

# Google Case Study 2: Cyclistic Bike Share Case Study

## Introduction

### Cyclistic Marketing Analysis

As a junior data analyst on the marketing analyst team at Cyclistic, your objective is to understand how casual riders and annual members use Cyclistic bikes differently. This analysis aims to develop insights that will inform a new marketing strategy to convert casual riders into annual members. To proceed, Cyclistic executives require data-driven recommendations supported by compelling data insights and professional visualizations.

The company seeks to answer the following three key questions for the future marketing program based on the dataset:

How do annual members and casual riders use Cyclistic bikes differently?

Usage Patterns: Analyze the differences in trip duration, frequency, and time of day/week between annual members and casual riders.

Trip Locations: Investigate if there are specific stations or routes that are more popular among annual members compared to casual riders.

Bike Types: Determine if there's a preference for certain types of bikes (standard, electric, etc.) among annual members versus casual riders.

Why would casual riders buy Cyclistic annual memberships?

Cost Savings: Highlight the potential cost savings for frequent riders if they switch to an annual membership.

Convenience: Emphasize the added convenience of an annual membership, such as priority access to bikes and more flexible usage terms.

Incentives: Identify what incentives (e.g., discounts, exclusive access to events) could motivate casual riders to become annual members.

How can Cyclistic use digital media to influence casual riders to become members?

Targeted Campaigns: Use data insights to create targeted digital marketing campaigns aimed at casual riders, emphasizing the benefits of an annual membership.

Engagement Strategies: Develop engagement strategies on social media platforms to build a community and foster a sense of belonging among casual riders.

Personalized Offers: Leverage digital media to deliver personalized offers and promotions to casual riders based on their usage patterns and preferences.

### Tools used in Analysis

- Data Cleaning: Microsoft Excel
- Data Analysis: Microsoft SQL Server
- Visualization: Tableau Public

### Data Analysis Process

To effectively analyze Cyclistic data and address key business questions while making informed recommendations, the data analysis process will adhere to the following key steps:

- 1) Ask relevant questions
- 2) Prepare the data for analysis
- 3) Process the data to ensure accuracy and relevance
- 4) Analyze the data to derive meaningful insights
- 5) Share findings through clear and actionable reports or presentations
- 6) Act upon the insights to drive strategic decisions and improvements.

These steps ensure a structured approach to deriving insights from data for informed business decisions at Cyclistic.

## Ask Phase:

To address the business task of understanding how annual members and casual riders use Cyclistic bikes differently, the following questions will guide the analysis:

1. Percentage of Casual Riders vs Annual Members
2. Proportion of Rides by Bike Type
3. Percentage of Rides per Month
4. Number of Rides per Day

5. Average Ride Duration by Day (in minutes)
6. Frequency of Rides per Hour
7. Percentage of Rides per Season
8. Proportion of Rides in Each Time Zone

### Prepare Phase:

- The dataset used for this project is public data provided by Motivate International Inc. It consists of monthly files covering the period from January 2022 to December 2022.
- Each file within the dataset contains 13 columns related to the bike rides.
- These columns provide detailed information about each ride, including the ride ID, rideable type, start and end station ID's and locations, coordinates, and membership type.

### Process Phase:

#### Cleaning in Excel:

During the process phase, data cleaning steps were carried out in Microsoft Excel. The following actions were performed:

1. Checking for Duplicates: Duplicate values were identified and removed using the built-in "Remove Duplicates" feature in Excel. This helped ensure data integrity and eliminate any duplicated entries.

2. Validating Column Values: The values in specific columns, such as `rideable_type` and `member_casual`, were verified for consistency and accuracy. Only valid values, including `classic_bike`, `docked_bike`, and `electric_bike` for `rideable_type`, and `casual`, `member` for `member_casual`, were retained.

3. Removing Blank Values: The dataset was checked for incomplete or blank values across all columns. Rows with missing values, particularly in columns like `start_station_name`, `start_station_id`, `end_station_name`, and `end_station_id`, were removed to ensure data completeness.

4. Removing Unwanted Columns: Columns such as `start_lat`, `start_lng`, `end_lat`, and `end_lng`, which were not relevant to the analysis, were removed from each file to streamline the dataset and focus on essential variables.

