Using School Immunization Rates and Venue Location Data to Choose Medical Clinic Locations

Kaif Patel

February 2, 2020

Introduction/Business Problem:

The business problem that we have is understanding how to choose the ideal location for opening a medical clinic. The audience that would be interested in solving this problem are medical professionals and medical clinic stakeholders. As stakeholders their primary interests would be due to public health concerns as well as their return on investment. We will be using the city of Toronto as an example to understand this problem and create an ideal solution. By assessing the supply and demand of medical services we can address this problem. One method that we can use to assess the demand is by looking at the immunization rates of all nearby schools. Then in order to understand the supply of medical support we will look at existing medical clinics in the local area.

Data:

The data used to solve this problem will be the Immunization Coverage for Students Data from the Toronto Open Data portal, and the Foursquare location data in order to search and count medical clinics near all Toronto schools. The immunization data has immunization rates for each individual school, as well as latitude and longitude values which we can then use to search for nearby medical clinic venues within a given radius.

Methodology:

The methods used in order to perform an exploratory data analysis involved cleaning and organizing the data into a form with more relevant information (Table 1). This involved averaging the vaccination rate across the different vaccines administered to find an approximation of vaccinated students per school. Then this the additive inverse of this number was found to find the approximate percentage of nonvaccinated students per school. This percentage was then multiplied by the enrolled population of the school to find the estimate for the total number of nonvaccinated students per school. After using the Foursquare API to find every medical clinic within 3 km of each school a frequency column was established. The number of nonvaccinated students was then divided by the frequency of medical clinics to obtain the nonvaccinated student population per medical clinic value. This value was used as a basis to gauge the supply and demand of medical clinics around schools with low vaccination rates. From this cleaned dataset (Table 2) the values can thus be explored and trends can be established.

	_id	School Name	Enrolled population	DTP coverage rate (%)	DTP Religious exemption rate (%)	MMR coverage rate (%)	MMR Religious exemption rate (%)	Lat	Lng
0	1	A Y Jackson S.S.	1027	89.0	1.1	96.5	1.1	43.805261	-79.366555
1	2	Academie Alexandre-Dumas	129	86.8	1.6	89.1	1.6	43.762419	-79.179765
2	3	Adam Beck Jr P.S.	309	96.1	3.6	95.8	3.6	43.683152	-79.288488
3	4	Africentric Alternative School	77	71.4	20.8	72.7	20.8	43.745424	-79.488261
4	5	Agincourt C.I.	1241	87.3	1.0	98.1	1.0	43.788874	-79.278910
...	...	...	...	...	...	...	...	...	...
801	802	York Memorial C.I.	876	86.8	2.5	96.3	2.5	43.690279	-79.476240
802	803	York Mills C.I.	1200	86.3	1.3	97.2	1.3	43.751529	-79.373524
803	804	Yorkview P.S.	256	93.8	2.3	93.0	2.3	43.772574	-79.435566
804	805	Yorkwoods P.S.	224	88.4	0.9	91.5	0.9	43.750660	-79.513885
805	806	Zion Heights Jr H.S.	592	94.4	1.0	96.5	1.0	43.797915	-79.371097

