

Problembeschreibung

In den letzten Jahren hat die Grippeimpfung eine immer größere Bedeutung im Gesundheitswesen erlangt. Die Erhebung und Analyse von Impfdaten ist essenziell, um die Impfstrategie zu verbessern und somit die Gesundheit der Bevölkerung zu fördern. Im Rahmen dieser Challenge besteht die Herausforderung darin, wöchentliche aggregierte Daten zu Grippeimpfungen aus verschiedenen Regionen Deutschlands zu analysieren und zu interpretieren.

Die Analyse umfasst verschiedene Dimensionen, einschließlich regionaler Unterschiede, demografischer Gruppen sowie der Abweichung von Empfehlungen. Ein tieferes Verständnis der Impftrends ermöglicht es, gezielte Maßnahmen zur Verbesserung der Impfquote und der öffentlichen Gesundheit zu entwickeln.

Ziele und Aufgabenstellung

Eure Aufgabe ist es, verschiedene Fragestellungen im Zusammenhang mit den Grippeimpfungen in Deutschland zu bearbeiten und dabei datengetriebene Insights zu generieren. Folgende Fragestellungen sollen eure Auswertung und Analyse stützen:

1. **Analyse von Impftrends:** Identifiziert zeitliche Trends bei den Impfungen.
2. **Regionale Unterschiede:**
 - a. Analysiert Unterschiede zwischen den verschiedenen Regionen (Nord, Süd, Ost, West).
 - b. Welche Regionen haben die höchsten/niedrigsten Impfquoten?
3. **Demografische Analyse:**
 - a. Untersucht Unterschiede in den Impfquoten nach Geschlecht und Altersgruppen.
 - b. Gibt es Unterschiede zwischen gesetzlich und privat Versicherten?
4. **Unterschiede in den Risikogruppen:**
 - a. Lässt sich etwas in der Auswertung der Risikogruppen erkennen?
5. **Analyse der Abrechnungsziffern:**
 - a. Untersucht die Verteilung und Nutzung der Abrechnungsziffern.
6. Gibt es weitere interessante Auffälligkeiten oder Anwendungsbereiche innerhalb des Datensatzes?

Beschreibung des Datensatzes

Der Datensatz enthält wöchentlich aggregierte Daten zur Grippeimpfung und umfasst die folgenden Attribute:

- **week:** Kalenderwoche der Datensammlung.
- **kvregion:** Kassenärztliche Vereinigungsregion.
- **region:** Regionale Unterteilung.
- **specialization:** Spezialisierung der impfenden Fachgruppe. Die Ziffern entsprechen der LANR-Nummer:
https://de.wikipedia.org/wiki/Lebenslange_Arztnummer#Aufbau
- **gender:** Geschlecht der geimpften Personen.
- **age_group:** Altersgruppe der geimpften Personen.
- **insurancecode:** Abrechnungsziffer entsprechend EBM bzw. GOÄ-Abrechnungsziffer.

- **insurancetype:** Versicherungstyp (GKV/PKV).
- **absolute:** Absolute Anzahl der Impfungen.
- **extrapolated:** Hochgerechnete Anzahl der Impfungen. Es handelt sich um eine Extrapolation des Verschreibers (Arztes). Die Gesamtheit ist die Anzahl der Ärzte pro Fachgebiet (und/oder KV-Region) in Deutschland. Dies wird durch die Anzahl der aktiven Ärzte im relevanten Zeitraum geteilt.

Im Datensatz mit den Risikogruppen (*Flu weekly with risk groups*) gibt es eine zusätzliche Spalte:

- **risk_groups:** Spezifische Risikogruppe (sofern vorhanden)

Bewertungskriterien

Eure Lösungen werden anhand folgender Kriterien bewertet:

- **Datenanalyse und -verständnis:** Wie gut wurden die Daten verstanden und sinnvoll aufbereitet?
- **Interpretation und Insights:** Welche wertvollen Erkenntnisse wurden gewonnen und wie plausibel sind diese (u.a. anhand der obenstehenden Fragestellungen)?
- **Kreativität und Originalität:** Wie innovativ sind die entwickelten Lösungsansätze?
- **Präsentation und Kommunikation:** Wie gut und klar wurden die erzielten Ergebnisse kommuniziert? Ist die Präsentation logisch strukturiert und benutzerfreundlich?

Hinweise und Ressourcen

- Nutzt gängige Datenanalysetools und Programmiersprachen wie z.B. Python, R, und entsprechende Bibliotheken (z.B. Pandas, Matplotlib, Seaborn).
- Denkt sowohl quantitativ als auch qualitativ und bezieht möglicherweise **externe Quellen/Studien ein, um eure Erkenntnisse zu stützen.**
- Die Daten sind anonymisiert und können ohne personenbezogene Rückschlüsse analysiert werden.