Predicting and Classifying Tree Health in New York City

Alexander Coles Archita Jayaram Manav Patel Mattias Larsson Thomas Larsson

Motivation/Introduction

The Problem

Inequitable Urban Greenery: Poorer and minority communities in NYC face adverse health impacts from insufficient tree maintenance and distribution.

Lack of Advanced Solutions: Current tools lack advanced analytics and algorithms to address disparities and predict future issues for targeted interventions.

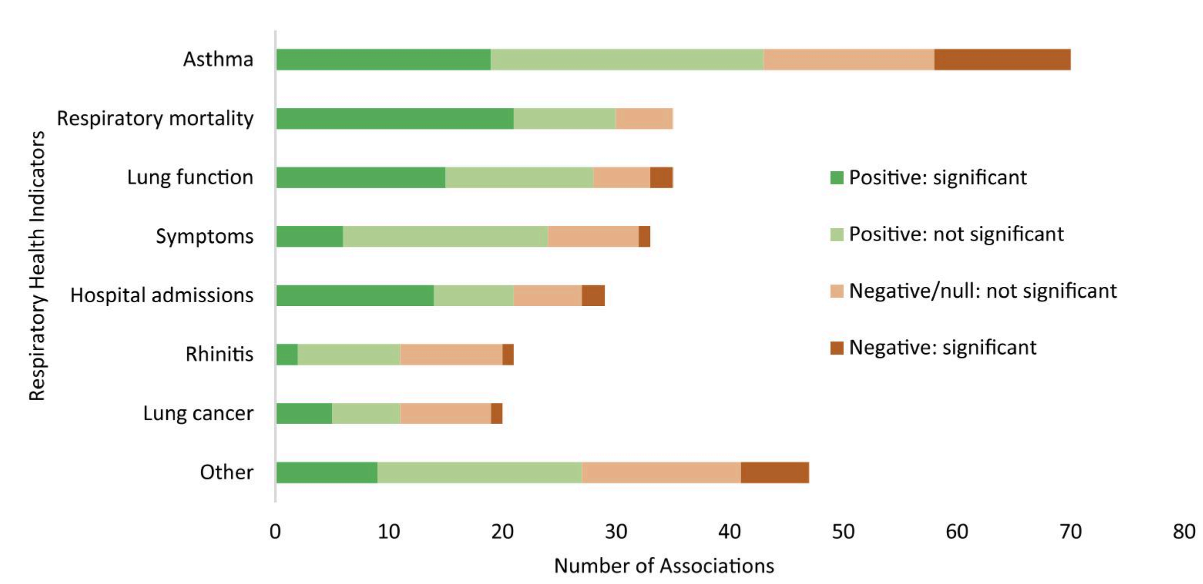
Importance

Public Health Impact: Insufficient urban greenery contributes to poorer air quality, higher urban heat, and negative mental health outcomes, disproportionately affecting vulnerable communities.

