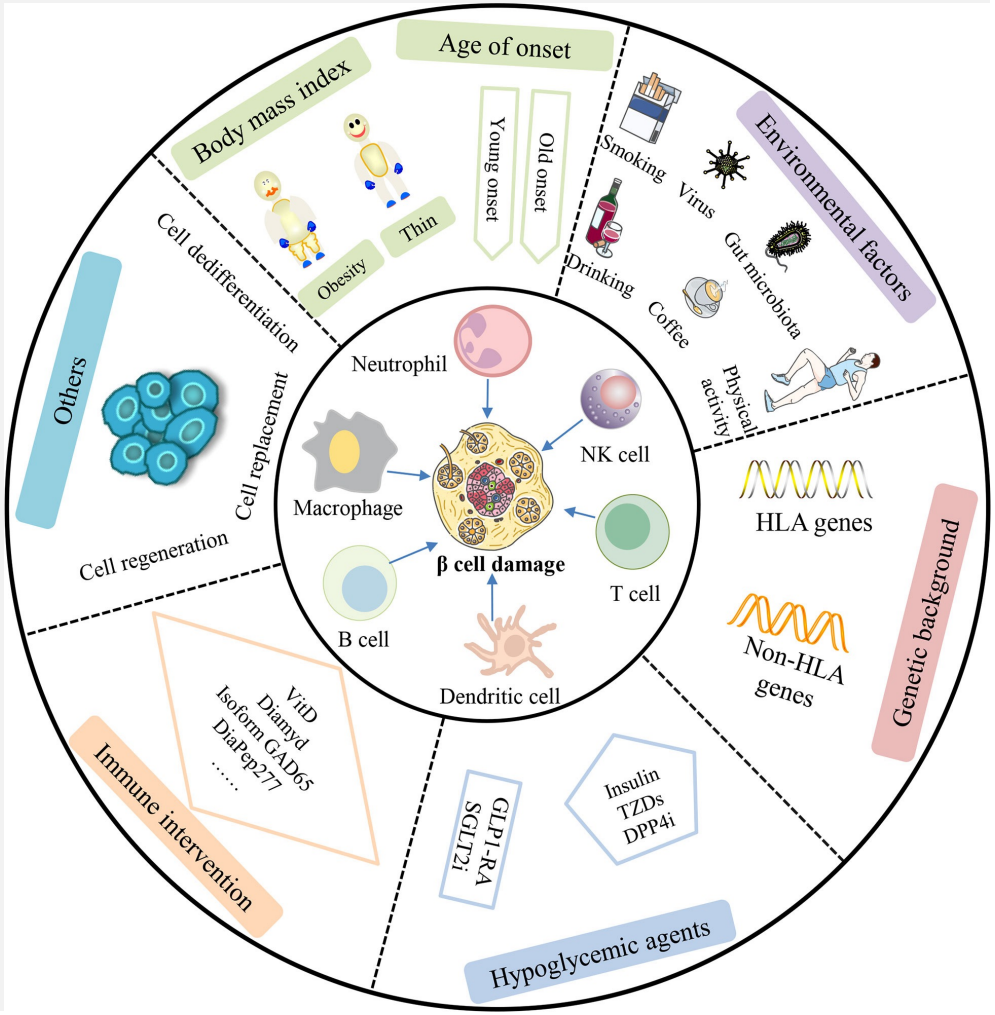
ANALYSIS OF DIABETES RISK FACTOR USING BRFSS 2015 DATA

Presenter: Md Arman Kaishar

Date: September 11, 2024

INTRODUCTION

- **Diabetes** is a global health issue affecting millions.
- Leads to serious complications like:
  - Heart disease
  - Vision loss
  - Kidney disease
- The significant **economic burden** on healthcare systems.
- Importance of early diagnosis and lifestyle interventions.

