DATA SCIENCE CAPSTONE

RESEARCHING CRIME RATES AND COFFEE SHOPS IN ORANGE COUNTY, CALIFORNIA, UNITED STATES OF AMERICA

KELLY ANDERSON – APRIL 2019

DESCRIPTION OF THE PROJECT:

Research to see if there is a correlation between the number of Starbucks and sushi restaurants and crime rates for cities in Orange County, California.

My questions that I would like to research, and answer are:

- Does the number of Starbucks and/or Sushi restaurants (per capita), have any correlation to crime rates?
- Of the surrounding cities in Orange County, which are similar to the city that I live in as far as sushi restaurants, Starbucks, and crime rates per capita

DATA COLLECTION AND METHODOLOGY

- Crime Rate and Population Data:
- https://en.wikipedia.org/wiki/Orange_County,_California

Cities by population and crime rates [edit]

Cities by population and crime rates [hide]

City	Population[91]	Violent crimes[91]	Violent crime rate per 1,000 persons	Property crimes[91]	Property crime rate per 1,000 persons
Aliso Viejo	48,999	43	0.88	415	8.47
Anaheim	344,526	1,279	3.71	10,070	29.23
Brea	40,253	74	1.84	1,292	32.10
Buena Park	82,505	206	2.50	2,066	25.04
Costa Mesa	112,635	254	2.26	4,079	36.21
Cypress	48,976	56	1.14	1,018	20.79
Dana Point	34,172	65	1.90	604	17.68
Fountain Valley	56,674	106	1.87	1,469	25.92

- Reference: United States Department of Justice, Federal Bureau of Investigation. Crime in the United States, 2012, Table 8 (California). Retrieved November 14, 2013.

DATA COLLECTION AND METHODOLOGY

Restaurant and Venue Data:

- Foursquare API to loop through cities to aggregate and consolidate by city

```
[269]: dfc.head()
```

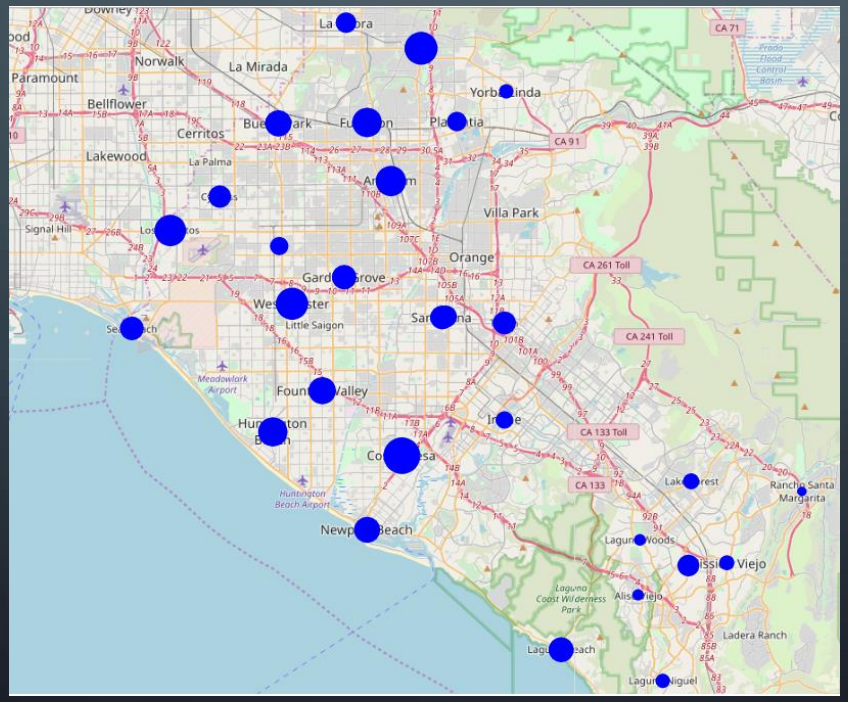
[269]:

	id	location.address	location.city	name
0	4b4e6d97f964a5206bed26e3	27020 Alicia Pkwy	Laguna Niguel	Starbucks
1	4ac8ea1cf964a52045bd20e3	23411 Aliso Viejo Pkwy	Aliso Viejo	Starbucks
2	4bca132a511f9521d5f5aec7	27072 La Paz Rd	Aliso Viejo	Starbucks
3	4b3bc69df964a520e97a25e3	28391 Marguerite Pkwy	Mission Viejo	Starbucks
4	4baa360af964a5209e533ae3	25630 Alicia Pkwy	Laguna Hills	Starbucks

```
[150]: dfc.to_csv('Starbucks_out.csv')
```

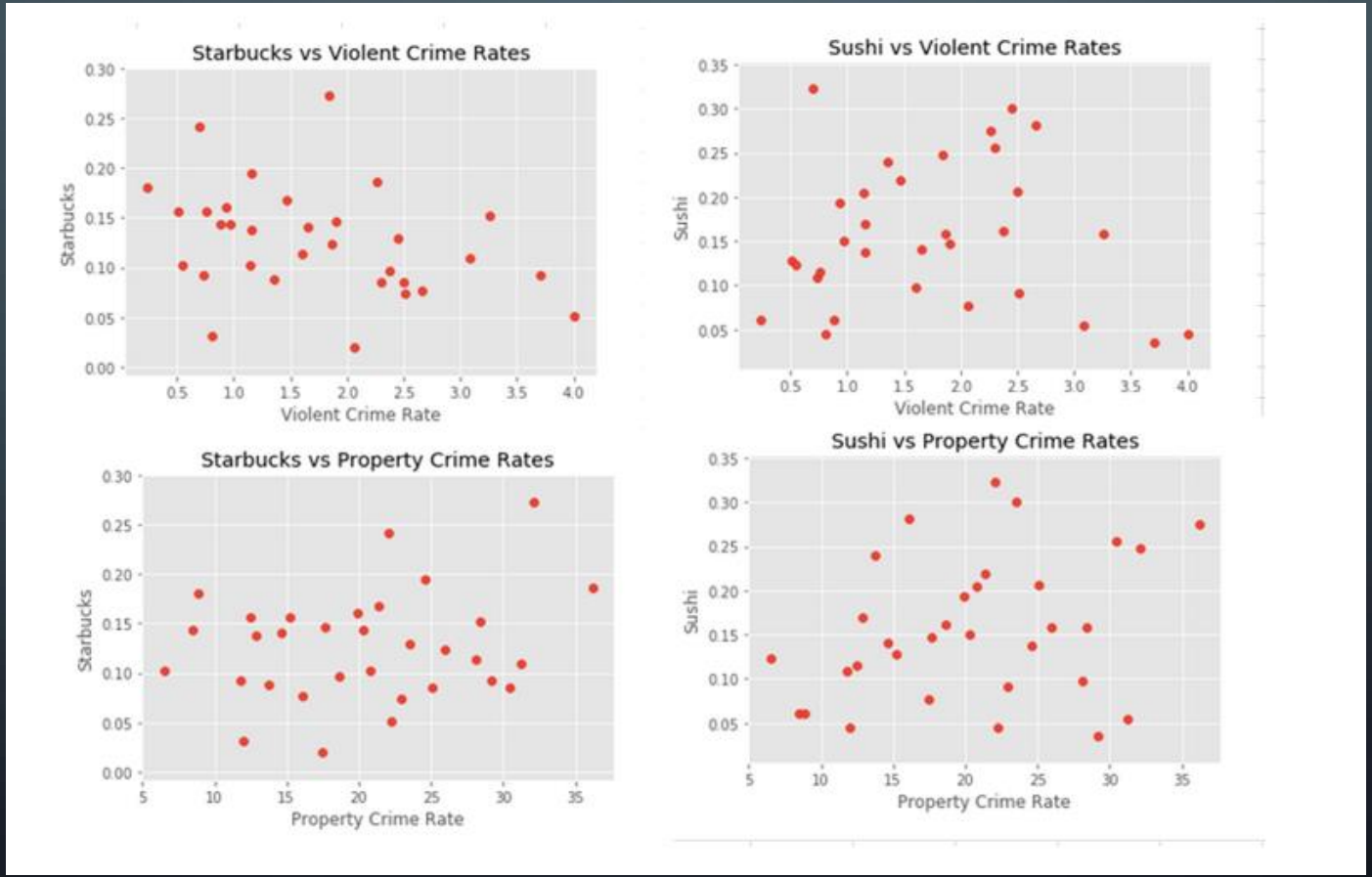
EXPLORATORY DATA ANALYSIS

- Reviewed datasets on a per capita basis with mapping and scatterplots:
- Here is an example of the crime rates per capita for the cities in Orange County:



EXPLORATORY DATA ANALYSIS

- Scatterplot graphs of dependent and independent variables:



METHODOLOGY

- Simple Linear Regression Analysis
- Multiple Linear Regression Analysis
- K-means Clustering

RESULTS

- No clear correlation between dependent and independent variables
- Sushi per capita vs Property Crime Rates

  Mean absolute error: 6.46

  Residual sum of squares (MSE): 68.77

  R2-score: -8.45

- Starbucks per capita vs Property Crime Rates

  Mean absolute error: 5.48

  Residual sum of squares (MSE): 46.52

  R2-score: -42.16

MULTIPLE LINEAR REGRESSION

- Using both values compared against the Property Crime rates we come up with the following values for the model:
- Residual sum of squares: 70.04
- Variance score: -0.97

CLUSTERING USING K-MEANS

- Four Clusters of cities for common features (Venues, Crime, and Population)

	Population	Violent_crime_rate	Property_crime_rate	Starbucks_PC	Sushi_PC
