No Show Appointment Project

I worked on “No Show Appointment” dataset -

Describing the features of the data

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. A number of characteristics about the patient are included in each row

questions that posed

Research Questions

1 - what is the overall appointment show-up vs. no show-up rate

2-what is the most frequent age?

3-What is the age group most affected by diabetes?

4 - what are the most feature that matter the most of ('Age', 'being alcoholic','Having an SMS', 'Gender', 'Scholarship'), to make the patient make it to his appiontment?

A description of what is did to investigate those questions

1- Looking for Data and detect the data that needs to be cleaned,

2-Fixing typos in data columns

3-Then starting Data Cleaning

Data Cleaning steps

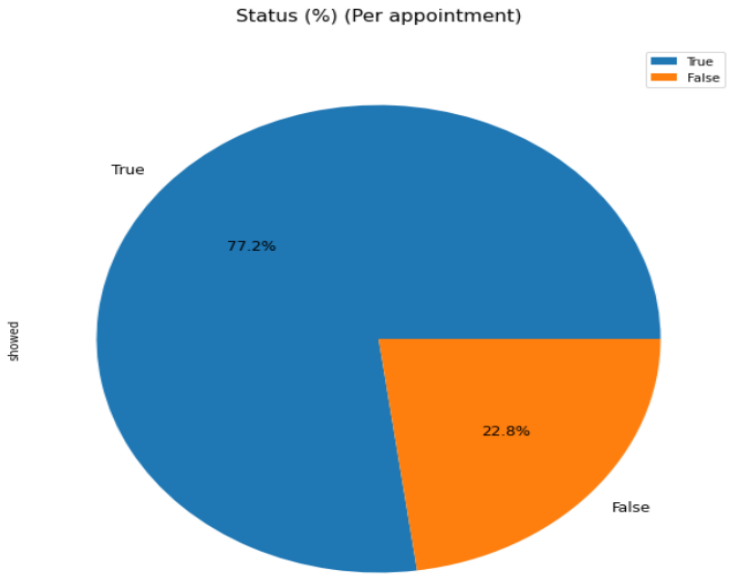
- *converting some columns that has date to a datetime datatype*
- *fixing the ages that are less then 0 by taking the mean of all ages and put those values into it*
- convert no show data to 0 and 1 instead of 'yes' and 'no' so we can handle using plot or goupby

Exploratory Data Analysis

Step on by answering Question 1

(what is the overall appointment show-up vs. no show-up rate?)

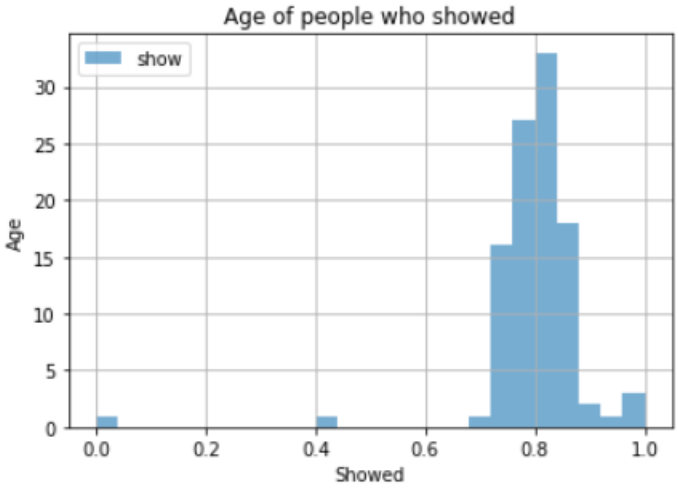
- *making a pie chart to have the percentage of who showed up and who is not*



Step on by answering Question 2

(what are the most feature that matter the most of ('Age', 'being alcoholic','Having an SMS', 'Gender', 'Scholarship'), to make the patient make it to his appiontment?)

- print the average age of people who show up and people who didn't
- making an histogram to see the realtion between the (age,alcoholic people, SMS recievers, Gender, Scholarship) and the people who came to thier appointment



- using group by function to find relations between features
- and then see the relation between alcoholic people and people who showed up and who didn't

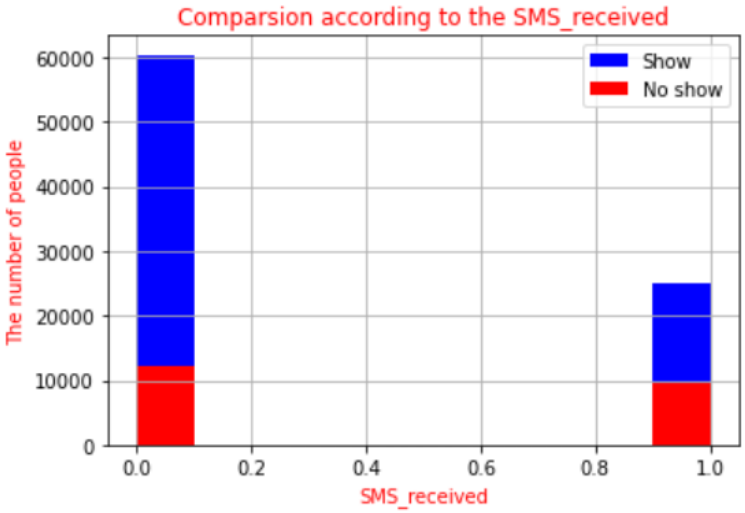
- # Summary statistics and plots communicating your final results

As we can see Age is the most factor that affect people who will attend or not!

