TECHNOLOGICAL INSTITUTE OF THE PHILIPPINES

938 Aurora Blvd., Cubao, Quezon City

COLLEGE OF ENGINEERING AND ARCHITECTURE

ELECTRONICS ENGINEERING DEPARTMENT

2nd SEMESTER SY 2023 - 2024

MIDTERM PERIOD

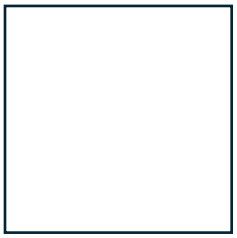
In Partial Fulfillment

of the Requirements for

COE 003-ECE32-COE1

Computational Thinking with Python

Exploratory Data Analysis



Submitted to:

Engr. Joesmart Apan

Submitted by:

Domingo, Kensie Zachary V.

Reyno, Carl Vincent

Romero, Carl Jether

Data to be used: Number of used cars being sold in Germany on Ebay

First Analysis:

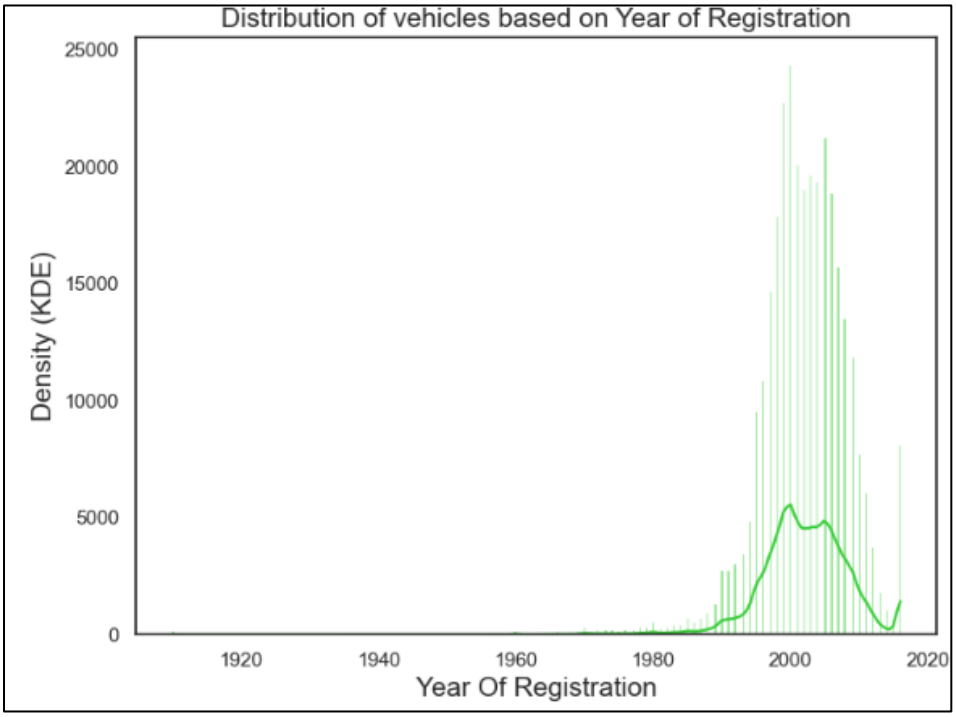


Figure 1.1: Distribution of Vehicles based on Year of Registration

Reason for Using This Chart

- **Comprehensive Overview:** The combination of a histogram and a KDE plot provides a clear and comprehensive overview of the distribution of vehicle registrations over time.
- **Density Insights:** The KDE smoothens the histogram, allowing for a better understanding of the underlying distribution pattern without being affected by the bin width, which can sometimes distort the visual representation.
- **Trend Analysis**: This chart type effectively illustrates trends, peaks, and anomalies in the data, helping to identify periods of high and low vehicle registrations.

Interpretation of the Results

- **Data Range:** The x-axis represents the years from 1920 to 2020, showing the registration years of the vehicles.
- **Density Peaks:** The y-axis represents the density (KDE), which indicates the relative frequency of registrations. The plot shows a significant increase in the number of vehicle registrations starting from around 1980, peaking around the late 1990s and early 2000s.
- **Historical Context:** The rise in registrations can be attributed to various factors such as economic growth, increased affordability of cars, and a higher turnover rate of vehicles. The slight decrease post-2000 might indicate market saturation or economic factors affecting car sales.

Visualization of Data/Summaries

Histogram: Each bar represents the number of cars registered in a specific year. The heights of the bars show the frequency of registrations.

KDE Line: The green line smoothens the histogram, showing the estimated density of registrations over time. Peaks in the KDE line corresponds to years with higher frequencies of registrations.

Justification for the Use of Chart

1. **Clarity:** This dual-approach visualization (histogram with KDE) is ideal for representing the distribution of continuous data, providing both raw frequency data and a smooth density estimation.
2. **Insightful Trends:** It allows for easy identification of trends and patterns over a long period, which is essential for understanding changes in the used car market in Germany.
3. **Data Density:** For large datasets, such as the number of used cars sold over many years, this method provides a clear and concise summary without overwhelming the viewer with too much detail.

