

[Open in app](#)

Search



Write

♦ Member-only story

# Los Angeles Neighborhood Analysis

Data Analysis and Visualizations using Python



Chaitanya Krishna Kasaraneni · Follow

Published in Towards Data Science · 5 min re



13

**Towards Data Science**

Your home for data science. A

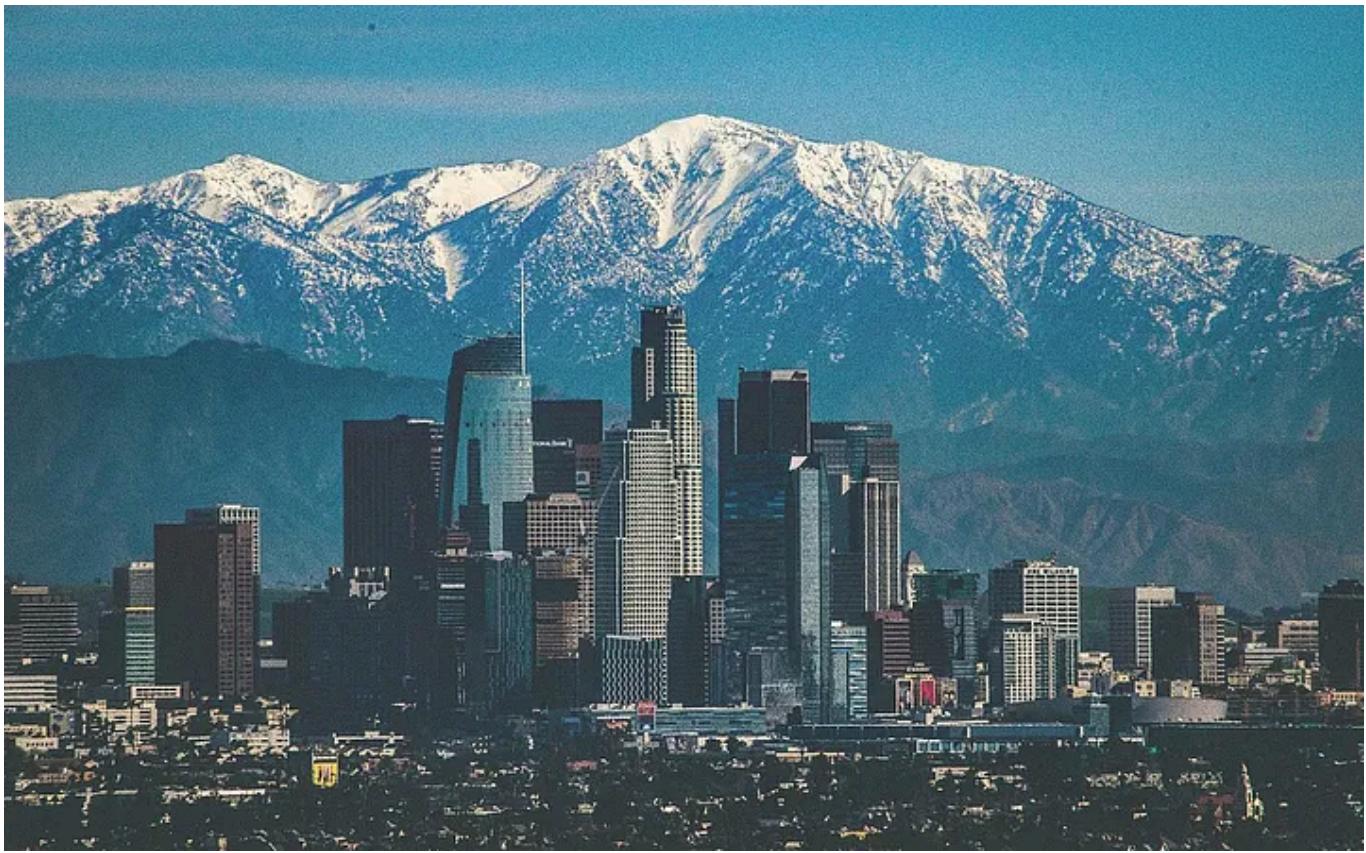
Medium publication sharing  
concepts, ideas and codes.

Followed by 685K people

[Follow](#)

*This project is a part of [“IBM Data Science Professional Certificate on Coursera”](#).*

*You can check my Jupyter Notebook on [GitHub](#)*



Los Angeles Sky Line (Image Source: [Wikipidea](#))

## Introduction

Los Angeles is a very vibrant city with a lot of neighborhoods, each with unique character. Some neighborhoods are quiet and cozy, has convenient store locations, while others offer a lot of fun and nightlife activities.

Choosing a neighborhood to live in or open a business can be a complicated task to do, but with the help of location data from Foursquare and crime data, we can make it a little bit easier.

## Business Problem

The objective of this capstone project is to analyze and select the best locations in the city of Los Angeles, California to choose a neighborhood to live in or open a new business. Using data science methodology and machine learning techniques like clustering, this project aims to provide

solutions to answer the business question: In the city of Los Angeles, California, what would be a better place to live in or start a business?

## Target Audience

- People interested in moving to Los Angeles and looking for a perfect neighborhood for their needs
- Business owners looking to expand their business to a new location
- A beginner data scientist who may use this research as an example

## Data

For this project, the following data is needed:

- List of neighborhoods in Los Angeles
- Latitude and longitude coordinates of neighborhoods to get the venue data
- Crime data in Los Angeles
- Venues Details

## Data Sources

### Location Data

First, we need to get a full list of all LA neighborhoods. Wikipedia article [List of districts and neighborhoods in Los Angeles](#) is a great place to start.

