Journal of Data Analysis and Information Processing, 2024, 12, 232-247
https://www.scirp.org/journal/jdaip
ISSN Online: 2327-7203
ISSN Print: 2327-7211

Analysis of Gestational Diabetes Mellitus (GDM) and Its Impact on Maternal and Fetal Health: A Comprehensive Dataset Study Using Data Analytic Tool Power BI

Shahistha Jabeen Hashim, Arthur McAdams

Department of Technology Management, University of Bridgeport, Bridgeport, USA
Email: shahistha941@gmail.com

How to cite this paper: Hashim, S.J. and McAdams, A. (2024) Analysis of Gestational Diabetes Mellitus (GDM) and Its Impact on Maternal and Fetal Health: A Comprehensive Dataset Study Using Data Analytic Tool Power BI. *Journal of Data Analysis and Information Processing*, **12**, 232-247.
https://doi.org/10.4236/jdaip.2024.122013

Received: March 18, 2024
Accepted: May 24, 2024
Published: May 27, 2024

Abstract

Gestational Diabetes Mellitus (GDM) is a significant health concern affecting pregnant women worldwide. It is characterized by elevated blood sugar levels during pregnancy and poses risks to both maternal and fetal health. Maternal complications of GDM include an increased risk of developing type 2 diabetes later in life, as well as hypertension and preeclampsia during pregnancy. Fetal complications may include macrosomia (large birth weight), birth injuries, and an increased risk of developing metabolic disorders later in life. Understanding the demographics, risk factors, and biomarkers associated with GDM is crucial for effective management and prevention strategies. This research aims to address these aspects comprehensively through the analysis of a dataset comprising 600 pregnant women. By exploring the demographics of the dataset and employing data modeling techniques, the study seeks to identify key risk factors associated with GDM. Moreover, by analyzing various biomarkers, the research aims to gain insights into the physiological mechanisms underlying GDM and its implications for maternal and fetal health. The significance of this research lies in its potential to inform clinical practice and public health policies related to GDM. By identifying demographic patterns and risk factors, healthcare providers can better tailor screening and intervention strategies for pregnant women at risk of GDM. Additionally, insights into biomarkers associated with GDM may contribute to the development of novel diagnostic tools and therapeutic approaches. Ultimately, by enhancing our understanding of GDM, this research aims to improve maternal and fetal outcomes and reduce the burden of this condition on healthcare systems and society. However, it's important to acknowledge the limitations of the dataset used in this study. Further research utilizing larger and more

diverse datasets, perhaps employing advanced data analysis techniques such as Power BI, is warranted to corroborate and expand upon the findings of this research. This underscores the ongoing need for continued investigation into GDM to refine our understanding and improve clinical management strategies.

Keywords

Gestational Diabetes, Visualization, Data Analytics, Data Modelling, Pregnancy, Power BI

1. Introduction

Gestational Diabetes Mellitus (GDM) is a significant concern during pregnancy due to its potential health implications for both the mother and the fetus. GDM, or gestational diabetes mellitus, is a condition characterized by irregular glucose levels during pregnancy. It is a prevalent complication, affecting 3% - 10% of pregnancies. Typically diagnosed between the 22nd and 26th weeks of gestation, GDM poses high-risk implications for both expectant mothers and infants. Potential complications include respiratory issues, metabolic disorders, premature delivery, and excessive fetal weight gain, which may complicate the birthing process. While GDM usually resolves after childbirth, women remain at an increased risk of developing type 2 diabetes, with a cumulative incidence ranging from 30% - 50% within 5 - 10 years following the initial pregnancy. Numerous studies suggest that early medical intervention in the first or second trimester can prevent high-risk complications associated with GDM [2].

Data cleaning is an essential prerequisite in the data analysis process to ensure the reliability and accuracy of results. This research paper presents a detailed account of the data cleaning steps undertaken on a dataset, highlighting the techniques employed to address various data quality issues. The dataset, consisting of medical and fetal information, was subjected to rigorous cleaning, restructuring, and transformation. The methodology and strategies implemented are outlined, with each step explained and exemplified. The paper also emphasizes the importance of data preprocessing in enhancing the overall quality of the dataset for subsequent analysis. The dataset consists of 600 pregnant women, with 74 diagnosed with GDM, 7 as prediabetic, and the remainder non-diabetic. Key demographic insights include age, race, and BMI distribution among the participants. All biomarkers with description and ranges from the dataset are listed in **Table 1**. In this investigation of Gestational Diabetes (GD), we leverage the capabilities of Microsoft Power BI to unearth valuable insights from the wealth of data within our organization. Power BI acts as a powerful tool that seamlessly connects various data sources, allowing us to bring together disparate sets of information.

Table 1. Data definitions of the columns used in dataset.

Field	Description	Ranges/Category
Participant ID	ID of Patients Visited	600 Patients
25OHD checked yes (1) or no (0)	25(OH)D is the only vitamin D metabolite that is used to determine whether a patient is vitamin D deficient, sufficient or intoxicated.	0-Not checked; yes-checked; Null-Not filled
Date Form Signed	Filled the details on these dates	Year-2013, 2014
Systolic BP (mmHg)	Systolic pressure is the maximum blood pressure during contraction of the ventricles.	Normal SBP/DBP: ≤120/80 mmHg; Prehypertension: 120 - 139/80 - 89 mmHg; and Hypertension: ≥140 and/or ≥90 mmHg
Diastolic BP (mmHg)	The diastolic reading, or the bottom number, *is the pressure in the arteries when the heart rests between beats*.	Normal-60 to 90; Hypertension > 90; Hypotension < 60
Pulse (bpm)	A pulse refers to the rhythmic expansion and contraction of an artery as blood is pumped through it by the heart.	Normal 60 to 100 bpm; Tachycardia > 100
Weight (kg)	Weight of the patient measured in kilograms	
Height (m)	Height of the Patient measured in meters	
BMI (kg/m²)	Body mass index (BMI) is a person's weight in kilograms divided by the square of height in meters.	Below 18.5. - Underweight; 18.5 - 24.9: Healthy Weight; 25.0 - 29.9: Overweight; ≥30.0: Obesity
Smoking 123	Smoking status of the patient	1-Current smoker 2-Ex smoker 3-Never smoked
Alcohol Intake	Alcohol use by the Patient. It is in units/week	1-took current week 0-didn't take current week
Ethnicity	Race of Patients	1 = White, 2 = mixed, 3 = Asian. 4 = black 5 = Chinese, 6 = other
GDM performed, GDM V1 & V2, Dx with GDM	Previous Gestational diabetes mellitus (GDM), if patient has a history for GDM. GDM screening is a common test performed during pregnancy to detect GDM, a type of diabetes that can develop during pregnancy. It is recommended that all pregnant women be screened for GDM between 24 and 28 weeks of pregnancy. Dx stands for Diagnosis.	normal = level below 140 mg/dL, glucose tolerance = 140 mg/dL to less than 190 mg/dL, gestational diabetes = 190 (mg/dl) and higher
GCT, GCT 1 h value > 7.5 mM, GCT 1 h value > 7.8 mM, GCT 1 h value > 7.2 mM, GCT 1 0 screening test.	The Glucose Challenge Test is a screening test used to detect Gestational Diabetes in pregnant women. The results have been taken if the GCT value > 7.5 mM, GCT 1 h value > 7.8 mM, GCT 1h value > 7.2 mM 1 hour after drinking the glucose solution.	If yes then 1, else 0
OGTT 0 H Value, OGTT 1H, OGTT 2 H.	The oral glucose tolerance test (OGTT) is one method of evaluating an individual's apparent insulin sensitivity and insulin resistance. This is done at visit 3. For this fasting is done before test.	Fasting glucose: less than 100 mg/dL (5.6 mmol/L) Normal 1 hour glucose: less than 200 mg/dL (11.1 mmol/L) Normal 2-hour glucose: less than 140 mg/dL (7.8 mmol/L) Normal.
PMHx	Previous Medical History	

