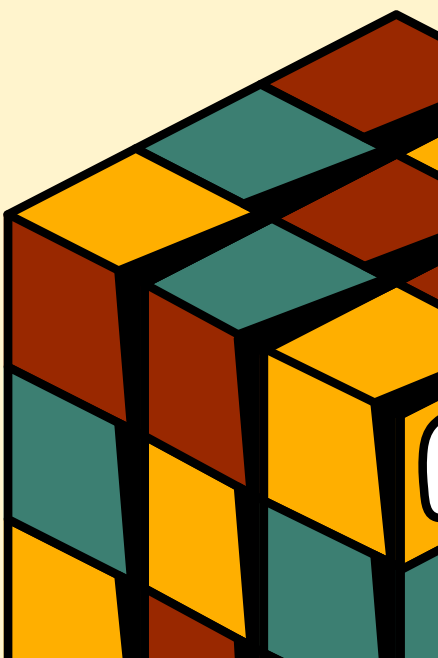
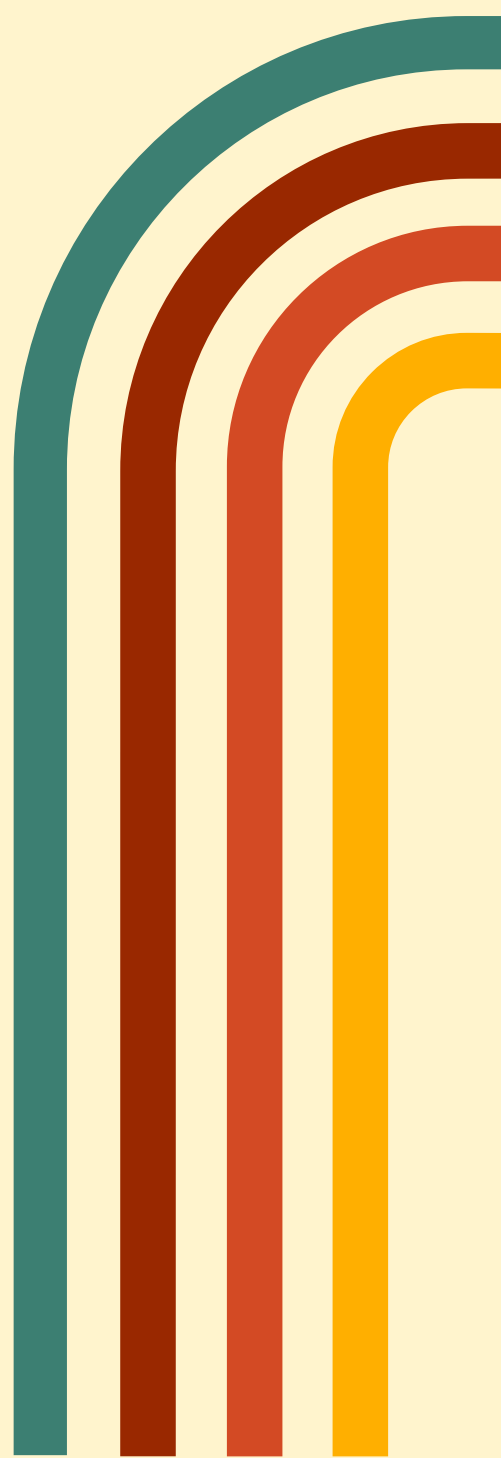


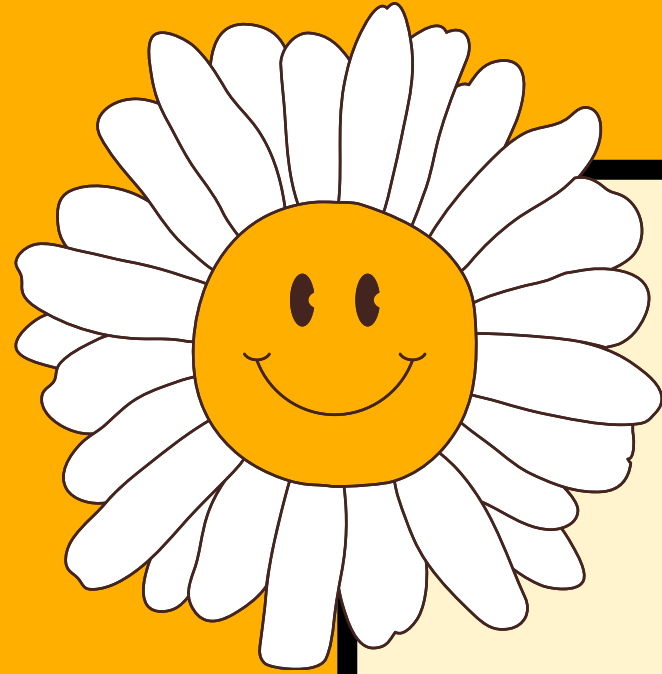
Prepared by group 6

Group 6 Project

DIABETES DATA SET ANALYSIS

18 April, 2024





Introduction

Our project aims to analyze, and draw conclusions from a data set that examines several different factors and their correlation with diabetes (National Institute of Diabetes and Digestive and Kidney Diseases). We recognize that this data set does not describe the criteria used for collecting the data including, exclusion factors.

Diabetes Analysis

Questions

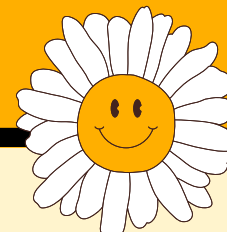
Which factors, such as BMI glucose, and blood pressure contributed the most to the outcome of diabetes?