5. Adding the Ride Length Column: A new column named `ride_length` was added to calculate the duration of each ride. The value in the `ride_length` column was obtained by subtracting the `started_at` timestamp from the `ended_at` timestamp.

6. Setting the Time Format: The `ride_length` column was formatted as "HH:MM".

7. Sorting the Table: The table was sorted in ascending order based on the `started_at` column to ensure data consistency.

By cleaning data in Excel, the dataset was refined, inconsistencies were addressed, and the ride length information was formatted appropriately for subsequent analysis.

## Data Transformation:

For data transformation, data processing was performed in SQL Server. The data from each month, spanning from January to December 2022, was imported and merged into a single table called “Annual\_trip\_data\_2022.” The steps involved in this process are as follows:

**Importing Data:** The monthly files containing the ride data were imported into SQL Server.

**Merging Data:** A new table named “Annual\_trip\_data\_2022” was created to store the consolidated data for the entire year. The data from the tables were merged into the “Annual\_trip\_data\_2022” table using the UNION ALL statement. SQL Query: Merge Data [(<https://github.com/JJRockzzzz/Google-Data-Analytics-Capstone-Project---2/blob/main/Merge%20Data.sql>)]

**Data Manipulation:** After merging data, a new table named “analyze\_annual\_trip\_data” was generated. This table includes additional columns such as “month,” “day,” “hour,” and “duration\_minutes” to facilitate the comparison and analysis of ride frequencies. SQL Query: Manipulate Data

[(<https://github.com/JJRockzzzz/Google-Data-Analytics-Capstone-Project---2/blob/main/Manipulate%20Data.sql>)]

## Analyze:

In the Analyze phase, we delve into the data to uncover insights and address the key findings related to how annual members and casual riders use Cyclistic bikes differently. The focus is on understanding their behavior, preferences, and patterns to inform marketing strategies aimed at

converting casual riders into annual members. To address the key findings, the following analyses were performed in SQL Server.

### 1. Percentage of Casual Riders vs Annual Members

	membership_type	ride_count	total_membership	membership_percentage
1	casual	1743720	4288562	40.66%
2	member	2544842	4288562	59.34%

### 2. Proportion of Rides by Bike Type

	bike_type	membership_type	individual_membership_count	total_membership	membership_percentage
1	classic_bike	member	1638280	2508658	65.31%
2	classic_bike	casual	870378	2508658	34.69%
3	docked_bike	casual	172580	172580	100.00%
4	electric_bike	member	906562	1607324	56.40%
5	electric_bike	casual	700762	1607324	43.60%

### 3. Percentage of Rides per Month

	membership_type	months	membership_rides	total_rides_per_month	membership_percentage
1	casual	January	710	1735	40.92%
2	member	January	1025	1735	59.08%
3	casual	February	14958	87638	17.07%
4	member	February	72680	87638	82.93%
5	casual	March	89638	283702	31.60%
6	member	March	194064	283702	68.40%
7	casual	April	91001	269002	33.83%
8	member	April	178001	269002	66.17%
9	casual	May	216918	494065	43.90%
10	member	May	277147	494065	56.10%
11	casual	June	287526	609763	47.15%
12	member	June	322237	609763	52.85%
13	casual	July	307582	633076	48.59%
14	member	July	325494	633076	51.41%
15	casual	August	265729	594286	44.71%
16	member	August	328557	594286	55.29%
17	casual	September	217468	525288	41.40%
18	member	September	307820	525288	58.60%
19	casual	October	148853	406299	36.64%
20	member	October	257446	406299	63.36%
21	casual	November	72359	251105	28.82%
22	member	November	178746	251105	71.18%
23	casual	December	30978	132603	23.36%
24	member	December	101625	132603	76.64%

#### 4. Number of Riders per Day

	membership_type	weekdays	ride_duration
1	casual	Sunday	27
2	member	Sunday	14
3	casual	Monday	24
4	member	Monday	12
5	casual	Tuesday	21
6	member	Tuesday	11
7	casual	Wednesday	20
8	member	Wednesday	12
9	casual	Thursday	21
10	member	Thursday	12
11	casual	Friday	22
12	member	Friday	12
13	casual	Saturday	26
14	member	Saturday	14

