Cyclistic Bike-Share

oogle Data Analytics

Capstone Project

About Me

HILMI ALFA RYAN

Fresh Graduate | Data Analyst

Ms. Excel | Python (Pandas, NumPy, Matplotlib, Seaborn, Plotly) | PostgreSQL

Tableau | Google Data/Looker Studio | Ms. Access

Go to Project

Capstone Project

Overview

It's the capstone project of **Google Data Analytics Professional Certificate** course through **Kominfo Fresh Graduate Digital Talent Scholarship** program. This capstone project involves me as a **Data Analyst** working in the **marketing analytics team** at the **Cyclistic Bike Share company**.

I will use **PostgreSQL** for database creation and joins between tables (datasets), as well as **Python (Pandas, NumPy, Matplotlib, & Plotly)** for cleaning, analyzing, and visualizing data because of its ease of analyzing large datasets.

E-Certificate

Code Python

Phases of Project Work

1

Ask

Define problems and objectives of the business by asking questions.

2

Prepare

Determines what data is needed, how to get the data, collect and store the data, ensure the data is credible and unbiased, and know the limitations of the data.

3

Process

Ensure that the dataset used is clean and relevant to the problem to be solved in order to avoid poor insights during data analysis.

4

Analyze

Analyze organized datasets to find insights that address business problems and tasks.

5

Share

Share findings from analysis through data visualization and communicate through presentations or dashboards.

6

Act

Deliver recommendations from the insights gained to reach business goals or business decision-making.

Previous **Next**

Outline (Click)

Introduction

Business Understanding

About Data

Data Cleaning & Manipulation

Summary of The Analysis

Visualization & Key Finding

Top 3 Recommendations

Previous

Next

Introduction

Cyclistic Bike Share

Cyclistic is a bike-sharing company based in Chicago, United States were in 2016, they launched a program offering short-term bike rentals which has grown to a fleet of **5,824 bikes** that are geotracked and locked into a network of **692 stations** across the city of Chicago. The bikes can be unlocked from one station and returned to any other station in the system anytime. **Cyclistic users** are more likely to **ride** for **leisure**, but about **30%** use them to commute to **work** each day. Cyclistic has flexible pricing plans: single-ride passes, full-day passes, and annual memberships. Customers who purchase **single-ride** or **full-day passes** are referred to as **casual riders**. Customers who purchase **annual memberships** are **Cyclistic members (annual members)**.

Scenario

Cyclistic's financial analysts have concluded that **annual members** are much **more profitable** than **casual riders**. Although the pricing flexibility helps Cyclistic attract more customers than usual, **Lily Moreno (the marketing director)** believes that expanding the number of annual members will be critical to future growth. Rather than creating a marketing campaign targeting all-new customers, Moreno believes there is a very good chance to convert casual riders into members. **Lily Moreno** wants **the marketing analysis team** to **design a new marketing strategy to convert casual riders into annual members**. Obviously, **the marketing program recommendations** from the marketing analytics team must be **approved** by the **Cyclistic executive team**, so the recommendations must be supported with insights and visualizations that are easily understood.

Business Understanding

Problem

Lily Moreno wants the marketing analysis team to **design a new effective marketing strategy to convert casual riders into annual members** based on questions that can guide marketing programs and assigned me to answer the question **"how do annual members (Cyclistic members) and casual riders use Cyclistic bikes differently?"**.

Objective

Design a **new marketing strategies** to **convert casual riders into annual members** based on **data-driven insights.**

Task

Analyze a dataset of **historical bike trips** over the past 12 months **(October 2021 - September 2022)** to identify **how annual members and casual riders use Cyclistic bikes differently** and deliver **recommendations** for **new marketing strategies**.

Deliverables

1. A **description** of all **data** sources used.
2. **Documentation** of any **cleaning or manipulation of data**.
3. A **summary** of the **analysis**.
4. Supporting **visualisations** and **key findings**.
5. Top three to four **recommendations** based on the analysis.

Previous | Back to Outline | Next

About Data

Description of Data

Datasets used are public data obtained from **Motivate International Inc.** under this **license** and can be found **here**. The dataset used is **structured data** composed of **rows (records)** and **columns (fields)** with **CSV (Comma-Separated Values)** file format. **Each record** represents **one Cyclistic bike trip** and **each CSV** file contains **one month** of **historical Cyclistic user trip dataset**. Since this project requires datasets for the **past 12 months** from **October 2021 to September 2022**, **12 CSV** files will be used and stored into a **database** in **PostgreSQL**.