For geolocation data, we will use Google's Geocoding API. To get more information about it, follow the Geocoding Developer Guide.

### Venues Data (Foursquare API)

Foursquare API provides information about venues and geolocation. We will use Foursquare API to get the venue data for LA neighborhoods.

Foursquare has one of the largest databases of 105+ million places and is used by over 125,000 developers. Foursquare API will provide many categories of the venue data such as name, location, hours, rating, prices, etc.

## Crime Data

To analyze criminal activity for each neighborhood we use Los Angeles Crime & Arrest Data: from Beginning 2020 to Present dataset from LA City Website. It contains information about location, time, category and other miscellaneous data from the LA Police Department.

## Analysis

### Location Data

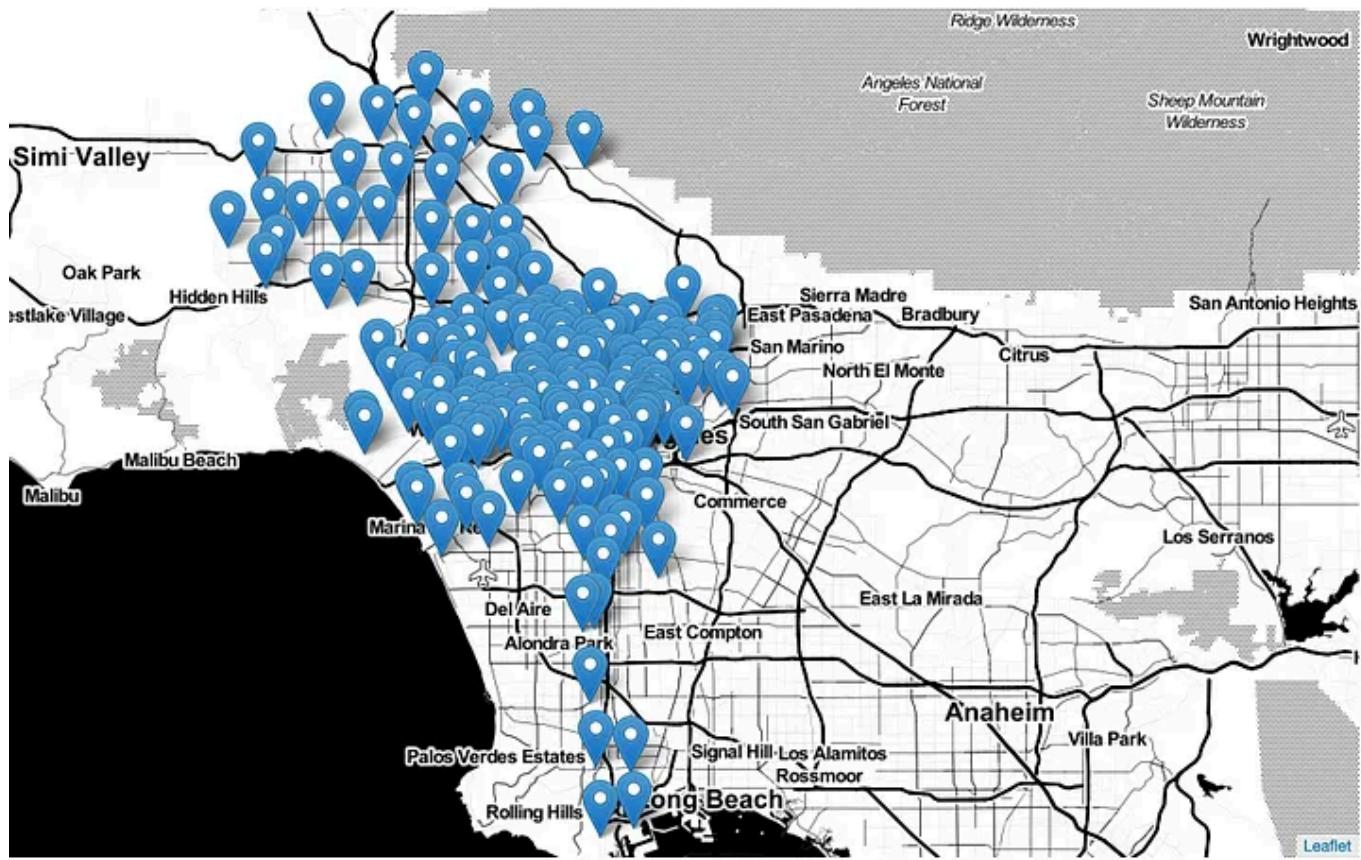
Using BeautifulSoup, a Python library used for pulling data out of HTML we parse the Wikipedia page to get the list of neighborhoods and districts in Los Angeles.

Using Google's Geocoding API, we collect the location data such as Latitudes and Longitudes of each neighborhood and store them into a pandas dataframe.

	Neighborhood	Latitude	Longitude
0	Angelino Heights	34.070289	-118.254796
1	Angeles Mesa	33.995562	-118.321977
2	Angelus Avenue	34.085498	-118.268441
3	Arleta	34.250459	-118.433835
4	Arlington Heights	34.042222	-118.318889
5	Arts District	34.041895	-118.232645
6	Atwater Village	34.117290	-118.261433
7	Baldwin Hills	34.006677	-118.350578
8	Baldwin Hills Crenshaw	34.010754	-118.337003
9	Baldwin Village	34.015091	-118.347656

DataFrame containing location data

Using Folium Geospatial visualization library, we then plot these Los Angeles neighborhoods on the map.



