

# Los Angeles Neighborhood Score

Matheus Schmitz  
Jameson Thai  
Yeonsoo (Emma) Park





## **Project Vision**

Creation of a comprehensive neighborhood quality of life comparison score for Los Angeles, along with an accompanying visualization tool, allowing those interested in learning more about the city or seeking to move into one of its neighborhoods to easily access relevant information.



# Neighborhood: Global Indexer

Find a real data management problem



Solve the problem we created for ourselves



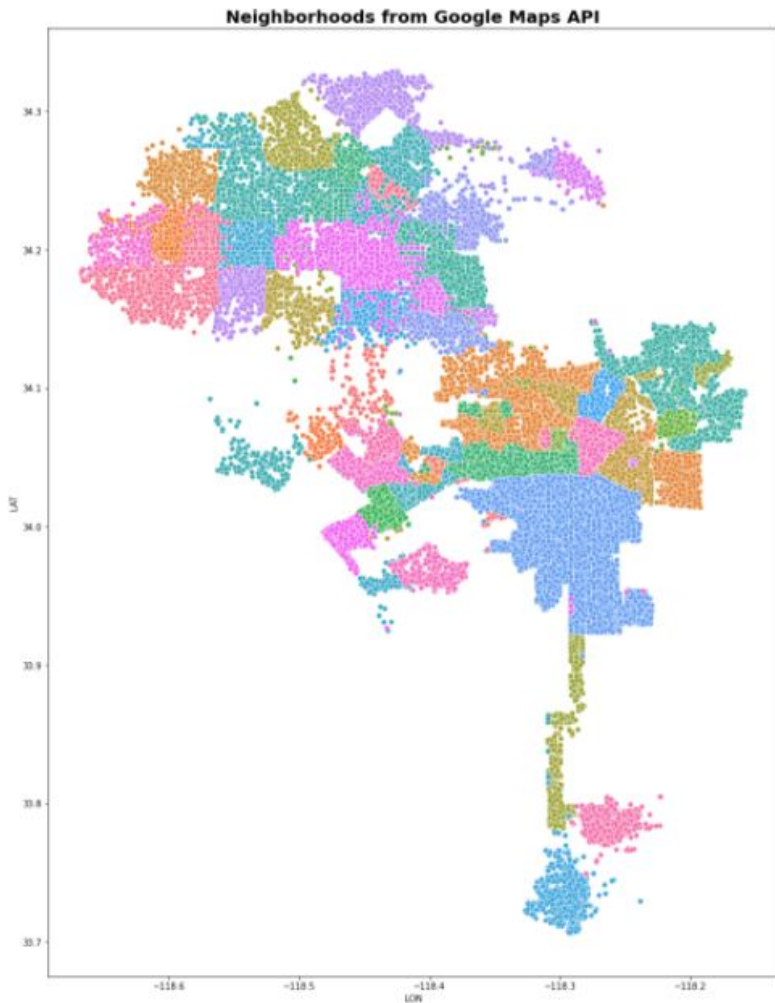
Index all data by neighborhood



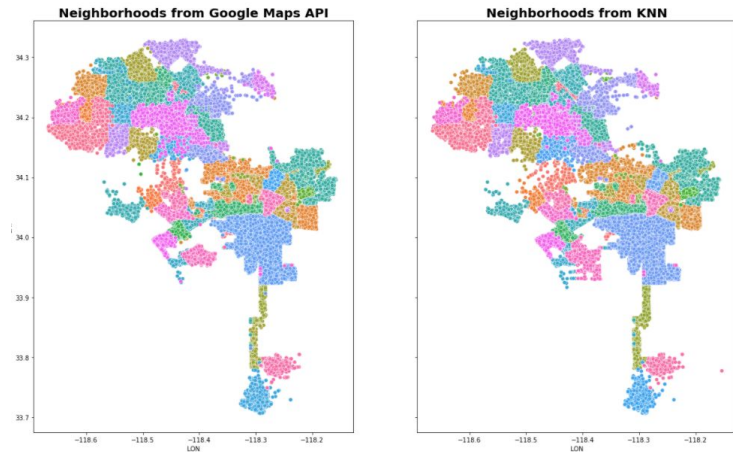
# Data Extraction with Google Maps

**Problem:** Not all datasets contained the neighborhoods associated with the records

**Solution:** Google Maps API



# Data Extraction with Machine Learning



**Problem:** Due to limitations on the Google Maps API, not all records on the crime dataset could be processed

**Solution:** Training a KNN algorithm to predict the neighborhoods for the rest of the records



# Data Modeling to Create Features

**Problem:** For the crime dataset, there was no feature which could be directly extracted from the dataset

**Solution:** Apply data modeling techniques to convert the “crime code” of each record into a feature which would be used to rank neighborhoods



# Databases



## Problem:

- 1) Multiple datasets to be stored
- 2) Team on different countries and time zones, needed all databases available online all times for collaboration

## Solution:

- 1) Individually storing the datasets on MySQL as they are finished
- 2) Hosting both databases online, ensuring availability for all team members



# Spark

## Problem:

- 1) Processing large crime dataset
- 2) Merging all datasets

## Solution:

- 1) Spark SQL for high level functions applying MapReduce
- 2) Individually load each dataset to Spark, then MapReduce all into a single tabular dataset indexed by neighborhood





