Cyclistic

Transforming Casual Riders into Annual Members

A Data-Driven Analysis

I - Mission

By Roxanne Desgagnés

Carl Hunley Jr

OVERVIEW

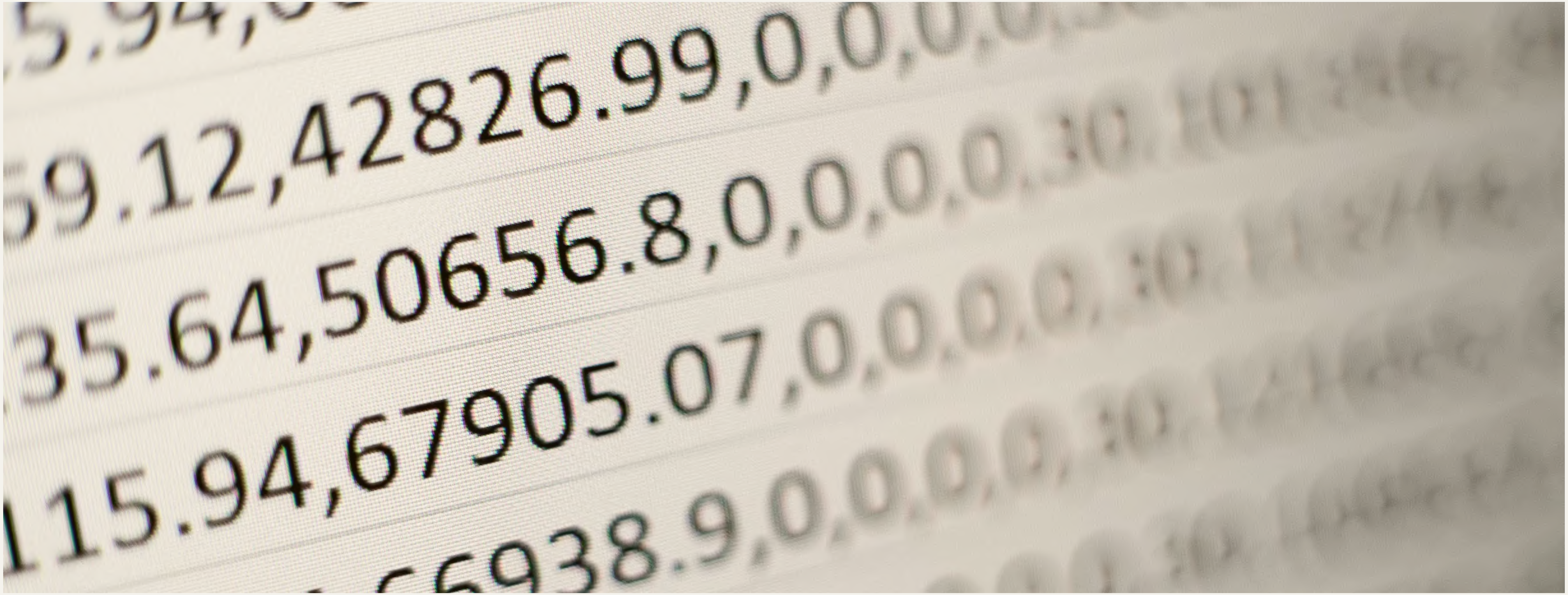
Business Task

Cyclistic is a fictional company modeled after a real service: the Divvy bike rental service in Chicago.

The primary goal of this analysis is to identify key behavioral differences between Cyclistic's casual riders and annual members, with the ultimate aim of converting casual riders into annual members.

By Carl Hunley Jr

II - Data Sources



By Mika Baumeister

DATA SOURCES

Data Origin

The dataset utilized for this analysis was obtained from an AWS folder shared by Lyft Bikes and Scooters LLC ('Bikeshare'), which operates Chicago's Divvy bike-sharing service.

The data adheres to the 'ROCCC' principles - Reliable, Original, Comprehensive, Current, Cited. it is publicly shared by the City of Chicago through Bikeshare, the primary stakeholders of the project. Sourced directly from the records of shared bicycles and docking stations managed by Lyft, the data is therefore highly credible and well-sourced.

By Aveedibya Dey

DATA SOURCES

Format of Data

The dataset is available in zipped .csv files, organized by year (from 2013 to 2025), and/or further subdivided by quarter or month.

Although the data is continuously updated, the 2023 dataset was selected for this analysis as it was the most complete at the outset, and to align with the original request dating back to 2019.

Initial inspection of the files in a spreadsheet revealed missing and incorrect data, necessitating cleaning and preparation prior to analysis.

By Dulcey Lima

DATA SOURCES

Data Details

The uncompressed .csv files for the 2023 dataset totaled 1.07 GB before cleaning.
The final consolidated .csv file was 0.98 GB.
The dataset encompasses all trips made by Bikeshare service users throughout 2023.
Key variables include a unique ride identifier, ride duration (departure and arrival time), start and end stations name, bike type, and user type (annual members or casual riders).

By Clique Studios

III - Data Cleaning & Preparation

By Max Bender

DATA CLEANING & PREPARATION

Context

The **12 .csv files**, containing **Bikeshare's 2023 operational data**, were decompressed and **consolidated into a unified dataset** using **R** and **RStudio**. Their considerable size precluded direct use in a spreadsheet.

The **raw dataset** comprises **10,050,983 observations** (approximately ten millions) across **14 variables**.

Initial examination of the raw data revealed **several issues that necessitated dataset cleaning**.

Given the inability to consult with the database managers, data stewards, or data engineers responsible for data collection to understand the nature of problematic entries and apply targeted modifications, **a rather rigorous cleaning approach was adopted**.

By Erica Fischer

DATA CLEANING & PREPARATION

Cleaning Process

Observations with missing data or inconsistencies (e.g., departure time later than arrival time) were removed.

