Classifying Crime Risk Using the K-Nearest Neighbor Classifier City of Leeds, England.

IBM Capstone Assignment

Justin Simpson

Introduction

- The city of Leeds and surrounding areas are located in the North of the United Kingdom within the county of West Yorkshire, approximately 170 miles north of London. Leeds has a diverse economic base and has a large multicultural population.

- Neighbourhoods within the Leeds area vary with respect to their composition in relation to both the levels and type of crime reported and the numbers and type of facilities, businesses and land use types that exiat in any given location

Business Problem

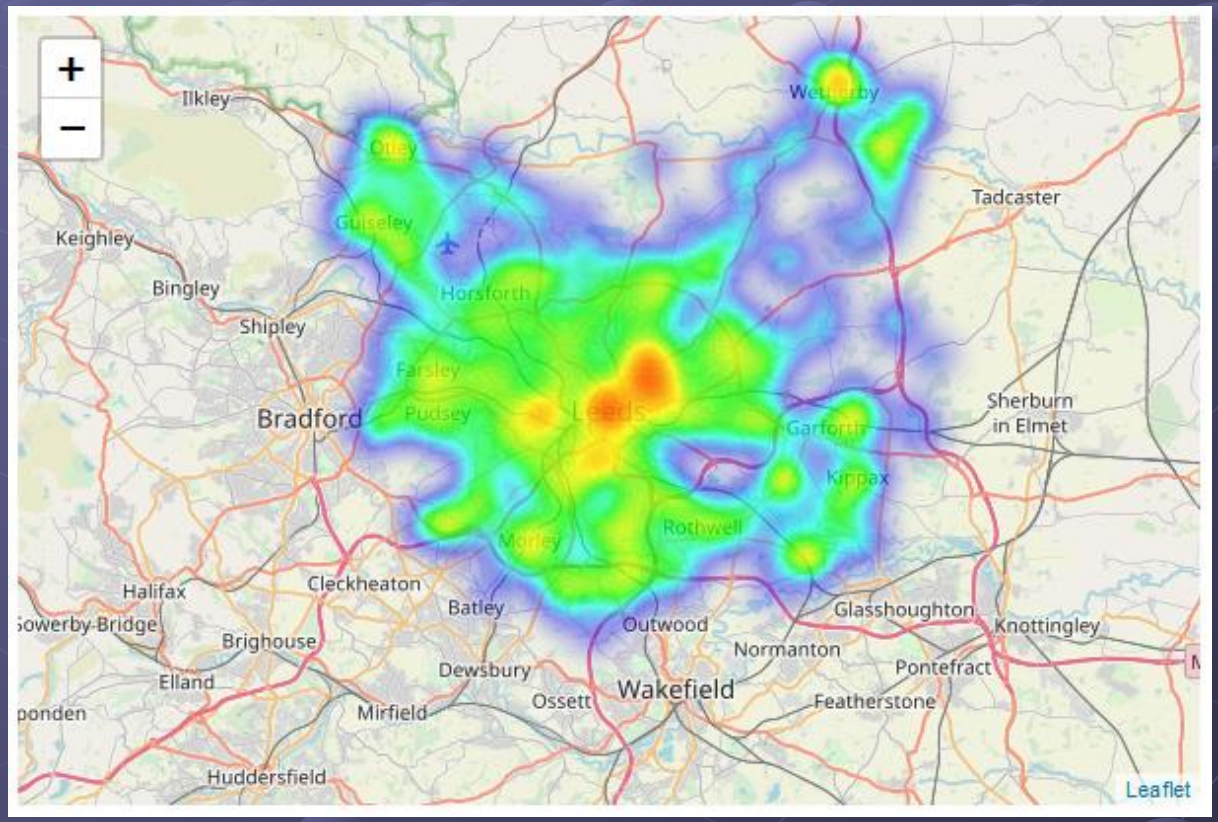
- An identifiable business problem relates to the ability to classify any specific geographic location within the Leeds area according to potential levels of crime based on the composition of the surrounding area. In other words:

