

Project: - No-show medical appointments

Table of contents:

- Introduction
 - Data wrangling
 - EDA analysis
 - Conclusion
-

1. Introduction

This dataset collects information from 100k medical appointments in Brazil and is focused on the question of whether or not patients show up for their appointment. The main question we are trying to answer here is why 30% of patients miss their scheduled appointment. We are trying to predict the most important factors that affect the attendance of the patient.

****Some questions we can ask to help us explore the data: ****

- Does the patient gender have a relation with the attendance?
- Does the neighborhood play a role in making patients don't show up?
"Location of the hospital"
- Which patients show up more? Does old age take care of their health more than youth?
- Does the disease type affect the patient's show up?

Column definition (described from author) :

- PatientId - Identification of a patient
- AppointmentID - Identification of each appointment
- Gender = Male or Female
- ScheduledDay = The day of the actual appointment
- AppointmentDay = The day someone called or registered the appointment
- Age = Age of the patient
- Neighbourhood = Where the appointment takes place.
- Scholarship = If the person receives welfare assistance (True or False)
- Hipertension = True or False
- Diabetes = True or False
- Alcoholism = True or False
- Handicap = True or False
- SMS_received = 1 or more messages sent to the patient.
- No-show = True or False.

No Missing values or duplicated ones

```

df.duplicated().sum()

[17]
...
0

df.isnull().sum()

[18]
...
PatientId      0
AppointmentID  0
Gender          0
ScheduledDay    0
AppointmentDay  0
Age            0
Neighbourhood  0
Scholarship     0
Hipertension    0
Diabetes        0
Alcoholism      0
Handcap         0
SMS_received    0
No-show         0
dtype: int64

Great! No missing data or duplicated ones.

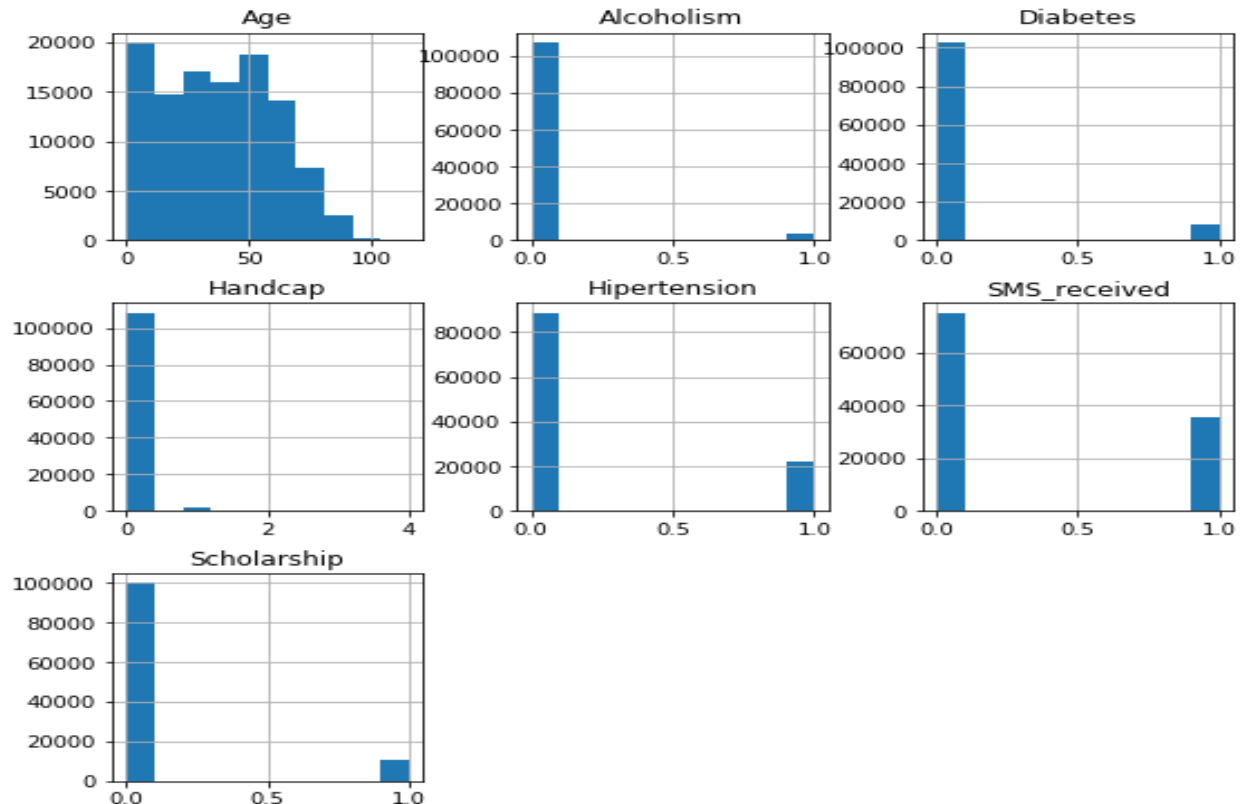
```