This cleaning reduced the dataset to 4,331,106 validated observations (approximately four million), a 56.9% decrease from the original. While such a significant reduction in a real-world project would demand thorough validation and issue resolution, for this exercise, the data will be used as is.

Three new columns were calculated and appended: trip duration, corresponding day of the week, and trip season, bringing the total number of variables to 17.

By Erica Fischer

DATA CLEANING & PREPARATION

Code with R & RStudio

For this project, **R** was chosen as the **primary language**, with **RStudio** serving as its **IDE**.
Several libraries were utilized during **data cleaning** and **preparation**, including **tidyverse** and **readr**.

Initially, the **various datasets** (one per month of the year 2023) are imported as **.csv files** and **merged** to form a **single dataframe**.

The **dataframe** is then **cleaned** of observations containing **missing data** and **new columns are calculated**, like the season and the day of the week corresponding to the date of the trip, and the travel time for each trip.

R Code

```
## Google Data Analytics Career Certificate - Cyclistic Case Study
## Analyzing Cyclistic/Divvy bike sharing data to understand usage and try to get more annual
members from casual users

## Import & Installation of Libraries
# ... (see full code)

## Building the main dataframe from .csv files in the working directory
df2023 <- list.files(pattern = "*.csv", full.names = T) %>%
  map_df(~read_csv(.))

## Getting rid of null values in the dataframe
df2023 <- drop_na(df2023)

## Time conversions, and creation of a new column with calculated travel time for each trip
df2023$started_at <- ymd_hms(df2023$started_at)
df2023$ended_at <- ymd_hms(df2023$ended_at)
df2023$travel_time <- df2023$ended_at - df2023$started_at
df2023$travel_time <- as.numeric(df2023$travel_time, units="mins")

## Removing irrelevant/falty travel times
df2023 <- subset(df2023, travel_time >= 1)

## Creation of new columns for day of the week and season
df2023$day_of_week <- strftime(df2023$started_at, "%A")
df2023$season <- time2season(df2023$started_at, out.fmt = "seasons")
df2023 <- df2023 %>%
   mutate(season = recode(season,
                          "spring" = "Spring",
                          "summer" = "Summer",
                          "autumn" = "Autumn",
                          "winter" = "Winter"),
                          season = factor(season, levels = c("Spring", "Summer", "Autumn", "Winter")))
```

DATA CLEANING & PREPARATION

Code with R & RStudio

The next phase involves data analysis and visualization.

Key metrics like total trips, average travel times, and top 10 destinations were calculated and plotted using ggplot2 for each user type to gain behavioral insights.

Further comparisons identified relevant information regarding casual riders and annual members, which was incorporated into the analysis. Less relevant findings were discarded.

The complete R code is available for review here: https://github.com/TonyChamCham/googleDACC_cyclistic/blob/main/241128_CaseStudy_Cyclistic.R

R Code

```
## Counting the number of trips by type of user
df2023 %>%
  count(member_casual) %>%
  ggplot() +
  geom_col(mapping = aes(x = member_casual, y = n, fill = member_casual)) +
  geom_text(mapping = aes(x = member_casual, y = n, label = n, fontface = "bold"),
            nudge_y = -50000) +
  theme(axis.text.x = element_text(face = "bold"), axis.title = element_text(face = "bold"), title =
element_text(face = "bold")) +
  labs(x = "User Type", y = "Number", title = "Number of service uses by user type")

## Average travel times by type of user
# Calculating Averages
averages <- df2023 %>%
  group_by(member_casual) %>%
  summarise(mean_time = mean(travel_time))

# Graph with percentages labels
ggplot(df2023, aes(x = member_casual, y = travel_time, fill = member_casual)) +
  geom_bar(stat = "summary", fun = "mean", position = "dodge") +
  geom_text(data = averages, aes(x = member_casual, y = mean_time, label = round(mean_time, 1)),
            vjust = -0.5, fontface = "bold") +
  theme(axis.text.x = element_text(face = "bold"),
        axis.title = element_text(face = "bold"),
        title = element_text(face = "bold")) +
  labs(x = "User Type",
       y = "Average travel time (min)",
       title = "Average travel time by user type")

## Extracting and plotting the top 10 end stations of casual users
df2023_casuals_shortlist_endstation <- select(df2023_casuals, end_station_name)
df2023_casuals_shortlist_endstation %>%
  count(end_station_name) %>%
  arrange(desc(n)) %>%
  top_n(10, n) %>%
  ggplot() +
  geom_col(mapping = aes(x = reorder(end_station_name, -n), y = n, fill = end_station_name)) +
  geom_text(aes(x = reorder(end_station_name, -n), y = n, label = n, fontface = "bold"),
            nudge_y = -500) +
  theme(axis.text.x = element_text(angle = 90, face = "bold", hjust = 1, vjust = 0.2),
        axis.title = element_text(face = "bold"),
        title = element_text(face = "bold")) +
  labs(x = "End station", y = "Number of trips", title = "Top 10 end stations of casual users")
```

