How does a bike-share navigate speedy success?

A Capstone Project for Google Data Analytics Certification

May 30, 2025

Presented By:

Quiambao, Eric Janssen P.

Table of Contents

Executive Summary

I analyzed Cyclistic's bike-share data to examine the behavioral differences between annual members and casual riders, with the goal of informing targeted digital marketing strategies to convert casual riders into loyal annual members. For a comprehensive view, I employed a full 12-month dataset for exploratory analysis using Google Sheets, Tableau, and SQL.

Throughout the analysis, I identified significant differences in ride length, peak usage times, and day-of-week patterns between the two customer groups (Casual riders and annual members). These observations imply that casual riders, who typically have shorter and more intermittent trips, might be best engaged with targeted marketing initiatives and reward promotions.

With these findings in mind, my top recommendations were to:

- Create targeted digital marketing campaigns that promote the value of an annual membership.
- Provide incentive trials or discounts to drive conversion.
- Increase customer interaction with usage-based customized messaging.

This project not only yielded actionable recommendations for Cyclistic's marketing plan but also demonstrated my technical skills across multiple data analytics tools, making it a strong addition to my professional portfolio.

Introduction

This capstone project analyzes the bike-share system of a fictional company known as Cyclistic, to reveal the behavioral distinctions between annual members and occasional riders. With the goal of determining actionable insights that can inform targeted marketing efforts, the analysis centers on converting occasional users into loyal, profitable annual members. For this analysis, I utilized a comprehensive 12-month dataset to capture seasonal trends and usage patterns using Google Sheets, Tableau, and SQL.

The project follows a structured six-step data analytics process—Ask, Prepare, Process, Analyze, Share, and Act—which ensures that every phase is methodically executed and thoroughly documented. By employing a diverse set of tools—Google Sheets for initial data manipulation, Tableau for interactive visualizations, and SQL for efficient data querying—I not only addressed critical business questions but also demonstrated versatility across multiple data analytics platforms.

This report details all stages of the process, from data collection and cleaning to in-depth analysis and final recommendations. The insights generated are intended to inform Cyclistic's strategic planning and maximize customer conversion through data-driven digital media campaigns, while also serving as a robust portfolio piece showcasing my technical proficiency.

Scenario

For this project, I played the position of a junior data analyst for the marketing team at Cyclistic, a leading bike-share operator in Chicago. Cyclistic had expanded to have more than 5,800 bikes at more than 600 docking stations with not just the standard, but also specialized bikes like reclining bikes, hand tricycles, and cargo bikes to accommodate different needs. While occasional riders propelled high usage through single-trip and all-day passes, the financial review indicated that annual members produced much greater profitability.

Working under Marketing Director Lily Moreno's guidance, I was challenged to examine a complete year of past trip data to discover how occasional riders and annual members utilized Cyclistic bikes differently. The findings that I drew from this analysis drove a targeted digital media campaign that aimed to convert occasional riders into loyal annual members. Using tools such as Google Sheets, Tableau, and SQL, I built a holistic, multi-tool analysis that not only addressed key business questions but also demonstrated my technical adaptability.

This situation paved the way for a data-driven transformation strategy to further Cyclistic's long-term development and cement its competitive advantage in the bike-share industry.

Methodology

I employed a multi-tool approach to ensure a comprehensive analysis of Cyclistic's bike-share data. My methodology was structured around the data analysis framework—Ask, Prepare, Process, Analyze, Share, and Act—taught to me in my Google Data Analytics Certification, which guided each phase of the project.This structured approach ensured that each stage of the project was methodically executed and thoroughly documented.

Process Overview

I began by clearly defining the business objectives and questions in the Ask phase. Next, I prepared and cleaned the data, ensuring its accuracy and consistency. I then processed the dataset to generate new variables (such as ride duration and day-of-week) and conducted exploratory analysis. Following that, I performed in-depth analysis, created visualizations, and finally, shared my findings and actionable recommendations.

