

Diabetes Dataset

What health factors matter for diabetes?

A word cloud featuring various medical and lifestyle terms in red text on a white background. The words are arranged in a non-uniform, overlapping manner. The largest word is 'glucose', positioned horizontally across the lower half. Other large words include 'bmi' (vertical), 'triglycerides' (horizontal), 'hd1' (horizontal), 'blood_pressure' (horizontal), 'hb_a1c' (horizontal), and 'ldl' (horizontal). Smaller words include 'family_history', 'diet_type', 'whr', 'hip_circumference', 'waist_circumference', 'hypertension', and 'medication_use'.

ldl

family_history

hypertension

diet_type

whr

medication_use

bmi

triglycerides

hd1

glucose

blood_pressure

hb_a1c

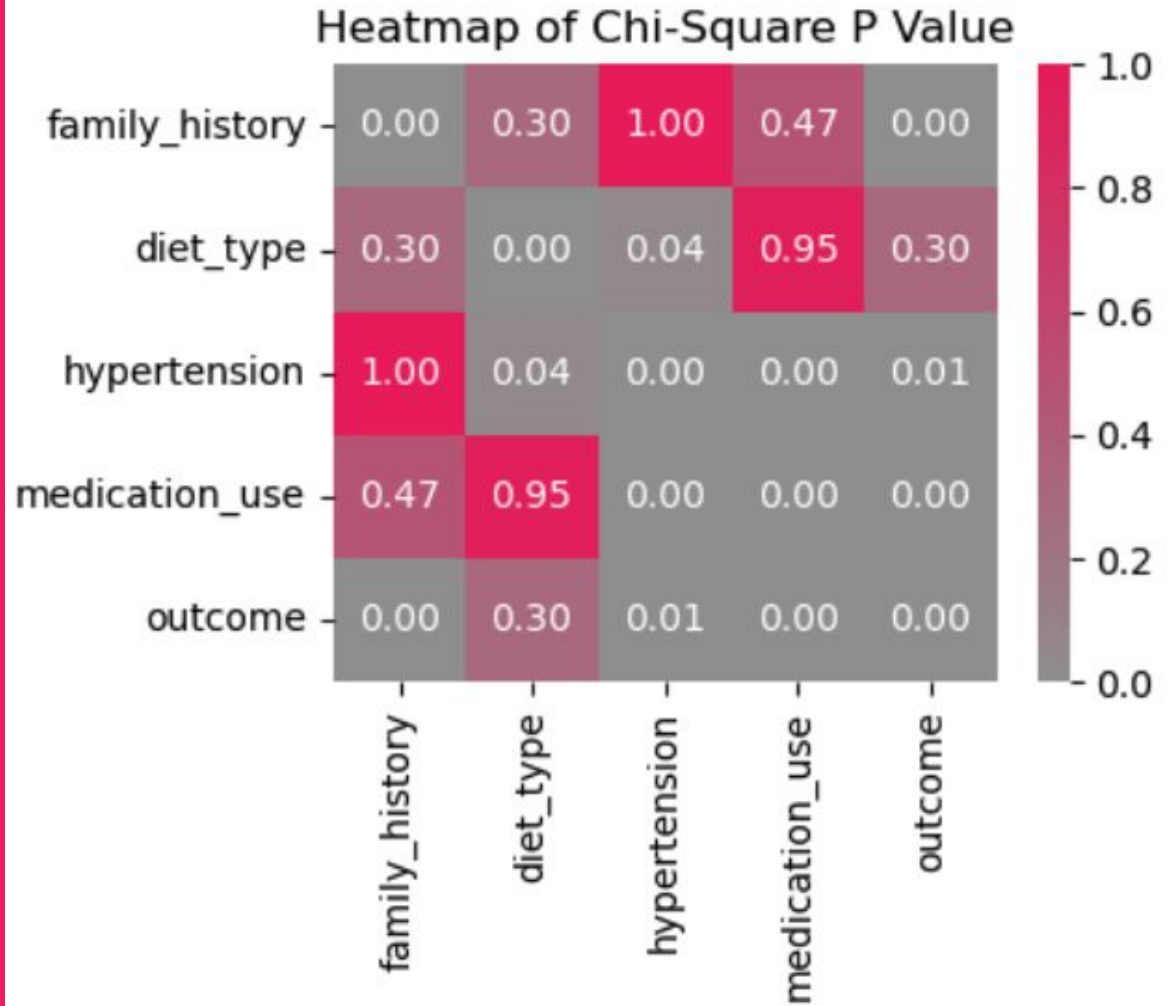
hip_circumference

waist_circumference

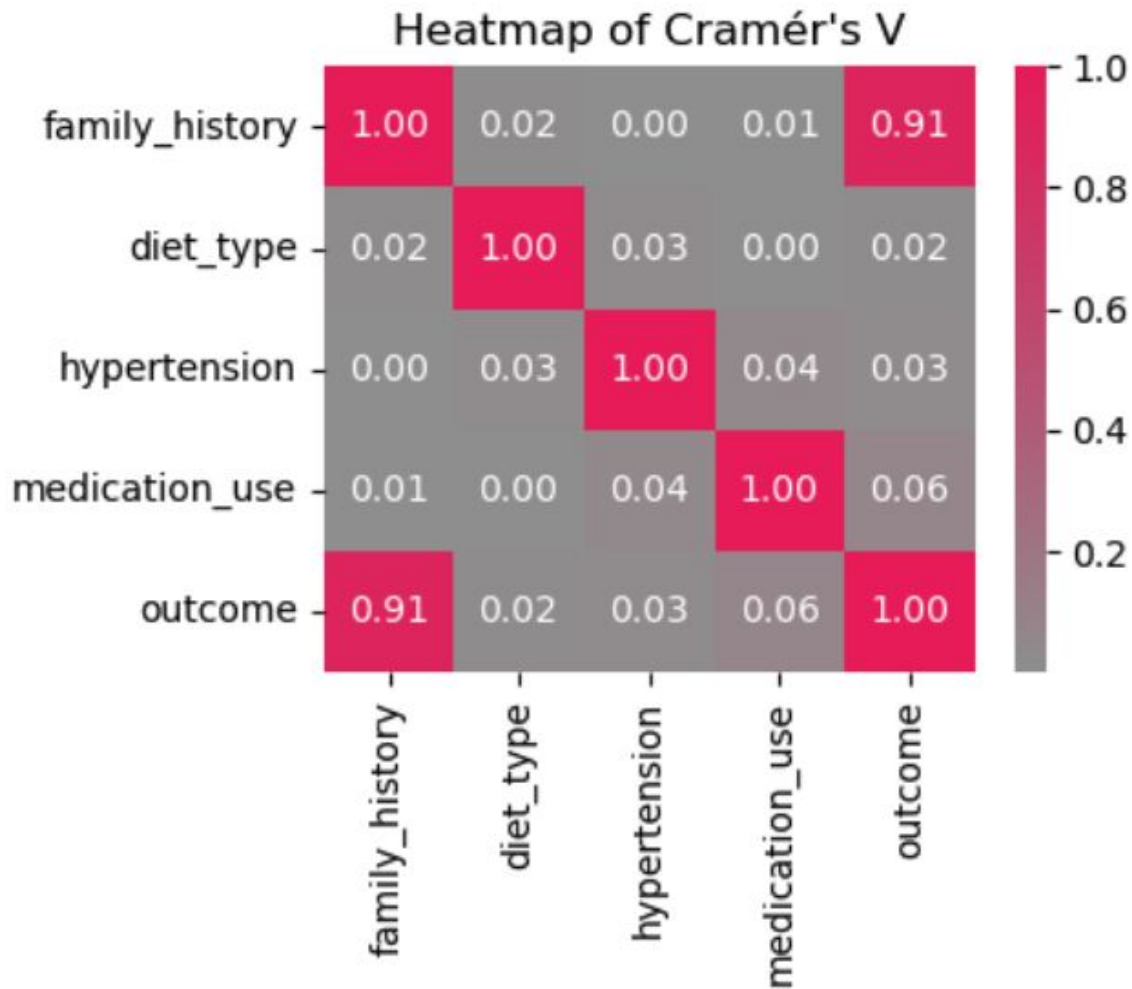
The Dataset:

- from www.kaggle.com
- 9538 patient entries
 - ~34% have diabetes (3282)
 - ~66% have no diabetes (6256)
- 17 columns check health factors, lifestyle and genetic factors for diabetes

Categorical Analysis [Chi-square p]



