Cyclistic Bike Share Analysis- Report

Capstone Project

DM13 – Business Analytics and Digital Marketing

Done By – Kannan Lankaraman

Problem Statement:

I am a junior data analyst working in the marketing analyst team at Cyclistic, a bike-share company in Chicago. The director of marketing believes the company's future success depends on maximizing the number of annual memberships. Therefore, my team wants to understand how casual riders and annual members use Cyclistic bikes differently. From these insights, my team will design a new marketing strategy to convert casual riders into annual members.

Metadata:

1. ride_id => unique ride id for each customer.
2. rideable_type => defines whether he rode electric or docked bike.
3. started_at => mentions the starting date and time of the customer rides.
4. ended_at => mentions the ending time.
5. start_station_name => mentions the starting station name.
6. end_station_name => mentions the ending station name.
7. start_station_id and end_station_id => mentions the id of the stations.
8. start_lat and start_lng => gives the coordinates for the starting point.
9. end_lat and end_lng => gives the coordinates for the ending point.
10. members => mentions whether he/she is a casual or a member(annual)

No. of columns = 13

No. of rows = 500,000 +

Data Cleaning Process Summary:

Column Assessment:

An initial evaluation of the dataset revealed no surplus columns that were unnecessary for the analysis.

Data Type Validation:

1. A thorough examination of each column's data types was conducted in the Power Query Editor. It was identified that the 'started_at' and 'ended_at' columns contained date and time values in text format.
2. The 'Transform' tab in Power Query Editor was utilized to convert these columns into the appropriate date/time format. All other columns retained their respective data types.

Duplicate Row Identification and Removal:

1. A check for duplicate rows was performed based on the primary key, 'ride_id'.
2. The 'ride_id' column was selected, and the 'Column Quality' dialogue box was accessed via the 'View' menu.
3. The analysis revealed that out of 1000 rows, 992 were unique, indicating the presence of 8 duplicate rows.
4. Duplicate rows were removed by right-clicking near the column name and selecting the 'Remove Duplicates' option.
5. Assurance was provided that no other columns contained duplicate values.

Elimination of Blank and Null Values:

1. Blank and null values were identified in columns 'rideable_type,' 'start_station_name,' 'end_station_name,' 'start_station_id,' and 'end_station_id.'
2. To address this issue, the 'Remove Empty' filter was applied to the respective columns, effectively eliminating both blank and null values.

Column Renaming:

To enhance column clarity, the 'members' column was renamed to 'membership.'

Value Replacement:

1. In the 'membership' column, the value 'member' was replaced with 'annual' to represent annual memberships.
2. This transformation was achieved by right-clicking the column header and selecting the 'Replace Values' option.

Value Formatting:

1. For improved visualization, formatting adjustments were made.
2. In the 'membership' column, values 'casual' and 'annual' that began with lowercase letters were formatted to start with uppercase letters. This was achieved by selecting the 'Transform' option and capitalizing the initial letter.

This comprehensive data cleaning process ensures that the dataset is well-prepared for subsequent analysis and visualization.

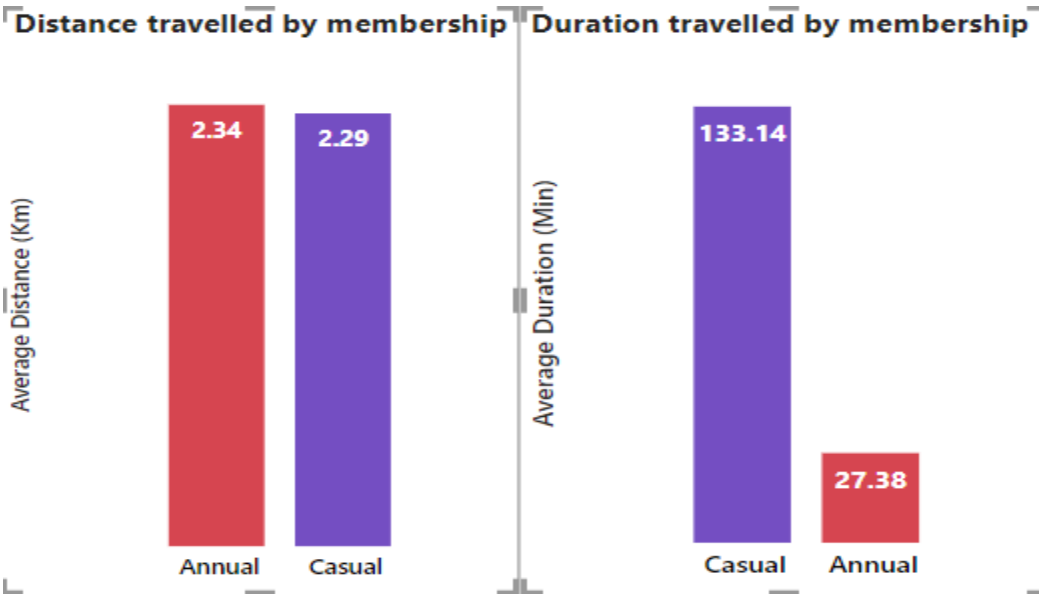
Summary:

