INNOMATICS®
RESEARCH LABS

INNOVATION. AUTOMATION. ANALYTICS

PROJECT ON

Title: Web Scraping and Exploratory Data Analysis of Electric Vehicles Data

INNOMATICS
RESEARCH LABS

Introduction:

- Electric Vehicles (EVs) use rechargeable batteries and electric motors instead of fuel engines.
- They produce zero tailpipe emissions, making them eco-friendly and cost-efficient.
- EVs convert over 85% of electrical energy into motion, compared to ~25% for petrol cars.
- Supported by advances in battery tech, charging stations, and green policies.
- EVs are shaping the future of clean, smart, and sustainable transportation.
- With Tesla's Roadster (2008) and Nissan Leaf (2010), EVs entered mainstream markets. Rapid improvements in battery range, charging infrastructure, and government incentives have made EVs globally popular.
- EVs are among the fastest-growing vehicle segments, with millions sold each year and major automakers fully transitioning to electric fleets.

Business Problem:

Electric vehicle information is dispersed across multiple automotive sources, making it difficult to obtain a unified view of prices, performance, and efficiency.

Manufacturers, policymakers, and consumers struggle to benchmark EV models due to inconsistent data on range, energy usage, and charging capabilities.

The lack of centralized analytics limits understanding of key market dynamics such as pricing trends, technological advancements, and brand competitiveness.

Buyers and sustainability advocates often lack accessible, data-driven insights to compare models and make informed purchasing decisions.

There is a growing need for comprehensive EV analytics to consolidate vehicle data, evaluate efficiency benchmarks, and support strategic decisions in the electric mobility sector.

Objectives:

- To analyze and compare electric vehicle models based on price, driving range, efficiency, and market segment.
- To provide data-driven insights that help consumers, manufacturers, and policymakers understand EV performance and market positioning.
- To explore brand-wise trends in pricing, energy efficiency, and vehicle types to identify leading manufacturers and innovation patterns.
- To study relationships between cost, range, and efficiency, revealing trade-offs and opportunities for improvement in EV design.
- To highlight market evolution and adoption patterns, supporting strategic decisions in the transition toward sustainable mobility.

Columns description:

S.No	Column Name	Description
1	**Car_Brand**	Name of the company or manufacturer of the car.
2	**Car_Model**	Specific name or model of the electric vehicle.
3	**Brand**	Manufacturer name (same as Car_Brand, used for grouping).
4	**Range_in_km**	Total distance (in kilometers) the EV can travel on a full charge.
5	**Efficiency (Wh/km)**	Energy consumed (in watt-hours) per kilometer — lower value means higher efficiency.
6	**Weight(kg)**	Total weight of the vehicle.
7	**Acceleration_time(0-100kmph)**	Time (in seconds) taken to accelerate from 0 to 100 km/h.
8	**Long_distance_total(km)**	Maximum range possible during long-distance driving.
9	**Battery(KWH)**	Battery capacity in kilowatt-hours.
10	**Fast_charge(KW)**	Maximum charging power supported (in kilowatts).
11	**Towing_weight(Kg)**	Maximum weight (in kilograms) the car can tow.
12	**Cargo_volume(L)**	Storage or luggage space in liters.
13	**Price_per_km(€/Km)**	Price per kilometer of driving range.
14	**Car_price_in_germany(€)**	Vehicle price in Germany (in Euros).
15	**Car_price_cat(€)**	Price category (e.g., Economy, Mid-range, Premium).
16	**Range_per_kWh**	Shows how many kilometers the car can travel per kWh of battery (efficiency indicator).
17	**Price_per_kWh**	Shows how much each kWh of battery costs in the vehicle (cost-efficiency measure).
18	**Charging_Speed_km_per_min**	Shows how many kilometers of range can be added per minute of fast charging.

Tools Used:


BeautifulSoup


pandas


.[RegEx]*


matplotlib


NumPy


seaborn


plotly
Express


INNOMATICS
RESEARCH LABS

Data Cleaning Steps:

- Checked for missing and null values on the entire data frame. Found no missing and null values

- Moved onto further steps like removing inconsistencies like having text values or special characters in numeric columns which prevents type conversion of numeric columns into int or float from object data type.

