

Criminality Analysis in and around Eindhoven based on Nightclub density, Vincent Rikkink

Abstract

In this project I will analyze the criminality in the communities (like Veldhoven, Best, Son en Breugel, etc.) that surround Eindhoven and within Eindhoven as well. The data will be normalized with the total number of citizens and then compared with the nightclub density in these communities centers equally normalized as above. The idea is to see whether there is a correlation between the number of nightclubs and the number of crimes in a big city and its surrounding neighbourhoods.

Business case

The idea is that the community councilors have a tool to predict crime development in relation to the number of nightclubs in a certain neighbourhood. Based on these results they might adopt a different security strategy (e.g. police surveillance) per Neighbourhood.

Data Sets

I will need the following datasets:

- Total number of citizens per community
- Crime numbers per community
- Extract the number of nightclubs in the center of the community

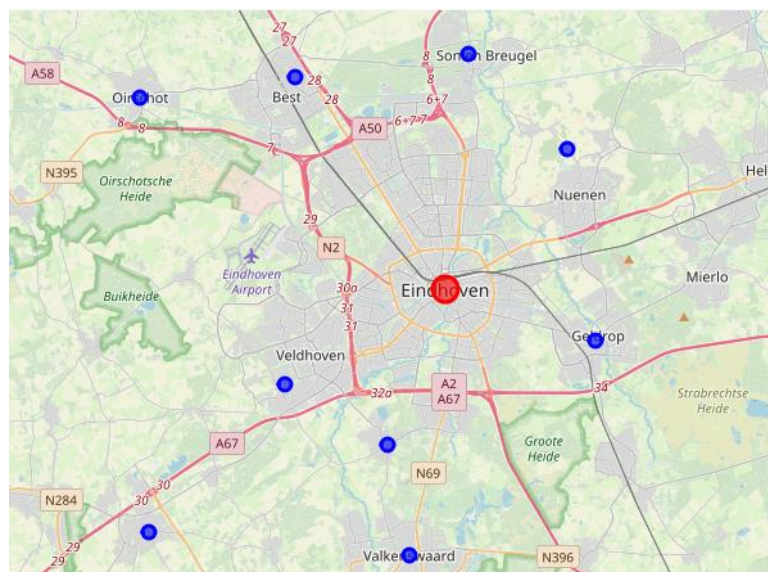
The first two data sets, I can obtain from the Dutch Office of Statistics (CBS.nl). The number of nightclubs can be obtained from searches on Foursquare.

Methodology

In a scatter plot I will present the results with on the x-axis the number of citizens and on the y-axis the crime number. Another scatter plot will present the night clubs against the number of citizens. The crime numbers per nightclub will be presented in a map. In this first week, only a map of Eindhoven and its Neighbourhoods will be presented.

Visualize Eindhoven and its Neighbourhoods

Eindhoven will be visualized as a big red dot and the neighbourhoods of Eindhoven as blue dot.



Data gathering and data analysis

The data has been extracted from the Central Office (Bureau) of Statistics (CBS) in the Netherlands. The individual files have been processed in a Cognitive Class Jupyter notebook due to the ease of data access and concatenated into a complete overview of registered crimes in Eindhoven and its surrounding neighbourhoods.

The data will be loaded, processes and the total number of registered crimes will be concatenated with the initial dataframe df.

Thereafter, the total population of Eindhoven and its Neighbourhoods (also extracted from the CBS website) will be loaded and added to the data frame df.

