AVE Trends in Intelligent Computing Systems

Studying Price Dynamics of Bus Services Using Machine Learning Algorithms

Parth Jani
Department of Business Architecture, Tech Operations, Molina Healthcare, Virginia, United States of America. jani.p1212@gmail.com

Desiya Nanban* **and Jennifer Selvan**
Department of Computer Science and Engineering, SRM Institute of Science and Technology, Ramapuram, Chennai, Tamil Nadu, India. dn4939@srmist.edu.in, js6485@srmist.edu.in

Nicholas Richardson
Department of Software Engineering, JPMorgan Chase, Chicago, Illinois, United States of America. nicrichardson322@gmail.com

R. Sivakani
Department of Artificial Intelligence and Data Science, Dhaanish Ahmed College of Engineering, Chennai, Tamil Nadu, India. sivakani@dhaanishcollege.in

R. Subhashni
Department of Computer Science and Applications, St. Peter's Institute of Higher Education and Research, Avadi, Chennai, Tamil Nadu, India. subhashniraj2018@gmail.com

*Corresponding author

Abstract: Today, bus services have become a very convenient way of travelling from one place to another in every household. Since the introduction of websites like Redbus.in, bookmytrip.com, etc., which have been incorporated into our lives, booking and travelling on buses has become much easier. However, the price dynamics for these bus services don't remain constant every single day, and they vary depending on the days as well as the time; for example, if a bus ticket is booked closer to the customer's departure date, it would be more expensive when compared to the other days before that. Initially, a bus service would set its price for a certain date. Observing this, the other bus services would set a price closer to or lesser than that to attract more customers. So, the main purpose of this study is to find the bus service that initially sets the price. In order to find which bus service initially sets the price, we will use a machine learning algorithm known as k-means clustering, which groups the different bus services into clusters with similar data fields.

Keywords: Machine Learning (ML); K-means Clustering; Feature Engineering (FE); Bus Services; Price Dynamics (PD); Price Category; Machine Learning Tools (MLT); Price Dynamics in Online.

Cite as: P. Jani, D. Nanban, J. Selvan, N. Richardson, R. Sivakani, and R. Subhashni, "Studying Price Dynamics of Bus Services Using Machine Learning Algorithms," *AVE Trends In Intelligent Computing Systems*, vol. 1, no. 1, pp. 54 –65, 2024.

Journal Homepage: https://avepubs.com/user/journals/details/ATICS

Received on: 03/09/2023, **Revised on:** 05/11/2023, **Accepted on:** 07/01/2024, **Published on:** 05/03/2024

1. Introduction

Bus services refer to public transportation systems that transport passengers from one location to another. Government agencies or private companies typically operate these services. Bus services provide an efficient and effective way for people to travel within or between cities and towns. Online bus services refer to platforms or websites that allow users to book bus tickets conveniently [12]. Users can search for available bus routes, select their preferred departure and arrival locations, choose a date and time, and book tickets online. Some online bus booking platforms also offer the option to select specific seat types, allowing passengers to select their preferred seating arrangements [13]. In this study, we will try to find the bus service that initially sets the price first, which influences other bus services to follow them, and set the price equal to or less than them to attract more customers.

The algorithm we will use to find which bus service sets the price first will be the k means clustering algorithm [14]. This algorithm follows unsupervised learning, meaning it aims to find patterns, structures, or relationships within the data independently without being provided with explicit input and output data (Figure 1).

Data fields: Seat Fare Type1, Seat Fare Type2, Bus, Recorded at, Service Date

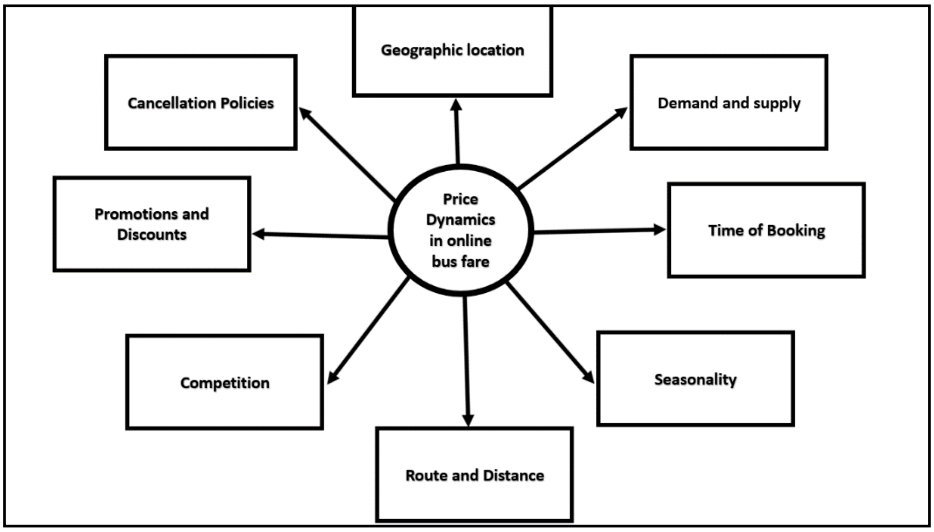

Figure 1: Price dynamics in online bus fare

In this study, we approach the challenge of understanding the pricing dynamics of bus services as an unsupervised learning problem. Our methodology's focal point is deploying the k-means clustering algorithm, a powerful tool in unsupervised learning. The underlying concept guiding our investigation is rooted in the premise that bus services tending to follow each other in pricing indicate similarities in their operational strategies.

