Battle of the Cities for Healthy Foods Café Pilot Launch

Toni Krowisz

June 29, 2020

1. Introduction

1.1 Problem Description and Background

One of our clients, ***LiveLong*** , has petitioned our firm to gather and analyze some research for healthy food cafes. They are requesting some recommendations on the best cities in which such cafes have the potential to be profitable and have a positive impact in that community. ***LiveLong' s*** goal is to target two cities for launching their initial health food cafes, and checking the results of revenue and positive impact in the community. The criteria is to target cities that have larger concentrations of venues for healthy choice activities with lowest rate of competition.

1.2 Data Description and Approach

Results from a study posted by Adam McCann, February 10, 2020 provides rankings for the healthiest and most unhealthy cities in the United States. A further analysis is done to make a determination of which cities may be chosen for analysis, according to their total health score rankings.

The top 25 healthiest cities are chosen from this previous study.
Within these cities, we gather Foursquare location data for venues that represent healthy choice activities, i.e. gyms and fitness centers, along with other types of eateries in the area. The healthy choice activities will give insight to the potential market in the neighborhoods. The other types of cafes in the area will give insight to potential competitors.

Of these venues, we gather data to understand visitor ratings. Some data visualizations are presented to give some insights. This information is used to determine a success factor rating of 1 or 0. 1 means that the visitor ratings for a specific venue is equal to or greater than the mean of all ratings. 0 means that the visitor ratings are less than the mean.

The location data for venues that qualify as successful are visually presented on a map.

Based on the nature of venues with positive success factor rating, two locations will be recommended for pilot launches of a healthy foods cafe . It is more likely that people who frequent healthy choice activity venues may be the best market for a healthy foods café.

2. Data acquisition and cleaning

2.1 Data sources

The data set for Healthiest and Unhealthiest cities is obtained from Healthiest and Unhealthiest Cities in America (research by posted by Adam McCann, Feb 10, 2020) is scraped from the webpage. There are a total of 174 cities in this dataset, with score rankings which include a total

health score for each listed city. This score is used to determine which of the 174 cities will be pulled to be part of the analysis. Because we presume that a healthy foods café may fare better in cities that are most health conscious. Location data (latitude and longitude), is gathered for each city, and this location information is used to pull fitness and restaurant venues for each of the 25 cities. These venues are all within a 5 mile radius of the location points used for each city. This represents the sample of data that will be analyzed to make recommendations.

2.2 Data cleaning

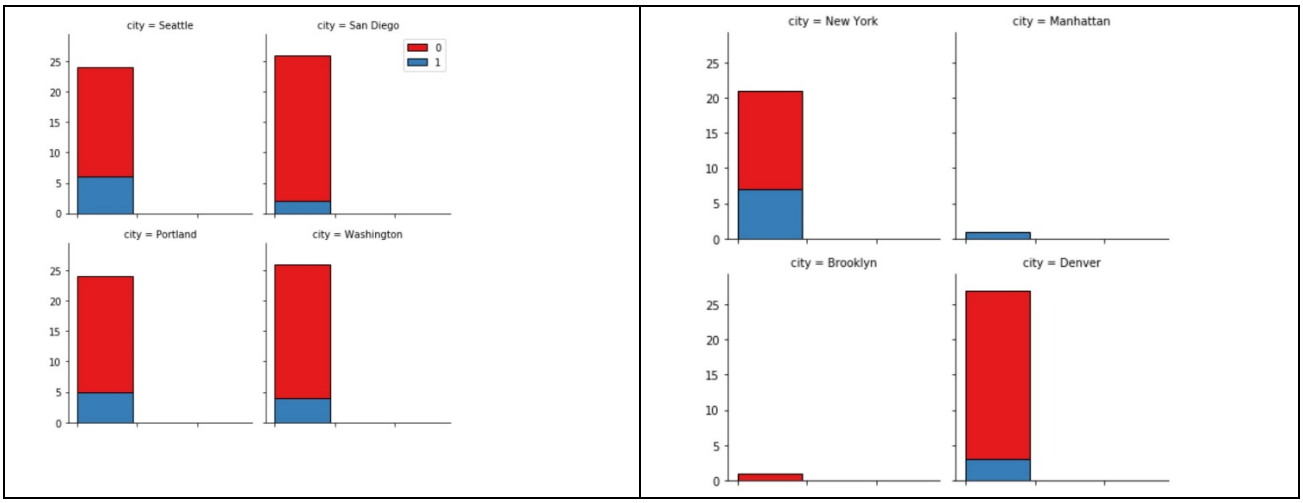
Once the data is pulled from the foursquare queries:

