bikeshare-analysis

August 30, 2022

1 Case Study: How Does a Bike-Share Navigate Speedy Success?

1.1 Guiding Questions of Analysis?

1. How do annual members and casual riders use Cyclistic bikes differently?

1.2 Data Gathering

The data gathered for this analysis is sourced from the **publicly open dataset** which is provided by the Company **Cyclistic** under the following **Data Licence Agreement**.

1.3 Data Cleaning

Start by Cleaning the data where there are any N/A values as well as rides that start and end at the exact coordinate values.

[3]: Rows drecreased by **1.507.429** from 5.901.463 to 4.394.034.

[4]: Check for matching observations (start and end stations should have the same amount of ID's and names)

Station Type	# of ID's	# of names	DIFF
Start	1130	1253	123
End	1158	1286	128

Since there is a clear difference in the numnber of names to stations there seems to be some kind of irregulatory, probably typos. Sincec we do not have an easy way to filter them out lets create a station dataframe with 1:1 id to name and only take the names with the largest number of appearances. afterwards we will replace the names in the cleaned dataframe with the stations that we found to be the most common names.

1.4 Data Preparation

- Adding formatted columns (hour of ride, weekday, month) to further slice data
- Calculating distance travlled (point-to-point)

1.4.1 Calculating the travel distane (Point-to-Point)

The haversine formula or haversine distance To calculate the distance on a sphere we can use the formula:

$$d = r * acos(sin(\Phi_1) * sin(\Phi_2) + cos(\Phi_1) * cos(\Phi_1) * cos(\Delta\lambda))$$

Where: r = radius of Sphere, Earth = ~6.371km Φ_1 = Latitude of start point Φ_2 = Latitude of end point $\Delta\lambda$ = Delta/Difference of longitude between end point and start point

Source: **movable-type.co.uk**

1.5 Data Analysis

Get a feel for the dataset with calculating some high-level statistics: * Average, Median, Max and Min for the ride time as well as the ride distance * most commonly observed time of day, weekday and month of ride

1.5.1 High Level Stats - 1

[12]:

The high-level statistics shows the following results: Number of observations: **4.394.034** Correlation between time and distance is: r = **0.09955**

Variable	Average	Median	Max	Min
ride length [t]	00:17:39	00:11:06	28 days 21:49:10	-1 days 21:50:55
ride distance [m]	2.221	1.667	1.189.522	0

Variable	Mode (most common)
hour of ride	17
day of ride	Saturday
month of ride	August
start station	Streeter Dr & Grand Ave
end station	Streeter Dr & Grand Ave

The data shows that there is a Min value of ride time that is negative, indicating that the end time was before the start time. The Min value of ride dist also shows 0 distance travelled (although that was supposed to be filtered out by same Lat + Long before).

Also there seems to be a stark outlier with over 1.189km travelled. On further inspection this was done with an electrik bike however the length of the bike ride was too short to make sense. Add new rule to filter out rides > 200km

Those data points are invalid and need to be filtered out before proceeding with more analysis.

1.5.2 Data Cleaning - Step 1

Filter out rows with the following conditios: * length of travel as timedelta < 0 (negative) OR * ride distane in m is <= 0 OR * ride distance in m is > 200.000

[14]:

Rows drecreased by **315** from 4.394.034 to 4.393.719.

1.5.3 High Level Stats - 2

[16]:

The high-level statistics shows the following results: Number of observations: 4.393.719 Correlation between time and distance is: r = 0.10411

Variable	Average	Median	Max	Min
ride length [t]	00:17:39	00:11:06	28 days 21:49:10	0 days 00:00:02
ride distance [m]	2.221	1.667	31.906	0

Time of ride	Mode (most common)
hour of ride	17
day of ride	Saturday
month of ride	August
start station	Streeter Dr & Grand Ave
end station	Streeter Dr & Grand Ave

Although there all rides with same start coordinates have been filtered out, there are still rides with no or very little distance travelled. Also there are rides that only last a few seconds. To get a beter understanding of "real" rides, e.g. rides that are used to travel somewhere and are not "accidentally" unlocked, a new rule will be applied: rides have to be farther than 100m and longer than 1 min.

1.5.4 Data Cleaning - Step 2

[18]:

Rows drecreased by **78.917** from 4.393.719 to 4.314.802.

In total Rows decreased by **1.586.661** from original dataset.

```
C:\Users\kemke\AppData\Local\Temp\ipykernel_9028\784846940.py:10:
SettingWithCopyWarning:


A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-
docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy
```

IMPROT DF for quick restart and PURGE OBSOLETE DATA

1.5.5 High Level Stats - 3

[23]:

The high-level statistics shows the following results: Number of observations: **4.314.802** Correlation between time and distance is: r = **0.10584**

Variable	Average	Median	Max	Min
ride length [t]	00:17:38	00:11:09	28 days 21:49:10	0 days 00:01:01
ride distance [m]	2.261	1.697	31.906	100

Time of ride	Mode (most common)
hour of ride	17
day of ride	Saturday
month of ride	August
start station	Streeter Dr & Grand Ave
end station	Streeter Dr & Grand Ave

1.6 Deep Dive - Statistics via Pivot

Taking a deeper look into the cleaned data and aggregating the data into different pivots

1.6.1 Pivot - Member-Type

[24]:
```
              count_of_rides        r  common_hour most_common_day  \
member_casual
Casual             1.753.742  0.07688           17        Saturday
Member             2.561.060  0.41390           17         Tuesday

              most_common_month  avg_time median_time  std_time  \
member_casual
Casual                   August  00:25:03    00:14:45  02:06:47
Member                   August  00:12:33    00:09:15  00:17:12

                      max_time  min_time  avg_dist  median_dist  std_dist  \
member_casual
Casual        28 days 21:49:10  00:01:01    2437.0       1884.0    1934.0
Member         1 days 00:52:55  00:01:01    2142.0       1582.0    1776.0

               max_dist  min_dist
member_casual
Casual          31906.0     100.0
Member          27488.0     100.0
```

1.6.2 Pivot - Rideable-Type

[25]:
```
              count_of_rides        r  common_hour most_common_day  avg_time  \
rideable_type
Classic Bike       2.863.190  0.29272           17        Saturday  00:16:47
Docked Bike          184.788  0.04627           16        Saturday  00:55:20
Electric Bike      1.266.824  0.46330           17        Thursday  00:14:03

              median_time  std_time         max_time  min_time  avg_dist  \
```

```
rideable_type
Classic Bike      00:11:12  00:29:30  1 days 00:59:25  00:01:01    2137.0
Docked Bike       00:25:51  06:15:34 28 days 21:49:10  00:01:10    2639.0
Electric Bike     00:10:00  00:14:44  0 days 07:59:55  00:01:01    2489.0

               median_dist  std_dist  max_dist  min_dist
rideable_type
Classic Bike        1608.0    1739.0   30186.0     102.0
Docked Bike         1973.0    2244.0   31906.0     102.0
Electric Bike       1899.0    1987.0   28741.0     100.0
```

1.6.3 Pivot - Month

[26]:

```
              count_of_rides        r  common_hour most_common_day  avg_time  \
month_of_ride
August               627.257  0.09377           17          Sunday  00:19:57
September            580.808  0.08451           17        Saturday  00:19:12
October              449.826  0.08579           17        Saturday  00:16:58
November             242.781  0.10864           17         Tuesday  00:13:35
December             167.342  0.06369           17        Thursday  00:13:45
January               76.473  0.05098           17        Thursday  00:12:36
February              84.707  0.11718           17          Monday  00:12:35
March                202.153  0.09013           17       Wednesday  00:16:07
April                255.224  0.22528           17        Saturday  00:15:37
May                  462.972  0.23225           17          Monday  00:18:34
June                 573.835  0.25961           17        Thursday  00:18:12
July                 591.424  0.15218           17        Saturday  00:18:22

              median_time  std_time         max_time  min_time  avg_dist  \
month_of_ride
August           00:12:37  01:33:06 28 days 21:49:10  00:01:01    2357.0
September        00:11:59  01:53:41 22 days 19:38:32  00:01:01    2331.0
October          00:10:33  01:40:30 28 days 06:25:01  00:01:01    2190.0
November         00:08:51  01:08:55 13 days 08:37:49  00:01:01    2002.0
December         00:08:34  02:10:13 21 days 02:40:33  00:01:01    1976.0
January          00:07:48  02:43:07 20 days 07:51:06  00:01:01    1780.0
February         00:08:04  00:58:38  7 days 13:45:58  00:01:01    1855.0
March            00:09:38  01:37:50 23 days 20:34:04  00:01:01    2134.0
April            00:09:48  00:36:39  5 days 05:45:58  00:01:01    2148.0
May              00:11:54  00:37:05  7 days 10:42:58  00:01:01    2343.0
June             00:12:08  00:32:01  4 days 15:12:52  00:01:01    2370.0
July             00:12:03  00:54:56 22 days 05:55:27  00:01:01    2381.0

               median_dist  std_dist  max_dist  min_dist
month_of_ride
August              1808.0    1878.0   31906.0     100.0
September           1769.0    1879.0   29389.0     100.0
```

```
October            1641.0   1792.0   28915.0    100.0
November           1492.0   1662.0   23446.0    101.0
December           1480.0   1642.0   22524.0    101.0
January            1325.0   1487.0   23263.0    102.0
February           1357.0   1579.0   24700.0    102.0
March              1581.0   1793.0   29734.0    102.0
April              1598.0   1792.0   21637.0    100.0
May                1760.0   1917.0   28036.0    101.0
June               1809.0   1897.0   30186.0    100.0
July               1802.0   1921.0   29685.0    101.0
```

