## CHANDIGARH UNIVERSITY

## UNIVERSITY INSTITUTE OF ENGINEERING

**DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING**



|  |  |
| --- | --- |
| **Submitted By:Kanishk Soni Submitted To:Sudhanshu Sharma** | |
| **Subject Name** | Kanishk Soni |
| **Subject Code** | 20CSP-317 |
| **Branch** | BE-CSE |
| **Semester** | 5 |

**EXPERIMENT-2**

**Student’s Name: Kanishk Soni UID: 20BCS9398**

**Section/Group: 707\_WM\_B Subject Code: 20CSP-317**

**Subject Name: ML Lab Date of performance:28/8/2022**

**Branch: BE CSE Semester:5th**

**Aim:** To perform data visualization on the given data-set.

**Objective:** To do Data Visualization.

**Software/Hardware Requirements:** Windows 7 & above version.

**Tools to be used:**

1. Anaconda Jupyter Notebook,
2. Plotly, Seaborn, Matplotlib.

**Introduction**

Data and information visualization is an interdisciplinary field that deals with the graphic representation of data and information. It is a particularly efficient way of communicating when the data or information is numerous as for example a time series.

**Dataset**

The data has been split into two groups:

* training set (train.csv)
* test set (test.csv)

The training set should be used to build your machine learning models. For the training set, we provide the outcome (also known as the “ground truth”) for each passenger. The model will be based on “features” like passengers’ gender and class.

The titanic dataset we have taken in to account for the EDA. Train.csv will contain the details of a subset of the passengers on board (891 to be exact) and importantly, will reveal whether they survived or not, also known as the “ground truth”.

**Code:**

import pandas as pd

import seaborn as sns

import matplotlib.pyplot as plt

df = pd.read\_csv('titanic-train.csv')

df.head()

df.columns

plt.scatter(df['age'], df['fare'])

plt.title('Visualisation')

plt.show()

plt.bar(df['passenger\_id'], df['fare'], color = 'red')

plt.show()

plt.scatter(df['survived'], df['fare'], c = df['passenger\_id'])

plt.title('Scatter with colors')

plt.show()

plt.hist(df['fare'])

plt.title('Histogram of Fare')

plt.show()

plt.boxplot('passenger\_id', data=df)

plt.show()

sns.scatterplot(x = 'passenger\_id', y = 'fare', data = df)

plt.title('Scatter plot using Seaborn')

plt.show()

sns.scatterplot(x = 'passenger\_id', y = 'parch', data = df, hue = 'sex')

plt.title('Scatter plot using Seaborn')

plt.show()

sns.lineplot(x = 'age', y = 'fare', data = df, hue= 'sex')

plt.title('Lineplot using Seaborn')

plt.show()

sns.lineplot(data = df.drop(['passenger\_id'], axis = 1))

plt.show()

sns.barplot(x = 'parch', y = 'age', data = df, hue= 'sex')

plt.show()

sns.histplot(x = 'age', data = df, kde= True, hue= 'sex', fill = False)

import plotly.express as xpy

fig = xpy.scatter(df, x = 'age', y = 'fare', color = 'sex')

