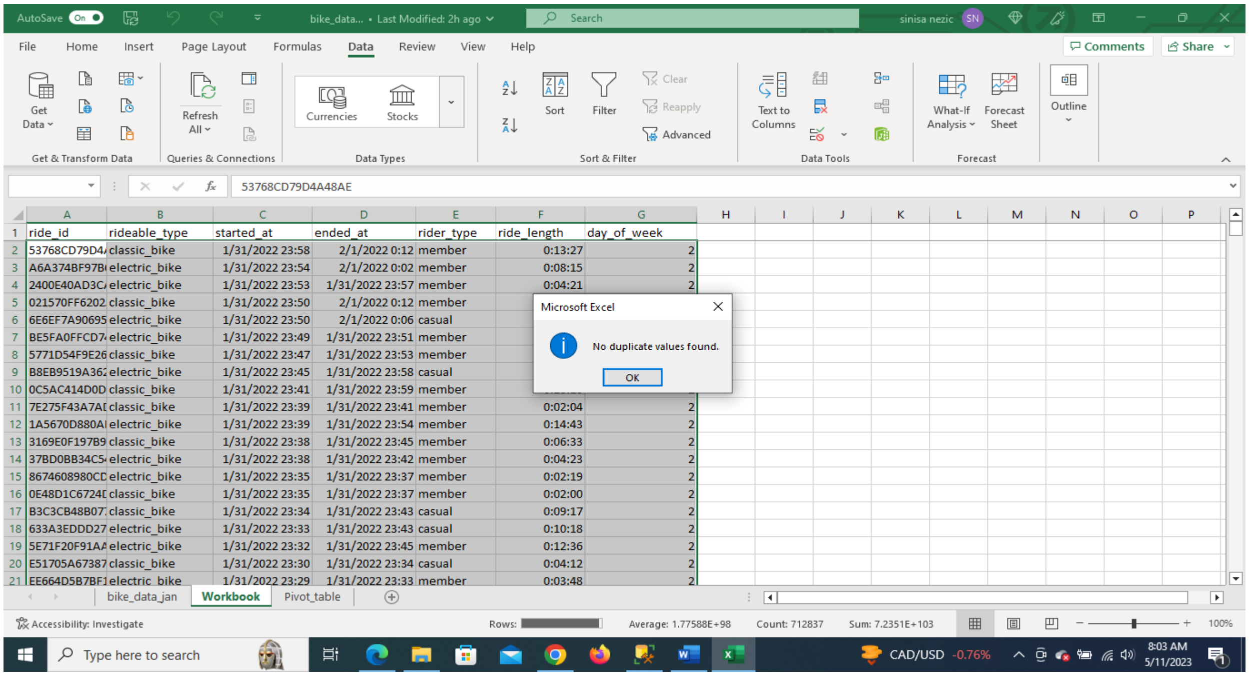
Exploratory Data Analysis in Microsoft Excel

- I created new column called “ride length” and set up values in “time” HH: MM: SS 37:30:55
- I created new column called “day of week” Where Sunday is 1 and Saturday is 7
- I deleted columns which I will not use for the analysis, making tables easy to read and understand.
- Using the function “find and replace”, I located all blanks in spreadsheet and deleted them.
- Calculated mean(average) and max values in column “ride length”.
- Calculated mode in column “day of week”, to discover the most frequently occurring value that appears in this column.
- Using Custom Filter function, I found outliers and deleted all the rows that are showing ride length <01.00 min and 24.00 h>
- Made pivot tables for each of twelve files to get initial insights of how these two types of riders use bikes differently.
- To be specific, for each month, I calculated average ride length, average ride length by day, and number of rides per day, and displayed results through charts to make it easier for stakeholders to understand.

Cleaning Process

Removing duplicates:



