

# Fast Food Restaurant Distribution Analysis based on neighbourhood for North America

Dhruvi Pandya  
7th July, 2019

## Background

With the total area of 9.834 million km<sup>2</sup> and population of 32.7 Crores[1], USA has the 500,000 [2] fast food restaurants in total. From the exit to the highway to every food court in the mall, suburbs, movie theatre you will find fast food outlets everywhere. These observations not only stand for countries like USA but for small countries like India as well. Since the last few years, there has been an explosion in the intake of fast food all over India, by almost 40%, spending almost 2.1% of per capita income on fast food[3]. This increase makes the fast food industry a very lucrative industry to venture into. There are multiple things one needs to consider before opening a new restaurant let alone a fast food restaurant. Type of food to offer, location, franchise, capital etc. Location becomes an important factor after the capital. One might say that if you are opening an outlet under franchise umbrella, the outlet is bound to be successful if the franchise is successful. But that is not really the case always. One of the reasons of failure is what if the target consumers group is not present in your outlet location? If the outlet doesn't have any traffic or visibility the business might not be successful. Since fast food restaurants usually target the age group of 15-45 years[5], and the fast food restaurants are usually located by highways, business streets and shopping centers and close to other fast food restaurants. Malls, colleges and universities, there has to be some pattern in the success of the restaurant and its location.

## Problem description

Analysis of successful fast food restaurants based on the neighbourhood they are located in helps in understanding the insights and relations present between the target consumers and the success of the restaurant. This project aims at unearthing above insights.

## Interest

Since this analysis is a general analysis between the target audience and the location, it remains in the interest of many industries like hotel industry, real estate which looks out for prime locations meaning what they are near to and what others attractions they are nearby. One such example is the rooms having a view are sold more as compared to other rooms in a hotel. Other such example is nowadays movie theatre are located in malls in contrast to the stand alone buildings in past.

## Data acquisition and cleaning

The dataset for the fast food restaurant is available at Kaggle [6]. It has the data about the outlet name, state, coordinates, and other location information. There are total 18,000 rows. The definition of the columns are as below:

Sr No.	Columns	Description
1	address	The address of the restaurant outlet, i.e. the street and the block number
2	categories	The type of food the restaurant serves, i.e. Fast Food
3	city	The city of the outlet
4	country	The country of the outlet, i.e. US
5	latitude	Latitude coordinate of the outlet
6	longitude	Longitude coordinate of the outlet
7	name	Name of the outlet
8	postalCode	PostalCode of the outlet
9	province	Province where outlet is located
10	websites	Website of the franchise/ outlet

The data is in two files which are combined to form a single table. The data doesn't have missing values but has many data entry discrepancies. Ideally data should have franchises with multiple outlets within a province and across the province as well. For Example, McDonalds has about 5000 outlets in the US. These outlets are present in the data with different names like McDonald's, Mc Donalds, McDonalds @ Old Atlanta, Mcdonald's, Mcdonald's - Walmart, Mcdonald's #25557, Mcdonalds Whitehouse. While they belong to the same franchise, such data

when used while grouping; will be interpreted as different franchise all together. 4 different cases were encountered while cleaning the data (i) special characters used in name (ii) multiple outlets at the same position (Food court like arrangement) (iii) Case mismatch (iv) province name appended in the name . First 3 cases constituents of about 25% of the data and the 4th case ~2%. Therefore the entries in the 4th case were ignored. For 2nd case, when there is a food court like location, more than one outlets are present at same location and same coordinates. Such data needed to be separated as we need to do a franchise wise location wise analysis.

After cleaning the data, the final rows count was around 18,000. The dataset still did not have the data regarding the neighbourhood (within 500m ) like what schools, malls, offices are present near it. This data was explicitly fetched from the FourSquare API for Places. Also additional provincial data like the total area and the population of the province was captured from web by scraping [8,9]. Additional data about the tourism in the province was also captured i.e. if the province belongs in the top 10 tourism province in the US, then it has major tourist inflow.

## Exploratory data analysis

In order to understand how best to select a location for your new fast food restaurant, location analysis of the successful food chains becomes essential giving insights into how neighbourhood plays the role in attracting customers. Following is the bar chart showing the total fast food restaurants per province.

