**Experian Code Challenge – Data Scientist - Targeting**

Completion Timeframe: 10 business days

Estimate time allocation: 2-4 hours

**Overview**

Congratulations for progressing to our technical assessment stage having performed well during your initial interviews. At Experian we put our best efforts into selecting the brightest minds to become a part of our data science and development teams.

Thank you for taking the time to work on this code challenge. Once completed it will allow us to understand your data science skills and how meticulous and creative you are, not only with code and outputs, but also:

* Quality
* Assumptions and approach to solve the problem
* Coverage
* Technology handling
* Creativity

**Objective**

Present a solution to a real problem during a sprint cycle (10 business days); you can use whatever language or framework you like most, any open source (we love open source) technology stack you like, preferably stable and well supported, to avoid any issues during the evaluation of your results.

**Instructions**

* Please make the results available at Github, Bitbucket or any other repository you prefer. However, you must provide clear instructions on how to run or deploy your solution
* Make sure you don’t forget any necessary library
* like many companies, we also use SQL to pull data out of the DBs, so it would be great if you include your queries and any data preparation scripts.
* Your code might be submitted to an internal set of tools that evaluate code quality (like Sonar or Veracode), so make sure you don’t use any deprecated library and please follow best practices
* **You can bring in any additional datasets that support your insights or create additional value. The limit is your creativity but remember: it must make sense.**

**Welcome to the Coffee Classification Challenge:**

Subu, our data scientist at Experian is originally from India and since he arrived in Australia, he has not been much into coffee (he just follow the team’s preferences sometimes), as he can’t find the perfect taste that reminds him of home. When Subu has coffee, he usually goes for the sweet, but not a too strong flavour, which makes him remember his childhood and brings him strong positive emotions.

Marcio, as a good Brazilian, grew up drinking coffee thrice a day. A ceremony he replicates in Australia. He particularly likes Brazilian coffee, from a region called “south of minas” (in Portuguese: sul de minas, which refers to the south of the “Minas Gerais” state), and specifically from a place called “Kaquend Farm” (fazenda kaquend, in Portuguese), which produces the best coffee beans in the country.

Marcio and Subu are going for a business trip to the Experian Brazil office to collaborate on a Mosaic project, and Marcio would like to take Subu to a coffee experience which will make him cry.

On their return to Australia, Marcio wants to bring back some coffee bags to give to the team as gifts:

* To Rafa (who is also Brazilian, and we will assume he has the same taste as Marcio)
* To James (who is from Hawaii and like coffee in a very American way: soft, no aroma, no flavor, and very sweet)

Marcio, as a good leader, wants to please so we need to find amongst the Brazilian coffee options, the coffee bags that will make his teammates very happy. He also wants to build a safety stock for himself.

There are also international options Marcio can buy from Duty Free, so non Brazilian options can be also considered.

**Your goal is to create a classification and a cluster of the coffees based on the information above, recommending to Marcio the coffees that are the most likely to help him achieve his ambitious goals, as well as bring him inspiring insights to help him glamourize his addiction.**

The suggested coffee database is available at <https://database.coffeeinstitute.org/>

We would love to see you bring any others. Remember, we live in a world of data!

There is a project on github you can use to get the data: <https://github.com/jldbc/coffee-quality-database/tree/master/data> (credits to the author [jldbc](https://github.com/jldbc))

You can use the already extracted datasets or tweak the scrapper to get the data by yourself, your choice.

***Good luck, Rockstar! Marcio.***