On-Time Performance and Transit Accessibility: A LeeTran Case Study

Jarius Hamid

UIN: 815227356

September 4th, 2024

Abstract

Access to transit has a significant impact on local communities as highlighted in the existing literature. Public transit is particularly crucial in serving underprivileged communities, where on-time performance (OTP) plays a vital role in ensuring reliable service and access. This analysis serves two main purposes. First, it provides a daily overview of LeeTran's route performance and offers suggestions for improvement. Second, it examines how socio-economic factors, such as median household income, are key determinants in deciding where to place bus stops. Using data analysis and regression modeling, I examine daily OTP at LeeTran and perform a cross-sectional analysis to explore how socio-economic factors, particularly median household income, influence transit access and steer bus stop placement toward low-income communities. Finally, I will propose potential solutions and policy recommendations for LeeTran to enhance its operations moving forward.

Keywords: Public Transit, Transit Accessibility, Data Analysis, Regression Modeling
JEL Classification: R41, R42, C10, C21

Contents

1 Introduction

The United States is known for being heavily reliant on personal vehicles and has a car-centric mindset to its infrastructure and transit solutions. This is a problem for a variety of reasons, the most important being inequality and inefficiency. Poverty affects an estimated 37 million Americans, and in a time where citizens are living paycheck to paycheck, transit costs can be substantial. For many low-income individuals, spending money on transportation costs is not realistic when other essential costs such as food, housing, and healthcare are more important. Due to this fact, many individuals rely on the current public transit infrastructure despite the fact it is not broadly developed.

Access to public transit is crucial for serving low-income communities across the United States. Many of these communities lack reliable vehicles and rely on public transit for daily activities such as commuting to work, shopping, running errands, and recreational activities. In Lee County, the importance of having robust public transit infrastructure is evident, but it remains underdeveloped compared to larger transit systems. When comparing transit in Lee County to larger transit agencies, there is a stark difference in both riders served and infrastructure. In more developed transit systems, it is not only utilized by low-income communities but also by the broader population as a viable alternative to personal vehicles. This highlights that there is a broader need for improved public transit access and infrastructure.

Globally we can see this difference as well. Transit typically accounts for 10 to 20 percent of urban trips in Western Europe, but only 2 percent in the United States (National Research Council, 2002). The European countries have more robust infrastructure to support transit agencies making living without a vehicle possible for many citizens. Comparing transit systems globally and studying other agencies is crucial for developing an efficient network that can serve the needs of communities. However, before we can see public transit flourish in America many political, financial, and social burdens must be overcome.

In this paper, I will use LeeTran as a Case study to conduct a dual analysis. First, to explore the on-time performance for their routes. Second, to analyze how transit accessibility is related to various socio-economic factors and discuss the crucial role on-time performance plays in delivering a reliable service to the communities served. This should also underscore the importance of robust transit infrastructure in ensuring accessibility and maintaining a dependable transit system.

1.1 About LeeTran

LeeTran is a department of the Lee County government and proudly provides public transportation, ADA paratransit service, Transportation Disadvantaged services and an employer vanpool program operated by Commuter Connect and the Florida Department of Transportation to the community of Lee County. LeeTran was established in 1975 and was created with the purpose of addressing the growing transportation needs of the community. They have an established network of bus routes covering much of Lee County including Fort Myers, Cape Coral, and Lehigh Acres. Some of the services LeeTran provides include:

- **Fixed Route:** The core of LeeTran's services, consisting of scheduled bus routes that run on regular intervals and follow predetermined paths.
- **Demand Response:** A service for individuals with disabilities or those who require specialized transportation. This service is typically reserved in advance and offers flexible routing.
- **Movement on Demand:** A service for individuals within MOD zones providing door-to-door transit service and also by connecting passengers to LeeTran fixed bus routes for continued journeys within Lee County.
- **Trolley Services:** In some tourist areas and specific neighborhoods, LeeTran operates trolley services that offer a more scenic and community-focused transit experience.
- **Express Routes:** These routes provide quicker travel between major hubs with fewer stops, aimed at reducing commute times for passengers.

The main focus for this report will be on fixed route, as this is the majority of LeeTrans network, and serves most of the local community. This is also the source for many of the challenges LeeTran faces. Fixed route on-time performance faces challenges due to factors such as traffic, weather, and operational issues. These difficulties are compounded by constraints relating to funding, budget, and political factors. This ultimately hinders the ability to expand services, maintain infrastructure, and make improvements.

2 Literature Review

The existing literature highlights the importance of transit accessibility and supportive infrastructure. It also shows that transit systems can increase access to jobs, education, and services (APTA, 2017). Leading to a significant positive impact on local economies as well, driving growth and development (FTA, 2010). This means greater economic mobility for residents, reduced transportation costs, and improved quality of life, making transit accessibility, infrastructure, and maintaining reliability vital components of sustainable urban development.

2.1 Transit Infrastructure

Transit infrastructure refers to the systems and facilities that support public transportation, including railways, bus routes, stations, and related amenities that enable efficient movement of people. Transit infrastructure must be developed in order to achieve improved socio-economic outcomes. Currently, the United States is far behind other countries in terms of transit infrastructure. "The transit systems in many European countries demonstrate the effectiveness of integrated public transportation networks in reducing traffic congestion and enhancing economic productivity, a contrast to the fragmented systems often seen in the United States" (National Research Council, 2002). The United States' fragmented transit infrastructure leads to a heavy reliance on gas-powered vehicles, resulting in a less efficient system.

2.2 Transit Accessibility

Transit accessibility refers to how easily riders and the public can reach and use public transportation services. Accessibility is important because public transit not only provides a service to those who rely on it, but also enhances equity, social inclusion, environmental benefits, and improved quality of life (Esri, 2021). Other studies show that improved transit accessibility also has additional effects. Litman (2021) finds that "quality public transportation and transit-oriented development lead to significant health benefits by reducing traffic accidents and pollution, promoting physical fitness, enhancing mental health, improving access to medical care and healthy food, and increasing affordability for lower-income households, thereby alleviating financial stress." Additionally, Lucas et al. (2016) describes a phenomenon where the lack of available or inadequate public transport services can lead to social exclusion of low-income groups and communities. Public Funding issues contribute to decreased accessibility as well. Insufficient and inconsistent funding for public transport leads to service cuts, which diminishes accessibility (Forum, 2021).

