NYC Food

Poverty Analysis

By Sehr Abrar

UN SDG Goal 1: No Poverty

Project Overview

- This project analyzes food poverty across New York City as part of **UN SDG Goal 1: No Poverty**.
- I combined multiple NYC datasets—**income distribution**, **census tract demographics**, **food assistance reports**, and **census block coordinates**—to understand where food insecurity is most severe.
- The goal was to **uncover structural patterns in poverty** and build **predictive models** that **estimate poverty rates** across neighborhoods to support targeted interventions.

Methodology

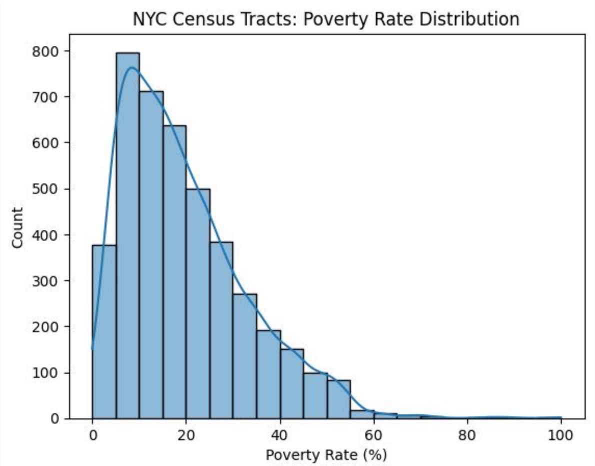
- I merged **four datasets** using census tract identifiers to create a unified dataset with income brackets, racial composition, transportation, occupations, poverty, child poverty, and food-assistance activity.
- I performed cleaning steps such as coercing numeric types (e.g., converting the Poverty column using pd.to_numeric).
- After preprocessing, I conducted exploratory data analysis, correlation studies, and borough-level comparisons.
- Finally, I trained and evaluated four models: Linear Regression, Random Forest, Gradient Boosting, and Hist Gradient Boosting.

