# Locating the next Chipotle franchise

Applied Data Science - Capstone

# Kanishk Mair

# Dec 24, 2020

1. Introduction

Chipotle Mexican Grill, Inc., often known simply as Chipotle, is an American chain of fast casual restaurants serving Mexican cuisine and was founded in the United States. It has its branches worldwide including nations like United Kingdom, Canada, Germany, and France. It has been growing every year and to continue the trend, this project aims to explore a feasible location.

1.1. Motivation

The franchise has been expanding and its locations are listed for new locations as a dataset. On a personal level, residing in Bloomington in Illinois, I wanted to explore where a new branch for the same can be introduced in the city. Also, comparison with other locations can help in finding out how the locations turned out to be successful.

1.2. Problem Statement

Based on the location data provided for the present locations, determine the next location for opening a new Chipotle franchise in the Bloomington, IL area.

1. Problem Dataset

2.1. Data Sources

As referenced earlier, the dataset was made available from [Kaggle](https://www.kaggle.com/jeffreybraun/chipotle-locations) and has the updated list of locations until 2020. The following inputs are available from the above:

* state
* location
* address
* latitude
* Longitude

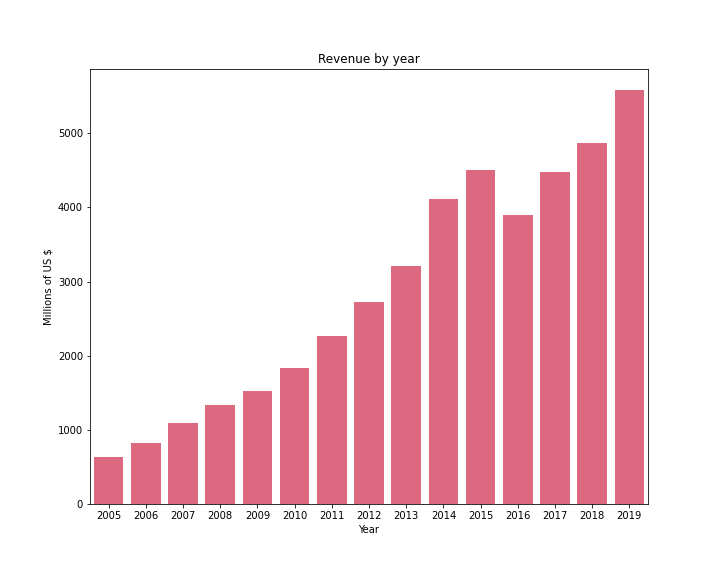
This data can be helpful to chart out the precise locations. However, we still need more data about the venue and its environment. For that, the information will be extracted through the [Foursquare API](https://developer.foursquare.com/docs/) and analyzed.

Additionally, there can be other sources for getting the statistics related to the same as:

1. [Number of locations](https://www.statista.com/statistics/221456/number-of-chipotle-restaurants/" \l ":~:text=Number%20of%20Chipotle%20Mexican%20Grill%20restaurants%20worldwide%202007%2D2019&text=This%20statistic%20shows%20the%20number,from%202%2C491%20the%20previous%20year.)
2. [Annual Revenue](https://www.macrotrends.net/stocks/charts/CMG/chipotle-mexican-grill/revenue)
3. [Top locations](https://www.apartmentguide.com/blog/best-cities-for-chipotle-lovers-in-america/)

2.2. Data Preparation

Firstly, we analyze the positive response the company has based on the stats provided above and check if it deems fit to open a new location. Next, based on the present



**Figure 1:** Revenue earned across all the locations

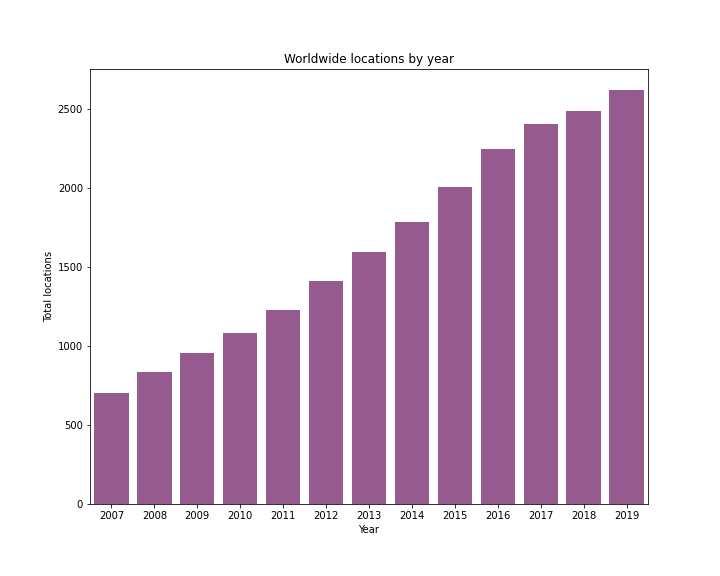
locations, the Foursquare API will be used to analyze the top locations and how they differ from other locations. From there onwards, required data will be analyzed based on those results.

