INNOMATICS®
RESEARCH LABS

INNOVATION. AUTOMATION. ANALYTICS

PROJECT ON

EV DATABASE

Title: Web Scraping and Exploratory Data Analysis of EV Database

Team Members

1. V. Roopa
2. K. Prasanthi
3. L. Rishitha

Preface:

INNOMATICS
RESEARCH LABS

Introduction:

- EV-Database is a popular online platform that provides detailed information about electric vehicles (EVs) from different brands and countries.

- It was launched in Europe to help people compare EVs easily using real-world usable data.

- EV-Database's main goal is to support EV buyers by giving clear specs like range, battery capacity, charging speed, efficiency, price, and performance.

- It also offers smart comparison tools, filters (range, price, charging, body type, etc.), and updated EV listings to help users choose the best electric car for their needs.

Business Problem:
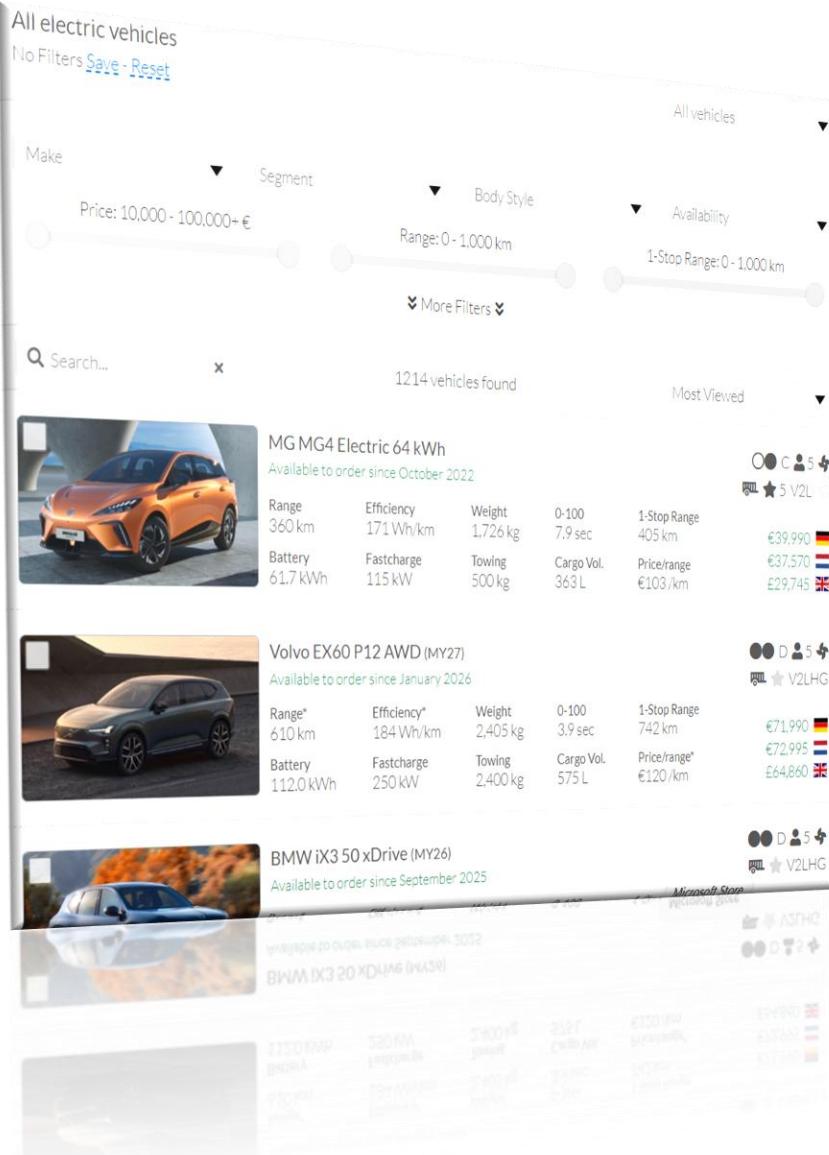
- EV information is scattered and inconsistent across many websites, brand pages, and review platforms.

- Customers struggle to compare electric vehicles properly because specifications like range, charging speed, and battery capacity are often shown differently by different sources.

- EV buyers lack clear, data-driven insights to identify which cars provide the best real-world range, best charging performance, and best value for money.

- There is a strong need for data-driven analysis to summarize EV performance, highlight market trends, and visualize comparisons across brands, price ranges, and vehicle types.

Objectives:

- To compare electric cars based on range, charging speed, efficiency, and price to help customers choose the best model.

- To identify the best value-for-money EVs by analyzing performance vs cost across brands and segments.

- To recommend the most suitable EVs for different needs (city use, long trips, family, budget) using data-driven insights.

Web Scraping

- EV Database was selected as the data source

- Used **browser developer tools** (Inspect Element) to locate relevant HTML tags Data.

- Used **BeautifulSoup and Requests** to extract data from EV Database.

- Sent **HTTP requests** to fetch web pages.

- Performed **Exploratory Data Analysis (EDA)** to identify trends.

Dataset:

price_per_	availability	Range	efficiency	weight	speed	Fast_charg	distance	cargo_vol	Battery	german_pr	netherland	uk_price	towing	driver_typ	No.of_Sea	Heat_Pum	Towbar	Rating	Bidirection	Load	Home	Grid
103	Available	360	171	1726	7.9	115	405	363	61.7	39990	37570	29745	500	Rear Whee	5	1	1	5	1	1	1	1
120	Available	610	184	2405	3.9	250	742	575	112	71990	72995	64860	2400	All Wheel	5	1	1	5	1	1	1	1
114	Available	610	178	2360	4.9	230	742	578	108.7	70900	70501	58755	2000	All Wheel	5	1	1	5	1	1	1	1
225	Discontinu	360	197	2440	6.1	102	388	500	71	77300	84998	69905	2500	All Wheel	5	1	1	5	0	1	1	1
128	Available	235	159	1365	9	67	261	185	37.3	34990	27999	23995	0	Front Whe	4	0	0	4	0	1	1	1
110	Discontinu	350	166	1811	7.3	82	375	385	58	39370	36939	34125	0	Rear Whee	5	1	0	5	0	1	1	1
122	Discontinu	455	165	2072	4.8	124	535	971	75	52990	53990	51990	1600	All Wheel	5	1	1	5	0	1	1	1
92	Available	585	145	2055	6.7	235	763	506	85	55859	53554	45615	1500	Rear Whee	5	1	1	5	1	1	1	1
126	Discontinu	445	169	2054	5	124	524	971	75	55970	52990	51990	1600	All Wheel	7	1	1	5	0	1	1	1
103	Discontinu	450	171	1946	7	115	488	385	77	46450	46989	39625	0	Rear Whee	4	1	0	5	1	1	1	1
146	Discontinu	430	174	2072	4	124	506	971	75	60970	58990	59990	1600	All Wheel	5	1	1	5	0	1	1	1
89	Available	445	135	1847	6.2	110	523	682	60	37970	36990	37990	1000	Rear Whee	5	1	1	5	0	1	1	1
135	Available	245	184	1645	7.9	51	235	400	45.1	31990	31690	30850	750	Front Whe	5	1	1	5	1	1	1	1
128	Discontinu	375	160	2003	5.9	110	438	971	60	45970	45990	44990	1600	Rear Whee	5	1	1	5	0	1	1	1
91	Available	455	171	1885	7.7	105	472	485	78	41390	41495	35995	1000	Front Whe	5	1	1	5	1	1	1	1
140	Available	440	208	2510	4.5	165	502	578	91.3	58990	58190	58990	1500	All Wheel	5	1	1	5	1	1	1	1
83	Available	580	136	1822	5.2	120	688	682	79	45970	45990	44990	1000	Rear Whee	5	1	1	5	0	1	1	1
97	Available	490	157	1953	7.2	130	574	952	77	47970	46990	49825	1600	Rear Whee	5	1	1	5	0	1	1	1
142	Available	500	188	2535	4.3	190	606	698	94	71281	72631	60350	2400	All Wheel	5	1	1	5	1	1	1	1
113	Available	380	158	1981	7.2	110	440	952	60	40970	39990	41990	1600	Rear Whee	5	1	1	5	0	1	1	1
96	Available	450	171	2115	6.6	120	493	470	77	44180	43990	35560	1000	Rear Whee	5	1	1	5	0	1	1	1
86	Discontinu	550	136	1822	5.2	124	655	682	75	44990	45990	44990	1000	Rear Whee	5	1	1	5	0	1	1	1
93	Available	360	181	1920	8	110	403	430	65	32400	31495	31495	750	Rear Whee	5	1	1	5	1	1	1	1
91	Available	335	155	1504	8	75	352	326	52	32900	30990	23945	500	Front Whe	5	1	1	4	1	1	1	1
101	Available	385	177	2015	7.5	130	448	651	68	38990	38990	49825	1600	Rear Whee	5	1	1	5	1	1	1	1
143	Available	460	172	2108	3.8	120	535	971	79	62970	62990	61990	1600	All Wheel	5	1	1	5	0	1	1	1

