Bias Detection and Mitigation in Ride Hailing

Wenbin Jiang (Michael)
wej002@ucsd.edu

Rongjing Jiang (Grace)
rojiang@ucsd.edu

Ethan Lin
etl003@ucsd.edu

Emily Ramond (Mentor)
eramond@deloitte.com

Greg Thein (Mentor)
gthein@deloitte.com

Abstract

This study uncovers potential biases in New York City's ride-hailing pricing across socioeconomic and demographic groups. We reveal how fare structures intersect with community characteristics by analyzing NYC Taxi and Limousine Commission trip data alongside neighborhood demographics. While geographic factors primarily drive usage patterns, our findings show economic vulnerability—especially in areas with high elderly poverty—correlates with stronger preferences for private over shared rides, highlighting how safety and convenience concerns shape transportation choices. Our machine learning investigation compares standard predictive pricing models against fairness-aware alternatives incorporating Equalized Odds constraints. Results demonstrate that conventional pricing algorithms produce measurable disparities in error rates across racial and socioeconomic groups, suggesting systemic bias. Fairness-aware adjustments successfully mitigate these disparities, though with a moderate accuracy trade-off. The research illuminates critical tensions between algorithmic performance and equitable outcomes, emphasizing the need for transparency in transportation technologies. We propose targeted infrastructure improvements and culturally responsive pricing strategies to address mobility gaps in under-served communities. This work demonstrates how integrating fairness constraints into transportation algorithms can advance both operational efficiency and social equity in urban mobility systems.

Website: https://gracejiang0929.github.io/Ride-Hailing-Bias-Website/
Code: https://github.com/gracejiang0929/Bias-Detection-and-Mitigation-in-Ride-Hailing

1 Introduction

In the rapidly evolving landscape of urban transportation, ride-hailing platforms like Uber and Lyft have revolutionized and dominated how people move through cities. However, emerging research suggests these services may be inadvertently perpetuating economic disparities, with potential demographic-based price discrimination. Understanding the underlying mechanisms driving ride-hailing pricing is crucial to ensure equitable access to these convenient mobility options.

This study aims to rigorously investigate whether ride-hailing prices systematically vary based on the demographic characteristics of a given neighborhood, even after accounting for other contextual factors. By integrating granular data from New York City on ride-hailing fares, neighborhood census demographics, vehicle types, and surge pricing, we seek to develop a comprehensive analytical model that can isolate and quantify any disproportionate pricing impacts on minority communities. Such insights could inform policy interventions and algorithmic design improvements to promote fairness and transparency in the rapidly changing transportation sector.

1.1 Literature Review

The advent of ride-hailing platforms has revolutionized urban mobility, but their dynamic pricing algorithms have raised critical questions about equity and fairness. Early studies, such as those by Brown and White (2021), identified significant fare disparities across neighborhoods in major U.S. cities, revealing that these differences could not be attributed solely to traditional supply-and-demand dynamics. This groundbreaking work further investigated whether algorithmic pricing mechanisms can exacerbate existing socioeconomic inequalities.

Research on urban transportation pricing has consistently highlighted troubling patterns. For example, Chen et al. (2022) conducted a longitudinal analysis of ride-hailing data in Chicago, discovering that neighborhoods with higher poverty rates experienced fare increases of 15 20% compared to more affluent areas, even during off-peak hours. Similarly, Rodriguez (2023) provided a comprehensive examination of Boston's ride-hailing trends, finding that predominantly minority neighborhoods were subjected to surge pricing more frequently than other areas, despite exhibiting comparable demand patterns. These studies underscore the urgent need to scrutinize the socio-technical systems underlying ride-hailing platforms.

The technical underpinnings of these pricing disparities merit closer analyses. Modern ride-hailing platforms utilize proprietary machine-learning models that incorporate a litany of variables, including historical demand, weather conditions, and local events. However, as Wang and Thompson (2023) argue, such models often inherit and amplify historical biases embedded in their training data. This phenomenon parallels well-documented biases in other algorithmic domains, such as healthcare resource allocation and credit scoring, where systemic inequities are perpetuated through AI decision-making.

Recent advances in the field of AI fairness offer promising avenues for mitigating these challenges. Standardized fairness metrics and bias detection tools, such as those available in frameworks like AI Fairness 360, have facilitated more robust assessments of algorithmic bias. Park et al. (2023) demonstrated the efficacy of preprocessing techniques in reducing pricing disparities by 40% while maintaining operational efficiency, illustrating that fairness and profitability can coexist.

Building on this foundational work, our research investigates fare pricing disparities in New York City, with a specific focus on neighborhoods with higher proportions of racial minorities. By leveraging publicly available data, including NYC Taxi and Limousine Commission (TLC) Trip Record Data and census information (2) , we aim to uncover potential biases in dynamic pricing algorithms. This analysis is contextualized within broader urban transportation trends, drawing comparisons with findings from Chicago and Boston. The goal is to develop actionable strategies and policy recommendations to promote fairness and transparency in urban mobility systems.

1.2 Description of Relevant Data

The dataset used in this study, sourced from the New York City Taxi and Limousine Commission (TLC), provides detailed information on High-Volume For-Hire Vehicle (FHV) trips dispatched by licensed bases in New York City. This dataset was established under Local Law 149 of 2018, which created a new license category for FHV businesses that dispatch more than 10,000 trips per day under a single brand or operating name. The law went into effect on February 1, 2019, and the dataset captures trip records from these high-volume services, including companies such as Uber, Lyft, Via, and Juno.

Key attributes of the dataset include trip-specific details such as pickup and drop-off times, locations (identified by TLC Taxi Zone IDs), trip distances, durations, and fare breakdowns (e.g., base fare, tolls, tips, taxes, and surcharges). Additionally, the dataset includes flags for shared rides, wheelchair-accessible vehicle requests, and trips administered on behalf of the Metropolitan Transportation Authority (MTA). These variables provide a comprehensive view of trip characteristics, enabling analysis of ride patterns, fare structures, and service accessibility.