We aim to categorize these akin bus services into distinct groups by embracing clustering, a technique intrinsic to unsupervised learning. This categorization facilitates a clearer understanding of the underlying patterns and streamlines our price analysis process, making it both more efficient and insightful. The unsupervised nature of this approach allows the algorithm to autonomously identify inherent relationships within the data without relying on predetermined labels, paving the way for a nuanced exploration of the intricate dynamics governing the pricing behaviours of different bus services. Through this lens, our study aims to unravel the latent structures in the pricing strategies of bus services, contributing to a more comprehensive comprehension of the industry landscape.

2. Literature review

Thomas et al. [1] concluded that Redbus is a volume-driven company intending to enter more Indian markets and improve customer service by building regional offices. Redbus' current strategy entails gathering bus tickets using any available means and then distributing them to customers via any available channel. Redbus has to present a growth strategy to deliver long-term, sustainable growth. From the beginning, maintaining positive relationships with bus drivers and winning over new drivers and agents has been difficult for redBus' owners.

Banoula [2] said that the study aimed to use a model to separate fare changes from other factors' effects. This study used a special survey to collect ridership data for 52 transit systems 24 months prior and 24 months following each fare change. The survey also collected monthly information on other factors that could affect ridership, such as gasoline prices, vehicle miles of service (VMS), labour strikes, etc. On average, a 10% hike in bus fares would lead to a 4% drop in ridership.

In another related study, Al-Masri [3] analyzed how riders responded to changes in train frequency, service levels, and other transportation options (such as cars). The results showed that riders responded twice as quickly to changes in those factors in the long term than in the short term. Therefore, attempts to balance budgets by raising fares and lowering service quality would likely lead to higher subsidies and deficits. The results showed that lower public transportation subsidies that lead to higher fares and lower service quality may lead to higher per-rider costs than would be true with higher total subsidies.

Saji [5] used computer models to predict how people would use different modes of transportation depending on different policies. They predicted that having just ride-sharing policies would be less successful and that having a mix of transit and ride-share policies would draw in more people. They also analyzed how auto prices and transit policies can work together to make travel more popular.

According to Mangione [6], the Politeknik Kota Kuala Terengganu (PKKT) Online Transport Booking System was aimed and emphasized that other fields of endeavour might benefit from the methodology and technology used in this new transportation system. Before boarding, the user who wishes to utilize the transport must apply to reserve the transport.

Sindu [7] concluded that in recent years, public transport operators have been attempting to substitute paper-based tickets for electronic media, and several countries have either implemented or are implementing electronic ticketing systems. The primary feature of electronic ticketing is the sale and storage of tickets in electronic devices. Nevertheless, the advantages of a comprehensive e-ticketing system for transport operators are difficult to assess, as the primary objective of e-ticketing is to enhance the quality of service. In financial terms, electronic ticketing could decrease administrative costs, as fewer cashiers would be required, fare processing times would be shortened, and passenger throughput would be increased.

According to Kadir et al. [8], the first suggestion is that online booking providers must offer the service following customers' preferences and safeguard the confidentiality of customers' data. Additionally, it is advised that the refund be processed within the time frame specified on the websites.

Intending to analyze and evaluate the degree of customer happiness, Zhang et al. [9] undertook a project titled "customer satisfaction on online bus ticket booking." The major goal of this study was to determine the degree of customer satisfaction with online bus ticket reservations. There were 110 responders in the sample. The study found that socioeconomic parameters, including age, educational attainment, occupation, and family's monthly income, directly affected customers' satisfaction with the online bus ticketing process.

Sulaiman et al. [10] attempt to concentrate on the motivational variables that impact online purchasing. India has a sizable population of technologically knowledgeable individuals who not only use the internet to browse but also to buy things that are offered online. Online marketing's core idea is to use the internet as a platform to draw clients and offer goods or services. This essay is a theoretical attempt to link important motivating aspects that affect internet purchasing. According to the study, no discernible differences exist in men's and women's motivating elements.

According to Jain et al. [11], "Online Booking" is a mobile application that enables the travel industry to book a trip online as well as make use of various travel services like "bus travel, car rental, hotel, and aeroplane ticket booking." As well as "global one-line purchases," which meet the needs of their users. Additionally, it uses self-service technology (SST) instead of the conventional method. Users of these services will have less anxiety by just using their mobile devices to book their travel rather than doing it in person.

3. Objective

- The main objective of studying the price dynamics of bus services is to find the bus services that set prices independently and those that adopt the price structure of the other available services.
- This study mainly provides a data-based approach to reveal these price dynamics.
- The goal of this study is to group bus services into clusters that have similar data fields.

3.1. Existing Methods

- Booking Time: The timing of your booking can affect the price. Booking well in advance may secure lower fares, while last-minute booking fares can be more expensive.
- Route and Distance: The distance of the journey and the route's popularity can impact prices. Longer routes and those connecting major cities have different pricing structures.
- Seat Type: RedBus often allows passengers to choose seat types (Ex, Sleeper, semi-sleeper, or seater). The price can differ based on the seat selection.
- Festivals or Special Occasions: Due to increased demand, ticket prices may be higher during festivals, holidays, or special occasions.

3.2. Proposed Method

To study the price dynamics of the bus services, we first divided the dataset according to the number of price categories. Due to this, we observed that the same buses had different price categories depending upon their date and time; the closer the booking is to the departure date, the more expensive the bus fare is compared to the one booked earlier.

We propose an "Elbow Method" to find the ideal number of clusters. In order to find the ideal number of clusters, we will use k-means clustering, which groups bus services with similar data fields.

If a new bus is added to the data sample, it will be classified into clusters; within the clusters, similarity indices like price and time will be used to identify the closest leader and follower (Figure 2).

