Investigate [No-show medical appointments] Dataset

Dataset Description

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.

● ‘ScheduledDay’ tells us on what day the patient set up their appointment.

● ‘Neighborhood’ indicates the location of the hospital.

● ‘Scholarship’ indicates whether or not the patient is enrolled in Brazilian welfare program Bolsa Família.

● The last column: it says ‘No’ if the patient showed up to their appointment, and ‘Yes’ if they did not show up.

Some questions we can ask for investigation this data set:

Q1: Did Age effect on attendance?!

Q2: Is there relation between age and Diseases?!

Q3: Is there relation between Neighborhood and Attendance?!

Q4: Did received SMS effect on attendance?!

Data Wrangling

In this section loading in the data, check for cleanliness, and then trim and clean your dataset for analysis.

Find that:

- No null values.
- The spelling of some names of columns is incorrect.
- Mean age is 37, max age is 115 and min age is -1 (must select all rows that have -1 age and clear them)***.***
- Some duplicate rows that have the ID and show status

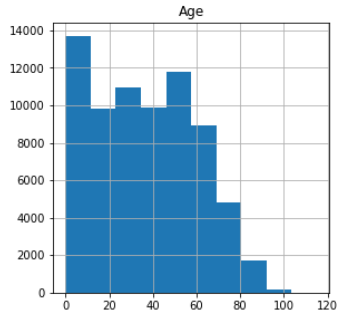
Data Cleaning

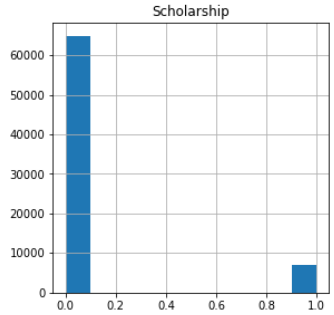
- Rename columns' names to correct names
- Delete row that have -1 age!
- Delete all duplicate rows that have the ID and show status

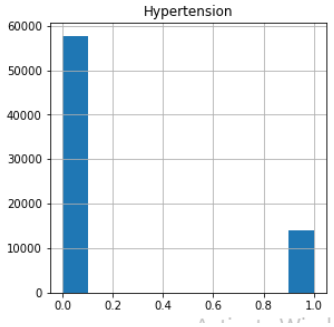
Exploratory Data Analysis

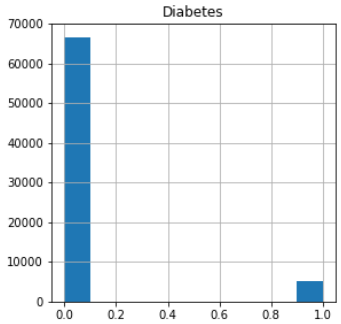
Compute statistics and **create visualizations**

