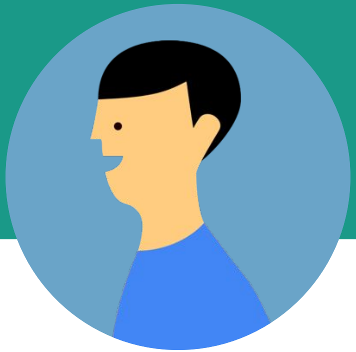
Dynamic Pricing for New York Subway Fares

Nghia Nim, Tengyu Song, Anik Dey, Cheryl Yan

ABOUT US!!!!



Tengyu Song

M.A. Statistics, 2nd Year

Anik Dey

B.S. Mathematics and Computer Science, Junior

Cheryl Yan

B.S. Operations Research, Junior

Nghia Nim

B.S. Computer Science and Mathematics, Junior

Problem Statement

Demand Management

Peak hours in the MTA system often see overcrowded trains, leading to uncomfortable travel conditions and increased wear on infrastructure. It can even lead to you missing your train!

Public Funding

There is always a need for public and social projects. We want to be able to guarantee revenue for these projects while not burdening residents with too much taxes.

Environmental and Traffic Benefits

By encouraging more people to use public transit during off-peak hours, the MTA can reduce the number of vehicles on the road, leading to less traffic congestion and decreased carbon emissions.

Benefits of our Solution

Balanced Ridership

The MTA can spread ridership more evenly throughout the day, reducing overcrowding during peak hours and underutilization during off-peak hours.

Affordability for Residents

Long-term passes (like 30-day passes) could be offered at discounted rates for residents, while short-term passes or single rides, often used by tourists, could be priced higher.

Potential for Tax Reduction

With increased revenue from dynamic pricing, from tourists and short timers, there might be less reliance on public funds to support the MTA. This could lead to potential tax reductions or allow for tax revenues to be reallocated to other essential public services.

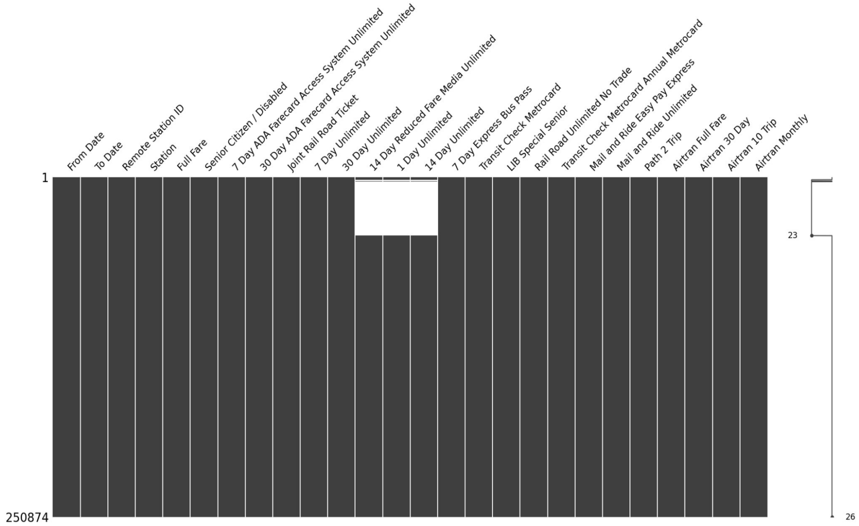
EDA

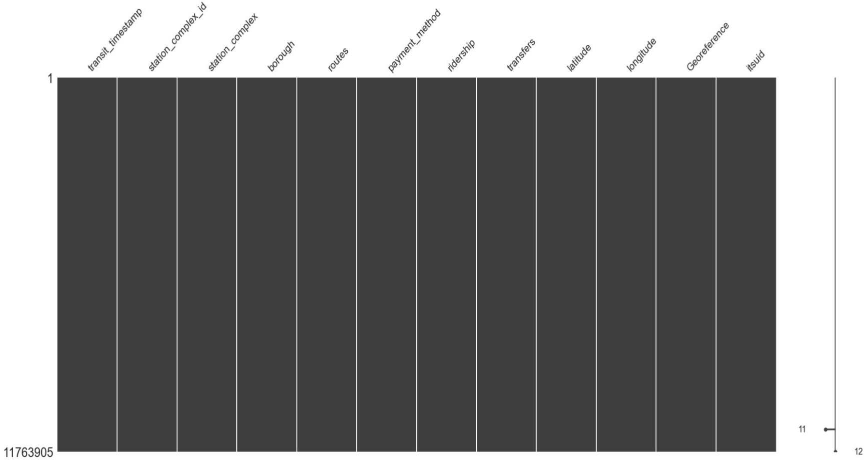
MTA Fare Card History - Beginning 2010, ends in 2021

Known problems:
Missing data, ‘To date’ and ‘From date’ flipping in some rows, Only a few weeks for 2021.