IV - Data Analysis

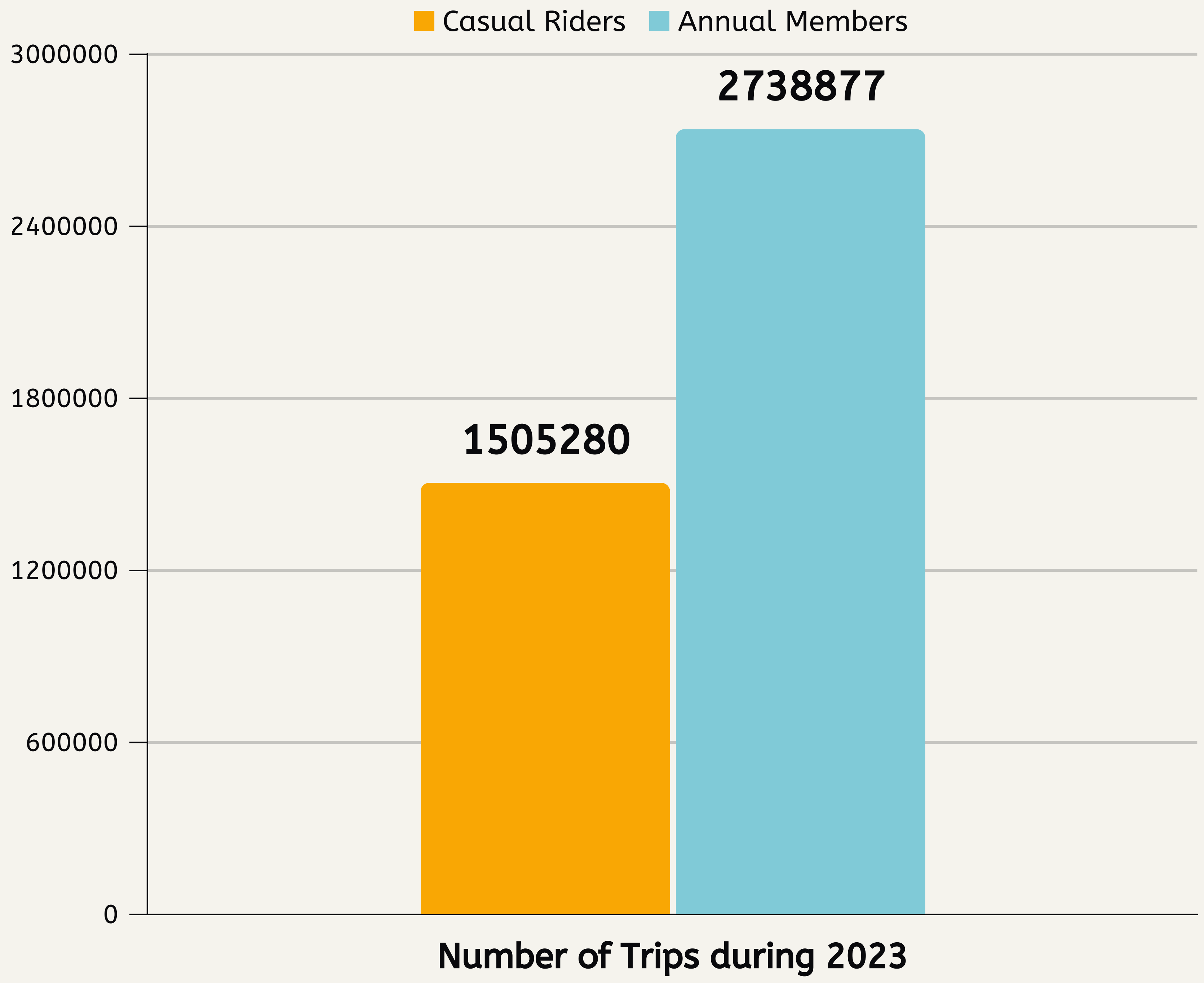
By UX Indonesia

DATA ANALYSIS

Trips By User Type

To contextualize the user base's service engagement, a count of trips by user type (casual riders vs. annual members) was performed for the entire year.

In 2023, annual members accounted for approximately 82.7% more service usage (about 2.8 million trips in the analyzed sample) than casual riders (about 1.5 million trips in the analyzed sample).



