2020

Cincinnati Crime Study

CORRELATIONS WITH FOURSQUARE DATA

ALEX M

I. Introduction and Business Problem

Public officials often wonder about and study the factors which may cause crimes to occur in their city. The United States is a prime example of a "melting pot" of many different cultures, religions, values, and ethnicities. As a result, there can be deep pockets of culture even within an individual city.

This is certainly the case with Cincinnati, Ohio. Ohio is a population center of the Midwest United States, having the 7th-ranked total population among all states while only being the 35th-largest by square mileage. While crime is certainly not fairly explained by race, ethnicity, or religion, it can be fairly predicted by behavioral factors. In this short study I examine one such factor: **alcohol use.**

The hypothesis is as follows: the larger the concentration of opportunities to drink alcohol within a certain area, the higher propensity for crimes. If this hypothesis is true, new residents of a city should steer clear of bars, gas stations, and liquor stores when choosing a place to live, because those are the establishments that offer opportunities for people to purchase and drink alcohol. Therefore the key audience of this study is **new residents of Cincinnati, or those looking to move between locations within Cincinnati.**

For this project I will be pulling data using the Foursquare API, and merging it with Cincinnati crime data. The final model development dataset will be split into training/testing sets and machine learning algorithms will be used to determine **if crime can be predicted by the presence of alcohol-centered venues.**

II. Data

There are two key sources of data for this project, as follows:

First, venue data will be pulled using the Foursquare API related to alcohol-centered establishments. This includes **bars, gas stations, and liquor stores.** The following queries in Python will be used to pull the data (Figure 1 below):

Figure 1: Foursquare Queries

```
[41]: #The first Foursquare query will be Bars within 30 miles of Cinci;
search_query1 = 'Bar'
radius = 50000
print(search_query1 + ' .... OK!')
```

```
Bar .... OK!
```

```
[42]: #The second Foursquare query will be Gas Stations within 30 miles of Cinci;
search_query2 = 'Gas Station'
radius = 50000
print(search_query2 + ' .... OK!')
```

```
Gas Station .... OK!
```

```
[43]: #The third Foursquare query will be Liquor Stores within 30 miles;
search_query3 = 'Liquor'
radius = 50000
print(search_query3 + ' .... OK!')
```

```
Liquor .... OK!
```

```
[44]: #Trying to find more liquor stores via query Spirits;
search_query4 = 'Spirits'
radius = 50000
print(search_query4 + ' .... OK!')
```

```
Spirits .... OK!
```

As can be observed in Figure 1 above, a separate query was utilized for “Bar”, “Gas Station”, “Liquor”, and “Spirits”. The purpose of two separate queries for liquor and spirits is because the name of the liquor store may likely include either key word. Note that the queries return the name and type of establishment as well as the zip code and latitude/longitude of the place. The number of each type of establishment will be aggregated by zip code.

Second, crime data will be pulled from the City of Cincinnati Online Crimes Database. In 2019, approximately 418,000 crimes were catalogued in the database. The data will be cleaned and again aggregated by zip code and neighborhood name, which are fields included in the dataset. Figure 2 below shows an example of the crime data.

Figure 2: Cincinnati Crime Data

Table Preview

View Data

)PE...	HATE...	DAY...	RPT_...	CPD_...	SNA_NEIGHBOR...	WEA...	HOU...	HOU...	ADD...
	N--NO BI...	SUNDAY	20	C. B. D. / ...	PENDLETON	40 - PERS...	1745	180	
	N--NO BI...	FRIDAY	419	WINTON ...	WINTON HILLS	U - UNKN...	440	450	
	N--NO BI...	FRIDAY	268	SAYLER P...	SAYLER PARK	99 - NONE	120	120	
	N--NO BI...	THURSDAY	385	PADDOC...	NORTH AVONDALE - P...	40 - PERS...	00	00	
	N--NO BI...	SATURDAY	481	ROSELAWN	ROSELAWN	99 - NONE	55	105	
	N--NO BI...	TUESDAY	467	BONDHILL	BOND HILL	99 - NONE	170	170	
	N--NO BI...	FRIDAY	52	WALNUT ...	WALNUT HILLS	40--PERS...	2140	2215	
	N--NO BI...	SATURDAY	385	PADDOC...	NORTH AVONDALE - P...	99 - NONE	00	60	
	N--NO BI...	TUESDAY	376	NORTH A...	AVONDALE	40 - PERS...	10	90	
	N--NO BI...	TUESDAY	203	WEST END	WEST END	U - UNKN...	170	030	
	N--NO BI...	THURSDAY	302	NORTH F...	NORTH FAIRMOUNT	20 - KNIF...	237	05	
	N--NO BI...	WEDNES...	271	WESTWO...	WESTWOOD	99 - NONE	015	525	
	N--NO BI...	SUNDAY	439	COLLEGE ...	COLLEGE HILL	40 - PERS...	210	220	
	N--NO BI...	MONDAY	303	SOUTH F...	EAST WESTWOOD	99 - NONE	40	430	