

Project Investigate Dataset: Hospital Appointments Data Analysis

By: I Gede Wibawa Cakramurti

Attached to this report, is the analysis for “No-show appointments” from Kaggle. The dataset collects information from 110,527 medical appointments in Brazil with certain characteristics, such as diseases, scholarships, and attendance.

In doing the analysis, there are two main questions:

1. Does the patients with hypertension, linked with age, are more likely to come as scheduled?

Because the data shows that there are more patients with hypertension than other diseases, I then curious to take this further. First, I looked whether there are correlation whether patients with hypertension linked to certain age. Then, I separate between patients with hypertension and attend the hospital as scheduled and patients without hypertension and not attend the hospital as scheduled. Turns out, there are no direct correlation.

2. Do patients that received sms are more likely to come as scheduled?

I think this is the obvious question that had to be asked. I count the mean between people that received sms, linked with age. Then, I checked whether some people with certain age has higher chance of receiving sms. Turns out, every age has the same ratio of receiving and not receiving sms. I then compare between patients showing and not showing linked with the mean of received sms. And there is a correlation that patients who received sms tend to show as scheduled.

In wrangling the data, I just looked at whether there are null values, different types, and duplicated rows. I also cleaned some bad data found when describing the data.

Summary Statistics

1. The mean of patients' age is 37.
2. Most patients didn't received scholarship showed by the described data that up to 75% still 0 or False.
3. Most patients didn't have hypertension, diabetes, alcoholism, nor handicap showed by the described data that up to 75% still 0 or False.
4. Not up to 50% patients received sms showed by the described data that up to 50% still 0 or False.

Conclusions

There are two main conclusions that I could take from this analysis:

1. Most patients are not showing to the hospital (around 4 out of 5 people are not coming) and there are no correlation when performing analysis of showing/not showing to patients with hypertension, linked to age.
2. Patients who received a sms about the schedule will more likely to show to the hospital than those who don't received the sms.

Resources: Udacity Courses

Thank you! :)