Doug Marcum

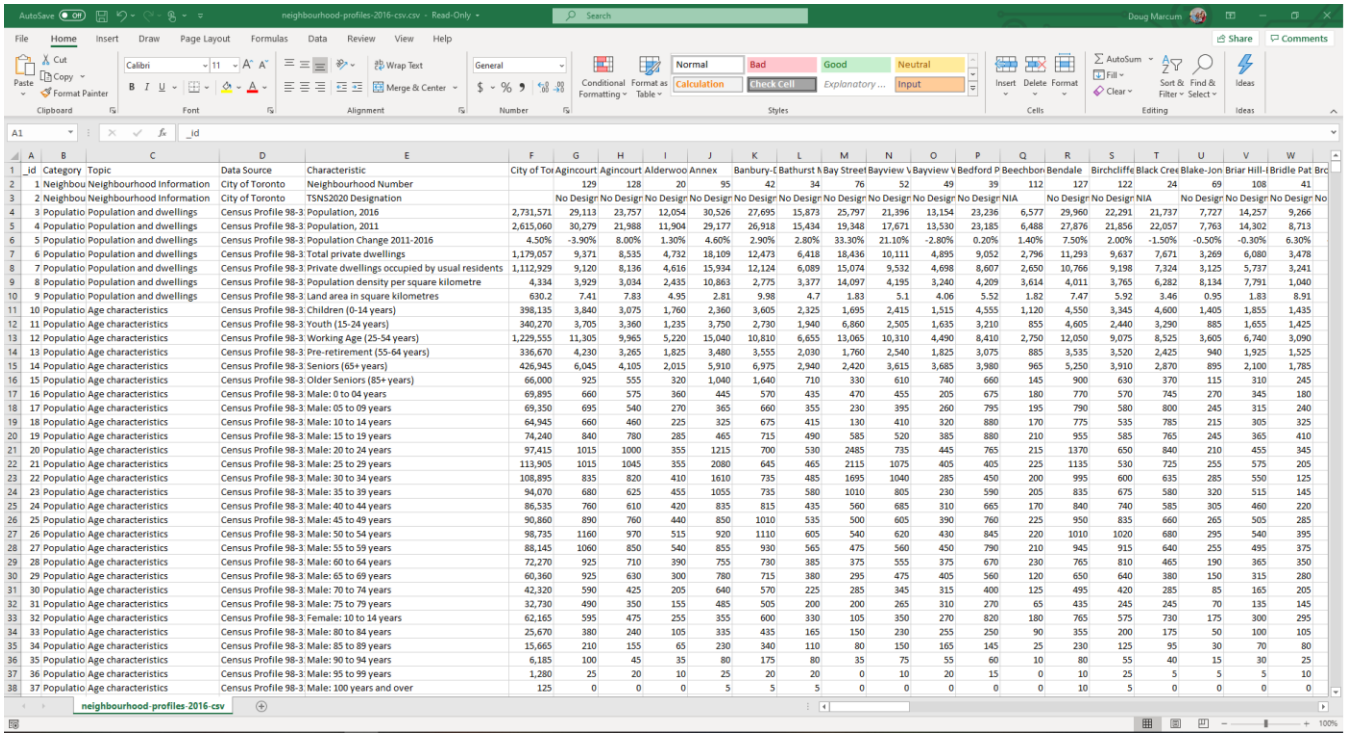
DSC 540 – Term Project: Milestone 1

After multiple attempts to find this much data about Indianapolis, then Chicago, then just about every other major city in the United States, I decided to venture outside of our boarders to our neighbors up North. There, in the open data sources of Canada, I discovered what I had been searching for this entire time. Posted below are links to the three data sets I will be using for the term project, a description of each, the relationship between each, and what will need to be accomplished to complete the five milestones.