The analysis of Cyclistic's historical bike trip data reveals distinct usage patterns between annual members and casual riders. While both segments peak in the summer months, members ride more often overall and primarily on weekdays, suggesting predominantly commuting use. Casual riders exhibit more leisure usage on weekends and around recreational destinations. Vehicle preferences also differ with casual riders less likely to use electric bikes.

Key Insights:

Distance and Duration:

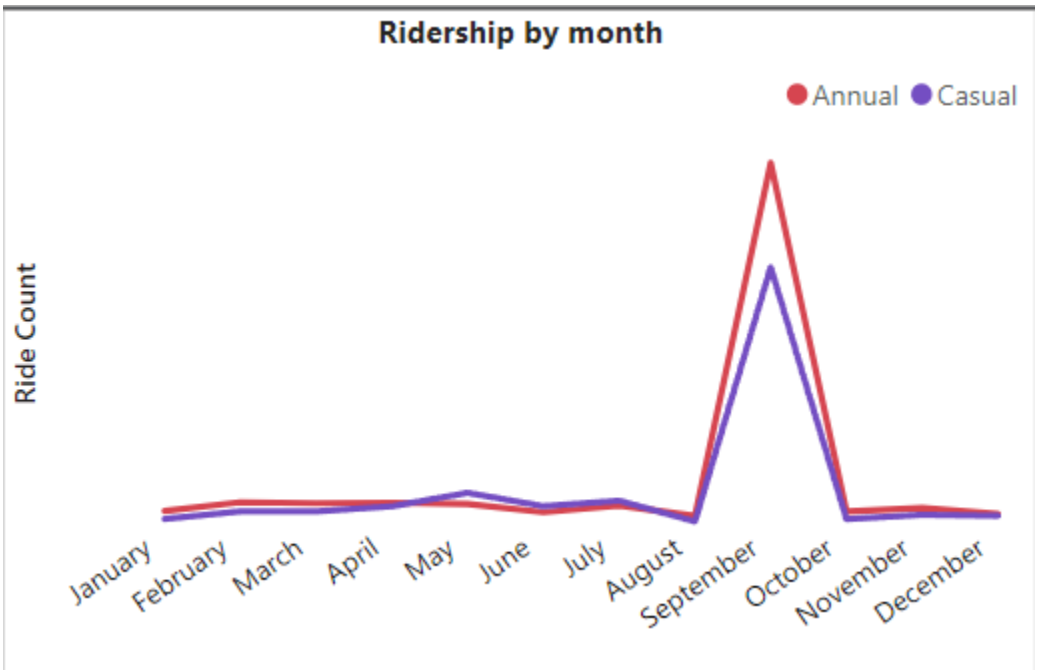
1. Annual members take longer rides on average - 2.34km vs 2.29km for casual riders.
2. The average duration for casual riders is longer at 133 minutes compared to 27 minutes for annual members.
3. The longer duration and shorter distance for casual riders indicates use for leisure and sight-seeing activities.
4. The shorter duration and longer distance for annual riders indicates use for commuting to office.



