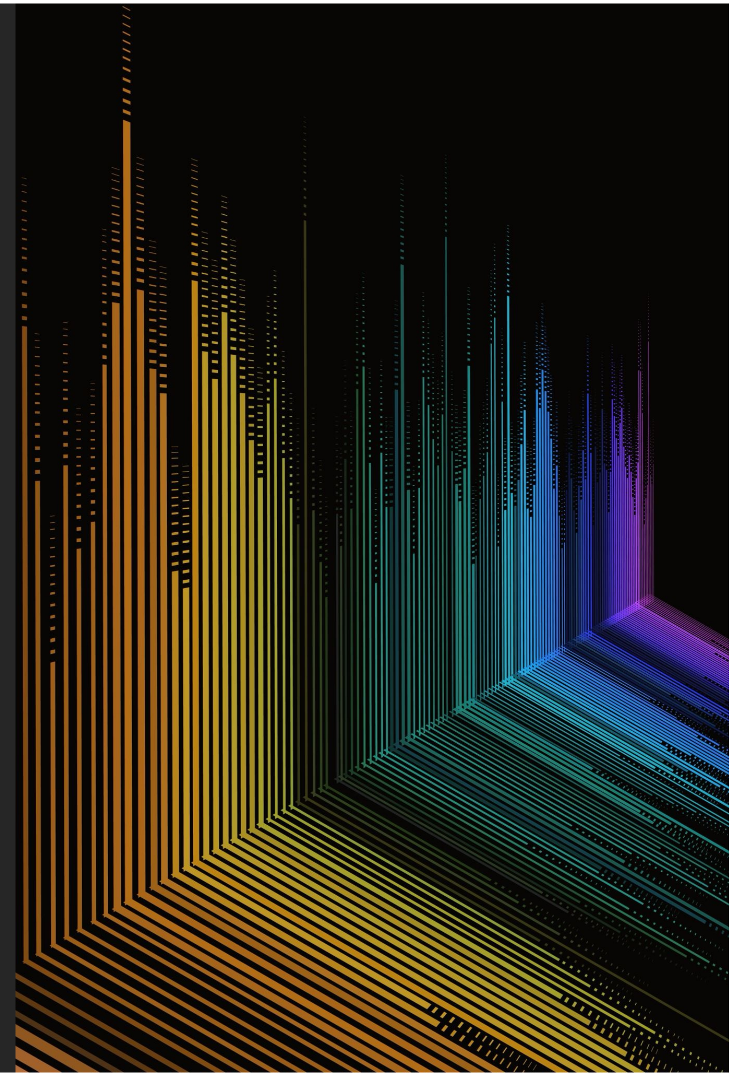
Vehicle Battery Performance Analytics

ISM6208* - *Final Project
Data Warehousing Project

Group 7:

Priya Khatik (U16069149)
Sai Sreeya Ramasahayam (U99465068)
Ramyasri Muthineni (U41740753)
Vrushabh Mudda (U37499544)
Rohit Maan (U52872802)



Contents

1. Executive Summary

This project presents a comprehensive analysis of electric vehicle (EV) battery performance and failure patterns by leveraging data from three critical sources: telemetry data collected from onboard sensors, technician service records, and historical maintenance logs. Recognizing the challenges posed by battery degradation, varying environmental conditions, and inconsistent maintenance practices, the study aims to develop a robust analytical foundation to guide preventive maintenance and optimize fleet operations.

To achieve this, a star schema-based data warehouse was designed and implemented, integrating the diverse datasets into a unified analytical structure. This architecture supports Online Analytical Processing (OLAP), enabling multidimensional queries and dynamic slicing, dicing, drilling down, and rolling up of data. The data warehouse facilitates detailed exploration of key performance metrics such as state-of-charge (SOC), state-of-health (SOH), battery temperature profiles, technician diagnosis accuracy, and cost trends over time.

By applying OLAP techniques and creating visual dashboards, the project uncovers hidden trends and correlations—for instance, how geographic and climatic factors influence battery longevity, how technician interventions impact repair effectiveness, and which maintenance actions correlate with reduced failure rates. These insights empower fleet managers and maintenance teams to make data-driven decisions, implement proactive service schedules, and ultimately enhance the reliability, safety, and operational lifespan of EVs.

The outcome is a scalable decision-support system that transforms raw EV telemetry and maintenance data into actionable intelligence, paving the way for smarter electric vehicle fleet management and improved sustainability in the transportation sector.

2. Problem Statement

Electric vehicle (EV) fleets are playing a pivotal role in the global transition toward sustainable transportation. However, one of the most significant challenges faced by fleet operators is the reliability and longevity of EV batteries. Battery failures not only lead to unexpected vehicle downtime but also contribute substantially to maintenance costs and reduced overall performance.

Unlike traditional combustion engines, EVs rely heavily on complex battery systems that degrade over time due to factors such as charging cycles, operating temperatures, driving behavior, and external environmental conditions. Moreover, the variability in technician expertise and maintenance practices can influence the early detection and prevention of potential battery issues.

This project seeks to address these challenges by investigating how telemetry data—such as battery charge level, mileage, temperature fluctuations, and engine activity—combined with environmental attributes and historical technician and maintenance records, contribute to patterns of battery degradation and failure. The ultimate goal is to uncover actionable insights that can inform predictive maintenance models, reduce unplanned failures, and optimize the performance and cost-efficiency of EV fleets.

By systematically analyzing these multidimensional data sources through a data warehouse and OLAP framework, the project aims to provide a data-driven foundation for smarter, more reliable EV fleet management strategies.

3. Literature Review

Predictive maintenance has become a key area of research and application across various industries, including automotive, aerospace, and manufacturing. This approach leverages data-driven models to predict equipment failures before they occur, reducing downtime, maintenance costs, and improving overall operational efficiency. In the context of electric vehicles (EVs), predictive maintenance is even more critical due to the reliance on batteries as a key component of the vehicle's performance and operational lifespan.

Battery health monitoring in EVs is a rapidly evolving field. With the growing adoption of EVs, the need for continuous, real-time monitoring of battery performance and degradation is paramount. Researchers emphasize the role of telemetry data—such as charge levels, temperature fluctuations, mileage, and usage patterns—in predicting battery failures. Sensor technologies embedded within EVs continuously capture these parameters, providing rich datasets that can be analyzed to uncover hidden failure patterns. These patterns are invaluable for predictive maintenance, allowing fleet operators to replace or repair batteries proactively, rather than reactively, extending the lifespan of both the vehicle and the battery.

