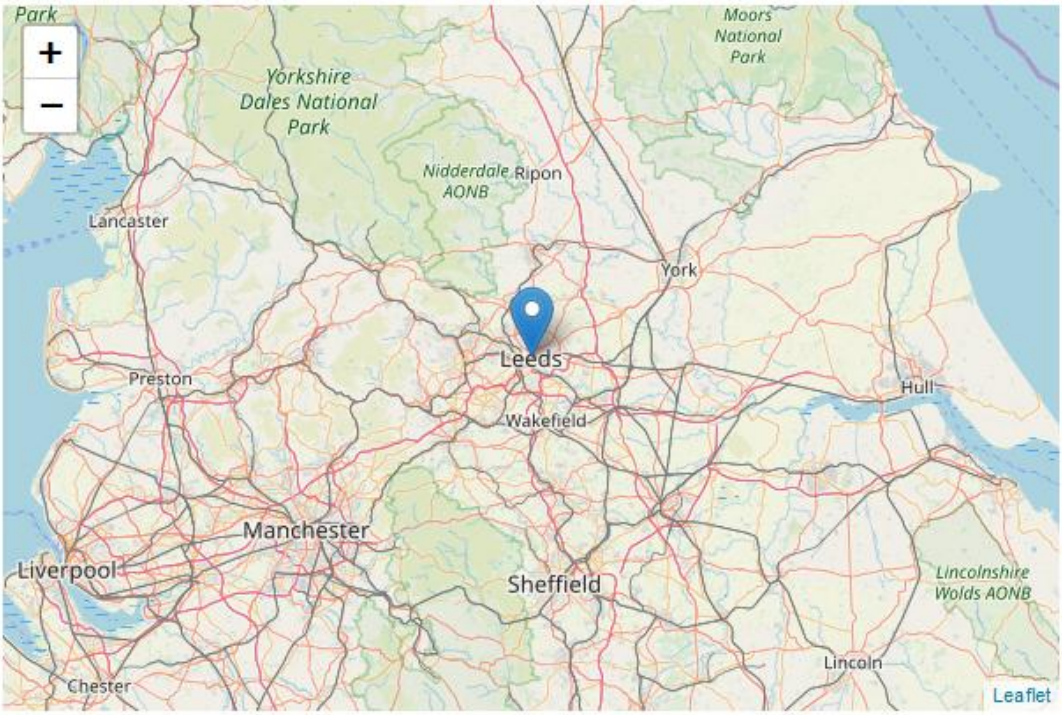
1. Introduction

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.

The City Of Leeds



2. 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.

3. Data

In order to answer the question posed, multiple data sets were required. These data sets related to crime within Leeds, geographic data relating to Leeds and location type details. To answer the question we need to obtain the number and types of venues and crimes which fall within a set distance of a geographical point.

Once retrieved and processed, the data will be used to create a classification model using the K-Nearest Neighbour classifier. The objective of the model is to classify any geographical point as high or low risk in relation to exposure to crime.