Limitations of Data

This dataset does not contain:

1. **Personally identifiable information**, such as names or email addresses of Cylistic users due to data privacy issues.
2. **Information** on **Cyclistic ticket fares**.
3. **Information** on the usage of **reclining bikes, hand tricycles,** and **cargo bikes**.
4. **Cyclistic user trip information** beyond the time of **October 2021 to September 2022**.

Credibility of Data

- **Reliable** because the data contains **accurate** and **unbiased information** based on **actual transactions** of each Cyclistic user's trip.
- **Original** because the data is collected from an **official source** that owns the **copyright** to the data.
- **Comprehensive** because this data contains **relevant information** to answer the questions in this project**.**
- **Current** as the data is **updated monthly.**
- **Cited** because this data is **public data** sourced from **official websites** provided by **Cyclistic Bike-Share** and **the city of Chicago government**.

The credibility of the data sources can be determined using the **ROCCC (Reliable, Original, Comprehensive, Current, Cited) system.*

Previous | Back to Outline | Next

About Data

Variables and Descriptions of Datasets

1.	ride_id	: unique id for each ride
2.	rideable_type	: type of bike used
3.	started_at	: start date and time of the ride
4.	ended_at	: end date and time of the ride
5.	start_station_name	: station name of the ride started
6.	start_station_id	: station id of the ride started
7.	end_station_name	: station name of the ride ended
8.	end_station_id	: station id of the ride ended
9.	start_lat	: station latitude when ride started
10.	start_lot	: station longitude when ride started
11.	end_lat	: station latitude when ride ended
12.	end_lot	: station longitude when ride ended
13.	member_casual	: user types

Data Cleaning & Manipulation

Documentation

1. **Merging 12 CSV file datasets** into 1 dataset with **Union All** using **PostgreSQL**. After this, **data cleaning** was done with **Python**.
2. **Removing end_station_name, end_station_id, end_lat,** and **end_lot** variables that are **not relevant for analysis**.
3. The **missing values** in the **start_station_name** and **start_station_id** variables were **not deleted** because it was necessary to discuss with the director and marketing analytics team where the stations that Cyclistic users use to pick up bicycles could be **recommended locations** for marketing campaigns.
4. There are **no duplicate values**.
5. **Creating a new variable, ride_duration,** which shows the **duration** of the user's trip in **minutes** using the .insert() function, by **calculating** the **time difference** between the **ended_at** and **started_at** variables.
6. **Outliers** in the **ride_duration** variable for **casual riders** are **numerous**. There is **no information** on the **limit** of the **maximum duration** of bike usage in the Cyclistic program, so the **median** value will be used instead of the **mean** value to determine the **trip duration** of users. That's because the **mean** value is strongly **influenced** by **outliers** while the **median** value is **not**.
7. **Filtering ride_duration** variables with values **below 1 minute** and **negative values (less than 0 minutes)** will not be used because it is impossible for the trip duration to be less than 1 minute or negative.
8. **Creating new variables**, namely **ride_hour, ride_day, ride_month, ride_year,** and **ride_season** which are added by **extracting** the **time** value from **started_at**. These variables were created to make the data easier to understand and as a **metric** to find out **the times that riders often start their trips**.

Previous | Back to Outline | Next

Summary Of The Analysis

Total Trips
Annual Members : **3.36 mio**
Casual Riders : 2.36 mio

Median Trip Duration
Annual Members : 9.1 min
Casual Riders : **14 min**

Busiest Season
Annual Members : **Summer**
Casual Riders : **Summer**

Longest Seasonal Trip Duration
Annual Members : **10 min in Summer**
Casual Riders : **15 min in Spring**

Busiest Month
Annual Members : **August 2022**
Casual Riders : **July 2022**

Longest Monthly Trip Duration
Annual Members : **10 min in June 2022**
Casual Riders : **16 min in May 2022**

Busiest Days of the Week
Annual Members : **Tuesday (Weekday)**
Casual Riders : **Saturday (Weekend)**

