ANALYSIS OF EV STATIONS ACROSS THE U.S.

Department of Applied Data Science, San Jose State University

DATA 230: Business Intelligence & Data Visualization

Prof. Venkata Duvvuri

Team Members (Group 7)

- Shatayu Thakur - 017451096
- Padmanabh Rathi - 018180006
- Rupeshkumar Motipawale - 018187845
- Mayuresh Pandey - 018196750

Business Problem

- Drive growth by targeting underserved regions based on data-driven trends.
- Identify gaps in sustainability efforts and promote the adoption of eco-friendly energy solutions in the EV charging network.
- Are there enough fast charging stations to meet the growing demand?
- Are charging stations dominated by a particular category of owners?

Dataset Used

- Source: U.S. Department of Energy – Alternative Fuels Data Center.
- URL: https://afdc.energy.gov
- Records Count: 78,837

Team Plan

- Data collection and cleaning.
- EDA and feature engineering.
- Dashboard design and visualization.
- Advanced analytics.
- Aesthetic Refinements.

EDA

- Null Value Check for Key Columns
- Understanding the data distribution
- Duplicate Check
- Colab file link : https://colab.research.google.com/drive/1oG_bJiAL5pFXMYQX18lJMk7y0tL6lrln?usp=sharing#scrollTo=faJbRYog4uvo

Null Value Check for Key Columns

Missing Data Overview:

	Missing Count	Missing Percentage
EV On-Site Renewable Source	78907	99.485595
EV Level1 EVSE Num	78617	99.119965
Federal Agency Name	78290	98.707685
Federal Agency ID	78290	98.707685
Federal Agency Code	78290	98.707685
Access Detail Code	74822	94.335246
Expected Date	73700	92.920633
Cards Accepted	71518	90.169577
Restricted Access	66811	84.235012
EV DC Fast Count	66799	84.219883
EV Pricing	63782	80.416063
Facility Type	59921	75.548131
Owner Type Code	58318	73.527076
EV Level2 EVSE Num	11389	14.359201
EV Network Web	9639	12.152808
Access Days Time	7986	10.068713
Open Date	441	0.556011
Date Last Confirmed	156	0.196684
Street Address	34	0.042867
EV Workplace Charging	11	0.013869
City	4	0.005043
Station Name	3	0.003782
Status Code	0	0.000000
Updated At	0	0.000000
State	0	0.000000
Access Code	0	0.000000
Longitude	0	0.000000
Latitude	0	0.000000
ID	0	0.000000
Geocode Status	0	0.000000
Fuel Type Code	0	0.000000
EV Network	0	0.000000
EV Connector Types	0	0.000000
Country	0	0.000000
ZIP	0	0.000000

Data Distribution (1)

Descriptive Statistics for Numerical Columns:

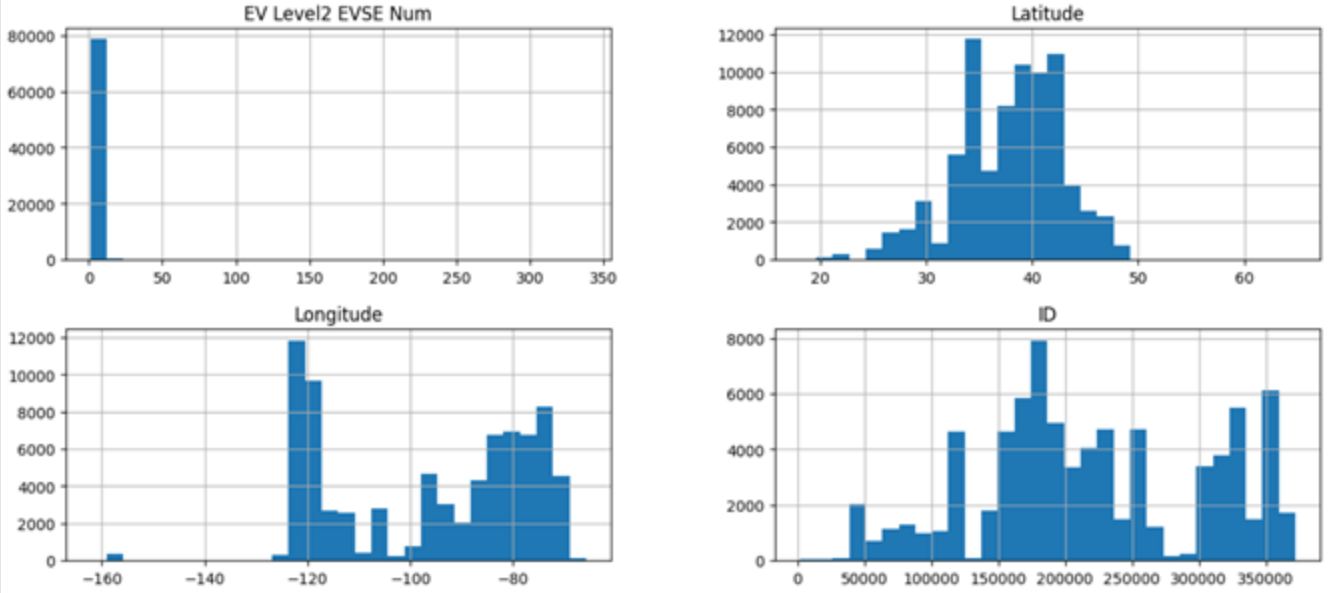
	EV Level2 EVSE Num	Latitude	Longitude
count	79315.000000	79315.000000	79315.000000
mean	2.551291	37.870088	-96.199288
min	1.000000	18.009854	-162.286348
25%	2.000000	34.043346	-117.881102
50%	2.000000	38.596430	-91.068251
75%	2.551291	41.591639	-78.656218
max	338.000000	64.852466	-65.756678
std	3.357435	5.041741	19.380892

