DI✓Y

Divvy Case Study

By: Tanner Quesenberry

Summary

This project is a case study analysis of <u>Divvy's bike-share program</u> that operates within the city of Chicago. Data based analytical tools and visualizations are used to draw insights and present conclusions to better inform stakeholders to make business related decisions. The following sections detail the critical thinking and problem-solving steps related to the project.

1. Business Task

The task at hand is to gain more insight into the ridership habits of customers and discover differences in how members and non-members use the bike share program. These insights will be used to drive company decision making related to transitioning from developing brand awareness/general appeal to actively converting casual riders to annual members. This will drive future growth as members are deemed more profitable than casual riders.

2. Data Description

The data for this analysis came from Divvy collected user data available through the following link: https://divvy-tripdata.s3.amazonaws.com/index.html. It was made available by Motivate International Inc. under this license. This case study covers a year's worth of ridership data from June 2020 to May 2021 which correlates to the 12 csv files from *'202006-divvy-tripdata.zip'* to *'202105-divvy-tripdata.zip'*. This data is anonymized and therefore does not contain any personally identifiable information.

3. Data Cleaning and Analysis Setup

Cleaning of the datasets for this analysis was done using R and RStudio due to the large sizes of each file. After the 12 csv had been downloaded, they were put into a project folder. Each file was read into a data frame appropriately named after the month/year using the read_csv() function. Each set of data contains the same 13 data points per observation for the applicable month. Prior to merging the 12 data frames into one, type casting was done for two columns on half the data frames to account for a change in how that data is recorded.

Once merged into one data frame, duplicate records were removed, unnecessary coordinated data was discarded, and columns were verified to only contain the correct inputs. Finally, new columns were created that parsed the start date/time, day of week and total length of the ride. Rows with ride lengths that were negative were discarded from the data frame. The cleaned data was saved to a new csv file for future use/backup.

- For a deeper look into the code, please refer to the R markdown file here.

4. Analyzing and Understanding the Data

Let's start the analysis by taking a look at some descriptive statistics in regards to ride lengths among all users.

```
Ride length (seconds)
Min:   1
Max:   3257001
Median:   843
Mean:   1617.189
```

Converting these measurements into minutes, the average length of a ride is about 27 minutes with a median ride time of 14 minutes. Now let's take a closer look at these statistics when broken down between members and casual riders.

```
Mean
    Rider  length_secs
1 casual     2562.0662
2 member      930.4542

Median
    Rider  length_secs
1 casual           1213
2 member            664

Max
    Rider  length_secs
1 casual       3257001
2 member       2476260

Min
    Rider  length_secs
1 casual             1
2 member             1
```

When categorized by rider type, we determine that casual riders tend to have far longer ride durations than members on average. This difference is quite large given times of 43 minutes to 16 minutes. One possible hypothesis for this is that members could be short distance commuters to work while casual riders are visitors to the city.

Analyze ridership by type and weekday

```
## # A tibble: 14 x 4
## # Groups:   member_casual [2]
##    member_casual weekday number_of_rides average_duration
##    <chr>         <ord>             <int>            <dbl>
##  1 casual        Sun              329920            2934.
##  2 casual        Mon              188393            2533.
##  3 casual        Tue              174391            2280.
##  4 casual        Wed              182540            2299.
##  5 casual        Thu              191733            2395.
##  6 casual        Fri              246620            2430.
##  7 casual        Sat              396397            2674.
##  8 member        Sun              307785            1048.
##  9 member        Mon              315457             895.
## 10 member        Tue              329375             875.
## 11 member        Wed              347149             889.
## 12 member        Thu              341939             877.
## 13 member        Fri              350544             913.
## 14 member        Sat              360531            1020.
```

Examining this rider difference deeper, we can look for differences and similarities between riders for each day of the week. Both casual riders and members tend to follow the same pattern regarding ride times, with the longest average durations occuring on the weekends. In terms of when rides are most common, members seem to steadily take more rides throughout the week before dropping off on Sundays. Casual riders on the other hand, take far more rides on the weekend compared to the weekdays. For example, there are over twice as many rides on Saturday than on Tuesday. Overall, members take far more rides in total than casual riders.

