Cyclistic Bike Share Data Analysis Project

Data Analyst: Mahmoud Al-Fouly
July 12th 2024

Introduction

- Objective: Analyze Cyclistic bike share data to identify trends and design marketing strategies to convert casual riders into subscribers.
- Importance: Understanding usage patterns can help Cyclistic increase revenue and customer loyalty.
- Project steps:
    - Data cleaning.
    - Exploratory data analysis (EDA).
    - More specific analysis based on user type.

Dataset Description

- Overview: The dataset includes information about bike rides taken by Cyclistic customers and subscribers.
- Key Columns: 'trip_id', 'bike_id', 'from_station_name', 'to_station_name', 'usertype', 'gender', 'birthyear'.

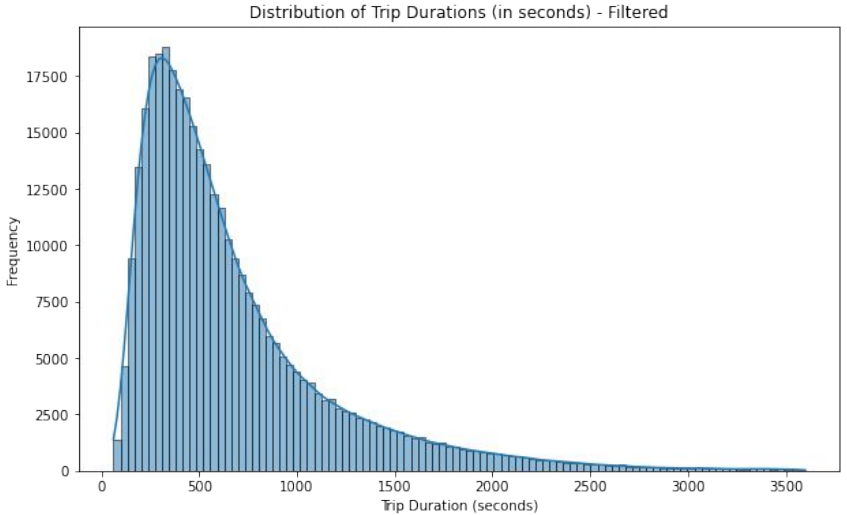
Data Cleaning and Preparation

- Key Steps:
  - Handled missing values.
  - Checked for duplicates.
  - Converted data types.

Exploratory Data Analysis

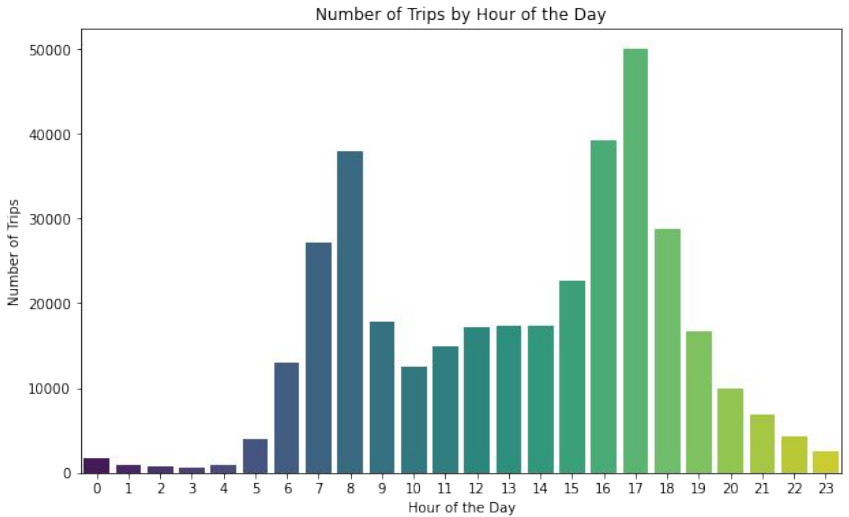
Trip Duration Distribution

- Histogram showing the distribution of trip duration.
- Insight: Most trips are relatively short, with a significant peak around a few hundred seconds.



