ECE143 TEAM 10

Optimal Bike Deployment

Chen Sun
Kejia Ruan
Youcheng Zhang
Cleavant Yang
Mansi Nanavati

Agenda

01 About the Dataset

02 Data Analysis

03 Strategies for Optimization

Capital Bikeshare dataset

A dataset tracking bike-sharing activity in Washington D.C., covering trip details, user behavior, and environmental factors spanning over 4 years (May 2020 - August 2024)

Key Features:

- Trip Details: Start/End times, durations, and station information.
- Weather Data: Temperature, precipitation and seasonal influence.
- Bicycle Station Data: Location, station type, rental frequency
- User Types: Subscriber vs. Casual user.

Why Use It?

Demand Forecasting

Station Optimization

User Behavior Analysis

Fun Facts:

Data volume of tens of millions

16,000,000 leased data

A total of 917 stations

Data Analysis

A

Rental Growth

Increase in rental numbers over the past four years

B

Weather Impact

Role of weather and seasons on rental numbers

C

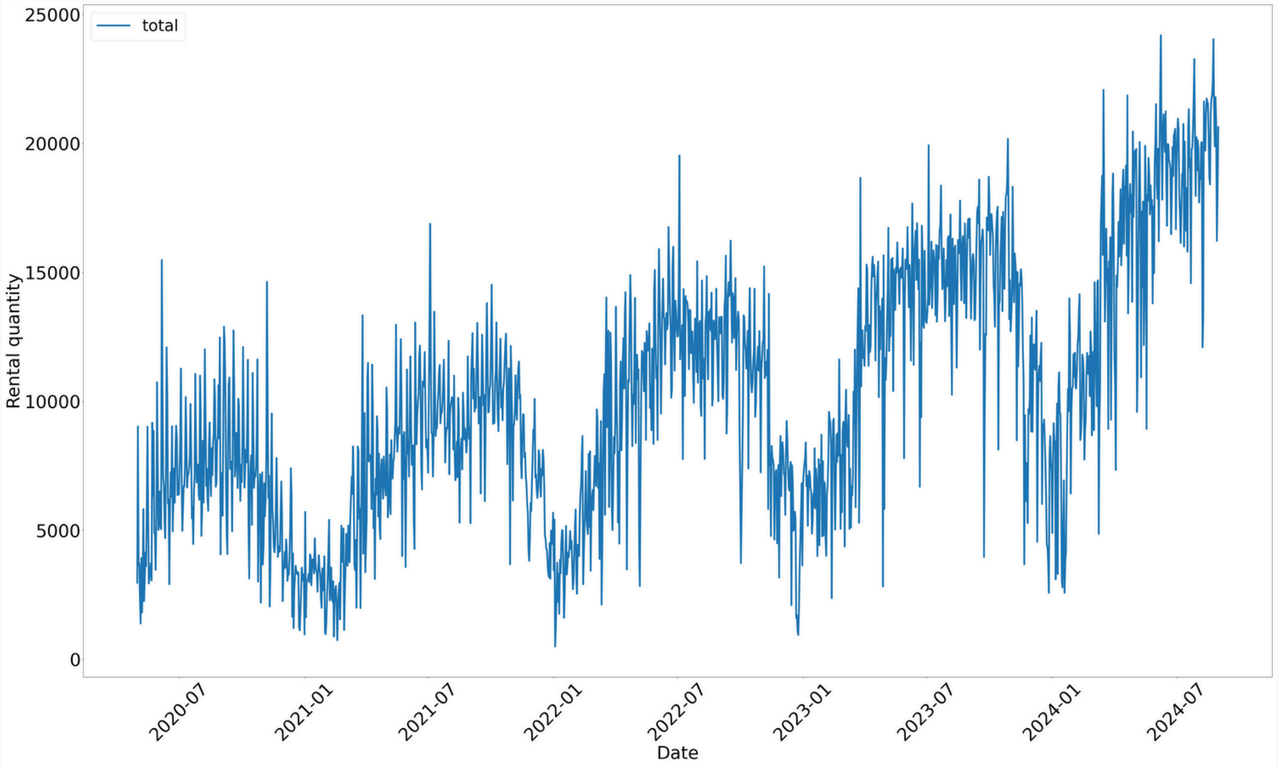
Ranking features

Analyze which features are important and rank them

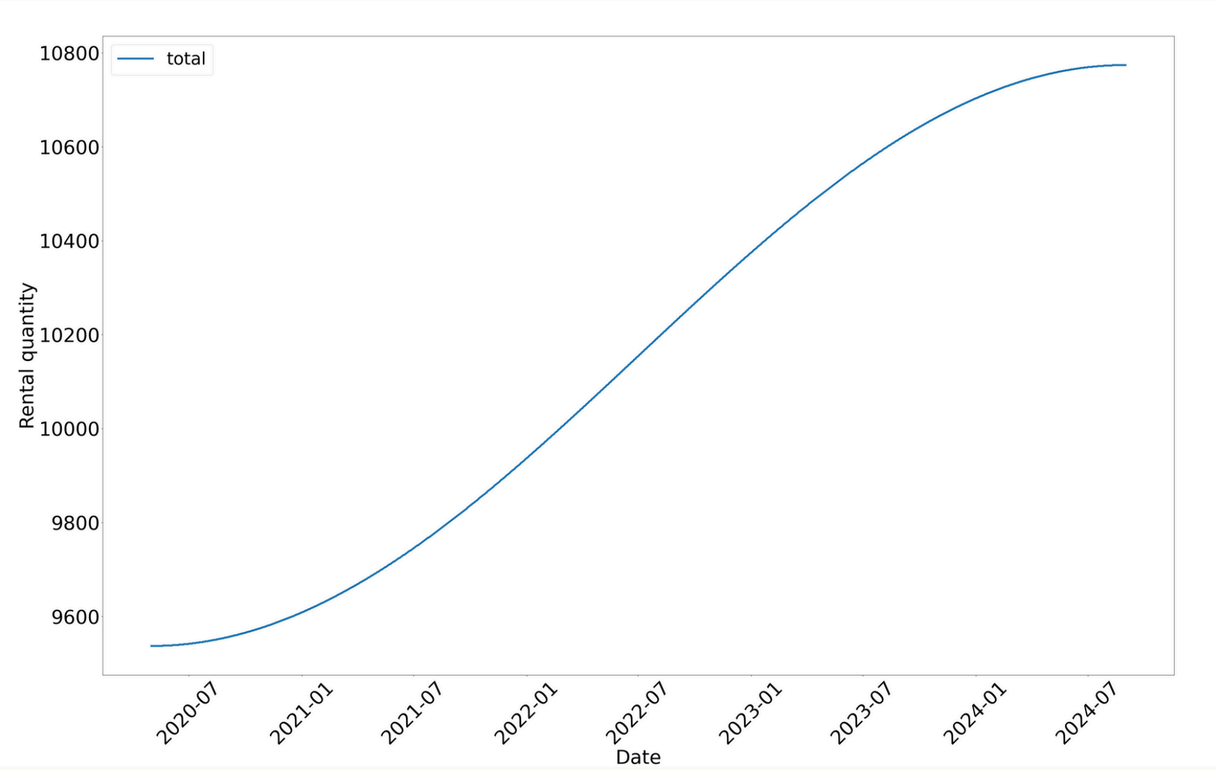
D

User Behavior

Analyze the difference between behaviors of members and casual riders



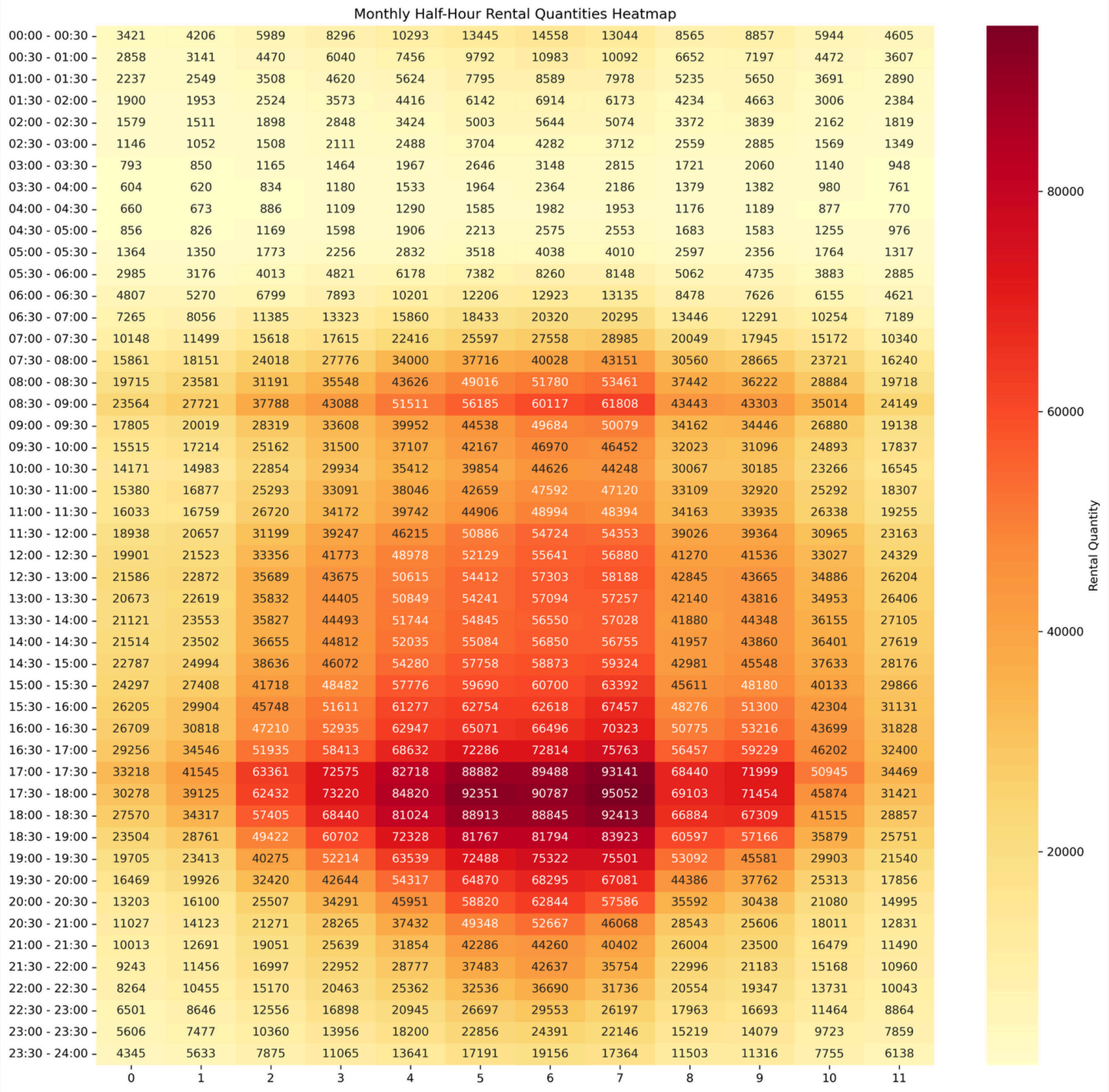
Total rental numbers over the past 4 years



Gaussian smoothed curve representing growth

Rental Growth

- Increase in rental over the 4 years with an annual growth rate of 12.5%
- Seasonal trends with peaks in summer and decline in winter
- Applied Gaussian Smoothing to analyze growth trends