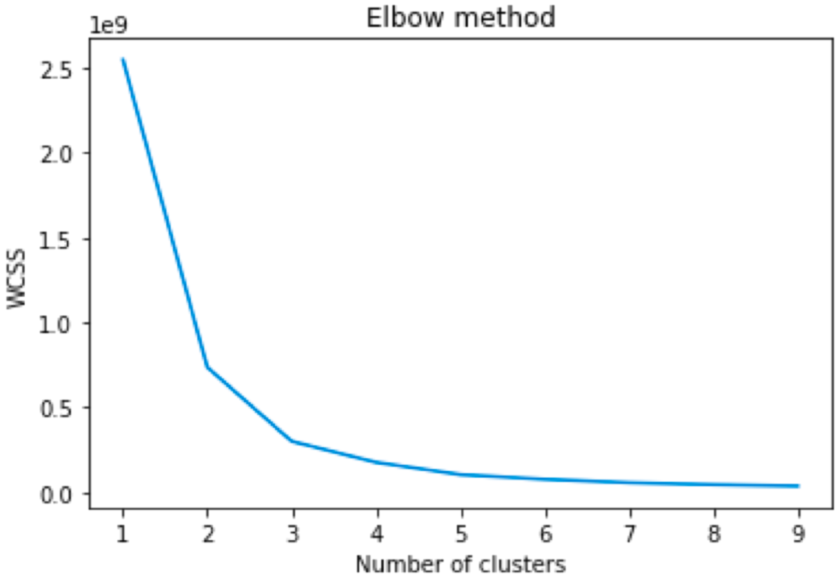

Figure 2: Elbow method for finding several clusters

Formula to calculate WCSS: WCSS(Within Cluster Sum of Squares) = Σ Pi in Cluster1 distance(Pi C1)^2 + Σ Pi in Cluster2distance(Pi C2)^2+ Σ Pi in Cluster3 distance(Pi C3)^2. Σ Pi in Cluster1 distance(Pi C1)2 is the sum of the squares of the distances between each data point and its cluster1 centroid.

3.3. Material needed

- Historical Price Data: Gather historical pricing data for red bus services. This data should include ticket prices for various routes and service classes over a significant period (e.g., months or years).
- Market Data: Collect market data, such as the number of passengers, competitors' pricing, and market trends. This data can help you contextualize price dynamics within the broader industry.
- Competitor Pricing Data: Analyse competitors' pricing strategies in the red bus industry. This data can help you identify pricing trends and benchmarks.
- Analytical Tools: Utilize statistical and analytical tools, such as spreadsheet software (e.g., Excel), data visualization tools (e.g., Tableau), and statistical software (e.g., R or Python with libraries like NumPy and Pandas), to analyze and visualize the data.
- Machine Learning and Predictive Models: Consider using machine learning models to predict future price dynamics based on historical data and market variables.

3.4. Architecture diagram

The process starts by determining the number of clusters, for k means clustering, using techniques like the "Elbow Method" to choose the appropriate clustering setup. Once the optimal number of clusters is identified, k means clustering is used to group the bus service data based on their characteristics. The data is then sorted by bus name, allowing for an examination of each bus service. Mean values for features such as price, time, and booking patterns are calculated for each bus service group, offering insights into the values within each group. Next, each bus service is assigned to a cluster whose centre point is closest to the values of its features, categorizing buses based on similarities in their attributes (Figure 3).

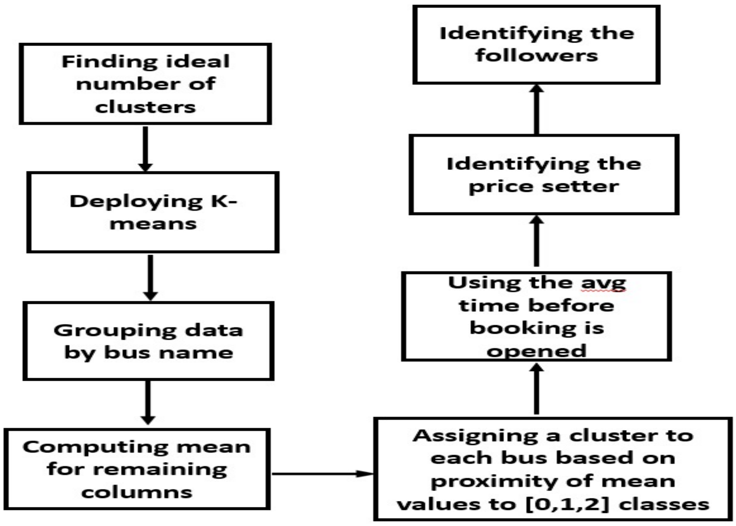

Figure 3: Architecture diagram

The average time taken before booking opens is included as a feature, enhancing the analysis with information about booking trends. In each cluster, the bus service sets pricing standards. Known as the "price setter". It is identified to uncover pricing strategies and market influences. Additionally, the analysis extends to identifying "followers" within each cluster who mirror the pricing and booking behaviours of the price setter. This thorough method allows a grasp of how bus service pricing and market trends unfold.

4. Methodology

4.1. Problem Formulation and Data Collection

- Define the research objectives. Gather historical data on bus fares, including ticket prices, route information, dates, and relevant contextual data (e.g., economic indicators, fuel prices, weather conditions).
- Ensure the data is clean, structured, and well-documented.

4.2. Feature Engineering

- Create relevant features that might impact bus fares, such as:
    - Time-based features (e.g., day of the week, month, season)
    - Historical fare data (e.g., previous fares on the same route)
    - External factors (e.g., inflation rates, fuel prices)
    - Demand-related features (e.g., passenger counts, holidays)
- We need a categorical variable that indicates the number of pricing tiers and an indicator of the type of seat (type 1 or type 2) to investigate the price dynamics of bus services and determine which bus service sets the price first.
- The gap between the recorded and service dates must also be determined, followed by the average price across all levels.
- This time window allows for studying and identifying the bus service that sets the price first.