Continued

Chronic Illness	History of Chronic diseases in patients, such as heart disease, cancer, diabetes, stroke, and arthritis	
Medications	Whether the patient is having any medications or not currently	Names of medications given
Hx GDM, glycosuria, 1st degree rel	1st degree rel - History of GDM in first degree relative of patient Hx GDM - History of GDM in patient Glycosuria is a condition where glucose is detected in urine at higher concentrations than normal.	Glucose concentration of 0.3 to 0.5 g/L (1500 to 2500 mg/L) in a random urine sample is considered within the normal range for glycosuria during pregnancy.
High Risk 10	Whether the patient considered as high risk or not based on Hx GDM, glycosuria, 1st degree rel	If yes then 1, else 0
Screening method	OGTT: Oral glucose tolerance test-75 g glucose load GCT: Glucose Challenge Test-50 g glucose load	1 = OGTT 2 = GCT
Vit D List used	Maternal vitamin D deficiency during pregnancy appears to be associated with an increased risk of gestational diabetes. Patients were asked usage of vitamin D supplements.	Yes/No
1st DASS score > 33	The DASS score refers to the Depression, Anxiety and Stress Scale, which is a self-reported questionnaire that assesses the severity of symptoms of depression, anxiety and stress in individuals.	A score of 33 or above on any of the DASS scales is generally considered to be in the "severe" range. If DASS > 33 then 1, else 0
Ref to Psych	If the patient has been referred to a psychiatrist	If yes then 1, else 0
Gestational Age, LMP	During pregnancy, gestational age (GA) is typically estimated based on the woman's last mensural period (LMP) and confirmed by ultrasound measurements.	GA is provided in weeks
EDD V1, US EDD	EDD in gestational diabetes refers to the estimated due date at the first antenatal visit during pregnancy. It is calculated based on the last menstrual period. US EDD refers to Ultrasound Estimated Due Date.	The actual date of delivery can vary, and it's considered normal for a baby to be born between 37 and 42 weeks of pregnancy.
V1 HbA1c (mmol/mol)	HbA1c (mmol/mol) refers to the measurement of the level of glycosylated hemoglobin (HbA1c) which shows the glucose level management for the past 3 months.	Diabetes > 48 mmol/mol (6.5%), normal values < 5.4% in the 1st and 2nd trimesters and values < 5.7% in the 3rd trimester
WCC	WCC stands for White Cell Count, which is a blood test that measures the number of white blood cells in the bloodstream.	WCC within the range of 5000 to 15,000 cells per microliter (μL) is considered normal during pregnancy.
Hb	Hb refers to the hemoglobin level in patients to detect anemia or other blood-related health issues.	First Trimester = 11.6 - 13.9 g/dL; Second Trimester = 9.7 - 14.8 g/dL; Third Trimester = 9.5 -15 g/dL
Platelets	Platelets are a type of blood cell that helps with clotting, and their count may be used to detect potential blood-related health issues during pregnancy.	Normal: 140,000 or 150,000 platelets per μL; below 100,000 thrombocytopenia; below 50,000 platelets, risk of excessive bleeding during delivery
Creatinine	Creatinine is a waste product that is filtered out of the blood by the kidneys, and its level in the blood may be used to detect potential kidney-related health issues during pregnancy.	1st Trimester = 0.4 - 0.7 mg/dL; 2nd Trimester = 0.4 - 0.8 mg/dL; 3rd Trimester = 0.4 - 0.9 mg/dL
Calcium	Calcium is an essential mineral that is important for many bodily functions, including bone health, muscle function, and nerve transmission.	Normal, 2.12 to 2.62 millimoles per liter (mmol/L)

Continued

Albumin	Albumin is a protein that is produced by the liver and helps to maintain fluid balance in the body. Its level in the blood may be used to detect potential health issues during pregnancy, including pre-eclampsia.	Normal: 35 to 55 grams per liter (g/L)
ALT	ALT (Alanine Aminotransferase) is a liver enzyme. Increased activities of ALT are an indicator of hepatocellular injury and associated with insulin resistance, metabolic syndrome, and type 2 diabetes.	1st Trimester (V1): 3 - 30 U/L, 2nd Trimester (V2): 2 - 33, 3rd Trimester (V3): 2 - 25
CRP	CRP is a pentameric protein synthesized by the liver, whose level rises in response to inflammation.	First Trimester (V1): Not reported Second Trimester (V2):0.4 - 20.3 Third Trimester (V3): 0.4 - 8.1
U albumin	Albuminuria, or the presence of high levels of albumin in urine, is a sign of kidney damage that may indicate an increased risk of developing complications during pregnancy.	less than 30 mg/g creatinine.
U protein	U protein, also known as urine protein or proteinuria, refers to the measurement of protein in urine. Proteinuria is a sign of kidney damage that may indicate an increased risk of developing complications during pregnancy.	proteinuria > 300 mg/d was associated with preterm birth, preeclampsia with severe features, and birthweight < 10th percentile.
U creatinine	U creatinine, also known as urine creatinine, refers to the measurement of creatinine in urine.	Normal: 7 to 13 millimoles per liter (mmol/L).
PCR	PCR creatinine, also known as urine protein-to-creatinine ratio, refers to the measurement of protein in urine, normalized to the concentration of creatinine in urine.	≥0.3 mg/mg Cr
EDD estimation method	2 methods: CRL: Crown_Rump_Length. BPD: Biparietal diameter is the cross-sectional diameter of the skull, also known as the fetal head circumference	CRL/BPD
Prescription for Caltrate sent?	This column talks about if the patient is prescribed Calcium supplementation (Caltrate) or not	If yes then 1, else 0
Date Prescription sent?	Date at which the supplementation was prescribed	Date Format
Miscarriage before 28/40	If the patient had miscarriage before the 28th week.	If yes then 1, else 0
Withdrew before 28/40	If the patient withdrew from participating in this research study before 28 weeks	If yes then 1, else 0
Screening DNA	DNA screening is a test that can determine if a woman has a higher chance of having a fetus with Down syndrome.	If yes then 1, else 0
Date Screening	Screening date	Date Format
1h glucose	Glucose levels one hour after drinking the glucose solution	1 hour glucose: less than 200 mg/dL (11.1 mmol/L) Normal
GCT Ref to DNS?	Whether the GCT abnormal tested patient was referred to Diabetes Nurse Specialist (DNS)	If yes then 1, else 0
OGTT Ref to DNS	Whether the OGTT abnormal tested patient was referred to Diabetes Nurse Specialist (DNS)	Yes-done; No-Not Done