53768CD79D4A48AE

	A	B	C	D	E	F	G
1	ride_id	rideable_type	started_at	ended_at	rider_type	ride_length	day_of_week
2	53768CD79D4	classic_bike	1/31/2022 23:58	2/1/2022 0:12	member	0:13:27	2
3	A6A374BF97B	electric_bike	1/31/2022 23:54	2/1/2022 0:02	member	0:08:15	2
4	2400E40AD3C	electric_bike	1/31/2022 23:53	1/31/2022 23:57	member	0:04:21	2
5	021570FF6202	classic_bike	1/31/2022 23:50	2/1/2022 0:12	member		
6	6E6EF7A90695	electric_bike	1/31/2022 23:50	2/1/2022 0:06	casual		
7	BE5FA0FFCD7	electric_bike	1/31/2022 23:49	1/31/2022 23:51	member		
8	5771D54F9E26	classic_bike	1/31/2022 23:47	1/31/2022 23:53	member		
9	B8EB9519A36	electric_bike	1/31/2022 23:45	1/31/2022 23:58	casual		
10	0C5AC414D0D	classic_bike	1/31/2022 23:41	1/31/2022 23:59	member		
11	7E275F43A7A	classic_bike	1/31/2022 23:39	1/31/2022 23:41	member	0:02:04	2
12	1A5670D880A	electric_bike	1/31/2022 23:39	1/31/2022 23:54	member	0:14:43	2
13	3169E0F197B9	classic_bike	1/31/2022 23:38	1/31/2022 23:45	member	0:06:33	2
14	37BD0BB34C5	electric_bike	1/31/2022 23:38	1/31/2022 23:42	member	0:04:23	2
15	8674608980CD	electric_bike	1/31/2022 23:35	1/31/2022 23:37	member	0:02:19	2
16	0E48D1C6724	classic_bike	1/31/2022 23:35	1/31/2022 23:37	member	0:02:00	2
17	B3C3CB48B07	classic_bike	1/31/2022 23:34	1/31/2022 23:43	casual	0:09:17	2
18	633A3EDDD27	electric_bike	1/31/2022 23:33	1/31/2022 23:43	casual	0:10:18	2
19	5E71F20F91AA	electric_bike	1/31/2022 23:32	1/31/2022 23:45	member	0:12:36	2
20	E51705A67387	classic_bike	1/31/2022 23:30	1/31/2022 23:34	casual	0:04:12	2
21	EE664D5B7BF1	electric_bike	1/31/2022 23:29	1/31/2022 23:33	member	0:03:48	2

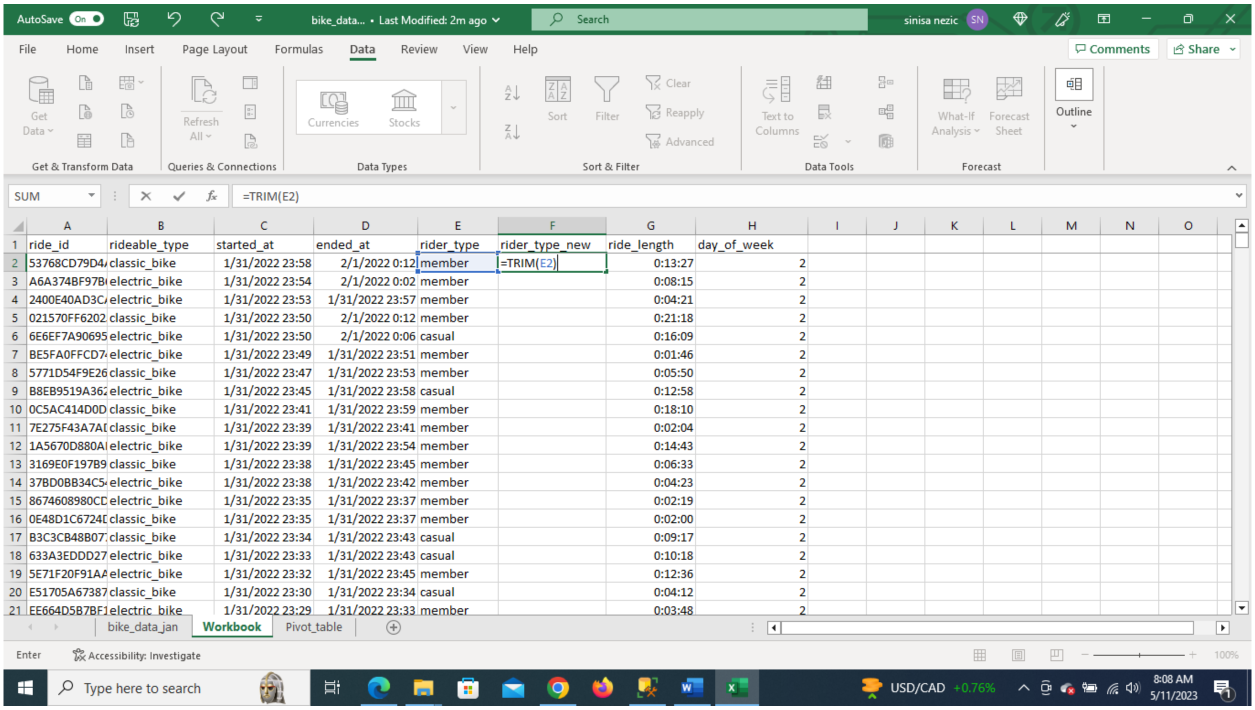
Microsoft Excel

No duplicate values found.

