



InSide

Innovation and Sicily Data Engine

Indicatori e correlazioni territoriali

Prima edizione

Indice

1. Introduzione	3
2. Dinamiche Territoriali in Sicilia: Un'Analisi del Biennio 2023-2024	4
3. Appendice Grafici con Interpretazione	9
4. Conclusioni e Prospettive per lo Sviluppo Territoriale Siciliano	89
5. Sintesi delle Dinamiche Socio-Economiche Emerse	90
6. Riflessioni sull'Evoluzione Temporale e Contestuale	91
7. Implicazioni Strategiche e Raccomandazioni	93
8. Considerazioni Finali	94

1. Introduzione

Il presente lavoro si concentra su un processo strutturato di analisi statistica automatizzata e generazione guidata di report descrittivi. L'obiettivo è offrire uno strumento che consenta di sintetizzare in maniera chiara, visiva e testuale l'informazione contenuta in dataset multivariati, con un'attenzione particolare all'evoluzione territoriale delle province siciliane rispetto al contesto nazionale.

Analisi dei dati

Una volta ottenuto il dataset, i dati vengono sottoposti a una fase preliminare di pulizia e normalizzazione, per garantire che tutte le variabili siano confrontabili e prive di distorsioni dovute a scale o valori mancanti.

Successivamente viene applicata una tecnica di riduzione della dimensionalità nota come PCA (Principal Component Analysis), che consente di proiettare il dataset originario in uno spazio bidimensionale. Tale trasformazione permette di mantenere la maggior parte della variabilità informativa e rende più agevole sia l'identificazione di pattern nei dati sia la loro rappresentazione grafica.

Su questa nuova rappresentazione viene eseguito un algoritmo di clustering non supervisionato, che individua gruppi omogenei di osservazioni. In particolare, è stato utilizzato l'algoritmo DBSCAN, che consente di identificare automaticamente gruppi di elementi simili tra loro e di riconoscere eventuali outlier.

Generazione automatica del report

Per ogni gruppo (cluster) individuato, il sistema calcola statistiche descrittive sintetiche che rappresentano le caratteristiche principali del gruppo stesso. Tali informazioni vengono poi fornite in input a un modello di linguaggio basato su intelligenza artificiale (GPT), che genera in maniera automatica un testo interpretativo in linguaggio naturale.

Il risultato è una descrizione testuale che illustra le peculiarità di ciascun gruppo, con attenzione alle differenze tra cluster e agli elementi comuni tra le osservazioni appartenenti allo stesso insieme.

Il report finale è composto da:

- Grafici delle proiezioni PCA, con evidenziazione visiva dei cluster;
- Tabelle con i dati trasformati e l'assegnazione di ciascuna osservazione a un gruppo;
- Descrizioni automatiche dei cluster;
- Evidenziazione specifica delle province siciliane all'interno delle visualizzazioni.

Obiettivo e vantaggi del sistema

L'intero processo consente di ottenere in modo automatizzato una sintesi visiva e testuale di dati complessi, rendendo accessibile l'analisi anche a figure non esperte di statistica o programmazione. La combinazione di metodi statistici avanzati e generazione automatica del linguaggio rende il sistema particolarmente adatto a contesti in cui sia necessaria una lettura rapida, interpretativa e territorialmente consapevole dei dati.

2. Dinamiche Territoriali in Sicilia: Un'Analisi del Biennio 2023-2024

Introduzione e Obiettivi dell'Analisi

Il presente report si propone di esplorare le tendenze e le interconnessioni tra diversi aspetti della vita socio-economica e ambientale nelle province siciliane, analizzando i dati disponibili per il biennio 2023-2024. L'obiettivo è quello di fornire una lettura chiara e accessibile delle dinamiche territoriali, evidenziando come determinati indicatori si influenzino reciprocamente e come queste relazioni evolvano nel tempo, con un focus sui cambiamenti più recenti. L'analisi mira a offrire spunti utili per comprendere meglio la realtà siciliana e orientare eventuali strategie di sviluppo territoriale.

Relazioni Dirette e Brevi Catene Causali

Le analisi dei dati rivelano diverse correlazioni significative che collegano indicatori chiave del territorio. Alcune di queste relazioni sono stabili nel tempo, mentre altre mostrano un rafforzamento o una debolezza nel periodo considerato.

Legami con il Mercato del Lavoro e il Benessere Sociale

Un tema ricorrente, particolarmente nel 2023 e 2024, è la forte interdipendenza tra il **tasso di occupazione** e una serie di altri

indicatori. Si osserva una chiara tendenza: dove il tasso di occupazione è elevato, si riscontra anche una maggiore **spesa delle famiglie**, un maggior numero di **beneficiari di reddito di cittadinanza (o assegni sociali)**, un minor numero di **giovani che non lavorano e non studiano (Neet)**, e una migliore **qualità della vita delle donne**.

- **Logica di Causalità Presunta:** Un mercato del lavoro dinamico e con buone prospettive di occupazione genera reddito disponibile, che a sua volta alimenta i consumi delle famiglie. Un maggior benessere generale e la percezione di opportunità lavorative possono contribuire a una riduzione della dipendenza da sussidi sociali e a una diminuzione della percentuale di giovani che non studiano né lavorano. La qualità della vita, intesa in senso lato, tende a migliorare in contesti economicamente più solidi.
- **Evoluzione Temporale:** Queste correlazioni appaiono robuste e confermate nel periodo analizzato. Nel 2024, in particolare, si nota un ulteriore rafforzamento del legame tra il **tasso di occupazione** e la **spesa delle famiglie**, suggerendo un aumento della fiducia e della capacità di spesa dei cittadini siciliani in aree con maggiore occupazione.

Il Divario Digitale e le sue Implicazioni

L'accesso a tecnologie avanzate, come la **banda larga** e la **banda ultra larga**, mostra una correlazione significativa con indicatori di sviluppo e modernizzazione. Dove la connettività digitale è più diffusa, si riscontrano tendenzialmente valori più

alti in termini di **depositi bancari** e, in un'ottica più ampia, un possibile impatto positivo sulla gestione finanziaria sia a livello individuale che aziendale.

- **Logica di Causalità Presunta:** Una maggiore penetrazione della banda larga favorisce l'adozione di servizi digitali, inclusi quelli bancari online. Questo può portare a una maggiore facilità di gestione del denaro, incentivare risparmi e investimenti, e contribuire a un'economia più digitalizzata.
- **Evoluzione Temporale:** La correlazione tra la disponibilità di banda larga e la presenza di depositi bancari si conferma, con un leggero rafforzamento nel 2023-2024, indicando una crescente importanza della digitalizzazione nei circuiti economici locali.

Sicurezza e Percezione di Incertezza

I dati del 2023 e 2024 evidenziano una forte connessione tra indicatori di criminalità e aspetti legati alla stabilità sociale e alla gestione delle cause legali. Si rileva una correlazione positiva tra **furti con strappo** e **rapine in pubblica via** con l'**indice di criminalità totale**. Allo stesso tempo, un elevato **indice di litigiosità** è strettamente legato alla **quota di cause pendenti ultratriennali**.

- **Logica di Causalità Presunta:** Un aumento dei reati predatori e una generale percezione di insicurezza tendono a correlare con un aumento della criminalità complessiva. La lentezza e la complessità del sistema giudiziario, evidenziate da un alto numero di cause pendenti e da un elevato indice di litigiosità, potrebbero

suggerire una società in cui le controversie sono frequenti e la risoluzione è problematica.

- **Evoluzione Temporale:** La correlazione tra indici di criminalità predatoria e l'indice di criminalità totale appare persistente. Nel 2024, si osserva un rafforzamento della relazione tra questi indicatori di insicurezza e la presenza di negozi e attività commerciali, suggerendo un impatto più diretto sulla vitalità economica delle aree urbane.

3. Dinamiche Territoriali Complesse

Oltre alle relazioni dirette, è possibile identificare percorsi di influenza più articolati che coinvolgono più indicatori, delineando scenari di sviluppo più complessi.

Il Ciclo della "Città Vivibile" e le Sfide Demografiche

Un insieme di fattori sembra descrivere un quadro di "città vivibile", dove la presenza di **ristoranti**, **librerie** e **medici di medicina generale** si lega a un contesto favorevole alla natalità e alla qualità della vita, ma che al contempo evidenzia sfide demografiche.

- **Percorso Logico:** La disponibilità di servizi culturali (ristoranti, librerie) e di assistenza sanitaria di base (medici di medicina generale) contribuisce a creare un ambiente urbano più attrattivo e vivibile. Tuttavia, questa vitalità sembra contrapporsi a un basso **quoziente di natalità** e a un elevato **indice di dipendenza degli anziani**. La presenza di molte **piste ciclabili** e un buon **ecosistema urbano** nel 2021 e 2022 sembrano

segnalare un'attenzione alla qualità ambientale e alla mobilità sostenibile, ma non compensano pienamente le criticità demografiche.

- **Evoluzione Temporale:** Sebbene la presenza di servizi sia generalmente positiva, nel 2021-2022 si nota una correlazione più stretta tra la presenza di questi servizi e un **tasso di natalità basso**. Questo potrebbe indicare che, nonostante la presenza di infrastrutture e servizi, altri fattori (come incertezza economica o sociale) scoraggiano la natalità. Il 2023 e 2024 vedono emergere una correlazione tra una minore natalità e una maggiore **solitudine**, suggerendo un possibile indebolimento dei legami sociali in alcune aree.

Crescita Economica e Disuguaglianze Territoriali

L'analisi dei dati mostra una correlazione tra la **retribuzione media annua** e il **valore aggiunto per abitante**, indicatori classici di benessere economico. Dove questi valori sono più alti, si osserva anche una maggiore **spesa delle famiglie**. Tuttavia, questo quadro di crescita economica non è esente da criticità legate alla **disuguaglianza del reddito netto** e alla **fragilità urbana**.

- **Percorso Logico:** Un aumento della retribuzione e del valore aggiunto per abitante in una provincia siciliana tendono a tradursi in una maggiore capacità di spesa da parte delle famiglie. Nel 2023-2024, emerge una connessione tra questa crescita economica e una maggiore presenza di **famiglie con ISEE basso**, suggerendo che i benefici della crescita non sono

distribuiti uniformemente, e un aumento dell'**indice di fragilità urbana**, che può indicare aree con maggiori difficoltà infrastrutturali e sociali nonostante la crescita economica generale. L'alto tasso di occupazione in queste aree è a volte accompagnato da un elevato numero di **beneficiari di reddito di cittadinanza** o assegni sociali, suggerendo un mercato del lavoro polarizzato.

- **Evoluzione Temporale:** Le correlazioni tra crescita economica e spesa delle famiglie rimangono forti. Nel 2024, si nota un rafforzamento del legame tra l'**indice di fragilità urbana** e il **tasso di occupazione**, indicando che anche in aree con occupazione, persistono criticità strutturali.

Sicurezza e Resilienza Sociale

I dati degli ultimi anni (2023-2024) mettono in luce un legame tra la percezione di sicurezza e la stabilità sociale. Un **indice di criminalità** elevato tende a correlare con un maggiore **indice di litigiosità** e un **tasso di mancata partecipazione al lavoro** più alto. Nel 2024, questo quadro è ulteriormente complicato dalla correlazione tra questo insieme di indicatori e un **alto indice di solitudine** e un **tasso di fecondità basso**.

- **Percorso Logico:** Un ambiente con maggiore criminalità e litigiosità può generare un senso di insicurezza e disagio sociale, che a sua volta potrebbe contribuire a un aumento della solitudine e a una diminuzione del desiderio di formare famiglie o avere figli. La difficoltà nell'inserimento lavorativo (tasso di

mancata partecipazione al lavoro) può esacerbare queste problematiche.

- **Evoluzione Temporale:** La connessione tra criminalità, litigiosità e mancata partecipazione al lavoro si rafforza nel 2023-2024. L'emergere di nuove correlazioni con la solitudine e la bassa natalità nel 2024 suggerisce un deterioramento delle condizioni sociali in alcune aree, che necessitano di attenzione specifica.

4. Conclusioni e Osservazioni Finali

L'analisi delle dinamiche territoriali in Sicilia per il biennio 2023-2024 rivela un panorama complesso e sfaccettato. Le correlazioni osservate evidenziano una forte interdipendenza tra il mercato del lavoro, il benessere sociale, la digitalizzazione, la sicurezza e la demografia.

Tendenze Costanti:

- Il legame tra un **tasso di occupazione** elevato e indicatori di benessere sociale (come maggiore spesa delle famiglie e minore percentuale di giovani Neet) è una costante, confermando l'importanza di un mercato del lavoro dinamico per la prosperità territoriale.
- La criminalità e la litigiosità tendono a correlare negativamente con la qualità della vita e la partecipazione sociale.

Cambiamenti Rilevanti:

- Il biennio 2023-2024 ha visto un **rafforzamento di molte correlazioni**, indicando una maggiore cristallizzazione di alcune dinamiche territoriali.

- È emersa una chiara tendenza nel 2023-2024 in cui la **crescita economica locale non sempre si traduce in una distribuzione equa del benessere**, come testimoniato dalla correlazione tra indicatori economici positivi e la presenza di famiglie con ISEE basso e un aumento dell'indice di fragilità urbana in alcune province.
- Le sfide demografiche, come il basso tasso di natalità e l'alto indice di dipendenza degli anziani, si sono mostrate persistenti e, in alcuni casi, accentuate, con nuove correlazioni emerse nel 2023-2024 tra questi fattori e la solitudine.

Implicazioni per lo Sviluppo Territoriale Siciliano:

1. **Investire nel Mercato del Lavoro:** Le politiche dovrebbero concentrarsi sulla creazione di posti di lavoro di qualità e sulla riduzione della disoccupazione giovanile, poiché questo ha un impatto diretto sul benessere delle famiglie e sulla dinamica economica generale.
2. **Colmare il Divario Digitale:** L'accelerazione nell'adozione della banda larga e ultra larga è cruciale per la modernizzazione economica e per facilitare l'accesso a servizi essenziali.
3. **Affrontare le Disuguaglianze:** Le politiche devono mirare a una distribuzione più equa dei benefici della crescita economica, affrontando la fragilità urbana e sostenendo le fasce di popolazione più vulnerabili.
4. **Promuovere la Coesione Sociale:** Le nuove correlazioni emerse nel 2023-2024, che legano insicurezza, solitudine e bassa

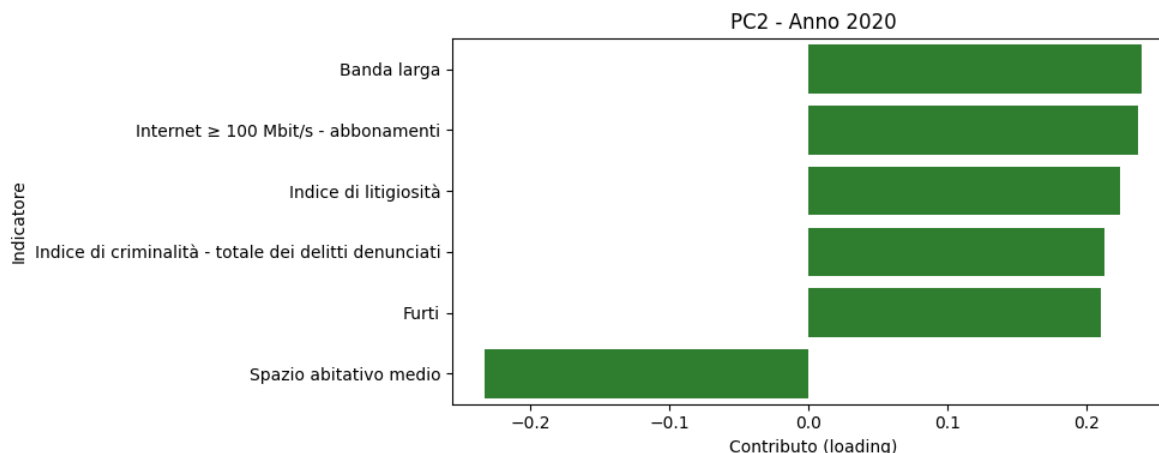
natalità, richiedono interventi mirati a rafforzare il tessuto sociale, migliorare la qualità della vita e promuovere un clima di fiducia.

5. Monitoraggio Continuo: L'evoluzione delle dinamiche territoriali suggerisce la necessità di un monitoraggio costante degli indicatori per adattare le strategie di intervento e rispondere tempestivamente ai cambiamenti in atto.

In sintesi, i dati suggeriscono che la Sicilia è un territorio in trasformazione, con aree che mostrano resilienza e progresso, ma anche con sfide significative che richiedono un'attenzione particolare per garantire uno sviluppo equo e sostenibile per tutti i suoi cittadini.

3. Analisi tecnica delle correlazioni su orizzonte quadriennale (2020-2024)

Barplot_PC2



L'immagine mostra un'analisi chiamata "PC2 - Anno 2020", che probabilmente si riferisce alla seconda componente principale di un'analisi multivariata dei dati socio-economici delle province italiane per l'anno 2020. Ogni barra orizzontale rappresenta un indicatore e la sua lunghezza, a destra o a sinistra dello zero, indica il "contributo" (o "loading") di quell'indicatore a questa seconda componente.

In parole semplici, questa "componente principale 2" raggruppa province che hanno caratteristiche simili rispetto a questi indicatori. Vediamo che ci sono due gruppi principali di indicatori:

- **Indicatori con contributo positivo (a destra dello zero):**
- **Banda larga:** Le province con valori elevati in questa componente tendono

ad avere un'ottima disponibilità di banda larga.

- **Internet $\geq$ 100 Mbit/s - abbonamenti:** Similmente, queste province hanno un alto numero di abbonamenti internet ad alta velocità.
- **Indice di litigiosità:** Questo indicatore ha un contributo positivo, anche se leggermente inferiore agli altri due. Significa che le province con un alto indice di litigiosità sono associate a questa componente positiva.
- **Indice di criminalità - totale dei delitti denunciati:** Anche questo ha un contributo positivo, simile all'indice di litigiosità.
- **Furti:** Questo è l'ultimo indicatore con contributo positivo, con un valore appena sopra lo zero.
- **Indicatori con contributo negativo (a sinistra dello zero):**

- **Spazio abitativo medio:** Questo è l'unico indicatore con un forte contributo negativo. Significa che le province con valori elevati in questa componente hanno, invece, uno spazio abitativo medio basso.

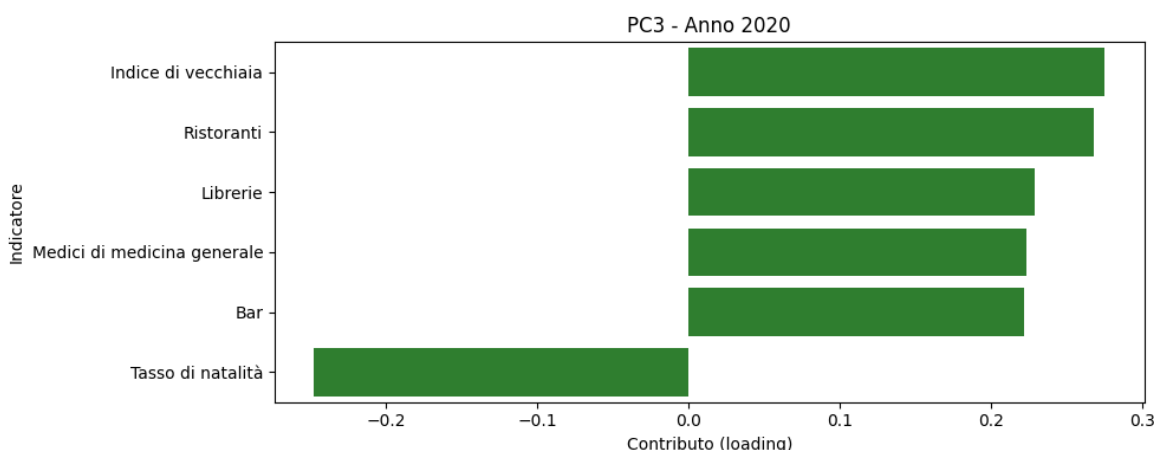
Cosa possiamo osservare da questi pattern?

Questa seconda componente principale sembra catturare una sorta di contrapposizione. Da un lato, abbiamo province che sono **ben connesse digitalmente** (banda larga, internet

veloce) e che presentano anche una **maggiore incidenza di criminalità e litigiosità**. Dall'altro lato, queste stesse province tendono ad avere uno **spazio abitativo medio più ristretto**.

In sintesi, la "PC2" sembra descrivere un fenomeno dove l'alta connettività digitale e un certo livello di disordine sociale (criminalità, litigiosità) si accompagnano a una minore disponibilità di spazio abitativo. Potrebbe indicare aree urbane o aree con un certo grado di sovrappollamento e con buone infrastrutture digitali.

Barplot_PC3



Questo grafico, intitolato "PC3 - Anno 2020", ci mostra come diversi indicatori socio-economici contribuiscono a definire una particolare dimensione o "componente" dei dati provinciali italiani per l'anno 2020. In termini semplici, è come se stessimo cercando di capire quale mix di caratteristiche rende una provincia "tipo" in questa specifica dimensione identificata.

Cosa rappresenta:

- **Indicatori (asse verticale):** Sulla sinistra troviamo diversi indicatori che riguardano la popolazione e i servizi:
- Indice di vecchiaia (quante persone anziane rispetto a quelle giovani)
- Ristoranti (numero di ristoranti)
- Librerie (numero di librerie)
- Medici di medicina generale (numero di medici)
- Bar (numero di bar)

- Tasso di natalità (quante nascite per 1000 abitanti)
- **Contributo (loading) (asse orizzontale):** La barra verde indica quanto ciascun indicatore "pesa" o contribuisce a definire questa componente (chiamata PC3).
- **Barre a destra dello zero (valori positivi):** Significano che un valore alto di quell'indicatore è associato a un punteggio alto in questa componente PC3.
- **Barre a sinistra dello zero (valori negativi):** Significano che un valore basso di quell'indicatore è associato a un punteggio alto in questa componente PC3 (o, equivalentemente, un valore alto dell'indicatore è associato a un punteggio basso nella PC3).
- **Barre più lunghe:** Indicano un contributo maggiore, sia positivo che negativo.

Evidenze e pattern che si notano:

1. Componente legata alla "vitalità" e ai servizi: Sembra che questa componente PC3 descriva qualcosa legato alla struttura demografica e alla presenza di attività commerciali e servizi.

- **Indicatori con forte contributo positivo:** "Indice di vecchiaia", "Ristoranti", "Librerie", "Medici di medicina generale" e "Bar" hanno tutti un contributo positivo. Questo suggerisce che le province che hanno un punteggio alto in questa componente tendono ad avere più anziani, più ristoranti, più librerie, più medici di base e più bar. In un certo senso, potremmo interpretarlo come una caratteristica di aree forse più

"sviluppate" o con una certa densità di popolazione e servizi, dove la popolazione è anche più anziana.

- **Indicatore con forte contributo negativo:** Il "Tasso di natalità" ha un contributo nettamente negativo. Questo conferma l'interpretazione precedente: le province che hanno un alto punteggio in questa componente PC3 (cioè, sono "ricche" di anziani, ristoranti, ecc.) tendono ad avere un basso tasso di natalità. Al contrario, le province con un basso punteggio in questa componente avranno probabilmente più nascite e meno anziani, ristoranti, ecc.

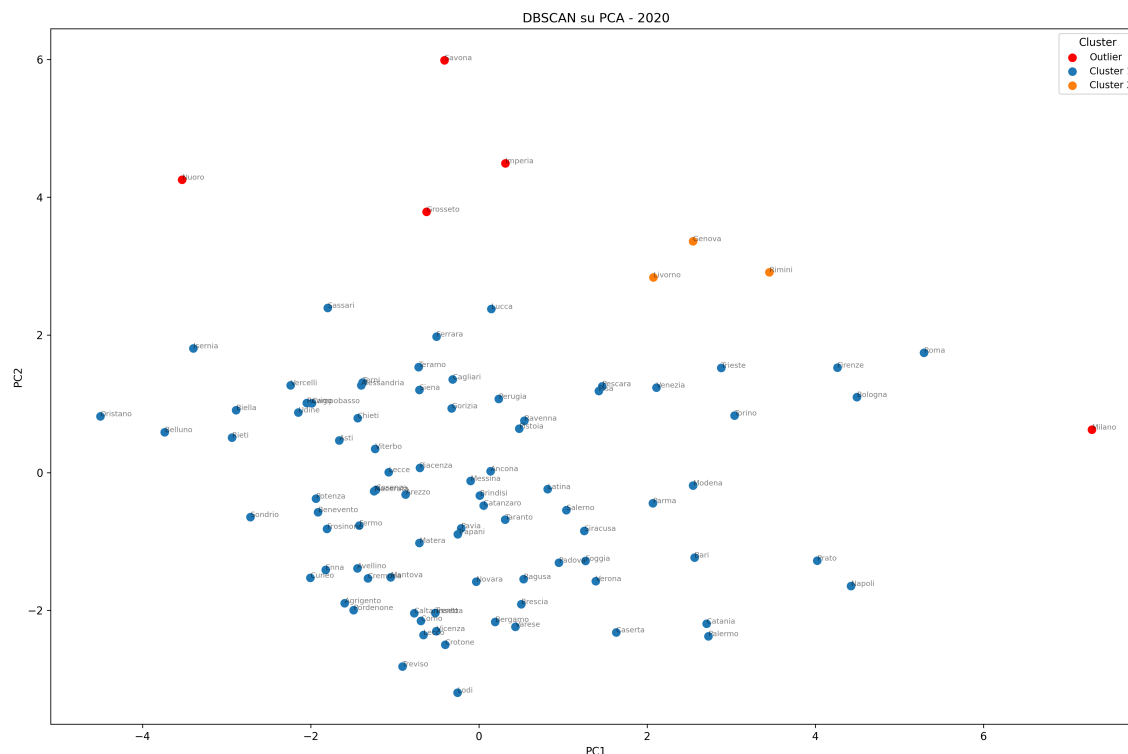
In sintesi per un analista territoriale:

Questa componente PC3 del 2020 ci dice che le province italiane si differenziano lungo un asse dove da un lato troviamo territori caratterizzati da una **popolazione più anziana**, una buona **presenza di servizi e attività commerciali** come ristoranti, librerie e bar, e un numero adeguato di medici di base. Dall'altro lato, si collocano le province con un **maggiore dinamismo demografico** (più nascite) e, presumibilmente, una minore densità di questi specifici servizi e una popolazione mediamente più giovane.

Potremmo definirla una componente che contrappone contesti di **"invecchiamento e consolidamento dei servizi"** a contesti di **"gioventù e potenziale di crescita demografica"**. Questo è un pattern classico che potremmo ritrovare nel divario tra centri urbani più maturi o aree turistiche consolidate (con più ristoranti, anziani) e aree più periferiche o rurali con

famiglie giovani o zone di nuova espansione (con più nascite).

dbscan_pca_2020



Questa immagine rappresenta una sintesi visuale di dati socio-economici delle province italiane nel 2020, elaborata tramite una tecnica chiamata DBSCAN applicata a dati ridotti con PCA (Analisi delle Componenti Principali). In pratica, abbiamo preso tante informazioni (PIL, disoccupazione, istruzione, ecc.) per ogni provincia e le abbiamo "schiacciate" in due dimensioni principali (PC1 e PC2) per poterle disegnare qui.

Ogni puntino sull'asse X (PC1) e sull'asse Y (PC2) rappresenta una provincia italiana. La posizione di un puntino indica come quella provincia si posiziona rispetto a queste due dimensioni socio-economiche. Le province che sono vicine tra loro in

questo grafico tendono ad avere caratteristiche socio-economiche simili.

Abbiamo identificato tre gruppi principali:

- **Cluster 1 (punti blu):** Questo è il gruppo più numeroso e rappresenta la maggior parte delle province italiane. Sono le province che si collocano in una fascia "media" per quanto riguarda le caratteristiche socio-economiche considerate. Non spiccano né in positivo né in negativo rispetto alla media generale delle province.
- **Cluster 2 (punti arancioni):** Questo gruppo è più piccolo e comprende province come Livorno, Rimini, Genova e Vicenza. Queste province mostrano caratteristiche socio-

economiche che le distinguono dal gruppo più numeroso, posizionandole in un modo specifico all'interno del panorama italiano. Ad esempio, potrebbero avere indicatori socio-economici più alti o più bassi rispetto alla media, ma in una combinazione particolare.

- **Outlier (punti rossi):** Questi sono i puntini rossi: Nuoro, Savona, Grosseto, Imperia e Milano. Le province etichettate come "Outlier" sono quelle che si discostano significativamente dalla maggior parte delle altre province. Hanno profili socio-economici unici e non rientrano facilmente negli schemi definiti dagli altri cluster. Milano, in particolare, è un outlier molto pronunciato, indicando che le sue caratteristiche socio-economiche sono nettamente diverse da quelle della maggior parte delle province italiane.

Evidenze e pattern che si notano:

1. **Concentrazione della maggioranza:**

La maggior parte delle province (Cluster 1) si aggrega in una zona centrale del grafico, dimostrando una certa omogeneità socio-economica a livello nazionale.

2. **Polarizzazione di alcune aree:** Si notano delle province che si posizionano in modo più isolato. Il Cluster 2 sembra indicare un'altra "configurazione" socio-economica, distinta dal gruppo principale.

Ad esempio, Livorno e Rimini sono abbastanza vicine tra loro, suggerendo similitudini.

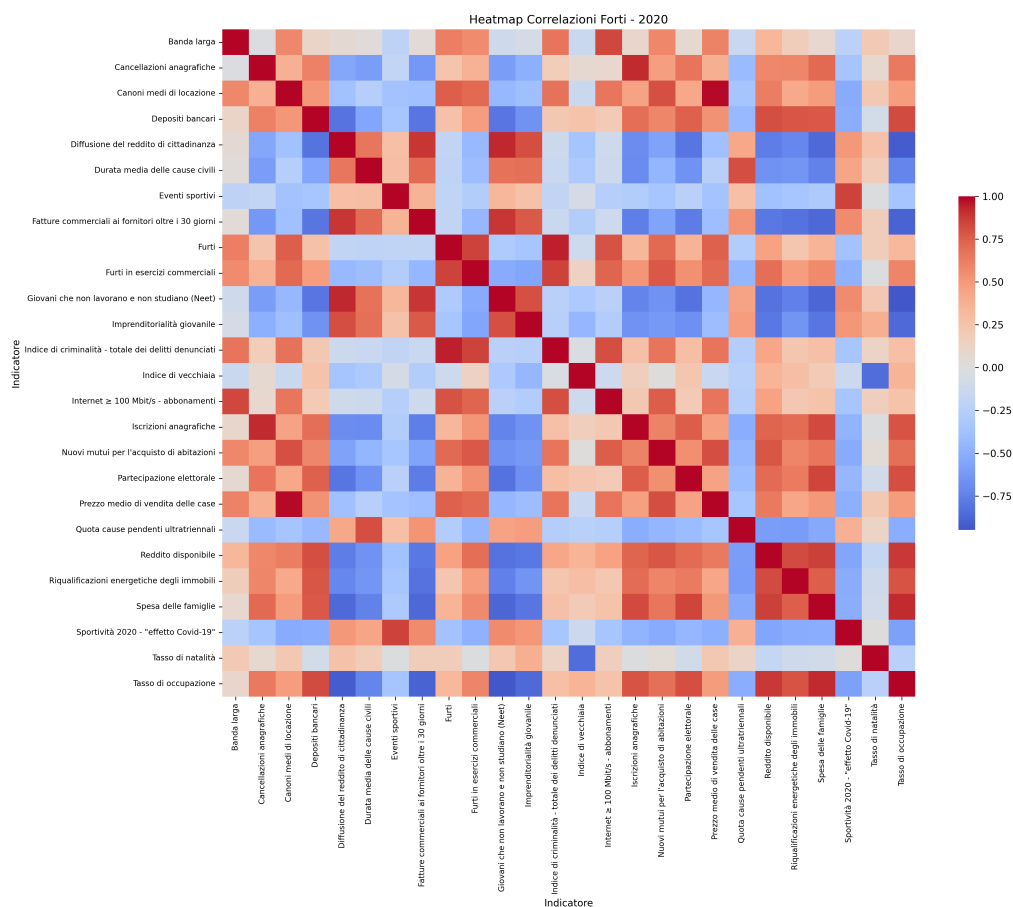
3. **Province uniche (Outlier):** Milano spicca in modo netto sulla destra, suggerendo un profilo socio-economico decisamente "elevato" o diverso rispetto alla media nazionale. Le altre province rosse (Nuoro, Savona, Grosseto, Imperia) sono anch'esse isolate, ma in posizioni diverse, suggerendo caratteristiche peculiari che non le accomunano né al gruppo principale né al Cluster 2. Grosseto e Imperia, ad esempio, si trovano in una zona simile, mentre Nuoro è più a nord-ovest.

4. **Distribuzione geografica implicita:**

Anche se il grafico non mostra direttamente la geografia, la posizione delle province potrebbe riflettere tendenze territoriali. Ad esempio, il fatto che Milano sia un outlier così marcato potrebbe essere legato al suo ruolo di polo economico. Bisognerebbe poi correlare queste posizioni con la mappa dell'Italia per capire meglio se ci sono raggruppamenti geografici che condividono queste caratteristiche.

In sintesi, questo grafico ci aiuta a identificare gruppi di province con profili socio-economici simili e a individuare quelle che si discostano dalla norma, offrendo uno spunto per analisi più approfondite sui fattori che determinano queste differenze.

heatmap_correlazioni_2020



Questa immagine è una "heatmap" che mostra la correlazione tra diversi indicatori socio-economici a livello provinciale in Italia nel 2020. In termini semplici, ogni quadrato colorato ci dice quanto due indicatori tendono a muoversi insieme.

Cosa rappresenta:

- **Le righe e le colonne:** Ogni riga e ogni colonna rappresenta un indicatore diverso. Pensa a questi indicatori come a diverse caratteristiche delle province italiane, come la disponibilità di banda larga, il tasso di criminalità, il reddito disponibile, il prezzo delle case, ecc.

- **I colori:** Il colore di ogni quadrato (l'incrocio tra una riga e una colonna) ci dice la forza e la direzione della relazione tra i due indicatori corrispondenti.
- **Rosso intenso:** Indica una correlazione positiva molto forte. Significa che quando un indicatore aumenta, anche l'altro tende ad aumentare significativamente.
- **Rosso chiaro/arancione:** Indica una correlazione positiva moderata o debole.
- **Bianco/grigio chiaro:** Indica una correlazione molto debole o quasi nulla. I due indicatori non sembrano influenzarsi molto a vicenda.