- ***"Can we determine if an area around a specific geographic point is likely to have an above or below average risk of crime, based on the numbers and types of locations surrounding it ?"***

- The ability to answer this question could benefit a range of potential stakeholders. People looking to move home could assess what the levels of crime would be at a potential address. Local Police could identify crime hotspots, while planning authorities could determine if any new developments or change of use applications would, lead to increased (or decreased) crime in the vicinity

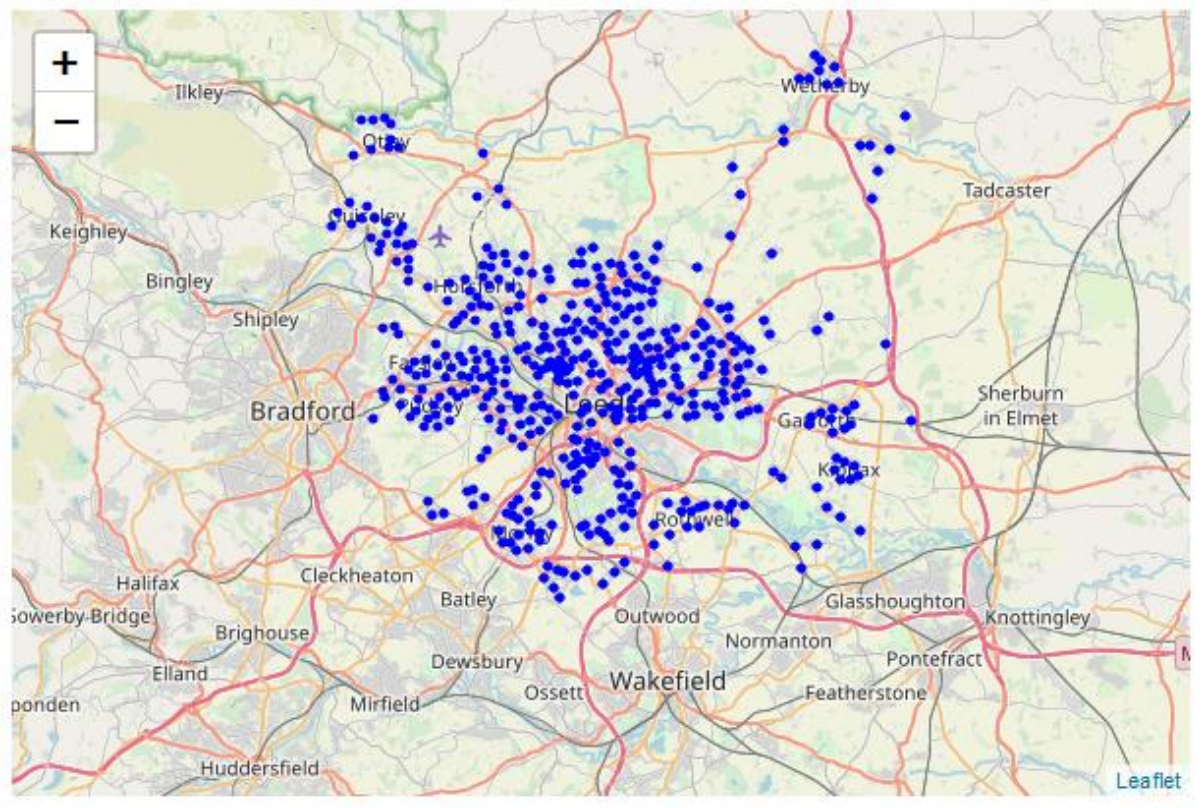
Crime Distribution In Leeds

- Frequency of crime varies by location, Shown below for recorded crimes in August 2019.