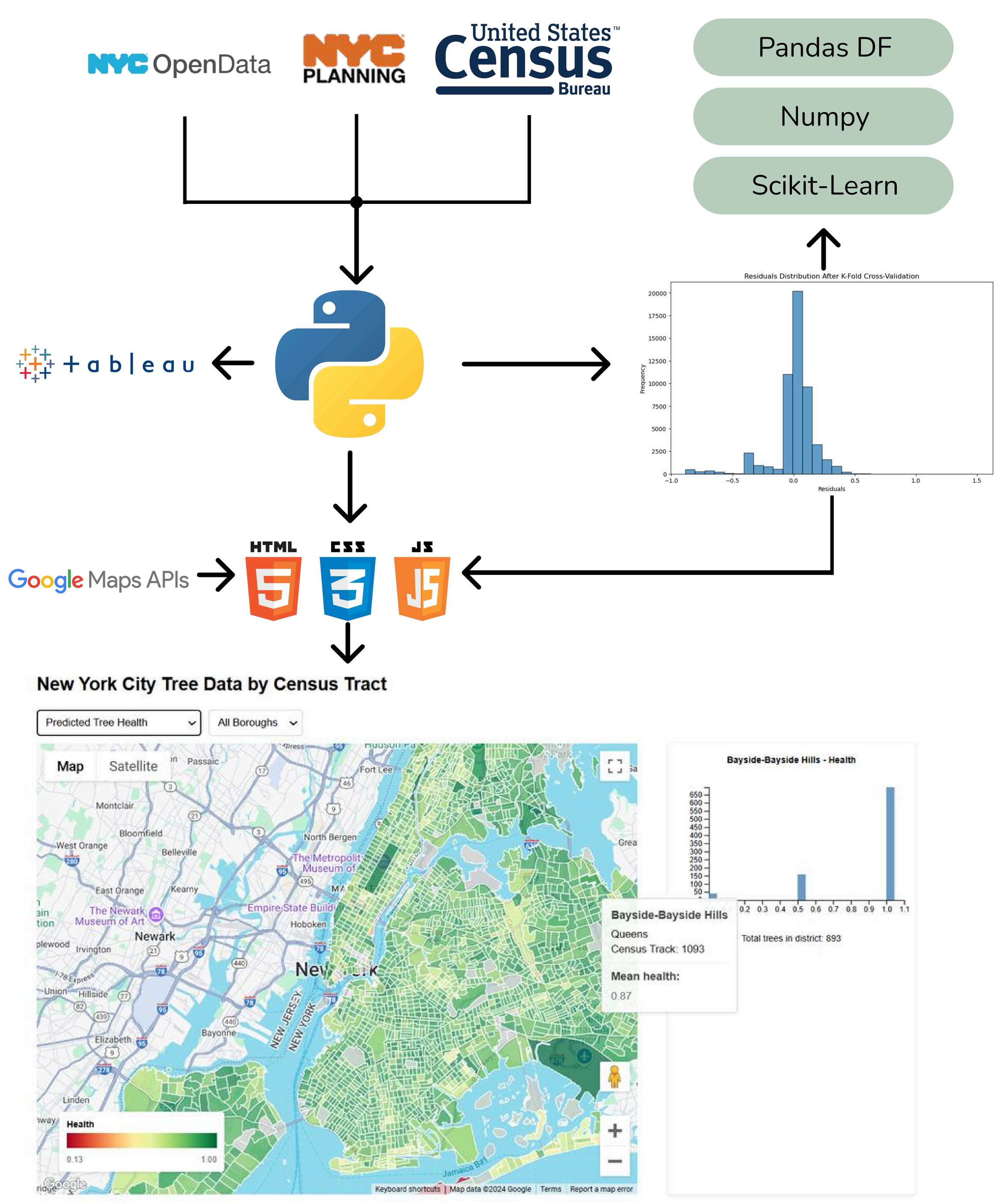
Environmental Equity: Addressing disparities in tree maintenance promotes fairness and improves the quality of life in under-served neighborhoods.

Sustainable Urban Planning: Equitable tree distribution and proactive maintenance are critical for creating resilient, livable cities in the face of climate change.

Green-space and Respiratory Association Chart



Graph highlighting the association between urban green-spaces and respiratory health indicators



Approaches

Interactive Map

An **interactive Choropleth Map** visualizes tree health alongside neighborhood data, allowing stakeholders to **explore data visually** for more targeted insights and decision-making.

Predictive Algorithm

An **interactive Choropleth Map** visualizes tree health alongside neighborhood data, allowing stakeholders to **explore data visually** for more targeted insights and decision-making.

How does it work?

1. Large tree health dataset, census dataset, and census tract JSON downloaded
2. Tree health dataset quantified by converting categorical variables to integer variables
3. Tree health dataset and census dataset cleaned by aligning census
4. Tree health dataset cleaned by making all dead trees have health 0 and 0 values in other variables instead of null, removed additional null values, and aligned the two datasets
5. Tree health dataset and census dataset joined on census tract and borough
6. Joined dataset used to train Decision Tree algorithm
7. D3 choropleth map created with toggle modes between baseline dataset information and algorithm implementation into map
8. D3 choropleth map overlaid over Google Maps API for greater visual clarity and usability
9. City planners are able to use the interactive map and adjust their mode/parameters to create different visualizations of tree health and socioeconomic relationships in New York, assisting with future city planning and investments

Why does it solve the problem?

Our method helps city planners to make informed investment decisions through

Pattern Identification
Links tree health to neighborhood conditions, revealing resource gaps.

Targeted Insights
Gives planners granular data to address green space disparities.

Scenario Exploration
Allows testing of outcomes for informed resource decisions.

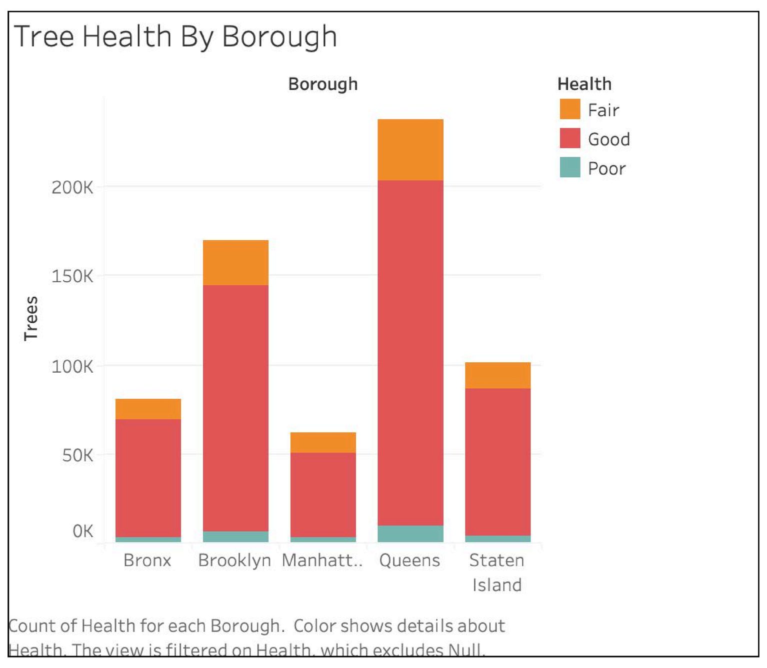
What's New?

