Coursera Capstone Project
The battle of Neighborhood - Phoenix, AZ

Doo Hwan Kim

Business Problem

This project is to find out the safest zip code in Phoenix area based on the total crime dataset in Phoenix, explore the neighborhoods of each zip code to find the 10 most common venues in each neighborhood and finally cluster the neighborhoods using k-mean clustering. I live in Phoenix and I'm curious how safe my place is and this will be helpful for someone who are looking to relocate to Phoenix, AZ

Data

Following data sources will be needed to extract/generate the required information:

- Phoenix Crime Dataset from 2017 to 2019: A dataset consisting of the crime statistics of each zip code in Phoenix.
  Dataset URL: https://www.phoenixopendata.com/dataset/crime-data
- Zip code with co-ordinates.
  Dataset URL:
  https://public.opendatasoft.com/explore/dataset/us-zip-code-latitude-and-longitude/

Data Explanation

Phoenix Crime Data

About this file

- INC NUMBER : Incident number
- OCCURRED ON : Start time
- OCCURRED TO : End time
- UCR CRIME CATEGORY : Crime type
- 100 BLOCK ADDR : Address
- ZIP : Zip Code
- PREMISE TYPE : Premise type

Data Explanation - cont’d

Zip code co-ordinates

About this file

- Zip
- City
- State
- Latitude
- Longitude
- Timezone
- Daylight savings time flag
- geopoint

Methodology

- Exploratory Data Analysis: Visualise the crime rates in the City of Phoenix based on zip codes to idenity the safest zip code and pick up one zip code to figure out what each type of crime happened in that specific zip code and eventually find the 10 most common venues in each zip code.

- Modelling: To help people find similar neighborhoods in the safest zip code we will be clustering similar neighborhoods using K - means clustering which is a form of unsupervised machine learning algorithm that clusters data based on predefined cluster size. We will use a cluster size of 10 for this project that will cluster the 40 neighborhoods into 10 clusters. The reason to conduct a K- means clustering is to cluster zip codes with similar venues together so that people can find the area of their interests based on the venues/amenities around each neighborhood.