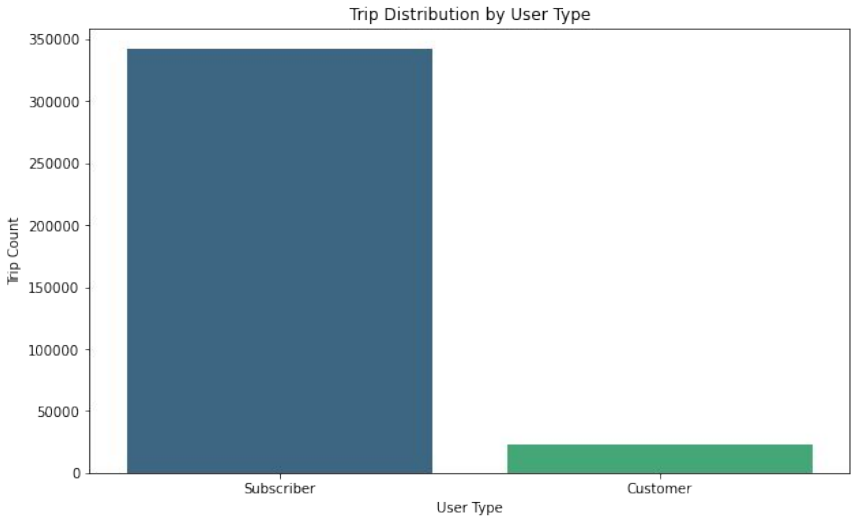
Number of Trips by Hour (Time) of the Day

- Bar plot showing the distribution of trips by the time of the day (hour) they start in.
- Insight: Peak times are in accordance with commuting times (around nine a.m and five p.m).



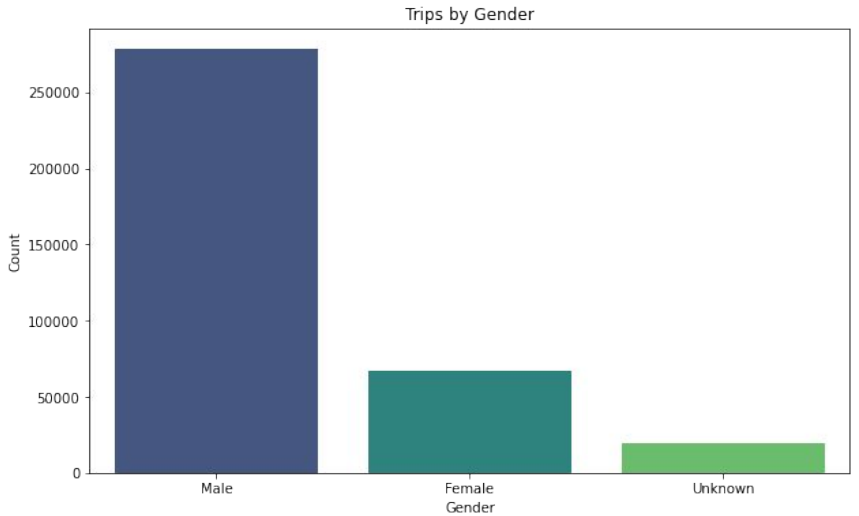
Trip Distribution by User Type

- Bar plot showing the distribution of the number of trips by user type.
- Insight: Most trips are taken by subscribers with over 300,000 trips. Customers take a minority of trips.

