(Do not share without authors' permission as it is under review for publication)

CIEN E4011
Big Data Analytics in Transportation Systems
Professor Xuan Sharon Di, PhD

Analyzing the effects of price change on NYC Ferry Ridership

Due Date: May 11th, 2023

Group One Members: Avikal Sagar (as6804)
Irtiza Sheikh (is2745)
Luke Demo (lbd2127)
Michelle Bermudez (mab2467)
Mohommad Rizwan (mr4269)
Samuel Choung (sc4988)

Remarks:

INDEX

TABLE OF FIGURES

Abstract

The NYC Ferry System was established on May 1st, 2017, with the East River and Rockaway Routes. The system was expanded to nine ferry routes at its peak, which connected all five boroughs in New York City while over 6 million people opted to take the ferry as their mode of transportation [1]. However, in 2022 only seven of the nine total ferry routes were operational, with one operating on a seasonal basis. The remaining two routes were discontinued or postponed.

Due to the expansion of the NYC Ferry System and its profitability, there have been questions regarding its financial viability and past expenses. To address these concerns, Mayor Eric Adams announced on July 14th, 2022 - five years after the launch of the first NYC Ferry - that there will be a change in the ferry fare system to allow for a more "equitable, accessible and fiscally sustainable citywide ferry system" [2].

The primary objective of this paper studies if there was a change in ridership after the $1.25 increase in the ferry fare for tourists and non-frequent ferry riders occurring in September 2022. Open-source datasets were used to analyze the ferry ridership, finding a slight decrease in passengers after the price change. This finding was supported by the application of the Machine Learning algorithm SARIMA, which was implemented to predict the expected number of passengers using historical data. However, the results do not provide a guarantee that this shift in ferry boardings was solely due to the pricing change as certain seasonal, COVID-related, and behavioral trends may persist.

Moreover, revenue was analyzed to understand the relationship between ticket price and route profitability. The price elasticity of demand of the five most popular routes was investigated, showing that for specific routes the number of tickets purchased operated independent of the price change.

Key Words: NYC Ferry, MTA Subway, Ridership, Price Elasticity, Time-series Analysis

1. Introduction

1.1 Project Background

NYC Ferry History

New York's first ferry dates to 1642 and crossed the East River between what today is known as Lower Manhattan and Brooklyn. On the west side of Manhattan, the first ferry to cross the Hudson River towards Jersey City was in 1661. By 1904, there were 147 ferries servicing the New York City area. However, due to the construction of the Williamsburg and Brooklyn Bridges and tunnels crossing the East River, coupled with car ownership and accessibility to subways, most ferries were discontinued by 1967 [3].

Fast forward to early 2011, the NYCDOT and NYCEDC (NYC Economic Development Corporation) released a Ferry Study [4] along the NYC Shores due to an uptick in residential buildings along the NYC rivers. Ferry service along the East River started in June 2011 with seven stops between Pier 11/Wall Street and East 34th Street. The demand for the ferry exceeded expectations by 250% within the first four months, with a total ridership of 350,000 from the expected 134,000. Unfortunately, due to Hurricane Sandy in 2012, the ferry system was severely damaged, and due to budget constraints from the government and the cost of operations, ferry service culminated on October 31st, 2014 [5].

In 2015, after NYC gained its strength following the devastating Hurricane Sandy, Mayor DeBlasio ordered a comprehensive ferry study to create and expand ferry routes that would service all five boroughs in NYC. The 2015 study determined that to increase ferry ridership, ferry fares will have to be lowered to $2.75, and the city would have to cover the remaining "subsidized" cost of $12.88 per rider [6], regardless of weekday or weekend. It should be noted that the previous ferry fares were $4 for weekdays and $6 for weekends [7]. The driving purpose of lowering the ferry rates was to make the ferry affordable and promote "transit equity" - making it the same price as the subway and bus, the two other popular modes of transportation in NYC. On May 1st, 2017, the East River and Rockaway routes launched with 6,400 riders on the first day. At the end of the first week, these two routes saw 49,000 riders. By July 26th, 2017, the NYC ferry system had 1 million riders - just two months after launching the new ferry system. It should be noted that the

Lower East Side Route was discontinued in 2020 and the Coney Island Route has been postponed indefinitely due to budget constraints.

NYC Ferry Today

By 2022, the NYC ferry system had expanded to six ferry routes in addition to the "shuttle" ferries to Staten Island, Governor's Island, and Rockaway Beach. These ferries saw roughly 6.1 million riders in 2022 alone.

Figure 1: NYC Ferry Routes

On July 17th, 2022, Mayor Eric Adams announced a change in the fare price for the NYC Ferry System to allow for a "more equitable, accessible and fiscally sustainable citywide ferry system" will start in September 2022.[2] Additionally, this new pricing system will be bringing the revenue up by $2 million [3] to compensate for the underreporting of $255 million of expenses for the ferry system which started in 2015. [5] This new fare pricing system is targeted to low-income New Yorkers and senior and disabled residents by offering them an even lower cost of $1.35 per ferry ride. Furthermore, everyday riders - people who rely on the ferry system as their only means of transportation - can opt to pay $27.50 up front for a 10-ticket package, which keeps the ferry fare at the previous $2.75. However, for visitors or infrequent riders, single ride fares increased to $4 starting in September 2022.

2. Objective

The goal of this project is to determine if there has been a change in ferry ridership after the $1.25 increase was implemented in September 2022. To accomplish our goal, we are using open datasets available in the NYC Open Data and MTA websites. Since the dataset is extensive, we are narrowing the dataset to include data from January 2022 until January 2023. However, we will be using the data available from 2017 and 2018 to validate our results and determine if they follow the same trend. We are not using data from 2019 or 2020 because of COVID-19 in ferry ridership. Once we have established if there has been a change in ridership after September 2022, we will analyze ridership within the ferry routes which saw the most ridership during this time. Figure 2 below shows the total ferry ridership from July 2017 until January 2021, divided into nine ferry routes.

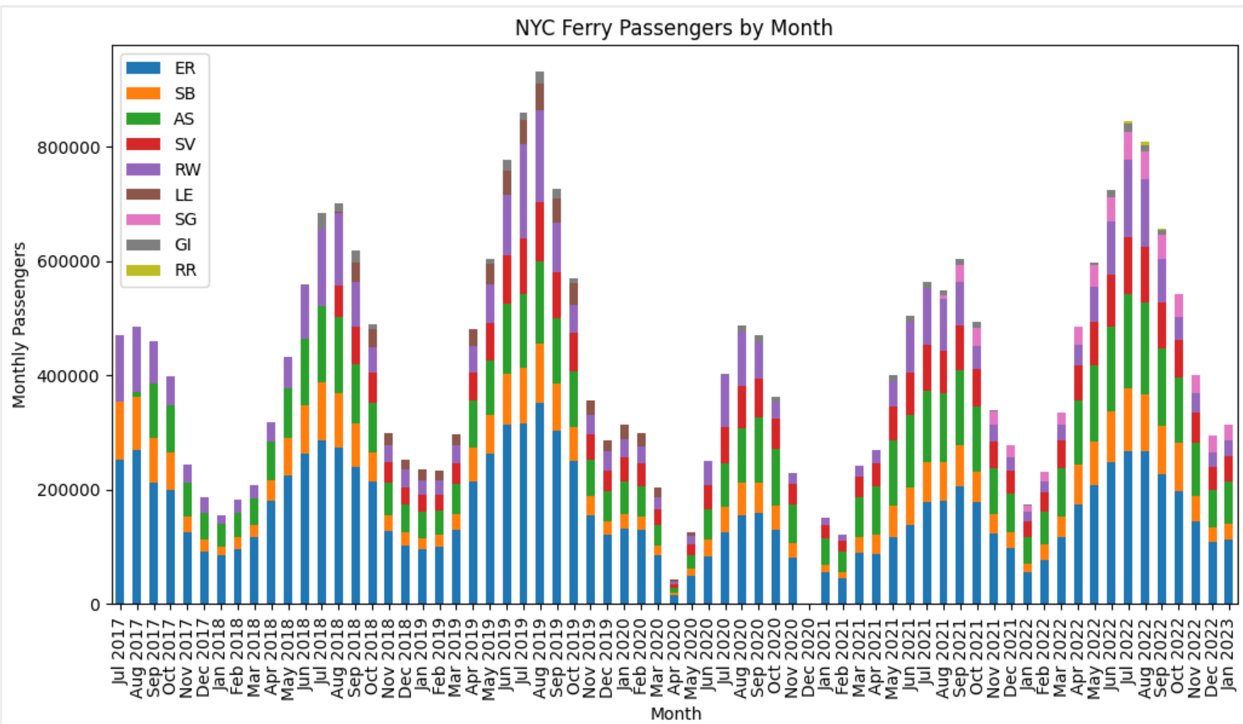


Figure 2: NYC Ferry Ridership

The purpose of analyzing the ferry routes is to determine if they are price inelastic - riders will pay whatever cost to use the ferry since it's most likely their only way of transportation. Additionally, we will be looking into the NYC Ferry revenue and determining if an increase in ridership signifies an increase in revenue, which was one of the reasons why Mayor Adams incorporated the new ferry fare system.

