ELECTERICAL VEHICLES ANALYSIS

Cloud based Data Engineering Project

DOCUMENTATION

Prepared by: Yousef Jaber Abdulaziz Abo Enaba
Date: 9/2/2025

Table of Contents

Contents

1. Data Overview

- This synthetic dataset contains **3022 records** of **electric vehicles** and **plug-in models** across many **brands** and **years**, capturing technical specifications, **performance**, **pricing**, **manufacturing** origin, **sales**, and safety-related attributes. Each row represents a unique vehicle listing identified by Vehicle_ID and described as follows:

Features	Description	Notes
Vehicle_ID	Unique identifier for each vehicle record.	Ignored
Manufacturer	Vehicle brand or OEM.	53 Manufacturer
Model	Specific model name/variant.	181 Model
Year	Model year associated with the record.	2015 To 2025
Battery_Type	Battery chemistry/technology used.	15 Types
Battery_Capacity_kWh	Nominal battery capacity in kilowatt-hours.	
Range_km	Claimed driving range on a full charge (kilometers).	
Charging_Type	Dominant charging interface or feature.	16 Types
Charge_Time_hr	Approximate time to charge (hours), context varies by charging method.	
Price_USD	Indicative vehicle price in US dollars.	
Color	Primary exterior color or finish.	55 Color
Country_of_Manufacture	Country where the vehicle is manufactured/assembled.	40 Country
Autonomous_Level	Automation capability level (e.g., 0–5), may include decimals for sub-levels.	0 To 5 Levels
CO2_Emissions_g_per_km	Tailpipe CO2 emissions in g/km (often 0 for BEVs); may be blank for missing data.	Ignore as all 0
Safety_Rating	Safety rating (scale not explicitly defined; often 1–5).	3 To 5 Levels
Units_Sold_2024	Units sold in 2024 for this model/variant (as recorded).	
Warranty_Years	Warranty duration in years.	3 To 5 Levels

2. Logical & Physical Model

1) Schema Type: Star Schema

The model follows a **Star schema** structure with a central **fact table** and **four** surrounding **dimension tables** that provide context.

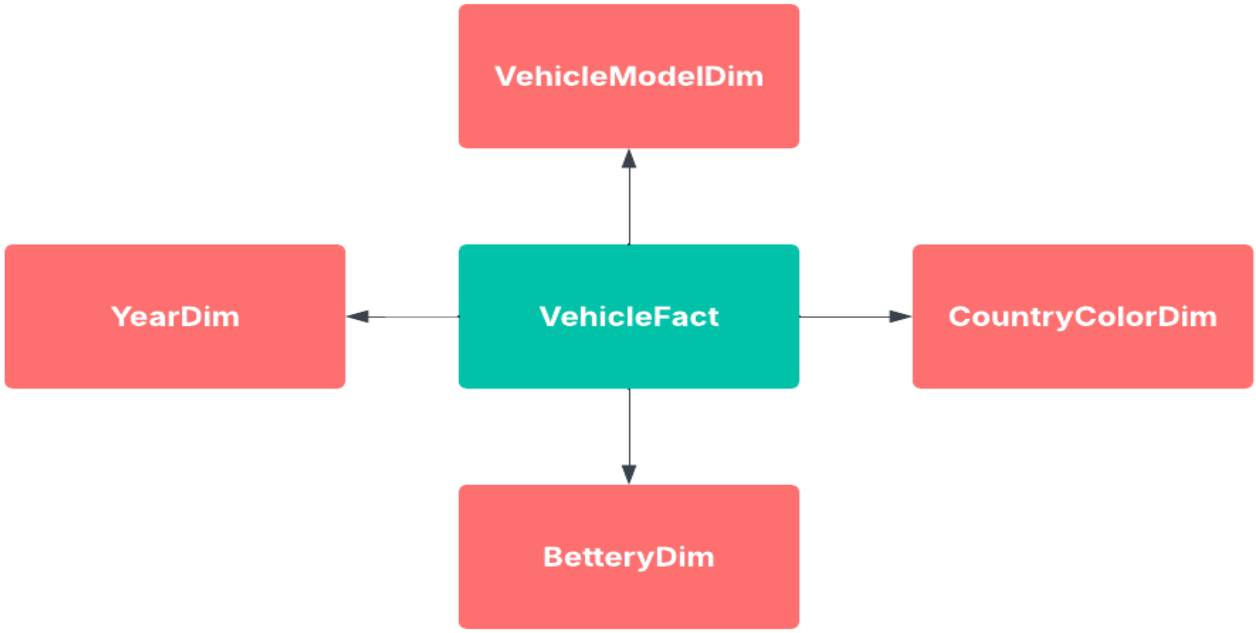
2) Fact Table:

- **VehicleFact:** Captures the **core measurable characteristics and performance indicators** of each electric vehicle in the dataset. It records both **technical specifications** and **market-related values**, which serve as the basis for analytical queries.
  - **VehicleID**
  - **BatteryCapacityKwh**
  - **RangeKM**
  - **ChargeTimeHR**
  - **PriceUSD**
  - **AutonomousLevel**
  - **SafetyRating**
  - **UnitsSold2024**
  - **WarrantyYears**
  - **ModelYearID** → Links to **YearDim** for temporal context.
  - **VehicleModelID** → Links to **VehicleModelDim** for manufacturer and model details.
  - **BatteryID** → Links to **BatteryDim** for battery and charging attributes.
  - **ManCountryID** → Links to **CountryColorDim** for country of manufacture.

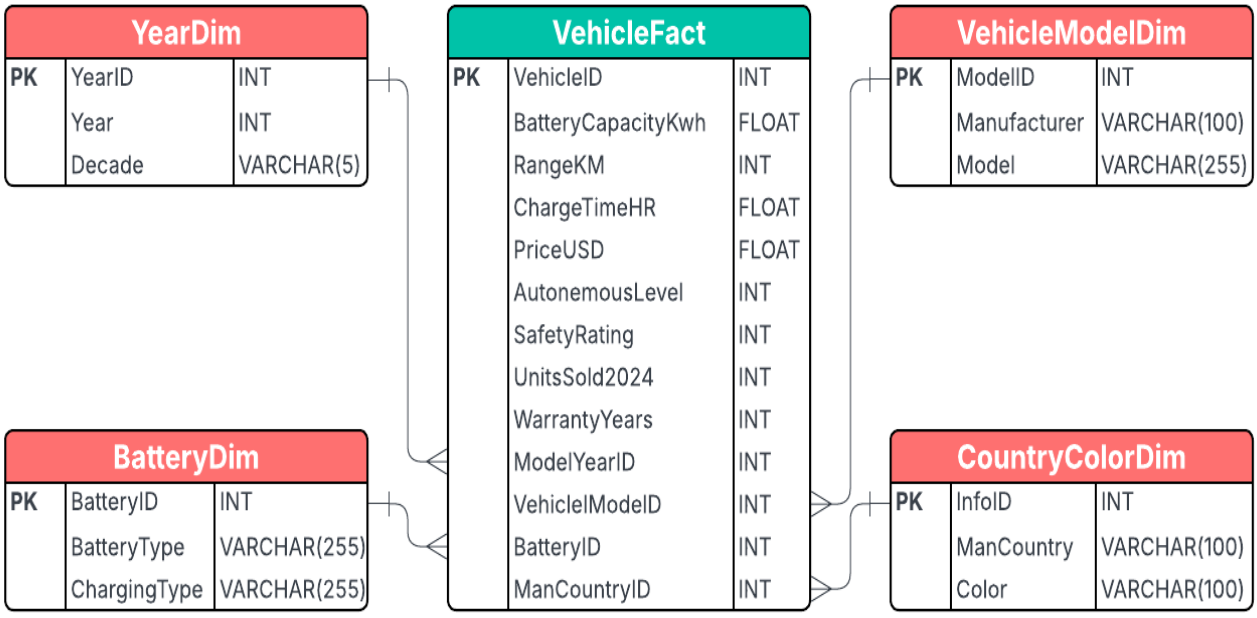
3) Dimensions Tables

- Describes the attributes of **electric vehicles**. Each dimension provides context to the central **VehicleFact** table:
  a) **YearDim:**
     - Captures the temporal aspect of vehicles (Model Year, Model Manufacturing Decade).
     - Allows analysis of EV trends across different model years and decades.
  b) **BatteryDim:**
     - Describes the battery technology and charging type.
     - Useful for comparing vehicle performance and technological advancements in batteries.
  c) **VehicleModelDim:** Stores manufacturer & model details.
     - Useful Enables comparison of vehicles across OEMs and models
  d) **CountryColorDim**
     - Contains manufacturing country and available color information.
     - Supports market analysis by region and aesthetic preferences.

