**DA Report**

*A report submitted in partial fulfilment of the requirements for the Award of Degree of*

**BACHELOR OF TECHNOLOGY**

**IN**

**ARTIFICIAL INTELLIGENCE & DATA SCIENCE ENGINEERING**

by,

**ONKAR JAMMA**

**ROLL NO 21**



**DEPARTMENT OF ARTIFICIAL INTELLIGENCE & DATA SCIENCE ENGINEERING**

**Pradnya Niketan Education Society, Pune.**

**NAGESH KARAJAGI *ORCHID* COLLEGE OF ENGGINEERING & TECHNOLOGY, SOLAPUR.**

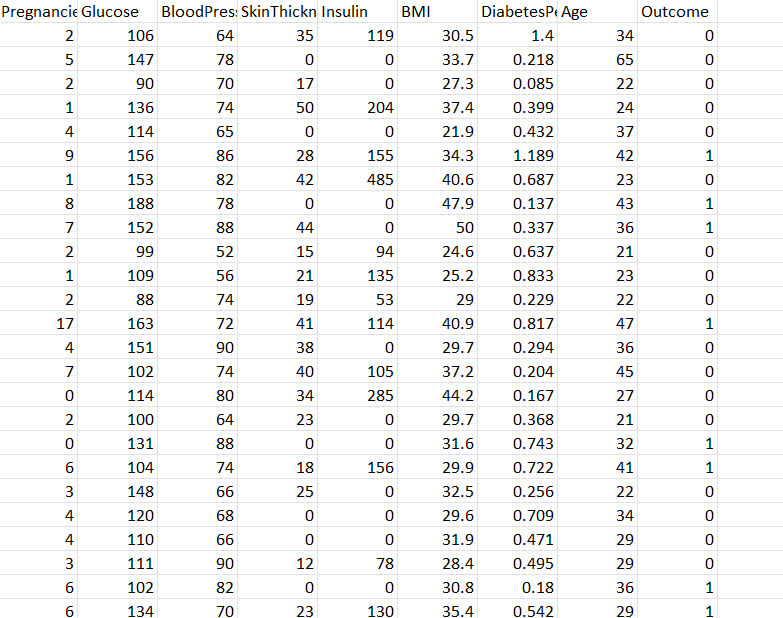
**2023-2024**

**Introduction**

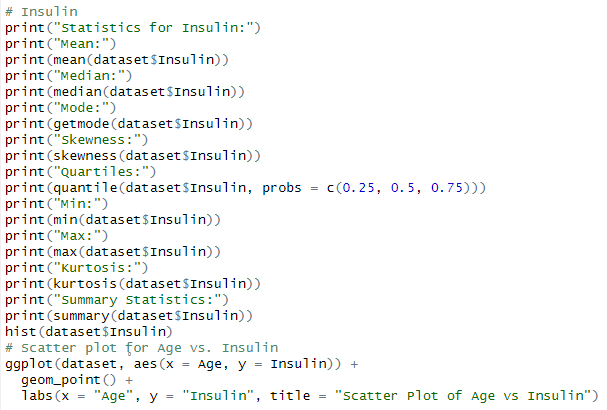
This report presents a comprehensive analysis of a dataset obtained from a study focusing on individuals diagnosed with diabetes. The dataset encompasses various demographic and health-related factors, including age, glucose levels, blood pressure, insulin levels, BMI (Body Mass Index), and diabetes pedigree function. Collected through meticulous record-keeping and medical assessments, this dataset aims to shed light on the health profiles and trends within this population.

The data provides valuable insights into the physiological and metabolic characteristics of individuals grappling with diabetes, a prevalent and complex health condition. Through graphical representations and statistical analyses, this report seeks to elucidate patterns, correlations, and trends within the dataset, offering insights that can inform healthcare practices, interventions, and public health strategies. Diabetes data analysis involves the systematic examination of various datasets related to diabetes, aiming to extract actionable insights for better understanding, prevention, and management of the disease. Through techniques such as data preprocessing, exploratory data analysis, and predictive modeling, analysts uncover patterns, risk factors, and outcomes associated with diabetes. This analysis facilitates the identification of high-risk populations, evaluation of interventions, and development of personalized treatment strategies. By leveraging machine learning algorithms and statistical methods, diabetes data analysis enables healthcare providers to predict disease progression, assess the effectiveness of interventions, and optimize patient outcomes. Moreover, it plays a vital role in population health management by identifying geographic disparities, social determinants of health, and trends in diabetes prevalence. Ensuring data privacy and security is paramount in handling sensitive health information, complying with regulations such as HIPAA. Overall, diabetes data analysis empowers stakeholders, including healthcare professionals, researchers, policymakers, and individuals affected by diabetes, to make informed decisions and improve the quality of care for those living with this chronic condition.

