IBM DATA SCIENCE PROFESIONAL CERTIFICATION

CAPSTONE PROJECT

ANALYZING POLLUTION IN LAHORE

BY

SAAD SHAKIL

Contents

1. Introduction

My idea is less of business problem and more of social issue. The issue concerns my current city Lahore which is second largest city of Pakistan and 24th largest city in world according to population. The current population stands around 11 million people in the metropolitan. Due to high density of city, it is face with sewer health problems. One of the major health issues is low quality of air due to excessive air pollution. According to IQAir AirVisual's 2018 World Air Quality Report, Lahore ranks at 10th most polluted city in the world with an average air quality index of 114.9. As a result of high pollution, citizen face many health problems. During the winter season, level of smog is so intense that visibility almost drops to zero. For those who are not familiar with the term smog; it is a fog made heavier and darker by smoke and chemical fumes. It is a mixture of smoke, gases, and chemicals that makes the atmosphere difficult to breathe and harmful for health.

Problem

For this project I will analyze the population density of Lahore neighborhoods. I will also incorporate fuel station in my analysis, which will give us better idea of areas where traffic flow is high thus resulting in pollution. It is assumed that in areas where there is fuel station, traffic flow will be high in those areas. Then I will compare population density and fuel station locations against forestation in the city, which will identify which areas of city need more forestation to reduce impact of pollution.

Stakeholders

The concerned stakeholders for this analysis would be general public, as they need to be aware for of need for forestation, a task to which every individual can easily contribute. Moreover, most important stakeholders would be government as it is responsible for well being of citizen. Moreover, fuel station companies are also targeted stakeholders, as they need to take responsibility for harmful effects their products are causing to general public.

2. Data

Various medium was used to extract the data. Wikipedia was used to get the list of towns in Lahore city. Official population figures were extracted from Pakistan Bureau of Statistics department. Population data at town level is not publicly available, therefore population statistics at district level was downloaded. Coordinates were extracted using Python's libraries such as Geocoder. Foursquare was used to extracted venues list data.

However, during the course of this project many challenges were faced to get hold of data. No Shapefile was available publicly which could plot Lahore city at Town/District level. Therefore, while plotting map of Lahore district/town level boundaries could were not generated in maps.

3. Methodology

The idea behind this project was to highlight areas of Lahore according to population density and forestation. By doing this we could identify which areas need more plantation so that impact of pollution can be reduced.

Therefore, different types of maps of Lahore city were generated to highlight the area of interests.

4. Results and Discussion

Lahore is world 24th largest city according to the population, spanning at area of 1,772 sq. km. The city is divided into 5 districts, which is further divided into 10 towns.
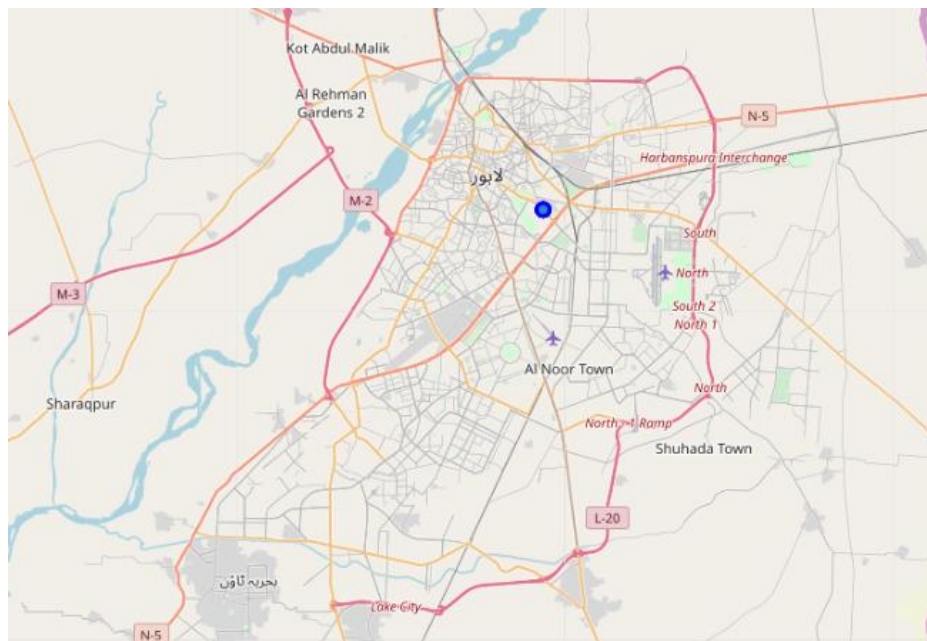
	Town
0	Ravi
1	Shalamar
2	Wagha
3	Aziz Bhatti
4	Data Gunj Buksh
5	Gulberg
6	Samanabad
7	Iqbal
8	Nishtar
9	Cantonment