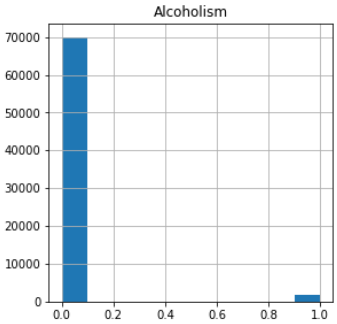
Data set Overview

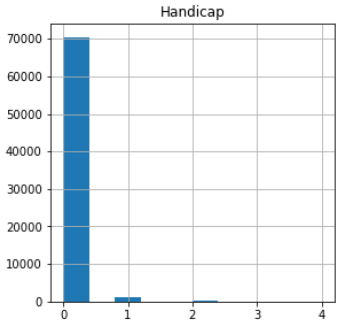


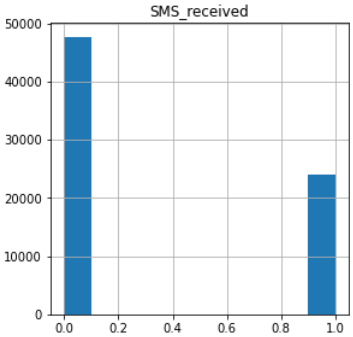










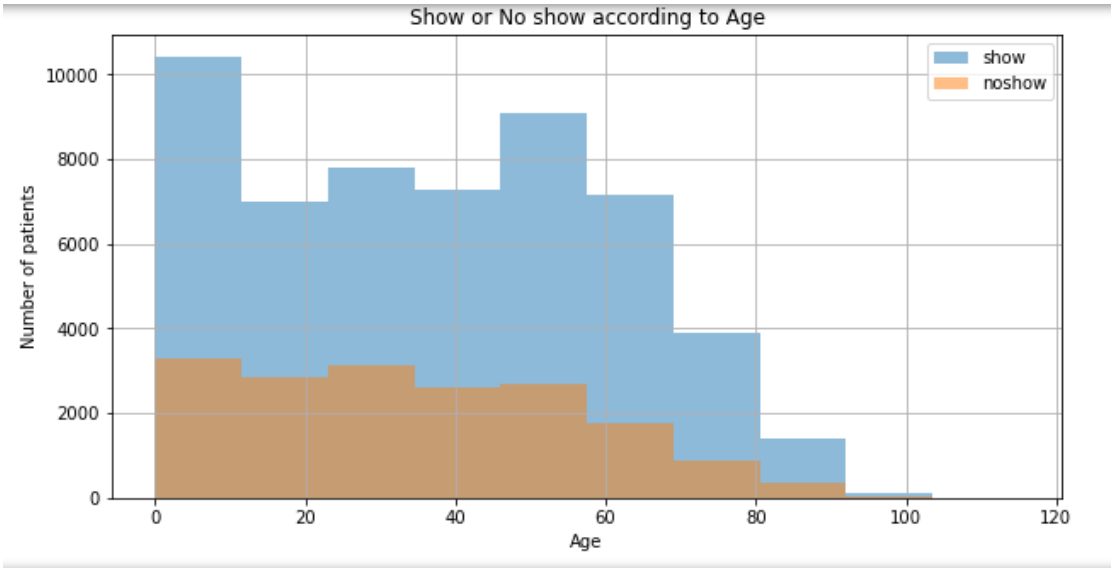


- Half of patients didn't receive SMS
- Almost of patients had Scholarship
- Almost of patients didn't suffer from Diabetes, Alcoholism or Handicap
- More than 10000 patients suffered from Hypertension

Visualizations

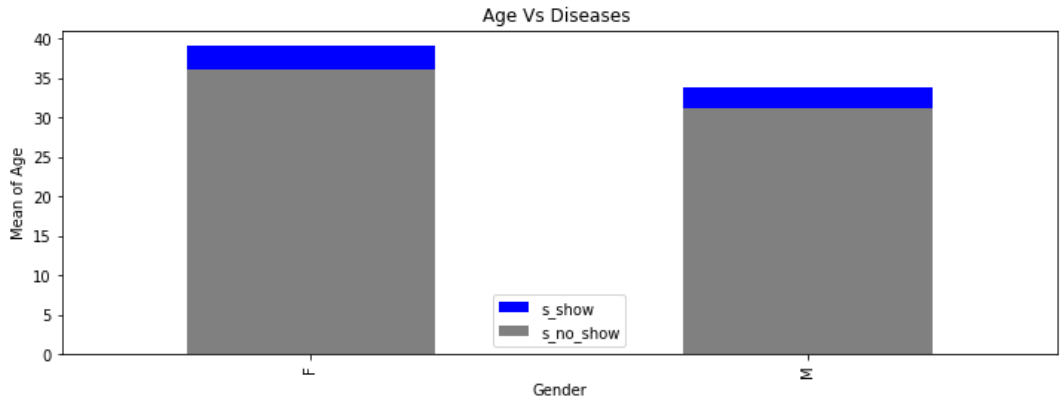
Figure to answer the questions which asked in the beginning

Q1: Did Age effect on attendance?!