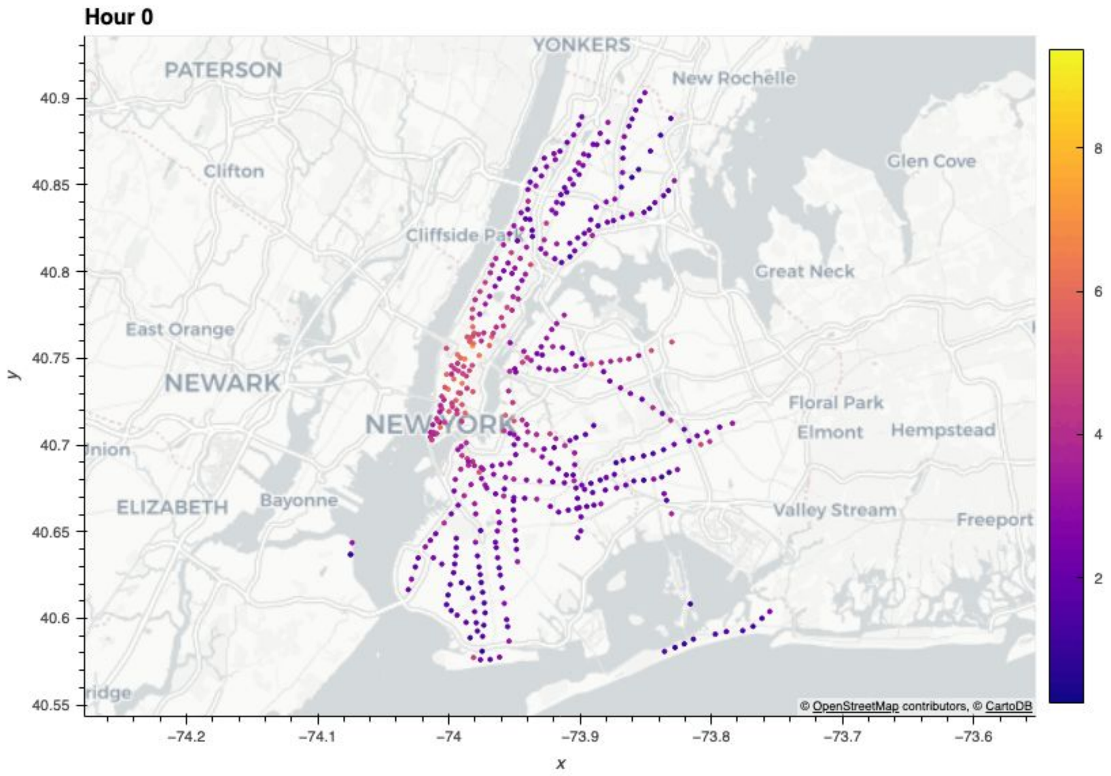
MTA Hourly Ridership - Beginning February 2022:

Known problems:
Some hour data for some station missing
E.g. No data for station ‘H022’ at 02/25/23 at 1 a.m.



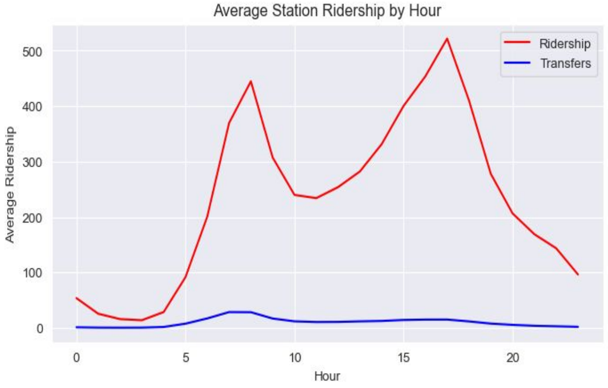


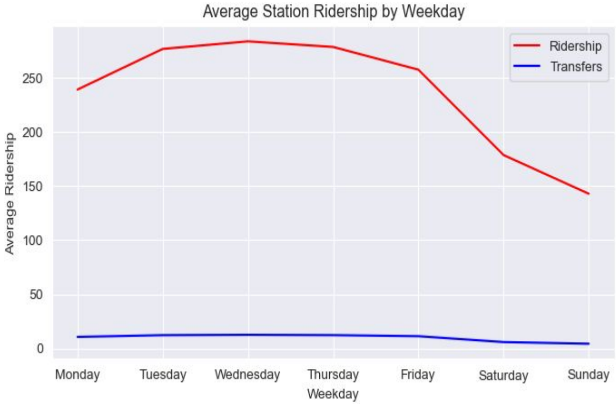
Average Ridership at Each NYC Station throughout Different Hours of the Day