- Used regex to remove special characters from numerical and categorical columns. And for text values in numerical columns like N/A, unknown which were present right from data extraction from the source, used replace method and filled np.NaN and after that filled those NaN values with median by checking skewness on particular columns

```
df['Price_per_km(€/Km)']=df['Price_per_km(€/Km)'].apply(lambda x:re.sub(r" /km","",x))
```

```
Last executed at 2025-11-07 19:53:59 in 50ms
```

```
df['Car_price_in_germany(€)']=df['Car_price_in_germany(€)'].apply(lambda x:re.sub(r"€","",x))
```

```
Last executed at 2025-11-07 19:53:59 in 63ms
```

```
# Step 1: Replace 'Unknown' with NaN for numerical consistency
df["Cargo_volume(L)"] = df["Cargo_volume(L)"].replace("unknown", np.nan)

# Step 2: Convert column to numeric
df["Cargo_volume(L)"] = pd.to_numeric(df["Cargo_volume(L)"], errors="coerce")

# Step 3: Replace missing (NaN) values with the median
df["Cargo_volume(L)"].median()
df["Cargo_volume(L)"].fillna(df["Cargo_volume(L)"].median(), inplace=True)
```

```
df['Range_per_kWh']=df['Range_in_km']/df['Battery(KWH)']
```

```
Last executed at 2025-11-07 19:54:01 in 15ms
```

```
df['Price_per_kWh']=df['Car_price_in_germany(€)']/df['Battery(KWH)']
```

```
Last executed at 2025-11-07 19:54:01 in 12ms
```

```
df['Charging_Speed_km_per_min']=(df['Fast_charge(KW)']/df['Efficiency (Wh/km)'])*60
```

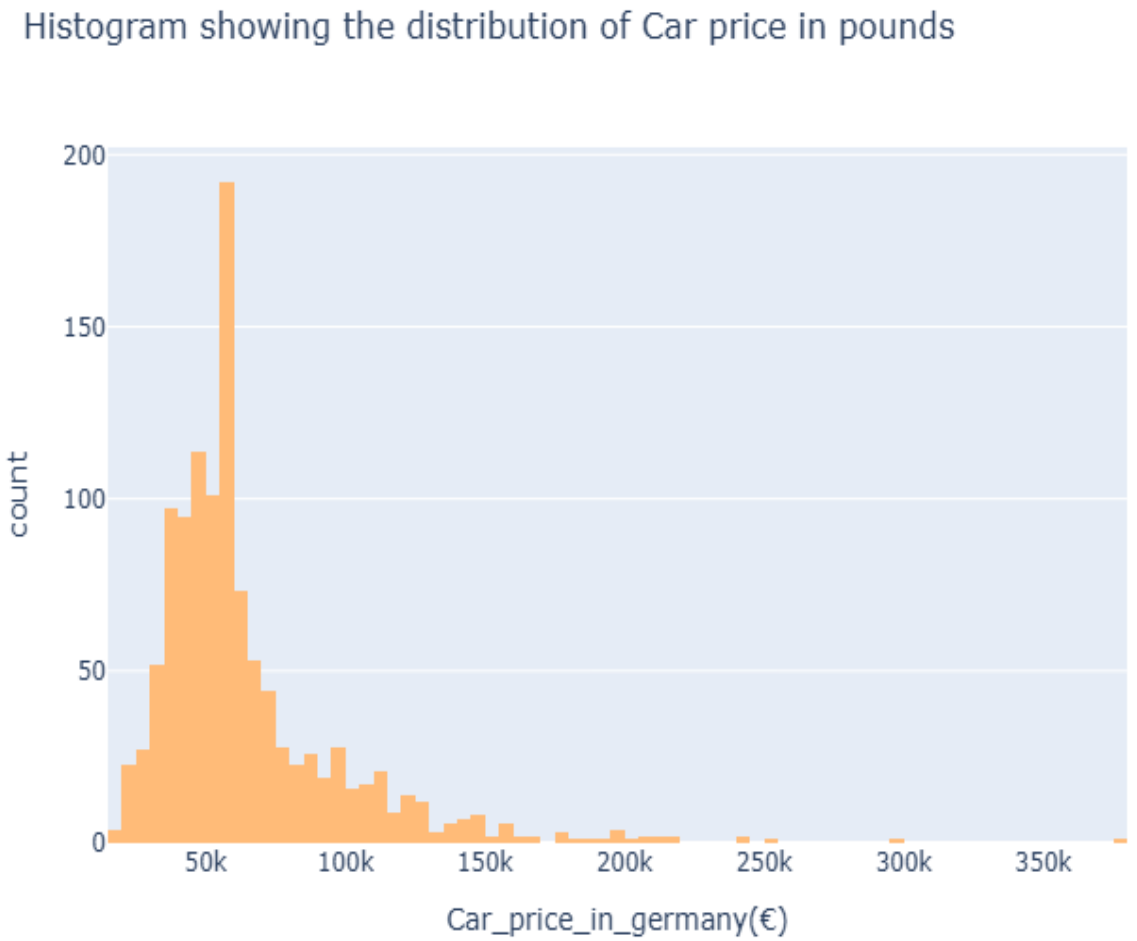
```
Last executed at 2025-11-07 19:54:01 in 18ms
```