1.6.4 Pivot - Member-Type & Month

[27]:

```
                            count_of_rides        r  common_hour  \
member_casual month_of_ride
Casual        August               308.612  0.07084           17
              September            266.430  0.06643           17
              October              173.455  0.06243           17
              November              64.296  0.07135           17
              December              41.098  0.05152           16
              January               11.539  0.04988           17
              February              13.738  0.06506           17
              March                 59.337  0.06383           17
              April                 82.006  0.16807           17
              May                  194.465  0.17729           17
              June                 261.320  0.20812           17
              July                 277.446  0.11493           17
Member        August               318.645  0.45973           17
              September            314.378  0.45343           17
              October              276.371  0.44390           17
              November             178.485  0.38527           17
              December             126.244  0.43902           17
              January               64.934  0.41840           17
              February              70.969  0.32736           17
              March                142.816  0.30966           17
              April                173.218  0.32382           17
              May                  268.507  0.41082           17
              June                 312.515  0.40993           17
              July                 313.978  0.45942           17

                            most_common_day  avg_time median_time  std_time  \
member_casual month_of_ride
Casual        August               Saturday  00:26:39    00:15:52  02:11:25
              September            Saturday  00:26:27    00:15:15  02:46:39
              October              Saturday  00:24:57    00:14:01  02:40:22
              November             Saturday  00:21:01    00:11:35  02:10:44
              December               Friday  00:23:41    00:11:14  04:21:22
```

	January	Saturday	00:26:08	00:09:55	06:58:20
	February	Monday	00:22:41	00:10:52	02:18:54
	March	Sunday	00:26:42	00:14:34	02:56:55
	April	Saturday	00:24:14	00:14:24	00:56:10
	May	Sunday	00:25:55	00:15:48	00:52:20
	June	Saturday	00:23:37	00:14:51	00:42:25
	July	Saturday	00:23:51	00:14:44	01:17:58
Member	August	Tuesday	00:13:27	00:10:08	00:15:51
	September	Thursday	00:13:03	00:09:48	00:16:10
	October	Saturday	00:11:58	00:08:54	00:15:18
	November	Tuesday	00:10:54	00:08:01	00:16:40
	December	Thursday	00:10:31	00:07:51	00:14:05
	January	Thursday	00:10:12	00:07:29	00:14:13
	February	Monday	00:10:38	00:07:37	00:18:36
	March	Wednesday	00:11:43	00:08:16	00:21:52
	April	Tuesday	00:11:33	00:08:17	00:20:51
	May	Monday	00:13:14	00:09:44	00:17:53
	June	Thursday	00:13:40	00:10:14	00:18:17
	July	Saturday	00:13:31	00:10:08	00:16:11

		max_time	min_time	avg_dist	median_dist \
member_casual	month_of_ride				
Casual	August	28 days 21:49:10	00:01:01	2483.0	1937.0
	September	22 days 19:38:32	00:01:01	2491.0	1933.0
	October	28 days 06:25:01	00:01:02	2390.0	1850.0
	November	13 days 08:37:49	00:01:01	2178.0	1666.0
	December	21 days 02:40:33	00:01:06	2116.0	1624.0
	January	20 days 07:51:06	00:01:05	1900.0	1495.0
	February	7 days 13:45:58	00:01:03	2083.0	1597.0
	March	23 days 20:34:04	00:01:03	2397.0	1828.0
	April	5 days 05:45:58	00:01:01	2419.0	1850.0
	May	7 days 10:42:58	00:01:01	2501.0	1917.0
	June	4 days 15:12:52	00:01:01	2457.0	1908.0
	July	22 days 05:55:27	00:01:01	2459.0	1894.0
Member	August	1 days 00:46:23	00:01:01	2235.0	1673.0
	September	0 days 21:58:24	00:01:01	2196.0	1632.0
	October	0 days 23:35:08	00:01:01	2065.0	1520.0
	November	1 days 00:20:34	00:01:01	1939.0	1409.0
	December	0 days 20:30:52	00:01:01	1931.0	1418.0
	January	0 days 23:00:00	00:01:01	1758.0	1289.0
	February	1 days 00:26:24	00:01:01	1811.0	1314.0
	March	1 days 00:38:29	00:01:01	2024.0	1479.0
	April	1 days 00:52:55	00:01:01	2020.0	1479.0
	May	1 days 00:35:38	00:01:01	2229.0	1641.0
	June	0 days 22:03:32	00:01:01	2297.0	1711.0
	July	1 days 00:36:42	00:01:01	2312.0	1714.0

member_casual	month_of_ride	std_dist	max_dist	min_dist
Casual	August	1937.0	31906.0	100.0
	September	1953.0	29389.0	102.0
	October	1883.0	28915.0	100.0
	November	1718.0	23446.0	102.0
	December	1690.0	22524.0	102.0
	January	1438.0	15667.0	102.0
	February	1707.0	24700.0	103.0
	March	1950.0	29734.0	102.0
	April	1952.0	21637.0	101.0
	May	2017.0	28036.0	101.0
	June	1941.0	30186.0	100.0
	July	1962.0	29685.0	102.0
Member	August	1810.0	26344.0	101.0
	September	1803.0	25227.0	100.0
	October	1720.0	23432.0	101.0
	November	1636.0	22073.0	101.0
	December	1624.0	20355.0	101.0
	January	1495.0	23263.0	102.0
	February	1549.0	21213.0	102.0
	March	1712.0	25216.0	102.0
	April	1696.0	21232.0	100.0
	May	1833.0	23343.0	101.0
	June	1855.0	24669.0	102.0
	July	1882.0	27488.0	101.0

1.6.5 Pivot - Member-Type & Day

[28]:

member_casual	day_of_ride	count_of_rides	r	common_hour	avg_time \
Casual	Monday	202.460	0.08057	17	00:25:30
	Tuesday	185.023	0.11569	17	00:21:31
	Wednesday	191.032	0.07499	17	00:21:15
	Thursday	215.972	0.06820	17	00:22:01
	Friday	240.037	0.07902	17	00:23:26
	Saturday	381.328	0.08328	15	00:27:47
	Sunday	337.890	0.06255	15	00:28:54
Member	Monday	360.537	0.42844	17	00:12:06
	Tuesday	403.168	0.42333	17	00:11:42
	Wednesday	401.858	0.41321	17	00:11:50
	Thursday	397.578	0.40522	17	00:12:01
	Friday	349.014	0.40803	17	00:12:13
	Saturday	337.896	0.40333	12	00:14:14
	Sunday	311.009	0.41339	15	00:14:17

median_time std_time max_time min_time \

member_casual	day_of_ride				
Casual	Monday	00:14:39	02:03:54	22 days 05:55:27	00:01:02
	Tuesday	00:12:40	01:15:14	10 days 18:02:58	00:01:01
	Wednesday	00:12:41	01:57:59	21 days 02:40:33	00:01:01
	Thursday	00:13:00	02:02:42	18 days 19:22:48	00:01:01
	Friday	00:13:55	01:52:18	15 days 17:09:54	00:01:01
	Saturday	00:16:55	02:12:40	28 days 06:25:01	00:01:01
	Sunday	00:17:08	02:37:23	28 days 21:49:10	00:01:02
Member	Monday	00:08:54	00:16:11	1 days 00:52:55	00:01:01
	Tuesday	00:08:45	00:16:00	1 days 00:22:02	00:01:01
	Wednesday	00:08:53	00:16:22	1 days 00:20:34	00:01:01
	Thursday	00:08:57	00:17:02	1 days 00:35:38	00:01:01
	Friday	00:09:03	00:16:52	1 days 00:15:18	00:01:01
	Saturday	00:10:28	00:19:26	0 days 23:37:03	00:01:01
	Sunday	00:10:23	00:18:39	1 days 00:46:23	00:01:01

		avg_dist	median_dist	std_dist	max_dist	min_dist
member_casual	day_of_ride					
Casual	Monday	2380.0	1828.0	1923.0	29038.0	100.0
	Tuesday	2319.0	1785.0	1843.0	28036.0	100.0
	Wednesday	2323.0	1807.0	1816.0	31906.0	100.0
	Thursday	2354.0	1827.0	1850.0	29389.0	101.0
	Friday	2359.0	1825.0	1846.0	30186.0	100.0
	Saturday	2565.0	1999.0	2019.0	29685.0	102.0
	Sunday	2561.0	1980.0	2050.0	29734.0	101.0
Member	Monday	2109.0	1538.0	1765.0	25227.0	101.0
	Tuesday	2090.0	1527.0	1752.0	23330.0	100.0
	Wednesday	2110.0	1554.0	1753.0	24669.0	101.0
	Thursday	2116.0	1555.0	1763.0	26344.0	101.0
	Friday	2092.0	1542.0	1739.0	27488.0	101.0
	Saturday	2260.0	1699.0	1827.0	25224.0	101.0
	Sunday	2253.0	1681.0	1838.0	25216.0	102.0

