Cyclistic Bikeshare Project

By Justin Okonkwo

(August 2021)

Overview

1. Background
2. Business Objective
3. Data Processing
4. Data Cleaning
5. Data Analysis
6. Data Visualization
7. Key Takeaways
8. Recommendations

Background

- **Cyclistic: Bike-Share company based in Chicago**

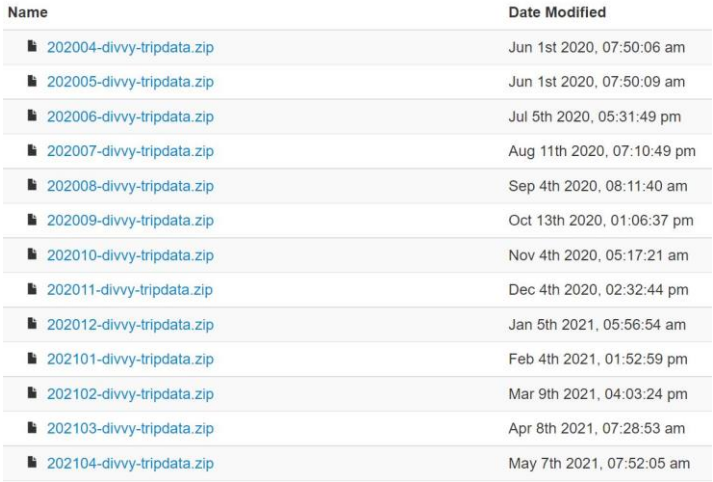
  - **Nearly 6000 bicycles and over 600+ bike stations within their program**
  - **Two different type of riders:**
    Casual and Members (Annual subscription)
  - **Director of marketing believes**
    future success -> maximizing annual memberships

Business Objective

- **Goal: Member vs Casual Riders**
  - **Identify how annual members and causal riders use Cyclistic bikes differently**
  - **Find trends/insights in the customer behavior for marketing direction on converting casual to members**
  - **Stakeholders:**
    Director of Marketing
    Cyclistic Executive Team
    Marketing Analytics Team

Data Processing

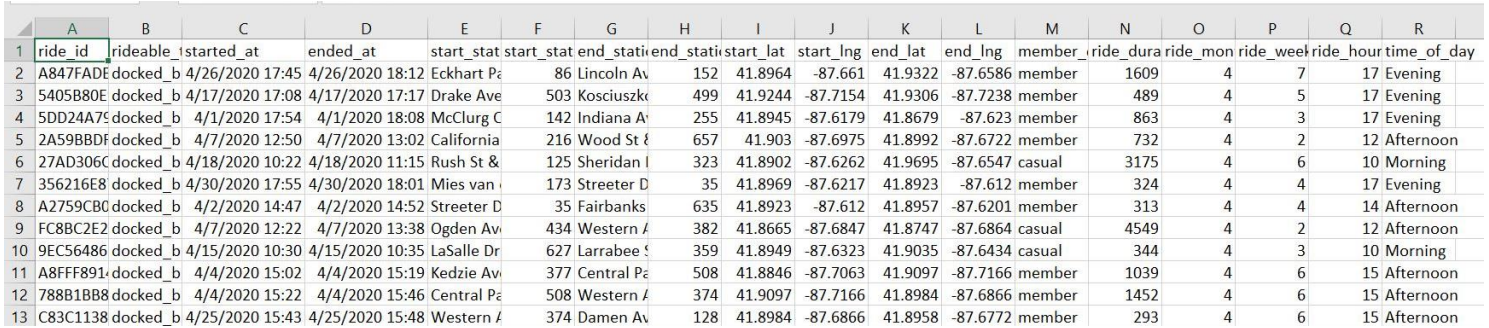
- **Datasets**
  - **Derived from divvy bike-share program, available by Motive International Inc.**
  - **Historical trip data from May 2020 – April 2021**
  - **Split into 12 CSV files, on month-to-month basis**
  - **Processed using Microsoft Excel then SQL due to memory constraints**



Name	Date Modified
202004-divvy-tripdata.zip	Jun 1st 2020, 07:50:06 am
202005-divvy-tripdata.zip	Jun 1st 2020, 07:50:09 am
202006-divvy-tripdata.zip	Jul 5th 2020, 05:31:49 pm
202007-divvy-tripdata.zip	Aug 11th 2020, 07:10:49 pm
202008-divvy-tripdata.zip	Sep 4th 2020, 08:11:40 am
202009-divvy-tripdata.zip	Oct 13th 2020, 01:06:37 pm
202010-divvy-tripdata.zip	Nov 4th 2020, 05:17:21 am
202011-divvy-tripdata.zip	Dec 4th 2020, 02:32:44 pm
202012-divvy-tripdata.zip	Jan 5th 2021, 05:56:54 am
202101-divvy-tripdata.zip	Feb 4th 2021, 01:52:59 pm
202102-divvy-tripdata.zip	Mar 9th 2021, 04:03:24 pm
202103-divvy-tripdata.zip	Apr 8th 2021, 07:28:53 am
202104-divvy-tripdata.zip	May 7th 2021, 07:52:05 am

