

# Einkaufsgewohnheiten in den USA

Ole Kepa, Fabian Elsner, Sören Bax

## Inhaltsverzeichnis

<b>1</b>	<b>Gendererklärung</b>	<b>1</b>
<b>2</b>	<b>Aufgabe und Daten verstehen</b>	<b>1</b>
<b>3</b>	<b>Beschreibung der Datenquelle</b>	<b>2</b>
3.1	Erste Datenquelle . . . . .	2
3.2	Zweite Datenquelle . . . . .	3
<b>4</b>	<b>Untersuchung der Daten</b>	<b>5</b>
4.1	Einlesen der Daten . . . . .	5
<b>5</b>	<b>Anwendung und Evaluation von ML-Modellen</b>	<b>5</b>
<b>6</b>	<b>Kritische Bewertung/ Ideen für weitere Analysen</b>	<b>5</b>
<b>7</b>	<b>Ehrenwörtliche Erklärung</b>	<b>5</b>

## 1 Gendererklärung

Aus Lesbarkeitsgründen wird in dieser Studienarbeit auf die verschiedene Ansprechweisen, sei es divers, männlich oder weiblich verzichtet. Alle Formulierungen sprechen gleichermaßen alle Geschlechter an.

## 2 Aufgabe und Daten verstehen

Thesen...

## 3 Beschreibung der Datenquelle

### 3.1 Erste Datenquelle

Wir lesen zunächst Daten ein .

```
shopping_trends <- read_csv('./data/shopping_trends.csv')
```

Nun werfen wir einen kurzen Blick auf die Datenstruktur:

```
shopping_trends
```

```
# A tibble: 3,900 x 19
```

```
  `Customer ID`   Age Gender `Item Purchased` Category Purchase Amount (USD~1
      <dbl> <dbl> <chr>   <chr>          <chr>          <dbl>
1             1    55 Male    Blouse         Clothing         53
2             2    19 Male    Sweater        Clothing         64
3             3    50 Male    Jeans          Clothing         73
4             4    21 Male    Sandals        Footwear         90
5             5    45 Male    Blouse         Clothing         49
6             6    46 Male    Sneakers        Footwear         20
7             7    63 Male    Shirt          Clothing         85
8             8    27 Male    Shorts          Clothing         34
9             9    26 Male    Coat           Outerwear         97
10            10    57 Male    Handbag        Accessori~         31
```

```
# i 3,890 more rows
```

```
# i abbreviated name: 1: `Purchase Amount (USD)`
```

```
# i 13 more variables: Location <chr>, Size <chr>, Color <chr>, Season <chr>,
```

```
# `Review Rating` <dbl>, `Subscription Status` <chr>, `Payment Method` <chr>,
```

```
# `Shipping Type` <chr>, `Discount Applied` <chr>, `Promo Code Used` <chr>,
```

```
# `Previous Purchases` <dbl>, `Preferred Payment Method` <chr>,
```

```
# `Frequency of Purchases` <chr>
```

Jede Zeile steht für irgendwas.

Variable	Typ	Bedeutung
<b>Customer ID</b>	dbl	Eindeutige Kunden Identifikationsnummer
<b>Age</b>	dbl	Alter des Kunden
<b>Gender</b>	chr	Geschlecht des Kunden
<b>Item Purchased</b>	chr	Gekauftes Produkt
<b>Category</b>	chr	Kategorie des gekauften Produkts

Variable	Typ	Bedeutung
<b>Purchase Amount</b>	dbl	Bezahlter Preis
<b>Location</b>	chr	
<b>Size</b>	chr	
<b>Color</b>	chr	
<b>Season</b>	chr	
<b>Review Rating</b>	dbl	
<b>Subscription Status</b>	chr	
<b>Payment Method</b>	chr	
<b>Shipping Type</b>	chr	
<b>Discount Applied</b>	chr	
<b>Promo Code Used</b>	chr	
<b>Previous Purchases</b>	dbl	
<b>Preferred Payment Method</b>	chr	
<b>Frequency of Purchases</b>	chr	

```
describe_tbl(shopping_trends)
```

```
3 900 (3.9k) observations with 19 variables
0 observations containing missings (NA)
0 variables containing missings (NA)
0 variables with no variance
```

Im ersten Datensatz gibt es 3.900 Instazen (Beobachtungen). Keine dieser Instanzen enthalten Werte ohne Angabe (NA), daher müssen wir das Datenset nicht aufgrund fehlender Variablen aufbereiten.

