**SECP1513**
**TECHNOLOGY AND INFORMATION SYSTEM**

# HEALTHCARE DATA SETS FOR DIABETES

SECTION 11


| NO. | NAME | MATRIC NO. |
|---|---|---|
| 1 | FARAH IZZAH BINTI MOHD AISHA NUDDIN | A21EC0022 |
| 2 | NOR ALYA BT MOHAMAD NOR | A21EC0105 |
| 3 | NURFAZLIANA SURAYA BINTI BAHARUDDIN | A21EC0118 |
| 4 | KHALISAH NAJAH BINTI NAWAWI | A21EC0038 |
| 5 | MUHAMAD HAIKAL BIN MAT ARMAN | A21ET0101 |


**LECTURER'S NAME** : DR. MUHAMMAD IQBAL TARIQ BIN IDRIS

**SUBMISSION DATE** : 6TH JANUARY 2022

# ORGANISATION CHART

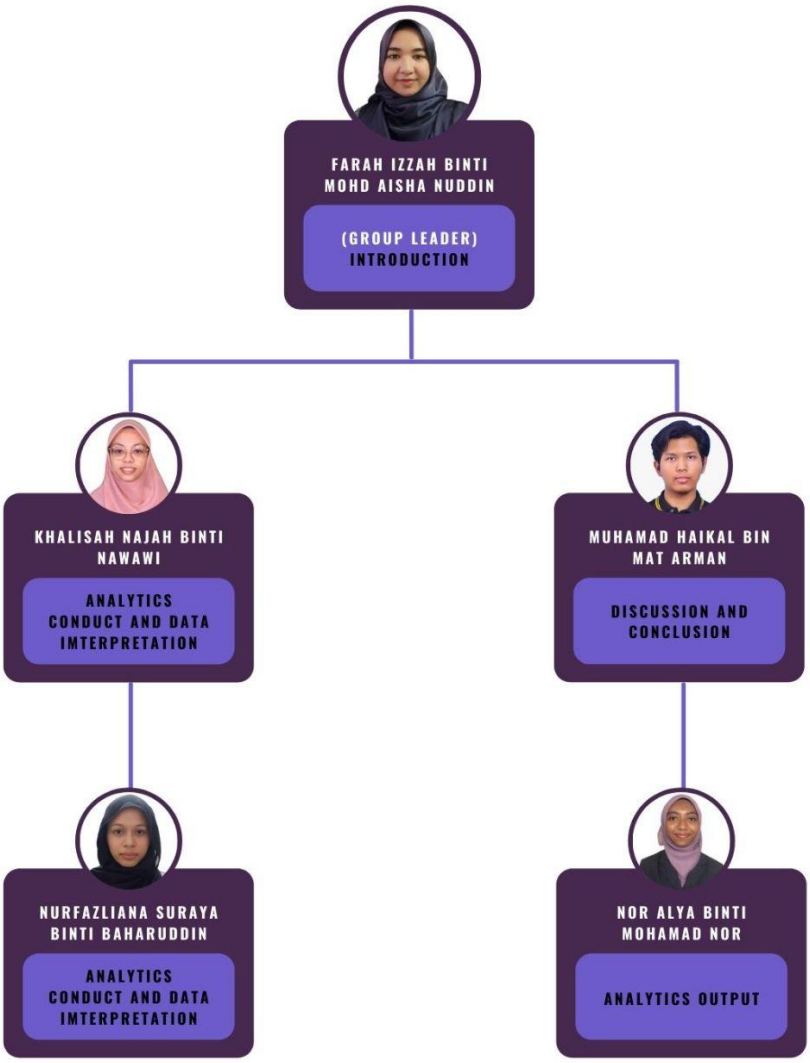


**TABLE OF CONTENT**

## INTRODUCTION

Data Analytics is a series of techniques aimed at extracting relevant and valuable information from extensive and diverse sets of data gathered from different sources and varying in sizes. It is also a process of analyzing raw datasets in order to derive a conclusion regarding the information they hold. Data Analytics processes and techniques may use applications operating on machine learning algorithms, simulation, and automated systems. Data Analytics exists in a lot of sectors such as retail, agriculture, and banking.

