**Finding a possible location and type for a restaurant in Los Angeles County, CA** By Felix Fichtner

Table of Contents

[1 Introduction - 2 -](#_Toc72683247)

[2 Data - 3 -](#_Toc72683248)

[3 Methodology - 4 -](#_Toc72683249)

[3.1 Initial Analysis - 5 -](#_Toc72683250)

[4 Results and Discussion - 6 -](#_Toc72683251)

[5 Conclusion - 6 -](#_Toc72683252)

# Introduction

In this capstone project I will try to give a recommendation on where to open a  
restaurant in Los Angeles County, CA. The location will be a community or area in the county, that is defined by its postal code. In addition, there shall be given a  
recommendation of whichs type of food venue could be opened, based on  
existing restaurants in the area and generally popular restaurants in the  
whole county.

The decision on where to open a restaurant can be based on many factors,  
depending on the target group. For example, one could look for very dense  
populated areas, or areas with lots of wealthy citizens. Even the median age  
of the population can play a role. Because of this, I will include census and economic data into the analysis.

I will make a final recommendation based on the following conditions, that :

* Find the area with a good balance between number of possible customers  
  and a high median income (income is slightly more important than population).
* The type of restaurant will be determined by the most recommended categories  
  of food venue in Los Angeles County, CA and the number of already existing  
  venues in the area, grouped by their categories.

# Data

The basis for the analysis will be the 2010 census data of all zip codes in Los Angeles County, CA. This census data provides information on the composition of the population, age distribution, and the total numbers. This dataset will be enhanced with the 2019 median household income per zip code. The income can also serve as factor in the recommendation on where to open a restaurant, how it should be set up, and what the target audience could be. As a third part geospatial data is added for every zip code, to be able to use Foursquare to explore existing food offerings in the communities of Los Angeles County, CA.

Table

Description automatically generated with medium confidence

Figure - subset of the prepared dataset

Using Foursquare, already existing food venues can be explored, and grouped. This way the competitors can be explored, for example by the food type or rating. It is also possible to identify market niches and give a recommendation for a type of restaurant that could be opened.

Table

Description automatically generated

Figure - Foursquare Data for Zip Code 90001

The following example illustrates how the data from Foursquare helps to analyze the food venues in a community. Using this API, one can determine that Mexican restaurants are the dominant food venue category in Florence-Graham, South Los Angeles.

Table

Description automatically generated

Figure - Food Venues grouped by Category in Florence-Graham

In detail, we will try to answer the following questions:

* What are the most recommended types of restaurants in the county?
* What is the distribution of the existing food venues categories in the area where we want to open a restaurant in?

The datasets are publicly available at:

* 2010 Los Angeles Census Data
  + https://www.kaggle.com/cityofLA/los-angeles-census-data
* Median Household Income by Zip Code in 2019
  + http://www.laalmanac.com/employment/em12c.php
* US Zip Code Latitude and Longitude
  + https://public.opendatasoft.com/explore/dataset/us-zip-code-latitude-and-longitude/information/

# Methodology

In this project we will focus on finding a suitable area for a new  
restaurant in Los Angeles County, CA. The areas are defined by their US Zip  
Code. In addition, we will look at the most  
recommended food venue categories throughout the country, to suggest which  
type of restaurant could be opened. There won’t be a specific location in  
the chosen area recommended.

In the first step we have merged three different datasets, that provide data  
on the different areas in Los Angeles County. With this data it is possible  
to group or cluster the areas using information like median income, number of  
households and number of inhabitants.

The second step in the analysis is to cluster (using k-means clustering) the  
areas in the county and to describe the individual clusters. Using this  
method we support the process of finding a single area that looks promising  
for a new restaurant.

The third step is to pick a cluster that fits the chosen criteria most. The  
area shall be chosen under the premise of finding a good balance between estimated median income and number of potential  
customers. So, the target is to find an area that has as many citizens as  
possible with the highest income possible. After an area was chosen, the  
distribution of local restaurant types in this area will be analysed.  
Combining this information with the categories of food venues that are  
popular throughout the whole county, a recommendation of the restaurant to open  
can be given.

