

Finding the perfect spot for a new store of school and office supplies

Friedrich Keppeler

Applied Data Science Capstone Project –
IBM Data Science Professional Certificate

Problem

- The owner of a large school and office supplies' chain (WD's Office & School Supplies) is interested in open a new branch in Toronto, Canada. His business plan is already well-established, and the address of new branches is defined according to four main points:
 - 1) Stores are always in the vicinity of elementary- and high-schools (their main clients are young students and their parents);
 - 2) Branches are locations with high average family income (i.e., larger than the population average)
 - 3) Stores are in neighborhoods with a high proportion of young people (10 to 50 years old);
 - 4) The density of people in the neighborhood where the new branch will be open has to be larger than the average of the city.

Goal

- To identify the best location in Toronto to open the new branch based on the directions provided by the owner.

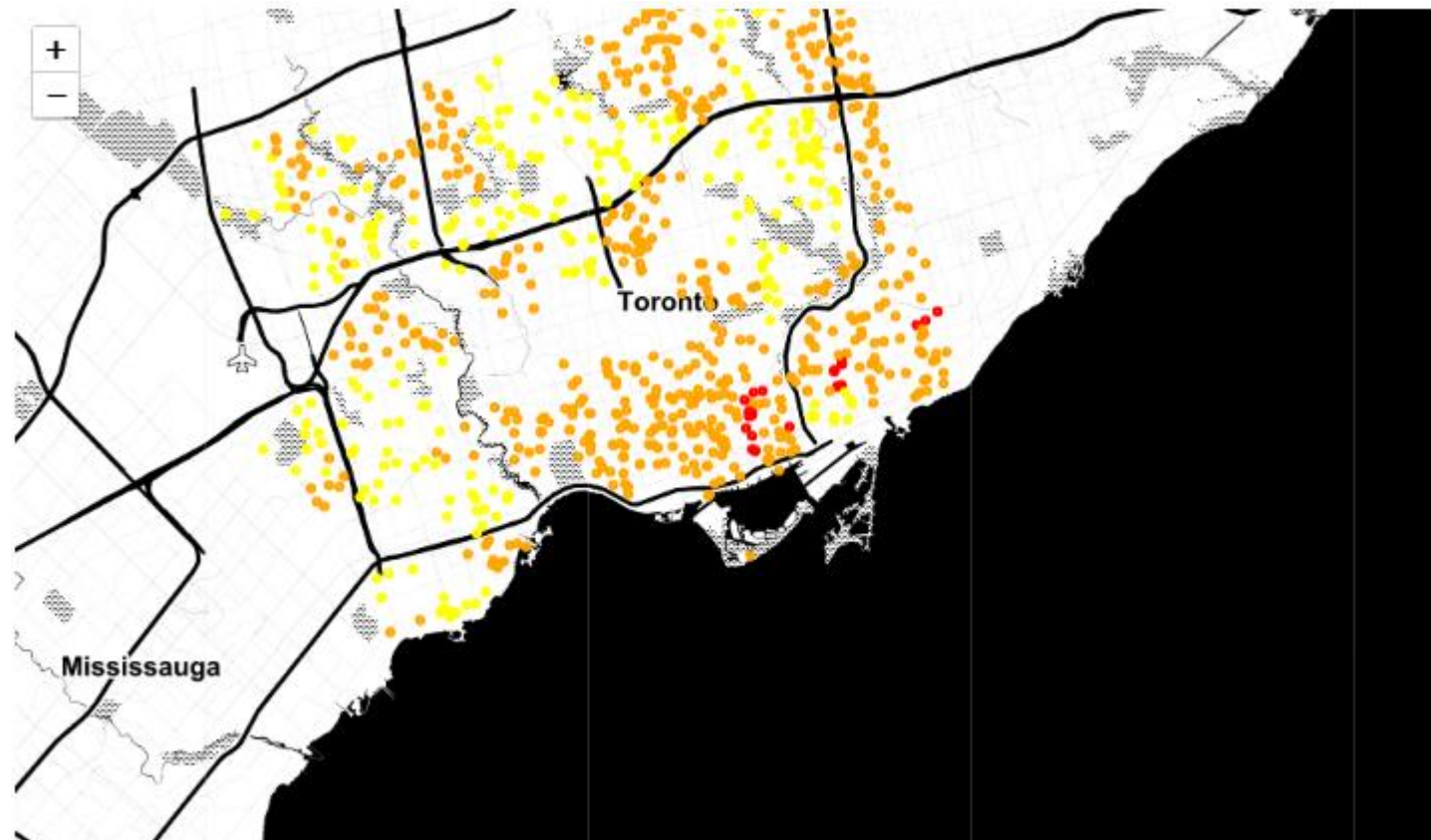


Approach

- Rank school locations according to average family income, percentage of age groups, and population density of the surrounding neighborhoods.
- Plot the location of each school location in a map and colored them according to the variables of interest (i.e., average family income, percentage of age groups, population density) and the overall ranking score.
- Explore the correlation between average family income, percentage of age groups, and population density to have a better understanding of the data.

