

# **Analysis of “No-Show Appointment “**

**which dataset would be analyzed?**

The “no-show appointments” data set was analyzed

## **1- Introduction:**

This data set contains data for 110.527 medical appointments its 14 associated variables (PatientId, AppointmentID, Gender, Age, Neighbourhood, Scholarship, Hipertension, Diabetes,Alcoholism, Handcap, SMS\_received,No-show) in Brazil. No-show variable shows which the patient showed up for the appointment or not, so 'Yes' mean that the patient did not show for the appointment. moreover, for more information over the scholaship, please visit this website ([https://en.wikipedia.org/wiki/Bolsa\\_Fam%C3%ADlia](https://en.wikipedia.org/wiki/Bolsa_Fam%C3%ADlia))

## **2- the questions that I posed?**

1-which area most has no show

2-is the gender and age affect the no show

3-how long between schedule day and appointment day and does it affect the result

4-does people who does not receive scholarship show more 'no show'

5-does receiving SMS affects 'no show'

6-are these features (Hypertension, Diabetes, Alcoholism, Handcap) affect the 'no show' result?

## 3- Data Wrangling

### 3-1-Assessing the data

<code>df.shape</code>	<code>#how many rows and columns</code>
<code>df.info()</code>	<code>#features and data types</code>
<code>sum(df.duplicated())</code>	<code>#number of deplicated rows</code>
<code>df.nunique()</code>	<code>#uniques values</code>
<code>df.isnull().sum()</code>	<code>#is there nodata values</code>

### 3-2- Preparing the data

1-#rename the column to lowercase

```
df.rename(columns=lambda x: x.strip().lower().replace("-", "_"), inplace=True)
```

2- #change the time from object to time

```
df['scheduledday'] = pd.to_datetime(df['scheduledday'])
```

```
df['appointmentday'] = pd.to_datetime(df['appointmentday'])
```

3- #extracting no show data only

```
df_yes=df.query("no_show== 'Yes'")
```

4- # difference between schedul time and appointment

```
df['difference_days']= df['appointmentday'].dt.date-df['scheduledday'].dt.date
```

5- #check some rows which as differences more than 30 days

```
df.iloc[110518]
```

6- # Creating a Function to remove columns

```
def remove_columns(columns):
```

```
    return df.drop(columns=columns, inplace=True)
```

```
remove_columns('appointmentid')
```

## Exploring with visuals

### 1- #number of nonshow

```
df.groupby('no_show')['patientid'].count().plot(kind='bar')
```

### 2- #number of male and female patients

```
df.groupby('gender')['patientid'].count().plot(kind='bar')
```

### 3- #exploring the range of ages

```
df['age'].plot(kind='hist')
```

### 4- #exploring the differences between scheduled date and appointment date

```
df.groupby('patientid')['difference_days'].count().plot(kind='hist')
```

```
plt.ylabel('Number of non-show')
```

```
plt.xlabel('Duration between schedul day and appointment day in days')
```

```
plt.title('Relation between duration between schedul day and appointment day in days and non-show');
```

### 5- #Exploring the patient ages and handicap degree

```
plt.scatter(df['handcap'], df['age'])
```

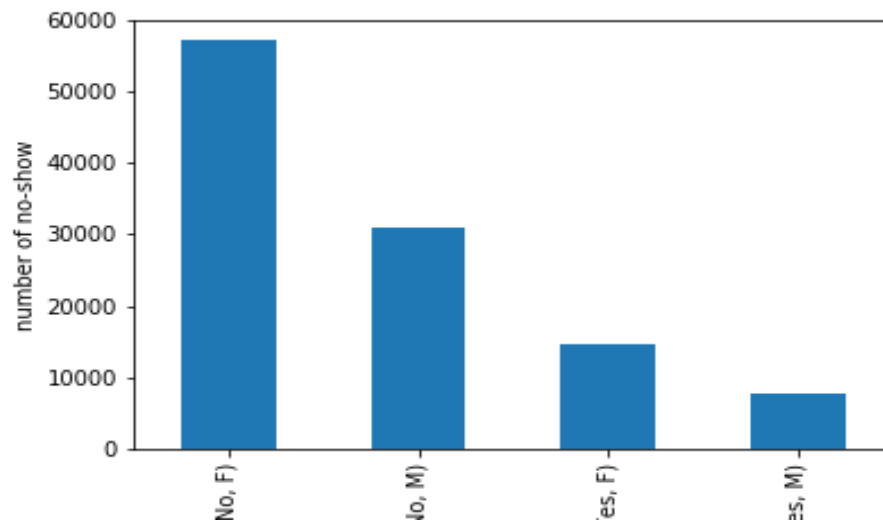
```
plt.xlabel('Handcap Degree')
```

```
plt.ylabel('Age in years')
```

```
plt.title('Exploring the patient ages and handicap degree')
```