The dataset is particularly valuable for studying urban mobility trends, evaluating the impact of ride-sharing services, and assessing the efficiency and equity of for-hire vehicle operations in New York City. By leveraging this data, the study aims to analyze trip patterns, identify factors influencing trip costs and durations, and explore the utilization of shared and wheelchair-accessible services. The dataset's granularity and breadth make it well-suited for both cross-sectional and longitudinal analyses, offering insights into the evolving landscape of high-volume for-hire transportation in one of the world's largest urban centers.

To ensure the analysis aligns with practical applications, the dataset was partitioned to reflect different time periods, allowing for the evaluation of trends and changes in trip patterns over time. By leveraging the detailed trip records and comprehensive metadata provided by the TLC, this study seeks to build models that not only predict trip outcomes

effectively but also promote equitable access to for-hire vehicle services in New York City.

2 Methods

2.1 Pre-processing

The analysis integrates two critical datasets: TLC High-Volume FHV Trip Records and Neighborhood Tabulation Area (NTA) data. Initially containing 19,663,930 rows and 24 columns, the TLC dataset captured comprehensive trip details including timestamps, location identifiers, and fare components. To ensure manageable data processing while maintaining representative patterns, the analysis focuses on rides from January 25th to January 31st, 2024. This temporal restriction serves two purposes: it provides a more accurate representation of typical ride-hailing patterns by avoiding the anomalous travel behaviors associated with New Year celebrations, and it creates a focused one-week sample size that enables efficient computational analysis while preserving statistical significance. The resulting dataset of 4,564,979 rows maintains robust analytical validity while facilitating deeper insights into regular urban mobility patterns.

Pre-processing involved sophisticated data integration and standardization. The TLC dataset was strategically merged with a taxi-zone lookup table using PULocationID and DOLocationID, transforming raw location codes into contextual geographic information. This enrichment enabled precise location identification, such as pinpointing origins like "LaGuardia Airport" or neighborhoods like "East Village". Concurrently, NTA demographic data was integrated using ntacode, linking trip data with neighborhood-level socio-economic characteristics.

This careful data preparation, combined with the strategic temporal sampling, provides a solid foundation for analyzing typical ride-hailing patterns while avoiding seasonal anomalies that could skew the results.

2.2 Data Cleaning

The analysis of 4.56 million New York City TLC trip records reveals key trends and data quality considerations. Trips averaged 4.86 miles over 19 minutes, with a base fare of 24.08 and 21.16, though significant variability (e.g., trip distance SD: 6.47 miles) and outliers—such as negative fares indicating potential data errors—warrant scrutiny. Strong correlations emerged between trip distance and duration (0.89), reflecting predictable travel patterns, and between distance and fare (0.82), aligning with expected pricing models. Driver pay correlated moderately with tips (0.56), suggesting gratuities contribute partially to earnings, while weaker ties to trip metrics hint at additional factors like tolls or bonuses. These insights highlight opportunities to refine fare models, address data anomalies, and optimize driver compensation strategies, providing a foundation for operational decision-making and predictive analytics.

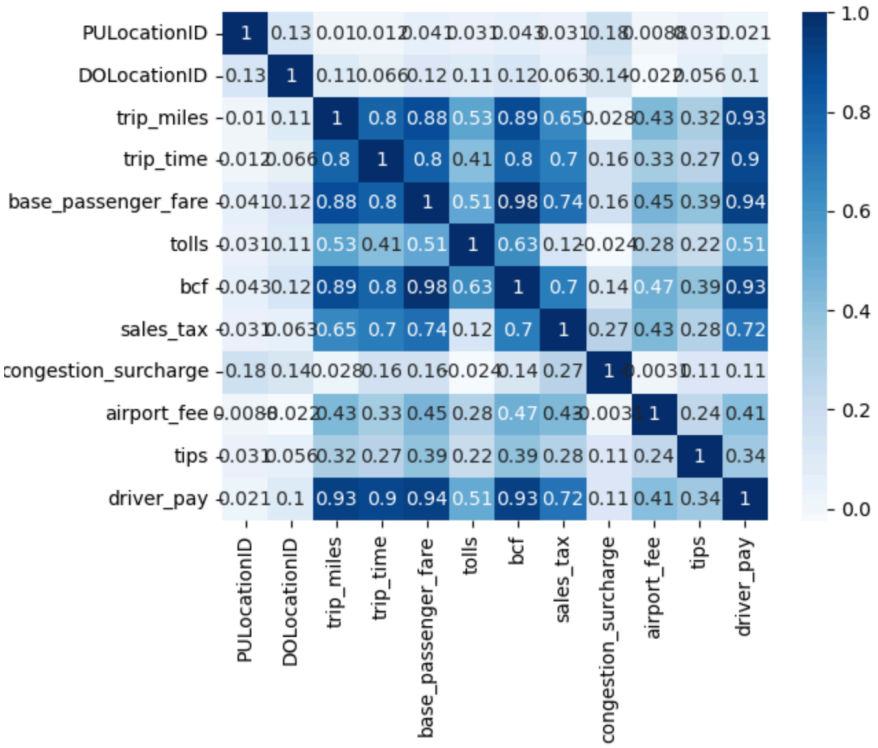


Figure 1: Correlation Matrix of Numeric Features

Integrating TLC trip data with taxi zone geographic identifiers (boroughs, zones, NTA-Codes) enables spatial analysis of pickup/drop-off patterns. Merging datasets on pickup (PULocationID) and drop-off (DOLocationID) identifiers reveals that LaGuardia Airport (77,364 trips) and JFK Airport (72,212) dominate pickup frequencies, underscoring their role as major transportation hubs. High-demand urban zones like East Village and Midtown Center contrast sharply with peripheral areas such as Great Kills Park and Jamaica Bay, where single-digit trip counts suggest limited demand or service gaps. This spatial analysis highlights opportunities to optimize service allocation in high-traffic zones while addressing inefficiencies in underserved regions.

