LAB EVALUATION

Q1 For the given insurance dataset justify the hypothesis

1. Does bmi of males differ significantly from that of females?
2. Is the proportion of smokers significantly different in different genders?

Q2 FOR the Airbnb NYC 2023 dataset

A. Data Cleaning & Preprocessing

1. Handle missing, duplicate, and inconsistent values.
2. Convert categorical variables appropriately (e.g., encode neighbourhoods).
3. Handle skewed distributions (log-transform prices, etc.).
4. Drop irrelevant or highly correlated features (justify your decision).

Part B: Exploratory Data Analysis (EDA)

1. Univariate, bivariate, and multivariate analysis.
2. Visualize price distribution across:
    - Room types
    - Neighborhood groups
    - Availability and review counts
3. Use box plots, violin plots, scatter matrices, and heatmaps to show relationships.

Part C: Analytical Questions (Critical Thinking Required)

1. What are the top 5 factors most correlated with high-priced listings?
2. Are there pricing anomalies? Can you spot over/underpriced listings?
3. Which neighborhoods are oversaturated or underserved?
4. Can you cluster neighborhoods based on listing behavior and pricing?
5. Is there a relationship between review scores and pricing strategy?

Q3 For the dataset that includes CO2 emissions from each energy resource starting January 1973 to July 2016. Answer the following

1. Check the stationary using ADF test and autocorrelation plot.
2. Forecast the target variable prediction using a suitable type of model.
3. Evaluate the different types of error residues to check fitness of good of the model.
4. Forecast the predictions for next 10 years on target variable.