Tools Used:

Beautiful Soup

pandas

.[RegEx]*

matplotlib

NumPy

seaborn

Data Cleaning Steps:
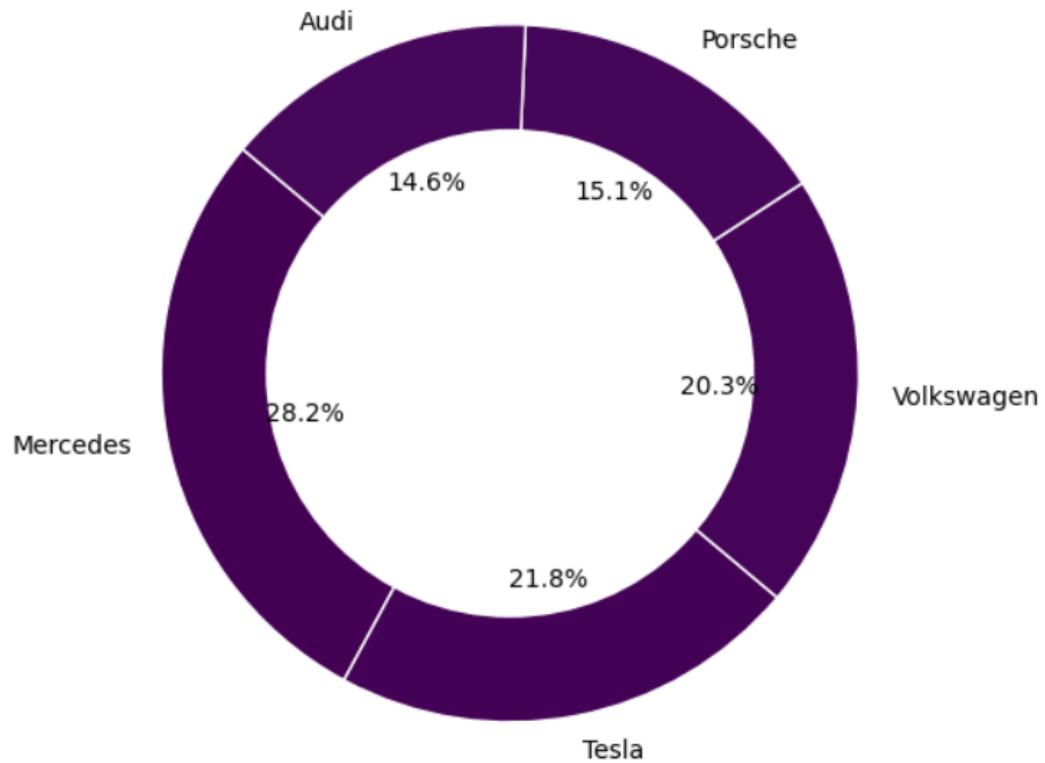
- Filled missing **Rating** values using mode.

- Cleaned and extracted numeric values from columns like **Range, Efficiency, Weight, Speed, Fast_charge, Distance, Battery, Cargo_vol, Towing** and converted them to proper numeric types.

- Standardized **German, Netherlands, UK prices** by removing symbols/commas and converting to numeric.

- Simplified text columns like **Driver_Type** and **Availability** into clean fields (**driver_type, availability_status**).

- Dropped unused raw columns after creating cleaned, analysis-ready columns.

INNOMATICS
RESEARCH LABS

Data Visualization:

- **Mercedes dominates** with the highest share (28.2%), indicating the widest model variety in the dataset.

- **Tesla ranks second (21.8%)**, showing strong presence and leadership in the EV-focused segment.

- **Volkswagen is close behind (20.3%)**, making competition tight among the top three brands.

- **Porsche (15.1%) and Audi (14.6%)** have smaller shares, suggesting more premium or niche positioning.

- **Market concentration is high** — the top 3 brands together account for nearly 70% of total models, showing dominance by a few key players.
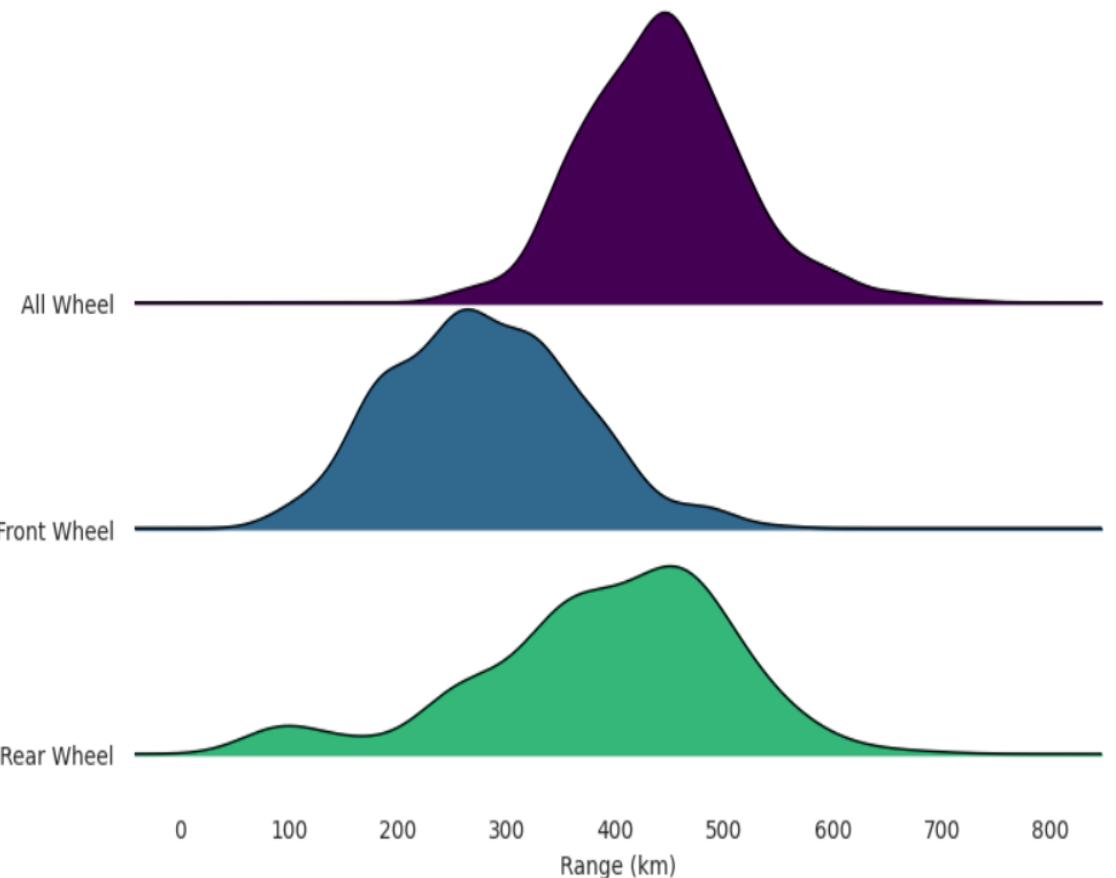
Top 5 Brands Distribution (Model Count)

