OkCupid

Date-A-Scientist

2

HELLO!

I am Vinicius Amim

I’m a junior data scientist. You can find me at vfamim@gmail.com.

3

OkCupid

This project analyzes data from online dating app OkCupid. These app give us access to a wealth of information that we've never had before about how different people experience romance.

The goal of this project is to try to predict some characteristics of the platform's users.

4

TABLE OF CONTENTS

5

ABOUT DATA

Continuous variable

- **age** - age of user
- **height** - height of user
- **income** -income of user

Date variable

- **last_online** - last login

Categorical variable

- **body_type** - body type of user
- **diet** - dietary information
- **drinks** - alcohol consumption
- **drugs** - drug usage
- **education** - educational attainment
- **ethnicity** - ethnic backgrounds
- **job** - employment description
- **location** - user locations
- **offspring** - children status
- **orientation** - sexual orientation
- **pets** - pet preferences
- **religion** - religious background
- **sex** - gender
- **sign** - astrological symbol
- **smokes** - smoking consumption
- **speaks** - language spoken
- **status** - relationship status

Short answer

- **essay0** - My self summary
- **essay1** - What I'm doing with my life
- **essay2** - I'm really good at
- **essay3** - The first thing people usually notice about me
- **essay4** - Favorite books, movies, show, music, and food
- **essay5** - The six things I could never do without
- **essay6** - I spend a lot of time thinking about
- **essay7** - On a typical Friday night I am
- **essay8** - The most private thing I am willing to admit
- **essay9** - You should message me if…

6

CENTRAL QUESTION

- I became interested in the status(in a relationship or not) and height column.
- Can we predict status based on the following features:
  - Body type
  - Diet, drinks, drugs and smoke?
- Can we predict height based on the following features:
  - User age
  - Status relationship

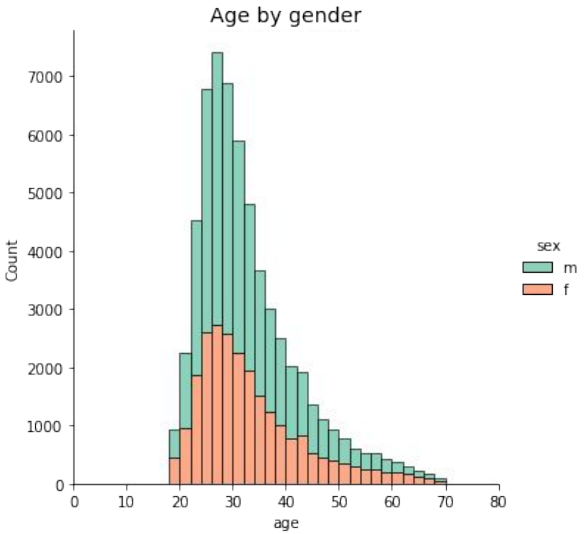
7

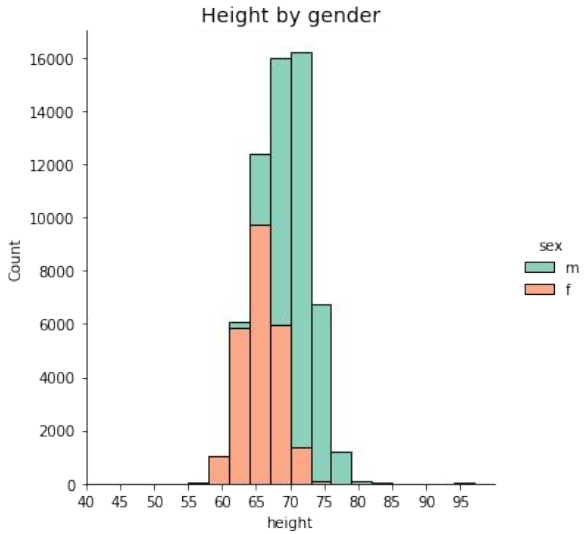
Data Visualization:

DataViz some features for Machine Learning models.

