Analytical Report on German Car Sales

05.05.2025

Emmanuella Oluwakemi Odiete
emmanuellaodiete04@gmail.com

Introduction

Cars are more than just machines, they're a reflection of our lives, they reflect innovation, efficiency, and consumer priorities. This report consists of insights from a dataset that features over 1,000 used cars from various brands and segments in the German market, including details like price, mileage, transmission, fuel type, and model year—offering valuable insights into market trends and buyer preferences.

Objective

The purpose of this project is to perform a comprehensive analysis of a used car dataset to identify trends, patterns, and relationships among various vehicle attributes. This analysis is aimed at supporting stakeholders—such as analysts, marketers, or pricing strategists—in understanding the factors that influence vehicle value and market behavior in the automotive sector.

Scope

This analytical report presents a structured examination of the used car market through four key dimensions, leveraging quantitative and qualitative insights to uncover market trends and valuation drivers:

1. **Brand Analysis**
    - Investigates the frequency and diversity of car brands within the dataset.
    - Explores the relationship between brand and pricing, mileage, and segment.
    - Identifies premium vs. economy brand positioning through pricing and specifications.
2. **Transmission Type Distribution**
    - Analyzes the distribution of automatic and manual transmissions.
    - Examines transmission preferences across brands and car segments.
    - Studies how transmission type correlates with price and mileage.
3. **Mileage Distribution**
    - Explores the spread of mileage across listed vehicles.
    - Detects potential outliers or inconsistencies in mileage reporting.
    - Evaluates the impact of mileage on car pricing.
4. **Price Evaluation**
    - Conducts a multivariable analysis of pricing.
    - Examines how attributes such as brand, mileage, fuel type, and condition affect vehicle cost.
    - Identifies possible market patterns, including undervalued or overvalued entries.

These focus areas allow for a multi-dimensional analysis of the dataset, providing a balanced view of both descriptive statistics and inferential relationships.

Dataset Overview

The dataset comprises structured information about a diverse set of used cars. Each row represents a unique vehicle listing, and each column describes an attribute or characteristic of the vehicle. The dataset includes a total of 8970 records and 10 relevant columns. Below is a brief description of each column:

I. Unnamed: 0: An auto-generated index column, not relevant for analysis.
II. Name: The name of the car, typically including the make and model (e.g., Volkswagen Tiguan).
III. Price: The listing price of the vehicle in US dollars.
IV. Description: A textual field containing additional seller notes or car details.
V. Trassmition: Indicates the vehicle's transmission type (e.g., Manual or Automatic).
VI. Model: The year the car was manufactured or registered (e.g., 2020, 2021).
VII. Milage: The total distance the car has traveled, measured in kilometers.
VIII. Fuell: The type of fuel the car uses (e.g., Petrol, Diesel, Hybrid, Electric).
IX. Brand: The manufacturer or brand of the vehicle (e.g., Volkswagen, Volvo).
X. Segment: The category or class of the car (e.g., Low-Range, Premium, Exotic), reflecting its market position.

Data Preprocessing & Cleaning

A. Initial Assessment

Upon review, the raw dataset exhibited several quality issues and redundancies that required attention before analysis. These issues included:

- Irrelevant columns that did not contribute to the analysis (e.g., internal identifiers or names).
- Minor typographical errors in column headers.
- Inconsistent or limited data entries for certain brands, potentially skewing results opportunities for column renaming to improve semantic clarity and accuracy.

A methodical cleaning process was implemented to ensure data consistency and usability.

B. Cleaning Steps

The following cleaning steps were conducted:

- **Identification of Missing Values:** These functions were used to detect the number of missing values in the dataset and ultimately removed.

```
cars.isnull().sum()
#check for missing value
```

```
Unnamed: 0      0
name            0
price           0
description    27
trassmition     0
model           0
milage          0
fuel            0
Brand           0
Segment         0
dtype: int64
```

```
cars=cars.dropna()
#drop empty cells
```

- **Column Elimination:** The columns 'name' and 'S/N' were dropped.

```
del cars["Unnamed: 0"]
#drop s/n column
```

```
del cars["name"]
#drop the name column
```