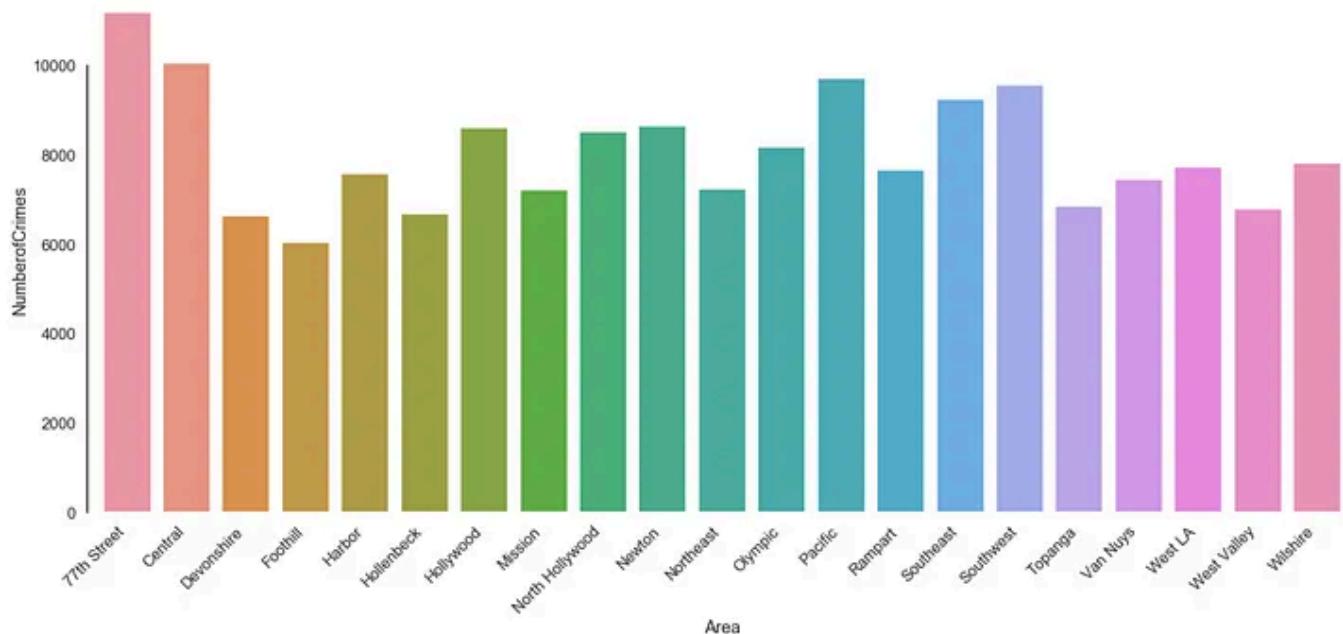
Los Angeles Neighborhoods plotted on a map

**Note:** For interactive maps, please open the Jupyter Notebook in [this website](#) by pasting the notebook's GitHub URL.

## Crime Data

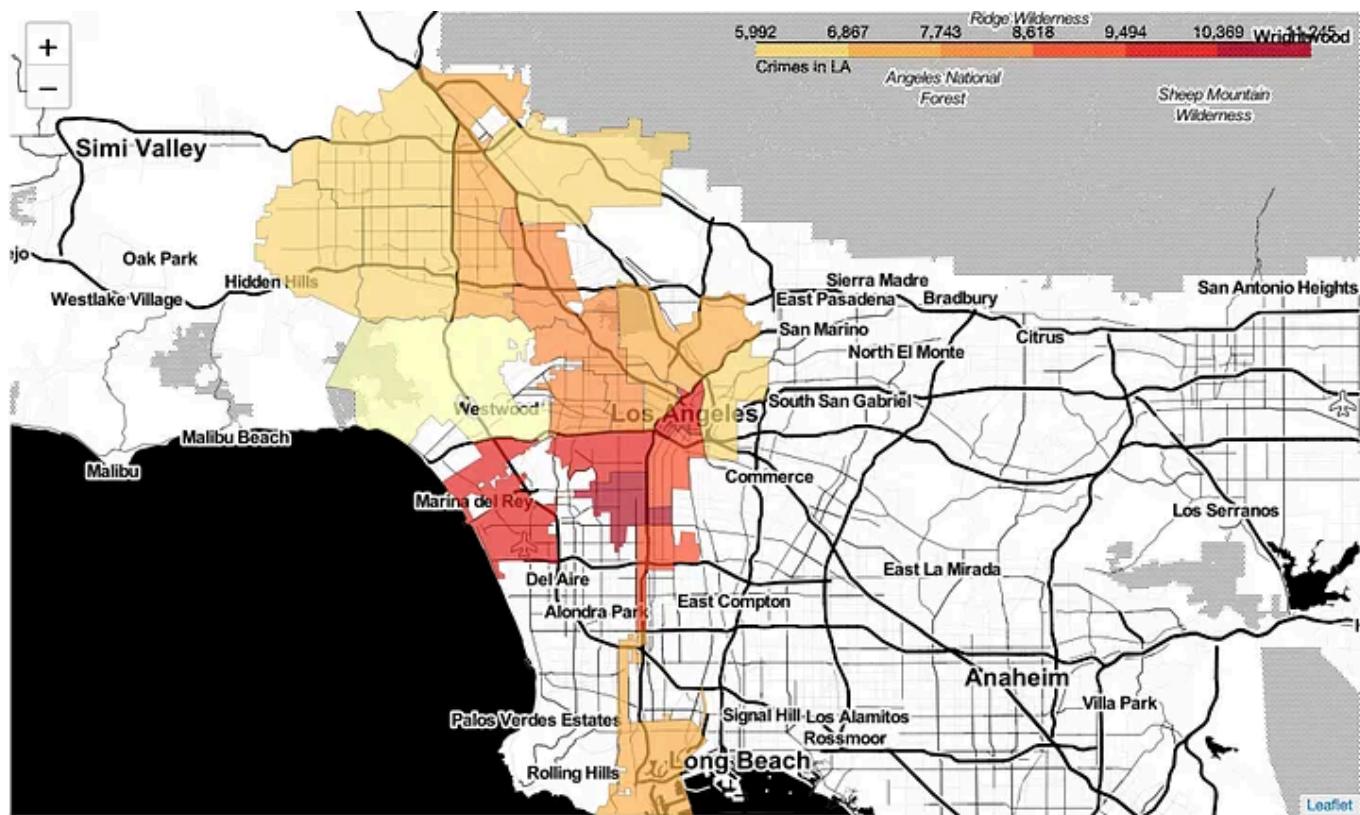
Now, we collect the crime data of 21 divisions of LAPD and load the data into a pandas dataframe. This data includes features like Date the case was reported, area, crime committed, etc.