Retailers have always been laser-focused on getting the right products into the hands of the appropriate customers in the retail sector. Predictive data analytics are now being utilized to hyper-personalize the entire customer experience online, not just offering purchase recommendations. Next, in Africa, agriculture is the largest industry, but much of the land is currently unutilized. A data-driven platform is being developed by a research cooperation to study risk-sharing in order to improve farming methods. To encourage lenders to grant lower-risk loans, data science helps anticipate the value of modern farming practices. As for the banking sector, predictive analytics is having a big impact on this banking industry as well. Financial institutions can use the insights they gather to provide consumers with value-driven services that are customized for each individual, rather than pushing out mass marketing. For example, an European bank trying to boost retention of inactive customers turned to machine learning algorithms to predict which customers were most likely to reduce their activity with the bank. The data-driven program helped create a targeted marketing campaign that lowered customer churn by 15 percent.

For the reflection, I think that Data Analytics is very advanced and useful because it has a very important use in each data analytics type. For example, in descriptive analytics, it describes the happening over time, such as whether the number of views increased or decreased and whether the current month's sales are better than the last one. Other than that, diagnostic analytics focuses on the reason for the occurrence of any event. It requires hypothesizing and involves a much diverse dataset. It examines data to answer questions, such as "Did the weather impact the selling of umbrellas?" or "Did the new ad strategy affect sales?". Predictive and

prescriptive analytics also have their important use and function. Not only that, but data analyst also has a lot of roles in the business such as studying the information, cleaning it from noise, assessing the quality of data and its sources, developing the scenarios for automation and machine learning, and lastly is to oversee the proceedings.

Based on the Industrial Talk 7, Microsoft Power BI is a collection of software services and apps that work together to turn unrelated sources of data into coherent, visually immersive, and interactive insights. Power BI lets you easily connect to your data sources, visualize and discover what's important, and share that with anyone or everyone you want. Microsoft Power BI is used when you want to run reports and surface insights based on a company's data. Power BI can connect to a wide range of data sets, and "tidies up" the info it's fed so that it can be better digested and understood. The reports and visuals generated from this data can then be shared with other users. Power BI can be used anywhere as long as you have your device with you and then you just access it through the internet or directly through the apps if you already downloaded it.

As a reflection, I think that the talk explained the meaning and importance of data visualization such that information can be visualized in a number of ways, each of which can provide a specific insight. It is also important to identify and understand the story you are trying to tell with a proper visualization. There are also many types of data that can be visualized such as quantitative, discrete, continuous and categorical. Other than that, There also more than one data relationships such as nominal comparison, time series, correlation, ranking, deviation, distribution and part-to-whole relationships. Chart types are also categorized as bar chart, pie chart, line chart, scatter plot chart, bubble chart, and heat map variations.

## 1.0 DATA ANALYTICS

### 1.1 Analytics Conduct

In this study, we are going to create data visualization using Microsoft Power BI with the secondary data we have obtained which is a database of people diagnosed with diabetes with a total of 768 respondents. However, some of the patients' information is not completed, resulting in missing variables for certain patients. To get the exact result, the data has been filtered to only show the results for people with diabetes, to know if either the variables are a factor for a person to be prone to diabetes.

