Data Cleaning Project

Prepared by: **Vyshnavi Patibandla**

This project focuses on cleaning and preprocessing two datasets — **AB_NYC_2019.csv** (Airbnb dataset) and **US_category_id.json** (YouTube data). The main objective is to ensure data accuracy, consistency, and reliability by performing essential data cleaning operations. Key steps include: • Handling missing values by imputing numeric and categorical columns.
• Removing duplicate records to maintain data integrity.
• Standardizing column names for uniformity.
• Detecting and treating outliers using the IQR method.
• Validating data quality through visual inspection (heatmaps). After cleaning, the datasets were free from missing, duplicate, and inconsistent data — ready for further analysis or modeling. This project enhanced data preprocessing, data visualization, and analytical thinking skills.

Tools Used: Python, Pandas, NumPy, Matplotlib, Seaborn

Date: November 2025