WEEK 9 – STATISTICS FOR DATA ANALYTICS ASSIGNMENT

Data Analytics Training Program

TASK OVERVIEW

In this week's assignment, you will work with **Sampling Techniques, Central Limit Theorem (CLT), Normal Distribution, and Z-Score**.
You are required to apply statistical concepts using **Python / Pandas** and interpret the results clearly.

You may use the **Sales / Superstore dataset** or any dataset used during the sessions.

TASK 1 : POPULATION & SAMPLING

Perform the following:

- Load the dataset using Pandas
- Treat the full dataset as **Population**
- Create a **Sample dataset** using:
    - Simple Random Sampling
- Display:
    - Population size
    - Sample size

TASK 2 : SAMPLING TECHNIQUES

Using the same dataset:

- Create:
    - Random Sample
    - Systematic Sample (every nth record)
- Compare:
    - Mean of population
    - Mean of samples
- Write a short note on:
    - Difference between population and sample results

TASK 3 : CENTRAL LIMIT THEOREM (CLT)

Perform the following steps:

- Take multiple samples of size **30**
- Calculate the **mean of each sample**
- Store sample means
- Plot the distribution of sample means
- Observe and explain:
    - How the distribution behaves
    - Relation to CLT

TASK 4 : NORMAL DISTRIBUTION ANALYSIS

Using any numerical column (Sales / Marks / Profit):

- Calculate:
    - Mean
    - Standard Deviation
- Plot:
    - Histogram of the data
- Check whether the data approximately follows a normal distribution
- Apply the **68–95–99.7 rule** and explain the result

TASK 5 : Z-SCORE CALCULATION

Perform the following:

- Calculate Z-Score for each value in the selected column
- Add a new column named Z_Score
- Identify:
    - Values with $Z > 3$ or $Z < -3$
- Mark them as **outliers**

TASK 6 : BUSINESS INSIGHTS

Answer the following questions in simple words:

- Why is sampling required in real-world data analysis?

- How does CLT help in analytics?
- Why is normal distribution important before hypothesis testing?
- How does Z-Score help in identifying unusual values?

SUBMISSION INSTRUCTIONS

1. Save your work as:
    - Jupyter Notebook (.ipynb) **or**
    - Python file (.py)
2. File naming format:
    - YourName_Week9_Statistics.ipynb
3. Create a **GitHub repository**
4. Upload:
    - Code file
    - Any plots generated
5. Share the **GitHub repository link** in the WhatsApp group

---

DEADLINE

📅 **Sunday – 11:00 PM**
