

Motoric 7.12.

Outliers

Problem

$$\begin{aligned}
 & \text{Age mean: } 5,30 \quad \} \quad \text{aktuell } 5,38 + (2 \cdot 0,67) = \\
 & \text{Age std: } 0,67 \\
 - \text{Age max: } 6,9 \quad & \xrightarrow{\quad} \quad 6,72 \\
 & \text{realer Wert oder wegen ungewöhnlicher} \\
 & \text{Alter}
 \end{aligned}$$

→ Outliers vorher entfernen?

→ $2 \times \text{mean} \rightarrow \text{value} \rightarrow \text{remove}$

→ $\text{std} \cdot 2 + \text{mean}$

→ only swimming lessons participate + age

CB Data

Zipcode	Gender	Age	Value
3151	Total	0-5	260
3151	Men	0-5	210

total - 0-5 - value men - 0-5 - value



260

210

CDS Data

ID	Gender	Leeftijd	Postcode	Perioden	Bevolking - 1
total		0-5	3191	2020	460
total		5-10	3191	2020	520
men		0-5	3191	2020	:
men		5-10	3191	2020	:
women		0-5	3191		:
women		5-10	3191		:

Men Df

Zip code total_0-5 total_5-10		
1	1	1
zip list_women		col_1 + col_2
Bevolking		
= where		
col_1 + column 2	is in the row +	
zipcode		

Tony

- validation, set
 - Research Paper
 - Plan time
 - Portfolio
- Feedback → send Tony

Cleaning

- convert columns \Rightarrow object to float

- remove Outliers \Rightarrow ign, mean + std

- T0, T1

- add z-jp codes

- calculate

- MQ Score \rightarrow 4-6, 7-10

- BMI category \rightarrow 4-6, 7-10

- MQ category \rightarrow 4-6, 7-10

- drop columns where no/1 unique value is found