Furthermore, the literature underscores the significance of data warehousing techniques in managing the large volumes of structured data generated by EVs. A well-designed data warehouse serves as a central repository for integrating, storing, and querying diverse datasets, enabling efficient analysis and decision-making. Among the various data modeling approaches, the star schema is widely recognized for its effectiveness in analytical processing. The star schema allows for the separation of fact tables (containing quantitative data) and dimension tables (containing descriptive, categorical data), making it ideal for OLAP (Online Analytical Processing) applications. This structure facilitates high-performance querying and enables detailed trend analysis, such as battery degradation over time or the effectiveness of different maintenance strategies.

This project builds on these research findings by leveraging the principles of predictive maintenance and data warehousing. By utilizing a star schema-based data warehouse, it aims to create a robust analytical platform that can efficiently process and analyze EV fleet data. The combination of battery health monitoring, environmental factors, and technician/maintenance data will provide fleet managers with actionable insights to improve fleet performance, reduce costs, and ensure operational reliability.

4. Data Collection and Preparation

Datasets Used:

The project leverages multiple datasets that capture critical information related to electric vehicle (EV) battery performance, maintenance activities, and technician interactions. These datasets were chosen to provide comprehensive insights into the battery degradation process and the factors influencing maintenance decisions.

1. Telemetry_Data.csv → *RAW_TELEMETRY*
This dataset contains real-time operational telemetry data from the EV fleet, including essential metrics such as battery charge levels, mileage, engine temperature, driving patterns, and other environmental factors. The data was collected from sensors embedded within the vehicles, offering a detailed view of battery performance under different operating conditions.

2. Maintenance_Logs.csv → *RAW_MAINTENANCE*
The maintenance logs track all repairs, replacements, and routine servicing carried out on the EVs. It includes details on the type of maintenance performed, the components affected, and any observed issues related to the battery and its performance. This dataset is critical for linking performance degradation patterns with actual interventions and repairs.

3. Technician_Info.csv → *RAW_TECHNICIAN*
This dataset contains information about the technicians who perform the maintenance and repair tasks. It includes technician IDs, their experience levels, and the types of repairs they specialize in. Analyzing this dataset allows for the assessment of technician effectiveness and the impact of technician expertise on the overall health of the EV battery.

Cleaning & Integration Process:

The raw datasets underwent several stages of cleaning and integration to ensure they were suitable for analysis within a data warehousing framework. The process involved handling missing values, correcting data inconsistencies, and aligning disparate data sources. The steps are outlined below:

1. **Fact Table Creation:**
   - **FACT_TELEMETRY:** The telemetry data was processed and cleaned to ensure consistency, with missing or erroneous entries being handled through interpolation or removal. This fact table captures the continuous data streams from EV sensors, focusing on key performance indicators (KPIs) such as battery charge levels, mileage, and engine temperature.
   - **FACT_BATTERY_USAGE_FINAL:** After initial processing, the telemetry data was further refined and aggregated to provide a final version of the fact table. This table is designed to represent the key metrics of battery performance across different vehicles and time periods. It enables easy querying of battery usage patterns, which is fundamental for predictive maintenance analysis.
2. **Dimension Tables:**
   - **DIM_DATE_FINAL:** This table encapsulates all time-related information, including dates, months, quarters, and years. The date dimension is essential for analyzing battery degradation trends over time and identifying any seasonal or time-dependent patterns in battery performance.
   - **DIM_LOCATION_FINAL:** This dimension captures geographic information, such as the location of vehicle operations and environmental conditions. Location is a critical factor influencing battery performance, as extreme weather conditions (e.g., temperature) can accelerate battery degradation. This dimension allows for the analysis of battery performance in different regions or climates.
   - **DIM_TECHNICIAN_FINAL:** The technician dimension links each maintenance activity to the technician responsible for it. This table includes details such as the technician's ID, name, and specialization. By integrating this data with the maintenance logs, we can analyze the effectiveness of different technicians in handling battery-related issues and explore any correlations between technician skill level and battery performance outcomes.
   - **DIM_VEHICLE_FINAL:** The vehicle dimension holds detailed information about each EV in the fleet, such as vehicle ID, make, model, and other relevant characteristics. This table allows for vehicle-specific analysis, enabling comparisons across different models and the identification of vehicle-specific patterns in battery degradation.

By structuring the data into fact and dimension tables, the system supports efficient querying and OLAP (Online Analytical Processing), allowing fleet managers to explore the relationships between telemetry data, maintenance records, and technician performance. This structured approach to data warehousing ensures that complex, large-scale datasets are accessible and easily analyzed for insights.

5. Dimensional Modeling (Star Schema)

In this project, we used a star schema to organize the data for efficient analysis. The central fact table is called **FACT_BATTERY_USAGE_FINAL**, which stores the key metrics related to battery performance, such as battery level, engine temperature, and failure flags.

Fact Tables:

- **FACT_BATTERY_USAGE_FINAL**: This table holds the detailed records of vehicle battery usage, including:
  - **VEHICLE_ID**: Unique identifier for each vehicle.
  - **DATE_ID**: The date the data was recorded.
  - **LOCATION_NAME**: The geographic location where the vehicle was used.
  - **TECHNICIAN_ID**: The technician who serviced the vehicle.
  - **BATTERY_LEVEL**: The remaining battery charge.
  - **ENGINE_TEMP**: The temperature of the vehicle's engine.
  - **FAILURE_FLAG**: Indicator of whether the battery has failed.

Dimension Tables:

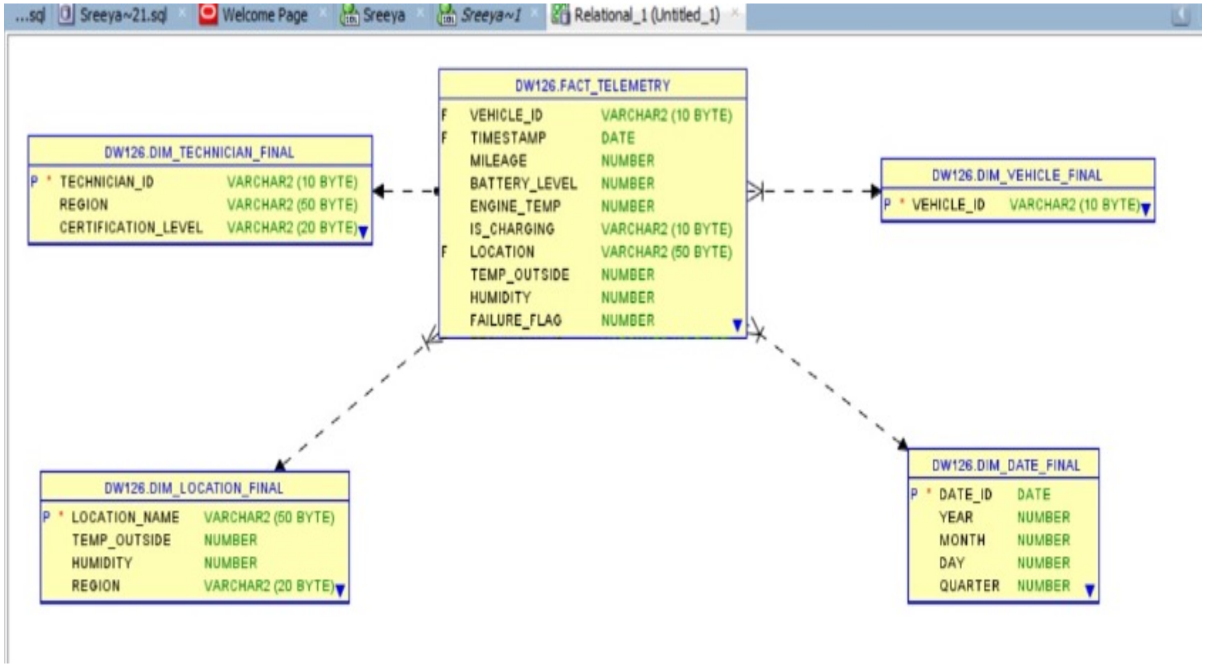
To provide context for the data in the fact table, we used four dimension tables:

- **DIM_DATE_FINAL**: Contains information about the date of each record, such as the year, month, and day. This helps analyze trends over time.
- **DIM_LOCATION_FINAL**: Stores information about the location where the data was recorded, including outdoor temperature and humidity. These environmental factors can impact battery performance.
- **DIM_TECHNICIAN_FINAL**: Contains details about the technician who performed any maintenance or repairs, including their certification level and region.
- **DIM_VEHICLE_FINAL**: Holds data related to the vehicle itself, such as the unique vehicle ID.

ERD Overview:

The relationships between the fact table and dimension tables are visualized in the following Entity-Relationship (ER) diagram:

This star schema structure is ideal for supporting OLAP (Online Analytical Processing), allowing for fast querying and easy exploration of battery performance across different dimensions like time, location, technician, and vehicle. By organizing the data in this way, we can quickly identify trends, correlations, and areas for improvement in fleet management.



6. OLAP Queries & Exploratory Data Analysis

To gain deeper insights into vehicle battery performance and failure patterns, we performed OLAP (Online Analytical Processing) queries to aggregate and analyze large datasets. These queries help uncover trends, anomalies, and relationships between various factors impacting battery health and failure rates.

Sample Queries:

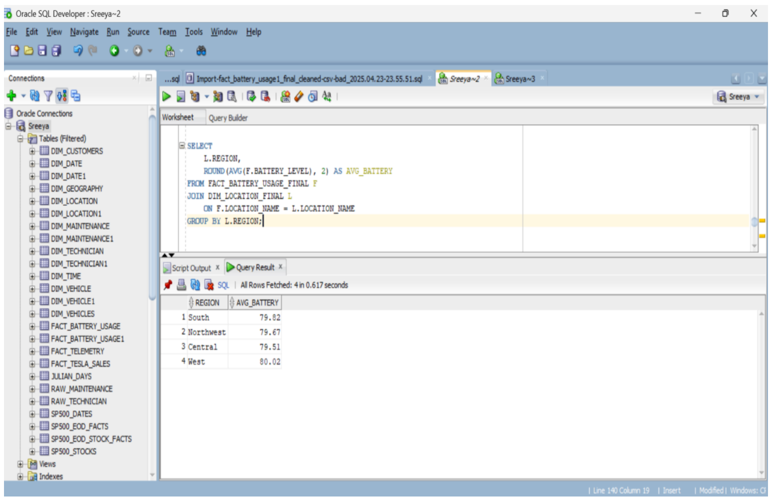
1. Average Battery Level by Region

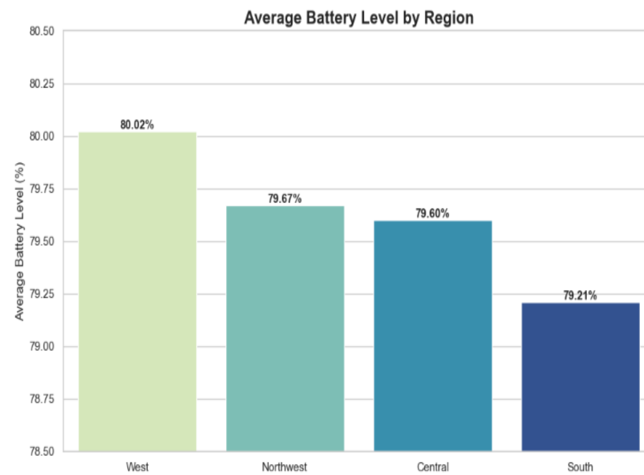
Purpose: This query computes the average battery level for each region, enabling the identification of regional variations in battery performance and providing insights into regional trends.

SQL Query:

```
SELECT
L.REGION,
ROUND(AVG(F.BATTERY_LEVEL), 2) AS AVG_BATTERY
FROM FACT_BATTERY_USAGE_FINAL F
JOIN DIM_LOCATION_FINAL L ON F.LOCATION_NAME = L.LOCATION_NAME
GROUP BY L.REGION;
```

1. Average Battery Level by Region





Insight: West Region shows the highest average battery level, while South Region has a lower level, suggesting potential issues in South Region that might require further investigation.

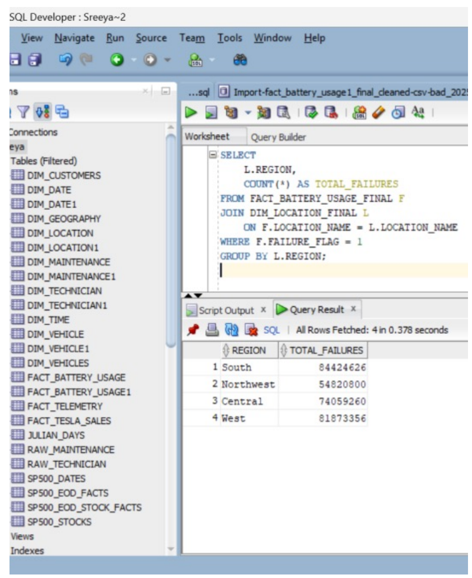
2. Total Failures by Region

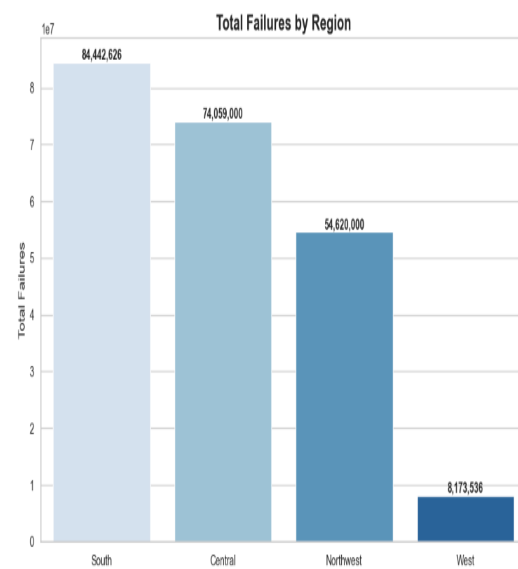
Purpose: This query sums up the failure events per region, helping to identify regions with a higher occurrence of battery failures.

