ISM 6208.020U24.50087
Data Warehousing

Summer 2024

US Electric Vehicle Population Data

Prof. Donald Berndt

University of South Florida

Muma College of Business Master of Science in

Business Analytics and Information Systems

Group Members

Neha Satpute

Gopi Kavuri

Krishna Sumanth Pulikonda

Narayan Rohith Reddy Pasham

Introduction

In response to the growing importance of sustainable transportation solutions, this project focuses on analyzing the electric vehicle (EV) population in the United States. Utilizing **Microsoft SQL Server Management Studio for data management** and a leading data visualization tool, this study aims to provide deep insights into the adoption trends and distributions of electric vehicles across various states. This project leverages modern data warehousing techniques to handle large datasets effectively, ensuring that data-driven decisions can be made to promote eco-friendly transportation alternatives.

Executive Summary

This project is designed to leverage cutting-edge data warehousing and analytics technologies to explore the electric vehicle market in the United States. By utilizing Microsoft SQL Server Management Studio and advanced data visualization tools, we aim to uncover significant patterns and trends that could influence future policies and business strategies. The insights derived from this analysis are intended to support stakeholders in making informed decisions that could accelerate the adoption of electric vehicles, thereby contributing to environmental sustainability and reducing carbon footprints.

Problem Statement

Despite the push for cleaner energy solutions, the adoption rate of electric vehicles varies significantly across different regions in the United States. Understanding the factors influencing these variations can help policymakers, manufacturers, and consumers make better decisions that could potentially accelerate the adoption of electric vehicles. This project seeks to analyze various demographic and economic factors to identify key drivers and barriers in the electric vehicle market.

About the Dataset

The dataset used in this project was sourced from Kaggle and contains comprehensive information on the electric vehicle population within the United States. It includes details such as vehicle type, make, model, year, and geographic information regarding the distribution and density of electric vehicles.

Detailed Description of the Dataset

1. **Vehicle Identification**:

- VIN (Vehicle Identification Number): This unique identifier for each vehicle allows for detailed tracking and record-keeping.
- Make: The brand of the vehicle, e.g., Tesla, Nissan, Chevrolet.
- Model: The specific model of the vehicle, e.g., Model S, Leaf, Bolt.
- Model Year: The production year of the vehicle, which helps in analyzing trends related to the release of new models and technology updates.

2. Vehicle Specifications:

- Type: Specifies whether the vehicle is a Battery Electric Vehicle (BEV) or a Plug-in Hybrid Electric Vehicle (PHEV).
- Battery Capacity: Measured in kWh, indicating the energy storage capacity of the vehicle's battery.
- Electric Range: The maximum distance the vehicle can travel on a single charge, measured in miles.

3. Registration Information:

- State: The U.S. state where the vehicle is registered, which is crucial for geographical trend analysis.
- County: More localized registration data, offering insights into urban versus rural adoption rates.
- ZIP Code: Even more granular data that can be used for detailed regional analysis and mapping.

4. **Usage Information**:

- Annual Mileage: Estimated yearly mileage, providing insights into how electric vehicles are used compared to conventional vehicles.
- Vehicle Purpose: Classifies the primary use of the vehicle, e.g., personal, commercial, or governmental use.

5. **Economic Indicators**:

- Purchase Type: Indicates whether the vehicle was purchased new or used, which can impact adoption rates.
- Federal Tax Credit Used: Whether the buyer utilized federal tax incentives for purchasing an electric vehicle, which can be a significant driver of sales.

6. **Environmental Impact**:

- Emissions Saved: An estimate of the reduction in carbon emissions due to the use of the electric vehicle, compared to an average gasoline vehicle. This data is crucial for analyzing the environmental impact of transitioning to electric vehicles.

7. **Date and Time Stamps**:

- Registration Date: The date on which the vehicle was registered, useful for tracking adoption trends over time.
- Last Update: The most recent update to the vehicle's record, providing information on the dataset's freshness and relevance.

Based on these columns in dataset, we can design a star schema for an ETL process. Here's a high-level approach:

Dimension Tables:

1. Vehicle Dimension: Dim_Vehicle
    - VehicleID (surrogate key)
    - VIN (business key)
    - Make
    - Model
    - Model Year
2. Location Dimension: Dim_Location
    - LocationID (surrogate key)
    - County
    - City
    - State
    - Postal Code
    - Vehicle Location
    - 2020 Census Tract
3. Utility Dimension: Dim_Utility
    - UtilityID (surrogate key)
    - Electric Utility

Fact Table:

Fact_EV_Sales :

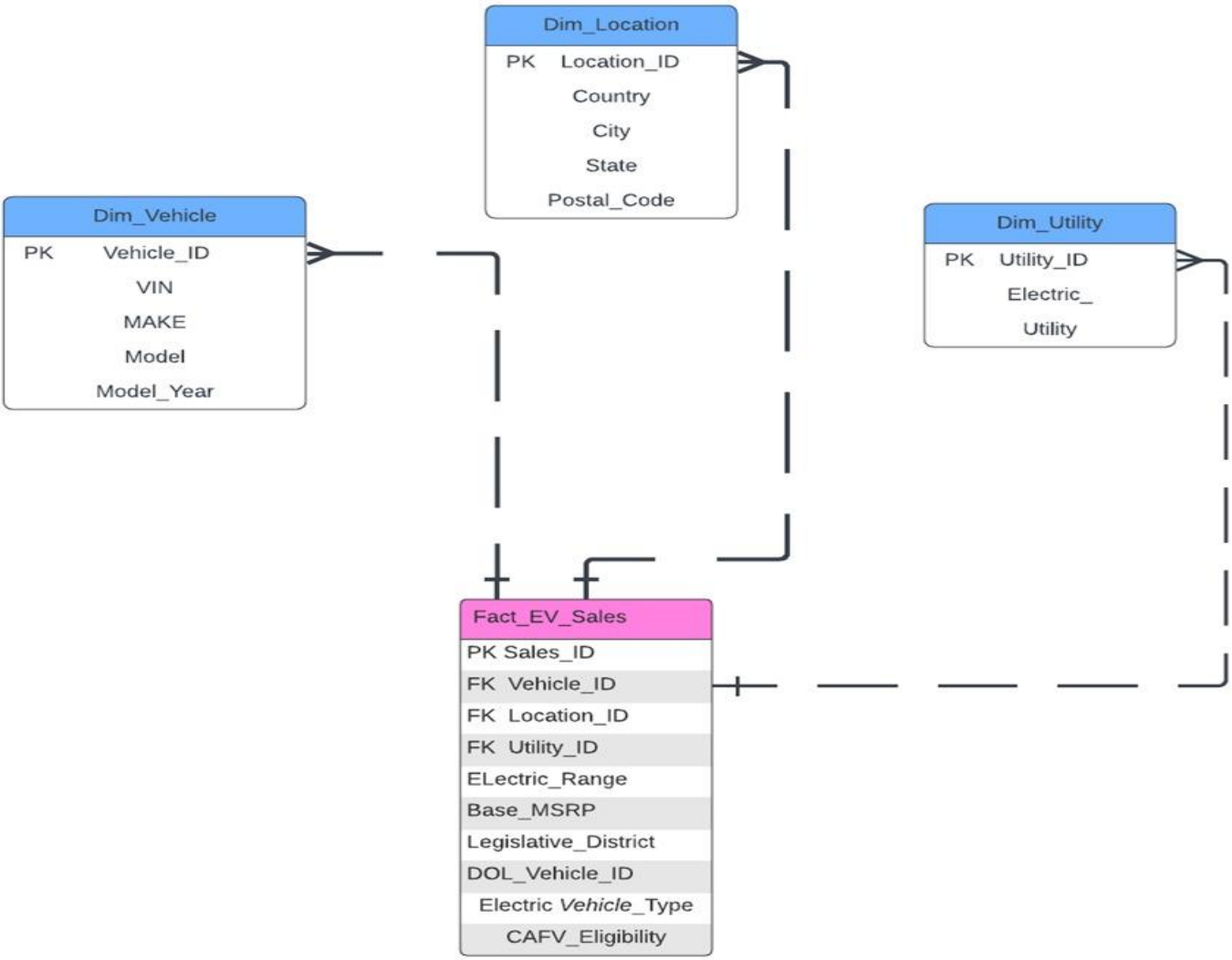
- SalesID (surrogate key)
- VehicleID
- LocationID
- UtilityID
- Electric Range
- Base MSRP
- DOL Vehicle ID
- Electric Vehicle Type
- CAFV Eligibility
- Legislative District