4.3. Data Splitting

Split the dataset into training, validation, and test sets to evaluate model performance effectively. The training set is used to train your machine learning models, the validation set is used to fine-tune hyper-parameters, and the test set is reserved for the final evaluation. Approach to finding the bus service setting the price first:

- Different bus services were grouped into clusters based on similarities in the data fields.
- The similarity in average time before booking that characterizes a cluster determines the follower and price setter.

4.4. Generalization of approach

- Classify a new bus into any of the clusters using the technique.
- Similarity indices like price and time determine a cluster's nearest leader and follower.
- Any new data can be added to our train data and test case.

5. Interpretability and Analysis

Interpreting models predicting bus service prices requires a multifaceted approach. The initial steps involve analyzing the importance of features using techniques like permutation importance or tree-based models. Correlation analysis uncovers features with significant impacts on prices, while visualization techniques, such as scatter plots and heat maps, reveal complex relationships. Model coefficients in linear models offer insight into the direction and magnitude of feature effects (Table 1).

Table 1: Dates vs Bus fares

Dates	Bus Fares
15 th July	Rs 900
16 th July	Rs 775
17 th July	Rs 650
18 th July	Rs 500
19 th July	Rs 1100

Partial Dependence Plots (PDPs) provide a nuanced understanding of individual feature influences while exploring interaction effects uncovers synergies or antagonisms [4]. Error analysis helps refine model performance by identifying residual patterns. Integrating domain knowledge enriches interpretation, guiding strategic decisions effectively. Through this comprehensive approach, stakeholders gain insights into the dynamic nature of bus service pricing, facilitating informed decision-making and strategy formulation (Figure 4).

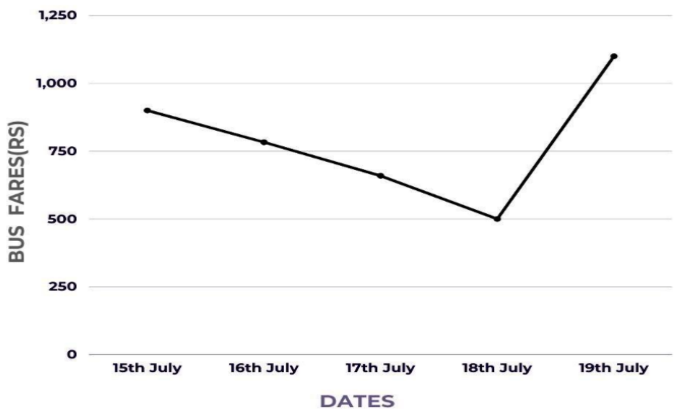

Figure 4: Values of bus fares on the corresponding date

5.1. Conclusion and Reporting

Summarise your findings, draw conclusions about the price dynamics of bus services, and provide recommendations or insights based on your analysis. Document your methodology and results in a report or presentation.

5.2. K means clustering

K-means clustering is a popular unsupervised machine learning algorithm for clustering data points into a predefined number of groups, known as clusters. The algorithm iteratively assigns each data point to the nearest cluster centroid and then recalculates the centroids based on the mean of all points assigned to each cluster. This process continues until the centroids no longer change significantly or a predefined number of iterations is reached.

K-means clustering aims to minimize the within-cluster sum of squares, the sum of squared distances between each data point and its assigned centroid. It's an iterative optimization algorithm and can be sensitive to the initial placement of centroids. Therefore, it's common to run the algorithm multiple times with different initializations and choose the result with the lowest within-cluster sum of squares.

$$\mathbf{J(C)} = \sum_{i=1}^{K} \sum_{j=1}^{N} \| X_j - C_i \|$$

K stands for cluster number.
N: the quantity of information points.
X: The group of informational points

Initialisation:Select the K number of clusters.Set the centroids of K clusters at random. These centroids represent the cluster centres.

Assigning: Locate the closest centroid, C i, for each data point, X j, and allocate X j to cluster X i.Assign each data point to the nearest centroid. This is usually based on the Euclidean distance, but other distance metrics can also be used. Each data point is assigned to the cluster with the nearest centroid.

$$\mathbf{Argmin}_{i} \, \|X_j - C_i\|^2$$

Update: In the K-means algorithm, after assigning each data point to the nearest centroid, the next step is recalculating the cluster centroids. This recalibration ensures that the ci=1:

$$c_i = 1 / |C_i| \sum x$$

ci is the centroid of cluster Ci
$|Ci|$ is the number of datapoint
X represents a datapoint in cluster Ci

Centroids accurately represent the centre of their respective clusters. Specifically, each centroid is updated to be the mean of all data points assigned to its cluster.

Repeat: Iterate steps 2 and 3, analyzing the correlation between features and prices, and visualizing relationships. Repeat until convergence, indicated by minimal centroid changes between iterations. This iterative process refines understanding of feature impacts on bus service prices, enhancing predictive accuracy and insights (Figure 5).