## 3.2 Zweite Datenquelle

Wir lesen erneut Daten ein .

```
shopping_behavior <- read_csv('./data/shopping_behavior_updated.csv')
```

Nun werfen wir einen kurzen Blick auf die Datenstruktur:

```
shopping_behavior
```

```
# A tibble: 3,900 x 18
  `Customer ID`   Age Gender `Item Purchased` Category Purchase Amount (USD~1
    <dbl> <dbl> <chr> <chr> <chr> <dbl>
1         1      55 Male   Blouse      Clothing      53
2         2      19 Male   Sweater     Clothing      64
3         3      50 Male   Jeans       Clothing      73
4         4      21 Male   Sandals     Footwear      90
5         5      45 Male   Blouse      Clothing      49
6         6      46 Male   Sneakers    Footwear      20
7         7      63 Male   Shirt       Clothing      85
8         8      27 Male   Shorts      Clothing      34
9         9      26 Male   Coat        Outerwear     97
10        10     57 Male   Handbag     Accessori~    31
# i 3,890 more rows
# i abbreviated name: 1: `Purchase Amount (USD)`
# i 12 more variables: Location <chr>, Size <chr>, Color <chr>, Season <chr>,
#   `Review Rating` <dbl>, `Subscription Status` <chr>, `Shipping Type` <chr>,
#   `Discount Applied` <chr>, `Promo Code Used` <chr>,
#   `Previous Purchases` <dbl>, `Payment Method` <chr>,
#   `Frequency of Purchases` <chr>
```

Jede Zeile steht für irgendwas.

Variable	Typ	Bedeutung
<b>Customer ID</b>	dbl	Eindeutige Kunden Identifikationsnummer
<b>Age</b>	dbl	Alter des Kunden
<b>Gender</b>	chr	Geschlecht des Kunden
<b>Item Purchased</b>	chr	Gekauftes Produkt
<b>Category</b>	chr	Kategorie des gekauften Produkts
<b>Purchase Amount</b>	dbl	Bezahlter Preis
<b>Location</b>	chr	
<b>Size</b>	chr	
<b>Color</b>	chr	
<b>Season</b>	chr	
<b>Review Rating</b>	dbl	
<b>Subscription Status</b>	chr	
<b>Payment Method</b>	chr	
<b>Shipping Type</b>	chr	
<b>Discount Applied</b>	chr	
<b>Promo Code Used</b>	chr	
<b>Previous Purchases</b>	dbl	
<b>Payment Method</b>	chr	

Variable	Typ	Bedeutung
<b>Frequency of Purchases</b>	chr	

```
describe_tbl(shopping_behavior)
```

```
3 900 (3.9k) observations with 18 variables
0 observations containing missings (NA)
0 variables containing missings (NA)
0 variables with no variance
```

Im zweiten Datensatz gibt es 3.900 Instanzen (Beobachtungen). Keine dieser Instanzen enthalten Werte ohne Angabe (NA), daher müssen wir das Datenset nicht aufgrund fehlender Variablen aufbereiten.

## 4 Untersuchung der Daten

### 4.1 Einlesen der Daten

Aufbereitung

## 5 Anwendung und Evaluation von ML-Modellen

Text

## 6 Kritische Bewertung/ Ideen für weitere Analysen

Text

## 7 Ehrenwörtliche Erklärung

Hiermit erklären wir, dass wir die vorliegende Studienarbeit (Produktstudie) selbständig angefertigt haben und die Bearbeiter der einzelnen Abschnitte wahrheitsgemäß angegeben haben. Es wurden nur die in der Arbeit ausdrücklich benannten Quellen und Hilfsmittel benutzt. Wörtlich oder sinngemäß übernommenes Gedankengut haben wir als solches kenntlich gemacht. Diese Arbeit hat in gleicher oder ähnlicher Form ganz oder teilweise noch keiner Prüfungsbehörde vorgelegen.