Our approach is innovative in a variety of ways such as:

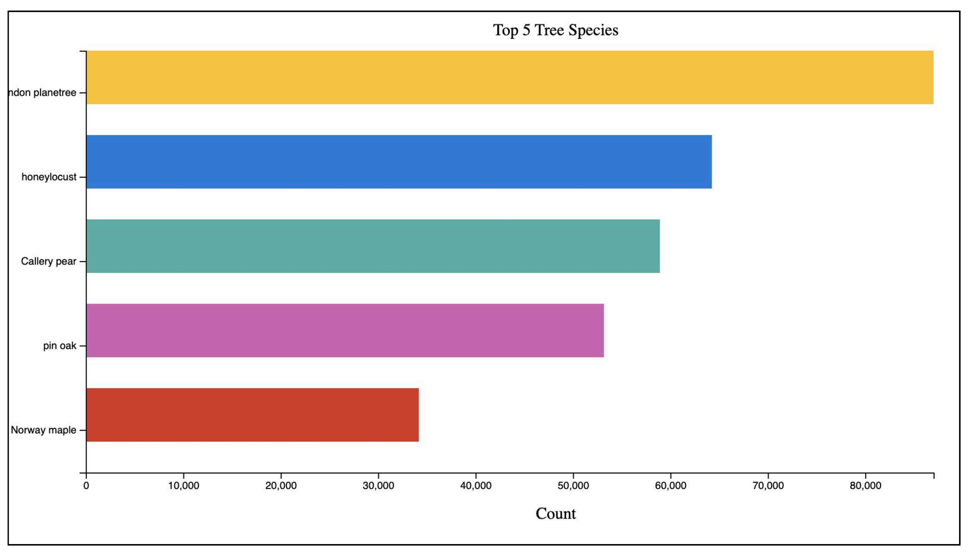
Integrated Machine Learning Models:
Combining multiple machine learning models (K-Means Clustering, Neural Networks, and Decision Trees) to predict tree health and socioeconomic conditions, providing a more comprehensive analysis

Holistic Visualization
Integrating both tree data and socioeconomic data into an interactive Choropleth Map to offer a dynamic view of geographic disparities, both current and predicted

Interactive Features:
Use of Google Maps API, tooltips, zoom feature, and more enhances the geographical context and user engagement



Count of Health for each Borough. Color shows details about Health. The view is filtered on Health, which excludes Null.



The Data

NYC Tree Data (2015 Street Tree Census)

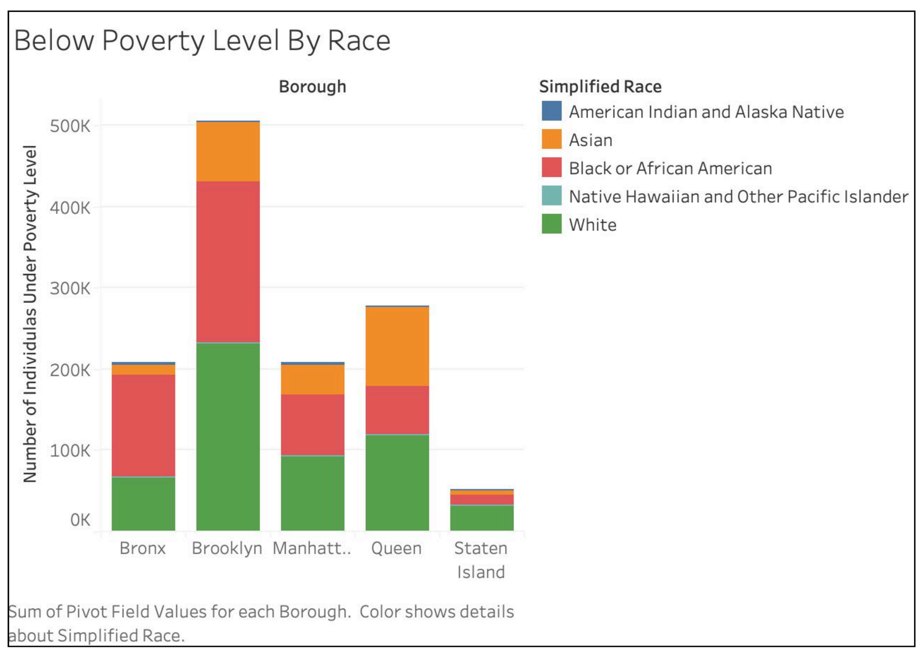
How did you get it?
- Downloaded from NYC Open Data platform

