ElectricPe Data Analysis and Analytics

Submitted By-
Darshan N Shenoy
20BLC1076
VIT, Chennai

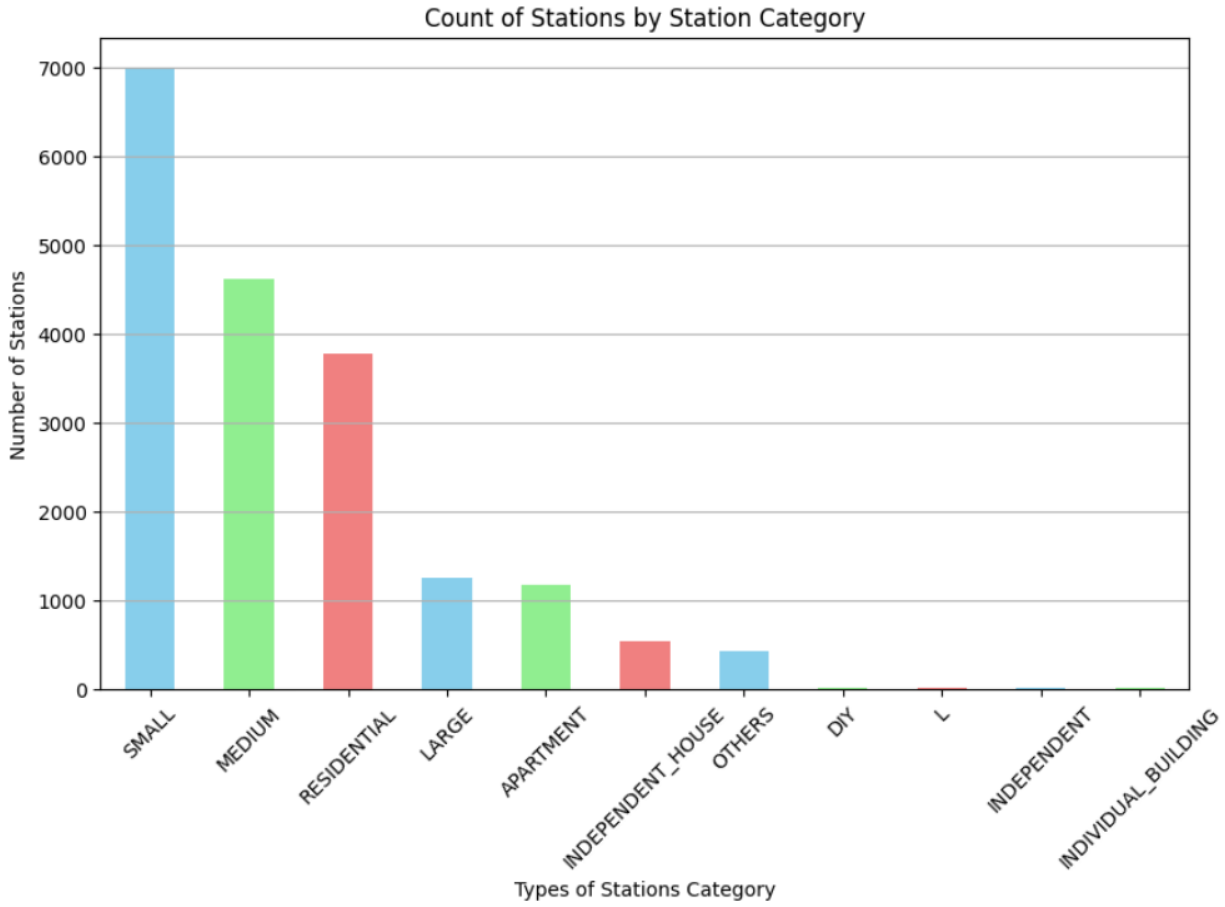
Data Preprocessing:

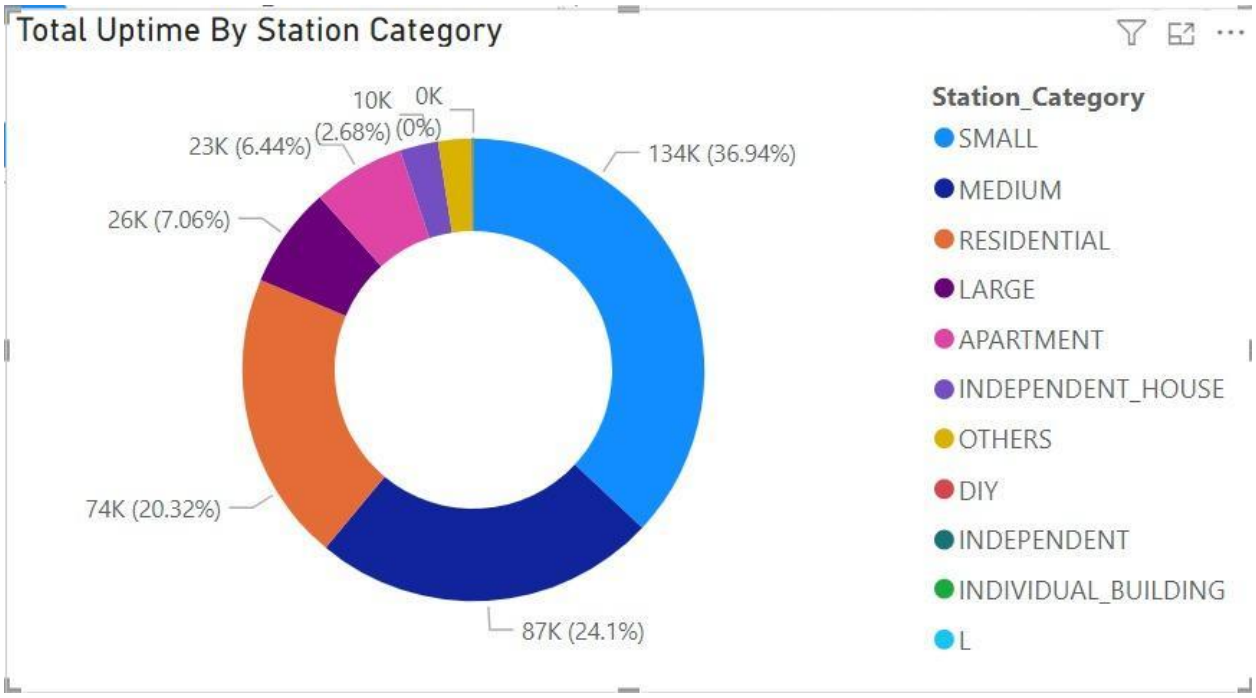
- Imported the required libraries.
- Determined and eliminated null values.
- In the "hour" columns, "True" values were changed to "1" and "False" values to "0".
- Determine the columns' distinct values.
- Number of rows associated with each station category and access has been identified.

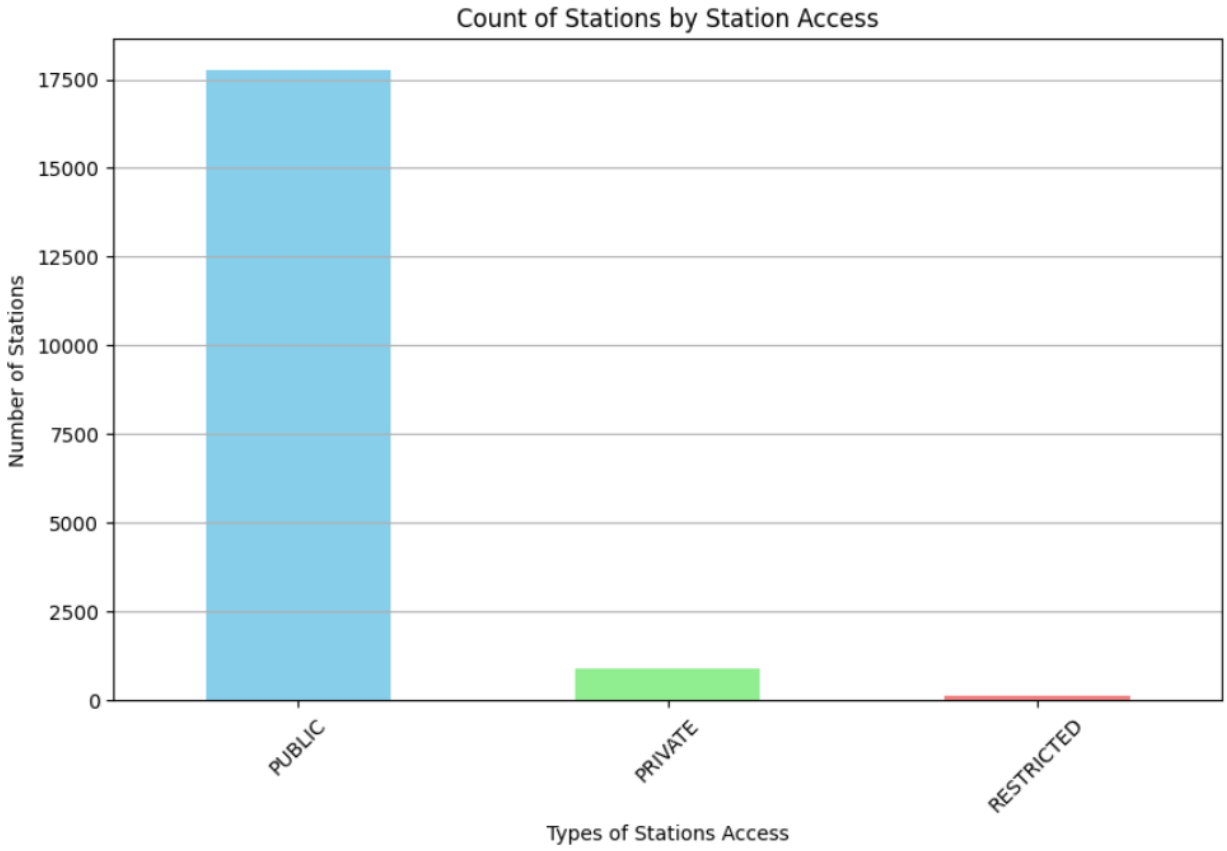
Screenshots:

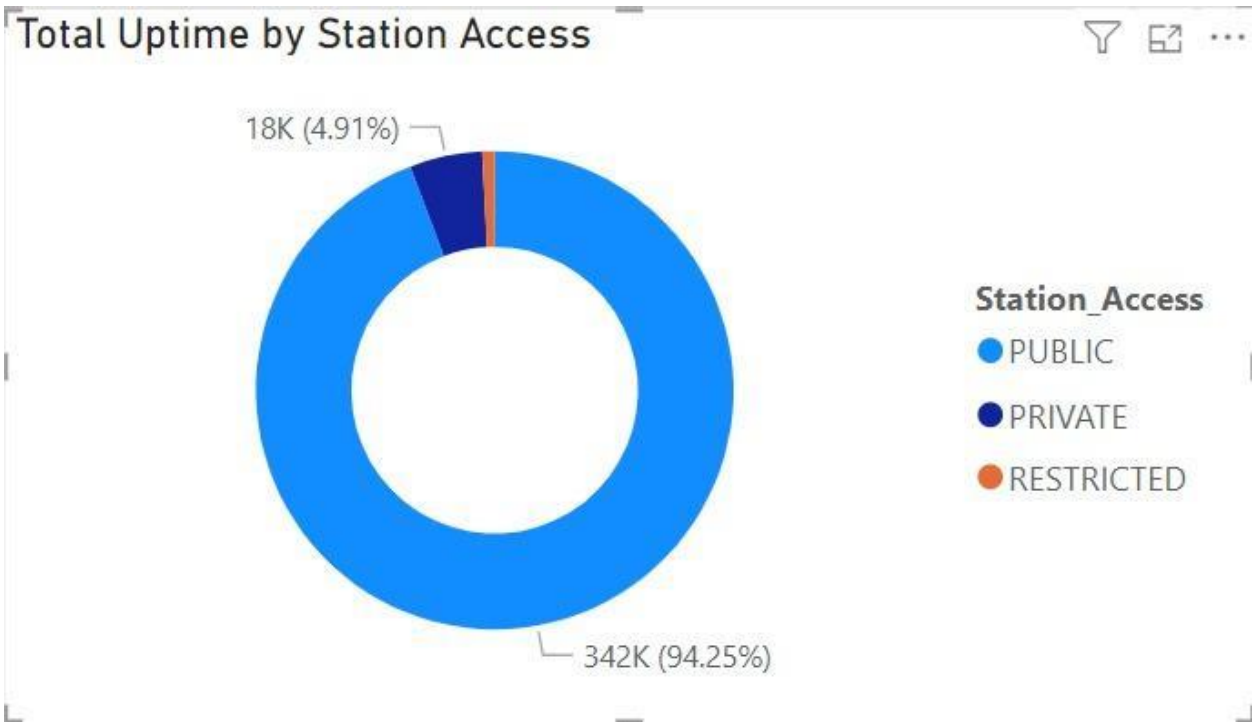
No. of rows for every station category-

```
SMALL                6980
MEDIUM               4605
RESIDENTIAL          3775
LARGE                1255
APARTMENT            1170
INDEPENDENT_HOUSE     540
OTHERS                425
DIY                    10
L                       5
INDEPENDENT             5
INDIVIDUAL_BUILDING     5
Name: Station_Category, dtype: int64
```




