iJRASET
International Journal For Research in Applied Science and Engineering Technology



INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 14 **Issue:** V **Month of publication:** May 2026

DOI: https://doi.org/10.22214/ijraset.2026.82051

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 14 Issue V May 2026- Available at www.ijraset.com

Ridership Analysis of Cyclistic Bike Users

Shruti Yadav[1], Asst. Prof. Mamta Ramesh Devre[2]
Department of Information Technology and Mathematics, The S.I.A. College of Higher Education, Thane, India

Abstract: ***Urban mobility systems generate large volumes of ride data, yet many organizations struggle to translate this data into actionable business insights. Bike-sharing services, in particular, face challenges in understanding user behavior and converting short-term users into long-term subscribers. This study analyzes Cyclistic bike-share ride data to identify behavioral differences between casual riders and annual members using a structured exploratory data analysis (EDA) approach. Monthly trip records were integrated, cleaned, and processed using R-based analytical tools to ensure data quality and consistency. The analysis focuses on temporal patterns (month, weekday, hour, and time of day), bike type preferences, and station usage trends. Results indicate that annual members predominantly use bikes for weekday commuting and exhibit consistent usage patterns, while casual riders are more active during weekends and leisure hours, reflecting recreational behavior. Distinct preferences in bike types and station locations further highlight segmentation between the two groups. Based on these findings, the study proposes targeted marketing strategies, including weekend promotions, location-based campaigns, and seasonal offers, to improve conversion of casual riders into annual members. The research demonstrates how data preprocessing, feature engineering, and visualization can transform raw operational data into meaningful insights. The proposed analytical framework provides a scalable and practical approach for supporting customer growth, optimizing resource allocation, and enabling data-driven decision-making in bike-sharing systems.***

Keywords: Bike-Sharing, Data Analytics, Exploratory Data Analysis (EDA), Customer Segmentation, Urban Mobility, RStudio, Tableau, Business Intelligence

I. INTRODUCTION

The rapid expansion of urban mobility solutions has significantly transformed how people navigate metropolitan environments. Shared transportation systems such as bike-sharing, scooter-sharing, and app-based rentals have emerged as cost-effective, flexible, and environmentally sustainable alternatives to traditional commuting [1],[2]. Among these, bike-sharing systems have gained widespread adoption due to their accessibility and suitability for short-distance travel, daily commuting, and recreational use [3]. As cities continue to promote sustainable mobility, data generated from these systems provides valuable opportunities to understand user behavior and optimize service delivery [4].

Cyclistic, a bike-sharing service provider, operates through multiple pricing models, including single-ride passes, day passes, and annual memberships. Its user base is broadly divided into casual riders, who use the service occasionally, and annual members, who subscribe for long-term usage. While both groups utilize the same infrastructure, their behavioral patterns differ significantly in terms of ride frequency, duration, timing, and preferences. From a business standpoint, annual members contribute more stable and predictable revenue streams, making their acquisition and retention a strategic priority [5]. However, the lack of a clear, data-driven understanding of how casual riders differ from members limits the effectiveness of marketing and conversion strategies [6].

The increasing availability of large-scale ride data enables organizations to apply data analytics techniques to uncover meaningful insights into customer behavior. Ride datasets containing attributes such as trip duration, time of travel, bike type, and station usage allow for comprehensive exploratory analysis. By leveraging analytical tools such as RStudio and Tableau, this study examines behavioral differences between casual riders and annual members across multiple dimensions, including temporal usage patterns, ride frequency, and bike preferences.

The primary objective of this research is to transform raw ride data into actionable insights that support strategic decision-making. Through systematic data cleaning, feature engineering, and visualization, the study identifies trends such as peak usage periods, differences in weekday and weekend activity, and variations in bike-type preferences. These insights provide a foundation for targeted marketing strategies aimed at increasing the conversion of casual riders into annual members. In a broader context, this study highlights the importance of data-driven approaches in modern business environments. By analyzing user behavior and identifying key patterns, organizations can optimize operations, enhance customer experience, and improve profitability. The findings not only contribute to Cyclistic's membership growth strategy but also demonstrate the practical application of data analytics in solving real-world business problems within the domain of urban mobility.

