Determining the Safest Districts to Reside In and Start Businesses in: Greater Manchester

Hadi Najafpour

December 28th 2019

Introduction

Background

Manchester is a lively place in the United Kingdom, full of music, sports and leisure activities which students and professionals alike thrive off of. Since the early 90s, the districts in the Greater Manchester county (1,276 km^2) have reaped the benefits of having Manchester become a capital of entertainment in the United Kingdom. As a result of this, business owners look at the potential to build their business there, and would like the determining factors in deciding which district it would be preferable to set up in.

Problem

Whilst business owners may find the county to be a good opportunity to set up shop in, crime rates and the general safety of civilians and their business do pose risks to their plans. In this report, we will look to find any correlation between the number of venues present in the metropolitan county, and the number of crimes committed throughout the month of June in 2019. This will be done to determine the safest districts for business owners or law-abiding residents to spend time in and build their businesses at.

Data

Sources

We will be using the following data to help draw us to our conclusions:

1. Foursquare Developers Access to venue data: https://foursquare.com/
2. Open Crime Data from https://data.police.uk/data/open-data/ using only data for the month of June 2019

Roughly 32,000 crimes are committed per month in Greater Manchester, giving around 384,000 crimes committed throughout the year. Using Foursquare data in tandem with the dataset acquired from https://data.police.uk/, we will seek to find any sort of linear relationship between Crimes Committed and Venues in the Greater Manchester area.

Cleaning & Appending

The Open Crime Dataset acquired will first be rid of any unwanted data, such as the crime identification number, who it was reported by, etc. The only pieces of data we will be taking into for this dataset are the districts in which the crime was committed, and we will furthermore determine the number of crimes committed per district in Greater Manchester for the month of June. The data included data regarding districts not in Greater Manchester, such as Cheshire East, Rossendale, Warrington, and West Lancashire, so the rows relating to crimes in these districts will be removed and disregarded as well.

Foursquare data will also be used in order to retrieve the number of venues - for leisure, entertainment, sports, food, etc. - in each district, to subsequently compare to the number of crimes there exists per district in a new dataset, and we shall append a column to this dataset consisting of the number of crimes per district in Greater Manchester.

Methodology

Retrieving Data from Data Sources

Firstly, we begin by importing data from Open Crime Data for Greater Manchester in the month of June, and storing it in a dataframe using the pandas library in order to use it for data analysis. Following the editing and cleaning of the dataframe, such as removing districts not in the Greater Manchester county, and disregarding all other data from the data source which will not be used, Figure 1 below shows the dataframe of the number of crimes committed for each district in Greater Manchester.

District	Number of Crimes Committed
Bolton	3610
Bury	1822
Manchester	8828
Oldham	2842
Rochdale	2510
Salford	2771
Stockport	2424
Tameside	2548
Trafford	1764
Wigan	2924

Figure 1: Number of Crimes Committed for each District in Greater Manchester for June 2019

GeoPy to Display Districts on Map

After importing the necessary libraries from GeoPy.GeoCoders called Nominatim, we are able to instantiate the names of the districts in Greater Manchester ourselves, and then use the library to find the coordinates of each district in Greater Manchester and create a dataframe for this data.

Following the retrieval of the coordinates, we use the imported library Folium to render a map, and add circular markers in the coordinates from the dataframe created previously for each district. Figure 2 shows the map formed.

Figure 2: Map showing markers on coordinates of each district in Greater Manchester

Retrieving Venues in each District using Foursquare API

A function is created in order to retrieve venues in each district within a 500km radius, and the function returns Area, Area Latitude, Area Longitude, Venue, Venue Latitude, Venue Longitude, and Venue Category. The only relevant categories of data we need from this dataframe are Area (District), and a count of the number of venues per district, which we will find using the groupby function on the Area Column, and this will be appended as a column to the dataframe showing crimes per districts, as in Figure 1. The dataframe of Districts and Venues is shown below.

Area	Venues
Bolton	41
Bury	48
Manchester	100
Oldham	26
Rochdale	19
Salford	18
Stockport	32
Tameside	4
Trafford	1
Wigan	48

Figure 3: Dataframe showing Venues per District in Greater Manchester

Results & Discussion

We start by visualizing the data gathered for crimes in Greater Manchester per district, and the venues in each district, from Figures 1 and 3. The corresponding bar plots for each dataframe are shown below.