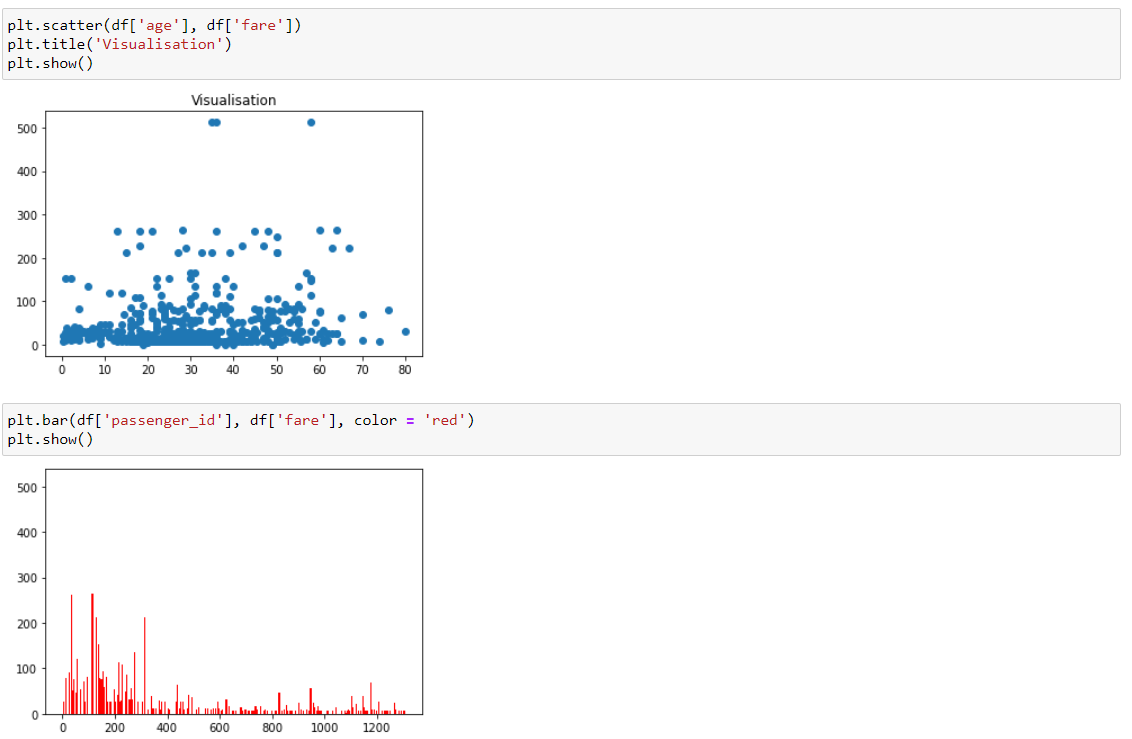
fig.show()

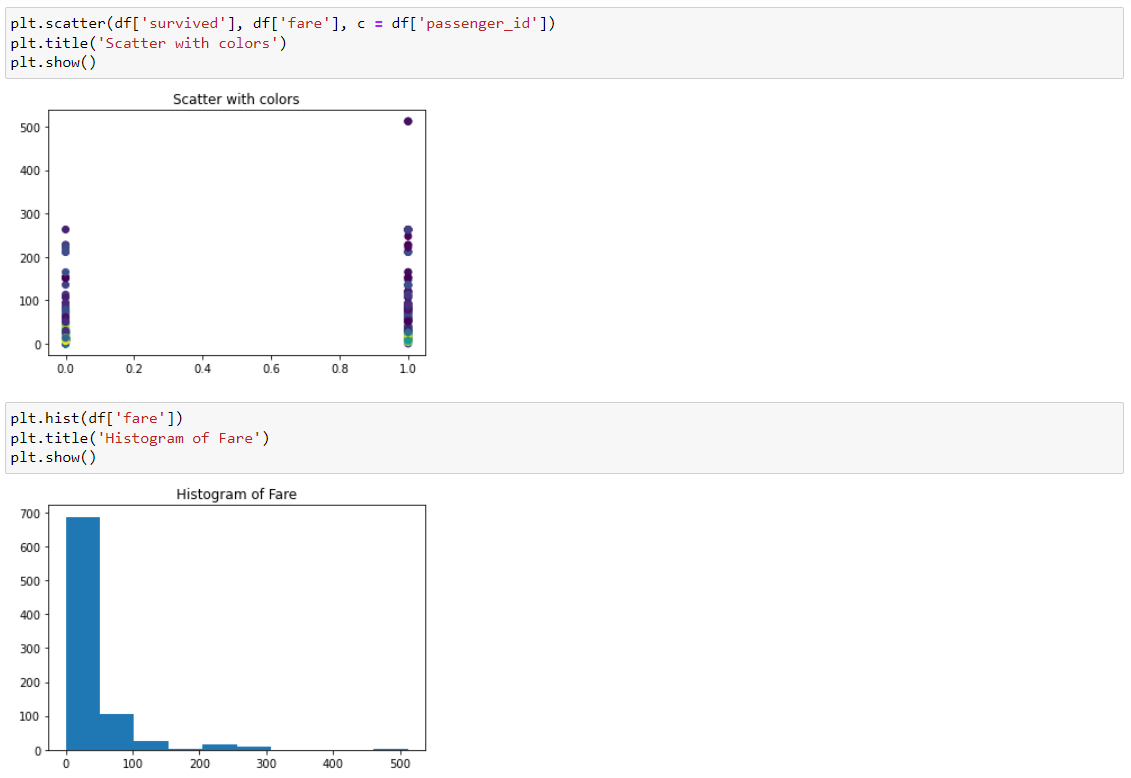
fig = xpy.line(data\_frame = df ,x = 'age', y = 'parch', title= 'Line plot using Plotly')

