

Disparate Impact of Artificial Intelligence Bias in Ridehailing Economy's Price Discrimination Algorithms

Akshat Pandey

George Washington University
Department of Computer Science
apandey0123@gwu.edu

Aylin Caliskan

George Washington University
Department of Computer Science
Institute for Data, Democracy & Politics
aylin@gwu.edu

ABSTRACT

Ridehailing applications that collect mobility data from individuals to inform smart city planning predict each trip's fare pricing with automated algorithms that rely on artificial intelligence (AI). This type of AI algorithm, namely a price discrimination algorithm, is widely used in the industry's black box systems for dynamic individualized pricing. Lacking transparency, studying such AI systems for fairness and disparate impact has not been possible without access to data used in generating the outcomes of price discrimination algorithms. Recently, in an effort to enhance transparency in city planning, the city of Chicago regulation mandated that transportation providers publish anonymized data on ridehailing. As a result, we present the first large-scale measurement of the disparate impact of price discrimination algorithms used by ridehailing applications.

The application of random effects models from the meta-analysis literature combines the city-level effects of AI bias on fare pricing from census tract attributes, aggregated from the American Community Survey. An analysis of 100 million ridehailing samples from the city of Chicago indicates a significant disparate impact in fare pricing of neighborhoods due to AI bias learned from ridehailing utilization patterns associated with demographic attributes. Neighborhoods with larger non-white populations, higher poverty levels, younger residents, and high education levels are significantly associated with higher fare prices, with combined effect sizes, measured in Cohen's d , of -0.32 , -0.28 , 0.69 , and 0.24 for each demographic, respectively. Further, our methods hold promise for identifying and addressing the sources of disparate impact in AI algorithms learning from datasets that contain U.S. geolocations.

CCS CONCEPTS

• General and reference → Empirical studies; • Computing methodologies → Supervised learning by regression; • Applied computing → Decision analysis.

KEYWORDS

AI ethics, algorithmic bias, disparate impact, geolocation, prediction, price discrimination

1 INTRODUCTION

Ridehailing applications are artificial intelligence (AI) based services that allow users to order rides to a specified location from the convenience of their phone. Transportation network companies,

such as Uber and Lyft, provide these services that can be found all over the world. In major U.S. cities, 21% of adults have used ridehailing services, and in the first five years of adoption, beginning in 2009, there were a total of 250 million estimated global users [20]. Unlike traditional taxi services, fare prices for ridehailing services are dynamic, calculated automatically using both the length of the requested trip as well as the demand and relative supply for ridehailing services in the area [16, 61, 63]. Potential AI bias in ridehailing applications, embedded in the algorithm and data, is a concern because of its possible disparate effect on people as well as the implications it has on the future of smart city automation and resource allocation.

In 2017, Uber launched a service called "Uber Movement", a platform on which cities can purchase and access transportation data collected by Uber with the goal of meeting future "smart city" aspirations [4]. The lack of extensive algorithmic transparency and the black box nature of ridehailing fare pricing algorithms leads to apprehension of whether they may be unintentionally exhibiting AI bias towards riders based on demographics or their neighborhoods. If there is bias in the algorithms and data being used in the development of smart cities, it may propagate outside of ridehailing platforms through data-driven policy making and automated resource allocation. Questions have been raised towards ridehailing providers' driverless car aspirations as well [9]. In addition to the economic strain placed on their workers [4, 9], automating smart city data collection by removing human drivers from the gig economy is concerning given the black box nature of dynamic pricing and supply allocation algorithms. As a result, it is still unclear how ridehailing providers impact traffic, supply and demand, and what data is being used to make automated pricing and supply allocation decisions.

In the context of price discrimination in ridehailing, we define AI bias as the difference in algorithmic outcomes for fare pricing associated with demographic characteristics of riders' neighborhoods. Such unintentional difference in outcomes, or disparate impact, is the current standard for legal determination of discrimination in the United States [29]. It was introduced in the case of *Griggs vs. Duke Power Co.* [1], which found African American employees were disparately impacted by employment exams, though unintentionally. AI systems have been known to adopt policies that display demographic disparity as well [43]. The disparity has been documented in algorithmic solutions to job candidate selection, online advertisements, recidivism prediction, and location-based pricing for college test preparation services [3, 7, 26, 44, 47, 59]. AI has also been shown to learn implicit biases documented by social psychologists, such as racial, gender, and other cultural biases, through

the statistical regularities of language [15], resulting in biased data representations that may propagate to downstream applications such as automated essay grading, university admissions, or resume screening for job candidate assessment [50].

Vendors of job candidate selection software have avoided claims of intentional discrimination by ensuring that their models do not contain legally protected attributes, such as race, religion, gender, or disability [49]. However, in the case of inadvertent discrimination, simply denying a model access to protected attributes does not guarantee an unbiased decision-making process, due to the presence of proxies to protected attributes [49]. Mehrabi et al. [47] denote 23 different ways that bias can exist in AI systems, more than 5 of which come from bias in the collection and interpretation of input data alone. Consequently, unless an AI system has been carefully vetted for the existence of bias, the existence of inevitable bias in the AI pipeline can be expected as the null hypothesis. This work is the first to investigate the fare pricing algorithms employed by ridehailing companies for evidence of AI bias. Results show that fare pricing based on demand relative to supply and trip duration leads to AI bias significantly associated with ethnic makeup, poverty levels, age, and education levels of neighborhoods.

Uber determines where to route their supply of gig workers by predicting demand for future rides using AI models that forecast based on historical demand information [38]. For fare pricing, Uber charges passengers based on the duration and length of their trips, as well as the times that riders use the application [61, 63]. In addition, Uber has been known to use the concept of a “surge multiplier” [16, 61], which increases prices depending upon demand relative to supply in a trip’s location [16, 34]. Ridehailing company Lyft also has a similar concept to the “surge multiplier”, called “primitime pricing” [5]. “Surge multipliers” vary based on independent “surge areas” within cities, which have independent “surge multipliers” [16]. Uber pricing varies based upon the location in the city from which rides begin and end. An anonymous Uber driver from the city of Chicago stated of the incentive areas, “low-income areas surged due to these zones”, leading to higher prices in those particular low-income neighborhoods. Pricing changes caused by incentive zones promoted by ridehailing service providers are troubling, as they may manipulate traffic, supply in neighborhoods, and ride forecasting, possibly motivating disparate impact and misrepresentation in future urban development.

This is the first analysis of AI bias in dynamic pricing algorithms of ridehailing, performed using a dataset of over 100 million rides from the city of Chicago from November 2018 to September 2019 [19]. Data from the city of Chicago was chosen because of a new city-wide law requiring ridehailing applications to disclose fare prices [21], making the Chicago ridehailing data the most complete dataset for understanding ridehailing fare pricing and price discrimination models currently available. The data includes fare prices and trip duration, as well as pickup and dropoff location, which can be used to connect the dataset to census tracts as designated by the American Community Survey (ACS) [13]. The ACS is an annual survey conducted by the federal government of the U.S. that collects demographic statistics about people residing in the U.S. census tracts. Census tracts are approximations of neighborhoods, subdivisions of a county designated by the U.S. Census Bureau consisting of about 1,200-8,000 people. Since “surge areas”

in ridehailing are not publicly available, fare pricing is examined by location using ACS census tracts [13].

The study presents a measurement of AI bias in the city of Chicago ridehailing dataset by quantifying associations between fare prices in neighborhoods and their corresponding demographic attributes. An application of random effects modeling drawn from meta-analysis literature computes city-level AI bias on fare pricing. This approach provides a thorough and repeatable blueprint for identifying and addressing bias in AI systems that produce outcomes using data from U.S. geolocations. Findings reveal that neighborhoods with large non-white populations, higher poverty levels, younger residents, and high education levels correspond to higher trip fares, with combined effect sizes, measured in Cohen’s d [23], of -0.32 , -0.28 , 0.69 , and 0.24 respectively. Our results indicate that fare pricing algorithms learning from demand, relative supply, and duration of trips may be the source of disparate impact in ridehailing. Neighborhood residents have no control over demand, supply, or trip duration, but are paying for the consequences of all three.

2 RELATED WORK

Current anti-discrimination laws do not account for AI bias as it may propagate in a predictive AI model [8]. Consequently, neither the model nor the ground truth data used to train the model may be available for examination of bias [6, 25, 57]. The lack of extensive algorithmic audit and the black box nature of ridehailing fare pricing algorithms leads to the concern of whether they may be unintentionally exhibiting AI bias towards riders based on their demographics or the properties of their neighborhoods.

We examine two different categories of related work in the space of algorithmic bias: research examining bias specifically in ridehailing and taxi service pickup rates [12, 31, 32], and research detailing how to ensure the fairness of supervised learning models [24, 28, 29, 51, 69]. We also examine two papers on the use of ACS demographic data rather than individual demographic attributes to make inferences about an individual [33, 55].

Dillahun et al. [27] recruited 13 low-income individuals to study Uber as passengers living in transportation-scarce environments. The analysis revealed barriers such as cost, limited payment methods, and low digital literacy, that make ridehailing services infeasible. Ge et al. [32] examine bias in Uber and Lyft pickups in the cities of Boston and Seattle by requesting human subjects to call for rides using Uber and Lyft, evaluating wait times and cancellation rates based on subject ethnicity through a user study. Their findings suggest male riders with African-American names are three times more likely to have rides cancelled and wait as much as 35% longer for rides. Brown [12] performs an extensive analysis of ridehailing applications and their effect on mobility in the city of Los Angeles, including an analysis of the different socioeconomic characteristics of the neighborhoods served by Lyft. Garz et al. [31] seek to bridge regulator and academic perspectives on underlying sources of harm for consumers and potential consequences such as high and hidden prices, post-contract exploitation, and discrimination. Unlike Ge et al. [32], our work does not rely on a user study, and unlike both Ge et al. [32] and Brown [12], our work is the first to provide an analysis on ridehailing trip fare pricing rather than frequency.

Raman et al. [51] studied the impact of rider-driver matching policy and redistributing income to reduce forms of inequality in ridehailing. The analysis revealed that policies minimizing the variance of income also maximize riders, especially during high demand periods. This finding indicates that reducing inequality can boost utility. Cohen et al. [24] analyze the application of fairness constraints in price discrimination algorithms' price, demand, consumer surplus, and no-purchase valuation. Feldman et al. [29] and Zemel et al. [69] detail methodologies for testing and removing bias from datasets, on the same datasets with different perspectives on fairness, disparate impact and statistical parity respectively. While the debiasing of a supervised model is not relevant to this work directly, these works describe methods for quantifying forms of bias and inequality. Generally, public access to white box models making decisions algorithmically or the data used to train them is not granted due to privacy regulations [25], as in prior cases of bias in online recruitment, online advertisements, and recidivism prediction [26, 44, 59]. The focus of our work is on measuring bias in individualized pricing for riders in neighborhoods defined by ACS. We use observable model outcomes and a partial set of input features, which can be a necessity if models and their training data cannot be made publicly available.