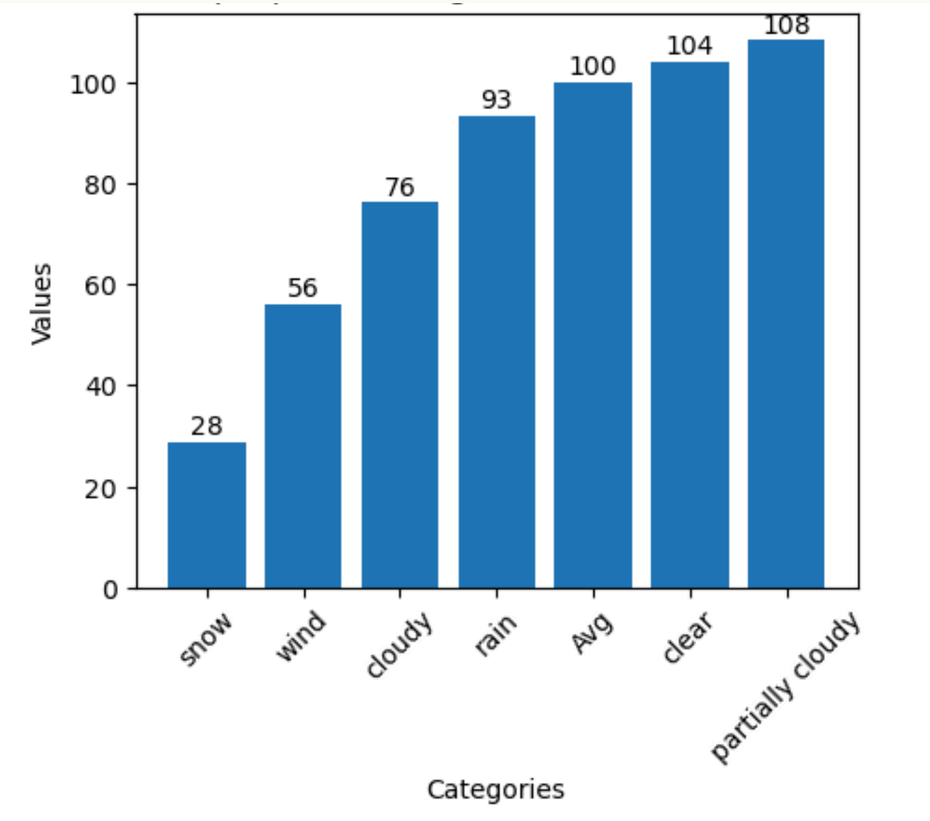
	0	1	2	3	4	5	6	7	8	9	10	11
00:00 - 00:30	3421	4206	5989	8296	10293	13445	14558	13044	8565	8857	5944	4605
00:30 - 01:00	2858	3141	4470	6040	7456	9792	10983	10092	6652	7197	4472	3607
01:00 - 01:30	2237	2549	3508	4620	5624	7795	8589	7978	5235	5650	3691	2890
01:30 - 02:00	1900	1953	2524	3573	4416	6142	6914	6173	4234	4663	3006	2384
02:00 - 02:30	1579	1511	1898	2848	3424	5003	5644	5074	3372	3839	2162	1819
02:30 - 03:00	1146	1052	1508	2111	2488	3704	4282	3712	2559	2885	1569	1349
03:00 - 03:30	793	850	1165	1464	1967	2646	3148	2815	1721	2060	1140	948
03:30 - 04:00	604	620	834	1180	1533	1964	2364	2186	1379	1382	980	761
04:00 - 04:30	660	673	886	1109	1290	1585	1982	1953	1176	1189	877	770
04:30 - 05:00	856	826	1169	1598	1906	2213	2575	2553	1683	1583	1255	976
05:00 - 05:30	1364	1350	1773	2256	2832	3518	4038	4010	2597	2356	1764	1317
05:30 - 06:00	2985	3176	4013	4821	6178	7382	8260	8148	5062	4735	3883	2885
06:00 - 06:30	4807	5270	6799	7893	10201	12206	12923	13135	8478	7626	6155	4621
06:30 - 07:00	7265	8056	11385	13323	15860	18433	20320	20295	13446	12291	10254	7189
07:00 - 07:30	10148	11499	15618	17615	22416	25597	27558	28985	20049	17945	15172	10340
07:30 - 08:00	15861	18151	24018	27776	34000	37716	40028	43151	30560	28665	23721	16240
08:00 - 08:30	19715	23581	31191	35548	43626	49016	51780	53461	37442	36222	28884	19718
08:30 - 09:00	23564	27721	37788	43088	51511	56185	60117	61808	43443	43303	35014	24149
09:00 - 09:30	17805	20019	28319	33608	39952	44538	49684	50079	34162	34446	26880	19138
09:30 - 10:00	15515	17214	25162	31500	37107	42167	46970	46452	32023	31096	24893	17837
10:00 - 10:30	14171	14983	22854	29934	35412	39854	44626	44248	30067	30185	23266	16545
10:30 - 11:00	15380	16877	25293	33091	38046	42659	47592	47120	33109	32920	25292	18307
11:00 - 11:30	16033	16759	26720	34172	39742	44906	48994	48394	34163	33935	26338	19255
11:30 - 12:00	18938	20657	31199	39247	46215	50886	54724	54353	39026	39364	30965	23163
12:00 - 12:30	19901	21523	33356	41773	48978	52129	55641	56880	41270	41536	33027	24329
12:30 - 13:00	21586	22872	35689	43675	50615	54412	57303	58188	42845	43665	34886	26204
13:00 - 13:30	20673	22619	35832	44405	50849	54241	57094	57257	42140	43816	34953	26406
13:30 - 14:00	21121	23553	35827	44493	51744	54845	56550	57028	41880	44348	36155	27105
14:00 - 14:30	21514	23502	36655	44812	52035	55084	56850	56755	41957	43860	36401	27619
14:30 - 15:00	22787	24994	38636	46072	54280	57758	58873	59324	42981	45548	37633	28176
15:00 - 15:30	24297	27408	41718	48482	57776	59690	60700	63392	45611	48180	40133	29866
15:30 - 16:00	26205	29904	45748	51611	61277	62754	62618	67457	48276	51300	42304	31131
16:00 - 16:30	26709	30818	47210	52935	62947	65071	66496	70323	50775	53216	43699	31828
16:30 - 17:00	29256	34546	51935	58413	68632	72286	72814	75763	56457	59229	46202	32400
17:00 - 17:30	33218	41545	63361	72575	82718	88882	89488	93141	68440	71999	50945	34469
17:30 - 18:00	30278	39125	62432	73220	84820	92351	90787	95052	69103	71454	45874	31421
18:00 - 18:30	27570	34317	57405	68440	81024	88913	88845	92413	66884	67309	41515	28857
18:30 - 19:00	23504	28761	49422	60702	72328	81767	81794	83923	60597	57166	35879	25751
19:00 - 19:30	19705	23413	40275	52214	63539	72488	75322	75501	53092	45581	29903	21540
19:30 - 20:00	16469	19926	32420	42644	54317	64870	68295	67081	44386	37762	25313	17856
20:00 - 20:30	13203	16100	25507	34291	45951	58820	62844	57586	35592	30438	21080	14995
20:30 - 21:00	11027	14123	21271	28265	37432	49348	52667	46068	28543	25606	18011	12831
21:00 - 21:30	10013	12691	19051	25639	31854	42286	44260	40402	26004	23500	16479	11490
21:30 - 22:00	9243	11456	16997	22952	28777	37483	42637	35754	22996	21183	15168	10960
22:00 - 22:30	8264	10455	15170	20463	25362	32536	36690	31736	20554	19347	13731	10043
22:30 - 23:00	6501	8646	12556	16898	20945	26697	29553	26197	17963	16693	11464	8864
23:00 - 23:30	5606	7477	10360	13956	18200	22856	24391	22146	15219	14079	9723	7859
23:30 - 24:00	4345	5633	7875	11065	13641	17191	19156	17364	11503	11316	7755	6138