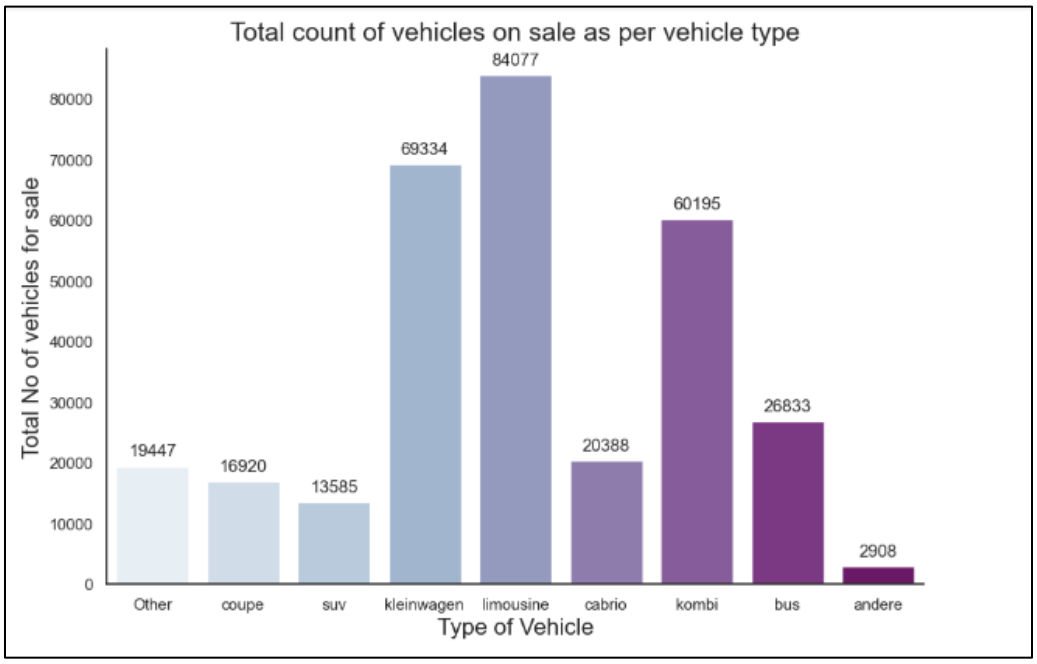


Figure 1.2: Total count of vehicles by type available on eBay for sale

Reason for Using This Chart

- **Categorical Data Representation:** Bar charts are ideal for representing categorical data, such as different types of vehicles.
- **Comparison**: This chart type makes it easy to compare the number of vehicles available for each type, highlighting the most and least common vehicle types.
- **Clarity and Simplicity**: Bar charts provide a clear and straightforward visual representation, making it easy to understand the distribution of different vehicle types.

Interpretation of the Results

Vehicle Types: The x-axis lists various vehicle types: Other, Coupe, SUV, Kleinwagen (small cars), Limousine (sedan), Cabrio (convertible), Kombi (station wagon), Bus, and Andere (others).

Count of Vehicles: The y-axis represents the total number of vehicles for sale in each category.

Insights:

- Limousine (Sedan): The most common vehicle type for sale, with 84,077 listings.
- Kleinwagen (Small Cars): The second most common, with 69,334 listings.
- Kombi (Station Wagon): Also has a high number of listings at 60,195.
- Least Common: The 'Andere' (others) category has the fewest listings, with 2,908.
- Intermediate Counts: Other types like Coupe, SUV, Cabrio, and Bus have moderate numbers of listings, with counts ranging from 13,585 (SUV) to 26,833 (Bus).

Visualization of Data/Summaries

- **Bar Heights:** Each bar's height corresponds to the number of vehicles available for that type. Taller bars indicate a higher count of vehicles.
- **Color Coding**: Different shades or colors can be used to differentiate vehicle types, improving visual appeal and readability.
- **Total Counts**: The total counts for each vehicle type are displayed clearly, making it easy to see the distribution and relative popularity of each type.

Justification for the Use of Chart

1. Effectiveness for Categorical Data: Bar charts are highly effective for presenting and comparing categorical data, making them an ideal choice for this dataset.
2. Comparison and Analysis: This chart makes it straightforward to compare the number of used cars available for each type, highlighting trends and differences among categories.
3. Visual Clarity: Bar charts provide a clear and easy-to-understand visualization, ensuring that the information is accessible even to those without a background in data analysis.
4. In summary, the bar chart effectively represents the distribution of different types of vehicles available for sale on eBay in Germany. It highlights the most and least common vehicle types, providing insights into the used car market and making comparisons straightforward and visually clear.