Categorical Analysis [Cramér's V]



family_history
outcome

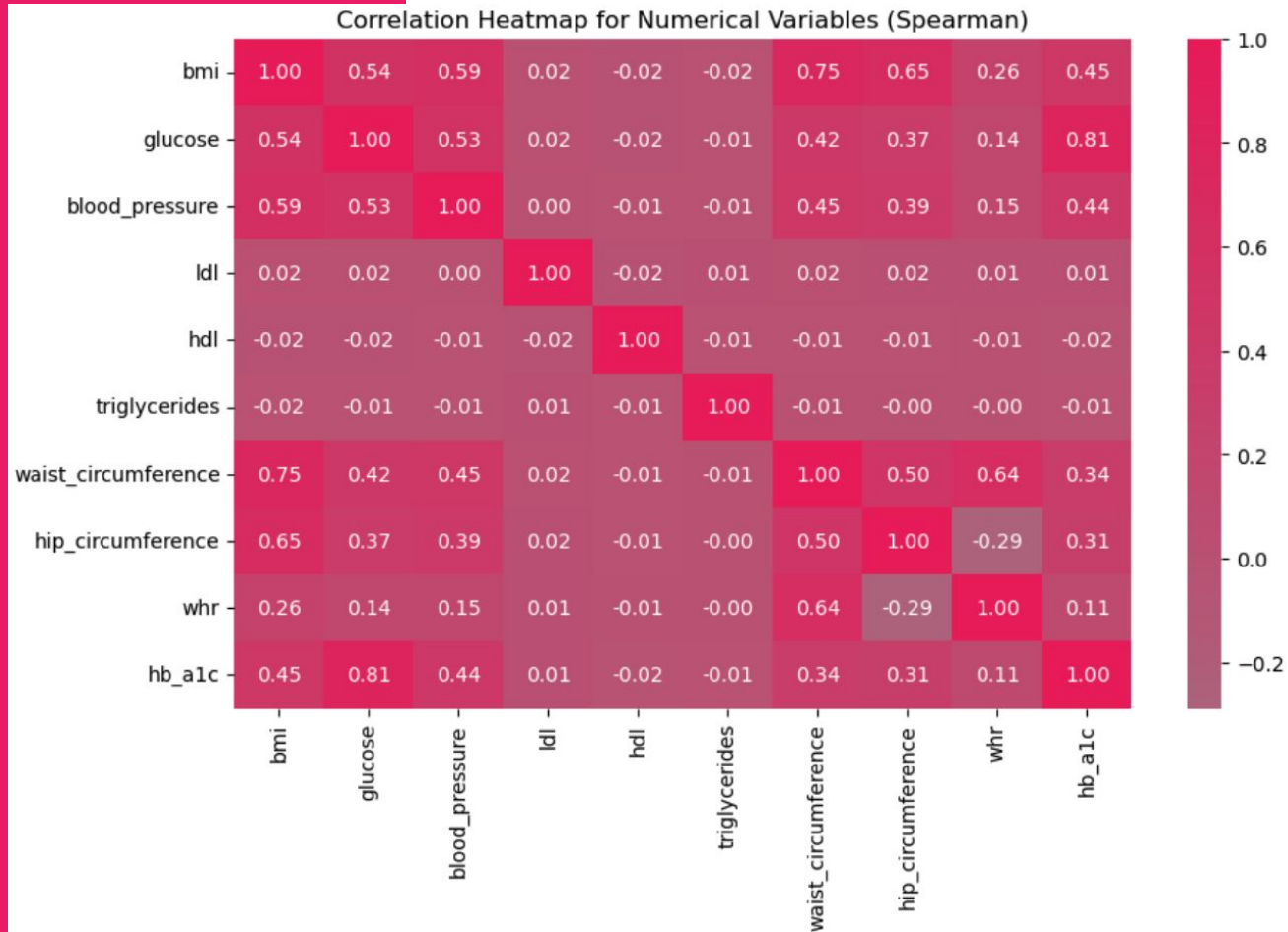
Chi-square p value is 0
Cramér's Value is 0.91

- We can reject H_0 that the two values are independent
- There is a strong association between both values