©IJRASET: All Rights are Reserved

II. PROBLEM IDENTIFICATION

Despite the rapid growth of shared mobility services, organizations continue to struggle with effectively leveraging ride data to support strategic decision-making. In the case of Cyclistic, a significant proportion of users are casual riders who generate short-term revenue but do not contribute to long-term business value comparable to annual members. While the company aims to increase its membership base, it lacks a clear, data-driven understanding of how casual riders differ from annual members in terms of usage behavior, frequency, and preferences. Without this insight, marketing strategies, pricing models, and promotional campaigns remain largely assumption-driven and risk being ineffective.

Existing approaches to data analysis further compound the problem. Manual processing of large-scale ride datasets is time-consuming, error-prone, and difficult to scale. Basic statistical summaries fail to capture deeper behavioral patterns such as seasonal demand fluctuations, peak usage hours, and differences in bike or station preferences. As a result, meaningful segmentation between customer groups is not achieved, limiting the ability to design targeted interventions. Additionally, static reporting methods restrict interactive exploration of data, reducing the potential to uncover hidden trends and insights.

Data quality issues, including missing values, duplicate records, and invalid entries, also pose significant challenges, as they can distort analytical outcomes if not properly handled. The lack of a structured and systematic analytical framework leads to weak pattern detection, inefficient processing, and reduced reliability of insights. Consequently, the absence of a robust data-driven system undermines both operational efficiency and marketing effectiveness.

To address these challenges, there is a need for a comprehensive analytics approach that can efficiently process large datasets, ensure data quality, and generate actionable insights. By systematically analyzing ride behavior and identifying key differences between casual riders and annual members, the system can support evidence-based strategies aimed at improving membership conversion and overall business performance [7].

III. LITERATURE SURVEY

A. Literature Review

The analysis of large-scale urban mobility data has gained significant attention in recent years, particularly in the context of shared transportation systems such as bike-sharing platforms. Studies on bike-sharing analytics emphasize the importance of data-driven decision-making to understand user behavior, optimize operations, and improve customer retention. Traditional data analysis approaches, however, often rely on basic aggregation techniques, which limit the ability to uncover deeper behavioral patterns among different user segments.

Early analytical approaches in mobility systems primarily utilized spreadsheet-based analysis and static reporting tools. These methods focused on descriptive statistics such as total ride counts, revenue summaries, and basic usage metrics. While useful for operational monitoring, such approaches fail to capture temporal patterns, user segmentation, and behavioral differences that are critical for strategic decision-making. In particular, the inability to distinguish between casual and long-term users restricts the effectiveness of targeted marketing strategies.

With the advancement of data analytics technologies, tools such as R and its ecosystem (including Tidyverse, Data.Table, and Lubridate) have enabled more sophisticated data processing and transformation capabilities [8]. These tools support efficient handling of large datasets, feature engineering, and time-based analysis, allowing researchers to extract meaningful insights from raw ride data. The use of date-time processing libraries has been particularly important in identifying patterns across months, weekdays, and hours, which are essential for understanding user behavior in mobility systems [9].

Visualization technologies have further enhanced the analytical process. Tools like Tableau and visualization libraries such as ggplot2 provide interactive and intuitive dashboards that facilitate exploration of complex datasets. Unlike static reports, these platforms enable dynamic filtering, comparison across user groups, and real-time interpretation of trends. Visualization-driven analysis has been shown to significantly improve the clarity and usability of insights, especially for business stakeholders.

Recent studies in customer analytics also highlight the importance of segmentation in improving marketing effectiveness. By categorizing users based on behavior, frequency, and preferences, organizations can design targeted strategies to increase engagement and conversion rates. In the context of bike-sharing systems, segmentation between casual riders and annual members is particularly critical, as these groups exhibit distinct usage patterns and revenue contributions.

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |

B. Literature Summary

The reviewed approaches indicate a clear transition from basic reporting systems to advanced data analytics frameworks in mobility services. Traditional methods provide only surface-level insights, whereas modern analytical tools enable deeper exploration of user behavior, temporal trends, and service usage patterns. The integration of data processing tools like RStudio with visualization platforms such as Tableau allows for scalable, efficient, and interpretable analysis of large ride datasets.