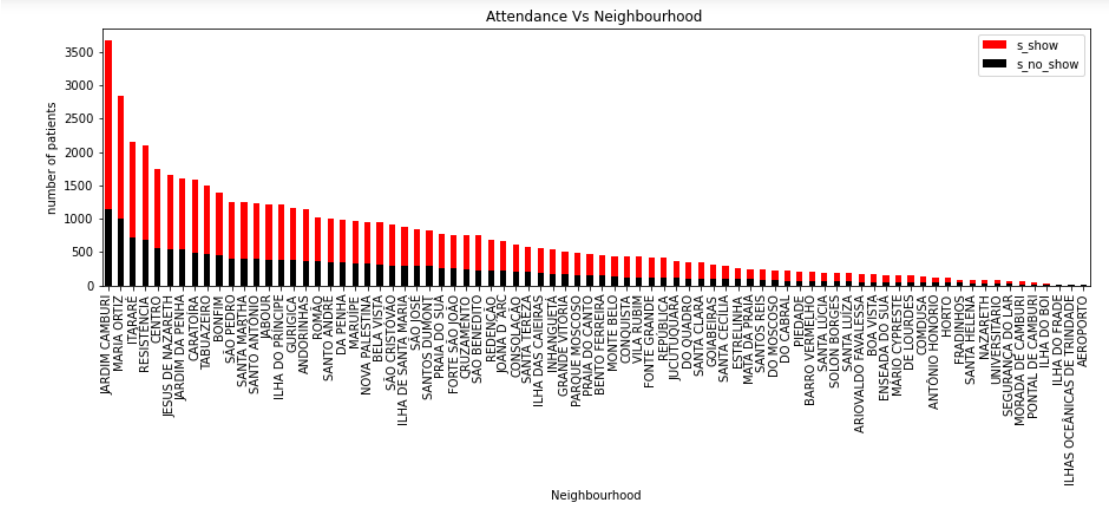
As shown from figure there is a high attendance rate from 1 to 20 years and it's so low from 90

Q2 : Is there relation between age and Diseases?!



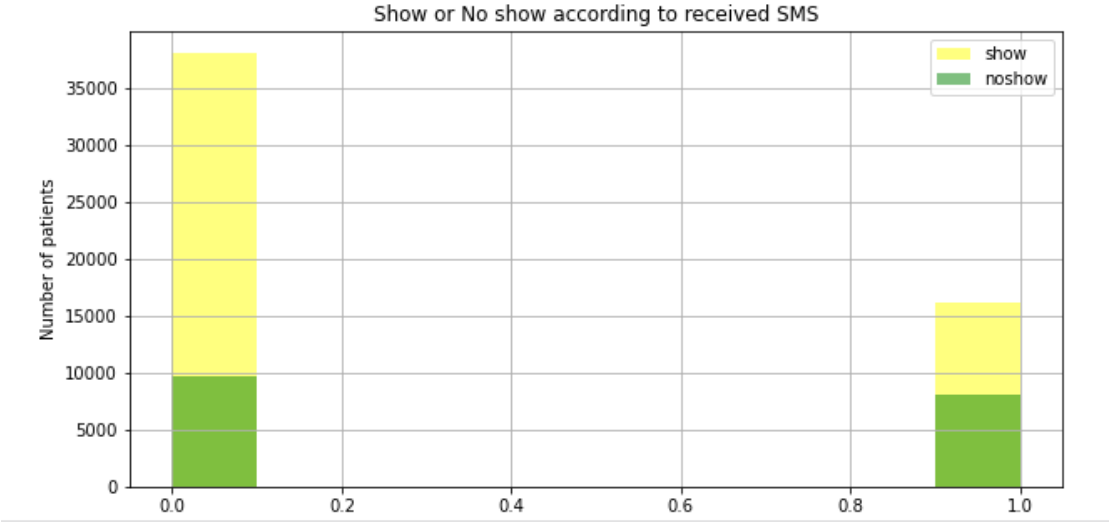
As shown there is not effected relation between age and diseases for each gender

Q3: Is there relation between Neighborhood and Attendance?!



There are place had a high attendance more than other, surly this is for reasons.

Q4: Did received SMS effect on attendance?!



As shown attendance of patients who did not receive SMS more who received SMS!

Conclusions

Throw this investigation (No-show appointments) data set we fine :

- Average of age of patients is 37 and max age is 115
- There is a high attendance rate from 1 to 20 years old and it's so low from 90 years old
- There are place had a high attendance more than other, surly for some reasons
- Attendance of patients who did not receive SMS more who received SMS!
- Half of patients didn't receive SMS
- Almost of patients had Scholarship
- Almost of patients didn't suffer from Diabetes, Alcoholism or Handicap
- More than 10000 patients suffered from Hypertension