	Date Last Confirmed	ID
count	79159	79315.000000
mean	2024-09-07 19:18:28.346492416	220069.654038
min	2019-12-12 00:00:00	1517.000000
25%	2024-10-11 00:00:00	165452.500000
50%	2024-12-02 00:00:00	205580.000000
75%	2024-12-02 00:00:00	308834.500000
max	2024-12-02 00:00:00	372237.000000
std	NaN	85900.785636

	Open Date	EV Workplace Charging
count	78874	79315.000000
mean	2021-03-28 11:03:17.631665408	0.019268
min	1995-08-30 00:00:00	0.000000
25%	2020-06-12 00:00:00	0.000000
50%	2021-10-23 00:00:00	0.000000
75%	2023-06-16 00:00:00	0.000000
max	2025-01-15 00:00:00	1.000000
std	NaN	0.137455

Data Distribution (2)

```
import matplotlib.pyplot as plt

# Plot histograms for all numerical columns
df_cleaned.select_dtypes(include=['float64', 'int64']).hist(figsize=(15, 10), bins=30)
plt.suptitle("Histograms for Numerical Columns", fontsize=16)
plt.show()
```



Duplicates Check

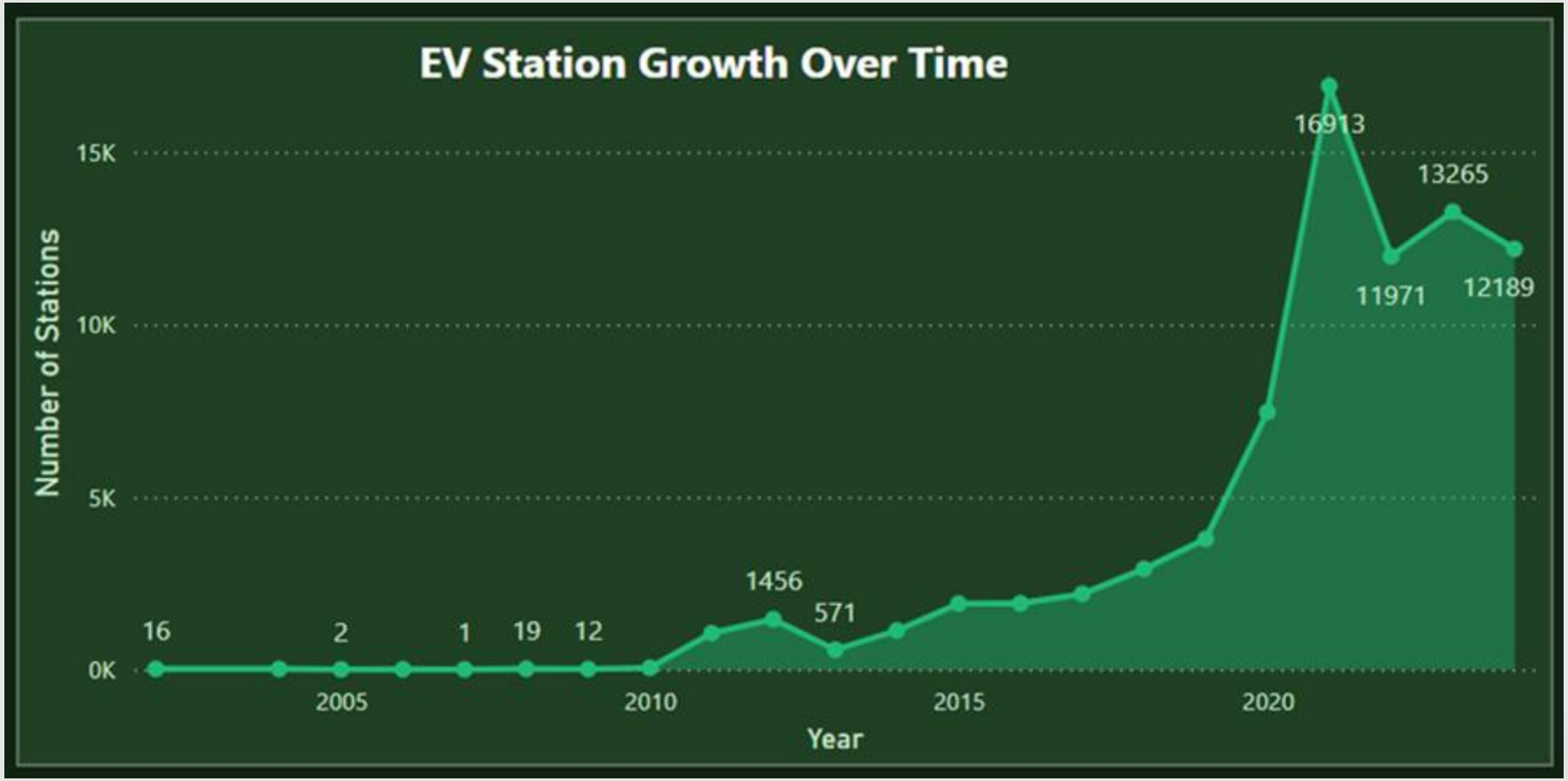
```
# Check for duplicate rows
print(f"Number of duplicate rows before removal: {df_cleaned.duplicated().sum()}")

# Remove duplicate rows
df_no_duplicates = df_cleaned.drop_duplicates()

# Display results
print(f"Shape before duplicate removal: {df_cleaned.shape}")
print(f"Shape after duplicate removal: {df_no_duplicates.shape}")
```

```
Number of duplicate rows before removal: 0
Shape before duplicate removal: (79315, 22)
Shape after duplicate removal: (79315, 22)
```

