ReDI School of Digital Integration

Data Analytics Final Project

Used Car Market Analysis- German Dataset

We use tech to connect human potential and opportunity with dignity & humility

Akinsope Idowu

PROBLEM STATEMENT

- Analyze the German used car market to understand factors influencing vehicle prices.
- Identify how brand, mileage, horsepower, year, and fuel type affect pricing.
- Support buyers, sellers, and dealerships with data-driven insights.

Data Cleaning & Preprocessing

- Removed duplicates and renamed columns.
- Handled missing values and corrected inconsistent entries.
- Standardized categorical variables (make, model, fuel, gear).
- Filtered unrealistic price, mileage, and horsepower values.

Exploratory Data Analysis (EDA)

- Explored distributions of key features (price, mileage, horsepower).
- Analyzed brand-level price differences and market composition.
- Examined correlations among numerical variables.
- Identified depreciation patterns and fuel-type trends.

DATA CLEANING AND PREPROCESSING

```
#Convert price from Int to Float
cars['price'] = cars['price'].astype(float)

# Remove duplicates from dataset
cars= cars.drop_duplicates()

# Remove extreme outliers from prices or mileage
cars = cars[(cars['price'] > 200) & (cars['price'] < 300000)]
cars = cars[cars['mileage'] < 500000]
```

```
#categorize columns into Numerical(numbers) and Categorical (non-numbers) Data
numerical_cols = ['price', 'mileage', 'horsepower', 'year']
categorical_cols = ['make', 'model', 'gear', 'fuel']
```

```
cars = cars.rename(columns={'hp': 'horsepower'})
```

```
#Fill nan values
cars['model'] = cars['model'].fillna('Unknown')
cars['gear'] = cars['gear'].fillna(cars['gear'].mode()[0])
cars['horsepower'] = cars['horsepower'].fillna(cars['horsepower'].median())
```


Total Cars
44,236
Average Price
€16,387
Average Mileage
72,117 km
Average Horsepower
133 HP
Most Common Make
Volkswagen
Most Common Model
Golf

