**Data Visualization**

**INTERACTIVE DATA VISUALIZATION PROJECT ON DIABETES US HOSPITAL**

**Assignment 2**

****

**Team Member**

**Sumit Sharma [10510947]  
Amit Goud [10505896]  
Jeet Udaiyar [10514319]**

**Introduction**

Our Assignment for B9DA106 are explained below with research and study on data visualization on data set using Diabetes Us Hospital where we have described and provide explanation about in details with reason ,aim, vision and queries on which data visualization is prepared using python and its library and have shown exposure on python.

**Dataset: DIABETES US HOSPITAL**

This information has been set up to dissect components identified with readmission just as different results relating to patients with diabetes. The dataset speaks to certain long stretches of clinical consideration at various US medical clinics and incorporated conveyance systems. It incorporates more than 50 highlights speaking to patient and medical clinic results. Data was removed from the database for experiences that fulfilled the accompanying criteria. (1) It is an inpatient experience (an emergency clinic confirmation). (2) It is a diabetic experience, that is, one during which any sort of diabetes was entered to the framework as a determination. (3) The length of stay was in any event 1 day and at most 14 days. (4) Laboratory tests were performed during the experience. (5) Medications were controlled during the experience. The information contains such traits as patient number, race, sexual orientation, age, affirmation type, time in clinic, therapeutic claim to fame of conceding doctor, number of lab test performed, HbA1c test outcome, analysis, number of drug, diabetic prescriptions, number of outpatient, inpatient, and crisis visits in the year prior to the hospitalization, and so forth.

Data can be found on [https://archive.ics.uci.edu/ml/datasets/Diabetes+130-US+hospitals+for+years+1999-2008#](https://archive.ics.uci.edu/ml/datasets/Diabetes+130-US+hospitals+for+years+1999-2008)

**TOOLS FOR VISUALIZATION: PYTHON, JUPYTER NOTEBOOK, TABLEAU**

**LIBRARIES: PANDAS, NUMPY, MATPLOTLIB, SEABORN**

**Aim:** The aim of this project is become familiar with the visualization implementing techniques after exploring the issues involved in implementing interactive visualization.

**Vision:** As mentioned in the aim of his assignment i.e., to gain familiarity of implementing visualization techniques for which we need a data source which can be used for visualization, we selected the database of diabetes patients in a hospital. This data source helped in visualizing different types of data. Data from this data source is visualized in different manners in this assignment. As the data source contains several patients’ details and it is an important information for people, so we used to visualize data from this data source. We used to visualize different dimensions of this data in the assignment for a better understanding of visualization techniques.

**Requirements for developing the Visualizations**

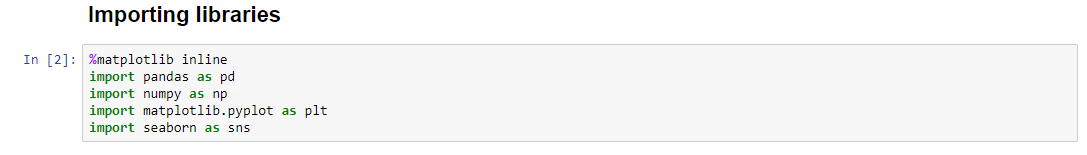
Below are the tools used to develop the, reports, datasets, design diagrams and visualizations:

1. Microsoft Word
2. Microsoft Excel (.csv)
3. Draw.io
4. Python
5. GitHub

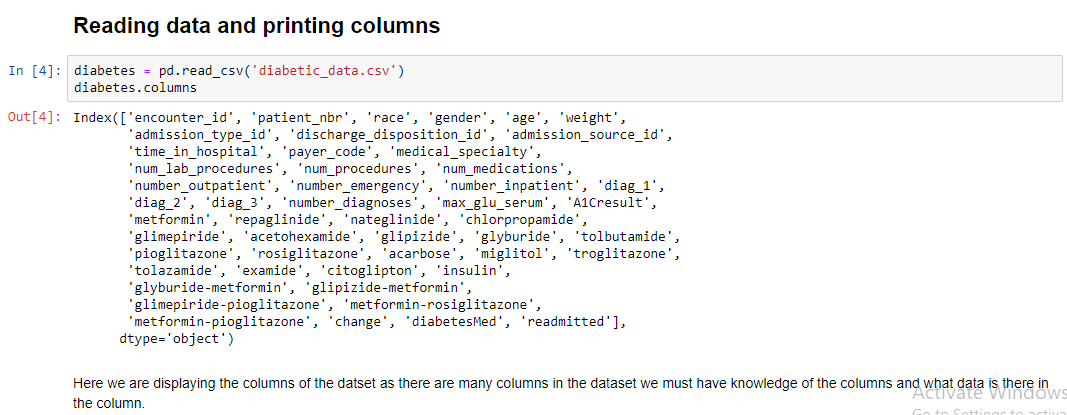
**Data visualizations based on below analytical questions**

1. Number of Male and Female diabetes patients in US?
2. Diabetes Patients from different Races?
3. Diabetes Patients by different age groups?
4. Patients attends by which medical specialist?
5. Metformin found the diabetics patients.
6. Repaglinide percentage in patients?
7. Insulin found the diabetics patients.
8. Improvement after treatment in patients?
9. Percentage of patient’s diabetes met after treatments.
10. Patients by admission types?
11. Patients who have NO chances to readmit.

In the following code snippets, we have imported python libraries which we are going to use in throughout data visualization.



Now I am reading CSV data file and displaying column’s names. Each column in data contains data which is somehow related to diabetes patients. The data is basically related to a hospital admission of diabetic patients. So, it could be called an inpatient or a diabetic encounter. The dataset contains several columns representing information of patients, hospitals performances etc.



**Data Set**

This information has been set up to dissect components identified with readmission just as different results relating to patients with diabetes.

The dataset speaks to certain long stretches of clinical consideration at various US medical clinics and incorporated conveyance systems. It incorporates more than 50 highlights speaking to patient and medical clinic results. Data was removed from the database for experiences that fulfilled the accompanying criteria.

(1) It is an inpatient experience (an emergency clinic confirmation).

(2) It is a diabetic experience, that is, one during which any sort of diabetes was entered to the framework as a determination.

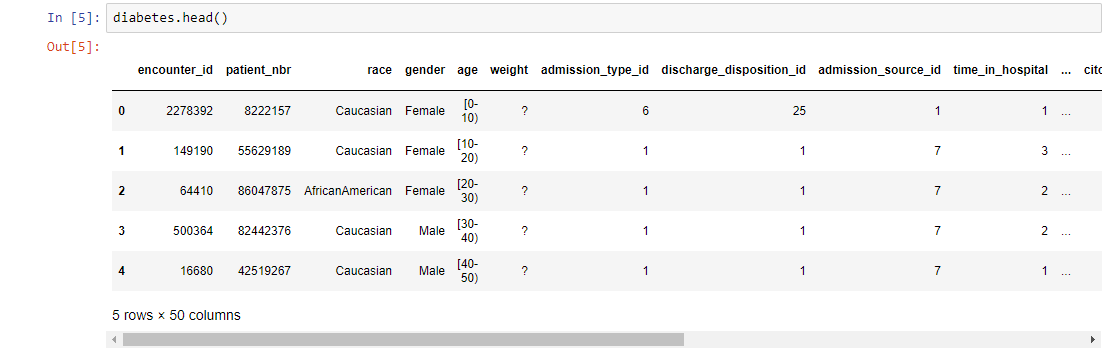
(3) The length of stay was in any event 1 day and at most 14 days.

(4) Laboratory tests were performed during the experience.

(5) Medications were controlled during the experience.

