

Lim Li Xuan <https://public.tableau.com/profile/li.xuan.lim#!/>

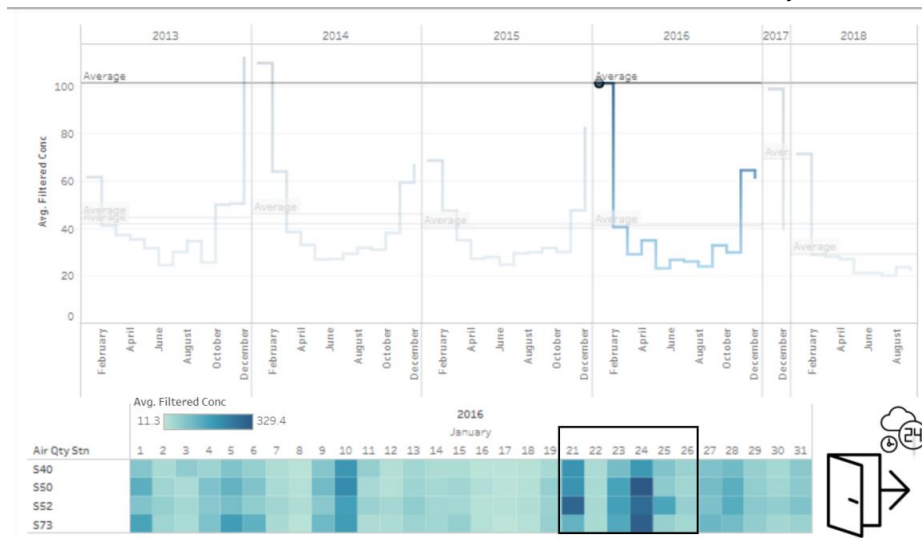
Interesting Findings/ Task

1A. Characterize the past and most recent situation with respect to air quality measures in Sofia City. With Respect to PM10 Concentrations taken, the average of each year is as shown StoryBoard 1

2013 (Daily)	2014 (Daily)	2015 (Daily)	2016 (Hourly)	2017 (Hourly)	2018 (Hourly)
44.42	46.19	41.07	41.09	Omitted because of missing data for Months Jan to Oct.	29.20

Comparing the PM10 Concentration of 2013 and 2018, there is a general decrease in PM10 Concentration over the years.

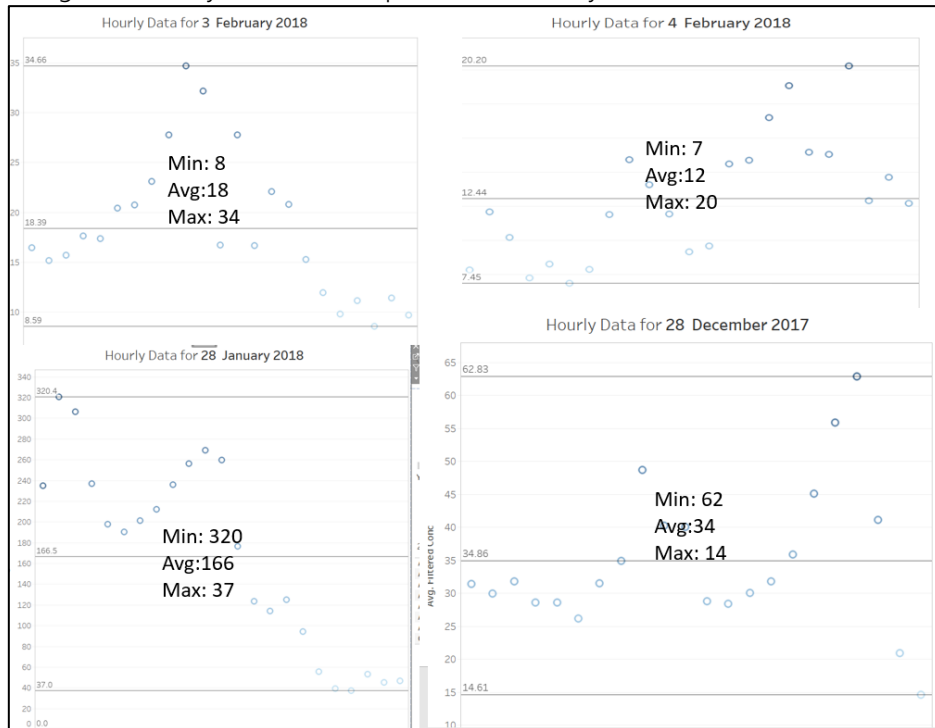
The concentration of PM10 usually peaks at the start and ends of the year, during January, November and December. This still holds true from 2013-2018, as shown in Story Board 1.



