Final Report

Impact of Fuel Prices and Metro Fares on Public Transport Usage in Istanbul

1. Motivation

As a resident of Istanbul and frequent user of the city's metro system, I have long been interested in understanding the external factors that influence public transport usage. During my time living in the Netherlands, I observed how fuel price fluctuations were sometimes met with adjustments in public transportation fares to influence commuter behavior. This sparked the idea to explore whether similar patterns could be observed in Istanbul. Through this project, I aim to combine personal interest with practical data science skills to evaluate the impact of fuel prices, metro fares, and vehicle sales on metro ridership trends in Istanbul.

2. Project Overview

This project analyzes the relationship between fuel prices, metro fares, vehicle sales, and public transport usage in Istanbul over a period from 2013 to 2024. The objective is to determine the key factors influencing annual metro ridership and to build predictive models capable of estimating future usage under varying economic conditions. Findings from this study are expected to provide valuable insights for policymakers and urban transport planners.

3. Hypotheses

The project tests the following hypotheses:

• H01: Fuel prices do not significantly impact metro ridership.

• H11: Increases in fuel prices lead to higher metro ridership.

• H02: Metro fare changes do not significantly affect metro ridership.

• H12: Metro fare increases lead to lower metro ridership.

• H03: Vehicle sales do not have a significant impact on metro ridership.

• H13: Increases in vehicle sales lead to a decline in metro ridership as more people opt for private transportation.

4. Data Sources

The data used in this project was obtained from the following sources:

• **Metro Ridership:** Yearly metro passenger statistics from Metro Istanbul and the Istanbul Metropolitan Municipality (IBB).

• **Fuel Prices:** Annual gasoline and diesel prices sourced from Aytemiz and TPPD.

• **Metro Fares**: Yearly fare data for full, student, and discounted categories from TUHIM (tuhim.ibb.gov.tr).

• **Vehicle Sales:** Annual automobile sales statistics from the Automotive Distributors Association (ODMD).

• **Population:** Istanbul population estimates from the Turkish Statistical Institute (TURKSTAT).

