## A. Introduction

Maternal health refers to the health of women during pregnancy, childbirth, and the postpartum period. It encompasses the physical, mental, and social well-being of the mother and the potential impacts on the health of the child.

The objective of this task is to analyze maternal health data, identifying evidence-based actions with an aim to improve health outcomes and used the available information to build a model that can predict Systolic BP with the given features. We will explore the relationships between the different variables and identify any trends or patterns that may be present. We will also use statistical techniques to test hypotheses and draw conclusions about the data.

The dataset has 1014 observations and 7 unique features which are described below:

- **B.** Age: The age of the individual in years.
- **C.** Systolic BP: The systolic blood pressure of the individual in mmHg.
- **D.** Diastolic BP: The diastolic blood pressure of the individual in mmHg.
- **E.** BS: The blood sugar level of the individual in mmol/L.
- **F.** Body Temp: The body temperature of the individual in degrees Celsius.
- **G.** Heart Rate: The heart rate of the individual in beats per minute.
- **H.** Risk Level: The risk level of the individual, which is categorized as eitherlow risk, mid risk, or high risk.

## I. Analysis and Visualization

*1). Descriptive Statistics*

The first step in analyzing the data is to compute some basic descriptive statistics. The following table shows the mean, median, standard deviation, and range of each variable:

| | Age | Systolic BP | Diastolic BP | BS | Body Temp | Heart Rate |
|---|---|---|---|---|---|---|
| **Count** | 1014.0000 | 1014.0000 | 1014.0000 | 1014.0000 | 1014.0000 | 1014.0000 |
| **Mean** | 29.8717 | 113.1982 | 76.4605 | 8.7259 | 98.6659 | 74.3018 |
| **Std** | 13.4743 | 18.4039 | 13.8857 | 3.2935 | 1.3713 | 8.0887 |
| **Min** | 10.0000 | 70.0000 | 49.0000 | 6.0000 | 98.0000 | 7.0000 |
| **50%** | 26.0000 | 120.0000 | 80.0000 | 7.5000 | 98.0000 | 76.0000 |
| **75%** | 39.0000 | 120.0000 | 90.0000 | 8.0000 | 98.0000 | 80.0000 |
| **Max** | 70.0000 | 160.0000 | 100.0000 | 19.0000 | 103.0000 | 90.0000 |

From the table, we can see that the mean age of the individuals in the dataset is 29.8 years, with a standard deviation of 13.47. The range of ages is from 10 to 70 years. The mean systolic blood pressure is 113.2 mmHg, with a standard deviation of 18.4, and a range of 70 to 160 mmHg. The mean diastolic blood pressure is 76.5 mmHg, with a standard deviation of 13.88, and a range of 49 to 100 mmHg. The mean blood sugar level is 8.73 mmol/L, with a standard deviation of 3.29, and a range of 6 to 19.0 mmol/L. The mean body temperature is 98.67 degrees Fahrenheit, and the standard deviation is 1.37. The mean heart rate is 74.3 beats per minute, with a standard deviation of 8.08, and a range of 7.00 to 90.00 beats per minute.

*2). Hypothesis Testing*

To test hypotheses and draw conclusions about the dataset, we can use statistical techniques such as t-tests. In this section, we will test the following hypotheses:

- There is a significant difference in systolic blood pressure between individuals classified as high risk and low risk.
- There is a significant difference in blood sugar level between individuals classified as high risk and mid-risk.