A key takeaway from the literature is the importance of combining data cleaning, feature engineering, and visualization to derive actionable insights. Studies consistently show that understanding differences between user segments particularly casual users and long-term subscribers is essential for designing effective business strategies. Additionally, interactive dashboards and visual analytics play a crucial role in translating complex data into meaningful business decisions.

C. Research Gap

Despite advancements in data analytics for mobility systems, several gaps remain. First, many existing approaches rely heavily on basic summaries and fail to perform detailed behavioral segmentation between different user groups. This limits the ability to design targeted marketing strategies aimed at increasing long-term customer retention.

Second, traditional systems often lack efficient handling of large-scale datasets, leading to time-consuming processing and potential data quality issues such as missing values and duplicates. Without proper preprocessing and validation, the reliability of insights may be compromised.

Third, there is limited integration between data processing and visualization tools in many analytical workflows. Static reporting systems do not provide the flexibility required for exploring patterns across multiple dimensions such as time, location, and user type.

Finally, most existing analyses focus on descriptive insights without translating them into actionable business strategies. There is a need for a structured analytical framework that not only identifies behavioral differences but also supports decision-making for customer conversion and retention.

The present study addresses these gaps by implementing a comprehensive data analytics approach that integrates data cleaning, feature engineering, segmentation, and interactive visualization. By focusing on behavioral differences between casual riders and annual members, the study provides actionable insights to support membership conversion strategies and improve business performance in bike-sharing systems [10].

IV. RESEARCH METHODOLOGY

A. Data Selection and Preparation Criteria

The analytical framework for this study was designed based on three key criteria to ensure meaningful, reliable, and business-relevant insights from the Cyclistic ride dataset [11]:

1) Relevance to business objectives: The dataset and variables were selected based on their ability to explain customer behavior and support the core objective of converting casual riders into annual members. Key attributes such as ride timestamps, membership type, bike category, and station usage were prioritized, as they directly reflect usage patterns, frequency, and customer engagement.
2) Analytical depth and scalability: The methodology emphasizes the use of scalable data processing and analytical techniques capable of handling large datasets (over 4.5 million records). Tools such as RStudio and libraries including Tidyverse, Data.Table, and Lubridate, etc. were utilized to enable efficient data manipulation, time-based feature extraction, and high performance computation. This ensures that the analysis moves beyond basic summaries and towards deeper behavioral insights.
3) Practical applicability of insights: The workflow was structured to ensure that analytical outputs are not only descriptive but also actionable. Visualization tools such as Tableau were incorporated to transform processed data into interactive dashboards, enabling clear interpretation and direct application in marketing and business strategy decisions.

These criteria ensured that the methodology integrates data reliability, analytical rigor, and business usability, forming a comprehensive approach to mobility data analysis.

B. Method of Analysis

The study follows a structured data analytics pipeline, transforming raw ride data into actionable insights through multiple stages of processing and analysis:

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |

1) Data Collection and Integration: Monthly ride datasets (August 2020 to July 2021) were collected and consolidated into a unified dataset. This step ensures completeness and continuity of data across time periods.
2) Data Cleaning and Validation: Data preprocessing techniques were applied to improve quality and reliability. Missing values were removed, duplicate records were eliminated, and invalid ride entries (e.g., zero or negative durations) were filtered out. This step ensures that subsequent analysis is based on accurate and consistent data.
3) Feature Engineering: New variables were derived from raw data to enhance analytical depth. These include ride length, month, weekday, hour, season, and time-of-day categories. Such transformations enable temporal and behavioral segmentation of users.
4) Exploratory Data Analysis (EDA): Comparative analysis was conducted between casual riders and annual members using statistical summaries and aggregated counts. Key dimensions analyzed include monthly trends, weekly patterns, hourly usage, bike preferences, and station popularity.
5) Visualization and Interpretation: Data visualization techniques were employed using Tableau and R-based plotting libraries to present insights in an interpretable format. Charts, dashboards, and comparative plots were used to highlight differences between user groups and identify trends.
6) Insight Generation and Recommendation: The final stage involved translating analytical findings into business insights. Behavioral differences between rider groups were interpreted to propose targeted strategies for increasing annual membership conversion.

