Hamoye Internship: Electric Vehicles Manufacturing

Team Prophet

21.09.2021

Team Prophet

1. Nimatallahi Masuud - Team Lead
2. Muhammad Salis - Model Deployment
3. Amit Purswani - Feature Engineering
4. Damola Agboola - EDA & Slide Curation
5. Kunle Adewole - Data Scrapping
6. Dauda Umar Faruq - Machine Learning
7. Abe Enoch - Machine Learning
8. Olamide Amoo - Model Deployment
9. Muhammed Jimoh - Project Manager
10. Alex Izuka - EDA

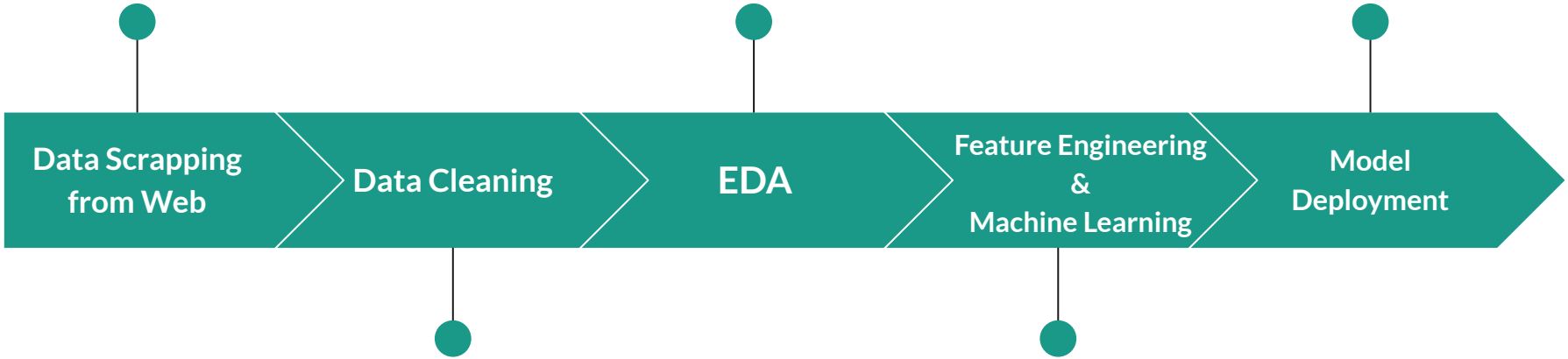
Problem Statement

- As EV technology is evolving continuously and gradually replacing the conventional automobiles all over the world.

- Their demand is increasing with each passing day, each EV brand is coming up with new improved utilities and enhanced performance.

- Develop an ML Model to predict the prices of EV based on their features.

Process Flow



Data Collection

Website: https://ev-database.org

Technology: Selenium for the web scrapping

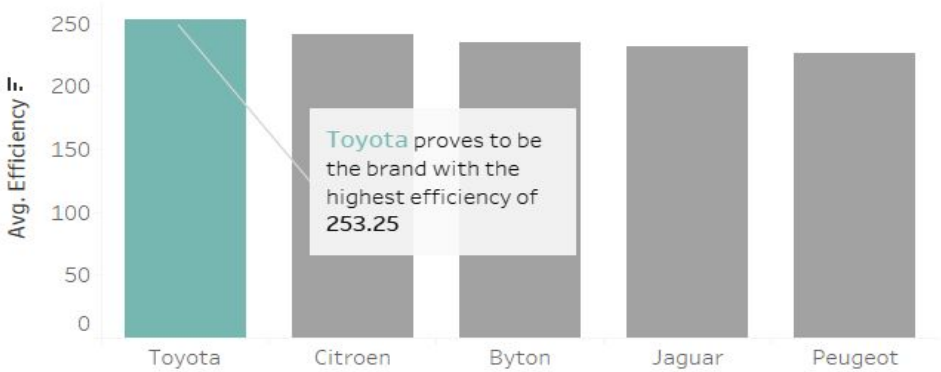
Result: 181 unique vehicles with car specifications

Features: acceleration, fast charge, efficiency, car speed, distance, seats, battery capacity, car model and car brand

