Data Analysis for starting a new restaurant in California.

## Capstone Project - The Battle of Neighborhoods

IBM Applied Data Science Capstone

Introduction

* California is a [state](https://en.wikipedia.org/wiki/U.S._state) in the [Pacific Region](https://en.wikipedia.org/wiki/Pacific_states) of the [United States of America](https://en.wikipedia.org/wiki/United_States_of_America).
* California is the [most populous U.S. state](https://en.wikipedia.org/wiki/List_of_states_and_territories_of_the_United_States_by_population) and the [third-largest](https://en.wikipedia.org/wiki/List_of_U.S._states_and_territories_by_area) by area, and is also the [world's thirty-fourth most populous](https://en.wikipedia.org/wiki/List_of_country_subdivisions_by_population) subnational entity.
* California's economy, with a [gross state product](https://en.wikipedia.org/wiki/Gross_state_product) of $3.0 trillion, is the [largest sub-national economy](https://en.wikipedia.org/wiki/List_of_country_subdivisions_by_GDP_over_200_billion_USD) in the world.
* California integrates foods, languages, and traditions from other areas across the country and around the globe.
* California's agriculture industry has the highest output of any U.S. state.



Problem

To start a new Restaurant in California which is more likely to give a good business.

Data

* Neighborhood of California

[California](https://www.california-demographics.com/cities_by_population) demographics by Cubit.

<https://www.california-demographics.com/cities_by_population>

* Geographical coordinates of neighborhood

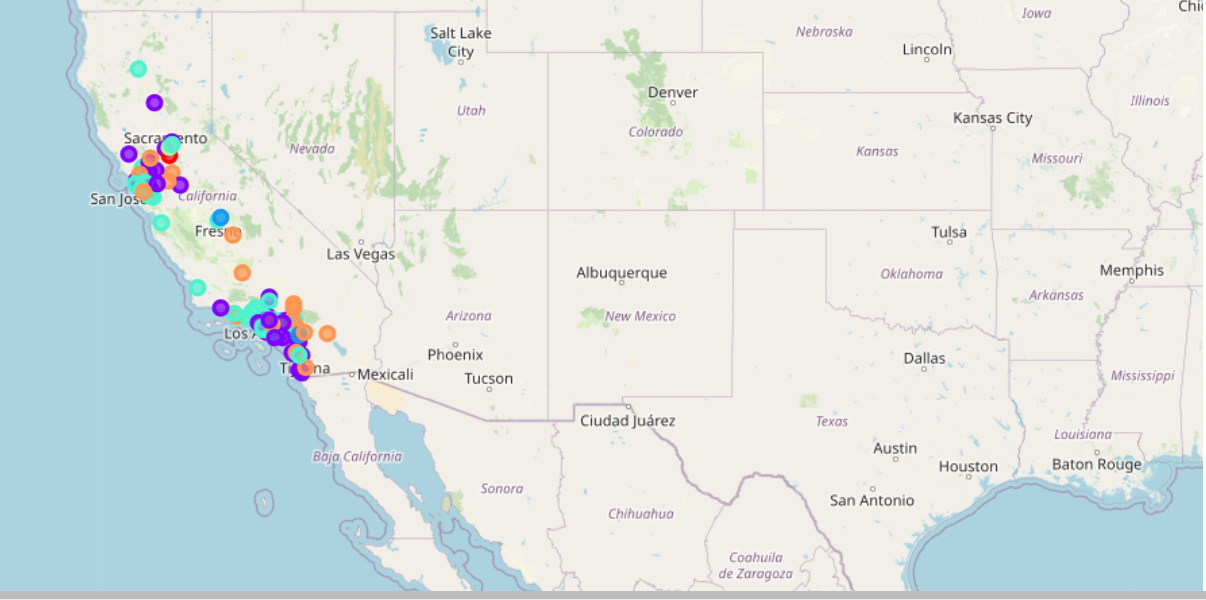
Using Geopy Library

* Venue Data from Foursquare API

Methodology

* Feature Extraction – one Hot encoding
* Unsupervised Learning – K-Means Clustering
* Plotting – Folium

Results



Visualization of clusters