- **Blu chiaro:** Indica una correlazione negativa debole. Quando un indicatore aumenta, l'altro tende a diminuire leggermente.
- **Blu intenso:** Indica una correlazione negativa molto forte. Significa che quando un indicatore aumenta, l'altro tende a diminuire significativamente.
- **La barra a destra (scala di colori):** Questa barra ci aiuta a interpretare i colori. I valori più alti (vicini a 1.00) rappresentano correlazioni positive forti, mentre i valori più bassi (vicini a -0.75) rappresentano correlazioni negative forti. Il valore 0.00 indica nessuna correlazione.

Evidenze e pattern notati:

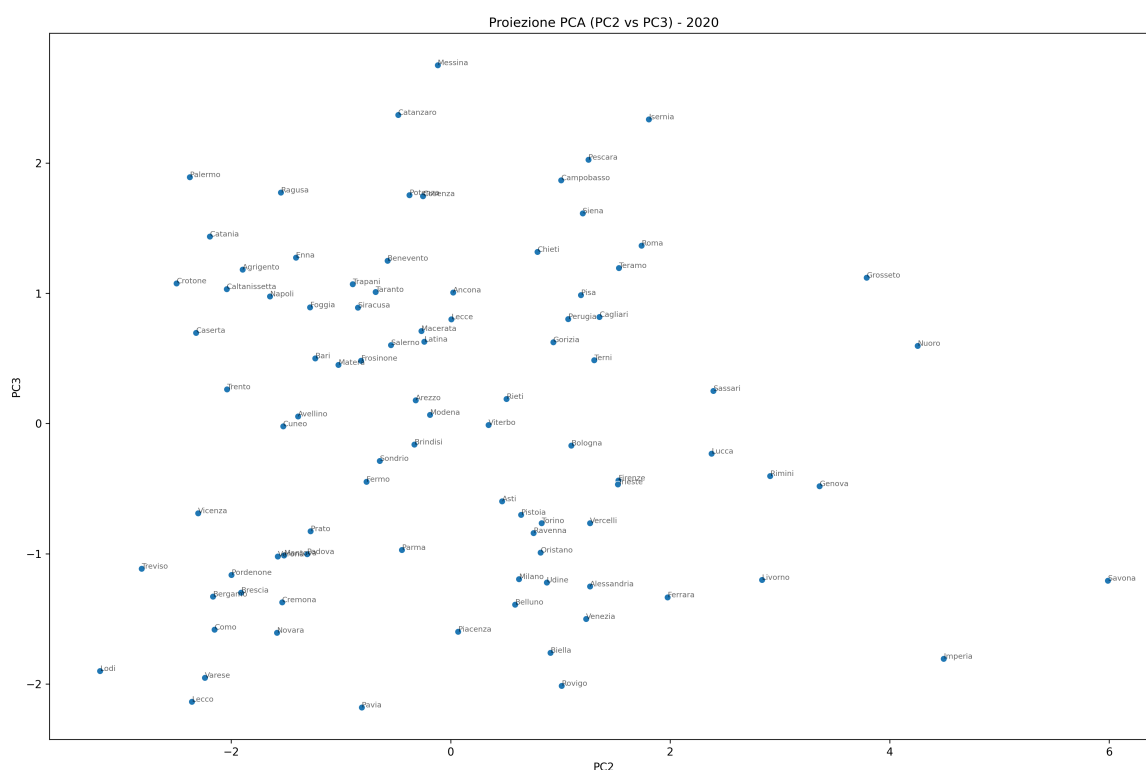
Osservando la mappa, possiamo notare diversi raggruppamenti e tendenze:

- **Indicatori legati all'economia e al benessere:** Si nota una forte correlazione positiva tra alcuni indicatori che suggeriscono prosperità economica. Ad esempio, "Prezzo medio di vendita delle case", "Depositi bancari" e "Reddito disponibile" sembrano essere strettamente legati. Le province con prezzi delle case più alti tendono anche ad avere più depositi bancari e redditi disponibili maggiori. Anche "Nuovi mutui per l'acquisto di abitazioni" si lega bene a questi.
- **Indicatori di attività economica e di spostamenti:** C'è una correlazione positiva piuttosto forte tra "Banda larga" e "Internet $\geq$ 100 Mbit/s - abbonamenti", il che è atteso. Interessa notare che la disponibilità di banda larga sembra correlata anche con alcuni indicatori di attività economica, come forse "Eventi sportivi" e persino "Sportività 2020 - 'effetto Covid-19'". Questo potrebbe suggerire che dove c'è buona connettività, c'è anche più propensione a determinate attività o stili di vita.
- **Indicatori di criminalità e difficoltà:** Si osservano correlazioni negative tra "Indice di criminalità - totale dei delitti denunciati" e alcuni indicatori di benessere. Ad esempio, una maggiore criminalità tende ad essere associata a un minore "Reddito disponibile" e a un minor "Prezzo medio di vendita delle case", indicando che le aree più difficili potrebbero presentare anche indicatori economici meno favorevoli. Anche "Furt" e "Furt in esercizi commerciali" mostrano correlazioni negative con indicatori di benessere.
- **Indicatori demografici e legali:** "Iscrizioni anagrafiche" e "Tasso di natalità" sembrano correlati positivamente, come ci si aspetterebbe. La "Durata media delle cause civili" mostra correlazioni più variegata, suggerendo che non sia un indicatore direttamente collegato a quelli di benessere o criminalità in modo netto.
- **Indicatori di disoccupazione/inattività:** "Giovani che non lavorano e non studiano (Neet)" e "Imprenditorialità giovanile" potrebbero mostrare relazioni interessanti, ma a prima vista sembrano più "isolati" in termini di correlazioni fortissime e nette con altri indicatori principali. Tuttavia, vale la pena indagare le correlazioni più deboli.

In sintesi, la heatmap ci offre una visione d'insieme su come diversi aspetti della vita economica e sociale delle province italiane siano interconnessi. Possiamo facilmente identificare cluster di province con caratteristiche simili basandoci su questi

pattern di correlazione. Ad esempio, le province con un'economia fiorente tendono ad avere prezzi immobiliari più alti, più depositi bancari e un reddito disponibile maggiore, mentre potrebbero avere tassi di criminalità inferiori.

scatter_pca_2020



Questo grafico mostra una proiezione delle province italiane in uno spazio a due dimensioni, ottenuto tramite un'analisi multivariata dei dati socio-economici del 2020. Immagina di avere tantissime informazioni su ogni provincia (reddito, occupazione, istruzione, ecc.). Questo grafico riduce tutta quella complessità in due "direzioni" principali, chiamate PC2 e PC3, che cercano di catturare la maggior parte delle differenze tra le province.

In pratica, ogni punto rappresenta una provincia, e la sua posizione ci dice qualcosa sulla sua "somiglianza" con altre province secondo queste due dimensioni. Province vicine tra loro tendono ad avere caratteristiche socio-economiche più simili, mentre quelle lontane sono più dissimili.

Cosa possiamo notare:

- **Un asse Est-Ovest (PC2):** L'asse orizzontale (PC2) sembra distinguere

tra province. Le province con valori più alti sull'asse PC2 (come Imperia, Savona, Livorno, Grosseto, ma anche alcune del Nord-Est come Udine, Belluno, Ravenna) sembrano avere caratteristiche diverse rispetto a quelle con valori più bassi (come molte province del Sud e alcune del Centro Italia come Viterbo, Rieti, Foggia). Senza sapere esattamente cosa misurano PC2 e PC3, possiamo ipotizzare che questo asse possa correlare con fattori come la ricchezza, lo sviluppo economico o forse la vicinanza a centri urbani importanti.

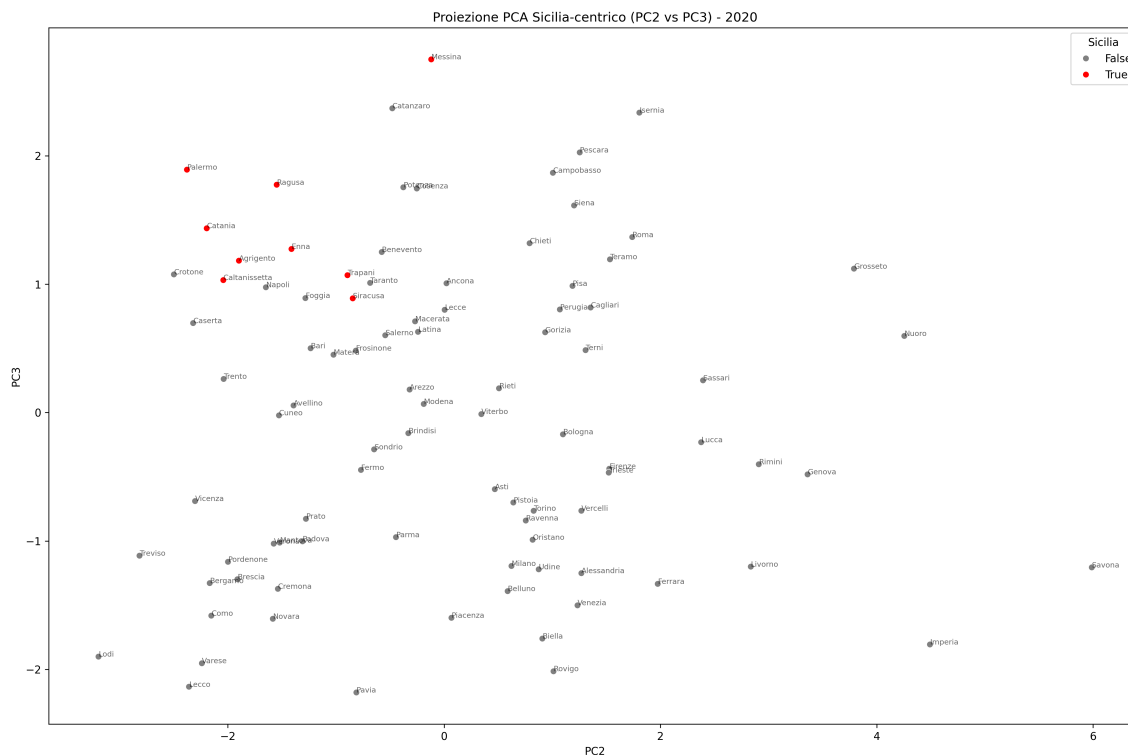
- **Un asse Nord-Sud (PC3):** L'asse verticale (PC3) sembra separare ulteriormente le province. Le province più in alto sull'asse PC3 (come Messina, Catanzaro, Agrigento, Caltanissetta, ma anche Roma, Campobasso) mostrano delle peculiarità rispetto a quelle più in basso (come Lodi, Varese, Lecco, Vicenza, Pordenone). Potrebbe essere che PC3 catturi differenze legate alla struttura economica, ad esempio un maggior peso del settore terziario o industriale al Nord, e un'economia più legata all'agricoltura o al turismo in altre aree.
- **Cluster Geografici (non perfetti):** Pur non essendo una mappa geografica, notiamo delle tendenze. Le province del Nord Italia tendono a concentrarsi nella parte alta e destra

del grafico, suggerendo un profilo socio-economico comune. Le province del Sud Italia tendono invece a posizionarsi più a sinistra e nella parte centrale-bassa del grafico. Le province del Centro Italia sono più sparse, mostrando una maggiore eterogeneità.

- **Esempi di "isole":** Alcune province si distinguono nettamente dalle altre per la loro posizione. Ad esempio, Imperia e Savona sono molto isolate verso destra, suggerendo caratteristiche socio-economiche uniche. Allo stesso modo, Messina e Catanzaro sono abbastanza distanti da altri gruppi.
- **Confronto Città Metropolitane:** Le grandi città metropolitane come Roma, Milano, Napoli, Torino sembrano avere posizioni che riflettono la loro centralità e complessità. Milano e Torino si trovano nella zona "nord-occidentale" del grafico. Roma è un po' più centrale. Napoli è più a sinistra, suggerendo una differenziazione socio-economica rispetto ai centri del Nord.

In sintesi, questo grafico ci offre una visione compressa e intuitiva delle somiglianze e differenze socio-economiche tra le province italiane nel 2020, permettendoci di identificare gruppi di province con profili simili e di osservare quali province si discostano maggiormente da questi profili.

scatter_pca_2020_sicilia



Questa immagine mostra le province italiane proiettate su due dimensioni, PC2 e PC3, che sono una sintesi di diversi dati socio-economici. Ogni punto rappresenta una provincia, con il suo nome etichettato.

La distinzione principale che salta all'occhio è il colore dei punti:

- **Grigio:** Indica le province che **non sono in Sicilia** (etichettato come "False" nella legenda).
- **Rosso:** Indica le province **in Sicilia** (etichettato come "True" nella legenda).

Cosa ci dice questa visualizzazione?

1. Posizionamento delle province siciliane: Le province rosse (Palermo, Messina, Agrigento, Caltanissetta, Catania, Enna, Ragusa, Siracusa, Trapani) sono raggruppate in un'area specifica del

grafico, tendenzialmente nella parte alta e sinistra. Questo suggerisce che, rispetto alle altre province italiane analizzate, condividono delle caratteristiche socio-economiche simili che le posizionano in questa zona.

2. Separazione tra Sicilia e altre province: Sembra esserci una certa separazione visiva tra le province siciliane e la maggior parte delle altre province. Le province grigie sono distribuite su tutto lo spazio del grafico, ma quelle rosse si concentrano in una nicchia. Questo indica che, in base ai dati analizzati e sintetizzati in queste due dimensioni, le province siciliane formano un gruppo con tendenze distinte.

3. Pattern interni al gruppo siciliano: All'interno del gruppo di province siciliane rosse, notiamo che non sono tutte

esattamente nello stesso punto, ma presentano una loro micro-variabilità. Ad esempio, Messina e Palermo si posizionano leggermente più in alto rispetto ad altre come Agrigento o Caltanissetta. Questo potrebbe riflettere sottili differenze tra queste province all'interno della regione stessa.

4. Relazione con gli assi (PC2 e PC3):

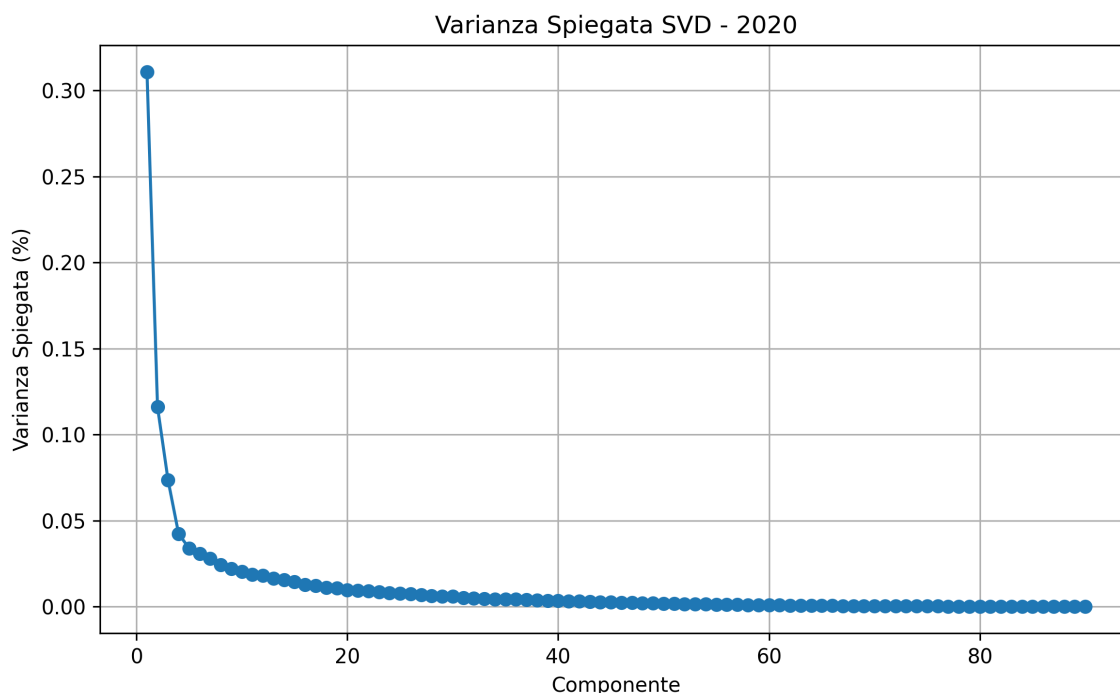
- L'asse **PC2** (orizzontale) sembra differenziare le province che si trovano più a sinistra (valori negativi) da quelle più a destra (valori positivi). Le province siciliane tendono ad avere valori di PC2 più bassi, indicando che sono in quella parte dello spazio che

rappresenta certe combinazioni di caratteristiche.

- L'asse **PC3** (verticale) fa lo stesso per le posizioni in alto (valori positivi) e in basso (valori negativi). Le province siciliane tendono anche ad avere valori di PC3 più alti rispetto alla media generale.

In sintesi, l'immagine ci mostra che le province siciliane, considerate nel loro insieme, tendono a posizionarsi in una determinata area di questo spazio socio-economico sintetico (con valori più alti sull'asse PC3 e più bassi sull'asse PC2), distinguendosi da molte altre province italiane. Le province grigie, invece, mostrano una maggiore eterogeneità di posizionamento.

svd_varianza_spiegata_2020



Questa immagine ci mostra quanto "peso" hanno le diverse dimensioni che emergono dall'analisi dei dati socio-

economici delle province italiane nel 2020. In pratica, è come se stessimo cercando di riassumere tante informazioni su ogni provincia (reddito, occupazione, istruzione, ecc.) in un numero minore di "fattori" o "componenti" che catturano la maggior parte delle differenze tra le province.

Cosa rappresenta il grafico:

- **Asse orizzontale (Componente):** Qui vediamo i diversi "fattori" che sono emersi dall'analisi, ordinati dal più importante al meno importante. Il primo componente è quello che spiega più differenze tra le province, il secondo meno, e così via.
- **Asse verticale (Varianza Spiegata (%)):** Questo indica quanta della differenza totale tra le province è spiegata da ciascun componente. Una componente che spiega il 30% significa che è responsabile del 30% delle variazioni osservate nei dati socio-economici.

Evidenze e pattern che si notano:

1. Dominanza dei primi componenti: Si nota chiaramente che i primi pochi componenti spiegano una parte molto consistente della varianza totale. Il primo componente da solo spiega oltre il 30% delle differenze tra le province. Il secondo e il terzo componente aggiungono ancora una buona fetta, arrivando complessivamente a spiegare circa il 45-50% (sommando i primi tre). Questo è un segnale molto forte.

2. Crollo rapido della varianza: Dopo i primi tre o quattro componenti, la quantità di varianza spiegata da ciascun componente successivo diminuisce

drasticamente e molto velocemente. Diventa quasi piatta.

3. Effetto "gomito" o "scree plot": La curva ha una forma tipica che ricorda un "gomito" o una "frattura". Questa forma indica che c'è una chiara separazione tra i componenti "principali" (quelli che spiegano molto) e quelli "secondari" (quelli che spiegano poco).

4. Molti componenti con poca informazione: Ci sono circa 90 componenti totali che sono stati analizzati (lo vediamo perché la linea arriva fino a circa 90 sull'asse orizzontale). Tuttavia, dopo i primissimi, quasi tutti gli altri componenti spiegano una quantità minima di varianza (meno dell'1% ciascuno).

Cosa ci dice questo dal punto di vista territoriale:

Questo grafico ci suggerisce che la complessità dei dati socio-economici delle province italiane nel 2020 può essere riassunta efficacemente con un numero relativamente piccolo di dimensioni. In altre parole, la maggior parte delle differenze che osserviamo tra le province in termini di indicatori socio-economici possono essere ricondotte a pochi grandi "fattori" sottostanti.

Ad esempio, potremmo ipotizzare che il primo componente rappresenti un indice generale di "sviluppo socio-economico", dove province con valori alti in questo componente sono quelle più ricche, con alta occupazione, alto livello di istruzione, ecc. Il secondo componente potrebbe rappresentare un altro aspetto, magari legato alla struttura economica (es. peso

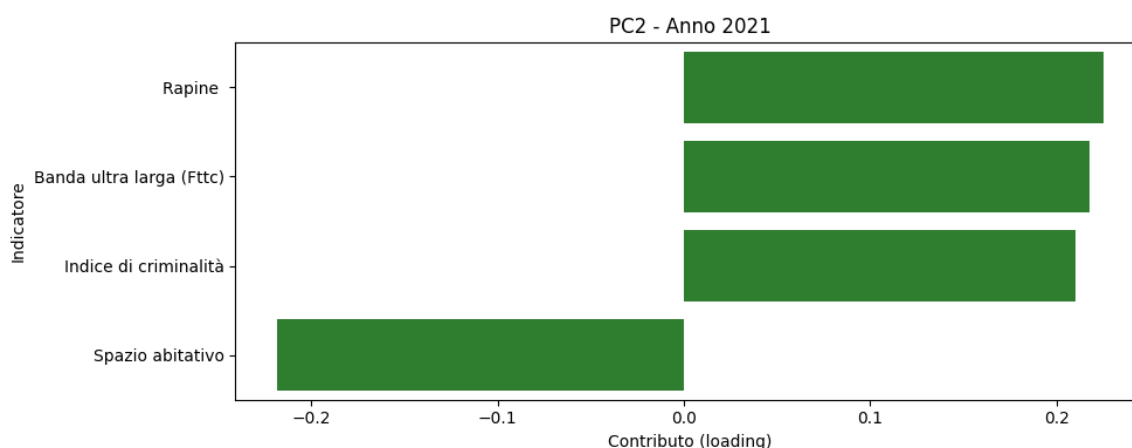
dell'industria vs. servizi) o alla demografia, e così via.

La rapida diminuzione della varianza suggerisce che scegliere solo i primi 3-5 componenti potrebbe già darci un quadro molto completo della situazione, riducendo la complessità senza perdere informazioni essenziali. I componenti successivi, invece, tendono a catturare sfumature molto specifiche o persino

rumore nei dati, e sono meno utili per una visione d'insieme.

Per un analista territoriale, questo significa che potremmo focalizzare la nostra analisi e le nostre strategie territoriali su queste poche dimensioni chiave, piuttosto che dover gestire decine di indicatori separati. Ci aiuta a semplificare la narrazione e a identificare i driver principali delle disuguaglianze o delle eccellenze tra le province.

barplot_PC2



Questa immagine mostra il contributo di diversi indicatori socio-economici a una componente principale (PC2) specifica, basata sui dati del 2021. In altre parole, questa "componente principale" è una sorta di riassunto statistico che cerca di cogliere una tendenza comune tra più indicatori.

Ecco cosa vediamo nel dettaglio:

- **Indicatori sull'asse verticale:** Sono elencati gli indicatori analizzati: "Rapine", "Banda ultra larga (Fttc)", "Indice di criminalità" e "Spazio abitativo".

- **Contributo (loading) sull'asse orizzontale:** Questa scala indica quanto ciascun indicatore contribuisce a definire la componente principale PC2. Valori più vicini allo zero indicano un contributo minore, mentre valori più distanti dallo zero (sia positivi che negativi) indicano un contributo più forte.

Cosa ci dicono i pattern:

- **Indicatori di criminalità e connettività in aumento:** Le barre per "Rapine", "Banda ultra larga (Fttc)" e "Indice di criminalità" sono tutte positive e si estendono verso

destra. Questo significa che un aumento in questi indicatori è associato a un valore più alto della componente PC2. In particolare, "Rapine" e "Banda ultra larga" hanno un contributo abbastanza simile e consistente, mentre l'"Indice di criminalità" è leggermente meno marcato. Possiamo quindi dire che le province che hanno valori elevati in questi tre indicatori tendono a posizionarsi in una certa direzione all'interno di questo riassunto statistico.

- **Spazio abitativo in diminuzione:** Al contrario, la barra per "Spazio abitativo" è negativa e si estende verso sinistra. Ciò implica che un aumento nello spazio abitativo per provincia è associato a un valore più basso della componente PC2. Le province che hanno più spazio abitativo pro capite tendono a posizionarsi in modo opposto rispetto a quelle con più rapine o più copertura di banda larga.

In sintesi per un analista territoriale:

La componente PC2 sembra catturare una dicotomia tra aree con caratteristiche

"urbane" o con problemi legati alla criminalità e alla connettività digitale, e aree con più spazio e forse meno densamente popolate o con caratteristiche rurali.

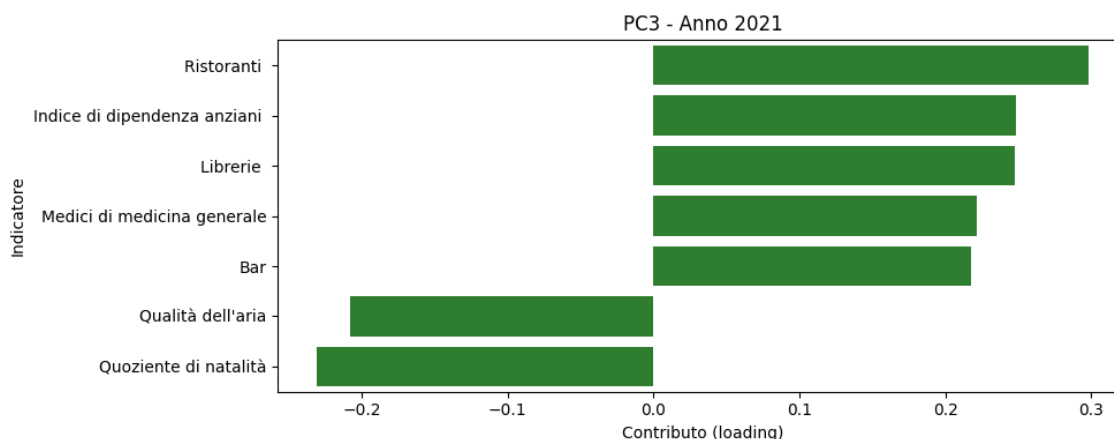
Le province che si posizionano in un estremo di questo asse (dove la PC2 è alta) potrebbero essere caratterizzate da:

- Una maggiore incidenza di rapine.
- Una buona diffusione della banda larga.
- Un indice di criminalità tendenzialmente più elevato.
- Uno spazio abitativo pro capite minore.

Al contrario, le province che si posizionano sull'altro estremo (dove la PC2 è bassa) avrebbero caratteristiche opposte: meno rapine, minore diffusione della banda larga, indice di criminalità più basso e più spazio abitativo.

Questo tipo di analisi ci aiuta a capire quali sono i fattori che variano insieme e che definiscono delle tendenze comuni tra le province italiane, permettendoci di raggrupparle o distinguerle in base a questi pattern.

barplot_PC3



Questa immagine mostra una rappresentazione di come diversi indicatori socio-economici contribuiscono a definire un "fattore" (denominato PC3) per le province italiane nell'anno 2021. In pratica, stiamo guardando quali elementi sono più legati tra loro per caratterizzare una certa dimensione di analisi delle province.

Gli indicatori sono elencati sull'asse verticale e il loro "contributo" (o peso) al fattore PC3 è mostrato sull'asse orizzontale. I valori positivi indicano che un aumento di quell'indicatore è associato a un valore più alto del fattore PC3, mentre i valori negativi indicano il contrario.

Cosa notiamo:

- **Indicatori con un forte contributo positivo:** "Ristoranti", "Indice di dipendenza anziani", "Librerie", "Medici di medicina generale" e "Bar" hanno tutti un contributo positivo al PC3. In particolare, "Ristoranti" è l'indicatore che contribuisce maggiormente (circa 0.3). Questo suggerisce che le province con più ristoranti, un maggior numero di anziani a carico della popolazione in

età lavorativa, più librerie, più medici di base e più bar tendono ad avere un valore più alto per questo fattore PC3. Potremmo interpretare questo fattore come legato alla **"vivacità dei servizi e della popolazione più anziana"** o un **"modello di sviluppo incentrato sui servizi e con una popolazione che tende a invecchiare"**.

- **Indicatori con un contributo negativo:** "Qualità dell'aria" e "Quoziente di natalità" hanno un contributo negativo. "Qualità dell'aria" ha un valore di circa -0.15 e "Quoziente di natalità" di circa -0.22. Questo significa che un miglioramento della qualità dell'aria e un aumento del tasso di natalità sono associati a un valore più basso di questo fattore PC3. In altre parole, le province con un'aria più pulita e più nascite tendono ad avere un valore più basso nel PC3.

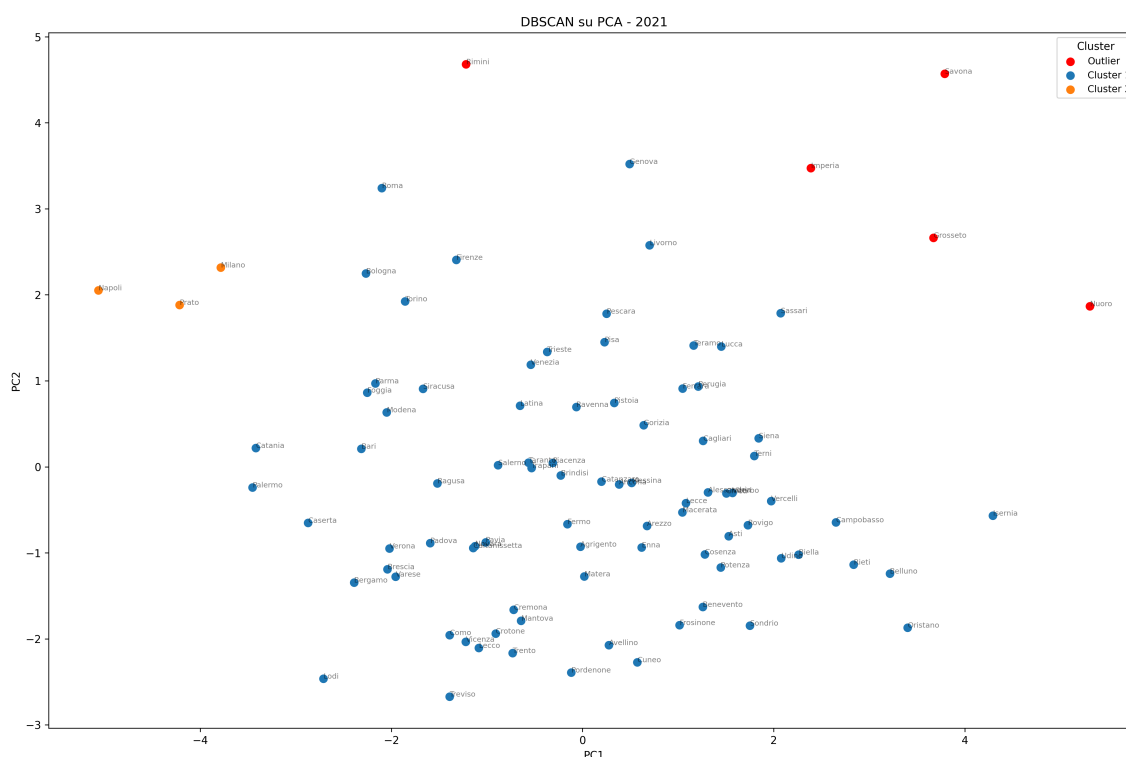
In sintesi per un analista territoriale:

Questo PC3 sembra catturare una dimensione che distingue le province: da un lato quelle con una forte presenza di attività legate ai servizi, una popolazione più anziana e una certa concentrazione di

servizi sanitari e ricreativi (ristoranti, bar, librerie, medici), e dall'altro quelle che potrebbero avere caratteristiche demografiche più dinamiche (più nascite) e una migliore qualità ambientale. Le province che si posizionano "in alto" su

questo fattore sono quelle con molti ristoranti, un alto tasso di dipendenza degli anziani, molte librerie, medici e bar. Al contrario, le province che si posizionano "in basso" sono quelle con un'aria migliore e un quoziente di natalità più alto.

dbscan_pca_2021



L'immagine rappresenta una sintesi visiva dei dati socio-economici di 70 province italiane, basata su un'analisi che ne ha raggruppato le caratteristiche. Ogni punto indica una provincia, posizionata secondo due dimensioni principali (PC1 e PC2) che cercano di catturare il maggior numero di informazioni dai dati originali.

Ecco cosa si può notare:

- **Suddivisione in gruppi (Cluster):** Le province sono state raggruppate in tre

categorie principali, distinguibili dai colori:

- **Cluster 1 (Blu):** La maggior parte delle province rientra in questo gruppo. Indica un insieme di province che condividono un profilo socio-economico simile, rappresentando la norma o la maggioranza.
- **Cluster 2 (Arancione):** Solo poche province (Napoli, Milano, Prato) appartengono a questo gruppo. Suggestisce un profilo socio-economico distinto da quello del gruppo blu, forse

orientato verso caratteristiche tipiche delle grandi aree urbane o metropolitane.

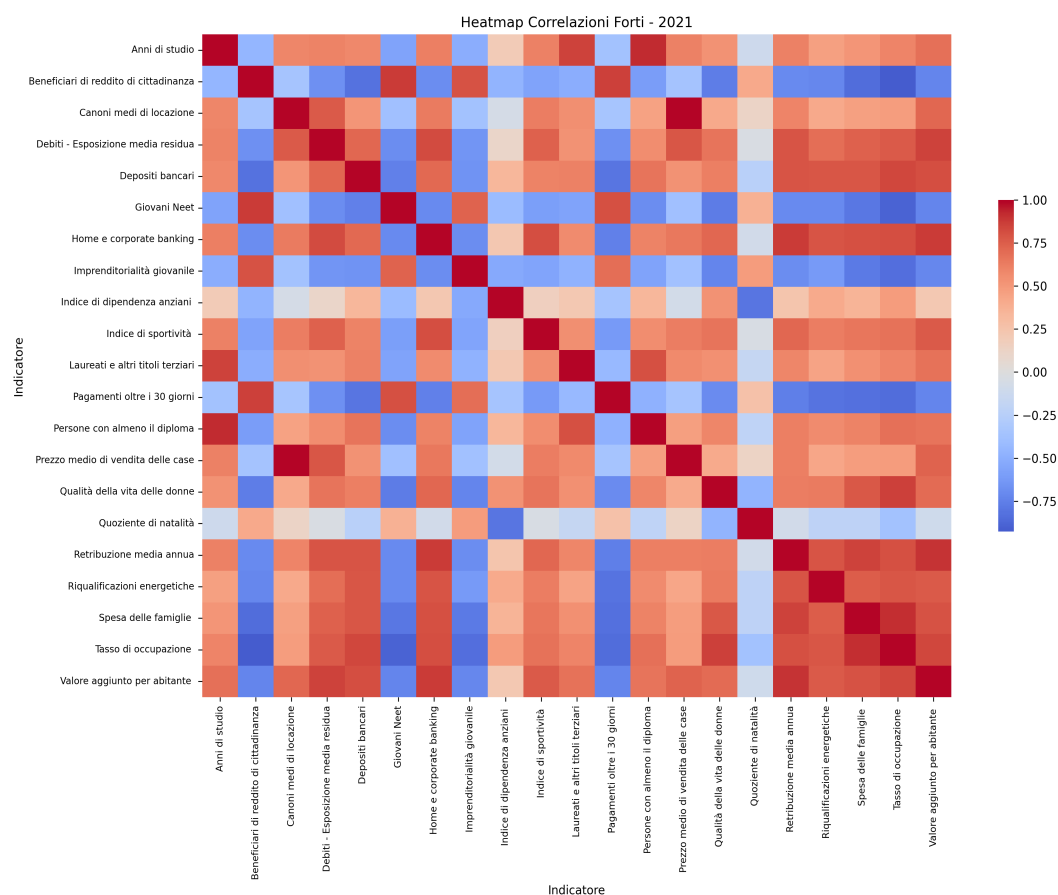
- **Outlier (Rosso):** Alcune province (Rimini, Imperia, Savona, Grosseto, Nuoro) sono state identificate come "outlier", cioè significativamente diverse dalle altre. Queste province presentano caratteristiche socio-economiche che le distinguono notevolmente dalla massa e potrebbero meritare un'analisi più approfondita per capirne le peculiarità.
- **Distribuzione geografica (implicita):** Sebbene il grafico non rappresenti le province sulla mappa d'Italia, la loro posizione sui due assi (PC1 e PC2) può suggerire una certa tendenza legata a fattori territoriali. Ad esempio, potremmo ipotizzare che una delle dimensioni possa catturare aspetti come il livello di industrializzazione, la densità abitativa, il reddito medio, o altri indicatori che variano a livello nazionale.
- **Province chiave:** Alcune province sono posizionate in modo particolare:
- **Milano e Napoli** sono nel cluster arancione, suggerendo che entrambe,

pur con le loro differenze, si discostano dal profilo medio delle province. Milano si trova in alto a destra, mentre Napoli è più a sinistra e in basso, indicando differenze interne al gruppo arancione.

