Predicting Air Quality Index in Cleveland, OH

Anna Funsten
Ben Jackany
Chris Caldarella

2021-05-21

Outline

1) Air Quality and Asthma
2) Data Source and Data Cleaning
3) Regression Models
4) Streamlit Application
5) Summary
6) Sources

Air Quality Measures

1) Carbon Monoxide
2) Nitrogen Dioxide
3) Ozone
4) 10um Particulate Matter
5) 2.5um Particulate Matter
6) Sulfur Dioxide

Air Quality Measures

Daily Air Quality Index = pollutant with highest individual air quality index for that day

Carbon Monoxide	0.5 ppm	6 AQI
Nitrogen Dioxide	36 ppb	34 AQI
Ozone	0.026 ppm	24 AQI
10um Particulate Matter	21 ug/m3 SC	19 AQI
2.5um Particulate Matter	9.6 ig/m3 LC	40 AQI
Sulfur Dioxide	5 ppb	7 AQI
Cumulative AQI		**40 AQI**

Air Pollution and Lungs

Headache and anxiety (SO_2)

Impacts on the central nervous system (PM)

Irritation, inflammation and infections

Asthma and reduced lung function (NO_2)

Chronic obstructive pulmonary disease (PM)

Lung cancer (PM, BaP)

Irritation of eyes, nose and throat

Breathing problems (O_3, PM, NO_2, BaP)

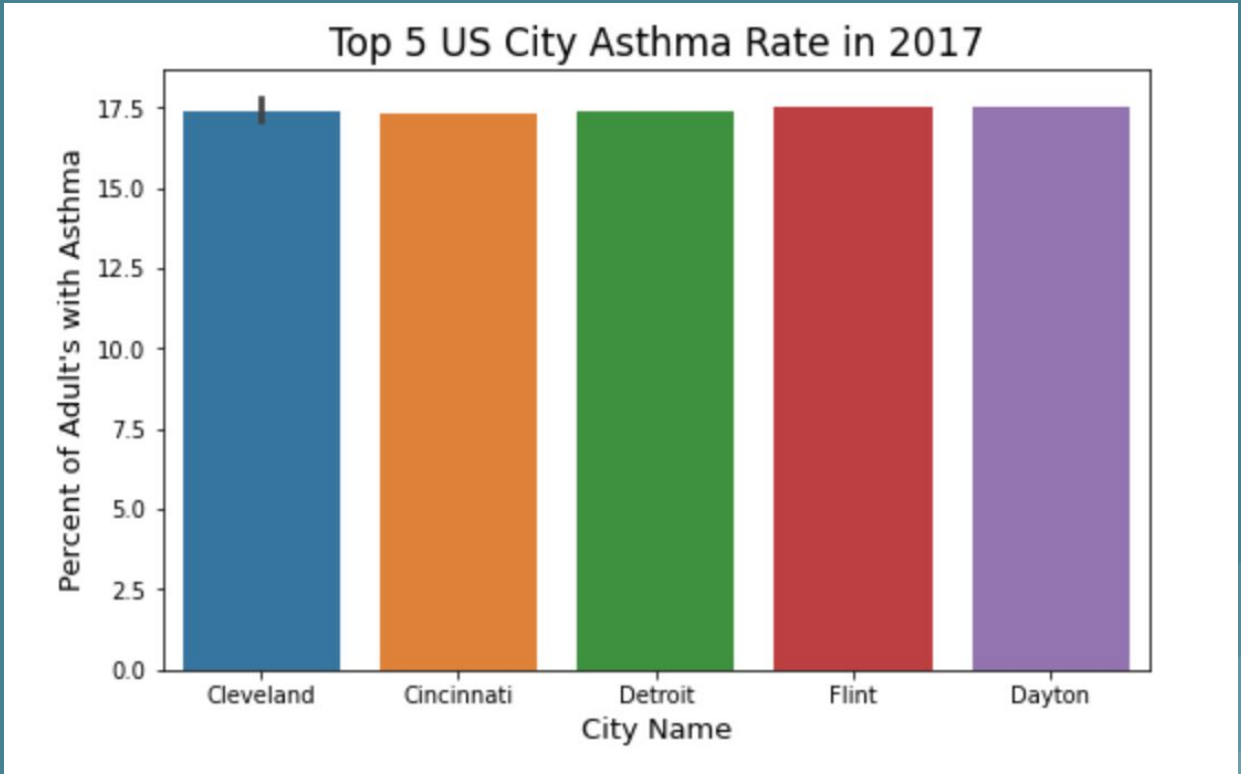
Cardiovascular diseases (PM, O_3, SO_2)