Stagging Table:

Staging_EV Table:

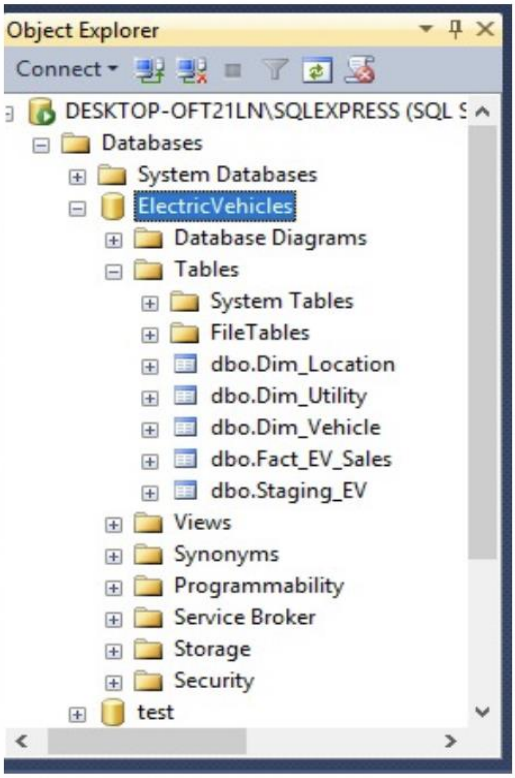
- VIN
- County
- City
- State
- Postal_Code
- Model_Year
- Make
- Model
- Electric_Vehicle_Type
- CAFV_Eligibility
- Electric_Range
- Base_MSRP
- Legislative_District
- DOL_Vehicle_ID
- Vehicle_Location
- Electric_Utility
- Census_Tract

ERD Diagram:


Dim_Location
PK Location_ID
Country
City
State
Postal_Code
Dim_Vehicle
PK Vehicle_ID
VIN
MAKE
Model
Model_Year
Dim_Utility
PK Utility_ID
Electric_
Utility
Fact_EV_Sales
PK Sales_ID
FK Vehicle_ID
FK Location_ID
FK Utility_ID
ELectric_Range
Base_MSRP
Legislative_District
DOL_Vehicle_ID
Electric Vehicle_Type
CAFV_Eligibility

ETL Process

1. Created staging table, dimensions and fact table




```
-- Create Location Dimension
CREATE TABLE Dim_Location (
    location_id INT IDENTITY(1,1) PRIMARY KEY,
    county VARCHAR(50),
    city VARCHAR(50),
    state CHAR(2),
    postal_code INT,
    vehicle_location VARCHAR(100),
    census_tract BIGINT,
    UNIQUE (county, city, state, postal_code)
);

-- Create Utility Dimension
CREATE TABLE Dim_Utility (
    utility_id INT IDENTITY(1,1) PRIMARY KEY,
    electric_utility VARCHAR(100),
    UNIQUE (electric_utility)
);
-- Create EV Sales Fact Table
CREATE TABLE Fact_EV_Sales (
    fact_id INT IDENTITY(1,1) PRIMARY KEY,
    vehicle_id INT,
    location_id INT,
    utility_id INT,
    electric_range INT,
    base_msrp INT,
    dol_vehicle_id BIGINT,
    electric_vehicle_type VARCHAR(50),
    cafv_eligibility VARCHAR(100),
```