2.3 Transit Reliability

All existing literature on addressing socio-economic concerns through transit highlights another critical aspect: the importance of maintaining reliable transit systems. The main measure of transit reliability is on-time performance, this reflects the effectiveness of the service (such as a bus or train) in adhering to its published schedule. It is an important aspect because it directly impacts customer satisfaction, ridership, and the effectiveness of your transit network (Mishra et al., 2023). Although there are other aspects of transit performance the cost for each varies. The cost of unreliable service may actually be greater than other costs including travel time (Bartee, 2007).

2.4 Empirical Assessments

Looking into effects on performance, it is influenced by the specific circumstances of each transit agency. Gonzalez et al. (1993) shows that the probability of on-time failures increases during PM peak periods, with longer headway and higher levels of passenger activity, and as buses progress further along their routes. While Gonzalez's study analyzes on-time performance (OTP) by specific time periods, this study utilizes daily data. It is worth noting that time-period data could enhance the analysis. Empirical Analyses on accessibility have similarly been done to assess its effect in an area on various socio-economic variables. Baker et al. (2019) found solid evidence of the causal effects of bus access on poverty and unemployment. In contrast, other research indicates that economic changes vary significantly across different cities. Overall, the existing literature on the subject highlights the importance of various aspects of transit like infrastructure, accessibility, and reliability. It also highlights how accessibility could affect socio-economic factors, as well as how this varies by agency and geographic location.

3 Data Overview and Methodology

The exploratory variables were sourced from multiple datasets: the OTP dataset and the Census Tract dataset. All data was cleaned, wrangled, processed, and visualized using R and various packages.

3.1 OTP Data Summary

The OTP data encompasses daily aggregates by route, as well as daily aggregates for other variables that are not broken down by route, covering the period from October 7, 2022, to June 30, 2024. The majority of this data came internally from the LeeTrans database, with the exception of accident counts which were collected via a data request from the Florida Department of Transportation, and rain and temperature data coming from Weather Underground. Weather Underground is a commercial weather service providing real-time weather information over the Internet. The LeeTran database contains all pertinent data through Clever Software and includes detailed information and reports on each bus route, such as schedules, ridership, on-time performance, and service disruptions. This data is normally used in the day-to-day operations of LeeTran in order to monitor various performance metrics. Although hourly data would have been ideal for analyzing OTP, daily aggregates were selected to account for other variables and maintain consistency.

The On-time Performance dataset includes a variety of variables for analysis. Below is a list of the key variables used in this study, with some unrelated variables used for calculations omitted.

- **Day**: Day of the week
- **Route**: Identifier for the transit route
- **On_time**: Percentage of trips from buses that were on time (aggregated at route level)
- **Avg_rain_inch**: Average rainfall in inches on the observation day (aggregated at day level)
- **Accident_count**: Number of accidents recorded on the day (aggregated at day level)
- **Tmax**: Maximum temperature, in degrees Fahrenheit (aggregated at day level)
- **Season**: Binary season of service (e.g., Off-Season, In-Season)
- **PC1**: First principal component from PCA analysis (describes relationship between number of miles round trip for a route and number of stops)

The majority of these variables are self-explanatory with the exception of the principal component created. The purpose of principal components analysis is to reduce the dimensionality of data, in this case, we had two variables with high collinearity, the number of miles round trip for a route, and the number of stops. Since routes with higher mileage are correlated with an increase in stops, the first principal component captures the majority of this relationship and will be used. Additionally, the "season" variable was created based on the month, with November to April classified as the "in-season" period and the remaining months considered "off-season." This distinction was made to capture the potential impact of seasonality on On-Time Performance (OTP).

Variable	Mean	SD	Median	Range	Skew	Kurtosis
day*	4.02	2.07	4.00	6.00	-0.02	-1.36
route*	12.73	7.28	12.00	24.00	0.14	-1.15
on_time	63.32	15.27	64.59	100.00	-0.56	0.37
avg_rain_inch	0.13	0.36	0.00	3.66	5.57	40.71
accident_count	83.72	21.18	84.00	121.00	-0.23	-0.51
tmax	84.56	7.36	85.00	49.00	-0.91	1.16
season	0.61	0.49	1.00	1.00	-0.45	-1.80
PC1	-0.00	1.35	-0.02	5.39	0.17	-0.43

Table 1: Summary Statistics

Based on the Summary Statistics, most of these variables have a relatively normal distribution with some having a slight skew. Variables like avg rain and tmax have extreme values due to extreme rain and weather events that happen often in southwest Florida.

An interesting trend in the On-Time Performance (OTP) data, as shown in Figure 1, is the variation in OTP across different days of the week. Higher OTP percentages are observed on days like Sundays, likely due to reduced traffic, and a reduced number of routes, which leads to smoother operations compared to other days. Future analyses could further break down the data by time of day. This would identify how peak hours impact OTP, as these peak periods likely drive much of the variation in daily averages. Certain routes also exhibit consistent OTP patterns due to factors such as road congestion, traffic conditions, and route-specific characteristics. This suggests that some routes inherently perform better than others, and these differences in OTP may be influenced by the unique conditions of each route.