OK

Average: 1.77588E+98 Count: 712837 Sum: 7.2351E+103

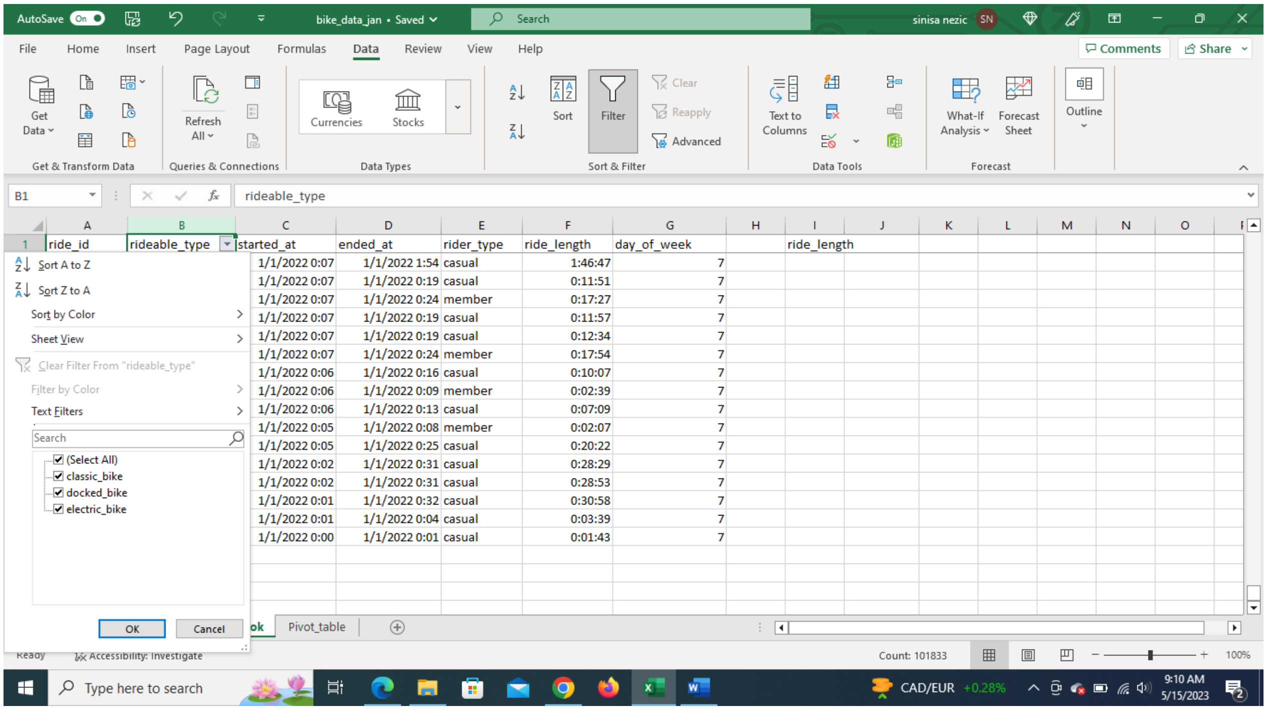
- Made sure that all the tables are consistent (column names etc.)
- Removed all the extra spaces using TRIM function. (Where applicable)



SUM =TRIM(E2)