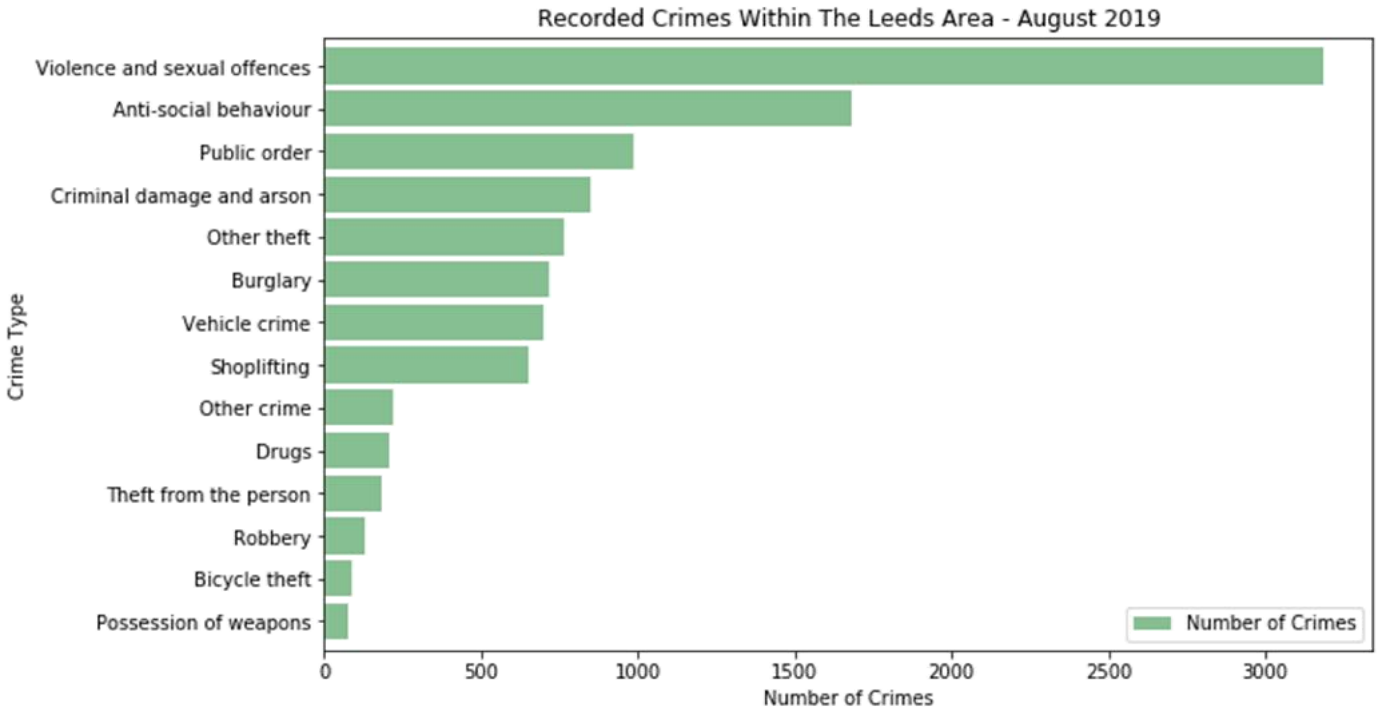
Crime Data:

Crime data relating to Leeds was obtained online from publicly available datasets published by the police.
(https://data.police.uk/data/)

This data was supplied by Police Force area and provided information with respect to the type of crime, the latitude and longitude of the crime and local area (LSOA) that the crime occurred within. The data was also supplied based on Month, in this case the most recent data-set was chosen, relating to August 2019.

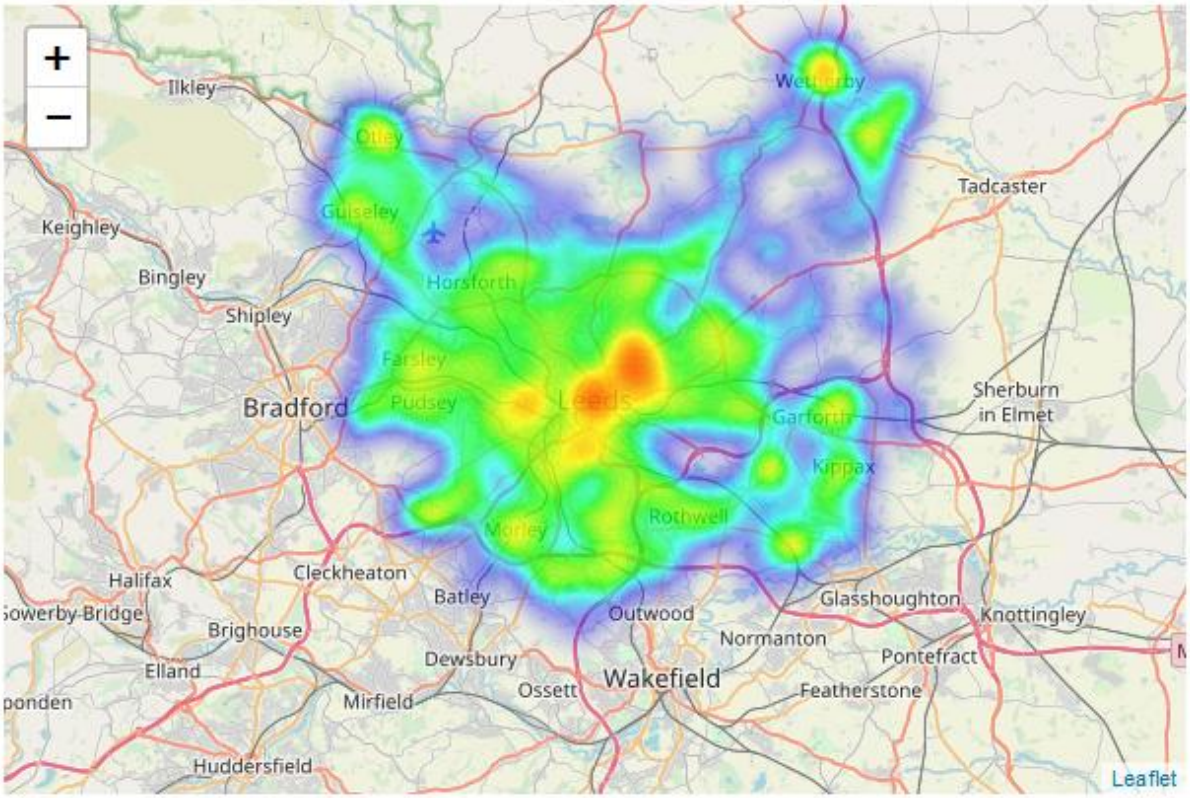
As data relating to the whole of West Yorkshire was included in the data-set, crimes relating to the Leeds areas had to be extracted, which was done based on the area codes that made up the Leeds area. The data was checked for missing values in any required columns (none were found)

