

# Data Analysis for starting a new restaurant in California.

Capstone Project - The Battle of Neighborhoods

IBM Applied Data Science Capstone

Varun Kumar Anand

30-Dec-2020

# Introduction

- California is a state in the Pacific Region of the United States of America.
- California is the most populous U.S. state and the third-largest by area, and is also the world's thirty-fourth most populous subnational entity.
- California's economy, with a gross state product of \$3.0 trillion, is the largest subnational economy 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.

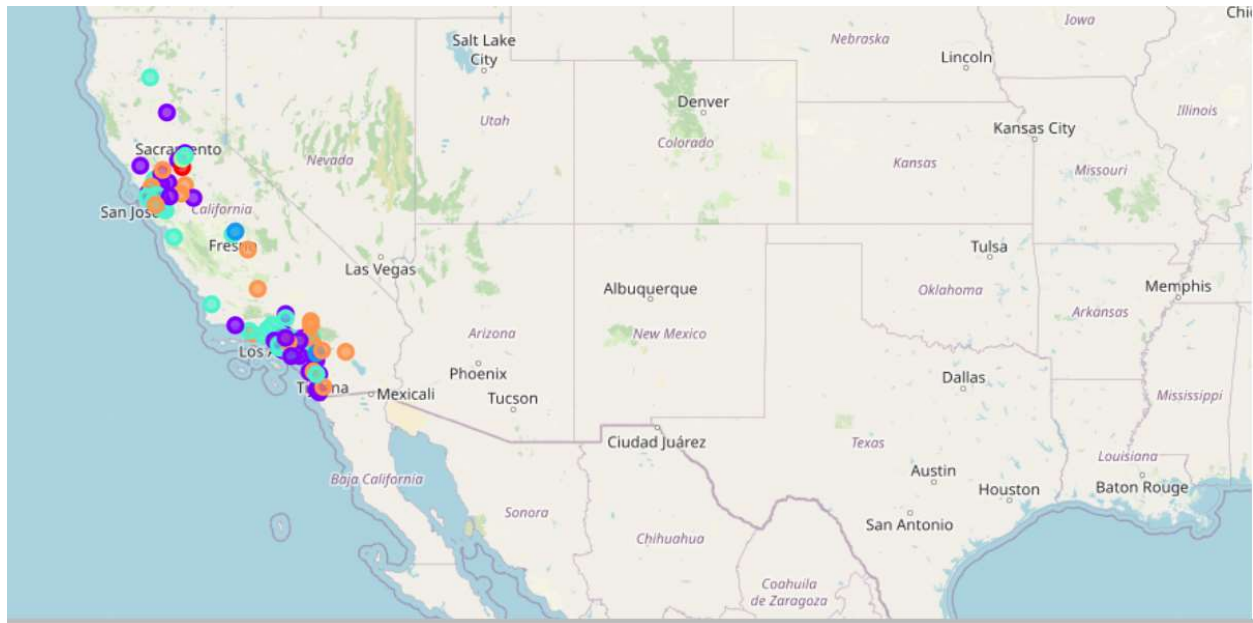
# Data

- Neighborhood of California  
California demographics by Cubit.  
[https://www.california-demographics.com/cities\\_by\\_population](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