Data Analytics List [15 Marks]

1. Create product advertising 'sales' Data set having 3 columns namely: ID, TV and Sales.(random 500 entries) Build a linear regression model by identifying independent and target variable. Split thevariables into training and testing sets. then divide the training and testing sets into a 7:3 ratio, respectively and print them. Build a simple linear regression model.

2. Create the following dataset in python

TID	*Items*
1	Bread, Milk
2	Bread, Diaper, Beer, Eggs
3	Milk, Diaper, Beer, Coke
4	Bread, Milk, Diaper, Beer
5	Bread, Milk, Diaper, Coke

Convert the categorical values into numeric format. Apply the apriori algorithm on the above dataset to generate the frequent itemsets andassociation rules. Repeat the process with different min_sup values.

3. Download the **Market basket dataset**.
Write a python program to read the dataset and display its information. Preprocess the data (drop null values etc.) Convert the categorical values into numeric format. Apply the apriori algorithm on the above dataset to generate the frequent itemsets and association rules.

4. Use the **Iris dataset**.
Write a Python program to view some basic statistical details like percentile,mean, std etc. of the species of 'Iris-setosa', 'Iris-versicolor' and 'Iris-virginica'. Apply logistic regression on the dataset to identify different species (setosa, versicolor, verginica) of Iris flowers given just 4 features: sepal and petal lengths and widths. Find the accuracy of the model.

5. Consider any text paragraph. Remove the stopwords. Tokenize the paragraph to extract words and sentences.

6. Create product advertising 'sales' Data set having 3 columns namely: ID, Radio and Sales.(random 500 entries) Build a linear regression model by identifying independent and target variable. Split thevariables into training and testing sets. then divide the training and testing sets into a 7:3 ratio, respectively and print them. Build a simple linear regression model.

7. . Create the following dataset in python

```
D= {{'water','Biscuit','Butter','orange'},
    {'Biscuit','water','Butter'},
    {'Biscuit','bread','orange','water'},
    {'Butter','Biscuit',orange},
    {'water', 'Butter','Biscuit'},
    {'bread','Butter'}
    {'water','Biscuit','orange'}}
```

Convert the categorical values into numeric format. Apply the apriori algorithm on the above dataset to generate the frequent itemsets andassociation rules. Repeat the process with different min_sup values

8. Consider any text paragraph. Tokenize the paragraph to extract words and sentences. Calculate the word frequency distribution and plot the frequencies.

9. Download the **groceries dataset.**
Write a python program to read the dataset and display its information.Preprocess the data (drop null values etc.) Convert the categorical values into numeric format. Apply the apriori algorithm on the above dataset to generate the frequent itemsets andassociation rules.

10. Consider the following review messages. Perform sentiment analysis on the messages.

i. I enjoy listening to music.

ii. I take a walk in the park everyday

11. Create 'User' Data set having 5 columns namely: User ID, Gender, Age, EstimatedSalary and Purchased. Build a logistic regression model that can predict whether on the given parameter aperson will buy a car or not.

12. Consider the following review messages. Perform sentiment analysis on the messages.

iii. I purchased headphones online. I am very happy with the product.

iv. I saw the movie yesterday. The animation was really good but the script was ok.

13. Download the **Market basket dataset**.
Write a python code to implement the apriori algorithm. Test the code on Market Basket dataset.

14. Create the following dataset in python

```
D= {{'eggs', 'milk','bread'},
    {'eggs', 'apple'},
    {'milk', 'bread'},
    {'apple', 'milk'},
    {'milk', 'apple', 'bread'}}
```

Convert the categorical values into numeric format. Apply the apriori algorithm on the above dataset to generate the frequent itemsets andassociation rules. Repeat the process with different min_sup values.

15. Create product advertising 'sales' Data set having 3 columns namely: ID, Newspaper and Sales. (random 500 entries) Build a linear regression model by identifying independent and target variable. Split thevariables into training and testing sets. then divide the training and testing sets into a 7:3 ratio, respectively and print them. Build a simple linear regression model.

16. Consider any text paragraph. Tokenize the paragraph to extract words and sentences. Plot the wordcloud ofthe text.

17. Create the following dataset in python

D= {{'butter',' bread',' milk','sugar'};
{'butter', 'flour',' milk','sugar'};
{'butter',' eggs',' milk',' salt'};
{'eggs'};
{'butter',' flour',' milk',' salt',' sugar'}}

Convert the categorical values into numeric format. Apply the apriori algorithm on the above dataset to generate the frequent itemsets andassociation rules. Repeat the process with different min_sup values.

18. Download the **groceries dataset.**
Write a python code to implement the apriori algorithm. Test the code on groceries dataset.

19. Build a multiple linear regression model for Fish Species Weight Prediction. (download dataset https://www.kaggle.com/aungpyaeap/fish-market?select=Fish.csv)

20. Consider the following review messages. Perform sentiment analysis on the messages.

i. I saw the movie yesterday. The animation was really good but the script was ok.
ii. I enjoy listening to music.