## 3.1 Initial Analysis

First, let us see what we can find out by analyzing the dataset. As population per Area and income are most important. I have sorted the dataset by these features, to make out some potential candidates for opening a food venue, according the our criteria.

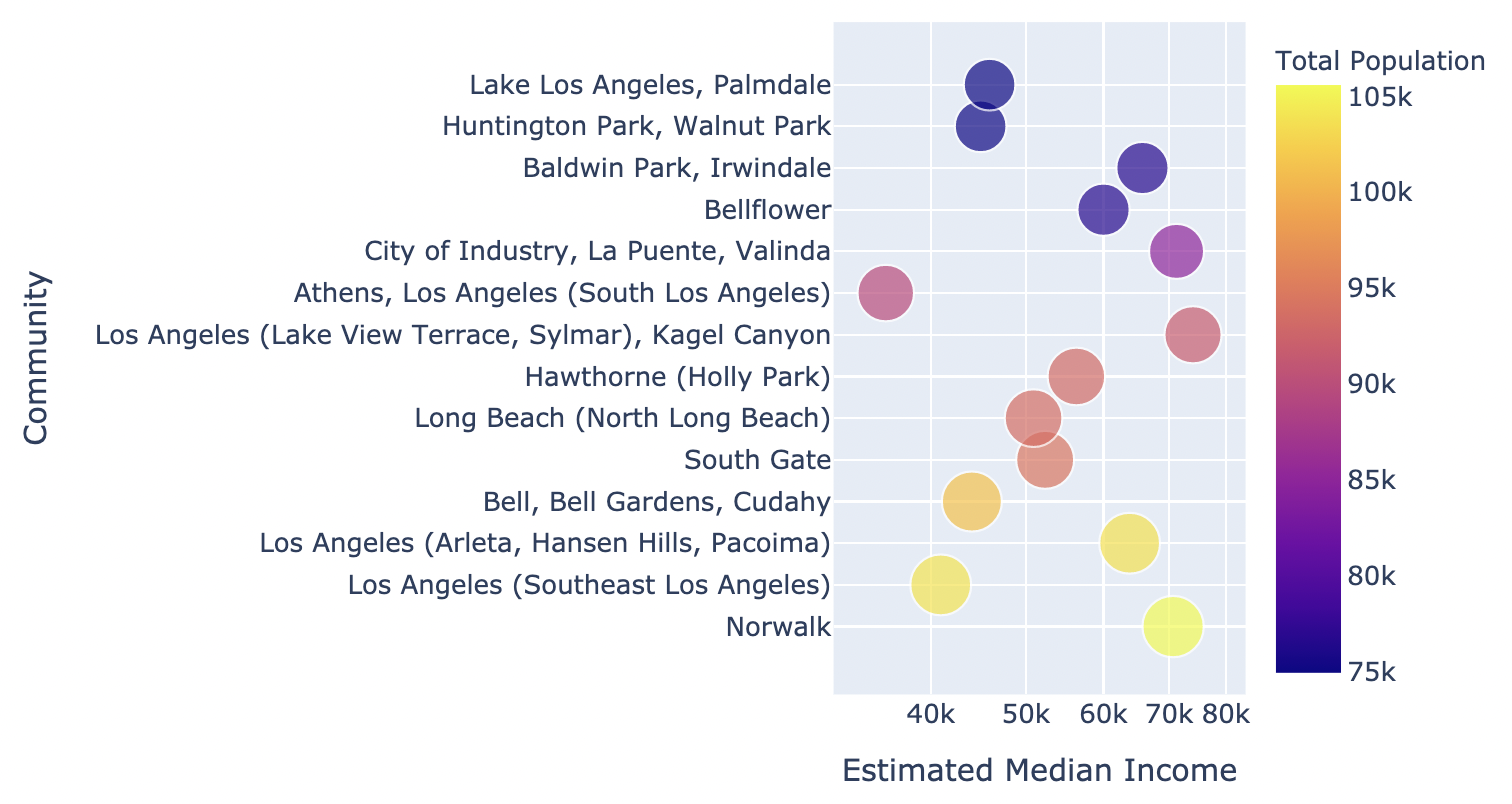


Figure - Income and Population per Area

In Figure 4 we find a combined representation of the top 15 areas by their population (bubble size), combined with their estimated median income (x axis). In this graph there are four areas that stand out:

|  |  |  |
| --- | --- | --- |
| **City** | **Population** | **Median Income** |
| Norwalk | 105,6 K | 70,7 K |
| Lake View Terrace, Sylmar | 91,7 K | 74 K |
| La Puente, Valinda | 85 K | 71,2 K |
| Hansen Hills, Pacoima | 104 K | 64 K |

Table - Areas with a good combination of income and population

These four cities / neighbourhoods seem to be suitable areas, based on their combination of population and median income. We are going to see, if this assumption is confirmed going forward.

