**Case Study: How does a bike-share navigate speedy success?**



**Summary**

As a junior data analyst in the marketing team at Cyclistic, a bike-share company in Chicago, I am tasked with analyzing data to uncover insights. Cyclistic stands out among bike-share programs with its diverse offerings, including reclining bikes, hand tricycles, and cargo bikes, ensuring accessibility for individuals with disabilities and those seeking alternative options. While the majority of riders prefer traditional bikes, around 8% choose the assistive options. Riders predominantly use Cyclistic for leisure purposes, although approximately 30% utilize the bikes for daily commuting. Customers who purchase single-ride or full-day passes are considered casual riders, while those with annual memberships are recognized as Cyclistic members.

The director of marketing believes that the company's future success hinges on increasing the number of annual memberships. To achieve this goal, my team aims to understand the distinct usage patterns between casual riders and annual members. Armed with these insights, we will develop a targeted marketing strategy to convert casual riders into loyal annual members. However, the final recommendations must gain approval from Cyclistic executives, necessitating compelling data insights and professional data visualizations to support our findings.

**Purpose**

The analysis aims to uncover the differences in how annual members and casual riders utilize Cyclistic bikes. By studying their behavior and usage patterns, we can gain valuable insights into the distinct characteristics of these customer segments. These insights will serve as the foundation for developing a targeted marketing campaign, specifically designed to convert casual riders into loyal annual members. The study will leverage historical bike trip data from the past year to identify relevant trends and patterns.

**Cyclistic Historical Trip Data**

The dataset used for analysis consists of Cyclistic's historical [trip data](https://divvy-tripdata.s3.amazonaws.com/index.html) from the past 12 months. The data is in a clean and usable format, stored as a .csv file. It includes accurate dates and times without any outliers, making it highly relevant to the problem at hand. The dataset has been generously provided by Motivate International Inc. under [a](https://ride.divvybikes.com/data-license-agreement) license. As public data, it offers an opportunity to explore and understand how various customer types are utilizing Cyclistic bikes.

**About Cyclistic**

Cyclistic is a fictional bike-share program that features 5,824 bicycles and 692 docking stations. Cyclistic sets itself apart by also offering reclining bikes, hand tricycles, and cargo bikes, making bike-share more inclusive to people with disabilities and riders who can’t use a standard two-wheeled bike. The majority of riders opt for traditional bikes; about 8% of riders use the assistive options. Cyclistic users are more likely to ride for leisure, but about 30% use them to commute to work each day.

**Key Individuals and Teams:**

**Lily Moreno**: The director of marketing and your manager at Cyclistic. Moreno is responsible for developing campaigns and initiatives to promote the bike-share program. This may involve utilizing email, social media, and other channels to reach customers and increase awareness.

**Cyclistic marketing analytics team**: A team of data analysts dedicated to collecting, analyzing, and reporting data that informs Cyclistic's marketing strategy. As a junior data analyst, you are part of this team, working towards understanding customer behavior and usage patterns to drive effective marketing campaigns.

**Cyclistic executive team**: The executive team at Cyclistic oversees the overall operations and strategic decision-making. They will ultimately decide whether to approve the recommended marketing program. Their focus on detail and analysis ensures that decisions align with Cyclistic's goals and objectives.

**Data Analysis Methodology**

The analysis followed a structured approach consisting of the six primary steps of data analysis:

* Ask: Defining the key questions and objectives of the analysis.
* Prepare: Gathering and organizing the required data for analysis.
* Process: Cleaning, transforming, and refining the data to ensure its quality and relevance.
* Analyze: Applying various analytical techniques and methods to extract insights from the data.
* Share: Communicating the findings and insights derived from the analysis effectively.
* Act: Taking informed actions based on the analysis results to drive decision-making and achieve desired outcomes.