#### 5. Average Ride Duration by Day (in minute)



	membership_type	weekdays	ride_duration
1	casual	Sunday	27
2	member	Sunday	14
3	casual	Monday	24
4	member	Monday	12
5	casual	Tuesday	21
6	member	Tuesday	11
7	casual	Wednesday	20
8	member	Wednesday	12
9	casual	Thursday	21
10	member	Thursday	12
11	casual	Friday	22
12	member	Friday	12
13	casual	Saturday	26
14	member	Saturday	14

## 6. Frequency of Rides per Hour

	membership_type	hours	rides_per_hour
1	casual	0	32940
2	member	0	24810
3	casual	1	21118
4	member	1	15302
5	casual	2	12579
6	member	2	8424
7	casual	3	7143
8	member	3	5195
9	casual	4	4631
10	member	4	6021
11	casual	5	8308
12	member	5	25005
13	casual	6	21487
14	member	6	73198
15	casual	7	37723
16	member	7	139245
17	casual	8	52020
18	member	8	162152
19	casual	9	54223
20	member	9	109477
21	casual	10	71550
22	member	10	101642
23	casual	11	93686
24	member	11	121731
25	casual	12	110675
26	member	12	140516

## 7. Percentage of Rides per Season

	membership_type	season	rides_per_season	total_rides_per_season	percentage_ride_per_season
1	casual	Autumn	438680	1182692	37.09%
2	member	Autumn	744012	1182692	62.91%
3	casual	Spring	397557	1046769	37.98%
4	member	Spring	649212	1046769	62.02%
5	casual	Summer	860837	1837125	46.86%
6	member	Summer	976288	1837125	53.14%
7	casual	Winter	46646	221976	21.01%
8	member	Winter	175330	221976	78.99%

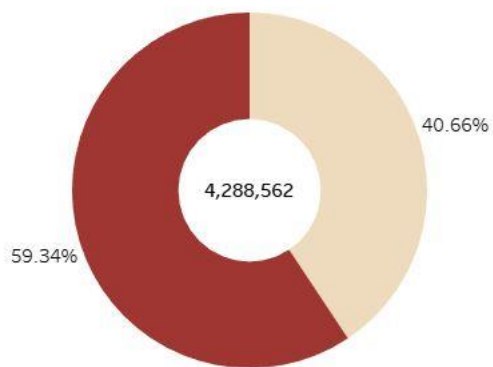
## 8. Percentage of Rides per Time Zone

	membership_type	time_zone	rides_per_time_zone	total_rides_per_time_zone	percentage_of_rides
1	casual	Afternoon	801400	1884805	42.52%
2	member	Afternoon	1083405	1884805	57.48%
3	casual	Evening	524912	1194147	43.96%
4	member	Evening	669235	1194147	56.04%
5	casual	Morning	330689	1038134	31.85%
6	member	Morning	707445	1038134	68.15%
7	casual	Night	86719	171476	50.57%
8	member	Night	84757	171476	49.43%

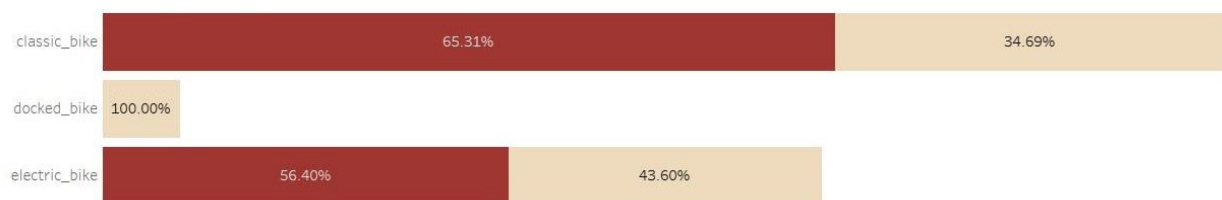
## Share:

The share phase presents the insights and findings derived from the analysis of Cyclistic Bike Share data using Tableau Public, a powerful data visualization tool. The analysis revealed several key findings:

### 1. Percentage of Casual Riders vs Annual Members



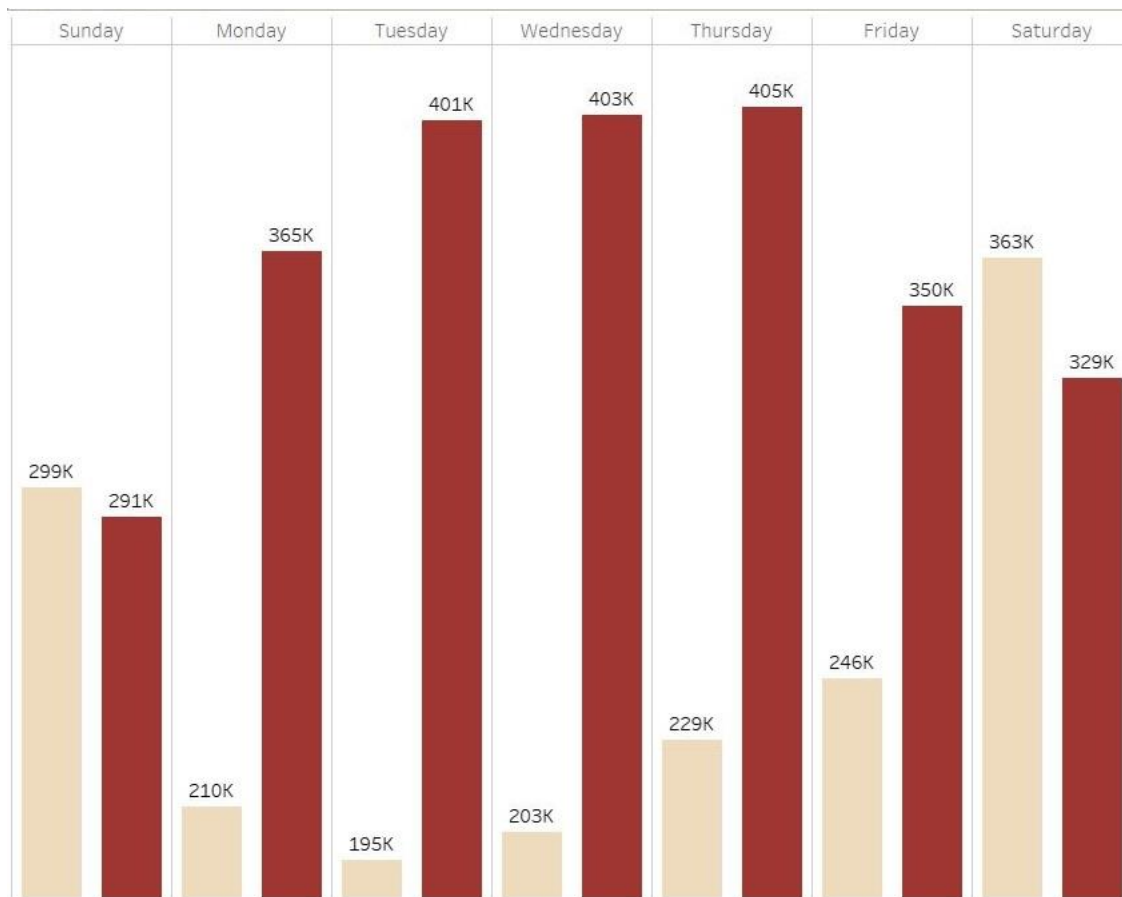
## 2. Proportion of Rides by Bike