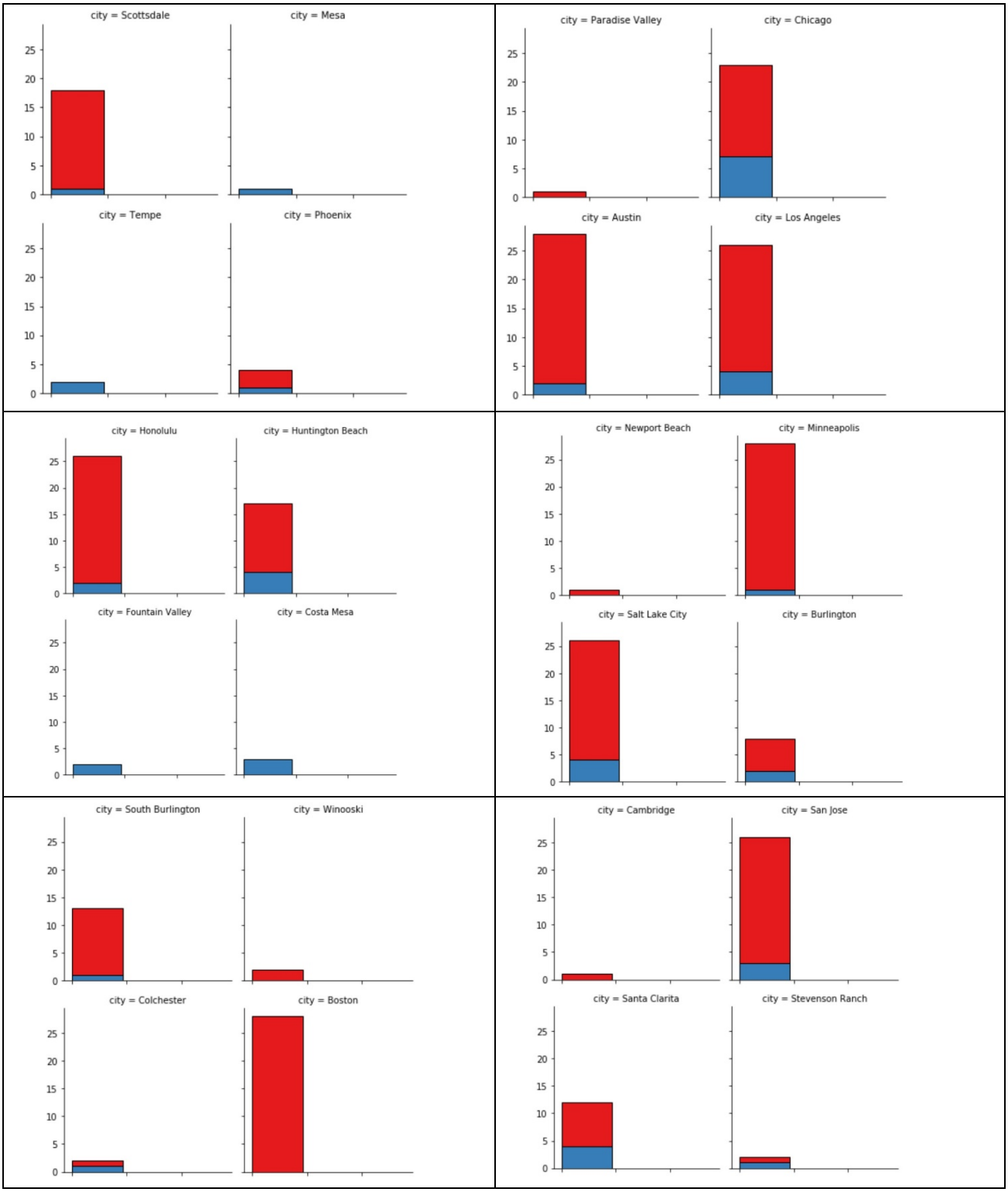
- Unwanted columns are filtered from the data.
- For cities that are near the country's border's, there may be venues picked up in the data that do not reside in the US. These venues are removed filtered from the datasets. This results in removal of 90 rows representing restaurants, and 79 rows representing fitness venues.
- A 'likes_count' column is added and initialized to value of 0.0. This column is populated via foursquare queries that will extract like counts for each remaining venue.
- A 'success_factor' column is added and initialized to 0. This column is populated by comparing the likes_count column value to the mean value of all likes_counts for each venue. If the likes_count column value is equal to or greater than the mean, it is set to 1 (success is true). If the likes_count value is less than the mean, then the value for that row is left at 0 (success is false)

