

NYC Borough Air Quality and Crime Report Statistics

Description

Dataset is a cumulative data frame that consists of statistics regarding both air quality and crime reports of the Boroughs in the City of New York, separated by each year between 200g-2020. The purpose of the dataset is to provide a comprehensive look at possible relationships between air pollution and number/types of crime committed in an urban setting. Content of the dataset that could help with this analysis includes, the average value of four types of air pollution (O3, NO2, SO2, PM2,5) within each Borough throughout the years, the number of total crimes committed categorized by age and gender groups, the number of certain types of crime committed (assault, larceny, drugs), etc.

Annual Statistics

Keywords

Air Pollution NYC Boroughs

NAAOS Standards

NYPD Crime Reports

Use cases (potential real-world application of the dataset):

- 1. What is the effect of Ozone pollution on person's mind?
- 2. Which pollutant type is most correlated with increase in crime?
- 3. Is there a pattern between crime and season? 4. Which demographic group is most affected by air quality?
- 5. Which Borough has the highest crime rate vs air quality?

How to use it?

Intended Use

- Intended Domain. Environmental research
- Intended Domain. Policy making.
- Intended Domain. Human psychology
- Intended Use. Analyzing trends and correlations between air pollution level and crime rates.
- Other Responsible Uses. Making policies on air quality standards and policing based on trends or correlations found.

Known Uses

Research/analysis To better understand what is needed for well-being of humanity.

Restrictions on Use

Not used for making prejudiced inferences for certain Boroughs: Data does not provide a clear cut picture of what each neighborhood is actually like. NOT used to make general assumption about air quality of the effect for different individuals: Data does not take into account many individual factors of a particular person.

① Do Not Use

Domain. Any in which data is not used for collective well-being of

Risky

Unknown

Number of issues

Risky

Safe

Unknown

For unethical reasons.

About the dataset

People NYC OpenData V Owned by Created by Group BB6 V City of New York > Maintained by Funding City of New York > City of New York > Management

Technical information	
Publish Date	2023-11-24
Format	CSV
Instances	340 recorded data for each Borough in each given year
Version	1
License	Public Domain
Collection timeframe	2009-2020
Collection process	^

Air pollutant data is collected by the NYCCAS real-time air quality monitor network reporting the air quality of the NYC Boroughs at different years. Crime rate data is collected by the NYPD Police Department consisting of arrest reports made across each NYC Boroughs throughout the year including demographic and type details.

Dataset access point

https://github.com/INFO-zo1-Fall-zoz3-Final/final-projects-sgiang1

Link to the data dictionary for the rows of the dataframe.

https://github.com/INFO-201-Fall-2023-Final/final-projectssgiang1/blob/main/data_wrangling/data_dictionary.md

Representation

Subpopulation: Borough, Pollutant, Age, Gender

Concerns about using data to make decisions/predictions at the individual level: The data does not account for many individual factors such as current health status and other environmental exposure, thus the data should not be used to make predictions at the individual level.

Grouping by location, gender, or age group may allow us to see some general trends and insights among each demographic group but will still not be representative of each

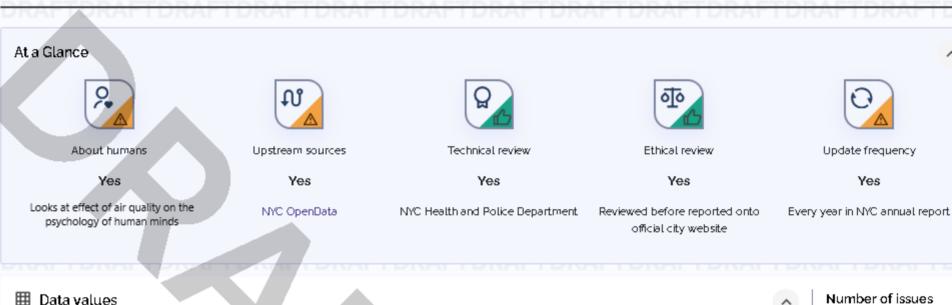
Other potential representation issues: For crime reports, there may be an overrepresentation of reports due to potential practices

