

# Advancing Fairness in Public Funding Using Domain Knowledge

Thomas Goolsby,<sup>1</sup> Sheikh Rabiul Islam,<sup>2</sup> Ingrid Russell<sup>3</sup>

University of Hartford<sup>1,2,3</sup>

goolsby@hartford.edu,<sup>1</sup> shislam@hartford.edu,<sup>2</sup> irussell@hartford.edu<sup>3</sup>

## Abstract

Artificial Intelligence (AI) has become an integral part of several modern-day solutions impacting many aspects of our lives. Therefore, it is of paramount importance that AI-powered applications are fair and unbiased. In this work, we propose a domain knowledge infused AI-based system for public funding allocation in the transportation sector by keeping potential fairness-related pitfalls in mind. In the transportation sector, in general, the funding allocation in a particular geographic area corresponds to the population in that area. However, we found that areas with high diversity index have a higher public transit ridership, and this is a crucial piece of information to consider for an equitable distribution of funding. Therefore, in our proposed approach, we use the above fact as domain knowledge to guide the developed model to detect and mitigate the hidden bias in funding distribution. Our intervention has the potential to improve the declining rate of public transit ridership which has decreased by 3% in the last decade. An increase in public transit ridership has the potential to reduce the use of personal vehicles as well as to reduce the carbon footprint.

## Keywords

domain knowledge, artificial intelligence, machine learning, federal funding, federal transit administration, public transportation, bias, fairness

## Introduction

Available public data establishes a set of criteria based on census data to determine how funding is tabulated and granted to federal transit agencies in major Urbanized Areas (UZAs) in the United States (Giorgis 2020). The current system takes into consideration a range of census-based criteria (Giorgis 2020) and supposed to take into consideration of protected attributes defined in Title VI of the Civil Rights Act of 1964 (Title VI 1964) among other determinants. This raises the question as to how and if it is possible to use AI-based systems to allocate federal funding in an equitable fashion while abiding by Title VI guidelines.

In this work, we investigate the federal allocation of funds for public transportation by keeping fairness issues in mind. When we talk about fairness in this paper, we are speaking to the mitigation of hidden bias that can be introduced inadvertently during the machine learning process. Ultimately, fairness in AI regarding this paper, looks to employ known techniques to eliminate hidden bias. Furthermore, the FTA is supposed to distribute public funds in an equitable fashion, as defined in Title VI of the Civil Rights Act of 1964, thus it is our goal to replicate that equity through using a machine learning approach that mitigates bias that may fabricate during the process. In the transportation sector, in general, the funding allocation in a particular geographic area corresponds to the population in that area. However, we found that areas with high diversity index have a higher public transit ridership and a crucial piece of information to consider for an equitable distribution of funding. Therefore, in our proposed approach, we use the above fact as domain knowledge to guide the developed model to detect and mitigate the hidden bias in funding distribution.

Domain knowledge is a high-level, abstract concept that encompasses the problem area. For example, in a car classification problem from images, the domain knowledge could be that a convertible has no roof, or a sedan has four doors, etc. However, encoding this domain knowledge in a black-box model is challenging. Bias can occur during data collection, data preprocessing, algorithm processing, or the act of making an algorithmic decision. Through the comparison of machine learning models with and without domain knowledge, this work measures the effectiveness of domain knowledge integration. We use different machine learning classifiers such as Random Forests (RF), Extra Trees (ET), and K-nearest neighbor, to name a few, for the experiments. We also use IBM AI Fairness 360 to detect and mitigate bias and evaluate different standard fairness metrics to further emphasize the effect of incorporating domain knowledge into our proposed approach.

In T. Kido, K. Takadama (Eds.), Proceedings of the AAAI 2022 Spring Symposium "How Fair is Fair? Achieving Wellbeing AI", Stanford University, Palo Alto, California, USA, March 21–23, 2022. Copyright © 2022 for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

## Background

A good amount of work has been conducted on the domain of bias and fairness in AI. Mehrabi et. al. developed a general survey exploring this topic. They emphasize the importance of a continuous feedback loop between data, algorithms, and users (Mehrabi et. al. 2021). This accentuates how susceptible AI algorithms are to bias. This bias can be introduced when data is collected. It's important to be aware of the kinds of bias that can occur as well.