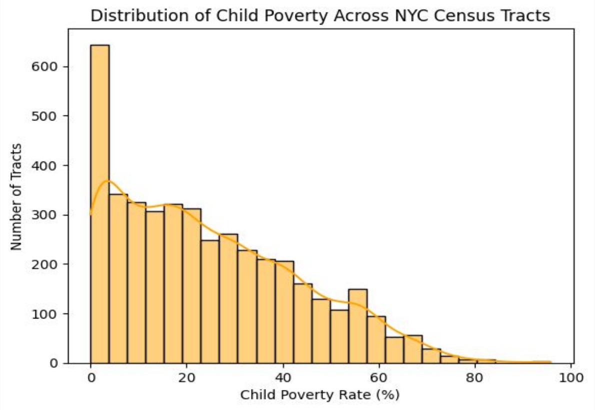
Exploratory Data Analysis Findings

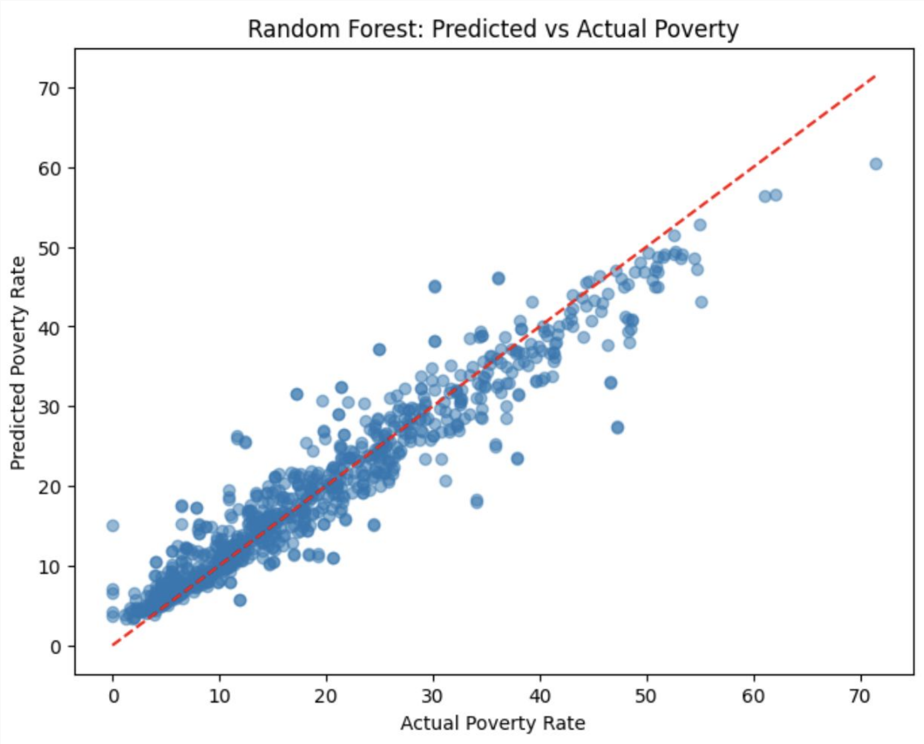
- Poverty rates are **low to moderate in most NYC neighborhoods**, but the **Bronx** consistently shows the **highest poverty levels**, while **Staten Island** shows the **lowest**.
- **Child poverty** is strongly correlated with **overall poverty (≈ 0.91)** and is more severe citywide.
- **Income** is inversely **relate**d to poverty (≈ –0.5), confirming inequality patterns.
- Racial and ethnic compositions differ sharply by borough, with the Bronx showing majority Black/Hispanic populations.
- **Food-assistance program usage spiked after 2020**, reflecting pandemic-driven food insecurity.
- These findings highlight structural disparities in poverty, income, demographics, and food access across the city.

Predictive Modeling

- I trained **four predictive models** to estimate neighborhood poverty rates: Linear Regression, Random Forest, Gradient Boosting, and Hist Gradient Boosting.
- **Linear Regression performed poorly ($R^2 \approx 0.57$)**, showing that poverty is too nonlinear for simple models.
- **Random Forest: $R^2 \approx 0.90$ (best performance)**
- **Hist Gradient Boosting: $R^2 \approx 0.86$**
- **Gradient Boosting: $R^2 \approx 0.75$**
- Feature importance indicated **strong influence** from occupation types, income brackets, transportation modes, and racial composition.
- **Random Forest** produced the tightest fit between actual vs. predicted poverty, making it the **most reliable model for policy planning**.


NYC Census Tracts: Poverty Rate Distribution
Count
Poverty Rate (%)


Distribution of Child Poverty Across NYC Census Tracts
Number of Tracts
Child Poverty Rate (%)


Random Forest: Predicted vs Actual Poverty
Predicted Poverty Rate
Actual Poverty Rate

Risk Management & Solution Development

One risk was **dataset inconsistency** due to **merging sources from different** years and **different geographic resolutions**. I mitigated this by aligning all data using census tract identifiers and carefully **converting non-numeric fields**.

Model bias was another risk, especially because **racial** and **socioeconomic variables are sensitive**. To address this, I **compared multiple model types** to see whether predictions were stable across algorithms.

Using the insights, I developed a solution direction that **prioritizes data-driven resource allocation**, where the **city could focus food-assistance expansion** in identified high-risk tracts. The predictive model also allows for simulations of future trends, supporting long-term planning for SDG 1.

Key Takeaways

- **Food poverty in NYC is unevenly distributed**, with the Bronx experiencing the highest vulnerability.
- **Child poverty is alarmingly high** and tightly tied to overall poverty.
- The **pandemic permanently increased reliance on food assistance systems**.
- **Tree-based ML models**, especially Random Forest, can **accurately estimate** neighborhood-level poverty.
- These insights support more equitable, targeted, and sustainable interventions aligned with UN SDG 1: No Poverty.