PICT, PUNE

DSBDAL

Assignment - 08

Name: Omkar Gaikwad
Batch: L1
Roll no: 31126

Title: Data Visualization - I

Problem Statement:

1. Use the inbuilt dataset 'titanic'. The dataset contains information about the passengers who boarded the unfortunate Titanic ship. Use seaborn library to see if we can find any patterns in the data.
2. Write a code to check how the price of the ticket (column name: fare) for each passenger is distributed by plotting a histogram.

Learning Objectives:

- To understand various visualization techniques using seaborn python library.
- To apply appropriate plotting techniques to visualize numeric data types (i.e 'Fare')
- Describe the observations made by using each plot/graph

Learning Outcomes:

- Students will be able to:
- Perform basic visualisations using appropriate graphs.
- make observation on the 'Fare' of the ticket in the dataset using plots

PICT, PUNE

Plots used in Analysis:

① Displot : It is used basically for univariant set of observations and visualizes it through a histogram.

② jointplot : Draw a plot of two variables with bivariate and univariate graphs. It basically combines different plots.

③ Pairplot : It represents pairwise relation across the entire dataframe. It basically creates a jointplot between every possible numerical columns in dataframe.

④ Rugplot : It plots data points in an array as sticks on an axis. It takes in a single column attribute. Instead of drawing histogram, it creates dashes all across the plot.

Observations:

- The displot() indicates that most of ticket fares were below 200
- The modal fare for passengers was around 5-15
- The majority of passengers between the age of 20-30 have paid a fair less than 15
- Average age of passengers is between 20-40

Conclusion:

Thus, we have successfully applied various visualisation techniques and inferred the variation of ticket fare in titanic dataset

Plots used in Analysis:

① Displot : It is used basically for univariant set of observations and visualizes it through a histogram.

② jointplot : Draw a plot of two variables with bivariate and univariate graphs. It basically combines different plots.

③ Pairplot : It represents pairwise relation across the entire dataframe. It basically creates a jointplot between every possible numerical columns in dataframe.

④ Rugplot : It plots data points in an array as sticks on an axis. It takes in a single column attribute. Instead of drawing histogram, it creates dashes all across the plot.

Observations:

- The displot() indicates that most of ticket fares were below 200
- The modal fare for passengers was around 5-15
- The majority of passengers between the age of 20-30 have paid a fair less than 15
- Average age of passengers is between 20-40

Conclusion:

Thus, we have successfully applied various visualisation techniques and inferred the variation of ticket fare in titanic dataset.