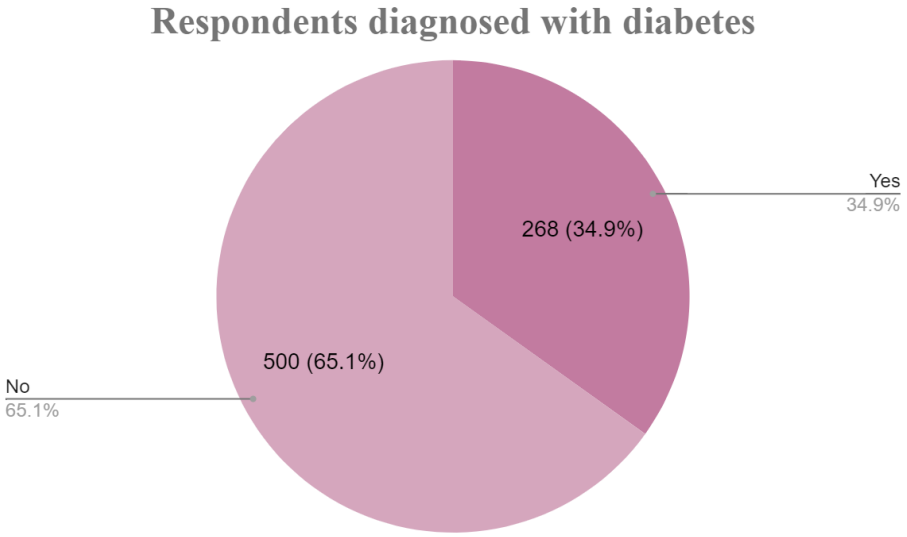



**Figure 1**: Pie chart of total of respondents diagnosed with diabetes or not

After filtering the patient list to be only those who are suffering from diabetes, we get a total of 268 respondents, from which we have collected their data. We have collected 5 variables from the respondents which are glucose levels, blood pressure level, insulin levels, BMI and age. Two questions have been construct to create the visualizations that tell a story about the data that we have obtained :

- Does a person's health status affect whether they are diagnosed with diabetes?
- Does age and weight play a part in the likeability of a person to get diabetes?

Therefore to get answers for these questions, we are to conduct a research on whether these variables are the factors for a person to be prone to diabetes.

1. **Dataset of respondent's glucose level**

| Range | Number of People |
|---|---|
| 0.00-19.90 | 2 |
| 19.91-39.80 | 0 |
| 39.81-59.70 | 0 |
| 59.71-79.60 | 1 |
| 79.61-99.50 | 13 |
| 99.51-119.40 | 54 |
| 119.41-139.30 | 63 |
| 139.31-159.20 | 51 |
| 159.21-179.10 | 45 |
| 179.11-199.00 | 39 |

**Table 1** : Respondent's glucose level

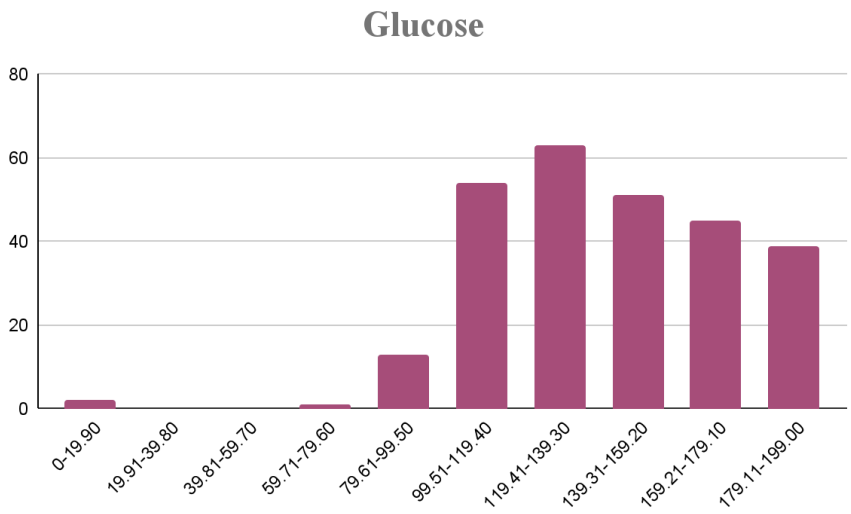



**Figure 2 :** Histogram of respondent's glucose level

**2. Dataset of respondents' blood pressure level**

| Range | Number of People |
|---|---|
| 0.00-12.20 | 16 |
| 12.21-24.40 | 0 |
| 24.41-36.60 | 1 |
| 36.61-48.80 | 2 |
| 48.01-61.00 | 20 |
| 61.01-73.20 | 86 |
| 73.21-85.40 | 94 |
| 85.41-97.60 | 38 |
| 97.61-109.80 | 8 |
| 109.81-122.00 | 3 |