Data - Sources and Descriptions

Toronto Neighborhood Profiles – CSV File

https://open.toronto.ca/dataset/neighbourhood-profiles/

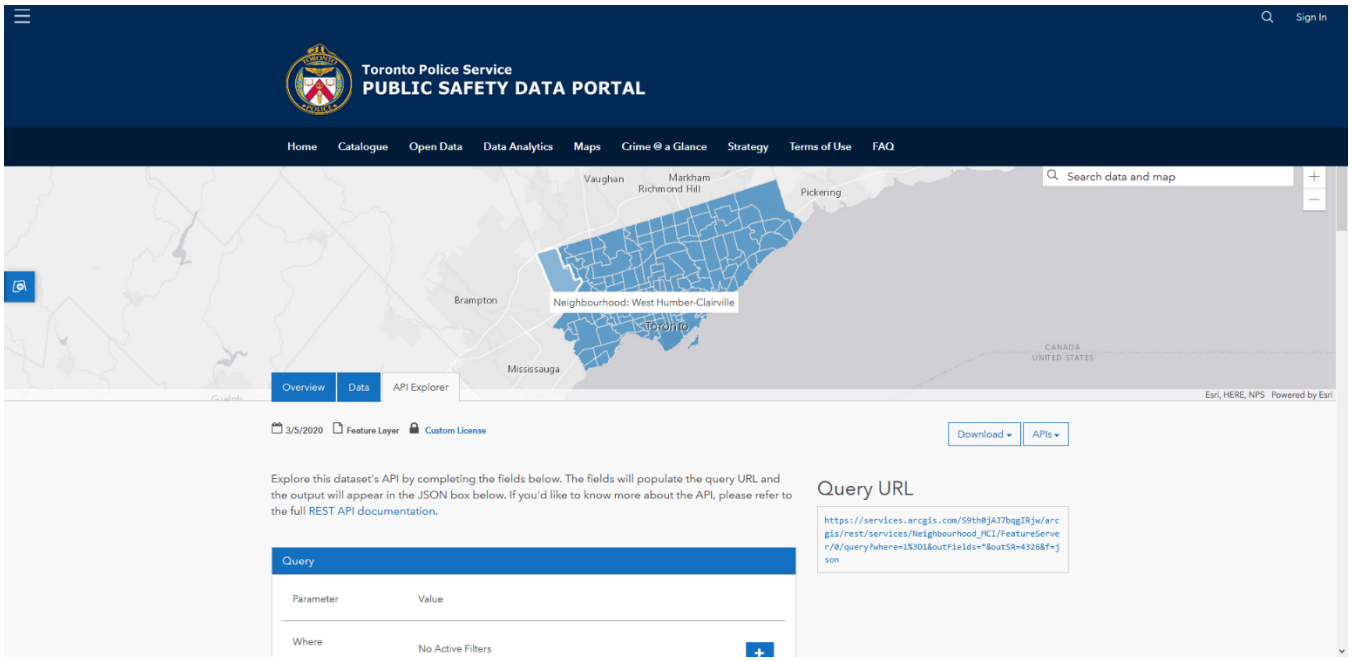


_id	Category	Topic	Data Source	Characteristic	City of Tor	Agincourt	Agincourt	Alderwoo	Annex	Banbury-E	Bathurst M	Bay Street	Bayview V	Bayview V	Bedford P	Beechbor	Bendale	Birchcliffe	Black Cree	Blake-Jon	Briar Hill-I	Bridle Pat	Bro
1	Neighbou	Neighbourhood Information	City of Toronto	Neighbourhood Number		129	128	20	95	42	34	76	52	49	39	112	127	122	24	69	108	41	
2	Neighbou	Neighbourhood Information	City of Toronto	TSNS2020 Designation		No Design	No Design	No Design	No Design	No Design	No Design	No Design	No Design	No Design	No Design	NIA	No Design	No Design	NIA	No Design	No Design	No Design	No
3	Populatio	Population and dwellings	Census Profile 98-3	Population, 2016	2,731,571	29,113	23,757	12,054	30,526	27,695	15,873	25,797	21,396	13,154	23,236	6,577	29,960	22,291	21,737	7,727	14,257	9,266	
4	Populatio	Population and dwellings	Census Profile 98-3	Population, 2011	2,615,060	30,279	21,988	11,904	29,177	26,918	15,434	19,348	17,671	13,530	23,185	6,488	27,876	21,856	22,057	7,763	14,302	8,713	
5	Populatio	Population and dwellings	Census Profile 98-3	Population Change 2011-2016	4.50%	-3.90%	8.00%	1.30%	4.60%	2.90%	2.80%	33.30%	21.10%	-2.80%	0.20%	1.40%	7.50%	2.00%	-1.50%	-0.50%	-0.30%	6.30%	
6	Populatio	Population and dwellings	Census Profile 98-3	Total private dwellings	1,179,057	9,371	8,535	4,732	18,109	12,473	6,418	18,436	10,111	4,895	9,052	2,796	11,293	9,637	7,671	3,269	6,080	3,478	
7	Populatio	Population and dwellings	Census Profile 98-3	Private dwellings occupied by usual residents	1,112,929	9,120	8,136	4,616	15,934	12,124	6,089	15,074	9,532	4,698	8,607	2,650	10,766	9,198	7,324	3,125	5,737	3,241	
8	Populatio	Population and dwellings	Census Profile 98-3	Population density per square kilometre	4,334	3,929	3,034	2,435	10,863	2,775	3,377	14,097	4,195	3,240	4,209	3,614	4,011	3,765	6,282	8,134	7,791	1,040	
9	Populatio	Population and dwellings	Census Profile 98-3	Land area in square kilometres	630.2	7.41	7.83	4.95	2.81	9.98	4.7	1.83	5.1	4.06	5.52	1.82	7.47	5.92	3.46	0.95	1.83	8.91	
10	Populatio	Age characteristics	Census Profile 98-3	Children (0-14 years)	398,135	3,840	3,075	1,760	2,360	3,605	2,325	1,695	2,415	1,515	4,555	1,120	4,550	3,345	4,600	1,405	1,855	1,435	
11	Populatio	Age characteristics	Census Profile 98-3	Youth (15-24 years)	340,270	3,705	3,360	1,235	3,750	2,730	1,940	6,860	2,505	1,635	3,210	855	4,605	2,440	3,290	885	1,655	1,425	
12	Populatio	Age characteristics	Census Profile 98-3	Working Age (25-54 years)	1,229,555	11,305	9,965	5,220	15,040	10,810	6,655	13,065	10,310	4,490	8,410	2,750	12,050	9,075	8,525	3,605	6,740	3,090	
13	Populatio	Age characteristics	Census Profile 98-3	Pre-retirement (55-64 years)	336,670	4,230	3,265	1,825	3,480	3,555	2,030	1,760	2,540	1,825	3,075	885	3,535	3,520	2,425	940	1,925	1,525	
14	Populatio	Age characteristics	Census Profile 98-3	Seniors (65+ years)	426,945	6,045	4,105	2,015	5,910	6,975	2,940	2,420	3,615	3,685	3,980	965	5,250	3,910	2,870	895	2,100	1,785	
15	Populatio	Age characteristics	Census Profile 98-3	Older Seniors (85+ years)	66,000	925	555	320	1,040	1,640	710	330	610	740	660	145	900	630	370	115	310	245	