Tools & Technologies

To achieve a thorough analysis, I utilized the following tools:

- **Google Sheets**

  I used Google Sheets for initial data cleaning and manipulation on a full 12-month dataset. This allowed me to quickly compute key metrics like ride duration and day-of-week, perform preliminary exploratory analysis with pivot tables, and generate basic charts that captured seasonal trends.

- **Tableau**

    I leveraged Tableau to create interactive dashboards that visually represented trends and usage patterns across the full year. Its dynamic visualization capabilities enabled a user-friendly exploration of data trends and supported the identification of actionable insights.

- **SQL**

    I utilized SQL with Google BigQuery for efficient data querying and aggregation across the full 12-month dataset. This ensured that the insights from the other tools were validated and that data was consistently and reliably aggregated.

This multi-tool methodology provided multiple perspectives on the data while showcasing my proficiency across diverse platforms. By validating findings through various analytical approaches, I tailored the analysis to comprehensively address the different facets of the project.

Ask

In this phase, I defined the business objectives and formulated the key questions that would guide my future analysis. My primary goal was to **understand how annual members and casual riders used Cyclistic bikes differently**, with the intention of helping in developing targeted marketing strategies to convert casual riders into loyal annual members. My approach is informed by the following guiding questions:

- What is the problem I am trying to solve?
- How can the insights from this analysis drive effective business decisions?

To achieve this, I focused on several critical questions to solve this goal:

1. How did the usage patterns of casual riders differ from those of annual members?
2. What factors might drive a casual rider to upgrade to an annual membership?
3. And which digital marketing strategies could effectively encourage this conversion?

For this case study, I concentrated on answering the first question. By examining and understanding the distinct usage behaviors between these two groups, I aimed to generate actionable insights that would support Cyclistic's strategic objectives and inform their marketing decisions.

Problems

As I embarked on planning the analysis, I understood several key problems that needed addressing in order to successfully answer our questions:

- **Differing Usage Patterns**

  I needed to determine the specific ways in which annual members and casual riders interact with the Cyclistic system. This included differences in ride duration, frequency, and time-of-day usage.

- **Conversion Challenges**

  Understanding why casual riders might hesitate to commit to an annual membership was essential. I aimed to uncover any possible barriers or motivational factors that could inform conversion strategies.

- **Data-Driven Decision-Making**

  To support effective marketing strategies, it was crucial to rely on solid data. This meant identifying any issues in data collection, ensuring its accuracy, and addressing any inconsistencies or gaps that could impact the analysis.

This foundational work set the stage for the rest of the project. By clearly defining the business task and outlining these challenges, I ensured that my analysis would yield actionable insights. Additionally, the project deliverables are structured to include:

- A clear statement of the business task.
- A description of all data sources used.
- Documentation of any cleaning or manipulation of data.
- A summary of the analysis.
- Supporting visualizations and key findings.
- The top three recommendations based on the analysis.

Prepare

For this phase, I began by gathering Cyclistic's historical trip data from multiple sources. I downloaded a full 12-month dataset to be used with tools such as Google Sheets, Tableau, and SQL. All files were stored in a dedicated project folder titled "Cyclistic_Case_Study," where I organized the data into clearly labeled subfolders for Raw Data, Processed Data, and Documentation. This structured approach ensured that each dataset could be easily accessed and differentiated, avoiding any confusion between the full-year perspective and the early-year subset used for advanced statistical analysis.

Once the datasets were downloaded, I verified that the data was organized consistently, with all CSV files containing the expected columns: trip_id, start_time, stop_time, bikeid, tripduration, from_station_id, from_station_name, to_station_id, to_station_name, user_type, gender, and birthday. I ensured that the files were formatted correctly, with date and time fields properly recognized so that subsequent calculations, such as ride duration and day-of-week extractions, could be performed without error. Sorting and filtering were applied as necessary to remove any incomplete or erroneous entries, and to make sure that the data was clean and standardized across the different sources.

