

Electric Vechicle Sales Analysis USING SQL And Machine learning

INTRODUCTION

OBJECTIVE

- Workflow of machine learning
- Source code
- Analysis using Sql

CONCLUSION

INTRODUCTION

The Electric Vehicle Sales Analysis project combines SQL, Python, and machine learning to analyze sales data and predict trends using the Random Forest model. This project showcases expertise in data querying, analysis, and predictive modeling for the growing EV market.

OBJECTIVE

Analyze and predict Electric Vehicle (EV) sales across states in India to gain insights into sales trends and support decision-making.

Key Deliverables:

- Understand historical trends in EV sales.
- Predict future EV sales based on specific inputs (year, state, vehicle type, and month)

WORKFLOW

1 Data Collection

I have Use a dataset named Electric Vehicle Sales by State in India. It has following Column

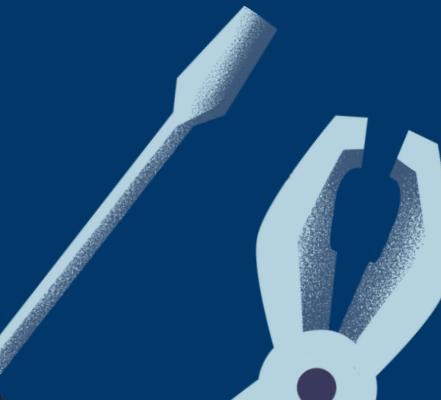
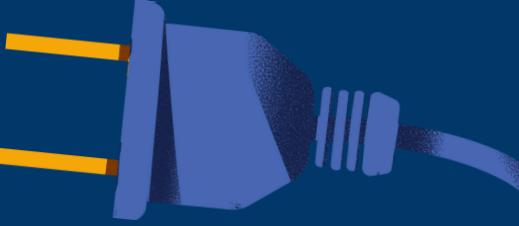
State: Name of the Indian state.

Vehicle_Type: Type of EV (e.g., 2W_Personal, 3W_Shared).

Year: Year of sales.

Month_Name: Name of the month.

EV_Sales_Quantity: Number of EVs sold.



WORKFLOW

2 Data Preprocessing

Encoding Categorical Variables:

- Map states to numerical values using a dictionary.
- Map vehicle types to numerical values using a dictionary.
- Map month names (Jan, Feb, etc.) to integers (1–12).

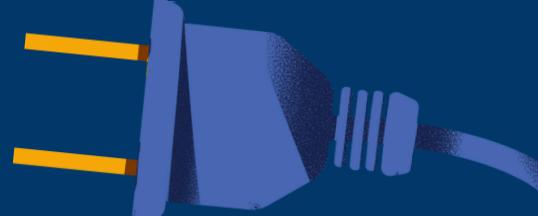
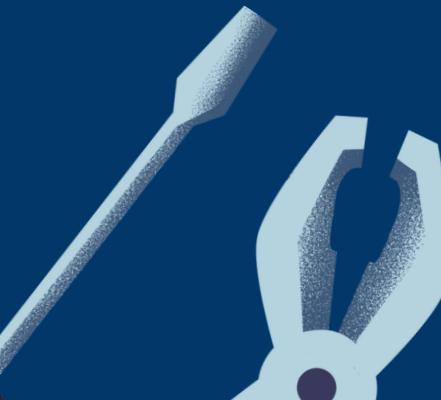
WORKFLOW

2 Data Preprocessing

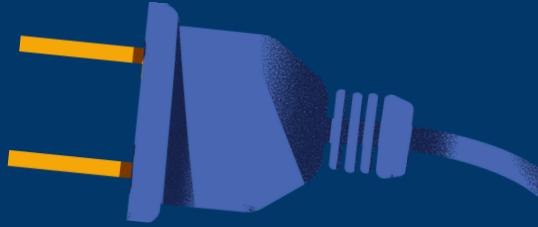
Feature Engineering:

Add encoded columns:

- State_encoded
- Vehicle_Type_encoded
- Month_Name_encoded



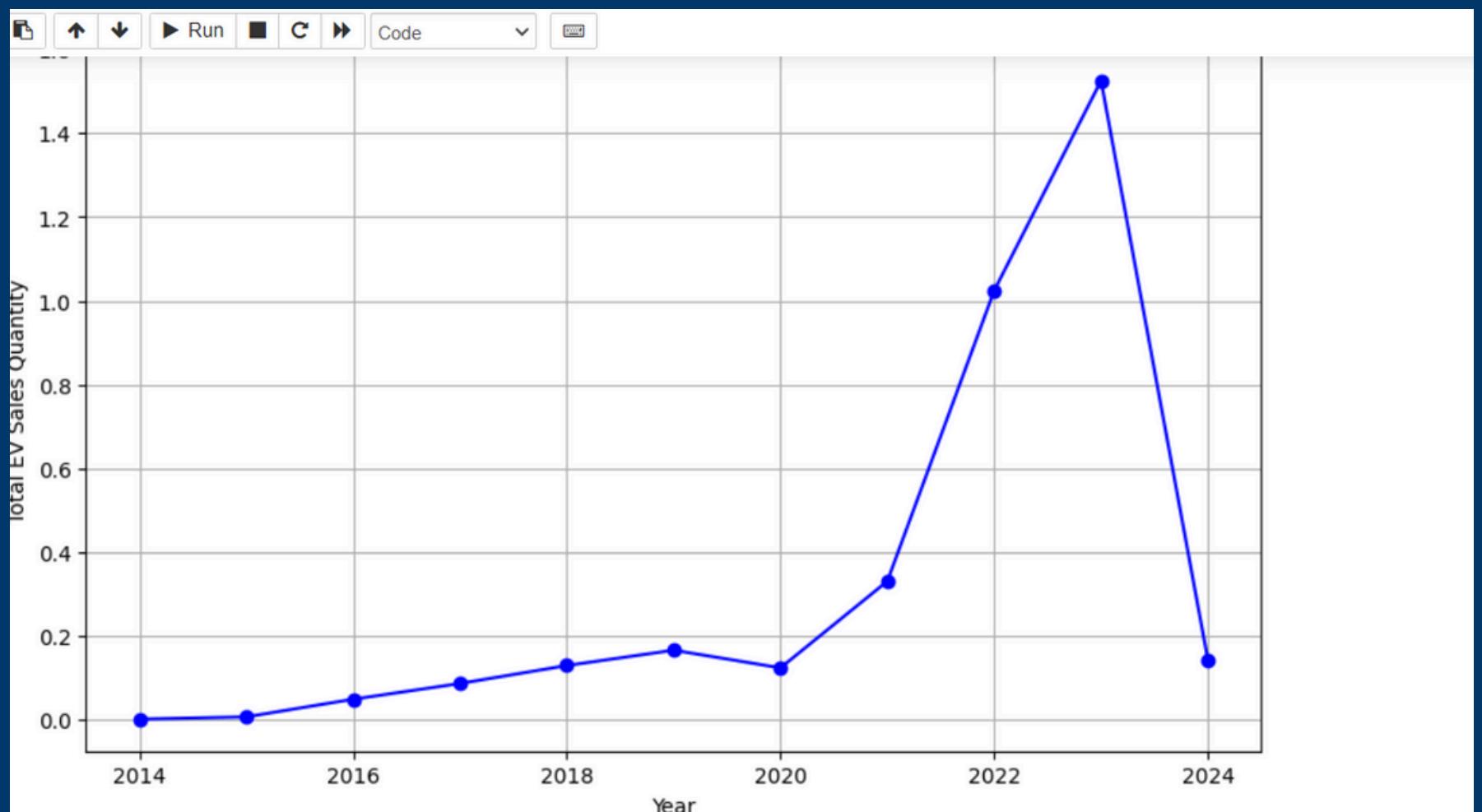
WORKFLOW



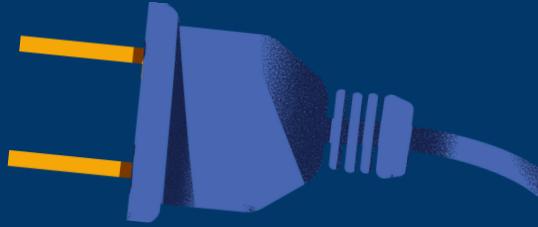
3 Exploratory Data Analysis (EDA)

1. Sales Trend by Year

Objective: Analyze the yearly trend of EV sales to understand growth patterns over time.