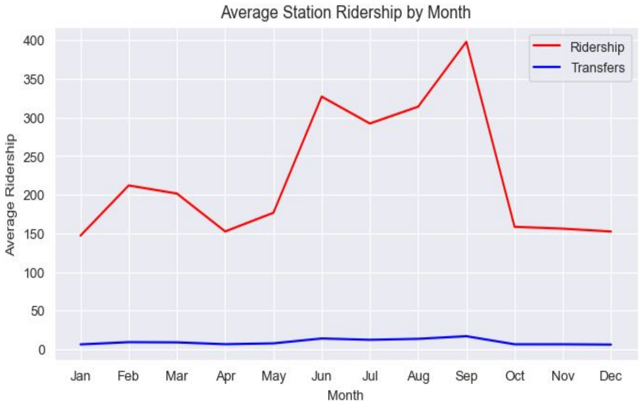


[Link to the interactive plot]

Peak & Off-peak Time







Peak hours: 7- 8 AM & 5-6 PM

Off-peak hours: before 5 AM, 11 AM - 12 PM & after 8 PM

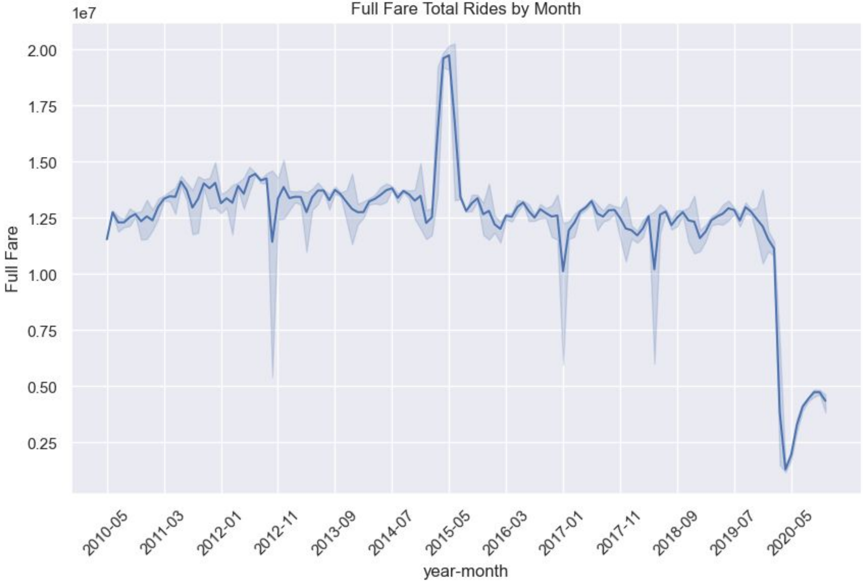
Peak day: Wednesday

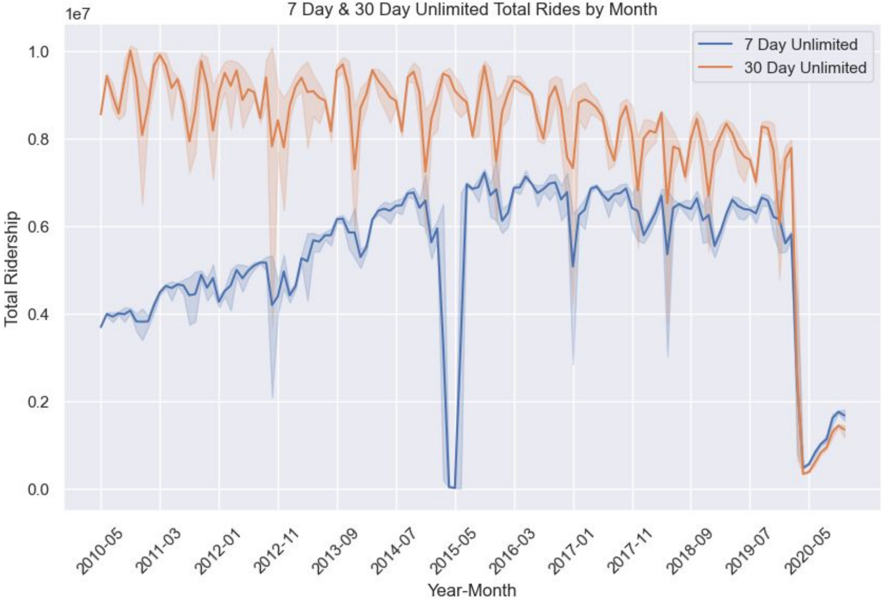
Off-peak day: Sunday

Peak month: Sep

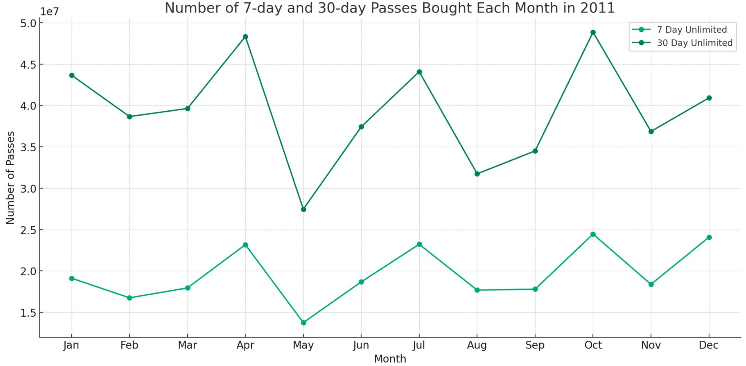
Off-peak months: Jan, Apr, Oct, Nov, Dec

