Chicago DIVVY Bike Sharing

July 2021-August 2022

Data Visualization

Başak Buluttekin - Matthieu Baillard

Ironhack Paris - Data Analytics - Project #4 - May 8, 2023

Overview

- Dataset description
- Challenges
- Process
- Learnings
- Improvements
- Highlights

Dataset

Chicago Divvy Bike Sharing, from Aug. 2021 to Jul. 2022

- 3.2M rides
  - Bike type (electric / classic / docked)
  - Start & end times (datetime)
  - Ride length (i.e. duration)
  - Year, month, Year-month
  - Start & end latitude / longitude, station names & ID
  - Client type (member / casual)

Credit: SUMC.

Challenges

- Dates to Days and Hours
- Long. / Lat. data lacking precision

(unable to convert to actual coordinates)

Process - Data Preparation

```
#Convert start and end columns to datetime format

bike['started_at'] = pd.to_datetime(bike['started_at'])
bike['ended_at'] = pd.to_datetime(bike['ended_at'])

#Add day of the week and hour of the day columns

bike['start_days'] = bike['started_at'].dt.day_name()
bike['hour'] = bike['started_at'].dt.hour
```

Process - Pivot Tables

rideable_type	classic_bike	docked_bike	electric_bike
Year-Month			
2021-08-01	295698	19359	153612
2021-09-01	257308	14259	157601
2021-10-01	169768	9095	162782
2021-11-01	74655	3018	102527
2021-12-01	47994	1817	68906
2022-01-01	25451	375	22464
2022-02-01	27520	456	25933
2022-03-01	64303	2889	70220
2022-04-01	83519	4271	99398
2022-05-01	173357	9662	161202
2022-06-01	232150	11930	195639
2022-07-01	211670	11926	263153

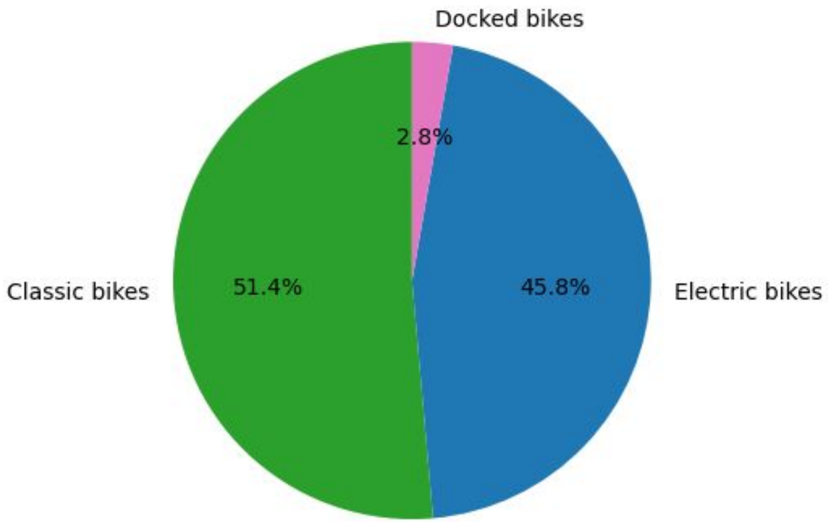
member_casual	casual	member
Year-Month		
2021-08-01	223790	244879
2021-09-01	195120	234048
2021-10-01	130931	210714
2021-11-01	51269	128931
2021-12-01	31923	86794
2022-01-01	8339	39951
2022-02-01	9504	44405
2022-03-01	39982	97430
2022-04-01	58857	128331
2022-05-01	141478	202743
2022-06-01	198548	241171
2022-07-01	229829	256920