**Table 2** : Respondents' blood pressure level

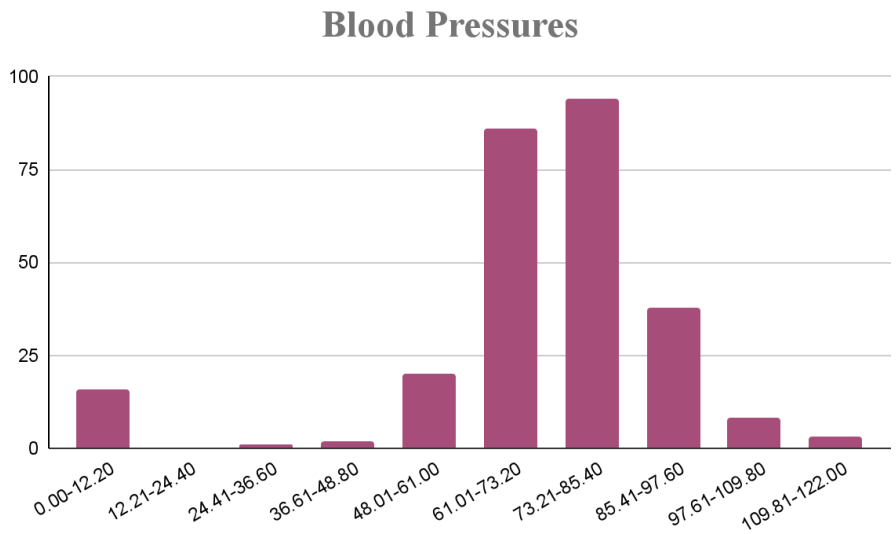



**Figure 3 :** Histogram of respondents' blood pressure level

### 3. Dataset of respondent's insulin level

| Range | Number of People |
|---|---|
| 0.00-84.60 | 147 |
| 84.61-169.20 | 56 |
| 169.21-253.80 | 36 |
| 253.81-338.40 | 14 |
| 338.41-423.00 | 3 |
| 423.01-507.60 | 6 |
| 507.61-592.20 | 4 |
| 592.21-676.80 | 1 |
| 676.81-761.40 | 0 |
| 761.41-846.00 | 1 |