Geronimus et al. [33] and Soobader et al. [55] test how the use of ACS demographic data to make inferences about individuals compares to the use of individual demographic data to make inferences in the context of predicting health outcomes. Geronimus et al. [33] and Soobader et al. [55]'s findings conclude that the results of inferences based upon aggregates can not be interpreted the same as the results of inferences based on individual demographic attributes. However our methodology does not attempt to make inferences based on neighborhood-wide demographic data, but rather uses bias found across neighborhoods to test bias on individual demographics. The role of our methodology for investigating algorithmic decision-makers rather than making further inferences using their outcomes provides the justification for the use of ACS data in our analyses. In addition, without access to training data, this approach is currently a scientifically justified possible option for analyzing fairness and bias.

3 DATA

This study mainly focuses on ridehailing data from the city of Chicago and ACS data for Chicago's census tracts. Data from taxis in Chicago are also used to compare ridehailing to taxi fare pricing and coverage, whenever a comparison is possible. The ridehailing and taxi datasets used in this study are collected from the city of Chicago's data portal [18, 19]. Prior to 2019, there was no comparable comprehensive ridehailing dataset available for analysis.

Ridehailing data

The ridehailing data contains 100,717,117 trips in the city of Chicago from November of 2018 to September of 2019, and the taxi data contains 19,496,933 ride samples in Chicago during the same time period. Shared trips, or rides that involve the pickup and dropoff of multiple individuals, are excluded from the ridehailing dataset. These trips were not included because they contain multiple riders, have multiple pickup and dropoff points as well as a

different pricing strategy than rides with only a single rider [66]. In addition, including them would limit comparisons of ridehailing services with taxi services, as the taxi services do not list a similar ride sharing option in the provided dataset. After excluding shared rides and removing any rows with missing data from the dataset, a total of ~68 million ridehailing rides remain.

The dataset contains information about the time, charges, and pickup and dropoff locations for each ride, with location being given in the form of ACS census tracts. ACS census tracts are subdivisions of a county designated by the U.S. Census Bureau consisting of about 1,200-8,000 people, whose boundaries are updated every 10 years.

The start and end times of trips are rounded to the nearest 15 minutes and fares are rounded to the nearest \$2.50. No rider characteristics are reported in order to preserve the privacy of the riders [17]. The dataset contains information about the trip duration, charges, and pickup and dropoff locations for each ride, with location being given in the form of ACS census tracts. Trips are not labelled by ridehailing service, however, Uber takes roughly 72% of the ridehailing market share in Chicago, and Lyft and Via take up 27% and 1% of the market share [21].

This work is quantifying differences in fare price across neighborhoods defined by census tracts and not on individual riders. The demographics of individual riders are not available, and an individual rider's demographics cannot be assumed based on the neighborhoods where they are picked up or dropped off. Instead AI bias measurements, as detailed in the methods section, are taken based on neighborhood demographics. Other works examining correlations between ridehailing trips and neighborhood demographics [12, 34, 41] also perform correlational analysis on ridehailing data and census demographic attributes in a similar fashion.

Additional Preprocessing: Ridehailing data

Prior to calculating combined effect sizes of bias (detailed in the methods section) on fare price per mile, trip price per mile, and duration per mile are averaged for each census tract, in order to compare census tracts to each other. Pickup and dropoff counts are summed for each census tract, and then divided by the square meters of each census tract. Pickup and dropoff counts are spatially aggregated to account for density of rides in a census tract, rather than raw frequency, as in similar ridehailing trip frequency analyses [12].

Preprocessing: Taxi data

Preliminary analysis of the taxi datasets revealed many fare prices inconsistent with the rates charged by Taxi cab companies in Chicago, possibly due to the variety of taxi cab companies in the city and the self-reported nature of the taxi fare pricing data. For example, Yellow Cab Chicago charges a base fare of \$3.25, with an additional charge of \$2.25 per mile [2]. Given that the taxi dataset contains around 13 million rides when rows with missing data are removed, representative of approximately 37 million miles worth of driving, assuming all rides were taken using Chicago Yellow Cab, the total expected fare price would be approximately \$125 million. However, the observed overall fare price is closer to \$143 million. Multiple rides cost much more in fare per mile than expected given

listed cab prices, with over 1,000 rides costing more than \$1,000 per mile, and the maximum fare price per mile being approximately \$90,000. Rides with outlier fare prices were detected and removed using a Gaussian Mixture Model (GMM) [54].

GMMs are a probabilistic model which assume data is generated from a mixture of a number of Gaussian distributions, and have been shown to be an effective approach for anomaly detection in prior work [42, 45]. GMMs require the specification of a number, k , of clusters to create. Taxi rides are clustered assuming 2 Gaussian distributions, one corresponding to normal rides, and another for anomalous rides, meaning k is set to 2. After removing rides clustered in the anomalous cluster, the new total observed fare price is approximately \$130 million, down from \$143 million, which is closer to the expected total fare of \$125 million. The new maximum observed fare price per mile is \$17, which is more consistent to expectations based on Yellow Cab rates [2] compared to the original maximum fare price per mile of about \$90,000. After removing outliers 10, 292, 081 taxi rides remain.

3.1 American Community Survey Data

ACS census tract demographic information is retrieved from the United States Census Bureau from the year 2018. 2018 data was used as this is the closest year to the time frame of the Chicago ridehailing dataset for which ACS census data was available. The average age of ridehailing users as of 2017 is 37 [20]. In our representation of age information for census tracts, age is split using a threshold of 40 because age is organized in multi-year brackets in ACS data (0-5 years old, 5-9 years old, etc), and the age 40 is the closest bracket age that can be used to divide the population into those above and those below the average ridehailing user age. A poverty level threshold income of \$25,000 is chosen based on the poverty line as defined by United States Department of Health and Human Services [14]. The threshold for median house prices is set at \$250,000 based on a median house price estimate conducted by Zillow in 2018 [70].

All source code and the data analysis performed in this work are available at <https://github.com/akshat0123/RShareChicago>.

4 METHODS

Effect size has been used as a metric to assess implicit human bias in prior works, including the Implicit Association Test [35], which measures differential association of stereotypical attributes and social groups based on nationality, gender, race, religion, and sexuality. More recently, it has also been used to reveal and quantify biases in unsupervised machine learning approaches that generate representations of language, namely word embeddings [15]. Cohen's d [23] is a widely-used metric for calculating effect size of bias, which compares the difference of means of effect between two groups, to determine the magnitude of the effect of a treatment. Cohen's d can be used to measure AI bias because it quantifies the difference in fare pricing between trips in neighborhoods with different demographic statistics that represent distinct social groups.

4.1 Measuring Effect Size of AI Bias

An effect size calculation measures the difference in observed outcomes of two populations resulting from a single demographic

intervention. As an example, given two groups of neighborhoods A and B , an effect size of -1 says the mean fare price in A is one standard deviation less than the mean fare price in B . In the context of the demographic attribute non-white percentage, if A includes neighborhoods less than 50% non-white, and B includes all remaining neighborhoods, effect size will quantify the difference in mean fare price between majority white (A) and majority non-white neighborhoods (B).

Lone effect sizes cannot however, quantify the overall association between a continuous-valued demographic attribute and corresponding observed outcomes. The standardized mean difference in fare price between majority white and majority non-white neighborhoods only quantifies how fare prices differ when the non-white percentage is 50%. A complete association measure of fare price and non-white percentage requires that the difference in fare price at every non-white percentage interval for the continuous-valued attribute be accounted for. As a result, combining multiple effect sizes from different demographic intervals provides a more extensive association measure between a demographic attribute and fare price.

In order to calculate the association between each neighborhood demographic attribute and fare price, a combined effect size is calculated for each demographic attribute using a random effects model [11]. Random effects models from the validated meta-analysis literature summarize effect sizes from across studies, as detailed in Equation 1, by weighting each result with the within-study and between-study variances [37]. In this work, random effects models are used to combine effect size measurements across demographic intervals to produce a comprehensive measurement of AI bias. Further information about combined effect sizes are provided in the appendices.

4.2 Calculating Combined Effect Sizes

Combined effect sizes are a measurement of association between a demographic attribute X with a value below a threshold t and target outcomes (fare price), Y . Groups X_{t-} and X_{t+} are subsets of the data where the demographic attribute X is below t (X_{t-}) and above t (X_{t+}). Comparing the groups X_{t-} and X_{t+} quantifies how much having an X value below t changes mean output values. With regard to AI bias in the case of ridehailing, this determines the magnitude of the difference in outcomes (fare pricing) when comparing two sets of neighborhoods. Examining effect sizes rather than computing correlation coefficients provides the advantage of being able to pinpoint bias at particular thresholds (t values) when necessary, and is also free of any assumption about the linearity or monotonicity of the data, when compared to Pearson [10] or Spearman [68] correlation. In addition, other methods of determining feature importance, like information gain [40], cannot be applied due to the continuous nature of X and Y in the context of this work.

Effect sizes calculated between groups X_{t-} and X_{t+} as t is increased from the minimum to the maximum observed values of X are combined using a weighted mean of effect sizes, where weights are assigned using a random effects model [11]. Figure 2 goes over an example of measuring effect size of the proportion of non-white residents in a neighborhood on fare pricing along with information for interpreting the chart's x and y axes values.

$$\begin{aligned}
& \text{Combined Effect Size (ces)} \\
ces(X, Y) &= \frac{\sum_{t=[t_{min}...t_{max}]} d(X, Y, t) \times w(X, Y, t)}{\sum_{t=[t_{min}...t_{max}]} w(X, Y, t)} \quad (1) \\
d(X, Y, t) &= \frac{\bar{Y}_{X_{t-}} - \bar{Y}_{X_{t+}}}{\sigma(Y)} \\
w(X, Y, t) &= \frac{1}{var(X, Y) + var_t(X, Y, t)} \\
var(X, Y) &= \frac{\sigma^2}{t=[t_{min}...t_{max}]} [d(X, Y, t)] \\
var_t(X, Y, t) &= \frac{|X_{t-}| + |X_{t+}|}{|X_{t-}| |X_{t+}|} + \frac{d(X, Y, t)}{2(|X_{t-}| + |X_{t+}| - 2)} \\
&\quad \times \left(\frac{|X_{t-}| + |X_{t+}|}{|X_{t-}| + |X_{t+}| - 2} \right)
\end{aligned}$$

Figure 1: This equation displays how combined effect sizes are calculated. Random effects models from the validated meta-analysis literature summarize effect sizes from across studies by weighting each result with the within-study and between-study variances [11]. t represents the threshold by which census tracts are being compared. For example, when comparing neighborhoods by non-white percentage, a threshold of 50 means fare pricing in majority non-white and majority white neighborhoods are being compared. $d(X, Y, t)$ corresponds to a calculation of Cohen’s d [23] at threshold t , and $w(X, Y, t)$ denotes the weight for the Cohen’s d calculation at threshold t . $var(X, Y)$ is the variance across all effect size calculations for a demographic attribute X , and $var_t(X, Y, t)$ is the variance of a single Cohen’s d calculation [22].

In meta-analysis, random effects models are used to compute a combined effect given a set of effect size calculations, where weights are assigned to effect sizes given the amount of information they contain [11]. A random effects model is used to summarize the effect sizes rather than computing a mean of effect sizes because as t increases, the sizes of groups X_{t-} and X_{t+} change, which effects the variance of each effect size calculation. A random effects model is also called a variance components model where the model parameters are random variables. Fixed effects model does not apply in this scenario since the group means are not fixed, as opposed to a random effects model in which the group means are a random sample from a population.

