Artificial Intelligence & Data Science

Assignment No: 8

Data Visualization I

1. Use the inbuilt dataset 'titanic'. The dataset contains 891 rows and contains information about the passengers who boarded the unfortunate Titanic ship. Use the 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.

Theory:-

Data visualization is the graphical representation of information and data. By using visual elements like charts, graphs, and maps, data visualization tools provide an accessible way to see and understand trends, outliers, and patterns in data.

In the world of Big Data, data visualization tools and technologies are essential to analyze massive amounts of information and make data-driven decisions.

The advantages and benefits of good data visualization:
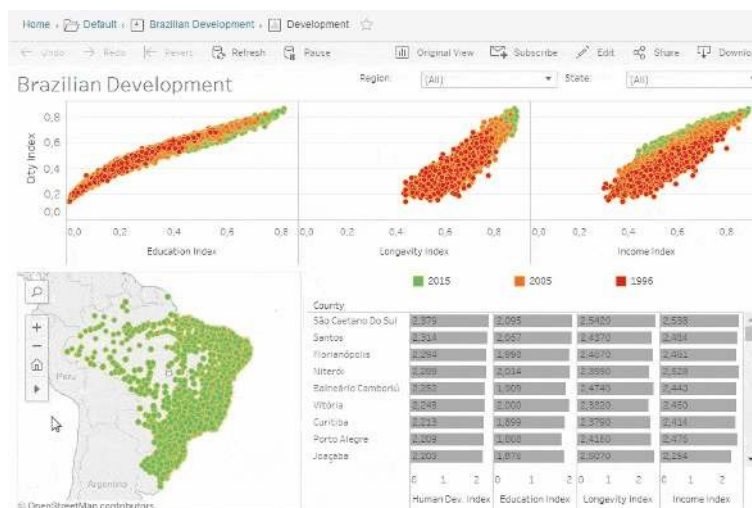
Our eyes are drawn to colors and patterns. We can quickly identify red from blue, square from circle. Our culture is visual, including everything from art and advertisements to TV and movies. Data visualization is another form of visual art that grabs our interest and keeps our eyes on the message. When we see a chart, we quickly see trends and outliers. If we can see something, we internalize it quickly. It's storytelling with a purpose. If you've ever

Artificial Intelligence & Data Science

stared at a massive spreadsheet of data and couldn't see a trend, you know how much more effective a visualization can be.

Big Data is here and we need to know what it says

As the "age of Big Data" kicks into high-gear, visualization is an increasingly key tool to make sense of the trillions of rows of data generated every day. Data visualization helps to tell stories by curating data into a form easier to understand, highlighting the trends and outliers. A good visualization tells a story, removing the noise from data and highlighting the useful information. However, it's not simply as easy as just dressing up a graph to make it look better or slapping on the "info" part of an infographic. Effective data visualization is a delicate balancing act between form and function. The plainest graph could be too boring to catch any notice or it make tell a powerful point; the most stunning visualization could utterly fail at conveying the right message or it could speak volumes. The data and the visuals need to work together, and there's an art to combining great analysis with great storytelling.

Algorithm:

Step 1: Download the data set of Titanic

(https://www.kaggle.com/uciml/iris)

Step 2: Importing Libraries

```
import pandas as pd
import numpy as np

import matplotlib.pyplot as plt
import seaborn as sns

dataset = sns.load_dataset('titanic')

dataset.head()
```

Step 3: Draw distributional plot

```
sns.distplot(dataset['fare'])
```

Step 4: Removal of Kernal Density line

```
sns.distplot(dataset['fare'], kde=False)
```

Step 4: Draw histogram

```
sns.distplot(dataset['fare'], kde=False, bins=10)
```

Step 5: The Joint Plot

      sns.jointplot(x='age', y='fare', data=dataset)

Step 6: sns.jointplot(x='age', y='fare', data=dataset, kind='hex')

Step 7: #The Pair Plot

      sns.pairplot(dataset)

Conclusion: Implemented successfully Simple Data visualization techniques using Python on Titanic dataset.