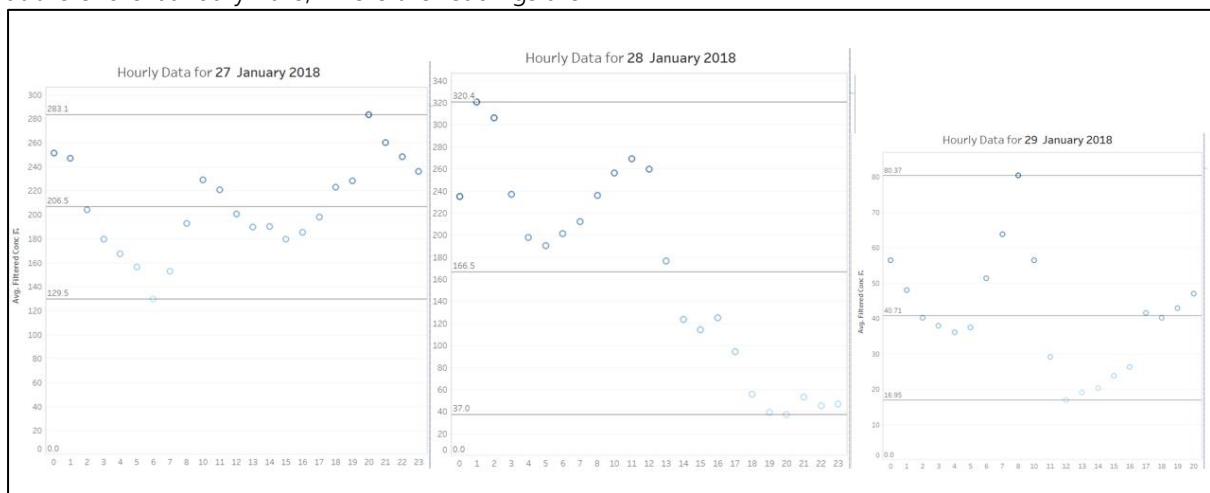
1B. What does a typical day look like for Sofia city? Do you see any trends of possible interest in this investigation?

Based on the hourly data [Storyboard 2],

On a typical day, concentration is of an average of about 40 across 2013-2018. The readings are never consistent throughout the day, with obvious peaks over the days



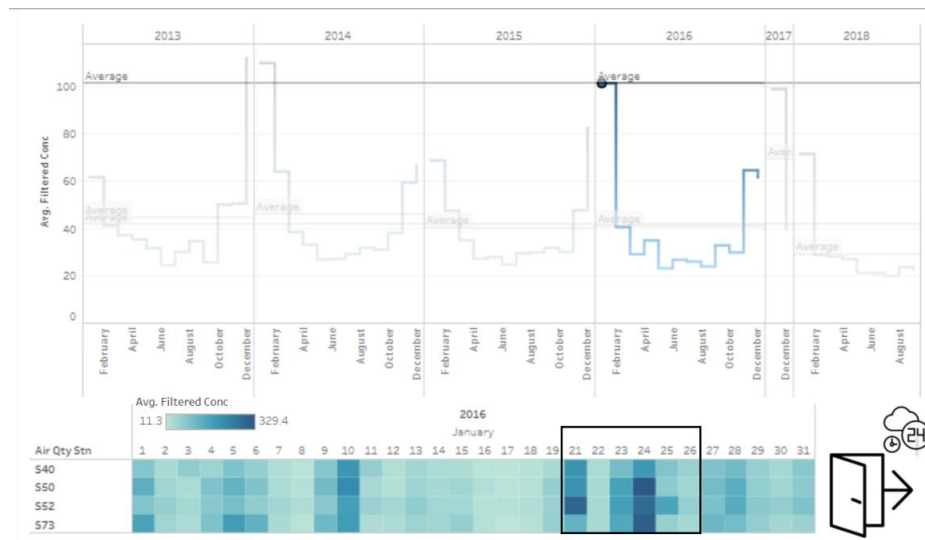
On a typical day, hourly readings are pretty consistent such that the maximum point is not a suspected anomaly, but is also accompanied by several other points. One possible point for investigation would be at the end of January 2018, where the readings are