1.6.6 Pivot - Member-Type & Ride-Type & Month

[29]:

			count_of_rides	r \
member_casual	rideable_type	month_of_ride		
Casual	Classic Bike	August	209.947	0.24366
		September	179.184	0.22760
		October	96.925	0.22574
		November	29.143	0.20890
		December	17.891	0.18177
…			…	…
Member	Electric Bike	March	47.855	0.47299
		April	58.940	0.58292
		May	79.763	0.59476
		June	86.882	0.62357

```
                         July                 107.186  0.65341

                                              common_hour  avg_time median_time  \
member_casual rideable_type month_of_ride
Casual        Classic Bike  August                   17.0  00:23:29    00:15:24
                            September                17.0  00:23:06    00:14:53
                            October                  17.0  00:22:36    00:14:05
                            November                 15.0  00:19:50    00:12:05
                            December                 14.0  00:20:52    00:11:53
…                                                     …       …         …
Member        Electric Bike March                    17.0  00:10:31    00:07:47
                            April                    17.0  00:10:27    00:07:54
                            May                      17.0  00:12:08    00:09:15
                            June                     17.0  00:12:13    00:09:30
                            July                     17.0  00:12:00    00:09:23

                                              std_time        max_time  min_time  \
member_casual rideable_type month_of_ride
Casual        Classic Bike  August         00:36:59 1 days 00:54:07  00:01:01
                            September      00:39:39 1 days 00:57:57  00:01:01
                            October        00:41:09 1 days 00:49:22  00:01:02
                            November       00:39:43 0 days 23:14:55  00:01:01
                            December       00:46:24 1 days 00:12:59  00:01:06
…                                             …              …        …
Member        Electric Bike March          00:12:45 0 days 07:58:32  00:01:01
                            April          00:10:27 0 days 07:20:46  00:01:01
                            May            00:11:08 0 days 07:45:51  00:01:01
                            June           00:11:00 0 days 05:47:04  00:01:01
                            July           00:10:26 0 days 07:05:03  00:01:01

                                           avg_dist  median_dist  std_dist  \
member_casual rideable_type month_of_ride
Casual        Classic Bike  August           2379.0       1885.0    1804.0
                            September        2375.0       1871.0    1814.0
                            October          2254.0       1747.0    1745.0
                            November         2036.0       1582.0    1593.0
                            December         1951.0       1531.0    1573.0
…                                             …            …        …
Member        Electric Bike March            2276.0       1689.0    1872.0
                            April            2290.0       1714.0    1843.0
                            May              2508.0       1912.0    1996.0
                            June             2589.0       1986.0    2037.0
                            July             2550.0       1956.0    2007.0

                                           max_dist  min_dist
member_casual rideable_type month_of_ride
Casual        Classic Bike  August          28476.0     102.0
```

```
                    September       29389.0      102.0
                    October         28915.0      102.0
                    November        21024.0      102.0
                    December        22524.0      103.0
…                                      …       …
Member        Electric Bike March           25216.0      103.0
                    April           18663.0      100.0
                    May             23343.0      101.0
                    June            24669.0      103.0
                    July            26003.0      101.0

[72 rows x 13 columns]
```

1.6.7 Pivot - Member-Type & most used Station

[30]:

```
                                                                   no_of_rides
member_casual station_id   station_name
Casual        13022        Streeter Dr & Grand Ave                      101345
              LF-005       DuSable Lake Shore Dr & North Blvd            50795
              13300        DuSable Lake Shore Dr & Monroe St             48169
              13008        Millennium Park                               47821
              13042        Michigan Ave & Oak St                         45227
Member        KA1503000043 Kingsbury St & Kinzie St                      49100
              TA1308000050 Wells St & Concord Ln                         43748
              TA1307000039 Clark St & Elm St                             43500
              KA1504000135 Wells St & Elm St                             38061
              WL-012       Clinton St & Washington Blvd                  37214
```

1.6.8 Pivot - Member-Type & Month & most used Station

[31]:

```
no_of_rides
member_casual month_of_ride station_id   station_name
Casual        August        13022        Streeter Dr & Grand Ave
18915
                            LF-005       DuSable Lake Shore Dr & North Blvd
11584
              September     13022        Streeter Dr & Grand Ave
15179
                            LF-005       DuSable Lake Shore Dr & North Blvd
7954
              October       13022        Streeter Dr & Grand Ave
7980
                            13008        Millennium Park
5229
              November      13022        Streeter Dr & Grand Ave
2746
                            13300        DuSable Lake Shore Dr & Monroe St
```

1679
December 13022 Streeter Dr & Grand Ave
1683
13008 Millennium Park
1267
January TA1307000039 Clark St & Elm St
216
KA1504000135 Wells St & Elm St
197
February 13022 Streeter Dr & Grand Ave
355
13008 Millennium Park
250
March 13022 Streeter Dr & Grand Ave
3674
13300 DuSable Lake Shore Dr & Monroe St
2055
April 13022 Streeter Dr & Grand Ave
5002
13300 DuSable Lake Shore Dr & Monroe St
2480
May 13022 Streeter Dr & Grand Ave
12295
13300 DuSable Lake Shore Dr & Monroe St
6312
June 13022 Streeter Dr & Grand Ave
15281
LF-005 DuSable Lake Shore Dr & North Blvd
8984
July 13022 Streeter Dr & Grand Ave
18042
LF-005 DuSable Lake Shore Dr & North Blvd
10006
Member August LF-005 DuSable Lake Shore Dr & North Blvd
6629
TA1308000050 Wells St & Concord Ln
6343
September TA1308000050 Wells St & Concord Ln
5839
TA1307000039 Clark St & Elm St
5381
October KA1503000014 Ellis Ave & 60th St
5975
TA1307000039 Clark St & Elm St
5155
November KA1503000014 Ellis Ave & 60th St
4208

```
                             KA1503000043 Kingsbury St & Kinzie St
4056
              December       KA1503000043 Kingsbury St & Kinzie St
2922
                             WL-012       Clinton St & Washington Blvd
2427
              January        KA1503000043 Kingsbury St & Kinzie St
1780
                             TA1305000032 Clinton St & Madison St
1382
              February       KA1503000071 University Ave & 57th St
2010
                             KA1503000014 Ellis Ave & 60th St
1903
              March          KA1503000043 Kingsbury St & Kinzie St
3280
                             WL-012       Clinton St & Washington Blvd
2788
              April          KA1503000043 Kingsbury St & Kinzie St
3745
                             WL-012       Clinton St & Washington Blvd
3316
              May            KA1503000071 University Ave & 57th St
4925
                             KA1503000014 Ellis Ave & 60th St
4873
              June           LF-005       DuSable Lake Shore Dr & North Blvd
5696
                             KA1503000043 Kingsbury St & Kinzie St
5395
              July           LF-005       DuSable Lake Shore Dr & North Blvd
5760
                             KA1503000043 Kingsbury St & Kinzie St
5333
```

1.6.9 Pivot - Member-Type & Day & most used Station

[32]:
```
no_of_rides
member_casual day_of_ride station_id   station_name
Casual        Monday      13022        Streeter Dr & Grand Ave
13005
                          13008        Millennium Park
6585
                          LF-005       DuSable Lake Shore Dr & North Blvd
6156
              Tuesday     13022        Streeter Dr & Grand Ave
8295
```