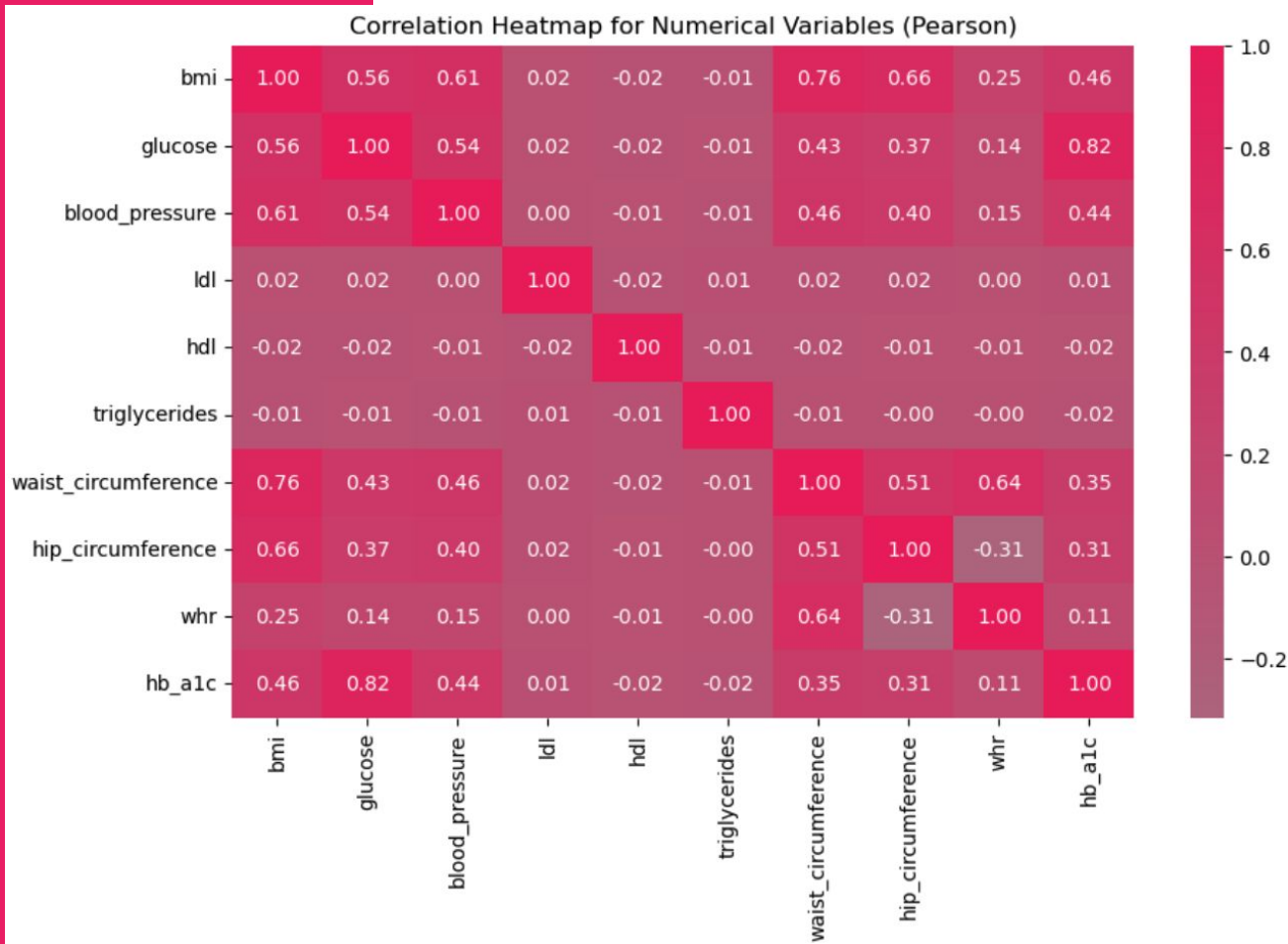
family_history outcome

		absolute_frequency	relative_frequency
Diabetes	Family History		
False	False	6256	100.0
True	False	397	12.1
	True	2885	87.9