Impacts on liver, spleen and blood (NO_2)

Impacts on the reproductive system (PM)

Air Quality Index (AQI) Values	Levels of Health Concern
0 to 50	Good
51-100	Moderate
101-150	Unhealthy for Sensitive Groups
151-200	Unhealthy
201-300	Very Unhealthy
301 to 500	Hazardous

Asthma Rate in Cleveland, OH



What is the relationship between daily pollutants and Air Quality Index (AQI)?

Can we create a model that accurately predicts Air Quality Index (AQI)?

Data Source and Data Cleaning

-2017-2019 daily pollutant data from EPA

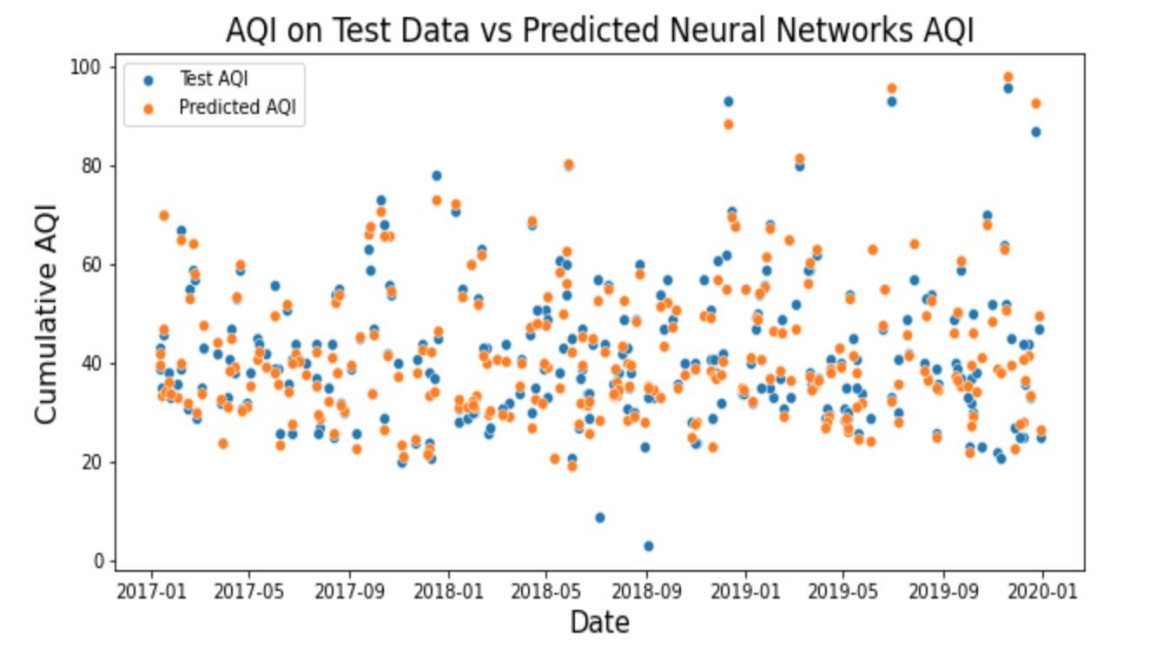
-Daily high and low temperature from Cleveland.com

-Calculated daily cumulative AQI

-Calculated average daily temperature

Model	RSME	R2 Score
Baseline (mean AQI)	15.441	.0000225
Linear Regression	8.236	0.762
Elastic Net	8.290	0.759
Lasso	8.290	0.759
Ridge	8.238	0.762
KNN Regressor	8.528	0.745
Neural Networks	**5.783**	**0.883**
Decision Tree Regressor	**7.223**	**0.817**