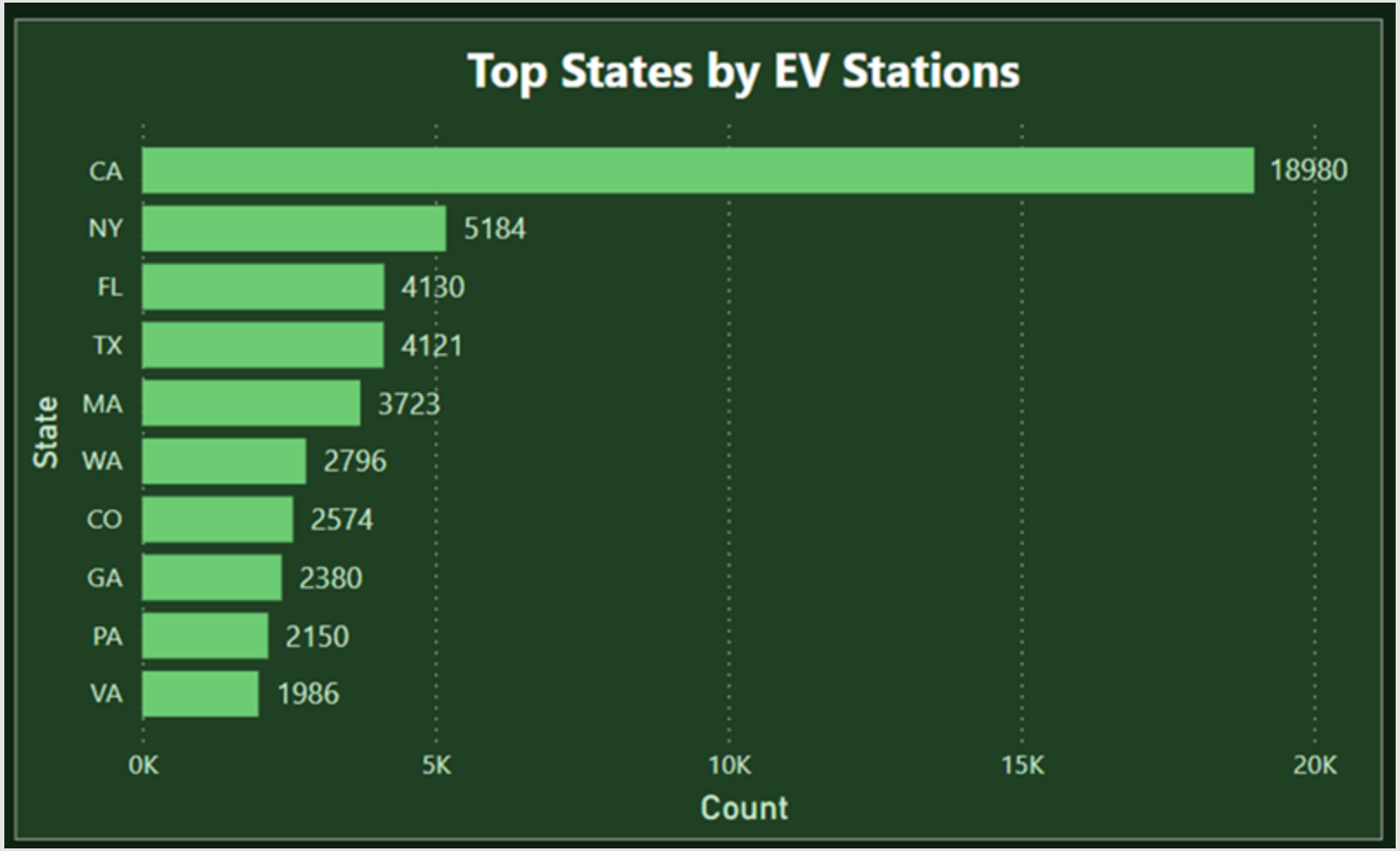
Analysis Performed

- Trend Analysis over Time.
- Access Type Distribution.
- Top States by EV Station Utilization
- Availability of Fast Chargers.
- Classification of Key Stakeholders.
- ML Forecasting of Future Expansion.