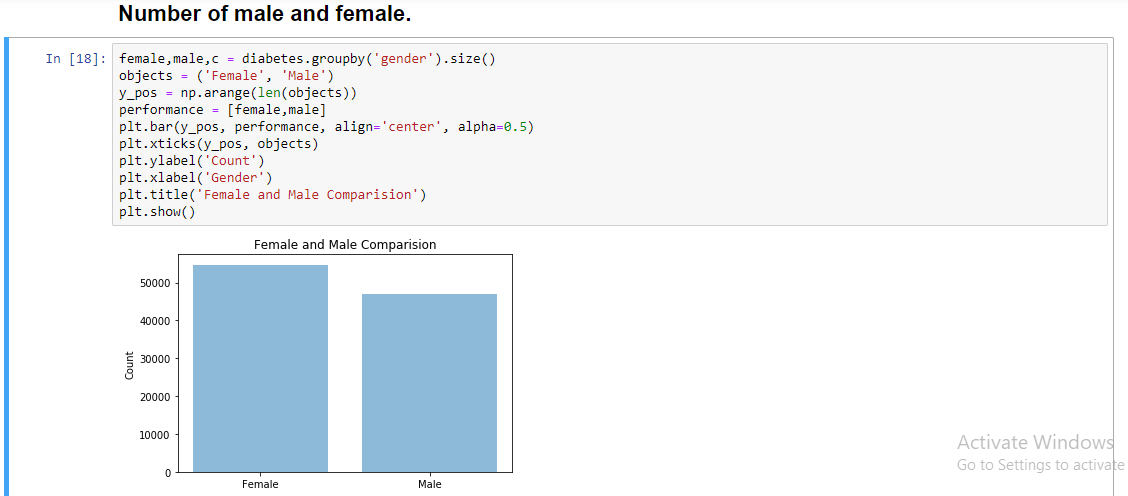
The information contains such traits as patient number, race, sexual orientation, age, affirmation type, time in clinic, therapeutic claim to fame of conceding doctor, number of lab test performed, HbA1c test outcome, analysis, number of drug, diabetic prescriptions, number of outpatient, inpatient, and crisis visits in the year prior to the hospitalization, and so forth.

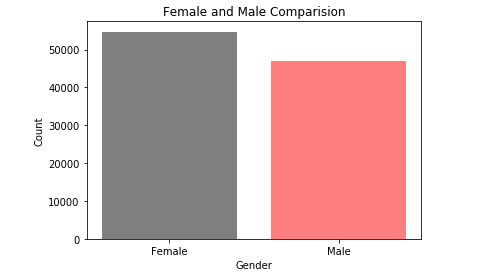
Printing 10 records form the dataset to get some idea how and what is stored in the columns. A few records of different diabetic patients are represented below. The following table is representing a little information about the treatment of admitted patients. The data is including information of both male and female diabetic patients.



**Gender comparison**

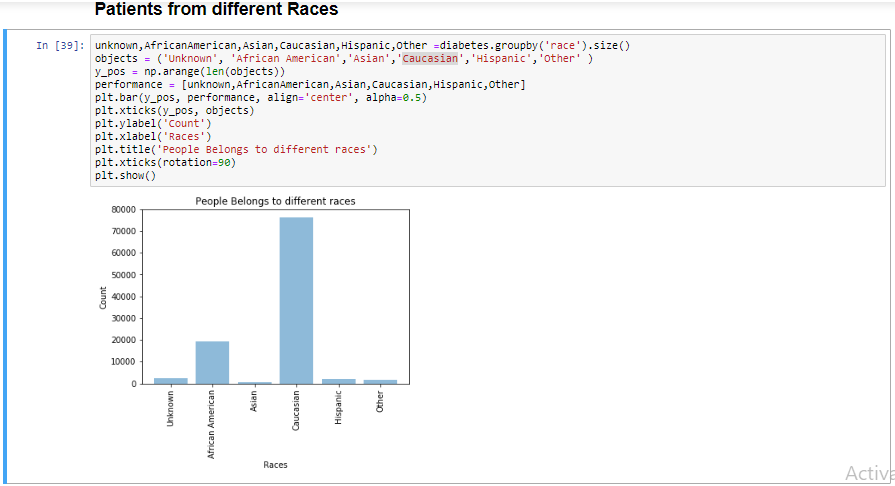
Comparing number of male and female diabetes patients. After performing data visualization operation, we got that females are more affected by the diabetes in comparison to male. Following graph is showing the comparison of male and female patients admitted in the hospital for their treatments. It shows that mostly females are affected from diabetes. Graph is used for a better visualization of this comparison.