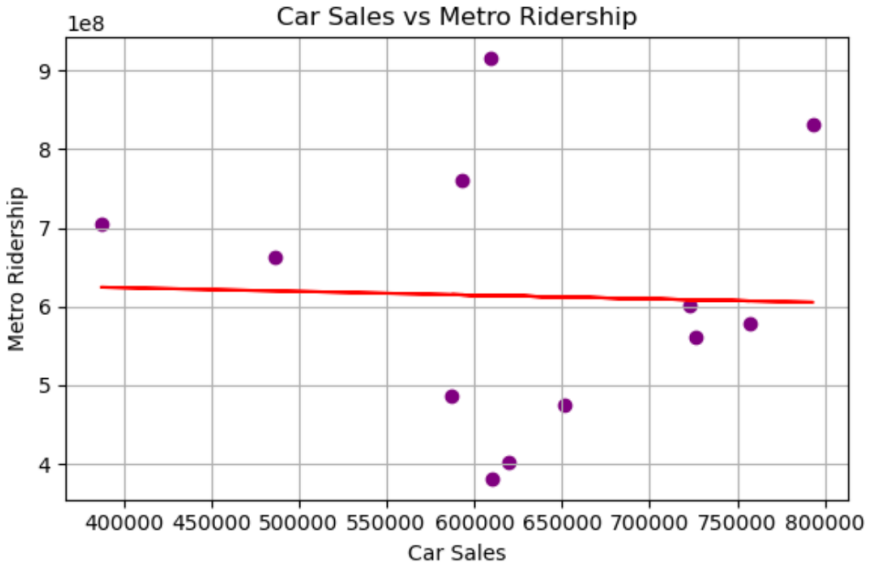
5. Exploratory Data Analysis

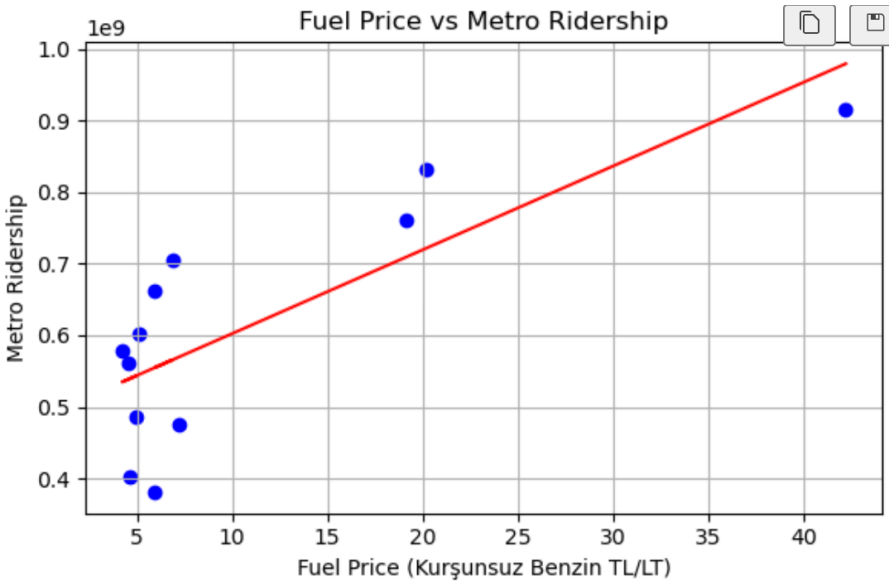
The datasets were cleaned, standardized, and merged by year. Where monthly data was available, it was aggregated into annual averages. Missing values, such as certain fare prices for early years, were imputed conservatively using the previous year's data. Exploratory visualizations revealed the following trends:

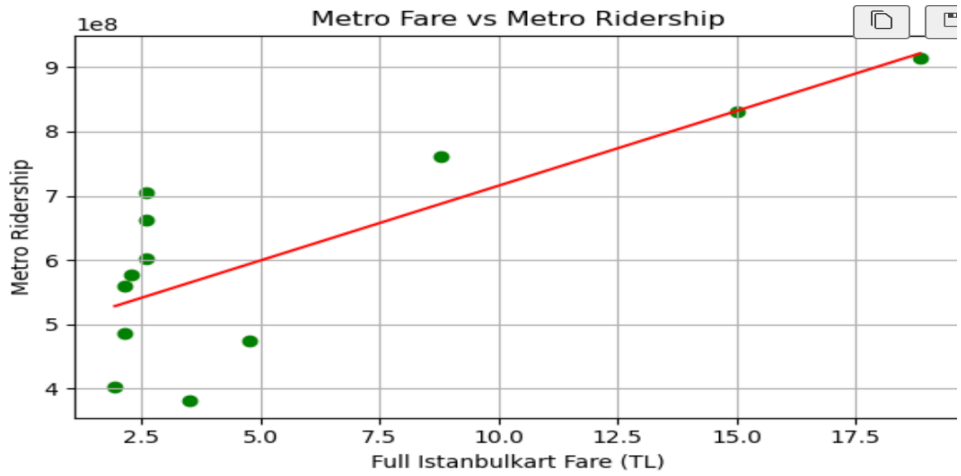
• A strong positive correlation between fuel prices and metro ridership.

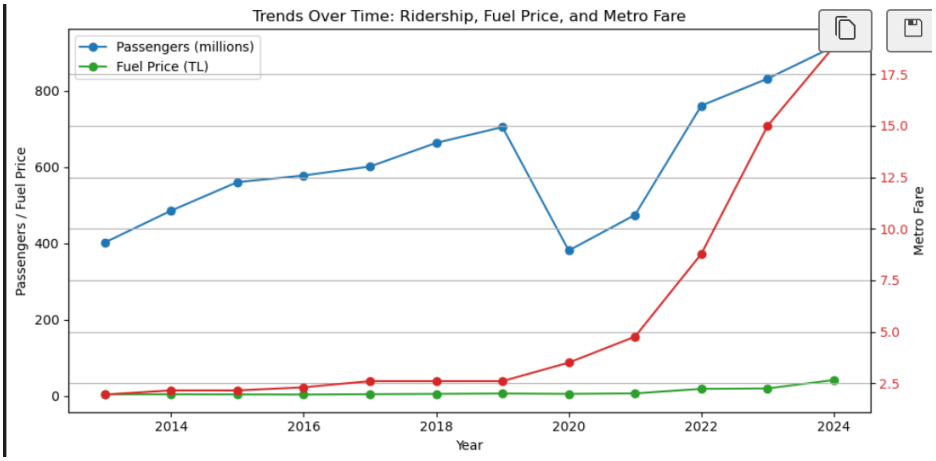
• Metro fare increases did not reduce ridership significantly, possibly due to inflation and lack of alternatives.