Counting the number of crimes for each community Police station and plotting it as a graph, we have:



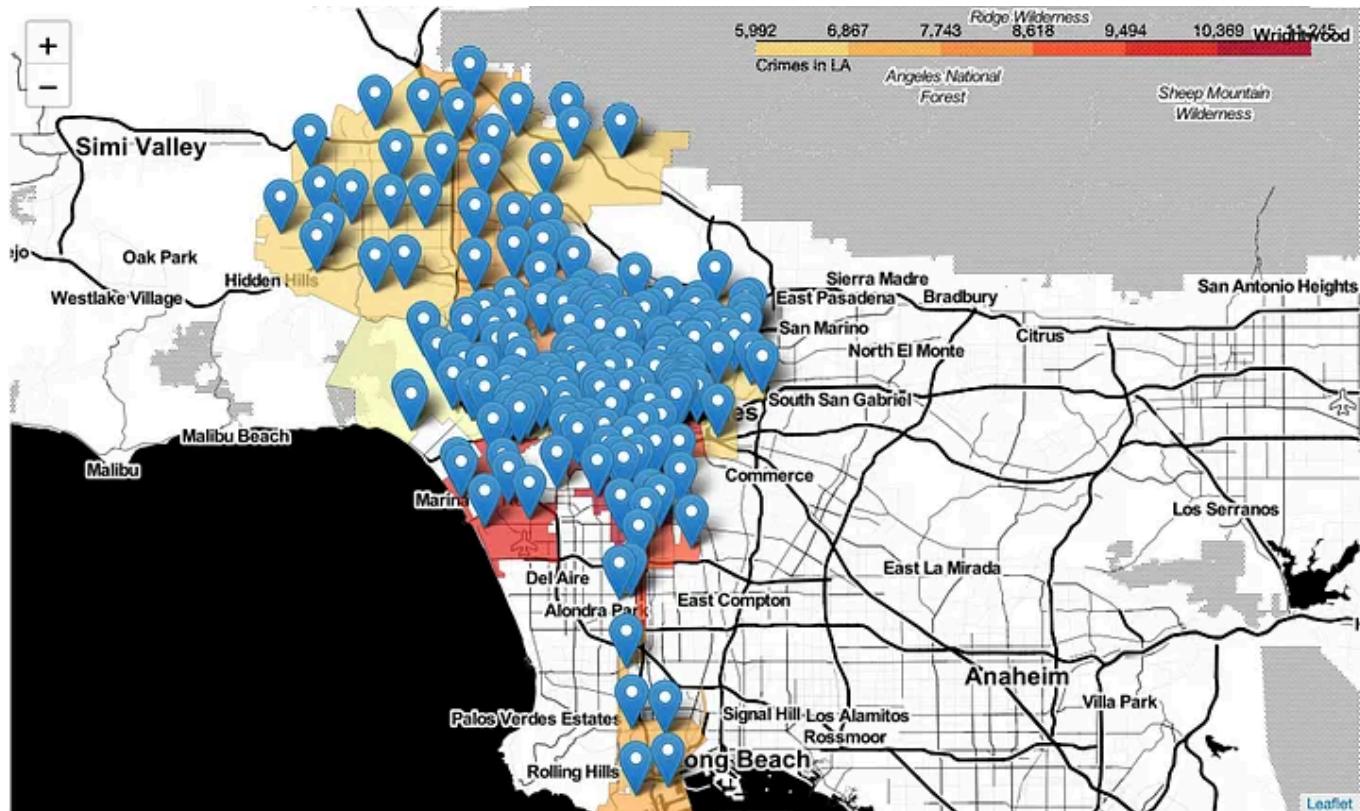
Graph for Area vs Number of Crimes

It can be seen that 77th Street division has the most number of reported cases followed by Central division. Let's plot this information on a choropleth map using folium library. The boundary data for LAPD divisions is taken from [this website](#)



