Week 5 Assignment – Statistics & Analytical Techniques

Data Analytics Training Program – NeoSkillz DTAA

Introduction

This assignment will test your understanding of the core **data analytics fundamentals** covered in **Week 5**, including:

✔ Descriptive Statistics
✔ Probability & Distributions
✔ Correlation Analysis
✔ Regression Analysis (Prediction)
✔ Hypothesis Testing (Decision Making)

You are required to analyze a dataset using these techniques and present findings in a structured format.

Dataset

Use **any one dataset** (or use your own) such as:

File Name	Purpose
StudentScores.xlsx	Education Analytics
Employee_Salary.xlsx	HR Analytics
Sales_Marketing.xlsx	Business Analytics
Health_Data.xlsx	Medical Analysis

Dataset should contain at least 2 numerical columns and 1 categorical column.

Tasks to Perform

1. Descriptive Statistics

Perform statistical analysis on your dataset:

- Calculate: **Mean, Median, Mode** (any one numerical column)

- Find: **Range & Standard Deviation**

- Create a **Frequency Distribution Table**

- Plot **Histogram** of any numerical data

➤ **Purpose** – Understand central tendency & spread of data.

---

2. Probability Analysis

Choose any **categorical column** (e.g., Region / Department / Class):

- Find probability of each category (P = Frequency / Total)
- Write minimum **3 probability-based questions** from your data
- Mention **Theoretical vs Experimental Probability**

➤ **Purpose** – Understand uncertainty & chance-based analysis.

---

3. Correlation Analysis

Identify **relationship between two numerical columns**:

- Use Excel / Python / Power BI to find the **Correlation Coefficient (r)**
- Create a **Scatter Plot**
- Interpret result:
    - Strong / Weak
    - Positive / Negative / No correlation

➤ **Purpose** – Understand how variables relate.

---

4. Regression Analysis (Prediction)

Perform **Simple Linear Regression** using Excel / Python:

[
$$y = mx + c$$
]

- Choose one **Independent Variable** (X)
- Choose one **Dependent Variable** (Y)
- Predict value for a **new input**
- Plot **Regression Line**

➤ **Purpose** – Use data for forecasting.

---

5. Hypothesis Testing (Decision Making)

Choose any one test:

Test	Use Case
T-Test	Compare two groups (e.g., Male vs Female salary)
Paired T-Test	Before & after improvement
Chi-Square Test	Categorical comparison
ANOVA	Compare 3+ groups
A/B Testing	Marketing / Website test

Steps:

1. Write **Null Hypothesis (H_0)** & **Alternative Hypothesis (H_1)**
2. Apply statistical test
3. Calculate **p-value**
4. Decision:
   - If **$p < 0.05$ → Reject H_0 (Significant difference)**
   - If **$p \geq 0.05$ → Fail to reject H_0 (No evidence)**

➤ **Purpose** – Make business decisions using data.

Submission Format

Prepare a **report in PDF/Word** with the following structure:

Section	Description
Title Page	Name, Dataset, Date
Dataset Description	What data is used & why
Task 1 – Statistics	Table + Graph + Explanation
Task 2 – Probability	Questions & Calculations
Task 3 – Correlation	r value + Scatter Plot
Task 4 – Regression	Prediction + Graph
Task 5 – Hypothesis Test	H_0, H_1, p-value, Conclusion
Final Insights	Findings & Recommendations
Learning Summary	What did you learn in Week 5