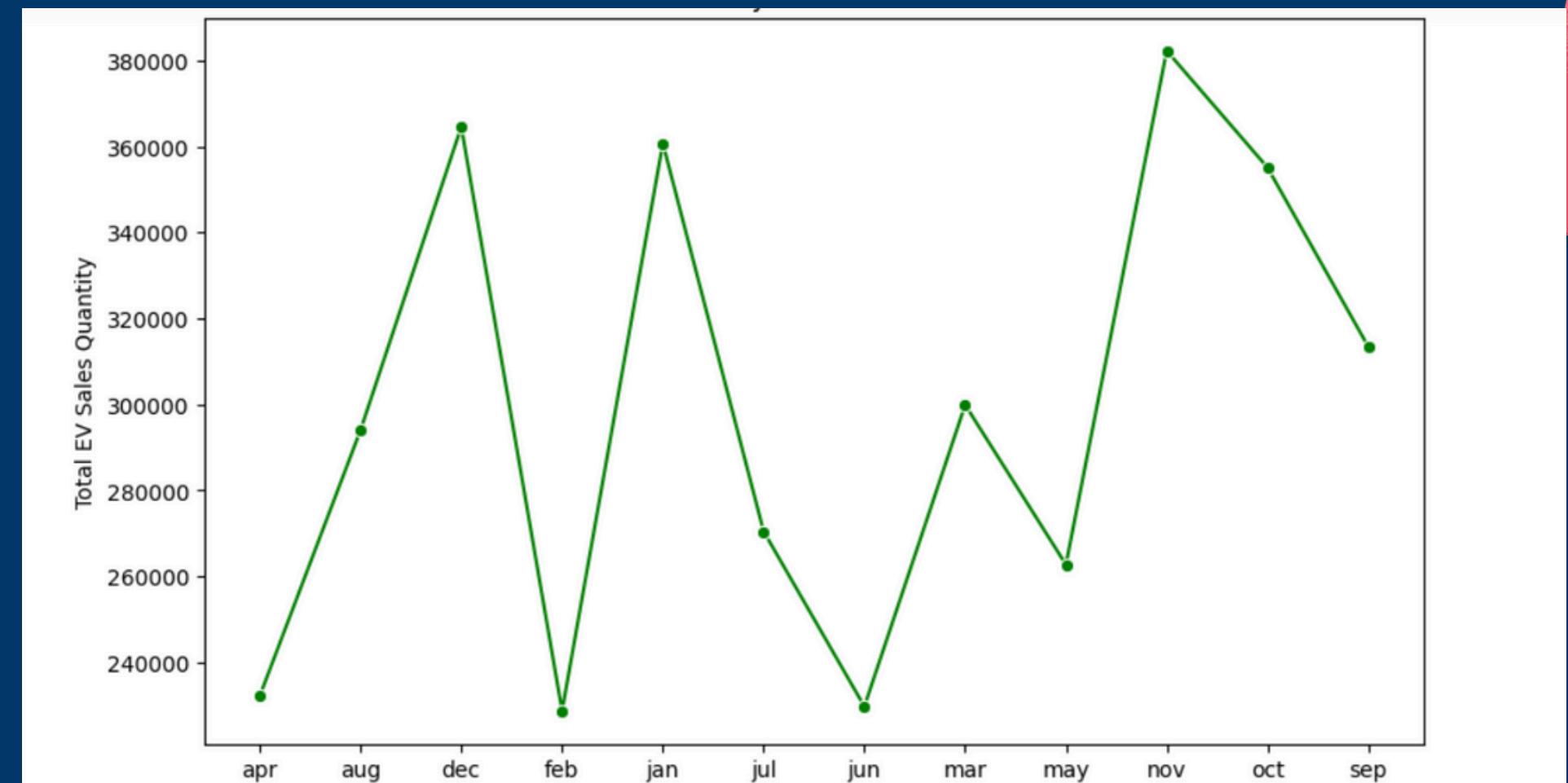
WORKFLOW



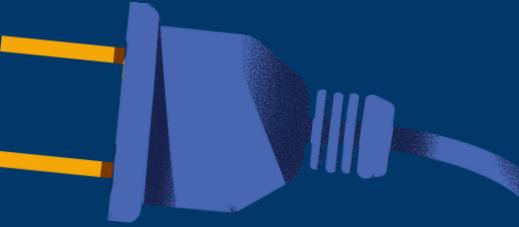
3 Exploratory Data Analysis (EDA)

1. Sales Trend by Month

Objective: Identify monthly variations in sales to uncover seasonal trends.



WORKFLOW



3 Exploratory Data Analysis (EDA)

Null Value Analysis

Objective: Ensure data quality by identifying and resolving missing values.

Checked for missing values using `isnull().sum()` in Pandas
Identified null values in the following columns:

State
Vehicle_Type
Month_Name



WORKFLOW



4 Model Building

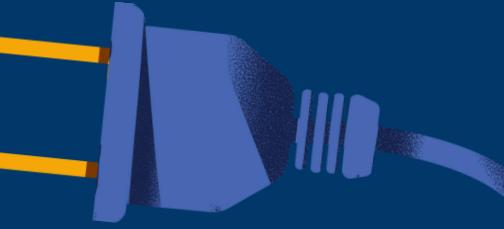
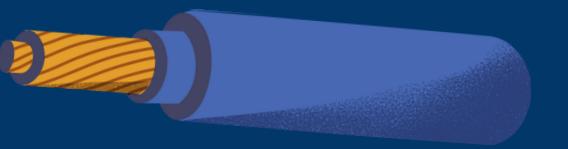
Features (X)

Year, Month_Name_encoded, State_encoded,
Vehicle_Type_encoded.

Target (y):

EV_Sales_Quantity.

WORKFLOW



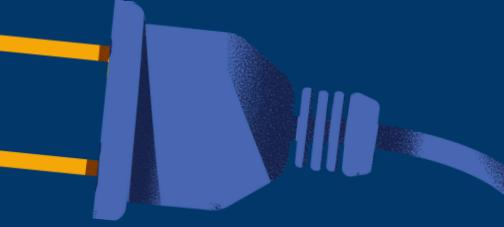
4 Model Building

Train-Test Split

Split the data into training (80%) and testing (20%) sets using `train_test_split`.



WORKFLOW



4 Model Building

Model Selection:

I have Use RandomForestRegressor for predictions due to its robustness for tabular data.



WORKFLOW



4 Model Building

why RandomForest model Choose

The Random Forest Regressor is a versatile and powerful ensemble learning method that combines multiple decision trees to improve predictive accuracy and control overfitting. It is especially effective for regression tasks where patterns in the data may be complex and nonlinear.



WORKFLOW



4 Model Building

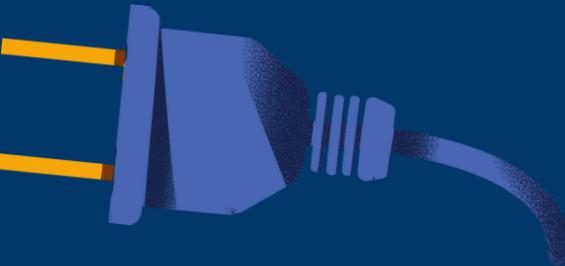
Model Training

Objective: Train a machine learning model to predict Electric Vehicle (EV) sales based on the provided features.

model.fit(X_train, y_train)



