Diabetes
Visualisation & Insights Report

Prepared by

[Team 10 – Mellitus]

Date: 22/11/24

Table of Contents

1 Overview 3
1.1 Background 3
1.2 Key Findings 3
1.3 Actionable Insights 3
2 Health Factors 3
2.1 Key Findings 3
2.2 Actionable Insights 4
3 Lifestyle Factors 4
3.1 Key Findings 4
3.2 Actionable Insights 4
4 Diabetes Risk 4
4.1 Key Findings 4
4.2 Actionable Insights 5
5 Summary 5
6. Advanced Data Insights 6
6.1 Model Evaluation Insights 7
6.2 Potential Implications for the Company 7

1 Overview

1.1 Background

Diabetes is a growing concern, and understanding its patterns is necessary for prevention and management. By analysing the data from 100,000 patients, this report dives into the key trends surrounding diabetes, offering actionable recommendations for patients. The goal here was to make the data meaningful and help inform ways to get better health outcomes.

1.2 Key Findings

- **Diabetes Prevalence**: Out of 100,000 patients, **95,650** are diabetic, which is an alarmingly high percentage.
- **BMI and Age**: The average BMI for diabetic patients is **27.0**, categorizing most as overweight, and the average age is **45 years**. This places middle-aged patients at the highest risk.
- **Gender Split**: Diabetes affects **55% females** and **45% males**, showing a slight predominance in females.
- **Family History**: About **24% of diabetic patients** have a family history of diabetes, pointing to the importance of genetics.

1.3 Actionable Insights

- Patients, especially middle-aged adults, should consider **regular diabetes screenings** to catch the condition early.
- Maintaining a **healthy BMI (below 25)** can significantly reduce the risk.
- Patients with a family history of diabetes should focus on **preventative measures**, such as healthy eating and physical activity, to offset genetic predisposition.

2 Health Factors

The relationship between diabetes, hypertension, and other health factors stood out in the data:

2.1 Key Findings

- **Hypertension**: Among diabetic patients, **83,780** did not have hypertension, but **16,220** did. While most diabetics are not hypertensive, the 16% who are face compounded risks.
- **BMI**: Overweight (BMI 25–30) and obese patients (BMI > 30) dominate the diabetic population, while underweight patients have the lowest prevalence.
- **Stress Levels**: Elevated and extreme stress levels were reported in **30,000 diabetic cases**, indicating that stress plays a significant role.

- **Alcohol Consumption**: Moderate alcohol consumption was observed in **44.96%** of diabetic patients, followed by light (**37.13%**) and heavy (**17.49%**) drinking.

2.2 Actionable Insights

- For those with **hypertension**, managing blood pressure through lifestyle changes or medication is key to lowering overall health risks.
- Patients should work towards achieving a **BMI below 25**, which could reduce the strain on their metabolic health.
- Managing stress through practices like mindfulness or therapy is essential for reducing diabetes risk.
- Limiting alcohol intake can help prevent further health complications.

3 Lifestyle Factors

Lifestyle factors are some of the most modifiable contributors to diabetes risk.

3.1 Key Findings

- **Physical Activity**: Sedentary lifestyles dominate among diabetic patients, especially in middle-aged and senior groups. Very active patients showed significantly lower prevalence rates.
- **Sleep Patterns**: Those sleeping fewer than 6 hours per night reported higher diabetes prevalence compared to patients sleeping 7–8 hours.
- **Social Media Usage**: Moderate and occasional social media users showed higher diabetes rates, possibly linked to sedentary behaviors.

3.2 Actionable Insights

- Patients should aim to get at least **150 minutes of moderate exercise per week**, such as brisk walking or cycling, to reduce risk.
- Improving **sleep hygiene** by creating consistent sleep schedules and limiting screen time before bed can help regulate blood sugar levels.
- Patients should balance their online and offline activities by incorporating more physical and social offline engagements.

4 Diabetes Risk

The dataset's predictive analysis sheds light on high-risk groups and potential interventions.

4.1 Key Findings

- **High vs. Low Risk**: Out of the diabetic population, **3,261 patients** are classified as high-risk due to their BMI, family history, and lifestyle, while **96,739 are low-risk**.

