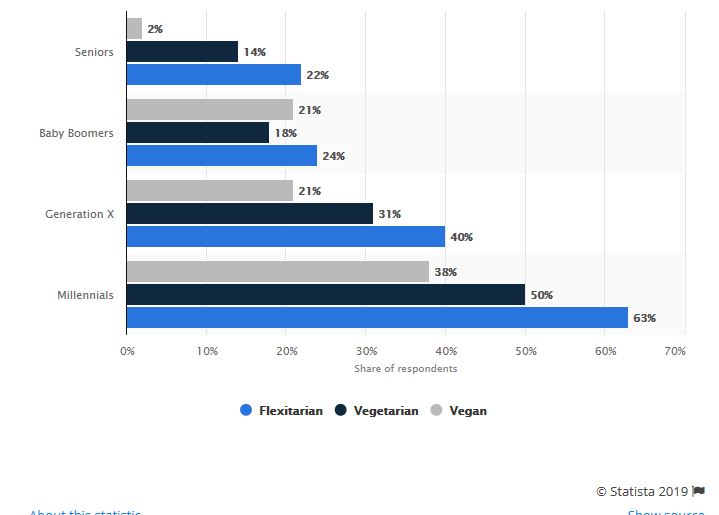
Lubna Jamal

Capstone Project for IBM Professional Certificate

**Introduction:**

**Background:** According to a [2018 survey](https://www.statista.com/statistics/875526/share-alternative-diet-us-generation/), 50% of millennials in the United States opt for a vegetarian diet, and 38% prefer to go completely vegan. The graph below shows the further delineation by consumer group.

****

As is evident, there is a big market for opening a vegan/vegetarian restaurant as more consumers are opting for meat-free dishes. Flexitarians are defined as those that mostly adopt a plant-based diet with occasional meat consumption.

**Problem:**

A client wants to invest money to set up a vegan/vegetarian restaurant in a US metro area. They would like to know some statistical information with regards to how many vegetarian restaurants currently exist there. They would like to know how many restaurants in major metro areas offer vegetarian meals. Similarly, how many offer vegan options. In order to cater to the most popular demand, they would also like to analyze which cuisines are served up at most vegetarian restaurants. This would enable them to cater their menu according to market demand for vegetarian/vegan options.

**Target Audience and Interest:**

The restaurant industry in the United States has seen an increase in sales and profit-margins. The labor costs are mostly holding steady, and that coupled with the fact the most American like to eat out frequently, have signaled profitable business for the restaurant industry. However, there are a huge number of restaurants that already exist. In order for them to reap the profits, a restaurant has to set itself apart from the fierce competition. Hence, being on the forefront of the plant-based diet/vegan trend will prove to be a good option for opening a new restaurant. **Hence, the target audience for this project would be any restaurant entrepreneur who would like to cater to the demand of a vegan/vegetarian restaurant in a major US metropolis.**

**Data Acquisition**:

A subset of the data set, outlining 18,000 vegetarian/vegan restaurants in the United States, has been acquired from [Kaggle](https://www.kaggle.com/datafiniti/vegetarian-vegan-restaurants). This will be used to analyse Furthermore, Foursquare API data will be used to extract trending vegan/vegetarian restaurants in one of the top shortlisted metro areas to scout out the competition.

**Data explanation and Cleaning:**

The data will be cleaned by dropping some irrelevant columns from the dataframe e.g.Date added, updated, facebook url etc. The final dateframe will contain the following column headers. Id, name,categories, city, cuisines, latitude, longitude, priceRangeMin,priceRangeMax, province, date opened.

There are some columns with missing data. The missing values for price ranges will be replaced with the mean prices of those columns.

**Data analysis and Visualization Plan:**

* A bar chart of date opened to number of restaurants will show a trend in timeline as to the increase of vegetarian options.

Correlation data:This will indicate the maximum and minimum that consumers are willing to pay for a vegetarian dish.

* Based on the information above, the metro with the least amount of vegan/vegetarian options will be shortlisted as the location for starting a new restaurant.
* FourSquare Data will be utilized to find trending restaurants in that metropolis.

**Methodology:**

Data was scraped from the web and put into a pandas dataframe. Then it was cleaned by dropping column variables that were not needed. The shape of the data indicated that there were 1000 entries and 47 columns. This data represented a subset of the bigger dataset of 18,000 entries. The index was reset to the ‘city’column so multicity analysis could be done. The cleaned data was then analyzed through various methods, including the describe(), groupby(), and the agg() to find the min, max and sum of the minimum price range of a vegan menu item. This was done to show prospective investors the pricing baseline for certain vegan items.

Data visualization was done in the form of bar and scatterplots to figure out relationships between different variables using the data analysis done before. Furthermore, a correlation matrix was also plotted. Foursquare API was also processed for the city of NY near the Conrad hotel to see specifically Indian food restaurants (since Indian Cuisine has a diverse range of Vegan and vegetarian options)

**Results:** By plotting the value counts one can see that New York has the biggest concentration of Vegan restaurants. Also, in terms of pricing, the minimum priced vegan menu item in New York is the cheapest, followed by San Diego and then Houston. In the State of New York, Albany offers much cheaper option than New York City, where the average minimum price is around $7 for a vegan item.

**Conclusions:**

New York City offers a good market for opening a new Vegan restaurant since consumer demand is already present. Pricing of menu should be between $7 -$22.

Sources:

Datasets

<https://www.kaggle.com/datafiniti/vegetarian-vegan-restaurants>

<https://www.kaggle.com/pavansanagapati/us-wages-via-zipcode>

<https://www.kaggle.com/paultimothymooney/map-of-restaurants-in-san-francisco>

<https://www.kaggle.com/jafethsm/what-affects-the-ratings>

Articles

<https://www.statista.com/statistics/875526/share-alternative-diet-us-generation/>

https://www.forbes.com/sites/sageworks/2018/01/26/restaurants-margins-are-fatter-but-competition-is-fierce/#2cf584c227f9