Continued

Delivered before 36/40 10	If the delivery was done before 36 weeks of GA.	1-yes 0-no
Abdominal Circumference	it is the measurement of the distance around a woman's abdomen at the level of her belly button, and it is used to estimate fetal growth during pregnancy.	Data given in cm
***Weight change* (%)**	% weight change from visit 1 to visit 3	%
***BP change* (%)**	% change in bp from visit 1 to visit 3	%
***HbA1c Change* (%), *HbA1c rise* > 10%, *HbA1c rise* > 5%**	% change in HbA1c from visit 1 to visit 3	If there is change, yes = 1/else no = 0
***Hb change* (%)**	% change in Hb from visit 1 to visit 3	%
***Create change* (%)**	% change in Creatinine from visit 1 to visit 3	%
***ALT change* (%)**	% difference in ALT levels in visit 1 and 3	%
***CRP change* (%)**	% difference in CRP levels in visit 1 and 3	%
***PCR change* (%)**	% difference in PCR in visit 1 and 3	%
Date of Data Collection	When was data collected: V1-Visit 1 (First Trimester); V2-Visit 2 (Second Trimester); V3-Visit 3 (Third Trimester); Fetal data – Visit 4 (after delivery)	Date Format
Attended GDM clinics	If the patient attended GDM clinic or not.	1-yes 0-no
Diagnosed with Vit D Deficiency	If the patient has deficiency of Vitamin D	1-yes 0-no
Took Vit D Supplements	If the patient takes Vit D supplements	1-yes 0-no
Number obstetric clinics	Obstetrics Clinic is a clinic that specializes in the care of women during pregnancy and childbirth and in the diagnosis and treatment of diseases of the female. This column talks about the number of times the patient visits these clinics.	Range in our dataset is from 1 - 10
Number GDM clinics	The number of times the patient visits the gestational diabetes clinic	Range in our dataset is from 1 - 8
Nutritional counselling	If the patient received nutritional counselling.	Yes/No 1 = Yes 0 = No
Glucose lowering therapies	If the patient takes any blood glucose lowering therapies	Yes/No 1 = Yes 0 = No
Insulin/Metformin	use of insulin or metformin for lowering the glucose levels.	Insulin/Metformin response
Gestational hypertension	Gestational hypertension is a condition happens when you only have high blood pressure during pregnancy and do not have protein in your urine or other heart or kidney problems.	Higher that 140/90 mmHg
Pre-eclampsia	Preeclampsia is a complication of pregnancy. With preeclampsia, you might have high blood pressure, high levels of protein in urine that indicate kidney damage (proteinuria), or other signs of organ damage.	Yes/No 1 = Yes 0 = No

Continued

Eclampsia	Eclampsia is the new onset of seizures or coma in a pregnant woman with preeclampsia. These seizures are not related to an existing brain condition.	Yes/No 1 = Yes 0 = No
HELLP	HELLP (Hemolysis, Elevated Liver enzymes and Low Platelets) syndrome is a life-threatening pregnancy complication usually considered to be a variant of preeclampsia.	Yes/No 1 = Yes 0 = No
Induction	Labor induction is prompting the uterus to contract during pregnancy before labor begins on its own for a vaginal birth.	If induction done = 1; else 0
Caesarean, LSCS	A C-section is an operation to deliver the baby through the mother's belly. A woman who has diabetes that is not well controlled has a higher chance of needing a C section,	If C section done = 1; else 0
Emergency	Whether the delivery was emergency or not	If emergency = 1; else 0
Instrumentation	Instrumental delivery is delivery by emergency C Section	If instrumentation = 1; else 0
Still-birth	Stillbirth is when a baby dies in the womb after 20 weeks of pregnancy.	If still birth = 1; else 0
Twins 10	Whether the patient is having twins	If twin birth = 1; else 0
Birth weight (kg)	Most babies born between 37 and 40 weeks weigh somewhere between 5 pounds, 8 ounces (2500 grams) and 8 pounds, 13 ounces (4000 grams). With gestational diabetes, higher chances of higher baby birth weight.	BW >5 pounds, 8 ounces: High birth weight BW< 3 pounds, 5 ounces: very low birthweight
Twins weight	Average twin weighs 5.5 pounds	No Data
Apgar 1 minute, Apgar 3 minutes	Apgar is a quick test performed on a baby at 1 and 5 minutes after birth. The 1-minute score determines how well the baby tolerated the birthing process. The 5-minute score tells the health care provider how well the baby is doing outside the mother's womb.	The Apgar score will range from 0 to 10. The higher your baby's Apgar score, the better they're doing after birth.
Birth Injury	Birth injury is defined as an impairment of the neonate's body function or structure due to an adverse event that occurred at birth.	Yes/No
Shoulder dystocia	Shoulder dystocia happens when one of the baby's shoulders gets stuck behind the mother's pubic bone (the bone behind the pubic hair) or sacrum (the bone at the back of the pelvis, above the tailbone) during birth.	Yes/No
Brachial plexus injury	An injury to the brachial plexus nerves means that brachial plexus may be stretched, compressed, or torn in a difficult delivery.	Yes/No
Other nerve injury	Nerve damage can occur when there is compression of the peripheral nerves during childbirth. carpal tunnel syndrome, femoral neuropathy, and post partial foot drop.	Yes/No
Clavicular fracture	A fractured clavicle in the newborn is a broken collar bone in a baby that was just delivered.	Yes/No
Skull fracture	maternal diabetes, including gestational diabetes, was associated with an increased risk of birth trauma, including skull fractures, in offspring.	Yes/No