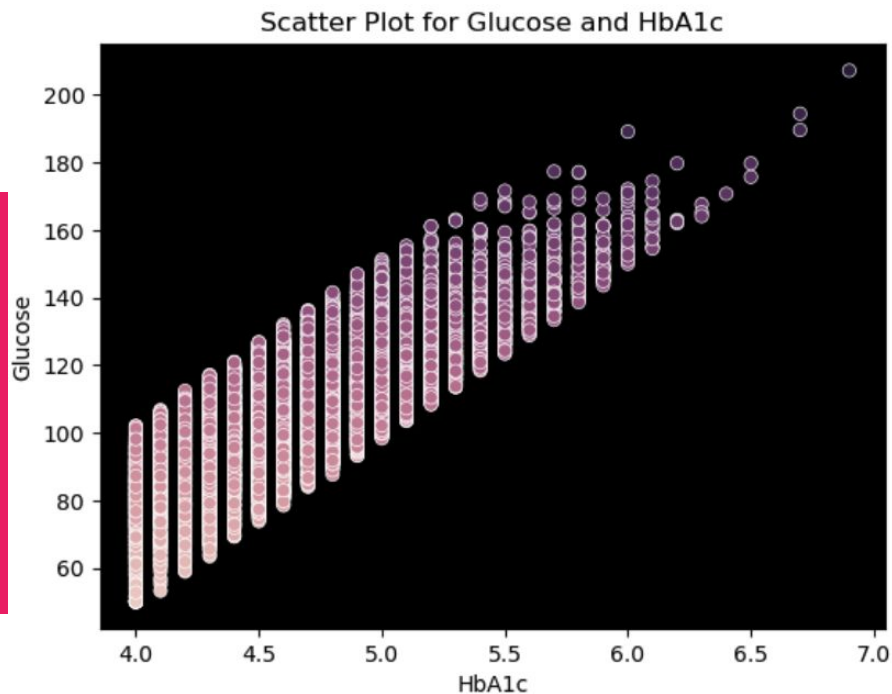
Numerical Analysis [Spearman]

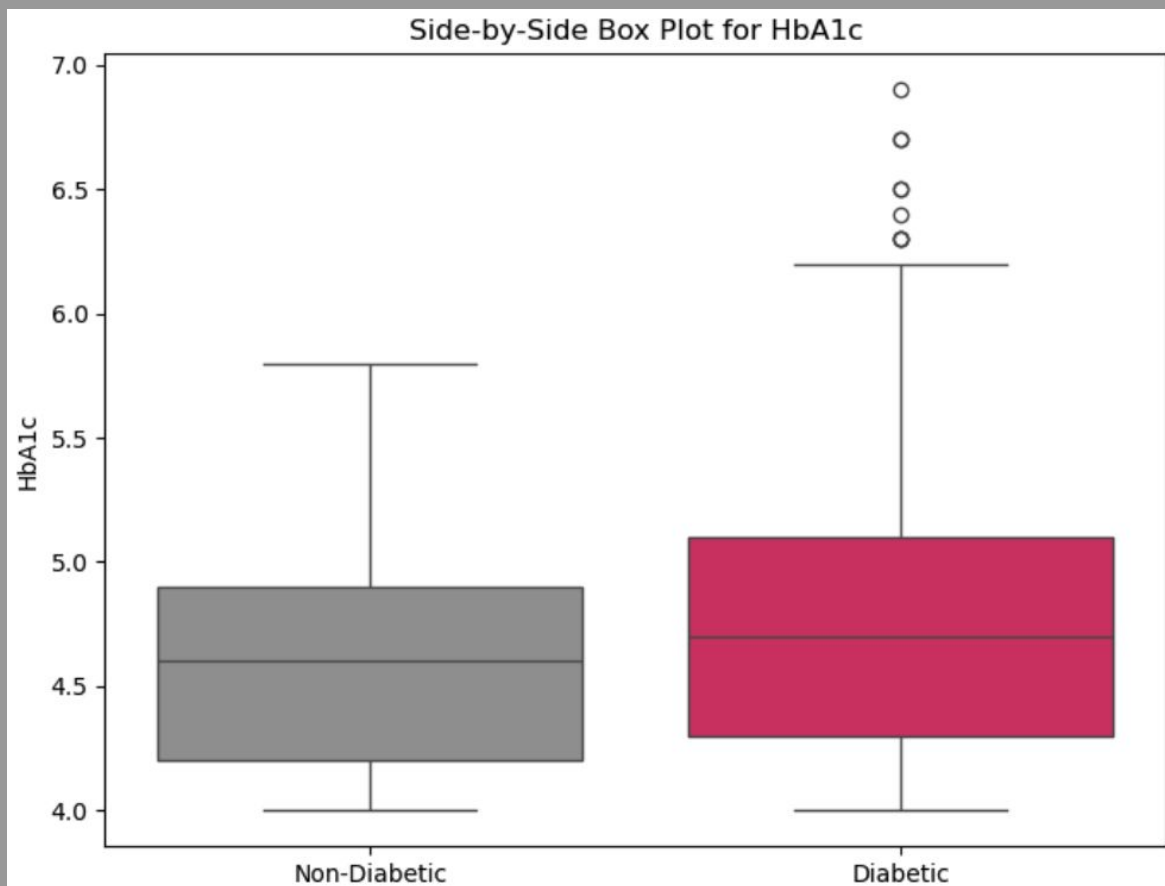


Numerical Analysis [Pearson]

