# 

# 

# 

BU Spark!

Bus Equity

DS 701 x Livable Streets Alliance

2023 Fall: Tools for Data Science

Team member: Xiang Li, Xinyu Zhang, Bohan Wang, Yu Liang

PM: Laurel Murphy TPM: Smriti Suresh



# Introduction

Bus equity is a significant issue affecting different areas in Boston. There are some gaps between different racial groups, especially for Black and LatinX Riders. This is a complicated problem that requires multiple approaches to address it.

In this report, we are going to focus on establishing safe and equitable public spaces while proposing transportation solutions in Massachusetts. It focuses on current plans and future plans to address bus commute time disparities in Massachusetts, particularly examining the impact of these inequalities on Black riders and LatinX communities. A significant difference in bus commute times between Black and White riders is highlighted, with Black riders spending an additional 64 hours per year on buses compared with White riders. Similarly, there is a time disparity of 10 hours in the LatinX community. The MBTA's services are not equitably distributed, with Black and Latino communities particularly facing longer commute times. This caused regional and racial disparities within the MBTA's bus service.

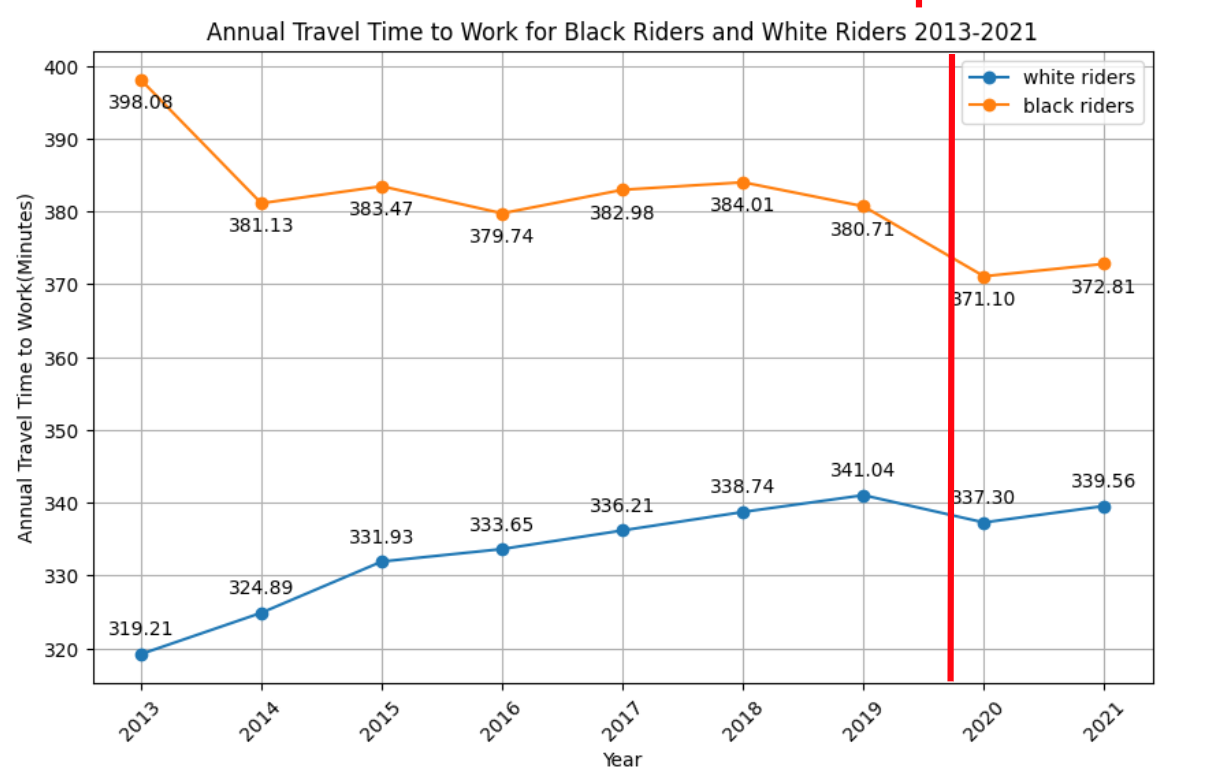
Our plan for this project is to do data analysis according to the client’s requirements. This would include hypothesis tests, geographic analysis, and time series analysis. We use the data from the American Community Survey, from the U.S. Census Bureau. For questions 1 and 4, we used the data from 2013 to 2021. For question 2m we used the data from 2016 to 2021. For question 3, we used the data from 2011 to 2021. We decided to use the 5-year estimate data as one-year data for each analysis question instead of directly using one-year data to ensure that the data is more accurate. This project mainly uses Python to analyze and visualize the data. Based on the visualization, we answer our client’s questions and give suggestions for further decision-making. We also utilize a linear regression model to forecast future trends in disparities, considering the potential ongoing effects of COVID-19. This can help in understanding the possible future scenarios. Our final goal is to conduct an analysis based on annual commute census data and aims to provide a concise overview of the progress or lack thereof in addressing this issue. We want to provide a foundation to support decision-making and policy development.

We also discuss the further analysis both on variables selection analysis and factors analysis.

# Visualizations and Analysis for Key Questions

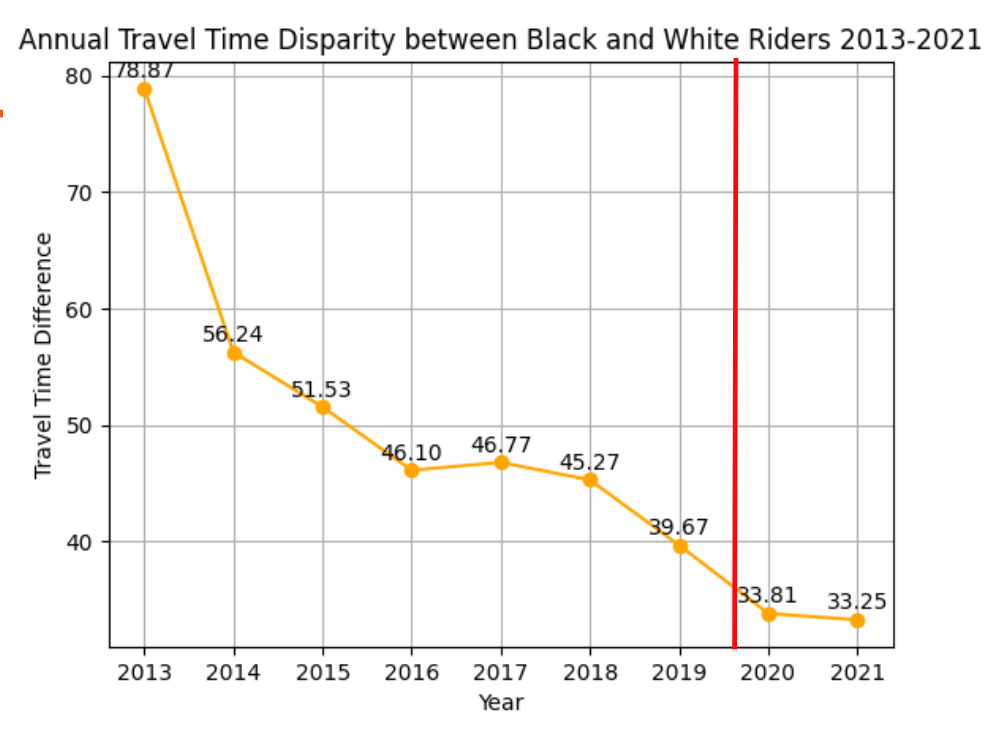
## Analysis for Question 1

How has the 64-hour bus commute disparity between Black and white riders evolved over the past five years, considering both pre-Covid and Covid-affected periods, and are there discernible patterns or changes in the trend?



*Figure 1: Average Travel Time to Work for Black Riders and White Riders 2013-2021*

From Figure 1, the annual commute time was continually increasing for white riders before COVID-19. For black riders, there was a sharp drop in annual commute time from 2013 to 2014, followed by small annual fluctuations from 2015 to 2019. But overall, annual commuting times for black riders are decreasing. Commute times for both black and white riders decreased significantly in the Covid-affected period, attributable to the impact of the Covid. The annual commute time for both black and white riders declined sharply in 2020 compared to the previous year. In 2021, as the world began to adapt to COVID-19, commute times for both groups increased slightly compared to 2020 but did not return to pre-Covid levels.



*Figure 2: Annual commute Time Disparity between Black and White Riders 2013-2021*

Figure 2 shows that the disparity in commute times has been gradually narrowing in the years before and during COVID-19. The most significant change occurred in 2020 when the disparity dropped significantly to 33.81 hours. This may be caused by the fact that riders' normal commuting patterns were disrupted due to lockdowns, increased remote working, and other factors associated with COVID-19. However, the data suggest that it may have a relatively greater impact on reducing commute times for black riders, resulting in a significant decline in 2020. In 2021, this disparity narrows to 33.25 hours, which means the disparity in commuting times between blacks and whites tends to balance in the Covid-affected period. However, without additional context or data for the years after 2021, it is unclear whether this equality trend will continue or return to pre-Covid patterns. Therefore, we conducted linear regression analysis in question 4 to predict the future pattern after COVID-19. Figure 2 clearly shows that the 64-hour commute time disparity has continued to narrow over the past five years. We updated the disparity in commute times between black and white riders in the past five years(2017-2021) from 64 hours to 40 hours by using the formula:

******

**Statistical analysis (T-Test):**

*Null hypothesis(H0):* Covid does not affect the difference in commute times between black and white riders. This means that the mean of the difference in commuting times before and after the pandemic is the same.

*Alternative hypothesis(H1):* Covid has an effect on the difference in commute times between black and white riders. This means that the mean of the difference in commuting time before and after the pandemic is different.

*The Result for T-tests:*

T-statistic: -4.33

P-value: 0.0047

*Interpretation:*

This P-value is less than 0.05. The t-test results allowed us to reject the null hypothesis at a 95% confidence interval, indicating that there is a statistically significant difference in the commute time disparities between Black and White riders before and during COVID-19. This result suggests that Covid did have an effect on the difference in commute times between black and white riders.

## 

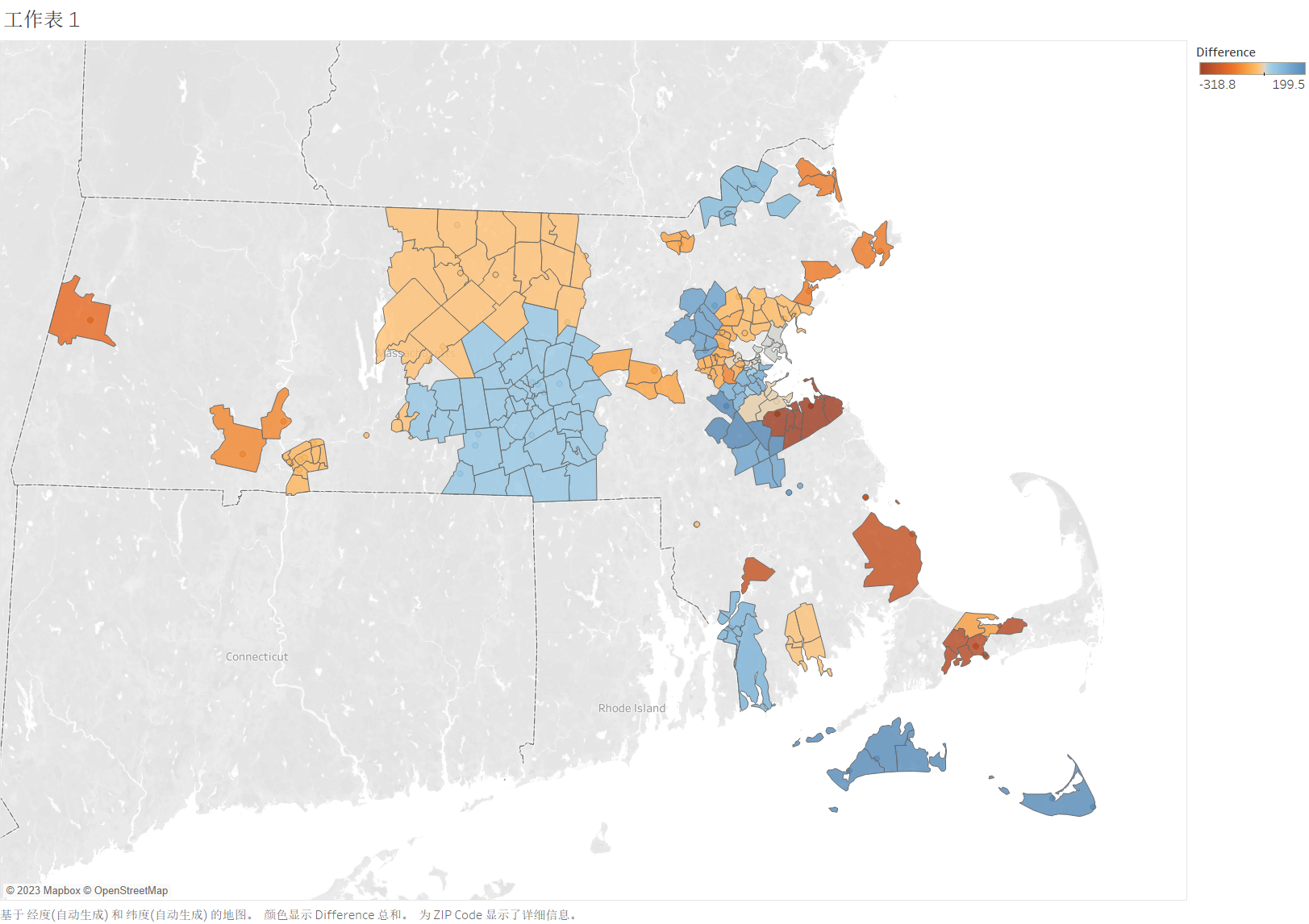
## 

## 

## 

## AnalysisforQuestion2

Does the 64-hour gap in bus commute times hold consistent across different geographic areas or does it vary significantly, and are there specific regions where the gap is more pronounced?

