

Case Study Cyclistic: members and casual riders preferences

AUTHOR: HANNA SLIASHYNSKAYA

DATE: 14-JAN-2024

Background of this case study

- I completed this case study as part of my Google Analytics Certificate programme.
- I analysed data of a fictional company called **Cyclistic**, a bike share company.
- I analysed a dataset that contained information about all rides that took place in the period of April to June 2020 (more than half a million lines of data) using SQL to identify trends and preferences of annual members and casual riders.



INTRODUCTION



Business task at hand

- Because the company's future depends on converting casual riders into annual members, marketing team is working on a strategy for converting casual riders into annual members.
- Understand how casual riders and annual members use Cyclistic bikes differently
- Design a new marketing strategy to convert casual riders into annual members



Project plan

- Data transformation and cleaning (steps + cleaning protocol)
- Data analysis (SQL)
- Visualizations (CSV exports to Excel for graphs)
- Recommendations (based on business task)

Data Transformation and Cleaning

1. Tables join (dataset consisted of 3 separate CSV files for each month of data)
2. Search and remove missing values
3. Calculate the length of rides from start and end time
4. Remove ≤ 0 values for length ride as outliers
5. Calculate the day of the week from date of the ride value



Analysis

- Descriptive statistics (number of rides, max and min values for ride length, average and median ride length)
- Calculate average ride length by type of membership
- Identify trends in weekday vs. weekend use of rent bikes by type of membership

