

Women's Clothing E-Commerce dataset revolving around the reviews

OPRACOWALI: AGNIESZKA DĄBROWSKA-MAŁYSZKO, WITOLD ŁAŃCUCKI, PAULINA SOKAL, IRENEUSZ TUR

Informacje wstępne:

- ▶ W zbiorze danych mam 23 486 wierszy i 10 zmiennych dla każdego z nich. Każdy z nich to jedna ocena klienta:
- ▶ Clothing ID: Integer
- ▶ Age: Integer
- ▶ Title: String
- ▶ Review Text: String
- ▶ Rating: Integer w zakresie od 1 do 5, gdzie 1 to najniższa ocena, a 5 najlepsza
- ▶ Recommended IND: Binary variable, gdzie 1 = ,polecam', a 0 = ,nie polecam'
- ▶ Positive Feedback Count: Integer
- ▶ Division Name: String
- ▶ Department Name: String
- ▶ Class Name: String

Plan prezentacji

- ▶ ANALIZA DANYCH LICZBOWYCH
- ▶ ANALIZA KOMENTARZY
- ▶ ANALIZA CZYNNIKOWA

Podstawowe dane statystyczne

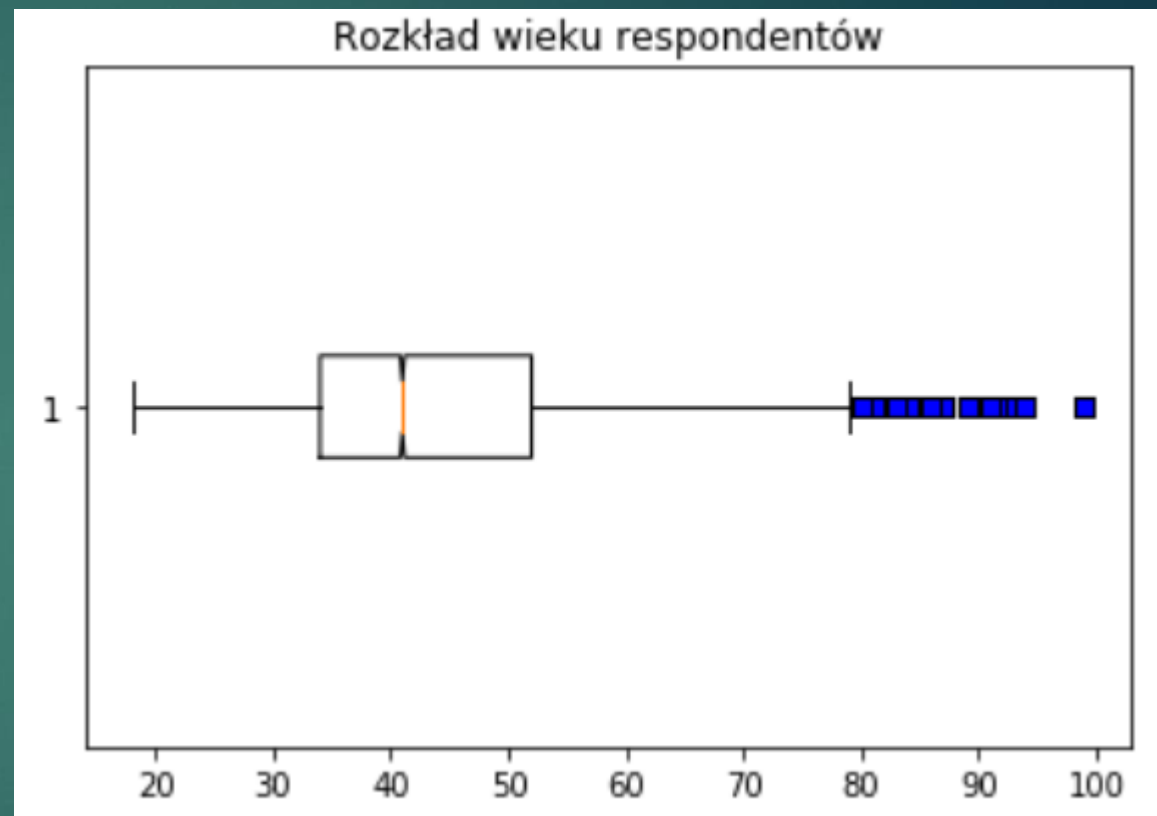
	Age	Rating	RecommendedIND	PositiveFeedbackCount
count	23486.000000	23486.000000	23486.000000	23486.000000
mean	43.198544	4.196032	0.822362	2.535936
std	12.279544	1.110031	0.382216	5.702202
min	18.000000	1.000000	0.000000	0.000000
25%	34.000000	4.000000	1.000000	0.000000
50%	41.000000	5.000000	1.000000	1.000000
75%	52.000000	5.000000	1.000000	3.000000
max	99.000000	5.000000	1.000000	122.000000