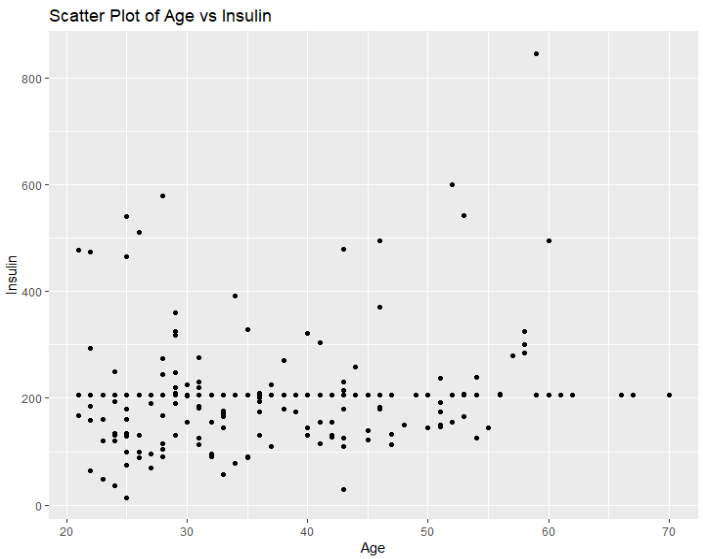
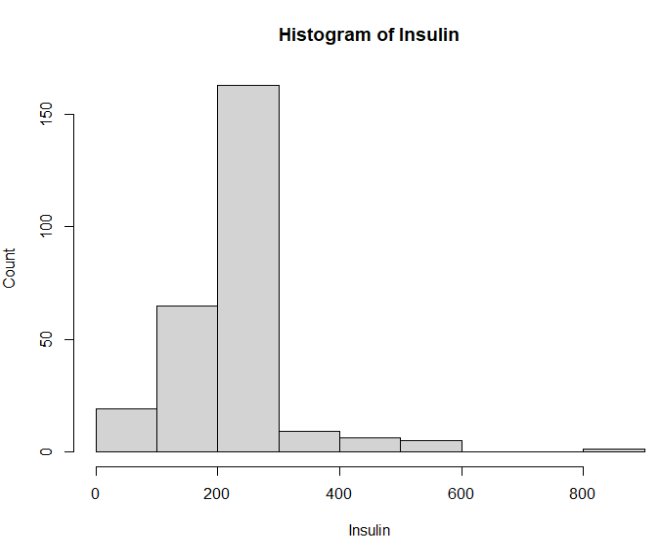
**View (DiabetesDataset)**