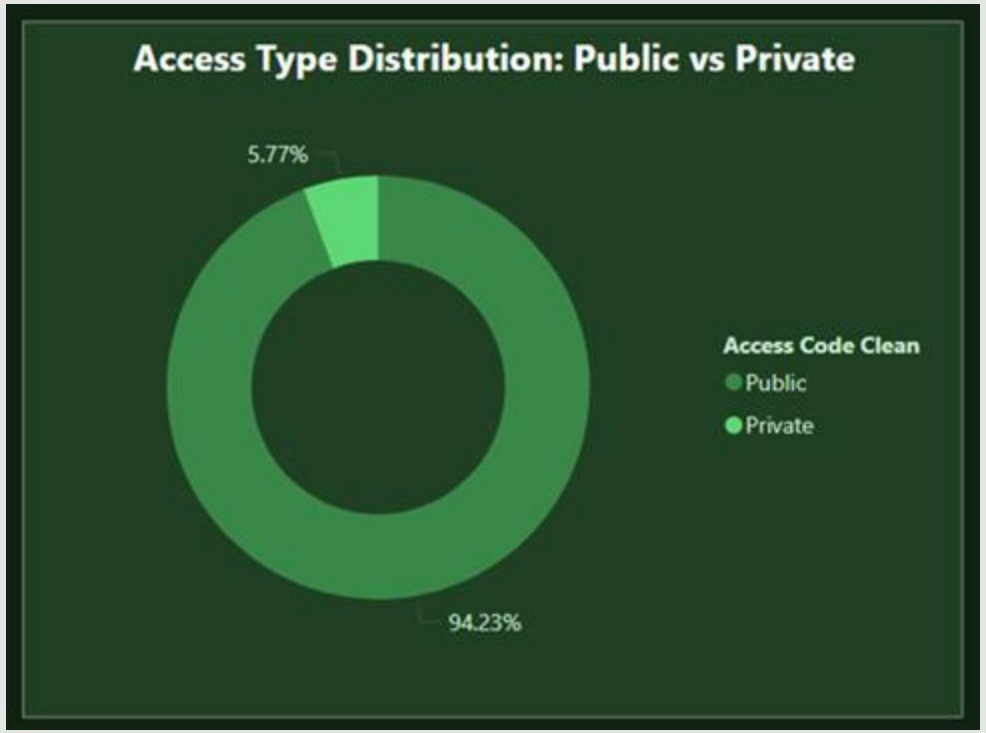
Area Chart

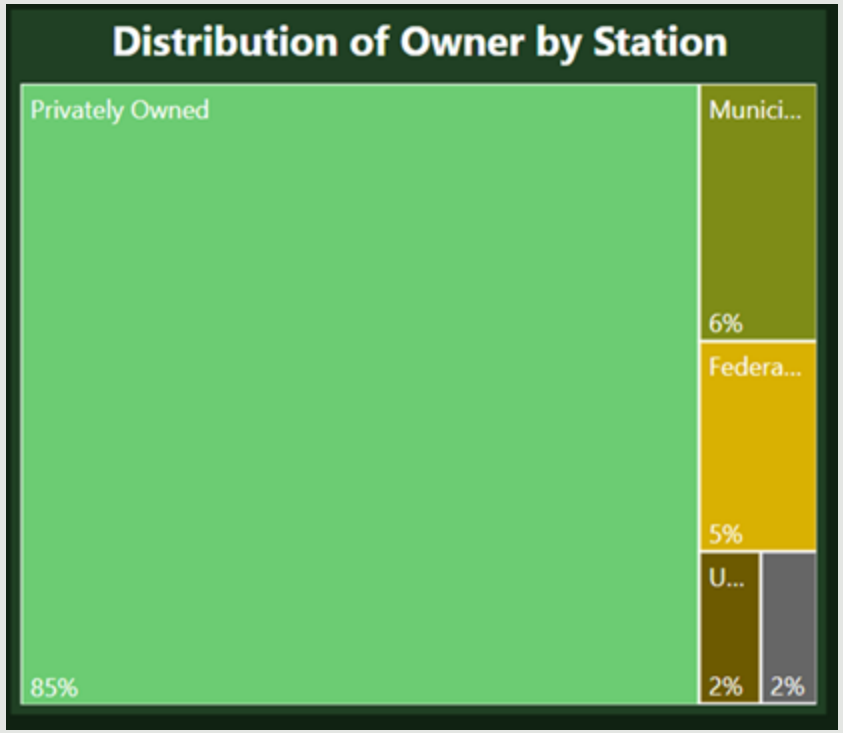

EV Station Growth Over Time
15K
10K
5K
0K
Number of Stations
16
2
1
19
12
1456
571
16913
13265
11971
12189
2005
2010
2015
2020
Year