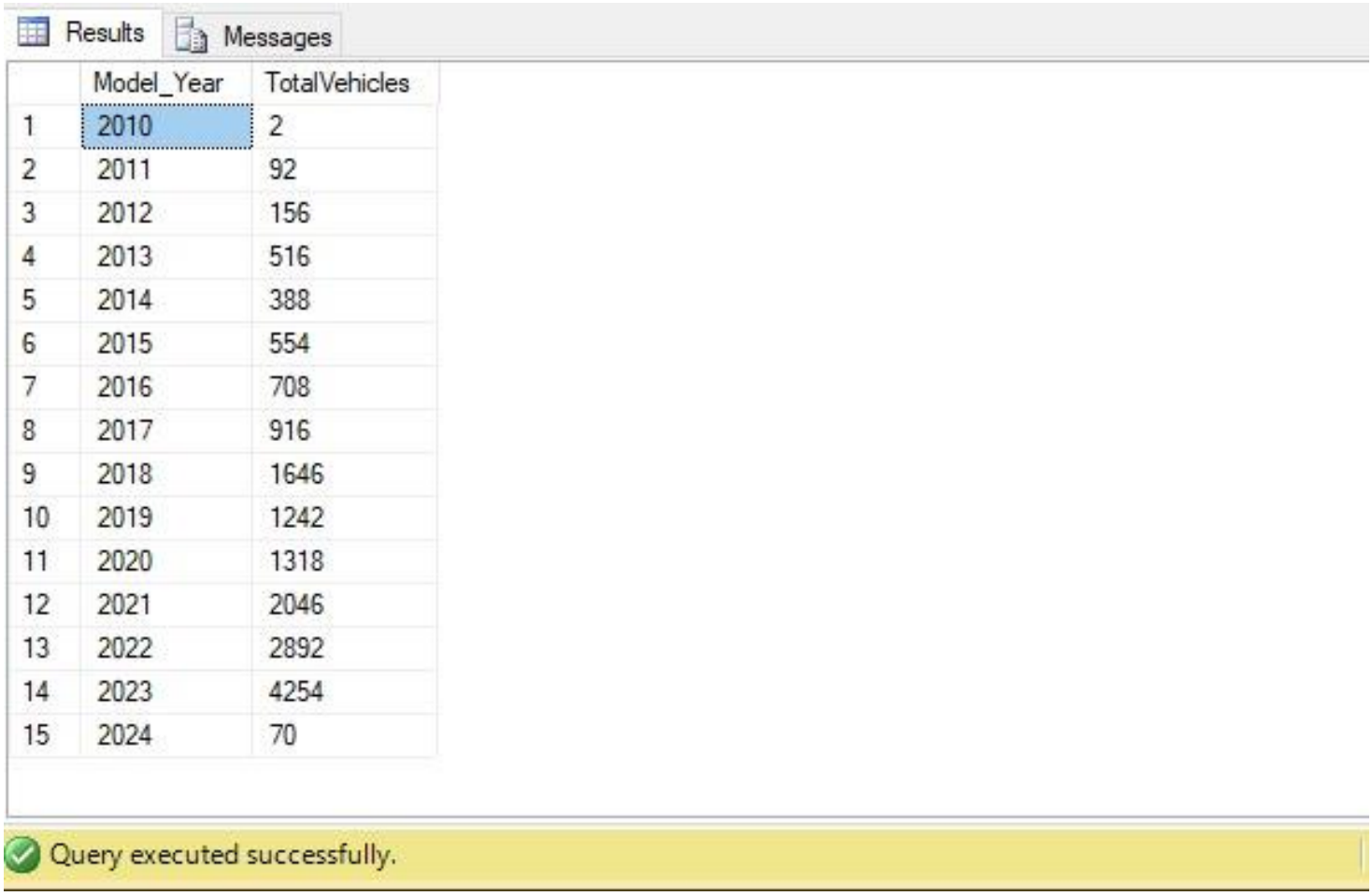
```sql
-- Create Location Dimension
CREATE TABLE Dim_Location (
    location_id INT IDENTITY(1,1) PRIMARY KEY,
    county VARCHAR(50),
    city VARCHAR(50),
    state CHAR(2),
    postal_code INT,
    vehicle_location VARCHAR(100),
    census_tract BIGINT,
    UNIQUE (county, city, state, postal_code)
);

-- Create Utility Dimension
CREATE TABLE Dim_Utility (
    utility_id INT IDENTITY(1,1) PRIMARY KEY,
    electric_utility VARCHAR(100),
    UNIQUE (electric_utility)
);
-- Create EV Sales Fact Table
CREATE TABLE Fact_EV_Sales (
    fact_id INT IDENTITY(1,1) PRIMARY KEY,
    vehicle_id INT,
    location_id INT,
    utility_id INT,
    electric_range INT,
    base_msrp INT,
    dol_vehicle_id BIGINT,
    electric_vehicle_type VARCHAR(50),
    cafv_eligibility VARCHAR(100),
```

```sql
USE ElectricVehicles;
-- Populate Vehicle Dimension
INSERT INTO Dim_Vehicle (vin, make, model, model_year)
SELECT DISTINCT VIN, Make, Model, Model_Year
FROM Staging_EV;

-- Populate Location Dimension
INSERT INTO Dim_Location (county, city, state, postal_code, vehicle_location, census_tract)
SELECT DISTINCT County, City, State, Postal_Code, Vehicle_Location, Census_Tract
FROM Staging_EV;

-- Populate Utility Dimension
INSERT INTO Dim_Utility (electric_utility)
SELECT DISTINCT Electric_Utility
FROM Staging_EV;

-- Populate Fact EV Sales Table
INSERT INTO Fact_EV_Sales (
    vehicle_id, location_id, utility_id, electric_range, base_msrp, dol_vehicle_id,
    electric_vehicle_type, cafv_eligibility, legislative_district
)
SELECT
    v.vehicle_id,
    l.location_id,
    u.utility_id,
    EV.Electric_Range,
    EV.Base_MSRP,
    EV.DOL_Vehicle_ID,
    EV.Electric_Vehicle_Type,
    EV.CAFV_Eligibility,
    EV.Legislative_District
```

ANALYTICAL QUERIES

1.Query to calculate total Vehicles by Model Year

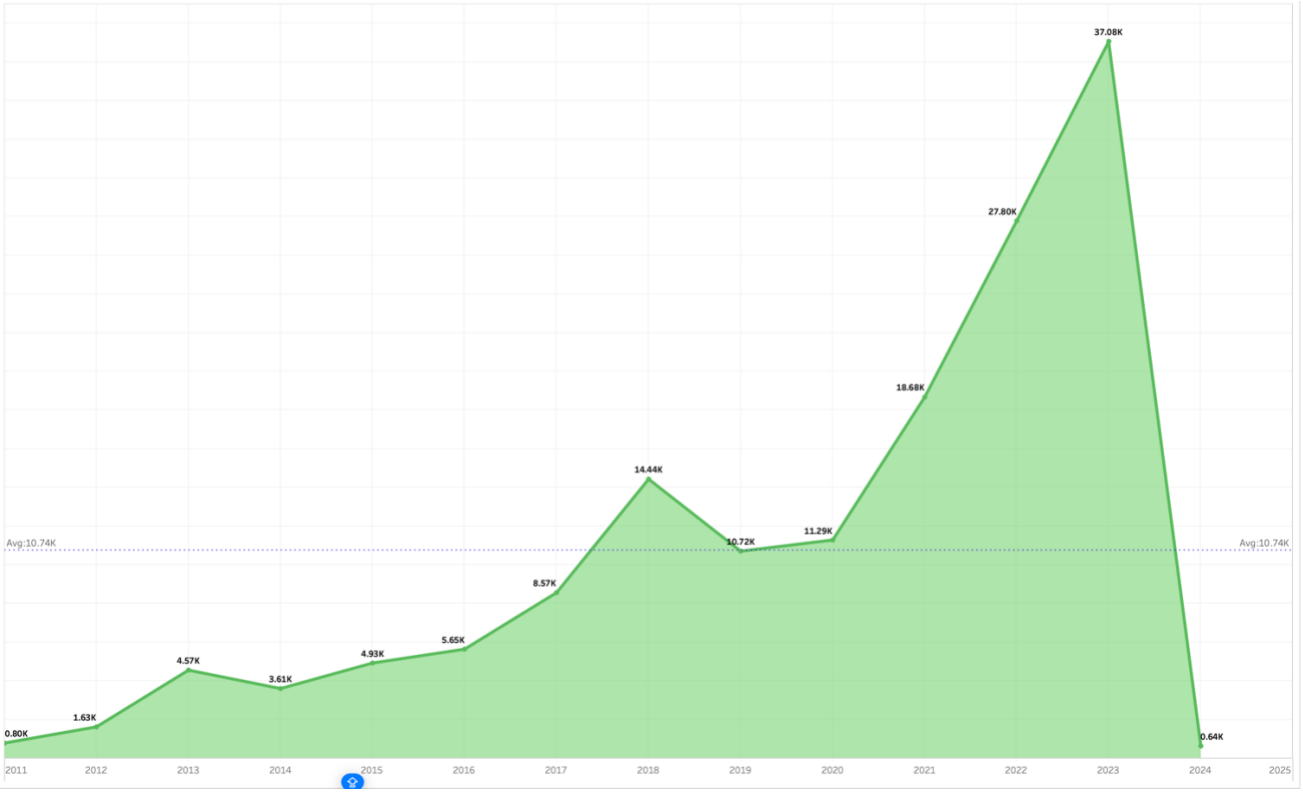
```
SELECT Model_Year, COUNT(*) AS TotalVehicles
FROM Staging_EV
GROUP BY Model_Year
ORDER BY Model_Year;
```