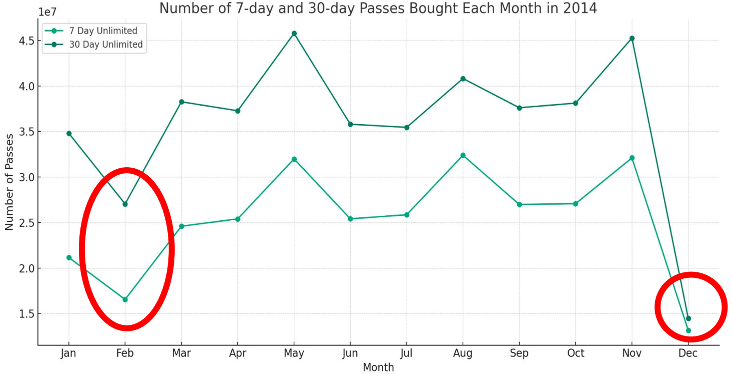
Full Fare & Unlimited Trend

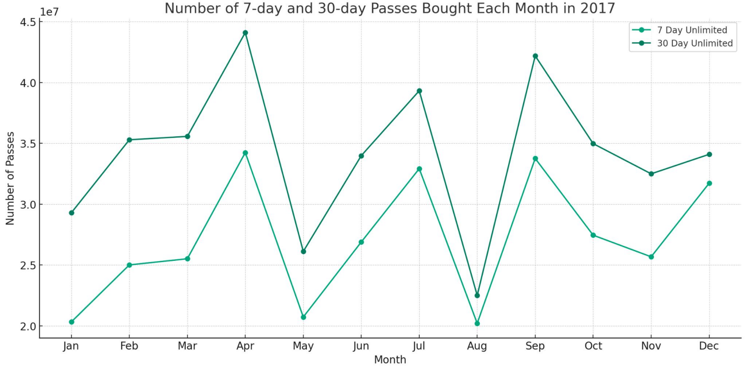


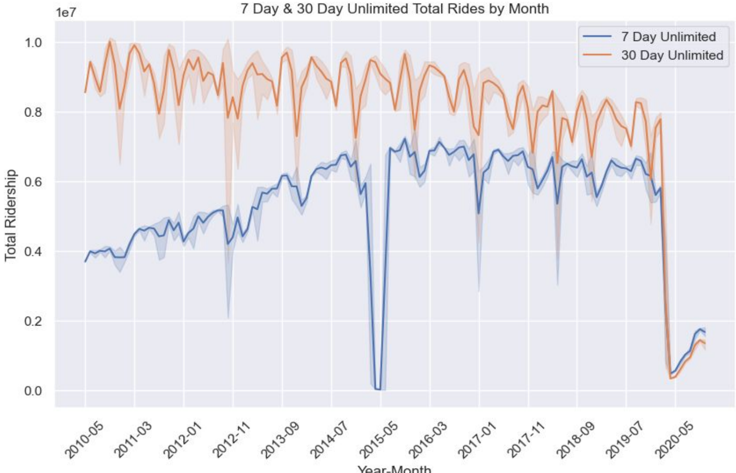


Strong Correlation between 7-day and 30-day Passes









Primitive Pricing Model

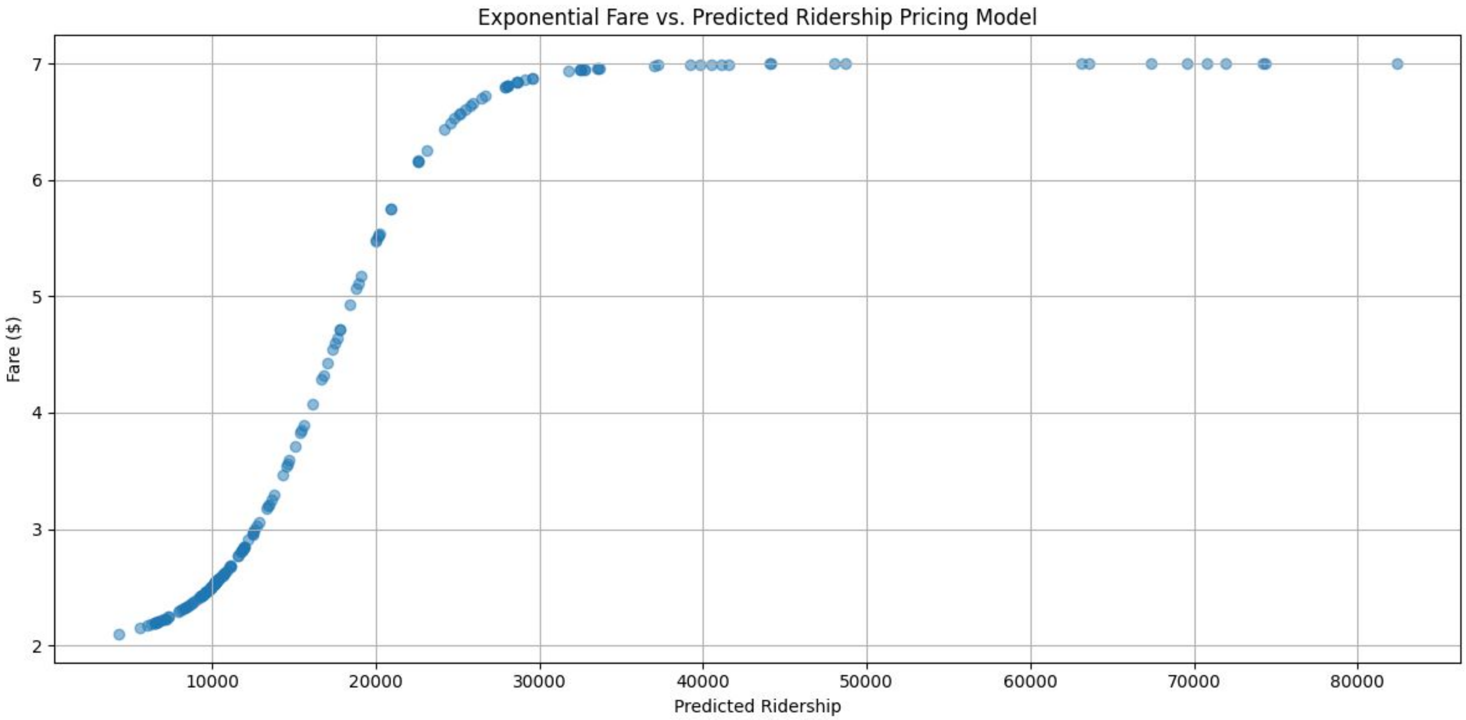
Modified Sigmoid-like Exponential Pricing Model

- Objective: Adjust fares dynamically based on predicted ridership, encouraging more ridership during off-peak times and managing high demand during peak times.