Daily Patterns

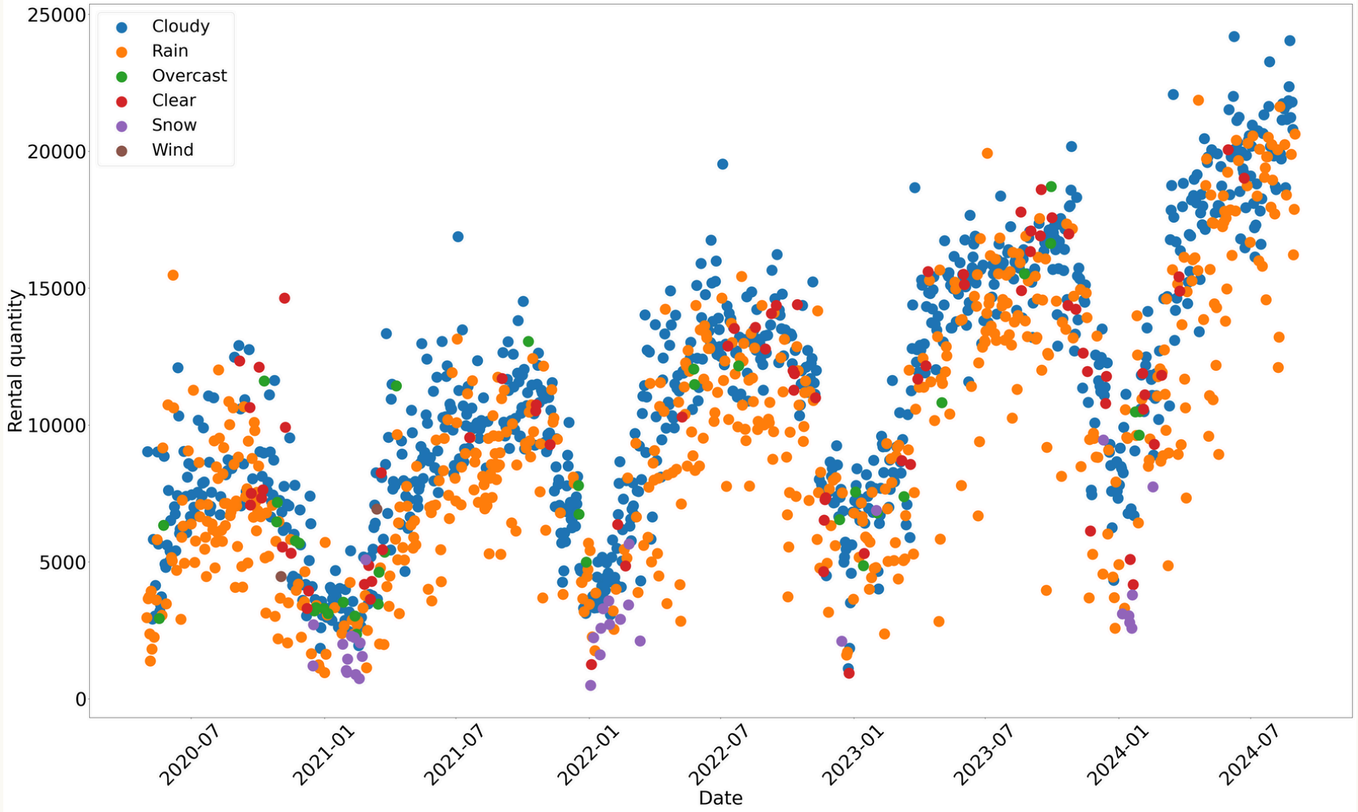
Rentals surge during morning and evening rush hours, with the evening peak being more prominent

Heatmap Insights

Aggregated rental records over 30-minute intervals reveal clear patterns of seasonal and daily rental variations



Ratio (in %) of average rental of a specific weather against the overall rental average



Shows the weather and rental volume of each day

Weather Impact

- Rentals drop to 28% of avg on **snowy days** - largest impact
- Rentals are as low as 56% of avg on windy days
- Cloudy and rainy days are also below avg
- Clear and partially cloudy days boost rentals

```
F-statistic: 26.840, P-value: 0.000
Differences between categories are statistically significant (reject the null hypothesis)
```

F-statistic among samples from all weather conditions

```
partially_cloudy    F-statistic: 0.560, P-value: 0.454
rain                F-statistic: 4.024, P-value: 0.045
cloudy              F-statistic: 10.340, P-value: 0.002
snow                F-statistic: 87.119, P-value: 0.000
wind                F-statistic: 2.299, P-value: 0.134
```

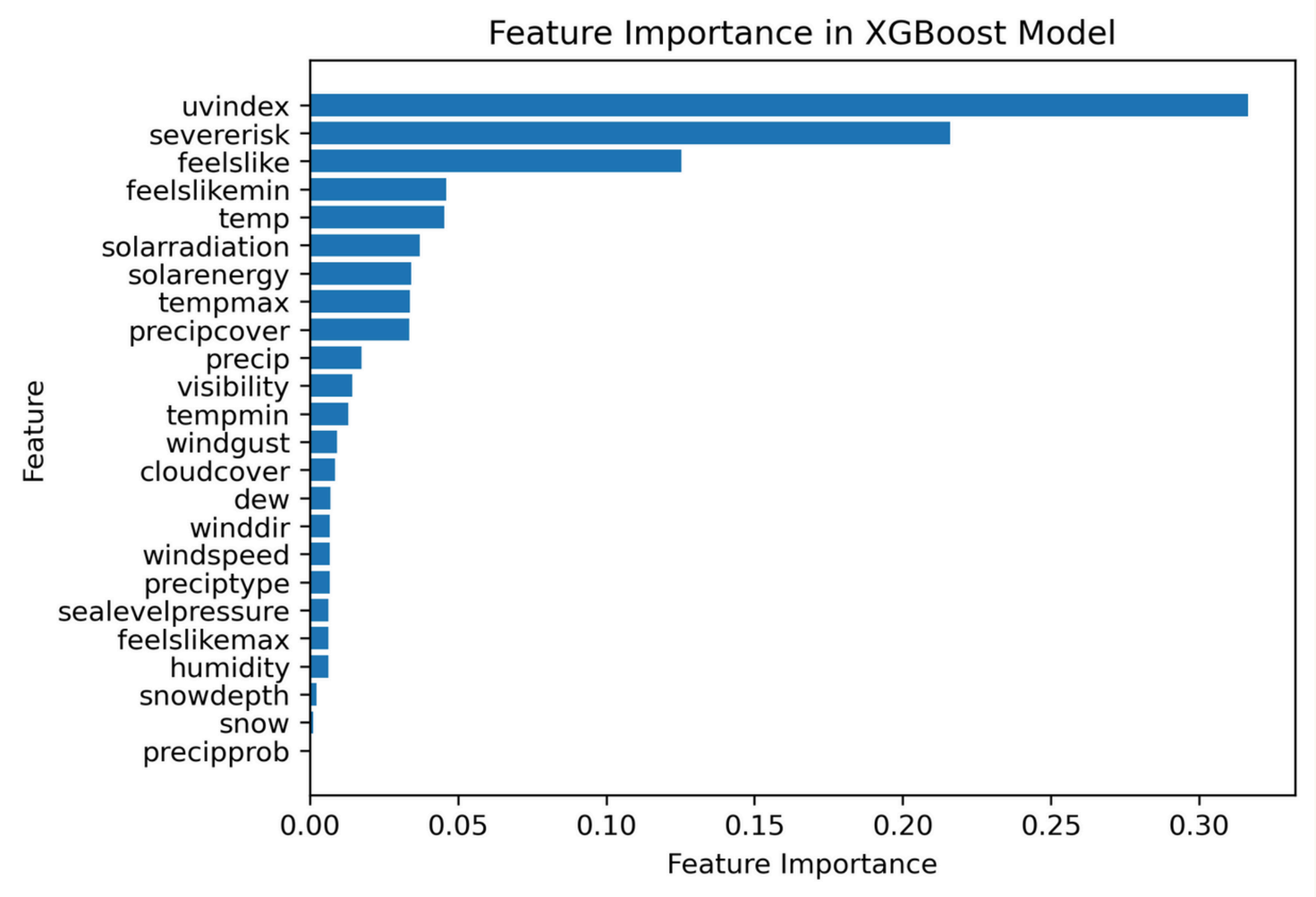
F-Statistic generated for each weather with samples from clear day

Weather Impact

F-statistic value of 26 confirms significant variance in rentals based on weather conditions.