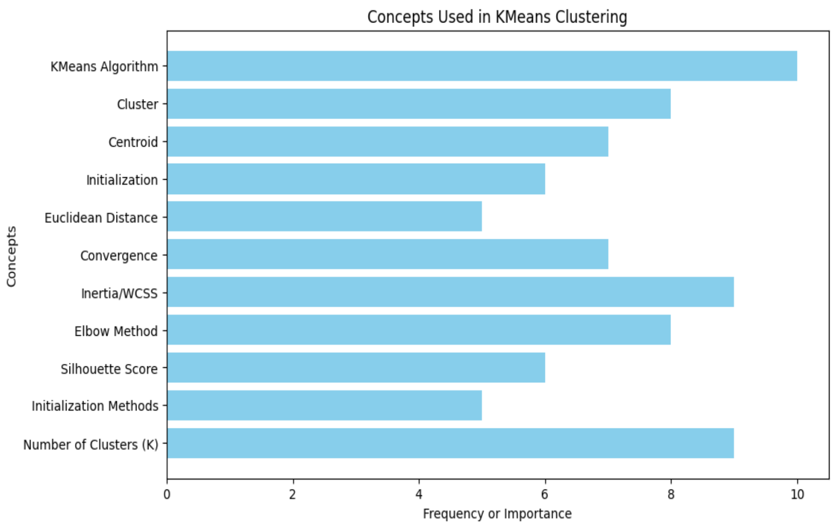

Figure 5: Concepts used in k means clustering

5.3. Implementation And Testing

```
# Import necessary libraries
import pandas as pd
import numpy as np
from statistics import mean
from sklearn.cluster import KMeans
import matplotlib.pyplot as plt

# Read data from ‘PricingData.csv’ into a DataFrame and preprocess it
df = preprocess_data(‘PricingData.csv’)

# Plot the Elbow method to determine the optimal number of clusters
plt.plot(range(1, 10), wcss)
plt.title('Elbow method')
plt.xlabel('Number of clusters')
plt.ylabel('WCSS')
plt.show()

# Fit KMeans with 3 clusters to the data
kmeans_3 = KMeans(n_clusters=3, random_state=0).fit(df_new)

# Assign cluster labels to the original DataFrame
df['cluster'] = kmeans_3.labels_

# Perform KMeans clustering and analyze clusters
analysis = analyze_clusters(df)

# Save analysis results to ‘ppt_approach.csv’
save_to_csv(analysis, 'ppt_approach.csv')

# Sort clusters and generate trail data
trail_version = generate_trail_data(analysis)

# Define a list of Bus IDs for output
output = ['d6fa79179fda2a77455794637f225962', '23400e84ea8d9f642252d1c343d26464', ...]

# Create a final DataFrame with ‘Follows’ and ‘Is followed by’ columns
final = {‘Follows’: follows, ‘Is followed by’: followed_by}
final_version = pd.DataFrame(final, index=output)

# Rename the index of the final DataFrame to ‘Bus’
final_version.index.rename('Bus', inplace=True)

# Save the final output to ‘RJX8235_Data miners_output.csv’
save_to_csv(final_version, 'RJX8235_Data miners_output.csv')
```

6. Results and discussions

6.1. Data Preprocessing

The preprocessing step prepares the pricing data from ‘PricingData.csv’ for clustering analysis. This involves handling missing values, standardizing numerical features, and encoding categorical variables. Before analysis, the dataset undergoes preprocessing steps to clean and format the data.

This may include handling missing values, outlier detection, and standardizing data formats. Additionally, data may be aggregated or filtered to focus on specific periods, routes, or bus IDs of interest.

6.2. Cluster Analysis and Output Generation

Following preprocessing, the refined dataset undergoes analysis using data mining techniques like KMeans clustering. These algorithms categorize the data into clusters, grouping buses with similar trajectories. By identifying commonalities among buses, distinct patterns and relationships emerge, shedding light on operational behaviours or route characteristics. KMeans clustering, specifically, partitions the data into K clusters based on centroids, aiding in discovering clusters representing buses traversing similar routes or exhibiting comparable operational trends. This segmentation facilitates deeper insights into bus dynamics, guiding decision-making processes for route optimization, resource allocation, and service enhancement initiatives (Figure 6).

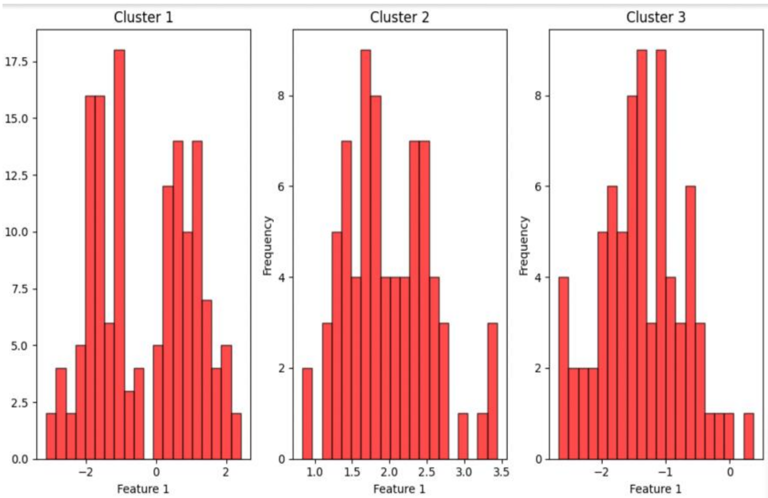

Figure 6: Frequency of clusters

6.3. Directional Relationships

Central to the analysis is exploring directional relationships among buses, particularly discerning ‘Follows’ and ‘Is followed by’ associations. This inquiry delves into both temporal and spatial proximities among buses, elucidating nuanced patterns of bus dynamics. Temporal proximity examines the sequential order of buses along routes, revealing instances where one bus closely follows another. Such observations unveil inherent bus-following behaviour, indicating potential operational dependencies or leader-follower dynamics within bus fleets. Spatial proximity analysis further enriches understanding by scrutinizing the physical closeness of buses along their trajectories (Figure 7).