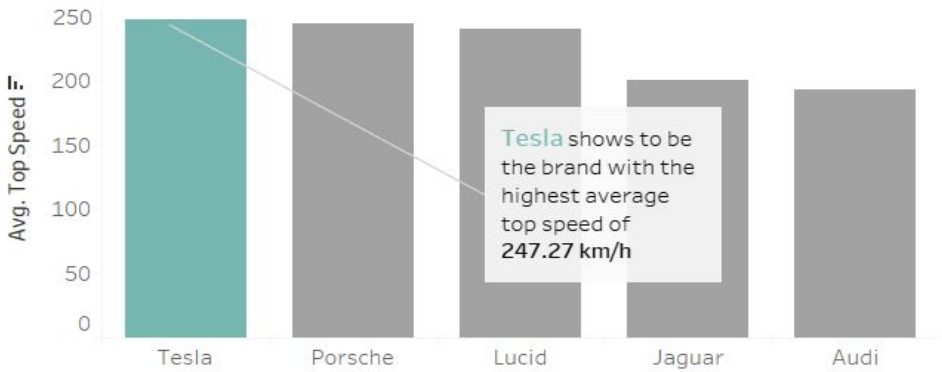
Technology: Tableau

EDA Dashboard on EV features

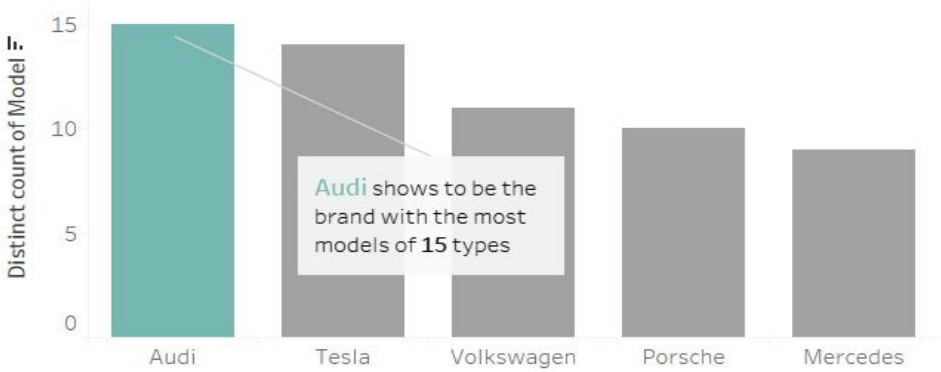
Top 5 EV brands with highest average efficiency



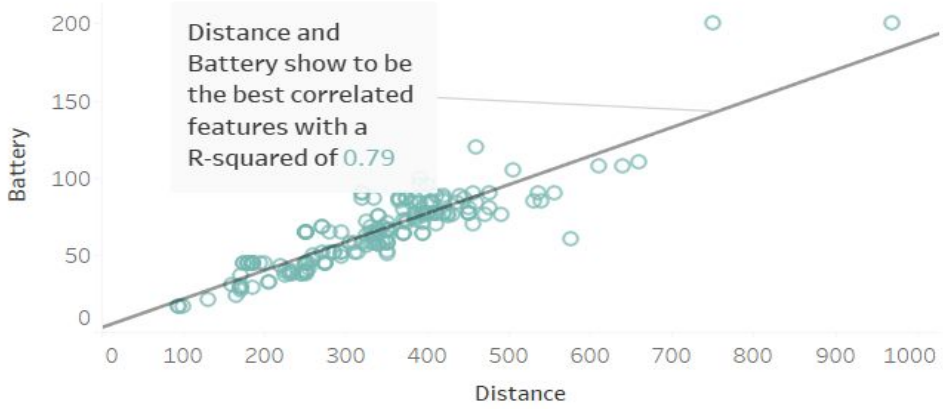
Top 5 EV brands with highest average top speed



Top 5 EV brands with the most models



Distance and Battery Correlation



Feature Engineering

Brand name: Brand value plays an important role in pricing.

Battery: Battery capacity is one of the crucial factor for comparing EV cars.

Distance: Range covered by the EV in single battery charge.

Fast charge: This determines the range of an EV with one hour of charging.

Efficiency: Here it specifically is the battery performance of the EV.

Acceleration, Top speed: These are regular features applicable to automobiles, EV or otherwise.

Model Training and Evaluation Experiment

S/N	Scenario	Model and R2 Score
1	Fast Charge Null values replaced with Median value German price Outliers removed German price Null values replaced with Median value Brand Names manually replaced with Ordinal numbers No. of Seats considered in model	Linear Regression 0.69 Random Forest 0.74 Xgboost 0.68
2	**Fast Charge Null values dropped** **German price Outliers retained (considering luxury cars)** **German price Null values dropped** Brand Names encoded with Ordinal Encoder (scikit lib) No. of Seats considered in model	Linear Regression 0.53 Random Forest 0.83 Xgboost 0.73
3	Fast Charge Null values dropped German price Outliers retained (considering luxury cars) German price Null values dropped **Brand Names dropped** No. of Seats considered in model	Linear Regression 0.54 Random Forest 0.86 Xgboost 0.77
4	**Fast Charge Null values replaced with zero** German price Outliers retained (considering luxury cars) **German price Null values calculated from other country prices** Brand Names dropped No. of Seats considered in model	Linear Regression 0.48 Random Forest 0.73 Xgboost 0.74