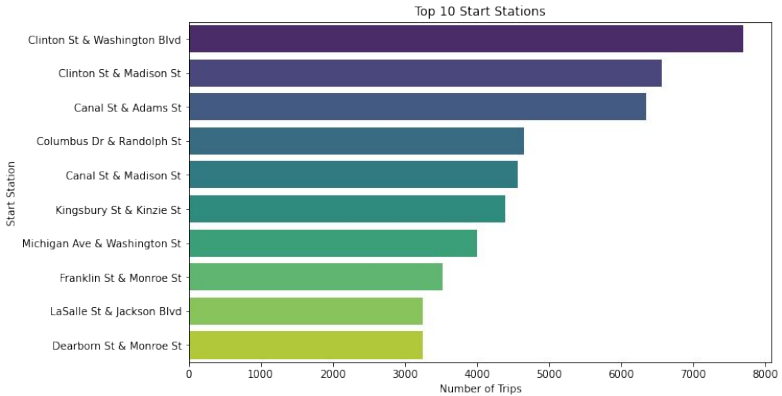


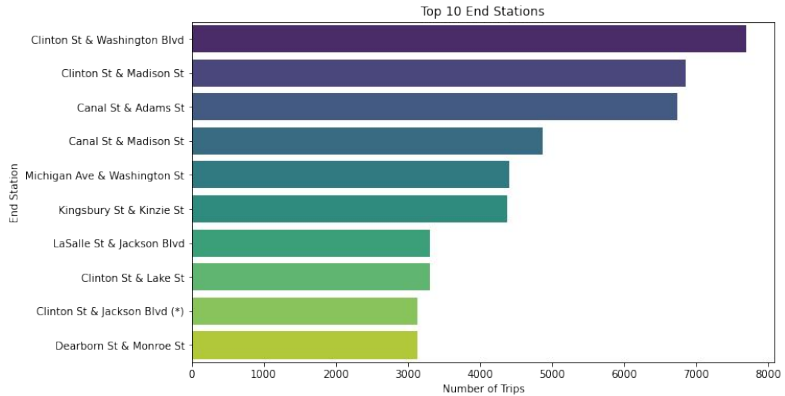
Trips by Gender

- Bar plot showing the distribution of trips by gender.
- Insight: Most trips are taken by male users. Female trips are much less. There's a significant number of trips that lack 'gender' data.



Top 10 Start & End Stations





More Specific Analysis Based On User Type

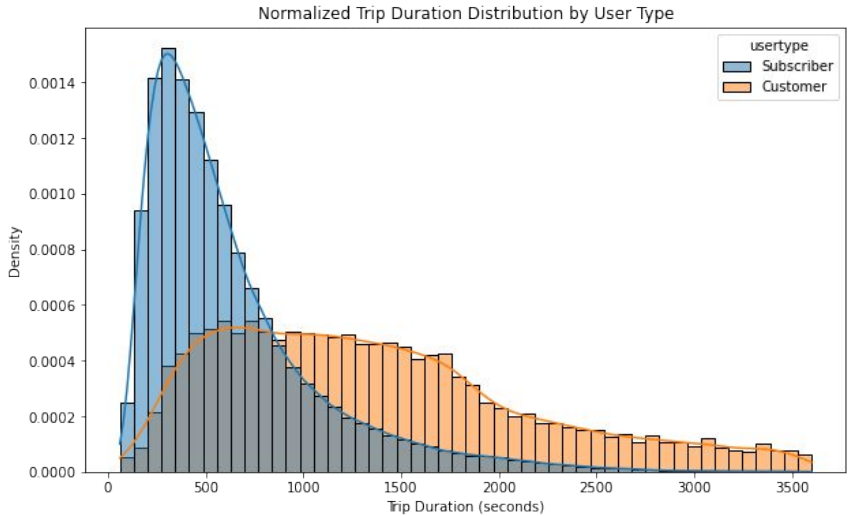
Normalized Trip Duration Distribution by User Type

Description:

- Histogram showing the distribution of trip duration by user type.

Insight:

1. Customers take longer duration trips than Subscribers.
2. Mean trip duration for Subscribers is 633.7 seconds and for customers is 1374.4 seconds.



