Data Analysis Report

Project: Data Cleaning & Missing Value Handling

This report summarizes the data cleaning and preprocessing steps applied to two datasets: House Prices Dataset and Medical Appointment No Shows Dataset. The main objective was to handle missing values and improve data quality.

Datasets Used

1. House Prices Dataset – Contains housing-related features such as price, bedrooms, and location details.
2. Medical Appointment No Shows Dataset – Contains medical appointment records and patient attendance.

Data Cleaning Methodology

- Identified missing values using Pandas.
- Visualized missing data patterns using bar charts.
- Applied median imputation for numerical columns.
- Applied mode imputation for categorical columns.
- Removed columns with extremely high missing values.
- Validated datasets after cleaning.

Results & Observations

After cleaning, both datasets had no missing values. Data consistency and reliability improved, making the datasets suitable for further analysis and machine learning tasks.

Conclusion

This task provided hands-on experience in data preprocessing and emphasized the importance of handling missing values to ensure high data quality.