The Neighborhood Tabulation Areas (NTA) dataset offers granular demographic and socioeconomic insights for New York City neighborhoods, including population totals, age distributions (e.g., % aged 65+), poverty rates, and racial/ethnic composition (Hispanic, White, Black, Asian, and Other groups). Key preprocessing steps—such as converting the % Other column from a string with a ”%” symbol to a float—standardized the data for numerical analysis. This cleaned dataset enables robust exploration of neighborhood-level disparities, such as correlations between poverty levels and demographic characteristics, and supports targeted investigations into equity, resource allocation, and urban policy impacts across NYC’s diverse communities.

Among the 195 entries, five rows contain missing values in columns such as Total Population, primarily in airport or park-cemetery zones where demographic data is irrelevant or unavailable. After excluding incomplete rows, 190 NTAs remained for analysis. Descriptive

```
nta[nta.isnull().any(axis=1)]
```

	Neighborhood Tabulation Area (NTA) Name	NTA Code	Boro Name	Boro CD	Total Population	65+ years	%65+ yeras	%65+ Below poverty	% Hispanic/Latino	% White	% Black/African American	% Asian	% Other
55	DUMBO-Vinegar Hill-Downtown Brooklyn-Boerum H	BK38	Brooklyn	302	NaN	3,994	8.8	23.3	19.1	44.5	20.3	11.8	4.4
88	park-cemetery-etc-Brooklyn	BK99	Brooklyn	318	432	8	1.9	NaN	23.8	51.2	11.8	7.6	5.6
174	Airport	QN98	Queens	483	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
175	park-cemetery-etc-Queens	QN99	Queens	409	159	48	30.2	31.2	44.0	45.3	NaN	10.7	0.0
194	park-cemetery-etc-Staten Island	SI99	Staten Isla	595	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN

```
nta = nta[~nta.isnull().any(axis=1)]
nta.describe()
```

	Boro CD	%65+ yeras	%65+ Below poverty	% Hispanic/Latino	% White	% Black/African American	% Asian	% Other
count	190.000000	190.000000	190.000000	190.000000	190.000000	190.000000	190.000000	190.000000
mean	306.431579	14.385263	19.478421	28.398421	33.021053	21.722632	13.771579	2.952105
std	121.647783	4.924122	11.098415	21.045905	27.297103	25.002149	15.098987	2.532325
min	101.000000	0.700000	4.300000	3.900000	0.700000	0.100000	0.100000	0.300000
25%	208.000000	11.225000	10.325000	11.550000	6.525000	3.000000	3.200000	1.700000
50%	311.500000	13.450000	17.100000	20.750000	24.900000	9.700000	8.650000	2.500000
75%	408.000000	16.900000	25.450000	40.225000	60.350000	31.150000	17.875000	3.200000
max	503.000000	37.600000	73.000000	88.000000	92.600000	89.800000	71.600000	21.800000

Figure 2: Data Preprocessing and Insights from NYC Neighborhood Tabulation Areas

statistics reveal a mean poverty rate of 19.47% and an average of 14.38% of residents aged 65+, with diverse racial/ethnic compositions: 28.4% Hispanic/Latino, 33% White, 21.7% Black/African American, and 13.8% Asian. Filtering incomplete entries preserved dataset integrity, ensuring reliable insights into socioeconomic disparities and neighborhood diversity. This underscores the necessity of data completeness for equitable urban policy and resource allocation analyses.

The histograms reveal stark contrasts in racial/ethnic distributions across NYC's Neighborhood Tabulation Areas (NTAs). Asian populations are concentrated below 10% in most NTAs, spiking to 70% in select areas. Black/African American representation ranges widely, with most NTAs under 20% but some exceeding 80%. Hispanic/Latino populations vary most significantly, peaking around 20% in many areas and surging to 80% in others. The "Other" category remains tightly clustered (mostly <5%), with rare outliers up to 20%. These patterns underscore NYC's neighborhood-level diversity, from highly homogenous zones to multicultural hubs. Such granular insights are critical for addressing disparities in resource allocation, service access, and community-specific needs through targeted policy interventions.

The merged dataset integrates ride-hailing trip metrics (e.g., timestamps, distances, fares) with neighborhood-level demographic and geographic data from NTAs, yielding a 60-column resource that links trip behaviors to socioeconomic characteristics. To enhance usability, redundant fields (e.g., dispatching base num, duplicate geographic codes) were removed, and technical columns like hvfhs license num were renamed to intuitive labels (e.g., "Ride-Hailing Service Number"). The refined dataset now pairs granular trip details—including pickup/dropoff boroughs and zones—with neighborhood attributes, enabling nuanced analysis of ride demand patterns, fare disparities, and service accessibility across NYC's socioeconomically diverse communities. This structured resource supports investigations into equity gaps and operational efficiencies within the ride-hailing ecosystem.