The data-set contained details of 10,441 crimes, grouped by 14 distinct crime classifications. Details of the breakdown of these crimes is shown in the chart below, which shows that Violence and Sexual Offences were by far the most common crimes recorded..



The distribution of these crimes was also variable with urban centre areas showing higher concentrations of crime, especially in the Leeds city centre areas. The distribution of crime is shown on the heat-map below. The orange, yellow areas having the highest crime levels and the blue the lowest.

Distribution of Crime Within Leeds – August 2019



Leeds Geographic Data:

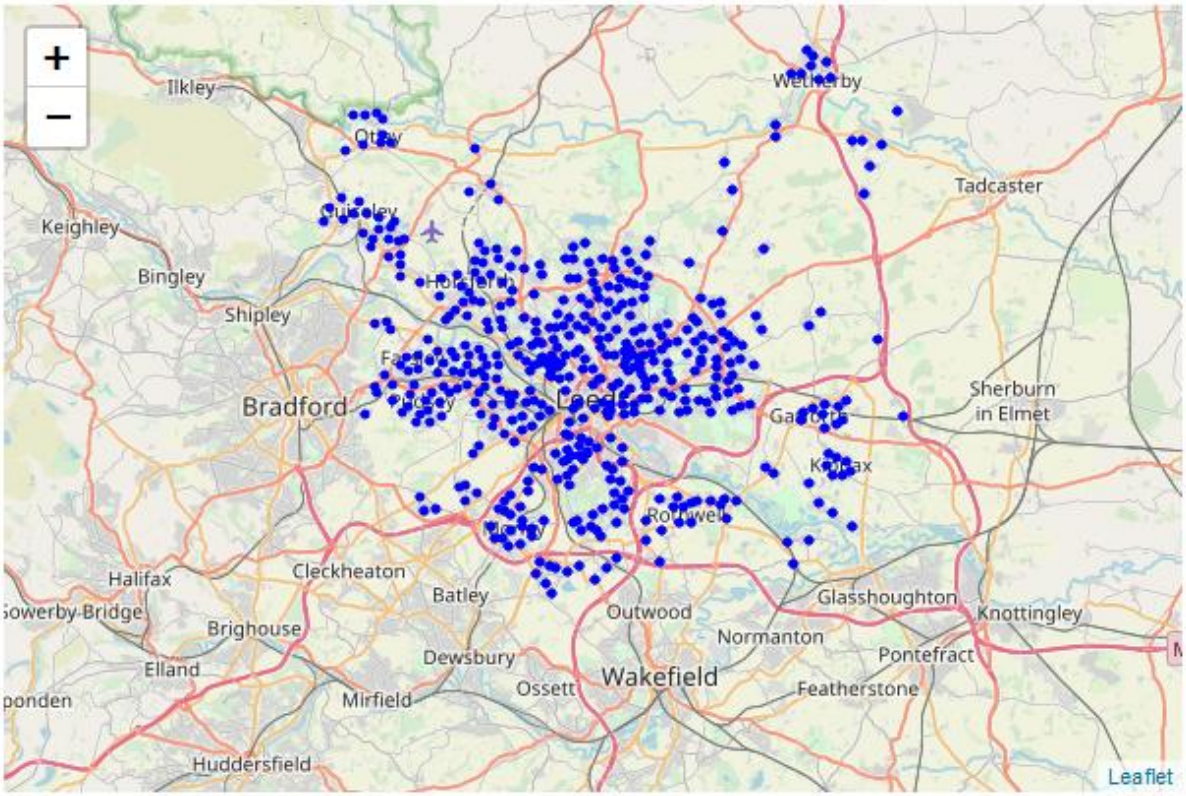
Publicly available Geographic Data relating to Leeds was obtained from The Office for National Statistics (https://geoportal.statistics.gov.uk/datasets/). This data comprised of geographic locations for the centroid of each LSOA (Lower Layer Super Output Area) within the Leeds area.