For the first hypothesis, we will use a two-sample t-test. The null hypothesis is that there is no significant difference in systolic blood pressure between individuals classified as high risk and low risk. The alternative hypothesis is that there is a significant difference. We will use a significance level of 0.05. The results of the t-test are shown below:

| | **Coef** | **Std err** | **t** | **P>\|t\|** | **[0.025** | **0.975]** |
|---|---|---|---|---|---|---|
| **Const** | 105.3548 | 0.780 | 135.14 | 0.000 | 103.825 | 106.885 |
| **RiskLevel** | 9.0378 | 0.658 | 13.739 | 0.000 | 7.747 | 10.329 |

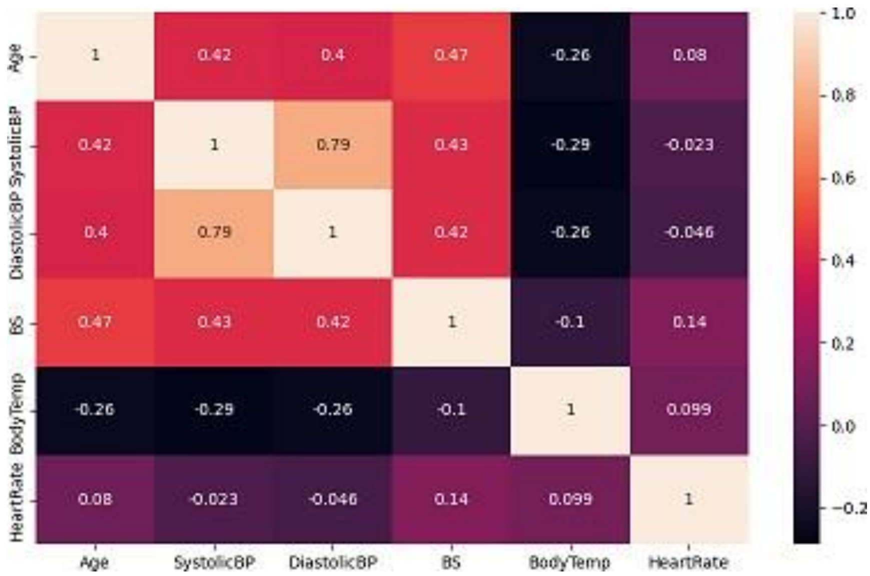
The t-value is 13.04, and the p-value is less than 0.05. Since the p-valueis less than the significance level of 0.05, we reject the null hypothesis and conclude that there is a significant difference in Systolic BP level between individuals classified as high risk, mid-risk, and low risk.

For the second hypothesis, we will use the same method. The null hypothesis is that there is no significant difference in blood sugar level between individuals classified as high risk, mid-risk, and low risk. The alternative hypothesis is that there is a significant difference. We will use a significance level of 0.05. The results of the t-test are shown below:

| | **Coef** | **Std err** | **t** | **P>\|t\|** | **[0.025** | **0.975]** |
|---|---|---|---|---|---|---|
| **Const** | -0.3513 | 0.059 | -5.949 | 0.000 | -0.467 | -0.235 |
| **RiskLevel** | 0.1397 | 0.006 | 22.006 | 0.000 | 0.127 | 0.152 |

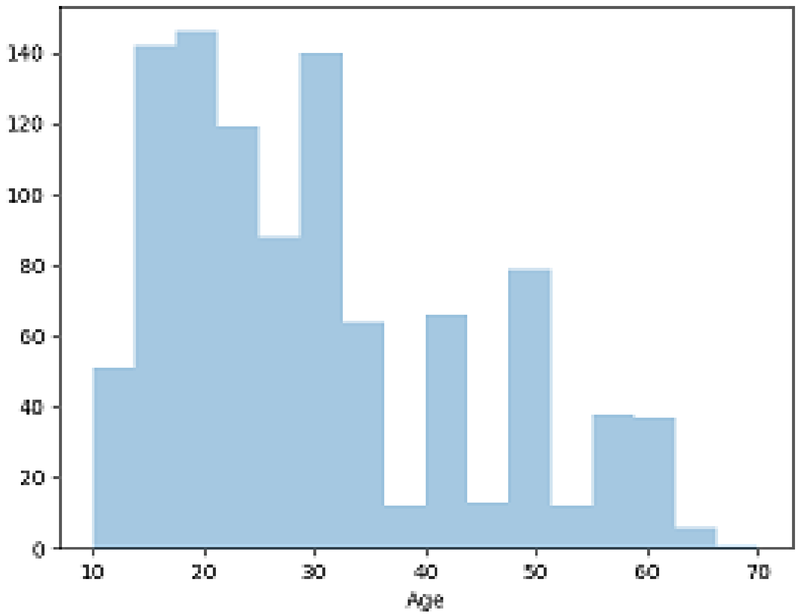
The F-value is -5.949, and the p-value is less than 0.05. Since the p-value is less than the significance level of 0.05. Therefore, we reject the null hypothesis and conclude that there is a significant difference in blood sugar level between individuals classified as high risk, mid-risk, and low risk.