As a result, what is the breakdown of the factors in relation to the outcome of diabetes



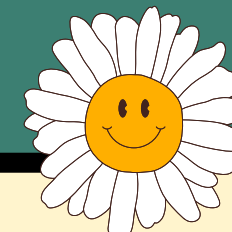
Hypothesis

We hypothesize that the factors of BMI and glucose will play a significant contribution to the outcome of patients having diabetes.



Assumptions going into this data set

There were three main assumptions associated with this data set. Firstly, we assumed all patients were assigned female at birth. Secondly, we assumed that all pregnancies were carried to full term. Lastly, we assumed that these pregnancies represented the standard typical case.



Methodology

Data Visualization

We primarily used bar charts, pie graphs, and scatter plots for visualization

We used scatter plots when looking at potential factors that may be correlated

Brainstorming Sessions:

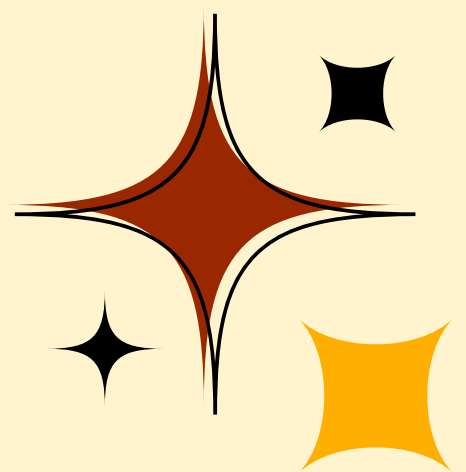
Collaborated and thought of ways to illustrate the data set that showed that it was a correlational relationship between factors.

We considered the different methods to illustrate the relationships between BMI, glucose, pregnancies, and age.

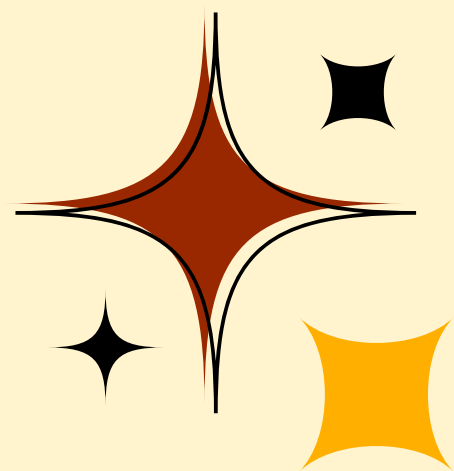
Testing and Refinement

Throughout the actual coding of this process, we realized that scatterplots for everything we initially wanted would not work (ex: Diabetes Pedigree Function). We had to continue to refine and find data collection methods that were simple and effective.

As we reviewed the visualizations we learned of a different story being told.



Factors To Compare



BMI and glucose

- We assumed that there would be a higher level of association between BMI and glucose because BMI levels are indicative of other predetermined health conditions, and one of them might be glucose.

Glucose and blood pressure

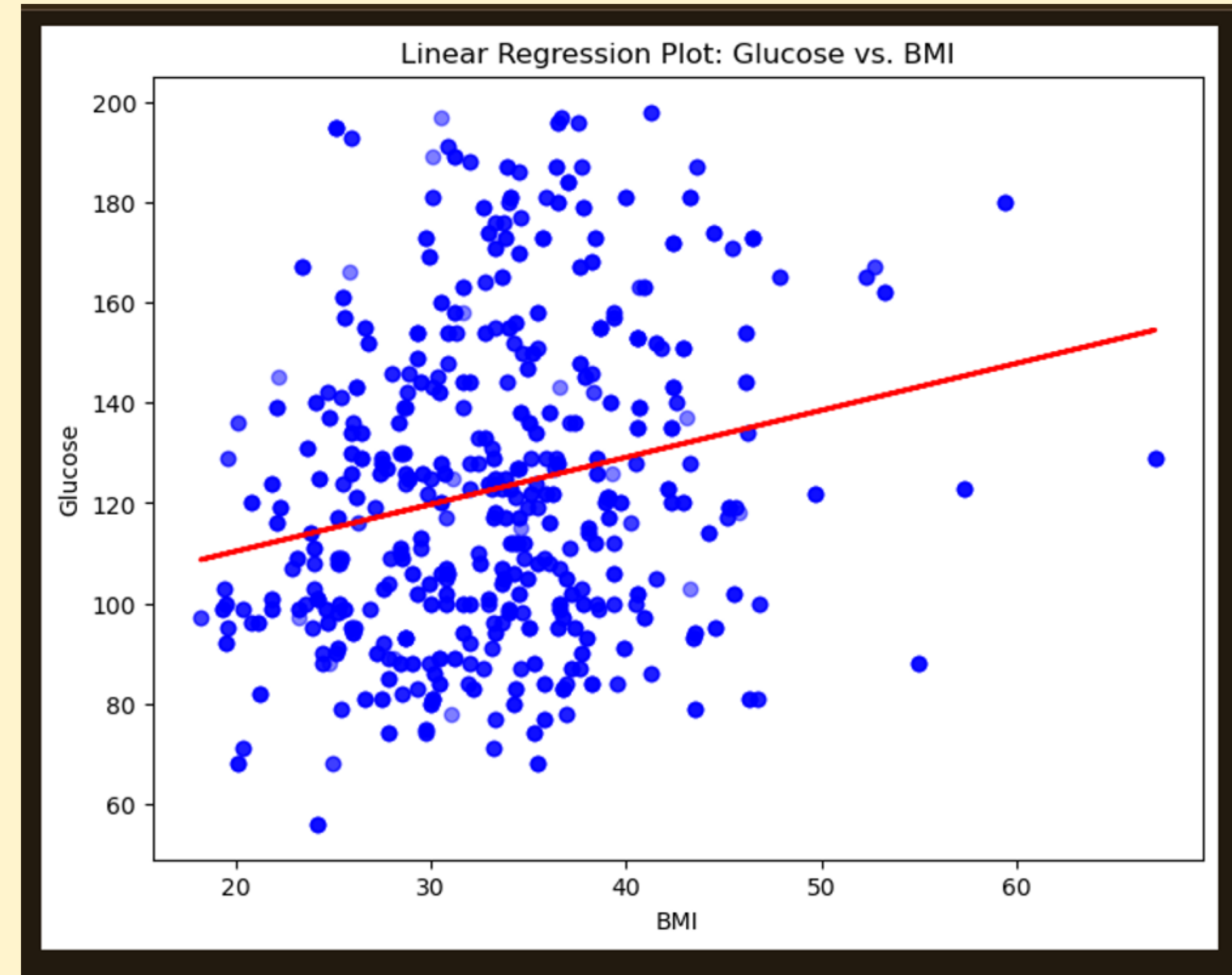
- We assumed that the scatterplot of glucose and blood pressure illustrates a level of correlation, indicating that higher glucose levels correspond to higher blood pressure rates.

