

Gruppenarbeit

1 Aufgabe 1: Übersicht und Datenqualität

1. Geben Sie eine Übersicht über den Datensatz mit den Funktionen `summary()` und `skim()`.
2. Visualisieren Sie die fehlenden Datenpunkte mit der Funktion `plot_missing()`.

2 Aufgabe 2: Erstellen neuer Variablen

1. Erstellen Sie eine neue Variable `Sport_wagen`, die den Wert 1 hat, wenn die PS > 300 sind, und 0 ansonsten.
2. Erstellen Sie eine neue Variable `Luxus_Auto`, die den Wert 1 hat, wenn der Preis > 200.000 und die PS > 400 sind.

3 Aufgabe 3: Filtern und Auswählen

1. Filtern Sie alle Autos, die sowohl `Sport_wagen` als auch `Luxus_Auto` sind, und speichern Sie die Ergebnisse in einer neuen Tabelle `luxury_sport`.
2. Filtern Sie alle Autos mit "Mercedes-Benz", deren Preis > 300.000 oder PS > 400 , und speichern Sie nur die Spalten `Model`, `Km`, `Ps`, `Price`. Nennen Sie die Tabelle `last_df`.

4 Aufgabe 4: Datenanalyse und Aggregation

1. Berechnen Sie die durchschnittlichen PS (`Ps`) für jede `Manufacturer`.
2. Berechnen Sie die Anzahl der Autos pro Jahr (`Year`) und visualisieren Sie die Ergebnisse: Nutzen Sie piping mit (`group_by()` dann `summarise(count=n())`)

5 Aufgabe 5: String-Verarbeitung

1. Filtern Sie alle Autos, deren `Model` mit "GL" beginnt, unter Verwendung von `str_detect()`.
2. Erstellen Sie eine neue Variable `SUV`, die den Wert "Ja" hat, wenn der `Model-Name` "GL" oder "X" enthält, und "Nein" ansonsten, mithilfe von `case_when()`.

6 Aufgabe 6: Statistische Berechnungen

1. Erstellen Sie ein Histogramm der PS-Werte (`Ps`) für alle Autos.

7 Aufgabe 7: Visualisierung

1. Erstellen Sie ein Streudiagramm von Preis (`Price`) gegen PS (`Ps`), farbcodiert nach `Manufacturer`.
2. Visualisieren Sie die Verteilung der Autos pro `Year` mit einem Balkendiagramm.

8 Aufgabe 8: Sortieren und Gruppieren

1. Sortieren Sie die Daten nach `Price` absteigend und zeigen Sie die obersten 10 Einträge an. (Nutzen sie bitte die Fct `arrange(desc(Variable)) %>% head(x)`)
2. Gruppieren Sie die Daten nach `Manufacturer` und berechnen Sie den maximalen Preis (`Price`) mit der Fct `summarise(max_price=max(Price, na.rm=TRUE))`

9 Aufgabe 9: Statistische Prüfung

1. Testen Sie, ob der Durchschnittspreis von Autos mit `PS > 300` signifikant höher ist als der von Autos mit `PS <= 300`. Nutzen sie bitte hierfür den `t.test()` Befehl um den Welch 2-Sample Test durchzuführen.
2. Berechnen Sie die Korrelation zwischen `PS (ps)` und `Preis (Price)`.

10 Aufgabe 10: Arbeiten mit Teilmengen

1. Erstellen Sie eine Teilmenge des Datensatzes mit den Spalten `Manufacturer`, `Model`, `Price`, und `ps`, in der nur Autos mit `Preis > 100.000` enthalten sind.
2. Speichern Sie die gefilterte Teilmenge als CSV-Datei.