- Province come **Roma, Firenze e Torino** sono nel gruppo blu principale, il che significa che il loro profilo socio-economico è considerato piuttosto "medio" rispetto all'insieme delle province analizzate, secondo le dimensioni considerate.
- Le province rosse (outlier) sono sparse e isolate, confermando la loro unicità. **Nuoro** è un outlier in alto a destra, mentre **Rimini** è in alto, **Imperia** e **Savona** più a sinistra, e **Grosseto** in basso.

In sintesi, questo grafico ci aiuta a capire quali province italiane hanno profili socio-economici simili e quali invece si discostano in modo significativo, identificando sia gruppi di province con caratteristiche condivise sia quelle che potrebbero essere considerate eccezioni nel panorama nazionale. Questo è utile per capire le diversità territoriali italiane non solo in termini geografici, ma anche in termini di caratteristiche socio-economiche.

heatmap_correlazioni_2021



Questo grafico, chiamato "Heatmap Correlazioni Forti", ci mostra quanto sono collegati tra loro diversi indicatori socio-economici nelle province italiane nel 2021. Immagina ogni quadratino come un punto di incontro tra due indicatori: il colore e l'intensità del colore ci dicono se questi due indicatori tendono a variare insieme e in che misura.

In pratica:

- **Rosso intenso:** Indica una forte correlazione positiva. Se un indicatore aumenta, anche l'altro tende ad aumentare in modo significativo.
- **Blu intenso:** Indica una forte correlazione negativa. Se un indicatore

aumenta, l'altro tende a diminuire in modo significativo.

- **Colori chiari (bianco/grigio chiaro):** Indicano una correlazione debole o quasi nulla. I due indicatori non sembrano influenzarsi molto a vicenda.
- **Colori intermedi (arancione, azzurro):** Indicano correlazioni più moderate, positive o negative a seconda della tonalità.

Cosa notiamo da questa mappa?

Osservando le aree con colori più caldi (rossi) e più freddi (blu), possiamo identificare alcuni pattern interessanti:

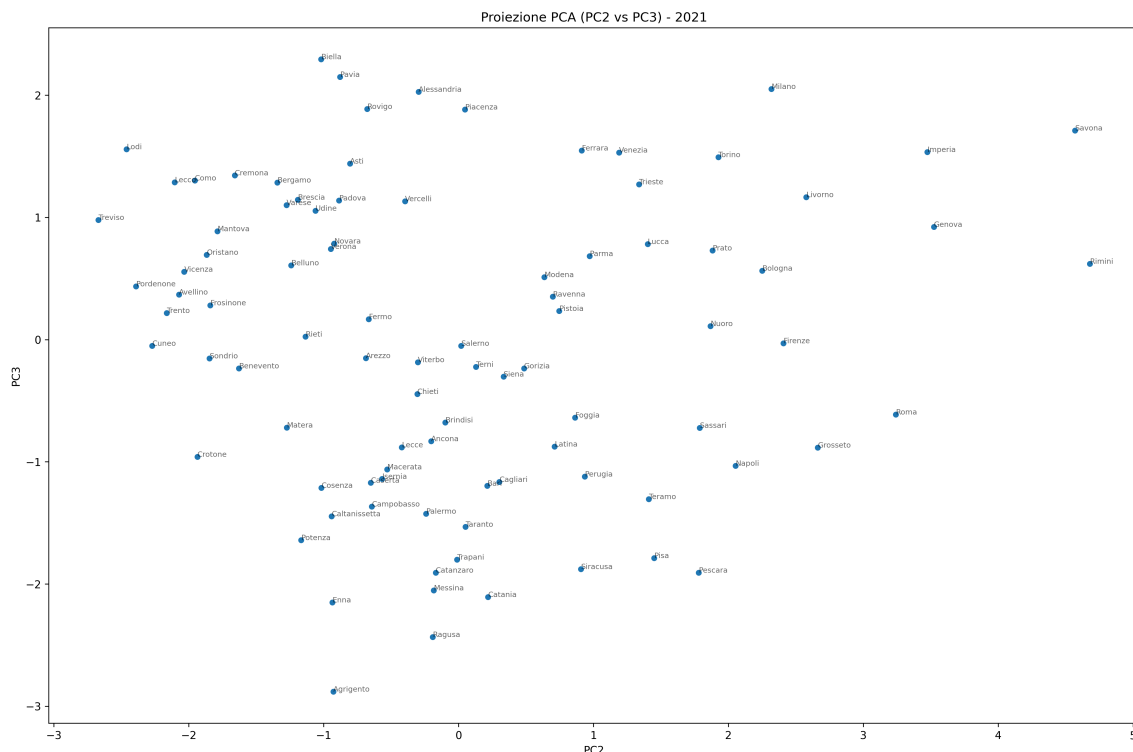
- **Istruzione e Finanza:** "Anni di studio" e "Laureati e altri titoli terziari" sembrano essere collegati positivamente con indicatori come "Deposti bancari" e "Home e corporate banking". Questo suggerisce che nelle province dove la popolazione è più istruita, ci sono spesso anche più risparmi e un maggiore uso di servizi bancari. C'è anche una correlazione positiva con il "Prezzo medio di vendita delle case", il che potrebbe indicare che aree più prospere con maggiore istruzione tendono ad avere anche un mercato immobiliare più costoso.
- **Disoccupazione e Indicatori Negativi:** "Giovani Neet" (i giovani che non studiano, non lavorano e non sono in formazione) ha delle correlazioni negative con indicatori come "Tasso di occupazione" (ovviamente) e "Valore aggiunto per abitante". Dove ci sono più giovani NEET, tendenzialmente c'è meno occupazione e un valore economico per persona inferiore.
- **Reddito e Consumi:** "Retribuzione media annua" mostra correlazioni positive con "Spesa delle famiglie" e "Valore aggiunto per abitante". Non sorprende, se le persone guadagnano di più, tendono a spendere di più e il valore economico della provincia aumenta.
- **Demografia e Servizi:** "Indice di dipendenza anziani" ha correlazioni

interessanti. Potrebbe essere collegato in modo inverso con indicatori di natalità e salute, anche se da questa heatmap non emergono collegamenti fortissimi e chiari in tal senso.

- **Un caso particolare:** Notiamo una correlazione molto forte e positiva tra "Retribuzione media annua" e "Prezzo medio di vendita delle case". Questo è un dato che salta all'occhio: dove gli stipendi sono più alti, i prezzi delle case sono significativamente più alti.
- **Indicatori poco correlati:** Alcuni indicatori, come "Beneficiari di reddito di cittadinanza" o "Riqualificazioni energetiche", sembrano avere correlazioni più deboli o miste con molti altri indicatori, suggerendo che la loro dinamica potrebbe essere influenzata da un insieme più complesso di fattori o che sono meno direttamente collegati a quelli rappresentati qui.

In sintesi, questa mappa ci aiuta a capire quali aspetti della vita socio-economica delle province italiane tendono ad andare di pari passo, fornendoci spunti per eventuali analisi territoriali più approfondite o per la formulazione di politiche mirate. Ad esempio, potremmo notare che intervenire sul livello di istruzione potrebbe avere effetti a cascata positivi sulla situazione finanziaria e immobiliare di alcune aree.

scatter_pca_2021



Questa immagine è una mappa che ci mostra come si posizionano le province italiane in base a due aspetti socio-economici riassuntivi, chiamati "PC2" e "PC3". Immagina che ogni provincia sia un punto su questa mappa, e la sua posizione ci dice quanto è simile o diversa dalle altre province per questi due aspetti.

Cosa rappresenta:

- **Le province:** Ogni pallino rappresenta una provincia italiana, ed è etichettato con il nome della provincia stessa.
- **Gli assi (PC2 e PC3):** Sono come due dimensioni che riassumono tanti dati socio-economici (come reddito, istruzione, occupazione, servizi, ecc.). Senza entrare nei dettagli tecnici di come questi "PC" vengono calcolati, possiamo interpretarli così:

- **Asse orizzontale (PC2):** Rappresenta un primo grande fattore di differenziazione tra le province.
- **Asse verticale (PC3):** Rappresenta un secondo grande fattore, diverso dal primo, che distingue ulteriormente le province.

Evidenze e pattern che si notano:

1. Concentrazioni geografiche, ma non solo:

- Si nota una **clusterizzazione delle province del Nord** in alto a sinistra. Province come Milano, Torino, Venezia, Ferrara, Bologna, Firenze, e quelle lombarde e venete sono raggruppate in un'area che possiamo definire di "valori alti" per entrambi i fattori PC2 e PC3. Questo suggerisce che queste province condividono

caratteristiche socio-economiche simili, posizionandosi in modo omogeneo.

- Al contrario, le province del **Sud Italia**, in particolare Sicilia, Calabria e Puglia, tendono a trovarsi nella parte bassa della mappa, con valori PC2 e PC3 generalmente più bassi. Questo indica che, nel complesso, presentano un profilo socio-economico differente rispetto alle province settentrionali.
- Ci sono poi aree di **dispersione** e province che si posizionano in modo intermedio. Ad esempio, le province del Centro Italia sembrano più sparse.

2. Province "eccellenti" o "particolari":

- **Milano** si trova nella posizione più "alta" e verso destra, suggerendo che eccelle in entrambi i fattori riassuntivi (PC2 e PC3). Potrebbe rappresentare un modello di sviluppo o benessere particolarmente elevato.
- Altre province come **Torino, Bologna, Firenze**, pur non raggiungendo le posizioni estreme di Milano, si posizionano comunque nella parte alta destra, indicando profili socio-economici robusti.
- Province come **Roma** si trovano in una posizione più isolata e verso sinistra, suggerendo un profilo che è distintivo rispetto ai cluster settentrionali e meridionali, ma che potrebbe avere caratteristiche uniche nel suo insieme.

3. Province "a rischio" o con particolari fragilità:

- Le province più in basso e a sinistra potrebbero rappresentare situazioni dove i fattori socio-economici riassunti da PC2 e PC3 sono meno

favorevoli. Le province del profondo Sud sembrano concentrarsi in quest'area.

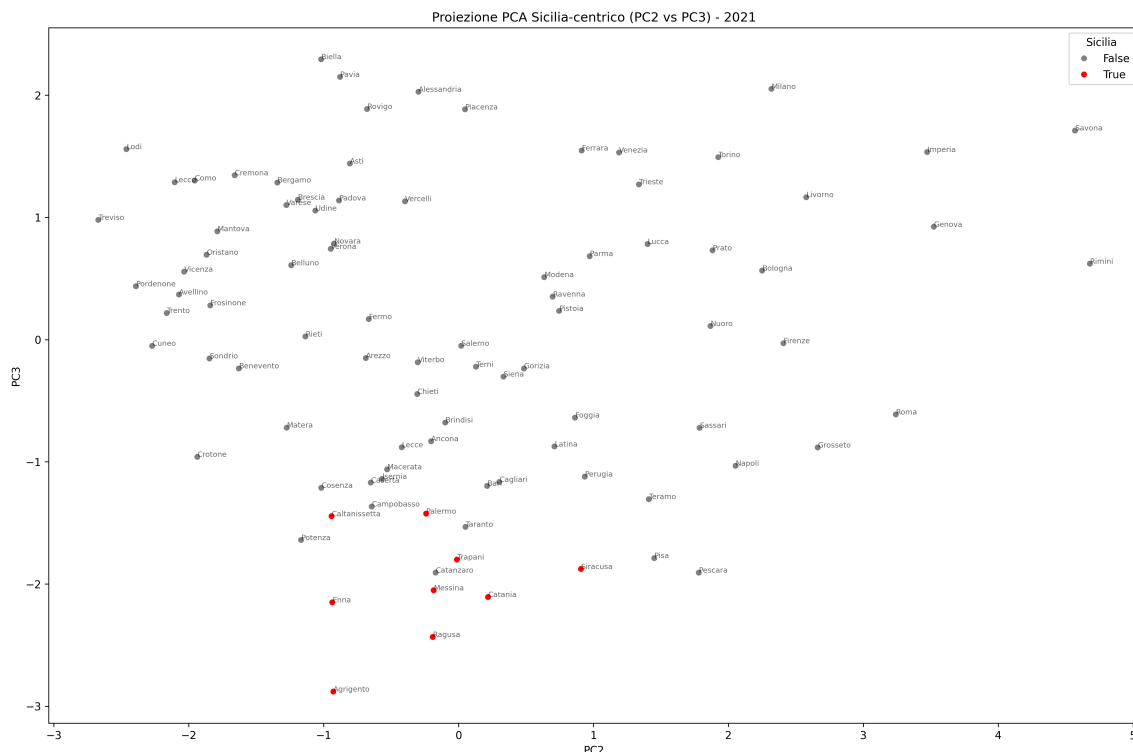
- La presenza di province isolate, come quelle più a sinistra della mappa, potrebbe indicare pattern di sviluppo specifici o problematiche che le rendono uniche rispetto alle altre.

4. La dimensione "est-ovest" (indicata da PC2): Osservando la distribuzione lungo l'asse orizzontale (PC2), sembra esserci una tendenza che va da destra verso sinistra. Le province più "a destra" (come Milano, Bologna, Firenze) potrebbero rappresentare un certo tipo di sviluppo, mentre quelle più "a sinistra" (come alcune province del Nord-Ovest o del Sud) potrebbero rappresentare profili differenti.

5. La dimensione "nord-sud" o "sviluppo" (indicata da PC3): L'asse verticale (PC3) sembra distinguere le province in base a un fattore che generalmente posiziona più in alto le province settentrionali e più in basso quelle meridionali, pur con molte eccezioni. Questo PC potrebbe quindi catturare una parte significativa delle disparità territoriali tradizionalmente associate alla divisione Nord-Sud.

In sintesi, questo grafico ci permette di visualizzare la "mappa socio-economica" delle province italiane in modo sintetico. Ci aiuta a identificare gruppi di province con caratteristiche simili, a mettere in evidenza le aree più virtuose e quelle potenzialmente più fragili, e a comprendere le principali dimensioni di variabilità socio-economica che distinguono il territorio italiano.

scatter_pca_2021_sicilia



Questo grafico rappresenta la posizione di diverse province italiane in base a due dimensioni principali (PC2 e PC3) derivate da un'analisi di dati socio-economici del 2021. In sostanza, ci aiuta a visualizzare come le province si differenziano o si assomigliano su alcuni aspetti economici e sociali.

Ogni punto sul grafico corrisponde a una provincia italiana, identificata dal suo nome. Le posizioni orizzontale (PC2) e verticale (PC3) dei punti indicano quanto ogni provincia si collochi su queste due dimensioni. Le province più vicine tra loro tendono ad avere caratteristiche socio-economiche più simili, mentre quelle più lontane sono più distanti.

Evidenze e Pattern:

- Distinzione tra Province Siciliane e Altre:** La colorazione rossa delle province siciliane (Agrigento, Caltanissetta, Catania, Enna, Messina, Palermo, Ragusa, Siracusa) è la prima cosa che salta all'occhio. Queste province sono raggruppate in un'area specifica del grafico, tendenzialmente nella parte inferiore sinistra e centrale, con valori negativi sia su PC2 che su PC3. Questo suggerisce che, secondo i dati analizzati, le province siciliane nel loro complesso presentano caratteristiche socio-economiche che le distinguono dalla maggior parte delle altre province italiane, che sono rappresentate in grigio.
- Le Province del Nord in Alto a Sinistra:** Osservando le province grigie (non siciliane), notiamo un gruppo di province del Nord Italia

(come Milano, Torino, Bologna, Firenze, Venezia, Ferrara) posizionate nella parte alta e destra del grafico, con valori positivi su PC2 e PC3. Questo indica che queste province sono caratterizzate da un insieme di fattori socio-economici che le differenziano da quelle siciliane. Ad esempio, potrebbero rappresentare indicatori di maggiore sviluppo economico, reddito più elevato o infrastrutture più avanzate, se questi sono stati gli elementi considerati nell'analisi.

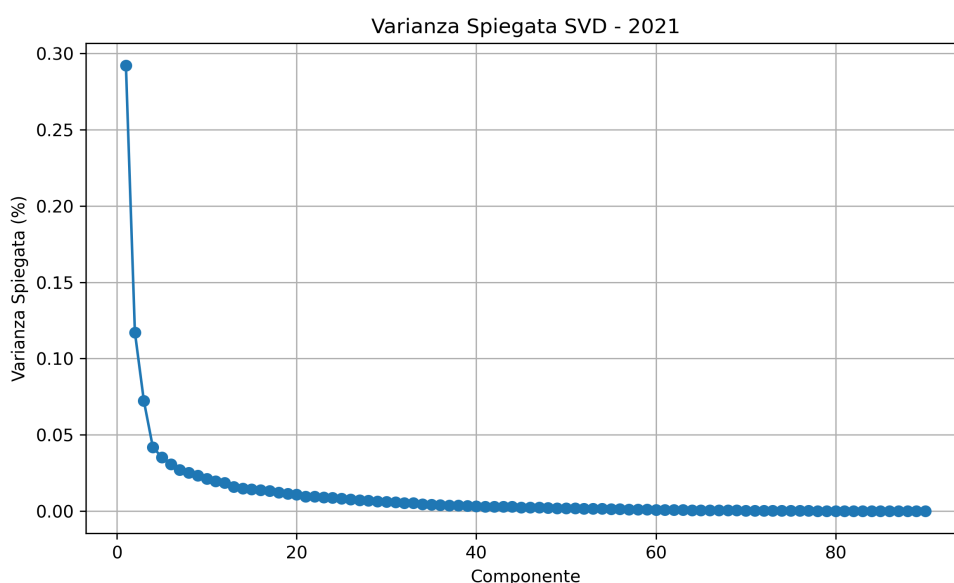
- **Province Intermedie o con Caratteristiche Miste:** Ci sono altre province grigie sparse nel grafico. Alcune si trovano in posizioni intermedie, suggerendo un mix di caratteristiche che non le posizionano né con il gruppo delle province settentrionali più sviluppate, né con quello delle province siciliane. Altre ancora, come Potenza (rappresentata in grigio ma vicina alle siciliane in

termini di posizionamento su PC2), mostrano una certa vicinanza a questo blocco, pur non facendone parte.

- **Polarità Generale:** Nel complesso, il grafico sembra evidenziare una certa polarità tra il blocco delle province siciliane (rosse) e un gruppo di province più settentrionali (grigie), suggerendo una divisione geografica e socio-economica significativa. Le altre province grigie si distribuiscono in uno spazio intermedio, delineando una sorta di "continuum" tra queste due estremità.

In sintesi, il grafico è uno strumento utile per identificare gruppi di province con profili socio-economici simili e per osservare le differenze regionali. Le province siciliane emergono come un gruppo con caratteristiche distinte rispetto a molte altre aree del paese, in particolare rispetto a quelle situate più a nord.

svd_varianza_spiegata_2021



Questo grafico, intitolato "Varianza Spiegata SVD - 2021", ci mostra quanto di questa complessità socio-economica delle province italiane viene "catturata" da una serie di dimensioni o fattori, ordinati per importanza. In termini pratici, è come se stessi cercando di descrivere le province italiane usando tanti indicatori diversi (reddito, occupazione, istruzione, ecc.). L'analisi SVD cerca di creare dei "riassunti" di questi indicatori, che chiamiamo "componenti", cercando di farli rappresentare il più possibile la varietà di situazioni che vediamo tra le province.

Cosa rappresenta il grafico:

- Sull'asse orizzontale ("Componente"), troviamo questi "riassunti" o fattori, che sono ordinati dal più importante (il primo, a sinistra) al meno importante.
- Sull'asse verticale ("Varianza Spiegata (%)"), vediamo quanta della "varietà totale" nei dati socio-economici viene spiegata da ciascuno di questi componenti.

Evidenze e pattern che si notano:

1. Forte concentrazione nelle prime componenti: La cosa più evidente è che le prime poche componenti spiegano una porzione molto grande della variabilità complessiva. Il primo componente da solo spiega circa il 30% della varianza, il secondo circa il 12%, e il terzo circa il 7%. Questo significa che le prime 2-3 dimensioni sono in grado di cogliere la maggior parte delle differenze socio-economiche tra le province italiane.

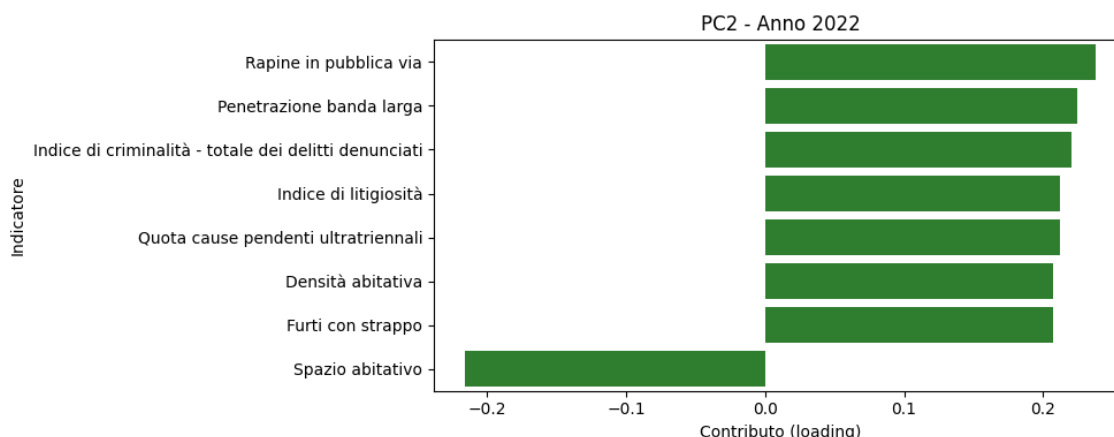
2. Diminuzione rapida: Dopo le prime componenti, la quantità di varianza spiegata da ciascuna componente successiva diminuisce molto rapidamente. Questo comportamento a "gomito" (o "scree plot" in termini più tecnici) è molto comune in analisi di questo tipo.

3. Informazione distribuita su molte componenti residue: Anche se le prime componenti sono le più importanti, notiamo che la varianza non si esaurisce immediatamente. Ci sono molte componenti successive che spiegano una piccola quantità di varianza, ma nel complesso possono comunque contribuire a una comprensione più fine delle differenze. Tuttavia, l'impatto di ciascuna di queste componenti è marginale.

In sintesi per un analista territoriale:

Possiamo dire che la geografia socio-economica italiana, per il 2021, può essere ragionevolmente descritta utilizzando un numero limitato di dimensioni chiave. Le prime 2-3 componenti sono quelle che ci danno la visione più chiara e sintetica delle principali disparità e omogeneità territoriali. Le componenti successive, pur presenti, aggiungono dettagli sempre più specifici e meno impattanti sul quadro generale, e probabilmente si riferiscono a sfumature o caratteristiche socio-economiche più particolari che non sono predominanti. Questo ci suggerisce che un'analisi territoriale che si concentra su pochi fattori principali, derivati da queste prime componenti, potrebbe essere molto efficace nel distinguere le province italiane dal punto di vista socio-economico.

barplot_PC2



Questa immagine ci mostra come alcuni indicatori socio-economici contribuiscono a definire una specifica dimensione o "fattore" (indicato come "PC2") per le province italiane nell'anno 2022. In altre parole, stiamo cercando di capire quali aspetti della realtà provinciale tendono ad andare insieme.

Ecco cosa notiamo:

- **Indicatori che "tirano nella stessa direzione":** La maggior parte delle barre sono verdi e si estendono verso destra, indicando un contributo positivo. Questo significa che se una provincia ha valori elevati in questi indicatori, tende a posizionarsi in un certo modo rispetto a questo "PC2". Parliamo di:
 - **Rapine in pubblica via:** Aree con più rapine tendono ad avere punteggi alti.
 - **Penetrazione banda larga:** Le province con una maggiore diffusione della banda larga sembrano contribuire positivamente.
 - **Indice di criminalità - totale dei delitti denunciati:** Anche un alto numero di denunce per reati contribuisce positivamente.

- **Indice di litigiosità:** Similmente, un indice di litigiosità elevato è associato a punteggi alti.
- **Quota cause pendenti ultratriennali:** Se ci sono molte cause legali che durano da più di tre anni, questo contribuisce.
- **Densità abitativa:** Le aree più densamente popolate hanno un contributo positivo.
- **Furti con strappo:** Analogamente alle rapine, i furti con strappo si associano a punteggi alti.
- **Indicatori che "tirano nella direzione opposta":** L'unica barra che si estende a sinistra, indicando un contributo negativo, è quella relativa allo **"Spazio abitativo"**. Questo significa che le province dove lo spazio abitativo è minore (quindi, presumibilmente, più affollate o con case più piccole) tendono ad avere punteggi bassi per questo "PC2".

In sintesi, questo "PC2" sembra catturare una dimensione che potrebbe essere interpretata come "Complessità urbana e criticità sociali". Le province che hanno punteggi alti in questo fattore sono probabilmente aree con alta densità

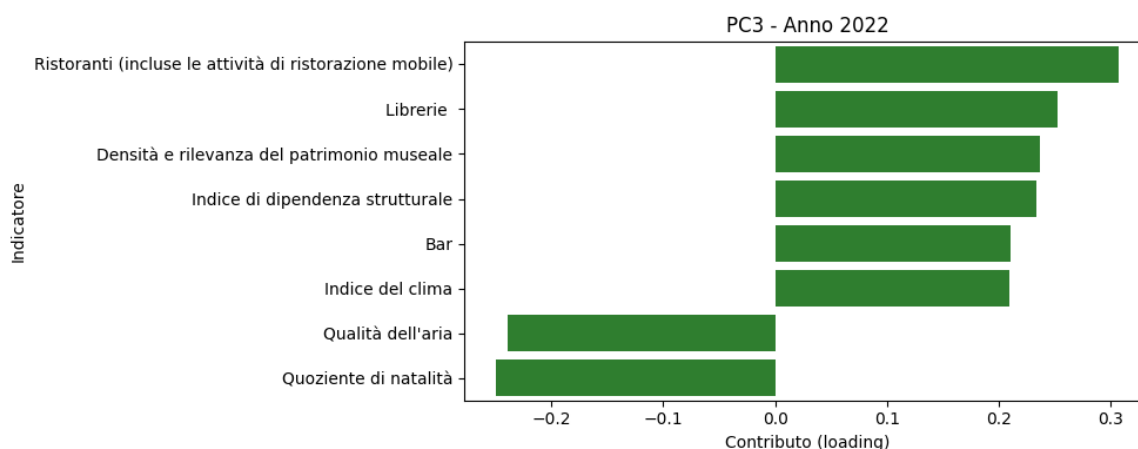
abitativa, dove si verificano più problemi legati alla criminalità, alla lunghezza dei processi e una maggiore diffusione della tecnologia (banda larga). Al contrario, province con punteggi bassi sono quelle dove c'è più spazio abitativo e, con ogni probabilità, minore pressione urbana e minore incidenza di questi specifici problemi.

È interessante notare come la penetrazione della banda larga sia inclusa

in questo quadro. Potrebbe suggerire che in alcune aree l'urbanizzazione e la tecnologia avanzata vanno di pari passo con un aumento delle criticità sociali, o che la tecnologia è particolarmente diffusa nelle aree urbane più "complesse".

Questi risultati ci aiutano a raggruppare le province in base a profili socio-economici simili, permettendoci di individuare aree con caratteristiche comuni per eventuali interventi o analisi più mirate.

Barplot_PC3



Questa immagine mostra l'analisi del terzo fattore ("PC3") per l'anno 2022, basata su vari indicatori socio-economici a livello di provincia italiana. Il grafico a barre orizzontali indica quanto ciascun indicatore contribuisce a definire questo fattore e in quale direzione.

Cosa rappresenta:

Immagina di voler creare una sintesi che descriva alcune delle differenze tra le province. Questo grafico ci mostra che il "PC3" cattura una dimensione specifica di queste differenze. Vediamo quali elementi sono fortemente associati a questo fattore

e se questa associazione è positiva (cioè, dove questi elementi sono presenti in abbondanza, il PC3 è alto) o negativa (cioè, dove questi elementi sono presenti in abbondanza, il PC3 è basso).

Evidenze e pattern notati:

- **Attrattività e offerta culturale/ricreativa:** Gli indicatori con il contributo più forte e positivo al PC3 sono "Ristoranti (incluse le attività di ristorazione mobile)", "Librerie" e "Densità e rilevanza del patrimonio museale". Questo suggerisce che il PC3 rappresenta in

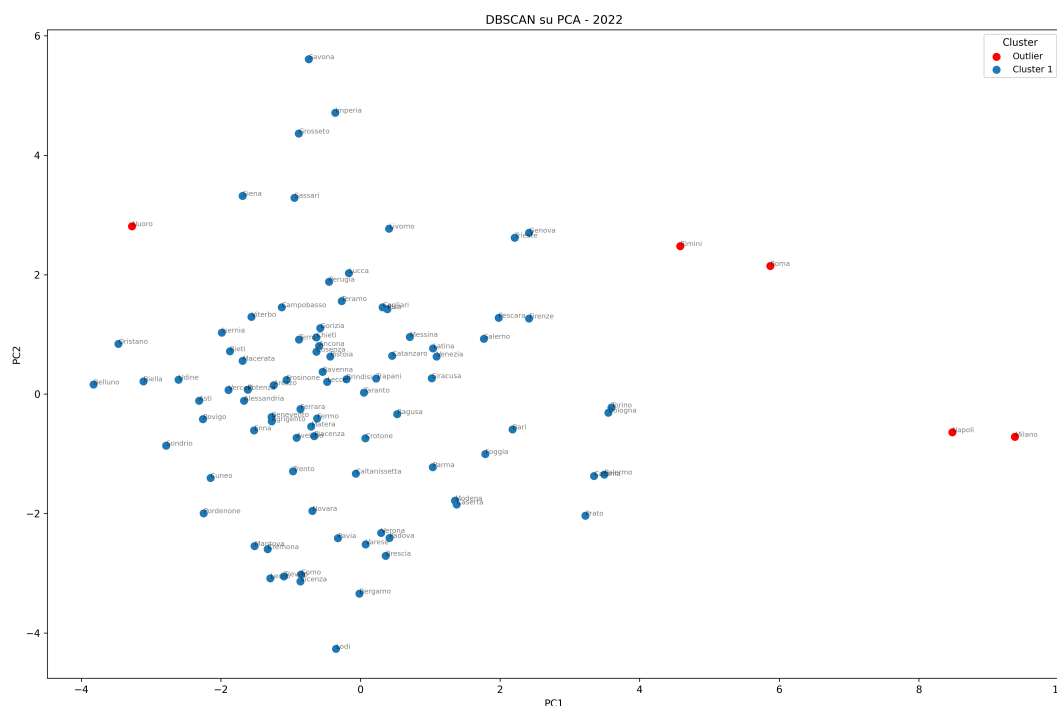
gran parte una dimensione legata all'offerta di servizi legati alla ristorazione, alla cultura e al turismo/ricreazione. Le province che eccellono in questi settori avranno un valore alto per questo fattore.

- **Spesa e qualità della vita "moderna":** Anche "Indice di dipendenza strutturale" e "Bar" hanno un contributo positivo, sebbene leggermente inferiore. L'"Indice di dipendenza strutturale" da solo potrebbe sembrare controintuitivo, ma in questo contesto potrebbe riflettere una popolazione con maggiore capacità di spesa per servizi non essenziali, o un'economia più orientata ai servizi che attrae determinate attività. La presenza di bar, come i ristoranti e le librerie, è un indicatore di vivacità sociale e offerta per il tempo libero.
- **Fattori ambientali e demografici meno rilevanti per questo specifico fattore:** D'altra parte, "Indice del clima", "Qualità dell'aria" e "Quoziente

di natalità" hanno contributi negativi o vicini allo zero. Questo significa che, per definire questo specifico fattore (PC3), questi indicatori hanno un peso minore. Anzi, dove la qualità dell'aria è migliore o il tasso di natalità è più alto, il PC3 tende ad essere più basso. Questo non significa che questi indicatori non siano importanti in generale, ma che non sono le caratteristiche distintive di questo particolare "pacchetto" di variabili socio-economiche che stiamo analizzando.

In sintesi, il PC3 del 2022 sembra definire una sorta di "attrattività territoriale" legata principalmente a servizi di ristorazione, cultura, tempo libero e un'economia orientata ai consumi, piuttosto che a fattori ambientali o demografici di base. Le province con un alto valore del PC3 sono probabilmente quelle più dinamiche sul fronte dei servizi e dell'offerta ricreativa.

dbscan_pca_2022



Questa immagine mostra il risultato di un'analisi applicata ai dati socio-economici delle province italiane nel 2022. In sostanza, ogni punto rappresenta una provincia italiana, e la sua posizione sull'asse orizzontale (PC1) e verticale (PC2) è determinata da un riassunto delle sue caratteristiche socio-economiche. Il colore dei punti ci dice se la provincia è stata considerata "normale" (blu) o "anomala" (rossa) secondo un certo algoritmo (chiamato DBSCAN).

Cosa vediamo:

- **La maggior parte delle province si raggruppa:** La maggior parte dei punti blu si trova concentrata in diverse aree del grafico. Questo significa che molte province condividono caratteristiche socio-economiche simili. Possiamo vederli

raggruppati più o meno al centro e verso sinistra.

