A Better City: Commuter Rail Analysis

By
Bhuvan Shivalingaiah Gowda (bhuvansg@bu.edu)
Neel Gangrade (neelg@bu.edu)
Ningxin Guan(nxguan@bu.edu)
Sumanth Hosdurg Kamath (sumanth@bu.edu)
Yuanchen Yin (yinthyg@bu.edu)

Analysis Paper
DS701 Fall 2024 at Boston University
BU Spark!

Table of Contents

Introduction

The project aims to build a comprehensive data repository and analysis framework for MBTA (Massachusetts Bay Transportation Authority) schedules, focusing on both current and historical data. This initiative will support policy-making and decarbonization strategies by providing insights into transit schedules, timing, and operational changes over time.

The project begins by aggregating data from the MBTA API and creating a database. The database will be uploaded to a cloud service for accessibility and scalability. Additionally, scripts will be developed to automatically fetch the most current data and retrieve archived data for analysis.

We will be focusing on answering the questions listed below:

- How have schedules shifted over time?
- What is the net number of trains operating per line?
- How do travel times vary across different schedules?
- How do station changes impact overall transit schedules?
- A breakdown of trains by time of day (early morning, peak morning, midday, afternoon peak, evening, late night)
- Identify/analyze the number of express trains per line
- Do we see incidents of a station changing fare zones?
- If cost info is included in the active GTFS feed & archives, can we compare the changes in fare cost over time?

We also plan to develop a streamlined pipeline to extract raw data from the MBTA archives and load it into BigQuery for seamless storage and analysis. Additionally, we will create a separate pipeline to upload the answers to each question directly into BigQuery, ensuring efficient organization and accessibility of insights.

Client:

A Better City represents a multi-sector group of nearly 130 business leaders united around a common goal: to enhance the Greater Boston region's economic health, competitiveness, equitable growth, sustainability, and quality of life for all communities. By amplifying the voice of the business community through collaboration and consensus-building, A Better City develops solutions and influences policy in three critical areas: 1) transportation and infrastructure, 2) land use and development, and 3) energy and the environment. A Better City is committed to building an equitable and inclusive future for the region that benefits and uplifts residents, workers, and businesses in Greater Boston.

Dataset Description

The dataset originated from the MBTA archives and is scraped for the latest version for each season and year. The scraped URL contains 31 tables containing information on routes, lines, schedules, trips, fares, etc. which helps analyze each project question. This raw dataset is also uploaded to BigQuery for feasible preprocessing and analysis.

Data Analysis

Q1) A breakdown of trains by time of day (early morning, peak morning, midday, afternoon peak, evening, late night)

This question aims to find the changes in trip number in each commuter rail line by time of day from 2019 to 2024. We define different times of the day by the following time period in the *time_of_day(time_str)* function: **AM Peak**: 6:00 AM - 9:00 AM, **PM Peak**: 5:00 PM - 7:00 PM, **Midday Base**: 9:00 AM - 4:00 PM, **Early AM**: 5:00 AM - 6:00 AM, **Late Evening**: 8:00 PM - 11:00 PM, **Night**: 11:00 PM - 5:00 AM, **Sunrise**: 5:00 AM - 6:00 AM, and **Midday School**: 12:00 PM - 2:00 PM. We break down this question into two parts:

- Changes in the number of trips by time of day for each CR line from 2019 to 2024.
- Changes in number of trips by time of day for each CR line across weekdays (Monday to Friday) and weekends (Saturday and Sunday).

1. **Changes in the number of trips by year**

In this problem, we use the lines.txt, trips.txt, calendar.txt, and stop_times.txt files to get the information of each trip, including the arrival time, route id, and calendar by weekday and weekend. We grouped the trip data from 2019 to 2024 by line, year, weekday, and time period. The number of trips per line was calculated for each year and time period. Trip counts were also broken down by weekdays (Monday to Friday) and weekends (Saturday and Sunday), observing any patterns in trip distribution.

We created the *get_weekday_services( )* function to filter trips that operate on weekdays and weekends(Monday to Sunday). By checking the service dates of each trip, the function determines which trips are running on weekdays. The output *trip_count* stores this information of time_period, route id, trip count, year, and season. We used trip_counts in the following plot_annual_trip_counts function to plot the annual trip count trends for different time periods. It aggregates the trip counts for each time period across different years and generates a line graph for each CR line. This helps analyze how trip counts change over the years for a particular CR line and shows the trends in transportation patterns.

2. **Changes in the number of trips by calendar**

In this section, we first merged the trip and calendar data by stop times, trips, and calendar. The get_mergeddata function will iterate through each season in each year and merge stop times and trips by trip id. The returned data frame contains information of stop times, trips, and calendar. We then get the merged calendar from 19 to 20 and go through each commuter rail line and analyze the train operation on weekdays and weekends. The output data frame contains the number of trips by time period for every weekday, classified by commuter rail line.

3. **Visualization Analysis**

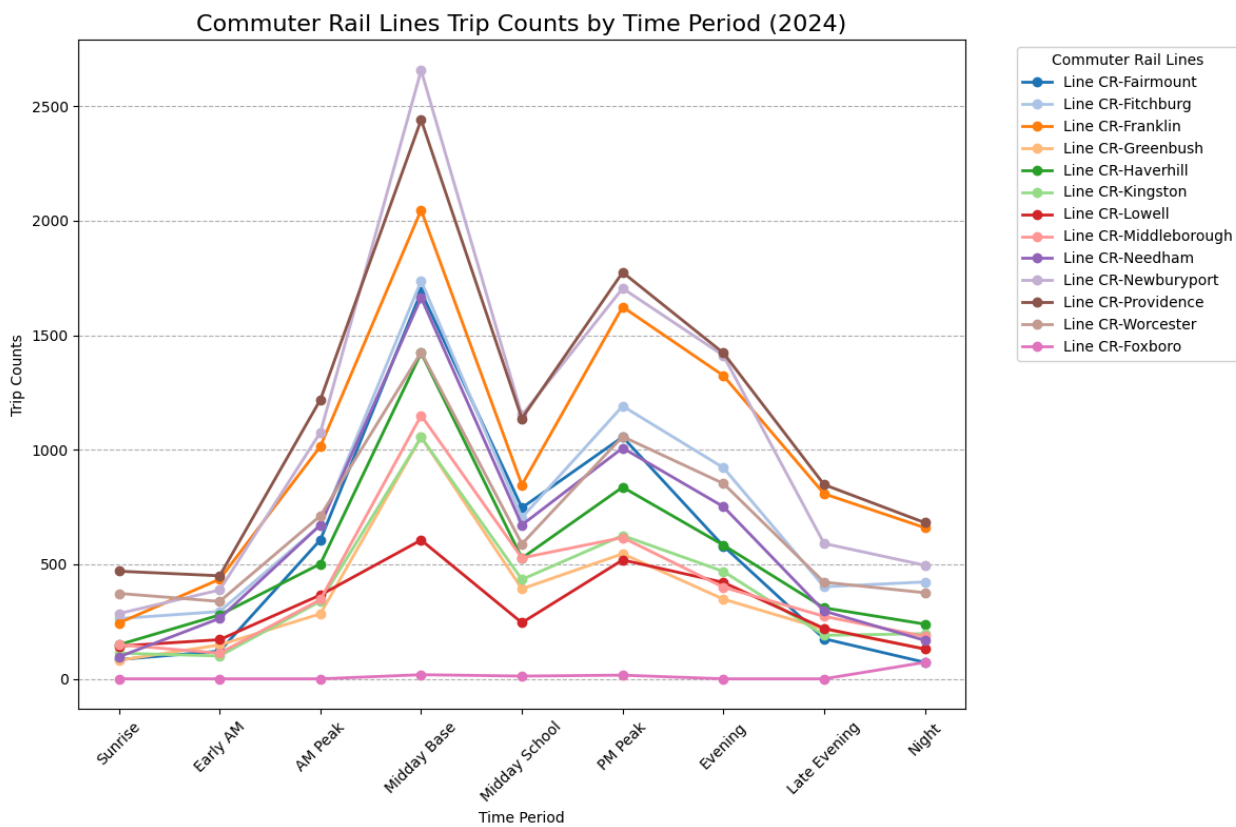


Figure1: The graph shows the number of trips by different time periods from 2019 to 2024 for all the commuter rail.

From the Trip Counts by Time Period graph (Figure 1), it is evident that for every commuter rail line, the Midday Base period usually has the highest number of trips, followed by the PM peak. This indicates a strong ridership demand during the morning and evening rush hour. In contrast, there are significantly fewer trips during sunrise and late night. This pattern is aligned with the typical daily travel behavior.

Figure 1 indicates a noticeable increase in trip number from 2019 to 2024. The purple number represents 2019 and the yellow line represents the trip number from 2024. As shown in the graph, we can see a gradual increase of trip number from 2019 to 2024. These changes may be caused by the endemic, which has a serious impact on the MBTA schedule during 2019 and 2020, as most of the people chose to work remotely from home. As conditions improved, the ridership demand slowly increased, and gradually grew back to normal in 2024. Graphs for all the commuter rail lines from 2019 to 2024 can be found in Appendix A.

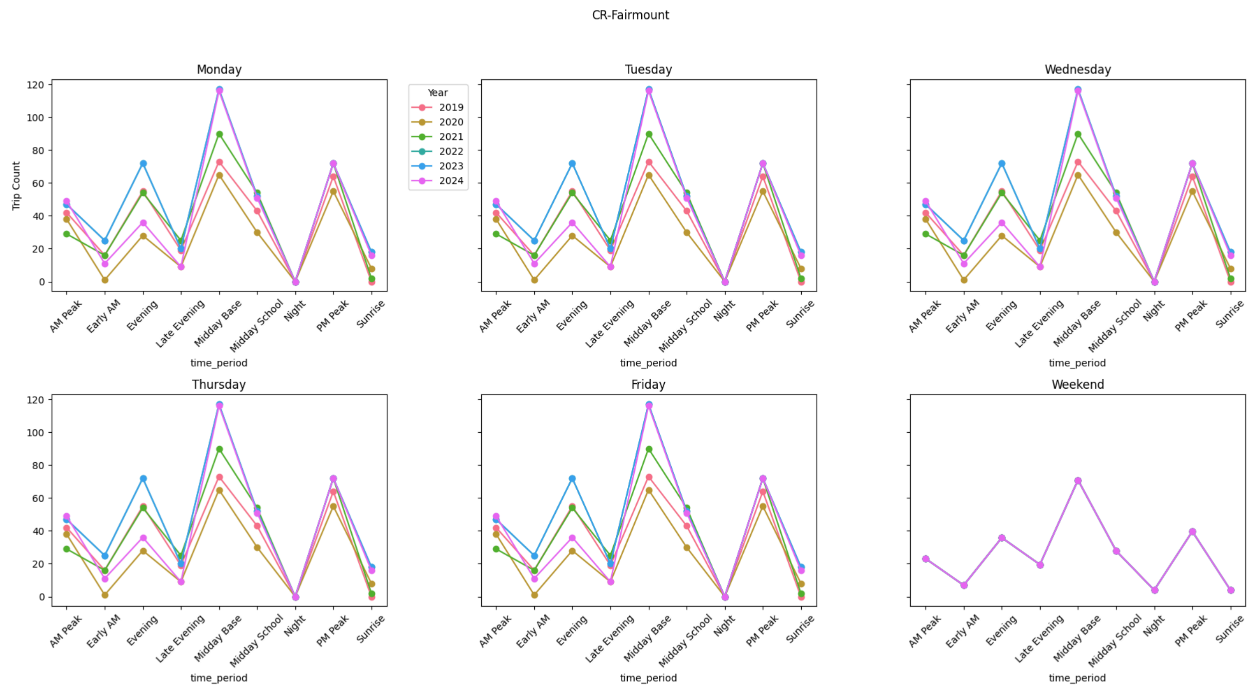


Figure2: The graph shows the number of trips by different time periods from 2019 to 2024 for Fairmount line by weekday.

Figure 2 is an example of the Fairmoung line from 2019 to 2024. This graph shows the number of trips by weekday, with x axis representing the time period and y axis representing the number of trips. The graph indicates that the weekday schedules (monday to friday) are very consistent, and the weekend has less number of trips. There are no trips on weekends from 2019 to 2023, which could be a result of the pandemic. Graphs for other commuter rail lines can be found in Appendix B.

There is an exception to this pattern. In the graph (see Appendix A) of the CR Lowell line, Worcester line, and Fitchburg line, 2019 has the most number of trips. There are several reasons that may contribute to this change. In 2020, some commuter rail lines such as the Lowell line underwent infrastructure improvement, which could result in decreases of trip numbers.

Q2) The number of express train for each line

This problem asks for the number of trains on each commuterial line. In MBTA GTFS, each train is presented by a unique trip id, therefore we can count the number of unique trips.

In this question, we count the number of stops for each unique trip and merge the trip id and the route id. For each route, we count for the max number of stops. If the number of stops for a trip is less than the maximum number of stops for that route, we assume that this trip has skipped stops, and therefore it is an express train.

Figure 3: The output graph shows the number of express trips on each commuter rail line from 2021 to 2024. The y-axis represents the trip counts and the x-axis represents the years.

As shown in the graph, 2022 has the highest number of express trains compared to other years. The line Providence, Kingston, Worcester, Needham, Fitchburg all have a decrease in number of express trips from 2022 to 2023. Line Foxboro has the lowest number of express trains and has a slow increase from 2021 to 2024.

Several factors may contribute to this schedule change. The increased number of express trains in 2022 could represent the post-pandemic recovery in ridership demand. Due to the impact of the pandemic from 2019 to 2021, the majority of people have to stay at home and work remotely. As the pandemic came under control, many people began to travel extensively after it reopened, leading to a ridership increase in 2022.

Q3) Do we see incidents of a station changing fare zones?

