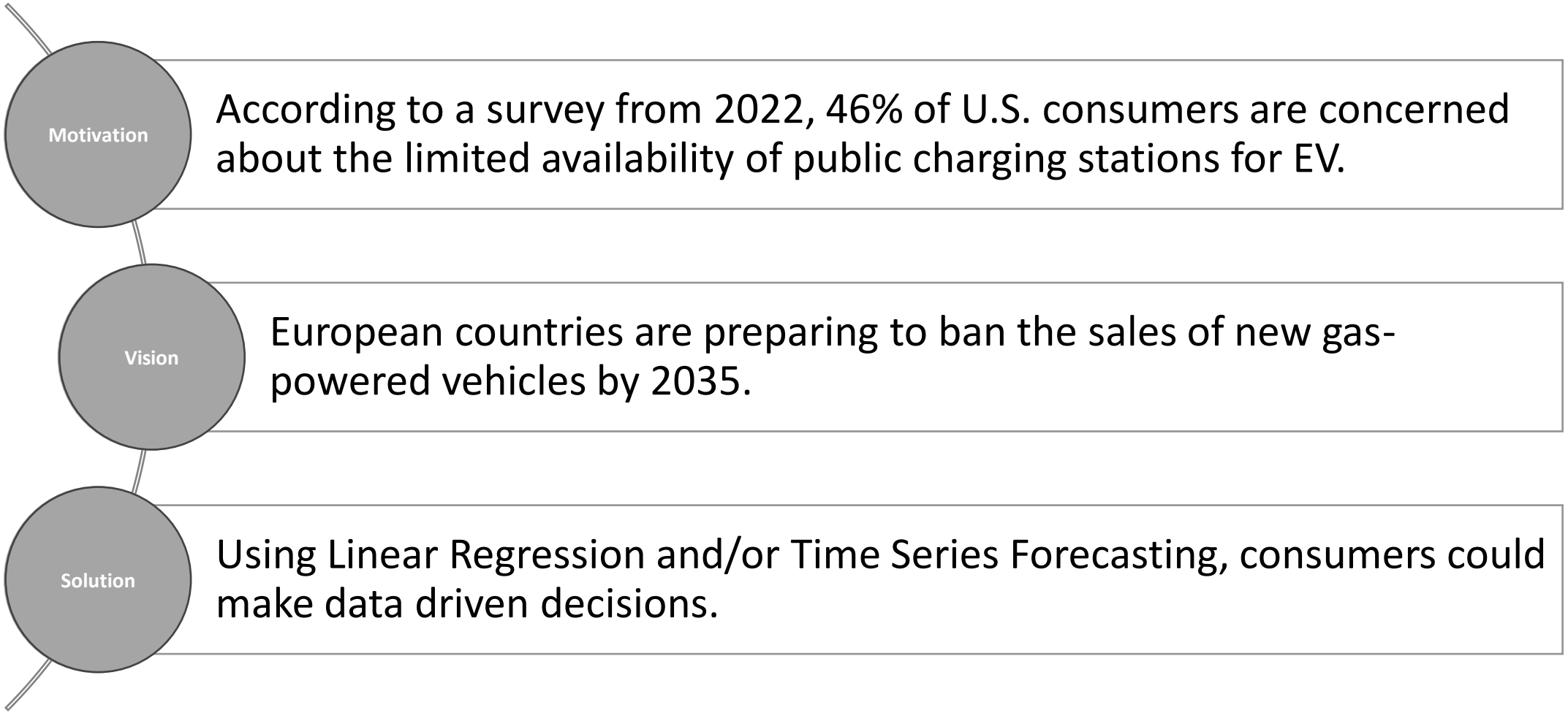
Sprint 3

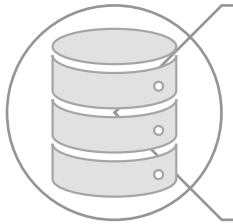
Capstone Project
Alternative Fuel Stations and Electric Vehicles Study

Michael Dorado
BrainStation Data Science Bootcamp
May 5th, 2024

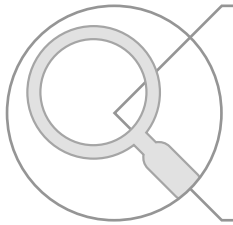
Project Overview



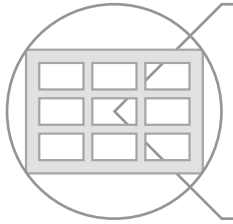
Dataset Overview



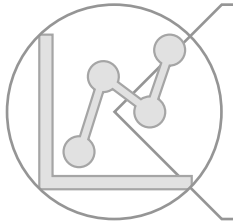
www.data.gov



Published by the National Renewable Energy Laboratory on July 31st, 2021.