In equation 1, $d(X, Y, t)$ corresponds to a calculation of Cohen’s d [23] at threshold t , and $w(X, Y, t)$ denotes the weight for the Cohen’s d calculation at threshold t . $var(X, Y)$ is the variance across all effect size calculations for a demographic attribute X , and $var_t(X, Y, t)$ is the variance of a single Cohen’s d calculation [22].

The statistical significance (p -value) for combined effect sizes are calculated using permutation testing [56], which measures the likelihood of the null hypothesis, by computing the probability that randomly selected groups of census tracts (with any t setting, from Equation 1) could produce a larger output than a combined effect size calculation. The permutation test is $Pr_i[ces(X_i, Y_i) > ces(X, Y)]$, where X_i and Y_i are a randomly selected subset of X and Y , which account for census-wide data for 1,159 census tracts for the ridehailing dataset and 820 census tracts for the taxi dataset. 1,000 iterations of the permutation test are run to calculate p -values for each reported AI bias measurement.

4.3 Interpreting Effect Size Charts

Figure 2 provides a guideline on how effect size charts can be interpreted. In Figure 2, when $A = 0.20$, the corresponding effect size is approximately -0.30 . This shows that neighborhoods that have less than 20% non-white residents have a mean fare price 0.30 standard deviations lower than neighborhoods that have more than 20% non-white residents. Every setting of A in the chart can be read similarly. When $A = 0.60$, the corresponding effect size is approximately -0.15 . This shows that neighborhoods that have less than 60% non-white residents have a mean fare price 0.15 standard deviations lower than neighborhoods that have more than 60% non-white residents. Using charts such as these, the neighborhoods with the large and small differences in mean fare pricing can be distinguished. In Figure 2, effect size is lowest approximately when A , or non-white percentage is approximately 5%. This shows that neighborhoods with less than a 5% non-white population have the largest difference in mean fare price, approximately -1.10 standard deviations away from fare prices in neighborhoods with greater than a 5% non-white population.

5 RESULTS

In relation to ridehailing, AI bias measures a difference in variables of the dataset based on neighborhood demography. For fare pricing, this difference in outcomes is measured as the difference in mean fare price between neighborhoods with different demographic characteristics. The standardized difference in mean fare price is Cohen’s d [23], or the effect size. Conventional small, medium, and large values of Cohen’s d are 0.20, 0.50, and 0.80. In Figure 3, effect size of AI bias measures the differential fare pricing associated with two groups of neighborhoods with different ratios of a demographic attribute. For example, the chart with the x -axis $A = \% \text{ Non-U.S. Citizens}$ in Figure 3, measures the effect of the variable percentage of immigrants in a neighborhood, on fare pricing.

AI bias effect sizes, summarized using random effects modeling, show that fares, trip frequency, and trip duration are all lower for trips beginning or ending in neighborhoods with older populations (see Tables 1 and 2). Trips beginning in neighborhoods with a smaller ratio of educated residents, as well as neighborhoods with lower home values also have lower fares, trip frequency, and trip durations (see Tables 1 and 2). Trips beginning or ending in neighborhoods with more non-white residents, or more residents living below the poverty line, have higher fares, as well as higher trip duration (see Table 1). However, in both of these groups of neighborhoods (non-white, high level of poverty), demand is lower

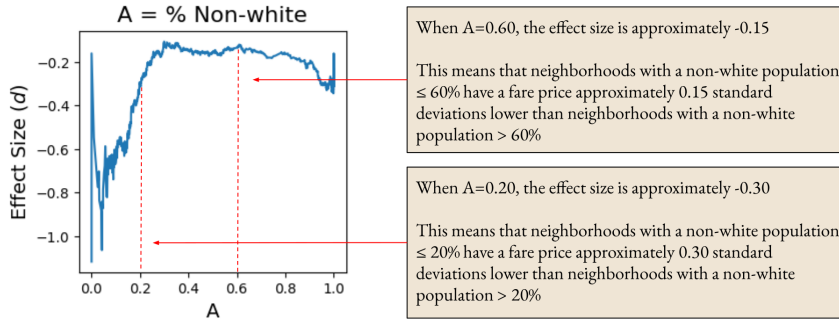


Figure 2: This figure displays how effect size charts in this paper can be interpreted. The chart displayed here shows the effect size of the non-white percentage of a neighborhood's population on ridehailing fare pricing. In this chart, when $A = 0.20$, the corresponding effect size is approximately -0.30 . This shows that neighborhoods that are less than 20% non-white have a mean fare price 0.30 standard deviations lower than neighborhoods that are greater than 20% non-white.

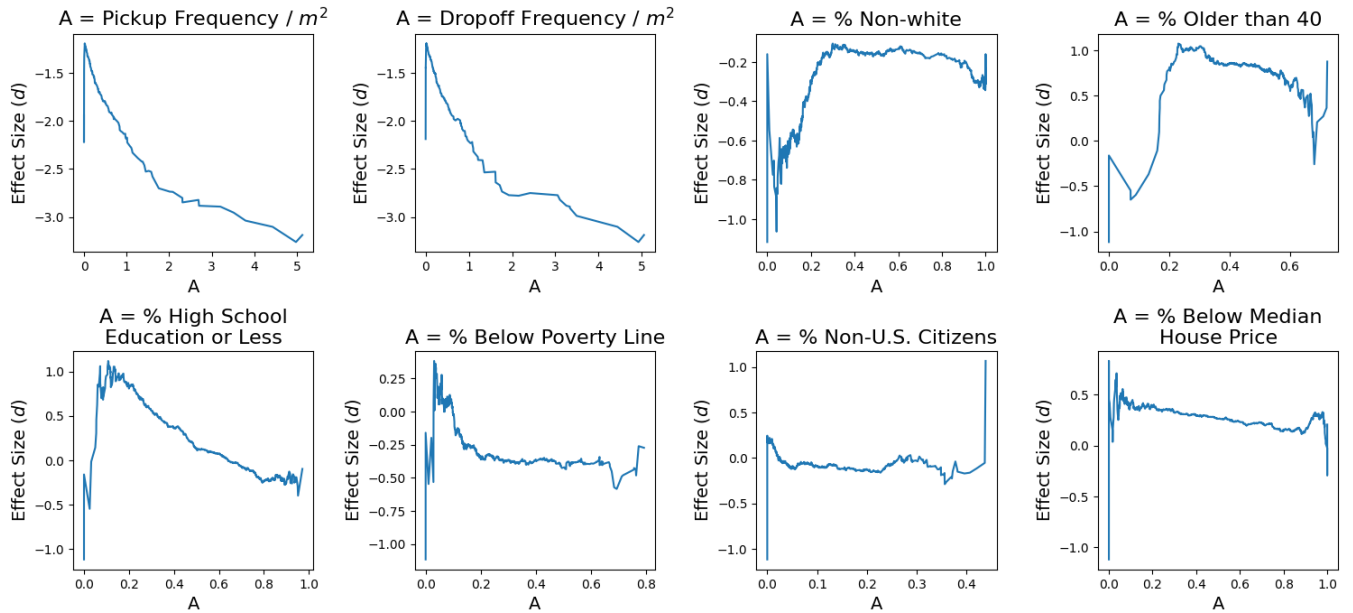


Figure 3: Effect sizes on fare price per mile are calculated for pickup and dropoff frequencies per neighborhood. Effect sizes on fare price per mile are also calculated for the percentage of non-white, over 40, high-school educated or less, below poverty line, and non-U.S. citizen populations, as well as the percentage of people living in homes below median home value in Chicago. Effect sizes are calculated using Cohen's d [23]. As an example, in the chart labelled " $A = \% \text{ Non-White}$ ", when $x = 0.2$, the effect-size is $y = -0.6 (d)$. This says the average fare price per mile for neighborhoods less than 20% non-white is 0.6 standard deviations less than the average fare price per mile for neighborhoods more than 20% non-white. Accordingly, the more the non-white populations in a census tract, the more expensive the fare price.

(see Table 2). The percentage of immigrants in a neighborhood does not show any significant association with fare pricing (see Table 1).

Pickup and dropoff frequencies have a strong positive association with fare pricing, meaning that neighborhoods with more frequent trips have higher prices (see Table 1). The positive relationship between pickup and dropoff frequencies and fare price is consistent with Uber's "surge pricing" strategy [65] - areas where there are more trips or demand are more likely to have relatively fewer drivers, or supply, and go into "surge", causing prices to increase. This pattern is not observed in neighborhoods with more non-white residents, or with more residents living below the poverty line, which is an anomaly. However, if there is very low supply of drivers in these neighborhoods, demand relative to the supply and

accordingly fare pricing might increase. Ridehailing providers do not publish data regarding where and why they route the supply of drivers and which areas they declare as unsafe no go zones. As a result, analyzing AI bias caused by algorithm's surge pricing objective that relies on supply is currently not possible. Having access to this information would make it possible to understand how algorithmic supply routing may affect the traffic conditions, relative demand, and the corresponding fare pricing.

The highest demand neighborhoods for ridehailing trips in Chicago are in the city's business district, known as "The Loop" (see Figure 4). Neighborhoods with a high percentage of non-white residents, and neighborhoods where much of the population live below

the poverty line, lie south of Chicago's high-demand business district (see Figure 4). Most of these neighborhoods (non-white, high poverty level) are concentrated in the southwest corner of Chicago, where trip durations are low, while others are just southwest of the business district, where trip durations are high (see Figure 4).

Combined effect size analysis for a taxi fare dataset from Chicago collected from the same time period as the ridehailing data shows that the demographic attributes of neighborhoods where taxi trips begin have no significant association with fare pricing (see Table 1). Taxi rides that end in neighborhoods with older populations, less educated residents, or lower value homes have cheaper fare prices (see Table 1). The association between taxi fare pricing and age of the population is weaker than the association between ridehailing fare pricing and age of the population. Taxi fare pricing has a stronger association with both college education levels and home values in a trip destination neighborhood than ridehailing fare pricing does.

6 DISCUSSION

Prior works have shown that younger people and college educated individuals use ridehailing services more frequently than older people and those without a college education [12, 20], which is reflected in the measurement of AI bias fare pricing as well as demand measured by trip frequency (see Table 1 and Table 2). In addition, for Chicago's census tracts, the percentage of population older than 40 or without a college degree in a neighborhood is negatively correlated with dropoff frequency (Pearson's $r = -.23, p < 10^{-4}$, $r = -.40, p < 10^{-4}$). Higher levels of demand combined with longer trip durations (see Tables 1 and Table 2) may explain why younger and highly college educated neighborhoods have higher fare prices.

The difference in fare pricing for neighborhoods with more non-white residents or more residents living below the poverty line could be occurring for a variety of reasons. These neighborhoods are associated with longer trip duration but also decreased demand, which doesn't directly explain the increased fare pricing. One potential reason for decreased demand is the low number of businesses and entertainment options in non-white and impoverished neighborhoods relative to the rest of the city. The non-white percentage and poverty level of a neighborhood are both negatively correlated with the number of jobs in arts, food, and entertainment (Pearson's $r = -.13, p < 10^{-4}$, $r = -.11, p < 10^{-3}$) as well as the total number of businesses in a neighborhood (Pearson's $r = -.27, p < 10^{-4}$, $r = -.25, p < 10^{-4}$). However, the number of ridehailing trips in a neighborhood is positively correlated with the number of businesses and jobs in arts, food, and entertainment (Pearson's $r = .84, p < 10^{-4}$, $r = .88, p < 10^{-4}$).