Data Visualization:
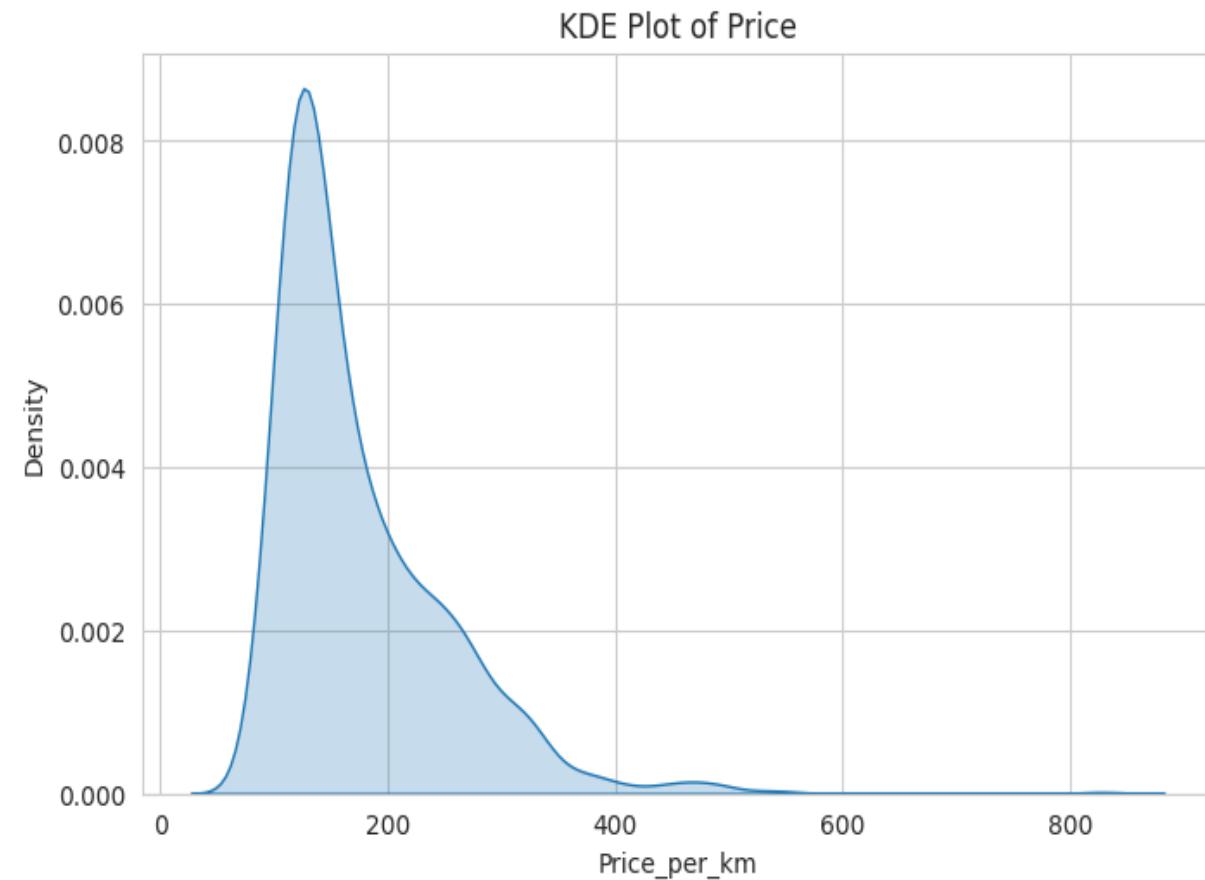
- **AWD vehicles offer the highest range** (around 450–500 km), indicating strong long-distance capability.

- **RWD vehicles provide balanced performance**, with range mostly between 350–450 km.

- **FWD vehicles have lower range** (200–350 km), suggesting focus on city or budget-friendly models.

- **AWD shows greater variability**, meaning it includes both mid-range and high-range EVs.

- **Overall trend: AWD > RWD > FWD**, proving that drivetrain type significantly impacts EV range and is crucial for performance-based conclusions.

Ridgeline Plot: Range Distribution by Driver Type
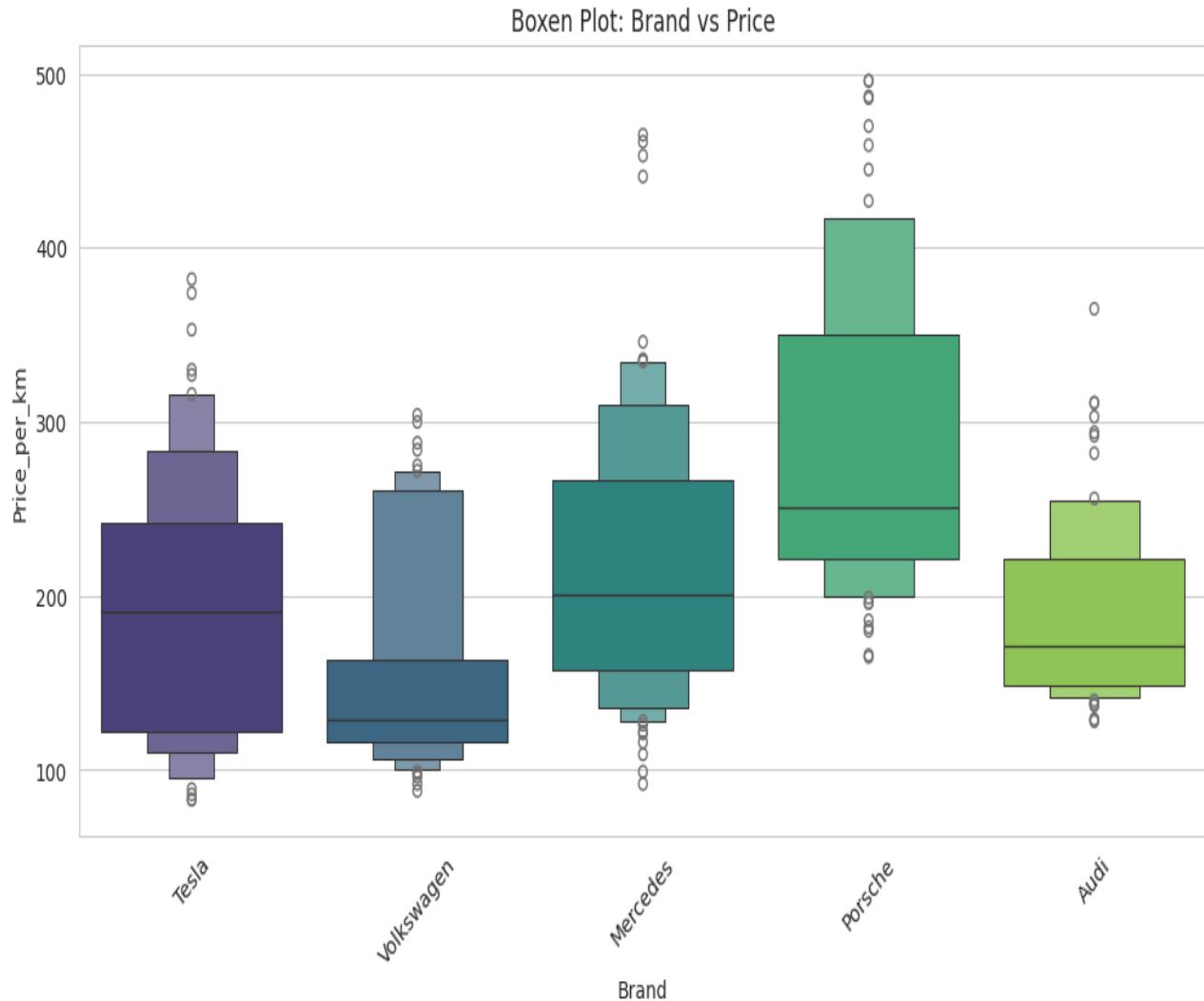
Data Visualization:

- **Most EVs are concentrated between 100–200 price per km**, showing strong affordability in the market.

- **The distribution is right-skewed**, meaning only a small number of vehicles have very high cost per km.

- **Peak density at lower-mid values** indicates that the majority of models are cost-efficient.

- **A long right tail (400+ price per km)** represents premium or luxury EV segments.

- **Overall conclusion:** The EV market in this dataset is largely cost-effective, with affordability dominating and only a few high-cost outliers.