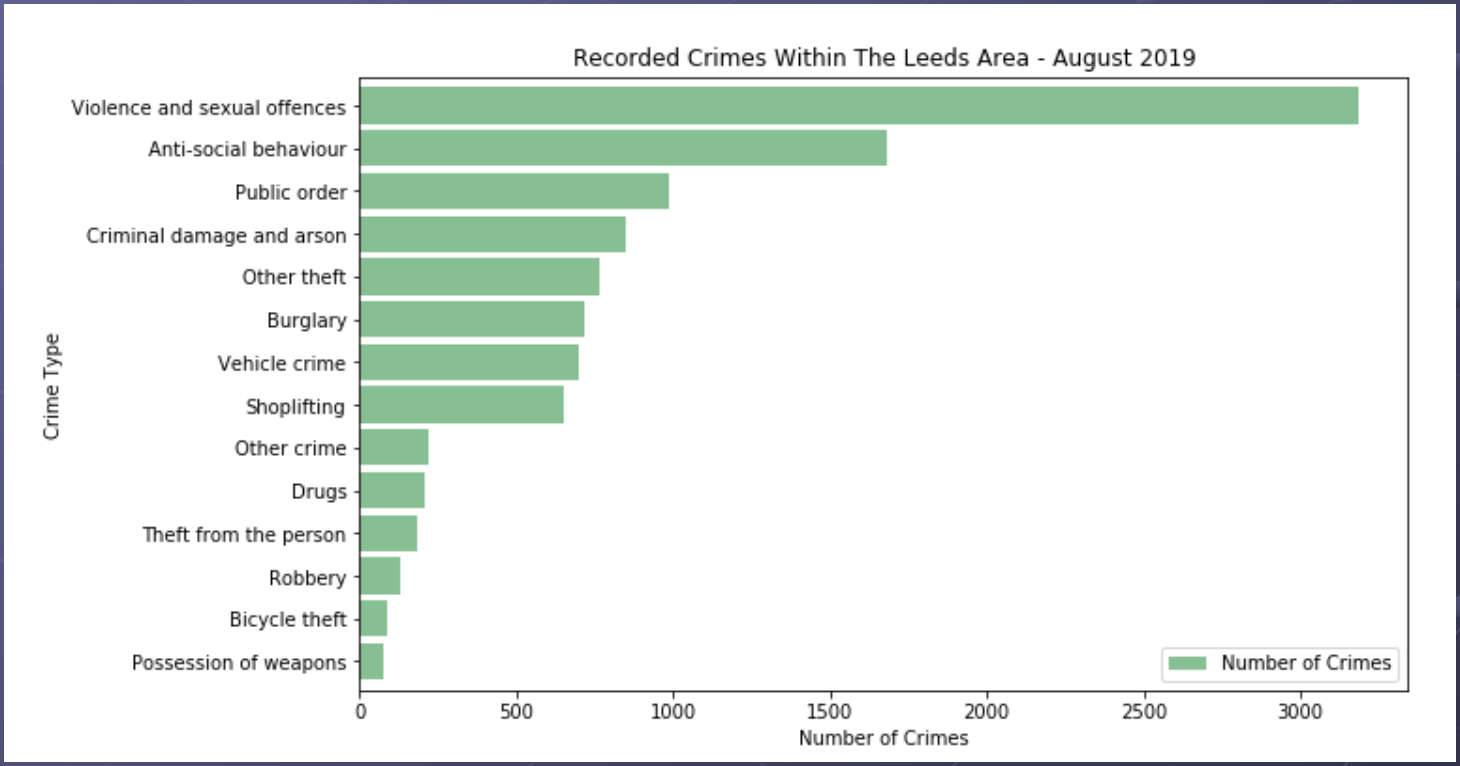
Location Points

- Over 400 locations were identified, based on exiting administrative centroids.