Original dataframe shape (56800, 65)



Modeling dataframe shape (316, 1)

```
1  df_original.shape
```

```
(56800, 65)
```

```
2  df_TM.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 47086 entries, 159 to 56793
Data columns (total 2 columns):
 #   Column     Non-Null Count  Dtype
---  ------     --------------  -----
 0   Open Date  45791 non-null  datetime64[ns]
 1   State      47086 non-null  object
dtypes: datetime64[ns](1), object(1)
memory usage: 1.1+ MB
```

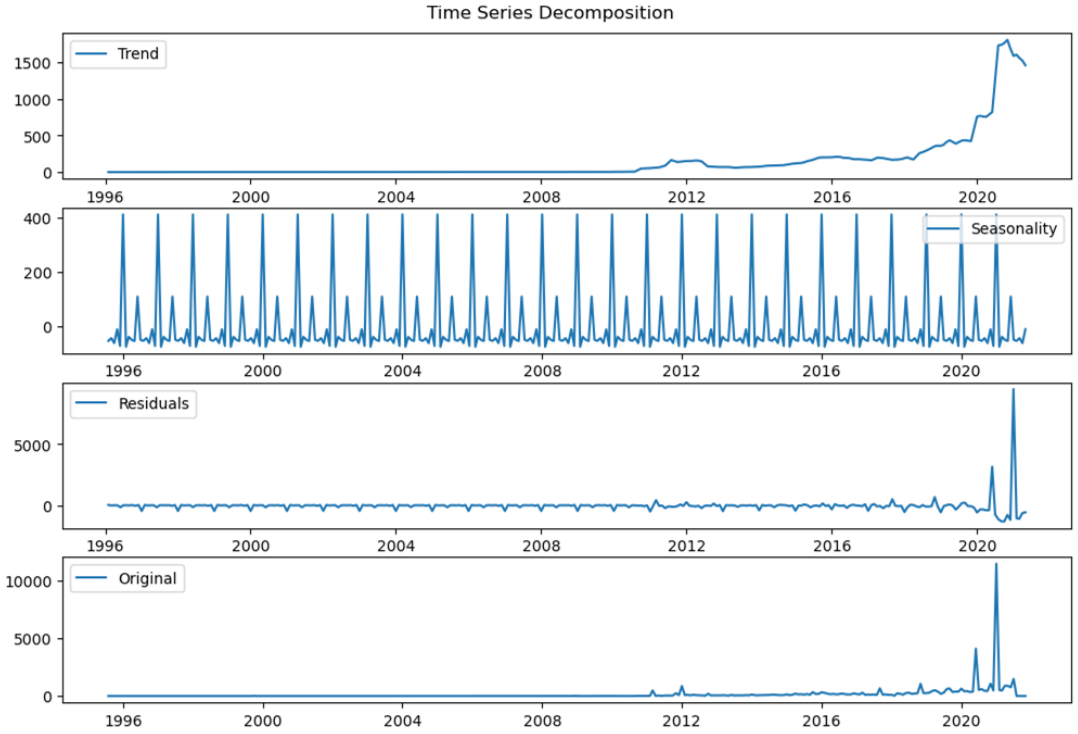
```
2  df_TM_dummies.info()
```

```
<class 'pandas.core.frame.DataFrame'>
DatetimeIndex: 45791 entries, 1999-10-15 to 2021-07-29
Data columns (total 53 columns):
 #   Column              Non-Null Count  Dtype
---  ------              --------------  -----
 0   State_AK            45791 non-null  uint8
 1   State_AL            45791 non-null  uint8
 2   State_AR            45791 non-null  uint8
```