# Summary statistics and plots communicating final results

1-#relation between gender and statues (show or no show):



Females present more “no-show” than males

no\_show gender

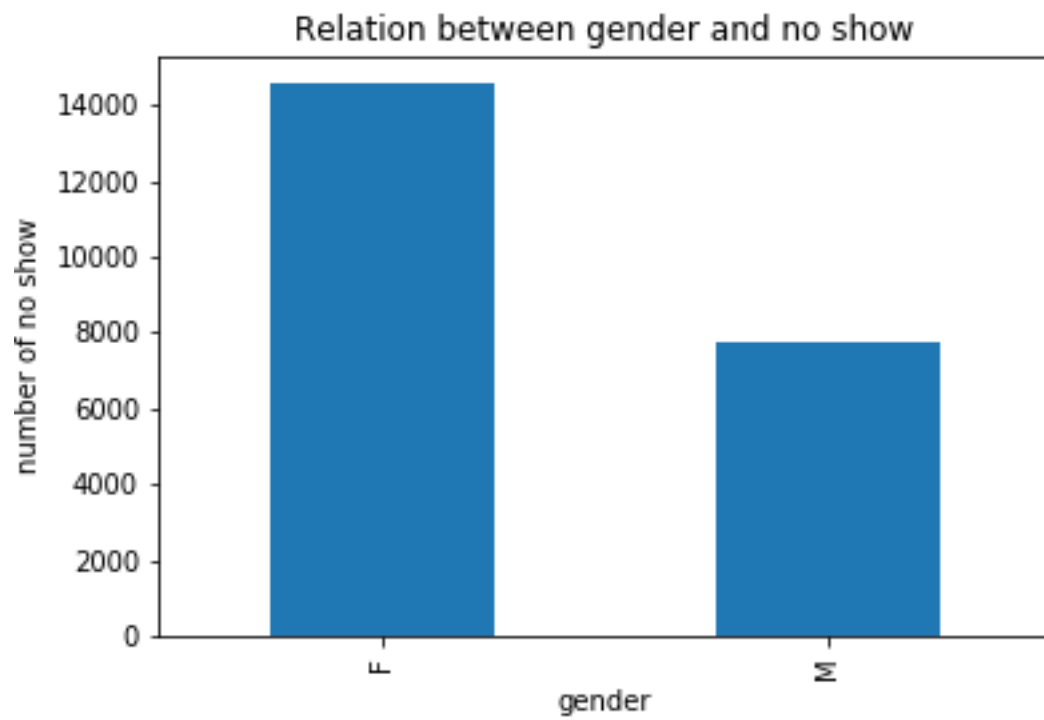
No F 57246

M 30962

Yes F 14594

M 7725

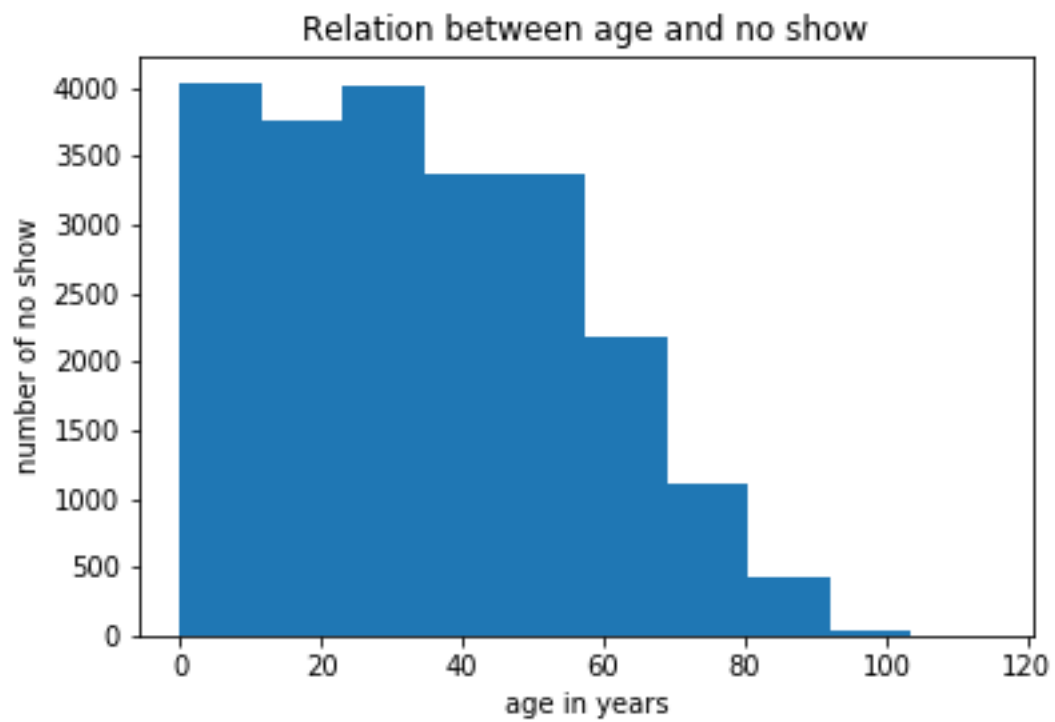
## 2-#relation between gender and no show



