Project 2. No Show Appointments

Load the Csv file 'KaggleV2-May-2016.csv' and create

import pandas as pd

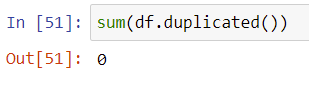
import matplotlib.pyplot as plt

import numpy as np

df = pd.read\_csv('KaggleV2-May-2016.csv')

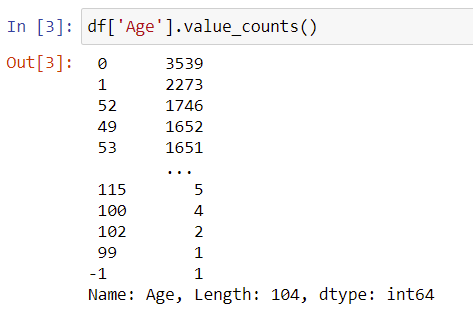
**1.Cleaning phase the data set**

**-** There is no duplicates in the dataset

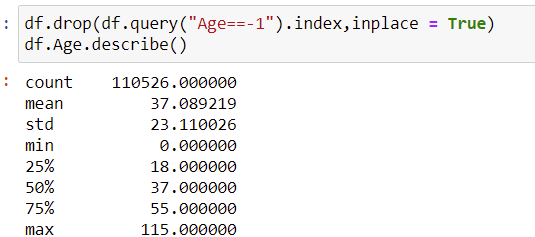


Identify each group age by the following syntax:

df[‘Age’].value\_counts()

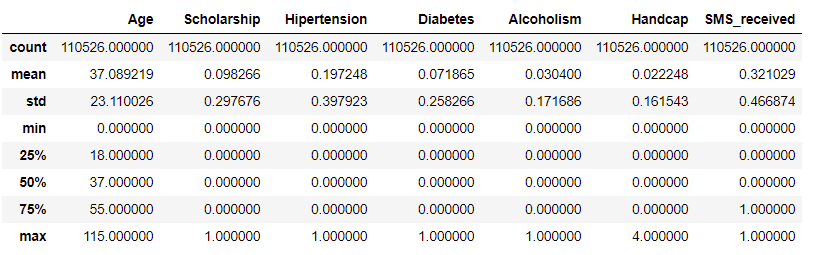


There is 1 group age of – 1, which is an outlier that we need to remove it from the data set



Removing the row that has the value = -1 in the column ‘Age’

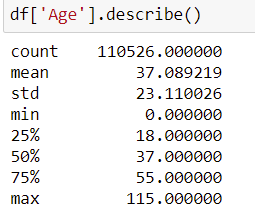
after cleaning the data, we have the following data set:



**2. Analysis**

Proposing the questions:

**2.1 Identify the group age.**



- The average age of the sample was 37 years old, in which the minimum age is 0 (less than 1 years old) and the highest was 115 years old.

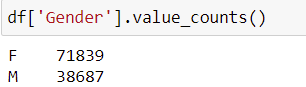
- 50% of the samples aged from 0 to 37 years old.

- 75% of the samples age from 0 to 55 years old.

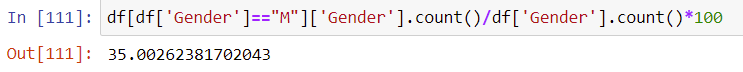
for showing up for the appointments.

The gender structure of the dataset

**2.2 Identify the gender**

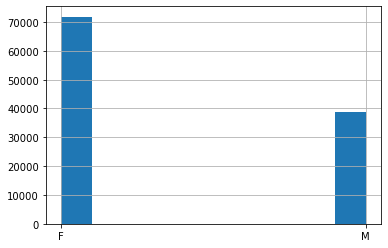






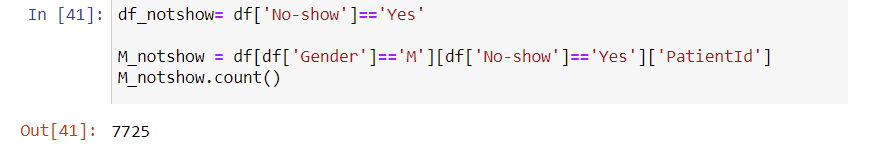
The female number was 71,839 which represent 65% the sample.

