

A photograph of a tree-lined street in New York City. The street is lined with red brick buildings on both sides. Large, leafy green trees are planted along the sidewalks, their branches arching over the street. Several cars are parked along the sides of the road. The overall scene is a typical urban street in NYC, characterized by its architecture and mature trees.

# Re-foresting NYC

Reid Majka  
Flatiron School Fall 2023


# Introduction & Objective

- Urban tree coverage is crucial to the success of urban environments<sup>1</sup>
  - \$60.1M in health impact reduction in NYC<sup>2</sup>
  - Stormwater mitigation
  - Urban heat effect
  - Increase property values
  - Community integration



# Introduction & Objective


- Using data from NYC Opendata, create a machine-learning model to predict whether a tree is in need of care and/or replacement
  - Provide NYC Parks department with an optimized model to prioritize trees needing care, minimizing resources to fix and increase the canopy across the city



**New York City Tree Map**  
Explore and Care For NYC's Urban Forest

**Map** | My Trees | Learn | Groups

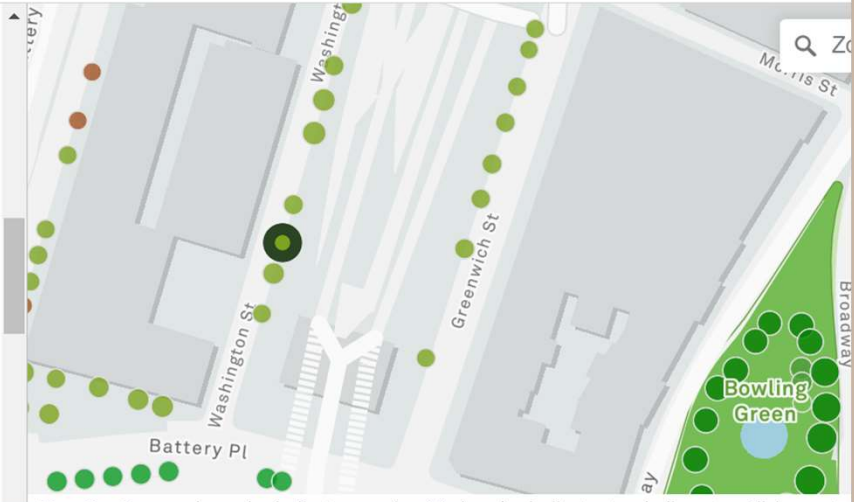
Text Size



© Cyclomedia

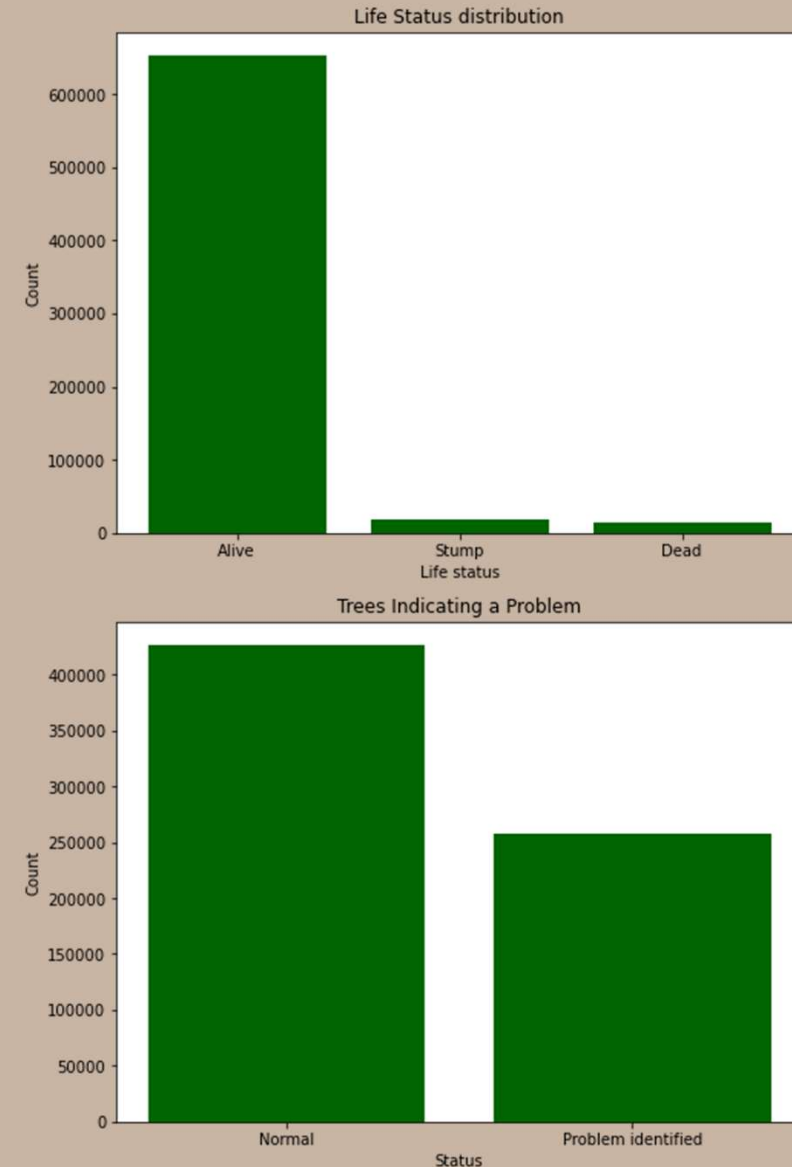
### Latest Inspection Results

Aug 29, 2023	A Forester has inspected this tree.
Aug 29, 2023	Work Order #21668659 created. A Parks contractor will prune the tree.



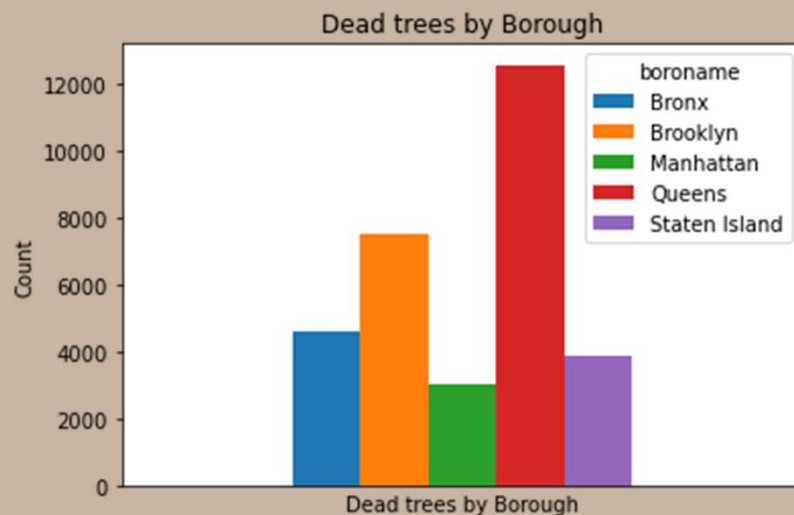
# Data Background

- Majority of trees are alive – this is good!
  - May cause some issues when creating a prediction model – to be discussed
- More % of all trees have “Problems” that could lead to future decay/death
  - Dataset range of health: No Problems, Problems (with a wide range of values, assuming all “problems” carry equal importance)
  - Determine trees with problems as a first-response to care



# Predicting if a Tree is Dead / Stump in plot

- Dataset binary classification: Alive or Dead/Stump