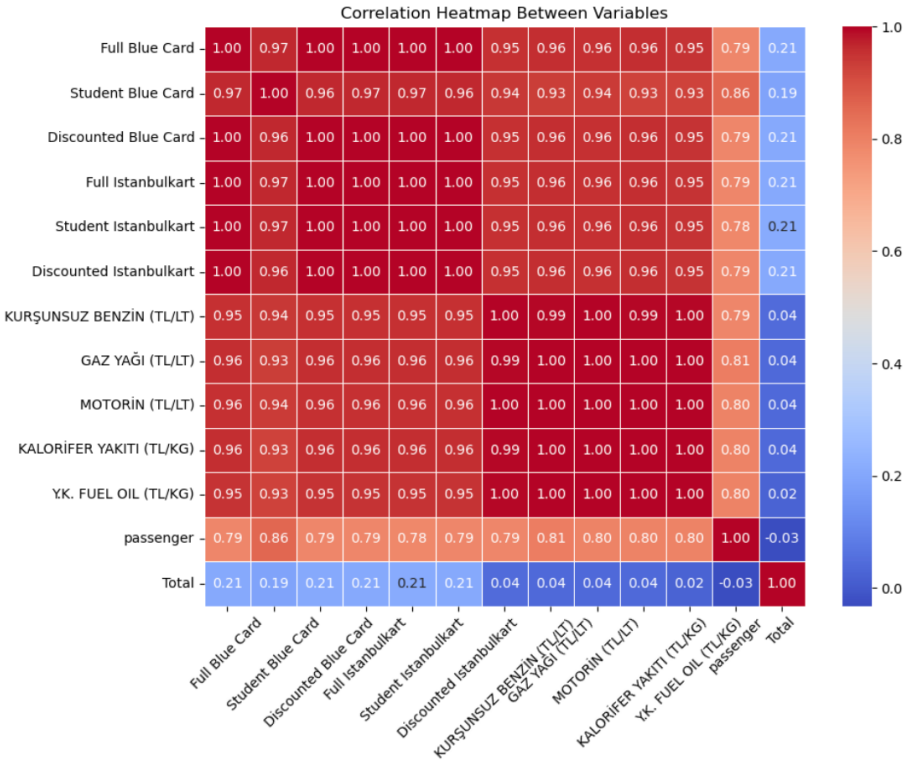
• Vehicle sales showed little to no relationship with public transport usage.




Fuel Price vs Metro Ridership
1e9
Metro Ridership
Fuel Price (Kurşunsuz Benzin TL/LT)


Metro Fare vs Metro Ridership
1e8
Metro Ridership
Full Istanbulkart Fare (TL)


Trends Over Time: Ridership, Fuel Price, and Metro Fare
Passengers (millions)
Fuel Price (TL)
Passengers / Fuel Price
Metro Fare
Year



6. Hypothesis Testing

Statistical tests were conducted using Pearson correlation and linear regression:

• Fuel Price vs. Metro Ridership:

Pearson r = 0.789

p-value = 0.0023

=> Significant positive relationship.

• Metro Fare vs. Metro Ridership:

Pearson r = 0.788

p-value = 0.0023

=> Significant, but unexpectedly positive.

• Vehicle Sales vs. Metro Ridership:

Pearson r = -0.032

p-value = 0.9206

=> No significant relationship.

7. Machine Learning Models

To further analyze the influence of these variables and predict future trends, two machine learning models were applied:

• Random Forest Regressor:

R^2 Score = 0.82

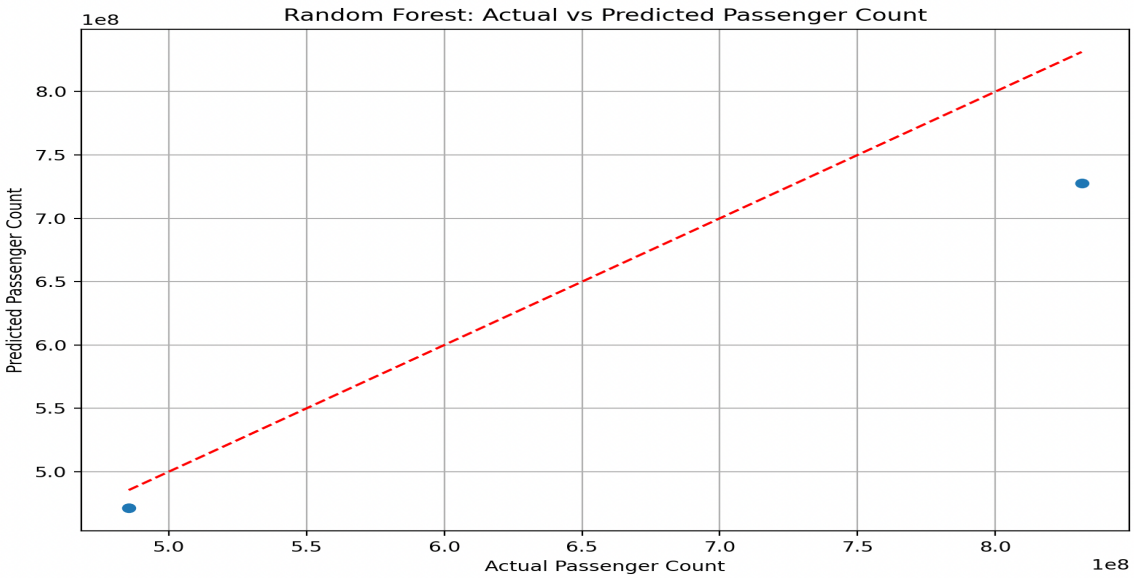
Mean Squared Error = 5.45×10^{15}

• XGBoost Regressor:

R^2 Score = 0.80

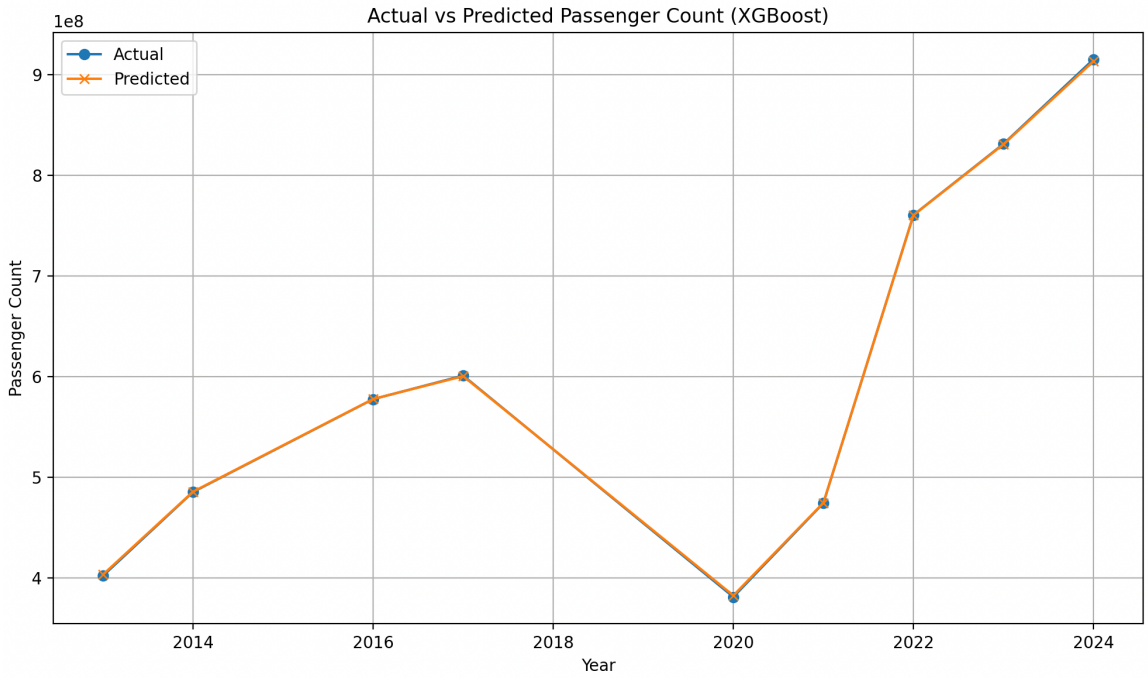
Mean Squared Error = 5.91×10^{15}

=> Both models identified average fare as the most impactful predictor of metro ridership. Fuel prices were secondary. Vehicle sales and population were not found to be influential within the model.



8. Forecasting and Scenario Testing

Using the XGBoost model, metro ridership was forecasted for the year 2025 under various economic scenarios. All scenarios, including higher fare, lower fuel price, and increased population, produced similar ridership predictions (around 474.8 million), indicating the model is primarily responsive to fare inputs.



9. Findings

• Fare pricing is the most effective lever for influencing public transport usage.

• Fuel prices impact ridership, but not as strongly as fare prices.

• Vehicle sales and population changes have minimal predictive power.

10. Limitations

• The dataset is limited to annual data, reducing granularity.

• Behavioral and socioeconomic factors were not included.

• External shocks (e.g. COVID-19) may introduce anomalies in trends.

11. Future Work

• Incorporating monthly or weekly ridership data.

• Including additional predictors like weather, economic indicators, and transport network expansions.

• Applying clustering or classification models to segment rider types.

12. Conclusion

This project demonstrates that metro fare levels are the most significant factor affecting ridership in Istanbul. While fuel prices do play a role, fare policy remains the most direct and influential tool for managing public transport demand. The findings from both data analysis and machine learning models can support evidence-based decisions in transportation planning.