**Table 3** : Respondent's Insulin level

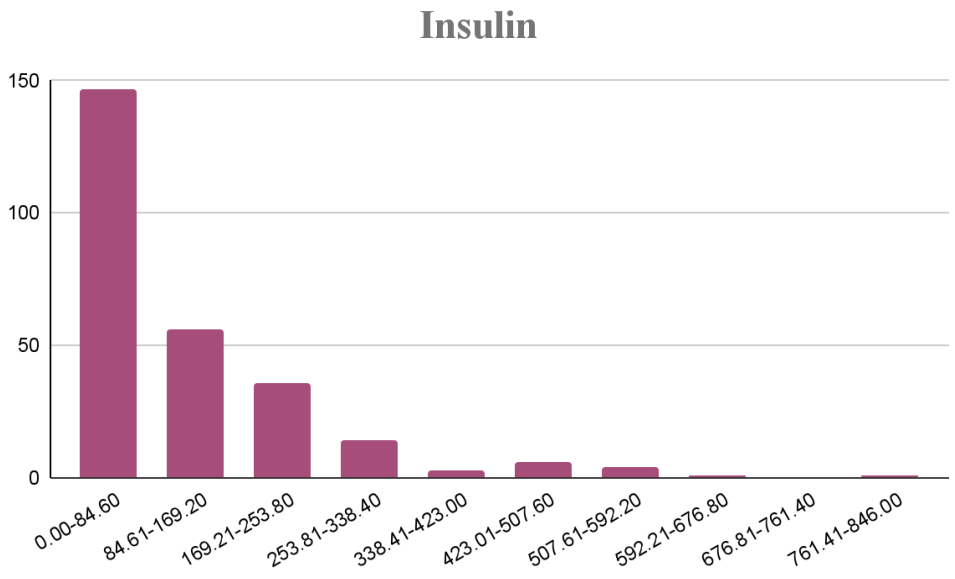
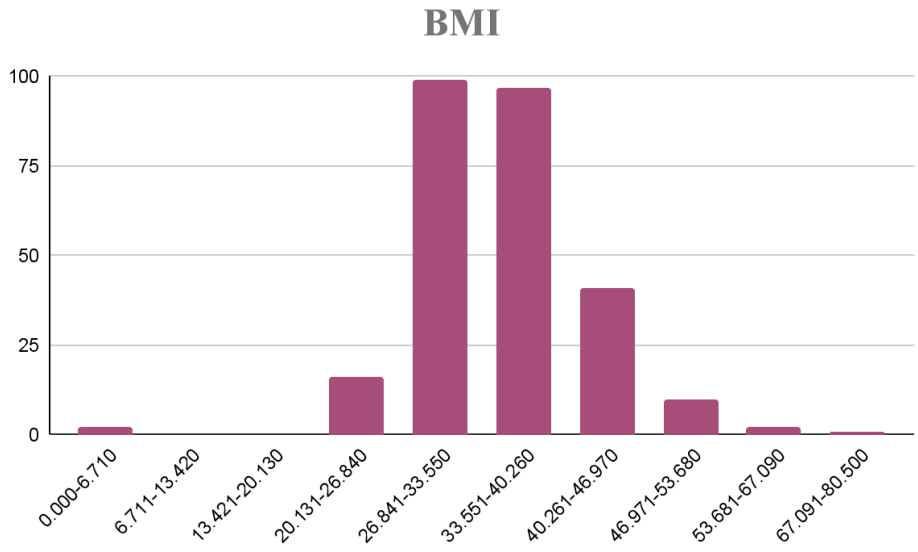



**Figure 4** : Histogram of respondent's Insulin level

**4. Dataset of respondents' BMI**

| Range | Number of People |
|---|---|
| 0.000-6.710 | 2 |
| 6.711-13.420 | 0 |
| 13.421-20.130 | 0 |
| 20.131-26.840 | 16 |
| 26.841-33.550 | 99 |
| 33.551-40.260 | 97 |
| 40.261-46.970 | 41 |
| 46.971-53.680 | 10 |
| 53.681-67.090 | 2 |
| 67.091-80.500 | 1 |

