

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

Question 1

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

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 appointment?

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

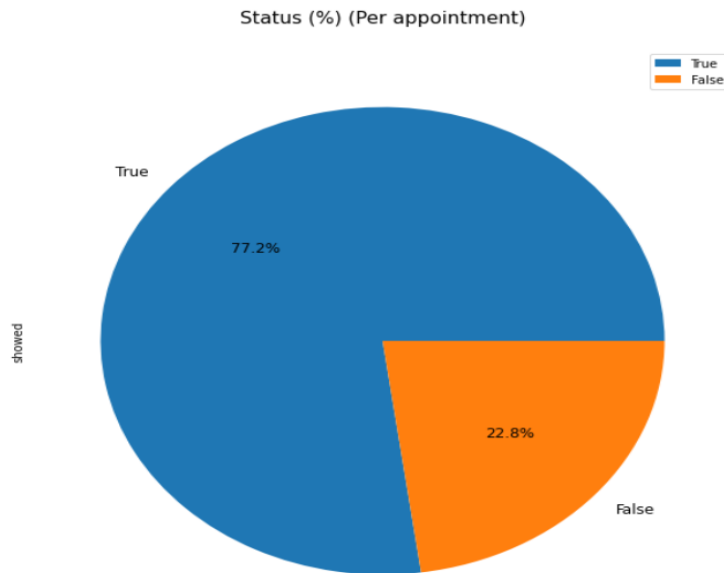
- *converting some columns that has date to a datetime datatype*
- *fixing the ages that are less then 0 by taking the mean of all ages and put those values into it*
- convert no show data to 0 and 1 instead of 'yes' and 'no' so we can handle using plot or goupby

Exploratory Data Analysis

Step on by answering Question 1

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

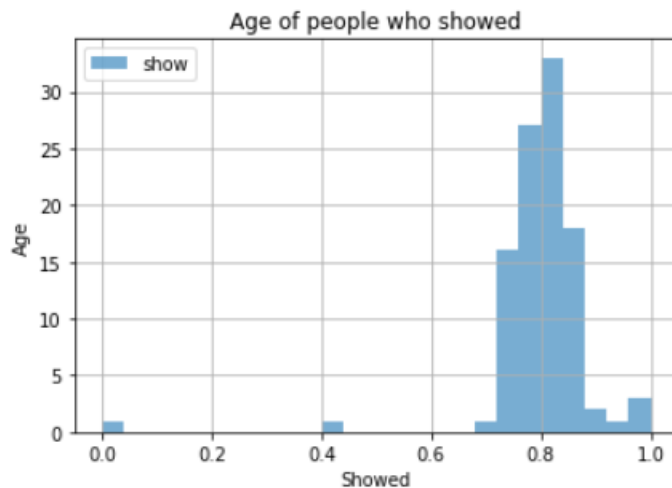
- *making a pie chart to have the percentage of who showed up and who is not*



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 relation between the (age,alcoholic people, SMS recievers, Gender, Scholarship) and the people who came to thier appointment



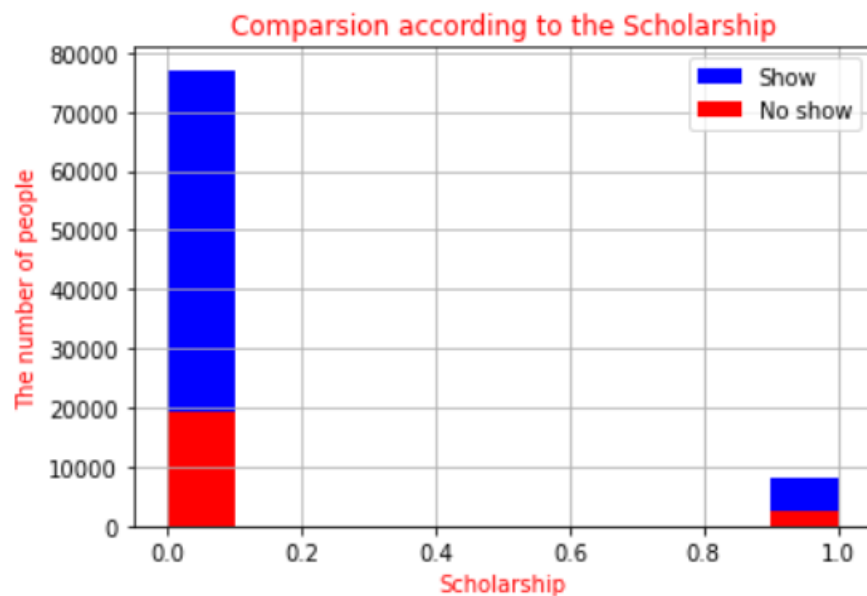
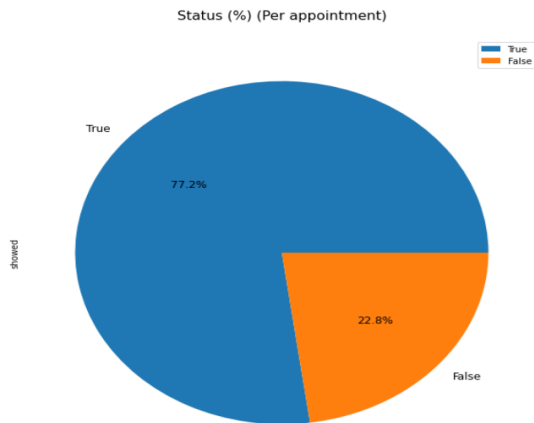
- using group by function to find relations between features
- and then see the relation between alcoholic people and people who showed up and who didn't

• 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
- 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.
- 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!



A list of Web sites, books, forums, blog posts, github repositories, etc. that you referred to or used in creating My submission

- <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