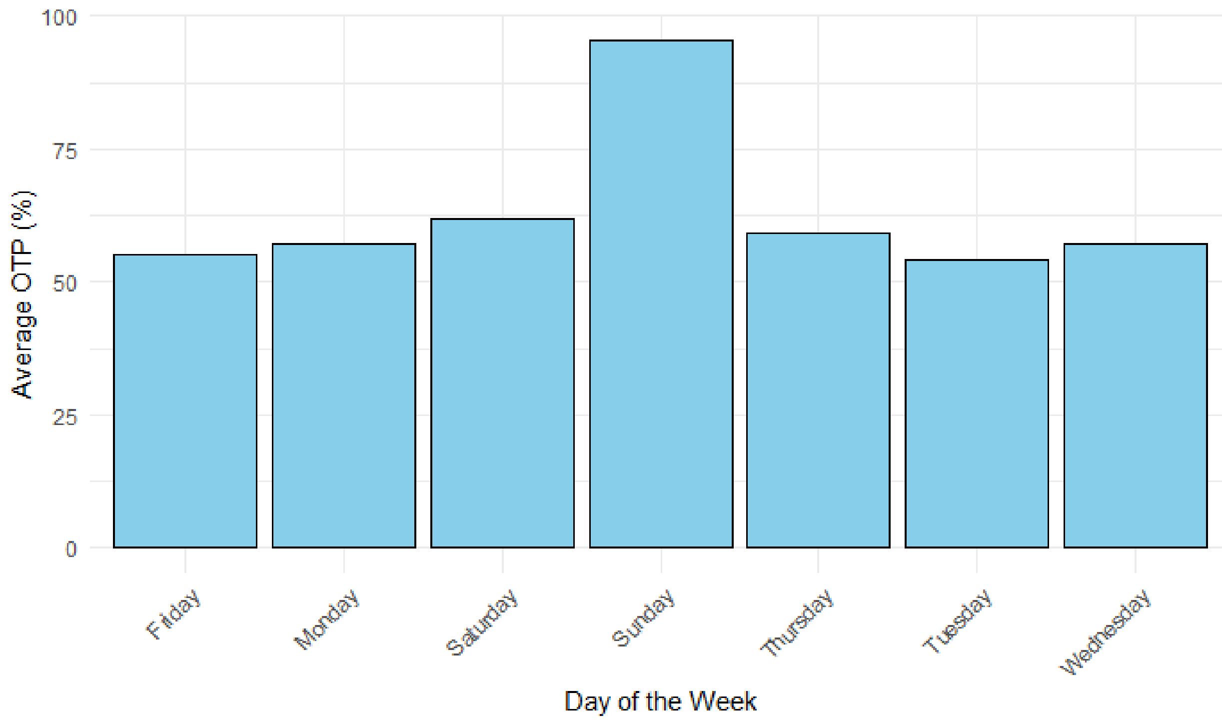

Figure 1: OTP by Day

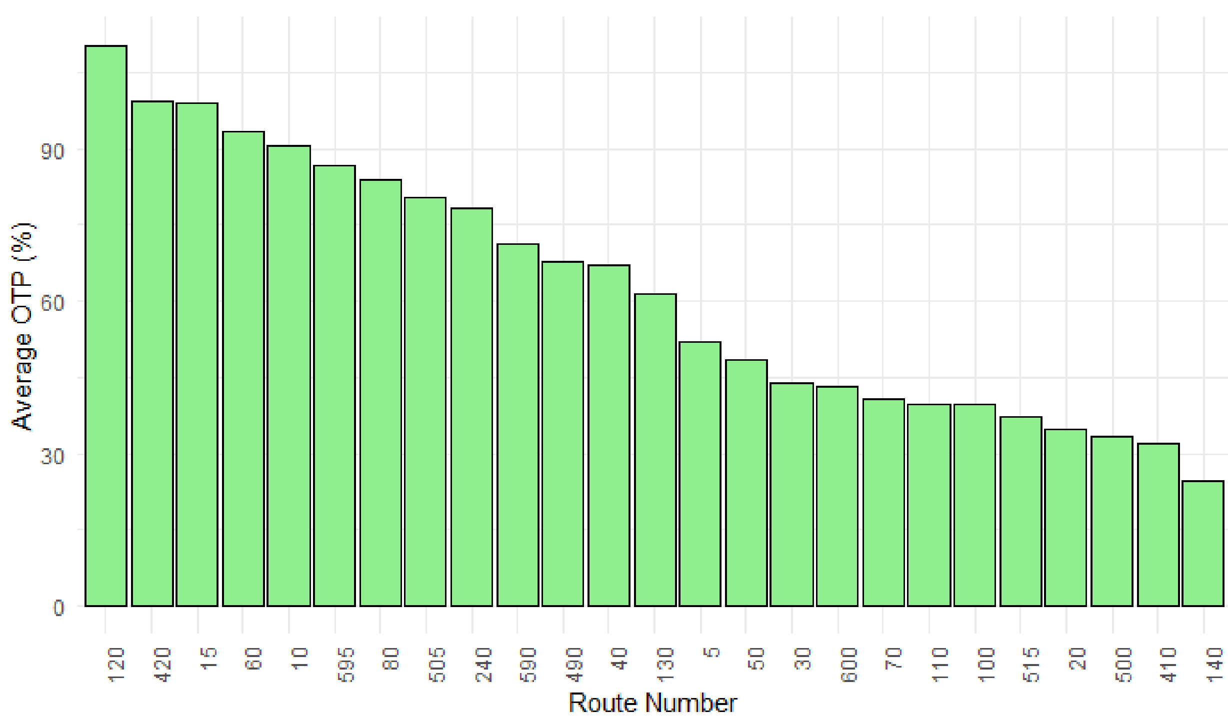

Figure 2: OTP by Route

Looking at the correlation matrix in Figure 3, we see that on-time and accident counts exhibit the strongest negative correlation, suggesting an increase in accidents is associated with a decrease in OTP. All other variables show negligible or weak relationships with both on-time and each other, indicating that they do not have a strong impact on OTP by themselves. Additionally, after viewing the scatter plots, it became evident that they were unnecessary to include, as they did not reveal any clear trends.

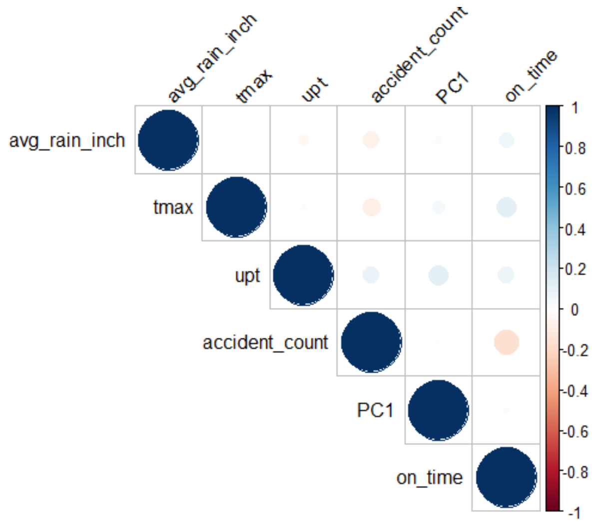

Figure 3: Correlation Matrix

Figure 4 shows a box plot comparing ranges of OTP across the seasonality variable, 0 being off-season and 1 being in-season. Here we can see a minor decrease in on-time performance based on seasonality. This would make sense as there is higher congestion on routes in peak season due to increased snowbirds.