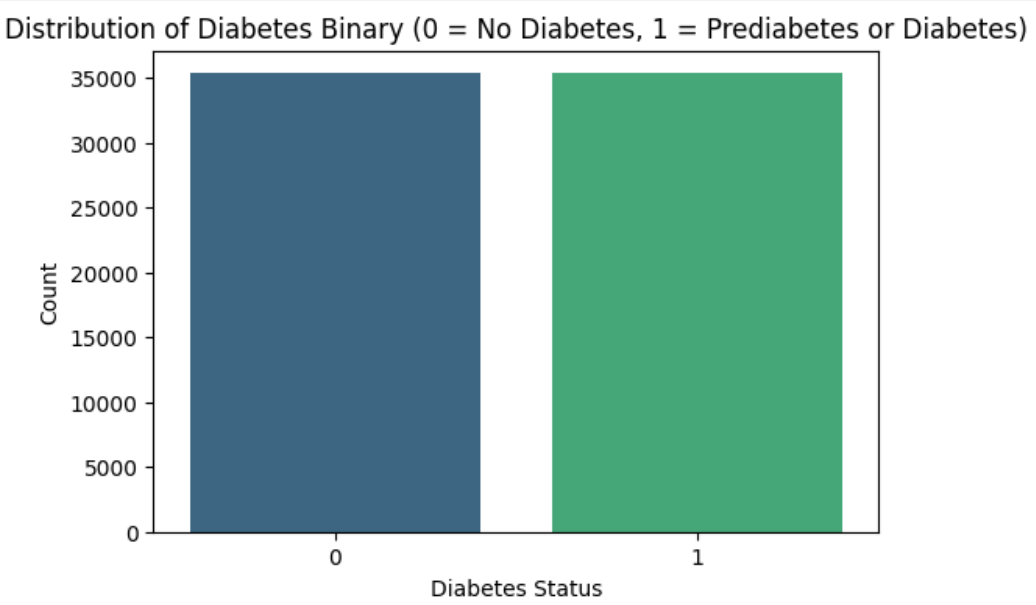


OBJECTIVES

- **Primary Goal**: Identify key risk factors predictive of diabetes.
- **Perform Exploratory Data Analysis (EDA)**: Use EDA techniques to visualize data and understand the relationships between different health and demographic variables and diabetes.
- Provide insights to aid in prevention and public health strategies.

DATA OVERVIEW

- **Source**: Behavioral Risk Factor Surveillance System (BRFSS) 2015.
- **Sample Size**: 70,692 respondents.
- **Balance**: 50% with no diabetes, 50% with prediabetes or diabetes.
- **Features**: 21 health and demographic variables.
- **Target Variable**: Diabetes_binary (0 = No diabetes, 1 = Prediabetes or diabetes).



KEY VARIABLES

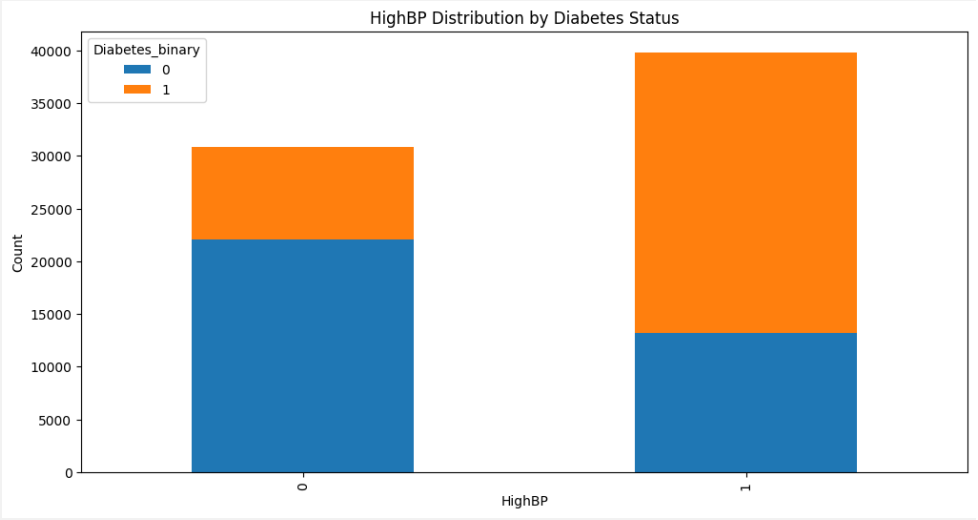
- **Health Conditions**: High Blood Pressure (HighBP), High Cholesterol (HighChol), Stroke History (Stroke), Heart Disease (HeartDiseaseorAttack)
- **Lifestyle Factors**: Body Mass Index (BMI), Physical Activity (PhysActivity), Smoking Status (Smoker), Diet (Fruit and Vegetable Intake)
- **Demographics**: Age, Education Level, Income

