Data Analysis: Urban Area and Air Pollution

Authors: Arian Mohamad-Hosaini, Daniel Yu
Grade 12, Earl Haig Secondary School

Abstract

This report is a research and analysis big data project which investigates Toronto's environmental condition and air pollution. Based on the data set of the amount of air pollution and related information within 140 neighborhoods in Toronto, we used Python and internet resources to investigate the correlation between these data. The purpose of this project is to collect data on the pollutant score from each non-industrial area in Toronto, and arrange and organize it to analyze factors related to environmental protection. This analysis was done by comparing the area of green space and traffic conditions with relation to pollutants in a neighborhood and compare data between neighborhoods. The significance of this project is that big data may help come up with solutions to environmental issues such as pollution, climate change, and waste, as big data may have a bigger role with resource management or environment monitoring in the future.

Keywords

Air pollution, Environment, Big Data

Introduction

There is no doubt that people living with environmental and green services are living healthier lives. The cleaner the air we breathe in, the healthier we are. However pollution may not be completely avoidable to push forward human development. With burning fuels producing air pollution, and waste chemicals producing water pollution, can people really have the sense of being able to minimize or control their pollutant production? Scientists and researchers who are concerned about the environment study the methods of controlling. As scientists, data analysts and the government are monitoring the city's environment through collected data every year, the data on pollution has become particularly essential.

When we received an Excel file of the environmental data of Toronto, we focused our ideas on transportation air pollution since it contains more relevant data. We decided to look for any possible relationship between major causes of air pollution, to major protecting procedures to amount of air pollution in the city. Of course, factories create a large amount of pollution. However we are trying to discover causes of pollution that are more subtle, and the emission of carbon monoxide (CO) and nitrogen monoxide (NO) from automobiles generally increase the amount of air pollution[1].

Major sources of air pollution in Toronto are the combustion of fossil fuel, gasoline, petroleum and industrial energy use. Different regions of Toronto have different sources of pollution. For example, the center of Toronto has a higher population than other areas, so the area may contain a high amount of air pollution from transportation. Therefore we will consider arranging our data carefully. There are other reasons of air pollution in Toronto, such as fire or polluting particles blown by wind, but in this case, changes of data will be small and will be considered insignificant for our research.

The tool we are using is Python because we have experience using it in Grade 11 and we thought it has more freedom to manipulate data. With the project, we are also looking forward to develop our skills as researchers and innovators, gain experience working with big data, and get a deeper insight into the important usage of big data by solving problems.

Method

We used more than one dataset from Wellbeing Toronto, which provided information of population, environment, and transportation status for every 140 neighborhoods in Toronto. For each category, we created a list, and compiled all data from the Excel file into each list. This ensures easy access of information. For every item in the list, its index number corresponds to the same index number in another list, which represents data from the same neighborhood.

In order to interpret the correlation between pollutants released to the air and other data, we decided to create many scatter plots, without considering the order of neighborhoods IDs they appear in the list. Since neighborhoods have different areas, we used population density (calculated by population/area(km^2)) to describe different parts of the city.

To properly find transportation's contribution to air pollution, neighborhoods have been segregated. Neighborhoods with more than one industrial facility is considered a have great influence on the measure of transportation pollution. We will be calling these industrial zones. Neighborhoods that are located in urban areas (which are densely populated) will be used to measure transportation pollution. In addition, two of the neighborhoods were removed from the lists, because they were outliers. Their population density was too large to fit on the graph and they were way beyond the average population density in Toronto.

By sorting and processing data, we did some comparisons and created a linear regression for the graph that has a visible trend. To investigate the amount of pollutant change in relation to population density, tree covered area, green space, and road volume, one of each graph was created.

Discussion

The amount of pollutants was measured by ChemTRAC adding 25 different hazardous chemical exposed to air. Looking at Figure 1, the plot shows the amount of pollutant change in relation to the increasing car density. The trend of pollution didn't go expectedly upward as the road volume increases. Instead there is a cluster in road volume from 2000 to 4000, which plots like a normal distribution. There are more neighborhoods within the range of 2000 to 4000 than other ranges. So this increases the likelihood of having a high amount of pollution. Other factors have influenced this plot greatly, since the plot has an unobvious increasing trend that seems to fall off. This can be explained by human action. Areas with greater car densities may be considered to be more 'needed to fix', so more trees are planted in the area that help reduce the amount of pollution. In Figure 2, the graph shows the amount of pollutant in relation to population density; however this graph is very similar to Figure 1. The trend of the plot is similar, and the increasing population density may also result in a greater release of pollutants. It seems that more people take care of their environment when the population densifies.