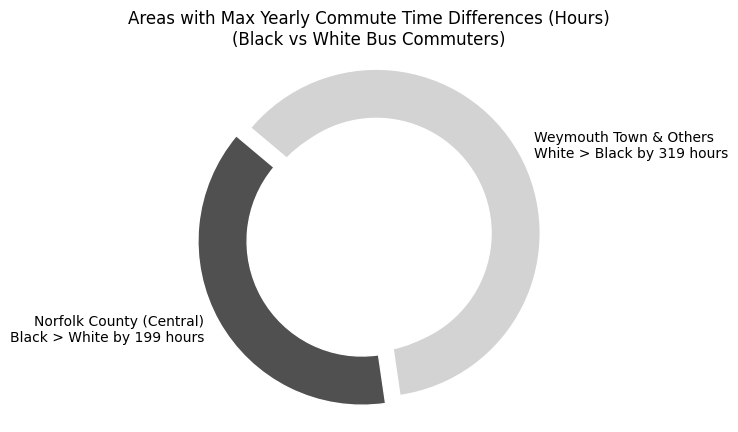


***Source: United States Census, 2016-2021***

*Figure 3: Annual commute Time (hours) Disparity Between Black And White Bus Riders 2016-2021*

Figure 3 is based on data from a total of PUMA areas from 2016 to 2021 in MA. The difference represents the hours gap that blacks spend commuting to work more than whites. The darker the red in the graph, the more whites in the area use buses cost time to commute than blacks, while the darker the blue, the more blacks in the area use buses cost time to commute than whites.

As we can see in Figure 3, the difference in commute times to work by bus for black whites is still relatively large in most areas. In central MA, the differences are smaller, but in the margins of MA, the disparities are larger. The results indicate that there are regional inequities that need to be addressed to ensure commute time equity in all regions.



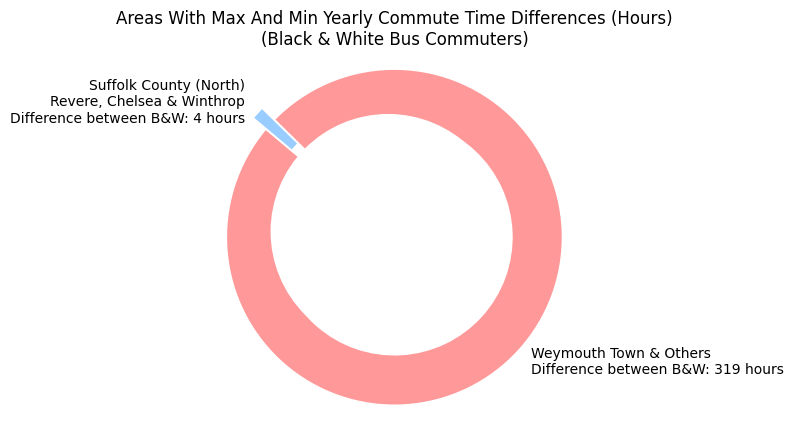
**Black Riders +199** hours in Norfolk County

**White Riders +319** hours in Weymouth Town

***Source: United States Census, 2016-2021***

*Figure 4: Areas with Max Yearly Commute Time Differences (Hours) (Black vs White Bus Commuters) 2016-2021*

In Figure 4 we find the largest black-white disparity in commute time to work by bus. Blacks have up to 199 more hours of commuting per year than whites in Norfolk County, while whites have up to 319 more hours of commuting per year than blacks in Weymouth Town. It can be concluded that for riding public transit, the more unfair to blacks in Massachusetts is Norfolk County and the more unfair to whites is Weymouth Town.



**The maximum difference between B&W 319** hours in Weymouth Town

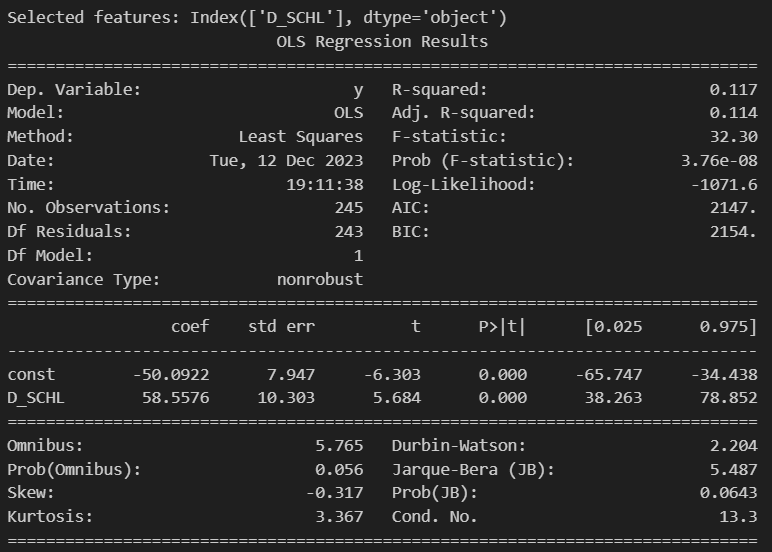
**The minimum difference between B&W 4** hours in Suffolk County

***Source: United States Census, 2016-2021***

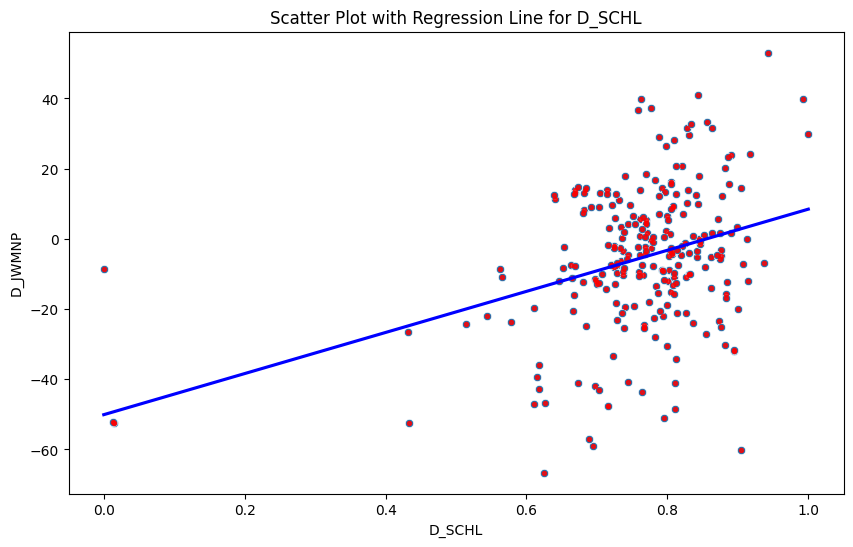
*Figure 5: Areas With Max And Min Yearly Commute Time Differences (Hours) (Black & White Bus Commuters) 2016-2021*

In Figure 5, we find the maximum and minimum time gaps for black whites commuting to work by bus. Weymouth Town has the largest time gap at 319 hours per year. Suffolk County has the smallest time gap at 4 hours per year. We find the maximum and minimum time gaps for black whites commuting to work by bus. Weymouth Town has the largest time gap at 319 hours per year. Suffolk County has the smallest time gap at 4 hours per year. It can be concluded that the most equitable area in Massachusetts for transit is Suffolk County and the least equitable is Weymouth Town. The further analysis will be finding out what are the main reasons to affect commuting time.

We try to find relationships between different *travel times* to work between black and white and [*Age, Education Attainment, Weeks worked during past 12 months, Salary income past 12 months*]. So, we did a regression analysis and found that the differences in *Education Attainment* between black riders and white riders are related to differences in *travel time to work.*



*Figure 6: Linear regression model with time gaps for black whites commuting to work by bus and schooling in yearly data*



*Figure 7: Linear regression with time gaps for black whites commuting to work by bus and schooling in yearly data*

* For D\_SCHL, the coefficient is 58.5576, suggesting a positive relationship with the dependent variable.
* The F-statistic tests the overall significance of the model. A very low Prob (F-statistic) value, as seen here, suggests that the model is statistically significant.
* In summary, the model has one predictor (D\_SCHL) which seems to have a significant positive relationship with the dependent variable. However, the R-squared value is relatively low, indicating that the model explains only a small portion of the variability in the dependent variable.

So, for the Suffolk County, Weymouth Town:

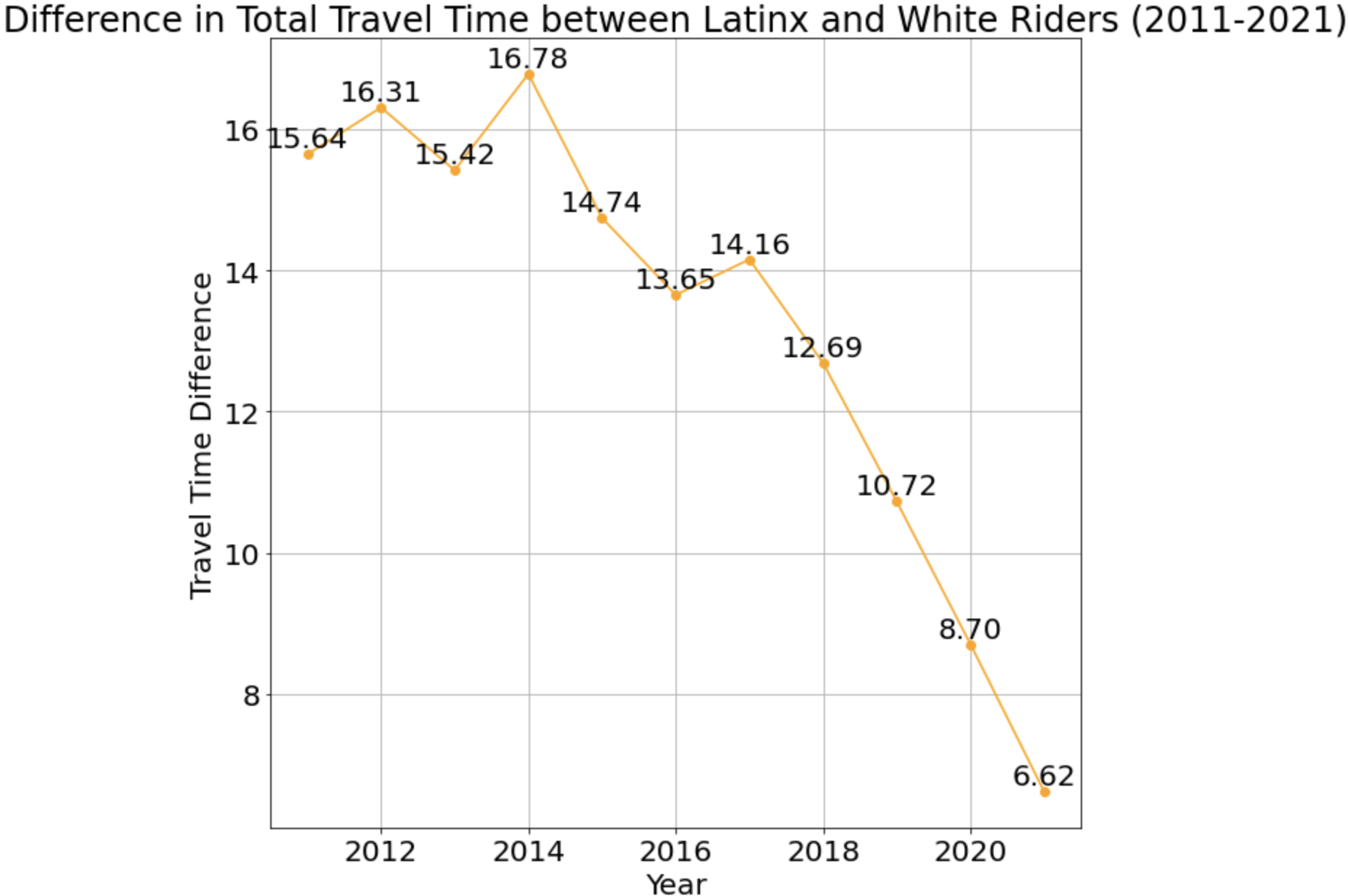
*Table 8: The time gaps for black whites commuting to work by bus and schooling in yearly data in the significant areas*

| **Location** | **Average D\_SCHL** | **Time Gap (hours/year)** | **Time Gap Description** |
| --- | --- | --- | --- |
| Suffolk County (North)--Revere, Chelsea & Winthrop Town Cities | -0.242531 | 4 | Smallest time gap at 4 hours per year |
| Weymouth Town, Braintree Town Cities, Hingham, Hull & Cohasset Towns | -2.832550 | 319 | Largest time gap at 319 hours per year |