Korelacje

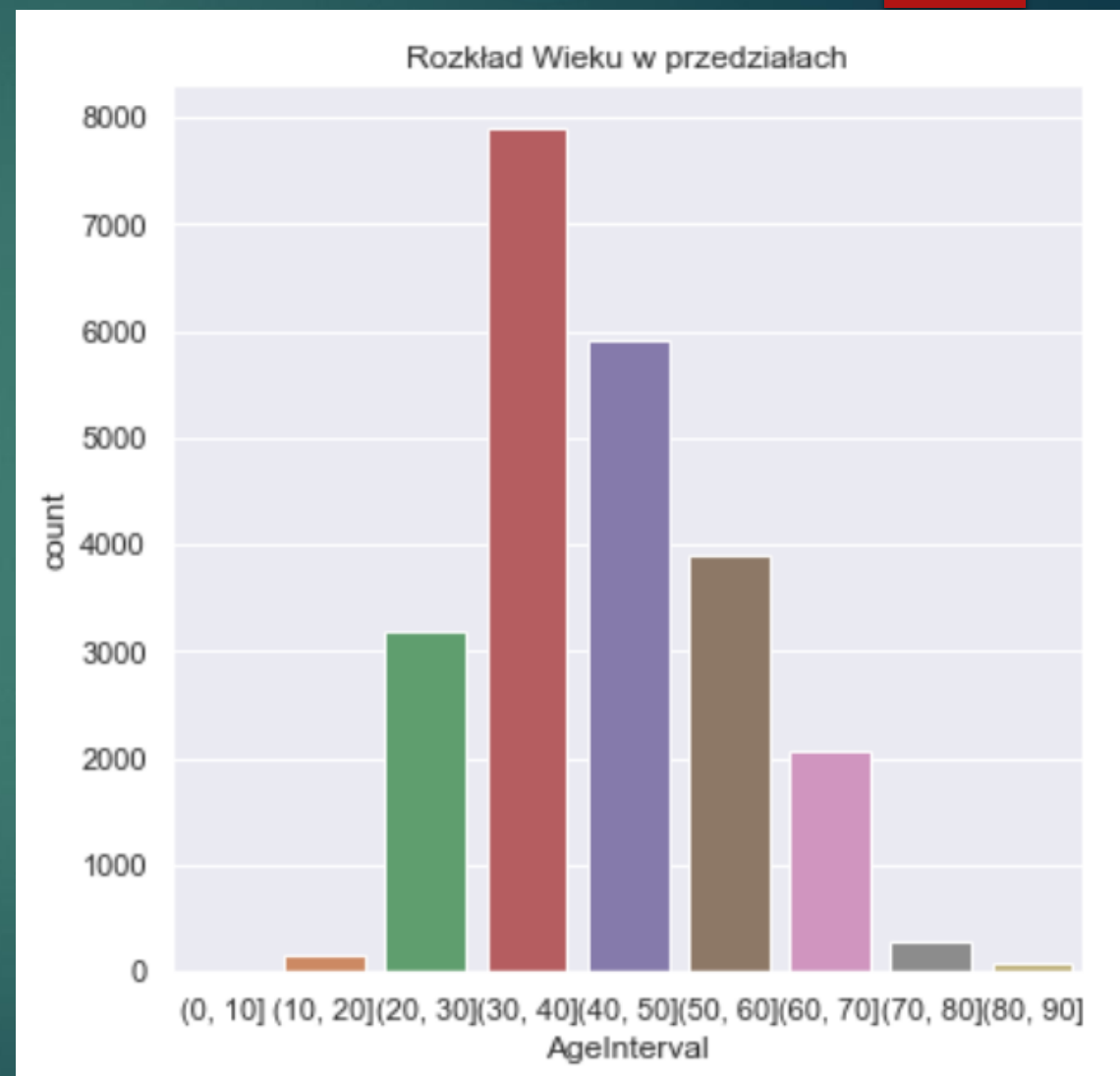
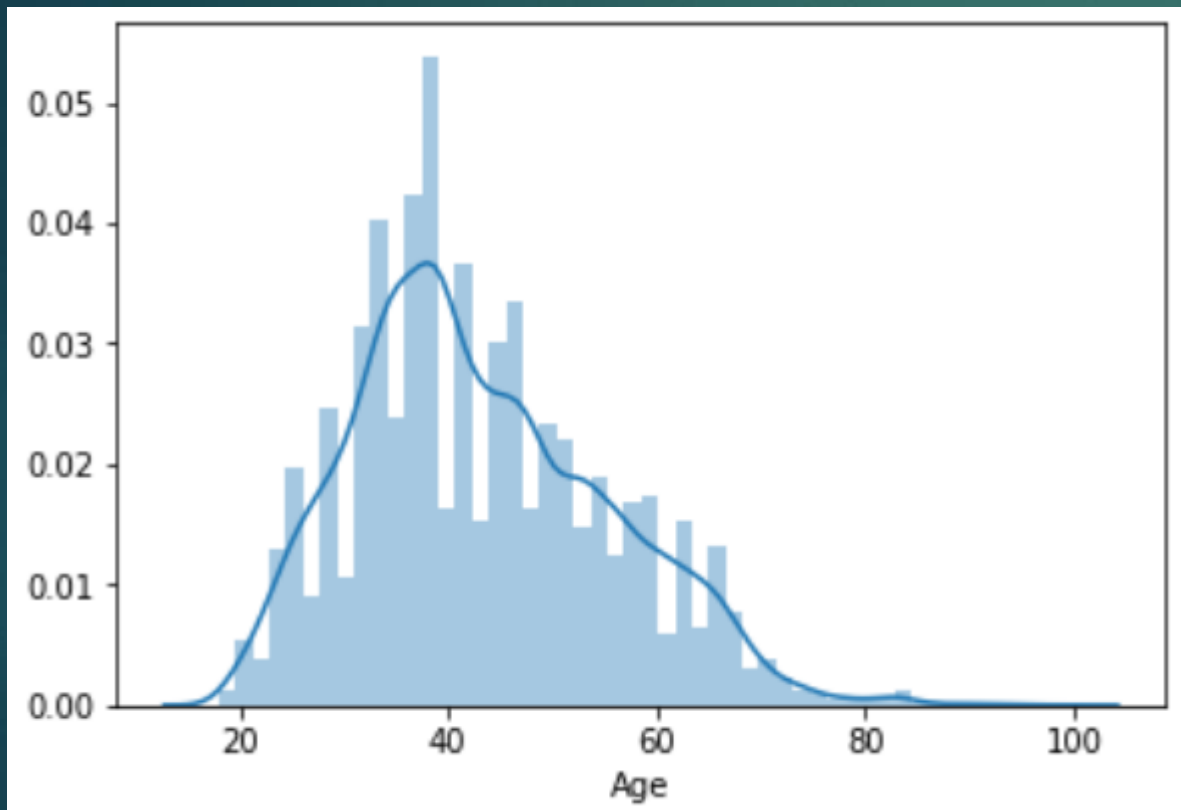
ClothingID	1	0.018	-0.019	-0.016	0.046	-0.25
Age	0.018	1	0.027	0.031	0.043	0.0042
Rating	-0.019	0.027	1	0.79	-0.065	0.017
RecommendedIND	-0.016	0.031	0.79	1	-0.069	0.011
PositiveFeedbackCount	0.046	0.043	-0.065	-0.069	1	-0.042
ClassNameNUM	-0.25	0.0042	0.017	0.011	-0.042	1
	ClothingID	Age	Rating	RecommendedIND	PositiveFeedbackCount	ClassNameNUM