Results | Messages

	Model_Year	TotalVehicles
1	2010	2
2	2011	92
3	2012	156
4	2013	516
5	2014	388
6	2015	554
7	2016	708
8	2017	916
9	2018	1646
10	2019	1242
11	2020	1318
12	2021	2046
13	2022	2892
14	2023	4254
15	2024	70

Query executed successfully.



Visualization for the above query

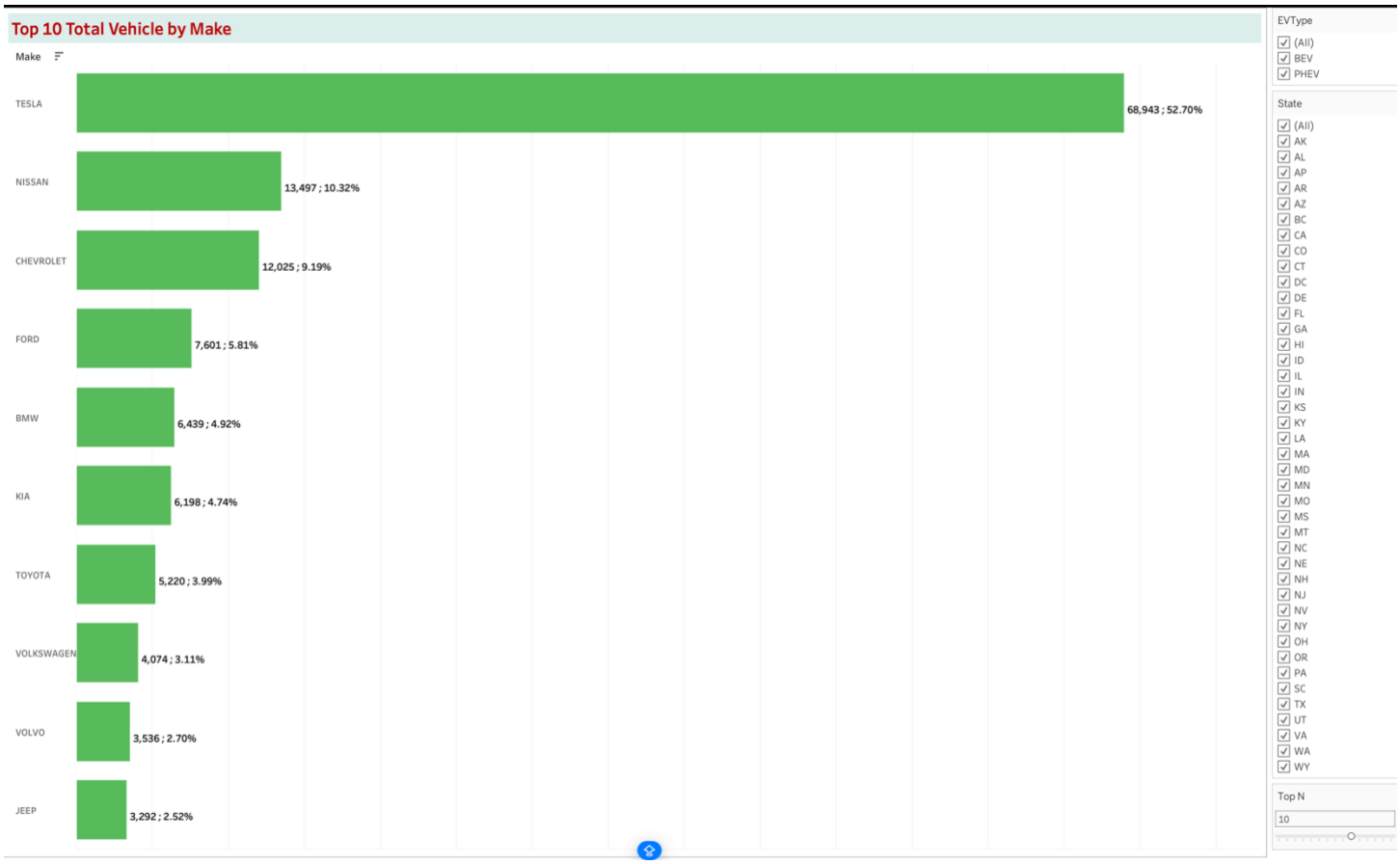
2.Query to find top 10 vehicle makes by number of vehicles registered

```
SELECT TOP 10 Make, COUNT(*) AS TotalVehicles
FROM Staging_EV
GROUP BY Make
ORDER BY TotalVehicles DESC;
```

Results | Messages

	Make	TotalVehicles
1	TESLA	7716
2	NISSAN	1670
3	CHEVROLET	1134
4	BMW	762
5	KIA	714
6	FORD	694
7	TOYOTA	526
8	HYUNDAI	444
9	VOLVO	436
10	VOLKSWAGEN	414