**Table 4 :** Respondents' BMI

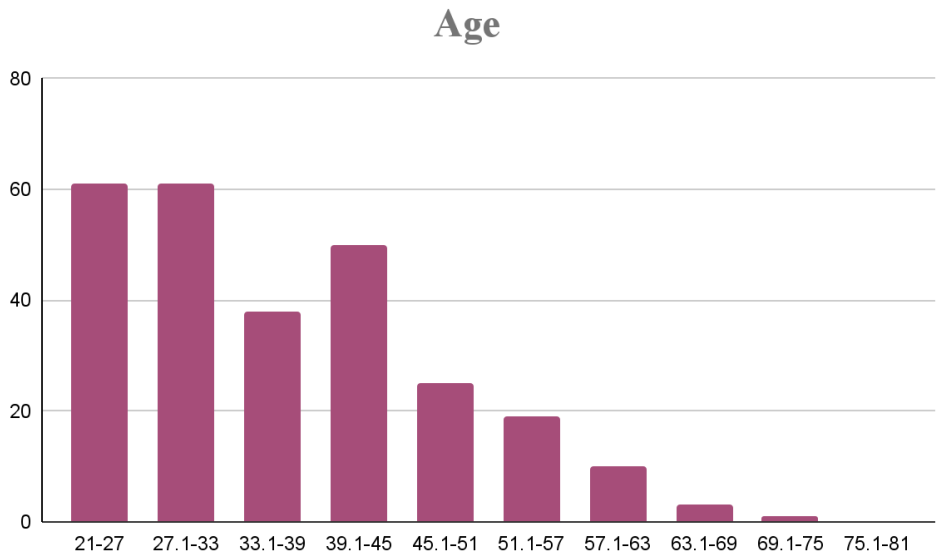



**Figures 5 :** Histogram of respondents' BMI

**5. Dataset of respondents' Age**

| Range | Number of People |
|---|---|
| 21-27 | 61 |
| 27.1-33 | 61 |
| 33.1-39 | 38 |
| 39.1-45 | 50 |
| 45.1-51 | 25 |
| 51.1-57 | 19 |
| 57.1-63 | 10 |
| 63.1-69 | 3 |
| 69.1-75 | 1 |
| 75.1-81 | 0 |

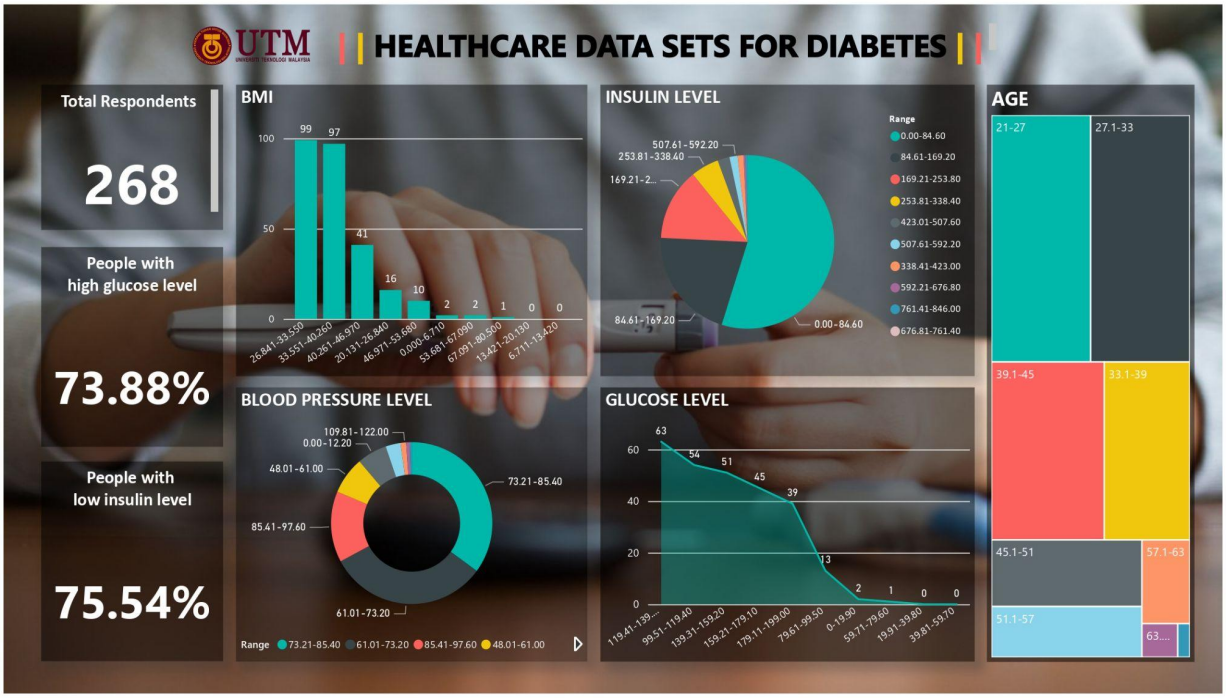
**Table 5 :** Respondents' Age




**Figures 6 :** Histogram of respondents' age

## 1.2 Analytics Output

After developing data visualization using data from Kaggle, we explore it and perform data analytics using Microsoft Power BI.