- Predictions on Life status resulted in 100% Accuracy
  - Dataset not diverse enough to use this target variable



Predictions on Life Status

	Dead	Alive
Actual Dead	25289	0
Actual Alive	0	521741
Predicted		

# Predicting Problems – Baseline model

- Factors included:
  - Community support (community board, steward(s), assessor)
  - Tree details (species, guards, trunk diameter, alive/dead status)
  - Geography (block location, district, borough, geo-coordinates)
- Predictions on Problems resulted in 98% precision, 88% Accuracy, and 70% recall
  - Aiming for higher recall, to minimize chance a tree with problems is classified as “normal”

Problem Classification Model

Actual Problems	No Problems	337,824	3,431
	Problems	62,127	143,648
		No Problems	Problems
		Predicted Problems	

# Optimized Logistic Regression Model

- Predictions on Problems resulted in 90% precision, 87% Accuracy, and 75% recall
- Fine-tuned model includes:
  - Increased iterations
  - Threshold of 35%
  - Lower regularization strength
- Critical factors: guards, root system, steward availability
- Decision tree modeling was assessed in process, and deemed unfit for success
- Balancing data set did not result in improvements

Problems Classification Model

Actual	No Problems	80,475	4,550
	Problems	13,186	38,547
		No Problems	Problems
		Predicted	

# Conclusion and next steps

- Prioritize re-planting at these sites with dead trees (or stumps)
  - Predictions on Health status vary – the model is good at predicting if the tree is alive, but more information is needed to predict if it is in sub-optimal health
  - Most importantly: just go outside and check! Call 311 if you see a tree in need.
- 



- Next Steps:
  - Partner with parks department to revisit trees in suboptimal health, and gather better data parameters in order to better tune the mode
  - Gather incremental data to increase performance:
    - Zoning district, air quality, demographics of surrounding population



# Questions