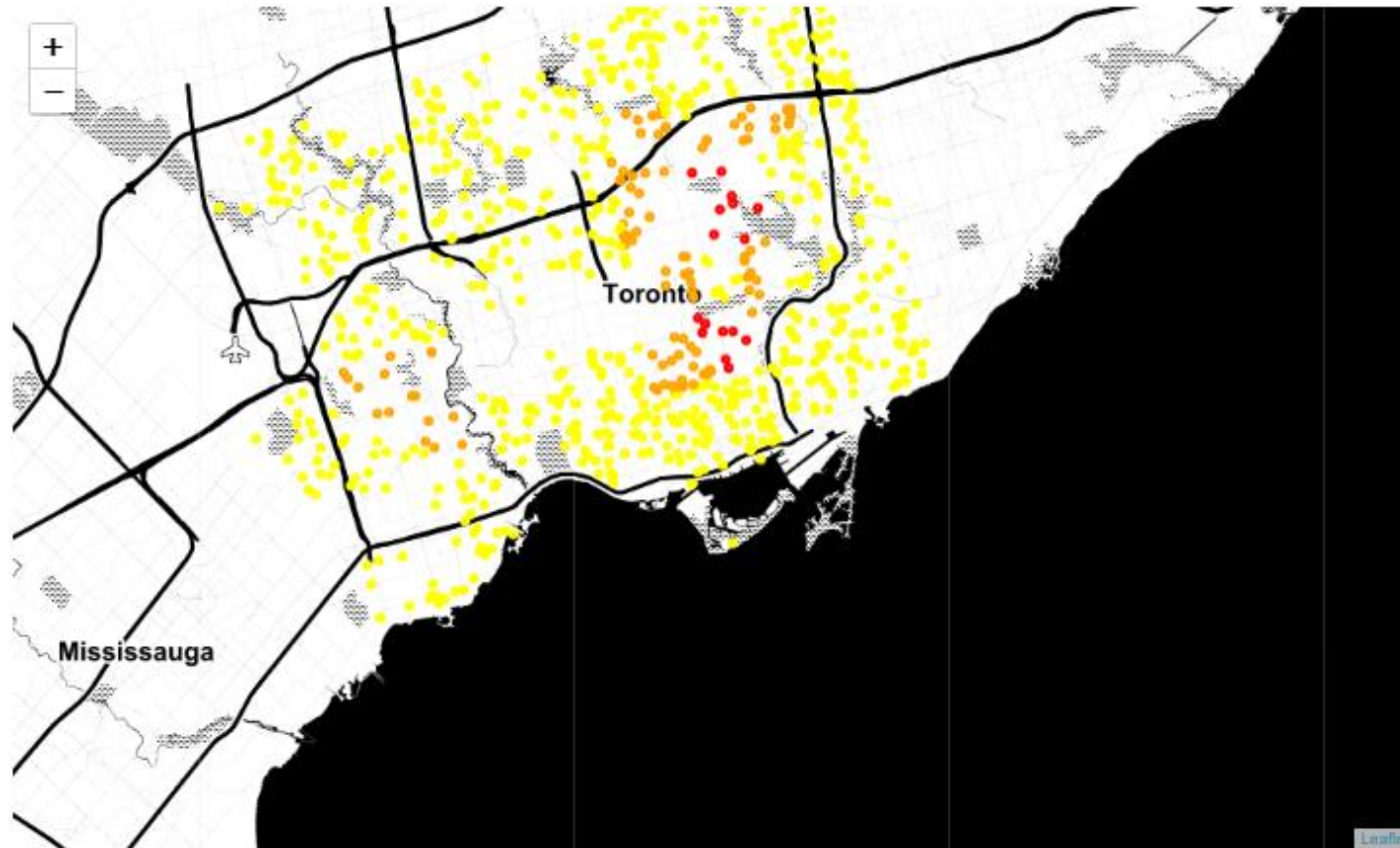
Results

- Population density in the surroundings of each school. Red points indicate high densities, orange points are associated with intermediate densities, and yellow points represent low population densities