In preparation for further analysis, the most recommended food venue categories in Los Angeles County were identified, using the Foursquare API. This was done by going through all available postal codes and querying up to 50 of the most recommended food venues in every area (sorted by their popularity) in a 1 km radius. The food venues were then grouped by their category, to see which are most often recommended.

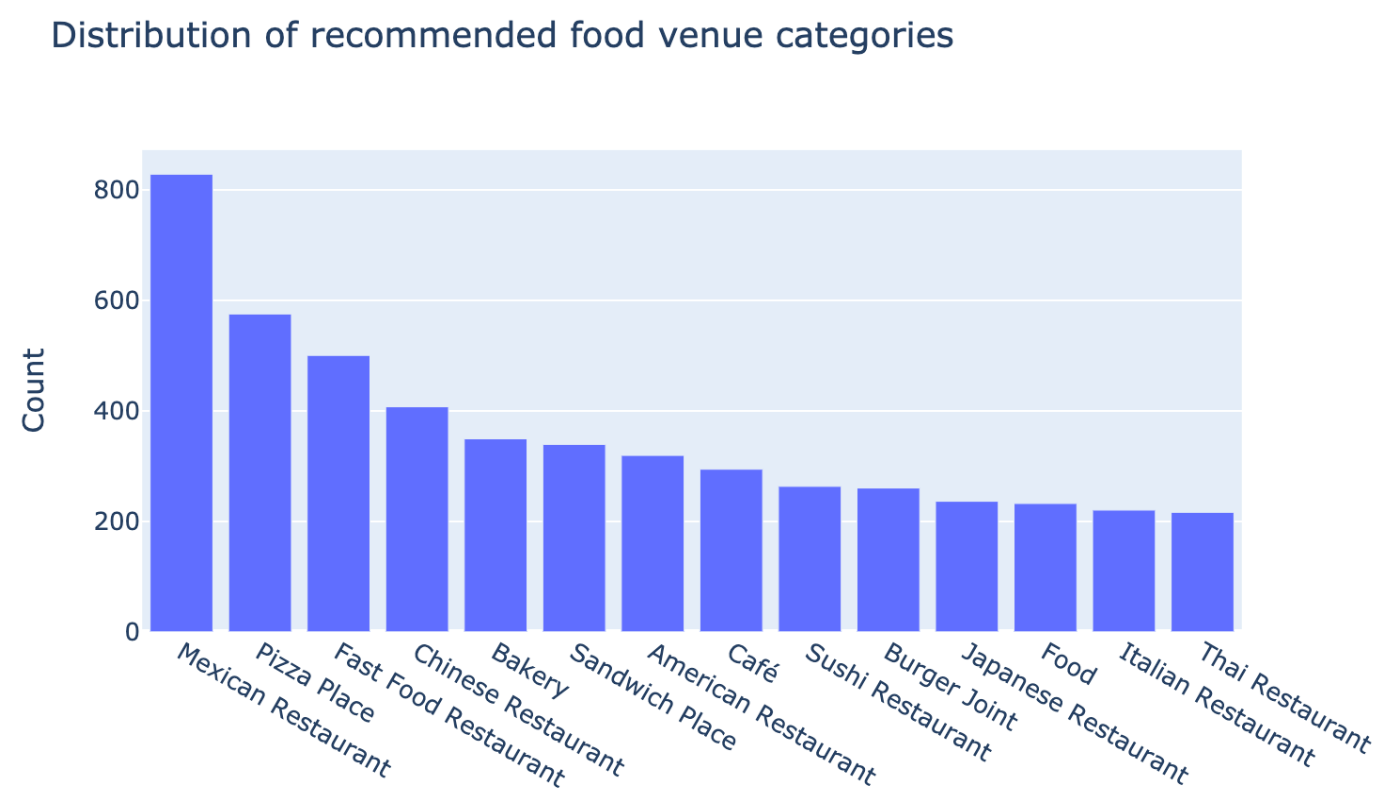


Figure - Distribution of recommmended Food Venue Categories

Mexican restaurants are by far the most often recommended type of food venue in Los Angeles County, CA. They are followed in some distance by pizza places, fast food and Chinese restaurants and bakeries.

