Fredericton Transit & Weather Analysis

Safia Naznin
Technology Management and Entrepreneurship
University of New Brunswick
Fredericton, Canada
safia.naznin@unb.ca

Sara Manha
Computer Science
University of New Brunswick
Fredericton, Canada
sara.manha@unb.ca

Saeid Metvaei
Civil Engineering
University of New Brunswick
Fredericton, Canada
s.metvaei@unb.ca

Rishabh Kalai
Computer Science
University of New Brunswick
Fredericton, Canada
rishabh.kalai@unb.ca

***Abstract*—This study delves into the analysis of historical data from Fredericton Transit and Fredericton Weather to unravel commuters' route preferences and underlying factors influencing route popularity. The overarching goal is to offer data-driven recommendations to the transit authority for optimizing the public transportation system in Fredericton. Through analysis, this project aims to refine the efficiency and convenience of public transit, ultimately improving the overall passenger experience. By leveraging insights gleaned from extensive data analysis, authorities can make informed decisions regarding route optimization and system improvements.**

***Index Terms*—transit, weather, data, analysis, feature selection, clustering, prediction**

I. Introduction

Transit systems and weather patterns intricately shape our daily lives, affecting our routines, schedules, and safety. The fusion of data analytics with these two critical elements—transit and weather—proffers an opportunity to revolutionize their efficiencies. This paper explores the transformative potential of leveraging data analytics in these domains. By dissecting commuter behavior, traffic flows, and historical patterns, data analytics empowers transit systems to optimize routes, predict congestion, and enhance overall efficiency. Simultaneously, in weather forecasting, data analytics decipher complex atmospheric data, and historical weather patterns, elevating the precision and accuracy of predictions. This paper delves into methodologies, and emerging trends illustrating how data analytics reshapes transit operations and weather forecasts. It aims to highlight the unprecedented advancements achievable by harnessing the power of data in these crucial aspects of modern society.

II. Literature Review

Fredericton partners with Stantec for a $90,000 strategic transit plan, aiming to enhance services and attract more riders. The study includes route reassessment, a potential second transit hub, and the integration of technologies like Wi-Fi. Emphasizing efficiency without route removal, the plan aligns with the city's vision of diverse and forward-looking transportation services. Public engagement will be key to optimizing the system for current and future needs. [1] Research highlights transit's significant economic benefits in Canada, surpassing mere numbers. Annual savings of $16.16 billion in operating costs, collision expenses, and health benefits far exceed the $12.86 billion in annual transit costs. Unquantified benefits like travel time, land use effects, and physical activity could further enhance these impacts. Future evaluations may uncover additional advantages, including improved access to jobs and services, family savings, and increased city wealth. As urbanization and global competition rise, investing in transit becomes crucial for community competitiveness, quality of life, and long-term sustainability. The true economic benefit extends beyond numbers, reaching every corner and generation of Canadians to come. [2]

This study leverages extensive smart card records and detailed meteorological observations to examine the influence of daily weather conditions on public transit usage. Conducted at fine spatiotemporal scales, the research establishes statistical associations between intra-day variations in public transit ridership and changes in weather conditions, considering both system-wide and station-level dynamics. Results reveal that an increase in humidity, wind, and rainfall is generally associated with a decrease in transit ridership, with variations across metro and bus systems. Weather impacts are more pronounced during off-peak hours, indicating commuting trips during peak hours are more resilient. Station-level models demonstrate the spatial distribution of weather effects, with more affected stations located within urban areas. Notably, the study highlights that frequent transit users, often labeled as 'captive riders,' exhibit increased transit use under adverse weather, enriching the understanding of individual responses to weather variations. While providing valuable empirical evidence, the study acknowledges limitations, such as the absence of route information and the need for a more extended study period for a comprehensive analysis of seasonal variations. [3] The intersection of Fredericton's strategic transit plan, the economic impact of transit in Canada, and the analysis of weather effects on public transit paints a clear picture of the vital role comprehensive transit planning plays in our communities. Fredericton's $90,000 partnership with Stantec reflects a forward-looking strategy, focusing on efficiency, technology, and alignment with the city's transportation vision. The economic analysis of transit in Canada goes beyond numbers, emphasizing health, environmental, and societal benefits crucial for community competitiveness and sustainability. The weather impact study, while revealing decreased ridership in specific conditions, sheds light on the resilience of commuting trips and the increased use by frequent transit users during adverse weather. These insights collectively guide the creation of adaptable and user-centric transit systems, ensuring their relevance and impact for further development within the city.