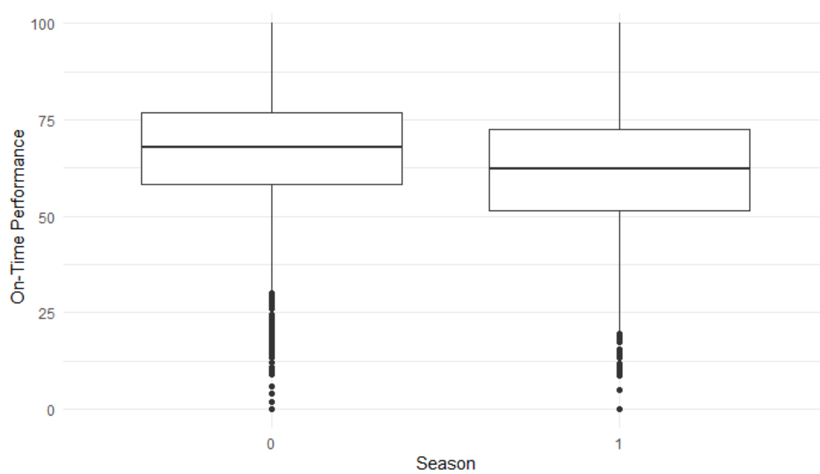

Figure 4: OTP by Season

3.2 Census Tract Data Summary

The Census Tract data was obtained from the 2019 American Community Survey for Lee County. "The US Census Bureau conducts various surveys that study households, businesses, schools, hospitals, etc. These statistics deliver valuable information for local officials and organizations who provide resources and services to your community" (USCB, 2024). Some of this data comes from the American Community Survey (ACS) which is an ongoing survey that occurs every year. Some of the data was broken up by census tract blocks but was then aggregated to the entire census tract before analysis. This will be used to perform a cross-sectional analysis of transit accessibility based on various socio-economic variables. Focusing on census tracts provides a deeper understanding of Lee County's data and how these various socio-economic variables vary from community to community.

The Census Tract dataset includes a variety of variables for analysis. The primary focus during data collection was to gather a set of socio-economic variables and use the distance to census tracts as a measure of accessibility. Below is a list of the key variables used in this study, with some unrelated variables used for calculations omitted. All of the variables used in the dataset are measured at the census tract level.

- **Tract**: Identifier for the census tract
- **Total_pop**: Total population
- **White_prct**: Percentage of the population that is White
- **Black_prct**: Percentage of the population that is Black
- **Asian_prct**: Percentage of the population that is Asian
- **Other_prct**: Percentage of the population that is of other races
- **Med_householdincome**: Median household income
- **Poverty_prct**: Percentage of the population living below the poverty line
- **Avg_workhours**: Average number of work hours per week for individuals
- **College_prct**: Percentage of the population with a college degree or higher
- **Employment_prct**: Percentage of the population that is employed
- **Distance_from_closest_stop_miles**: Average distance (in miles) from the closest transit stop

Most of the variables listed are self-explanatory for this dataset. However, it is worth noting the process in which distance from closest stop was calculated. The distance was measured from the center of the census tract to the nearest bus stop. This distance serves as our accessibility metric, so understanding how it was calculated is crucial.

Variable	Mean	SD	Median	Range	Skew	Kurtosis
tract	279.25	227.88	302.01	897.99	0.41	-0.79
total_pop	4469.50	2757.29	3768.00	12957.00	1.11	0.83
white_prct	85.67	14.97	89.50	93.12	-2.34	7.79
black_prct	8.00	12.52	3.42	85.45	3.34	14.76
asian_prct	1.60	2.24	0.99	18.17	3.74	20.61
other_prct	2.82	4.35	0.94	29.56	2.69	9.74
med_householdincome	64266.14	23202.54	59015.75	131764.75	1.15	1.82
poverty_prct	12.71	8.78	10.60	50.80	1.55	2.73
avg_workhours	38.66	2.07	38.70	16.50	-0.88	3.81
college_prct	30.70	14.87	28.20	69.07	0.67	-0.21
employment_prct	40.06	10.54	41.44	55.51	-0.71	0.19
distance_from_closest_stop_miles	1.56	2.40	0.69	18.01	3.52	15.83

Table 2: Summary Statistics

Table 2 displays the Summary Statistics for the census tract data set. Within this dataset, there is variation across census tracts in terms of demographic and socio-economic characteristics. There is significant variation in terms of total population, income levels, and racial composition, with some tracts showing extreme values for these variables. This makes sense as each census tract has a unique population with different characteristics, representing broad diversity in socio-economic conditions and access to services across tracts.

Comparing Figures 5, 6, and 7, they highlight the total population, distance to the nearest bus stop, and poverty percentage by census tract. The distance map provides valuable insights into areas that Lee-Tran currently services. It appears that many of the areas currently serviced are predominantly lower-income communities. However; it does highlight some neighborhoods that are currently under served. Based on my preliminary data exploration, since many communities and tracts within Lee County exhibit significant differences, this emphasizes the diversity within our dataset. Further investigation is needed to identify where LeeTran could improve its service coverage to better meet the needs of these communities.