Continued

Humeral fracture	Humeral fractures are the second common long bone fractures of the neonatal period after clavicle	Yes/No
Other birth injury	Cerebral palsy, caput succedaneum and cephalohematoma, Neonatal Intracranial Hemorrhage, hydrocephalus, cervical Dystonia, hemiplegia, Hemorrhagic Stroke, Neonatal Stroke, HIE, Infant Seizures.	Yes/No
Antenatal steroid use	Antenatal steroid use refers to the administration of corticosteroids to pregnant women before delivery to promote fetal lung maturation.	Yes/No
Perinatal death	Perinatal death refers to the death of a fetus or newborn during the perinatal period, which includes the time from 20 weeks of gestation through the first week of life.	Yes/No
Fetal hypoglycemia	Infants born to women with poorly controlled gestational diabetes were more likely to experience fetal hypoglycemia (low glucose levels).	Yes/No
Fetal jaundice	Fetal jaundice, also known as neonatal hyperbilirubinemia, refers to a condition where a newborn baby's skin and eyes appear yellow due to a high level of bilirubin in the blood.	Yes/No
Fetal phototherapy	Fetal phototherapy is a treatment for neonatal hyperbilirubinemia, a condition where a newborn baby's skin and eyes appear yellow due to a high level of bilirubin in the blood.	Yes/No
Total bilirubin	Total bilirubin is a measure of the amount of bilirubin in the blood, which is an important marker for neonatal jaundice or hyperbilirubinemia.	Yes/No
Surfactant use	Surfactant is a substance that helps to reduce surface tension in the lungs, allowing them to expand and contract more easily during breathing.	If surfactant used, Yes/else, no
RR>60	Respiratory rate (RR) is a measure of the number of breaths a person takes per minute.	RR > 60 is a sign of respiratory distress or failure
SCBU	It is Special Care Baby Unit for the care and treatment of newborn babies that are ill or premature.	Yes/No
Cong malformation	Congenital malformation refers to physical or structural abnormalities that occur in a fetus during pregnancy. These abnormalities can affect any part of the body, including the heart, brain, limbs, or internal organs.	Yes/No
Spina bifida	Spina bifida is a birth defect that occurs when the spinal cord and surrounding tissues do not develop properly in the womb.	Yes/No
Cleft palate	Cleft palate is a congenital condition in which there is an opening or split in the roof of the mouth	Yes/No
Cleft Lip	Cleft lip is a congenital deformity in which the upper lip fails to fuse properly during fetal development, resulting in a split or opening in the lip	Yes/No
Hydrocephalus	Hydrocephalus is a condition characterized by an accumulation of cerebrospinal fluid (CSF) within the brain, which leads to an increase in intracranial pressure (ICP).	Yes/No

Continued

Cardiac murmur	A cardiac murmur is a whooshing or swishing sound heard during a heartbeat, caused by turbulent blood flow in the heart	Yes/No
Cardiac anomaly	A cardiac anomaly refers to any structural defect or abnormality in the heart or its major blood vessels. It can affect the heart's function and cause problems with blood flow and oxygen supply to the body's organs and tissues.	Yes/No
Tetralogy of Fallot	Tetralogy of Fallot is a congenital heart defect that involves four abnormalities in the heart's structure, leading to inadequate oxygen supply to the body	Yes/No
Transposition of Great arteries	Transposition of the great arteries is a congenital heart defect in which the two main arteries leaving the heart are switched	Yes/No
Septal defect	Septal defect refers to a type of congenital heart defect that involves an abnormal opening in the septum (wall) between the heart's chambers.	Yes/No
Another cardiac anomaly	Other Cardiac Related Problems	Yes/No
Epidural	Epidural is a type of pain relief medication that is administered to women during labor and childbirth	Yes/No
Spinal	Spinal anesthesia is a type of regional anesthesia that is used to provide anesthesia or numbing of the lower body during surgical procedures or childbirth	Yes/No

2. Data Collection

The method of data collection for the study on gestational diabetes involved accessing data from https://physionet.com/, which provided a dataset consisting of information on 600 pregnant women. This dataset included details such as age, race, BMI, diabetes status (gestational diabetes, prediabetic, non-diabetic), and multiple visits for each woman throughout the pregnancy and post-delivery. To ensure the accuracy of the sample data, the following measures have been taken:

Data Source Validation: Physionet.com is a reputable source for physiological data, which increases the likelihood of the data being accurate and reliable. However, we have conducted preliminary assessments to ensure the dataset's integrity and validity.

Data Cleaning: Prior to analysis, we performed data cleaning processes to identify and rectify any errors, inconsistencies, or missing values in the dataset. This step helped improve data quality and accuracy.

Sample Representativeness: The sample of 600 pregnant women may have been selected to represent a diverse population to ensure the generalizability of the findings. This includes considering factors such as age, race, BMI, and diabetes status to reflect the broader population of pregnant women.

Statistical Analysis: Statistical techniques have been employed to detect outliers, assess distributions, and examine patterns within the data. These analyses

helped identify anomalies and ensure the accuracy and consistency of the sample data.

3. Literature Review

Medical informatics had its roots in the 1950s, initially emerging in the United States and later extending its influence on Europe and developing Eastern countries. The inaugural scholarly work outlining the concept of utilizing computer technologies in medicine was introduced by Robert S. Ledley and Lee Browning Lusted in 1959 [2]. As authors describe, we have seen the development of technology and health together and these both domains cannot be separated. In recent years, the introduction of data analytics to large amounts of healthcare data collected on daily basis opened numerous new opportunities and challenges in the field of medical informatics. By definition, healthcare informatics refers to the process of leveraging information technologies to improve the quality of healthcare [3]. The shift to digitizing healthcare data is a direct outcome of the evolution and revolution of big data. The substantial increase in data volume in recent years prompted the identification of a distinct domain known as big data. In the realm of information technology, the term "big data" commonly refers to vast sets of data that surpass the capacity and complexity manageable by traditional databases. Healthcare activities generate large amounts of data. Analytical procedures have been used to derive actionable judgments from data management technologies [4]. It promises us the power of early detection, prediction, prevention, and helps us to improve the quality of life [5]. Numerous machine learning models have been developed to predict pregnancies. In addition to these models, employing an analytical approach could prove to be an efficient and cost-effective method. Therefore, this paper investigates the use of analytics through power BI analytical tool on the patients with gestational diabetes.

4. Data Modeling

To facilitate effective data analysis, the dataset is divided into five distinct tables as shown in **Figure 1**, visit 1, visit 2, visit 3, visit 4, and difference between visit 1 and visit 3, allowing for focused examination of each pregnancy visit and comparisons between specific visits.

