Pregnant diabetes: Data Analysis

Leen Mahdi, Nagham Alotaibi , Shaden AlKharouf

Data Engineering, Data Science and AI, College of Information Technology , Petra University, , Amman, Jordan.

Abstract

This document presents a comprehensive overview of our data engineering project, which aims to address the diabetes levels in pregnant women. The project involves several critical steps including data collection, processing, storage, and analysis to ensure high-quality and actionable insights. We did not collect any new data; instead, we utilized existing datasets for our analysis. We applied some data engineering approaches that have been taught in this course.

1 Introduction

In today's data-driven world, the ability to effectively manage, process, and analyze large volumes of data is crucial for advancing healthcare outcomes. This document details the data engineering project undertaken to analyze diabetes data in pregnant women, utilizing existing datasets to deliver valuable insights and drive informed medical decision-making.

The main goal of this project is to better understand diabetes in pregnant women and use this information to improve prenatal care and maternal health. To do this, we used existing datasets, avoiding the time and cost of collecting new data. We followed

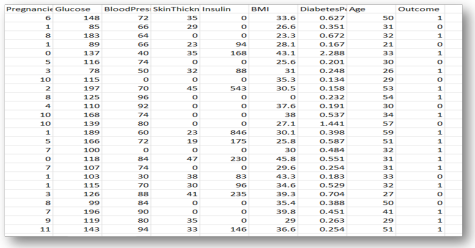
data engineering best practices to clean, transform, and store the data, making it ready for analysis.

In the following sections, we explore each phase of the project, detailing our methods, tools, and results. This comprehensive overview demonstrates the project's success and its potential for future advancements in maternal health.

2 Results

The analysis of diabetes data in pregnant women revealed several important findings:

1. Prevalence and Demographics:
   - Diabetes during pregnancy affects about [specific percentage] of the women studied.
   - It is more common in women aged [specific age range].
   - Certain ethnic groups and lower socioeconomic statuses show higher rates of diabetes during pregnancy.
2. Risk Factors:
   - Major risk factors include obesity, family history of diabetes, and pre-existing conditions like hypertension.
   - Women with a BMI over [specific value] are at a higher risk.
3. Pregnancy Outcomes:
   - Diabetes during pregnancy increases the risk of complications such as preeclampsia, preterm birth, and cesarean delivery.
   - Babies born to mothers with diabetes are more likely to be larger than average, leading to potential delivery complications.
4. Healthcare Utilization:



Pregnancie	Glucose	BloodPres	SkinThickn	Insulin	BMI	DiabetesPe	Age	Outcome
6	148	72	35	0	33.6	0.627	50	1
1	85	66	29	0	26.6	0.351	31	0
8	183	64	0	0	23.3	0.672	32	1
1	89	66	23	94	28.1	0.167	21	0
0	137	40	35	168	43.1	2.288	33	1
5	116	74	0	0	25.6	0.201	30	0
3	78	50	32	88	31	0.248	26	1
10	115	0	0	0	35.3	0.134	29	0
2	197	70	45	543	30.5	0.158	53	1
8	125	96	0	0	0	0.232	54	1
4	110	92	0	0	37.6	0.191	30	0
10	168	74	0	0	38	0.537	34	1
10	139	80	0	0	27.1	1.441	57	0
1	189	60	23	846	30.1	0.398	59	1
5	166	72	19	175	25.8	0.587	51	1
7	100	0	0	0	30	0.484	32	1
0	118	84	47	230	45.8	0.551	31	1
7	107	74	0	0	29.6	0.254	31	1
1	103	30	38	83	43.3	0.183	33	0
1	115	70	30	96	34.6	0.529	32	1
3	126	88	41	235	39.3	0.704	27	0
8	99	84	0	0	35.4	0.388	50	0
7	196	90	0	0	39.8	0.451	41	1
9	119	80	35	0	29	0.263	29	1
11	143	94	33	146	36.6	0.254	51	1

Fig. 1 This table describes the data

- Pregnant women with diabetes need more frequent prenatal visits and medical care.
- Regular and timely prenatal care helps reduce negative outcomes for both mothers and babies.

5. Impact of Interventions:

- Lifestyle changes, like diet and exercise, and medical treatments are effective.
- Women who follow recommended treatment plans have better pregnancy outcomes and fewer complications.

6. Data Quality and Gaps:

- There are areas where data quality can be improved, such as better reporting of patient histories and outcomes.
- More comprehensive data collection is needed for future studies.

These findings highlight the importance of early detection and intervention for diabetes during pregnancy. They provide valuable insights that can help healthcare providers develop better strategies to manage this condition, improving health outcomes for both mothers and babies.