Bar Chart

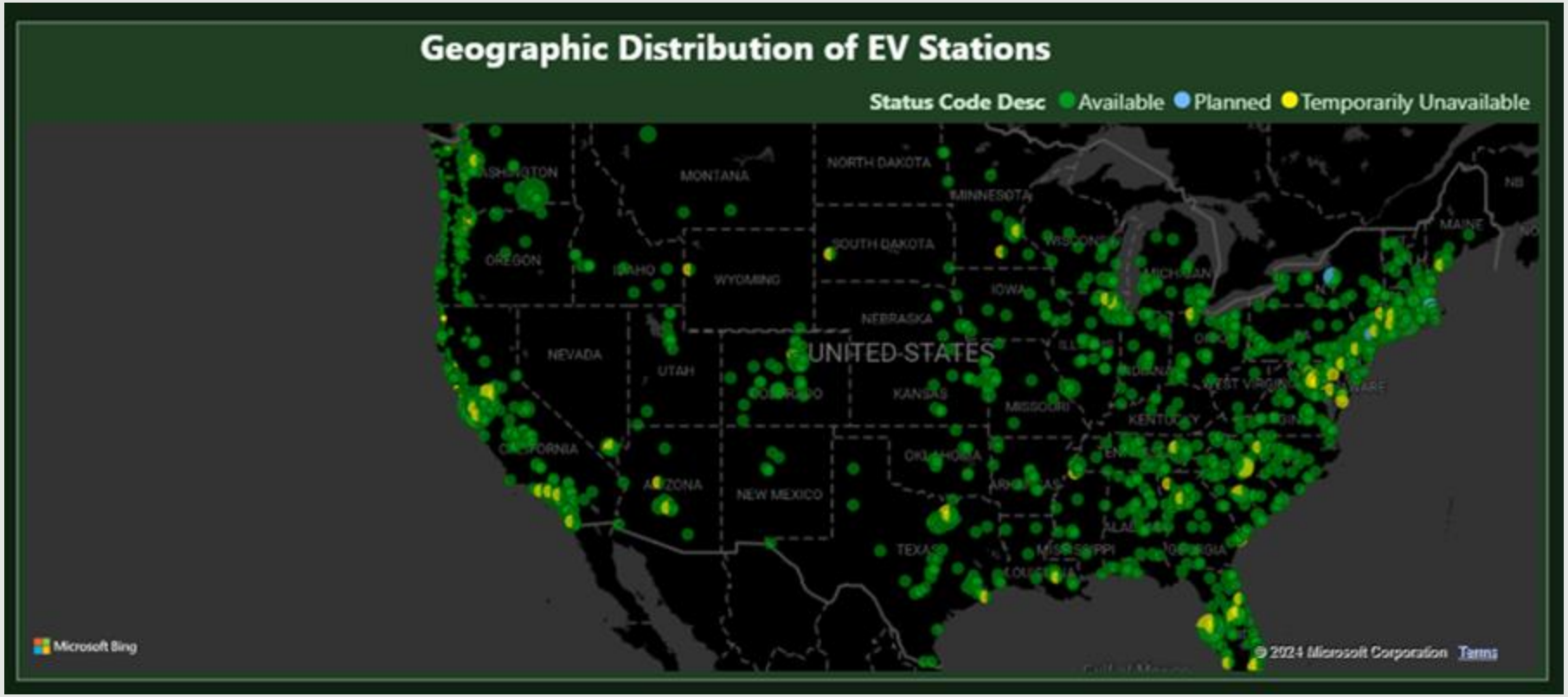

Top States by EV Stations
CA
NY
FL
TX
MA
WA
CO
GA
PA
VA
18980
5184
4130
4121
3723
2796
2574
2380
2150
1986
State
0K
5K
10K
15K
20K
Count