Seeing as unique the interaction between data and users is, there are two biases in particular that apply to the data that we are working with. One of those biases is the omitted variable bias, which occurs when one or more important variables are left out of the model (Riegg 2008, Mustard 2003, Clarke 2005). A simple example of this type of bias in play could be with an algorithm that is trained to predict when users will unsubscribe from a company's service. A possible omitted variable here could be that a strong competitor enters the market that the algorithm was unaware of (Mehrabi et. al. 2021). The introduction of this competitor would be the omitted variable, which would then lead to bias being introduced in the algorithm when it tries to predict when a particular customer would unsubscribe. The other important form of bias is aggregation bias. Aggregation bias occurs when a one-size-fits-all model is used for groups with different conditional distributions (Suresh and Guttag 2019). Both the omitted variable bias and aggregation bias are unique in machine learning applications since they are technical biases that can occur at any point in the machine learning process. This leads to them being particularly difficult to counteract. The authors of this work discussed how the introduction of discrimination in AI is unique since it is a direct interaction between data and users. Again, domain knowledge is being used to attempt to counteract specific instances of bias like this.

Furthermore, it is important to understand the problematic nature of introducing racial categories to machine learning. Programmers face a unique dilemma in this problem domain since they can either be blind to racial group disparities or be conscious of those racial categories (Benthall et. al. 2019). However, regardless of which path the programmer chooses to go down, both options ultimately reify the negative and inaccurate implications of race in society. Moreover, this observes differences in races in the United States, which is inherently problematic. Race differences are created by ascribing race classifications onto individuals who were previously racially unspecified. This ultimately leads to the newly racially classified individuals being linked to stereotyped and stigmatized beliefs about

non-white groups (Omi and Winant 2014). With observing domain knowledge in the allocation of federal funds, we must be extremely cautious of these implications. Link and Phelan provide a clear definition of what stigma is. They define stigma as "the co-occurrence of labeling, stereotyping, separation (segregation), status debasement, and discrimination" (AI Fairness 360 2021). By understanding the systemic instillment of stigma in racial categories, this work will look for ways to introduce fair domain knowledge without reifying those dangerous stigmas. This ultimately leads to some implications in the development of a fair AI algorithm for allocating federal funds for public transportation.

Public transit agencies are supposed to abide by the Title VI of the Civil Rights Act of 1964. The Federal Transit Agency (FTA) follows closely with the rules written in Title VI which protects people from discrimination based on race, color, and national origin in programs and activities receiving federal financial assistance (Title VI 1964). Within this work, we also abide by these laws to develop a legally applicable AI for allocating federal funds, and investigate the disparities. A fair and unbiased AI algorithm for allocating federal funds for public transportation could further help combat the national decline in public transit ridership. William J Mallet from the Congressional Research Service emphasized that public transit ridership has declined nationally by 7% over the last decade (Mallett et. al. 2018). Competing transportation options like personal vehicles, ride-sourcing (e.g., Uber), and bike-sharing are partially at the forefront of the national decline. Some solutions proposed by this work are incentive funding, raising user fees on personal automobiles, and improving general funding for public transportation (Mallett et. al. 2018). That is where this work comes in; to attempt and answer the question of if an AI algorithm embedded with fairness can contribute to a more equitable solution.

## Experiments and Results

This project explores how domain knowledge can be integrated to ensure fairness in AI. A publicly available dataset on the allocation of federal funds to public transportation agencies is being used (Giorgis 2020). This dataset is the basis on which this exploration and application of machine learning is being used. The dataset includes official data from 2014-2019 on 449 Federal Transit Agency (FTA) defined public transportation agencies in the continental United States, Alaska, Hawaii, and Puerto Rico. The dataset is read into RStudio using R version 4.1.0 and Python version 3.8.0. The R programming language is being used in a simple R script while Python is being used in

isolated code chunks within an R markdown (Rmd) file. For bias detection and mitigation, we use IBM AI 360 fairness open-source toolkit (AI Fairness 360 2021).

