# Business Problem

This capstone project is going to give a recommendation in which community (defined by the postal code) one should open a restaurant in Los Angeles County, CA. In addition, there will be a recommendation given which food venue category could be chosen, based on census data and data of the existing food places in the community, and the whole county.

# 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

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

Sources:

* Median Household Income by Zip Code
  + http://www.laalmanac.com/employment/em12c.php
* Los Angeles Census Data
  + https://www.kaggle.com/cityofLA/los-angeles-census-data
* US Zip Code Latitude and Longitude
  + https://public.opendatasoft.com/explore/dataset/us-zip-code-latitude-and-longitude/information/
* Find the best areas for a new restaurant
  + A lot of inhabitants
  + More on the wealthy side
* Find what type of restaurant to open
  + City -> 50 recommendetions per Area -> Categories -> most recommended
* Visualize the food venues in the selected area (Norwalk)
* Compare some of the venues in the chosen cluster

Was als nächstes?

* Clustering aufbereiten und richtige Schlüsse ziehen
* Recommendations per Area für ganz LA abfragen und sichern

Danach

* 3 Gebiete vergleichen und 1 empfehlen
* 1 Gebiet komplett auswerten und eine Empfehlung geben, basierend auf recommendations in ganz LA
* Grafiken überlegen