Visualization for the above query



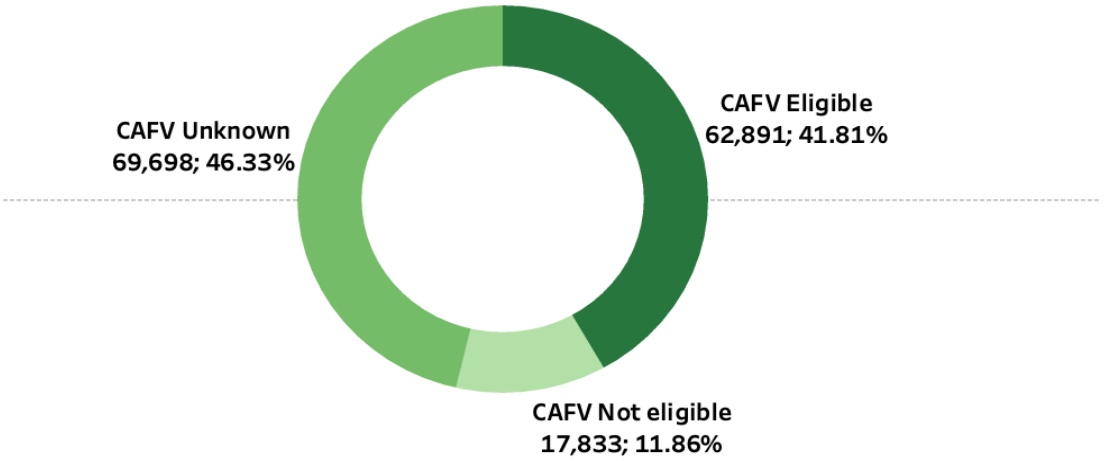
3.Query to calculate total vehicles by CAFV Eligibility

```
SELECT CAFV_Eligibility, COUNT(*) AS TotalVehicles
FROM Fact_EV_Sales
GROUP BY CAFV_Eligibility
ORDER BY TotalVehicles DESC;
```

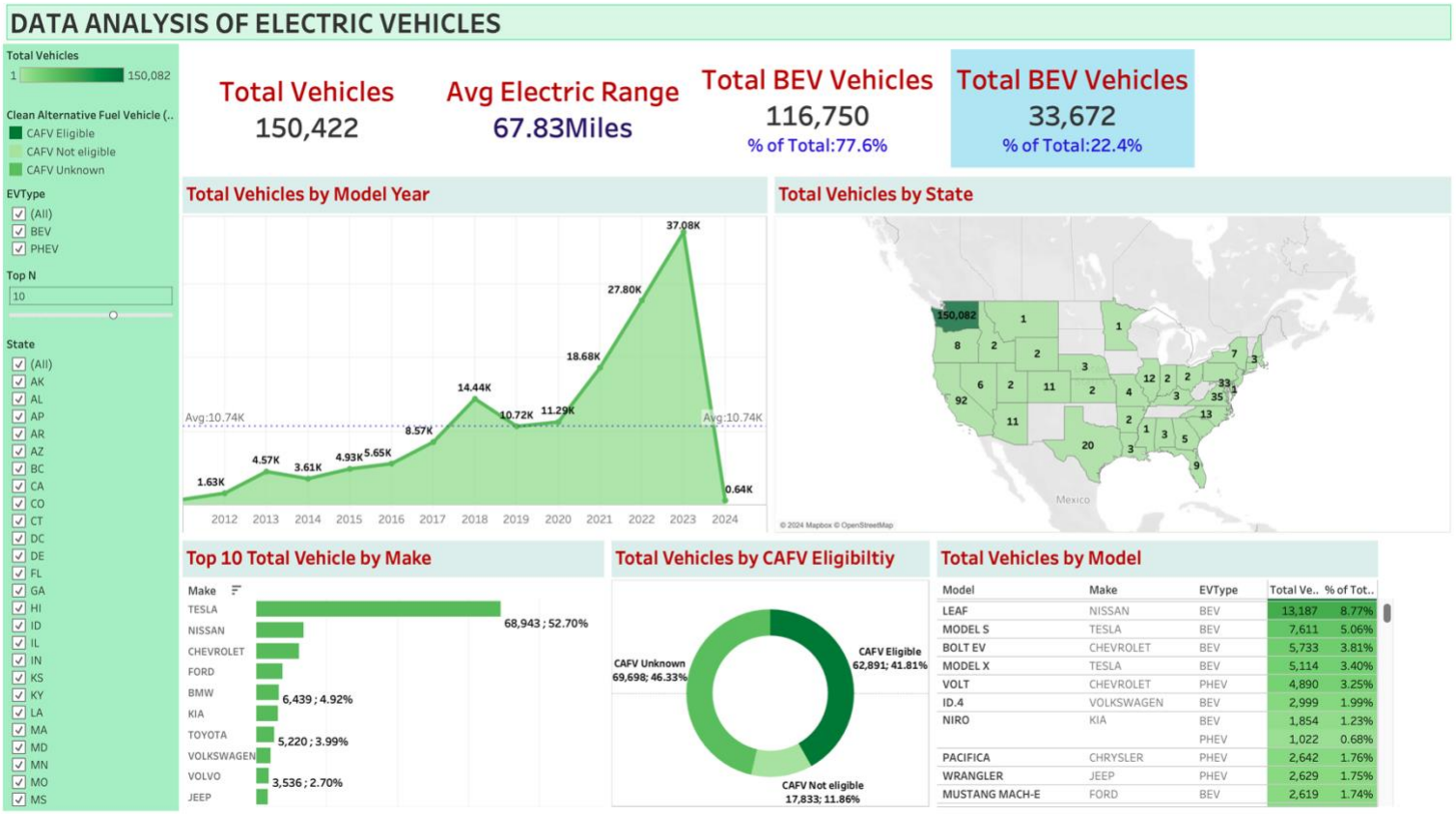
Results | Messages

	CAFV_Eligibility	TotalVehicles
1	Eligibility unknown as battery range has not be...	7588
2	Clean Alternative Fuel Vehicle Eligible	7424
3	Not eligible due to low battery range	1788

