Projektname

Datenbericht

Letzte Änderung: 24.08.2021

# Rohdaten

## Übersichtstabelle der Rohdatensätze

|  |  |  |
| --- | --- | --- |
| **Datensatz Name** | **Quelle** | **Speicherort** |
| Dataset 1 History of Pandemics in Switzerland | Kaspar Staub and Wiebke Weber from  leaddata.ch | Uploaded Excel file: 1\_History\_Pandemics.xlsx in GitHub Repo: [Project\_PODSV/Data at main · plospen1/Project\_PODSV](https://github.com/plospen1/Project_PODSV/tree/main/Data) |
| Dataset 2 | … | … |

## Details Dataset 1

* **Description**:  
  Annual mortality statistics in Switzerland from 1880 to 2022, including pandemic-related deaths and excess mortality estimates.
* **Source Details:**

Historical data from Switzerland

COVID-19 data from Switzerland

* **Data Acquisition:¨**

File: 1\_History\_Pandemics.xlsx, contact Person Wiebke Weber

* **Legal Aspects:**

The data is publicly available and free to use for educational and analytical purposes.

* **Data Governance:**
* Category: Public
* No personal data is included.
* **Access Information:**
* Provided as an Excel file. On GitHub, linked at Speicherort.
* Can be accessed using common tools (Excel, Python pandas, etc.)

### Datenkatalog

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Spaltenindex** | **Spaltenname** | **Datentyp** | **Wertebereich / Validierung** | **Beschreibung** |
| 1 | Jahr | Integer | 1880–2022 | Year of observation |
| 2 | Todesfälle\_Grippe\_100000 | Float | ≥ 0 | Influenza deaths per 100,000 population |
| 3 | Todesfälle\_Covid\_100000 | Float | ≥ 0 | COVID-19 deaths per 100,000 population |
| 4 | TodesfälleGrippe | Float | ≥ 0 | Total number of influenza deaths |
| 5 | TodesfälleCOVID | Float | ≥ 0 | Total number of COVID-19 deaths |
| 6 | Population | Integer | ≥ 0 | Total population of Switzerland |
| 7 | Überasterblichkeit\_Alles | Float | Any value (%) | Total excess mortality percentage |
| 8 | ÜberasterblichkeitPlus | Float | ≥ 0 | Positive excess mortality percentage |
| 9 | ÜberasterblichkeitMinus | Float | ≤ 0 | Negative excess mortality percentage |

## Datenqualität Dataset 1

Methodology

* The analysis was performed using automated data profiling tools such as ydata.
* Descriptive statistics and a Pearson correlation matrix were used.
* The dataset includes 143 annual observations from 1880 to 2022.
* There are 9 variables total: 7 numeric and 2 categorical.

General Statistics

* Number of variables: 9
* Number of observations: 143
* Time span: 1880–2022
* Missing cells: 33.3% of all values
* No duplicate rows or invalid entries
* Mean year: 1951
* Mean population: approximately 5.2 million

Key Variables and Values

* TodesfälleGrippe (Influenza deaths): 3 to 21,491 per year, mean = 767, 2.1% missing.
* TodesfälleCOVID: 3 values only, 97.9% missing.
* Todesfälle\_Grippe\_100000: 0.0385 to 554 per 100k, mean = 18.2, minimal missing.
* Todesfälle\_Covid\_100000: 3 values, 97.9% missing.
* ÜbersterblichkeitMinus: -13.8% to 0%, mean = -2.87%, 52.4% missing.
* ÜbersterblichkeitPlus: 0% to 49.3%, mean = 3.74%, 46.2% missing.
* Übersterblichkeit\_Alles: nearly complete, -13.8% to 49.3%, mean = 0.66%.

Interpretation and Insights

* The dataset is ideal for long-term influenza mortality analysis.
* COVID-19 data is too incomplete for reliable modeling.
* Excess mortality is a valuable measure for both direct and indirect pandemic impact.
* Population normalization (per 100k) allows comparison across time.
* Negative excess mortality years may reflect successful health measures or mild seasons.

Data Quality Assessment

* Completeness: moderate (mainly missing values in COVID fields).
* Consistency: high (no duplicates or major outliers).
* Usability: excellent for flu and excess mortality; limited for COVID.

Conclusion

* Long-term flu mortality trends
* Excess mortality during pandemics
* Data-driven storytelling

## Details Dataset 2

* Beschreibung, welche Informationen die Daten enthalten
* Details zur Datenquelle
* Informationen zur Datenbeschaffung: Beschreibung und eventuell Verweise auf Ressourcen (Download-Scripts, Tools, Online-Services, …)
* Rechtliche Aspekte zur Nutzung der Daten, Lizenzen etc.
* Data Governance-Aspekte: Kategorisierung der Daten aufgrund Geschäftsinterner Vorgaben, z.B. Öffentlich, Geschäftsrelevant, Persönlich
* Wie kann auf die Daten zugegriffen werden: Beschreibung, Scripts, Tools, …
* Falls zutreffend: Unterteilung in abhängige (Zielgrösse, Regressand) und unabhängige (Regressor) Variablen
* …

### Datenkatalog Dataset 2

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Spaltenindex | Spaltenname | Datentyp | Werte (Wertebereich, ev. Validierungsregeln) | Kurze Beschreibung | Primary/ Foreign Key |
| 1 |  |  |  |  |  |
| 2 |  |  |  |  |  |
| … |  |  |  |  |  |

Beispiel eines Schemas von Daten in einer relationalen Datenbank:

Ein Bild, das Text, Screenshot, Schrift, Zahl enthält.

KI-generierte Inhalte können fehlerhaft sein.

### Datenqualität Dataset 2

Explorative Datenanalyse

* Beschreibung der Vorgehensweise und Verweise auf verwendete Ressourcen (Scripts, Tools, Software, …) und Output-Dateien -> Nachvollziehbarkeit und Sicherstellung der Reproduzierbarkeit
* Zusammenfassende Ergebnisse: Visualisationen und statistische Kenngrössen
* Erkenntnisse und relevante Zusammenhänge im Bezug auf die Modellierungs- und Analyseziele
* Einschätzung der Datenqualität

# Prozessierte Daten

## Übersichtstabelle der Prozessierten Daten

|  |  |  |
| --- | --- | --- |
| Name | Input-Datensätze | Speicherort |
| Processed Dataset 1 | Dataset 1, Dataset 2 | Link und/oder kurze Beschreibung des Ortes, wo die Daten abgelegt wurden |
| Processed Dataset 2 | Dataset 2 |  |

## Details Processed Dataset 1

* Beschreibung, welche Informationen die Daten enthalten
* Details und Gründe zu den Prozessierungsschritten -> Nachvollziehbarkeit und Sicherstellung der Reproduzierbarkeit
* Wie kann auf die Daten zugegriffen werden: Beschreibung, Scripts, Tools, …
* …

## Details Processed Dataset 2

* Beschreibung, welche Informationen die Daten enthalten
* Details und Gründe zu den Prozessierungsschritten
* Wie kann auf die Daten zugegriffen werden: Beschreibung, Scripts, Tools, …
* …