KING'S OWN INSTITUTE*

Success in Higher Education

INDIVIDUAL ASSIGNMENT COVERSHEET

Family Name: TASNIM Given Name: SABABA

Student Number: 20028086 Lecturer's/ Tutor's Name: Dr. Firoz Anwar

Subject Code & Name: ICT370 Data Analytics T225

Assignment Title: Individual Progress Report and Reflection-Submission Week

Declaration

(This declaration must be completed by the student or the assignment will not be marked.)

I certify the following:

- I have read and understood the *Student Academic Misconduct Policy*.
- This assignment is my own work based on my personal study and or research.
- I have acknowledged all material and sources used in the preparation of this assignment including any material generated in the course of my employment.
- The assignment has not previously been submitted for assessment.
- I have not copied in part or in whole or otherwise plagiarised the work of other students.
- I have read and I understand the criteria used for assessment.
- The assignment is within the word and page limits specified in the unit outline.
- The use of any material in this assignment does not infringe the intellectual property / copyright of a third party.
- I understand that this assignment may undergo electronic detection for plagiarism, and an anonymous copy of the assignment may be retained on the database and used to make comparisons with other assignments in future.
- By completing this coversheet in full and submitting this assignment electronically, I am bound by the conditions of the KOI's *Student Academic Misconduct Policy* and the declaration on this coversheet.

sababa
Signature

20 / 08 / 2025
Date

Assignment Receipt

Family Name: TASNIM Given Name: SABABA

Student Number: 20028086 Lecturer's/ Tutor's Name: Dr. Firoz Anwar

Subject Code & Name: ICT370 Data Analytics T225

Assignment Title: Individual Progress Report and Reflection-Submission Week

sababa
Signature

20 / 08 / 2025
Date

1) Executive Summary

This report analyzes a specific real-world dataset on diabetes risk in order to address a total of five business questions: two of them are univariate, two are bivariate, and one is a multivariate analysis. The analysis synthesizes these insights into a dashboard. All descriptive charts are justified appropriately. In addition, a Week-by-Week progress log and evidence checklist are submitted to comply with rubrics.

2) Dataset & Data Dictionary

Column	Type	Description
gender	Categorical	Sex of the individual (Female, Male, Other)
age	Numeric (years)	Age in years
hypertension	Binary (0/1)	Presence of hypertension
heart_disease	Binary (0/1)	Presence of heart disease
smoking_history	Categorical	Smoking behaviour (never, current, former, etc.)
bmi	Numeric	Body Mass Index
HbA1c_level	Numeric	HbA1c % value
blood_glucose_level	Numeric	Random blood glucose measurement
diabetes	Binary (0/1)	Target flag: 1 = diabetes present, 0 = otherwise

3) Business Questions

1. What does the age distribution of the cohort look like?
2. How prevalent is diabetes in this dataset?
3. How does diabetes prevalence vary by smoking history?
4. How do glucose levels differ by diabetes status?
5. How do age and BMI interact in relation to diabetes rate (with HbA1c as an additional indicator)?

4) Analysis:

4.1. Histogram of patient ages

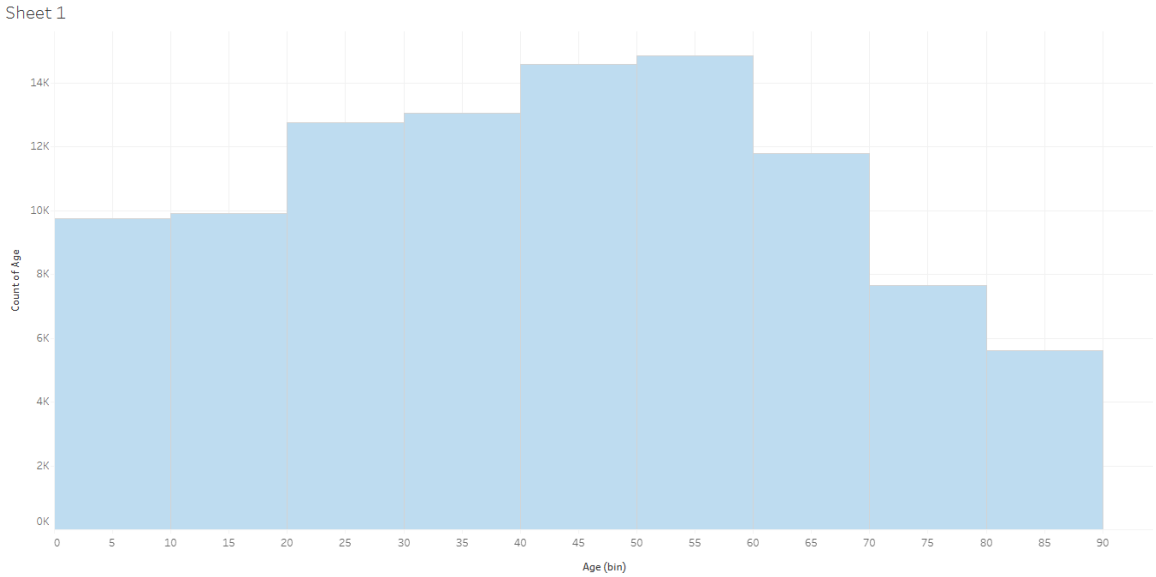

Figure 1: Histogram of age distribution of patients

4.2 Diabetes Bar

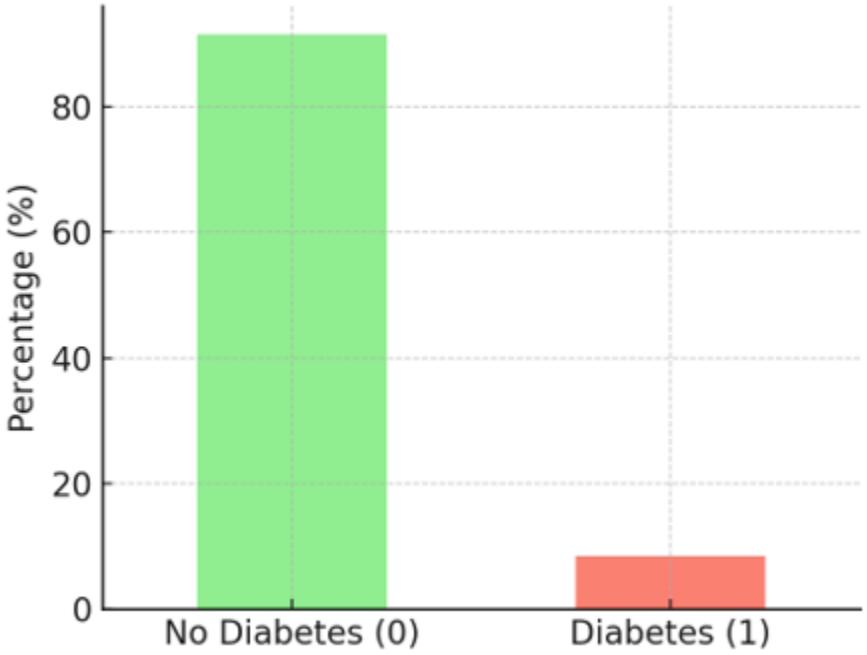

Figure 2: Bar chart showing proportion of diabetes vs non-diabetes percentage.

4.3. Diabetes by Smoking History

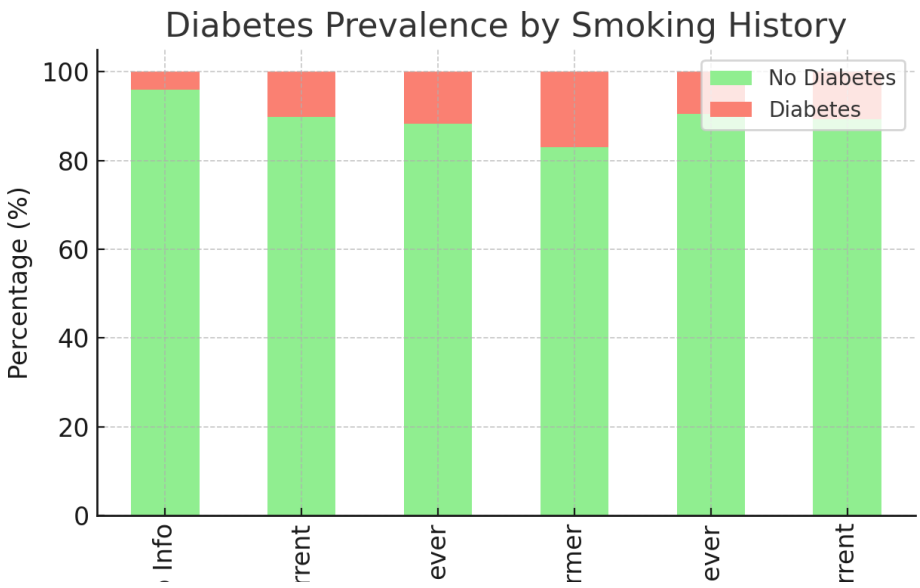

Figure 2: Stacked bar showing diabetes rates across level of smoking categories.

