Number of issues

Risky

Unknown



# NYC Borough Air Quality and Crime Report Statistics

NYPD Crime Reports

## 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.

## Keywords

Air Pollution Annual Statistics NYC Boroughs

### NAAOS Standards

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?
- 5. Which Borough has the highest crime rate vs air quality?
- 4. Which demographic group is most affected by air quality?

## Known Uses

How to use it?

Intended Use

Intended Domain. Environmental research

Intended Domain. Human psychology

Intended Domain. Policy making.

pollution level and crime rates.

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

Other Responsible Uses. Making policies on air quality standards and

Intended Use. Analyzing trends and correlations between air

policing based on trends or correlations found.

### 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
- For unethical reasons.

## About the dataset

## People NYC OpenData V Owned by Created by Group BB6 V Maintained by City of New York > 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:

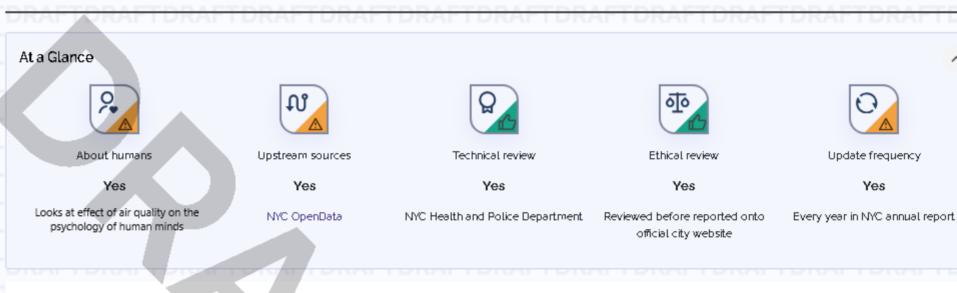
For crime reports, there may be an overrepresentation of reports due to potential practices of predictive policing, which ethically is unfair and biased since it disproportionately targets low-income neighborhoods and high minority areas. This may also be caused by different strictness in enforcement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

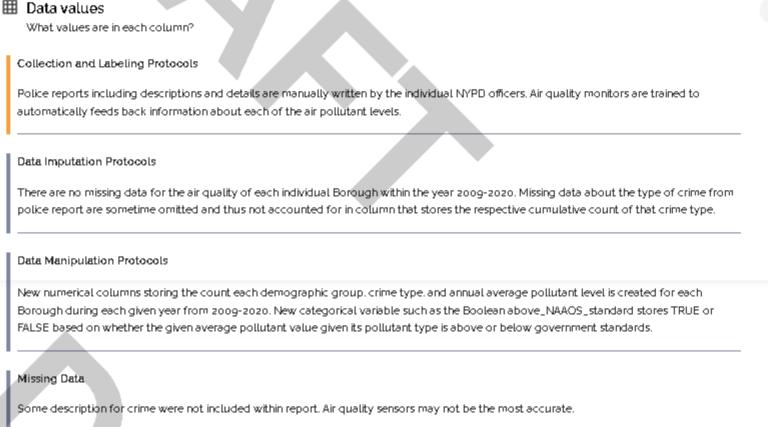
| Alert Count                     | 4  |
|---------------------------------|----|
| Completeness                    | (  |
| lpsteam: Air Quality dataframe  |    |
| Underrepresentation             | 1  |
| Inaccurate Prediction           | 2  |
| Jpstream: NYPD dataframe        |    |
| Racial bias                     | 1  |
| Socioeconomic Bias              | 1  |
| Human Rights Principles         |    |
| Privacy                         | R  |
| Accountability                  | R  |
| Safety and Security             | R  |
| Transparency and Explainability | R  |
| Fairness and Non-discrimination | NR |
| Human Control of Technology     | NR |
| Professional Responsibility     | R  |
| Promotion of Human Values       | R  |