Prior works [20, 52] have found that the most common reason for using ridehailing services is for leisure activities, such as going to bars and restaurants. Residents of neighborhoods that are non-white or have high poverty levels may have to travel further to business district neighborhoods in order to pursue the same leisure activities as neighborhoods with lower non-white or below poverty level populations. Demand and fare pricing in neighborhoods with more businesses and entertainment is higher than in other areas of the city (see Figure 4). In addition, the non-white percentage and poverty level of neighborhoods are both positively correlated with

the number of workers in arts, food, and entertainment (Pearson's $r = .21, p < 10^{-4}$, $r = .34, p < 10^{-4}$). Many of these workers may also be traveling to neighborhoods where there are more arts and entertainment options for their employment. Overall, traveling to or from high demand neighborhoods during rush hour might increase fare pricing for trips originating or ending in neighborhoods with larger non-white or below poverty level populations.

Combined effect size analysis indicates that trips for neighborhoods with more non-white and below poverty level residents have longer trip times (see trip seconds/mile columns in Table 2). Increased trip times may be related to trips to higher traffic areas of the city with more businesses and entertainment options. The number of jobs in arts, food, and entertainment have a positive correlation with traffic and car accidents (Pearson's $r = 0.40, p < 10^{-4}$, $r = 0.59, p < 10^{-4}$). Similarly, the number of businesses in a neighborhood positively correlates with both traffic and car accidents (Pearson's $r = 0.58, p < 10^{-4}$, $r = 0.81, p < 10^{-4}$).

One potential reason for increased fare pricing in non-white and high poverty level neighborhoods is being subject to surge pricing due to low supply of drivers if the supply of drivers are not distributed uniformly, relative to the demand distribution, around the city. If drivers avoid these neighborhoods and the demand is relatively higher than supply, the ridehailing algorithm's dynamic pricing objective based on the surge principle will perpetuate social bias by increasing the algorithmic fare price. Since the data and corresponding algorithm for ridehailing supply allocation are not available to the public, we are not able to measure the magnitude of bias associated with surge pricing that depends on supply information.

We discuss in detail how another possible reason for high fares in neighborhoods with more non-white residents is driver prejudice and fear of neighborhoods with high crime rates. Ge et al. [32] found that ridehailing users in Boston with African-American names had rides cancelled more than twice as often as those with white sounding names. The same study also found that in Seattle, riders with African-American names had almost a 35% increase in waiting times. If drivers or the supply allocation algorithm are avoiding non-white and high poverty level neighborhoods, this will cause an increase in fare pricing due to relatively low supply.

Uber drivers have also been reported to avoid high-crime areas on multiple occasions. So called "no-go zones", which are areas avoided by Uber drivers, have been identified in Rio de Janeiro [58], Johannesburg [53], and Atlanta [39]. Uber drivers avoid these areas due to fear of becoming victim of a crime. In 2017, Uber tested a new feature allowing drivers to filter out areas of cities in which they did not wish to drive [36] and currently, Uber still allows drivers to set a destination twice a day [64], which matches them to riders going toward their specified destination.

The percentage of non-white residents in a neighborhood and percentage of below poverty level residents of a neighborhood in Chicago are both positively correlated with crime rates (Pearson's $r = 0.23, p < 10^{-4}$, $r = 0.27, p < 10^{-4}$). Crime rates are positively correlated with fare prices at both trip pickup location (Pearson's $r = 0.47, p < 10^{-4}$) and fare prices at trip dropoff location (Pearson's $r = 0.49, p < 10^{-4}$) as well. Allowing drivers to choose destinations, as well as driver preferences, in conjunction with demand and relative supply based pricing could be causing

Attribute	Ridehailing Fare Price/Mile				Taxi Fare Price/Mile			
	Pickup		Dropoff		Pickup		Dropoff	
	Combined Effect Size (d)	p	Combined Effect Size (d)	p	Combined Effect Size (d)	p	Combined Effect Size (d)	p
Pickup Frequency / m^2	-1.57	$< 10^{-3}$	-1.59	$< 10^{-3}$	0.24	0.04	-0.96	$< 10^{-3}$
Dropoff Frequency / m^2	-1.57	$< 10^{-3}$	-1.57	$< 10^{-3}$	-0.15	0.21	-0.88	$< 10^{-3}$
% Non-white	-0.22	0.02	-0.32	$< 10^{-3}$	0.18	0.11	0.07	0.50
% Older than 40	0.66	$< 10^{-3}$	0.69	$< 10^{-3}$	-0.04	0.72	0.52	$< 10^{-3}$
% High School education or less	0.24	0.01	0.15	0.17	0.05	0.65	0.41	$< 10^{-3}$
% Below Poverty Line	-0.19	0.05	-0.28	$< 10^{-3}$	0.01	0.94	0.05	0.59
% Non-U.S. Citizens	-0.10	0.35	-0.07	0.54	-0.02	0.83	0.04	0.71
% Below Median House Price	0.23	0.02	0.19	0.06	0.02	0.85	0.32	$< 10^{-3}$

Table 1: Combined effect sizes on fare price per mile by neighborhood attributes - Combined effect sizes scores are shown for the fare price per mile given a set of neighborhood attributes. Combined effect sizes are weighted using a random-effects model [11]. The “Ridehailing” column contains combined effect sizes calculated for ridehailing trip fares, and the “Taxi” column contains combined effect sizes calculated for “Taxi” trip fares. “Pickup” and “Dropoff” columns designate fare price per mile when being picked up or dropped of in a neighborhood and “p” presents the p-value for effect size calculations.

Attribute	Ridehailing Trip Frequency / m^2				Ridehailing Trip Seconds/Mile			
	Pickup		Dropoff		Pickup		Dropoff	
	Combined Effect Size (d)	p	Combined Effect Size (d)	p	Combined Effect Size (d)	p	Combined Effect Size (d)	p
% Non-white	0.25	$< 10^{-3}$	0.23	$< 10^{-3}$	-0.14	0.17	-0.31	$< 10^{-3}$
% Older than 40	0.38	$< 10^{-3}$	0.38	$< 10^{-3}$	0.67	$< 10^{-3}$	0.70	$< 10^{-3}$
% High School Education or less	0.69	$< 10^{-3}$	0.69	$< 10^{-3}$	0.20	0.05	0.08	0.43
% Below Poverty Line	0.27	$< 10^{-3}$	0.26	$< 10^{-3}$	-0.13	0.19	-0.25	$< 10^{-3}$
% Non-U.S. Citizens	-0.09	0.21	-0.11	0.13	-0.21	0.02	-0.11	0.32
% Below Median House Price	0.42	$< 10^{-3}$	0.43	$< 10^{-3}$	0.30	$< 10^{-3}$	0.17	0.08

Table 2: Combined effect sizes on trip frequency and trip seconds per mile by neighborhood attributes - combined effect sizes are shown given a set of neighborhood attributes. The “Ridehailing Trip Frequency / m^2 ” column contains combined effect sizes of neighborhood demographic attributes on the frequency of trips in a neighborhood, and the “Ridehailing Trip Seconds/Mile” column contains combined effect sizes measuring the effect size of neighborhood attributes on the time taken per mile for trips in a neighborhood. “Pickup” and “Dropoff” columns designate frequency and trip time when being picked up or dropped of in a neighborhood and “p” presents the p-value for effect size calculations.

prices to rise in areas of high crime due to a lack of driver supply. A comprehensive safety report released by Uber [60] found a total 7 cases of fatal physical assault against drivers out of 1.3 billion rides taken in the United States in 2018. Ridehailing companies should be concerned about algorithmic disparate impact if neighborhoods are being avoided due to fear of crime, direct driver prejudice, or the supply allocation algorithm. 14% of the rides in Chicago examined in this study are in majority non-white neighborhoods.

The ridehailing data provided by the city of Chicago is a list of rides, and not a list of ride requests, which are not public. Since ride requests are not available, trip frequency in a neighborhood only partially accounts for demand. Ridehailing companies have detailed information about demand and how the supply of gig workers is routed around the city based on monetary incentives that is not publicly available.

Uber and Lyft both offer bonuses to drivers to drive in particular areas of cities. Uber offers drivers the ability to earn bonuses by taking rides in “boost zones” [62] and Lyft similarly offers “personal power zones” [46]. Figure 5, provided by an anonymous gig worker,

displays how “boost zones” and “surge areas” work from the perspective of a driver. Blue areas with bonuses listed on them (e.g. “15 for 3”) are boost zones, where drivers receive bonuses when they complete rides. Red areas are in surge, meaning that high demand compared to available supply has caused prices in the area to rise. The driver that provided the screenshot claimed “low-income areas surged after bonuses were introduced by the ridehailing platform in blue zones”. The area to the left of the boost zones, “Austin”, is a community area in Chicago that contains a population that is 95.00% non-white, 55.50% high school educated or less, and 39.20% below the poverty line [30]. **Artificial manipulation of supply by ridehailing companies, incentivizing drivers to shift to other parts of the city, could also be a cause for increased prices by artificially inflating demand in neighborhoods with high percentages of non-white and below poverty level populations.**

The main findings of this analysis of Chicago ridehailing data are that fare pricing is higher for neighborhoods with younger, highly college educated, non-white, and high below poverty level populations. The difference in pricing for younger and highly college

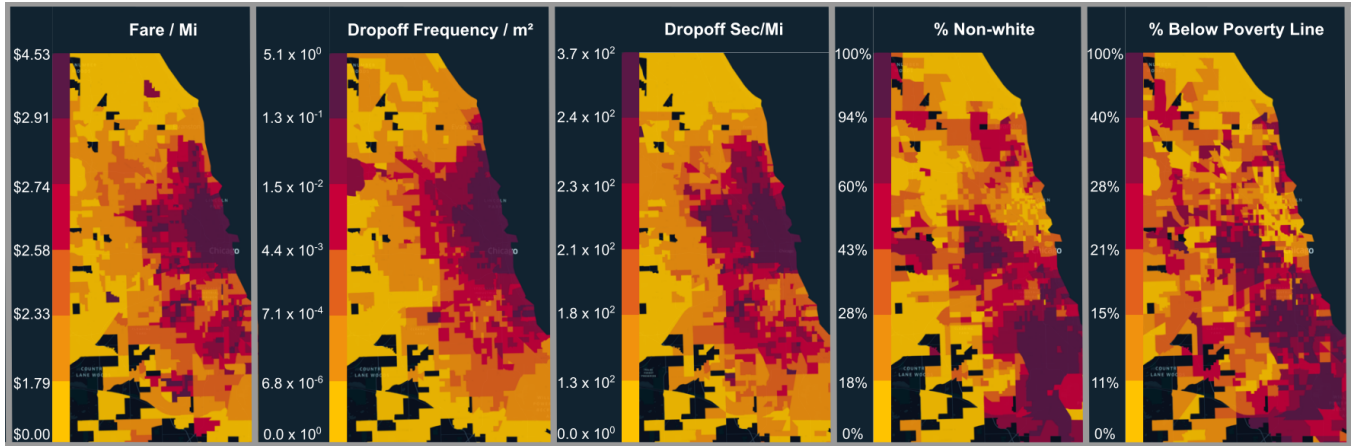


Figure 4: The heat maps depict, in order, fare pricing, ride frequency, ride duration, ethnic makeup, and poverty levels across the city of Chicago. The fare pricing map shows average fare per mile prices in a census tract in dollars, the dropoff count map shows the average number of dropoffs per square meter in each census tract, and the ride duration map shows the average number of seconds per mile for trips in a census tract. The remaining maps show the amount of each demographic labelled in their respective titles by percentage in each census tract. The color coding for each map is scaled by 6-quintiles, with darker colors representing higher values and lighter colors representing lower values for each respective neighborhood attribute. Non-white neighborhoods and neighborhoods with a large population living below the poverty line have higher fare prices (combined effect size $d = -0.32, -0.28$). Non-white and impoverished neighborhoods also show lower demand (combined effect size $d = 0.25, 0.27$), but increased trip duration (combined effect size $d = -0.31, -0.25$). Increased fare prices in these areas are anomalous since low demand can cancel increased duration’s effect in fare pricing. Consequently, ridehailing algorithms’ notion of surge pricing based on relative supply and demand might be the underlying source of bias leading to disparate impact.