Best Model: Neural Networks



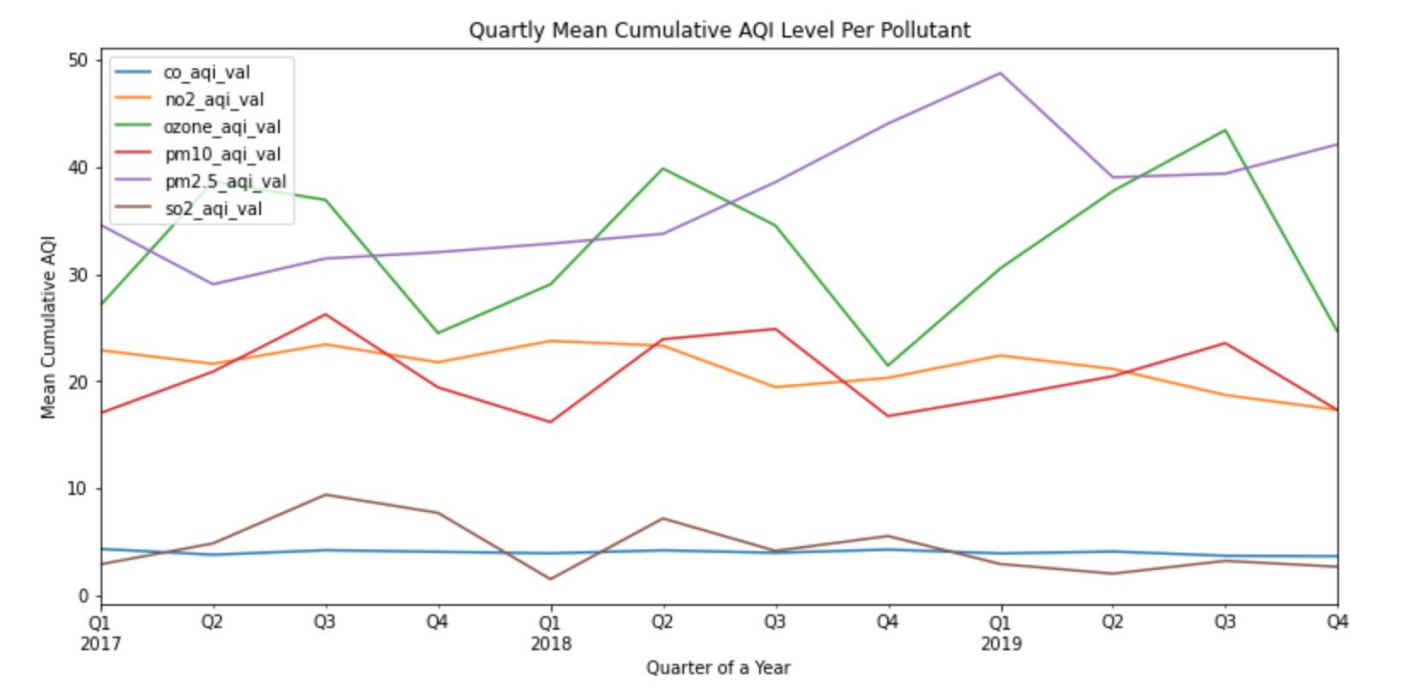
-5 Dense Layers:
50, 200, 50, 50, 1

-Early Stopping: 5

-Batch Size: 100

-Epochs: 100

Important Features Identified Decision Tree Regressor Model



- PM 2.5 values
- Ozone values

Streamlit

Summary

- **Best Model: Neural Networks**
  - RSME: 5.783
  - R2: .883
- **Recommendation:**
  - Pay attention to Ozone and PM 2.5 values regarding overall air quality
- **Limitations**
  - Data specific to Cleveland, OH
- **Further Research**

Sources

https://www.eea.europa.eu/themes/air/health-impacts-of-air-pollution

https://www.cnet.com/health/air-quality-index-how-to-tell-if-the-air-quality-is-bad-in-your-area/

https://ecolink.com/info/six-common-air-pollutants/

https://www.epa.gov/outdoor-air-quality-data/download-daily-data

https://www.cleveland.com/datacentral/2008/09/cleveland_weather_history_find.html

https://www.airnow.gov/sites/default/files/2020-05/aqi-technical-assistance-document-sept2018.pdf

https://www.epa.state.oh.us/portals/27/ams/plans/NCORE-GTCraig.pdf

https://www.breeze-technologies.de/blog/what-is-an-air-quality-index-how-is-it-calculated/