WORKFLOW



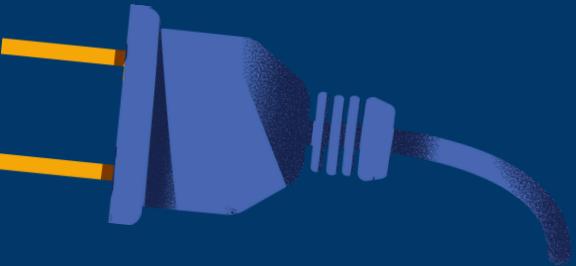
5 Model Evaluation Predictions

- The trained Random Forest Regressor model was used to predict EV sales for the test dataset (X_{test})
- The predictions were stored in y_{pred} .

Mean Absolute Error (MAE)

Mae was used to evaluate the accuracy of the model's predictions. The calculated MAE is 41.21, indicating that the model's predictions deviate by an average of 41.21 EV sales from the actual values.

WORKFLOW



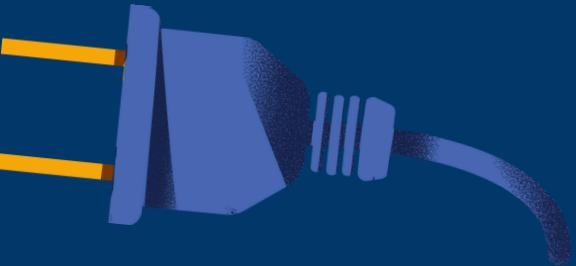
User Input and Prediction

Interactive Input:

Accept user inputs for
Year (e.g., 2025)
Month (e.g., 1–12)
State (e.g., Uttar Pradesh)
Vehicle Type (e.g., 2W_Personal)

Validate user inputs and encode them
using mappings.

WORKFLOW



User Input and Prediction

Predictive Output:

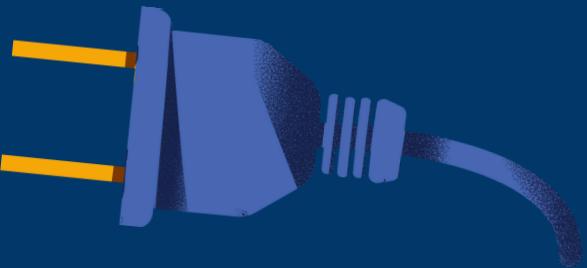
Use the trained model to predict EV_Sales_Quantity based on user inputs.

Display the predicted sales quantity.

SOURCE CODE

[https://drive.google.com/drive/folders/
1fbv-
aZE3hwOWhUJvfXovetRSvbtI75WO?
usp=sharing](https://drive.google.com/drive/folders/1fbv-aZE3hwOWhUJvfXovetRSvbtI75WO?usp=sharing)

ANALYSIS USING SQL



Total Sales

```
SELECT  
    SUM(EV_Sales_Quantity) AS Total_Sales  
FROM  
    newdemo.`electric vehicle`
```

	Total_Sales
▶	3593811

Annual EV Sales Growth by State

```
SELECT year,State,sum(EV_Sales_Quantity) ,lag(sum(EV_Sales_Quantity))over(partition by State order by Year)  
| as  
prev_year_sale,  
((sum(EV_Sales_Quantity) -lag(sum(EV_Sales_Quantity))over(partition by State order by Year) )  
/lag(sum(EV_Sales_Quantity))over(partition by State order by Year) )*100  
asAnnual_Sales_growth  
FROM newdemo.`electric vehicle`  
group by year,State
```

	year	State	Sales_Quantity	prev_year_sale	Annual_Sales_growth
▶	2014	Andaman & Nicobar Island	0	NULL	NULL
	2015	Andaman & Nicobar Island	0	0	NULL
	2016	Andaman & Nicobar Island	0	0	NULL
	2017	Andaman & Nicobar Island	0	0	NULL
	2018	Andaman & Nicobar Island	20	0	NULL
	2019	Andaman & Nicobar Island	2	20	-90
	2020	Andaman & Nicobar Island	36	2	1700
	2021	Andaman & Nicobar Island	92	36	155.55555555555557
	2022	Andaman & Nicobar Island	23	92	-75
	2023	Andaman & Nicobar Island	26	23	13.043478260869565
	2024	Andaman & Nicobar Island	3	26	-88.46153846153845
	2014	Andhra Pradesh	12	NULL	NULL
	2015	Andhra Pradesh	8	12	-33.33333333333333
	2016	Andhra Pradesh	20	8	150
	2017	Andhra Pradesh	0	20	100