educated neighborhoods may be explained by an analysis of demand and trip duration (see Table 2). The differences in pricing for neighborhoods with more non-white and below poverty level residents may be explained by trip duration, but can not be explained by low demand (see Table 2 and Figure 4). Raghavan and Barocas [49] point out that even in the case when a business argues disparate impact caused by an algorithmic decision maker is justified due to it being the result of a necessary business purpose, the business may still be held liable if an alternative decision making process could perform the same purpose without disparate impact. For the case of algorithmic hiring, Raghavan and Barocas [49] suggest that vendors of such models should, at minimum, know the protected classes of individuals during training to prevent disparate impact. For the case of ridehailing, in order to prevent disparate impact, ridehailing companies should at the very least consider the demography, effect of prejudice and algorithmic allocation on supply, transit geography, and economic geography of the cities in which they operate when designing fare pricing algorithms. Moreover, measuring bias, simulating fairness, or Raman et al. [51]’s approaches to eliminate inequity while increasing utility might help ridehailing providers design a more equitable price discrimination algorithm.

To extend this analysis, obtaining the supply allocation algorithm, as well as data regarding the total number of ride requests and supply in each neighborhood would allow for a more comprehensive understanding of the relationship between demand and fare pricing. Current analysis is dependent upon trip frequencies

in neighborhoods, which does not exhaustively account for relative rider demand. Information about what rides are declined, and how price changes for a rider after their initial request is declined may further explain why prices are high in areas where there is relatively lower demand. Understanding how fare prices change in neighborhoods when driver incentive zones are activated can also be used to explain why higher fare prices occur in neighborhoods with relatively low demand. Explicit detail of how, when, and where an incentive zone is introduced, and how these zones impact trip frequency would also shed light on how ridehailing companies may be inadvertently manipulating demand and relative supply, and as a result, fare prices.

7 CONCLUSION

This study demonstrates that dynamic pricing algorithms of ridehailing learn AI bias via the ridehailing utilization patterns in neighborhoods with diverse demographic makeup. The methodology used in measuring AI bias in ridehailing can be adapted to examine bias in datasets of geolocation based predictive AI models that have continuous outcomes such as pricing.

According to the publicly available data, ridehailing fare pricing algorithms that partially use trip frequency and duration information in decision making have generated a pricing scheme that results in disparate impact given neighborhood demographics in the city of Chicago. From the time period between November 2018 and September 2019, neighborhoods with younger, more college educated, poorer, and more non-white residents faced higher fares

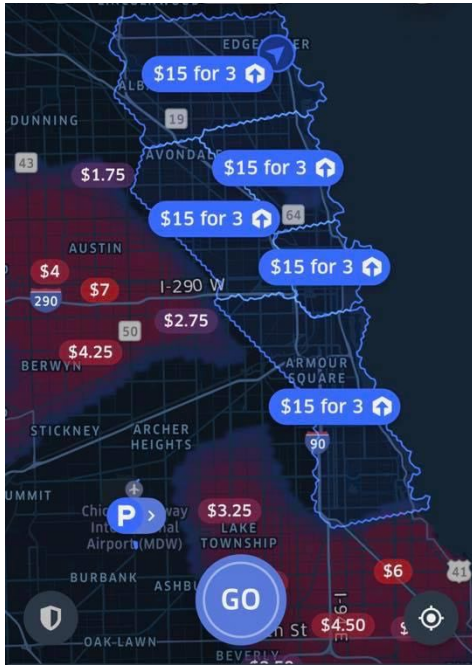


Figure 5: This image displays an anonymous driver’s view when driving for Uber. Blue areas with bonuses listed on them (e.g. “15 for 3”) are boost zones, where drivers are incentivized to receive bonuses when they complete rides. Red areas are in surge, meaning that low supply of drivers in these areas resulted in relatively high demand, which in turn caused prices to rise in the area.

using ridehailing applications in the city of Chicago. Moreover, current and past pricing may affect future demand, which may in turn change pricing. Several iterations of this feedback cycle may cause bias to propagate from fare pricing to service, as the supply of drivers is continuously shifted to higher demand areas.

In addition to existing prejudice, if the social demography, transit geography, and economic geography of a city are not taken into account in the development of dynamic pricing algorithms, there may be disparate impact in pricing outcomes. If disparate impact in outcomes is not accounted for, it will remain in the data when it is re-purposed for smart city development, where it may cause further disparate impact depending on how the data is used. From a policy perspective, lack of appropriate transparency measures regarding data collection and AI decision making enforced upon companies may be contributing to disparate impact. Ultimately, algorithmic bias cannot be mitigated without algorithmic transparency.

8 ACKNOWLEDGEMENTS

We want to thank Robert Pless and the anonymous reviewers for their useful feedback.

REFERENCES

[1] 1971. *Griggs v. Duke Power Co.*, 401 U.S. 424 (Mar 1971). <https://supreme.justia.com/cases/federal/us/401/424/>

[2] 2019. Yellow Cab Chicago Rates. <https://yellowcabchicago.com/rates/>

[3] Julia Angwin, Surya Mattu, and Jeff Larson. 2015. Asian Students Are Charged More for Test Prep. <https://www.theatlantic.com/education/archive/2015/09/princeton-review-expensive-asian-students/403510/>

[4] Kafui Attah, Katie Wells, and Declan Cullen. 2019. “We’re building their data”: Labor, alienation, and idiocy in the smart city. *Environment and Planning D: Society and Space* 37, 6 (2019), 1007–1024. <https://doi.org/10.1177/0263775819856626>

[5] Siddhartha Banerjee, Ramesh Johari, and Carlos Riquelme. 2016. Dynamic Pricing in Ridesharing Platforms. *SICom Exch.* 15, 1 (Sept. 2016), 65–70. <https://doi.org/10.1145/2994501.2994505>

[6] Iain Barclay, Alun D. Preece, Ian Taylor, and Dinesh C. Verma. 2019. Quantifying Transparency of Machine Learning Systems through Analysis of Contributions. *CoRR abs/1907.03483* (2019). arXiv:1907.03483 <http://arxiv.org/abs/1907.03483>

[7] Solon Barocas, Moritz Hardt, and Arvind Narayanan. 2019. *Fairness and Machine Learning*. fairmlbook.org. <http://www.fairmlbook.org>

[8] Solon Barocas and Andrew D. Selbst. 2016. Big Data’s Disparate Impact. *California Law Review* 104, 3 (2016), 671–732. <http://www.jstor.org/stable/24758720>

[9] Aaron Benanav. 2020. Why Uber’s business model is doomed. *The Guardian* (Aug 2020). https://www.theguardian.com/commentisfree/2020/aug/24/gig-economy-uber-lyft-insecurity-crisis?CMP=fb_cif

[10] Jacob Benesty, Jingdong Chen, Yiteng Huang, and Israel Cohen. 2009. Pearson correlation coefficient. In *Noise reduction in speech processing*. Springer, 37–40.

[11] Michael Borenstein. 2012. *Introduction to meta-analysis*. Wiley.

[12] Anne Elizabeth Brown. 2018. *Ridehail Revolution: Ridehail Travel and Equity in Los Angeles*. Ph.D. Dissertation. UCLA.

[13] US Census Bureau. 2020. American Community Survey (ACS). <https://www.census.gov/programs-surveys/acs/>

[14] Sylvia M Burwell. 2015. Annual Update of the HHS Poverty Guidelines. <https://www.federalregister.gov/documents/2015/01/22/2015-01120/annual-update-of-the-hhs-poverty-guidelines>

[15] Aylin Caliskan, Joanna J Bryson, and Arvind Narayanan. 2017. Semantics derived automatically from language corpora contain human-like biases. *Science* 356, 6334 (2017), 183–186.

[16] Le Chen, Alan Mislove, and Christo Wilson. 2015. Peeking Beneath the Hood of Uber. In *Proceedings of the 2015 Internet Measurement Conference* (Tokyo, Japan) (IMC ’15). Association for Computing Machinery, New York, NY, USA, 495–508. <https://doi.org/10.1145/2815675.2815681>

[17] City of Chicago. 2017. How Chicago Protects Privacy in TNP and Taxi Open Data. <http://dev.cityofchicago.org/opendata/dataportal/2019/04/12/tnp-taxi-privacy.html>

[18] City of Chicago. 2020. Taxi Trips. <https://data.cityofchicago.org/Transportation/Taxi-Trips/wrvz-psew>

[19] City of Chicago. 2020. Transportation Network Providers - Trips. <https://data.cityofchicago.org/Transportation/Transportation-Network-Providers-Trips/m6dm-c72p> Accessed: 2020-05-05.

[20] Regina R. Clewlow and Gouri S. Mishra. 2017. *Disruptive Transportation: The Adoption, Utilization, and Impacts of Ride-Hailing in the United States*. Institute of Transportation Studies, Working Paper Series qt82w2z91j. Institute of Transportation Studies, UC Davis. <https://ideas.repec.org/p/cdl/itsdav/qt82w2z91j.html>

[21] CNBC. 2019. Uber’s carpool pricing strategy is revealed by Chicago fare data. <https://www.cnbc.com/2019/11/26/ubers-carpool-pricing-strategy-is-revealed-by-chicago-fare-data.html>

[22] Robert; Coe. 2002. It’s the effect size, stupid: what effect size is and why it is important. <http://www.leeds.ac.uk/educol/documents/00002182.htm> Presented at the Annual Conference of the British Educational Research Association, University of Exeter, Exeter, United Kingdom, September 12–14, 2002.

[23] Jack Cohen. 1988. *Statistical power analysis for the behavioral sciences*. L. Erlbaum Associates.

[24] Maxime Cohen, Adam N Elmachet, and Xiao Lei. 2019. Price Discrimination with Fairness Constraints. *Available at SSRN 3459289* (2019).

[25] European Commission. 2016. Art. 9 GDPR – Processing of special categories of personal data. <https://gdpr-info.eu/art-9-gdpr/>

[26] Amit Datta, Michael Carl Tschantz, and Anupam Datta. 2015. Automated experiments on ad privacy settings: A tale of opacity, choice, and discrimination. *Proceedings on privacy enhancing technologies* 2015, 1 (2015), 92–112.