Total Uptime By Station Category
10K (2.68%)
0K (0%)
23K (6.44%)
134K (36.94%)
26K (7.06%)
74K (20.32%)
87K (24.1%)
Station_Category
SMALL
MEDIUM
RESIDENTIAL
LARGE
APARTMENT
INDEPENDENT_HOUSE
OTHERS
DIY
INDEPENDENT
INDIVIDUAL_BUILDING
L


Count of Stations by Station Access
Number of Stations
17500
15000
12500
10000
7500
5000
2500
0
PUBLIC
PRIVATE
RESTRICTED
Types of Stations Access

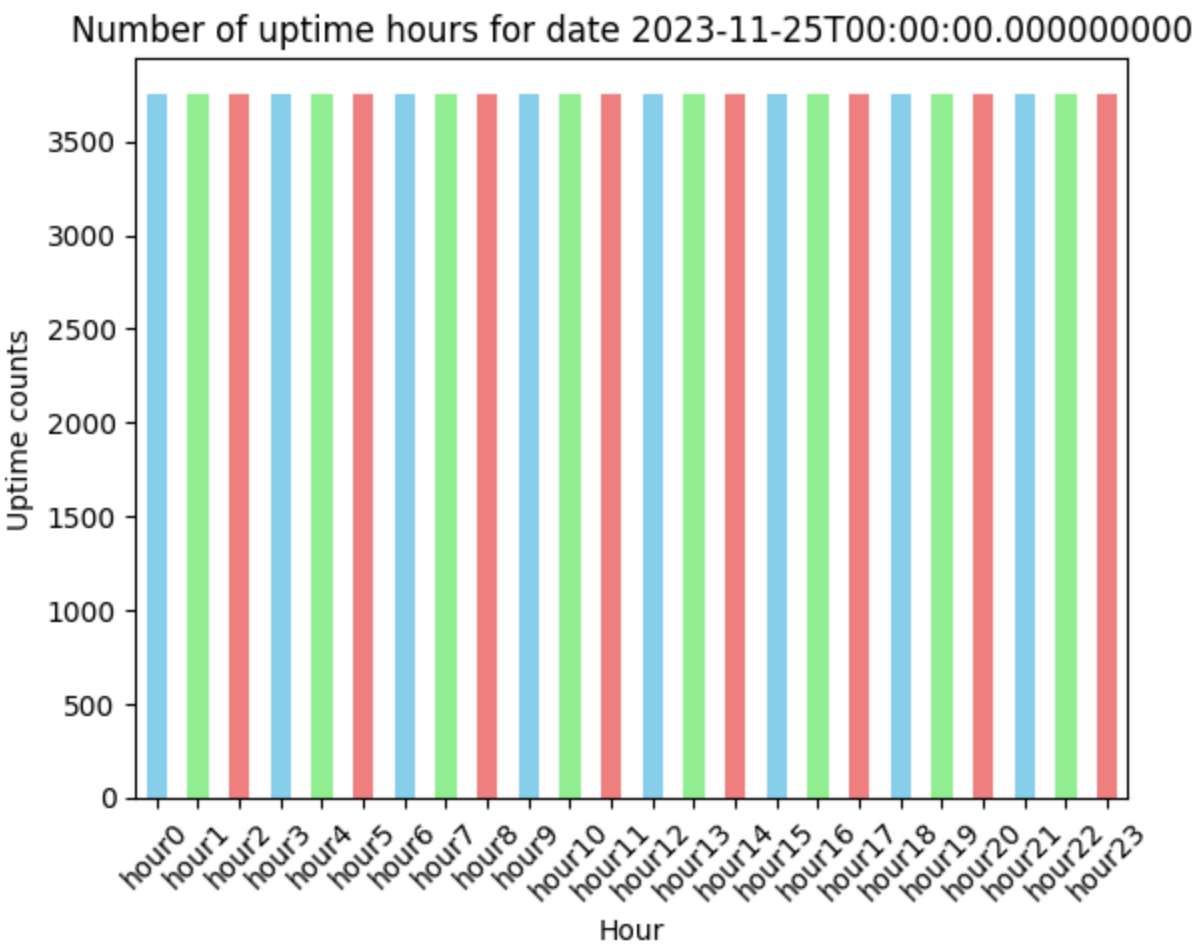


Task 1: Daily and Hourly Uptime Analysis

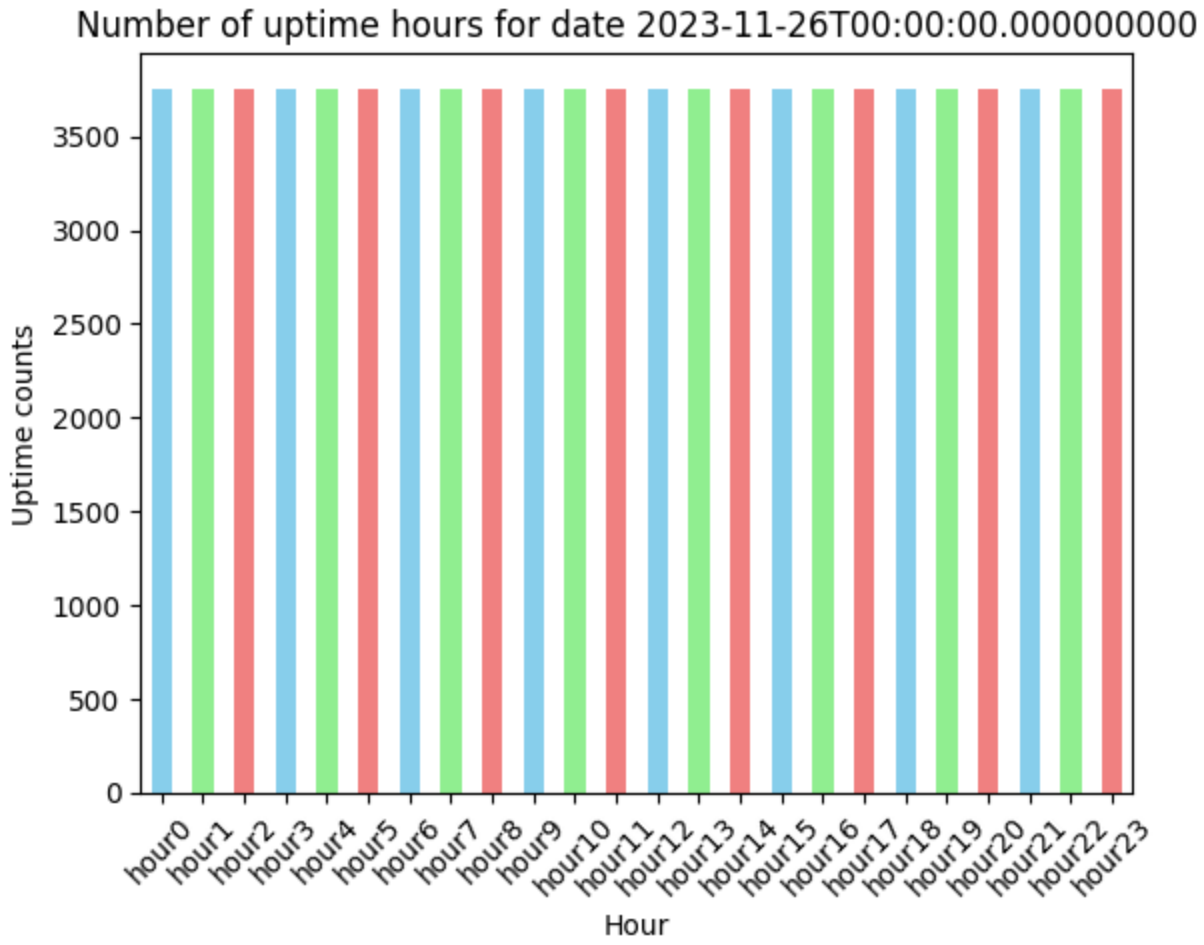
- Determined how many hours each Station Category had uptime.
- Determined how the uptime hours were distributed for different dates.
- There was no discernible pattern of uptime. There was little to no difference in uptime between different days and hours.

Screenshots:

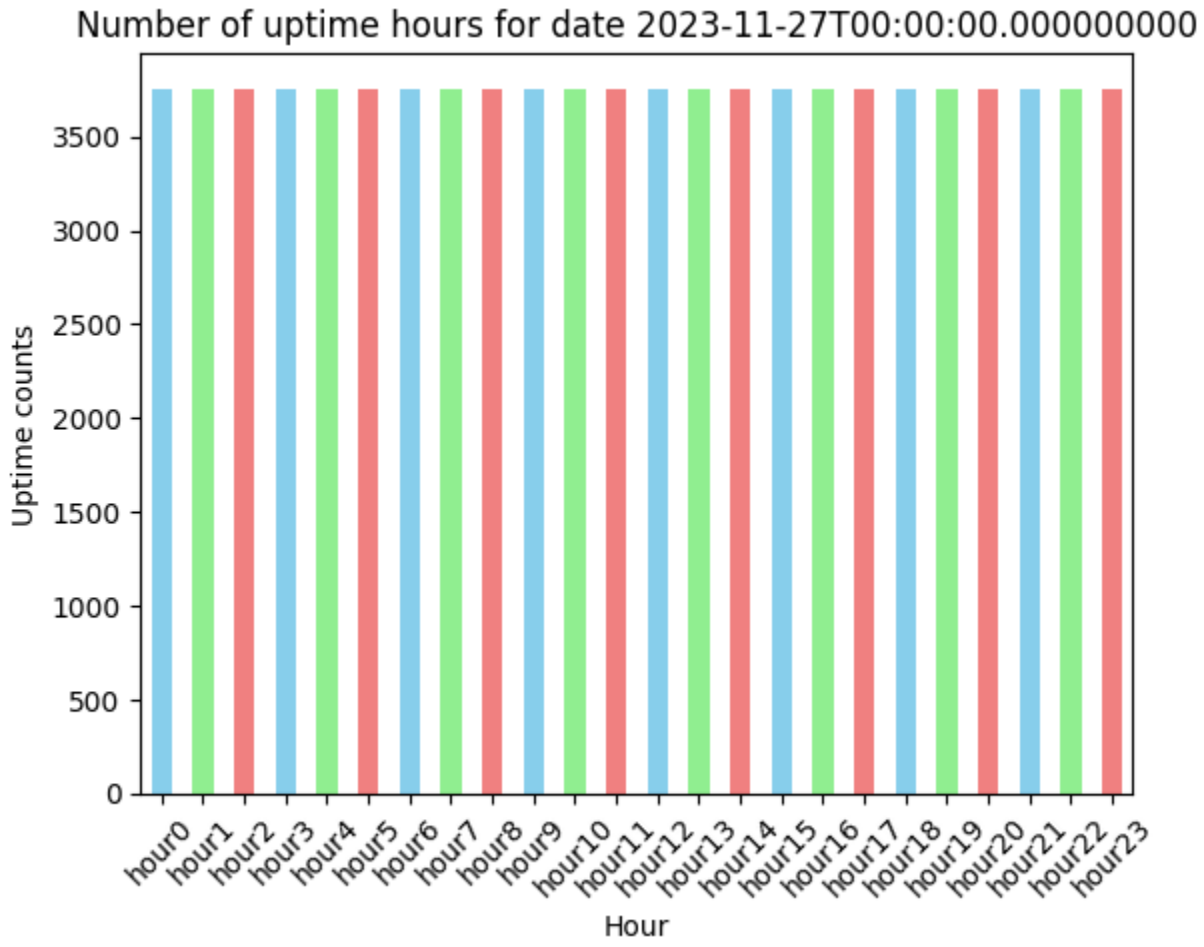
No. of uptime hours for 25th Nov for all the stations-


