

Applied Data Science Capstone Project

Business Problem and Data Sources

Ritobrata Ghosh¹

¹University of Calcutta
Kolkata, West Bengal, India

ritobrata-ghosh[at]outlook[dot]com

***Abstract.** This is the first phase of the capstone project where I discuss the business problem and mention the sources of the data.*

1. Business Problem

1.1. Stakeholder Information

There is a VC firm which is looking into investing in eateries in **California City, CA**. They have applications from chefs in their hands.

1.2. Stakeholder Problem

They are willing to invest in multiple eateries of different types which will face the least resistance and will be highly profitable.

1.3. Project Goal

The goal is to suggest them a list of investment options which clearly outlines the best eatery type and its respective county based on their criteria.

2. Sources of Data

For solving this problem, data from four sources will be leveraged.

2.1. Location Data

Location data titled "**California Counties**" provided in **California Open Data Portal** provided by **Government of California** for the geographical location data. This data is in .csv format.

2.2. Venues Data

The **Foursquare** API for information about established restaurants based on location.

2.3. Population Data

County-wise population data from **US Government Census site**.[\(File Link\)](#). This file is in .xlsx format. Only the latest data (year 2019) has been kept, and it has been turned into a CSV file for further cleaning.¹

¹Annual Estimates of the Resident Population for Counties in California: April 1, 2010 to July 1, 2019 (CO-EST2019-ANNRES-06) Source: U.S. Census Bureau, Population Division Release Date: March 2020

2.4. Economic Data

County-wise Real GDP data provided by [Bureau of Economic Analysis, U.S. Department of Commerce](#). ([File Link](#)). This data is also in .xlsx format. Irrelevant data has been truncated and the file has been converted to CSV format for further cleaning.

3. Example of Data Use

Example of data use has been shown in [this notebook](#).

With the data collected from these sources, analysis will be carried out for solving the business problem.