**Nanodregree Data Science I**

Project 2: Investigate a Dataset

Lucas Amorim Bonini

30/09/2018

# Dataset characteristics

This dataset collects information from 100k medical appointments in Brazil and is focused on the question of whether patients show up for their appointment. Several characteristics about the patient are included in each row.

## Data Dictionary

* 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.
* ScheduledDay = The day of the actual appointment, when they have to visit the doctor.
* AppointmentDay = 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.
  + Indicates whether or not the patient is enrolled in Brasilian welfare program Bolsa Família.
* Hipertension = True or False
* Diabetes = True or False
* Alcoholism = True or False
* Handcap =  Number of disabilities a person has. (1, 2, 3, 4)
* SMS\_received = 1 or more messages sent to the patient.
* No-show = True or False.
  + Be careful about the encoding of the last column: it says **‘No’** if the patient showed up to their appointment, and **‘Yes’** if they did not show up.

Source: <https://www.kaggle.com/joniarroba/noshowappointments>

# Questions posed

Which gender has the highest attendance?

Do people who receive SMS tend to attend more?

The greater the number of disabilities, the greater the chances of attending?

What is the relationship between the date it was marked and the attendance record?

Which neighborhood shows the highest attendance rate?

Which age group has the highest attendance?

What month were there more appearances?

# Description

1. Which gender has the highest attendance?

<https://pandas.pydata.org/pandas-docs/stable/>

proporção