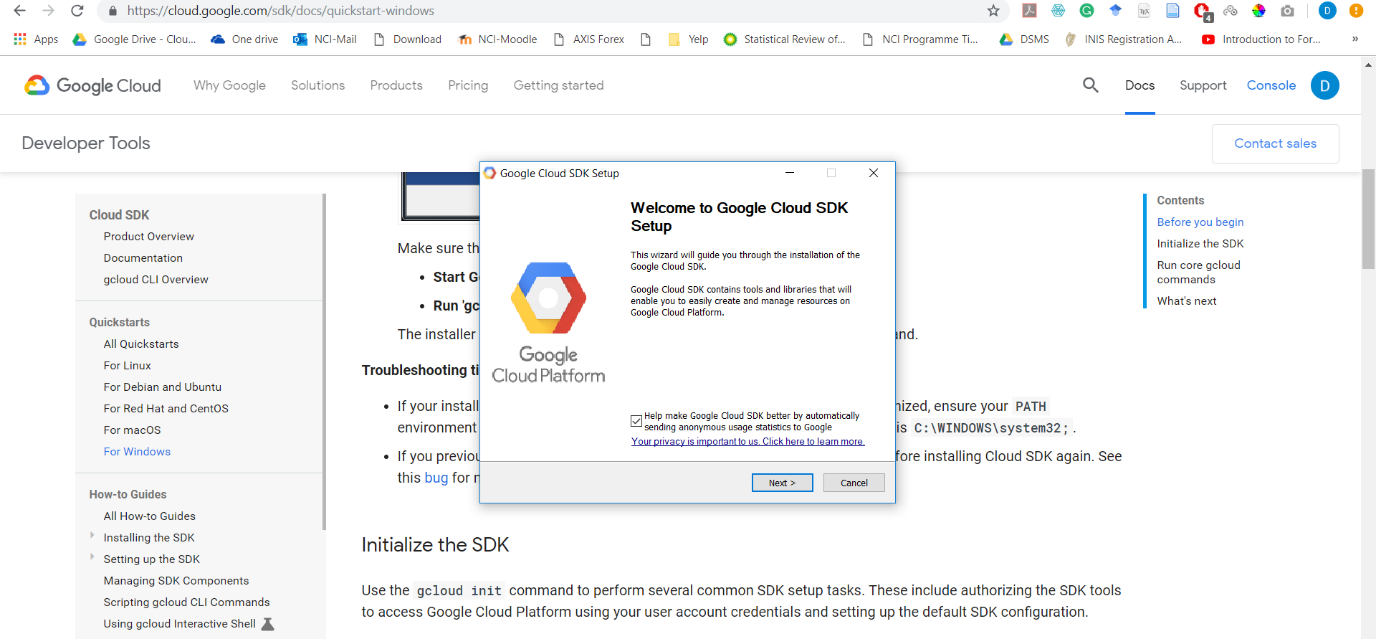
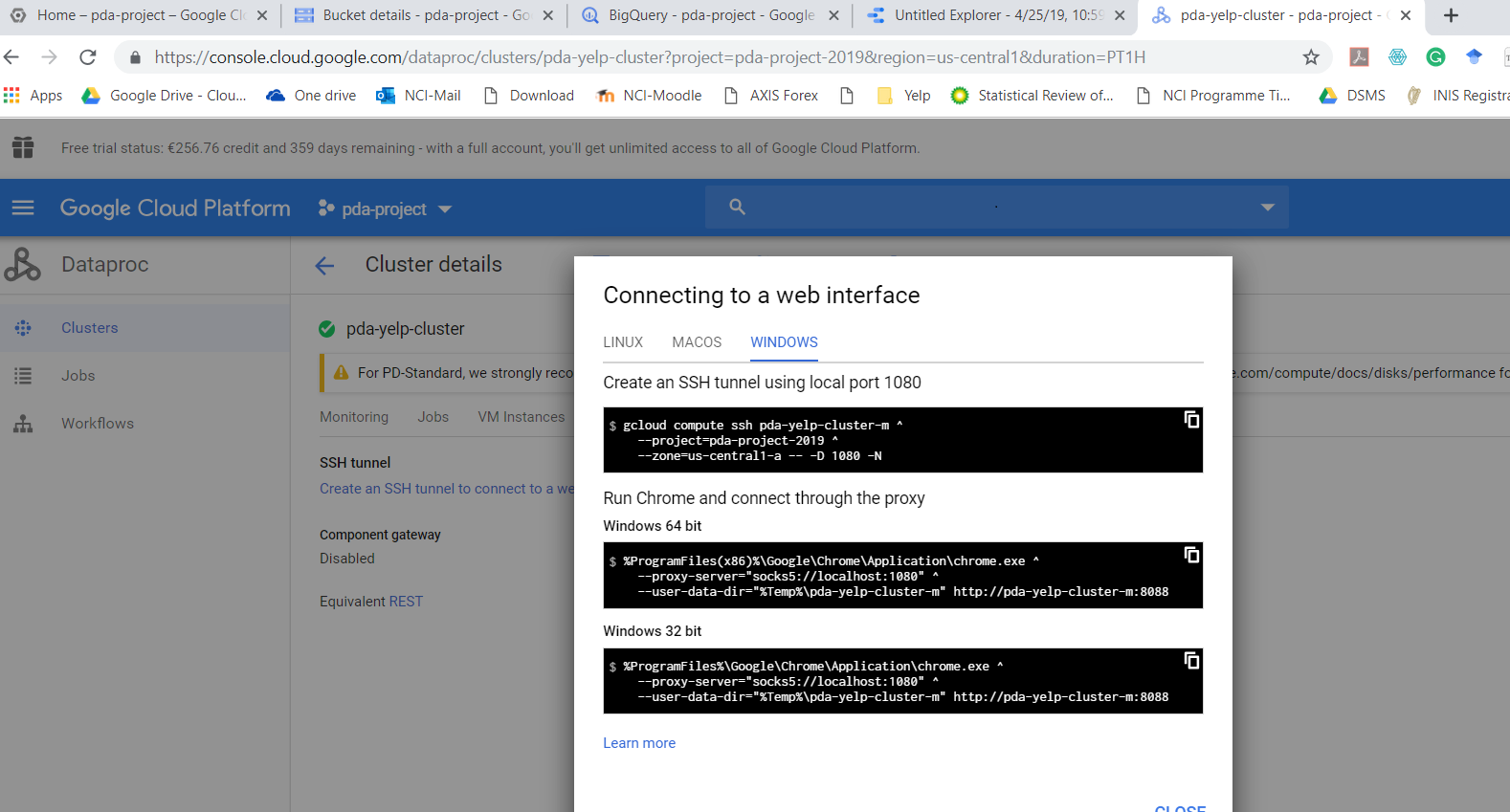
Steps followed for Pyspark on Google cloud platform