[27] Tawanna R Dillahun, Vaishnav Kameswaran, Linfeng Li, and Tanya Rosenblatt. 2017. Uncovering the values and constraints of real-time ridesharing for low-resource populations. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*. 2757–2769.

[28] Cynthia Dwork, Moritz Hardt, Toniann Pitassi, Omer Reingold, and Richard S. Zemel. 2011. Fairness Through Awareness. *CoRR abs/1104.3913* (2011). arXiv:1104.3913 <http://arxiv.org/abs/1104.3913>

[29] Michael Feldman, Sorelle Friedler, John Moeller, Carlos Scheidegger, and Suresh Venkatasubramanian. 2014. Certifying and removing disparate impact. arXiv:1412.3756 [stat.ML]

[30] Chicago Metropolitan Agency for Planning. 2020. Community Data Snapshots. <https://www.cmap.illinois.gov/data/community-snapshots>

- [31] Seth Garz, Xavier Giné, Dean Karlan, Rafe Mazer, Caitlin Sanford, and Jonathan Zinman. 2020. *Consumer protection for financial inclusion in low and middle income countries: Bridging regulator and academic perspectives*. Technical Report. National Bureau of Economic Research.
- [32] Yanbo Ge, Christopher R Knittel, Don MacKenzie, and Stephen Zuepf. 2016. *Racial and Gender Discrimination in Transportation Network Companies*. Working Paper 22776. National Bureau of Economic Research. <https://doi.org/10.3386/w22776>
- [33] Arline T. Geronimus, John Bound, and Lisa J. Neidert. 1996. On the Validity of Using Census Geocode Characteristics to Proxy Individual Socioeconomic Characteristics. *J. Amer. Statist. Assoc.* 91, 434 (1996), 529–537. <https://doi.org/10.1080/01621459.1996.10476918>
- [34] Raymond Gerte, Karthik C. Konduri, and Naveen Eluru. 2018. Is There a Limit to Adoption of Dynamic Ridesharing Systems? Evidence from Analysis of Uber Demand Data from New York City. *Transportation Research Record* 2672, 42 (2018), 127–136. <https://doi.org/10.1177/0361198118788462> arXiv:<https://doi.org/10.1177/0361198118788462>
- [35] Anthony G. Greenwald, Debbie E. McGhee, and Jordan L. K. Schwartz. 1998. Measuring individual differences in implicit cognition: The implicit association test. *Journal of Personality and Social Psychology* 74, 6 (1998), 1464–1480. <https://doi.org/10.1037/0022-3514.74.6.1464>
- [36] Andrew J. Hawkins. 2017. Uber is giving some drivers more control over where they drive. <https://www.theverge.com/2017/8/16/16157810/uber-drop-off-area-filter-control-drivers>
- [37] Larry V Hedges. 1983. A random effects model for effect sizes. *Psychological Bulletin* 93, 2 (1983), 388.
- [38] Jeremy Hermann. 2019. Scaling Machine Learning at Uber with Michelangelo. <https://eng.uber.com/scaling-michelangelo/>
- [39] Jamie Kennedy. 2019. Rideshare drivers label "no-go-zones" in Metro Atlanta. https://www.cbs46.com/news/rideshare-drivers-label-no-go-zones-in-metro-atlanta/article_e6a7a44e-157d-11ea-83bd-5b5c6a0addf7.html
- [40] Solomon Kullback. 1997. *Information theory and statistics*. Dover Publ.
- [41] Patricia S. Lavieri, Felipe F. Dias, Natalia Ruiz Juri, James Kuhr, and Chandra R. Bhat. 2018. A Model of Ridesourcing Demand Generation and Distribution. *Transportation Research Record* 2672, 46 (2018), 31–40. <https://doi.org/10.1177/0361198118756628> arXiv:<https://doi.org/10.1177/0361198118756628>
- [42] R. Laxhammar. 2008. Anomaly detection for sea surveillance. In *2008 11th International Conference on Information Fusion*. 1–8.
- [43] Nicol Turner Lee. 2018. Detecting racial bias in algorithms and machine learning. *Journal of Information, Communication and Ethics in Society* (2018).
- [44] Nicol Turner Lee, Paul Resnick, and Genie Barton. 2019. Algorithmic bias detection and mitigation: Best practices and policies to reduce consumer harms. *Brookings Institute: Washington, DC, USA* (2019).
- [45] Lishuai Li, R. Hansman, Rafael Palacios, and Roy Welsch. 2016. Anomaly detection via a Gaussian Mixture Model for flight operation and safety monitoring. *Transportation Research Part C: Emerging Technologies* 64 (03 2016), 45–57. <https://doi.org/10.1016/j.trc.2016.01.007>
- [46] Lyft. 2018. Personal Power Zones. <https://help.lyft.com/hc/en-us/articles/115012926807-Personal-Power-Zones>
- [47] Ninareh Mehrabi, Fred Morstatter, Nripsuta Saxena, Kristina Lerman, and A. Galstyan. 2019. A Survey on Bias and Fairness in Machine Learning. *ArXiv abs/1908.09635* (2019).
- [48] City of Chicago. 2020. City of Chicago: Data Portal: City of Chicago: Data Portal. <https://data.cityofchicago.org/>
- [49] Manish Raghavan and Solon Barocas. 2019. Challenges for mitigating bias in algorithmic hiring. <https://www.brookings.edu/research/challenges-for-mitigating-bias-in-algorithmic-hiring/>
- [50] Manish Raghavan, Solon Barocas, Jon Kleinberg, and Karen Levy. 2020. Mitigating bias in algorithmic hiring: Evaluating claims and practices. In *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency*. 469–481.
- [51] Naveen Raman, Sanket Shah, and John P Dickerson. 2020. Data-Driven Methods for Balancing Fairness and Efficiency in Ride-Pooling. (2020).
- [52] Lisa Rayle, Danielle Dai, Nelson Chan, Robert Cervero, and Susan Shaheen. 2016. Just A Better Taxi? A Survey-Based Comparison of Taxis, Transit, and Ridesourcing Services in San Francisco. *Transport Policy* 45 (01 2016), 168–178. <https://doi.org/10.1016/j.tranpol.2015.10.004>
- [53] Staff Reporter. 2017. No-Go Zones Declared As Attacks On Uber Drivers Intensifies. https://www.huffingtonpost.co.uk/2017/07/18/no-go-zones-declared-as-attacks-on-uber-drivers-intensifies_a_23034863/
- [54] Douglas Reynolds. 2009. *Gaussian Mixture Models*. Springer US, Boston, MA, 659–663. https://doi.org/10.1007/978-0-387-73003-5_196
- [55] Mah-jabeen Soobader, Felicia B LeClere, and Wilbur Hadden. 2001. Using aggregate geographic data to proxy individual socioeconomic status: does size matter? *American Journal of Public Health* 91, 4 (Apr 2001), 632–636. <https://doi.org/10.2105/ajph.91.4.632>
- [56] Larissa Stanberry. 1970. Permutation Test. https://link.springer.com/referenceworkentry/10.1007/978-1-4419-9863-7_1186
- [57] Martin Strobel. 2019. Aspects of Transparency in Machine Learning. In *Proceedings of the 18th International Conference on Autonomous Agents and MultiAgent Systems* (Montreal QC, Canada) (AAMAS '19). International Foundation for Autonomous Agents and Multiagent Systems, Richland, SC, 2449–2451.
- [58] Rosa Sulleiro. 2019. No Uber, no problem: Brazil app drives into Sao Paulo's no-go zones. <https://phys.org/news/2019-04-uber-problem-brazil-app-sao.html>
- [59] Latanya Sweeney. 2013. Discrimination in online ad delivery. *Queue* 11, 3 (2013), 10–29.
- [60] Uber. 2019. . 17–17 pages.
- [61] Uber. 2020. How are fares calculated? <https://help.uber.com/riders/article/how-are-fares-calculated/?nodeId=d2d43bbc-f4bb-4882-b8bb-4bd8acf03a9d>
- [62] Uber. 2020. How does Boost work? <https://help.uber.com/driving-and-delivering/article/how-does-boost-work?nodeId=e2e70956-42a0-4845-8d9c-746b9164e8a1>
- [63] Uber. 2020. How Uber's dynamic pricing model works. <https://www.uber.com/en-GB/blog/uber-dynamic-pricing/>
- [64] Uber. 2020. Set a Driver Destination. <https://help.uber.com/driving-and-delivering/article/set-a-driver-destination?nodeId=f3df375b-5bd4-4460-a5e9-afd84ba439b9>
- [65] Uber. 2020. Surge pricing. <https://marketplace.uber.com/pricing/surge-pricing>
- [66] Uber. 2020. Understanding UberPool fares and fees. <https://www.uber.com/en-GB/blog/london/london-pool-fares/>
- [67] Center for Economic Studies US Census Bureau. 2010. US Census Bureau Center for Economic Studies Publications and Reports Page. <https://lehd.ces.census.gov/data/>
- [68] Jerrold H Zar. 2005. Spearman rank correlation. *Encyclopedia of Biostatistics* 7 (2005).
- [69] Rich Zemel, Yu Wu, Kevin Swersky, Toni Pitassi, and Cynthia Dwork. 2013. Learning Fair Representations. In *Proceedings of the 30th International Conference on Machine Learning (Proceedings of Machine Learning Research)*, Sanjoy Dasgupta and David McAllester (Eds.), Vol. 28. PMLR, Atlanta, Georgia, USA, 325–333. <http://proceedings.mlr.press/v28/zemel13.html>
- [70] Zillow. 2020. Chicago IL Home Prices & Home Values. <https://www.zillow.com/chicago-il/home-values/>

9 APPENDICES

9.1 Calculating Combined Effect sizes

Equation 1 displays how combined effect sizes are calculated using a random effects approach. Taking the demographic neighborhood attribute “% Below Poverty Line” as an example, the combined effect size is a summary of the corresponding chart in Figure 6 (row 2 column 2). When $t = 30$, $d(X, Y, t)$ calculates the effect size on fare price of being picked up in a neighborhood with 30% of the population below the poverty line versus being picked up in a neighborhood with 70% of the population above the poverty line (see Equation 1). X_{t-} are the neighborhoods with a percentage of the population below the poverty line less than t , and X_{t+} are the neighborhoods with a percentage of the population below the poverty line greater than t . The combined effect size summarizes the effect sizes calculated between groups X_{t-} and X_{t+} on fare price Y as t is increased from the minimum to the maximum observed values of X .

Additional Datasets: Businesses, Traffic, Crashes, Transit, Groceries, Crime, and Arts/Entertainment in Chicago

Datasets regarding business, average daily traffic, crashes, subway stations, grocery stores, and crimes in Chicago were collected from the Chicago Data Portal [48]. All data in these datasets were used to count frequencies of businesses, cars, crashes, subway stations, grocery stores, and crimes in each census tract where corresponding data was available. No preprocessing steps were performed on these datasets.