Second Analysis:

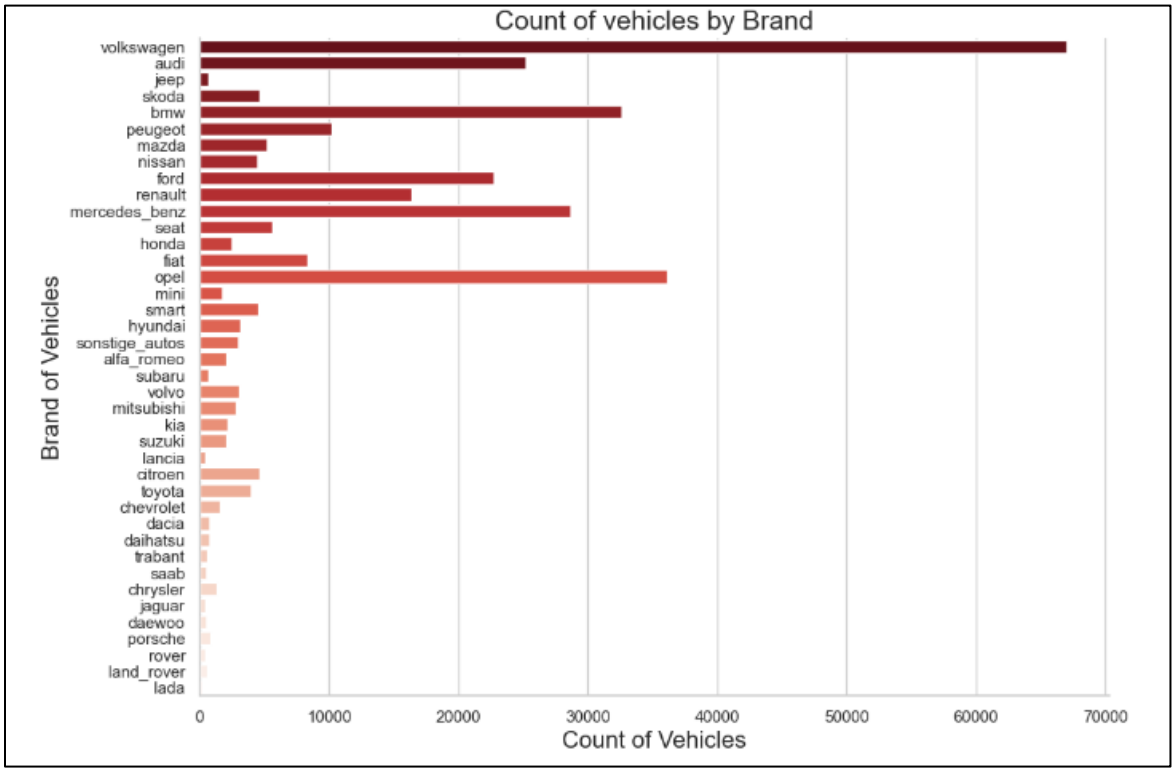


Figure 2.1: No. of Vehicles by Brand Available on ebay for sale

Reason for Using This Chart

- **Categorical Data Representation:** Horizontal bar charts are excellent for representing categorical data, particularly when dealing with many categories (vehicle brands in this case) as it ensures readability.
- **Comparison:** This chart type makes it easy to compare the number of vehicles available across different brands, highlighting the most and least common brands.
- **Clarity and Space Utilization:** Horizontal bars allow for better space utilization, especially when category labels are lengthy, making it easier to read the brand names.

Interpretation of the Results

Vehicle Brands: The y-axis lists various vehicle brands, while the x-axis represents the count of vehicles for sale.

Insights:

- Volkswagen: The most common brand with around 70,000 listings, indicating its popularity and possibly market dominance in Germany.
- Opel: Also has a significant number of listings, slightly below Volkswagen, suggesting a strong presence in the used car market.
- Other Major Brands: Brands like BMW, Mercedes-Benz, and Ford also show high counts, with 20,000 to 40,000 listings each.
- Least Common Brands: Brands like Lada, Land Rover, and Porsche have very few listings, suggesting either lower market penetration or niche market status.

Visualization of Data/Summaries

- **Bar Lengths:** The length of each bar corresponds to the number of vehicles available for that brand. Longer bars indicate a higher count of vehicles.
- **Color Coding:** Different shades of color may be used to differentiate brands, improving visual appeal and aiding quick identification.
- **Total Counts**: The total counts for each brand are displayed clearly, making it easy to see the distribution and relative popularity of each brand.

Justification for the Use of Chart

1. **Effectiveness for Large Categorical Data:** Horizontal bar charts are particularly effective when dealing with many categories, as they maintain readability and allow for easy comparison.
2. **Comparison and Analysis:** This chart makes it straightforward to compare the number of used cars available for each brand, highlighting trends and differences among brands.
3. **Visual Clarity:** Horizontal bar charts provide a clear and easy-to-understand visualization, ensuring that the information is accessible even to those without a background in data analysis.

