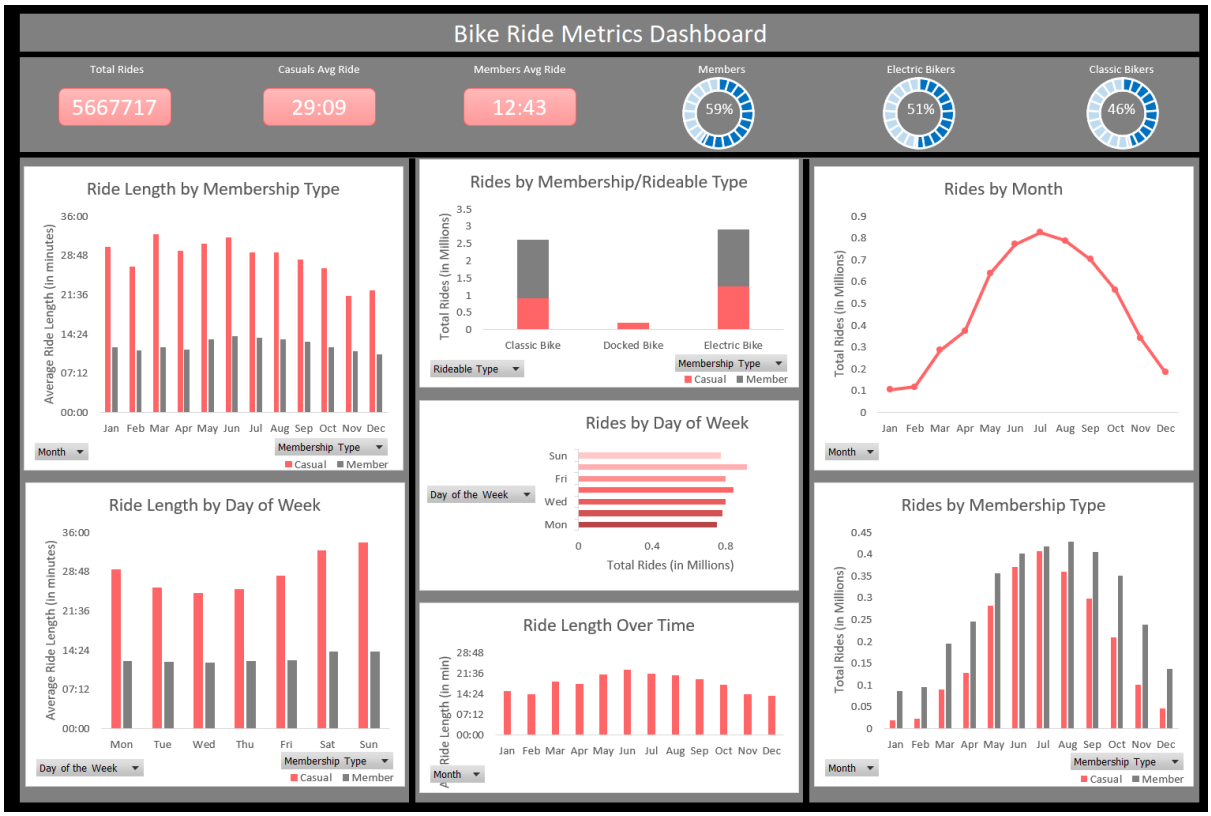
BIKE ANALYTICS EXPLORATION PORTFOLIO PROJECT

Investigating Bike Sharing Data for Strategic Insights



by
Sachin Dahiya
Master of Applied Statistics, Louisiana State University, 2023

Table of Contents

Abstract

The Bike Analytics Exploration Portfolio Project delves into the rich dataset of Cyclistic, a bike-share company in Chicago, to uncover valuable insights into the behaviour and preferences of casual riders and annual members. The project's primary aim is to identify strategic opportunities for maximising annual memberships and improving the company's long-term success.

Using Cyclistic's historical trip data from January to December 2022, this analysis explores various aspects of bike usage, including ride lengths, rideable types, frequency of rides by membership type and day of the week, and the potential impact of digital media on member conversions. Key findings include significant differences in ride lengths between casual riders and annual members, with casual riders averaging 29:09 minutes per ride compared to members' average of 12:43 minutes.

Insights into rider behaviour by day of the week indicate that Saturdays and Sundays attract the longest rides, averaging between 32:37 and 34:03 minutes for casual riders, while members show slightly shorter rides on these days. The analysis also highlights the potential of digital media as a catalyst for converting casual riders into annual members, emphasising the need for targeted marketing strategies.