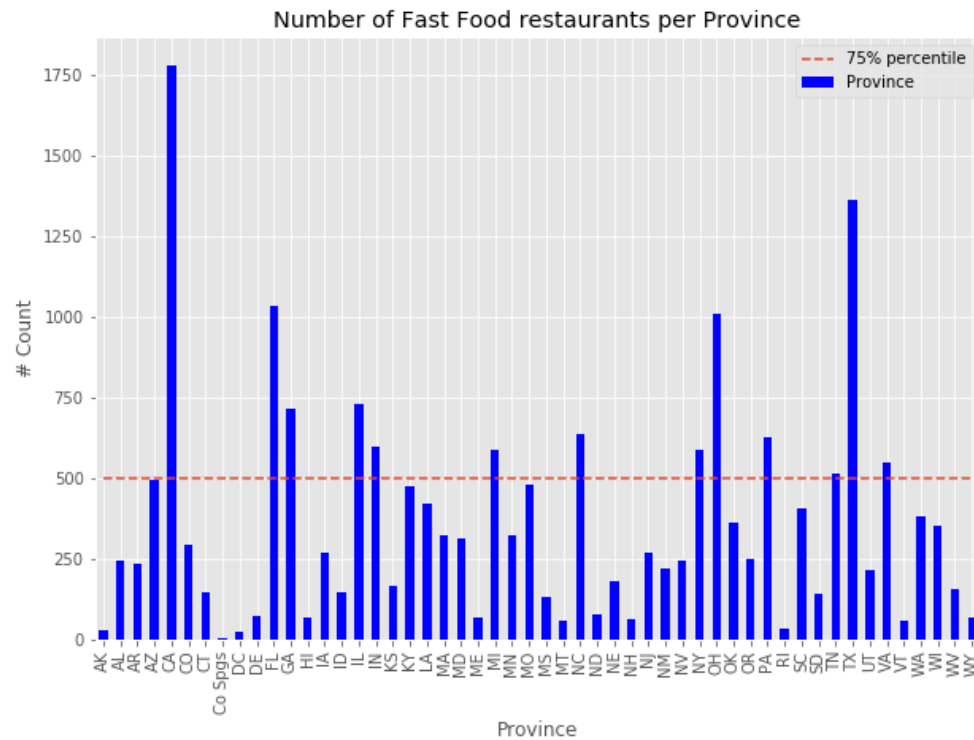


Figure (i)

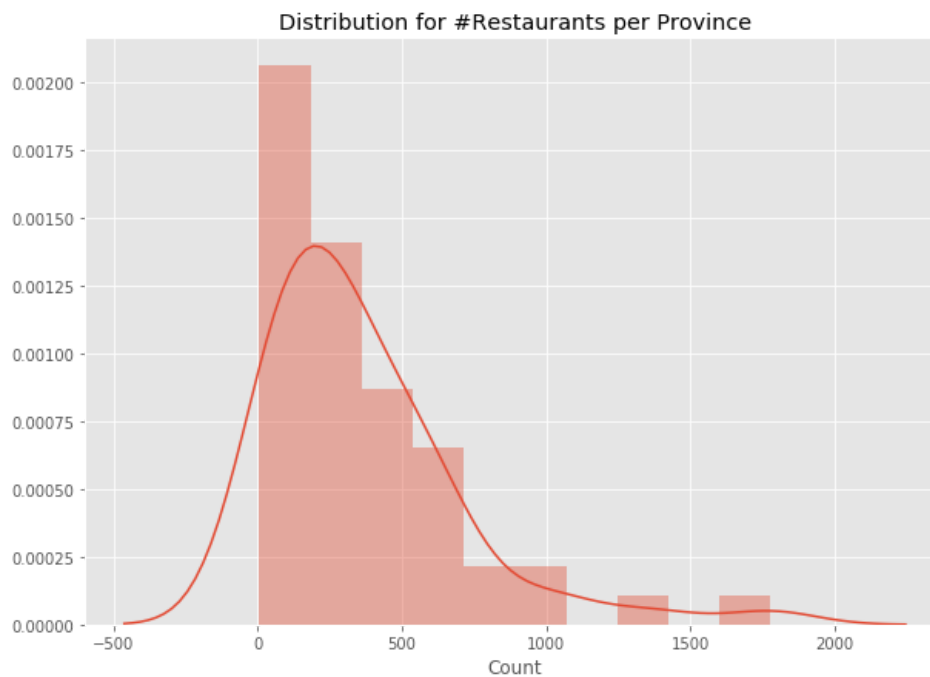


Figure (ii)

As is clear from the figure (i) and (ii), 60% of the provinces have restaurant count between 200-500. About 25% of total restaurant distribution is concentrated in a few provinces. Refer figure (i), the dotted line shows the 75 percentile of the restaurant distribution i.e. all the markers below the line constituent the 75% of distribution. 13 Province forms the 25% of the distribution.