Top 5 States with Maximum EV Sales for Each Vehicle Category

```
• | SELECT  
    state,  
    Vehicle_Category,  
    SUM(EV_Sales_Quantity) AS Total_Quantity_Sold  
FROM  
    newdemo.`electric vehicle`  
GROUP BY state , Vehicle_Category  
ORDER BY SUM(EV_Sales_Quantity) DESC  
LIMIT 5
```

	state	Vehide_Category	Total_Quantity_Sold
	Uttar Pradesh	3-Wheelers	641906
	Maharashtra	2-Wheelers	342236
	Karnataka	2-Wheelers	278671
	Bihar	3-Wheelers	184394
	Tamil Nadu	2-Wheelers	181735

Month-on-Month EV Sales Performance Comparison

```

SELECT *
FROM (
    SELECT
        Month_Name,
        SUM(EV_Sales_Quantity) AS total_sales,
        LAG(SUM(EV_Sales_Quantity)) OVER (ORDER BY FIELD(Month_Name, 'January', 'February', 'March', 'April', 'May', 'June', 'July',
        'August', 'September', 'October', 'November', 'December')) AS pre_month_sales,
        (SUM(EV_Sales_Quantity) - LAG(SUM(EV_Sales_Quantity)) OVER (ORDER BY FIELD(Month_Name, 'January', 'February', 'March',
        'April', 'May', 'June', 'July', 'August', 'September', 'October', 'November', 'December'))) /
        LAG(SUM(EV_Sales_Quantity)) OVER (ORDER BY FIELD(Month_Name, 'January', 'February', 'March', 'April', 'May', 'June',
        'July', 'August', 'September', 'October', 'November', 'December')) * 100 AS MOM
    FROM
        newdemo.`electric vehicle`
    GROUP BY
        Month_Name
) AS t1
ORDER BY FIELD(Month_Name, 'January', 'February', 'March', 'April', 'May', 'June', 'July', 'August', 'September', 'October', 'November', 'December');

```

	Month_Name	total_sales	pre_month_sales	MOM
jan	360703	NULL	NULL	
feb	228739	360703	-36.58522385452852	
mar	299888	228739	31.10488373211389	
apr	232194	299888	-22.573093955076562	
jun	229754	232194	-1.0508454137488479	
jul	270473	229754	17.72286880750716	
aug	294022	270473	8.706599179955116	
sep	313433	294022	6.601886933630817	
oct	355083	313433	13.288326372781423	
nov	382217	355083	7.641593655567854	
dec	364558	382217	-4.620150333449323	
may	262747	364558	-27.927243401598652	

Monthly EV Sales Distribution by Vehicle Type and Category

```
SELECT  
    Vehicle_Type,  
    Vehicle_Category,  
    Month_Name,  
    SUM(EV_Sales_Quantity) AS month_sales  
FROM  
    newdemo.`electric vehicle`  
GROUP BY Vehicle_Type , Vehicle_Category , Month_Name  
ORDER BY Vehicle_Type , Vehicle_Category , Month_Name
```