1. **Ask**

* Business Objective: Develop effective marketing strategies to convert casual riders into annual members by gaining insights into the distinct usage patterns of annual members and casual riders of Cyclistic bikes.
* Director of Marketing - Lily Moreno: Lily Moreno reports to the Cyclistic executive team and is responsible for developing marketing campaigns. She has assigned the marketing analytics team the task of understanding the differences in Cyclistic bike usage between casual riders and annual members. Lily firmly believes that the company's future success relies on maximizing the number of annual memberships. Her specific goal is to design marketing strategies that effectively convert casual riders into annual members.
* Marketing Analytics Team: This team comprises data analysts who are responsible for collecting, analyzing, and reporting data to guide marketing strategies. They report directly to Lily Moreno.
* Junior Analyst (Report Writer): As a new member of the marketing analytics team, my role is to answer the question: How do annual members and casual riders use Cyclistic bikes differently? I will utilize the most recent 12 months of Cyclistic bike share data to provide detailed analysis and present the top three recommendations to Lily Moreno in a comprehensive report.
* Business Task: Using the latest 12 months of Cyclistic bike share data, the objective is to examine and identify the distinct ways in which annual members and casual riders utilize Cyclistic bikes. Lily Moreno seeks the top three recommendations based on the analysis and expects a detailed report outlining the findings.

1. **Prepare**

* Data Source: The data used for analysis is Cyclistic's historical [trip data](https://divvy-tripdata.s3.amazonaws.com/index.html) from the last 12 months. The data is in a usable .csv format and is organized into monthly files. For this analysis, we will utilize the data from June 1, 2022 to May 31, 2023.
* Data Quality (ROCCC): Reliability, Originality, and Credibility: The data is sourced directly from Divvy Bikes bike share for this fictitious case. It is the original data collected by Divvy and properly cited, ensuring its reliability and credibility.
* Comprehensive: Ample data is available to conduct the study, ensuring its comprehensiveness.
* Bias and Missing Data: There are cases of missing data in the dataset, and it would be important to investigate the reasons behind these null values. In a real-world scenario, we would communicate with the team responsible for data collection and bike station maintenance to confirm the reliability and address any potential bias in data collection. For the purposes of this case, we have removed the null values.
* Licensing, Privacy, Security, and Accessibility: The data is provided under license from Motivate International Inc. There is no personally identifiable information in the dataset, ensuring privacy and security.
* Data integrity was maintained throughout the process. The data was downloaded and stored locally, and individual CSV files were directly imported into R Studio from their original folder to minimize data replication and transfer. Data types, total number of rows, and other checks were performed at various stages of preparation and processing to ensure data integrity and avoid unintended data loss.

1. **Process**

* Tools Selection: Considering the combined size of all 12 datasets, performing data cleaning in a spreadsheet would be time-consuming and slow. Therefore, I have chosen R as the tool of choice. Using R allows me to efficiently handle both data wrangling and analysis, including visualizations, within the same platform.
* Data Transformation: The process of transforming the data includes tasks such as data migration, data warehousing, data integration, and data wrangling. These steps are crucial to prepare the data for further analysis and insights.
* Data Error Checking: To ensure data quality, it is essential to check the data for any errors or inconsistencies. This involves examining the data for missing values, outliers, or any other issues that may affect the accuracy and reliability of the analysis.
* Data Cleaning: Cleaning the data involves various techniques and procedures to address errors, inconsistencies, and missing values identified during the error checking stage. By applying data cleaning techniques, we aim to improve the overall quality and integrity of the dataset.
* Documentation of the Cleaning Process: It is important to document the steps taken during the data cleaning process. This documentation helps ensure transparency and reproducibility of the analysis. It allows others to understand the actions performed on the data and provides a reference for future analysis or audits.

1. **Analyze**

* Data Aggregation and Analysis: After preparing and cleaning the data, the next step is to aggregate the data and perform calculations to identify trends and relationships. This involves summarizing the data at different levels of granularity and deriving meaningful insights.
* Aggregating the Data: Data aggregation involves grouping the data based on specific variables or categories. This allows us to summarize and analyze the data at a higher level, providing a broader perspective on the patterns and trends within the dataset. Aggregation can be performed using various statistical functions such as sum, average, count, etc.
* Calculating Trends and Relationships: Once the data is aggregated, we can apply statistical and analytical techniques to identify trends and relationships within the dataset. This may involve calculating metrics such as growth rates, proportions, correlations, or performing more advanced analyses such as regression or clustering.
* Exploring Data Visualization: Data visualization plays a crucial role in understanding trends and relationships in the data. By visualizing the aggregated data using charts, graphs, or other visual representations, we can gain insights that might not be apparent from raw numbers alone. Data visualization enhances the interpretation and communication of findings to stakeholders.
* Interpreting Results: Analyzing the aggregated data and calculated metrics helps us gain a deeper understanding of the dataset. By interpreting the results, we can identify significant trends, patterns, and relationships that can inform decision-making and drive actionable insights.
* Communicating Findings: The insights and findings derived from the data analysis need to be effectively communicated to stakeholders. This involves creating clear and concise reports, visualizations, or presentations that highlight the identified trends and relationships. Presenting the information in an understandable and visually appealing manner helps stakeholders make informed decisions based on the analysis.