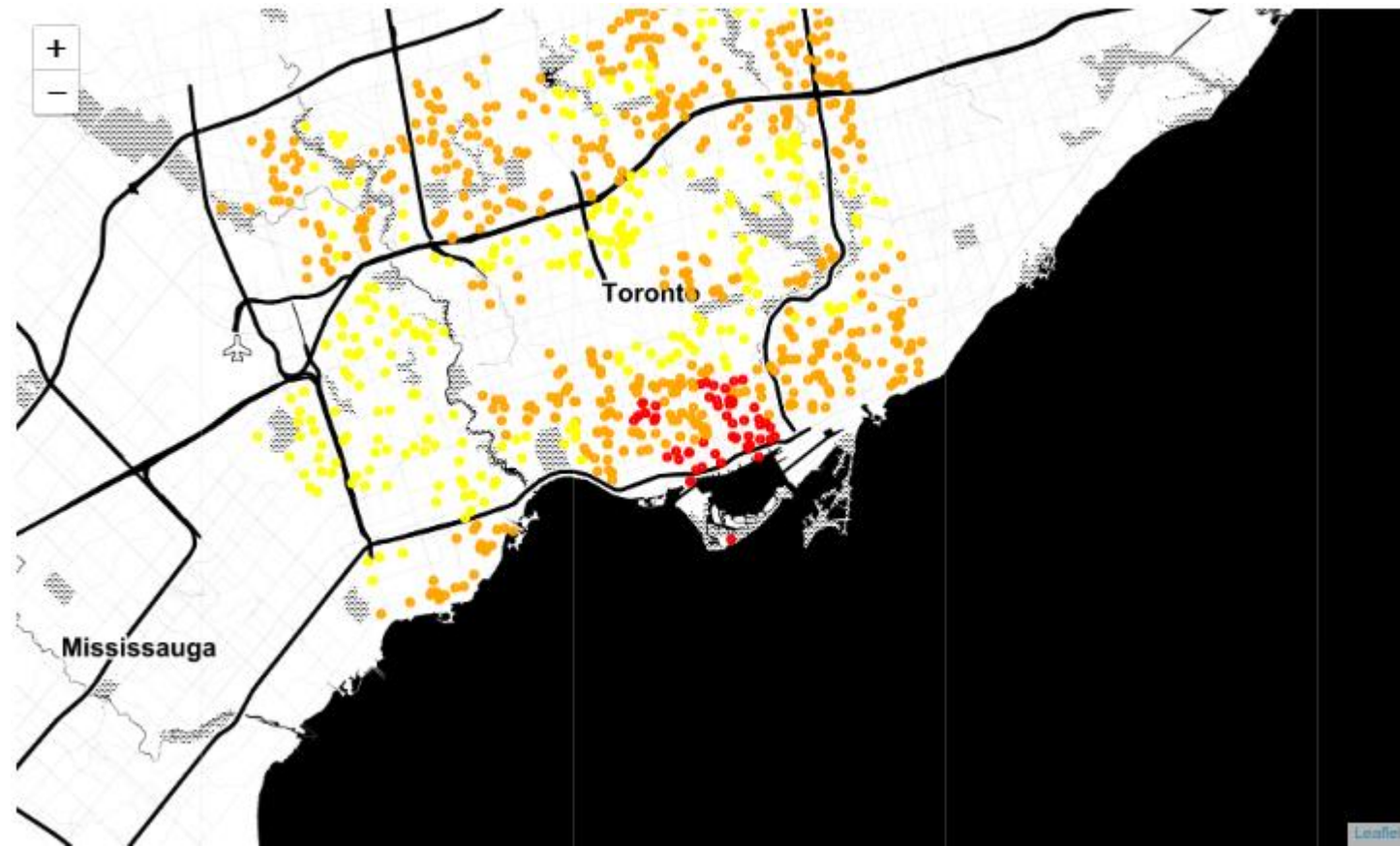
Results

- Average family income of the neighborhoods wherein the schools are located. Red points indicate high family incomes, orange points are associated with intermediate family incomes, and yellow points represent low family incomes.