Pregnancy

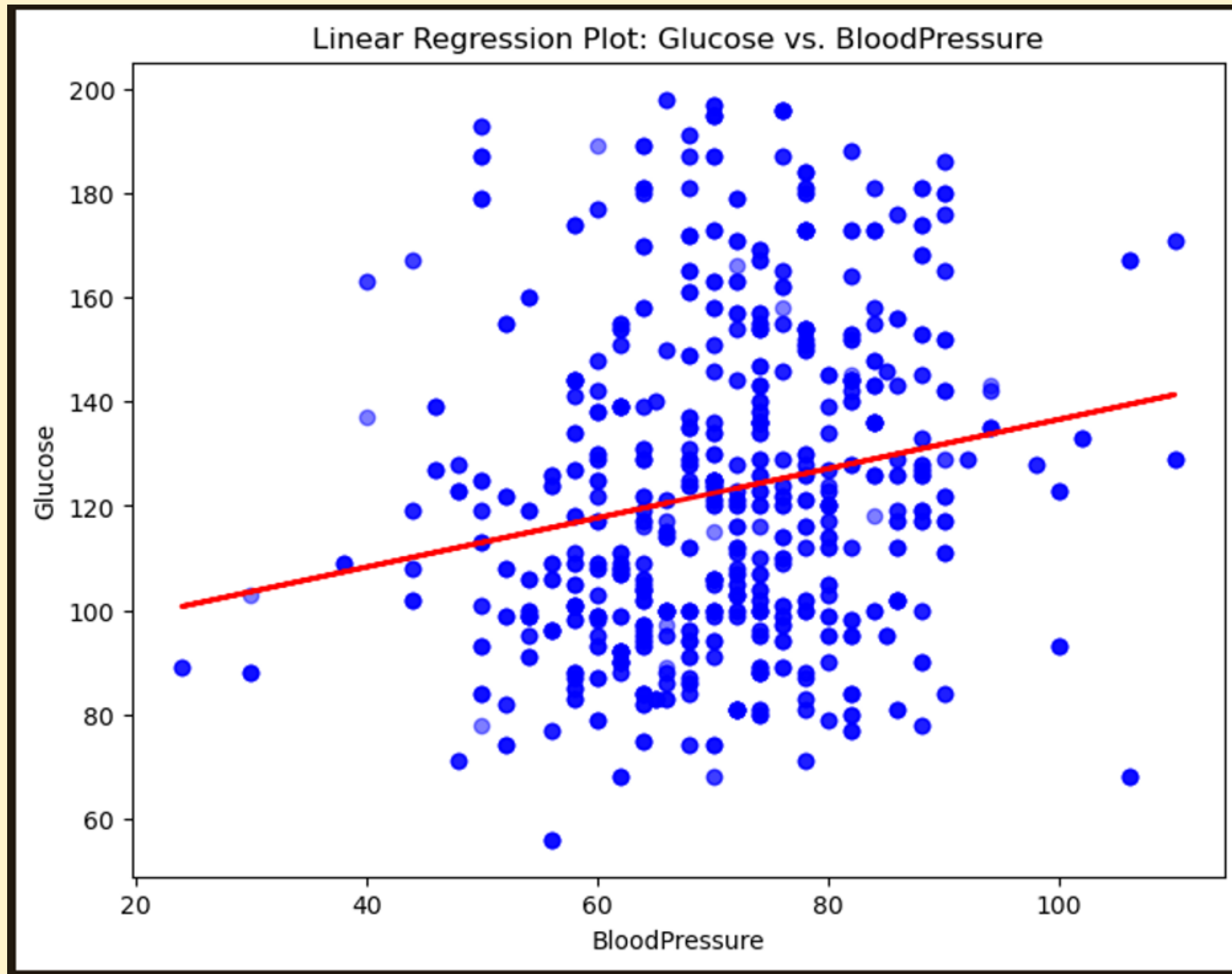
- We assumed that there would be a higher incidence of diabetes among those with multiple pregnancies.

