

Francisco Francillon
Professor Jairam
CIS 3121 MSA[34793]
Final Project Report
12/04/21

The quality of the air we breathe has a long, and short term effect on our health. According to the National Geographic, short term effects on bad air quality towards humans include “discomfort such as pneumonia or bronchitis. They also include discomfort such as irritation to the nose, throat, eyes, or skin”. Long term effects include heart disease, lung cancer, and respiratory disease such as emphysema... also cause long-term damage to people’s nerve, brain, kidneys, liver, and other organs” and according to the National Geography article ‘nearly 2.5 million people die worldwide each year from the effects of outdoors or indoor pollution’. Effect of air pollution varies from children to adults. Children and the elderly are more at risk of developing health conditions. The air quality also affects the environment; they can contaminate the surface of bodies of water and soil, kill crops, kill young trees and other plants. “Hazy air pollution can even muffle sounds”. The bigger effect of air quality is the warming of the planet. ‘Global warming is an environmental phenomenon caused by natural and anthropogenic air pollution. It refers to rising air and ocean temperatures around the world’. Carbon dioxide from our burning of fossil fuels gets trapped in the atmosphere and warms up the average temperature of the earth. There are also natural greenhouse gases from things such as volcanoes. Less developed nations and large cities tend to have more air pollution. The article states that some of the world’s most populated cities are Karachi, Pakistan, New Delhi, India; Beijing, China; Lima, Peru; and Cairo, Egypt... many developing nations also have air pollution problems. Los Angeles, California, is nicknamed Smog City” [1].

For my research topic I wanted to focus on air quality in the United States and observe how they differ based on population and other factors such as wind. My motivation for doing this is because I am a living organism living on this planet and would like to be part of the change to curb global warming. I want to start locally so I chose to only focus on the United States for this research, but also include the European Air Quality Index alongside the Chinese Index for any discrepancies. I divided the United States into three regions eastern cities, southeast cities, and west coast cities. I chose to do that because I also wanted to see if there is any correlation by region. I also understand that different regions have different weather patterns, so I want to see if those differences will have an effect on the air quality.

I selected four data sources for this project. Those data sources include one API and three websites. The API of choice for this project is IQAir. This API goal is to give people access to timely air quality information. They are the “world’s largest free real-time air quality information platform and engage an ever growing number of global citizens, organizations, and governments” [2]. The other three data sources are three separate wikipedia pages providing cities, state, and population of the different regions of the United States. The east coast cities are located on the shoreline on the Atlantic Ocean, the original thirteen colonies of Britain are also located on the east coast. [3]. Big cities such as New York City is in this region Next is the southeast region of the United States. The climate in the south east is generally humid subtropical. I focus on the largest cities which tend to be midsize [4]. The west coast cities are much diverse and include both mid and big cities. The population varies, and states like California are considered to be some of the most polluted State In the United States. I believe

that the west coast region of the United state will be some of the most polluted due to prior knowledge, and I also believe that more polluted cities will be more polluted. In addition I also think wind plays a factor in pollution and that windy city will have less pollution.

The code for this project was really complex. I faced many constraints, and limitations. One constraint I had was with my API. The free version of this API allows for a maximum ten thousand calls a month and a max of five calls per minute. The API also only allows a maximum of 500 calls per day, limiting the amount of API calls I can make [6]. Because of my 5 calls per minute limitation my code includes the import time module to utilize the time.sleep module to add a fourteen second delay on each request. Due to this the full code takes around 30 minutes to run to completion. I also imported the itertools to allow me to iterate to multiple indexes in two different lists. The request module to send HTTPS requests to both the API and website that I scrapped. The Json module allowed me to format the json file I grabbed from the API. the pandas as pd allow for the creation of dataframe. Next the bs4 BeautifulSoup helps parse the HTML document, and extract data from the HTML file. Finally the matplotlib and pyplot allow for the visualization of data.

The global variable consists of eight lists, six of those lists are empty and data will be appended to them. The url list stores the three websites that I scrapped from. The city, population, and state hold data from the url link. The wind, aqicn, aqius, and url_list are used for the API. The wind stores the wind data, aqicn stores the Air quality of the Chinese model, and aqius stores the European model air quality.