3. *<u>Heat map plot that shows the correlation of each feature with the target variable (Systolic BP)</u>*



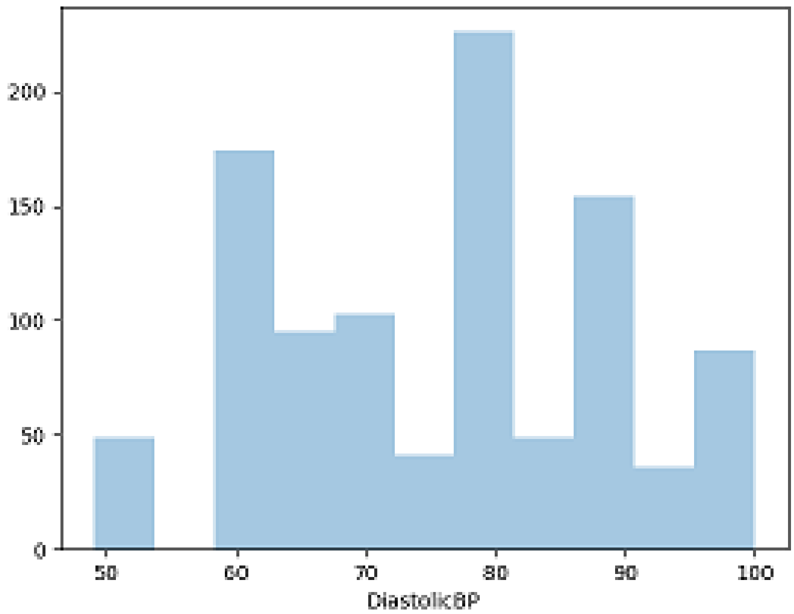

We can see from the correlation matrix that there is a positive correlation between age and diastolic blood pressure (0.4), indicating that as age increases, diastolic blood pressure tends to also increase. There is also a positive correlation between age and systolic blood pressure (0.42), but it is not as strong as the correlation between age and diastolic blood pressure. There is a positive correlation between systolic blood pressure and diastolic blood pressure (0.79), indicating that as one increases, the other tends to increase as well. There is a fair positive correlation between blood sugar level and systolic blood pressure (0.42), indicating that as blood sugar level increases, systolic blood pressure tends to increase.