- **Duplicated Value:** These functions were used to identify and drop duplicate rows.

```
cars.duplicated().sum()
#recheck for duplicate
```

```
1006
```

```
cars=cars.drop_duplicates(keep="first")
#drop duplicate but keep the first occurrence
```

- **Standardizing Column Names:** A misspelled column label, “transsmition”, was corrected to “transmission” to align with standard terminology and improve readability.

```
cars.rename(columns={"trassmition":"transmission"}, inplace=True)
#renamed trassmition column to transmission
```

- **Selective Row Removal:** Rows where the value in the brand column was "seat" were removed. This decision was made to maintain a consistent brand representation and eliminate low-frequency outliers that could disproportionately influence brand-level summaries.

```
cars.loc[cars["Brand"]=="Seat"]
#view rows where the Brand is seat
```

	price	description	transmission	model	milage	fuel	Brand	Segment
0	26118	Sportstourer 1.4 E-hybrid Fr Led Acc Navi	Automatic	2021	52080	Hybrid	Seat	Low-Range
1008	43768	Fr 2.0 Tdi 150 Dsg Nav Shz Kessy Ehk ...	Automatic	2024	10	Diesel	Seat	Low-Range
1009	19904	St "fr"_climatr_sitzh_led_sw_abstandstemp..	Manual	2019	42500	Petrol	Seat	Low-Range
1010	36682	Style 1.5 Tsi 150 Dsg Led Nav Kam Pdc...	Automatic	2024	10	Petrol	Seat	Low-Range
1011	13802	75ps Chic 4-tur Klima+sihzg+pdc+hu7/26+alu+t	Manual	2019	28850	Petrol	Seat	Low-Range
...	...	...	...	...	...	...	...	...
8896	26197	Fr 1.5 Tsi Opf*automatik*ruckfahrkamera...	Automatic	2024	16400	Petrol	Seat	Low-Range
8897	19071	1.0 Tsi Fr Digital-tacho R.kamera Keyless	Manual	2019	57000	Petrol	Seat	Low-Range
8898	24547	Style 1.0 Tsi Dsg Lm Led W-paket Navi	Automatic	2024	9986	Petrol	Seat	Low-Range

```
cars = cars.loc[cars["Brand"] != "Seat"]
#drop rows where brand is seat by reassigning columns where brand is not seat
```

- **Validation of Data Types and Structure:** The dataset was reviewed to confirm that the data types in numerical columns (e.g., price, mileage) and categorical columns (e.g., brand, transmission, fuel_type) were correctly represented. No major missing data was observed in critical columns, indicating a relatively complete dataset suitable for analysis.

C. Resulting Dataset Summary

Following the above cleaning steps, the final dataset used for analysis contains 7,784 rows and 8 columns. The cleaned columns are as follows:

- Price
- Description
- Transmission

- Model
- Mileage
- Fuel
- Brand
- Segment

```
cars
```

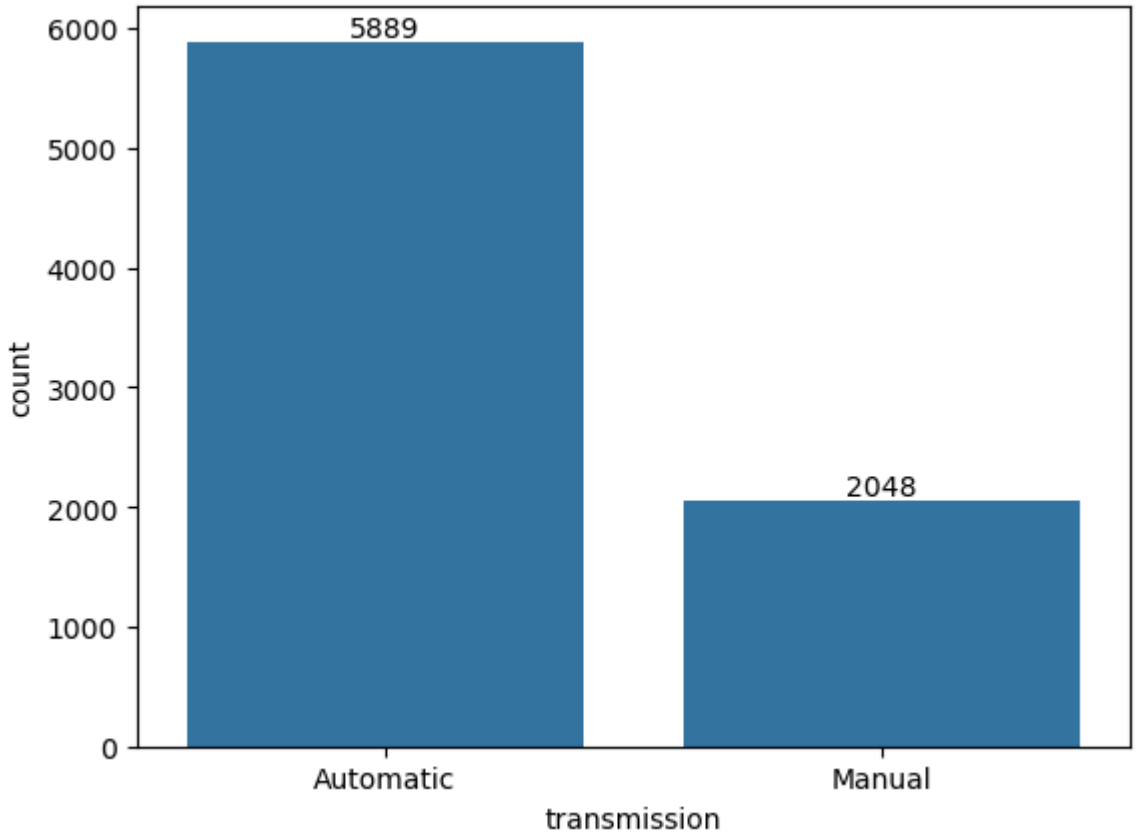
	price	description	transmission	model	milage	fuel	Brand	Segment
1	27777	T6 Awd Recharge Plug-in Momentum Pro Ahk	Automatic	2020	107217	Hybrid	Volvo	Premium
2	20392	2.0 Tdi Navi/sound/bmt,comfortline,1.hand	Manual	2017	123000	Diesel	Volkswagen	Premium
3	30967	2.0 Tdi Highline R-line 4mo. Led+navi+acc	Automatic	2019	68100	Diesel	Volkswagen	Premium
4	22892	Join Start-stopp 2hand Ahk Garantie	Manual	2018	105958	Diesel	Volkswagen	Premium
5	26760	2.0 Tdi/matrix/4motion/ambiente/200ps	Automatic	2021	127000	Diesel	Volkswagen	Premium
...	...	...	...	...	...	...	...	...
8965	95691	E-hybrid	Automatic	2022	14800	Hybrid	Porsche	Ultra-Luxury/Exotic
8966	47533	Diesel / Sportdesign / Voll / 1.hd.	Automatic	2016	99870	Diesel	Porsche	Ultra-Luxury/Exotic
8967	84648	Coupe E-hybrid*head-up*360°cam*panorama*	Automatic	2021	36500	Hybrid	Porsche	Ultra-Luxury/Exotic

Exploratory Data Analysis (EDA) & Data Visualization

Through statistical summaries and visual exploration, this section characterizes the dataset's composition and identifies significant patterns to guide deeper analysis and strategic decision making.

Transmission Mode Analysis

Automatic cars significantly outnumber manual cars in the dataset, with 5,889 automatic and 2,048 manual vehicles recorded. This translates to approximately 74% of the cars being automatic, indicating a strong consumer preference for convenience and ease of driving, especially in high-traffic or urban environments. In contrast, manual cars, which make up about 26%, may attract performance-oriented drivers or those seeking more affordable vehicle options.