Task 1 in pyspark : find out popular dishes using dussines.csv and tip.csv of yelp dataset

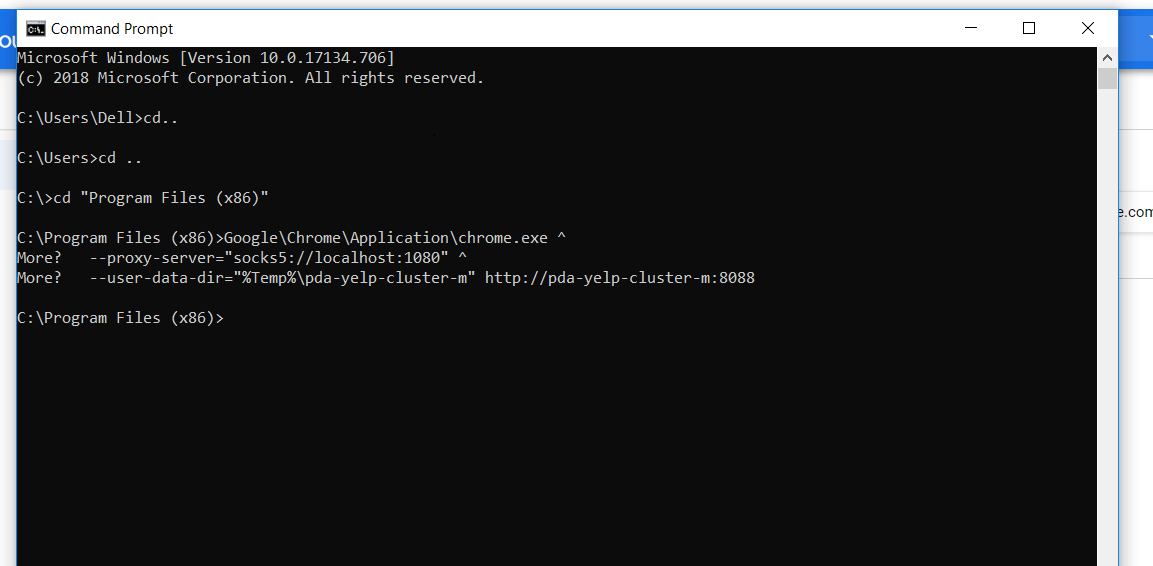
1. Cluster with anaconda and Jupiter, Bucket creation already done in gcp setup
2. Install GCP SDK



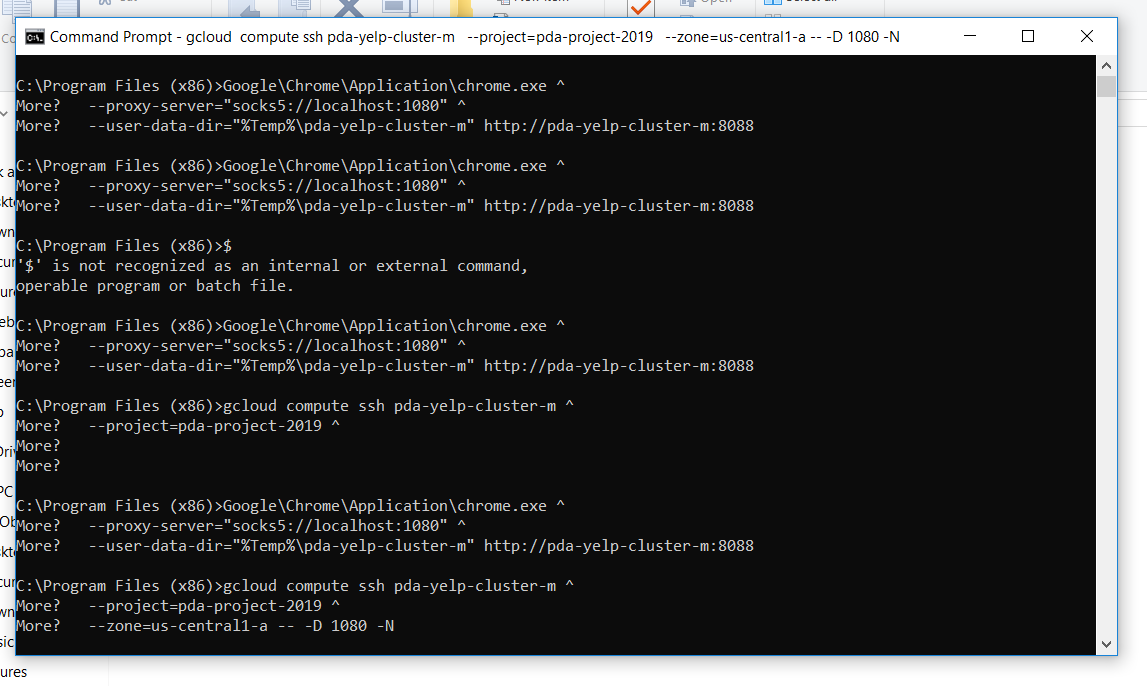
1. Connecting to web interface of cluster for pyspqrk coding in jyupter