- **Alcune province sono considerate "speciali" (Outlier):** Ci sono alcuni punti rossi che si trovano molto distanti dagli altri. Questi sono le province identificate come "outlier", ovvero che hanno un profilo socio-economico significativamente diverso dalla maggior parte delle altre province. In questo caso, vediamo le province di **Milano, Napoli, Roma e Nuoro** come outlier.
- **PC1 e PC2 come riassunti di caratteristiche:** Gli assi PC1 e PC2 non rappresentano direttamente una singola variabile (come il PIL o la disoccupazione), ma piuttosto una combinazione di molte di esse. Immaginateli come due dimensioni sintetiche che ci aiutano a "vedere" le somiglianze e le differenze tra le

province basandoci su tutti i dati analizzati.

Evidenze e pattern che si notano:

1. Concentrazione geografica e di caratteristiche: Notiamo una forte densità di province blu nella zona centrale sinistra del grafico. Questo suggerisce che ci sono ampi gruppi di province che, da un punto di vista socio-economico aggregato, si assomigliano molto. Senza conoscere esattamente cosa rappresentino PC1 e PC2, possiamo ipotizzare che questo gruppo raccolga province con un certo "profilo standard".

2. Isolamento degli Outlier: Le province rosse (Milano, Napoli, Roma, Nuoro) sono chiaramente separate dal gruppo principale. Questo indica che queste province hanno dei tratti socio-economici che le distinguono in modo marcato.

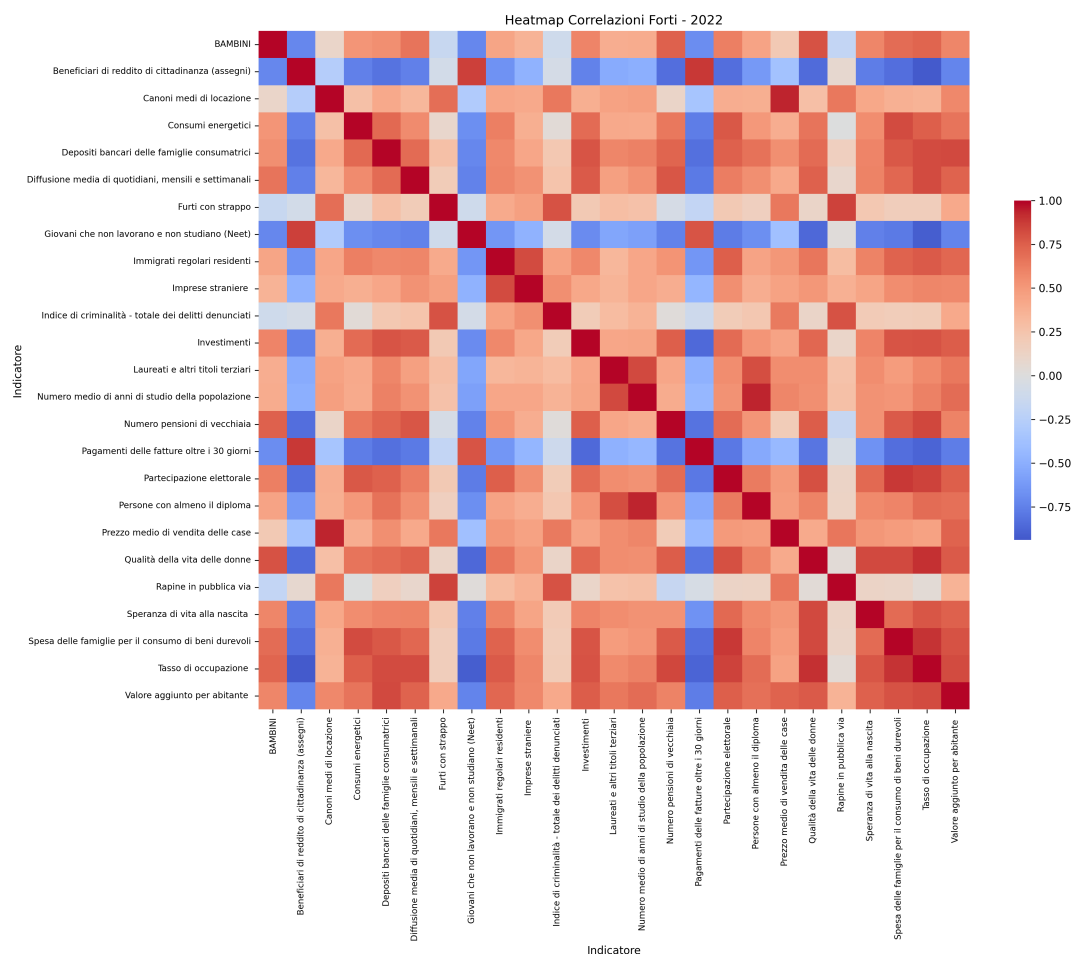
- **Milano e Napoli** si trovano nella parte destra, con Milano più a destra e Napoli più in alto. Potrebbero rappresentare estremi su una delle dimensioni sintetiche (PC1 o PC2).
- **Roma** si trova anch'essa nella parte destra, ma leggermente diversa da Milano e Napoli.

- **Nuoro**, invece, si trova in alto a sinistra, molto isolata dal gruppo principale e anche dai altri outlier. Questo suggerisce che il suo profilo socio-economico è unico e diverso sia dalla maggioranza delle province che dalle altre anomalie individuate.

3. Possibili cluster all'interno del gruppo principale: Pur essendoci una concentrazione generale, si possono intravedere alcune aree dove i punti blu sembrano aggregarsi un po' di più, suggerendo la possibile esistenza di ulteriori sotto-gruppi (cluster) all'interno del gruppo più ampio, ma il rosso indica che l'algoritmo DBSCAN ha deciso di marcarli come esterni a gruppi più definiti.

In sintesi, questo grafico ci offre una mappa visuale delle province italiane basata sui loro profili socio-economici, evidenziando quelle che si discostano maggiormente dalla media e permettendoci di cogliere similitudini tra le altre. Le province rosse richiedono un'analisi più approfondita per capire quali specifici fattori socio-economici le rendano così particolari.

heatmap_correlazioni_2022



Questa immagine è una heatmap che visualizza la **forza delle correlazioni tra diversi indicatori socio-economici delle province italiane nel 2022**. In parole semplici, ci mostra quanto questi fattori tendono ad andare insieme (o in direzioni opposte) nelle diverse aree del paese.

Ecco come interpretarla:

- **La griglia:** Ogni casella rappresenta la relazione tra due indicatori. L'indicatore sulla sinistra è quello sulla riga, mentre quello in cima è quello sulla colonna.
- **I colori:** La scala di colori è fondamentale:
 - **Rosso e tonalità calde (verso l'alto della scala):** Indicano una forte **correlazione positiva**. Significa che quando un indicatore è alto in una provincia, è probabile che anche l'altro indicatore sia alto in quella stessa provincia, e viceversa. Più è intenso il rosso, più è forte questa relazione.
 - **Blu e tonalità fredde (verso il basso della scala):** Indicano una forte **correlazione negativa**. Quando un indicatore è alto, è probabile che l'altro sia basso, e viceversa. Più è intenso il blu, più è forte questa relazione inversa.
 - **Colori neutri (bianco/grigio chiaro al centro della scala):** Suggestiscono

una **correlazione debole o quasi nulla**. I due indicatori tendono a variare in modo indipendente.

Cosa notiamo (evidenze e pattern):

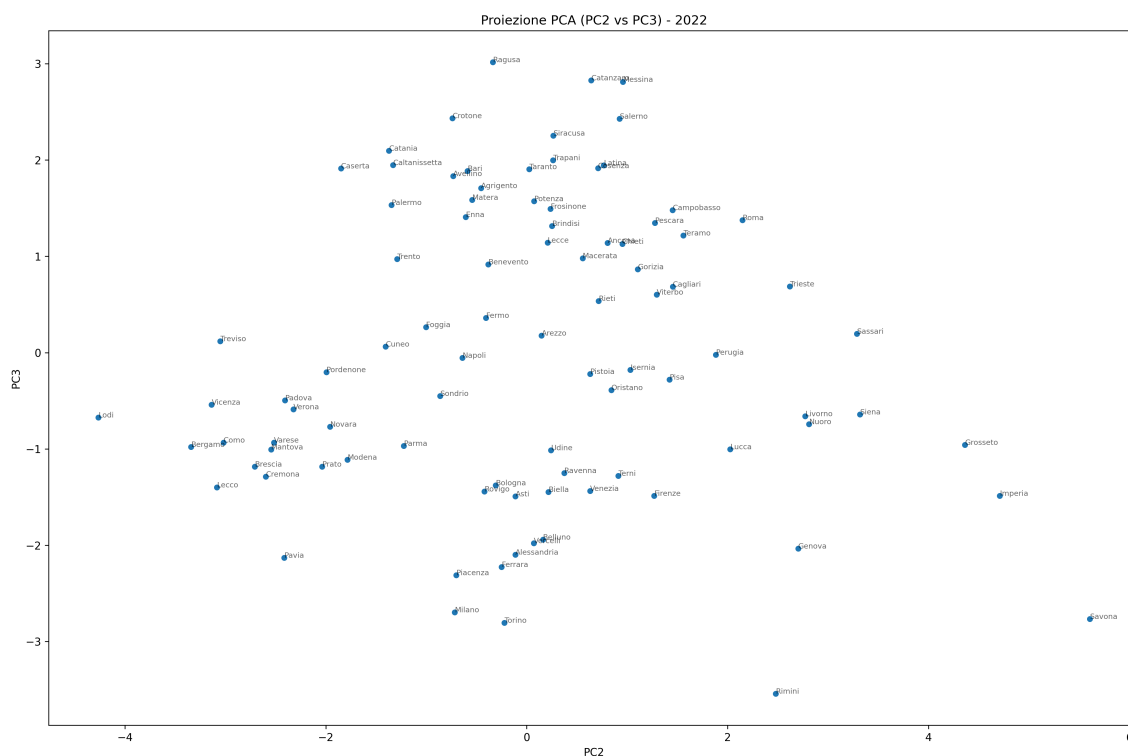
Guardando le aree più rosse e più blu, possiamo individuare delle connessioni interessanti:

- **Fattori economici legati al benessere:** Si osserva una forte correlazione positiva (rosso intenso) tra "Depositi bancari delle famiglie consumatrici", "Consumi energetici" e "Spesa delle famiglie per il consumo di beni durevoli". Questo suggerisce che le province con famiglie che accumulano più risparmi tendono anche ad avere maggiori consumi energetici e a spendere di più in beni durevoli. Anche il "Valore aggiunto per abitante" sembra correlato positivamente con questi indicatori, il che è logico.
- **Istruzione e occupazione:** C'è una chiara correlazione positiva tra "Laureati e altri titoli terziari", "Numero medio di anni di studio della popolazione" e "Tasso di occupazione". Le province con una popolazione più istruita tendono ad avere tassi di occupazione più elevati. Analogamente, "Persone con almeno il diploma" mostra una correlazione positiva con l'occupazione.
- **Indicatori demografici:** "Bambini" sembra avere correlazioni moderate, ma notiamo una tendenza a una correlazione positiva con "Speranza di vita alla nascita". Province con più bambini potrebbero anche essere associate a una maggiore aspettativa di vita.

- **Criminalità e fattori sociali:** L'"Indice di criminalità - totale dei delitti denunciati" mostra alcune correlazioni negative interessanti. Ad esempio, appare correlato negativamente (blu) con "Canoni medi di locazione" e "Prezzo medio di vendita delle case". Questo potrebbe suggerire che dove gli affitti e i prezzi delle case sono più alti, la criminalità tende ad essere meno diffusa (o viceversa). C'è anche una correlazione negativa con la "Qualità della vita delle donne", il che è un dato da approfondire.
- **Giovani e NEET:** La categoria "Giovani che non lavorano e non studiano (Neet)" mostra alcune correlazioni negative con indicatori di benessere economico e occupazione, come ci si potrebbe aspettare.
- **Importazione e residenti stranieri:** Le "Imprese straniere" sembrano avere una correlazione positiva con "Immigrati regolari residenti", il che è un dato atteso.

In sintesi, questa mappa ci offre una visione d'insieme delle interconnessioni tra vari aspetti della vita nelle province italiane. Ci aiuta a capire quali fenomeni tendono ad accadere insieme, fornendo spunti per analisi territoriali più mirate e per la formulazione di politiche basate su evidenze concrete. Ad esempio, potremmo notare che intervenire sull'istruzione potrebbe avere ricadute positive sull'occupazione, o che politiche che mirano a ridurre il costo della vita potrebbero avere un impatto indiretto sulla criminalità.

scatter_pca_2022



Questa immagine mostra una rappresentazione bidimensionale delle province italiane basata su un'analisi di dati socio-economici. Le due dimensioni su cui sono state proiettate le province, etichettate come PC2 e PC3, sono il risultato di un processo matematico chiamato Analisi delle Componenti Principali (PCA). In termini semplici, la PCA cerca di riassumere molte informazioni complesse (come indicatori di reddito, occupazione, istruzione, etc.) in un minor numero di nuove "variabili" o assi che catturano la maggior parte della variabilità presente nei dati originali.

In questo grafico, ogni punto rappresenta una provincia italiana, e la sua posizione è determinata dai suoi valori lungo questi due assi sintetici, PC2 e PC3. L'idea è che

province che si trovano vicine tra loro in questo spazio bidimensionale tendono ad avere caratteristiche socio-economiche simili, mentre quelle più distanti sono più dissimili.

Evidenze e Pattern che si Notano:

- **Raggruppamenti Geografici e Socio-Economici:** Si osservano chiari raggruppamenti di province.
- **Nord Italia:** Le province del nord Italia tendono a disporsi nella parte alta e a destra del grafico. In particolare, province come Milano, Torino, Bologna, Verona, Padova, Venezia, Trieste, Udine, e quelle lombarde e venete in generale, appaiono concentrate in quest'area. Questo suggerisce che queste province

condividono un set di caratteristiche socio-economiche che le differenzia dal resto del paese. Potremmo ipotizzare che questi assi (PC2 e PC3) catturino dimensioni legate a un maggiore sviluppo economico, reddito più elevato, migliori tassi di occupazione, o un'economia più diversificata e orientata ai servizi o all'industria avanzata.

- **Centro Italia:** Alcune province del centro Italia, come Roma, Firenze, Perugia, e quelle toscane, si posizionano in una zona intermedia, a volte più vicine al gruppo del nord, a volte più isolate.
- **Sud Italia e Isole:** Le province del Sud Italia e delle isole, come Napoli, Bari, Palermo, Catania, Reggio Calabria, Lecce, Taranto, e quelle siciliane e sarde, tendono a concentrarsi nella parte bassa e sinistra del grafico, e anche più a destra nella parte bassa. Questo indica una tendenza a raggrupparsi e a differenziarsi significativamente dalle province settentrionali. La loro posizione suggerisce che potrebbero condividere caratteristiche socio-economiche diverse, forse legate a un minor reddito pro capite, tassi di disoccupazione più elevati, o una struttura economica differente.
- **Province Specifiche:** Alcune province sembrano essere posizionate in modo più isolato, come ad esempio Lodi, Savona, Grosseto o Imperia, il che potrebbe indicare un profilo socio-economico unico o particolari caratteristiche che le distinguono dalle province vicine.
- **Gradiente Nord-Sud Implicito:** Sebbene non sia un grafico geografico

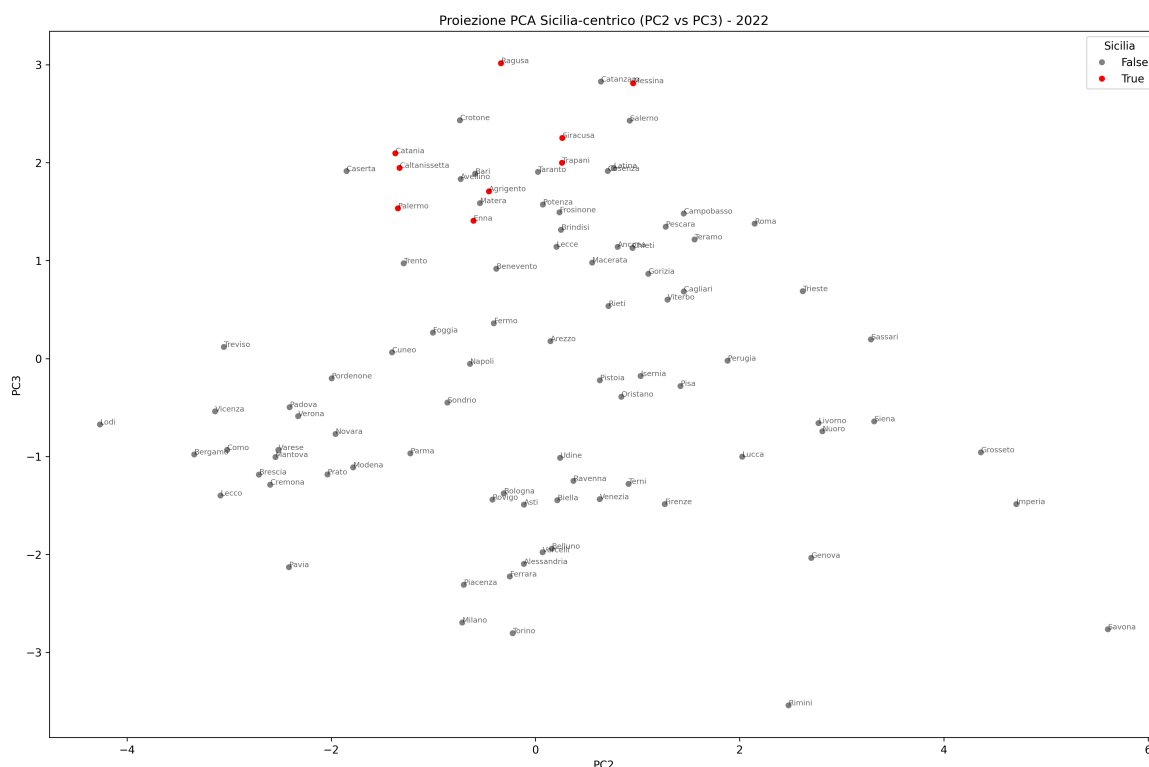
diretto, la disposizione generale delle province suggerisce un certo gradiente socio-economico che corre da nord a sud. Le province più "avanzate" (secondo le dimensioni catturate dai PC) sono prevalentemente al nord, mentre quelle con caratteristiche diverse sono più presenti al sud.

- **Variabilità tra Province:** Il grafico mostra una notevole dispersione dei punti, il che indica che le province italiane, pur avendo alcuni raggruppamenti, presentano anche una notevole eterogeneità nelle loro caratteristiche socio-economiche.
- **Relazione tra PC2 e PC3:** L'orientamento dei punti nel grafico ci dà un'idea di come i due assi (PC2 e PC3) interagiscono nel definire il profilo delle province. Ad esempio, alcune province al nord hanno valori alti sia su PC2 che su PC3, mentre altre province, pur essendo nel nord, potrebbero avere valori più bassi su uno dei due assi, spostandole leggermente all'interno del gruppo settentrionale. Allo stesso modo, per le province del sud, potremmo notare che alcune hanno valori simili su entrambi gli assi, mentre altre si differenziano.

In sintesi, questo grafico offre una "fotografia" delle province italiane attraverso lenti socio-economiche aggregate, evidenziando come gruppi di province abbiano profili simili e come queste somiglianze e differenze si traducano in posizioni distinte in uno spazio ridotto ma informativo. È uno strumento utile per identificare cluster territoriali con caratteristiche omogenee o per analizzare le principali dimensioni di

differenziazione socio-economica nel paese.

scatter_pca_2022_sicilia



L'immagine rappresenta una proiezione di dati socio-economici delle province italiane nel 2022, utilizzando due componenti principali (PC2 e PC3). Ogni punto su questo grafico corrisponde a una provincia, e la sua posizione è determinata dai suoi valori su queste due componenti derivate dall'analisi.

Le province sono colorate in base all'appartenenza alla Sicilia:

- **Grigio:** Province non siciliane.
- **Rosso:** Province siciliane.

Evidenze e pattern notati:

1. Raggruppamento delle province siciliane: Si osserva un chiaro raggruppamento delle province siciliane (quelle rosse) in una zona specifica del grafico, concentrata prevalentemente nella parte superiore sinistra. Questo suggerisce che queste province condividono caratteristiche socio-economiche simili, che le differenziano dalle altre province italiane rappresentate in grigio.

2. Separazione geografica e socio-economica: Il grafico sembra separare le province siciliane da molte altre province italiane. Le province siciliane sono posizionate in un'area relativamente

compatta, mentre le province non siciliane sono più sparse su tutto il grafico. Questo potrebbe indicare differenze strutturali o di performance socio-economica tra la Sicilia e il resto d'Italia, così come definito dalle variabili considerate nell'analisi.

3. Posizionamento delle province siciliane: All'interno del gruppo delle province siciliane, si notano alcune distinzioni. Ad esempio, Ragusa, Siracusa, Trapani e Palermo sono posizionate in aree leggermente diverse, suggerendo che, anche all'interno della regione, ci sono sfumature nelle loro caratteristiche socio-economiche.

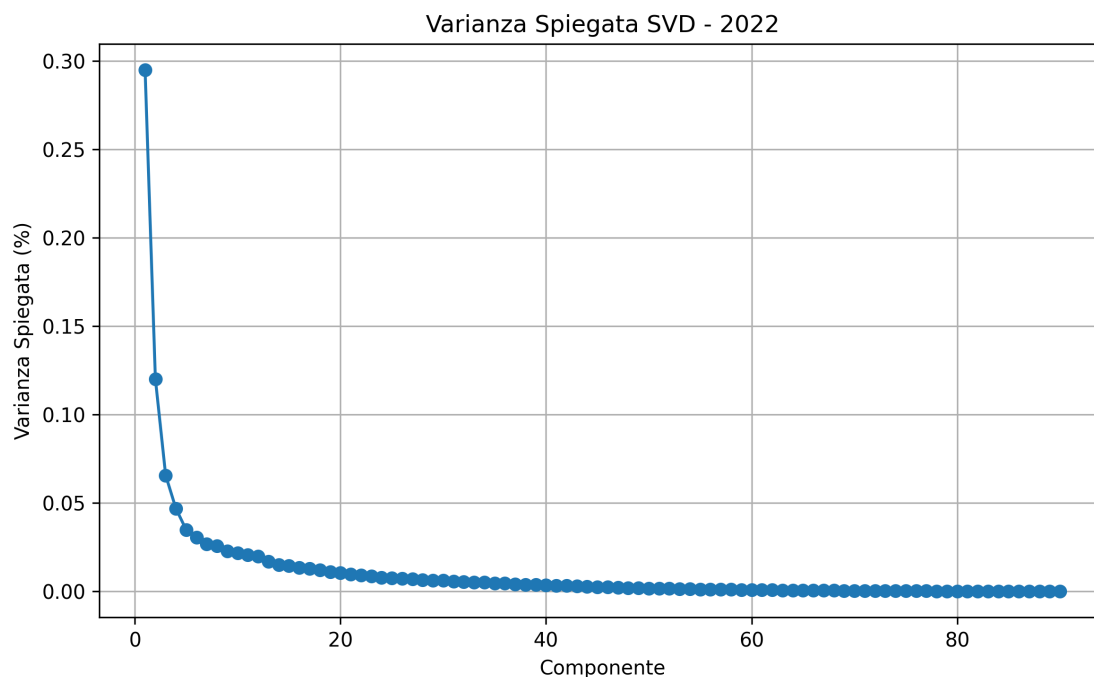
4. Cluster di province non siciliane: Osservando le province in grigio, si possono identificare diversi cluster. Ad esempio, sembra esserci un raggruppamento di province nel nord-ovest (Milano, Torino, Genova, ecc.), un altro nel centro (Roma, Firenze, Bologna) e un altro ancora nel sud e isole non

siciliane (Napoli, Bari, Cagliari). Questo pattern riflette in parte la tradizionale divisione geografica e socio-economica del paese.

5. Posizione delle province "estreme": Alcune province, come quelle in alto a destra (es. Trieste) o in basso a sinistra (es. Lodi), si trovano isolate. Questo indica che le loro caratteristiche socio-economiche, come misurate da queste due componenti, sono piuttosto distanti da quelle della maggior parte delle altre province e in particolare dalla media delle province siciliane.

In sintesi, questo grafico è uno strumento utile per visualizzare come le province italiane si posizionano in base a fattori socio-economici complessi, evidenziando in particolare l'esistenza di un pattern distintivo per le province siciliane e mostrando al contempo la diversità interna ed esterna alla regione.

svd_varianza_spiegata_2022



Questa immagine, intitolata "Varianza Spiegata SVD - 2022", illustra l'efficacia con cui diversi "componenti" (cioè, gruppi di caratteristiche socio-economiche) riescono a descrivere la variabilità presente nei dati delle province italiane. In parole semplici, ci dice quanti fattori principali sono necessari per catturare la maggior parte delle differenze che vediamo tra le province in termini di indicatori socio-economici.

Cosa rappresenta:

- **L'asse orizzontale (asse X)** mostra i "componenti". Questi sono essenzialmente delle dimensioni sintetiche che riassumono diverse variabili socio-economiche. Ad esempio, il primo componente potrebbe rappresentare un mix di fattori legati alla ricchezza e allo sviluppo economico, il secondo un mix

di fattori legati alla demografia e così via.

- **L'asse verticale (asse Y)** indica la "Varianza Spiegata" in percentuale. Questo valore ci dice quanta parte dell'intera diversità dei dati originali viene catturata da ciascun componente. Un componente che spiega un'alta percentuale di varianza è molto importante per comprendere le differenze tra le province.

Evidenze e pattern:

1. **Forte concentrazione di informazione nei primi componenti:** Osserviamo che il primo componente (a sinistra sull'asse X) spiega una porzione molto significativa della varianza totale, circa il 30%. Questo indica che esiste un fattore principale molto potente che differenzia le province italiane. Anche il secondo componente è rilevante, spiegando circa il 12-13% della varianza.

2. Rapido calo dell'importanza dei componenti successivi: Subito dopo i primi due, l'importanza di ogni componente successivo diminuisce drasticamente. Il terzo componente spiega circa il 5%, il quarto circa il 3%, e così via. Dopo i primi 5-10 componenti, l'apporto di ciascun nuovo componente alla spiegazione della varianza diventa molto piccolo, quasi trascurabile.

3. Indizio per la riduzione delle dimensioni: Questo andamento "a gomito" o "a scree plot" è un classico segnale che indica che gran parte dell'informazione essenziale nei dati può essere riassunta utilizzando un numero relativamente ridotto di componenti. Se dovessimo scegliere un numero di componenti per rappresentare i dati in modo efficiente, potremmo considerare di fermarci tra i primi 5 e i primi 15, a seconda di quanto "spazio" di variabilità si vuole conservare.

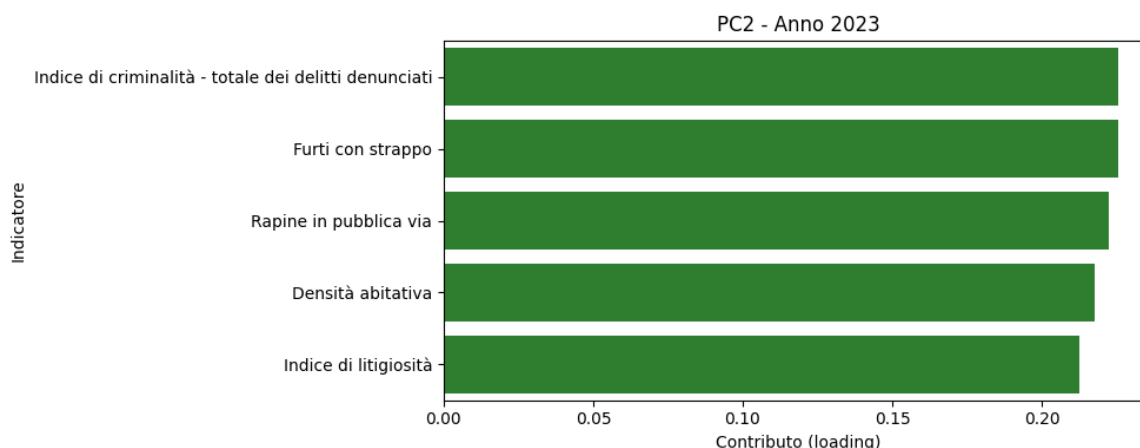
4. Presenza di un rumore di fondo (potenziale): I numerosi componenti alla fine del grafico che spiegano solo una

frazione minuscola di varianza potrebbero rappresentare aspetti specifici e molto localizzati, oppure "rumore" nei dati che non cattura pattern socio-economici ampi e diffusi.

In sintesi per un analista territoriale:

Il grafico ci suggerisce che, nel 2022, le province italiane si differenziano principalmente lungo un paio di assi di variabilità socio-economica. Il primo asse è quello più dominante e spiega quasi un terzo di tutte le differenze che riusciamo a misurare. Ci sono poi altri fattori che contribuiscono, ma con un peso decrescente. Se dovessimo costruire un modello o una sintesi per capire le differenze territoriali, potremmo concentrarci sui primi 5-10 fattori principali, perché sono quelli che catturano la stragrande maggioranza delle informazioni distintive. I fattori successivi sono di minore importanza generale e potrebbero essere considerati solo se stiamo cercando di analizzare dettagli molto specifici o anomalie territoriali.

barplot_PC2



Questo grafico a barre rappresenta il contributo di diversi indicatori socio-economici a un particolare fattore, etichettato come "PC2 - Anno 2023". In termini semplici, ci dice quali di questi indicatori pesano di più nel definire questo fattore (PC2), che a sua volta è stato derivato da un'analisi complessa sui dati delle province italiane.

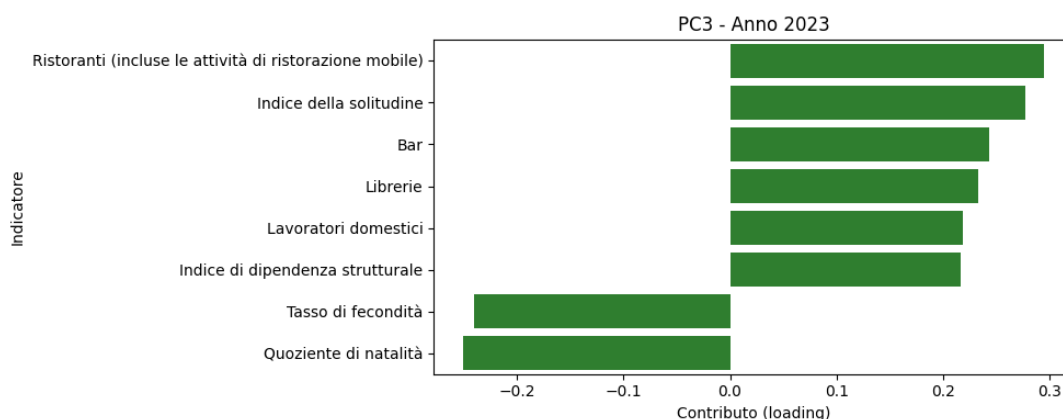
Ecco cosa si nota:

- **Indicatori di criminalità e litigi a parità di peso:** Tutti gli indicatori legati alla criminalità (indice di criminalità totale, furti con strappo, rapine in pubblica via) e anche l'indice di litigiosità mostrano contributi molto simili e significativi. Le barre sono quasi della stessa lunghezza, attestandosi intorno a 0.20. Questo suggerisce che questi indicatori sono tutti ugualmente importanti nel definire il fattore PC2.
- **Densità abitativa leggermente inferiore:** La densità abitativa contribuisce un po' meno rispetto agli indicatori di criminalità e litigiosità, anche se il suo valore (circa 0.19) è

comunque rilevante e molto vicino agli altri.

- **Un quadro di "criticità sociali" generalizzato:** L'insieme di questi indicatori (criminalità, litigiosità, densità abitativa) suggerisce che il fattore PC2 potrebbe rappresentare una sorta di "indice di criticità sociale" o "indice di complessità sociale e di sicurezza". Le province che presentano valori elevati in questo PC2 tendono ad avere un'alta incidenza di delitti denunciati, furti, rapine, un elevato numero di contenziosi legali e anche una maggiore densità di popolazione.
- **Utilità per l'analisi territoriale:** Per noi analisti territoriali, questo grafico è molto utile. Ci indica che se vogliamo capire le province che si distinguono per queste caratteristiche (ad esempio, quelle che necessitano di maggiori interventi in termini di sicurezza, giustizia o pianificazione urbanistica), dobbiamo prestare attenzione a tutte queste variabili in modo congiunto. Non c'è un singolo indicatore che sovrasta nettamente gli altri, ma piuttosto un insieme di fattori che sembrano operare insieme.

barplot_PC3



Questo grafico mostra le caratteristiche che contribuiscono maggiormente a definire la terza componente principale (PC3) dei dati socio-economici delle province italiane per l'anno 2023. In altre parole, ci dice quali indicatori tendono a variare insieme e caratterizzano un certo tipo di province.

Cosa rappresenta:

- Sull'asse verticale (Indicatori) ci sono diversi fattori socio-economici: ristoranti, indice di solitudine, bar, librerie, lavoratori domestici, indice di dipendenza strutturale, tasso di fecondità e quoziente di natalità.
- Sull'asse orizzontale (Contributo) vediamo quanto ciascun indicatore "pesa" nella definizione di questa componente (PC3). Un valore positivo indica che all'aumentare di questo indicatore, aumenta anche la PC3, e viceversa per valori negativi.

Evidenze e Pattern:

Notiamo una chiara divisione tra gli indicatori:

- **Indicatori con un forte contributo positivo:** "Ristoranti (incluse le attività di ristorazione mobile)", "Indice della solitudine", "Bar", "Librerie" e "Lavoratori domestici" hanno tutti un valore positivo che si spinge verso destra. Questo suggerisce che le province con un alto numero di ristoranti, un alto tasso di solitudine, molti bar e librerie, e un elevato numero di lavoratori domestici tendono a posizionarsi in un certo modo lungo questa componente PC3. Possiamo interpretare questo gruppo come indicatori che potrebbero essere

associati a una certa **vitalità sociale ed economica, ma anche a forme di disagio o complessità demografica (solitudine).**

- **Indicatori con un contributo negativo:** "Tasso di fecondità" e "Quoziente di natalità" hanno valori negativi e si trovano sulla sinistra del grafico. Questo significa che le province con un basso tasso di fecondità e un basso quoziente di natalità tendono a posizionarsi in modo opposto rispetto agli indicatori del gruppo precedente lungo questa componente. Si tratta di indicatori legati alla **dinamica demografica della popolazione.**
- **Indice di dipendenza strutturale:** Si colloca in una posizione intermedia con un contributo positivo moderato. Questo indicatore, che solitamente misura il peso della popolazione non attiva (anziani e bambini) sulla popolazione attiva, sembra avere un legame positivo con la prima categoria di indicatori, ma meno forte.