Data Analysis

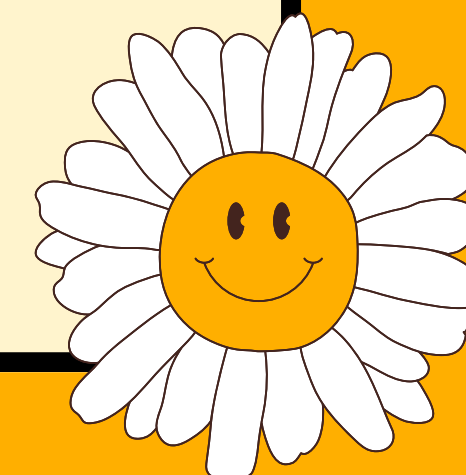
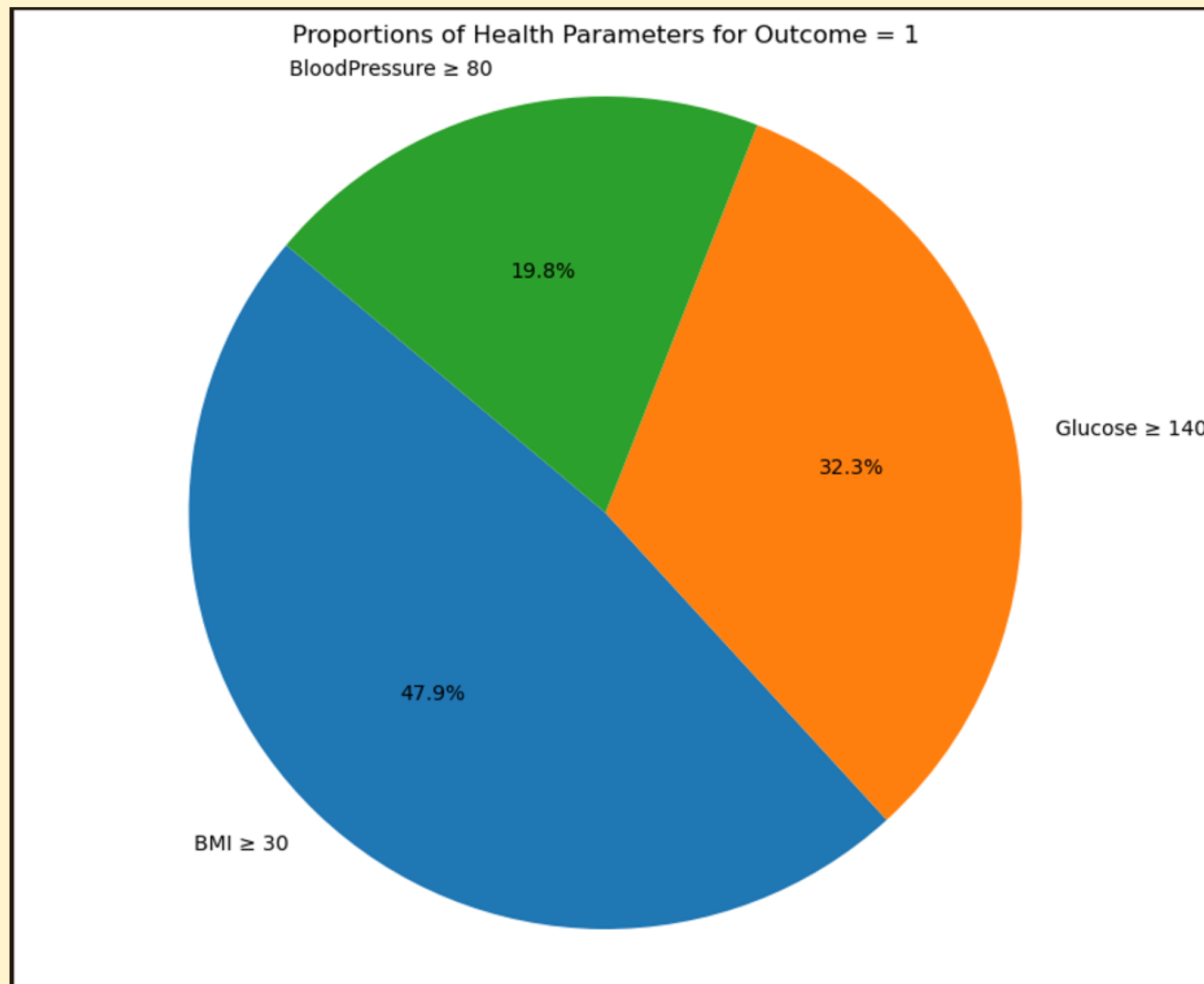
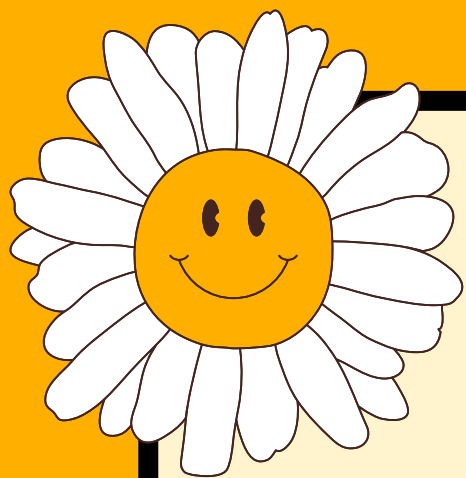
This scatter plot looks at the relationship between BMI and Glucose. This was an interesting plot to look at because it illustrated that the preconceived notion that a patient having a higher BMI will translate to a higher glucose level is wrong in which the concentrated range of BMI for the patients was 30–40 BMI.



