

Aufgabenblatt 1: Visualisierung des Auto Datensatzes

Tim Lukas

October 20, 2025

1 Visualisierung: Preisverteilung nach Karosserieform

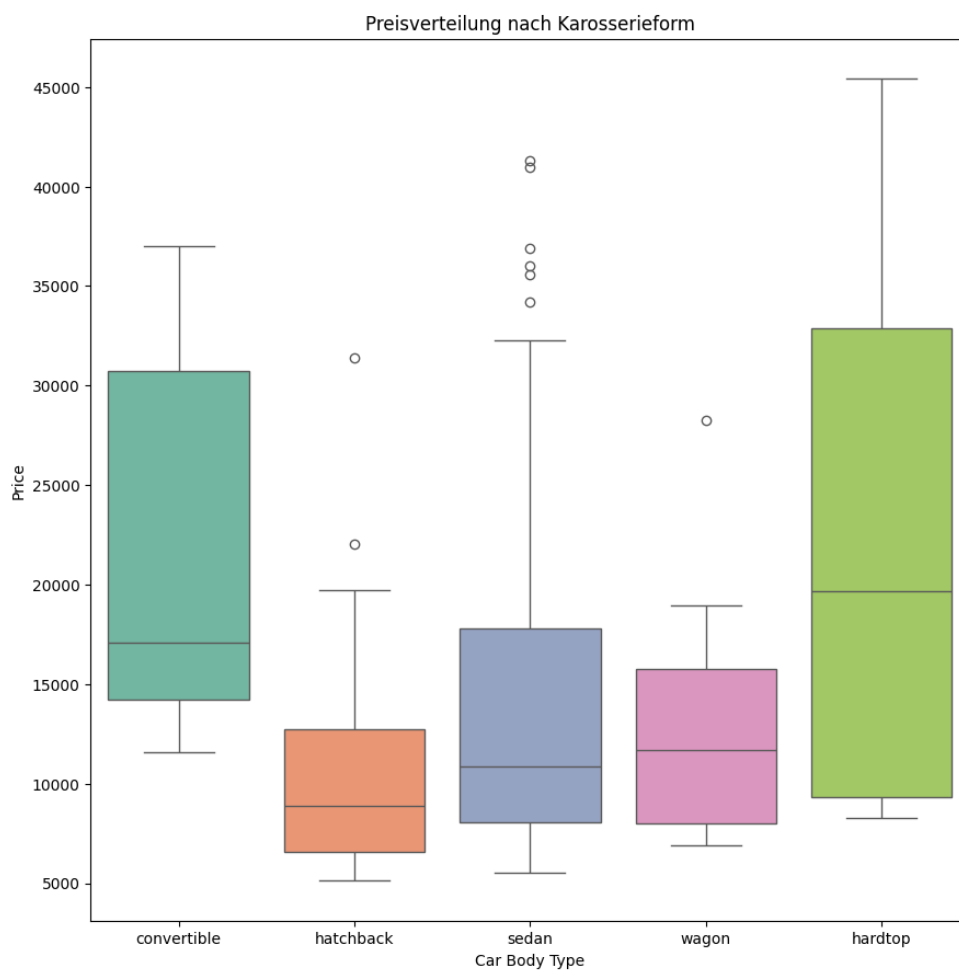


Figure 1: Boxplot der Preisverteilung nach Karosserieform

Beschreibung der Visualisierung Die Abbildung zeigt die Preisverteilung der verschiedenen Karosserieformen als Boxplot. Aus der Visualisierung lassen sich der Durchschnittspreis jeder Karosserieform, die Minimal- und Maximalwerte, das obere und untere Quartil sowie gegebenenfalls einzelne Ausreißer ablesen.

Begründung der Wahl Für die Darstellung als Boxplot habe ich mich entschieden, da sowohl zentrale Tendenzen (Median), als auch die Streuung der Preise einschließlich Extrema und Ausreißer abgebildet werden. So lassen sich Unterschiede in der Preisverteilung klar erkennen.