Data Exploration

Phoenix Crime Data with Zip code

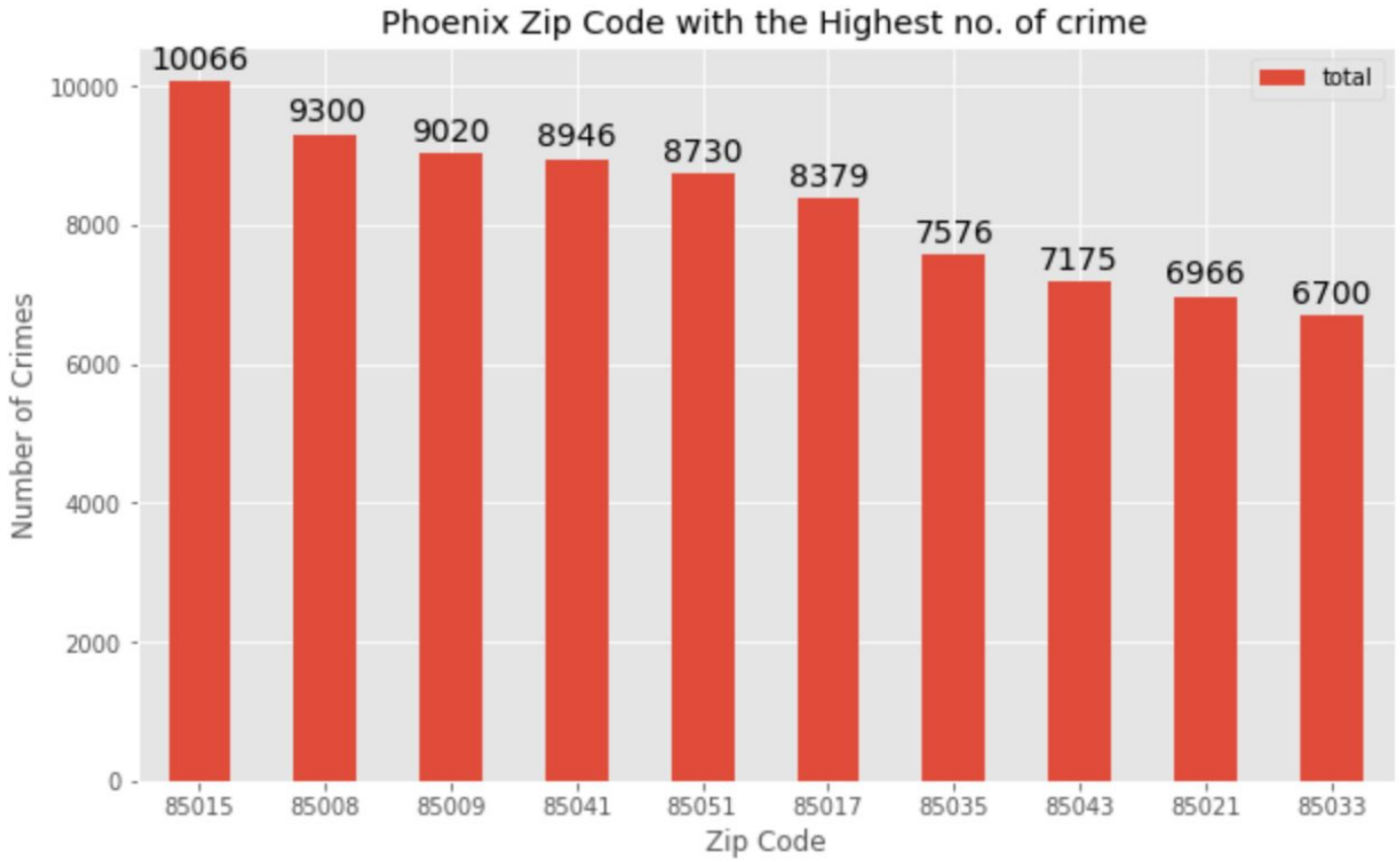
crime_type	zip_code	AGGRAVATED ASSAULT	ARSON	BURGLARY	DRUG OFFENSE	LARCENY-THEFT	MOTOR VEHICLE THEFT	MURDER AND NON-NEGLIGENT MANSLAUGHTER	RAPE	ROBBERY
0	85003	264	26	302	355	1227	205	3	80	127
1	85004	228	11	248	198	1517	163	6	50	141
2	85006	432	51	712	439	1868	395	12	137	209
3	85007	452	38	388	573	1378	288	11	76	204
4	85008	831	84	1128	943	4773	924	25	195	397