C. *Comparison and Analytical Outcomes*

The structured methodology enabled a systematic comparison between casual riders and annual members, revealing distinct behavioral patterns across multiple dimensions:

1) Temporal behavior differences: Casual riders exhibit higher activity during weekends and afternoon/evening hours, indicating leisure-oriented usage. In contrast, annual members show consistent weekday and morning usage patterns, reflecting commuting behavior.
2) Seasonal and demand patterns: Both user groups demonstrate peak activity during summer months (June–August), highlighting the influence of seasonal factors on ridership demand.
3) Product preference variation: Casual riders show a stronger preference for docked bikes, whereas annual members predominantly use classic bikes, indicating differing usage needs and familiarity levels.
4) Spatial usage patterns: Station-level analysis reveals that casual riders concentrate around tourist and central locations, while members primarily use stations associated with residential and business areas.

From this comparative analysis, three major patterns emerge:

a) Behavioral segmentation is strongly linked to time-based variables such as weekday and hour, making temporal targeting highly effective for marketing strategies.
b) Usage intensity among casual riders indicates untapped potential for conversion, particularly during peak periods such as weekends and summer months.
c) Differences in bike and station preferences highlight the need for personalized and location-based marketing approaches.

These findings validate the effectiveness of the proposed analytical framework. By systematically transforming raw ride data into structured insights, the methodology provides a robust foundation for data-driven decision-making and targeted membership conversion strategies in bike-sharing systems.

V. DISCUSSION

A. *Synthesis of Analytical Findings*

The analysis of large-scale ride data reveals a clear and consistent behavioral segmentation between casual riders and annual members. Across temporal, spatial, and usage dimensions, both groups exhibit distinct patterns that reflect different underlying motivations. Annual members demonstrate structured and routine-based usage, characterized by higher activity during weekdays, particularly in the morning and evening hours, indicating strong alignment with daily commuting needs. In contrast, casual riders show irregular but concentrated usage patterns, with peak activity during weekends, afternoons, and summer months, suggesting recreational and leisure-oriented behavior. Seasonal trends further reinforce these differences, as both user groups exhibit increased activity during June, July, and August. However, the intensity of usage among casual riders during these months highlights their responsiveness to favorable weather and tourism cycles.

©IJRASET: All Rights are Reserved

Additionally, differences in bike preferences docked bikes for casual users and classic bikes for members indicate variations in convenience, accessibility, and familiarity with the system. Station-level analysis also supports this segmentation, with casual riders favoring central or tourist-heavy locations, while members predominantly use stations associated with residential and business districts.

B. Business Implications

These findings have direct implications for strategic decision-making. Contrary to the assumption that casual riders represent low-value users, the analysis shows that they are highly active during predictable periods and locations. This positions them as a high-potential segment for conversion into annual members. The presence of strong temporal and spatial patterns enables targeted marketing interventions rather than broad, ineffective campaigns.

For instance, the concentration of casual rider activity during weekends and evenings provides an opportunity to deploy time-specific promotional strategies. Similarly, the high usage of tourist-centric stations suggests that location-based marketing, such as in-app offers or QR-based promotions at key stations, could significantly influence conversion rates. Seasonal demand patterns indicate that summer months are particularly suitable for intensive membership campaigns, as user engagement is naturally elevated.

C. Operational and Analytical Considerations

From a methodological standpoint, the study demonstrates the effectiveness of exploratory data analysis in extracting actionable insights from large datasets without relying on predictive modeling. The processing of over 4.5 million ride records, combined with rigorous data cleaning and validation, ensures the reliability of the findings. The use of visualization tools further enhances interpretability, allowing complex behavioral patterns to be communicated clearly.

However, the absence of demographic and contextual variables limits the depth of behavioral interpretation. Factors such as age, income, occupation, weather conditions, and special events could significantly influence ride behavior but are not captured in the dataset. As a result, the analysis focuses primarily on observable usage patterns rather than underlying causal drivers.

