Project notes

The data I analyzed for this project is called No-show appointment dataset. This dataset includes the records of 100k medical appointment in Brazil. There are several variables in this dataset. They are PatientID, AppointemntID, Gender, ScheduledDay, AppointmentDay, Age, Neighborhood, Scholarship, Hipertension, Diabetess, Alcoholism, Handcap, SMS_received, and No-Show. In this project, I am going to explore which factors affect the No-show of appointment. I choose Gender, ,SMS_received, hypertension, diabetes as my independent variables and variable No-show as my dependent variables. My analysis will focus on 3 questions that I pose.

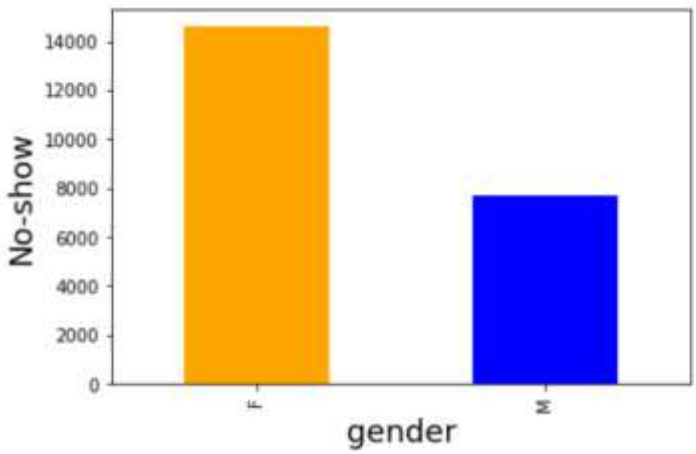
1. Whether gender is related to No-show of appointment?
2. Whether receiving SMS message would affect the rate of No-show?
3. Whether Hipertension and Diabetes affect the rate of No-show?

Data wrangling:

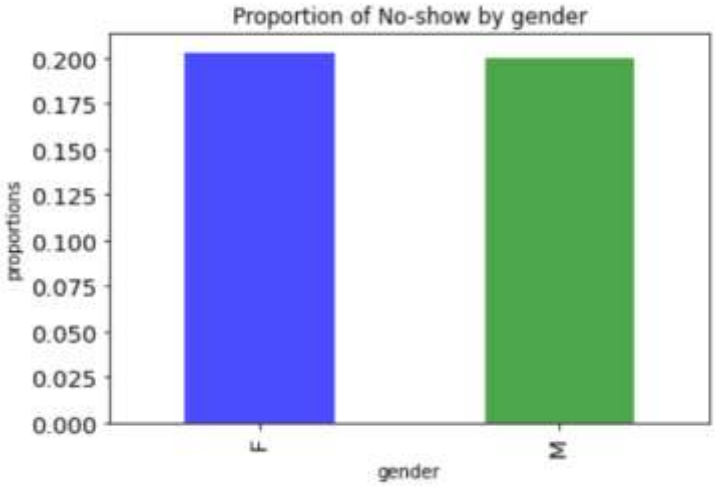
1. First, upload original dataset to Jupiter notebook, then read it in pandas.
2. Check the shape of dataset, data type of each variable, data info, unique values of each variables, and duplicates
   (After checking, found there is no missing values, no incorrect data types, no duplicates)
3. Start cleaning data, as no missing values, no need to fill not the null value with the mean of the column, also no duplicates so no need to check for duplicates to drop those rows. So here, only need to drop some columns not needed for my project. For example, PatientID, AppointmentID, ScheduledDay, Appointment Day, Neighborhood are specific to patients. So drop these columns and saved the dataset. .

Q1: Whether gender is related to No-show of appointment?

Summary: From the bar graph (refer to graph 1) , we can see that female are more likely to showing up for the appointment, however, if we take look at the proportion (refer to graph 2) , we found there is no slight difference between these two genders. Conclusion would be gender does not affect the rate of no show.



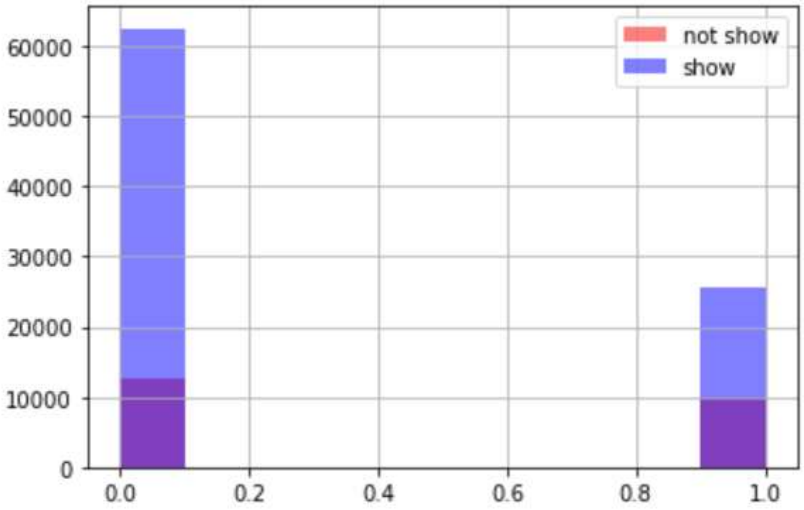
(graph1)



(graph 2)

Q2: Whether receiving SMS message would affect the rate of No-show?

Summary statistics: No-show for sms_received and not sms_received are 0.43 and 0.29, use histgram to see the distribution of female and male, found most patients who come to the appointment are those not received the message. So it is able to conclude that receiving message would increase rate of no_show of the appointment.



4. **Whether Hipertension and Diabetes affect the rate of No-show?**
   Summary statistics: The no show for patients having or not having diabetes are based on the data below. There is a pie graph visualize the results, but we cannot conclude that diabetes affect the rate of no show. Also take look at the hipertension, the rate of hipertension for patients who come to the appointment is 0.2. The rate of hypertension

for patients who do not make the appointment is 0.16. Below is the bar graph for both groups. So hipertension does affect the rate of no show.

```
:  Diabetes  No-show
   0         No         81695
             Yes        20889
   1         No          6513
             Yes         1430
```