



Impact of Air Quality Index and Pollutant Exposure on Respiratory Mortality

A Big Data Analytics Approach using Dask

Supervisor: Dr. Muhammad Zulkifl Hasan
Co-Supervisor: Hussain Ijaz

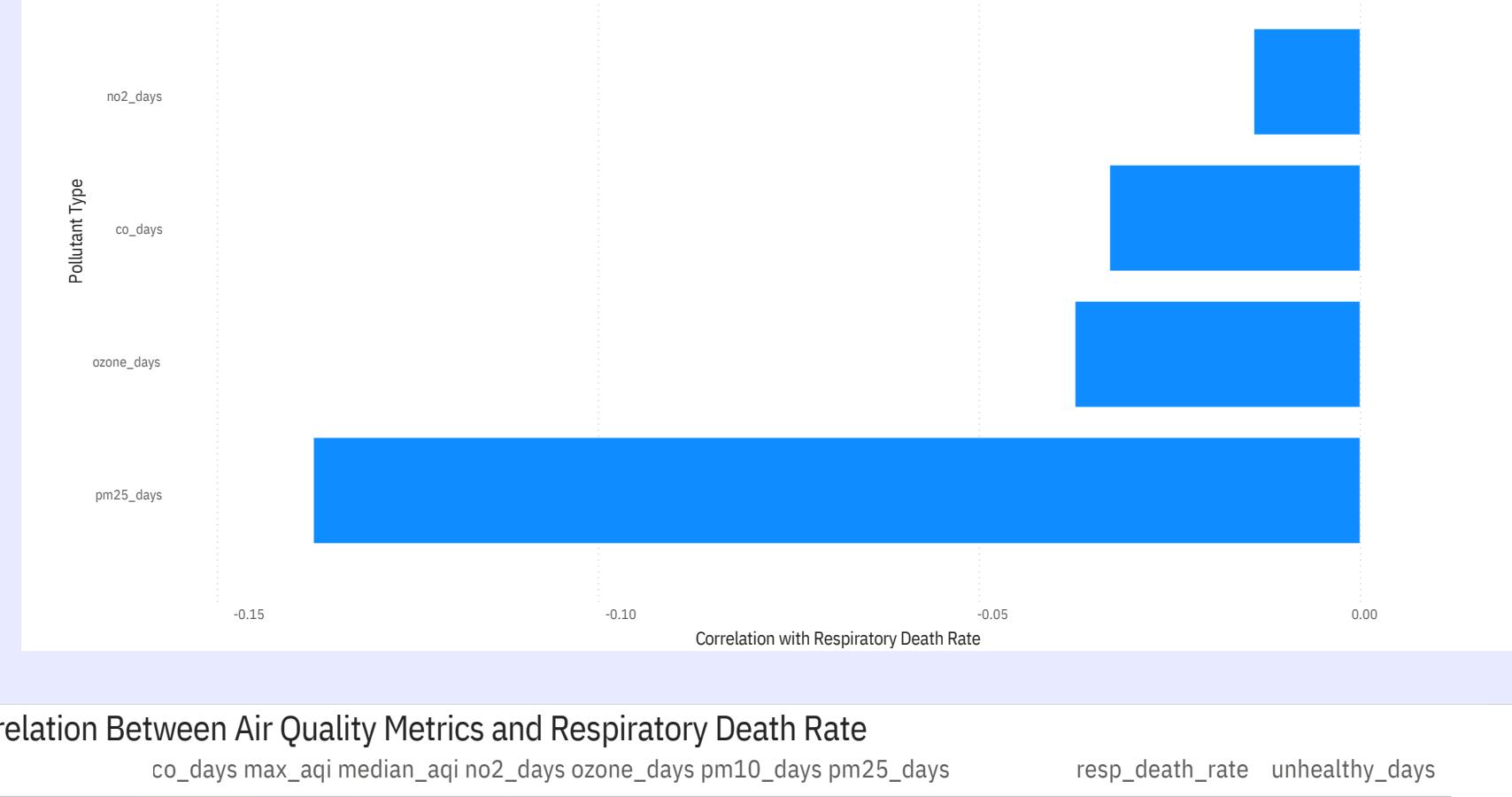
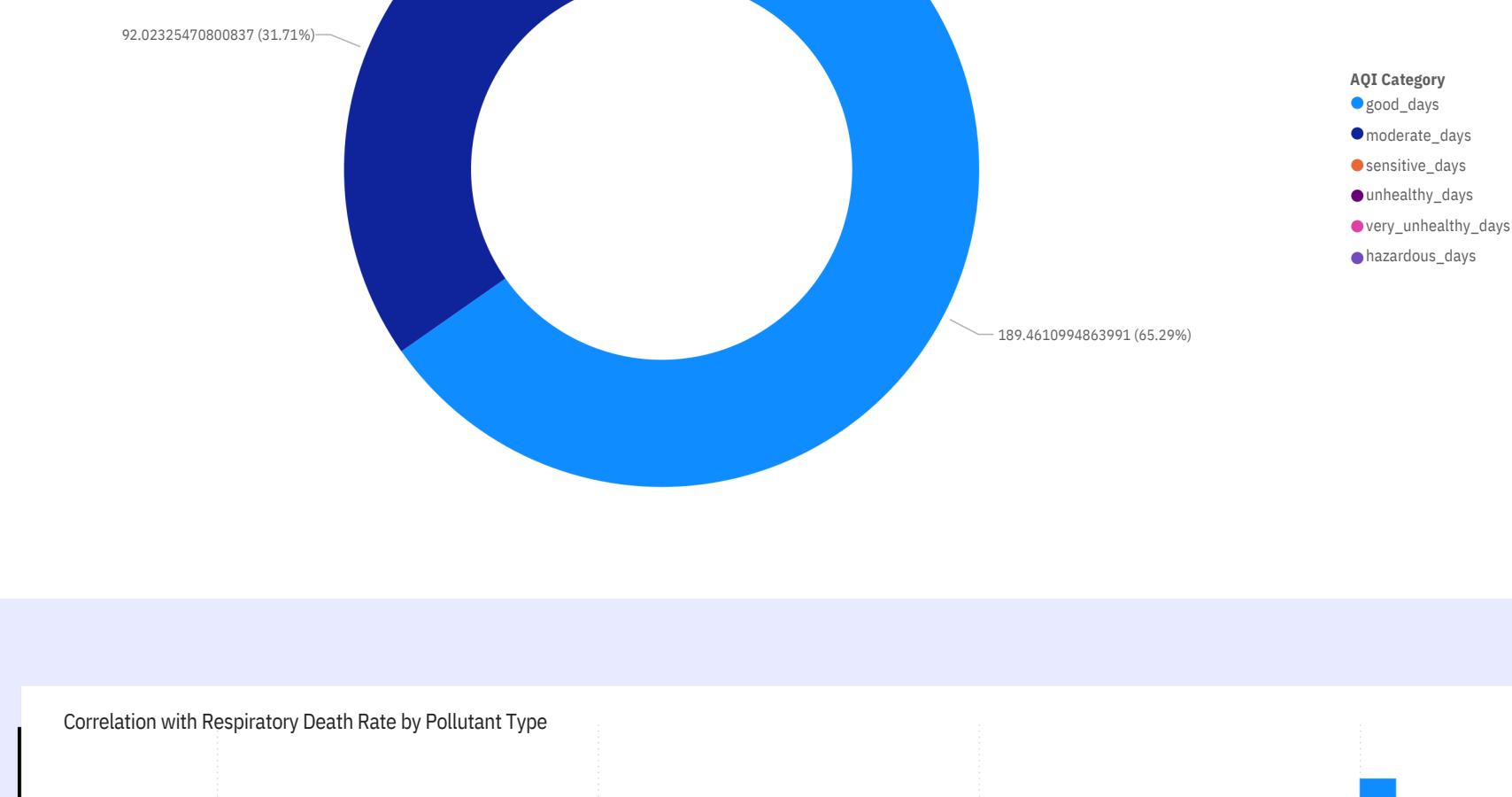
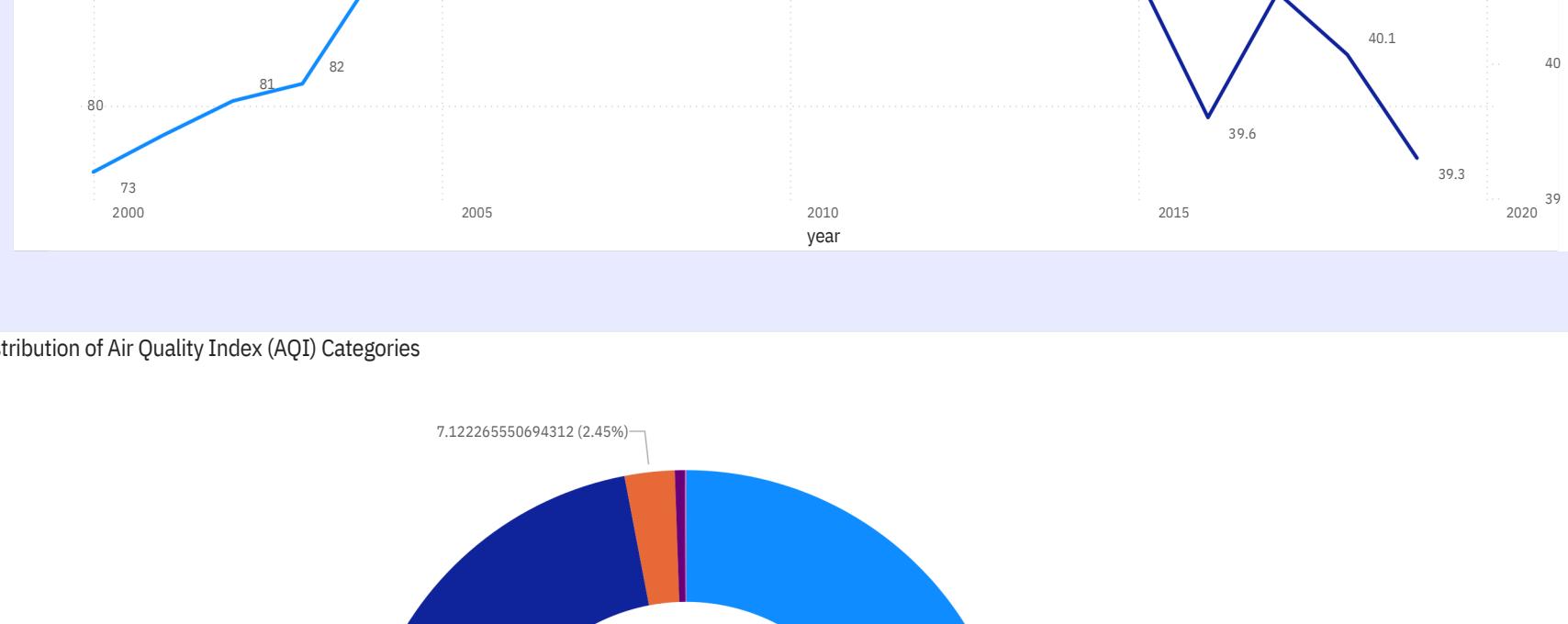
M. Arham¹, Alishba Zahid¹
¹CE Department, ITU