To address concerns regarding credibility and data integrity, I reviewed the accompanying data dictionaries and metadata provided with the datasets. These documents clarified the definitions and formats of each variable, and confirmed that the data was made available by Motivate International Inc. under the appropriate license. This validation process helped ensure that the data was not only reliable and accurate but also free from issues of bias. Additionally, I made sure to comply with data privacy guidelines by only using aggregated and non-personally identifiable information, as the dataset does not allow the connection of pass purchases to credit card details.

By meticulously organizing and verifying the data, I created a robust foundation for the subsequent phases of the analysis. The prepared datasets are now ready for further cleaning, processing, and analysis, which will ultimately support actionable insights into how different customer types use Cyclistic bikes

Process

I began by importing Cyclistic's historical trip data into Google Sheets and organized my workbook into multiple sheets to manage the dataset efficiently. I renamed the original sheet as "Raw Data" and standardized the column naming conventions (using a consistent word_word format) to maintain clarity. To accommodate the spreadsheet's cell limit, I removed extraneous columns such as trip_id, bike_id, from_station_id, and to_station_id.

Next, I ensured that the date and time fields—start_time and end_time—were correctly formatted by applying the Date Time format, rather than relying on the default automatic format. With these fields properly formatted, I added a new column called ride_length, which calculates the duration of each trip by subtracting start_time from end_time. I observed inconsistencies between the provided trip_duration and the computed ride_length, so I opted to remove the trip_duration column and rely solely on the accurately calculated ride_length.

To further enhance my analysis, I inserted a day_of_week column to categorize each trip based on the day it occurred. I also created a start_hour column to help determine peak usage times. Using conditional formatting, I checked for anomalies—such as negative values in ride_length—and confirmed that there were none.

Once the data was fully cleaned and transformed, I renamed the "Raw Data" sheet to "Cleaned Data" to reflect its finalized state. Based on this Cleaned Data, I then created a new sheet titled "Pivot Tables & Charts." In this sheet, I constructed several pivot tables to explore key trends and insights:

- **Average Ride Length by User Type -** This pivot table compares the average duration of rides between casual riders and annual members.
- **Ride Count by Day of the Week and User Type** - This table displays the number of rides for each day, segmented by user type, to highlight daily usage trends.
- **Average Ride Length by Day of the Week and User Type -** This pivot table illustrates how ride durations vary across the week for each user group.
- **Ride Count by Start Hour and User Type** - This table shows the distribution of rides by the hour they started, indicating peak usage times for both user types.

I added comments to each pivot table to record my insights and created charts based on these tables to visually represent the key findings. This structured, multi-sheet approach in Google Sheets allowed me to efficiently clean, process, and analyze the data while ensuring that all insights were clearly documented and reproducible for the Cyclistic case study

Analyze

In this phase, I examined Cyclistic's bike-share data to uncover key behavioral differences between annual members (Subscribers) and casual riders (Customers). By leveraging pivot tables and interactive dashboards in Google Sheets and Tableau Public, I identified that Customers tend to take longer rides predominantly on weekends, while Subscribers exhibit shorter, more frequent rides during weekdays. These insights provide a clear foundation for developing targeted marketing strategies to convert casual riders into loyal, annual members.

Sheets Analyzation

Based on the pivot table outputs created in Google Sheets, several key insights have emerged that directly address our primary question: How do Subscribers and Customers use Cyclistic bikes differently? The analysis reveals that Customers average ride durations between 27 to 33 minutes and are predominantly active on weekends, suggesting a leisure or recreational use. In contrast, Subscribers average shorter ride durations (around 11 to 13 minutes) and show higher ride counts during weekdays, particularly from Monday to Friday, indicating a commuter usage pattern.

