You are a Data Analyst at a reputable property company based in Colombo, where the demand for rental properties is rapidly growing. In response to this rising market, your company is keen to leverage data to gain a deeper understanding of the current rental landscape in Colombo. Your goal is to provide valuable insights into rental price trends, apartment features, and the key factors influencing rental prices. To begin this journey, your team has collected a dataset comprising various apartments from different neighbourhoods across Colombo.

DatasetFeatures:

a. **Apartment_ID**: Unique identifier for each apartment listing
b. **Neighborhood**: Name of the neighbourhoods
c. **Rental_Price**: Monthly rental price (in USD)
d. **Size_in_Sqft**: Size of the apartment in square feet
e. **Distance_to_City_Center_KM:** Distance to the Colombo Fort Station
f. **Bedrooms**: Number of bedrooms
g. **Bathrooms**: Number of bathrooms
h. **Furnished**: Whether the apartment is furnished
i. **Building_Type**: Type of the building

(1) Descriptive Statistics:

a. Generate summary statistics tables which consist of extreme values (min & max), standard deviation, median and quartile elements for all the numerical features in the dataset.

b. Generate summary statistics tables for all the categorical features in the dataset.

(2) Data Visualization:

a. Create a visual representation displaying the average rental price for each neighbourhood, categorized by building type.

b. Create a **histogram** of the *Rental_Price* to show the distribution of prices.

c. An array of plots of variable pairs provides very interesting insights into the distribution of data values and their correlation. Correlation coefficients between a pair of attributes illustrates a linear relationship between each other. Visualise the full array of plots of variable pairs for the numerical features in the given dataset.

d. Create a **boxplot** to compare the distribution of *Rental_Price* for *furnished* vs. *unfurnished* apartments.

e. Create a **heatmap** to visualize the correlation between numerical features.

(3) Explorative Data Analysis:

Provide all the analysis, visualizations, and discussion for the EDA questions below.

3. A Understand Rental Price Trends

a. What is the average rental price across different neighborhoods?

b. How does rental price vary based on the size of the apartment (sq ft)?

c. Is there a relationship between *Rental_Price* and *Distance_to_City_Center*?

d. Do furnished apartments have a higher rental price compared to unfurnished ones?

e. How does rental price vary by building type?

f. Calculate the probability of finding any apartment in Downtown with a rental price below Rs 500,000/= using the normal distribution assumption. Provide all the calculations and the visual diagram. Assume 1 USD = Rs 304.00

3.B Apartment Features & Their Influence

a. What is the distribution of apartment sizes across neighborhoods?

b. How does the number of bedrooms and bathrooms impact on rental pricing?

c. What is the proportion of furnished vs. unfurnished apartments in each neighborhood?

d. Do larger apartments tend to have more bathrooms and bedrooms, or is there variation?

3.C Neighborhood Comparisons

a. Which neighborhoods have the highest and lowest rental prices?

b. How does the size of apartments vary by neighborhood?

c. Are there certain neighborhoods where furnished apartments are more common?

3.D Market Dynamics & Outliers

a. Are there any outliers in rental pricing? (e.g., extremely high or low rents)

b. Are there neighborhoods where rental prices do not align with apartment sizes?

c. Are there trends in rental pricing based on proximity to the city center?

3.E Correlations & Patterns

a. What are the strongest factors influencing rental prices?

b. Is there a correlation between apartment size and rental price?

c. Does distance from the city center significantly impact rental price?

3.F Write down three additional hypotheses for data analysis that you would like to test, and provide all the calculations, visualizations, and discussions.