	A	B	C	D	E	F	G	H
1	ride_id	rideable_type	started_at	ended_at	rider_type	rider_type_new	ride_length	day_of_week
2	53768CD79D4	classic_bike	1/31/2022 23:58	2/1/2022 0:12	member	=TRIM(E2)	0:13:27	2
3	A6A374BF97B	electric_bike	1/31/2022 23:54	2/1/2022 0:02	member		0:08:15	2
4	2400E40AD3C	electric_bike	1/31/2022 23:53	1/31/2022 23:57	member		0:04:21	2
5	021570FF6202	classic_bike	1/31/2022 23:50	2/1/2022 0:12	member		0:21:18	2
6	6E6EF7A90695	electric_bike	1/31/2022 23:50	2/1/2022 0:06	casual		0:16:09	2
7	BE5FA0FFCD7	electric_bike	1/31/2022 23:49	1/31/2022 23:51	member		0:01:46	2
8	5771D54F9E26	classic_bike	1/31/2022 23:47	1/31/2022 23:53	member		0:05:50	2
9	B8EB9519A36	electric_bike	1/31/2022 23:45	1/31/2022 23:58	casual		0:12:58	2
10	0C5AC414D0D	classic_bike	1/31/2022 23:41	1/31/2022 23:59	member		0:18:10	2
11	7E275F43A7A	classic_bike	1/31/2022 23:39	1/31/2022 23:41	member		0:02:04	2
12	1A5670D880A	electric_bike	1/31/2022 23:39	1/31/2022 23:54	member		0:14:43	2
13	3169E0F197B9	classic_bike	1/31/2022 23:38	1/31/2022 23:45	member		0:06:33	2
14	37BD0BB34C5	electric_bike	1/31/2022 23:38	1/31/2022 23:42	member		0:04:23	2
15	8674608980CD	electric_bike	1/31/2022 23:35	1/31/2022 23:37	member		0:02:19	2
16	0E48D1C6724	classic_bike	1/31/2022 23:35	1/31/2022 23:37	member		0:02:00	2
17	B3C3CB48B07	classic_bike	1/31/2022 23:34	1/31/2022 23:43	casual		0:09:17	2
18	633A3EDDD27	electric_bike	1/31/2022 23:33	1/31/2022 23:43	casual		0:10:18	2
19	5E71F20F91AA	electric_bike	1/31/2022 23:32	1/31/2022 23:45	member		0:12:36	2
20	E51705A67387	classic_bike	1/31/2022 23:30	1/31/2022 23:34	casual		0:04:12	2
21	EE664D5B7BF1	electric_bike	1/31/2022 23:29	1/31/2022 23:33	member		0:03:48	2

- Made sure that all dates are in the same format.
- Used FILTER function to be sure that there is not any error or unusual values.