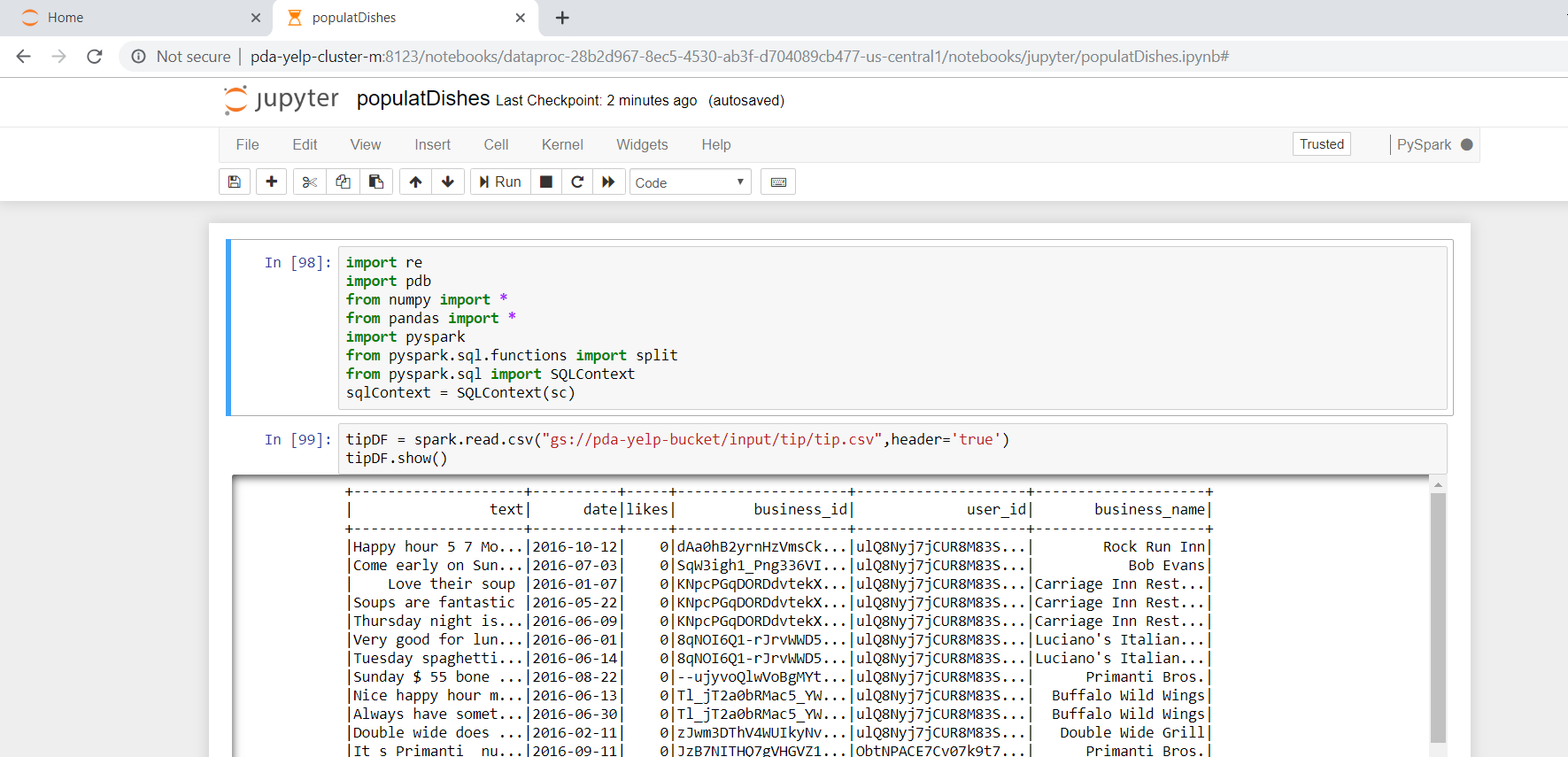
1. Establish ssh to cluster



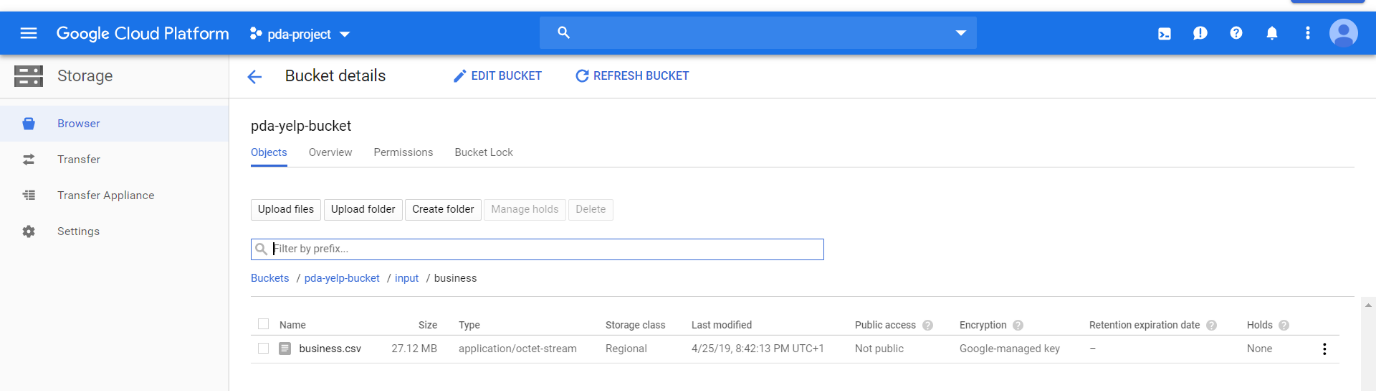
1. Open web interface