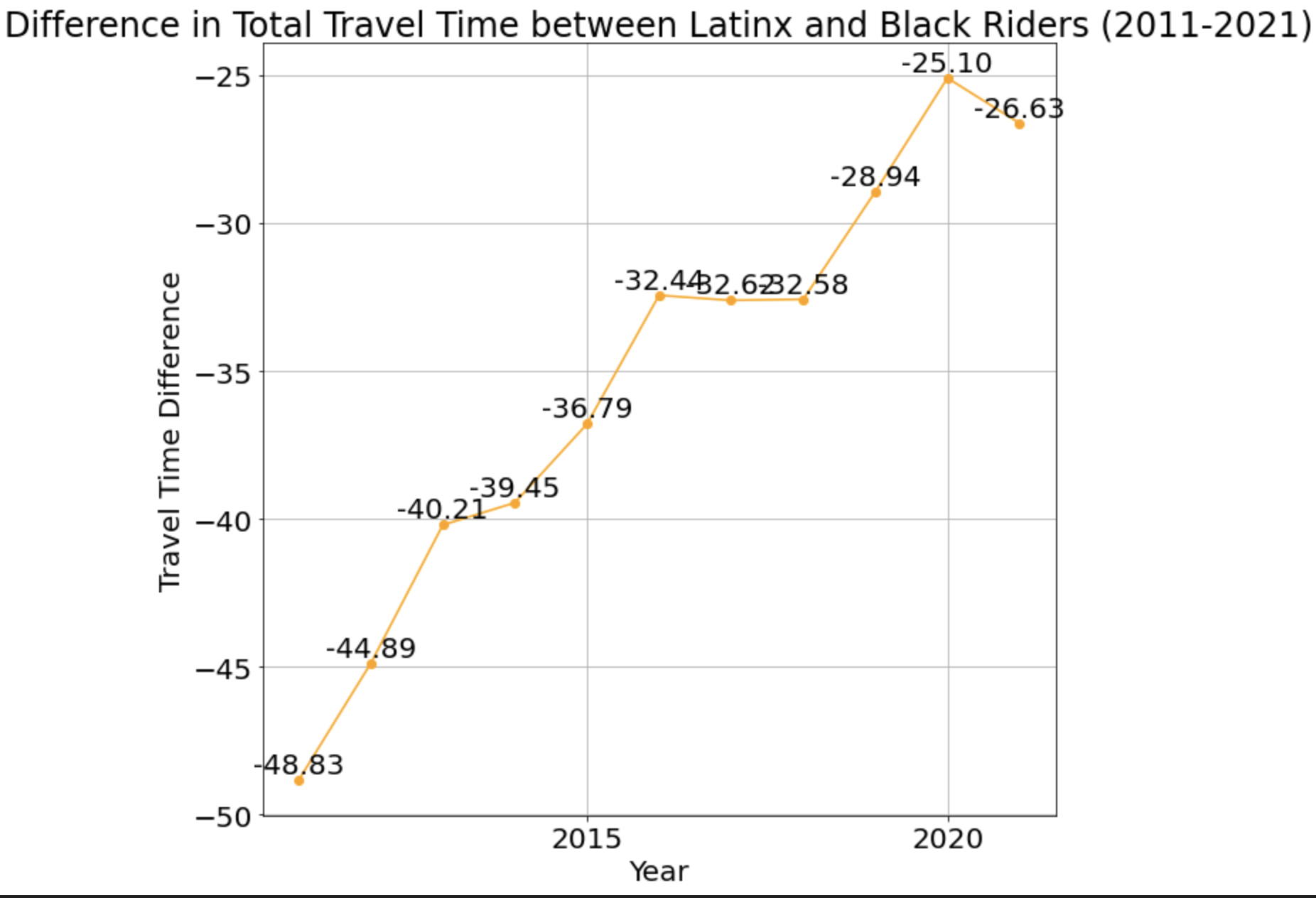
In Table 8, we observed not only stark regional disparities in commute times but also a notable correlation with educational differences. Suffolk County, with a minimal annual commute time gap of 4 hours and a low average difference in education levels (Average D\_SCHL), represents a model of relative equity. Conversely, in Weymouth Town, White commuters endure up to 319 more hours annually, coinciding with a larger average difference in education levels. These findings underscore the need to address not only the racial and regional inequities in public transit but also the intertwined impact of educational disparities in Massachusetts.

## Analysis for Question 3

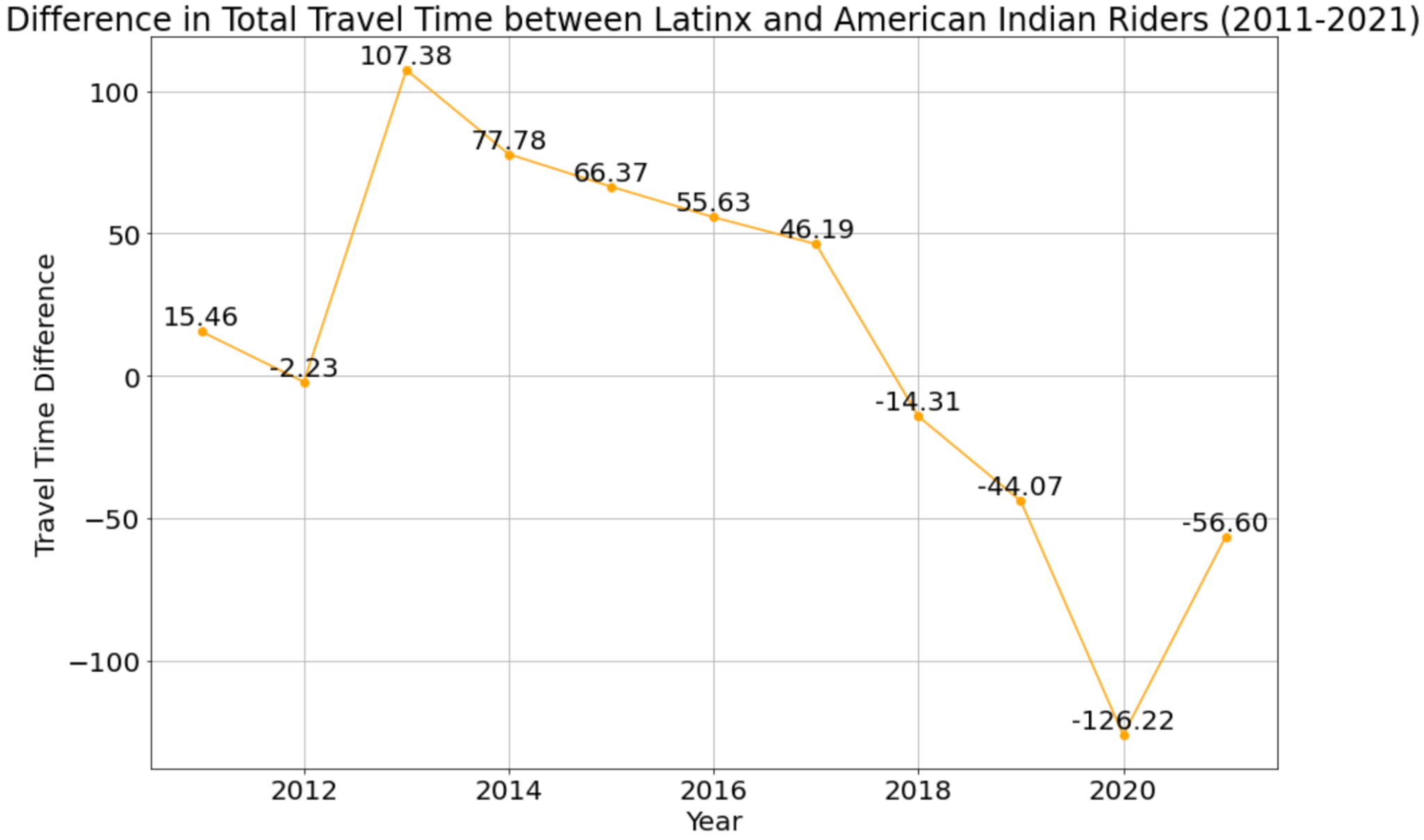
How does the 10-hour bus commute disparity observed in LatinX communities compare with other racial groups, and are there common factors contributing to these discrepancies across different demographics?

****

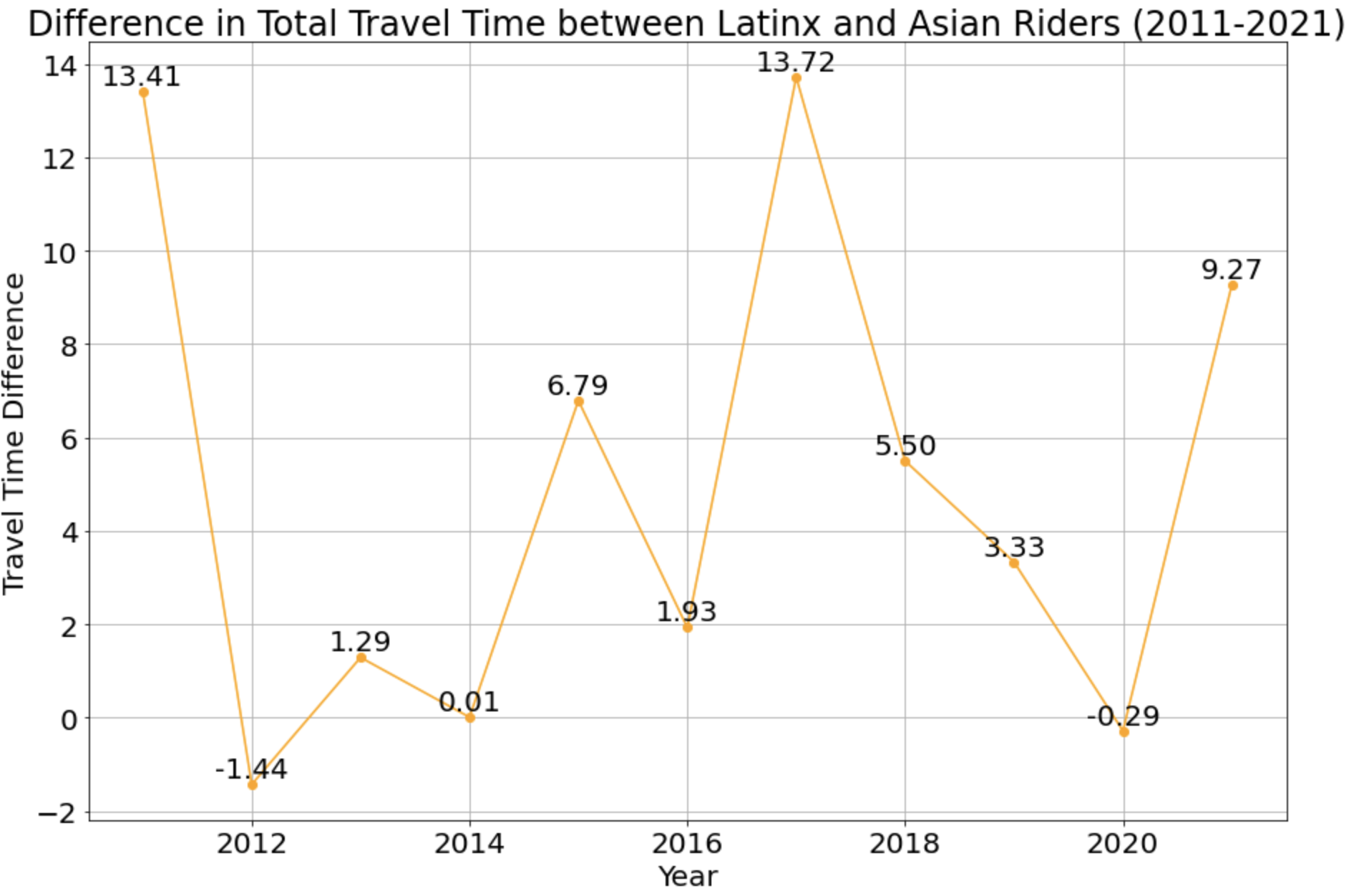
*Figure 9: Difference in Total Travel Time between LatinX and White Riders(2011-2021)*

