

EDA Practice Case Studies

Case Study 1: URL Dataset (Titanic Dataset)

Dataset URL:

<https://raw.githubusercontent.com/mwaskom/seaborn-data/master/titanic.csv>

Tasks:

1. Load dataset from URL using Pandas.
2. Identify rows, columns, data types, and target variable.
3. Handle missing values and justify method.
4. Perform distribution analysis (Age, Fare).
5. Detect outliers using IQR method.
6. Create correlation heatmap and pairplot.
7. Perform categorical analysis (Sex, Pclass vs Survived).
8. Write key insights about survival patterns.

Case Study 2: SQL Dataset (Customer Churn)

Table Name: customers

Columns:

customer_id, age, gender, city, annual_income,
spending_score, membership_years, churn

Tasks:

1. Create SQLite database and insert records.
2. Load data using pd.read_sql().
3. Identify numerical and categorical variables.
4. Handle missing values and duplicates.
5. Perform distribution analysis (income, spending score).
6. Detect outliers using IQR.
7. Analyze relationship between income, membership_years and churn.
8. Provide business recommendations.