1. Exploratory Data Analysis

3.1. Feasibility check

In order to check the feasibility of opening a new chain, let’s first check the company’s growth. The revenue earned by Chipotle is plotted in (**Figure 1**). On an annual basis, the company’s revenue has seen a consistent rise, albeit a slight slump in 2016. This is a positive sign considering that the company would be more likely to expand.

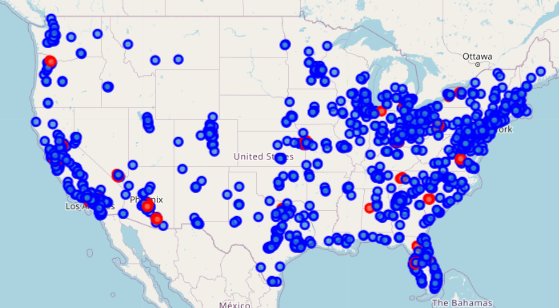
We can also look at the number of branches in operation worldwide (**Figure 2**). This is a more helpful parameter since it helps answering our objective. The chart is monotonously increasing which shows the company’s intent to expand further. From both of these observations, we can conclude that if there is a potential location with possible growth, then a franchise can be opened.



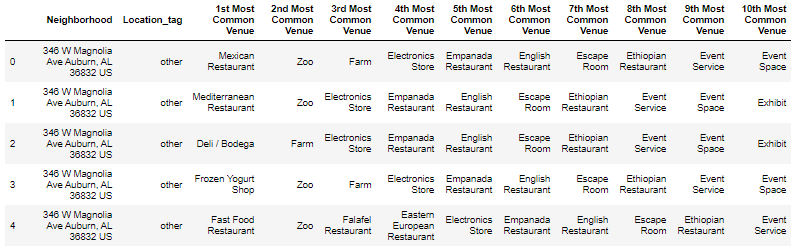
**Figure 2.** Number of locations worldwide

3.2. Present locations

Before moving on to choosing our location, we analyze the present locations. A map of current locations along with the top performing ones are charted (**Figure 3**) which show the dispersion of their locations along prominent cities.



**Figure 3.** Locations of Chipotle stores within the US. Here, the red locations are the top performing locations while other are marked in blue



**Figure 4.** Top 10 most common venue in a neighborhood

1. Methodology

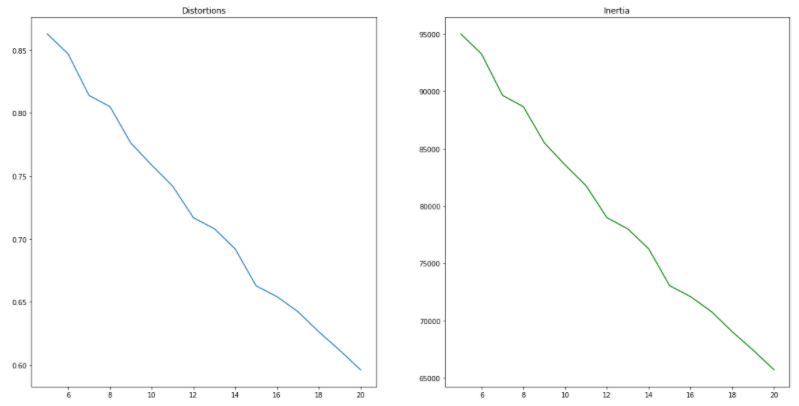
4.1. Most common venues

Using the Foursquare API, we can analyze these venues and get more insights into the locations/settings they are presently operating in. We provide the geographic parameters and find all the nearby venues. After this data generation, there were ~700 various kinds of venue, like various *Restaurants, Stores, Offices, etc.* For pre-processing we select only the venues which were mentioned more than 5 times (a hyper-parameter).

Next, the frequency counts of the data was made to understand which venue was most common for a particular location. We also tag whether the location is the top performing franchisee or not to analyze the differences. Lastly, the top 10 most common venue in that neighborhood was determined. The final data will be assigned the clusters (**Figure 4**).

4.2. Clustering the data

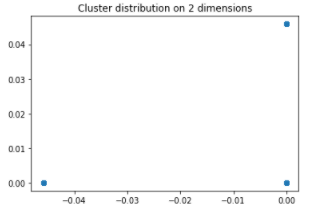
Now that we have the frequency values, we can cluster our data to determine what distinguishes various locations. To determine the number of clusters, the Elbow method was applied, however, the distortion never stabilized even for 30 clusters (**Figure 5**).



