Activity 1: Research & Implement Key Statistical Measures Using the English Premier League Dataset

---

Step 1: Quick Research (10 Minutes)

Task:

1. Research and write down the formulas for the following statistical measures:
    - Mean (Average)
    - Median
    - Mode
    - Range
    - Variance
    - Standard Deviation

Step 2: Implement in Python (20-25 Minutes)

Dataset: Use the English Premier League dataset. Focus on the following columns:

- **GF (Goals For)** – Goals scored by the team.
- **GA (Goals Against)** – Goals conceded by the team.
- **xG (Expected Goals)** – Metric that predicts the likelihood of scoring based on shot quality.
- **Poss (Possession %)**
- **Attendance** – Number of spectators.
- **Sh (Shots Taken)** – Total shots taken by the team.

Task:

1. Manually implement the formulas for **mean, median, mode, range, variance, and standard deviation in Python**, without using built-in Pandas methods. Calculate these measures for the columns above (or any of your choosing).
2. Compare results for at least two different teams or matches in terms of:
    - **Goals For (GF)**
    - **Expected Goals (xG)**
    - **Shots Taken (Sh)**
    - **Possession (Poss)**

3. Visualize the data using Matplotlib or Seaborn (e.g., **histograms, boxplots**).

4. Find correlations between variables, such as:

    - Does high xG (Expected Goals) always lead to more Goals For (GF)?
    - Is there a correlation between Possession and Goals For?

Step 3: Analysis & Discussion (15 Minutes)

Discussion Questions:

1. Does high possession always lead to more Goals For (GF)?
    - Consider teams with high possession but low goal output. Discuss why possession doesn't always correlate with goals.
2. How do outliers affect statistics like mean and variance?
    - Explore how unusually high or low values in metrics like Goals For (GF) or Shots Taken (Sh) can impact the mean and variance of a team's performance.