III. Project Benefits

The comprehensive analysis of commuters' route preferences and the identification of influential factors behind route popularity, assigning bus stops to each passenger through historical data are pivotal in aiding transit authorities to make informed decisions. By leveraging these insights, authorities can implement strategic improvements across various facets of the public transportation system in Fredericton. This encompasses optimizing routes, refining schedules, enhancing infrastructure, and even potentially introducing innovative services or incentives that align closely with commuters' preferences. Such data-driven enhancements not only elevate the efficiency and convenience of public transit but also contribute to reducing congestion, environmental impact, and commuting times. Moreover, this approach fosters a more sustainable and commuter-centric transit system, increasing overall user satisfaction and encouraging greater public reliance on public transportation.

IV. Dataset Details

The study draws its data from two primary sources: Fredericton Open Data [4] and Fredericton Weather Stats [5], covering a substantial period from November 24, 2019, to December 31, 2022. This extensive data set spans over three years, 1 month, and 8 days. Comprising a total of 58 features and exceeding 220,000 records, the data set was analyzed using Python, Google Colab, and Excel. The utilization of these tools facilitated in-depth exploration and comprehensive analysis of the substantial volume of data, enabling the extraction of meaningful insights into commuters' route preferences and associated factors influencing route popularity within Fredericton's public transportation system.

V. Pre-processing Data Methods

The project unfolds starting with exploratory data analysis (EDA) of the Fredericton Transit data set, to uncover underlying trends, correlations, and dependencies within the data set. The integration of the Fredericton Weather data set enriches the analysis enabling an assessment of user's impact on commuters' route selections.

We merged transit and weather data, aiming to understand how diverse factors, like weather conditions, influence the popularity and usage patterns of different transit routes. We computed various metrics for each route on a daily, weekly, and monthly basis. These metrics include the total number of sessions and unique users utilizing each route within specific time frames. Additionally, we establish thresholds to categorize routes into 'Very Less Popular', 'Less Popular', 'Moderately Popular', and 'Highly Popular' based on daily session counts. We also assigned seasons (Winter, Summer, Fall) according to the month. We then segregate data based on popularity metrics and generate separate data sets that contain unique user counts and usages for each transit route. To prevent skewed analysis, we removed outliers in route usage data and categorized routes into three levels of popularity by their usage metrics and incorporated it into the data set under the column 'overall_popularity'.

Subsequently, we refined our data set to ensure the data set's accuracy and relevance for analysis by removing potential duplicate entries, filtering, and removing data. We addressed missing weather attribute values is a crucial step for our data set. We imputed these missing values with means derived from weekly or monthly averages, ensuring the preservation of seasonal trends in the data set and not affecting the distribution of data.

We then identified entries with identical "User_ID," "Route," "Start_Date," and "Start_Time" and if they occurred within a 5-minute interval, we removed those entries to eliminate potential duplicates or instances of rapid reboarding. Next, we focused on transit working hours on a weekday and removed entries where "Start_Time" is before 6:15 AM or after 11:00 PM. Furthermore, entries corresponding to Sundays were also excluded from the data set, recognizing that transit patterns on Sundays often differ from typical weekdays. We also converted the time values into a 24-hour format ensuring uniformity in representation, eliminating confusion between AM and PM designations prevalent in a 12-hour clock, and facilitating more efficient data analysis and comprehension.

VI. Feature Engineering

We enriched the data set by adding 'Boarding Stop' to each user by assuming that the user boards the bus on their phones three minutes before the scheduled departure, using this and the Fredericton bus schedule images which we converted to CSV format. [6] We then used the bus schedule to allocate the bus stop by using the start time of the transit and adjusting it by adding 3 minutes assuming the user boards 3 minutes before the scheduled departure. If the resulting minute value is more than 59, it sets it to 59 to keep it within the valid minute range. This adjustment simulates the boarding time being 3 minutes before the actual departure time. It then finds the closest scheduled time after this adjusted start time from the schedule and determines the corresponding boarding stop for that time.