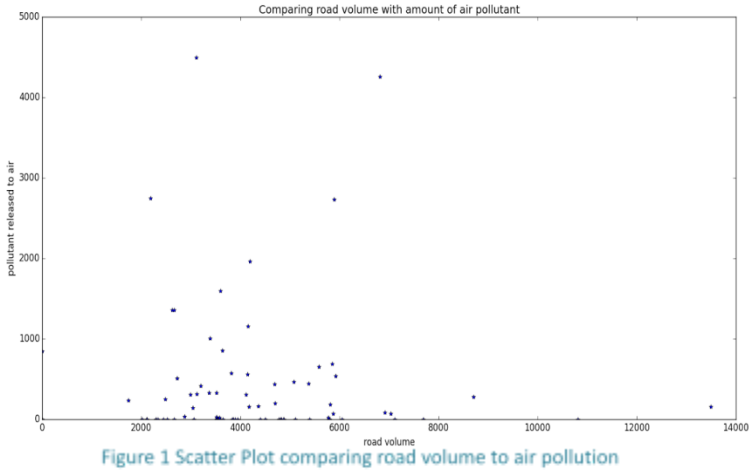

Figure 1 Scatter Plot comparing road volume to air pollution

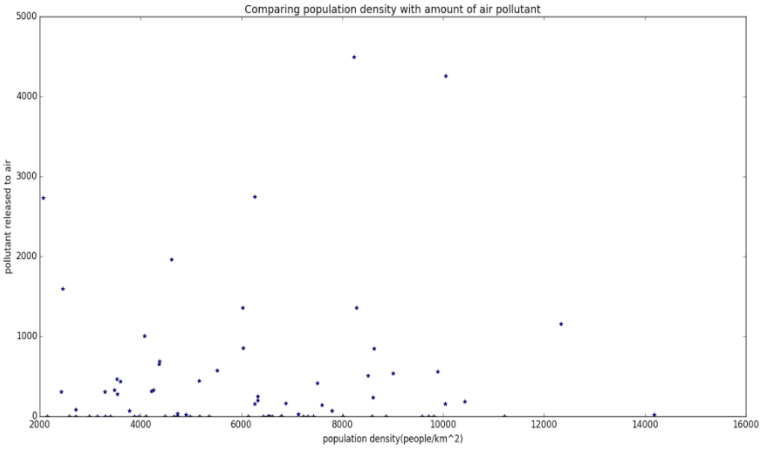

Figure 2 Scatter Plot comparing population densities to air pollution

The best natural air purified comes from trees. Figure 3 compares the amount of pollutant change with amount of area covered by tree and area of green space. The change follows our expectancy this time. The more trees and green spaces there are, the less pollution. Also notice the significant changes due to a remarkable drop by building more green spaces. It seems that the lack of green spaces and trees is one factor that contributes to the problem of air pollution.

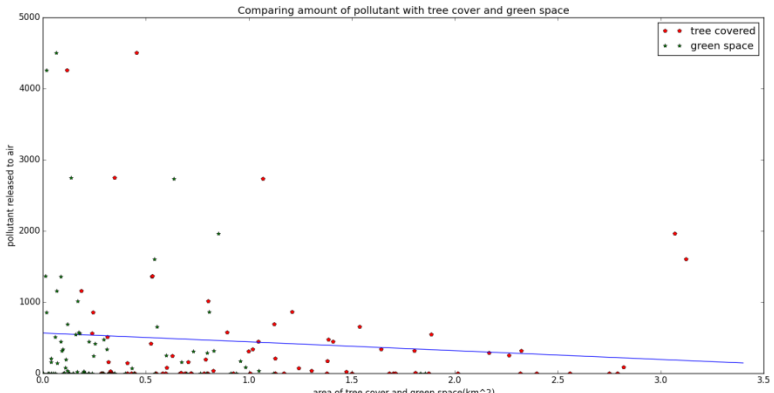

Figure 3 Scatter Plot comparing green space and tree covered area to air pollution