This question asks us to show any stations changing their fare zone in the past 4 years (2021-2024). A station is a type of location or physical structure that contains one or more platforms. We used the stops table (stops.txt) to find the answer to this question. The first filtering step involves filtering out the rows which do not correspond to Commuter Rails. This is done by using the *zone_id* column which represents the fare zone. Only those rows are kept whose *zone_id* value starts with a 'CR'. Then the next filtering is done based on the location type. Since we only need to analyse stations that change fare zones, we filter the dataset to only have those columns whose *location_type* = 1 since that corresponds to stations. After that we combine the data for all the seasons in a year into individual datasets. The individual datasets for each year are then merged together into one single dataset. Now, we just compare the *zone_id* for each station each year and shortlist the stations whose fare zones change.

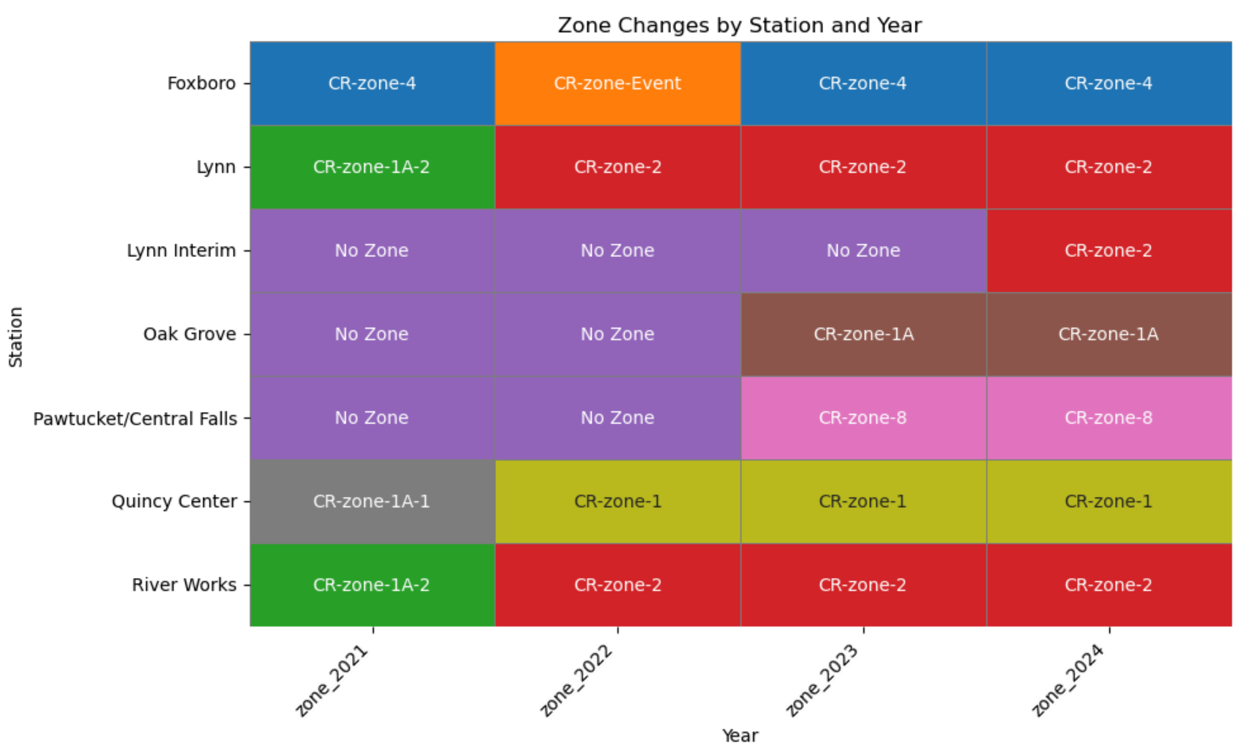


Figure 4: This graph shows the zone changes by station and year.

This visualization shows the stations on the y-axis and the years on the x-axis. We can clearly see that after 2022, all stations have a consistency in their fare zones. The reason behind that is the end of the COVID-19 pandemic as we moved from 2021 to 2022. There is also a special fare zone called 'CR-zone-Event' which is only introduced at the times of major events in Boston to redirect the crowd going to the event and ensure a smoother service for everyone in the city. An example of one of these major events is the Taylor Swift concert which was held at the Gillette Stadium.

Zone	One-Way	Reduced One-way	Monthly Pass	Monthly mTicket
Zone 1A	$2.40	$1.10	$90.00	$80.00
Zone 1	$6.50	$3.25	$214.00	$204.00
Zone 2	$7.00	$3.50	$232.00	$222.00
Zone 3	$8.00	$4.00	$261.00	$251.00
Zone 4	$8.75	$4.25	$281.00	$271.00
Zone 5	$9.75	$4.75	$311.00	$301.00
Zone 6	$10.50	$5.25	$340.00	$330.00
Zone 7	$11.00	$5.50	$360.00	$350.00
Zone 8	$12.25	$6.00	$388.00	$378.00
Zone 9	$12.75	$6.25	$406.00	$396.00
Zone 10	$13.25	$6.50	$426.00	$416.00

Table 1: This table shows the fare zone from zone 1A to zone 10

The Commuter Rail is divided into 11 fare zones, from Zone 1A to Zone 10. Stations in metro Boston are in Zone 1A, and every Zone beyond that indicates each station's distance from Boston. Your Commuter Rail fare depends on which Zones your boarding and exiting stations are located in. If your trip begins or ends at a station located in Zone 1A, your fare is based on the location of the other station you are traveling to or from. For example, a one-way trip from Concord (Zone 5) to North Station (Zone 1A) is a Zone 5 fare. You will also pay a Zone 5 fare for a one-way trip from North Station to Concord.

Q4) Can we compare the changes in fare cost over time?

This study explores the changes in fare costs for commuter rail services over time using data from the MBTA for the years 2023 and 2024. The primary objective is to identify trends and variations in fare costs across different seasons and routes. By leveraging datasets containing information on fare products, fare leg rules, and routes, this analysis aims to compare fare changes across the available time frame and understand the impact of these changes on commuter rail services.

The preprocessing began by consolidating data from all seasons (spring, summer, fall, and winter) for the years 2023 and 2024 into a single data frame. The fare_products.txt, fare_leg_rules.txt, and routes.txt files were used to extract and filter data relevant to commuter rail services. The filtering process involved using the route_desc column in the routes dataset, selecting only records where route_desc == 'commuter rail'.

To simplify the analysis, the focus was narrowed to cash fare payment data, as cash, credit/debit, and mTicket payment methods all had identical fare amounts. The dataset was further cleaned by removing null values and duplicates to ensure accuracy.

The fare calculation involved isolating trips originating within Boston (from_area_id = 'area_commuter_rail_zone_1a') and determining the fare to the last station on each commuter rail line. Each line's final station was identified using to_area_id values specific to its zone, as detailed below:

***CapeFlyer*: area_cf_zone_hyannis**
***Fairmount*: area_commuter_rail_zone_2**
***Fitchburg*: area_commuter_rail_zone_8**
***Worcester*: area_commuter_rail_zone_8**
***Middleborough*: area_commuter_rail_zone_8**
***Needham*: area_commuter_rail_zone_2**
***Newburyport*: area_commuter_rail_zone_8**
***Providence*: area_commuter_rail_zone_10**
***Greenbush*: area_commuter_rail_zone_6**
***Haverhill*: area_commuter_rail_zone_7**
***Kingston*: area_commuter_rail_zone_8**
***Franklin*: area_commuter_rail_zone_6**
***Lowell*: area_commuter_rail_zone_6**

The fare between the respective zones was calculated by linking the origin and destination zones for each line, laying the groundwork for analyzing fare changes across seasons and years.

The final dataset compared fare costs over time by examining average fares for each commuter rail line across seasons and years. The results were stored in a structured data frame containing columns for line_id, season_year, and the calculated fare values, facilitating a comprehensive analysis of fare cost trends. Since the data is available only for 2 years, there are no fare changes across periods. So, the table in the appendix below presents the fare to travel each route.

Q5/7) How have schedules shifted over time and how do travel times vary across different schedules?

This study examines how commuter rail schedules have evolved over time and how travel times vary among plans from 2019 to 2024. The study focuses on finding patterns and trends in scheduling and travel times using datasets that contain information on calendars, trips, stop times, and routes. Understanding how schedules have changed over the years and seasons and how these modifications have affected travel times on various routes is the main goal.

Cleaning and filtering the datasets to concentrate on commuter rail data was the first stage of the investigation. To save only entries pertinent to commuter rail services, the routes dataset was filtered using the route_desc column. To match schedule data with commuter rail routes, the calendar, trips, and stop_times records were then combined. Invalid or missing data were found and eliminated from important columns like departure_time and arrival_time to guarantee correctness. A comparative study over the time period under examination was made possible by structuring the resultant dataset into a nested format by year and season. The groundwork for identifying patterns in schedule changes and travel time variances was established by this preprocessing.

To analyze the variation in travel times across different schedules, the dataset containing columns such as route_id, trip_id, arrival_time, and departure_time was utilized. Each route_id corresponds to a specific route, which comprises multiple trips identified by unique trip_id values. For each trip, there are multiple arrival_time and departure_time entries corresponding to different stops.

To calculate the duration of a specific trip, the first recorded arrival_time and the last recorded departure_time were identified, and their difference was computed to obtain the total duration of the trip. This process was repeated for all trips associated with a given route, resulting in a collection of trip durations for each route. The average trip duration for a given route was then calculated by averaging the durations of all trips associated with that route.

This analysis was carried out for every route in the dataset across multiple seasons and years. The results were stored in a new DataFrame containing columns for route_id, season_year, and the average trip duration.

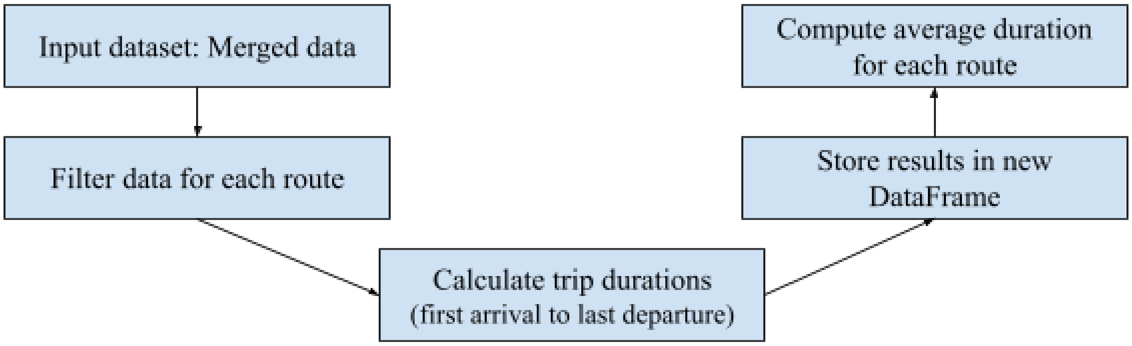


Figure 5: Analysis flowchart

```
      route_name season-year  average_trip_duration_minutes
0    CR-Fairmount    Fall2019                      28.577778
1    CR-Fairmount    Fall2020                      28.857143
2    CR-Fairmount    Fall2021                      28.989474
3    CR-Fairmount    Fall2022                      28.989474
4    CR-Fairmount    Fall2023                      28.444444
..            ...         ...                            ...
303     CapeFlyer  Spring2023                     114.125000
304     CapeFlyer  Spring2024                     114.125000
305     CapeFlyer  Summer2023                     114.125000
306     CapeFlyer  Summer2024                     114.125000
307     CapeFlyer  Winter2024                       7.000000
```

Figure 6: Final DataFrame showing the routes and its respective trip duration through season-years

Figure 6 is the final DataFrame that we used to plot the line graph that shows us how the duration of the trips, for a given route, changes through the season-year.

Figure 7 visualizes the **average trip duration (in minutes)** for all commuter rail routes over time, using chronological seasons as the x-axis. Each line represents a unique rail route, with its trajectory highlighting changes in average trip duration across different seasons from **Fall 2019 to Summer 2024**. Notable observations include relatively consistent durations for most routes, indicating stable scheduling, while certain routes exhibit significant changes. For instance, **CR-Foxboro** shows a notable **decrease in duration**, which could indicate operational improvements or route-specific optimizations. Conversely, the **CapeFlyer** line stands out as erratic, with significant drops and spikes, especially toward the end of the timeline, suggesting inconsistencies in its scheduling or data anomalies.

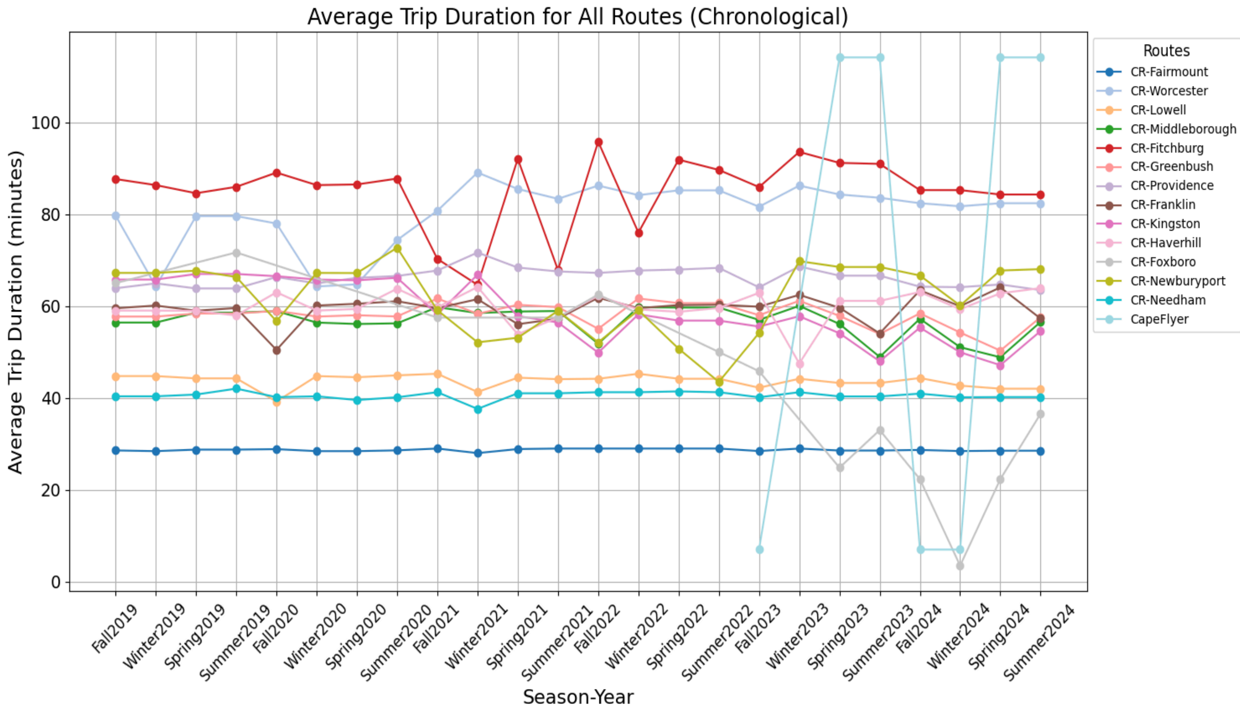


Figure 7: Visualization of 14 commuter routes throughout each season from Fall 2019 to Summer 2024

This plot helps us understand and visualize how the durations for all the 14 routes have changed over season-year.