Used Cars in 2024

To identify pre-owned vehicles in the year 2024, we established a selection criterion of vehicles with over 100 miles on the odometer. This threshold accounts for typical test-driving and dealership movement while excluding brand-new, undriven inventory. Our analysis reveals 921 qualifying vehicles in the dataset that meet these parameters.

```
used_cars = cars.loc[(cars['model'] == 2024) & (cars['milage'] > 100)]
used_cars
#rows of used cars in 2024 i.e mileage > 100
```

	price	description	transmission	model	milage	fuel	Brand	Segment
7	38297	R-line Pano Kam Navi Iq-light 19'' Black	Automatic	2024	6123	Petrol	Volkswagen	Premium
18	95137	S*acc,ahk,bose,pano,tempolim.,14-wege,360°	Automatic	2024	14300	Petrol	Porsche	Ultra-Luxury/Exotic
20	34797	Pure Tech Gt Eat *7-sitze*acc*led*upe:49	Automatic	2024	8231	Petrol	Peugeot	Mid-Range
78	60937	Sportback 2xs-line Pano Ahk360° Kamera	Automatic	2024	7130	Petrol	Audi	Luxury
88	74427	Avant 40 Tdi Quattro Edition One Lp: 82.515,-	Automatic	2024	4000	Diesel	Audi	Luxury
...	...	...	...	...	...	...	...	...
8898	24547	Style 1.0 Tsi Dsg Lm Led W-paket Navi	Automatic	2024	9986	Petrol	Seat	Low-Range
8923	50417	Techno 220 Long Range	Automatic	2024	800	Not specified	Renault	Unknown
8924	47704	E-tech Esprit Alpine 220 Aktion Gewerbe*	Manual	2024	2500	Not specified	Renault	Unknown
8925	51704	E-tech Esprit Alpine 220 Long Range	Manual	2024	2500	Not specified	Renault	Unknown
8948	140927	4s ''21-zoll Innodrive Headup Hd-matrix L	Automatic	2024	6900	Not specified	Porsche	Ultra-Luxury/Exotic

921 rows × 8 columns

Brand & Price Analysis

i. Brand with the Highest Total Price in 2024: Analysis reveals Volkswagen as the premium revenue-generating brand in 2024's used car market, commanding the highest total sales value.

```
cars.loc[cars["model"]==2024].groupby("Brand")[["price"]].sum().sort_values("price", ascending=False).head(1)
#brand with the highest total price in 2024
```

	price
Brand	
Volkswagen	23547924

ii. Unique Car Brands: The dataset features 51 distinct automotive brands, representing the full spectrum of market positioning—from high-volume mass-market manufacturers to exclusive luxury marques. This breadth of brand representation:

1. Highlights market segmentation, with clear differentiation between value driven and premium buyer preferences.
2. Indicates competitive diversity, demonstrating that multiple brands thrive across varied price points and consumer demographics.
3. Suggests stable demand across market tiers, reinforcing the used-car sector's resilience.

```
cars["Brand"].nunique()
#count of unique brands of cars
```

```
51
```

```
cars["Brand"].unique()
#the different unique brands of cars
```

```
array(['Volvo', 'Volkswagen', 'Toyota', 'Renault', 'Porsche', 'Peugeot',
       'Mini', 'Mercedes-benz', 'Hyundai', 'Fiat', 'Dacia', 'Citroen',
       'Bmw', 'Audi', 'Nissan', 'Land', 'Suzuki', 'Smart', 'Tesla',
       'Skoda', 'Ssangyong', 'Opel', 'Mitsubishi', 'Mazda', 'Maserati',
       'Lotus', 'Lexus', 'Lamborghini', 'Kia', 'Jeep', 'Jaguar', 'Honda',
       'Ford', 'Ferrari', 'Ds', 'Dodge', 'Cupra', 'Chevrolet', 'Cadillac',
       'Bentley', 'Aston', 'Alpina', 'Alfa', 'Abarth', 'Rolls-royce',
       'Polestar', 'Morgan', 'Mg', 'Subaru', 'Mclaren', 'Lynk'],
      dtype=object)
```

The presence of both mainstream and luxury brands enables cross-segment benchmarking, revealing pricing trends, depreciation patterns, and consumer loyalty metrics critical for strategic decision-making.

iii. Brand with the Highest Total Price (Overall): Brand dominance is consistent even beyond 2024, where Volkswagen holds the highest total price. This reinforces its position as a key player in the market.

```
cars.groupby("Brand")[["price"]].sum().sort_values("price", ascending=False).head(1)
#car brand with the highest total price
```

Brand	price
Volkswagen	54365163

Mileage Analysis

Mileage is a strong indicator of vehicle usage and affects both pricing and buyer perception. For clearer insights, cars were grouped into defined mileage brackets. This classification allowed for a straightforward visualization of how mileage is distributed across the dataset.