An interest of investigation would be what caused the rise in levels from 26 Jan and what happened after 10AM ON 28 January to have cause the rapid decrease in concentration back to safe levels. However, the daily meteorological information is unable to provide further information on this.

1C. What anomalies do you find in the official air quality dataset?

1. There were extreme values that were beyond the concentration of 300 in 2017, which was questionable, given the average was only around 40. After taking a closer look, multiple different stations took turns to hold the most extreme scores (beyond 300), on the 24 of January 2018. Between 12AM to 1AM. Hence, a variable Percentile filter is created to filter anomaly based on research to decide on the threshold we want to use for future predictions.



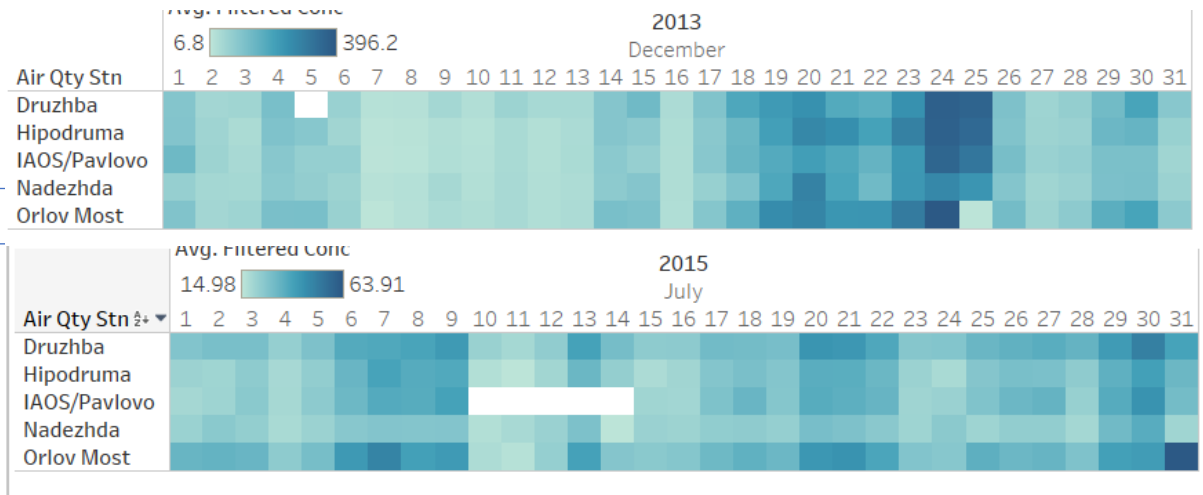
2. Missing data in 2017 : excluded from [Visual 1: Analysis across years] to prevent additional bias to the dataset
3. Air Quality STA-BG0079A and STA-BG0054A Had a lot of blanks. After researching, some of the Air Quality Stations closed down between 2013-2018. Removed to reduce inconsistencies and bias. For the rest of the stations, on some years and months, readings were also not absolutely consistent.

1D. How do these affect your analysis of potential problems to the environment?

1. Unable to plot a trend line spanning from 2016-2018 using hourly data.
2. By not having full data of stations, I resorted to finding the average of each time bin. However, there may be inaccuracies in the averaged-data as some only had 1 data point.