Characteristics:
- Size on disk: ~150 MB
- # of records: ~700,000 tree observations
- Sample Feature: Health, Stewardship, Census Tract

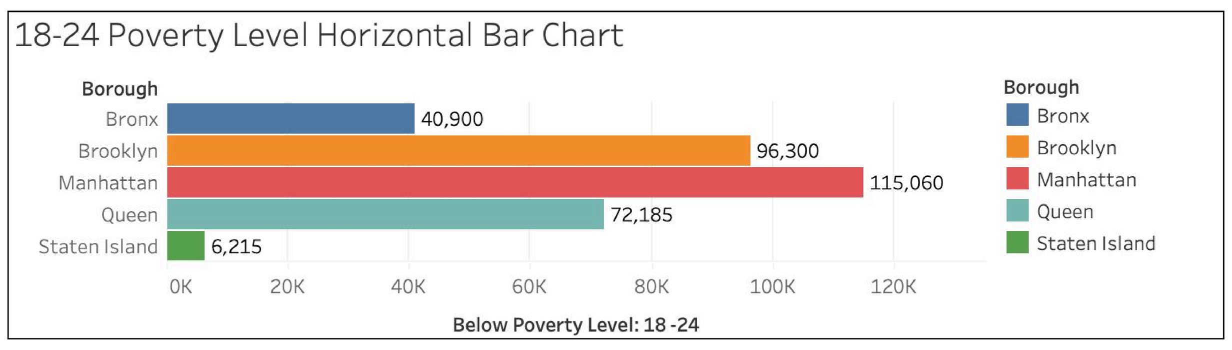
ACS NYC Census Tract Data

How did you get it?
- Downloaded from the U.S. Census Bureau's website

Characteristics:
- Size on disk: ~10 MB
- # of records: Approx. 2000 census tracts in NYC
- Sample Feature: Poverty Percentage by Age, Race, Education Level, and more



Sum of Pivot Field Values for each Borough. Color shows details about Simplified Race.



Experiments

Evaluated model efficiency through training the models and considering accuracy and other metrics.

User study implemented to evaluate effectiveness of the visualizations through clarity, interpretability , and engagement.

User study implemented to evaluate usability of the interactive map with features such as zoom, tooltip, toggle feature and more.

Results

K-Means Clustering
- Training/Test Split: 70:30
- Metrics:
  - MSE: 0.2
  - R^2: 0.75

Decision Tree/Random Forest
- Training/Test Split: 70:30
- Metrics:
  - MSE: 0.15
  - R^2: 0.82
  - MSE (Random Forest): 0.13
  - R^2 (Random Forest): 0.84
- Socioeconomic Metrics:
  - MSE: 0.039
  - R^2: 0.93

Neural Network (Tree Health)
- Cross-Validation: 5-fold
- Metrics:
  - MSE: 0.39
  - R^2: 0.40

Positive Feedback: The project received an 8/10 for clarity, 10/10 for interpretability, and 8/10 for engagement.

Clorepth Map Adjustments: Based on feedback, the color gradient was adjusted for better interpretation, and the map was overlaid on Google Maps to increase geographic context and user engagement.

Supporting Graphs Improvement: Graphs were updated to include attached numbers for better interpretability.

Usability Enhancements: The project received an 8/10 for usability with adjustments made for better zoom functionality in dense data areas and improved tooltip stability for better visibility and performance.

Comparison

Model Complexity & Comparison:
- We implemented and compared multiple models, including K-Means Clustering, Decision Trees, Random Forests, and Neural Networks, to capture different patterns in the data.

Evaluation & Optimization:
- We used multiple metrics (MSE, R^2) to evaluate model performance and tuned hyperparameters to optimize each model's accuracy.

User Testing:
- We conducted user testing to ensure the effectiveness of our visualizations, gathering direct feedback from participants.

Grading & Feedback:
- Our visualizations were graded on clarity, interpretability, engagement, and usability, receiving positive ratings overall.

Iterative Improvements:
- Based on feedback, we adjusted usability features (such as the zoom), the color gradient, overlaid the choropleth Map on Google Maps, and more.

How does the team's method compare to other methods:
- This approach stands out because it combines machine learning with interactive visualization, offering dynamic scenario exploration and more actionable insights than static tools.