Introduction where you discuss the business problem and who would be interested in this project.

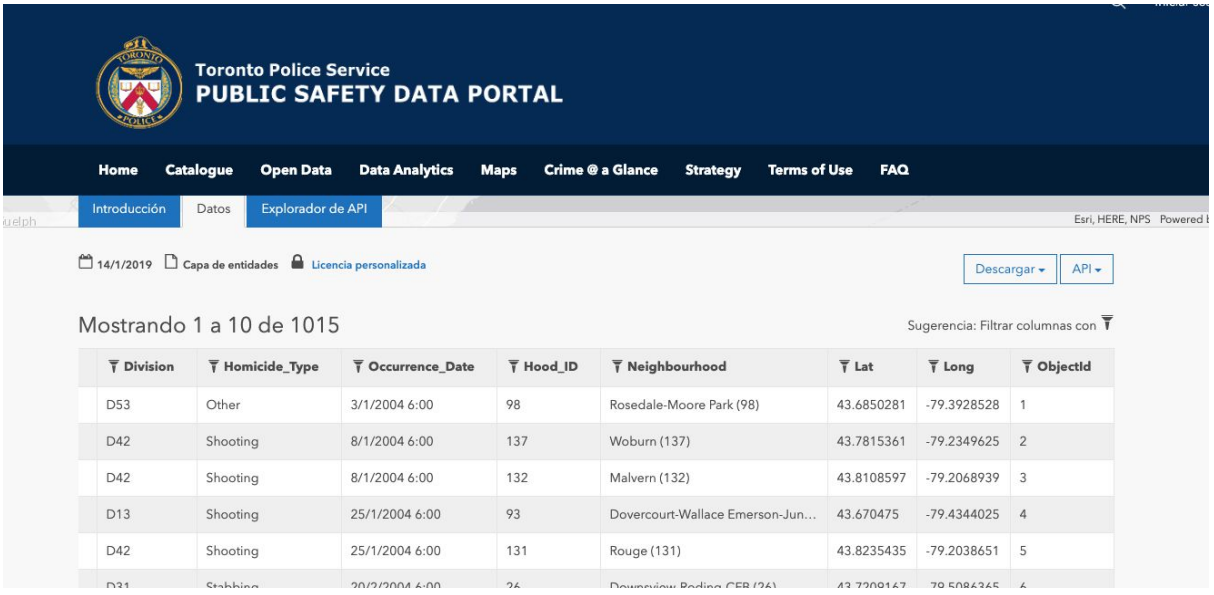
My partner and I moved to Toronto, and we want to start a new business.
The business we want to set up is a Stationery shop, in which we want to sell books and school supplies for children, but with an ecological point of view, trying to sell products with recycled material.

However, we don't know the city of Toronto very well, therefore we want to analyze the situation of the city, and for this we want to look for:

- A safe place
- Near schools
- That there is not much competition

Data where you describe the data that will be used to solve the problem and the source of the data.
To solve the information needs we have about the city of Toronto we need the following data sources.

- A safe place - We need a crime index in the regions of the city of Toronto



Division	Homicide_Type	Occurrence_Date	Hood_ID	Neighbourhood	Lat	Long	ObjectId
D53	Other	3/1/2004 6:00	98	Rosedale-Moore Park (98)	43.6850281	-79.3928528	1
D42	Shooting	8/1/2004 6:00	137	Woburn (137)	43.7815361	-79.2349625	2
D42	Shooting	8/1/2004 6:00	132	Malvern (132)	43.8108597	-79.2068939	3
D13	Shooting	25/1/2004 6:00	93	Dovercourt-Wallace Emerson-Jun...	43.670475	-79.4344025	4
D42	Shooting	25/1/2004 6:00	131	Rouge (131)	43.8235435	-79.2038651	5