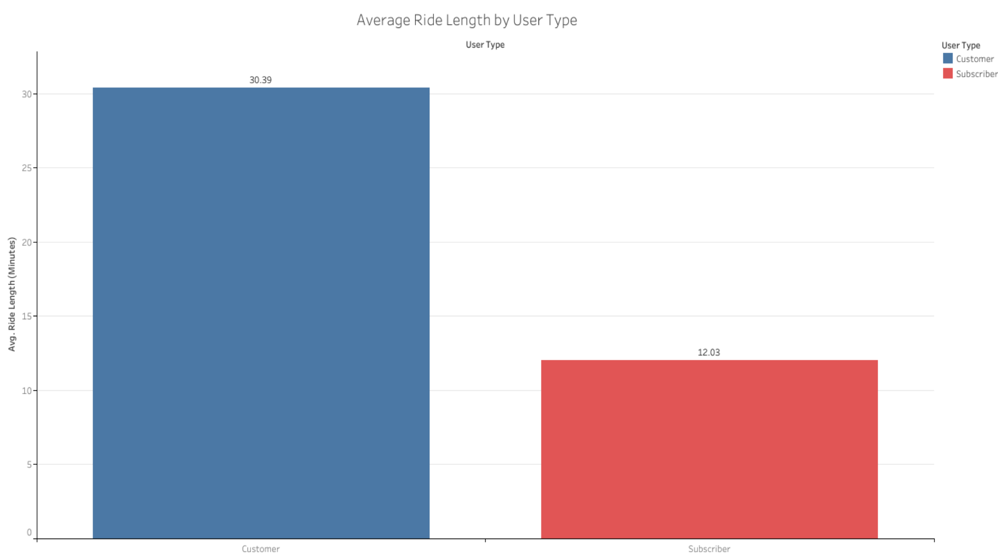
The time-of-day analysis further reinforces these distinctions. Subscribers primarily ride during the early and late working hours—consistent with typical commuting schedules—while Customers display a more uniform riding pattern during working hours with a notable surge on weekends. These findings not only highlight significant behavioral differences but also present a clear opportunity to tailor digital marketing strategies. For instance, campaigns could emphasize the convenience and cost-efficiency of annual memberships for weekday commuters, while weekend promotions could be designed to attract casual riders.

Tableau Documentation

After the initial analysis in Sheets, I imported the “Clean Data” into Tableau Public and verified that all fields were imported correctly. I observed that the ride_length field was interpreted as a datetime value, so I created a calculated field to convert ride_length into minutes for accurate analysis.

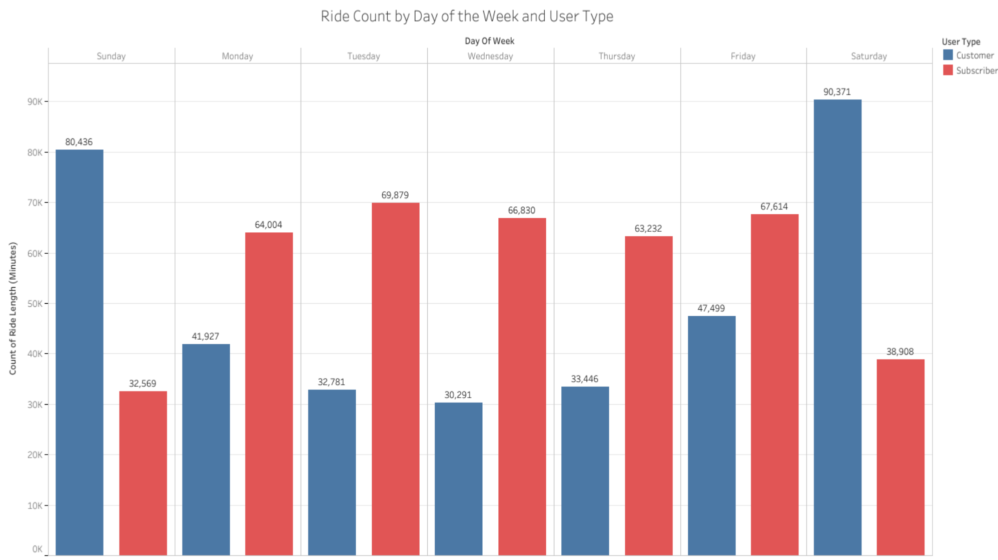
I then recreated the key charts that I had developed in Google Sheets:

- Average Ride Length by User Type



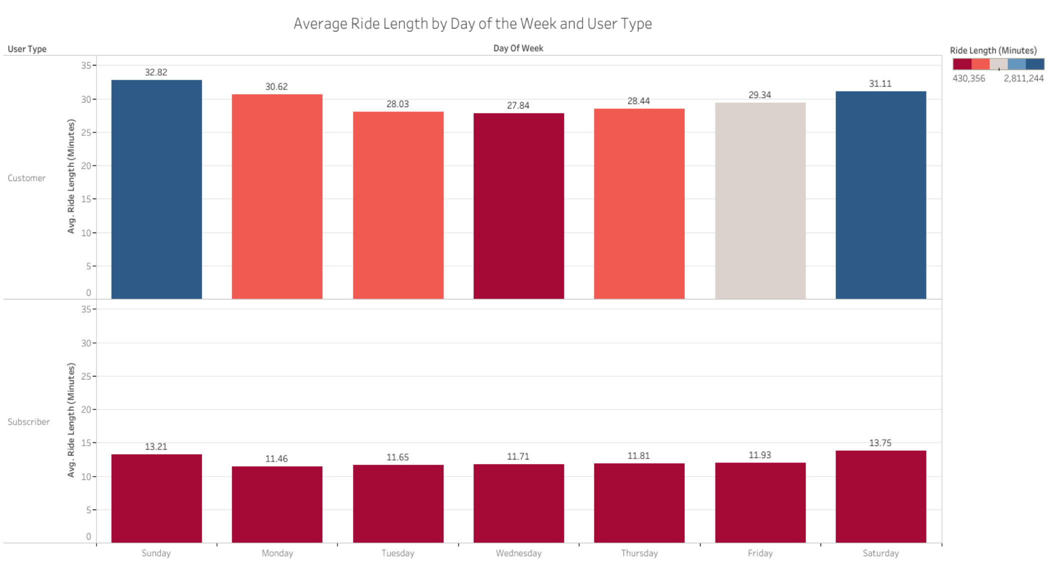
I set up a bar chart with user_type on the Columns shelf and the average ride_length on the Rows shelf. I applied user_type as the Color field and enabled labels to always show, clearly demonstrating that Customers average a longer ride duration than Subscribers.

- Ride Count by Day of the Week and User Type



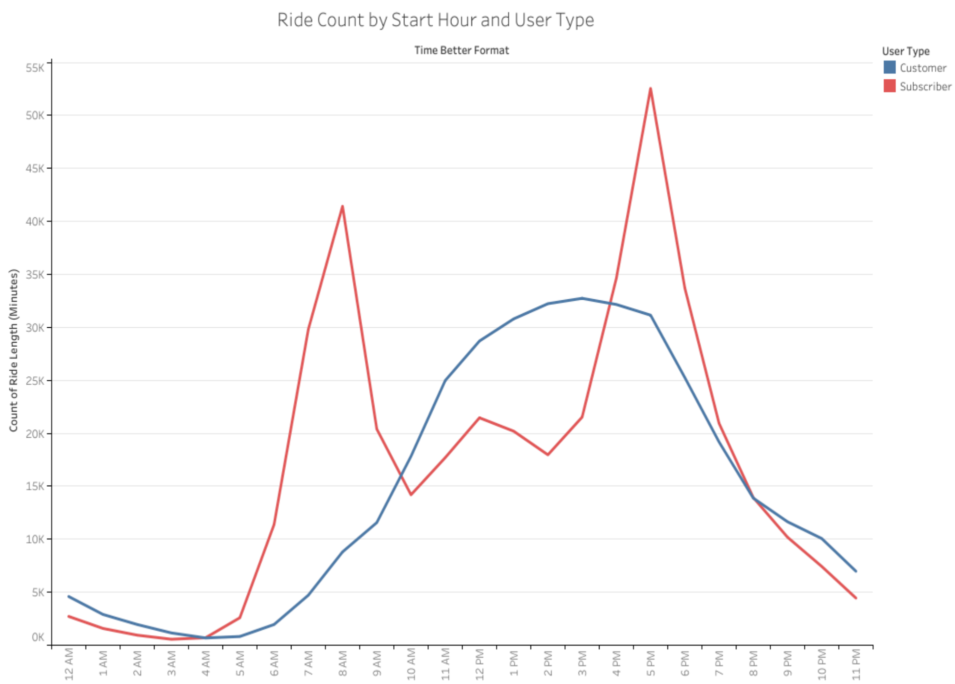
Using a bar chart, I placed day_of_week on the Columns shelf and the count of ride_length on the Rows shelf, with user_type as the Color field. Labels were enabled to provide clear insights on daily ride volumes.