- We don't need patient id or appointment id for analysis, so we will drop them.
 - Age can't be negative so we will drop the -1
- ```
df = df[df['Age'] >= 0] # drop negative age
```
- Rename incorrect columns names

```
df = df.rename(columns={'Handcap': 'Handicap',
 'Hipertension': 'Hypertension'})
```

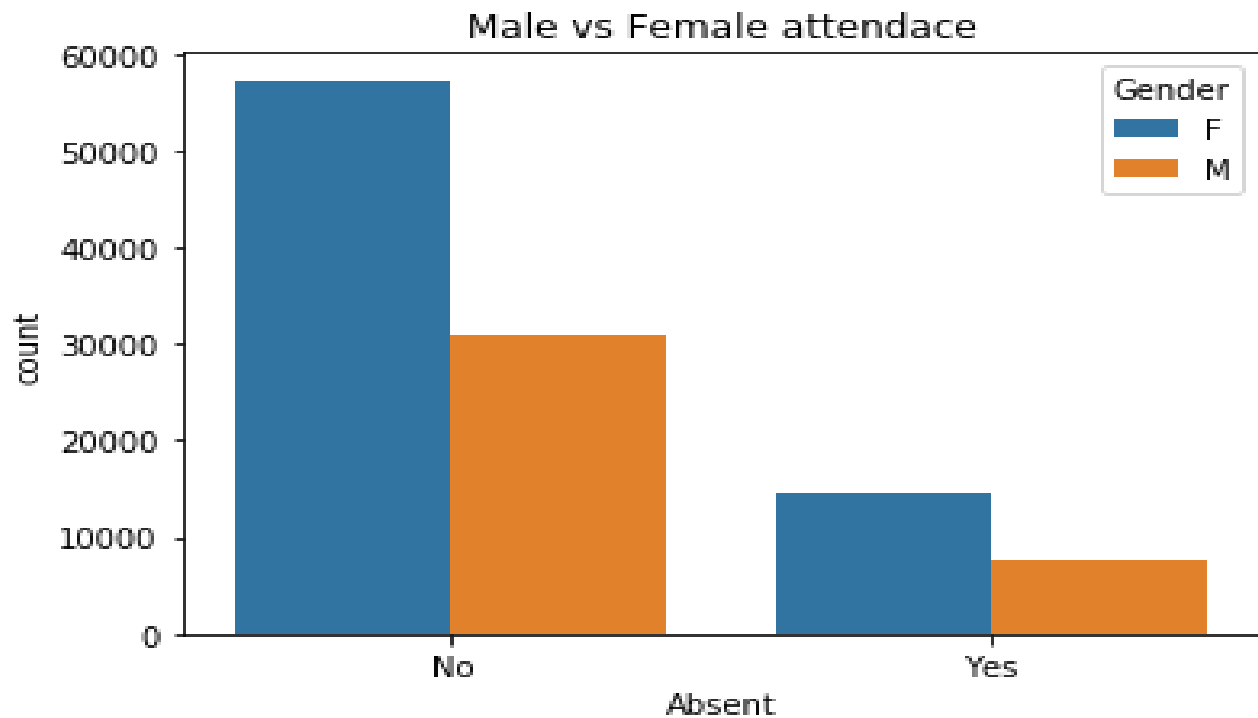
- rename the No-show column to avoid misleading

```
df = df.rename(columns={'No-show': 'Absent'})
```