By performing data aggregation, calculations, and analysis, we can uncover valuable insights and patterns within the dataset, enabling us to make data-driven decisions and recommendations.

1. **Share**

Effective Sharing of Data Analysis Insights through R Visualization

The "Share" phase of the data analysis methodology is a critical step where we communicate the findings and insights derived from the analysis to stakeholders and decision-makers. In this case, we utilized R, a powerful programming language for statistical analysis and data visualization, to create impactful visualizations that effectively communicate the results of our analysis.

Key Findings and Insights:

1. Usage Patterns and Trends:

* Through our data analysis using R, we uncovered valuable insights into the usage patterns and trends of the Cyclistic bike riders. Our visualizations highlighted the disparities between Member and Casual users, peak riding times, popular stations, and seasonal trends. These insights provide a comprehensive understanding of the bike usage dynamics, allowing for targeted marketing strategies and operational optimizations.

1. User Behavior and Preferences:

* Using R visualization, we examined user behavior and preferences. We explored factors such as the average ride duration, preferred bike types (classic, electric, or docked), and weekday versus weekend usage patterns. These insights shed light on user preferences and can guide decisions related to bike fleet management, promotions, and user experience enhancements.

1. Conversion Opportunities:

* Our analysis identified potential opportunities to convert Casual riders into annual members. By analyzing the usage patterns, we can develop targeted marketing strategies and incentives tailored to the specific needs and preferences of Casual users. This can help increase membership sign-ups and drive revenue growth.

Sharing Strategies and Recommendations:

To effectively share the data analysis insights and maximize their impact, we recommend the following strategies:

1. Compelling Visualizations:

* Utilize R's visualization capabilities to create visually appealing and informative charts, graphs, and interactive dashboards. Visualizations should focus on key findings and insights, making complex data more accessible and engaging for the intended audience.

1. Executive Summary Report:

* Prepare a concise executive summary report that summarizes the key findings, insights, and recommendations derived from the analysis. The report should be clear, concise, and visually appealing, using R-generated visualizations to support the narrative.

1. Stakeholder Presentations:

* Conduct presentations to share the analysis results with stakeholders and decision-makers. Use R visualizations to illustrate the key insights, making use of storytelling techniques to effectively communicate the significance and implications of the findings.

1. Data Storytelling:

* Craft a compelling data-driven narrative that tells the story of the analysis. Use R visualizations to guide the storytelling process, presenting the data in a way that engages the audience and helps them connect with the insights and recommendations.

1. Interactive Online Platforms:

* Leverage R Shiny or other web-based visualization tools to create interactive online platforms that allow stakeholders to explore the data analysis results in a self-guided manner. These platforms can provide additional context, drill-down capabilities, and interactivity to enhance the understanding of the insights.

The "Share" phase of the data analysis methodology is crucial for ensuring the impact of the analysis results. By leveraging R's visualization capabilities, we can effectively communicate the key findings and insights derived from the analysis. Through compelling visualizations, executive summary reports, stakeholder presentations, and interactive online platforms, we can engage stakeholders, drive informed decision-making, and enable the implementation of strategies to achieve the desired outcomes.

1. **Act:** Key Findings from the Analysis:

* Ride Frequency and Duration: Casual riders have fewer rides compared to members, but their average ride lengths are higher.
* Bike Preference: The majority of riders prefer classic bikes, followed by electric bikes.
* Riding Patterns: Casual riders tend to ride more on weekends and in the late afternoon around 5 PM.
* Seasonal Influx: The third quarter experiences a higher influx of casual riders, indicating that many of them are likely tourists or visitors.
* Popular Station: The most popular station records more than double the number of rides compared to the second most popular station.

These insights provide valuable information about the behavior and preferences of casual riders and can be used to inform targeted marketing strategies aimed at converting them into annual members.