Q6) What is the net number of trains operating per line?

We analyzed data from the trips.txt dataset, covering 2019 to 2024. The focus was on commuter rail services. We filtered the data to include only commuter rail trips. To get the answer to this question more accurately, each unique Trip ID was counted once. This method ensured no duplicates were included. The approach provided an accurate representation of train operations. It also highlighted the total number of distinct trips on each line. The results give a clear understanding of train usage across the network. This analysis offers valuable insights for planning and resource allocation. It supports future improvements in commuter rail services.

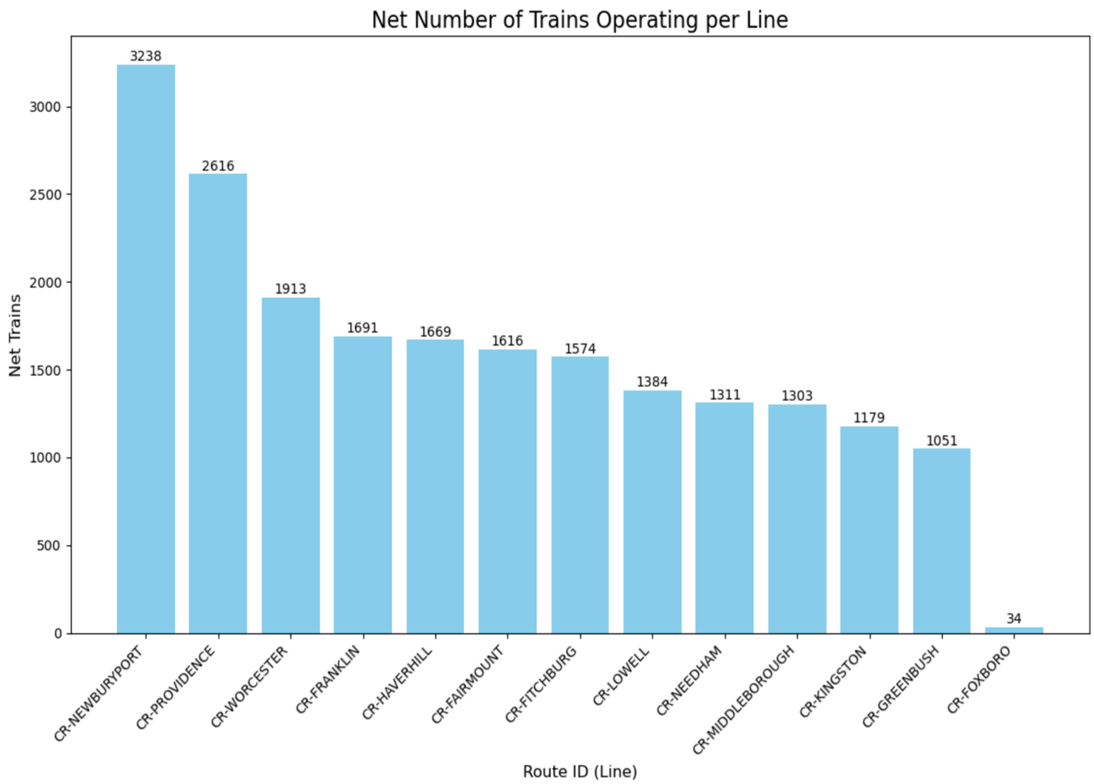


Figure 8: This chart shows the net number of trains operating on each commuter rail line. (2019-2024)

This chart shows the net number of trains operating on each commuter rail line. The CR-Newburyport line has the most trains, with 3,238 trips. The CR-Providence line follows, operating 2,616 trains. The CR-Worcester line has 1,913 trains, while the CR-Franklin line has 1,691. These numbers highlight their importance in the network. In contrast, the CR-Foxboro line has only 34 trips. This is because it is a special line that operates exclusively during events. The data shows clear differences in train operations across the lines. These differences likely reflect variations in passenger demand, service priorities, and special use cases like the CR-Foxboro line.

In 2019, the growth rate was 0% because it was the starting year. In 2020, the number of trains grew by 22.6%, marking a significant increase. However, in 2021, growth dropped sharply to -15.9%, indicating a decline in operations. The trend reversed in 2022 with an 18.7% increase, showing signs of recovery. In 2023, growth declined again, reaching -8.2%. By 2024, growth surged to 25.8%, the highest in the period. These shifts highlight fluctuations in train operations, likely influenced by external factors such as COVID-19.

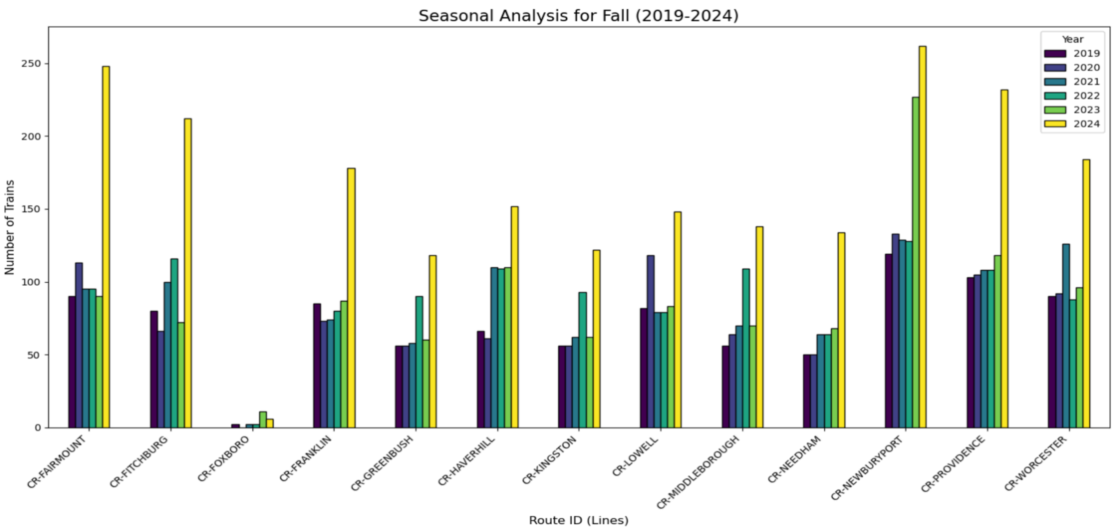


Figure 9: This bar chart shows the seasonal analysis for fall from 2019 to 2024. We will contain the visualizations for the other three seasons in appendix C.

This graph highlights the number of trains operating on each commuter rail line. Most lines show a steady increase in train numbers over the years. The highest values are consistently in 2024. The CR-Newburyport and CR-Providence lines have the most growth, reflecting increased demand or service expansion. The CR-Foxboro line remains low, as it only operates during events. Lines like CR-Fitchburg and CR-Haverhill show smaller but steady increases. The data reveals a positive trend in fall operations, with some lines growing faster than others. These changes likely result from evolving demand and adjustments in service.

The seasonal analysis graph and the year-over-year growth are closely linked. The seasonal graph shows an increase in train numbers during fall for most lines, especially in 2024. This aligns with the 25.8% growth shown in the year-over-year growth for the same year. Declines in certain years, like 2021 and 2023, also match the negative growth rates of -15.9% and -8.2%. These patterns show that changes in annual growth directly affect seasonal train operations. Together, the graphs highlight how yearly trends shape seasonal performance across the network.

BigQuery Pipeline:

As part of our automation process, we developed two streamlined pipelines to manage and analyze data from the MBTA archives effectively. These pipelines leverage BigQuery, a powerful data warehouse tool, to ensure seamless data integration and facilitate analysis.

1. Uploading Raw Data to BigQuery: The first pipeline focuses on uploading the raw data extracted from the GTFS archives into BigQuery. This data includes comprehensive transit information such as schedules and routes for multiple years. Using the BigQuery pipeline, we categorize the data by year, creating separate datasets for each. Within these datasets, all tables corresponding to different seasons are combined, ensuring an organized structure for future processing.
2. Uploading Cleaned and Analyzed Data to BigQuery: The second pipeline is dedicated to processing the raw data into meaningful insights by addressing specific analytical questions. After cleaning and analyzing the data, the results are uploaded to BigQuery under a dedicated dataset named *analysis_data*. Each analytical question, addressing key transit-related concerns such as fare changes, schedule shifts, or travel times, corresponds to a distinct table within the *analysis_data* dataset. This structure enables stakeholders to easily access and interpret the insights derived from the data.

The working of the two scripts are explained below:

- The *bigquery_pipeline.py* script can be used to upload all the raw data for each season and year from the MBTA archives to BigQuery. The command to do so is:
    - *python bigquery_pipeline.py --year_range 2019-2024 --project ds-better-city-commuter*
    - *--year_range* is used to specify the range of years for which you want the data to be uploaded.
    - *--project* specifies the project ID of your project on Google Cloud Platform.
- The *bigquery_cleaned_pipeline.py* script is used to upload the dataset corresponding to a specific base question to BigQuery under the analysis_data Dataset. The command to do so is:
    - *python3 bigquery_cleaned_pipeline.py --project ds-better-city-commuter --question_num q4*
    - The possible values of the argument --question_num are 'q1', 'q2', 'q3', 'q4', 'q5', 'q6', 'q7'. These values correspond to the particular questions in the following manner:
        - q1: A breakdown of trains by time of day (early morning, peak morning, midday, afternoon peak, evening, late night)
        - q2: Identify/analyze the number of express trains per line
        - q3: Do we see incidents of a station changing fare zones?
        - q4: If cost info is included in the active GTFS feed & archives, can we compare the changes in fare cost over time?
        - q5: How have schedules shifted over time?

        - q6: What is the net number of trains operating per line?
        - q7: How do travel times vary across different schedules?

Conclusion:

The project involves developing scripts to automate data fetching, enabling efficient updates and management of datasets. It also includes creating visualizations of MBTA schedule data to uncover insights and patterns. A BigQuery pipeline is implemented to ensure non-technical clients can easily access and interact with the data. Additionally, a detailed tutorial and transition document are provided to guide users in self-serving the data, ensuring smooth future analyses and knowledge transfer.

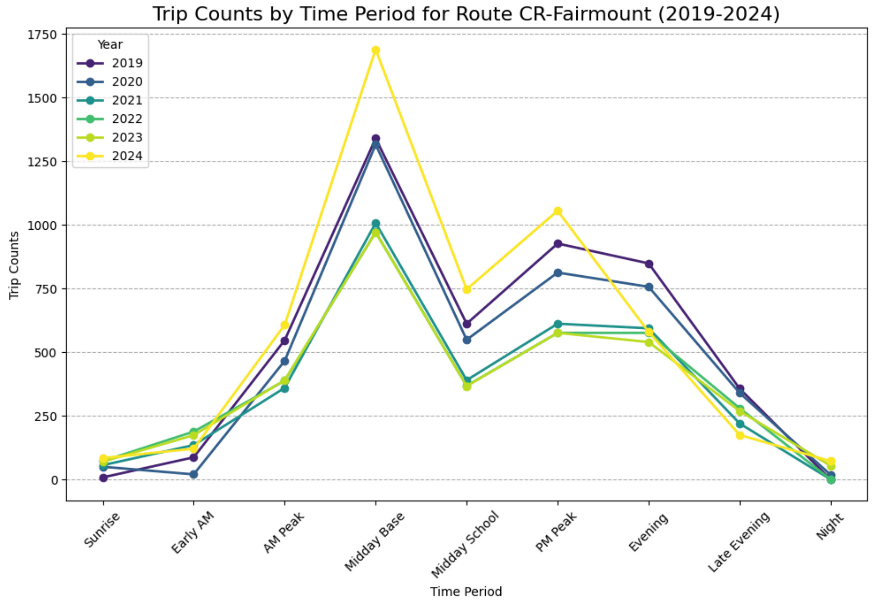
Contributions:

- Neel Gangrade:
    - Do we see incidents of a station changing fare zones? (Question 3)
    - BigQuery Pipeline
- Yuanchen Yin:
    - A breakdown of trains by time of day (early morning, peak morning, midday, afternoon peak, evening, late night) (Question 1)
    - Number of express train on each commuter rail line (Question 2)
- Ningxin Guan:
    - What is the net number of trains operating per line? (Question 6)
    - E-poster
- Sumanth Kamath:
    - How have schedules shifted over time and how do travel times vary across different schedules?
    - Preliminary Analysis Document Drafting
- Bhuvan Gowda:
    - Can we compare the changes in fare cost over time?(Question 4)
    - Report Drafting

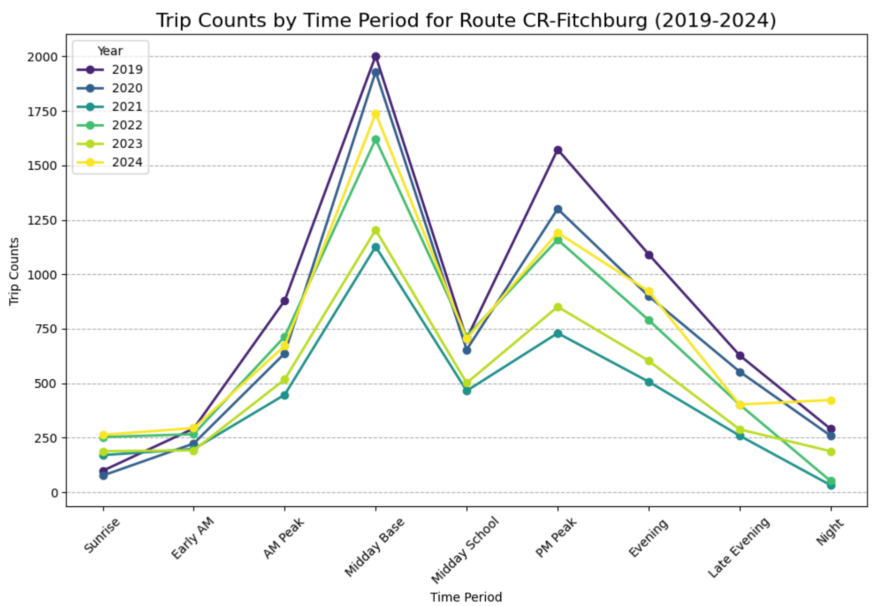
References

Massachusetts Bay Transportation Authority. *General Transit Feed Specification (GTFS) Data*. MBTA, [2019-2024], www.mbta.com/developers/gtfs.