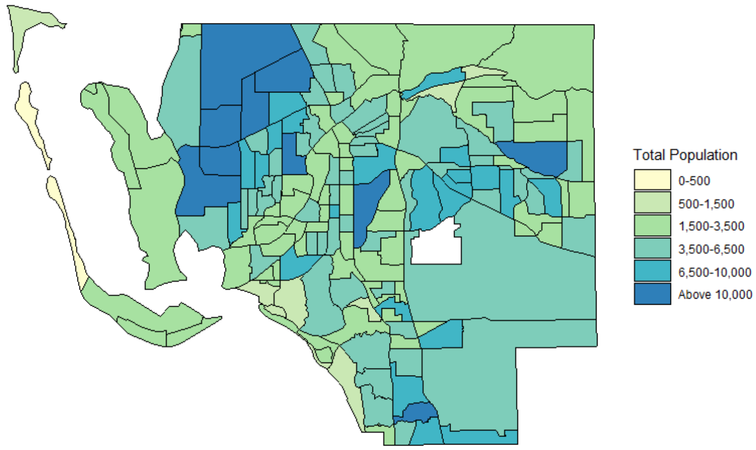

Figure 5: Total Population by Tract

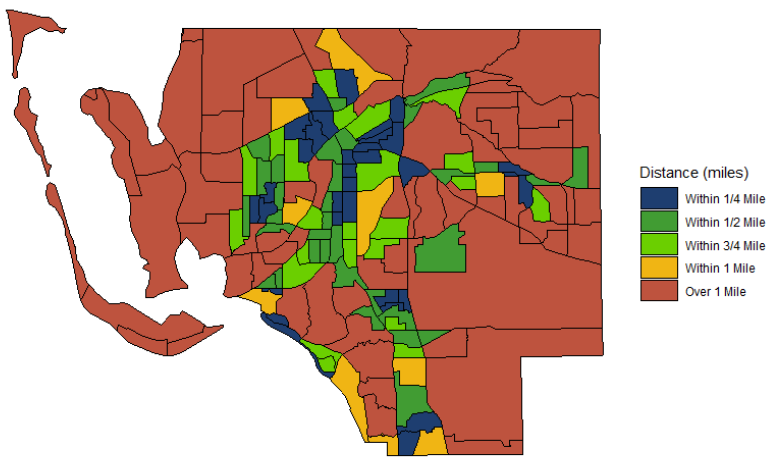

Figure 6: Distance to Nearest Bus Stop by Tract

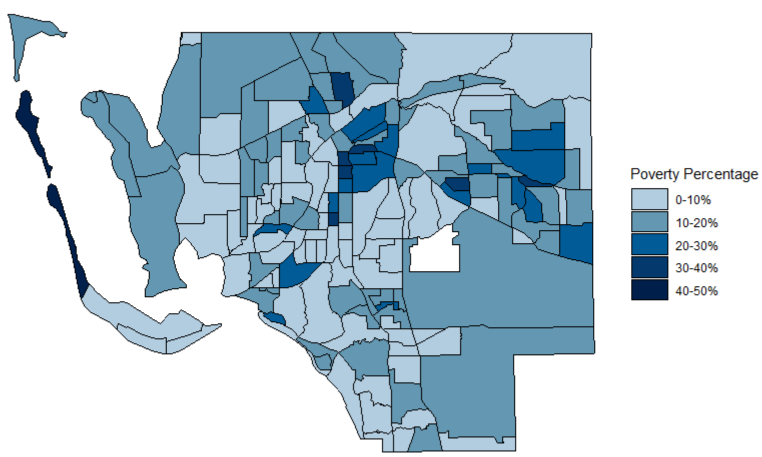

Figure 7: Poverty Percentage by Tract

The following correlation matrix in Figure 8 highlights some key relationships in the preliminary analysis that shows some trends in the data. First off, median household income has a strong positive correlation with college percentage (.70). This suggests areas with a higher percentage of college graduates lead to higher median incomes. Conversely, poverty has the opposite effect and is negatively correlated with income (-.47), which conceptually makes sense. College percentage is also negatively correlated with poverty (-.57) reinforcing these observations. Average work hours show a low correlation across the board. This suggests there is not a strong enough linear relationship to other exploratory variables by itself.

There are also significant inverse relationships in race percentages as we can see as the population of white % increases other minority races decrease and vice versa. Due to the collinearity among the racial variables, we decided to remove White %, to reduce dimensionality and mitigate multicollinearity. Population is highly correlated to poverty and employment, but since they are independent variables in the models it will be okay to include population as a control.

Finally, our main independent variable to be used, distance to stop, shows a positive correlation with median income (0.34) and college percentage (0.17). However; it has generally weak correlations with other variables. This does imply that areas with increased distances to bus stops have slightly higher income and educational attainment levels. This would fall in line with how LeeTran currently operates, by primarily targeting areas with low income.

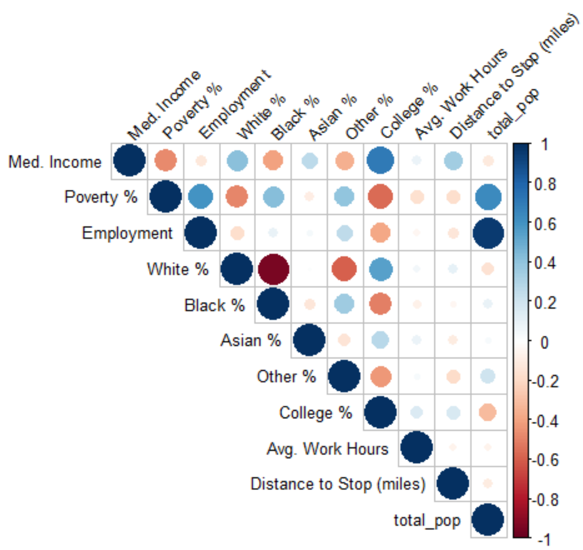

Figure 8: Census Tract Correlation Matrix

The scatter plots in Figure 9 show another necessary exploration step and highlight the effect of our main independent variable, distance to stop, on our various socio-economic variables used for the models. Even though we captured some of these relationships in the correlation matrix it seemed important to take a more granular look at the observations.

In the graph, it is evident there is a positive relationship between median household income and distance to stop. This suggests that, generally, the farther you are from a stop, the higher your income tends to be. However, there appear to be outliers in the data that could skew this relationship. Specifically, some census tracts with significantly higher median incomes are located farther away from stops. It is still important to include this variable for the cross-sectional analysis. Poverty and Employment seem to have a slight negative correlation to distance. Moving forward, it must be investigated further in the model to understand how distance affects these socioeconomic variables when taking into account other controls and factors.