Figure 1 illustrates the count of passengers getting on the bus at various stops. The Windsor/Montgomery stop records the highest boarding numbers and is located outside the

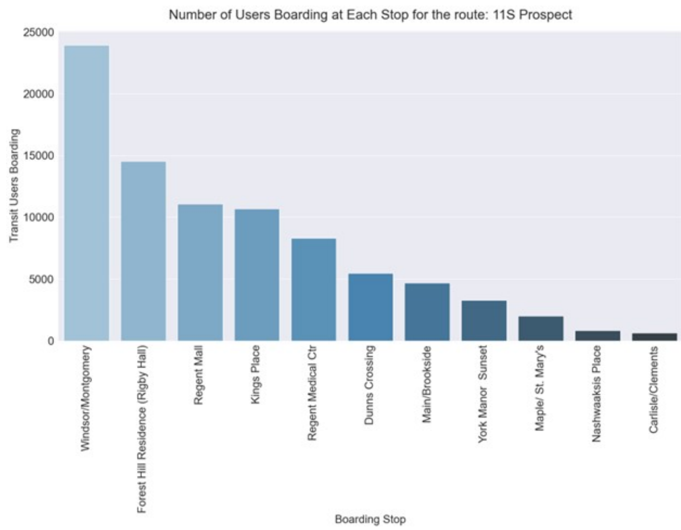

Fig. 1. Distribution of Boardings at Bus-Stops

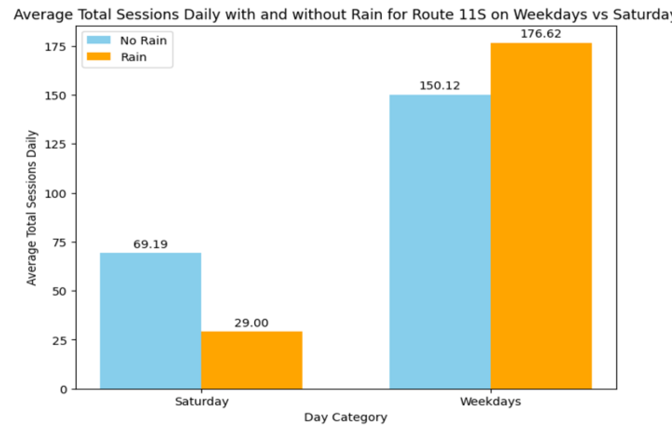

Fig. 2. Rain Anomaly

university, suggesting a high likelihood of students using the Hotspot transit app. This assures us that our approximation for the Boarding Stop is reasonably accurate.

Using the cleaner and enriched data set, we proceeded to the following step for feature selection to get more accurate insights into transit usage patterns during operational hours on regular weekdays. At its essence, this project revolves around the development of clustering analysis and a predictive model tailored to categorize bus routes into distinct popularity segments with similar bus usage behaviors. The model harmonizes aspects of popular routes, peak travel times, and user preferences as discussed in the following sections.

VII. Feature Selection

In the pursuit of identifying the key factors influencing transit ridership, two distinct feature selection methods were employed – the Gini Index Method and Correlation Analysis. The primary objective was to pinpoint specific weather-related attributes that could potentially impact transit ridership patterns. Despite the application of rigorous feature selection techniques, the analysis yielded no discernible impact on transit ridership. Initial attempts to isolate influential factors proved inconclusive, prompting a shift in focus.

Recognizing the limitations of the initial feature selection, emphasis was redirected towards understanding temporal patterns. This shift was prompted by the hypothesis that weather might have a differing impact on weekdays compared to Saturdays. The rationale behind this lies in the understanding that on weekdays, transit usage could be more necessity-driven, such as commuting to work or school. In contrast, on Saturdays, transit usage might lean towards discretionary activities like recreational trips to the mall. Consequently, adverse weather conditions on Saturdays, being non-essential, could lead to plan cancellations. Contrary to expectations, the study revealed minimal variation in transit ridership between weekdays and Saturdays.

The majority of features show consistent patterns throughout the week. Nevertheless, a notable rain anomaly is evident in Figure 2. On Saturdays, when rain occurs, there is a decrease in the total daily sessions compared to rain-free Saturdays. Conversely, on weekdays, even in the presence of rain, the total daily sessions show an increase.

VIII. Clustering Analysis of Transit Usage

The motivation behind using clustering analysis stems from the need to enhance the efficiency of the city's transit system. By understanding the diverse patterns in bus usage among transit users, we aim to provide decision-makers with valuable insights to optimize service planning and resource allocation. This data-driven approach ensures that limited resources are maximized to benefit most users.

Using clustering analysis, we were able to identify distinct user segments with similar bus usage behaviors. This, in turn, facilitates targeted decision-making by providing insights into popular routes, peak travel times, and user preferences. The potential analyses include route optimization, service improvements, targeted marketing, and efficient resource allocation, ultimately leading to a more user-centric and optimized public transit system.

