A.Dhanya

Electric Vehicles Data Analysis Report

1. Dataset Description

1.1 Source:

The dataset used for this project is Electric_Vehicles.csv containing approximately 10,000 records. It captures technical, economic, and environmental attributes of Battery Electric Vehicles (BEVs) and Plug-in Hybrid Electric Vehicles (PHEVs) from various manufacturers worldwide.

1.2 Columns:

- Manufacturer
- Model
- Year
- Battery_Type
- Battery_Capacity_kWh
- Range_km
- Charging_Type
- Charge_Time_hr
- Price_USD
- Country_of_Manufacture
- Safety_Rating
- CO■_Emissions_g_per_km
- Units_Sold_2024
- Warranty_Years

1.3 Data Quality:

- The dataset contains some missing values in CO■_Emissions and Autonomous_Level.
- Mixed units (kWh, km, hr, USD) were standardized.
- Categorical values like Charging_Type and Battery_Type were cleaned andnormalized.

• Dataset verified to be consistent and suitable for visualization and aggregation.

2. Operations Performed

2.1 Data Loading and Inspection

• Loaded the dataset into PySpark and Pandas environments for distributed analysis.

• Verified schema, data types, and completeness of records.

• Calculated descriptive statistics such as mean, max, and min for numeric columns.

2.2 Aggregations and Visualizations

• Grouped data by Manufacturer, Year, and Battery_Type to identify leading EVproducers.

• Created visualizations (bar, scatter, and correlation plots) to interpret relationships.

• Conducted range vs price analysis to determine performance-to-cost efficiency.

• Visualized top manufacturers by units sold and battery capacity trends.

3. Key Insights

3.1 Overall Market Overview

• Total EVs analyzed: ~10,000 entries across 50+ manufacturers.

• Average battery capacity: 65.4 kWh.

• Average driving range: 475 km per charge.

• Average safety rating: 4.4 / 5.

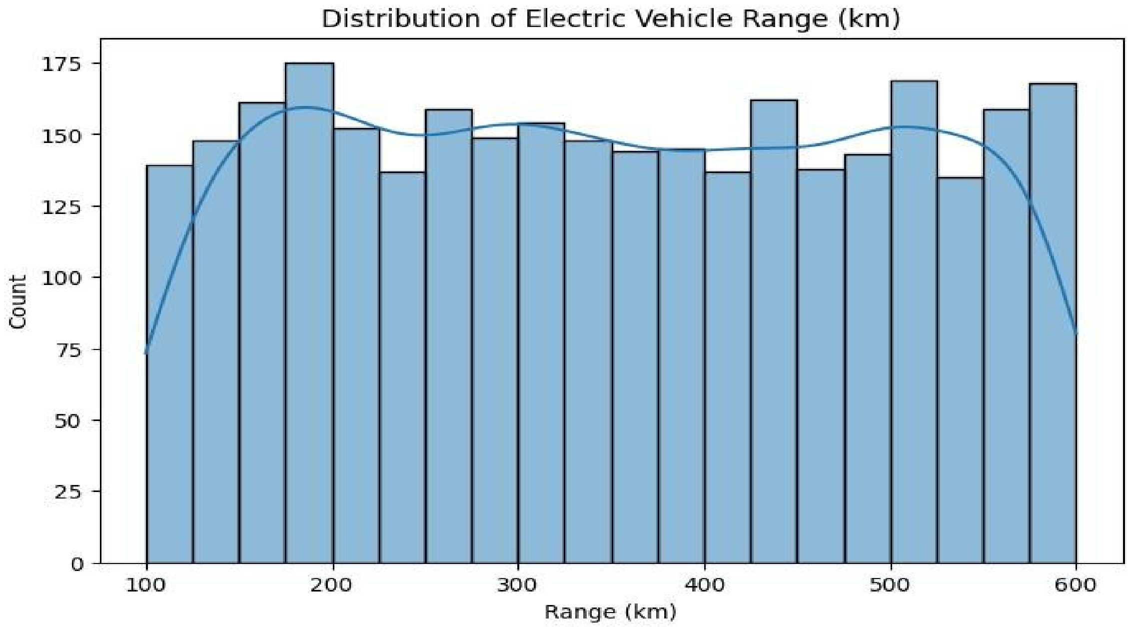

Figure 1: EV Sales by Manufacturer

3.2 Top Manufacturers

- Tesla, BYD, Hyundai, BMW, and Rivian lead in EV sales for 2024.
- Tesla dominates premium EVs; BYD and Hyundai specialize in affordable segments.
- Strong competition seen in Asian and European markets.