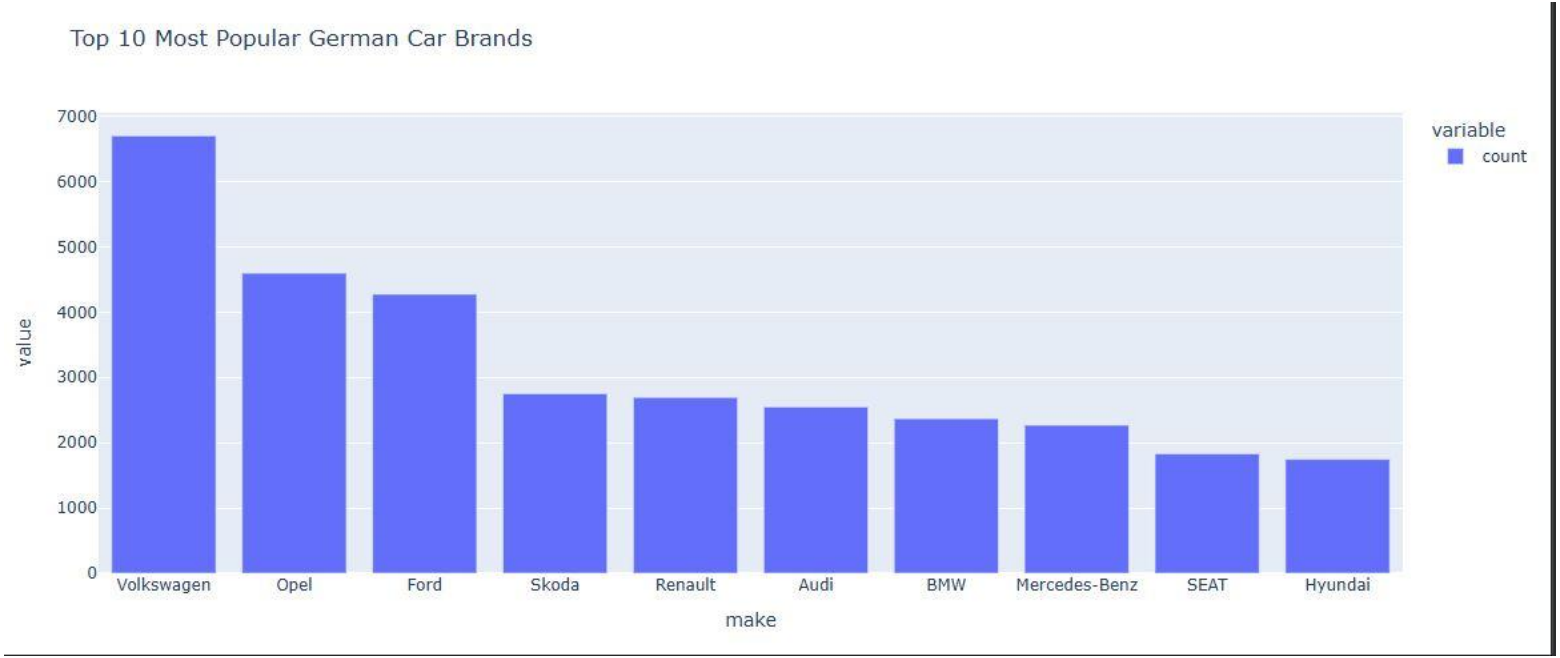
FEATURES & FUNCTIONALITY

Data Cleaning	Fixes raw dataset, handles missing/incorrect values
KPIs	Summarizes key numerical & categorical insights
Bar Charts	Shows popular makes/models
Boxplots	Shows distribution across top brands
Histograms	Shows entire dataset distribution
Pie Chart	Shows gear-type proportions

