**Read Me:**

**Background:**

* We designed this data pipeline and reusable analysis structure to help restaurants identify where they should open up a new restaurant with several factors in mind
  + Societal factors in the surrounding region
    - Income data, Crime data
  + Competition in a specific area
    - Consumer preferences, exact location of competition, competition’s category

**General Solution Steps:**

* Ingest data via R, Python, and direct click download
* Transform data using the Open Refine Scripts
* GCP storage through MySQL workbench (where additional transformation occurs)
  + Optional processing/storage in Neo4J and MongoDB
* Manipulate in R and Tableau to produce analyses

**Solution Files Present here:**

* Sample data sets: Yelp, Income, Crime, Austin Neighborhoods, and Unique ZIP Codes
* Scripts: SQL, R, Python, OpenRefine
* Files: Tableau workbook, Final PPT

**Specifically How to Replicate:**

* Create a GCP instance and connect your MySQL workbench to it
* Ingest the Yelp.csv file into your MySQL through the File Wizard
* Copy, paste, and run the SQL scripts in this order:
  + Script to create Income
  + Scripts to Create Income and Crime tables
  + Script to Create Unique ZIP
  + Split Yelp Data
  + Scripts for Indexes and foreign keys
  + OLAP Model
* Connect your GCP to Tableau and replicate the reports