2.1 Conceptual Model



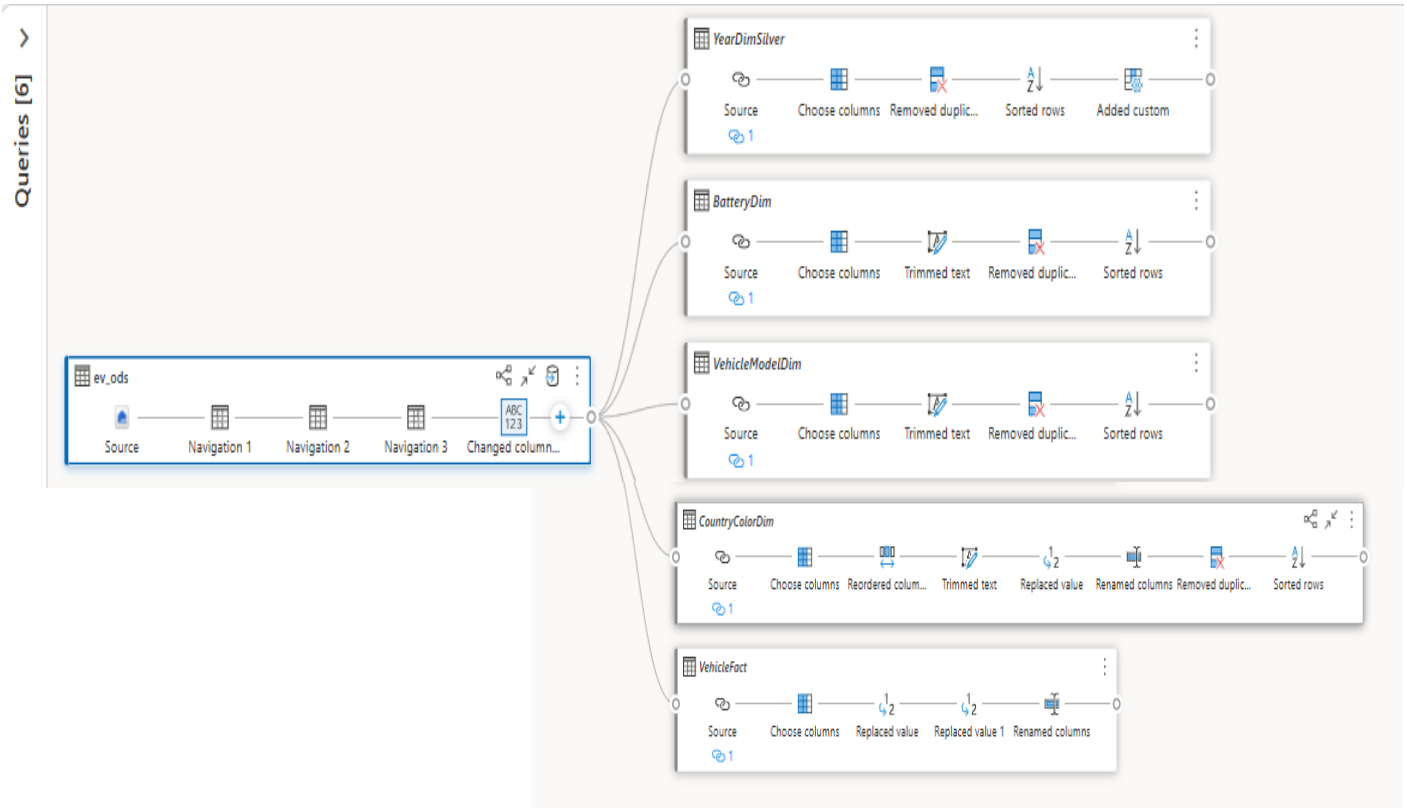
2.2 Logical Model



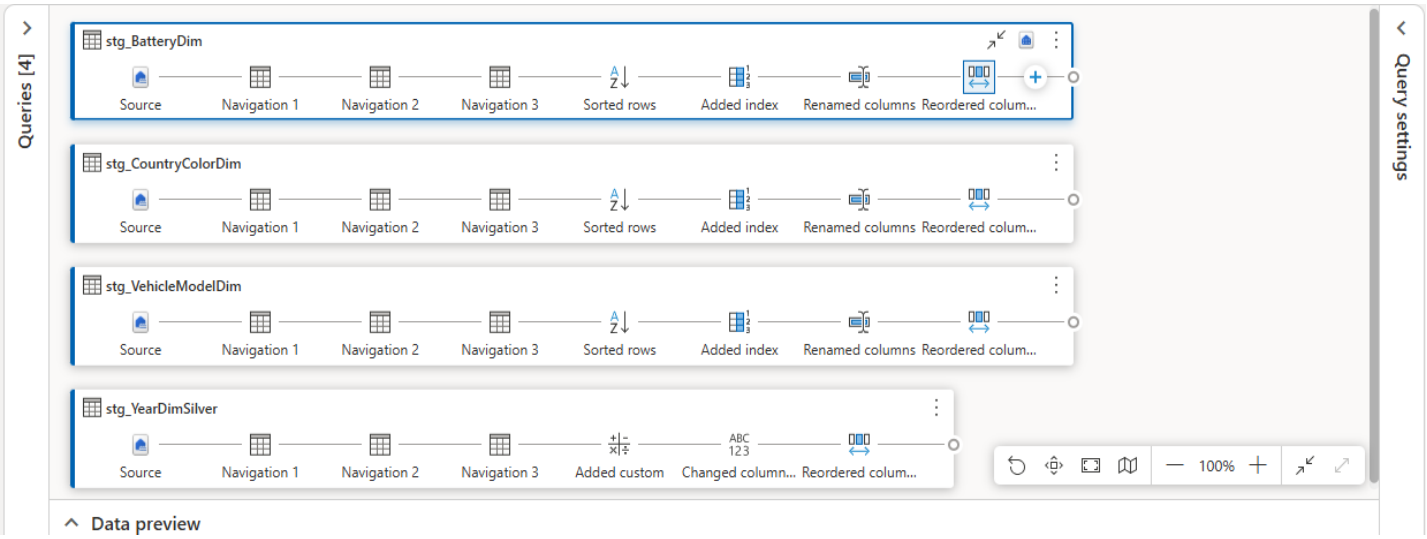
3. Fabric Pipeline

• In this Project I have followed some steps that let's me create the Data models mentioned before, these steps are:

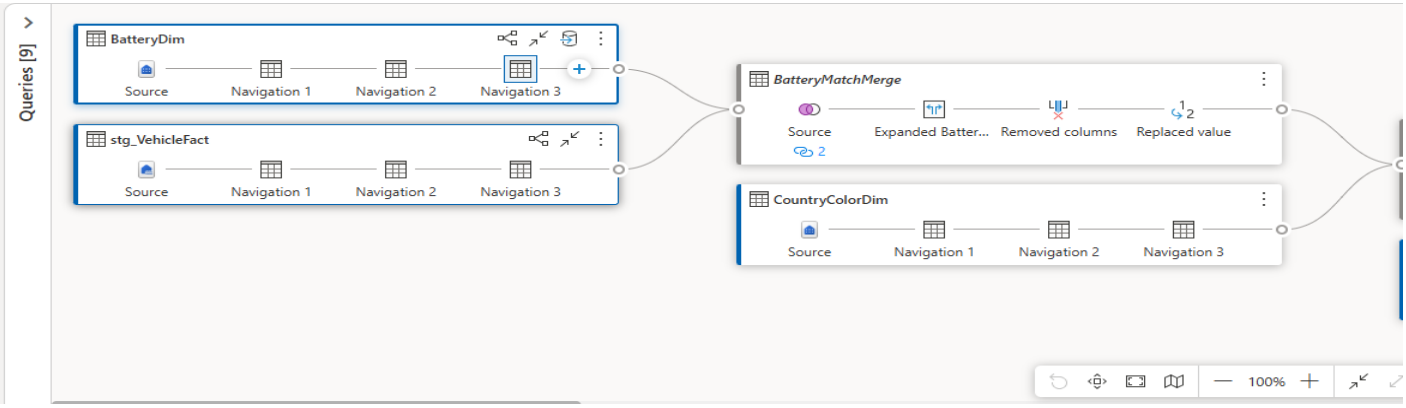
1) **Created a Data Lakehouse:**
   - Set up a **Lakehouse** to serve as the main storage unit for the project.
2) **Uploaded Raw Data**
   - Uploaded the **CSV file** containing the EV dataset into the **Files directory** of the Lakehouse, preparing it for use in the data pipeline.
3) **Converted CSV into a Table**
   - Transformed the uploaded CSV file into a **structured table** within the Lakehouse for easier querying and downstream processing.
4) **Built a Dataflow Gen2 (EV Silver Dataflow)**
   - Created a **Dataflow Gen2** named **"EV Silver Dataflow."**
   - This dataflow includes the **dataset source** and **referenced queries** used to design the project's **fact and dimension tables** (as defined in the star schema).