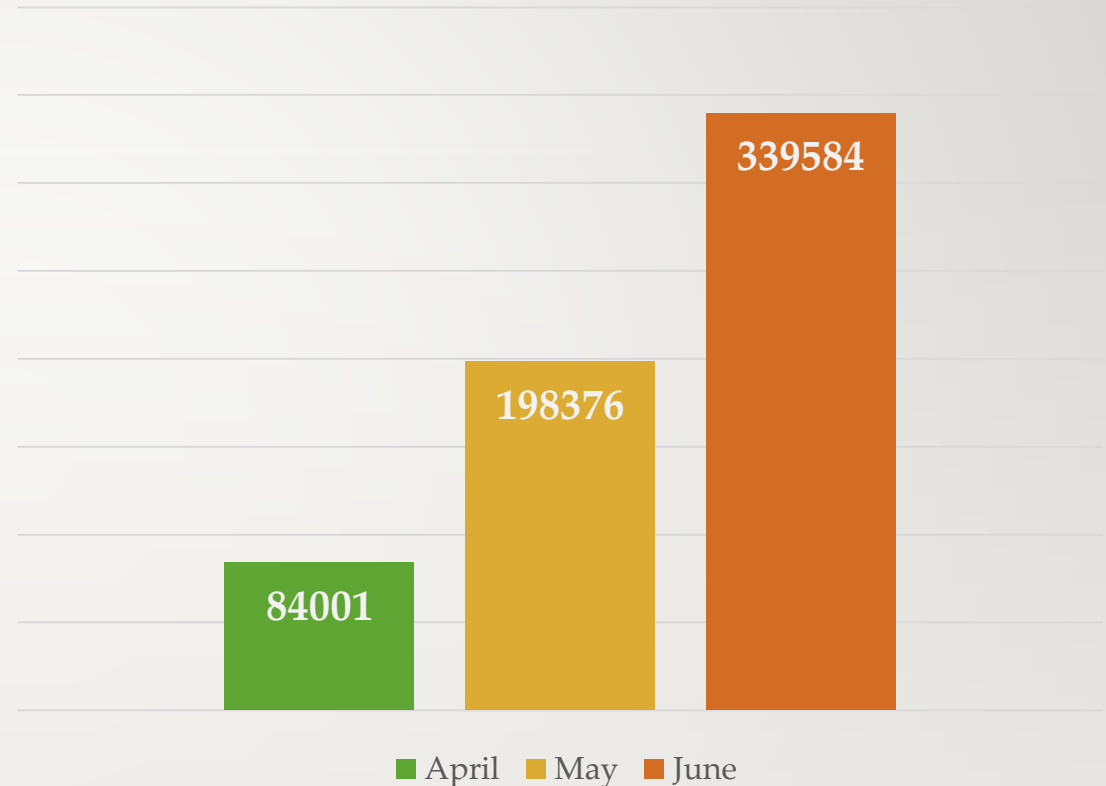


RESULTS

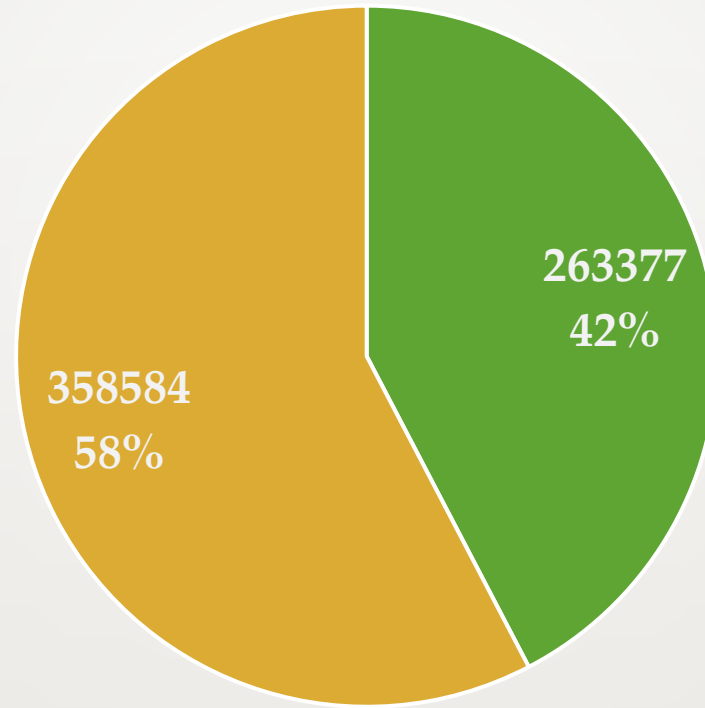
Descriptive statistics

- **621.961 rides** over 3 months
- Across **613 bike stations**
- Min ride length: 1 minute
- Max ride length: 6842 minutes
- Median ride length: **18 minutes**
- Average ride length: **31 minutes**