The project's results provide actionable recommendations for Cyclistic, such as leveraging the popularity of electric bikes among casual riders, focusing on peak months for targeted marketing campaigns, and optimising digital media channels to encourage member conversions. By implementing these recommendations, Cyclistic can enhance its marketing efforts, increase annual memberships, and ultimately drive sustainable growth.

Keywords: Bike analytics, bike-share, Cyclistic, casual riders, annual members, ride lengths, rideable types, digital media, marketing strategies, membership conversions, data analysis.

1. Introduction

1.1. Project Overview

The Bike Analytics Exploration Portfolio Project aims to analyse and extract insights from Cyclistic's historical trip data to gain a deeper understanding of how casual riders and annual members utilise the bike-sharing service differently. By examining patterns and trends within the dataset, the project seeks to identify strategic opportunities that can help Cyclistic maximise annual memberships and drive future growth.

1.2. Objective of the Project

The primary objective of this project is to investigate the behavioural differences between casual riders and annual members of Cyclistic. By analysing various metrics such as ride lengths, rideable types, frequency of rides by membership type and day of the week, and the potential impact of digital media, the project aims to address the following research questions:

1. How do annual members and casual riders use Cyclistic bikes differently?
2. Why would casual riders consider purchasing Cyclistic annual memberships?
3. How can Cyclistic utilise digital media to influence casual riders to become members?

The significance of this project lies in several key areas. Firstly, maximising annual memberships is crucial for Cyclistic's future success. By understanding the differences between casual riders and annual members, the company can tailor its marketing efforts to effectively target and convert casual riders into long-term members. This strategic approach will drive growth and profitability for Cyclistic.

Secondly, analysing the data will enable Cyclistic to enhance its marketing strategies. By gaining insights into the usage patterns and preferences of casual riders and annual members, the company can develop more targeted and personalised marketing campaigns. This

will allow Cyclistic to communicate the value of annual memberships more effectively, increasing the likelihood of conversion.

Furthermore, the project's findings will contribute to improving the overall customer experience. By understanding how casual riders and annual members utilise the bikes and stations, Cyclistic can optimise bike availability, station locations, and rideable options to better meet the needs of different rider segments. This will result in a more satisfying and tailored experience for customers.

Lastly, the project aims to explore the role of digital media in influencing casual riders' decision-making process. By studying how digital media can be utilised to attract and convert casual riders into members, Cyclistic can develop effective digital marketing strategies. This will enable the company to leverage the power of digital channels to reach a wider audience and drive membership growth.

Overall, the objective and significance of this project revolve around uncovering valuable insights that will inform strategic decisions, enhance the customer experience, and drive increased membership conversions for Cyclistic. By aligning its strategies with the preferences and behaviours of casual riders and annual members, Cyclistic can position itself as a leading bike-share provider and achieve sustainable growth in the competitive market.

1.3. Data and Bike Company Description

Cyclistic is a prominent bike-share program in Chicago, offering a diverse range of over 5,800 bicycles and 600 docking stations. In addition to traditional bikes, Cyclistic also provides options such as reclining bikes, hand tricycles, and cargo bikes, making the service more inclusive for individuals with disabilities and those who require specialised bike options. The program has gained popularity since its successful launch in 2016 and currently operates a geotracked fleet of 5,824 bicycles across 692 stations in Chicago.

Cyclistic's marketing strategy has focused on building general awareness and attracting a broad range of consumers through flexible pricing plans, including single-ride passes, full-day passes, and annual memberships. The finance analysts at Cyclistic have determined that annual members contribute significantly more to the company's profitability compared to casual riders. As a result, the project aims to explore how Cyclistic can convert casual riders into annual members by gaining insights into their preferences, motivations, and behaviours.

The project will utilise Cyclistic's publicly available historical trip data for the year 2022, obtained from Motivate International Inc. The dataset consists of monthly data with information on ride details such as ride IDs, rideable types (electric bikes, classic bikes, docked bikes), start and end timestamps, station locations, and membership status (casuals or members). By analysing this comprehensive dataset, we can uncover valuable insights that will inform strategic decisions and marketing efforts for Cyclistic.

