Amrita School of Computing

Department of Computer Science & Engineering

19CSE304 Foundations of Data Science

Lab Sheet 4- Exploratory Data Analysis

Exercise I Understand EDA

Open the below link and understand the visual analysis of the dataset House price

https://www.kaggle.com/code/ekami66/detailed-exploratory-data-analysis-with-python

Takeaways: Understand the following

- How is each attribute distributed?

- Distribution of all the features by plotting them

- Which features are strongly correlated?

- Identify features which have a strong relationship to a house price?

- Feature to feature relationship

- How many features can be reduced by visualizing heatmap?

- Distribution of quantitative features

- Categorical to Quantitative relationship

Exercise II Exploratory Visual Analysis

1. For the dataset AirQuality.csv do the following

2. Find the shape & structure of your dataset.

3. What variables does the dataset contain?

4. What is the data type of the column AQI_Bucket?

5. How are all the features distributed?

6. Are there any notable data quality issues? If so, handle it appropriately.

7. Are there any relationships among the variables?

8. Identify outliers in the dataset.

9. Use LOF method to remove outliers. (Refer lecture slides for code.)

10. Analyze the continuous values by creating histograms to understand the distribution

11. Identify correlation between features in the dataset.

12. Which features are strongly correlated?

13. Identify features which have a strong relationship to AQI

14. Find the correlation between PM2.5 and PM10?

15. Plot Feature to feature relationship

16. How many features can be reduced by visualizing heatmap? Show the visualization.

17. Plot the distribution of quantitative features

18. Show the categorical to Quantitative relationship using box plot and plot their distribution

Exercise III EDA Document Preparation

Prepare a document with 10 or more captioned visualizations detailing your most important insights from the data. Each visualization you can put as an image accompanied with a title and descriptive caption (1-4 sentences long) describing the insight(s) learned from that view.