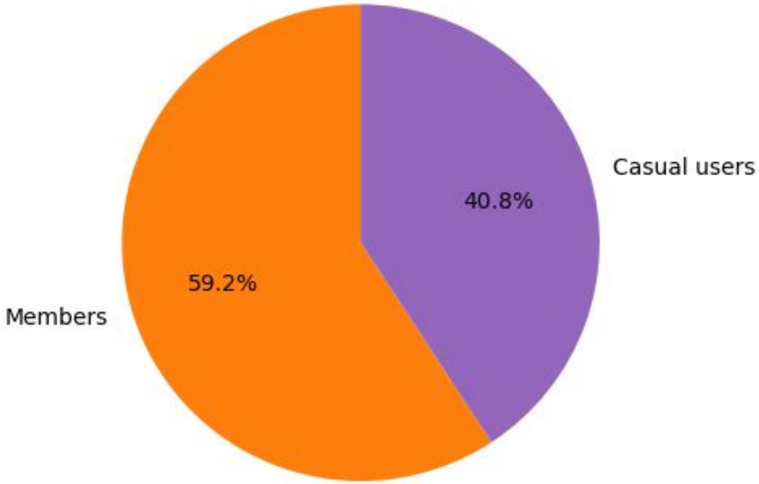
start_days	Friday	Monday	Saturday	Sunday	Thursday	Tuesday	Wednesday
member_casual	member	member	member	member	member	member	member
hour							
0	2856	1780	5070	6027	2051	1423	1627
1	1636	1052	3956	4202	934	714	784
2	845	590	2241	2647	448	410	401
3	479	399	1316	1486	367	293	284
4	720	691	793	1061	667	695	678
5	2952	3151	1167	1172	3394	3765	3770
6	8144	8513	2798	2445	9311	10381	10085
7	14155	15610	5446	4078	18579	20420	19757
8	15675	17818	8909	6456	21482	22459	22364
9	11130	10904	12738	10451	12929	12638	12765
10	10063	9554	15625	14505	10144	9696	9885
11	12617	11869	18016	16992	12036	11712	11406
12	15196	13814	19178	18672	14568	14384	14128
13	15089	13613	18978	18289	13819	13847	13721
14	15233	13406	18712	18360	13477	13078	13372
15	17907	16613	18753	19125	17552	17416	16730
16	22663	24505	18564	18888	25814	26803	26150
17	25796	31726	18009	17765	33946	36325	35302
18	22007	25043	17029	16118	27507	28035	28411
19	16249	17546	13996	12555	19632	19204	19542
20	11473	11809	10405	9059	13800	12911	13739
21	8702	8360	8883	6943	11361	9864	10391
22	7709	5473	8332	5423	8820	6631	7360
23	6391	2957	7433	3368	5317	3501	4188

```
pandas.cut(x, bins, right=True, labels=None, retbins=False, precision=3,
include_lowest=False, duplicates='raise', ordered=True)    [source]
```

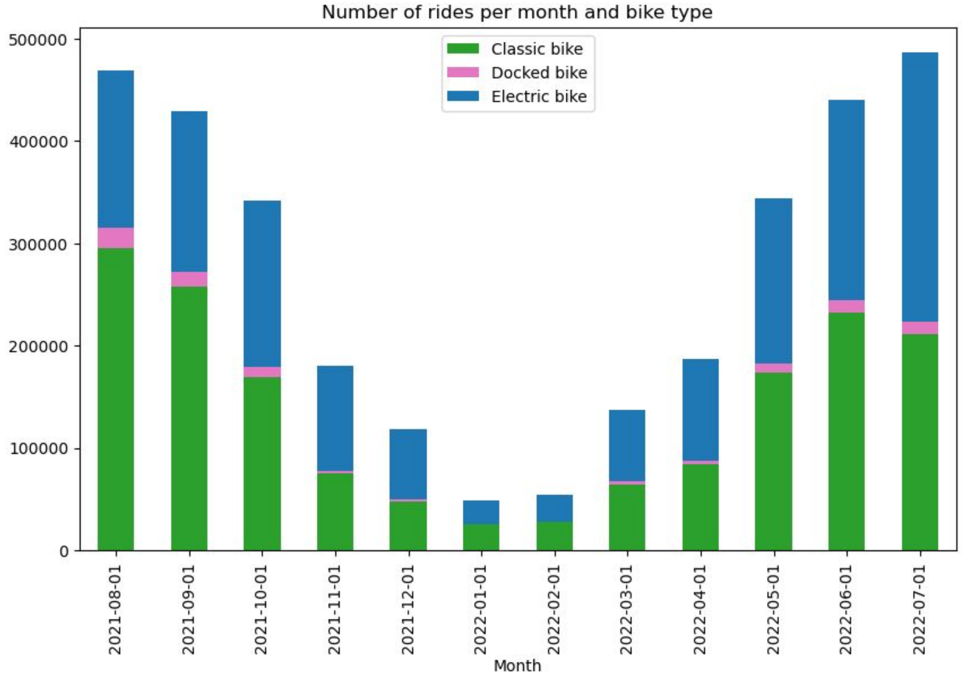
start_days	Monday	Tuesday	Wednesday	Thursday	Friday	Saturday	Sunday
hour_bin_label							
0-1	2832	2137	2411	2985	4492	9026	10229
2-3	989	703	685	815	1324	3557	4133
4-5	3842	4460	4448	4061	3672	1960	2233
6-7	24123	30801	29842	27890	22299	8244	6523
8-9	28722	35097	35129	34411	26805	21647	16907
10-11	21423	21408	21291	22180	22680	33641	31497
12-13	27427	28231	27849	28387	30285	38156	36961
14-15	30019	30494	30102	31029	33140	37465	37485
16-17	56231	63128	61452	59760	48459	36573	36653
18-19	42589	47239	47953	47139	38256	31025	28673
20-21	20169	22775	24130	25161	20175	19288	16002
22-23	8430	10132	11548	14137	14100	15765	8791

