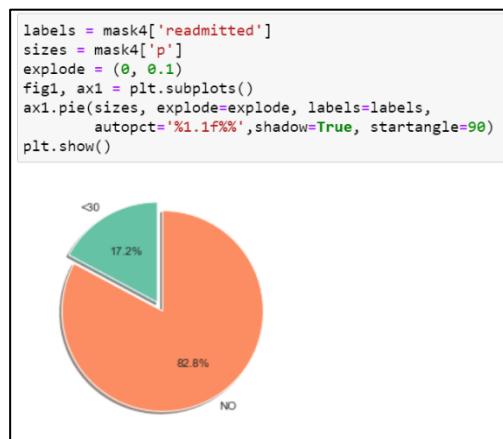# MVP

Diabetes Patients Readmitted within 30 days:

This project aims to highlight the risk factors may affecting on the rate of unscheduled readmission for diabetes patients within 30 days after discharge.

By applying EDA, the following part of this documents will demonstrate some findings in the data.

1- Re-admitted.
   The below graph shows that the datasets have 17% of patients readmitted after discharge.

```python
labels = mask4['readmitted']
sizes = mask4['p']
explode = (0, 0.1)
fig1, ax1 = plt.subplots()
ax1.pie(sizes, explode=explode, labels=labels,
        autopct='%1.1f%%',shadow=True, startangle=90)
plt.show()
```


2- Age.
   In this graph, we notice that most of patient is between 40-90 year. The highest rate of readmitted patients is in the range 50-80 year.

```python
fig = plt.gcf()
fig.set_size_inches(10, 6)
sns.set_style('darkgrid')
sns.set_palette('Set2')
sns.barplot(y='count',x='age',hue='readmitted',data=a_age,ci=500)
plt.title('Patients Age')
plt.xlabel('Age Group')
plt.ylabel('Total number')
sns.despine()
plt.show()
```
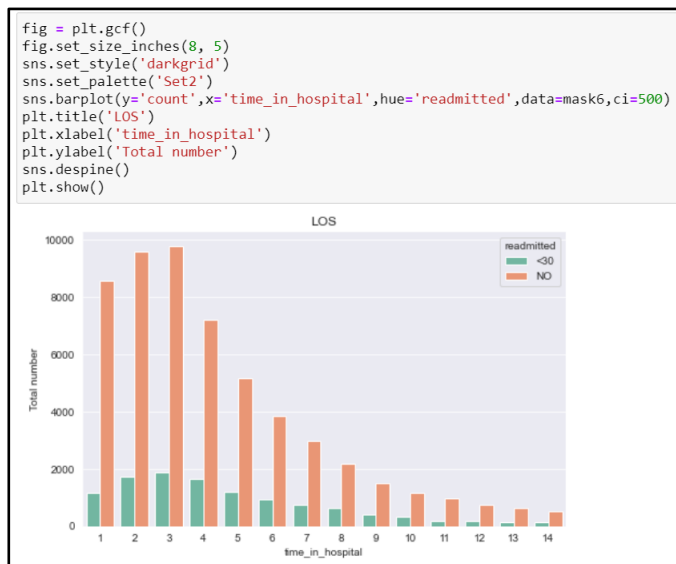
3- Gender.
Comparing the readmitted class with gender shows the risk of
readmission for female is higher probability than male.

```
fig = plt.gcf()
fig.set_size_inches(7, 4)
sns.set_style('darkgrid')
sns.set_palette('Set2')
sns.barplot(y='count',x='gender',hue='readmitted',data=a_gender,ci=500)
plt.title('Patients Gender')
plt.xlabel('Gender')
plt.ylabel('Total number')
sns.despine()
plt.show()
```


4- Time in hospital.
The following chart shows that the probability of readmitted is
increasing by decreasing the length of stay.

```
fig = plt.gcf()
fig.set_size_inches(8, 5)
sns.set_style('darkgrid')
sns.set_palette('Set2')
sns.barplot(y='count',x='time_in_hospital',hue='readmitted',data=mask6,ci=500)
plt.title('LOS')
plt.xlabel('time_in_hospital')
plt.ylabel('Total number')
sns.despine()
plt.show()
```

5- Race.

The below chart shows the risk of readmission is increasing in Caucasian compared to other population.

```
fig = plt.gcf()
fig.set_size_inches(8, 5)
sns.set_style('darkgrid')
sns.set_palette('Set2')
sns.barplot(y='count',x='race',hue='readmitted',data=mask7,ci=500)
plt.title('Race')
plt.xlabel('race')
plt.ylabel('Total number')
sns.despine()
plt.show()
```


Using Random Forest algorithm, the below graph shows the top 5 most significant factors of re-admission within 30 days.

```
features[features['importance']>0.05].plot(kind="bar",color='#5dbb7c')
```
`<AxesSubplot:xlabel='feature'>`
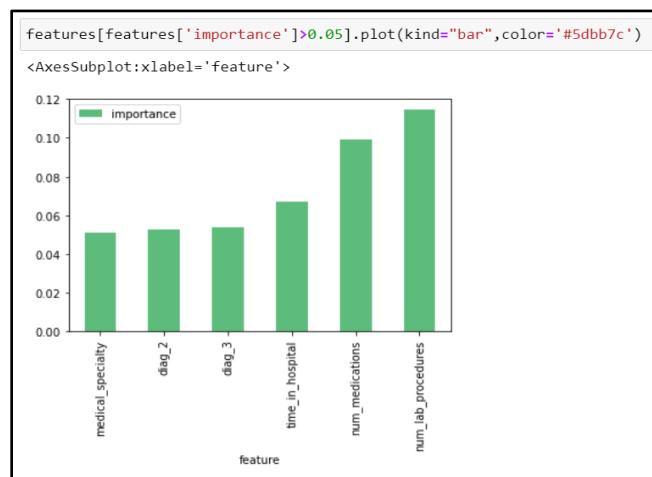
In addition, for next phase will using different algorithms and improving the model and enhancing the performance.