INNOMATICS RESEARCH LABS

Data Visualization:

1.1 Univariate Analysis - Histogram

- The Histogram tells us the distribution of price of car in Euros
- Most of the cars priced between 45k-60k Euros targeting budget/Entry level customers
- There are very a smaller number of cars priced above 200k-350k Euros



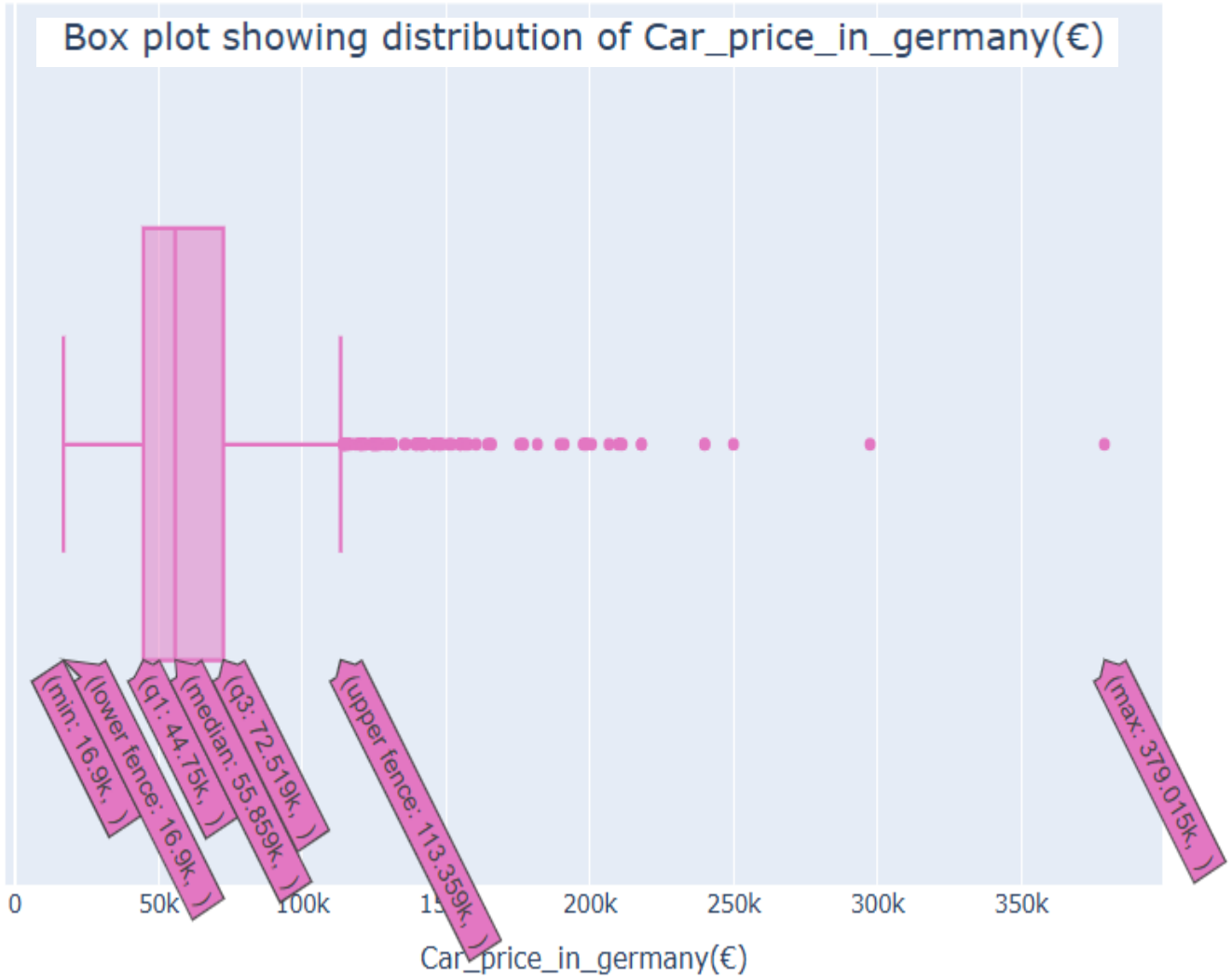
INNOMATICS RESEARCH LABS

Data Visualization:

1.2 Univariate Analysis- Boxplot

- Here we are using box plot to visualize the distribution of numerical data i.e., car price in Germany

- Like the presence of extreme price values in the column. As we can see there are some extreme points on the right side due to which box plot got squeezed to left side

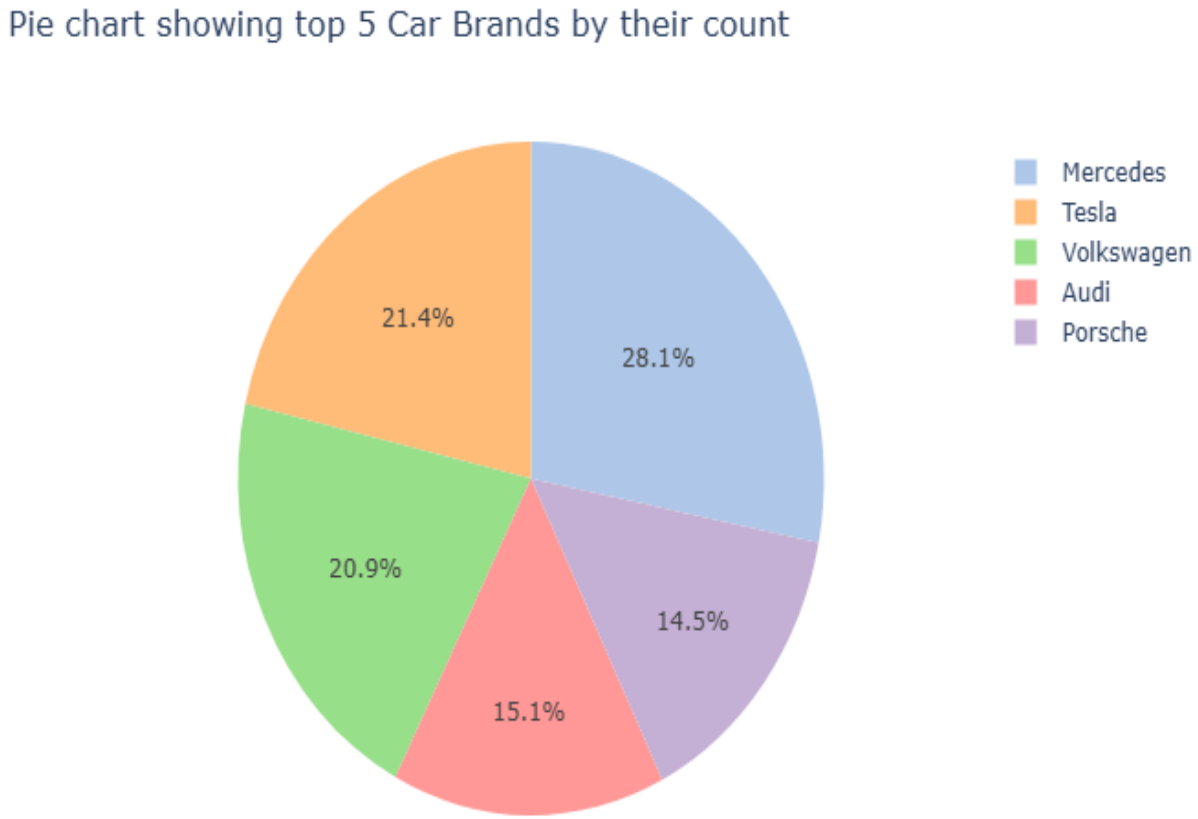


INNOMATICS RESEARCH LABS

Data Visualization:

1.3 Univariate Analysis- pie chart

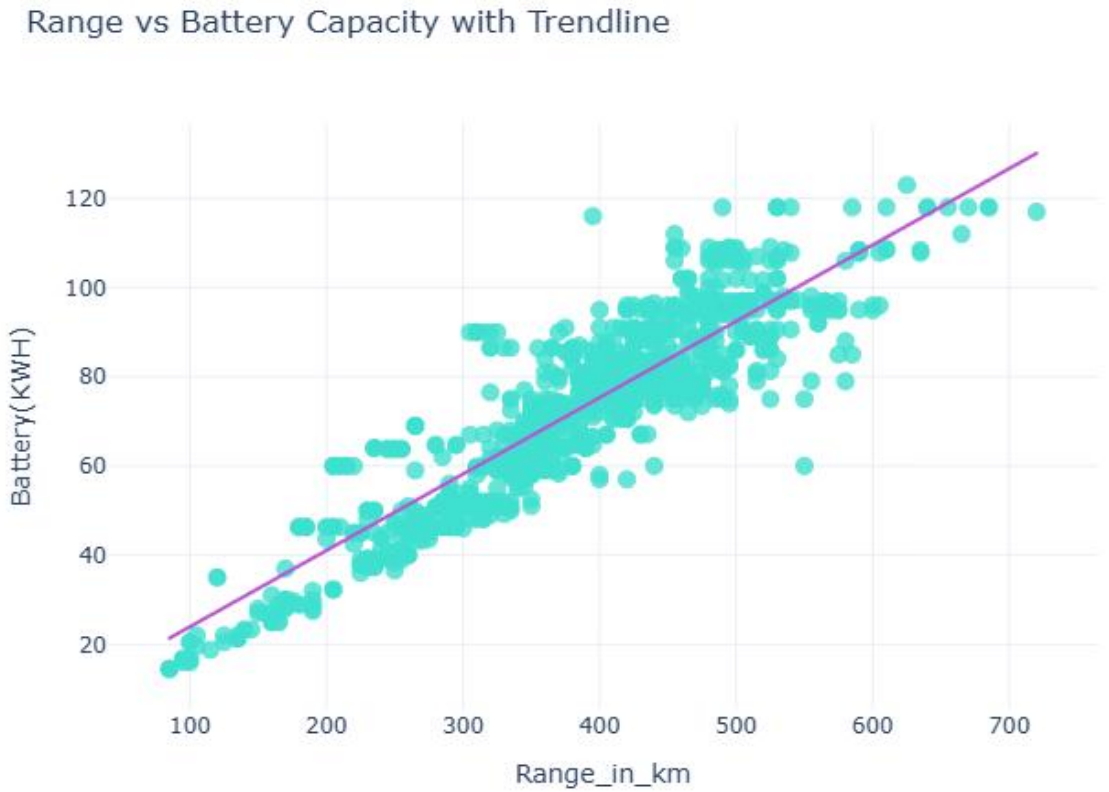
- Pie chart showing how the different car brands share the overall segment
- Mercedes has 110 cars by count, Tesla has 84, Volkswagen 82, Audi 59, and Porsche 57
- There are about 74 Car brands available in our dataset but for now we're considering only top 5



Data Visualization:

2.1 Bivariate Analysis- pie chart

- Battery capacity and range are linearly correlated.
- A larger battery means the vehicle can supply power for a longer time, allowing it to travel farther before needing a recharge.
- However, the relationship is not perfectly linear because range also depends on factors like Vehicle weight, Motor efficiency, Aerodynamics, and Driving conditions



	Range_in_km	Battery(KWH)
Range_in_km	1.000000	0.900471
Battery(KWH)	0.900471	1.000000