****

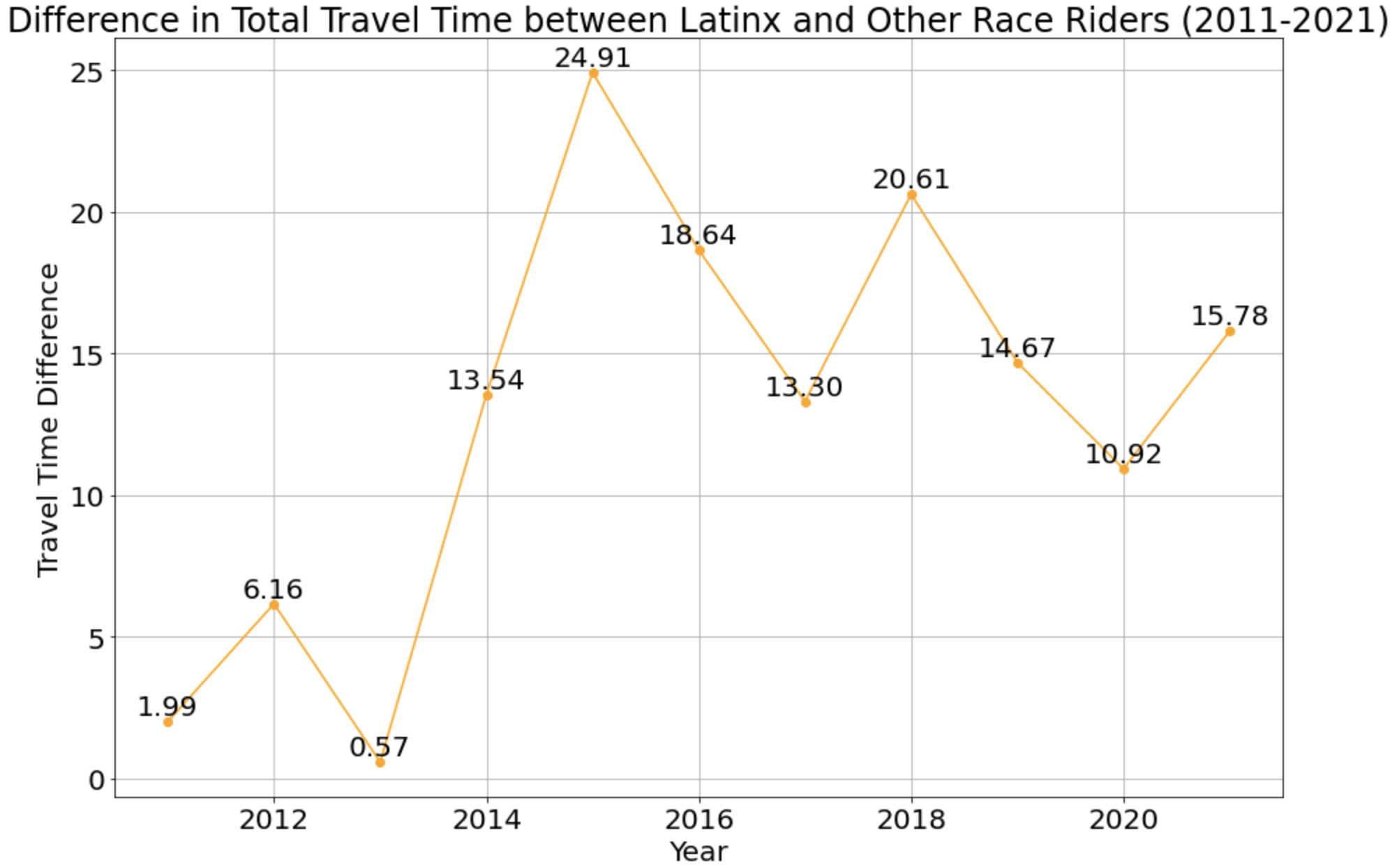
*Figure 10:Difference in Total Travel Time between LatinX and Black Riders(2011-2021)*

****

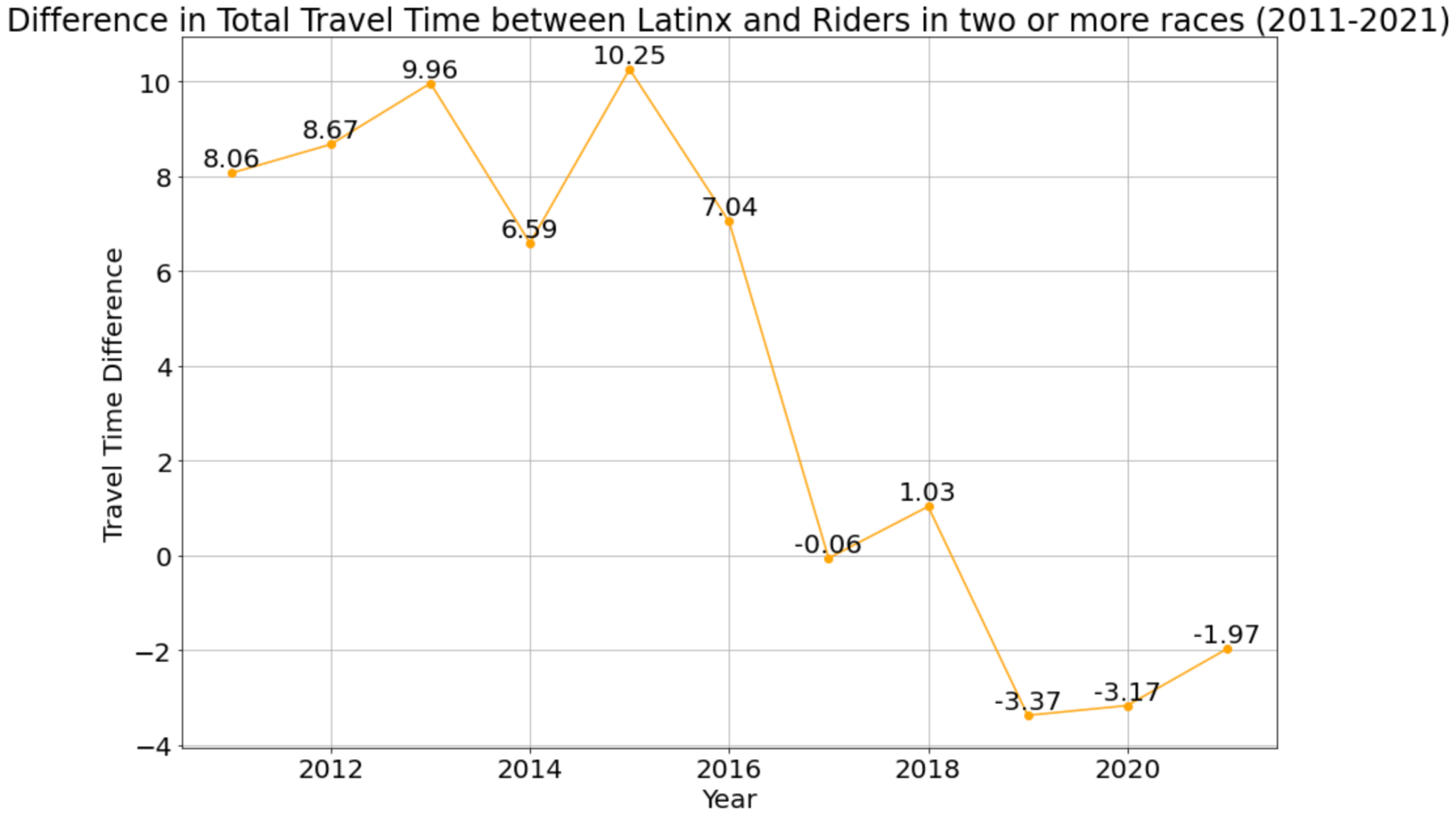
*Figure 11: Difference in Total Travel Time between LatinX and American indian Riders(2011-2021)*

****

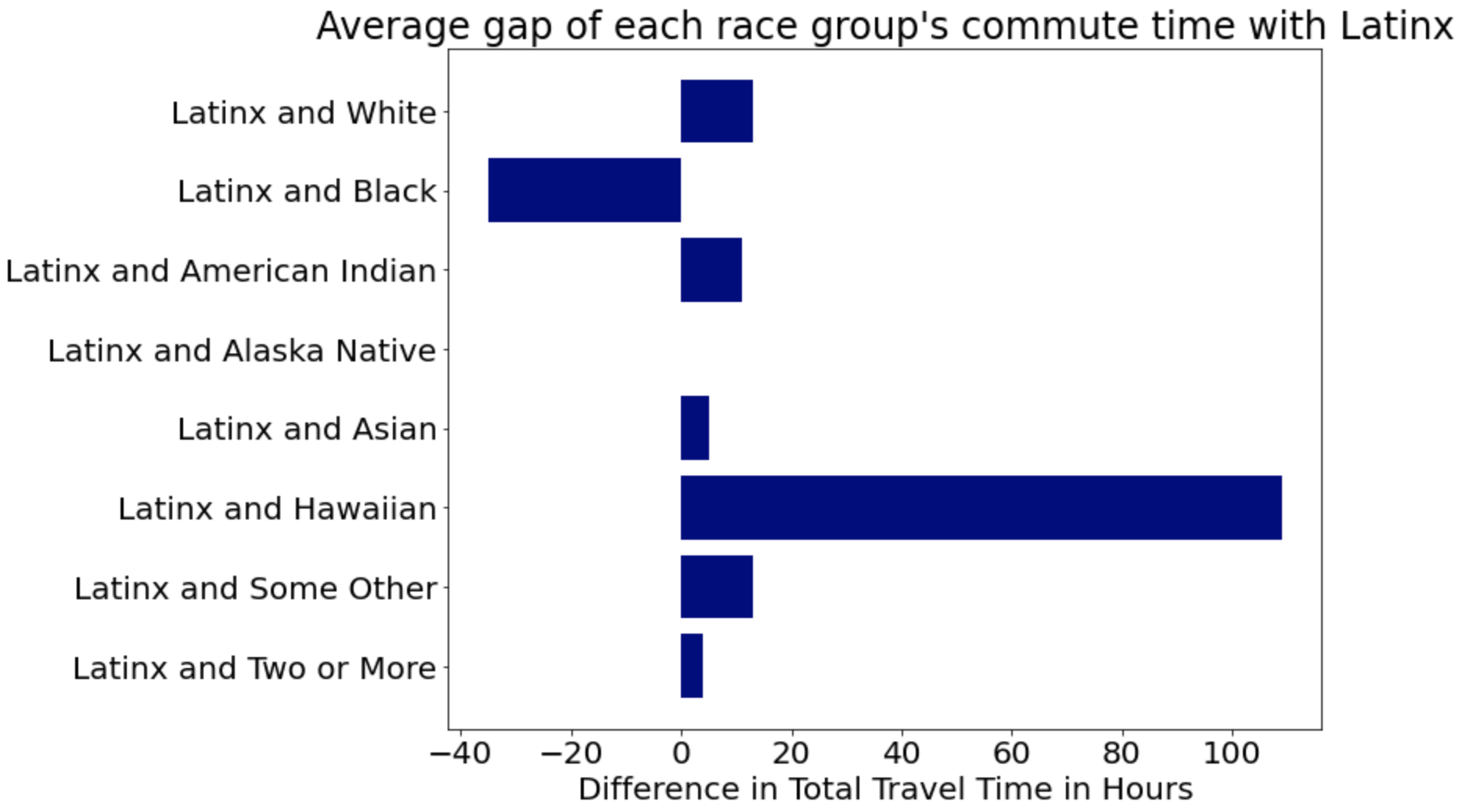
*Figure 12: Difference in Total Travel Time between LatinX and Asian Riders(2011-2021)*

****

*Figure 13: Difference in Total Travel Time between LatinX and Other Race Riders(2011-2021)*

****

*Figure 14: Difference in Total Travel Time between latinX and Other Race Riders(2011-2021)*

****

*Figure 15: Average Gap of Each Race Group’s Commute Time with LatinX*

We are using the five years of data from the year 2011 to the year 2021 to calculate the average commute time of each group by the formula:

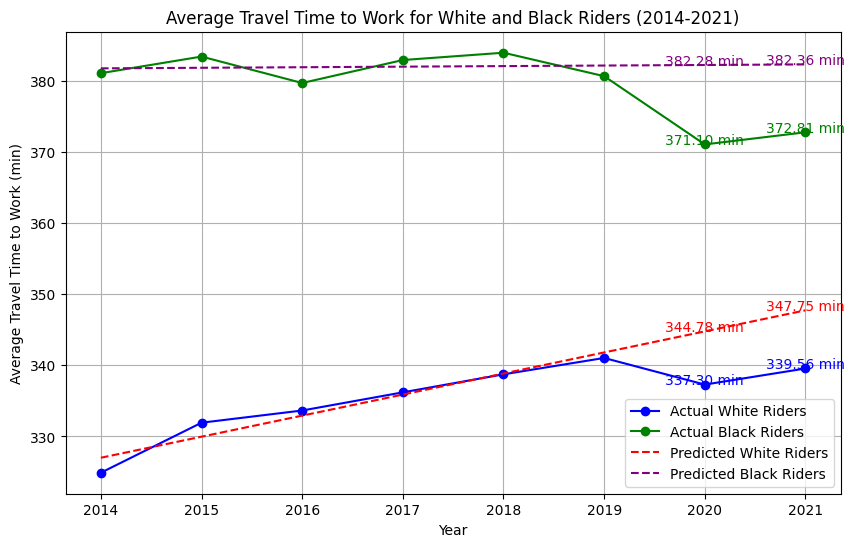
******

Based on the above plots, we can see there is a very significant gap between Latinx communities and Hawaiian communities which is a 109-hour gap. Compared to each group, there is an increased gap in the Two-or-more race rider group, the American Indian rider group. However, for other groups, there is a decreased gap overall. One interesting point to notice is that the Black group rider always has a longer commuting time than Latinx communities. In the next part, we would like to aim for some methods to address this point.

## Analysis for Question 4

To what extent did the COVID-19 pandemic impact the bus commute times for both Black and white riders, and how has this impact contributed to the existing disparity in hours spent on buses?

This question is based on conclusions from question 1. Figure 16 is the average commute time to work for black and white riders. The blue and green line represents the actual commute time, while the dashed lines indicate the time predicted by the linear model. For both Black and White riders, the actual commute times in 2020 and 2021 were lower than predicted. This suggests that the pandemic likely reduced commute times for both racial groups, possibly due to less traffic congestion as more people worked from home and there were fewer jobs to commute to because of lockdowns and economic slowdown.



*Figure 16: Average commute Time to Work for White and Black Riders*



*Figure 17: Disparity in commute Time between Black and White Riders with Linear Regression*

Figure 17 illustrates the disparity in commute time between Black and White riders. We can see that the actual disparity in 2020 and 2021 was greater than predicted. While commute times decreased for both groups, they did not decrease equally. The actual data for 2020 and 2021 show a smaller decrease or a plateau compared to the prediction, which means that the pandemic slowed down the trend toward decreasing disparity.

For 2020: The difference is approximately 6.98%.

For 2021: The difference is significantly higher at approximately 20.92%.

The larger difference in 2021 could imply that the changes in commute patterns due to the pandemic in 2020 may have extended in 2021, further affecting commute disparities.

## Analysis for Question 5

Have there been any policy interventions or initiatives aimed at reducing the bus commute disparity, and can data analysis determine the effectiveness of such measures over the years?

**Policy Recommendation:**

1. Develop data-driven regional planning that prioritizes route planning and service improvements with the largest time gaps. Collect and create better data, such as up-to-date transit route data, complete census data, or more scientific sampling. Utilize advanced data analytics tools and methods, such as machine learning and artificial intelligence, to predict more accurate future commuting patterns and demand. This will help policymakers better understand and anticipate long-term trends, leading to more effective strategies. Aim to make commute times equitable for all races.
2. Work deeply with communities with large time gaps(for example, the communities that are majority black) and prioritize understanding their specific needs and challenges. Also, adjust more appropriate service policies based on their feedback. Create a more open and interactive platform. Allow community members to participate directly in the transportation planning and decision-making process. This could be done by conducting online forums, community meetings, or using social media tools such as opening a Facebook account. The policy of adding more buses in the report of 64 *Hours: Closing the Bus Equity Gap* can be a good choice.
3. Redefine the Key Bus Routes program and adopt more flexible guidelines to allow for greater service frequency for routes with high rates of underserved riders. Provide customized transportation solutions based on the specific needs and characteristics of different communities. For example, more frequent bus service for certain communities or demand-responsive transportation for remote areas.
4. Establish long-term strategic planning. Policies should be adaptive and forward-looking, given the ongoing impacts of the COVID-19 disease and changing work and commuting patterns. This includes considering long-term trends and potential future scenarios for public transportation planning based on the impact of the epidemic on transportation equity and based on relevant regression analysis.

**Determine the Effectiveness of Previous Policy:**

The presentations on the [MBTA website](https://www.mbta.com/projects/bus-network-redesign) detail the improved plans for the bus network redesign from 2018 to 2022. MBTA has [identified High Priority Corridors](https://cdn.mbta.com/sites/default/files/2020-11/2020-11-23-fmcb-18-bus-transformation-update-bus-network-redesign.pdf) and redesigned bus networks. These initiatives are reflected in our data as well.

**For Pre-Covid:** According to analysis for question 1, we can find that the difference between the commuting time of black and white riders decreased year by year in the pre-Covid period. This shows that previous policies have been effective in reducing the inequality of commuting time for different racial groups in Massachusetts during the pre-Covid period.

**For COVID-Affected & After-COVID:** Due to factors such as the COVID-19, it's challenging to determine the direct correlation between these policies and the reduction in commute times for the COVID-19 period and after-Covid period. While the policies aimed to enhance the bus system, external factors like the pandemic could have significantly influenced commute patterns and times.

# Areas for Further Analysis

**Economic factors:** Examine if the bus commute time gap and policy effectiveness vary by specific regions within Massachusetts, considering socioeconomic and demographic factors.

Gather data on income levels, employment types, educational attainment, and age distribution for different regions in Massachusetts. Collect information on transportation policies implemented in Massachusetts, focusing on those aimed at reducing commute times and improving bus services. Use statistical methods to analyze the relationship between socio-economic factors and bus commute times. Look for patterns in how commute times and policy effectiveness vary among the regions in Massachusetts.

For example, explore whether areas with a higher proportion of working-class population exhibit different trends from areas with fewer working-class residents. This may involve comparing different counties, such as Weymouth County, in different economic contexts.

**Long-term effects:** MBTA's optimization of public transportation starts in 2019, and there are new plans implemented from 2023. We can extend the research timeframe to analyze the long-term impact of policy and COVID-19. This will capture the full scope of MBTA's optimization initiatives and the pandemic's impact.

A further distinction can be made in the future between the impact of policy changes and the COVID-19 pandemic on commuting patterns. Collect data on COVID-19 severity and restrictions in different regions over time, then examine how the pandemic affected public transportation usage and commute times. This may involve analyzing differences in areas of different pandemic severity or the timing of policy implementation. Highlighting the multiple factors contributing to the observed trends, and acknowledging that pandemic-related changes and policy shifts (such as changing bus routes) can all affect people's commuting behavior.

**Distinguish the impacts of policy and pandemic:** Comparing trends in regions with different levels of pandemic severity or timing of policy implementations.

Through a detailed examination of regions experiencing varying pandemic intensities and policy implementation schedules, we will explore how these factors, both individually and in conjunction, shaped public transport usage and commute times. For example, we can analyze regions with early/late policy adoption and high/low COVID-19 severity, then correlate transportation data with health and policy data to understand the interplay between policy implementation, pandemic trends, and commuting patterns. By carefully analyzing and comparing these variables, we hope to gain valuable insights into the effectiveness of policies during pandemic conditions and inform future transportation planning efforts for navigating similar situations.

# Limitations

1. Finding the correct data is the biggest limitation we have so far. We did not acquire the correct data until mid-October. This process is seriously slowing down our normal schedule which results in us not having enough time to do further analysis. We also put the graph drawn using inappropriate data in the appendix part for reference and to provide our timeline for this project.
2. We also have a problem while processing the data. This is no limit to the data missing and data type errors when processing the data. The amount of data is not like census data, but if it is sampled data lacks uniform management. For example, the data we used had some missing values for the travel time to work part and there is some data that is not collected in the correct type which needs transformation. This may result in an inaccurate evaluation of the commute gap between different groups.
3. We can only access data up to 2021, so it is difficult for us to analyze the long-term impact of COVID-19 on public transportation.
4. Since the data is census data, it is very large. The data over the years has led to very confusing labeling categories, for example, the PUMA geo labels don't include the full MA region for each year and are different for each year. The definitions are also different for pre and post-2020.

# Conclusion

Overall, our report reveals that commute times in Massachusetts by different races, particularly between blacks and LatinX communities with other races. During the COVID-19 in 2020 and 2021, commute times for both black and white riders were lower than expected, suggesting that bus commute times were reduced, possibly due to changes related to COVID-19. The results also highlight the need for a resilient and equitable transport system. This provides insight into how the transport system works under unconventional circumstances.

Our report notes that there are also large differences between regions, with the largest differences in areas such as Norfolk County and Weymouth Township. Our analysis notes that these differences are caused by the education level. An interesting finding is that the difference between black and white riders' commuting times is smaller in the central region of Massachusetts. However, the difference is large at the edge region of Massachusetts. The central region may be more developed and invested in, thus having better and more efficient public transport systems. However, in the marginal areas of Massachusetts, the level of infrastructure may be relatively low due to limited resources or insufficient investment. The report also noted a significant difference in commute time between Latinx communities and Hawaiian communities 109 hours. For the other groups except for the rider group of two or more races and the American Indian rider group, the overall disparity in bus commute time decreased.

While the MBTA has taken many steps to address these issues, such as redesigning the transportation network, the effect of these measures is difficult to assess due to the impact of COVID-19. Our projections suggest that these commuting time gaps are likely to persist if more appropriate policies are not implemented. The differences in commuting times between different regions reflect uneven regional planning and service quality, underscoring the importance of regional analysis and improvement. The report makes specific recommendations for data-based regional planning, prioritizing areas with large disparities in commute times, and recommending the redefinition of key bus routes to increase the frequency of service. We recommend further research on economic factors, policies, and the long-term impact of the epidemic in the future to more comprehensively address transport inequalities.

Our report highlights racial commuting time disparities in the Massachusetts transit system and makes specific strategies and recommendations to reduce these disparities. At the same time, it also demonstrates a profound insight into the impact of COVID-19 on public transport and provides directions for future research, policy-making and decision-making.

# Appendix

* **Pre-Process:**

We find the corresponding data by using the US Census data. As the requirement from our client, we will focus on these three areas:  **Roxbury, Dorchester, and Mattapan**. We follow the graph to access these three places by their census id.



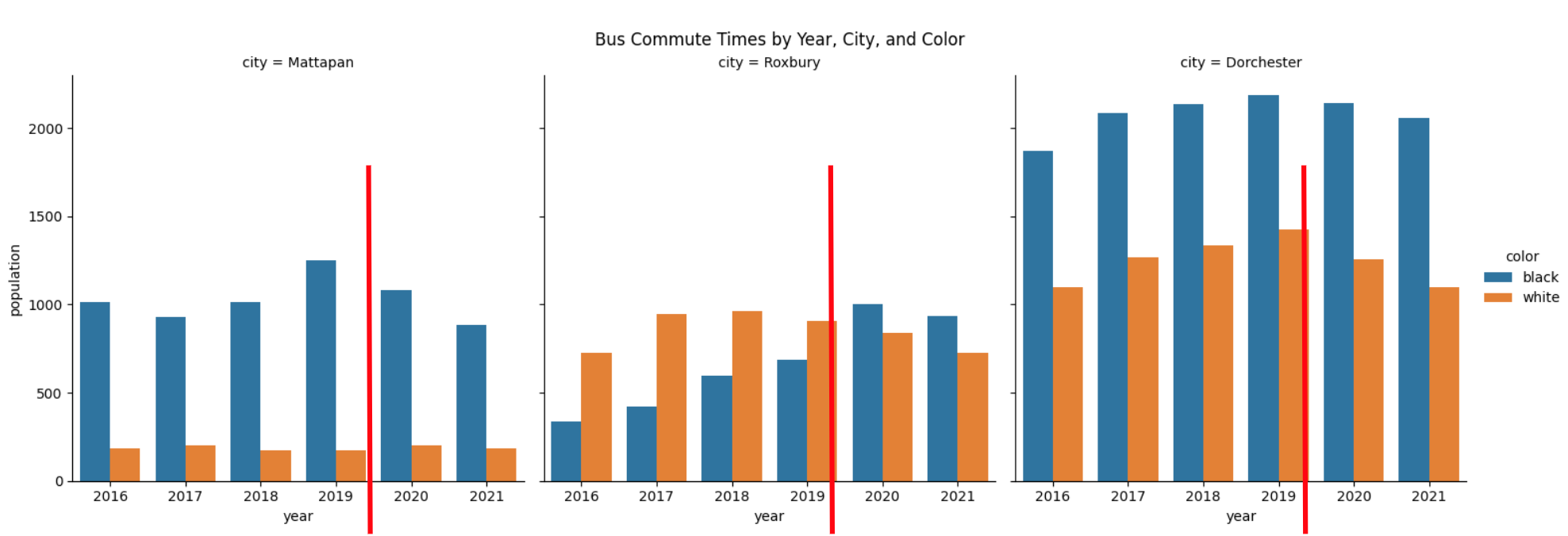
* **Additional Visualization for Question 1 Analysis:**

**Demographic information for Black and White riders of public transportation**

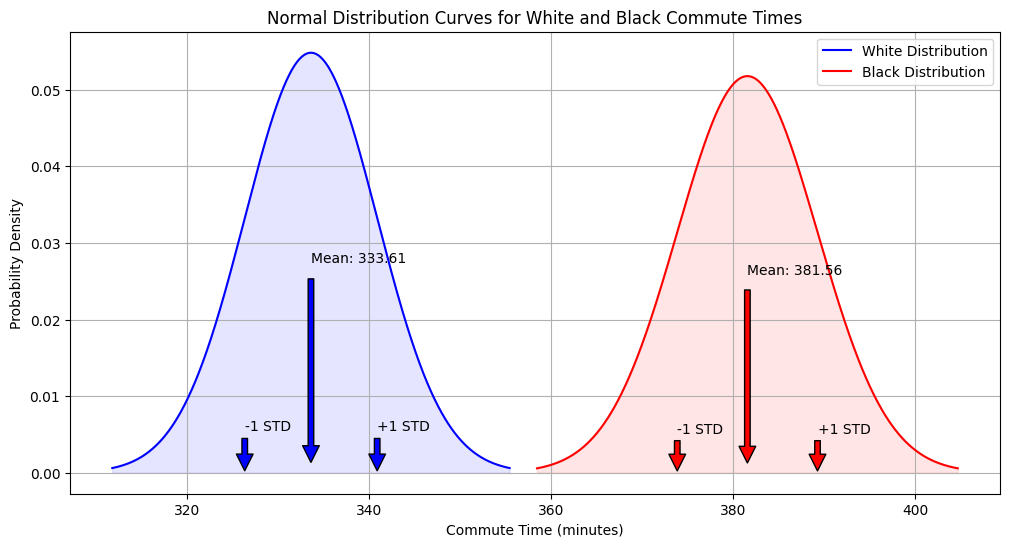
**Source:**

[MEANS OF TRANSPORTATION TO WORK BY TRAVEL TIME TO WORK](https://data.census.gov/table/ACSDT5Y2020.B08134?q=B08134&g=050XX00US25025$1400000)

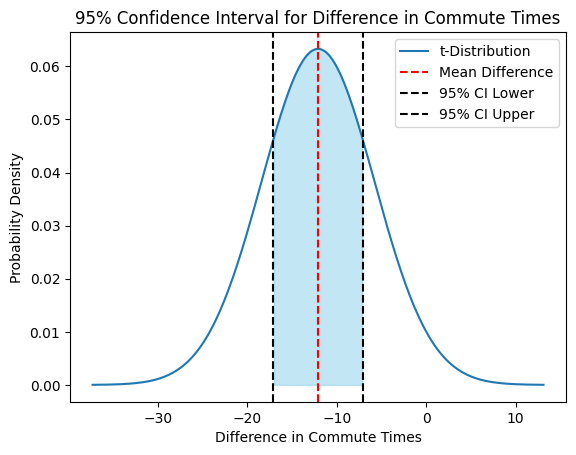
[United States Census Bureau](https://www.census.gov/programs-surveys/acs/microdata/access.html)

****

*Figure 1: Bus Commute Population Difference between Black and White Riders*

**

*Figure 2: Normal Distribution Curves for Whites and Black Commute Times*

**

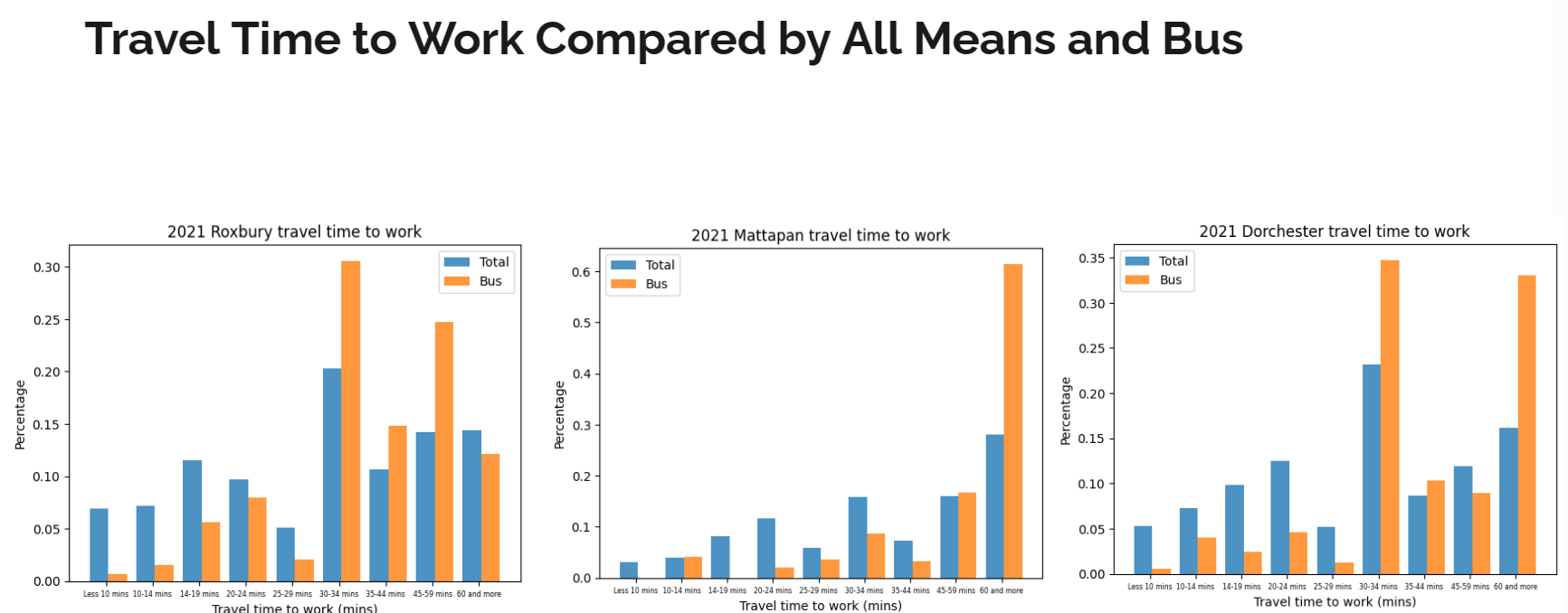
*Figure 3: 95% Confidence Interval for Difference in Commute Times for T-distribution*

* **Additional Visualization for Question 2 Analysis:**

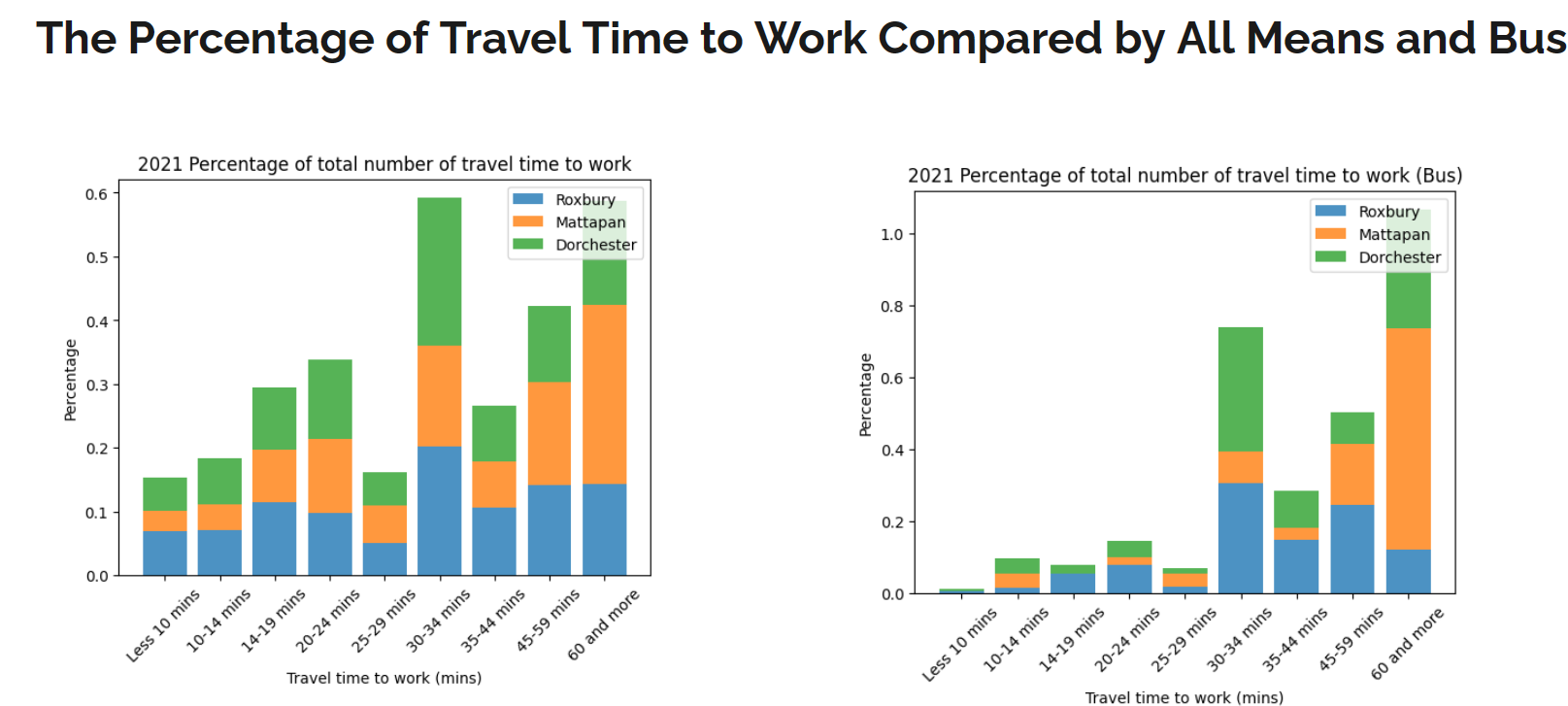
**Commuting time in different areas.**

**Source:**

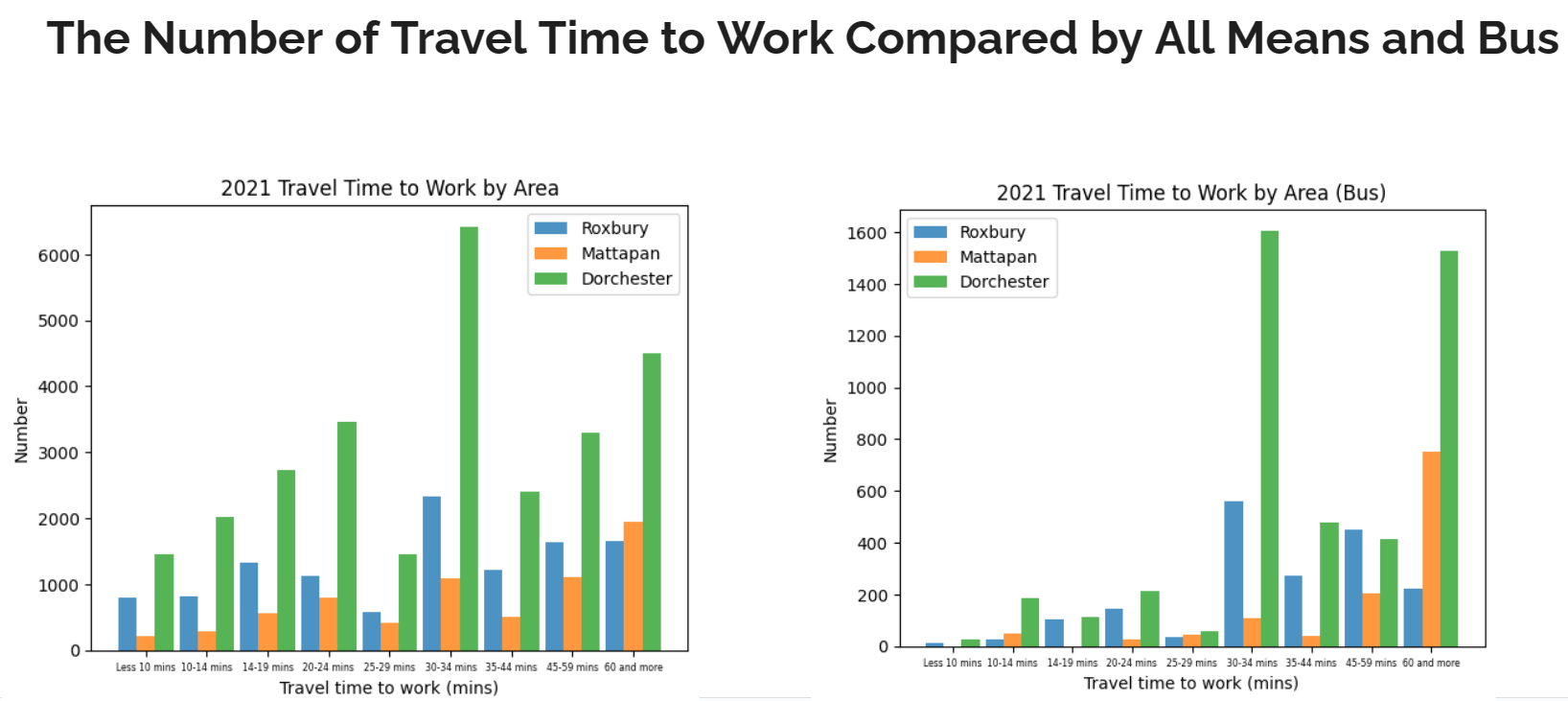
[B08134: Means of Transportation to Work by Travel Time to Work](https://data.census.gov/table/ACSDT5Y2021.B08134?q=B08134:%20Means%20of%20Transportation%20to%20Work%20by%20Travel%20Time%20to%20Work&g=050XX00US25025$1400000)



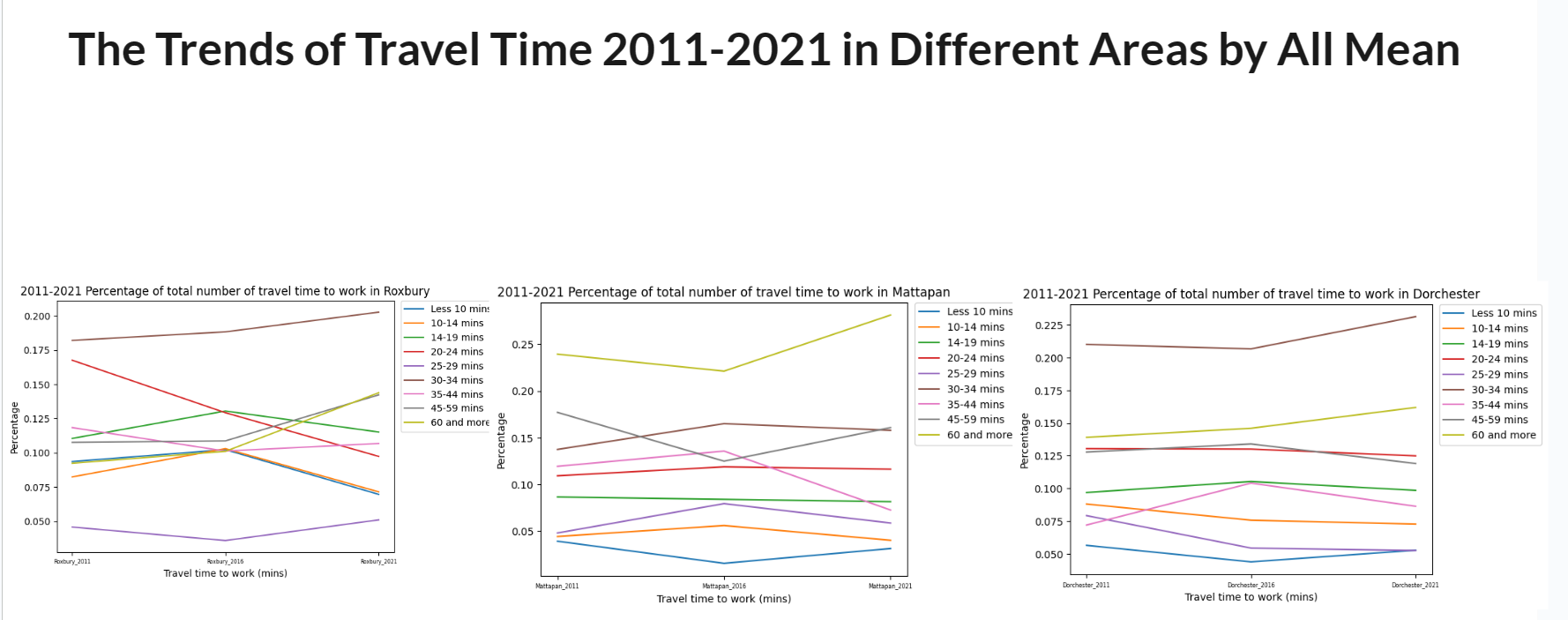
*Figure 4: Travel Time to Work Compared by All Means and Bus*

**

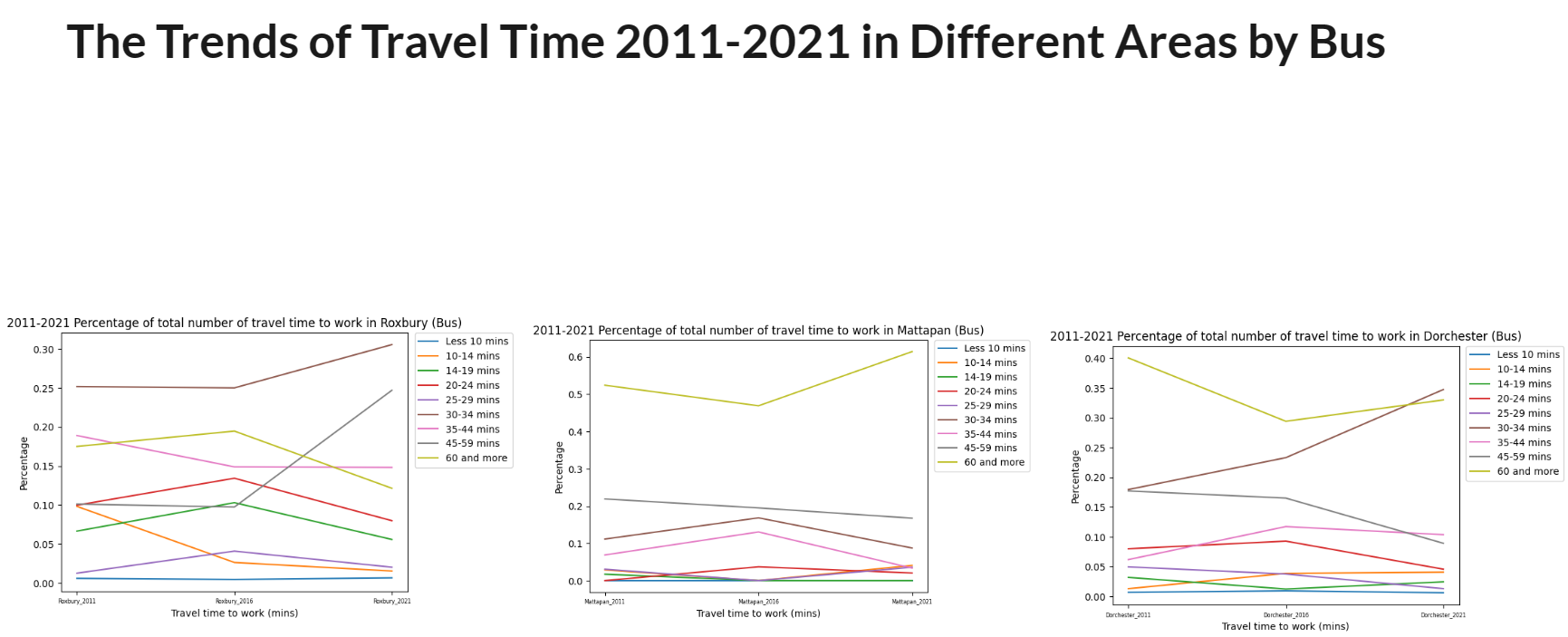
*Figure 5: The Percentage of Travel Time to Work Compared by All Means and Bus*

**

*Figure 6: The Number of Travel Time to Work Compared by All Means and Bus*

**

*Figure 7: The Trends of Travel Time 2011-2021 in Different Areas by All Mean*

**

*Figure 8: The Trends of Travel Time 2011-2021 in Different Areas by Bus*

* **Further analysis:**

We conducted a regression analysis and selected variables that have the greatest impact on commuting time for further analysis:

1. JWMNP: Journey to Work Minutes
2. PUMA: Public use Microdata Area
3. RACIP: Race
4. AGEP: Age
5. SCHL: Schooling
6. WKW: Weeks Worked
7. SEX
8. OCCP: Occupation
9. INDP: Industry

* **Top Positive Influencing Variable for Each Category:**

1. PUMA\_4903
2. RAC1P\_2
3. SEX\_2
4. OCCP\_6320.0
5. INDP\_290.0

* **Top Negative Influencing Variable for Each Category:**

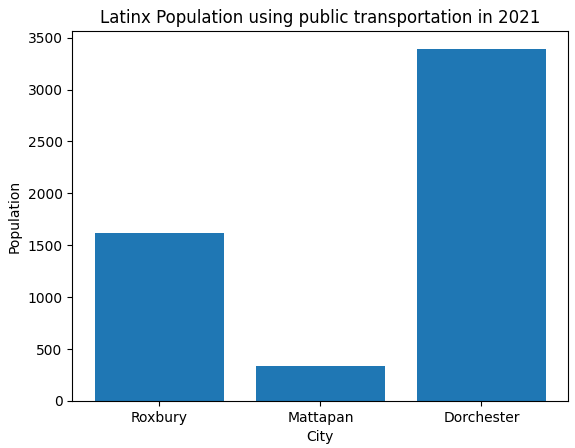
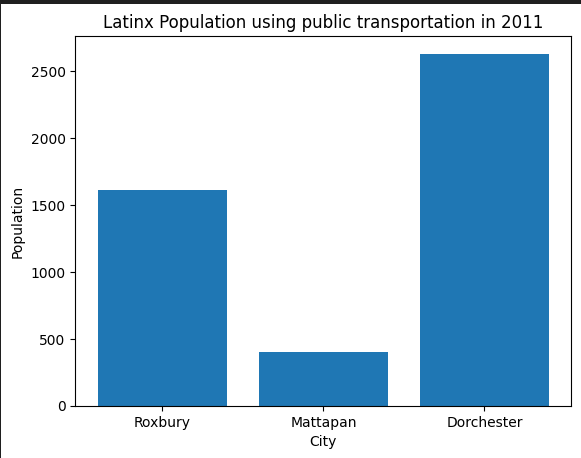
1. PUMA\_200
2. OCCP\_6765.0
3. INDP\_2880.0

* **Additional Visualization for Question 3 Analysis:**

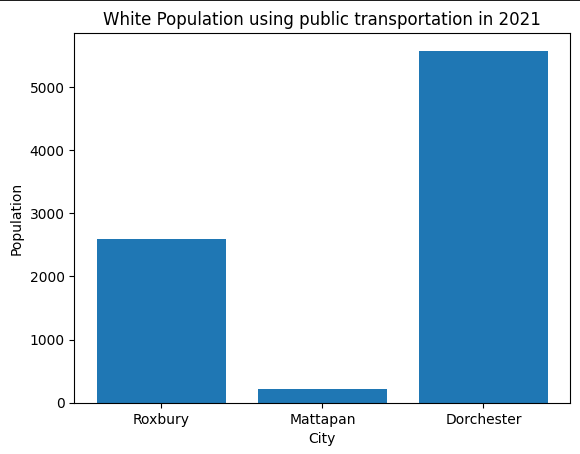
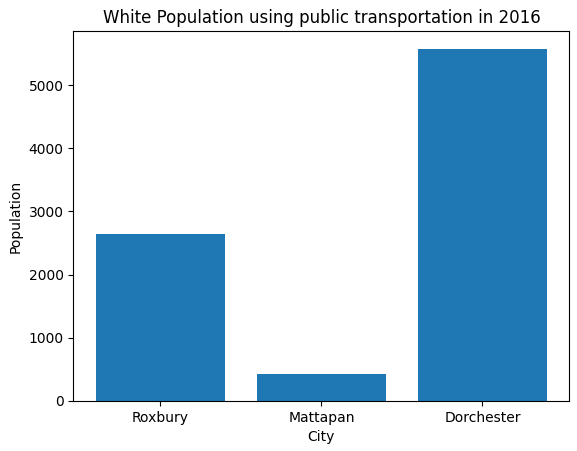
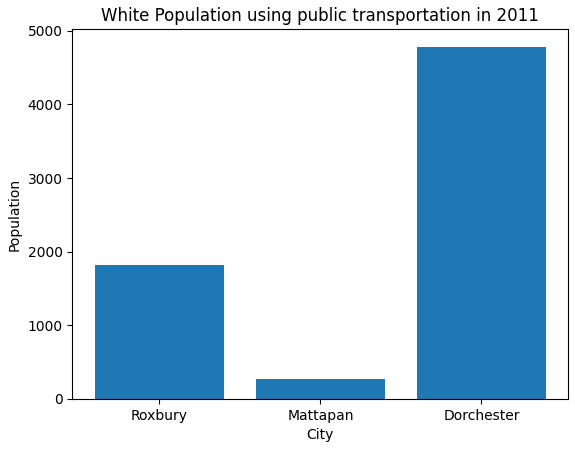
**Demographics information**