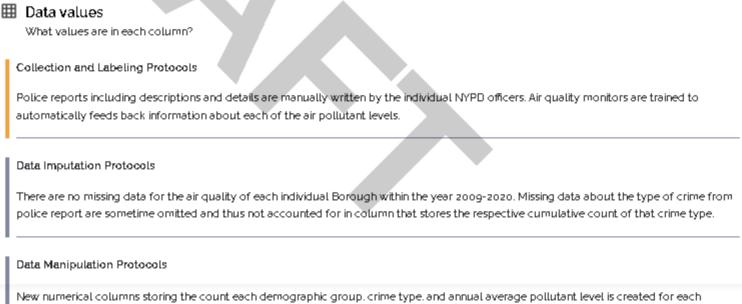
Promotion of Human Values

of predictive policing, which ethically is unfair and biased since it disproportionately targets low-income neighborhoods and high minority areas. This may also be caused by different strictness in enforcement between different demographics. For air quality, only areas with sensors installed within are correctly represented, thus not all of New York City is

Alert Count Completeness Upsteam: Air Quality dataframe Underrepresentation Inaccurate Prediction Upstream: NYPD dataframe Racial bias Socioeconomic Bias 8 Human Rights Principles Privacy Accountability Safety and Security Transparency and Explainability Fairness and Non-discrimination Human Control of Technology Professional Responsibility

Inference risks



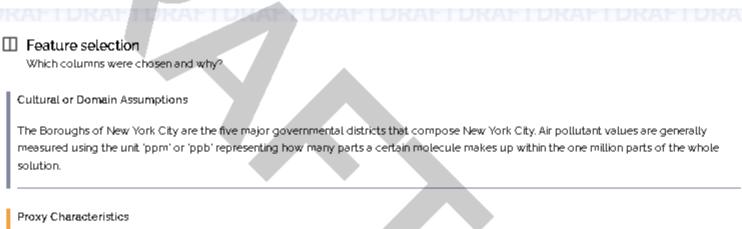


Borough during each given year from zoog-zozo. New categorical variable such as the Boolean above_NAAOS_standard stores TRUE or FALSE based on whether the given average pollutant value given its pollutant type is above or below government standards. Missing Data

Some description for crime were not included within report. Air quality sensors may not be the most accurate.

Raw Data

https://catalog.data.gov/dataset/



Person's age may have influence on likelihood to commit crimes. Air quality may affect the crime rate within a given Borough.

Planning Representation

Reports were based on an observational study as opposed to a survey.

Domain Knowledge

The Boroughs of New York City are the five major governmental districts that compose New York City. Air pollutant values are generally measured using the unit 'ppm' or 'ppb' representing how many parts a certain molecule makes up within the one million/billion parts of the whole solution.

Opstream sources Are there known risks in datasets upstream? Intended Use Familiarity

NYC OpenData: Very familiar. Data are collected through observation and are intended to be used by government.

officials such as policy makers or researchers along with the general public to analyze.

Data Collection Familiarity

 NYC OpenData; Very familiar. Representation due to oversampling in certain neighborhoods may cause bias within the data. Areas marked with predictive policing may consist of overrepresentation of a certain demographic while other areas may be underrepresented. Areas with less air quality sensor networks such as more remote places. are subject to underrepresentation as the air there is not recorded as often.



General risks Any additional risks?

Individual Information

Consent

Yes. Individual privacy is respected and private information is secured in government records.

Generalized Inferences

Data is only from NYC, thus data will reflect mostly the trends present in said state. Rural neighborhoods may not be represented as much as urban ones due to lack of resources installed there to record data on.

Generalized Inferences - Mitigation

Data should only be used to make inferences on the mentioned Boroughs within New York.

Sensitive Content

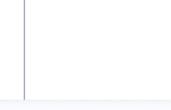
Other Known Issues

The data might incorrectly or disproportionally represent a certain demographic group in a bad light.

Documented Known Issues

https://catalog.data.gov/dataset

The data frame only encapsulate, on the grand scale, a small time frame around only a decade of year, thus, the data may not be representative of the actual causal or effect between air quality and crime.



Number of issues

Risky

Unknown