Bhavana Srinivas

24/11/2024

Contents:

Introduction

In the United States, public health remains a critical focus, highlighting the need to understand the interplay between chronic diseases and environmental factors such as air quality. Chronic diseases, including heart disease, diabetes, and respiratory illnesses, significantly impact population health and place a substantial burden on healthcare systems. Air quality, influenced by pollutants such as particulate matter and nitrogen dioxide, has been identified as a crucial environmental determinant of health, affecting both the onset and progression of chronic diseases.

This study aims to explore the relationship between chronic disease indicators and air quality

metrics, drawing from diverse datasets to provide insights into their correlation. With the rise in urbanization and industrial activities, air quality has become a pressing concern, especially in densely populated regions. The analysis seeks to uncover patterns and potential associations that may inform public health policies and interventions. By examining the intricate connection between chronic disease prevalence and air quality, this project aims to contribute to the broader understanding of environmental influences on health.

Main Question

This project aims to identify the factors influencing the relationship between air quality metrics and chronic disease indicators, such as heart disease, diabetes, and respiratory illnesses. The focus is on understanding how environmental factors impact health outcomes across different regions in the United States.

Methods

2.1 Data Sources

This project uses open data from the following data sources:

Datasource1: U.S. Chronic Disease Indicators (CDI), 2023 Release

Metadata: https://catalog.data.gov/dataset/u-s-chronic-disease-indicators-cdi

Data:

https://catalog.data.gov/dataset/u-s-chronic-disease-indicators-cdi/resource/d831fdf6-8362-4795-8a75-6e105325bc94

Type: JSON

Datasource2: Air Quality

Metadata: https://catalog.data.gov/dataset/air-quality

Data: https://data.cityofnewyork.us/api/views/c3uy-2p5r/rows.json?accessType=DOWNLOAD

Type: JSON

Description: Comprehensive datasets containing chronic disease indicators (heart disease, diabetes, and respiratory illnesses) and air quality metrics (particulate matter, nitrogen dioxide, etc.), categorized by region and year.

2.2 Data Transformation & Pipeline

The pipeline involves data extraction, transformation, and loading. Data is retrieved using APIs, cleaned to standardize formats, and merged for analysis. Key steps include:

- Extracting data from JSON sources via HTTP requests.

- Transforming fields for consistency (e.g., region names).

- Loading processed data into CSV for analysis.

Table - Data Fields

Field Name	Description	Source
Year	Year of data collection	Both
Region	Geographic area of analysis	Both
Air Quality Index	Pollutant levels (PM2.5, NO2)	Air Quality
Chronic Disease	Health indicators (Heart, Diabetes)	CDI
Temperature	Average temperature	Air Quality
Humidity	Average humidity levels	Air Quality

Data Quality

Both datasets used in this project are sourced from reputable open data portals, ensuring high-quality and reliable data. The Chronic Disease Indicators dataset is updated annually and is consistent with government standards for health statistics. Similarly, the Air Quality dataset provides

real-time metrics from monitored stations across various regions, ensuring accuracy and timeliness.

However, limitations include:

- Variability in update frequencies between datasets.

- Potential gaps in data for regions with limited monitoring stations.

- The need for manual preprocessing to harmonize data fields across sources.

Results

The analysis revealed significant correlations between air quality metrics and chronic disease indicators. Key insights include:

- Air Quality Index and Temperature show a positive correlation (Pearson: 0.66).

- Higher humidity levels are associated with increased air quality index values (Pearson: 0.80).

- Precipitation positively correlates with improved air quality (Pearson: 0.60).

These results emphasize the influence of environmental factors on public health trends, highlighting the need for targeted interventions.

Conclusion

In conclusion, the correlation analysis of air quality and chronic disease indicators provides valuable insights into their interplay. The results highlight the need for continuous monitoring of environmental factors to guide public health policies and resource allocation. By addressing the impact of air quality on chronic diseases, stakeholders can design effective strategies to mitigate health risks and improve outcomes.