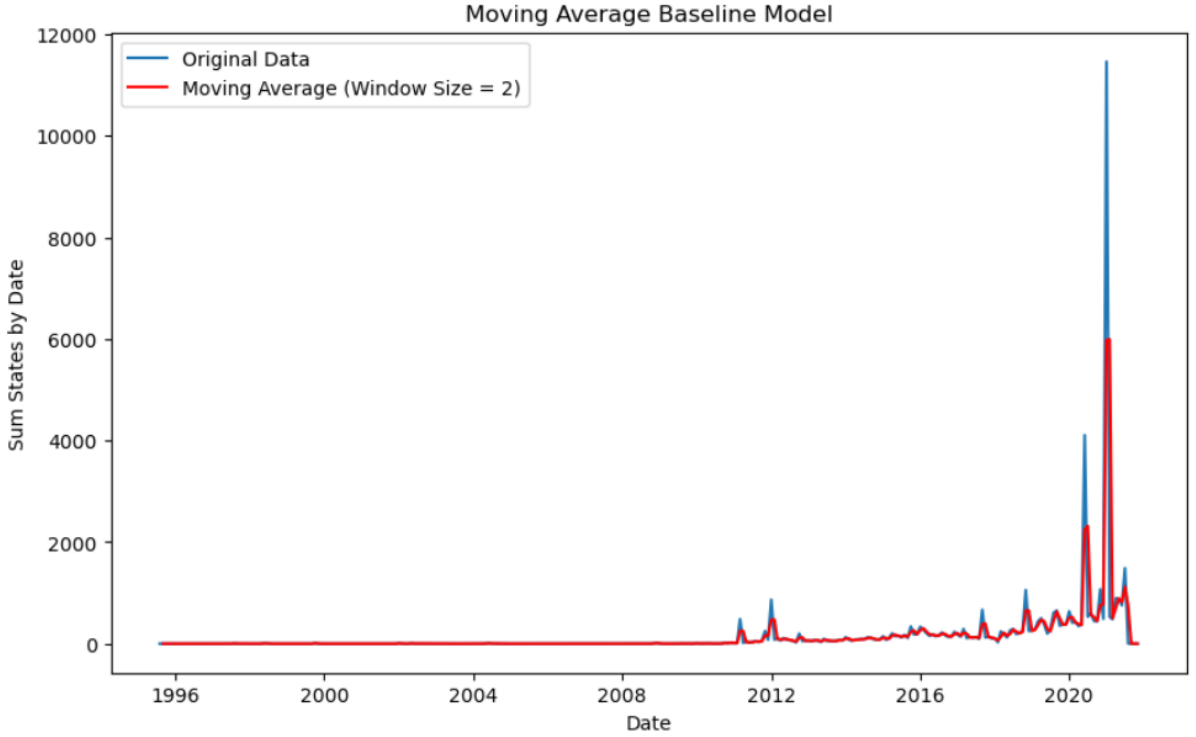
```
2  df_TM_monthly.info()
```

```
<class 'pandas.core.frame.DataFrame'>
DatetimeIndex: 316 entries, 1995-08-01 to 2021-11-01
Freq: MS
Data columns (total 1 columns):
 #   Column             Non-Null Count  Dtype
---  ------             --------------  -----
 0   Sum States by Date  316 non-null    int64
dtypes: int64(1)
memory usage: 4.9 KB
```

Important EDA Findings & Baseline Model



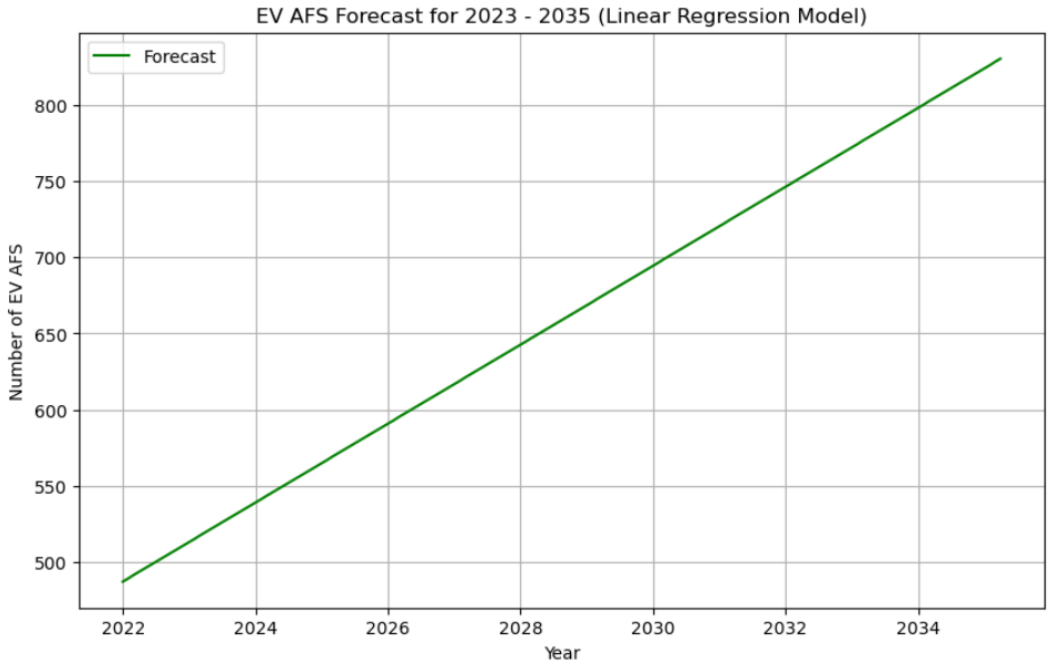
- Trend is positive.
- Seasonality is highest during January and June.
- Residual shows slight seasonality, notable variance at the end.
- Data is non-stationary, its statistical properties, such as mean and variance, change over time.



