Projektname

Datenbericht

Letzte Änderung: 24.08.2021

# Rohdaten

## Übersichtstabelle der Rohdatensätze

|  |  |  |
| --- | --- | --- |
| **Datensatz Name** | **Quelle** | **Speicherort** |
| Dataset 1 History of Pandemics in Switzerland | BAG???? | Uploaded Excel file: 1\_History\_Pandemics.xlsx in GitHub Repo |
| Dataset 2 | … | … |

Falls zutreffend: Entity-Relationship Diagram

## Details Dataset 1

* 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
* …
* **Description**:  
  Annual mortality statistics in Switzerland from 1880 to 2022, including pandemic-related deaths and excess mortality estimates.
* **Source Details:**
* Historical data: ????.
* COVID-19 data: ????
* **Data Acquisition:**
* File: 1\_History\_Pandemics.xlsx
* **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.
* Can be accessed using common tools (Excel, Python pandas, etc.)

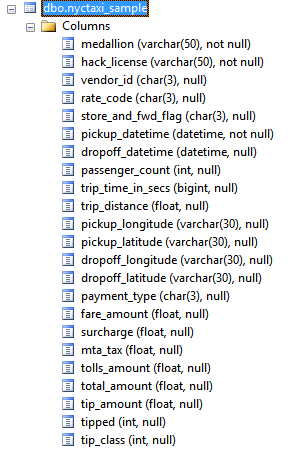
### Datenkatalog

Beispiel eines Data Dictionary

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **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 |

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

Beispiel eines Schemas von Daten in einer relationalen Datenbank:



## Datenqualität Dataset 1

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
* **Explorative Data Analysis (EDA):**
* Tools used: Python (pandas, matplotlib, bokeh)
* Scripts available: Visualization scripts created in Jupyter Notebook.
* Steps:
  + Check for missing values.
  + Basic statistical summary (describe()).
  + Visualizations: bar plots, line plots, excess mortality scatter plots.
* **Summary Results:**
* Data spans 143 years.
* Some missing values, mainly for pandemic-specific fields.
* COVID-19-related fields (Todesfälle\_Covid\_100000, TodesfälleCOVID) are populated only for 2020–2022.
* **Insights:**
* 1918 (Spanish Flu) shows the highest historical excess mortality.
* COVID-19 pandemic impact (2020) is clearly visible but less extreme compared to 1918.
* **Assessment of Data Quality:**
* Overall good: data is consistent and plausible.
* Missing values are explainable (e.g., COVID-19 only after 2020).
* No major anomalies detected in historical data.

## Details Dataset 2

* …

### Datenkatalog Dataset 2

### Datenqualität Dataset 2

# 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, …
* …