Data Analysis

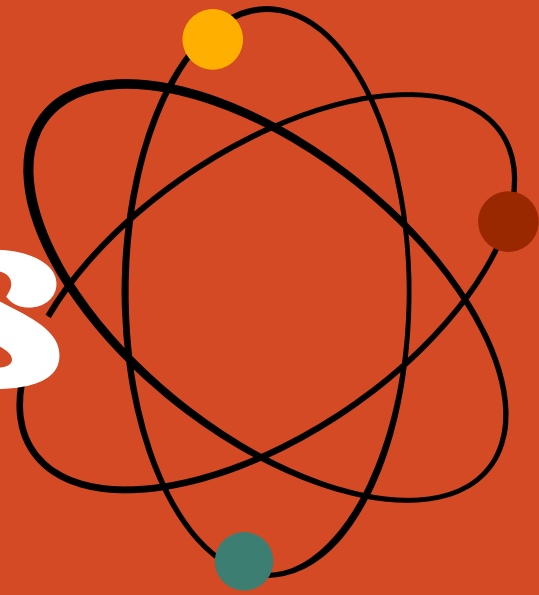


Glucose and blood pressure's relationship is linear in which there is a more concentrated rate of blood pressure and glucose at the center of the plot (in which the higher the blood pressure rate the higher the glucose rate.) which is what we suspected.



