UDACITY

Project 2: showing up in medical appointments

Introduction:

The dataset collects information from 100k medical appointments in Brazil and is focused on the question of whether patients show up for their appointment:

- 'ScheduledDay' tells us on what day the patient set up their appointment.
- 'Neighborhood' indicates the location of the hospital.
- 'Scholarship' indicates whether the patient is enrolled in Brasilian welfare program Bolsa Família.
- 'age' indicates the age of patient
- 'No-show' indicates 'No' if the patient showed up to their appointment, and 'Yes' if they did not show up.
- 'SMS_received' indicate if the patients have received SMS massage before their appointment
- 'Gender' Male Or Female
- 'Hipertension' known as high blood pressure. Part of patient's medical history(0,1).
- 'Diabetes' Part of patient's medical history(0,1)
- 'Alcoholism' drinking of alcohol that results in significant mental or physical health problems(0,1)
- 'Handcap' Part of patient's medical history(0,1)

Main dependent variable is appointment status 'No-Show'; the rest of variable will be the independent ones

Questions:

I collect 3 questions to investigate about:

- does the Gender have an important role in a person to come to his appointment?
- does the location of the hospital Neighborhood would be reason for the patient to come to his appointment?
- if the patient has diseases is that will cause the patient to show up to his appointment?

In my point of view for question one the gender female specially will care to show in appointments more than male , and for question two that the location of the hospital Neighborhood would be reason for the patient to come to his appointment I guess that there is some locations have more population than other locations in brazil and for that last question if the patient has diseases is that will cause the patient to show up to his appointment , in my opinion , patients will care more , since the disease may increase if they neglect their appointments, so they are keen to come to their appointment

Data Wrangling:

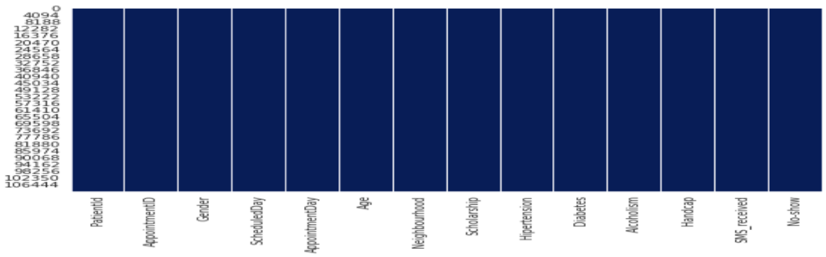
Duplicate:

```
# remove the Duplicate row
df.drop_duplicates(inplace=True)
```

d:

Null:

```
#The heatmap under shows how much data are missing! , there's no missing data
fig, ax = plt.subplots(figsize=(9,5))
sns.heatmap(df.isnull(), cbar=False, cmap="YlGnBu_r")
plt.show()
```



UDACITY

Normalization:

```
# replace NO to 1 & YES to 0:the number easy to handle more than string
df["No-show"].replace({"No":1, "Yes":0}, inplace=True)
print(df)
```

```
             PatientId  AppointmentID Gender  ... Handcap SMS_received  No-show
0         2.987250e+13        5642903      F  ...       0            0        1
1         5.589980e+14        5642503      M  ...       0            0        1
2         4.262960e+12        5642549      F  ...       0            0        1
3         8.679510e+11        5642828      F  ...       0            0        1
4         8.841190e+12        5642494      F  ...       0            0        1
...                ...            ...    ...  ...     ...          ...      ...
110522    2.572130e+12        5651768      F  ...       0            1        1
110523    3.596270e+12        5650093      F  ...       0            1        1
110524    1.557660e+13        5630692      F  ...       0            1        1
110525    9.213490e+13        5630323      F  ...       0            1        1
110526    3.775120e+14        5629448      F  ...       0            1        1

[110527 rows x 14 columns]
```

