Codebook for SF Bay Area Bike Share Dataset

# Data Overview

## Credentials

This data set can be downloaded from Kaggle data sets.

The direct link to data is: <https://github.com/Asafamitai/IDC-BDA-Exercises/blob/master/KaggleV2-May-2016_with_new_cols_sat_may5_1126am.csv>

Business goal

This data was collected to answer the question:

1. What are the characteristics of a NO SHOW patient?
2. How can we predict a potential NO SHOW?
3. Can we prevent a NO SHOW case?

## Data description

This data set is a data frame of 58 variables over 110527 rows. Each row represents an appointment of a specific patient.

There are missing values in the data (for example zip code) .

# Variables description

|  |  |  |  |
| --- | --- | --- | --- |
| Variable Name | Description | Type | Possible values |
| PatientId | The patient ID – will repeat it self if a patient returns to have another appointment. | float64 | Positive numbers |
| AppointmentID | An appointment ID – is unique for each row. | int64 | Positive numbers |
| Gender | Male/female. Female is the greater proportion, woman takes way more care of they health in comparison to man | object |  |
| ScheduledDay | The date and time of the scheduling of an appointment. | object | LongDate+Time |
| AppointmentDay | The day of the appointment | object | LongDate+Time |
| Age | How old is the patient | int64 | Positive numbers |
| Neighbourhood | Where the appointment takes place | object | text |
| Scholarship | True of False . Observation, this is a broad topic, consider reading this article <https://en.wikipedia.org/wiki/Bolsa_Fam%C3%ADlia> | int64 | positive |
| Hipertension | True = 1 False = 0 | int64 | positive |
| Alcoholism | True = 1 False = 0 | int64 | positive |
| Handcap | True = 1 False = 0 | int64 | positive |
| SMS\_received | 1 or more messages sent to the patient. | int64 | positive |
| No-show | Ture of False | object |  |

PatientId - Identification of a patient

AppointmentID - Identification of each appointment

Gender = Male or Female . Female is the greater proportion, woman takes way more care of they health in comparison to man.

DataMarcacaoConsulta = The day of the actuall appointment, when they have to visit the doctor. DataAgendamento = The day someone called or registered the appointment, this is before appointment of course.

Age = How old is the patient.

Neighbourhood = Where the appointment takes place.

Scholarship = Ture of False . Observation, this is a broad topic, consider reading this article <https://en.wikipedia.org/wiki/Bolsa_Fam%C3%ADlia>

Hipertension = True or False

Diabetes = True or False

Alcoholism = True or False

Handcap = True or False

SMS\_received = 1 or more messages sent to the patient.

No-show = True or False.