Nanodregree Data Science I

Project 2 - Investigate a Dataset

Lucas Amorim Bonini

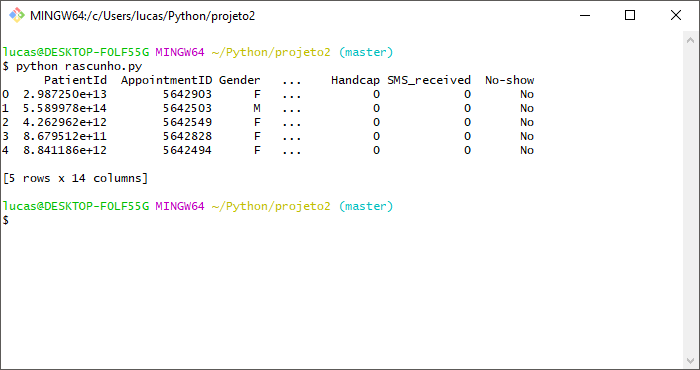
30/09/2018

# 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.

* ‘ScheduledDay’ tells us on what day the patient set up their appointment.
* ‘Neighborhood’ indicates the location of the hospital.
* ‘Scholarship’ indicates whether or not the patient is enrolled in Brasilian welfare program [Bolsa Família](https://www.google.com/url?q=https://en.wikipedia.org/wiki/Bolsa_Fam%25C3%25ADlia&sa=D&ust=1536765425666000).
* 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

* A statement of the question(s) you posed
* A description of what you did to investigate those questions
* Documentation of any data wrangling you did
* Summary statistics and plots communicating your final results