Deliverable 2 Presentation

Background

- With the increase in electric vehicle use, this project will be focusing on the sales of electric vehicles globally.
- This requires a supervised learning approach and there will be some **linear regression** modeling included to track the data.
- This will help answer questions as to whether the switch to electric vehicles will withhold globally, or if it was a quick burst that still requires further development before a majority of the world population begins adopting electric vehicles as their daily commuter cars.

Data Source

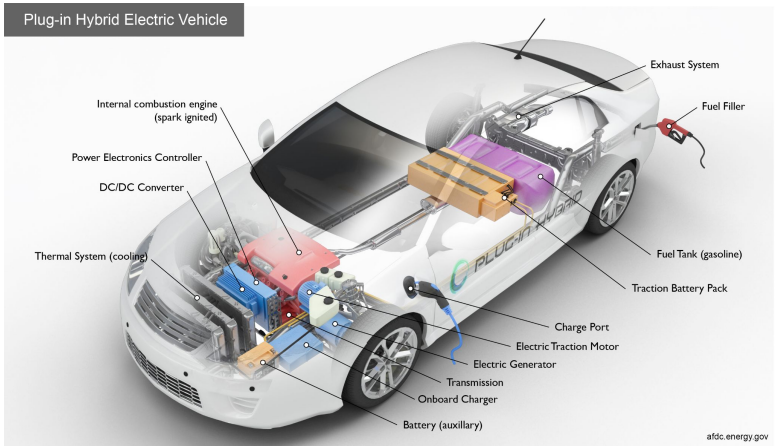
Edson Marin. (2023). Historic sales of electric vehicles [Data set]. Kaggle. https://doi.org/10.34740/KAGGLE/DSV/5672792

- The dataset includes 8 columns including region, category (historical), parameter (electric vehicle sales), mode (cars), powertrain (battery electric vehicle (BEV) or plug-in hybrid electric vehicle (PHEV)), year, unit (vehicles), and value (amount of vehicle sales)
- It is a single table provided by a data scientist who gathered the data from the Global EV Outlook, an annual publication

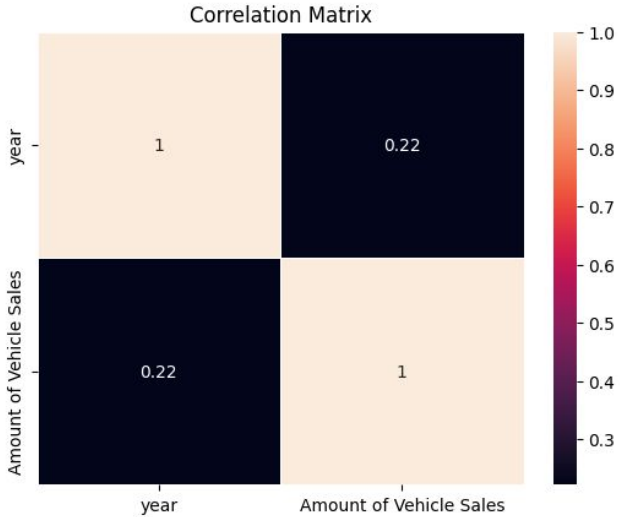
region	category	parameter	mode	powertrain	year
Belgium 3% China 3% Other (782) 94%	1 unique value	1 unique value	1 unique value	BEV 53% PHEV 47%	2010
Australia	Historical	EV sales	Cars	BEV	2011
Australia	Historical	EV sales	Cars	BEV	2012
Australia	Historical	EV sales	Cars	PHEV	2012
Australia	Historical	EV sales	Cars	PHEV	2013