4.4. Blood Glucose vs Diabetes boxplot comparison

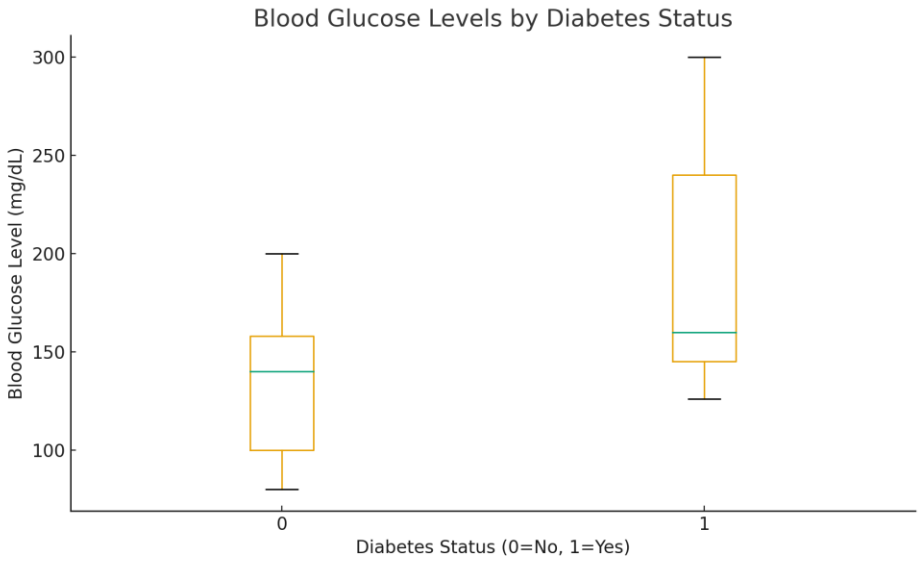

Figure 3: Boxplot showing blood glucose levels by diabetes status

4.5. Age × BMI Heatmap

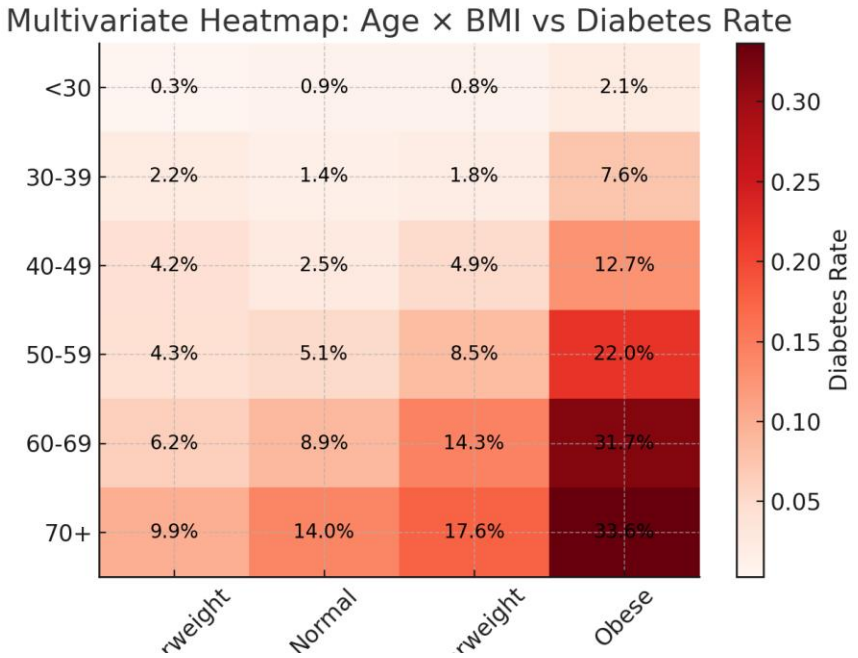

Figure 4: Heatmap of higher diabetes rates as age, weight and BMI in the dataset.

5) Dashboard

Dashboard compositions contains the below.

- KPI Card: diabetes prevalence.
- Age histogram (left) and Glucose box plot (right).
- 100% stacked bar: diabetes by smoking history.
- Heatmap: Age, BMI, and diabetes rate with HbA1c overlay.

6) Justification of Methods and Charts

- Histogram: best for numeric distribution (age).
- KPI/bar: clear for binary prevalence (diabetes).
- 100% stacked bar: proportion comparisons (smoking).
- Box plot: numeric distributions between groups (glucose).
- Heatmap: interaction effects across variables -age, BMI, HbA1c.

7) Progress Per Week

- Week 2 to 4: Defined business questions and validated columns in the dataset.
- Week 6: Dataset imported and visuals were drafted.
- Week 8: built smoking stacked bar and glucose boxplot, multivariate heatmap.

8) Limitations and Next Steps

- Data may not accurately reflect the general population.
- Smoking 'No Info' group should be treated with caution.
- Descriptive results are not causal statements.
- Next: add gender, hypertension and heart disease slicers, refine dashboard filters.

09) Conclusion & Recommendations

Analysis confirms prevalence of diabetes is higher with advancing age, increasing BMI, and smoking. Blood glucose and HbA1c continue to be clinically significant.

Recommendations:

- Target older and obese adults.
- Increase interventions for smoking cessation.
- Recommend regular screening for glucose and HbA1c levels.
- Continue to use Business Intelligence (BI) tools to maintain and monitor these dashboards.

10) References

- Sarker, I.H. (2021) 'Data science and analytics: An overview from data-driven smart computing, decision-making and applications perspective', *SN Computer Science*, 2(5), p. 377. doi:10.1007/s42979-021-00765-8.
- Charkaoui, A. and Jabraoui, S. (2024) '20 years of scientific study on business intelligence and decision-making performance: A bibliometric analysis', *Journal of Information Systems Engineering and Business Intelligence*, 10(3), pp. 408–421. doi:10.20473/jisebi.10.3.408-421.
- Ozaydin, B., Zengul, F., Oner, N. and Feldman, S.S. (2020) 'Healthcare research and analytics data infrastructure solution: A data warehouse for health services research', *Journal of Medical Internet Research*, 22(6), p. e18579. doi:10.2196/18579.