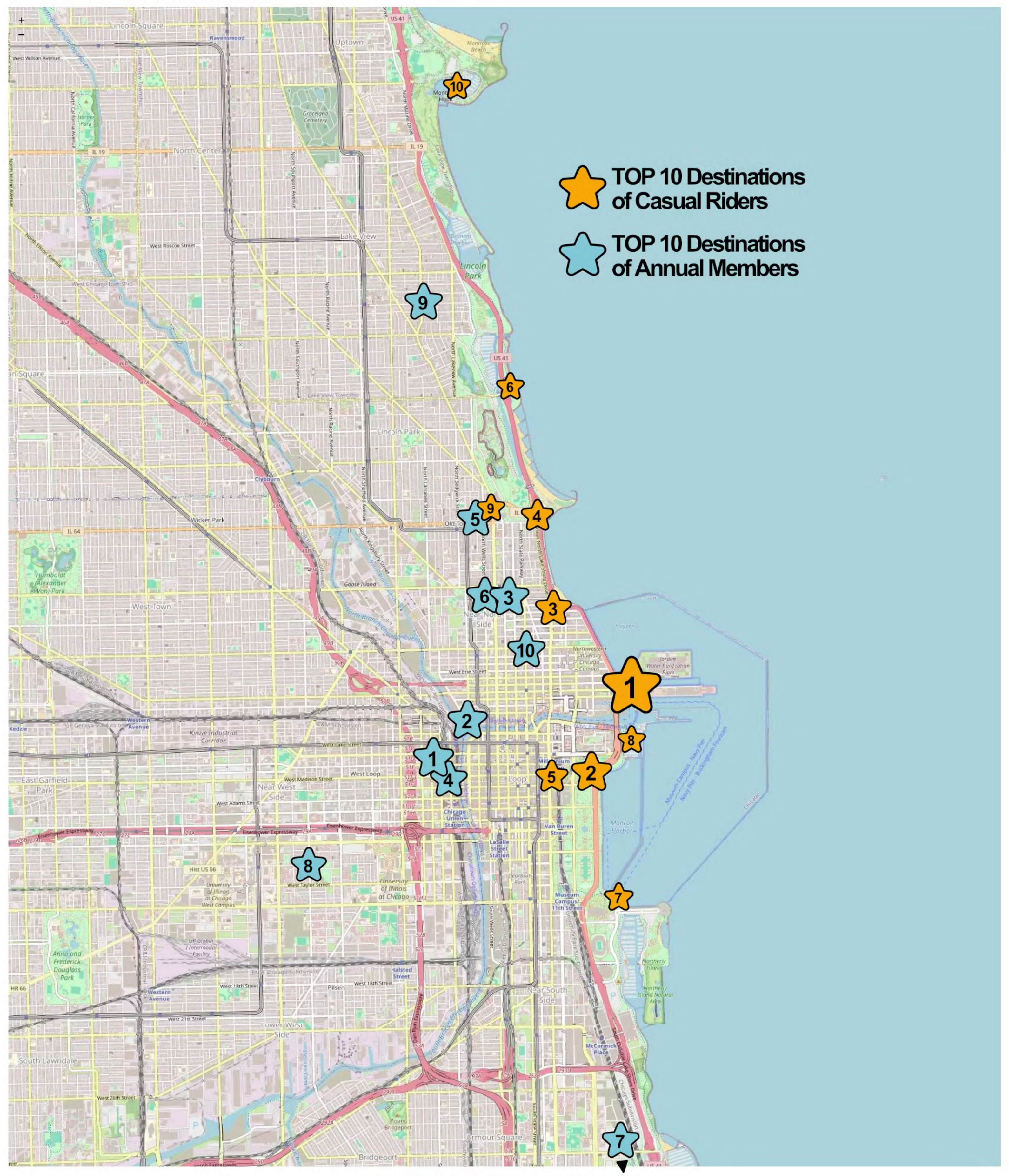
DATA ANALYSIS

Top 10 Destinations

Analyzing usage differences between the two user types, the most frequent arrival stations reveal distinct patterns.

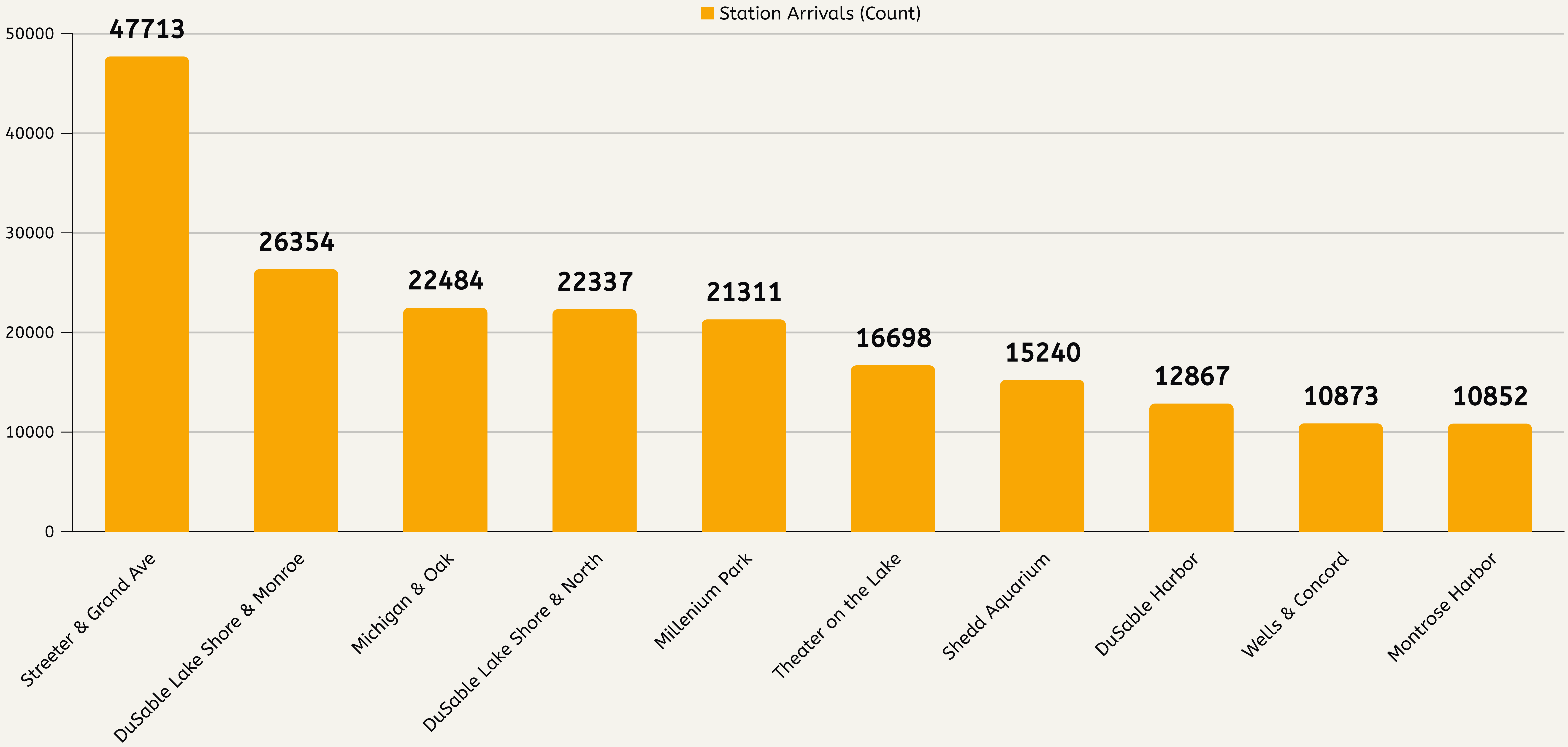
Most of the top 10 destinations for annual members are located in Downtown Chicago's business district, with Clinton St & Washington Blvd (25,475 arrivals) being the primary station, closely followed by Kingsbury St & Kinzie St (24,312 arrivals).

