

# Dokumentation: In-Vehicle Coupon Recommendation Datensatz

---

## 1. Beschreibung des Datensatzes

Der Datensatz 'in-vehicle-coupon-recommendation IHK.csv' enthält Informationen über Fahrten, Mitfahrer, Wetter, Gutscheine und demografische Merkmale von Personen. Ziel ist es, vorherzusagen, ob ein Fahrgast einen in-vehicle Coupon annimmt (target = 1) oder nicht (target = 0).

## 2. Spaltenübersicht und Erklärungen

Spalte	Erklärung
<b>destination</b>	Zielort der Fahrt (z. B. Home, Work, No Urgent Place)
<b>passanger</b>	Mitfahrer (Alone, Friend(s), Partner, Kid(s))
<b>weather</b>	Wetterbedingungen (Sunny, Rainy, Snowy)
<b>temperature</b>	Temperatur (numerisch, z. B. 55, 80)
<b>time</b>	Uhrzeit der Fahrt (z. B. 2PM, 10AM, 6PM)
<b>coupon</b>	Art des Gutscheins (z. B. Restaurant(<20), Coffee House, Bar, Carry out & Take away, Restaurant(20-50))
<b>expiration</b>	Gültigkeit des Gutscheins (2h, 1d)
<b>gender</b>	Geschlecht (Male, Female)
<b>age</b>	Alter (numerisch, z. B. 21, 46)
<b>maritalStatus</b>	Familienstand (Single, Married, Unmarried partner, Divorced)
<b>has_children</b>	Hat Kinder (0 = nein, 1 = ja)
<b>education</b>	Bildungsstand (z. B. Bachelors degree, Some college - no degree, High School Graduate)
<b>income</b>	Einkommensgruppe (z. B. \$37500 - \$49999, \$100000 or More)
<b>car</b>	Hat Auto (never, less1, 1~3, 4~8, gt8, unknown)
<b>target</b>	Zielvariable: Gutschein angenommen (1) oder nicht (0)

### 3. Datenqualität

Anzahl fehlender Werte pro Spalte:

destination: 0

passanger: 0

weather: 0

temperature: 0

time: 0

coupon: 0

expiration: 0

gender: 0

age: 0

maritalStatus: 0

has\_children: 0

education: 0

income: 0

car: 12576

Anzahl Duplikate: 74

Datentypen der Spalten:

- destination: object
- passanger: object
- weather: object
- temperature: int64
- time: object
- coupon: object
- expiration: object
- gender: object
- age: object
- maritalStatus: object
- has\_children: int64
- education: object
- income: object
- car

### 4. Vorverarbeitungsschritte

1. Einlesen der CSV-Datei mit Semikolon als Trennzeichen.
2. Auswahl und Umbenennung relevanter Spalten.
3. 3. Spalte Car wird entfernt weil zu viele fehlende Werte in de spalte sind.

4. Missing Value mit most value befüllt da es nicht viele fehlende werte gibt

#### **4.1 Modellierung**

1. Umwandlung von Texten in Zahlen (Label Encoding für Machine Learning).
2. Aufteilung in Trainings- und Testdaten für Machine Learning.

#### **5. Machine Learning-Vorbereitung**

Für Machine Learning werden alle kategorialen Spalten (z. B. coupon, gender) in numerische Werte umgewandelt. Die Zielspalte 'target' wird als Integer (0 oder 1) verwendet. Die Daten werden in Trainings- und Testdaten aufgeteilt (z. B. 30 % Training, 70 % Test).

#### **6. Erste Analyseergebnisse**

Verteilung der Zielvariable (target):

1: 7210  
0: 5474