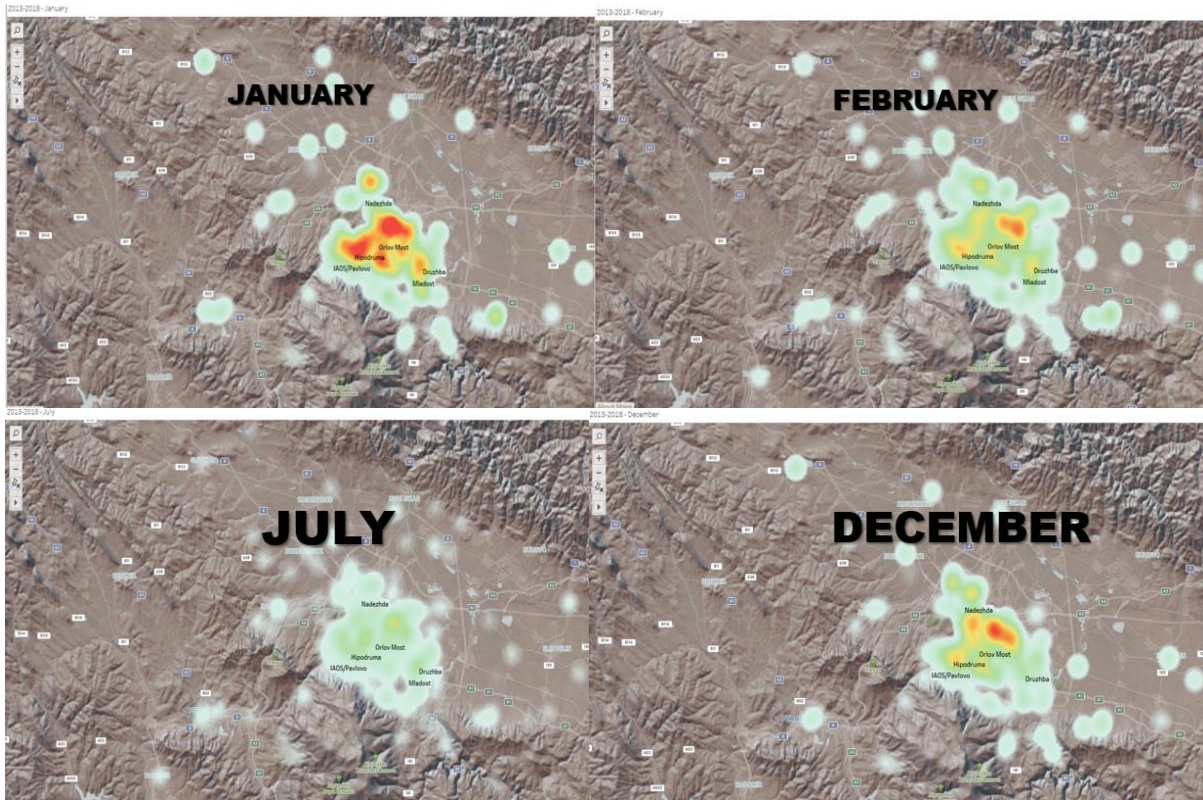
2A. Characterize the sensors' coverage, performance and operation. Are they well distributed over the entire city? Are they all working properly at all times? Can you detect any unexpected behaviors of the sensors through analyzing the readings they capture? (6 Images, 800 Words)

1. Station Orlov Most has erratic readings that does not follow the norms. It may have been faulty. It eventually stopped service halfway through 2015. Its geographis location is also such that it is being surrounded by other sensors, hence it does not make sense why it can have such different readings at times.