Number of uptime hours for date 2023-11-25T00:00:00.000000000
Uptime counts
0
500
1000
1500
2000
2500
3000
3500
hour0
hour1
hour2
hour3
hour4
hour5
hour6
hour7
hour8
hour9
hour10
hour11
hour12
hour13
hour14
hour15
hour16
hour17
hour18
hour19
hour20
hour21
hour22
hour23
Hour

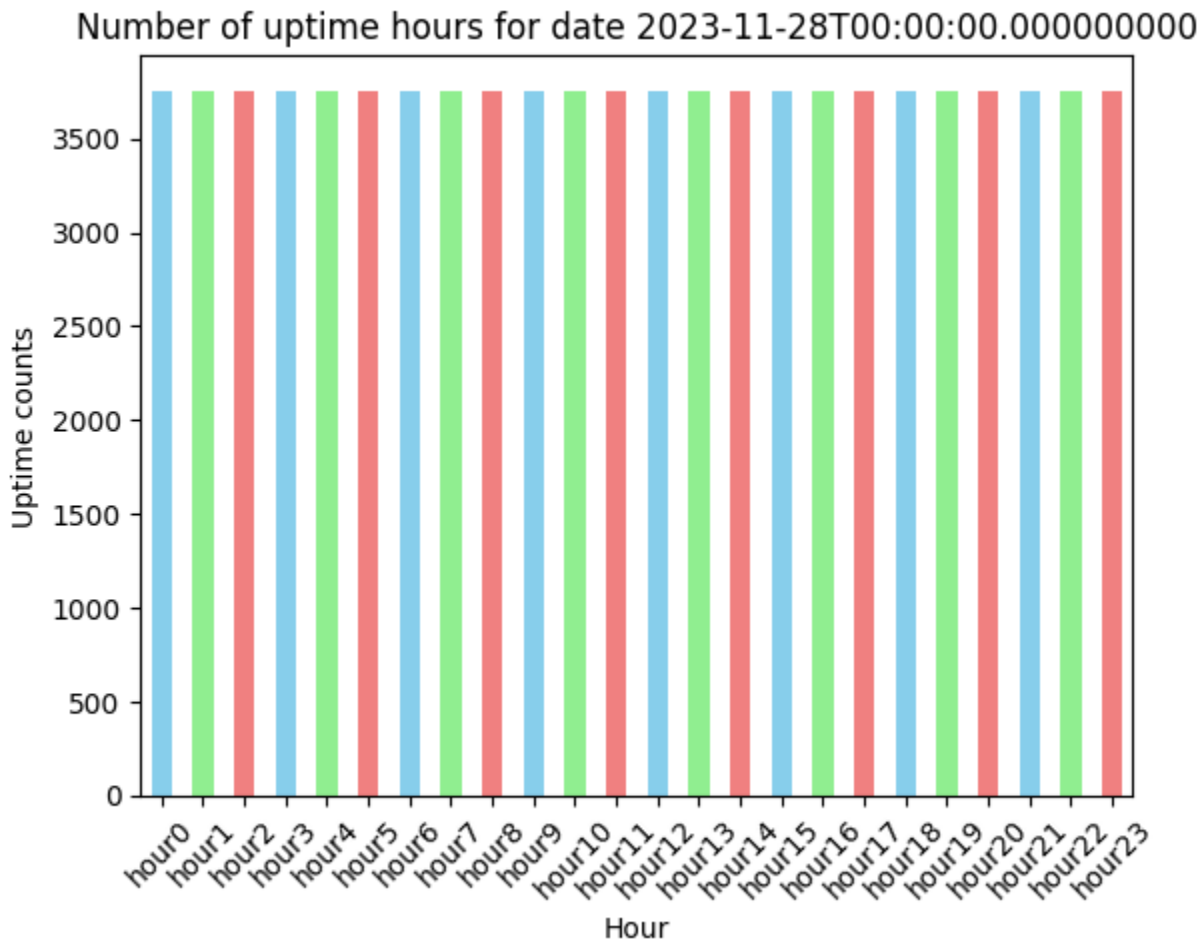
No. of uptime hours for 26th Nov for all the stations-



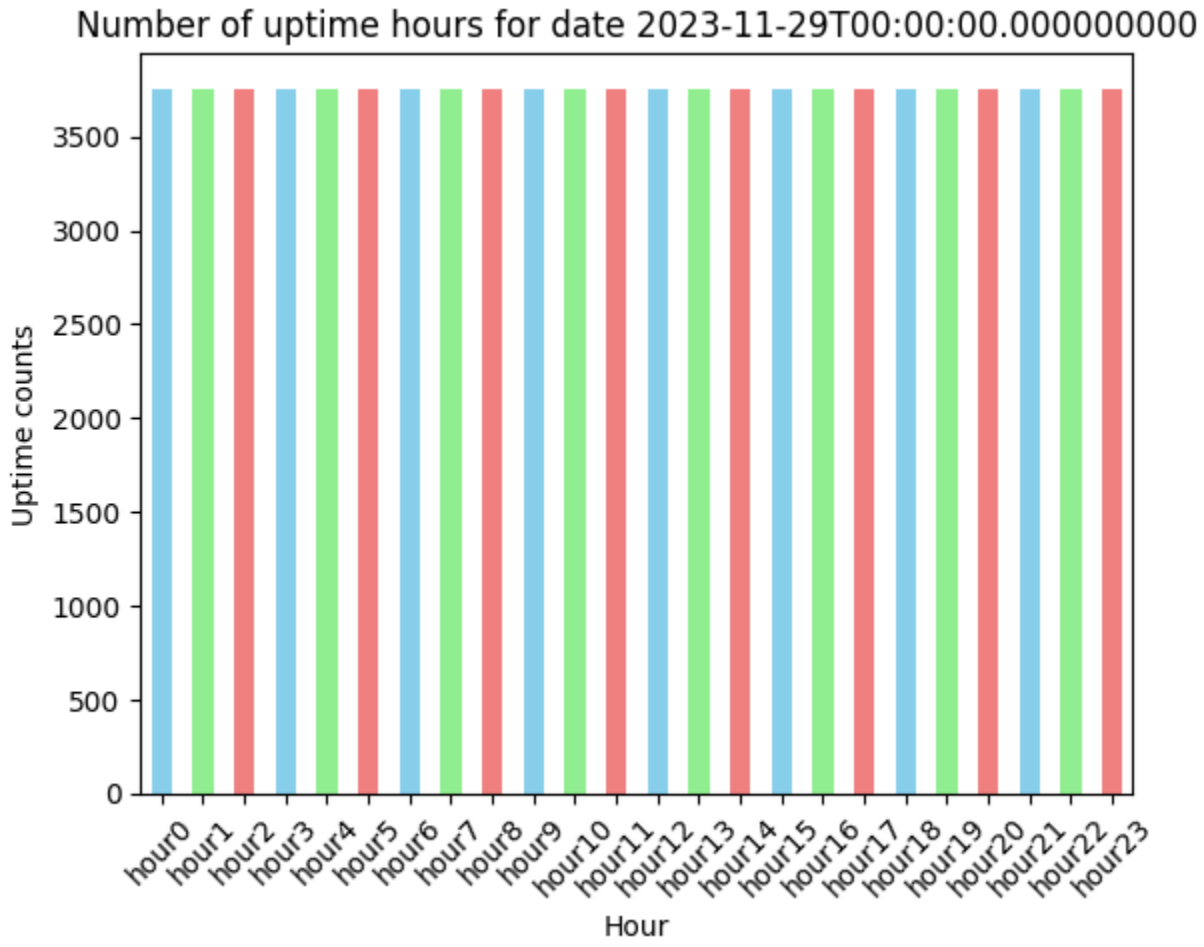
No. of uptime hours for 27th Nov for all the stations-



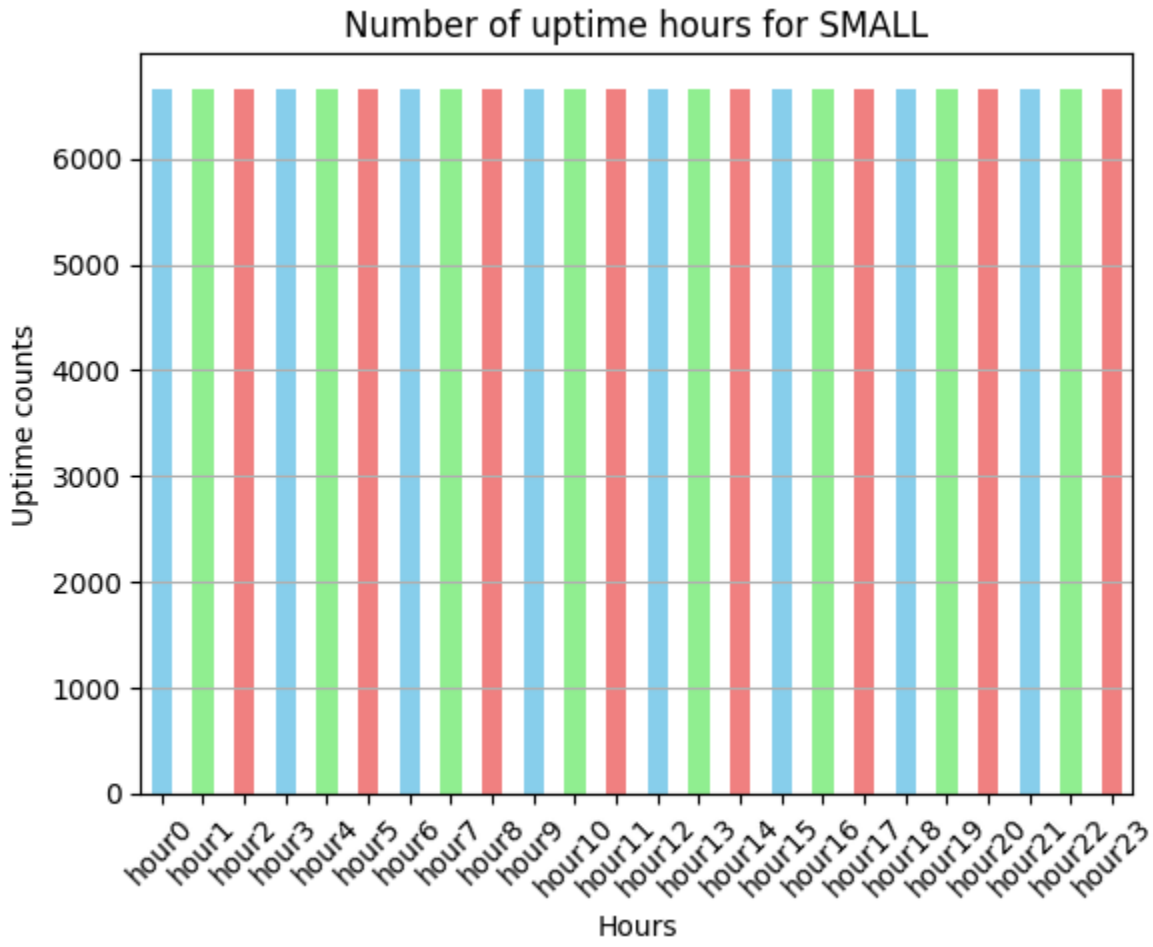
No. of uptime hours for 28th Nov for all the stations-