Another Insight Based on Daily Trip Duration in a Week
Trip duration by **casual riders** is **longer** than by **annual members** on **every day,** but **trip duration** by **annual members** tends to be **consistent** during **weekdays.**

Busiest Hour
Annual Members : **5 pm**
Casual Riders : **5 pm**

Most Used Bike Types
Annual Members : **Classic Bike**
Casual Riders : **Electric Bike**

Busiest Station to Pick Up Bikes
Annual Members : **Kingsbury St & Kinzie St**
Casual Riders : **Streeter Dr & Grand Ave**

Previous | Back to Outline | Next

Summary Of The Analysis

Based on the analysis results, it is clear how casual riders and annual members use Cyclistic bikes differently

Casual Riders

Casual riders tend to use **Cyclistic** during **summer** (especially in **August 2022**) and on **weekends,** also **they** have **longer trip duration** than **annual members**.
These factors suggest that **casual riders** use **Cyclistic** for **leisure.**

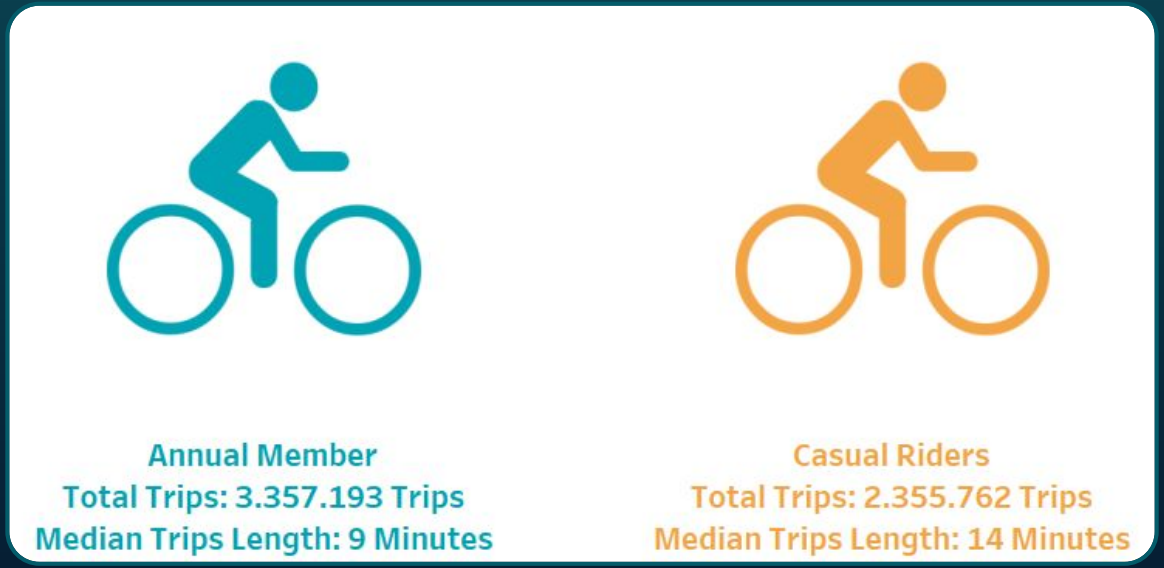
Annual Members

Annual members tend to use **Cyclistic's bike-sharing system** during **summer** (especially in **July 2022**) and on **weekdays**, also **their trip duration** is more **consistent** during **weekdays**. However, **their trip duration** is **shorter** than **casual riders**.
Based on these factors, it shows that **they** use **Cyclistic** for **commuting to work/school** or **running errands** on **weekends**.