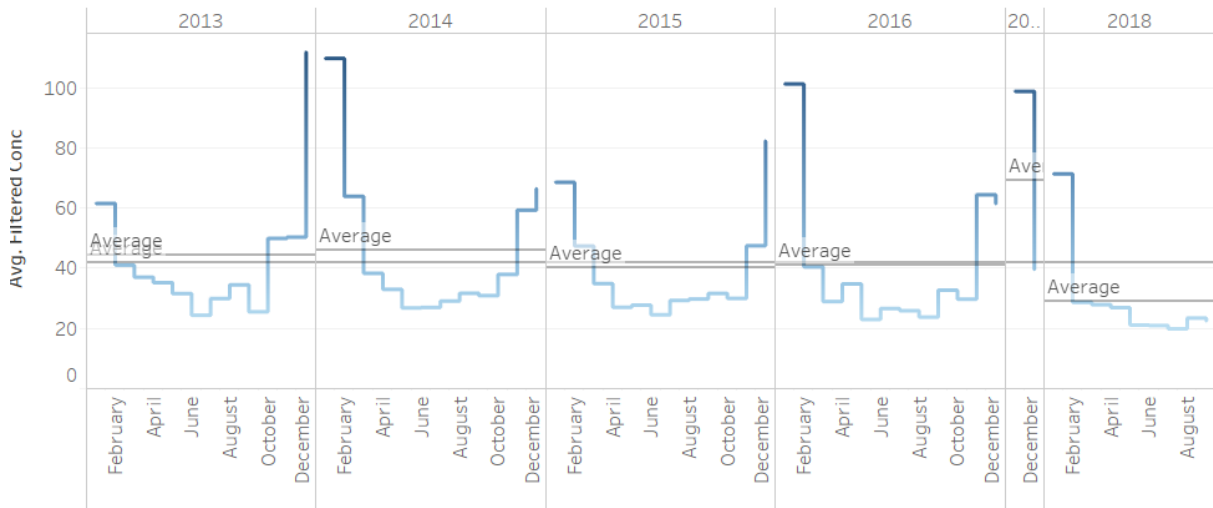


2B. Now turn your attention to the air pollution measurements themselves. Which part of the city shows relatively higher readings than others? Are these differences time dependent? Limit your response to no more than 6 images and 800 words.


1. The temperatures might be Month Dependent / Temperature Dependent
Monthly Averaged readings (From 2013-2018, Aggregated by average)



Judging by this, the trend of concentration is month dependent.



Following the investigation into these months, the underlying factor could be temperature.

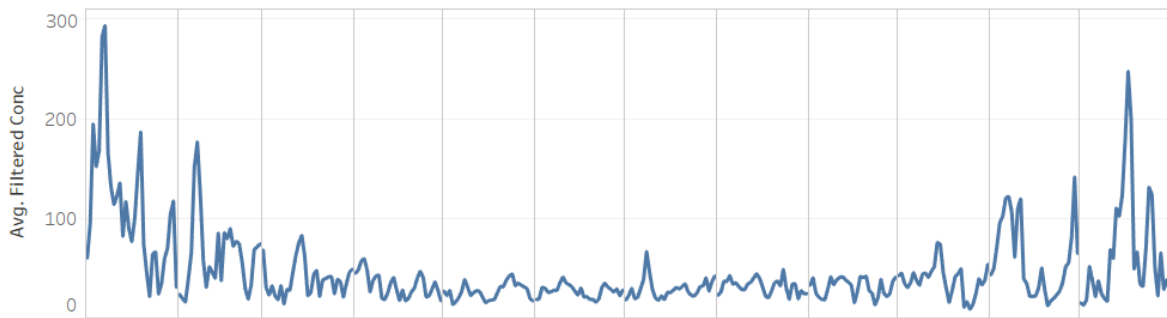
Select A Year 

2014

Daily Readings for 2014 

Bottom Pct.

Top Pct.