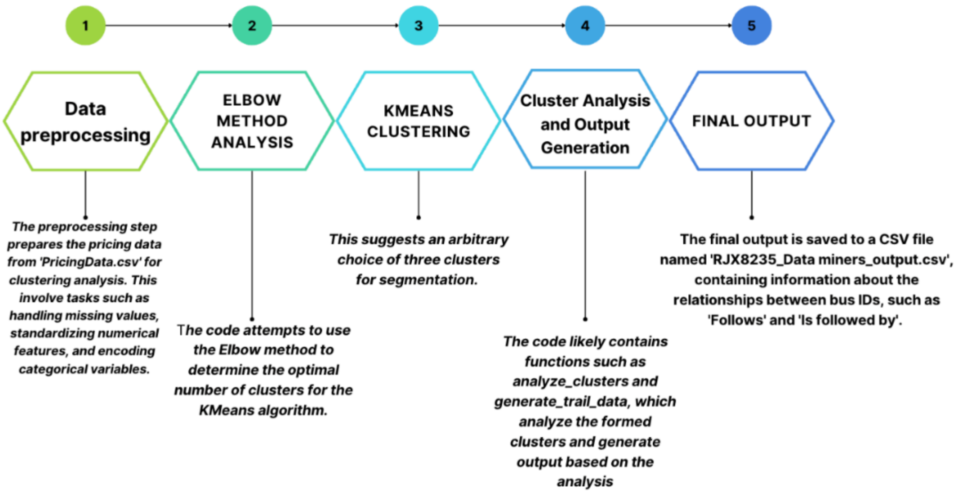

Figure 7: Mind map

Identifying bus convoys, characterized by tight spatial clustering, elucidates instances where buses travel nearby over extended distances. These convoy formations may stem from operational strategies, traffic conditions, or route constraints, presenting opportunities for efficiency improvements or congestion mitigation measures. The analysis unveils multifaceted directional relationships among buses by integrating temporal and spatial perspectives, fostering insights into operational dynamics and network efficiency. Such insights empower transit authorities to optimize route scheduling, enhance service reliability, and implement strategic interventions to improve passenger experience and system performance. Moreover, understanding directional relationships facilitates proactive management of bus fleets, enabling agile responses to disruptions and streamlining operations for enhanced transit services.

6.4. Final Output

The 'RJX8235_Data miners_output.csv' file, representing the culmination of the analysis, is expected to contain crucial insights into the relationships between different bus IDs within the transportation network. The 'Follows' and 'Is followed by' columns serve as vital indicators of directional associations among buses, elucidating which buses trail others and which are trailed. This directional information unveils nuanced behavioural dynamics within the bus fleet, shedding light on operational dependencies, convoy formations, and leader-follower patterns. By discerning these relationships, transit agencies gain actionable intelligence to optimize various aspects of system management. For instance, understanding which buses follow specific routes or maintain close spatial proximity allows for refined scheduling and fleet deployment strategies, minimizing service gaps and enhancing overall reliability. Furthermore, identifying buses that frequently lead or follow others aids in pinpointing potential operational inefficiencies or congestion hotspots, prompting targeted interventions to streamline traffic flow and improve service efficiency. Ultimately, the insights derived from the 'RJX8235_Data miners_output.csv' file empower transit authorities to make data-driven decisions, enhancing system performance, reducing operational costs, and enhancing the overall passenger experience. These findings lay the foundation for proactive and responsive transportation management strategies, ensuring the continued optimization and sustainability of the transit network.

6.5. After using k means clustering, the data is grouped by bus services having similar data

This section imports pandas for data handling, numpy for numerical computations, sklearn.cluster for KMeans clustering, and matplotlib.pyplot for visualization. Subsequently, the code reads data from the CSV file 'PricingData.csv' into a DataFrame named df. This step is crucial for subsequent analysis. The preprocessing phase likely involves handling missing values, scaling features, or encoding categorical variables to prepare the data for clustering. Following data preprocessing, the code aims to employ the Elbow method to determine the optimal number of clusters for KMeans. The code continues by fitting the KMeans algorithm with 3 clusters to the preprocessed data, suggesting an attempt to partition the data into three distinct groups based on similarity (Figure 8).