Supporting Visualizations & Key Findings

Total Trips & Median Trip Length

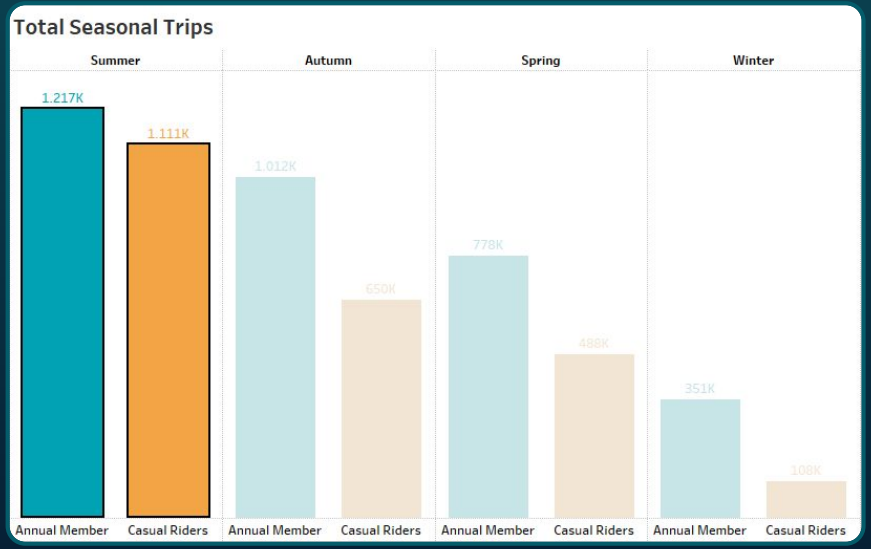
Annual members recorded **more trips** using Cyclistic than **casual riders**. However, it is the **casual riders** who spend **more time** using Cyclistic than the **annual members**.

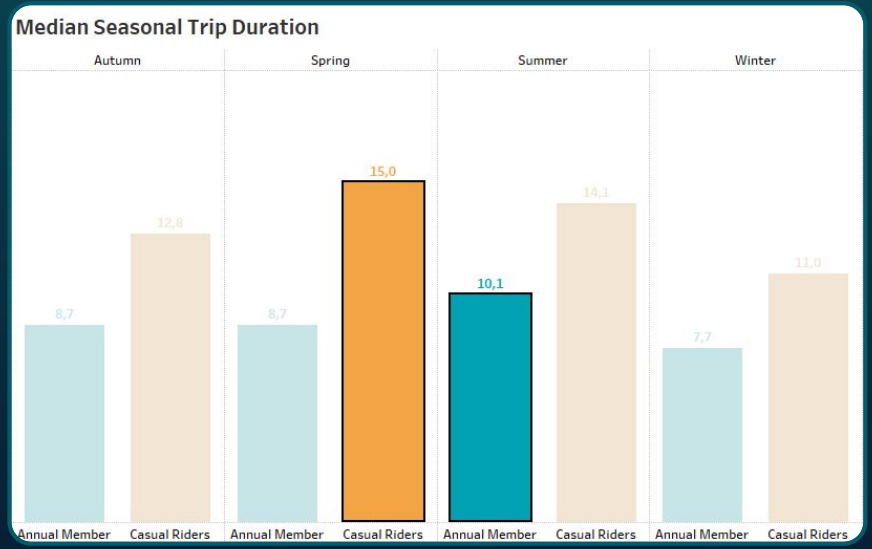


Supporting Visualizations & Key Findings

Total Trips & Median Trip Duration Seasonal

The **busiest season** for **both types of users** to trip using Cyclistic is in **summer**. **Annual members** spend **more time** cycling in the **summer** and **casual riders** in the **spring**.