### Data Preprocessing

This dataset (Giorgis 2020) is being preprocessed into a summarized form which gives totals for individual transit agencies per year. The data started off with 42 columns and 36,656 rows. Empty columns and rows are deleted which then leads to the dataset containing 40 columns and 18,673 rows. The overall dataset is then split up into separate data containers for individual years; thus, producing six separate datasets for six individual years (2014-2019). Each of the six datasets contains 13 columns and anywhere from 440-444 rows depending on the year. Finally, separate data containers are combined back into a single data container which now consists of summarized data for every given FTA UZA per year. This summarized data container containing all data from 2014-2019 has 13 columns and 2,615 rows.

Furthermore, the measure of operating expenses is converted to classes in which supervised machine learning can take place. Operating expense classes are determined by examining the distribution of operating expenses across transit agencies. It was found that the distribution was skewed towards the lower end (< \$100,000,000). However, it is also found that the total amount of operating expenses for a specific transit agency has a high correlation, roughly 95%, with the population of its service area. These are the factors that lead to the current distribution of operating expense level classes. Data is currently being utilized from the 2020 national census, specifically diversity indices at the state and county level. Data engineering techniques are being used to incorporate both state and county-level diversity indices into the summarized public funds' allocation dataset.

To evaluate the fairness of the models with domain knowledge, diversity index by county had to be sorted into classes. Diversity index by county is being used as the primary form of domain knowledge here since it provides a clearer vision of the diversity across populations. The diversity index serves as a measure of how likely it is that two individuals chosen at random from a population are from different races and ethnic groups (Bureau et. al. 2021). The diversity index is bound between 0 and 1 where a 0-value indicates that everyone in the population has the same racial and ethnic characteristics. While a value closer to 1 indicates that everyone in the population has different racial and ethnic characteristics (Bureau et. al. 2021). Therefore, we observe diversity index by county for each of the 449 FTA-defined public transportation agencies, and found it as

an effective incorporation of census-based domain knowledge. To convert the diversity index by county into class, the distribution of the values is being evaluated. As seen in Figure 1, there is a great number of observations (roughly 55%) that have a diversity index between 0.25 and 0.5.

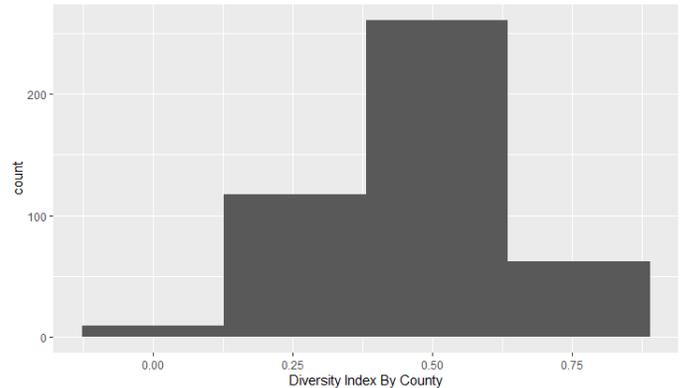
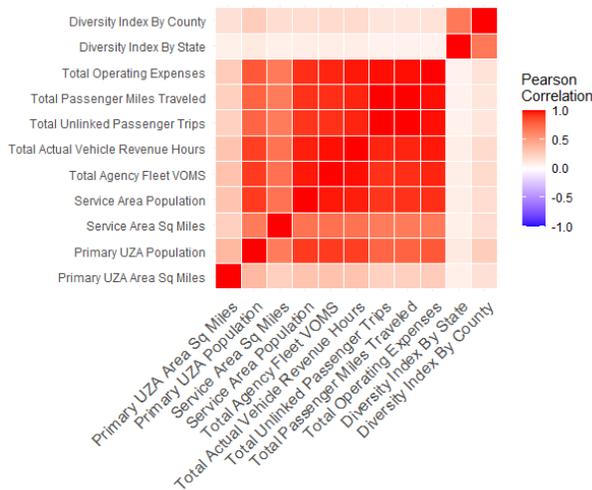