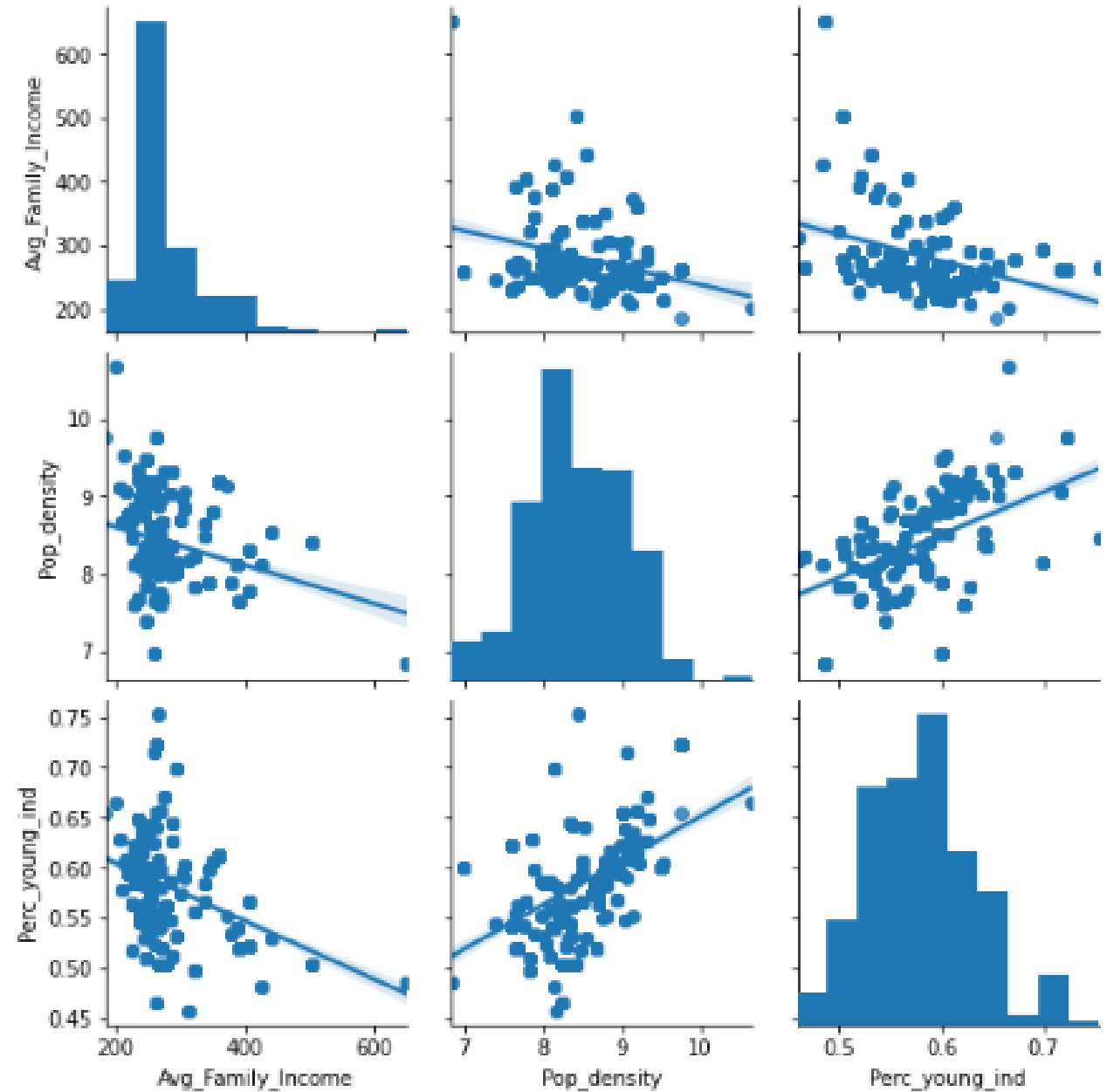
Results

- Percentage of young individuals in the neighborhoods wherein the schools are located. Red points indicate high percentages, orange points are associated with intermediate percentages, and yellow points represent low percentages.



