Towards a Better LA Metro Rail: A Comprehensive Proposal for SMART

10.20.2023

Group 4

Viswadatta Chinta - Unstucker

Domonick Demas - Research god

Joseph Lee - Liaison of affairs

Nirbhay Singh Naruka - Content creator

Prit Shah - Supreme team leader

Overview

In response to the growing concerns surrounding urban congestion and the need for sustainable transportation solutions in Los Angeles, this project proposal to SMART outlines a comprehensive strategy to enhance the efficiency, ridership, and overall appeal of the LA Metro light rail. Drawing on data-driven insights, we aim to identify key challenges and opportunities, proposing actionable measures that align with SMART's mission to foster metropolitan transportation systems that are environmentally friendly, fiscally responsible, and highly efficient for the residents of Los Angeles.

Data

Provide links or citations to the datasets you plan to use.

Finance and Budget - LA Metro
Records the finances of LA Metro over the course of the years
Crime Data from 2020 to Present | Los Angeles - Open Data Portal (lacity.org)
(to find the correlation between crime in LA and the usage of public transport)
This dataset reflects incidents of crime in the City of Los Angeles dating back to 2020. This data is transcribed from original crime reports that are typed on paper and therefore there may be some inaccuracies within the data. Address fields are only provided to the nearest hundred block in order to maintain privacy. This data is as accurate as the data in the database.

Who are the data providers? What do their organizations do? Why did they create these datasets?

LA Metro and LA city websites record and visualize the raw data to keep note of the crimes, and metro riders and keep track of the finances of various ticket pricing as well as how they affect revenue each year.

How was the data collected?

This data was collected through the sites to make a large CSV after cleaning and organizing the columns of the main required fields and to make the correlation between them.

What are the cases, i.e., the rows of the datasets?

The cases we considered are the ridership, finances, metro ridership, crime rate, and how the gasoline prices affected the ridership past 3 years from 2020 - 2022.

What variables will you most likely use as your target variables? What do these variables measure, and how are they quantified?

Ridership - The number of people riding the Metro rail over the course of the year

Crime rate near LA Metro - The rate of different types of crimes near the metro lines which might affect the ridership the most.

What variables will you consider using as predictors? What do these variables measure, and how are they quantified?

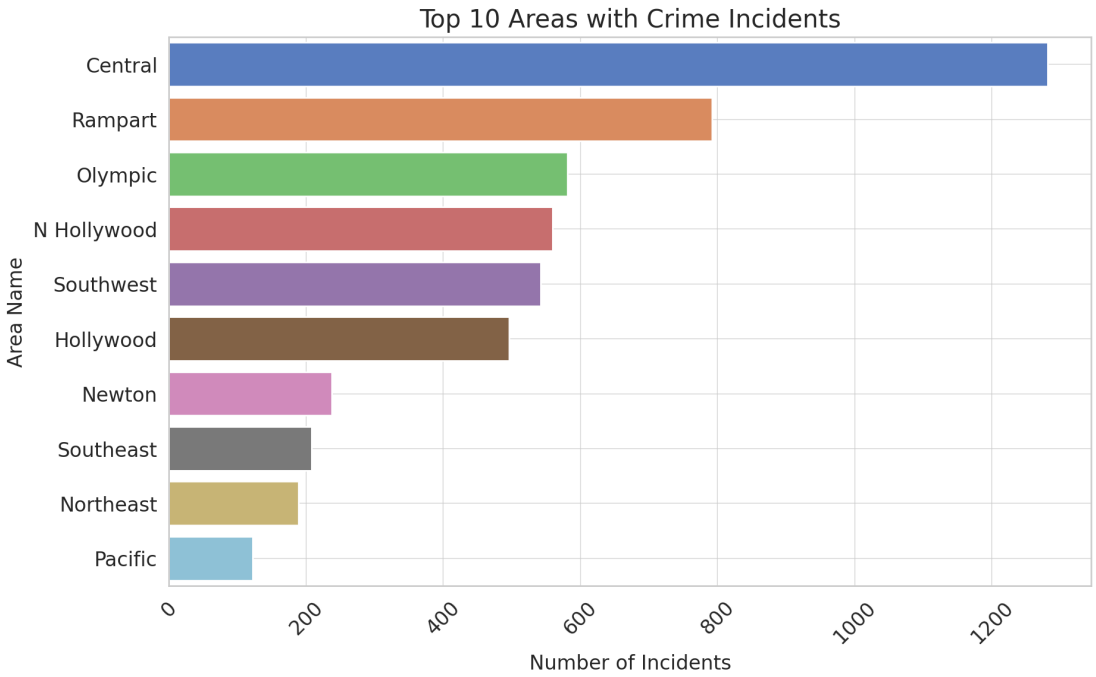
- Crime rate - Amount of different crimes in the area
- Average gasoline prices - The price of gasoline in the LA county
- Hours lost driving due to freeway traffic - Amount of time spent in traffic
- Dollars lost due to traffic - Opportunity cost lost in traffic

What data cleaning or manipulation might be necessary to prepare this data for modeling?

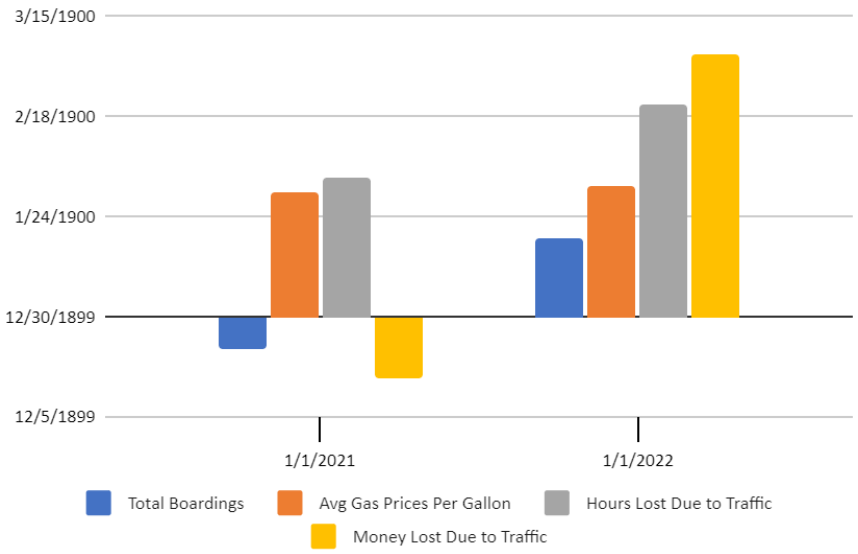
Cleaning crime data based on the area near metro rails. This will help us understand how the crime data is correlated with the ridership of metro rail.
Cleaning the ridership data to show based on the year-to-year correlation between crime rate, ticket prices, and gasoline prices changed the yearly revenue of the LA metro. Manipulating the columns to merge the crime rate and ridership data to have a correlation based on the areas of interest for LA metro lines.

Total Number of Crimes Reported Each Month

Here's the individual bar plot showing the top 10 areas with the most crime incidents. The "Central" area has the highest number of reported incidents, followed by "Rampart" and "Olympic".



There was a definite change in total boardings from 2021 to 2022, while hours and money lost due to traffic and gas prices all significantly increased.



Date	Total Boardings	Avg $/GL	Hours Lost	Money Lost
1/1/2020	51918949	$3.04	46	$1,142
1/1/2021	47866883	$3.98	62	$968
1/1/2022	57300995	$5.28	95	$1,601

EDA Analysis

	Ridership	Crime
Minimum	4461122	974
Maximum	57717922	1386
Mean	17567790.11	1184.67
Median	9260318	1194
Mode	4461122	974
1st Quartile	6171352.25	974
3rd Quartile	22585121.5	1386
Standard Deviation	17468178.77	173.21
Variance	3.05137E+14	30000.94

Research Questions

What major questions do you expect to be able to answer with this data?