ABSTRACT

Air pollution is a major public health concern worldwide. This study analyzes the relationship between air quality indicators and respiratory death rates across counties using large-scale AQI data. Distributed data processing with Dask was employed to handle multi-year, multi-regional data efficiently. Statistical analysis and machine learning models were applied to identify trends, pollutant impacts, and spatial patterns. Results highlight the role of PM2.5 exposure and regional disparities in respiratory mortality, emphasizing the need for data-driven environmental health policies.

RESULTS



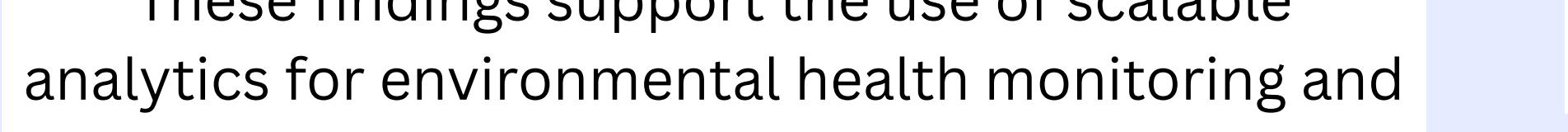
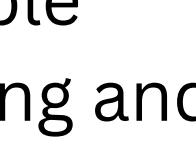
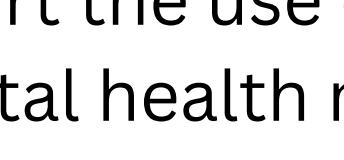
	co_days	max_aqi	median_aqi	no2_days	ozone_days	pm10_days	pm25_days	resp_death_rate	unhealthy_days
co_days	1.00	0.00	-0.12	0.01	-0.05	0.01	-0.04	-0.01	-0.01
max_aqi	-0.02	1.00	0.04	-0.01	0.08	-0.33	0.30	-0.07	0.09
median_aqi	-0.12	-0.01	1.00	-0.01	-0.03	-0.02	-0.06	-0.03	0.43
no2_days	0.01	0.01	-0.01	1.00	-0.03	-0.02	-0.06	-0.03	0.09
ozone_days	-0.05	0.01	0.08	-0.03	1.00	-0.25	-0.56	-0.04	0.08
pm10_days	0.01	0.14	-0.33	-0.02	-0.25	1.00	-0.15	0.03	-0.03
pm25_days	-0.04	-0.01	0.30	-0.06	-0.56	-0.15	1.00	-0.12	0.02
resp_death_rate	-0.01	-0.07	-0.08	-0.03	-0.04	0.03	-0.12	1.00	-0.10
unhealthy_days	-0.01	0.09	0.43	0.09	0.08	-0.03	0.02	-0.10	1.00

