**CSR Consulting - Written Report**

**Project 1 - August, 14, 2021**

**Rich Kirschenheiter, Stefany Lima, Christian A. Reyes**

**Hypothesis:**

Vernon Hills, IL has the potential to host a new high-end restaurant, using data insights to understand competition and underserved market segments.

**Motivating Questions:**

* Who would visit this high-end restaurant?
* What tax bracket (income level) would these guests be a part of?
* What type of food would this new high-end restaurant serve?

**Data Collection Process:**

Data was collected using the following public APIs;

* Census Data
  + [Census quick fact data](https://www.census.gov/quickfacts/fact/table/US/PST045219)
* Yelp Data
  + [Yelp Fusion API, Business Endpoints](https://www.yelp.com/developers/documentation/v3/get_started)

**Data Exploration and Cleanup:**  
 The Census data was a wealth of information in which we had to target by index to filter to relevant topics for the scope of this project. The CSV file did import NaN columns that also required dropping. Then the formatting had to be replaced to drop %, commas and dollar signs to allow for moving from object to float.

The Yelp Fusion API is a public API that provides a range of consumer and business data. We explored their Business Endpoints data to collect restaurant information (business name, location, price, review data, delivery services). The data could be pulled into a .json file, manipulated in the Jupyter Notebook environment, and saved into .csv files. Some businesses did not have certain data points such as price and delivery services, so we had to account for these errors in the code by using ‘?’ as a placeholder in the pandas dataframes.

**Data Analysis:**

The Census Data reinforced our decision with 9,735 Households with a Median income of $104,199. In addition to this the four surrounding cities of Lake Forest, Mundelein, Lincolnshire, and Libertyville share similar demographics. In addition, on average the households across these cities have a 90%+ households with a broadband internet subscription.

Yelp Fusion API returns a maximum of 1000 data points before returning an error. We set our location radius to approximately 2.5 miles from the Vernon Hills Village Hall, a central location in the town, which returned 217 businesses. Examining this data revealed the type of restaurants located in our selected area. We examined the average and total ratings for all restaurants in this area, to determine the online activity for our client base. We intended to look at price range averages for each category to further support census income data, but not all businesses displayed a price. 

**Conclusions:**

The Census data reaffirmed that Vernon Hills not only was an optimal location for a high end restaurant with a large well off surrounding community. Due to the high adoption rate of broadband in our potential customer base, special attention is required to our internet/website exposure to ensure high end experience.  

The business data obtained from Yelp Fusion API reveals an underserved market segment in the restaurant industry of Vernon Hills, IL. We saw a significant representation of businesses categorized as groceries, Mexican, Chinese, pizza, and burgers. Only one business was categorized as steak and another as seafood. We observed that most restaurants, 25.8% , received a 3.5/5 star rating. This reveals the potential for a high-end restaurant that serves steaks and seafood.

**Limitations:**

The popularity of restaurants was determined by the rating it received on Yelp. However, it is not accurate to compare a highly rated restaurant with a low number of reviews received versus a restaurant with the same rating but a higher number of reviews.

The Yelp API pulled restaurants under categories that oftentimes overlapped with each other. For example, a restaurant could fall under the “burgers” or “hotdogs” or even “tradamerican”. We would have liked more time to explore grouping these categories to paint a more accurate picture. One of our initial questions dealt with price range. However, the API pull would not return price values [$, $$, $$$] so we had to remove them. It would have been interesting to compare price with category for correlation.

We also acknowledge that not all food establishments are represented on Yelp. They may not have a Yelp profile or respond to reviews. We would have liked to further analyze demographic information such as age group and race to determine if certain categories were preferred.

**Implications:**

Based on Census demographic and income data, as well as Yelp business and consumer data, we conclude that Vernon Hills, IL can be home for a successful high-end steakhouse restaurant. With more time and resources, our data analysis can be expanded to create a superior report to garner investor interest.