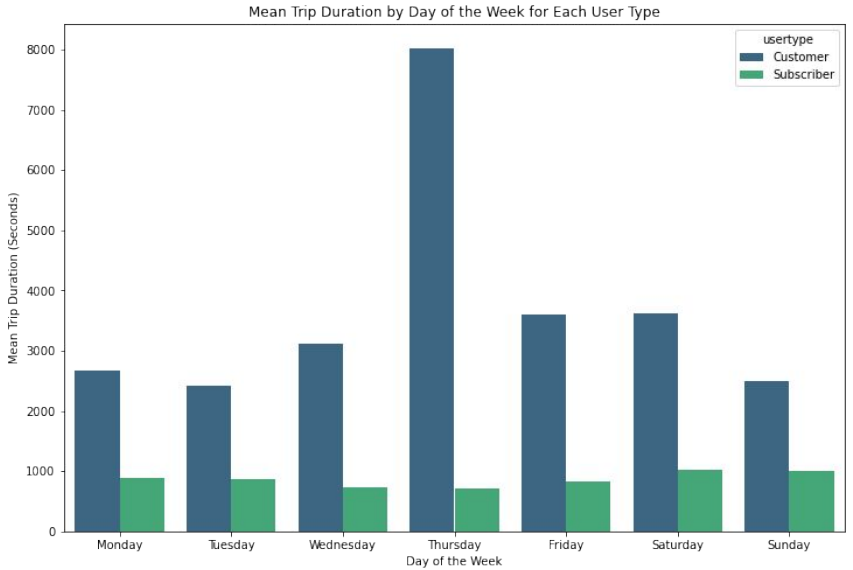
Mean Trip Duration by Day of the Week & User Type

Description:

- Bar chart showing mean trip duration in each day of the week for each user types.

Insight:

- For subscribers, the average trip duration is stable in each day of the week. For customers, the day with the lengthiest trips is Thursday.



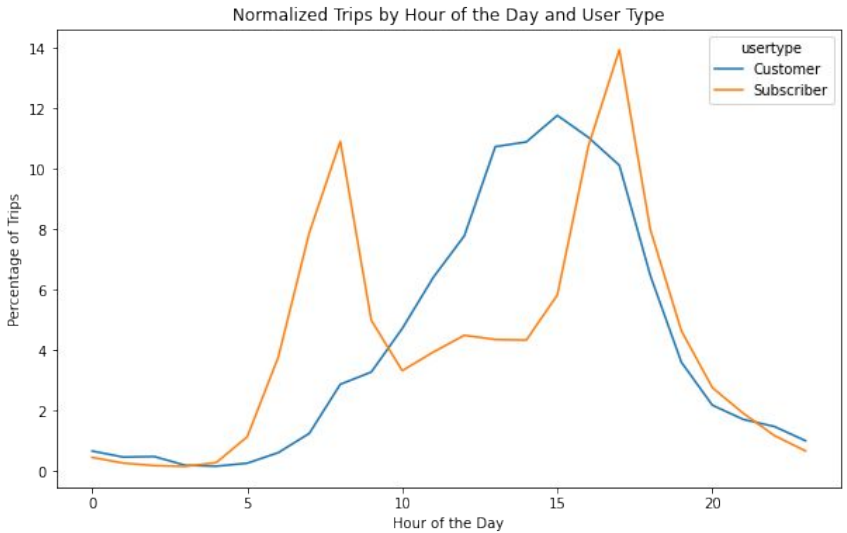
Normalized Trips by Hour of the Day & User Type

Description:

- Line plot showing the differences in trips by hour of the day between the two user types.

Insights:

1. Subscribers are most likely to take trips at the times of going to and leaving work (nine and five).
2. Customers trips are usually between 10 am and 7 pm.



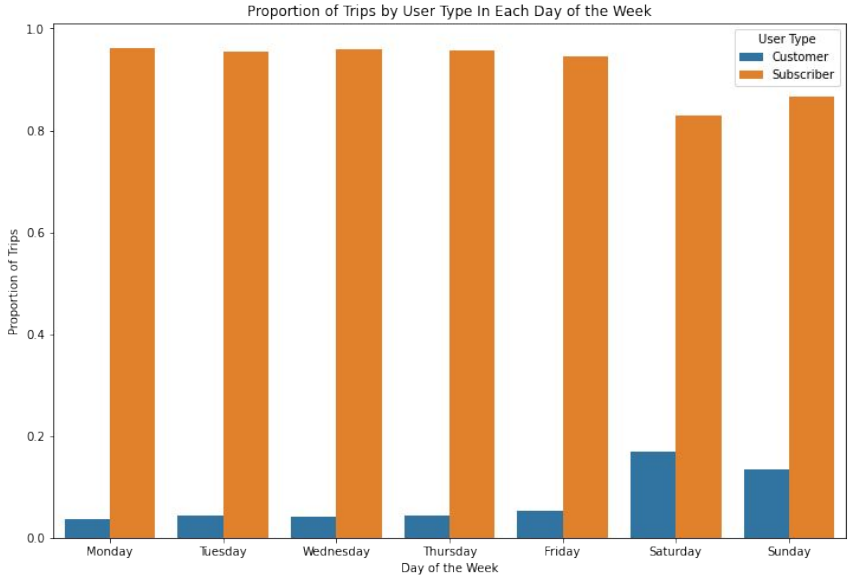
Proportion of Trips by User Type & Day of the Week

Description:

- Bar plot showing the proportion of trips taken by each user type (subscribers and customers) for each day of the week.

