

Project: Investigate a Dataset - [Hospital Appointment No-Show Dataset]

Table of Contents

- [Introduction](#)
- [Data Wrangling](#)
- [Exploratory Data Analysis](#)
- [Conclusions](#)

Introduction

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

In this project we will answer some question on why patients miss their scheduled appointment. We will be predicting some of the factors that affecting patient attendance in the Hospital.

Question(s) for Analysis

Does the patient Gender and Age has any effect on attendance? Does receiving SMS remainder increase the chance of a patient showing for their appointment? Does the location or neighborhood of the patient play a role in missing or showing up for appointment? Does the disease type affect the patient's that show up or missed appointment?

Exploratory Data Analysis

Research Question 1 (Are young people showing up for appointment more? Are women more likely to showup for schedule appointments than men?)

In this stage we were able to gather our Data from a CSV file and explore the first few rows to see the sturcture of the data and it's general properties. We then check to see the dimation of the data, checked the data for duplicated values, missing values and possible errors; We found some duplicated values in the 'PatiendId' column and remove them. checked the dataset type to see if there are any further missing data to handle. And lastly we were able to see some statistical information about the data on min, max, and mean. And finally we cleaned the data by removing duplicates, unnecessary data, correcting the columns names and dropping unnecessary columns.

Now that I have trimmed and cleaned my data, it's time to move on to exploration. Computing statistics and creating visualizations with the goal of addressing the research questions that I posed in the Introduction section.

Conclusions

Having finished the analysis we found that Neighbourhood has the highest effect on attendance having some neighbourhood having the greater number of patients and also having the greater number of showing.

Also Age has a greater effect on the showing rate as the age range 0:10 has the highest number of shows and the age range 40:100 and above has lower shows.

Limitations

There may be some patterns that I wasn't able to identify due to my lack of expertise. However, from the data we know that if the hospital does not capture the patients details correctly the patient will not receive an SMS but if patient details were captured correctly the chances of the person coming for hospital appointments will increase. So we need to search for more factors to help patient remember their appointments and show up

There is no clear correlation between showing and chronic diseases, gender and enrolment in the welfare program.