- Near schools - Number of schools and their location

```
dataframe = json_normalize(venues)
pd.DataFrame(dataframe).head()
```

location.formattedAddress	location.labeledLatLngs	location.lat	location.lng	location.neighborhood	location.postalCode	location.state	name	referralId	venuePage.id
[575 Bay St. (at Dundas St. W), Toronto ON M5G...	[{'label': 'display', 'lat': 43.65564568498175...	43.655646	-79.382737	NaN	M5G 2C5	ON	Ted Rogers School of Management	v-1574759203	NaN
[155 College St. (at McCaul St.), Toronto ON M...	[{'label': 'display', 'lat': 43.65923191282564...	43.659232	-79.393254	NaN	M5T 3M7	ON	Dalla Lana School of Public Health	v-1574759203	NaN
[105 St. George St. (University of Toronto), T...	[{'label': 'display', 'lat': 43.6647972002443,...	43.664797	-79.398576	NaN	M5S 3E6	ON	Rotman School of Management	v-1574759203	NaN
[439 University Ave. (at Dundas St. W), Toront...	[{'label': 'display', 'lat': 43.65453291877178...	43.654533	-79.387831	NaN	NaN	ON	Ontario Public School Boards' Association	v-1574759203	NaN
[64 Baldwin Street, Toronto	[{'label': 'display', 'lat':						Beverly	v-	

- That there is not much competition - Number of stores related to my business

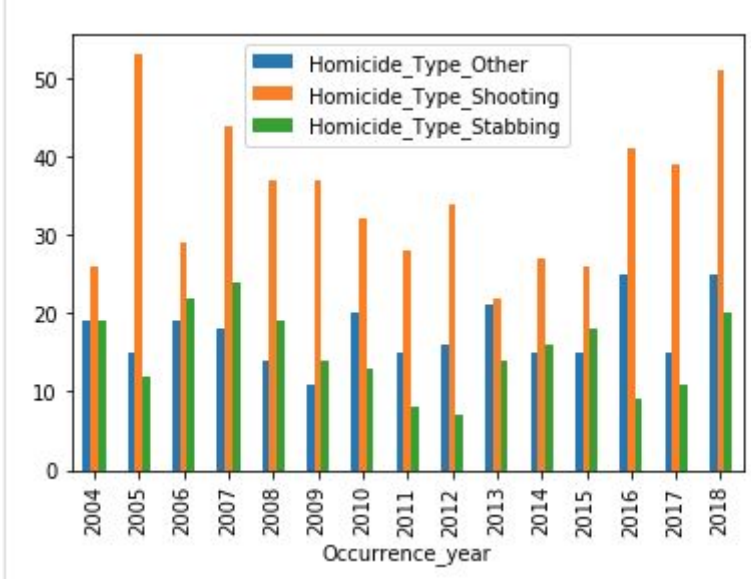
```
pd.DataFrame(dataframe).head()
```

Out[65]:

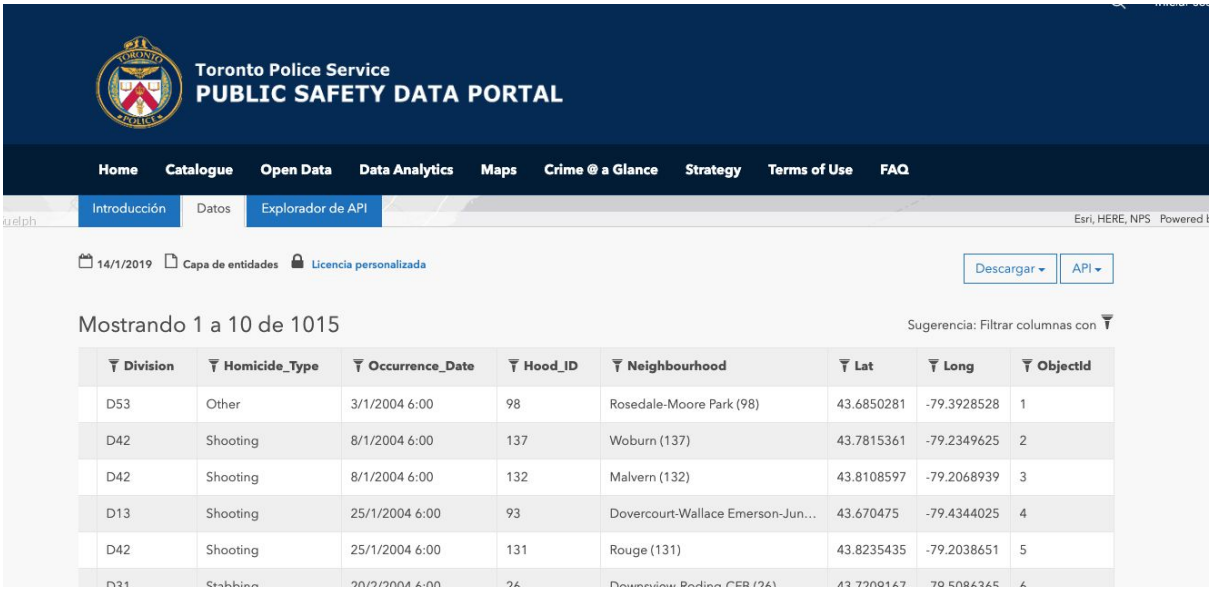
ition.distance	location.formattedAddress	location.labeledLatLngs	location.lat	location.lng	location.postalCode	location.state	name	referralId	venuePage.id
2174	[887 Queen St W (nr Walnut Ave.), Toronto ON M...	[{'label': 'display', 'lat': 43.645665, 'lng':...	43.645665	-79.411644	M6J 1G5	ON	The Paper Place	v-1574759403	102345330
1133	[191 Baldwin St, Toronto ON M5T 1M1, Canada]	[{'label': 'display', 'lat': 43.654556, 'lng':...	43.654556	-79.401260	M5T 1M1	ON	Lit Paper & Glass	v-1574759403	NaN
1927	[99 Yorkville Ave, Toronto ON, Canada]	[{'label': 'display', 'lat': 43.67094261959884...	43.670943	-79.391872	NaN	ON	Paper Things	v-1574759403	NaN
1615	[1 Yonge St, Toronto ON M5E 1W7, Canada]	[{'label': 'display', 'lat': 43.6429251, 'lng':...	43.642925	-79.374186	M5E 1W7	ON	Crepe Paper Crafts	v-1574759403	496121219
2369	[116 Massey Street, Toronto ON, Canada]	[{'label': 'display', 'lat': 43.64497586917162...	43.644976	-79.413871	NaN	ON	Paper Bag Records	v-1574759403	NaN

Methodology section which represents the main component of the report where you discuss and describe any exploratory data analysis that you did, any inferential statistical testing that you performed, if any, and what machine learnings were used and why.

The first thing I wanted to do is analyze the crime rate of the city of Toronto and check if it is a safe city analyzing the type of crimes that have been made.

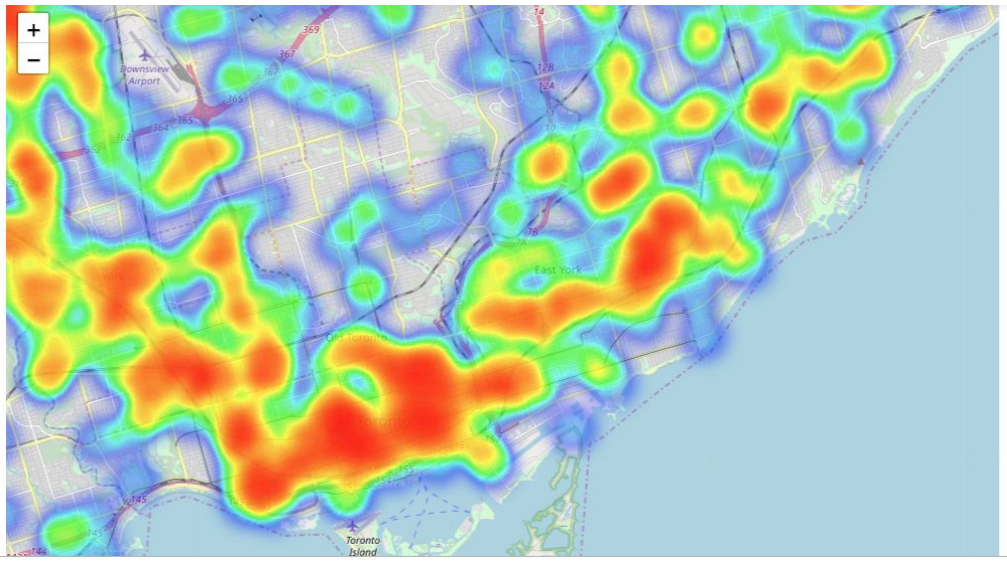


To do this I got a data source from http://data.torontopolice.on.ca/ and with that information and we analyzed this data and we saw in which places of the city they had commented.