	Lahore district
0	LAHORE CANTT
1	LAHORE CITY
2	MODEL TOWN
3	RAIWIND
4	SHALIMAR
5	Total

Due to limitations of unavailability of data population of Lahore at town level could not be extracted. Therefore, towns were categorized under districts for purpose of population distribution.

	Lahore district	Population	Towns
0	LAHORE CANTT	1636342	Cantonment, Aziz Bhatti
1	LAHORE CITY	3655774	Iqbal, Gulberg, Samanabad
2	MODEL TOWN	2698235	Nishtar
3	RAIWIND	855626	Raiwand
4	SHALIMAR	2280308	Data Gunj Buksh, Shalamar, Wagha

As per 2017 census, total population of Lahore stands around 11 million people.
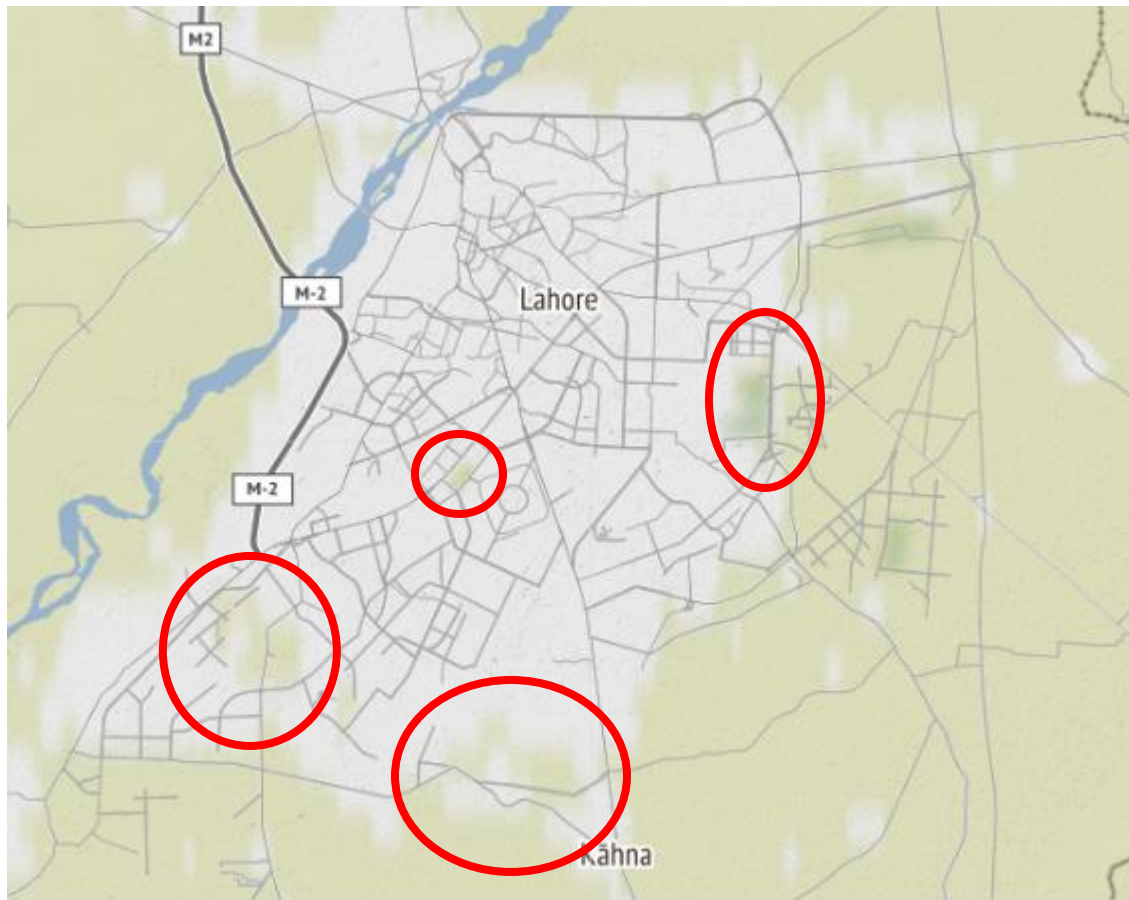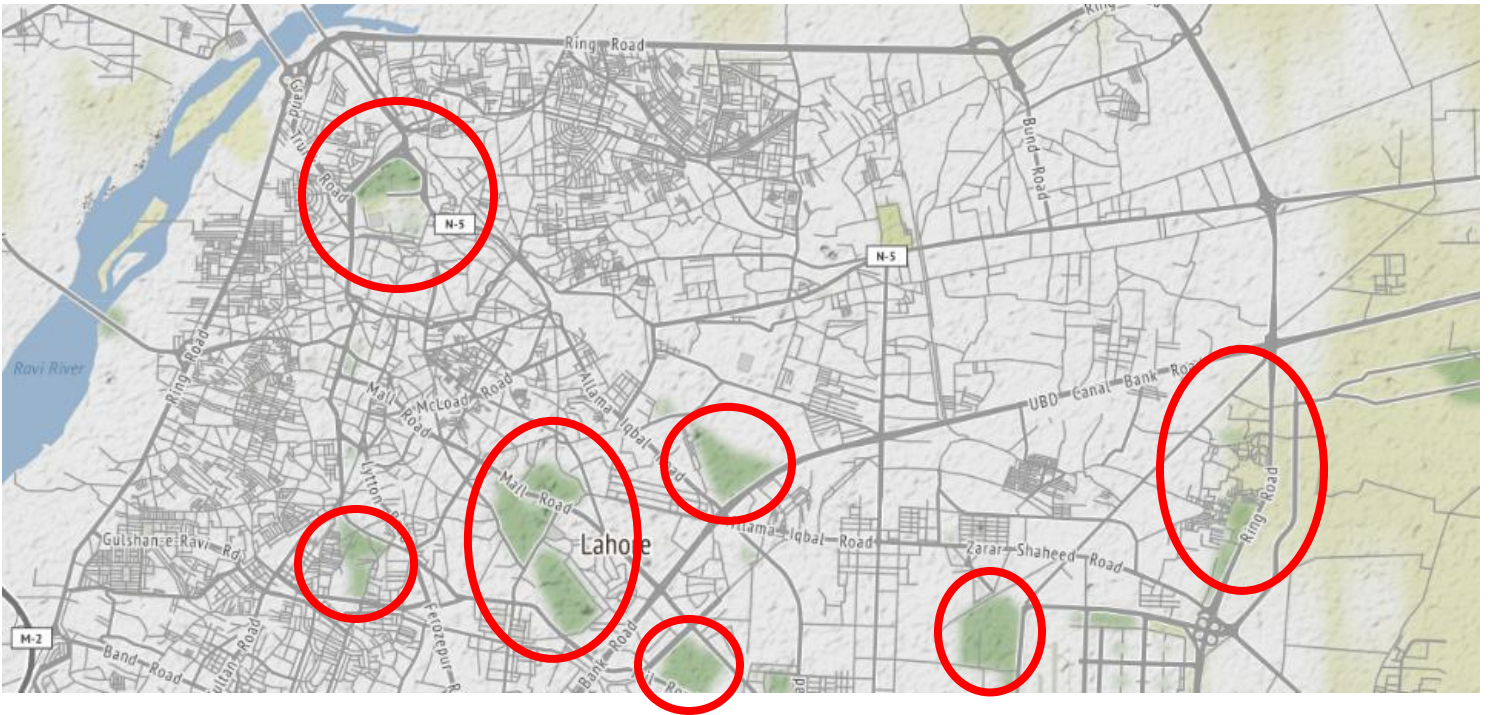
Map of Lahore:

Due to limitation of data, such at Shapefile of Lahore city, boundaries of city could not drawn at district/town level. In order to draw boundaries at town/district level detailed data of coordinates is needed, which unfortunately was not available despite extensive search.

For our analysis, we needed to highlight 'Green' areas of city which would identify level of forestation in the city. For this purpose, map of Lahore city was generated in 'Stamen Terrain' to get the desired results.

The circled areas marked below display the forest/green areas as generated by Folium library maps.