Donut Chart & Tree Map



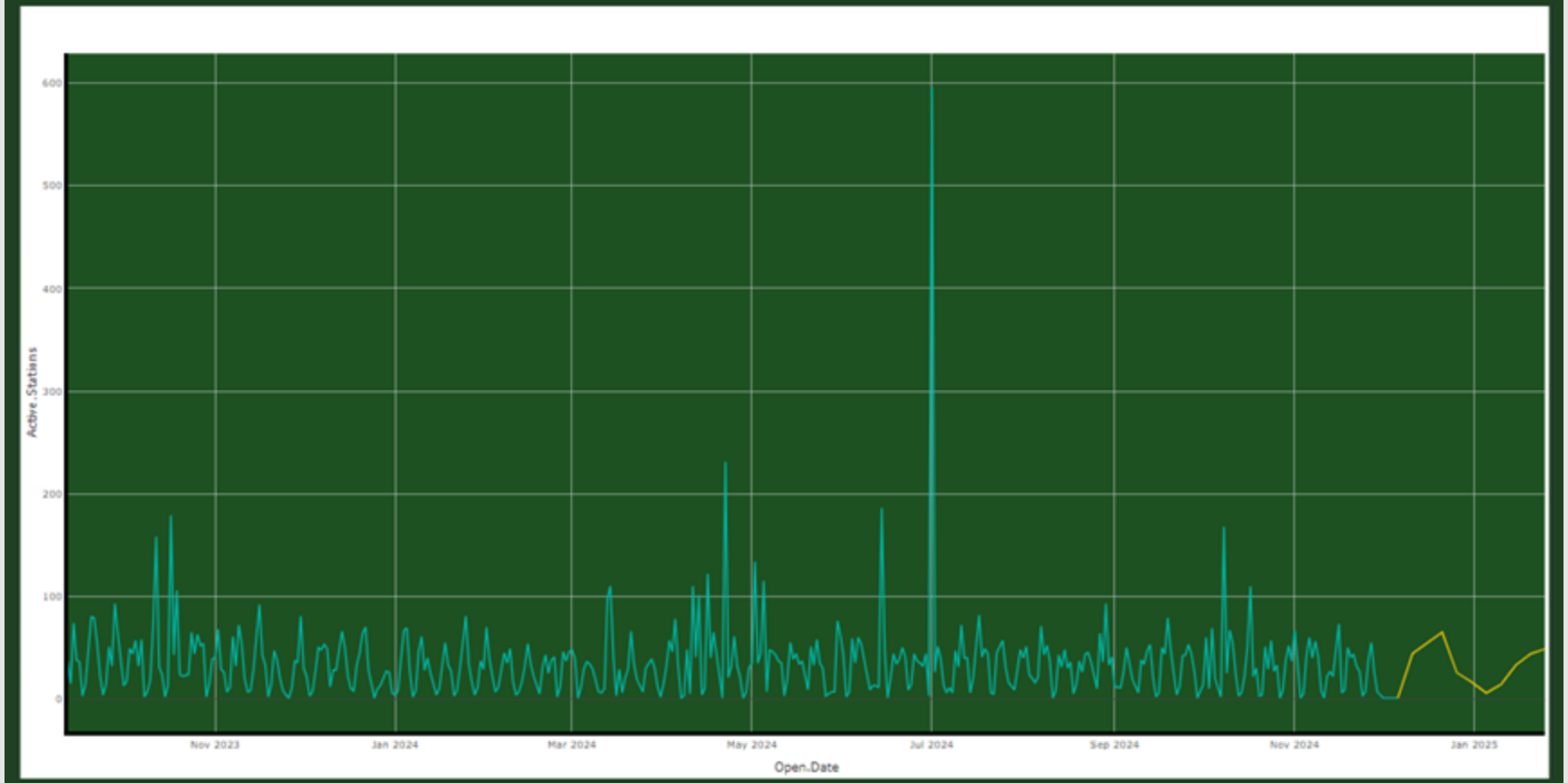


Geospatial Map

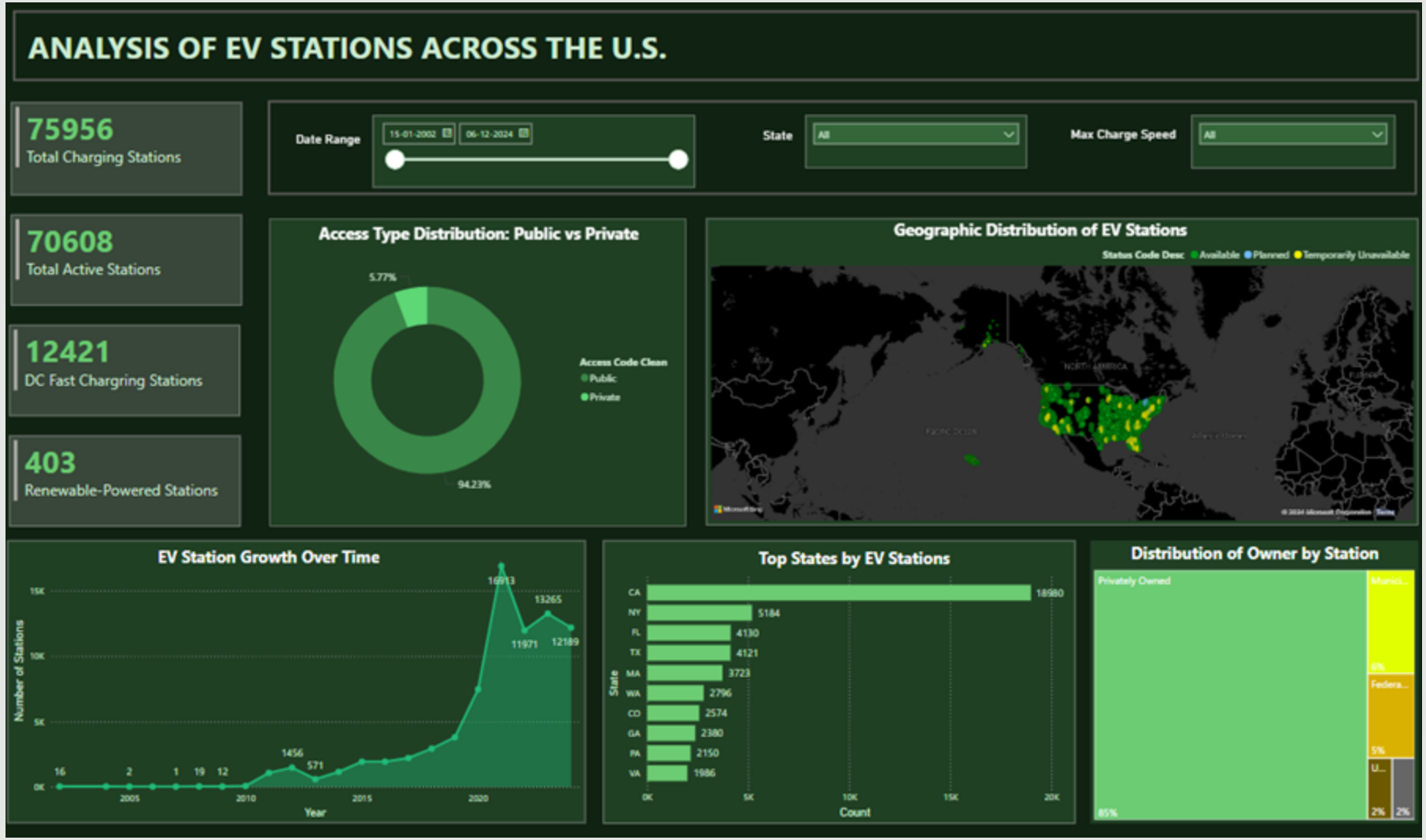


ML Prediction using MAQ Chart