- **Diet and BMI Correlation**: Atkins, pescatarian, and ketogenic diets correlate with higher BMIs, while vegan and Mediterranean diets are associated with lower BMIs.
- **Pregnancy**: Diabetic pregnancies averaged **0.78 pregnancies per patient**, emphasizing the need for gestational diabetes monitoring.

4.2 Actionable Insights

- High-risk patients should **consult healthcare providers** for tailored plans to manage their risk factors.
- Shifting to diets like **Mediterranean or plant-based diets** could help lower BMI and improve metabolic health.
- Pregnant patients should undergo regular **gestational diabetes screenings** and follow dietary recommendations.

5 Summary

After analysing the dataset, it's clear that diabetes is a complex issue influenced by health metrics, lifestyle, and genetics. To summarise the top factors:

1. **Physical Activity Level**
   Sedentary lifestyles are the dominant behaviour among patients with diabetes. The data clearly shows that the less active someone is, the higher their risk. On the other hand, patients who are moderately or very active have significantly lower diabetes prevalence. This makes physical activity one of the strongest predictors of diabetes.
2. **BMI (Body Mass Index)**
   The average BMI for diabetic individuals is **27.0**, which falls in the overweight category. Obesity (BMI > 30) is a major driver, with higher BMI levels strongly correlating with increased diabetes risk.
3. **Stress Level**
   Elevated stress levels are a consistent trend among diabetic individuals. The data highlights that **30,000 cases** are linked to extreme or elevated stress. Stress impacts the hormonal balance, and over time, it can contribute to the development of diabetes.
4. **Alcohol Consumption**
   Heavy alcohol consumption stands out as a prevalent behaviour among patients with diabetes, followed closely by moderate drinking. Over **44% of diabetic cases** are associated with moderate alcohol use, suggesting that alcohol consumption is a lifestyle factor that cannot be overlooked.
5. **Family History of Diabetes**
   Approximately **24% of diabetic cases** involve individuals with a family history of the condition. This highlights the importance of genetics as a risk factor. While it's not something patients can control, it underscores the need for preventive care and early monitoring for those with a family history.

Conclusion and Recommendations

For patients at risk of developing diabetes, the data points to actionable steps that can significantly lower risk:

1. **Maintain a Healthy BMI**
    - Aim for a BMI below 25 through a balanced diet and regular physical activity.
2. **Increase Physical Activity**
3. **Adopt a Healthier Diet**:
    - Adopt diets rich in vegetables, whole grains, and lean proteins, such as a mediterranean diet.
    - Reduce processed foods and high-carb meals that contribute to weight gain.
4. **Improve Sleep Hygiene**:
    - Try to sleep for 7–8 hours per night.
5. **Manage Stress and Mental Health**:
    - Have hobbies that reduce stress levels.
6. **Limit Alcohol Consumption**
7. **Focus on High-Risk Groups**:
    - Patients with a family history of diabetes, sedentary lifestyles, or poor dietary habits should prioritize preventative care, including regular health screenings.

By making small but consistent changes, patients can lower their diabetes risk and improve their overall quality of life.

There are significant implications for the diabetes drugs company. The high prevalence of diabetes among males with sedentary lifestyles, elevated stress, and heavy alcohol consumption presents a substantial market opportunity. By focusing on the top predictors, the company can develop targeted medications and personalised treatment plans tailored to this demographic. This data-driven approach can enhance patient outcomes, expand market reach, and reinforce the company's position as a leader in diabetes care.

6. Advanced Data Insights

We also did further advanced data analysis to determine the best course of actions for the client.

For instance, the analysis identifies the following as the most significant predictors of diabetes, ranked by their importance:

1. **Physical Activity Level**: Sedentary behaviour is the strongest predictor of diabetes.
2. **Weight**: Higher weight correlates closely with diabetes prevalence, emphasizing the role of obesity in disease progression.

3. **Stress Level**: Elevated stress levels substantially increase the likelihood of developing diabetes.
4. **Family History of Diabetes**: A strong genetic link suggests a need for tailored solutions for those with a family history.
5. **BMI**: Higher BMI values show a strong association with diabetes prevalence.
6. **Alcohol Consumption**: Heavy alcohol use emerges as a key lifestyle factor impacting diabetes risk.

