**DATA TEST**

**Duration:** Half-day.

**Objective:** Test the candidate ability to industrialize, no data science skills or clustering results would be asses.

**Data:** Static geographical information of CityBike‘s stations in Brisbane (“*Brisbane\_CityBike.json*”)

**Instructions:**

Propose a code to perform a clustering based on either the location or characteristics of bike stations. Any clustering library can be used no assessment is done on the choice of the library. The data is a subset provided by the business.

It should be developed as close as a real industrialized process. In production this code should be launched daily on 10 Go of data, the choice of the platform is up to you it can be run on a Spark cluster or Kubernetes Cluster or Azure function.

**Languages:** Either Scala or Python3 (or C# if on Azure function).

**Deliveries:** The delivery of this test should contain:

* Code: A zip file containing all your code or a link to a repository
* Readme: How the job is launched, and short description of your program
* Output: directory with the clustering result
* Backlog: The User Stories & Tasks (done or to do) related to industrializing this code.