DATA ANALYSIS APPROACH

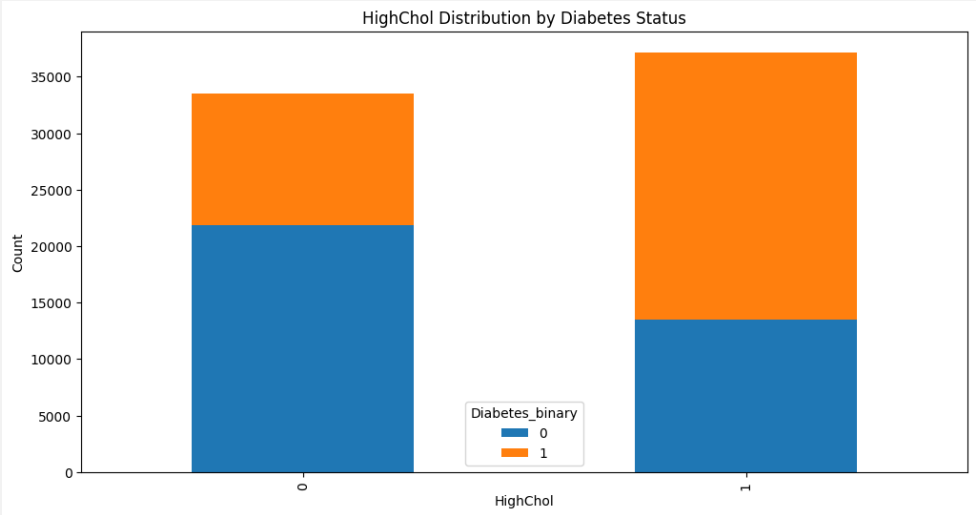
- **Univariate Analysis**: Examine individual feature distributions.
- **Correlation Analysis**: Identify strong predictors.
- **Predictive Modeling**:
  - Logistic Regression
  - Decision Trees

UNIVARIATE ANALYSIS FINDINGS

- High Blood Pressure (**HighBP**): A larger portion of individuals with diabetes have high blood pressure compared to those without diabetes.



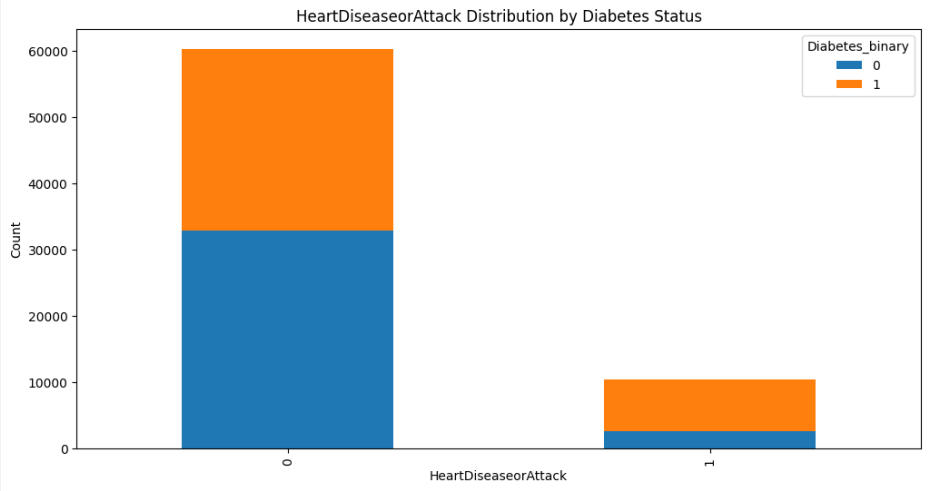
- High Cholesterol (**HighChol**): A higher proportion of individuals with diabetes have high cholesterol.