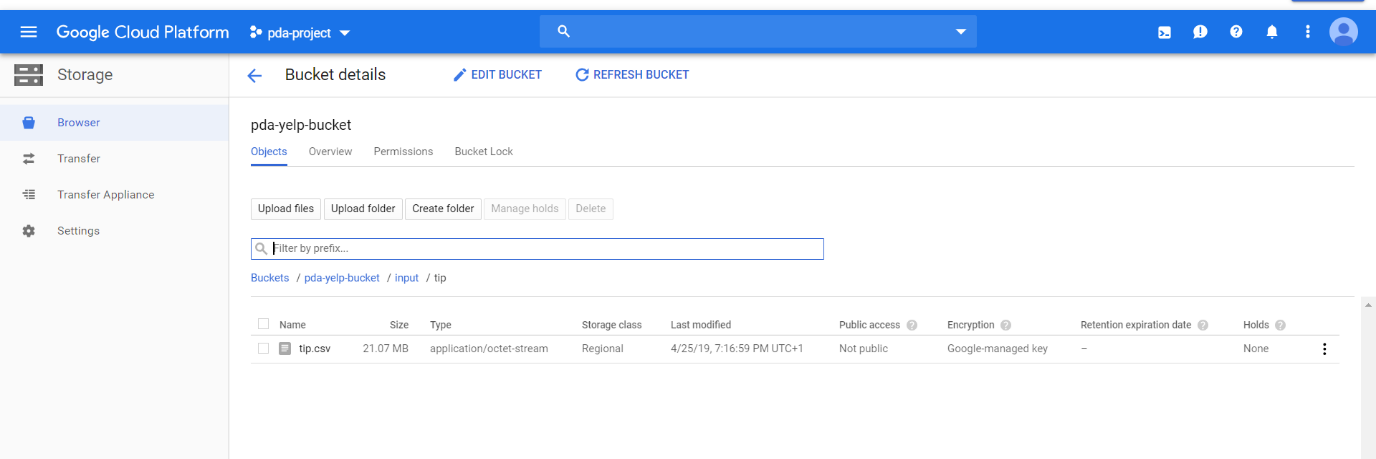


1. Coding on jyputer

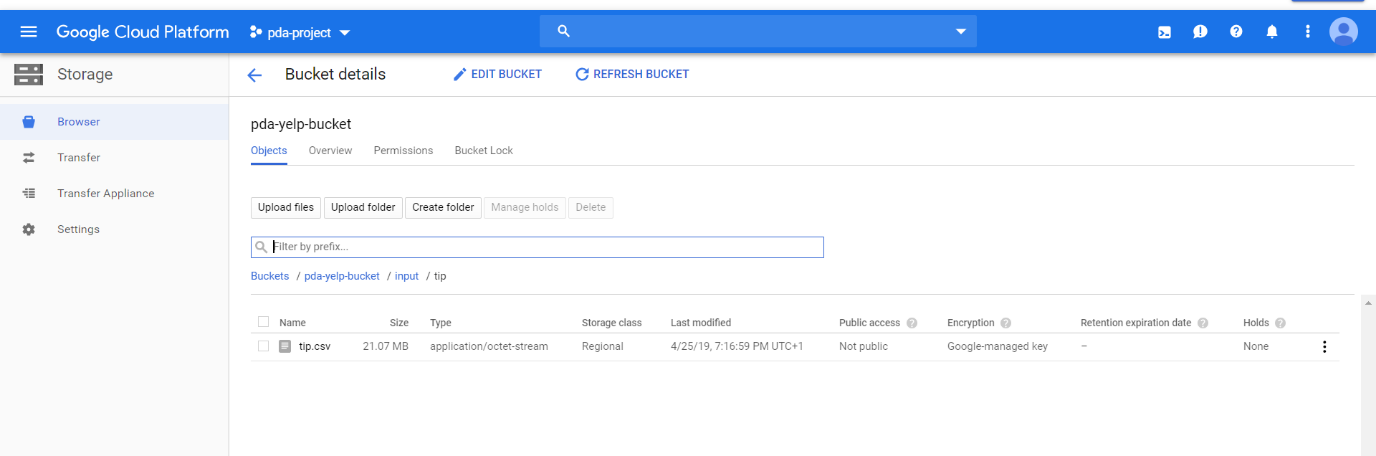


1. Input files on bucket

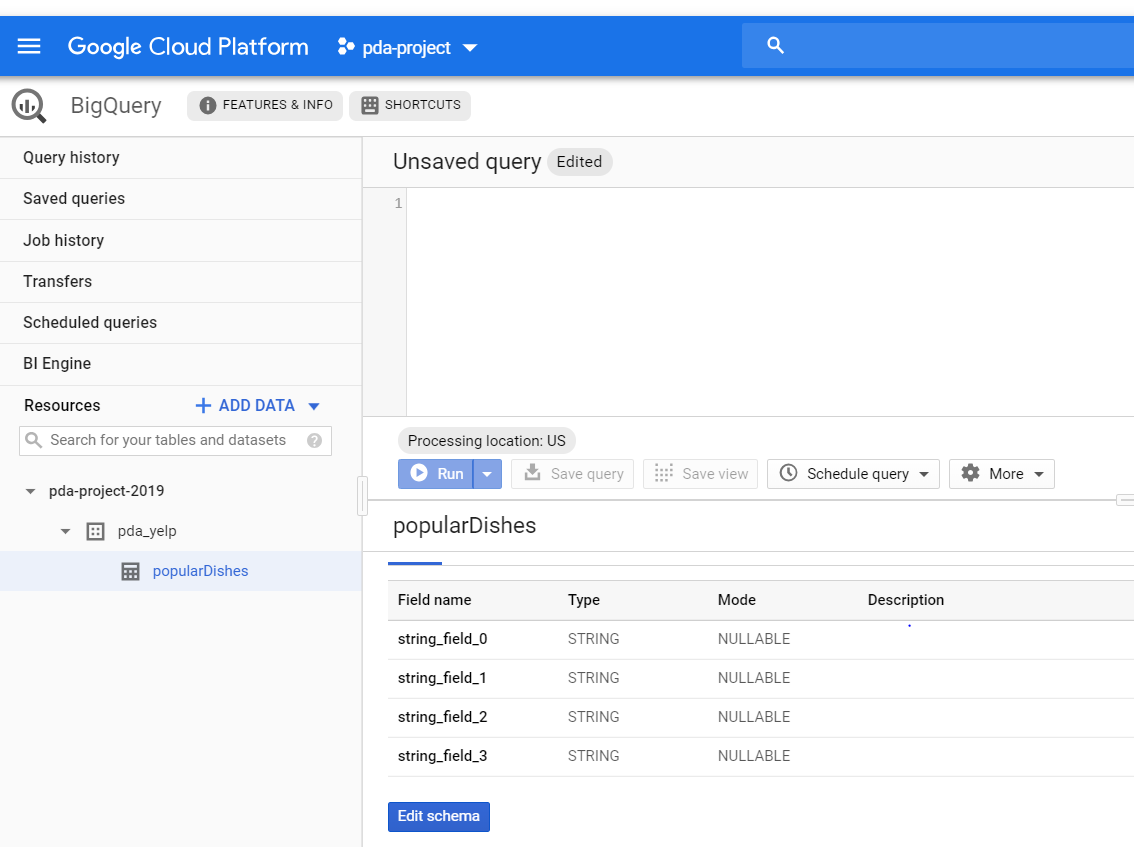