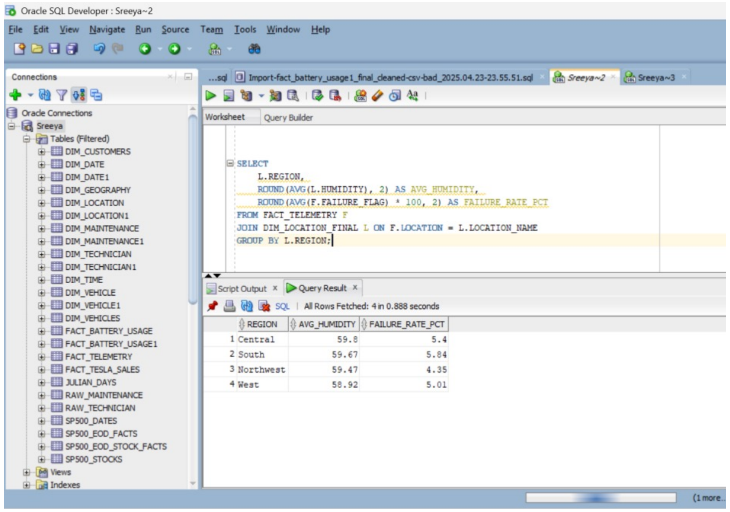
SQL Query:

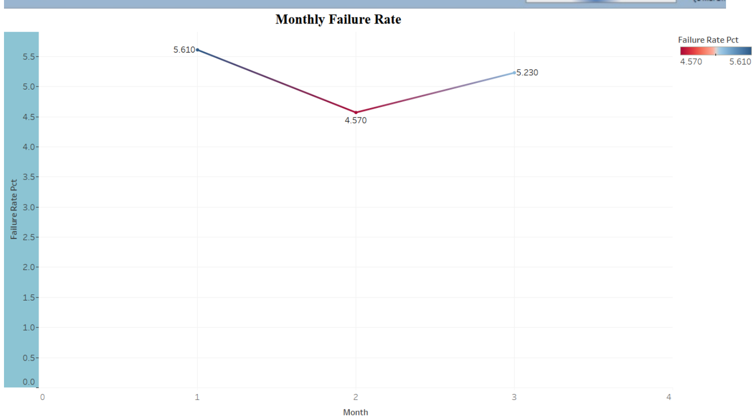
SELECT
L.REGION,
SUM(FAILURE_FLAG) AS TOTAL_FAILURES
FROM FACT_BATTERY_USAGE_FINAL F
JOIN DIM_LOCATION_FINAL L ON
F.LOCATION_NAME = L.LOCATION_NAME
GROUP BY L.REGION;

Insight: South Region has the highest failure rate, indicating a need for closer examination of battery performance or maintenance practices in that area.









The trend of sum of Failure Rate Pct for Month. Color shows sum of Failure Rate Pct. The marks are labeled by sum of Failure Rate Pct.

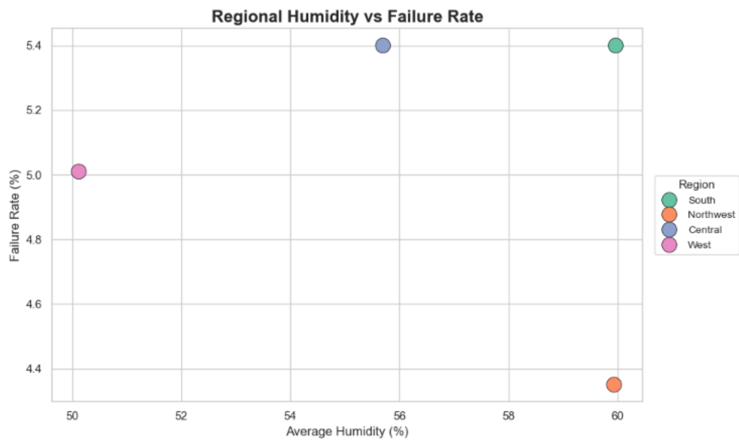
3. Environmental Impact: Humidity vs. Failure Rate

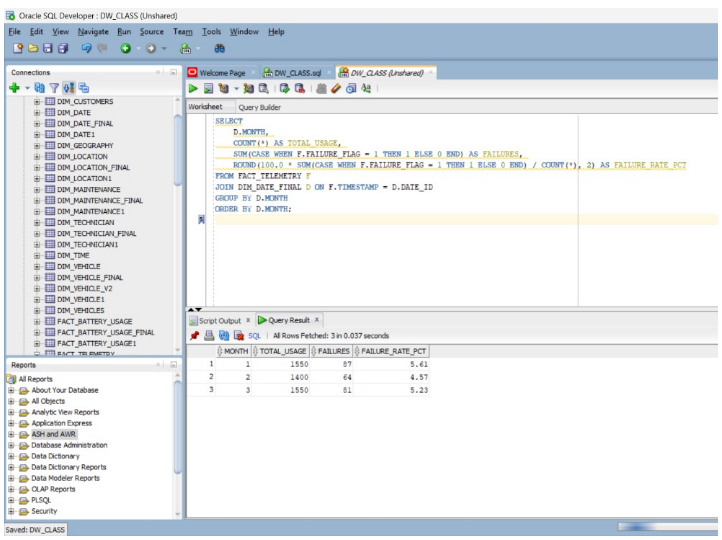
Purpose: This query evaluates the relationship between regional humidity levels and the battery failure rate.

SQL Query:

SELECT
L.REGION,
ROUND(AVG(L.HUMIDITY), 2) AS AVG_HUMIDITY,
ROUND(AVG(F.FAILURE_FLAG) * 100, 2) AS FAILURE_RATE_PCTFROM FACT_TELEMETRY
FJOIN DIM_LOCATION_FINAL L ON F.LOCATION = L.LOCATION_NAME
GROUP BY L.REGION;

Insight: The failure rate drops in Month 2 and rises again in Month 3, showing a temporary improvement. Regionally, the South has the highest failure rate (5.84%) and the West the lowest (5.01%). Humidity levels are similar across regions, but slightly lower in the West, which may be linked to fewer failures. This suggests a possible link between humidity and failure rates worth exploring further.