Number of rides by month



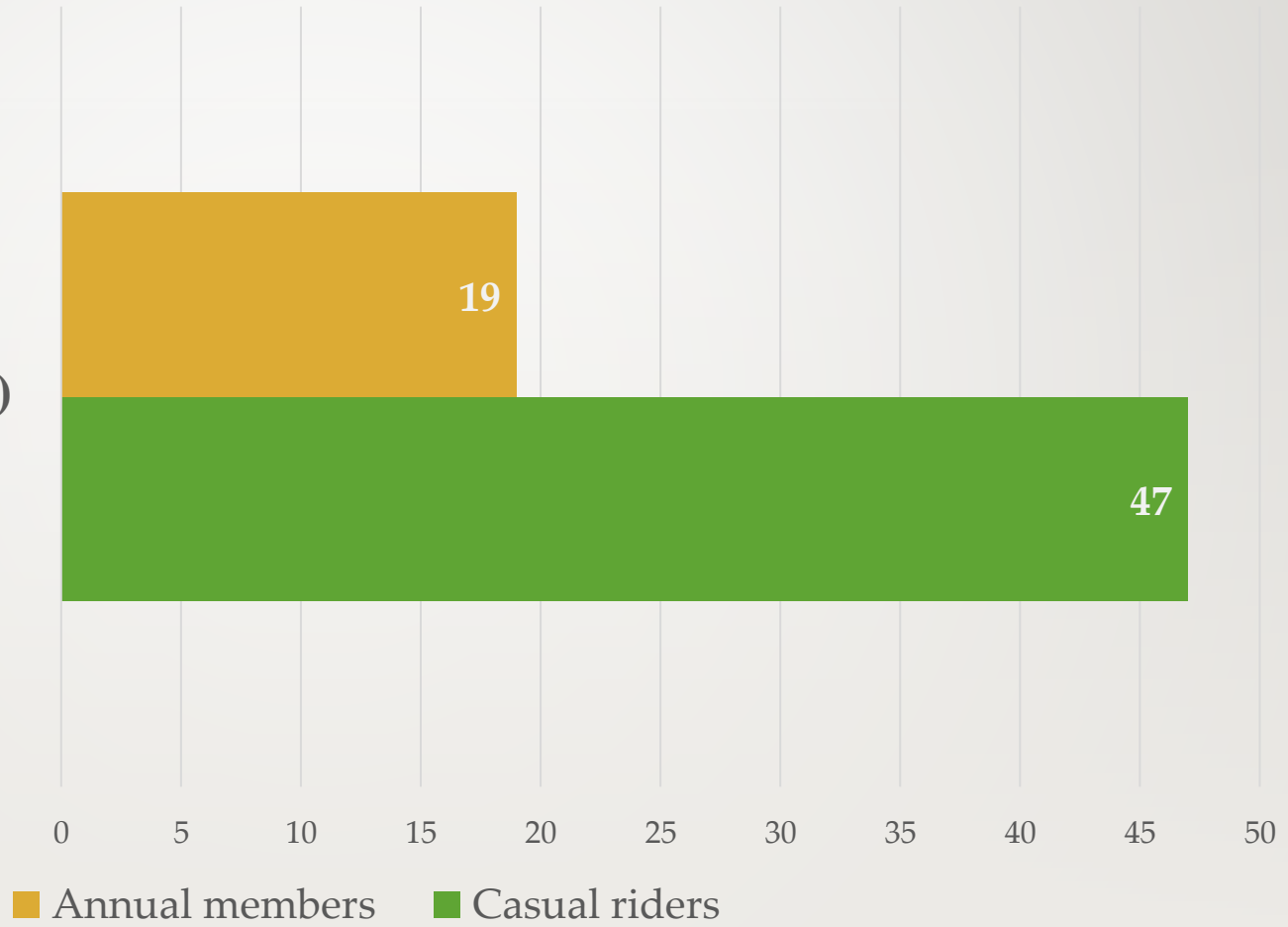
All rides April-June 2020 by membership



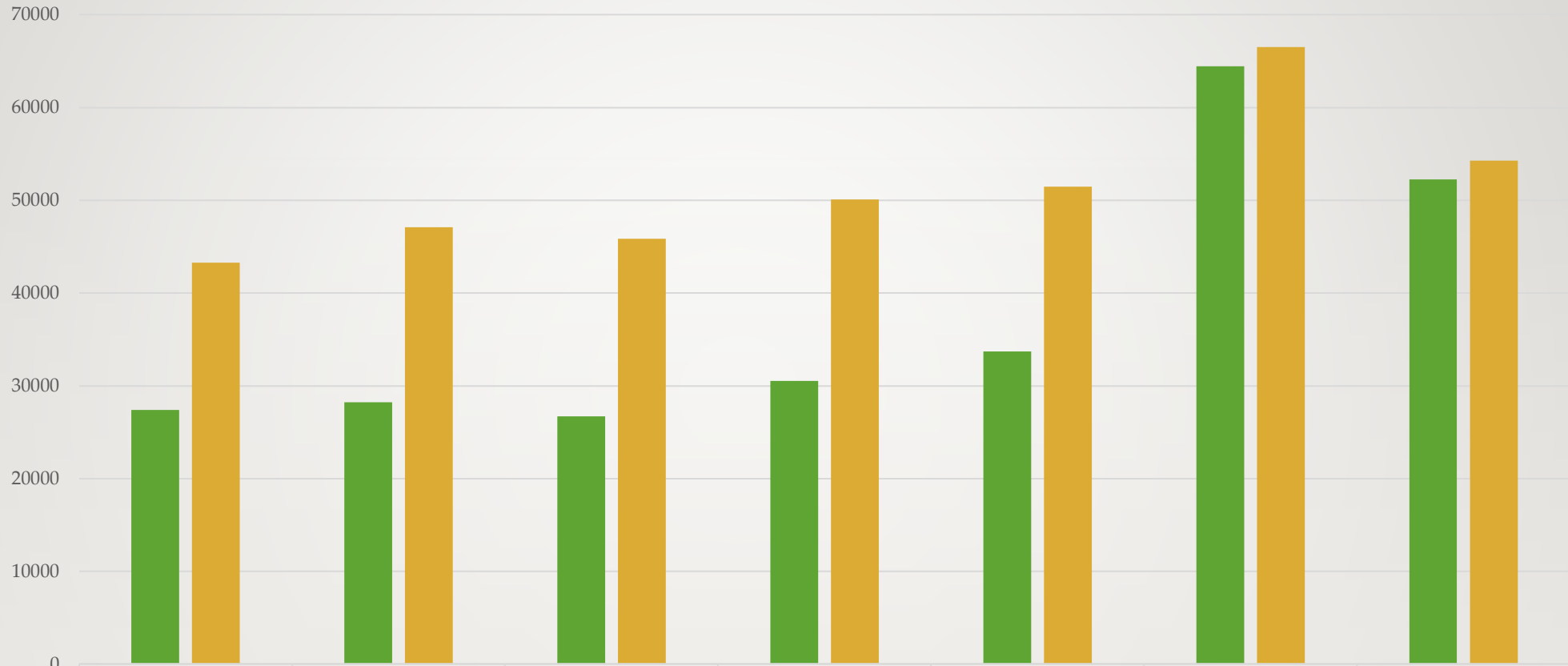
■ Casual riders ■ Annual members

How do members and casual riders use Cyclistic differently?

Average ride length (mins)



Ride preference by weekday



- Casual riders
- Annual members

	Mon	Tue	Wed	Thu	Fri	Sat	Sun
Casual riders	27416	28246	26728	30551	33729	64471	52236
Annual members	43268	47101	45848	50073	51460	66544	54290

The story

- Casual riders tend to use bikes **for leisure purposes** and when they have time – **Saturday and Sunday** rides are most common, as well as **longer length of use of bikes (47 mins)** indicates that they grab a bike when they feel that they have time for the ride.
- Annual members use bikes **both for daily commute** and **for leisure** (on weekends length of rides for annual members are slightly longer with 31 minutes on average).



Recommendations

- Offer limited-time discounts for annual memberships (e.g. send the campaign out on a weekend when casual riders use bikes most and limit the time the promotion is available for a sense of urgency)
- Consider offering guaranteed longer ride duration in high season and on weekends for annual members
- Consider analysing survey casual riders (see below)

Further analyses needed

- Analyse longer-term data to examine seasonal use preferences: in doing so, we can identify in which season casual riders use bikes the most and arrange a promotional campaign for that time)
- Survey casual riders for their preferences: what would be the deal-breaker for them to switch to annual membership? Who are these users and what additional services can we offer them to make membership even more attractive?
- Analyse casual riders' costs and prepare some promotional material to demonstrate that annual membership is more cost-effective.