		LF-005	DuSable Lake Shore Dr & North Blvd	5593
		13008	Millennium Park	4030
	Wednesday	13022	Streeter Dr & Grand Ave	8850
		LF-005	DuSable Lake Shore Dr & North Blvd	5692
		13042	Michigan Ave & Oak St	3858
	Thursday	13022	Streeter Dr & Grand Ave	10072
		LF-005	DuSable Lake Shore Dr & North Blvd	6440
		13008	Millennium Park	5010
	Friday	13022	Streeter Dr & Grand Ave	12921
		13008	Millennium Park	6440
		13300	DuSable Lake Shore Dr & Monroe St	5771
	Saturday	13022	Streeter Dr & Grand Ave	25955
		13300	DuSable Lake Shore Dr & Monroe St	13351
		13042	Michigan Ave & Oak St	11547
	Sunday	13022	Streeter Dr & Grand Ave	22247
		13300	DuSable Lake Shore Dr & Monroe St	11497
		LF-005	DuSable Lake Shore Dr & North Blvd	10715
Member	Monday	KA1503000043	Kingsbury St & Kinzie St	8045
		WL-012	Clinton St & Washington Blvd	6262
		TA1307000039	Clark St & Elm St	6184
	Tuesday	KA1503000043	Kingsbury St & Kinzie St	8751
		WL-012	Clinton St & Washington Blvd	8329
		TA1305000032	Clinton St & Madison St	7170
	Wednesday	KA1503000043	Kingsbury St & Kinzie St	

```
8332
                          WL-012       Clinton St & Washington Blvd
7960
                          TA1305000032 Clinton St & Madison St
6871
             Thursday     KA1503000043 Kingsbury St & Kinzie St
7638
                          WL-012       Clinton St & Washington Blvd
7544
                          TA1305000032 Clinton St & Madison St
6536
             Friday       KA1503000043 Kingsbury St & Kinzie St
6520
                          TA1308000050 Wells St & Concord Ln
5909
                          TA1307000039 Clark St & Elm St
5781
             Saturday     TA1308000050 Wells St & Concord Ln
7331
                          13179        Clark St & Lincoln Ave
6156
                          TA1307000039 Clark St & Elm St
5863
             Sunday       TA1308000050 Wells St & Concord Ln
6157
                          LF-005       DuSable Lake Shore Dr & North Blvd
6002
                          TA1308000001 Theater on the Lake
5561
```

1.6.10 Pivot - Member-Type & Ride-Type & most used Station

[33]:
```
no_of_rides
member_casual rideable_type station_id   station_name
Casual        Classic Bike  13022        Streeter Dr & Grand Ave
57559
                            LF-005       DuSable Lake Shore Dr & North Blvd
34245
                            13042        Michigan Ave & Oak St
27200
                            13300        DuSable Lake Shore Dr & Monroe St
27005
                            13008        Millennium Park
24830
              Docked Bike   13022        Streeter Dr & Grand Ave
23815
                            13008        Millennium Park
```

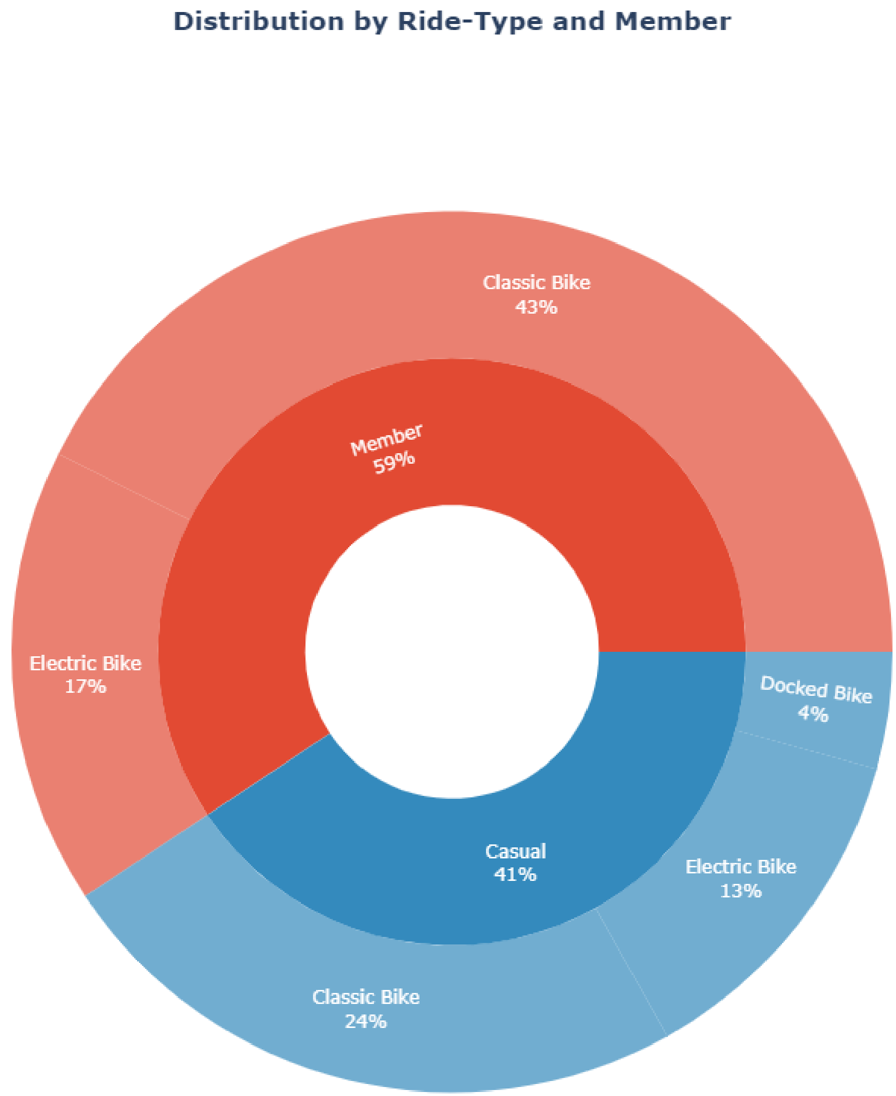
```
11906
                           13300        DuSable Lake Shore Dr & Monroe St
10975
                           15544        Shedd Aquarium
9138
                           13042        Michigan Ave & Oak St
8026
              Electric Bike 13022        Streeter Dr & Grand Ave
19971
                           13008        Millennium Park
11085
                           TA1308000050 Wells St & Concord Ln
10608
                           13300        DuSable Lake Shore Dr & Monroe St
10189
                           13042        Michigan Ave & Oak St
10001
Member        Classic Bike  KA1503000043 Kingsbury St & Kinzie St
34308
                           TA1307000039 Clark St & Elm St
31554
                           TA1308000050 Wells St & Concord Ln
31106
                           KA1503000014 Ellis Ave & 60th St
30064
                           KA1503000071 University Ave & 57th St
29644
              Electric Bike KA1503000043 Kingsbury St & Kinzie St
14792
                           WL-012       Clinton St & Washington Blvd
13432
                           TA1308000050 Wells St & Concord Ln
12642
                           TA1307000039 Clark St & Elm St
11946
                           TA1305000032 Clinton St & Madison St
11636
```

1.6.11 EXPORT PIVOTS TO EXCEL

1.7 Deep Dive - Visualization

Let’s first take a look at the distribution of rides used by member and casual riders as well as what type of ride they use