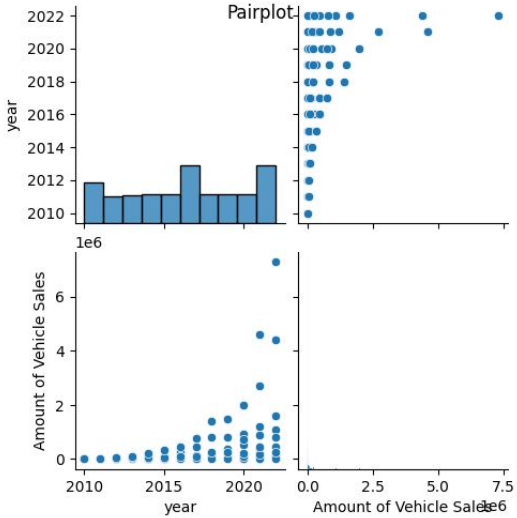
Data Cleanup

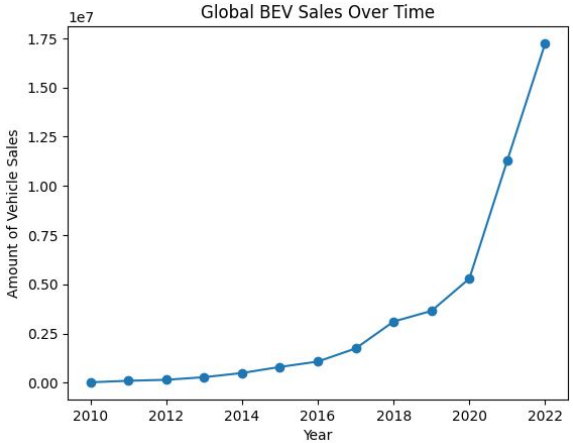
- Data was cleaned to exclude PHEVs from the data, since they still use gasoline and are at the intersection of being both a gasoline and electric vehicle



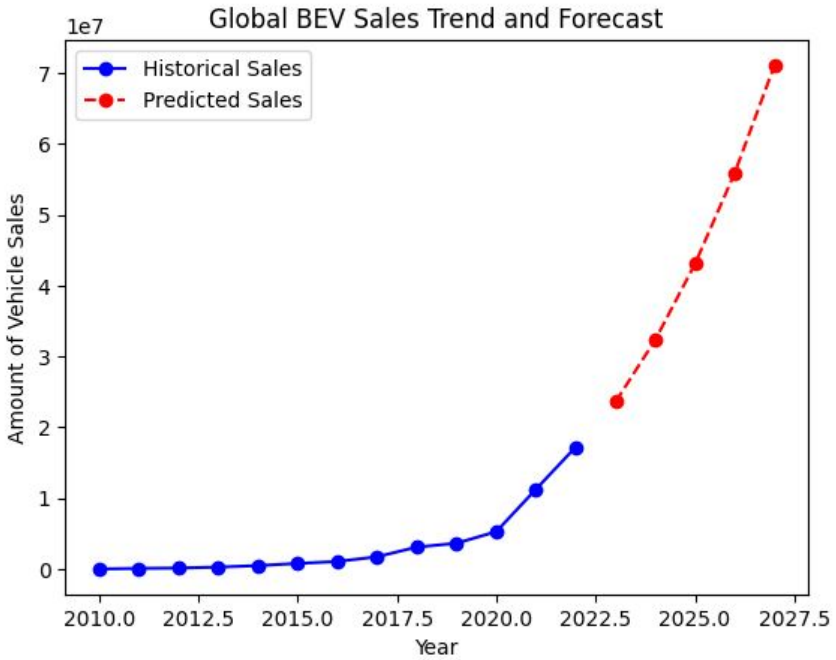
Exploratory Data Analysis







Linear Regression Modeling



- **Mean Absolute Error**: 567197.4254807692
- **Mean Squared Error**: 482526090743.5142
- **Root Mean Squared Error**: 694640.9797467424
- **R2 Score**: 0.9806169173306662

Addressing Potential Over/Underfitting

Split data into “train” and “test” and changed the year to a more recent one (2020)

- **Train MAE**: 66064.39808154042
- **Train RMSE**: 176897.8715094852
- **Train R2 Score**: 0.06423091079501086
- **Test MAE:** 456650.1284851498
- **Test RMSE**: 1177860.2047293456
- **Test R2 Score**: -0.026757109574435667

Analysis

- The linear regression model was able to predict how BEV sales will continue to increase in the coming years, up to the year 2027
- Different evaluation metrics also showed that with the high R2 value, the model is supported by the data
- Summary Table:

Metrics	Train	Test
MAE	66064.398082	4.566501e+05
RMSE	176897.871509	1.177860e+06
R2 Score	0.064231	-2.675711e-02

Conclusion

- The model is supported by data that BEV sales will increase globally in the coming years
- This data was shown to be underfitted in the last code section, meaning that the model may in fact be underwhelming compared to how global BEV sales will actually increase
- EDA can help greatly in determining which model to use, and that showing the evaluation metrics along the way may paint a better picture of what is happening in the data
- Next steps → Seeing how data compares now that some states in the US are removing the EV tax rebates