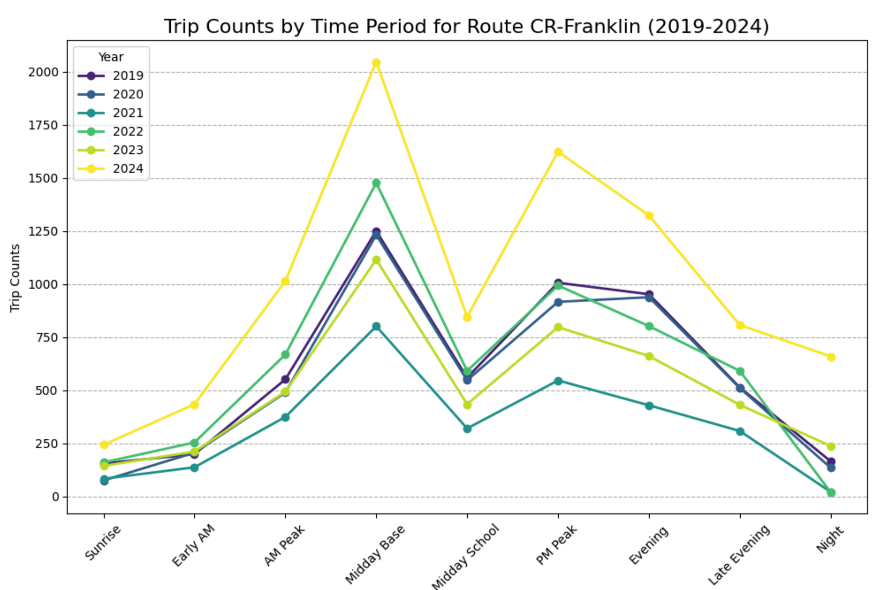
Appendix A



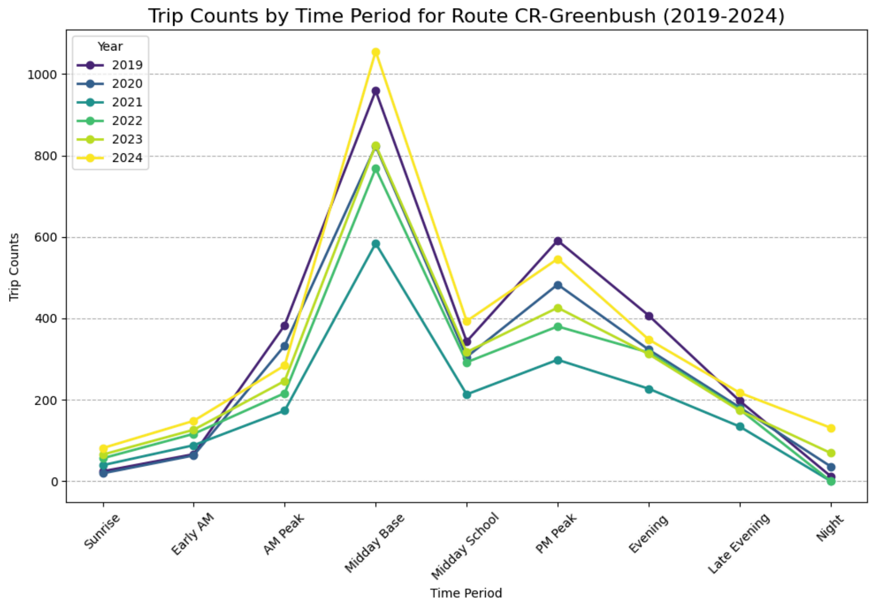
Trip Counts by Time Period for Route CR-Fairmount (2019-2024)
Year
2019
2020
2021
2022
2023
2024
Trip Counts
0
250
500
750
1000
1250
1500
1750
Sunrise
Early AM
AM Peak
Midday Base
Midday School
PM Peak
Evening
Late Evening
Night
Time Period

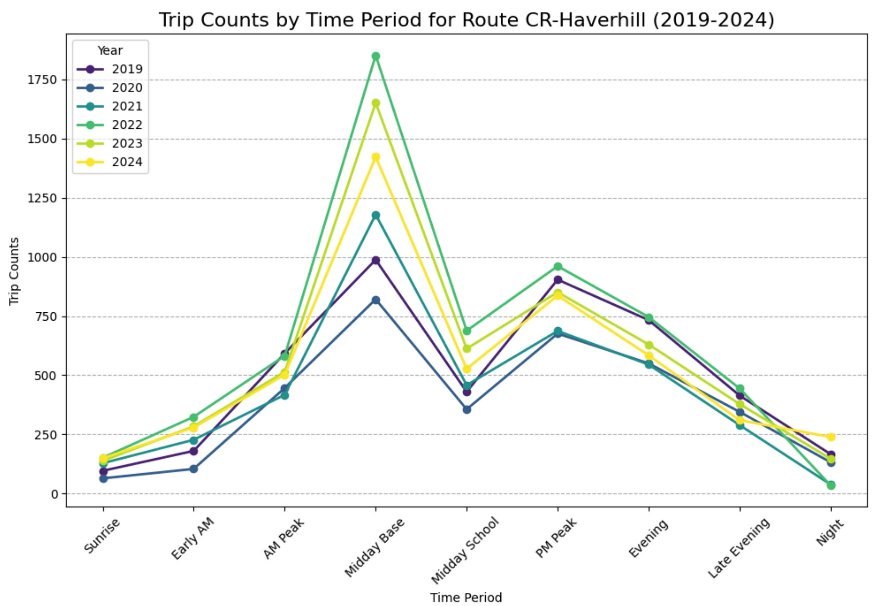


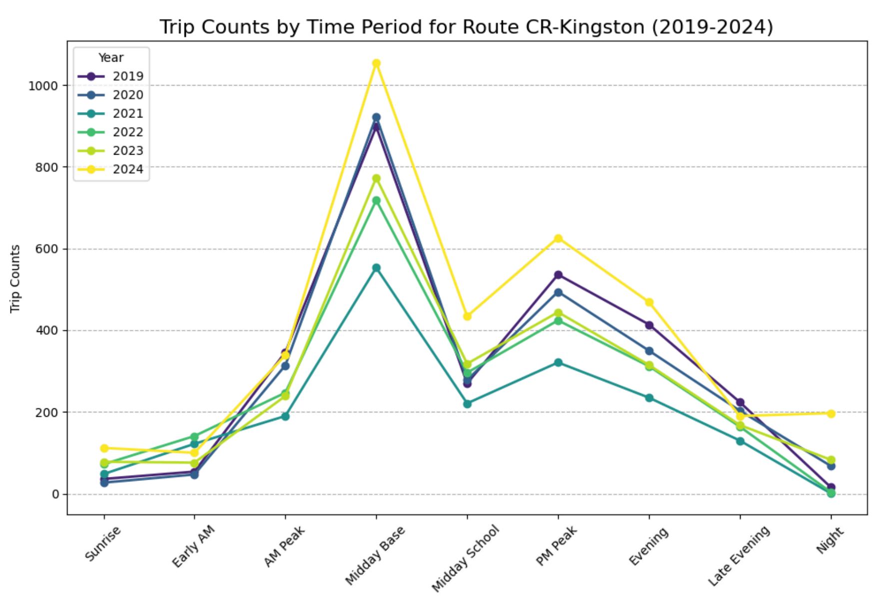
Trip Counts by Time Period for Route CR-Fitchburg (2019-2024)
Year
2019
2020
2021
2022
2023
2024
Trip Counts
0
250
500
750
1000
1250
1500
1750
2000
Sunrise
Early AM
AM Peak
Midday Base
Midday School
PM Peak
Evening
Late Evening
Night
Time Period

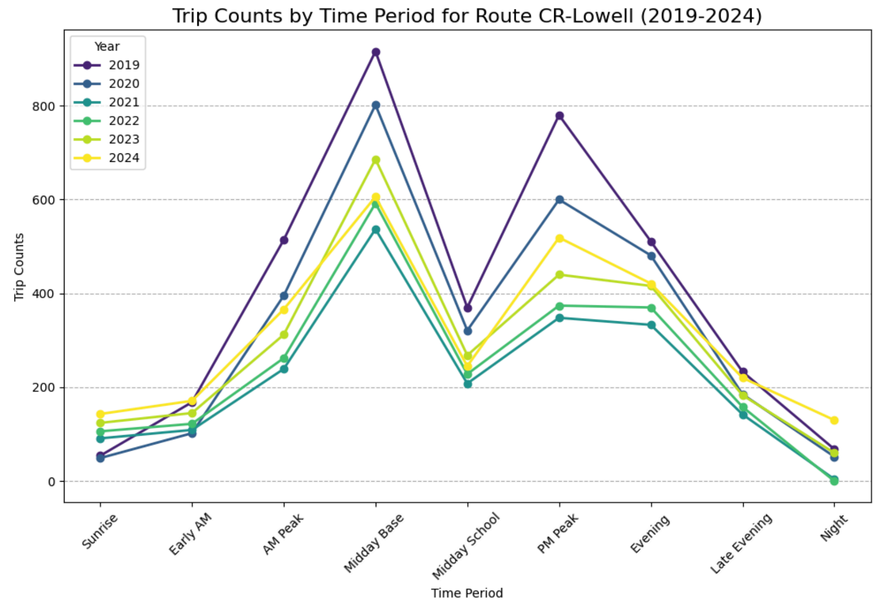


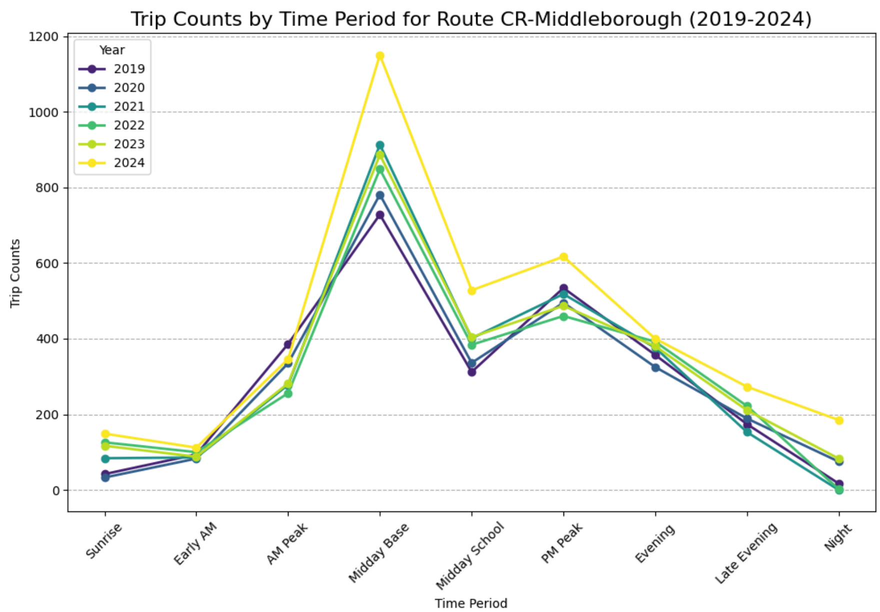
Trip Counts by Time Period for Route CR-Franklin (2019-2024)
Year
2019
2020
2021
2022
2023
2024
Trip Counts
0
250
500
750
1000
1250
1500
1750
2000
Sunrise
Early AM
AM Peak
Midday Base
Midday School
PM Peak
Evening
Late Evening
Night