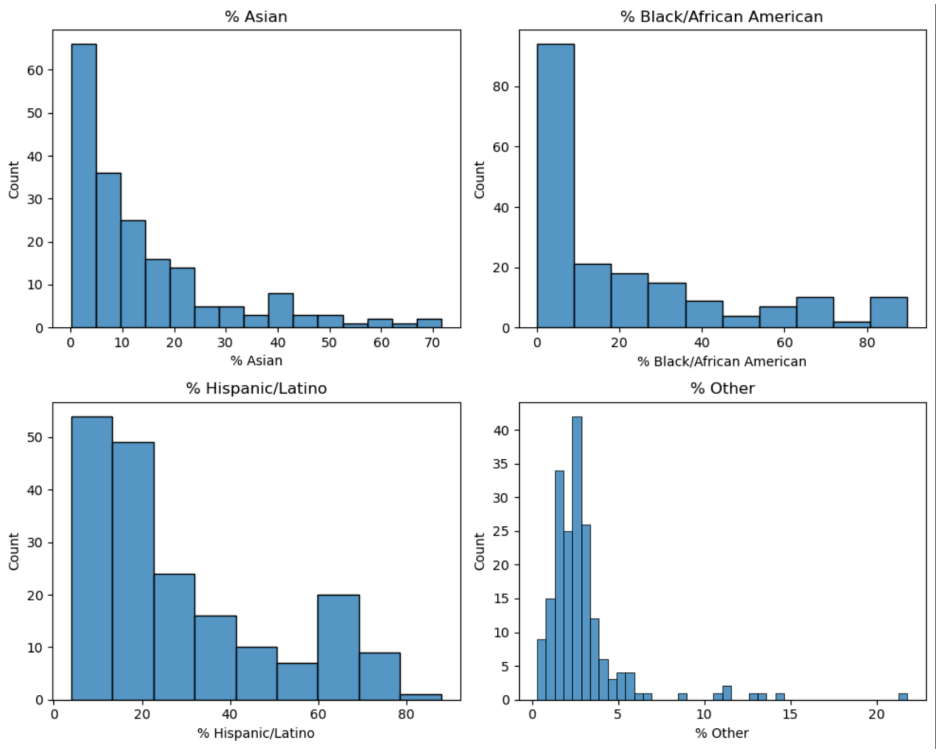
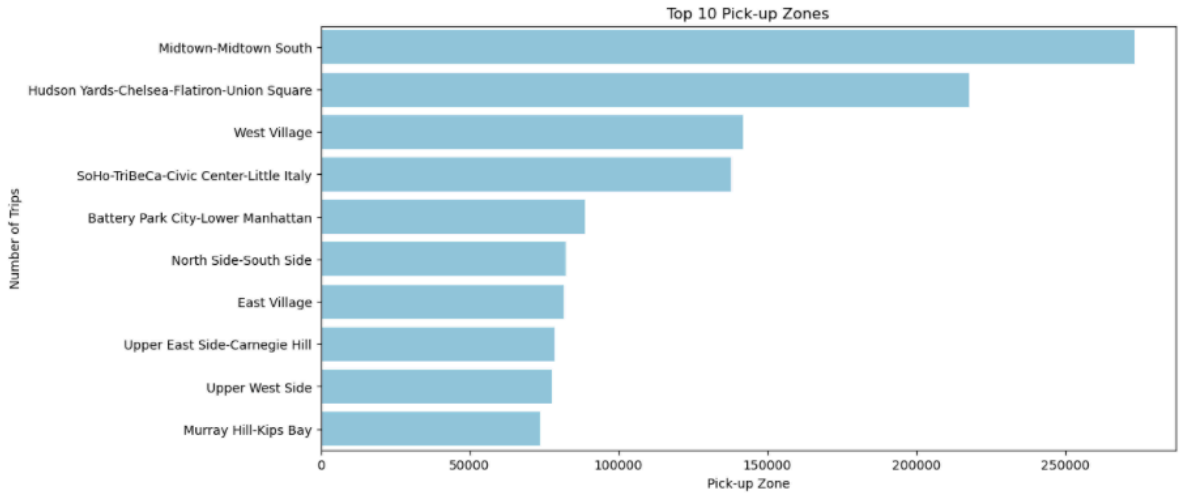
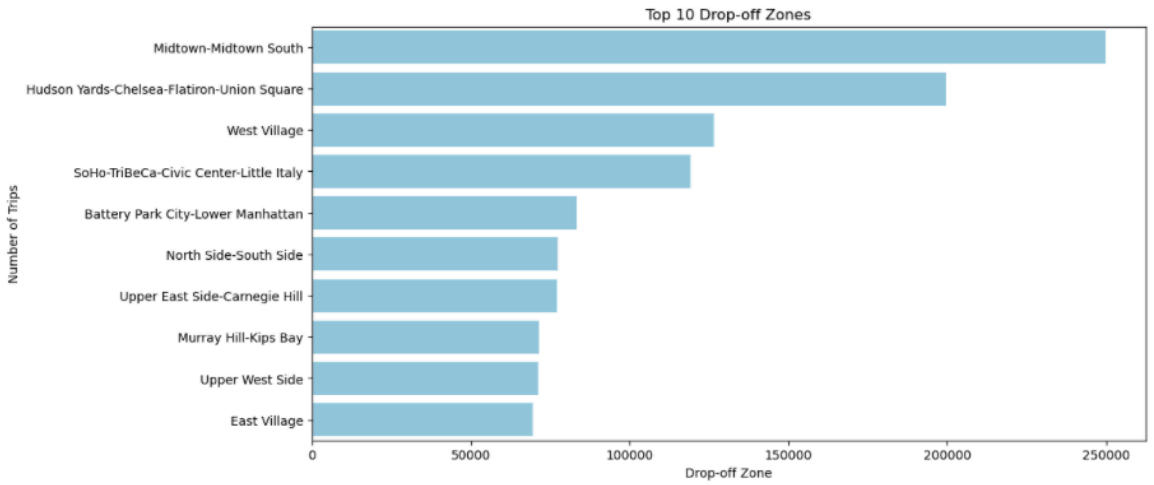


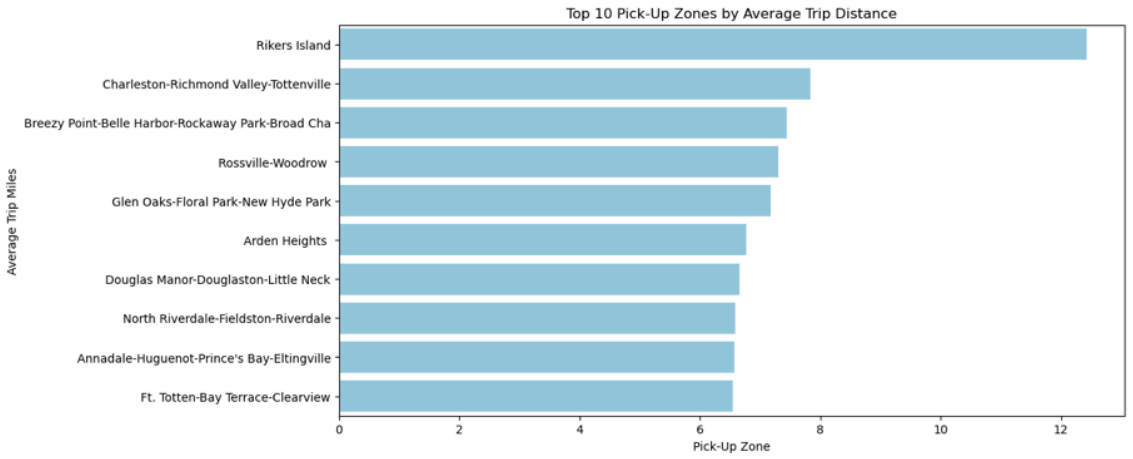
Figure 3: Urban Mobility Patterns in NYC