Conversely, the top destination for casual riders (by a significant margin) is Streeter Dr & Grand Ave station at Navy Pier (47,713 arrivals), a hub for cultural venues, entertainment, and green spaces. Subsequent top destinations further confirm this trend, consistently being near the city's green spaces and/or cultural sites.



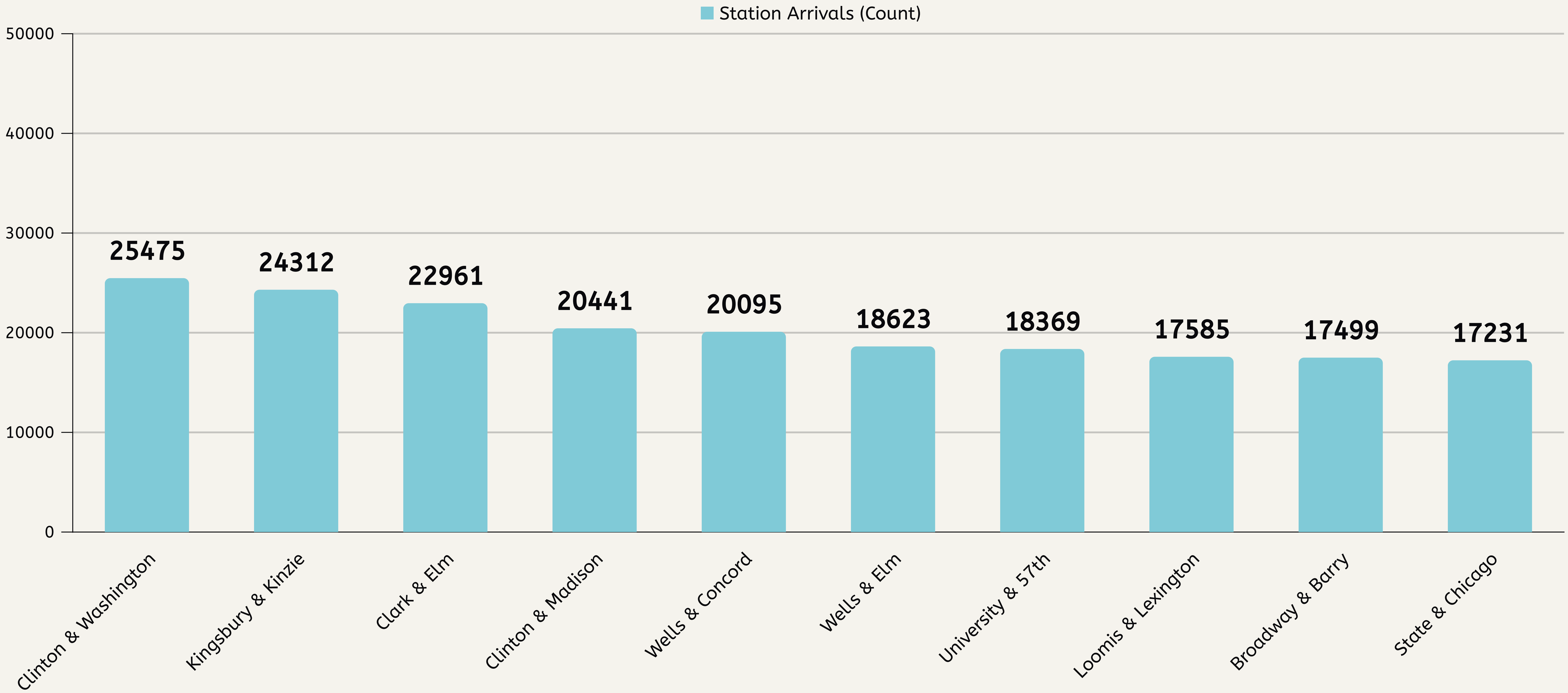
DATA ANALYSIS

Top 10 Destinations of Casual Riders



DATA ANALYSIS

Top 10 Destinations of Annual Members



DATA ANALYSIS

Observations on Destination

By comparing the total 2023 trips for each user category (approx. 2.7 million for annual members and 1.5 million for casual riders) with the total arrivals in their respective top 10 destinations (approx. 202,000 for annual members and 206,000 for casual riders), we can infer that annual members' destination long tail decreases less rapidly than that of casual riders.

This suggests casual riders focus on a more limited set of destinations (primarily parks and recreation), whereas annual members' destinations are more numerous, varied, and likely work-related, given their concentration in the business district.