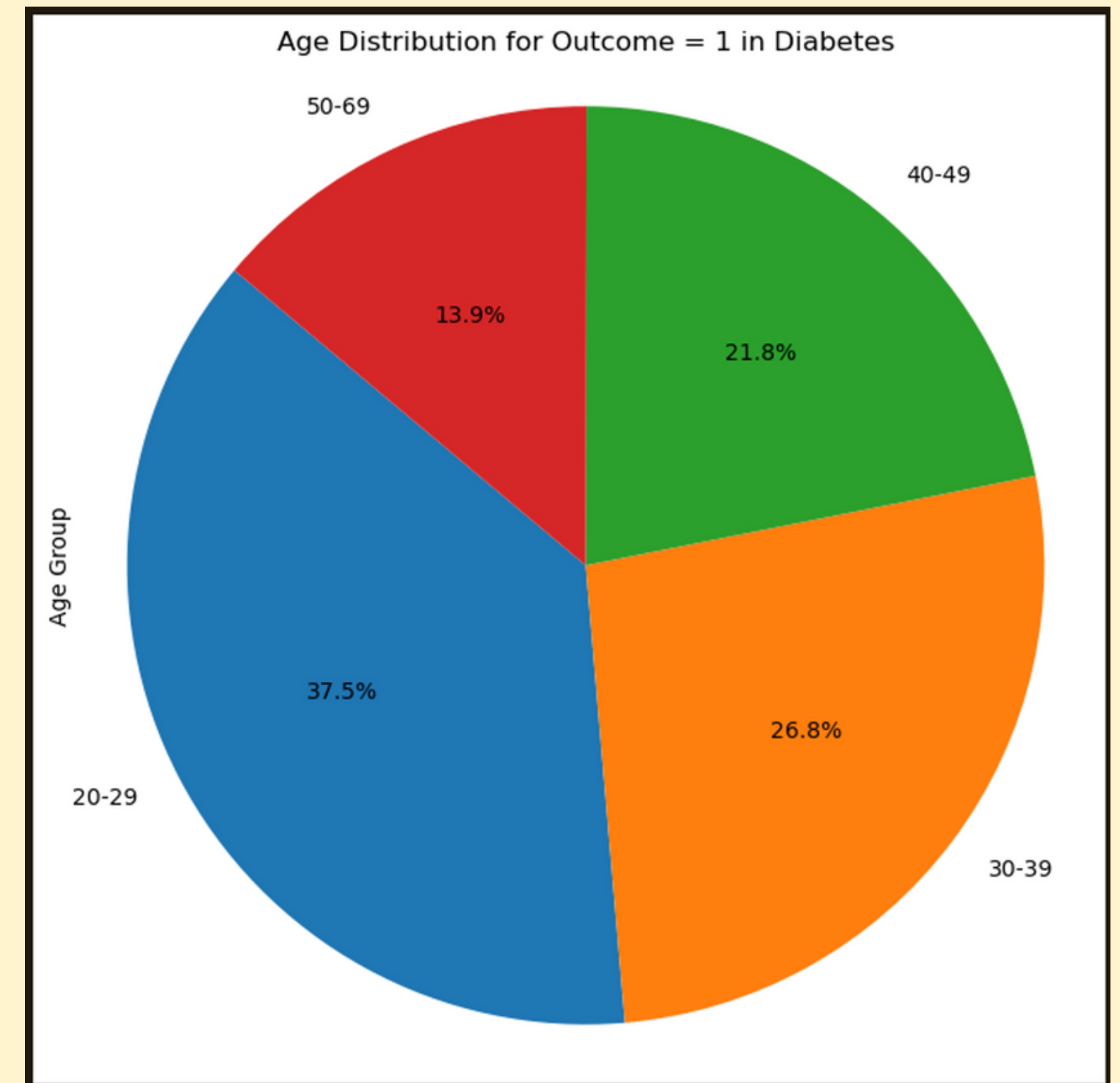
Three Factors

Data Analysis



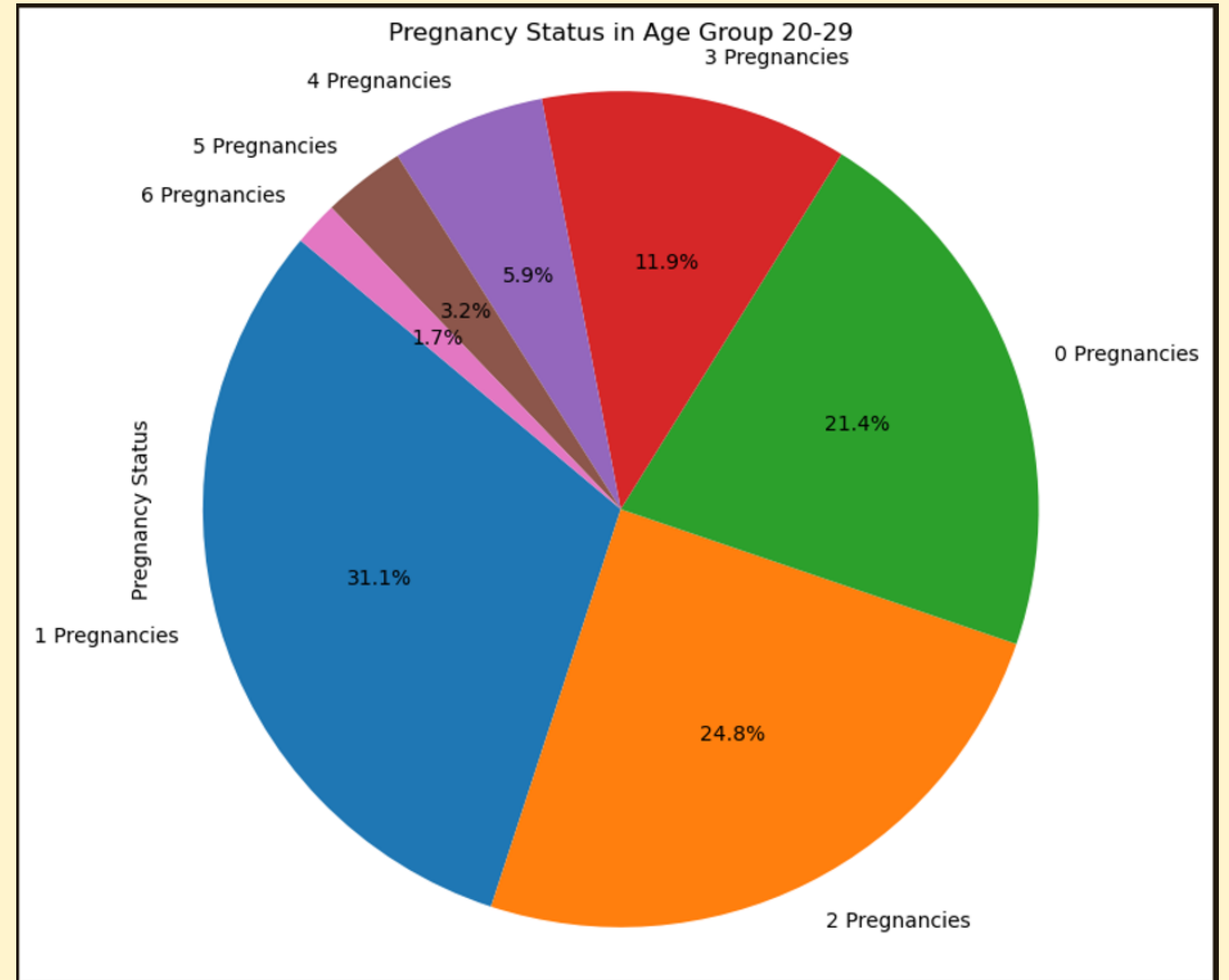
Age distribution of the likelihood to developing diabetes

This pie chart was not what we suspected the age distribution to look like, with the leading age being 20–29 years, while the second leading age is 30–39 years old as we suspected. This may be due to some unmeasurable factors.