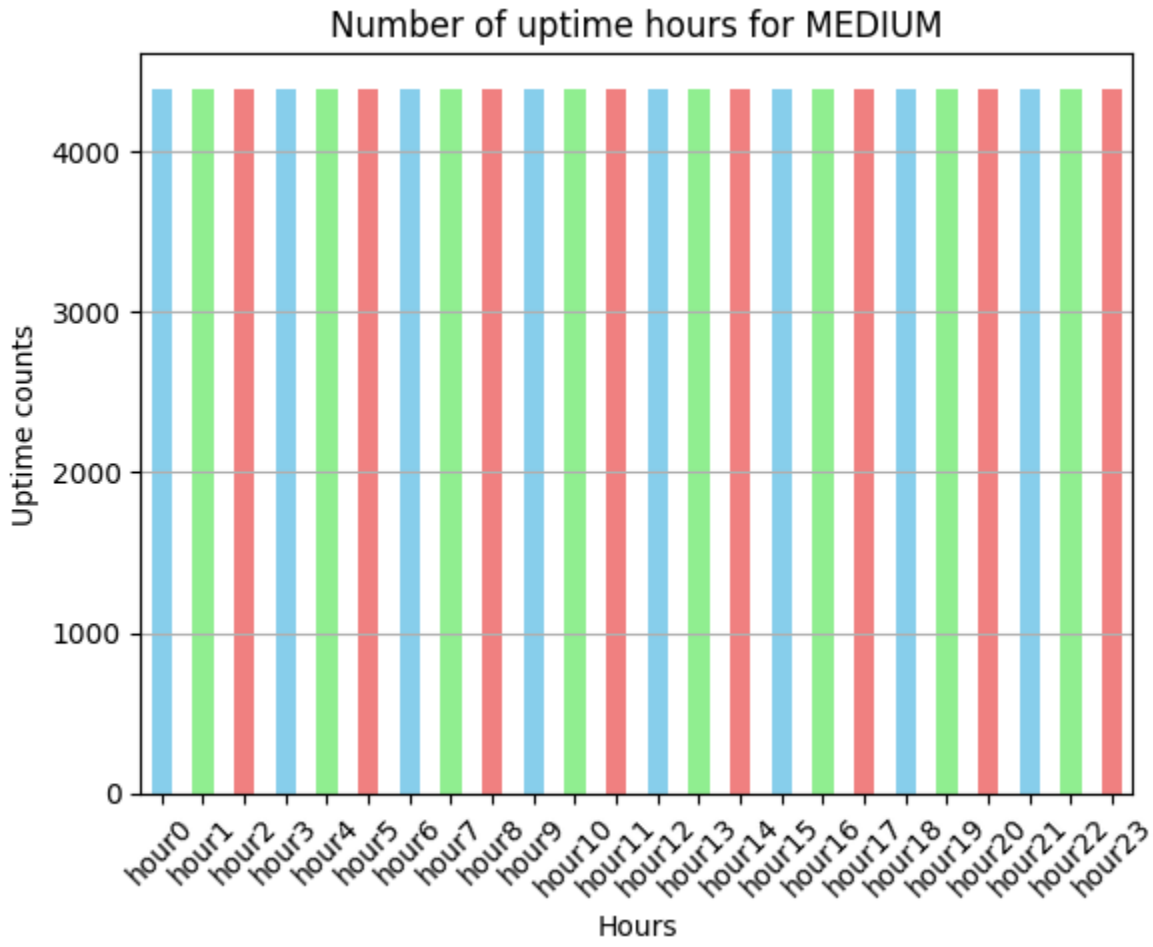
No. of uptime hours for 29th Nov for all the stations-



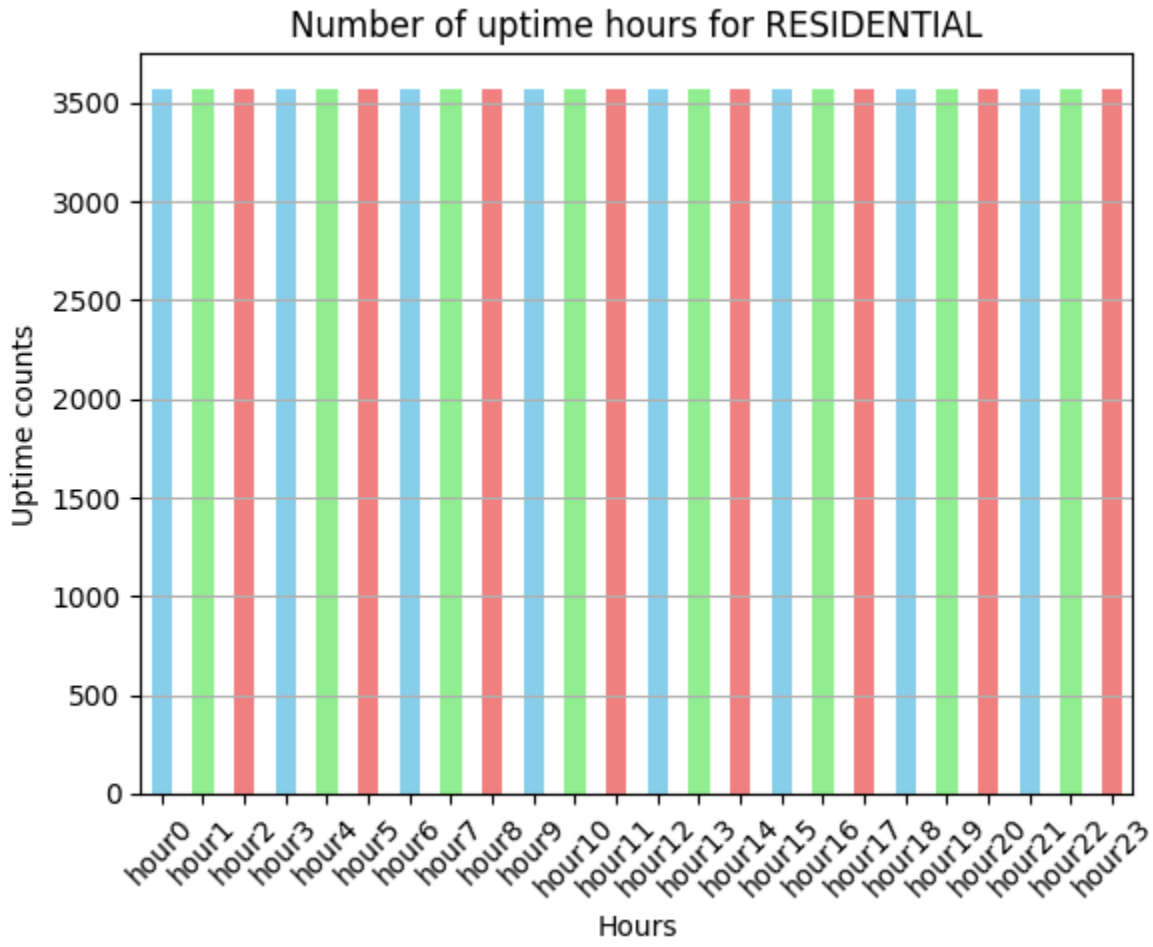
No. of uptime hours for small station category for all the dates-



No. of uptime hours for medium station category for all the dates-



No. of uptime hours for residential station category for all the dates-



Task 2: Performance Comparison of Charger Stations:

- Total uptime of various stations was identified.
- 245 Outlier stations were identified considering all the dates. Some were having very high uptime, and some very low uptime -
    - Station ID: 42997455-b138-4108-ab1b-b207e0dcd416 (Residential, Public) has 11625 hours uptime
    - Station ID: badbd3e9-ffaf-47c5-87a5-7a8c8270f4e5 (Residential, Public) has 7440 hours uptime
    - Station ID: 634177e0f44c2d250a0bfa42 (Independent House, Public) has 465 hours uptime
    - Station ID: 61b65436896e59f10647ada7 (Independent House, Public) has 465 hours uptime.

Screenshots:

Stations with maximum and minimum uptimes-

```
Station with maximum total uptime: Salarpuria Anugraha
Maximum total uptime: 11625
Station with minimum total uptime:  Desai Radiant Developers
Minimum total uptime: 465
```

Stations with highest uptimes-

StationID	Station_Name	Station_Category	Station_Access	
42997455-b138-4108-ab1b-b207e0dcd416	Salarpuria Anugraha	RESIDENTIAL	PUBLIC	11625
badbd3e9-ffaf-47c5-87a5-7a8c8270f4e5	Sattva Park Cubix	RESIDENTIAL	PUBLIC	7440
623aeea2896e59f10634cbeb	SJR BLUE WATERS	RESIDENTIAL	PUBLIC	7440
923fa229-286a-4ba7-a24f-a486c13e0606	SOBHA DREAM ACRES PHASE 1	RESIDENTIAL	PUBLIC	6975
628b7230968b1a96e425c939	Krishna Kunj Gardens	LARGE	PUBLIC	6975

Stations with lowest uptimes-

StationID	Station_Name	Station_Category	Station_Access	total_hour
634177e0f44c2d250a0bfa42	DSR PRATHAMAM APARTMENT	SMALL	PUBLIC	465
61b65436896e59f10647ada7	Pradeep	INDEPENDENT_HOUSE	PUBLIC	465
61e6f864896e59f106dd9597	RAMESH KUMAR	INDEPENDENT_HOUSE	PUBLIC	465
61e6fac1896e59f106ddca8d	VENKATESH	INDEPENDENT_HOUSE	PUBLIC	465
61f0f5b5896e59f1069c82d9	Rohith R Raiker	INDEPENDENT_HOUSE	PUBLIC	465

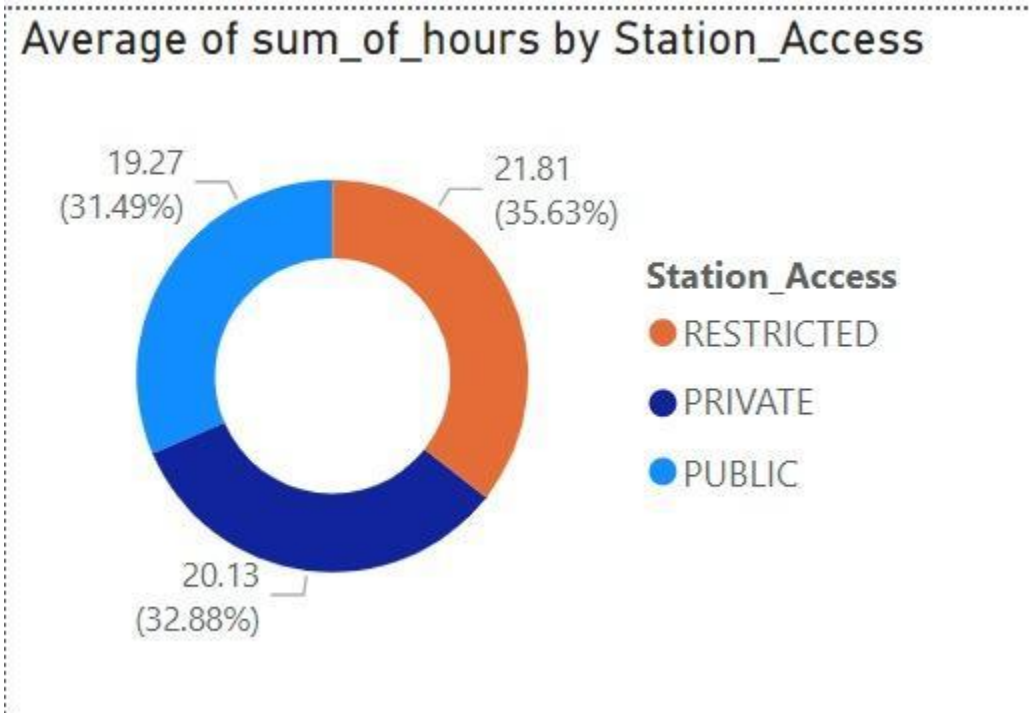
Reasons for outliers:

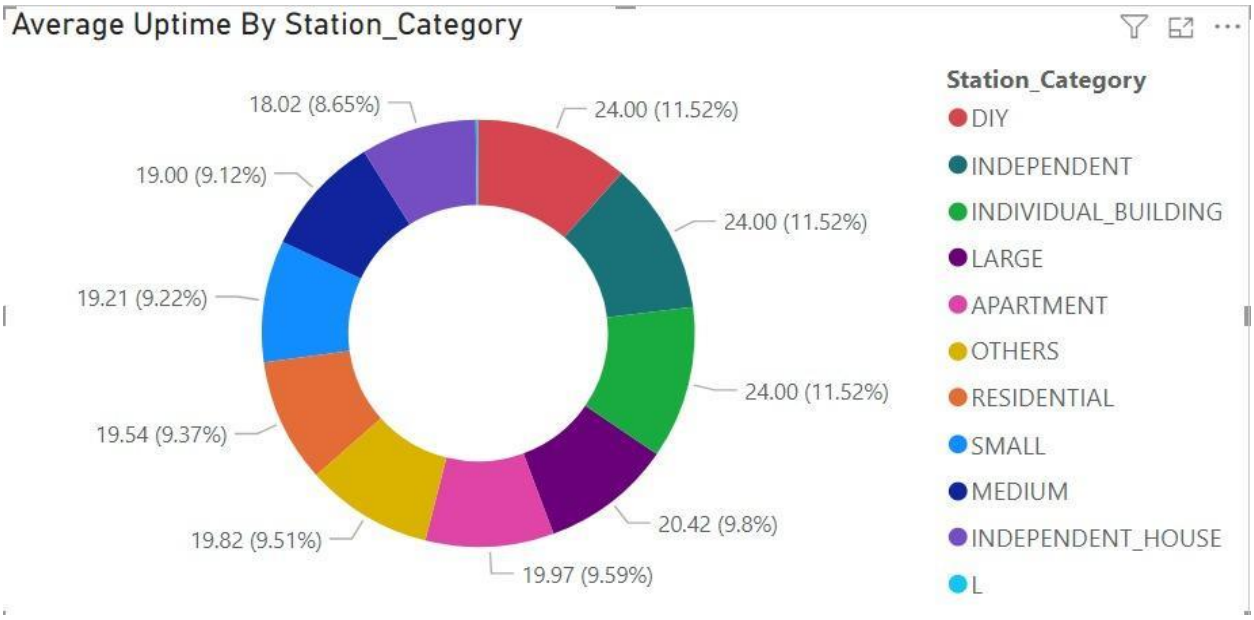
- The reason for high uptime could be -
    - The charger is present in residential areas, and most people in that area own an EV and are highly dependent on the charger. So, they take proper care of the charger and perform timely maintenance.
    - There are very few chargers around the area. So, most people are dependent on the charger.
    - Software is updated and tuned for the optimal performance of the charging station.
    - The charger is present at a very prime location, where many people use it. So, routine maintenance is properly performed.
- The reasons for low uptime could be -
    - Power supply issues
    - Bad weather conditions
    - Maintenance activity around the time
    - Equipment or software malfunction

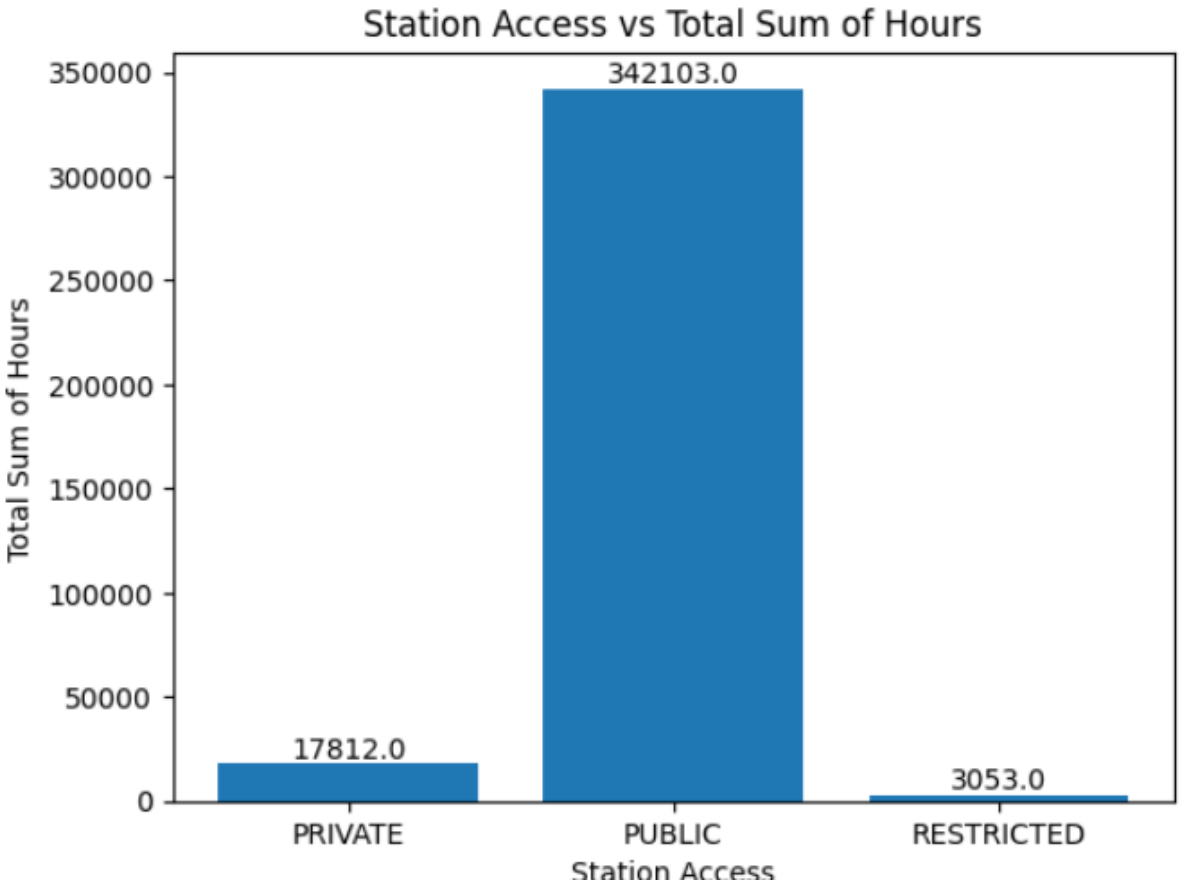
Task 3: Correlation with Station Characteristics

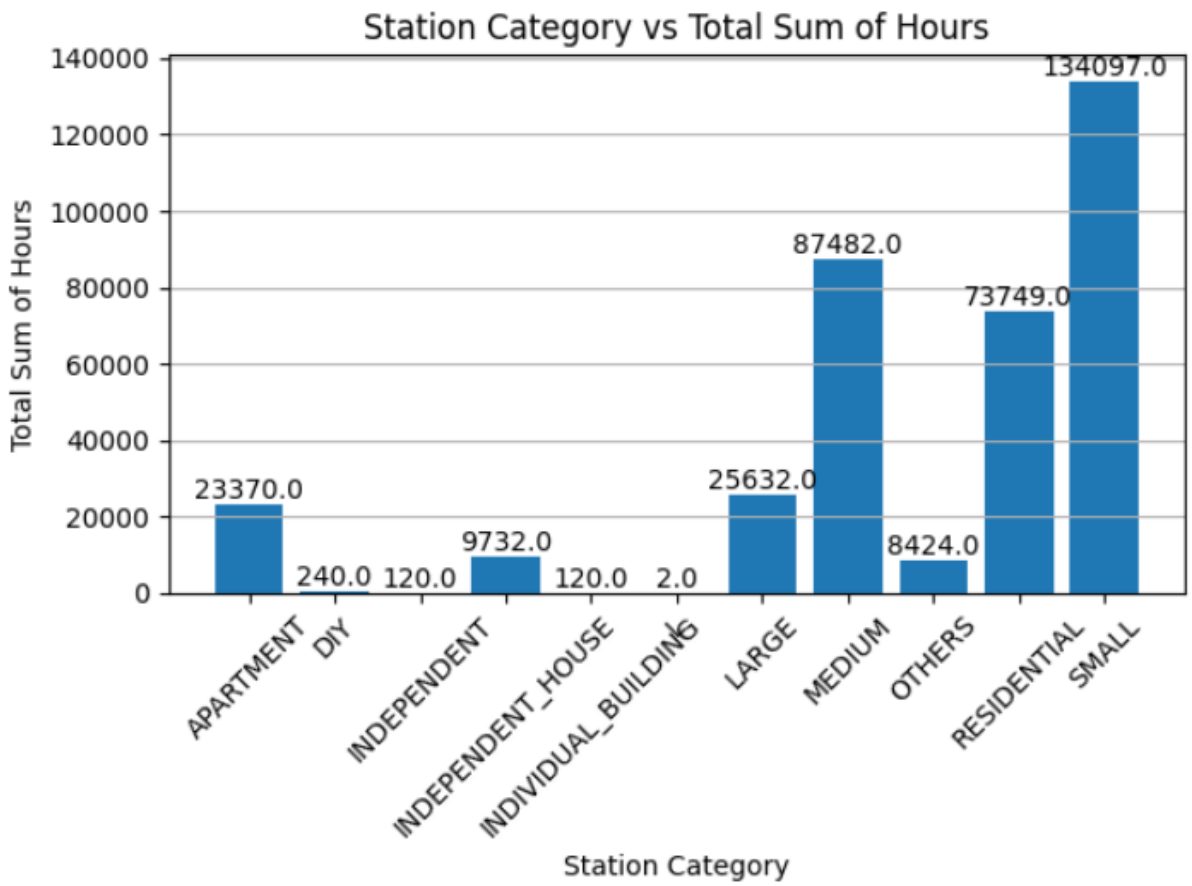
- Restricted Station Access has the highest average uptime with 21.81 hours followed by Private with 20.13 hours. Public Station Access has the lowest average uptime with 19.27 hours.
- DIY, Independent, and Independent Building Station Categories all have the highest average uptime (24 hours). L Category has the lowest average uptime (0.4 hours)