Kilka słów o wieku

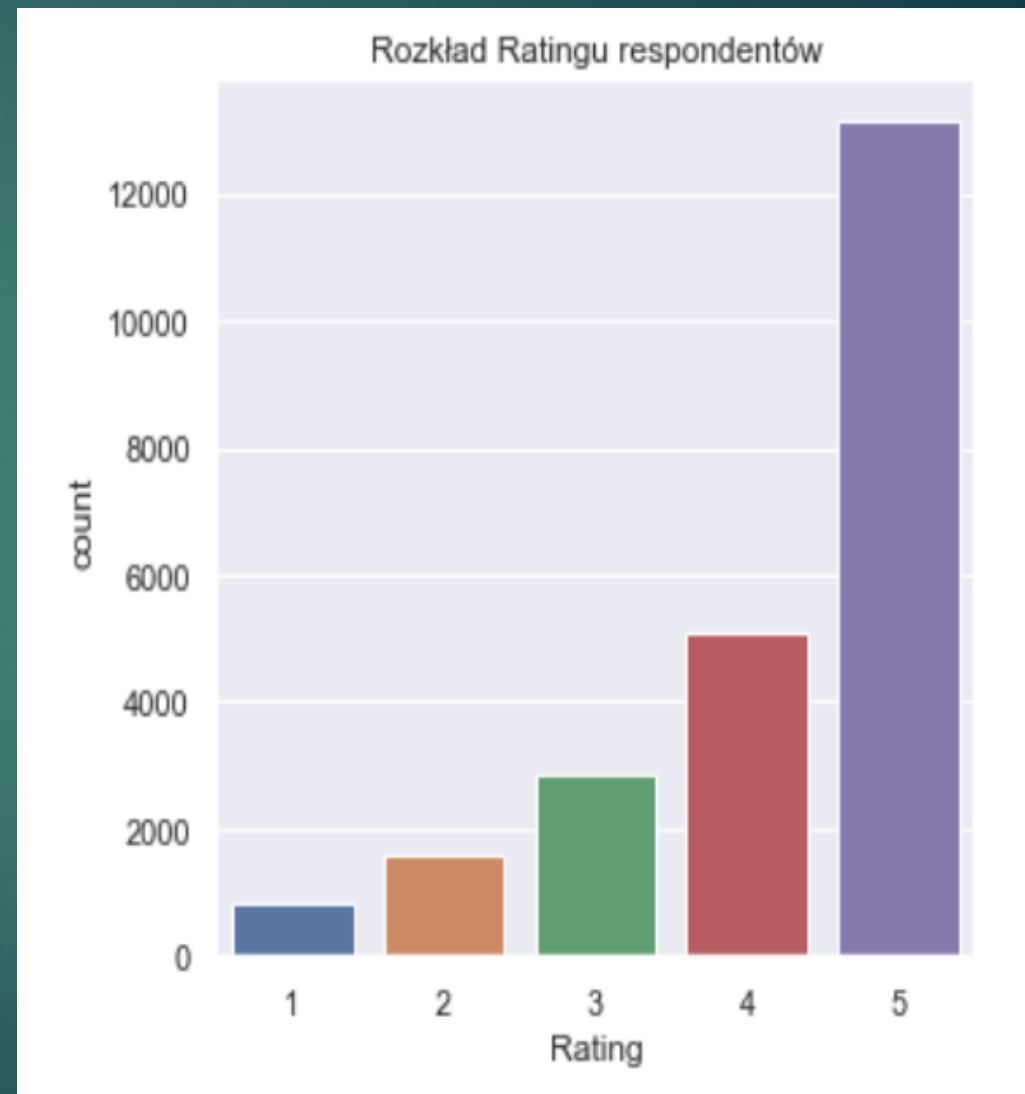
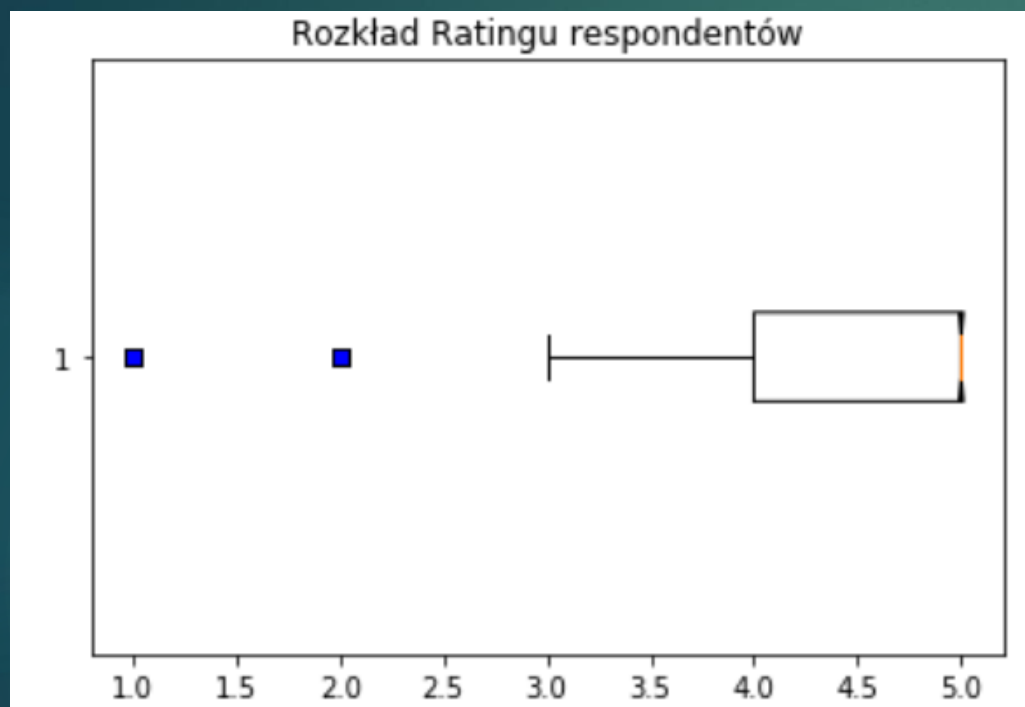
	AgeInterval	Rating	Percentage
0	(0, 10]	0	0.000000
1	(10, 20]	152	0.647553
2	(20, 30]	3186	13.573041
3	(30, 40]	7912	33.706812
4	(40, 50]	5908	25.169344
5	(50, 60]	3891	16.576492
6	(60, 70]	2057	8.763260
7	(70, 80]	281	1.197120
8	(80, 90]	86	0.366378



