

DATA ANALYSIS ON NO-SHOW APPOINTMENTS

INTRODUCTION

The dataset contains over 100,000 medical appointments in Brazil and is aimed at answering the following questions:

- Which gender is likely to show up for an appointment?
- Does age affect a patient showing up for an appointment?
- Did patients who got the scholarship show up?
- Does receiving SMS increase the chances of a patient showing up?

The dataset has 13 dependent variables with 'No_Show' as the target variable.

INSPECTING THE DATASET

After viewing the data, I observed that there were 110527 rows and 14 columns with no entry missing. There were no duplicate values too.

Also, I realized some column names were wrongly displayed. So, I renamed them and converted all the headings into lower case for clarity and simplicity.

From inspection, I observed that the "patientId" and "appointmentId" were of float and integer datatypes respectively. So, I converted them to strings.

Per analysis, I found that the handicap had five unique values, with 0 and 1 constituting 98% of the total. So, I converted the 'handicap' column to binary (0 and 1) using a lambda function.

I converted the "scheduledday" and "appointmentday" to date using the datetime function.

I also dropped the ages less than 0.

With the **Exploratory Data Analysis (EDA)**, I based my research on four areas on the dataset: Gender, Age, Scholarship, and SMS notification.

From the EDA, more females booked medical appointment than males. Further, patients below the ages of forty years showed up less often. Surprisingly, the number of patients who got the scholarship for free medication did not show up. Finally, it appeared that the proportion of patients that showed up after receiving SMS were less than those that did not received the notification.