22BCS034

22BCS039

22BCS064

22BCS065

22BCS066

22BCS308

21BCS086

**Lab 3: Data Charts – Univariate Analysis**

**Prelab Questions**

1. **Define univariate analysis and give an example.**

* Univariate analysis focuses on analyzing and interpreting a single variable.  
  Example: Analyzing the distribution of exam scores in a class.

1. **Why are histograms preferred for visualizing continuous data?**

* Histograms effectively display the frequency distribution of continuous data, showcasing patterns and distribution shapes.

1. **List the key differences between bar charts and histograms.**

* **Bar Charts**: Display categorical data; bars are separated.
* **Histograms**: Display continuous data; bars are adjacent.

1. **How can outliers affect univariate visualizations?**

* Skew the visualization, distorting the data representation.
* Obscure patterns in the main dataset.

1. **What are the benefits of adding a kernel density estimate (KDE) to histograms?**

* KDE provides a smooth curve representing the data’s distribution, making it easier to identify patterns and compare distributions.

**In-Lab Details**

**Objective**:

* Visualize and interpret single-variable distributions using histograms and KDE plots.

**Resources**:

* Python (Jupyter Notebook).
* Libraries: Matplotlib, Seaborn.
* Dataset: performance.csv with columns for scores and departments.

**Program Code**:

Code: [*https://colab.research.google.com/drive/1gxTPNP3VGj4TDwyD5LxRwhiDNbUMjUdP?usp=sharing*](https://colab.research.google.com/drive/1gxTPNP3VGj4TDwyD5LxRwhiDNbUMjUdP?usp=sharing)

**Expected Output**:

* Histogram showing score distribution.
* KDE overlay for smoother visualization.

**Postlab Questions**

1. **How does increasing the number of bins in a histogram affect the visualization?**

* Increasing bins provides more detail but may introduce noise. Fewer bins simplify the visualization but may obscure patterns.

1. **What does the area under the KDE curve represent?**

* The total area under the KDE curve equals 1, representing the entire dataset’s probability distribution.

1. **Compare the benefits of using KDE over histograms for data visualization.**

* KDE provides a smoother, continuous representation of the data.
* Avoids arbitrary bin selection issues inherent in histograms.

1. **Why is it important to consider the scale of the x-axis when plotting univariate data?**

* The x-axis scale affects interpretation by influencing the granularity and comparability of visualizations.

**5.Suggest scenarios where bar charts are more suitable than histograms.**

* Bar charts are suitable for categorical data analysis, such as comparing sales by region or product categories.