(a) (b)

**Figure 5.** Elbow method plot of (a) Distortion, and (b) Inertia with varying ‘K’ in KNN

However, the results of cluster and its reduced version in 2D are as follows:



|  |  |
| --- | --- |
| Cluster | Count |
| 1 | 11,001 |
| 2 | 475 |
| 3 | 458 |
| 4 | 356 |
| 5 | 475 |

The clustering algorithm doesn’t succeed since there are multiple venues and based on the number of clusters provided, it just separates those that have the highest frequency. The 2D plot of the data shows that most of the data has low frequency for all dimensions and gets classified as cluster 1. Then, the remaining clusters have 0 frequency in all except that cluster’s dimension. For instance, if a venue is very common, then that dimension has its own cluster created. While predicting, if a venue has high occurrence of it, then it’ll be classified into that dimension. However, if a cluster has diverse set of results, then it goes into the main cluster and this is the most common case.

With feature engineering, we can remove some of the less common features but still it would result in more distinct but overfit clusters. Based on the cluster interpretation, we can instead see what are the top locations for the top and normal locations are.

|  |  |  |
| --- | --- | --- |
| Rank | Top Locations | Other Locations |
| 1 | Clothing Store | Mexican Restaurant |
| 2 | Mexican Restaurant | Clothing Store |
| 3 | Coffee Shop | Sandwich Place |
| 4 | Sandwich Place | Fast Food Restaurant |
| 5 | Pizza Place | Pizza Place |
| 6 | American Restaurant | American Restaurant |
| 7 | Cosmetics Shop | Mobile Phone Shop |
| 8 | Bar | Furniture / Home Store |
| 9 | Shipping Store | Convenience Store |
| 10 | Jewelry Store | Frozen Yogurt Shop |

**Table 1.** Comparison of the most important venues at Chipotle’s Top and Other locations

From the above, we see that the top ranked locations are similar to other locations but the difference must be in the higher population frequenting those locations. The main takeaway is that all locations are situated close to Mexican restaurants and then other common restaurants. After that there are different shops and businesses close to the venues.

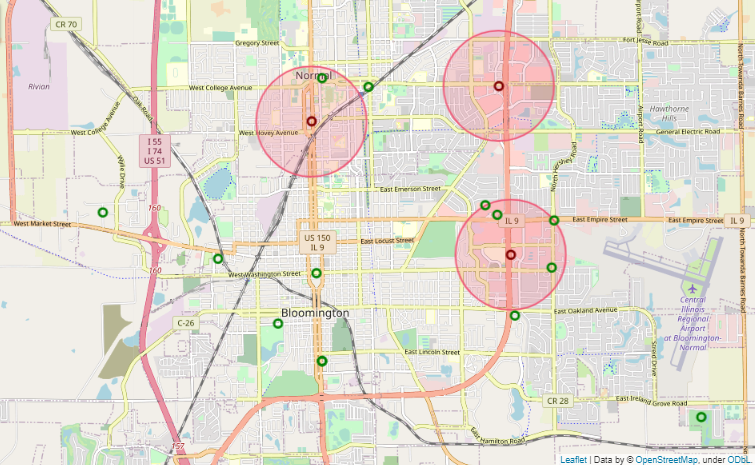
4.3. Locations in Bloomington

Based on the cluster analysis, our new location must be close to the Mexican and other restaurants. Since the city is spread in a radius of more than 8 km, the data exploration with Foursquare will be done for 10 km of radius.

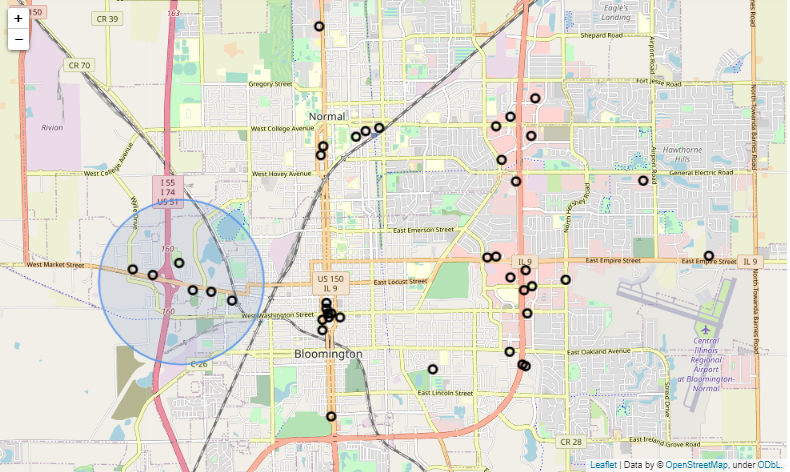
First, by exploring the neighborhood to find locations where the Mexican restaurants are located (**Figure 6**), we find that there are 3 existing franchisees of Chipotle operating. We also find that possible areas to explore are in the South and South-West regions of the city.