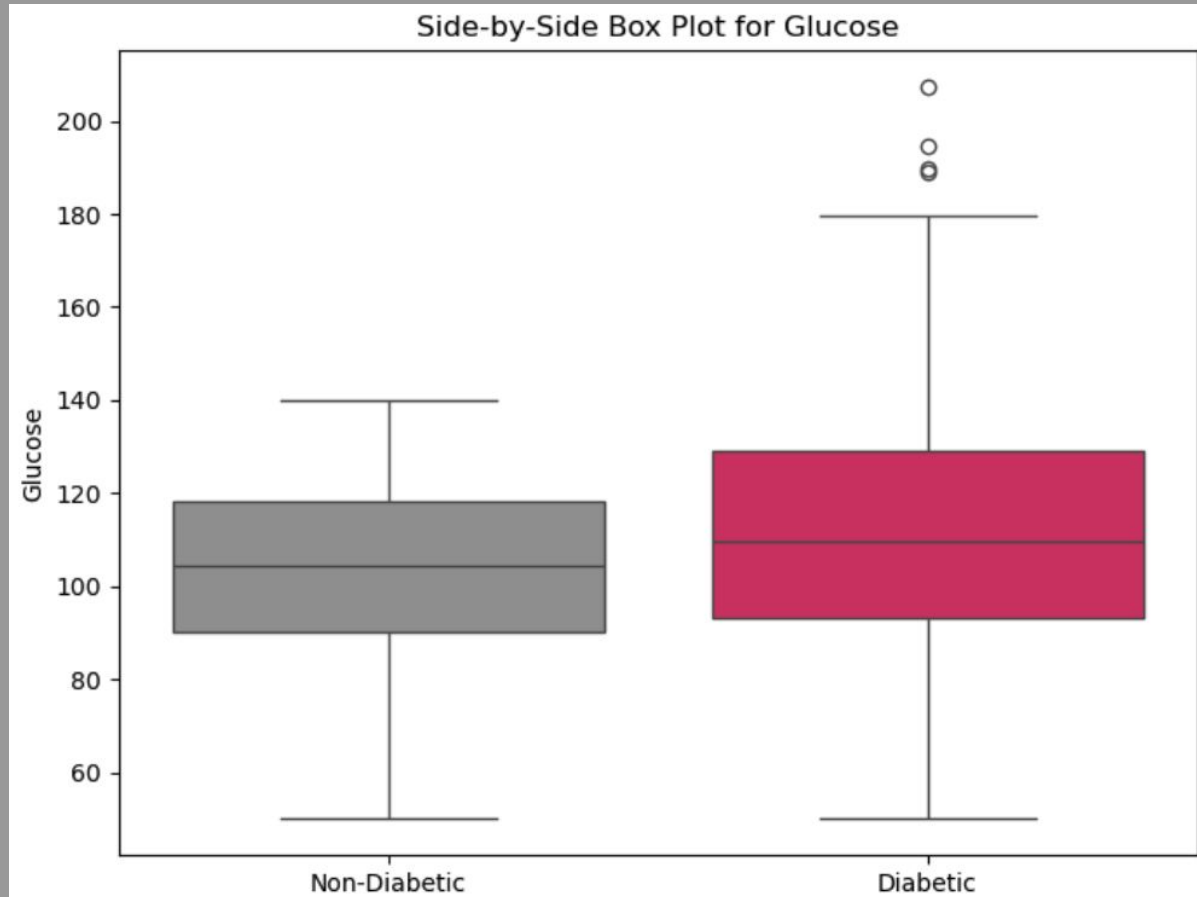


HbA1c Glucose





HbA1c



GLUCOSE

CONCLUSION

- Only Family History is strongly associated with having Diabetes
- Only Glucose is correlated with long term blood sugar HbA1c

My Wish

The outcome column to be more differentiated with Diabetes Type I or II, Pre-Diabetic etc.

A binary column did not allow to collect many results.

The End