	Crime	Eindhoven	Eersel	Best	Waalre	Valkenswaard	Geldrop	Nuenen	SonenBreugel	Oirschot	Veldhoven
0	Misdrijven, totaal	18600	700	1100	475	1220	1565	705	665	675	1370
1	1 Vermogensmisdrijven	11535	380	645	330	610	915	435	450	430	775
2	1.1 Diefstal/verduistering en inbraak	10315	325	540	235	475	775	375	375	355	620
3	1.1.1 Diefstal en inbraak met geweld	175	5	10	5	5	5	15	10	5	10
4	1.1.2 Diefstal en inbraak zonder geweld	10140	320	530	225	475	765	360	370	350	610

Adding the total number of citizens and the location information of the neighbourhoods resulting finally in the following df.

	lat	lng	Crimes	Citizens
place				
Best	51.510017	5.398662	1100	29497
Geldrop	51.422199	5.559182	1565	39252
Eindhoven	51.439265	5.478633	18600	229126
Oirschot	51.503239	5.315549	675	18558
Eersel	51.358637	5.320131	700	18778
Nuenen	51.486030	5.544007	705	23019
Valkenswaard	51.350688	5.459450	1220	30654
Son en Breugel	51.517813	5.491435	665	16753
Veldhoven	51.407706	5.392731	1370	44925
Waalre	51.387640	5.447687	475	17075

Data Manipulations and Assumptions

The data frame will be extended with 3 additional columns:

1. Relative Crime This is the number of crimes committed per 1000 habitants. The assumption here is that in larger cities crimes will proportionally increase with the number of citizens.
2. Distance This is an indication of the radius of the center of a city/neighbourhood. I will assume that the center radius of a city with 10k citizens is 300m. The radius follows a square root increase with the increase of citizens. So a city with 40k citizens will have an assumed radius of 600m. And a city of 200k, a radius of $300 \cdot \sqrt{20}$.
3. The nightclubs in the centers of the individual neighbourhoods and Eindhoven extracted from FourSquare searches.

	lat	lng	Crimes	Citizens	RelCrime	Radius	Bars
place							
Best	51.510017	5.398662	1100	29497	37	515	5
Geldrop	51.422199	5.559182	1565	39252	39	594	4
Eindhoven	51.439265	5.478633	18600	229126	81	1436	50
Oirschot	51.503239	5.315549	675	18558	36	408	7
Eersel	51.358637	5.320131	700	18778	37	411	2
Nuenen	51.486030	5.544007	705	23019	30	455	0
Valkenswaard	51.350688	5.459450	1220	30654	39	525	24
Son en Breugel	51.517813	5.491435	665	16753	39	388	0
Veldhoven	51.407706	5.392731	1370	44925	30	635	8
Waalre	51.387640	5.447687	475	17075	27	392	1

Data Visualization

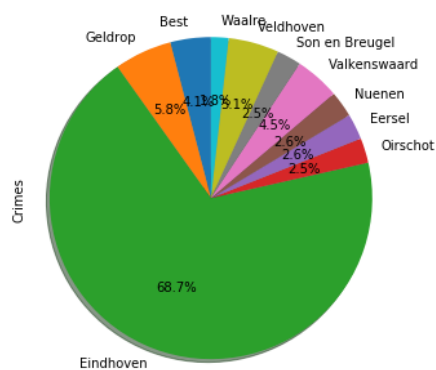
DISCLAIMER

The number of Neighbourhoods is far too low to make general conclusions. The idea is only to show the possibilities in case bigger data sets would be used. In above case it might be very well possible that a 2nd order regression would better fit than a linear one.

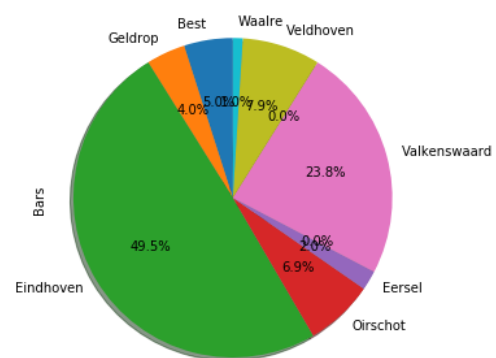
Nightclub and Crime distribution

The next plots show pie charts of the Nightclub and Crime distribution :

