Data Analysis Project - Visualization with Matplotlib and Seaborn

# (Coded by M.Essa)

# Introduction

This project demonstrates how to analyze a discrete dataset of student scores using Python. We will perform the following data visualizations using Matplotlib and Seaborn: Bar Plot, Pie Chart, and Dot Plot. The dataset includes the scores of students Essa, Danish, Amber, Asim, and Rida in three subjects: Math, Science, and English.

# Dataset

The dataset consists of the following student names and their scores in three subjects. This data is represented in a tabular format.

data = {  
 'Student': ['Essa', 'Danish', 'Amber', 'Asim', 'Rida'],  
 'Math': [85, 78, 92, 70, 88],  
 'Science': [90, 82, 85, 72, 91],  
 'English': [78, 75, 89, 83, 77]  
}

The data is then converted into a DataFrame using the pandas library, which allows for easy data manipulation and analysis.

# Bar Plot: Comparing Scores Across Subjects

A bar plot is used to visualize and compare the performance of each student across the subjects: Math, Science, and English. The following code creates a bar plot.

df.set\_index('Student').plot(kind='bar', figsize=(10, 6))  
plt.title('Student Scores in Different Subjects')  
plt.ylabel('Scores')  
plt.xticks(rotation=45)  
plt.show()

# Pie Chart: Proportion of Total Scores by Subject

A pie chart is used to display the proportion of total scores for each subject. This helps in understanding which subject contributes more to the overall scores.

subject\_totals = df[['Math', 'Science', 'English']].sum()  
plt.figure(figsize=(7, 7))  
plt.pie(subject\_totals, labels=subject\_totals.index, autopct='%1.1f%%', startangle=140)  
plt.title('Total Scores Proportion by Subject')  
plt.show()

# Dot Plot: Individual Scores by Subject

The dot plot visualizes individual student scores across the three subjects. It provides a clear view of how each student performed in each subject.

melted\_df = df.melt(id\_vars='Student', var\_name='Subject', value\_name='Score')  
plt.figure(figsize=(10, 6))  
sns.stripplot(data=melted\_df, x='Student', y='Score', hue='Subject', jitter=True, size=8, palette='Set1')  
plt.title('Dot Plot of Scores by Student and Subject')  
plt.ylabel('Scores')  
plt.show()