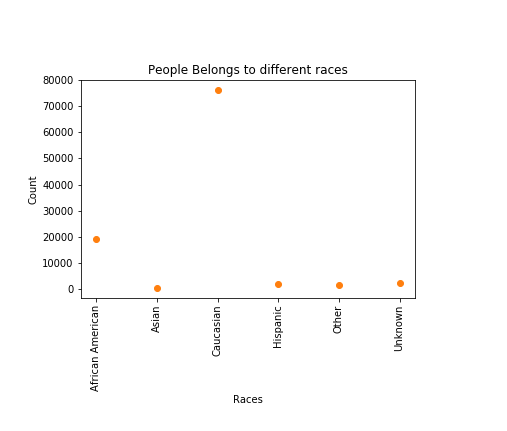




**Patients from different races**

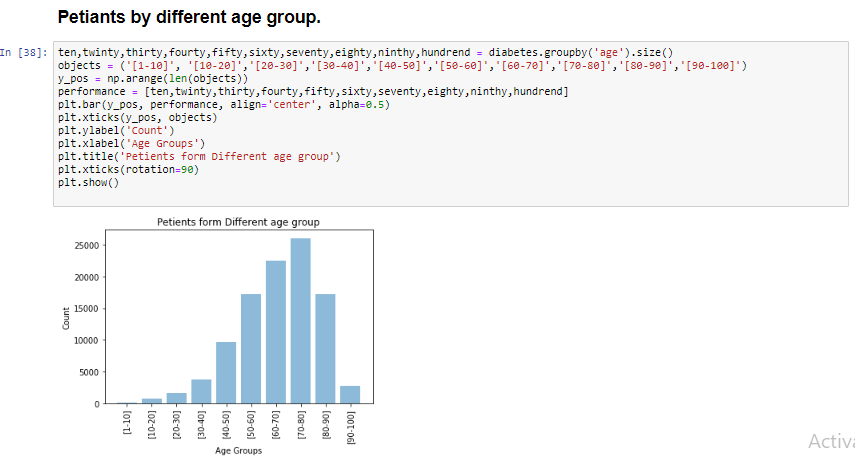
Now we are comparing number of diabetes patients belongs to different races. The patients admitted in the hospitals are from different races such as African, American, Asian, Caucasian, Hispanic, unknown and others. But after getting results we can say that largest number of patients belongs to Caucasian and least number of patients belongs to Asians. According to the graph, patients who are most admitted in hospitals belonged to races African America and Caucasian. Even some of the patients were not recognized by their races so we mentioned one of the races’ categories as unknown. From this comparison we concluded that Caucasians are mostly get affected from diabetes.