KDE Plot of Price
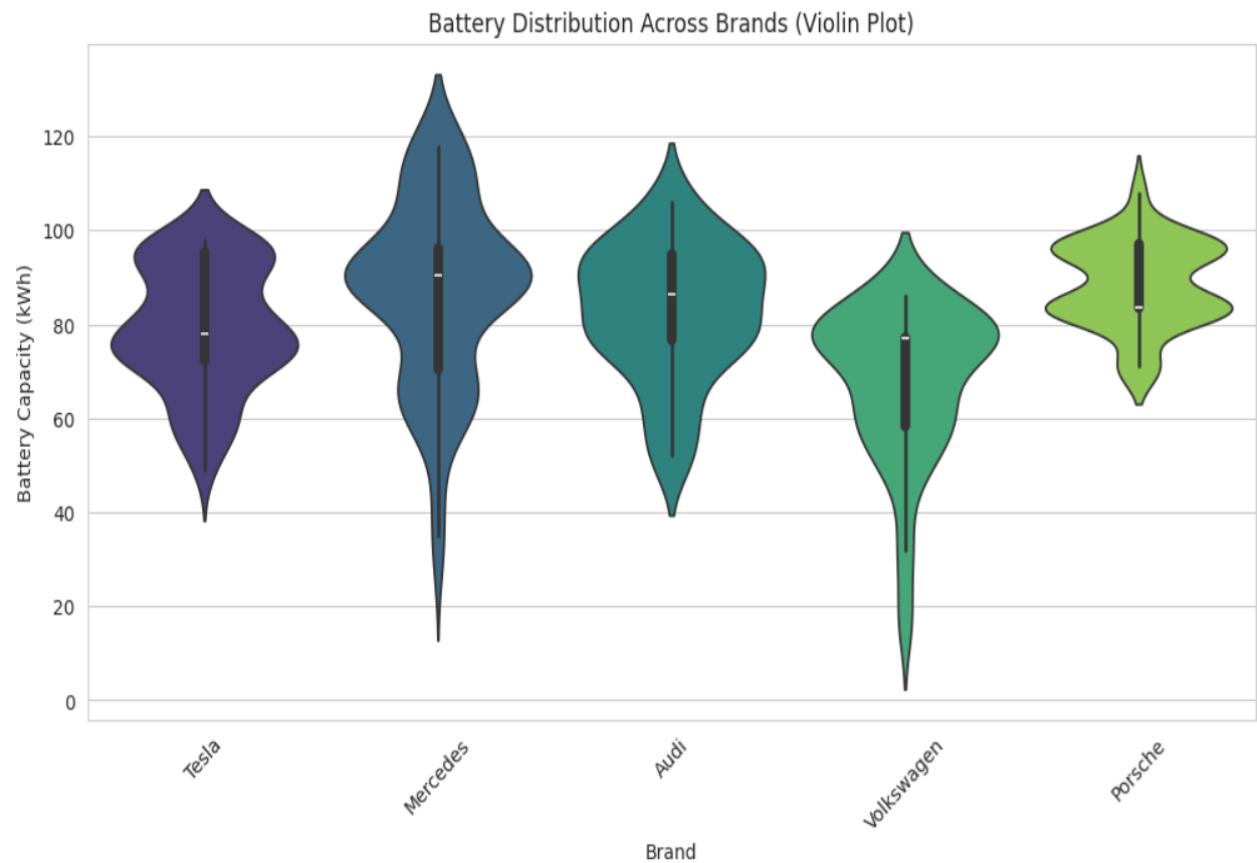
Data Visualization:

- **Porsche** has the **highest median price per km** (most premium).

- **Volkswagen** has the **lowest price per km** (most affordable).

- **Tesla** and **Mercedes** sit in the **mid-to-high range**.

- **Porsche** and **Mercedes** show **high price variation** across models (more outliers).

- **Conclusion:** Clear split between **luxury brands (high cost/km)** and **budget-friendly brands (low cost/km)**.

Boxen Plot: Brand vs Price
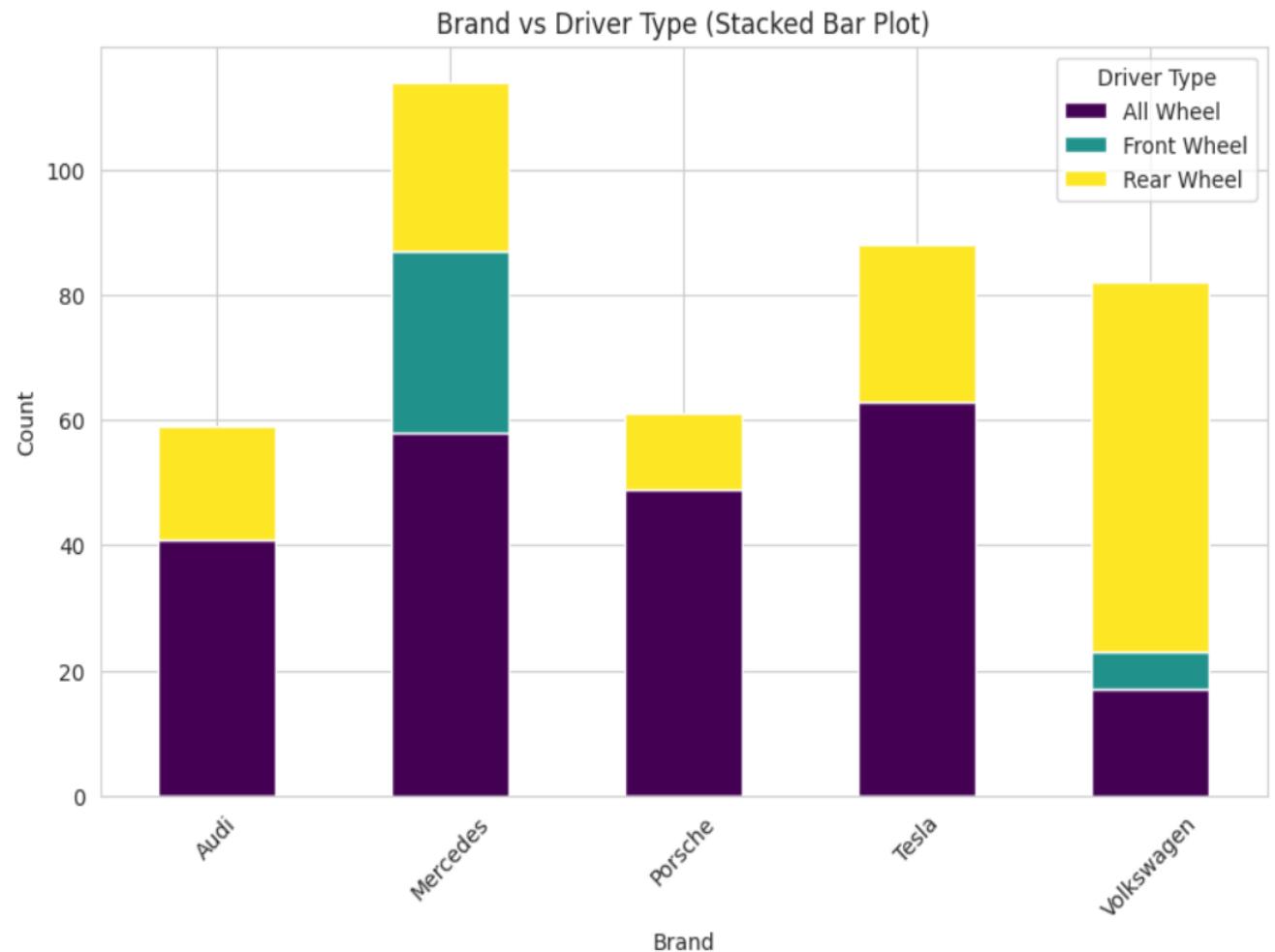
Data Visualization:

- **Mercedes** has the **widest battery range** (most diverse models).

- **Porsche** has **consistently high battery capacities** (premium focus).

- **Audi & Tesla** are **balanced**, mostly **mid-to-high capacity**.

- **Volkswagen** has **lower/moderate battery capacities** (mid-range EVs).