****

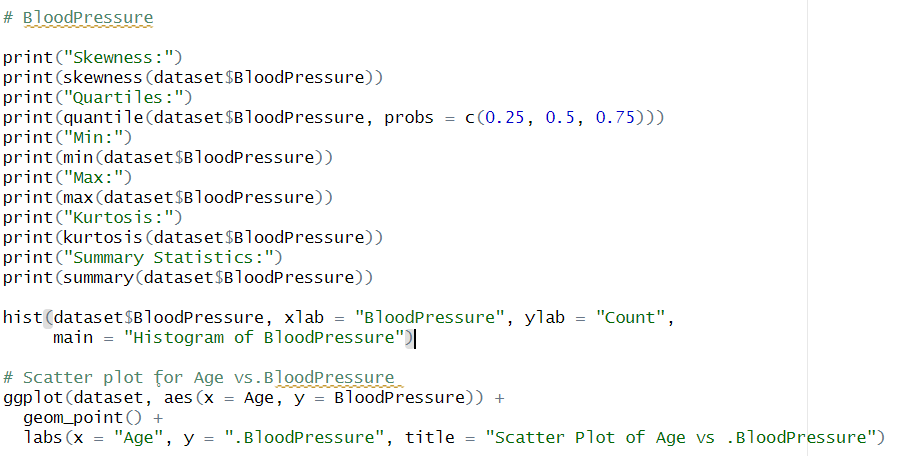
**Insulin Level Analysis**



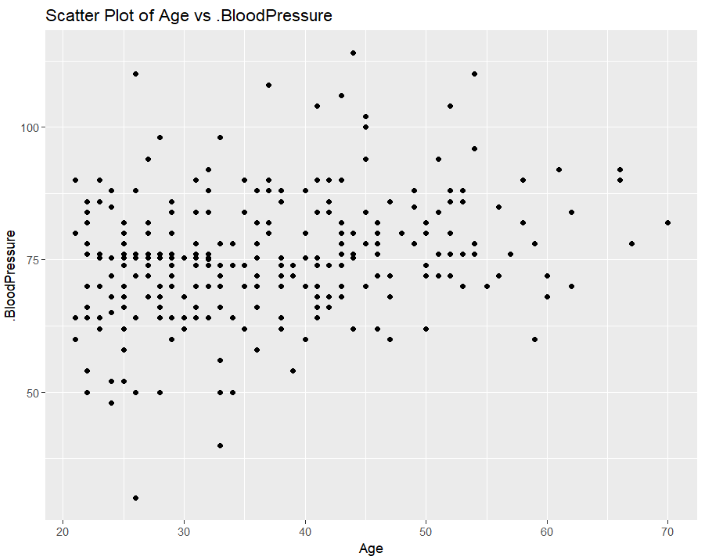
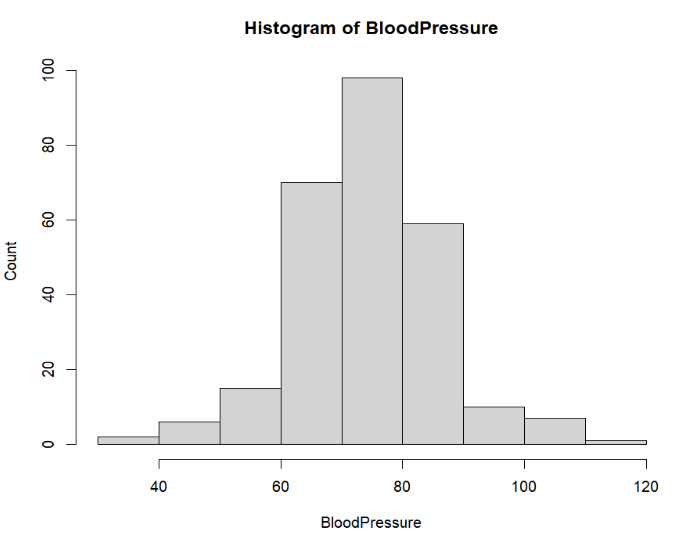
The analysis of insulin levels in individuals with diabetes reveals several key findings. The **mean insulin level** is **206.8462 units**, with a median and mode at the same value. The distribution is positively skewed, indicating a concentration of values around the mean. **Quartile analysis** shows clustering around the median. The dataset spans a wide range from 14 to 846 units, with a high kurtosis value suggesting potential outliers. A **histogram** visually **illustrates the distribution**. Additionally, a **scatter plot** examining the **relationship between age and insulin levels** shows the **dispersion of data points** and any **potential patterns or trends**. This provides **insights into how insulin levels vary with age**, which can be **valuable for understanding the progression of diabetes and informing treatment strategies.**