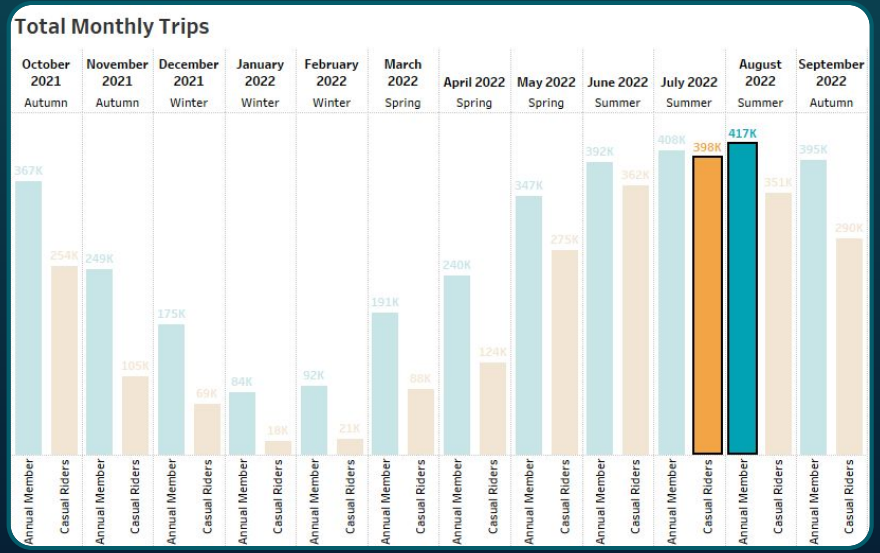


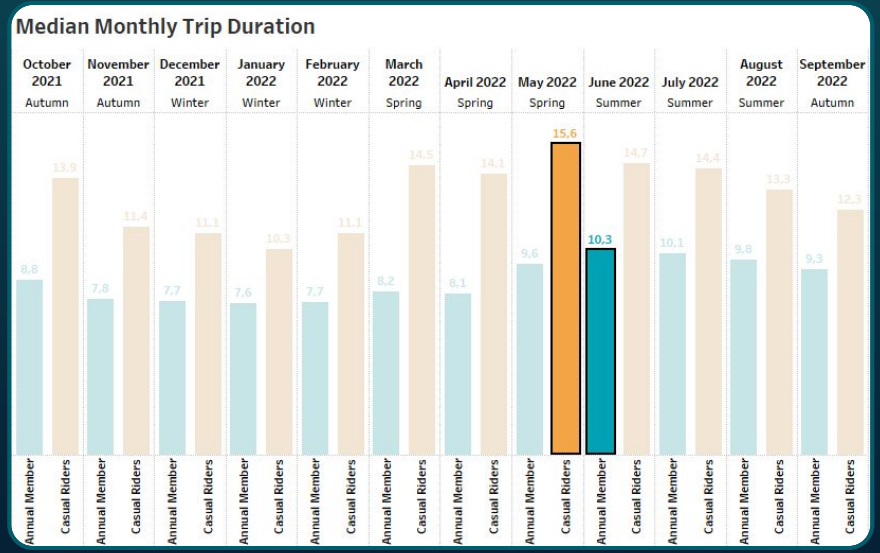
Previous | Back to Outline | Next

Supporting Visualizations & Key Findings

Total Trips & Median Trip Duration Monthly

The **busiest month** for Cyclistic usage by **annual members** was in **August 2022** while for **casual riders** it was in **July 2022**. But, **annual members** spent **more time** on bike trips in **June 2022** while **casual riders** in **May 2022**.



