CS632P Python Programming Midterm Requirement

Group Number:

Topic selected:

Problem statement: :

Data set selected:

Midterm requirements:

- You should be able to explain clearly about your dataset, what and why you are analyzing the dataset for (particular problem), what setbacks or limitations you had, what are your findings(results).
- Your presentation must have (Introduction, Problem definition or hypothesis, dataset used, your EDA backed by appropriate visualizations, results and conclusion)
- Use of python concepts like Data Types, Functions, Loops, OOPs concept, File Handling, Exception Handling(if any), libraries are expected.
- Submission : Python files, Presentation

EDA:

1. Loading data in to DataFrames.
2. Check the Data Types of your data columns.
3. Drop any NULL, missing values or unwanted columns.
4. Drop duplicate values.
5. Check for outliers using a box plot or histogram.
6. Plot features against each other using a pair plot.
7. Use a HeatMap for finding the correlation between the features(Feature to Feature).
8. Use a scatter plot to show the relationship between 2 variables.
9. Merging two Data Frames.
10. Slicing Data of a particular column value (like year, month, filter values depending on the categorical data)
11. Representing data in matrix form.
12. Upload data to Numerical Python (NumPy)
13. Select a slice or part of the data and display.
14. Use conditions and segregate the data based on the condition (like show data of a feature(column) >,<,= a number)
15. Use mathematical and statistical functions using libraries.
16. Select data based on a category(categorical data based).
17. Libraries expected to try(minimum 4 required): Pandas, Numpy, SciPy, Seaborn, Matplotlib
18. Box Plot, Scatter Plot, Pair Plot, Z-score, Histogram, Heatmaps.