**Team Formation**: Teams have a limit of 4 members. You are free to form your own team. Participants, however, need not form a team: those without a team will be placed into one based on complementary skill sets.

**Data:** The primary dataset for the Datathon will be hosted as Google BigQuery tables. We’ve included sample credentials to access BigQuery at the bottom of this document. We strongly encourage you to use them and familiarize yourself with accessing data from BigQuery in your local computation environments, and with running queries in the BigQuery environment.

BigQuery provides a SQL interface and supports the most common SQL operations.

BigQuery can be accessed using the web interface or through 3rd party libraries. Sample starter code is provided in python (pandas) and R.

On the day of the event you will be provided credentials (different from the one below) to a google account that has full access to the Bigquery tables and the projectid and table name.

**Required Skills**: The skills a team must possess include

1. Basic SQL (Including aggregate and window functions)
2. Programming (To conduct Data Analysis as well as API Access)
3. Familiarity with Google Cloud ecosystem, particularly BigQuery.
4. Data Visualization and Basic Machine Learning Concepts
5. Business Acumen, Creativity & Presentation Skills

**Credentials:**

Username: datathon01@gmail.com

Password: correlationone

**Advanced Preparation**: While not required, we would recommend that you familiarize yourself with [Quandl](https://www.quandl.com/). A large number of datasets on Quandl are available for free, and Quandl has a well documented [API](https://www.quandl.com/docs/api) for pulling data. Please note that there exist tools for fetching data from Quandl into R, Python, documentation for which can be found on Quandl’s homepage.