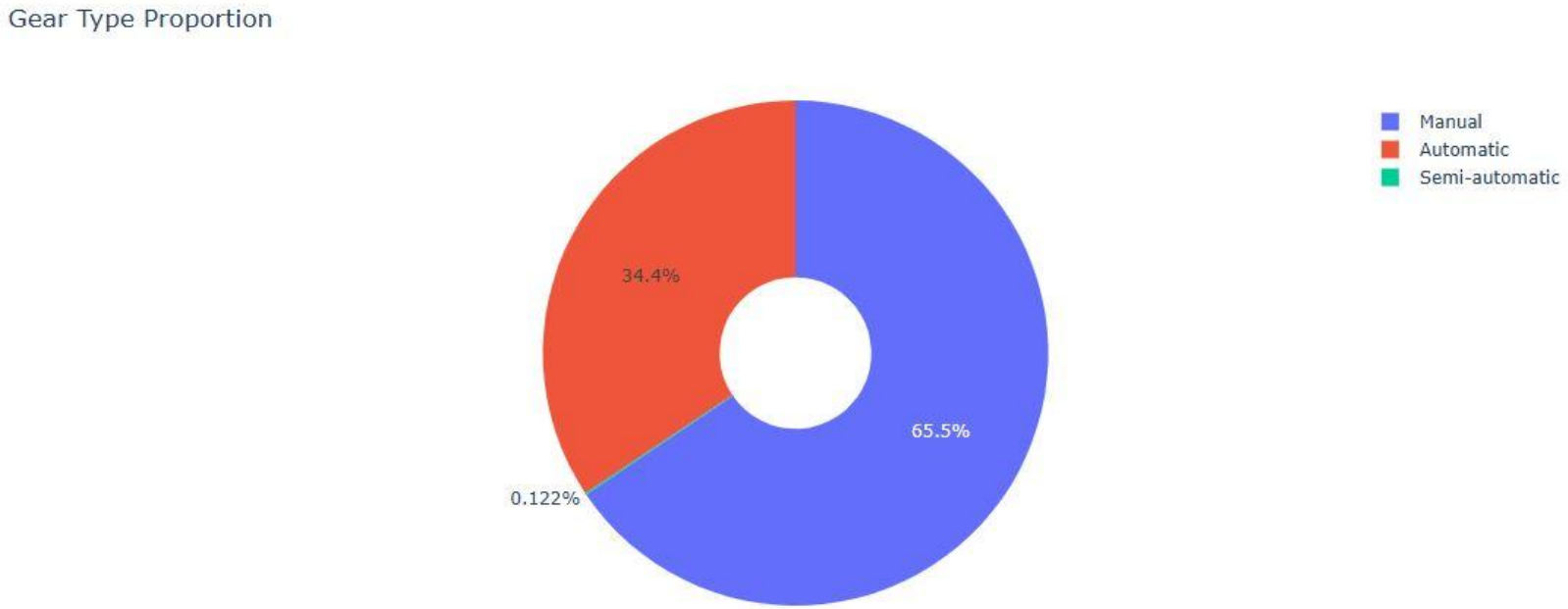
FEATURES & FUNCTIONALITY

Scatter Plots	Shows relationships between key attributes
Correlation Heatmap	Detects patterns and correlations
Widgets	Enables interactive variable selection
Dashboard Filtering	Allows brand-specific analysis
CSV Export	Downloads cleaned .csv file to local drive

