

Data Science Capstone

Where to Open my Peruvian Restaurant?

Adolfo G. Ramirez-Aristizabal

Intro

In this project I will present methodology for finding the best locations to open a Peruvian restaurant in Los Angeles. The problem is framed from the perspective of helping a business owner and their investors. Currently the business owner has had great success in his own restaurant. The restaurant was opened over ten years ago and has since gone from a small ma and pa shop into a distinguished food brand in the local area. Currently the owner has been talking to investors to open a second location in Los Angeles, so that they can start having a successful franchise chain for their Peruvian food service. What we know is that Los Angeles would be very receptive to this type of food and there is a lot of commercial traffic to tap into. The problem is figuring out a location that would ease the success of the new restaurant. We will use public data sets for cities and their zip codes inside Los Angeles county to tap into Four Square data for current commercial venue location data. We have some simple criteria to figure out where our first new location should open. We know that Los Angeles county is big and diverse which gives us a lot of options, but we also want to stand out and be able to have sustainable business without having to over invest in publicity. The first two criteria we consider is to find a location that has a cluster of successful venues already. Two, we also don't want to compete with other related restaurants, so we need to find a cluster of venues that don't have Peruvian related restaurants.

Data

The first dataset we work with is coming from well documented location data from the county. Many city planners and researchers already use these data sets to find insights about such a large and highly populated areas. The dataset provides all cities in the county along with all zip codes and their latitude and longitude. Using the Four Square API we can then pass that information to retrieve venue location data. This will allow us to see what the top venues per location are. We will need to figure out how to appropriately cluster venues based on their description so that we don't get the top 5 or 10 venues to be the same or highly similar because of slightly different descriptions.