Figure 1: Histogram of diversity index by county

Therefore, since the distribution looked as such with 4 bins, the diversity index by county was split into 4 classes. The first class is “Very Low”, which constitutes all observations that have a diversity index greater than or equal to 0 and less than 0.25. The next class was “Low” which is made up of observations that have a diversity index greater than or equal to 0.25 while also less than 0.5. Then the “Moderate” class includes all observations that have a diversity index greater than or equal to 0.5 and less than 0.75. Finally, the last class was “High” which included the remaining observations, or those that have a diversity index that is greater than or equal to 0.75 and less than 1 (since this is the maximum value possible). These class bounds are also supported by the fact that the diversity index by county had the largest correlation with the population of a particular UZA. It’s found that the correlation between these two values is 0.26, which again was the highest correlation that diversity index by county had with any other variable in the data set (see Figure 2). Furthermore, one of the variables that has the highest correlation with primary UZA population is unlinked passenger trips (0.76). Total unlinked passenger trips serves as an FTA defined measure of public transportation ridership. Therefore, we can see the relation here that urban areas with higher population tend to have higher public transit ridership as well as a higher diversity index by county.



**Figure 2: Heatmap of all numeric variables in data set**

Furthermore, 7 out of the top 10 UZA is from the top 10 diverse states (Jensen et. al. 2021) – Hawaii, California, Nevada, Maryland, District of Columbia, Texas, New Jersey, New York, Georgia, and Florida (Jensen et. al. 2021).

**Table 1. Top 10 UZA with the highest ridership**

New York-Newark, <b>NY-NJ-CT</b>
Los Angeles-Long Beach-Anaheim, <b>CA</b>
Chicago, <b>IL-IN</b>
Washington, <b>DC-VA-MD</b>
San Francisco-Oakland, <b>CA</b>
Boston, <b>MA-NH-RI</b>
Philadelphia, <b>PA-NJ-DE-MD</b>
Seattle, <b>WA</b>
Miami, <b>FL</b>

Although the diversity index of a county has the highest correlation (.26) with the population of UZA, it has a comparatively low correlation (.14) with total operating expenses in that area. This finding encourages us to develop an equitable distribution technique.

**Model Creation**

Both the R and Python programming languages are being used to create machine learning models on the dataset. R is primarily being used to preprocess the dataset while Python is being used to develop classification models using a 70/30 training and test set split. Random forest, extra trees, and k-nearest neighbor models without domain knowledge (i.e., without considering diversity index) are being developed and analyzed. The scikit-learn package is being used to develop Python-based supervisor learning model (Random

Forest), while the class package is being used to develop the k-nearest neighbor algorithm in R. For the models without domain knowledge, 12 columns are being used. The 11 predictors are all numeric values and some of the variables include Primary UZA Population, Total Unlinked Passenger Trips, and Total Passenger Miles Traveled to name a few. These 11 predictors are being used to predict Total Operating Expenses, which serves as a general measure of how much money a specific FTA transportation agency is receiving/spending. The models with domain knowledge have 12 predictors: the same 11 predictors as the models without domain knowledge, plus our variable representing Diversity Index by County employed as domain knowledge. The goal of measuring the accuracy, precision, recall, and ROC performance metrics was to take a trivial look as to if incorporating domain knowledge into some simple classification models will drastically affect those values. As seen in Table 1, the accuracy, precision, recall, and ROC metrics are calculated, each of which has a value of 0.99X. The metrics with domain knowledge (i.e., after incorporating diversity index as encoded domain knowledge) deviated only slightly from the metrics produced by the models without domain knowledge. The largest difference between metrics of models with and without domain knowledge can be seen in the K-Nearest Neighbor models. The average difference between models without domain knowledge minus the models with domain knowledge is 0.00265. This difference is negligible and expected considering the overall societal impact from it.

**Table 2. Accuracy precision, recall, and ROC metrics for Random Forest models w/ and w/o domain knowledge**

Random Forest	Without Domain Knowledge	With Domain Knowledge
Accuracy	0.99492	0.99490
Precision	0.99488	0.99501
Recall	0.99492	0.99490
ROC	0.99998	0.99998

**Table 3. Accuracy precision, recall, and ROC metrics for Extra Trees models w/ and w/o domain knowledge**

Extra Trees	Without Domain Knowledge	With Domain Knowledge
Accuracy	0.99619	0.99618
Precision	0.99617	0.99622
Recall	0.99619	0.99618
ROC	0.99998	0.99998

**Table 4. Accuracy, precision, recall, and ROC metrics for K-Nearest Neighbor models w/ and w/o domain knowledge**

K-Nearest Neighbor	Without Domain Knowledge	With Domain Knowledge
Accuracy	0.99111	0.98854
Precision	0.99099	0.98849
Recall	0.99111	0.98854
ROC	0.99952	0.99655

