

Uitleg

Datatypen en -structuren

- **Kwantitatieve data:** Numerieke waarden (bijv. aantallen, meetwaarden). Te visualiseren met histogrammen, boxplots.
- **Kwalitatieve data:** Categorieën of tekst (bijv. landen, productnamen). Te visualiseren met staafdiagrammen, woordwolken.
- **Gestructureerde data:** Tabellen in CSV, databases. Helpt bij kolom- en rijstructuur.
- **Ongestructureerde data:** Tekstbestanden, logs, afbeeldingen. Vereist extractie of transformatie.

Data bronnen

- **Open data portals:** CBS, Eurostat, World Bank.
- **API's:** RESTful endpoints, JSON- of XML-response.
- **Web scraping:** Automatisch gegevens ophalen van webpagina's met tools als BeautifulSoup.
- **Eigen data:** Sensoren, enquêtes of experimenten.

Data importeren

- In Python: gebruik `pandas.read_csv()`, `pandas.read_json()`, `sqlalchemy` voor databases.
- In Excel: data import wizard, Power Query voor complexere bewerkingen.

Data cleaning technieken

- **Ontbrekende waarden:** Identificeren (`isnull()`), verwijderen (`dropna()`) of imputeren (gemiddelde, mediaan).
- **Outliers:** Detectie via statistische methoden (IQR, z-score), visuele inspectie (boxplot).
- **Duplicaten:** Identificatie (`duplicated()`), verwijderen (`drop_duplicates()`).
- **Normalisatie/standaardisatie:** Schalen van kolommen (min-max, z-score) voor vergelijkbare eenheden.

Data profiling

Maak een eerste verkennende analyse:

- **Statistische samenvatting:** `describe()` toont mean, median, min-max.
 - **Distributies bekijken:** Histogrammen, density plots.
 - **Relaties onderzoeken:** Scatterplot matrix, correlatiematrix.
-

Revision #1

Created 13 June 2025 09:50:31 by Admin

Updated 13 June 2025 09:51:28 by Admin