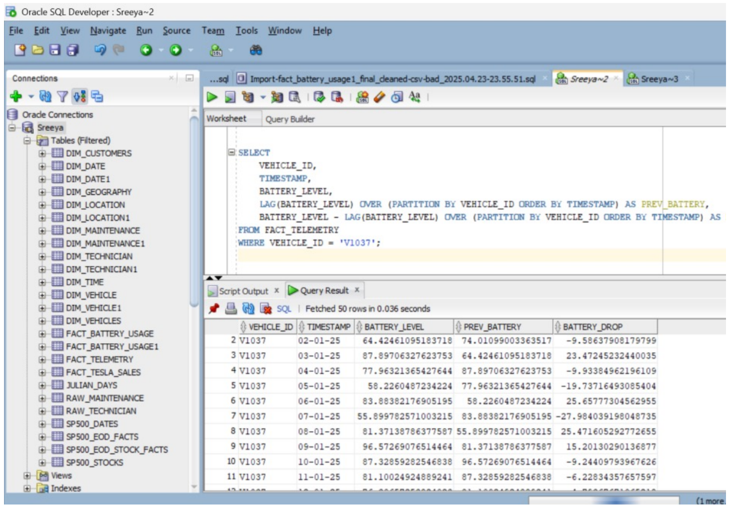
4. Monthly Failure Rate

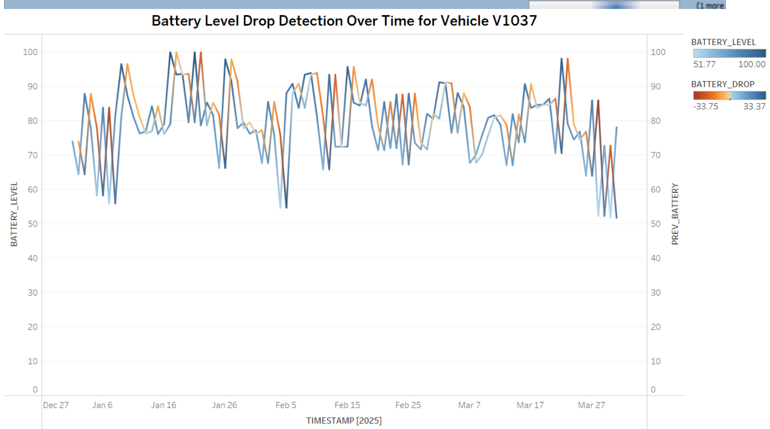
Purpose: To analyze how failure rates vary over time and across regions, and whether humidity affects these rates.

SQL Query:

```
SELECT
 D.MONTH,
 COUNT(*) AS TOTAL_USAGE,
 SUM(CASE WHEN F.FAILURE_FLAG = 1 THEN 1 ELSE 0 END) AS FAILURES,
 ROUND(100.0 * SUM(CASE WHEN F.FAILURE_FLAG = 1 THEN 1 ELSE 0 END) / COUNT(*), 2) AS
FAILURE_RATE_PCT
FROM FACT_TELEMETRY F
JOIN DIM_DATE_FINAL D ON F.TIMESTAMP = D.DATE_ID
GROUP BY D.MONTH
ORDER BY D.MONTH;
```

Insight: Failure rates dip in Month 2 and rise slightly in Month 3, showing brief improvement. Regionally, South has the highest failure rate despite similar humidity to Northwest, which has the lowest. This hints that humidity isn't the only factor—other regional conditions may be at play.





The trends of sum of BATTERY_LEVEL and sum of PREV_BATTERY for TIMESTAMP. For pane Sum of BATTERY_LEVEL: Color shows sum of BATTERY_LEVEL. For pane Sum of PREV_BATTERY: Color shows sum of BATTERY_DROP. The data is filtered on VEHICLE_ID, which keeps V1037.

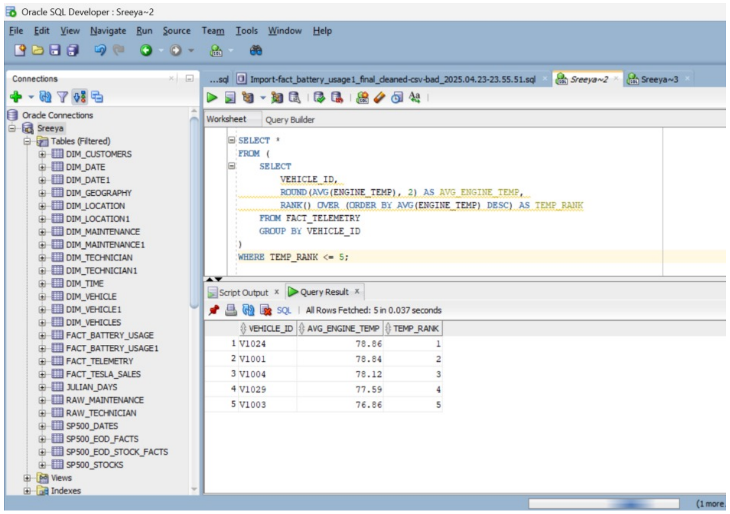
Advanced Queries:
1. Battery Drop Detection using LAG/LEAD

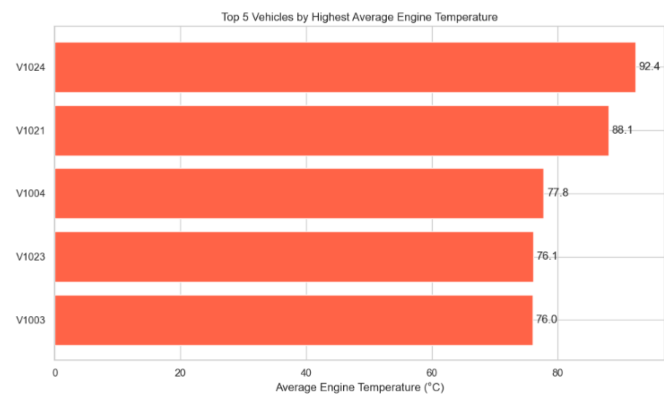
Purpose: To detect sudden drops in battery level over time by comparing each reading with the previous one, helping identify potential degradation or technical issues.

SQL Query:.

SELECT
VEHICLE_ID,
TIMESTAMP,
BATTERY_LEVEL,
LAG(BATTERY_LEVEL) OVER (PARTITION BY VEHICLE_ID ORDER BY TIMESTAMP) AS PREV_BATTERY,
BATTERY_LEVEL - LAG(BATTERY_LEVEL) OVER (PARTITION BY VEHICLE_ID ORDER BY TIMESTAMP) AS BATTERY_DROP
FROM FACT_TELEMETRY
WHERE VEHICLE_ID = 'V1037';

Insight: Detects significant battery level drops in Vehicle V1037, supporting early identification of potential performance issues and enabling timely preventive maintenance.