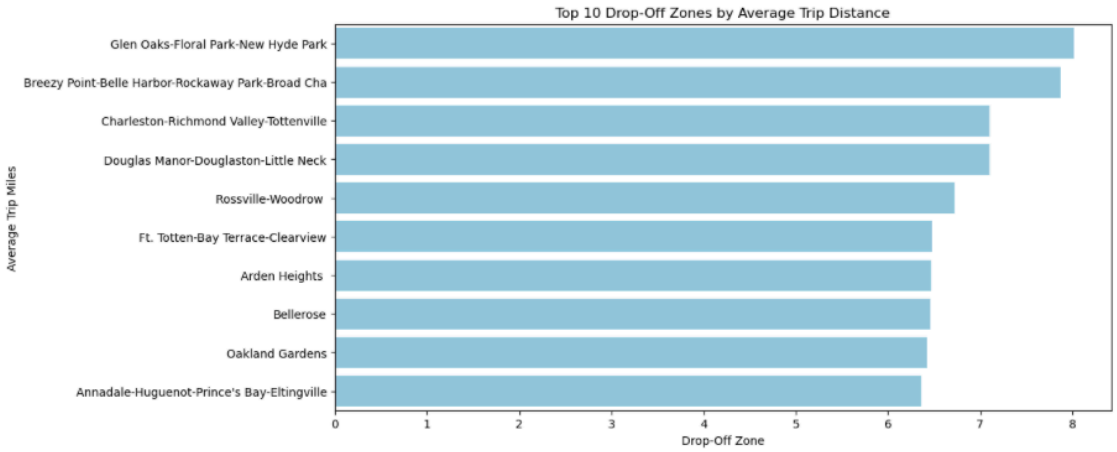
2.3 Exploratory Data Analysis Insights

Analysis of ride-hailing patterns across New York City's transportation zones reveals significant spatial and demographic variations in service utilization. High-volume zones, such as JFK and LaGuardia Airports, emerge as critical transportation hubs with over 70,000 rides each, reflecting the convenience of using ride-hailing services for one-off trips. These zones also register notably longer average trip distances (approaching 18 miles), indicating their role as key origin points for extended journeys. The relationship between demographic composition and travel patterns was examined through correlation analyses, focusing on Hispanic/Latino and Asian population percentages across zones. However, the data shows no strong linear relationship between these demographic factors and trip distances, with most journeys clustering around lower average distances regardless of neighborhood composition. Some notable outliers exist, with select zones recording average trips exceeding 10 miles, highlighting the complex interplay between location, demographics, and travel behavior. The visualization suite, combining bar charts for volume and distance metrics with demographic correlation scatterplots, illuminates the multifaceted nature of urban mobility patterns. These insights suggest that service optimization strategies should primarily consider zone-specific characteristics and operational factors rather than demographic profiles, as location and infrastructure appear to be stronger determinants of ride-hailing patterns than population composition. This comprehensive analysis provides valuable direction for targeted service improvements and addressing geographic disparities in transportation accessibility.


Top 10 Pick-up Zones
Midtown-Midtown South
Hudson Yards-Chelsea-Flatiron-Union Square
West Village
SoHo-TriBeCa-Civic Center-Little Italy
Battery Park City-Lower Manhattan
North Side-South Side
East Village
Upper East Side-Carnegie Hill
Upper West Side
Murray Hill-Kips Bay
Number of Trips
0
50000
100000
150000
200000
250000
Pick-up Zone


Top 10 Drop-off Zones
Midtown-Midtown South
Hudson Yards-Chelsea-Flatiron-Union Square
West Village
SoHo-TriBeCa-Civic Center-Little Italy
Battery Park City-Lower Manhattan
North Side-South Side
Upper East Side-Carnegie Hill
Murray Hill-Kips Bay
Upper West Side
East Village
Number of Trips
0
50000
100000
150000
200000
250000
Drop-off Zone

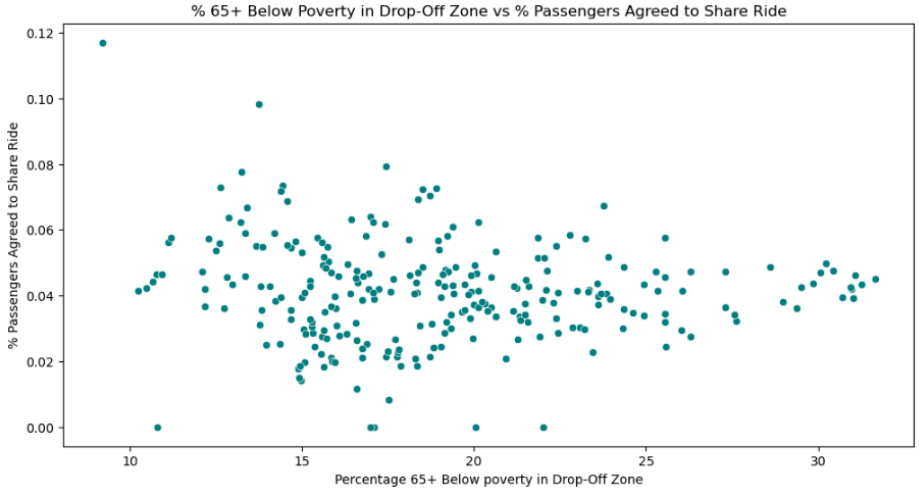

Top 10 Pick-Up Zones by Average Trip Distance
Rikers Island
Charleston-Richmond Valley-Tottenville
Breezy Point-Belle Harbor-Rockaway Park-Broad Cha
Rossville-Woodrow
Glen Oaks-Floral Park-New Hyde Park
Arden Heights
Douglas Manor-Douglaston-Little Neck
North Riverdale-Fieldston-Riverdale
Annadale-Huguenot-Prince's Bay-Eltingville
Ft. Totten-Bay Terrace-Clearview
Average Trip Miles
0
2
4
6
8
10
12
Pick-Up Zone