8

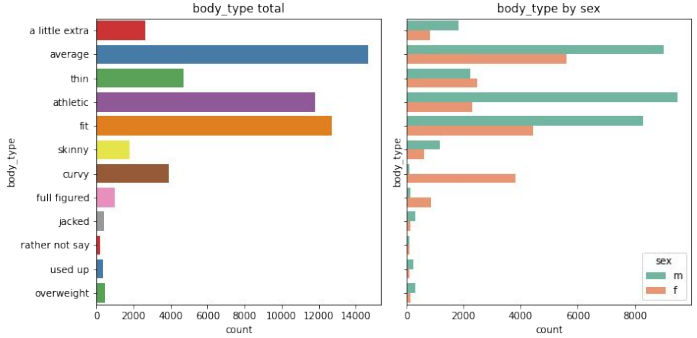
CONTINUOUS CHARTS

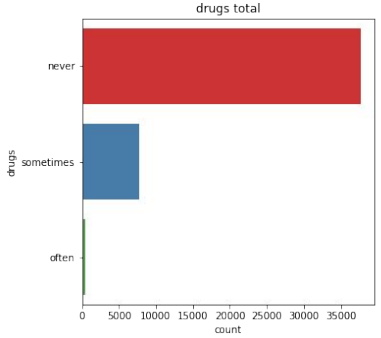


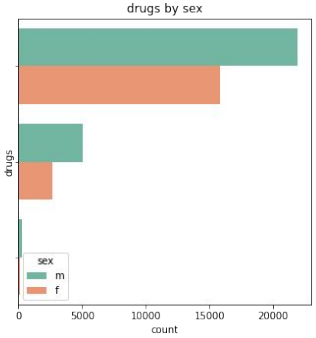


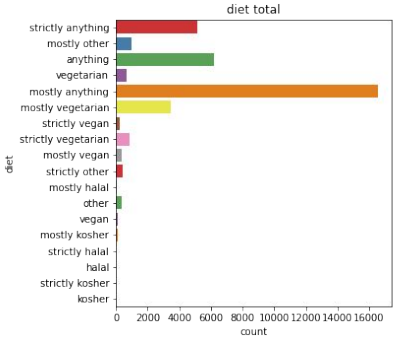
9

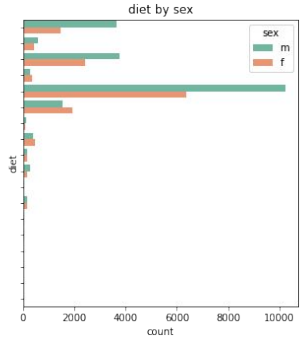
CATEGORICAL CHARTS

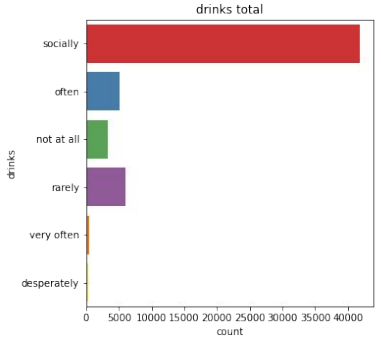


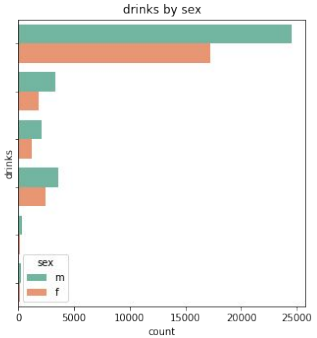


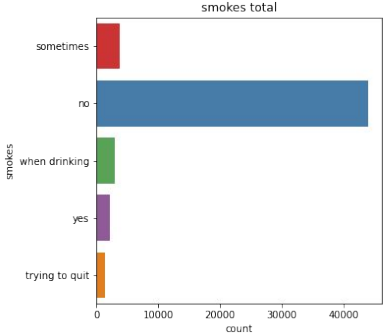


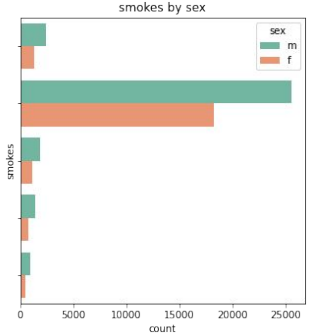












10

Machine Learning:
Classification

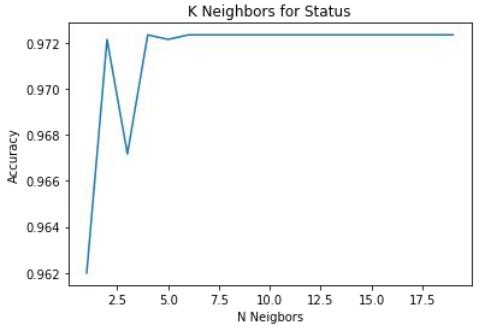
Whoa! Let’s predict the status relationship. Whether user is single or not.

11

CLASSIFICATIO N APPROACHES

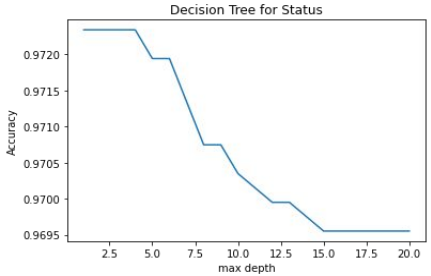
K-Nearest Neighbors Classifier:

- Range of N Neighbors values from 1 to 17.5
- Predictions were very accurate



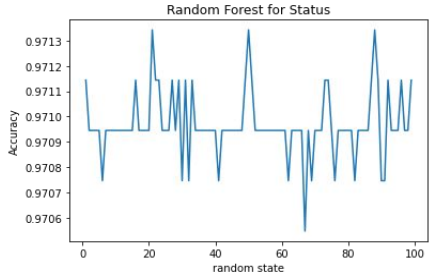
Decision Tree:

- Range of Max Depth values from 1 to 20
- Predictions were very accurate