In sintesi per noi analisti territoriali:

La PC3 sembra catturare una contrapposizione tra province che presentano una maggiore densità di attività legate al tempo libero e alla socialità (ristoranti, bar, librerie), magari con un indice di solitudine più elevato, e province che si distinguono per una dinamica demografica più debole (bassa natalità e fecondità). Le province con più "vita sociale" e magari anche maggiore solitudine tendono a essere quelle dove nascono meno bambini. L'indice di dipendenza strutturale si allinea in parte con la prima tendenza, suggerendo che le aree con più servizi e attività potrebbero essere

anche avere una struttura demografica che comporta una maggiore dipendenza.

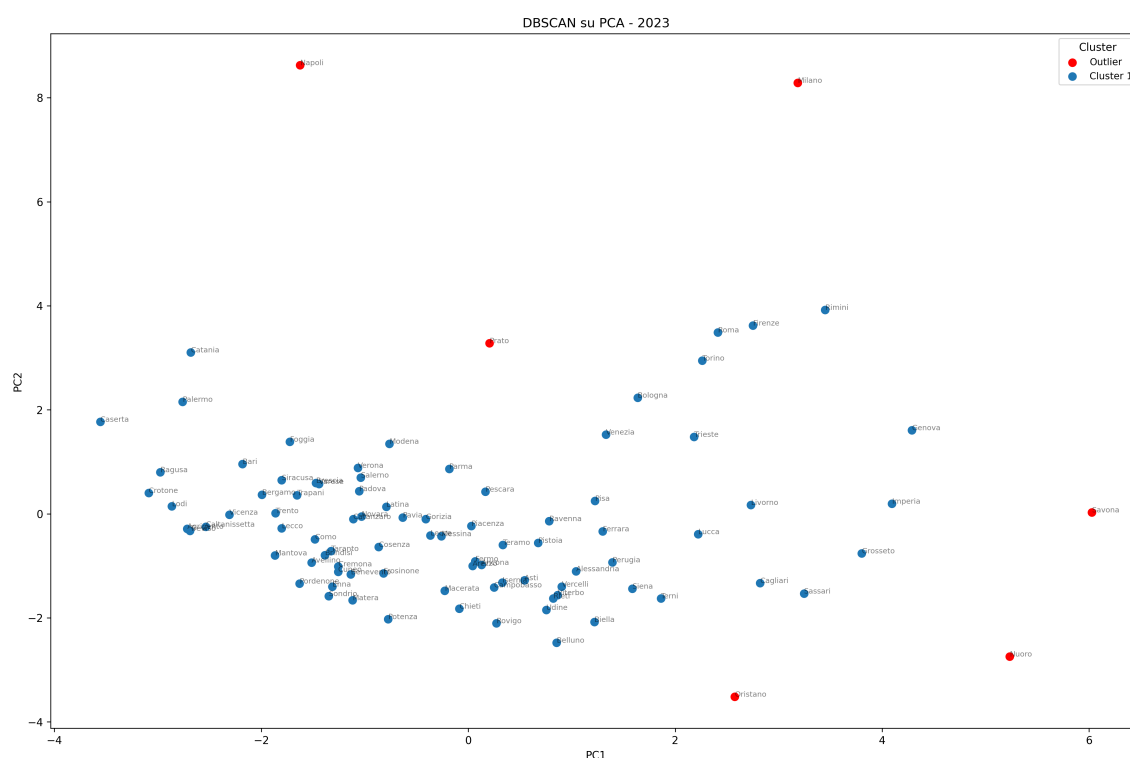
Potremmo quindi usare questa componente PC3 per identificare province che:

1. Sono aree urbane o turistiche con molta offerta di servizi e una popolazione che magari affronta sfide sociali come la solitudine.

2. Sono aree con una demografia in declino.

È interessante notare come questi indicatori si raggruppino in questo modo, offrendo spunti per analisi più approfondite su quali contesti territoriali stiamo osservando.

dbscan_pca_2023



Questo grafico a dispersione mostra le province italiane in uno spazio a due dimensioni, ridotto da dati socio-economici complessi (come indicato dal titolo "DBSCAN su PCA - 2023"). In pratica, le posizioni delle province sono state semplificate per evidenziare pattern e relazioni. Ogni punto rappresenta una provincia, con il suo nome associato.

Le coordinate dei punti, chiamate PC1 e PC2, sono il risultato di una riduzione di dimensionalità (PCA). Immagina che PC1 e PC2 siano due indicatori sintetici che riassumono la maggior parte delle informazioni presenti nei dati socio-economici originali. Province vicine in questo grafico tendono ad avere caratteristiche socio-economiche simili, mentre quelle lontane sono più dissimili.

Cosa notiamo in termini di pattern e evidenze:

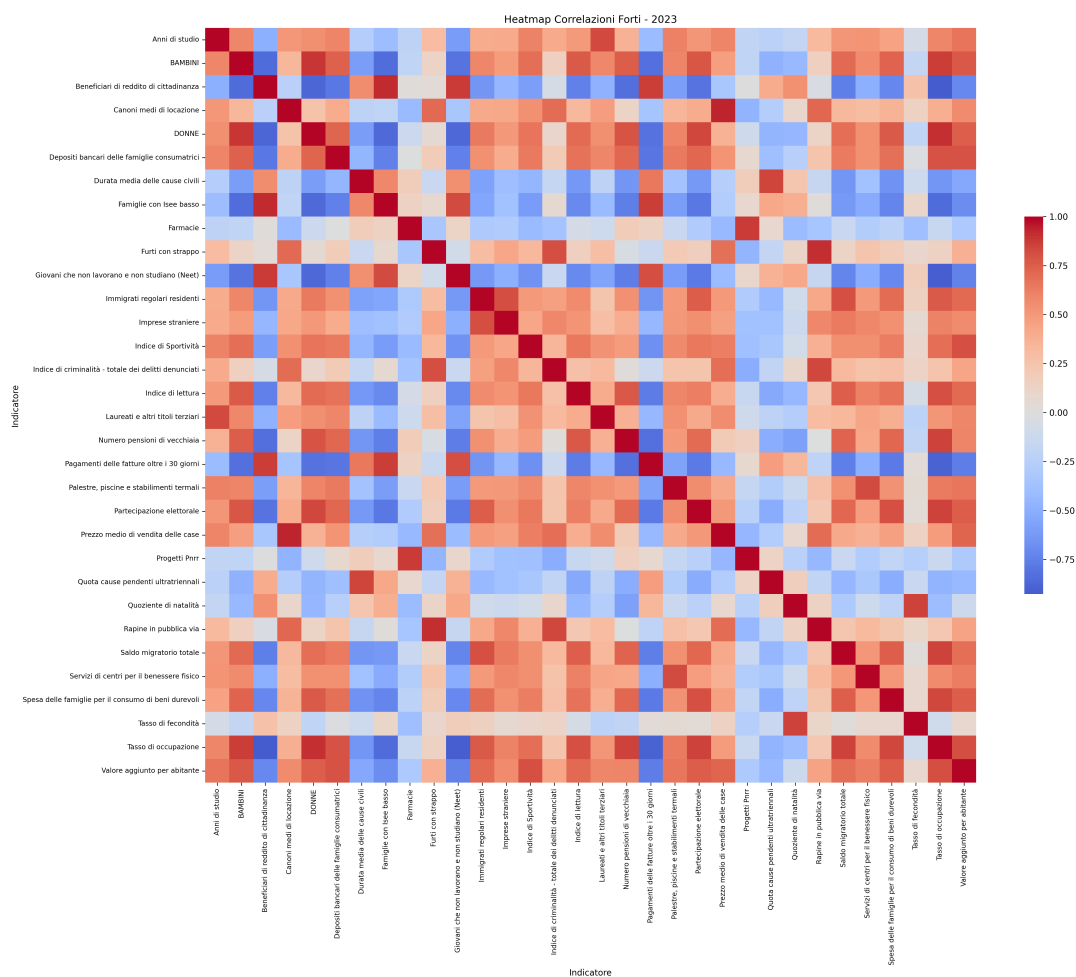
- **Clustering delle Province:** La maggior parte delle province (rappresentate dai punti blu, etichettati come "Cluster 1") si raggruppa in diverse aree, suggerendo la presenza di gruppi di province con profili socio-economici omogenei. Possiamo osservare una concentrazione di province in alto a sinistra, una serie più sparpagliata al centro e una certa dispersione verso destra. Questo indica che esistono zone d'Italia con caratteristiche simili.
- **Outlier (Punti Rossi):** Sono state identificate alcune province come "Outlier" (punti rossi). Queste sono province che si distinguono significativamente dalle altre, avendo caratteristiche socio-economiche uniche o non riconducibili ai pattern generali osservati. Nel grafico, vediamo diverse di queste province isolate:
- **Napoli:** Situata in alto a sinistra, molto distante dal gruppo principale. Questo suggerisce che Napoli abbia un profilo socio-economico particolarmente distintivo, diverso dalla maggior parte delle altre province.
- **Milano:** Anch'esso in una posizione isolata, in alto a destra. Questo potrebbe indicare un profilo economico e sociale diverso dal resto del paese, potenzialmente legato al suo ruolo di centro economico e finanziario.
- **Prato:** Posizionato più o meno al centro del grafico ma lontano da altri punti, suggerisce anche una sua specificità.

- **Nuoro e Savona:** Si trovano nella parte inferiore destra del grafico, molto isolate, indicando caratteristiche socio-economiche peculiari rispetto alla maggior parte delle altre province.
- **Oristano:** Anch'esso in basso a destra e isolato.
- **Distribuzione Geografica (Implicita):** Sebbene il grafico sia basato su dati socio-economici e non geografici diretti, potremmo fare delle ipotesi. Ad esempio, la forte presenza di province del Nord nel raggruppamento centrale e verso destra potrebbe suggerire alcune tendenze regionali, mentre la posizione isolata di alcune province del Sud (Napoli) o di isole (Nuoro) potrebbe evidenziare delle peculiarità. Tuttavia, è importante ricordare che la posizione qui deriva da variabili socio-economiche, non da coordinate geografiche pure.
- **Differenziazione tra Cluster 1 e Outlier:** La distinzione tra punti blu e rossi è netta. I punti rossi rappresentano province che l'algoritmo DBSCAN ha ritenuto non appartenere a nessun cluster denso, definendole appunto "rumore" o anomalie. Questo è un elemento chiave per identificare aree che richiedono un'analisi più approfondita e specifica.

In sintesi, il grafico ci offre una visione semplificata e aggregata delle province italiane basata sui loro profili socio-economici. Ci mostra che esistono dei raggruppamenti naturali di province con caratteristiche simili, ma soprattutto evidenzia un numero limitato di province che si discostano in modo significativo da

questi raggruppamenti, meritando un'attenzione particolare per comprendere le loro specificità.

heatmap_correlazioni_2023



Questa mappa di calore (heatmap) mostra le correlazioni tra diversi indicatori socio-economici e demografici per le province italiane nel 2023. Ogni casella rappresenta la forza e la direzione della relazione tra due indicatori:

- **Rosso intenso:** Indica una correlazione positiva molto forte. Significa che quando un indicatore aumenta, anche l'altro indicatore

tende ad aumentare in modo significativo.

- **Rosso chiaro:** Indica una correlazione positiva moderata.
- **Bianco/grigio chiaro:** Indica una correlazione debole o quasi nulla. I due indicatori si influenzano poco o per nulla.
- **Blu chiaro:** Indica una correlazione negativa moderata. Quando un

indicatore aumenta, l'altro tende a diminuire leggermente.

- **Blu intenso:** Indica una correlazione negativa molto forte. Quando un indicatore aumenta, l'altro tende a diminuire in modo significativo.

Evidenze e pattern principali che si notano:

- **Istruzione e reddito:** C'è una chiara correlazione positiva tra "Anni di studio", "Laureati e altri titoli terziari" e indicatori come "Valore aggiunto per abitante" e "Tasso di occupazione". Dove il livello di istruzione è più alto, mediamente si registrano anche maggiori redditi e occupazione.
- **Demografia e spesa:** Si osserva una correlazione positiva tra "Bambini", "Famiglie con Isee basso" e indicatori legati alla spesa per beni durevoli e consumi. Le famiglie con più figli e un ISEE basso tendono a spendere di più in questi beni, probabilmente per necessità.
- **Criminalità e disagio sociale:** "Indice di criminalità - totale dei delitti denunciati" mostra correlazioni positive moderate con indicatori di disagio come "Beneficiari di reddito di cittadinanza" e "Giovani che non lavorano e non studiano (Neet)". Questo suggerisce che le aree con maggiori difficoltà economiche e sociali potrebbero anche registrare livelli più alti di criminalità.
- **Immigrazione e occupazione:** C'è una correlazione positiva tra "Immigrati regolari residenti" e "Tasso di occupazione", anche se non fortissima. Questo suggerisce che la presenza di immigrati regolari è

associata, in media, a una maggiore occupazione nelle province.

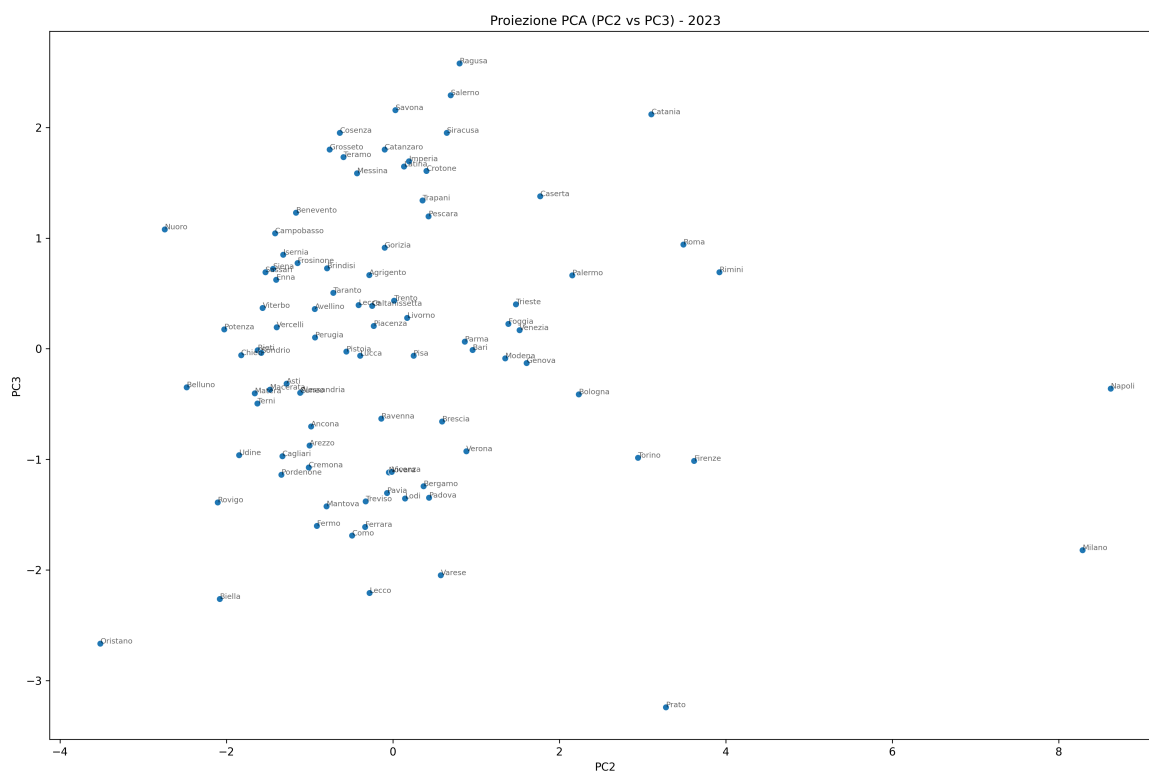
- **Servizi e qualità della vita:** Gli indicatori legati a servizi come "Farmacie" e "Palestre, piscine e stabilimenti termali" mostrano alcune correlazioni positive con indicatori di benessere, come il "Valore aggiunto per abitante". Le province con più servizi tendono ad essere anche più ricche.
- **Fenomeni demografici opposti:** Si nota una correlazione negativa, anche se non estrema, tra "Tasso di natalità" e "Numero pensioni di vecchiaia". Dove nascono più bambini, mediamente ci sono meno anziani (o proporzionalmente meno pensionati rispetto alla popolazione totale), il che è un pattern demografico atteso.
- **Mercato immobiliare:** Il "Prezzo medio di vendita delle case" sembra correlato positivamente con indicatori di benessere economico e istruzione. Le aree con prezzi immobiliari più alti tendono ad avere livelli di istruzione e reddito più elevati.
- **PNRR e impatto:** L'indicatore "Progetti Pnrr" mostra correlazioni variabili con altri indicatori, alcune positive (ad esempio con la spesa per beni durevoli) e altre meno evidenti. Questo è un dato interessante da approfondire per capire quali tipi di progetti stanno avendo maggiore impatto e su quali settori.
- **Incertezze e sfide:** Indicatori come "Quota cause pendenti ultratriennali" e "Rapine in pubblica via" mostrano correlazioni negative con indicatori di benessere economico, suggerendo che le province con maggiori criticità in questi ambiti potrebbero essere quelle

con minori risorse o maggiori problemi sociali.

In sintesi, la mappa evidenzia come diversi aspetti della vita socio-economica e demografica delle province italiane siano

interconnessi. Si possono identificare macro-aree di forza (ad esempio, istruzione-reddito-occupazione) e aree di potenziale fragilità (criminalità-disagio sociale, dipendenze da flussi migratori per l'occupazione).

scatter_pca_2023



Questa immagine è una rappresentazione grafica delle province italiane basata su dati socio-economici, dove ogni provincia è rappresentata da un punto. L'analisi ha ridotto molte variabili socio-economiche complesse a solo due dimensioni, chiamate "PC2" e "PC3". Immagina che queste due dimensioni siano come dei nuovi assi che ci aiutano a capire come le province si assomigliano o differiscono in base a questi aspetti socio-economici sintetizzati.

Cosa rappresenta:

- **Asse orizzontale (PC2):** Rappresenta una dimensione socio-economica che è il risultato della combinazione di diversi indicatori. Senza conoscere i dettagli precisi dell'analisi, possiamo interpretare che province con valori PC2 simili tendono ad avere caratteristiche socio-economiche simili lungo questa dimensione. Province con valori PC2 più alti potrebbero rappresentare, ad esempio, aree con maggiore sviluppo economico, maggiore concentrazione di attività produttive o determinate

dinamiche demografiche. Al contrario, valori PC2 più bassi potrebbero indicare aree con caratteristiche opposte.

- **Asse verticale (PC3):** Allo stesso modo, rappresenta un'altra dimensione socio-economica, anch'essa una combinazione di più indicatori originali. Province con valori PC3 simili si assomigliano per quanto riguarda questa particolare combinazione di fattori. Valori PC3 più alti potrebbero indicare, ad esempio, province con un certo livello di benessere, infrastrutture avanzate o particolari aspetti del mercato del lavoro. Valori PC3 più bassi suggerirebbero caratteristiche differenti.
- **Le province:** Ogni punto sull'immagine è una provincia italiana. La sua posizione è determinata dai suoi valori sulle dimensioni PC2 e PC3.

Evidenze e Pattern Notati:

1. Raggruppamenti Geografici

Impliciti: Nonostante l'analisi sia basata su dati socio-economici e non geografici diretti, si osservano alcuni raggruppamenti che suggeriscono differenze tra Nord, Centro e Sud Italia, e anche tra aree metropolitane e province meno popolate o con economie diverse.

- **Nord-Ovest:** Province come Milano, Torino, Firenze, Bologna, Genova sembrano posizionarsi in aree con valori PC2 e PC3 tendenzialmente più alti rispetto ad altre, suggerendo caratteristiche di sviluppo economico e/o demografico più accentuate. Si nota un certo addensamento di province del Nord (es. Bergamo,

Padova, Vicenza, Verona, Treviso, Venezia, Udine, Piacenza, Parma, Modena, Brescia, Mantova, Cremona, Pavia, Lodi, Varese, Como, Lecco, Sondrio, Biella, Novara, Vercelli, Asti) in un'area con valori di PC2 e PC3 moderati o leggermente positivi.

- **Centro:** Roma si posiziona da sola in un'area con valori di PC2 positivi e PC3 vicino allo zero. Altre province del Centro come Perugia, Terni, Arezzo, Siena, Pistoia, Lucca, Pisa, Livorno, Grosseto, Ancona, Macerata, Pesaro e Urbino, Rieti, Viterbo, Frosinone, Latina, Cassino (se è quella presente) tendono a posizionarsi in aree con valori di PC2 e PC3 medi, talvolta con una certa dispersione.
- **Sud e Isole:** Le province del Sud e delle Isole mostrano una maggiore dispersione. Alcune province, come Napoli, Palermo, Bari, Taranto, Lecce, Foggia, Catanzaro, Cosenza, Messina, Catania, Siracusa, Ragusa, Trapani, Enna, Caltanissetta, Agrigento, Nuoro, Oristano, Sassari, Potenza, Campobasso, Benevento, Avellino, Caserta, Salerno, Reggio Calabria, Crotone, si collocano in aree con valori di PC2 e PC3 più bassi, suggerendo caratteristiche socio-economiche medie o inferiori rispetto alle province del Nord più sviluppate, secondo le dimensioni considerate. In particolare, province come Biella e Oristano sembrano posizionarsi in aree con valori PC2 e PC3 molto bassi.

2. Province con Caratteristiche Distintive:

- **Milano:** Si trova in una posizione piuttosto isolata, con valori di PC2 e PC3 elevati, suggerendo che possieda

caratteristiche socio-economiche uniche o molto accentuate rispetto alla media delle altre province, probabilmente legate al suo ruolo di grande centro economico e finanziario.

- **Napoli:** Si posiziona in un'area con PC2 moderatamente positivo e PC3 molto basso, distinguendosi da altre grandi città del Sud o da quelle del Nord.
- **Roma:** Come Milano, è un punto di riferimento, posizionata con PC2 positivo e PC3 vicino allo zero.
- **Province "Estreme":** Alcune province come Nuoro, Oristano, Biella, Taranto, Palermo, Reggio Calabria, Crotone, si trovano in posizioni più isolate, suggerendo che abbiano profili socio-economici che le differenziano in modo più marcato dal gruppo centrale.

3. Presenza di Cluster e "Zone Grigie":

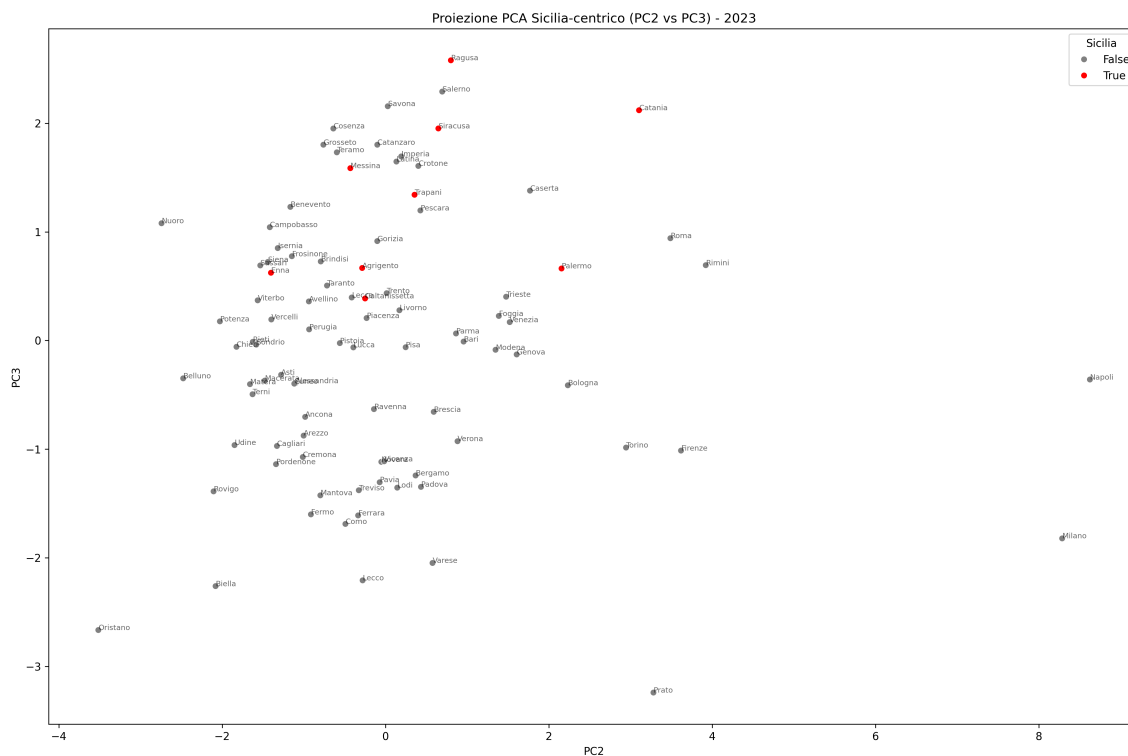
- C'è un'area più densamente popolata di punti, soprattutto tra PC2 da -2 a 2 e

PC3 da -1 a 2. Questo indica che molte province condividono caratteristiche socio-economiche "medie" o simili lungo queste due dimensioni.

- Le province più isolate, quelle con valori estremi su uno o entrambi gli assi, sono quelle che si discostano maggiormente da questo "gruppo centrale".

In sintesi, questo grafico ci permette di visualizzare come le province italiane si collocano su due grandi "dimensioni" socio-economiche. Le province più a destra e in alto tendono ad avere caratteristiche socio-economiche che questa analisi identifica come più sviluppate o particolari, mentre quelle più a sinistra e in basso tendono ad avere caratteristiche considerate "meno sviluppate" o semplicemente diverse. Ci aiuta a identificare gruppi di province simili e quelle che sono outliers, offrendo una prospettiva sintetica per confronti territoriali.

scatter_pca_2023_sicilia



Questa immagine è una rappresentazione grafica di un'analisi che ha cercato di riassumere le caratteristiche socio-economiche delle province italiane in due dimensioni principali. Immagina di avere tante informazioni per ogni provincia (reddito, occupazione, istruzione, ecc.) e che questa analisi le abbia "comprese" per farcele vedere più facilmente.

Ecco cosa rappresenta:

- **Le province italiane:** Ogni punto sull'immagine è una provincia italiana. Sono etichettate con i loro nomi.
- **Due dimensioni riassuntive (PC2 e PC3):** Gli assi orizzontale (chiamato PC2) e verticale (chiamato PC3) non rappresentano direttamente una singola variabile socio-economica, ma

sono combinazioni di molte di esse. Pensa a loro come a due "direzioni" principali lungo le quali le province differiscono in termini di benessere e sviluppo. Province che si trovano vicine sull'asse orizzontale (PC2) condividono caratteristiche simili per quanto riguarda queste combinazioni, così come quelle vicine sull'asse verticale (PC3).

- **Punti Rossi vs Grigi:** La differenza principale qui è legata alla Sicilia. I punti rossi rappresentano le province siciliane, mentre i punti grigi rappresentano le province del resto d'Italia.

Evidenze e pattern che si notano:

1. Le province siciliane sono in gruppi distinti: Osservando i punti rossi, notiamo che non sono sparsi casualmente, ma tendono a raggrupparsi in determinate aree del grafico. Questo suggerisce che le province siciliane hanno profili socio-economici che le rendono simili tra loro, e spesso le differenziano dalle altre province italiane.

2. Province siciliane con posizioni particolari:

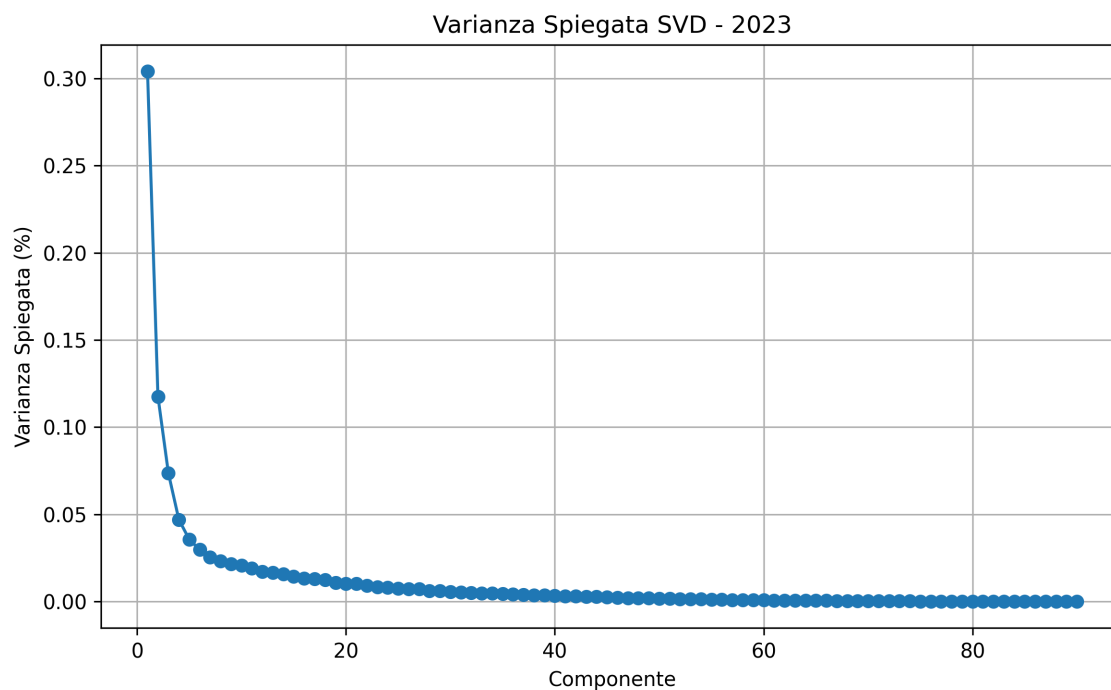
- **Catania e Ragusa:** Si trovano nella parte alta e destra del grafico, indicando che, secondo queste dimensioni riassuntive, potrebbero avere caratteristiche socio-economiche che le distinguono nettamente dalle altre province.
- **Siracusa e Trapani:** Sono un po' più a sinistra e in basso rispetto a Catania e Ragusa, ma comunque chiaramente identificate come province siciliane (rosse).
- **Messina:** Si posiziona leggermente a sinistra rispetto a Siracusa e Trapani.
- **Agrigento:** Si trova nella parte sinistra e centrale, con un valore di PC2 negativo e di PC3 vicino a zero.
- **Palermo:** Anche Palermo è un punto rosso, ma si trova in una posizione più isolata nella parte alta e centrale, distinguendosi ulteriormente.

3. Province "non siciliane" (grigie) più distribuite: I punti grigi, che rappresentano le altre province italiane, sono molto più dispersi su tutto il grafico. Questo indica una maggiore variabilità socio-economica tra le province del resto del paese. Non c'è un unico "cluster" per le province non siciliane, suggerendo che le differenze tra loro possono essere più sfumate o distribuite lungo direzioni diverse.

4. Possibili cluster di province "non siciliane": Sebbene più disperse, si possono intuire alcune aree dove le province grigie si addensano, come ad esempio un gruppo nella parte centrale/sinistra (province come Varese, Lecco, Sondrio, Biella, Novara, Cremona, etc.), e un altro gruppo nella parte destra (province come Milano, Torino, Napoli, ecc.).

In sintesi, questo grafico ci dice che, considerando le variabili socio-economiche analizzate, le province siciliane tendono ad avere profili distinti e tendenzialmente raggruppabili, differenziandosi dalle province del resto d'Italia, che invece mostrano una maggiore eterogeneità. I punti rossi ci evidenziano appunto le peculiarità, forse in termini di sviluppo o di tipologia di economia, delle province siciliane rispetto al quadro generale nazionale.

svd_varianza_spiegata_2023



Questa immagine mostra un grafico che illustra quanto "peso" o "importanza" hanno le diverse componenti estratte da un'analisi dei dati socio-economici delle province italiane. In pratica, stiamo cercando di capire quali aspetti principali emergono da un insieme complesso di informazioni economiche e sociali a livello provinciale.

Cosa rappresenta:

- **Asse orizzontale ("Componente"):** Rappresenta le diverse dimensioni o fattori che l'analisi ha identificato. Pensa a queste componenti come a delle "riassunti" di diversi indicatori. Ad esempio, la prima componente potrebbe rappresentare un mix di reddito e occupazione, la seconda potrebbe riguardare servizi e infrastrutture, e così via. Il grafico ne mostra un buon numero, fino a circa 90.

- **Asse verticale ("Varianza Spiegata (%)):** Indica quanta parte della variazione totale dei dati socio-economici viene catturata da ciascuna componente. Una componente che spiega una grande percentuale di varianza è molto importante perché riassume molte delle differenze osservate tra le province.

Evidenze e pattern:

1. **Forte concentrazione della varianza nelle prime componenti:** La curva inizia molto alta e scende drasticamente nei primi punti. Questo è il pattern più evidente e importante. Significa che le prime poche componenti spiegano la maggior parte delle differenze che osserviamo nei dati socio-economici delle province italiane.
- La **prima componente** (a sinistra) da sola cattura circa il 30% della varianza

totale. Questo ci dice che esiste un fattore primario che differenzia fortemente le province italiane in termini socio-economici. Probabilmente rappresenta il livello generale di sviluppo o benessere.

- La **seconda componente** aggiunge un altro bel pezzo, arrivando forse a spiegare circa il 12-13% della varianza. Questo è ancora significativo.
- La **terza e la quarta componente** continuano a contribuire, ma con un impatto via via minore.

2. Diminuzione rapida e poi stabilizzazione: Dopo le prime 5-10 componenti, la curva si appiattisce notevolmente. Questo suggerisce che le componenti successive spiegano sempre meno varianza, diventando via via meno rilevanti per descrivere le caratteristiche socio-economiche delle province.

