Task – Diabetes Prediction Analysis

Dataset

You'll use a **real dataset** of patient health data for diabetes prediction. With features like:

- **Age**, **Gender**, **BMI**, **Hypertension**, **HbA1c**, **Blood Glucose**, **Hypertension**, **Heart Disease**, **Smoking History**, and a **Diabetes Status** (Positive/Negative) indicator https://www.kaggle.com/datasets/iammustafatz/diabetes-prediction-dataset.

This dataset is perfect for practicing both exploratory analysis and statistical testing.

Task Steps

1. Data Loading & Initial Exploration

- Load the dataset (CSV) into a Pandas DataFrame.
- Display the first few rows, data types, and summary statistics using `.describe()` and `.dtypes`.
- Check and handle any missing values (if present).

2. Descriptive Analysis (NumPy, Pandas, Seaborn)

- Compute average values for **BMI**, **Hypertension**, **Glucose**, and **HbA1c**.
- Compute and interpret the **correlation matrix** among these health indicators.
- Group the data by **Diabetes Status** and compare the mean values of these variables.
- Visualizations using Seaborn:
    - Histogram of **BMI** distributions.
    - Boxplot of **Glucose** levels grouped by **Diabetes Status**.
    - Scatter plot with regression line: **HbA1c vs Blood Glucose** (`sns.regplot`).
    - Heatmap showing the correlation matrix.

3. Hypothesis Testing

1. **Z-Test**
    - **Claim:** The average **BMI** in this population = **25**.
    - Conduct a **one-sample Z-test** to assess if the sample diverges significantly from this claim.
2. **T-Test**
    - **Question:** Is the average **age** different between diabetic-positive and diabetic-negative groups?
    - Use an **independent two-sample t-test** and interpret the result.
3. **Chi-Square Test**
    - **Question:** Is **Smoking History** related to **Diabetes Status**?
    - Create a contingency table and perform a **Chi-Square test of independence**.

4. Summary & Insights

Write **5–7 bullet points** summarizing your findings:

- Which indicators differ most between diabetic and non-diabetic groups?
- What were the results of the Z-test, T-test, and Chi-Square test?
- What insights might healthcare professionals draw from your analysis?