**Figures 7 :** Data analytics using Microsoft Power BI

### 1.3 Data Interpretation

We can observe that glucose level does have a role in whether a person has diabetes or not based on the results for variables we collected among individuals diagnosed with diabetes. According to the data gathered, many diabetic individuals had high glucose levels, ranging from 119.41 mg/dL and above. After eating, a normal person's glucose level is 90 to 110 mg/dL, whereas persons who haven't eaten for at least eight hours (fasting) have a glucose level of less than 100 mg/dL. As an outcome, we can deduce that those with a blood sugar level of 110 mg/dL or more are more likely to develop diabetes than others.To support my claim, the percentage of persons with high glucose levels is 73.88 percent, which includes nearly all of the respondents; the calculation used to determine the percentage of people with high glucose levels is shown below.

| Range | Number of People |
|---|---|
| 0-19.90 | $\frac{2}{268} \times 100 = 0.75\%$ |
| 19.91-39.80 | $\frac{0}{268} \times 100 = 0.00\%$ |
| 39.81-59.70 | $\frac{0}{268} \times 100 = 0.00\%$ |
| 59.71-79.60 | $\frac{1}{268} \times 100 = 0.37\%$ |
| 79.61-99.50 | $\frac{13}{268} \times 100 = 4.85\%$ |
| 99.51-119.40 | $\frac{54}{268} \times 100 = 20.15\%$ |
| 119.41-139.30 | $\frac{63}{268} \times 100 = 23.51\%$ |
| 139.31-159.20 | $\frac{51}{268} \times 100 = 19.03\%$ |
| 159.21-179.10 | $\frac{45}{268} \times 100 = 16.79\%$ |
| 179.11-199.00 | $\frac{39}{268} \times 100 = 14.55\%$ |

**Table 6** : Calculation of percentage for respondents' glucose level

$$23.51 + 19.03 + 16.79 + 14.55 = 73.88\%$$

Therefore, it is proved that high glucose levels cause people to be more prone to diabetes.

Next, from the result for variables we have collected amongst patients diagnosed with diabetes, it shows that the insulin level of the respondents also helps to determine the possibility that they are diabetis. Insulin is a vital part of metabolism. If there's not enough insulin, glucose can't get into your cells and it will stay in the bloodstream instead. This will lead to hyperglycemia, which means blood glucose levels that are too high. Diabetic patient pancreatic cells that normally produce insulin are destroyed and lead them to have low insulin levels. Having an insulin level less than 100 mg/dL is counted as not normal and too low. To further prove my statement, the percentage of people with low insulin levels is 75.54% which is composed of almost all of the respondents. Below is the calculation done to know the percentage of people with low insulin levels.

| Range | Number of People |
|---|---|
| 0.00-84.60 | $\frac{147}{268} \times 100 = 54.65\%$ |
| 84.61-169.20 | $\frac{56}{268} \times 100 = 20.90\%$ |
| 169.21-253.80 | $\frac{36}{268} \times 100 = 13.43\%$ |
| 253.81-338.40 | $\frac{14}{268} \times 100 = 5.22\%$ |
| 338.41-423.00 | $\frac{3}{268} \times 100 = 1.12\%$ |
| 423.01-507.60 | $\frac{6}{268} \times 100 = 2.24\%$ |
| 507.61-592.20 | $\frac{4}{268} \times 100 = 1.49\%$ |
| 592.21-676.80 | $\frac{1}{268} \times 100 = 0.37\%$ |
| 676.81-761.40 | $\frac{0}{268} \times 100 = 0.00\%$ |
| 761.41-846.00 | $\frac{1}{268} \times 100 = 0.37\%$ |

**Table 7** : Calculation of percentage for respondents' insulin level

$54.64\% + 20.90\% = 75.54\%$

Thus, it is proved that low levels of insulin are the reason for people to get diabetes.