As you can see, level of forestation/greenery inside the city is little to none. There are very few parks in the city as show in above pictures. Most of the greenery is found at east and south side of city.
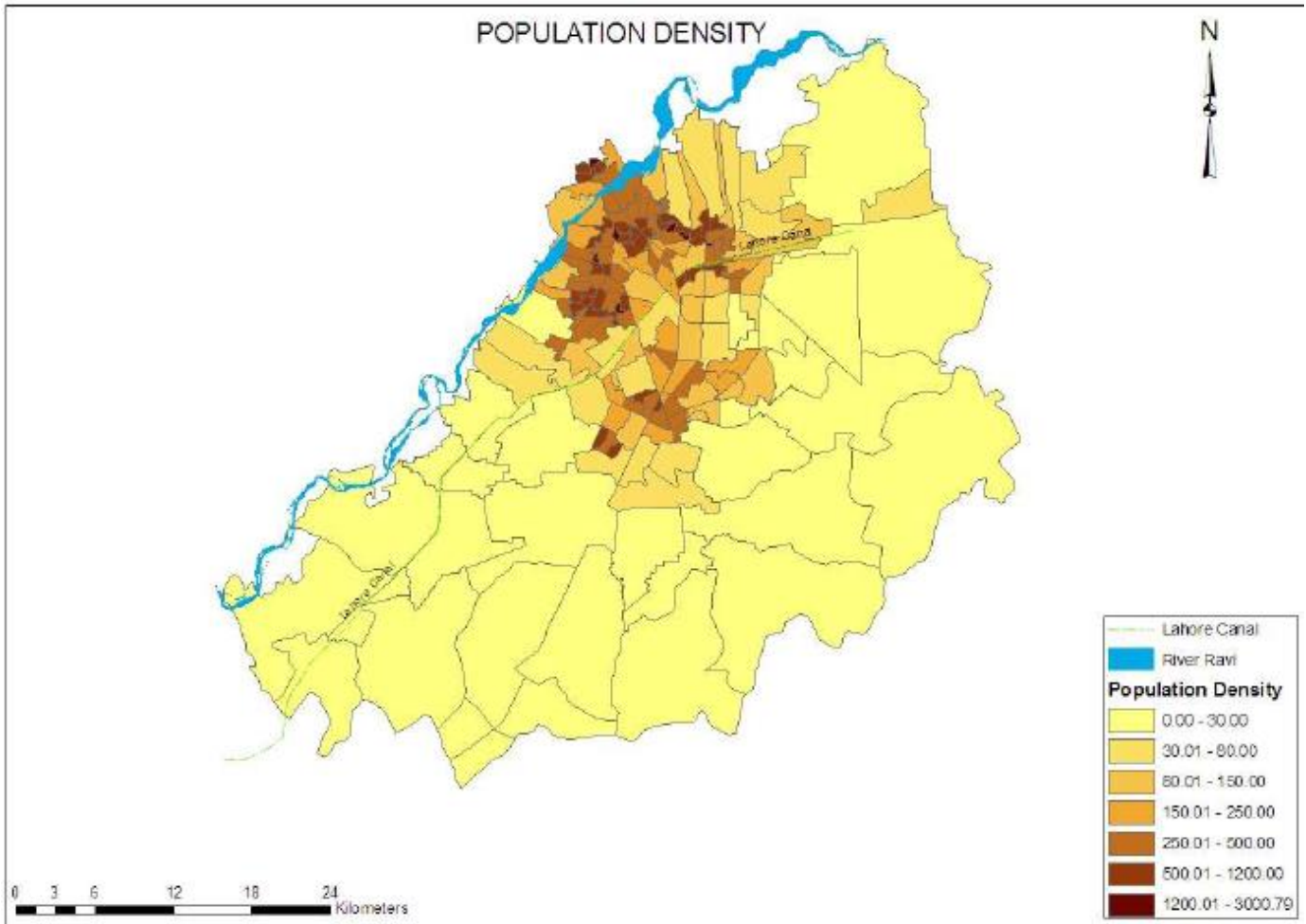
Looking at the population density of Lahore, northern part of city is highly dense.

Comparing the population density map and Greenery map, we can clearly see that at high population density areas there is no greenery at all.