Data Processing

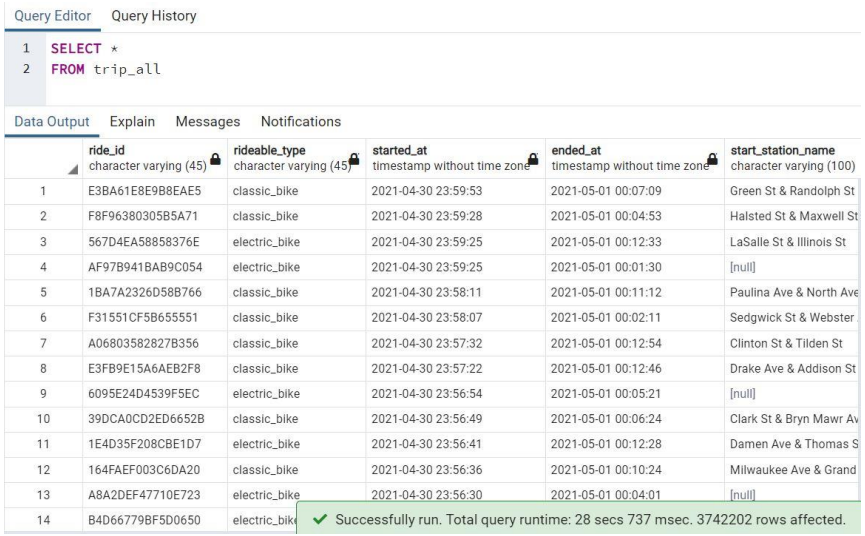
- **Microsoft Excel**
  - **13 original columns and more than 3 million rows total**
  - **Used formulas to create new columns for further analysis**
    Ex. ride_duration, ride_month, time_of_day, etc.



	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R
1	ride_id	rideable_t	started_at	ended_at	start_stat	start_stat	end_statio	end_statio	start_lat	start_lng	end_lat	end_lng	member_	ride_dura	ride_mon	ride_week	ride_hour	time_of_day
2	A847FADE	docked_b	4/26/2020 17:45	4/26/2020 18:12	Eckhart Pa	86	Lincoln Av	152	41.8964	-87.661	41.9322	-87.6586	member	1609	4	7	17	Evening
3	5405B80E	docked_b	4/17/2020 17:08	4/17/2020 17:17	Drake Ave	503	Kosciuszk	499	41.9244	-87.7154	41.9306	-87.7238	member	489	4	5	17	Evening
4	5DD24A79	docked_b	4/1/2020 17:54	4/1/2020 18:08	McClurg C	142	Indiana Av	255	41.8945	-87.6179	41.8679	-87.623	member	863	4	3	17	Evening
5	2A59BBDF	docked_b	4/7/2020 12:50	4/7/2020 13:02	California	216	Wood St &	657	41.903	-87.6975	41.8992	-87.6722	member	732	4	2	12	Afternoon
6	27AD306C	docked_b	4/18/2020 10:22	4/18/2020 11:15	Rush St &	125	Sheridan I	323	41.8902	-87.6262	41.9695	-87.6547	casual	3175	4	6	10	Morning
7	356216E8	docked_b	4/30/2020 17:55	4/30/2020 18:01	Mies van	173	Streeter D	35	41.8969	-87.6217	41.8923	-87.612	member	324	4	4	17	Evening
8	A2759CB0	docked_b	4/2/2020 14:47	4/2/2020 14:52	Streeter D	35	Fairbanks	635	41.8923	-87.612	41.8957	-87.6201	member	313	4	4	14	Afternoon
9	FC8BC2E2	docked_b	4/7/2020 12:22	4/7/2020 13:38	Ogden Av	434	Western A	382	41.8665	-87.6847	41.8747	-87.6864	casual	4549	4	2	12	Afternoon
10	9EC56486	docked_b	4/15/2020 10:30	4/15/2020 10:35	LaSalle Dr	627	Larrabee S	359	41.8949	-87.6323	41.9035	-87.6434	casual	344	4	3	10	Morning
11	A8FFF891	docked_b	4/4/2020 15:02	4/4/2020 15:19	Kedzie Av	377	Central Pa	508	41.8846	-87.7063	41.9097	-87.7166	member	1039	4	6	15	Afternoon
12	788B1BB8	docked_b	4/4/2020 15:22	4/4/2020 15:46	Central Pa	508	Western A	374	41.9097	-87.7166	41.8984	-87.6866	member	1452	4	6	15	Afternoon
13	C83C1138	docked_b	4/25/2020 15:43	4/25/2020 15:48	Western A	374	Damen Av	128	41.8984	-87.6866	41.8958	-87.6772	member	293	4	6	15	Afternoon

Data Processing

- **PgAdmin4 (PostgreSQL)**
  - **Compared to Excel, SQL has capacity to analyze full dataset**
  - **Created tables for dataset, imported CSV files, and merged (UNION)**
  - **Ability to clean and analyze data, comprehensively**



