HUMBER

Introduction to Data Analytics – Lab 2

Data Visualization

Objective

Practice creating various types of graphs using Matplotlib based on different data and feature types.

Instructions:

Make sure your code is error and warning free before submitting this activity. Use comments inside the code to provide explanations as required.

Write the entire activity in a Python file called <YOUR_STUDENT_NUMBER>_lab2.py.

Dataset:

Titanic dataset (contains information about passengers on the Titanic, including survival status, age, fare, class, and more)

To load the data use:

```
import pandas as pd

dataset url =
'https://raw.githubusercontent.com/datasciencedojo/datasets/master/titanic.csv'

df = pd.read_csv(url)
```

HUMBER

Introduction to Data Analytics – Lab 2

Steps

1. Load the Titanic Dataset:
    a. Use the provided URL to load the Titanic dataset into a Pandas DataFrame.
    b. Drop columns that are not needed for the visualization exercise (Name, Ticket, Cabin).
    c. Fill missing values in the Age column with the median age and in the Embarked column with the mode.
2. Create a Histogram:
    a. Plot a histogram to show the distribution of passenger ages.
    b. Use 30 bins, and make the bars slightly transparent for better visualization.
    c. Label the axes and add a title.
3. Create a Scatter Plot:
    a. Plot a scatter plot to show the relationship between Age and Fare.
    b. Make the points slightly transparent.
    c. Label the axes and add a title.
4. Create a Bar Chart:
    a. Plot a bar chart to show the count of passengers in each passenger class (Pclass).
    b. Sort the passenger classes in ascending order.
    c. Use different colors for the bars. Label the axes and add a title.
5. Create a Box Plot:
    a. Plot a box plot to show the age distribution within each passenger class (Pclass).
    b. Use Seaborn for easier creation of the box plot.
    c. Label the axes and add a title.
6. Create a Line Chart:
    a. Sort the DataFrame by Fare.
    b. Plot a line chart to show the fare against passenger ID (sorted by fare).
    c. Label the axes and add a title.

Submission

Upload the Python file and the screenshots as raw files for this activity.