6.1 Model Evaluation Insights

- **Random Forest and XGBoost**:
    - These advanced machine learning models consistently outperform others in predictive accuracy, making them ideal for identifying high-risk patients. Their ability to handle complex data makes them valuable in uncovering multifactorial diabetes risks.
- **Logistic Regression**:
    - While less accurate, logistic regression provides a simple and efficient solution for quick analyses or when computational resources are limited.

6.2 Potential Implications for the Company

1. Market Opportunity

The data shows that diabetes predominantly affects males with sedentary lifestyles, high stress, and heavy alcohol consumption. By focusing on these characteristics, the client can develop medications or treatment plans specifically designed for this demographic. For example:

- Medications that address metabolic issues linked to physical inactivity and obesity.
- Stress-modulating supplements or treatments targeting diabetes-related complications.

2. Targeting Using Top Predictors

The top six predictors - physical activity level, weight, stress, family history, BMI, and alcohol consumption—offer a clear roadmap for intervention strategies. Combining these insights with predictive modelling, you can:

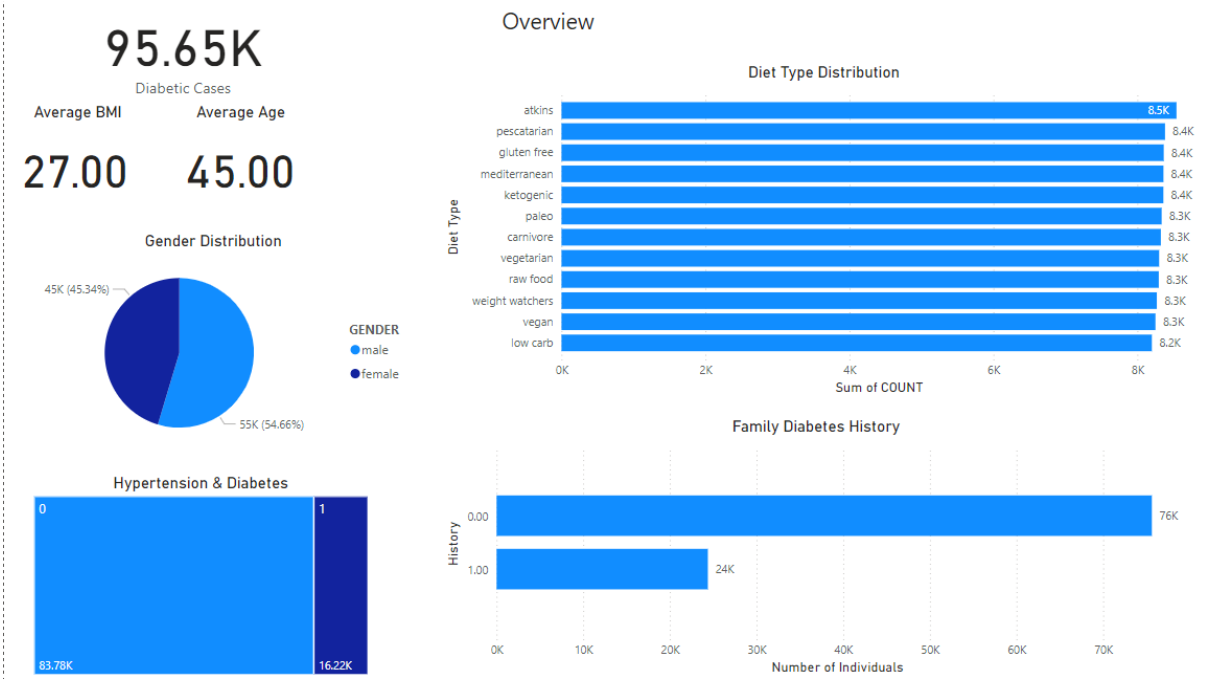
- Develop and market solutions tailored to individuals with these characteristics.
- Offer personalized treatment regimens based on identified risk factors.

