



# Executive Summary

12.13.2023

**Group 4**

**Viswadatta Chinta**

**Domonick Demas**

**Joseph Lee**

**Nirbhay Singh Naruka**

**Prit Shah**



## Introduction and Problem Specification

Our final proposal for SMART aims to provide in-depth insights into the dynamics of post-pandemic urban mobility, with an emphasis on the Los Angeles metropolitan area. We intended to investigate how diverse factors interact to impact public transportation utilization by examining datasets comprising numerous socioeconomic indicators and public transportation variables from June 2021 to December 2023. This period was especially noteworthy because of the changing environment of urban movement in the aftermath of COVID-19, with elements such as pandemic limitations, economic developments, and social dynamics all playing important roles.

In our detailed study for SMART, we have carefully selected a set of key variables that are crucial for a deeper understanding of public transportation trends in the post-pandemic era. Our dataset primarily includes monthly counts of passengers using Metro and rail services. To add context, we've included monthly unemployment rates, offering insights into how the local economy's health might influence public transportation choices. Recognizing the impact of fuel costs, we also considered variations in gas prices, understanding their potential role in shaping transit preferences. Given the significant effect of the COVID-19 pandemic on travel behaviors, we incorporated data on monthly COVID-19 case counts, to gauge both its direct and indirect influences on public mobility. We also looked at crime rates to understand how perceptions of safety might affect transit usage. Finally, we analyzed fare pricing within the Los Angeles Metro Transit Authority (LA MTA), to see how changes in fares might sway ridership levels.

## Trends and Correlations

In our analysis, a key finding was the positive correlation between the number of COVID-19 cases and public transportation ridership. However, it likely indicates that as COVID-19 restrictions were lifted and people started to return to their normal routines, there was a corresponding increase in the use of public transportation, despite the rising number of cases. This suggests that the willingness or necessity to use public transit grew as the immediate concerns of the pandemic began to ease. Conversely, an inverse relationship was noted between unemployment and public transportation use, indicating that economic factors play a significant role in transit decisions. As unemployment rates increased, a corresponding decrease in public transportation usage was observed. Furthermore, our correlation matrix provides a complete perspective of how numerous elements interact with one another, such as crime rate, petrol costs, unemployment rates, COVID-19 case counts, and ridership.

## Contributions

As a collaborative effort within our team, the selection of these specific variables for the SMART project was a thoughtful and strategic process. Months were chosen to account for potential seasonality, recognizing that transportation patterns may vary throughout the year. The incorporation of the Crime rate and Unemployment serves to gauge the overall urban environment, impacting public perception and the economic context in which transportation systems operate. Gas prices were considered due to their direct influence on travel costs, affecting



the affordability of public transportation. In the current global context, including COVID cases is vital to understanding the pandemic's impact on travel behavior and safety concerns. Meanwhile, Ridership and Rail Fare are central to SMART's mission, directly reflecting the public's reliance on public transportation and the system's financial sustainability. This collaborative approach ensures a holistic analysis, providing SMART with actionable insights to optimize urban transportation in terms of efficiency, convenience, environmental friendliness, and fiscal responsibility.

## Data Collection

**Finance and Budget - LA Metro:** The data source, published by the Los Angeles County Metropolitan Transportation Authority (Metro), serves as a reliable reference for various financial documents. As the public transportation agency for Los Angeles County, California, Metro provides access to budget and financial reports, the Comprehensive Annual Financial Report, Metro's annual audits, and the National Transit Database. These resources have been instrumental in helping us comprehend Metro's primary financial sources. Consequently, this understanding has enabled us to focus our data collection efforts specifically on light rails.

**Metro Ridership:** This website is an official source of ridership data from the Los Angeles County Metropolitan Transportation Authority (Metro). It allowed us to view the monthly passenger counts across different modes of transit, such as Metro, rail, and bus services. We got the ridership data of each light rail line from June 2021 to June 2023, which was required for our analysis. The data included the total monthly ridership which was used to measure the public transit utilization and to correlate it with other variables in their dataset.

**Crime Data from 2020 to Present | Los Angeles:** This dataset contains crime incidents reported in the City of Los Angeles since 2020. We downloaded the data as a csv file and filtered it to include only the areas within approximately 3 miles of the metro rail stations and their adjacent neighborhoods. This way, we could focus on the impact of crime rates on public transit.

**Los Angeles Retail Regular Gas Price:** Our analysis aimed to examine the relationship between gas prices and LA metro ridership. We collected the monthly gas price data from 2020 to 2023 by manually scraping the website, since we could not download the full dataset without a subscription. We used this data to train our model and test our hypothesis.

**COVID Data:** We obtained the COVID-19 cases data from USAFacts<sup>1</sup>, a reliable source of public information. We wanted to explore how the pandemic and the related lockdown measures influenced the public transit demand in Los Angeles. We assumed that people would be less likely to use the metro light rails due to the health risks and the travel restrictions. We processed the data to convert the daily number of cases in LA County to a monthly average. This way, we could compare the Covid-19 trends with the ridership data over time.

**Unemployment Data:** We obtained the unemployment rate data from charts, a trustworthy source of economic indicators. We speculated that the unemployment rate would have a negative impact on ridership, as fewer people would need to travel by metro if they were out of work.

## Summary Statistics

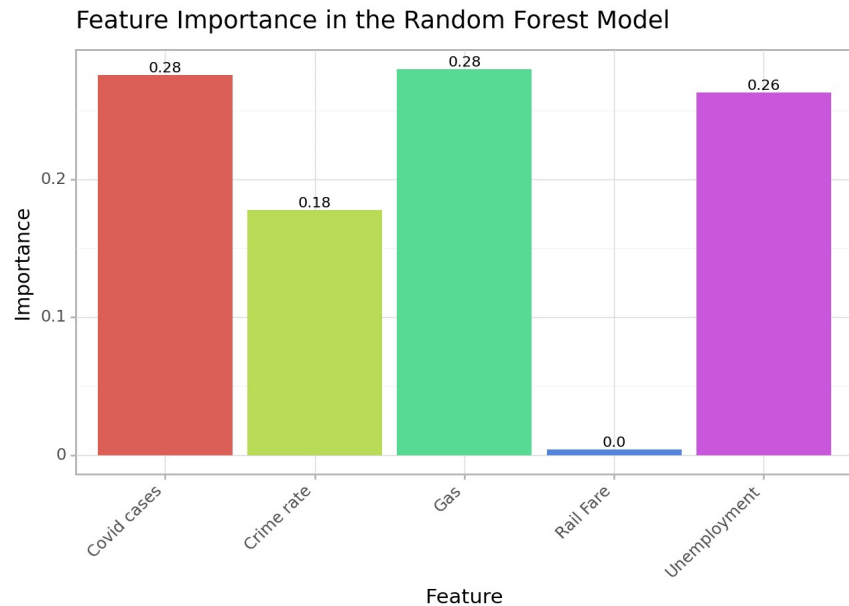
Parameter	Count	Mean	Min	0.25	0.5	0.75	Max	Std
Crime rate	25	19017.96	17153	18374	18762	19784	20416	877.62
Gas	25	5.19	4.38	4.68	5.02	5.6	6.51	0.63
Unemployment	25	5.86	4.4	4.8	5.1	6.5	9.9	1.62
Covid cases	25	76874.6	36164	43777	86382	99026	108439	27072.38
Ridership	25	4744448.88	4004502	4505616	4718506	5067135	5509095	370891.7
Rail Fare	25	93.28	72	100	100	100	100	12.2

The summary statistics reveal important insights into our dataset. Firstly, the stable crime rate, with a mean of around 19,018 incidents and a relatively tight range, suggests a consistent security environment. Secondly, the gas price averages around 5.19, reflecting the cost of transportation, which is critical to our riders. Thirdly, the unemployment rate varies with an average of 5.86, signaling potential impacts on ridership. Fourthly, the wide range in COVID-19 cases, averaging around 76,875, highlights the uncertainty and challenges we faced due to the pandemic. Our ridership data, with an average of 4,744,449 riders, shows some fluctuations, possibly influenced by these factors. Lastly, our rail fare data remains relatively stable with an average of 93.28.

## Model Procedure and Results

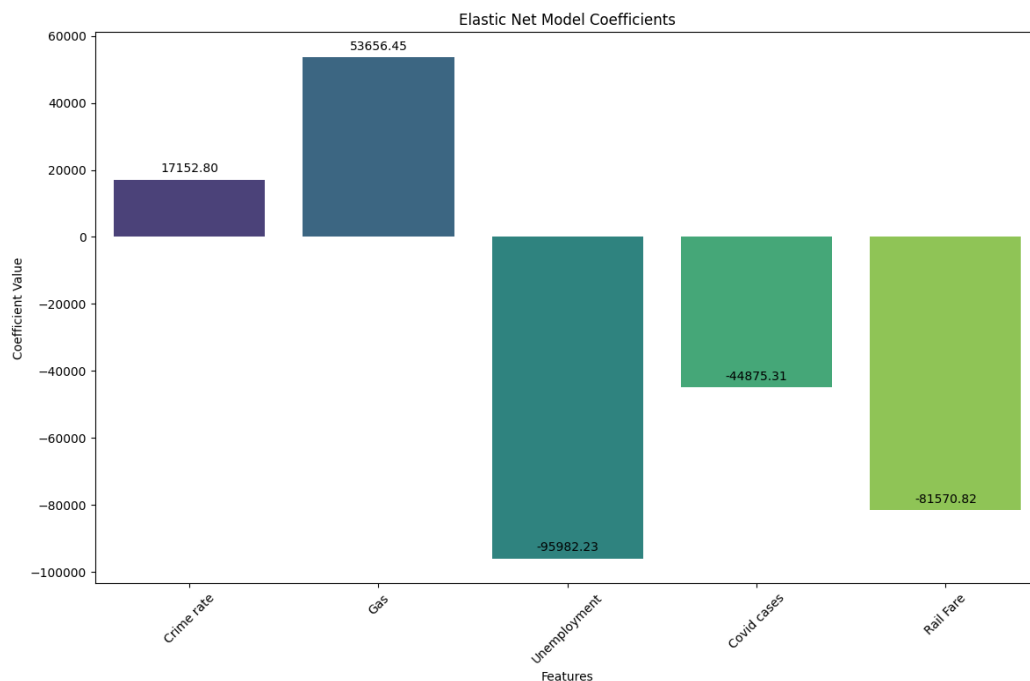
### Random Forest Model

The Random Forest model emerged as the most accurate in predicting public transportation ridership, with the lowest RMSE of 228901.79 among the evaluated models. This ensemble method, known for its robustness and ability to capture complex, non-linear relationships, outperformed Linear Regression, Ridge, Lasso, and Elastic Net in our analysis. For SMART's strategic planning, this model offers valuable insights, particularly in understanding how various factors like crime rate, gas prices, and unemployment rate influence ridership. Regular updates with new data and further analysis of the most influential features are recommended to maintain the model's relevance and support informed decision-making.



In our Random Forest model analysis, we found that COVID cases, gas prices, and unemployment rates are the most influential factors in predicting public transportation ridership, each accounting for approximately 28%, 28%, and 26% of the model's predictive power, respectively. This highlights the strong correlation between public transportation usage and these key economic and public health indicators. Interestingly, the model indicated that rail fare, which remained constant during the study period, had no significant impact on ridership predictions. These insights are crucial for strategic decision-making, suggesting that adjustments to transit services should consider changes in gas prices, unemployment trends, and public health conditions to effectively meet ridership demands and adapt to changing circumstances.

## Elastic Net Model

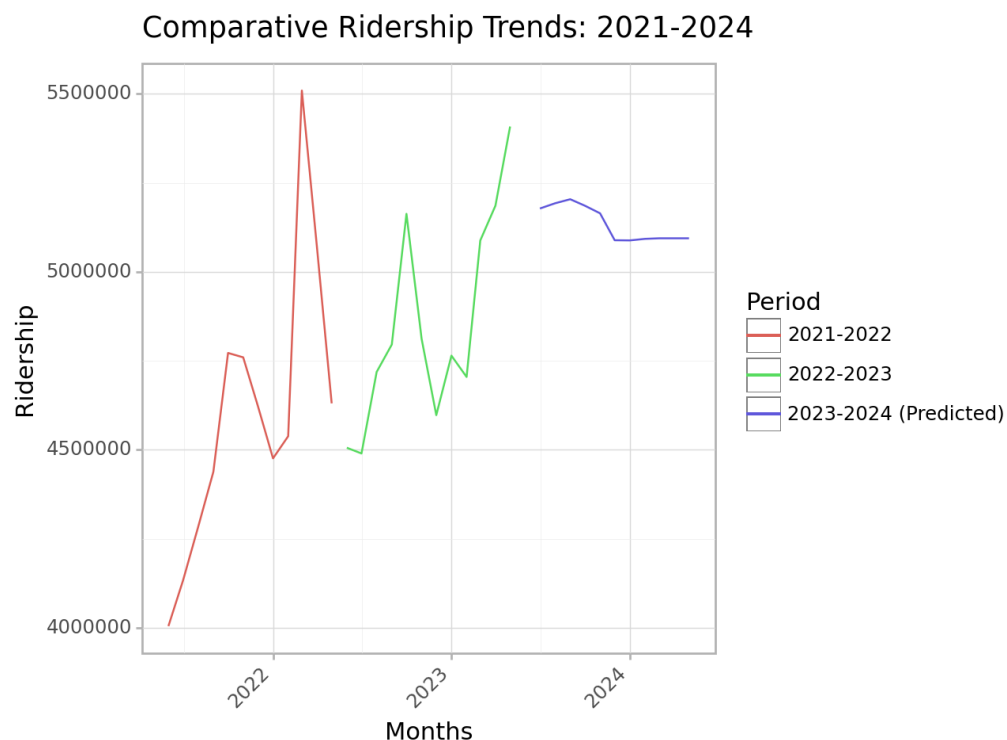


The graph presents the factors derived from an Elastic Net regression analysis. This analysis reveals key predictors' impact on the outcome. Gas prices show the strongest positive influence, indicating increased outcomes with higher prices, likely due to cost-related consumer behavior changes. Unemployment exhibits a significant negative correlation with the outcome, while Covid cases also negatively affect the outcome, possibly due to pandemic effects. Rail fare is negatively linked to outcomes, signifying sensitivity to pricing. Crime rate, with a small positive coefficient, has a minor impact on outcomes compared to other factors.

## Conclusions and Recommendations

We chose Random Forest as our primary modeling technique due to its balance of simplicity and explanatory power. Our model's aim was not just to forecast ridership but also to provide actionable insights for service planning, resource allocation, and policy formulation. The model's predictive capability is essential for scenario planning, allowing SMART to respond proactively to shifts in ridership patterns.

The graph below shows ridership trends for the years that were analyzed post-COVID lockdown, 2021-2023, and also the predicted years, 2023-2024. There was a slight decrease in ridership in the winter months of 2023-2024, but a pattern of sustained ridership was predicted. To improve ridership, we recommend SMART focus on advertisements near gas stations and also in cities where unemployment rates were higher than the county average. Fare prices did not seem to affect ridership as they have been static at a monthly rate of \$72. If this were to rise, it may affect ridership negatively, so we would recommend that fair prices not increase.





## Citations

<https://www.metro.net/about/financebudget/>

<https://isotp.metro.net/MetroRidership/YearOverYear.aspx>

<https://data.lacity.org/Public-Safety/Crime-Data-from-2020-to-Present/2nrs-mtv8>

[https://ycharts.com/indicators/los\\_angeles\\_retail\\_price\\_of\\_regular\\_gasoline](https://ycharts.com/indicators/los_angeles_retail_price_of_regular_gasoline)

<https://usafacts.org/visualizations/coronavirus-covid-19-spread-map/>

[https://ycharts.com/indicators/los\\_angeles\\_ca\\_unemployment\\_rate#:~:text=Los%20Angeles%2C%20CA%20Unemployment%20Rate%20is%20at%205.30%25%2C%20compared,but%20are%20with out%20a%20job.](https://ycharts.com/indicators/los_angeles_ca_unemployment_rate#:~:text=Los%20Angeles%2C%20CA%20Unemployment%20Rate%20is%20at%205.30%25%2C%20compared,but%20are%20with out%20a%20job.)