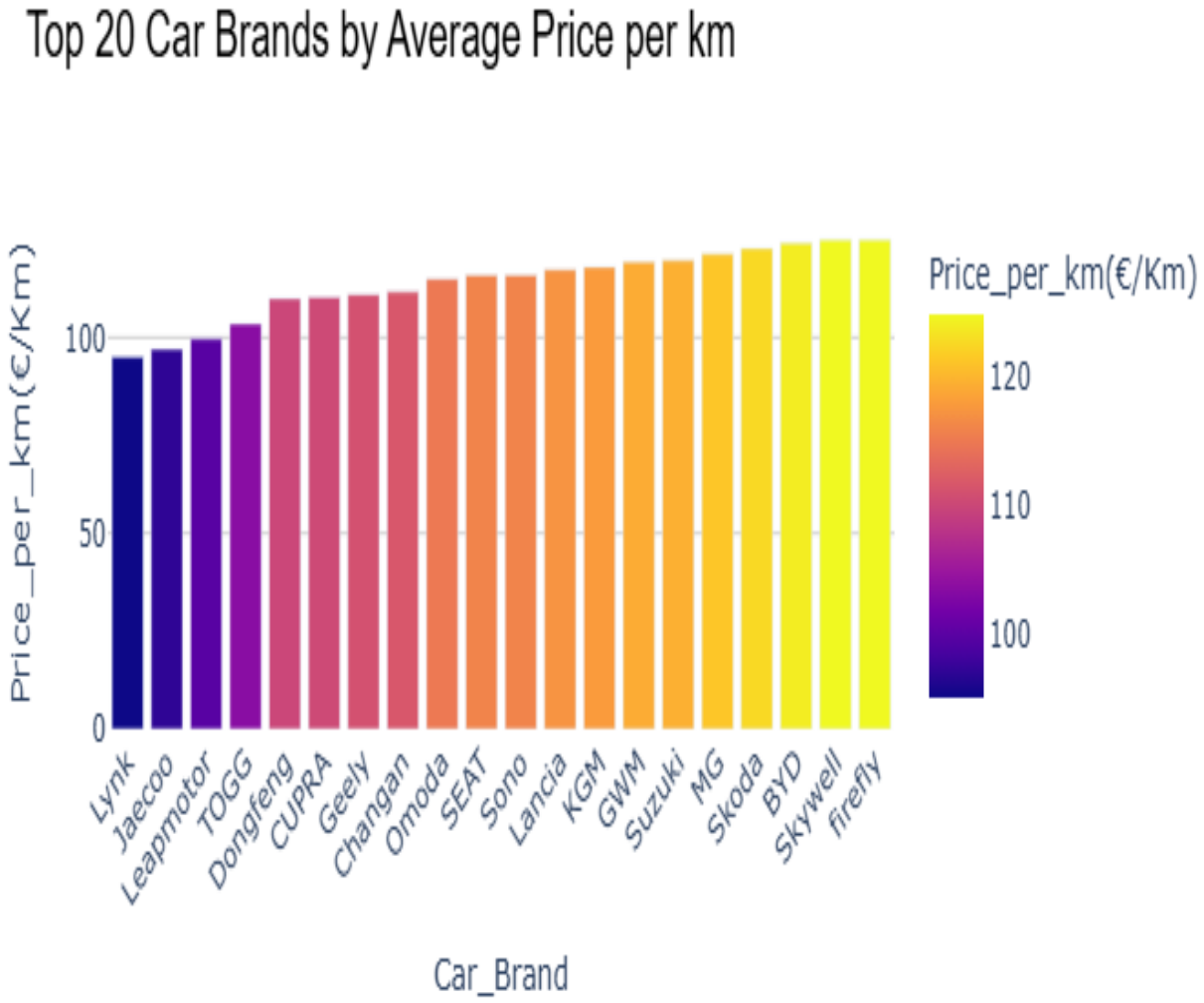
INNOMATICS RESEARCH LABS

Data Visualization:

2.2 Bivariate Analysis-Bar chart

- Brands like Lynk, Jaecoo, Leapmotor, and TOGG have the lowest price per km, meaning they offer more cost-efficient driving. it means they offer better mileage (higher efficiency) — in other words, you spend less money per kilometer traveled.

- High-end brands appear toward the right (higher Price_per_km).