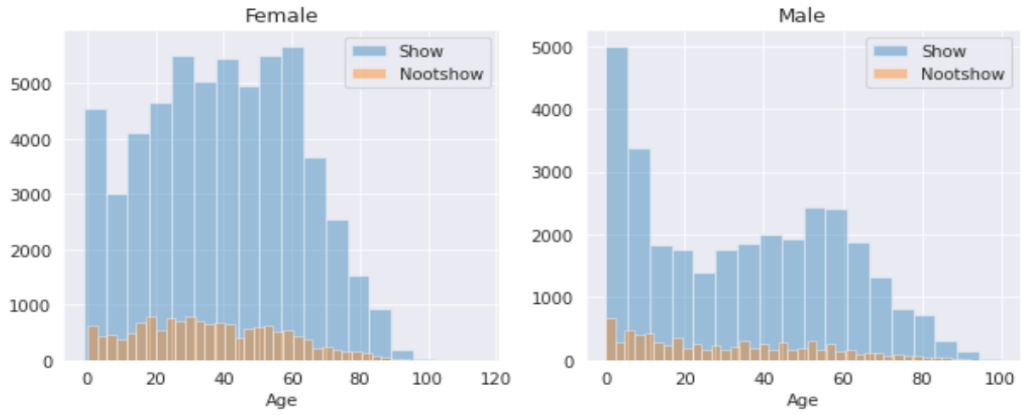
Remove unwanted columns:

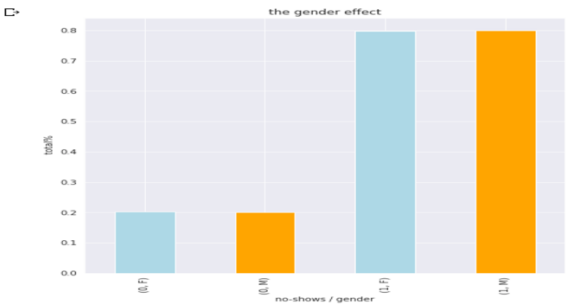
```
[10] #  remove 'PatientId' and 'AppointmentID' :  unwanted columns .
     df.drop(['AppointmentID','PatientId'],axis=1,inplace=True)
     df.head(1)
```

	Gender	ScheduledDay	AppointmentDay	Age	Neighbourhood	Scholarship	Hipertension	Diabetes	Alcoholism
0	F	2016-04-29T18:38:08Z	2016-04-29T00:00:00Z	62	JARDIM DA PENHA	0	1	0	0

Exploratory Data Analysis:

1. **does the Gender have an important role in a person to come to his appointment?**