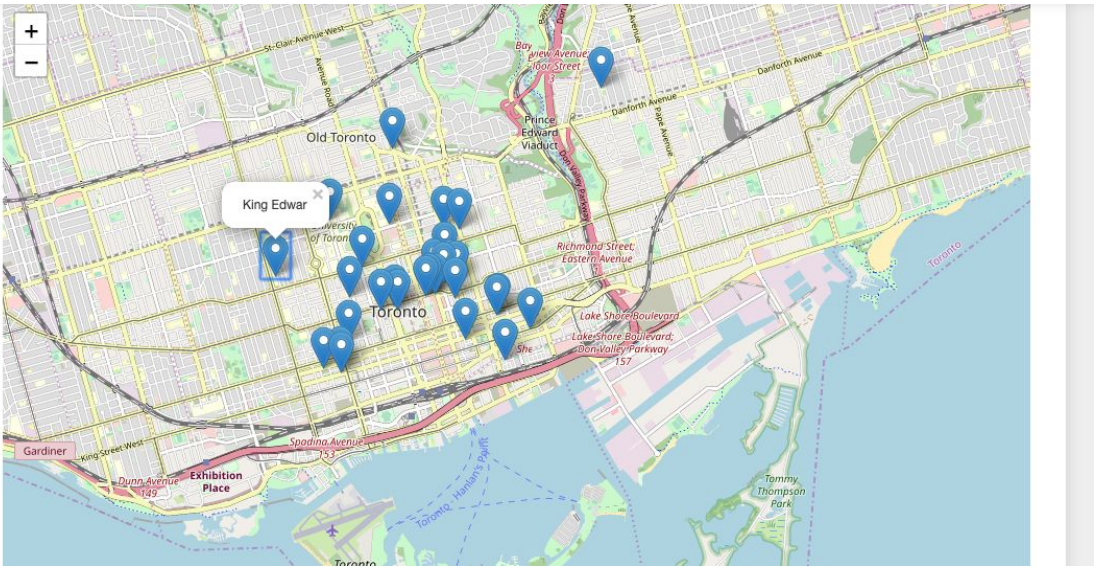
Heat map of homicidies in Toronto:

Out[133]:



Next I made a call to the Foursquare API to place on the map the schools and those similar to mine in the city.

Out[50]:



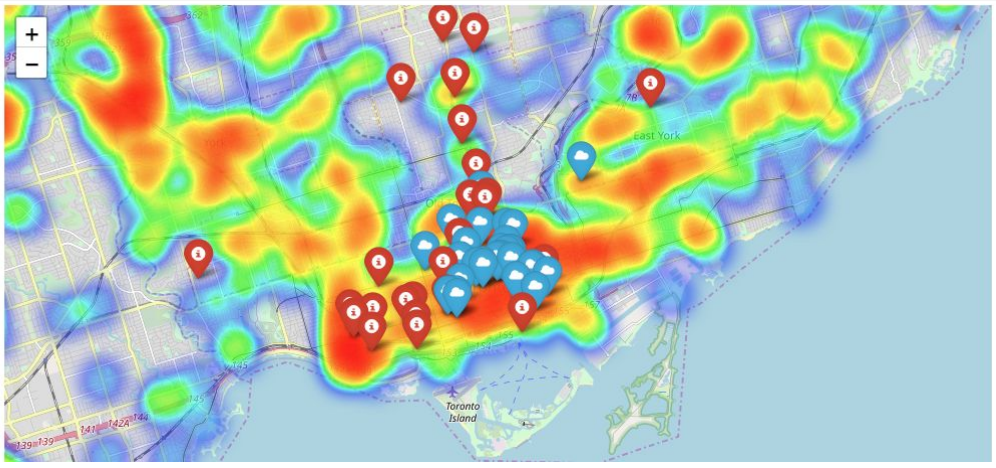
Place all the data on a map to visually see the best place to place my store.