Females present more “no-show” than males

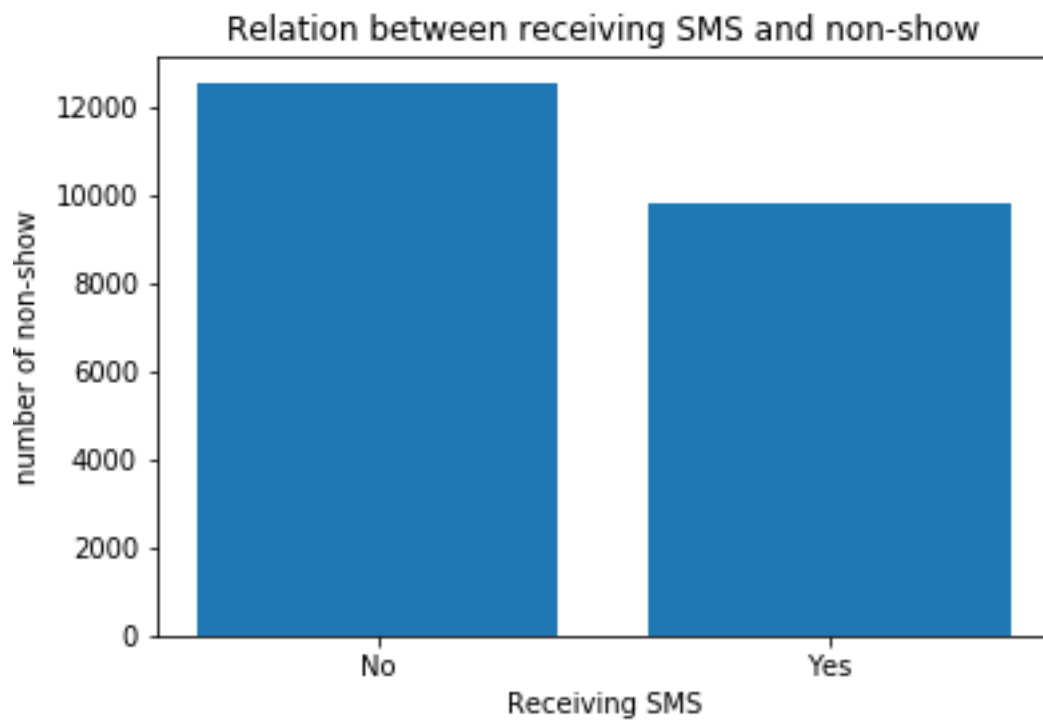
no_show	gender	
Yes	F	14594
	M	7725

