**High Impact Skills Development Program**

**in Artificial Intelligence, Data Science, and Blockchain**

**Module 3: Data Visualization**

Lab 5: Advanced Data Visualization Using Python and ChatGPT

Instructor: Dr. Arham Muslim

Assistant Professor, SEECS, NUST

# Objective:

In this lab, you will learn to create and customize visualizations using Python, focusing on Matplotlib and advanced visualizations with the assistance of ChatGPT. The session is designed to progressively increase in difficulty, starting with basic plotting and customization, and moving towards generating and modifying advanced visualizations such as a boxplot, heatmap, and Sankey diagram using ChatGPT

# Materials Needed:

* Python and any preferred IDE
* ChatGPT

# Dataset:

* Superstore Dataset

# Lab Work:

## Task 1: Creating a Basic Bar Plot with Matplotlib

* Load and Inspect the Superstore Dataset
  + Begin by loading the Superstore dataset into a Pandas DataFrame.
  + Inspect the first few rows of the dataset using the ‘head()’ method to familiarize yourself with the structure and the available columns.
  + Code Example:

import pandas as pd

superstore\_data = pd.read\_csv('superstore\_dataset.csv')

print(superstore\_data.head())

* Plotting a Basic Bar Plot
  + Using Matplotlib, create a bar plot to visualize **total sales for each product category**.
  + Begin by grouping the data by the 'Category' column and summing the 'Sales' column.
  + Use this aggregated data to generate a bar plot that displays the sales for each category.
  + Code Example:

import matplotlib.pyplot as plt

category\_sales = superstore\_data.groupby('Category')['Sales'].sum()

category\_sales.plot(kind='bar')

plt.title('Total Sales by Category')

plt.xlabel('Category')

plt.ylabel('Sales')

plt.show()

* Customizing the Bar Plot
  + Enhance the bar plot by customizing its appearance.
  + Change the color of the bars, add data labels on top of each bar, and customize the plot's title and axes labels to make the visualization more informative and visually appealing.
  + Code Example:

plt.bar(category\_sales.index, category\_sales.values, color='skyblue')

plt.title('Total Sales by Category', fontsize=16, fontweight='bold')

plt.xlabel('Category', fontsize=12)

plt.ylabel('Sales ($)', fontsize=12)

plt.xticks(rotation=45)

for i, value in enumerate(category\_sales.values):

    plt.text(i, value + 500, f'{value:.2f}', ha='center')

plt.show()

## Task 2: Creating and Customizing Boxplots with Matplotlib

* Introduction to Boxplots
  + Learn the concept of boxplots, which are used to summarize the distribution of a dataset.
  + create a boxplot to visualize the **distribution of sales across different product sub-categories**.
  + Code Example:

sub\_category\_sales = superstore\_data.groupby('Sub-Category')['Sales'].sum()

plt.figure(figsize=(12, 8))

plt.boxplot(sub\_category\_sales.values, vert=False, patch\_artist=True)

plt.title('Sales Distribution by Sub-Category')

plt.xlabel('Sales')

plt.yticks([1], ['Sub-Categories'])

plt.show()

* Customizing the Boxplot
  + Customize the boxplot by adjusting its aesthetics: Add colors, change the orientation, and include statistical annotations like the mean and outliers.
  + Modify the boxplot's appearance to make it more informative.
  + Code Example:

plt.figure(figsize=(12, 8))

plt.boxplot(sub\_category\_sales.values, vert=False, patch\_artist=True,

            boxprops=dict(facecolor='lightblue', color='darkblue'),

            whiskerprops=dict(color='darkblue'),

            capprops=dict(color='darkblue'),

            flierprops=dict(marker='o', color='darkblue', alpha=0.5))

plt.title('Sales Distribution by Sub-Category', fontsize=16, fontweight='bold')

plt.xlabel('Sales ($)', fontsize=12)

plt.yticks([1], ['Sub-Categories'])

plt.show()

## Task 3: Generating a Heatmap using ChatGPT

* Generate Heatmap Code Using ChatGPT
  + Use ChatGPT to generate basic Python code for creating a heatmap of **sales data across different regions and categories**.
  + Copy the generated code into your Python environment and run it to produce a basic heatmap.
* Customizing the Heatmap
  + Customize the heatmap generated by ChatGPT.
  + Modify the color palette, add annotations, and adjust the layout for better readability.
  + Experiment with different options to improve the heatmap's visual impact.

## Task 4: Creating a Boxplot with ChatGPT Assistance

* Generate a Boxplot Code Using ChatGPT
  + Request ChatGPT to generate code for creating a boxplot that visualizes the **distribution of profits across different ship modes**.
  + Implement the generated code to produce the boxplot.
* Customizing the Boxplot
  + Customize the boxplot by modifying the color scheme, adding notches, and tweaking the axis labels and title.
  + Adjust the plot's aesthetics to make the visual representation more insightful.

## Task 5: Creating a Sankey Diagram using ChatGPT Assistance

* Generate Basic Sankey Diagram Code Using ChatGPT
  + Use ChatGPT to generate code for creating a Sankey diagram that visualizes the **flow of products from different categories to their respective regions**.
  + Implement the generated code in your Python environment.
* Customizing the Sankey Diagram
  + Enhance the Sankey diagram by customizing the node colors, labels, and flow values.
  + Experiment with different design options to improve clarity and visual appeal.
  + Adjust the layout to better fit the dataset and make the flow between categories and regions more evident.

# Additional Resources for Self-Learning:

* [Introduction to Plotting with Matplotlib in Python](https://www.datacamp.com/tutorial/matplotlib-tutorial-python)
* [Create a Bar Chart in Python using Matplotlib](https://datatofish.com/bar-chart-python-matplotlib/)
* [Box Plot in Python using Matplotlib](https://www.geeksforgeeks.org/box-plot-in-python-using-matplotlib/)
* [How to create Heatmaps using matplotlib.pyplot](https://medium.com/@SeanAT19/how-to-create-heatmaps-using-matplotlib-pyplot-db9ff94d25e8)
* [Sankey diagram in Python](https://plotly.com/python/sankey-diagram/)