Choropleth Map based on Number of Crimes

Adding the neighborhoods from previously stored neighborhoods dataframe

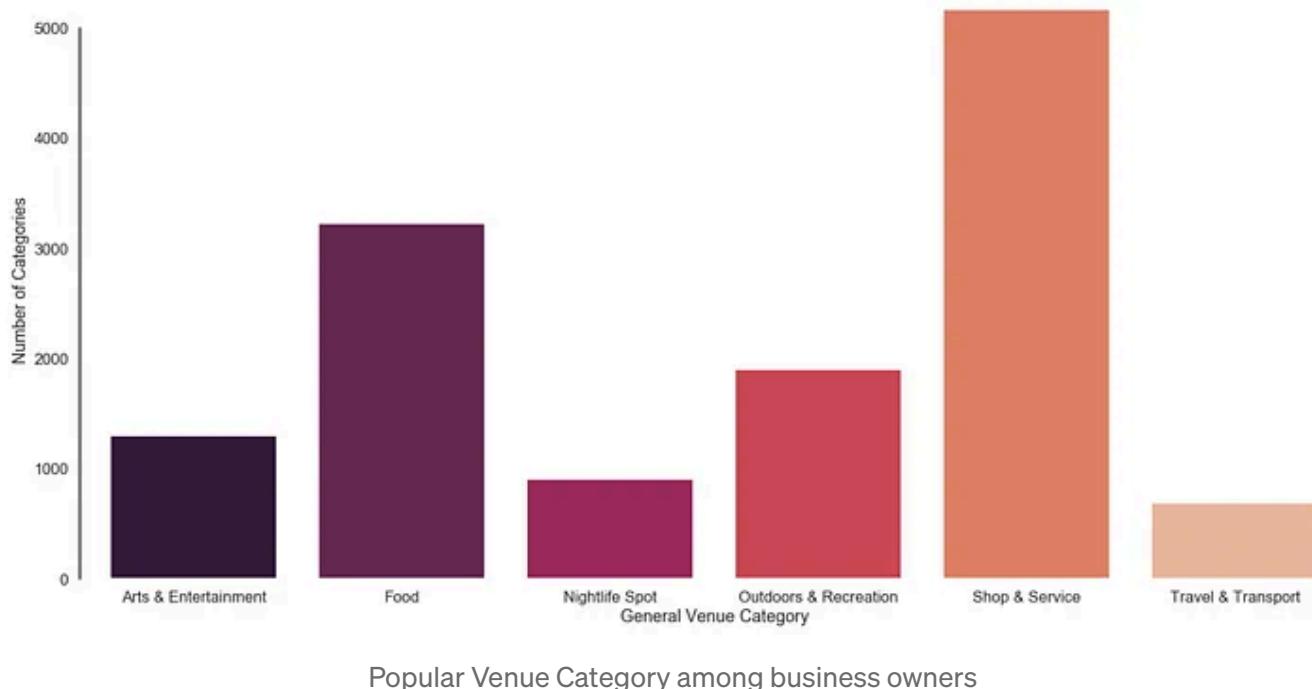


## Venue Data

Taking the neighborhoods information from the Location data, we gather the data of venues using Foursquare API and load it into a pandas dataframe. We then classify the venues into 6 general categories:

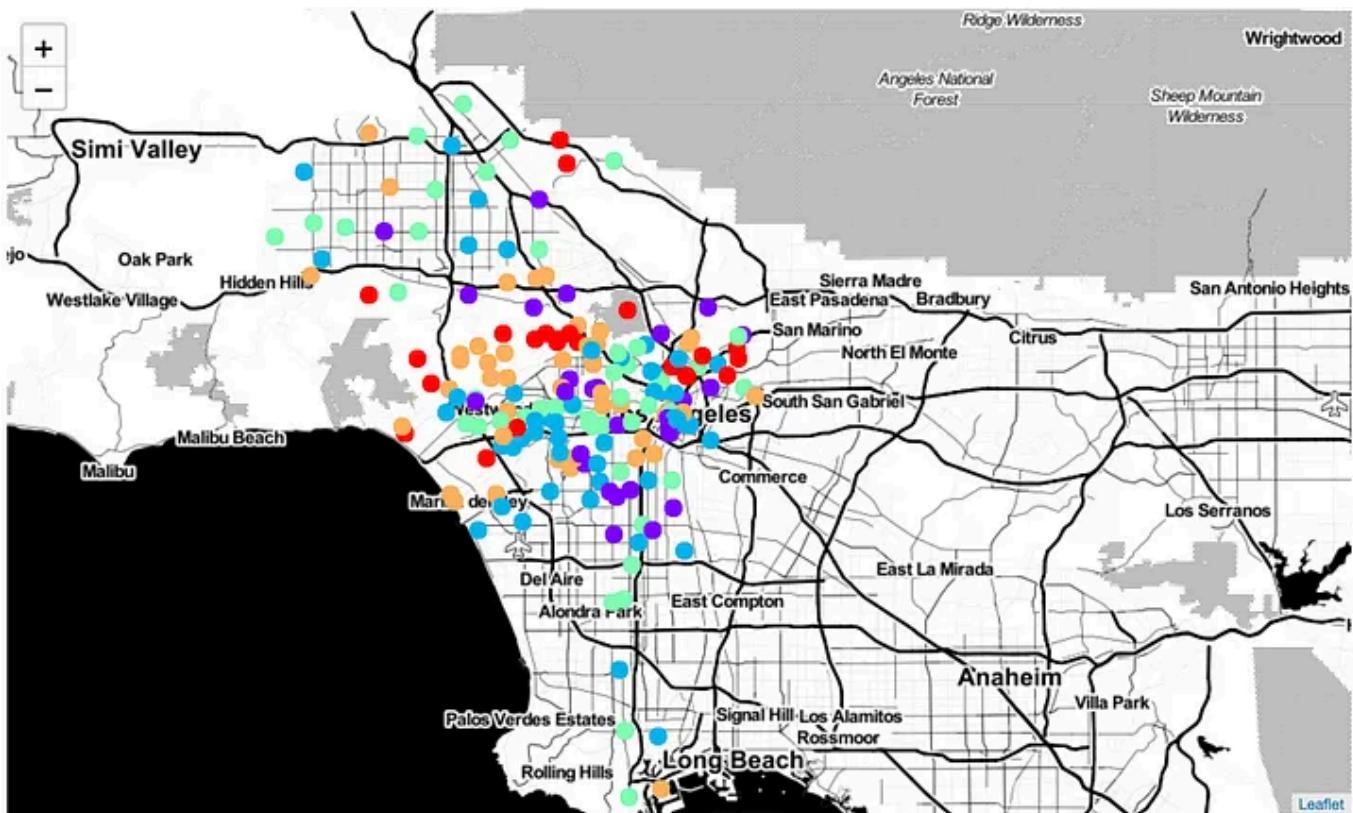
- Shop & Service
- Outdoors & Recreation
- Travel & Transport
- Food
- Nightlife Spot
- Arts & Entertainment