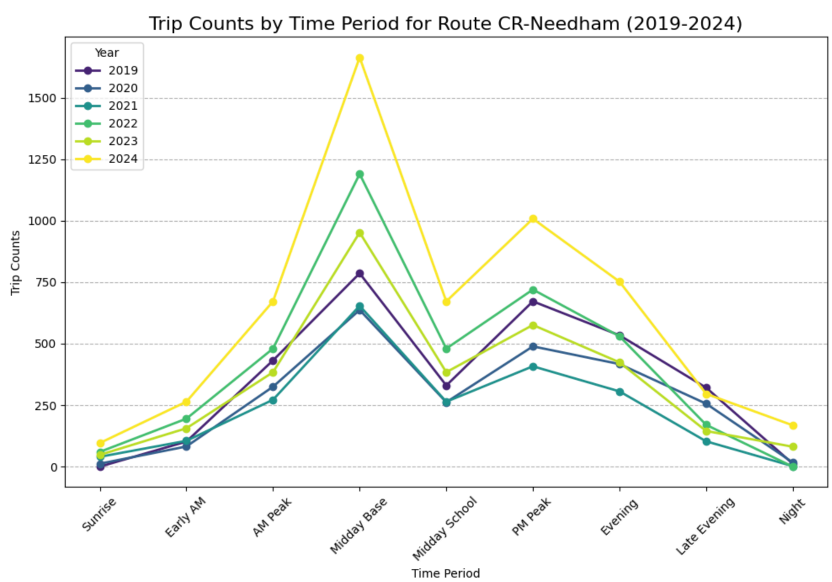

Trip Counts by Time Period for Route CR-Greenbush (2019-2024)
Year
2019
2020
2021
2022
2023
2024
Trip Counts
0
200
400
600
800
1000
Sunrise
Early AM
AM Peak
Midday Base
Midday School
PM Peak
Evening
Late Evening
Night
Time Period

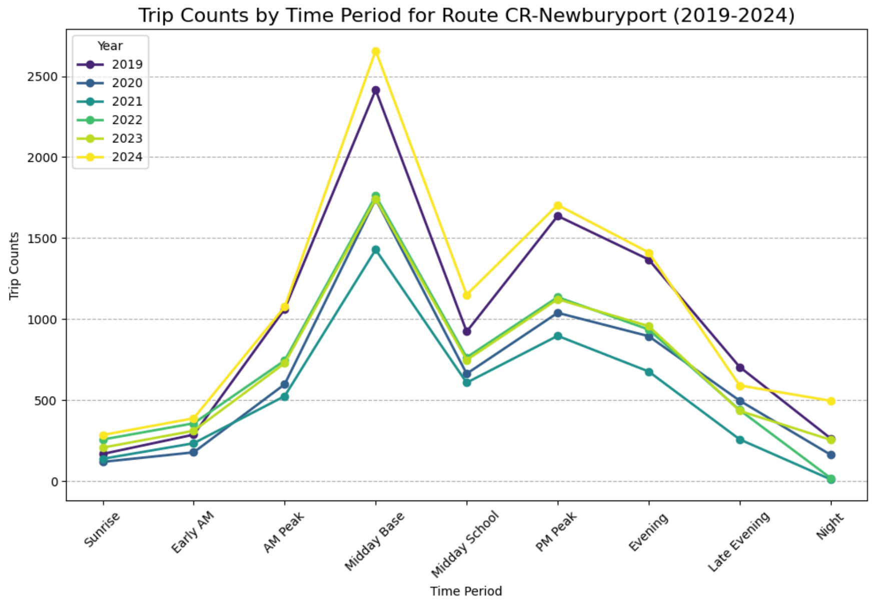

Trip Counts by Time Period for Route CR-Haverhill (2019-2024)
Year
2019
2020
2021
2022
2023
2024
Trip Counts
0
250
500
750
1000
1250
1500
1750
Sunrise
Early AM
AM Peak
Midday Base
Midday School
PM Peak
Evening
Late Evening
Night
Time Period

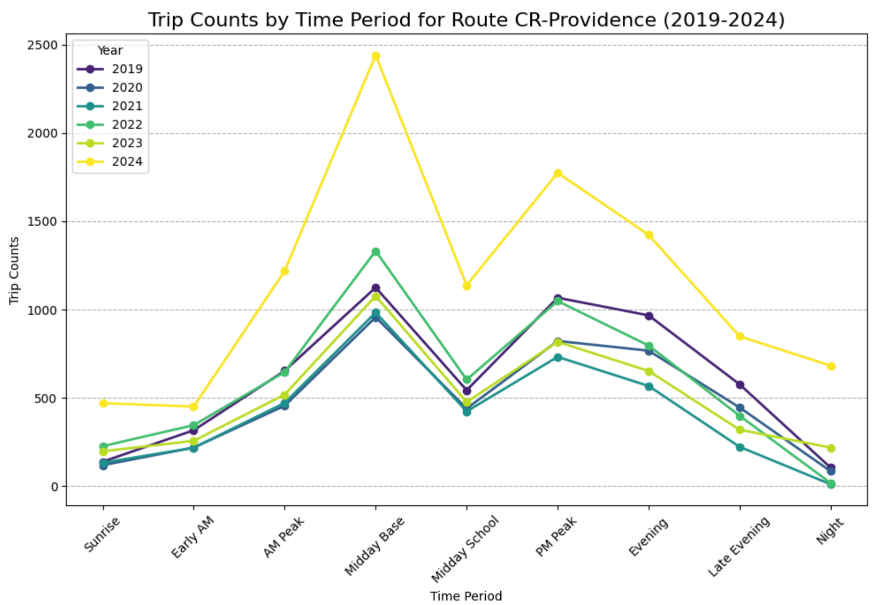

Trip Counts by Time Period for Route CR-Kingston (2019-2024)
Year
2019
2020
2021
2022
2023
2024
Trip Counts
0
200
400
600
800
1000
Sunrise
Early AM
AM Peak
Midday Base
Midday School
PM Peak
Evening
Late Evening
Night


Trip Counts by Time Period for Route CR-Lowell (2019-2024)
Year
2019
2020
2021
2022
2023
2024
Trip Counts
Time Period
Sunrise
Early AM
AM Peak
Midday Base
Midday School
PM Peak
Evening
Late Evening
Night

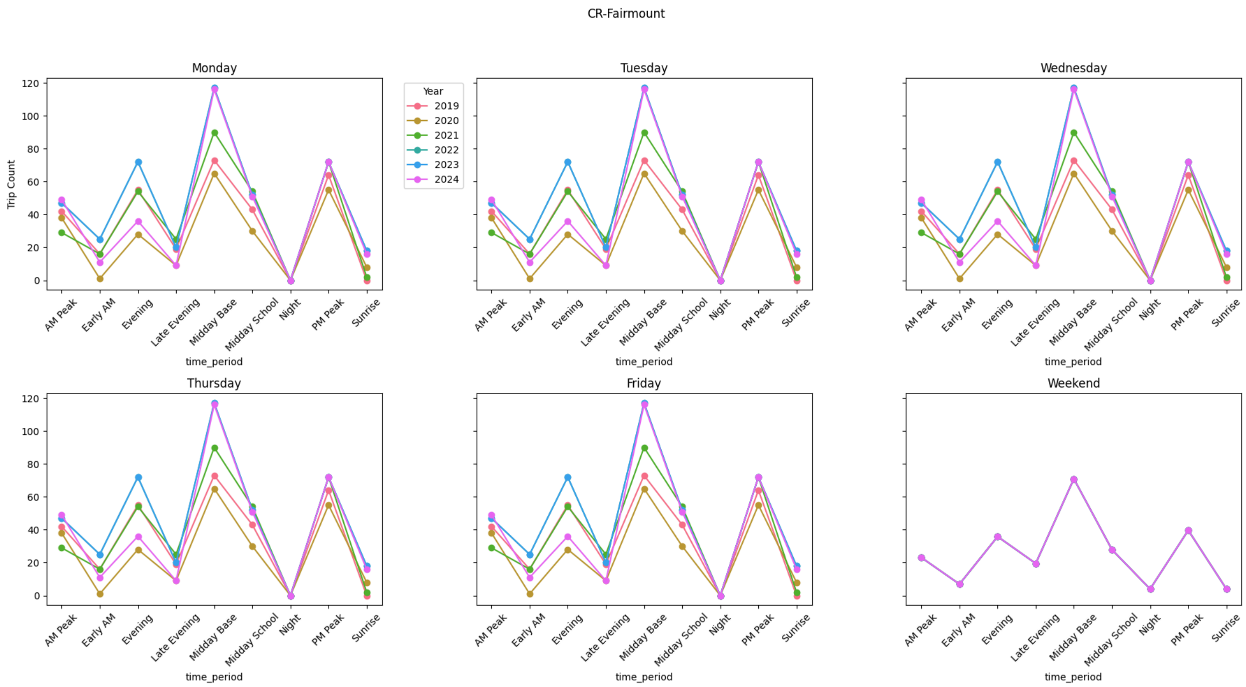

Trip Counts by Time Period for Route CR-Middleborough (2019-2024)
Year
2019
2020
2021
2022
2023
2024
Trip Counts
Time Period
Sunrise
Early AM
AM Peak
Midday Base
Midday School
PM Peak
Evening
Late Evening
Night

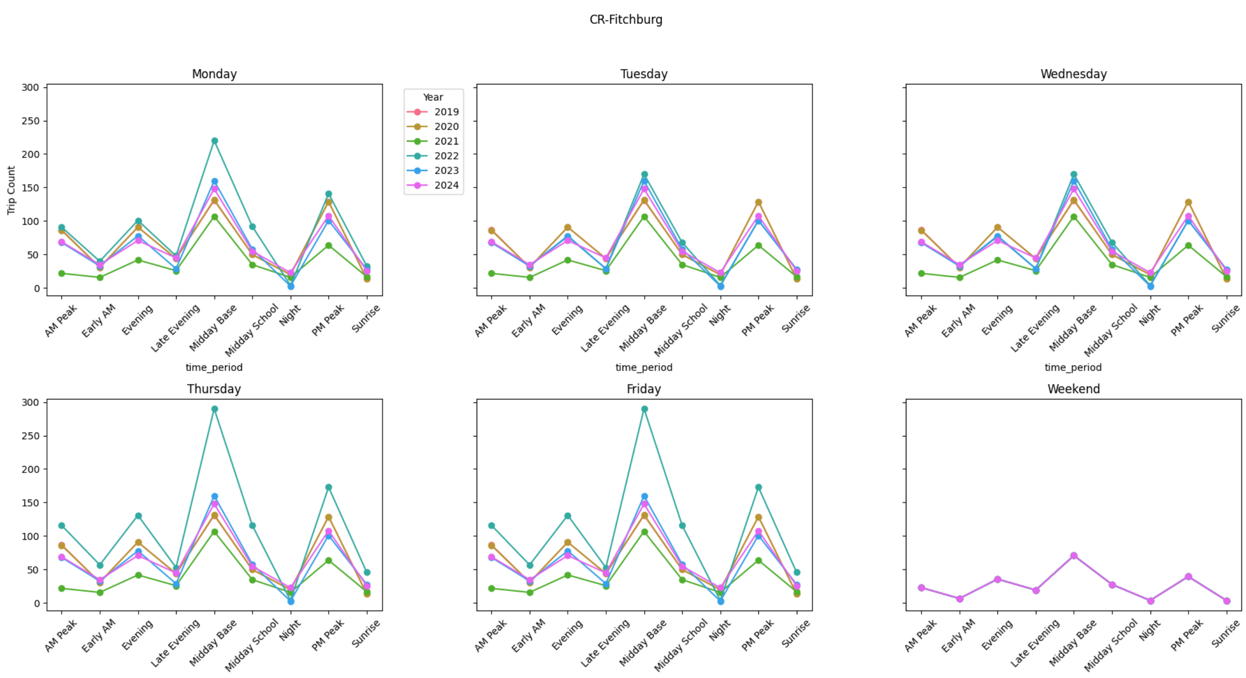

Trip Counts by Time Period for Route CR-Needham (2019-2024)
Year
2019
2020
2021
2022
2023
2024
Trip Counts
Time Period
Sunrise
Early AM
AM Peak
Midday Base
Midday School
PM Peak
Evening
Late Evening
Night

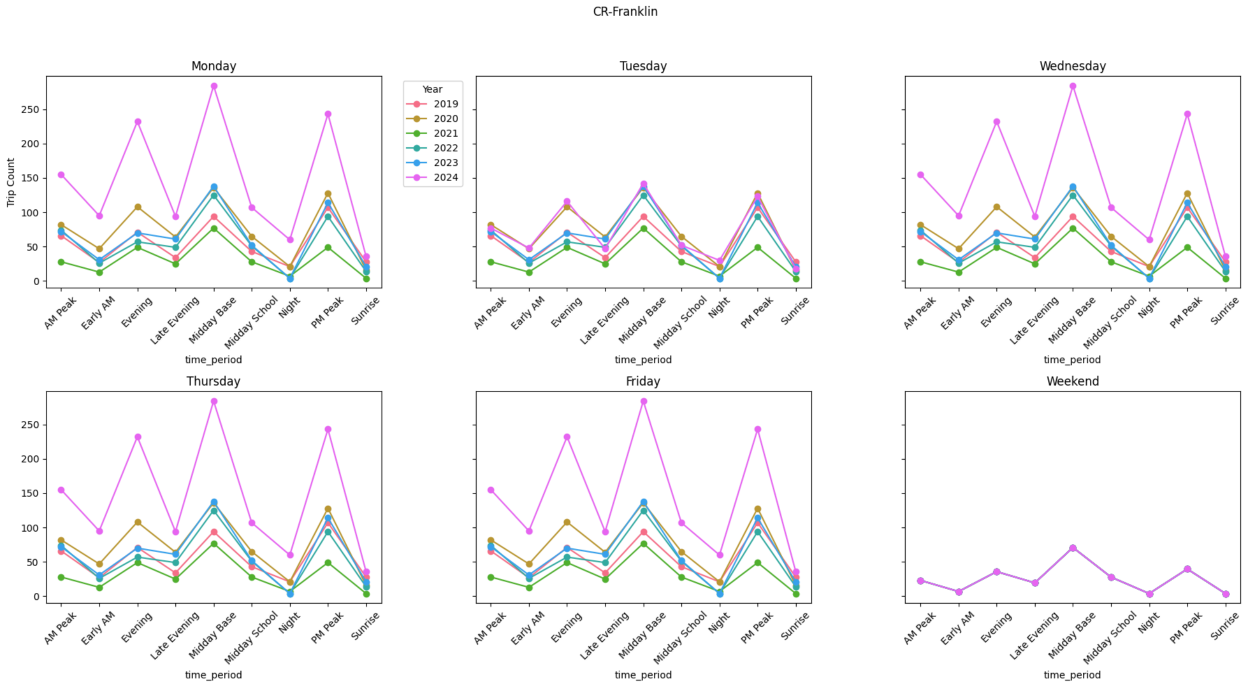

Trip Counts by Time Period for Route CR-Newburyport (2019-2024)
Year
2019
2020
2021
2022
2023
2024
Trip Counts
0
500
1000
1500
2000
2500
Sunrise
Early AM
AM Peak
Midday Base
Midday School
PM Peak
Evening
Late Evening
Night
Time Period