16	Populatio	Age characteristics	Census Profile 98-3	Male: 0 to 04 years	69,895	660	575	360	445	570	435	470	455	205	675	180	770	570	745	270	345	180	
17	Populatio	Age characteristics	Census Profile 98-3	Male: 05 to 09 years	69,350	695	540	270	365	660	355	230	395	260	795	195	790	580	800	245	315	240	
18	Populatio	Age characteristics	Census Profile 98-3	Male: 10 to 14 years	64,945	660	460	225	325	675	415	130	410	320	880	170	775	535	785	215	305	325	
19	Populatio	Age characteristics	Census Profile 98-3	Male: 15 to 19 years	74,240	840	780	285	465	715	490	585	520	385	880	210	955	585	765	245	365	410	
20	Populatio	Age characteristics	Census Profile 98-3	Male: 20 to 24 years	97,415	1015	1000	355	1215	700	530	2485	735	445	765	215	1370	650	840	210	455	345	
21	Populatio	Age characteristics	Census Profile 98-3	Male: 25 to 29 years	113,905	1015	1045	355	2080	645	465	2115	1075	405	405	225	1135	530	725	255	575	205	
22	Populatio	Age characteristics	Census Profile 98-3	Male: 30 to 34 years	108,895	835	820	410	1610	735	485	1695	1040	285	450	200	995	600	635	285	550	125	
23	Populatio	Age characteristics	Census Profile 98-3	Male: 35 to 39 years	94,070	680	625	455	1055	735	580	1010	805	230	590	205	835	675	580	320	515	145	
24	Populatio	Age characteristics	Census Profile 98-3	Male: 40 to 44 years	86,535	760	610	420	835	815	435	560	685	310	665	170	840	740	585	305	460	220	
25	Populatio	Age characteristics	Census Profile 98-3	Male: 45 to 49 years	90,860	890	760	440	850	1010	535	500	605	390	760	225	950	835	660	265	505	285	
26	Populatio	Age characteristics	Census Profile 98-3	Male: 50 to 54 years	98,735	1160	970	515	920	1110	605	540	620	430	845	220	1010	1020	680	295	540	395	
27	Populatio	Age characteristics	Census Profile 98-3	Male: 55 to 59 years	88,145	1060	850	540	855	930	565	475	560	450	790	210	945	915	640	255	495	375	
28	Populatio	Age characteristics	Census Profile 98-3	Male: 60 to 64 years	72,270	925	710	390	755	730	385	375	555	375	670	230	765	810	465	190	365	350	
29	Populatio	Age characteristics	Census Profile 98-3	Male: 65 to 69 years	60,360	925	630	300	780	715	380	295	475	405	560	120	650	640	380	150	315	280	
30	Populatio	Age characteristics	Census Profile 98-3	Male: 70 to 74 years	42,320	590	425	205	640	570	225	285	345	315	400	125	495	420	285	85	165	205	
31	Populatio	Age characteristics	Census Profile 98-3	Male: 75 to 79 years	32,730	490	350	155	485	505	200	200	265	310	270	65	435	245	245	70	135	145	
32	Populatio	Age characteristics	Census Profile 98-3	Female: 10 to 14 years	62,165	595	475	255	355	600	330	105	350	270	820	180	765	575	730	175	300	295	
33	Populatio	Age characteristics	Census Profile 98-3	Male: 80 to 84 years	25,670	380	240	105	335	435	165	150	230	255	250	90	355	200	175	50	100	105	
34	Populatio	Age characteristics	Census Profile 98-3	Male: 85 to 89 years	15,665	210	155	65	230	340	110	80	150	165	145	25	230	125	95	30	70	80	
35	Populatio	Age characteristics	Census Profile 98-3	Male: 90 to 94 years	6,185	100	45	35	80	175	80	35	75	55	60	10	80	55	40	15	30	25	
36	Populatio	Age characteristics	Census Profile 98-3	Male: 95 to 99 years	1,280	25	20	10	25	20	20	0	10	20	15	0	10	25	5	5	5	10	
37	Populatio	Age characteristics	Census Profile 98-3	Male: 100 years and over	125	0	0	0	5	5	5	0	0	0	0	0	10	5	0	0	0	0	