My personal interest lies in these 13 provinces where 25% of the restaurants are located. Following are the 13 provinces and number of restaurants they have.

Sr No.	province	restaurants_in_province
1	CA	1778
2	TX	1364
3	FL	1033
4	OH	1009
5	IL	728
6	GA	715
7	NC	635
8	PA	628
9	IN	599
10	NY	588
11	MI	586
12	VA	549
13	TN	515

California has the highest number of fast food restaurants followed by Texas and Florida. Let us take a look at the geographical location of these restaurants in USA.

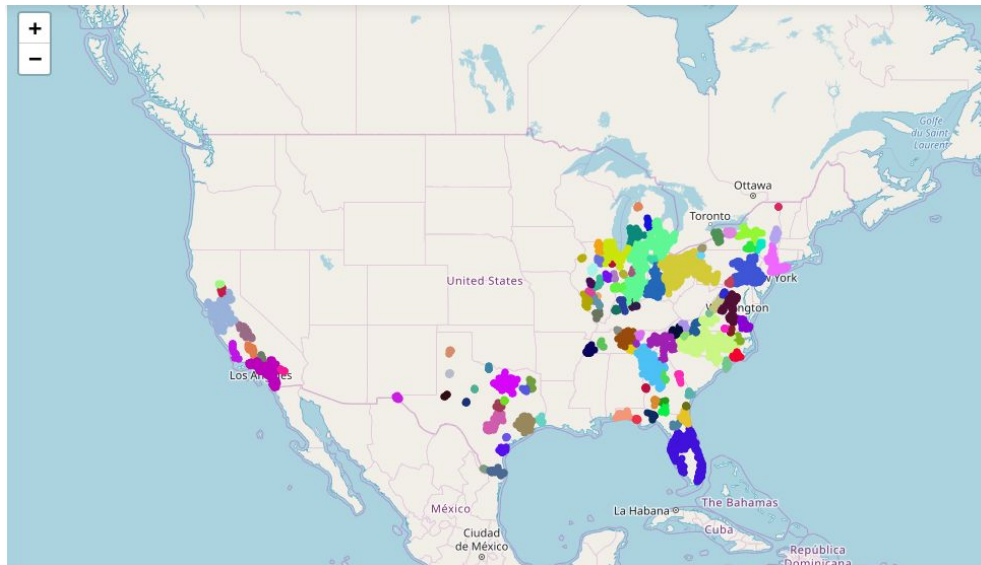


Figure (iii)

The location of these 13 province where apart from California and Texas all other provinces are clustered together definitely has some insight. With the demographics of these provinces it could be anything from the number of tourist places to the number of universities and malls it

has. Or it could be related to the population of the province or the number of highways in the province as it is a common known fact that fast food restaurants are mostly situated around the highways. To understand this relationship, a metadata for these provinces were scraped from the internet. This meta data comprises of the following variables:

<b>area_km2</b>	Area per sq km of the province
<b>population</b>	Population of the province
<b>province</b>	Code of the province
<b>is_tourist</b>	Is Province in top 10 tourist places in USA ?
<b>restaurants_in_province</b>	Total number of fast food restaurants in province
<b>rest_per_km2</b>	Total restaurants per sq km
<b>rest_per_1000people</b>	Total restaurants per 1000 people

The pairwise correlation of these variables shows the following outcome:



Figure (iv)

As can be seen in the above correlations, the variables like population and the area\_km2 does not directly affect the restaurant distribution. Observing the graph of variable restaurants\_in\_province to area\_km2 and populations, we can safely say the number of restaurants increase with increase in the population or the land area. While the graph of restaurants\_per\_province to rest\_per\_km2 and rest\_per\_1000\_people tells us that even though the number of restaurants in the province increase, the franchise wouldn't want to over cater the population by cluttering many restaurant together.