- **Conclusion: Mercedes & Porsche = luxury/high battery, Volkswagen = practical/efficient mid-range**

Battery Distribution Across Brands (Violin Plot)
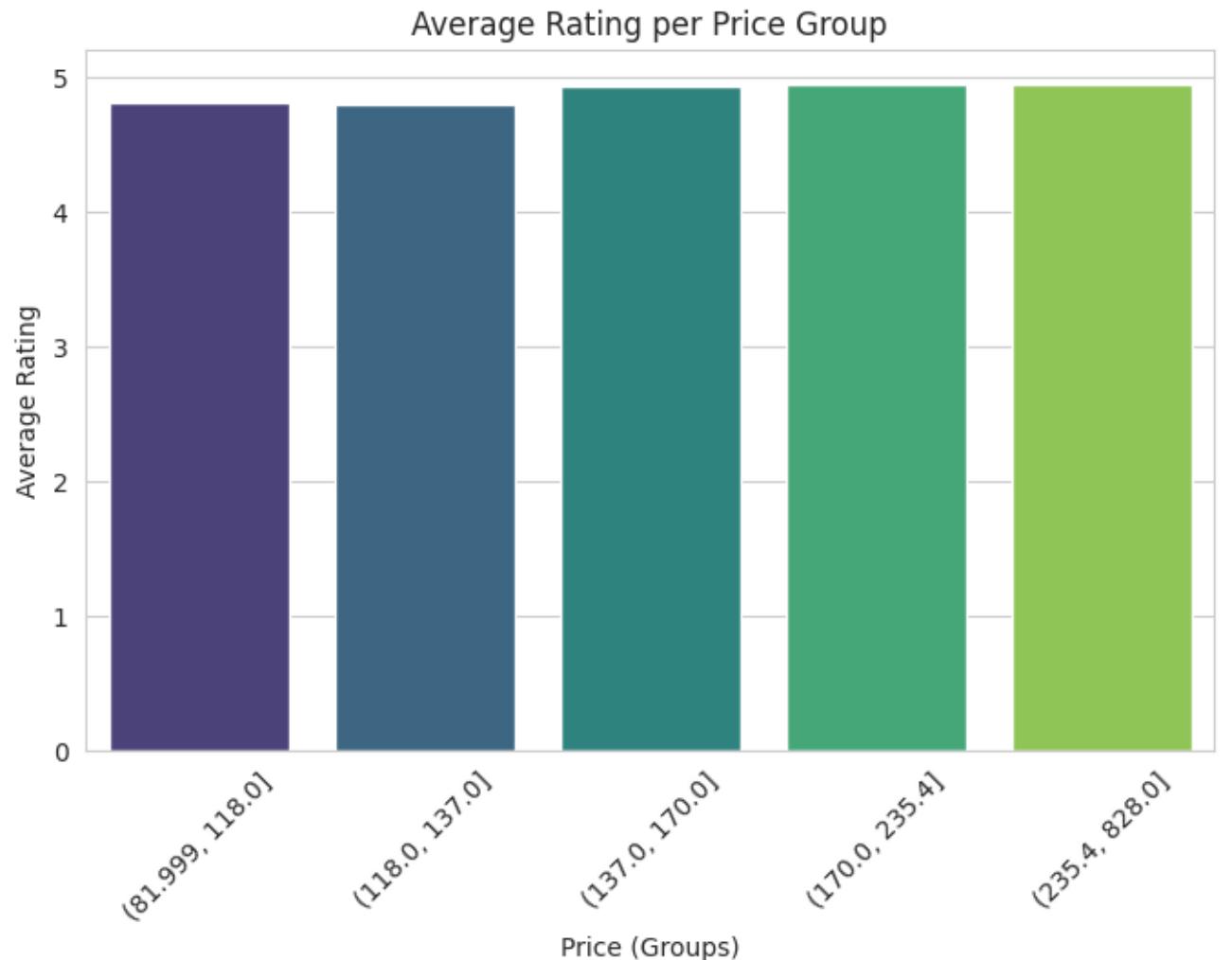
Data Visualization:

- **AWD dominates** across brands, especially **Tesla & Mercedes**.

- **Mercedes** has the **most diverse drivetrain mix** overall.

- **Volkswagen** has a **higher RWD share** compared to AWD.

- **FWD is rare**, seen mainly in **Mercedes & Volkswagen**.

- **Conclusion:** Premium brands focus on **AWD performance**, while **Volkswagen** follows a more **balanced/RWD-oriented** approach.


Brand vs Driver Type (Stacked Bar Plot)
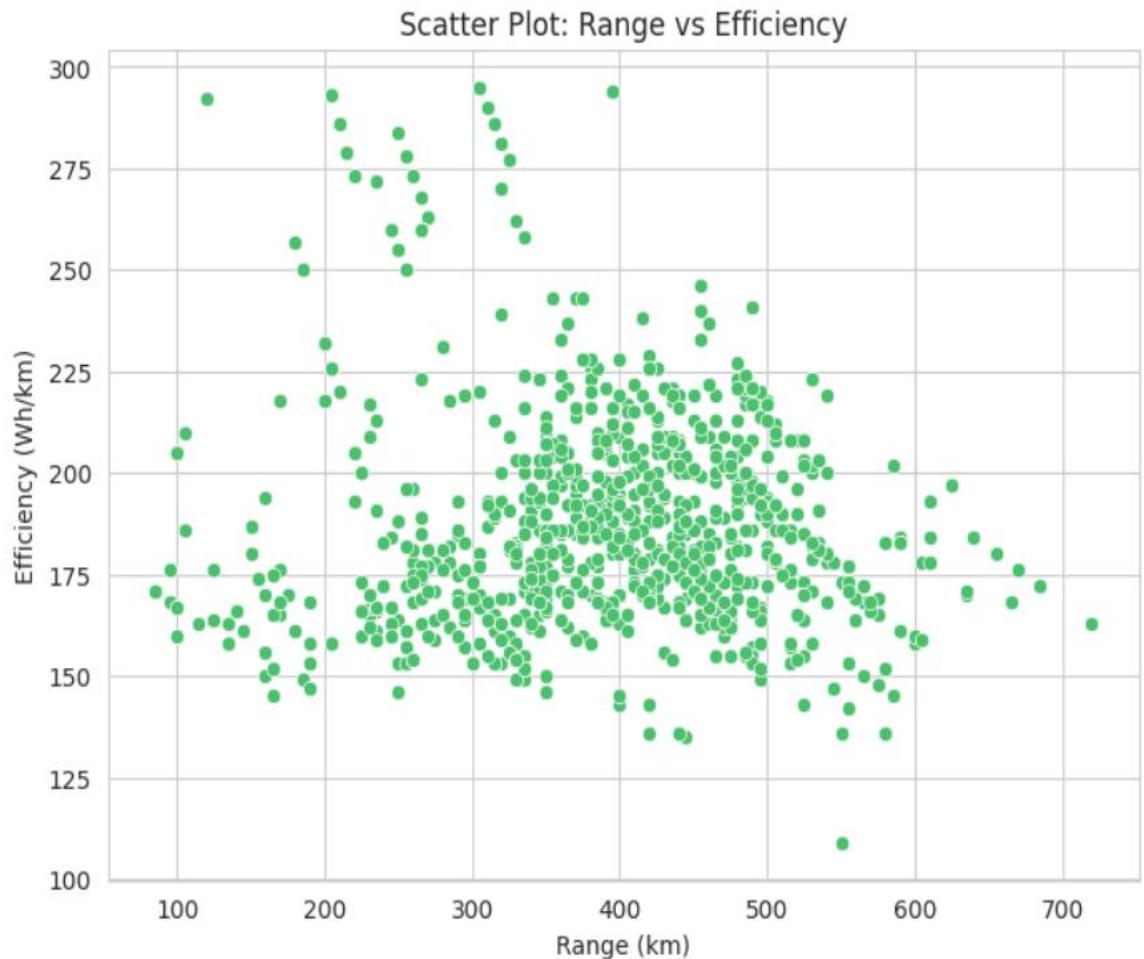
Data Visualization:

- Ratings are **high across all price groups** (~4.8–5.0).

- **Premium price groups** have **slightly higher ratings**.

- Rating differences are **very small**, so price/km doesn't strongly affect satisfaction.

- **Mid-to-high price segments** show the **highest average ratings**.

- **Conclusion:** Satisfaction is strong overall, with a **small advantage for premium EVs**.