### 3- relation between age and no show



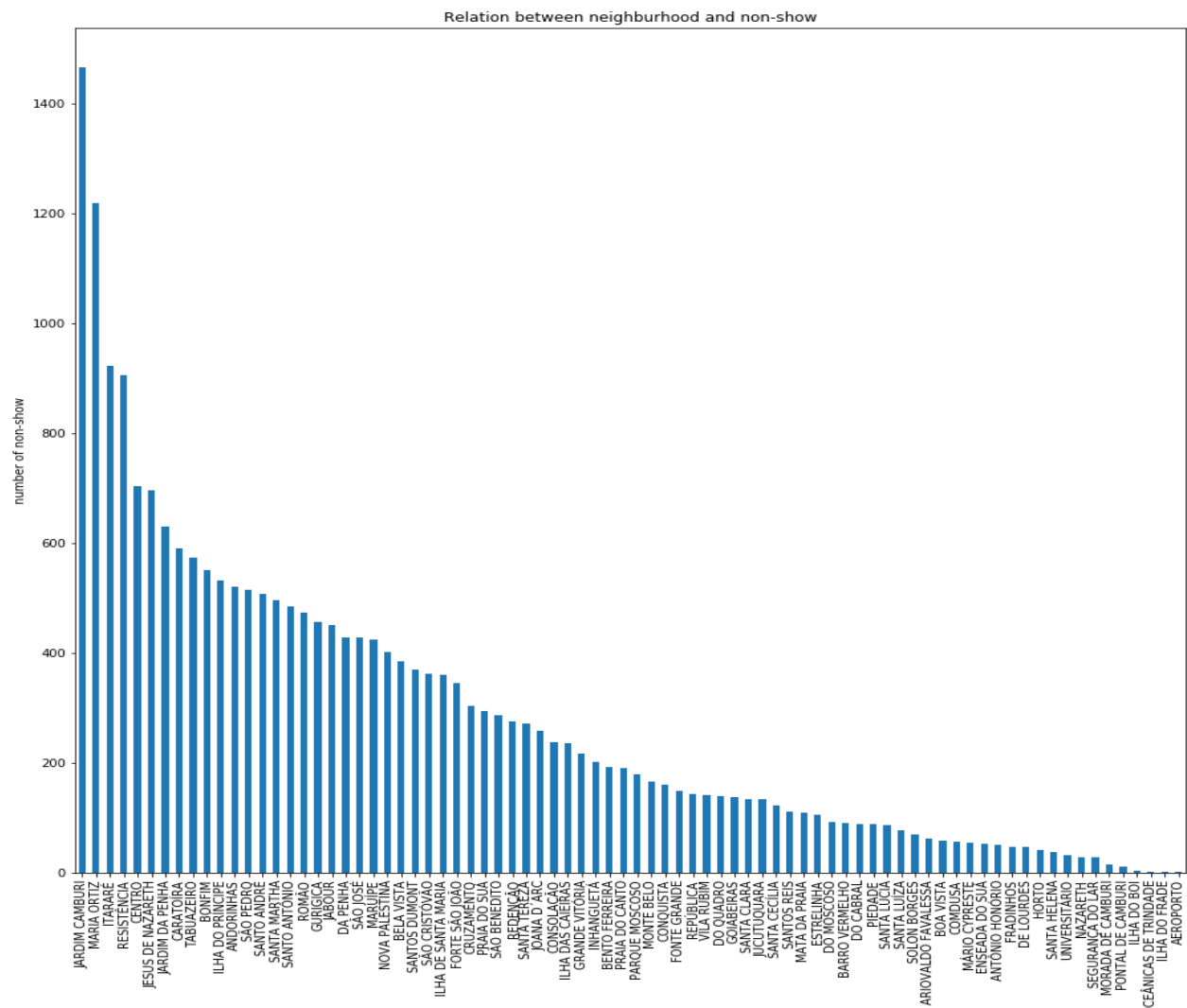
People from 0 to 30 years old show more “no-show” than older people.