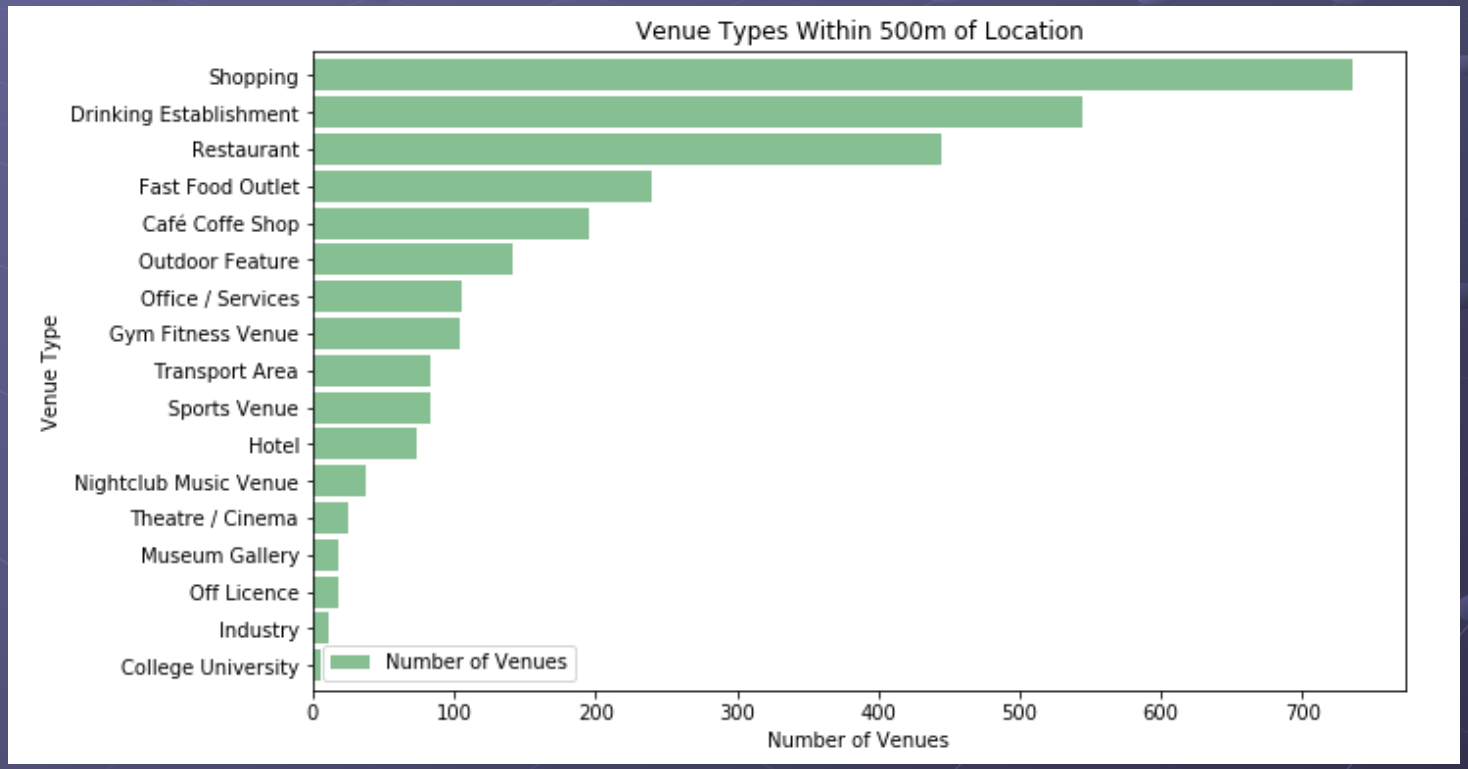
Crimes Within 500m

- Crimes within 500m of each centroid were identified..