There are four functions, one for the web scraping, and the other for the Dataframe and API the last two were for the visualization. the page variable sends the request to the url in the first index, soup variable parse through this data. The letGetit finds the class and table, and the correct table variable locates the specific table that will be scrapped. Next the for loops retrieve the city, state, and population data. This process gets repeated two more times for each url. The index of the url list is specified, and the respected variables are changed to one and two to show the respected url being requested. The function title DataFrame removed some of the retrieved data from the list due to the bus from the code, the API was unable to find those cities so I removed them. I also deleted index 105 to 185 due to my API constraint. If I kept all those cities it would have taken close to two hours to make all those API calls. Next I store the collected data from the websites to a dataframe called pd, and name the column city, state, and population respectively. I also changed the population to float to be able to analyze better. Next the key variable stores my API key, then the for loop creates the url to be used for the api call and stores it in the url_list. The for loop after data is used to make the API call by retrieving the url from the url_list, and retrieve the China air quality model, the European model, and the wind respectively. Each of these data are appended to their respected list. The time.sleep deley each call for 14 second, allowing four calls every minute. The second data frame stores the data from the API call, and then a third dataframe to merge the data from the first and second dataframe. Next I printed the first 10 rows of the third data frame, and used the describe method to print the descriptive statistic of the data frame. Then the third dataframe was saved as a csv file. Finally the last two functions were used to create three graphs, two scatter, and one bar graph showing the air quality vs population, air quality vs wind, and air quality vs city.

My findings were really surprising. Based on my bar graph, cities in the western part of the United States seem to have higher pollution in contrast to cities in the east coast. Irvine California has the highest with 166, and the smallest was also in the west coast with Fontana California at 0. I did not find any relation with air quality and population, what I did discover

was western cities tend to have worse quality. There was zero relationship with wind and air quality. Wind does not seem to have an effect on the air quality. Based on my results I can conclude that population is one of the factors that play into air quality, as there are more factors, such as the region you are located in. New York City with 8 million people has a rating of 41 in contrast Springfield Massachusetts has a population of 153,606 people has a rating of 48. These discrepancies demonstrate the other factors that could be at play. The results of my research are below:

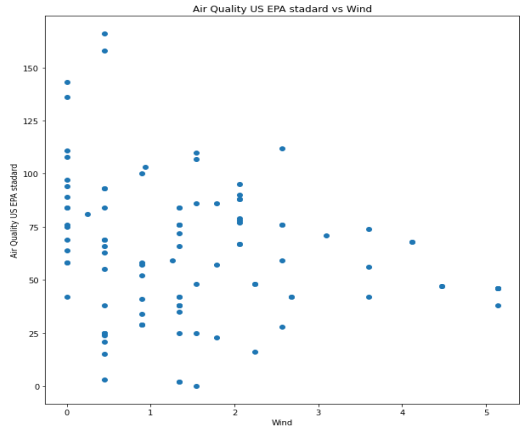
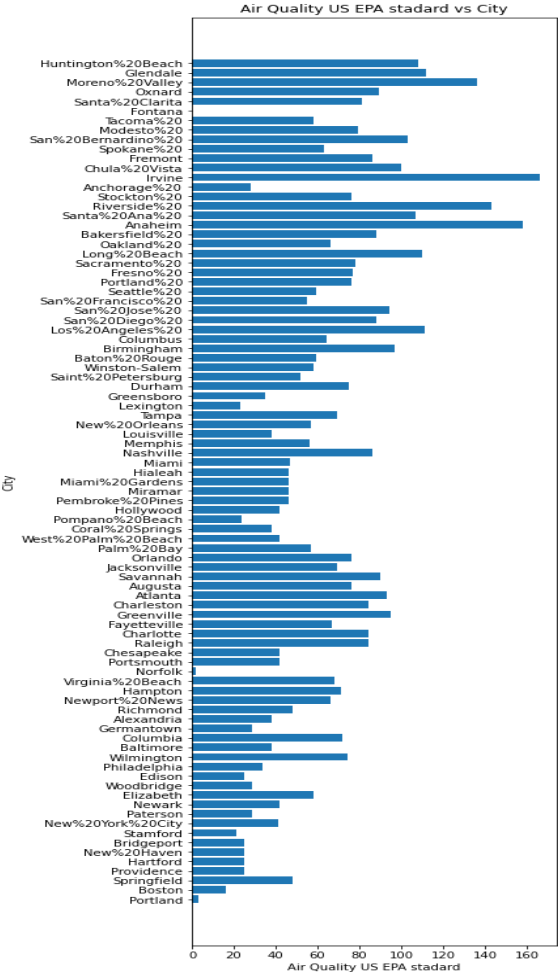
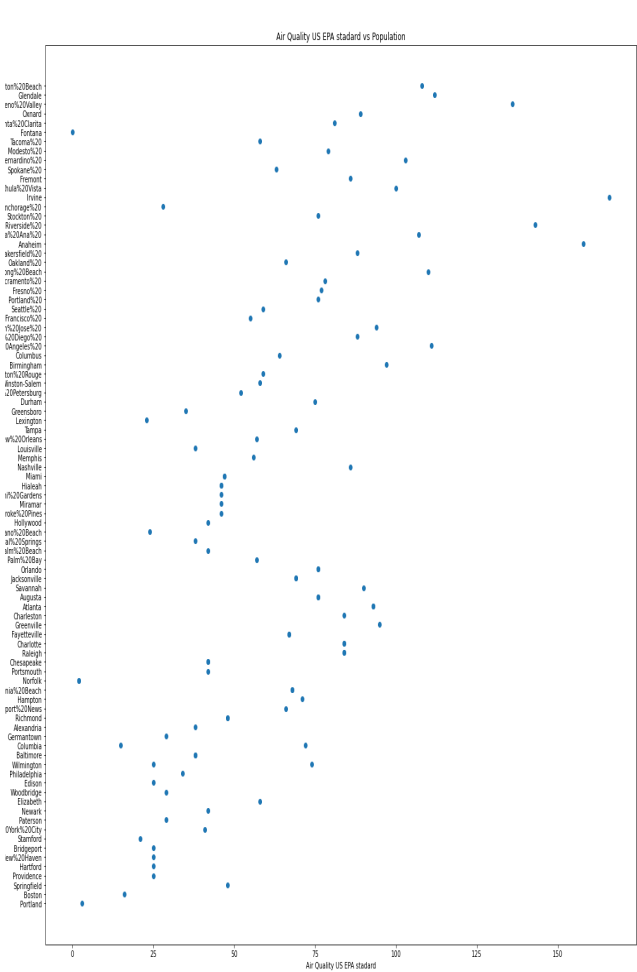
Statistical Description