Random Forest:

- Range of Random State values from 1 to 100
- Predictions were very accurate



12

Machine Learning:
Regression

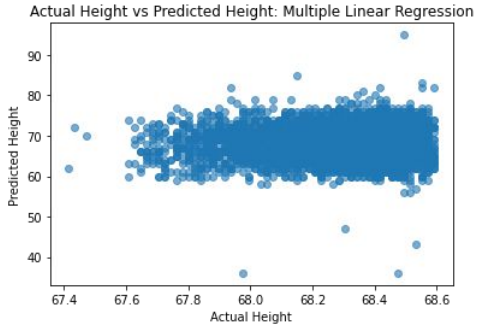
Now let’s predict the height.

13

REGRESSION APPROACHES

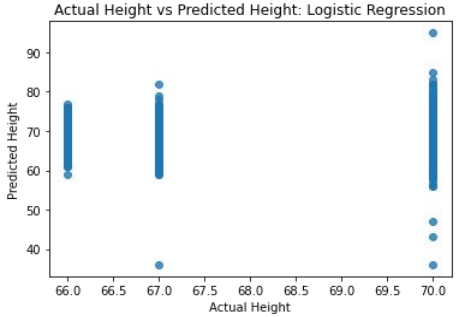
Multiple Linear Regression:

- Predictions were not very accurate



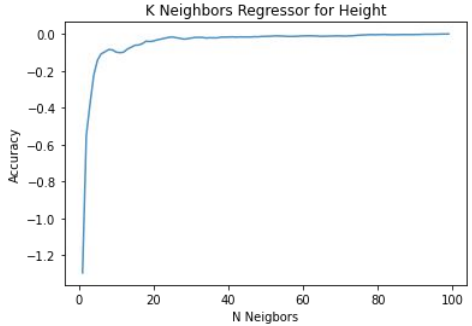
Logistic Regression:

- Predictions were not very accurate



K-Nearest Neighbors Regression:

- KN Regressor approaches were not very accurate



14

CONCLUSION

- The regression model for predicting **height** were mostly unsuccessful in showing any strong correlation.
- Classification model for predicting **status** (single or not) based in some habits showed a high accuracy, highlighting by K-Neighbors Classifier results.

THANKS!

Any questions?

You can find me at

- LinkedIn
- vfamim@gmail.com