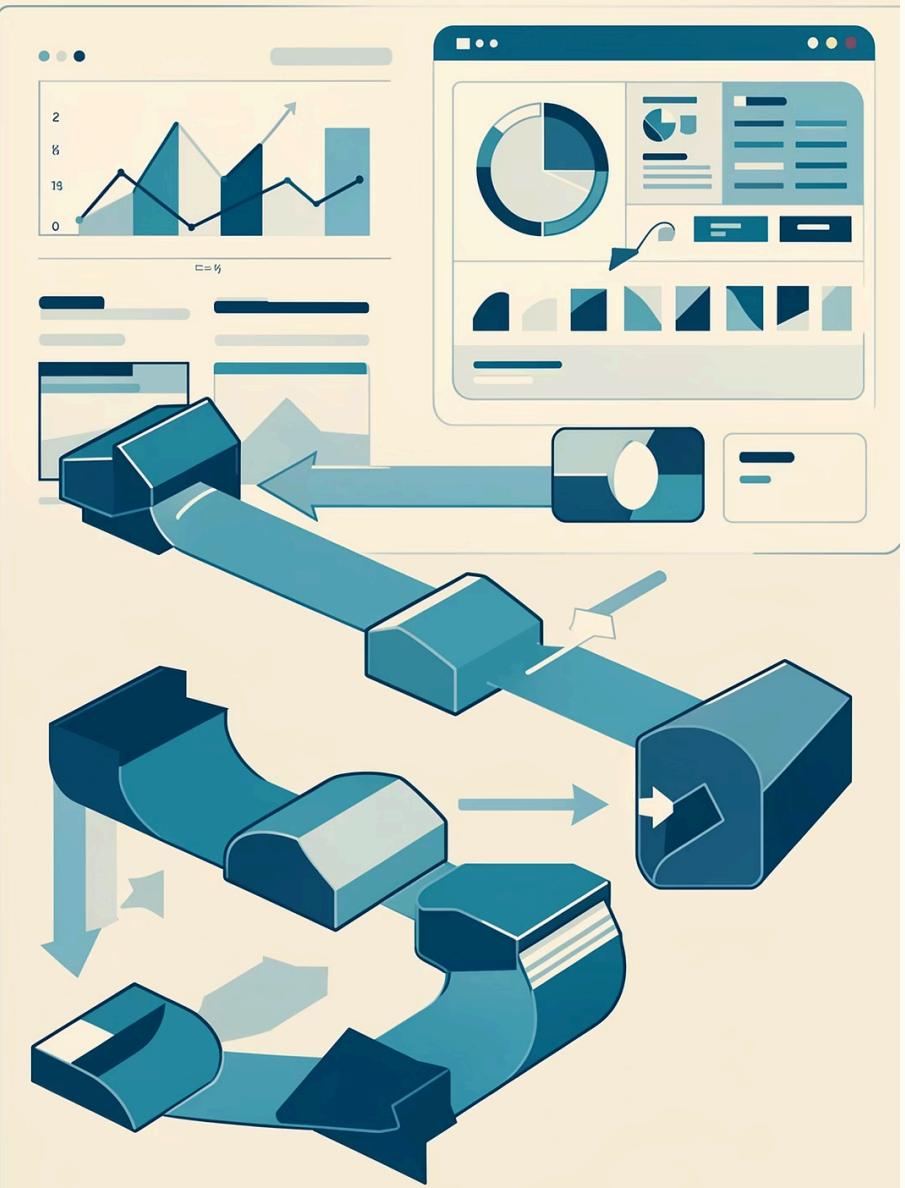


Data Cleaning and Preprocessing

Data Analytics Internship

Anisha More

Dataset Overview

Initial State

Raw dataset with inconsistencies

- Missing values present
- Duplicate records found
- Formatting issues

Missing Value Handling

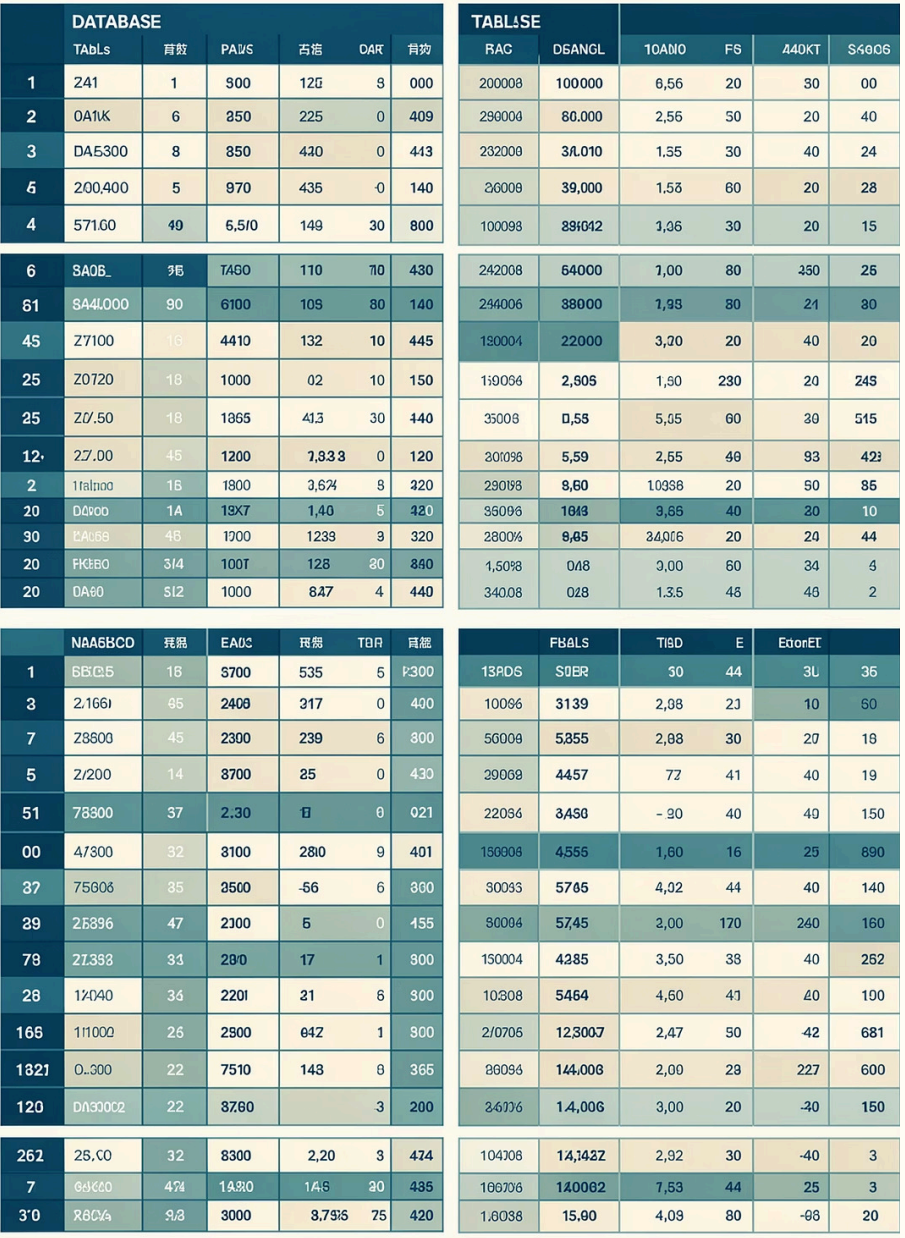
- Identified null values
- Used mean/median imputation
- Removed rows with excessive missing data