By Carl Hunley Jr

DATA ANALYSIS

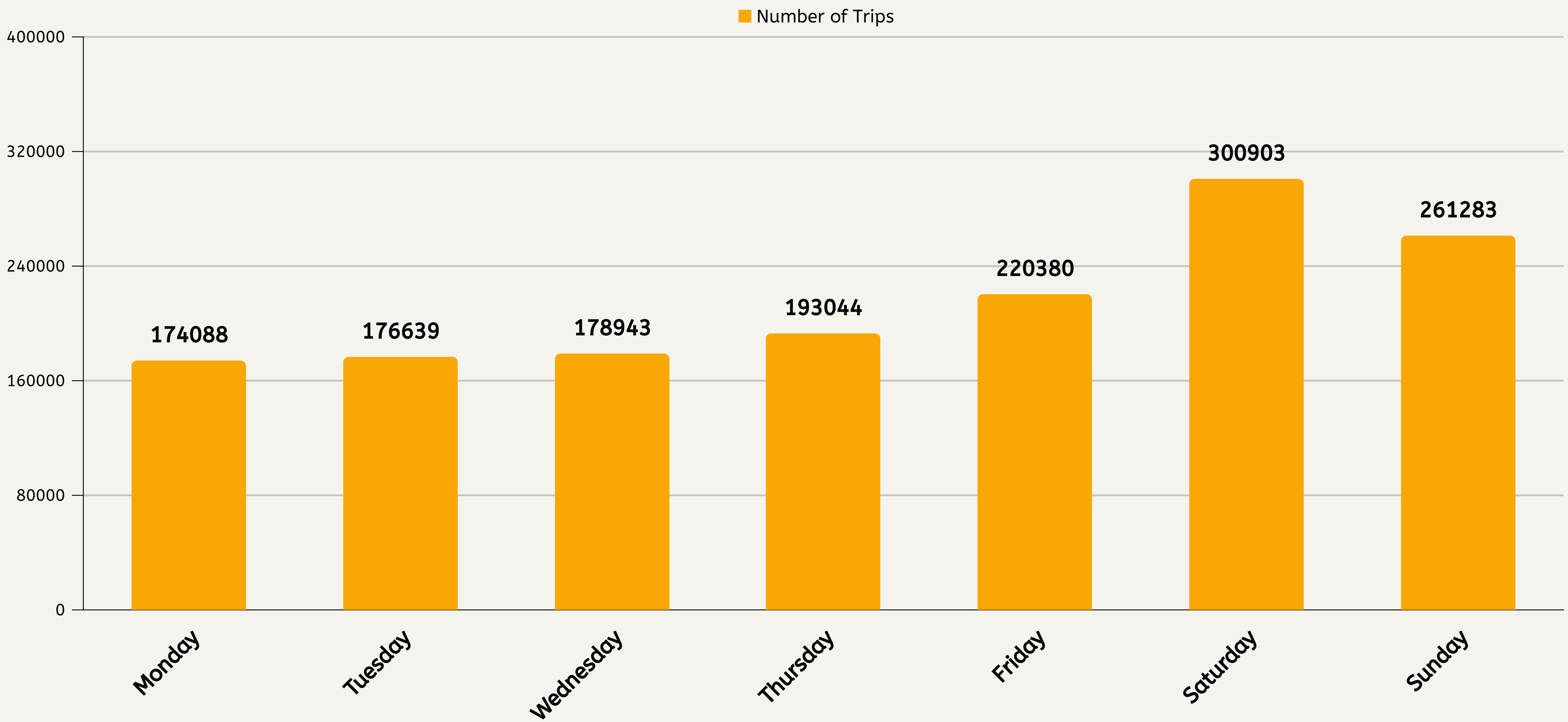
Observations on Destination

This differentiation is further confirmed by analyzing the preferred days of the week for each user type: annual members primarily use the service during weekdays, whereas casual riders predominantly use it on weekends, as illustrated in the graphs on the following pages.