Kilka słów o wieku

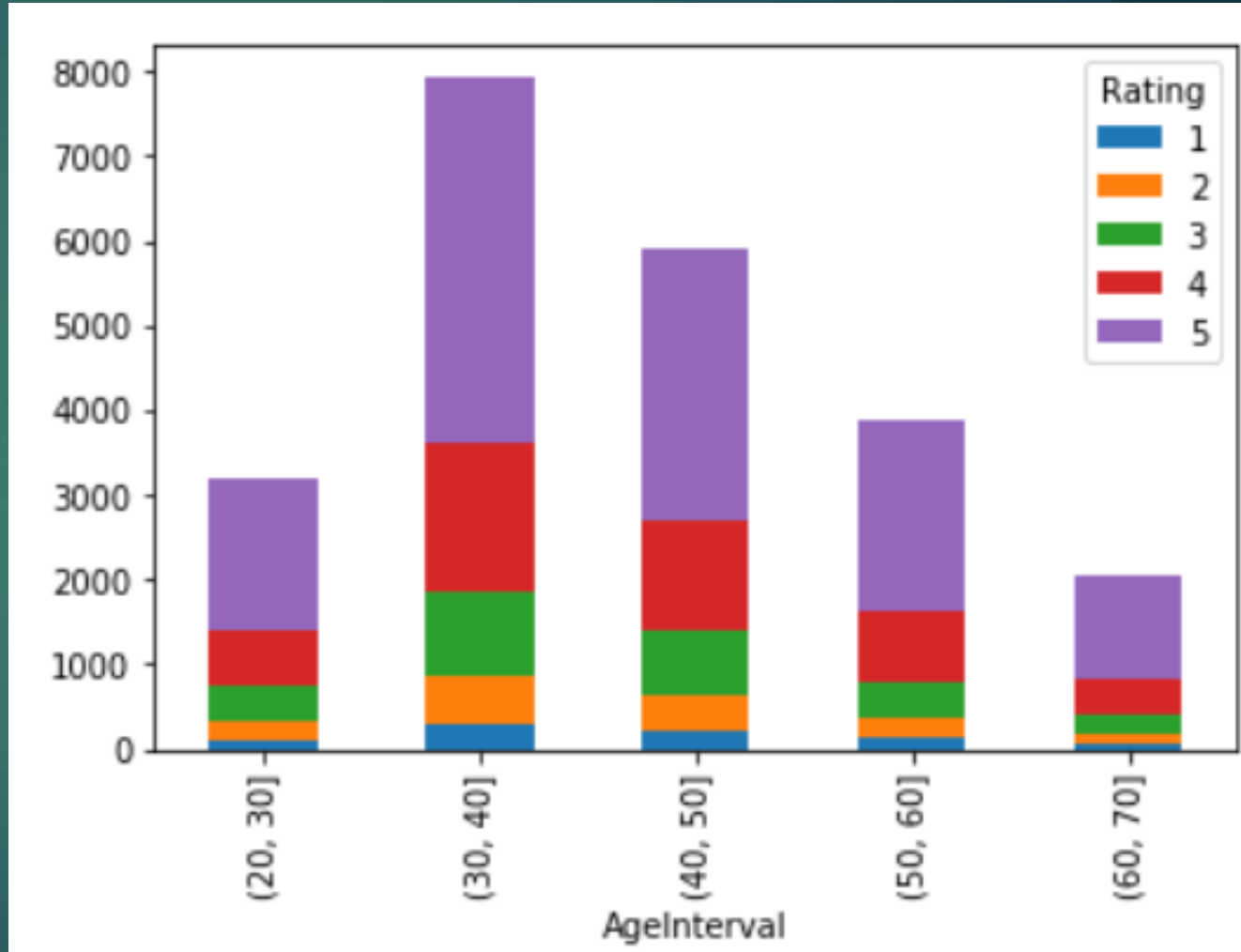


Rating jako miara zadowolenia

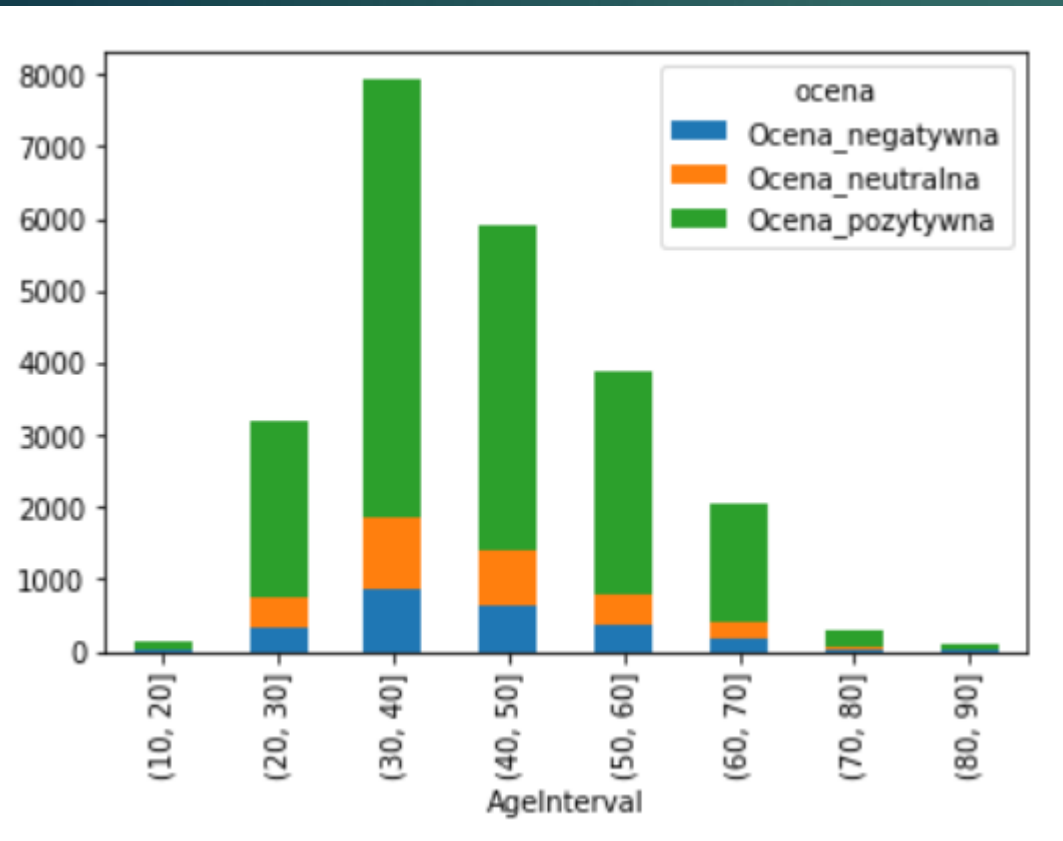


Rating a wiek

AgeInterval	Rating	
(20, 30]	1	3.452605
	2	6.591337
	3	13.433773
	4	20.935342
	5	55.586943
(30, 40]	1	3.538928
	2	7.128413
	3	12.815976
	4	22.143579
	5	54.373104
(40, 50]	1	3.537576
	2	7.109005
	3	12.677725
	4	22.494922
	5	54.180772
(50, 60]	1	3.829350
	2	5.988178
	3	10.562837
	4	20.997173
	5	58.622462
(60, 70]	1	3.646087
	2	5.153136
	3	10.500729
	4	19.834711
	5	60.865338