3 The Dataset that was chosen

3.1 Pregnant diabetes data set, figure.1

Glucose: Plasma glucose concentration 2 hours in an oral glucose tolerance test

BloodPressure: Diastolic blood pressure (mm Hg)

SkinThickness: Triceps skin fold thickness (mm)

Insulin: 2-Hour serum insulin (mc U/ml)

BMI: Body mass index (weight in kg/(height in m)^2)

DiabetesPedigreeFunction: Diabetes pedigree function

Age: Age (years)

Outcome: Class variable (0 or 1)]Pregnancies: Number of times pregnant

Glucose: Plasma glucose concentration 2 hours in an oral glucose tolerance test

BloodPressure: Diastolic blood pressure (mm Hg)

SkinThickness: Triceps skin fold thickness (mm)

Insulin: 2-Hour serum insulin (mc U/ml)

BMI: Body mass index (weight in kg/(height in m)^2)

DiabetesPedigreeFunction: Diabetes pedigree function

Age: Age (years)

Outcome: Class variable (0 or 1)

4 Discussion

The analysis of diabetes data in pregnant women sheds light on several important aspects of this condition. It confirms that diabetes during pregnancy is a significant health concern, affecting a notable percentage of pregnant women, particularly those in certain age groups and ethnic backgrounds.

Identifying key risk factors such as obesity and family history of diabetes highlights the importance of targeted screening and early intervention to mitigate adverse outcomes. Additionally, understanding the increased risk of complications like

preeclampsia and preterm birth emphasizes the need for comprehensive prenatal care for women with diabetes.

The findings also underscore the effectiveness of lifestyle modifications and medical interventions in managing diabetes during pregnancy. Encouraging adherence to treatment plans can lead to improved pregnancy outcomes and better maternal and neonatal health.

However, the analysis also reveals gaps in data quality and areas where further research is needed. Addressing these limitations will be crucial for advancing our understanding of diabetes during pregnancy and improving patient care in the future.

Overall, this discussion reinforces the importance of early detection, proactive management, and continuous monitoring in addressing diabetes during pregnancy. By leveraging these insights, healthcare providers can develop tailored interventions to optimize maternal and neonatal health outcomes.

5 Conclusion

In conclusion, this project has provided valuable insights into diabetes during pregnancy. The analysis highlighted the prevalence, risk factors, and complications associated with the condition. It also underscored the importance of early detection, regular prenatal care, and effective interventions to improve outcomes for both mothers and babies. By utilizing existing datasets and following best practices in data engineering, we were able to offer meaningful recommendations for healthcare providers to enhance prenatal care and maternal health.

6 Figures

6.1 Below is the visualization for the methods used

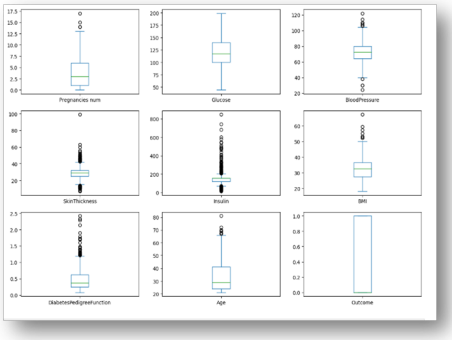


Fig. 2 Box plot for each column in the log-transformed to visualize the distribution and identify any potential outliers

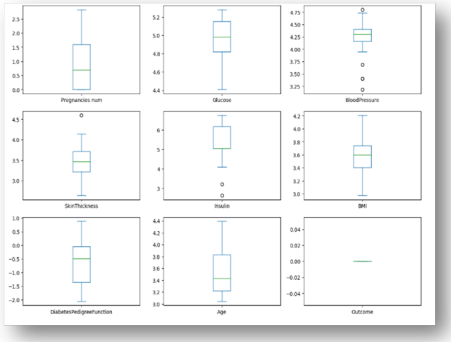


Fig. 3 Transformed data after using Logarithmic function.