Type



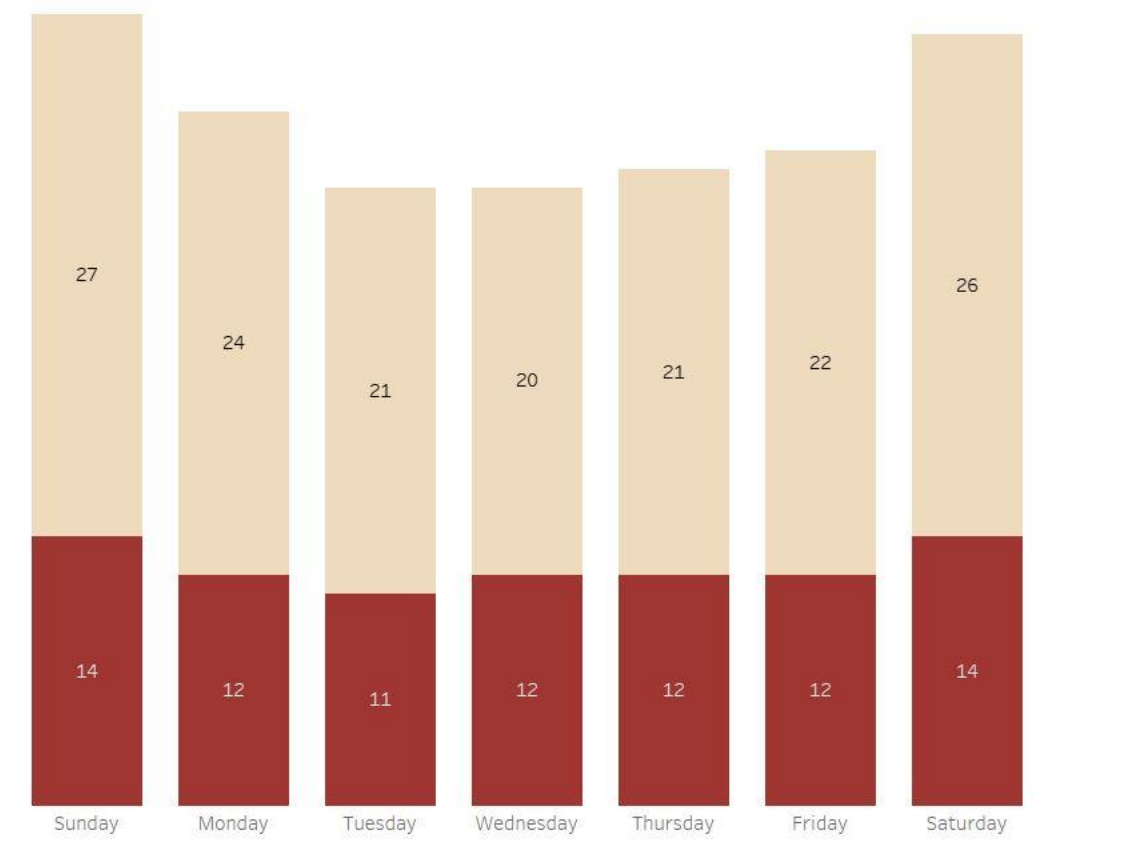
## 3. Percentage of Rides per Month



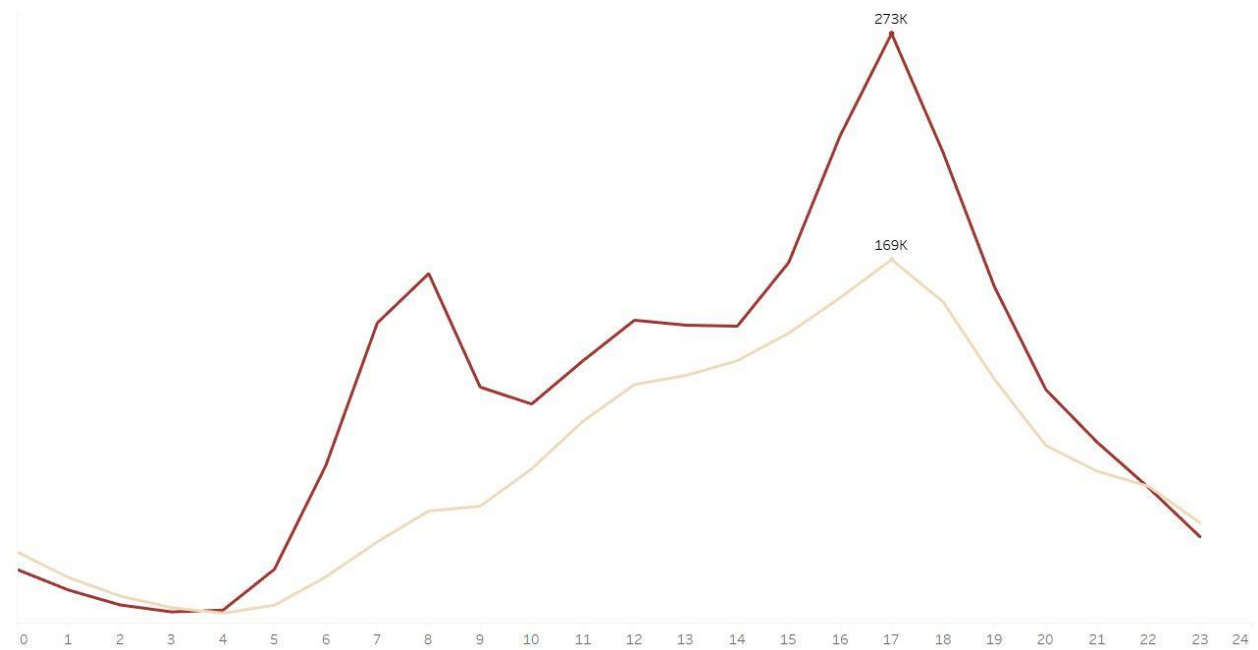
