Training Report – Day 9

Topic Covered Today:

- Introduction to **Descriptive Statistics** in Python
- Mean, Median, and Mode
- Standard Deviation
- Scatter Plot Visualization
- Percentiles

Key Learning:

Mean, Median, Mode:

- **Mean (Average):** The sum of all values divided by the total number of values.
  Example: For [10, 20, 30, 40], mean = (10+20+30+40)/4 = 25.
- **Median:** The middle value when the dataset is arranged in order.
  Example: For [5, 10, 15, 20, 25], median = 15.
- **Mode:** The most frequently occurring value in a dataset.
  Example: For [2, 4, 4, 6, 8], mode = 4.

Standard Deviation:

Standard deviation measures the **spread or variation** of data from the mean.

- Low standard deviation → data is close to mean.
- High standard deviation → data is spread out.

Formula:
[
\sigma = \sqrt{\frac{\sum (x_i - \mu)^2}{N}}
]

In Python, we can calculate using `numpy.std()`.

Scatter Plot:

A scatter plot is used to **visualize the relationship between two variables**. Each point on the plot represents a pair of values.

In Python (using matplotlib):

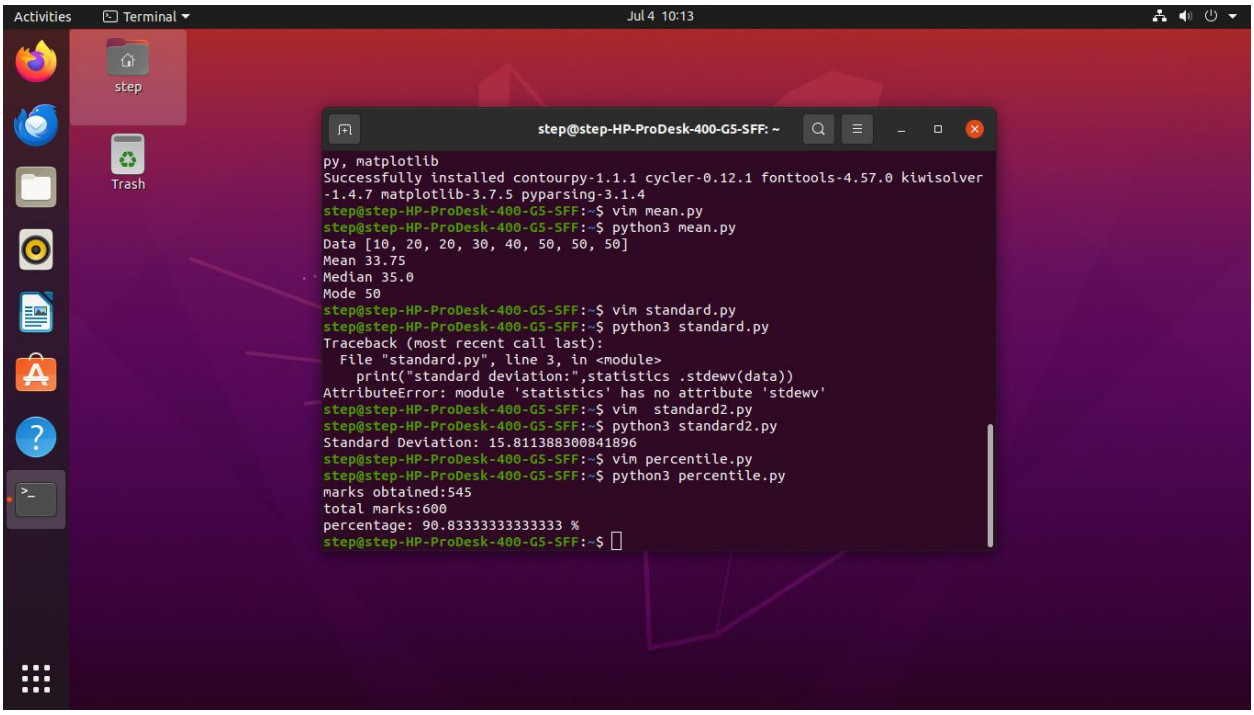
```
import matplotlib.pyplot as plt

x = [5, 7, 8, 7, 6, 9, 5, 6]
y = [99, 86, 87, 88, 100, 86, 103, 87]

plt.scatter(x, y)
plt.title("Scatter Plot Example")
plt.xlabel("X values")
plt.ylabel("Y values")
plt.show()
```