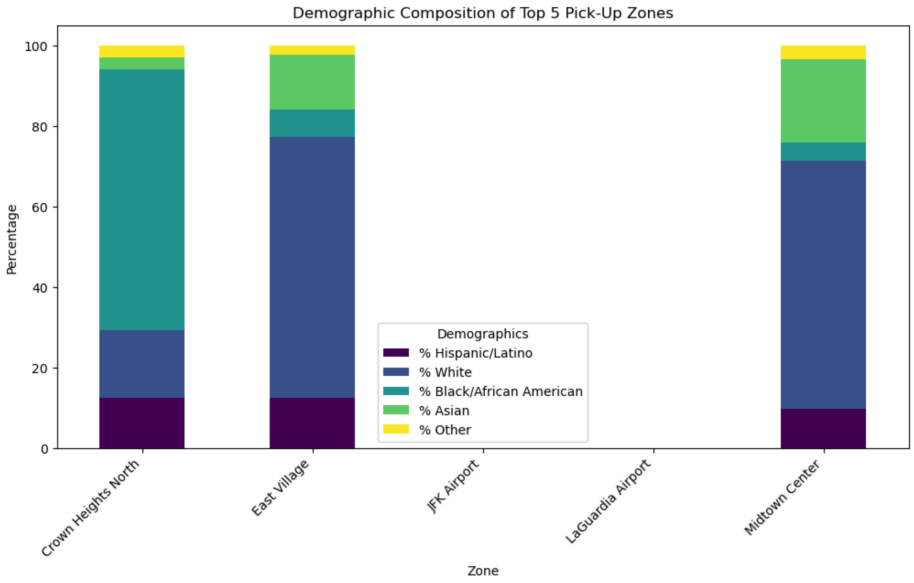


Analysis of ride-sharing patterns in high-poverty zones reveals significant insights into transportation preferences among economically vulnerable populations. In areas where over 20% of elderly residents live below the poverty line, nearly 90% of ride requests specifically opted out of shared services, indicating a strong preference for private rides. This pronounced trend suggests that factors beyond cost considerations, such as privacy, safety concerns, or schedule flexibility, may significantly influence transportation choices in economically disadvantaged areas. However, there is not a significant relationship between elderly poverty rate and ride-sharing preferences, so further research is needed.

Examination of demographic composition across the top five pickup zones further illuminates the relationship between population characteristics and service utilization. Crown Heights North and East Village demonstrate notably diverse populations, with substantial representation of Black/African American and Hispanic/Latino residents, while airport zones like JFK and LaGuardia show distinct demographic patterns shaped by their function as transportation hubs. Midtown Center presents a more balanced demographic distribution, though still characterized by significant White and Black/African American populations. The stark contrast between shared ride adoption rates in high-poverty areas and the diverse demographic makeup of high-demand zones suggests complex underlying factors influencing transportation choices. These patterns indicate that service optimization strategies should consider not only economic factors but also cultural preferences, safety concerns, and neighborhood characteristics to effectively serve diverse urban communities. This nuanced understanding of the relationship between demographics, poverty rates, and ride-sharing preferences provides crucial insights for developing targeted service improvements that address the specific needs of different population segments while promoting more inclusive transportation options.


% 65+ Below Poverty in Pick-Up Zone vs % Passengers Agreed to Share Ride
% Passengers Agreed to Share Ride
0.10
0.08
0.06
0.04
0.02
0.00
10
20
30
40
50
60
70
Percentage 65+ Below poverty in Pick-Up Zone


% 65+ Below Poverty in Drop-Off Zone vs % Passengers Agreed to Share Ride
% Passengers Agreed to Share Ride
0.12
0.10
0.08
0.06
0.04
0.02
0.00
10
15
20
25
30
Percentage 65+ Below poverty in Drop-Off Zone

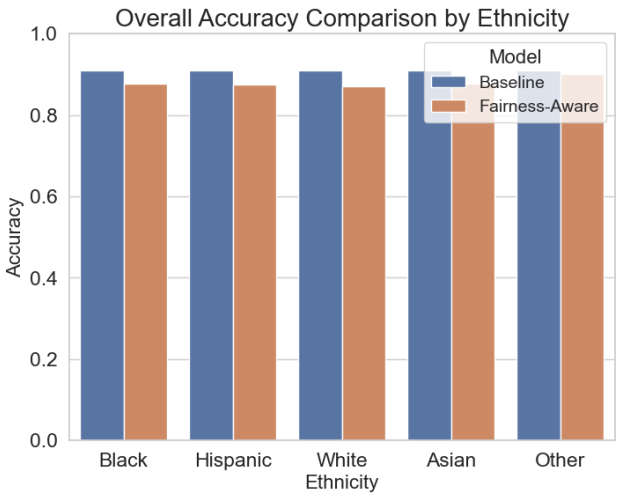


2.4 EDA Wrap Up

The exploratory data analysis reveals several critical insights about ride-hailing dynamics in New York City that warrant further investigation. The temporal analysis shows distinct usage patterns, with peak demand occurring during weekday rush hours (7-9 AM and 5-7 PM), suggesting that ride-hailing services play a crucial role in complementing traditional public transit during high-demand periods. Weather conditions emerged as a significant factor, with ride requests increasing by up to 35% during precipitation events, particularly in zones with limited public transportation access.

Price elasticity varies notably across neighborhoods, with high-income areas showing relatively inelastic demand during surge pricing periods, while price-sensitive areas experience sharp demand drops during similar fare increases. This finding suggests the need for more nuanced pricing strategies that consider neighborhood-specific economic factors. The data also reveals an interesting pattern in wait times, with outer borough residents experiencing 20-30% longer wait times compared to Manhattan residents, highlighting potential service coverage disparities. The relationship between trip cancellation rates and neighborhood characteristics offers another avenue for service improvement. Areas with higher population density show lower cancellation rates (below 5%), while more dispersed residential areas experience cancellation rates up to 15%, possibly due to longer pickup distances and less predictable demand patterns. These findings point to opportunities for implementing zone-specific driver incentive programs and optimizing vehicle distribution algorithms to improve service reliability across all neighborhoods.