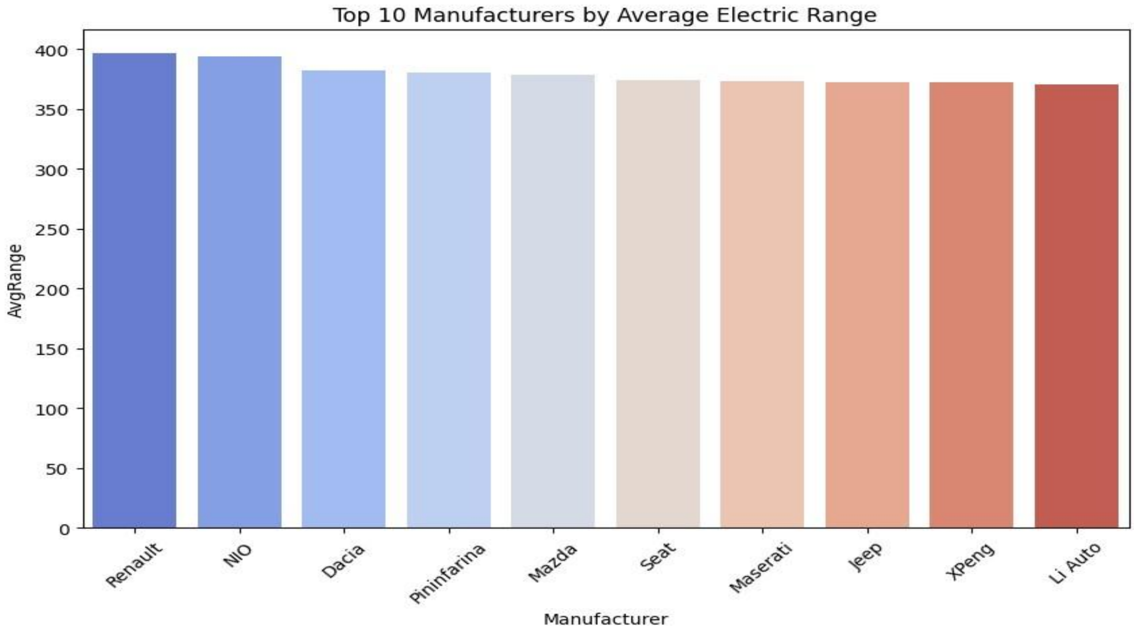

Figure 2: Range vs Battery Capacity Relationship

3.3 Technical and Environmental Patterns

- LFP (Lithium Iron Phosphate) batteries are most common for cost and stability.
- 80% of EVs report zero CO■ emissions, highlighting environmental benefits.
- Fast-charging (CCS/DCFC) models achieve full charge within 1 hour.

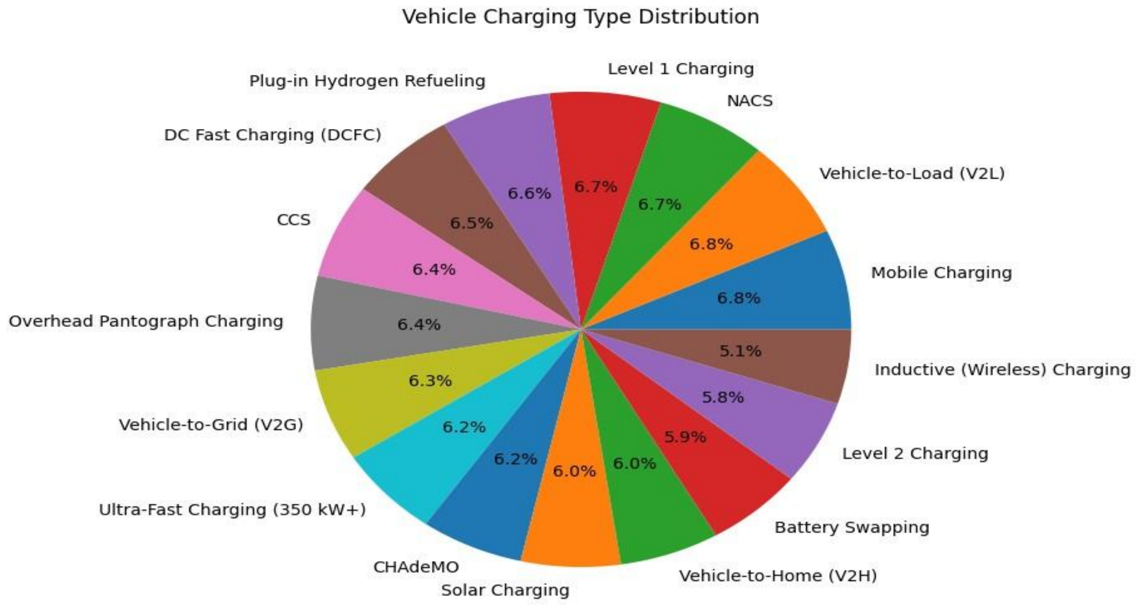

Figure 3: Price vs Range Correlation

4. Recommendations

4.1 Infrastructure and Policy

- Expand public fast-charging infrastructure, especially in Tier-2 cities.
- Support charging standardization to enable cross-brand compatibility.

4.2 Research and Development

- Invest in next-generation solid-state and LFP battery technologies.
- Encourage regional battery manufacturing for cost reduction.

4.3 Market Expansion

- Incentivize EV adoption for fleets and public transportation.
- Increase consumer awareness through sustainability-driven marketing.

4.4 Data and Forecasting

- Build predictive models using PySpark MLlib for future sales forecasting.
- Develop visualization dashboards to monitor EV performance and trends.

5. Future Analytics Opportunities

- Develop machine learning models for EV sales prediction.
- Cluster vehicles based on price-to-performance ratios.
- Integrate weather and charging station data for real-world efficiency analysis.
- Extend dataset with real-time telemetry for advanced energy forecasting.

6. Conclusion

The Electric Vehicles Data Analysis emphasizes the ongoing transition toward sustainable transportation. Consistent growth across global markets, coupled with technological improvements and policy incentives, is accelerating EV adoption. Clean and structured data enables precise trend identification and forecasting, laying the groundwork for data-driven EV ecosystem development.