No-Show Appointments Dataset

Introduction:

This dataset contains 100k medical appointments in Brazil and is focused on the question of whether or not patients show up for their appointment and a number of characteristics about the patients who make the appointments, through the analysis processes I will put a structure for what I need from data, what the questions I want to answer through the data, then I will clean the dataset to can be ready for analysis without any bias, so we can get right answer from the dataset, finally, I will use Python libraries like pandas, NumPy and matplotlib to help me to clean and analysis dataset to can answer these questions.

Here some questions we want to answer from dataset:

Q1- Is there a correlation between missing the appointment and the age of the patient?
Q2- Is there a correlation between missing the appointment and the Gender of the patient?
Q3- Is there a correlation between missing the appointment and the neighborhood?
Q4- Is there a correlation between missing the appointment and not receiving an SMS message?

Data Wrangling:

Firstly, discover dataset shape and data types to correct wrong data type, and I discover the dataset has (110,527 record) and (14 column), and has two columns with incorrect types ScheduledDay and AppointmentDay So, we will convert them to DateTime types, then if any missing value in dataset and didn't find any missing value, In next step I use discrep to see any unnormal value in dataset, from discrep method I find the min age is -1 so I will drop it, in last step Investigate if there is duplicate records in the dataset, so no duplicate value in data set.

Exploration Phase:

General Exploration:

- The proportion of females is greater than that of males.
- the period of time is starting from 2015-11-10 for scheduling dates and from 2016-04-29 for appointments until -> 2016-06-08.
- The Age range contains nearly all ages (from 0 to 115) and the majority is a very young age and between 20 and 55.
- There are 81 neighborhoods where the appointment takes place.
- Most of the patients have No: Scholarship 90%, Hypertension 80%, Diabetes 93%, Alcoholism 97%, Handcap 98% (('according to the data')).
- 68% of patients did not receive SMS.
- Nearly 20% of patients missed their appointment.

Exploratory Data Analysis:

- The main goal of this analysis is to find why patients miss their appointment. and this will be done by finding the relationships between the patients who missed and attended their appointment and the other characteristics in the dataset.
- **Q1- Is there a correlation between missing the appointment and the age of the patient?**
    - Results:
      There is no strong correlation between age alone and missing the appointment:
      The mean of patients missing an appointment across all ages (19.85%) is close to the overall average (20.19%)
      After some exploration we could say:
      patients from 12 to mid-30s are more likely to miss their appointments
      kids and patients in their 60s, 70s and 80s are more likely to show up to their appointments: The reason for that could be because there is someone who is taking them to their appointments.
- **Q2- Is there a correlation between missing the appointment and the Gender of the patient?**
    - Results:
      we could say that there is no correlation between gender alone and missing the appointment.
      The percentage of male and female patients missing their appointments is the same and same as the general average.
- **Q3- Is there a correlation between missing the appointment and the neighborhood?**
    - Results:
      Most of the neighborhoods are close to the same general percentage of 20%
      In general, the more patient coming from the neighborhood the more patient will miss their appointments

The neighborhoods with most patients showing to their appointments have low patients' numbers. So, we couldn't conclude that these neighborhoods have a higher chance that patients won't miss their appointments.

- **Q4- Is there a correlation between missing the appointment and not receiving an SMS message?**
    - Results:
      Most patients didn't receive an SMS.
      The patients who received an SMS have a higher percentage of missing the appointments 27.6% to those who didn't 16.7%.
      For the patients who missed their appointments, 43% of them received an SMS.
      According to the data, receiving an SMS, didn't make patients more likely to show to their appointments.

Conclusions:

We investigated the correlation between some of the characteristics of the patients and whether they missed or showed to their appointment. And these are the results of these analyses:

- patients from 12 to mid-30s are more likely to miss their appointments
- kids and patients in their 60s, 70s, and 80s are more likely to show to their appointments
- There is no correlation between gender alone and missing the appointment.
- The neighborhoods with most patients showing to their appointments have low patients' numbers. So, we couldn't conclude that these neighborhoods have a higher chance that patients won't miss their appointments.
- Receiving an SMS did not increase the chance that the patient would show to their appointment.

Limitations:

- The period when the data collected was just nearly two months, and this short period doesn't make the data reliable to make a strong correlation
- There was no explanation of how the data was collected, and if it was collected in a way that makes these analyses more reliable.
- The Handcap and SMS_received columns weren't explained clearly on the dataset page.
- The dataset didn't explain why the patient comes to the doctor, which may give an additional explanation about why they miss their appointments.

Data Base Source:

https://www.kaggle.com/ahmedmohameddawoud/no-show-appointments