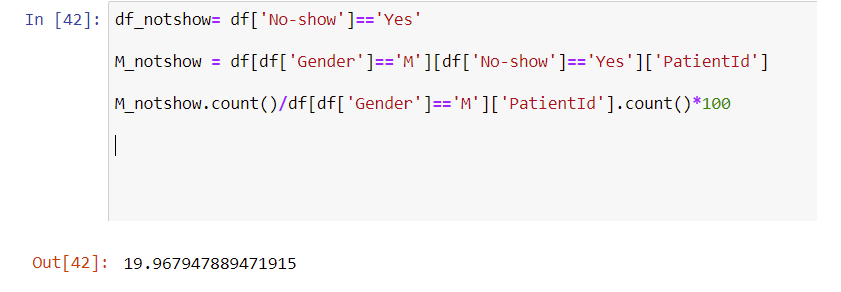
The male number was 38,687 which represent 35% the sample.



**2.3 Finding the relationship between the gender and the No-show results**

**2.3.1 Male number that not showing at the appointment**

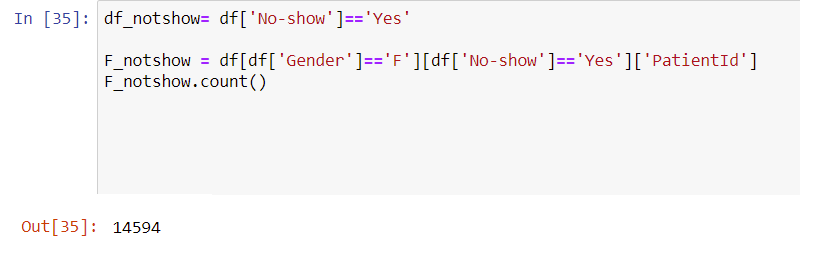
****

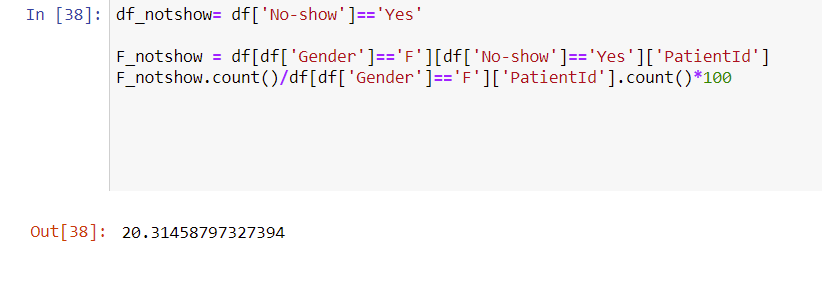
****

There was 7,725 male out of 38,687 that not appearing in the appointments. The percentage was 19.97%

**2.3.2 Female number that not showing at the appointment**

Number of female that didn’t show up:

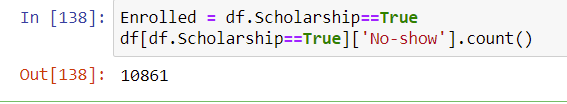


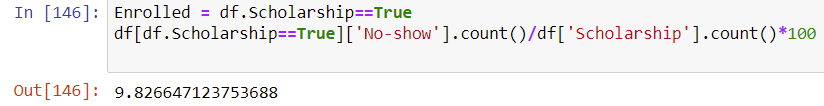


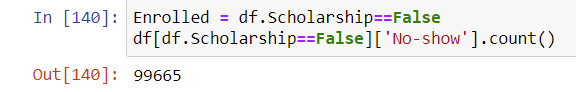
There was 14,594 female out of 71,839 that not appearing in the appointments. The percentage was 20.3%

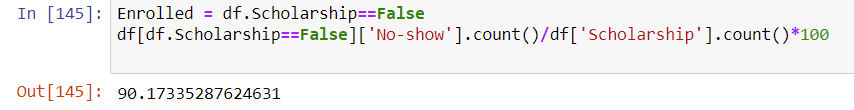
**Conclusion:** a little bit higher of male rate that not showing at the appointments compared with the female. Therefore, Gender is not the factor that affecting the results of the appoinments.