A. Introduction to K-Means Algorithm

1) K-Means Overview: The K-means algorithm is a widely used clustering technique that groups data points into distinct clusters based on similarity in their features. It operates iteratively, assigning data points to the nearest cluster centroid and updating centroids until convergence. K-means minimizes the variance within clusters, making it a robust method for forming well-defined clusters. In this project, K-means was chosen for its simplicity and scalability, making it suitable for our large data set of bus users. Its efficiency in forming clear clusters aligns with our goal of uncovering distinct patterns in bus usage behavior. The following steps demonstrate the function of K-Means:

- Initialization: Random placement of cluster centers (centroids).

- Assignment Step: Data points are assigned to the nearest centroid.
- Update Step: Centroids are recalculated as the means of assigned data points.
- Iteration: Repeat the assignment and update steps until convergence.

B. Implementation of Clustering Analysis

Our data set comprises bus user information, concentrating on two crucial sets of features outlined in Figure 3. The initial set involved aspects associated with the bus route usage and frequency of each user's utilization of specific bus routes. The second set detailed the times at which users boarded buses, organizing their usage patterns into morning, mid-day, and evening periods.

	user_id	10N Carlisle	116 Kings Place	11S Prospect	13S Prospect	14N Barkers Point	15S Hanwell	16N Marysville	17S Regent	18 Silverwood	Counts for Mornings	Counts for Mid-day	Counts for Evenings
0	111111410	0	0	0	0	0	0	0	0	1	0	1	0
1	111111513	0	0	0	0	0	0	0	0	2	0	2	0
2	111111688	0	0	0	0	0	1	1	0	0	2	0	0
3	111111721	8	0	0	4	2	0	4	10	0	5	17	6
4	111111797	0	0	0	0	0	0	13	2	0	1	14	0
...	...	...	...	...	...	...	...	...	...	...	...	...	...
5822	211595206	2	0	0	0	0	0	0	0	0	1	1	0
5823	211595348	0	1	0	0	0	0	0	1	0	0	2	0
5824	211595555	1	0	0	0	0	0	0	0	0	1	0	0
5825	211595620	0	0	0	0	0	0	0	2	0	0	1	1
5826	211595661	0	0	0	0	0	1	0	0	0	0	0	1

Fig. 3. Key Feature Groups in Bus User Data

The optimal parameters for clustering were determined by carefully selecting the number of clusters as 8 and setting the random state to 28. This decision was made thoughtfully, aiming to strike a balance between maximizing the Silhouette score and ensuring an adequate number of clusters to unveil specific patterns within the data. The Silhouette score, which evaluates the cohesion and separation of clusters, serves as a metric for assessing how well-defined the clusters are, with higher scores indicating clearer boundaries between clusters.

Fig. 4. Silhouette Scores Across Varying Numbers of Clusters

As shown in Figure 4, the visual illustration of the Silhouette score corresponding to each cluster count provides a clear depiction of the ideal number of clusters based on the established criteria.

	10N Carlisle	116 Kings Place	11S Prospect	13S Prospect	14N Barkers Point	15S Hanwell	16N Marysville	17S Regent	18 Silverwood	Counts for Mornings	Counts for Mid-day	Counts for Evenings
0	3.597929	0.835850	3.259707	1.239646	0.668680	0.744392	0.894953	1.466350	6.169111e-02	5.318594	5.464193	1.532571
1	130.275862	25.620690	379.517241	26.689655	11.724138	21.482759	10.862069	22.724138	3.448276e-02	275.862069	254.068966	82.620690
2	155.910112	3.337079	22.640449	27.280899	4.617978	5.314607	4.089888	4.033708	1.123596e-02	165.617978	45.112360	8.224719
3	20.694348	7.870819	36.395617	11.143022	7.782007	8.600923	7.385236	12.843137	3.033449e-01	49.283737	47.379469	14.101499
4	25.166667	8.866667	33.733333	20.733333	12.833333	4.100000	127.533333	199.766667	6.666667e-02	189.266667	189.033333	39.466667
5	12.428571	392.285714	38.857143	4.857143	0.571429	1.571429	0.428571	5.285714	1.428571e-01	330.714286	90.857143	31.857143
6	43.355705	15.167785	160.093960	31.302013	11.906040	10.194631	6.879195	12.261745	1.476510e-01	103.919463	130.852349	51.510067
7	22.900000	36.950000	33.850000	38.400000	203.000000	199.500000	12.800000	27.050000	1.387779e-17	234.050000	223.450000	100.600000