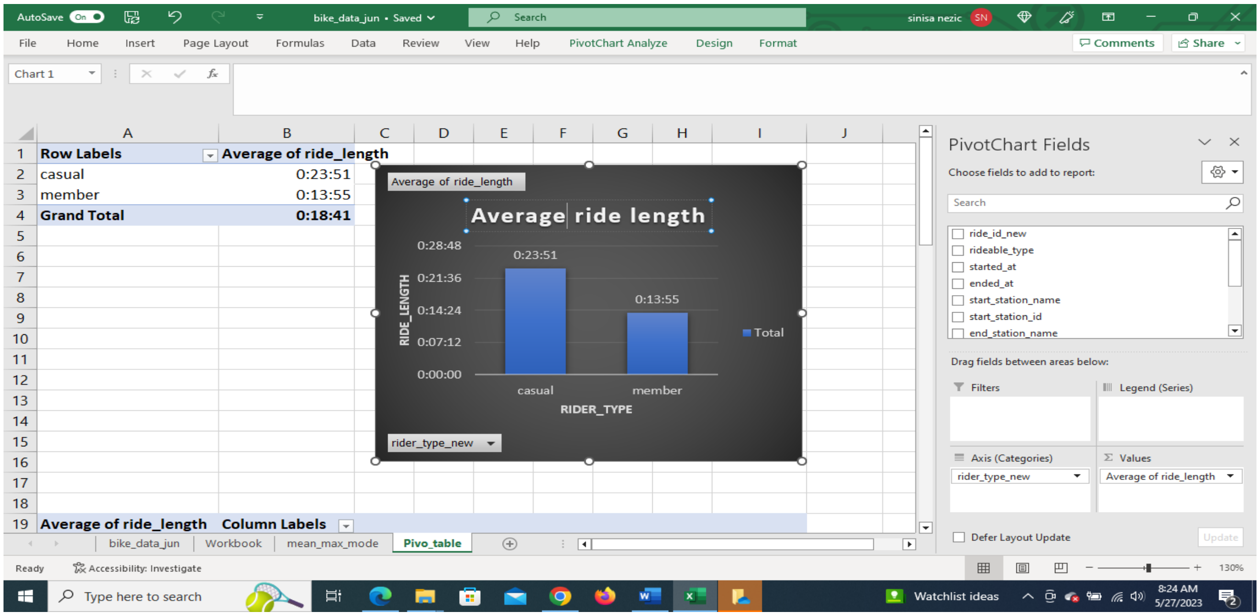


B1 rideable_type

☑ (Select All)
☑ classic_bike
☑ docked_bike
☑ electric_bike

ride_id	rideable_type	started_at	ended_at	rider_type	ride_length	day_of_week		ride_length
		1/1/2022 0:07	1/1/2022 1:54	casual	1:46:47	7		
		1/1/2022 0:07	1/1/2022 0:19	casual	0:11:51	7		
		1/1/2022 0:07	1/1/2022 0:24	member	0:17:27	7		
		1/1/2022 0:07	1/1/2022 0:19	casual	0:11:57	7		
		1/1/2022 0:07	1/1/2022 0:19	casual	0:12:34	7		
		1/1/2022 0:07	1/1/2022 0:24	member	0:17:54	7		
		1/1/2022 0:06	1/1/2022 0:16	casual	0:10:07	7		
		1/1/2022 0:06	1/1/2022 0:09	member	0:02:39	7		
		1/1/2022 0:06	1/1/2022 0:13	casual	0:07:09	7		
		1/1/2022 0:05	1/1/2022 0:08	member	0:02:07	7		
		1/1/2022 0:05	1/1/2022 0:25	casual	0:20:22	7		
		1/1/2022 0:02	1/1/2022 0:31	casual	0:28:29	7		
		1/1/2022 0:02	1/1/2022 0:31	casual	0:28:53	7		
		1/1/2022 0:01	1/1/2022 0:32	casual	0:30:58	7		
		1/1/2022 0:01	1/1/2022 0:04	casual	0:03:39	7		
		1/1/2022 0:00	1/1/2022 0:01	casual	0:01:43	7		

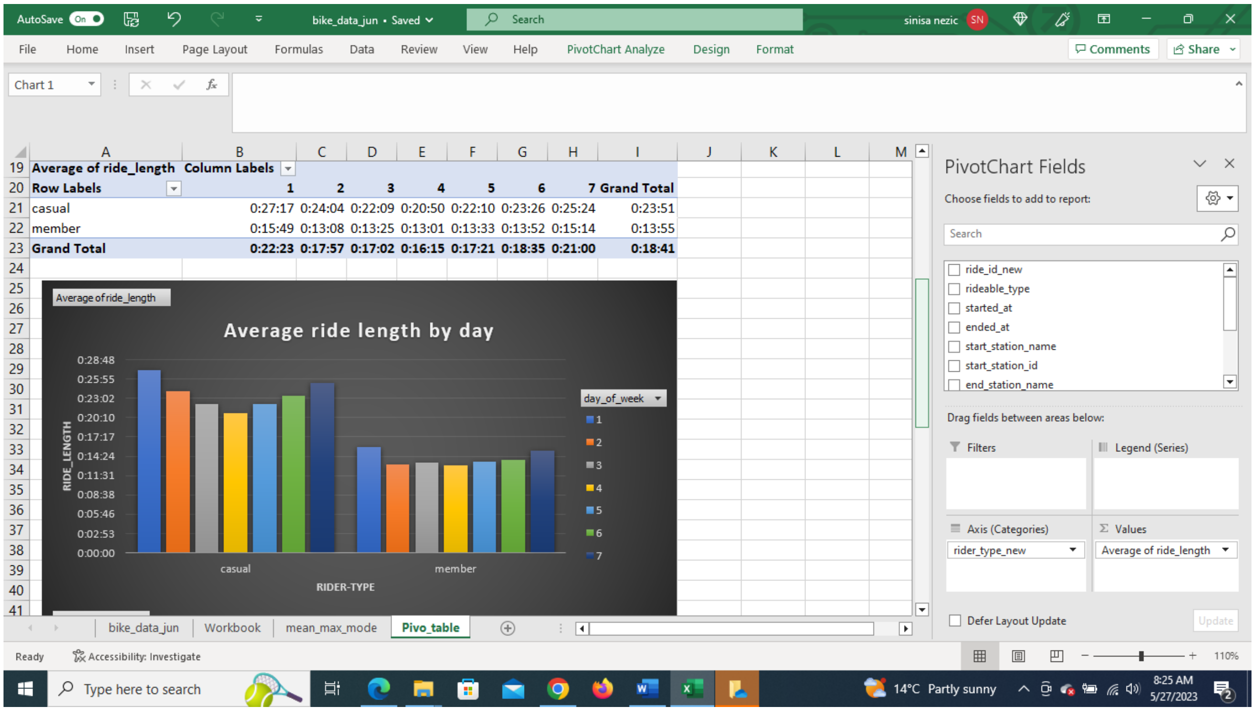
Average ride length:



Row Labels	Average of ride_length
casual	0:23:51
member	0:13:55
Grand Total	**0:18:41**

Average of ride_length Column Labels

Average ride length by day (Where #1 is Sunday, #7 is Saturday)

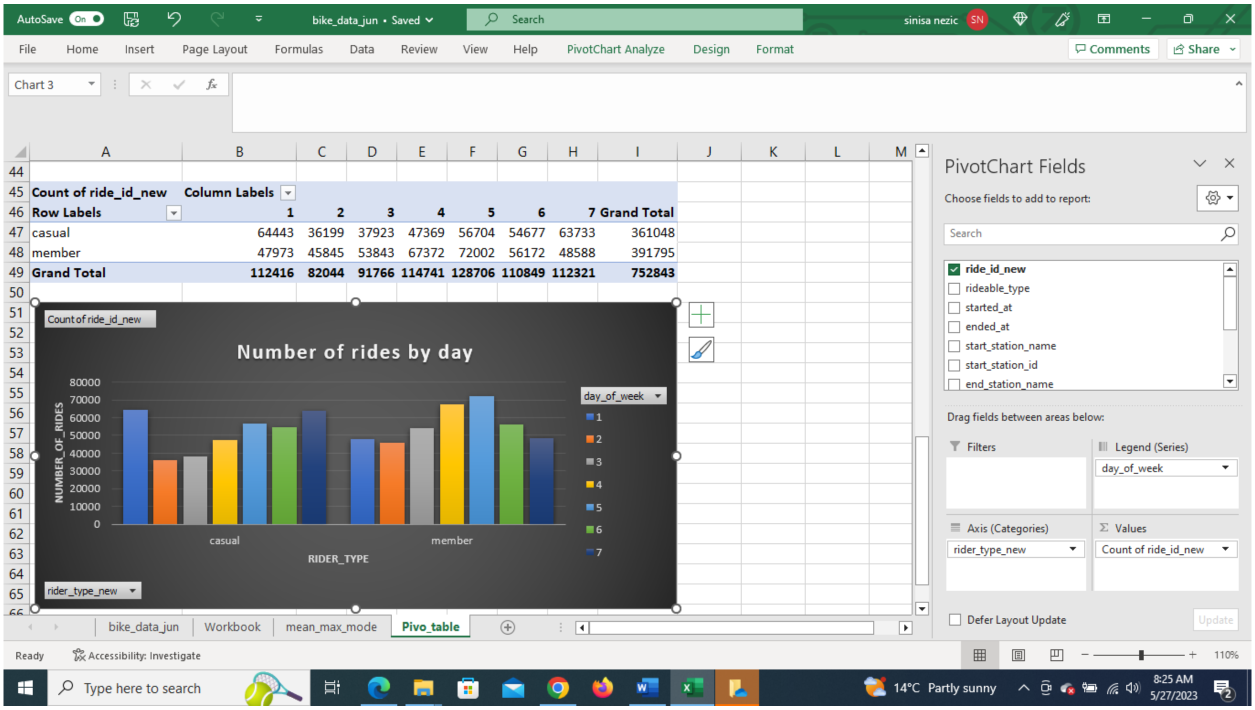


Average of ride_length	Column Labels							
Row Labels	1	2	3	4	5	6	7	Grand Total
casual	0:27:17	0:24:04	0:22:09	0:20:50	0:22:10	0:23:26	0:25:24	0:23:51
member	0:15:49	0:13:08	0:13:25	0:13:01	0:13:33	0:13:52	0:15:14	0:13:55
Grand Total	0:22:23	0:17:57	0:17:02	0:16:15	0:17:21	0:18:35	0:21:00	0:18:41

Axis (Categories): rider_type_new

Values: Average of ride_length

Number of rides per day (Where #1 is Sunday, #7 is Saturday)



Count of ride_id_new	Column Labels							
Row Labels	1	2	3	4	5	6	7	Grand Total
casual	64443	36199	37923	47369	56704	54677	63733	361048
member	47973	45845	53843	67372	72002	56172	48588	391795
Grand Total	112416	82044	91766	114741	128706	110849	112321	752843

Legend (Series): day_of_week

Axis (Categories): rider_type_new

Values: Count of ride_id_new