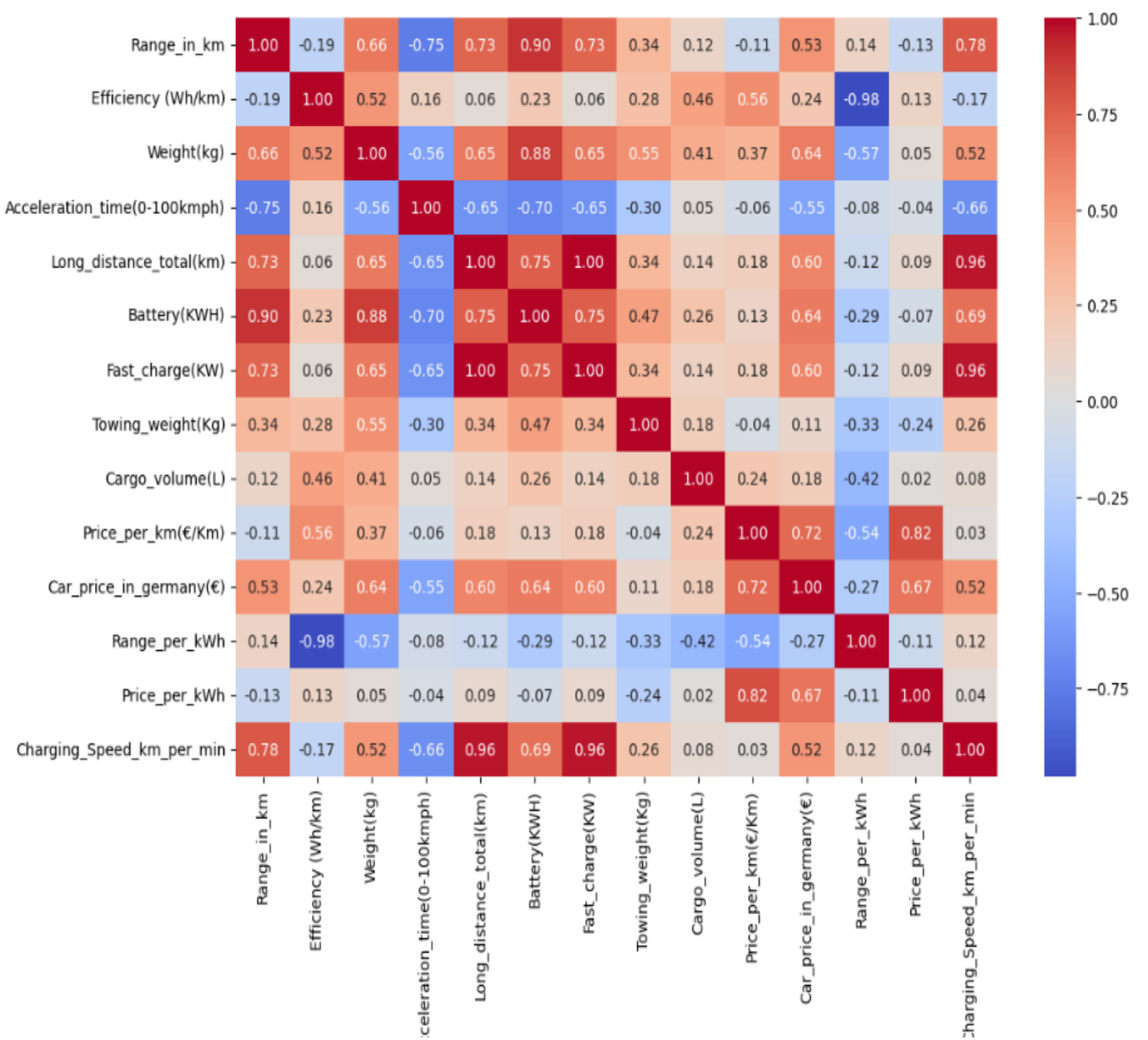


INNOMATICS RESEARCH LABS

Data Visualization:

3.1 multivariate Analysis- Heatmap

- The Correlation matrix or Heatmap showing the correlation of one numeric variable with another
- Long_distance_total(Km) and Fast_charge(KW) $r = 1.00 \rightarrow$ Perfect positive correlation.
- Long_distance_total(Km) and Charging_Speed_km_per_min $r = 0.96$
- Fast_charge(KW) and Charging_Speed_km_per_min $r = 0.96$
- Efficiency (Wh/km) and Range_per_kWh $r = -0.98 \rightarrow$ Very strong negative correlation.