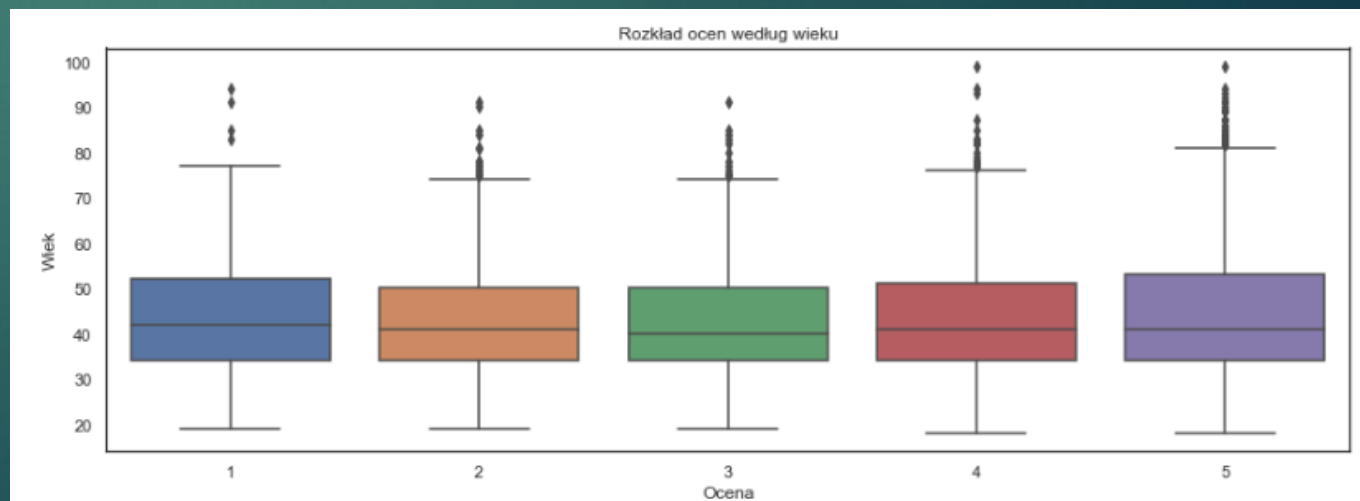
Rating a wiek



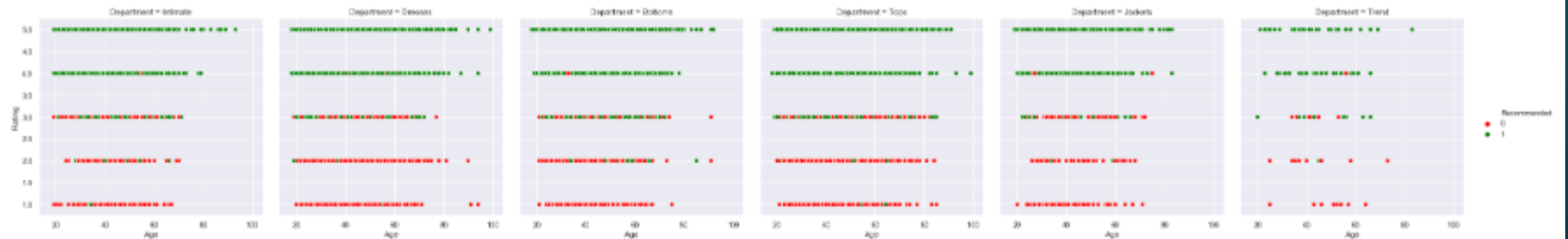
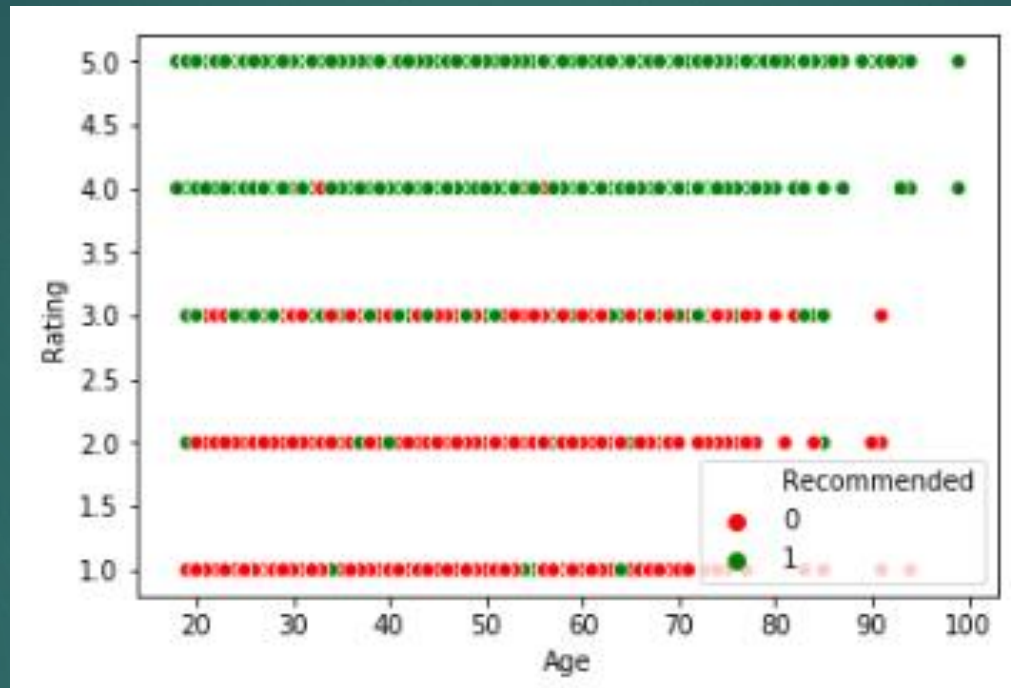
Struktura ocen wewnątrz przedziałów jest zbliżona.

Struktura ocen:

- około 80% to oceny pozytywne
- 10% to oceny neutralne
- 10% to oceny negatywne



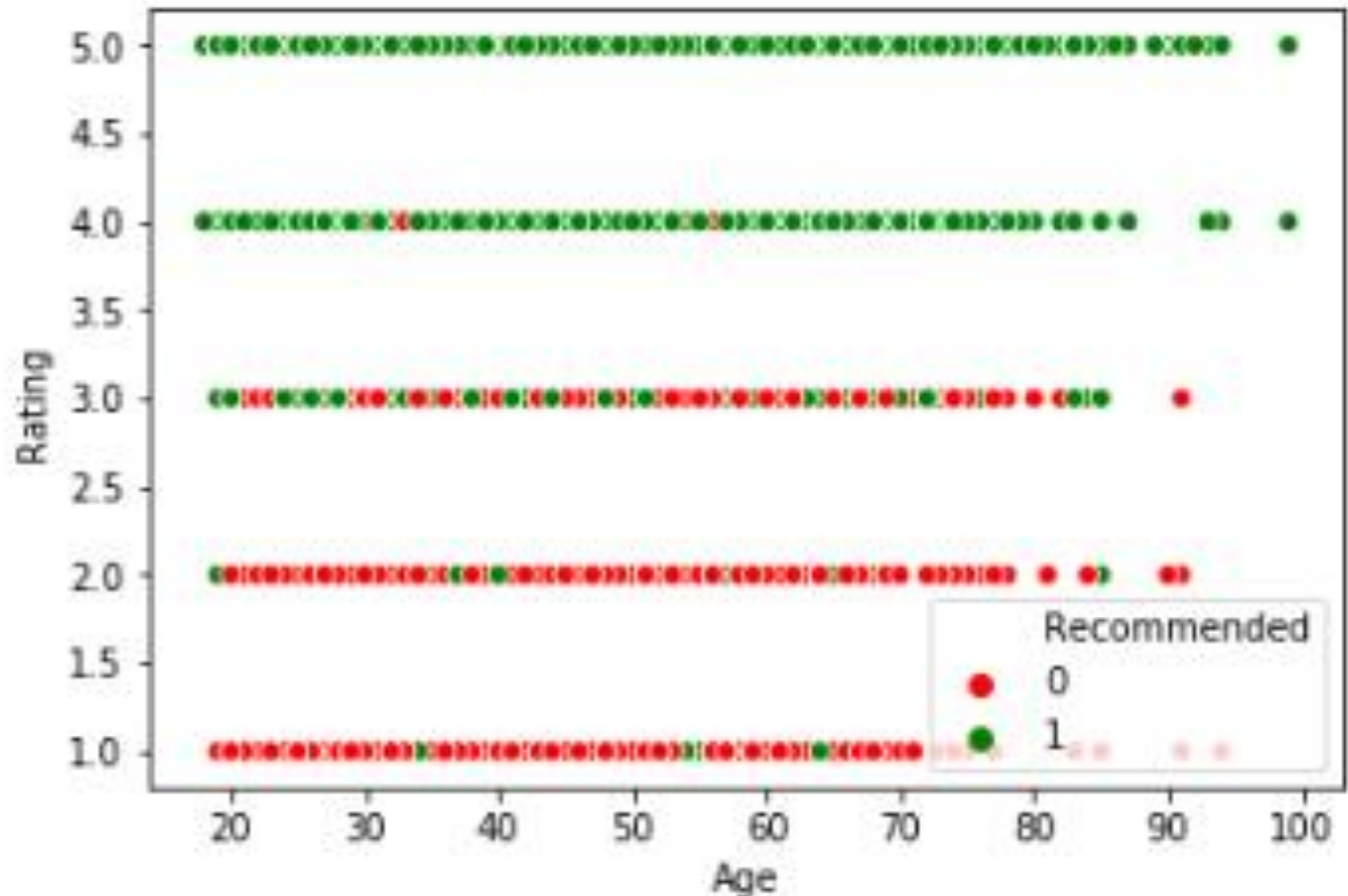
Polecenie produktu jako miara zadowolenia