**Fairness Evaluation Preprocessing**

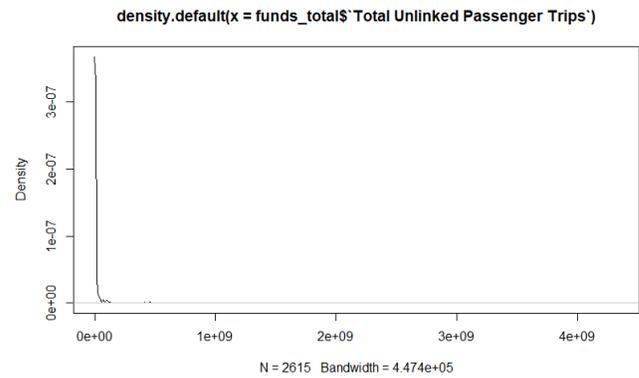
For evaluating fairness in the models with domain knowledge, the IBM AI 360 tool is being used. We use the R package of this tool for our experiment. To begin the process of evaluating fairness, the data set needs to be converted into a binary representation of itself. The most important columns are being chosen to be present in the fairness evaluation. These variables were deemed the most important since they all presented the highest correlation with the variable being predicted: Total Operating Expenses. Furthermore, these variables are all numeric values which are imperative to the development of classification models that can be evaluated using the IBM AI 360 (AI Fairness 360 2021). Considering that all these variables are numeric values, it is much clearer where to set bounds when converting variables to binary representations.

This includes the population of the UZA in which transit agencies exist in, total unlinked passenger trips, year, and operating expense level. Since the operating expense level is already broken into classes (Low, Medium, High), a separate column is made for each. For example, there is one column labeled “Operating Expense Level Low”, which has a 1 in this column in the operating expenses are categorized as "Low" and a 0 in every other row. A little more nuance is being taken to convert the UZA population and total unlinked passenger trips columns to binary representations. The density of both these variables shows a heavy concentration of observations at the lower end (Figures 3 & 4).

Since both these variables have such many observations near the lower end of the range, the ranges for the classes are being chosen to reflect this trend. For UZA population, three classes are being created to split this column into a binary representation. The following is the range for each class for the UZA population:

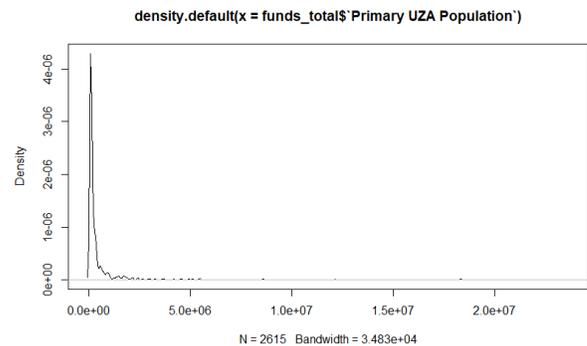
- Low: population [0, 250]
- Medium: population [250K, 1M]
- High: population [1M, MAX]

**Figure 3: Density plot for Primary UZA Population**



**Figure 4: Density plot for Total Unlinked Passenger Trips**

A very similar idea is being used to split up total unlinked passenger trips into classes. The National Transit Database (NTD) and the FTA provided the explain that unlinked passenger trips are the number of boardings on public transportation vehicles in a fiscal year for a specific transportation agency (Federal Transit Administration 2021). Transit agencies must count each passenger that boards their vehicles, regardless of how many vehicles the



passenger boards from origin to destination (Federal Transit Administration 2021). Similar to previous variables, 3 classes are being created with the following ranges for each:

- Low: total unlinked passenger trips [0, 5M]
- Medium: total unlinked passenger trips [5M, 100M]
- High: total unlinked passenger trips [100M, MAX]

The Year variable is also being split up into a binary representation. The year in this data set ranges from 2014 to 2019. Thus, a separate column for each year is being made where a value of 1 means the specific observation is from that year. The last variable that is converted to a binary representation is, of course, the diversity index by county. Simply, for this column, a value of 1 is given if the diversity

index is categorized as “Moderate” or “High” and a value of 0 is the diversity index is categorized as “Very Low” or “Low”. Figure 5 provides a snapshot of the data after all variables are done being converted to binary representations. The data set still has 2,615 rows, however, the binary data set has 16 columns.

