**Servando Portillo – ID: South Carolina**

**CIS 3330**

**CODE 10 – Yelp API Text Analysis**

**DATA RETRIEVAL AND ANALYSIS PLAN**

**List of Insights to focus on:**

* **Value**: Any reviews that indicate the customer’s opinion on pricing, and food quality/appearance.
* **Service:** Any indications of poor service received by the customers.
* **Establishment:** The environment and conditions of the facilities.

**Search Parameters:**

* **Terms** – I am interested in how my national cuisine is faring in El Paso, so searching for keywords such Mexican, Hispanic, Tex-Mex. Would allow me to identify restaurants that are related to those keywords and maybe find repeat results and focus on those establishments.
* **Location** – Since the assignment’s purpose is to simulate what somebody interested in opening a similar business would look out for, it would make sense to focus on a location that is familiar such as El Paso.
* **Sorting** – It has been proven that people are more likely to leave feedback when they’ve had negative experiences, focusing on those negative experiences would allow for valuable insight on what to avoid or correct in terms of business practices. Looking for popular restaurants that have real and quality negative feedback would allow for valuable insight on what to avoid when starting this venture. As such, focusing on the review count will be of use to check what the most popular restaurants are doing wrong.

**How will that information be used?**

It will be simpler to identify recurring occurrences of the events that gave rise to those unfavorable evaluations by filtering the information. Prior to the restaurant's debut, these patterns will make it possible to put in place the appropriate preventative measures or corrective actions. A successful launch will be possible if you can learn from others' mistakes without making your own.

**Text Analysis**

By using sentiment-type text analysis, it will be possible to get the general feeling or sentiment of the reviews that are being processed. This will permit further analysis by allowing one to focus only on the reviews that have been marked with lower scores.

**TEXT INSIGHTS REPORT**

Using ‘vaderSentiment' filtering the reviews was possible, I was able to find just the reviews with a low score by searching for 80 restaurants with the phrase 'Mexican' and three reviews per restaurant due to Yelp API restrictions. The reviews weren't complete because the Yelp API only gives abridged reviews, which made it challenging to analyze some of the evaluations because it was unable to gain their full context. Web scraping would be the ideal solution to this issue, however it is prohibited by their terms of service, so it was not tried.

Using NLTK stop words to remove words that do not contribute too much insight into the review, I was able to gather some information from the reviews:

* **Value**: There were complaints of the food not being up to par with the customer expectations, descriptions of the appearance, texture of the food appeared several times. There are complaints about the price of the food in some instances.
* **Service**: Some occurrences of poor service can be read in the reviews, people being disrespected, long wait times, wrong orders being delivered, lack of utensils, lack of menus, having to make requests several times, long wait times.
* **Establishment**: Some occurrences of establishments being hard to find, improper care and maintenance of the facilities, the environment not being up to par with their expectations, loudness.

**CONCLUSION**

In general, poor food and service were the main topics of complaints. When food is either excessively expensive or of poor quality, people feel disappointed since they expected to get more for their money. Authenticity is another trait that is desired. Another area of opportunity is in the provision of services; it is crucial to ensure that clients are treated with respect and that their needs are met promptly and accurately. Another essential component that clients consider is the location of the business. clients feel more at ease in a business that has a suitable atmosphere, cares about cleanliness, and is easy to get to.

After gaining some experience, I was able to recognize the potential value of text analysis in the hands of an expert. When trying to enter an industry and try to uncover gaps that you could fill that would create a competitive advantage, being able to analyze vast amounts of information and acquire important insights is helpful. There were certain restrictions, though.

I initially tried to get 15 negative reviews from the top 3 restaurants with high review counts and try to see what it would take to fill the gaps that they’re not taking care of. However, only three reviews per restaurant can be obtained with the current access I was given, this made me change my approach and instead got the three most recent reviews from 80 restaurants and filtered them to only get the negative ones as interpreted by the vaderSentiment library. This returned a reduced number of reviews and most of them were truncated by the API, meaning that some of the reviews seem to be negative but with the missing context it’s possible that these are positive reviews overall.

I considered utilizing web scraping techniques to get better insights but found out that it’s prohibited by Yelp to do any web scraping, so no attempts to do so were made.

**Resources**

Valence Aware Dictionary and Sentiment Reasoner - <https://github.com/cjhutto/vaderSentiment>

Natural Language Toolkit - <https://www.nltk.org/>

Yelp API - <https://docs.developer.yelp.com/docs/fusion-intro>

Text Analysis material/code resources were provided during class for educational purposes: CIS 3330 - by Dr. Villacis Calderon