1.7.1 Distribution of rides



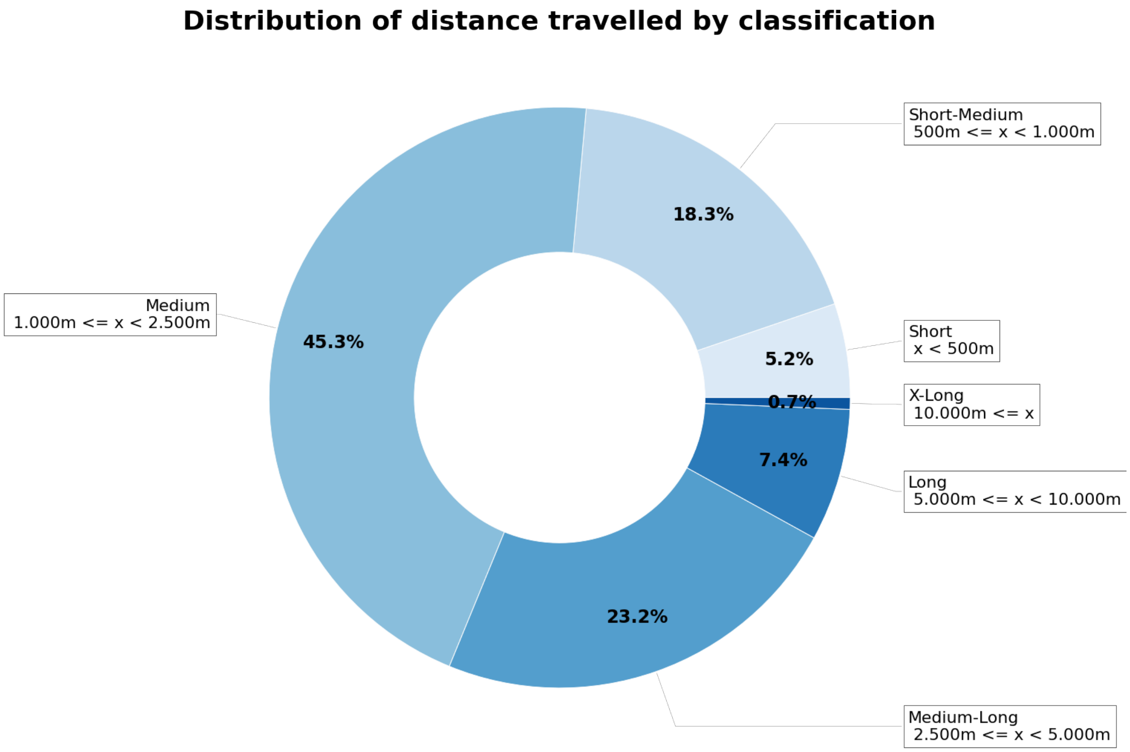
[37]:

Both parties seem to have a preference for the class bike instead of an electric bike. This preference is strongest for the members who prefer to use the classic bike for 71.7% of their rides vs casual riders who use it for 58.5% of their rides.

Also casual riders are the only ones who use the option of the docked bike.

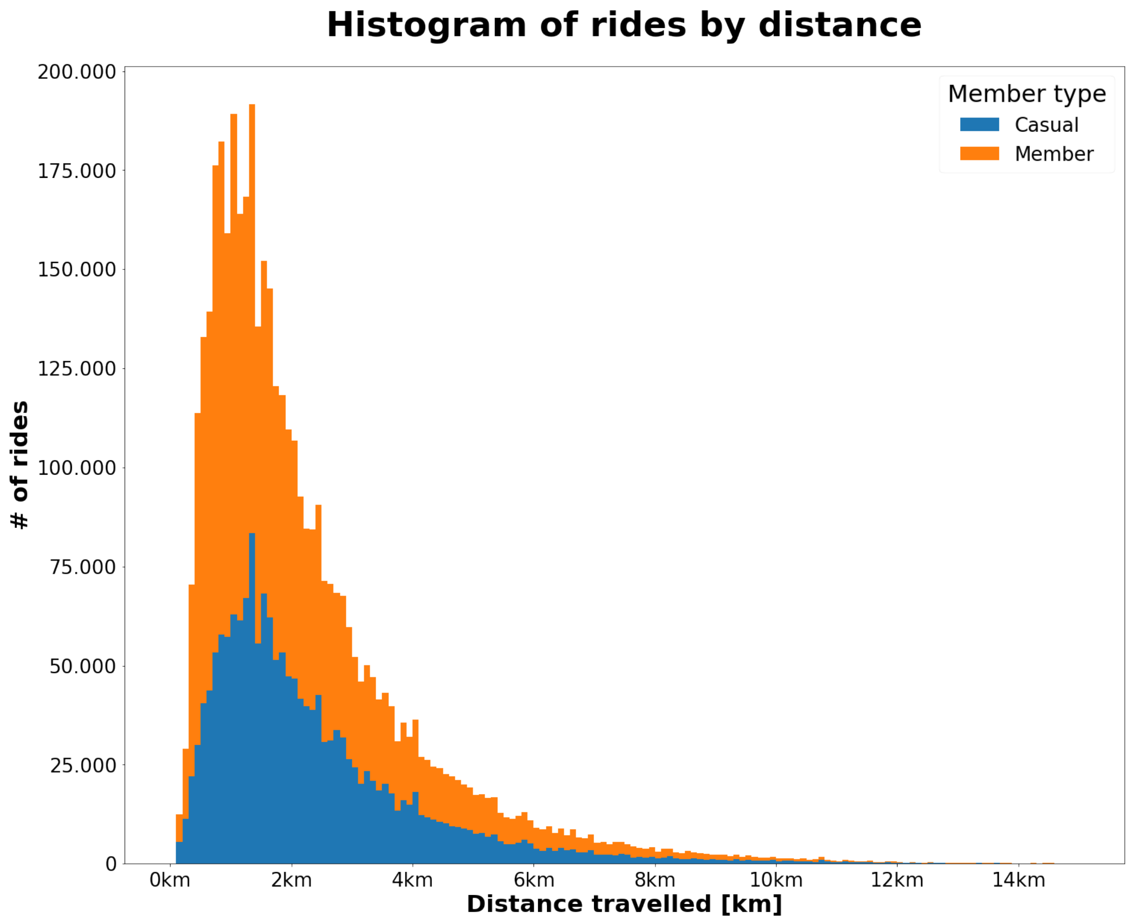
1.7.2 Ride distance

Next we will look at a quick distribution of the distance classified into 6 categories from Short to X-Long (see annotations).



But is there a difference between the two groups and their ride distance?

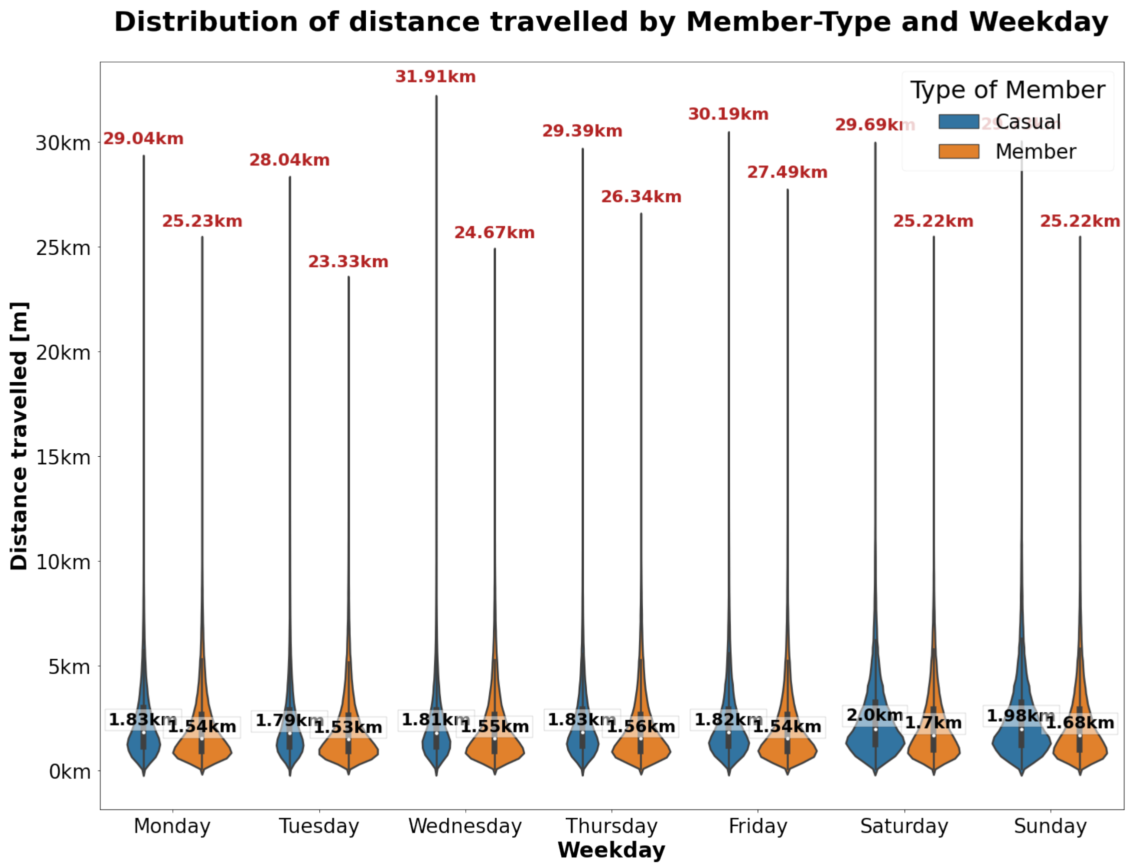
Histogram



The histogramm suggests that the casual riders ride slightly longer distances. Also it can be seen very well that the members use Cyclist more often.

But is there a difference in ride behaviour (number of rides, as well as distance) in the weekdays?