With this data, we aim to understand why some LA residents avoid the Metro. Concerns might include service reliability, cleanliness, and safety. We'll compare the Metro's appeal against other transport options in terms of cost and convenience. We'll also explore if certain areas or demographics feel underserved by the Metro. Analyzing peak hours and station popularity will further clarify usage patterns.

How will answers to those questions help the client make actionable decisions?

Using the data, SMART can make important changes to the LA Metro Rail. If people complain about the trains being dirty, unsafe, or late, the Metro can work on fixing these issues. If some areas or groups of people need better access to the Metro, new routes can be added or changed. If tickets are too expensive, the Metro can think about offering discounts or changing prices. If people don't know much about the Metro, more advertising can help. Lastly, talking to communities and understanding their needs can make the Metro better for everyone.

How will you validate that your answers to the questions are correct?

First, we'll compare our results with historical data, ensuring that our conclusions align with past trends and patterns. Additionally, we can further validate our conclusions by cross-referencing our findings with other studies or data sources on the same topic. Using social media to gather insights about rider sentiment will provide first-hand feedback. To make our results more credible, we'll utilize statistical tests, ensuring that our conclusions are significant and not merely the result of random variations.

Ethics

What individuals or groups could be helped by your analysis?

Our analysis seeks to benefit various groups in Los Angeles. Regular Metro riders can expect improved services, while non-users might be drawn to a better Metro experience, potentially easing road traffic. LA Metro management will gain insights for smarter decision-making and service enhancement. Local businesses near stations may see increased foot traffic. Importantly, a boosted public transport system can lead to fewer cars on roads and cleaner air for all.

What individuals or groups could be harmed by your analysis?

While aiming to enhance the Metro rail, our analysis may pose challenges for some. Transport alternatives like buses, taxis, and ride-shares might see fewer users. Residents near Metro stations

could face overcrowding and noise, especially at busy times. Additionally, improper data handling could risk privacy breaches.

Are there any natural biases you see in the data collection process?

One potential bias is selection bias, where the data sources available might not represent the entire spectrum of the LA Metro's operations or ridership. There's also the risk of historical bias, as past data might have been influenced by events or conditions that are no longer relevant. Aggregation bias can occur if data is overly generalized, potentially missing out on nuances or specific trends within subsets of the data. Lastly, confirmation bias remains a concern; even when working with existing datasets, we tend to seek out and interpret data in ways that confirm our pre-existing beliefs or hypotheses, which can unintentionally influence our analysis.

What information in your dataset, if any, should be protected for privacy reasons?

Personal identifiers, such as names, addresses, phone numbers, and other directly identifiable details, should either be anonymized or removed to protect individuals' identities. While opinions and feedback are central to our analysis, they should be presented in a way that they cannot be linked back to specific individuals, ensuring confidentiality.

What will you do to ensure your analysis is used for ethical purposes?

To ensure our analysis is used ethically, we'll be transparent about our goals and how we interpret the data. We'll use data only from reputable sources that have been collected responsibly. While analyzing, we'll avoid making assumptions that could lead to biased results. We'll respect the intentions of the original data collectors, like LA Metro, and ensure our findings align with the academic standards of our school. Any insights or recommendations will be made with the public's best interest in mind, and we'll always be open to feedback to ensure our work remains accountable and ethical.

References

https://transitmap.net/los-angeles-metro-k-line-2022/

https://www.dailynews.com/2023/06/16/downtown-regional-connector-train-opens-after-10-years-and-1-8-billion/

https://www.foxla.com/news/los-angeles-drivers-commute-traffic-hours-worst-corridor

https://www.laalmanac.com/energy/en12.php

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

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

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

https://citytransit.uitp.org/los-angeles/average-daily-ridership-during-a-work-day/suburban-rail