806 rows × 9 columns

Table 1. Original dataframe without any data cleaning

	School Name	Enrolled population	DTP coverage rate (%)	MMR coverage rate (%)	Lat	Lng	freq	Average vaccine coverage rate (%)	Average vaccine noncoverage rate (%)	Average vaccine noncoverage ratio	Nonvaccinated students	Nonvaccinated students/clinic
0	Monsignor Percy Johnson Catholic H.S.	936	84.9	94.7	43.720695	-79.572017	1	89.80	10.20	0.1020	95.47	95.47
1	Josyf Cardinal Slipyj C.S.	395	84.8	84.8	43.659475	-79.566025	1	84.80	15.20	0.1520	60.04	60.04
2	St. Basil The Great College	1258	86.2	95.1	43.727168	-79.533451	2	90.65	9.35	0.0935	117.62	58.81
3	St. Mother Teresa Catholic Academy	413	83.5	90.1	43.807243	-79.217789	1	86.80	13.20	0.1320	54.52	54.52
4	St Benedict C.S.	436	86.5	89.4	43.720695	-79.572017	1	87.95	12.05	0.1205	52.54	52.54
...	...	...	...	...	...	...	...	...	...	...	...	...
801	Monsignor Fraser College - St. Martin	16	62.5	100.0	43.667175	-79.364426	67	81.25	18.75	0.1875	3.00	0.04
802	Rosedale Jr P.S.	151	98.7	98.7	43.677656	-79.381686	53	98.70	1.30	0.0130	1.96	0.04
803	Montrose Jr P.S.	112	99.1	98.2	43.658682	-79.418844	53	98.65	1.35	0.0135	1.51	0.03
804	Kimberley Jr P.S.	160	100.0	100.0	43.682620	-79.299128	13	100.00	0.00	0.0000	0.00	0.00
805	Anson S Taylor Jr P.S.	122	100.0	100.0	43.804292	-79.261561	6	100.00	0.00	0.0000	0.00	0.00

806 rows × 12 columns

Table 2. Final dataframe after data cleaning

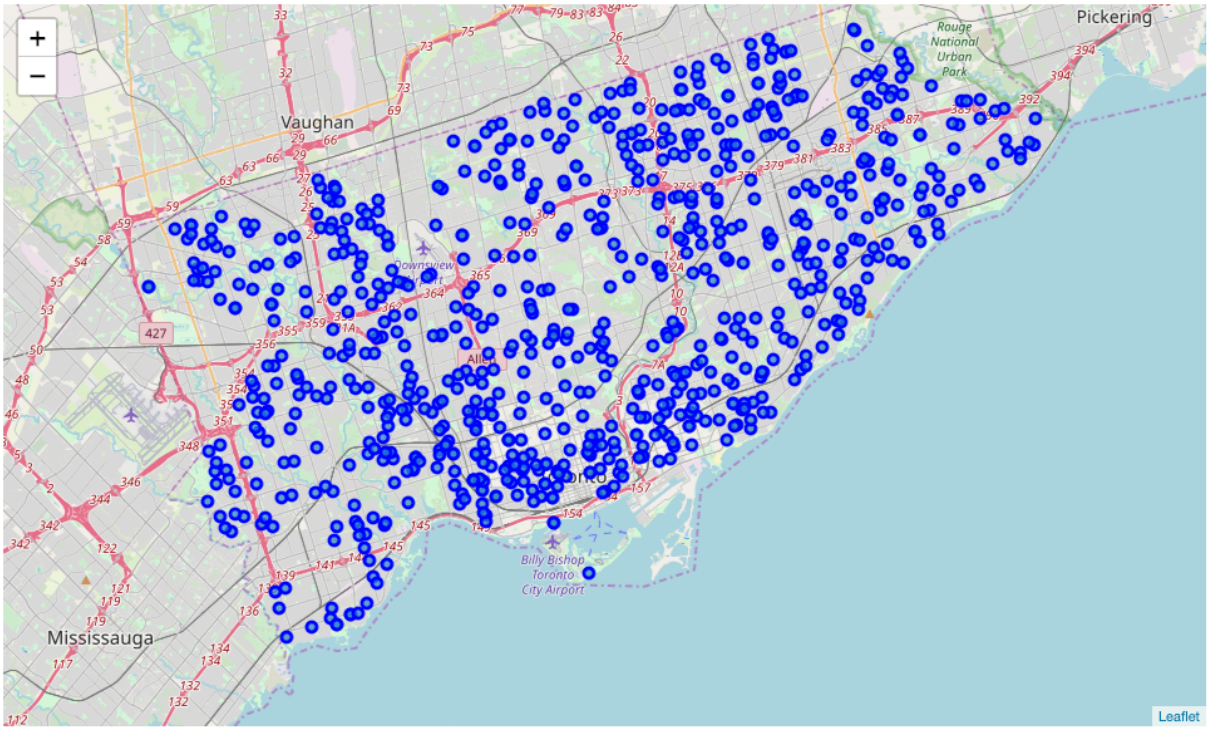

Figure 1. Map showing vaccination rate and population of every school in Toronto. The blue markers represent the schools.