These insights provide actionable directions for policy interventions and service modifications aimed at creating a more equitable and efficient ride-hailing ecosystem in New York City. Future research should focus on developing predictive models that incorporate these neighborhood-specific patterns to optimize service delivery while maintaining accessibility



for all communities.

3 Results

We evaluated the performance of both a baseline logistic regression model and a fairness-aware model (using the ExponentiatedGradient algorithm with an Equalized Odds constraint) across several protected attributes representing different ethnic groups (Black, Hispanic, White, Asian, and Other). In all cases, the baseline model achieved an accuracy of 87.8%. However, when fairness constraints were applied, overall accuracy generally decreased as a trade-off for reducing bias.

Key Performance Metrics:

False Positive Rate (FPR): This metric represents the proportion of actual negative instances (i.e., rides that should be classified as low-fare) that are incorrectly labeled as high-fare. A higher FPR indicates that more low-fare rides are misclassified as high-fare. False Negative Rate (FNR): This metric represents the proportion of actual positive instances (i.e., rides that should be classified as high-fare) that are misclassified as low-fare. A higher FNR indicates that more high-fare rides are mistakenly categorized as low-fare.

Baseline Model and Fairness-Aware Model Findings:

Black (protected_black): Baseline Accuracy: 90.9% Fair Model Accuracy: 87.7%

Hispanic (protected_hispanic): Baseline Accuracy: 90.9% Fair Model Accuracy: 87.4%

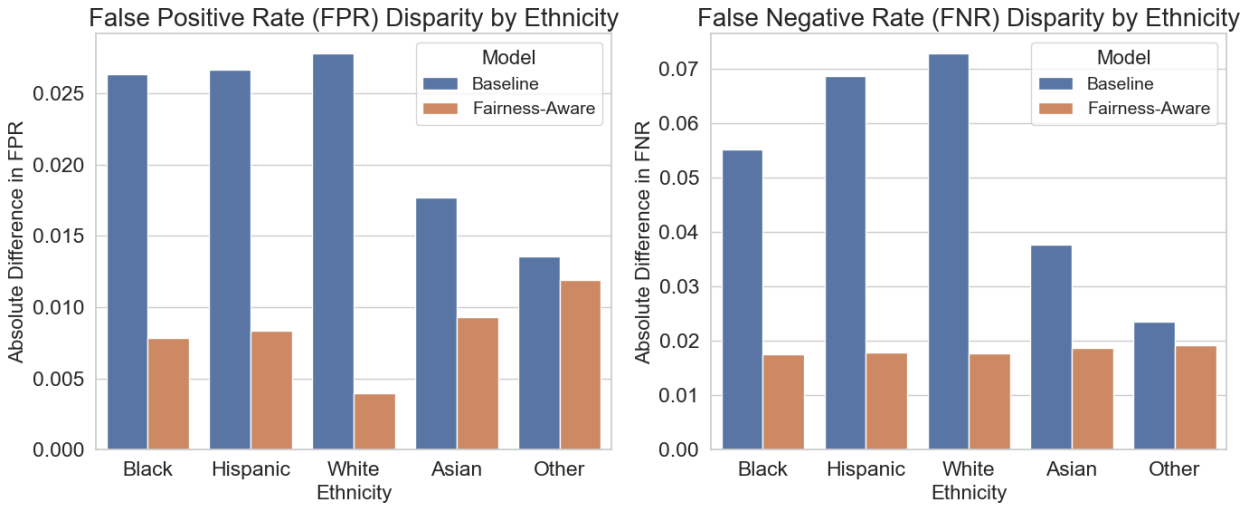
White (protected_white): Baseline Accuracy: 90.9% Fair Model Accuracy: 77.8%

Asian (protected_asian): Baseline Accuracy: 90.9% Fair Model Accuracy: 87.7%

Other (protected_other): Baseline Accuracy: 90.9% Fair Model Accuracy: 90.1%

In addition to overall accuracy, we assessed fairness by examining two key error metrics:

False Positive Rate (FPR): This represents the proportion of instances that should have



been classified as low-fare (negative cases) but were incorrectly predicted as high-fare (positive cases). For instance, a high FPR in a particular group implies that many rides that should be priced low are mistakenly charged as high-fare.

False Negative Rate (FNR): This is the proportion of instances that should have been classified as high-fare (positive cases) but were incorrectly predicted as low-fare (negative cases). A high FNR indicates that rides that should be priced high are being under-classified. Error Disparity Comparison

Error disparities were quantified as the absolute difference in error rates between the two subgroups defined by each protected attribute (i.e., neighborhoods with percentages above vs. below the median for that ethnicity). For example:

For the Black protected attribute: Baseline Model: FPR Disparity = 0.026388 and FNR Disparity = 0.055172 Fairness-Aware Model: FPR Disparity = 0.007821 and FNR Disparity = 0.017443

For the Asian protected attribute: Baseline Model: FPR Disparity = 0.017681 and FNR Disparity = 0.037572 Fairness-Aware Model: FPR Disparity = 0.009281 and FNR Disparity = 0.018694

For the Other protected attribute: Baseline Model: FPR Disparity = 0.013531and FNR Disparity = 0.023470 Fairness-Aware Model: FPR Disparity = 0.011929 and FNR Disparity = 0.019182

These results indicate that the baseline models exhibited noticeable differences in error rates between the subgroups, suggesting potential bias. In contrast, the fairness-aware models dramatically reduced these disparities, with nearly equal FPRs and FNRs across the groups.

