

Investigate a Dataset: no show appointments-kaggle2-may-2016

Introduction

- 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 Familia.
- Be careful about the encoding of the last column: it says 'No' if the patient showed up to their appointment, and 'Yes' if they did not show up. We will address that in the next steps so as not to get confused

Questions

- attendance ratio
- which features affect in attendance

Data Wrangling

- In this part, we are going to view the data
- There is no Missing Data
- There is no duplicate Data
- There is a wrong age as it cannot be less than zero
- Finding wrong age
- Finding Duplicate attendance status for patients

Data Wrangling

- Deleting wrong age
- Deleting Duplicate attendance status for patients
- Deleting unimportant columns
- Fixing columns name

The data is now ready to explore

Exploratory Data Analysis

Now that we've trimmed and cleaned the data, we move on to exploration. Calculate statistics and create visualizations address the questions we asked in the Introduction section.

- **Research Question 1 (attendance rate)**
 - It turns out that the attendance rate is high
 - Let's explore what are the features affecting attendance rate
- **Research Question 2 (Gender affection in attendance rate)**
 - Gender has no obvious effect on attendance rate
- **Research Question 3 (age affection in attendance rate)**
 - At a young age, the interest in attendance increased
- **Research Question 4 (Gender & Age affection in attendance rate)**
 - There is no Affection
- **Research Question 5 (Neighbourhood affection in attendance rate)**
 - It is clear that the patient's place of residence reflects positively and negatively on his response to attendance
- **Research Question 6 (Scholarship affection in attendance rate)**
 - Scholarship has no effect on attendance rate
- **Research Question 7 (Hypertension & Diabetes & Age affection in attendance rate)**
 - These chronic diseases do not have a clear effect on attendance rate
- **Research Question 8 (Alcoholism Affection in attendance rate)**
 - Alcoholism has no clear effect on attendance rate
- **Research Question 9 (SMS_received affection in attendance rate)**
 - SMS_received has no clear effect on attendance rate

Conclusions

- It turns out that the attendance rate is high
- Gender has no obvious effect on attendance rate
- At a young age, the interest in attendance increased
- It is clear that the patient's place of residence reflects positively and negatively on his response to attendance
- Scholarship has no effect on attendance rate
- These chronic diseases do not have a clear effect on attendance rate
- Alcoholism has no clear effect on attendance rate
- SMS_received has no clear effect on attendance rate