Labels					
0	62238.0	1.391667	26.043333	0.204258	0.232801
1	212854.5	3.031667	27.035000	0.098625	0.080372
2	79833.7	0.826000	12.322000	0.128471	0.110207
3	48913.7	2.062000	20.938000	0.095249	0.203220

CLUSTER PROFILES

- Group 0- Medium Population, Medium crime, and lots of Starbucks and Sushi
- Group 1 - High population, highest crime, Lowest Starbucks and Sushi
- Group 2 - Medium population, lowest crime, medium Starbucks and Sushi
- Group 3 - Lowest population, Medium crime, Low Starbucks and high Sushi

SIMILAR CITIES TO LAGUNA BEACH, CA (HOME)

	City	Population	Violent_crime_rate	Property_crime_rate	Starbucks_PC	Sushi_PC	Labels
3	Buena Park	82505	2.50	25.04	0.084843	0.206048	3
5	Cypress	48976	1.14	20.79	0.102091	0.204182	3
6	Dana Point	34172	1.90	17.68	0.146319	0.146319	3
7	Fountain Valley	56674	1.87	25.92	0.123513	0.158803	3
12	La Habra	61731	2.38	18.63	0.097196	0.161993	3
13	Laguna Beach	23283	2.45	23.54	0.128849	0.300649	3
17	Lake Forest	79166	1.35	13.74	0.088422	0.240002	3
18	Los Alamitos	11728	2.30	30.44	0.085266	0.255798	3
22	Placentia	51778	2.07	17.50	0.019313	0.077253	3
28	Stanton	39124	2.66	16.10	0.076679	0.281157	3

OBSERVATIONS AND RECOMMENDATIONS

- No Clear Correlation
- More detailed data required for:
  - Larger sample size
  - Less variance for larger cities

CONCLUSION

- Even though some cities have a lot of Starbucks and Sushi restaurants, there is no indication that they have less crime.
- The clusters of cities are helpful for finding other places that fans of these venues, and low crime would like.
- Thanks! Kelly