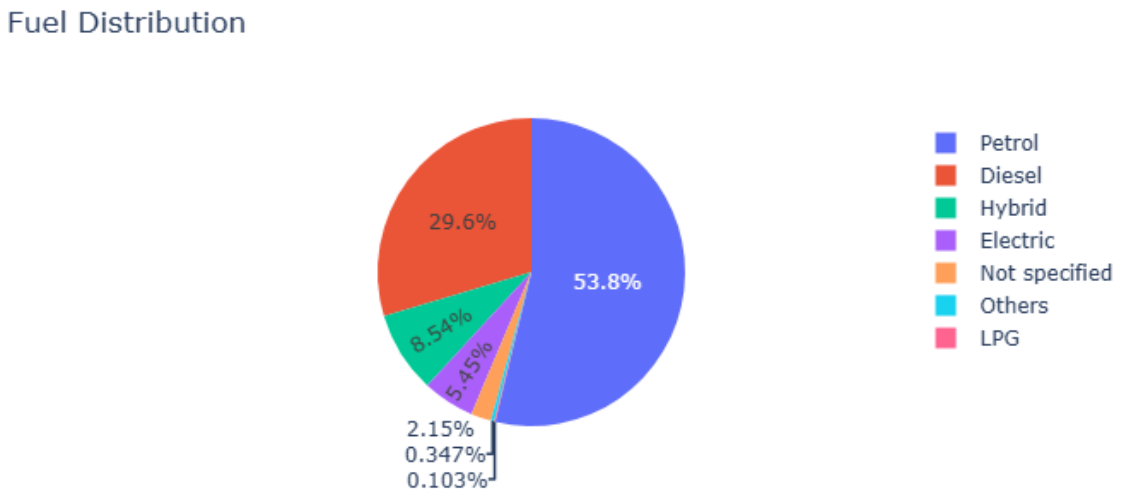
Most vehicles fall within the 20k–50k km and 50k–100k km ranges, suggesting moderate usage typical of resold cars. Interestingly, the 0–5k km bracket also has a high count, suggesting a notable presence of nearly new or minimally driven cars, possibly from test drives or short-term usage. Meanwhile, 5k–20k km and 100k–150k km classes appear less common. This distribution highlights a healthy spread across usage levels, with demand or availability strongest for moderately used vehicles.

```python
def milage_class(x):
    if x <= 5000:
        return "0-5k"
    elif 5000 < x <= 20000:
        return "5k-20k"
    elif 20000 < x <= 50000:
        return "20k-50k"
    elif 50000 < x <= 100000:
        return "50k-100k"
    elif 100000 < x <= 150000:
        return "100k-150k"
    else:
        return "Above 150k"
#assign milage class based on range
```

```python
cars["milage"].apply(milage_class)
# Apply the mileage classification to the 'milage' column to assign each car to a mileage class
```

Fuel Type Distribution

Petrol-powered cars lead the dataset at 53.8% of listings, followed by diesel vehicles at 29.6%. Alternative fuel types—hybrid (8.54%) and electric (5.45%)—comprise a much smaller share, indicating that greener technologies have yet to gain widespread traction. This distribution mirrors current fuel preferences: traditional options still prevail, while the transition to sustainability remains gradual.



Segment-Wise Total Price Analysis

Car segments show distinct value and price dynamics. The Luxury segment dominates, reflecting high demand for premium features and brand prestige. Next is the Premium segment, which attracts buyers seeking a blend of performance and upscale amenities. Low-Range and Mid Range cars each hold a similar share, indicating steady interest from budget-conscious and middle-income consumers.