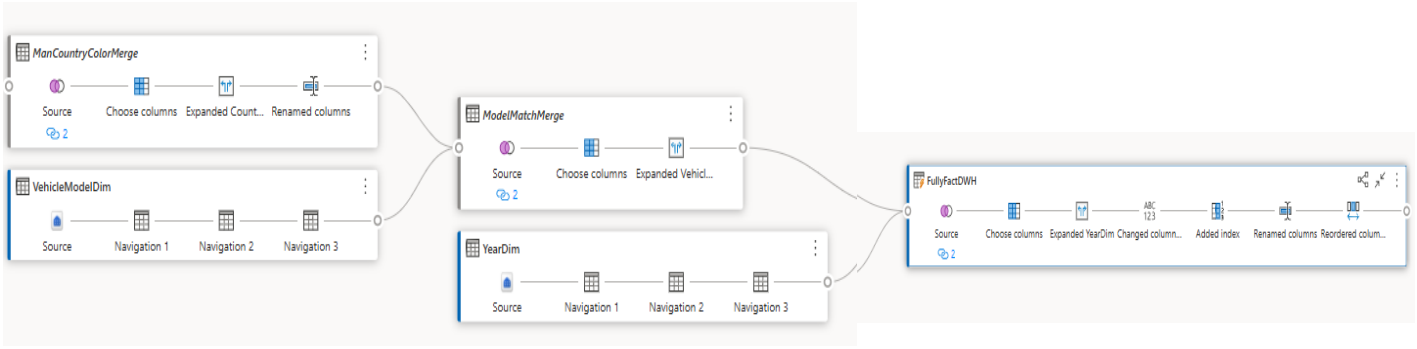


5) Created Dataflow Gen2 for **Gold** Layer (**EV Dimensions Gold**)
   - Built a new Dataflow Gen2 named “EV Dimensions Gold.”
   - In this step, the tables from the staging layer (Silver) in the Lakehouse were transferred into predefined tables within a pre-created Warehouse.
   - Assigned and set the IDs for each dimension to ensure proper relationships and consistency across the star schema.



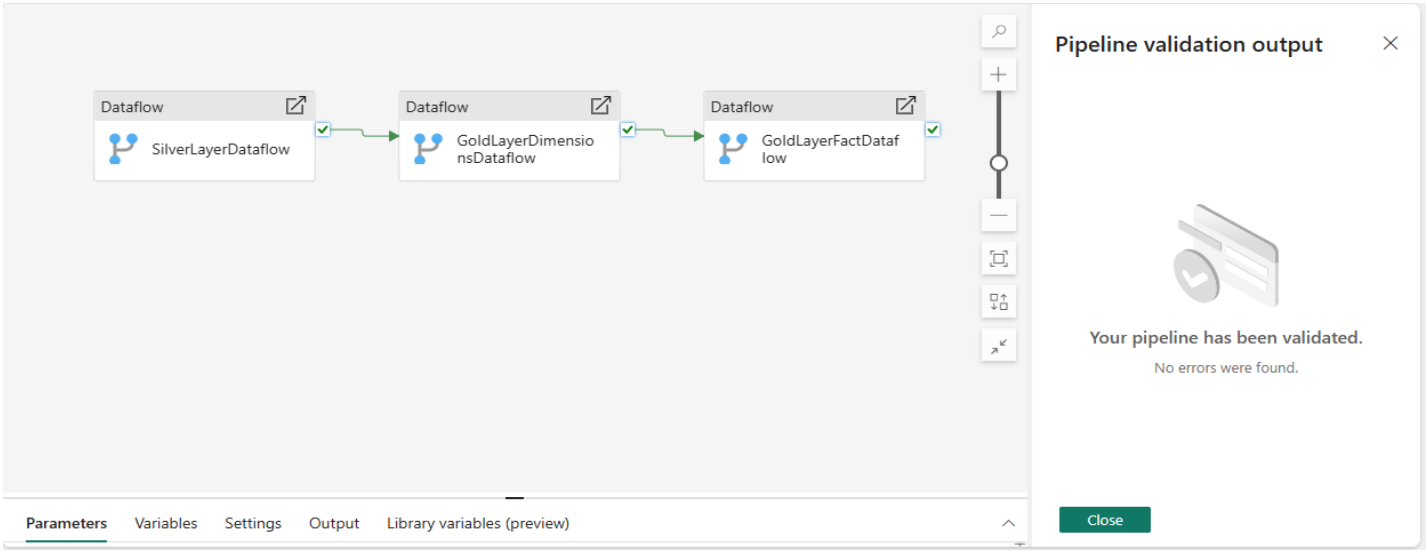
6) **Created Dataflow Gen2 for Fact Table (EV Gold Fact)**
   - Developed a **Dataflow Gen2** named **“EV Gold Fact.”**
   - Selected the **numerical features** from the dataset to serve as **fact measures** (e.g., battery capacity, range, price, sales, etc.).
   - Performed the necessary **joins** between the staged data (from the Silver layer) and the **dimension tables in the Gold layer** to ensure referential integrity.
   - Defined the **destination** as a **pre-created Fact table** in the Warehouse to store the fact data.
   - Configured an **index** for the samples within the Fact table to optimize performance and maintain unique identification of records.