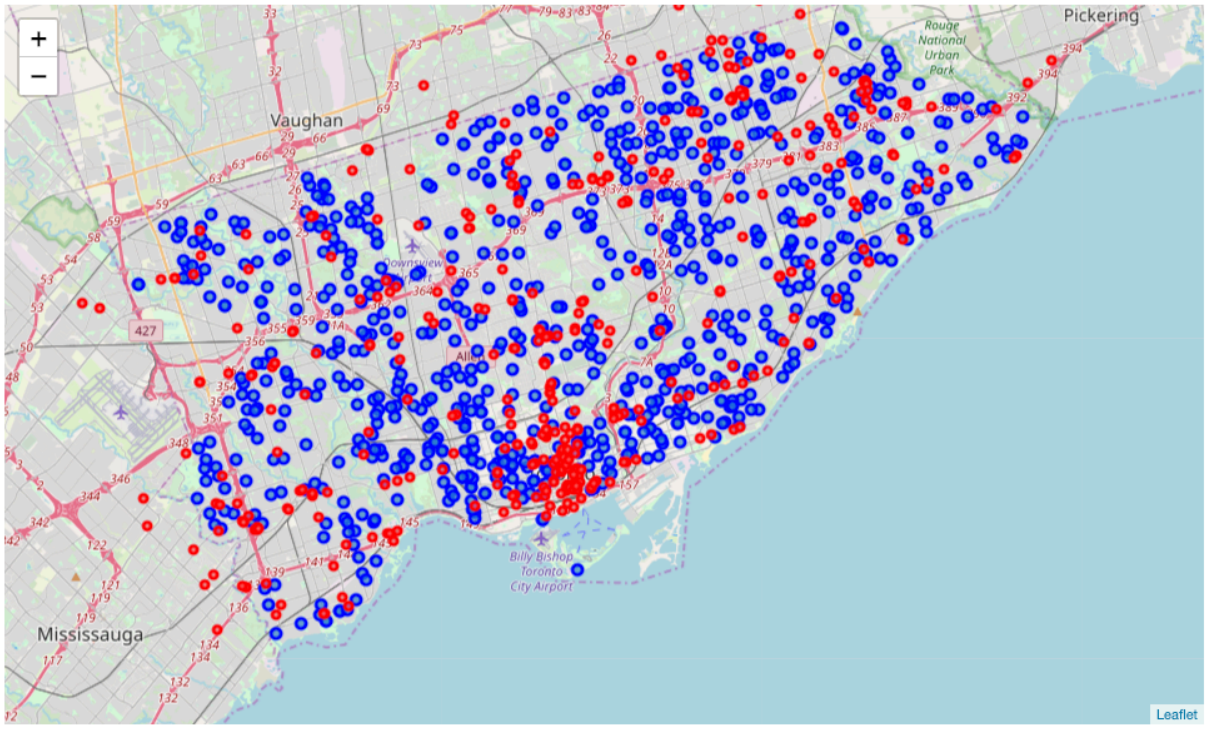

Figure 2. Map showing estimated nonvaccinated students/clinic for each school, and every medical clinic within 3 km of each school in Toronto. The blue markers represent the schools and the red markers represent the medical clinics.