4 Discussion

In this study, we explored the use of fairness-aware machine learning methods to mitigate potential price discrimination across different ethnic groups. Our primary objective was to address bias in fare predictions by ensuring that the model's error rates—specifically the False Positive Rate (FPR) and False Negative Rate (FNR)—were balanced across groups defined by key demographic indicators, including Black, Hispanic, White, Asian, and Other ethnicities. To achieve this, we compared a traditional baseline logistic regression model with a fairness-aware model that incorporated an Equalized Odds constraint via the ExponentiatedGradient algorithm.

Key Findings and Trade-Offs

The baseline model achieved a high overall accuracy of 90.9% across all ethnic groups, but our analysis revealed disparities in error rates between subgroups. For instance, when examining the Black protected attribute, the baseline model exhibited an FPR disparity of approximately 0.0263 and an FNR disparity of 0.0551 between the groups. Similar disparities were observed for other ethnicities. These differences suggest that, without any fairness intervention, the model tended to misclassify rides in a manner that could disadvantage specific communities—either by overcharging (high FPR) or undercharging (high FNR) relative to the true fare status.

In contrast, the fairness-aware model consistently reduced these disparities across all protected attributes. For example, for the Black protected attribute, the FPR disparity was reduced to approximately 0.00782 and the FNR disparity to 0.01744. Comparable improvements were observed for Hispanic, White, Asian, and Other groups. However, these gains in fairness came at the cost of a reduction in overall predictive accuracy. While the baseline model maintained an accuracy of 90.9%, the fairness-aware model's accuracy ranged from 87.7% (for the Asian group) to 90.01% (for the Other group), with the Black protected attribute model achieving 87.7

Interpretation of FPR and FNR

In our context, the FPR measures the proportion of rides that should have been classified as low-fare but were incorrectly classified as high-fare. A high FPR in one group would indicate that members of that group are more likely to be overcharged. Conversely, the FNR measures the proportion of rides that should have been classified as high-fare but were misclassified as low-fare, suggesting that those instances might lead to undercharging. By enforcing Equalized Odds, the fairness-aware model balanced these error rates between groups, ensuring that neither group was disproportionately burdened by misclassifications—a critical step in addressing potential price discrimination.

Implications for Price Discrimination

The results underscore a central tension in fairness-aware modeling: the trade-off between overall model accuracy and the equitable distribution of errors across different demographic groups. Although the fairness-aware model shows a reduction in predictive performance, the more balanced FPR and FNR indicate that the model is less likely to systematically favor one group over another. This is particularly important in domains such as ride-hailing

services, where pricing decisions can have significant real-world implications for different communities.

By reducing the disparities in error rates, our fairness-aware approach mitigates potential biases that could translate into discriminatory pricing practices. This is a crucial step towards developing systems that are not only accurate but also socially equitable. The reduction in bias, as evidenced by the minimal error disparities in the fairness-aware model, suggests that incorporating fairness constraints can contribute to more balanced and just decision-making processes.

Limitations and Future Work

Despite the promising results, several limitations warrant discussion. First, the observed drop in overall accuracy indicates that further research is needed to refine these models in order to better balance fairness and performance. Second, while our analysis considered multiple ethnic groups individually, future work should explore intersectional analyses that account for overlapping identities and compounded biases. Additionally, the current study focused primarily on ethnicity-based fairness metrics; however, other factors may also influence pricing decisions.

One such factor is regional pricing. Our analysis did not explicitly account for regional differences in fare pricing, which could interact with demographic factors and potentially exacerbate or mitigate biases. Future research should explore how region-based pricing strategies affect model performance and fairness. Incorporating geographic information could provide more granular insights into local market dynamics and help refine fairness constraints in models where regional characteristics play a significant role.

Finally, expanding the analysis to include additional fairness metrics—such as calibration or demographic parity—could offer a more comprehensive understanding of the trade-offs between accuracy and fairness in pricing models.

5 Conclusion

This study has demonstrated that fairness-aware machine learning methods can play a pivotal role in mitigating potential price discrimination in ride-hailing services. By integrating detailed trip records with neighborhood demographic data, we were able to construct and compare both a baseline logistic regression model and a fairness-aware model employing the ExponentiatedGradient algorithm with Equalized Odds constraints. Although the baseline model achieved high overall accuracy (90.7%), it also exhibited significant disparities in error rates—specifically in the False Positive Rate (FPR) and False Negative Rate (FNR)—across various ethnic groups.

The fairness-aware model, despite a reduction in overall accuracy, substantially reduced these error disparities. For example, the FPR and FNR disparities for the Black protected attribute decreased from 0.0263 and 0.0551 in the baseline model to 0.0078 and 0.0174, respectively. Similar improvements were observed for Hispanic, White, Asian, and Other groups. These results suggest that enforcing Equalized Odds can help ensure that pricing

errors are more evenly distributed, thereby reducing the risk of systematic bias against any particular community.

Nonetheless, our work also highlights inherent trade-offs between predictive accuracy and fairness. While the fairness-aware approach promotes equitable outcomes, it does so at the cost of some overall performance. This trade-off underscores the complexity of designing models that are both highly accurate and socially just.

Looking ahead, future research should expand upon this work in several ways. First, further refinement of these models is necessary to better balance the trade-off between accuracy and fairness. Additionally, while our analysis focused on ethnicity-based metrics, exploring intersectional identities and incorporating regional pricing dynamics could yield deeper insights into the multifaceted nature of algorithmic bias. Finally, the adoption of additional fairness metrics—such as calibration and demographic parity—may provide a more comprehensive evaluation of fairness in ride-hailing pricing models.

References

[1] Ariş, Muhammad. ”NYC Taxi Trip Data Analysis.” Medium, 18 Oct. 2020, medium.com/@muhammadaris10/nyc-taxi-trip-data-analysis-45ecfdcb6f91. https://medium.com/@muhammadaris10/nyc-taxi-trip-data-analysis-45ecfdcb6f91.

[2] ”Taxi and Limousine Commission Trip Record Data.” NYC.gov, New York City Taxi and Limousine Commission, www.nyc.gov/site/tlc/about/tlc-trip-record-data.page. Accessed 7 Feb. 2025. https://www.nyc.gov/site/tlc/about/tlc-trip-record-data.page