**Blood Pressure level Analysis**

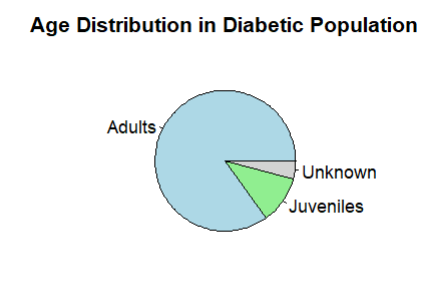
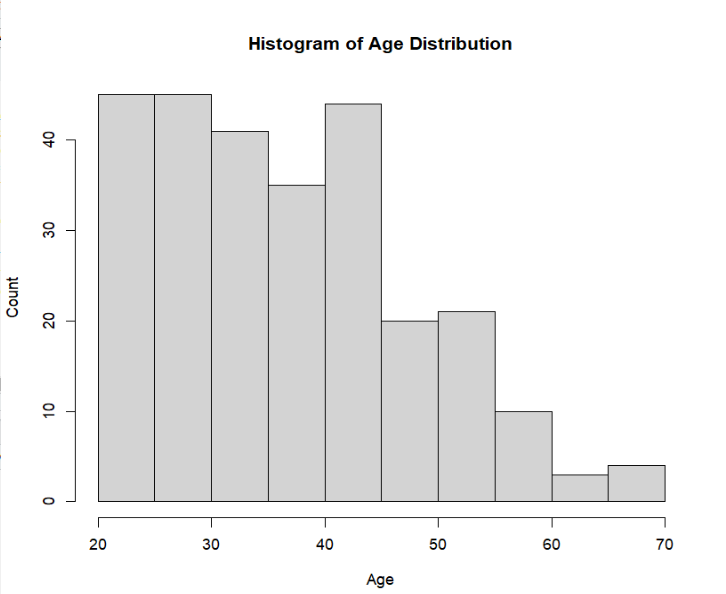


In our analysis of blood pressure levels among individuals with diabetes, we found several key insights. The **mean blood pressure level** was approximately **75.32 mmHg**, with the median at the same value, indicating a **central tendency** around this measure. **Quartile analysis** revealed that **25%** of individuals had a blood pressure of **68.00 mmHg or lower (Q1)**, while **75%** had a blood pressure of **82.00 mmHg or lower (Q3)**. The mode of the distribution was **70 mmHg**, suggesting a **concentration of observations** around this value. Blood pressure levels ranged from a minimum of 30 mmHg to a maximum of 114 mmHg, with a slight **right skew** indicated by a skewness value of approximately 0.08. The **kurtosis** value of approximately **1.21** suggested a distribution with relatively heavy tails, possibly indicating **platykurtic characteristics**.