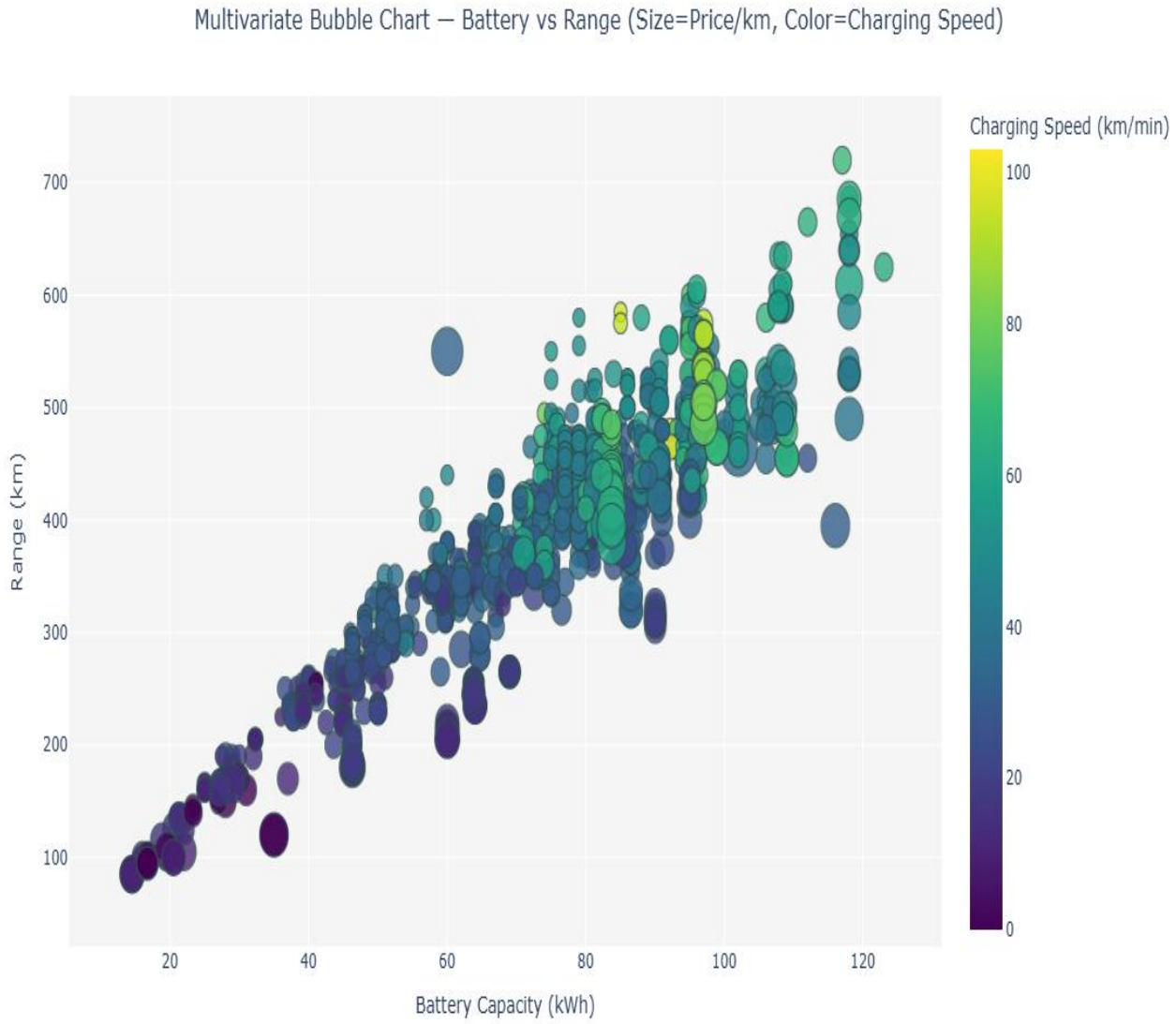


INNOMATICS RESEARCH LABS

Data Visualization:

3.2 multivariate Analysis- Scatter plot

- X-axis (Range_per_kWh): Efficiency — more distance per energy.
- Y-axis (Price_per_kWh): Cost efficiency — how much each kWh “costs”.
- Bubble size (Battery): Represents energy capacity and Color: Car pricing segment.
- Range increases with battery capacity, showing a strong positive correlation — most EVs above 80 kWh deliver 500–700 km range
- Higher charging speeds (green–yellow bubbles) are mostly seen in mid-to-large battery capacities (70–120 kWh), indicating optimized fast-charging technologies in premium EVs.



INNOMATICS RESEARCH LABS

Key Business Questions:

- Which brands offer the most energy-efficient and high-performing electric vehicles?
- How does an EV's price relate to its range, battery capacity, and overall efficiency?
- Which models provide the best value for money, based on price per kilometer or price per kWh?
- How does charging speed vary across different brands and battery sizes?
- What is the impact of vehicle weight and acceleration on energy efficiency and range?
- Which price category (Economy, Mid-range, Premium) offers the most balanced mix of performance, range, and cost?

Conclusion:

1.Growing Range and Efficiency Improvements
Electric vehicles show continuous progress in driving range and energy efficiency, with many modern models exceeding 400–500 km per charge, reflecting strong advancements in battery technology

2.Brand Leadership and Market Diversity
Tesla, BMW, Hyundai, and Volkswagen lead the EV market with diverse offerings across segments — from compact city cars to luxury SUVs — showcasing a mature and competitive landscape.

3.Price and Performance Balance
While premium models deliver maximum range and acceleration, mid-range EVs now offer the best balance of cost, efficiency, and range, making electric mobility more accessible to mainstream buyers.

4.Charging Innovation and Convenience
Fast-charging technology continues to improve — some models now add 200–300 km of range in under 30 minutes, helping reduce charging anxiety and supporting long-distance travel.

5.Efficiency Influencers:
Weight and Design Data shows that lighter vehicles achieve better efficiency (Wh/km), while heavier, performance-focused models trade efficiency for power and speed — highlighting the design trade-off between range and acceleration.

Experience– Web Scraping & Data Analysis:

- Electric Vehicle data is obtain from match data from multiple websites using Python (BeautifulSoup, pandas).
- Cleaned and standardized data.
- Created visualizations showing its performance .
- Learned end-to-end data workflow: **scraping → cleaning → analysis → visualization.**

Challenges – Web Scraping & Data Analysis:

- Missing or mismatched match details.
- Handling duplicate and unclean data.
- Overcoming request blocks and slow scraping.
- Choosing the most meaningful visuals for insights.

INNOMATICS
RESEARCH LABS