Analyzing further, locations of fuel stations in the city was extracted from Foursquare. The results are depicted the map below. Purpose of incorporating gas stations in my study was because presence of a gas station is an indication of traffic flows. Gas stations are usually located in areas where traffics flow is high. Higher traffic flow indicated higher pollution!
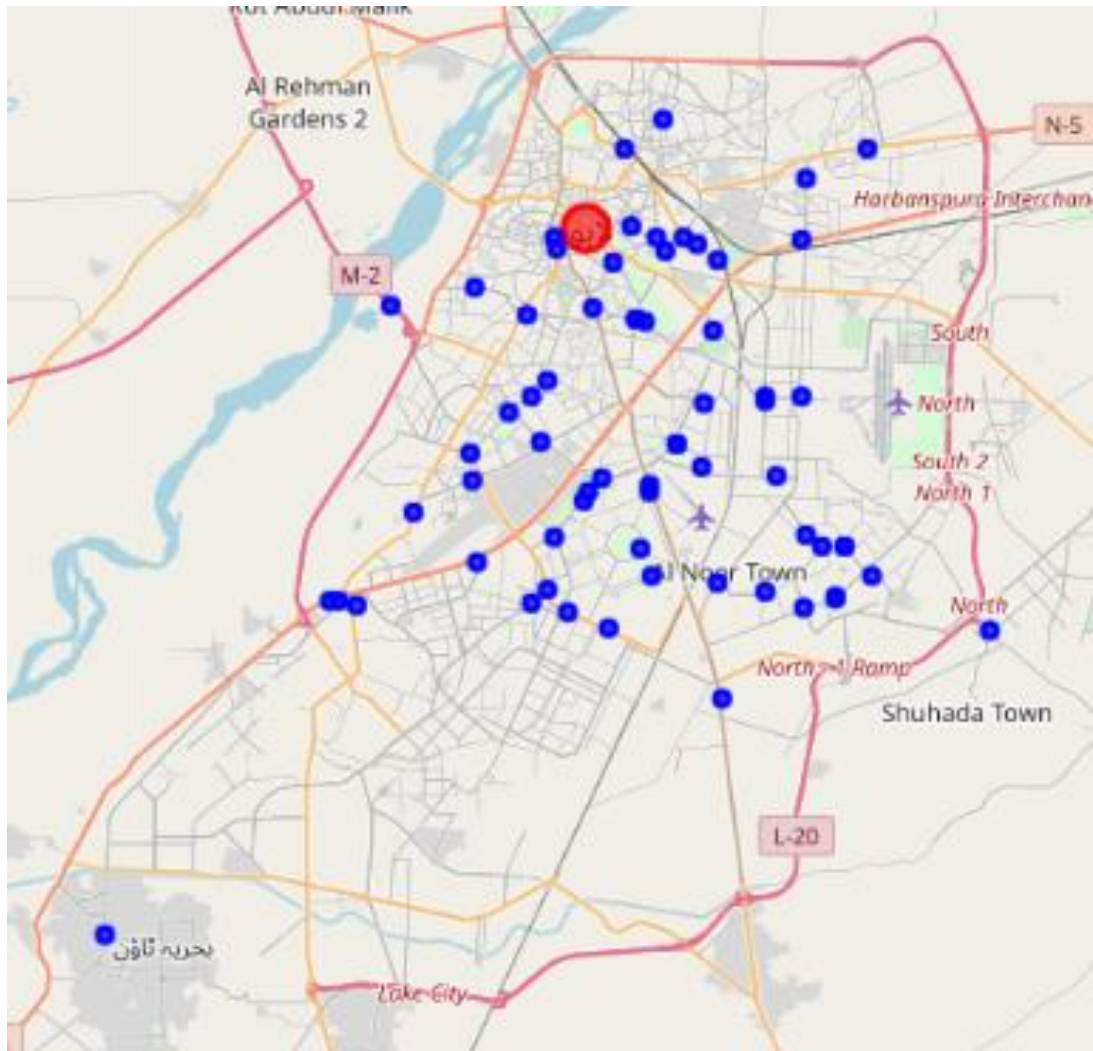
Gas stations are evenly distributed throughout the city. However, concentration of gas station in middle of city seems higher. Northern part of city, which is highly populated with very less greenery, has presence of handful of gas stations as well.

5. Conclusion

On conclusion, core reason of high pollution in city is lack of forestation and greenery such as parks. Overall, green areas in city are very few. Moreover, alarming situation is present in northern part of city where population in very dense and there are hardly any parks and forestation there.

At central part of city, population is also dense, and there are high number of gas stations. Green area again is little to none. Traffics flow is high in this area assuming on presence of gas stations in that area, thus leading to higher pollution.

Responsible authorities should immediately address the pollution problem by building more parks and forest areas in northern and central areas of city. Moreover, curbing pollution responsibility can be shared by individual citizen as well as they can plant trees outside their homes and neighborhood.