Insights:

1. Subscriber trip percentages are stable on weekdays and decrease on weekends.
2. Customer trip percentages peak on weekends.



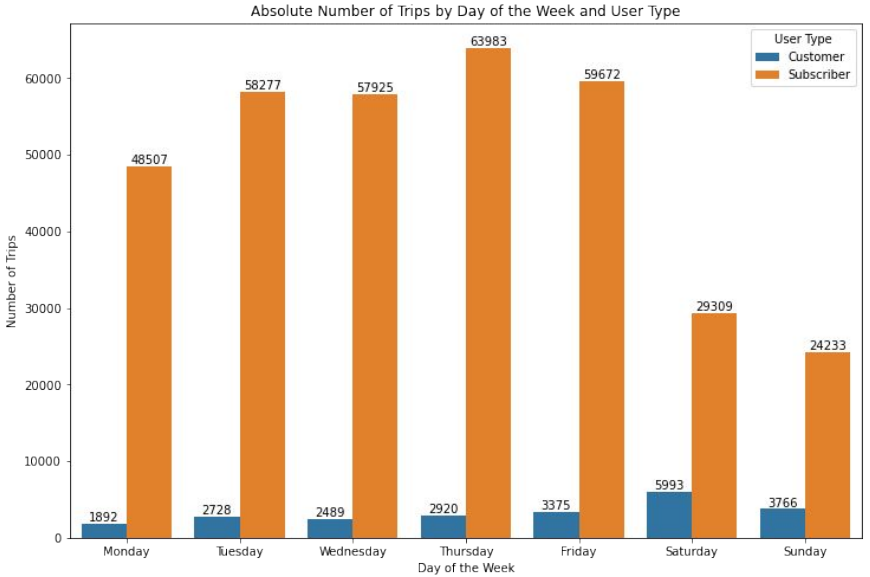
Absolute Number of Trips by User Type & Day of the Week

Description:

- Bar plot showing the absolute number of trips taken by each user type (subscribers and customers) in each day of the week.

Insights:

- The increase in the percentage of customer trips on weekends is due to an increase in the number of trips taken by customers, coupled with a decrease in the number of trips taken by subscribers on these days.



Top Start & End Stations For Both User Types

For subscribers:

- Start stations:
  1. Clinton St & Washington Blvd
  2. Clinton St & Madison St
  3. Canal St & Adams St
- End stations:
  1. Clinton St & Washington Blvd
  2. Clinton St & Madison St
  3. Canal St & Adams St

For customers:

- Start stations:
  1. Streeter Dr & Grand Ave
  2. Lake Shore Dr & Monroe St
  3. Shedd Aquarium
- End stations:
  1. Streeter Dr & Grand Ave
  2. Lake Shore Dr & Monroe St
  3. Millennium Park