Bus	seat type	difference	number_of_prices	avg_price	cluster
56f71d6aca7fc3a5db009a462d9e502a	1	92.5358052	1.726190476	904.309524	0
9656d6395d1f3a4497d0871a03366078	1	67.671227	1.853503185	774.210191	0
10598f3c00cb10feb6c5a369ea5d4331	1	10.3902966	2.816742081	760.782805	0
b8d1710db82f66126ca3f540f2ad2f08	1	4.73739566	1.881469115	750.212855	0
f663c68449dda5634f3797fd9c4c3ad6	1	2.92074842	1.454148472	723.481805	0
94e6adf5189531e3a52686bb17c2b2ad	1	2.66232531	2.279503106	781.832298	0
a0243f8c4129f3a468ede17e832a7a9a	1	2.16304167	2.56	803.3	0
5efeca02cbcf91c6230fd894cfae0b1f	1	1.80532294	1.02200489	778.557457	0
c3001e2a3fcf58a3f9881a6635d8765a	1	1.08864413	1.81147541	809.221311	0
f4b9954bf711461b37cab613fdcb8807	1	0.68091639	1.558823529	790.323529	0
d1854bef9a416d20150912c61a1fb9e1	0	0.58285619	1	762.045952	0
78690969ed345320991aba0cd59e0733	1	0.40528183	1.979927007	883.257299	0
28fd7bdc6b80666877c3e9dd41e0ae5a	1	0.35749113	1.723404255	795.744681	0
bae2b9f85a7c3e8eaea1a12ac8be7af2	1	0.21754115	2.481481481	838.123457	0
8a96c2026a520cfd595767af6e5974ef	1	0.0588553	1.901960784	756.895425	0
8239a4d7ab3b3de7711c6cb7748229bf	1	0.01177053	2.799711816	839.18732	0
29b77a9e58d23bf43daa780eb8b7db65	0	-0.0098501	1.552631579	826.631579	0
22f694bbcef788c5f2f4d44ad39fcdbd	1	-0.1479695	2.337190083	821.008815	0
5580f995d6f4d3bcceca7e2db6c77bf7	0.49793218	-0.1589548	1.936311001	748.954163	0
28af5b39ca85472e76714235b77a08c6	1	-0.2006073	2.054901961	830.184314	0
a70bfd3145f3c601b7b8371845be6e57	1	-0.2421484	2.93627451	729.689542	0
bb34a6d5a6c5e806c2bdc7afc63e4397	1	-0.265417	1.016359918	824.642127	0
4f442421f3255cd732698b660003791e	0	-0.5324981	1.598409543	724.572565	0
3b9b3497d23b8212272f1b4c1f70bbfa	1	-0.7357837	2.571428571	867.71164	0
a85170e65d1c03e1ea78595c0d8deb80	1	-1.8589344	2.528673835	814.831541	0
6ebe14c775a983e43b07c55e6b71d77d	1	-1.953082	1.966666667	927.679762	0
09d3a01cf347bce0b92631414af3fea8	0	-2.2716673	1	754.574163	0
ab479dab4a9e6bc3eaefe77a09f027ed	0	-2.5210678	1	747.854031	0
3014ebaddbddfbfc27bf6d8958851aa5	1	-5.4113592	2.41025641	836.118234	0
9d2a5d655e5000921f3591f6cd7908d3	1	-5.8785539	1.901960784	785.294118	0

Figure 8: Output after clustering

This section imports pandas for data handling, numpy for numerical computations, sklearn.cluster for KMeans clustering, and matplotlib.pyplot for visualization. Subsequently, the code reads data from the CSV file 'PricingData.csv' into a DataFrame named df. This step is crucial for subsequent analysis. The preprocessing phase likely involves handling missing values, scaling features, or encoding categorical variables to prepare the data for clustering. Following data preprocessing, the code aims to employ the Elbow method to determine the optimal number of clusters for KMeans. The code continues by fitting the KMeans algorithm with 3 clusters to the preprocessed data, suggesting an attempt to partition the data into three distinct groups based on similarity.

6.6. Discussion

The findings of this study offer valuable insights into the dynamic interplay between buses within transportation networks, unveiling intricate relationships that underpin operational dynamics and system performance. Understanding these bus relationships is pivotal for transit agencies seeking to optimize service delivery, alleviate congestion, and enhance passenger experience. Transit authorities can make informed decisions to improve system reliability, operational efficiency, and overall service quality by comprehending the nuances of bus interactions. At the core of this study is the recognition of the multifaceted nature of bus relationships, which extend beyond mere spatial and temporal proximities. The analysis reveals complex behavioural patterns, including bus-following behaviour, convoy formations, and operational dependencies. By deciphering these patterns, transit agencies gain valuable insights into the underlying dynamics of their bus fleets, enabling them to devise strategies for more effective management and operation of transportation systems. One of the key implications of this study is the potential for enhancing service reliability. Transit agencies can optimize route scheduling and fleet deployment by identifying instances of bus-following behaviour and convoy formations to minimize service gaps and reduce passenger wait times.

Moreover, understanding the temporal dynamics of bus relationships allows agencies to anticipate disruptions and proactively adjust operations to maintain service continuity. For example, real-time data analysis could enable adaptive routing strategies, diverting buses to alleviate congestion or bypass delays, thereby mitigating the impact on service reliability. Reducing congestion is another critical aspect addressed by this study. By recognizing patterns of bus interactions that contribute to congestion, such as convoy formations or inefficient routing, transit agencies can implement measures to optimize traffic flow and alleviate bottlenecks. Strategic adjustments to bus schedules, routes, and stop locations can help distribute passenger loads evenly across the network, reducing overcrowding and improving overall system efficiency. Additionally, leveraging insights from bus relationships enables agencies to coordinate traffic signal prioritization and implement transit signal priority (TSP) strategies, facilitating smoother traffic flow and reducing travel times for buses and other road users. Enhancing passenger experience is a fundamental goal of transportation system management, and the insights gained from this study offer valuable opportunities for achieving this objective.

By optimizing service reliability and reducing congestion, transit agencies can provide passengers with more predictable and efficient travel experiences. Moreover, by identifying and addressing operational inefficiencies, such as bus bunching or irregular headways, agencies can improve service frequency and consistency, enhancing passenger satisfaction and loyalty. Furthermore, leveraging real-time data and predictive analytics enables agencies to provide passengers with up-to-date information on bus arrival times, service disruptions, and alternative routes, empowering them to make informed travel decisions and reducing the frustration associated with uncertainty. Future research in this area could explore the temporal dynamics of bus relationships in greater detail, incorporating real-time data and advanced analytical techniques to adjust operations in response to changing conditions adaptively. By developing dynamic models of bus interactions and leveraging real-time data streams, researchers can devise proactive strategies for managing bus fleets, optimizing routing and scheduling, and mitigating the impact of disruptions.