Fig. 5. Cluster Profiles: Average Attribute Values for Each Cluster

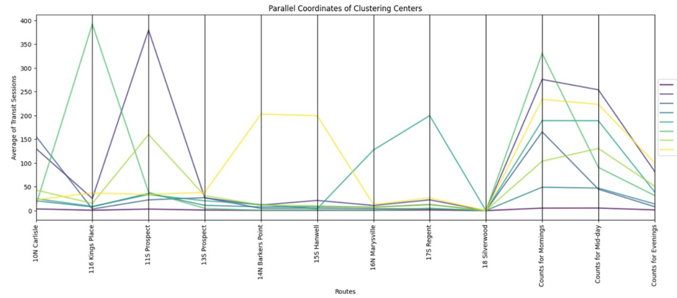

Fig. 6. Cluster Profiles: Visual Representation of Average Attribute Values for Each Cluster

Upon executing the K-means model, users were categorized into 8 distinct clusters, unveiling unique usage patterns among individuals. Figure 5 and Figure 6 supply a visual representation and average attribute values for each cluster, offering a comprehensive overview of their respective characteristics.

Cluster	Number of People	# of Transit Sessions	Contribution in Transit Data %
0	4635	59101	22.16
1	29	18239	6.84
2	89	20224	7.58
3	868	98084	36.77
4	30	12984	4.87
5	7	3195	1.20
6	149	43405	16.27
7	20	11489	4.31

Fig. 7. Cluster Distribution Summary

In Figure 7, there's an overview of how individuals are distributed across clusters, their overall bus usage, and the percentage each cluster contributes to the transit records. Notably, clusters 0 and 3 host a substantial number of individuals, yet they lack clear defining traits. However, a closer look unveils important details: those in cluster 3 consistently prefer the 11S Prospect route, indicating its significance within the transit system as depicted in Figure 8.

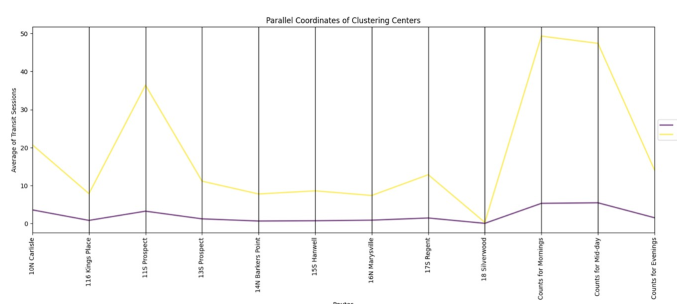

Fig. 8. Route Preference within Clusters 0 and 3

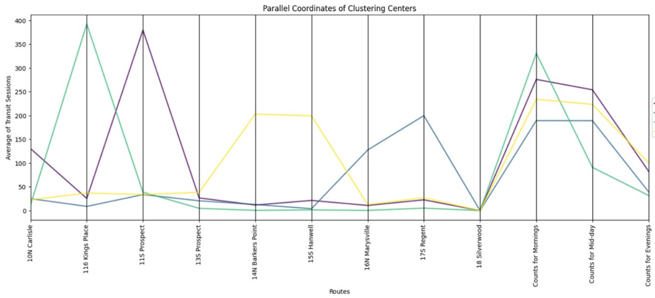

Fig. 9. Route Engagement for Users Within Clusters 1, 4, 5, and 7

Among these clusters, smaller cohorts exhibit highly specialized usage patterns on specific bus routes as shown on Figure 9. For instance, cluster 1 predominantly uses route 11S Prospect, while clusters 4, 5, and 7 engage distinctively with routes 116 Kings Place, 16N and 17S, 14N and 15S, respectively. These specialized user groups present a strategic advantage for transit authorities. By focusing on these cohorts, transit authorities can conduct targeted route-specific surveys and implement tailored adjustments to enhance the experience for users along these routes. This nuanced approach can significantly elevate service quality, meeting the unique needs of these user segments and contributing to an optimized and user-centric public transit system.

C. Application of the Results

In leveraging the insights gained from our clustering analysis, we identified clusters 1, 3, and 7 as individuals who exhibit a higher frequency of bus route 11S Prospect usage as shown in Figure 10.

