Text

Description automatically generated

**CSE303: Lab Assignment – 1 (Tasks on Lab 04)**

**Individual Assignment**

|  |  |
| --- | --- |
| Student ID | 2019-1-60-024 |
| Student Name | Adri Saha |
| Course Name | Statistics for Data Science |
| Course Code | CSE303 |
| Section | 01 |

Submitted to:

Dr. Mohammad Rezwanul Huq

Associate Professor,

Department of Computer Science & Engineering

Date of submission: 12 July 2021

**Topic: Exploratory Data Analysis using Pandas**

Lab Tasks

1. How many rows and columns this dataframe has? Print this information.

Text

Description automatically generated

1. Describe (numerical summary) the time and amount column. Print this information.

Text

Description automatically generated

1. There are 31 columns in the dataset. Compute some statistical measures like mean, median, standard deviation, variance using Pandas Function for at least two columns. Print this information.

Text

Description automatically generated

1. Show the Box Plot of Time and Amount column. Also print the value of Q1, Median, Q3, IQR. Are there any outliers? Explain your answer and print it.

Text

Description automatically generated

1. Show the Histogram of Time and Amount column. Print the value of the Skewness and Kurtosis using appropriate Pandas functions. Comment on the type of the data distribution and print it.

Text

Description automatically generated

1. Find the percentage of records with class value = 0 (Non-Fraudulent) and class value = 1 (Fraudulent). Print this information.

Text

Description automatically generated

1. Show the result you have got in 6 using a Histogram.

Chart, histogram

Description automatically generated

1. Show the result you have got in 6 using a Bar chart. Create the bar chart on the percentage value, not on the total number of occurrences.

Chart, bar chart

Description automatically generated

1. Show the Histrogram (data distribution) of a few other columns (your choice) showing both positive and negative skew and also leptokurtic and platykurtic data distribution. So, you should display at least four Histograms.

Chart, bar chart

Description automatically generated

**Figure:** V23- Negative or left-skewed, V24- platykurtic kurtosis, V28- positive or right skewed, V12- Leptokurtic kurtosis

1. Find the highest positive correlation among all attributes. While finding the correlation, use appropriate code, not manually. Print this information accordingly.

Text

Description automatically generated

1. Support your findings in Question 10 using a Scatter Plot.

Chart, scatter chart

Description automatically generated

1. Find the highest negative correlation among all attributes. While finding the correlation, use appropriate code, not manually. Print this information accordingly.

Text

Description automatically generated

1. Support your findings in Question 12 using a Scatter Plot.

Chart, scatter chart

Description automatically generated

1. Create a Box Plot of the Amount Column.

Chart

Description automatically generated

1. Now create two other box plots side by side. The first one will show the Amount column value for which the class value = 0 (Non-Fraudulent) and the second one will show the Amount column value for which the class value = 1 (Fraudulent). Do you find any particular pattern by just considering Amount column. Explain your answer and print it accordingly.

Chart, box and whisker chart

Description automatically generated

Text

Description automatically generated

Text

Description automatically generated