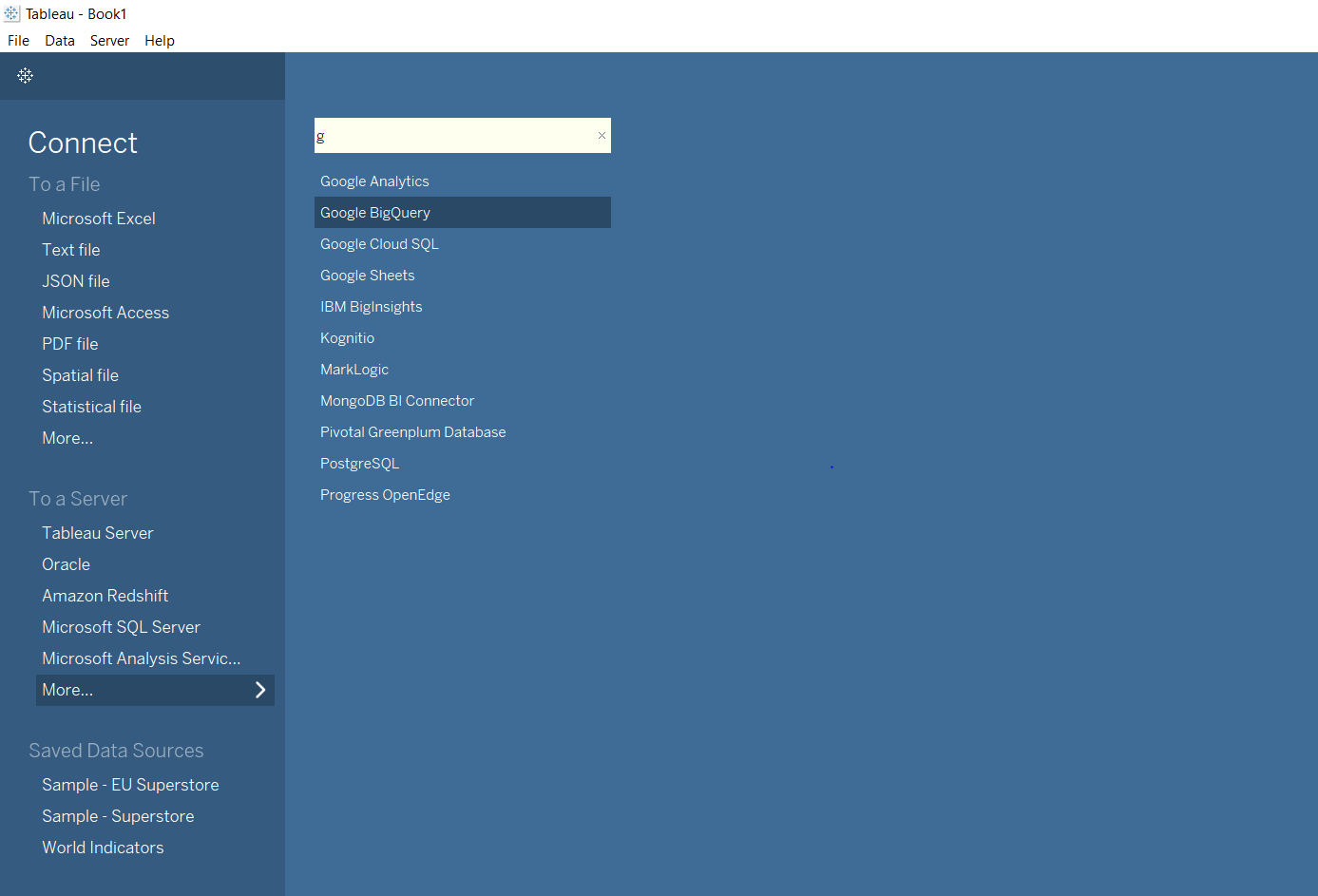
1. Output file generated on bucket

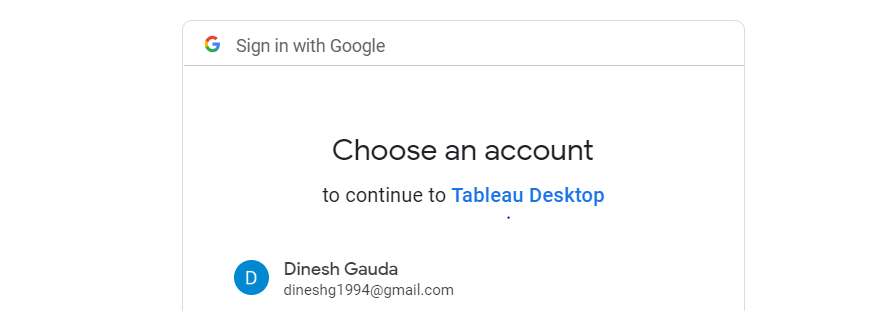