Conclusion

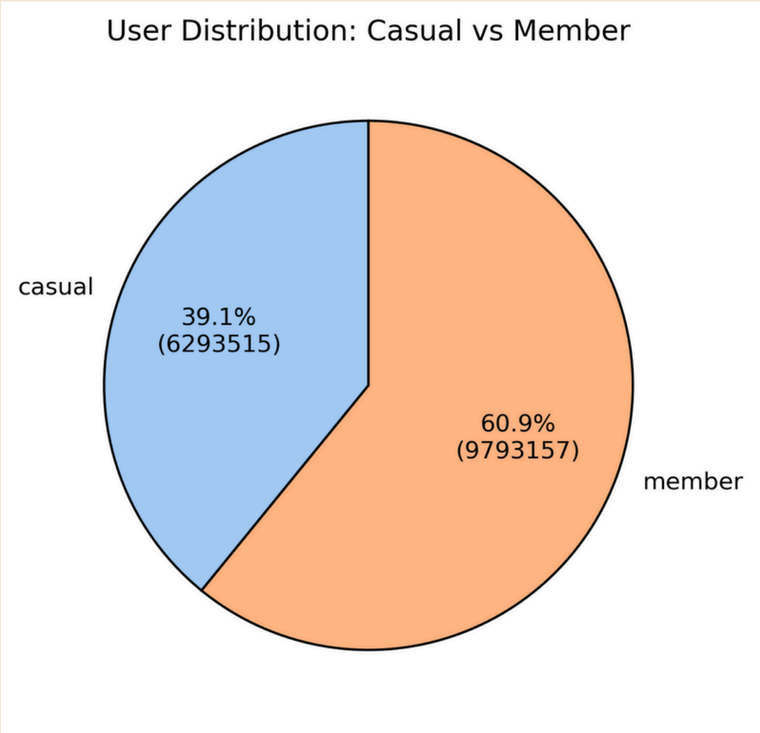
- Weather has a huge impact on travel
- To optimize, we identify key weather-impact features and model future bicycle usage.



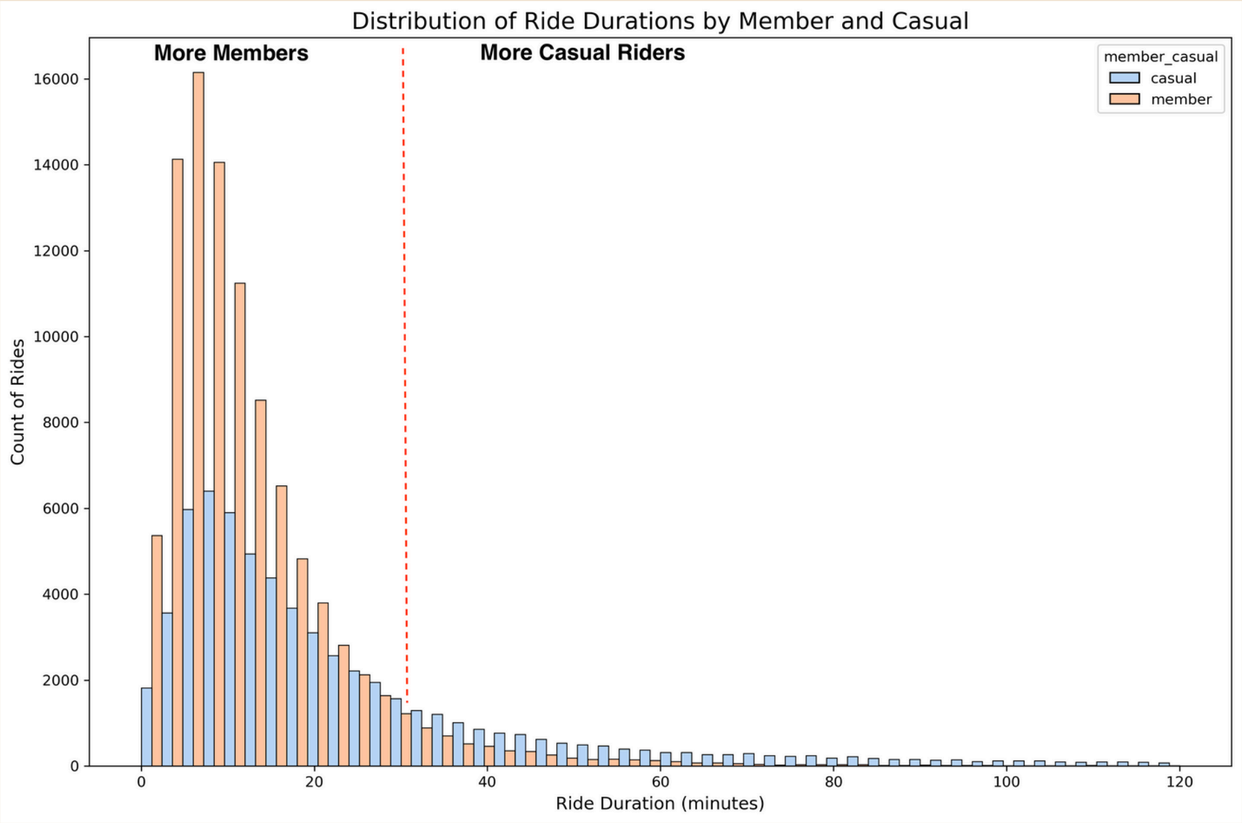
Ranking Features

- Measured importance of features using XGBoost Model and ranked them.
- Important features: temperature, UV index, and precipitation
- We leverage the important features to predict future bike usage based on weather conditions

