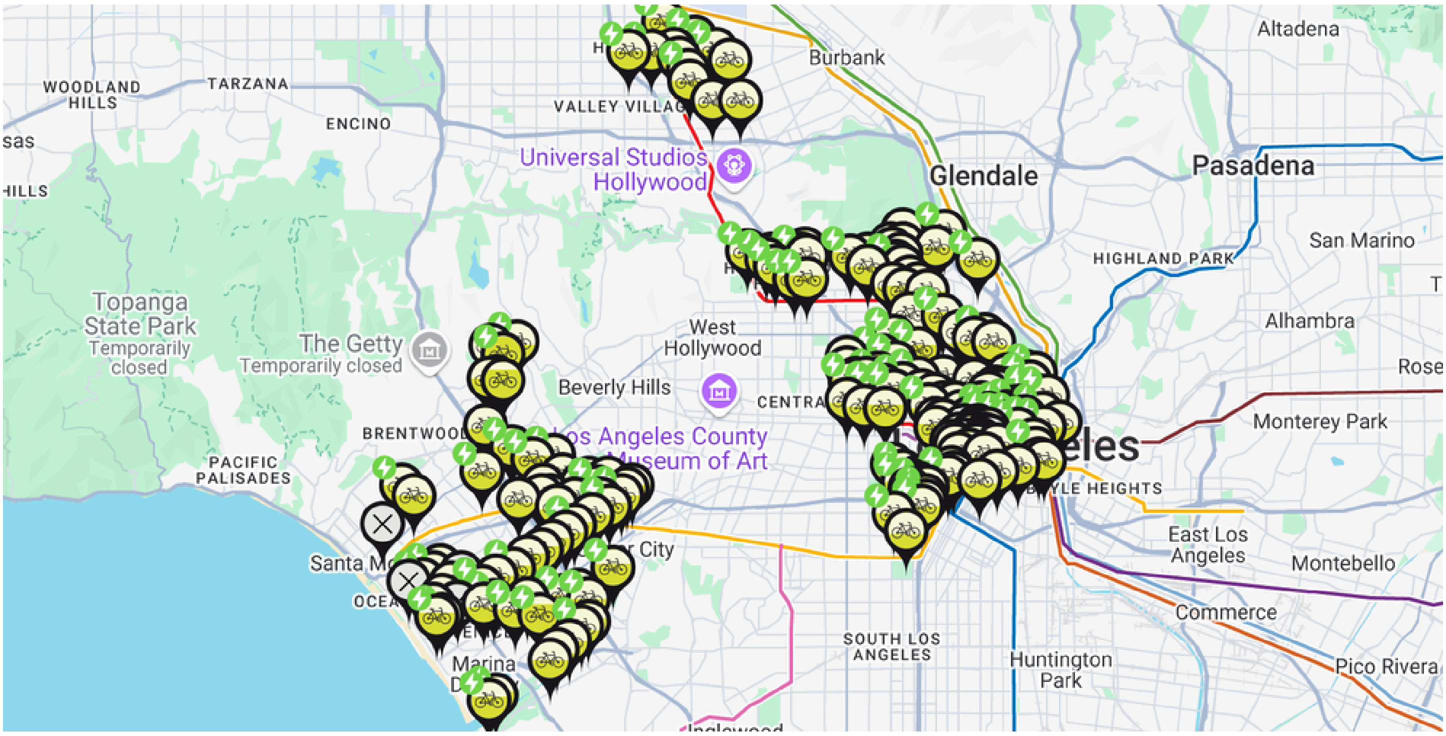
Capstone 2 Project Proposal

LA Metro Station Relocation

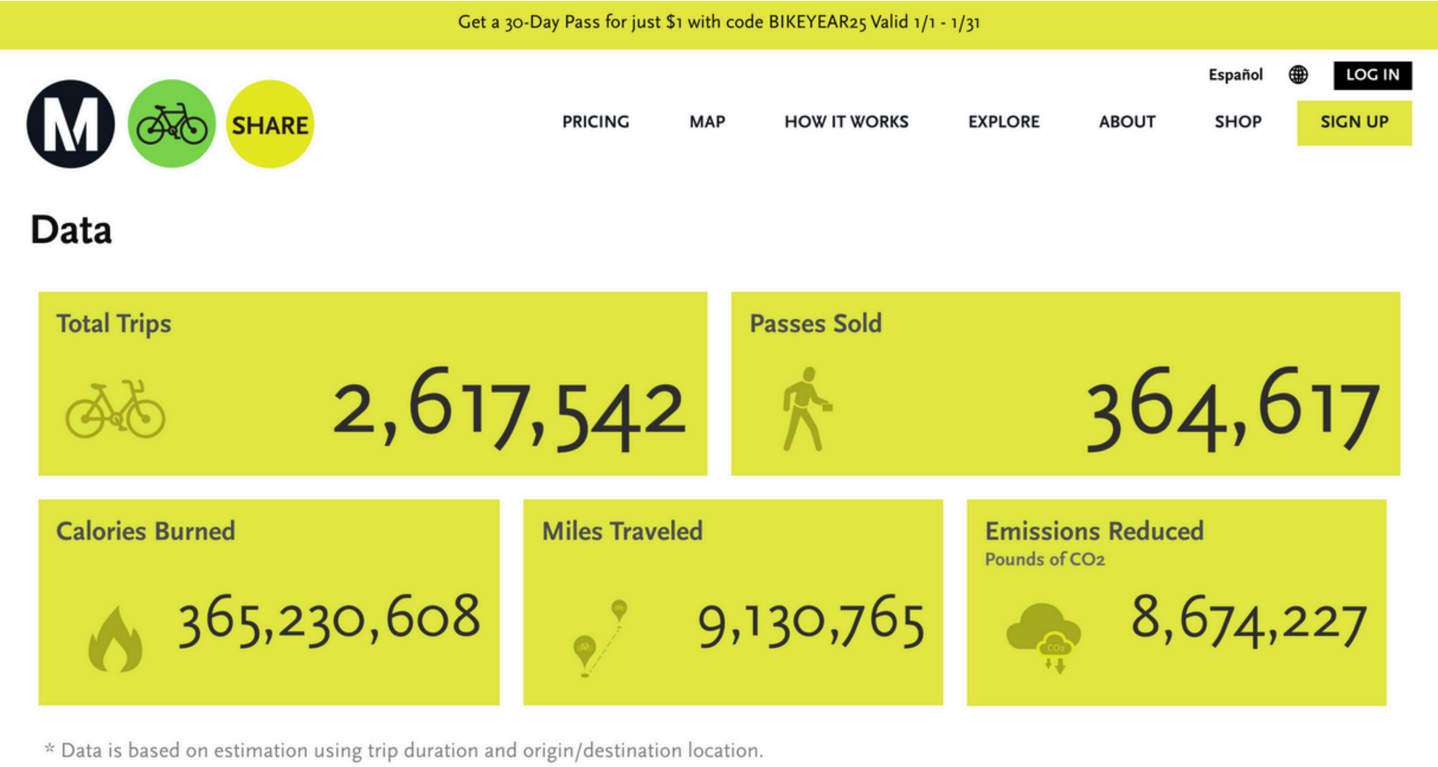
Started in 2016, the Los Angeles Metro bikes have been an essential part for the public transportation and infrastructure when it comes to bicycles. Over the last 8 years LA Metro has introduced over 220 Metro Bike Share stations across LA County. You can find Metro Bikes in each of these service areas: Downtown LA, Central LA, Hollywood, North Hollywood, the Westside and Westwood. **The goal is to analyze station performance using regression models and deep learning techniques to identify underutilized stations and provide actionable recommendations for relocations to optimize usage and accessibility.**



Main Objectives:

- **Station Performance Analysis**: Evaluate the performance of bike stations based on key metrics like trip frequency, duration, and station utilization to identify underperforming locations.
- **Demand Prediction and Optimization**: Use regression and deep learning models to predict demand across regions and pinpoint high-demand areas where station relocations or expansions can increase ridership and system efficiency.
- **Geospatial and User Behavior Insights**: Integrate geographic data (e.g., Bus and Tain station locations, points of interest) with user behavior patterns (e.g., passholder types, trip durations) to provide data-driven recommendations for improving station placements and enhancing overall accessibility.
- **Weather Analysis:** Discover underlying weather trends, analyze quarterly performance of stations based on historical weather data to understand the Trip volume and external factors that might affect station performance.

The Data



The Data for this project is real world data, publicly available and is collected from :

The official LA Metro Website (Trip and Station information)

https://bikeshare.metro.net/about/data/

Additional Data Sources

https://data.lacounty.gov/maps/lacounty::la-county-points-of-interest/about

A summary of Schools, Hospitals, Art facilities and Attractions around Los Angeles

https://data-lahub.opendata.arcgis.com/datasets/6679d1ccc3744a7f87f7855e7ce33395_1/explore

A visual map of the Metro Rail Line Stops

https://developer.metro.net/gis-data/

Geospatial Data for Metro Rail and Bus Stations

https://open-meteo.com/en/docs/historical-forecast-api#start_date=2025-01-12

A Weather API to extract historical weather information

Below are the columns of the stations and trips .csv files:

trip_id	• Locally unique integer that identifies the trip
duration	• Length of trip in minutes
start_time	• Date/time when trip began, ISO 8601 format local time
end_time	• Date/time when trip ended, ISO 8601 format local time
start_station	• Station ID where trip originated
end_station	• Station ID where trip terminated
bike_id	• Locally unique integer that identifies the bike
plan_duration	• Number of days that plan the passholder is using
trip_route_category	• Round-Trip or One-Way
passholder_type	• Plan name
bike_type	• Standard, electric assist or smart bikes
start_lon/start_lat	• coordinates of starting station
end_lon/end_lat	• coordinates of ending station

Enhanced Data Improvements

1. Streamlining the process of fetching, organizing, and **analyzing weather data from the API** by automating API calls, aggregating data for multiple locations, and structuring it into a unified format."
2. The **integration of bus and metro train station data** into the analysis to enhance insights by identifying patterns of accessibility and proximity. By automating the process of combining geospatial data for public transit stops with existing rideshare and weather data, I aim to uncover relationships between transit infrastructure and station performance. This will provide a more comprehensive understanding of factors influencing ridership and optimize station relocations.

3. *The **relationship between station performance and nearby attractions**. By automating the process of combining geospatial data for POIs such as museums, parks, and commercial hubs with existing rideshare and weather data, I aim to identify patterns of demand influenced by proximity to key locations. This will provide a deeper understanding of ridership behavior and support data-driven recommendations for station relocations.*

Data Management

Data	Key Info	Use
Trip Data (pass holders, bike types, duration)	☐ All Stations and Trip Data ☐ Customer demographics ☐ Includes Bike types	• Analyze Trip Behaviour • Understand individual Station performance
Bike Station Data (Station Name, Station ID, Geospatial info)	☐ Name of Station with unique station ID ☐ Coordinates of each Station	• Visualize Stations on a map • Interpret Stations by Trip Volume
Bus and Train Station Data (all LA Metro Bus and Train stations)	☐ Includes map view ☐ Coordinates of all Stations ☐ Bus and Train Stops	• Compute distance from Bike stations to Train e.g. • Understand station behavior based on other public transportation
Points of Interest (Dataset with location of Attractions, points of interest and more)	☐ Detailed information ☐ 7 separate data sets ☐ Coordinates available	• Understand how attractions or other public points of interest contribute to trip volume of stations
Historical Weather Data (using an API)	☐ Information about temperature, wind and perception	• Enhance quarter analysis by adding weather data

The Machine Learning Approach

Machine learning is an ideal approach for analyzing the performance of Metro bike stations due to its ability to uncover complex relationships within the data. The performance of these stations is influenced by numerous interconnected factors, such as trip frequency, user demographics, station proximity to points of interest, and weather conditions. These relationships are often non-linear and challenging to model manually. Additionally, the dataset encompasses a large scale of information, including data from numerous stations, thousands of trips, and additional features like weather and geospatial information. Machine learning efficiently processes such large datasets, enabling the discovery of hidden patterns that inform actionable insights. By leveraging historical data, machine learning models can predict underperforming stations and simulate the impact of their relocation, offering a data-driven approach to optimization. Moreover, the outputs from these models provide stakeholders with concrete recommendations, facilitating better decision-making and enhancing the overall efficiency of the bike-sharing system.

Tools and Technology

Tool	Purpose	Expected Insights
Linear Regression	Establishes a baseline and helps understand the relationship between station performance and key features.	Provides interpretable coefficients to identify the most influential factors affecting station utilization, guiding feature engineering for advanced models.
Random Forest	Captures non-linear relationships and handles feature interactions effectively.	Offers feature importance rankings and accurate predictions while providing a robust benchmark for evaluating more complex models.
Feedforward Neural Network (FNN)	Identifies complex, non-linear relationships between features using fully connected layers.	Optimizes predictive performance by capturing intricate dependencies across input features, such as the interaction between weather, POIs, and trip frequency, for robust recommendations.
Convolutional Neural Network (CNN)	Extracts spatial features and relationships from geospatial data (e.g., station proximity to POIs or transit stops).	Reveals how spatial configurations, such as the clustering of bike stations near busy hubs, affect usage, enabling data-driven insights into the optimal placement of underutilized stations.

Timeline

Week 1	Week 2	Week 3	Week 4
Data Wrangling	Data Wrangling		
	EDA and Feature Engineering	EDA and Feature Engineering	
		Model Training and Development	Model Training and Development