- Average Ride Length by Day of the Week and User Type

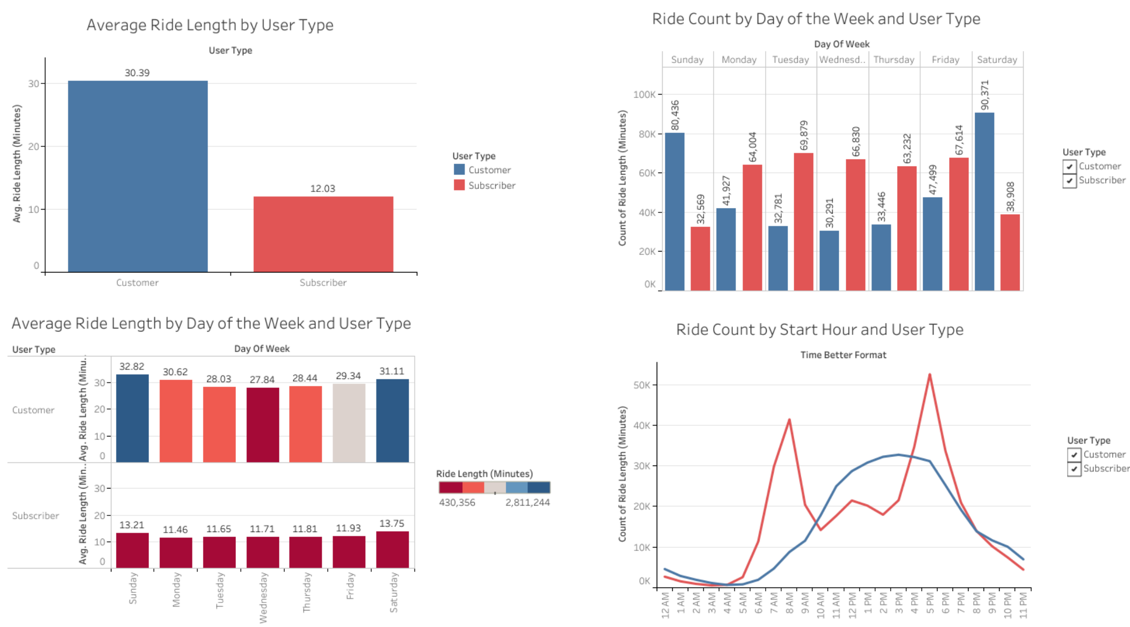


I created a bar chart with a heat map style by setting day_of_week on the Columns shelf, user_type on the Rows shelf, and visualizing the average ride_length with color intensity. Labels were added to facilitate quick interpretation of the data.

