Medical Appointments No Shows

Why do 30% of patients miss their scheduled appointments?

The dataset collects information from more than 100k medical appointments in Brazil and is focused on the question of whether or not patients show up for their appointment. The Project Medical Appointments No Shows is investigating Medical Appointment No Shows dataset which contains historical data for more than 110K appointments made across different medical facilities in Brazil for more than 60k patients, for each record there are 14 Variables, metadata related to appointments date, patients gender, age, medical condition, social support coverage and facilities and 1 TARGET variable "Wither the patient attended the appointment or not. (Based on a kaggle dataset)

Project Target

What is the explanation for a person making a doctor appointment, receives all the instructions and no-show. Who to blame?

Data Definition

PatientId: Identification of a patient that is unique for each person.	Age: How old is the patient.	Alcoholism: T or F.
AppointmentID: Identification of each appointment.	Neighborhood: Where the appointment takes place.	Handcap: T or F.
Gender: Male or Female.	Scholarship: 1 or 0 (this is a program in Brazil to support poor people with their cost of living),	SMS_received: 1 or more messages sent to the patient.
AppointmentDay: The day of the actual appointment, when they have to visit the doctor.	Hipertension: T or F.	No-show: "Yes" or "No" ("No" means they showed up on their appointments while "Yes" means they didn't!).
ScheduledDay: The day someone called or registered the appointment, this is before appointment of course.	Diabetes: T or F.	

Data Cleaning

• Edit the "No-show" column to be in positive form instead of negativity.

• Edit the "ScheduledDay", and "AppointmentDay" columns datatype to be Datetime.

• Edit the "PatientId", and "AppointmentID" columns datatype to be String.

• Delete row with negative age value.

Algorithms

- Classification> Logistic Regression Model.
- Regression > Linear Regression Model.

Questions for Analysis

• What is the percentage of patients who show up on their appointments vs. who not?

• Is one gender more committed to medical schedules than another?

• Where do most appointments take place?

• Is the duration between registration and appointment affect the ability to show up?

• Are patients who received SMS messages reminding them of the appointment likely to attend?

• What is the percentage of patients diagnosed with diabetes, hypertension, alcoholism and disability?

• Is drinking alcohol a cause of missing appointments?

• Do older patients more committed to medical schedules than others?

Tools

• **Pandas**: a library offers data structures and operations for manipulating numerical tables and time series.

• **Numpy**: a library used for working with arrays. It also has functions for working in domain of linear algebra, fourier transform, and matrices.

• **Matplotlib**: a plotting library for the Python programming language and its numerical mathematics extension NumPy.

• **Seaborn**: a data visualization library built on top of matplotlib and closely integrated with pandas data structures in Python.