Data Exploration

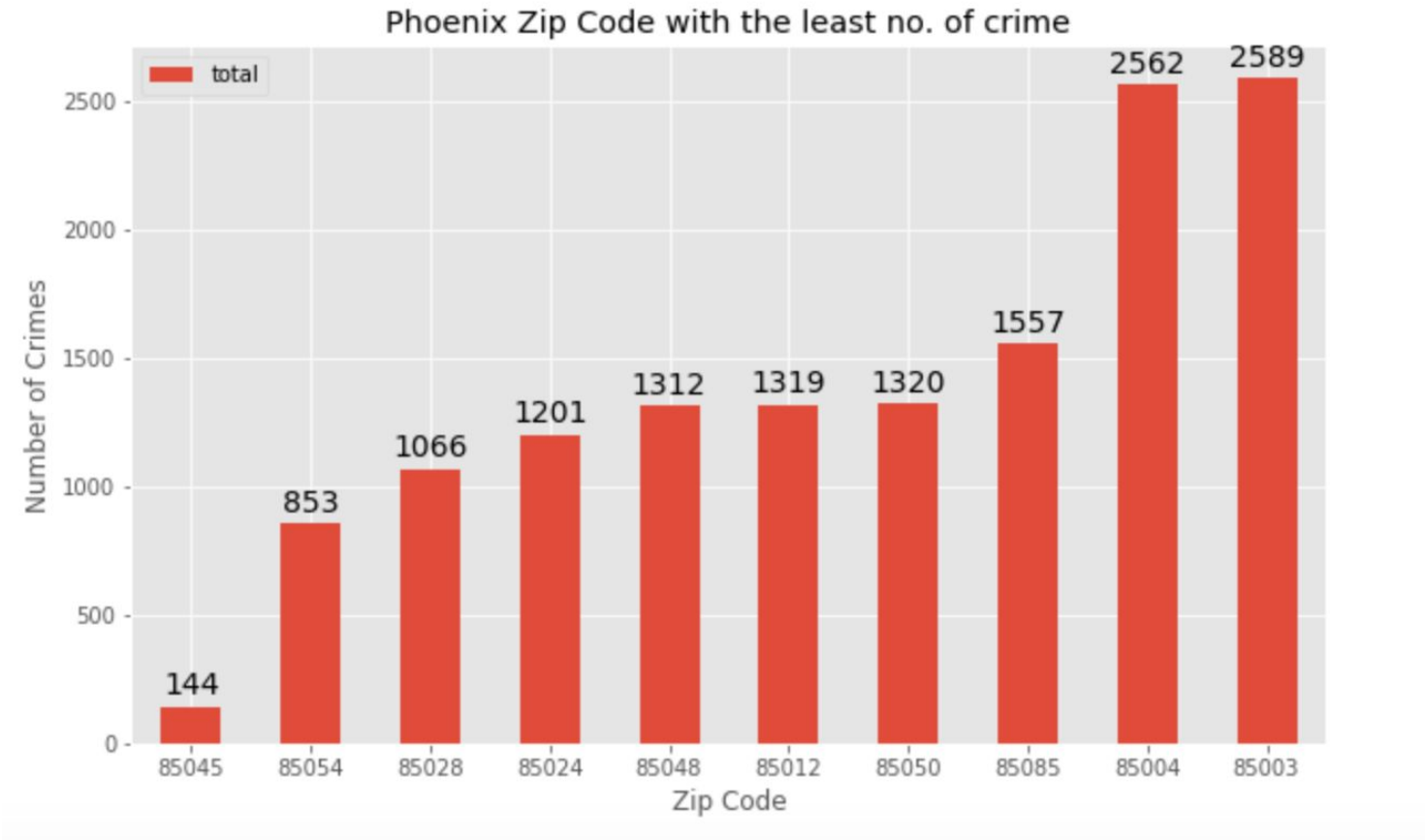
Zip code with co-ordinates

	zip_code	city	State	lat	lng
0	85001	Phoenix	AZ	33.703967	-112.351835
1	85031	Phoenix	AZ	33.493496	-112.171080
2	85041	Phoenix	AZ	33.386995	-112.100760
3	85013	Phoenix	AZ	33.507110	-112.084830
4	85017	Phoenix	AZ	33.514092	-112.122850
5	85023	Phoenix	AZ	33.638271	-112.093410
6	85042	Phoenix	AZ	33.381234	-112.027663
7	85032	Phoenix	AZ	33.624140	-112.004160
8	85051	Phoenix	AZ	33.559783	-112.133610
9	85035	Phoenix	AZ	33.472492	-112.187500
10	85020	Phoenix	AZ	33.563663	-112.055190
11	85004	Phoenix	AZ	33.451093	-112.070570
12	85045	Phoenix	AZ	33.299740	-112.098120
13	85029	Phoenix	AZ	33.598841	-112.120230
14	85034	Phoenix	AZ	33.437772	-112.028100
15	85006	Phoenix	AZ	33.466392	-112.048750
16	85015	Phoenix	AZ	33.506890	-112.102620