- We used Random Forest as a base model for predicting riderships which is then fed into our pricing model.

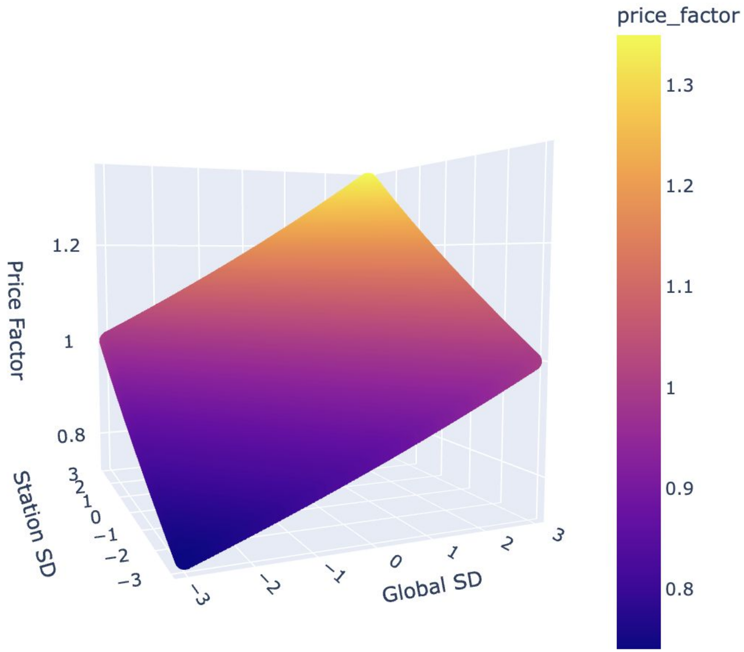
Data Processing

- Extracted day, month, and hour from the timestamp.
- One-hot encode the stations.
- Aggregated ridership data by day, month, hour, location.
- Calculated total ridership by summing different fare types.