By Evan Wise

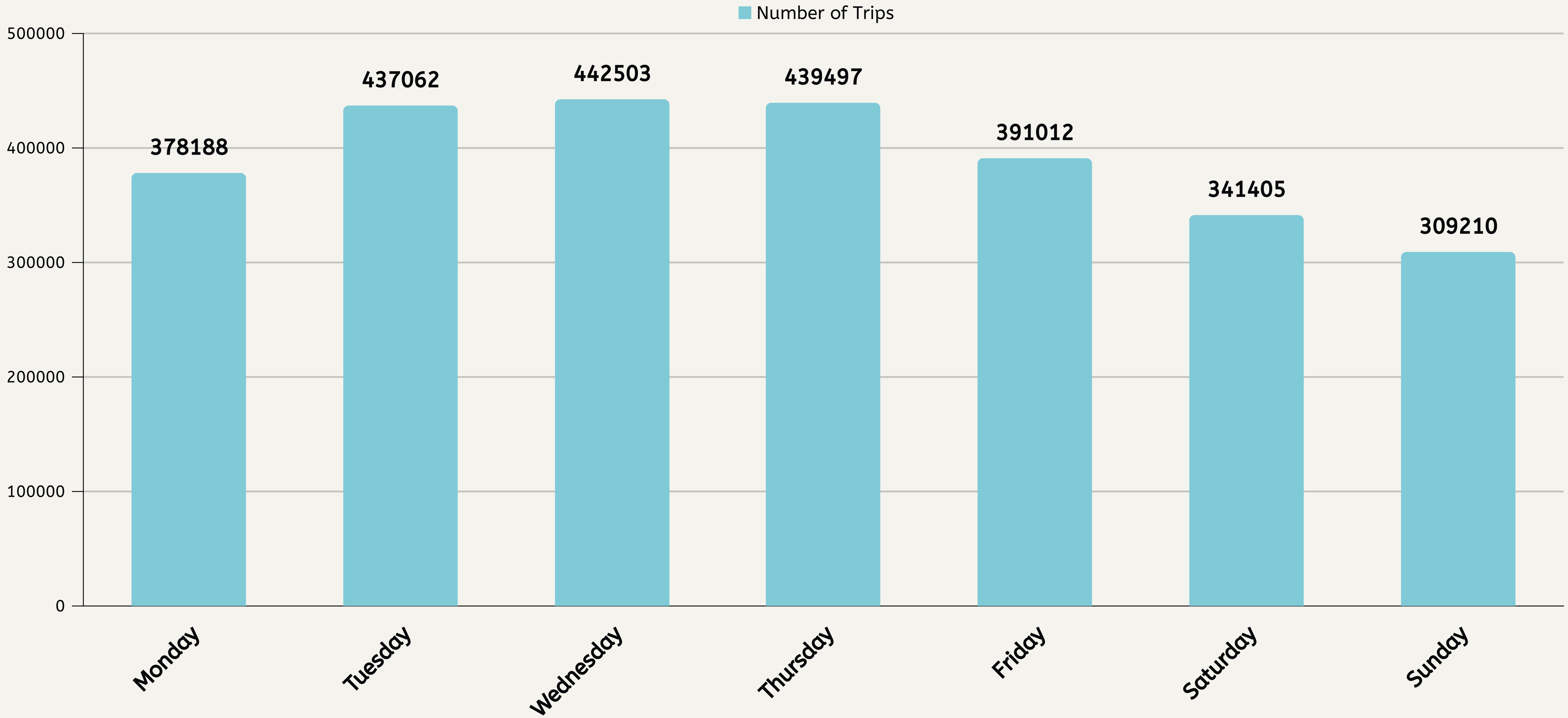
DATA ANALYSIS

Casual Rider Trips by Day of Week



DATA ANALYSIS

Annual Members Trips By Day of Week



DATA ANALYSIS

Seasonality

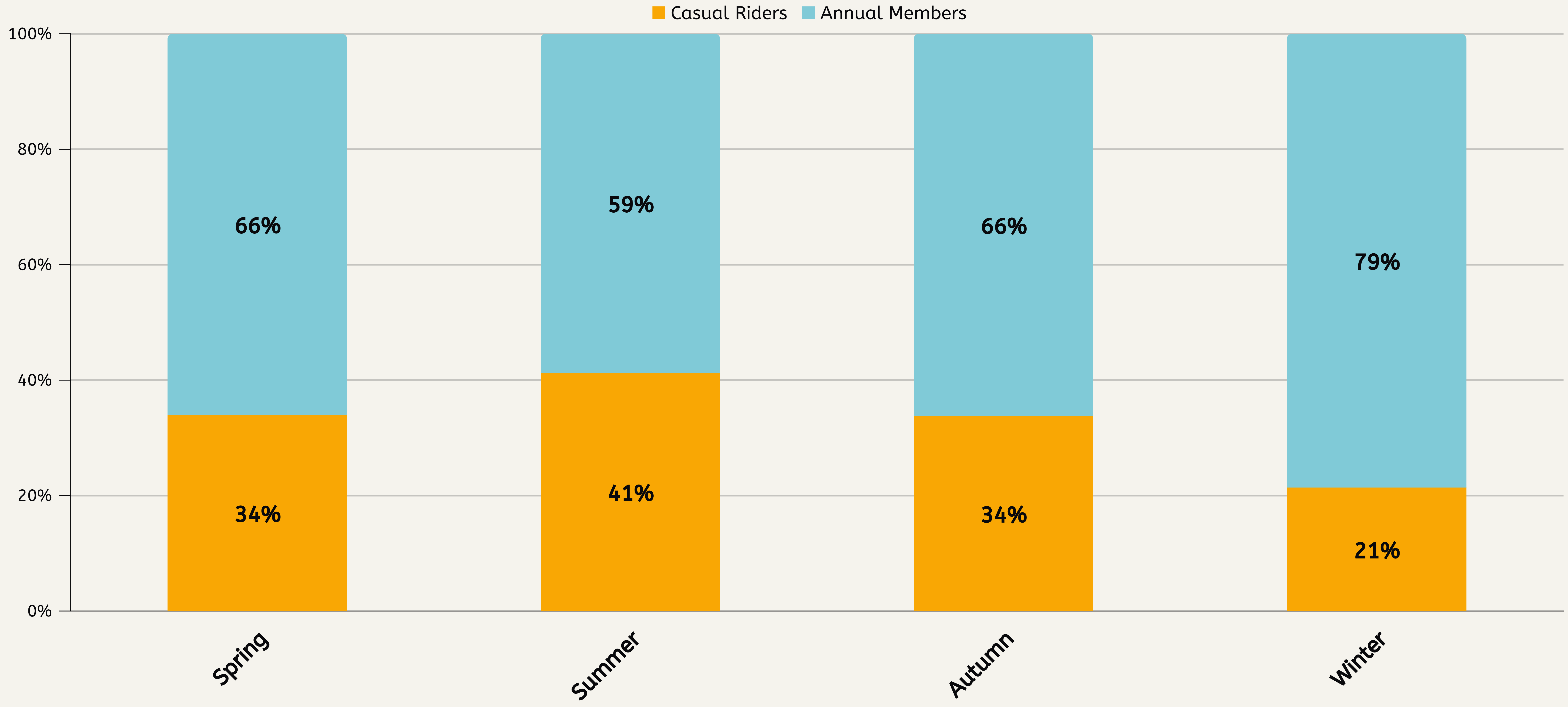
Analyzing the seasonality of trips taken by casual riders and annual members (next slide), we observe that the proportion of trips made by casual riders increases in summer and decreases in winter.

This further confirms their use of the service for leisure, as they are more active in warmer months and avoid harsh winters. Conversely, annual members consistently use the service for commuting, regardless of the season.