**2.3 Identify people involving the Brasilian wellfare**

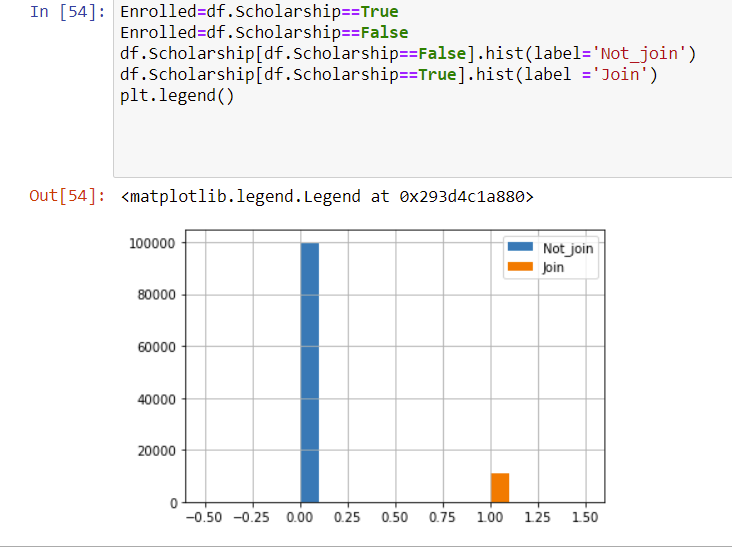






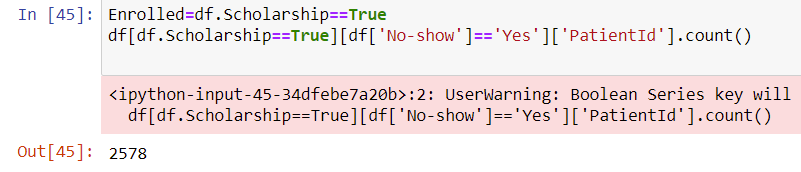


There are 10,861 (9.8%) people joining in the Brasilian and 99,665 (90%) not joining the program.

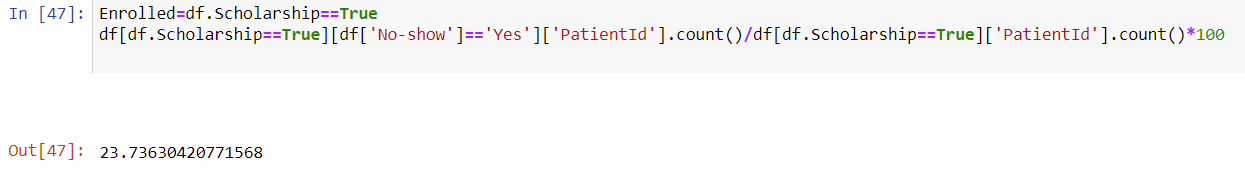


**2.4 Relationship between the wellfare program with the No-show results**

**2.4.1 Number of people joining the wellfare program that didn’t show up at the appointments**

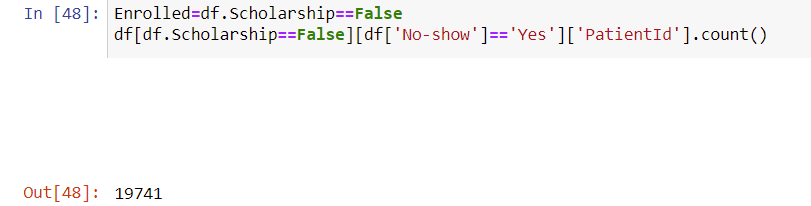


2.4.2 **The percentage of people joining the wellfare program that didn’t show up at the appointments**

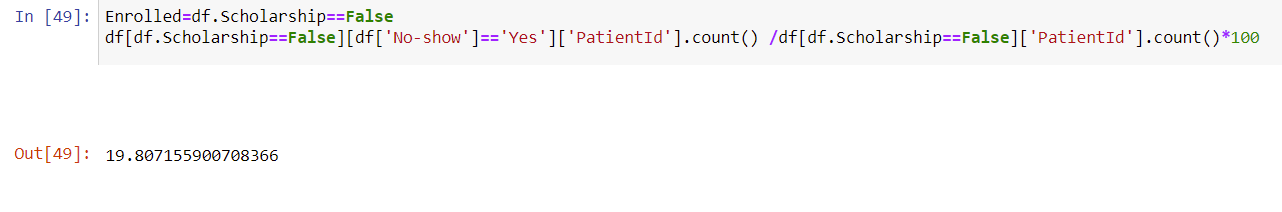


There are 2,578 people **joining the program** didn’t show up at the appointmes. The percentage was 23.74%

**2.4.3 Number of people not joining the wellfare program that didn’t show up at the appointments**



**2.4.4 The percentage of people not joining the wellfare program that didn’t show up at the appointments**



There are 19,741 out of 99,665 (results in 2.3) not joining the program didn’t show up at the appointmes. The percentage was 19.8 %