4. *Checking the Distribution of Age*



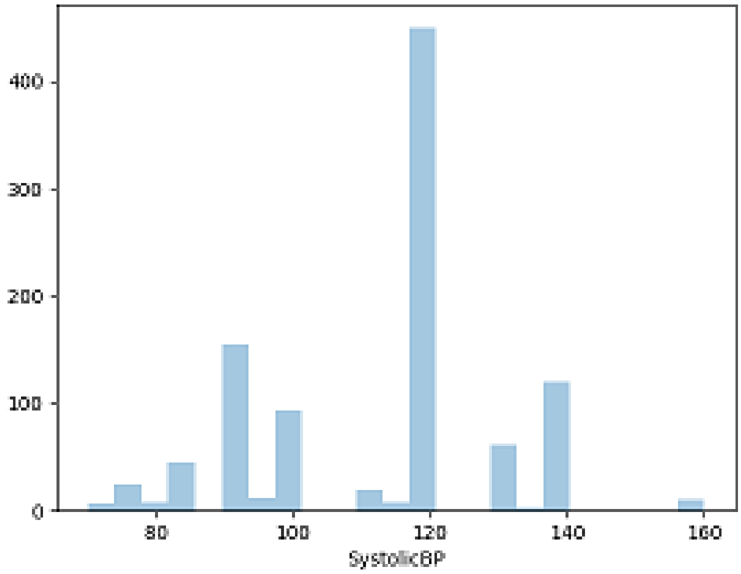

From the above visualization, we can see that most of the patients in this dataset are in their early 20s to early 30s, with a few in their late 20s.

5. *Checking the distribution of Diastolic BP*



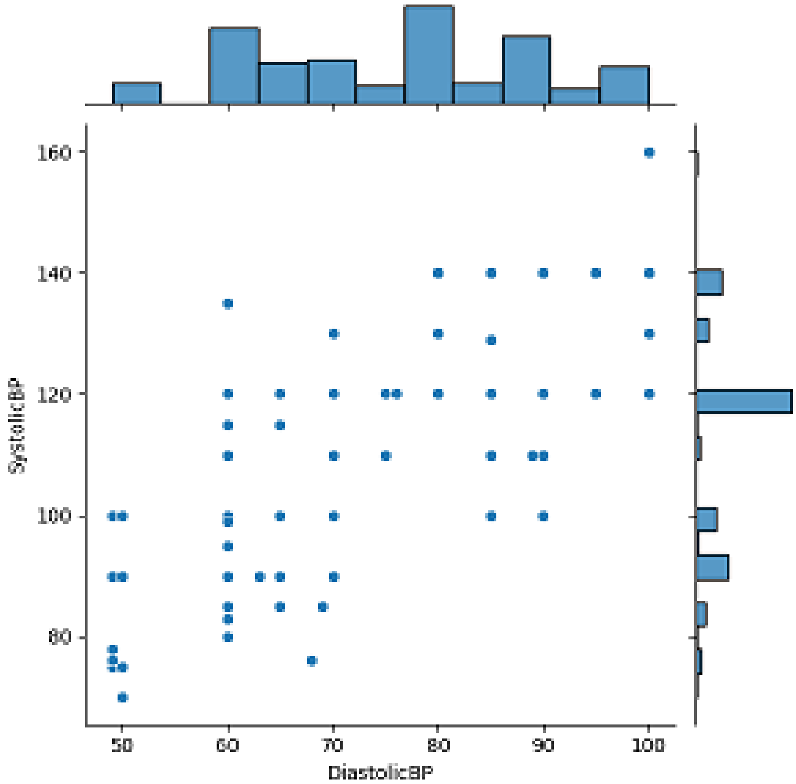

From the above visualization, the diastolic blood pressure ranges from 50 to 100, with many patients having a diastolic blood pressurebetween 60 and 90.