	start_station_name	station_count
	<chr>	<int>
1	Streeter Dr & Grand Ave	47998
2	Lake Shore Dr & Monroe St	39332
3	Clark St & Elm St	35979
4	Theater on the Lake	35931
5	Lake Shore Dr & North Blvd	32400
6	Millennium Park	30379
7	Wells St & Concord Ln	29947
8	Michigan Ave & Oak St	28521
9	Indiana Ave & Roosevelt Rd	27858
10	Wells St & Elm St	27328

Taking a quick glance at starting location, Streeter Dr & Grand Ave is the most popular spot to rent a bike. The rest of the list shows the top 10 locations for all riders. Let's break it down by rider type next.

```
      member_casual  start_station_name           station_count
       <chr>          <chr>                               <int>
 1     casual         Streeter Dr & Grand Ave            36557
 2     casual         Lake Shore Dr & Monroe St          28231
 3     casual         Millennium Park                    24804
 4     casual         Theater on the Lake                18561
 5     casual         Michigan Ave & Oak St              18361
 6     casual         Lake Shore Dr & North Blvd         16867
 7     casual         Indiana Ave & Roosevelt Rd         15883
 8     casual         Michigan Ave & Lake St             13926
 9     casual         Shedd Aquarium                     13868
10     casual         Clark St & Elm St                  13348
11     member         Clark St & Elm St                  22631
12     member         Wells St & Concord Ln              17679
13     member         Theater on the Lake                17370
14     member         Broadway & Barry Ave               17307
15     member         Dearborn St & Erie St              17186
16     member         Kingsbury St & Kinzie St           17083
17     member         St. Clair St & Erie St             16771
18     member         Wells St & Elm St                  16521
19     member         Wells St & Huron St                16113
20     member         Lake Shore Dr & North Blvd         15533
```

When broken down by rider type, we find some interesting stats. For members, Clark St & Elm St is the most popular starting station but not by much. Most of the members' station counts are very evenly distributed. Casual riders however skew heavily towards their top 3 station locations. These locations all correspond to tourist type areas in the city of Chicago.
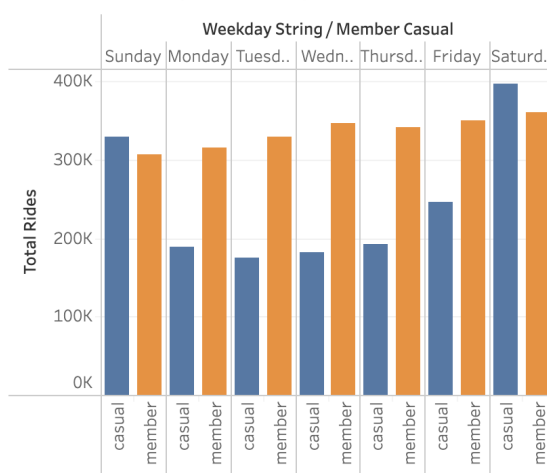
5. Visuals

Visualizations were made using Tableau (interactable here) and R.
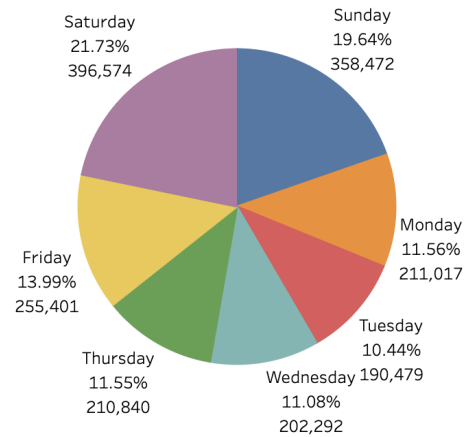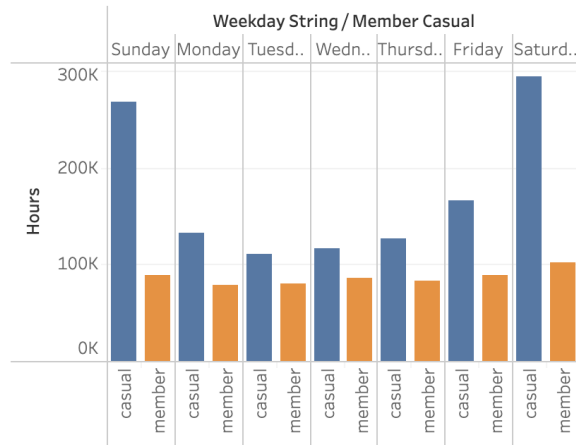