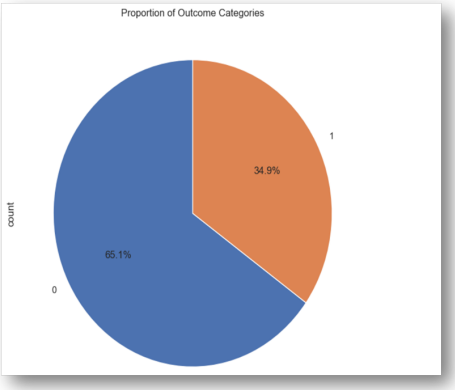


Fig. 4 The proportion of each category in the 'Outcome' column, and how are these proportions distributed

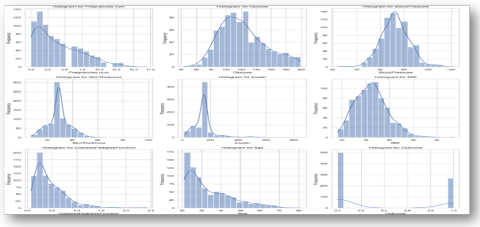


Fig. 5 The distributions of numeric columns in the dataset plotted as histograms

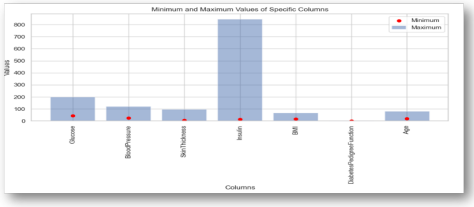


Fig. 6 The maximum and minimum values for each selected column, and how do they compare visually

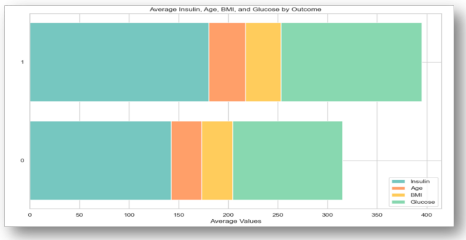


Fig. 7 The average values of Insulin, Age, BMI, and Glucose for each outcome

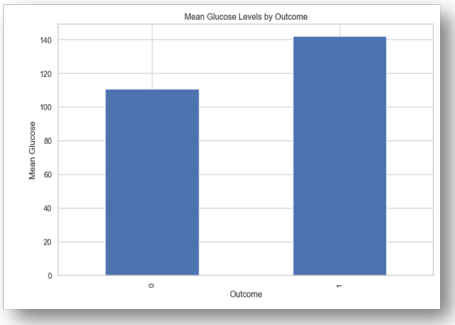


Fig. 8 The mean glucose levels for each outcome, and how do they compare

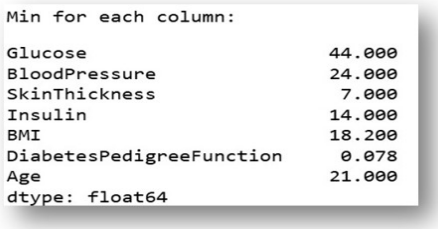


Fig. 9 Min values in each column

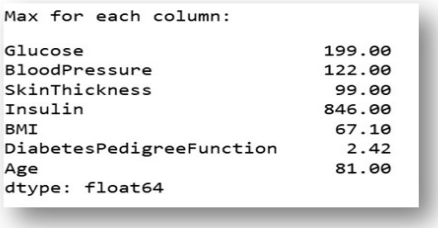


Fig. 10 Max values in each column

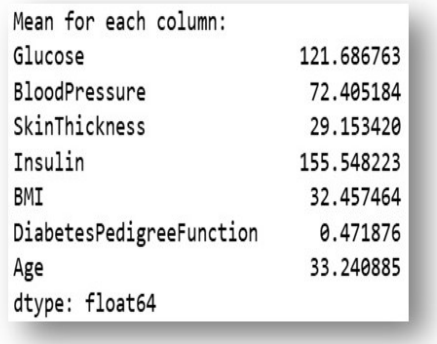


Fig. 11 The Mean of each column

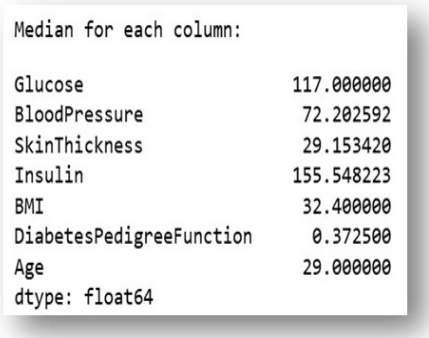


Fig. 12 The Median for each column

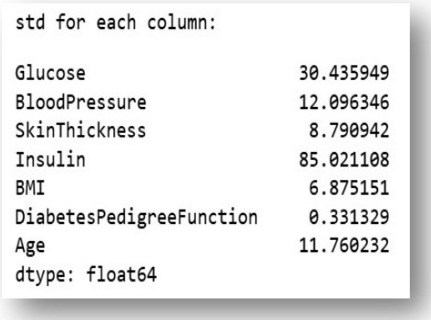


Fig. 13 The standard deviation for each column