## 2.0 DISCUSSION

Based on data of respondents diagnosed with diabetes,from the total of 768 respondents we can see that 268 (34.9%) of the respondents suffer from diabetes and 500 (65.1%) of the respondents are exempt from diabetes.

Based on data of glucose level that have been collected from the respondents in table 1, there are 2 respondents with glucose levels of 0.00-19.0. The trend of data keeps escalating from glucose level of 59.71-79.60 with only one respondent, glucose level of 99.51-119.40 with 54 respondents and glucose level of 119.41-139.30 with 63 respondents. Then, data declined from glucose level of 139.31-159.20 with 51respondents, glucose level of 159.21-179.10 with 45 respondents and glucose level of 179.11-199.00 with 39 respondents.

Based on table 2 that shows blood pressure level of respondents, data records 16 respondents with blood pressure of 0.00-12.20. The data start to show the increasing of people from blood pressure of 24.41-36.60 with only one respondent, blood pressure of 36.61-48.80 with 2 respondents, 48.01-61.00 with 20 respondents, 61.01-73.20 with 86 respondents and 73.21-85.40 with 94 respondents which is the highest number of people recorded. Then the data declined from 85.41-97.60 with 38 respondents, 97.61-109.80 with 8 respondents and 109.81-122.00 with only 3 respondents.

Based on the data that been gathered in table 3, shows declining of trend with 147 respondents with insulin level of 0.00-84.60, 56 respondents with insulin level of 84.61-169, 36 respondents with insulin level of 169.21-253.80, 14 respondents with 253.81-338.40 of insulin level , 3 respondents with 338.41-423.00 of insulin level. Data show an increase for insulin level of 423.01-507.60 with 6 people then declined to 4 respondents with 507.61-592.20 insulin level and one respondent with insulin level of 592.21-676.80 and 761.41-846.00.

From Table 4 that collects respondent's BMI, we can see that 2 people with BMI range 0.000-6.710 and increasing to 16 respondents with BMI range 20.131-26.840 and 99 respondents with BMI range 26.841-33.550. Data declined to 97 respondents with BMI range

33.551-40.260 then to 41 respondent with BMI range 40.261-46.970, 10 people with BMI range 46.971-53.680,
2 respondents with BMI range 53.681-67.090 and one respondent with 67.091-80.500.

Table 5 collects the respondent's age, age of 21-27 and 27.1-33 both have 61 respondents while 38 respondents with the age of 33.1-39 and 50 people within the age of 39.1-45. Trend shows decreasing in respondents from the range of age of 51.1-57 with 19 people, age within 57.1-63 with 10 respondents, 63.1-69 with 3 respondents and within age of 69.1-75 with only one respondent.

## 3.0 CONCLUSION

Based on the question of whether a person's health status affects whether they are diagnosed with diabetes, we can relate it with calculations made in table 6 and conclude that it have been proved that high glucose levels cause people to be more prone to diabetes and for the question of age and weight play a part in the likeability of a person to get diabetes have been proved in table 7. In conclusion, the health, age and weight of a person play a part in the likeability of a person to get diabetes.

**REFERENCE**


1. Rauch, S. (2021, June 7). *Top Four Industries Using Data Analytics to Thrive.* simplilearn. https://www.simplilearn.com/industries-using-data-analytics-to-thrive-article

2. Hart, M. (2021, September 22). *What Is Power BI?* Microsoft. https://docs.microsoft.com/en-us/power-bi/fundamentals/power-bi-overview

3. Team, S. (2021, May 11). *What are normal blood glucose levels?* The Checkup. Retrieved January 5, 2022, from https://www.singlecare.com/blog/normal-blood-glucose-levels/#:%7E:text=A%20normal%20blood%20glucose%20level,90%20to%20110%20mg%2FdL.