Query Editor | Query History

```
1  SELECT *
2  FROM trip_all
```

Data Output | Explain | Messages | Notifications

	ride_id character varying (45)	rideable_type character varying (45)	started_at timestamp without time zone	ended_at timestamp without time zone	start_station_name character varying (100)
1	E3BA61E8E9B8EAE5	classic_bike	2021-04-30 23:59:53	2021-05-01 00:07:09	Green St & Randolph St
2	F8F96380305B5A71	classic_bike	2021-04-30 23:59:28	2021-05-01 00:04:53	Halsted St & Maxwell St
3	567D4EA58858376E	electric_bike	2021-04-30 23:59:25	2021-05-01 00:12:33	LaSalle St & Illinois St
4	AF97B941BAB9C054	electric_bike	2021-04-30 23:59:25	2021-05-01 00:01:30	[null]
5	1BA7A2326D58B766	classic_bike	2021-04-30 23:58:11	2021-05-01 00:11:12	Paulina Ave & North Ave
6	F31551CF5B655551	classic_bike	2021-04-30 23:58:07	2021-05-01 00:02:11	Sedgwick St & Webster
7	A06803582827B356	classic_bike	2021-04-30 23:57:32	2021-05-01 00:12:54	Clinton St & Tilden St
8	E3FB9E15A6AEB2F8	classic_bike	2021-04-30 23:57:22	2021-05-01 00:12:46	Drake Ave & Addison St
9	6095E24D4539F5EC	electric_bike	2021-04-30 23:56:54	2021-05-01 00:05:21	[null]
10	39DCA0CD2ED6652B	classic_bike	2021-04-30 23:56:49	2021-05-01 00:06:24	Clark St & Bryn Mawr Av
11	1E4D35F208CBE1D7	electric_bike	2021-04-30 23:56:41	2021-05-01 00:12:28	Damen Ave & Thomas S
12	164FAEF003C6DA20	classic_bike	2021-04-30 23:56:36	2021-05-01 00:10:24	Milwaukee Ave & Grand
13	A8A2DEF47710E723	electric_bike	2021-04-30 23:56:30	2021-05-01 00:04:01	[null]
14	B4D66779BF5D0650	electric_bik			

✔ Successfully run. Total query runtime: 28 secs 737 msec. 3742202 rows affected.

Data Cleaning

- **Methods used:**
  - **Check for duplicates**
  - **Identify null values**
  - **Inaccurate rows**
  - **Drop outliers – IQR Method**

Query Editor | Query History

```
1  SELECT *
2  FROM trip_all
3  WHERE ride_duration < 0
4  order by started_at asc
```

Data Output | Explain | Messages | Notifications

at	start_lng real	end_lat real	end_lng real	member_casual character varying (45)	ride_duration bigint	ride_month bigint	ride_weekday bigint	ride_hour bigint	time_of_day character varying (45)
614	-87.6762	41.9659	-87.7008	member	-70	6	1	10	Morning
908	-87.6315	41.9105	-87.6531	casual	-65	6	1	10	Morning
494	-87.6545	41.9491	-87.6486	member	-57	6	1	10	Morning
222	-87.6389	41.908	-87.6315	member	-53	6	1	10	Morning
117	-87.6268	41.9367	-87.6368	member	-36	6	1	10	Morning
828	-87.6612	41.8653	-87.6179	casual	-29	6	1	10	Morning
688	-87.6577	41.9617	-87.6546	member	-23	6	1	10	Morning
117	-87.6268	41.9139	-87.6488	member	-11	6	1	10	Morning
894	-87.6293	41.9069	-87.6262	member	-243	6	1	10	Morning
664	-87.5657	41.781	-87.5761	member	-2	6	1	10	Morning
263	-87.6308	41.984	-87.6523	member	-9	6	1	10	Morning
665	-87.6884	41.9665	-87.6884	member	-4	6	1	10	Morning

Data Analysis

- **Descriptive Analysis**
  - **Summary Statistics**
  - **Ride Count**
  - **Ride Duration**
  - **Top 5 Stations by Member_Casual**

Query Editor | Query History

```
1  SELECT 'count' as summary_statistics,
2      count(*) as rd_stats
3  FROM trip_all
4  UNION
5  SELECT 'total',
6      sum(ride_duration)
```

Data Output | Explain | Messages | Notifications

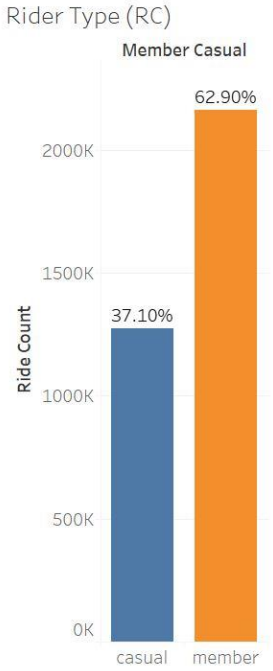
	summary_statistics text	rd_stats double precision
1	min	-1742998
2	mean	1467.54927980905
3	median	858
4	count	3742202
5	Q3	1581
6	Q1	467
7	max	3257001
8	total	5491865850
9	std_dev	22353.83654037