The Census of Population is held across Canada every 5 years and collects data about age and sex, families and households, language, immigration and internal migration, ethnocultural diversity, Aboriginal peoples, housing, education, income, and labor. City of Toronto Neighborhood Profiles use this Census data to provide a portrait of the demographic, social and economic characteristics of the people and households in each City of Toronto neighborhood. The profiles present selected highlights from the data, but these accompanying data files provide the full data set assembled for each neighborhood. Each data point in this file is presented for the City's 140 neighborhoods, as well as for the City of Toronto as a whole. The data is sourced from a number of Census tables released by Statistics Canada. The general Census Profile is the main source table for this data, but other Census tables have also been used to provide additional information. Additional information can be found here: https://www.toronto.ca/city-government/data-research-maps/neighbourhoods-communities/neighbourhood-profiles/

Neighborhood Crime Rate (Boundary File) – API

https://data.torontopolice.on.ca/datasets/af500b5abb7240399853b35a2362d0c0_0?geometry=-80.685%2C43.542%2C-78.072%2C43.890



Toronto Neighborhoods Boundary File includes 2014-2018 Crime Data by Neighborhood. Counts are available for Assault, Auto Theft, Break and Enter, Robbery, Theft Over and Homicide. Data also includes four-year averages and crime rates per 100,000 people by neighborhood based on 2016 Census Population.

Narcity News Article on Toronto Crime Rates – Website for scraping via BeautifulSoup

https://www.narcity.com/ca/on/toronto/news/toronto-neighbourhoods-ranked-by-how-dangerous-they-are-right-now-based-on-2018-crime-rates

This article provides a breakdown of the most common crime occurrence in each neighborhood, as well as the police district the neighborhood resides. This was one of the few sources providing the neighborhoods tied to their districts.

With these three datasets, there are 140 variables (each neighborhood) and approximately 2400 rows across them.

Relationships

The common relationship running through all three data sets is the neighborhood id (name). This one to many relationship occurs on with each. With this common relationship, there is belief that patterns will emerge across each neighborhood pertaining to crime and different demographics.

Steps to Accomplish Five Milestones

In order to accomplish all five milestones, the data is going to need to be scrubbed and analyzed. The formatting does not align between the three sources, so attention will need to be made in an effort to not cause spurious relationships. Additionally, from a quick review of the neighborhood naming across the three, there appears to be a very slight inconsistency in the naming of each. While only less than five, this connection is vital and will certainly need to be resolved. Looking ahead to the fifth milestone, the visualization of this data is something I am looking forward to a great deal. The mix of statistics and geography should allow for several striking charts and plots. Overall, the project is going to take a great deal of continued focus and new skills to be completed.