Average Ride Time (Weekday)
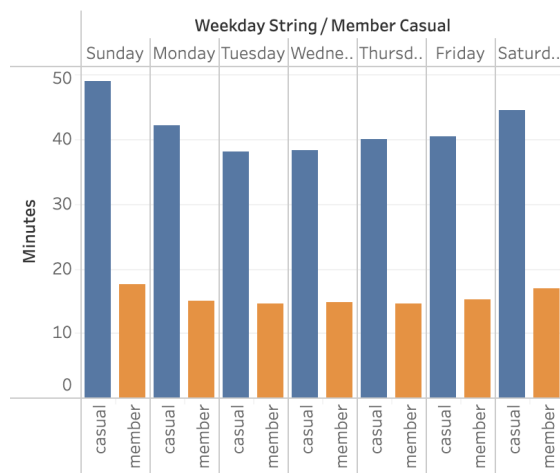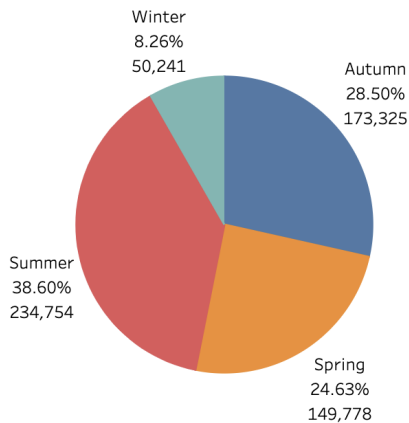

Total Rides By Weekday

These graphs really help visualize the differences between longer durations of casual riders vs the frequency at which members use the bikes. We also notice the stark difference in that casual riders rent bikes for weekend use while member usage is evenly distributed.

Total Weekday Ride Hours by Membership

Total Ride Hours Percentage (Weekday)

Weekday String / Member Casual

Sunday | Monday | Tuesd.. | Wedn.. | Thursd.. | Friday | Saturd..

Hours: 300K, 200K, 100K, 0K

casual / member (for each day)

Saturday 21.73% 396,574
Sunday 19.64% 358,472
Friday 13.99% 255,401
Monday 11.56% 211,017
Thursday 11.55% 210,840
Wednesday 11.08% 202,292
Tuesday 10.44% 190,479

Again, this bar chart shows that casual riders ride for longer durations than members and heavily leans to weekend usage. Total ridership overall seems to be pretty evenly distributed with a slight edge to the weekend.

Total Seasonal Member Ridership

Total Seasonal Casual Ridership

Winter 8.26% 50,241
Autumn 28.50% 173,325
Summer 38.60% 234,754
Spring 24.63% 149,778

Winter 2.43% 29,516
Autumn 21.86% 266,054
Summer 50.76% 617,695
Spring 24.96% 303,712
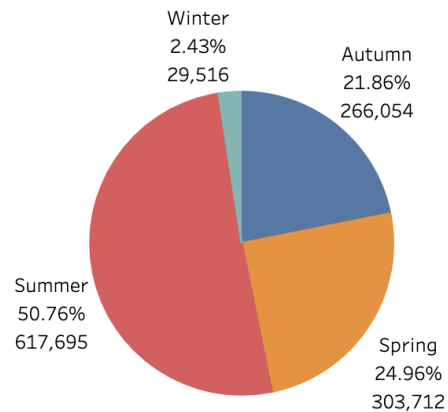
Seasonal differences reveal that over half of the casual ridership occurs during the summer months of June to August while winter rides are very slim. Members tend to ride more during the summer as well but not to the extent of casual riders.