- Rolling Moving Average with a 2-month window was used for the Baseline.
  - Mean % Error = -132.92%
    - Underestimates the original data
    - 2-month window size provides finer granularity and sensitivity variations in the data.

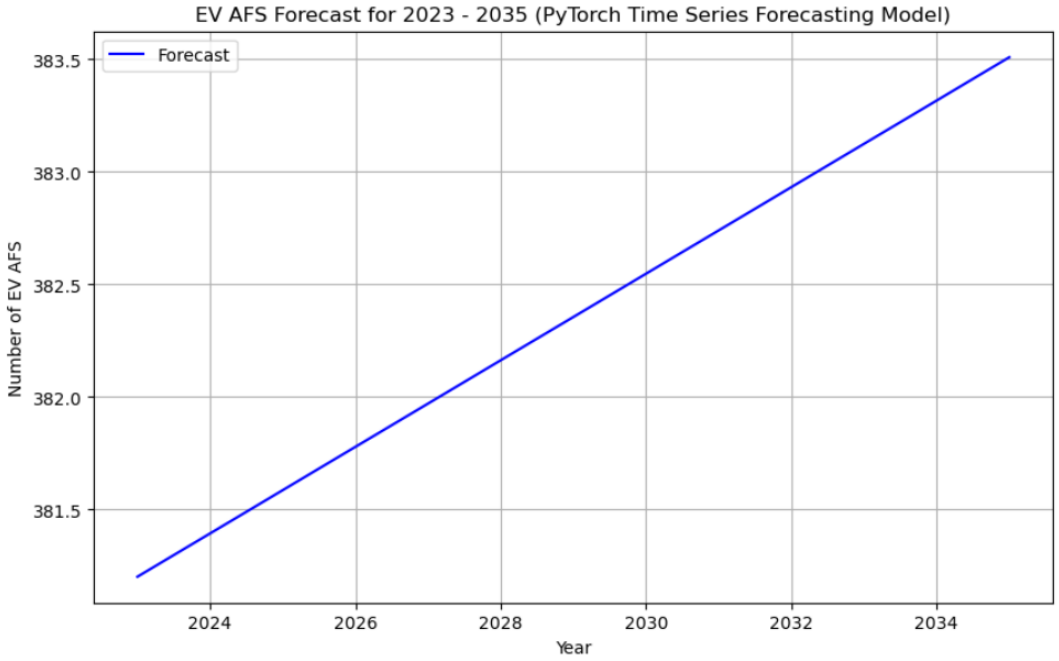
Modeling

Linear Regression Model



- R^2 = 7.8%
  - Correlation between Sum States by Date Variability and Time Index is low.
- From this chart we can visualize the future trend, 2023 through 2035, is a positive trend, with a ***(m = 25).***

PyTorch Time Series Forecasting



- The Mean Squared Error (MSE) on the testing data was approximately 213125136, significant discrepancy.
- From this chart we can visualize the trend of the number of predicted AFS over the years 2023 through 2035. It is a positive with a ***(m = 0.1).***

Additional Findings

	Vehicles by 2022	Vehicles by 2035	% Change	Number of Stations	Vehicles per Station (2022)	Fill-up/Charge Up Time
Gas Powered Vehicles	270 million (approximately)	250 million (rough estimate)	-11.11%	150,000	1,800 vehicles per 1 gas station	5 min (6 gas powered vehicles for every 1 EV)
Electric Vehicles	2.5 million (approximately)	40 million (rough estimate)	+1,500%	57,200	43 vehicles per 1 EV station	30 min (1 EV for every 6 gas powered vehicle)

- Consider:
  - For each EV charged, there are 6 gas powered vehicles fueled
  - EV market in the US expected to grow 1,500% by the year 2035
  - Although the current ratio of EV per EV Station is better than the gas powered vehicle infrastructure (even with the turnaround factored in), to maintain a healthy ratio, the growth of EV stations would have to grow by at least 130% to keep a 300 EV per 1 EV station ratio.

Conclusion

Findings	• Charging stations infrastructure must have a yearly rate of change of ~6.5% to keep a healthy vehicles to charging stations infrastructure ratio.
Model	• Linear regression slope = 25 • Pytorch Time Series Forecast slope = 0.1
Conclusion	• EV Consumers would face charging stations infrastructure challenges.
Considerations	• Keep in mind my dataset has reliability opportunities. • low R^2, high p-value, not a proper MSE.