Ridership by Month:

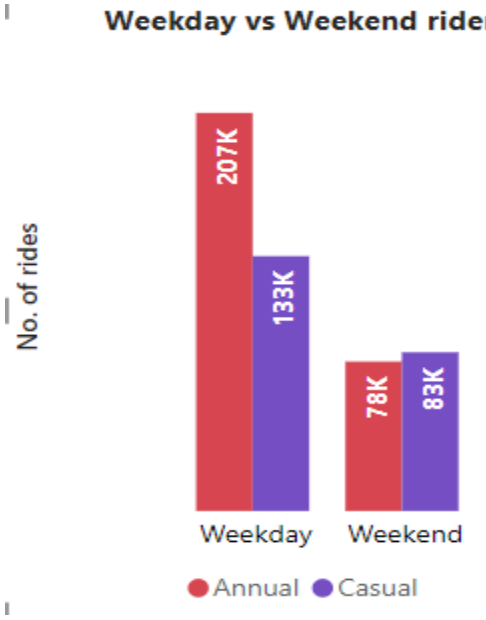
1. Annual membership ridership peaks in September with over 150,000 rides.

2. Casual ridership also peaks in September, but the peak is lower at around 130,000 rides.
3. The annual member peak in September of nearly 190,000 rides seems abnormal.
4. This spike is likely due to annual membership renewals occurring in September as the riding season winds down.
5. Also, it can be due to the fact that September is a summer season in Chicago so more casual riders may take bike for sightseeing and leisure activities.
6. Other than September, the consistent base of only 5,000-10,000 rides per month with a steep fall in winter months.



Ridership by Weekday vs Weekend:

1. Both annual members and casual riders use Cyclistic more during the weekdays than weekends.
2. However, casual ridership is more evenly distributed between weekdays and weekends compared to annual members.
3. Over 75% of member rides are on weekdays while only 60% of casual rides are weekdays.
4. As the ridership of annual members are high on weekdays, the casual ridership is more than annual ridership on weekends.
5. This suggests predominantly commuter use (annual members) and utilitarian use (casual riders)



Heat Map Analysis:

1. Highest ridership is on weekdays from (7-9am) and peaks around (5-7pm) when most riders take their bike ride, indicating commute times.
2. A smaller evening peak occurs around (8-9pm), likely representing social/dinner trips.
3. Late night (12-5am) and very early morning (2-5am) have lowest ridership.
4. Fridays have notably lower ridership than other weekdays, particularly during commuting times. People may work from home more on Fridays.
5. Weekend usage peaks around (10am-4pm).
6. Weekends see more leisurely daytime usage rather than concentrated commuting.

Annual Members

7. Most pronounced peaks during weekday commute times (7-9am) and (5-7pm). Same evening peak around (8-9pm).
8. Weekend usage is significantly lower and more evenly distributed between (10am-6pm). Members primarily use bikes for commuting.
9. Very low late night and early morning use on all days. Members don't typically use bikes for late night transportation.

Casual Riders

10. Weekday peaks are less extreme than annual members, showing less commuting focus. Steadier weekday usage.
11. Weekends see higher relative usage than members - more recreation focused. Peak weekend times are (1-5pm).

12. Higher late night (12-4am) usage than members, may represent social trips and nightlife transportation.

Overall Heat Map:

Hour	Sunday	Monday	Tuesday	Wednesday	Thursday	Friday	Saturday
0	1277	411	587	624	743	305	1166
1	744	196	353	313	414	146	633
2	420	98	192	177	221	69	317
3	193	62	131	74	138	52	166
4	149	101	128	131	163	69	103
5	324	576	433	509	562	348	146
6	985	1911	1372	1778	1976	1155	469
7	1860	3400	2311	3195	3410	2144	1034
8	2503	3307	2352	3400	3766	2294	1907
9	3408	2282	1761	2294	3263	1733	2993
10	4766	2430	1875	2218	3837	1754	4146
11	6833	3098	2744	2779	4870	2222	5407
12	6341	3847	3373	3476	5856	2834	6219
13	6579	3871	3433	3555	5535	2772	6329
14	6483	3709	3879	3573	5583	2867	6419
15	7481	4499	4447	4592	6435	3583	6534
16	8055	6221	5860	6024	7575	4419	6201
17	8515	9031	8095	9083	9352	5088	5874
18	7648	8393	7597	8381	8247	4368	5258
19	4910	5522	4893	5014	5701	2832	3787
20	2842	3095	2765	3183	3416	1697	2450
21	1967	2140	2059	2465	2559	1253	2020
22	1502	1863	1666	2184	2289	1444	2208
23	890	1414	1086	1519	1864	1355	1941

Annual Members Heat Map:

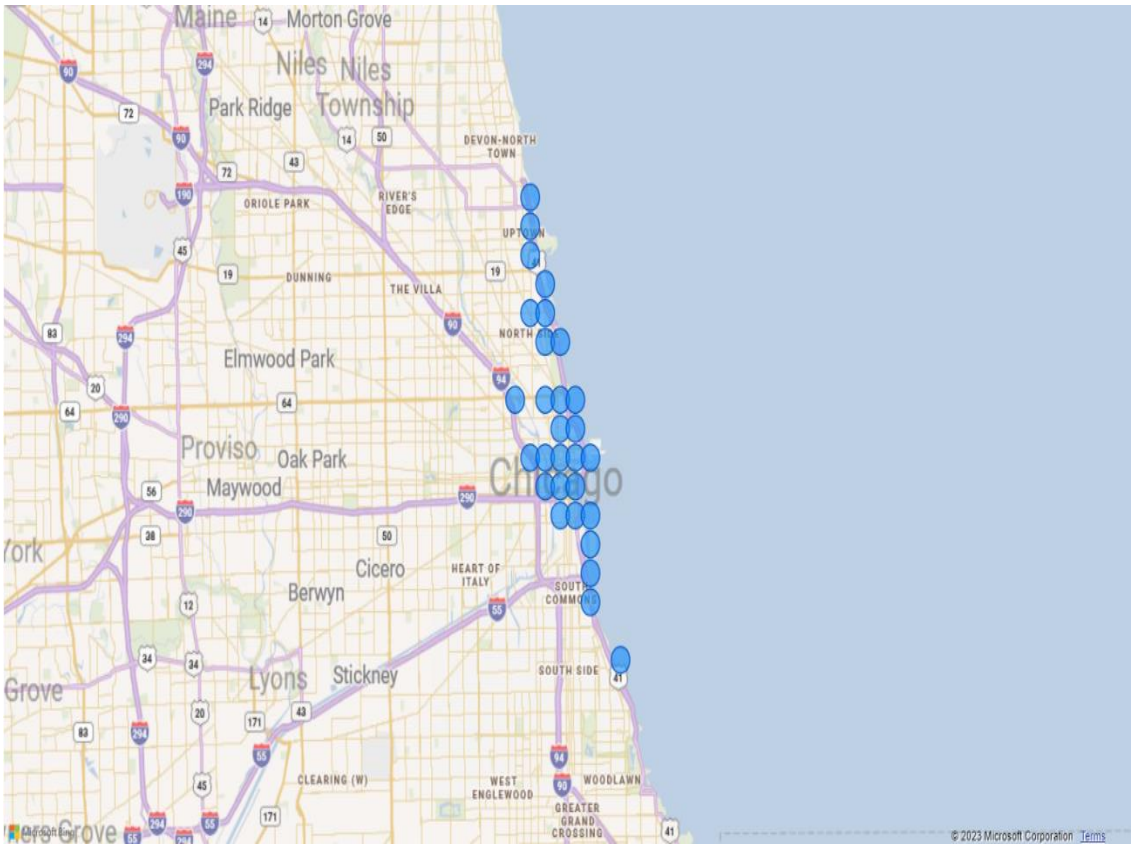
Hour	Sunday	Monday	Tuesday	Wednesday	Thursday	Friday	Saturday
0	415	142	182	233	245	132	362
1	224	56	107	124	95	44	212
2	125	23	46	65	52	31	111
3	58	20	34	34	34	18	58
4	69	59	62	71	80	38	43
5	223	454	343	410	414	270	75
6	755	1563	1128	1489	1595	946	322
7	1386	2774	1912	2616	2700	1741	716
8	1709	2622	1855	2696	2843	1764	1204
9	2032	1646	1295	1689	2127	1222	1686
10	2613	1617	1282	1505	2224	1139	2007
11	3616	2006	1699	1854	2776	1327	2413
12	3247	2468	2032	2343	3332	1702	2637
13	3268	2389	1906	2293	2874	1544	2617
14	3182	2280	2129	2228	2746	1602	2543
15	3724	2780	2412	2806	3405	2052	2596
16	4234	4024	3403	3902	4261	2633	2493
17	4712	5810	4918	5796	5610	3048	2406
18	4319	5339	4382	5284	4732	2501	2296
19	2745	3512	2850	3096	3214	1673	1604
20	1557	1810	1497	1840	1844	918	1013
21	972	1183	1017	1328	1294	592	752
22	674	886	713	999	1045	612	757
23	361	592	418	597	693	536	652