## 4. Number of Rides per Day



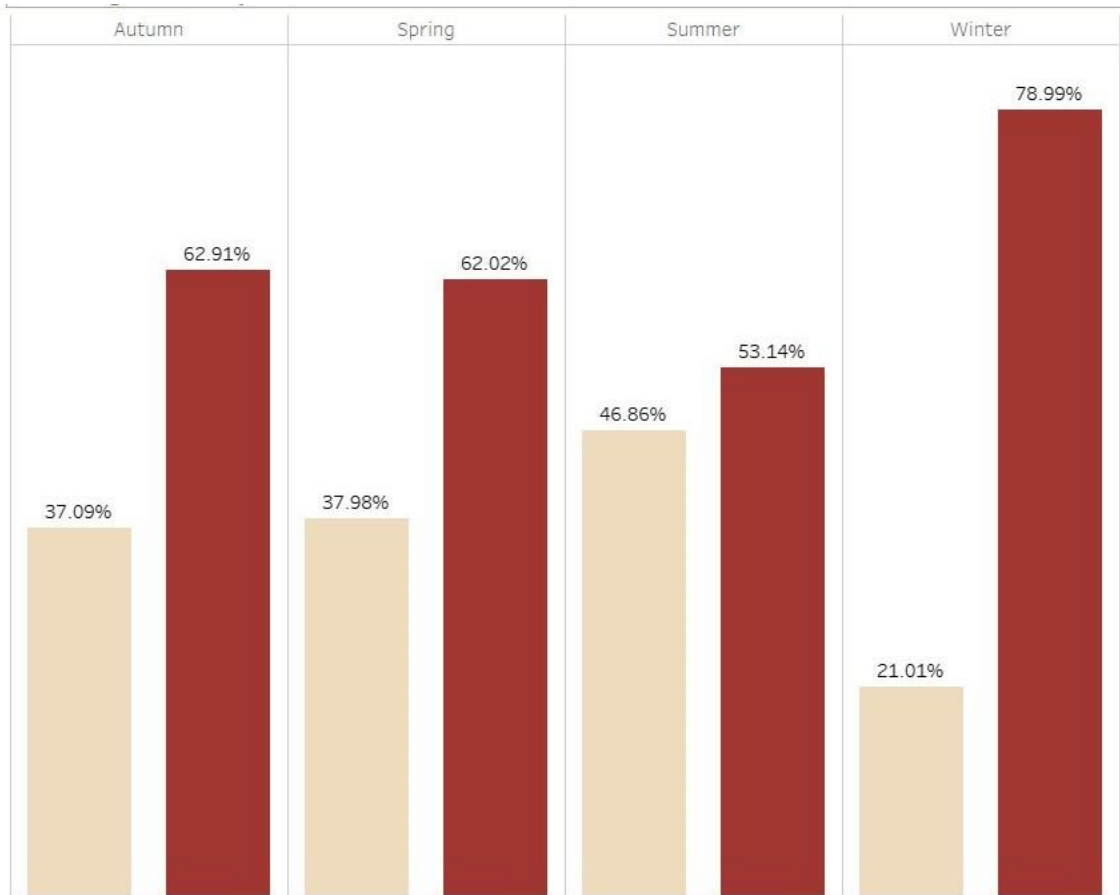
##### 5. Average Ride Duration by Day (in minutes)