	School Name	Enrolled population	DTP coverage rate (%)	MMR coverage rate (%)	Lat	Lng	freq	Average vaccine coverage rate (%)	Average vaccine noncoverage rate (%)	Average vaccine noncoverage ratio	Nonvaccinated students	Nonvaccinated students/clinic
0	Monsignor Percy Johnson Catholic H.S.	936	84.9	94.7	43.720695	-79.572017	1	89.80	10.20	0.1020	95.47	95.47
1	Josyf Cardinal Slipyj C.S.	395	84.8	84.8	43.659475	-79.566025	1	84.80	15.20	0.1520	60.04	60.04
2	St. Basil The Great College	1258	86.2	95.1	43.727168	-79.533451	2	90.65	9.35	0.0935	117.62	58.81
3	St. Mother Teresa Catholic Academy	413	83.5	90.1	43.807243	-79.217789	1	86.80	13.20	0.1320	54.52	54.52
4	St Benedict C.S.	436	86.5	89.4	43.720695	-79.572017	1	87.95	12.05	0.1205	52.54	52.54
5	St John Vianney C.S.	282	86.9	86.2	43.737976	-79.553846	1	86.55	13.45	0.1345	37.93	37.93
6	Michael Power/St Joseph H.S.	1818	84.6	95.5	43.659290	-79.581984	5	90.05	9.95	0.0995	180.89	36.18
7	Hollycrest Middle School	475	90.7	94.3	43.656104	-79.582708	1	92.50	7.50	0.0750	35.62	35.62
8	Father John Redmond Catholic S.S.	1187	83.0	94.4	43.595141	-79.516539	4	88.70	11.30	0.1130	134.13	33.53
9	Richview C.I.	1035	84.8	96.8	43.678869	-79.538912	3	90.80	9.20	0.0920	95.22	31.74
10	Weston C.I.	894	84.2	95.3	43.703816	-79.509422	3	89.75	10.25	0.1025	91.64	30.54
11	Brebeuf College School	852	81.6	90.6	43.801516	-79.403038	4	86.10	13.90	0.1390	118.43	29.61
12	St Angela C.S.	397	77.6	81.1	43.753387	-79.595791	3	79.35	20.65	0.2065	81.98	27.33
13	West Humber C.I.	1016	85.2	96.1	43.730922	-79.586705	4	90.65	9.35	0.0935	95.00	23.75
14	St Andrew C.S.	550	79.3	86.4	43.744931	-79.582382	4	82.85	17.15	0.1715	94.33	23.58
15	Father Henry Carr Catholic S.S.	850	83.8	94.5	43.735842	-79.592378	4	89.15	10.85	0.1085	92.22	23.06
16	Lester B Pearson C.I.	1291	88.0	98.3	43.803334	-79.225399	4	93.15	6.85	0.0685	88.43	22.11
17	Martingrove C.I.	1088	86.8	97.1	43.673383	-79.560740	4	91.95	8.05	0.0805	87.58	21.90
18	Sir Oliver Mowat C.I.	1048	86.5	97.1	43.779302	-79.142238	4	91.80	8.20	0.0820	85.94	21.48
19	Albert Campbell C.I.	1072	88.2	96.8	43.809014	-79.273407	4	92.50	7.50	0.0750	80.40	20.10
20	St Bernard C.S.	478	86.6	88.7	43.702506	-79.500608	3	87.65	12.35	0.1235	59.03	19.68

Table 3. Subset of cleaned dataframe displaying the top schools in terms of nonvaccinated students per clinic