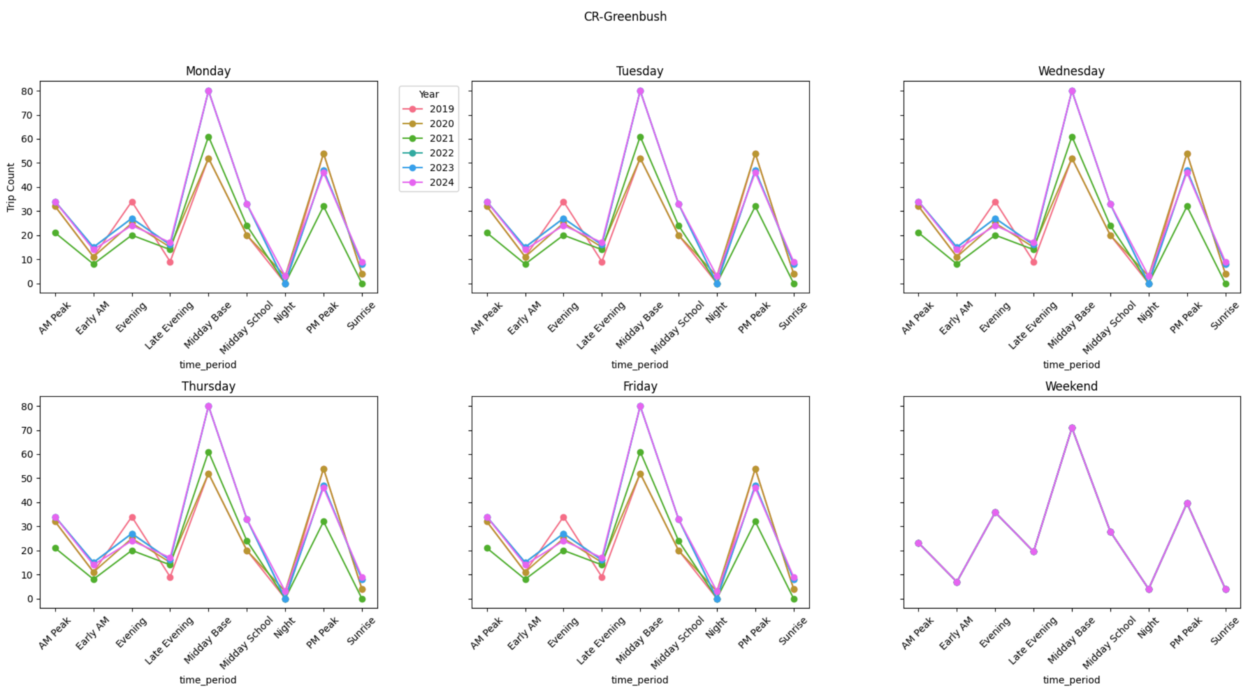

Trip Counts by Time Period for Route CR-Providence (2019-2024)
Year
2019
2020
2021
2022
2023
2024
Trip Counts
0
500
1000
1500
2000
2500
Sunrise
Early AM
AM Peak
Midday Base
Midday School
PM Peak
Evening
Late Evening
Night
Time Period

Appendix B

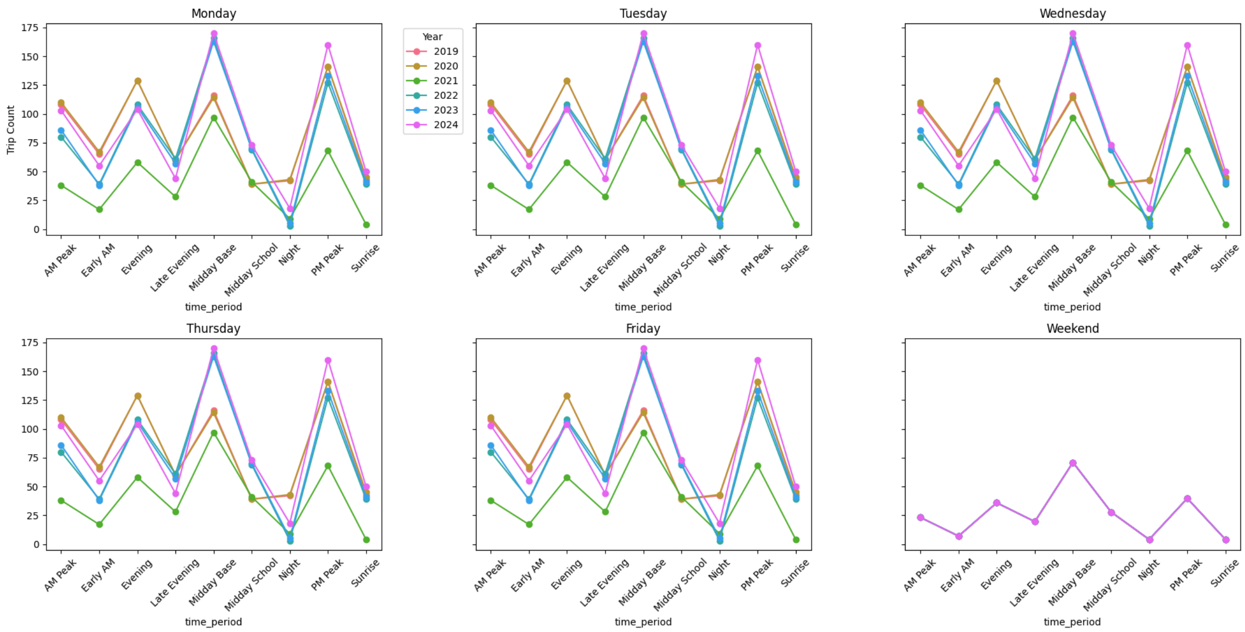

CR-Fairmount
Monday
Tuesday
Wednesday
Thursday
Friday
Weekend
Trip Count
time_period
Year
2019
2020
2021
2022
2023
2024
AM Peak
Early AM
Evening
Late Evening
Midday Base
Midday School
Night
PM Peak
Sunrise


CR-Fitchburg
Monday
Tuesday
Wednesday
Thursday
Friday
Weekend
Trip Count
time_period
Year
2019
2020
2021
2022
2023
2024
AM Peak
Early AM
Evening
Late Evening
Midday Base
Midday School
Night
PM Peak
Sunrise

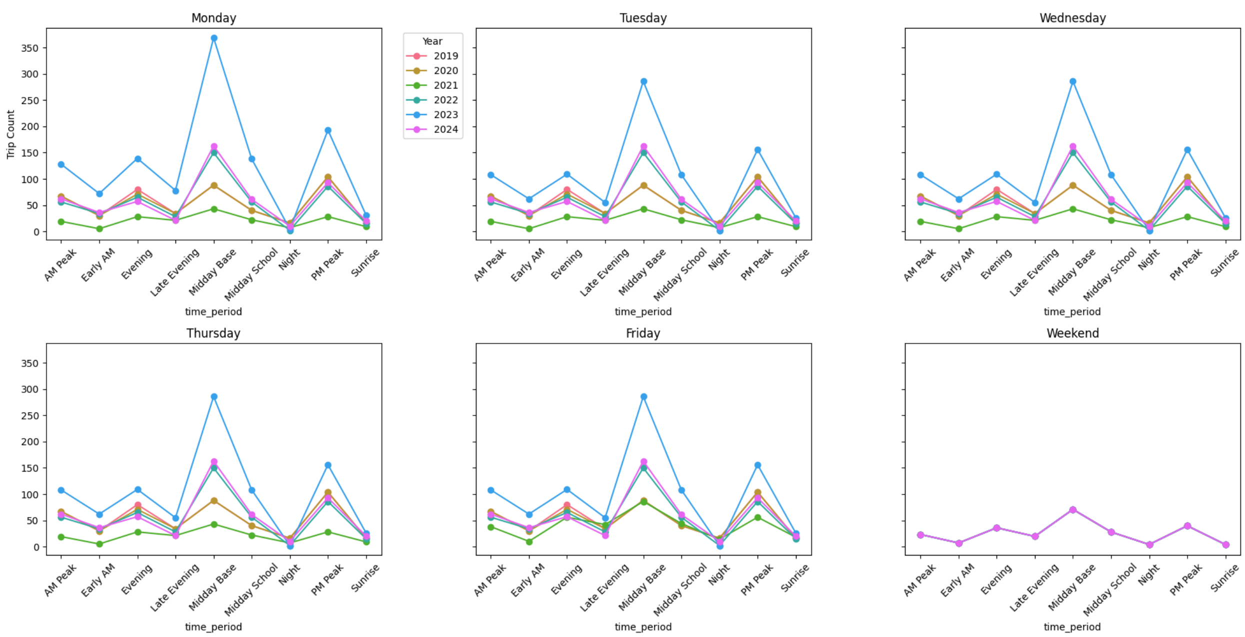

CR-Franklin
Monday
Tuesday
Wednesday
Thursday
Friday
Weekend
Year
2019
2020
2021
2022
2023
2024
Trip Count
time_period
AM Peak
Early AM
Evening
Late Evening
Midday Base
Midday School
Night
PM Peak
Sunrise

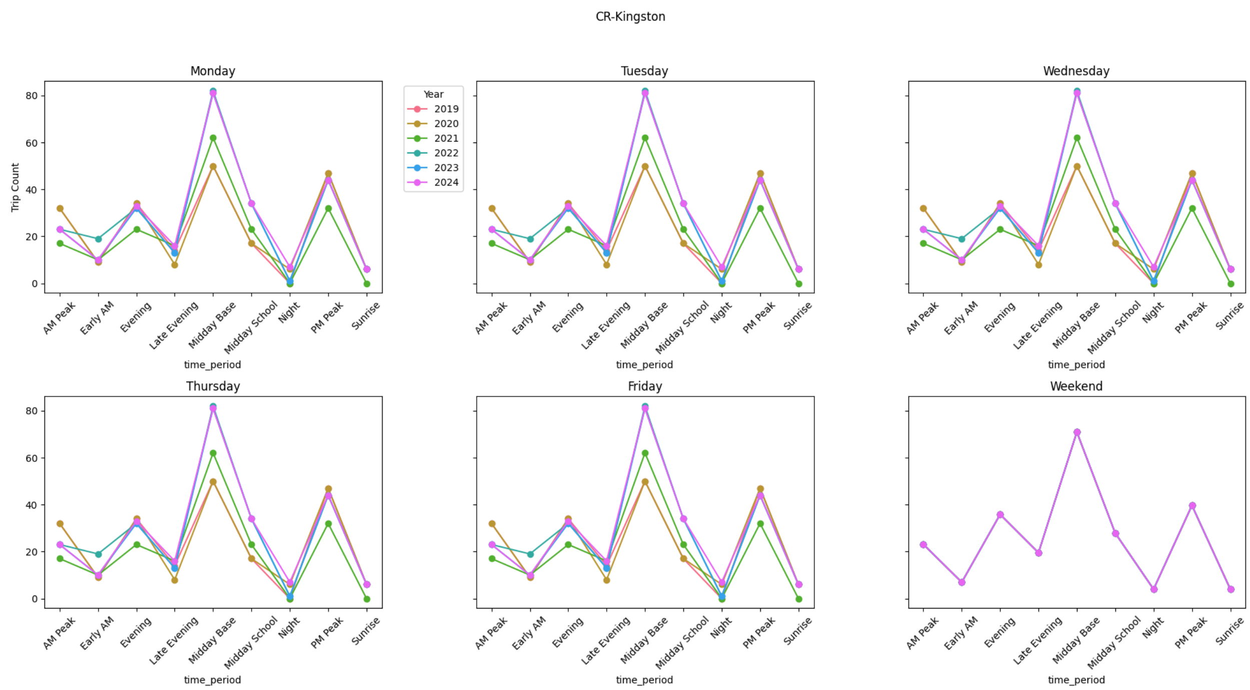

CR-Greenbush
Monday
Tuesday
Wednesday
Thursday
Friday
Weekend
Year
2019
2020
2021
2022
2023
2024
Trip Count
time_period
AM Peak
Early AM
Evening
Late Evening
Midday Base
Midday School
Night
PM Peak
Sunrise