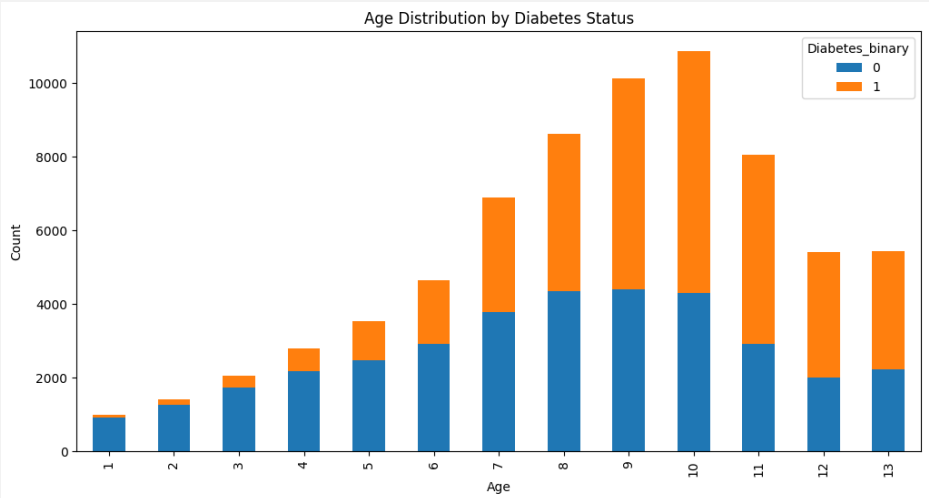


UNIVARIATE ANALYSIS FINDINGS

- **Heart Disease or Attack**: A larger proportion of individuals with diabetes have a history of heart disease or heart attacks.

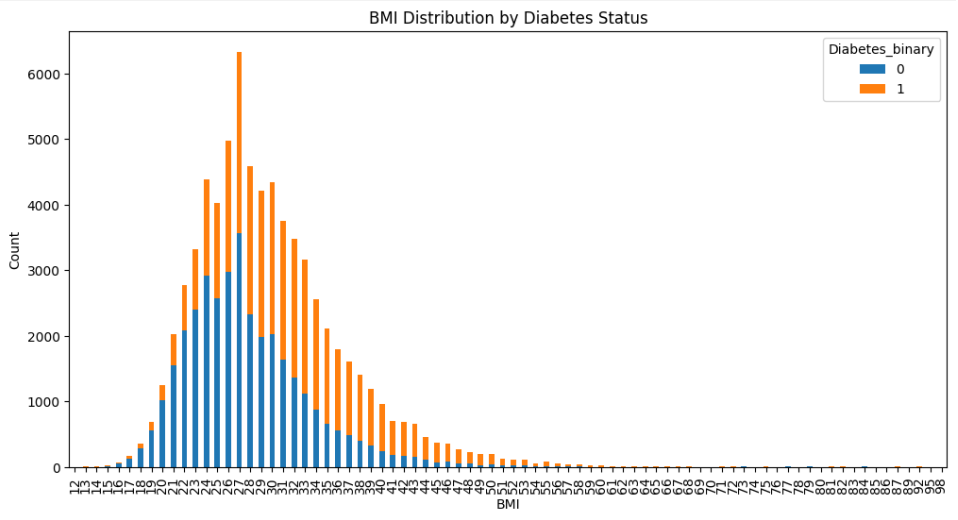
- **Age**: Older age groups might have a higher proportion of individuals with diabetes.



