**Steps for Analytics:**

IMPORTANT: In addition to my dataset, my analytics will also require a secondary dataset. Please follow the data ingest and ETL instructions under “/covid-19” to obtain and reproduce the national covid-19 dataset.

1. Create Hive tables for analytics with the two datasets. First login to Hive using beeline command line prompt, then run the following commands into your terminal:
   1. use NetID;
   2. create external table healthy\_diet (country string, alcohol\_percentage float, protein\_percentage float, grains\_percentage float, fruits\_percentage float, vegetables\_percentage float, dairy\_percentage float, fats\_percentage float, sugars\_percentage float, other\_percentage float, products\_ratio float, fats\_ratio float, obesity\_rate float, undernourished\_rate float) row format delimited fields terminated by ',' location '/user/NetID/project/’;
   3. create external table covid (country string, confirmed int, death int, recovered int, active int, death\_rate double, recovered\_rate double, one\_week\_increase int, one\_week\_increase\_rate double)

row format delimited fields terminated by ‘,’

location ‘/user/NetID/project’;

1. Open Tableau Desktop (download if necessary), and download the Hive ODBC driver for Cloudera Hadoop.
2. Log out from NYU Dumbo. Re-login with port forwarding using the command “ssh -L 4483:babar.es.its.nyu.edu:10000 NetID@dumbo.hpc.nyu.edu” then connect to Hadoop using the Cloudera Hadoop.

Graphical user interface, application

Description automatically generated

1. Select NetID in “Schema” using Exact Match.
2. We are looking at the correlations between each food group in the healthy\_diet table against the confirmed COVID-19 cases and recovery rate. Repeat the following steps for the following 5 columns: alcohol percentage, obesity rate, undernourished rate, vegetables percentages, and animal product to vegetable product ratio.
   1. Click “New Custom SQL” to import data source.
   2. Use the following custom query: “select t1.country, t1.alcohol\_percentage, t2.confirmed, t2.death, t2.recovered, t2.active, t2.recovered\_rate from NetID.healthy\_diet t1 inner join NetID.covid t2 on t1.country = t2.country” to create a data relationship.
   3. Go to the Sheet 1 tab to create a symbol map.
   4. Plot a symbol map using the “Show Me” tool. Set the following details for your symbol map:
      1. Details: Country on the map
      2. Color: Alcohol percentage
      3. Size: Confirmed cases.
   5. Adjust color schemes from red to green for maximum contrasting. In the example under “/screenshots/screenshots\_WilliamHuang” you can see that alcohol percentage is redder when higher and greener when lower.
   6. Change the size dimension to the recovery\_rate field for a similar symbol map that displays alcoholic percentages of diet to COVID-19 recovery rates. You can similarly play with the various COVID-19 dimensions against alcohol percentages for similar mappings.