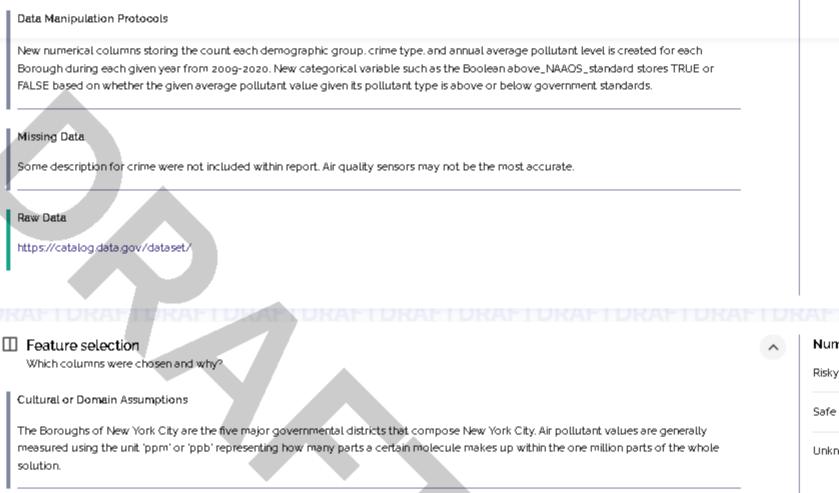
misdemeanor, violation, <18, 18-24, 25-44, 45-64, 65+, drug\_use, larceny, DUI, assault,

total\_crime, crime\_per\_value, above\_NAAQS\_standard, start\_season

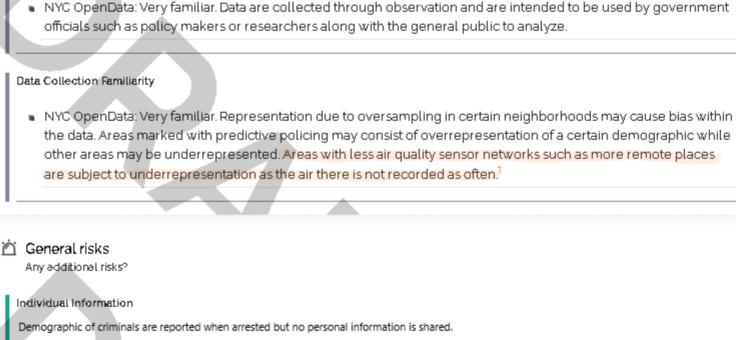
### Inference risks





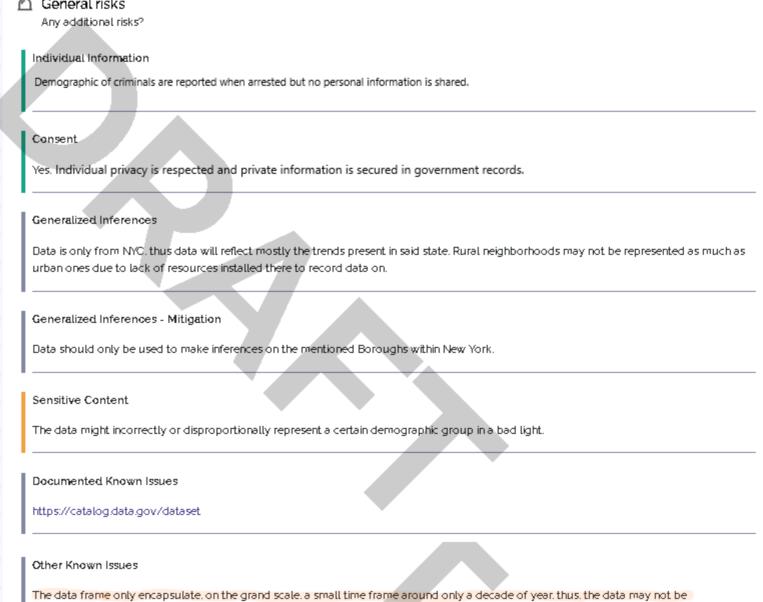






representative of the actual causal or effect between air quality and crime.

Intended Use Familiarity





Number of issues Risky

Unknown