Out[170]:

Then to be able to make the correct analysis we must insert all the information in a map to be able to see the ideal location.

```
world_map.add_child(incidents)
```

Out[55]:



4. results

Then I use the k means algorithm to create 4 clusters of the schools we have around:

Cluster 1:

```
In [88]: cluster_1.head()
```

Out[88]:

	name_x	name_y
1	0	Dalla Lana School of Public Health
2	0	Ontario Public School Boards' Association
3	0	Beverly Junior Public School
4	0	Rotman School of Management
11	0	The IELTS School of Toronto

Cluster 2:

```
In [89]: cluster_2.head()
```

Out[89]:

	name_x	name_y
26	1	SCHOOL Restaurant
29	1	Regal Road Junior Public School

Cluster 3:

```
In [90]: cluster_3.head()
```

Out[90]:

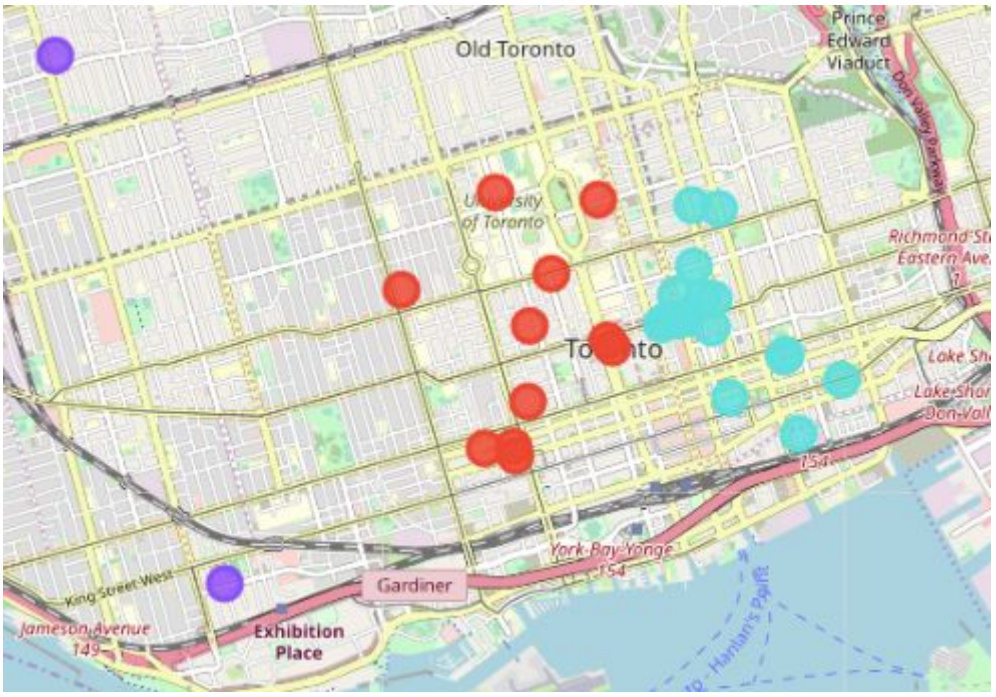
	name_x	name_y
0	2	Ted Rogers School of Management
5	2	George Brown School of Design
6	2	Ryerson School of Interior Design
7	2	Ted Rogers School of Management Patio
8	2	Canadian National Ballet School

Cluster 4:

```
In [91]: cluster_4.head()
```

Out[91]:

	name_x	name_y
25	3	Jackman Avenue Public School Junior



We return to compare the data with crime rates:

Conclusion

After all the information collected with the analysis we see that cluster 1 is the one that has the best characteristics to be able to start a business like ours.

- Reason 1: It is the cluster with the most schools with respect to its crime rate
- Reason 2: It's where we have the least competition