Verwendete Datenattribute

- **price** - Fahrzeugpreis in US-Dollar
- **Car Body Type** - Karosserieform

Beantwortbare Fragen

- Welche Karosserieform weist den höchsten (**hardtop**) bzw. den geringsten Durchschnittspreis auf(**hatchback**)?
- Welche Karosserieform hat eine eher geringe Preisspanne (**hatchback**, **wagon**), bzw. welche eine sehr große (**hardtop**)?
- Welche Karosserieformen zeigen eine große Preisstreuung oder Ausreißer

2 Visualisierung: Beziehung zwischen Leistung und Kraftstoffeffizienz

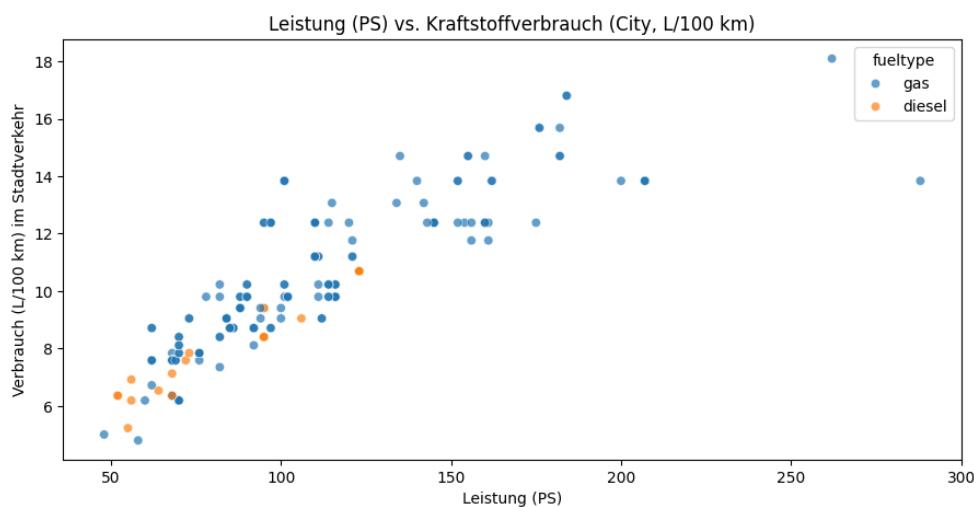


Figure 2: Scatterplot: Leistung vs. Kraftstoffeffizienz (City MPG)

Beschreibung der Visualisierung Die Abbildung zeigt die Beziehung zwischen der Motorleistung (PS) und der Kraftstoffeffizienz (*Liter pro 100 Kilometer*) als Scatterplot. Jeder Punkt der Abbildung repräsentiert ein Fahrzeug, die Farbe des jeweiligen Punktes gibt an, ob das Fahrzeug mit Benzin oder Diesel betrieben wird.

Begründung der Wahl Mithilfe eines Scatterplots kann die Korrelation zwischen zwei numerischen Variablen, in diesem Fall Motorleistung und Kraftstoffeffizienz, gut dargestellt werden, da sich Trends, Zusammenhänge und Außreiser unmittelbar visuell erkennen lassen. Durch die farbliche Unterscheidung der Punkte nach Kraftstofftyp wird zudem sichtbar, ob sich Benzin- und Dieselfahrzeuge hinsichtlich ihres Verbrauchsverhaltens unterscheiden.

Verwendete Datenattribute

- **horsepower** - Motorleistung (PS)
- **citympg** - Kraftstoffeffizienz (Miles per Gallon im Stadtverkehr)
- **fueltype** - Kraftstoffart (Benzin oder Diesel)

Zusätzliche Berechnung Da in Deutschland der Verbrauch typischerweise $L/100k$ angegeben wird, habe ich die Miles Per Gallon im Stadtverkehr in die entsprechenden $L/100k$ umgerechnet. Dafür habe ich folgende Formel verwendet:

Grundlage für die Umrechnung sind folgende Werte:

$$1 \text{ mile} = 1.609344 \text{ km}, \quad 1 \text{ gallon} = 3.785411784 \text{ L}.$$