2. Data Collection and Preparation

2.1. Data Source

The primary data source for this project is Cyclistic's historical trip data, which is publicly available and provided by Motivate International Inc. The dataset includes information on ride_id, rideable_type, started_at, ended_at, start_station_name, start_station_id, end_station_name, end_station_id, start_lat, start_lng, end_lng, and member_casual. The data spans the entire year of 2022 and is divided into monthly datasets.

2.2. Data Preparation and Cleaning

Before conducting the analysis, the data underwent a series of preparation and cleaning steps. This involved handling missing values, ensuring data consistency, and creating additional variables to enhance analysis. For instance, a new variable named ride_length was calculated by finding the difference between the ended_at and started_at columns to determine the length of each ride. Additionally, a variable named "day of the week" was created to identify the specific day when each ride started.

2.3. Data Analysis Tools Used

To analyse the data and derive insights, various tools were utilised in this project:

2.3.1. Excel

Excel was employed to create a comprehensive dashboard that visualises key performance indicators (KPIs) and graphs. The dashboard presents metrics such as total rides, average ride length for casual riders and members, rideable type distribution, and other relevant information. It provides a clear overview of the data and facilitates easy interpretation.

2.3.2. R

R, a programming language for statistical analysis, was used to perform in-depth data analysis. Packages such as "hms" were utilised to handle time data for the variable ride_length, "tidyverse" to summarise data by membership type and day of the week, and

"epiDisplay" to identify the busiest day of the week in terms of total rides. R allowed for advanced statistical calculations and data manipulation to extract meaningful insights.

2.3.3. MySQL

MySQL, a relational database management system, was employed to import and analyse the data for each month individually. By leveraging SQL queries, the data was analysed to obtain insights into various aspects such as ride frequency, rideable type preferences, and more. MySQL provided a structured environment to query and manipulate the data efficiently.

By employing a combination of Excel, R, and MySQL, this project utilised a comprehensive methodology to analyse the Cyclistic trip data effectively. These tools allowed for data cleaning, visualisation, statistical analysis, and database querying, enabling the extraction of valuable insights for the subsequent stages of the project.

3. Project Findings and Recommendations

3.1. Analysis of Annual Members and Casual Riders

The analysis of Cyclistic's historical trip data revealed significant differences in the behaviour and preferences of annual members and casual riders. Casual riders exhibited longer average ride lengths compared to annual members, indicating a greater inclination towards leisurely rides. On the other hand, annual members showed shorter ride lengths, suggesting more utilitarian use of the bikes for commuting purposes. Understanding these differences provides valuable insights for developing targeted marketing strategies.

3.2. Factors Influencing Casual Riders to Purchase Annual Memberships

Through the analysis, several factors were identified that could influence casual riders to purchase annual memberships. These factors include the convenience of bike availability, the cost-effectiveness of long-term membership, the potential for enhanced customer benefits, and the ability to customise ride options. By emphasising these factors in marketing campaigns and highlighting the advantages of annual memberships, Cyclistic can incentivise casual riders to convert into loyal members.

3.3. Leveraging Digital Media for Converting Casual Riders into Members

The analysis also revealed the potential of digital media in influencing casual riders to become annual members. By leveraging digital marketing channels such as social media, targeted advertisements, and personalised promotions, Cyclistic can effectively communicate the value proposition of annual memberships to the right audience. Utilising data-driven strategies and engaging content, the company can capture the attention of casual riders and encourage them to make the transition.

3.4. Recommendations for Marketing Strategies

Based on the findings, several recommendations can be made to optimise Cyclistic's marketing strategies:

3.4.1. Tailored Messaging

Develop targeted marketing campaigns that highlight the unique benefits of annual memberships, such as convenience, cost-effectiveness, and customer rewards. Craft persuasive messages that resonate with casual riders' preferences and emphasise the advantages of long-term commitment.

3.4.2. Seamless User Experience

Enhance the user experience by ensuring bike availability at popular locations, optimising station placement, and offering a variety of rideable options. By providing a seamless and enjoyable riding experience, Cyclistic can further attract and retain members.

3.4.3. Digital Marketing Optimisation

Invest in digital marketing efforts, utilising social media platforms, online advertisements, and email campaigns to reach and engage casual riders. Implement targeted strategies based on demographic and behavioural data to maximise conversion rates.

3.4.4. Incentivisation Programs

Develop loyalty programs and incentives that reward frequent riders and encourage casual riders to upgrade to annual memberships. Offering discounts, exclusive perks, or bonus ride credits can motivate casual riders to make the transition.