Visualization for the above query



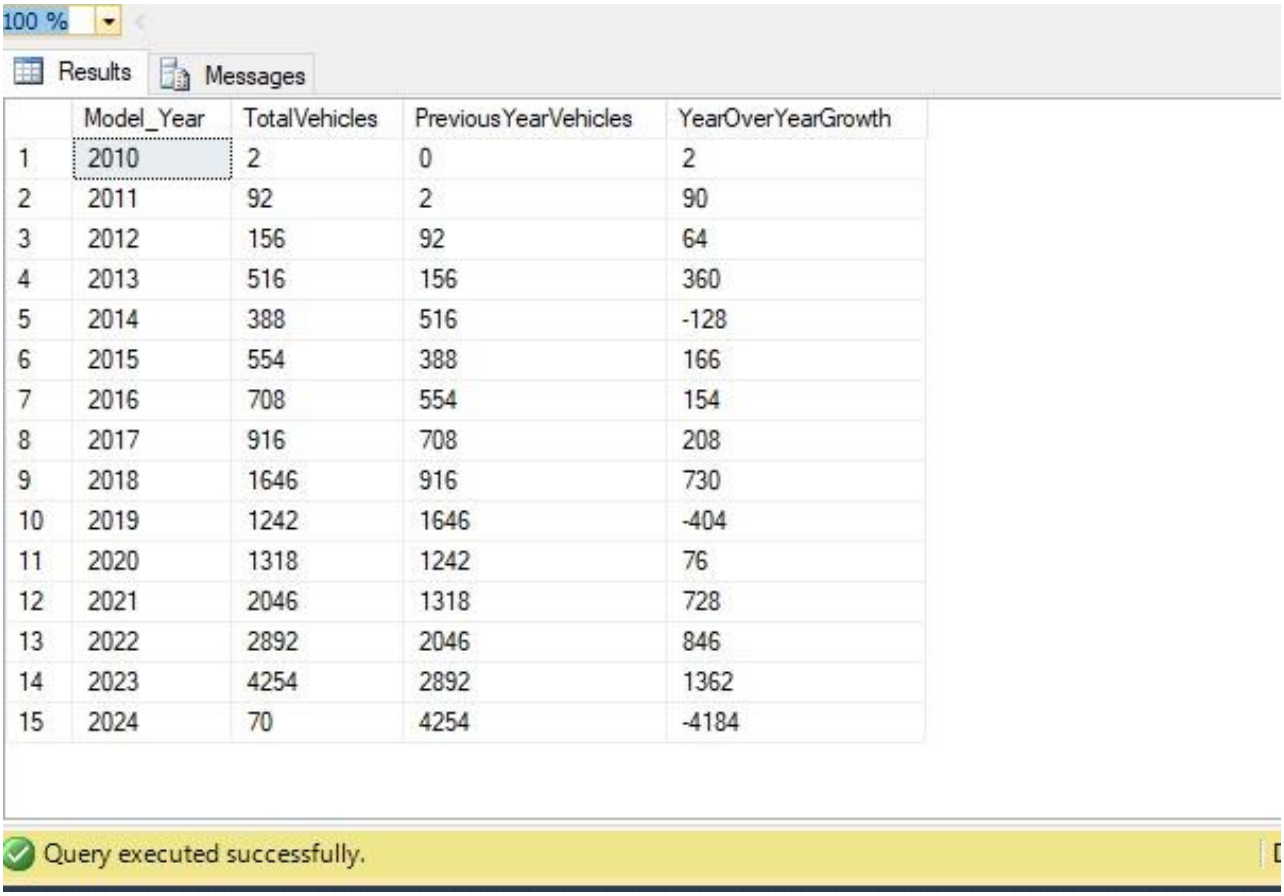
DASHBOARD FOR THE SAME DATA



More Analytical Queries

1.Electric Vehicle Growth over Years

```
SELECT Model_Year,
       TotalVehicles,
       PreviousYearVehicles,
       TotalVehicles - PreviousYearVehicles AS YearOverYearGrowth
FROM (
    SELECT d.Model_Year,
           COUNT(*) AS TotalVehicles,
           LAG(COUNT(*), 1, 0) OVER (ORDER BY d.Model_Year) AS PreviousYearVehicles
    FROM Fact_EV_Sales f
    JOIN Dim_Vehicle d ON f.Vehicle_ID = d.Vehicle_ID
    GROUP BY d.Model_Year
) t
ORDER BY Model_Year;
```



100 %

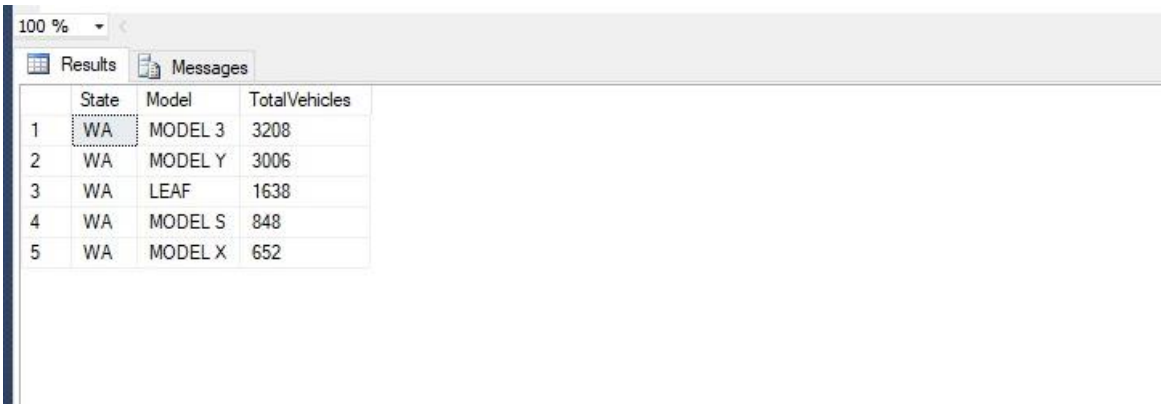
Results | Messages

	Model_Year	TotalVehicles	PreviousYearVehicles	YearOverYearGrowth
1	2010	2	0	2
2	2011	92	2	90
3	2012	156	92	64
4	2013	516	156	360
5	2014	388	516	-128
6	2015	554	388	166
7	2016	708	554	154
8	2017	916	708	208
9	2018	1646	916	730
10	2019	1242	1646	-404
11	2020	1318	1242	76
12	2021	2046	1318	728
13	2022	2892	2046	846
14	2023	4254	2892	1362
15	2024	70	4254	-4184

Query executed successfully.