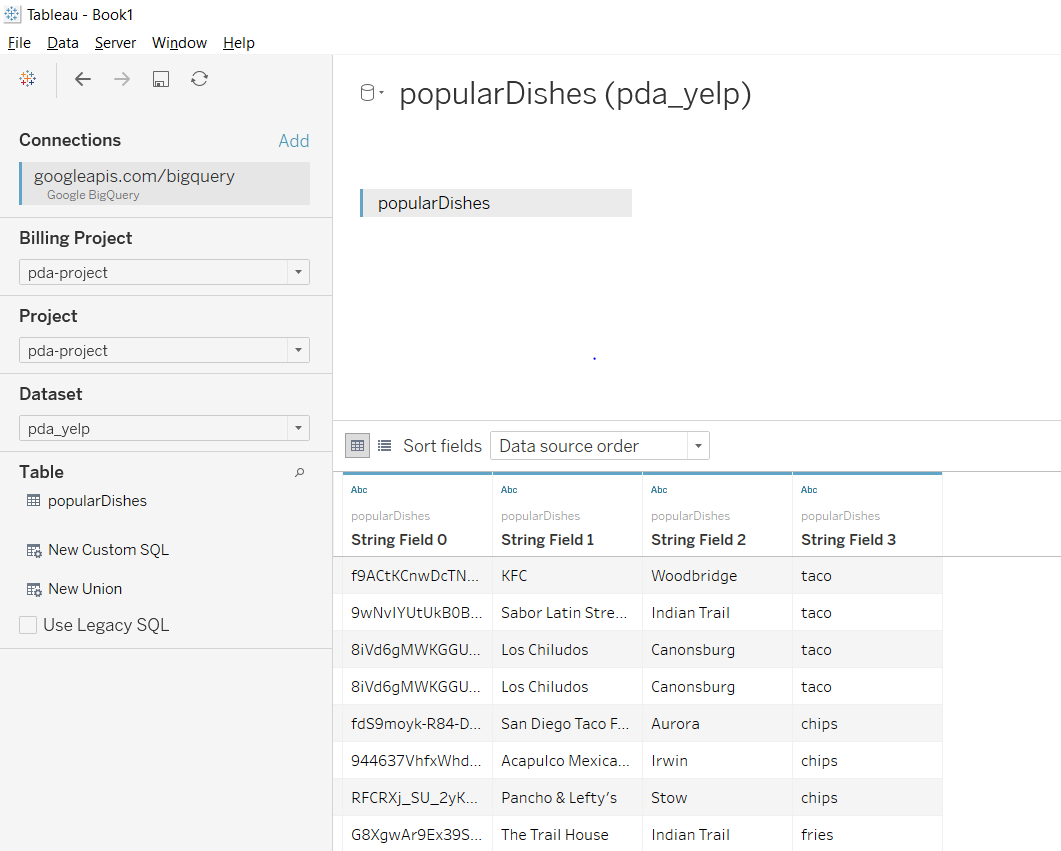
1. create table in biq query



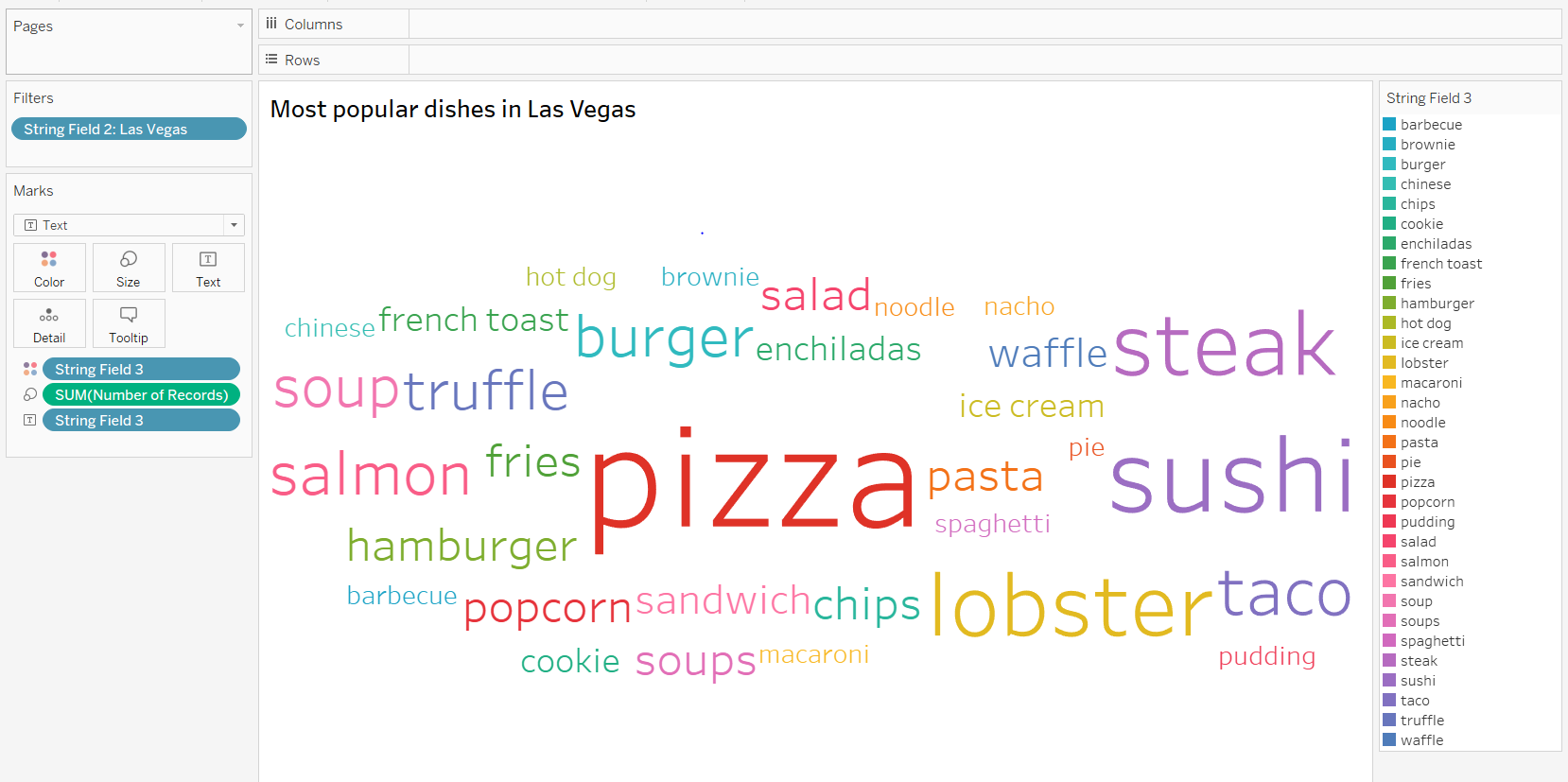
1. connect tableau with google biq query