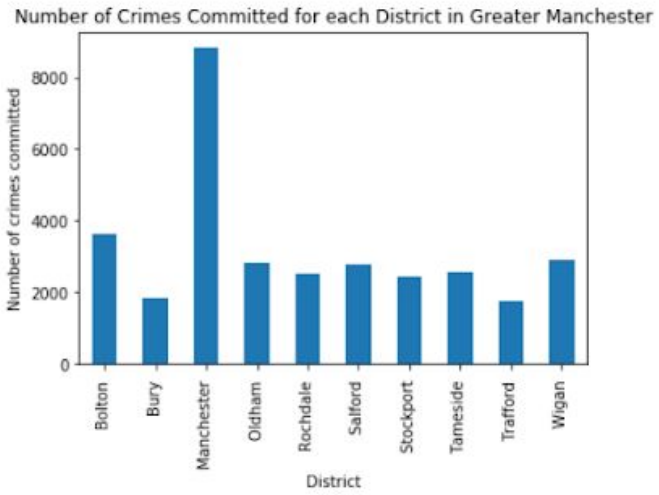

Figure 4: Bar plot showing the Number of Crimes Committed for each District in Greater Manchester during June 2019

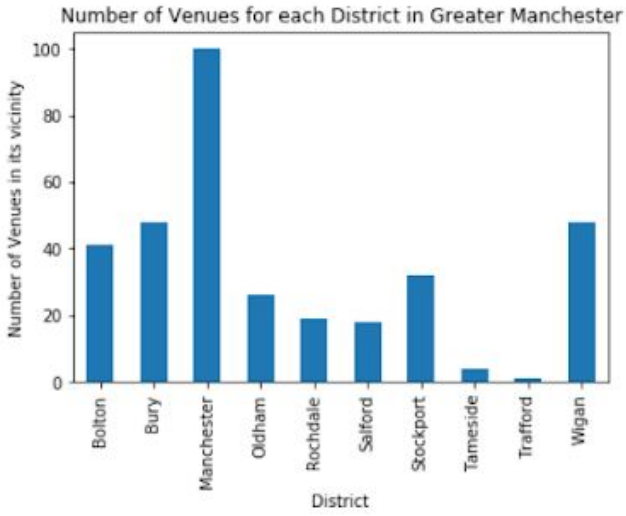

Figure 5: Bar plot showing the Number of Venues for each District in Greater Manchester

From simple visual inference, we are able to tell there may be some correlation between venues in each district and the number of crimes committed. But we will seek to validate these claims using Python libraries.

Firstly, we will combine both dataframes into one, and create a new column which calculates the number of crimes per venue in each district. Figure 6 shows the resulting dataframe.

Area	Venues	Crimes	Crimes per Venue
Bolton	41	3610	88.048780
Bury	48	1822	37.958333
Manchester	100	8828	88.280000
Oldham	26	2842	109.307692
Rochdale	19	2510	132.105263
Salford	18	2771	153.944444
Stockport	32	2424	75.750000
Tameside	4	2548	637.000000
Trafford	1	1764	1764.000000
Wigan	48	2924	60.916667

Figure 6: Dataframe showing Venues, Crimes Committed and Crimes per Venue for each District in Greater Manchester for June 2019

Based on the number of Crimes per Venue column in Figure 6, it is clear to see that Bury (with 38 crimes per venue) is at least 1.5x as safe as the next safest district, Wigan (61). Based on these results, the districts can be ordered in terms of safety.

1. Bury
2. Wigan
3. Stockport
4. Bolton
5. Manchester
6. Oldham
7. Rochdale
8. Salford
9. Tameside
10. Trafford

Next, we will use the .corr() function for the columns Venues and Crimes, in order to find the degree of correlation between the two datasets. The resulting dataframe is shown below.

	Venues	Crimes
Venues	1.00000	0.83636
Crimes	0.83636	1.00000

Figure 7: Dataframe showing the correlation between Venues and Crimes in Greater Manchester

As can be seen from Figure 7, there is a roughly 84% correlation between crimes committed in Greater Manchester and the number of venues in the county. This confirms our intuition that crime does relate to the number of venues in an area.

In order to visually confirm our findings, we will create a regression plot of our data of crime and venues. This plot is shown below in Figure 8.

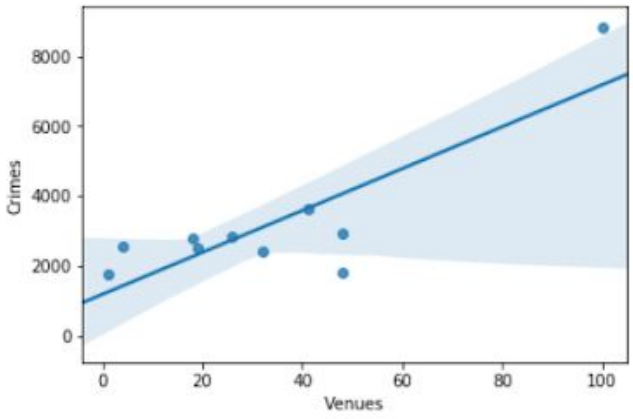

Figure 8: Regression Plot of Venues and Crimes in Greater Manchester

As shown in Figure 8, there is a positive linear relationship between the two features, Venues and Crimes. This confirms our initial assumptions and findings, that as venues in a given district increase, the number of crimes increase as well.

Conclusion

To conclude, we can suggest that entrepreneurs or business owners who seek to formally open up a business in Greater Manchester to set up shop in Bury, Greater Manchester. It has the fewest crimes per venue for a given district in the county, and as the county itself is fairly accessible from any of the districts, the commute from the city centre (Manchester) to Bury would not be a particularly gruesome one. Students, business owners, and civilians alike can enjoy the numerous venues available in the district whilst knowing they are safer than they would be in other parts of the county.