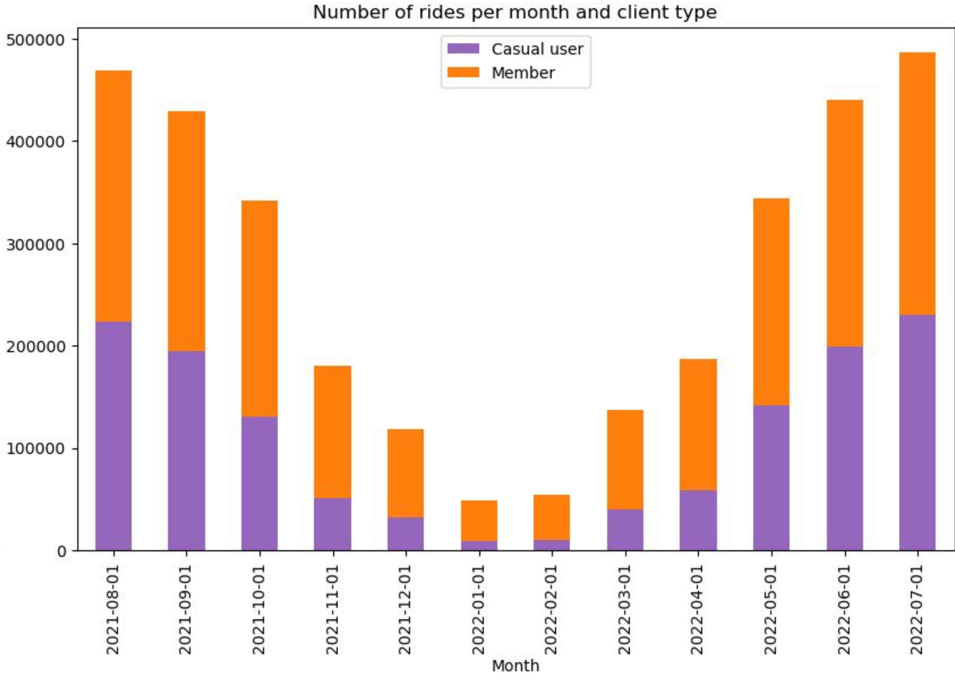
Rides per Bike and User Types



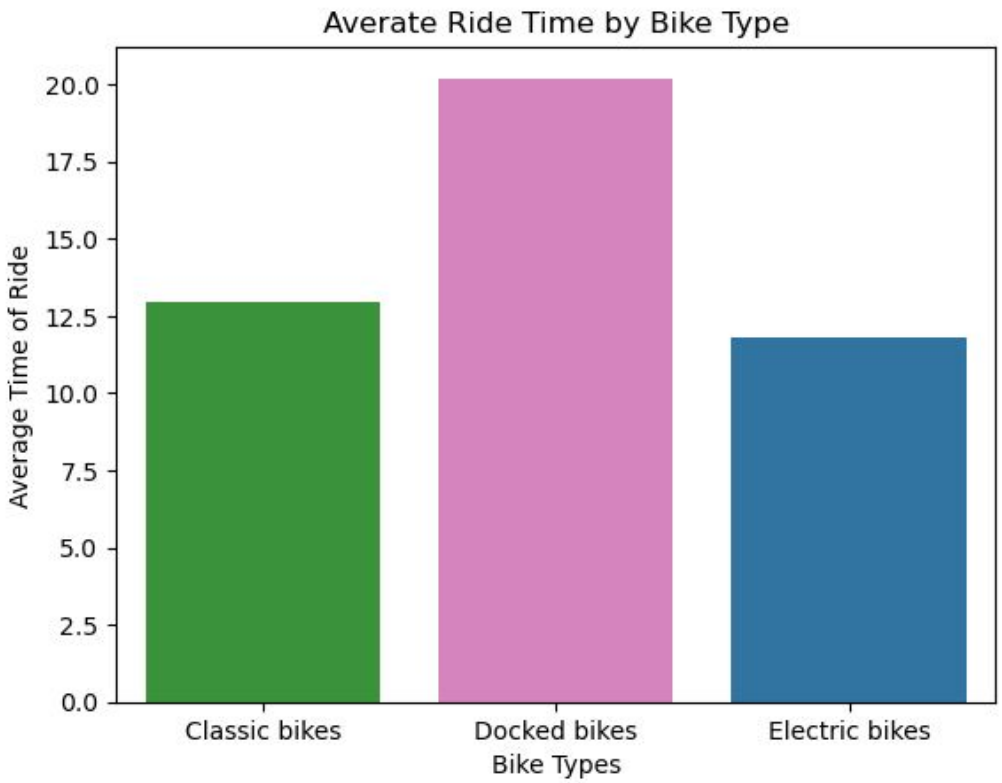


Seasonality per Bike and User Types

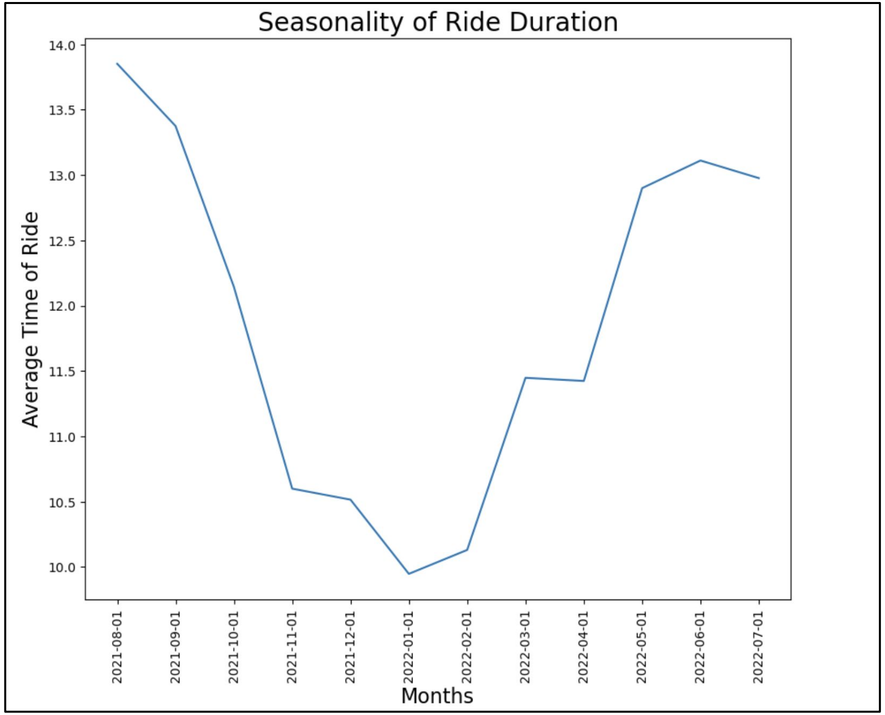


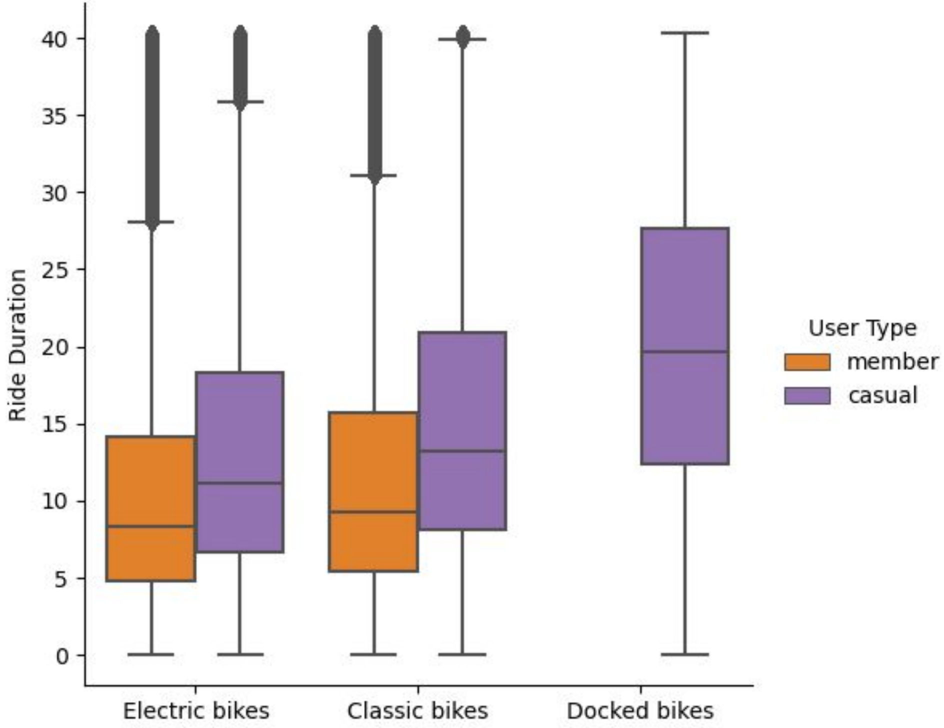


Average Ride Duration per Bike type

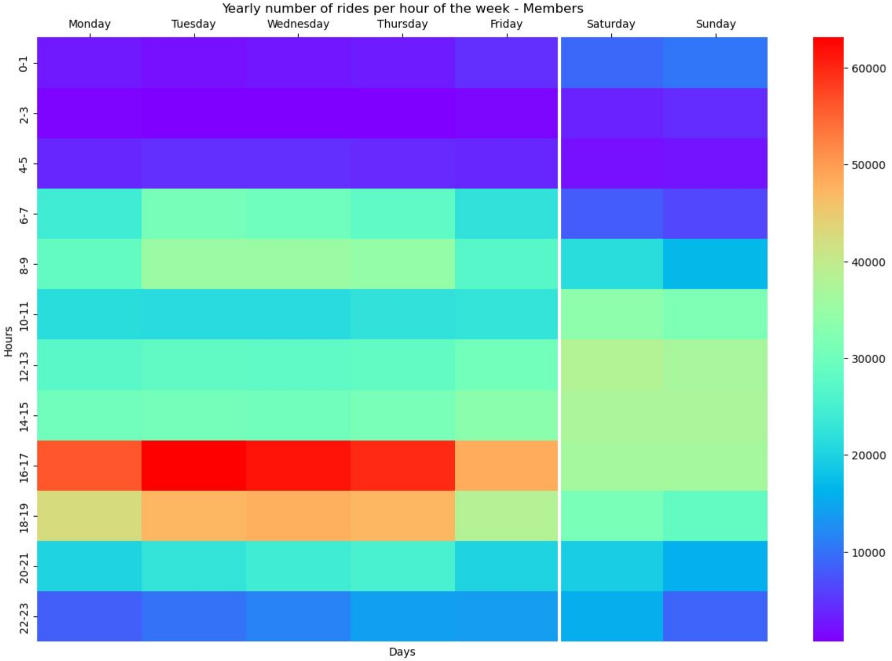