The data provided the locations of 482 geographic points within the Leeds area. Each LSOA code was unique and this was combined with a further set of data which assigned more user friendly Electoral Ward names to the area (Although each ward could consist of many LSOA codes, it provided a reference that would be potentially more meaningful).

These centroid locations were used as they fell within the geographical area that this project covered and related to already established points. The distribution of these points can be seen in the following map.

These points will be used as the basis for future analysis and comparison between areas.

Distribution of Geographic Points Within Leeds



Crimes within 500m Radius of Geographic Point

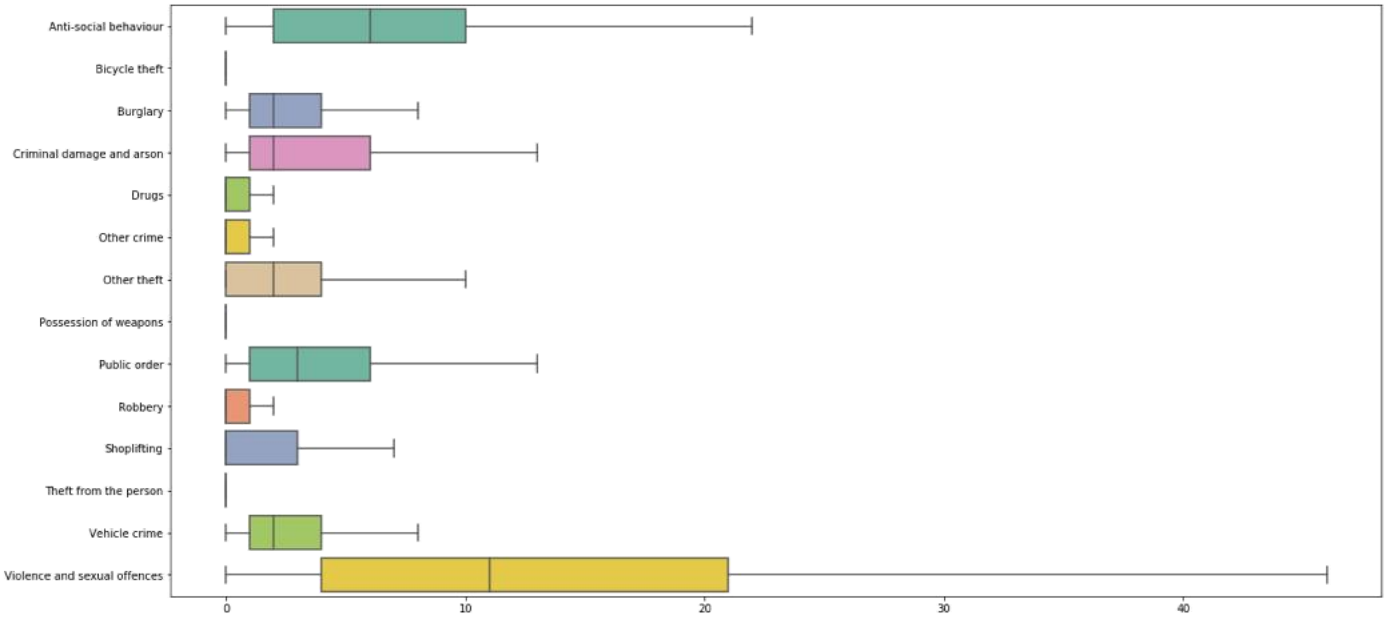
The next stage of data manipulation was to determine the number and types of crimes that fell within a radius of a given point. For each geographical location the crime data was iterated over, and if the crime fell within a radius of 500m from the geographical point, it was added to that points