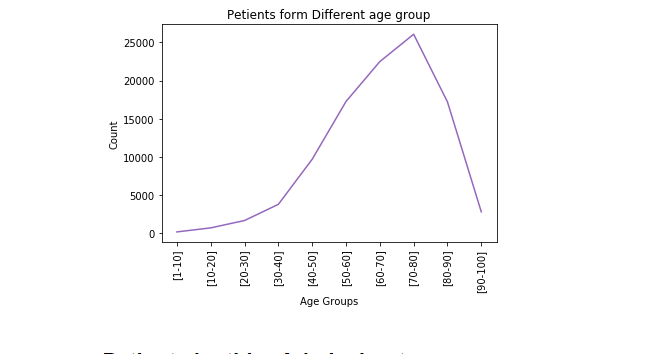




**Patients of different age groups**

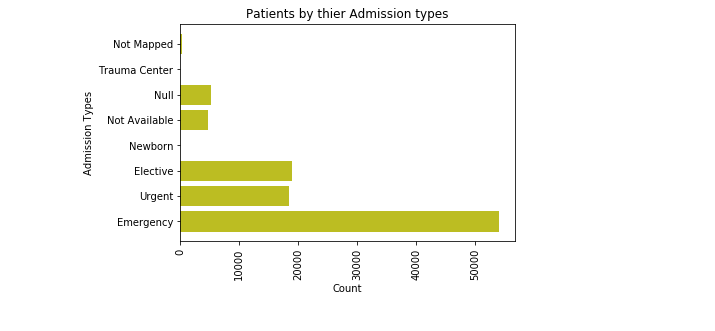
After performing data visualization on different age group, we get the following result. Highest number of diabetic’s patients are belonged to 70-80. In the ascending order, we can write it in the following order [1-10], [10-20], [20-30], [90-100], [30-40], [40-50], [50-60], [60-70], [70-80], [80-90]. To represent this comparison, we used to type graphs for better understanding and visualization. From both visualizations, it is represented that people after age 60 are mostly get affected from diabetes. The age group which is less affected from diabetes is 1-10 which includes infants and children.





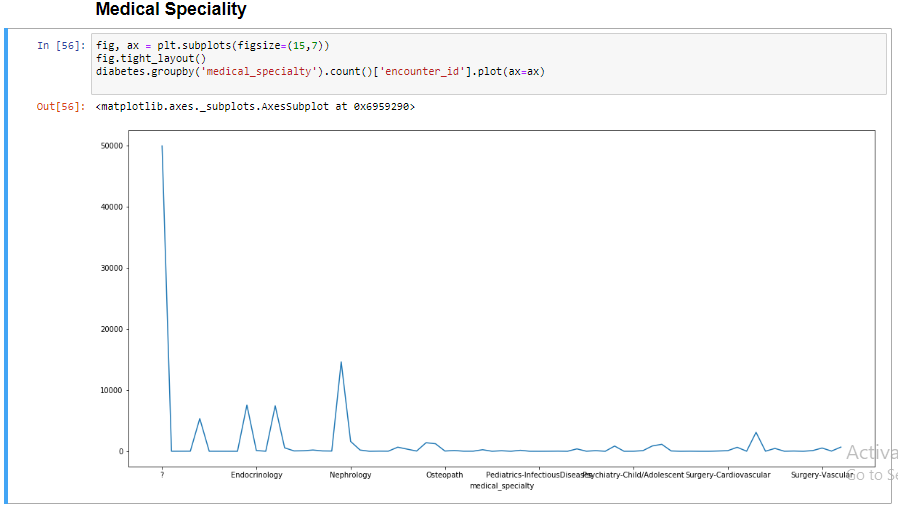
**Patients’ admission types**

Now we have visualized data based on the type of admission of the patient get admitted in the hospital. In the dataset admission type Ids are given which represents different type of admissions in the hospital. After visualization we can see that maximum number of patients are admitted in the hospital with emergency code. It means that some people are getting highly affected from this disease and admitted in the hospital for urgent treatment. Man, of the patients are also admitted in urgent code which means that they need to start their treatment on urgent bases so that this disease can not affect them highly.



**Medical specialist**

Following graph, gives information about which patients attends by which medical specialist. There are different medical specialists who treated several diabetic patients such as endocrinology, Nephrology, ode path and others. The results of this analysis are shown in the following graph. From this result it is concluded that most of the patients are treated through nephrology specialist.



**Metformin percentage**

Following pie graph gives information about how the metformin can be found the diabetics patients. After visualization we get the following results.

1. 80% patients do not contain metformin.

2. 18% patients contain steady metformin.

3. 1% patients contain UP metformin.

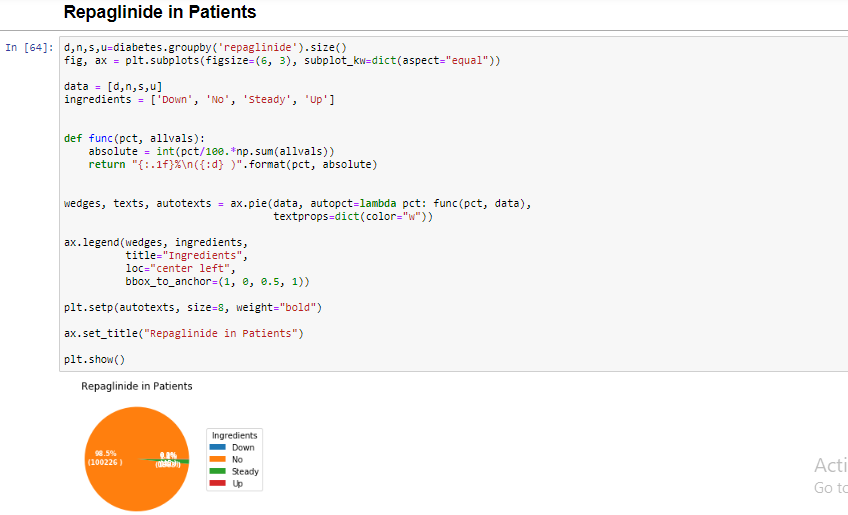
4. 0.4% patients contain DOWN metformin.