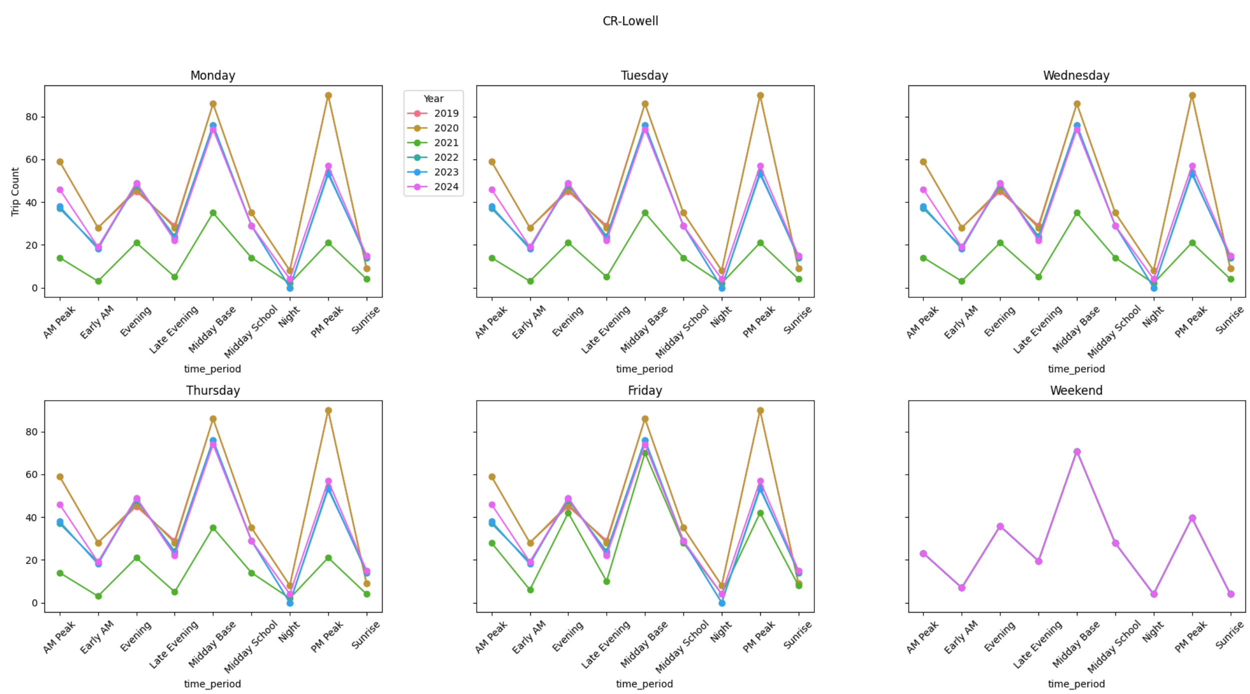
CR-Worcester

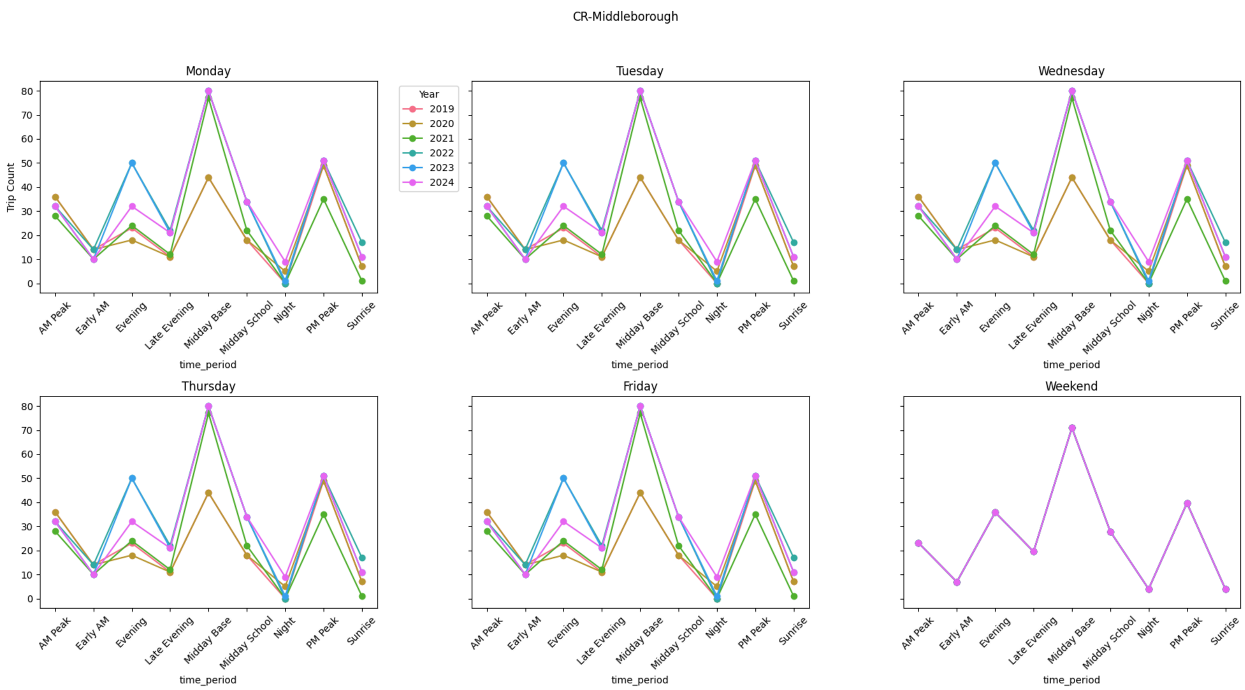


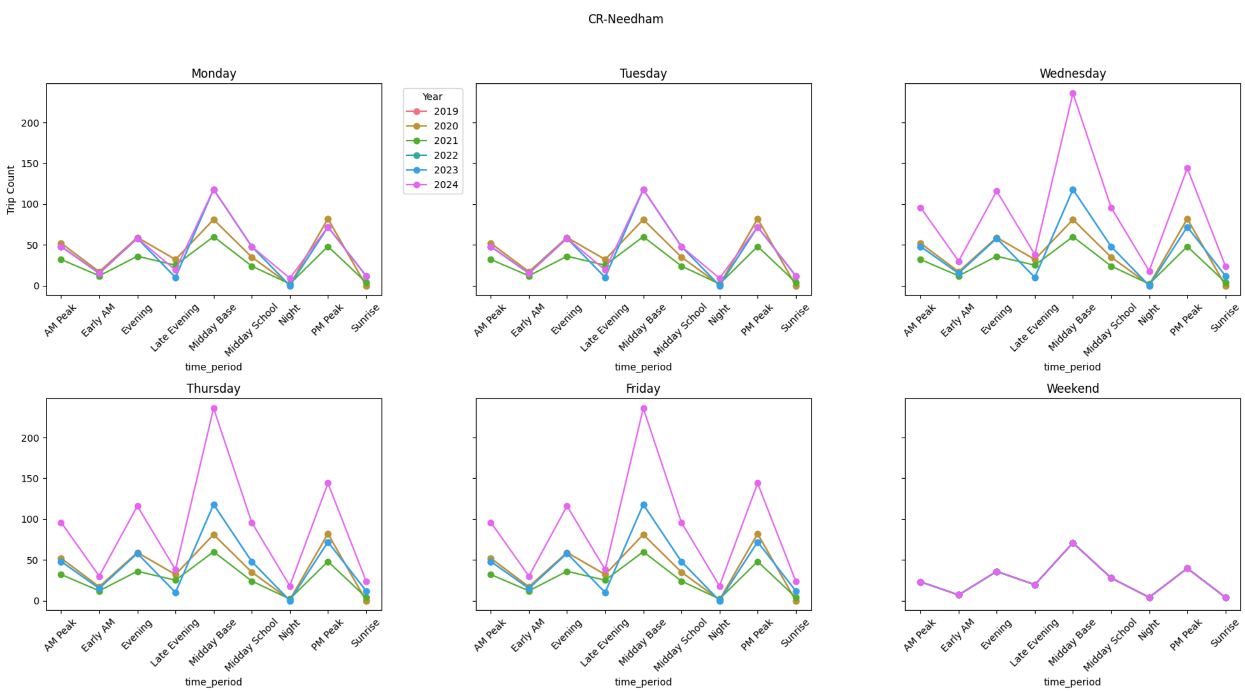
CR-Haverhill



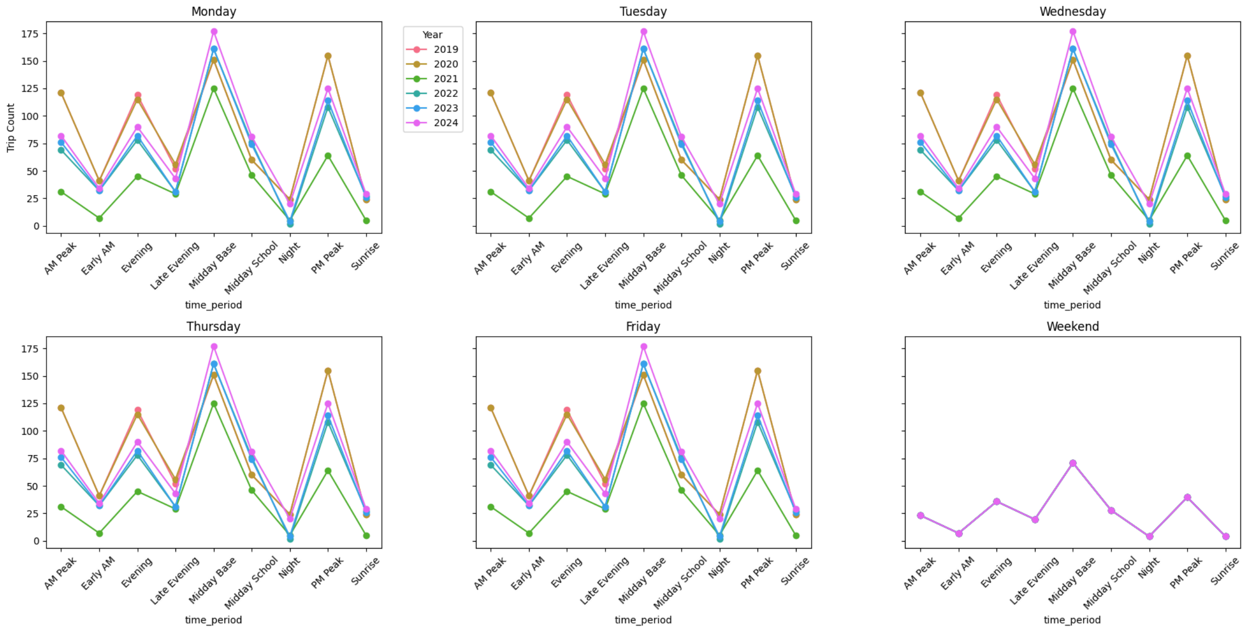

CR-Kingston
Monday
Tuesday
Wednesday
Thursday
Friday
Weekend
Year
2019
2020
2021
2022
2023
2024
Trip Count
time_period
AM Peak
Early AM
Evening
Late Evening
Midday Base
Midday School
Night
PM Peak
Sunrise
0
20
40
60
80


CR-Lowell
Monday
Tuesday
Wednesday
Thursday
Friday
Weekend
Year
2019
2020
2021
2022
2023
2024
Trip Count
time_period
AM Peak
Early AM
Evening
Late Evening
Midday Base
Midday School
Night
PM Peak
Sunrise
0
20
40
60
80

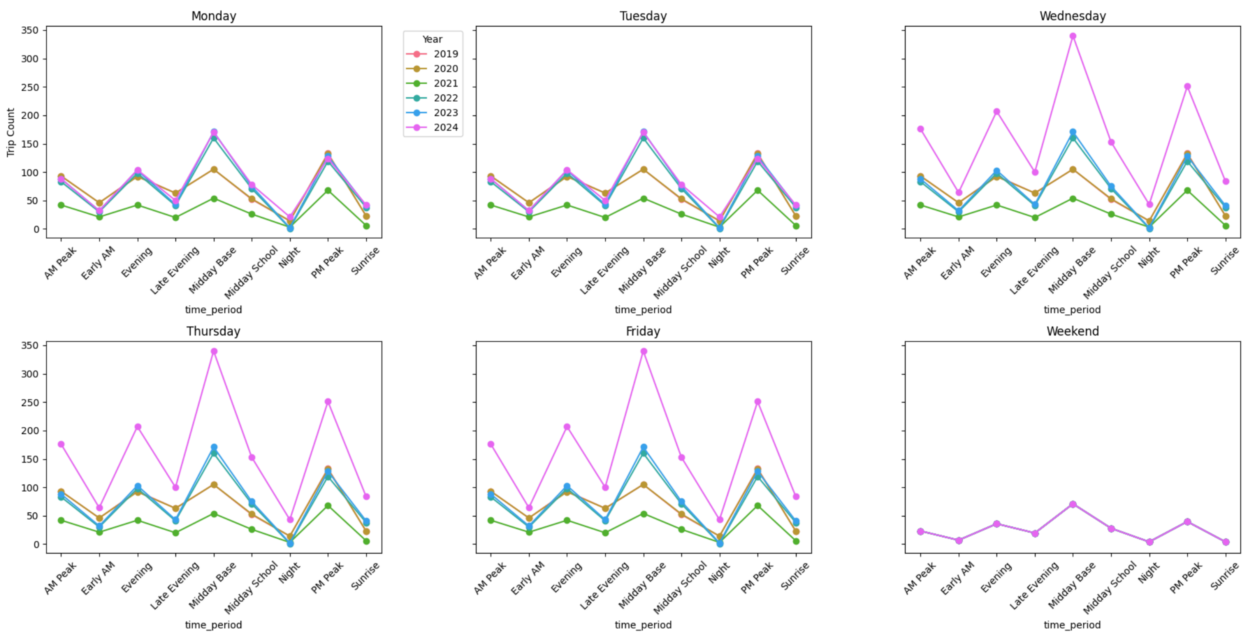

CR-Middleborough
Monday
Tuesday
Wednesday
Thursday
Friday
Weekend
Year
2019
2020
2021
2022
2023
2024
Trip Count
0
10
20
30
40
50
60
70
80
AM Peak
Early AM
Evening
Late Evening
Midday Base
Midday School
Night
PM Peak
Sunrise
time_period


CR-Needham
Monday
Tuesday
Wednesday
Thursday
Friday
Weekend
Year
2019
2020
2021
2022
2023
2024
Trip Count
0
50
100
150
200
AM Peak
Early AM
Evening
Late Evening
Midday Base
Midday School
Night
PM Peak
Sunrise
time_period