Duplicate Removal

Detected duplicate records

Removed duplicates

Ensured data uniqueness



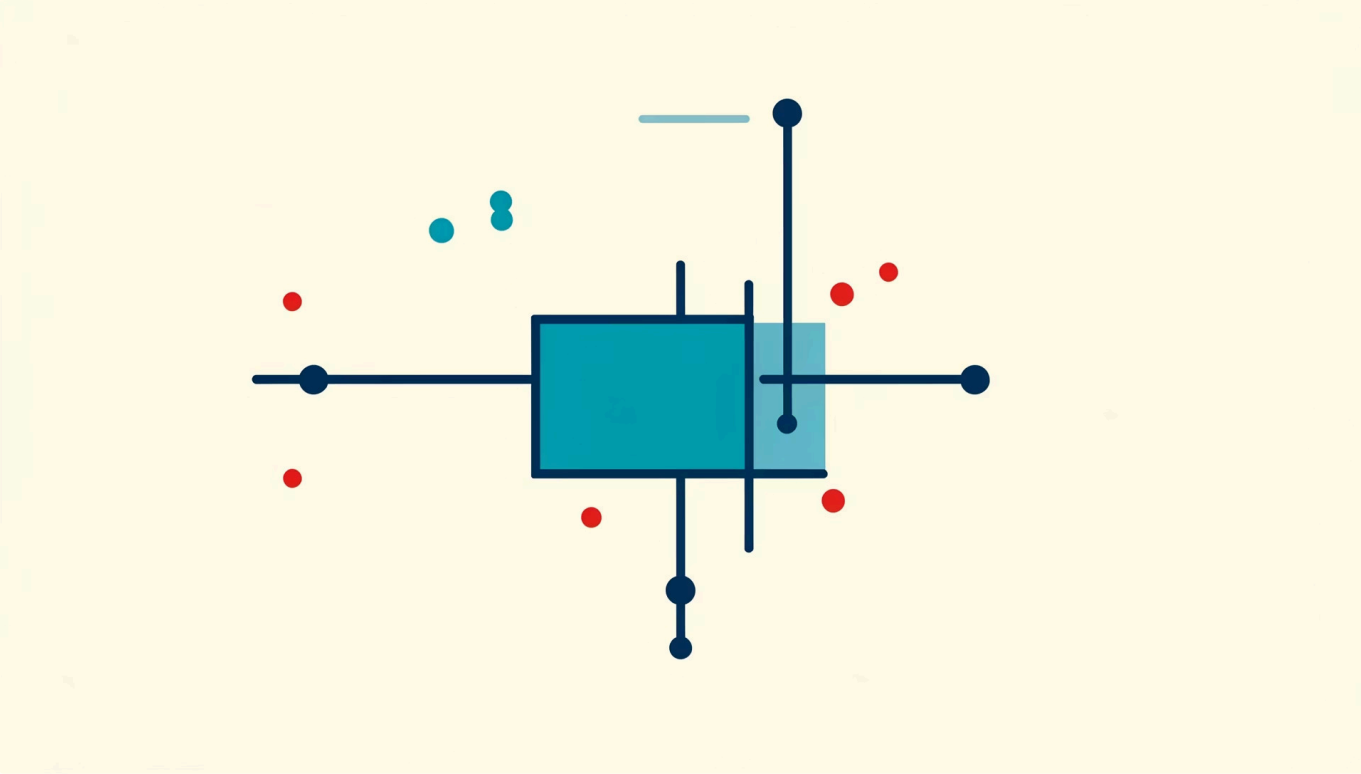
Standardization

Standardized date formats

Unified categorical values

Corrected inconsistent entries

Outlier Detection



01

Used boxplots for detection

02

Identified extreme values

03

Treated outliers appropriately

Final Outcome

Clean and structured dataset

Improved data quality

Ready for analysis

🎤 Speak: