Identifying Congestion Patterns in the NYC MTA

By: Hyunjin Kim

NYC MTA

Objective: To identify factors contributing to congestion and develop strategies to mitigate the issue.

Data Sources

Four sources were used:

Historical MetroCard History data

US Holidays

Historical Gasoline Prices

Historical Weather data

Key Columns:

From Date - To Date

Station

Total Fares for the Week

Contains Holiday

Gas Price

Total Snowfall/ Total Rainfall

Data Assumptions and Limitations

Handling Missing Values

- MetroCard History:
  - Null values replaced with 0
  - Assumption: No passengers used the specific fare type in the given time period

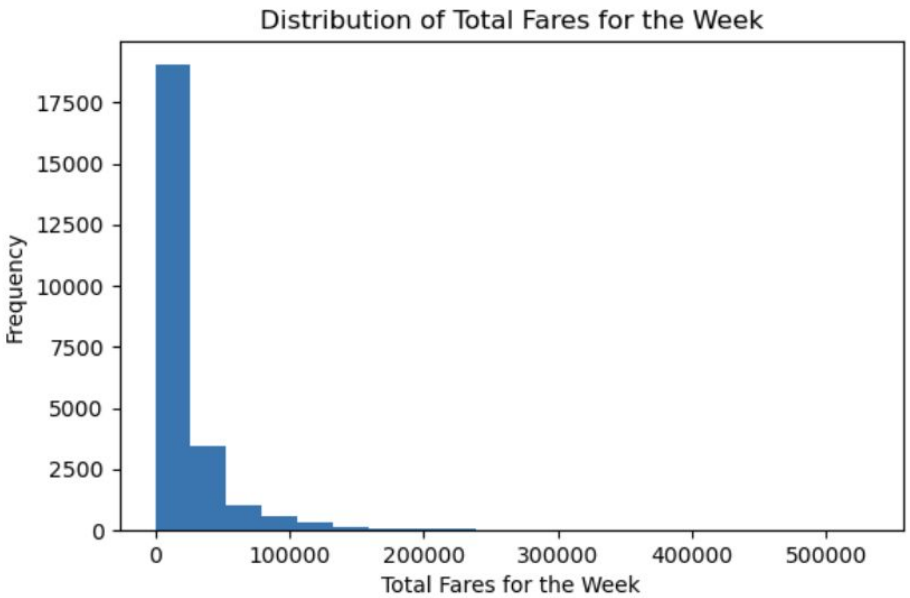
- Gas Price:
  - Forward fill used to impute missing values
  -Assumption: Gas prices remain constant until the next recorded price change.

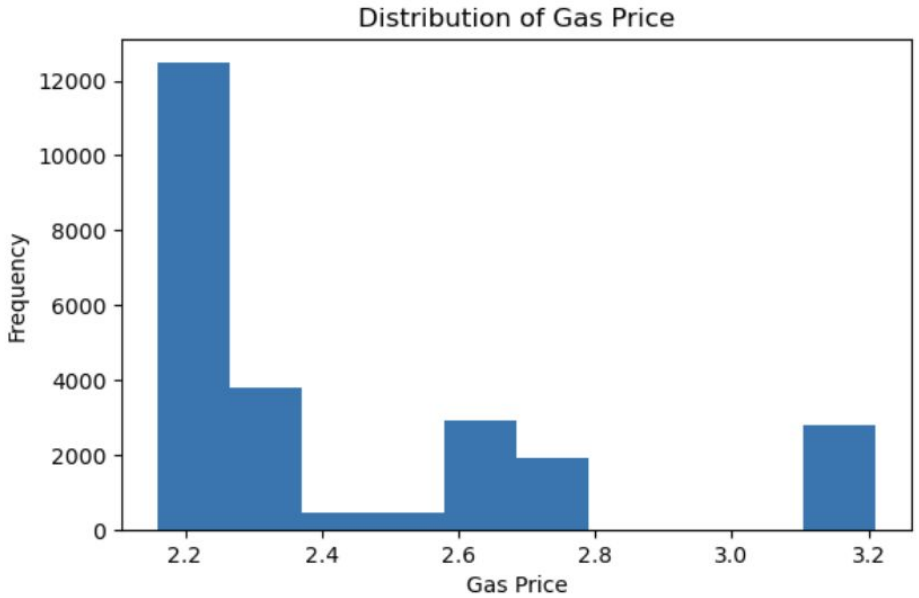
Handling Missing Values

- Weather Data:
  - Focus only on snowfall and rainfall
  - Assumption: Snow and rain have a more significant impact on ridership

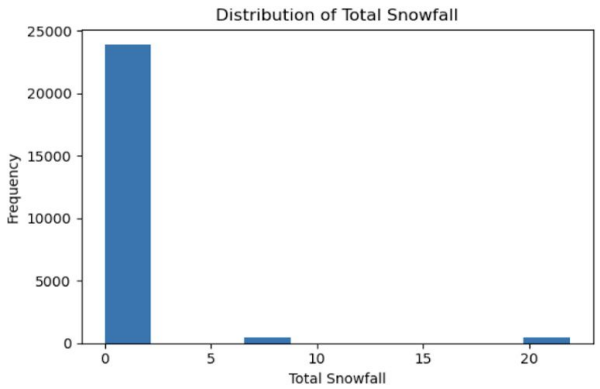
- Limitations
  - Assumptions simplify analysis but may not capture all nuances
  - Other weather factors or events may influence ridership patterns

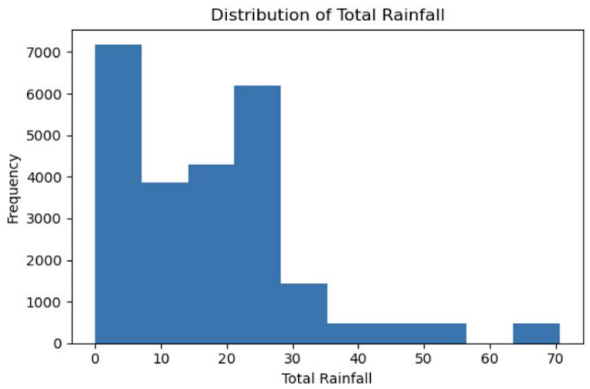
EDA - Part 1


Distribution of Total Fares for the Week
Frequency
17500
15000
12500
10000
7500
5000
2500
0
0
100000
200000
300000
400000
500000
Total Fares for the Week


Distribution of Gas Price
Frequency
12000
10000
8000
6000
4000
2000
0
2.2
2.4
2.6
2.8
3.0
3.2
Gas Price

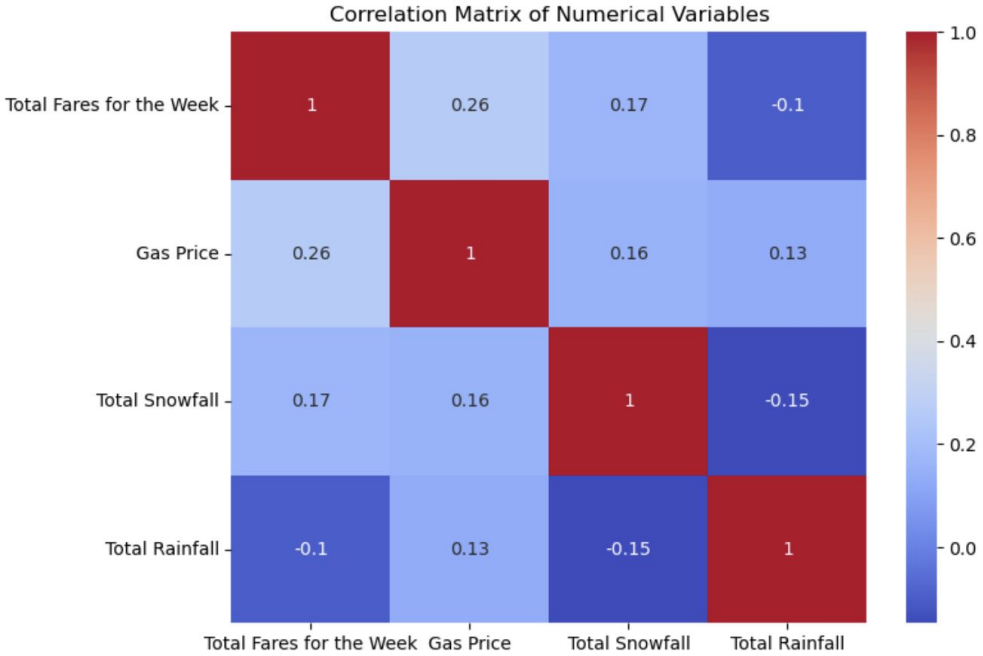
EDA - Part 2





- Most of the data is right skewed
- We will need to select a model that is less sensitive to distribution.

Correlation Matrix

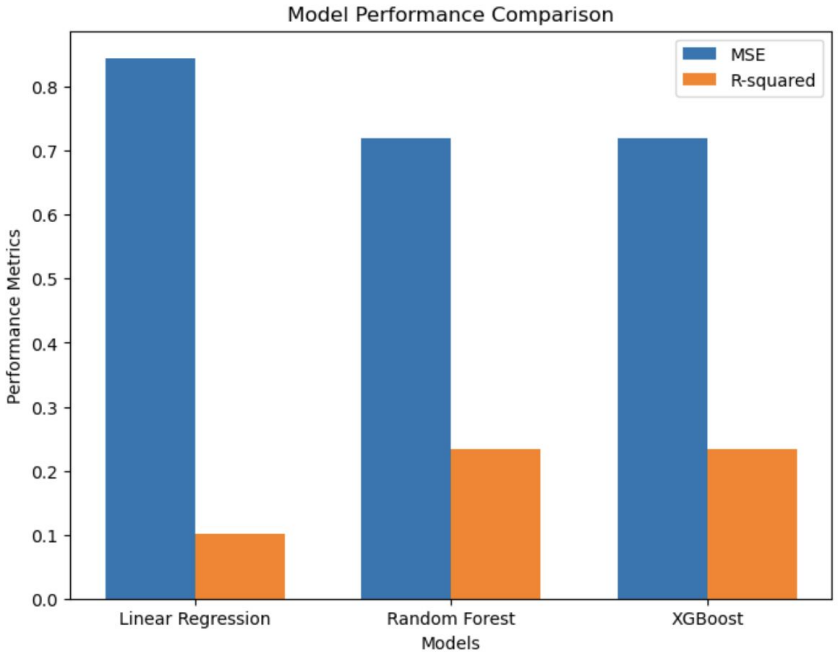


- There are no really strong correlations
- Rain seems to have a negative effect on ridership
- Gas price seems to have the largest positive effect on ridership

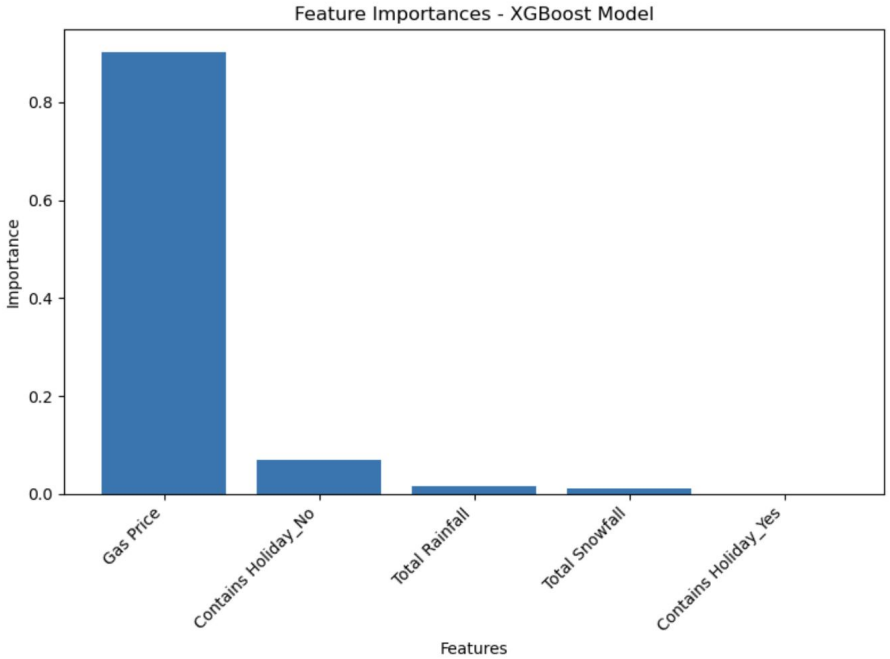
Model Development

MSE - Lower value means better accuracy

R-squared - How well does the model explain variation, numbers closer to 1 are better.



Feature Importance



- Higher gas prices are associated with increased MTA ridership and total fares
- Weeks without holidays tend to have lower total fares compared to weeks with holidays
- Weather factors have relatively lower importance

Key Findings

- Higher gas prices = increase in ridership
- Weeks with holidays experience higher ridership
- Weather factors have little to no impact in predicting congestion

Recommendations

- Implement dynamic pricing that adjusts fares based on gas prices
  - Encourage the use of public transport during times of higher gas prices

- Focus on managing congestion during times of gas price changes and holiday periods
  - Increase service frequency or implement holiday specific pricing to mitigate the impact of holidays

- Offer incentives for off-peak travel

Future Work

Short Term (1-3 months):

- Expand the feature set
  - Explore additional features that may impact congestion e.g. Economic indicators, population density, major events (e.g. sporting events, concerts)

Medium Term (3-6 months):

- Analyze time series patterns
  - Investigate daily, weekly, or seasonal variations in congestion
- Experiment with advanced machine learning models
  - Explore deep learning or ensemble methods to capture complex relationships

Long Term (6-12 months):

- Develop a congestion management dashboard
  - Create a user-friendly interface that displays real-time congestion levels
- Implement continuous monitoring and model updates
  - Regularly update the models with new data to capture changes in ridership patterns

Thank you!
Questions, comments

Hyunjin Kim:

example@email.com

917-000-0000