The exact numbers of the top 10 categories are:

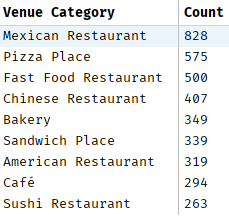


Figure - Top 10 Food Venue Categories

## Clustering the Dataset

## 3.3 Selecting a Cluster and analyzing the local Conditions in selected Areas

# Results and Discussion

Our analysis shows that there is a high variability in population, income  
and median age in the different areas of Los Angeles County, CA. So it was  
possible to identify multiple areas that fit the criteria of a relatively high  
median income and a high population. Using census and income data of all  
areas in Los Angeles County, we did a clustering to identify similar areas.  
By analysing the formed clusters there have been three areas identified  
that fit the criteria best. Norwalk, Lake View Terrace in Sylmar and Hansen  
Hills in Pacoima.

They are slightly different in income, population and age and also very  
different in their local distribution of available food venues. The final  
area could be picked on which of these factors matters to the stakeholders  
most. I will pick Norwalk as the final area for a food venue, because it  
offers the best combination of a high population and a good income out  
of these three areas.

By analysing the most recommended food venue categories across the whole  
county, we found that mexican restaurants are by far the most often  
recommended venue. After them pizza places, fast food restaurants,  
chinese restaurants and bakeries follow in that order. To give a  
recommendation for a food venue to open in Norwalk, we can compare the local  
distribution of food venues what was recommended the most in the county. By  
looking at Norwalk we found that there are already many mexican  
restaurants (12) and fast food restaurants (15). Pizza places (7) and  
chinese restaurants (5) area also recommended in a higher number, so  
opening a restaurant in one of those categories would be better, but there  
is still some competition. What stands out is that there is currently  
only one bakery recommended by Foursquare in Norwalk. Looking at the  
distribution of recommendations in the county, bakeries are the fifth  
most recommended venue category. Because of this, I would recommend  
opening a bakery in Norwalk, CA to the stakeholders.

Purpose of this analysis was to identify a possible area and food venue type  
for a new restaurant in Los Angeles County, CA based on a very limited  
amount of factors. Analysing census data and existing food venues is only  
one part on the way to find a location for opening a new restaurant. Other  
factors that also play a role are for example available spaces, rent  
costs, other venues in the area. This analysis serves as a starting point  
for finding possible locations, but further analysis needs to be done by  
the stakeholders.

# Conclusion

The purpose of this project was to find a possible location for a new  
restaurant in Los Angeles County, CA. The desire from the stakeholders was  
to identify locations that offer a good balance between median income and  
number of inhabitants, although income shall be rated slightly more important  
than population. In addition, the idea was to identify possible food  
venue categories by comparing recommended venues across the country with the  
local venues in the different areas. So for this there were census and  
income data combined to identify areas that fit the criteria. A clustering  
was performed, to group the communities in Los Angeles County using their  
income and population. Then the cluster was chosen that fit the former  
mentioned criteria the most. From this cluster the top 3 areas were chosen,  
that had the best balance between income and population. This way the best  
three candidates for a new restaurant location were identified.

The next step was to analyze the local food venue categories. For this the  
local venues in a 4 km radius were identified and grouped. This grouping was  
then compared with the distribution of the most recommended food venue  
categories across the whole county. In doing so, opportunities for new  
restaurants in any of the three chosen communities have been identified.

The final decision can be made by the stakeholders, based on the  
recommendations given in this project. This decision for a locality can be  
based on income, population or median age of the areas. The decision for a  
venue category can be based on popular venues across the county, and the gaps  
in the local food offerings that have been identified.