The business license dataset contains 54,734 businesses across 824 census tracts in Chicago. The traffic dataset is a collection of 1,279 average daily traffic counts for the number of vehicles on streets at select locations in Chicago. The traffic dataset includes average daily counts for locations across 562 census tracts. The crash dataset lists information about each traffic crash to occur on city streets within Chicago, comprised of data for 429,415 crashes across 843 census tracts. The transit dataset contains a list of subway stops on Chicago’s “L” metro system, consisting of 300 stops across 112 census tracts. The grocery store dataset contains a list of grocery stores in Chicago, used to estimate how many Chicagoans live in food deserts, and lists 506 grocery stores across 342 census tracts.

The crimes dataset contains a list of all crimes that were recorded in Chicago from November 2018 to September 2019, the same time period as the trips in the ridehailing dataset. The crimes dataset consists of 237,426 crimes across 838 census tracts. The number of crimes in a neighborhood is positively correlated with fare prices in neighborhoods. The correlation between number of crimes and fare price, calculated using Pearson’s r is 0.47, with a p-value less than 10^{-4} , when examining neighborhoods where trips begin. The correlation between number of crimes and fare price, calculated using Pearson’s r is 0.49, with a p-value less than 10^{-4} , when examining neighborhoods where trips end.

Data regarding jobs and workers in the arts, food, and entertainment was collected from the Longitudinal Employer-Household Dynamics (LEHD) program of the U.S. Census Bureau [67]. The

number of jobs in the arts, food, and entertainment as well as the number of workers in the arts, food, and entertainment were collected using the LEHD variables NAICS 71 (Arts, Entertainment, and Recreation) and NAICS 72 (Accommodation and Food Services). The LEHD data contains counts for jobs and workers in arts, food, and entertainment for 1,314 census tracts in and around Chicago.

Extended Results: Fare Pricing

Figure 6 and Figure 7 display effect size charts of AI bias in fare pricing associated with trip frequency, ethnicity, age, education, poverty level, citizenship, and home value in Chicago neighborhoods. Figure 6 displays the effect sizes calculated for a neighborhood when it is the origin of a trip. Figure 7 displays the effect sizes calculated for a neighborhood when it is the destination of a trip.

Effect sizes of pickup and dropoff frequency on fare price suggest that neighborhoods with lower trip frequencies also have lower fares. Trip frequencies represent ride requests that were accepted by ridehailing applications, or instances where consumer demand was met. However riders whose trips were not accepted are not represented in this dataset, which is why trip frequency only accounts for a portion of demand. Still, the positive relationship between the visible demand in this dataset (trip frequency) and fare price, shows that neighborhoods with higher demand have higher fare prices. From Figures 6 and 7, charts “A = Pickup Frequency / m^2 ” and “A = Dropoff Frequency / m^2 ” show that effect sizes decrease as pickup and dropoff frequencies in a census tract increase. This means that the price is always lower for neighborhoods with lower pickup and dropoff frequencies when compared to neighborhoods with higher pickup and dropoff frequencies. This is consistent whether considering a neighborhood the origin of a trip, or the destination, shown by the similar trends for these two charts across both Figures 6 and 7.

Effect sizes of the non-white percentage in a neighborhood and fare price show that neighborhoods that on average, neighborhoods that are less white have higher fare prices than neighborhoods that are more white. Considering effect size chart “A = % Non-white” in Figures 6 and 7, effect sizes increase as the percentage of non-white population in a neighborhood increase to about 40%, and then effect sizes level out. However, because all effect sizes are negative, Figures 6 and 7 show that the difference between average fare price of a more non-white neighborhood is always higher than the average fare price of a less non-white neighborhood. This is consistent regardless of whether neighborhoods are considered as the origin or the destination of a trip.

Looking at effect size chart “A = % Older than 40” in Figure 6, almost all effect sizes are positive, meaning being picked up in neighborhoods with a younger population generally always costs more than being picked up in neighborhoods with an older population. The only case when neighborhoods with more older residents pay more are when the percentage of population older than 40 is less than approximately 10%. However the combined effect size (see Table 3) is negative, reflecting the overall trend of cheaper prices in neighborhoods with more residents over 40. The same results can be seen when being dropped off in neighborhoods with younger populations versus neighborhoods with older populations, shown in Figure 7.

Fare prices decrease as the percentage of population in a neighborhood that has a high school diploma or less increases. Fare prices increase as the percentage of population in a neighborhood below the poverty line increases. Effect size charts “% High School Education or Less” and “% Below Poverty Line” in Figures 6 and 7 have similar decreasing effect sizes regardless of whether examining these demographics at the origin or destination of a trip. However the effect sizes for “% High School Education or Less” are positive until $A \approx 0.70$. This shows that average fare price is generally more expensive for neighborhoods with high levels of college education, confirmed by Table 3. The effect size in the “% Below Poverty Line” chart becomes negative approximately when $A \approx 0.15$. This shows that average fare price is generally more expensive for neighborhoods with high levels of their population living below the poverty line, confirmed by Table 3. These trends are consistent for both demographic attributes (high school education and poverty level) regardless of whether the attributes are being considered at trip origin or destination.

Fare price does not change significantly with the percentage of U.S. citizens in a neighborhood. Fare prices decrease as the percentage of population living in homes below the median house price in a neighborhood increases. Effect size charts “% Non-U.S. Citizens” and “% Below Median House Price” in Figures 6 and 7 both show little variance regardless of the interval being examined. The effect sizes in “% Non-U.S. Citizens” chart remains at approximately ~ 0.0 , which is reflected in the combined effect sizes for the demographic (see Table 3), which are all insignificant. This means there is no significant association between the size of the immigrant population in a neighborhood and fare prices. The effect sizes in chart “% Below Median House Price” remains consistently positive, meaning that fare prices for neighborhoods with low home values are always lower than neighborhoods with high home values. Effect sizes for both demographic attributes (non-U.S. citizen, below median house price) are consistent regardless of whether the attributes are being considered at trip origin or destination.

Figure 8 and Figure 9 display the difference in average fare price per mile for two groups of neighborhoods at various demographic intervals. Figure 6 displays the differences in average fare price per mile for neighborhoods when they are the origin of a trip. Figure 7 displays the differences in average fare price per mile for neighborhoods when they are the destination of a trip. These figures can be used to determine the actual fare price difference per mile at specific demographic intervals. For example, in Figure 6, looking at the “ $A = \% \text{ Older than } 40$ ” chart, when $A = 0.20$ the average fare price per mile for neighborhoods with greater a than 20% population older than 40 is approximately \$2.50, represented in orange. The average fare price per mile for neighborhoods with less than a 20% population older than 40 is approximately \$3.00, represented in blue.

Extended Results: AI Bias on Fare Pricing with ACS Error Margins

ACS demographic statistics are estimated values, calculated using voluntary surveys taken throughout the United States [13]. In addition to demographic estimates, ACS also provides standard error margins indicating how far estimates may deviate from ground

truth demographic statistics. In order to account for the margin of error in ACS demographic estimates, all combined effect size calculations were repeated using minimum possible demographic estimates as well as maximum possible demographic estimate. Minimum possible demographic estimates were calculated by subtracting the error margin from each demographic statistic, and maximum possible demographic estimates were calculated by adding the error margin to each demographic statistic. These combined effect sizes for each demographic are shown in Table 3. Combined effect sizes calculated using minimum and maximum demographic estimates are shown in the “Min Est” and “Max Est” columns in Table 3.

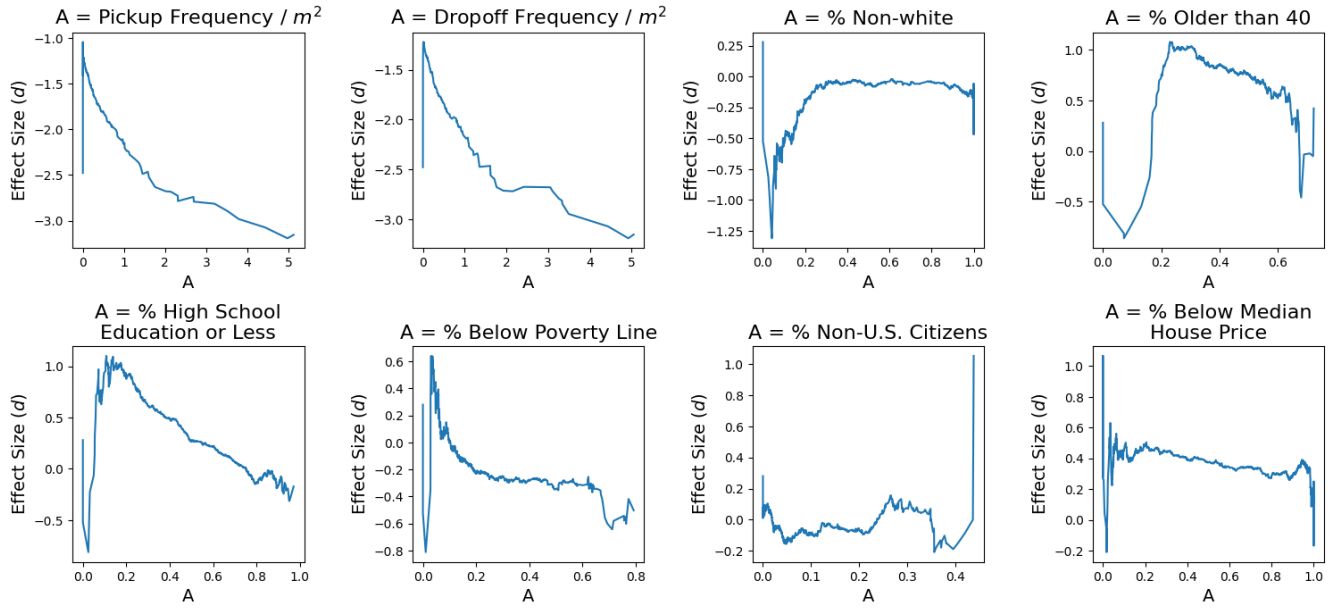


Figure 6: Effect sizes on fare price per mile at pickup location - Effect sizes on fare price per mile are calculated for pickup and dropoff frequencies per neighborhood. Effect sizes on fare price per mile are also calculated for the percentage of non-white, over 40, high school educated or less, below poverty line, and non-U.S. citizen populations, as well as the percentage of people living in homes below median home value in Chicago. Effect sizes are calculated using Cohen's d [23]. The raw fare prices generating these effect size charts can be observed in detail in Figure 8.

