Title: Data Science Assignment 3 Data Visualization
Student Name: Minahil Irfan

Overview

Data visualization is the graphical representation of information and data. It helps identify patterns, trends, and insights that may not be obvious in raw numbers.

Key Libraries:

- **Matplotlib:** Used for basic visualizations such as bar charts, histograms, and scatter plots.
- **Seaborn:** Built on top of Matplotlib, providing more attractive and informative statistical plots.

Common Plots:

- **Bar Chart:** Compares categories or groups.
- **Histogram:** Shows the distribution of a numerical variable.
- **Scatter Plot:** Displays relationships between two variables.
- **Box Plot:** Highlights data spread and outliers.
- **Heatmap:** Displays correlations using color gradients.

Importance:
Visualization allows better storytelling with data. It converts raw data into meaningful insights, helping in decision-making and hypothesis formation.

Task Overview

The goal of this assignment was to create five different plots from the cleaned dataset, describe the insights gained from each, and complete the first stage of Exploratory Data Analysis (EDA).

Activity

Step 1: Load Cleaned Dataset

- Imported libraries: *pandas*, *matplotlib*, and *seaborn*.
- Loaded the cleaned dataset twitter_sentiment_cleaned.csv.
- Mapped sentiment labels (0 = Negative, 4 = Positive) for easier readability.

Step 2: Visualization and Insights

Plot 1: Sentiment Distribution

- Created a bar chart showing the count of positive and negative tweets.
- **Insight:** Positive tweets slightly outnumber negative ones, indicating a generally optimistic sentiment trend.

Plot 2: Tweet Length Distribution

- Used a histogram to visualize tweet lengths.
- **Insight:** Most tweets fall between 40 and 120 characters, showing that users prefer concise expressions.

Plot 3: Most Active Users

- Generated a bar plot of the top 10 users by tweet count.
- **Insight:** A few users contribute significantly more tweets, suggesting high engagement or automated posting behavior.

Plot 4: Tweet Length by Sentiment

- Created a box plot comparing tweet lengths between positive and negative sentiments.
- **Insight:** Negative tweets tend to be slightly longer, possibly reflecting more detailed expressions of opinion.

Plot 5: Common Words in Positive vs Negative Tweets

- Used word frequency counts to display the most common words for each sentiment.
- **Insight:** Positive tweets often include uplifting or casual terms, while negative tweets contain more critical or emotional words.

Step 3: Project Milestone

- Completed the first Exploratory Data Analysis (EDA) through visualization and pattern discovery.