Data Analytics 1

Assignment 1

Data Visualization

Release : 11 August 2023
Deadline : 18 August 2023 (11:55 pm)

The Objective of this assignment is to initiate your introduction to visualization and develop an understanding of data processing using Python.

- For this assignment, you have been provided with a dataset containing information about both residential and commercial properties. This dataset encompasses a variety of property types, including apartments, houses, plots, and other related categories.
Now Imagine yourself in the role of a data scientist, tasked with sharing your discoveries from the datasets with investors who are seeking housing 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
Link For dataset is here

- As investors have diverse budget limits, split the overall opportunities into three different ranges. Analyze the distribution of the budget ranges : **8 Marks**
- Now since the budgets ranges are categorized , provide the investors with a high level summary of the whole data using the budget range dimensions. (What method/tool would you use ?) : **7 Marks**

- Some investors want to compare and analyze the investment opportunities in Hyderabad and bangalore . For instance, compare and analyze the size and cost of houses that can be purchased in each of these cities. : **20 Marks**
- Certain investors seek to compare investment oppurtunities in less expensive and expensive localities within each city.(for instance : compare the avg size of high budget houses in expensive localities and affordable localities) : **20 Marks**
- Some prefer larger area properties with relatively lesser budgets : 1**5 Marks**
- Some investors are interested in knowing the hotspot for their offices in Mumbai and Ahmedabad. : **25 Marks**
- 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)

Note :

- 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 notebook

Submission format:

- Submit a zip folder named <assignment1_teamId> containing a single file <assignment1_teamId>.ipynb