Conclusions

• As we can see sending an SMS for the appiontment is not neccessary the right option to make sure that the patient will come

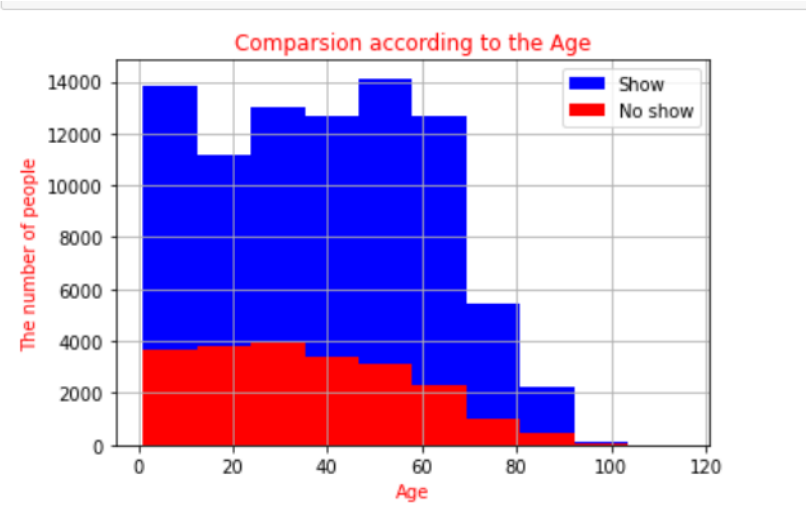
As we see



• about 22.8% of people that schedule an appointment did not make it to thier appointment

• our investigation the Age is the most important factor that decided if a patient would come or not the average of age for people who will be most likely to show up is 39.07518726482 ,and the average age for people who are not likely to show up is 35.329151291512915.

As we see



•most of people who has Scholarship are most likely to miss thier appointments with a percentage of 76.2% of showing and patients who don't have a scholarship have the percentage 80.1%

• the features such as different gender or alcoholic is not a factor to decide if the person would come to his appointment or not!

Limitations:

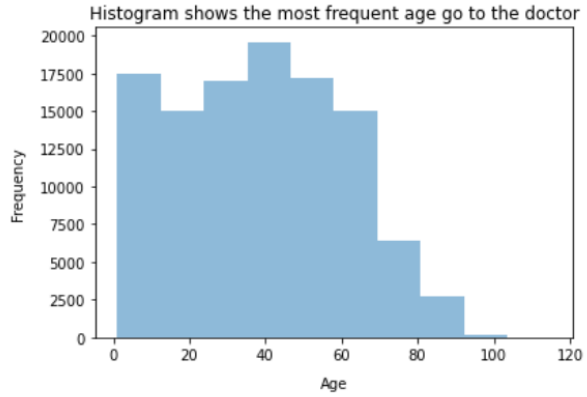
There are several factors that are missing that could be valuable in determining what is the most important feature that influences turning up for an appointment, such as whether the patient is employed or not, or whether the patient has a series of medical issues. There were some nonsensical facts, such as patients who were under the age of one.'

a) The most frequent age go to the doctor is 1 year

see the chart below for more understanding

```
the most frequent age is :  37.08887421173107
```

```
Text(0.5, 1.0, 'Histogram shows the most frequent age go to the doctor')
```



b) Also the age group most affected by diabetes is

between 95:99

see the picture below for more understanding

```
Diabetes   Age
0.000000   1.000000     2272
           2.000000     1617
           3.000000     1513
           4.000000     1298
           5.000000     1489
                        ...
1.000000   96.000000       2
           97.000000       1
           98.000000       2
           99.000000       1
37.088874  37.088874    3540
Name: Diabetes, Length: 194, dtype: int64
```

Reference:

- https://www.coursera.org/articles/data-analytics-books-for-beginners
- https://www.investopedia.com/terms/d/data-analytics.asp
- https://github.com/RahmaElbana/python_task.git
- https://github.com/RahmaElbana/python_project/blob/0d570ec1d970bb83eeb5f05af62634a9f9417013/Untitled3-checkpoint.ipynb