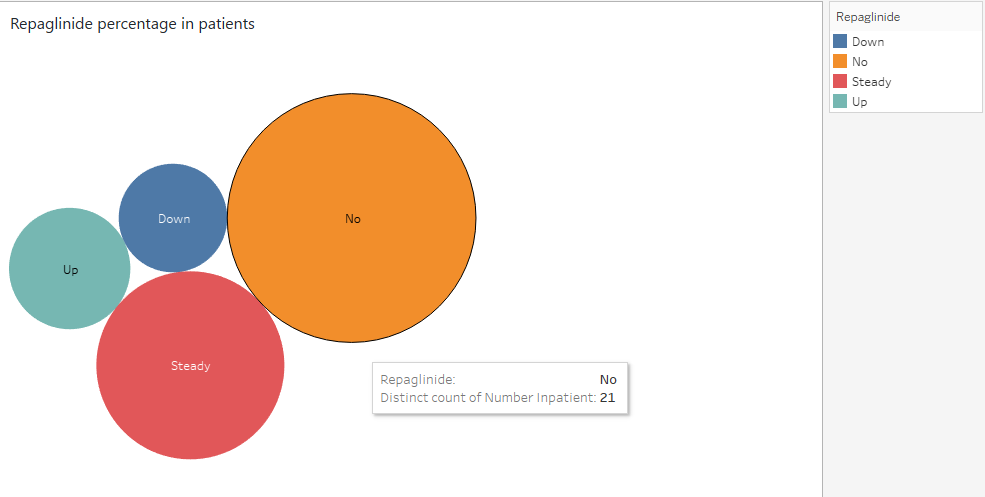
For better visualization of these results we used pie chart. Pie chart is an effective way for representing data in percentage form as it uses to make different sections for representing different percentages. In the following pie chart, we included the percentage of metformin which is found in diabetic patients. From this analysis, we found that very less patients i.e., 0.4% contains metformin and most of the patients i.e., 80% do not contain metformin.



**Repaglinide percentage**

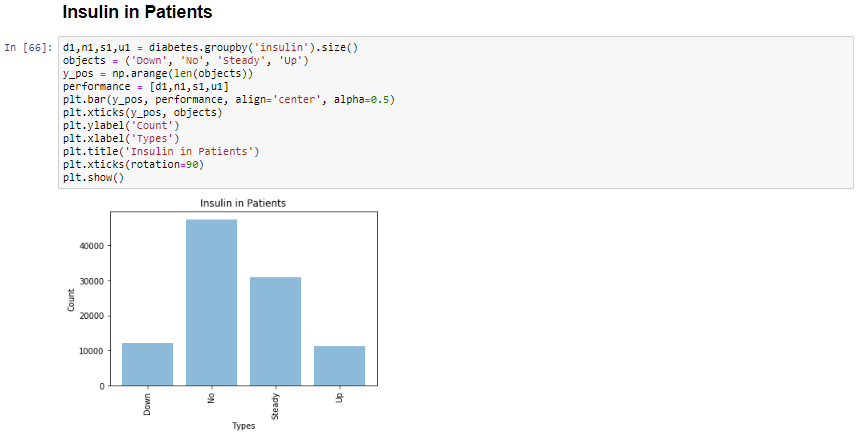
Following chart represents information of repaglinide percentage in patients. From this analysis, we found that most of the diabetic patients i.e., 98.5% do not contain repaglinide. We used a pie chart for representing his percentage also.

