Project Proposal: Statistical Analysis of Real Estate Prices

**1. Project Title**

Statistical Analysis of Real Estate Pricing Patterns

**2. Objective**

The primary objective of this project is to analyze real estate prices based on various attributes such as the number of bedrooms, furnishing status, payment options, and location (cities). This analysis will help to identify significant factors that influence real estate prices and to understand the distribution and variability of these prices.

**3. Background**

Real estate markets are influenced by multiple factors, including location, property size, furnishing status, and payment methods. Understanding these factors can provide valuable insights for buyers, sellers, and real estate professionals. By performing a detailed statistical analysis, we can identify patterns and trends in real estate pricing, as well as the impact of different attributes on these prices.

**4. Scope**

The scope of this project includes:

- Importing and cleaning a real estate dataset.

- Converting relevant columns to appropriate data types.

- Handling missing values and outliers.

- Splitting data based on payment methods.

- Performing statistical tests to compare prices across different categories.

- Visualizing the data to highlight significant findings.

**5. Methodology**

1. Data Import and Cleaning

- Load the dataset using the pandas library.

- Convert relevant columns (e.g., Price, Bedrooms, Bathrooms, Area) to numeric data types.

- Handle missing values by replacing 'Unknown' with NaN and dropping rows with critical missing data.

2. Data Exploration

- Preview the dataset to understand its structure and summary statistics.

- Replace missing values and drop rows with missing data in critical columns.

3. Outlier Detection and Treatment

- Use the Interquartile Range (IQR) method to identify and remove outliers in the 'Price' column.

- Plot the distribution of prices after outlier removal.

4. Data Splitting and Statistical Analysis

- Split the data based on payment options (cash vs. installment) and perform an independent t-test to compare prices.

- Split the data based on the number of bedrooms and perform an independent t-test to compare prices.

- Split the data based on furnishing status and perform a Mann-Whitney U test to compare median prices.

5. Variance Analysis

- Analyze the variance in prices across the top three cities using ANOVA.

- Visualize price distributions across different categories using histograms and boxplots.

6. Tools and Technologies

- Programming Language:Python

- Libraries:Pandas, NumPy, Scipy, Matplotlib, Seaborn

- Development Environment: Jupyter Notebook

7. Expected Outcomes

- A clean and well-structured dataset ready for analysis.

- Detailed statistical analysis of real estate prices based on different attributes.

- Identification of significant factors influencing real estate prices.

- Visual representations of price distributions and variability across different categories.

- Insights into whether cash payments result in lower prices compared to installments, the impact of the number of bedrooms on prices, the effect of furnishing status on prices, and price variability across top cities.

Conclusion

This project aims to provide a comprehensive statistical analysis of real estate prices, leveraging various statistical methods to uncover significant insights. The findings from this project will be beneficial for stakeholders in the real estate market, including buyers, sellers, and analysts.