User Behavior



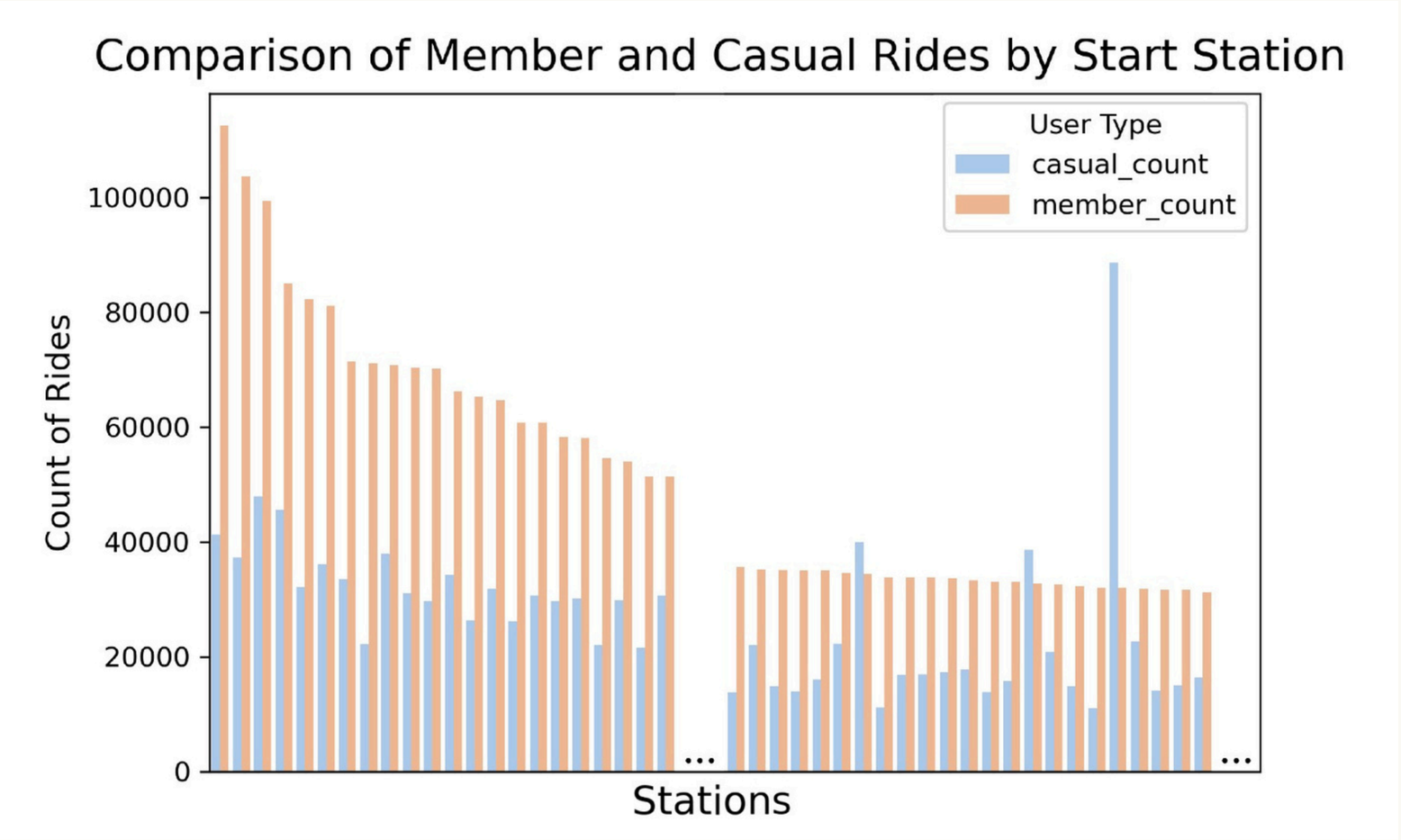
Members outnumber casual riders



Most rides are under 20 mins, with members taking rides < 30 mins and casual riders taking rides > 30 mins.

Conclusion:

Frequent short-trip users (ex: commuters) are likely to become members, while occasional long-ride users stay non-members.



No. of members and casual riders by start station (sorted by member count in descending order, taking top 100)

User Behavior

- Most stations have more member riders than casual riders.
- At some stations, casual riders outnumber members.
- Stations with more casual riders account for 37.03% of all stations.

Strategy for Optimization

01 Forecasting Model

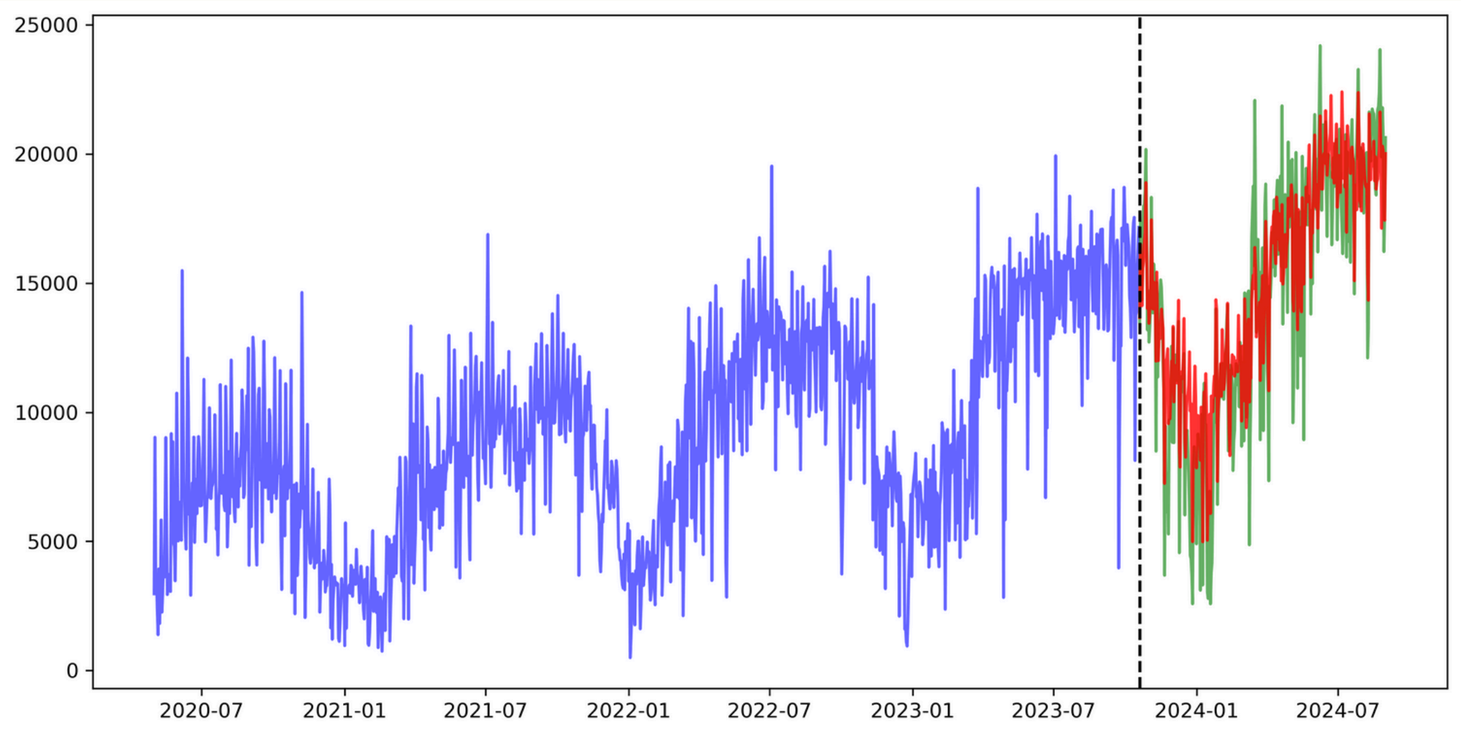
Regression model that predicts future bicycle usage