Casual Riders Heat Map:

Hour	Sunday	Monday	Tuesday	Wednesday	Thursday	Friday	Saturday
0	1277	411	587	624	743	305	1166
1	744	196	353	313	414	146	633
2	420	98	192	177	221	69	317
3	193	62	131	74	138	52	166
4	149	101	128	131	163	69	103
5	324	576	433	509	562	348	146
6	985	1911	1372	1778	1976	1155	469
7	1860	3400	2311	3195	3410	2144	1034
8	2503	3307	2352	3400	3766	2294	1907
9	3408	2282	1761	2294	3263	1733	2993
10	4766	2430	1875	2218	3837	1754	4146
11	6833	3098	2744	2779	4870	2222	5407
12	6341	3847	3373	3476	5856	2834	6219
13	6579	3871	3433	3555	5535	2772	6329
14	6483	3709	3879	3573	5583	2867	6419
15	7481	4499	4447	4592	6435	3583	6534
16	8055	6221	5860	6024	7575	4419	6201
17	8515	9031	8095	9083	9352	5088	5874
18	7648	8393	7597	8381	8247	4368	5258
19	4910	5522	4893	5014	5701	2832	3787
20	2842	3095	2765	3183	3416	1697	2450
21	1967	2140	2059	2465	2559	1253	2020
22	1502	1863	1666	2184	2289	1444	2208
23	890	1414	1086	1519	1864	1355	1941

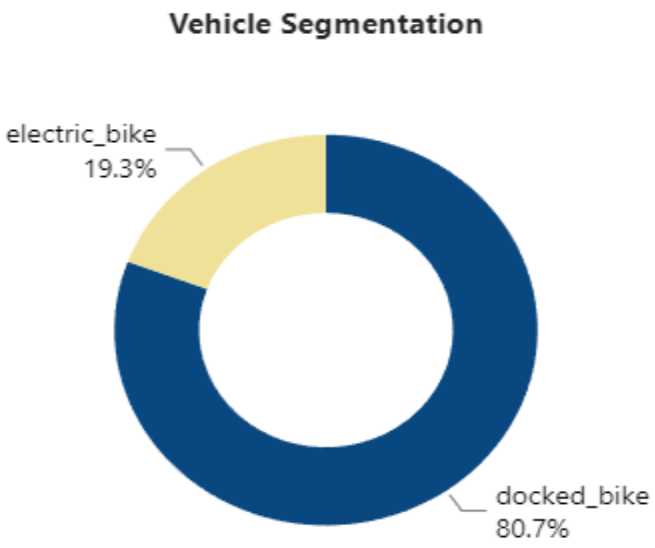
Popular Destinations:

1. Casual riders favor recreational locations like the Lake Shore, Grand Ave, and the Theater on the Lake. This aligns with leisure use.
2. They also more frequently visit coastal areas for sightseeing.
3. Annual members frequent major commuting spots like Clark/Elm, Erie/St. Clair, and Kinzie/Kingsbury.
4. This shows their primary end location is downtown where majority of offices are located.
5. There is one recreational spot, the Theater on the Lake, that is popular across members and casual riders.



Vehicle Segmentation:

1. Annual Riders use docked bikes for 81.51% times and electric bikes for 18.49% times.
2. Causal Riders use docked bikes for 79.63% times and electric bikes for 20.37% times.
3. This traditional docked bike is the most popular among both the annual members and casual riders.
4. Electric bikes have slightly higher adoption with annual members.
5. Safety concerns, price sensitivity, and terrain challenges may limit e-bike adoption.



Recommendations:

1. Offer reduced winter pricing and bundled monthly passes to smooth seasonal drops and incentivize casual riders to commit to annual plans.
2. Market weekend and full-day sightseeing passes to casual riders to balance weekday/weekend use and showcase membership benefits.
3. Expand electric bike availability near recreational hotspots and promote via social media to attract tech-savvy casual riders.

The recommendations focus on converting casual riders to annual members by incentivizing winter/off-peak usage, highlighting membership benefits for leisure trips, and targeting promotional efforts around electric bikes and recreational destinations.