Average Rating per Price Group
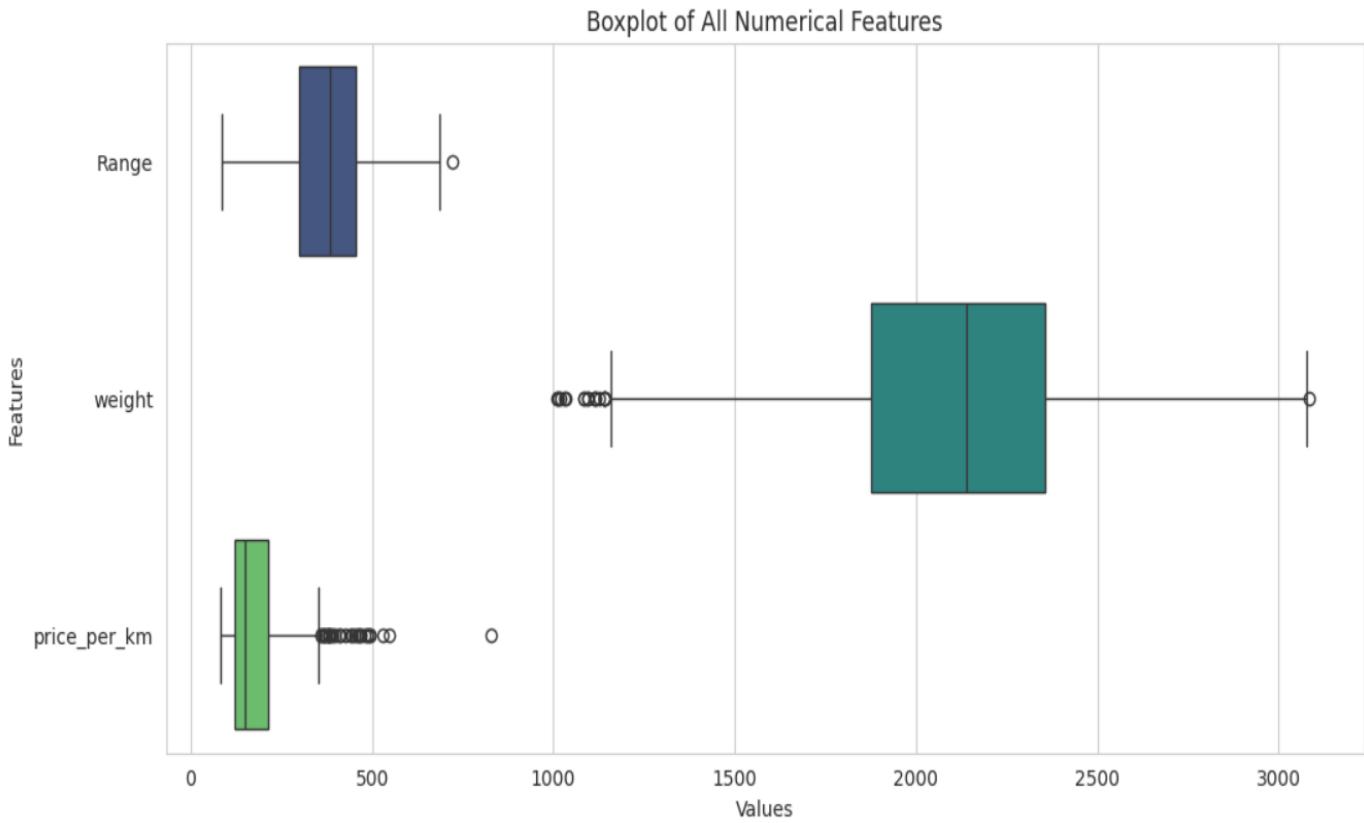
Data Visualization:

- Range and efficiency show a **moderate positive trend** (mostly 170–210 Wh/km).

- Most EVs fall in the **300–500 km range** segment.

- **Higher range ≠ always better efficiency**, showing improved tech.

- Some **high Wh/km outliers (250+)** appear at moderate ranges (likely performance models).

- **Conclusion:** EVs are well-optimized for **mid-to-high range with reasonable efficiency**.


Scatter Plot: Range vs Efficiency
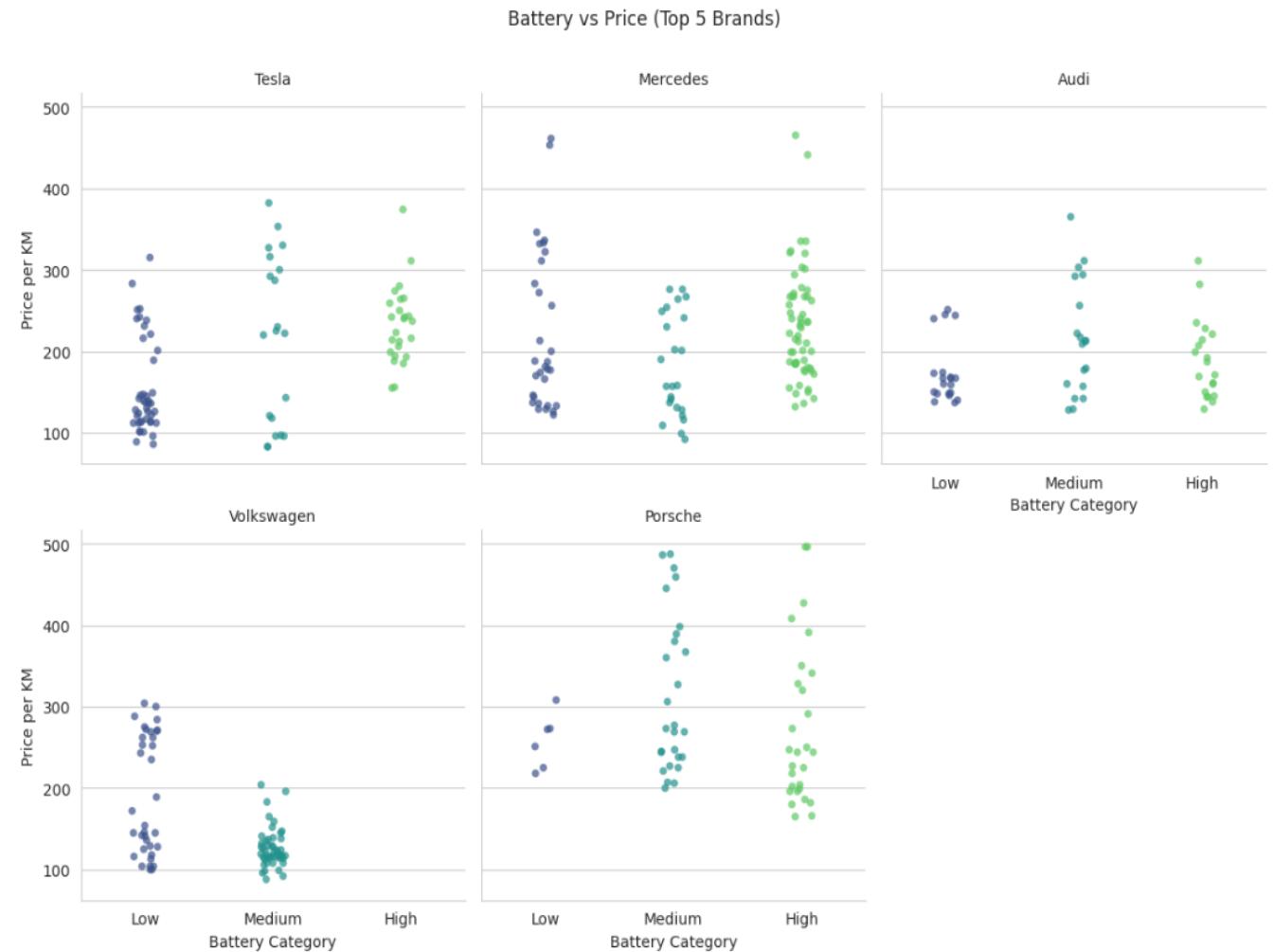
Data Visualization:

- **Weight varies the most**, with a wide spread and many outliers.

- **Range is moderately spread**, mostly mid-range with a few high outliers.

- **Price per km is right-skewed**, with many high-end outliers (premium EVs).

- EVs have a **high median weight** due to batteries.

- **Conclusion:** Differences are driven more by **weight and pricing** than by range.


