

Assignment No: 10

Problem Statement:- Data Visualization III

Download the Iris flower dataset or any other dataset into a DataFrame.(e.g.,<https://archive.ics.uci.edu/ml/datasets/Iris>). Scan the dataset and give the inference as:

List down the features and their types (e.g., numeric, nominal) available in the dataset.

Create a histogram for each feature in the dataset to illustrate the feature distributions.

Create a box plot for each feature in the dataset.

Compare distributions and identify outliers.

Theory:-

Data Visualization: Data visualization is the practice of translating information into a visual context, such as a map or graph, to make data easier for the human brain to understand and pull insights from. The main goal of data visualization is to make it easier to identify patterns, trends and outliers in large data sets. The term is often used interchangeably with others, including information graphics, information visualization and statistical graphics.

Data visualization is one of the steps of the data science process, which states that after data has been collected, processed and modeled, it must be visualized for conclusions to be made. Data visualization is also an element of the broader data presentation architecture (DPA) discipline, which aims to identify, locate, manipulate, format and deliver data in the most efficient way possible.

Examples of data visualization

In the early days of visualization, the most common visualization technique was using a Microsoft Excel spreadsheet to transform the information into a table, bar graph or pie chart. While these visualization methods are still commonly used, more intricate techniques are now available, including the following:

- infographics
- bubble clouds
- bullet graphs
- heat maps
- fever charts
- time series charts

Some other popular techniques are as follows.

Line charts: This is one of the most basic and common techniques used. Line charts display how variables can change over time.

Area charts: This visualization method is a variation of a line chart; it displays multiple values in a time series -- or a sequence of data collected at consecutive, equally spaced points in time.

Scatter plots: This technique displays the relationship between two variables. A scatter plot takes the form of an x- and y-axis with dots to represent data points.

Treemaps: This method shows hierarchical data in a nested format. The size of the rectangles used for each category is proportional to its percentage of the whole. Treemaps are best used when multiple categories are present, and the goal is to compare different parts of a whole.

Population pyramids: This technique uses a stacked bar graph to display the complex social narrative of a population. It is best used when trying to display the distribution of a population.

Common data visualization use cases

Common use cases for data visualization include the following:

Artificial Intelligence & Data Science

Sales and marketing: Research from the media agency Magna predicts that half of all global advertising dollars will be spent online by 2020. As a result, marketing teams must pay close attention to their sources of web traffic and how their web properties generate revenue. Data visualization makes it easy to see traffic trends over time as a result of marketing efforts.

Politics: A common use of data visualization in politics is a geographic map that displays the party each state or district voted for.

Healthcare: Healthcare professionals frequently use choropleth maps to visualize important health data. A choropleth map displays divided geographical areas or regions that are assigned a certain color in relation to a numeric variable. Choropleth maps allow professionals to see how a variable, such as the mortality rate of heart disease, changes across specific territories.

Scientists: Scientific visualization, sometimes referred to in shorthand as SciVis, allows scientists and researchers to gain greater insight from their experimental data than ever before.

Finance: Finance professionals must track the performance of their investment decisions when choosing to buy or sell an asset. Candlestick charts are used as trading tools and help finance professionals analyze price movements over time, displaying important information, such as securities, derivatives, currencies, stocks, bonds and commodities. By analyzing how the price has changed over time, data analysts and finance professionals can detect trends.

Logistics. Shipping companies can use visualization tools to determine the best global shipping routes.

Algorithm:-

Step 1: Download the data set of Iris Flower

Step 2: Importing Libraries

```
import numpy as np
import pandas as pd
```

Step 3: Reading the dataset

```
df = pd.read_csv("iris-flower-dataset.csv")
df
```

Step 4: Perform statistical analysis on data

```
df.mean()
df.median()
df.std()
df.min()
df.max()
df.describe()
```

Step 5: Print number of features and their datatypes

```
column = len(list(df))
column
df.info()
np.unique(df["species"])
```

Step 6: Data Visualization-Create a histogram for each feature in the dataset to illustrate the feature distributions. Plot each histogram.

```
import seaborn as sns
import matplotlib
import matplotlib.pyplot as plt
```

Artificial Intelligence & Data Science

```
%matplotlib inline

fig, axes = plt.subplots(2, 2, figsize=(16, 8))

axes[0,0].set_title("Distribution of First Column")
axes[0,0].hist(df["sepal_length"]);

axes[0,1].set_title("Distribution of Second Column")
axes[0,1].hist(df["sepal_width"]);

axes[1,0].set_title("Distribution of Third Column")
axes[1,0].hist(df["petal_length"]);

axes[1,1].set_title("Distribution of Fourth Column")
axes[1,1].hist(df["petal_width"]);
```

Step 7: Create a boxplot for each feature in the dataset. All of the boxplots should be combined into a single plot. Compare distributions and identify outliers.

```
data_to_plot = [df["sepal_length"],df["sepal_width"],df["petal_length"],df["petal_width"]]

sns.set_style("whitegrid")
# Creating a figure instance
fig = plt.figure(1, figsize=(12,8))

# Creating an axes instance
ax = fig.add_subplot(111)

# Creating the boxplot
bp = ax.boxplot(data_to_plot);
```

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