	population	Air Quality China MEP standard	Air Quality US EPA standard	wind
count	1.050000e+02	105.000000	105.000000	105.000000
mean	4.481483e+05	28.047619	62.209524	1.528000
std	9.021373e+05	19.570082	32.124366	1.395115
min	6.641700e+04	0.000000	0.000000	0.000000
25%	1.548230e+05	14.000000	41.000000	0.450000
50%	2.389420e+05	23.000000	59.000000	1.340000
75%	4.527450e+05	40.000000	84.000000	2.060000
max	8.398748e+06	112.000000	166.000000	5.140000

Df3 first 10 rows

	city	state	population	Air Quality China MEP standard	Air Quality US EPA standard	Wind
0	Portland	Maine	66417	1	3	0.45
1	Boston	Massachusetts	694583	5	16	2.24
2	Springfield	Massachusetts	153606	17	48	1.54
3	Providence	Rhode%20Island	179335	9	25	0.45
4	Hartford	Connecticut	122105	9	25	1.54
5	New%20Haven	Connecticut	130418	9	25	0.45
6	Bridgeport	Connecticut	144900	9	21	1.34
7	Stamford	Connecticut	129775	7	41	0.45

8	New%20York%20City	New%20York	8398748	14	21	0.89
9	Paterson	New%20Jersey	145627	10	41	0.89



Bibliography

1. National Geographic Society. "Air Pollution." *National Geographic Society*, 9 Oct. 2012, <https://www.nationalgeographic.org/encyclopedia/air-pollution/>.
2. "Empowering the World to Breathe Cleaner Air." *IQAir*, <https://www.iqair.com/about-iqair>.
3. "East Coast of the United States." *Wikipedia*, Wikimedia Foundation, 6 Dec. 2021, https://en.wikipedia.org/wiki/East_Coast_of_the_United_States.
4. "Southeastern United States." *Wikipedia*, Wikimedia Foundation, 29 Nov. 2021, https://en.wikipedia.org/wiki/Southeastern_United_States.
5. "List of Largest Cities on the United States West Coast." *Wikipedia*, Wikimedia Foundation, 18 May 2021, https://en.wikipedia.org/wiki/List_of_largest_cities_on_the_United_States_West_Coast.
6. "Select the API Plan Which Best Fits Your Needs." *Select the API Plan Which Best Fits Your Needs*, <https://www.iqair.com/us/air-pollution-data-api/plans>.