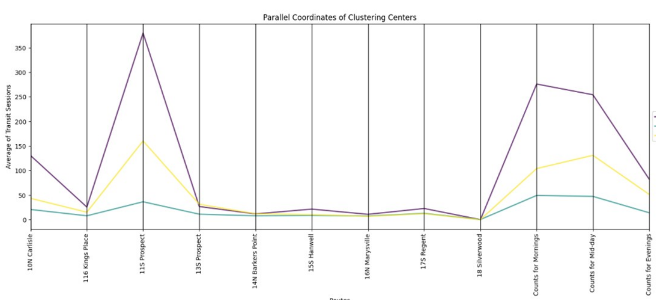

Fig. 10. Clusters 1, 3, and 7: Elevated Frequency of Bus Route 11S Prospect Usage

Building upon this distinction, we performed detailed calculations using the transit records of these clusters, incorporating some simplifying assumptions to find the average number of individuals from these clusters boarding each bus at various times of the day as illustrated in Figure 11.

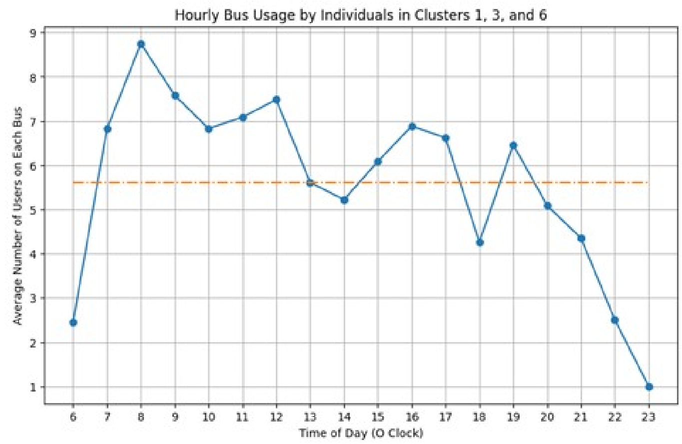

Fig. 11. Average Individuals Boarding Buses by Time: Clusters 1, 3, and 7 Analysis

This information presents a strategic opportunity for transit authorities to consider augmenting bus frequencies during peak hours, such as the morning rush from 7 AM to 9 AM for the 11S Prospect route. Aligning service adjustments with the behavior of loyal users enhances their satisfaction, improves service reliability, reduces wait times, and significantly enhances the overall transit experience for these dedicated route users. Crucially, the significance of clustering becomes apparent when comparing all users' transit records versus the behavior of loyal users for a specific route.

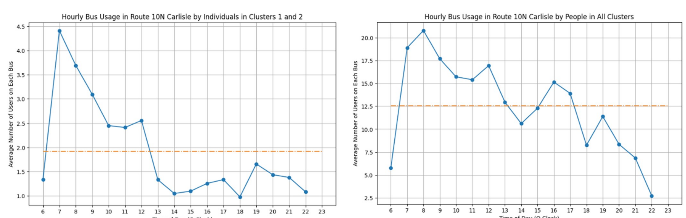

Fig. 12. Peak Hour Usage for All Users Vs. Loyal Individuals on Route 10N

As illustrated in Figure 12, while overall transit records might reveal several hourly peak times for route 10N Carlisle, loyal individuals (those who are in clusters 1 and 2) predominantly exhibit a single peak around 7 AM. Understanding these differences empowers authorities to make informed decisions, catering either to enhancing the satisfaction of loyal users or enticing regular users to become loyal, aligning schedule adjustments with distinct user behaviors for a more user-centric transit system.

IX. Regression Analysis of Transit Usage

Understanding and predicting weekly transit usage is crucial for efficient resource allocation, service planning, and enhancing overall transit system performance. Weekly patterns in transit usage often exhibit distinct variations influenced by factors such as workdays, weekends, and recurring events. This predictive insight can empower transit authorities in Fredericton to optimize schedules and allocate resources effectively.

A. Utility of Seasonal ARIMA

Seasonal Auto-Regressive Integrated Moving Average (SARIMA) is a time series forecasting model that extends the ARIMA model to account for seasonality. It is particularly effective for capturing seasonal trends in transit usage because it considers both the auto-regressive and moving average components of the time series and the seasonal component. This allows SARIMA to capture the patterns and fluctuations that occur at regular intervals, such as daily, weekly, or monthly seasonality in transit usage data. By incorporating these seasonal factors into the model, SARIMA can provide more accurate forecasts for time series data with strong seasonal patterns.

