Exercise Based on Pandas, Matplotlib, Seaborn or Scikitlearn

Case Study1: Diabetes Analysis

1. Import the diabetes dataset
2. Read the top 5 records from dataset
3. Read the bottom 5 records
4. Identify the dependent & independent variable in the dataset
5. Display total number of records and attributes present in the dataset
6. Display more information about all attributes in the dataset
7. Display statistical information of a dataset
8. Get the list of all attributes present in the dataset.
9. Check & treat the NULL values in the dataset
10. Check & treat for Duplicate record present in the dataset
11. Check & treat outliers in dataset.
12. Generate the count plot for dependent variable in the dataset.
13. Generate the histplot for all attributes present in the dataset
14. Generate the heatmap and convey important independent variable affecting diabetes.
15. Generate the pairplot to check whether dataset is linearly or non linearly seperable
16. Apply feature scaling to scale the data in range.
17. Divide the dataset into dependent and Independent variables.
18. Split the dataset into train & test set with ratio of 80:20
19. Save all the charts generated during this case study
20. Provide your observation for every chart.

Note: Here outcome is a dependent variable.

Case Study 2: Loan Data Analysis

1. Import the diabetes dataset
2. Read the top 5 records from dataset
3. Read the bottom 5 records
4. Identify the dependent & independent variable in the dataset
5. Display total number of records and attributes present in the dataset
6. Display more information about all attributes in the dataset
7. Display statistical information of a dataset
8. Get the list of all attributes present in the dataset.
9. Check & treat the NULL values in the dataset
10. Check & treat for Duplicate record present in the dataset
11. Check & treat outliers in dataset.
12. Generate the count plot for dependent variable in the dataset.
13. Generate the histplot for all attributes present in the dataset
14. Generate the heatmap and convey important independent variable affecting diabetes.
15. Generate the pairplot to check whether dataset is linearly or non linearly seperable
16. Apply feature scaling to scale the data in range.
17. Divide the dataset into dependent and Independent variables.
18. Split the dataset into train & test set with ratio of 80:20
19. Save all the charts generated during this case study
20. Provide your observation for every chart

Note: Here deposit is a dependent variable.