2.Top 5 Model in Each State

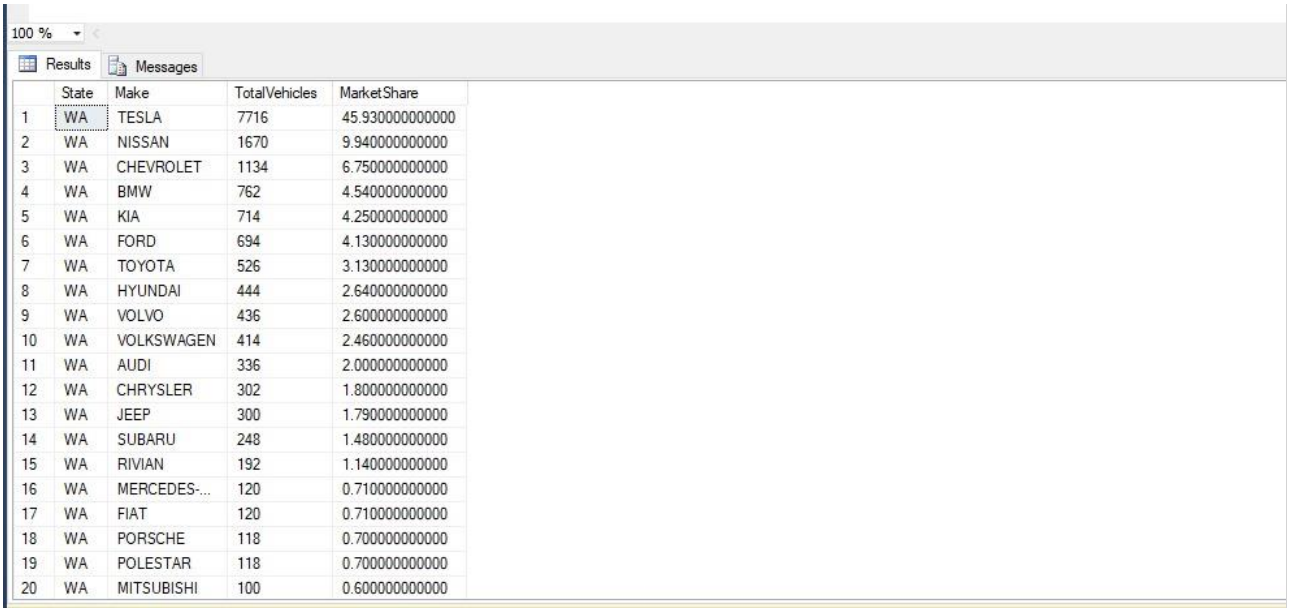
```
SELECT State, Model, TotalVehicles
FROM (
   SELECT l.State, v.Model,
        COUNT(*) AS TotalVehicles,
        ROW_NUMBER() OVER (PARTITION BY l.State ORDER BY COUNT(*) DESC) AS rn
   FROM Fact_EV_Sales f
   JOIN Dim_Vehicle v ON f.Vehicle_ID = v.Vehicle_ID
   JOIN Dim_Location l ON f.Location_ID = l.Location_ID
   GROUP BY l.State, v.Model
) AS SubQuery
WHERE rn <= 5
ORDER BY State, TotalVehicles DESC;
```



	State	Model	TotalVehicles
1	WA	MODEL 3	3208
2	WA	MODEL Y	3006
3	WA	LEAF	1638
4	WA	MODEL S	848
5	WA	MODEL X	652

3.State-wise Market Share of each Make

```
SELECT State, Make, TotalVehicles,

     ROUND((TotalVehicles * 100.0 / SUM(TotalVehicles) OVER (PARTITION BY State)), 2) AS
MarketShare
FROM (
   SELECT l.State, v.Make, COUNT(*) AS TotalVehicles
   FROM Fact_EV_Sales f
   JOIN Dim_Vehicle v ON f.Vehicle_ID = v.Vehicle_ID
   JOIN Dim_Location l ON f.Location_ID = l.Location_ID
   GROUP BY l.State, v.Make
) AS SubQuery
ORDER BY State, MarketShare DESC;
```



100 %

Results | Messages

	State	Make	TotalVehicles	MarketShare
1	WA	TESLA	7716	45.930000000000
2	WA	NISSAN	1670	9.940000000000
3	WA	CHEVROLET	1134	6.750000000000
4	WA	BMW	762	4.540000000000
5	WA	KIA	714	4.250000000000
6	WA	FORD	694	4.130000000000
7	WA	TOYOTA	526	3.130000000000
8	WA	HYUNDAI	444	2.640000000000
9	WA	VOLVO	436	2.600000000000
10	WA	VOLKSWAGEN	414	2.460000000000
11	WA	AUDI	336	2.000000000000
12	WA	CHRYSLER	302	1.800000000000
13	WA	JEEP	300	1.790000000000
14	WA	SUBARU	248	1.480000000000
15	WA	RIVIAN	192	1.140000000000
16	WA	MERCEDES-...	120	0.710000000000
17	WA	FIAT	120	0.710000000000
18	WA	PORSCHE	118	0.700000000000
19	WA	POLESTAR	118	0.700000000000
20	WA	MITSUBISHI	100	0.600000000000

4.Average Electric Range by Vehicle type, Highlighting Outlier

```
SELECT Model, Electric_Vehicle_Type, Electric_Range,
      AVG(Electric_Range) OVER (PARTITION BY Electric_Vehicle_Type) AS AvgRange,
      CASE
         WHEN Electric_Range > AVG(Electric_Range) OVER (PARTITION BY Electric_Vehicle_Type)
THEN 'Above Average'
         WHEN Electric_Range < AVG(Electric_Range) OVER (PARTITION BY Electric_Vehicle_Type)
THEN 'Below Average'
         ELSE 'Average'
      END AS RangeCategory
FROM (
   SELECT v.Model, v.Electric_Vehicle_Type, f.Electric_Range
   FROM Fact_EV_Sales f
   JOIN Dim_Vehicle v ON f.Vehicle_ID = v.Vehicle_ID
) AS SubQuery
ORDER BY Electric_Vehicle_Type, Electric_Range;
```