7) Created a Data Pipeline

- Designed a Data Pipeline in Microsoft Fabric to orchestrate and connect all the created Dataflows (Silver, Gold Dimensions, and Gold Fact).
- This setup allows all transformations and data movements to be executed in sequence or triggered at once, ensuring a smooth end-to-end workflow from raw data ingestion to the finalized Fact and Dimension tables.

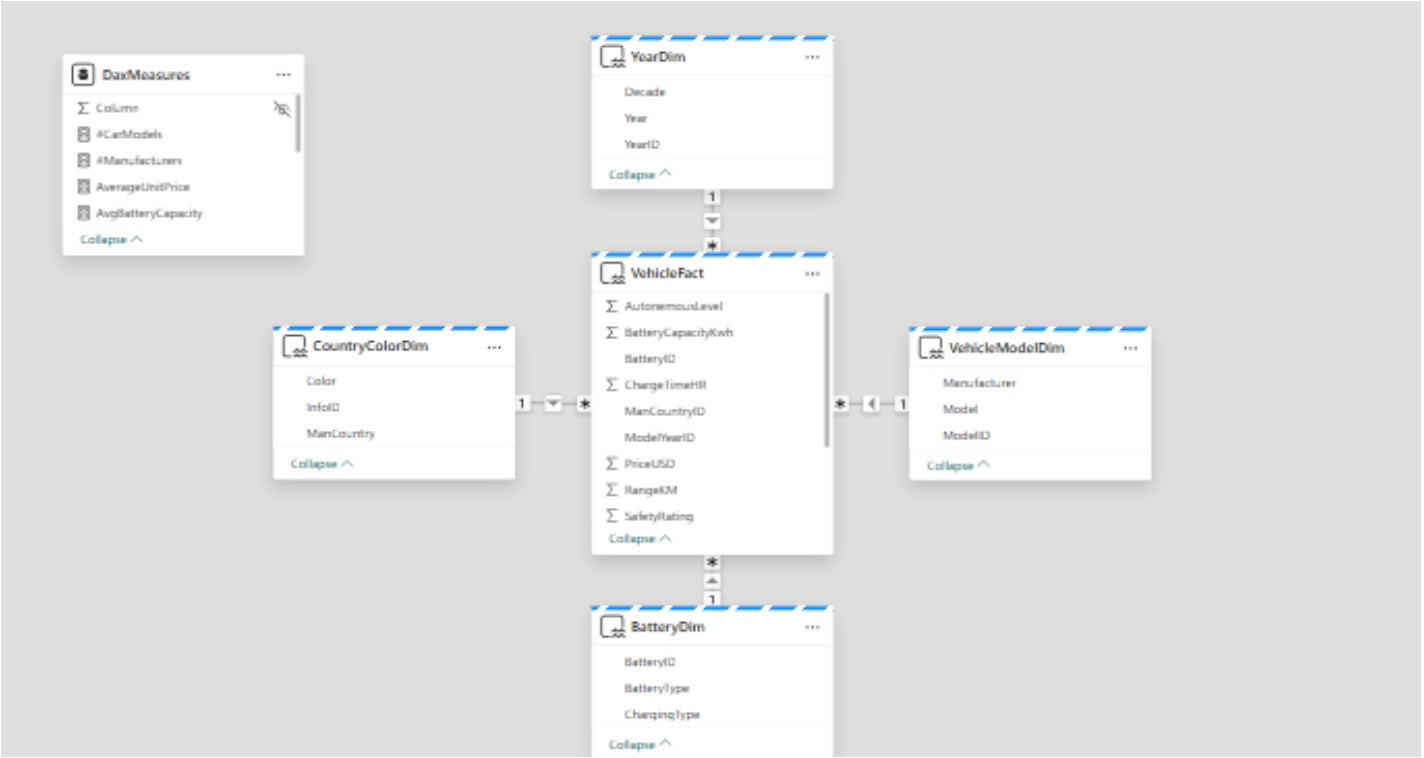


4. Semantic Model

The **semantic layer** was developed to enhance the user experience and empower self-service analytics. This layer provides business-friendly metadata by abstracting complex relationships and raw data structures into intuitive models.

After Filling the Datawarehouse We have created a new semantic model as follows:

1) Determining the Tables for the semantic model.
2) Creating the relationships between the Dimensions and the fact.