APPENDIX

SQL QUERIES

Data Transformation and Cleaning

- Tables join (dataset consisted of 3 separate CSV files for each month of data)

```
1 CREATE TABLE `black-resource-410015.cyclistic_data.merged_data` AS
2 SELECT * FROM `black-resource-410015.cyclistic_data.april`
3 UNION ALL
4 SELECT * FROM `black-resource-410015.cyclistic_data.may`
5 UNION ALL
6 SELECT * FROM `black-resource-410015.cyclistic_data.june`;
7
```

- Search and remove missing values

```
1 DELETE FROM `black-resource-410015.cyclistic_data.merged_data`
2 WHERE end_station_id IS NULL OR end_station_name= ''
3 OR start_station_id IS NULL OR start_station_name= ''
```

- Calculate the length of rides from start and end time

```
1 UPDATE `black-resource-410015.cyclistic_data.merged_data`  
2 SET Ride_length = TIMESTAMP_DIFF(ended_at, started_at, MINUTE)  
3 WHERE Ride_length IS NULL;
```

- Remove ≤ 0 values for length ride as outliers and top-100 longest rides

```
DELETE FROM `black-resource-410015.cyclistic_data.merged_data`  
WHERE Ride_length IS NULL OR Ride_length = 0 OR Ride_length < 0 OR Ride_length > 6842;
```

- Calculate the day of the week from date of the ride value

```
1 CREATE OR REPLACE TABLE `black-resource-410015.cyclistic_data.merged_data` AS  
2 SELECT *,  
3 CASE  
4   WHEN EXTRACT(DAYOFWEEK FROM started_at) = 2 THEN 'Monday'  
5   WHEN EXTRACT(DAYOFWEEK FROM started_at) = 3 THEN 'Tuesday'  
6   WHEN EXTRACT(DAYOFWEEK FROM started_at) = 4 THEN 'Wednesday'  
7   WHEN EXTRACT(DAYOFWEEK FROM started_at) = 5 THEN 'Thursday'  
8   WHEN EXTRACT(DAYOFWEEK FROM started_at) = 6 THEN 'Friday'  
9   WHEN EXTRACT(DAYOFWEEK FROM started_at) = 7 THEN 'Saturday'  
10  WHEN EXTRACT(DAYOFWEEK FROM started_at) = 1 THEN 'Sunday'  
11 END AS week_day,  
12 FROM `black-resource-410015.cyclistic_data.merged_data`
```

- Descriptive statistics (number of rides, max and min values for ride length, average and median ride length)

```
1 SELECT COUNT(*) AS row_count,  
2 COUNT(DISTINCT start_station_id) AS number_of_start_stations,  
3 COUNT(DISTINCT member_casual) AS memberships_types,  
4 MIN(Ride_length) AS min_ride_length,  
5 MAX(Ride_length) AS max_ride_length,  
6 AVG(Ride_length) AS average_ride_length,  
7 FROM `black-resource-410015.cyclistic_data.merged_data`
```

- Calculate average ride length by type of membership

```
1 SELECT  
2     ROUND(AVG(CASE WHEN member_casual = 'casual' THEN ride_length END), 2) AS avg_ride_casual,  
3     ROUND(AVG(CASE WHEN member_casual = 'member' THEN ride_length END), 2) AS avg_ride_member  
4 FROM `black-resource-410015.cyclistic_data.merged_data`  
5 WHERE member_casual IN ('casual', 'member');
```

- Identify trends in weekday vs. weekend use of rent bikes by type of membership

```
1 SELECT  
2 member_casual,  
3 COUNTIF(week_day = 'Monday') AS Mon,  
4 COUNTIF(week_day = 'Tuesday') AS Tue,  
5 COUNTIF(week_day = 'Wednesday') AS Wed,  
6 COUNTIF(week_day = 'Thursday') AS Thu,  
7 COUNTIF(week_day = 'Friday') AS Fri,  
8 COUNTIF(week_day = 'Saturday') AS Sat,  
9 COUNTIF(week_day = 'Sunday') AS Sun,  
10 FROM `black-resource-410015.cyclistic_data.merged_data`  
11 GROUP BY member_casual  
12 ORDER BY member_casual;
```