Within a city, there are certain regions where land is allotted for business while remaining might be for residential or other purposes. Thus, by exploring the location of all other restaurants, we can spot other locations where commercial activity occurs.

Searching for other restaurant locations where other restaurants are located (**Figure 7**), we notice that there is a large cluster of restaurants where there will be high competition while remaining locations are spread and are too few. On the other hand, the area in the

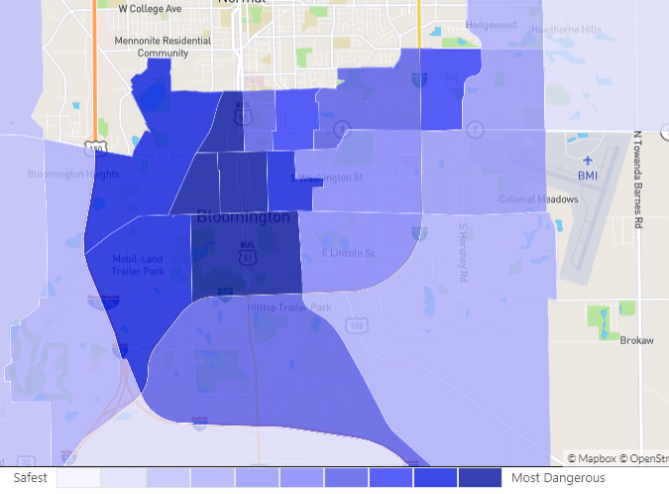
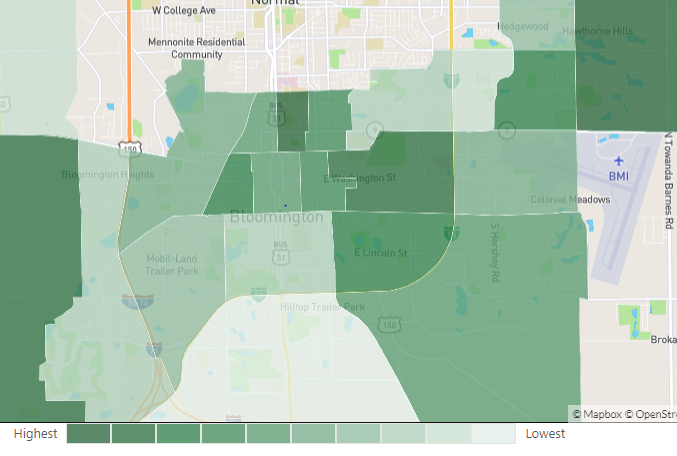


**Figure 6.** Location of all the Mexican restaurants in Bloomington. In red are the Chipotle locations with a circle of 1km radius too close for establishing a new branch



**Figure 7.** Location of all other restaurants in Bloomington. The highlighted region in blue is the location suggested

South-West part of the city are spread out and is more preferable. Finally, let’s look at other important aspects like the crime and cost of locations (**Figure 8**). From the choropleth map, it’s seen that the Downtown area with most of the other businesses in the South has some of the lowest land prices in the area. However, it is one of the Most Dangerous places crime-wise. The South-West part of the city is slightly more expensive, but is more safer and preferable for the business to setup. The other establishments are: Bob Evans Restaurant, KFC, Cousin's Restaurant, Country Pride Restaurant which too are similar to Chipotle in terms of expensiveness and the target population.

(a) (b)

**Figure 8.** Choropleth map of Bloomington showing (a) the crime rate, and (b) the realty prices to setup a business

1. Conclusion

The objective of this project was to determine the next location to be chosen as a possible candidate for a Chipotle outlet. Using the location dataset for in operation Chipotle’s locations, we discovered various venues. Although there was no clarity on the clusters made, a deeper analysis of the small clusters created revealed that they are most common surrounding venues. From this, we were able to conclude that the most common venues in the cluster were Mexican and other restaurants. Based on this and keeping in mind the location of presently running Chipotle outlets, the final location was selected in the South-West part of the city as seen in Figure 7. The location’s safety and land price was also analyzed.

The future steps must entail improved data collection and data about the city to carry out a thorough analysis along with investors cost and other constraints.

Finally, a detailed analysis of site, cost and footfalls in the region must be done to prepare a formal draft proposal