



# SILVER OAK UNIVERSITY

## EDUCATION TO INNOVATION

Date : .....

Page No. ....

### Assignment For Excel

#### Data Analysis

Title : Electric Vehicle population Data  
Analysis & predictive Modeling

Name : Kishan Mudaliar

Date : 24/9/29

Course : DATA Analysis.

#### INTRODUCTION:-

This Report presents a Comprehensive analysis for Electric Vehicle population data dataset, which specifies Battery Electric Vehicles & plug-in hybrid Electric Vehicles registered in Washington state. The primary objective is to perform a structured data analysis starting with data cleaning, moving through exploratory data analysis to Uncover trends & distribution. Concluding with development & evaluation of a linear regression model to predict a vehicle's electric range.



Section / **SILVER OAK UNIVERSITY**  
EDUCATION TO INNOVATION

Date : .....  
Page No. ....

Q) Data Cleaning Questions

- 1) Based on categories there are total 3 missing values. 1) Vehicle location - 1 Value
  - 2) Electric Utility - 1 Value
  - 3 ) 2020 Census 1 is null , 1 Value.
- 2.) Boys MRP: A Value of 0 in Coloumn indicates missing Value. These Should Be handled by either filtering Out records or by Using an imputation Strategy.
- 3.) Stocker Reserve: A Value of 0 in Coloumn Negate the claim in missing Or Dauge has been not recorded. as Some Vehicles in dataset have an Unknwown Status for truck And Singularity . For Analysis , These needs to be like further Out or you can use an imputation Method.
- 3.) No duplicate Records in the dataset. If duplicates were present they would be managed by selecting & removing them to ensure that each row represent a unique vehicle & does not skew the analysis.

4) VTNs can be anonymized by applying a crypto graphic mask from like SHA 256 to VTN Algo identifiable mask for each VTN which can be used as Unique identifier for analysis while maintaining privacy.

5) The Vehicle Location Column contains a single string information point (Latitude, Longitude). This can be cleaned by pruning the string to extract Numerical (Longitude & Latitude) Values which can be stored in two separate Columns (Latitudes & Longitudes).

## Section 2 Data Exploration:

- Top 5 Models: TESLA, NISSAN, BMW, KIA & CHEVROLET.
- Top 5 Models: LEAF, MODEL 3, MODELS, PREMIUM & XS.

2) EV Distribution is heavily Concentrated in a few Countries.

- Kitap County has most registrations in the dataset.

**SILVER OAK UNIVERSITY**  
**EDUCATION TO INNOVATION**

- Q) Based on data we have been a general growth in trend, with a significant increase in registration from 2018 to 2020. Then a dip in recent yrs (2021-2022). & then a sharp increase in 2023-25.
- 4) The Avg. Electric range of EVs in the market is approx 114-26 miles.
- 5) Top 5 vehicles (Catering in concern) are available for Clean Attentat
- 6) There is significant variation in electric range across different makes & models - Tesla Model Y (291 mi), CHEVROLET Bolt EU (239 mi), Tesla Model 3 (223 mi), Tesla Model X (229 mi), Tesla Model S (222 mi).
- 7) The Avg. Box weight follows:
- TESLA MODEL 1 S : \$ 69,000
  - BMW 330i : \$ 45,600
- 8) The data shows a strong trend in EU adoption in more densely populated urban & suburban areas. The top countries for EV are China, India & United States, which are home to major cities like Seattle, Benton, Olympia,

- 4. Fictional Regression Model**
- 1) By establishing a relationship b/w it the dependent Variable & Independent Variables.  
 The Model Learns the weight for each feature, allowing it to predict the electric range of a New Vehicle based on its Specifications.
- 2) Features that can be used to predict Electric Range Model year, Base MSRP, Make & Model.  
 It is important to note that the Base MSRP & Electric Range Columns will need to be cleaned & imputed to be useful in the Model.
- 3) Categorical Variables like Make Model Cannot be used directly in a linear regression Model because they are not numerical. They must be converted into a Numerical format using a technique called One hot encoding.
- 4) The R<sup>2</sup> Score or coefficient of determination is a statistical measure that represents the proportion of variance for dependent Variables that is explained by independent Variables in the Model. A Score of 1.0 Indicate the Model perfectly Predicts the dependent Variables while a Score of 0.0 suggest that the Model Capable None of Variability.



- Q) The influence of Base MSRP on Electric Range is determined by it's coefficient in linear regression Model. A positive coefficient would indicate that as the Base MSRP increases, the Electric Range is also expected to increase.
- b) Feature Engineering :- Create New feature from existing ones. For Ex :- Combine Make & Model into a single categorical feature.
- 2) Handling Missing data! Use more advanced imputation techniques to fill in the missing Electric Range & Base MSRP values.
- 3) Handling Outliers! Identify & Manage Outliers in the data as they can disproportionately affect the Model's performance.
- Q1. Yes a trained linear regression Model can be used to predict the range of New EV Models you would input the New Model's independent Variable into the trained Model & it would output a predicted. However the Model's accuracy would depend on how well the New Model's feature align with the data the Model was trained on.

## Conclusion:

EV adoption is rapidly increasing with Tesla leading Market in both Volume & Range. Gov incentives play a crucial role, and urban regions are the strongest adopters while Vehicle Cost does not directly predict range, brand & technology play a major role in performance.