Lastly, we will investigate subway ridership within a 5-block radius of the ferry stations. The purpose of this is to find the effect of the increase in ferry fare on subway ridership - are people within this area taking the subway rather than the ferry due to the price increase? To do this, we will combine the ferry dataset provided by NYC Open Data with the Subway Turnstile data found on the MTA website. We will be narrowing down the subway turnstile dataset to the dates mentioned above and filtering for the specific subway stations within the vicinity of ferry stops.

2.1 Limitations of Objective

One of the main limitations of our project is that we only had four months of data after the $1.25 fare increase in September 2022. Since these trends require some time to take effect, we can't accurately evaluate if there was a change in ridership. However, based on the data from 2017 and 2018, the data during our evaluation period seems to follow the same pattern. One way we could work around this is to implement machine learning to predict the ridership for the rest of 2023 based on the previous trends.

Another main issue we had with our report came down to the data cleaning. To merge the ferry and subway datasets, we had to create a dataset with subway lines and stops - which we had to do manually using google maps. Another constraint is that not everyone would necessarily use the subway - some people opted for the bus, taxi, or Citibike.

2.2 COVID-19 Effect

In this section, we will describe the impact of COVID-19 on ferry ridership. The first case of COVID-19 was reported in New York on March 1st, 2020, and the number of cases quickly skyrocketed, significantly impacting the public transportation system, including NYC Ferry. According to the result of the NYC Ferry September 2020 Survey [8], the percentage of commuters using the ferry dropped to 43% compared to 73% in 2019 during the peak periods. This decline was expected as many people continued to work from home, while schools and universities were closed during this period. Figure 3 compares the weekly ferry ridership from 2020 and 2019.

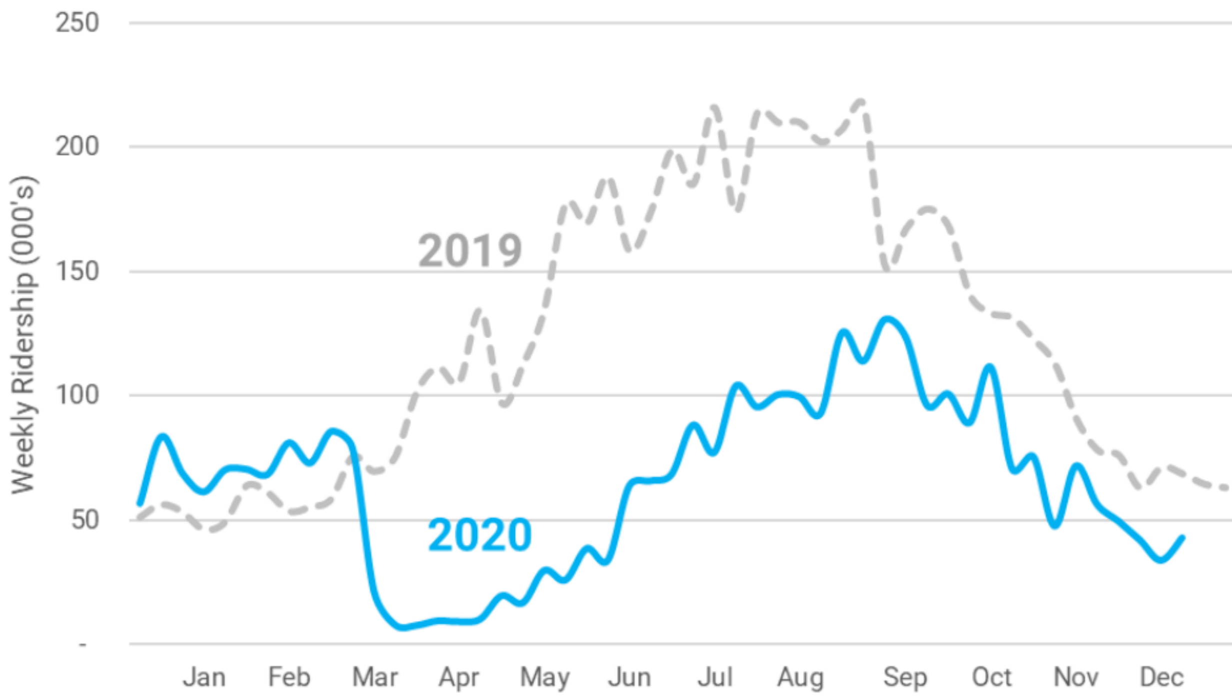


Figure 3: Effect of COVID-19 on Ferry Ridership

3. Data

3.1 Data Organization

Our first task after sourcing the datasets from the NYC Open Data and MTA websites was to clean and organize the datasets. To do this, we first focused on the Ferry Ridership dataset and then the Subway Ridership dataset.

Ferry Ridership

The NYC Ferry Ridership dataset retrieved from the NYC Open Data website consists of hourly records of NYC Ferry Ridership starting on July 1st, 2017, and ending on January 31st, 2023. This dataset includes hourly entries, which are categorized according to the day of the week (weekday or weekend), route names, ferry landings, and the directions of the ferry (Southbound or Northbound). This rich dataset consists of over 1.8 million records, proving to be a substantial enough dataset for our study.

Given the richness of the Ferry Ridership dataset, we had to perform data cleaning techniques to easily manipulate the data before properly executing the codes. The first step in cleaning the dataset was to determine that the data was in the form of a Pandas data frame after running an analysis in Python. We then proceeded to run some code to find out the type of data within the dataset and if there were any empty or null values. This crucial step allowed us to identify 73 missing values in the number of boardings and several direction blanks. Since these null and blank values don't affect the total ferry ridership, we decided to omit them.

After cleaning the data, we discovered the overall ferry ridership from 2017 until 2023. We then categorized the dataset into the nine different Ferry Routes - East River, South Brooklyn, Astoria, Soundview, Rockaway, Lower East Side, St. George, Governors Island, and Rocket Rockaway.

To acquire a deeper comprehension of the ridership for each route, we dove deeper and organized the ridership for each route separately. The purpose of this was to determine if all the ferry routes followed the same trends and to visualize which route had little to no effect on the overall ferry ridership -we can conclude that the RR (Rockaway Rocket) is a seasonal Ferry Route that was introduced in 2022 to perform as a “shuttle” between Manhattan and Rockaway Beach as seen in Figure 4 below.

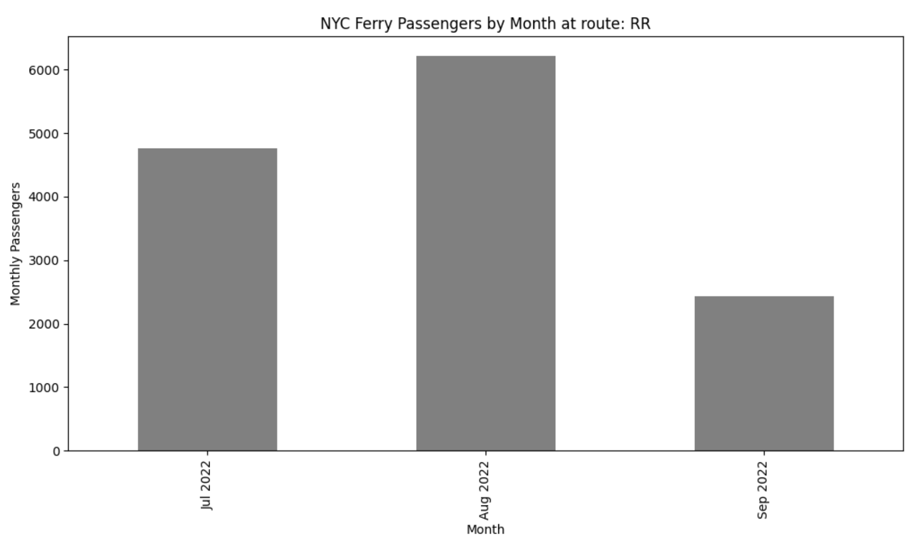


Figure 4: Rockaway Rocket Ferry Ridership

We also categorized the dataset based on the ferry direction - whether it was northbound or southbound along the East and Hudson Rivers. Given that the figures for the northbound and southbound routes below (Figures 5 and 6) have the same trends, we can infer that riders use the ferry for their daily commute.