Let us plot the count of each venue category

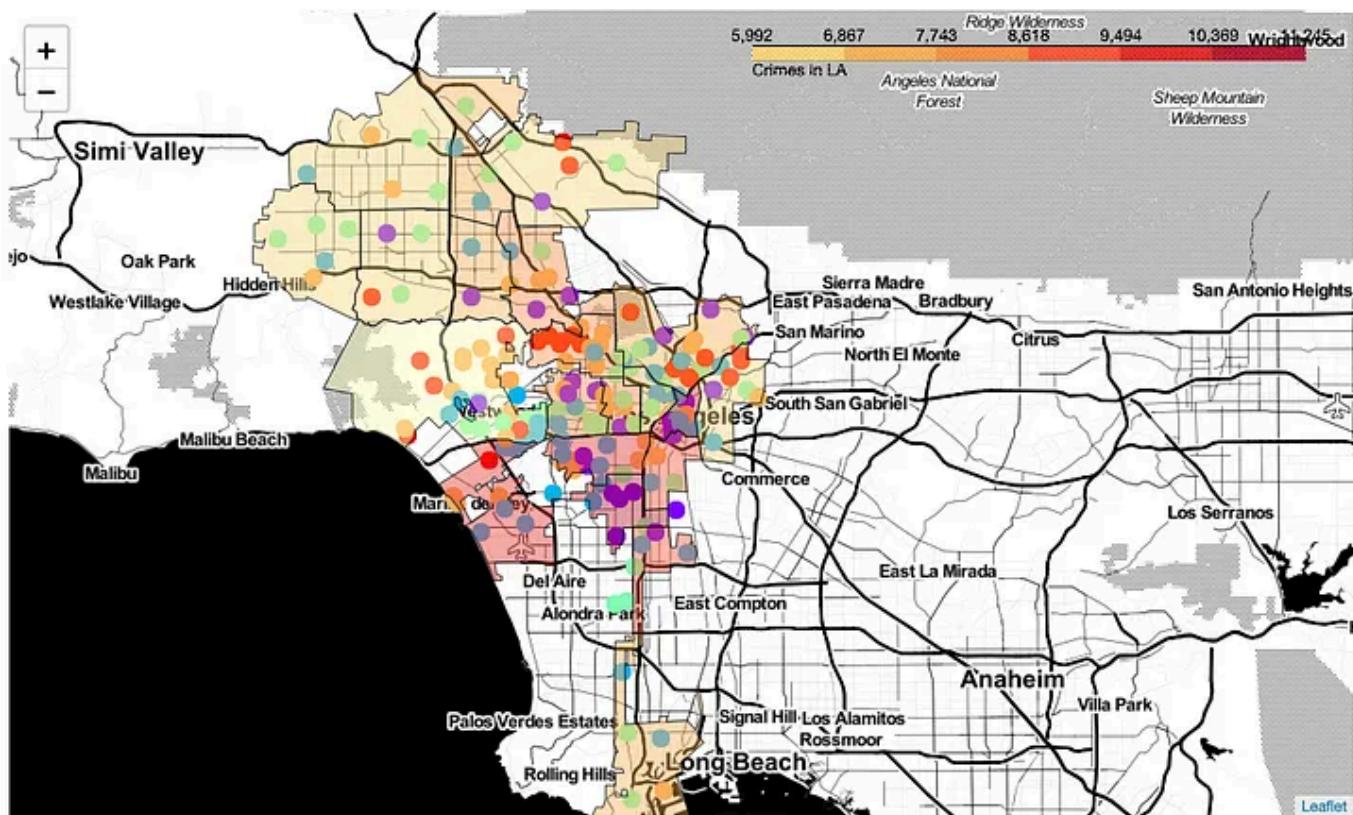


Of all the categories, “Shop & Service” venue category has many outlets i.e. popular among business owners, followed by “Food”

Clustering these venues into 5 clusters using k-Means algorithm and plotting them onto a map, we have