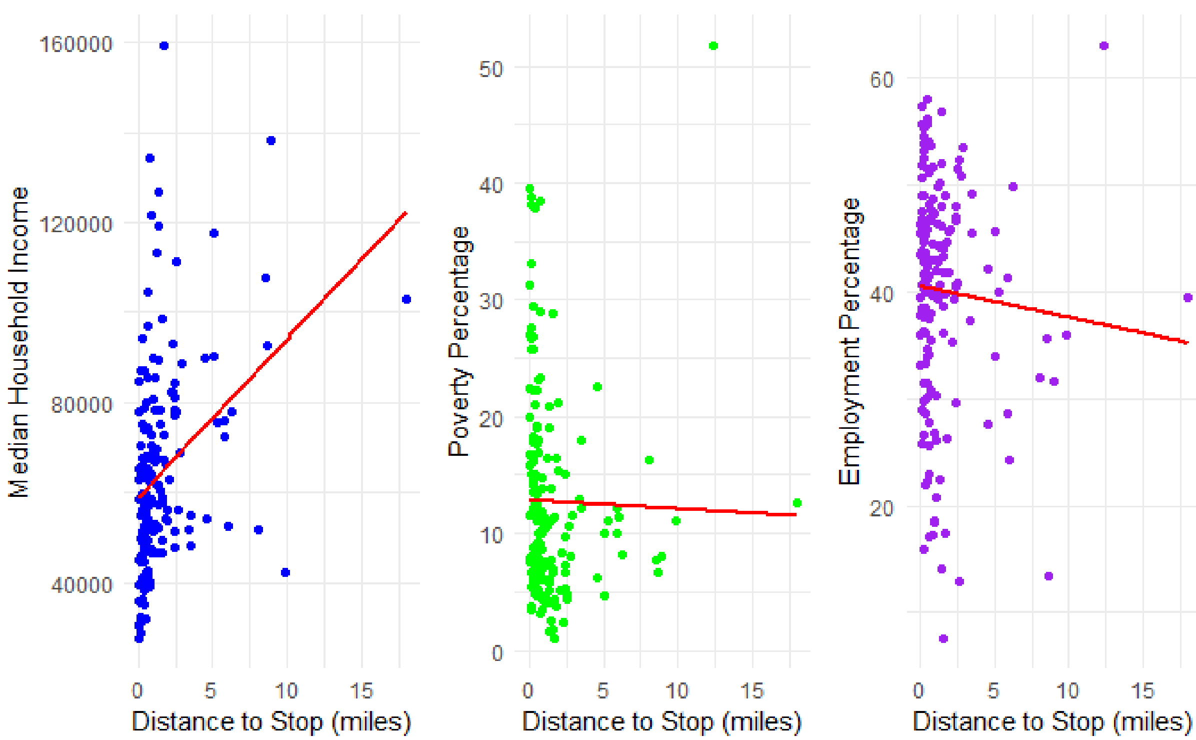

Figure 9: Census Tract Scatter Plots

4 On-Time Performance Analysis

The following is the Empirical Model used for the OLS analysis and displays the contextual understanding of how the model should work. The analysis will follow this empirical model in two different ways. The first model will follow the formula shown, however; the second model will be fixed at the route level accounting for unique route characteristics. In analyzing the models, the primary focus was on understanding the effect of each variable on on-time performance, rather than quantifying the exact numerical impact.

$$OnTime =$$
$$\beta_0 + \beta_1 Rain^2 + \beta_2 MaxTemp + \beta_3 UPT + \beta_4 AccidentCount + \beta_5 Route + \beta_6 Season + \beta_7 PC1 + \epsilon$$

On-time performance is the output variable we are measuring the effects on. Beta 0 is the intercept or the baseline when all values are set to zero. Beta 1 is the coefficient for the average rainfall variable. The rain variable was squared to account for the exponential effect that extreme weather, rainfall, has on the output. Beta 2 is the coefficient for maximum temperature; Beta 3 is the coefficient for unlinked passenger trips; Beta 4 represents the coefficient for accident counts; Beta 5 is the coefficient for the categorical route variable; Beta 6 is the coefficient for the binary seasonality variable; and Beta 7 corresponds to the coefficient for the first principal component (PC1) from the principal component analysis output. Epsilon represents the error term.

4.1 Results

Table 3 shows the output of the two regression models we created. The Regular OLS model, and the OLS model with fixed effects at the route level. Based on this table we can see the impacts of various predictors on the output variable.

In the OLS model without fixed effects, all variables were significant at the .1% level, indicating that each one has a meaningful contribution to explaining on-time performance. The exploratory variables had different effects on OTP as well, rainfall, temperature, and unlinked passenger trips surprisingly had a positive impact on OTP. While accident count, route, season, and PC1 had a negative coefficient. This could be due to some variables having a non-linear relationship with OTP that are not captured by the linear OLS model. Further analysis, including potential model specification checks or exploring interaction terms, is needed to better understand these effects.

The fixed effects model had different results. Here, only accident count remained significant at the .1% level, while rainfall and seasonality were significant at the 5% level. This suggests that, after accounting for route-specific effects, some of the originally significant predictors in the OLS model became less influential, with only a few variables maintaining statistical significance. Accident count and PC1 were the only variables hurting OTP in this model with the rest of the coefficients being positive.

In both models, variables like season and accident count show consistent and significant negative impacts on on-time performance. These results make sense conceptually as accidents can affect traffic and bus speeds leading to lower on-time performance. Accident count could also explain some of the variance that might be attributed to a traffic-related variable, which is not included in the model. Since accidents lead to increased congestion, they may capture some of the variability typically associated with traffic conditions. Similarly, the seasonality variable represents how there is increased traffic and congestion in these peak seasonal months affecting performance.

These models capture a low percentage of the variation in the data. The OLS Adjusted R squared is around 11%, and the OLS with fixed effects is around 31%. This means there are other predictors of on-time performance that are not included within this model. So these models are not a good measure of all the variables that affect on-time performance, but rather a glimpse into a few that do.

To improve this model, other variables should be explored and added to this model. Capturing traffic data, which was challenging to acquire and thus omitted, would be crucial for measuring On-Time Performance (OTP), especially on a route-by-route basis. Additionally, analyzing data at a more granular, time-specific level could offer a more quantifiable understanding of individual route performance. While this analysis does not provide a meaningful gauge of quantifiable impact, it highlights the key areas that influence on-time performance (OTP). This can serve as a step in the right direction for LeeTran as they can explore these variables further and pinpoint the most influential factors of OTP.