The tourist state factor might not be very much visible in the pairwise plot; but it is very evident in the figure (iii). Consider the following figure just for California which is the house to Hollywood and Silicon Valley. Let us check what impact these world class attractions have on the fast food restaurants:



Figure (v) California

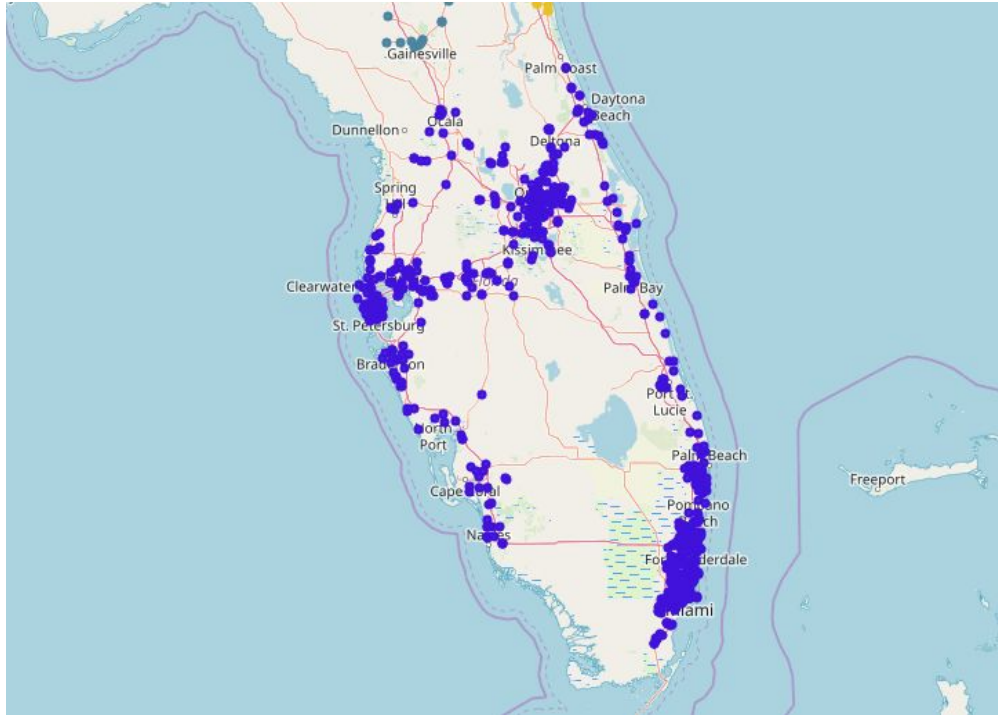


Figure (vi) Florida

As can be seen from the maps, one major cluster of the restaurants are located near the silicon valley near San Jose and Mountain View while the other cluster is at the Los Angeles and Malibu. Similarly, in Florida as can be seen in figure (vi), all the restaurants are located near the sea shores, hence attracting the tourists. This is an insight for a tourist province.

The 13 provinces in the top 25% percentile province contains non-tourist province as well.

Consider Indiana Province. Its at the bottom, both population wise and area wise in the list of 13 provinces. Consider the following figure:



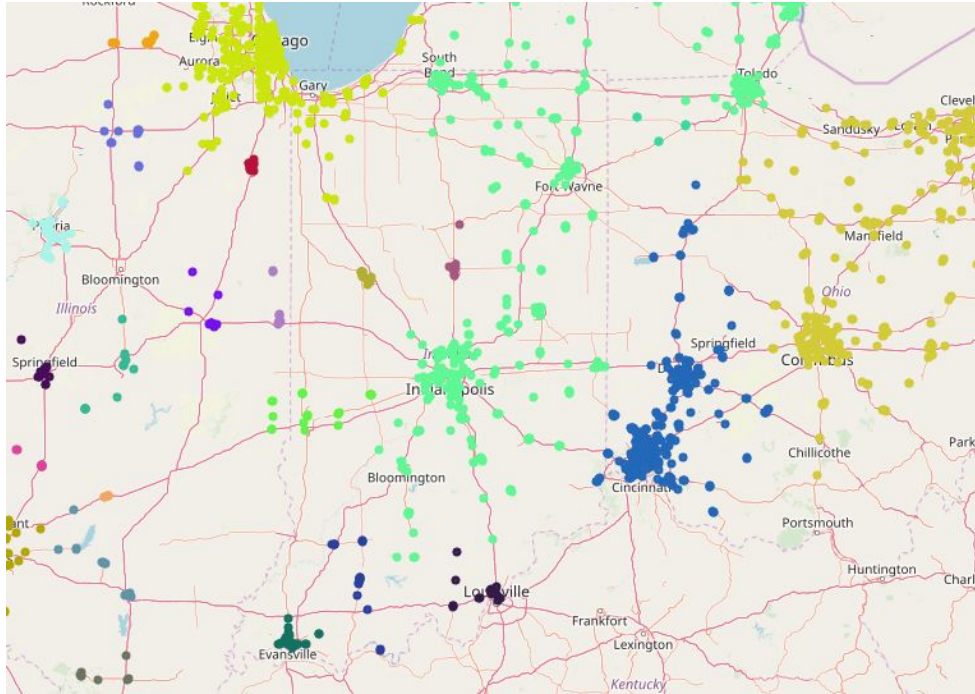


Figure (vii)