2.3 Data Visualization

At this point, we are able to do some visualizations to see what the data is telling us. Success is true = 1. Success is false = 0.

Fig 2.1 Fitness Success True/False compare




city = Scottsdale
city = Mesa
city = Tempe
city = Phoenix
city = Paradise Valley
city = Chicago
city = Austin
city = Los Angeles
city = Honolulu
city = Huntington Beach
city = Fountain Valley
city = Costa Mesa
city = Newport Beach
city = Minneapolis
city = Salt Lake City
city = Burlington
city = South Burlington
city = Winooski
city = Colchester
city = Boston
city = Cambridge
city = San Jose
city = Santa Clarita
city = Stevenson Ranch
0
5
10
15
20
25

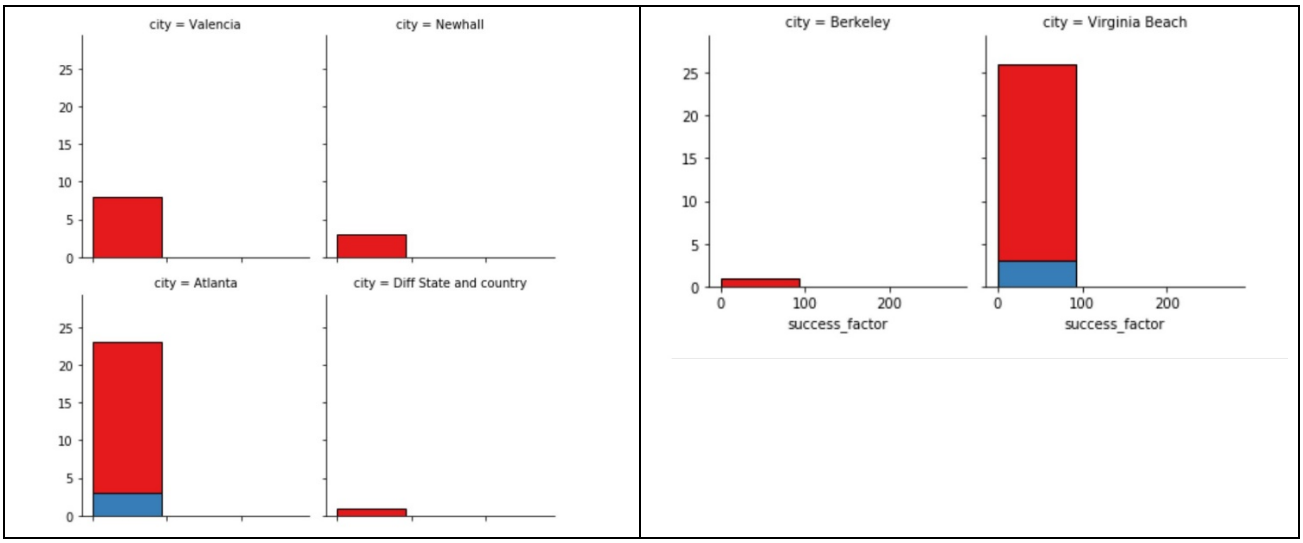
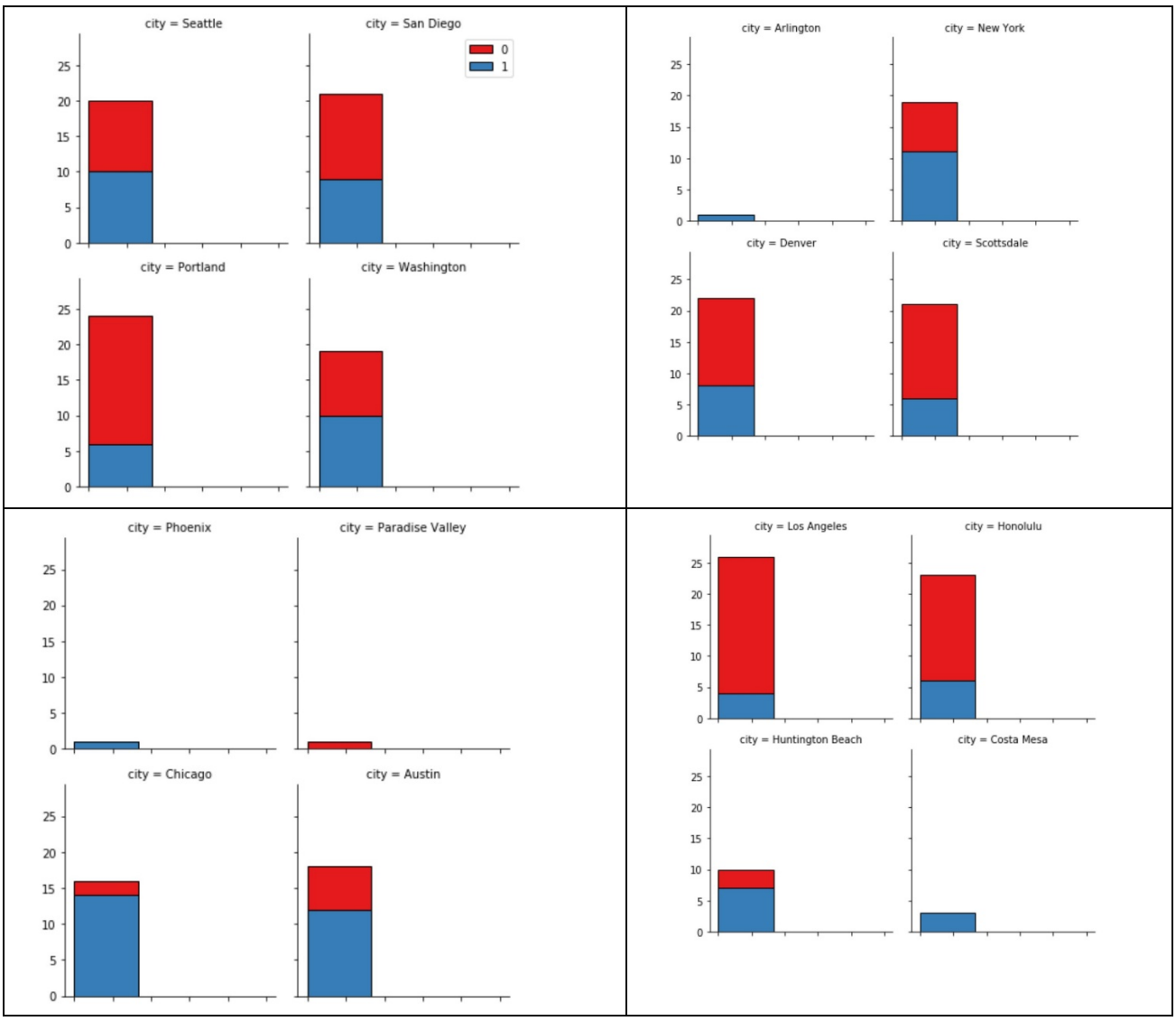
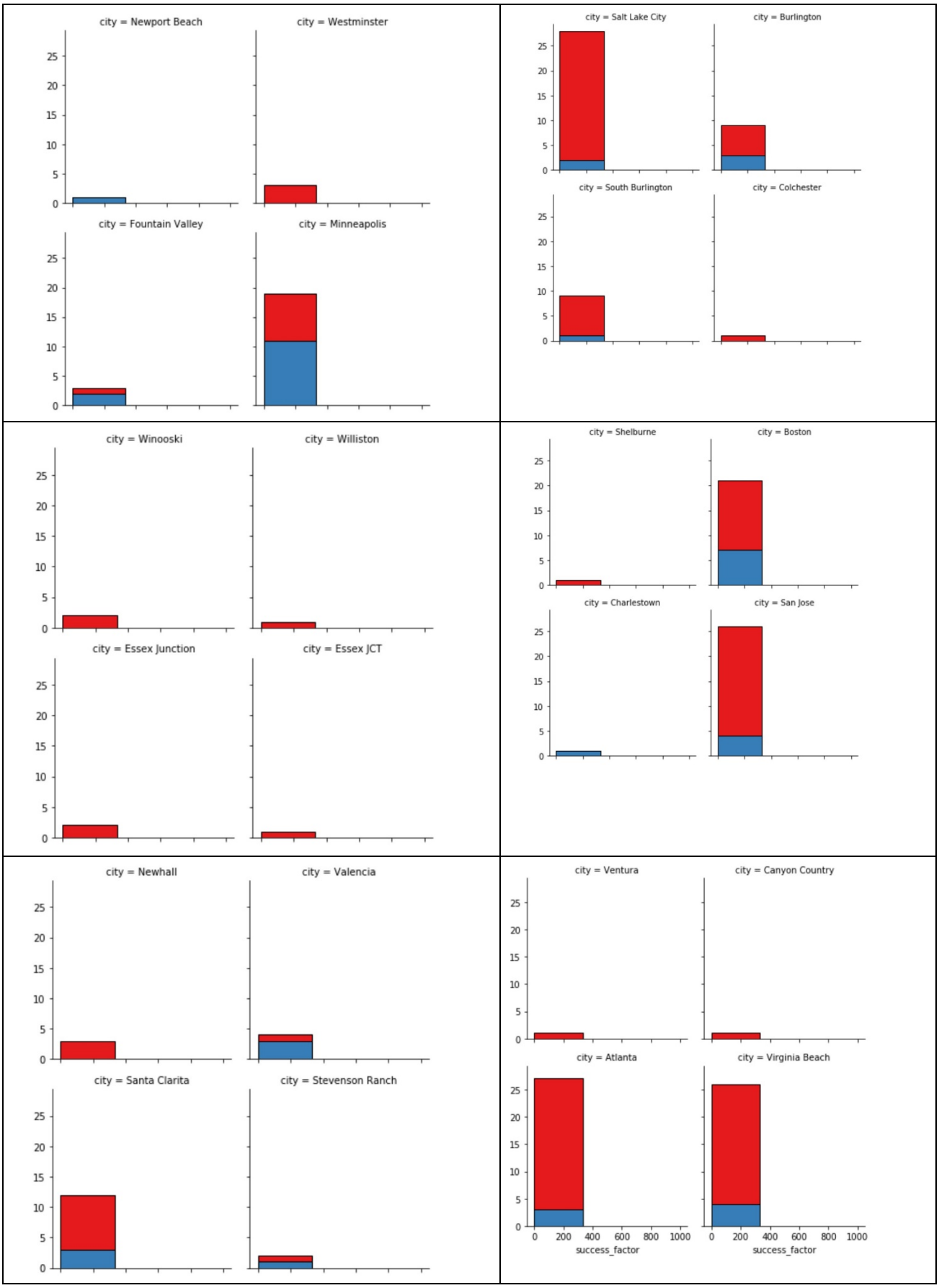