High Diversity Index by County	High Operating Expenses	Medium Operating Expenses	Low Operating Expenses	High Primary UZA Population	Medium Primary UZA Population
1	0	0	1	0	0
0	1	0	0	0	1
1	1	0	0	1	0
0	1	0	0	1	0
1	0	1	0	0	0
0	1	0	0	0	1
1	1	0	0	0	1
0	1	0	0	1	0
1	1	0	0	1	0
0	1	0	0	1	0

Figure 5: Snapshot of the binary representation of data

### Fairness Metrics Calculation

We create a new R script to calculate the desired fairness metrics. A simple definition of a fairness metric, as provided in the documentation of the IBM AI 360 tool, is a quantification of unwanted bias in training data or models (AI Fairness 360 2021). The fairness metrics that are being evaluated in this project are *statistical parity difference*, *disparate impact*, *equal opportunity difference*, and the *Theil index*. A brief definition for each observed fairness metric is as follows:

- *Statistical parity difference*: the difference in the rate of favorable outcomes received by the unprivileged group to the privileged group.
- *Disparate impact*: the ratio of the rate of a favorable outcome for the unprivileged group to that of the privileged group.
- *Equal opportunity difference*: the difference of true positive rates between the unprivileged and the privileged groups.
- *Theil index*: measures the inequality in benefit allocation for individuals.

These four fairness metrics were chosen based on the information provided on the IBM AI 360 tool. Furthermore, these four metrics specifically evaluate privileged versus unprivileged groups in terms of individual and group fairness. Regarding this project, we are looking at the

distribution of funds between FTA transportation agencies that are based in a county with a high diversity index ( $\geq 0.75$ ). By observing these specific fairness metrics, we can see how favorable outcomes, or higher federal funding, may be unequally distributed among privileged and unprivileged groups.

Furthermore, we chose to employ the IBM AI 360 toolkit as it provided a compact and efficient collection of fairness evaluation libraries. The problem area of the project is perfectly encapsulated in the recommended uses of the toolkit. The creators of the IBM AI 360 toolkit explain that the toolkit should be used in very limited settings, one of which is allocation assessment problems with well-defined protected attributes (AI Fairness 360 2021). This project’s problem area deals with allocation of funds. Moreover, and more importantly, the dataset being used for the fairness evaluation has a well-defined protected attribute, which is diversity index by county, since as we have explained earlier in the paper, has unintentional bias defined by the FTA and protected by Title VI of the Civil Rights Act of 1964.

The reweighing function is our tool of choice in the IBM AI 360 toolkit as it assigns weights to training set tuples instead of changing class labels (Kamiran et. al. 2012). This is favorable since we want to analyze how diversity index by county plays a role in the mitigation of bias in this problem.

In the R environment, the “aif360” library is being used, which includes all the metrics and capabilities provided by the IBM AI 360 project. The project library is loaded into the R environment and the binary data set from Figure 5 is also loaded in. To run any metric calculations with this library, any R data frames must be converted into an *aif* data set, which asks for the protected attribute, the privileged (i.e., reference group) and unprivileged value for the protected attribute, and the target variable. For our case, the target variable is the “Operating Expense Level High” column. To reiterate, a value of 1 is given in this column if the observation is considered to have “High” operating expenses, or operating expenses of more than \$1,000,000,000. The protected attribute in this project is the *diversity index by county* column that was added as a piece of domain knowledge. To capture the nature of the protected attribute, the privileged group are observations that have a value of 0, or “Very Low” and “Low” diversity indices, and the unprivileged group are observations that have a value of 1, or “Moderate” and “High” diversity indices.

The IBM AI 360 library uses underlying classification models to help develop and calculate fairness metrics. Since the IBM AI 360 library uses classification models, we need two data sets to compare the true data with the predicted data. Thus, we have one *aif* data set that is the

raw binary data, and another that is nearly identical, however, the “Total Operating Expenses High” variable was predicted by a simple logistic regression model (this is called the newly classified dataset). The reweighing technique (Kamiran et. al. 2012, Aif360 2021), which modifies the weights of different training examples, is being used to help mitigate any bias that is present in this project. The IBM AI 360 tool includes a reweighing option that modifies the weight of different training instances. The reweigh algorithm is being applied to both the original binary data set as well as the classified data set. Once both data sets are reweighed, the fairness metrics can be calculated and compared to the original data. Graphs are being produced to show the difference and improvement after bias is mitigated through reweighing. Figures 6, 7, 8, and 9 show the comparison of fairness metrics between the original data and the reweighed data that has bias mitigated.