By implementing these recommendations, Cyclistic can effectively drive membership conversions, increase customer loyalty, and position itself as a leading bike-sharing provider. The insights gained from the analysis provide a solid foundation for strategic decision-making and enable the company to optimise its marketing efforts for long-term growth and success.

4. Project Results and Visualisations

4.1. Bike Ride Metrics Dashboard

As part of the project, a comprehensive Bike Ride Metrics Dashboard was created to visualise and analyse the key performance indicators (KPIs) and trends in Cyclistic's bike sharing data. The dashboard provides an intuitive interface for exploring the data and gaining valuable insights into the usage patterns and behaviours of riders.

The Bike Ride Metrics Dashboard can be accessed at the following link: Bike Ride Metrics Dashboard. The dashboard offers charts and graphs that allow users to explore various metrics and trends related to bike rides, membership types, ride lengths, and more.

Users can navigate through the different tabs in the dashboard to access specific visualisations and summaries of the data. By utilising the Bike Ride Metrics Dashboard, stakeholders within Cyclistic can gain a comprehensive understanding of the company's bike sharing operations and make informed decisions to optimise resources, enhance marketing strategies, and improve the overall customer experience.

4.2. Insights from KPIs and Graphs

The analysis of the data and visualisations in the dashboard yielded several important insights:

- Total rides: The total number of rides in the dataset is 5,667,717, indicating a significant level of bike usage within the Cyclistic program.
- Casual riders average ride length: Casual riders have an average ride length of 29 minutes and 9 seconds, indicating that they tend to use the bikes for longer durations.
- Member riders average ride length: Members have an average ride length of 12 minutes and 43 seconds, suggesting that they use the bikes for shorter trips compared to casual riders.

- Percentage of members: Approximately 59% of the riders are members, highlighting the importance of attracting and retaining annual memberships for the success of Cyclistic.
- Rideable type preference: Electric bikes are the preferred rideable type, with 51% of riders choosing them, followed by classic bikes at 46%. This information can help in managing the fleet and ensuring an adequate supply of the preferred rideable types.
- Average ride length by membership type: Casual riders consistently have longer average ride lengths compared to members throughout the year, indicating that they use the bikes for more extended periods.
- Average ride length by day of the week: Casual riders have longer average ride lengths on weekends (Saturdays and Sundays), while members have shorter average ride lengths. Wednesdays tend to have the shortest average ride lengths for both casual riders and members.
- Frequency of rides by membership/rideable type: Members predominantly ride classic bikes, while casual riders favour electric bikes. Docked bikes are used the least by both membership types.
- Frequency of rides by day of the week: Saturdays observe the highest number of riders, while Mondays have the fewest riders. Understanding these patterns can help allocate resources and plan marketing initiatives accordingly.
- Average ride length over time: The months of May, June, and July have the longest average ride lengths, while February, November, and December have the shortest. This information can guide resource allocation and marketing efforts during peak and off-peak seasons.
- Frequency of rides by month: June, July, and August are the months with the highest number of riders, indicating higher demand during the summer season.

- Frequency of rides by membership type: June, July, and August have the highest ridership for both members and casual riders, suggesting a peak season for bike usage.

These insights, derived from the KPIs and visualisations in the Bike Ride Metrics Dashboard, provide valuable information for decision-making, resource allocation, and marketing strategies within Cyclistic. They enable the company to make data-driven decisions that enhance customer experience, improve operational efficiency, and drive membership growth.

5. Validation and Consistency Check

5.1. Validation in MySQL

To ensure the accuracy and consistency of the results obtained in the data analysis, a validation process was conducted using MySQL. The data for each month was imported individually into MySQL, and the same analyses and calculations performed in Excel and R were replicated in the MySQL environment. This allowed for a comparison of the results obtained across different platforms.

The validation process involved cross-checking the key metrics, such as total rides, average ride length, and ridership by membership type and rideable type, between the Excel, R, and MySQL analyses. By comparing the results obtained from each platform, any discrepancies or inconsistencies could be identified and addressed.

5.2. Validation in R

In addition to MySQL, the data analysis was also validated using R. The same datasets for each month were imported into R, and the analyses and calculations performed in Excel and MySQL were replicated using R's data manipulation and visualisation capabilities. By comparing the results obtained in R with those from Excel and MySQL, any inconsistencies or discrepancies could be detected. This validation process ensured that the findings and insights derived from the data analysis were consistent across multiple analytical platforms.