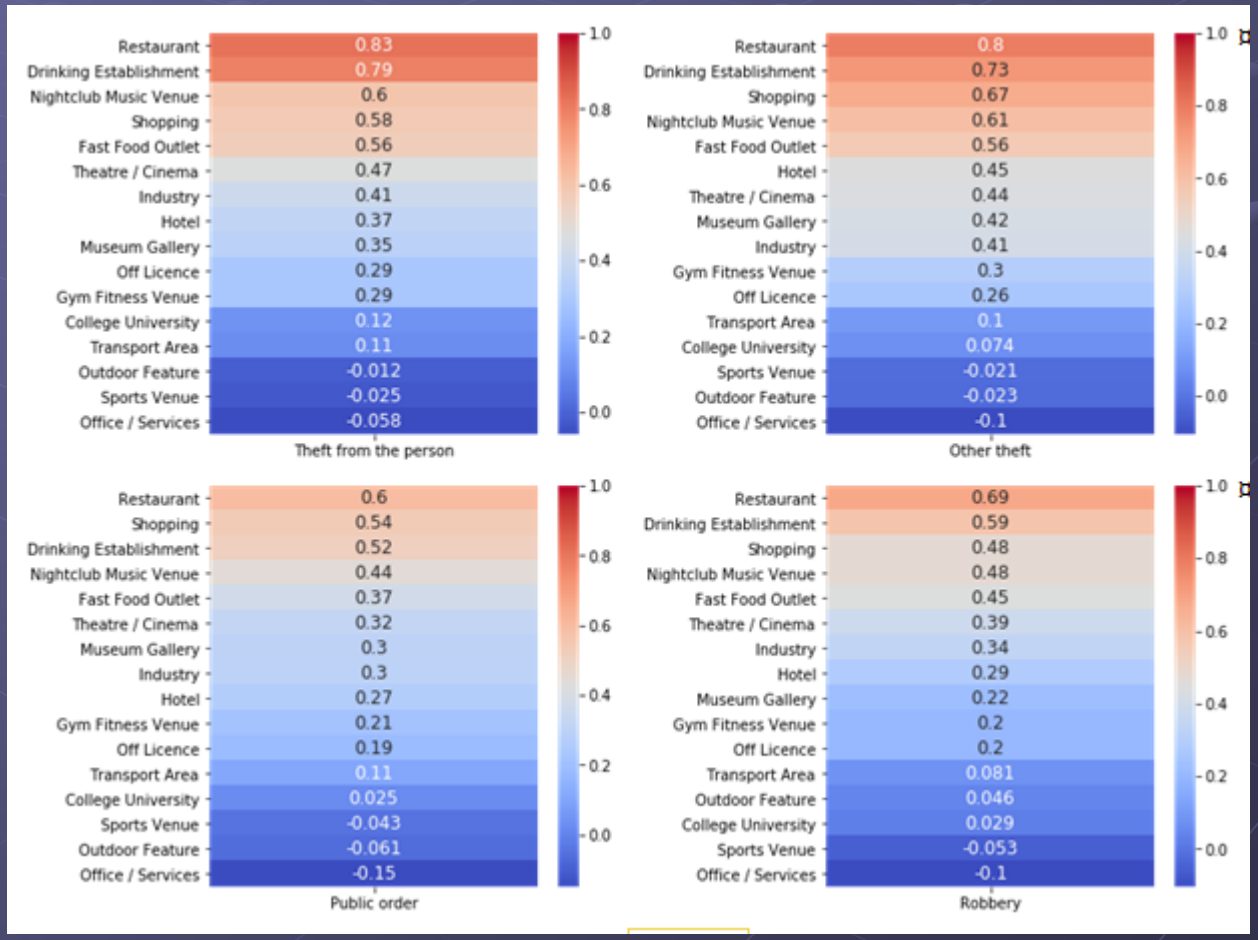
Locations Within 500m

- Locations within 500m of each centroid were identified..



Relationships

- Relationships between crimes and location type were identified.