Select A View

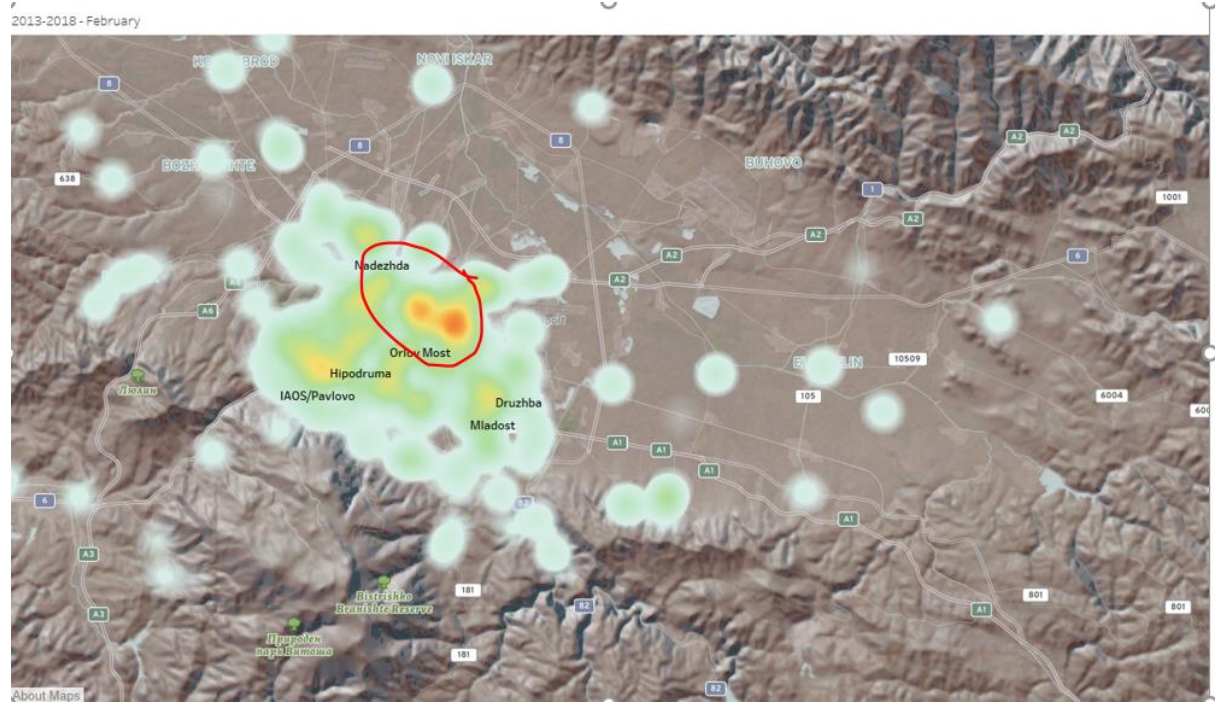
Temperature



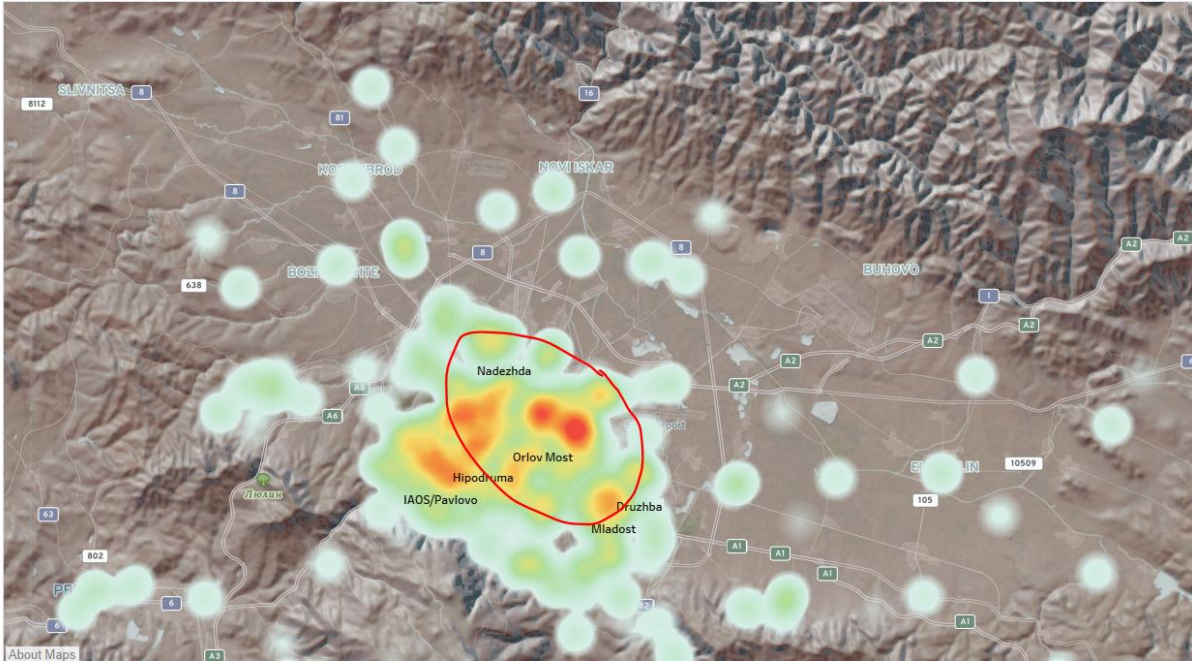
Based on further research, this is a phenomena of temperature inversion that traps smog and thus contributes to pollution. <https://sciencing.com/temperature-inversions-influence-air-pollution-10038430.html>