5. Methodology

Data cleaning plays a pivotal role in ensuring the integrity and usability of a dataset for analytical purposes. In this research, we present a systematic approach to data cleaning and preprocessing applied to a complex dataset comprising medical and fetal data. The paper also emphasizes the importance of data preprocessing in enhancing the overall quality of the dataset for subsequent analysis. The following sections describe the specific steps and techniques employed to address data quality issues and enhance the dataset's utility. The methodology and strategies implemented are outlined, with each step

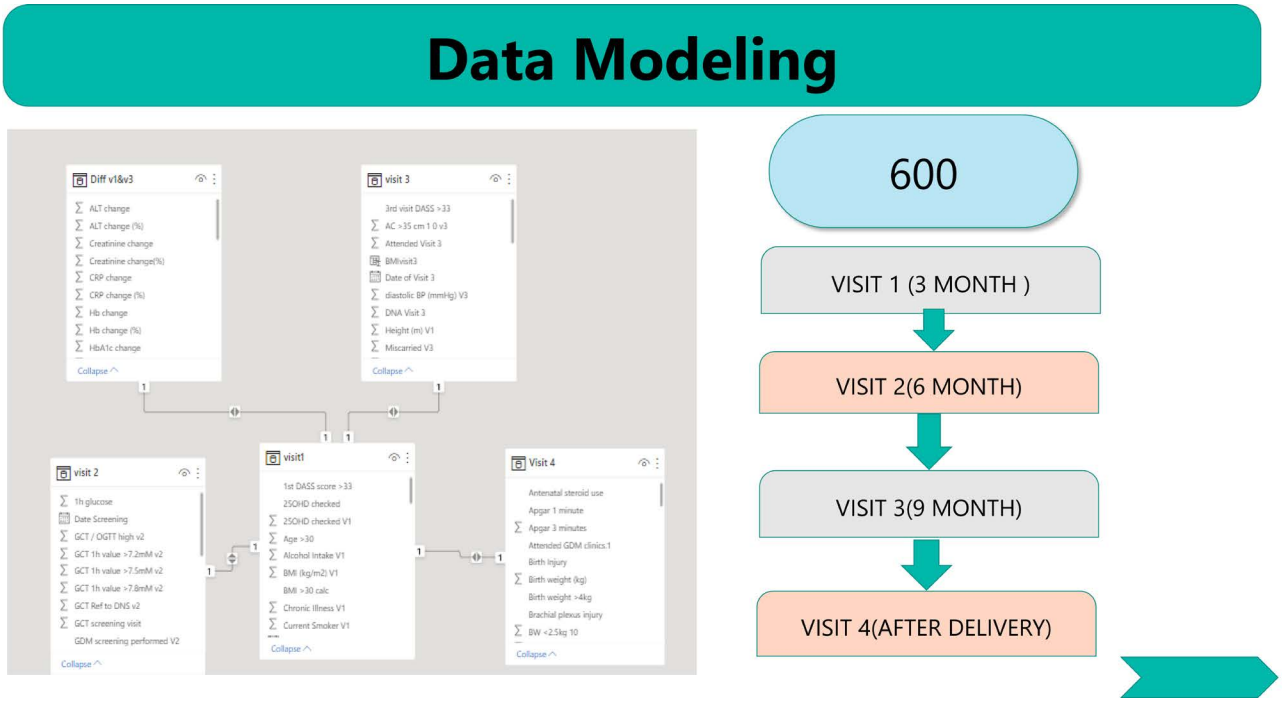

Figure 1. Data modeling.

explained and exemplified. Firstly, the dataset was partitioned into five sub-tables: Visit 1, Visit 2, Visit 3, Difference in Visit 1 and 3, and Fetal Details. The steps are:

1) Removal of Null and Redundant Columns Null and redundant columns were removed to simplify the dataset, enhance clarity, and reduce computational overhead.

2) Decimal Point Rounding Columns with excessive decimal points were rounded to maintain consistency and precision in numeric values.

3) Uniform Case Conversion Inconsistent case types (e.g., "Yes" and "yes") were standardized to ensure uniformity.

4) Data Type Transformation Column data types were adjusted based on the content of each column, ensuring accurate representation.

5) Column Name Modification Column names were modified for improved clarity and ease of interpretation.

6) Outlier Detection and Handling Outliers in column content were identified and addressed through various methods, such as correcting decimal places and recalculating erroneous values.

7) Unwanted Column Removal Columns containing repetitive or unnecessary information were removed to streamline the dataset.

8) Text Value Replacement Conditional columns were created to replace text values with appropriate ones, addressing inconsistencies in data entry.

9) Delimiter Standardization Columns with multiple delimiters were processed to achieve a uniform data pattern.

10) Symbol Replacement Symbols, such as "<.>", were replaced with values

slightly higher or lower to facilitate data analysis.

11) Column Splitting Columns were split using delimiters to consolidate data into two categories (e.g., "Yes" and "No").

12) Generalization of Values Similar values were generalized or replaced to ensure consistency and reduce complexity. For instance, ethnicities were categorized into five major groups.

6. Demographics

Demographics of the Dataset are shown in Figure 2. The dataset comprises 600 pregnant women, with 50 diagnosed with GDM, 7 with prediabetes, and the remainder as non-diabetic. Many of the patients were aged 30 and above, and most belonged to the white race. The dataset encompasses four critical visits throughout pregnancy and includes a significant percentage of women with a high BMI.

7. Identification and Risk Factor Analysis

Figure 3 shows the analysis revealed 50 GDM cases, with risk factors including obesity (33%), family history of type 2 diabetes, advanced maternal age (29 women