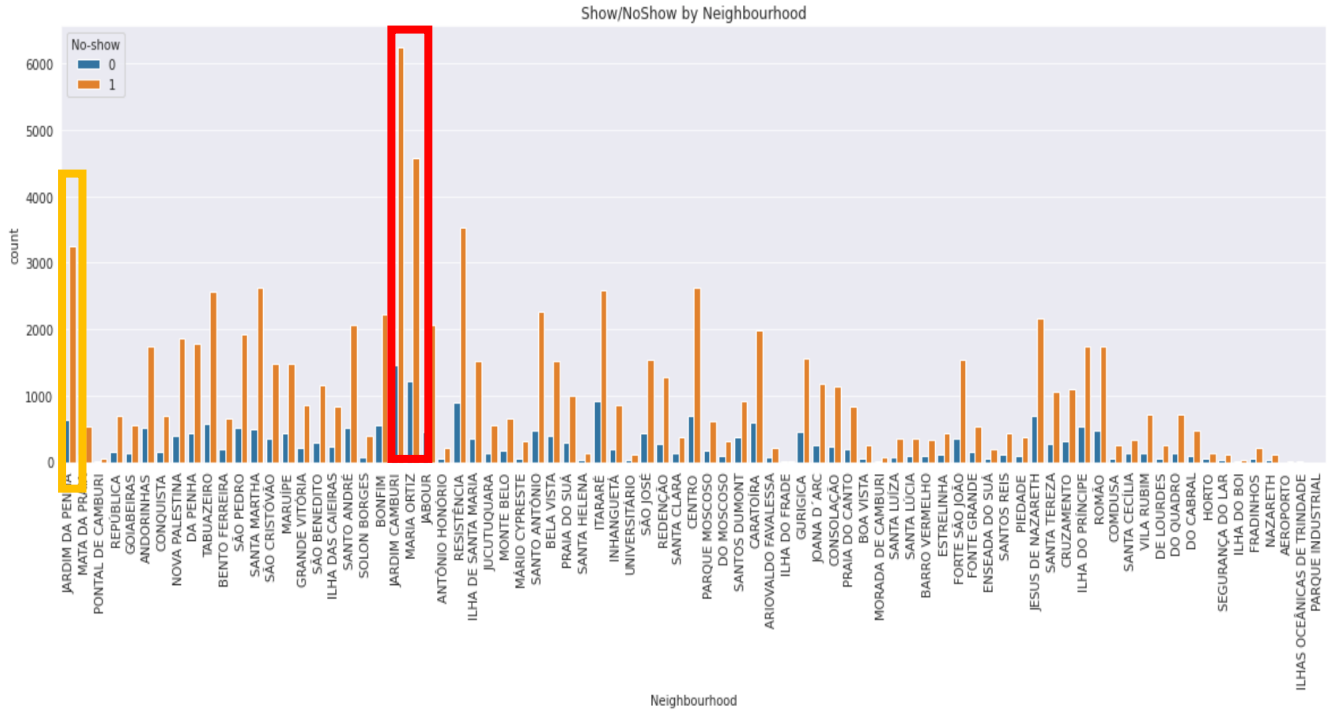


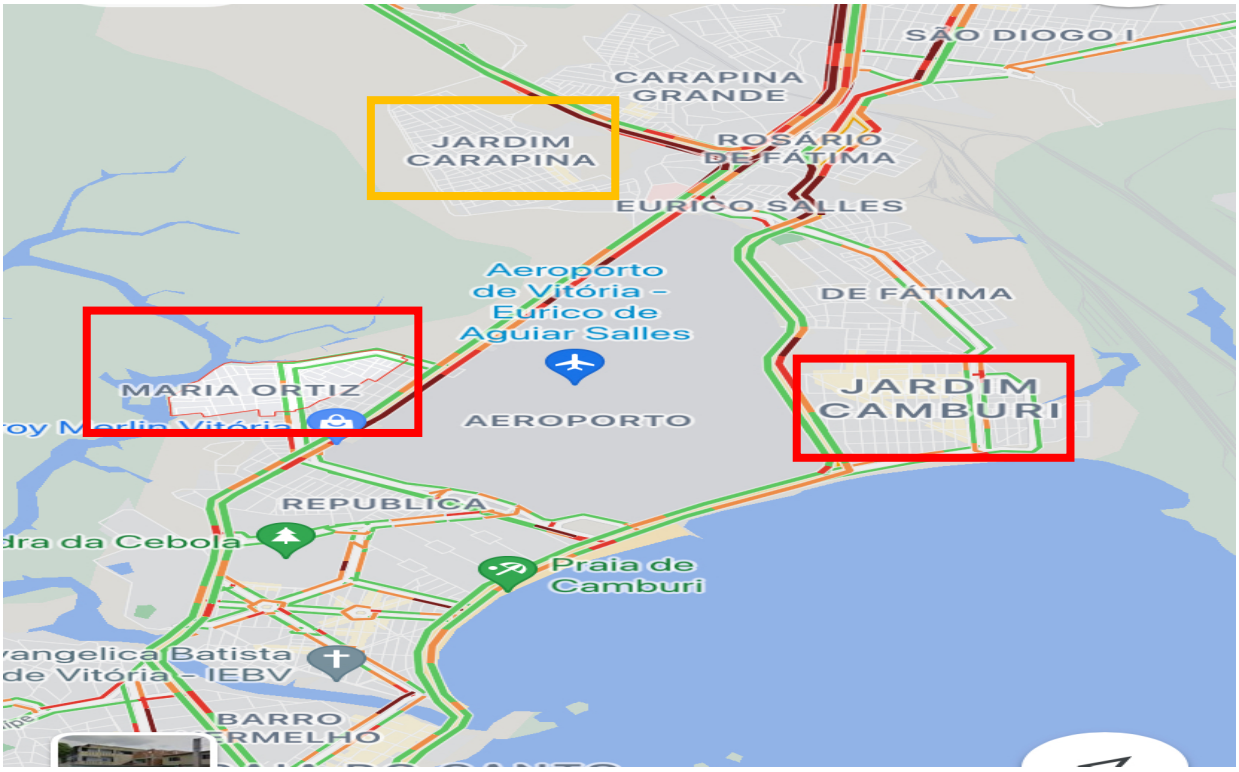


	counts	%
F	71840	0.649977
M	38687	0.350023

Weirdly, we can say that both proportion are the same, there is no difference between male and female in addition to the number of females more than the number of males but all the results are similar , Therefore, gender does not have an important role in a person to come to his appointment

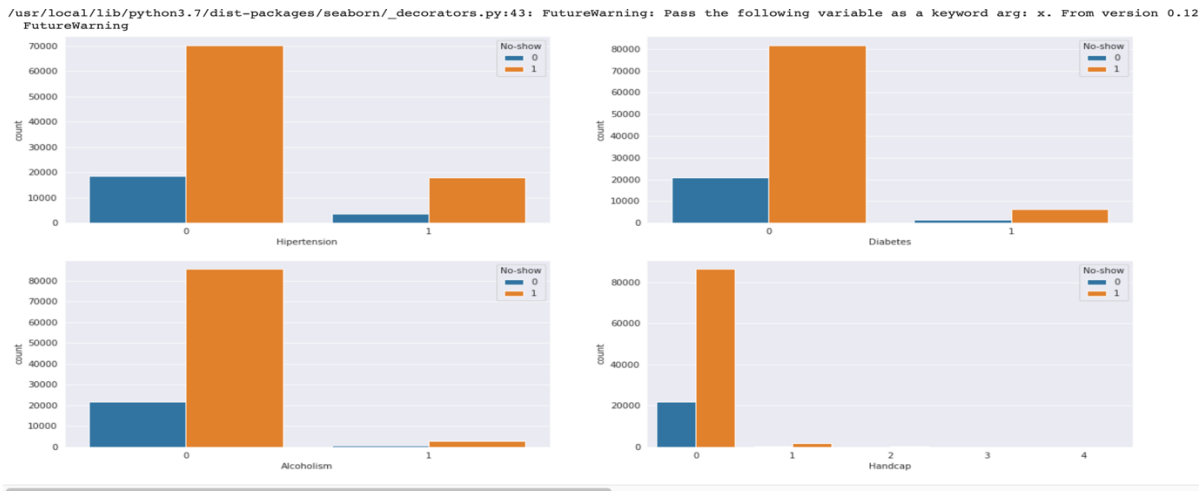
2. **does the location of the hospital Neighborhood would be reason for the patient to come to his appointment?**





The neighborhood is reason for coming the patient to his appointment. After looking at the neighborhood through Google Maps, it became clear to me that it is considered east of Brazil, where the international airport is located. The proximity of the neighborhoods to the airport increases the popularity of the hospital, and this leads to the popularity of the hospital, and the increases to attend the patient to his appointment.

3. **if the patient has diseases is that will cause the patient to show up to his appointment?**



We can see from the graphs that there is a large number who do not suffer from chronic diseases. However, a large percentage of those who suffer from the disease attend the appointment, so we conclude that the disease is related to attending the appointment.

Conclusions:

After downloading the data, we create three questions to start analyzing them. At first we cleaned the data from impurities and made sure that there was no duplicated row in addition to the normalization to make it easier to deal with, after that we started analyzing one question after another and we noticed, that a number of women's appointments is more than men, and with their large number, both proportion are the same and there is no difference between male and female, From the importance of the above feature, we can see that the neighborhood helps us in determining whether the patient who booked an appointment will show up on his appointment as well as in case he has a chronic disease

Limitation:

One of the limitations that I faced is that the size of the data is too large , in addition to that some variables need a lot of normalization in order to deal with them easily

Resources:

1 – google maps
2- https://www.kaggle.com/joniarroba/noshowappointments