#### 4- #relation between receving sms with the data and no show



People who received SMS show more than people who do not receive SMS

## 5- #which area receive much no show

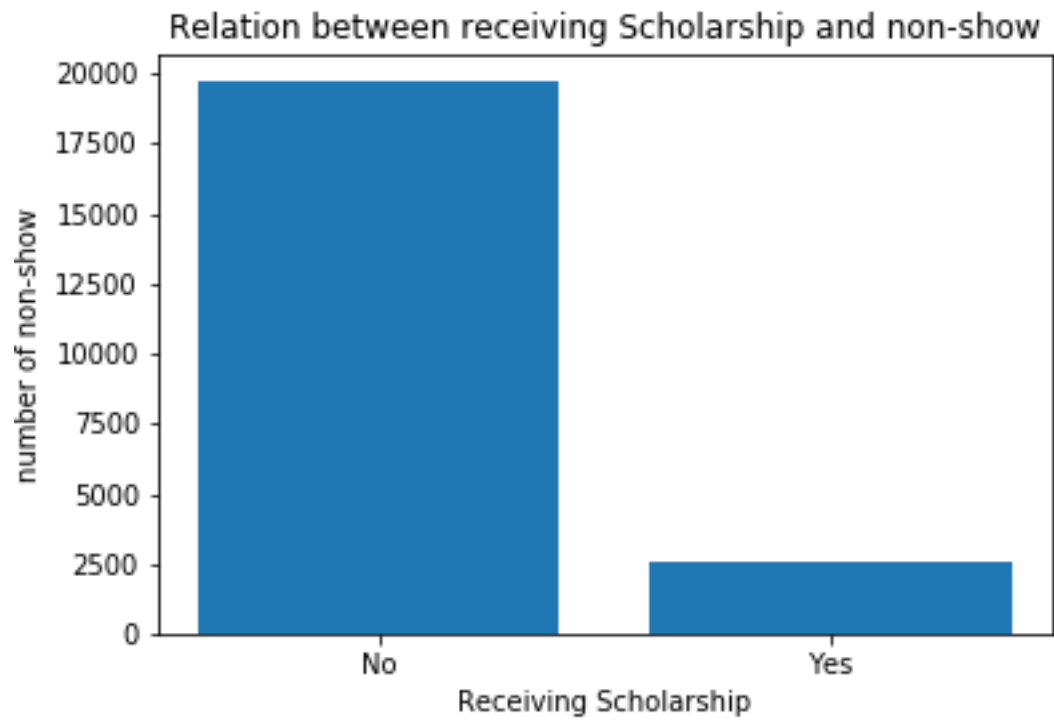


neighbourhood	
JARDIM CAMBURI	1465
MARIA ORTIZ	1219
ITARARÉ	923
RESISTÊNCIA	906
CENTRO	703
JESUS DE NAZARETH	696
JARDIM DA PENHA	631
CARATOÍRA	591