Plotting these cluster on the LAPD divisions choropleth map



Clusters of Venues plotted on the Choropleth Map based on Number of Crimes

## Cluster Analysis

### Cluster — 1

	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue
0	North Beverly Park	Outdoors & Recreation	Food	Arts & Entertainment	Travel & Transport	Nightlife Spot	Shop & Service
1	Cheviot Hills	Outdoors & Recreation	Food	Shop & Service	Nightlife Spot	Arts & Entertainment	Travel & Transport
2	Crestwood Hills	Outdoors & Recreation	Arts & Entertainment	Food	Shop & Service	Travel & Transport	Nightlife Spot
3	Elysian Heights	Outdoors & Recreation	Shop & Service	Arts & Entertainment	Nightlife Spot	Travel & Transport	Food
4	Elysian Park	Outdoors & Recreation	Arts & Entertainment	Food	Shop & Service	Nightlife Spot	Travel & Transport

Cluster — 1 DataFrame

### Cluster — 2

	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue
0	Atwater Village	Shop & Service	Food	Arts & Entertainment	Outdoors & Recreation	Nightlife Spot	Travel & Transport
1	Baldwin Hills Crenshaw	Shop & Service	Food	Travel & Transport	Outdoors & Recreation	Arts & Entertainment	Nightlife Spot
2	Chesterfield Square	Shop & Service	Food	Travel & Transport	Outdoors & Recreation	Arts & Entertainment	Nightlife Spot
3	Chinatown	Shop & Service	Food	Arts & Entertainment	Travel & Transport	Outdoors & Recreation	Nightlife Spot
4	Crenshaw	Shop & Service	Food	Arts & Entertainment	Outdoors & Recreation	Travel & Transport	Nightlife Spot

Cluster — 2 DataFrame

## Cluster — 3

	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue
0	Angelino Heights	Shop & Service	Food	Arts & Entertainment	Nightlife Spot	Outdoors & Recreation	Travel & Transport
1	Angeles Mesa	Shop & Service	Food	Nightlife Spot	Travel & Transport	Outdoors & Recreation	Arts & Entertainment
2	Angelus Avenue	Shop & Service	Food	Arts & Entertainment	Outdoors & Recreation	Nightlife Spot	Travel & Transport
3	Arts District	Shop & Service	Arts & Entertainment	Food	Outdoors & Recreation	Nightlife Spot	Travel & Transport
4	Beverly Grove	Shop & Service	Food	Outdoors & Recreation	Arts & Entertainment	Travel & Transport	Nightlife Spot
5	Boyle Heights	Food	Shop & Service	Nightlife Spot	Outdoors & Recreation	Travel & Transport	Arts & Entertainment

Cluster — 3 Data Frame

## Cluster — 4

	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue
0	Arlata	Shop & Service	Food	Outdoors & Recreation	Travel & Transport	Arts & Entertainment	Nightlife Spot
1	Arlington Heights	Shop & Service	Food	Arts & Entertainment	Nightlife Spot	Outdoors & Recreation	Travel & Transport
2	Beverlywood	Shop & Service	Food	Outdoors & Recreation	Travel & Transport	Arts & Entertainment	Nightlife Spot
3	South Broadway & Manchester Avenue	Shop & Service	Food	Nightlife Spot	Travel & Transport	Outdoors & Recreation	Arts & Entertainment
4	Canoga Park	Shop & Service	Food	Outdoors & Recreation	Travel & Transport	Arts & Entertainment	Nightlife Spot

Cluster — 4 Data Frame

## Cluster — 5

	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue
0	Baldwin Hills	Outdoors & Recreation	Arts & Entertainment	Nightlife Spot	Shop & Service	Food	Travel & Transport
1	Baldwin Village	Shop & Service	Nightlife Spot	Outdoors & Recreation	Food	Arts & Entertainment	Travel & Transport
2	Baldwin Vista	Shop & Service	Outdoors & Recreation	Food	Arts & Entertainment	Nightlife Spot	Travel & Transport
3	Beachwood Canyon	Outdoors & Recreation	Arts & Entertainment	Nightlife Spot	Shop & Service	Travel & Transport	Food
4	Bel Air	Food	Outdoors & Recreation	Nightlife Spot	Shop & Service	Arts & Entertainment	Travel & Transport

Cluster — 5 DataFrame

## Observations

- All the venues can be grouped into 5 clusters
- Of all the clusters, Cluster 1 has least number of neighborhoods (23) and “Outdoor & Recreation” venue category is the most popular among neighborhoods in Cluster-1
- “Shop & Service” venue category is the most popular among neighborhoods in clusters 2,3 & 4
- Among all the venue categories, “Shop & Service” is the most popular category
- “Food” seems to be the second popular venue category followed by “Entertainment”
- The neighborhoods that come under Pacific, 77th Street and Southwest LAPD community divisions have higher number of crimes recorded

## Conclusion

In this project, we analyzed the neighborhoods in Los Angeles. The neighborhoods data was scraped from Wikipedia using BeautifulSoup. Then

using Google's Geocoding API and Folium maps, we plotted these neighborhoods on a map.

Next, we analyzed LAPD crime data and plotted it on a Choropleth map along with the neighborhoods to know the neighborhoods where the number of crimes is higher

Then using Foursquare API, we gathered the details of venues in the neighborhoods and divided them into 5 clusters. Finally, we plotted these on a map along with the crimes map.

- **Note:** For interactive maps, please open the Jupyter Notebook in [this website](#) by pasting the notebook's GitHub URL.