Although the Ultra-Luxury/Exotic category is niche, it commands a significant total value, revealing a market for high-end collectors and affluent buyers. The smallest contribution comes from the Unknown segment, likely due to unclassified or incomplete data. Overall, luxury and premium vehicles account for the largest portion of the market's total value.

```
cars.groupby('Segment')[['price']].sum().sort_values("price",ascending=False)
#total price of each segment
```

Segment	price
Luxury	99101509
Premium	74835174
Low-Range	44807440
Mid-Range	44729131
Ultra-Luxury/Exotic	16938169
Unknown	6720781

Key Insights

1. **Fuel Types:** The dataset includes various fuel types such as Hybrid, Electric, LPG, and Diesel, reflecting both market diversity and a gradual transition to greener vehicles. Petrol is the most common, followed by Diesel and Hybrid, indicating a slowly evolving shift toward alternative fuels.
2. **Segment Distribution:** Cars are categorized into tiers like Low-Range, Mid-Range, Premium, and Ultra-Luxury/Exotic, highlighting a mix of affordability and luxury. Premium dominates with 2,307 listings, followed by other classes—suggesting a strong market focus on high-value vehicles.
3. **Model Year Range:** The model years range from 2017 to 2024, supporting time-based trend analysis. Most listings are from 2020 to 2024, indicating a predominantly modern used-car inventory.
4. **Mileage Distribution:** Mileage values vary significantly, providing insights into vehicle usage and depreciation trends across the market.
5. **Brands Dominance:** Brands like Volkswagen, Volvo, and Seat are prominent, with Volkswagen appearing most frequently (1,836 listings), reflecting either strong domestic loyalty or market saturation.
6. **Transmission Trends:** Both Automatic and Manual transmissions are well represented, allowing for comparative analysis; automatic dominates with over 74% of listings.
7. **Price Range:** The average price is approximately $36,287, with listings ranging from $10,013 to a striking $698,540—possibly indicating outliers such as luxury models or mispriced entries.
8. **Mileage Variation Reflects Diverse Usage:** The median mileage is around 27,073 km, with a wide range from 0 to 150,000 km—useful for analyzing depreciation and condition-based pricing strategies.

Recommendations

1. **Fix Brand-Segment Inconsistencies**

Some budget brands are labelled under the Premium segment and this may reduce trustworthiness unless verified. Audit and reclassify brands incorrectly assigned to premium segments to align with their actual market positioning. This ensures data integrity and avoids misleading buyers. Clear brand-segment mapping builds credibility and improves user trust in the platform.

2. **Automate Segment Classification Using Car Attributes**

Replace manual segment tagging with a rule-based or machine-learning system that uses price, model, and features to assign segments. This improves classification accuracy and helps customers better understand what to expect. It also prevents mislabeling that could confuse or deter users.

3. **Adjust Pricing Based on Mileage and Vehicle Age**

Vehicles with mileages over 100,000 km are priced comparably to low-mileage ones in some cases, suggesting pricing inconsistencies that can drive away informed buyers. Implement a smart pricing tool that considers mileage and model year to suggest appropriate price ranges for sellers. This prevents high-mileage cars from being overpriced and ensures buyers feel confident in value.

4. **Boost Visibility of Alternative Fuel Vehicles**

While electric and hybrid options exist in the data, they are underrepresented, potentially missing out on eco-conscious buyer segments. Create dedicated filters, campaigns, and educational content for electric and hybrid cars to attract sustainability-minded buyers. These efforts can increase engagement and expand the market for alternative fuel listings. Highlight long-term cost savings to influence buyer interest.

5. **Improve Brand Diversity in Listings**

Volkswagen frequently appears in listings, potentially overshadowing other brands. Implement personalized recommendations to diversify visible listings. Consider offering incentives or discounted listings to underrepresented brands for better balance.

6. **Emphasize Unique Selling Points in Listings**

Cars with similar specs are hard to distinguish in listings. Enable listings to showcase standout features (e.g., sunroof, Apple CarPlay) and visually highlight them for easier comparison. This makes it easier for buyers to differentiate similar vehicles. Use recommender systems to surface best-value options based on user preferences.

Limitations

This analysis is based on a single-timeframe dataset, primarily from 2024, which limits the ability to observe trends over time. The removal of low-frequency brands, such as Seat, may have improved consistency but reduced brand diversity. Additionally, the dataset lacks detailed vehicle features like safety ratings or optional upgrades, which could have enhanced the depth of pricing and value analysis.