Dependent Variable:	On Time	
Model:	(1) OLS No FE	(2) OLS with Route Fixed Effects
Variables		
I(avg_rain_inch^2)	1.5970***	0.3084*
	(0.4181)	(0.1189)
tmax	0.0903***	0.0808
	(0.0259)	(0.0538)
upt	0.0028***	0.0009
	(0.0004)	(0.0019)
accident_count	-0.1114***	-0.1121***
	(0.0071)	(0.0220)
route	-0.0194***	
	(0.0008)	
season	-2.6394***	-2.1417*
	(0.4067)	(1.1975)
PC1	-1.8227***	
	(0.1232)	
Fit statistics		
Observations	9,841	9,848
R^2	0.1129	N/A
Adjusted R^2	0.1123	0.3132
Within R^2	N/A	0.0542
RMSE	14.39	12.6

Clustered (route) standard-errors in parentheses
*Signif. codes: ***: 0.001, **: 0.01, *: 0.05*

Table 3: Comparison of OLS Model Results

5 Transit Accessibility Analysis

The census tract analysis will include three separate models, each using a different socio-economic variable as the dependent variable: Median Household Income, Employment Percentage, and Poverty Percentage. The goal of this analysis is to examine how transit accessibility, measured by the distance to the nearest bus stop, relates to socio-economic outcomes in Lee County, and to identify the main determinant for prioritizing service in low-income areas through a cross-sectional analysis of census tracts. The following is the empirical model that will be used for the models, this provides context for how the OLS Regression will be applied.

$$SocioeconomicVariable = \beta_0 + \beta_1 DistancefromStop + \beta_2 TotalPop + \beta_3 Black\% + \beta_4 Asian\% + \beta_5 Other\% + \beta_6 College\% + \beta_7 AvgWorkHours + \epsilon$$

The output variable will be determined by three different socioeconomic variables used in each model: Median Household Income, Employment Percentage, and Poverty Percentage. Beta 0 is the intercept or the baseline when all values are set to zero. Beta 1 corresponds to the coefficient for the distance to the bus stop variable, Beta 2 represents the coefficient for the total population, and Beta 3, 4, and 5 represent the coefficients for the race control groups. Beta 6 refers to the coefficient for the percentage of college graduates, and Beta 7 represents the coefficient for the change in average work hours. Epsilon represents the error term. No interaction effects were modeled beyond the standard equation.

5.1 Results

Table 4 displays the output of the 3 comparison models across socio-economic outcomes. Based on this, the relationship of transit accessibility by census tract can be assessed.

In the first model, median household income was used as the dependent variable. Both distances from stop and college % had a positive relationship with income and were significant at the .1% level. The distance from stop coefficient suggests for every 1-mile increase in distance the median income goes up by approximately $2603. Similarly, for college % for every percentage increase in college attainment income increased by $964. The total population also had a positive relationship, but only significant at the 5%

level. overall, this model had the highest performance with an Adjusted R-squared of 55%, suggesting over half of the variability in the output is explained by the exploratory variables.

In the second model, Poverty % was used as the dependent variable. Our main independent variable distance, did not have a significant relationship to poverty. Other control variables like black % were significant at the .1% level and had a positive relationship, an increase in the Black population within census tracts is associated with a higher poverty percentage, indicating that Black communities are more affected by poverty compared to other racial groups. College % had a negative relationship, indicating an increase in college attainment leads to a decrease in poverty, this was also significant at the .1% level. Average work hours had the same negative relationship, but this was significant at the 1% level. Additionally, the intercept was significant at the .1% level suggesting that there is a base level of poverty present across census tracts when all variables are equal to zero. This model had the second highest performance with the model accounting for 47% of the variability in the outcome variable.

The third model, Employment % was used as the dependent variable. Our main independent variable distance, was not significant in impacting poverty in this model. College % had a positive relationship to employment and was significant at the .01%, this highlights how increased educational attainment increases employment. Other control variables like Asian % were significant at the .1% level and had a positive relationship, suggesting an increase in the Asian population contributed to an increase in employment. Additionally, Black % had a significant negative relationship at the 5% level, suggesting that black communities in Lee County are affected by a decreased level of employment. The intercept was significant at the 1% level suggesting that there is a base level of employment present across census tracts when all variables are equal to zero. This model had the poorest performance with the model accounting for 34% of the variability in the outcome variable.

Overall, median household income is the main determinant in transit accessibility for LeetTran. This makes sense since LeeTran has a primary focus of serving low-income communities. However, they do not focus on communities based on employment or impoverished. Taking into account the preliminary data exploration, population isn't directly a factor either.

Dependent Variable:	Socio-Economic Outcomes		
Model:	(1) Med. Household Income	(2) Poverty %	(3) Employment %
Variables			
Intercept	25039.95	43.78^{***}	39.91^{**}
	(23630.78)	(9.77)	(12.89)
Distance from Closest Stop (miles)	2603.41^{***}	0.09	0.31
	(560.81)	(0.22)	(0.29)
Total Population	0.97^{*}	-0.00031	0.00036
	(0.48)	(0.0002)	(0.0003)
Black %	-126.82	0.23^{***}	-0.16^{*}
	(115.99)	(0.05)	(0.06)
Asian %	1079.90	0.21	1.33^{***}
	(582.68)	(0.24)	(0.31)
Other %	-124.13	0.19	0.30
	(322.41)	(0.13)	(0.18)
College %	964.54^{***}	-0.24^{***}	-0.43^{***}
	(109.47)	(0.045)	(0.06)
Average Work Hours	25.68	-0.66^{**}	0.25
	(608.27)	(0.25)	(0.33)
Fit Statistics			
Observations	158	158	158
Multiple R-squared	0.5714	0.4798	0.3711
Adjusted R-squared	0.5522	0.4566	0.343
Residual Std. Error	15530	6.47	8.54
F-statistic	29.71	20.69	13.23
p-value	< 2.2e-16	< 2.2e-16	2.38e-13

*Signif. codes: ***: 0.001, **: 0.01, *: 0.05*

Table 4: Comparison of Regression Model Results

6 Conclusion

This analysis highlighted the various aspects that can affect on-time performance, as well as the main determinants that guide LeeTrans Transit Accessibility. This is a good starting point on what areas LeeTran can improve on. I have a few recommendations including ensuring a more reliable network, expanding the network to target a wider community, using other agencies as a benchmark, and advocating for pro-transit infrastructure. These areas will help LeeTran grow as an agency and improve service quality. Additionally, improving data management, operational involvement, and tracking rider feedback would provide a clearer understanding of service impact and who is being served.

