**Interim Submission Report**

**Finding Summary:**

In the course of our interim analysis, several key findings have emerged, along with notable challenges and adjustments to our original plan:

1. **Key Findings:**

* The dataset contains a comprehensive collection of real estate pricing data, including various attributes such as property features, location, and sale prices.
* Initial exploration revealed instances of missing values and duplicate entries, highlighting the need for thorough data cleaning to ensure the integrity of our analysis.
* Univariate analysis uncovered interesting patterns in the distribution of key variables, providing valuable insights into the factors influencing house prices.

1. **Challenges encountered:**

* Dealing with missing values and anomalies posed challenges during the data cleaning process, requiring careful consideration and implementation of appropriate techniques.
* Ensuring clear documentation and comments within the Jupyter Notebook to explain each step of the analysis proved to be time-consuming but essential for transparency and reproducibility.
* Coordinating data storage and version control on GitHub necessitated careful organization and adherence to best practices to maintain a structured and accessible repository.

1. **Adjustments to the original plan:**

* Due to unexpected complexities in the dataset, additional time was allocated for data cleaning and preprocessing to ensure the quality of our analysis.
* We refined our approach to univariate analysis, focusing on key variables such as house prices and property features to uncover meaningful insights.

**Insights from Univariate analysis:**

Our initial univariate analysis yielded several insightful observations regarding the distribution of key variables:

1. **House Price Distribution:**

* The distribution of house prices exhibits a right-skewed pattern, indicating a wide range of property values in the market.
* Outliers were observed at the higher end of the price spectrum, suggesting the presence of luxury properties or exceptional market conditions in certain areas.

1. **Location Demographics:**

* Analysis of location demographics revealed varying trends in different neighborhoods, with some areas showing higher demand and property values compared to others.
* Factors such as proximity to amenities, schools, and transportation hubs appeared to influence property prices significantly.

1. **Property Features:**

* Exploration of property features, such as the number of bedrooms, bathrooms, and square footage, highlighted the diversity of housing options available in the market.
* Certain features, such as larger living spaces or additional amenities like swimming pools or garages, appeared to command higher prices.

1. **Temporal Trends:**

* Preliminary examination of temporal trends suggested seasonal fluctuations in house prices, with peak buying periods coinciding with specific times of the year.
* Further analysis is needed to understand the underlying factors driving these temporal variations and their implications for market dynamics.

**Submitted By:**

**Geet Govind**