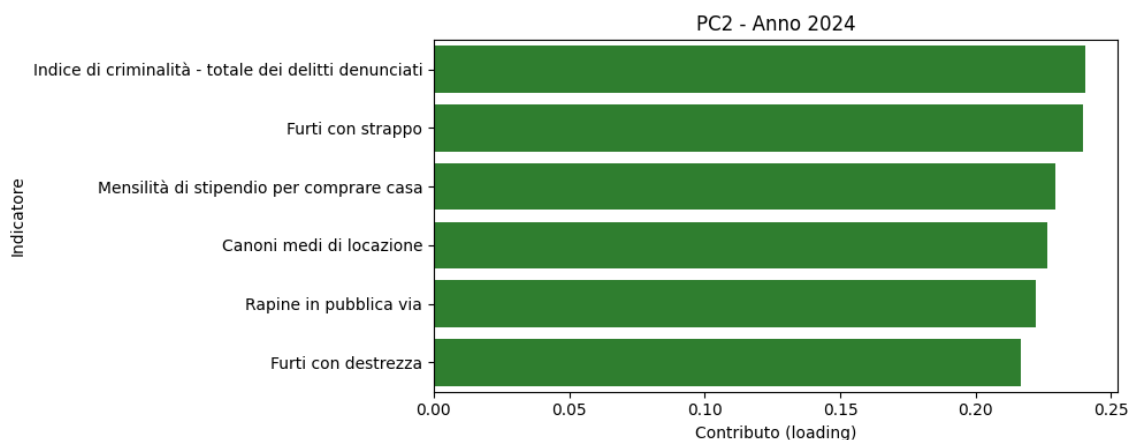
3. Presenza di molte componenti con basso impatto: Anche se le prime componenti sono dominanti, il grafico mostra che ci sono molte altre componenti (fino alla fine) che, singolarmente, spiegano una porzione molto piccola della

varianza. Questo può indicare che ci sono molte sfumature e dettagli specifici nell'andamento socio-economico delle province, ma che sono meno generali e omogenei rispetto a quelli catturati dalle prime componenti.

In sintesi per un analista territoriale:

Questa analisi ci dice che per capire le principali differenze socio-economiche tra le province italiane, ci dobbiamo concentrare su un numero relativamente piccolo di fattori chiave. La prima componente è il driver principale, probabilmente legato al "livello di sviluppo" generale. Le successive componenti aggiungono dettagli importanti ma via via meno dominanti. Se dovessimo fare una sintesi, potremmo dire che c'è una forte "struttura" socio-economica nelle province italiane, dominata da pochi elementi fondamentali, ma anche una certa eterogeneità che emerge nelle componenti meno importanti. Questo ci aiuta a capire dove focalizzare i nostri sforzi di analisi e, potenzialmente, di intervento, partendo dai fattori più influenti.

barplot_PC2



Questo grafico a barre orizzontali mostra come diversi indicatori socio-economici contribuiscono a definire una specifica dimensione (denominata "PC2") per l'anno 2024. Ogni barra rappresenta un indicatore e la sua lunghezza ci dice quanto quell'indicatore è importante nel definire questa dimensione.

In pratica, questo grafico ci sta dicendo che le province italiane che ottengono un punteggio elevato in questa dimensione "PC2" tendono ad avere contemporaneamente:

- **Un indice di criminalità più alto:** Questo include un maggior numero totale di delitti denunciati.
- **Maggior numero di furti con strappo.**
- **Una mensilità di stipendio necessaria per comprare casa più alta:** In altre parole, il costo della casa in proporzione al reddito è più elevato.
- **Canoni medi di locazione più elevati:** Affittare una casa costa di più.

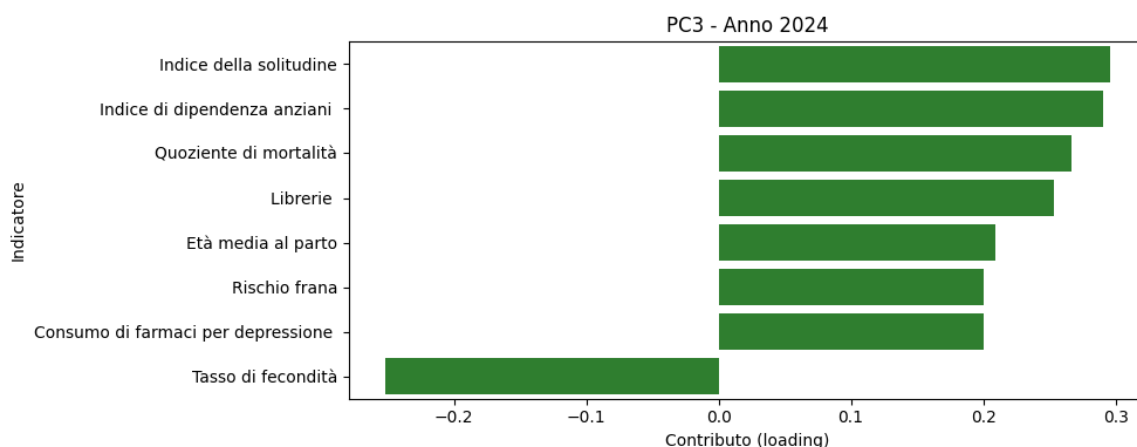
- **Maggior numero di rapine in pubblica via.**
- **Maggior numero di furti con destrezza.**

Pattern e Evidenze:

Il pattern più evidente è che la dimensione "PC2" sembra essere fortemente correlata a **indicatori di insicurezza e costo della vita elevato**. Tutte le barre presenti nel grafico hanno un contributo positivo e di magnitudo simile (tra circa 0.20 e 0.24), suggerendo che questi fattori non sono isolati ma tendono a presentarsi insieme.

In sintesi, le province che si collocano "in alto" in questa dimensione "PC2" sono quelle dove si registrano più reati di vario tipo e dove vivere è economicamente più impegnativo, sia per l'acquisto che per l'affitto di un'abitazione. Questo è un dato importante per capire le sfide territoriali, ad esempio per interventi di sicurezza, politiche abitative o per valutare l'attrattività economica delle diverse aree del paese.

barplot_PC3



Questo grafico a barre rappresenta il contributo di diversi indicatori socio-economici a un fattore specifico, denominato "PC3", per l'anno 2024. Ogni barra indica quanto un determinato indicatore "pesa" in questo fattore.

In parole semplici, stiamo guardando quali elementi sono più fortemente associati a questo fattore "PC3".

Cosa si nota:

- **Indicatori con contributo positivo (barre a destra dello zero):** Questi sono gli indicatori che, quando aumentano, contribuiscono ad aumentare anche il valore di questo fattore "PC3". Notiamo che:
- **"Indice della solitudine" e "Indice di dipendenza anziani"** sono gli indicatori che hanno il contributo positivo più elevato. Questo suggerisce che le province con un alto tasso di solitudine e un'alta percentuale di anziani dipendenti tendono ad avere un valore più alto per questo fattore "PC3".
- Anche **"Quoziente di mortalità", "Librerie" ed "Età media al parto"** contribuiscono positivamente, ma in misura minore rispetto ai primi due.
- Il **"Rischio frana"** ha un contributo positivo, ma il suo impatto è leggermente inferiore rispetto agli indicatori appena citati.
- Infine, il **"Consumo di farmaci per depressione"** ha un contributo positivo, ma è quello con il valore più basso tra quelli positivi.
- **Indicatori con contributo negativo (barra a sinistra dello zero):** Questo indicatore, quando aumenta, tende a diminuire il valore del fattore "PC3".

- L'unico indicatore con un contributo negativo significativo è il **"Tasso di fecondità"**. Questo significa che le province con un alto tasso di fecondità tendono ad avere un valore più basso per questo fattore "PC3".

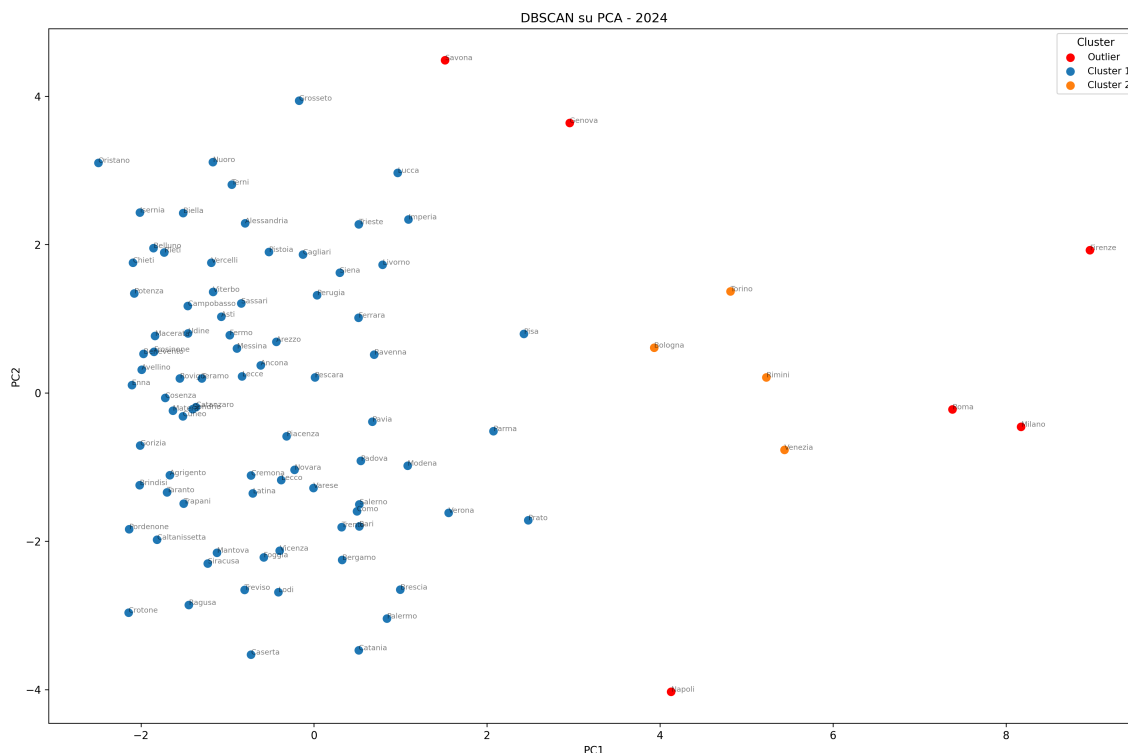
Pattern principali:

1. **Invecchiamento e isolamento sociale legati al "PC3":** C'è una forte associazione tra invecchiamento della popolazione (alto indice di dipendenza anziani, alta età media al parto, alto quoziente di mortalità) e problematiche sociali come la solitudine, con un impatto positivo su questo fattore.
2. **Salute mentale e "PC3":** Un elevato consumo di farmaci per la depressione è associato a questo fattore, suggerendo che potrebbe essere un indicatore di disagio o fragilità in queste aree.
3. **Dinamiche demografiche opposte:** Il basso tasso di natalità (basso tasso di fecondità) è associato a un aumento di questo fattore, mentre un tasso di natalità più alto lo riduce.

In sintesi, il fattore "PC3" sembra catturare una dimensione legata alla **fragilità socio-demografica e alla presenza di problematiche sociali e di invecchiamento**, dove un più alto indice di solitudine, dipendenza degli anziani e mortalità, unito a una bassa natalità, sono caratteristiche predominanti. Le librerie, sorprendentemente, contribuiscono positivamente a questo fattore, il che potrebbe essere interpretato in vari modi (magari aree più "abitabili" o con più servizi che attirano anche fasce più anziane della popolazione, o un indice legato a una minore vivacità culturale

percepita in contrapposizione a contesti più dinamici e giovani).

dbscan_pca_2024



Questo grafico rappresenta una mappa delle province italiane basata su diverse variabili socio-economiche, ridotte a due dimensioni principali (PC1 e PC2) per semplificare la visualizzazione. Ogni punto sull'asse rappresenta una provincia, e la sua posizione è determinata dai suoi valori aggregati su queste due dimensioni.

Ci sono tre gruppi di province evidenziati:

- **Cluster 1 (Punti blu):** Questo è il gruppo più numeroso e contiene la maggior parte delle province italiane. Queste province tendono ad avere caratteristiche socio-economiche simili tra loro. La loro distribuzione in questo spazio bidimensionale indica

che formano una "massa" centrale di province con profili medi.

- **Cluster 2 (Punti arancioni):** Questo gruppo è più piccolo e si trova in una zona distinta del grafico, separata dal cluster principale. Le province in questo gruppo (come Bologna, Torino, Rimini, Venezia) hanno caratteristiche socio-economiche che le differenziano dalla maggior parte delle altre. Potrebbero rappresentare province con economie più dinamiche, più orientate ai servizi o con caratteristiche specifiche che le rendono uniche rispetto alla media nazionale.

- **Outlier (Punti rossi):** Questi sono i punti che si trovano più distanti sia dal cluster principale che dal cluster 2. Sono province che presentano caratteristiche socio-economiche particolarmente distintive e non facilmente riconducibili agli altri gruppi. Nel grafico, vediamo alcune di queste province come Roma, Milano, Firenze, Napoli e Savona. Questi punti isolati suggeriscono che queste province hanno profili unici che le rendono anomale rispetto al quadro generale. Ad esempio, Milano e Roma, pur essendo grandi centri, potrebbero avere pattern socio-economici che le distinguono in base alle variabili considerate. Firenze e Napoli, anch'esse significative, potrebbero avere caratteristiche peculiari che le isolano. Savona appare come un outlier anche a livello geografico, sebbene sia stata individuata dall'algoritmo sulla base dei dati.

Pattern e Evidenze Notate:

1. Concentrazione Generale: La maggior parte delle province italiane (Cluster 1) si distribuisce in un'area relativamente compatta, suggerendo una certa omogeneità socio-economica diffusa sul territorio, quando si considerano le dimensioni ridotte.

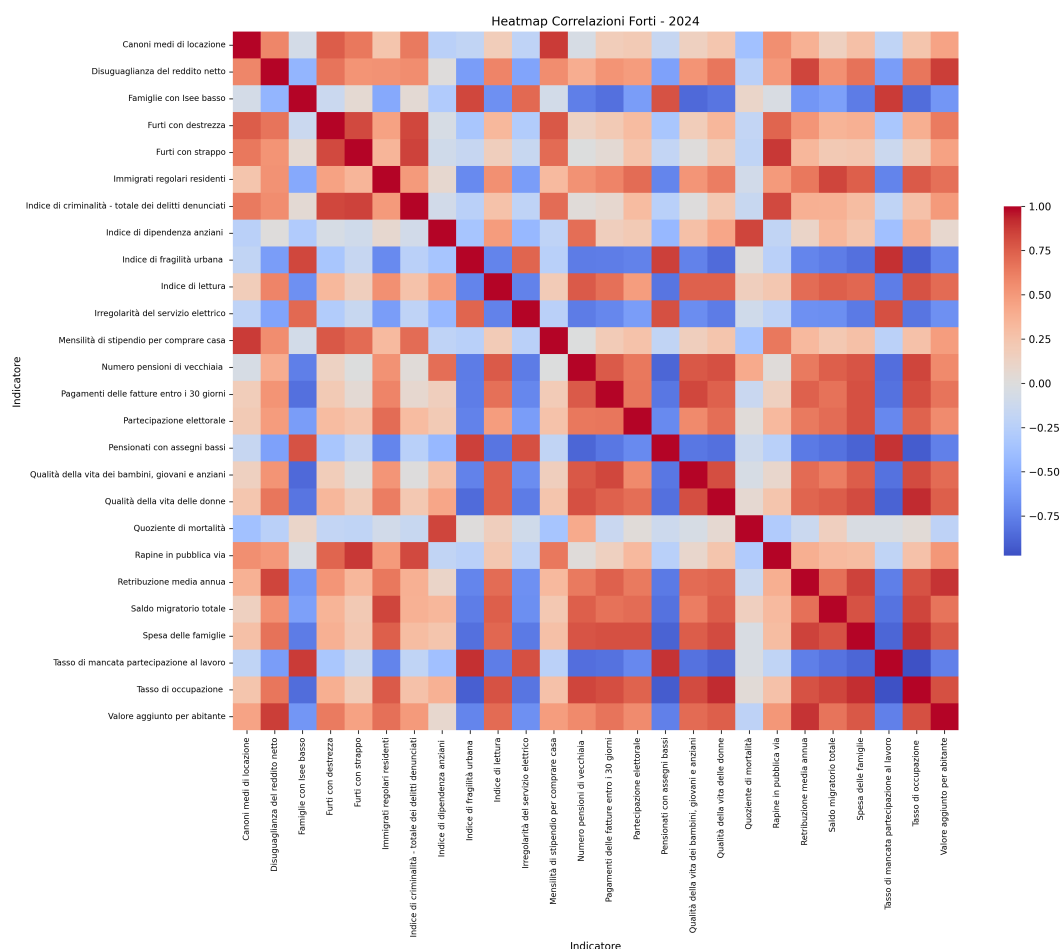
2. Distinzione del Nord-Est (probabile): Le province arancioni (Cluster 2) sembrano raggrupparsi, con molte di esse

situate in aree geografiche che tradizionalmente includono città importanti del nord Italia (come Bologna, Torino, Venezia). Questo potrebbe indicare un modello di sviluppo economico o sociale distinto rispetto ad altre aree.

3. Eccezionalità delle Grandi Città e Capitali: Le province rosse (Outlier), che includono le maggiori città metropolitane (Milano, Roma, Napoli) e un centro importante come Firenze, emergono come "speciali". La loro distanza dagli altri gruppi evidenzia che, nonostante la loro importanza, presentano profili socio-economici che non rientrano nei pattern medi o in quelli del cluster più dinamico, secondo le variabili analizzate. Questo potrebbe essere dovuto a economie molto sviluppate ma con una forte presenza di settori specifici (es. turismo per Firenze, pubblica amministrazione per Roma), o a dinamiche demografiche e sociali particolari.

4. Potenziale per Ulteriori Indagini: Le province identificate come outlier e quelle nel cluster 2 meritano un'attenzione particolare. Analizzare le variabili socio-economiche specifiche che guidano queste distinzioni può fornire insight preziosi per politiche territoriali mirate. Ad esempio, si potrebbe voler capire cosa rende una provincia come Savona un outlier, o quali sono i fattori chiave che differenziano il Cluster 2 dal Cluster 1.

heatmap_correlazioni_2024



Questo grafico è una mappa di calore che visualizza le correlazioni tra diversi indicatori socio-economici a livello provinciale per l'anno 2024. Ogni cella rappresenta la relazione tra due indicatori: il colore indica la forza e la direzione di questa relazione, mentre la posizione ne indica gli elementi specifici.

La scala dei colori va dal blu (correlazione negativa, cioè al diminuire di un indicatore, l'altro tende a diminuire, o al contrario, al diminuire di uno, l'altro tende ad aumentare) al rosso (correlazione positiva, cioè i due indicatori tendono a muoversi nella stessa direzione). Il bianco indica una correlazione neutra o molto

debole. I valori più intensi di rosso e blu indicano correlazioni più forti.

Osservando la mappa, possiamo notare alcuni pattern interessanti:

- **Legami tra indicatori economici:** Si notano forti correlazioni positive tra indicatori come "Retribuzione media annua", "Valore aggiunto per abitante" e "Tasso di occupazione". Questo è prevedibile, poiché un'economia più forte tende a riflettersi in stipendi più alti e maggiore impiego. Anche i "Canoni medi di locazione" sembrano correlati positivamente con questi indicatori, suggerendo che le aree con

economie più vivaci tendono ad avere anche costi abitativi maggiori.

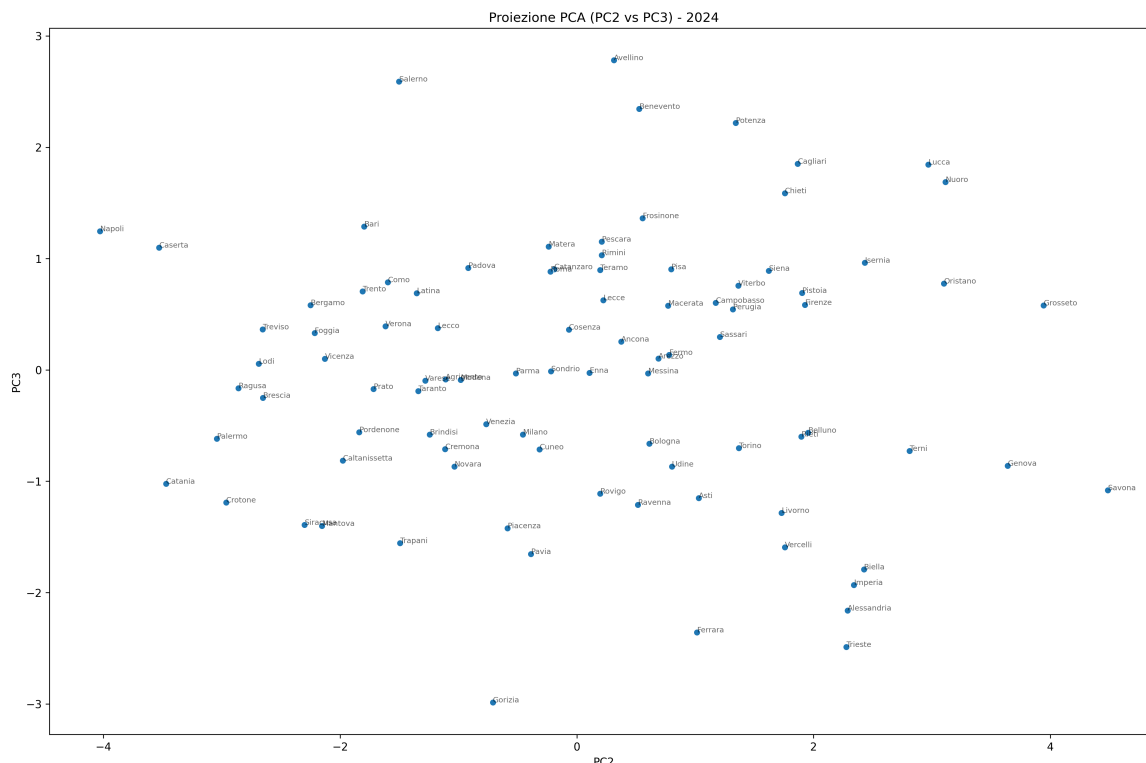
- **Segnali di disagio socio-economico:** Indicatori come "Disuguaglianza del reddito netto", "Famiglie con Isee basso" e "Pensionati con assegni bassi" mostrano correlazioni spesso negative con gli indicatori di prosperità economica. Questo indica che dove c'è più povertà e disuguaglianza, l'attività economica generale tende ad essere meno dinamica. Allo stesso modo, "Indice di criminalità - totale dei delitti denunciati" e "Rapine in pubblica via" mostrano correlazioni positive con indicatori di difficoltà, come le famiglie con Isee basso.
- **Indicatori demografici e sociali:** L'"Indice di dipendenza anziani" e l'"Indice di lettura" sembrano avere correlazioni miste. Ad esempio, l'"Indice di dipendenza anziani" mostra una debole correlazione negativa con la "Qualità della vita dei bambini, giovani e anziani", suggerendo che dove ci sono molti anziani a carico, potrebbe esserci una pressione su altri aspetti del benessere

sociale. L'"Indice di lettura" appare positivamente correlato con indicatori di istruzione o benessere, ma non in modo estremamente marcato.

- **Particolarità:** Notiamo alcune correlazioni che potrebbero richiedere un'analisi più approfondita. Ad esempio, la correlazione tra "Indice di fragilità urbana" e alcuni indicatori di criminalità o difficoltà economica potrebbe suggerire che le aree urbane più vulnerabili tendono a concentrare problematiche sociali ed economiche. Allo stesso modo, la correlazione tra "Tasso di mancata partecipazione al lavoro" e altri indicatori di debolezza economica merita attenzione.

In sintesi, questa heatmap ci offre una panoramica immediata delle interconnessioni tra diverse sfaccettature della vita socio-economica nelle province italiane. Ci permette di identificare a colpo d'occhio quali fenomeni tendono a verificarsi insieme, sia in termini positivi (crescita e benessere) che negativi (difficoltà e marginalità), fornendo spunti preziosi per politiche territoriali mirate.

scatter_pca_2024



Questo grafico è una rappresentazione visiva di come le province italiane si posizionano rispetto a due dimensioni socio-economiche aggregate, chiamate qui PC2 e PC3. Immagina queste dimensioni come "super-indicatori" che riassumono una serie di dati complessi. Le province che si trovano vicine tra loro in questo grafico condividono caratteristiche socio-economiche simili secondo questi due indicatori.

Cosa vediamo:

- **Posizionamento delle province:** Ogni punto sul grafico rappresenta una provincia italiana e il suo nome è indicato accanto al punto. La posizione orizzontale del punto indica il suo valore sull'indicatore PC2, mentre la posizione verticale indica il suo valore sull'indicatore PC3.

Evidenze e pattern che possiamo notare:

1. Raggruppamenti Geografici e Socio-Economici:

- **Nord-Est e Nord-Ovest:** Si nota una concentrazione di province del Nord-Est (come Treviso, Lodi, Venezia, Verona, Vicenza, Padova, Udine) e del Nord-Ovest (come Milano, Torino, Como, Bergamo, Brescia, Varese, Novara, Cremona) che tendono a posizionarsi nella parte destra e centrale-alta del grafico (valori positivi per PC2 e a volte anche per PC3). Questo potrebbe indicare che queste aree tendono a condividere certe caratteristiche socio-economiche.
- **Centro Italia:** Le province del centro Italia (come Firenze, Siena, Perugia,

Ancona, Arezzo, Terni) si trovano in una posizione intermedia, spesso con valori misti per PC2 e PC3.

- **Sud Italia e Isole:** Le province del Sud Italia e delle Isole (come Napoli, Caserta, Salerno, Bari, Lecce, Catanzaro, Palermo, Catania, Trapani, Cagliari, Nuoro, Oristano) tendono a concentrarsi nella parte sinistra del grafico (valori negativi per PC2) e nella parte bassa (valori negativi per PC3), suggerendo un'altra possibile area con caratteristiche socio-economiche omogenee, ma diverse da quelle del Nord. Alcune province del Sud (come Avellino, Benevento, Potenza, Chieti) si trovano leggermente più in alto, indicando differenze all'interno del Mezzogiorno.

2. Province "Estreme":

- Alcune province si trovano molto a sinistra (es. Napoli, Caserta, Vicenza), suggerendo che si distinguono significativamente dagli altri secondo l'asse PC2.
- Altre si trovano molto in alto (es. Avellino, Belluno) o molto in basso (es. Gorizia, Crotone, Caltanissetta), indicando posizioni "estreme" rispetto all'asse PC3.

3. Relazione tra PC2 e PC3:

- Sembra esserci una tendenza generale per cui le province con valori alti in PC2 tendono ad avere valori medi o bassi in PC3, e viceversa. Ad esempio, molte province del Nord (alto PC2) si trovano nella parte centrale-alta o centrale del grafico per PC3, mentre

molte province del Sud (basso PC2) si trovano nella parte centrale-bassa.

- Tuttavia, ci sono anche delle eccezioni, come alcune province del Nord-Est (es. Treviso) che si posizionano in alto sia in PC2 che in PC3, o alcune province del Sud (es. Potenza) che si distinguono per un valore più alto in PC3.

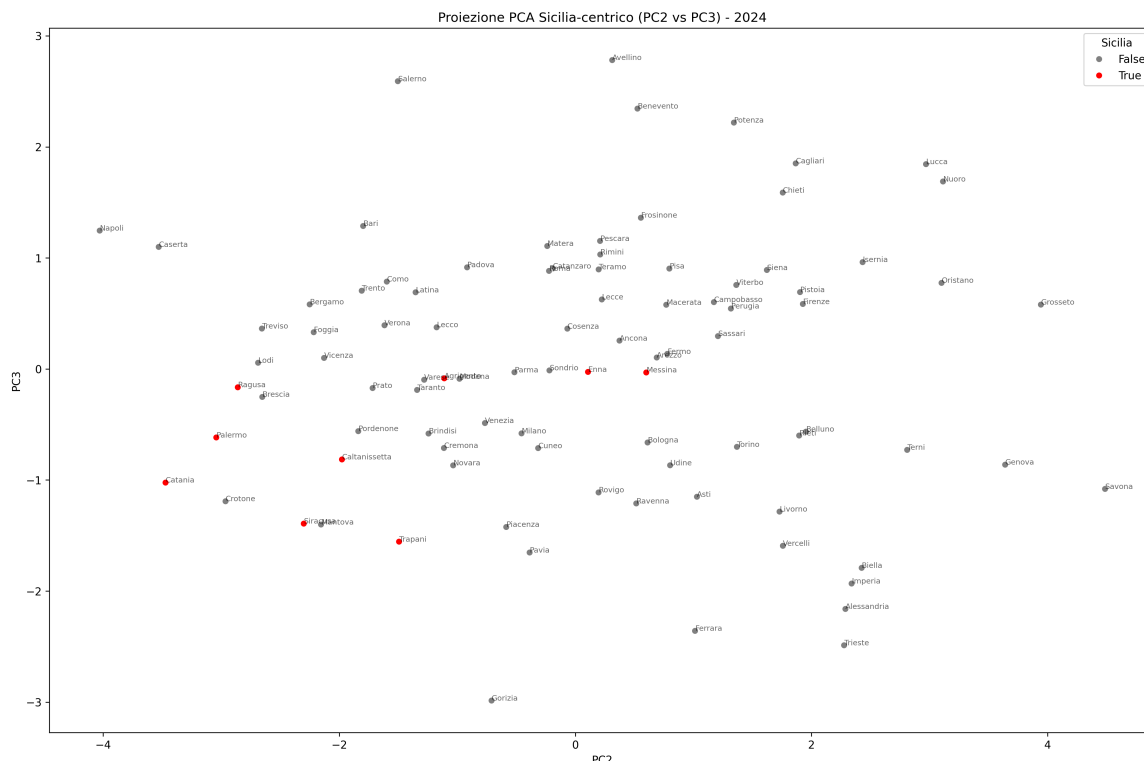
In sintesi per un analista territoriale:

Questo grafico ci permette di "mappare" le province italiane non solo in base alla loro posizione geografica, ma soprattutto in base a come si differenziano o assomigliano in termini di un insieme di fattori socio-economici. Possiamo identificare cluster di province che sembrano condividere profili simili, il che può essere utile per:

- **Politiche mirate:** Capire quali aree sono più simili tra loro può aiutare a disegnare interventi socio-economici più efficaci e omogenei.
- **Analisi comparative:** Possiamo confrontare province apparentemente diverse per capire dove risiedono le loro similitudini o differenze in termini di questi indicatori aggregati.
- **Identificazione di peculiarità:** Province che si trovano isolate in questo spazio di rappresentazione potrebbero rappresentare casi particolari da indagare più a fondo.

In sostanza, questo grafico riduce la complessità di molti dati socio-economici in un'unica immagine bidimensionale, rendendo più immediate le relazioni e le differenze tra le province italiane.

scatter_pca_2024_sicilia



Questo grafico mostra le province italiane posizionate su due assi principali, chiamati PC2 e PC3. Questi assi sono il risultato di un'analisi che cerca di riassumere molte informazioni socio-economiche su ciascuna provincia in soli due valori. In parole semplici, le province che si trovano vicine tra loro nel grafico tendono ad avere caratteristiche socio-economiche simili, mentre quelle lontane sono più diverse.

Il grafico è colorato per evidenziare le province della Sicilia: quelle rosse indicano la Sicilia ("True"), mentre quelle grigie sono tutte le altre province italiane ("False"). L'anno di riferimento è il 2024.

Cosa notiamo:

- **Le province siciliane tendono a raggrupparsi:** Si nota chiaramente che le province rosse (Messina,

Palermo, Catania, Caltanissetta, Trapani, Enna, Ragusa, Siracusa) sono concentrate principalmente nella parte sinistra del grafico, soprattutto nel quadrante inferiore sinistro. Questo suggerisce che, in base ai dati analizzati, queste province condividono caratteristiche socio-economiche che le distinguono dal resto d'Italia.

- **Divisione Est/Ovest o Nord/Sud:** Osservando la disposizione generale delle province grigie (non siciliane), possiamo intravedere una certa tendenza. Le province più a sinistra del grafico (con valori negativi sull'asse PC2) sembrano essere più concentrate al Sud e nelle isole (come Napoli, Bari, Matera, Cagliari, Crotona), mentre quelle più a destra (con valori positivi sull'asse PC2) si trovano

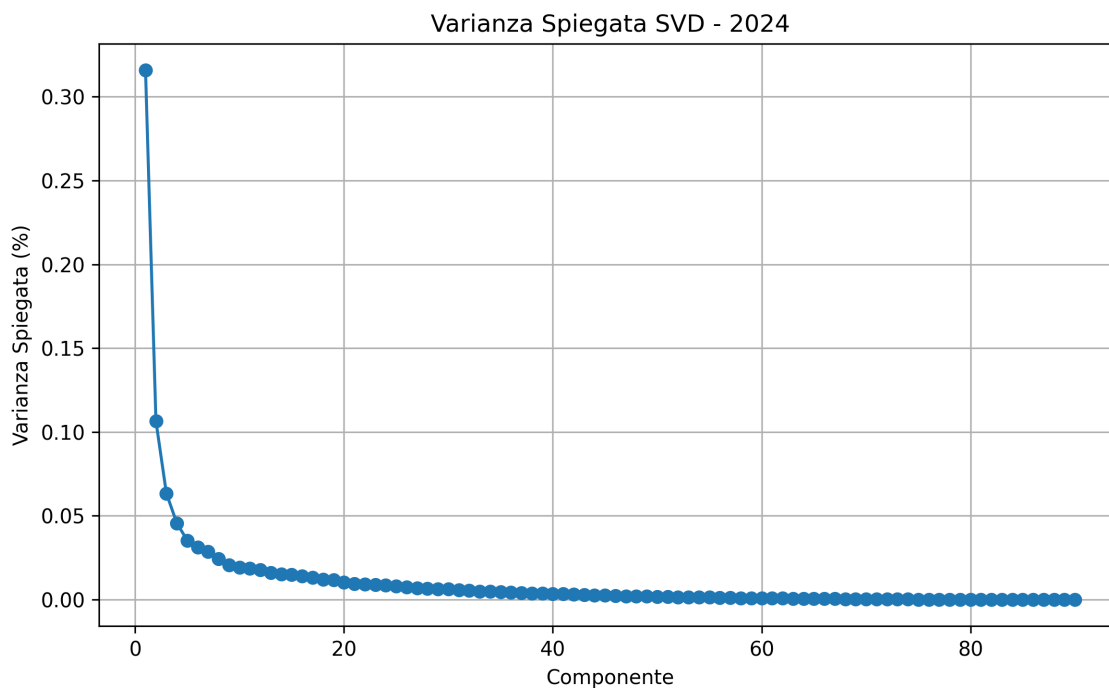
prevalentemente al Nord e Centro Italia (come Milano, Bologna, Torino, Firenze, Venezia).

- **Caratteristiche uniche di alcune province:** Alcune province, sia siciliane che non, si trovano isolate rispetto ai gruppi principali. Ad esempio, province come Nuoro o Oristano (Sardegna) si trovano nella parte alta sinistra, mentre altre come Belluno o Gorizia sono più isolate nella parte destra o centrale. Questo indica che potrebbero avere profili socio-economici unici.
- **Sovrapposizione e differenze all'interno della Sicilia:** Anche se le province siciliane tendono a raggrupparsi, si notano anche delle leggere differenze tra di loro. Ad

esempio, Trapani e Siracusa sono più a destra rispetto a Catania e Enna, suggerendo che, pur rimanendo all'interno di un cluster "siciliano", presentano sfumature di caratteristiche socio-economiche differenti.