6.1 Maintaining Reliable Transit

The biggest issue LeeTran currently faces is ensuring service is reliable. The analysis revealed several factors that can influence OTP, as well as numerous variables that should be included in the model to account for the remaining variance. Any other potential factors should be further examined to determine if they can be improved through operational or planning adjustments.

One potential factor to be analyzed in the future is driver behavior. Since drivers behave differently, having better-enforced standards and schedule adherence can assist with OTP. Some drivers may start their routes late or may not adhere to the scheduled route pattern. A good start for this would be to keep track of schedule adherence on the operational side and enforce standards for all drivers. Driver performance could be displayed in break areas to highlight employees who closely adhere to starting on time, and incentive programs could be developed around this. Ultimately, this would need to be led by operational management to drive change in driver behavior.

Another recommendation would be to add more buses on key routes to reduce headway. This potential solution is difficult due to funding constraints. Based on the Transit Development Plan, which states that "routes should never be removed," when a route is underperforming, adding additional service should be considered as a necessary step to improve performance before any other action is taken LeeTran (2020). Any removal would also lead to the disenfranchisement of regular riders. Ideally, by focusing on expanding the network, rather than limiting it, LeeTran can continue to grow through this avenue.

Although beneficial to save money, splitting interlines to create separate bus routes can also enhance performance. This should be considered as a trade-off between budget constraints and performance. LeeTran should align its approach to balancing budget constraints and service changes with its long-term goals and mission. By doing so, it can ensure that decisions are made in a way that supports both financial sustainability and the quality of service provided to the community. Separating blocks helps alleviate congestion and ultimately creates smoother operations between interconnected routes.

Finally, continuous monitoring of performance is essential to maintain strong performance amidst changing conditions. This involves tracking OTP at a granular, hourly level, and understanding all factors that contribute to a bus being delayed. These kinds of reports can be viewed in the various software LeeTran has licenses to, including Clever Reports and Clever CAD. Utilization across departments is important and can help maintain monitoring. Since many factors are involved, LeeTran should focus on what is attainable on the operational side and plan for future improvements on the planning side.

6.2 Operational and Policy Implications

Coordination between county departments would be necessary to have the best service possible. As reviewed in the literature, having a supportive infrastructure surrounding transit enhances its effectiveness and serviceability. This includes dedicated bus lanes, efforts to attract new riders beyond the current target demographic, emphasis on the importance of infrastructure development at the county level, and exploring other mixed modes of transport, such as rail, to complement the system. However; as it currently stands this is a tall task, so a broad change in county priorities would be required to secure additional budget allocation for these ideas. However, it is crucial for LeeTran to continuously advocate for these ideas to garner support. Focusing on developing a community-driven, sustainable, and efficient network should be the goal, as it moves us toward the future of transit.

The analysis shows that low-income areas are the main determinant for the current bus network. Expanding service to high-population areas, in addition to low-income communities, would be beneficial for LeeTran. Based on the census tract maps in the Data Summary section, many high-population, higher-income communities present opportunities where service could be expanded. One potential solution for these communities is a neighborhood circulator, as residents often remain within the area and visit similar points of interest. More detailed analysis of these communities is needed, but this approach could help increase LeeTran's ridership while expanding service beyond the current network.

During my research, I found a lack of data on riders themselves, as well as insufficient data management practices to ensure data quality. Currently, LeeTran relies on a third-party vendor to survey riders for feedback for the Transit Development Plan. However, I believe that conducting these surveys internally more frequently and implementing changes based on rider feedback would greatly benefit the service. This approach would better highlight the needs of riders and should always be a key consideration in decision-making. Additionally, LeeTran's database is missing some important metrics and contains inaccuracies in others. A comprehensive data management overhaul to ensure data quality would significantly improve data analysis and contribute to overall service improvements. Certifying the APCS at LeeTran and appointing dedicated personnel to oversee data management would be crucial steps in this direction.

Finally, comparing LeeTran to other transit agencies of similar size would be highly beneficial. Many of these agencies have implemented strategies and solutions that LeeTran can adopt to improve their service. Ultimately, this effort requires strong county support and fresh, innovative ideas that can drive change and uncover new ways to transport people. Remaining stagnant with changes compared to other agencies and external transit services will hinder growth. Thus, LeeTran must embrace new ideas to remain competitive and provide a high-quality service that unites the community.

These recommendations can be made to improve service quality and the transit network itself, positively impacting the community, environment, and economy. Ideally, these improvements would lead to the development of a transit network that serves the entire community, enabling seamless connections for all. With an improved infrastructure to support its growth, the system would offer more reliable service and eventually become one of the main modes of transportation for residents. By taking these actions, LeeTran has the opportunity to become a model for transit systems in the region and redefine transit in Southwest Florida.

References

APTA (2017). The economic impact of public transportation investment.

Baker, R. A. et al. (2019). Access to public transportation and its impact on poverty and unemployment in atlanta, georgia. *Urban Studies Institute*.

Bartee, T. (2007). Transit reliability and its impact on transit usage. *National Center for Transit Research Center for Urban Transportation Research*.

Esri (2021). Driving toward equity in addressing community mobility.

Forum, I. T. (2021). Future public transport funding: New approaches and strategies.

FTA (2010). Transit and the economy: The economic impact of transit investment.

Gonzalez, J. et al. (1993). The role of on-time performance in public transport: A literature review. *Transportation Research Part A: Policy and Practice*, 27(6):525–533.

LeeTran (2020). Transit development plan. Prepared for Lee County Transit.

Litman, T. (2021). Evaluating public transportation health benefits. Technical report, Victoria Transport Policy Institute.

Lucas, K., Verlinghieri, E., and Mattioli, G. (2016). Transport and social exclusion: Where are we now? *Transport Policy*, 49:1–10.

Mishra, R., Pulugurtha, S. S., and Mathew, S. (2023). Examining associations with on-time performance... *Multimodal Transportation*, 2(4):100094.

National Research Council (2002). Chapter: Transit use, automobility, and urban form. In Council, N. R., editor, *Making Transit Work: Insight from Western Europe, Canada, and the United States*. National Academies Press.

USCB (2024). United states census data.