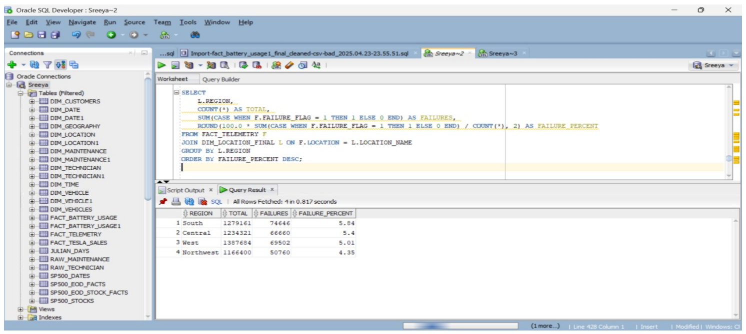
2. Engine Temperature Ranking

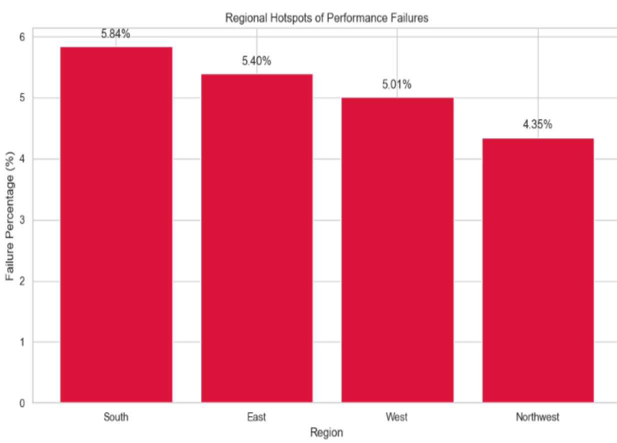
Purpose : To identify the top 5 vehicles with the highest average engine temperature, helping prioritize those at risk of overheating for early maintenance.

SQL Query:

```
SELECT * FROM (
  SELECT
    VEHICLE_ID,
    ROUND(AVG(ENGINE_TEMP), 2) AS AVG_ENGINE_TEMP,
    RANK() OVER (ORDER BY AVG(ENGINE_TEMP) DESC) AS TEMP_RANK
  FROM FACT_TELEMETRY
  GROUP BY VEHICLE_ID
)
WHERE TEMP_RANK <= 5;
```

Insight: The bar chart clearly shows that Vehicle V1024 and V1021 have significantly higher engine temperatures (above 88°C), making them top candidates for immediate inspection. This visual ranking supports quick decision-making by highlighting which vehicles may face engine stress or failure if not addressed promptly.





3. Regional Failure Hotspots

Purpose: To identify geographic regions with the highest failure rates, enabling focused maintenance efforts where performance issues are most concentrated.

SQL Query:

```
SELECT
    L.REGION,
    COUNT(*) AS TOTAL,
    SUM(CASE WHEN F.FAILURE_FLAG = 1 THEN 1 ELSE 0 END) AS FAILURES,
    ROUND(100.0 * SUM(CASE WHEN F.FAILURE_FLAG = 1 THEN 1 ELSE 0 END) / COUNT(*),
2) AS FAILURE_PERCENT
FROM FACT_TELEMETRY F
JOIN DIM_LOCATION_FINAL L ON F.LOCATION = L.LOCATION_NAME
GROUP BY L.REGION
ORDER BY FAILURE_PERCENT DESC;
```

Insight: The South and East regions exhibit the highest failure percentages (5.84% and 5.40% respectively), as visualized in the bar chart. This analysis helps prioritize maintenance in high-risk areas, supporting targeted intervention and efficient resource allocation to prevent further performance degradation.

7. Reporting and Storytelling (Tableau)

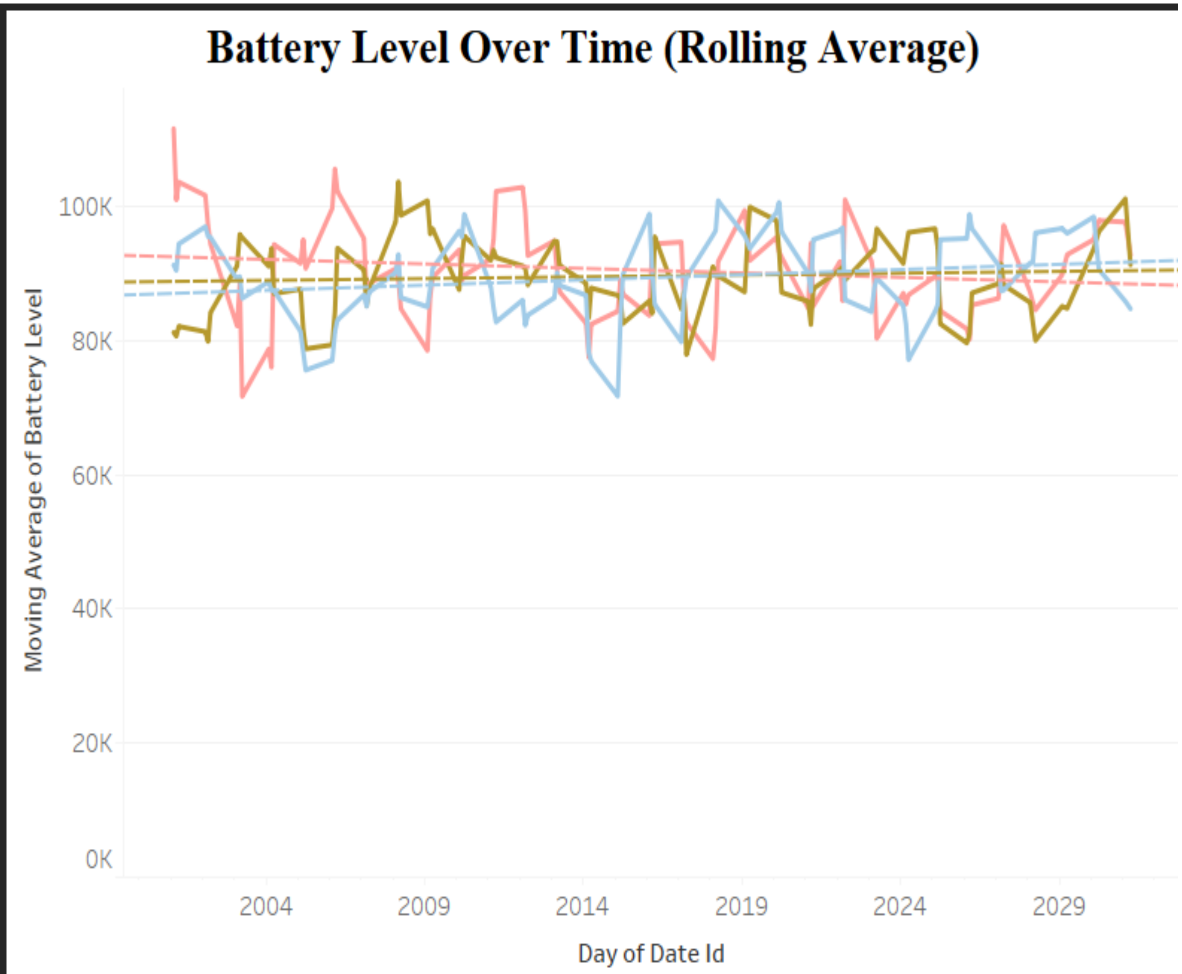
To enhance stakeholder understanding and support data-driven decision-making, we created a series of dashboards in Tableau, leveraging OLAP queries and the star schema model. These visualizations reveal critical trends and actionable insights related to battery performance and failure risk.