Polecenie produktu jako miara zadowolenia

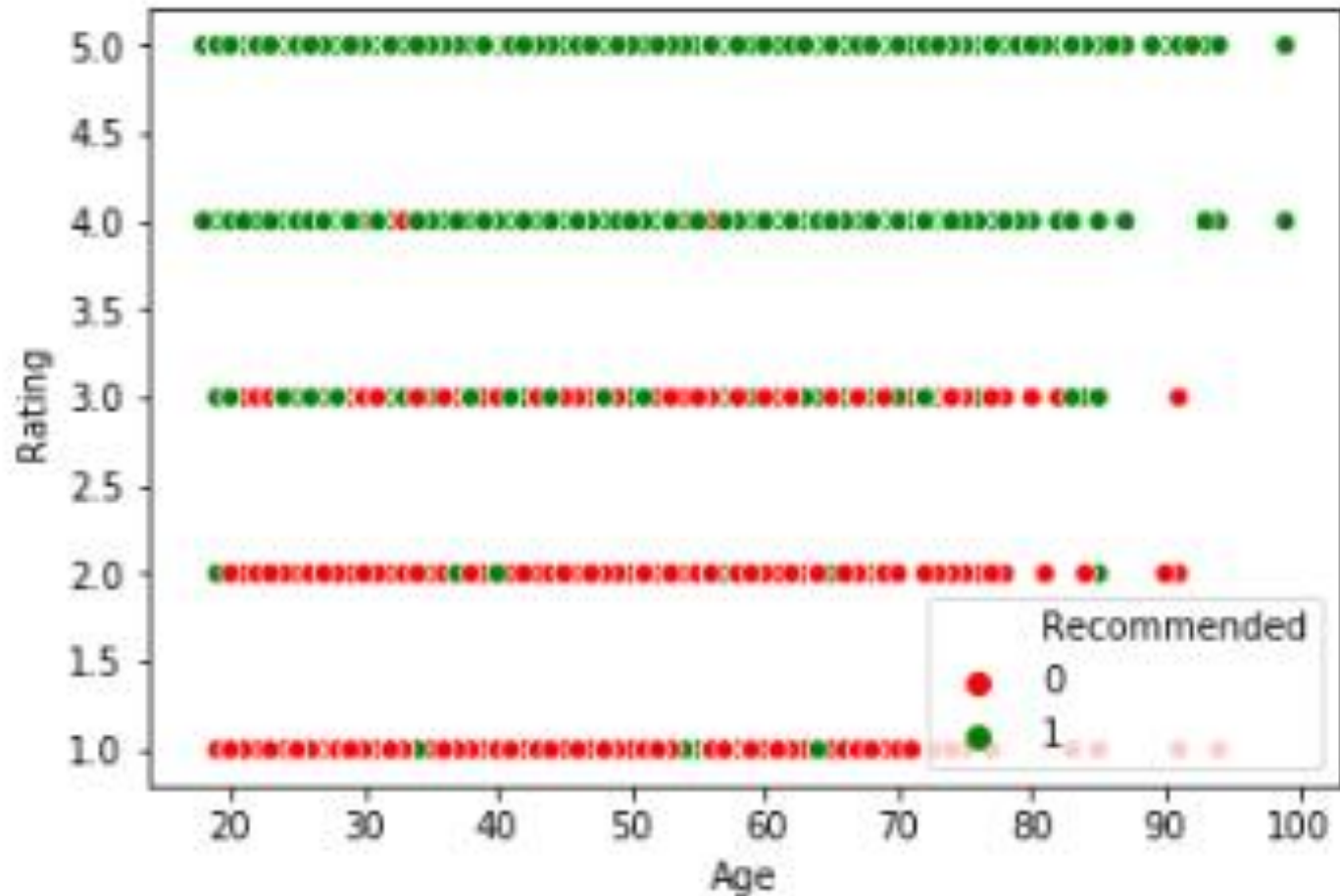
Prawdopodobieństwo polecenia produktu pod warunkiem wystawienia oceny > 3 wynosi 99%

Prawdopodobieństwo polecenia produktu pod warunkiem wystawienia oceny ≤ 3 wynosi 25%



