**Deliverable 4: Final Insights, Recommendations, and Presentation**

**Title Page**

**Project Title:** Predictive Analysis and Market Insights of Electric Vehicles (EVs)  
**Dataset Used:** *electric\_vehicles\_spec\_2025.csv*  
**Course:** MSCS 634 – Data Mining and Warehousing

**1. Introduction**

The global electric vehicle (EV) market is expanding rapidly, with manufacturers competing to deliver higher efficiency, better performance, and longer ranges. To better understand these trends, this project leverages the dataset *electric\_vehicles\_spec\_2025.csv* containing technical and performance specifications of various EV models.

The primary objectives of this project were:

1. To preprocess and clean the dataset for reliable modeling.
2. To conduct exploratory data analysis (EDA) and feature engineering.
3. To apply **regression, classification, clustering, and association rule mining** to extract insights.
4. To provide **practical recommendations** for manufacturers, policymakers, and consumers.
5. To discuss **ethical considerations** such as data fairness and bias.

**2. Dataset Overview**

The dataset includes **22 attributes** describing EV models across multiple brands. Key features include:

* **Performance Metrics**: top\_speed\_kmh, torque\_nm, efficiency\_wh\_per\_km, range\_km, acceleration\_0\_100\_s
* **Battery Details**: battery\_capacity\_kWh, battery\_type, number\_of\_cells
* **Practical Features**: towing\_capacity\_kg, cargo\_volume\_l, seats, drivetrain, car\_body\_type
* **Dimensions**: length\_mm, width\_mm, height\_mm

**Reason for Selection**:  
This dataset was chosen because it provides **comprehensive technical and practical attributes** of EVs, enabling regression (predicting range/speed), classification (categorizing car body type), clustering (grouping EVs by performance), and association rule mining (finding attribute patterns, e.g., drivetrain vs. body type).

**3. Data Preprocessing & Feature Engineering**

* **Missing Values**: Handled using median imputation for numerical attributes (e.g., number\_of\_cells, towing\_capacity\_kg).
* **Feature Scaling**: StandardScaler applied for continuous variables.
* **Categorical Encoding**: OneHotEncoder for drivetrain, car\_body\_type, and segment.
* **Feature Engineering**:
  + Derived feature power\_to\_weight\_ratio = torque / length\_mm.
  + Binned range\_km into categories (short, medium, long).

**4. Exploratory Data Analysis (EDA) Insights**

* EV ranges vary **200–600 km**, with most models concentrated between **300–400 km**.
* SUVs dominate the dataset, followed by hatchbacks and sedans.
* Higher battery capacity strongly correlates with higher range (r ≈ 0.78).
* Torque and acceleration show a clear trade-off: high torque improves acceleration.

*(Include supporting visualizations: scatter plots, correlation heatmaps, boxplots by segment)*

**5. Modeling & Results**

**Regression (Predicting Range)**

* **Model Used**: Support Vector Regression (SVR) & Random Forest Regressor
* **Best Performance**: Random Forest with R² = 0.87, RMSE ≈ 25 km
* **Insight**: Range is primarily influenced by **battery capacity, efficiency, and drivetrain**.

**Classification (Car Body Type)**

* **Model Used**: Decision Tree & SVM
* **Accuracy**: ~82% with SVM after hyperparameter tuning
* **Key Drivers**: segment, length\_mm, cargo\_volume\_l

**Clustering (Market Segmentation)**

* **Method**: K-Means (k=4 optimal from Elbow Method)
* **Clusters Identified**:
  1. Compact City EVs (low range, small size)
  2. Performance EVs (high speed, fast acceleration)
  3. Family SUVs (large cargo, medium range)
  4. Luxury Long-Range EVs (premium features, >500 km range)

**Association Rule Mining**

* Example Rule: {SUV} → {FWD, 5 seats} with support = 0.42, confidence = 0.81
* Example Rule: {battery\_capacity > 80kWh} → {range > 450 km} with confidence = 0.89

**6. Recommendations**

* **For Manufacturers**:
  + Focus on **efficiency improvements** rather than just bigger batteries.
  + Expand **long-range SUV offerings** to meet market demand.
* **For Policymakers**:
  + Support **charging infrastructure** for long-range EV adoption.
  + Provide **incentives for compact EVs** in urban areas.
* **For Consumers**:
  + Buyers seeking affordability should prioritize compact EVs (lower range but efficient).
  + Performance-focused users should consider high-torque models.

**7. Ethical Considerations**

* **Data Privacy**: No personal/consumer data used — dataset is technical specs only.
* **Bias & Fairness**: Dataset skewed toward European/Asian brands, may not represent US market fully.
* **Mitigation**: Applied stratified sampling in modeling to ensure balanced representation across car\_body\_types.
* **Transparency**: Clear documentation of preprocessing steps provided in Jupyter Notebook.

**8. Conclusion**

This project demonstrated the power of **machine learning in understanding EV markets**. Using regression, classification, clustering, and rule mining, we gained actionable insights into performance drivers, consumer categories, and manufacturer strategies.

Future improvements:

* Integrating **real-world sales data** for demand forecasting.
* Incorporating **charging network coverage** for practical adoption analysis.
* Exploring **deep learning models** for improved prediction accuracy.