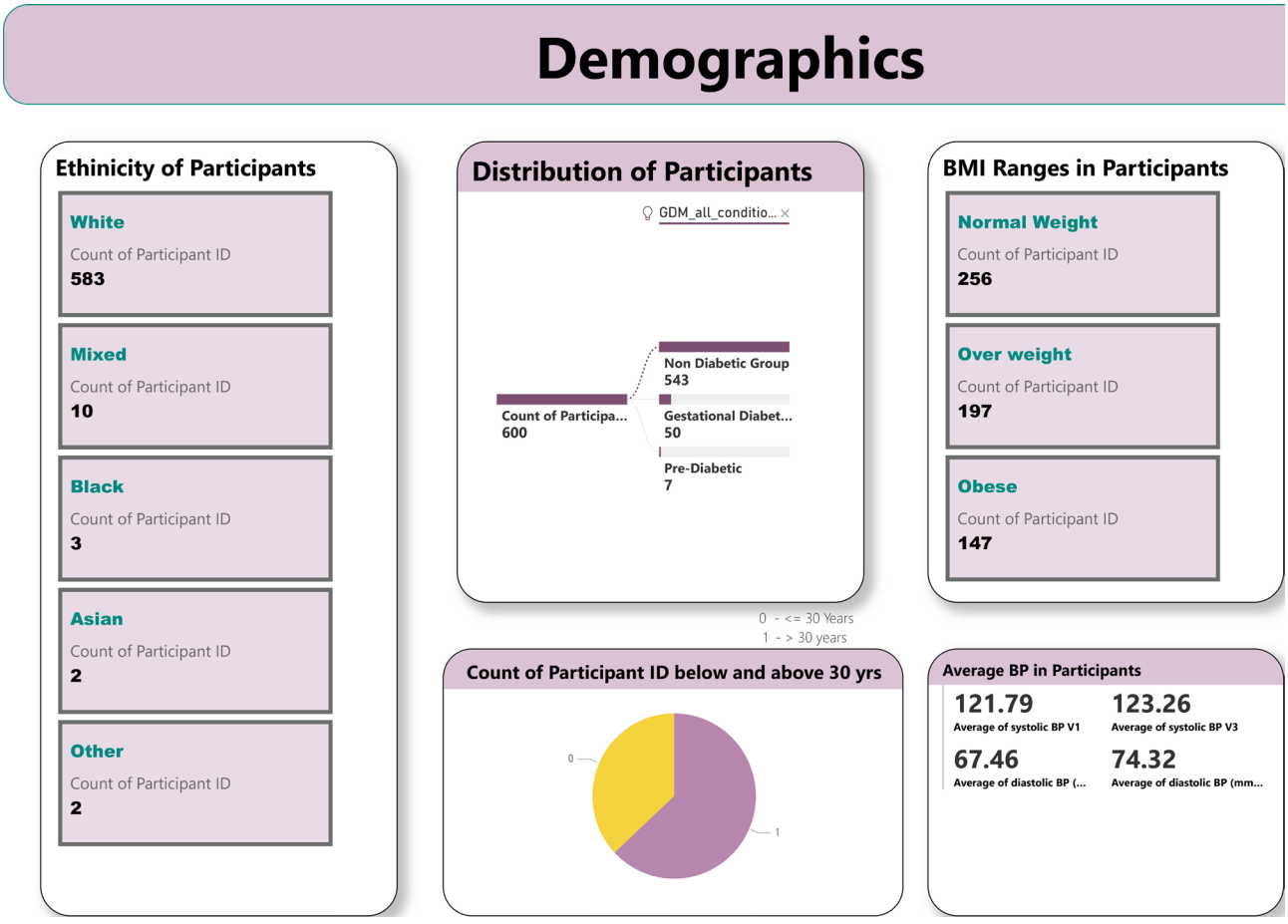

Figure 2. Demographics of dataset.

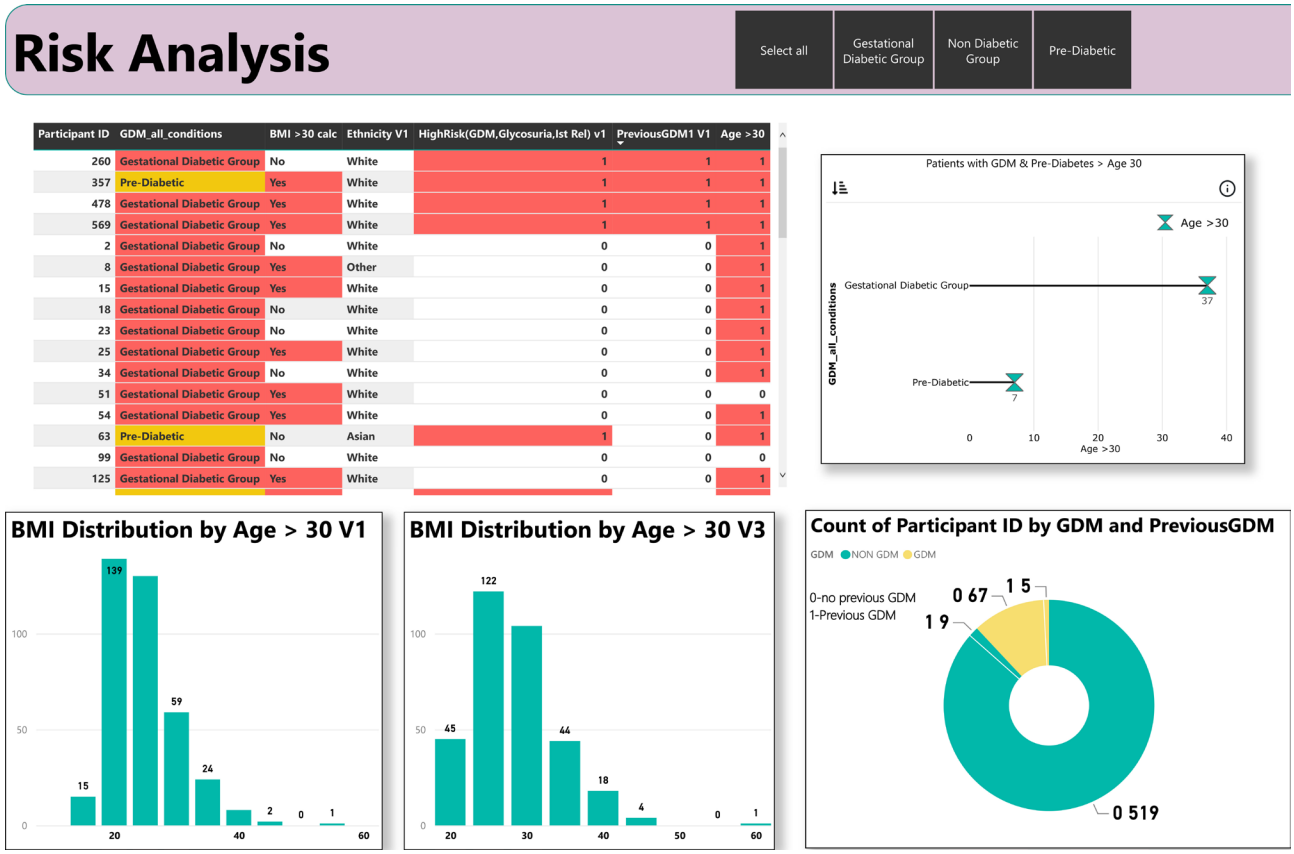

Figure 3. Risk analysis.

above 30), and limited instances of prior GDM. Ethnicity's impact was inconclusive due to skewed representation [6].

In addition, the study identified 50 GDM cases based on oral glucose tolerance tests (OGTT)/Glucose Challenge Tests (GCT) during the second trimester (V2), corroborated by the "Dx with GDM" column in the V4 table as shown in Figure 4. Risk factors for GDM, including obesity, family history, previous GDM history, age, and ethnicity, were analyzed, providing critical insights into potential predispositions [7].

Figure 5 shows analysis of various biomarkers, such as liver enzymes (ALT), kidney markers (albumin and creatinine), platelets, vitamin D, and others, were examined to understand their associations with GDM. Notable findings include elevated ALT levels in GDM patients, potential kidney-related complications, reduced platelet counts, and a correlation between vitamin D deficiency and GDM.

The analysis of maternal complications revealed several maternal complications associated with GDM as shown in Figure 6, including cesarean sections (C-sections), gestational hypertension, and the use of interventions during delivery. However, there were no instances of preeclampsia or eclampsia in the dataset, emphasizing the need for further exploration [8].