Boxplot of All Numerical Features
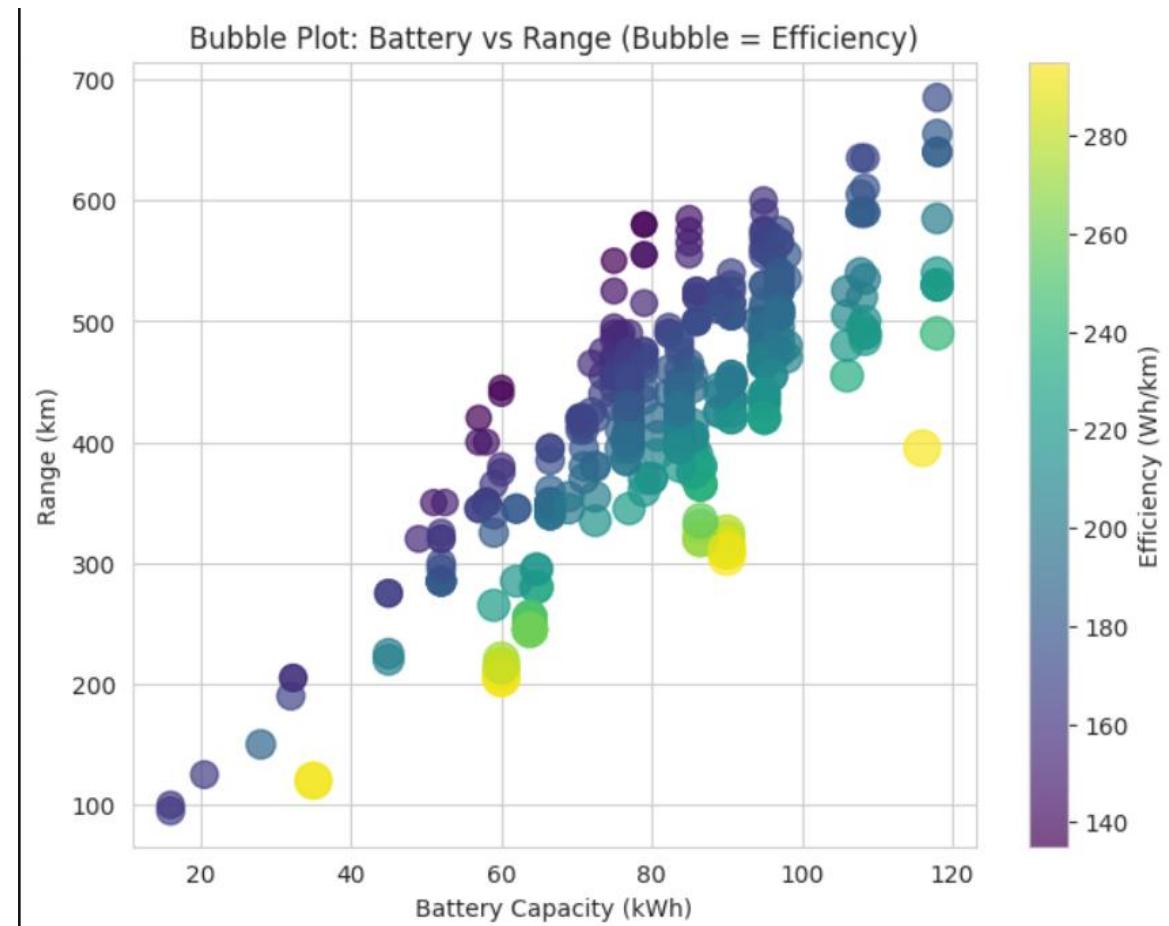
Data Visualization:

- **Price per km rises** from **Low → Medium → High battery** categories.
- **Porsche** is the **most expensive** in every battery group.
- **Mercedes & Tesla** are **mid-to-high priced**, mainly in Medium/High batteries.
- **Volkswagen** stays **comparatively affordable**, even in Medium batteries.
- **Conclusion:** Bigger batteries cost more, but brands differ — **Porsche luxury**, **Volkswagen value**, **Tesla/Mercedes balanced**.

Battery vs Price (Top 5 Brands)
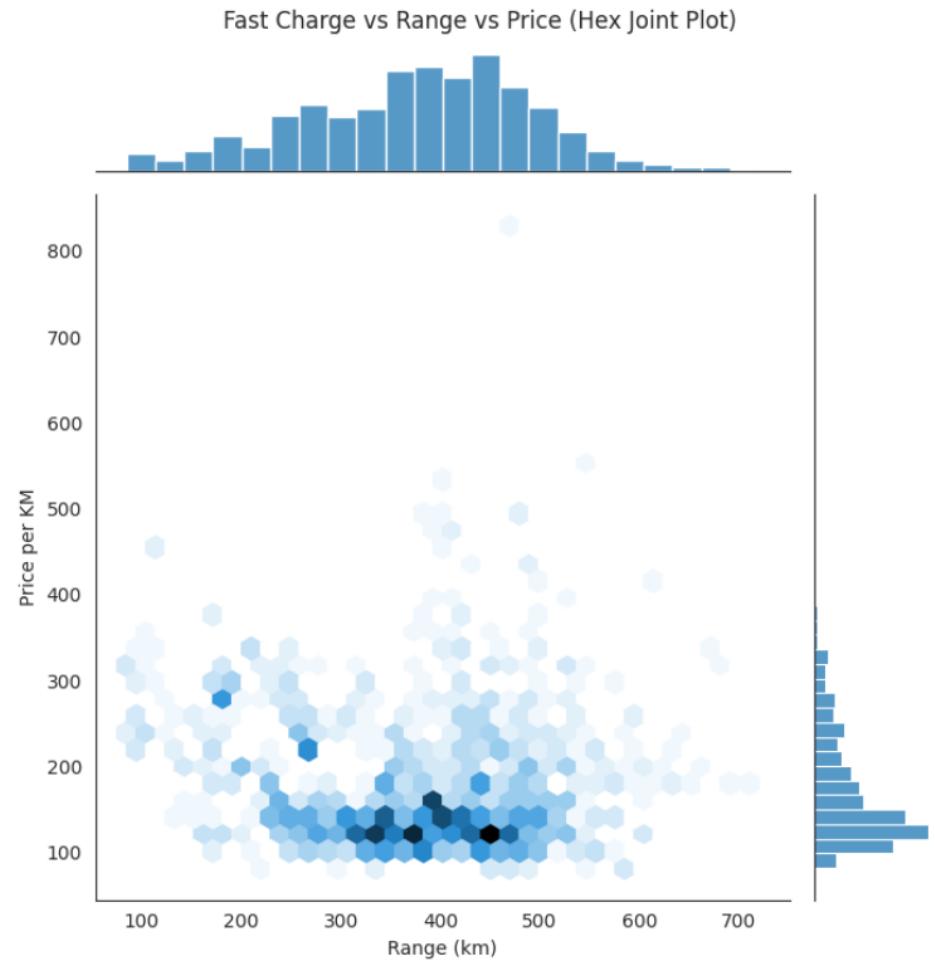
Data Visualization:

- **Battery capacity and range** have a **strong positive relationship**.

- Most EVs cluster at **60–100 kWh** with **350–550 km range**.

- **More efficient EVs (lower Wh/km)** get better range even with moderate batteries.

- Bigger batteries usually mean more range, but **efficiency strongly affects results**.

- **Conclusion: Battery size drives range**, but **efficiency maximizes performance**.

Bubble Plot: Battery vs Range (Bubble = Efficiency)
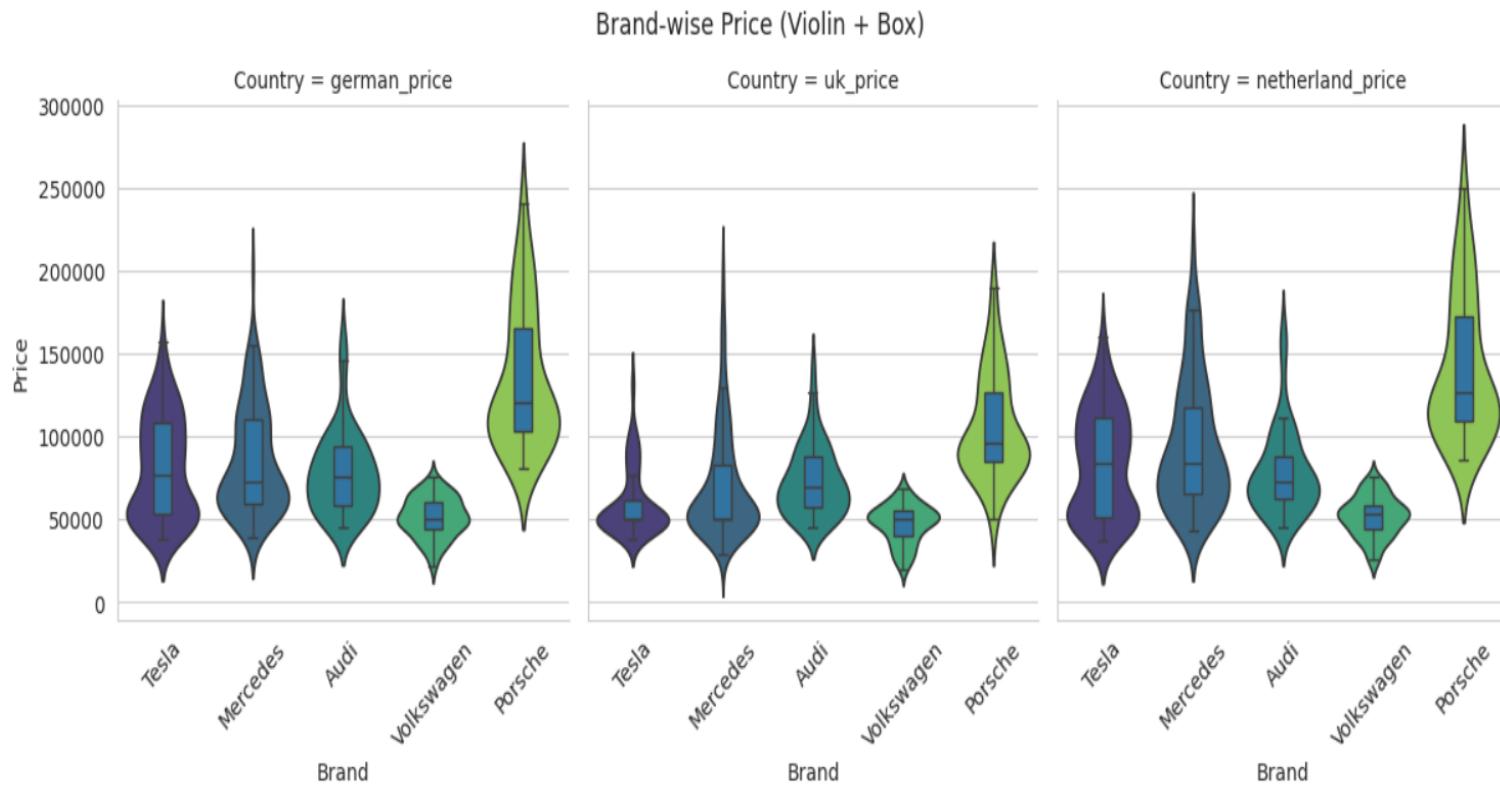
Data Visualization:

- Most EVs cluster in **300–500 km range** and **100–200 price/km** (main market).

- **Range and price/km** show a **moderate positive trend**.

- **400+ price/km outliers** are rare (luxury models).

- Many **mid-to-high range EVs stay affordable**, showing good tech balance.

- **Conclusion:** Market focuses on **strong range + reasonable cost**, with only a small ultra-premium

Fast Charge vs Range vs Price (Hex Joint Plot)
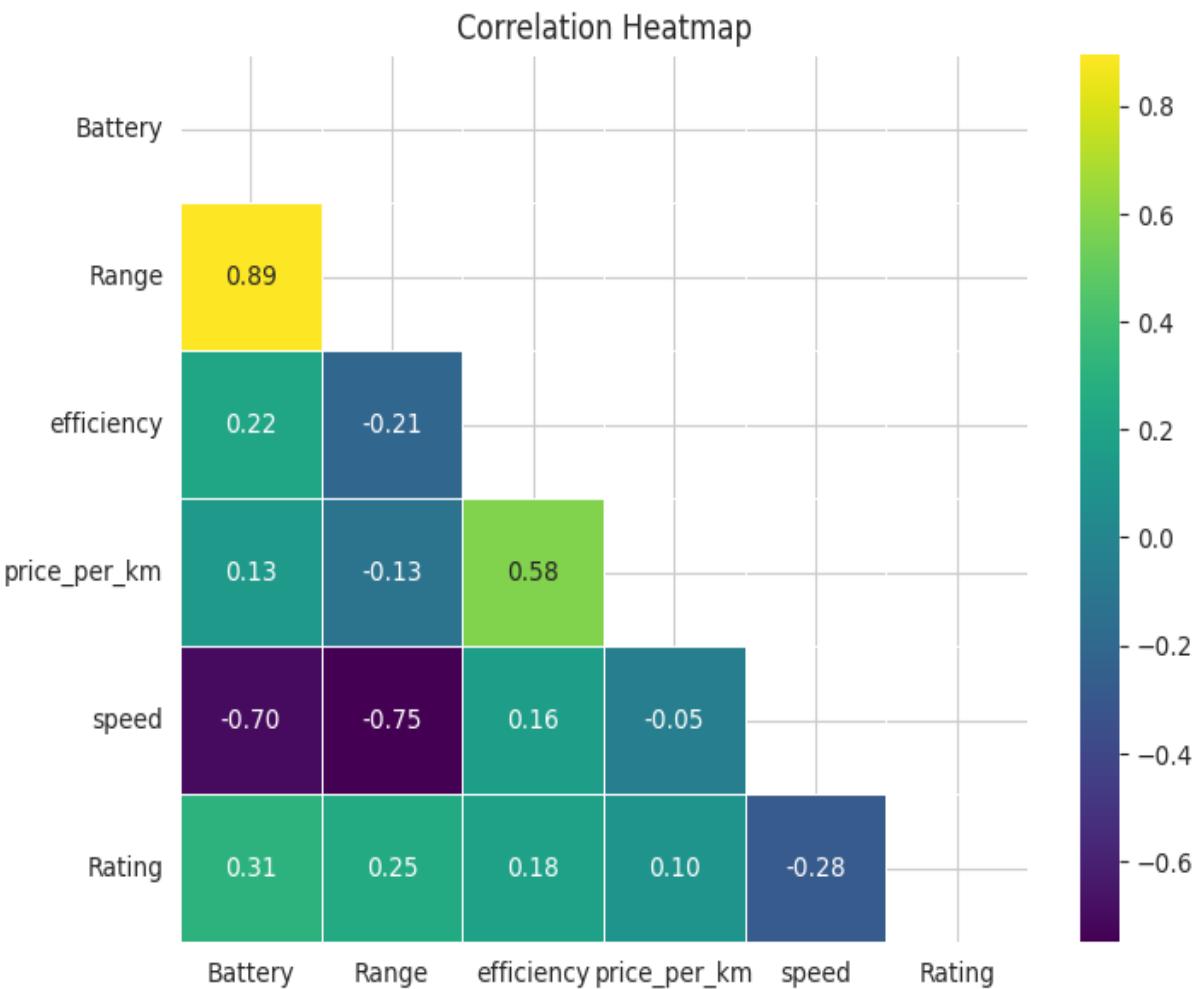
Data Visualization:

- **Porsche** has the **highest median price** in Germany, UK, and Netherlands.

- **Volkswagen** has the **lowest median price** across countries.

- **Mercedes, Audi, and Tesla** sit in the **mid-to-high price range**.

- **Porsche & Mercedes** show **higher price variation** (wider model range).

- **Conclusion:** Brand pricing is consistent globally — **Porsche luxury**, **Volkswagen budget**, others **mid-premium**.

Brand-wise Price (Violin + Box)

Data Visualization:

- **Battery ↔ Range** is **very strongly positive (0.89)**.

- **Speed** is **strongly negative** with **Battery (-0.70)** and **Range (-0.75)**.

- **Efficiency ↔ Price/km** is **moderately positive (0.58)**.

- **Ratings** slightly increase with **Battery (0.31)** and **Range (0.25)**.

- **Conclusion: Battery drives range**, **speed reduces performance**, and **efficiency affects cost**.

Correlation Heatmap

Conclusion:

Based on the comparative analysis of range, charging speed, efficiency, and price, the following conclusions were derived:

- Battery capacity strongly determines range, but efficiency decides real-world performance. The best EVs are not the ones with the biggest batteries, but the ones that balance battery size and efficiency.

- Mid-range EVs (350–500 km range, 60–100 kWh battery) offer the best value-for-money**, as they provide strong performance without extreme pricing.

- Premium brands like Porsche deliver high performance but at significantly higher price per km, making them suitable for luxury/performance buyers rather than value-focused customers.

- Tesla and Mercedes provide the best performance-to-cost balance, combining good range, strong efficiency, AWD options, and high customer ratings.

- Volkswagen stands out in affordability, making it ideal for budget-conscious and city-focused users.

INNOMATICS
RESEARCH LABS

Final Recommendations Based on Customer Needs

For City Use (Daily Commute):

- Efficient, lower battery EVs (Volkswagen models) with lower price per km.

For Long Trips:

- High-range AWD models (Tesla / Mercedes mid-range models).

For Families:

- Balanced range + battery + comfort (Mercedes / Tesla).

For Budget Buyers:

- Volkswagen (best affordability with reasonable range).

For Performance / Luxury Buyers:

- Porsche (high battery, high range, premium pricing).

Experiences & Challenges:

- Extracted large datasets from static web pages using Python (Beautiful Soup, Requests).

- Faced challenges with inconsistent data formats, units, and missing values during data cleaning.

- Automated the scraping process to handle multiple pages efficiently and reduce manual effort.

- Converted mixed text fields into numerical values for accurate analysis and visualization.

- Standardized pricing data across countries to ensure consistent comparisons.

- Ensured ethical data collection and maintained data accuracy throughout the analysis process.

THANK YOU