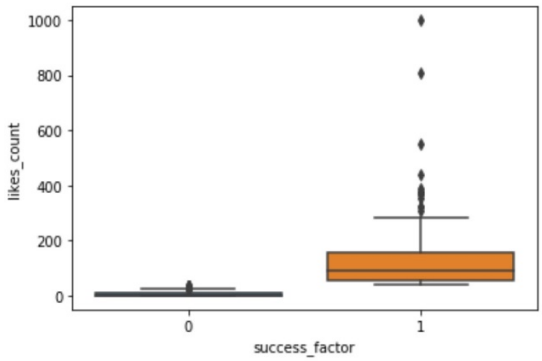
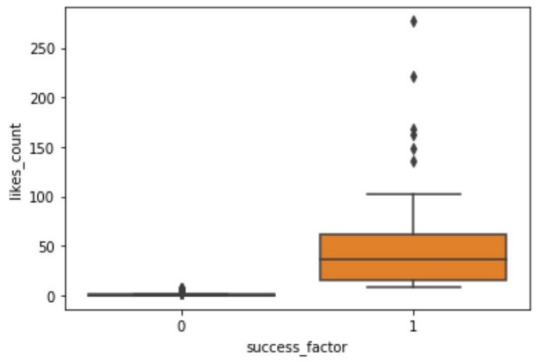
Fig.2.2 Restaurants Success True/False compare





The above visualizations suggest that most types of restaurants are more likely to be successful

