

POC HOUSE DATA

PROJECT SUMMARY:

For House Dataset I'm responsible for cleaning, analyzing, and preparing a cleaned housing dataset. I began by loading and inspecting the raw data, identifying inconsistencies, missing values and non-standard entries. After cleaning the dataset, I performed EDA to show the relationship between features. Observed outliers using IQR method and handled using clipping techniques. The Final output is clean, well-structured dataset for machine learning models

SPECIFIC ACTIONS & TECHNOLOGIES USED

- ✓ Used pandas, Numpy, Matplotlib.pyplot, seaborn to Loading and manipulating enhancing plot and visualization.
- ✓ Replaced all the missing values with 'NAN' and nan's with median and mode where it required.
- ✓ Converted datatypes of the features which they are suitable to
- ✓ Detected and handled outliers in numeric features using the **IQR method** and ".clip()"
- ✓ Performed **univariate and bivariate analysis** using **Seaborn** and **Matplotlib** visualizations
- ✓ Used a **correlation heatmap** to identify strong predictors of sold_price
- ✓ Exported a cleaned data set and stored it a "**cleaned_house_data.csv**" did all the visualizations on uncleaned data because of outliers it won't lead to miss-calculations, and I can have a clear observation of outliers