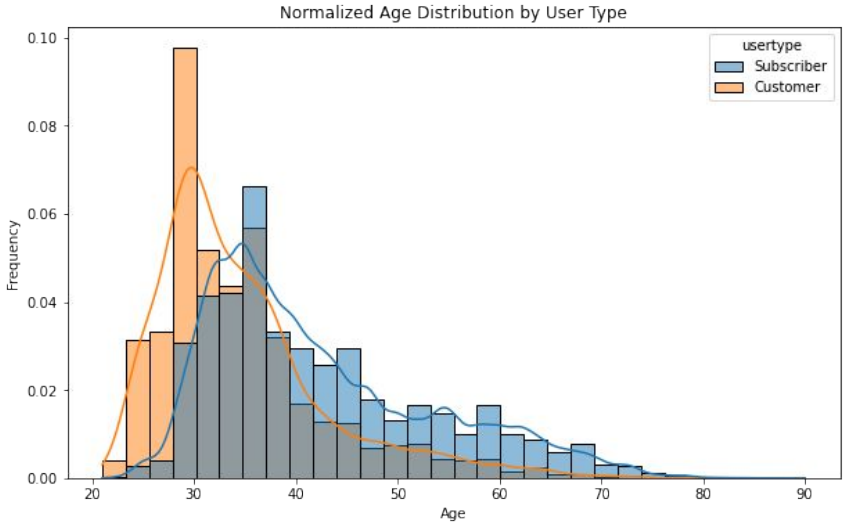
Normalized Age Distribution by User Type

Description:

- Bar chart showing the differences in age distribution between the two user types.

Insight:

1. Younger people are more likely to be customers.
2. Older people are more likely to be subscribers.



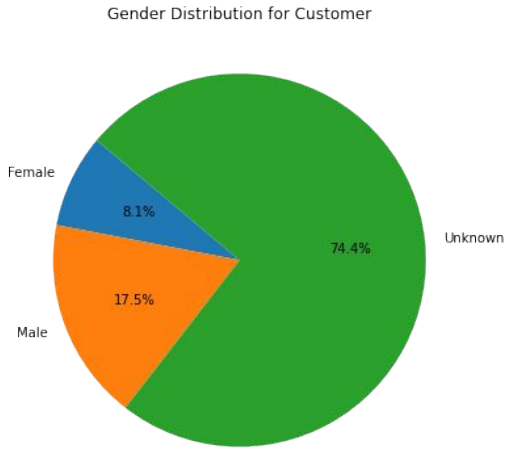
Gender Distribution For Each User Type

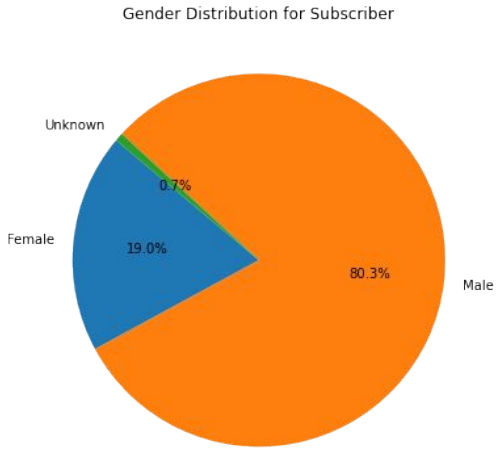
Description:

- Two pie charts showing the distribution of users based on gender in each user type.

Insights:

1. Predominantly Male Subscribers.
2. Low Female Participation in both user types.
3. High Percentage of unknown gender among customers.





Recommendations and Summary

Recommendations to the Marketing Team

1. Weekend Advertising: Focus advertising efforts on weekends (Saturdays and Sundays), as these are the peak days for customer trips.
2. Peak Hour Advertising: Schedule ads around 3 p.m., which is the peak time for customer trips.
3. Target Key Stations: Place ads at top customer start and end stations such as 'Streeter Dr & Grand Ave', 'Lake Shore Dr & Monroe St', and 'Shedd Aquarium'.
4. Highlight Subscription Benefits: Emphasize the financial advantages of subscribing, particularly for users who take longer trips.
5. Youth-Centric Campaigns: Target younger audiences, especially those under thirty, as they form the majority of the customer base.
6. Data Collection Improvement: Implement strategies to collect comprehensive gender and birthyear data from customers to enhance future analyses.

Summary of the Project

- I started the project with loading and cleaning the data.
- During exploratory data analysis, I did some analyses, like: trip duration distribution; trips by hour of the day; and user type distribution.
- In the more specific analysis, I compared between the two user types according to some factors, like: age, gender, trip duration, and most days they take trips in.
- Lastly, I provided the marketing team in Cyclistic with some recommendations based on the analysis.

Contact Me

For business enquiries, contact me here:

- Email: mahmoudalfouly2019@gmail.com
- LinkedIn: Link

Thank You