

Hauspreise

EDA und Regression

Ole Kepa, Fabian Elsner, Sören Bax

Inhaltsverzeichnis

1 Gendererklärung	1
2 Aufgabe und Daten verstehen	1
3 Beschreibung der Datenquelle	2
4 Untersuchung der Daten	2
4.1 Vorbereitung	2
4.1.1 Laden der Pakete	2
4.1.2 Einlesen der Daten	2
4.2 Einlesen der Daten	4
5 Anwendung und Evaluation von ML-Modellen	4
6 Kritische Bewertung/ Ideen für weitere Analysen	4
7 Ehrenwörtliche Erklärung	4

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

Super Datenquelle

4 Untersuchung der Daten

4.1 Vorbereitung

4.1.1 Laden der Pakete

Wir laden zunächst die benötigten Pakete.

```
library(tidyverse)
library(explore)
```

4.1.2 Einlesen der Daten

Wir lesen zunächst Daten ein .

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

Wir werfen einen Blick auf die ersten 20 Einträge:

```
head(shopping_trends, 20)
```

A tibble: 20 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
11	11	53	Male	Shoes	Footwear	34

12	12	30	Male	Shorts	Clothing	68
13	13	61	Male	Coat	Outerwear	72
14	14	65	Male	Dress	Clothing	51
15	15	64	Male	Coat	Outerwear	53
16	16	64	Male	Skirt	Clothing	81
17	17	25	Male	Sunglasses	Accessori~	36
18	18	53	Male	Dress	Clothing	38
19	19	52	Male	Sweater	Clothing	48
20	20	66	Male	Pants	Clothing	90

```
# 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>
```

Wir lesen erneut Daten ein .

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

Wir werfen einen Blick auf die ersten 20 Einträge:

```
head(shopping_behavior, 20)
```

```
# A tibble: 20 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
11        11     53 Male   Shoes          Footwear         34
12        12     30 Male   Shorts         Clothing         68
13        13     61 Male   Coat           Outerwear        72
14        14     65 Male   Dress          Clothing         51
15        15     64 Male   Coat           Outerwear        53
```

16	16	64 Male	Skirt	Clothing	81
17	17	25 Male	Sunglasses	Accessori~	36
18	18	53 Male	Dress	Clothing	38
19	19	52 Male	Sweater	Clothing	48
20	20	66 Male	Pants	Clothing	90

```

# 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>

```

4.2 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.