3) After Building the Relationships, we have created some measured in a Dax Created Table just for the measures which holds some measures like follows:

a. `#CarModels = DISTINCTCOUNT(VehicleFact[VehicleIModelD])`

b.
```
#Manufacturers =
        CALCULATE (
            DISTINCTCOUNT ( VehicleModelDim[Manufacturer] ),
            TREATAS ( VALUES ( VehicleFact[VehicleIModelD] ),
        VehicleModelDim[ModelID] )
        )
```

c. **AverageUnitPrice** = *[UnitsSoldPrice] / [TotalUnitsSold]*

d. **AvgBatteryCapacity** = *AVERAGE(VehicleFact[BatteryCapacityKwh])*

e. **AvgRangeKM** = *AVERAGE(VehicleFact[RangeKM])*

f. **AvgRechargingHr** = *AVERAGE(VehicleFact[ChargeTimeHR])*

g. **DistinctBatteryTypes** =

```
CALCULATE (
    DISTINCTCOUNT ( BatteryDim[BatteryType] ),
    CROSSFILTER ( VehicleFact[BatteryID],
BatteryDim[BatteryID], BOTH )
)
```

h. **DistinctChargingTypes** =

```
CALCULATE (
    DISTINCTCOUNT ( BatteryDim[ChargingType] ),
    CROSSFILTER ( VehicleFact[BatteryID],
BatteryDim[BatteryID], BOTH )
)
```

i. **TotalUnitsSold** = *SUM(VehicleFact[UnitsSold2024])*

j. **UnitsSoldPrice** = *SUMX(VehicleFact, VehicleFact[PriceUSD] * VehicleFact[UnitsSold2024])*

5. Power BI Overview

After Creating the semantic layer with its measures and relationships, It's time to Create the Power BI dashboard provides an interactive visualization layer for the data warehouse. It enables users to monitor key performance indicators, analyze customer loyalty trends, and explore flight and demographic data. Each report page is designed to deliver clear, actionable insights with user-friendly navigation and filtering options:

The created dashboard contains 4 pages as follows:

1. Home Page.
2. Data Overview
3. EV Sales Analysis
4. Battery TypesAnalysis

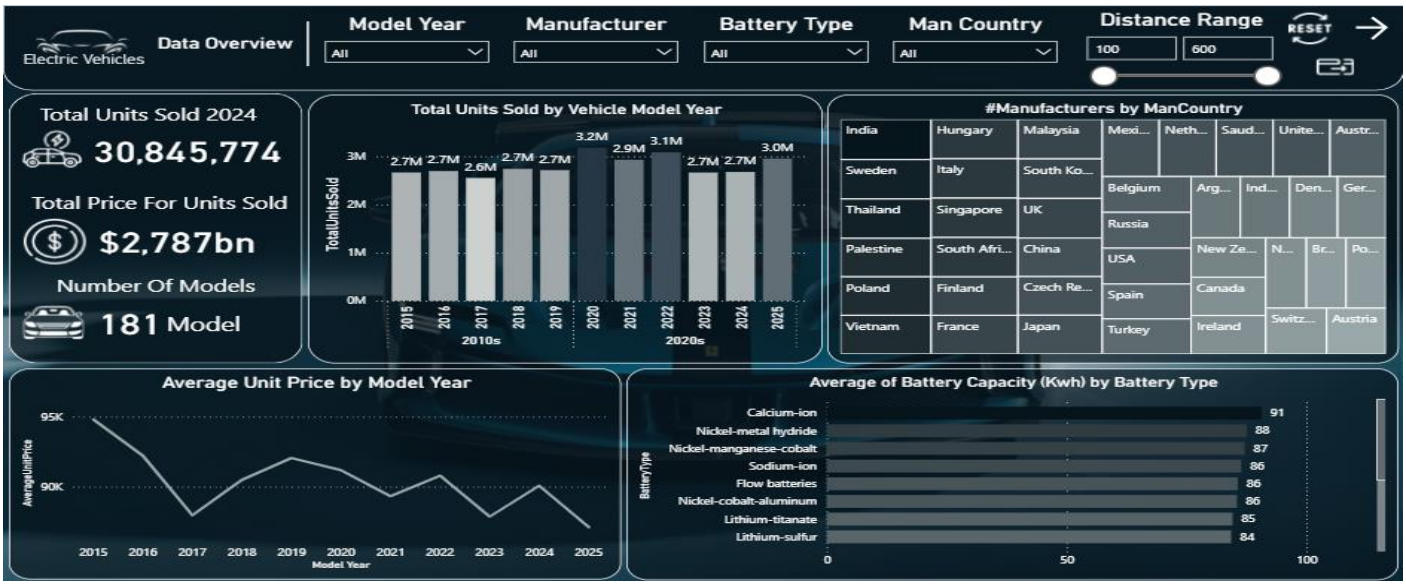
These Pages are described and shows as follows:

1) Home Page:

- It is the page that shows the Project title and contains the Navigation Buttons to the other Pages within the Dashboard.