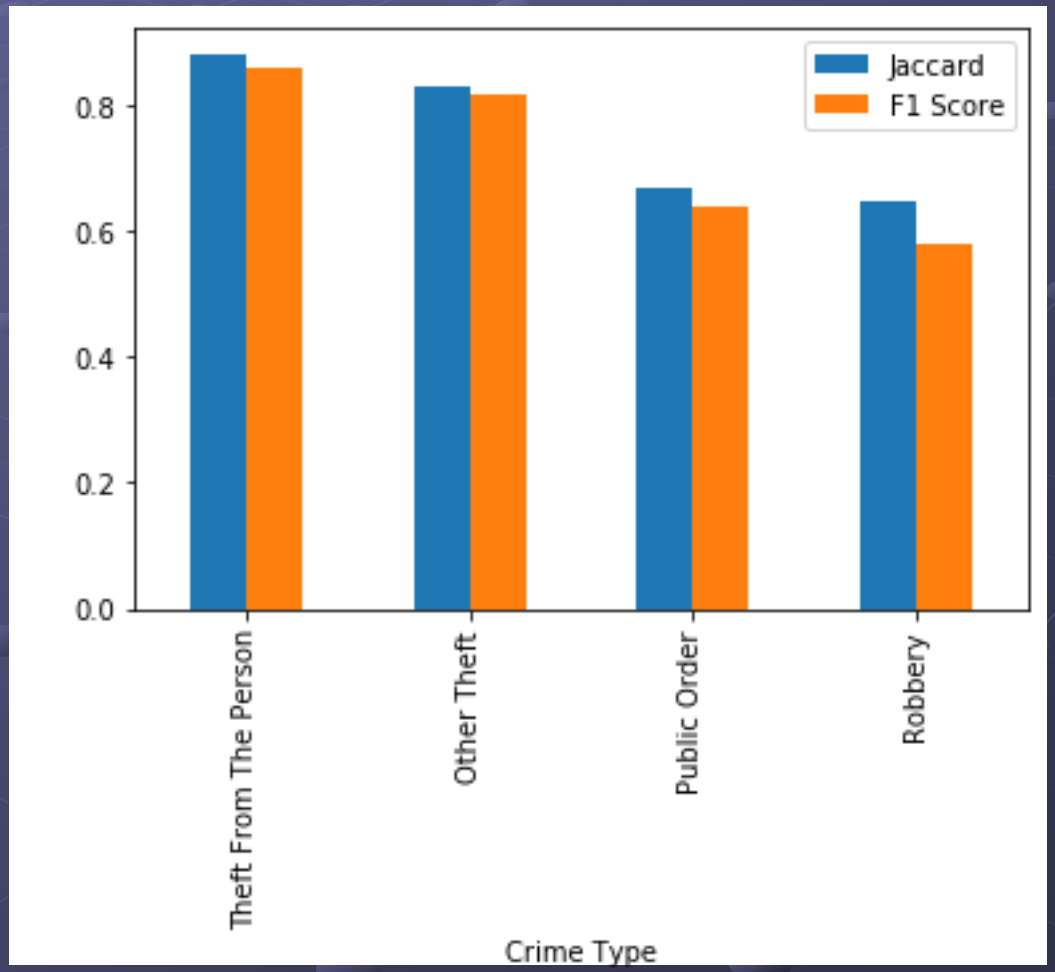


Modelling

- The data was modelled utilising the K-Nearest Neigbour classification model.

Results

- The accuracy of the models were high in relation to all four crime types tested.



Discussion

- The implementation of the models generated good results with a high accuracy when classifying the four crime types chosed. With some confidence the model could classify a given geographic location as a high or low crime risk in relation the certain crime types.

- The availability of the location data however would act as an influencer in these results. Using Foursquare as the location data provider resulted in relatively low numbers of location types being returned. Using an alternative source such as Google Places, may provide for a higher number of locations being included in the study. This would lead to a more representative reflection of what locations were in the vicinity and make for a more accurate data set and a more robust model.

- This could also mean that further crime types could be included in the modelling as opposed to just four. If this was the case, investigation into the more frequently occurring crime types, not included in this study, could be made (eg Violence and Sexual Offences).

Discussion (cont'd)

- The crime data used was the available published data, which was already categorised If data relating to Violence and Sexual offences could be broken down further this may have an impact on the data relationships and the modelling outcomes. Ie do the crimes camcel each other out when looking at them in relation to nearby locations)

- Additional classification models could also be tested in order to determine if a more accurate model could be utilised.

Conclusion

- In conclusion the implementation of the K-Nearest Neighbour classification model provided a high level of accuracy when classifying a location as high or low risk in relation to a number of crime types.
- The level of accuracy and validity of the models could potentially be improved if more complete information on nearby locations could be utilised and certain crime types being broken down into further categories.
- Using additional classification models may also result in greater levels of accuracy being returned.