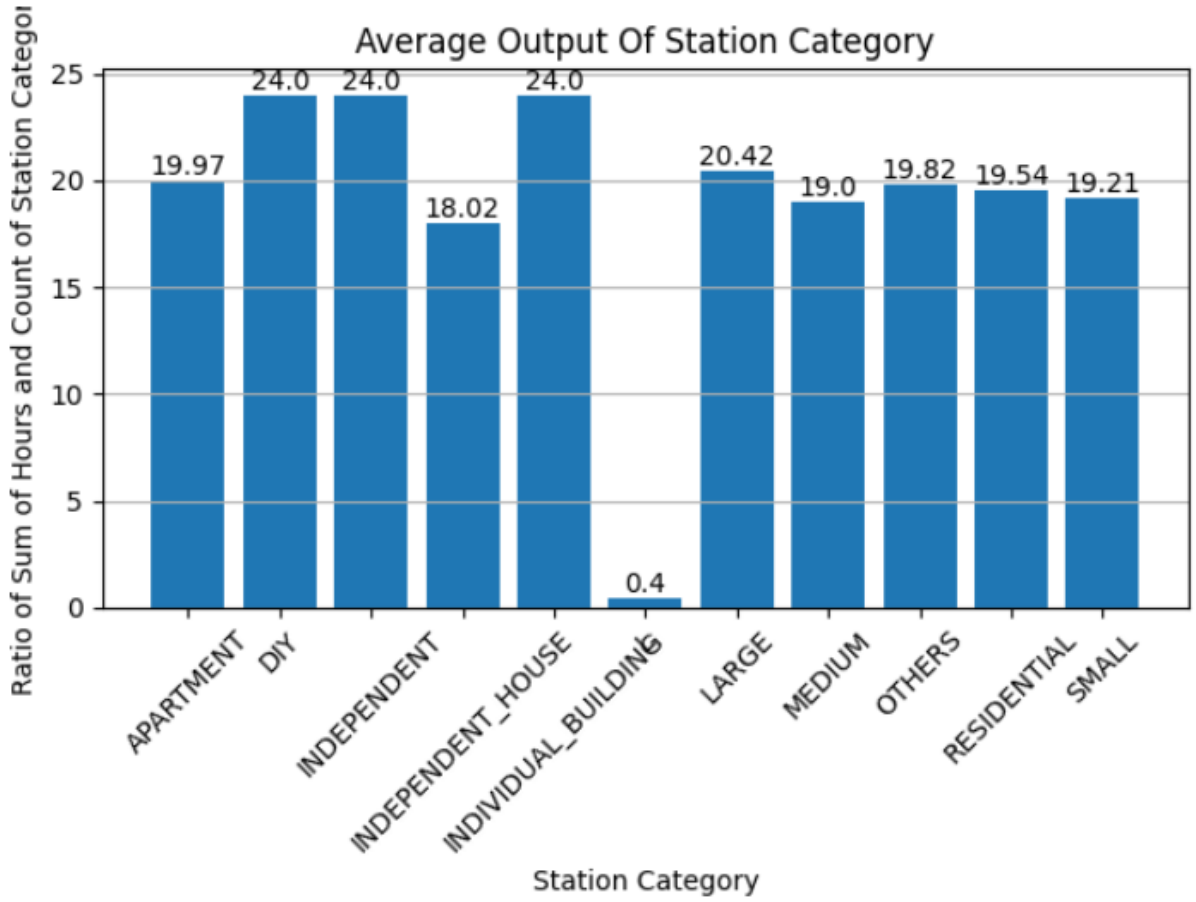
Screenshots:




Average Uptime By Station_Category
18.02 (8.65%)
24.00 (11.52%)
19.00 (9.12%)
24.00 (11.52%)
19.21 (9.22%)
24.00 (11.52%)
19.54 (9.37%)
20.42 (9.8%)
19.82 (9.51%)
19.97 (9.59%)
Station_Category
DIY
INDEPENDENT
INDIVIDUAL_BUILDING
LARGE
APARTMENT
OTHERS
RESIDENTIAL
SMALL
MEDIUM
INDEPENDENT_HOUSE
L


Station Access vs Total Sum of Hours
350000
300000
250000
200000
150000
100000
50000
0
Total Sum of Hours
342103.0
17812.0
3053.0
PRIVATE
PUBLIC
RESTRICTED
Station Access


Station Category vs Total Sum of Hours
Total Sum of Hours
Station Category
140000
120000
100000
80000
60000
40000
20000
0
23370.0
240.0
120.0
9732.0
120.0
2.0
25632.0
87482.0
8424.0
73749.0
134097.0
APARTMENT
DIY
INDEPENDENT
INDEPENDENT_HOUSE
INDIVIDUAL_BUILDING
LARGE
MEDIUM
OTHERS
RESIDENTIAL
SMALL


Average Output Of Station Category
Ratio of Sum of Hours and Count of Station Categor
25
20
15
10
5
0
19.97
24.0
24.0
18.02
24.0
0.4
20.42
19.0
19.82
19.54
19.21
APARTMENT
DIY
INDEPENDENT
INDEPENDENT_HOUSE
INDIVIDUAL_BUILDING
LARGE
MEDIUM
OTHERS
RESIDENTIAL
SMALL
Station Category

Task 4: Predictive Uptime For Future Uptime -

- Things to keep in mind:
    - The date 30th November has null values, so when null values were dropped, there were no rows left for 30th Nov.
    - No major difference was found for uptime of all the chargers over all the days and all the hours. So, Sequential models (RNN, LSTM) were not used.
    - No major difference was found for uptime depending on the hour
    - Dataset was found to be highly imbalanced. High number of uptime and low number of downtime hours were present.
- Necessary libraries were imported for classification.
- Categorical features need to be encoded. One hot encoding would highly increase the dimensionality of the data. So, label encoding has been applied.
- Necessary features that could be used to train the predictive model were identified.
- Training and testing data was split in an 80:20 ratio.
- Using data resampling (upsampling and downsampling) and SMOTE, various models were trained. The new resampled dataset was trained on various models, and highest accuracy was obtained as 82%.
- Without resampling of data, accuracy for SVM was obtained at around 79%. The accuracy for random forest in this case was obtained as 91%.
- Feature importance were identified, where the uptime was found to be highly dependent on the station name, rather than station category and station access.
- Classification report was visualized, in which model was found to be working fine, with not much imbalance.

Screenshots:

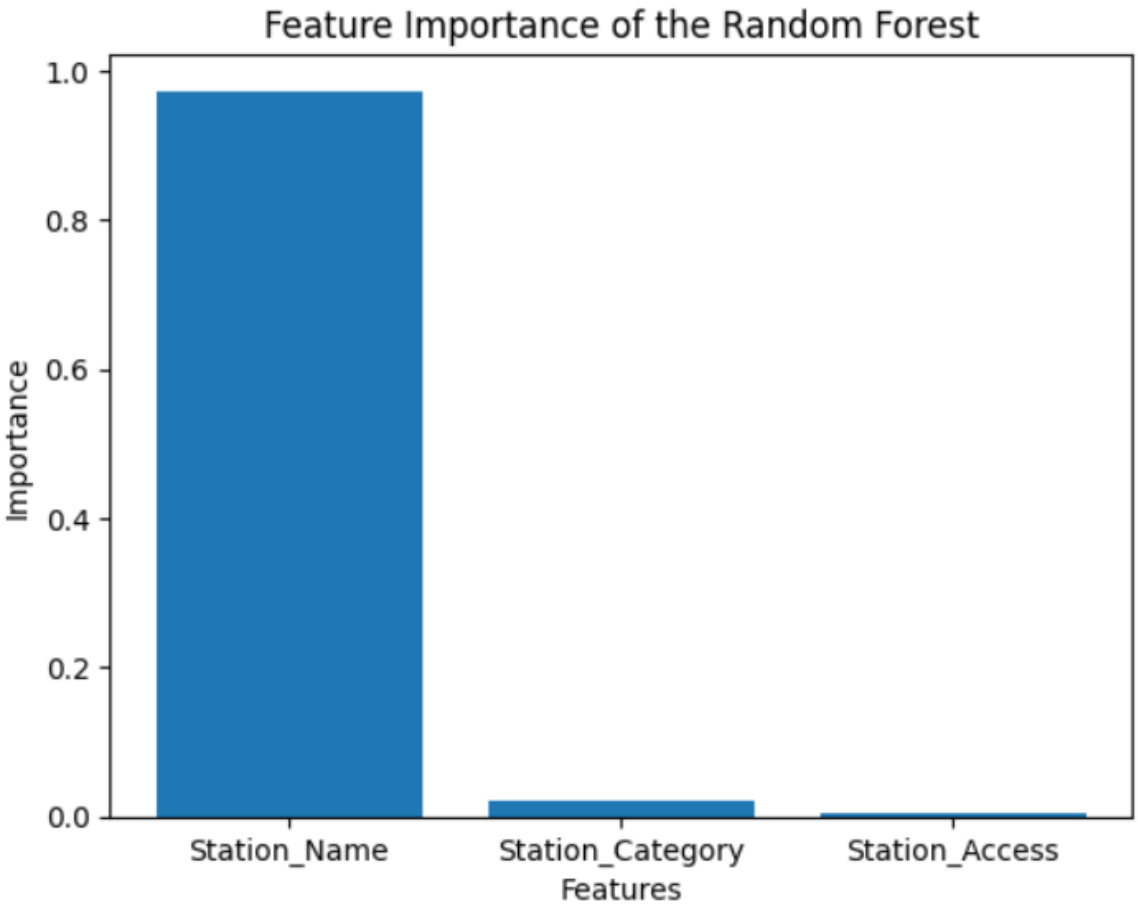
Accuracy of each hour using SVM model with Gaussian Kernel-

```
Validation Accuracy of hour0: 80.00 %
Validation Accuracy of hour1: 80.21 %
Validation Accuracy of hour2: 80.03 %
Validation Accuracy of hour3: 79.95 %
Validation Accuracy of hour4: 79.92 %
Validation Accuracy of hour5: 80.00 %
Validation Accuracy of hour6: 79.97 %
Validation Accuracy of hour7: 78.67 %
Validation Accuracy of hour8: 78.59 %
Validation Accuracy of hour9: 78.56 %
Validation Accuracy of hour10: 78.46 %
Validation Accuracy of hour11: 78.54 %
Validation Accuracy of hour12: 78.75 %
Validation Accuracy of hour13: 78.54 %
Validation Accuracy of hour14: 78.43 %
Validation Accuracy of hour15: 78.24 %
Validation Accuracy of hour16: 78.56 %
Validation Accuracy of hour17: 80.16 %
Validation Accuracy of hour18: 80.37 %
Validation Accuracy of hour19: 80.51 %
Validation Accuracy of hour20: 80.45 %
Validation Accuracy of hour21: 80.27 %
Validation Accuracy of hour22: 80.43 %
Validation Accuracy of hour23: 80.19 %
```

Accuracy of each hour using Random Forest Algorithm-

```
Validation Accuracy of hour0: 92.01 %
Validation Accuracy of hour1: 92.04 %
Validation Accuracy of hour2: 91.90 %
Validation Accuracy of hour3: 92.04 %
Validation Accuracy of hour4: 92.06 %
Validation Accuracy of hour5: 91.85 %
Validation Accuracy of hour6: 91.82 %
Validation Accuracy of hour7: 91.29 %
Validation Accuracy of hour8: 91.19 %
Validation Accuracy of hour9: 90.97 %
Validation Accuracy of hour10: 91.00 %
Validation Accuracy of hour11: 91.16 %
Validation Accuracy of hour12: 91.26 %
Validation Accuracy of hour13: 90.89 %
Validation Accuracy of hour14: 90.89 %
Validation Accuracy of hour15: 90.76 %
Validation Accuracy of hour16: 90.95 %
Validation Accuracy of hour17: 91.58 %
Validation Accuracy of hour18: 91.56 %
Validation Accuracy of hour19: 91.64 %
Validation Accuracy of hour20: 91.80 %
Validation Accuracy of hour21: 91.88 %
Validation Accuracy of hour22: 91.72 %
Validation Accuracy of hour23: 91.77 %
```