Vehide_Type	Vehide_Category	Month_Name	month_sales
2W_Personal	2-Wheelers	apr	129352
2W_Personal	2-Wheelers	aug	135837
2W_Personal	2-Wheelers	dec	176215
2W_Personal	2-Wheelers	feb	113260
2W_Personal	2-Wheelers	jan	185693
2W_Personal	2-Wheelers	jul	119905
2W_Personal	2-Wheelers	jun	99066
2W_Personal	2-Wheelers	mar	161671
2W_Personal	2-Wheelers	may	151651
2W_Personal	2-Wheelers	nov	203124
2W_Personal	2-Wheelers	oct	179585
2W_Personal	2-Wheelers	sep	140981
2W_Shared	2-Wheelers	apr	594
2W_Shared	2-Wheelers	aug	517
2W_Shared	2-Wheelers	dec	1527
2W_Shared	2-Wheelers	feb	1182
2W_Shared	2-Wheelers	jan	2099
2W_Shared	2-Wheelers	jul	352

Identify the Most Popular EV Vehicle Type by State

```
SELECT
    state, Vehicle_Type, SUM(EV_Sales_Quantity) AS total_sales
FROM
    newdemo.`electric vehicle`
GROUP BY State , Vehicle_Type
ORDER BY SUM(EV_Sales_Quantity) DESC
LIMIT 5
```

state	Vehide_Type	total_sales
Uttar Pradesh	3W_Shared_LowSpeed	596325
Maharashtra	2W_Personal	342231
Karnataka	2W_Personal	267178
Tamil Nadu	2W_Personal	181735
Bihar	3W_Shared_LowSpeed	174880

State wise vechicle type sale

```
SELECT  
    state, Vehicle_Type  
FROM  
    newdemo.`electric vehicle`  
GROUP BY State , Vehicle_Type  
ORDER BY SUM(EV_Sales_Quantity) DESC
```

state	Vehide_Type
Uttar Pradesh	3W_Shared_LowSpeed
Maharashtra	2W_Personal
Karnataka	2W_Personal
Tamil Nadu	2W_Personal
Bihar	3W_Shared_LowSpeed
Gujarat	2W_Personal
Assam	3W_Shared_LowSpeed
Rajasthan	2W_Personal
Delhi	3W_Shared_LowSpeed
Kerala	2W_Personal
Delhi	2W_Personal
Uttar Pradesh	2W_Personal
Odisha	2W_Personal
Madhya Prad...	2W_Personal
Andhra Prad...	2W_Personal
Rajasthan	3W_Shared_LowSpeed
Madhya Prad...	3W_Shared_LowSpeed
Chhattisgarh	2W_Personal
West Bengal	3W_Shared_LowSpeed
Haryana	3W_Shared_LowSpeed
Uttarakhand	3W_Shared_LowSpeed
Maharashtra	4W_Personal
Bihar	2W_Personal

Forecast EV Sales for the Next Year

```
SELECT year,sum(EV_Sales_Quantity)as year_wise_sales ,lead(sum(EV_Sales_Quantity))over(order by year ) as  
next_year_sales  
FROM newdemo.`electric vehicle` group by year|
```

year	year_wise_sales	next_year_sales
2014	2392	7805
2015	7805	49855
2016	49855	87420
2017	87420	130254
2018	130254	166819
2019	166819	124684
2020	124684	331498
2021	331498	1024723
2022	1024723	1525179
2023	1525179	143182
2024	143182	NULL

YOY Growth

```
SELECT year,sum(EV_Sales_Quantity)as current_sales ,lag(sum(EV_Sales_Quantity))over(order by year ) as prev_year_sales ,(sum(EV_Sales_Quantity)-lag(sum(EV_Sales_Quantity))over(order by year )) /lead(sum(EV_Sales_Quantity))over(order by year ) *100  
as YOY  
FROM newdemo.`electric vehicle` group by year
```

year	current_sales	prev_year_sales	YOY
2014	2392	NULL	NULL
2015	7805	2392	10.857486711463244
2016	49855	7805	48.101121024937086
2017	87420	49855	28.839805303483963
2018	130254	87420	25.676931284805686
2019	166819	130254	29.326136473003757
2020	124684	166819	-12.710483924488233
2021	331498	124684	20.1824297883428
2022	1024723	331498	45.452042022608495
2023	1525179	1024723	349.52438155634087
2024	143182	1525179	NULL