Fetal Complications Fetal complications observed in babies born to mothers

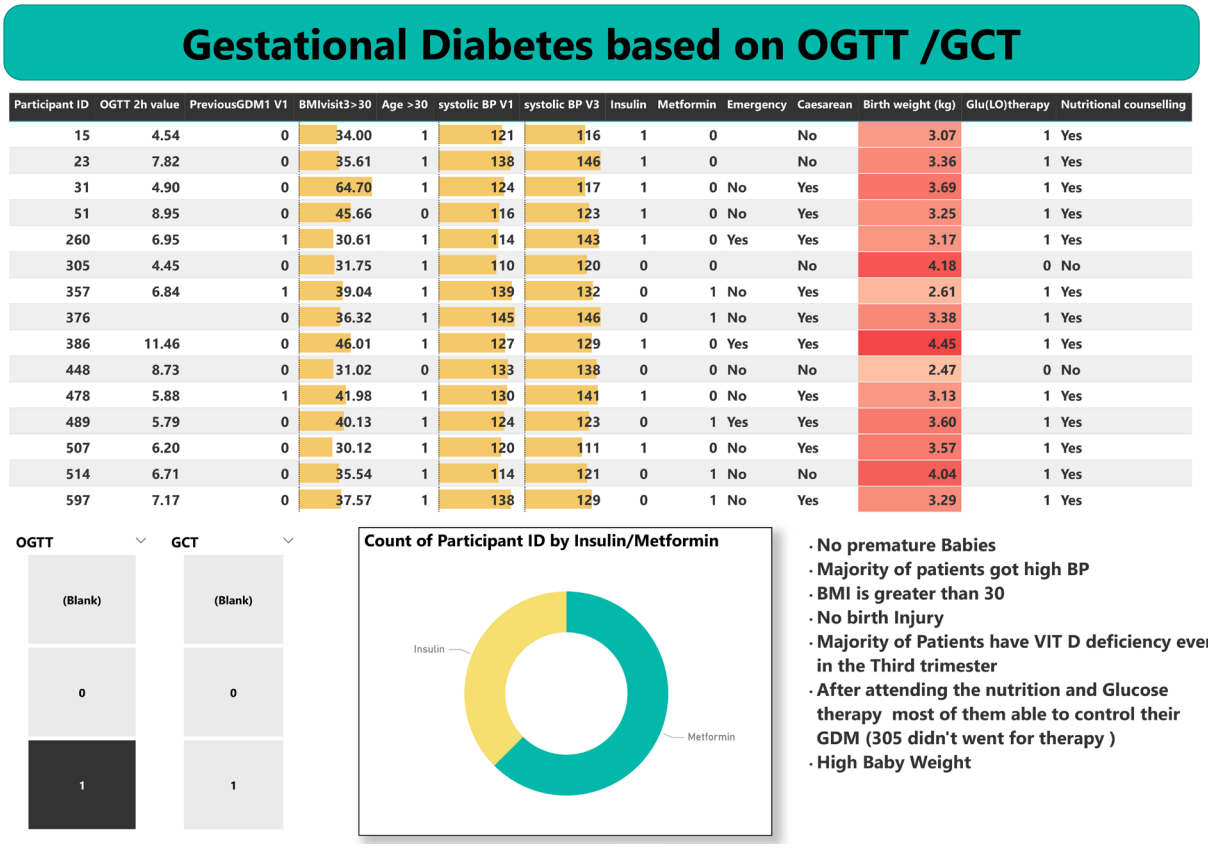

Figure 4. OGTT/GCT visualization visualize based on other biomarkers.

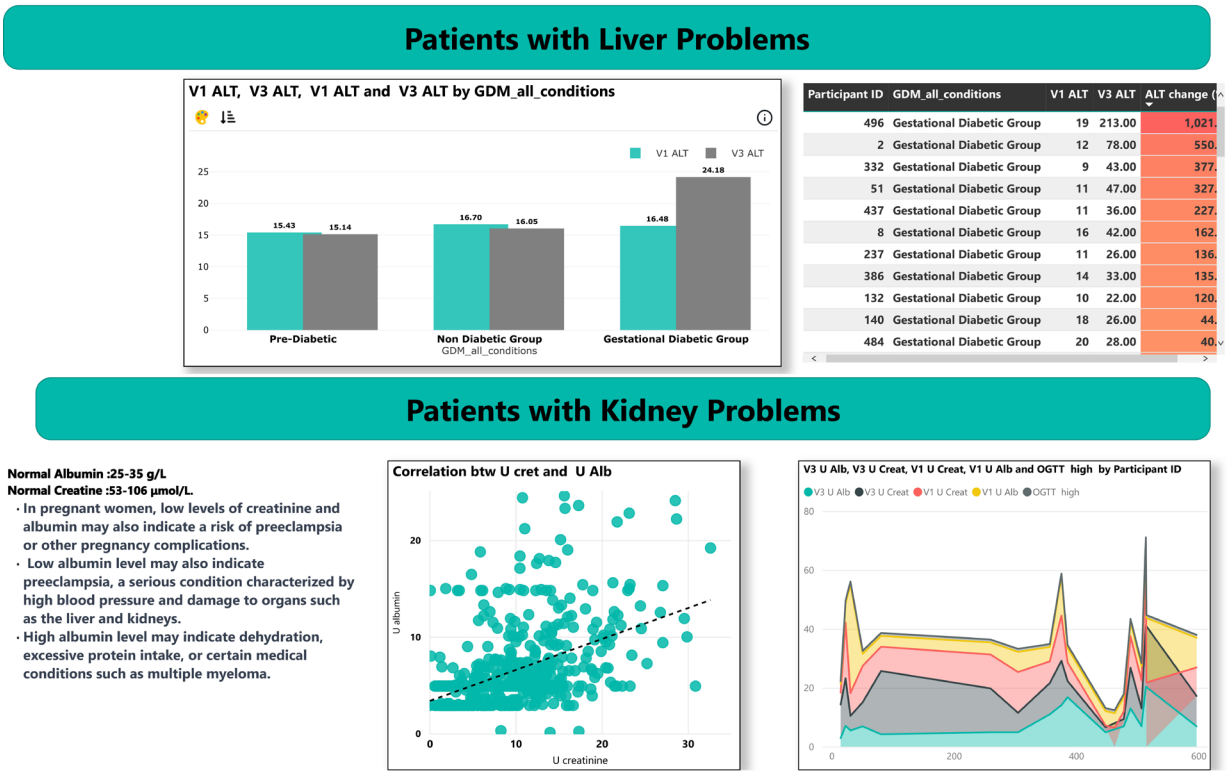

Figure 5. Visualization of liver enzymes.