Violinplot for member-type and weekday

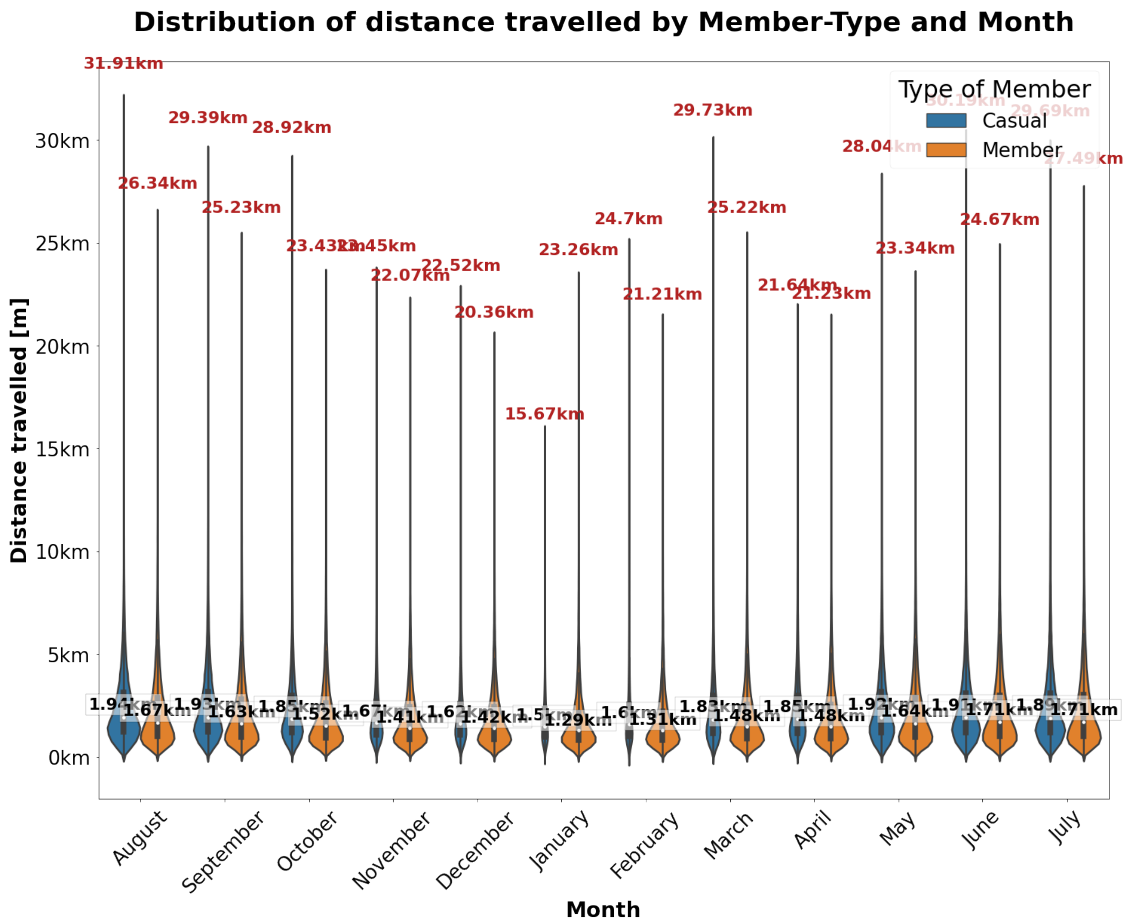


The violinplot of the ridedistance grouped by weekday and member type clearly shows that the number of rides for the casual members is significantly less in the workweek days (Mo-Fr) as shown by the slimmer widths of the violins. On weekends the casual riders even outweigh the members.

For both parties it also shows that the weekend comes with increased ride distances.

Next to the weekday is there also a difference when it comes to the months or the season?

Violinplot for member-type and month

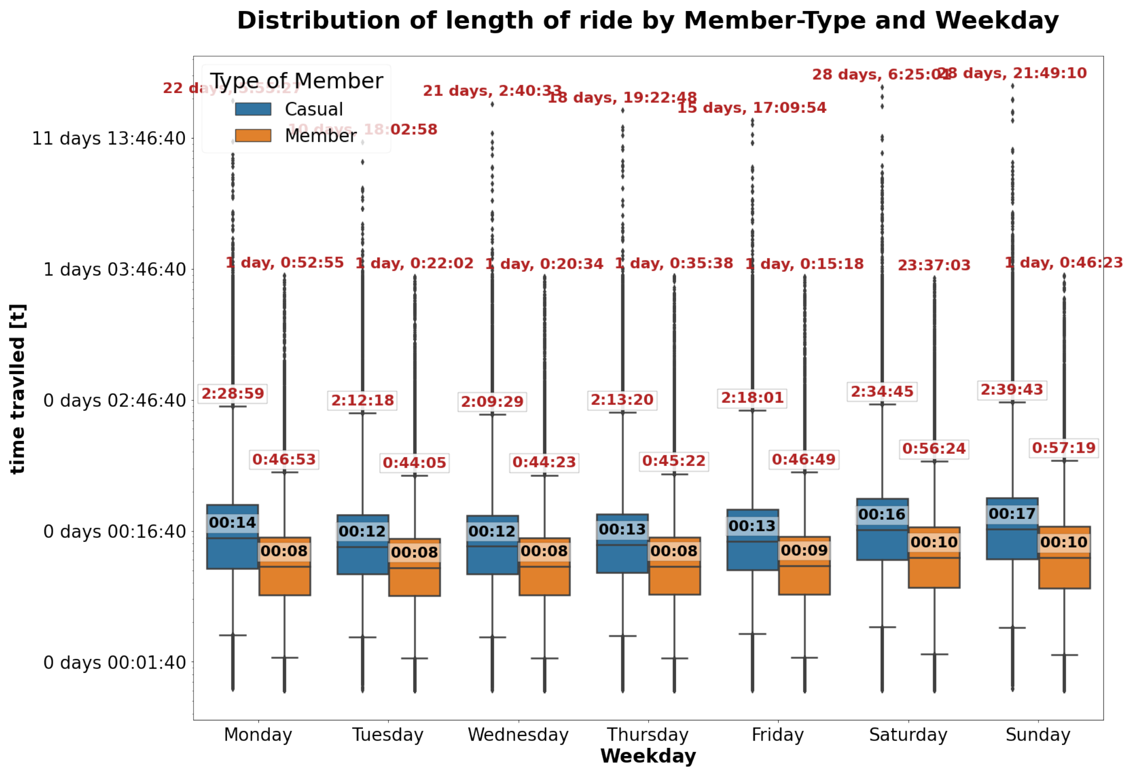


The above violinplot is similar to the one before with the only diffeerence that the weekdays are substituted by the months. Here we also see a drastic reduction in rides (width of violin) for the cold months starting from October and stretching to March. January and February are especially low volume for the casual riders whereas the members still ride almost as much as during the summers.

Also for both parties the distance travelled is shorter during the winter months.

1.7.3 Ride length [t]

Next let's look at the difference in riding behaviour for the length of the ride.