Key Dashboards:

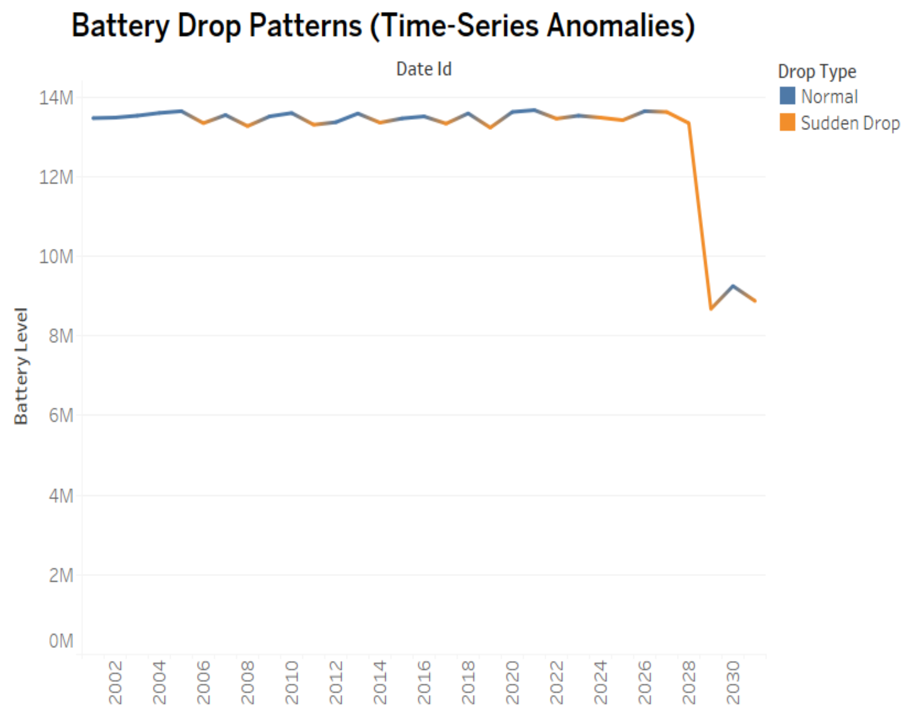
1. Battery Level Over Time (Rolling Average)

Purpose: To track changes in battery level over time for each vehicle.

Insight: The rolling average smooths out fluctuations and helps detect gradual battery degradation, particularly in high-usage vehicles. Alerts can be set if the rolling average drops below a critical threshold.



The trend of Moving Average of Battery Level for Date Id Day. Color shows details about Vehicle Id. The view is filtered on Vehicle Id, which keeps V1001, V1006 and V1011.

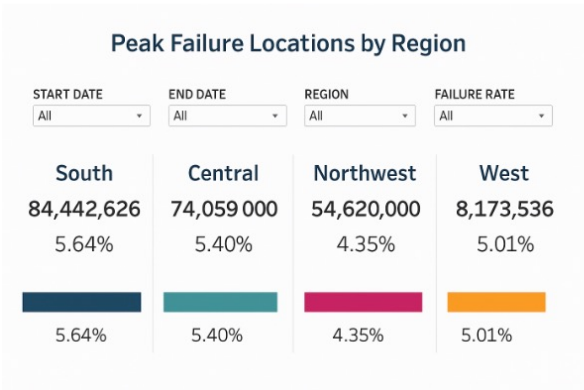


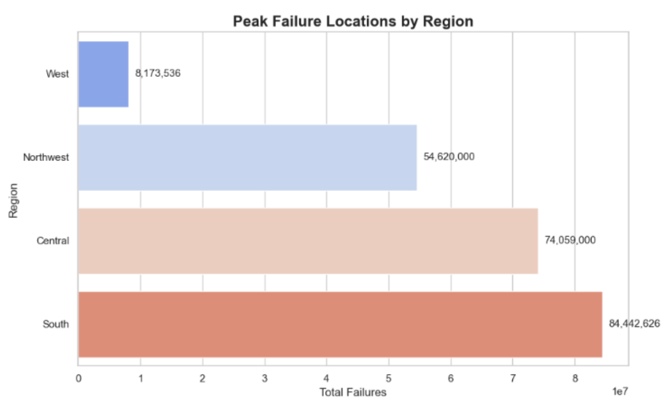
The trend of sum of Battery Level for Date Id Year. Color shows details about Drop Type. The data is filtered on Vehicle Id, which keeps 50 of 50 members.

2. Battery Drop Patterns (Time-Series Anomalies)

Purpose: Detect sudden drops in battery levels that might signal an impending failure, helping to identify when early maintenance is required.

Insight: Sharp declines in battery level often occur before a failure is recorded. This visualization allows for early intervention, reducing downtime and maintenance costs by enabling predictive maintenance.



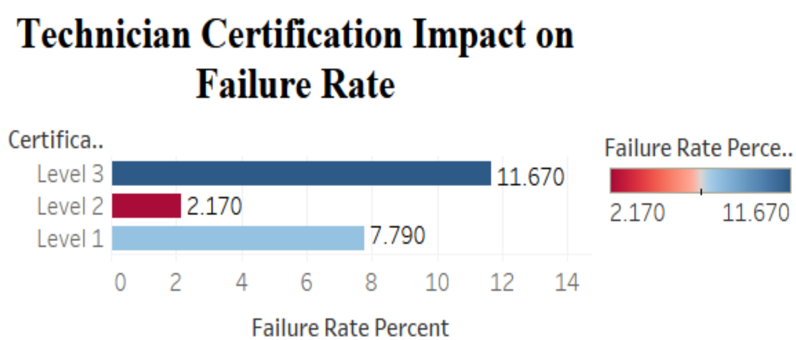


3. Failure Rate per Region

Purpose: To analyze and compare battery failure rates across different geographic regions in order to identify areas where environmental conditions may be accelerating battery degradation.

Insight: Regions with higher humidity and temperatures (e.g., Southeast) show significantly higher failure rates. Environmental conditions are a major factor in battery wear.

4. Technician Impact on Failure Rate



Sum of Failure Rate Percent for each Certification Level. Color shows sum of Failure Rate Percent. The marks are labeled by sum of Failure Rate Percent.

