

# 1. Introduction

## 1.1. Background

It is crucial factor to pick a good location of a coffee shop as it can be considered a vital factor for its survival within the first years of its opening. It would be safe to say that within the first years when there is no usual customer base and loyalty, location of the coffee shop makes it visible to the passing by people and to the customer audience. However, we should not forget about the fact that it is most necessary to make it available to the target customers and make it in easily accessible location.

## 1.2. Business Problem

The aim of the client is to open a small coffee shop for the youth, especially the university students and young adults as a quiet place to visit, to study, to have productive time or just to hang out with the groupmates. In this project the California state is considered and Los Angeles city is aimed location. With the usage of data we will also be able to see why it is a good decision to open this shop exactly in Los Angeles.

The main objective of this Capstone Project is to locate this coffee shop in optimal location where mostly university students are located. Using the California Census Data we will consider the densest population for the 20-30 years old US Citizens across neighborhoods, furthermore discussing the borough location where the most universities in LA is located.

For the data analysis thorough California Census Data and Los Angeles Education Facilities data is used from the [Los Angeles Open Data Portal](#). For geolocating our objects Foursquare API will be utilized.

## 1.3. Interests

Other businesses and non-governmental organizations with similar target audience (youth, university students) can utilize this analysis to understand the densest youth population across California.

## 2. Data acquisition and Cleaning

### 2.1. Data Sources

Estimated population by Census Tract, city name, ethnicity, gender and age group data for Los Angeles County citizens is obtained from this [link](#); all officially registered education facilities list in Los Angeles city is taken from this [link](#), both from the Los Angeles Open Data portal. Geographical coordinates, latitude and longitude data for each neighborhood in Los Angeles is obtained using this [link](#) from the LA City Geohub website.

Foursquare API explore function is used to receive the facilities across the neighborhoods of Los Angeles according to the coordinates. For more information on developer applications of Foursquare API this [link](#) can be useful.

### 2.2. Data cleaning

### 2.3. Feature selection

## 3. Exploratory Data Analysis

### 3.1. Folium mapping

## 4. Predictive Modeling

### 4.1. Data Pre-processing

### 4.2. k-Means Clustering

## 5. Conclusions

## 6. Future directions

## 7. References