What is the data

Data Mining Cup 2015, PRUDSYS

features

feature	scale
dateOfBirth	○
deliveryDate	○
color	□
creationDate	○
orderDate	○
state	□
customerID	□
itemID	□
manufacturerID	□
orderItemID	□
salutation	□
price	×
size	+
returnShipment	□

481,000 clothes orders

Missing values
No   Yes

Additional processing
No   Yes

Measurement scales
○ Interval
□ Nominal
+ Ordinal
× Ratio

Salutation:
highly uneven data

ReturnShipment:
balanced by organizator

Data preparation

Missed values:

- deliveryDate: omitted
- dateOfBirth: omitted
- color: omitted

All ID features omitted

Standartizing size dimension
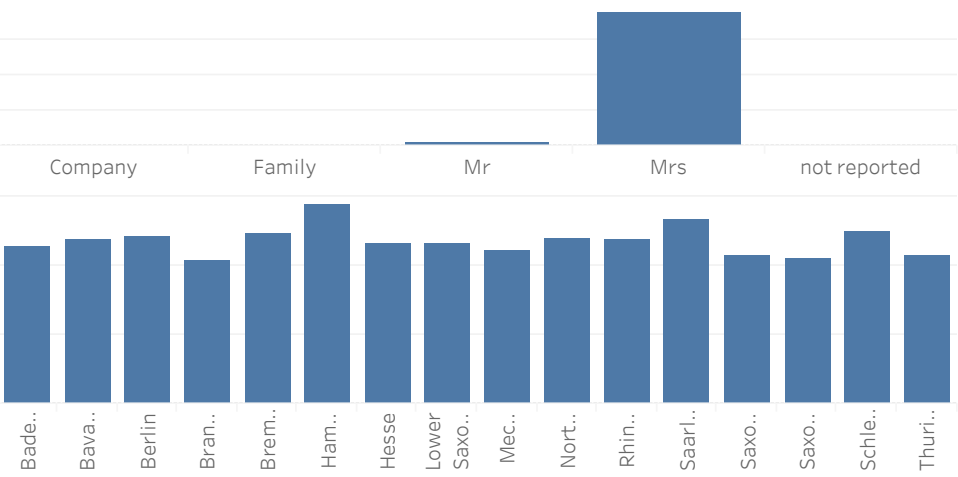
Top 10 colors

