Los Angeles Neighbourhood analysis for setting up for Indian cuisine

1. Introduction

Los Angles is very vibrant, and it is multi-cultural and famous for Hollywood entertainment industry. Los Angeles is known for its Mediterranean climate and its sprawling metropolis. Los Angeles lies in a basin in Southern California, adjacent to the Pacific Ocean, with mountains as high as 10,000 feet (3,000 m), and deserts.

I had chance to live in Rosemead, Los Angeles for more than year. Hence, I have fair idea of the LA neighbourhoods and food chains in greater LA. This data science project is intended to analyse various communities, population, area income limits, age and ethnic groups and pre-existing restaurants and come up with five possible location recommendations for starting Indian cuisine restaurant.

1.1 Problem statement

This data science project is expected to recommend best possible five locations to start Indian cuisine by studying the greater LA demographics, neighbourhood communities, income and various other factors including existing restaurants and their popularity.

The greater LA is one of the largest cities in united states, and there are different community-based neighbourhoods, it is important to apply data science and arrive at the best possible location for the success of the business.

1.2 Data Acquisition

There are various organizations provides vital information about the Greater Los Angeles demographics including income limits, population, age group and current restaurants with geocodes in neighbourhoods and their corresponding ratings

By applying statistical analysis and ML techniques, it is obvious that meaningful insights can be built and arrive at the best locations for stating Indian cuisine.

1.3 Data Sets

Data Source	Details
https://usc.data.socrata.com/ 	✓ Total population – For each neighbourhood (from 2010 till 2018) ✓ Rental price for each neighbourhood ✓ Age Distribution ✓ Area income limits ✓ Census track location for each neighbourhood ✓ Households (Single/Family/size)
Foursquare 	✓ Existing restaurants, ratings and reviews
Google Geocode API	Leverage google Geocode API for getting the geolocations of the neighbourhood (only if needed, most of the geocode details are already available provided by https://usc.data.socrata.com/

1.4 Sample Datasets

Geo code for Neighbourhoods

GEOID	Tract	Tract Number	Neighborhood	Location	Latitude	Longitude
1400000US06037101110	Census Tract 1011.10, Los Angeles County, California	101110	Tujunga	(34.2595555, -118.293602)	34.25956	-118.294
1400000US06037102103	Census Tract 1021.03, Los Angeles County, California	102103	Shadow Hills	(34.224155, -118.354339)	34.22416	-118.354
1400000US06037102105	Census Tract 1021.05, Los Angeles County, California	102105	Sun Valley	(34.210852, -118.3480495)	34.21085	-118.348
1400000US06037102107	Census Tract 1021.07, Los Angeles County, California	102107	Shadow Hills	(34.2412955, -118.3292705)	34.2413	-118.329
1400000US06037103101	Census Tract 1031.01, Los Angeles County, California	103101	Sunland	(34.274431, -118.30714)	34.27443	-118.307
1400000US06037103102	Census Tract 1031.02, Los Angeles County, California	103102	Sunland	(34.262834, -118.30683)	34.26283	-118.307
1400000US06037103200	Census Tract 1032, Los Angeles County, California	103200	Lake View Terrace	(34.2745565, -118.3439025)	34.27456	-118.344
1400000US06037103300	Census Tract 1033, Los Angeles County, California	103300	Shadow Hills	(34.255439, -118.3527775)	34.25544	-118.353

Age Distribution

Policy A	Dataset	Variable	Year	Percent	Count	Tract	Tract N	Neighb	GEOID	Locatio	Row ID	Date	Denom	Denom	tor Descrip
Demograp	Age Distribution	Population Ages 18-24	2013	5.656411	296	Census Tr	405400	Vincent	1400000U	(34.09233	Population	1/1/2013	5233	Total Population	
Demograp	Age Distribution	Population Ages 18-24	2013	10.46875	871	Census Tr	432500	El Monte	1400000U	(34.09491	Population	1/1/2013	8320	Total Population	
Demograp	Age Distribution	Population Ages 18-24	2013	9.909199	251	Census Tr	408004	West Covi	1400000U	(34.03781	Population	1/1/2013	2533	Total Population	
Demograp	Age Distribution	Population Ages 18-24	2013	7.356459	123	Census Tr	550400	Downey	1400000U	(33.93576	Population	1/1/2013	1672	Total Population	
Demograp	Age Distribution	Population Ages 18-24	2013	12.61785	629	Census Tr	533401	Maywood	1400000U	(33.99366	Population	1/1/2013	4985	Total Population	
Demograp	Age Distribution	Population Ages 25-34	2013	25	15	Census Tr	532400	Vernon	1400000U	(34.00227	Population	1/1/2013	60	Total Population	
Demograp	Age Distribution	Population Ages 18-24	2013	13.42503	693	Census Tr	541802	Lynwood	1400000U	(33.90983	Population	1/1/2013	5162	Total Population	
Demograp	Age Distribution	Population Ages 18-24	2013	13.16644	852	Census Tr	532900	Florence-F	1400000U	(33.97905	Population	1/1/2013	6471	Total Population	