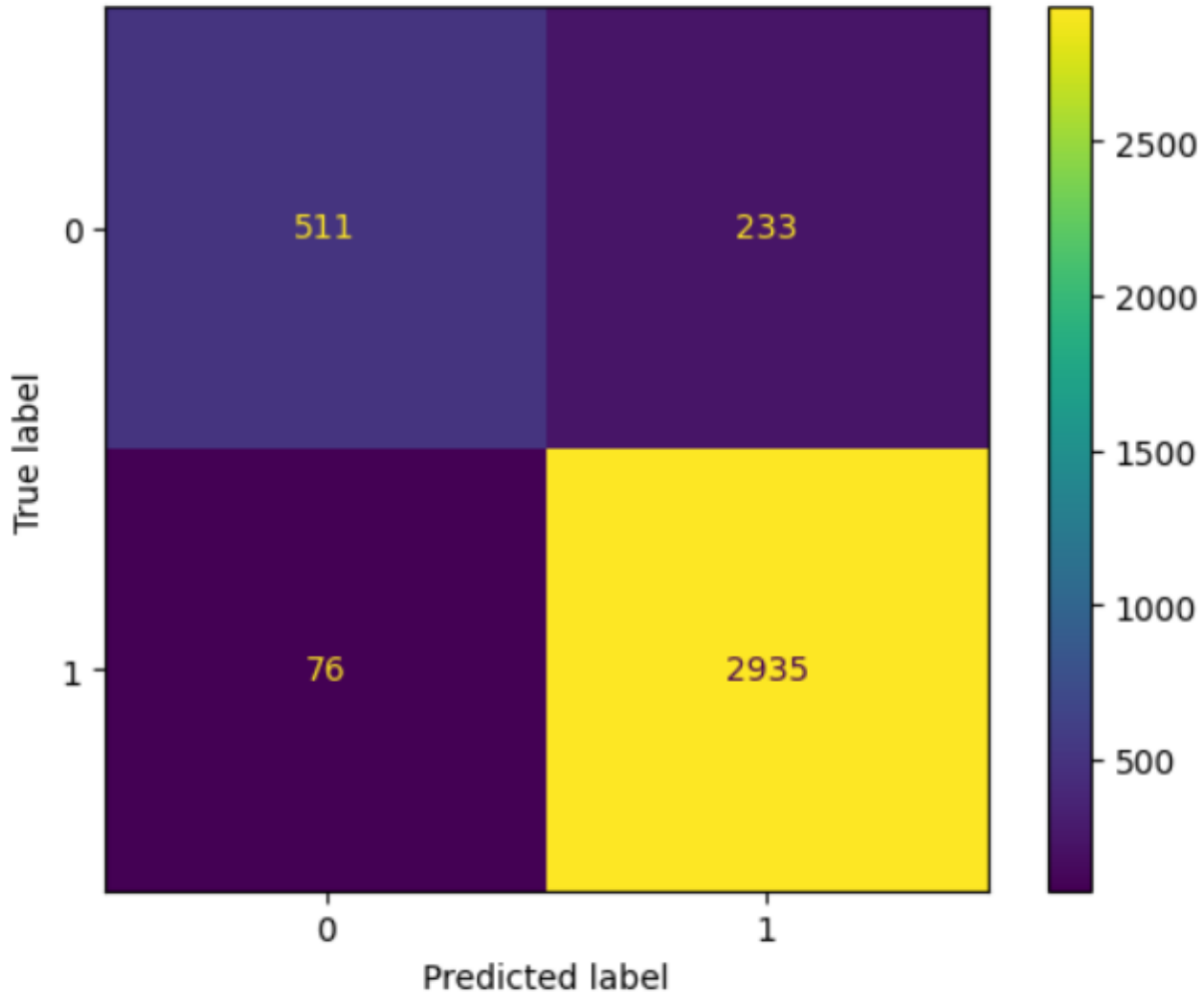
Feature Importance Plot for random forest model-



Classification Report for random forest model-

	precision	recall	f1-score	support
0.0	0.87	0.69	0.77	744
1.0	0.93	0.97	0.95	3011
accuracy			0.92	3755
macro avg	0.90	0.83	0.86	3755
weighted avg	0.92	0.92	0.91	3755

Confusion Matrix for random forest model-



Task 5: Open Ended Exploration

- Additional plots have been plotted on Power BI to analyze data better.

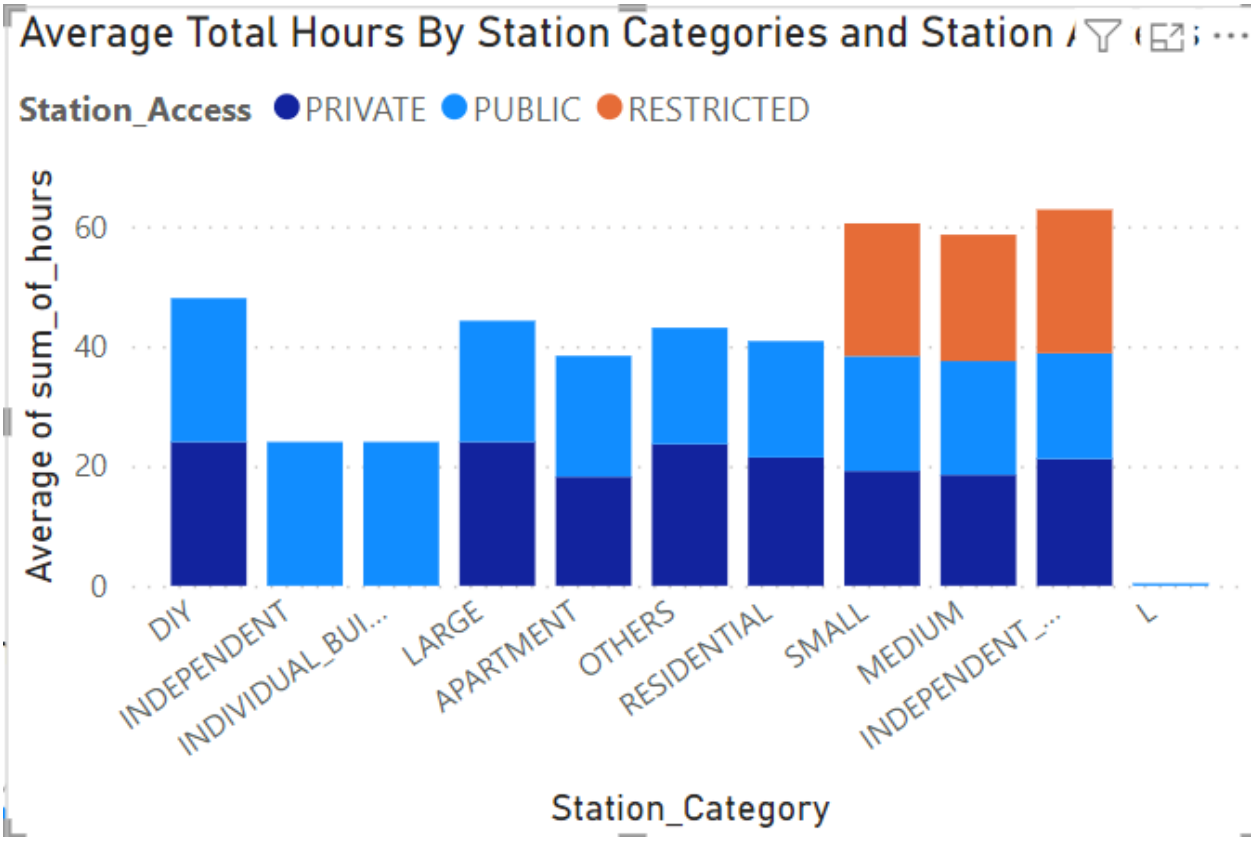


Figure 1 (Bar graph)

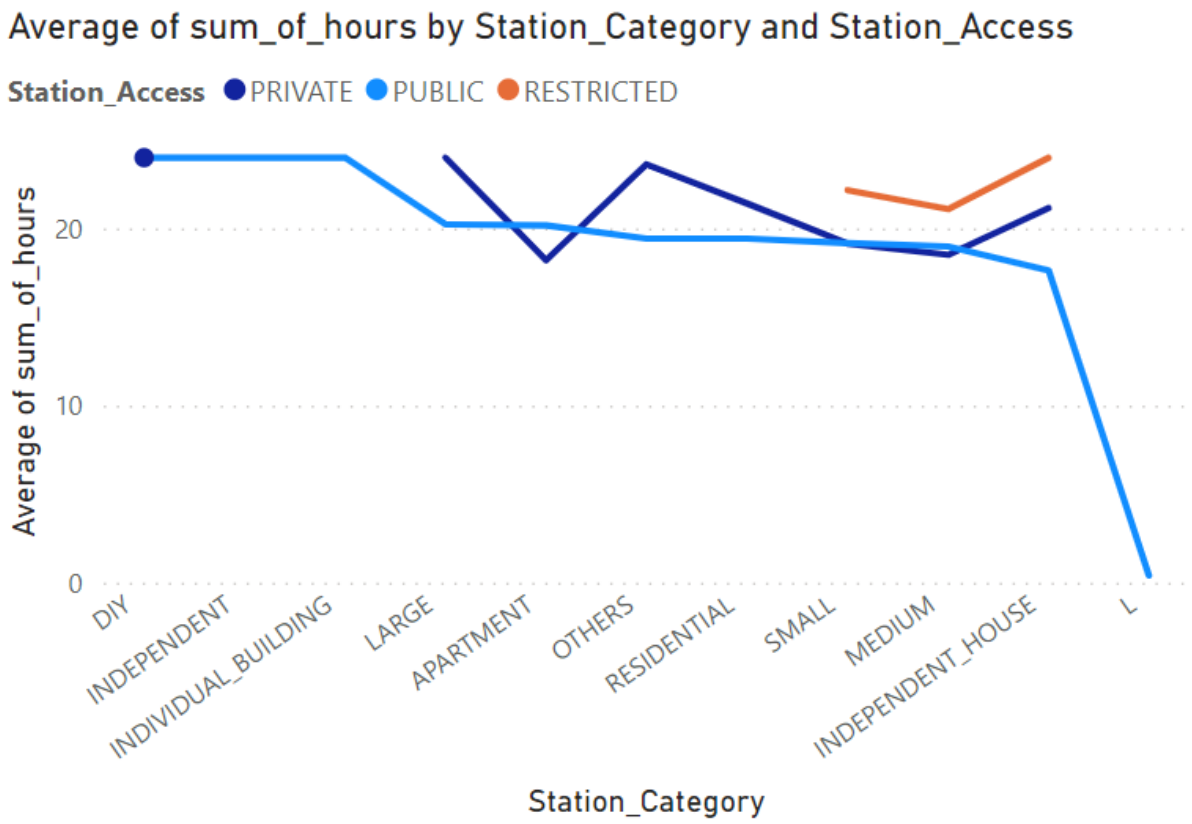


Figure 2 (Line plot)

From Figure 1 and Figure 2 we observe DIY and Independent have the highest Average sum of hours. While L and Independent House have the lowest sum of hours.