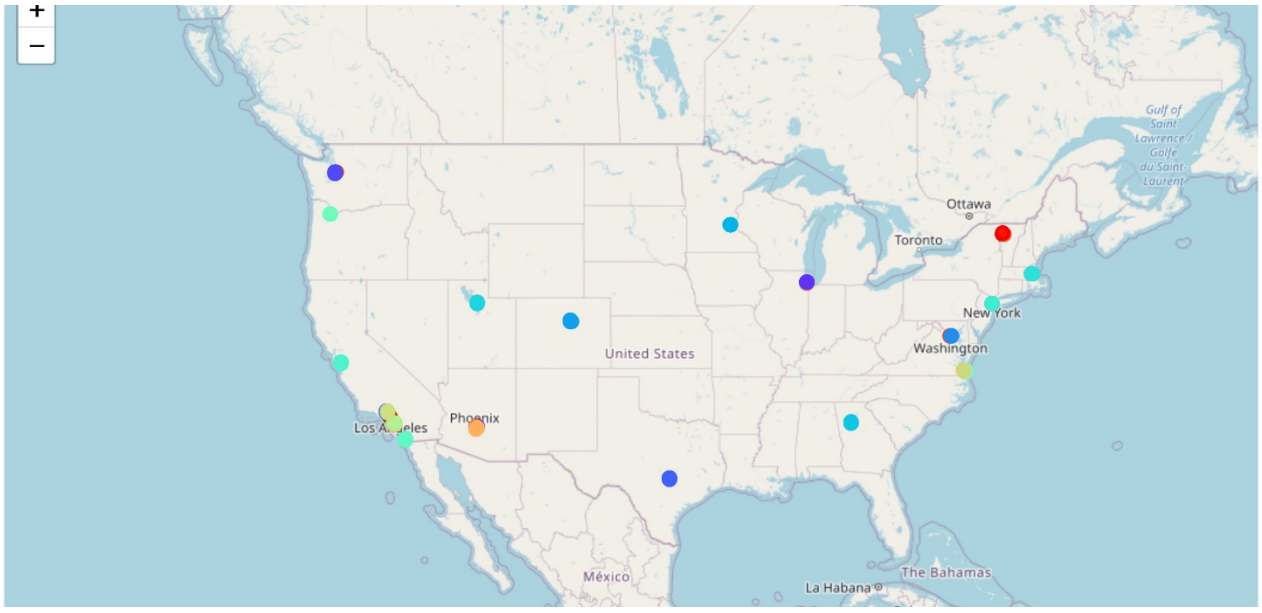
than the fitness venues are. The success rate appears to be higher among restaurants.
The following box plots indicate that many of the 'successful' venues appear to be outliers in regards to like_counts. These also suggest that restaurants are more popular than the fitness venues.
Fig 2.3

Like Counts for restaurants	Like Counts for Fitness Venues
	

3. Cluster Analysis

3.1 Kmeans Cluster

Kmeans Clustering model is used to create a visual to get a bigger picture of the healthy 'hotspots' in the country. The clusters are based on the cities captured in the dataset (combined successful restaurants and fitness venues into one dataset for clustering). Per the map, the healthy hubs appear to be mainly on the eastern and western coasts of the United States.



3.2 Observations and City Recommendations

- Based on like counts, the top cities for fitness center success (we deem this market to be most suitable for a healthy foods cafe) are:
    - New York, NY
    - Chicago, IL
    - Atlanta, GA
- Based on the like counts, the top cities for restaurant success (we look at this for competition potential) are:
    - New York, NY
    - Austin, TX
    - Boston, MA
    - Scottsdale AZ
- Based on the histogram plots for success factor comparisons, the top cities for both fitness center success and restaurant success are:
    - Seattle, WA
    - New York, NY

4. Conclusions

In this study, I used foursquare data to develop a mechanism for making recommendations for the pilot launch of a healthy foods café. Cities for review were selected based off a health score ranking from prior research dataset. Analysis was done using data visualization, and the kmeans cluster model was applied to get an overall understanding of where the best hotspots in the country may be.

Per the cluster map visualization, Seattle, WA and New York, NY are both part of the hubs along the east and west coast. These two cities also have a healthy success factor rate among both fitness centers and restaurants, which indicates that there is a target market in these areas. These are the two cities that we are recommending for pilot launch of a healthy foods café.