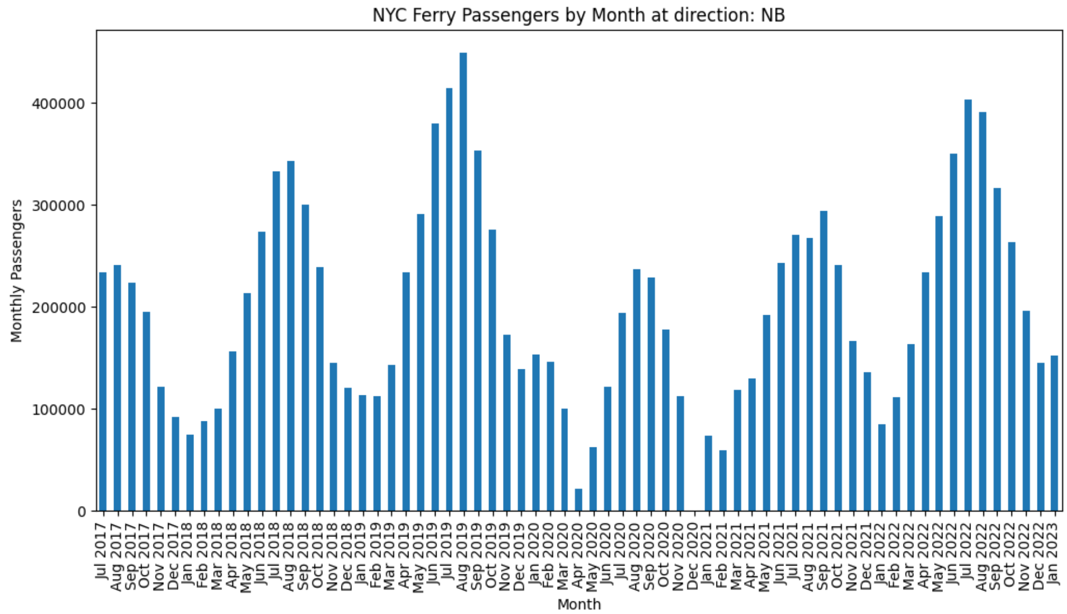


Figure 5: Northbound Ferry Ridership

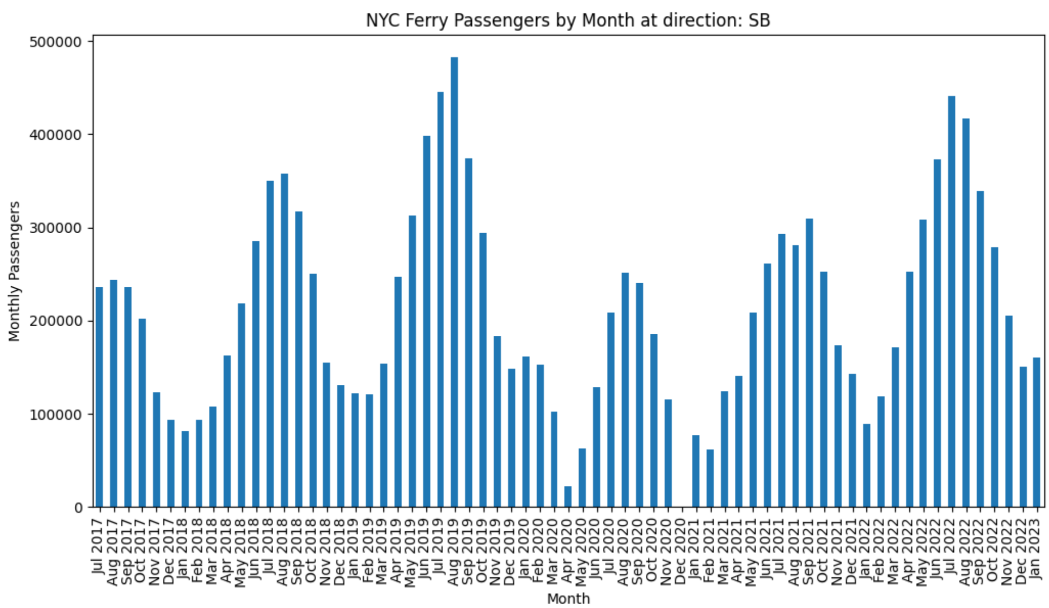


Figure 6: Southbound Ferry Ridership

Moreover, we observed that the five ferry routes with the most ridership were the East River, South Brooklyn, Astoria, Soundview, and Rockaway. The total ridership since July 2017 for these five routes is tabulated below in Chart 1.

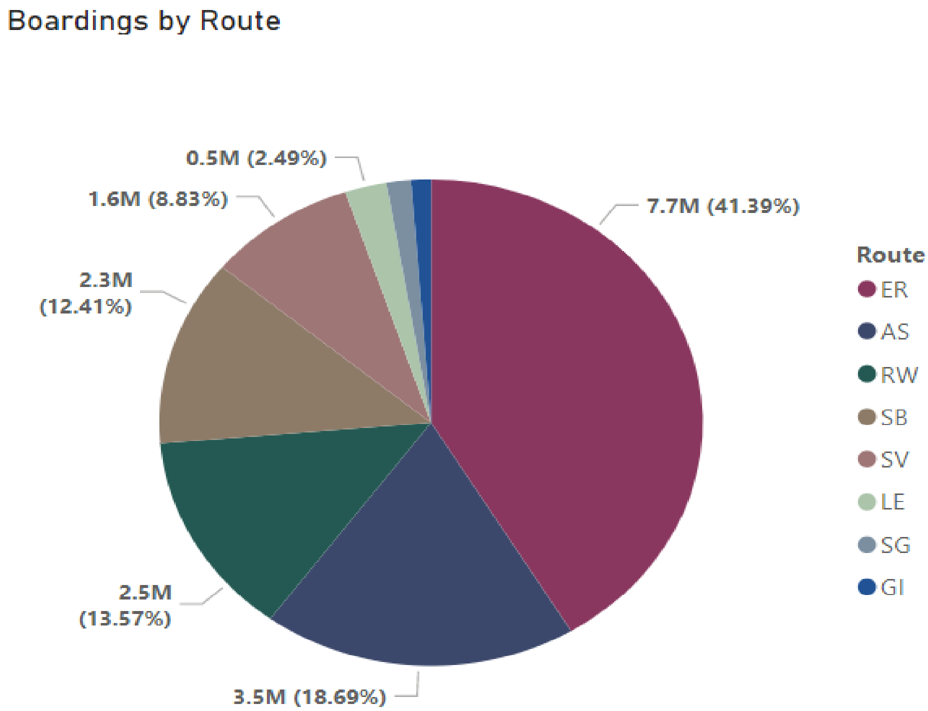


Chart 1: Total Ferry Ridership Route Distribution

The last task regarding organizing the dataset was to create an Excel sheet that tabulated all the ferry landings for each route. We manually entered the coordinates for each ferry landing into the Excel sheet to find the subway station names and coordinates within a 5-block radius of the ferry landing, as discussed in the following sections. Tables 1 and 2 below show a sample of the chart we created for the East River route.

Ferry			
Ferry Route	**Dock Name**	**Ferry Latitude**	**Ferry Longitude**
ER	Hunters Point South	40.741800	-73.961300
ER	Greenpoint	40.731500	-73.963190
ER	North Williamsburg	40.721050	-73.965130
ER	South Williamsburg	40.708570	-73.970313
ER	Dumbo / Fulton Ferry	40.703760	-73.995710

Table 1: Ferry Landing Information

Subway			
Station Name	**Turnstile STA Name**	**Station Latitude**	**Station Longitude**
Vernon Blvd-Jackson Av	VERNON-JACKSON	40.742626	-73.953581
Greenpoint Av	GREENPOINT AV	40.731352	-73.954449
Nassau Av	NASSAU AV	40.724635	-73.951277
Marcy Av	MARCY AV	40.708359	-73.957757
York St	YORK ST	40.699743	-73.986885

Table 2: Subway Station Information

Subway Ridership

Given the complexity of the NYC MTA Subway system and data, we decided to narrow down the dataset to the dates corresponding to our study - January 2022 until January 2023. This dataset contains the total number of entries and exits at each subway station in all five boroughs with the respective subway lines with the date and time for each recorded entry and exit.

The original dataset on the MTA website was in the weekly format. To analyze and understand the effects of the price change on ferries and compare the subsequent effect with the MTA subway ridership, we had to narrow down our study to data from January 2022 to January 2023. Since downloading weekly data multiple times for the entire year is cumbersome, we wrote a Python script to extract all data in the given timeframe using the MTA website's URL. The file was saved automatically. as a .csv file for use during the analysis.

We used the subway turnstile data to find if there was an increase in subway ridership after the ferry fare price increased. Therefore, we had to aggregate both ferry and subway data. Once we created the new dataset, we had to find subway stations within a 5-block radius of the ferry landings using Google Maps and create an Excel sheet with the station names. An important step was to ensure that the subway station names were in the same format for both the ferry and turnstile datasets. After establishing the station names, we filtered them from the January 2022 - 2023 dataset. Lastly, we decided to separate the subway ridership for the five ferry routes with the most ridership - East River, South Brooklyn, Astoria, Soundview, and Rockaway. Tables 3 and 4 demonstrate the before and after of the data cleaning.

	C/A	UNIT	SCP	STATION	LINENAME	DIVISION	DATE	TIME	DESC	ENTRIES	EXITS
0	A002	R051	02-00-00	59 ST	NQR456W	BMT	1/29/2022	3:00:00	REGULAR	7682427	2659659
1	A002	R051	02-00-00	59 ST	NQR456W	BMT	1/29/2022	7:00:00	REGULAR	7682432	2659667
2	A002	R051	02-00-00	59 ST	NQR456W	BMT	1/29/2022	11:00:00	REGULAR	7682443	2659708
3	A002	R051	02-00-00	59 ST	NQR456W	BMT	1/29/2022	15:00:00	RECOVR A	7682464	2659740
4	A002	R051	02-00-00	59 ST	NQR456W	BMT	1/29/2022	19:00:00	REGULAR	7682515	2659786
5	A002	R051	02-00-00	59 ST	NQR456W	BMT	1/29/2022	23:00:00	REGULAR	7682535	2659800
6	A002	R051	02-00-00	59 ST	NQR456W	BMT	1/30/2022	3:00:00	REGULAR	7682542	2659808
7	A002	R051	02-00-00	59 ST	NQR456W	BMT	1/30/2022	7:00:00	REGULAR	7682543	2659819

Table 3: Subway Data before Cleaning

	STATION	datetime	MONTHLY BOARDINGS
0	21 ST	2022-01-31	89701049
1	21 ST	2022-02-28	834002369
2	21 ST	2022-03-31	932826201
3	21 ST	2022-04-30	906634079
4	21 ST	2022-05-31	939916257
5	21 ST	2022-06-30	918533077
6	21 ST	2022-07-31	947142015
7	21 ST	2022-08-31	950035256

Table 4: Subway Data after Cleaning

3.2 Issues with Data

It is typical for large open-source datasets to contain errant values, mismatches, out-of-order sequences, or more problems. To address these concerns a data-cleaning procedure was established. A series of tests and solutions were conducted on each row that contained the seven variable types: Date, Hour, Route, Direction, Stop, Boardings, Weekday/Weekend.