CR-Newburyport

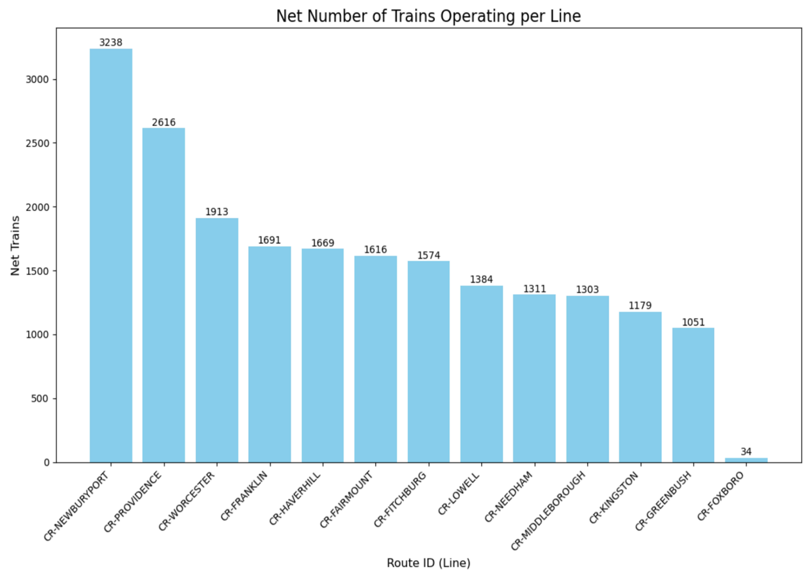


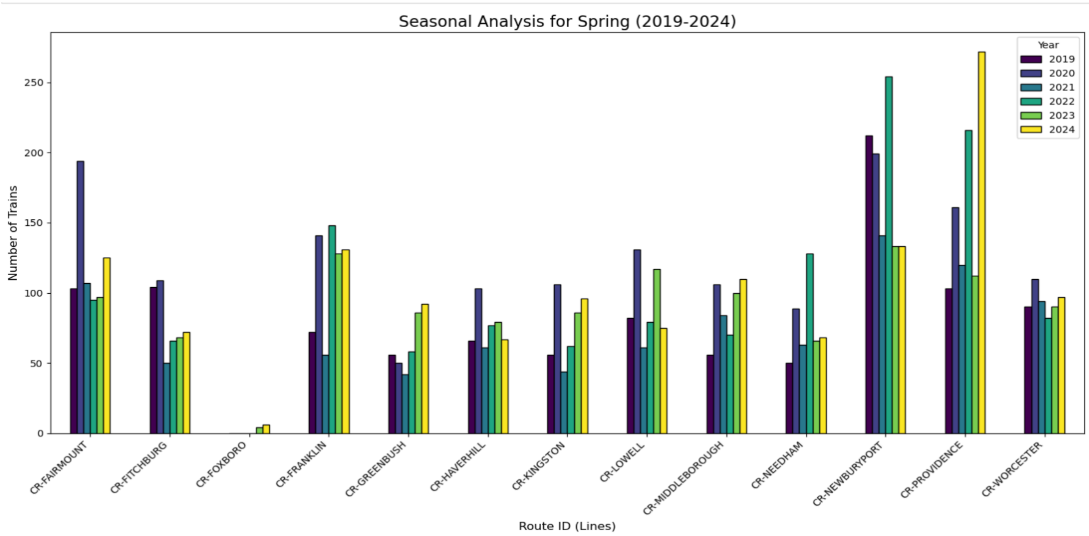
CR-Providence

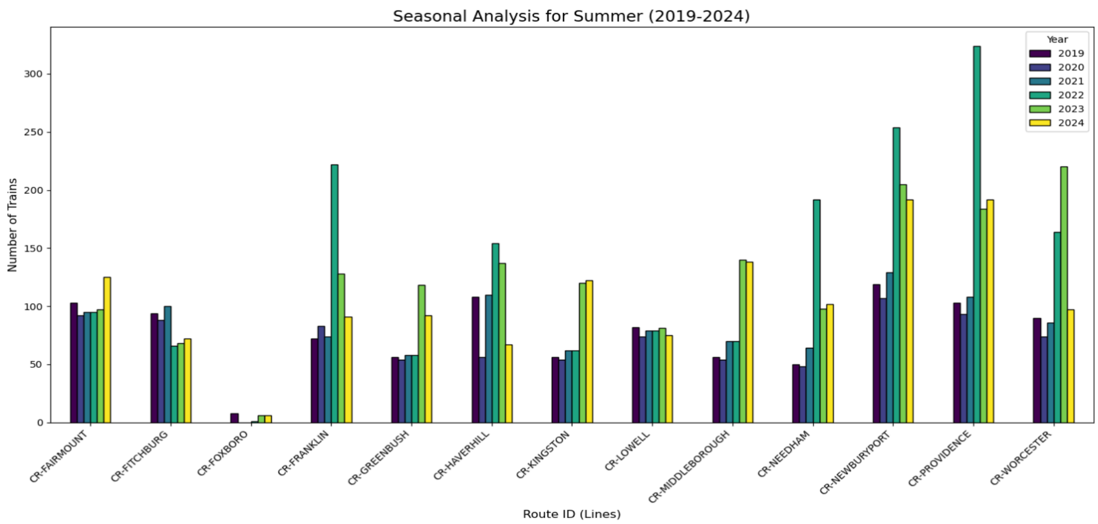


Appendix C

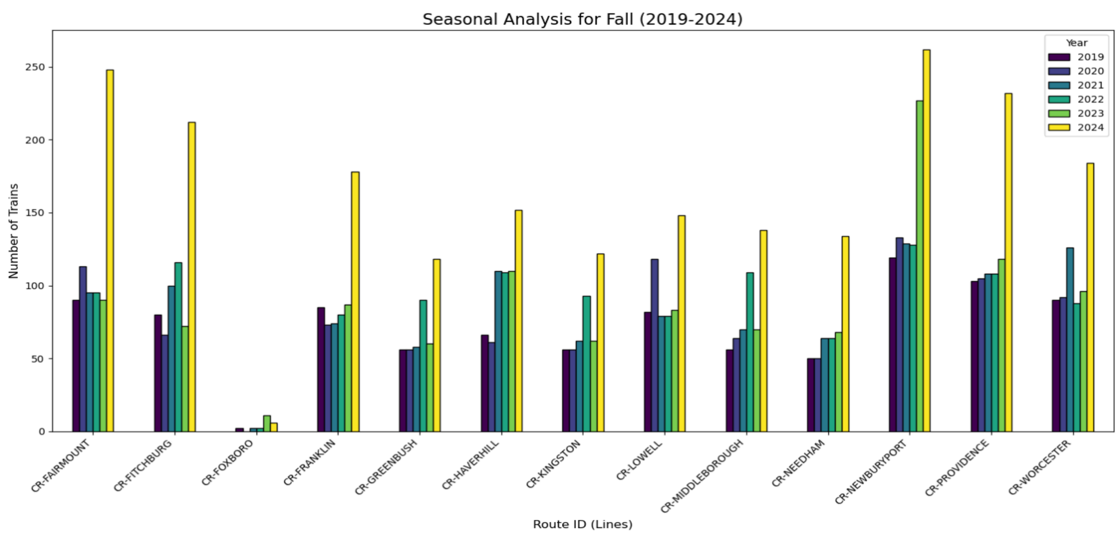
What is the net number of trains operating per line?



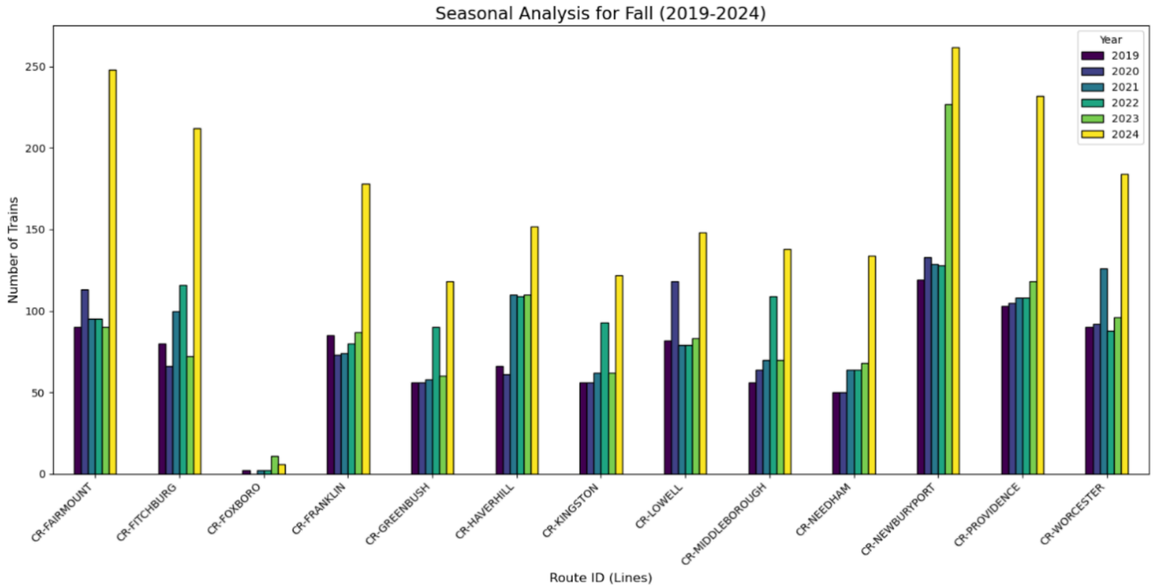




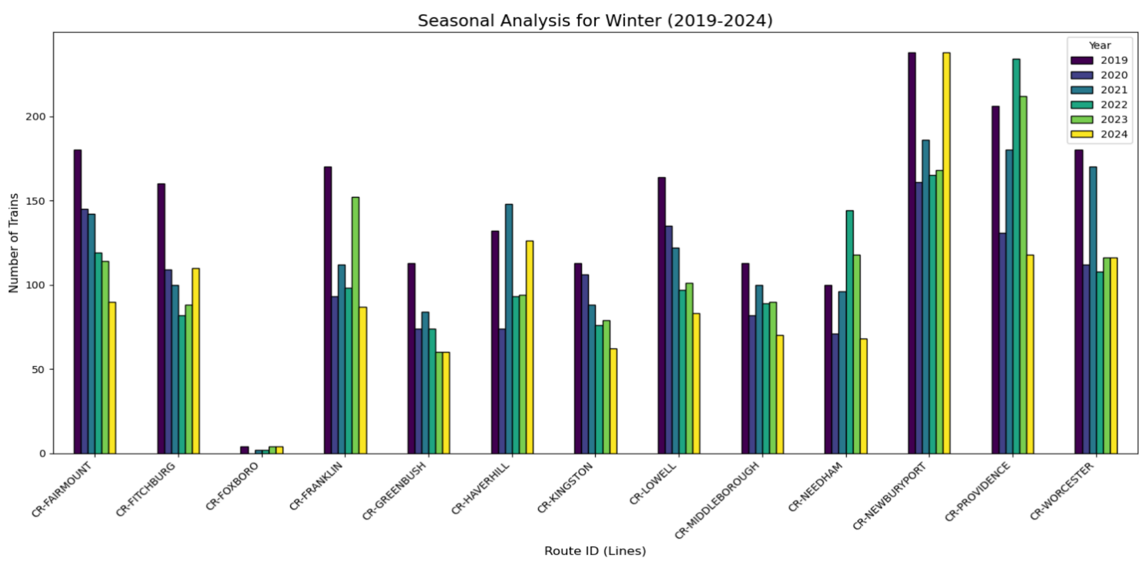
Seasonal Analysis for Summer (2019-2024)
Year
2019
2020
2021
2022
2023
2024
Number of Trains
CR-FAIRMOUNT
CR-FITCHBURG
CR-FOXBORO
CR-FRANKLIN
CR-GREENBUSH
CR-HAVERHILL
CR-KINGSTON
CR-LOWELL
CR-MIDDLEBOROUGH
CR-NEEDHAM
CR-NEWBURYPORT
CR-PROVIDENCE
CR-WORCESTER
Route ID (Lines)



Seasonal Analysis for Fall (2019-2024)
Year
2019
2020
2021
2022
2023
2024
Number of Trains
CR-FAIRMOUNT
CR-FITCHBURG
CR-FOXBORO
CR-FRANKLIN
CR-GREENBUSH
CR-HAVERHILL
CR-KINGSTON
CR-LOWELL
CR-MIDDLEBOROUGH
CR-NEEDHAM
CR-NEWBURYPORT
CR-PROVIDENCE
CR-WORCESTER
Route ID (Lines)



Seasonal Analysis for Fall (2019-2024)
Year
2019
2020
2021
2022
2023
2024
Number of Trains
0
50
100
150
200
250
CR-FAIRMOUNT
CR-FITCHBURG
CR-FOXBORO
CR-FRANKLIN
CR-GREENBUSH
CR-HAVERHILL
CR-KINGSTON
CR-LOWELL
CR-MIDDLEBOROUGH
CR-NEEDHAM
CR-NEWBURYPORT
CR-PROVIDENCE
CR-WORCESTER
Route ID (Lines)



Seasonal Analysis for Winter (2019-2024)
Year
2019
2020
2021
2022
2023
2024
Number of Trains
0
50
100
150
200
CR-FAIRMOUNT
CR-FITCHBURG
CR-FOXBORO
CR-FRANKLIN
CR-GREENBUSH
CR-HAVERHILL
CR-KINGSTON
CR-LOWELL
CR-MIDDLEBOROUGH
CR-NEEDHAM
CR-NEWBURYPORT
CR-PROVIDENCE
CR-WORCESTER
Route ID (Lines)

Appendix D

	route_id	amount
0	CapeFlyer	22.00
1	CR-Providence	13.25
2	CR-Fitchburg	12.25
3	CR-Worcester	12.25
4	CR-Kingston	12.25
5	CR-Middleborough	12.25
6	CR-Newburyport	12.25
7	CR-Haverhill	11.00
8	CR-Franklin	10.50
9	CR-Greenbush	10.50
10	CR-Lowell	10.50
11	CR-Fairmount	7.00
12	CR-Needham	7.00