Uber Data Analysis

* Create account on console.cloud.google.com
* Cloud Storage for storing the file
* Create bucket
* Switch to the fine-graine access from the permission tab to make the file public
* Click to 3 dots and click edit access > add entry > public, allusers, readers

To create MAZE inside virtual machine

* Navigate to compute engine from search bar.
* Enable compute engine
* Click on create instance > choose standard with 4GPU and 16GB memory
* Click ssh and once connection is done:
* sudo apt-get update -y
* sudo apt-get install python3-distutils
* sudo apt-get install python3-apt
* sudo apt-get install wget
* wget https://bootstrap.pypa.io/get-pip.py
* sudo python3 get-pip.py
* sudo pip3 install pandas
* to install mage > navigate to maze data tool on search bar > <https://www.mage.ai> > go to github repo > sudo pip3 install mage-ai
* start mage project > mage start project\_name
* run the vm from external 34.105.219.245:port > 34.105.219.245:6789
* if its not running > we need to allow this port from our instance so > click nic in instances > click firewall > create firewall rule > give details > targets to all instances to the network > give ip range > 0.0.0.0/0 > check tcp and give port : 6789 > mage ui opens
* create project on mage
* click new > click standard (batch)
* load data > click data loader > python > api > give name
* copy the uber\_data url from cloud storage and paste into the data loaders api url
* click play button to run the code
* click Transformer > python > generic > give name to the transformer
* inside Transformer code > add to top : import pandas as pd
* now copy all the transformation data from jupyter notebook from local that we did
* after the success in Transformer, now go to Data Exporter for BigQuery Data. The data are transmitted as data dictionary for processing on BigQuery.
* For connecting to BigQuery, we must configure the yaml configuration. Navigate to utils > io\_config.yaml > Provide details to GOOGLE\_SERVICE\_ACC\_KEY
* To gate the GOOGLE\_SERVICE\_ACC\_KEY, naviagate to API & Services > credentials > create credentials > service account > give name to service account name > give access to BigQuery Admin > Click done.
* Click on created account link > click key tab > add key > create new key > click json. A file with details will download. Using the details in file, give all the details to io\_config.yaml and save the yaml file
* Navigate to BigQuery from search bar
* Create new dataset from … vertical dots with displayed with project name.
* Copy the dataset name and paste it into the Data Exporter, table\_id. table\_id consists with project-name\_dataset\_name\_table\_name. Deal with it correctly
* Covert the previously json data that is passed from uber transformer to big query to DataFrame using DataFrame(data[‘fact\_table’])
* If no module name ‘google.clooud’ occurred, open the vm instance into another terminal and type sudo pip3 install google-cloud and also, sudo pip3 install google-cloud-bigquery
* Now, run the Data Exporter BigQuery. This time the table creation is successful. Check if the table is created or not from fact-table. Navigate to BigQuery > Go to earlier created dataset > click fact\_table > click preview > all the data are populated.
* Create the analytical table selecting only the required columns using the select and join.
* Go to looker studio to build the dashboard
* Click blank report > BigQuery > Authorize (for the first time) > add to report
* Design the report as desire