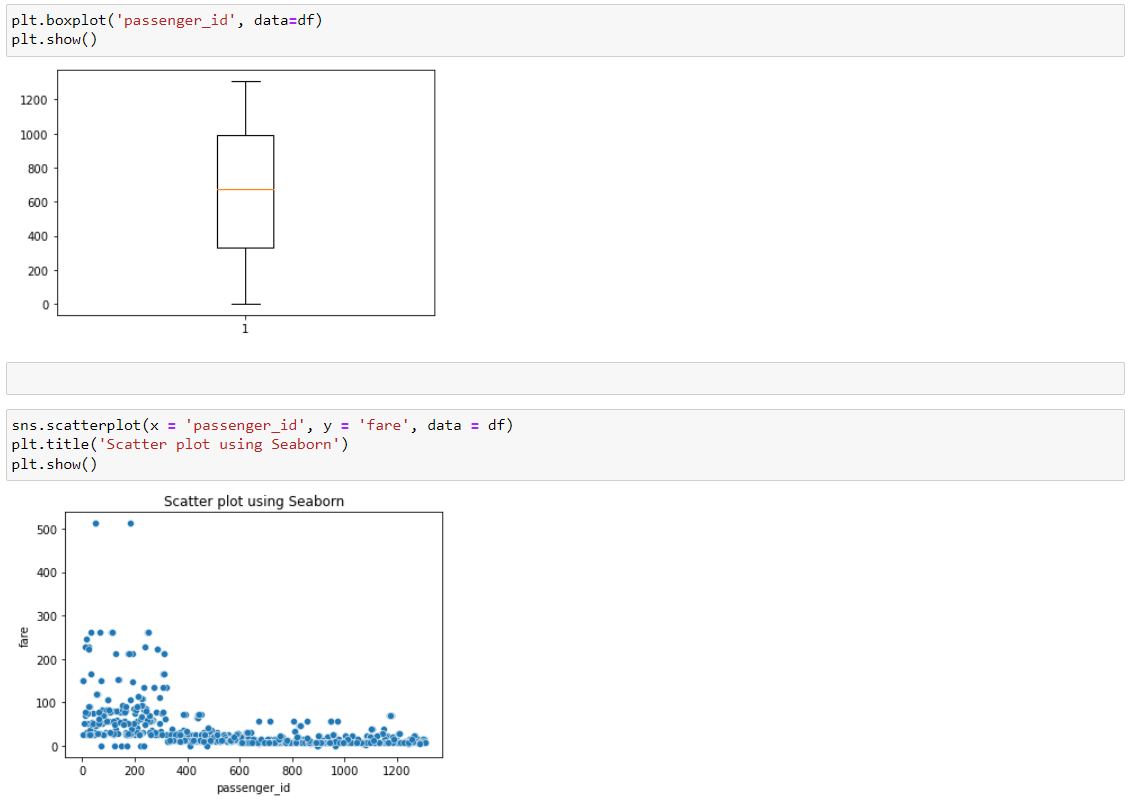
fig.show()

