The Restaurant Bubble

Utilizing Yelp API to understand restaurant closings in Los Angeles

**Restaurant Industry Overview**

In the US, the industry employees 14.7M people (10% of the workforce). There are over 1M restaurant locations in the US, with 7% just in California.

Margins for restaurants are slim. With rising labor costs and rent as well as market saturation, restauranteurs need to take a deeper dive into analytics to make smarter decisions regarding new ventures. The entrance of ‘disruptors’ in the industry (like DoorDash, UberEATs) will continue to threaten the industry.

**Executive Summary**

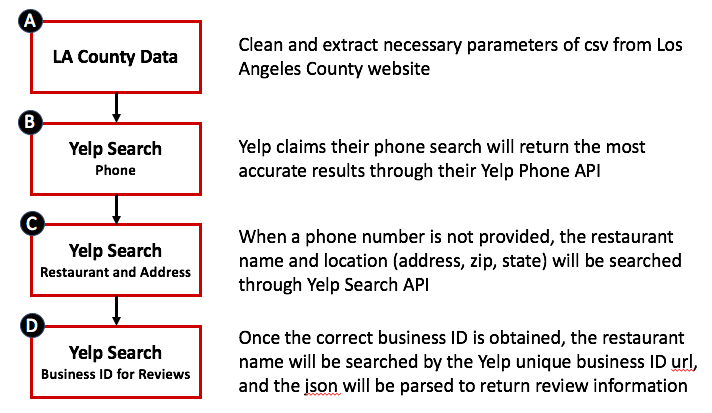
Given the changing landscape,

**Data Acquisition**

*Reference notebooks 1.01, 1.02, 2.01, 2.02*

There are two datasets that I will be working with:

1. LA County Dataset (A)
2. Yelp Dataset (B, C, D)

****

**LA County Data Acquisition Issues**

The most challenging aspect of this problem was data acquisition. Yelp considers data to be clean in a different manner than someone would expect.

* Suite or Store Number: A pound sign will throw off our ‘best match’ Yelp results when we use their API. Regex will help solve some of these issues.
* Multiple Inspections: One restaurant can be inspected multiple times, or have multiple inspections for more than one kitchen. Groupbys and dictionaries will help solve some of these issues.
* Misspellings: A restaurant may have multiple inspections where the name is misspelled. There are some solutions we can implement, but this will be part of the error.
* Non-restaurant inspections: entertainment/sporting venues, airports, grocery stores, are all included in the LA county data. Comprehensive dictionaries can help solve these issues, if they can be identified.

The cleaner the data is, the more successful our Yelp API calls will be.

**Yelp Data Acquisition Issues**