(LSOA) data.

The data was then manipulated and aggregated in order for each LSOA point to be shown as a row and crime types as a columns. Showing the number of crimes by type within 500m of each geographic point..

The distribution of these crimes by area can be seen in the following chart. Outliers have been removed for illustrative purposes, but left within the data as they reflect valid data.

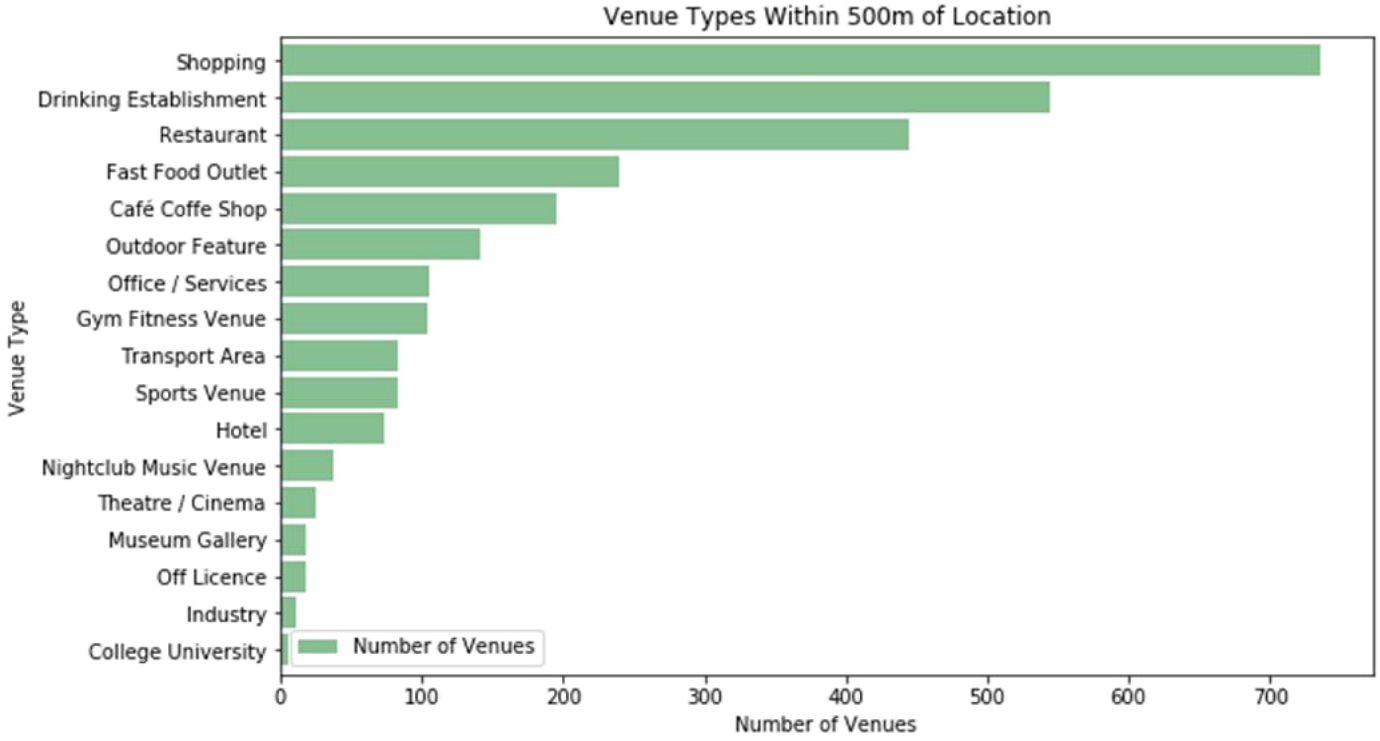
Box Plot Showing disribution of Crime Types by Area