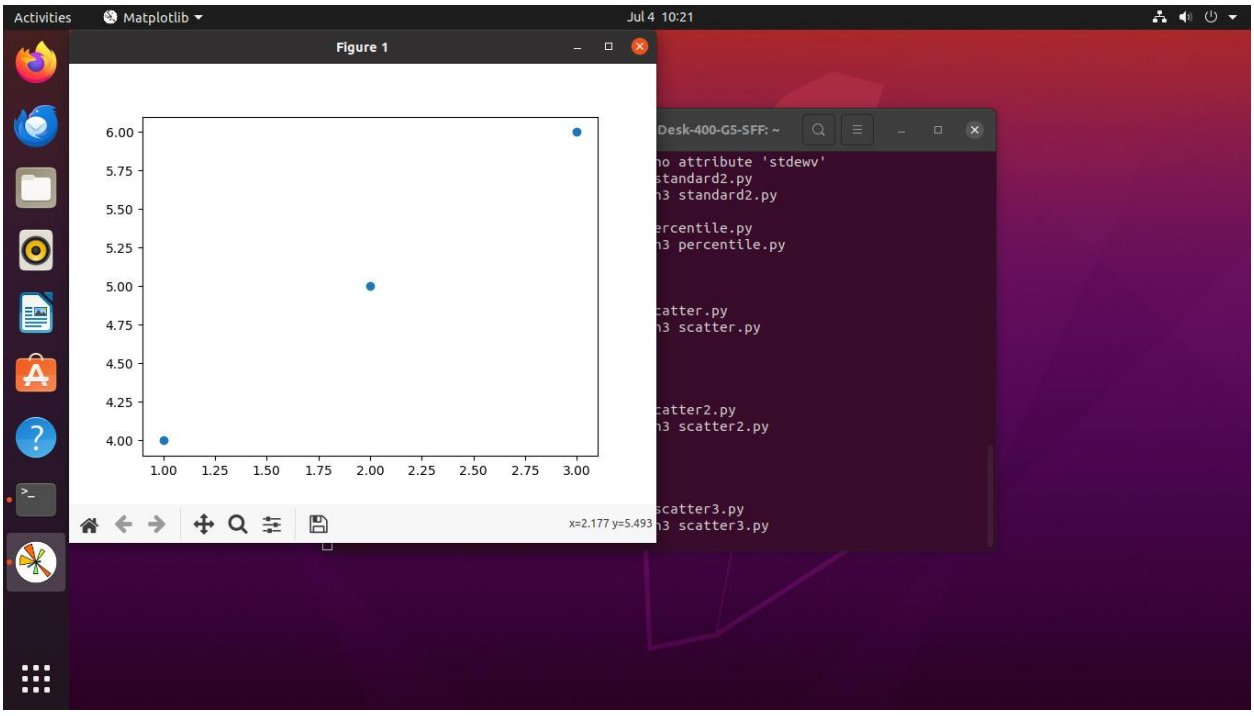
This shows how variables correlate with each other.

Percentile:

A percentile indicates the **value below which a given percentage of observations fall**.

- Example: 50th percentile = median.
- 90th percentile means 90% of values lie below that point.
- In Python:
- `import numpy as np`
- `data = [1, 2, 3, 4, 5, 6, 7, 8, 9, 10]`
- `np.percentile(data, 90)  # gives the 90th percentile`





Activities / Assignments:

- Practiced calculating mean, median, mode using **NumPy** and **statistics** library.
- Calculated standard deviation of a dataset in Python.
- Plotted a **scatter plot** between two variables using Matplotlib.

- Found 25th, 50th, and 75th percentiles of a dataset.
- Prepared notes on the use of statistics in analyzing datasets before applying ML models.

Personal Reflection for Day 9:

Today's session helped me understand the role of **statistics in data science and machine learning**. I realized that mean, median, and mode provide the central tendency of data, while standard deviation shows how spread out the data is.

Visualizing data with scatter plots made the concepts much clearer because I could see patterns and relationships between variables. Learning about percentiles showed me how data can be divided and analyzed for insights.

Overall, this session gave me strong fundamentals in statistics, which are essential for understanding datasets before building AI/ML models.