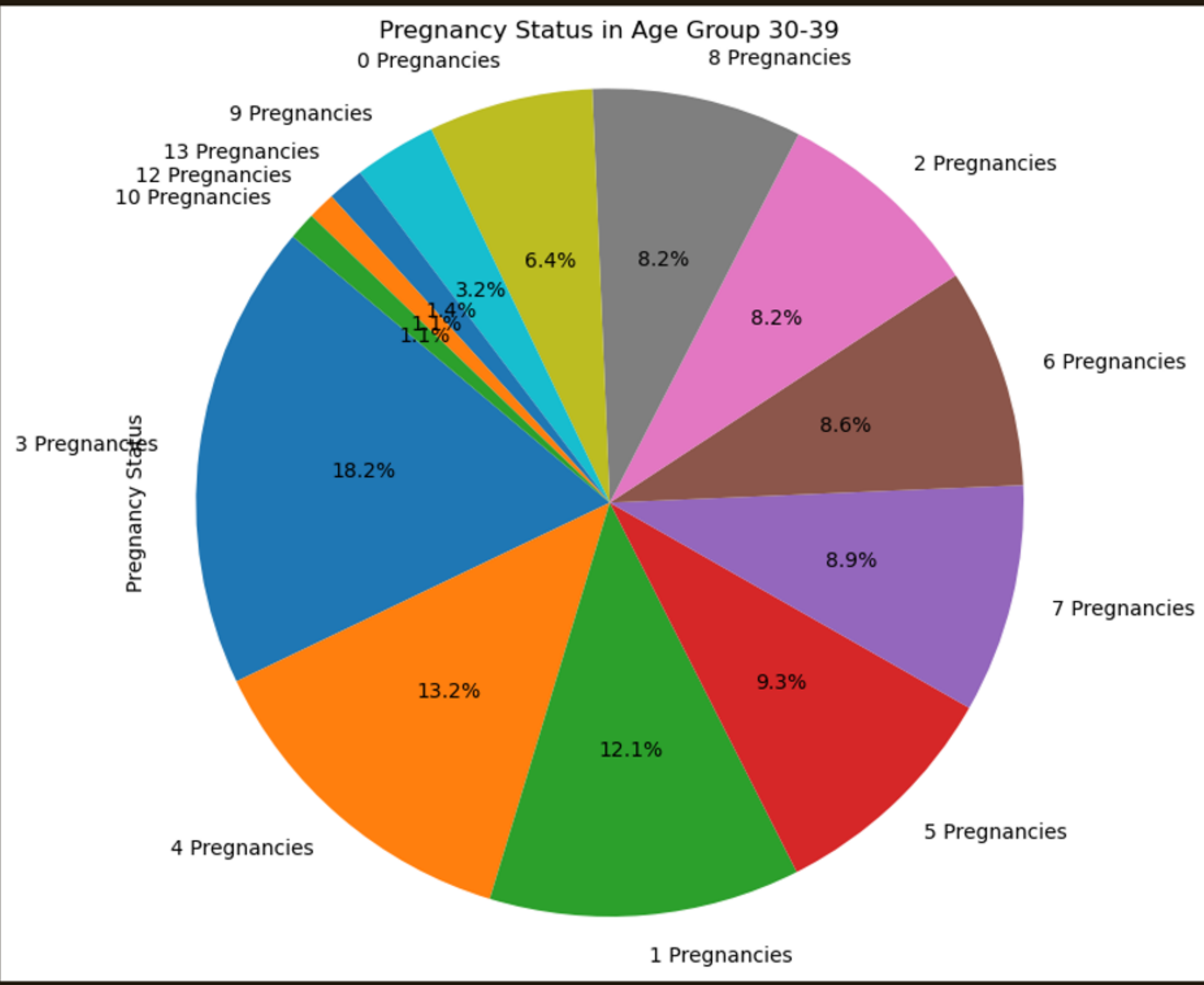


Data Analysis

This pie chart is looking at the analysis of that the amount of pregnancies within the specific age group of 20–29. This was as expected with most patients having 1–2 pregnancies at that time.