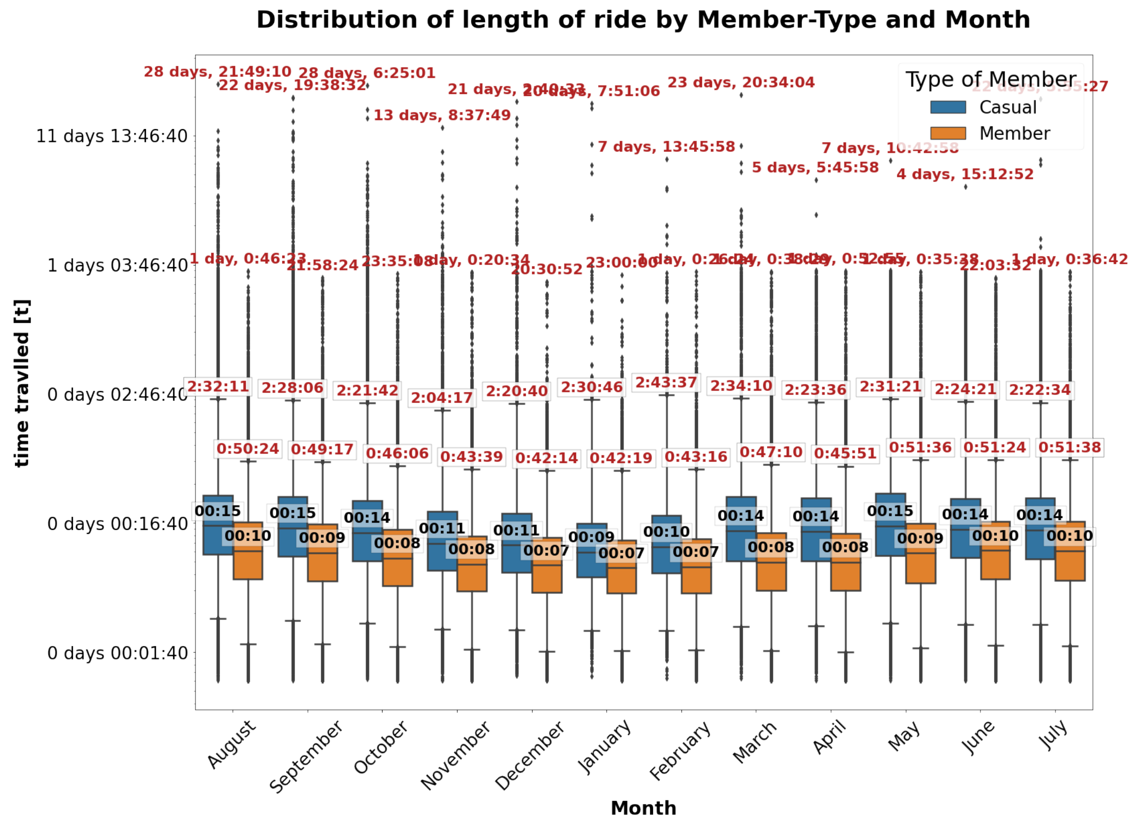


[44]: The above boxplot shows that the casual riders spent a significant time longer on the bike - by 54.4% to be exact (Casual: 00:14:25 vs Member: 00:09:20).

In the case of the ride length for Top 1% of rides (99th Percentile) this is even more drastic: Casual: 02:22:22 vs Member: 00:48:45.

So far this shows a clear difference in riding behaviour during the weekend, but is there also more differences throughout the year?

Boxplot - Ride-length by Member-Type and Month



[46]: The general difference in the ride length persisted throughout the year comparing the two groups. There is howevere a synchronous change in ride length throughout the year for both. During the winter months (Oct - Mar) the length of the ride drops by a significant ammount (ca. 30%) for both. The drop is slightly larger for the casual riders.

This seems to be innline with the previous findings that also the distance travelled is reduced in those months - however not by 30%. During the winter months the riders seem to hurry more to get to their destination. Whereas in the summer the motto seems to be "the ride is the goal" in Winter it shifts to "the destination is the goal".

1.7.4 Relation of distance and ride length

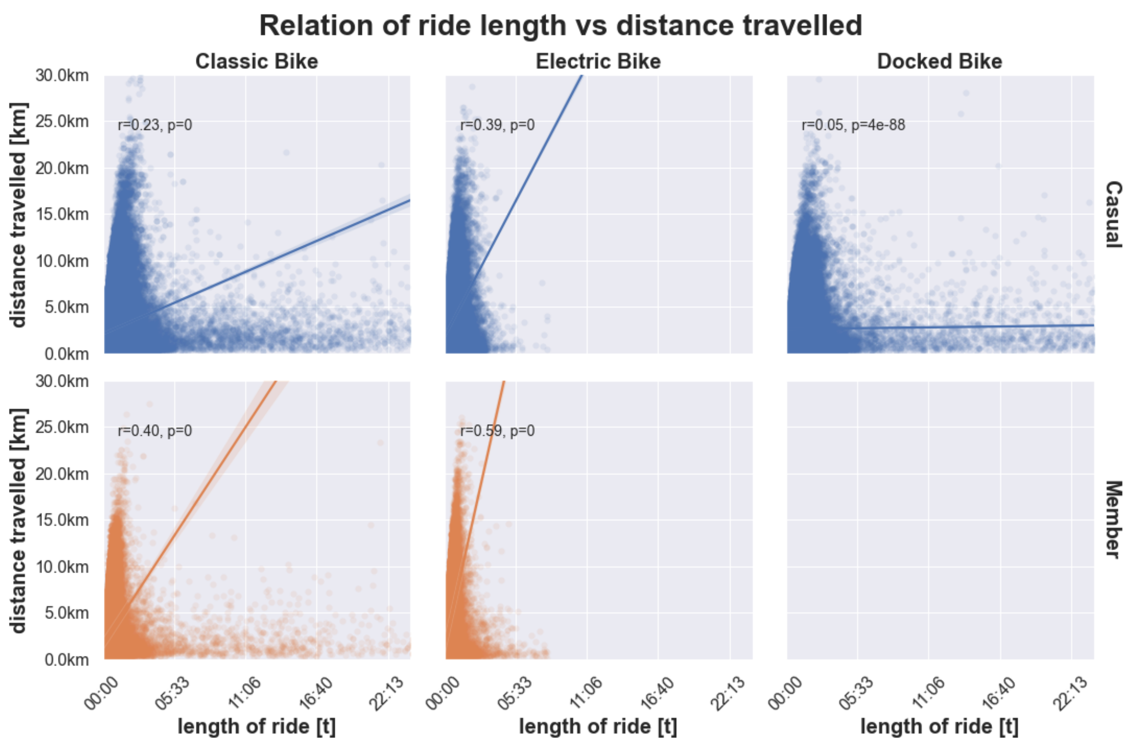
Now let's take a look at the relationship between the distance of the ride and the length. By now we know that the distance travelled is slightly longer for casual riders, but the time needed for this more than the distance would suggest.

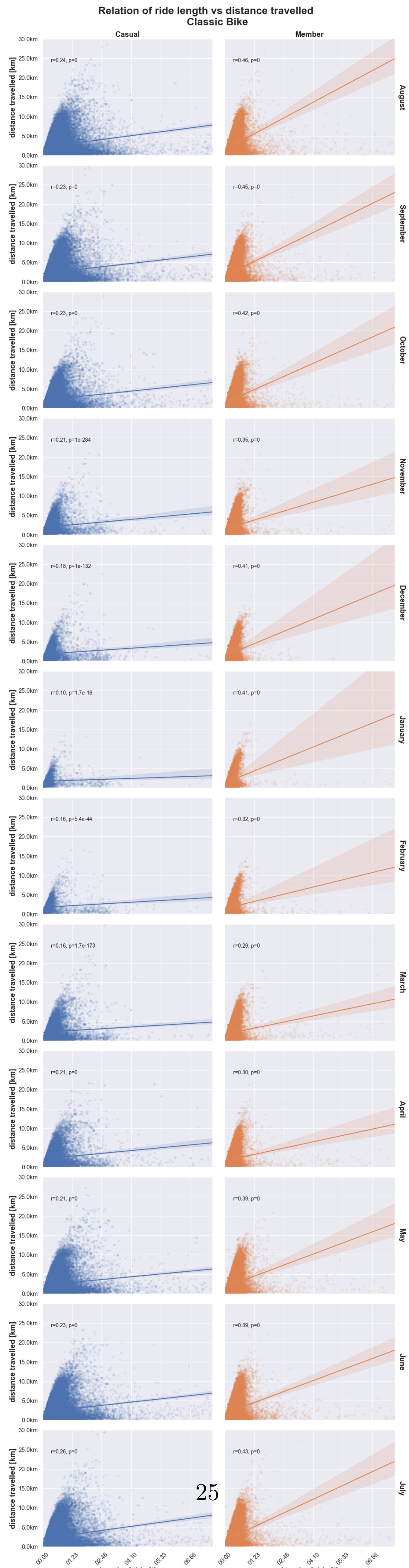
If there is a correlation between the ride distance and length of the ride, I expect it to be stronger for the members who travel almost as much as the casual riders in distance, but do it in a faster manner.

Also the docked bike has some gigantic outliers with the ride length (> 20 days). To get a good grasp lets take a look at the relationship of distance and ride length grouped by member-type and ride-type.

