Project Insights & Key Findings

Based on statistical analysis, hypothesis testing, and visual exploration performed on the healthcare dataset.

1. Smoking Status and Diabetes

- The **chi-square test of independence** indicated a statistically significant association between smoking habits and diabetes prevalence.
- Smokers and former smokers showed a **higher proportion of diabetes cases** compared to non-smokers.
- This suggests that smoking behavior may be a contributing risk factor for diabetes.

> **Insight:** *Smoking status is not independent of diabetes occurrence, supporting rejection of the null hypothesis.*

2. BMI Differences Between Diabetic and Non-Diabetic Individuals

- The **independent t-test** revealed a significant difference in mean BMI between diabetic and non-diabetic groups.
- Diabetic individuals generally had a **higher average BMI**.
- Box plots visually confirmed a higher median BMI among diabetic patients.

> **Insight:** *Higher BMI is strongly associated with diabetes, highlighting obesity as a major health risk factor.*

3. Age Distribution and Confidence Interval Interpretation

- The confidence interval for age provided a reliable estimate of the population mean age.
- The relatively narrow interval suggests **low variability and good sampling stability**.
- Age distribution appeared approximately normal, validating the use of parametric tests.

> **Insight:** *The dataset provides a statistically reliable age representation for inference.*

4. Age Groups and Glucose Levels (ANOVA)

- The **ANOVA test** showed significant differences in glucose levels across age groups.
- Older age groups exhibited **higher median glucose levels**.
- Box plots supported the statistical findings by showing upward trends with age.

> **Insight:** *Glucose levels vary significantly across age groups, indicating age as an important factor in diabetes risk.*

5. Covariance Between Age and BMI

- Covariance analysis showed a **positive value**, indicating that age and BMI tend to increase together.
- While covariance confirmed direction, it did not quantify strength.

> **Insight:** *As age increases, BMI tends to increase, suggesting lifestyle and metabolic changes over time.*

6. Correlation Between Age and BMI

- Pearson correlation analysis revealed a **moderate positive correlation** between age and BMI.
- Scatter plots with regression lines showed a clear upward trend.

> **Insight:** *Age and BMI are positively correlated, reinforcing the role of aging in weight-related health risks.*

7. Visualization-Driven Understanding

- Graphical analysis (box plots, scatter plots, heatmaps) made statistical relationships easier to interpret.
- Visuals consistently supported numerical test results.
- Heatmaps highlighted strong correlations among metabolic indicators such as glucose, BMI, and cholesterol.

> **Insight:** *Combining statistical tests with visualization improves clarity and confidence in conclusions.*

8. Overall Statistical Interpretation

- Most hypothesis tests resulted in **rejection of the null hypothesis**, indicating meaningful relationships between variables.
- The dataset exhibits realistic health patterns commonly observed in medical studies.
- The use of multiple statistical methods ensured robust and reliable conclusions.

> **Final Insight:** *The project demonstrates that statistical analysis can effectively uncover hidden patterns in healthcare data, supporting data-driven decision-making and risk assessment.*