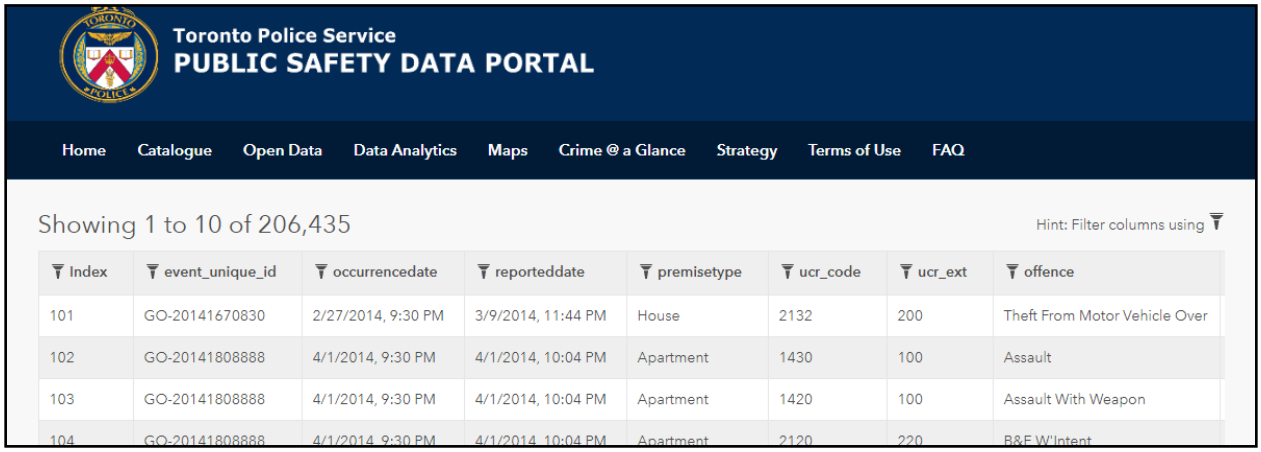
DATA

Data Description

The data in "**Public Safety Data Portal"** is available is .csv, geojson format for public use. Geojson file format will be used here in the analysis as it represents the real-time data and will be able to make this data analysis viable for a long time. The typical data set in the Police portal looks like the one in **Figure 1 : Toronto city- Crime data set**.



Index	event_unique_id	occurrencedate	reporteddate	premisetype	ucr_code	ucr_ext	offence
101	GO-20141670830	2/27/2014, 9:30 PM	3/9/2014, 11:44 PM	House	2132	200	Theft From Motor Vehicle Over
102	GO-20141808888	4/1/2014, 9:30 PM	4/1/2014, 10:04 PM	Apartment	1430	100	Assault
103	GO-20141808888	4/1/2014, 9:30 PM	4/1/2014, 10:04 PM	Apartment	1420	100	Assault With Weapon
104	GO-20141808888	4/1/2014, 9:30 PM	4/1/2014, 10:04 PM	Apartment	2120	220	B&E W'Intent

Figure 1 : Toronto city- Crime data set


```
Out[12]: [{'type': 'Feature',
           'properties': {'Index_': 7801,
            'event_unique_id': 'GO-20152165447',
            'occurrencedate': '2015-12-18T03:58:00.000Z',
            'reporteddate': '2015-12-18T03:59:00.000Z',
            'premisetype': 'Commercial',
            'ucr_code': 1430,
            'ucr_ext': 100,
            'offence': 'Assault',
            'reportedyear': 2015,
            'reportedmonth': 'December',
            'reportedday': 18,
            'reporteddayofyear': 352,
            'reporteddayofweek': 'Friday    ',
            'reportedhour': 3,
            'occurrenceyear': 2015,
            'occurrencemonth': 'December',
            'occurrenceday': 18,
            'occurrencedayofyear': 352,
            'occurrencedayofweek': 'Friday    ',
            'occurrencehour': 3,
            'MCI': 'Assault',
            'Division': 'D14',
            'Hood_ID': 79,
            'Neighbourhood': 'University (79)',
            'Long': -79.4052277,
            'Lat': 43.6569824,
            'ObjectId': 7001},
           'geometry': {'type': 'Point', 'coordinates': [-79.4052277, 43.6569824]}}]
```


Figure 2 : Geojson data of crimes in Toronto

The data set when imported from the portal as Geojson file and explored, the output looks like the one in **Figure 2 : Geojson data of crimes in Toronto**. The data represents the type of crime, reported and occurance date/month/time, Longitude and Latitude of the crime location. The longitude and latitude fields of the crime dataset will be used to see the number of venues within 500 meters of the locations using Foursquare API.

```
results = requests.get(url).json()
results
```


```
{'meta': {'code': 200, 'requestId': '5e915efeed78b8001b03b8d6'},
 'response': {'headerLocation': 'Corktown',
  'headerFullLocation': 'Corktown, Toronto',
  'headerLocationGranularity': 'neighborhood',
  'totalResults': 45,
  'suggestedBounds': {'ne': {'lat': 43.6587599045, 'lng': -79.3544279001486},
   'sw': {'lat': 43.6497598955, 'lng': -79.36684389985142}},
  'groups': [{'type': 'Recommended Places',
    'name': 'recommended',
    'items': [{'reasons': {'count': 0,
       'items': [{'summary': 'This spot is popular',
         'type': 'general',
         'reasonName': 'globalInteractionReason'}]},
      'venue': {'id': '54ea41ad498e9a11e9e13308',
       'name': 'Roselle Desserts',
       'location': {'address': '362 King St E',
        'crossStreet': 'Trinity St',
        'lat': 43.653446723052674,
        'lng': -79.3620167174383,
        'labeledLatLngs': [{'label': 'display',
          'lat': 43.653446723052674,
          'lng': -79.3620167174383}],
        'distance': 143,
```


Figure 3: Four square API data for identified location data

And the foursquare API data looks like the one in **Figure 3: Four square API data for identified location data**. This data gives the venue name, address, latitude and longitude and distance from the requested locations. The data gathered from the Foursquare API will be compared with crime data and correlation analysis is carried out.

```
In [13]: ColumnNames = ['offence','reportedyear','reportedmonth','reportedday','reporteddayofyear', 'reporteddayofweek','reportedhour','occurrenceyear','occurrencemonth','occurrenceday','occurrencedayofyear','occurrencedayofweek','occurrencehour','MCI','Division','Hood_ID','Neighbourhood','Longitude','Latitude']
```

Figure 4 : Relevant columns extracted for Analysis

Out of all the data obtained from the portal, only few fields will be input into the data frame for further analysis as shown in **Figure 4 : Relevant columns extracted for Analysis.**