People from these neighborhoods show “no show” more than other neighborhoods

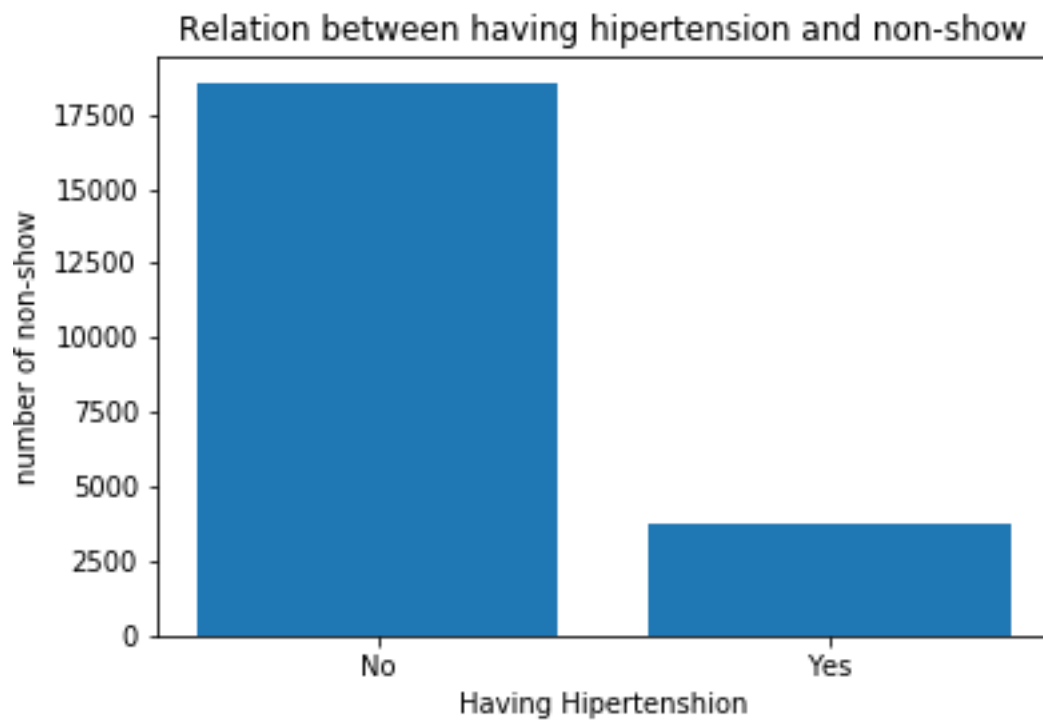


## #6- #relation between scholarship and no show



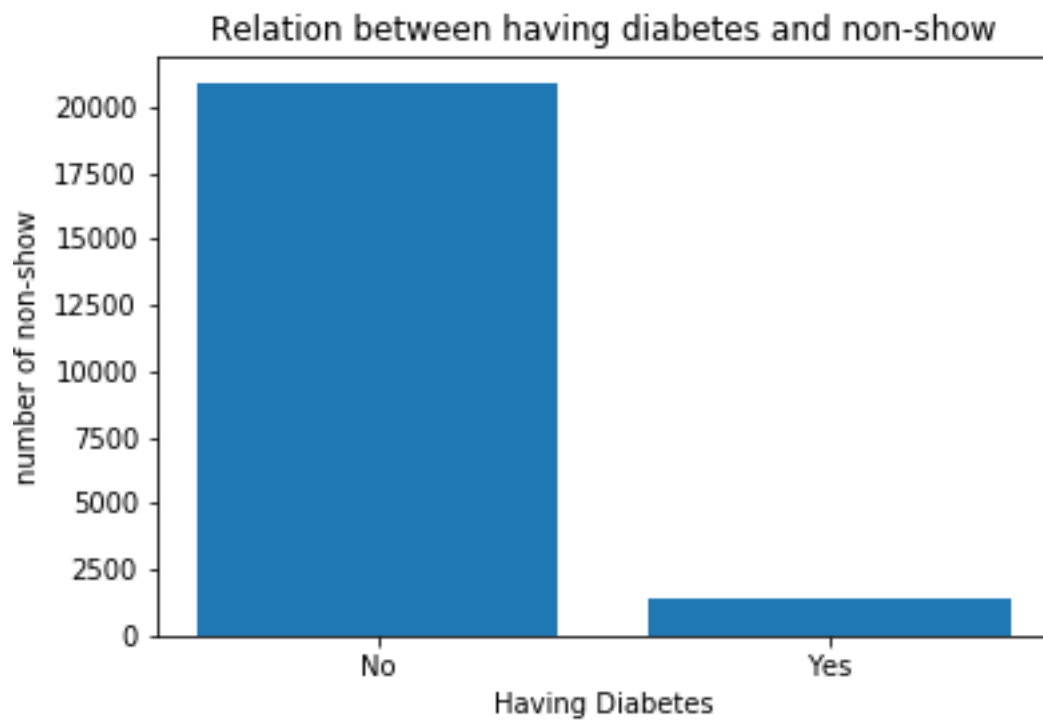
People who do not receive scholarship show more “no show” than people who receive scholarship