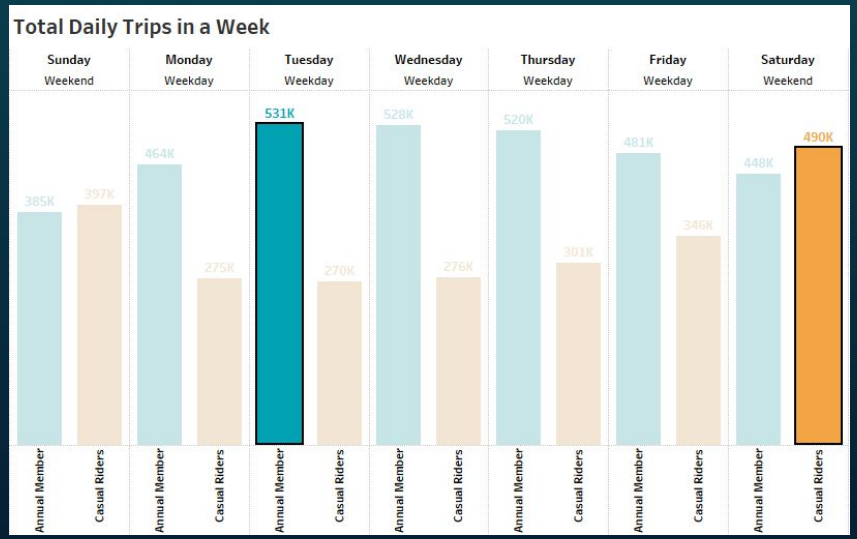


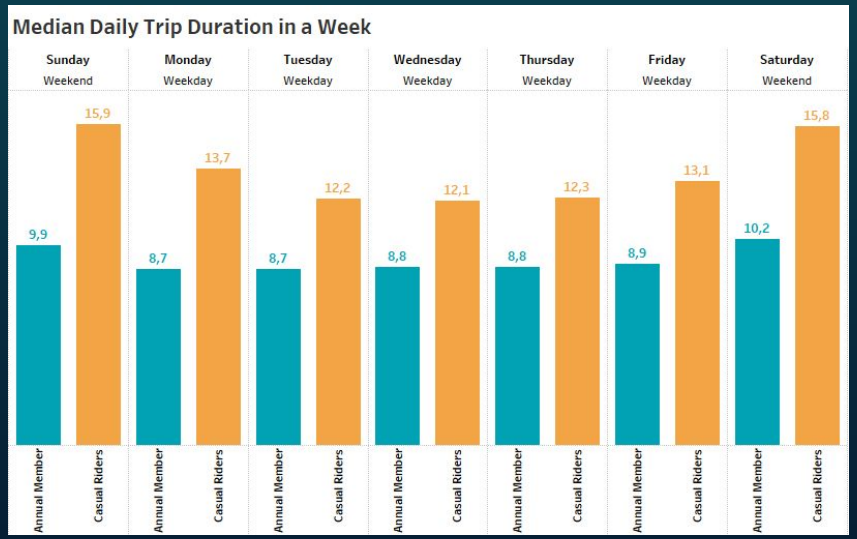
Previous | Back to Outline | Next

Supporting Visualizations & Key Findings

Total Trips & Median Trip Duration Daily in a Week

Saturday (weekend) is the **busiest day** of Cyclistic usage by **casual riders**, while **Tuesday (weekday)** is the **busiest day** for **annual members**. **Trip duration** by **casual riders** is **longer** than by **annual members** on **every day,** but **trip duration** by **annual members** tends to be **consistent** during **weekdays.** It indicates that **casual riders** use Cyclistic for **leisure** while **annual members** use Cyclistic for **work or school.**



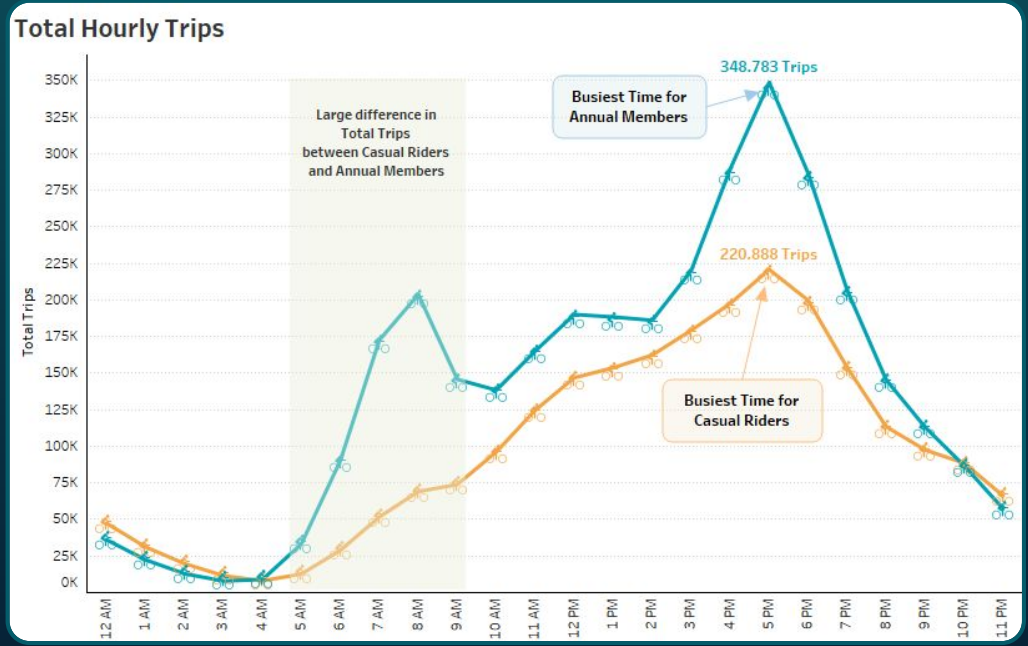


Previous Back to Outline Next

Supporting Visualizations & Key Findings

Total Trips & Median Trip Duration Daily in a Week

5 pm is the **busiest time** for **both users** to start a trip using Cyclistic, which may be them returning from **work or leisure** using Cyclistic. **Annual members** were recorded as using Cyclistic more from **4 am to 9 pm**, while **10 am to 3 am** were **more trips** made by **casual riders** and perhaps they used Cyclistic to **leisure** during night hours in Chicago. The **large difference** in **total trips** between **annual members** and **casual riders** at **5 am to 9 am** with **more trips** made by **annual members** suggests that **annual members** are commuting to **work or school** at these hours using Cyclistic.