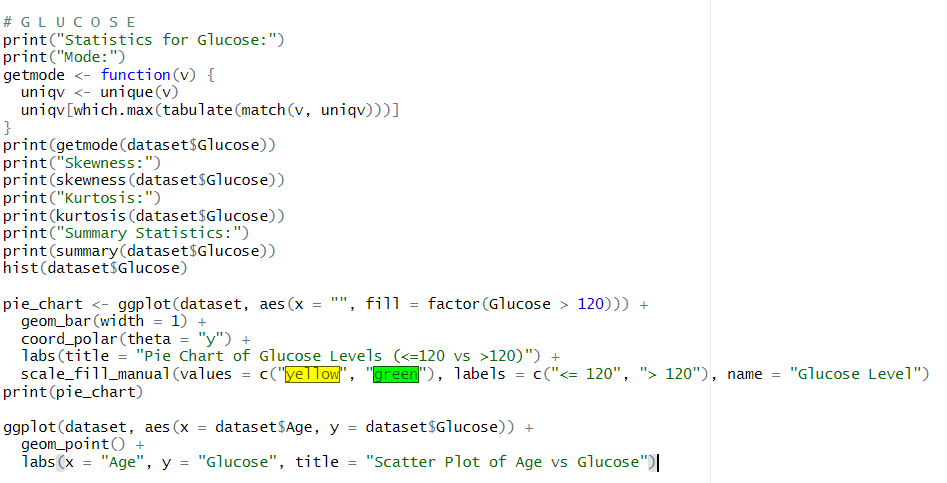
 

**Age Distribution Analysis**



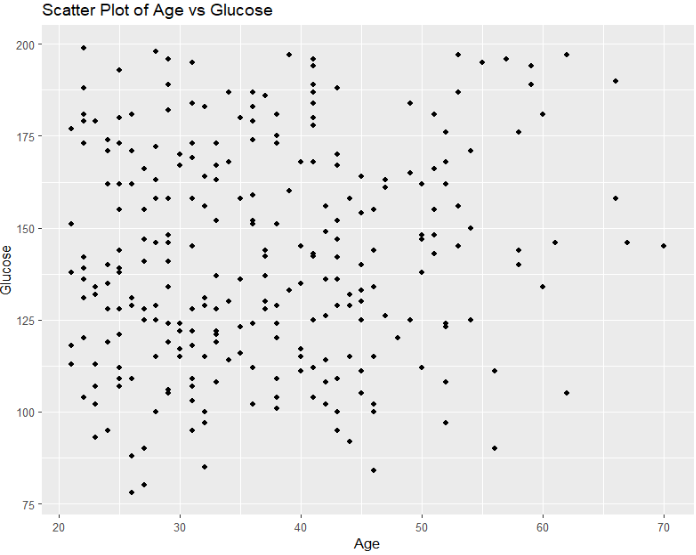
The age distribution analysis of individuals diagnosed with diabetes uncovers several key insights. The majority of the population, comprising **84.95%**, consists of **adults**, indicating a mature and stable diabetic community crucial for long-term health management. Additionally, the presence of **juveniles**, accounting for **10.92%**, signifies ongoing reproductive activity within the community, underscoring the importance of monitoring their survival into adulthood for sustainable population dynamics. Conversely, a smaller subset, comprising 4.13% of observations, falls into the 'unknown' category, highlighting potential areas for refinement in data collection methods. Statistically, the analysis reveals a **mean age** of **37.07 years**, with a median of 36 years and a mode of 25 years. The distribution displays a slight **right-skewed** pattern (skewness = 0.575), with quartiles at 28, 36, and 44 years. The age range spans from 21 to 70 years, with a relatively flat distribution indicated by a kurtosis of -0.383. This comprehensive examination offers valuable insights into the age demographics of the diabetic population, guiding healthcare planning and management strategies..  