The validation and consistency checks in both MySQL and R provided assurance that the project results were reliable and accurate. Any inconsistencies or discrepancies discovered during the validation process were carefully investigated and resolved to ensure the integrity of the findings presented in the project report.

6. Conclusion

Through the comprehensive analysis of Cyclistic's bike-sharing data and the exploration of key metrics and trends, this project has provided valuable insights into the behaviour and preferences of annual members and casual riders. The findings highlight important differences between these two rider segments, shedding light on opportunities to increase annual memberships and optimise marketing strategies.

The analysis revealed that annual members are more profitable for Cyclistic compared to casual riders. Understanding the factors that influence casual riders to purchase annual memberships and leveraging digital media channels can significantly enhance conversion rates. By tailoring marketing efforts and offering targeted incentives, Cyclistic can effectively attract and convert casual riders into loyal annual members.

The Bike Ride Metrics Dashboard, developed using Excel, provides a user-friendly interface to visualise and analyse the key performance indicators and trends in bike usage. The dashboard offers valuable insights into average ride lengths, ridership by membership type and rideable type, and the frequency of rides by day of the week and month.

The validation and consistency checks performed in MySQL and R confirm the reliability and accuracy of the project results. The findings were consistent across different analytical platforms, ensuring the integrity of the insights and recommendations.

Overall, this project has contributed to a deeper understanding of Cyclistic's bike-sharing data and provided actionable recommendations to maximise annual memberships, optimise marketing strategies, and enhance the customer experience. By implementing these recommendations, Cyclistic can position itself for long-term growth, profitability, and success in the competitive bike-sharing market.

7. Limitations and Future Scope

Despite the valuable insights and recommendations obtained from this project, it is important to acknowledge the limitations and consider future avenues for improvement. The following limitations should be taken into account:

1. Data Limitations: The analysis was conducted based on the historical trip data provided by Cyclistic for the year 2022. While this dataset offers a comprehensive view of bike usage patterns, it is limited to a single year. Future analyses could benefit from additional years of data to capture long-term trends and seasonality.
2. External Factors: The analysis focused primarily on internal factors such as rider behaviour and preferences. External factors such as weather conditions, local events, and competitor activities were not explicitly considered. Future research could incorporate external data sources to gain a more holistic understanding of the bike-sharing market dynamics.
3. Data Granularity: The available data provided information at a monthly level, limiting the ability to analyse more granular time periods, such as daily or hourly patterns. Fine-grained insights into peak usage times and specific user behaviours could further enhance marketing strategies and operational decisions.
4. Sample Bias: The analysis was based on the data provided by Cyclistic and may not represent the entire population of bike-share users. The findings and recommendations should be interpreted with the understanding that they are derived from a specific sample and may not be fully generalisable.

The future scope for this project includes the following:

1. Integration of Real-time Data: Incorporating real-time data streams can provide up-to-date insights and enable proactive decision-making. Monitoring live data on bike

availability, station occupancy, and user feedback could enhance operational efficiency and customer satisfaction.

2. Advanced Predictive Modelling: By applying machine learning and predictive analytics techniques, it is possible to forecast demand patterns, predict churn rates, and personalise marketing efforts. These advanced modelling approaches can provide actionable insights for targeted interventions and improved customer retention.
3. A/B Testing and Experimentation: Conducting controlled experiments and A/B testing of different marketing strategies can help validate the effectiveness of specific interventions. This iterative approach allows for data-driven decision-making and continuous optimisation of marketing campaigns.
4. Integration of User Surveys and Feedback: Collecting user feedback through surveys or app-based feedback mechanisms can provide qualitative insights into user preferences, satisfaction, and suggestions for improvement. Incorporating such feedback into the analysis can enhance the understanding of user motivations and inform strategic decisions.

Addressing these limitations and exploring the future scope will contribute to a more comprehensive understanding of Cyclistic's bike-sharing operations and enable data-driven decision-making for sustained growth and customer satisfaction.

8. References

1. R Core Team. (2021). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.
2. Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L. D., François, R., ... & Henry, L. (2019). Welcome to the tidyverse. Journal of Open Source Software, 4(43), 1686.
3. Wickham, H. (2019). tidyverse: Easily Install and Load the 'Tidyverse'. R package version 1.3.0.
4. Therneau, T. (2021). A Package for Time-to-Event Analysis in R. R package version 2.3-2.