Calculating all four desired fairness metrics shows that mitigating bias through reweighing leads to either metrics being the same, or slightly improving the value. As seen in all graphs, both the original data and the mitigated data are within the fair range. Statistical parity difference (i.e., discrimination) was reduced to .035 from .051 using domain knowledge (see Figure 6). Statistical parity, also called demographic parity, ensures each group has an equal probability of being assigned to the positive predicted class.

By mitigating bias, we can produce fairness metrics that are closer to true fairness, which is a value of 0 for statistical parity difference, equal opportunity difference, and Theil index, and a value of 1 for disparate impact. Currently, we are infusing the diversity index as domain knowledge. However, in the future, we would also like to investigate the infusible domain knowledge more by examining other criteria such as native language spoken, and family income.



Figure 6: Statistical parity difference of original vs mitigated data

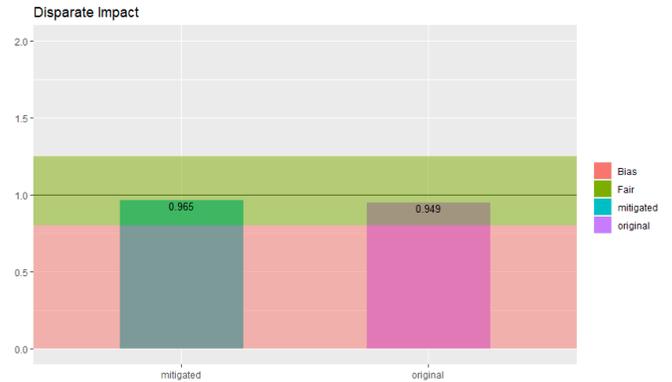


Figure 7: Disparate impact of original vs mitigated data

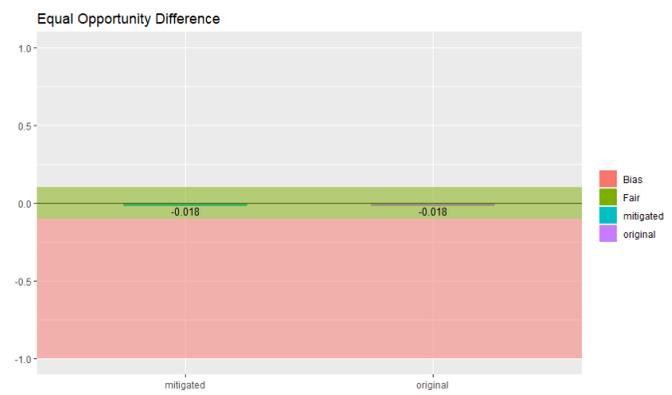


Figure 8: Equal opportunity difference of original vs mitigated data

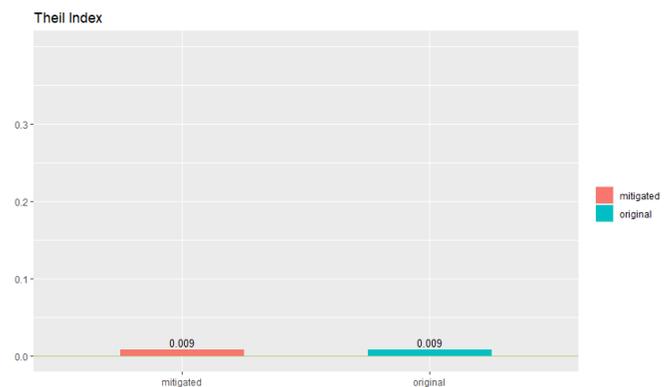


Figure 9: Theil index of original vs mitigated data

## Contributions and Future Works

By investigating the implications of domain knowledge on creating fair decision-making, this work explores how true fairness in AI can be achieved within the application of public funding allocation. This work investigates how