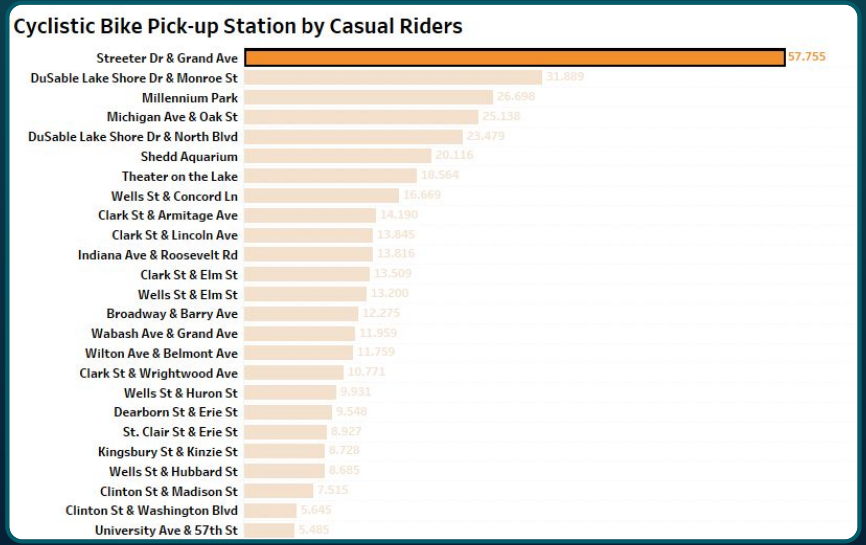


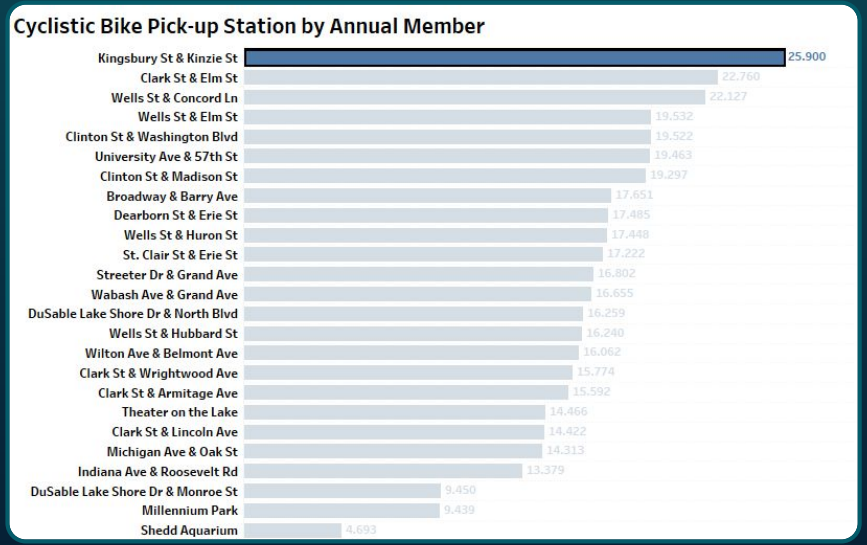
Previous | Back to Outline | Next

Supporting Visualizations & Key Findings

Cyclistic Bike Pick-up Station

Streeter Dr & Grand Ave are the **busiest stations** for **casual riders** to pick up Cyclistic bikes while **annual members** prefer **Kingsbury St & Kinzie St** stations to pick up Cyclistic bikes.