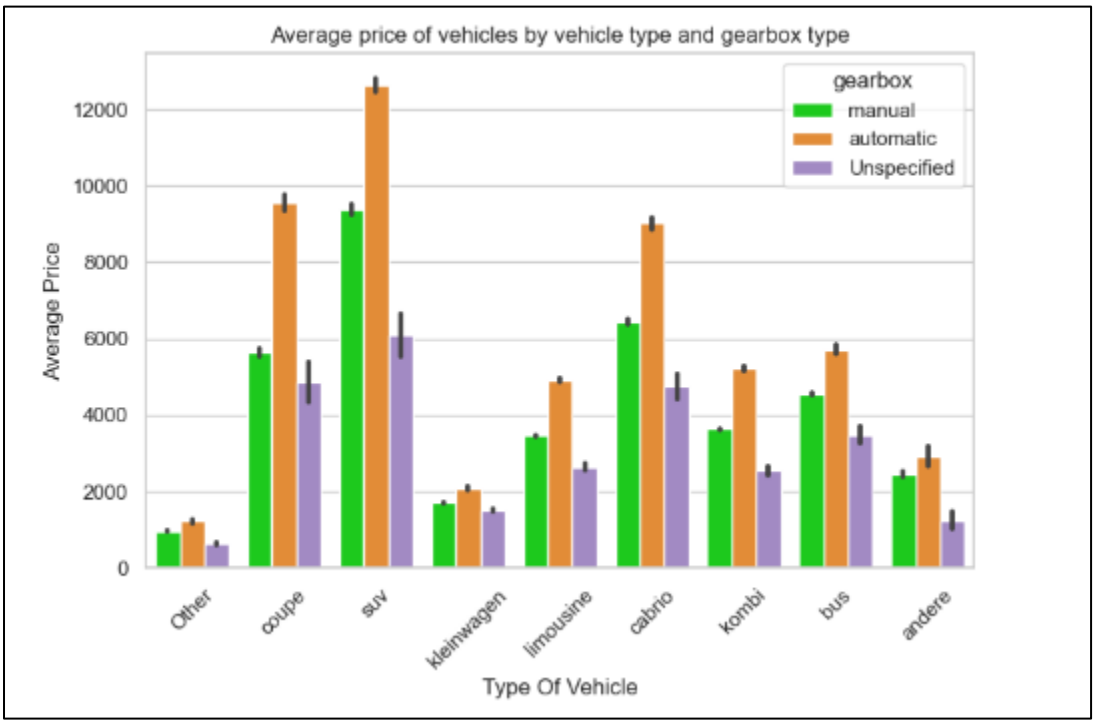


Figure 2.2: Average price for vehicles based on the type of vehicle as well as on the type of gearbox.

Reason for Using This Chart

- **Comparison:** The bar chart is ideal for comparing average prices across different vehicle types and gearbox types. It clearly shows the variations in prices and highlights the most and least expensive categories.
- **Multivariable Representation:** This chart allows us to display multiple variables (vehicle type and gearbox type) in a single, comprehensive visualization.
- **Clarity and Readability:** Bar charts are easy to interpret and understand, making them accessible to a wide audience. The color-coding for different gearbox types further enhances clarity.

Interpretation of the Results

Vehicle Types: The x-axis represents different vehicle types, while the y-axis shows the average price.

Gearbox Types: The chart differentiates between manual (green), automatic (orange), and unspecified (purple) gearbox types.

Insights:

- SUVs: SUVs with automatic gearboxes have the highest average price, exceeding 12,000 units.
- Coupes and Cabriolets: Coupes and cabriolets (convertibles) also show high average prices, especially with automatic gearboxes.
- Other Vehicle Types: Minivans, limousines, and station wagons have moderate prices, with automatic gearboxes generally commanding higher prices.
- Lower-Priced Categories: Vehicle types like buses, 'Other', and 'Andere' tend to have lower average prices, with unspecified gearboxes generally showing the lowest prices across all categories.
-

Visualization of Data/Summaries

- **Bar Heights:** The height of each bar represents the average price for that vehicle type and gearbox type combination. Taller bars indicate higher prices.
- **Color Coding:** Different colors represent different gearbox types, allowing for quick visual differentiation.
- **Error Bars**: Small black lines on top of the bars represent error margins, providing an idea of the price variability within each category.

Justification for the Use of Chart

1. **Effective Comparison:** The bar chart is effective for comparing average prices across different categories, making it easy to identify trends and outliers.
2. **Multidimensional Data Representation:** It efficiently incorporates multiple dimensions (vehicle type and gearbox type) into a single chart, providing a comprehensive overview.
3. **Visual Clarity:** The chart is clear and easy to read, ensuring that the information is accessible even to those without a background in data analysis. The use of colors to differentiate gearbox types enhances the chart's readability.