- Ride Count by Start Hour and User Type



I developed a line chart with start_hour on the Columns shelf and the count of ride_length on the Rows shelf, applying user_type as the Color field. This visualization revealed that Subscribers ride predominantly during the early and late working hours, whereas Customers peak during working hours.



To enhance the dashboard, I added a main title and integrated interactive filters—specifically, filters for start_hour and user_type on the Ride Count by Start Hour chart, and a user_type filter on the Ride Count by Day of the Week chart. This interactivity allows users to drill down into specific time periods and segments for more targeted insights.

BigQuery (SQL) Analysis

To further validate and expand on these insights, I conducted an SQL analysis in Google BigQuery using the table cyclisticcasestudy-454609.Cyclistic_Case_Study.Cleaned_Data. Key queries and their results include:

Overall Ride Duration Summary

- Minimum ride duration: 60 seconds (1.00 minute)
- Maximum ride duration: 86340 seconds (1439.00 minutes)
- Average ride duration: 1239.19 seconds (20.65 minutes)

Ride Duration Summary by User Type

Customers

- Ride count: 356,751
- Minimum: 60 seconds (1.0 minute)
- Maximum: 86340 seconds (1439.0 minutes)
- Average: approximately 1823.58 seconds (30.39 minutes)

Subscribers

- Ride count: 403,036
- Minimum: 60 seconds (1.0 minute)
- Maximum: 85860 seconds (1431.0 minutes)
- Average: approximately 721.91 seconds (12.03 minutes)

Ride Duration Summary by Day of the Week and User Type

The query results revealed detailed trends across days. For example, on Fridays, Customers averaged approximately 29.34 minutes per ride (with 47,499 rides), while Subscribers averaged about 11.93 minutes (with 67,614 rides). Similar patterns were observed across other days, confirming that Customers generally take longer rides, especially on weekends, whereas Subscribers maintain shorter ride durations on weekdays.

Ride Count by Start Hour and User Type

The query output for start_hour shows the distribution of rides throughout the day. For instance, at hour 0, Customers had 4,591 rides while Subscribers had 2,714; as the day progresses, distinct peaks and troughs are observed that align with the previously identified commuting and leisure patterns.

Analysis Summary

The comprehensive multi-tool analysis confirms significant behavioral differences between Subscribers and Customers. Customers engage in longer, weekend-centric rides—averaging about 30 minutes per ride—while Subscribers take shorter rides—averaging around 12 minutes—and are more active on weekdays, especially during typical commuting hours. These insights, validated across Google Sheets, Tableau Public, and BigQuery SQL, form a robust data-driven foundation for developing targeted digital marketing strategies aimed at converting casual riders into loyal annual members. Moving forward, these findings will directly inform actionable recommendations and strategic actions in the next phase of the project.

Share

For the Share phase, I disseminated my project results using two complementary approaches:

1. **GitHub Repository with Detailed Documentation**

    All code, SQL queries, pivot tables, and visualizations have been published in a well-organized GitHub repository. This repository includes a comprehensive README that provides an overview of the project, details the methodology, and summarizes the key insights.

    It also contains images and descriptions (in documentation/here) of the dashboards and visualizations, ensuring that every step of the process is transparent and easily reproducible for anyone reviewing the project.

2. **Interactive Tableau Public Dashboard**

    I published an interactive dashboard on Tableau Public that allows stakeholders to explore the data and insights dynamically. The dashboard is accessible via a direct link provided in the repository and project documentation.

    It offers an engaging, user-friendly experience, enabling viewers to drill down into key metrics such as ride duration, daily and hourly trends, and user type differences. Accompanying the dashboard is a brief explanation of the project context and the significance of the visualized insights.

Together, these methods ensure that the project outcomes are accessible to both technical and non-technical audiences, facilitating collaboration, feedback, and further analysis.

Act

In this phase, I translate the insights from my analysis into actionable steps. Building on the data-driven findings, I have formulated solutions, drawn final conclusions, and developed targeted recommendations that address the specific behavioral differences between Subscribers and Customers. These actions are designed to guide Cyclistic's marketing strategy and drive customer conversion, paving the way for a more effective and engaging service.

