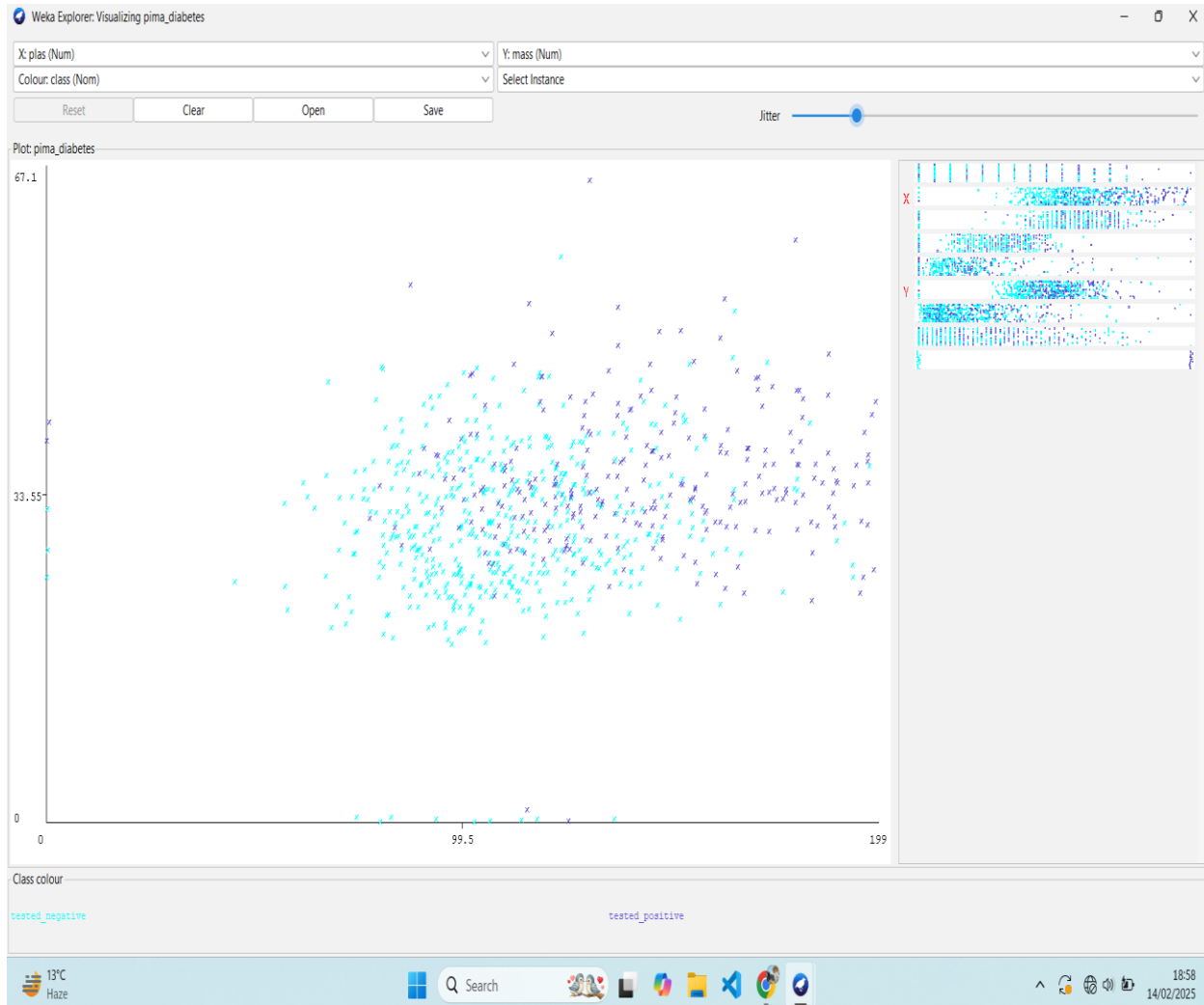


Lab Report on - Data Visualization -Lab 3

Submitted By: Aaditya Sapkota, BDDSc Roll no:37

1. Here is the screenshot of the Scatter Plot (Plasma Glucose Concentration Vs BMI)



Below are the answers to the questioned mentioned in the Lab Exercise 3:

1. Identify the visible trends or clusters in the scatter plot.

= There is a slight positive trend suggesting that as plasma glucose concentration increases, BMI also increases and vice-versa. This is a general trend but the correlation is not very strong and large amount of data points exist that deviate from this trend.

2. Answer the following questions:

- **Is there a correlation between Plasma glucose concentration and BMI?**
= Yes, there is a weak positive correlation between plasma glucose concentration and BMI. There are plenty of data points that deviate from this general trend.
- **Are there any outliers in the dataset? If yes, what might they indicate?**
= Yes, there are a few outliers in the dataset. There are points with high plasma glucose concentration but relatively low BMI, while few data points with moderate BMI but very low plasma glucose concentration levels.

They might indicate the inaccuracy and error during data entry, and there might be the observations of individuals with specific conditions that lead to anomaly in the scatter plot.

- **How does the distribution of data points differ between patients who tested positive and negative for diabetes?**

=

Positive (Dark Blue): These data points seem to be more clustered around the upper right portion of the plot indicating individuals tested positive for diabetes might have both higher plasma glucose concentrations and BMI.

Negative (Light Blue): These data points seem to be more clustered in the lower left to the center region of the plot indicating that individuals tested negative for diabetes might have lower plasma glucose concentrations and lower BMI accordingly and sometimes even relatively lower cases.

3. Findings and insights from the scatter plot.

= Evaluating the scatter plot (Plasma Glucose Concentration Vs BMI), there seems to be a weak positive correlation, indicating that as plasma glucose concentration increases, BMI also tends to increase. Diabetes being positive, these points are more concentrated on upper left portion of plot(dark-blue) while the diabetes being negative, these points are more concentrated on lower left portion of plot(light-blue).

Data points of scatterplot are roughly centered to middle of plasma glucose concentration while slightly lower than middle for BMI. There are a few outliers in the scatter plot, which include points with very high plasma glucose levels but relatively low BMI as vice-versa. In Conclusion, the scatter plot given had less predicted power, other factors should also be under consideration.