

WEEK 4 LECTURE NOTE ACTIVITY 1

Objective: Explore pairwise relationships and their correlation visually.

1. **Steps: Dataset Preparation:** Provide the students with a dataset having at least two numerical variables (e.g., advertising budget vs. sales).
2. **Python Script:** Load the dataset using pandas.

Calculate the correlation between two selected variables.

Use matplotlib or seaborn to create a scatter plot, overlaying the correlation value as text.

3. **Discussion:** Plot the scatter plot agrees with the correlation value.

Solution

```
import pandas as pd
import matplotlib.pyplot as plt

# Load dataset
data = pd.read_csv("sales_data.csv")

# Select two variables
x = data["Advertising"]
y = data["Sales"]

# Calculate correlation
correlation = x.corr(y)
print(f"Correlation: {correlation:.2f}")

# Scatter plot
plt.scatter(x, y, alpha=0.7)
plt.title(f"Scatter Plot (Correlation: {correlation:.2f})")
```

```
plt.xlabel("Advertising Budget")
```

```
plt.ylabel("Sales")
```

```
plt.show()
```