Basierend auf diesen Werten ergibt sich:

$$\begin{aligned} 1 \text{ mpg} &= \frac{1 \text{ mile}}{1 \text{ gallon}} \\ &= \frac{1.609344 \text{ km}}{3.785411784 \text{ L}} \\ &= 0.4251437 \frac{\text{km}}{\text{L}} \end{aligned}$$

$$\Rightarrow 1 \frac{\text{km}}{\text{L}} = \frac{1}{0.4251437} \text{ mpg} = 2.35215 \text{ mpg}$$

$$\Rightarrow 1 \text{ mpg} = 0.425144 \frac{\text{km}}{\text{L}}$$

$$\text{Da } 1 \text{ L}/100 \text{ km} = \frac{100}{\text{km}/\text{L}},$$

$$\begin{aligned} \text{folgt: } \text{L}/100 \text{ km} &= \frac{100}{0.425144 \times \text{mpg}} \\ &= \frac{235.215}{\text{mpg}} \end{aligned}$$

Die Umrechnung konnte im Python-Skript also mit folgender Formel umsetzen:

$$\boxed{\text{L}/100 \text{ km} = \frac{235.215}{\text{mpg}}}.$$

Beantwortbare Fragen

- Gibt es einen Zusammenhang zwischen den PS eines Autos und dem Benzinverbrauch? - **Ja, es lässt sich ein klarer Zusammenhang zwischen den Werten erkennen.**
- Existiert ein Unterschied zwischen dem leistungsabhängigen Verbrauch von Fahrzeugen, die mit Diesel betrieben werden, gegenüber Fahrzeugen, die mit Benzin betrieben werden? - **Nein, es ist kein Unterschied zwischen Diesel- und Benzinfahrzeugen erkennbar.**

3 Visualisierung: Preisstruktur nach Marke und Karosserieform

Hierarchische Preisstruktur nach Marke und Karosserie

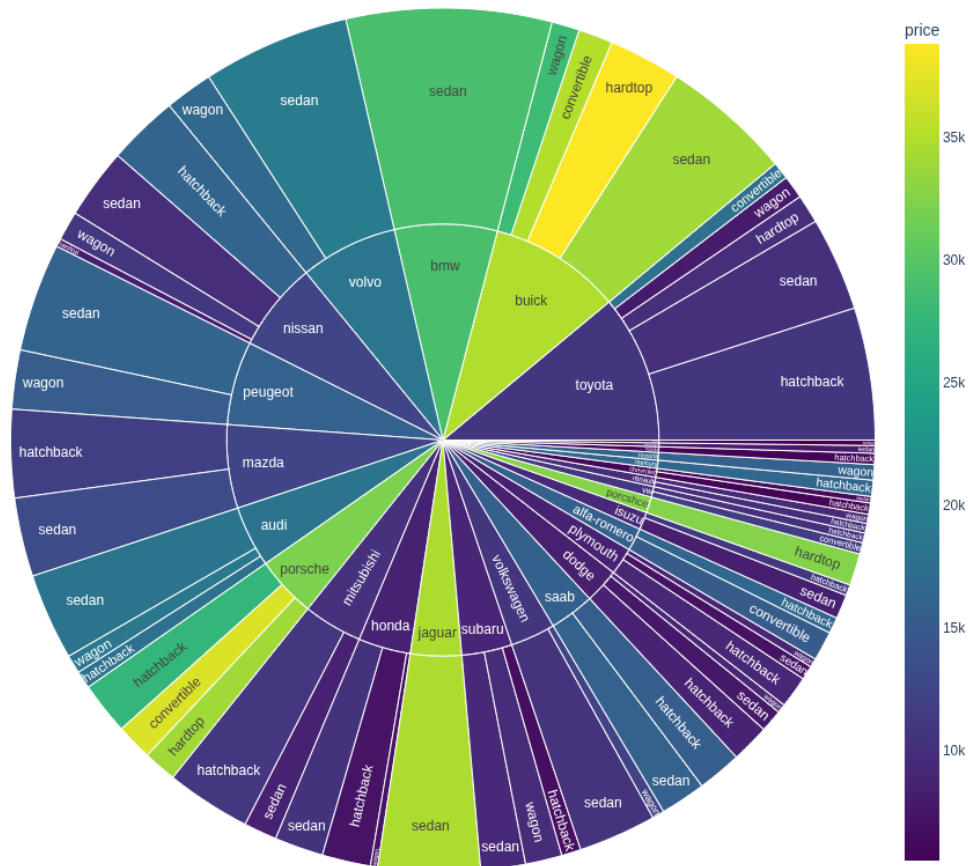


Figure 3: Sunburst-Diagramm: Preisstruktur nach Marke und Karosserieform

Beschreibung Das Sunburst-Diagramm zeigt die hierarchische Preisstruktur der Fahrzeuge nach Marke und Karosserieform. Jeder Sektion im äußeren Ring repräsentiert eine Karosserieform innerhalb einer Marke, während die Sektionen des inneren Rings die unterschiedlichen Marken repräsentieren. Durch Interaktivität (in der erstellten Version) kann eine Marke ausgewählt werden, um eine Detailsansicht zu erhalten.