In sintesi, il grafico ci mostra che le province siciliane, nel complesso, presentano un pattern socio-economico distinto rispetto alla media delle altre province italiane, tendendo a collocarsi in una specifica area dello spazio definito dai fattori analizzati (PC2 e PC3). C'è anche una chiara divisione tra province meridionali e centro-settentrionali, anche se con molte eccezioni e sfumature.

svd_varianza_spiegata_2024



Questo grafico illustra come i diversi aspetti (componenti) estratti dall'analisi dei dati socio-economici delle province italiane contribuiscono a spiegare la variabilità totale di questi dati. In parole semplici, ci dice quanto "informazione" c'è in ciascun aspetto individuato.

Cosa rappresenta:

- **Asse Orizzontale (Componente):** Indica i diversi "aspetti" o "fattori" che emergono dall'analisi dei dati socio-economici. Questi aspetti sono ordinati dal più importante al meno importante. La prima componente rappresenta il fattore che spiega la maggior parte delle differenze tra le province, la seconda componente spiega la parte rimanente più significativa, e così via.
- **Asse Verticale (Varianza Spiegata (%)):** Misura quanto ciascuna componente è in grado di "spiegare" le differenze osservate nei dati socio-economici. Un valore più alto significa che quella componente è più rilevante nel descrivere le caratteristiche delle province.

Evidenze e pattern notati:

1. Dominanza delle prime componenti:

Si osserva un calo molto ripido della varianza spiegata subito dopo le prime componenti. La prima componente da sola spiega circa il 30% della variabilità totale dei dati. La seconda componente aggiunge un altro buon 10-12%, e la terza circa il 5-6%. Questo significa che solo questi primi pochi aspetti riescono a catturare una fetta considerevole delle differenze

complessive tra le province italiane in termini socio-economici.

2. Coda lunga di importanza marginale:

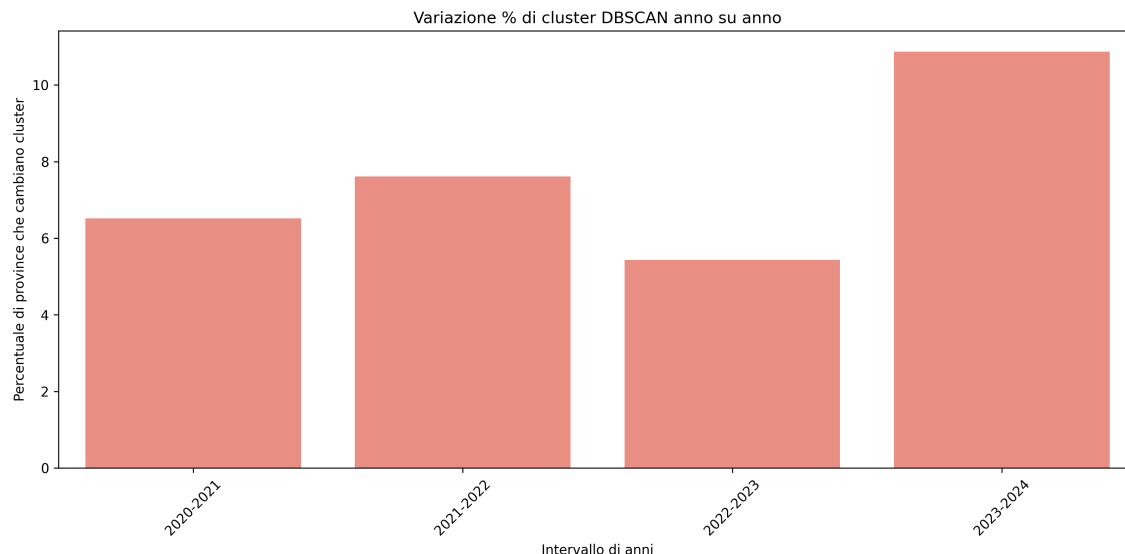
Dopo le prime 3-4 componenti, la varianza spiegata da ogni ulteriore componente diventa progressivamente molto piccola, attestandosi su valori vicini allo zero. Ciò indica che le componenti successive hanno un impatto minimo nel descrivere le differenze tra le province. Molte di queste componenti spiegano solo una frazione infinitesimale delle variazioni totali.

3. Possibilità di sintesi:

Questo pattern suggerisce che i dati socio-economici delle province italiane possono essere efficacemente sintetizzati in un numero relativamente piccolo di fattori (probabilmente tra le prime 3 e 10 componenti, a seconda del livello di dettaglio desiderato). Per esempio, se volessimo semplificare la rappresentazione delle province, potremmo concentrarci sulle prime 3-5 componenti perché già queste racchiudono la maggior parte dell'informazione rilevante. Le componenti successive potrebbero rappresentare rumore o aspetti molto specifici e poco generalizzabili.

In sintesi, l'analisi mostra che esiste una forte concentrazione della variabilità dei dati socio-economici nelle prime componenti individuate. Questo è un segnale positivo per chi vuole semplificare l'analisi e comprendere i driver principali delle differenze territoriali, in quanto un numero limitato di fattori riesce a dare un quadro molto rappresentativo della situazione.

cambiamenti_cluster_percentuale



L'immagine mostra una serie di barre che rappresentano la **variazione percentuale annuale delle province italiane che cambiano cluster**, come risultato di un'analisi DBSCAN sui dati socio-economici. In termini più semplici, ci dice quante province, ogni anno, si spostano da un gruppo socio-economico all'altro.

Ecco cosa si nota:

- **Periodo di analisi:** Il grafico copre quattro intervalli annuali: 2020-2021, 2021-2022, 2022-2023 e 2023-2024.
- **Stabilità iniziale con crescita:** Nel periodo 2020-2021, circa il 6.5% delle province ha cambiato cluster. L'anno successivo (2021-2022), questa percentuale è aumentata leggermente, attestandosi intorno al 7.7%.
- **Rallentamento nel 2022-2023:** Si osserva una diminuzione della percentuale di province che cambiano cluster nel periodo 2022-2023, scendendo a circa il 5.5%. Questo potrebbe indicare un periodo di maggiore stabilità nel

raggruppamento socio-economico delle province.

- **Forte accelerazione nel 2023-2024:** L'ultimo periodo analizzato, 2023-2024, mostra un **incremento significativo** nella percentuale di province che cambiano cluster, raggiungendo circa il 10.5%. Questo è il valore più alto dell'intero periodo e suggerisce che un numero considerevole di province ha subito uno spostamento di gruppo socio-economico.

In sintesi, i dati suggeriscono una certa dinamicità nel corso dei quattro anni, con un picco di cambiamenti nel raggruppamento socio-economico delle province italiane registrato nell'ultimo anno considerato (2023-2024). Questo potrebbe essere un indicatore di cambiamenti economici o sociali in atto che influenzano la distribuzione delle province nei cluster.

Questo grafico mostra l'evoluzione temporale a cui appartiene ciascuna provincia italiana, basata su un'analisi di dati socio-economici. Ogni riga rappresenta una provincia e la colonna colorata a destra identifica il "cluster" di appartenenza, ovvero un gruppo di province che condividono caratteristiche socio-economiche simili. Il colore blu occupa la maggior parte dello spazio, indicando che la maggior parte delle province appartiene a un cluster omogeneo e che per molte di esse, questa appartenenza non è cambiata significativamente nel periodo analizzato (dal 2020 al 2024 circa, come indicato sull'asse orizzontale).

Cosa si nota:

- **Predominanza di un Cluster Blu:** La maggior parte delle province è rappresentata in blu. Questo suggerisce che c'è un gruppo consistente di province con caratteristiche socio-economiche simili che rimangono stabili nel tempo. Questo potrebbe rappresentare il "gruppo di maggioranza" o il comportamento "normale" della maggior parte delle province italiane in base ai dati analizzati.
- **Piccoli Spostamenti di Colore:** In alcune righe (province), si notano delle barre colorate di colori diversi (marrone, turchese, viola, rosa, giallo, rosso, verde, arancione) che appaiono per specifici periodi di tempo. Questi cambiamenti indicano che alcune province hanno temporaneamente o permanentemente modificato il loro cluster di appartenenza, segnalando quindi una variazione nelle loro

caratteristiche socio-economiche rispetto al gruppo principale (blu).

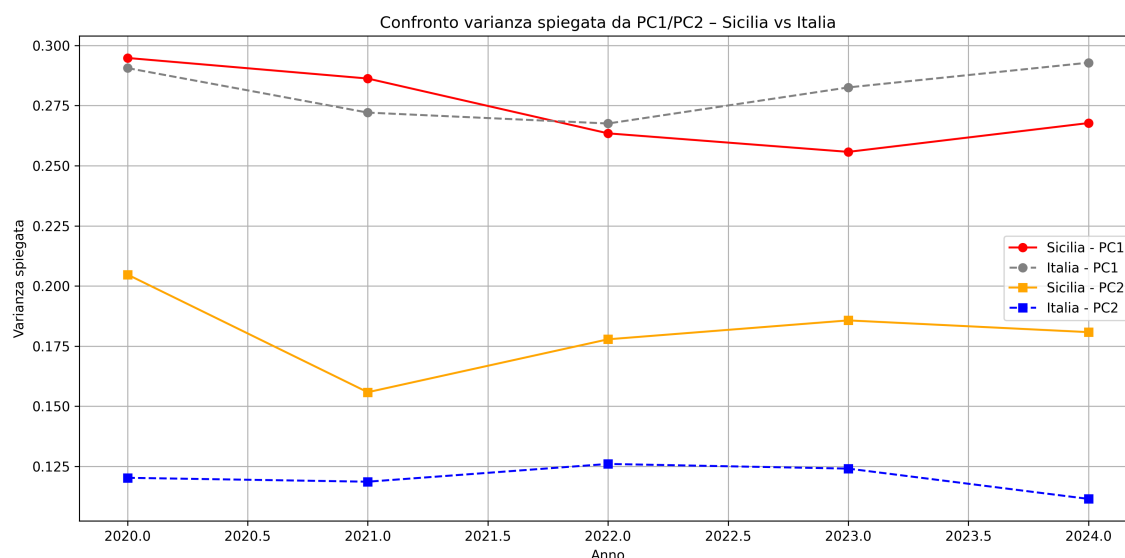
- **Province "Fuori Norma" (Outlier):** La barra colorata più in alto, etichettata "Outlier", e le altre barre colorate sulla destra, indicano province che sono state classificate in cluster diversi dal principale, o che presentano caratteristiche che le rendono uniche rispetto alla maggioranza. Ad esempio, si vedono province che passano dal blu a un altro colore per un certo periodo.
- **Durata dei Cambiamenti:** L'estensione orizzontale delle barre colorate indica la durata del periodo in cui una provincia ha appartenuto a un cluster diverso. Ad esempio, Grosseto (in turchese) e Livorno (in marrone) mostrano un periodo relativamente lungo di appartenenza a un cluster diverso. Altre province, come Genova, mostrano un periodo più breve.
- **Specifici Momenti di Cambiamento:** Si possono individuare momenti specifici in cui i cambiamenti avvengono. Ad esempio, intorno al 2023-2024, ci sono alcuni cambiamenti di cluster che emergono.
- **Mancanza di Cambiamenti Drastici per Molte Province:** La stragrande maggioranza delle province rimane nel cluster blu per tutto il periodo osservato, suggerendo una certa stabilità strutturale a livello socio-economico in molte aree del paese.

In sintesi, il grafico ci mostra che esiste un gruppo dominante di province con profili socio-economici omogenei. Tuttavia, ci sono anche delle eccezioni e delle dinamiche più fluide per un numero limitato di province, che evidenziano la

diversità del territorio italiano e la possibilità di variazioni nel tempo delle loro condizioni. Dobbiamo approfondire le

province che cambiano cluster per capire quali indicatori socio-economici abbiano guidato questi spostamenti.

confronto_varianza_sicilia_vs_italia.png



Il grafico confronta, su base annua, quanto le due componenti principali (PC1 e PC2) riescono a spiegare la varianza dei dati socio-economici, distinguendo tra Sicilia e l'Italia nel suo complesso. Le componenti principali sono in sostanza modi di riassumere le informazioni contenute in tanti indicatori diversi: PC1 cattura le tendenze più importanti e diffuse, mentre PC2 quelle secondarie ma comunque significative.

Ecco cosa notiamo:

- **PC1: La tendenza generale dei dati socio-economici in Sicilia è leggermente più variabile e complessa rispetto alla media nazionale.**
- La linea rossa rappresenta la Sicilia per la PC1, mentre quella grigia tratteggiata rappresenta l'Italia.
- All'inizio del periodo (2020), la Sicilia mostrava una varianza spiegata dalla PC1 leggermente superiore all'Italia, indicando che le sue dinamiche socio-economiche "principali" erano più marcate.
- Successivamente, sia la Sicilia che l'Italia hanno visto una diminuzione della varianza spiegata dalla PC1, suggerendo una stabilizzazione o un allineamento delle tendenze generali. Tuttavia, la linea rossa (Sicilia) tende a diminuire più rapidamente rispetto alla linea grigia (Italia) nella prima parte del periodo, per poi stabilizzarsi e risalire leggermente verso la fine, mantenendo comunque un livello di varianza spiegata dalla PC1 tendenzialmente inferiore alla media nazionale verso il 2023-2024.
- **PC2: La componente "secondaria" della varianza socio-economica**

mostra pattern distinti tra Sicilia e Italia.

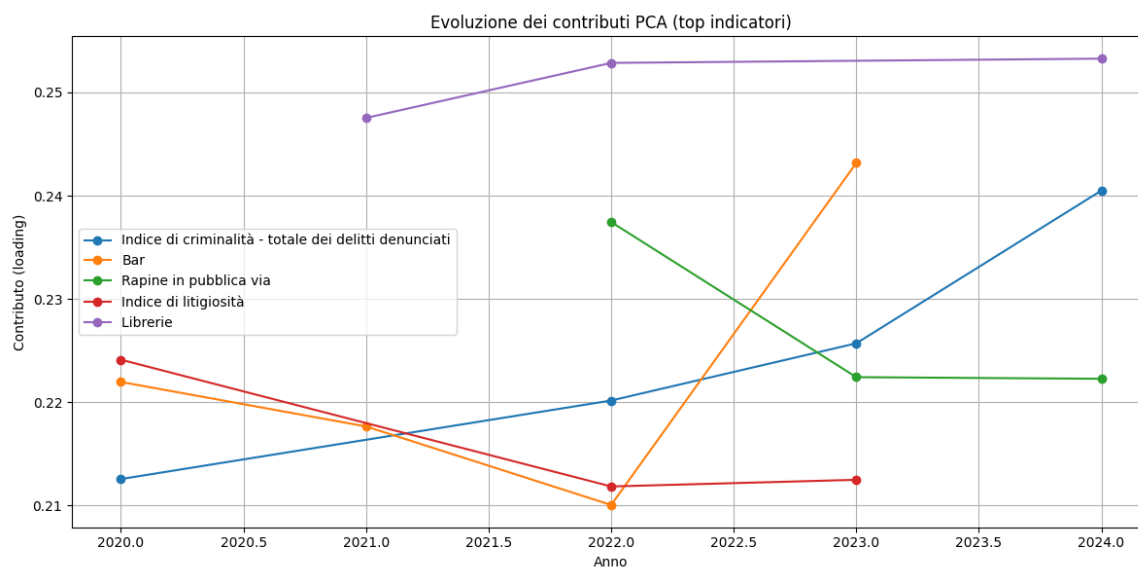
- La linea arancione rappresenta la Sicilia per la PC2, mentre quella blu tratteggiata rappresenta l'Italia.
- Inizialmente (2020), la Sicilia mostrava una varianza spiegata dalla PC2 significativamente più alta (circa 0.20) rispetto all'Italia (circa 0.12). Questo suggerisce che, in Sicilia, c'erano dinamiche socio-economiche "minori" ma comunque importanti da catturare, che non erano così presenti o marcate a livello nazionale.
- Tra il 2020 e il 2021, la varianza spiegata dalla PC2 in Sicilia è diminuita bruscamente, avvicinandosi a quella italiana.
- Dopo il 2021, la varianza spiegata dalla PC2 in Sicilia è aumentata gradualmente, superando di nuovo quella italiana verso la fine del periodo.
- La linea blu (Italia per la PC2) è rimasta relativamente stabile, con una

leggera oscillazione ma senza variazioni sostanziali, mantenendosi sempre su livelli più bassi rispetto alla Sicilia nella maggior parte del periodo.

In sintesi:

Il grafico ci dice che le dinamiche socio-economiche più ampie e trasversali (PC1) sono state inizialmente più accentuate in Sicilia rispetto alla media nazionale, per poi allinearsi e diminuire. Per quanto riguarda le dinamiche più specifiche e "di dettaglio" (PC2), la Sicilia ha mostrato una maggiore complessità e un maggiore peso di queste componenti minori all'inizio del periodo, che si è poi attenuata per poi risalire, evidenziando una certa peculiarità regionale rispetto all'andamento più omogeneo a livello italiano. La tendenza generale è una convergenza verso dinamiche più simili tra Sicilia e Italia per la PC1, mentre per la PC2 la Sicilia sembra avere pattern più dinamici e distinti.

contributi_indicatori_temporali



Questo grafico mostra l'evoluzione temporale di cinque indicatori socio-economici considerati "top" in un'analisi su dati provinciali italiani. Ogni linea rappresenta un indicatore, e l'altezza di ogni punto sulla linea indica quanto quel particolare indicatore "pesa" o contribuisce in un certo anno. In altre parole, una linea più alta significa che quell'indicatore è più rilevante in quell'anno rispetto ad altri in quell'analisi.

Ecco cosa si nota:

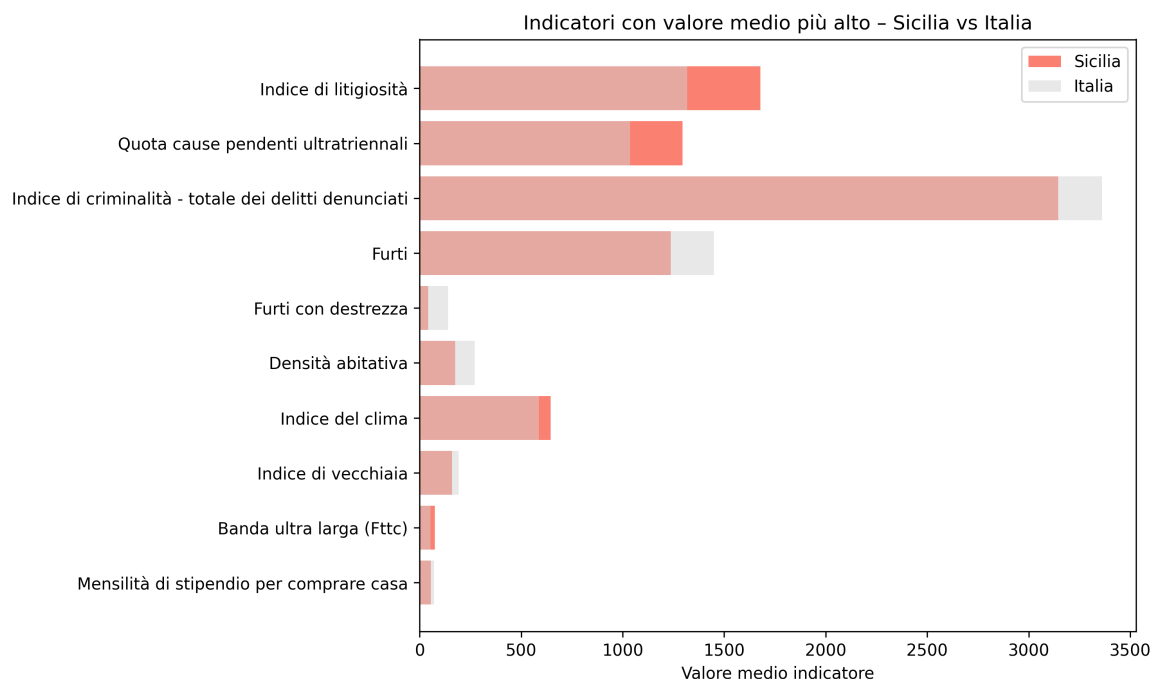
- **Indicatori stabili o in leggera crescita:** Le "Librerie" (linea viola) mostrano un trend nettamente in crescita fino al 2021, per poi stabilizzarsi a un livello elevato. Anche l'"Indice di criminalità - totale dei delitti denunciati" (linea blu) mostra una tendenza generale all'aumento, seppur più contenuta.
- **Indicatori in calo o altalenanti:** L'"Indice di litigiosità" (linea rossa) mostra un trend decrescente costante fino al 2022, per poi stabilizzarsi a un livello inferiore. Le "Rapine in pubblica via" (linea verde) hanno un

andamento molto volatile: crescono tantissimo nel 2022 per poi crollare e stabilizzarsi su valori bassi. I "Bar" (linea arancione) hanno un andamento simile alle rapine, con un picco nel 2023 e poi una diminuzione.

- **Divergenze significative:** Nel 2022 e 2023 si notano forti divergenze. Mentre alcuni indicatori (come le librerie e l'indice di criminalità generale) continuano a mantenere o aumentare la loro rilevanza, altri (come le rapine) subiscono un calo drastico, e i bar hanno un picco improvviso per poi ridiscendere.

In sintesi, questo grafico evidenzia come la rilevanza di certi fattori socio-economici nel contesto delle province italiane cambi nel tempo. Alcuni sembrano indicare un peggioramento (criminalità generale in crescita), altri una diminuzione di criticità (litigiosità), mentre altri ancora mostrano comportamenti imprevedibili e potenzialmente legati a eventi specifici (rapine e bar nel 2022-2023). La stabilità delle librerie a livelli alti suggerisce un loro ruolo costante e importante nell'analisi.

indicatori_valori_medi_sicilia_vs_italia



Questo grafico confronta la media di alcuni indicatori socio-economici tra la Sicilia e il resto dell'Italia. Ogni riga rappresenta un indicatore specifico, e per ciascuno sono mostrate due barre: una di colore rosato chiaro che indica il valore medio in Sicilia e una di colore grigio che rappresenta il valore medio in Italia. L'asse orizzontale indica il "Valore medio indicatore", dando un'idea dell'entità del valore.

Cosa notiamo:

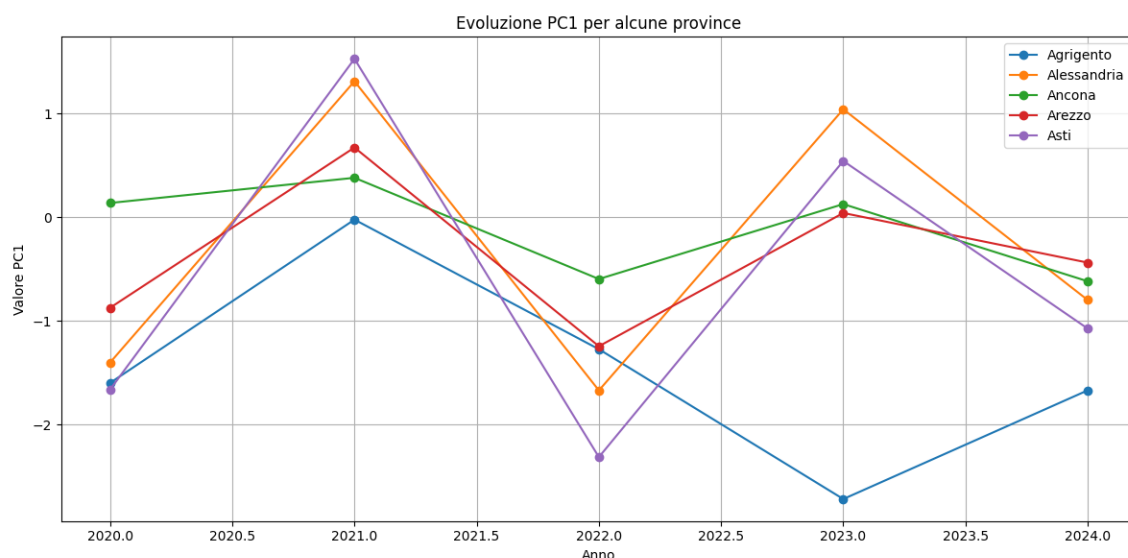
- **Criminalità e Litigiosità:** La Sicilia mostra valori medi significativamente più alti rispetto alla media italiana per quanto riguarda l'"Indice di criminalità - totale dei delitti denunciati" e l'"Indice di litigiosità". Le barre siciliane per questi indicatori sono molto più lunghe di quelle italiane, indicando una maggiore incidenza.

- **Furti:** Anche per i "Furti" e i "Furti con destrezza", la Sicilia presenta valori medi superiori alla media nazionale, anche se il divario sembra meno marcato rispetto ai delitti denunciati in generale.
- **Queste cause pendenti:** La "Quota cause pendenti ultratriennali" è più elevata in Sicilia rispetto alla media italiana.
- **Clima:** L'"Indice del clima" è più alto in Sicilia. La barra siciliana è composta da una parte rosata preponderante e una piccola parte arancione, suggerendo una componente specifica o una particolare misurazione all'interno di questo indice.
- **Indicatori di Povertà e Accesso a Servizi:**
 - La "Densità abitativa" è più bassa in Sicilia rispetto alla media italiana.
 - L'"Indice di vecchiaia" è più basso in Sicilia.

- La "Banda ultra larga (Fttc)" è meno diffusa in Sicilia rispetto alla media nazionale, con una barra siciliana molto corta rispetto a quella italiana.
- La "Mensilità di stipendio per comprare casa" indica che in Sicilia serve un numero maggiore di mensilità di stipendio per comprare casa rispetto alla media italiana.

In sintesi, il grafico evidenzia delle criticità in Sicilia in termini di criminalità, litigiosità, e accesso a determinate infrastrutture digitali, oltre a una minore facilità nell'accedere all'abitazione. Su altri aspetti, come il clima e l'indice di vecchiaia, si osservano differenze meno nette o invertite rispetto ad altri indicatori.

pc1_province_temporali



Questo grafico mostra l'andamento di un indice socio-economico, chiamato "Valore PC1", nel tempo (anni dal 2020 al 2024) per cinque diverse province italiane: Agrigento, Alessandria, Ancona, Arezzo e Asti.

Il "Valore PC1" è un indice riassuntivo che cattura diverse variabili socio-economiche. Valori più alti indicano una situazione socio-economica generalmente migliore o più favorevole, mentre valori più bassi suggeriscono il contrario.

Cosa osserviamo:

- **Andamenti diversi nel tempo:** Le province mostrano traiettorie molto differenti.
- **Agrigento (blu):** Inizia con un valore basso nel 2020, sale leggermente nel 2021, crolla nel 2022, per poi risalire nel 2023 e scendere di nuovo nel 2024. Mostra una notevole volatilità.
- **Alessandria (arancione):** Parte da un valore medio-basso, ha un picco forte nel 2021, poi scende significativamente nel 2022, risale nel 2023 e scende ancora nel 2024. Anch'essa è abbastanza variabile.
- **Ancona (verde):** Presenta un andamento più stabile e

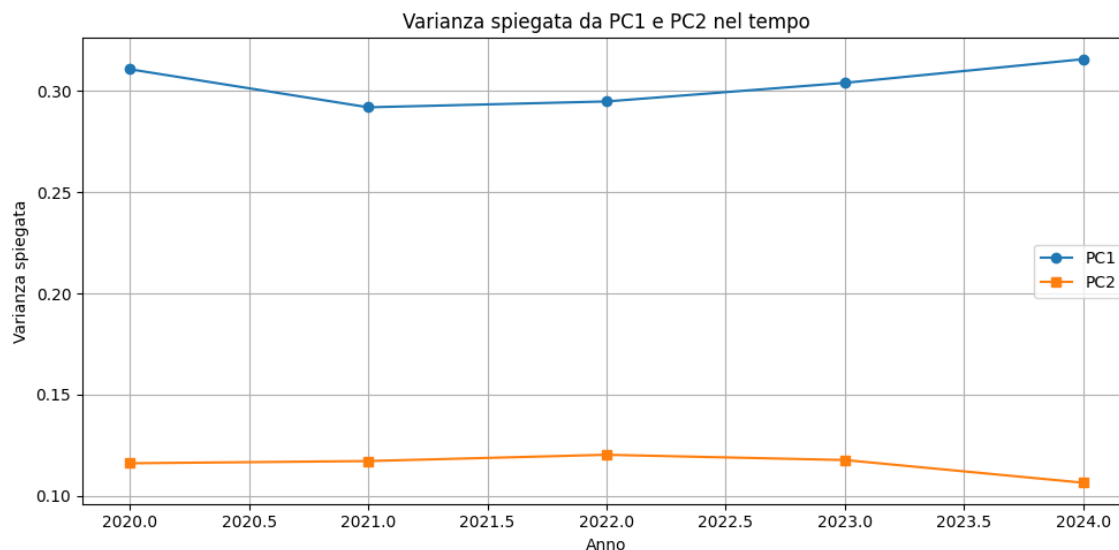
tendenzialmente in crescita nella prima parte del periodo, con un calo nel 2022 e poi una ripresa più contenuta nel 2023 e 2024.

- **Arezzo (rossa):** Inizia con un valore relativamente basso, cresce fino al 2021, poi cala nel 2022, risale nel 2023 e diminuisce nel 2024. Ha un andamento simile ad Alessandria ma con valori generalmente inferiori.
- **Asti (viola):** Inizia con il valore più basso nel 2020, mostra un forte aumento nel 2021, un drastico calo nel 2022, e poi una risalita nel 2023 e 2024, raggiungendo valori simili ad altre province.
- **Confronto tra province:**
- Nel 2020, Asti e Agrigento partono con i valori più bassi, mentre Ancona ha il valore più alto tra le province che iniziano in positivo.
- Nel 2021, Alessandria e Asti raggiungono i picchi più alti, mentre Agrigento sale ma rimane sotto zero. Ancona e Arezzo si attestano su valori medi.

- Il 2022 sembra essere un anno di forte contrazione per quasi tutte le province, in particolare per Asti, Alessandria e Agrigento. Ancona e Arezzo calano ma in modo meno drastico.
- Dal 2023 in poi, si osserva una ripresa generale, ma con pattern che divergono ancora. Agrigento e Asti mostrano una risalita decisa, mentre Alessandria e Arezzo mostrano un calo dopo il picco del 2023, e Ancona un andamento più contenuto.

In sintesi, il grafico ci mostra che il quadro socio-economico, rappresentato dal Valore PC1, non è omogeneo tra queste province italiane e che ci sono state dinamiche distinte tra il 2020 e il 2024, con alcuni anni (come il 2021 e il 2022) che hanno segnato picchi e cali significativi e differenziati. È interessante notare come alcune province abbiano recuperato terreno nel 2023/2024 dopo i cali del 2022, mentre altre mostrano un nuovo trend discendente verso la fine del periodo analizzato.

varianza_spiegata_temporale



Questo grafico mostra l'andamento nel tempo della **capacità di spiegazione** di due dimensioni principali (chiamate PC1 e PC2) nell'analisi dei dati socio-economici delle province italiane. In parole semplici, ci dice quanto queste due dimensioni riescono a "catturare" la variabilità generale dei dati che stiamo osservando per ogni provincia, e come questa capacità cambia di anno in anno.

Ecco cosa possiamo osservare:

- PC1 domina la scena:** La linea blu, che rappresenta PC1, è costantemente molto più alta della linea arancione (PC2). Questo significa che la prima dimensione (PC1) da sola riesce a spiegare una **frazione significativamente maggiore** della variabilità dei dati socio-economici rispetto alla seconda dimensione (PC2). Immaginiamo che PC1 sia un fattore molto importante e generale che influenza molti aspetti economici e sociali delle province.
- PC1 è sostanzialmente stabile, con un leggero aumento:** La linea blu non scende mai sotto lo 0.29 e arriva quasi a 0.32 nel 2024. Questo indica che la capacità esplicativa di PC1 è **molto consistente nel tempo**. C'è una leggera tendenza all'aumento negli ultimi anni, suggerendo che questo fattore generale potrebbe essere diventato ancora più rilevante o che i dati evidenziano ulteriormente la sua importanza.
- PC2 ha un ruolo secondario e stabile:** La linea arancione (PC2) si mantiene sempre intorno a 0.11 - 0.12. Questo conferma che PC2 cattura una parte molto minore della variabilità totale. Il suo andamento è **relativamente piatto**, con una leggera flessione alla fine del periodo di osservazione. Questo suggerisce che questa seconda dimensione, qualunque cosa rappresenti nel dettaglio (forse un fattore più specifico o settoriale), non ha subito grandi cambiamenti nella sua capacità di

spiegazione del fenomeno analizzato in questo lasso di tempo.

- **Focus su PC1:** Dato che PC1 spiega una quota così elevata della variabilità, per capire le dinamiche socio-economiche delle province italiane, è fondamentale concentrarsi principalmente su cosa rappresenta

questa prima componente. La sua stabilità e leggera crescita nel tempo indicano un fattore di fondo che continua a essere preponderante e forse si sta rafforzando. PC2, invece, potrebbe aggiungere dettagli interessanti, ma la sua influenza sul quadro generale è limitata.

scatter_sicilia_2020



L'immagine rappresenta una proiezione su due dimensioni (PC1 e PC2) di province italiane, basata su un'analisi di dati socio-economici del 2020. Ogni punto sull'immagine corrisponde a una provincia, e la sua posizione è determinata dalla combinazione di vari indicatori socio-economici. Le province della Sicilia sono evidenziate in rosso, mentre le altre sono in grigio.

Cosa osserviamo:

- **Raggruppamenti Geografici e Socio-Economici:** Le province che si trovano vicine in questo grafico tendono ad avere caratteristiche socio-economiche simili. Per esempio, notiamo che molte province del nord-ovest (come Torino, Milano, Bologna) si posizionano in un'area del grafico con valori di PC1 generalmente più alti. Al contrario, province del sud e delle isole, in particolare quelle siciliane evidenziate in rosso, tendono

a posizionarsi in aree con valori di PC1 più bassi.

