Einkaufsgewohnheiten in den USA

Ole Kepa, Fabian Elsner, Sören Bax

Inhaltsverzeichnis

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

3 Beschreibung der Datenquelle

Laden wir zuerst die beiden Datensets:

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

Nun werfen wir einen kurzen Blick auf die Datenstrukturen. Zunächst einmal die Datenstruktur des Datensets “Shopping_trends”:

```
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 einen *********

Variable	Typ	Bedeutung
Customer ID	dbl	Eindeutige Kunden Identifikationsnummer
Age	dbl	Alter des Kundens
Gender	chr	Geschlecht des Kundens
Item Purchased	chr	Gekauftes Produkt
Category	chr	Kategorie des gekauften Produkts
Purchase Amount	dbl	Bezahlter Preis
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	

Variable	Typ	Bedeutung
Frequency of Purchases	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 Datensatz "Shopping_trends" gibt es 3.900 Instazen (Beobachtungen). Keine dieser Instanzen enthalten Werte ohne Angabe (NA), daher müssen wir das Datenset nicht aufgrund fehlender Variablen aufbereiten. Nun schauen wir auf die Datenstruktur des Datensets "Shopping_behavior":

```
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 *******.

Variable	Typ	Bedeutung
Customer ID	dbl	Eindeutige Kunden Identifikationsnummer
Age	dbl	Alter des Kundens
Gender	chr	Geschlecht des Kundens
Item Purchased	chr	Gekauftes Produkt
Category	chr	Kategorie des gekauften Produkts
Purchase Amount	dbl	Bezahlter Preis
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	
Payment Method	chr	
Frequency of Purchases	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 Datensatz “Shoping_behavior” gibt es ebenfalls 3.900 Instazen (Beobachtungen). Zudem gibt wa qiwswe bwi jwswe Instanz keine enthalten Werte ohne Angabe (NA). Daher müssen wir, wie beim vorherigen Datensatz keine fehlenden Variablen aufbereiten.

7.2 Ideen für weitere Analysen

8 Quellen

Datenset: https://www.kaggle.com/datasets/zeesolver/consumer-behavior-and-shopping-habits-dataset/data

9 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.

Ole Kepa Ole Kepa Ole Kepa