Welcome

Welcome to the Data Science candidate test.

In this test you will help Wheelie Wonka, a hypothetical bike sharing company in Boston, solve a critical business problem. As such, you are encouraged to tackle the problem with this client in mind. However, please bear in mind that the assessment is heavily focussed on technical abilities and the approach you choose to solve the problem.

Challenge

Wheelie Wonka put their users first. To enhance their mobile app experience, they would like to show bike availability in the future to their users. An important component in achieving this goal is to predict, at the beginning of a bike trip, the​ *trip duration*​ and ​*ending bike station*​.

To make these predictions, Wheelie Wonka has given us data about the bike trips, the bike stations and weather information as recorded by a nearby weather station.

For the test, using the attached data, please carry out the following tasks:

* Go through a preliminary exploration of the data and choose ​**one**​ of these two prediction tasks:
  + Predict the trip duration
  + Predict the trip ending station
* Explore and derive insights from the data for the purpose of your chosen prediction.
* Develop and evaluate a machine learning model to make the type of prediction you chose.
* Write between 300 and 500 words along with visualizations to justify your workflow — for example, why did you choose a certain predictive model?

Data

The data for this test is a set of CSV files with the following self-explanatory names:

* hubway\_stations.csv
* hubway\_trips.csv
* weather.csv

You can find these files in the data folder, together with two documents that you may use to better understand the data.

The data is not perfect and sometimes inconsistent, as you would expect real-world data to be. This is part of the challenge of this test. The bike trip data comes from the Hubway public bicycle sharing system in Boston, United States, and the weather data can be pulled from the NOAA.

Evaluation Criteria

Data scientists at GA must be strong across a wide set of knowledge areas including general business administration, statistics, machine learning, optimization, data wrangling, visualisation, programming and presentation writing. As such, we will evaluate this test on the above topics.

At GA, we use Python extensively for data science. You may choose either of these languages, but we explicitly ask that you do not use any external analytics software such as, but not limited to, Alteryx, Tableau, Qlik or SAP.

Note that we value the correct use of statistical and machine learning methods as much as the clarity and structure of the code implementation.

In addition, alongside the code, we expect clear written explanations for the data science methods that you use and the choices that you make. These explanations should form a narrative that explains how you reach the goal of predicting either the trip duration or the ending station.

There are many ways to predict the trip duration or the ending station given the available data and we do not expect you to cover all of them. Instead, we prefer ​*one or a few well-implemented* and ​*well-explained ways* to make such predictions. In particular, we would like to see a complete data science analysis, going from the exploratory analysis to model evaluation.

While it is important to get good predictions, the way the methods are applied, explained and justified are more important.

Deadline

From the receipt of this test, you will have ​*a week*​ to complete it. If you are not able to start working on this test on the day you receive it, please send us an email ​*as soon as possible*​ to let us know when you will start.

How to submit

Once finished the test, please compress your files (​*code* and​*instructions* on how to run) in a zip file, and email it back to the email address you received it from. Feel free to either include all your writing, code, and plots in a Jupyter Notebook, or write .py​ ​files accompanied by ​.txt​ or ​.md​ files, or any other medium you find effective.

For any question or clarification, feel free to write an email back to the email address you received it from.

We hope you do well!