# Data Retrieval

**Problem:** There are multiple relevant features and multiple neighborhoods.  
Need to have query flexibility to extract partial combinations

**Solution:** Developed a firebase query function which can customize itself based on received inputs, allowing flexible real-time queries to be executed over a variety of features and neighborhoods



# Data Visualization

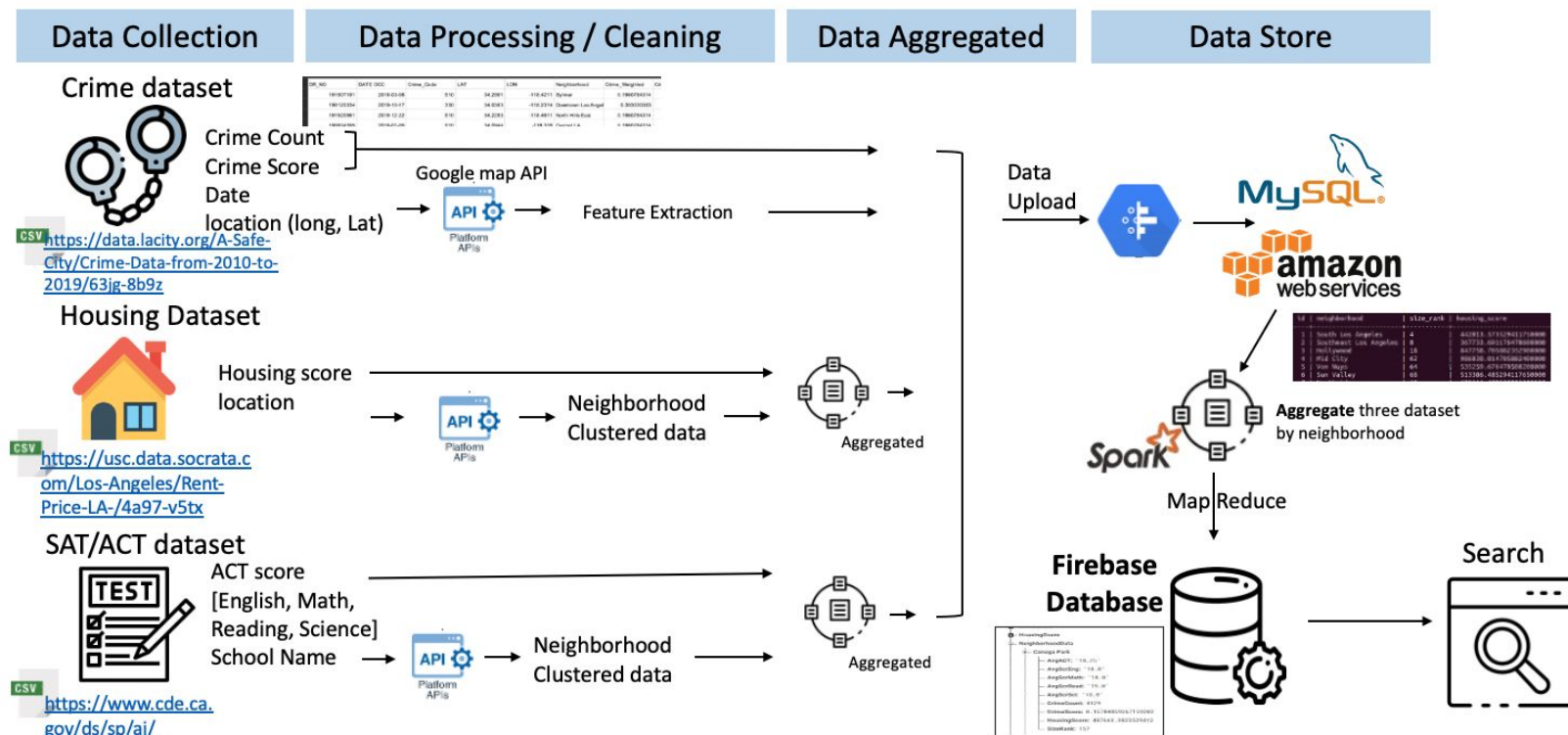
**Problem:** Hard to understand and compare the distribution of neighborhood score across neighborhood in Los Angeles

**Solution:** Utilize a GeoJSON (python library) to visualize selected neighborhood on the map in real time



Image : yellow color shows the neighborhood user selected

# Architecture Diagram





# DEMO

1. Checkboxes & Query
2. Features
3. GeoJson
4. Missing Values

```
df_output.idxmax()
```

CrimeCount	South Los Angeles
CrimeScore	The Flats
HousingScore	Bel Air
SizeRank	Playa Vista
AVG_SCR_ENG	Mid City
AVG_SCR_MATH	Mid City
AVG_SCR_READ	Mid City
AVG_SCR_SCI	Mid City
AvgACT	Mid City

dtype: object

```
df_output.idxmin()
```

CrimeCount	Fox Hills
CrimeScore	Sunkist Park
HousingScore	Watts
SizeRank	South Los Angeles
AVG_SCR_ENG	Westlake North
AVG_SCR_MATH	North Hollywood
AVG_SCR_READ	Westlake North
AVG_SCR_SCI	Westlake North
AvgACT	Westlake North

dtype: object



# Future Work

- Ensuring all neighborhood geolocations map to the map
- Color coding the map by score in a given feature
- Incrementing more quality of life aspects
- Neighborhood recommender system



**Thank You**