785	Lord Lansdowne Jr P.S.	174	95.4	96.0	43.658953	-79.401910	69	95.70	4.30	0.0430	7.48	0.11
786	Niagara Street Jr P.S.	125	92.0	96.8	43.643937	-79.408253	68	94.40	5.60	0.0560	7.00	0.10
787	Monsignor Fraser College - Midland	16	87.5	100.0	43.801069	-79.285717	10	93.75	6.25	0.0625	1.00	0.10
788	Contact Alternative School	35	68.6	94.3	43.653766	-79.390394	72	81.45	18.55	0.1855	6.49	0.09
789	Cottingham Jr P.S.	100	96.0	96.0	43.680379	-79.395609	44	96.00	4.00	0.0400	4.00	0.09
790	Monsignor Fraser College - Isabella	14	42.9	71.4	43.669799	-79.377153	67	57.15	42.85	0.4285	6.00	0.09
791	Seed Alternative	31	80.6	90.3	43.662078	-79.348686	48	85.45	14.55	0.1455	4.51	0.09
792	Downtown Alternative School	66	92.4	90.9	43.647736	-79.370416	66	91.65	8.35	0.0835	5.51	0.08
793	Ogden Jr P.S.	116	96.6	94.8	43.650014	-79.395127	76	95.70	4.30	0.0430	4.99	0.07
794	School of Life Experience	28	92.9	96.4	43.679281	-79.322772	20	94.65	5.35	0.0535	1.50	0.07
795	Downtown Vocal Music Academy of Toronto	47	87.2	89.4	43.651221	-79.402637	76	88.30	11.70	0.1170	5.50	0.07
796	Alternative Scarborough Education 1	15	86.7	100.0	43.766645	-79.254796	15	93.35	6.65	0.0665	1.00	0.07
797	Subway Academy II	19	78.9	78.9	43.655765	-79.395705	72	78.90	21.10	0.2110	4.01	0.06
798	Victoria Park Elementary	92	98.9	100.0	43.712689	-79.298151	11	99.45	0.55	0.0055	0.51	0.05
799	Bennington Heights Elementary	129	99.2	98.4	43.693882	-79.369553	32	98.80	1.20	0.0120	1.55	0.05
800	Native Learning Centre	15	66.7	86.7	43.663608	-79.379178	67	76.70	23.30	0.2330	3.49	0.05
801	Monsignor Fraser College - St. Martin	16	62.5	100.0	43.667175	-79.364426	67	81.25	18.75	0.1875	3.00	0.04
802	Rosedale Jr P.S.	151	98.7	98.7	43.677656	-79.381686	53	98.70	1.30	0.0130	1.96	0.04
803	Montrose Jr P.S.	112	99.1	98.2	43.658682	-79.418844	53	98.65	1.35	0.0135	1.51	0.03
804	Kimberley Jr P.S.	160	100.0	100.0	43.682620	-79.299128	13	100.00	0.00	0.0000	0.00	0.00
805	Anson S Taylor Jr P.S.	122	100.0	100.0	43.804292	-79.261561	6	100.00	0.00	0.0000	0.00	0.00

Table 4. Subset of cleaned dataframe displaying the bottom schools in terms of nonvaccinated students per clinic

Results:

From the cleaned dataframe we can see the top and bottom schools in terms of the population of nonvaccinated students/clinic. It should be noted that the top schools tend to be high population schools away from the city center, many of which are high schools and catholic schools. The bottom schools tend to be low population schools closer to the city center, many of which are alternative or elementary public schools.