1. Empty values – each entry into the set was tested to discover if there was no value entered. 73 entries were found and were removed.
2. Blank values – similarly, each entry was processed to determine if there exists only empty space (i.e., "   ") instead of a numerical or text-based value.

These two processing tools allowed us to accurately parse the existing data to only contain values that have a physical significance. We wanted to ensure that when one person entered the ferry on a certain date that it would be directly and accurately represented by one entry on the time-history graph. The two cleaning functions lead to Figure 7 (a).

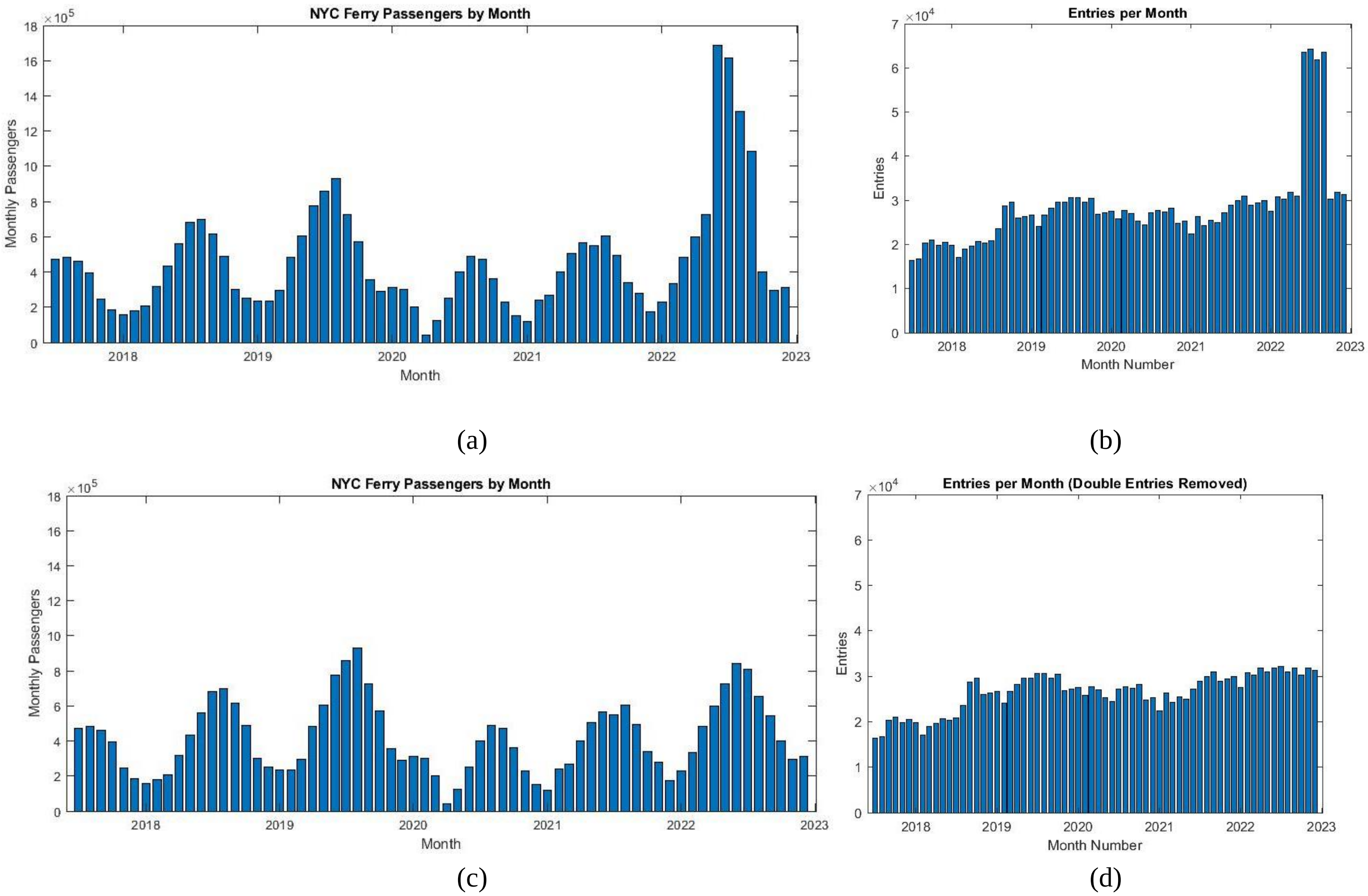


Figure 7: Ferry Ridership after removing blank entries {(a) and (b)}, and then after removing double entries {(c) and (d)}

It initially appears that there was a significant increase in ridership in June 2022. Perhaps this could be true; summer month trends increase and all public transit post-covid has risen. However, intuitively, does it make sense for 1,000 people to consistently board a ferry each day in May, then 2,000 people each day in June? Behaviorally, what would spark this? We reviewed available online sources and found no impetus. Consequently, this demanded further investigation into the data. It was discovered that a series of double entries existed from June 2022 through September 2022. This was confirmed both in Python and in MATLAB, where individual entries were compared (e.g., duplicated entries were found on dataset lines 150 and 300). The data was corrected and is displayed in Figures 7 (c) and 7 (d). Figure 7 (b) shows that there was an abnormally high number of entries for those months, and Figure 7 (d) returns the data to a consistent trend. This process elucidates the need for applying a top-down approach to data analysis. It is important to zoom out and see the whole picture to detect where numbers could be incorrect.

4. Analysis

4.1 PowerBI

PowerBI is a business analytics service by Microsoft that provides interactive visualizations and business intelligence capabilities with a simple interface for end-users to create reports and dashboards [9]. It is one of the most popular business intelligence tools available today used by numerous industries and organizations because of its accessibility for everyone in an organization to access and visually process shared data. With most engineering firms procuring the Microsoft Office 360 suite as a mandatory loadout for work desktops and laptops, the PowerBI application is just an install away.

Features

Power BI can connect to various data sources, including Excel spreadsheets, SQL Server databases, and other cloud-based data sources. Given that our study consists of working with Ferry and Subway Ridership datasets, we were able to generate rudimentary reports with ease which can easily be queried using radio buttons, slicer menus, and sliders.

The Data Transformation feature within PowerBI allowed us to transform, clean, and reshape data effortlessly with Power Query, which enabled us to make informed decisions based on accurate data.

PowerBI's Data Visualization tool consists of many customizable charts, tables, and graphs. This feature allowed us to create powerful visualizations to convey our findings effectively. The foundation of any good PowerBI dashboard is the scorecard of the overall project. In our case, our home dashboards serve as the 30,000ft view of the entire ferry system, which can be broken down by lines based on user selections.

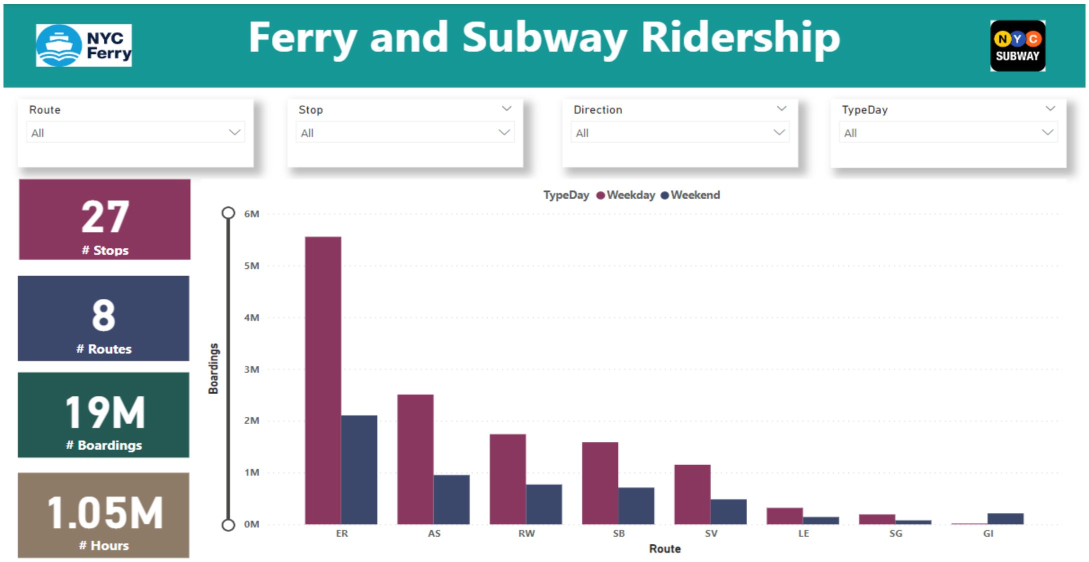


Figure 8: Ferry and Subway Ridership Scorecard

For those within an organization or research team that cannot easily process walls of data presented in .csv or .xls format, wants to avoid producing visuals for a static dataset, and easily track new changes(in our case, daily boarding or turnstile data as they are reported weekly), the home dashboards are a great way to have a connected and ever-changing dataset show these visuals with updated data since the PowerBI dashboard is always connected to a shared database.