Data Visualization

- **Software: Tableau**
  - **Imported final CSV file for visualizations**
  - **Easier to see and share insights vs. tables/spreadsheets**
  - **Ride Count, Ride Duration**

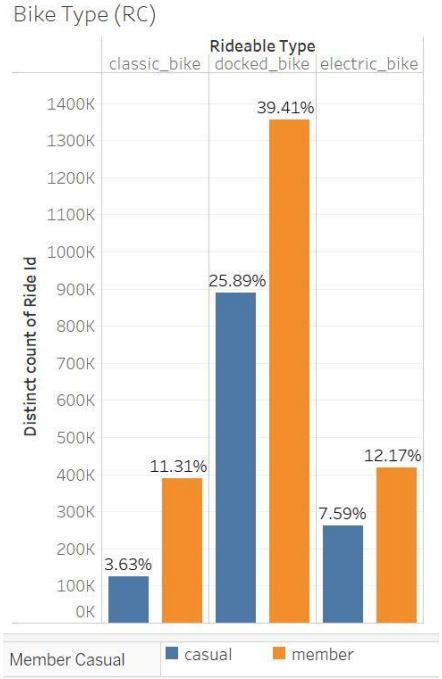
Data Visualization

Ride Count by Member/Casual:



Nearly 2/3 rides are by member riders

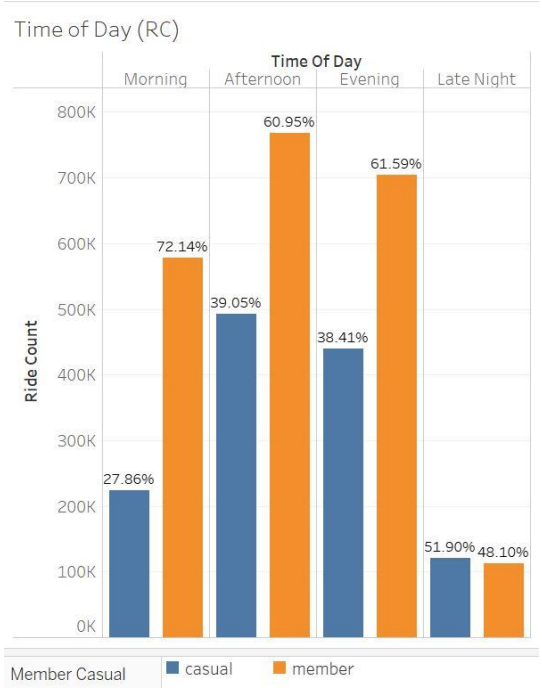
Ride Count by Bike Type:



Highest usage: Docked Bike (approx. 65%)

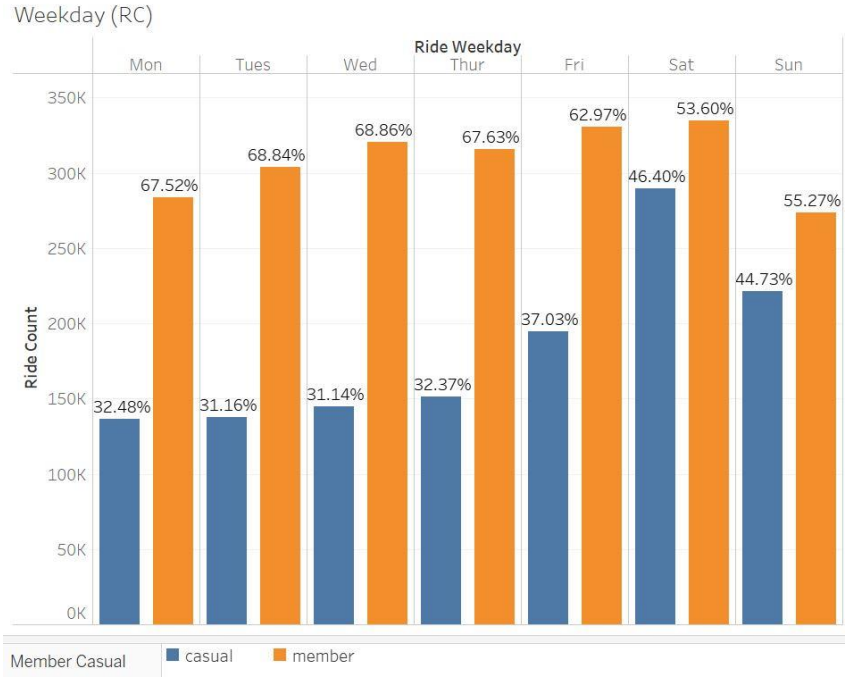
Data Visualization

Ride Count by Time of Day:



Casual riders less active during Late Nights/Mornings

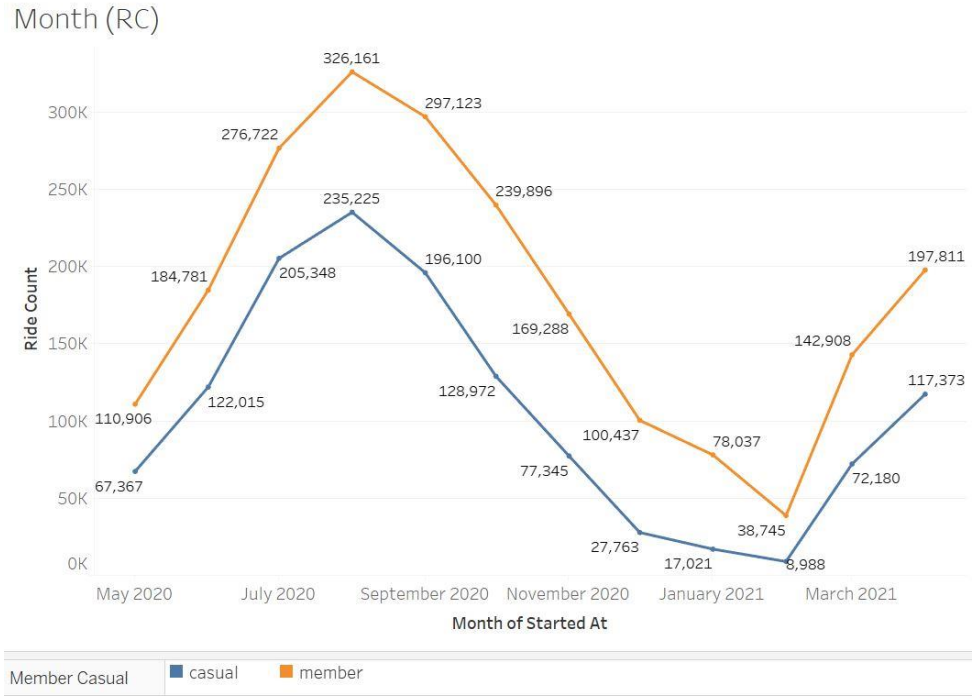
Ride Count by Day of the Week:



Increased rate of casual riders during weekends vs. weekdays

Data Visualization

Ride Count over last 12 months:

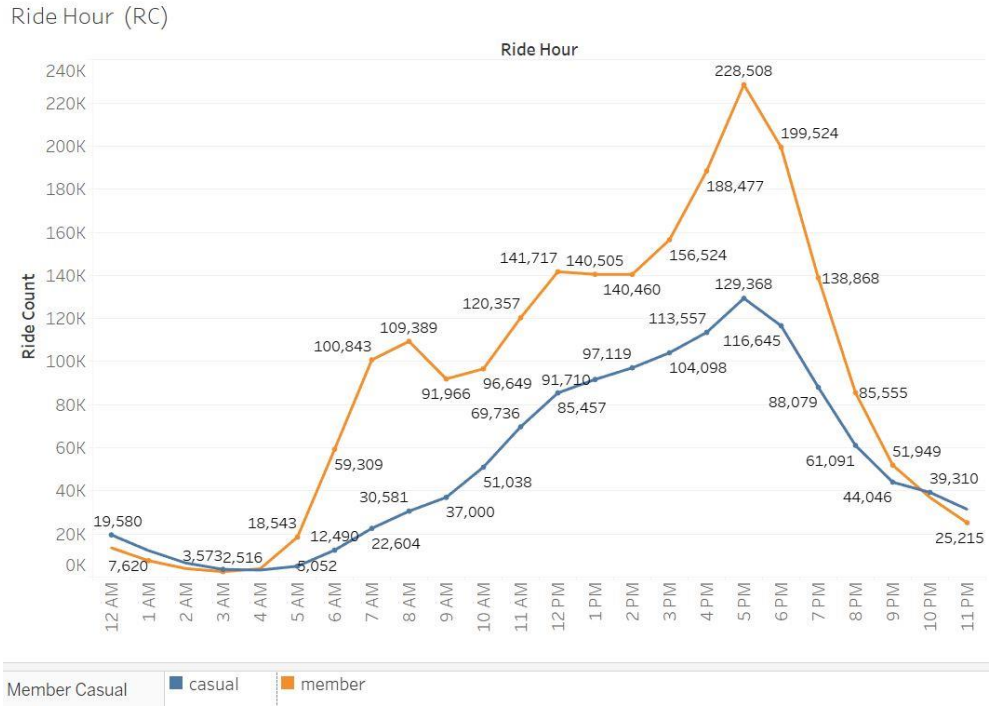


Peak ride count during Summer (July, Aug, Sept)

Lowest ride count during Winter (Dec, Jan, Feb)

Data Visualization

Ride Count by Ride Hour:

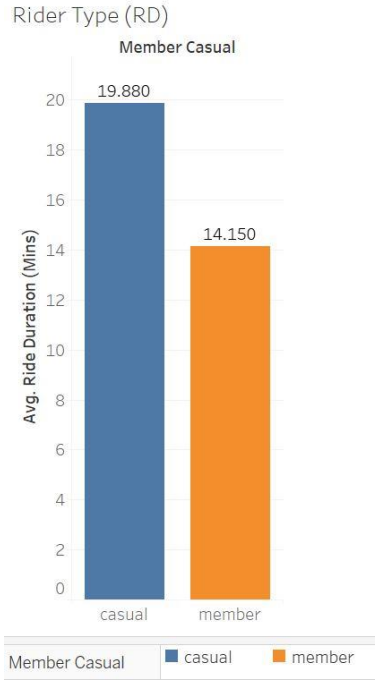


Peak ride count during 4-6 PM

Higher usage for casual riders during Late Night (10PM-2AM)