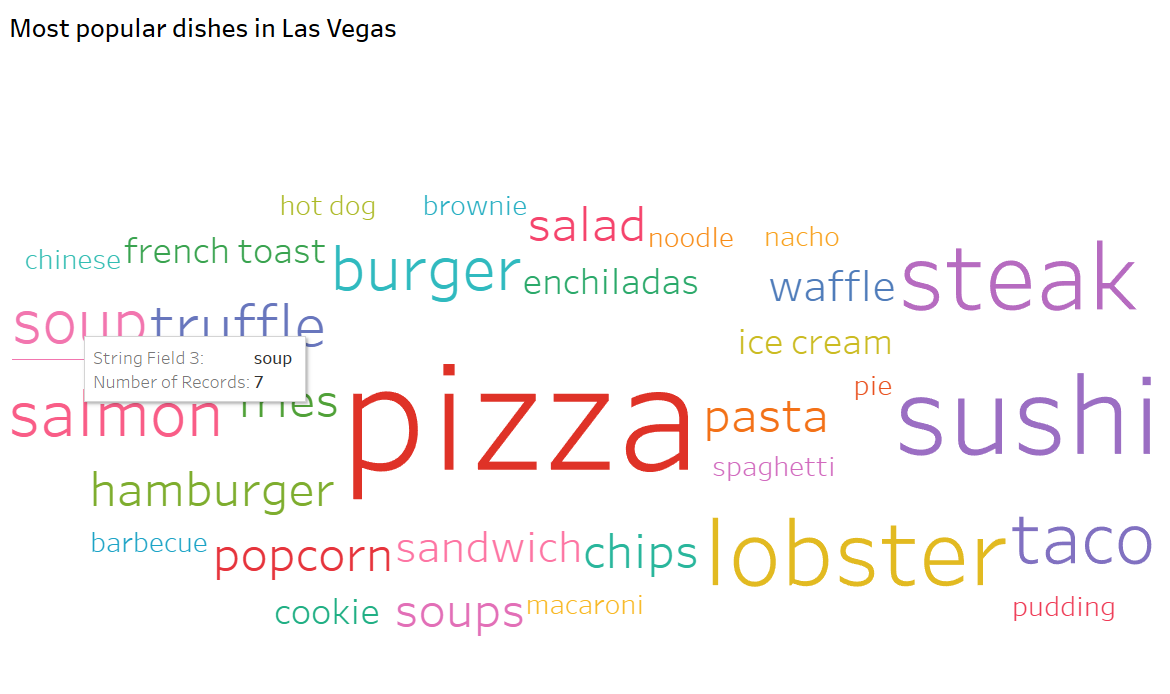




1. creating visualization



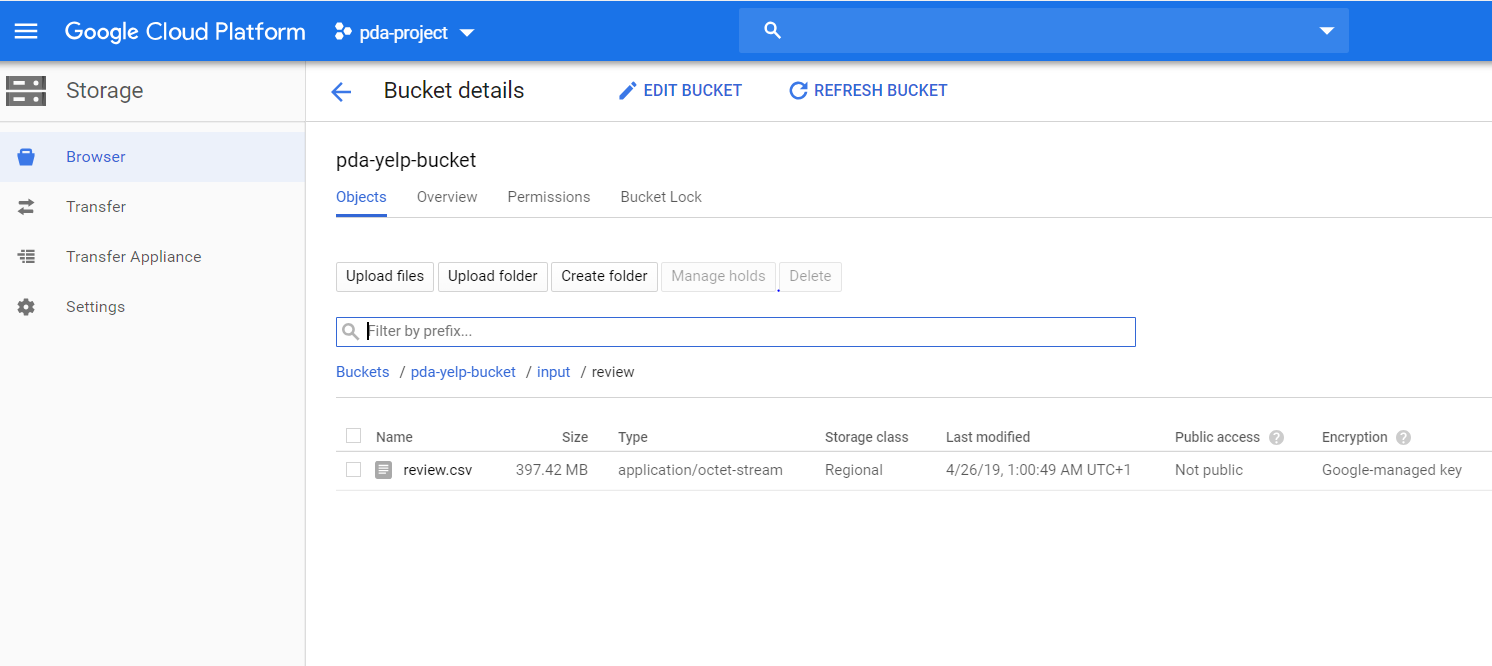
1. Final visualisation

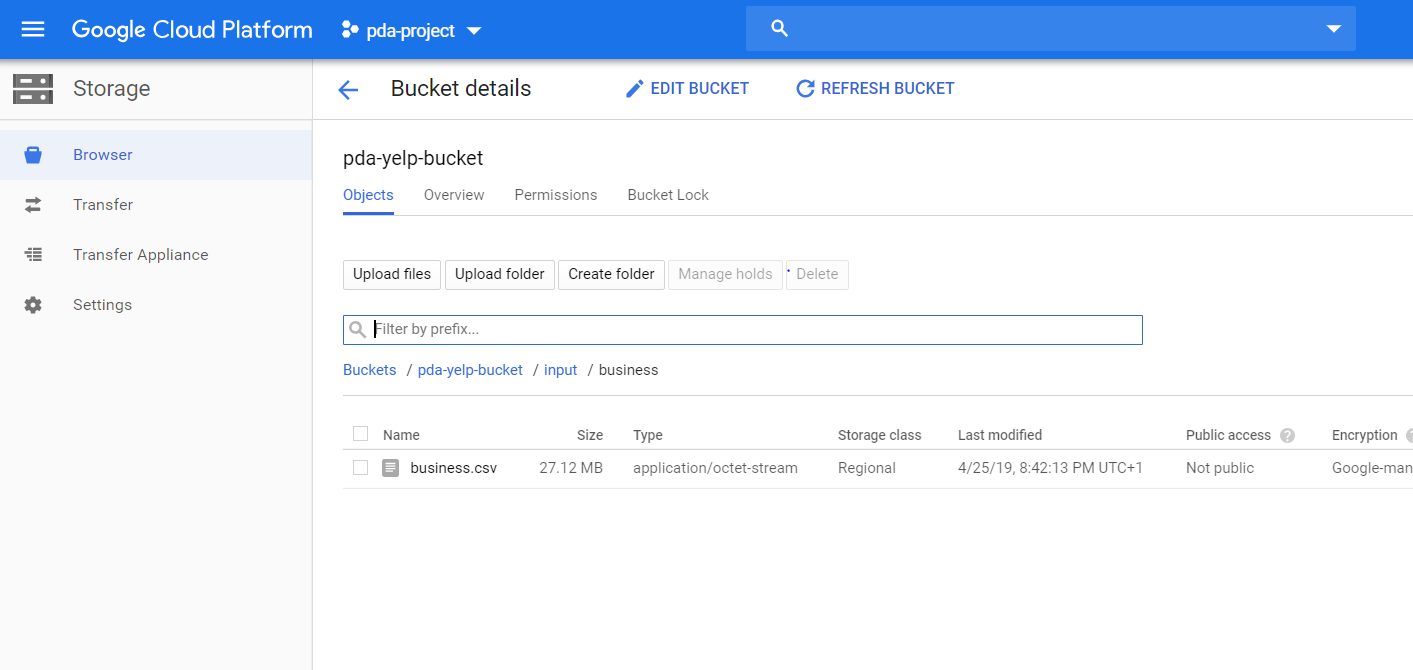


Task 2 in pyspark : find Sentiment score of reviews.