D. Limitations and Challenges

Despite the robustness of the dataset, several limitations must be acknowledged. The analysis is descriptive in nature and does not establish causal relationships between variables. Additionally, the lack of customer feedback data restricts the ability to understand user motivations beyond inferred behavior. External factors such as weather variations, city events, or policy changes are also not incorporated, which may affect the generalizability of the results.

Another challenge lies in translating analytical insights into real-world business actions. While patterns are clearly identified, their successful implementation depends on organizational capabilities, marketing execution, and user responsiveness. Without controlled experimentation or A/B testing, the effectiveness of recommended strategies cannot be fully validated.

E. Strategic Outlook

The study highlights the critical role of data-driven decision-making in modern mobility services. By identifying when, where, and how different user groups interact with the system, organizations can design targeted interventions that maximize conversion and customer retention. Future work should focus on integrating predictive analytics, incorporating external data sources, and validating strategies through real-world experimentation.

Overall, the discussion establishes that behavioral segmentation, when supported by large-scale data analysis, provides a strong foundation for strategic planning. The insights generated not only address the immediate business objective of increasing membership conversions but also contribute to a broader understanding of user behavior in shared mobility systems.

VI. CONCLUSION

The growing reliance on data-driven decision-making in urban mobility highlights the need for systematic analysis of user behavior to support business strategy. This study demonstrates how large-scale ride data can be effectively utilized to understand behavioral differences between customer segments in a bike-sharing system. Through comprehensive exploratory data analysis, clear distinctions were identified between annual members and casual riders in terms of usage patterns, temporal behavior, bike preferences, and station selection.

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |

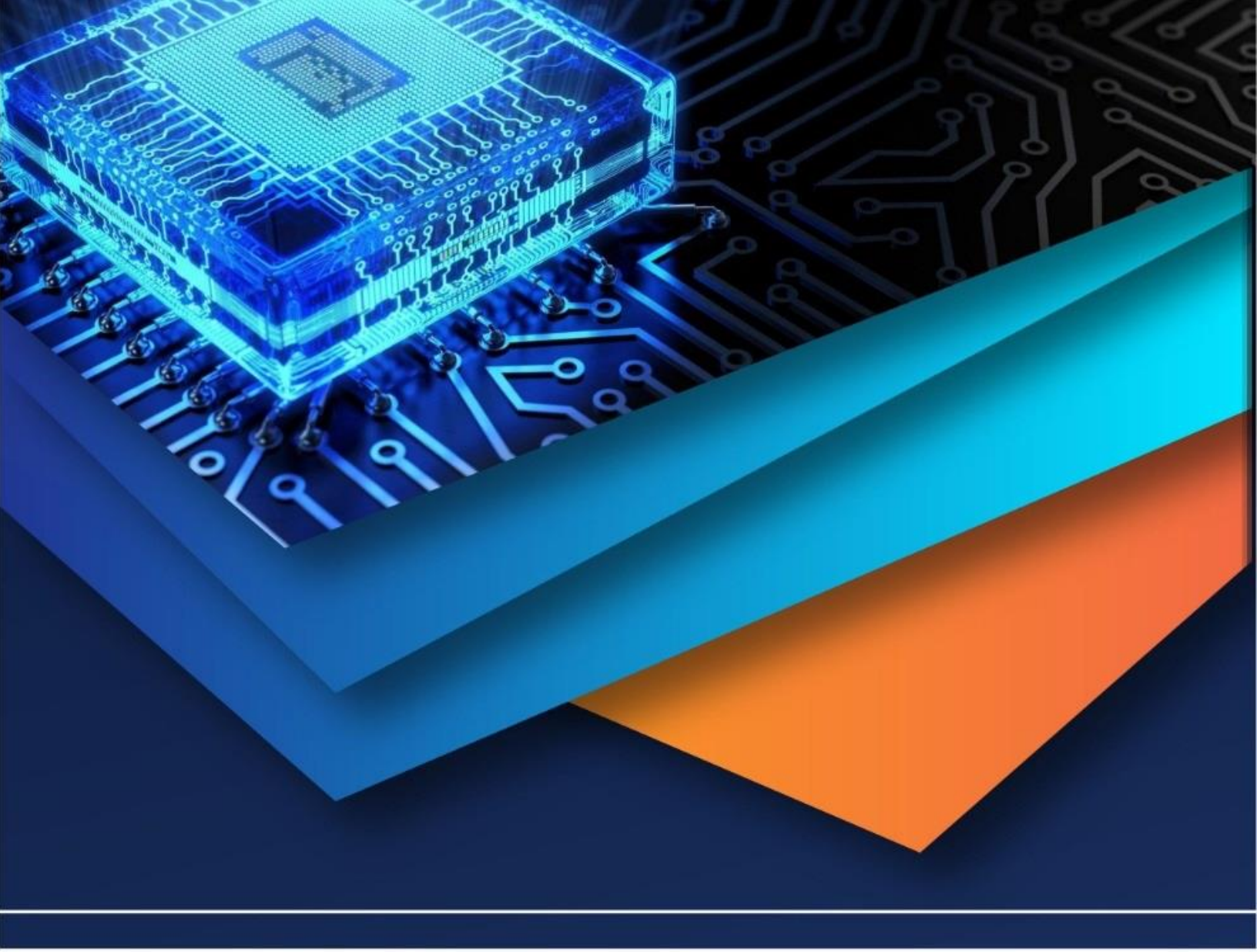
The findings reveal that annual members exhibit consistent, routine-based usage aligned with weekday commuting and peak-hour travel, while casual riders display irregular, leisure-oriented behavior concentrated during weekends, afternoons, and seasonal peaks. These differences indicate that casual riders are not merely occasional users but represent a high-potential segment for conversion into long-term subscribers. Patterns such as increased summer demand, preference for specific stations, and distinct bike usage further reinforce the opportunity for targeted, data-driven marketing interventions.