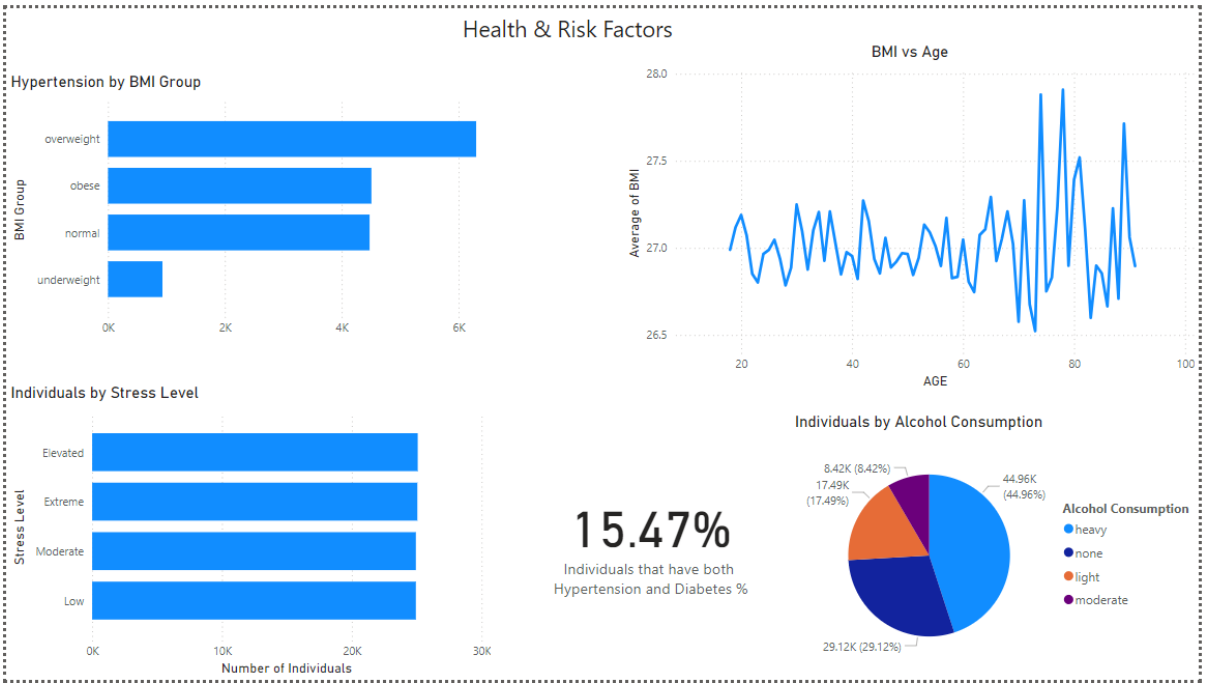
3. Predictive Analytics for Early Detection

By leveraging advanced models like **Random Forest** and **XGBoost**, the client can accurately identify high-risk individuals. These tools can support:

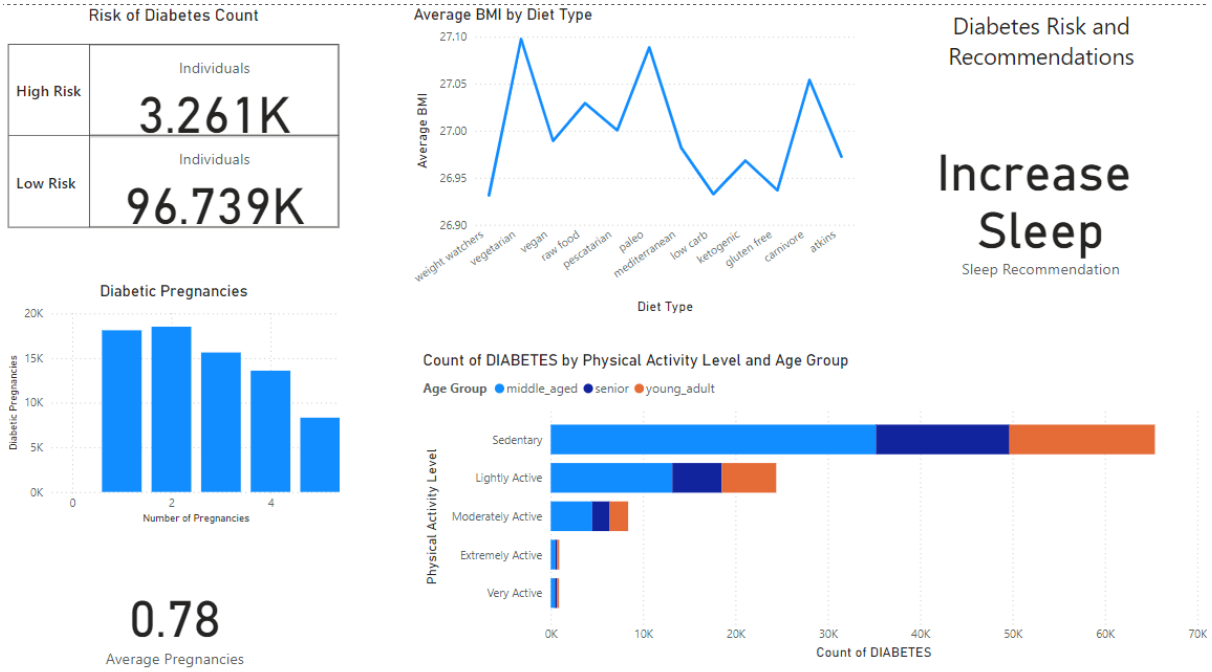
- Early intervention programs where your drugs can play a preventative or mitigating role.
- Collaboration with healthcare providers to integrate predictive algorithms into patient management systems.

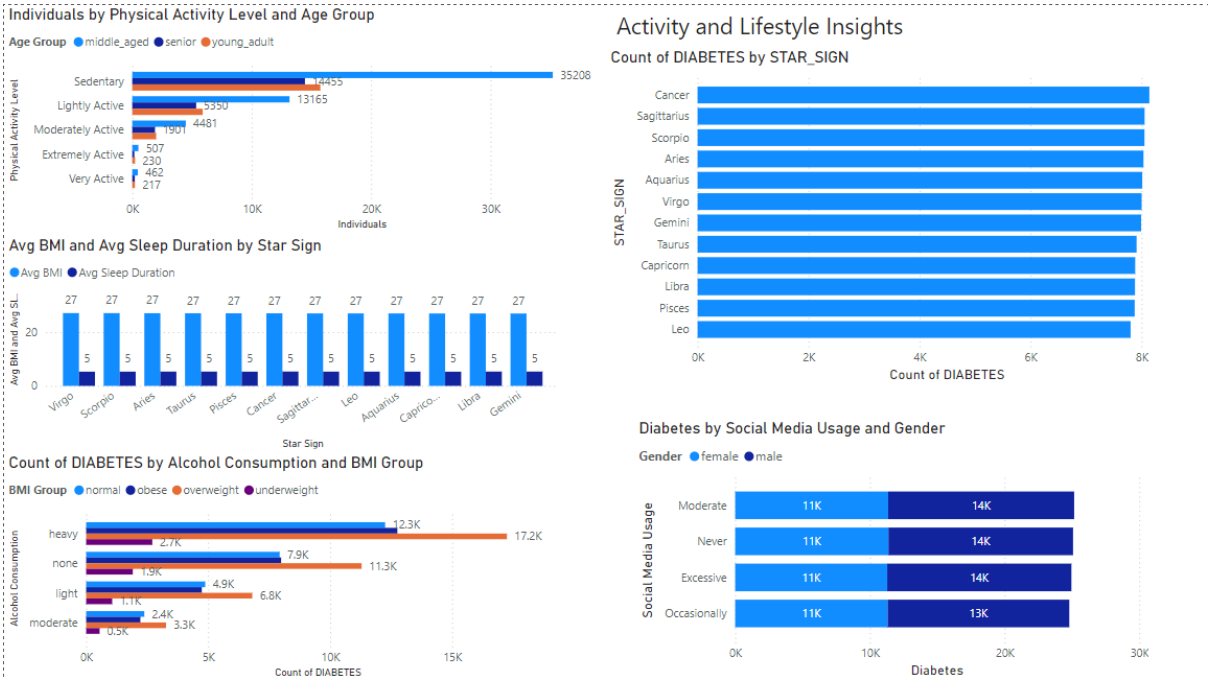
Appendix:

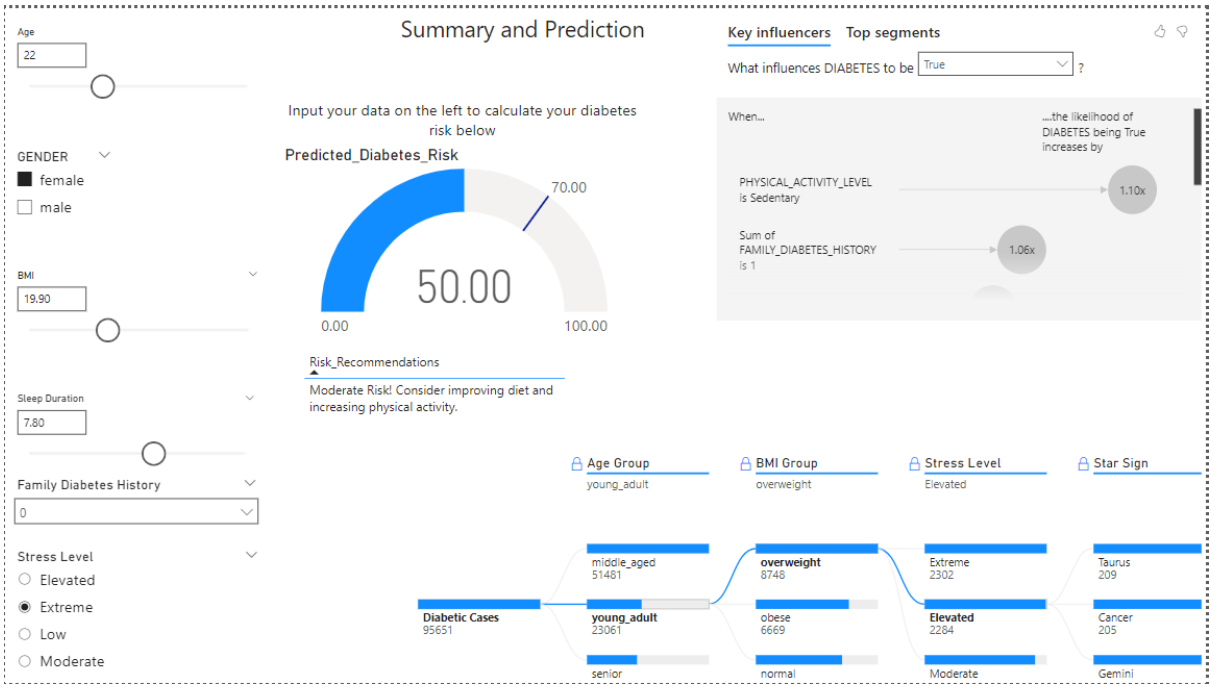


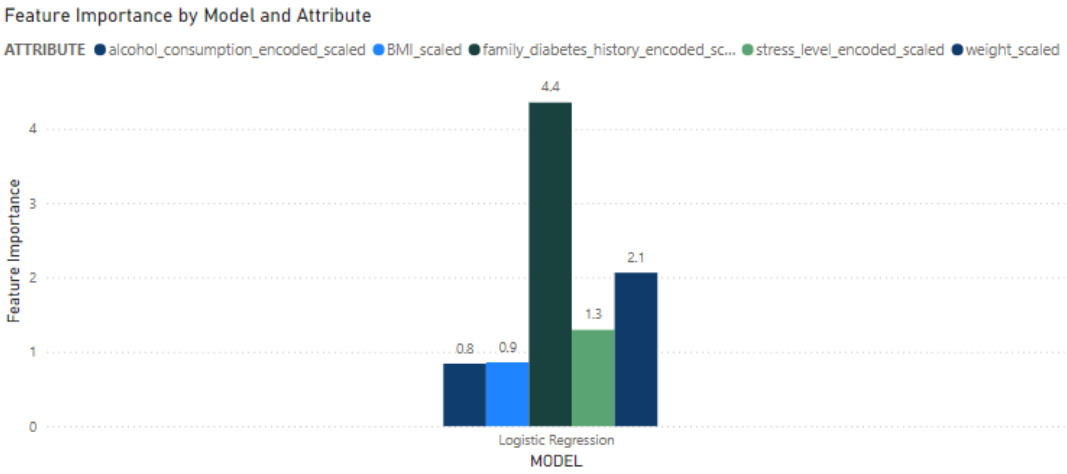


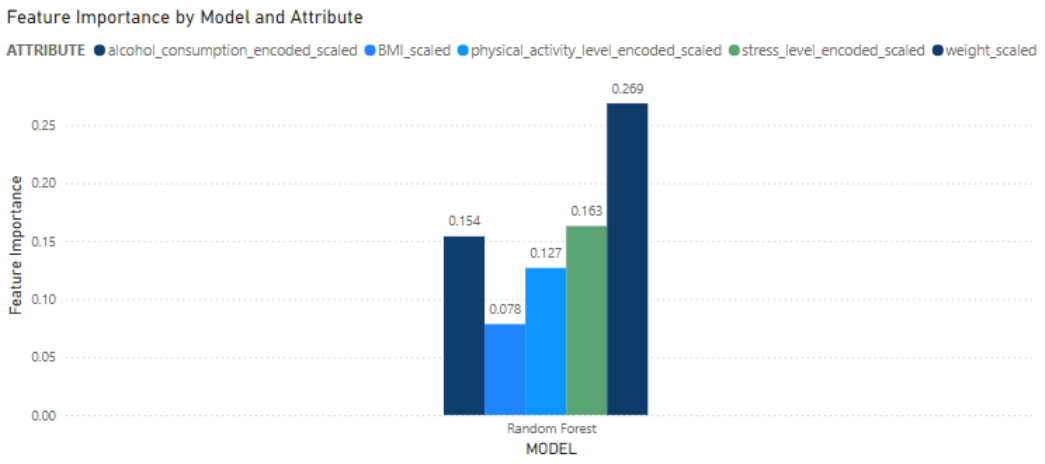
Date: 22/11/24

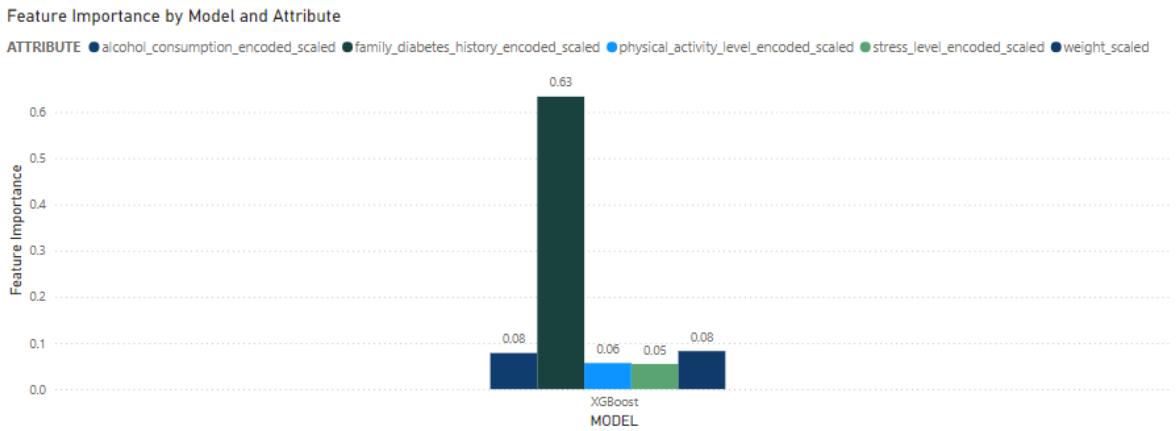

Risk of Diabetes Count
High Risk
Individuals
3.261K
Low Risk
Individuals
96.739K
Average BMI by Diet Type
Average BMI
27.10
27.05
27.00
26.95
26.90
weight watchers
vegetarian
vegan
raw food
pescatarian
paleo
mediterranean
low carb
ketogenic
gluten free
carnivore
atkins
Diet Type
Diabetes Risk and Recommendations
Increase Sleep
Sleep Recommendation
Diabetic Pregnancies
Diabetic Pregnancies
20K
15K
10K
5K
0K
0
2
4
Number of Pregnancies
0.78
Average Pregnancies
Count of DIABETES by Physical Activity Level and Age Group
Age Group middle_aged senior young_adult
Physical Activity Level
Sedentary
Lightly Active
Moderately Active
Extremely Active
Very Active
0K
10K
20K
30K
40K
50K
60K
70K
Count of DIABETES