02 Scheduling strategy

Scheduling strategies on a daily and monthly basis

03 Targeted Promotions

Increasing the conversion rate from casual riders to members via promotions



Rental quantity v/s Time

```
Mean Squared Error (MSE): 3907222.9902083008
Mean Absolute Error (MAE): 1458.3266245041264
R² Score: 0.8432569983242107
```

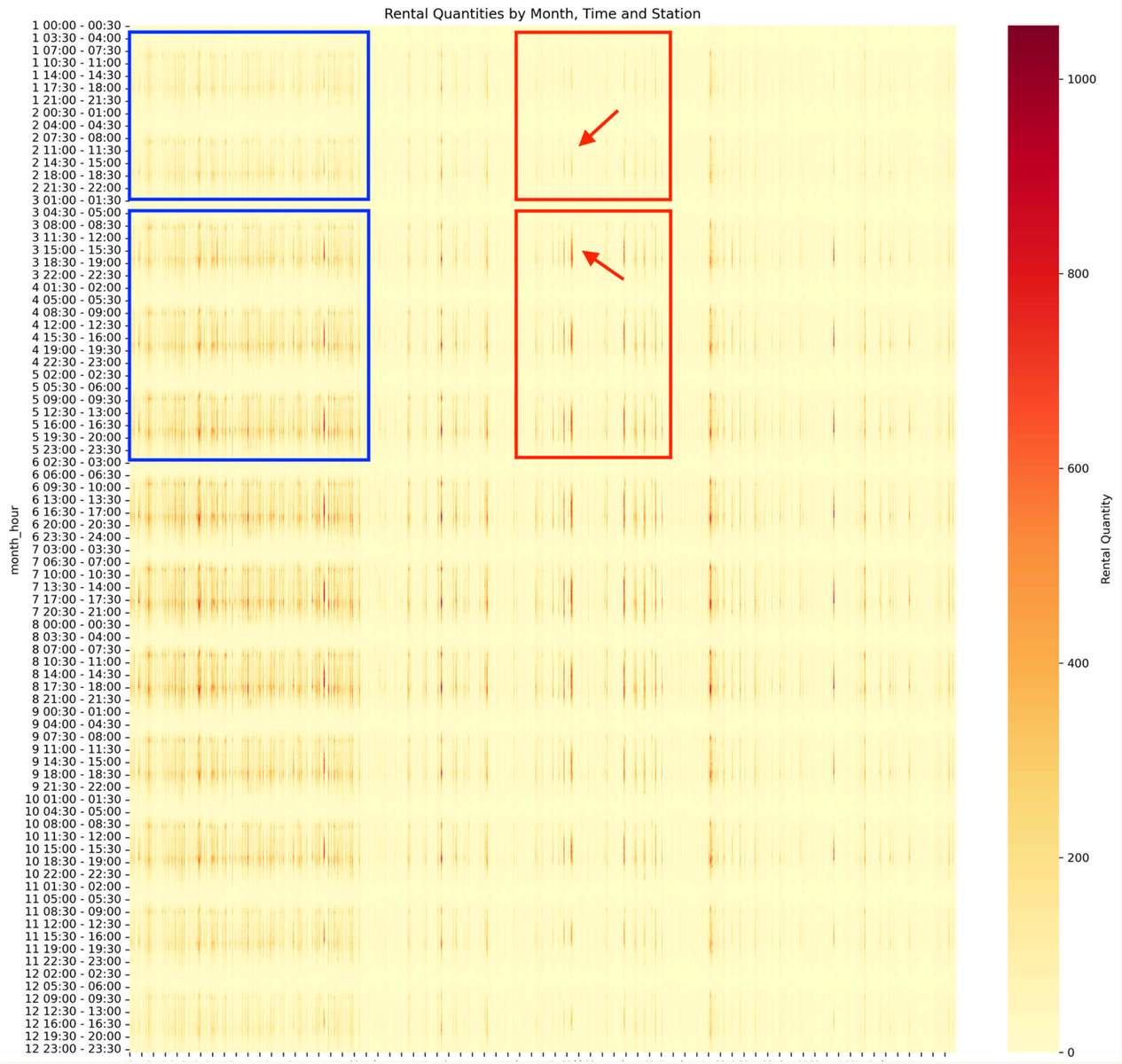
Forecasting Model

- The Prophet model is well-suited for time series forecasting with seasonal patterns and trend changes.
- **Achieved a fit of 84.3%** with a mean absolute error **MAE of 1458**.
- Low MAE value shows that the model works well with data
- Given an average rental volume of 10,156, the model's **accuracy exceeds 85%**, calculated as (10,156 - 1,458) / 10,156 * 100%.

Forecasting Model

Here are some strategies to follow:

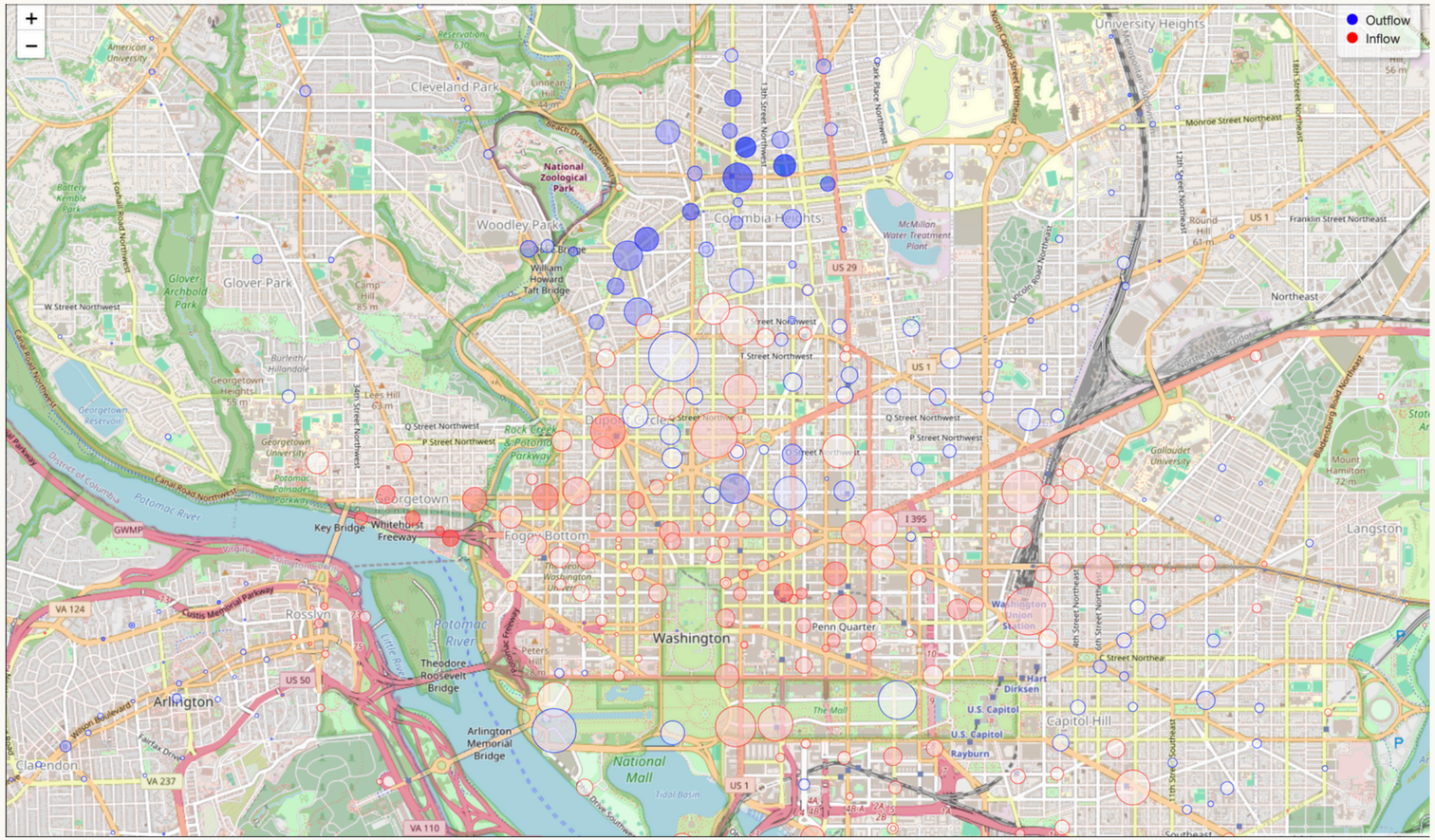
- On a daily basis, predict the bike usage based on weather forecast and deploy accordingly
- On a monthly basis, calculate and optimize the monthly rental volume since the data is greatly affected by season



