Codebook for No Show Dataset Data Overview

## Credentials

The direct link to data is: <https://www.kaggle.com/joniarroba/noshowappointments>

## Business goal

* Predict no-show probability for a patient
* Find the variables that affect our business question the most
* Reduce no-show
* Better utilize the medical team
* Save MONEY!

Once we are able to predict no show, 2 methods can be applied:

* Reduce amount of no shows
* Double booking (as a derivative)

## Data description

This data set is a data frame of 15 variables over more than a 110k rows. Each row represents an unique appointment of patient ID and doctor.

Neighborhoods poverty level derived from the index of quality of urban municipality (IQU). The IQU, developed by the Instituto Polis (Sao Paulo, SP, Brazil): <http://legado.vitoria.es.gov.br/regionais/indicadores/iqu/iqu.asp>

## 

## Variables description

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **ID** | | **Variable Name** | **Description** | **Type** | **Possible values** |
| 1 | PatientId | | Patient ID - random UUID | Numeric | Random ID |
| 2 | | AppointmentID | Appointment ID - random UUID | Numeric | Random ID |
| 3 | | Gender | Gender of the patient | Textual | F=Female M=Male |
| 4 | | ScheduledDay | Scheduled day of appointment | DayTime |  |
| 5 | | AppointmentDay | Actual appointment day | DayTime |  |
| 6 | | Age | Age of the patient | Numeric | -1(NA) : 115 |
| 7 | | Neighbourhood | Neighbourhood name in Vitória | Textual | Name |
| 8 | | Poverty | Poverty of the neighbourhood | Numeric | Scale of 0 (NA) – 5 (Highest) |
| 9 | | Scholarship | Social welfare scholarship received? | Binary | 0=No  1=Yes |
| 10 | | Hipertension | Hypertension existence | Binary | 0=No  1=Yes |
| 11 | | Diabetes | Diabetes existence | Binary | 0=No  1=Yes |
| 12 | | Alcoholism | Alcoholism existence | Binary | 0=No  1=Yes |
| 13 | | Handcap | Amount of handicaps a person presents | Numeric | 0 : 4 |
| 14 | | SMS\_received | SMS reminder received? | Binary | 0=No  1=Yes |
| 15 | | No-show | Show/No show to the appointment | Textual | No/Yes |
| 16 | | x\_coor | X coordinate | coordinate |  |
| 17 | | y\_coor | Y coordinate | coordinate |  |
| 18 | | waiting\_time | Number of days from schedule to actual appointment date | Numeric | 0-150 |
| 19 | | same\_day | Number of appointments in same day per PatientId | Numeric | 0-10 |
| 20 | | ever\_before | Number of ever before appointments per PatientId | Numeric | 0-10 |

## Business questions

1. Does the age and medical condition of the patient affect the no-show?
2. Is there a relationship between no-show to a specific doctor?
3. Are there more no-shows on a specific time of the year?
4. Is there a relationship between the duration from schedule to appointment to the probability of no-show?
5. Does the patient neighborhood can predict a no-show?

## Who needs to review the business question?

The medical center management

## Related links

<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3962267/>

<https://rstudio-pubs-static.s3.amazonaws.com/260312_b0f48fe809e44515a168cf5a2ce5ac75.html>