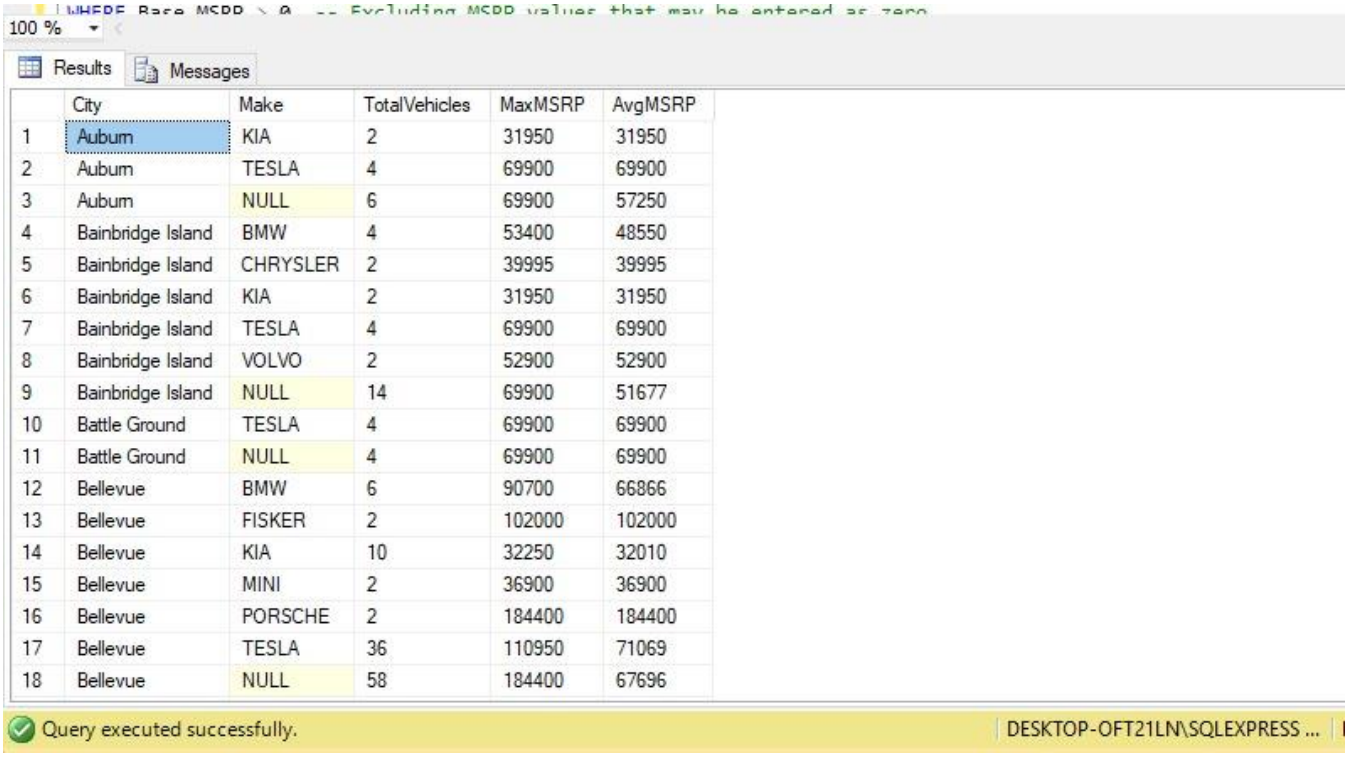
00 %

Results | Messages

	Model	Electric_Vehicle_Type	Electric_Range	AvgRange	RangeCategory
1	MODEL Y	Battery Electric Vehicle (BEV)	0	82	Below Average
2	MODEL 3	Battery Electric Vehicle (BEV)	0	82	Below Average
3	ID.4	Battery Electric Vehicle (BEV)	0	82	Below Average
4	BOLT EV	Battery Electric Vehicle (BEV)	0	82	Below Average
5	MODEL Y	Battery Electric Vehicle (BEV)	0	82	Below Average
6	MODEL 3	Battery Electric Vehicle (BEV)	0	82	Below Average
7	MODEL 3	Battery Electric Vehicle (BEV)	0	82	Below Average
8	MODEL Y	Battery Electric Vehicle (BEV)	0	82	Below Average
9	MODEL Y	Battery Electric Vehicle (BEV)	0	82	Below Average
10	R1S	Battery Electric Vehicle (BEV)	0	82	Below Average
11	MODEL Y	Battery Electric Vehicle (BEV)	0	82	Below Average
12	MODEL Y	Battery Electric Vehicle (BEV)	0	82	Below Average
13	MODEL Y	Battery Electric Vehicle (BEV)	0	82	Below Average
14	MODEL 3	Battery Electric Vehicle (BEV)	0	82	Below Average
15	IONIQ 5	Battery Electric Vehicle (BEV)	0	82	Below Average
16	MODEL Y	Battery Electric Vehicle (BEV)	0	82	Below Average
17	LEAF	Battery Electric Vehicle (BEV)	0	82	Below Average
18	MODEL Y	Battery Electric Vehicle (BEV)	0	82	Below Average
19	MODEL Y	Battery Electric Vehicle (BEV)	0	82	Below Average
20	MUSTA...	Battery Electric Vehicle (BEV)	0	82	Below Average
21	R1T	Battery Electric Vehicle (BEV)	0	82	Below Average

5.Detail Report by City and Make with Aggregates

```
SELECT
    l.City,
    v.Make,
    COUNT(*) AS TotalVehicles,
    MAX(f.Base_MSRP) AS MaxMSRP,
    AVG(f.Base_MSRP) AS AvgMSRP
FROM
    Fact_EV_Sales f
JOIN
    Dim_Vehicle v ON f.Vehicle_ID = v.Vehicle_ID
JOIN
    Dim_Location l ON f.Location_ID = l.Location_ID
WHERE
    f.Base_MSRP > 0  -- Excluding MSRP values that may be entered as zero
GROUP BY
    ROLLUP (l.City, v.Make)
ORDER BY
    CASE WHEN l.City IS NULL THEN 1 ELSE 0 END, l.City,  -- Handling NULLs last for City
    CASE WHEN v.Make IS NULL THEN 1 ELSE 0 END, v.Make;  -- Handling NULLs last for Make
```



WHERE Base MSRP > 0 -- Excluding MSRP values that may be entered as zero

100 %

Results | Messages

	City	Make	TotalVehicles	MaxMSRP	AvgMSRP
1	Auburn	KIA	2	31950	31950
2	Auburn	TESLA	4	69900	69900
3	Auburn	NULL	6	69900	57250
4	Bainbridge Island	BMW	4	53400	48550
5	Bainbridge Island	CHRYSLER	2	39995	39995
6	Bainbridge Island	KIA	2	31950	31950
7	Bainbridge Island	TESLA	4	69900	69900
8	Bainbridge Island	VOLVO	2	52900	52900
9	Bainbridge Island	NULL	14	69900	51677
10	Battle Ground	TESLA	4	69900	69900
11	Battle Ground	NULL	4	69900	69900
12	Bellevue	BMW	6	90700	66866
13	Bellevue	FISKER	2	102000	102000
14	Bellevue	KIA	10	32250	32010
15	Bellevue	MINI	2	36900	36900
16	Bellevue	PORSCHE	2	184400	184400
17	Bellevue	TESLA	36	110950	71069
18	Bellevue	NULL	58	184400	67696

Query executed successfully. DESKTOP-OFT21LN\SQLEXPRESS ...

Future Scope of the Electric Vehicle Population Analysis Project

1. Predictive Analytics:

- Develop models to forecast EV adoption and simulate market penetration scenarios based on policy changes and economic factors.

2. Integration with Renewable Energy Data:

- Analyze the impact of charging infrastructure and renewable energy sources on EV adoption and environmental outcomes.

3. Enhanced Geographic Analysis:

- Conduct detailed comparisons of urban versus rural adoption and expand the dataset internationally for broader insights.

4. Consumer Behavior Studies:

- Investigate EV usage patterns and consumer attitudes through surveys and social media analysis.

5. Technological Advancements:

- Assess how innovations in battery and autonomous technologies influence EV performance and adoption.

6. Regulatory Impact Assessment:

- Evaluate the effects of regulations and policies on EV adoption and perform economic impact studies.