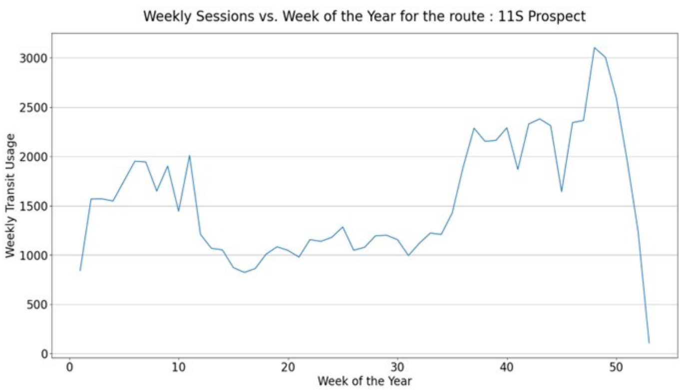

Fig. 13. Weekly Trends over the Year

The weekly trends as shown in Figure 13 for the transit usage for the route '11S Prospect' show a sharp increase around August demonstrating the surge in usage due to new international students starting university and then continuing to increase until the end of the year after which there is a sharp decline demonstrating holiday period when most users travel.

B. SARIMA Prediction based on the Weekly Transit Usage

Aiming to capture these seasonal trends, we used the Seasonal ARIMA (SARIMA) model to generate predictions for the '11S Prospect' route's weekly transit usage. SARIMA incorporates auto-regressive, moving average, and seasonal components, making it well-suited for capturing the intricate patterns observed in transit data with periodic fluctuations.

- Auto-Regressive (AR): This component looks at the previous week's transit usage to predict the current week. It considers how past values directly influence future values.
- Moving Average (MA): This part focuses on the average of past prediction errors, helping to smooth out any random fluctuations in the data.
- Seasonal Component: SARIMA considers the repeating patterns that occur at regular intervals, such as weekly variations in transit usage. This is crucial for capturing the influence of factors like workdays, weekends, or specific events that may impact transit demand differently each week.

SARIMA is a good choice for predicting weekly transit usage, even in the presence of data anomalies like those caused by the COVID-19 pandemic, for several reasons:

1) Adaptability to Seasonal Changes: SARIMA inherently accounts for seasonality, making it suitable for capturing regular fluctuations in transit patterns. Even during unprecedented events like COVID-19, where transit usage deviated drastically from normal patterns, SARIMA can adapt and still provide reasonable predictions based on historical trends.
2) Robustness to Short-Term Anomalies: SARIMA incorporates both autoregressive and moving average components, enabling it to handle short-term anomalies or sudden changes in transit behavior. While COVID-19 led to abrupt and significant disruptions, its ability to smooth out random fluctuations helps it maintain reliability in predicting patterns over the long term.

C. Model Calibration and Validation

It is imperative to highlight the calibration and validation processes undertaken. The model was calibrated using historical weekly transit usage data, with a specific focus on the identified seasonal patterns. Subsequently, validation was performed to assess the model's accuracy and reliability in capturing the observed trends. The hyperparameters were tuned for the following:

- Non-Seasonal Hyperparameters (p, d, q)
  - p (AR, Autoregressive): Determines how many past weeks' transit usage should be considered to predict the current week. A higher p captures longer-term trends, useful if there are persistent patterns week after week.
  - d (I, Integrated): Specifies how many differences between consecutive weeks are needed to make the transit usage data stationary. It addresses any underlying trends or seasonality.
  - q (MA, Moving Average): Represents the number of past prediction errors that should influence the current week's prediction. A higher q considers more short-term fluctuations in transit usage.
- Seasonal Hyperparameters (P, D, Q, s)
  - P (Seasonal AR): Similar to p, but for the seasonal component. It indicates how many past weeks with the same season (e.g., same week of the year) should be considered for predicting the current week's seasonal pattern.
  - D (Seasonal I): Like d, but for the seasonal component. It represents the number of differences between consecutive seasonal weeks needed to make the seasonal transit usage data stationary.
  - Q (Seasonal MA): Like q, but for the seasonal component. It determines how many past seasonal prediction errors should influence the current week's prediction.
  - s (Seasonal Step): Defines the length of the seasonal pattern. In the context of weekly transit usage, s=12

implies a yearly seasonality since there are typically 12 months in a year.