By Evan Wise

DATA ANALYSIS

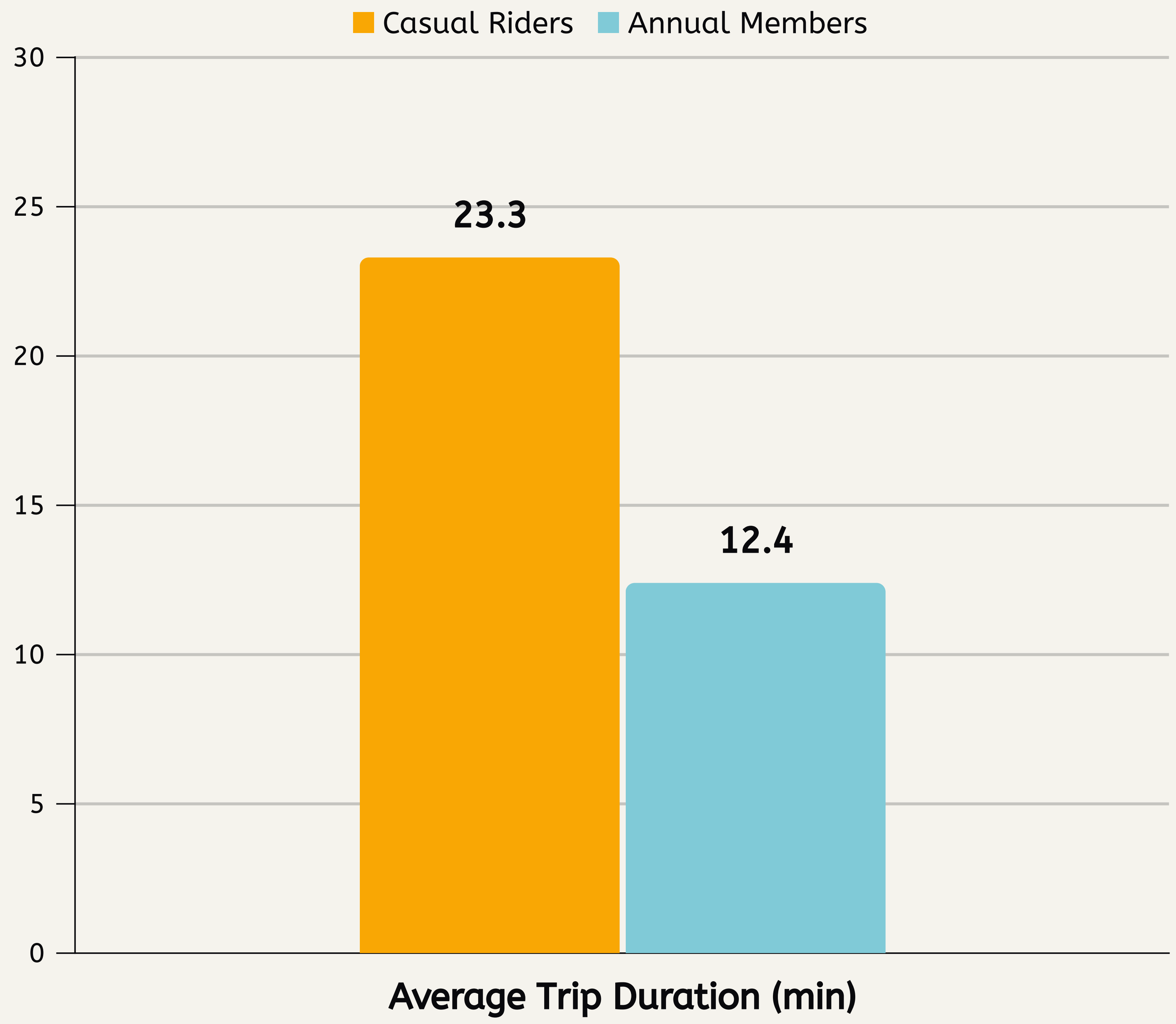
User Type Proportion Across Seasons



DATA ANALYSIS

Average Travel Time

Finally, comparing the average travel times between user types reveals that casual riders spend approximately twice as much time per trip biking than annual members.
This further supports the interpretation of casual riders' trips as recreational and annual members' trips as utilitarian.



DATA ANALYSIS

Key Findings

These observations collectively outline two distinct user profiles: the annual member is an active individual who primarily uses their bike for commuting to work. Their commutes are typically quick, frequent (several times a week), year-round, and predominantly directed towards the city's business center.

The casual rider, conversely, primarily uses the service on weekends (likely due to weekday work commitments) for leisure trips. These rides are typically at a leisurely pace, mainly targeting entertainment and cultural destinations, and occur more frequently in summer than in winter.

By Marco Verch

V - Recommandations

By Lance Anderson

RECOMMENDATIONS

Marketing Recommendations

Leveraging the established profiles of both casual riders and annual members, we can devise several marketing initiatives aimed at encouraging the conversion of casual riders to annual members.

Understanding the habits and behaviors of casual riders facilitates crafting messages that resonate with them.

By Steven Vance

RECOMMENDATIONS

First Recommendation

Given that casual riders predominantly utilize the service on weekends, an intermediate subscription specifically for weekend use could be introduced.
This subscription would be more financially attractive than pay-per-use options and could also be offered in couple and family versions, with additional benefits.
Once subscribed for weekend use, and naturally encouraged to use the service more frequently due to the subscription, the transition to a full annual membership would be facilitated.

By Praswin Prakashan

RECOMMENDATIONS

Second Recommendation

An advertising campaign promoting Cyclistic as an alternative to usual commuting methods would be displayed at the busiest stations frequented by casual riders.

As existing users, they would be more readily convinced than new users, provided that the friction points hindering their full adoption are addressed.

Surveys addressing these friction points would be conducted at these same busiest stations.

By Mariusz Pierog

RECOMMENDATIONS

Third Recommendation

Given that casual riders generally favor recreational use of the service, partnerships with leisure and cultural venues in the city can be established.

Cyclistic will promote these partners, notably through its app. In return, partners will offer exclusive discounts, free tickets, or invitations to special events, available exclusively to annual members.

Offering rewards directly aligned with the known interests of casual riders will more effectively incentivize them to subscribe to the service.

By Chris Nguyen

VI - Conclusion

By Muzammil Soorma

CONCLUSION

Converting Casual Riders

Data analysis reveals that casual riders and annual members represent two distinct user profiles, rather than a single profile at different stages of service engagement.

Converting casual riders can only be achieved by tailoring strategies to their specific profile, offering solutions aligned with their unique interests, issues, and needs, and by fine-tuning communication directed at them.

By Steven Vance

Thanks for your time!

VII - Contact

By Pedro Lastra

CONTACT

Let's get in touch!

I would be delighted to answer any questions or receive any feedback regarding this case study. Additionally, feel free to connect with me. I am always open to making new contacts!

LinkedIn
https://www.linkedin.com/in/antony-champane-045a6a68/

Email
antony.champane@outlook.fr

By Nathan Dumlao