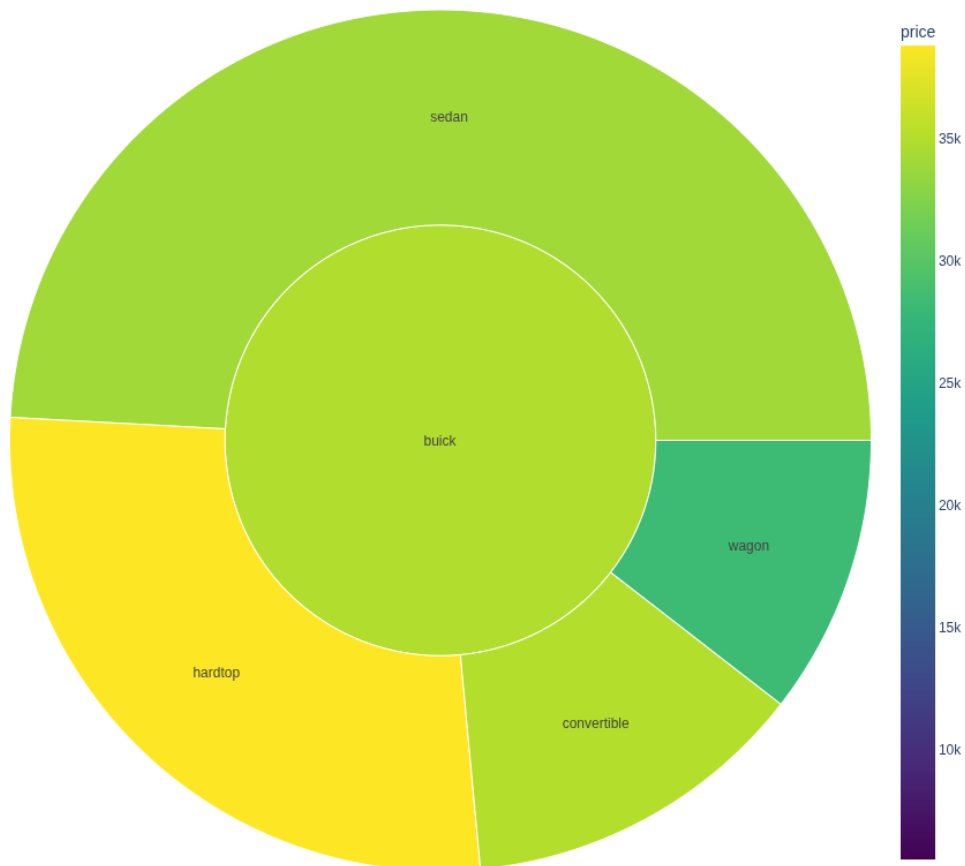


Figure 4: Detailansicht: Preisstruktur der Karosserieformen einer Marke

Beantwortbare Fragen

- Welche Marken haben einen besonders hohen durchschnittlichen Fahrzeugpreis? - **buick, jaguar**
- Welche Marken haben einen besonders niedrigen durchschnittlichen Fahrzeugpreis? - **honda, subaru**
- Wie unterscheiden sich die Fahrzeugpreise einer Marke abhängig von der Karosserieform?
- Welche Karosserieformen haben markenübergreifend durchschnittlich höhere bzw. geringere Durchschnittspreise?

4 Verwendete Tools

Für die Erstellung der Visualisierungen wurde Python mit folgenden Libraries verwendet:

- **pandas** - Einlesen und Verarbeitung der CSV-Daten
- **matplotlib** - Erstellen statischer Diagramme (Grundlage für Scatterplots und Boxplots)

- **seaborn** - Erweiterte grafische Darstellung und Stiloptimierung. Genutzt für Scatterplots und Boxplots und die Anpassung von Farben und Transparenz.
- **plotly.express** - Erstellung interaktiver Diagramme (Sunburst)