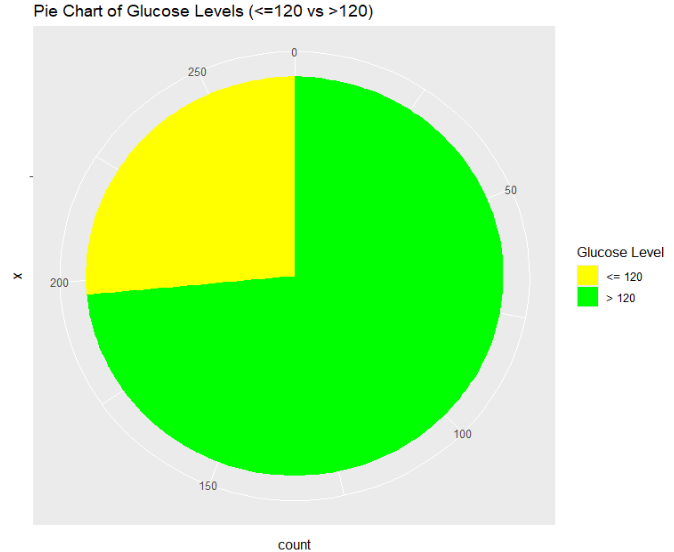
**Glucose Level Analysis**

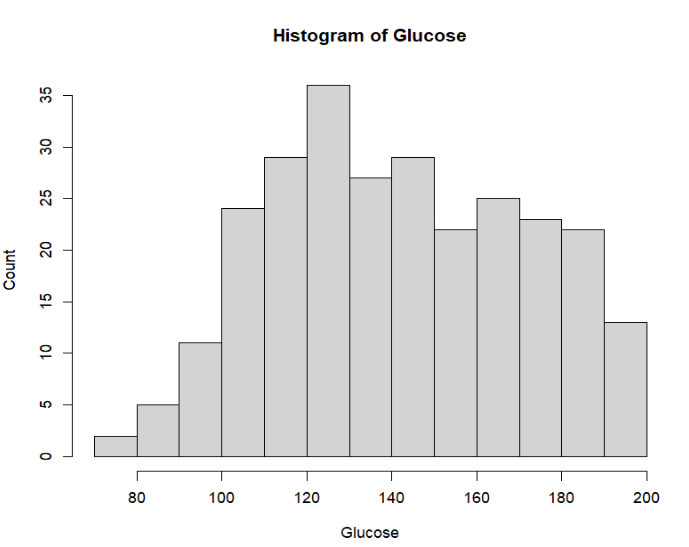


* The analysis of glucose levels in individuals with diabetes reveals a mean level of 142.32 mg/dL and a median of 140.5 mg/dL.
* **Quartile analysis** indicates glucose levels ranging from 78 to 199 mg/dL.
* The **distribution of glucose levels** shows a slight positive skewness and negative kurtosis, suggesting a relatively flat distribution with fewer extreme values.
* **Visualizations** such as histograms and scatter plots provide insights into the distribution and relationship of glucose levels with other variables.
* A **pie chart** complements the analysis by depicting glucose counts from 0 to 269, where counts from 0 to 195 are coloured in green, representing glucose levels lower than 120 mg/dL, while counts beyond 195 are coloured in yellow, representing glucose levels equal to or above 120 mg/dL.

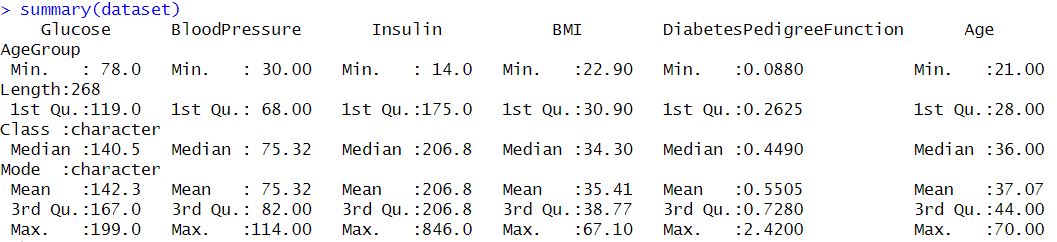
In summary, the analysis provides a comprehensive understanding of glucose regulation in individuals with diabetes, offering insights into their metabolic health and informing clinical management strategies.







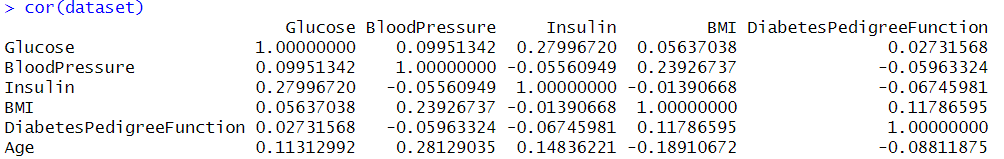
**Diabetes Dataset Summary**



**Correlation Analysis**

"Correlation analysis" is a statistical technique used to measure the strength and direction of the relationship between two or more variables. It helps identify whether and how strongly variables are related to each other. In the context of data analysis, correlation analysis is valuable for understanding the associations between different variables in a dataset.

Correlation analysis in the diabetes dataset reveals the degree of association between variables such as glucose levels, blood pressure, insulin levels, BMI, age, and diabetes pedigree function. Strong correlation coefficients near +1 or -1 indicate a robust linear relationship, while coefficients close to 0 suggest weak or negligible correlation between variables.



**Conclusion**

The analysis of the diabetes dataset provides a multifaceted understanding of the health profile and interrelationships among key variables in individuals diagnosed with diabetes. Predominantly comprising adults, the dataset exhibits a diverse range of clinical parameters, including glucose levels, blood pressure, insulin levels, BMI, age, and diabetes pedigree function. Glucose levels, a critical indicator of diabetes management, display a mean of 142.32 mg/dL, with a range spanning from 78 to 199 mg/dL, reflecting the variability in glycemic control among the population. Blood pressure levels are clustered around a mean of 75.32 mmHg, indicating overall stability within the cohort. Insulin levels exhibit significant variability, underscoring the heterogeneous treatment needs among individuals with diabetes.

Correlation analysis reveals associations between these variables, offering insights into their interrelationships and potential implications for diabetes management. Strong correlations between certain variables, such as glucose levels and insulin levels, highlight the intricate physiological mechanisms underlying diabetes pathophysiology. Additionally, demographic factors such as age and BMI demonstrate varying degrees of association with clinical parameters, providing valuable insights into the multifactorial nature of diabetes. This holistic understanding enables healthcare practitioners to tailor interventions effectively, addressing individualized needs and optimizing diabetes care strategies to improve health outcomes and quality of life for individuals living with diabetes.