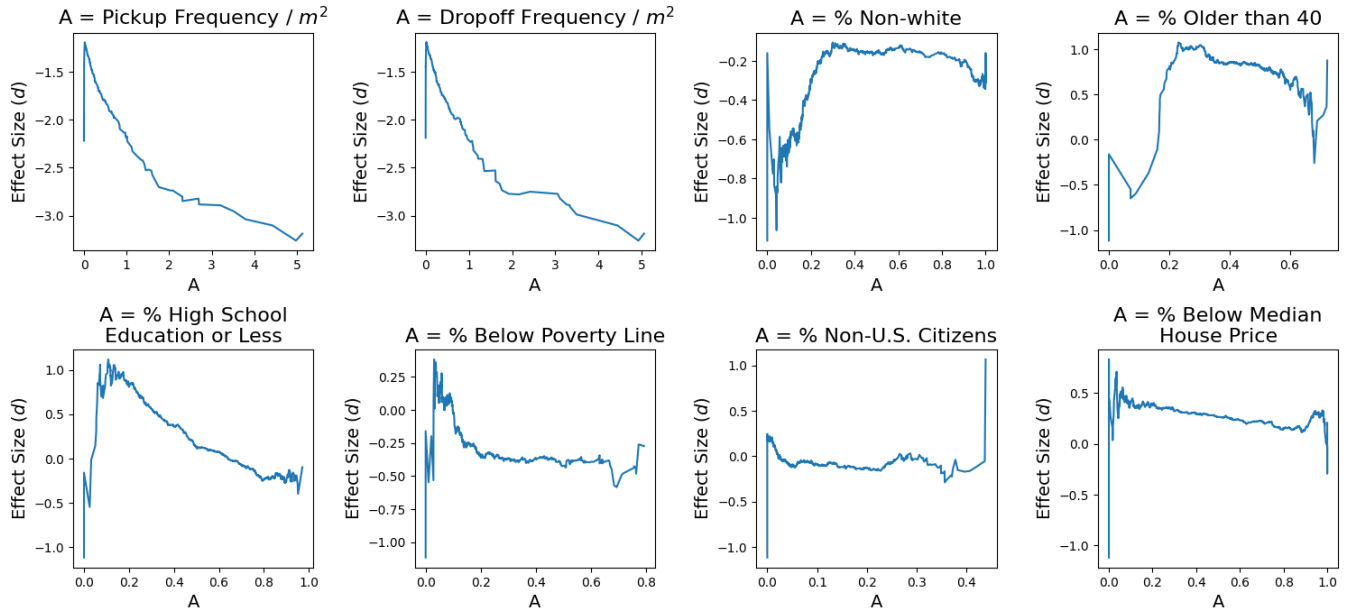


Figure 7: Effect sizes on fare price per mile at dropoff location - Effect sizes on fare price per mile are calculated for pickup and dropoff frequencies per neighborhood. Effect sizes on fare price per mile are also calculated for the percentage of non-white, over 40, high school educated or less, below poverty line, and non-U.S. citizen populations, as well as the percentage of people living in homes below median home value in Chicago. Effect sizes are calculated using Cohen's d [23]. The raw fare prices generating these effect size charts can be observed in detail in Figure 9.

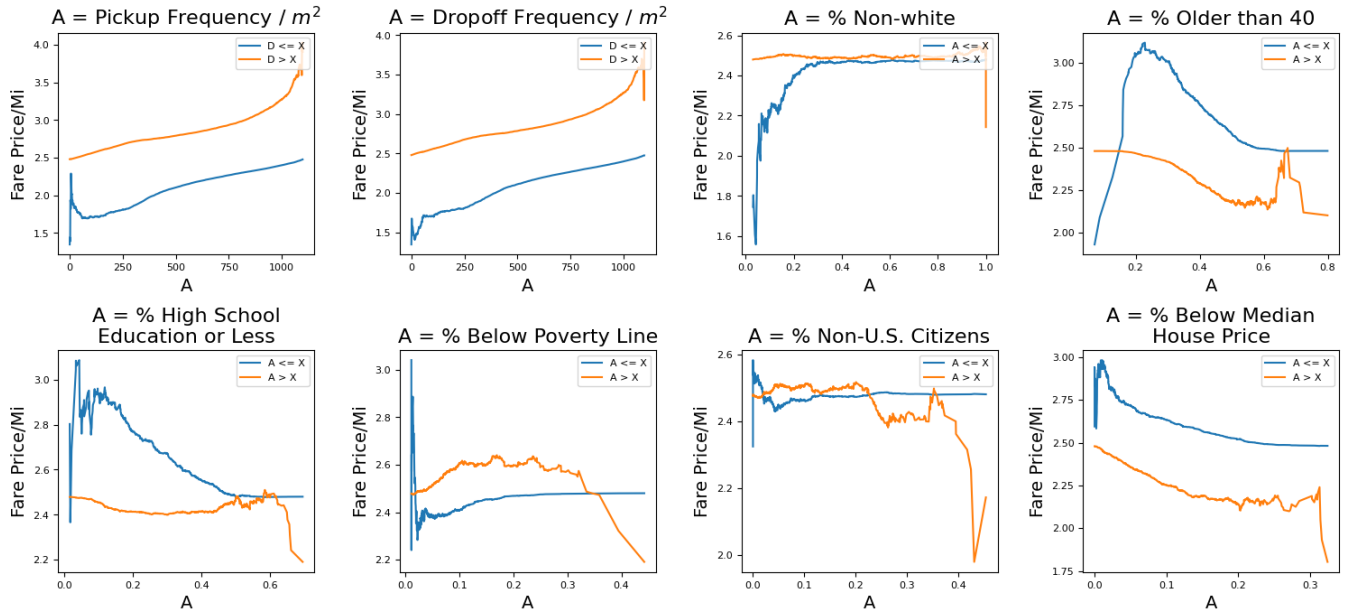


Figure 8: Raw price differences at pickup location - The change in fare price per mi for neighborhoods with $A \leq X$ and $A > X$ is shown as the value of neighborhood attribute A is increased. X refers to the values on the x axis of each chart. Raw price differences are calculated for pickup and dropoff frequencies per neighborhood. Raw price differences are also calculated for the percentage of non-white, over 40, high school educated or less, below poverty line, and non-U.S. citizen populations, as well as the percentage of people living in homes below median home value in Chicago.

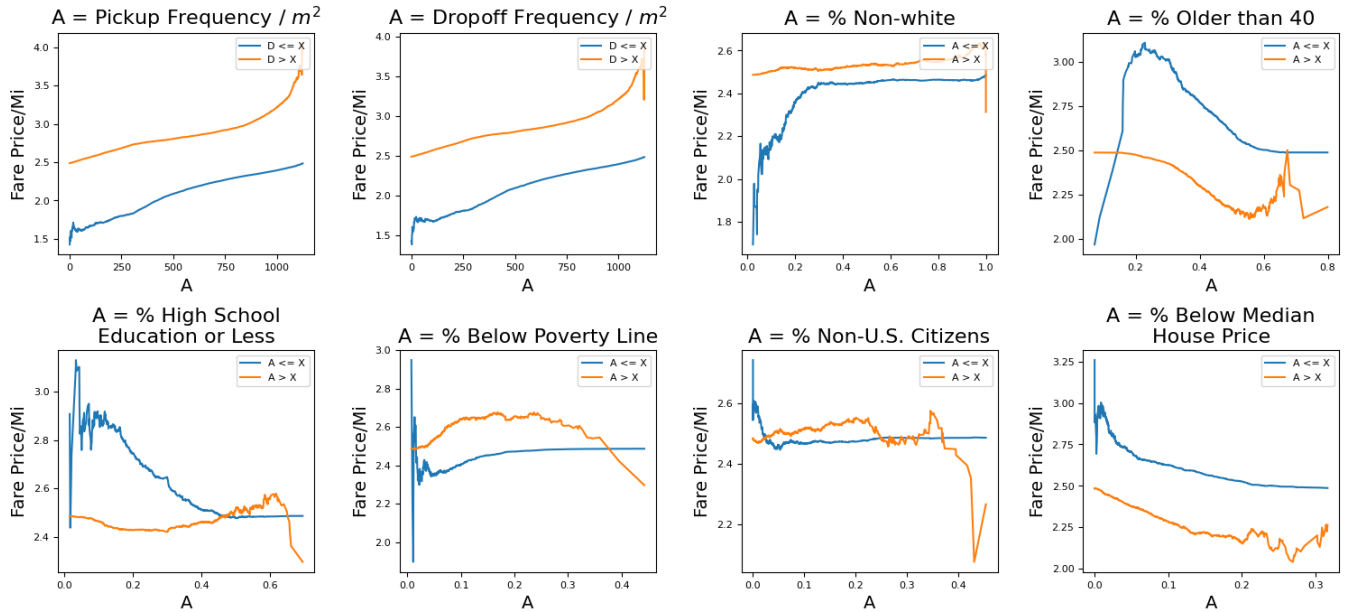


Figure 9: Raw price differences at dropoff location - The change in fare price per mi for neighborhoods with $A \leq X$ and $A > X$ is shown as the value of neighborhood attribute A is increased. X refers to the values on the x axis of each chart. Raw price differences are calculated for pickup and dropoff frequencies per neighborhood. Raw price differences are also calculated for the percentage of non-white, over 40, high school educated or less, below poverty line, and non-U.S. citizen populations, as well as the percentage of people living in homes below median home value in Chicago.

Pickup						
Attribute	Min Estimate		Estimate		Max Estimate	
	Combined	p	Combined	p	Combined	p
	Effect Size (d)		Effect Size (d)		Effect Size (d)	
Pickup Frequency / m^2	-1.57	$< 10^{-3}$	-1.57	$< 10^{-3}$	-1.58	$< 10^{-3}$
Dropoff Frequency / m^2	-1.57	$< 10^{-3}$	-1.57	$< 10^{-3}$	-1.57	$< 10^{-3}$
% Non-white	-0.13	0.20	-0.22	0.02	-0.23	0.02
% Older than 40	0.64	$< 10^{-3}$	0.66	$< 10^{-3}$	0.67	$< 10^{-3}$
% High School Education or Less	0.34	$< 10^{-3}$	0.24	0.01	0.30	$< 10^{-3}$
% Below Poverty Line	-0.11	0.31	-0.19	0.05	-0.08	0.45
% Non-U.S. Citizens	-0.07	0.51	-0.10	0.35	0.01	0.93
% Below Median House Price	0.24	0.01	0.23	0.02	0.27	0.02

Dropoff						
Attribute	Min Estimate		Estimate		Max Estimate	
	Combined	p	Combined	p	Combined	p
	Effect Size (d)		Effect Size (d)		Effect Size (d)	
Pickup Frequency / m^2	-1.59	$< 10^{-3}$	-1.59	$< 10^{-3}$	-1.59	$< 10^{-3}$
Dropoff Frequency / m^2	-1.57	$< 10^{-3}$	-1.57	$< 10^{-3}$	-1.57	$< 10^{-3}$
% Non-white	-0.29	$< 10^{-3}$	-0.32	$< 10^{-3}$	-0.35	$< 10^{-3}$
% Older than 40	0.66	$< 10^{-3}$	0.69	$< 10^{-3}$	0.70	$< 10^{-3}$
% High School Education or Less	0.25	$< 10^{-3}$	0.15	0.17	0.19	0.06
% Below Poverty Line	-0.18	0.07	-0.28	$< 10^{-3}$	-0.21	0.04
% Non-U.S. Citizens	-0.01	0.93	-0.07	0.54	0.03	0.79
% Below Median House Price	0.28	$< 10^{-3}$	0.19	0.06	0.19	0.06

Table 3: Combined effect sizes on fare price per mile by neighborhood attributes - combined effect sizes (column “Effect size”) are shown for the fare price per mile given a set of neighborhood attributes. “Pickup” and “Dropoff” columns designate fare price per mile when being picked up or dropped of in a neighborhood and “p” presents the p-value for effect size calculations. ACS data contains estimates for demographic statistics by census tract. Additionally, ACS provides a standard error margin for the demographic statistics estimates. Results calculated using the estimate with the error subtracted, as well as the estimate with the error added are shown in the “Min Estimate” and “Max Estimate” columns.