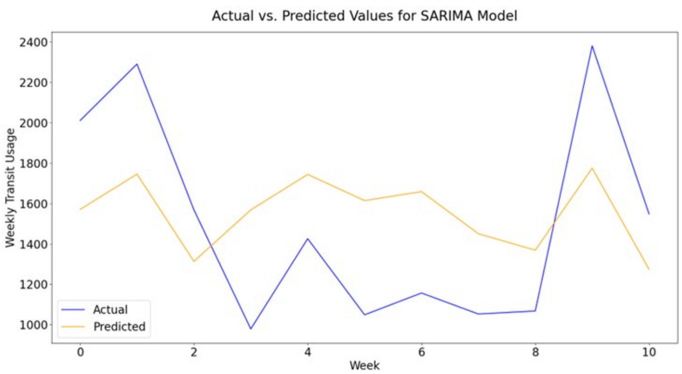

Fig. 14. Actual vs Predicted Values for SARIMA Model

Given the nature of transit operations where bus schedules can't be drastically altered daily, an RMSE (Root Mean Square Error) of 33.8% and an accuracy of 66.2% may be acceptable for optimizing weekly resource allocation. The model as shown in Figure 14 provides a reasonable understanding of transit patterns, allowing for effective resource planning, even if individual predictions show some variability. The focus should be on capturing broader trends and weekly variations, aligning well with the practical constraints of maintaining a consistent bus schedule. While fine-tuning the model is an option, the current performance metrics still offer valuable insights for optimizing transit resource allocation within the confines of a weekly planning cycle.

Conclusion

Transit systems are crucial, ensuring efficient connections and smooth mobility essential for our everyday activities and societal function. In conducting a holistic analysis of the Transit System in Fredericton, we employed K-Means clustering to identify distinct user segments with unique Transit Usage Behavior. Simultaneously, to capture the seasonal trends inherent in the transit usage pattern, we complemented our analysis by incorporating Seasonal AutoRegressive Integrated Moving Average (S-ARIMA) modeling. This addition allowed us to discern and account for the temporal variations, providing a more comprehensive understanding of transit user segments and their anticipated behaviors. By integrating K-Means clustering and S-ARIMA, our approach offers a nuanced perspective on transit dynamics, enabling tailored strategies and interventions to enhance the overall efficiency and satisfaction of the transit system in Fredericton.

X. Limitations

The data set exhibits potential class imbalance, given its focus solely on users accessing Fredericton's transit system through the Hotspot App. This limitation introduces a bias in the data, possibly skewing our understanding. Additionally, the data set might not fully capture the impact of the COVID-19 pandemic on transit ridership, potentially limiting the applicability of our findings. Moreover, there are uncertainties surrounding the assignment of boarding stops, which rely on estimations rather than in-depth analysis. This lack of detailed analysis could affect our interpretation of transit patterns. Lastly, considering the inherent limitations of the Hotspot App, there's a possibility that the data's comprehensiveness and accuracy might be compromised, ultimately impacting the reliability of the insights we derive.

XI. Future Improvements

Within our project, several opportunities for improvement such as augmenting our data set by incorporating details about boarding stops and bus-stop descriptions hold the potential to offer valuable insights into bus-occupancy rates. This addition could significantly enhance our understanding of the efficiency of various bus stops. Moreover, delving into transit patterns, particularly analyzing data from bus stops, presents an opportunity to identify optimal locations for establishing new bus depots or hubs. This analytical approach, based on transit patterns, contributes significantly to the strategic planning of transit infrastructure, optimizing its efficiency and effectiveness.

References

[1] New study aims to help increase the use of public transit in Fredericton — CBC News. (2017). Retrieved from https://www.cbc.ca/news/canada/new-brunswick/public-transit-fredericton-1.4460334

[2] CUTA Transit Economic Impact in Canada: More Than Numbers. Retrieved from https://cutaactu.ca/wp-content/uploads/2021/01/final_issue_paper_50_cuta_v2.pdf

[3] M. Zhou, D. Wang, Q. Li, Y. Yue, W. Tu, and R. Cao, "Impacts of weather on public transport ridership: Results from mining data from different sources," Transportation Research Part C: Emerging Technologies, vol. 75, pp. 17–29, 2017, doi: https://doi.org/10.1016/j.trc.2016.12.001.

[4] Fredericton Transit Histories 2019 2022. (n.d.). Retrieved from https://data-fredericton.opendata.arcgis.com/datasets/Fredericton::fredericton-transit-histories-2019-2022/about

[5] Data Download for Fredericton. (n.d.). Retrieved from https://fredericton.weatherstats.ca/download.html

[6] Schedule Maps. (n.d.). Retrieved from https://www.fredericton.ca/en/resident-services/fredericton-transit/schedule-maps