Data Science

Data Visualization

Data Analysis

Los Angeles



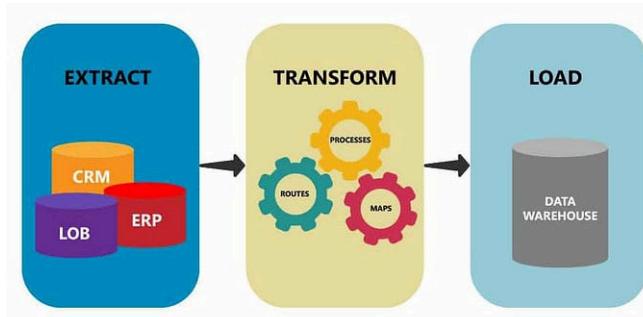
**Written by Chaitanya Krishna Kasaraneni**

77 Followers · Writer for Towards Data Science

Software Engineer — Data at Egen <https://chaitanyak.me>

Follow

## More from Chaitanya Krishna Kasaraneni and Towards Data Science



Chaitanya Krishna Kasaraneni in Analytics Vidhya

### Understanding ETL Pipeline

An introduction to phases of ETL pipeline in Data Science

5 min read · Apr 26, 2020

59

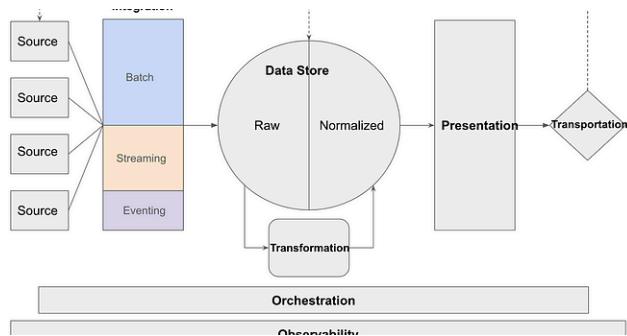
Cristian Leo in Towards Data Science

### The Math behind Adam Optimizer

Why is Adam the most popular optimizer in Deep Learning? Let's understand it by diving...

16 min read · Jan 30, 2024

2.4K 20



Dave Melillo in Towards Data Science

## Building a Data Platform in 2024

How to build a modern, scalable data platform to power your analytics and data science...

9 min read · Feb 5, 2024

1.6K 26

Chaitanya Krishna Kasaraneni in Towards Data Science

## Build kNN from scratch - Python

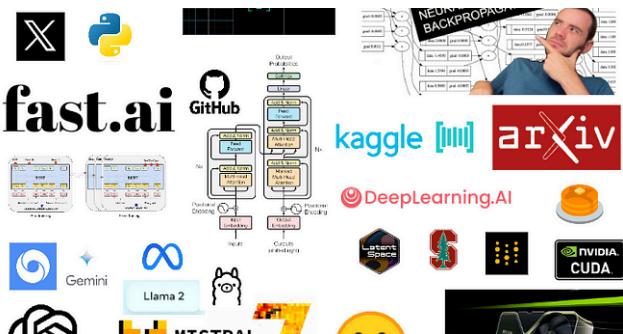
With k-Fold cross-validation (from scratch)

· 6 min read · Sep 12, 2020

See all from Chaitanya Krishna Kasaraneni

See all from Towards Data Science

## Recommended from Medium





Benedict Neo in bitgrit Data Science Publication



Cristian Leo in Towards Data Science

## Roadmap to Learn AI in 2024

A free curriculum for hackers and programmers to learn AI

11 min read · Feb 21, 2024

4.3K

54



## The Math behind Adam Optimizer

Why is Adam the most popular optimizer in Deep Learning? Let's understand it by diving...

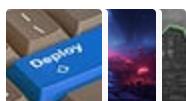
16 min read · Jan 30, 2024

2.4K

20



## Lists



### Predictive Modeling w/ Python

20 stories · 973 saves



### Practical Guides to Machine Learning

10 stories · 1157 saves



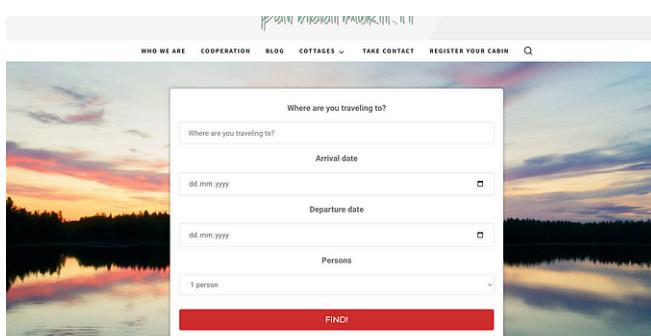
### ChatGPT prompts

45 stories · 1214 saves



### Coding & Development

11 stories · 485 saves



Artturi Jalli

## I Built an App in 6 Hours that Makes \$1,500/Mo

Copy my strategy!

· 3 min read · Jan 23, 2024



Mokkup.ai

## The Future of Data Visualization: 2024 and Beyond

In this blog post, we will discuss the latest trends in data visualization and explore how...

7 min read · Dec 15, 2023

12.1K

145



•••

799

4



•••



Sameerah\_writes

## Retail Store Price Optimization- An Analysis of Product Information.

A data project on optimizing product category prices for a retail store using Power...

14 min read · Jan 25, 2024

102

1



•••

Nathan Rosidi

## Data Science in 2024—What Has Changed

What has changed in the data science landscape, and what are the challenges of th...

4 min read · Jan 29, 2024

687

10



•••

See more recommendations