federal agencies like the FTA could apply AI in the process of allocating funds. In general, the allocation of FTA funds corresponds to the population in an area (i.e., UZA). However, it is found that areas with a higher diversity index have higher public transport ridership. Our proposed domain knowledge infused approach can reduce statistical parity difference which helps to ensure each group has an equal probability of being assigned to the positive predicted class. Finding the right domain knowledge is very challenging. Going forward, we want to incorporate and investigate the impact on other protected variables (e.g., native language spoken, family income), and find a way to enhance the infusible domain knowledge that reduces different disparities. An increase in public transit ridership has the potential to reduce the use of personal vehicles as well as to reduce the carbon footprint. A quantitative analysis of this possibility could be another direction of research.

## References

- [1] Giorgis, J. D. (2020). *Federal Funding Allocation* [Dataset]. United States Department of Transportation- FTA Federal Funding Allocation Since 2014, from <https://catalog.data.gov/dataset/federal-funding-allocation>
- [2] Title VI of Civil Rights Act of 1964, Pub.L. 88-352, 78 Stat. 241 (1964).
- [3] Mehrabi, N., Morstatter, F., Saxena, N., Lerman, K., & Galstyan, A. (2021). A survey on bias and fairness in machine learning. *ACM Computing Surveys (CSUR)*, 54(6), 1-35.
- [4] Benthall, S., & Haynes, B. D. (2019, January). Racial categories in machine learning. In *Proceedings of the conference on fairness, accountability, and transparency* (pp. 289-298).
- [5] Mallett, W. J. (2018, March). *Trends in public transportation ridership: Implications for federal policy* (No. R45144). Congressional Research Service.
- [6] Bureau, U. S. C. (2021, October 14). Racial and ethnic diversity in the United States: 2010 census and 2020 census. *Census.gov*. Retrieved November 22, 2021, from <https://census.gov/library/visualizations/interactive/racial-and-ethnic-diversity-in-the-united-states-2010-and-2020-census.html>
- [7] Jensen, E., Jones, N., Rabe, M., Pratt, B., Medina, L., Orozco, K., & Spell, L. (2021, August 12). *The chance that two people chosen at random are of different race or ethnicity groups has increased since 2010*. *Census.gov*. Retrieved November 24, 2021, from <https://www.census.gov/library/stories/2021/08/2020-united-states-population-more-racially-ethnically-diverse-than-2010.html>
- [8] Kamiran, F., & Calders, T. (2012). Data preprocessing techniques for classification without discrimination. *Knowledge and Information Systems*, 33(1), 1-33.
- [9] *Aif360 Algorithms Preprocessing Reweighing*. [aif360.algorithms.preprocessing.Reweighing](https://aif360.algorithms.preprocessing.Reweighing) - aif360 0.4.0 documentation. (n.d.). Retrieved November 24, 2021, from <https://aif360.readthedocs.io/en/latest/modules/generated/aif360.algorithms.preprocessing.Reweighing.html>.
- [10] Riegg, S. K. (2008). Causal inference and omitted variable bias in financial aid research: Assessing solutions. *The Review of Higher Education*, 31(3), 329-354.
- [11] Mustard, D. B. (2003). Reexamining criminal behavior: the importance of omitted variable bias. *Review of Economics and Statistics*, 85(1), 205-211.
- [12] Clarke, K. A. (2005). The phantom menace: Omitted variable bias in econometric research. *Conflict management and peace science*, 22(4), 341-352.
- [13] Suresh, H., & Gutttag, J. V. (2019). A framework for understanding unintended consequences of machine learning. *arXiv preprint arXiv:1901.10002*, 2.
- [14] Omi, M., & Winant, H. (2014). *Racial formation in the United States*. Routledge.
- [15] AI Fairness 360. Retrieved November 25, 2021, from <https://aif360.mybluemix.net/>
- [16] Federal Transit Administration Office of Budget and Policy. (2021, December 13). *National Transit Database 2021 policy manual*. Retrieved January 24, 2022, from [https://www.transit.dot.gov/sites/fta.dot.gov/files/2021-12/2021-NTD-Reduced-Reporting-Policy-Manual\\_1-1.pdf](https://www.transit.dot.gov/sites/fta.dot.gov/files/2021-12/2021-NTD-Reduced-Reporting-Policy-Manual_1-1.pdf)