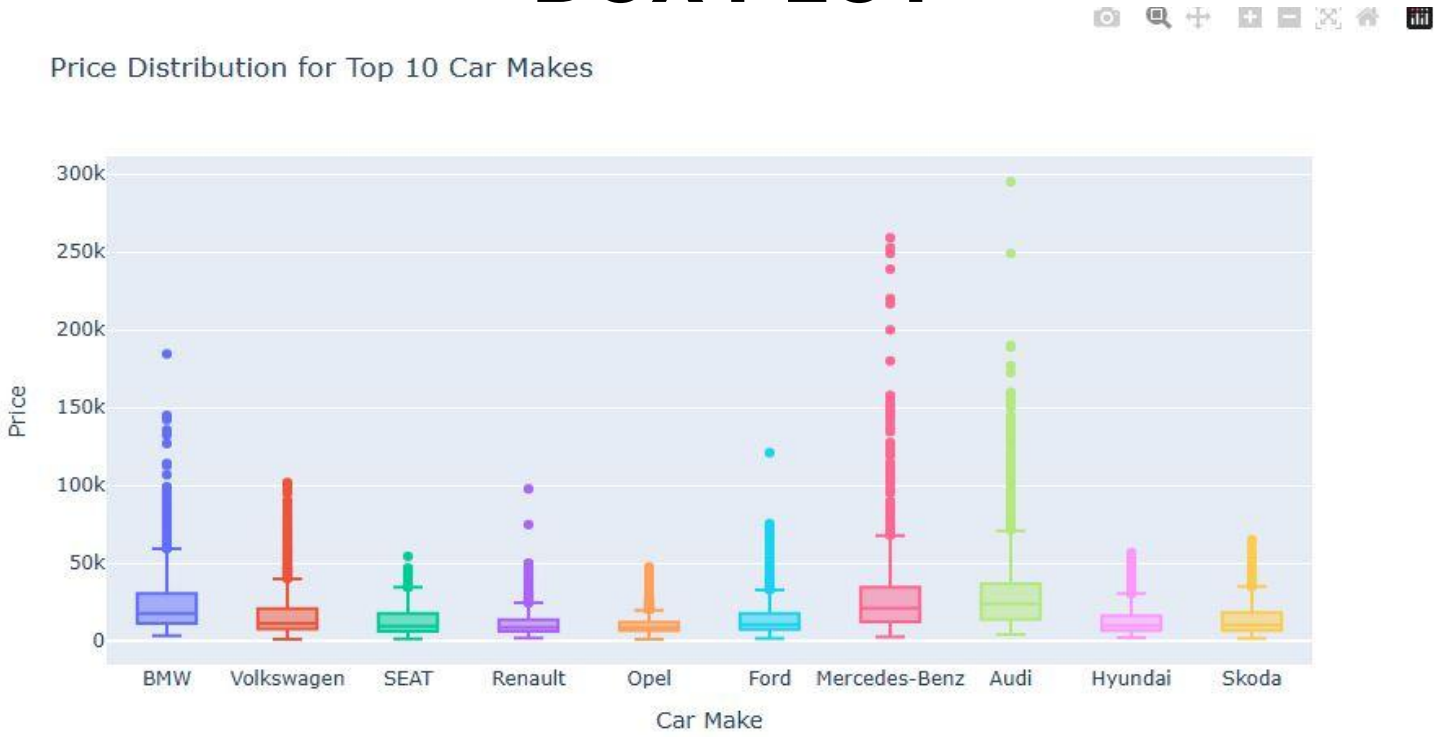
BAR CHART


Top 10 Most Popular German Car Brands
7000
6000
5000
4000
3000
2000
1000
0
value
Volkswagen
Opel
Ford
Skoda
Renault
Audi
BMW
Mercedes-Benz
SEAT
Hyundai
make
variable
count

PIE CHART


Gear Type Proportion
34.4%
65.5%
0.122%
Manual
Automatic
Semi-automatic

BOX PLOT


Price Distribution for Top 10 Car Makes
300k
250k
200k
150k
100k
50k
0
Price
BMW
Volkswagen
SEAT
Renault
Opel
Ford
Mercedes-Benz
Audi
Hyundai
Skoda
Car Make

SCATTER PLOT

Ethical Considerations & Limitations

- Potential sampling bias from a single platform (AutoScout24).
- Lack of key features such as location, accident history, and car condition.
- Risk of misinterpretation if insights are used for unfair pricing.
- Data quality issues: outliers, missing values, categorical inconsistencies.

CORELATION


Correlation heatmap (numeric features)
price
mileage
horsepower
makeYear
1
−0.3455217
0.7990036
0.4697259
−0.3455217
1
−0.02016765
−0.6764372
0.7990036
−0.02016765
1
.1815353
0.4697259
−0.6764372
0.1815353
1
x: horsepower
y: horsepower
color: 1
price
mileage
horsepower
makeYear
1
0.8
0.6
0.4
0.2
0
−0.2
−0.4
−0.6

Results & Real-World Impact

1. Identification of High-Demand Brands & Models

- KPI cards highlight the **most sold models** and **most active years** in the dataset.
- German brands dominate the listings, especially Audi, BMW, and Mercedes.

Real-World Impact:
Sellers and marketplaces can prioritize buying/selling cars that have consistently high demand and turnover.

2. Better Understanding of Car Value Retention

- Scatter plots highlight how some brands retain value even with higher mileage.
- Buyers can use these insights to make smarter purchase decisions (more value per euro).

Real-World Impact:
Individual buyers can choose brands that depreciate slower and make cost-effective purchases.

Results & Real-World Impact

3. Dashboard-Style Insights Through KPIs & Widgets

- Interactive widgets help users compare price, mileage, and horsepower distributions by brand.
- Categorical KPI cards condense important information into quick summaries.

Real-World Impact:
Businesses can embed these features into dashboards for real-time monitoring of market trends and stock performance.

Challenges & Solutions

1.Outliers Affecting Distribution

Challenge:

Extreme values in price, mileage, and horsepower skewed the plots.

Solution:

Removed unrealistic values (e.g., mileage > 400,000, horsepower > 1200) to improve accuracy of visualizations.

2.Missing Data in Model, Gear, and Horsepower

Challenge:

Some rows had missing values for key fields, affecting both charts and KPIs.

Solution:

- Filled model with `'Unknown'`
- Filled gear with mode
- Replaced horsepower with median

This preserved data integrity and avoided unnecessary row removal.

Conclusions & Future Improvements

- Mileage, brand, horsepower, and year strongly affect pricing.

  Data provides meaningful insights into German market dynamics.
- Future work: add more features, expand region scope, include time-based trends.
- Develop a more interactive dashboard for real-time analysis.

Potential Uses

1. Customer Decision Making

Customers shopping for used cars can benefit from:

- Understanding which brands maintain value over time
- Identifying the best year/mileage combinations

2. Academic or Portfolio Use

Students and data analysts can use this project to demonstrate:

- Data cleaning skills, Exploratory data analysis, Dashboard creation, Interactive widgets and KPI design
- Visualization with Plotly

Thank you!