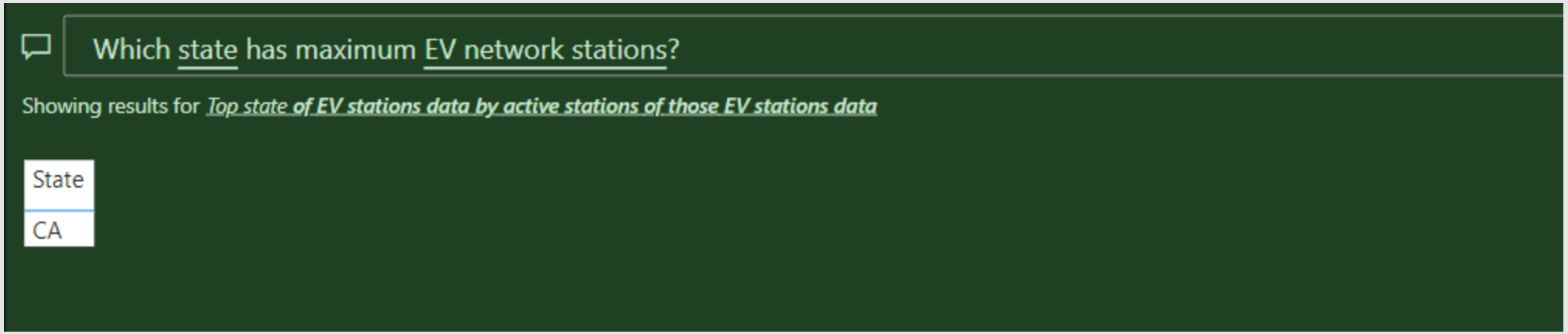
- Number of Station Prediction using R



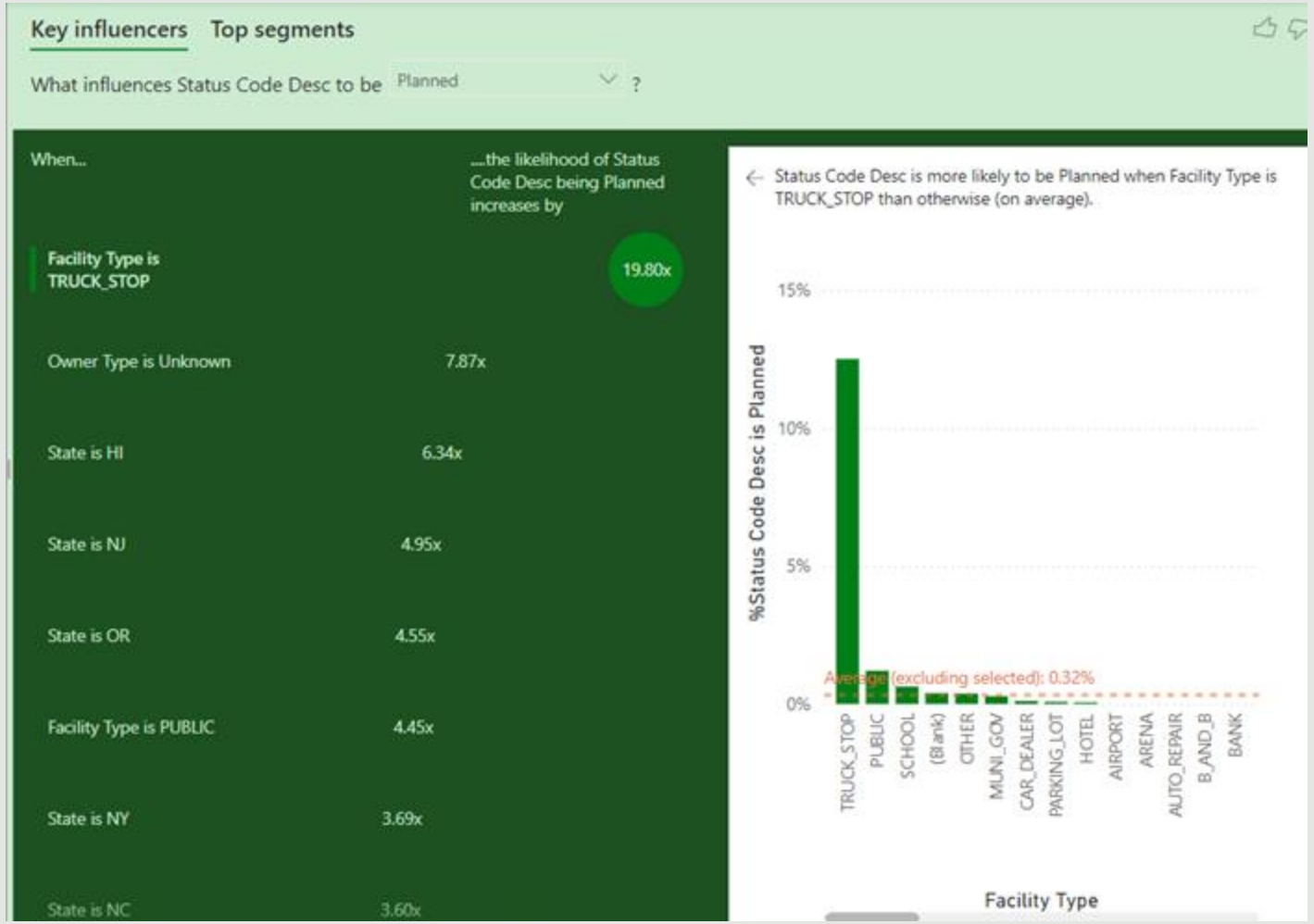
Dashboard View (Page 1)