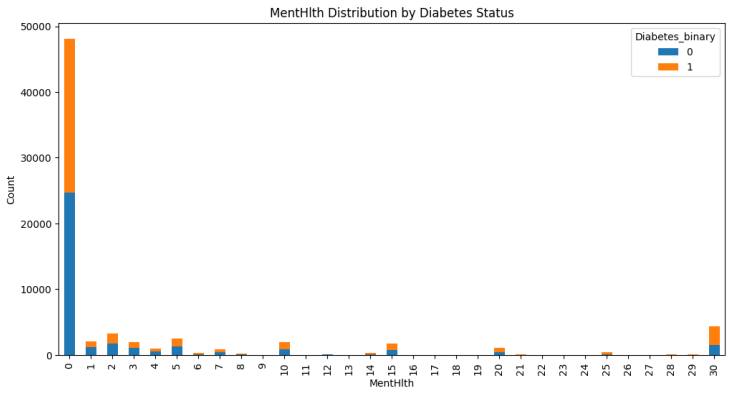


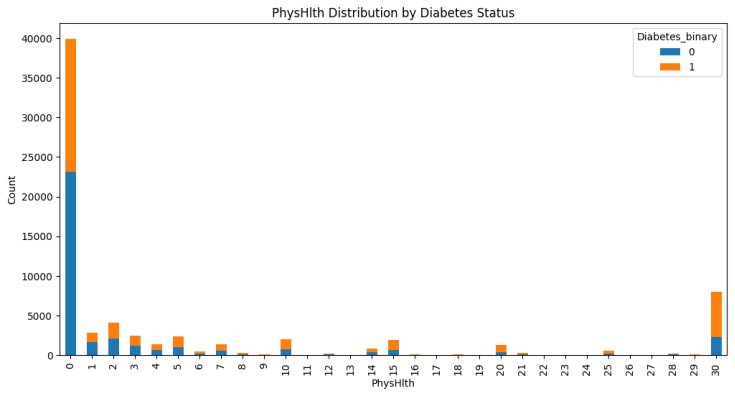
UNIVARIATE ANALYSIS FINDINGS

- The Body Mass Index (**BMI**): Higher BMI values are associated with a higher prevalence of diabetes.



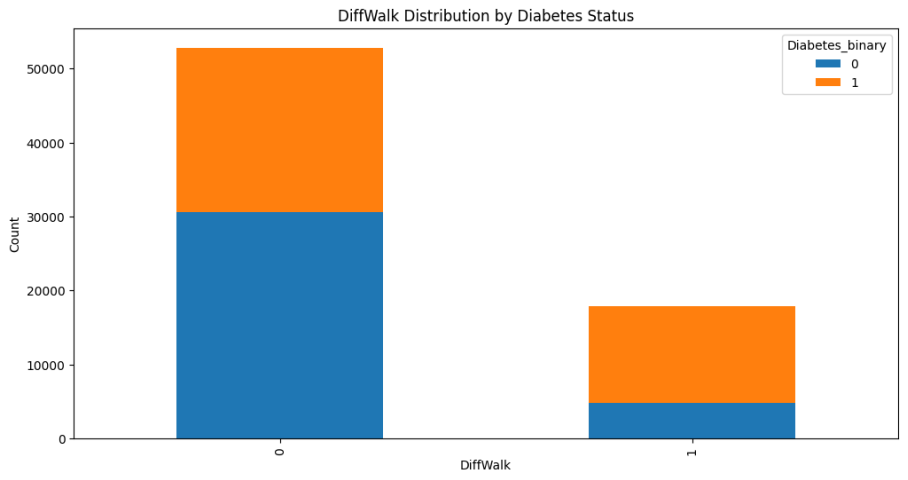
- **MentHlth** (Mental Health) & **PhysHlth** (Physical Health): A higher proportion of individuals with poor mental and physical health report diabetes.