2. Elevation Dependent

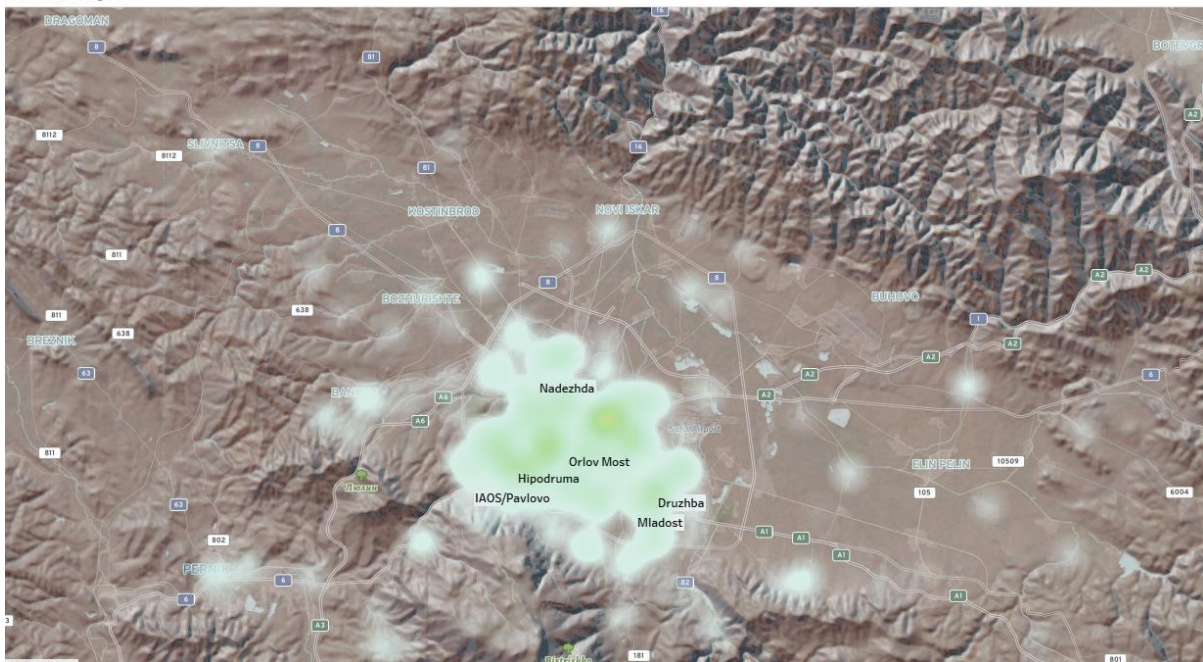
As shown on the terrain map, the lower left shows the map which has a higher elevation and usually would have lower readings.



2013-2018 - March



2013-2018 - August



3. You are required to reveal the relationships between the factors mentioned above and the air quality measure detected in Task 1 and Task 2. Limit your response to no more than 5 images and 600 words

By looking through the whole project, come plausible insights would be :

1. Temperature Inversion (Shown Above)
2. Elevation (Shown Above)
3. Wind speed (May also be caused by Elevation too)

On still sunny days when pollution levels can build up due to a lack of wind to disperse the pollution.

Old vehicles with gasoline engines, old diesel engines and the use of solid fuels for heating are the three main sources for the dangerous pollution in Bulgaria. Even though waste has been falling over the years, a drop in 195 kg per capita since 2008, incineration has increased. Just about a third of Sofia's waste goes to recycling and composting, while the remainder is turned into RDF for further incineration, or landfilled. These reasons are mainly habits formed since the past and require huge transformations to current processes and strategies. Unless Sofia city or Bulgaria does so, pollution is predicted to not improve.

As analysed, the current sources of pollution, coupled with geographical factors of being surrounded by valleys, would eventually emphasize the impact of meteorological conditions on Sofia.