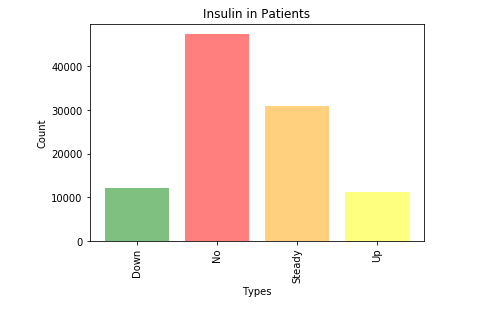




**Insulin in patients**

The followig graph is representing the ratio of insulin that we found in diabetc patients after analsis. From the following graph, it is clear that most of the patients do nt contain insulin. Many of patients are steady with containing insulin.





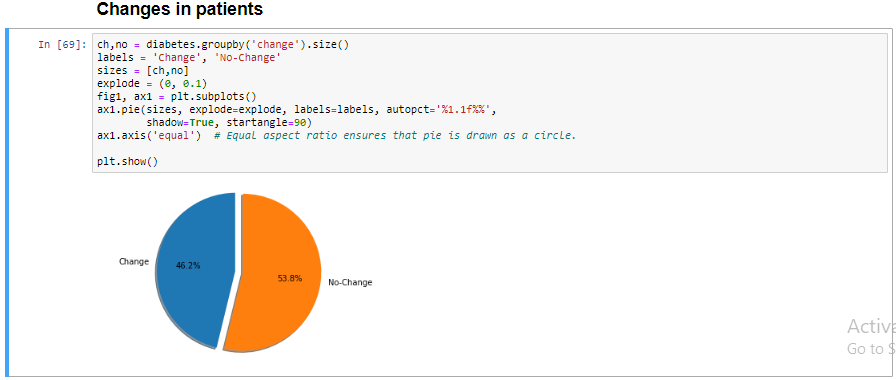
**Improvement after treatment**

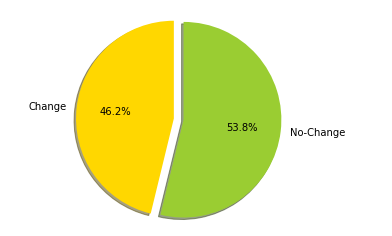
Following data visualization results represents percentage of patients changed or not changes after treatments.

1. 53.8% patients do not change after treatment.

2. 46.2% patients changed after treatment.

So, more than half patients have not recovered from their disease and only 46.2% patients have improved after treatment.





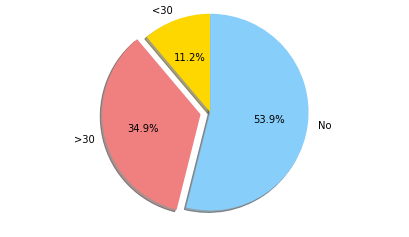
Following data visualization results represents percentage of patient’s diabetes met after treatments.

1. 23% patients YES after treatment.

2. 77% patients No after treatment.



**Patients who have NO chances to readmit.**

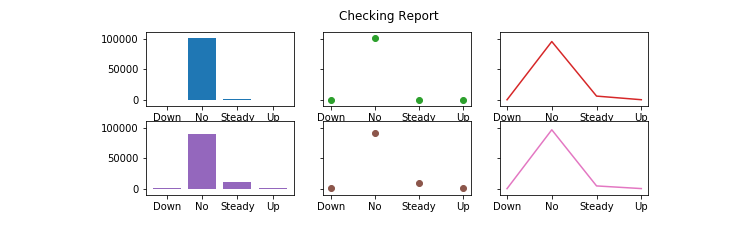


Patients who have more than 30% chances to readmit is 34.9%

Patients who have less than 30% chances to readmit is 11.1%

Patients who have NO chances to readmit is 53.9%

**Report for different gradient in blood sample of patients**



**Below is the work was distributed among the team members**

* Dataset research
* Data visualization queries
* Python coding
* GitHub configuration
* Report and documentation
* Video presentation

It was the great learning and experience developing this project in team each team member has given his 100% to achieve the task assign. By this exposure to different tools and technology will help us in future for industrial work.