**ETL Project - LA Restaurant Violations vs Yelp Rating**

**Data Sets**

Restaurant and Market Health violations and Inspection - From: Kaggle - LA Health Inspection Report

**Objectives**

* **E:** Reading csv files for inspections and violations. Scrapped Yelp for restaurant rating.
* **T:** Iterate inspections and insert violations into inspection
* **L:** Store data in Json form to load into Mongo DB

**Summary**

When we first started out it was difficult to nail down how we were going to Extract our data. Eventually, we settled on the idea of grabbing 2 CSV files from Kaggle. These files are records of restaurant health inspections in LA.

Our first challenging goal was to Transform the data into a convenient format for analysis. With these two files separate it would be difficult for anyone to gather useful conclusions from the Datasets. We used Pandas to consolidate every individual violation of a restaurant. Then, it is nested inside a single column of the original master DataFrame.



Our goal at this point was to query yelp for the review data on each restaurant inspected. After this is retrieved its then added onto the master DataFrame. The motivation behind this is to give the Analyst more to work with.

Finally, we converted the DataFrame into a Mongo DB friendly JSON format. We simply looped the inserts into the Mongo connection.

**Outputs**