Despite certain limitations, including the absence of demographic variables, external influencing factors, and predictive modeling, the study successfully transforms raw operational data into actionable business insights. It establishes a strong analytical foundation for future enhancements, including predictive conversion models, customer segmentation, and real-time decision-support systems. Overall, the research underscores the strategic value of exploratory analytics in identifying growth opportunities and optimizing service delivery. By leveraging behavioral insights and aligning them with targeted campaigns, bike-sharing services like Cyclistic can improve membership conversion, enhance customer engagement, and achieve sustainable long-term growth in competitive urban mobility markets.

REFERENCES

[1] Pucher, J., & Buehler, R. (2008). Making Cycling Irresistible: Lessons from The Netherlands, Denmark and Germany. *Transport Reviews*, *28*(4), 495–528. https://doi.org/10.1080/01441640701806612
[2] S. Shaheen, S. Guzman, and H. Zhang, "Bikesharing in Europe, the Americas, and Asia," *Transportation Research Record*, 2010. http://dx.doi.org/10.3141/2143-20
[3] Oliver O'Brien, James Cheshire, Michael Batty, Mining bicycle sharing data for generating insights into sustainable transport systems, Journal of Transport Geography, Volume 34, 2014, Pages 262-273, ISSN 0966-6923, https://doi.org/10.1016/j.jtrangeo.2013.06.007.
[4] Data Science for Business: What You Need to Know About Data Mining and Data-Analytic Thinking by Foster Provost, Tom Fawcett.
[5] Fishman, E. (2016). Bikeshare: A Review of Recent Literature. *Transport Reviews*, *36*(1), 92–113. https://doi.org/10.1080/01441647.2015.1033036
[6] An Introduction to Statistical Learning by Gareth James , Daniela Witten , Trevor Hastie , Robert Tibshirani
[7] *Data Mining: Concepts and Techniques* by J. Han, M. Kamber, and J. Pei.
[8] ggplot2: Elegant Graphics for Data Analysis by Hadley Wickham
[9] Grolemund, G., & Wickham, H. (2011). Dates and Times Made Easy with lubridate. *Journal of Statistical Software*, *40*(3), 1–25. https://doi.org/10.18637/jss.v040.i03
[10] Kimball, R., & Ross, M. (2013). The Data Warehouse Toolkit. Wiley.
[11] Google Data Analytics Professional Certificate. Capstone Case Study: Cyclistic Bike-Share Analysis. Coursera.

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089 (24*7 Support on Whatsapp)