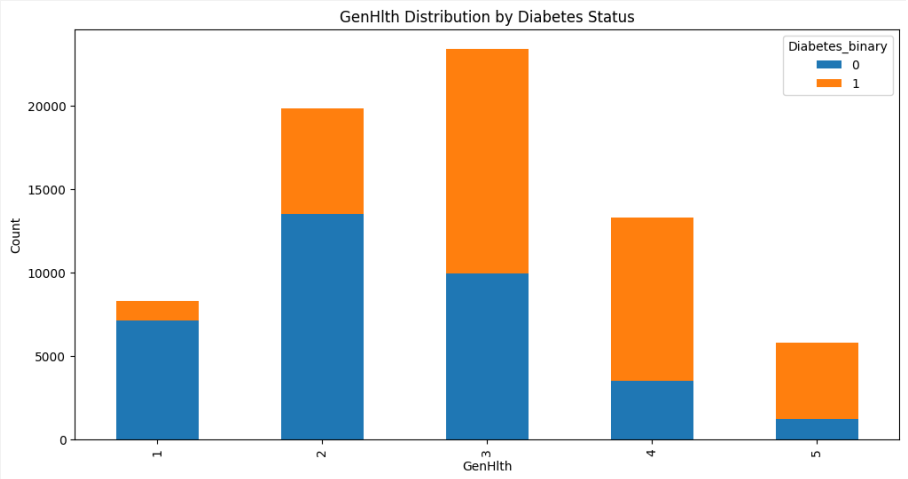


UNIVARIATE ANALYSIS FINDINGS

- **DiffWalk** (Difficulty Walking): A larger proportion of diabetic individuals may report serious difficulty walking, compared to non-diabetic individuals.

