2. Data Acquisition and Cleaning

2.1 Data Acquisition

The first data source of the project uses Denver crime data from kaggle website that shows the crime data in each Neighborhood in London.

The dataset contains the following columns:

- OFFENSE_TYPE_ID: which gives which type of offense it is.
- NEIGHBORHOOD_ID: This will give the Neighborhood in which offense reported
- IS_CRIME : This showing the offense is a crime
- Is_TRAFFIC: which shows if the offense is traffic offense
- REPORTED_DATE: This include crime reported date & time

The second data source of data Latitude & Longitude is extracted using geocoder for each Neighborhood.

The third data source is the list of nearby venues for Neighborhoods of the Denver and Venues information using Foursquare.

Leveraged Foursquare to extract the following data.

Venue Name
Venue Latitude
Venue Longitude
Venue category

2.2 Data Cleaning

From the Denver crime data, the crimes during the most recent year 2019 are only selected.

	OFFENSE_CODE	OFFENSE_TYPE_ID	OFFENSE_CATEGORY_ID	NEIGHBORHOOD_ID	IS_CRIME	year
1568	1208	robbery-residence	robbery	east-colfax	1	2019
23326	5499	traf-other	all-other-crimes	jefferson-park	1	2019
23327	2404	theft-of-motor-vehicle	auto-theft	west-highland	1	2019
23328	2303	theft-shoplift	larceny	city-park	1	2019
23329	2999	criminal-mischief-mtr-veh	public-disorder	montbello	1	2019

The major categories of crime are pivoted to get the total crimes per Neighborhood as per category.

	Neighborhood	aggravated-assault	all-other-crimes	Earson	auto-theft	burglary	drug-alcohol	larceny	murder	other-crimes-against-persons	public-disorder	robbery	sexual-assault	theft-from-motor-vehicle	white-collar-crime	Total
0	athmar-park	33	207	1	63	28	98	58	1	48	81	16	9	86	21	750
1	auraria	15	271	1	19	24	28	97	0	52	100	11	8	35	11	672
2	baker	28	227	6	72	49	120	262	0	55	133	21	7	123	20	1123
3	barnum	35	212	2	40	27	47	31	1	46	91	17	10	51	11	621
4	barnum-west	13	100	3	41	14	42	11	1	16	60	2	4	38	4	349

The second data of Neighbourhood Latitude and Longitude data extracted using Geocoder.

Neighborhood	aggravated-assault	all-other-crimes	Earson	auto-theft	burglary	drug-alcohol	larceny	murder	other-crimes-against-persons	public-disorder	robbery	sexual-assault	theft-from-motor-vehicle	white-collar-crime	Total	latitude
wellshire	1	4	0	6	12	3	5	0	4	13	0	0	18	2	68	39.658484
indian-creek	1	6	0	9	4	1	11	1	5	13	3	0	26	4	84	39.684761
country-club	3	22	0	12	34	2	17	0	3	11	3	0	33	1	141	39.720100
skyland	6	69	1	13	11	8	16	2	11	22	1	3	26	5	194	39.755713
rosedale	4	30	0	17	19	21	26	0	7	23	1	4	49	5	206	39.739236

The third data of nearby venues of Neighborhood using Foursquare API.

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category	Cluster Label	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Ven
0	wellshire	39.658484	-104.941353	Wellshire Golf Course	39.658276	-104.941329	Golf Course	2	Golf Course	Athletics & Sports	Playground	Park	Yc Stu
1	wellshire	39.658484	-104.941353	Mamie D. Eisenhower Park	39.658946	-104.937348	Park	2	Golf Course	Athletics & Sports	Playground	Park	Yc Stu
2	wellshire	39.658484	-104.941353	Wellshire Inn	39.658247	-104.941277	Golf Course	2	Golf Course	Athletics & Sports	Playground	Park	Yc Stu
3	wellshire	39.658484	-104.941353	Eisenhower Rec Center	39.659563	-104.937773	Athletics & Sports	2	Golf Course	Athletics & Sports	Playground	Park	Yc Stu
4	wellshire	39.658484	-104.941353	Eisenhower Park	39.659608	-104.936572	Playground	2	Golf Course	Athletics & Sports	Playground	Park	Yc Stu