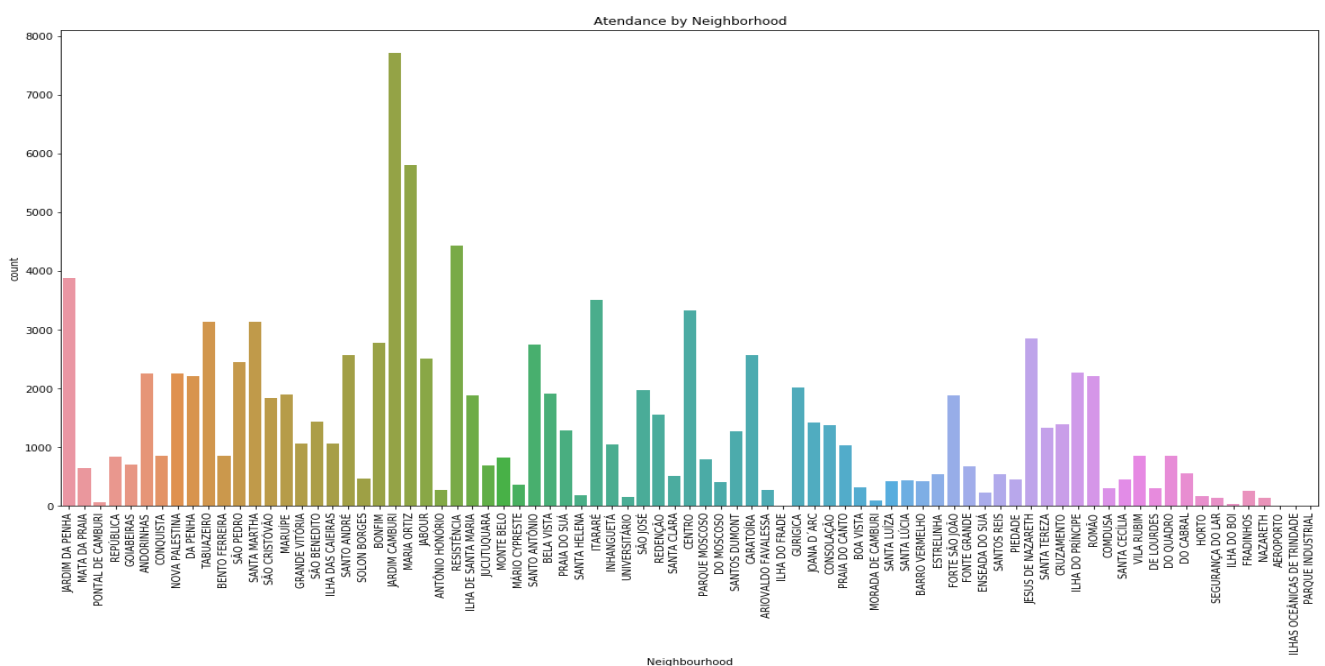
# Questions

1. Does the patient gender have a relation with the attendance?

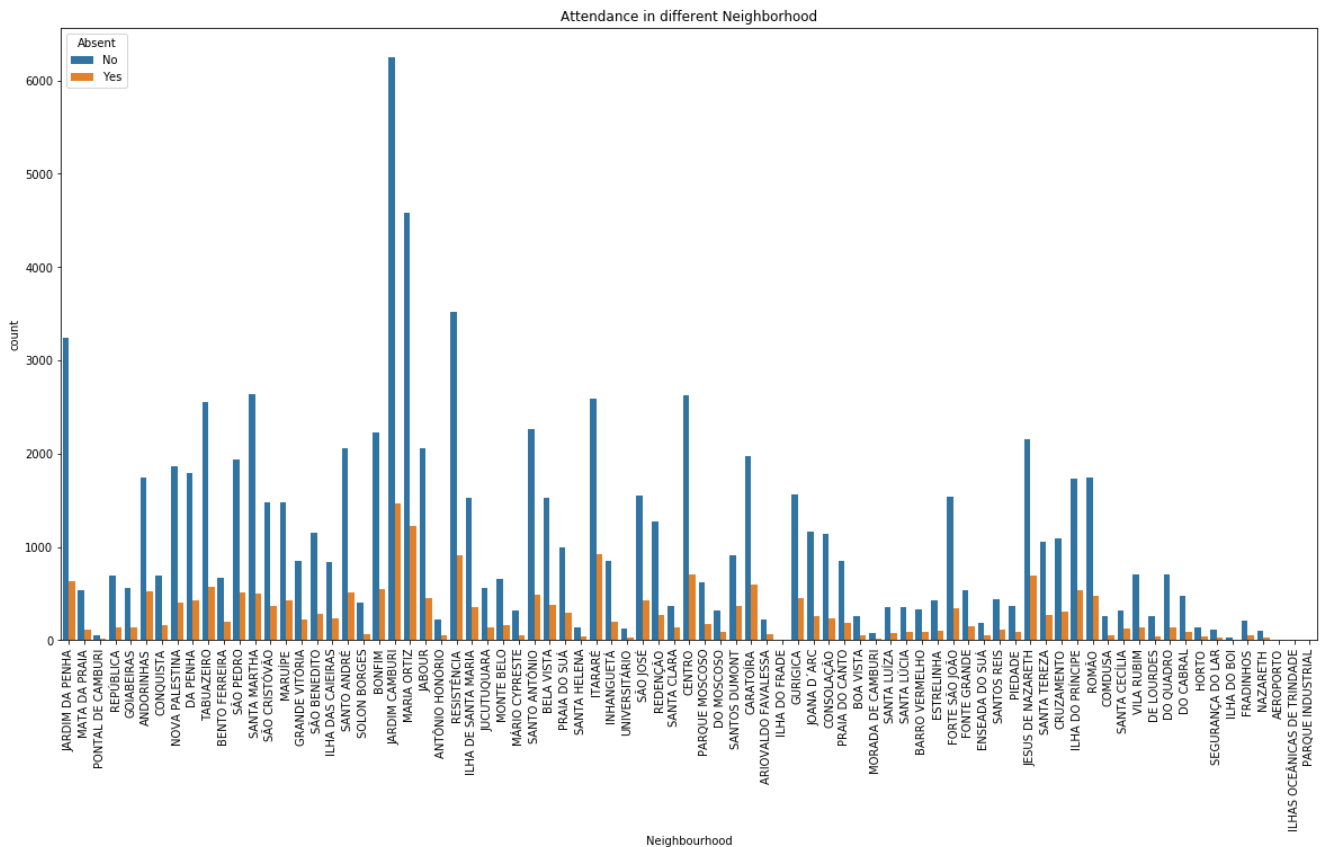


The number of females show up is greater than the males. May be because we have more data of females but that also show that they visit hospitals more in general.