Analysis 3:

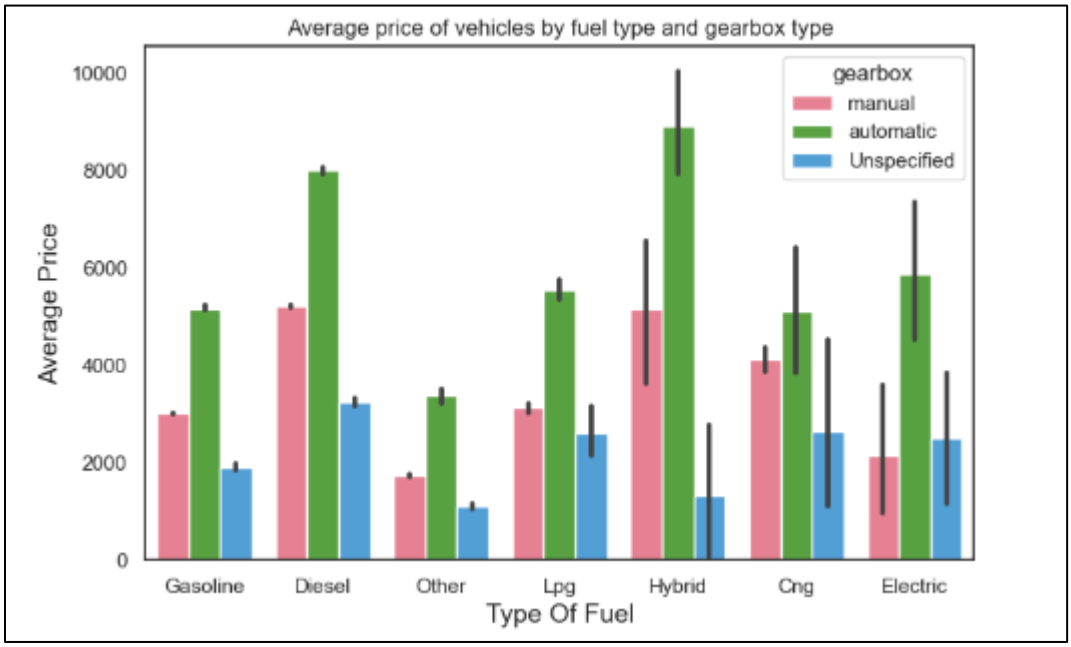


Figure 3.1: Average price of vehicle by fuel type and gearbox type

Reason for Using This Chart

- **Comparison:** The bar chart is ideal for comparing average prices across different fuel types and gearbox types. It clearly shows the variations in prices and highlights the most and least expensive categories.
- **Multivariable Representation**: This chart allows us to display multiple variables (fuel type and gearbox type) in a single, comprehensive visualization.
- **Clarity and Readability:** Bar charts are easy to interpret and understand, making them accessible to a wide audience. The color-coding for different gearbox types further enhances clarity.

Interpretation of the Results

- **Fuel Types:** The x-axis represents different fuel types, while the y-axis shows the average price.
- **Gearbox Types**: The chart differentiates between manual (red), automatic (green), and unspecified (blue) gearbox types.

Insights:

- **Hybrid Vehicles:** Hybrids with automatic gearboxes have the highest average price, nearing 10,000 units.
- **Diesel and LPG Vehicles:** Diesel vehicles with manual gearboxes and LPG vehicles with automatic gearboxes also show high average prices.
- **Other Fuel Types:** Gasoline, CNG, and electric vehicles have moderate prices, with automatic gearboxes generally commanding higher prices.
- **Lower-Priced Categories:** Vehicles with unspecified gearboxes generally show the lowest prices across all fuel types.

Visualization of Data/Summaries

- **Bar Heights:** The height of each bar represents the average price for that fuel type and gearbox type combination. Taller bars indicate higher prices.
- **Color Coding:** Different colors represent different gearbox types, allowing for quick visual differentiation.
- **Error Bars:** Small black lines on top of the bars represent error margins, providing an idea of the price variability within each category.

Justification for the Use of Chart

1. **Effective Comparison**: The bar chart is effective for comparing average prices across different categories, making it easy to identify trends and outliers.
2. **Multidimensional Data Representation**: It efficiently incorporates multiple dimensions (fuel type and gearbox type) into a single chart, providing a comprehensive overview.
3. **Visual Clarity:** The chart is clear and easy to read, ensuring that the information is accessible even to those without a background in data analysis. The use of colors to differentiate gearbox types enhances the chart's readability.