## 6. *Checking the Distribution of Systolic BP*



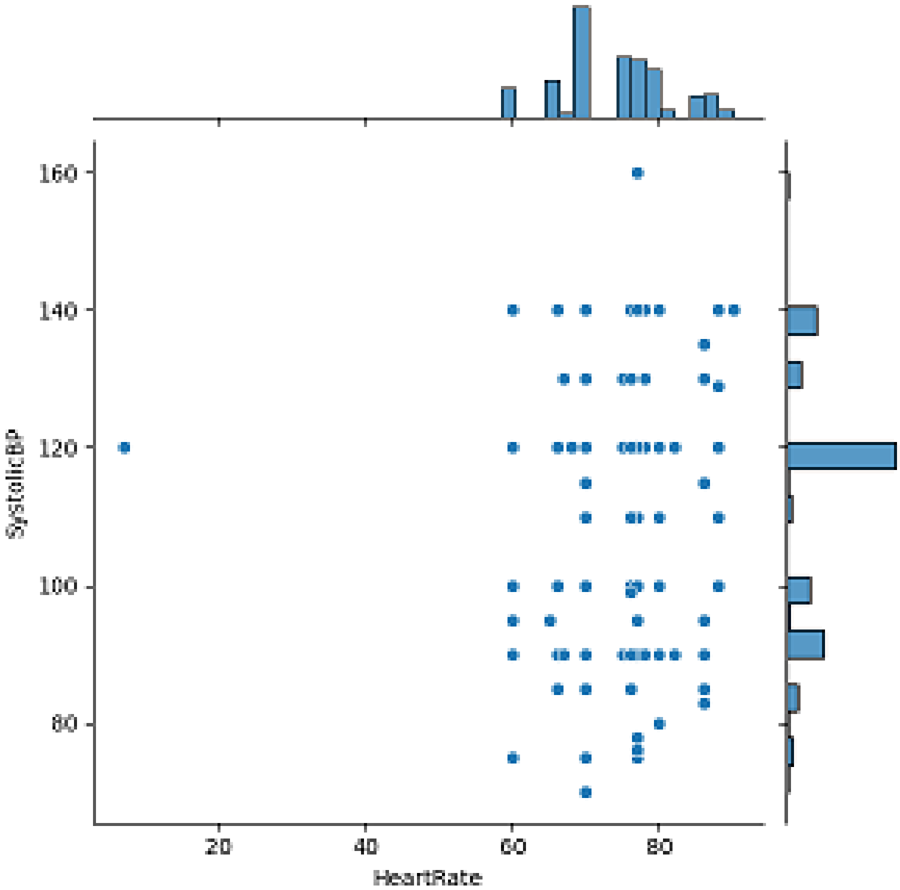

From the above visualization, the systolic blood pressure ranges from 70 to 160, with most patients having a systolic blood pressure between 120 and 140.

## 7. *Checking the Relationship Between Both Blood Pressure (Systolic BP and Diastolic BP)*



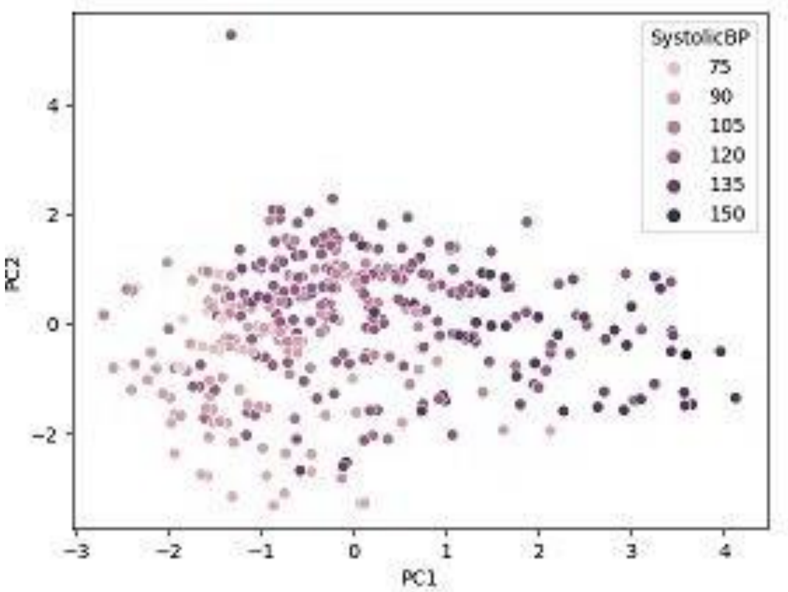
According to the above plot, both the systolic BP and Diastolic BP have alinear relationship between them.