2. Does the neighborhood play a role in making patients don't show up? "Location of the hospital"

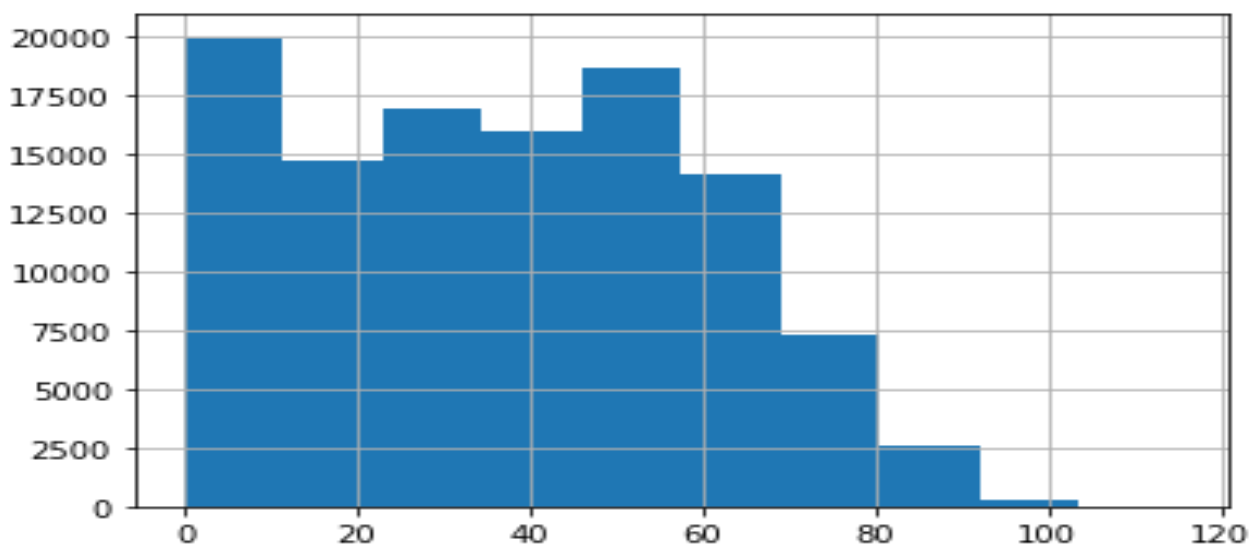


**We see that some neighborhood have more people show up for their appointment and this indicates that this area have increase in diseases**

