

1 What can we do with ‘size’?

1. Size count for each product. More than 40% of products have size count equal to 1. Is that possible that these product are on sale/clearance? A new variable can be construct based on the size count of a product.

size count	1	2	3	4	5	6	7	8	9	10	11	12	13	14	16	Total
# product	1965	940	596	482	396	118	72	51	33	33	26	23	18	16	1	4770

2. There are some special size that only occur in one or two products, e.g. “YSM 125-135”, “YM 135-147,5”, “YLG 147,5-157,5”, and “YXL 157,5-167,5” correspond to a Under Armour product (for kids). Special size like this could be used to identify product information. In this example, all the corresponding observations have a mainCategory equal to 15, thus I suspect that 15 represents ‘kids’, and 1 and 9 represent ‘men’ and ‘women’.
3. Some sizes are less informative and could be treated as the same category, e.g. “M(38/40)” and “M(38-42)”. Categories like this could be useful when combining with other features.

2 Re-clustering

1. XS (5)
 - XS (30-34); XS ; XS (32/34); XS (32); XS/S \Rightarrow XS
2. S (4)
 - S; S (34-38); S (34/36); S (36) \Rightarrow S
3. M (9)
 - M (38-42); M (38/40); M (40); 38/40 (M / L); M (38 - 42); M; M/L \Rightarrow M
 - 102 (M); 24 (M) \Rightarrow Jako.M
4. L (9)
 - L; L (40/42); L (42-46); L (43 - 46); L/XL (39-47); L (42-47); L (44) \Rightarrow L.1
 - L/K; L/T \Rightarrow L.2
5. XL (14)
 - XL; XL (44/46); XL (46-50); XL (46-48,5) \Rightarrow XL.11
 - L/XL (39-47); XL/T; 8 (XL) \Rightarrow XL.12
 - 2XL; 2XL/T \Rightarrow XL.2
 - 3XL; 3XL/T ; 28 (3XL) \Rightarrow XL.3
 - 4XL \Rightarrow XL.4
 - 30 (5XL) \Rightarrow XL.5
6. Others (5)
 - 0 (Bambini) \Rightarrow O.B
 - 1 (Junior); 01 Junior \Rightarrow O.J
 - 2 (Senior); 02 Senior \Rightarrow O.S
7. Under Armour for kids (4)
 - YSM 125-135 \Rightarrow UA.S
 - YM 135-147,5 \Rightarrow UA.M
 - YLG 147,5-157,5 \Rightarrow UA.L
 - YXL 157,5-167,5 \Rightarrow UA.XL
8. Nike Special (5)

- XS (116-128) \Rightarrow N.XS
- S (128-140) \Rightarrow N.S
- M (140-152) \Rightarrow N.M
- L (152-158) \Rightarrow N.L
- XL (158-170) \Rightarrow N.XL

9. Numbered Size 2-11

- 2 \Rightarrow 2
- 3 \Rightarrow 3
- 4 \Rightarrow 4
- 5 \Rightarrow 5
- 6 \Rightarrow 6
- 7 \Rightarrow 7
- 8 \Rightarrow 8
- 9 \Rightarrow 9
- 10 \Rightarrow 10
- 11 \Rightarrow 11

10. Numbered 29-48 2/3

- 29 \Rightarrow 29
- 30 \Rightarrow 30
- 31;31,5 \Rightarrow 31
- 32 \Rightarrow 32
- 33;33,5 \Rightarrow 33
- 34 \Rightarrow 34
- 35;35,5 \Rightarrow 35
- 36;36 2/3;36,5 \Rightarrow 36
- 37;37 1/3;37,5 \Rightarrow 37
- 38;38 2/3;38,5;19 (38) \Rightarrow 38
- 39;39 1/3;39,5 \Rightarrow 39
- 40;40 2/3;40,5 \Rightarrow 40
- 41;41 1/3;41,5 \Rightarrow 41
- 42;42 2/3;42,5 \Rightarrow 42
- 43;43 1/3;43,5 \Rightarrow 43
- 44;44 2/3;44,5 \Rightarrow 44
- 45;45 1/3;45,5 \Rightarrow 45
- 46;46 2/3;46,5 \Rightarrow 46
- 47;47 1/3;47,5 \Rightarrow 47
- 48;48 2/3;48,5 \Rightarrow 48

11. Numbered Size 104-176

- 104 \Rightarrow 104
- 116 \Rightarrow 116
- 128;0 (128) \Rightarrow 128
- 134 \Rightarrow 134
- 140;1 (140);10 (140) \Rightarrow 140

- $146 \Rightarrow 146$
- $152;2 (152) \Rightarrow 152$
- $158 \Rightarrow 158$
- $164;14 (164);3 (164) \Rightarrow 164$
- $176;16 (176) \Rightarrow 176$

12. Numbered Size Range(25-30)-(47-50)

- $1 (25-30) ; 00 (27-30) \Rightarrow 25-30$
- $0 (31-33) ; 1 (31-34) ; 2 (31-34) ; \Rightarrow 31-34$
- $1 (33-36) ; 1 (34-36) ; 35 - 38 ; 35/38 ; 3 (35-38) ; 2 (35-38) ; 2 (37-39) ; 10 (36-40) ; 37 - 40 ; 2 (37-40) \Rightarrow 35-40$
- $39 - 42 ; 39-42 ; 39/42 ; 3 (39-42) ; 4 (39-42) ; 3 (40-42) ; 3 (41-43) ; 41 - 44 ; 12 (41-45) ; 4 (43-45) \Rightarrow 41-45$
- $43 - 46 ; 43-46 ; 4 (43-46) ; 5 (43-46) ; 43/46 ; 4 (44-46) ; 45 - 47 \Rightarrow 44-46$
- $45-48 ; 14 (46-48) ; 5 (46-48) ; 5 (47-49) ; 47/49 ; 47 - 50 ; 6 (47-50) ; \Rightarrow 47-50$

13. Numbered Size Range(116-122)-(164-176)

- $116-122 ; 6/8 (116-128) ; 116/128 \Rightarrow 116-128$
- $10/12 (140-152) ; 140/152 \Rightarrow 140-152$
- $14/16 (164-176) ; 164/176 \Rightarrow 164-176$

14. Exceptions

- Note : 9 items (pid) without size information