Additionally, exploring the integration of emerging technologies, such as connected and autonomous vehicles (CAVs) and intelligent transportation systems (ITS), holds promise for further enhancing the efficiency and reliability of bus operations. In conclusion, the findings of this study illuminate the dynamic nature of bus relationships and their profound implications for transportation system management. Transit agencies can make informed decisions to improve service reliability, reduce congestion, and enhance passenger experience by understanding how buses interact within the network. Leveraging insights from bus relationships enables agencies to optimize route scheduling, alleviate bottlenecks, and provide passengers with more predictable and efficient travel experiences. Future research could explore the temporal dynamics of bus relationships and develop adaptive strategies for managing bus fleets in response to changing conditions, ultimately advancing the efficiency and effectiveness of public transportation systems.

7. Conclusion

In conclusion, our investigation into the price dynamics of bus services through the lens of machine learning algorithms has yielded promising results and important implications for both service providers and consumers. Incorporating machine learning, particularly predictive modelling, is a pivotal advancement in the bus service industry. This sophisticated technology facilitates accurate price forecasting and empowers operators to implement dynamic pricing strategies that respond in real-time to fluctuations in demand. The result is an adaptive pricing mechanism that ensures optimal revenue generation for service providers, marking a paradigm shift in traditional pricing approaches. Furthermore, the data-driven approach afforded by machine learning extends beyond pricing dynamics. It delves into understanding and predicting consumer behaviour, offering operators a nuanced understanding of passenger preferences. With such insights, operators can personalize offerings, tailoring services to meet individual needs and preferences. This personalization enhances customer satisfaction and fosters brand loyalty, positioning bus services as more customer-centric entities. The study also sheds light on the potential for optimizing operational efficiency through machine learning applications. Operators can make informed decisions regarding route adjustments and scheduling by strategically analyzing historical demand patterns. This proactive approach enables operators to align their services more closely with actual demand, minimizing inefficiencies and maximizing resource utilization. Consequently, machine learning becomes crucial in pursuing operational excellence within the bus service sector. Integrating machine learning into pricing strategies provides a competitive advantage and establishes a flexible framework for operators to adapt swiftly to evolving market conditions. Analyzing vast datasets in real-time empowers operators to make informed decisions on pricing adjustments and service modifications, ensuring a proactive response to market dynamics.

Acknowledgement: The support of all my co-authors is highly appreciated.

Data Availability Statement: The research contains data related to Price Dynamics of Bus Services analytics and associated metrics. The data consists of views and dates as parameters.

Funding Statement: No funding has been obtained to help prepare this manuscript and research work.

Conflicts of Interest Statement: No conflicts of interest have been declared by the author(s). Citations and references are mentioned in the information used.

Ethics and Consent Statement: The consent was obtained from the organization and individual participants during data collection, and ethical approval and participant consent were received.

References

1. S. Thomas, B. Pathak, and P. Vyas, "The growth of online bus ticketing industry: RedBus route to success in the Indian market," Int. J. Bus. Manag., vol. 9, no. 11, pp. 1-17, 2014.
2. M. Banoula, "K-means clustering algorithm: Applications, types, and demos [updated]," Simplilearn.com, 22-Apr-2020. [Online]. Available: https://www.simplilearn.com/tutorials/machine-learning-tutorial/k-means-clustering-algorithm. [Accessed: 24-May-2023].
3. A. Al-Masri, "How does k-Means Clustering in Machine Learning work?," Towards Data Science, 14-May-2019. [Online]. Available: https://towardsdatascience.com/how-does-k-means-clustering-in-machine-learning-work-fdaaaf5acfa0. [Accessed: 21-May-2023].
4. "Kaggle: Your machine learning and data science community," Kaggle.com. [Online]. Available: https://www.kaggle.com/. [Accessed: 21-May-2023].
5. B. Saji, "Elbow method for finding the optimal number of clusters in K-means," Analytics Vidhya, 20-Jan-2021. [Online]. Available: https://www.analyticsvidhya.com/blog/2021/01/in-depth-intuition-of-k-means-clustering-algorithm-in-machine-learning. [Accessed: 21-May-2023].
6. D. Mangione, "6 steps to implementing dynamic pricing in a bus company," Turnit.com. [Online]. Available: https://blog.turnit.com/6-steps-to-implementing-dynamic-pricing-in-bus-company. [Accessed: 21-May-2023].
7. S. Sindu, "Dynamic pricing model using price multipliers for online bus ticket booking," krj, vol. 1, no. 2, pp. 23–30, 2014.
8. R. A. Kadir, Y. Shima, R. Sulaiman, and F. Ali, "Clustering of public transport operation using K-means," in 2018 IEEE 3rd International Conference on Big Data Analysis (ICBDA), Shanghai, China, 2018.
9. Y. Zhang, W. Fan, H. Chen, H. Sheng, and Z. Wu, "Extending online travel agency with adaptive reservations," in On the Move to Meaningful Internet Systems 2007: CoopIS, DOA, ODBASE, GADA, and IS, Berlin, Heidelberg: Springer Berlin Heidelberg, pp. 285–299, 2007.
10. A. Sulaiman, J. Ng, and S. Mohezar, "E-ticketing as a New Way of buying tickets: Malaysian perceptions," J. Soc. Sci., vol. 17, no. 2, pp. 149–157, 2008.
11. A. K. Jain, M. N. Murty, and P. J. Flynn, "Data Clustering: A Review," ACM Computing Surveys, vol. 31, no. 3, pp. 264–323, 1999.
12. H. Zhou, "K-Means Clustering," in Learn Data Mining Through Excel, pp. 37–52, 2023.
13. D. Yang, J. Myung, S. Lee, and J. Song, "Comparison of k-Mean Clustering with Missing Data," The Korean Data Analysis Society, vol. 25, pp. 2131–2142, 2023.
14. R. Pan, C. Zhong, and J. Qian, "Balanced Fair K-Means Clustering," IEEE Transactions on Industrial Informatics, vol. 99, no.1, pp. 1–10, 2023.