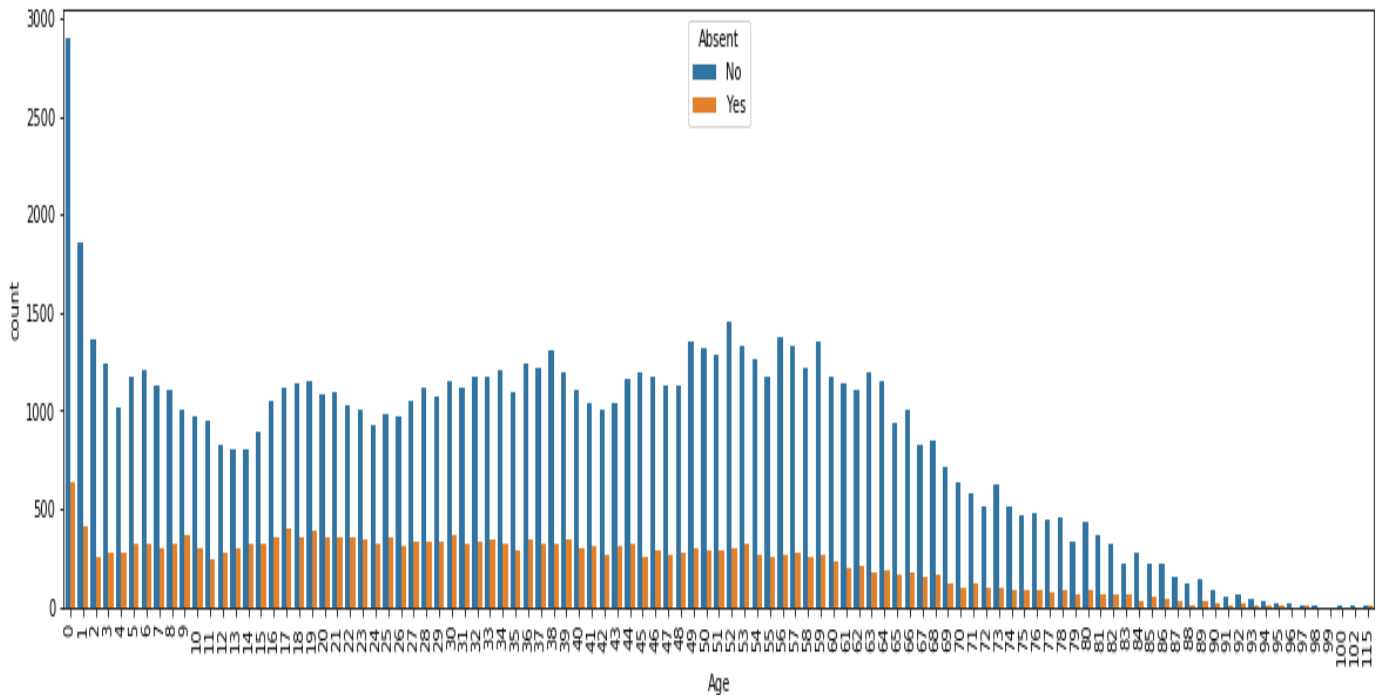


**In most neighborhoods patients attend more in the more the area where there are more disease**

**3. Which patients show up more? Does old age take care of their health more than youth?**

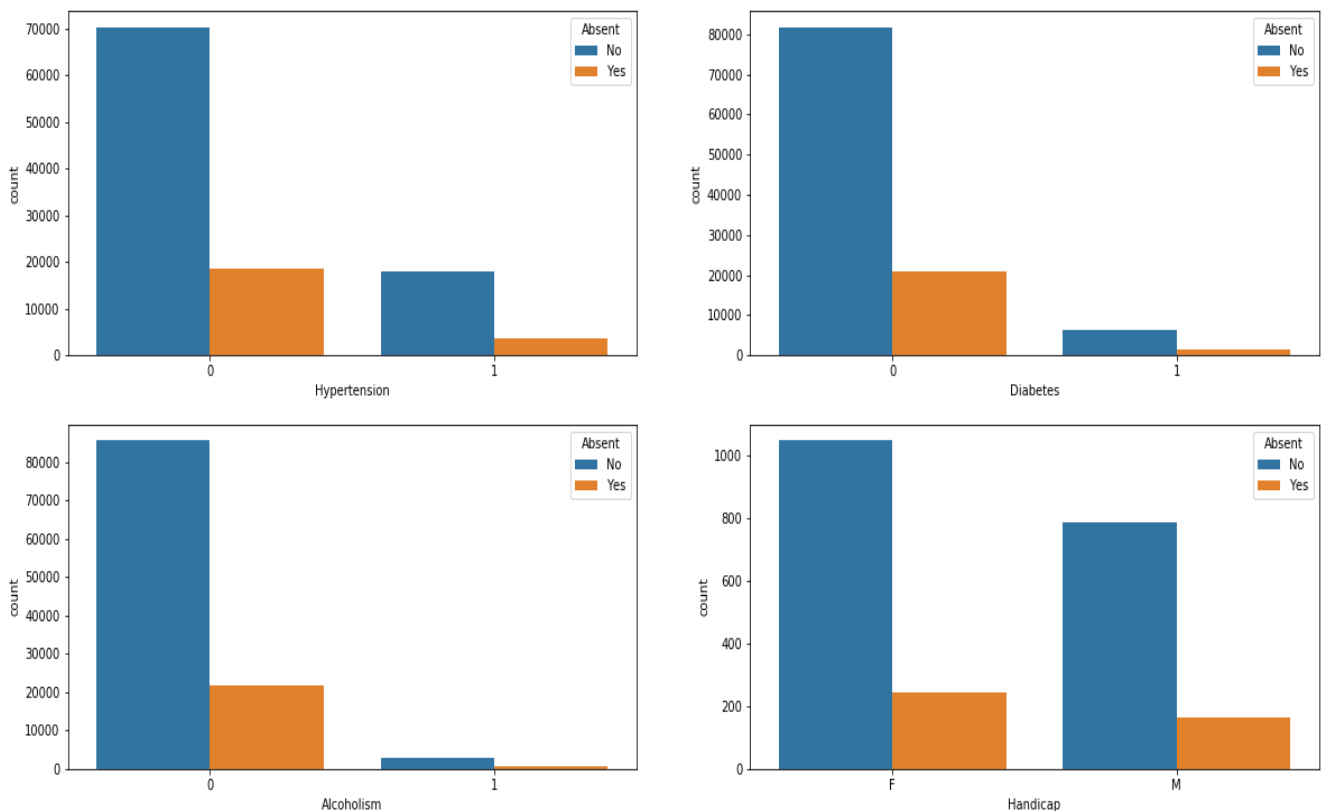


**We see that most of the patients in the data are youth**



This shows that the ratio are close but youth still show up more which the oppsite of what we argued at the beginning

#### 4. Does the disease type affect the patient's show up?



We see that most of them don't have a disease and show up for appointment but we notice that patients of hypertension show up either when they are infected or not which is a mark that hypertension will probably show up more.

## **Conclusions**

Now we can see the factors that affect the absence of the patients more clearly. The gender and age are the most important factor as we saw earlier that female and youth show up for their appointment more than male and old people. Neighborhood and hypertension come after gender and age as there are some neighborhoods that the diseases are spread and patients with hypertension tend to show up if they have it or not. So we need to search for more factors to help patient remember their appointments and show up.