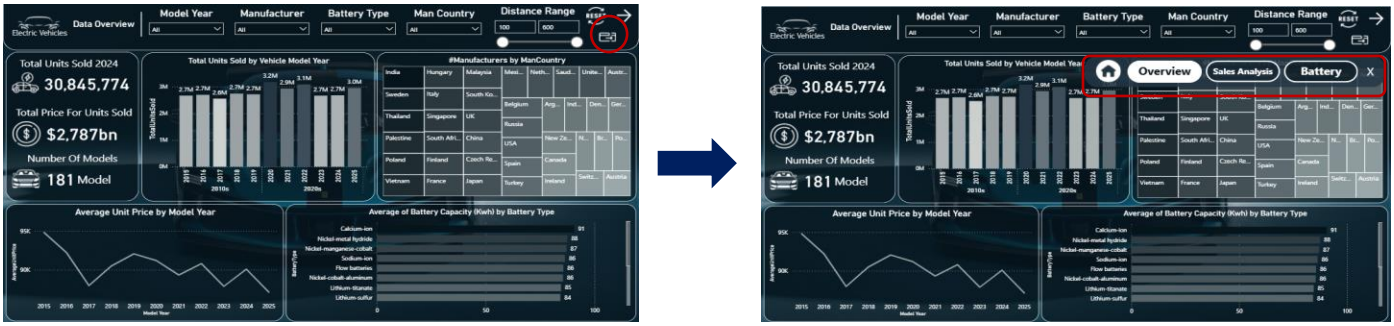
2) Overview Page:

- This Page contains a Historical Overview For the data like:
  - Different Electrical Vehicles With there model and KPIs.
  - The Sales Information for units sold in 2024.
  - The total price for all the EV units sold in 2024.

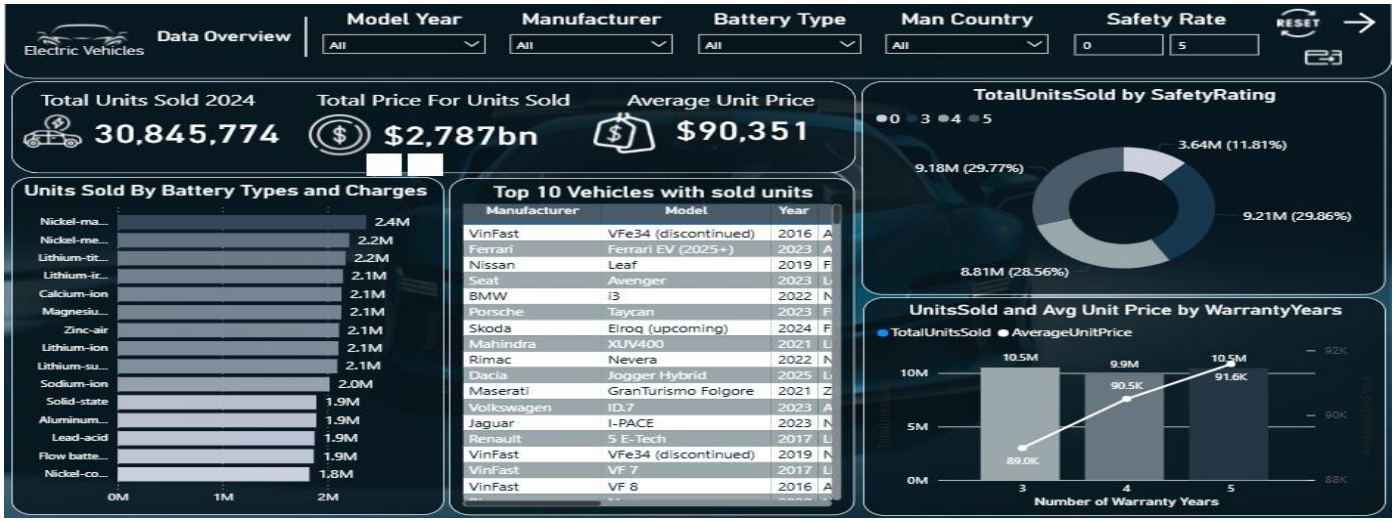


► The Navigation for each page can be accessed through the navigation button exists in the page as follows.

➢ **The Page Navigation:**



4) EV SalesAnalysis



3) Battery Analysis