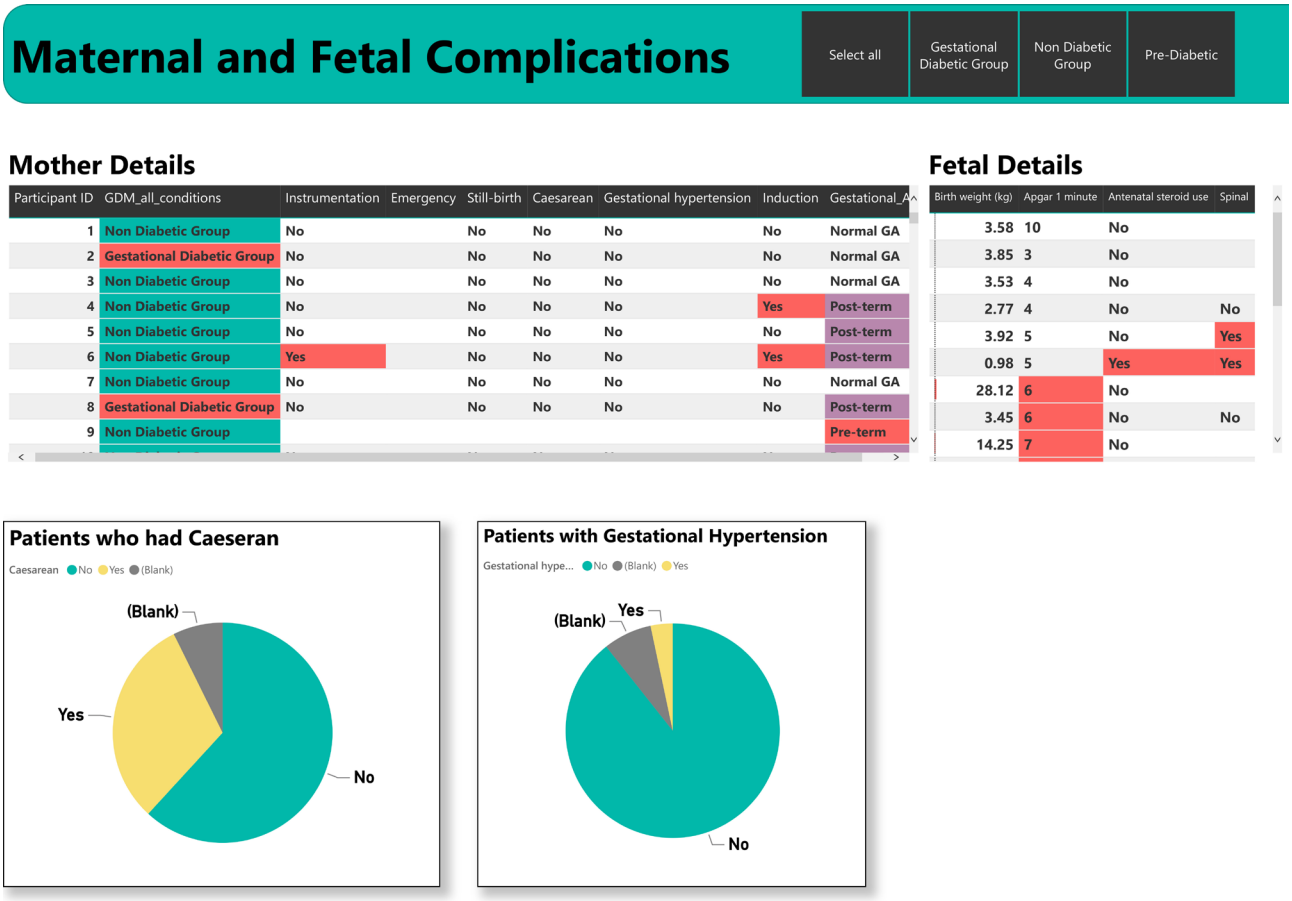

Figure 6. Visualization of maternal and fetal complication.

with GDM included macrosomia, low Apgar scores, antenatal steroid use, and abnormal spinal test results. These findings underscore the importance of managing GDM to mitigate fetal risks [9].

8. Conclusion

This comprehensive analysis offers valuable insights into the demographics, risk factors, biomarkers, and complications associated with GDM in a sizable healthcare dataset. Despite its contributions, this study acknowledges the limitations of the dataset and highlights the necessity for further research with larger sample sizes to gain a more nuanced understanding of GDM and its consequences. This research serves as a foundation for improved management and care of GDM during pregnancy, ultimately benefiting both mothers and their infants.

Conflicts of Interest

The authors declare no conflicts of interest regarding the publication of this paper.

References

[1] El-Rashidy, N., ElSayed, N.E., El-Ghamry, A. and Talaat, F.M. (2022) Prediction of Gestational Diabetes Based on Explainable Deep Learning and Fog Computing. *Soft*

Computing, **26**, 11435-11450. https://doi.org/10.1007/s00500-022-07420-1

[2] Masic, I. (2014) Five Periods in Development of Medical Informatics. *Acta Informatica Medica*, **22**, 44-48.

[3] Wang, F. and Stiglic, G. (2015) Data Analytics in Healthcare Informatics. 2015 *International Conference on Healthcare Informatics*, Dallas, TX, 21-23 October 2015, 444. https://doi.org/10.1109/ICHI.2015.62

[4] Awrahman, B.J., Aziz Fatah, C. and Hamaamin, M.Y. (2022) A Review of the Role and Challenges of Big Data in Healthcare Informatics and Analytics. *Computational Intelligence and Neuroscience*, **2022**, Article ID 5317760. https://doi.org/10.1155/2022/5317760

[5] Rehman, A., Naz, S. and Razzak, I. (2022) Leveraging Big Data Analytics in Healthcare Enhancement: Trends, Challenges and Opportunities. *Multimedia Systems*, **28**, 1339-1371. https://doi.org/10.1007/s00530-020-00736-8

[6] Razo-Azamar, M., *et al.* (2023) An Early Prediction Model for Gestational Diabetes Mellitus Based on Metabolomic Biomarkers. *Diabetology & Metabolic Syndrome*, **15**, Article Number: 116. https://doi.org/10.1186/s13098-023-01098-7

[7] Belsti, Y., *et al.* (2023) Comparison of Machine Learning and Conventional Logistic Regression-Based Prediction Models for Gestational Diabetes in an Ethnically Diverse Population; the Monash GDM Machine Learning Model. *International Journal of Medical Informatics*, **179**, 105228. https://doi.org/10.1016/j.ijmedinf.2023.105228

[8] Snyder, B.M., *et al.* (2020) Early Pregnancy Prediction of Gestational Diabetes Mellitus Risk Using Prenatal Screening Biomarkers in Nulliparous Women. *Diabetes Research and Clinical Practice*, **163**, 108139. https://doi.org/10.1016/j.diabres.2020.108139

[9] Cooray, S.D., Boyle, J.A., Soldatos, G., Wijeyaratne, L.A. and Teede, H.J. (2019) Prognostic Prediction Models for Pregnancy Complications in Women with Gestational Diabetes: A Protocol for Systematic Review, Critical Appraisal and Meta-Analysis. *Systematic Reviews*, **8**, Article Number: 270. https://doi.org/10.1186/s13643-019-1151-0