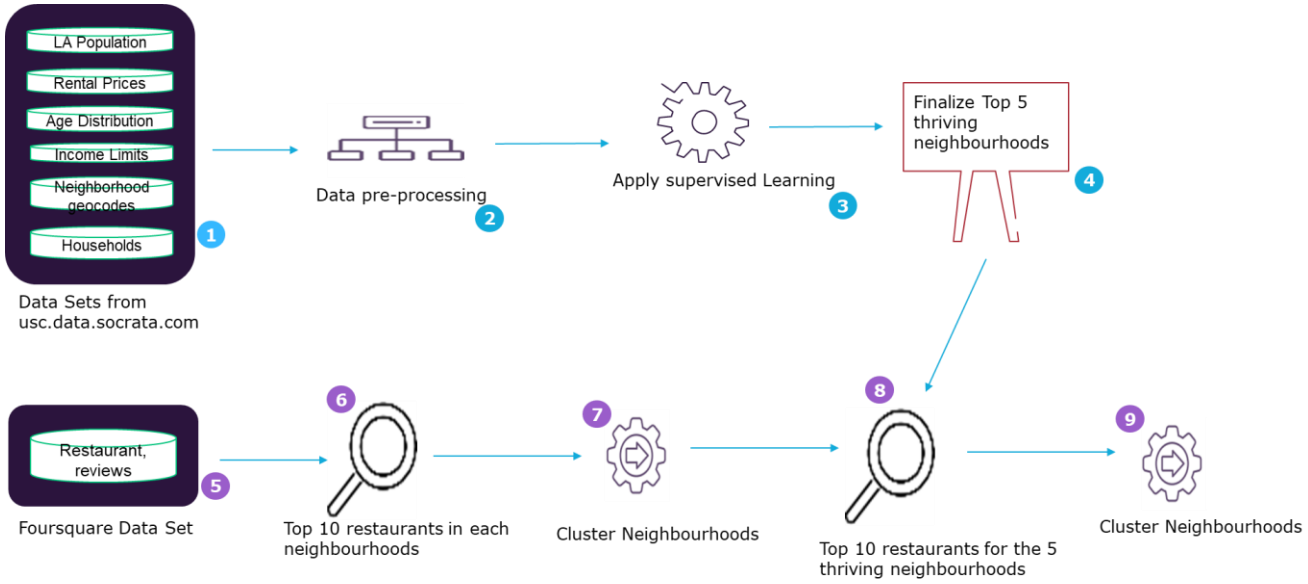
Total population

Policy Area	Dataset	Variable	Year	Count	Tract	Tract N	Neighborhood	GEOID	Locatio	Row ID	Date
Demography	Total Population	Total Populatio	2010	5017	Census Tr	101110	Tujunga	1400000U	(34.25947	Total_Pop	1/1/2010
Demography	Total Population	Total Populatio	2010	1841	Census Tr	102103	Shadow Hills	1400000U	(34.22508	Total_Pop	1/1/2010
Demography	Total Population	Total Populatio	2010	1525	Census Tr	102105	Sun Valley	1400000U	(34.20987	Total_Pop	1/1/2010
Demography	Total Population	Total Populatio	2010	3503	Census Tr	102107	Shadow Hills	1400000U	(34.24047	Total_Pop	1/1/2010
Demography	Total Population	Total Populatio	2010	2917	Census Tr	103101	Sunland	1400000U	(34.27324	Total_Pop	1/1/2010
Demography	Total Population	Total Populatio	2010	4470	Census Tr	103102	Sunland	1400000U	(34.26338	Total_Pop	1/1/2010
Demography	Total Population	Total Populatio	2010	3378	Census Tr	103300	Shadow Hills	1400000U	(34.25738	Total_Pop	1/1/2010

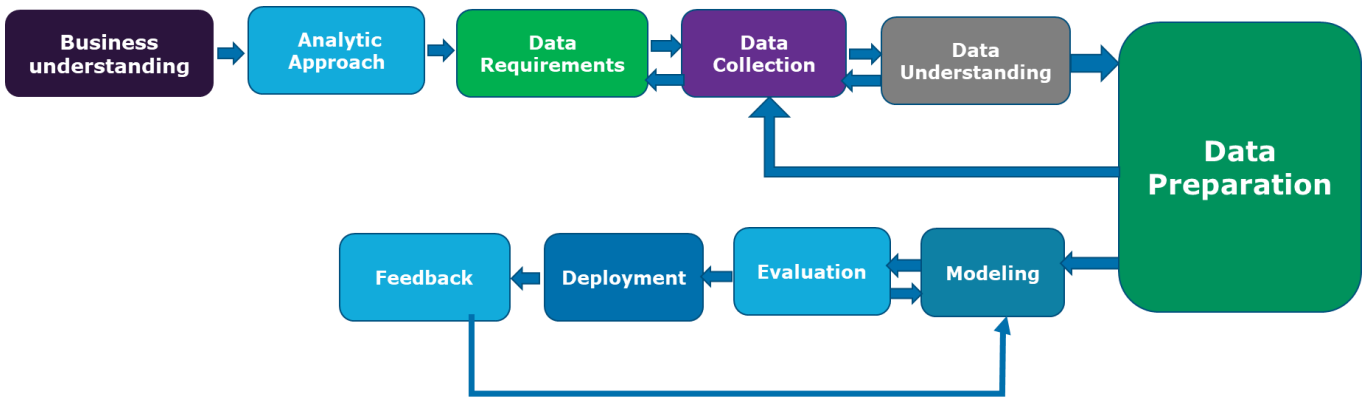
1.5 Data Processing and Machine Learning techniques

1. All the data sets from usc.data.socrata.com has GEOID which is the primary key to join all the data sets.
2. The restaurants, ratings data will be mapped using “Neighbourhood names” and then linked with data source mentioned in the point no 1.

The below diagram depicts step-by-step process flow which will be followed during the course of implementation.



Following Data Science methodology will be applied for building ML model, evaluate and deployment.



1.6 References

The Neighbourhood Data for Social Change (NDSC) platform
https://usc.data.socrata.com/stories/s/htr6-r22g

List of districts and neighbourhoods in Los Angeles
https://en.wikipedia.org/wiki/List_of_districts_and_neighborhoods_in_Los_Angeles

Author: Genesis Robinson