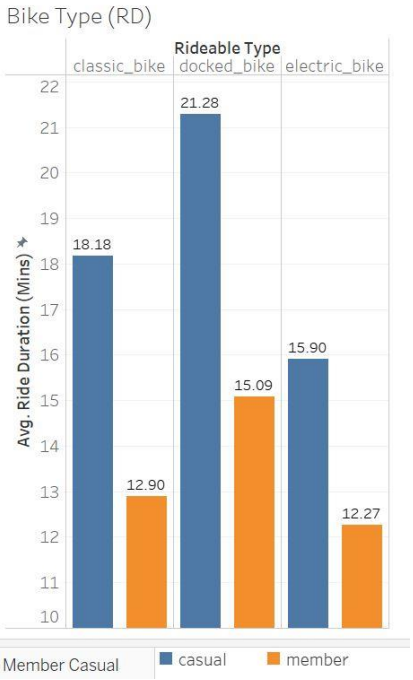
Data Visualization

Ride Duration by Member/Casual:



Casual rides are longer by avg. of 5 mins

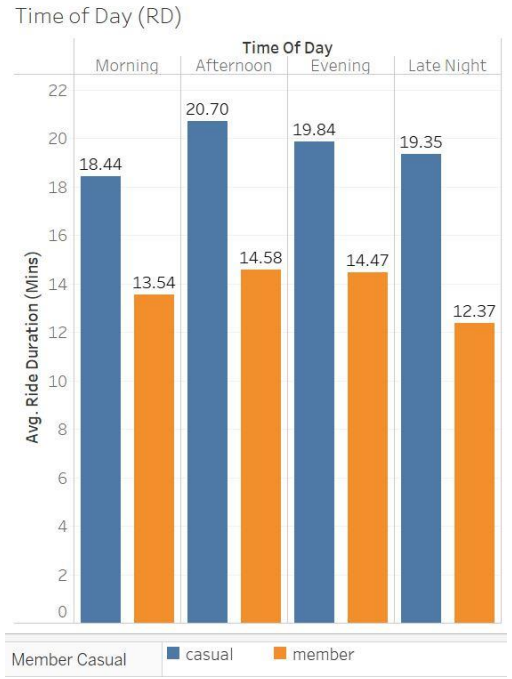
Ride Duration by Bike Type:



Longer usage: Docked Bike

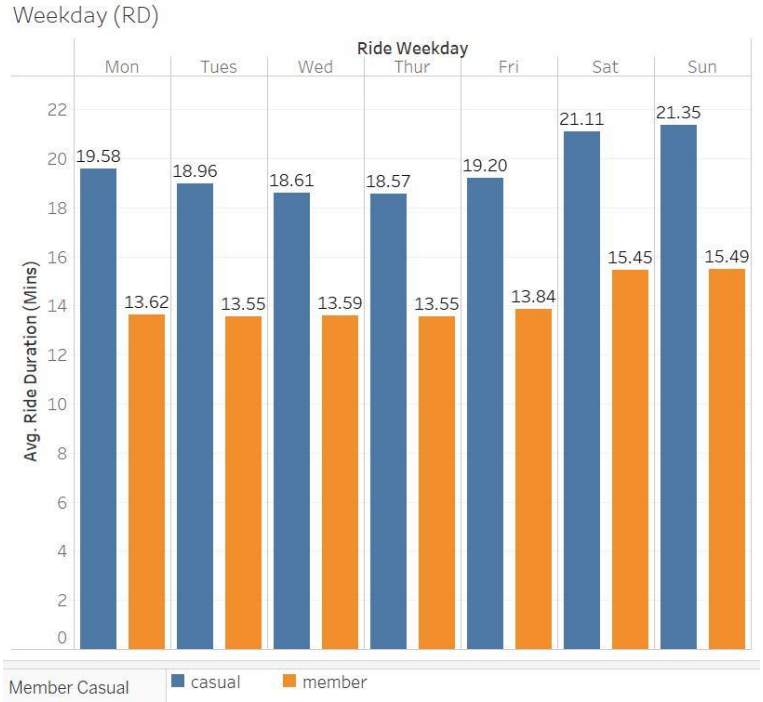
Data Visualization

Ride Duration by Time of Day:



Not much difference across the time of day

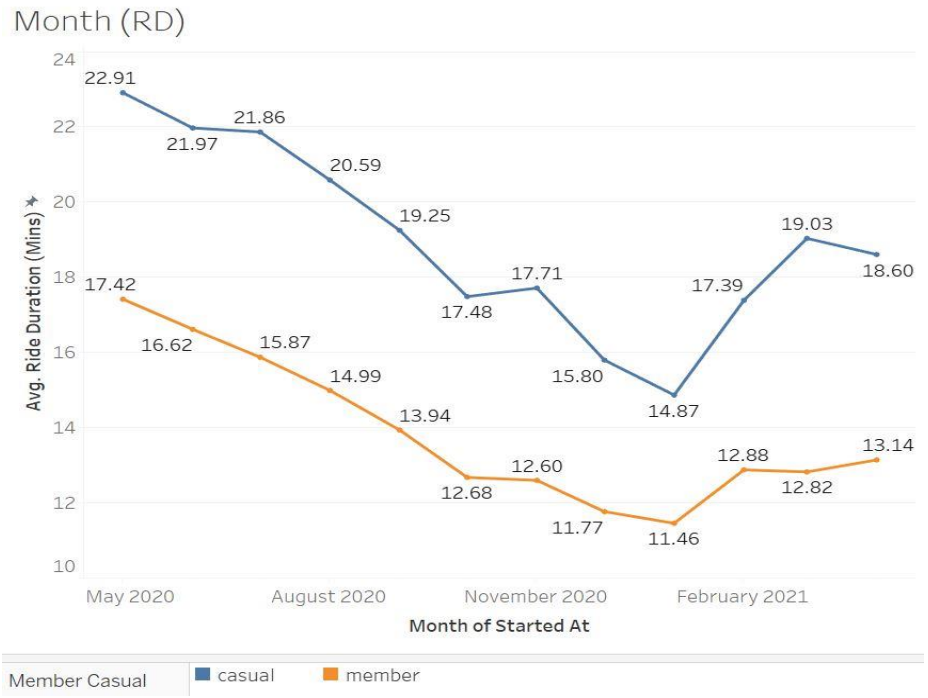
Ride Duration by Day of the Week:



Increased duration during weekends vs. weekdays

Data Visualization

Ride Duration over last 12 months:

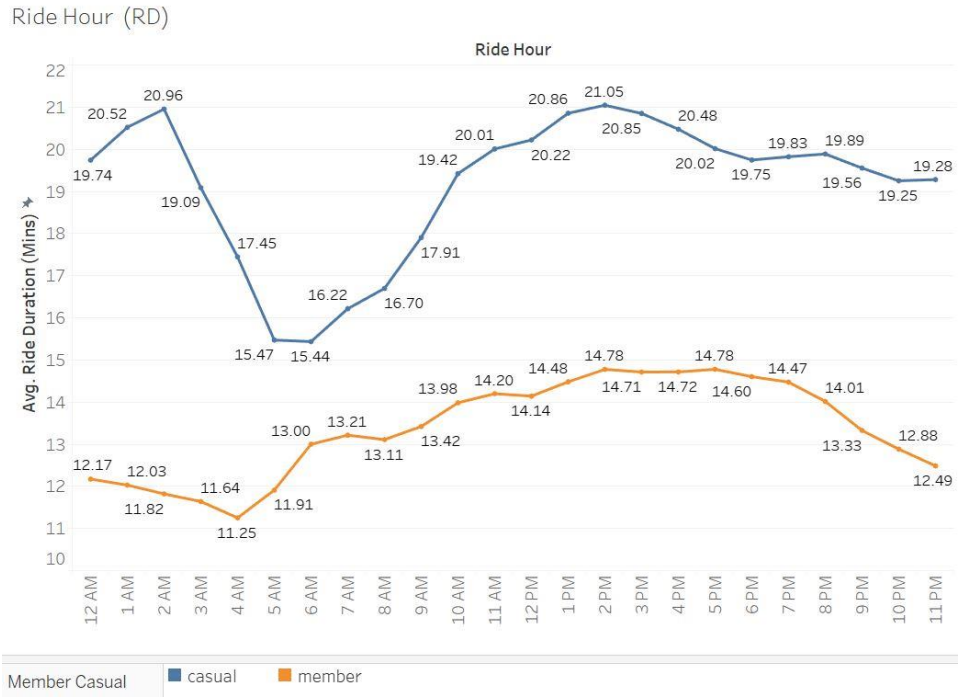


Higher ride duration during Summer (May, June, July)

Lowest ride duration during Winter (Dec, Jan, Feb)

Data Visualization

Ride Duration by Ride Hour:



Decrase in ride duration for casual riders 2AM - 6AM

Data Visualization

Top Casual Start Stations:

Top Start Stations

Start Station Na..	Member Casual: casual	member
Streeter Dr & Gr..	21,168	9,623
Lake Shore Dr & ..	17,352	9,610
Millennium Park	14,658	5,024
Theater on the L..	13,557	15,419
Lake Shore Dr & ..	11,635	14,110
Michigan Ave & ..	11,452	8,758
Clark St & Elm St	11,332	21,194
Indiana Ave & Ro..	10,532	11,152
Wells St & Conco..	9,885	16,160
Clark St & Lincol..	9,536	12,808
Michigan Ave & L..	9,382	8,158
Shedd Aquarium	9,279	4,357
Clark St & Armit..	9,200	14,390
Wells St & Elm St	9,131	15,107
Wells St & Everg..	8,730	10,865

Top Station:
Streeter Dr & Grand Ave

Top Start Station:

Top Start Stations

Start Station..	
Clark St & Elm St	32,526
Streeter Dr & Gr..	30,791
Theater on the L..	28,976
Lake Shore Dr & ..	26,962
Wells St & Conco..	26,045
Lake Shore Dr & ..	25,745
Wells St & Elm St	24,238
Broadway & Barr..	24,175
Dearborn St & Er..	23,774
Clark St & Armit..	23,590
Wells St & Huron..	23,128
Clark St & Lincol..	22,344
Indiana Ave & Ro..	21,684
Kingsbury St & Ki..	21,335
St. Clair St & Erie..	21,322
Desplaines St & ..	20,849

Top Station:
Clark St & Elm St

Top Member Start Stations:

Top Start Stations

Start Station Na..	Member Casual: casual	member
Clark St & Elm St	11,332	21,194
Broadway & Barr..	7,786	16,389
Wells St & Conco..	9,885	16,160
St. Clair St & Erie..	5,292	16,030
Dearborn St & Er..	7,758	16,016
Theater on the L..	13,557	15,419
Kingsbury St & Ki..	5,980	15,355
Wells St & Elm St	9,131	15,107
Wells St & Huron..	8,170	14,958
Clark St & Armit..	9,200	14,390
Desplaines St & ..	6,664	14,185
Lake Shore Dr & ..	11,635	14,110
Columbus Dr & R..	7,304	13,139
Lake Shore Dr & ..	7,891	12,924
Larrabee St & W..	7,909	12,850

Top Station:
Clark St & Elm St

Data Visualization

Top Casual End Stations:

Top End Stations

End Station Name	Member Casual casual	member
Streeter Dr & Gr..	22,371	9,429
Lake Shore Dr & ..	16,294	9,593
Millennium Park	15,893	5,450
Theater on the L..	14,664	15,033
Lake Shore Dr & ..	12,087	13,657
Michigan Ave & ..	11,856	8,835
Clark St & Elm St	10,708	21,646
Indiana Ave & Ro..	10,604	10,569
Wells St & Conco..	10,030	16,556
Clark St & Lincol..	9,608	12,490
Clark St & Armit..	8,997	13,234
Michigan Ave & L..	8,918	8,680
Wells St & Elm St	8,649	14,379
Michigan Ave & ..	8,472	6,804
Wells St & Everg..	8,424	10,863

Top Station:
Streeter Dr & Grand Ave

Top End Station:

Top End Stations

End Station ..
Clark St & Elm St	32,354
Streeter Dr & Gr..	31,800
Theater on the L..	29,697
Wells St & Conco..	26,586
Lake Shore Dr & ..	25,887
Lake Shore Dr & ..	25,744
Broadway & Barr..	24,767
Dearborn St & Er..	24,231
St. Clair St & Erie..	23,367
Wells St & Elm St	23,028
Clark St & Armit..	22,231
Clark St & Lincol..	22,098
Wells St & Huron..	21,831
Millennium Park	21,343
Larrabee St & W..	21,277
Wabash Ave & Gr..	21,196

Top Station:
Clark St & Elm St

Top Member End Stations:

Top End Stations

End Station Name	Member Casual casual	member
Clark St & Elm St	10,708	21,646
St. Clair St & Erie..	6,148	17,219
Broadway & Barr..	8,195	16,572
Wells St & Conco..	10,030	16,556
Dearborn St & Er..	7,702	16,529
Kingsbury St & Ki..	5,198	15,537
Theater on the L..	14,664	15,033
Wells St & Elm St	8,649	14,379
Wells St & Huron..	7,846	13,985
Wabash Ave & R..	6,979	13,667
Lake Shore Dr & ..	12,087	13,657
Larrabee St & W..	7,992	13,285
Clark St & Armit..	8,997	13,234
Lake Shore Dr & ..	8,016	13,141
Desplaines St & ..	6,238	13,069

Top Station:
Clark St & Elm St

Key Takeaways

- **Docked Bike is the most popular option for both riders**
- **Consistency between both start and end stations, relative to rider type**
- **Member/Causal trends by Ride Count**
  - ***Difference in ride count by time of year – Summer vs. Winter***
  - ***Spikes in ride count from 4PM – 6PM, but higher casual rate during 'Late Night' (10PM – 2AM)***
  - ***Increase ride count towards weekends (Fri-Sun) over weekdays (Mon-Thurs)***
- **Member/Causal trends by Ride Duration**
  - ***Casual ride duration > Member ride duration, by average of 5 mins***
  - ***Decrease in casual ride duration from 2AM – 6AM***

Recommendations

- **Marketing strategies/discounts towards casual trends (membership perks)**
  - ***Weekends, Summer, Late Night (10PM – 2AM) discounts***
  - ***Discounts towards longer ride duration (majority of casual riders)***
  - ***Exclusive or easier access to the popular 'Docked Bike' vs. rest***
  - ***Discounts/promotions related to top casual stations or nearby businesses***
- **Data Suggestions**
  - ***Ride ID consistency***
  - ***Both start and end station ID consistency***
  - ***Permissible customer data/demographics under proper ethics and data governance (age, gender, income, etc.)***

Questions/Comments