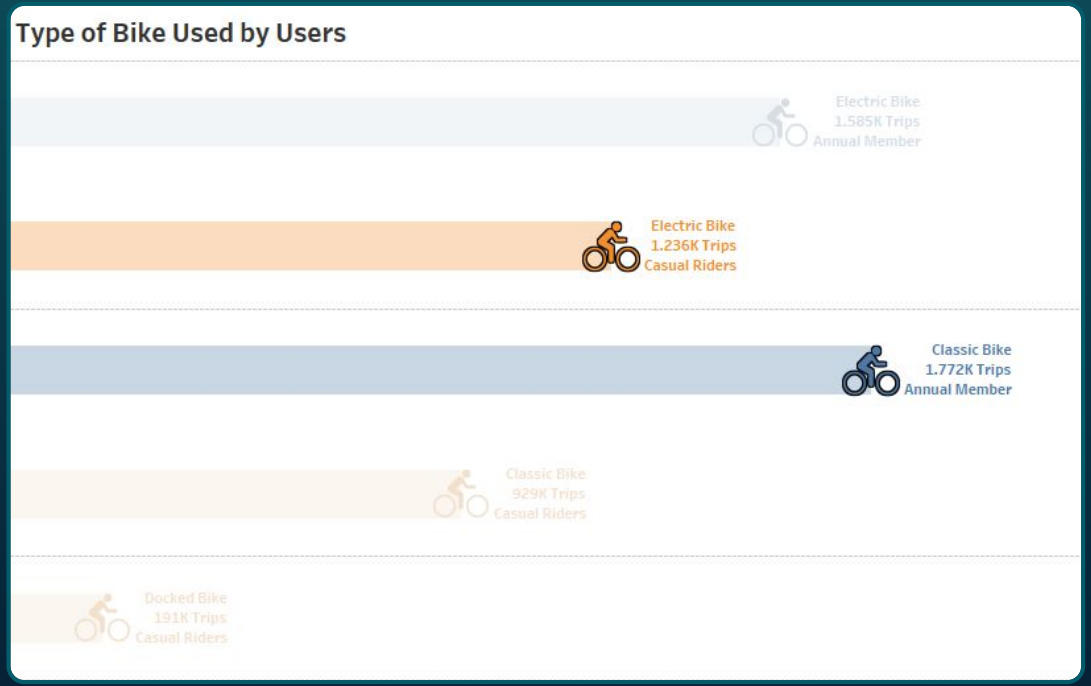


Previous | Back to Outline | Next

Supporting Visualizations & Key Findings

Type of Bike Used by Users

Casual riders prefer to use **electric bike** while **annual members** prefer to use classic bike for their trips. **Docked bike** are only used by **casual riders** and this needs to be discussed during team meetings.



Previous Back to Outline Next

Top 3 Recommendations

for Converting Casual Riders to Annual Members Based on Analysis Insights

1. **Promotional Campaigns for Annual Membership Sign-Ups are Only During The Summer Period or July.** Promotions may include **discounts** for **annual membership sign-ups** or **additional benefits** specific to being an **annual member** such as **priority access** to a desired **bike** or **free Cyclistic Bike Share exclusive merchandise** such as bags. The promotion aims to **attract** and **encourage casual riders** to become **annual members** based on the insight that **casual riders** make **more trips** during **summer** and **July**.

2. **Third-party service discounts for annual members**. Cyclistic Bike Share can negotiate discounts with third party services such as **entertainment venues or local restaurants** and can be campaigned on **weekends** or **at 5 pm** as **casual riders** are more likely to make Cyclistic trips at these times. The campaign could be carried out through advertisements at **stations** that are **most frequently used** by **casual riders** to pick up Cyclistic bikes. This discount could encourage **casual riders** to **sign up** for **Cyclistic membership** and make **annual members** more satisfied with Cyclistic's services, thus also potentially **retaining them**.

3. **Develop a point reward system for annual members. Points** can be **awarded** to **annual members** who **use Cyclistic frequently**, **spend a lot of time** with **Cyclistic**, **share stories** about their **Cyclistic trips** or the **benefits** of being an **annual member** on **social media**, and successfully **invite other people** or **casual riders** to **sign up** for **Cyclistic membership** through the concept of **referral marketing**. This system can be implemented through **Cyclistic's mobile app** to make it easier to track points accumulated by **annual members** and can be redeemed directly for **membership renewal discounts** or **prizes** such as **exclusive Cyclistic Bike Share merchandise**. This can be a **strong incentive** for **annual members** to **continuously use Cyclistic** and **invite others** to **sign up** for **Cyclistic membership**.

Thank You!

oogle Data Analytics

Capstone Project

About Me

Back to Project