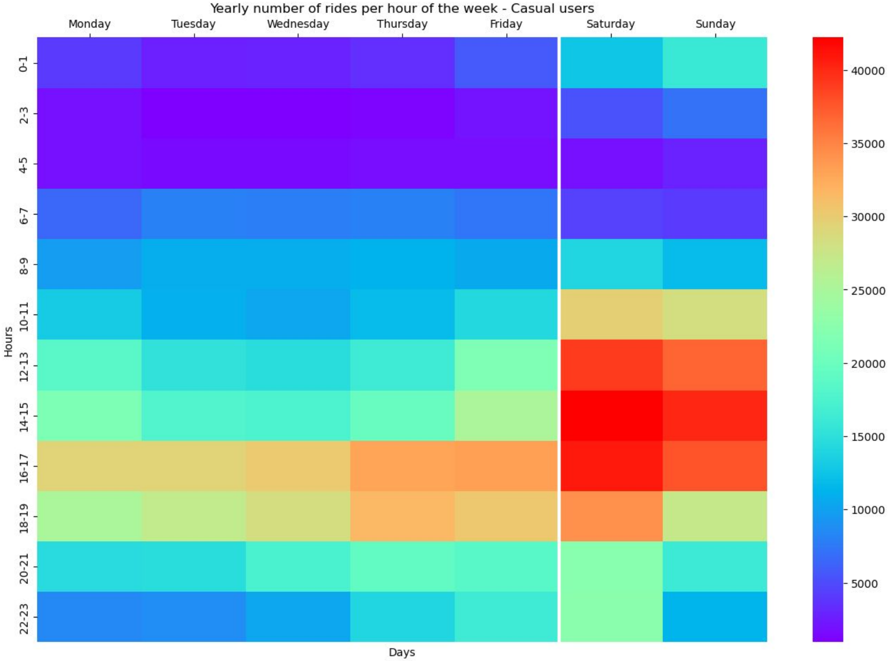
Ride Duration: Average per month
Distribution per bike and client types





Average weekly schedule / Members vs. Casual users





Improvements

With more time, we could explore libraries to further analyze

- Stats on distances
- Map of 10 most used stations and highlight some of the most common routes
- Check if some areas have significant differences in user or bike types

We could get more insights if the data included

- Bike ID #
- Member ID #

Highlights

- Challenge of having less variable data in dataset.
- Importance to understand the dataset clearly.