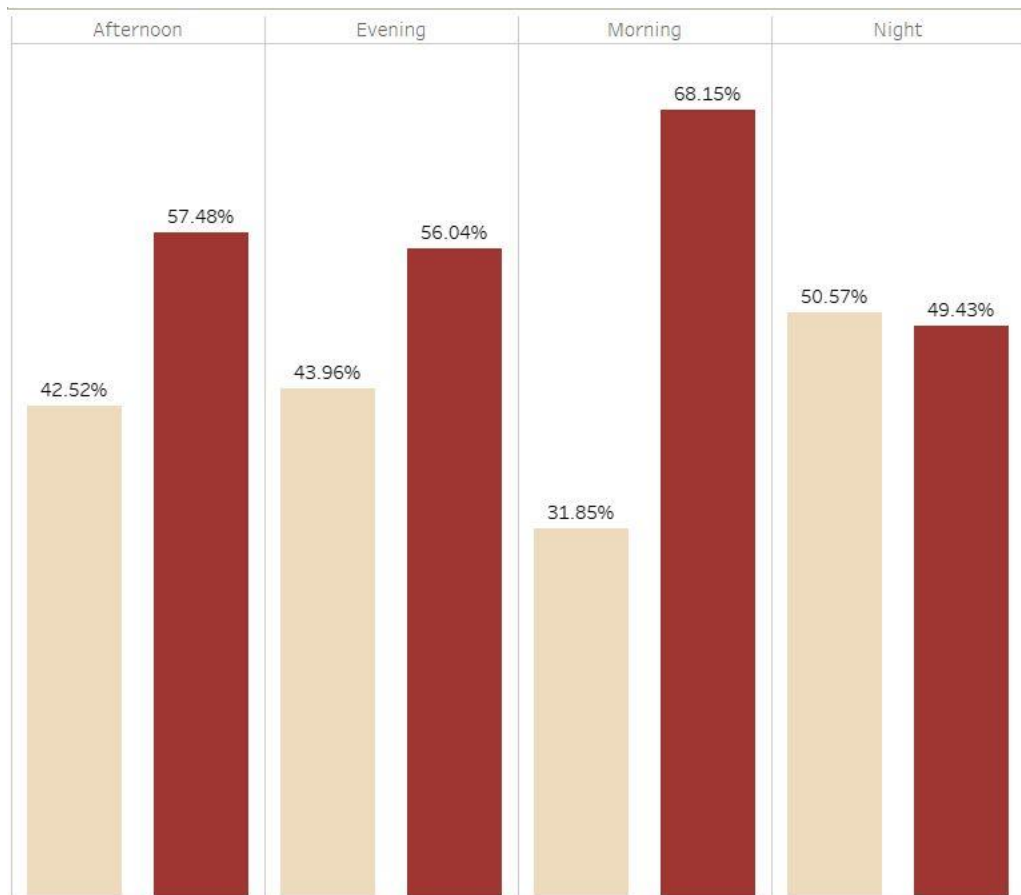
## 6. Frequency of Rides per Hour



## 7. Percentage of Rides per Season



8. Proportion of Rides in each Time Zone



Act :

### Key Takeaways

Based on the findings of my analysis, several key takeaways can be derived:

- 1) Bike Type Preference: Annual members show a higher preference for classic and electric bikes, while casual riders are more inclined toward docked bikes.
- 2) Monthly Variation: Annual members consistently have more rides than casual riders each month. February sees the highest proportion of annual members' rides, while July has the highest proportion of casual rides.

- 3) Weekday vs. Weekend Usage: Annual members tend to ride more on weekdays, whereas casual riders have higher usage on weekends.
- 4) Ride Duration: Casual riders have longer average ride durations compared to annual members.
- 5) Peak Usage Hour: Both annual members and casual riders experience peak usage at 5 p.m., with annual members having a higher number of rides.
- 6) Seasonal Variation: Members have the highest proportion of rides in all seasons, but winters see the highest proportion of annual members' rides, while summers see the highest proportion of casual rides.
- 7) Time Zone Distribution: Annual members dominate the morning, afternoon, and evening time zones, while casual riders have a higher proportion of rides during the night.

### Recommended Strategies

- 1) Targeted Promotions: Offer exclusive discounts or incentives for casual riders to upgrade to annual memberships, emphasizing the benefits and cost savings of long-term commitment.
- 2) Improved Bike Variety: Expand the fleet of classic and electric bikes to cater to the preferences of casual riders and enhance their overall riding experience.
- 3) Customized Membership Plans: Introduce tailored annual membership plans specifically for docked bike users, providing additional perks and advantages to encourage their transition.

By implementing these strategies, Cyclistic can enhance user experience, increase annual memberships, and better meet the diverse needs of its riders.