Primitive Pricing Model



Hybrid Dynamic Pricing Model

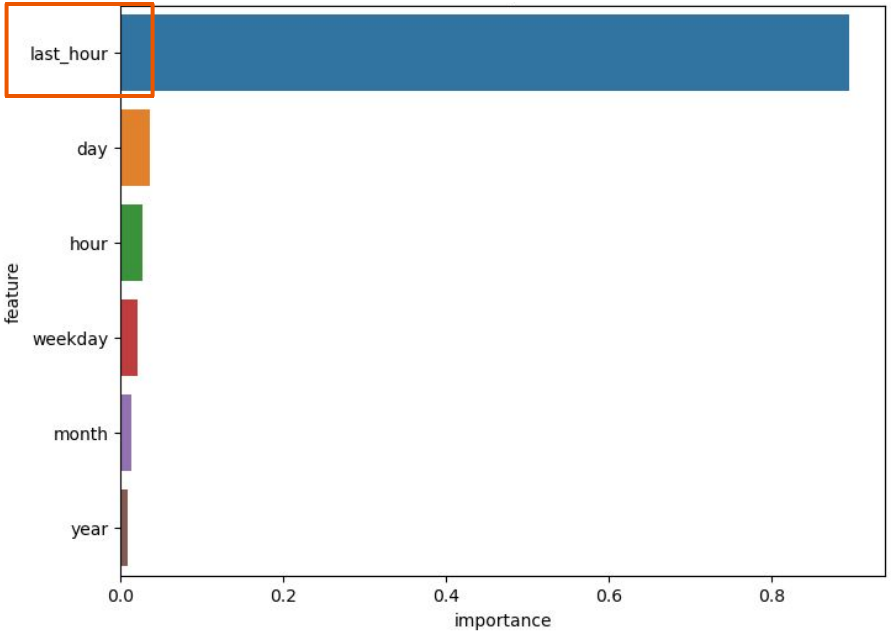


Key idea:

- If a station has higher average ridership than other stations, it will have a higher price factor
- For each station, the price factor at rush hour will be higher than normal hour.
- We can give a price factor for any station if we can **predict the demand for the next hour.**