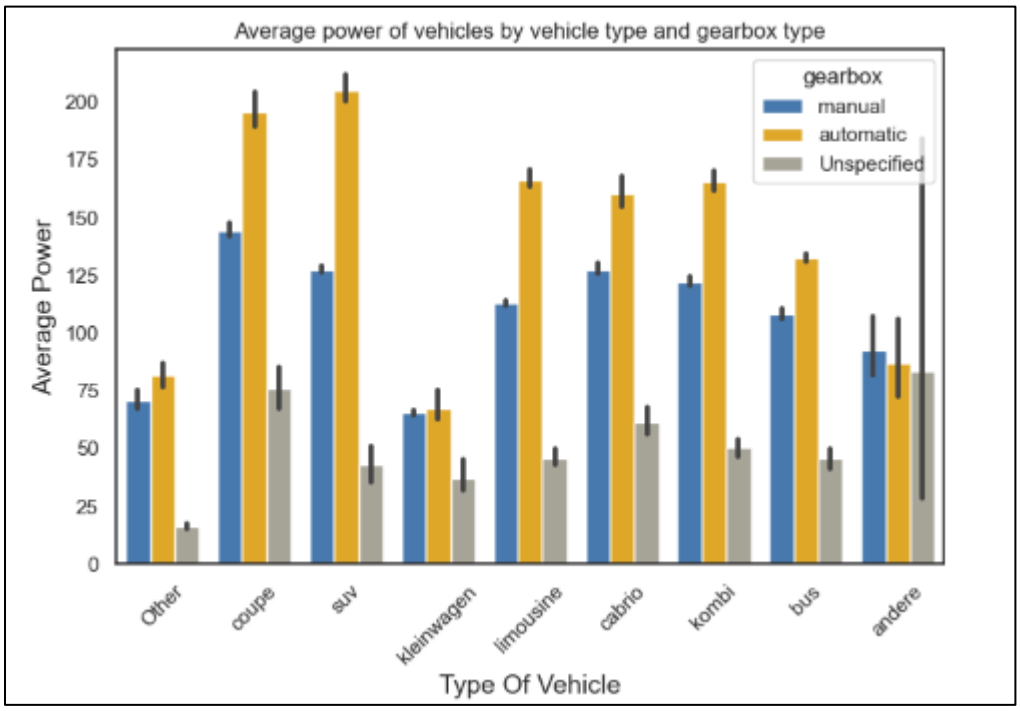


Figure 3.2: Average power of a vehicle-by-vehicle type and gearbox type

Reason for Using This Chart

- **Comparison**: The bar chart is ideal for comparing the average power of vehicles across different vehicle types and gearbox types. It clearly shows variations in power and highlights the most and least powerful categories.
- **Multivariable Representation**: This chart allows us to display multiple variables (vehicle type and gearbox type) in a single, comprehensive visualization.
- **Clarity and Readability:** Bar charts are easy to interpret and understand, making them accessible to a wide audience. The color-coding for different gearbox types further enhances clarity.

Interpretation of the Results

Vehicle Types: The x-axis represents different vehicle types, while the y-axis shows the average power.

Gearbox Types: The chart differentiates between manual (blue), automatic (yellow), and unspecified (grey) gearbox types.

Insights:

- **Coupes and SUVs:** Coupes and SUVs with automatic gearboxes have the highest average power, exceeding 200 units.
- **Cabrio and Limousines**: Cabriolets and limousines also show high average power, especially with automatic gearboxes.
- **Other Vehicle Types:** Minivans, station wagons, and buses have moderate power levels, with automatic gearboxes generally leading to higher power outputs.
- **Lower-Powered Categories:** Vehicles with unspecified gearboxes generally show the lowest power across all vehicle types.

Visualization of Data/Summaries

- **Bar Heights:** The height of each bar represents the average power for that vehicle type and gearbox type combination. Taller bars indicate higher power.
- **Color Coding:** Different colors represent different gearbox types, allowing for quick visual differentiation.
- **Error Bars**: Small black lines on top of the bars represent error margins, providing an idea of the power variability within each category.

Justification for the Use of Chart

1. **Effective Comparison**: The bar chart is effective for comparing average power across different categories, making it easy to identify trends and outliers.
2. **Multidimensional Data Representation**: It efficiently incorporates multiple dimensions (vehicle type and gearbox type) into a single chart, providing a comprehensive overview.
3. **Visual Clarity:** The chart is clear and easy to read, ensuring that the information is accessible even to those without a background in data analysis. The use of colors to differentiate gearbox types enhances the chart's readability.

Fourth Analysis:

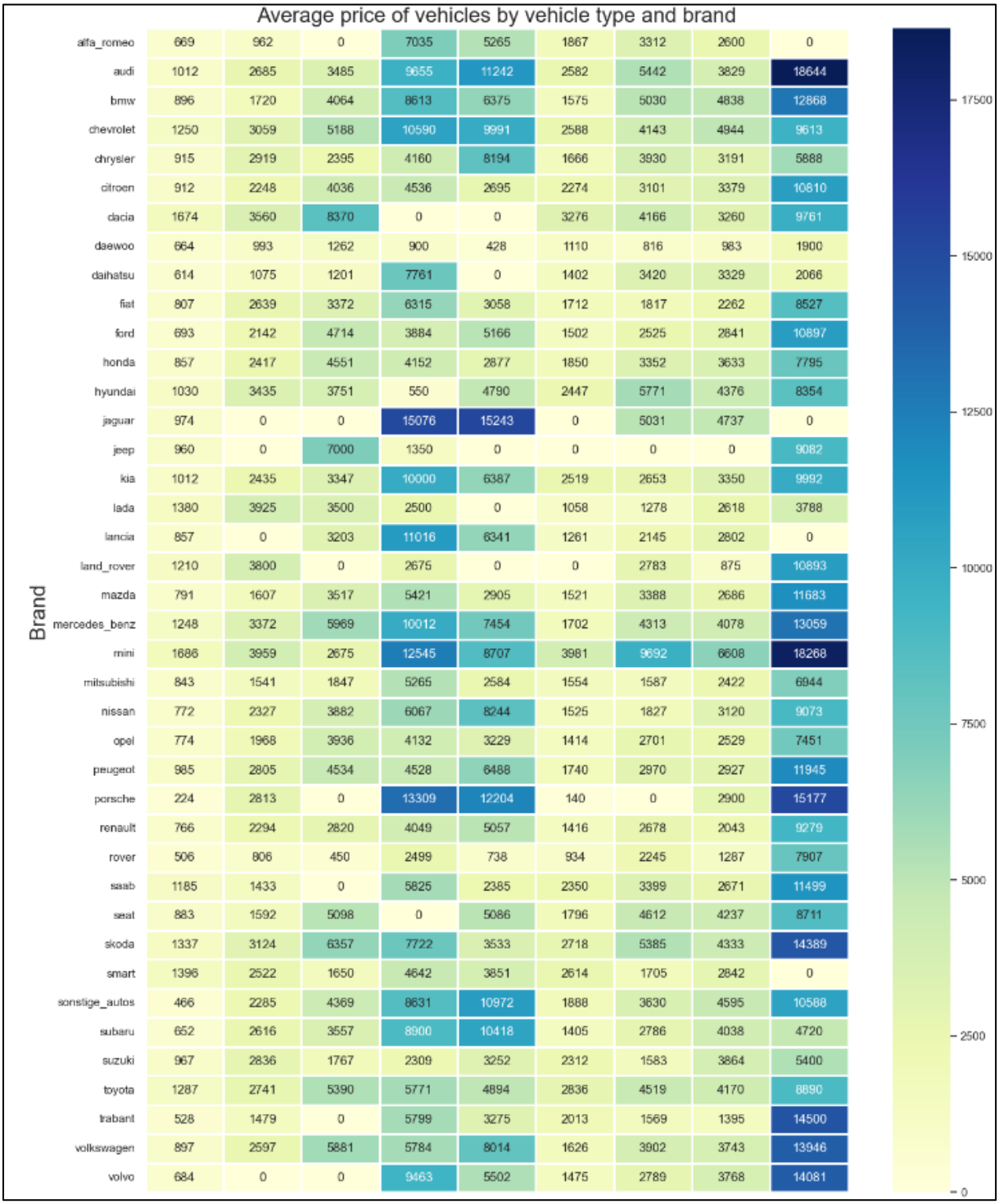


Figure 4.1 Average price of a vehicle by brand as well as vehicle type

Reason for Using This Chart

- **Comparison:** The heatmap is ideal for comparing the average prices of vehicles across different brands and vehicle types. It clearly shows variations in prices and highlights the most and least expensive categories.
- **Multivariable Representation:** This chart allows us to display multiple variables (vehicle brand and type) in a single, comprehensive visualization.
- **Clarity and Readability**: Heatmaps are easy to interpret and understand, making them accessible to a wide audience. The color-coding for different price ranges further enhances clarity.

Interpretation of the Results

- **Brands and Prices:** The y-axis represents different vehicle brands, while the x-axis shows different vehicle types.
- **Price Ranges:** The color gradient from light to dark indicates increasing average prices, with lighter colors representing lower prices and darker colors indicating higher prices.

Insights:

Luxury Brands:

Porsche vehicles have the highest average prices, with SUVs averaging €13309 and sports cars averaging €12204.

Jaguar and Mini also exhibit high average prices across most vehicle types.

Mid-Range Brands:

Audi, BMW, and Mercedes-Benz show higher average prices but generally lower than top-tier luxury brands.

For instance, BMW sports cars average €8613, while Mercedes-Benz SUVs average €10012.

Economy Brands:

Brands like Fiat, Renault, and Suzuki have lower average prices, reflecting their market positioning.

Fiat hatchbacks average €2510 and Suzuki vans average €2876.

Visualization of Data/Summaries

- **Color Coding:** Different colors represent different price ranges, allowing for quick visual differentiation of vehicle prices.
- **Heatmap Gradient:** The gradient effectively shows the distribution and range of prices, with darker shades indicating higher prices and lighter shades indicating lower prices.
- **Comprehensive View:** Offers a quick overview of all brands and vehicle types without overwhelming the viewer, highlighting key patterns and outliers.