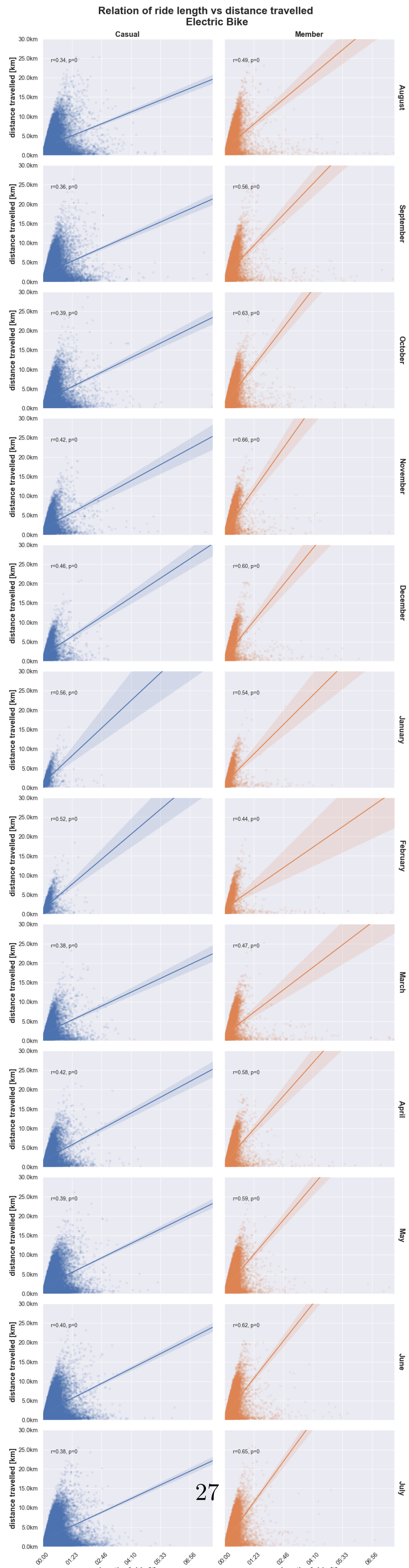
Linear Model of rides

```
C:\Users\kemke\AppData\Local\Temp\ipykernel_9028\1318206295.py:2:
SettingWithCopyWarning:


A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-
docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy
```





Relation of ride length vs distance travelled
Classic Bike
Casual
Member
August
September
October
November
December
January
February
March
April
May
June
July
r=0.24, p=0
r=0.46, p=0
r=0.23, p=0
r=0.45, p=0
r=0.23, p=0
r=0.42, p=0
r=0.21, p=1e-284
r=0.35, p=0
r=0.18, p=1e-132
r=0.41, p=0
r=0.10, p=1.7e-16
r=0.41, p=0
r=0.16, p=5.4e-44
r=0.32, p=0
r=0.16, p=1.7e-173
r=0.29, p=0
r=0.21, p=0
r=0.30, p=0
r=0.21, p=0
r=0.39, p=0
r=0.23, p=0
r=0.39, p=0
r=0.26, p=0
r=0.43, p=0
distance travelled [km]
length of ride [t]



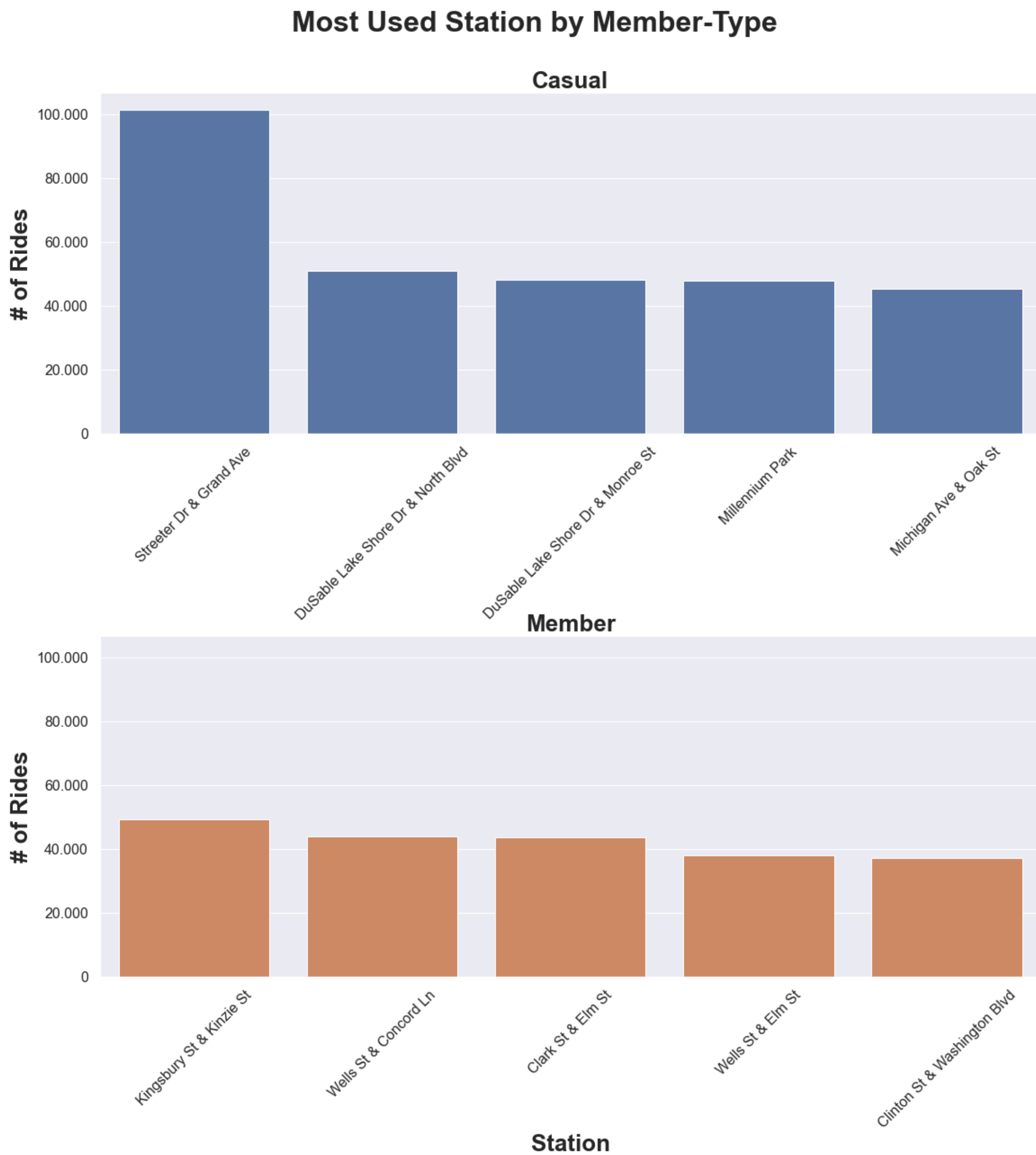
Relation of ride length vs distance travelled
Electric Bike
Casual
Member
August
September
October
November
December
January
February
March
April
May
June
July
r=0.34, p=0
r=0.49, p=0
r=0.36, p=0
r=0.56, p=0
r=0.39, p=0
r=0.63, p=0
r=0.42, p=0
r=0.66, p=0
r=0.46, p=0
r=0.60, p=0
r=0.56, p=0
r=0.54, p=0
r=0.52, p=0
r=0.44, p=0
r=0.38, p=0
r=0.47, p=0
r=0.42, p=0
r=0.58, p=0
r=0.39, p=0
r=0.59, p=0
r=0.40, p=0
r=0.62, p=0
r=0.38, p=0
r=0.65, p=0
distance travelled [km]
30.0km
25.0km
20.0km
15.0km
10.0km
5.0km
0.0km
00:00
01:23
02:46
04:10
05:33
06:56
length of ride [t]

1.7.5 Stations and routes used

Now let's take a look if there is a geographical difference as well. Let's look at the most used stations for this and then plot it into the map.

Barchart - Visited Stations

```
c:\Users\kemke\AppData\Local\Programs\Python\Python310\lib\site-
packages\seaborn\axisgrid.py:337: UserWarning:

The `size` parameter has been renamed to `height`; please update your code.

c:\Users\kemke\AppData\Local\Programs\Python\Python310\lib\site-
packages\seaborn\axisgrid.py:670: UserWarning:

Using the barplot function without specifying `order` is likely to produce an
incorrect plot.
```



Map Scatterplot - Visited Stations Now let's visualize the difference in used stations and plot it onto the map. The size of the marker will indicate how frequently the station is visited by each member-type.

Map Scatter- and Lineplot - Visited Stations and used routes Now let's visualize the difference in used stations and the typical routes taken by the riders. For this however the number of combinations for the routes (station to station) will be too big and the plot would end up quite confusing.

We will limit the lines plotted to visualize the routes taken therefore to the top 200 routes for each member type.

```
(259344, 4)
```

The plot shows a clear concentration of the casual riders near the lake side and in or between the several parks.

For the members there are several little hubs or networks. One which is near downtown and a big one near the university.

1.8 Conclusion

For the members the Cyclist bike seems to be more integrated into their daily life and used for work or similar things. They ride their bike throughout the year with little change and are not wasting time when they are riding.

In contrast the casual riders show a clear tendence towards riding as a leisure activity. They ride their bikes mostly on the weekens and during the hotter months, especially in summer. While they ride mostly in the parks they seem to stroll around which is why there is also a weaker correlation between the ride time and the distance of the ride. When winter hits, this however changes and the cold temperatures seem to shift the behaviour. For the few remaining casual riders the bike becomes more of a vehicle to get from point A to B instead of a long stroll through the new cihlly cold park.