DIVVY
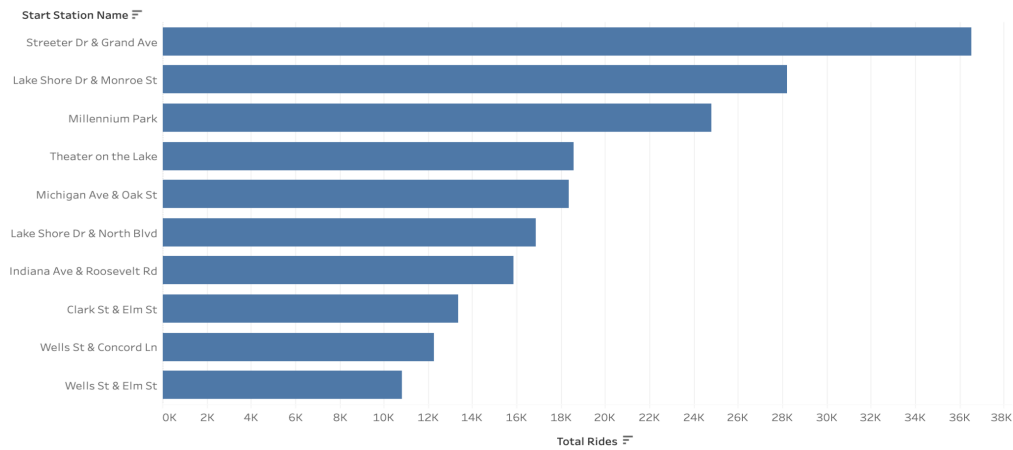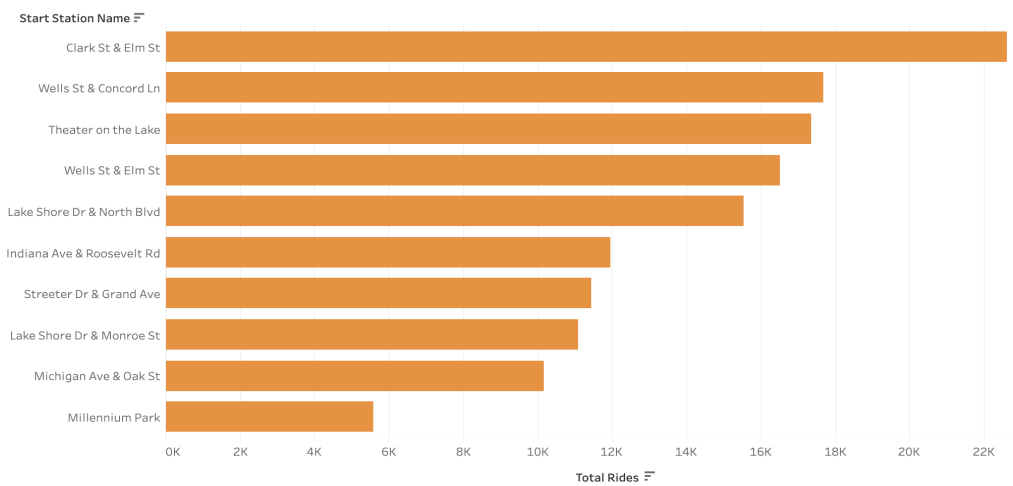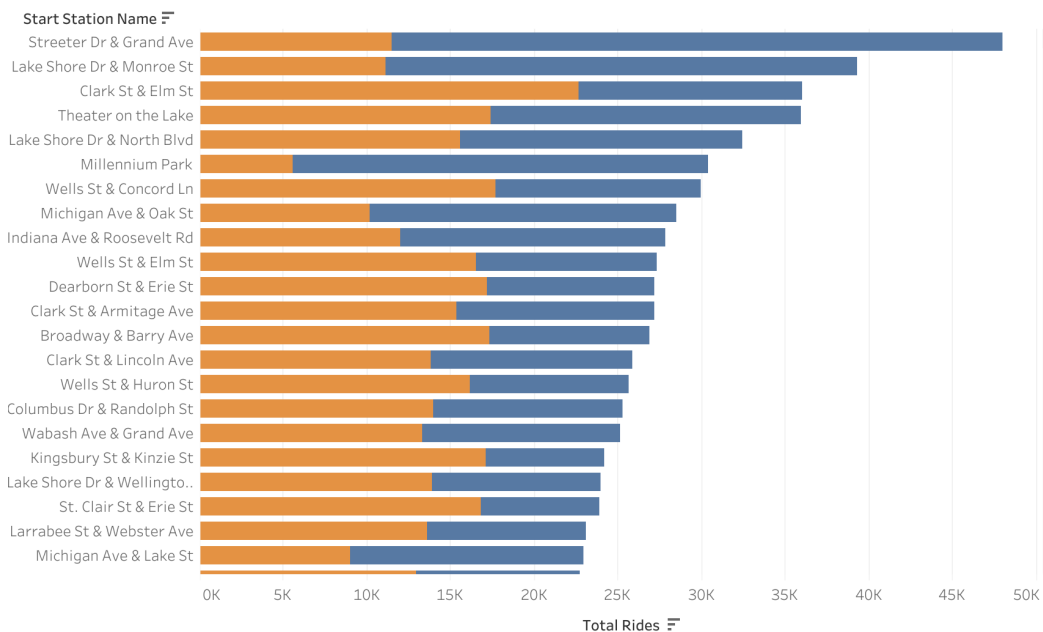
Top 10 Start Locations for Casual Riders