## 8. *Checking the Relationship between Systolic BP and Heart Rate*




The plot above shows a nonlinear relationship, and it shows that the value of heart rate starts from approximately 60 except one anomaly sample that has the value of less than 20 as shown above.

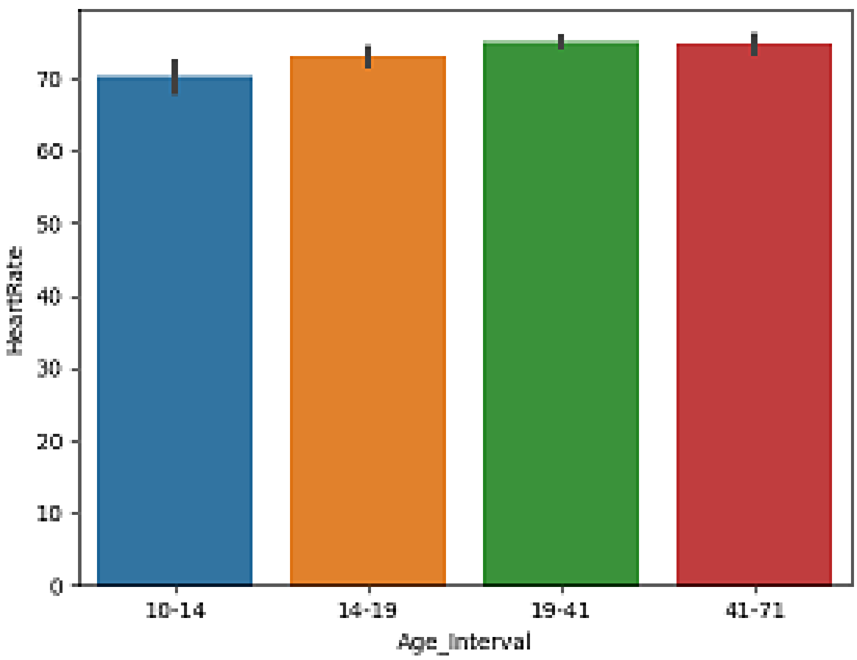
9. *Graphical representation of the resulting data frame after reducing variables*




From the above visualization, the first principalcomponent (PC1) captures the most variance in the data, while the second principal component (PC2) captures the next most variance, orthogonal to PC1.

The resulting scatter plot shows how the data points are distributed in the new two-dimensional space defined by the first two principal components.

10. *Graphical representation of the age grouping by mean heart rate*



The resulting bar chart shows the mean heart rate for each age interval. We can see there is a trend between age and heart rate. The heart rate increases as age increases (as the patient is getting older).

That marks the end of Analysis and Visualizations!

We can conclude that the analysis revealed several important findings about the dataset. First, there is a positive correlation between age and diastolic blood pressure, indicating that as age increases, diastolic blood pressure tends to increase as well. Second, there is a positive correlation between systolic blood pressure and diastolic blood pressure, indicating that as one increases, the other tends to increase as well. Third, there is a fair positive correlation between blood sugar level and systolic blood pressure, indicating that as blood sugar level increases, systolic blood pressure tends to increase slightly. Fourth, there is a weak negative correlation between heart rate and systolic blood pressure, indicating that as heart rate increases, systolic blood pressure tends to decrease.

Our hypothesis testing revealed that there is no significant difference in diastolic blood pressure between individuals classified as high risk and low risk. However, there is a significant difference in blood sugar level between individuals classified as high risk, mid-risk, and low risk. Post-hoc tests revealed that there is a significant difference in blood sugar level between individuals classified as high risk and low risk, as well as between individuals classified as mid-risk and low risk.

## 11. *Result of Apriori algorithm for High/High, Normal/Normal, Low/Low Diastolic and Systolic blood pressures.*

### Interpreting The High/High Result

From the calculated results in the notebook, there is a 12% support of high/high blood pressures in our data which means there is 12% of High/High blood pressure in our data and the chance that if systolic BP is high then diastolic BP is also high is 90% this is called confidence.