CONCLUSION

This study demonstrates how big data analytics can uncover meaningful patterns between air quality and respiratory mortality. While correlations at the county level are weak, machine learning and spatial analysis reveal the significant role of PM2.5 and regional exposure disparities. These findings support the use of scalable analytics for environmental health monitoring and policy planning. Literature Review

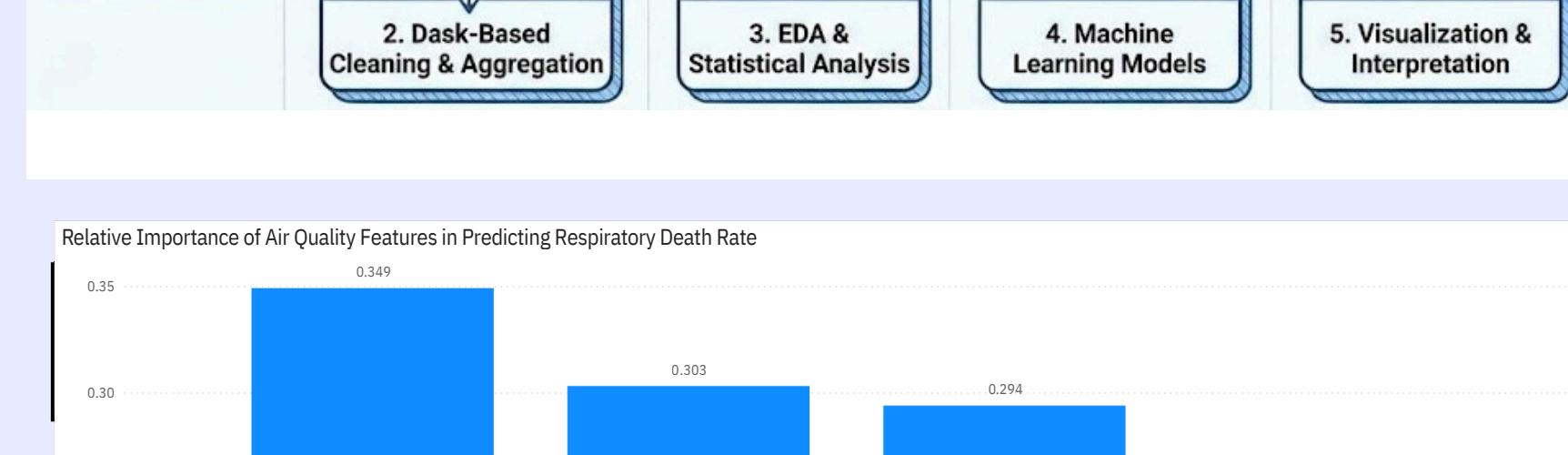
Prepared by: Muhammad Arham & Alishba Zahid | Big Data Analytics | 2025

Tools: Dask • Python • Power BI • Tableau



LITERATURE REVIEW

- Prolonged exposure to poor air quality is linked to respiratory diseases such as asthma, COPD, and lung cancer.
- PM2.5 and ozone are consistently identified as major contributors to respiratory mortality.
- Air Quality Index (AQI) is widely used as a population-level indicator of pollution exposure.
- Prior studies often rely on small datasets; scalable analytics remain limited.



REFERENCES

- J. Smith, "AQI relation to respiratory death rate," Kaggle Dataset, 2020. [Online].
- World Health Organization, Air Pollution and Child Health: Prescribing Clean Air, Geneva, Switzerland, 2018.
- S. Liang et al., "Big data analytics for environmental health monitoring," IEEE Access, vol. 8, pp. 134 567–134 578, 2020.
- A. McGovern et al., "Using machine learning to improve air quality forecasting," IEEE Computer, vol. 50, no. 7, pp. 66–75, Jul. 2017.