Solutions

Based on the analysis, several solutions emerge to leverage the behavioral differences identified between Subscribers and Customers. The primary solution is to develop data-informed digital marketing campaigns tailored to each group. For example, targeted promotions can be designed for weekday commuters (Subscribers), highlighting the cost-effectiveness and convenience of annual memberships, while incentive trials or discounts could be offered on weekends to encourage casual riders (Customers) to convert to membership. Additionally, implementing personalized, usage-based messaging could help further engage both groups by addressing their unique ride patterns.

Conclusion

The comprehensive analysis confirms significant differences in usage behavior: Customers tend to engage in longer rides, predominantly on weekends, while Subscribers take shorter, more frequent rides during weekdays, particularly during commuting hours. These findings indicate that casual riders are more likely to use the service for leisure, whereas Subscribers rely on it for daily commuting. The insights derived from Google Sheets, Tableau, and BigQuery (SQL) create a robust data-driven foundation that not only addresses the primary business question but also highlights clear opportunities for optimizing Cyclistic's marketing strategy.

Recommendations

1. Develop Targeted Digital Campaigns

Focus on promoting the benefits of annual memberships for weekday commuters by emphasizing convenience, cost savings, and efficiency. Tailor messages to highlight how annual memberships support daily travel needs.

2. Introduce Incentive Trials or Discounts

Create promotions specifically designed to convert casual riders by offering limited-time discounts or trial memberships. This approach can lower the barrier for conversion and encourage loyal membership adoption.

3. Enhance Personalized Customer Engagement

Utilize data-driven insights to implement personalized messaging strategies. For example, use peak ride times and user behavior patterns to deliver targeted communications that resonate with both Subscribers and Customers, ultimately boosting engagement and conversion rates.

Appendices/Documentation

Step-by-Step Data Cleaning and Transformation Process

Google Sheets

- Imported the raw data and renamed the original sheet as “Raw Data”
- Standardized column names and removed unnecessary columns
- Applied Date Time formatting to start_time and stop_time
- Calculated ride_length by subtracting start_time from stop_time
- Created additional columns (day_of_week and start_hour) to categorize data by date and time
- Applied conditional formatting to check for anomalies (e.g., negative ride lengths)
- Renamed the sheet to “Cleaned Data” once processing was complete
- Created a “Pivot Tables & Charts” sheet to summarize and visualize key trends

Tableau Public

- Imported “Cleaned Data” and corrected ride_length from datetime to minutes using a calculated field
- Recreated pivot table visualizations (e.g., Average Ride Length by User Type, Ride Count by Day of the Week, etc.) and added interactive filters

BigQuery (SQL)

- Loaded the cleaned dataset and executed SQL queries to compute summary statistics
- Converted ride_length (TIME) to total seconds using EXTRACT functions to calculate min, max, and average ride durations, both in seconds and minutes

- Aggregated data by user_type, day_of_week, and start_hour to validate and extend findings

Assumptions, Limitations, and Challenges

Assumptions

- The ride_length calculation (stop_time minus start_time) accurately represents the trip duration
- The derived fields (day_of_week, start_hour) effectively capture usage patterns

Limitations

- The dataset spans a full 12 months for Sheets, Tableau, and SQL, while a Q1 subset was planned for R (to be integrated in future work)
- Data privacy guidelines restrict the inclusion of personally identifiable information

Challenges

- Managing large datasets within Google Sheets' cell limits necessitated removal of extraneous columns
- Converting TIME fields to numerical values in BigQuery required careful use of EXTRACT functions
- Ensuring consistent formatting and data integrity across multiple tools was critical and required thorough documentation

References

[1] “Original Dataset,” *Docs*. https://divvy-tripdata.s3.amazonaws.com/index.html (accessed Mar. 22, 2025).