Model Training and Evaluation Experiment

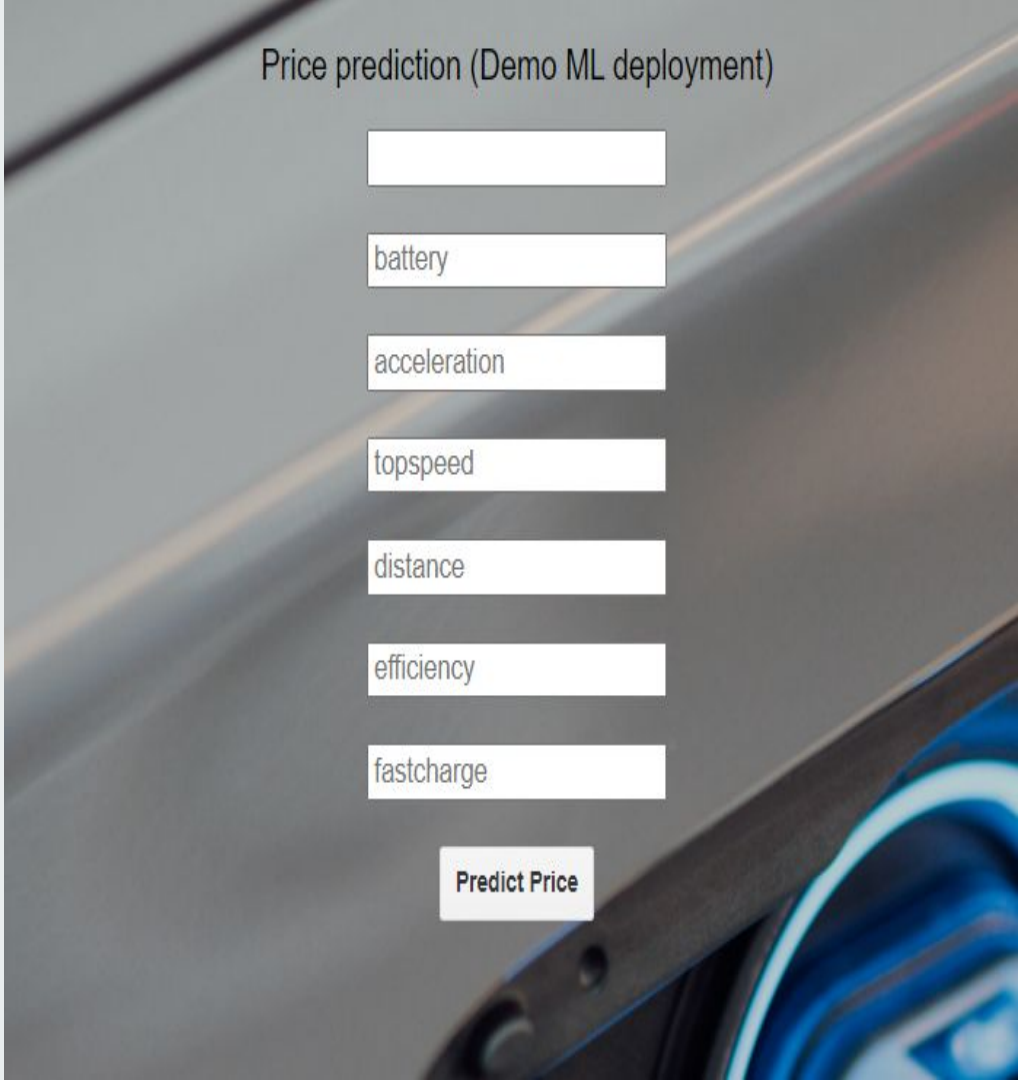
5	Fast Charge Null values replaced with zero German price Outliers retained (considering luxury cars) German price Null values calculated from other country prices **Brand Names encoded with Dummy variables method** **No. of Seats not considered in model**	Linear Regression 0.82 Random Forest 0.72 Xgboost 0.71 (**Final Iteration**)

Model Deployment

URL: https://ev-price-app.herokuapp.com

Technology: Heroku

Result: Predicts prices of EV vehicles



Challenges Faced

- We wanted a fairly large data-set to train the model but we got a smaller data-set and, in the interest of time we went ahead with the smaller data-set.
- Time Constraints for better research

Conclusion

- We have developed a basic ML model which can fairly predict the EV prices given the brand name and specific values of various features.

- As a continued learning exercise, we would like to train our model with a larger data-set and with more no. of features.

Thank you