Indiana is not a major tourist attraction province. But as can be seen in the map, there are clusters of restaurant around some road of networks. This could be some mall, offices hub, schools, parks or any of the places attracting the crowd. Following figure shows the location of the restaurants in Indiana and some major public places like schools, universities, office hubs, malls, parks etc. The Red markers are the public spaces and the blue markers are the restaurants. Almost 96% of the locations overlap.

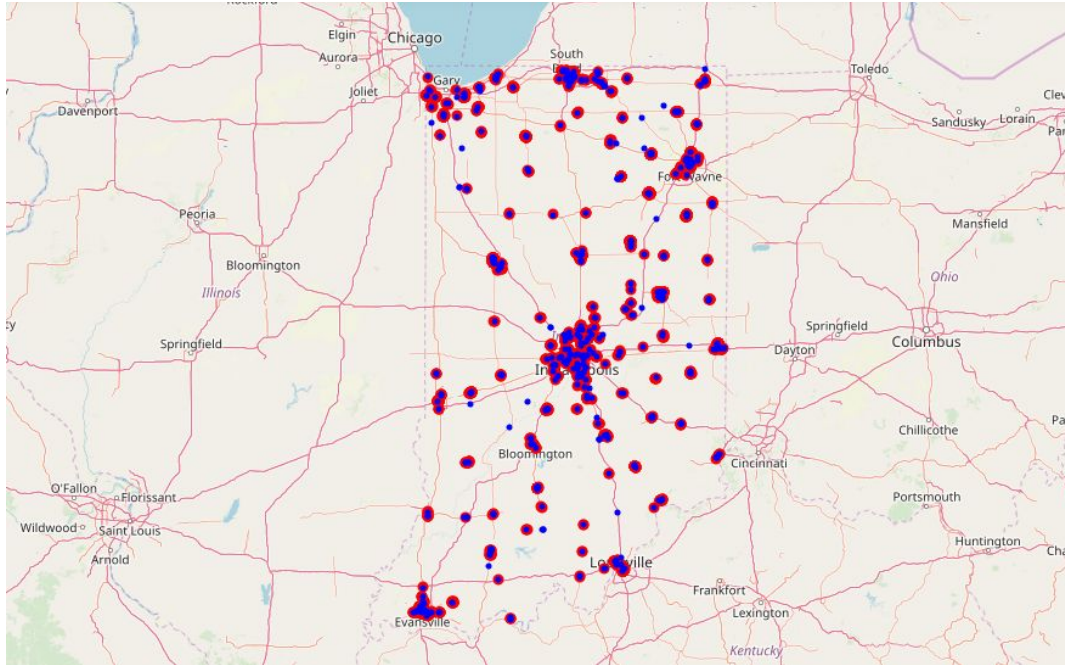


Figure (viii) Indiana

The data about the surrounding public places was fetched with the FourSquare API. This data was fetched using the venue API, which returns the places with location information which matches the given search query within the given radius, in our case its 500 meters. So, it becomes evident that the restaurant locations are selected keeping in mind the crowd attracting places nearby.

Let us see how these crowded places affects the restaurant distribution for other provinces:

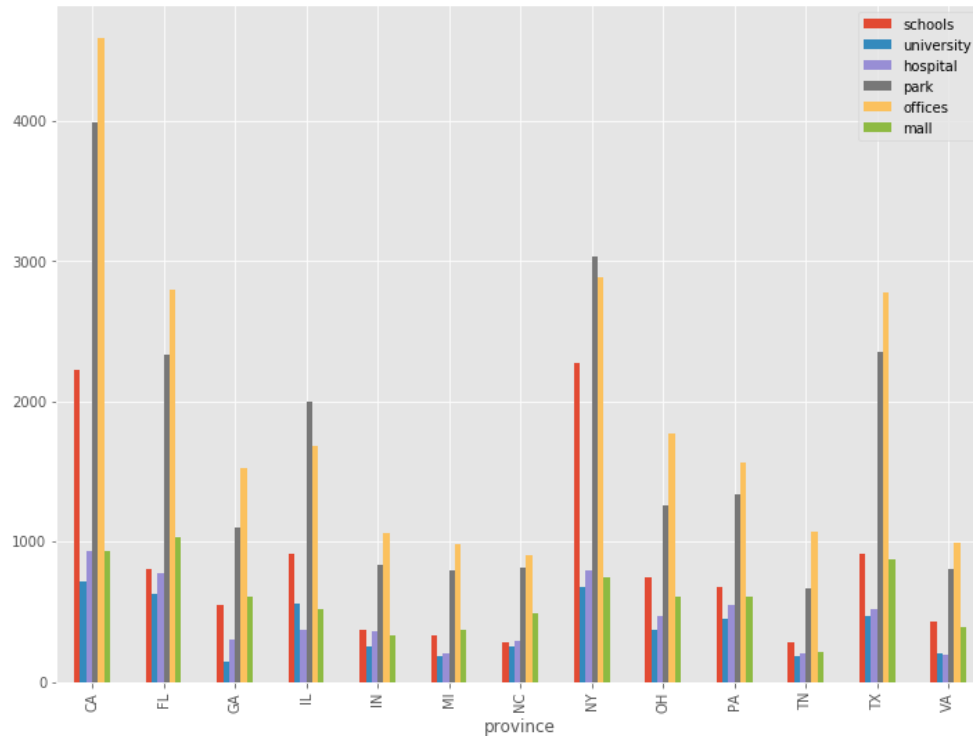


Figure (ix)

The above chart shows that in general fast food restaurants target the places nearby parks and offices. It also gives us the insight about the age group of the target crowd of the restaurants. California is a business and start up hub, attracting the young generation between 20-45 yrs old. Similarly is the New York Province with high number of offices and schools surrounding the restaurants.

## Conclusion

In this study, I have analysed how existing successful fast food restaurants have taken the hidden neighbourhood data in consideration for selecting a profitable location for their outlets. I have taken into consideration the points which I highlighted in the beginning of the study and cross verified with the actual distribution. This type of study can be used to make a classification of a potential profitable location for a new restaurant. That type of classification problem falls under the umbrella of positive and unsupervised classification as we do not have many negative/ failed cases of restaurant in the data. Also, this type of analysis helps in providing new directions for the marketing and service providing to the restaurant and the hotel industry.

## Future Direction

In this Case Study, there are many possibilities of further analysis, like classifying potential profitable locations. Furthermore, one can do analysis based on the individual franchise like McDonalds, Subway etc to understand how these giants select their locations for restaurants and what are the scenarios which lead to failed cases.

## Webliography

[1] [www.google.com](http://www.google.com)

[2]

<https://www.google.com/url?sa=t&rct=j&q=&esrc=s&source=web&cd=2&cad=rja&uact=8&ved=2ahUKEwjVsYm7pNriAhVI8HMBHaZOCtcQFjABegQIDRAE&url=http%3A%2F%2Fwww.pamf.org%2Fteen%2Fhealth%2Fnutrition%2Ffastfood.html&usg=AOvVaw3H21p9wWg4PpKcc86YmmQN>

[3] <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4484093/>

[4]

<https://smallbusiness.chron.com/things-consider-before-opening-fast-food-restaurant-17437.html>

[5] [https://www.bplans.com/fast\\_food\\_restaurant\\_business\\_plan/market\\_analysis\\_summary\\_fc.php](https://www.bplans.com/fast_food_restaurant_business_plan/market_analysis_summary_fc.php)

[6] <https://www.kaggle.com/datafiniti/fast-food-restaurants>

[7] <https://www.worldatlas.com/articles/the-most-visited-states-in-the-us.html>

[8] [https://en.wikipedia.org/wiki/List\\_of\\_U.S.\\_states\\_and\\_territories\\_by\\_area](https://en.wikipedia.org/wiki/List_of_U.S._states_and_territories_by_area)

[9]

[https://en.wikipedia.org/wiki/List\\_of\\_states\\_and\\_territories\\_of\\_the\\_United\\_States\\_by\\_population](https://en.wikipedia.org/wiki/List_of_states_and_territories_of_the_United_States_by_population)