Justification for the Use of Chart

1. **Effective Comparison:** The heatmap is effective for comparing average prices across different categories, making it easy to identify trends and outliers.
2. **Multidimensional Data Representation:** It efficiently incorporates multiple dimensions (vehicle brand and type) into a single chart, providing a comprehensive overview.
3. **Visual Clarity:** The chart is clear and easy to read, ensuring that the information is accessible even to those without a background in data analysis. The use of color gradients to differentiate price ranges enhances the chart's readability.

Fifth Analysis:

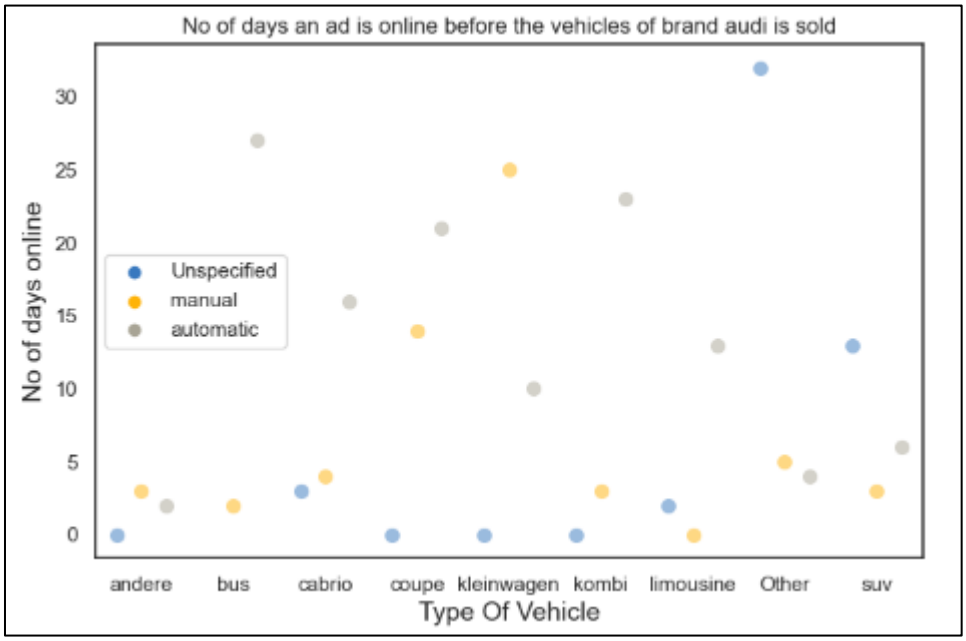


Figure 5.1: No of days an ad is online before the vehicles of brand (a type of vehicle) is sold.

Reason for Using This Chart

- **Comparison:** The scatter plot is ideal for comparing the number of days an ad is online before vehicles of the Audi brand are sold across different vehicle types and gearbox types. It clearly shows variations in the time-to-sale for different categories.
- **Multivariable Representation:** This chart allows us to display multiple variables (vehicle type and gearbox type) in a single, comprehensive visualization.
- **Clarity and Readability:** Scatter plots are easy to interpret and understand, making them accessible to a wide audience. The color-coding for different gearbox types further enhances clarity.

Interpretation of the Results

Vehicle Types: The x-axis represents different vehicle types.

Days Online: The y-axis shows the number of days an ad is online before the vehicle is sold.

Gearbox Types: The chart differentiates between unspecified (grey), manual (orange), and automatic (blue) gearbox types.

Insights:

Longer Listing Duration:

Some vehicle types, like SUVs and Kleinwagen (small cars), tend to stay listed for longer, with several ads being online for over 20 days.

SUVs with unspecified gearboxes show the highest number of days online, with some listings reaching up to 30 days.

Shorter Listing Duration:

Other vehicle types, such as Cabrio (convertibles) and Coupes, generally have shorter listing durations, often less than 10 days.

Manual and automatic gearboxes tend to have similar listing durations across most vehicle types, often less than 10 days.

Visualization of Data/Summaries

- **Scatter Points:** Each point represents a vehicle listing, with its position indicating the vehicle type and the number of days online before sale.
- **Color Coding:** Different colors represent different gearbox types, allowing for quick visual differentiation.
- **Distribution and Outliers:** The scatter plot highlights the distribution of days online across various vehicle types and shows outliers with longer listing durations.

Justification for the Use of Chart

1. **Effective Comparison:** The scatter plot is effective for comparing the number of days online across different vehicle types and gearbox types, making it easy to identify trends and outliers.
2. **Multidimensional Data Representation:** It efficiently incorporates multiple dimensions (vehicle type and gearbox type) into a single chart, providing a comprehensive overview.
3. **Visual Clarity:** The chart is clear and easy to read, ensuring that the information is accessible even to those without a background in data analysis. The use of colors to differentiate gearbox types enhances the chart's readability.