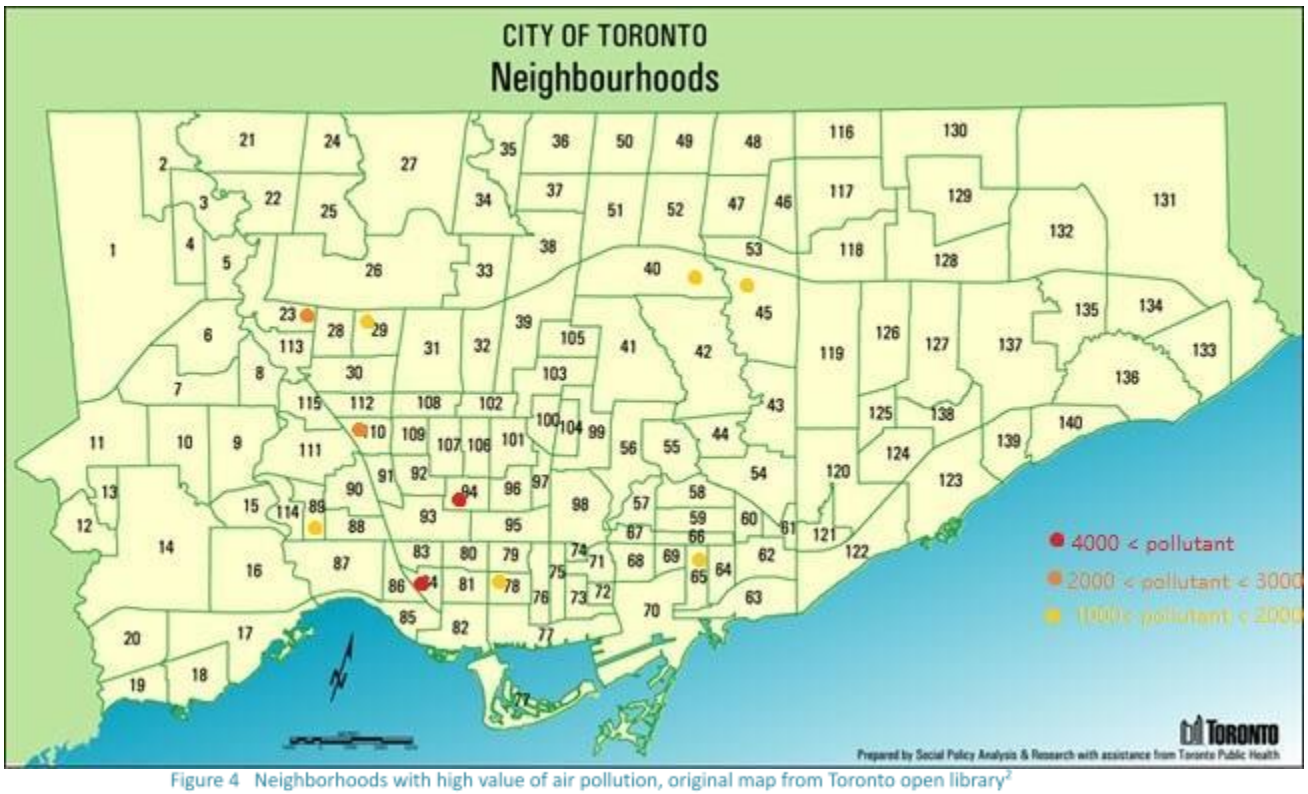

Figure 4 Neighborhoods with high value of air pollution, original map from Toronto open library[2]

In Figure 4, the map shows the neighborhoods of Toronto labeled with their ID numbers: yellow, orange and red. Each colour specifies the amount of pollution with ranges of 1000 to 2000, 2000 to 4000, and greater than 4000. Neighbourhoods with similar amounts of pollutants are quite close to each other, especially neighbourhoods with red and orange dots. Neighborhoods with yellow markers are not as close to each other, although they surround the center of Toronto.

Conclusion

We have proven that the amount of pollution decreases as the amount of green spaces and trees increase. Through our discovery, the solution to air pollution was building green spaces and planting trees to purify the air. The method of reducing air pollution appears to be nature itself, but human ingenuity can also improve eco-friendly technology. Energy-efficient cars and artificial gas filters are just some examples of eco-friendly technology that have the potential to economize resources like fuel and gasoline.

Despite how our future goes, our research indicates that areas closer to each other are more likely to get polluted together. Reasons to why urban area's air is cleaner are due to many variables, but based on the dataset we are given, we were able to track city of Toronto's health, and concluded how environment is being affected.

Bibliography

1."Cars, Trucks, and Air Pollution." *Union of Concerned Scientist.* N.p., n.d. Web. 19 Dec. 2015. <http://www.ucsusa.org/clean-vehicles/vehicles-air-pollution-and-human-health/cars-trucks-air-pollution#.VnaFSPlViko>

2."Toronto Neighborhoods List." N.p., n.d. Web. 20 Dec. 2015. <http://www1.toronto.ca/wps/portal/contentonly?vgnextoid=100c861b9fdb1410VgnVCM10000071d60f89RCRD>