**Source:**

[B08105IMEANS OF TRANSPORTATION TO WORK (HISPANIC OR LATINO)](https://data.census.gov/table/ACSDT5Y2021.B08105I?q=B08105&g=050XX00US25025$1400000) (data from 2011, 2016, 2021)



*Figure 9:LatinX population using public transportation in 2011, 2016 and 2021*



*Figure 10:White population using public transportation in 2011, 2016 and 2021*

* **Additional Visualization for Question 4 Analysis:**

**Commuting time distribution**

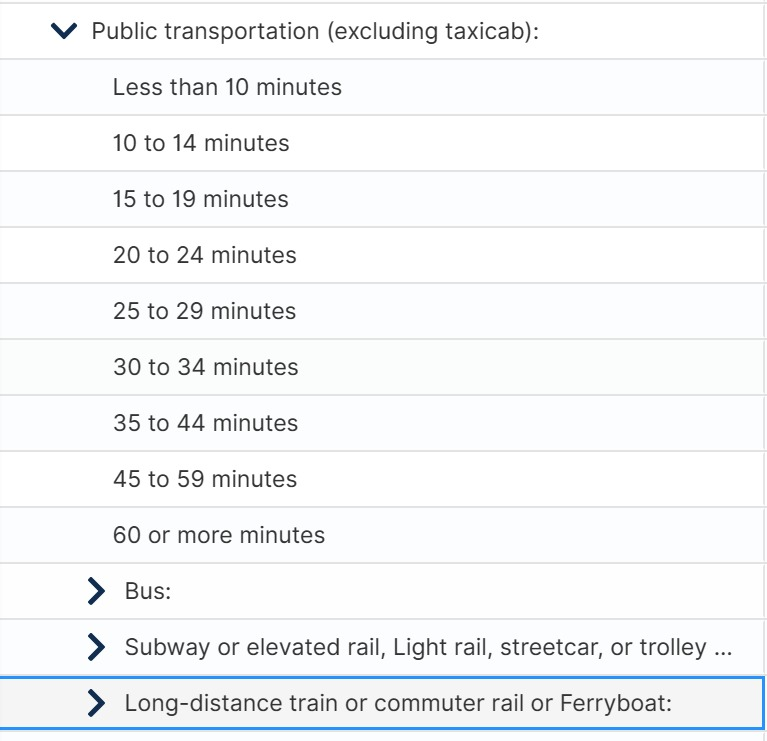
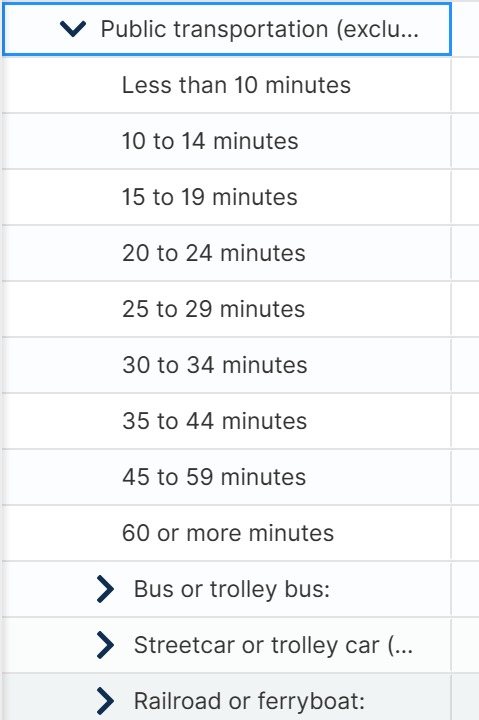
**Source:**

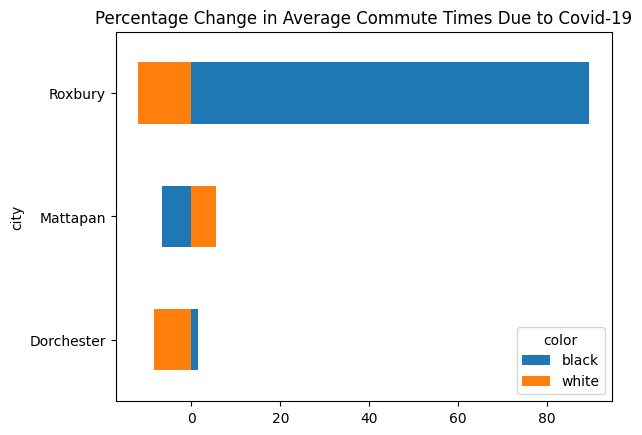
[B08134: Means of Transportation to Work by Travel Time to Work](https://data.census.gov/table/ACSDT5Y2021.B08134?q=B08134:%20Means%20of%20Transportation%20to%20Work%20by%20Travel%20Time%20to%20Work&g=050XX00US25025$1400000) (2016-2021)

[B08105A: MEANS OF TRANSPORTATION TO WORK(WHITE ALONE)](https://data.census.gov/table/ACSDT5Y2021.B08105A?q=B08105&g=050XX00US25025$1400000) (2019-2021)

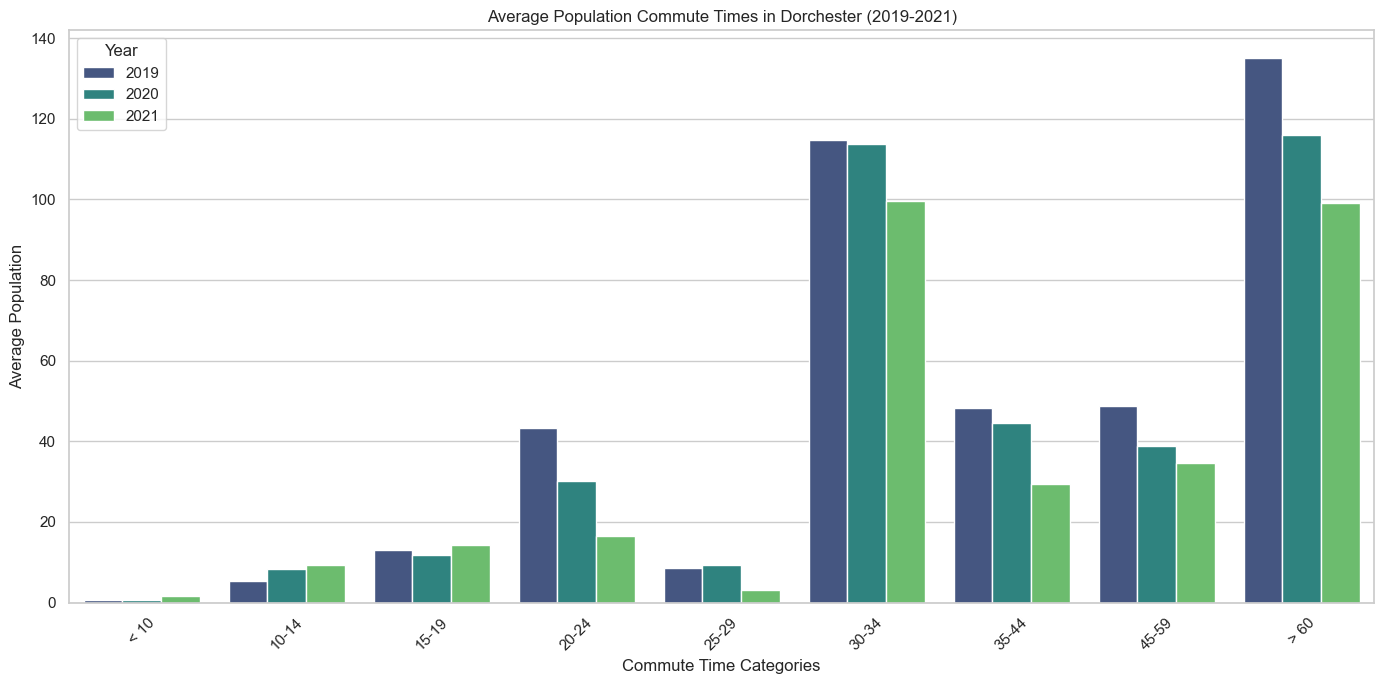
[B08105B | MEANS OF TRANSPORTATION TO WORK (BLACK OR AFRICAN AMERICAN ALONE)](https://data.census.gov/table/ACSDT5Y2021.B08105B?q=B08105&g=050XX00US25025$1400000)

(2019-2021)

****



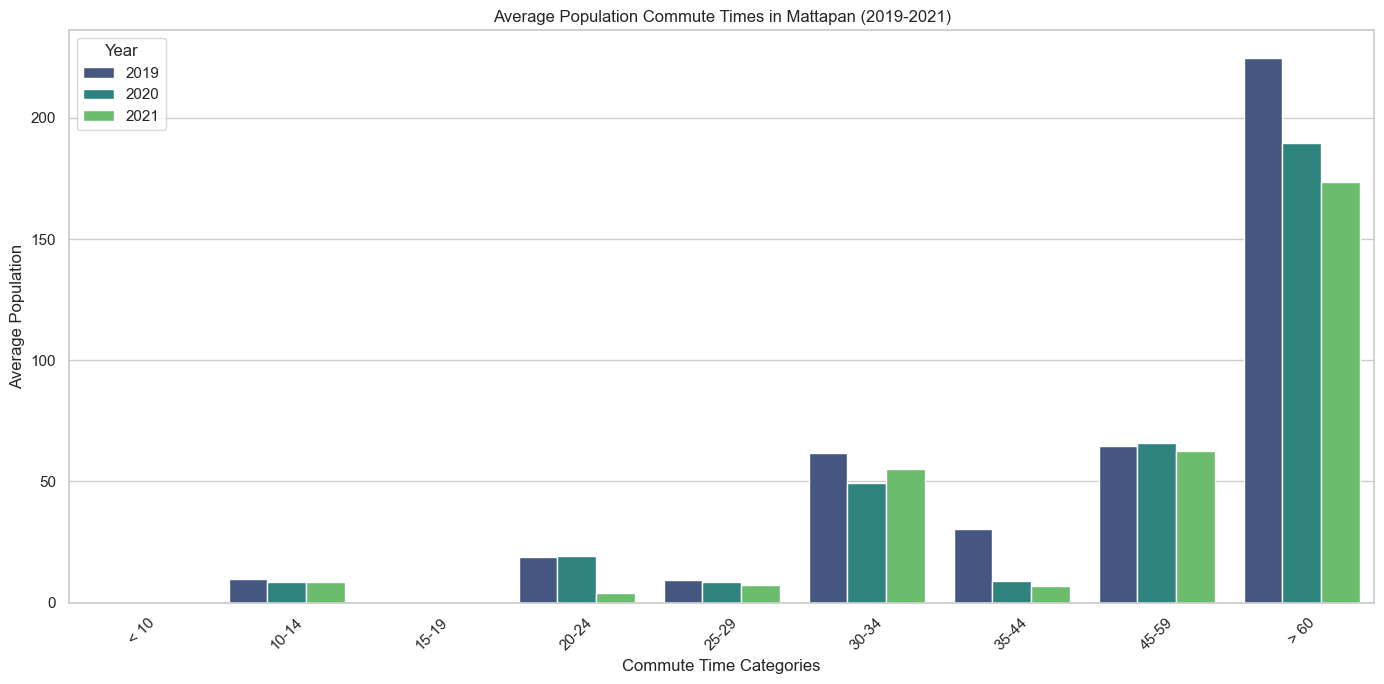
*Figure 11:Percentage change in average commute times due to Covid-19*



*Figure 12:Average population commute times in Dorchester(2019-2021)*



*Figure 13:Average population commute times in Roxbury(2019-2021)*



*Figure 14:Average population commute times in Mattapan(2019-2021)*