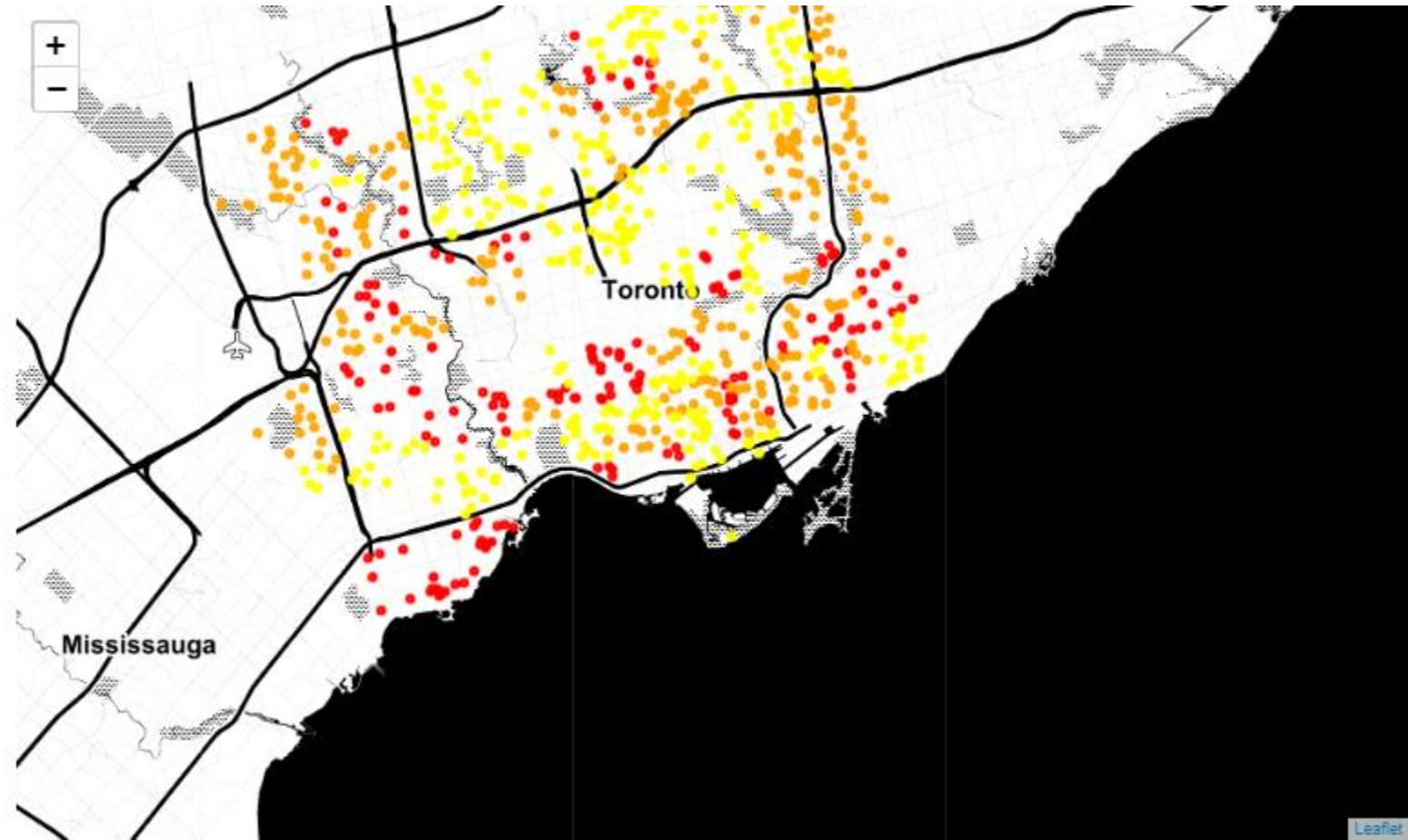
Results

- Correlations patterns between average family income, population density, and percentage of young adults.



Results

- The best locations are the ones with low ranking scores. Red points indicate regions with low ranking scores, orange points represent intermediate ranking scores, and yellow points are locations with high ranking scores. In this cases, the lower the ranking the better the location.



Results

- Top locations according to my rank based on average family income, population density and percentage of young adults.

NAME	SCHOOL_LEVEL	SCHOOL_TYPE	BOARD_NAME	SOURCE_ADDRESS	SCHOOL_TYPE_DESC	rank__Avg_Family_Income	rank_Pop_density	rank_Perc_young_ind	Overall_rank
St Simon Catholic Elementary School	NaN	ES	Toronto Catholic District School Board	24 Strathburn Blvd	English Separate	1.0	1.0	1.0	3.0
Suon Academy	NaN	PR	NaN	5 Queenslea Ave	Priv	1.0	1.0	1.0	3.0
St Basil The Great College School	NaN	ES	Toronto Catholic District School Board	20 Starview Lane	English Separate	1.0	1.0	1.0	3.0
Pelmo Park Public School	NaN	EP	Toronto District School Board	180 Gary Dr	English Public	1.0	1.0	1.0	3.0

Main conclusions

- Average income was negatively related to population density and the percentage of young individuals. On the other hand, the population density was positively correlated with the percentage of young individuals.
- Locations near St Simon Catholic Elementary School, Suon Academy St Basil, The Great College School, and Pelmo Park Public School had the lowest ranking scores and are the best options to open the new branch of WD's Office & School Supplies in Toronto.