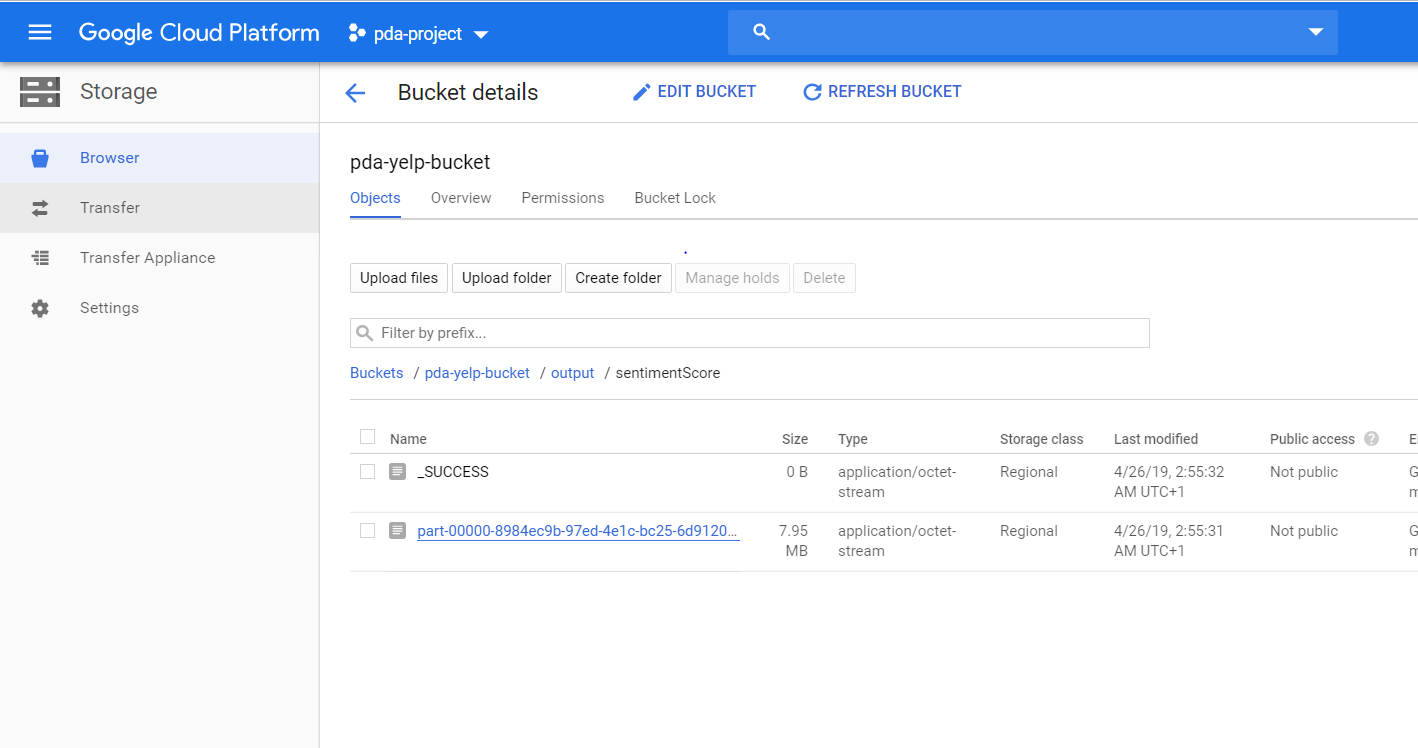
Steps followed:

1. Same as above steps 1-5
2. Coding in jyupter to find sentiment score
3. Input files of bucket





1. Output file generated on bucket



1. follow steps 9-11 of pyspark task 1
2. Final visualisation