The ArcGIS feature allowed us to visualize different values overlaid on an automatically generated overview map of the route stops and the breakdown bubbles. In this case, the breakdown of northbound traffic, compared to southbound traffic, shows the general popularity of the directional flow of each station. Counts and Line Graph visualizations show trends throughout a selectable range of dates and ferry stops which correlates to the closest subway station or group of stations to be analyzed.

The Calendar heat map visualization in Figure 9 below, which is built by Tallan, allows us to visually see a heatmap of ferry boardings and turnstile entries, sliceable by stops, subway stations, counts of each Year/Month/Day based on selection, and a visual comparison of the counts based on how dark the entry is compared to the other dates within the year.

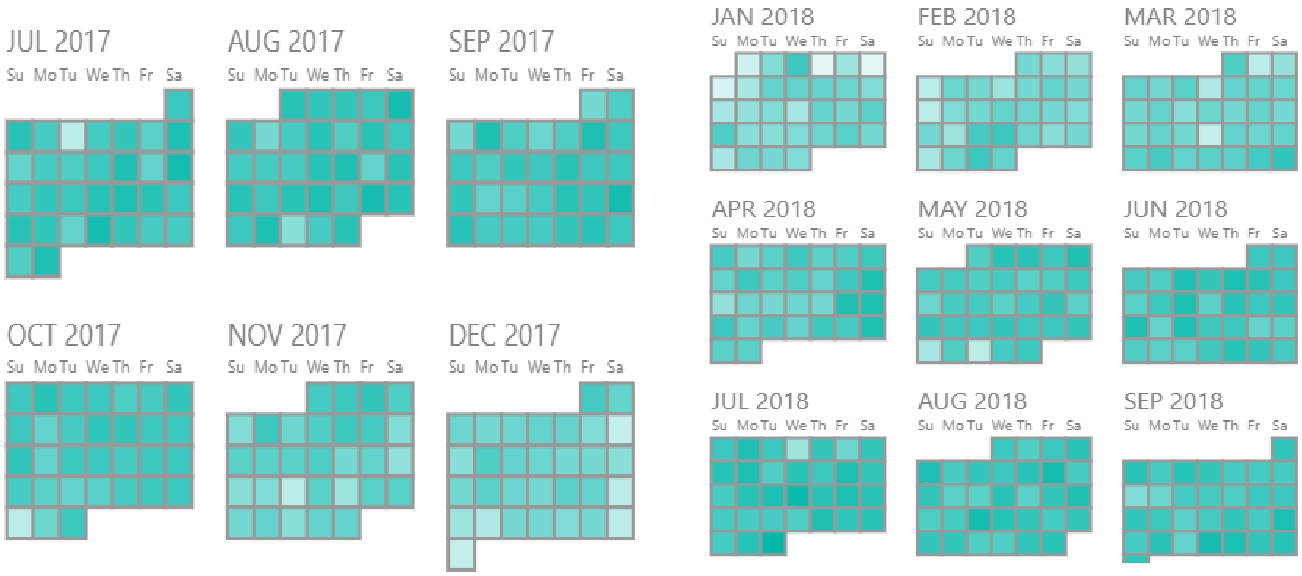


Figure 9: PowerBI Calendar Heatmap

Given that our team consists of multiple members, the collaboration attribute made sharing reports and dashboards with other users, allowing for efficient decision-making. When new data is entered into the shared database, every member can see up-to-date information and changes as the weeks progress throughout the year.

Analysis

After cleaning both datasets, we loaded them into Power BI. Since we are analyzing five ferry stops, we created a table in which we noted the latitude and longitude of these stops and the nearest subway stations corresponding to these ferry stops. Using PowerBI, we created a visualization of the ferry landings and the subway stations within a 5-block radius.

The Calendar heatmaps are an alternate way of viewing a general trend besides using the typical line graphs to show the same data counts. In our research, the calendar heat maps showed expected general trends and noticed anomalies within the data that required environmental context. In Figure 10 below, the calendar heat map of Hunters Point from January 2019 to January 2020 shows a heatmap that is replicable between all ferry stops in the system, but the heatmap is unusually void of medium to darker colors towards the end of March 2020, which we now know was the start of the Covid 19 pandemic. The lighter-colored and white boxes towards the end of March.

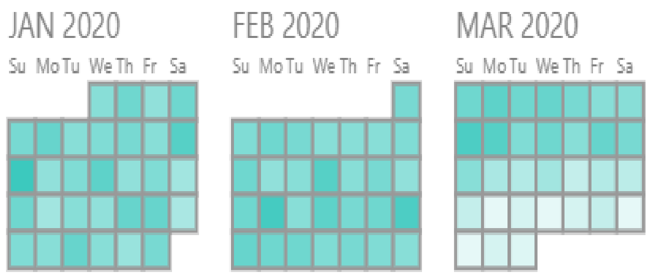


Figure 10: Hunters Point Calendar Heatmap

Microsoft's built in Time Series Forecasting functionality for PowerBI uses a regular interval of hours, days, months, and years to use Machine Learning to predict future points. The machine learning algorithm used is ETS AAA seasonal algorithm for time series forecasting which combines additive error, additive trend, and additive seasonality to the dataset's mean, variance, and other statistical properties. By setting our forecasting options to 30 additional points (representing the next month), 20 seasonality points, and a 95% confidence level, the ferry boarding data was reasonably accurate to the actual ridership for the following month, as seen in Figure 11 for the Greenpoint Stop below. Since the algorithm is limited by its sample size, monthly ridership data was the reasonably minimal subset of data to analyze, especially when it came to the lower count numbers/intervals of the ferry data. It can only be assumed that days of inclement weather affected ferry ridership much more than the subway turnstile data.

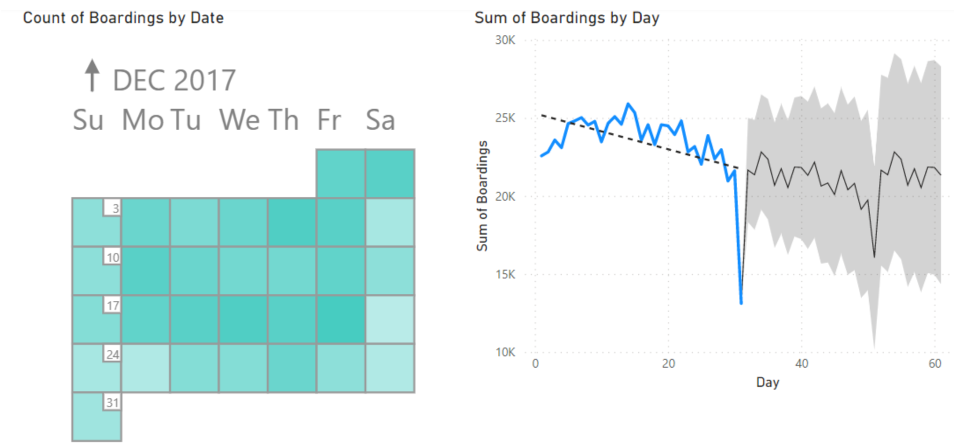


Figure 11: Greenpoint Stop Forecasting

Limitations

Microsoft's Power BI has certain limitations on the quantity of data that can be loaded and processed in a single report or dashboard. Since we cleaned and created new datasets, we did not face this challenge. Additionally, we used a free version of PowerBI which didn't allow us to leverage the advanced features only available in the premium version.

PowerBI's data modeling capabilities are not as advanced as other enterprise-level business intelligence tools, making it cumbersome to create complex data models. Although PowerBI offers a wide range of customizable charts and graphs, we cannot create completely custom visualizations, which may limit our ability to create unique and specialized reports and dashboards.

PowerBI may not be able to connect to the data source, or there may be authentication or permission issues. This can happen if the data source is not accessible, the connection string is incorrect, or the user does not have the required permission to access the data. Given that we retrieved our datasets from open sources such as NYC Open Data and Kaggle we bypassed data connection complications.

PowerBI's use case is primarily as a monitoring application for those that need to monitor general trends. If organizations like the MTA New York City Transit or NYC Ferry wanted to utilize PowerBI, it would be to monitor the accumulation of data as it progresses and to check for abnormalities as they pop up. Drastic swings in graphs or heat charts can hint towards missing data or poorly formatted data. But PowerBI's built in forecasting and Machine learning tools are limited to a very few selectable variables of a single pre-chosen algorithm that may not be suited for the dataset.