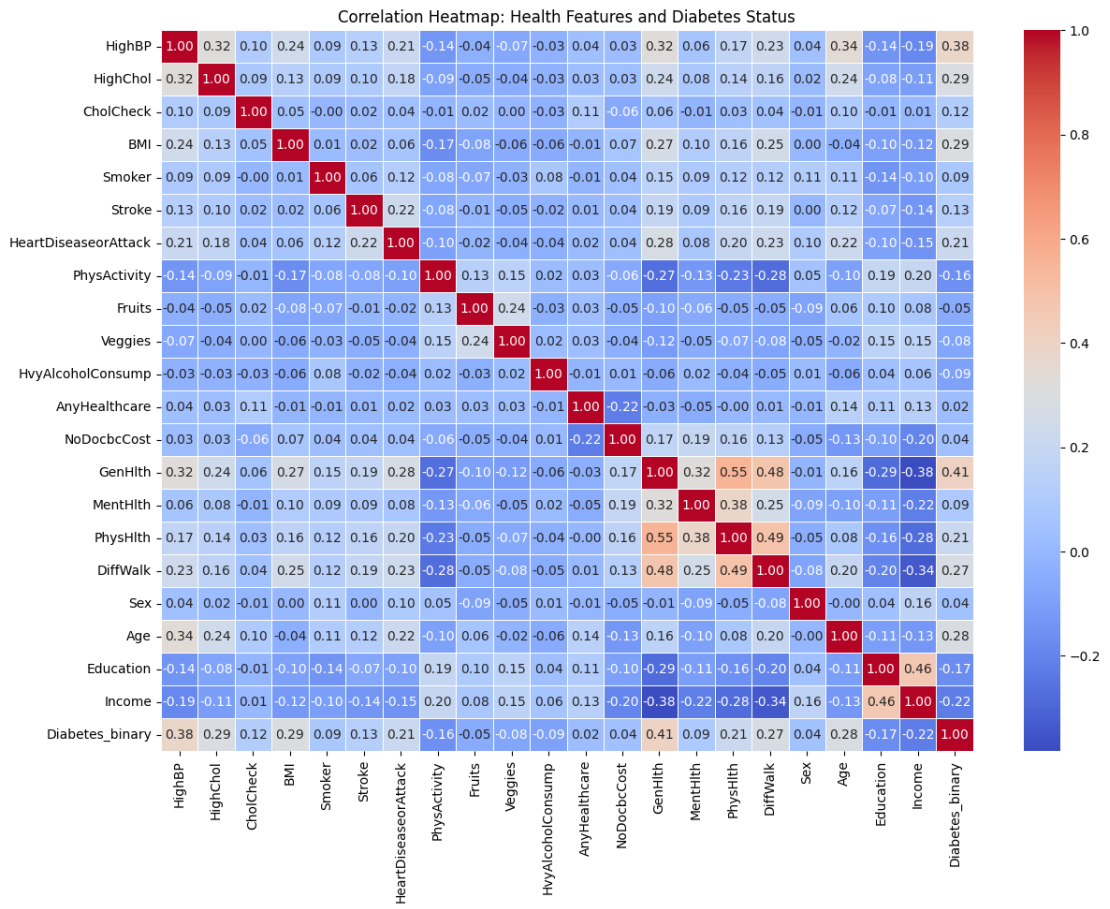


- **GenHlth** (General Health): Individuals with poor general health are more likely to have diabetes.



CORRELATION ANALYSIS

- **Strong Positive Correlations** with diabetes:
  - Body Mass Index (BMI)
  - High Blood Pressure (HighBP)
  - Poor General Health (GenHlth)
  - High Cholesterol (HighChol)
- **Negative Correlation**:
  - Physical Activity (PhysActivity)
- **Interpretation**:
  - Indicates that obesity and hypertension are significant risk factors.
  - Regular physical activity is protective against diabetes.



LOGISTIC REGRESSION RESULTS

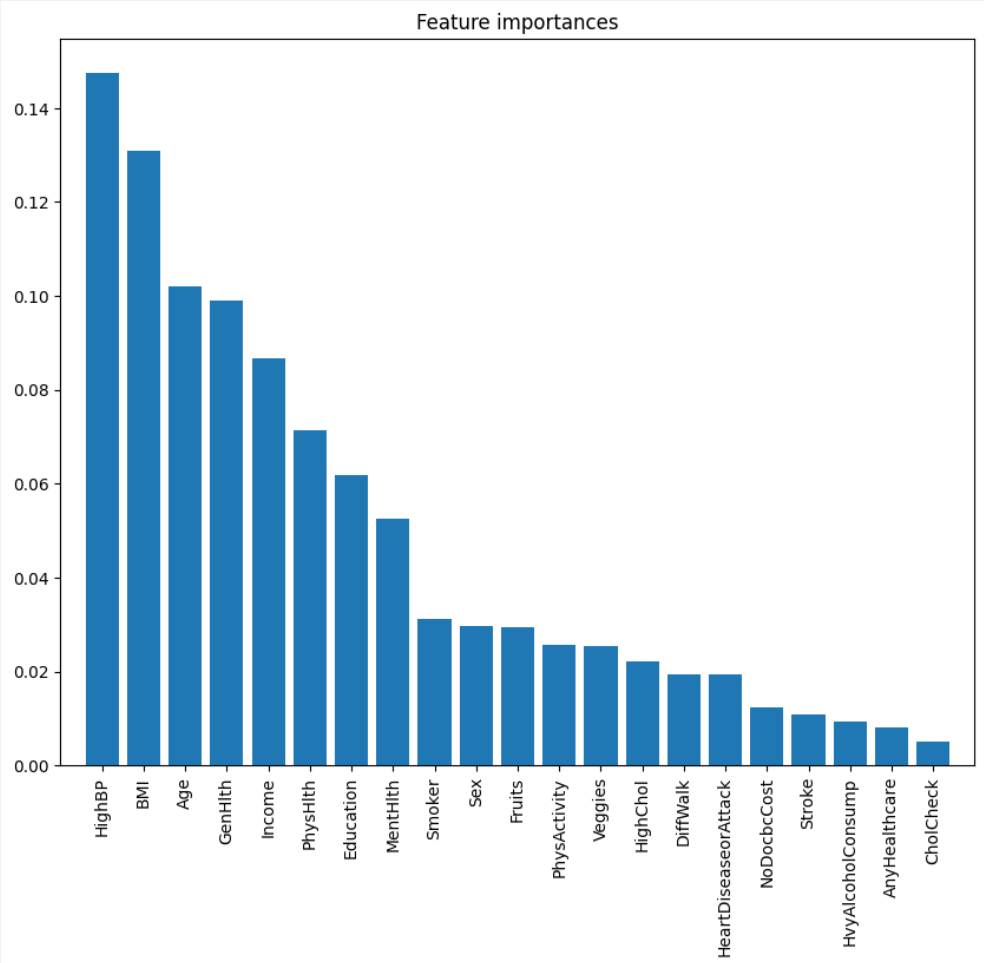
- **Key Predictors Identified**:
  - High BMI
  - Poor General Health
  - High Blood Pressure
  - High Cholesterol
- **Model Performance**:
  - Accuracy: 74.9 %
- **Implication**:
  - This confirms the importance of these factors in predicting diabetes risk.

```
Accuracy: 0.7494341757827235
                  Feature  Importance
2               CholCheck    1.266101
0                  HighBP    0.730723
1                HighChol    0.588805
13                GenHlth    0.586278
17                    Sex    0.254799
6   HeartDiseaseorAttack    0.254429
5                 Stroke    0.200057
18                   Age    0.151503
16              DiffWalk    0.118226
3                    BMI    0.075434
11         AnyHealthcare    0.048238
12           NoDocbcCost    0.015399
4                 Smoker    0.006707
7           PhysActivity   -0.003170
14              MentHlth   -0.004524
15              PhysHlth   -0.008167
19             Education   -0.026339
8                 Fruits   -0.051924
20                Income   -0.058910
9                Veggies   -0.093262
10    HvyAlcoholConsump   -0.731033
```