### 7- #relation between hypertension and no show



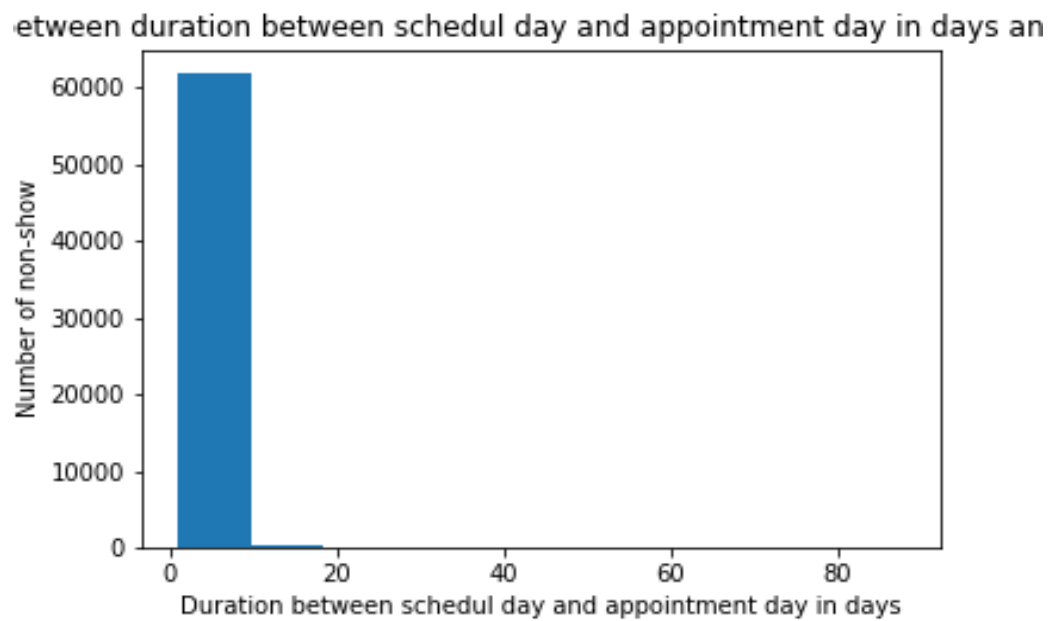
People who have hypertension show less “no show” than people who have it

## 8- #relation between diabetes and no show



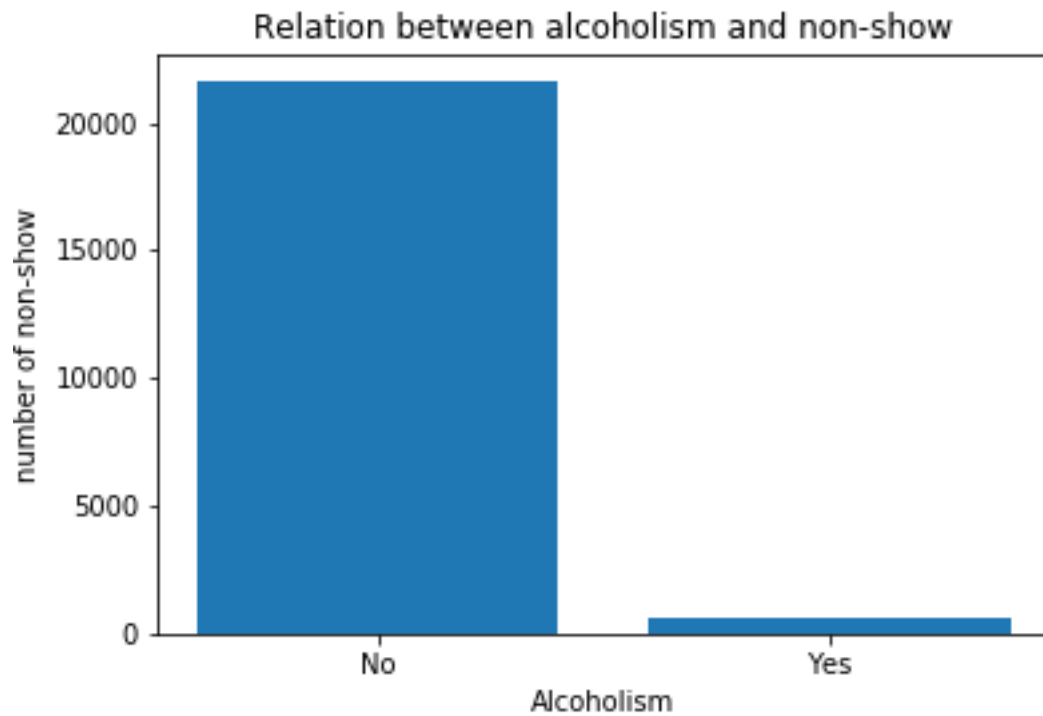
People who have diabetes show less “no show” than people who have it

