

DATA SCIENCE ENGINEER ASSESSMENT

For this assessment, the candidate is expected to **ATTEMPT** both questions.

SECTION A (Data Ingestion and API Communication) - Python

1. We want to ingest github data of techies or github users in the Machine learning Field in the following locations on github. (**Lagos, Nigeria, Rwanda**). It is expected that you use the **github search for users API** for this task. The final result for this task should be a dataframe that contains details of Github Users who do **Machine Learning** from the three location tags indicated above. A sample search Keyword for users would be like “**Machine Learning location: Lagos**”
 - a. The expected data points to be ingested or seen in the data frame include name, login/username, Bio, Blog, Company, email, followers(count of followers), following(count of following), id, url, and location.)

HINTS - Check out using **Pygithub** or **Github3.py** Python Packages for this task which can be installed using pip. As these packages can be used to interact with Github APIs directly.

SECTION B (Data Storage and Reporting) – SKLEARN/SQL

1. Make a **simple cluster analysis using mean shift** clustering and give a **brief explanation** on what each cluster might mean with this data.
2. **Create a writer using PyMySQL** or any other **SQL** client you might be more comfortable with, to **write this data into a database**.

Please make a pull request to this repository containing you IPYNB notebook and any other rough sheet of document you think might be relevant.