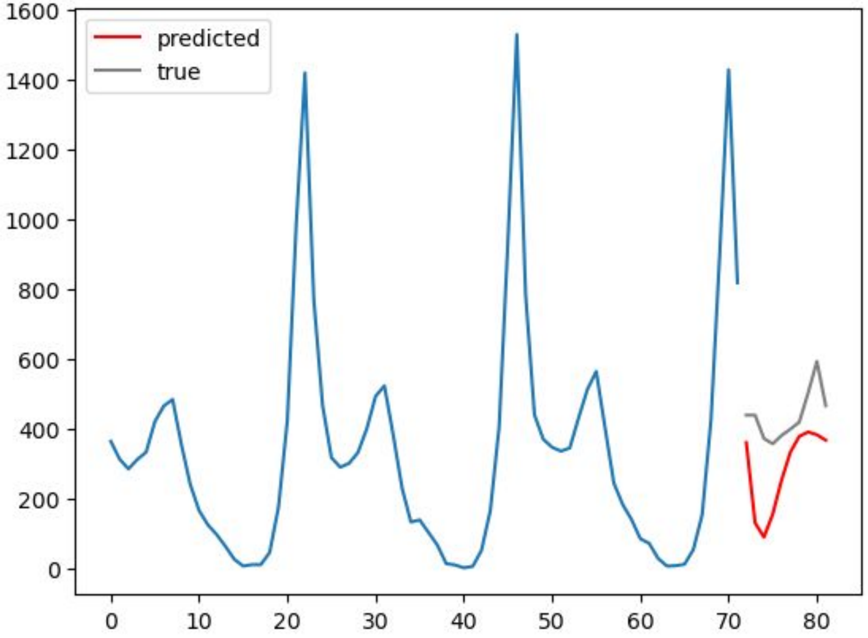
Demand Prediction Model

Feature importance from Random Forest Baseline



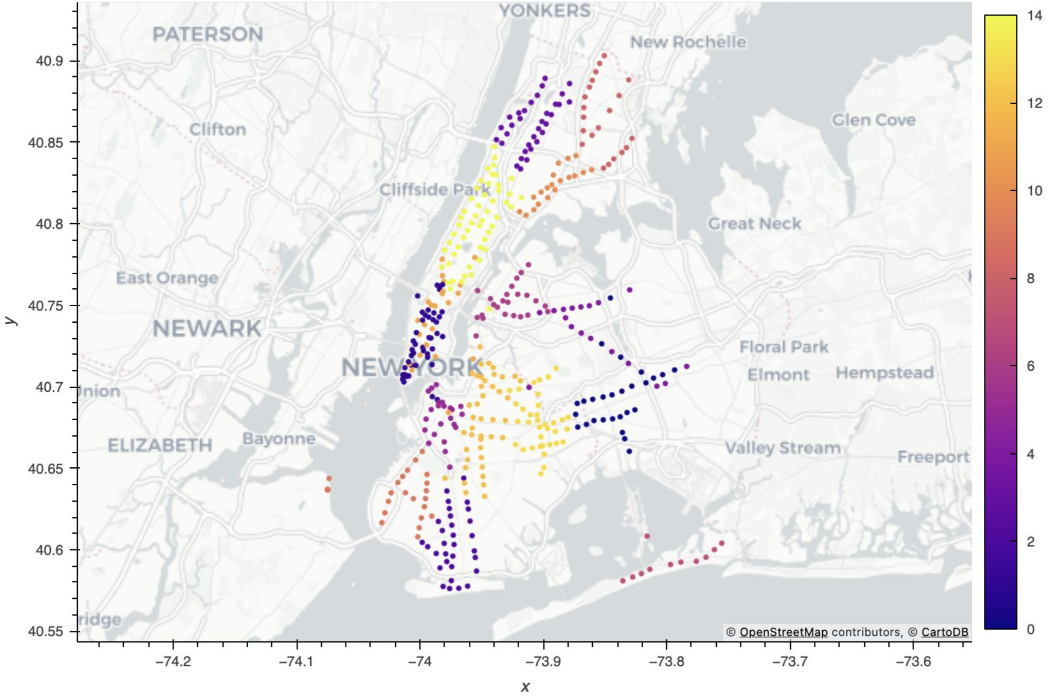
$R^2 = 0.80$

Prediction example from ARIMA(2,2,2)
Using past 72 hour window data



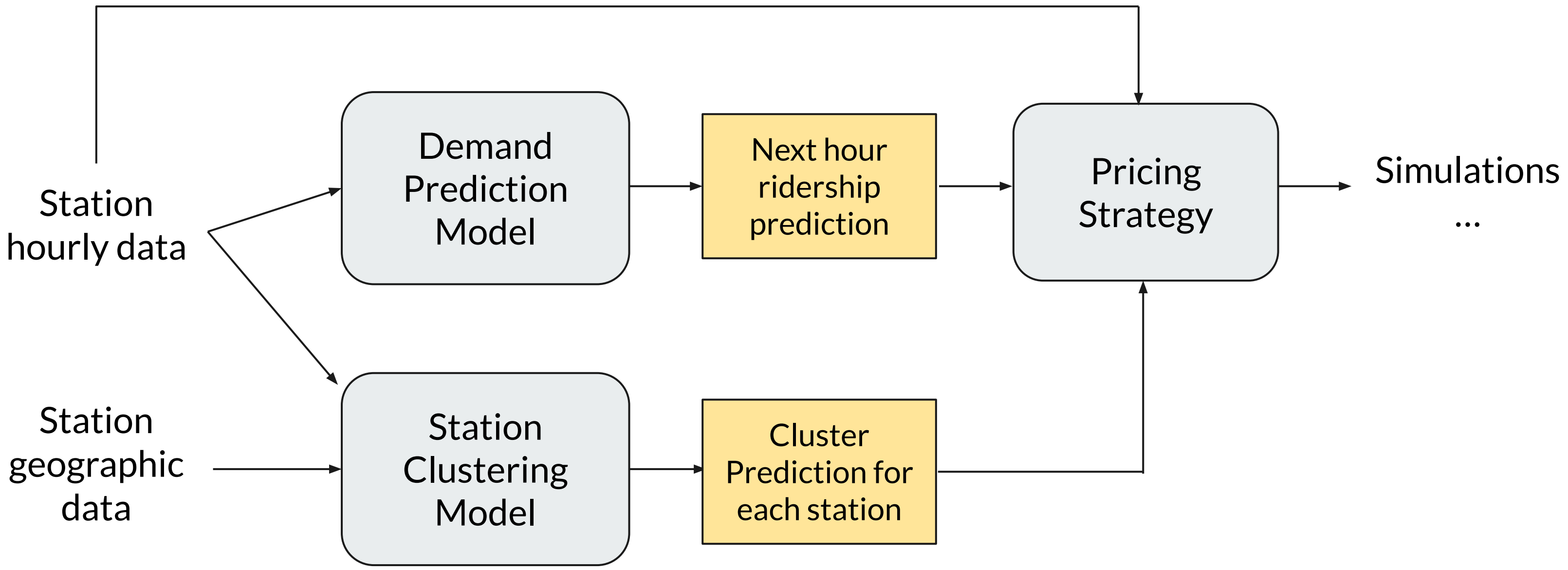
$R^2 = 0.86$

Clustering Stations

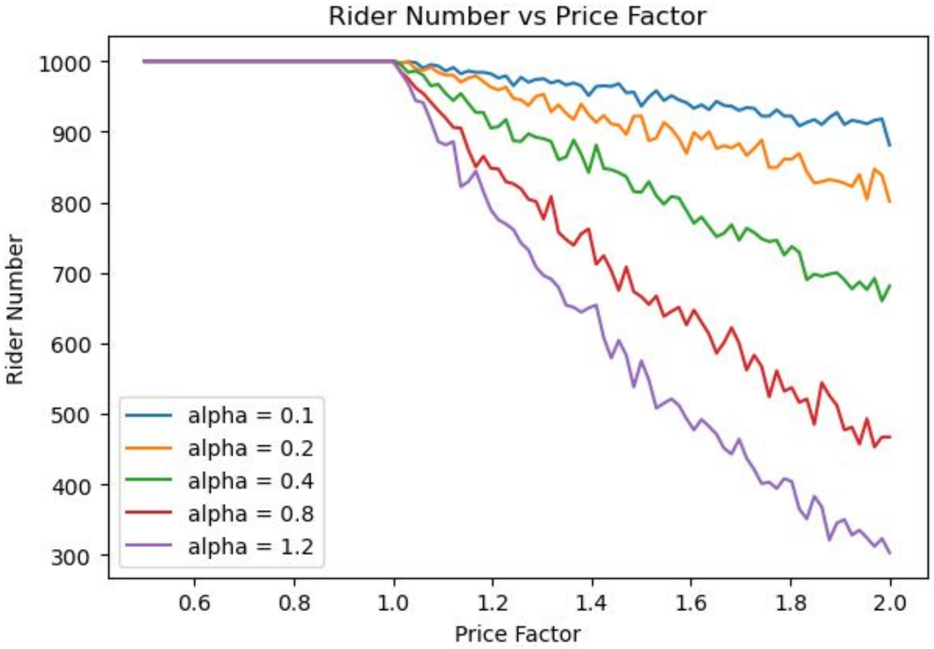


- To avoid exploitation, we do not want stations that are close to each other have very different price.
- Based on both geographic and hourly ridership data, we used spectral clustering to cluster all the stations into 15 clusters.
- In the clustered pricing, subway stations of the same cluster will share a same price.

Model Framework for Dynamic Pricing



Simulation Setup



- To verify our pricing strategy, we simulated it on certain time frame to compare the revenue using dynamic pricing and current pricing.
- We simulate how the rider will respond to the price change using a well defined exponential function. The hyperparameter α represent how much resistance riders have against price change.

Simulation Result

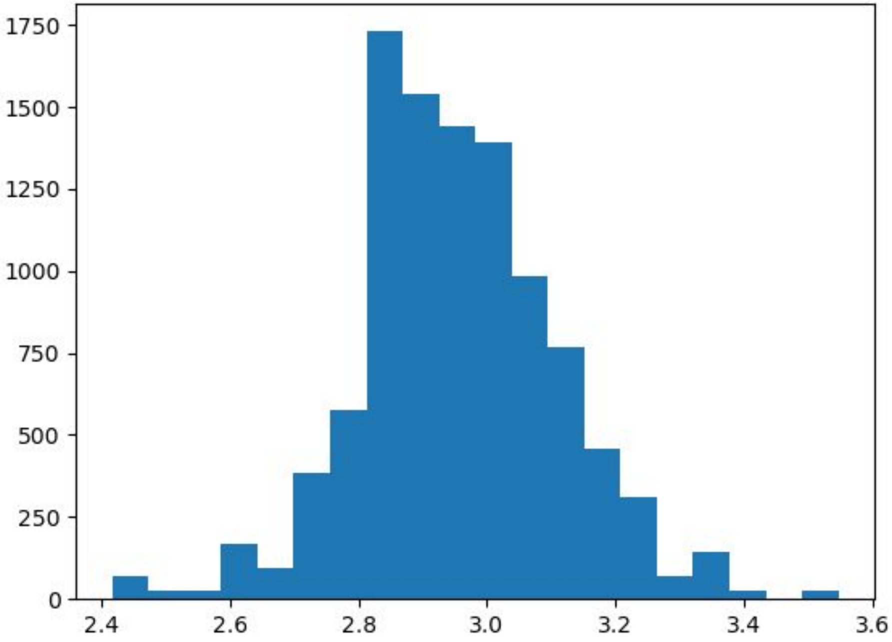
*Under default hyperparameter setting, randomly select 100 stations for testing.

alpha	Base Revenue	Revenue using Dynamic Pricing	Revenue using Clustered Dynamic Pricing
0.2	90596.0	95840.7	**96025.0**
0.4	90596.0	94238.7	**94929.5**
0.8	90596.0	91836.0	**92742.5**
1.2	**90596.0**	88857.2	90381.4

Even under very extreme case, our model performs very well.

Example:

Here is the distribution of our model's non-cluster pricing for all NYC subway stations at 5/1/2023 (Mon), 12:00 p.m.



Top 5 Stations
Times Sq-42 St
Grand Central-42 St
14 St-Union Sq
34 St-Herald Sq
Fulton St

Conclusions

- Our model performs well in predicting riderships and clustering stations
- Fares are adjust based on demand and location prediction, ensuring efficient resource allocation.
- Our pricing model can generate more revenue than current pricing of MTA.
- Our model provides lower fares during less busy times which encourages off-peak traveling and can encourage businesses to allow more flexible working hours.
- Higher fares during peak times can help manage overcrowding.

Thank you!!!!

Any questions?