## Relation between the differences between scheduled date and appointment date and no show:



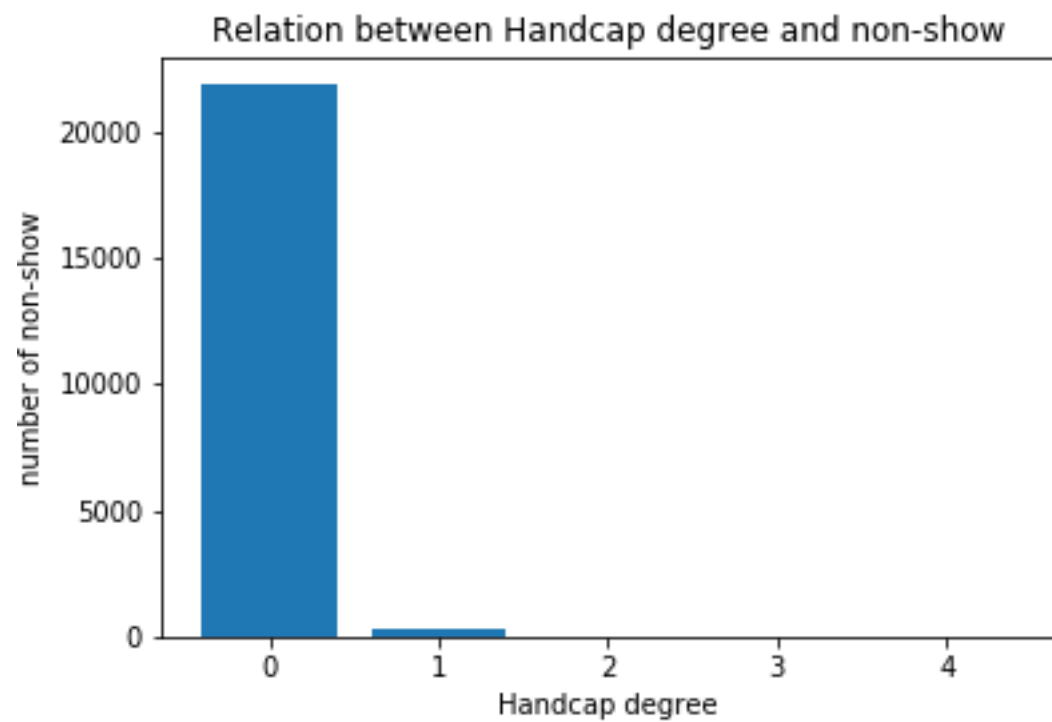
Most of “no -show” appointments were in the duration between 0 and 10 days

### 9- #relation between alcoholism and no show



People who are alcoholism show less “no show” than people who have it

### 10- #relation between handicap and no show



People who are not hand cap show “no show” to appointments more than people who are hand cap

## Conclusions

- Number of males that did not show up for the appointments is less than the female's number
- People with the age range from 0 to 40 years showed higher no show for their appointments regards order people.
- There is a big difference between people who does not receive scholarship and did not show up and people who receive scholarship and did not show up.
- people who did not receive SMS before their appointments showed higher no show than people who received SMS and did not show up.
- Some areas such as (JARDIM CAMBURI, MARIA ORTIZ) significantly showed higher no show than other areas.
- People with diseases such as (have diabetes, hypertension, hand cap, alcoholism) showed less no-show regarding health people.
- The difference between the schedule date and the appointment date does not have significant impact on no show data

## Limitations

There is a couple of limitations that obstacle our analysis:

- Most of the variables are categorical data which limit the ability to analysis them.
- Conclusions based on relations between different variables cannot be drawn as the most of variables were categorical data.