DECISION TREE FINDINGS

- **Top Features**:
  - Body Mass Index (BMI)
  - General Health (GenHlth)
  - Age
  - Physical Health (PhysHlth)

- **Decision Rules**:
  - High BP, Higher BMI, and poor general health significantly increase diabetes risk.



INSIGHTS

- **Obesity (High BMI)** is a major risk factor.
- **Lifestyle Factors**:
  - Lack of physical activity increases risk.
  - Healthy diet (fruits and vegetables) is protective.
- **Health Conditions**:
  - Hypertension and high cholesterol are prevalent among diabetics.
- **Socioeconomic Factors**:
  - Lower income and education levels correlate with higher diabetes prevalence.

RECOMMENDATIONS

- **Promote Healthy Lifestyles**:
  - Encourage regular physical activity.
  - Advocate for balanced diets rich in fruits and vegetables.
- **Implement Routine Screenings**:
  - Regularly monitor BMI, blood pressure, and cholesterol levels.
- **Target High-Risk Groups**:
  - Focus on communities with lower socioeconomic status.
  - Provide education and resources for diabetes prevention.
- **Enhance Public Health Policies**:
  - Support programs that reduce healthcare access barriers.

LIMITATIONS & FUTURE WORK

- **Data Limitations**:
  - Self-reported data may contain biases.
  - Data from 2015 may not reflect current trends.
- **Future Research**:
  - Utilize more recent datasets.
  - Explore additional predictive models (e.g., Random Forest, SVM).
  - Investigate the impact of other potential risk factors.

CONCLUSION

- **Summary**:
  - Identified key risk factors for diabetes.
  - Emphasized the role of BMI, general health, and lifestyle choices.

- **Final Thoughts**:
  - Addressing these factors can significantly reduce diabetes prevalence.
  - Collaboration between individuals, communities, and policymakers is essential.

REFERENCES

- Centers for Disease Control and Prevention (CDC)
- Behavioral Risk Factor Surveillance System (BRFSS) Data

THANK YOU