If the data source has formatting issues, such as missing values, inconsistent data types, or invalid characters may lead to errors when PowerBI tries to load the data or create visuals. To circumvent these issues, our team carefully configured the data source and connection settings in PowerBI and tested the dataset thoroughly before finalizing our findings.

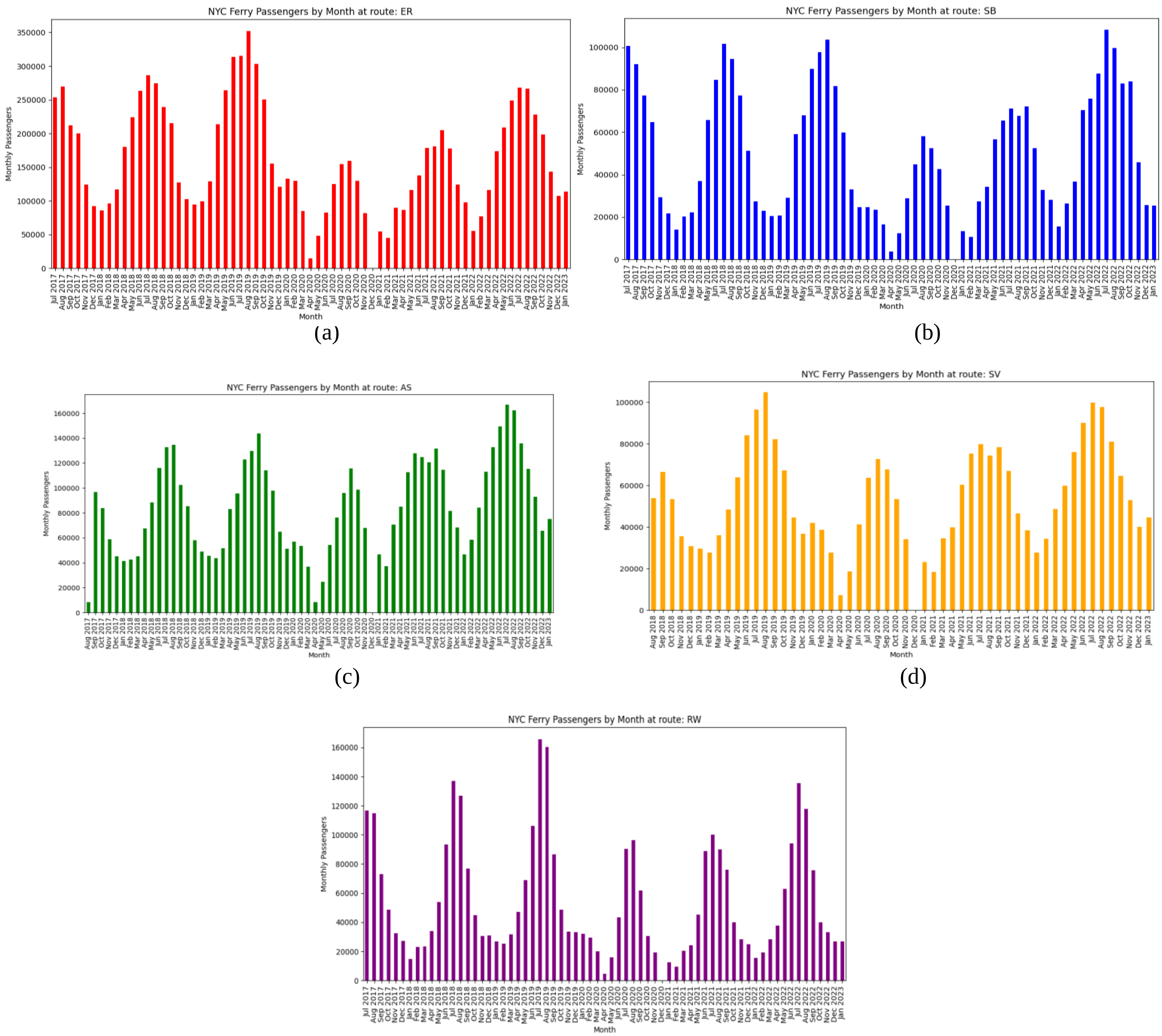
4.2 Python

Ferry Ridership

As mentioned in the previous sections, the main aim of our analysis for the NYC ferry data was to understand the effects of price changes - implemented in September 2022 - in each of the five main routes. The first step was to visualize the monthly boardings from 2017 and 2018 until the end of 2022, as shown in Figures 12 (a-e) in the following page.

Quite evidently, the generally observed trend in each plot was a sharp decrease in ferry boardings after August 2022. The two main factors contributing to this trend are seasonal effects

- cold weather drives lesser demand for in-water and partially open modes of transportation, as in this case, and the suspected price change.

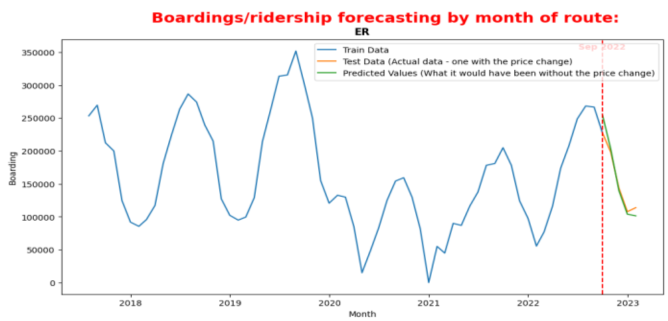


(e)

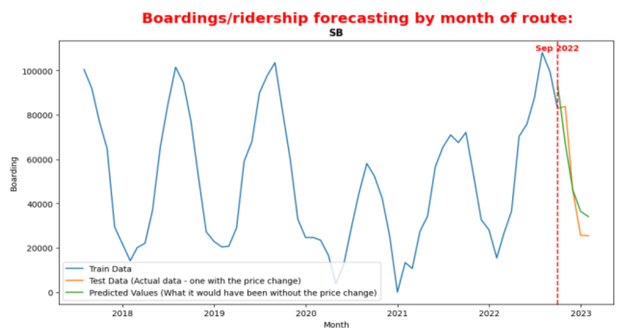
Figure 12:NYC Ferry Ridership per route {(a)-(e)}

To better understand the precise effects caused by price change solely, we used time-series forecasting models through Machine Learning algorithms, such as Seasonal Autoregressive

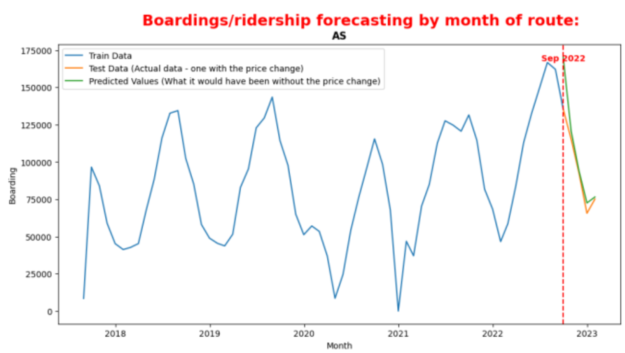
Integrated Moving Average (SARIMA). The motivation behind using such predictive models was to train it on data before the price change, that is, until August 2022, to learn the general trends, patterns, and cycles observed in the dataset due to seasonal effects (blue line). And then to make predictions on the test data (starting from September 2022), which shows what the boardings *could have been* had there been no price change (orange line). We also parallelly plotted the actual observations - boardings with price change (green line). This gives us a clear idea regarding the effects of price changes on ferry boardings per route. To understand this in a more simplistic manner, results are shown in Figures 13 (a)-(e) below.



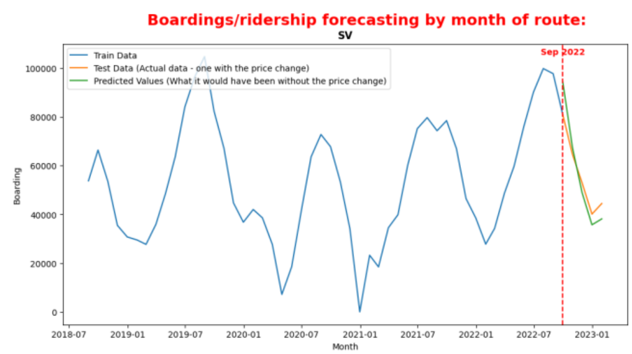
(a)



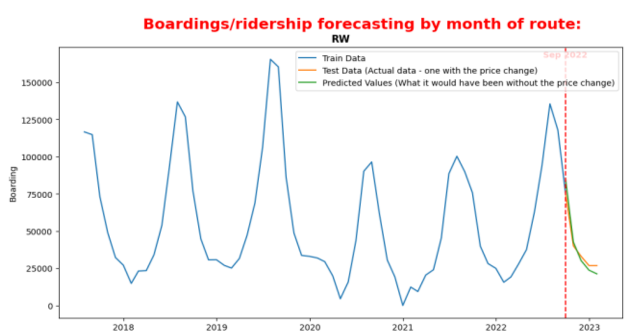
(b)



(c)



(d)



(e)

Figure 13: Ferry Ridership Forecasting {(a)-(e)}

To summarize:

- green line lies **fully above** the orange line: Route is price ELASTIC
- green line lies **fully below** the orange line: Route is price INELASTIC
- green line lies partially above & subsequently partially below the orange line: Route reversed the initial elastic effects
- green line lies partially below & subsequently partially above the orange line: Route reversed the initial inelastic effects
- green line and orange line fully coincide: Route is price INELASTIC (more precisely, constant)

Based on the plots generated through time-series forecasting, we can classify each of the routes with a certain amount of confidence in the following manner:

Route	Trend observed (in Figure 13)	Conclusion
ER	Green line above orange line	Price elastic
SB	Green line mostly below orange line	Price inelastic
AS	Green line above orange line	Price elastic
SV	Green line above orange line followed by vice versa	Price elastic effects recovered
RW	Coincide most of the time	Price inelastic

Table 5: Summary of Ferry Ridership Trends

Now that we assessed the price elasticity or inelasticity of each of the five main routes, it was also worth investigating the exact revenue or income variations after the fare increase. An inelastic price route, coupled with the fare increase, is more straightforward to understand as the revenue would likely increase or at least continue to be the same (not considering other factors as weather effects, that are present in our case, which is considered and explained more thoroughly in the next paragraph). However, we can't conclude the same regarding price elastic routes. In these cases, there is almost always an immediate decrease in revenue to a staggering loss in demand. In our case study, we can observe that the revenue constantly decreases for all routes, regardless of if the price is elastic or inelastic. This is due to the weather changes, as colder weather contributes to less residents or tourists taking ferries due to open decks and wind chill effects over water bodies.

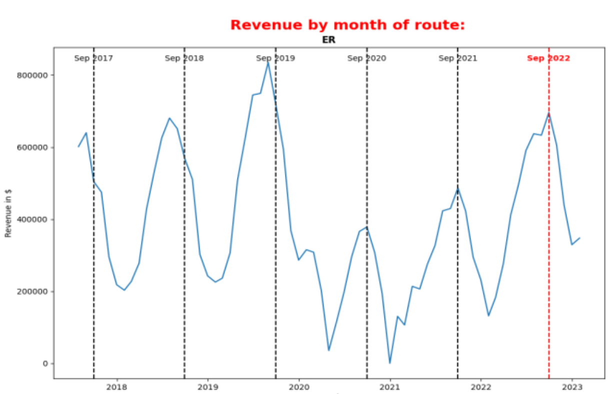
We precisely determined the revenue effects that are solely due to the fare increase by focusing on the slope of revenue in the same time frame (August to November) for all years from 2017 until 2022. If the slope of the line fell sharply, or mathematically decreased, in 2022 - compared to all the other years - then we can further validate that the route is price elastic. If the decrease wasn't as sharp, or followed a similar pattern to previous years, we can say that the route is price inelastic and the revenue decrease is as usual, purely due to weather effects, and that the changes in fare prices have nothing to do with it.

Time	Tourists/One-time traveler (A)	Frequent traveler (B)	Senior Citizens & kids under 5 (C)
Before September 2022*	2.75$	2.00$	2.75$
After September 2022**	4.00$	2.75$	1.35$

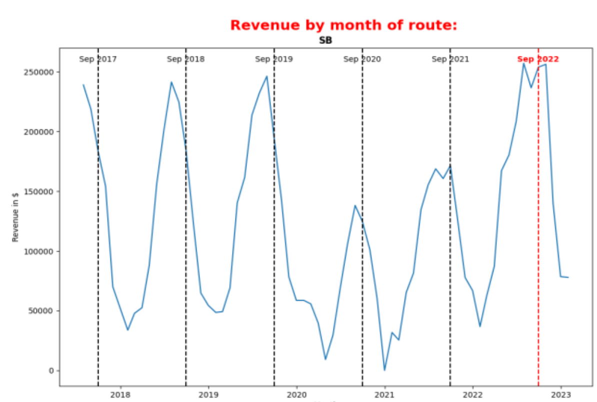
*Percentage share: before 09/22 - (A) + (C) = 50% & (B) = 50%
**Percentage share: after 09/22 - (A) = 39%, (B) = 48% & (C) = 13%

Table 6: Ferry Ridership Revenue Parameters

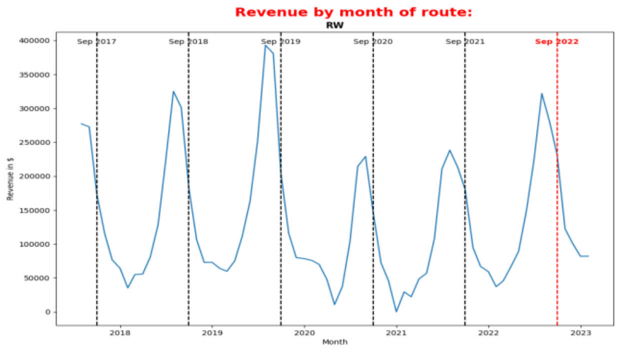
To calculate the revenue for NYC ferries as accurately as possible, the exact schemes of prices per category of riders before and after September 2022 were used and are seen in Figure 14 (a-e) below. NYC ferries charge less if you are a frequent traveler and more if you are a one-time traveler or irregular traveler. We then proceeded to determine the share of each category of riders to calculate the total revenue.



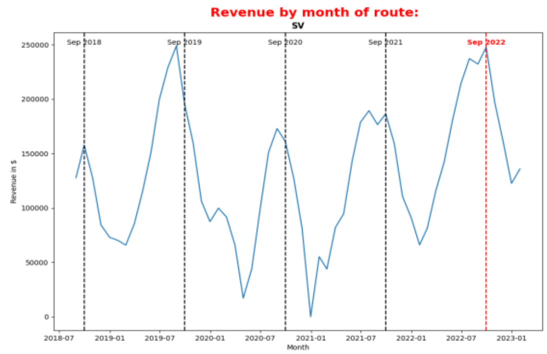
(a)



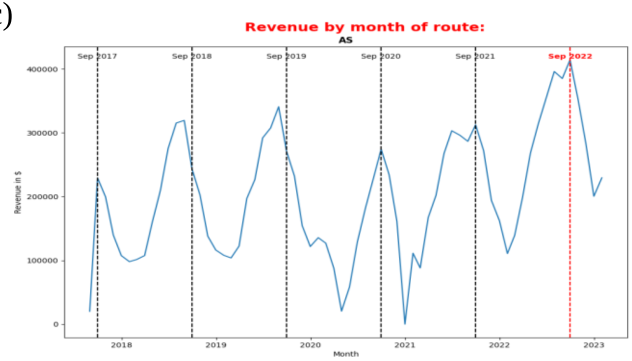
(b)



(c)



(d)



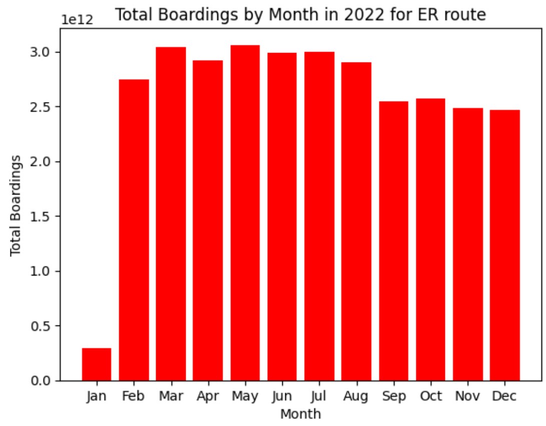
(e)

Figure 14: Ferry Ridership Revenue by Month {(a)-(e)}

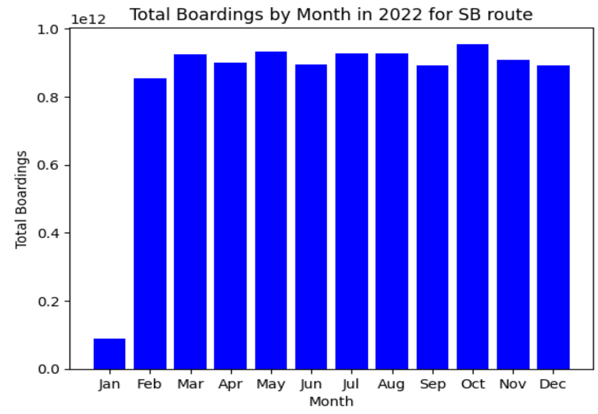
Subway Ridership

After analyzing the NYC ferry ridership data, we came to conclusions regarding the elasticity of the five major routes. Most commonly, we observed that the routes were price elastic, barring a few exceptions. But that poses yet another question. Considering that NYC is the hub for major industries such as finance, technology, education, and healthcare, while also considering the fast-paced, dynamic, and constantly growing economy, it is safe to assume that the demand for transportation and commute did not decrease as seen in the case of ferries. This is due to the price increase and weather changes, causing a shift in the crowd and its travel demand to other modes of transportation.