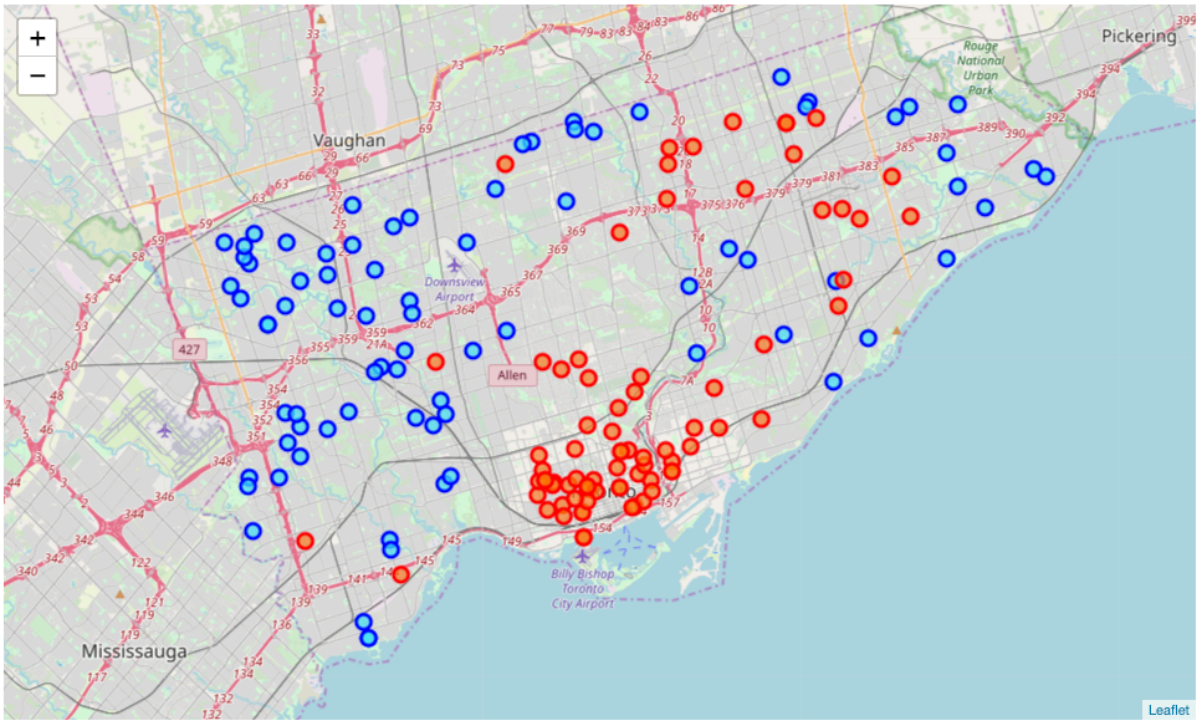

Figure 3. Map representing the population of nonvaccinated students and nonvaccinated students/clinic for each school in Toronto. The blue markers represent the top 10% of schools in Toronto in terms of highest nonvaccinated student/clinic rate, and the red markers represent the bottom 10% of schools. From an investor standpoint: the blue markers would be ideal locations to open a medical clinic, and the red markers would be non-ideal locations with heavy competition.

Discussion:

A few important observations is that the top schools tend to be in the lower income areas of Toronto, some of which have high immigrant populations. This doesn't suggest that immigrants are less likely to get vaccinated, as most of the elementary schools within those areas have reasonably high vaccination rates. This may be due to these schools not having very strict vaccination requirements, and since these high schools have an overall older population, many of the students may have the autonomy to choose whether or not to be vaccinated, but without parental oversight and maturity they may not be considering the vaccines as totally necessary. Furthermore, 8/10 of the top schools in terms of nonvaccinated students/clinic are catholic schools. This suggests that these populations may be more prone to opting out of vaccinations because of religious beliefs, or perhaps due to the less stringent vaccination policies of these schools. The bottom schools in terms of nonvaccinated students/clinic tend to be alternative schools and elementary public schools. The reason for this may be due to many alternative school students possibly being older and more mature, and so they have a greater awareness for the necessity of vaccinations. The elementary public schools on the other hand also have low nonvaccinated students/clinic values, possibly due to greater parental oversight.

My recommendation for investors looking to open a medical clinic would thus be to focus on areas outside of the city center, with a relatively lower income demographic. The west end of Toronto seems to be a particularly ideal location to open a medical clinic.

Conclusion:

From this analysis we can see the top areas of interest to a possible investor looking to open a medical clinic in the city of Toronto. A noteworthy argument against this analysis however is that this analysis makes the assumption that there is a strong correlation between vaccination rates and disease prevalence in the local area. In reality there may only be occasional outbreaks. Furthermore, the Foursquare venue data is not entirely accurate and may be making our data dirty to some degree. For example, many Goodlife Fitness gyms have been miscategorized as medical clinics, which have been removed, but other venues may also be miscategorized. The results may also be skewed because of a really high number of medical clinics per school in the downtown area.

That being said, this analysis does discover some interesting trends. It can be noted that the worst schools in terms of high numbers of nonvaccinated students/clinic tend to be high population schools in some of the lower income areas of the city. Furthermore the data also shows that the schools with the lowest rates of nonvaccinated students/clinic tend to be low population schools, possibly alternative and/or private schools with stringent vaccination requirements. This is supported by the fact that many of these low population schools are in downtown, which should normally have higher populations. This data thus suggests a possible correlation between vaccination rates and income.