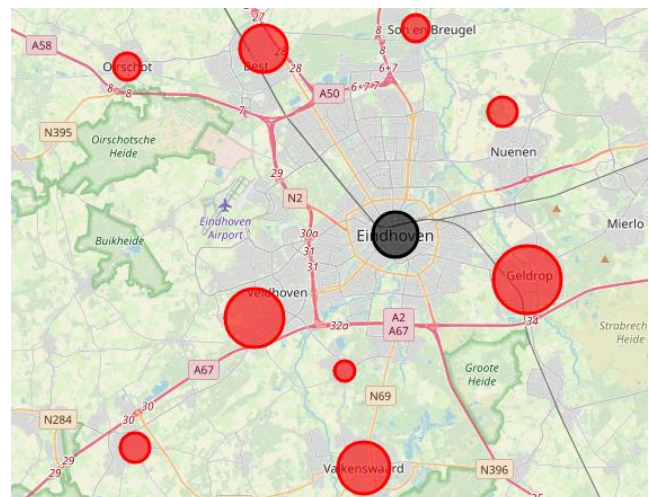
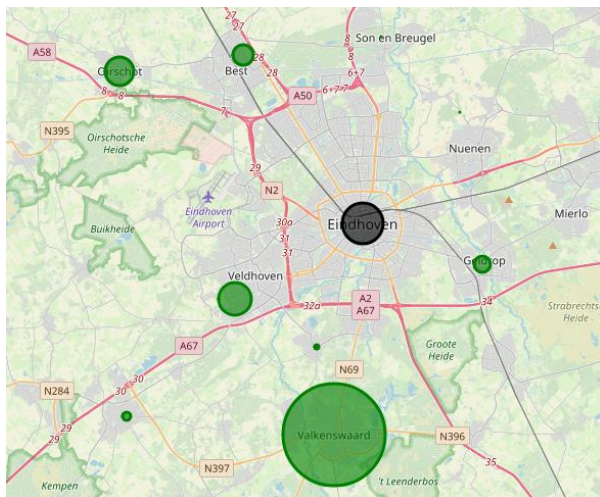
Crime distribution in Eindhoven and Neighbourhoods



Nightclub distribution in Eindhoven and Neighbourhoods



The next two figures give a spatial overview of the Nightclubs and Crimes in the Eindhoven region. The left figure show the Nightclub density in which Eindhoven has been shown in black (as an outlier). The right figure shows the Crime density and again Eindhoven in black.



Results

Different analysis shows a high correlation between Bars, Crimes and Citizens. Eindhoven is with respect to number of nightclubs, citizens and crimes an outlier with respect to its surrounding neighbours. Valkenswaard has a relatively high number of nightclubs compared to its neighbours.

Discussion

The high correlation does not necessarily mean that there also is a causality. For example the high correlation between nightclubs and crimes can also be due to the number of citizens. Comparing Eindhoven with its surrounding neighbourhoods is difficult due to its size. A follow up project could exclude Eindhoven, or split it up in different subareas which would have comparable sizes as Eindhoven. Also it would be interesting to extend the number of neighbourhoods with all the neighbourhoods around Eindhoven or even extend it to national scale. Another point to should be addressed is whether the information extracted from FourSquare has sufficient precision. First due to the fact that not all nightclubs are extracted and second because of the assumption that a center has been predicted based on population size of the corresponding city/neighbourhood.

Conclusion

A direct conclusion is difficult to draw due to the limited number of samples, the outlier Eindhoven and the precision of the extraction of data from FourSquare. The visualization on maps but equally in tables and plots gives a better insight in for example the distribution of nightclubs. For example, there are far more nightclubs on the west side of Eindhoven than on the east side. While crime is more evenly spread around Eindhoven with a slightly more south versus north distribution. But again, a more elaborated set of neighbourhoods would probably give a more complete picture. If there is any recommendation to make towards for example the police, than it would be based on the historical crime numbers rather than on a combination with other parameters.

Personal note

In this Capstone project I realized that these tools are so powerful that you can easily and very fast come to an analysis and a visualization of your results. I strongly believe that these kind of toolsets need to be common practice in a wide range of disciplines to support decision making in all domains. In the near future, I will absolutely going to apply these tools in my profession.