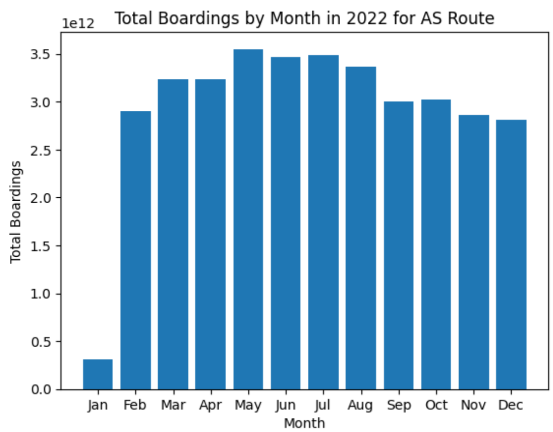
Hence, in this study, we chose to perform an elementary analysis of the subsequent effects on MTA subway ridership. We mainly focused on the subway routes within a 5-block radius of the ferry routes, and we can see the boardings in Figure 15 (a-e) below. These have been previously discussed under the "Data" section as well.



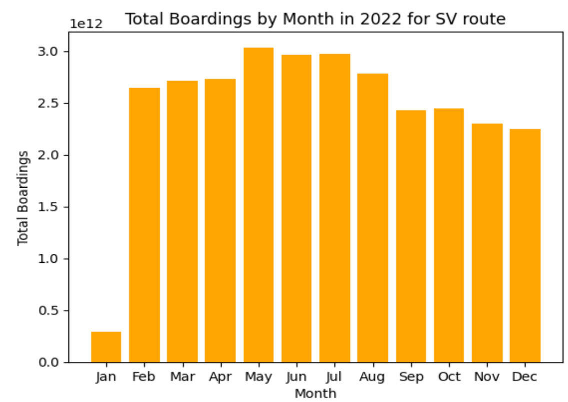
(a)



(b)



(c)



(d)

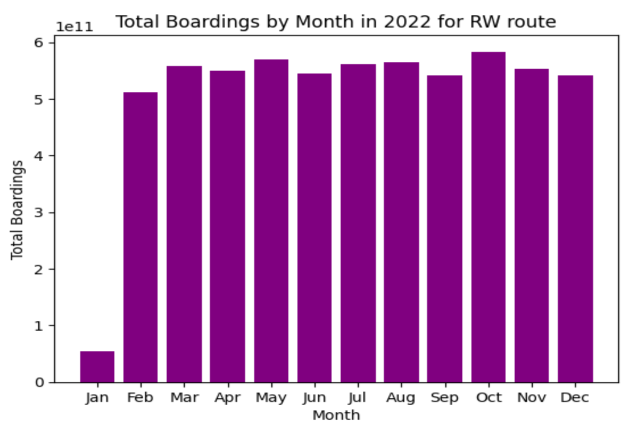


Figure 15: Ferry Boardings by Month {(a)-(e)}

5. Conclusion

The goal of this paper was to explore if there was a change in ridership after the $1.25 increase in the ferry fare for tourists and non-frequent ferry riders in September 2022. As stated in the previous sections, we can see that there has been a shift in ferry ridership amongst the five routes we selected for our study - East River, South Brooklyn, Astoria, Soundview, Rockaway, Lower East Side, St. George, Governors Island, and Rocket Rockaway. Based on our analysis, we can confidently say that there has been a decrease in ferry ridership after the fare price increase. However, we cannot attribute the change in ridership solely to the price increase, as stated in the previous sections. Some other factors that could have contributed to this decrease in ferry ridership are weather conditions during the colder months.

5.1 Results

If we compare our study period (January 2022 -2023) with the data from 2017, 2018, and 2019, we can see that the four years follow the same trendline - a peak in the summer months and then a gradual decrease during the colder months. However, if we compare the ferry ridership of September 2017, 2018, 2019, and 2022, we can conclude that there has been a consistent increase in ferry ridership until 2019. The decrease in ferry ridership in September 2022 can be attributed to the COVID-19 effect on mass transportation in the NYC Metro Area. The values in the table below are approximate values from the data shown in Figure 2, located in the Objective Section.

Month	Ferry Ridership
September 2017	450,000
September 2018	600,000
September 2019	700,000
September 2022	620,000

Table 7: Ferry Ridership in September

Once we concluded that there was a slight decrease in ferry ridership after September 2022, we wanted to understand which routes were affected the most. Therefore, we ran a Machine Learning algorithm with time-series forecasting models to predict boardings that *could have been* affected, with and without a price change. Through this machine learning process, we were able to establish that as the price increased, the demand decreased either completely, or at some point in time - except for the South Brooklyn route. Therefore, proving them to be price elastic.

Moving into Subway ridership - which was used to understand the sudden shift in ferry ridership - allows us to suggest that the sudden increase in subway ridership within a 5-block radius of the ferry landings was due to the increase in the ferry fare price. However, this is not guaranteed based on the limitation of the study mentioned in previous sections and the lack of data. We would need to analyze a substantial amount of data after the price increase to fully understand the implications of the ferry fare price increase. Additionally, we would need to understand the demographics of the area to truly understand this behavioral change.

5.2 Validation

We saw that time-series forecasting through Machine Learning validated the results we visualized for the ridership on the NYC ferries. The following hyper-parameters were established for our existing model:

1. Order: which refers to the non-seasonal component of the model and is programmed to (1, 1, 1)
2. Seasonal order: which refers to the seasonal component of the model and is programmed to (0,1,1,12).

Although pretty accurate, we can make further improvements by applying a hyperparameter tuning process.

Furthermore, alternative methods of time-series forecasting can be tried and tested. As mentioned previously the Seasonal Autoregressive Integrated Moving Average (SARIMA) was used. Facebook Prophet and a simple ARIMA model are other options that can be utilized for validation purposes.

5.3 Looking Ahead

This study serves as an initial insight into the price versus ridership discussion. Many factors obfuscate direct conclusions using the available data. However, this work has demonstrated that the ferry price did not create a significant shift or transformation in ridership. It is possible that a group of people who were already riding the ferry chose to seek alternate means of travel. Although, it is also possible that there was a concurrent upward trend in ridership on the yearly scale. Such a variation would not be able to be noticed by even the most advanced ML models because real human decision-making data was not accurate for the most dramatic COVID period. Consequently, a series of behavior polls could have been, and continue to be, carried out to assist the public's willingness to ride the ferry at the increased price. Psychological studies could show trends that would enable unification into a single regression-prediction model with ongoing data-driven passenger metrics. This future work would serve as one step further in capturing the Ferry's price elasticity of demand.

Additionally, the analysis presented here showing total boarding per route can be used to determine the efficacy of a specific route. Moreover, this data can be used by legislators to analyze the route's cost-effectiveness. The SARIMAX algorithm serves as a highly beneficial tool for comparing real-time revenue with predicted revenues. Further work could be done to incorporate short-term weather patterns (e.g., daily, or weekly trends) into the algorithm to effectively eliminate environmental and operational variables. Such a predictive model could aid in facilitating any future investigation into price changes.

6. References

1. NYC Ferry, Cruising towards another record-setting year (March 2023) Retrieved from: https://www.ferry.nyc/blog/cruising-towards-another-record-setting-year
2. City of New York, (July 2022) Retrieved from: https://www.nyc.gov/office-of-the-mayor/news/505-22/mayor-adams-phase-one-nyc-ferry-forward-vision-more-equitable-accessible-#/0
3. Wikipedia (August 2021) Retrieved from: https://en.wikipedia.org/wiki/NYC_Ferry
4. NYC Economic Development Corporation (2010) Retrieved from: (https://sallan.org/pdf-docs/EDC_FerryFeasibility.pdf
5. New York Times (October 2011) Retrieved from: https://www.nytimes.com/2011/10/17/nyregion/east-river-ferry-service-exceeds-expectations.html
6. Gothamist (July 2022) Retrieved from: https://gothamist.com/news/nyc-ferry-service-to-launch-discounted-fare-program-in-september-adams-says
7. New York Times (June 2011) Retrieved from: https://www.nytimes.com/2011/06/14/nyregion/east-river-ferry-service-begins-with-7-stops.html
8. Economic Development Corporation. (2021). NYC Ferry: Economic Development Corporation. Retrieved from https://council.nyc.gov/budget/wp-content/uploads/sites/54/2021/02/Economic-Development-Corporation-NYC-Ferry-Report-2-of-2.pdf
9. SimpliLearn (April 2023) Retrieved from: https://www.simplilearn.com/tutorials/power-bi-tutorial/what-is-power-bi

7. Appendix

Work	Avikal	Irtiza	Luke	Michelle	Rizwan	Sam
Introduction & Objectives				✓		
Hypothesis formulation	✓	✓	✓	✓	✓	✓
Data Extraction - ferries				✓		
Data Extraction - subway	✓			✓		
Data Labeling & Annotation		✓	✓	✓	✓	✓
Data Merging & Correlation	✓	✓	✓	✓	✓	✓
Data discrepancy analysis & cleaning			✓			
PowerBI - GIS mapping of routes					✓	✓
PowerBI - ferries dashboard					✓	
PowerBI - subway dashboard					✓	
PowerBI - limitations demonstration						✓
Python - ferries overall viz	✓		✓			
Python - ferries route wise viz	✓					
Python - COVID 19 analysis	✓	✓				
Python - ferries ridership time series forecasting (ML)	✓					
Python - ferries revenue calc.	✓	✓				
Python - subway viz & validation	✓					
Results				✓		
Conclusion			✓	✓		
Report formatting				✓		
Slides - PPT		✓	✓			

YouTube link to presentation recording:
https://www.youtube.com/watch?v=YNMmc-TYLGo&ab_channel=IrtizaShafi