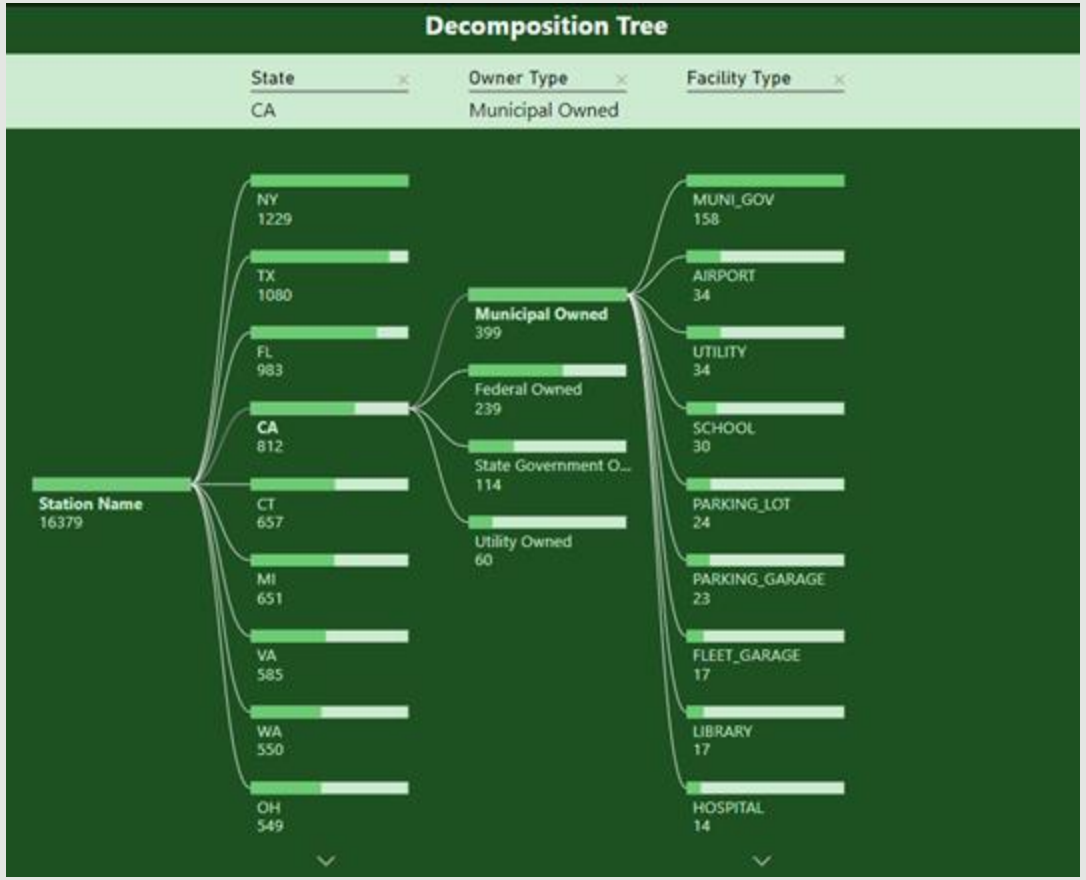
Additional Features - Q & A



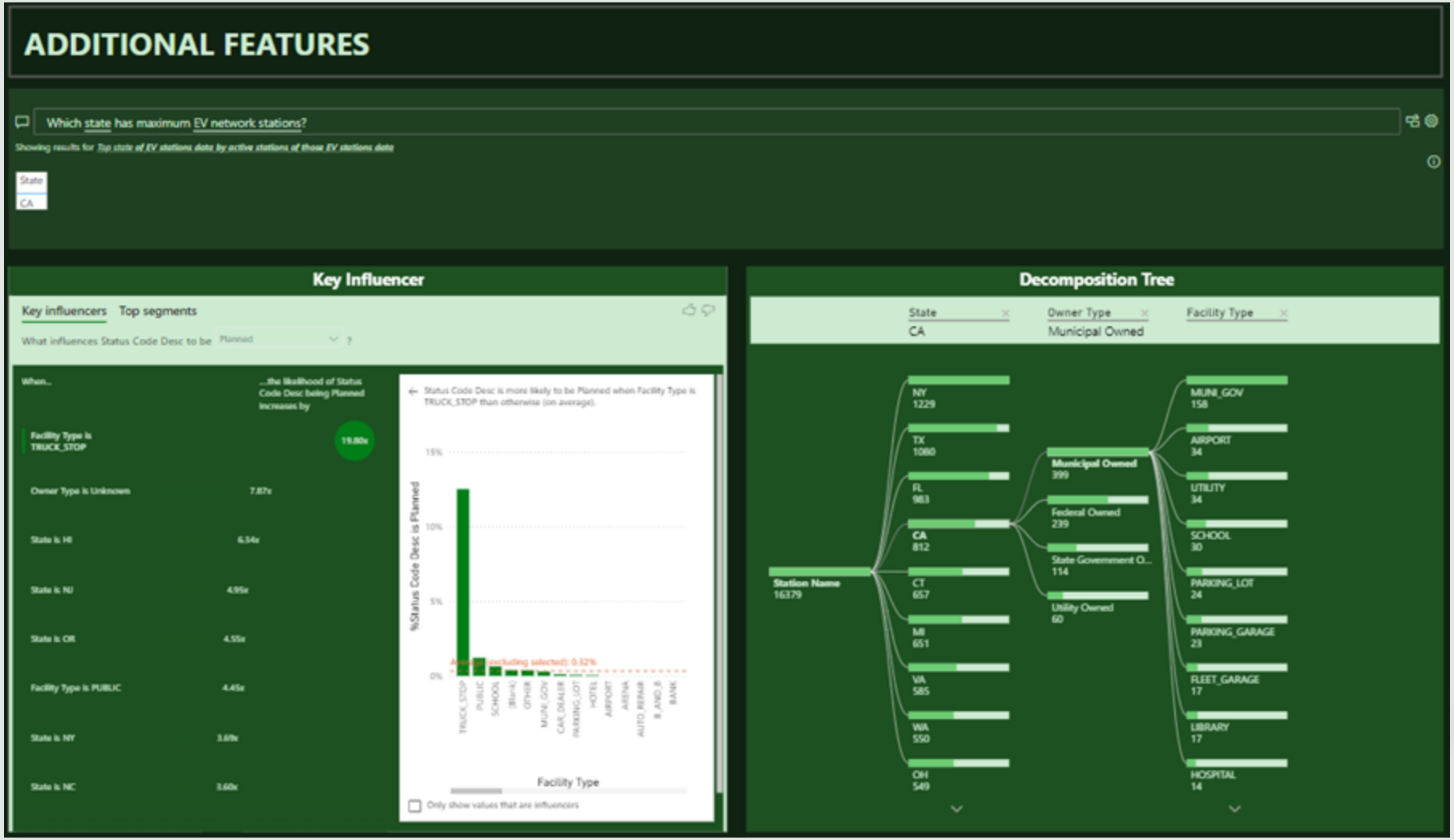
Additional Features - Key Influencers



Additional Features - Decomposition Tree



Dashboard View (Page 2)



Demonstration

https://app.powerbi.com/groups/me/reports/2df8ca23-ccda-44ce-bf88-e7c20369b9b2/56dcbb3684d7de188dac?experience=power-bi

Advanced Features

- DAX Metrics
- Q&A Feature
- Key Influencers
- Decomposition Tree

Conclusion

- California leads the U.S. in EV charging stations, showcasing its commitment to EV infrastructure.
- 95% of stations are publicly accessible, ensuring wide user access.
- EV charging stations have grown steadily since 2010, with a significant surge after 2019.
- Over 85% are privately owned, emphasizing the role of private entities in expansion.
- Only 18% offer DC Fast Charging, and less than 1% use renewable energy, highlighting opportunities for improvement in speed and sustainability.

Future Scope

- Integration with Real-Time Data to monitor station utilization and downtime.
- Use geo-spatial analysis and clustering techniques for recommending optimal locations for new stations.
- Incorporate user reviews for service improvements.

Thank You !