Start Station Name

Station	
Streeter Dr & Grand Ave	
Lake Shore Dr & Monroe St	
Millennium Park	
Theater on the Lake	
Michigan Ave & Oak St	
Lake Shore Dr & North Blvd	
Indiana Ave & Roosevelt Rd	
Clark St & Elm St	
Wells St & Concord Ln	
Wells St & Elm St	

0K 2K 4K 6K 8K 10K 12K 14K 16K 18K 20K 22K 24K 26K 28K 30K 32K 34K 36K 38K
Total Rides

Top 10 Start Locations for Members

Start Station Name

Station	
Clark St & Elm St	
Wells St & Concord Ln	
Theater on the Lake	
Wells St & Elm St	
Lake Shore Dr & North Blvd	
Indiana Ave & Roosevelt Rd	
Streeter Dr & Grand Ave	
Lake Shore Dr & Monroe St	
Michigan Ave & Oak St	
Millennium Park	

0K 2K 4K 6K 8K 10K 12K 14K 16K 18K 20K 22K
Total Rides

Total Starting Locations with Membership Status

Start Station Name

Station	
Streeter Dr & Grand Ave	
Lake Shore Dr & Monroe St	
Clark St & Elm St	
Theater on the Lake	
Lake Shore Dr & North Blvd	
Millennium Park	
Wells St & Concord Ln	
Michigan Ave & Oak St	
Indiana Ave & Roosevelt Rd	
Wells St & Elm St	
Dearborn St & Erie St	
Clark St & Armitage Ave	
Broadway & Barry Ave	
Clark St & Lincoln Ave	
Wells St & Huron St	
Columbus Dr & Randolph St	
Wabash Ave & Grand Ave	
Kingsbury St & Kinzie St	
Lake Shore Dr & Wellingto..	
St. Clair St & Erie St	
Larrabee St & Webster Ave	
Michigan Ave & Lake St	

0K 5K 10K 15K 20K 25K 30K 35K 40K 45K 50K
Total Rides

Finally, let's take a visual look at the starting station preferences among users. As noted in the earlier analysis, casual riders gravitate more towards stations near tourist destinations such as Millenium Park. Member locations are more evenly distributed and are not as clustered around tourist areas.

6. Recommendations

With all data, there are limitations. Since the data used in this analysis was anonymized, determining the relationship between Divvy users who live in Chicago vs. outside city limits is not possible. If it were possible to use this personal information along with things like socioeconomic status, insights could be made on price points, promotions, and geographical recruiting.

Given the data that was available, the following are recommendations for growth:

1. Target promotions to casual riders during the summer months of June to August, as that is when 46% of all rides take place.
2. Casual riders are far more likely to use the service on weekends making it a prime time to advertise to them.
3. Casual riders on average ride for longer durations than members. Offer a promotion that highlights the savings of becoming a member.
4. Place fliers or advertisements at the most popular locations for renting Divvy bicycles such as Millenium Park or Theater on the Lake.