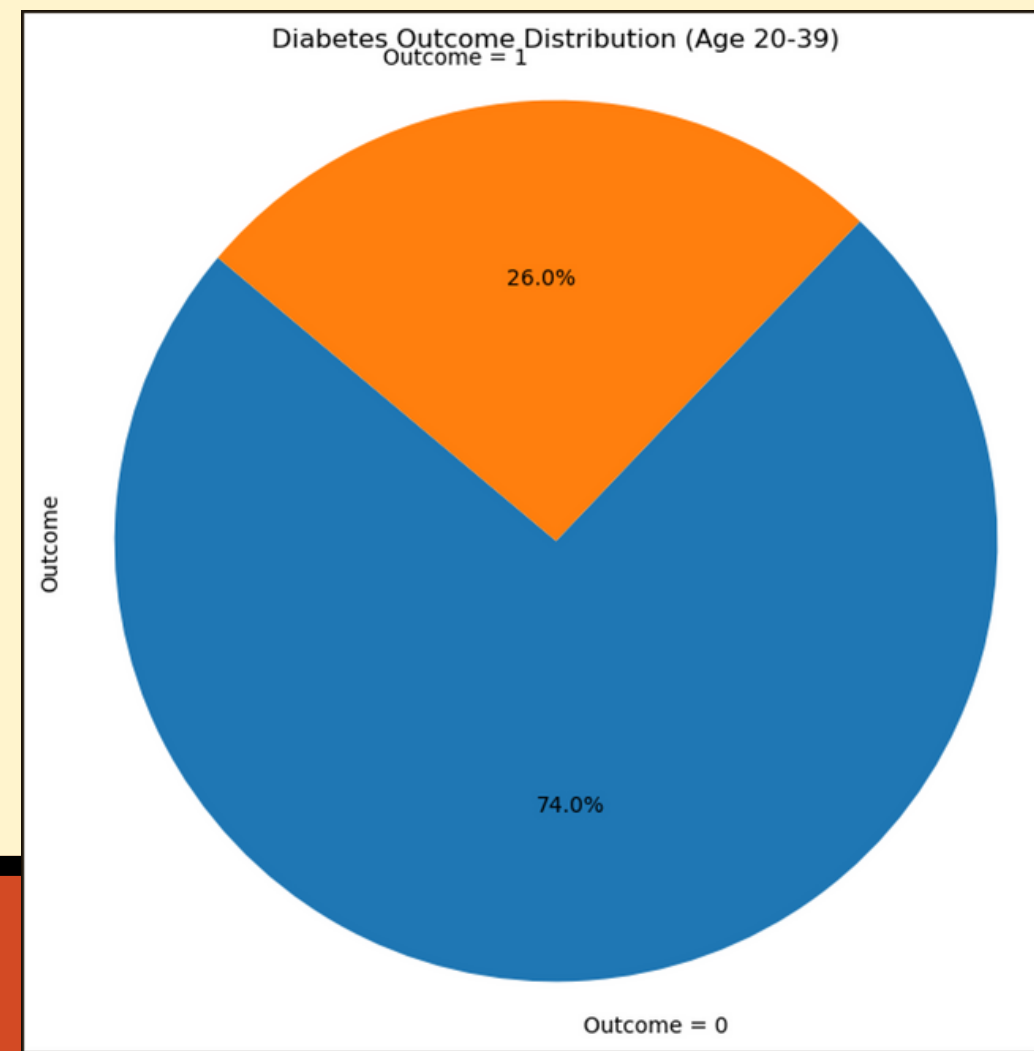
Data Analysis



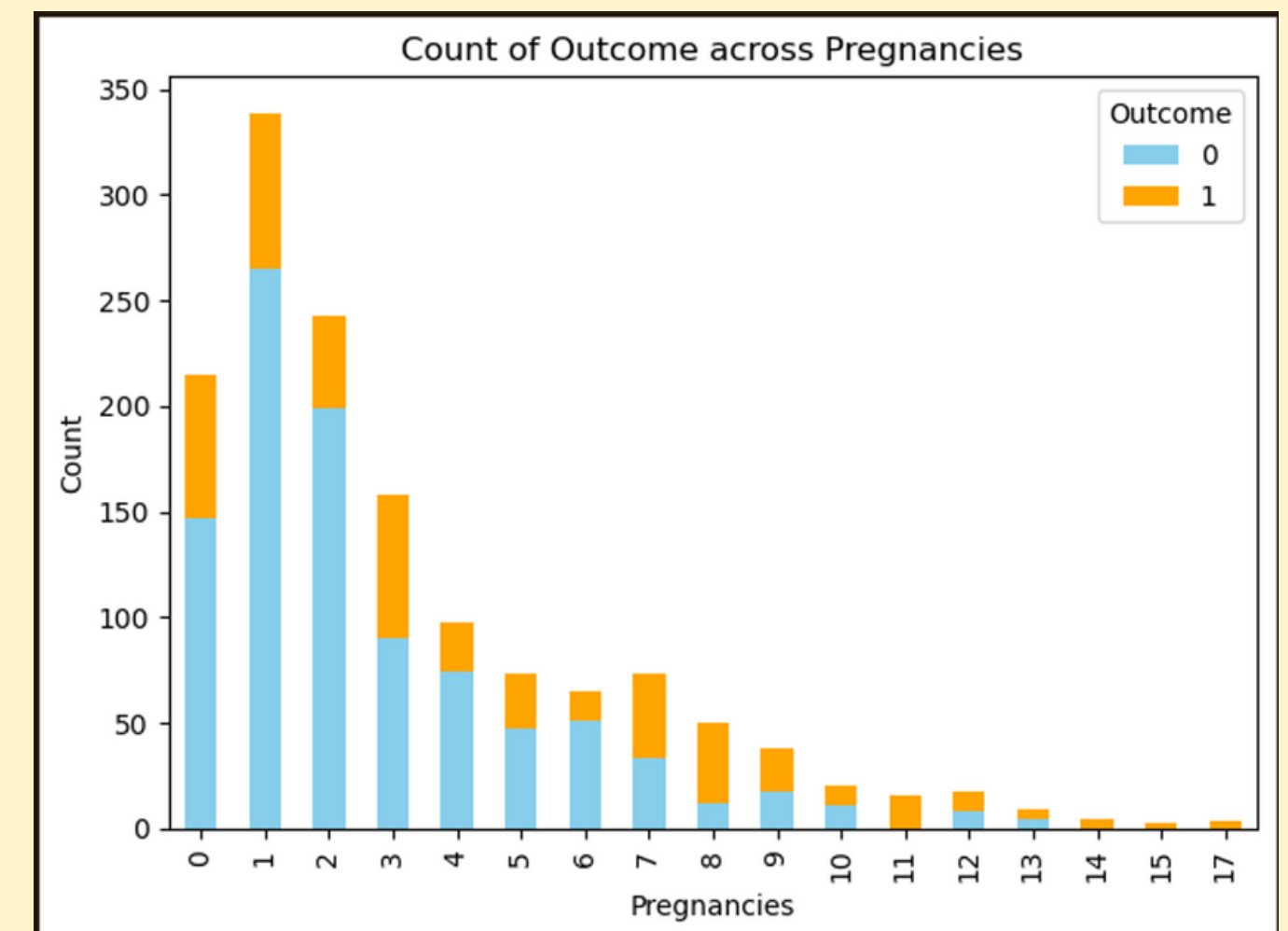
This pie chart illustrates that the pregnancy count within the age group of 30–39 year old are not as equally distributed as we would have thought them to be. We assume that this is a result of pregnancies after the age of 36 are considered to be at risk.

Data Analysis

This pie chart illustrates the outcome distribution between specific ages [range of 20-39] of being pregnant and not pregnant. Within this data set it showed that 26% had Diabetes, while on the other hand the outcome of not having diabetes is about 74%.



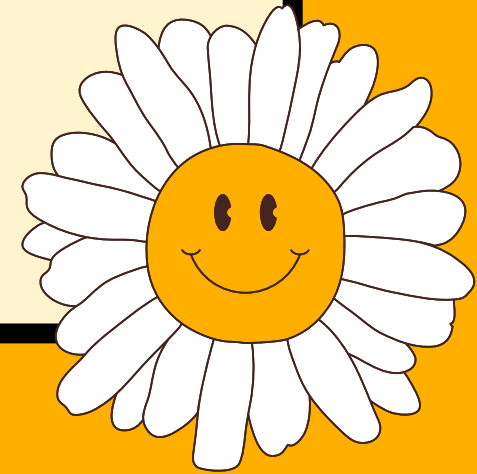
This bar graph also illustrates the same as the pie chart, however it's breaking it down by number of pregnancies and whether or not they have diabetes.





Conclusion

Through analyzing this data set we found out that our hypothesis was wrong, and that the factors of the BMI and glucose were not strongly associated enough to cause patients to have diabetes, while being pregnant. This was particularly noted through our first analysis in which we compared, and BMI and glucose levels.



Thank
You