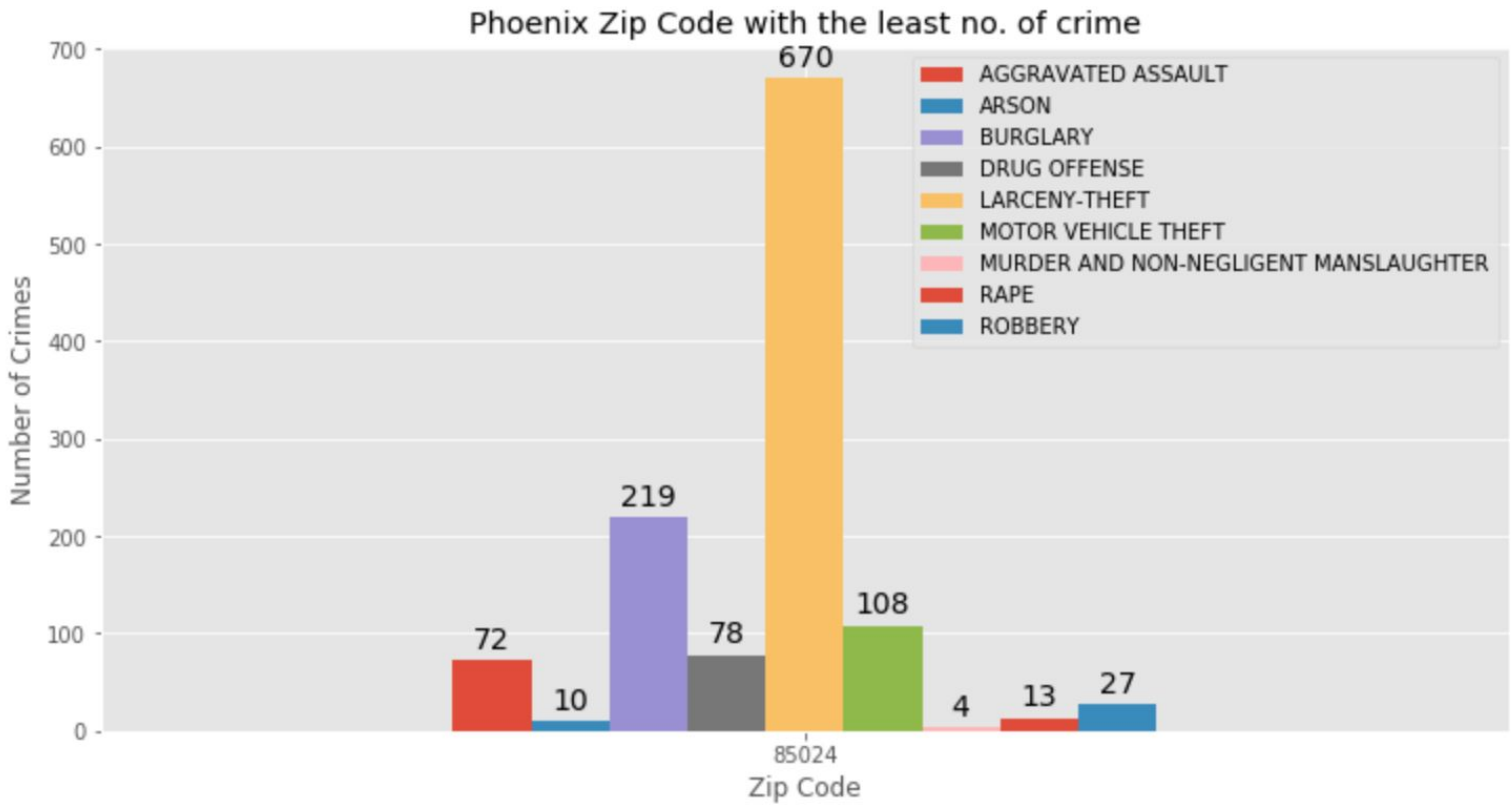
Data Exploration & Analysis



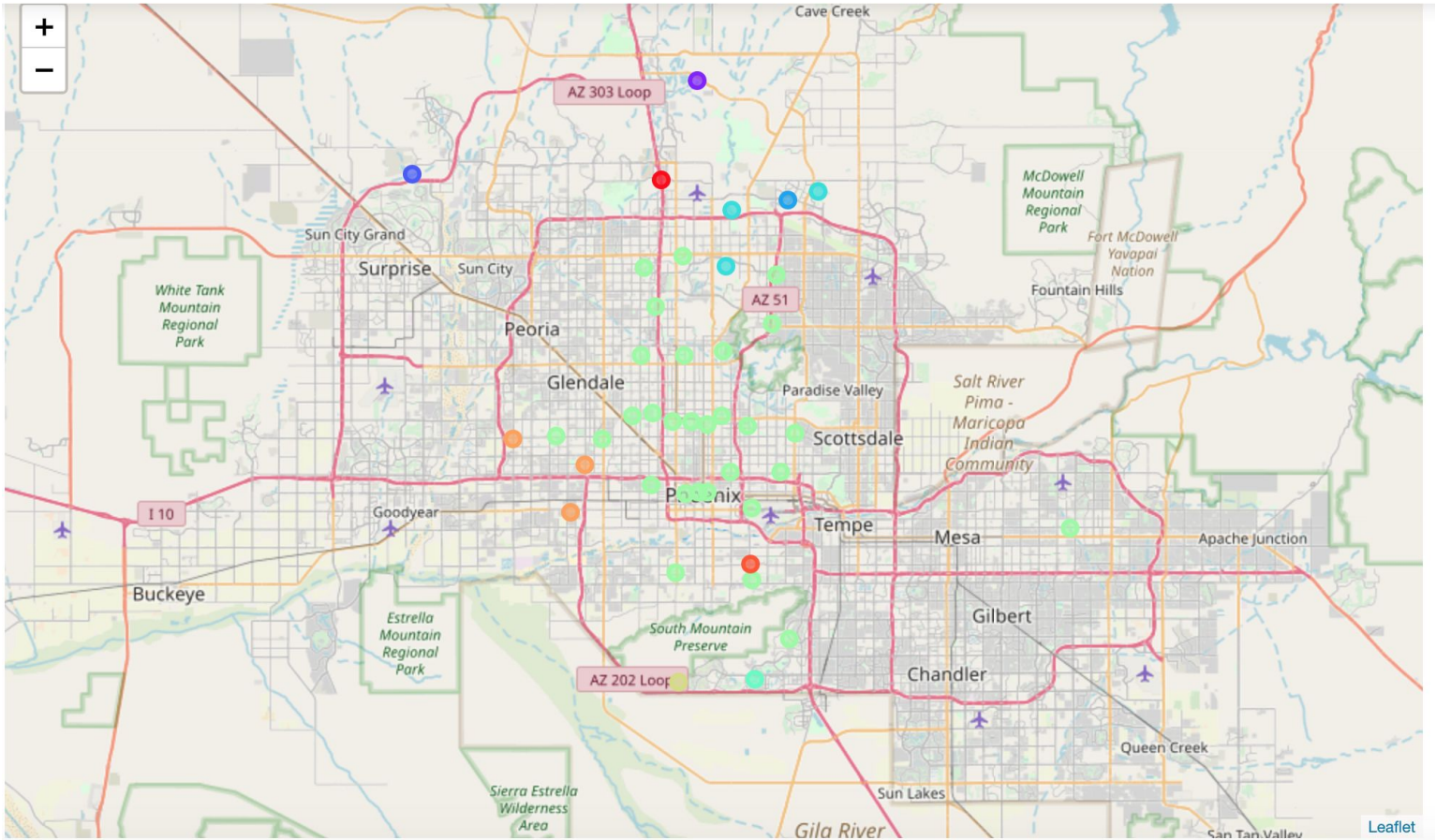
Data Exploration & Analysis



Data Exploration & Analysis



Data Exploration - Modeling(Clustering)



Results and Discussion

The aim of this project is to help people who want to relocate to the safest zip code in Phoenix, AZ using the data of crime, we can easily find the safest zip code in Phoenix based on that we can use the open dataset from Foursquare API to get information of all the venues of selected zip code, in the end, using machine learning algorithm to finish the cluster task, people can choose the place to live based on what kind of venues they like the most.

Conclusion

One limitation -> I tried to find out neighborhood information in each zip code within Phoenix area but, I couldn't find it. Clustering based on zip codes are less related in terms of venue finding because it's too broad for finding out the correlation. But, if someone already lives in Phoenix might be wondering how safe current their location is then this project will help them out to find out some clue about that for sure.