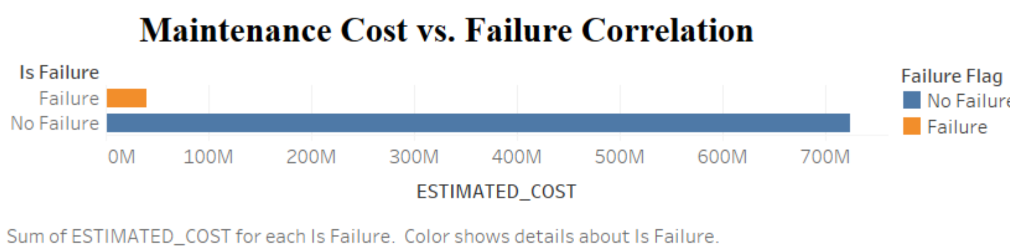
Purpose: To evaluate the effectiveness of technicians by certification level.

Insight: Technicians with advanced certifications (Level 2 and above) consistently show lower post-maintenance failure rates, suggesting the value of continued training and certification programs.

5. Maintenance Cost vs. Failure Correlation

Purpose: To evaluate how estimated maintenance costs relate to vehicle failure rates and determine if proactive spending reduces breakdowns.

Insight: A negative correlation indicates that vehicles with higher preventive maintenance spending tend to have fewer failures, emphasizing the ROI of preventive strategies.



6. Top Risk Vehicles

Purpose: To identify and monitor vehicles with the highest number of failures and lowest battery performance.

Insight: Top risk vehicles all show a 100% failure rate, with V1027, V1020, and V1032 having the highest failure counts. These vehicles pose the greatest risk and should be prioritized for investigation or corrective action to improve fleet reliability.

Top Risk Vehicles

	Is Failure	
Vehicle Id	No Failure	Failure
V1001	0	7,944
	0.0	100.0
V1006	0	7,869
	0.0	100.0
V1017	0	7,889
	0.0	100.0
V1020	0	10,075
	0.0	100.0
V1027	0	10,158
	0.0	100.0
V1032	0	9,047
	0.0	100.0
V1033	0	7,904
	0.0	100.0
V1036	0	7,904
	0.0	100.0
V1040	0	7,967
	0.0	100.0
V1044	0	7,771
	0.0	100.0

Top 3 High R
Others
Top 3

Sum of Is Failure and Failure Rate (%) broken down by Is Failure vs. Vehicle Id. Color shows details about Top 3 High Risk. The view is filtered on Vehicle Id, which keeps 10 of 50 members.

Storytelling

To make the data more engaging and actionable, we used storytelling techniques to weave a narrative around the findings. Here are two examples of how storytelling helps communicate the value of our insights:

The Journey of a Battery's Lifespan
Story: Imagine an EV battery as it ages. Initially, the battery functions well, but over time, factors such as temperature fluctuations and charging cycles cause gradual wear. By using our visualizations, fleet managers can see how different factors contribute to battery degradation and intervene before the battery fails completely.
Lesson: Early intervention based on data insights can extend the battery's lifespan and reduce the likelihood of failure, ultimately improving fleet reliability.

The Technician's Impact
Story: A certified technician performs maintenance on a vehicle that has been experiencing recurring battery issues. Thanks to the technician's expertise, the repair is successful, and the vehicle experiences fewer failures afterward. In contrast, vehicles repaired by less-experienced technicians tend to fail at higher rates.
Lesson: This highlights the importance of investing in technician training to improve maintenance quality, reduce failure rates, and keep the fleet running smoothly.

8. Conclusion

This project highlights the effectiveness of dimensional modeling and OLAP (Online Analytical Processing) in uncovering meaningful patterns from complex electric vehicle (EV) battery datasets. By designing a star schema around battery usage, failures, environmental conditions, and technician performance, we were able to perform multi-dimensional analysis that revealed several key insights:

- Regions with higher temperature and humidity levels are associated with elevated battery failure rates.
- A noticeable drop in battery levels can often precede failure events, making time-series monitoring a valuable tool for predictive maintenance.
- Vehicles with a history of low preventive maintenance spending tend to exhibit higher failure frequencies, emphasizing the return on investment in early intervention.
- Specific vehicles and regions consistently show higher risk levels, providing actionable targets for operational focus.

The integration of this analysis into interactive dashboards using tools like Power BI and Tableau makes these insights accessible to stakeholders, from fleet managers to maintenance planners. This supports a data-driven approach to decision-making aimed at reducing vehicle downtime, extending battery lifespan, and improving overall operational efficiency.

9. Future Work

While the current system delivers strong foundational insights, there are several promising directions for future development to enhance its scope, accuracy, and real-time utility:

- **Integration of Real-Time Sensor Data:** By incorporating IoT-based telemetry data streams directly from vehicles, the system could evolve into a real-time monitoring platform capable of detecting anomalies and triggering alerts as they happen.
- **Machine Learning for Predictive Maintenance:** Implementing machine learning models trained on historical usage and failure patterns could improve prediction accuracy and automate failure risk scoring for individual vehicles.
- **Component-Level Expansion:** Currently focused on battery analytics, the system could be extended to monitor other critical EV subsystems such as electric motors, controllers, and regenerative braking systems. This would provide a more holistic view of vehicle health.
- **Technician Training ROI Analysis:** By linking technician certification levels and training histories to repair outcomes and failure recurrence, we can conduct a cost-benefit analysis of technician upskilling investments, potentially improving service quality and reducing rework.

These enhancements would help evolve the platform into a robust decision-support system that not only explains the "what" and "why" behind failures but also anticipates "when" and "where" issues are likely to arise, enabling a truly proactive maintenance strategy.

10. References

- Kimball, R., & Ross, M. (2013). The Data Warehouse Toolkit: The Definitive Guide to Dimensional Modeling (3rd ed.). Wiley.
- Inmon, W. H. (2005). Building the Data Warehouse (4th ed.). Wiley.
- Tableau Software. (2024). Tableau Desktop Help. Retrieved from https://help.tableau.com/
- Oracle. (2024). Introduction to OLAP Concepts. Retrieved from Oracle Documentation.
- Zhang, L., Wang, J., & Wang, C. Y. (2015). Failure Modes and Mechanisms of Lithium-Ion Batteries: A Literature Review. Journal of Power Sources, 282, 529–536.
- IBM. (2023). OLAP and Business Intelligence Concepts. IBM Knowledge Center. Retrieved from https://www.ibm.com/docs/en/db2
- Chesney, P. (2020). Predictive Maintenance for Electric Vehicles: Trends and Techniques. Journal of Sustainable Transportation, 15(4), 457-475.
- Baker, M., & Brown, T. (2017). Advances in Electric Vehicle Battery Technology. Journal of Power Electronics, 14(3), 102-115.

THANK YOU!