Month and time v/s Stations

Scheduling Strategy

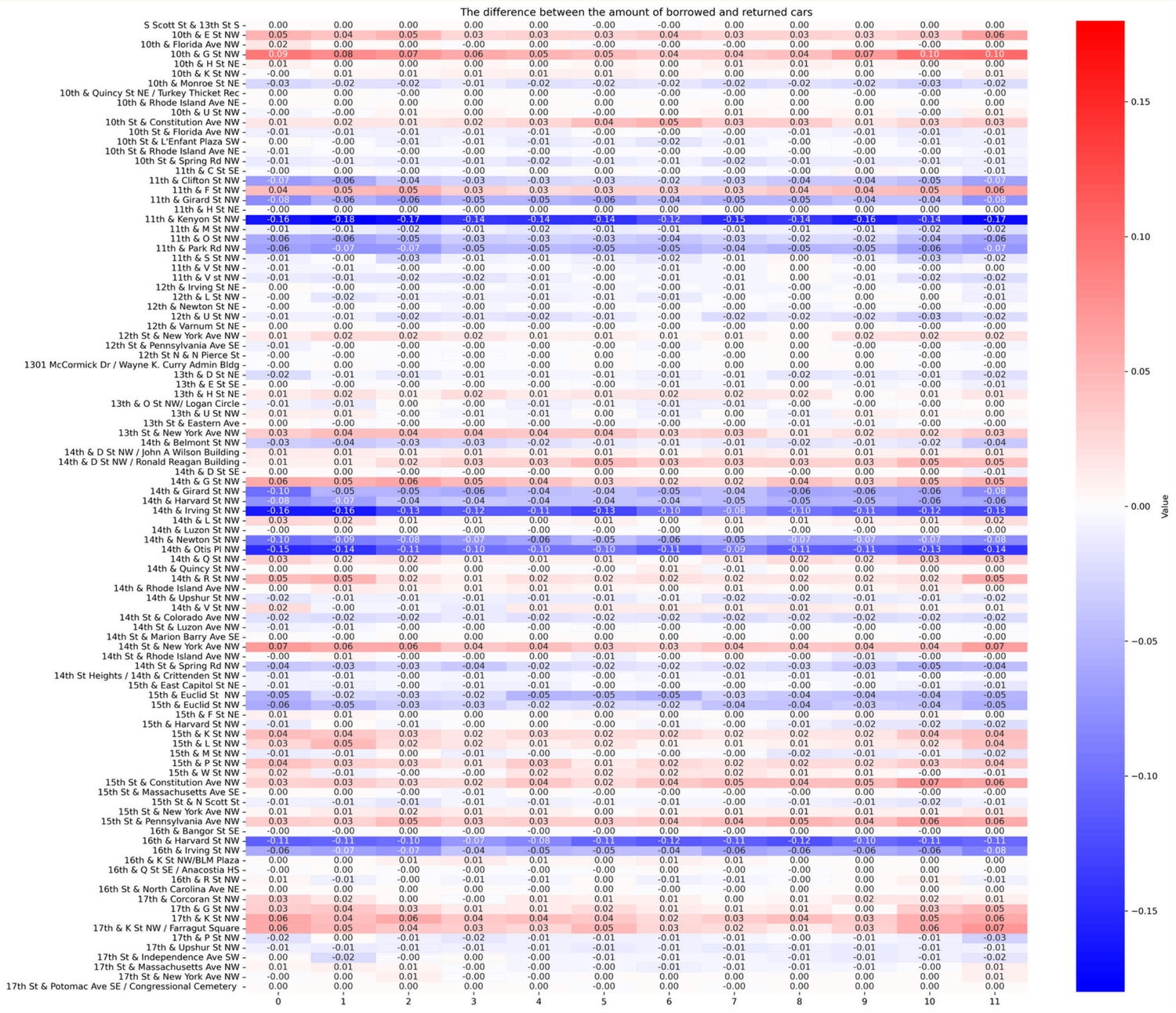
- The heat map shows that peak hours are consistent across stations (morning and evening), so no intra-day scheduling between stations is needed.
- Stations in the red box have lower rentals in January and February but see increase in March, emphasizing monthly adjustments in station allocation.



Darker blue stations -> Higher outflow of bikes
Darker red stations -> Higher inflow

Scheduling Strategy

- Mapped bike inflow and outflow using the pickup and drop off counts for every station in the dataset.
- Bike inflows higher in city center
- Outflows higher in suburbs, like Columbia Heights and Woodley Park



Scheduling Strategy

Conclusion

- Before each month, transfer bicycles from red stations (high inflow) to blue stations (high outflow) based on predicted monthly changes.

Targeted Promotions

- Increasing the conversion rate from casual riders to members via promotions
- Advertisements: Focus on attracting short-distance non-members by placing advertisements at office area stations.
- Offering discounts: Offer membership discounts for rides over an hour to convert casual long-distance riders into members.

Takeaways

- **Ranking features:** Dataset is quite vast so it is important to pick features that best represent data
- **Time-series forecasting model prediction:** Bike usage is heavily dependent on weather conditions and seasons, leading us to plan daily deployment based on weather and peak hours, and monthly planning based on seasons.
- **Literally mapping the data:** Identifying inflow and outflow stations is useful for logistics

Any questions?