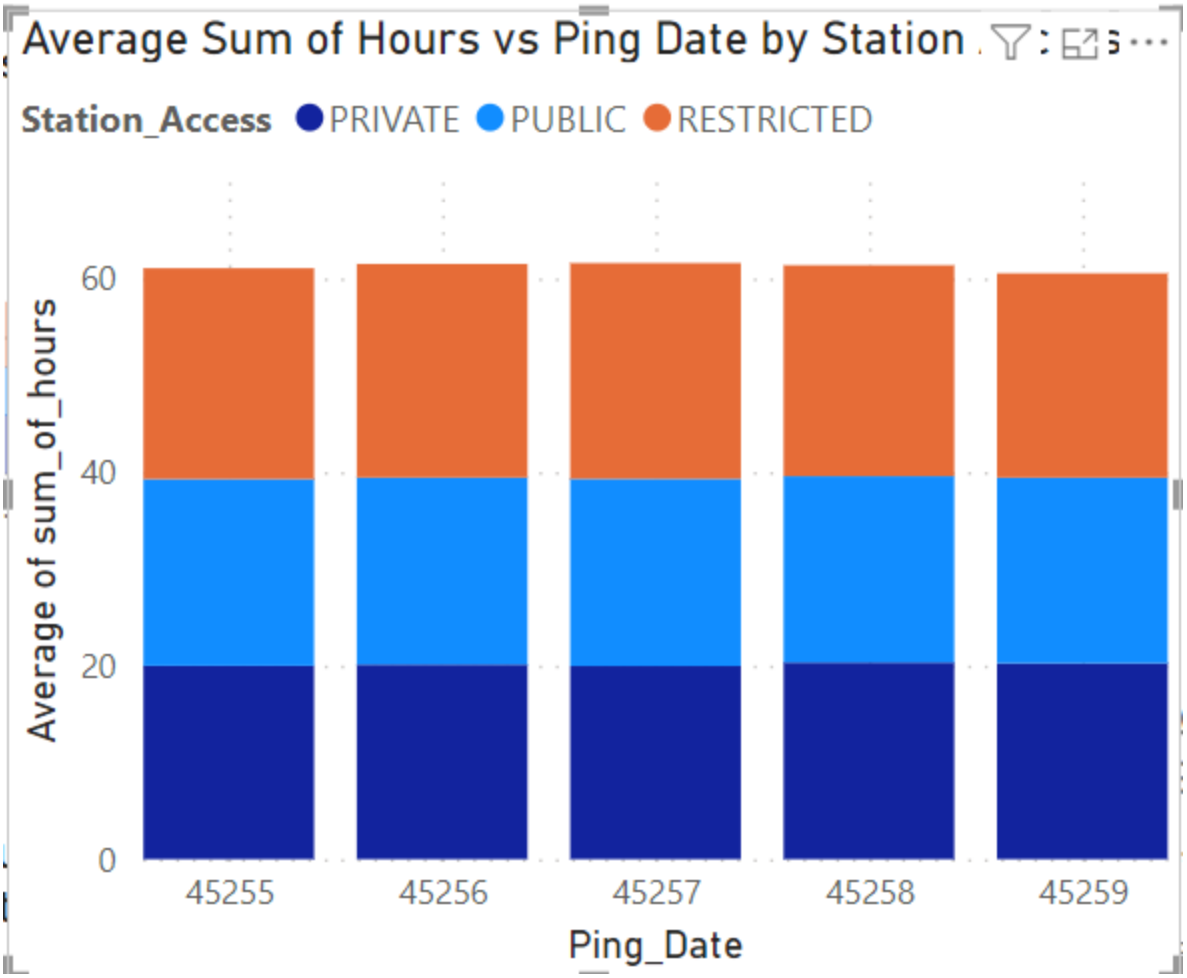


Figure 3

From Figure 3 we infer that for any given date, the Average sum of hours with all Station Access combined is almost the same.

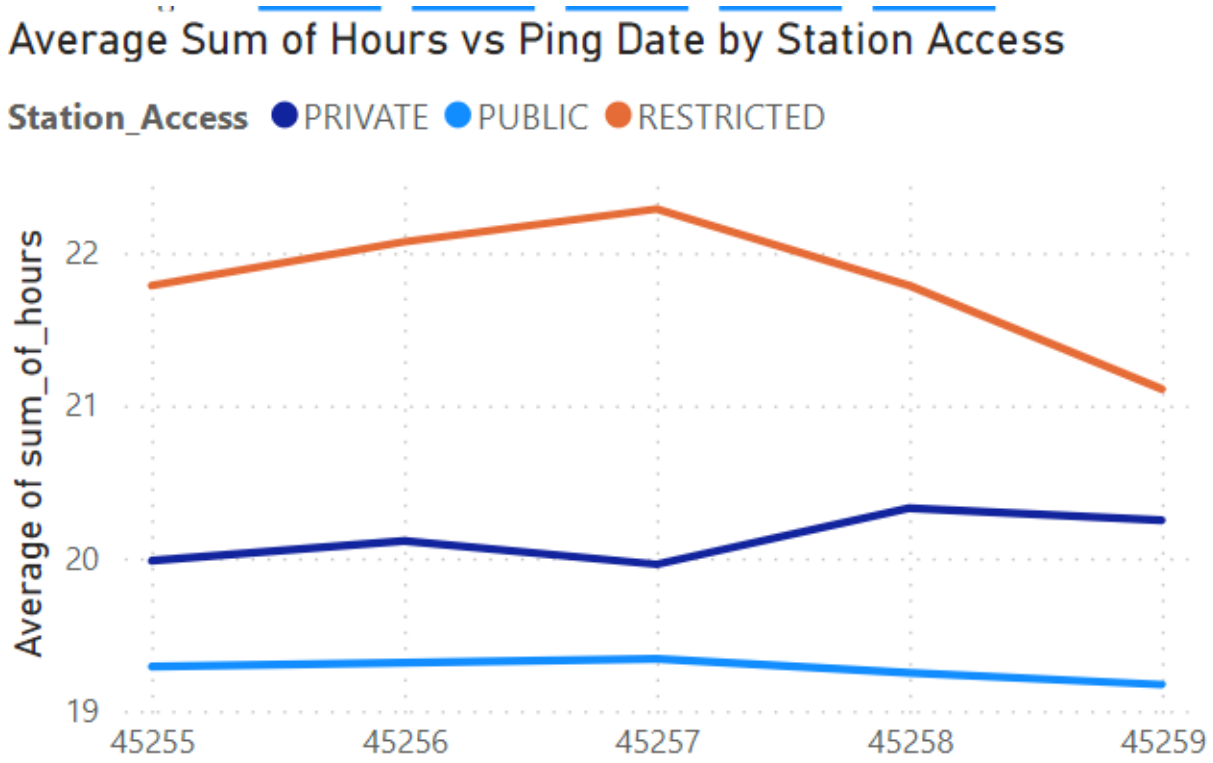


Figure 4

From Figure 4 we can infer that Restricted has a higher sum of hours for any Ping Date, followed by Private and Public.

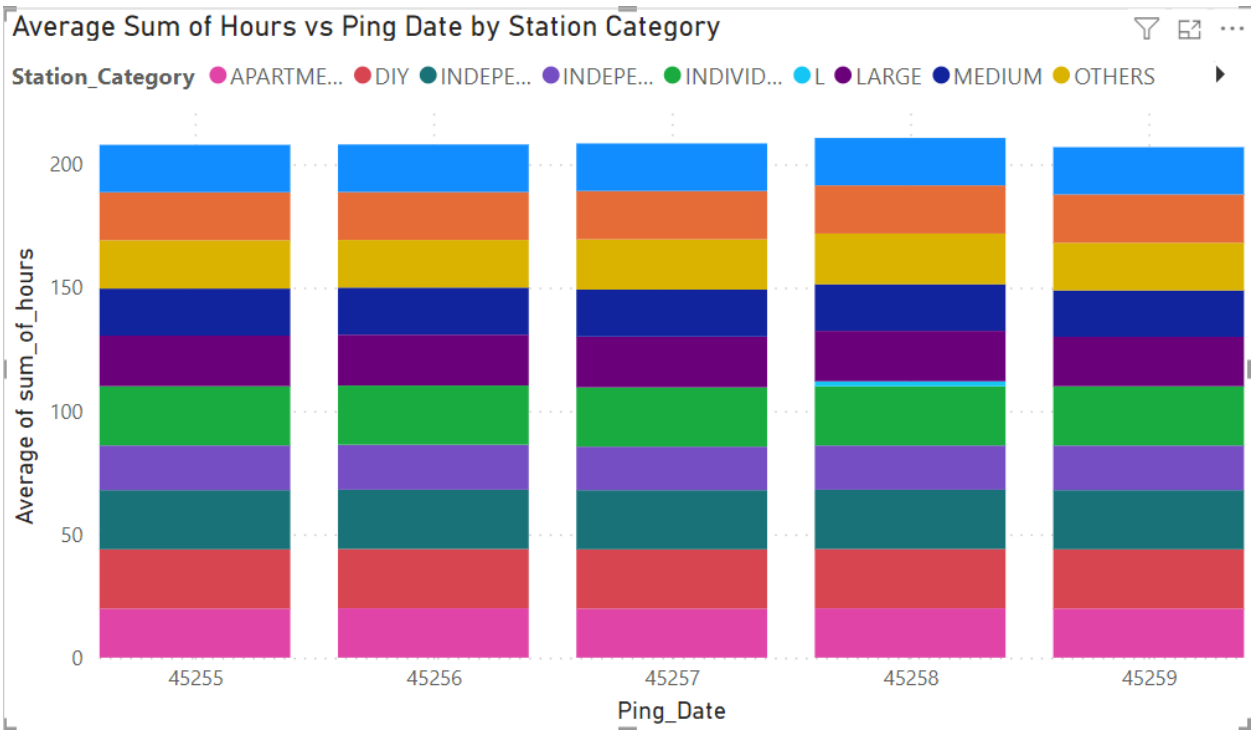


Figure 5

From Figure 3 we infer that for any given date, the Average sum of hours with all Station Categories combined is almost the same.

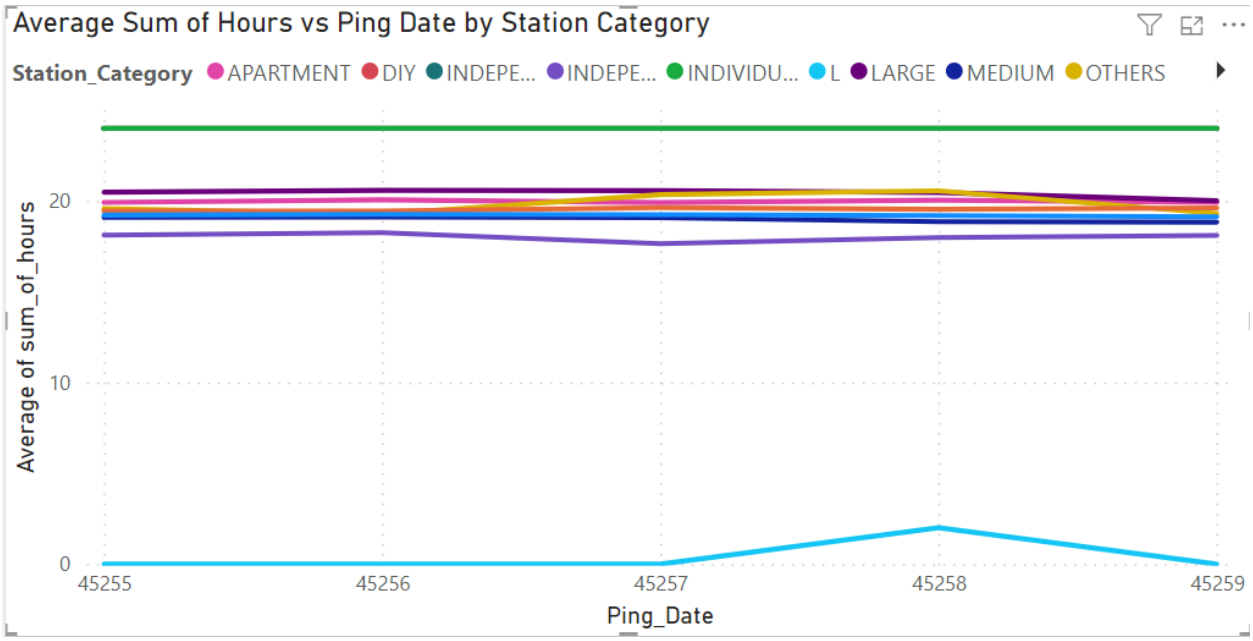


Figure 6

From Figure 6 we infer that Individual buildings have higher Average Sum of Hours while L has the lowest Average sum of hours.

Key takeaways:

1. Maintain L category charging stations. They have the least average uptime. Issues with them must be analyzed.
2. The company should focus on DIY, Independent, and Individual Building Station Categories. These are the stations with the highest average uptimes, and thus have the potential to generate the maximum revenue for the company.
3. Restricted Station Access generates the highest average revenue. But, they are the least in number. So, the company should focus on increasing their numbers to improve the revenue.

Source Code Link-

https://colab.research.google.com/drive/1En8VaE4NpOA-2HbPlQlubvJaQzfnhSVd?usp=sharing

https://colab.research.google.com/drive/1pxssEDOZozgf0Zm3nmsjcgECM7L8OagD?usp=sharing

https://colab.research.google.com/drive/1DogfjkJ2A399Uqfs0t2MpIrwvmQXFkGZ?usp=sharing