Individuals by Physical Activity Level and Age Group
Age Group middle_aged senior young_adult
Physical Activity Level
Sedentary
4455
35208
Lightly Active
5350
13165
Moderately Active
1901
4481
Extremely Active
507
230
Very Active
462
217
0K
10K
20K
30K
Individuals
Avg BMI and Avg Sleep Duration by Star Sign
Avg BMI
Avg Sleep Duration
Avg BMI and Avg SL...
27
20
0
5
Virgo
Scorpio
Aries
Taurus
Pisces
Cancer
Sagittar...
Leo
Aquarius
Caprico...
Libra
Gemini
Star Sign
Count of DIABETES by Alcohol Consumption and BMI Group
BMI Group normal obese overweight underweight
Alcohol Consumption
heavy
12.3K
17.2K
2.7K
none
7.9K
11.3K
1.9K
light
4.9K
6.8K
1.1K
moderate
2.4K
3.3K
0.5K
0K
5K
10K
15K
Count of DIABETES
Activity and Lifestyle Insights
Count of DIABETES by STAR_SIGN
STAR_SIGN
Cancer
Sagittarius
Scorpio
Aries
Aquarius
Virgo
Gemini
Taurus
Capricorn
Libra
Pisces
Leo
0K
2K
4K
6K
8K
Count of DIABETES
Diabetes by Social Media Usage and Gender
Gender female male
Social Media Usage
Moderate
11K
14K
Never
11K
14K
Excessive
11K
14K
Occasionally
11K
13K
0K
10K
20K
30K
Diabetes


Summary and Prediction
Age
22
GENDER
female
male
BMI
19.90
Sleep Duration
7.80
Family Diabetes History
0
Stress Level
Elevated
Extreme
Low
Moderate
Input your data on the left to calculate your diabetes risk below
Predicted_Diabetes_Risk
70.00
50.00
0.00
100.00
Risk_Recommendations
Moderate Risk! Consider improving diet and increasing physical activity.
Key influencers
Top segments
What influences DIABETES to be
True
?
When...
...the likelihood of DIABETES being True increases by
PHYSICAL_ACTIVITY_LEVEL is Sedentary
1.10x
Sum of FAMILY_DIABETES_HISTORY is 1
1.06x
Age Group
young_adult
BMI Group
overweight
Stress Level
Elevated
Star Sign
Diabetic Cases
95651
middle_aged
51481
young_adult
23061
senior
overweight
8748
obese
6669
normal
Extreme
2302
Elevated
2284
Moderate
Taurus
209
Cancer
205
Gemini


Feature Importance by Model and Attribute
ATTRIBUTE
alcohol_consumption_encoded_scaled
BMI_scaled
family_diabetes_history_encoded_sc...
stress_level_encoded_scaled
weight_scaled
Feature Importance
4
3
2
1
0
4.4
2.1
1.3
0.8
0.9
Logistic Regression
MODEL


Feature Importance by Model and Attribute
ATTRIBUTE
alcohol_consumption_encoded_scaled
BMI_scaled
physical_activity_level_encoded_scaled
stress_level_encoded_scaled
weight_scaled
Feature Importance
0.25
0.20
0.15
0.10
0.05
0.00
0.269
0.163
0.154
0.127
0.078
Random Forest
MODEL



ACCURACY	MODEL	AVG_ROC_AUC	F1_SCORE	MAX_ROC_AUC	MIN_ROC_AUC	PRECISION	RECALL
0.87	Logistic Regression	0.95	0.93	0.95	0.94	0.99	0.87
0.98	XGBoost	0.98	0.99	0.98	0.98	0.99	0.99
0.98	Random Forest	0.98	0.99	0.98	0.97	0.99	0.99