Data features -> number of days

Useful features

state
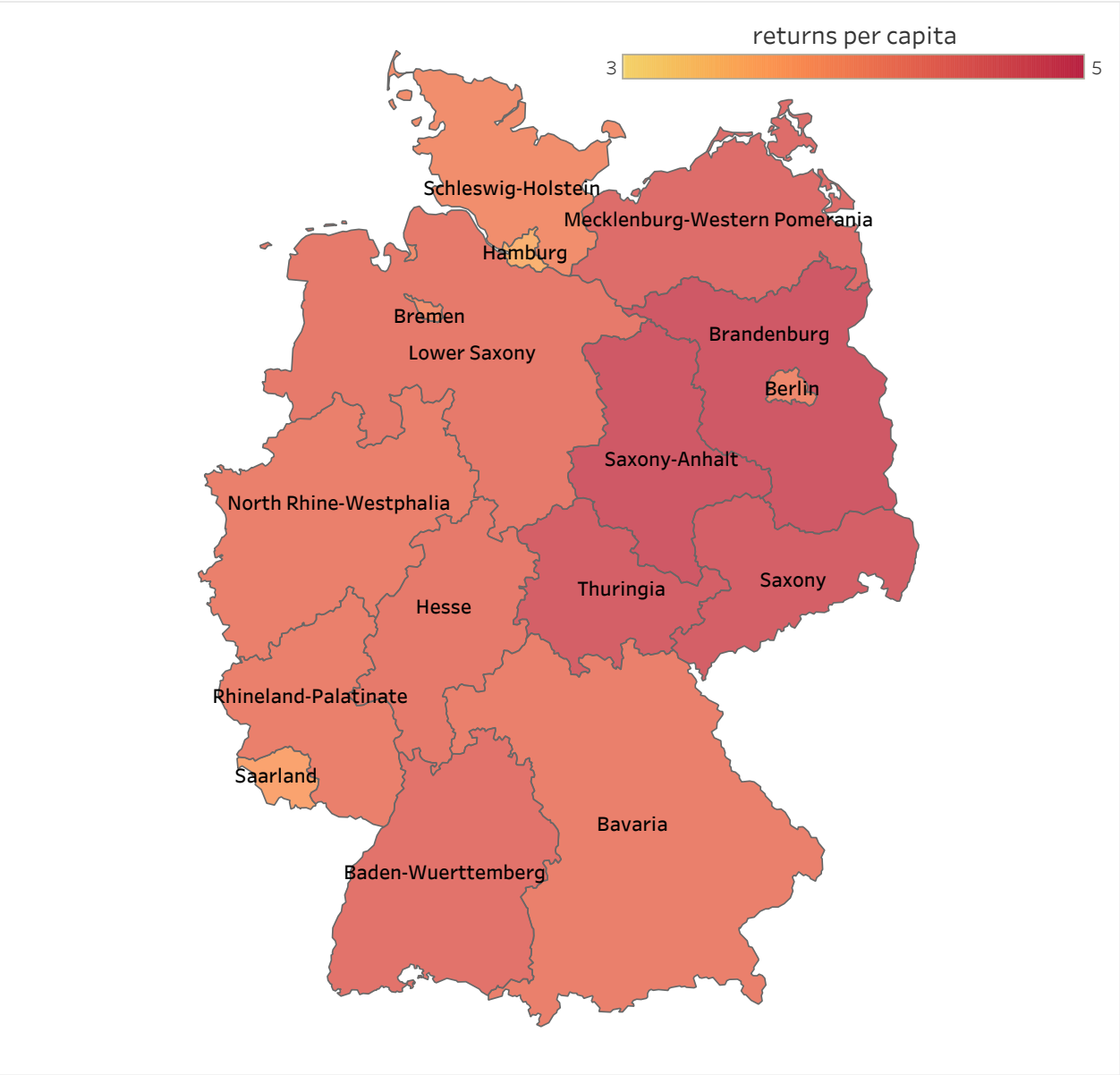
color

size

number of customers
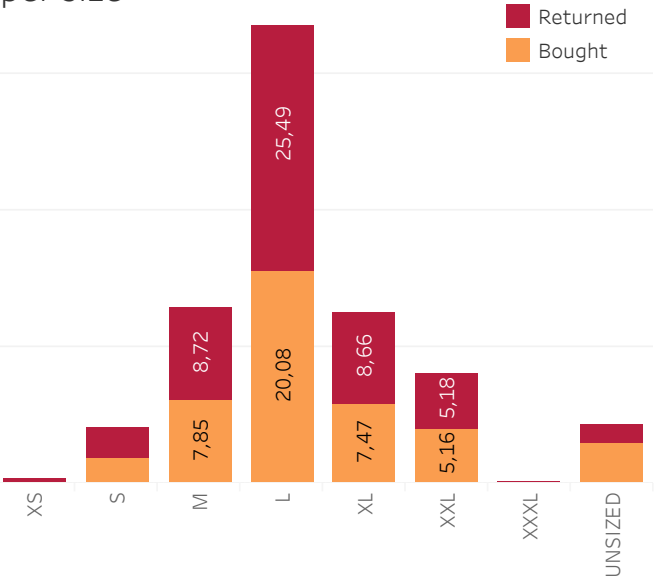
number of items

number of returns

Findings

Company   Family   Mr   Mrs   not reported

Bade.. Bava.. Berlin Bran.. Brem.. Ham.. Hesse Lower Saxo.. Mec.. Nort.. Rhin.. Saarl.. Saxo.. Saxo.. Schle.. Thuri..

How much Germans return

per capita

returns per capita

3   5

Schleswig-Holstein

Mecklenburg-Western Pomerania

Hamburg

Bremen

Lower Saxony

Brandenburg

Berlin

North Rhine-Westphalia

Saxony-Anhalt

Hesse

Thuringia

Saxony

Rhineland-Palatinate

Saarland

Bavaria

Baden-Wuerttemberg

per size

- Returned
- Bought

XS	S	M	L	XL	XXL	XXXL	UNSIZED

25,49

8,72   7,85

20,08

8,66   7,47

5,18   5,16

Top 10 colors bought with return percentage

green

grey
55,5 %

black
54,5 %

blue
54,1 %

white

brown

red