Data Analytics 1
Assignment 1: Data Visualization

Release: 08 August 2023
Deadline: 16 August 2023 (11:55 pm)

The objective of this assignment is to introduce you to data visualization and develop an understanding of data processing using Python.

For this assignment, you have been provided with a detailed dataset containing information about various real estate properties. This dataset encompasses a variety of property types and features across different cities.

Imagine yourself in the role of a data scientist, tasked with sharing your discoveries from the dataset with investors who are seeking real estate investment opportunities. Keep in mind that each investor has a unique investment strategy. Take into account the specified criteria given below by the investors and create relevant visual representations accordingly.

1. As investors have diverse budget limits, split the overall opportunities into three different price ranges. Analyze the distribution of these price ranges across cities. (5 Marks)

2. Provide investors with a high-level summary of the whole data using the price range dimensions. Consider factors like property type, city, and key amenities. What method/tool would you use to present this information effectively? (5 Marks)

3. Some investors want to compare investment opportunities in Thane and Mumbai. Analyze and visualize the differences in property types, sizes (using Carpet Area), and prices between these two cities. (10 Marks)

4. Certain investors seek to compare investment opportunities in less expensive and more expensive localities within each city. For instance, compare the average Carpet Area of high-budget properties in prime locations (isPrimeLocationProperty) versus non-prime locations. (15 Marks)

5. Some investors prefer larger area properties with relatively lower budgets. Identify and visualize properties that offer the best value in terms of Carpet Area per unit of Price across different cities. (10 Marks)

6. Some investors are interested in knowing the hotspots for their offices in Mumbai and Thane. Analyze commercial properties in these cities, considering factors like Price, Carpet Area, Amenities, and Location. (10 Marks)

7. Some investors want to analyze the impact of various amenities (e.g., Swimming Pool, Gymnasium, Club House) on property prices in Mumbai and Thane. Which amenities seem to add the most value? Are there differences in amenity preferences between the two cities? (marks 10)

8. Investors want investigate how the Possession Status and "Availability Starts From" dates affect property prices. Are there significant price differences between ready-to-move-in properties and under-construction ones? How does this vary between Mumbai and Thane? (marks 10)

9. Analyze the impact of developers on property prices and features. Are there certain developers associated with higher-end properties or better amenities? (marks 15)

10. Code quality (5 Marks)

    Note: You are expected to write the code in a vectorized way for data preprocessing using pandas instead of writing individual for loops (wherever possible)

11. Investor Report (5 Marks)

    Based on your analyses from the previous questions, create a comprehensive report for potential investors. This report should synthesize your findings and present clear, actionable insights about investment opportunities in the real estate market.

Additional Notes:

- For any of the above parts, it is encouraged to give your insights with multiple visualizations.
- You are free to use any tools like box plots, pivot tables, scatter plots, etc.
- Mention your inference/analysis for every question/visualization in the markdown in the notebook.
- Make use of relevant columns such as Amenities, Furnished Type, Bathroom, Parking, and other property features to provide deeper insights.

Submission format:

- Submit a zip folder named `<assignment1_teamId>` containing files `<assignment1_teamId>.ipynb` and `<assignment1_teamId_report>`