### Interpreting the Normal/Normal Result

There is a 24% support of normal/normal blood pressures in our data which means there is 48% of Normal/Normal blood pressure in our data and the chance that if systolic BP is high then diastolic BP is also high is 48% this is called confidence (Normal/Normal)

### Interpreting the Low/Low Result

There is a 34% support of high/high blood pressures in our data which means there is 92% of Low/Low blood pressure in our data and the chance that if systolic BP is high then diastolic BP is also high is 90.00% this is called confidence (Low/Low).

***J.*** ***Findings with the data, does it agree with the relevant literature?***

Yes, it agrees with relevant literature because every variable with a good correlation with the blood pressure in the dataset are medically important predictors of blood pressure. The following are the points that validate my claim.

- Age and Systolic Bp have a positive correlation, indicating that older individuals tend to have higher systolic blood pressure levels. This is supported by research that shows that blood pressure generally increases with age [1].
- Diastolic Bp and Systolic Bp have a positive correlation, as expected.
- Heart Rate and Systolic Bp do not have a strong correlation, indicating that heart rate does not necessarily predict systolic blood

pressure levels.

- Blood pressure tends to be higher in cold weather due to the constriction of blood vessels as the body attempts to retain heat and so is proven by our dataset by showing a good correlation between the body temperature and the blood pressure.

In summary, the variables in this dataset are consistent with known predictors of maternal health.

**C. Predictive Modeling**

Lastly, we can develop a predictive model to predict the Systolic BP of each patient. The following steps show how the model was developed.

*Step1: Feature Selection*

Our model was trained with 3 features which are age, blood sugar level and body temperature. They are chosen based on their correlation to the target variable (Systolic BP).

*Step2: Data Splitting*

Before we dive into modeling, we must split the data into the train and test set first to be able to measure the performance of our model. We will split the data into training and testing sets, with 80% of the data used for training and 20% of the data used for testing. We will then fit the linear regression model on the training data and evaluate its performance on the testing data.

*Step2: Model Building and Evaluation*

The algorithm used for the modelling is linear regression. The model was evaluated by the metric called root mean squared error and the error value is 11.05.

**Recommendation based on the dataset**

Based on the available data, here are some recommendations that can be made to government agencies to improve healthcare.

1. Regular health check-ups: Regular health check-ups should be encouraged for the general population to monitor blood pressure, body temperature, and heart rate. This can help identify any potential health risks early on and allow for timely interventions.
2. Blood pressure monitoring: It can lead to serious health complications. Therefore, healthcare providers should regularly monitor blood pressure in patients, especially those at high risk.
3. Diabetes screening: Diabetes is a chronic condition that affects millions of people worldwide. Healthcare providers should screen patients for diabetes regularly, especially those at high risk, such as those with a family history of diabetes or who are overweight.
4. Health education and awareness: Educating people about the importance of maintaining a healthy lifestyle, including regular exercise, proper nutrition, and stress management.

5. Access to healthcare: Access to healthcare is essential for ensuring that everyone has access to the medical services they need. Government agencies should work to improve access to healthcare services, especially in underserved communities, by increasing funding for healthcare facilities and programs, providing incentives to healthcare providers.

**Conclusion**

In this report, we analyzed a dataset consisting of 1014 observations and 7 variables, including age, blood pressure, blood sugar level, body temperature, heart rate, and risk level. We explored the dataset using various graphical and numerical methods, including histograms, box plots, correlation matrices, and descriptive statistics. We also tested two hypotheses using statistical techniques, including a two-sample t-test.

**Refrences:**


1. National Institute on Aging. High Blood Pressure. https://www.nia.nih.gov/health/high-blood-pressure
2. American Heart Association. Understanding Blood Pressure Readings. https://www.heart.org/en/health-topics/high-blood-pressure/understanding-blood-pressure-readings