Venues within 500m Radius of Geographic Point

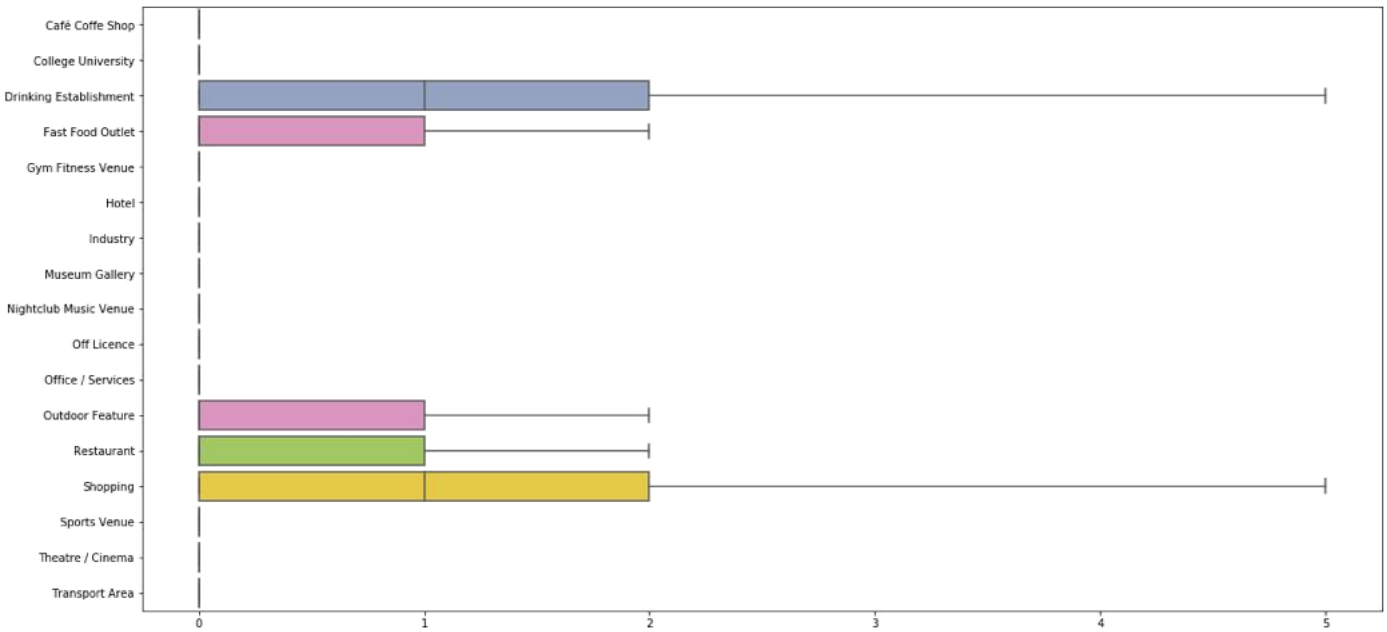
Collating venues types within a 500m radius of each point was the next step to be undertaken. Data available from Foursquare (https://foursquare.com/) was interrogated, with details of up to 100 locations within a radius of 500m being retrieved for each geographical point.

230 Unique venue types were retrieved through this process. In order to reduce this number to a more manageable and meaningful list of venue types, the 230 unique venues were mapped to a revised list of 17 high level categories. Numbers of venues within each category are shown below



As with the crime data, this venue data was aggregated and manipulated in order to show venue types by column and the number of each venue by area (row). The chart below shows the distribution of these venues by area (outliers have been removed, but remain in the data as they are valid)

Box Plot Showing the Distribution of Venue Type by Area



Amalgamantion of Data

The final step in the preparation of data was to combine all the data into one table. The location data was merged with the crime data already process. A final data set containing all the data

required to proceed with the analysis was created, detailing crime and location type numbers within 500m of each geographic location.