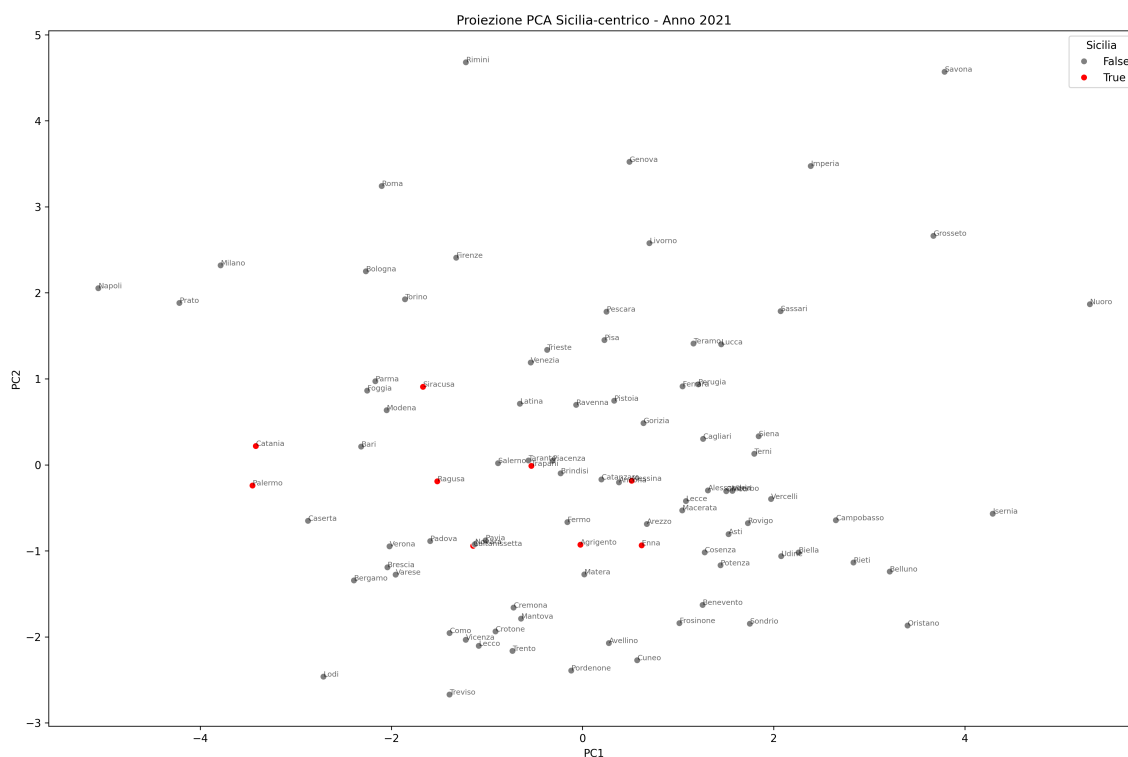
- **Distanziamento delle Province Siciliane:** Le province siciliane evidenziate in rosso (Agrigento, Caltanissetta, Catania, Enna, Messina, Palermo, Ragusa, Siracusa) sembrano formare un gruppo distinto. Alcune sono isolate, altre vicine tra loro, ma nel complesso si posizionano in un'area diversa rispetto a molte province del centro-nord. Questo suggerisce che queste province condividono caratteristiche socio-economiche che le differenziano significativamente da altre aree del paese, almeno per quanto riguarda gli indicatori utilizzati nell'analisi.
- **Eccezioni e Outlier:** Ci sono alcune eccezioni interessanti. Ad esempio, Lucca, una provincia della Toscana (non siciliana), è evidenziata in rosso, indicando che condivide delle caratteristiche socio-economiche con le province siciliane, almeno secondo questa specifica analisi. Allo stesso modo, sebbene la maggior parte delle province siciliane siano in rosso, se analizziamo più attentamente, alcune province del sud come Taranto, Lecce,

Potenza, Brindisi, Catanzaro, e perfino alcune del nord come Pavia e Foggia, sono colorate in grigio, mostrando che non tutte le province meridionali sono "vicine" tra loro in questo spazio di caratteristiche.

- **Indicatori Latenti (PC1 e PC2):** PC1 e PC2 sono dimensioni che riassumono le principali variazioni nei dati socio-economici. Sebbene non sia specificato cosa rappresentino esattamente, possiamo inferire che il PC1 sembra distinguere in modo netto tra province del nord/centro e quelle del sud/isole. Il PC2 potrebbe catturare altre dimensioni di variabilità, come ad esempio differenze tra aree urbane più sviluppate e aree più rurali o con un diverso tessuto produttivo.

In sintesi, il grafico ci aiuta a visualizzare come le province italiane si differenziano in termini di profili socio-economici. Le province siciliane, con alcune eccezioni e analogie con altre aree, tendono a collocarsi in un segmento particolare di questo spazio di caratteristiche, evidenziando potenziali aree di specificità territoriale che potrebbero richiedere politiche o interventi mirati.

scatter_sicilia_2021



Questa immagine rappresenta una **mappa delle province italiane basata sulle loro caratteristiche socio-economiche, semplificata in due dimensioni**. Immagina di avere un sacco di informazioni su ogni provincia: PIL, reddito, tasso di occupazione, livello di istruzione, ecc. Questa mappa cerca di riassumere tutte queste informazioni in un grafico dove la posizione di ogni puntino (che rappresenta una provincia) ti dice qualcosa sulla sua somiglianza con altre province in termini di questi fattori.

In parole semplici, province che si trovano vicine tra loro in questo grafico tendono ad avere profili socio-economici simili, mentre quelle lontane sono più diverse.

Cosa vediamo:

- **Puntini e nomi:** Ogni puntino è una provincia italiana, con il suo nome associato.
- **Assi (PC1 e PC2):** Gli assi non sono le classiche coordinate geografiche (nord, sud, est, ovest). PC1 (Asse 1) e PC2 (Asse 2) sono delle **combinazioni sintetiche di tutte le variabili socio-economiche che abbiamo analizzato**. L'asse PC1 tende a catturare la maggior parte della variabilità, e il PC2 quella rimanente. Quindi, il valore sull'asse PC1 ci dice quanto una provincia si posiziona su uno spettro socio-economico principale, e il PC2 su uno secondario.
- **Colori:** Qui sta il succo dell'analisi specifica di questa mappa:
- I puntini **grigi** rappresentano le province considerate "False" rispetto a un certo criterio (probabilmente legato alla Sicilia, dato il titolo).

- I puntini **rossi** rappresentano le province considerate "True". In questo caso, vediamo chiaramente che le province rosse sono **Catania, Siracusa, Palermo, Ragusa, Enna e Agrigento**. Sono tutte province siciliane. Questo indica che l'analisi ha cercato di isolare delle caratteristiche che distinguono le province siciliane da quelle non siciliane, o che all'interno del campione analizzato, queste province siciliane presentano un profilo omogeneo o particolare.

Pattern ed Evidenze:

1. Raggruppamento delle province siciliane: Le province siciliane evidenziate in rosso sono concentrate in una specifica area del grafico, soprattutto nella parte centrale e inferiore sinistra. Questo suggerisce che queste province, nel complesso dei dati analizzati nel 2021, hanno un profilo socio-economico che le accomuna e le distingue dalle altre province grigie.

2. Separazione dei gruppi: Si nota una certa separazione tra il gruppo delle province siciliane (rosse) e la maggior parte delle province grigie. Questo potrebbe indicare che le province siciliane, sulla base dei dati utilizzati, presentano delle peculiarità socio-economiche che le differenziano.

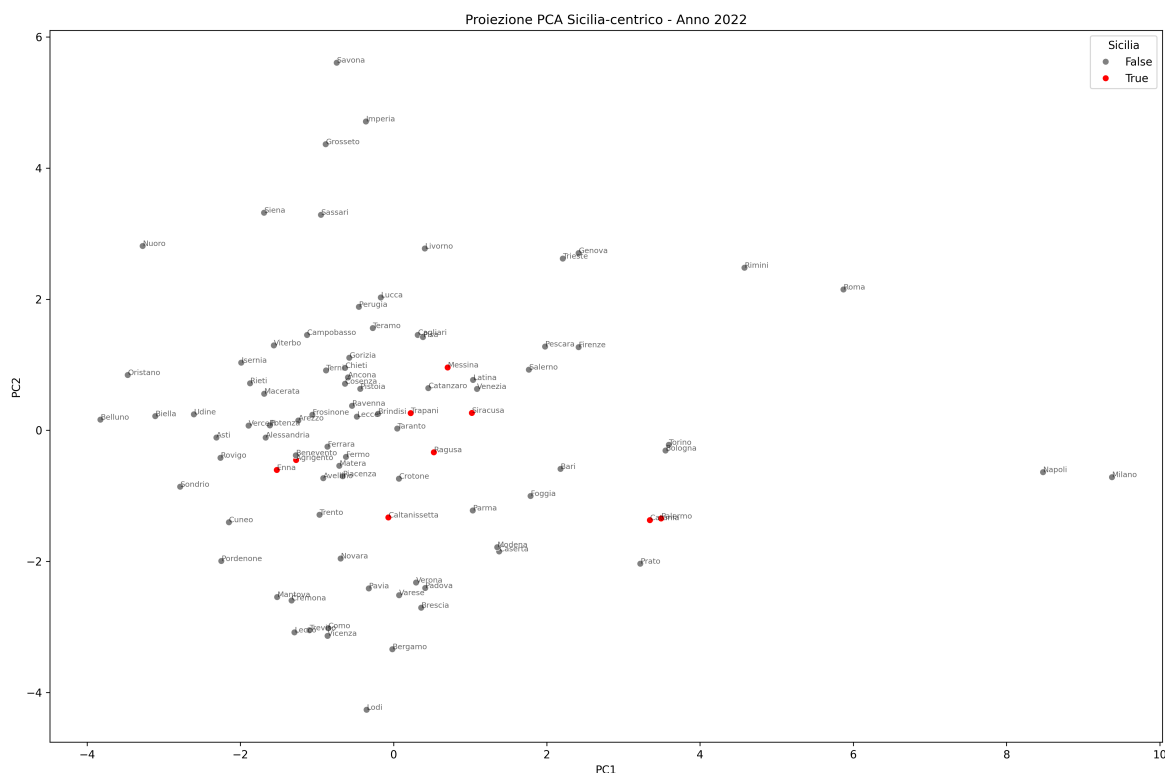
3. Posizionamento delle altre province:

Le province grigie sono più sparse sul grafico. Questo è normale, perché l'Italia è ricca di diversità socio-economiche tra le varie province. Alcune province grigie sono vicine tra loro (ad esempio, quelle del Nord-Est come Trieste, Venezia, Vicenza, Trento, Treviso, e poi quelle del Centro come Roma, Firenze, Bologna), suggerendo somiglianze tra loro.

4. Unica provincia "True" fuori dal cluster principale: Notiamo anche che "Agrigento" è marcata in rosso, posizionandosi leggermente più a sinistra e in basso rispetto al blocco più compatto di Siracusa, Catania, Enna. Questo potrebbe indicare una leggera sfumatura nel profilo socio-economico di Agrigento rispetto alle altre province siciliane identificate come "True", pur rimanendo all'interno della stessa area generale del grafico.

In sintesi, questo grafico è uno strumento visivo utile per capire come le province italiane si differenziano e si raggruppano sulla base di indicatori socio-economici. L'evidenza più forte qui è che le province siciliane selezionate mostrano una certa omogeneità e tendono a distinguersi dal resto del campione analizzato.

scatter_sicilia_2022



Questa immagine mostra una rappresentazione grafica delle province italiane basata su dati socio-economici per l'anno 2022. Le province sono disposte su un grafico a due dimensioni dove gli assi (chiamati PC1 e PC2) rappresentano delle combinazioni di queste variabili socio-economiche che catturano la maggior parte della loro variabilità.

In parole semplici, le province che si trovano vicine tra loro in questo grafico condividono caratteristiche socio-economiche simili, mentre quelle più distanti sono più differenti.

Cosa si nota:

- **Le province siciliane in rosso:** Le province evidenziate in rosso (Palermo, Trapani, Siracusa, Ragusa, Enna, Caltanissetta, Messina) sono

tutte in Sicilia. Questo indica che, secondo le variabili analizzate, queste province hanno un profilo socio-economico distinto rispetto alla maggior parte delle altre province italiane, che sono rappresentate in grigio.

- **Concentrazione delle province non siciliane:** La maggior parte delle province grigie (cioè quelle non siciliane) tende a raggrupparsi in diverse aree del grafico.
- Un gruppo di province si trova nella parte alta e centrale-sinistra del grafico, con valori di PC1 che vanno da circa -2 a +2 e valori di PC2 che variano da circa 0 a +4. Queste province potrebbero condividere caratteristiche comuni legate, ad esempio, a livelli di reddito, occupazione o industrializzazione.

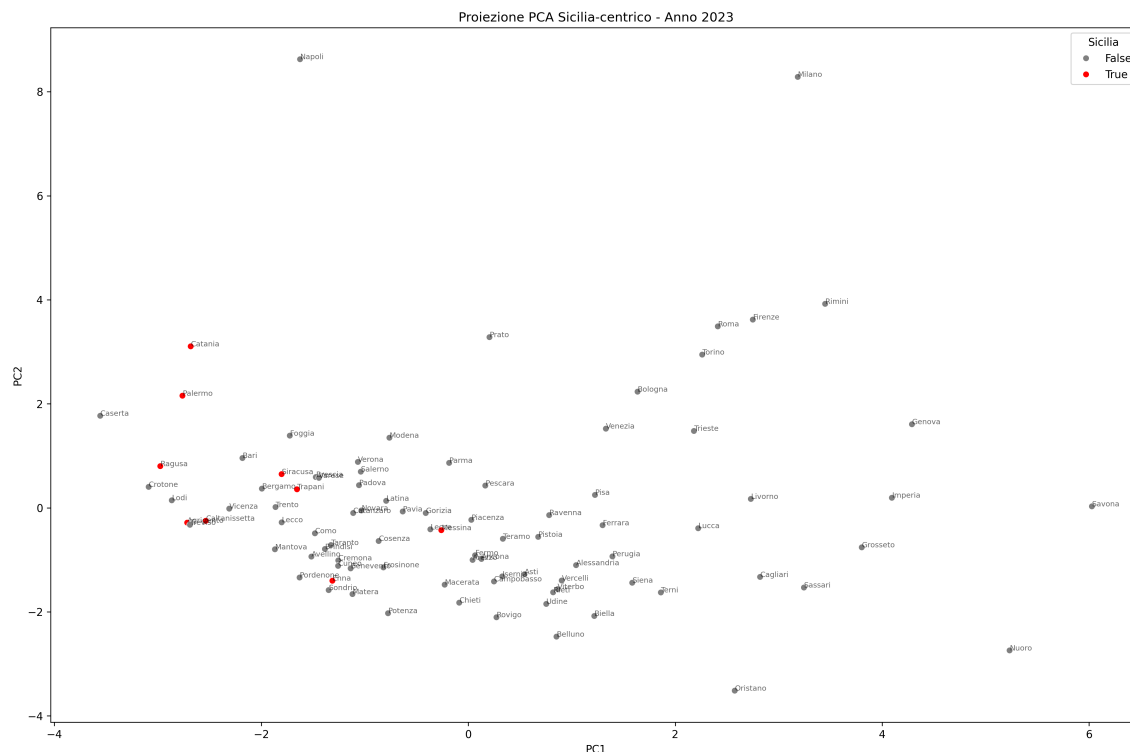
- Un altro gruppo si estende più verso destra, con valori di PC1 più elevati (intorno a 0-8) e valori di PC2 più bassi (tra -2 e +2).
- Ci sono anche alcune province più isolate, sia nel gruppo delle province non siciliane che in quello siciliano, suggerendo profili unici.
- **Posizionamento delle province siciliane:** Le province siciliane in rosso non sono tutte raggruppate in un unico punto, ma si trovano in posizioni diverse tra loro, pur essendo tutte distinte dal gruppo principale delle province grigie. Ad esempio, Palermo e Trapani sono vicine tra loro con un PC1 più alto e un PC2 negativo, mentre Siracusa e Ragusa sono più a est con PC1 positivi. Messina è situata a una posizione intermedia. Questo

potrebbe indicare che anche all'interno della Sicilia esistono differenze socio-economiche, ma nel complesso la regione presenta un'identità socio-economica specifica rispetto al resto del paese.

In sintesi:

Il grafico ci mostra che la Sicilia, come regione, tende a distinguersi dal resto delle province italiane per quanto riguarda i suoi indicatori socio-economici nel 2022. Le province siciliane, pur mostrando alcune differenze interne, sono posizionate in modo diverso rispetto alla maggior parte delle altre province, che invece tendono a formare dei cluster, suggerendo una certa omogeneità socio-economica all'interno di questi gruppi.

scatter_sicilia_2023



Questo grafico rappresenta la posizione di diverse province italiane in uno spazio bidimensionale basato su un'analisi dei loro dati socio-economici. Le due dimensioni (asse orizzontale PC1 e asse verticale PC2) sono dei "riassunti" di molteplici indicatori socio-economici, come reddito, occupazione, istruzione, demografia, ecc.

In termini semplici, province che si trovano vicine in questo grafico condividono caratteristiche socio-economiche simili, mentre quelle lontane sono più dissimili. L'analisi è "Sicilia-centrica", il che suggerisce che la posizione della Sicilia e delle sue province è stata presa come punto di riferimento principale per organizzare lo spazio.

Evidenze e pattern principali:

- Le province siciliane rosse (True):**
 Le province di Catania, Palermo, Ragusa e Siracusa sono evidenziate in rosso. Questo indica che, secondo questa specifica analisi, sono considerate più simili tra loro in termini di caratteristiche socio-economiche rispetto a molte altre province italiane. La vicinanza di questi punti rossi conferma questa coerenza interna. C'è anche una provincia rossa, Agrigento, che è leggermente più isolata ma comunque vicina al gruppo delle altre province siciliane.
- Le province siciliane grigie (False):**
 Le province siciliane non evidenziate in rosso (come Trapani e Caltanissetta) sono rappresentate in grigio, indicando che le loro

caratteristiche socio-economiche le rendono più diverse dalle altre province siciliane rosse, o le avvicinano maggiormente ad altre aree geografiche italiane. Trapani, ad esempio, è vicina a province del sud-ovest della penisola.

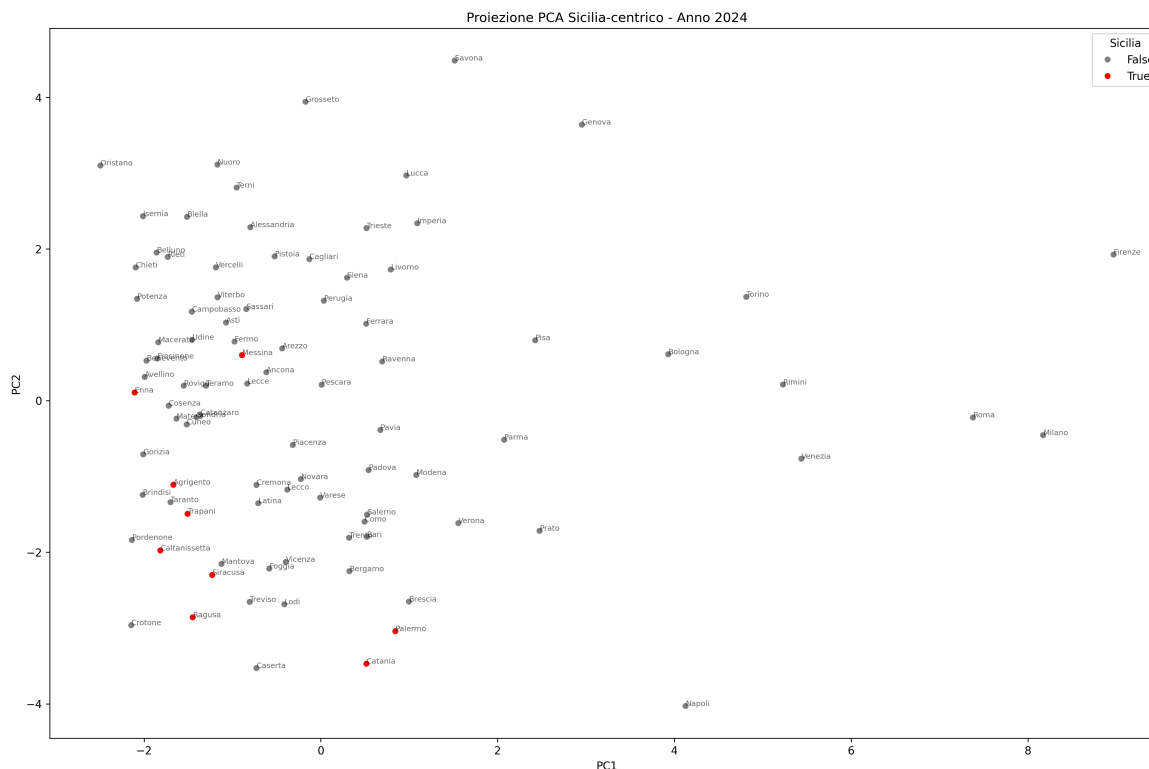
- **Le province del nord:** Si nota una concentrazione di province del nord Italia (come Milano, Torino, Bologna, Venezia, Trieste, Genova) che tendono ad occupare la parte destra e superiore del grafico. Questo indica che, nel complesso, queste province presentano pattern socio-economici che le distinguono dalle province del sud.
- **Le province del centro-sud:** Le province del centro-sud Italia (escluse quelle siciliane rosse) tendono a posizionarsi nella parte centrale e sinistra del grafico. Si osserva una certa aggregazione, ad esempio, delle

province campane (Napoli, Caserta) e pugliesi (Bari, Foggia).

- **Le province "anomale" o di confine:** Alcune province appaiono più isolate o in posizioni intermedie, suggerendo che le loro caratteristiche socio-economiche non le allineano perfettamente con i grandi raggruppamenti geografici o regionali (ad esempio, Grosseto, Livorno, alcune province del nord-est come Gorizia e Udine).

In sintesi, il grafico ci mostra come le province italiane si differenzino sulla base di indicatori socio-economici. Evidenzia una certa omogeneità tra alcune province siciliane (quelle rosse), le differenzia dal resto delle province dell'isola, e conferma una generale separazione tra le aree settentrionali e quelle centro-meridionali del paese, pur con le dovute sfumature e province intermedie.

scatter_sicilia_2024



Questa immagine mostra le province italiane posizionate su un grafico in base ai loro dati socio-economici. Le posizioni sono state ricavate utilizzando una tecnica che riduce la complessità di molti dati in soli due "fattori" principali, chiamati PC1 e PC2. In parole povere, il grafico ci dice che le province con posizioni simili sul grafico tendono ad avere caratteristiche socio-economiche simili.

Le province sono rappresentate da punti, con il nome della provincia scritto accanto. La legenda in alto a destra ci dice che i punti grigi rappresentano le province che non sono in Sicilia, mentre i punti rossi rappresentano le province siciliane.

Cosa notiamo:

- **Le province siciliane (rosse) tendono a raggrupparsi in basso a**

sinistra del grafico. Questo significa che queste province condividono caratteristiche socio-economiche simili, che sono distinte da quelle di molte altre province italiane. Ad esempio, Enna, Caltanissetta, Trapani, Agrigento, Ragusa, Siracusa, Catania e Palermo sono tutte posizionate nella parte inferiore sinistra.

- **Esiste una separazione generale tra le province del Nord e quelle del Sud.** Guardando la disposizione, sembra che le province del Nord (come Milano, Torino, Bologna, Venezia) tendano a posizionarsi nella parte destra del grafico (valori alti di PC1), mentre molte province del Sud si trovano più a sinistra o in basso. La Sicilia, come detto, è un esempio di questo raggruppamento meridionale.

- **Il PC1 (asse orizzontale) sembra catturare una differenza significativa tra province.** Le province sulla destra hanno valori alti di PC1, mentre quelle sulla sinistra hanno valori bassi. Questo potrebbe rappresentare, ad esempio, un divario tra aree economicamente più sviluppate e meno sviluppate, o differenze in altri indicatori socio-economici che sono stati analizzati.
- **Il PC2 (asse verticale) sembra catturare un'altra dimensione di differenziazione.** Le province in alto hanno valori alti di PC2, mentre quelle in basso hanno valori bassi. Questo potrebbe riflettere altri tipi di caratteristiche che distinguono le province, magari legate a urbanizzazione, demografia o altro.
- **Alcune province siciliane sono posizionate in modo leggermente diverso rispetto ad altre, ma rimangono comunque nella zona "sud-occidentale" del grafico.** Ad esempio, Messina e Enna sono leggermente più a destra rispetto ad altre siciliane, ma comunque lontane dalle province settentrionali.

In sintesi, questo grafico ci mostra che le province italiane, basandosi sui loro dati socio-economici, non sono distribuite casualmente. Ci sono pattern chiari: le province siciliane mostrano una certa omogeneità e si distinguono dalle altre, e si intravede una possibile dicotomia tra Nord e Sud Italia in termini di caratteristiche socio-economiche complessive.

4. Conclusioni e Prospettive per lo Sviluppo Territoriale Siciliano

Il presente report ha fornito un'analisi dettagliata delle dinamiche socio-economiche e ambientali che hanno caratterizzato le province italiane, nel quadriennio 2020-2024 (con particolare enfasi nel biennio 2023-2024), avvalendosi di un approccio strutturato di analisi statistica automatizzata e generazione guidata di report. L'obiettivo era quello di sintetizzare in modo chiaro e accessibile l'informazione complessa contenuta in dataset multivariati, evidenziando pattern, correlazioni e l'evoluzione temporale dei fenomeni osservati.

5. Sintesi delle Dinamiche Socio-Economiche Emerse

L'analisi ha rivelato un panorama territoriale articolato, dove indicatori chiave si influenzano reciprocamente in relazioni sia dirette che complesse. I principali filoni emersi riguardano l'interdipendenza tra mercato del lavoro e benessere sociale, il crescente impatto della digitalizzazione, le sfide legate alla sicurezza e alla coesione sociale, e le complesse dinamiche demografiche, il tutto in un contesto di disuguaglianze persistenti.

1. Lavoro e Benessere Sociale: Un mercato del lavoro dinamico con elevati tassi di occupazione si conferma come il pilastro del benessere territoriale. È stata osservata una forte e robusta correlazione tra l'occupazione, la spesa delle famiglie e la riduzione del fenomeno NEET (giovani che non studiano, non lavorano e non sono in formazione). Nel 2024, questo legame si è ulteriormente rafforzato, suggerendo un aumento della fiducia e della capacità di spesa nelle aree con maggiore occupazione.

2. Digitalizzazione e Sviluppo Economico:

La diffusione di infrastrutture digitali avanzate, in particolare la banda larga e ultralarga, mostra una correlazione crescente con indicatori di modernizzazione economica, come i depositi bancari. Questa tendenza, rafforzatasi nel biennio, sottolinea il ruolo cruciale della connettività per favorire la gestione finanziaria e l'innovazione a livello locale.

3. Sicurezza e Coesione Sociale:

Gli indici di criminalità (furti, rapine) e di litigiosità continuano a essere correlati con una generale percezione di insicurezza e problemi nella risoluzione delle controversie legali (cause pendenti ultratriennali). Nel 2024, è emerso un rafforzamento di questa correlazione con la vitalità economica delle aree urbane, suggerendo un impatto più diretto della sicurezza sul tessuto commerciale. Inoltre, l'analisi ha evidenziato come l'insicurezza e la litigiosità si leghino a un più alto tasso di mancata partecipazione al lavoro, e, nel 2024, all'incremento della solitudine e a un basso tasso di fecondità, delineando un quadro di progressiva fragilità sociale.

4. Demografia e Servizi Territoriali:

Il concetto di "città vivibile", caratterizzato da una ricca offerta di servizi (ristoranti, librerie, medici), coesiste con sfide demografiche significative, quali un basso tasso di natalità e un elevato indice di dipendenza degli anziani. Particolarmente critiche sono le correlazioni emerse nel 2023-2024, che associano una minore

natalità a una maggiore solitudine, indicando un potenziale indebolimento dei legami sociali e un impatto sulle dinamiche riproduttive.

5. Crescita Economica e Disuguaglianze: Nonostante l'evidenza di una crescita economica in termini di retribuzione media annua e valore aggiunto per abitante, l'analisi ha rivelato che i benefici non sono distribuiti equamente. La correlazione tra indicatori economici positivi e la presenza di famiglie con ISEE basso, unitamente all'aumento dell'indice di fragilità urbana, suggerisce un mercato del lavoro polarizzato e l'esistenza di sacche di vulnerabilità strutturale anche in contesti di apparente prosperità.

6. Riflessioni sull'Evoluzione Temporale e Contestuale

L'analisi sull'evoluzione temporale (2020-2024) ha messo in luce aspetti cruciali della dinamica territoriale italiana e, in particolare, siciliana.

Le analisi di riduzione della dimensionalità (PCA) hanno

costantemente dimostrato che la stragrande maggioranza della variabilità socio-economica delle province italiane è spiegata da un numero ristretto di componenti principali (circa il 30-32% per la PC1 e il 10-13% per la PC2). Questo indica che le differenze territoriali sono guidate da pochi fattori strutturali di fondo, spesso legati al livello generale di sviluppo e a dinamiche specifiche settoriali o demografiche.

Le province siciliane hanno mostrato un profilo socio-economico distintivo, aggregandosi costantemente in cluster separati o in aree specifiche dei grafici PCA in tutti gli anni analizzati (2020-2024). Questo evidenzia una specificità regionale che le differenzia dalla media nazionale, rendendo necessaria un'analisi e una pianificazione mirate. In particolare, il confronto tra la varianza spiegata per la Sicilia e quella per l'Italia ha indicato una maggiore complessità delle dinamiche "secondarie" (PC2) in Sicilia, suggerendo che la regione è influenzata da fattori più specifici e dinamici rispetto alla media nazionale.

L'evoluzione dei cluster di appartenenza delle province ha rivelato una stabilità generale per la maggior parte del territorio, ma con un'accelerazione significativa dei cambiamenti nel biennio 2023-2024 (un aumento di circa il 10.5% delle province che cambiano cluster). Questo denota un periodo di maggiore dinamismo e trasformazione, che potrebbe indicare nuove sfide o opportunità emergenti, con alcune province che mostrano una notevole

volatilità nel loro posizionamento socio-economico.

Infine, il confronto tra gli indicatori medi per la Sicilia e l'Italia ha consolidato l'immagine di una regione con criticità persistenti. La Sicilia presenta valori medi significativamente più alti in termini di indice di criminalità totale, indice di litigiosità e quota di cause pendenti ultratriennali. La penetrazione della banda ultralarga rimane inferiore alla media nazionale, e il numero di mensilità di stipendio necessarie per comprare casa è più elevato. Nonostante un indice di vecchiaia complessivamente inferiore alla media nazionale, la regione affronta sfide demografiche legate alla bassa natalità e a un crescente indice di dipendenza degli anziani in alcune aree, accompagnate da un emergente indice di solitudine, come evidenziato dalle componenti principali degli ultimi anni.

7. Implicazioni Strategiche e Raccomandazioni

Le dinamiche osservate suggeriscono la necessità di un approccio multidimensionale e territorialmente consapevole per lo sviluppo della Sicilia:

1. Potenziare il Mercato del Lavoro di Qualità: Le politiche devono continuare a focalizzarsi sulla creazione di posti di lavoro stabili e qualificati, con particolare attenzione alla riduzione della disoccupazione giovanile e femminile. Un mercato del lavoro inclusivo è il motore primario del benessere e della fiducia economica delle famiglie.

2. Accelerare la Transizione Digitale: È imperativo colmare il divario digitale, garantendo un accesso capillare alla banda larga e ultralarga. Questo non solo modernizza l'economia e facilita i servizi bancari e digitali, ma può anche fungere da catalizzatore per l'innovazione e la riduzione delle disuguaglianze territoriali.

3. Contrastare le Disuguaglianze e la Fragilità Urbana: La crescita economica non deve lasciare indietro le fasce più vulnerabili della popolazione. Sono necessarie politiche mirate per la redistribuzione del reddito, il sostegno alle famiglie con ISEE basso e la riqualificazione delle aree urbane più

fragili, garantendo che i benefici dello sviluppo siano equamente condivisi.

4. Rafforzare la Coesione Sociale e la Sicurezza: Gli indicatori di criminalità e litigiosità richiedono interventi volti a migliorare la sicurezza percepita e reale dei cittadini, rafforzando l'efficienza del sistema giudiziario. Le nuove correlazioni con solitudine e bassa natalità evidenziano la necessità di investire nel tessuto sociale, promuovendo iniziative che favoriscano l'inclusione, il supporto alle famiglie e il contrasto all'isolamento.

5. Politiche Demografiche Mirate e Sostegno ai Servizi: Affrontare il calo demografico e l'invecchiamento della popolazione richiede un sistema integrato di servizi per la famiglia e per gli anziani, con un'attenzione particolare alla qualità della vita e al benessere sociale. È essenziale creare contesti che incentivino la natalità e contrastino l'isolamento.

6. Monitoraggio Costante e Flessibilità: L'evoluzione dinamica delle correlazioni e dei cluster sottolinea l'importanza di un monitoraggio continuo degli indicatori socio-economici. La capacità di adattare rapidamente le strategie di intervento in base ai cambiamenti osservati è fondamentale per garantire uno sviluppo equo e sostenibile.

8. Considerazioni Finali

Questa **Prima Edizione** costituisce il primo riferimento nazionale per una lettura **olistica e replicabile** delle interrelazioni territoriali. Il valore scientifico è fondato su un impianto metodologico trasparente—armonizzazione delle fonti, riduzione dimensionale, clustering non supervisionato, interpretazione controllata—che rende risultati **tracciabili, verificabili e confrontabili** tra territori. Il report opera come **infrastruttura di evidenze** a supporto di politiche pubbliche coerenti e valutabili.

Il cuore innovativo è **tecnologico**: integrazione di dati eterogenei, tecniche di apprendimento statistico e **intelligenza artificiale** per la lettura dei pattern consentono di identificare **interconnessioni non ovvie** tra “sintomi” territoriali lontani—lavoro giovanile, ecosistemi di servizi, dotazioni digitali, sicurezza, dinamiche demografiche—aprendo la strada a **soluzioni territoriali e sociali** più creative ed efficaci. La piattaforma analitica funziona da **radar dei segnali deboli**: coglie accoppiamenti inattesi, descrive nodi sistemici e suggerisce leve di intervento che trascendono i silos settoriali.

La **robustezza** delle correlazioni è sostenuta da un disegno conoscitivo che privilegia:

- **coerenza temporale** (stabilità delle relazioni lungo la serie storica e in diverse congiunture),
- **convergenza tra fonti** (indicatori indipendenti che puntano nella stessa direzione),
- **sensibilità controllata** (risultati consistenti al variare di normalizzazioni, parametri e soglie).

Questi principi garantiscono che le associazioni emerse non siano effetti spuri, ma **evidenze strutturate** utili alla definizione delle priorità e alla misurazione degli impatti.

Il risultato è uno **standard nazionale** per la lettura sistemica dei territori: un linguaggio comune tra amministrazioni centrali, Regioni e città, capace di orientare risorse e strumenti con maggiore **equità, tempestività e accountability**. Questa Prima Edizione fissa la soglia qualitativa: un **bene informativo condiviso** che rende l’azione pubblica più intelligente perché più informata, e più innovativa perché capace di **connettere ciò che, finora, non veniva messo in relazione**.