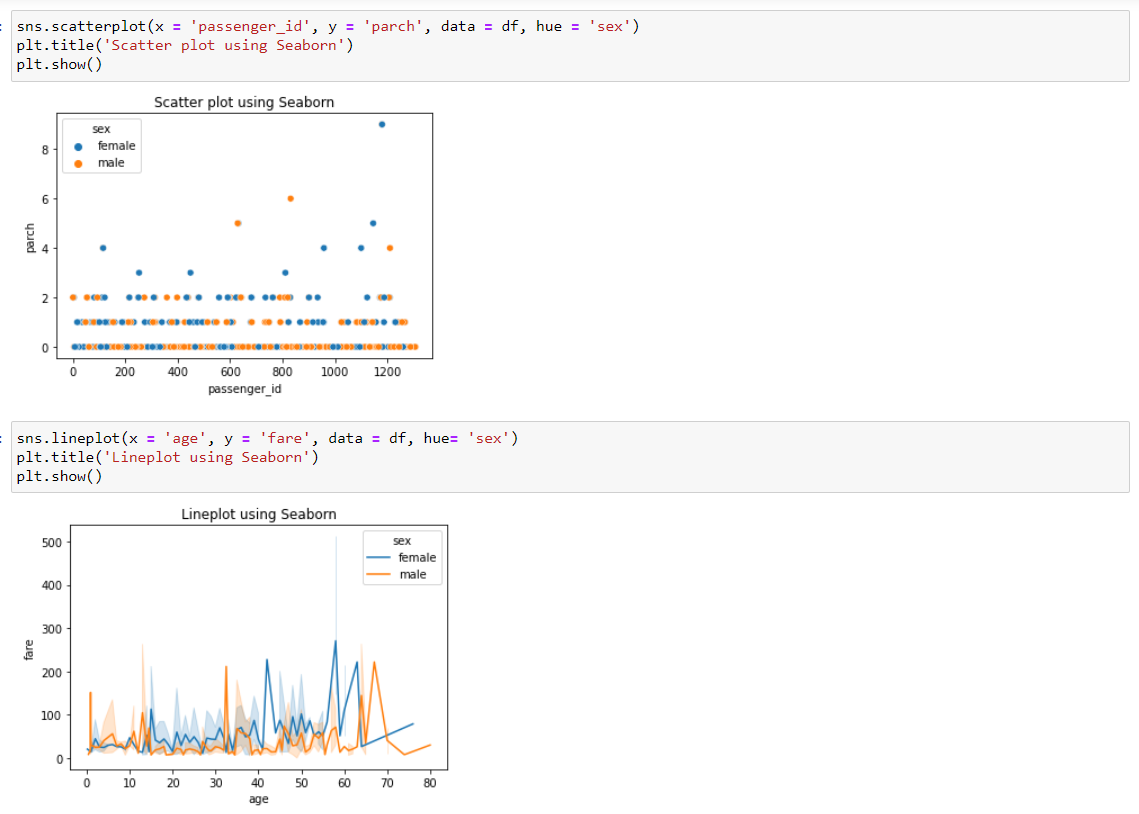
**Output:**

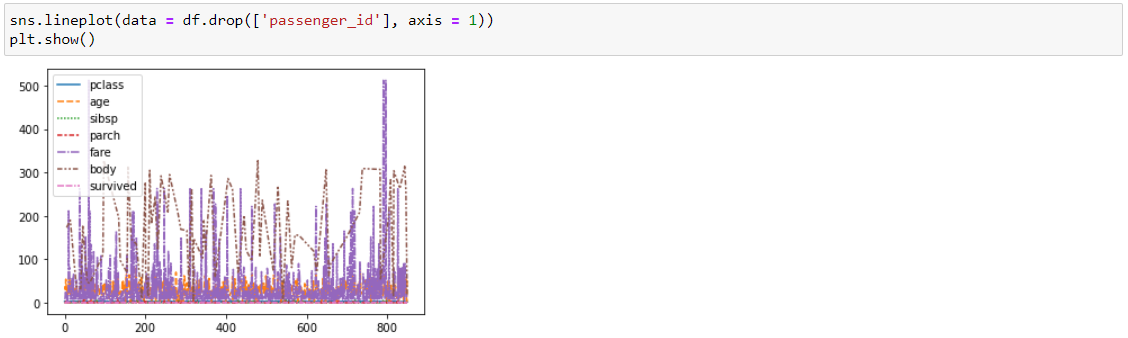




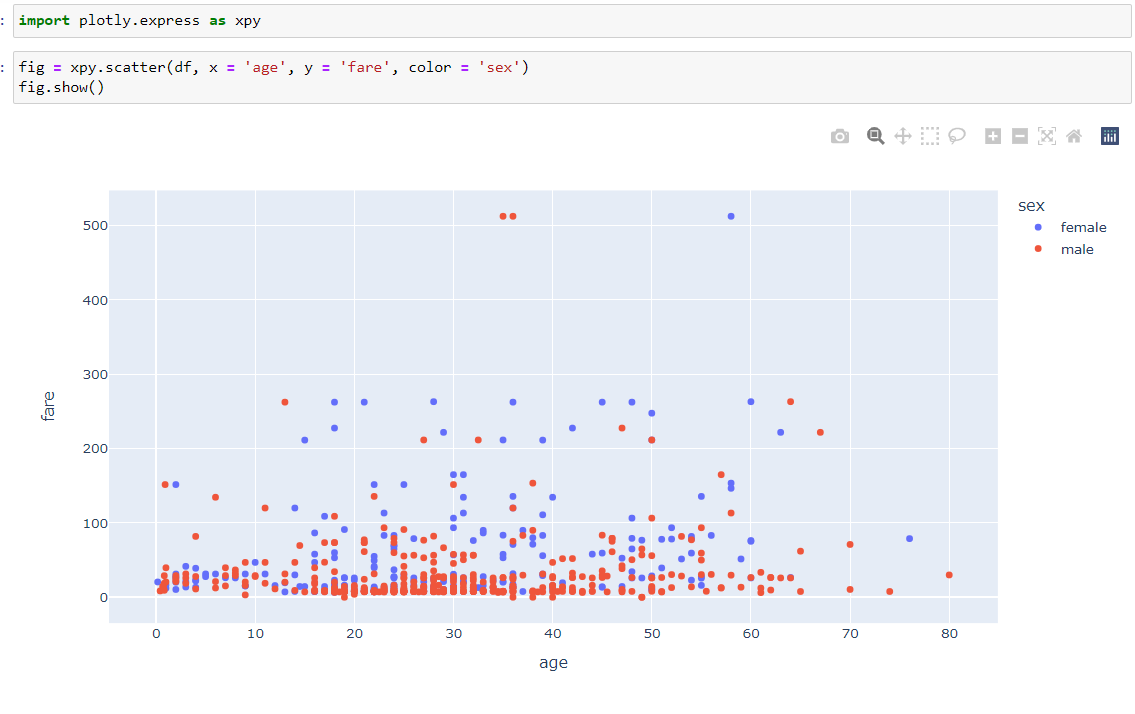


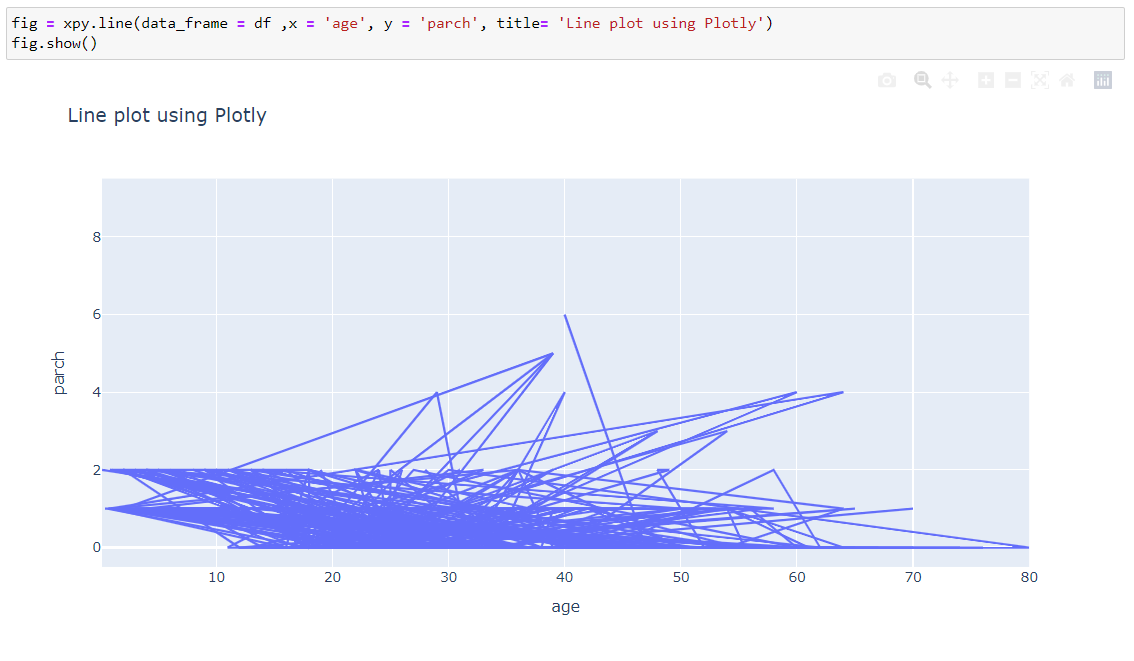












**Learning Outcomes:**

1. We learnt about how to analyze the data via Data Visualization on the dataset given to us.
2. We learnt different graph in matplotlib, plotly and seaborn.