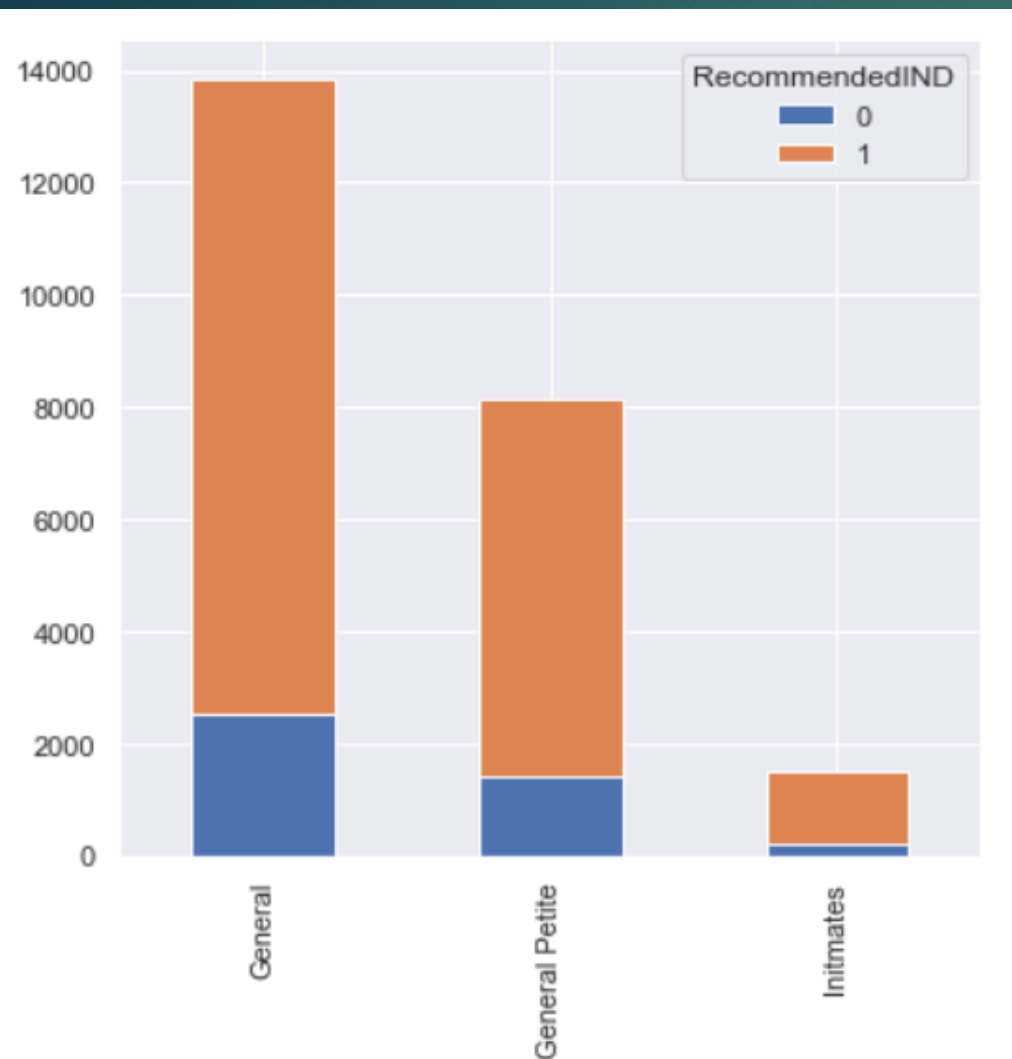
Polecenie produktu jako miara zadowolenia

Prawdopodobieństwo polecenia produktu pod warunkiem, że respondent ma mniej niż 30 lat wynosi 82%



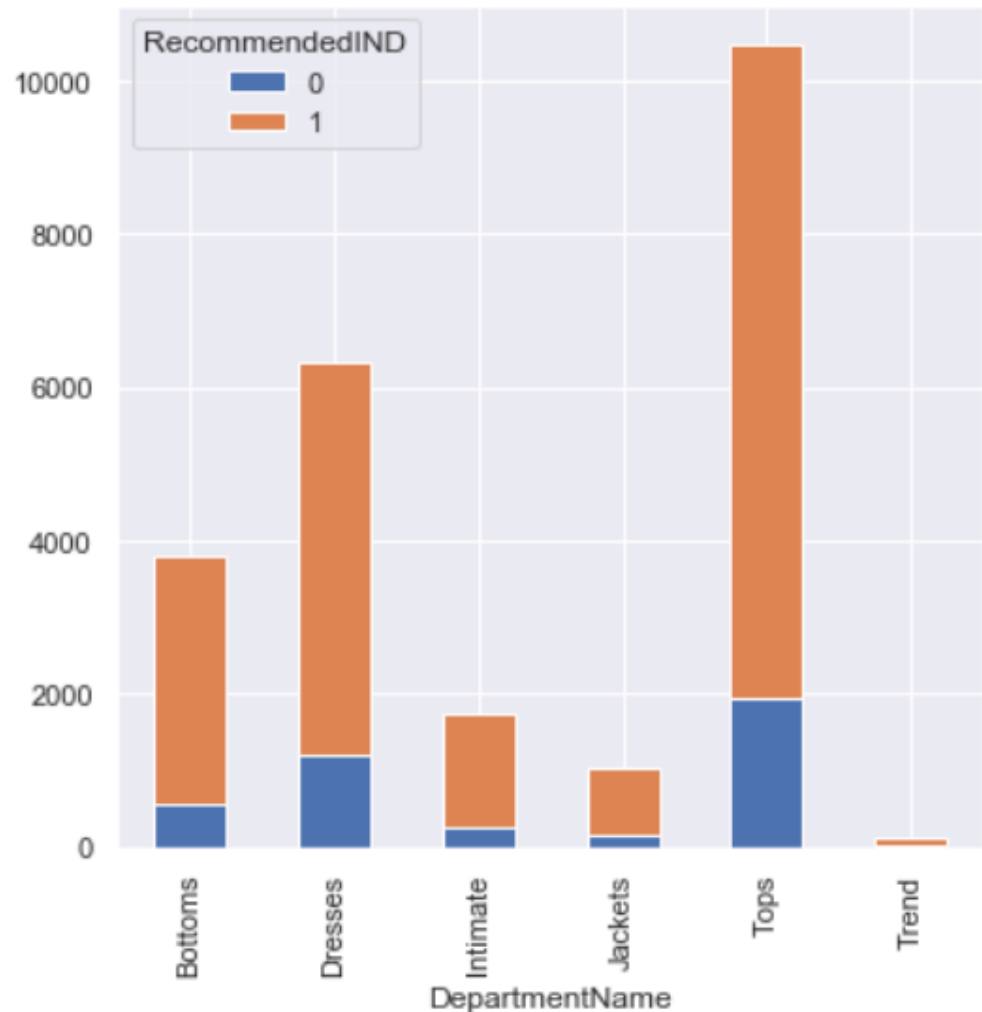
Prawdopodobieństwo polecenia produktu pod warunkiem, że respondent ma więcej niż 50 lat wynosi 84%

Polecenie produktu jako miara zadowolenia – działły:



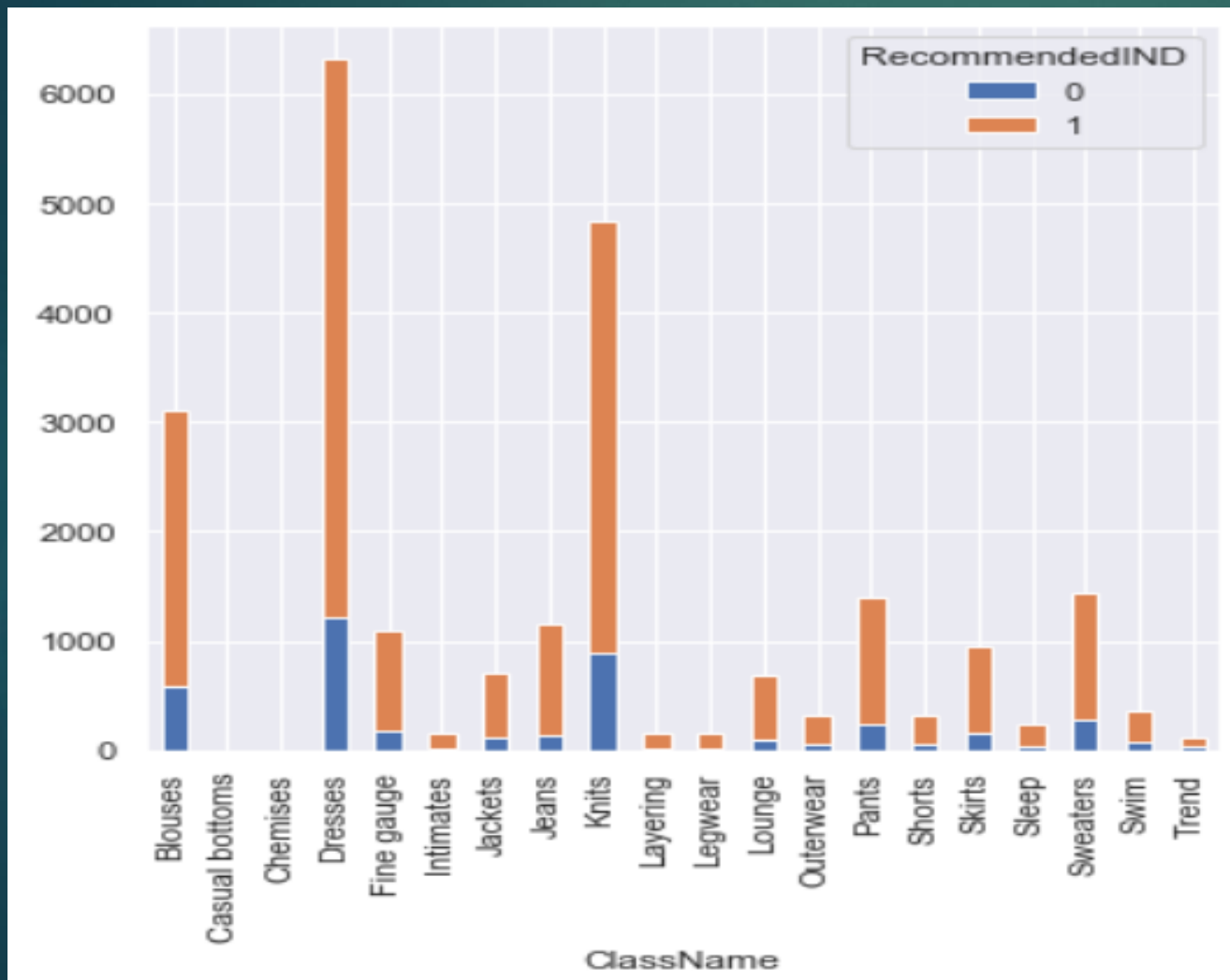
DivisionName	RecommendedIND	
General	0	18.317690
	1	81.682310
General Petite	0	17.401478
	1	82.598522
Initmates	0	14.780293
	1	85.219707

Polecenie produktu jako miara zadowolenia – departamenty:



DepartmentName	RecommendedIND	
Bottoms	0	14.872335
	1	85.127665
Dresses	0	19.180250
	1	80.819750
Intimate	0	14.985591
	1	85.014409
Jackets	0	16.375969
	1	83.624031
Tops	0	18.484906
	1	81.515094
Trend	0	26.050420
	1	73.949580

Polecenie produktu jako miara zadowolenia – klasy ubrań:



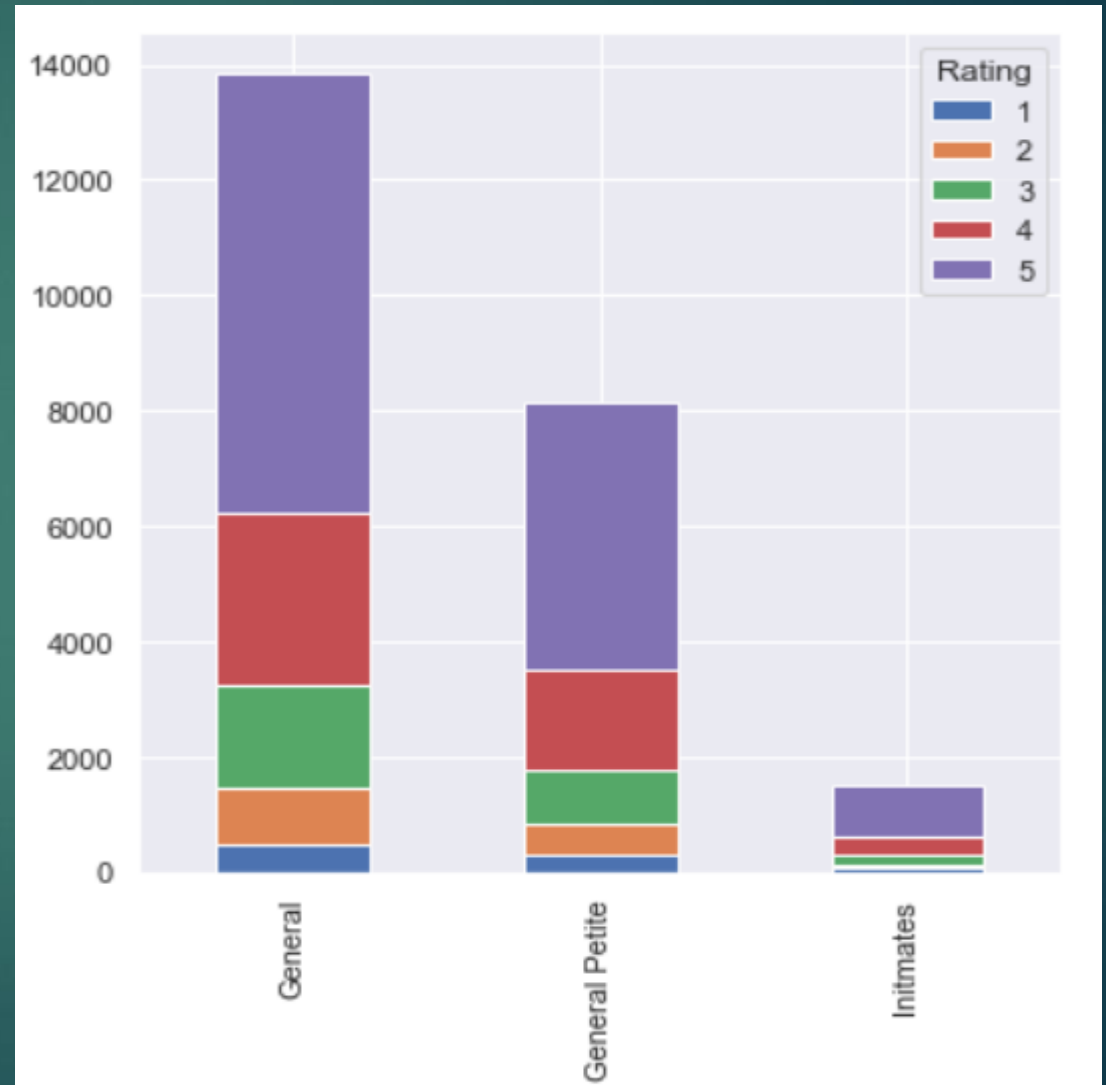
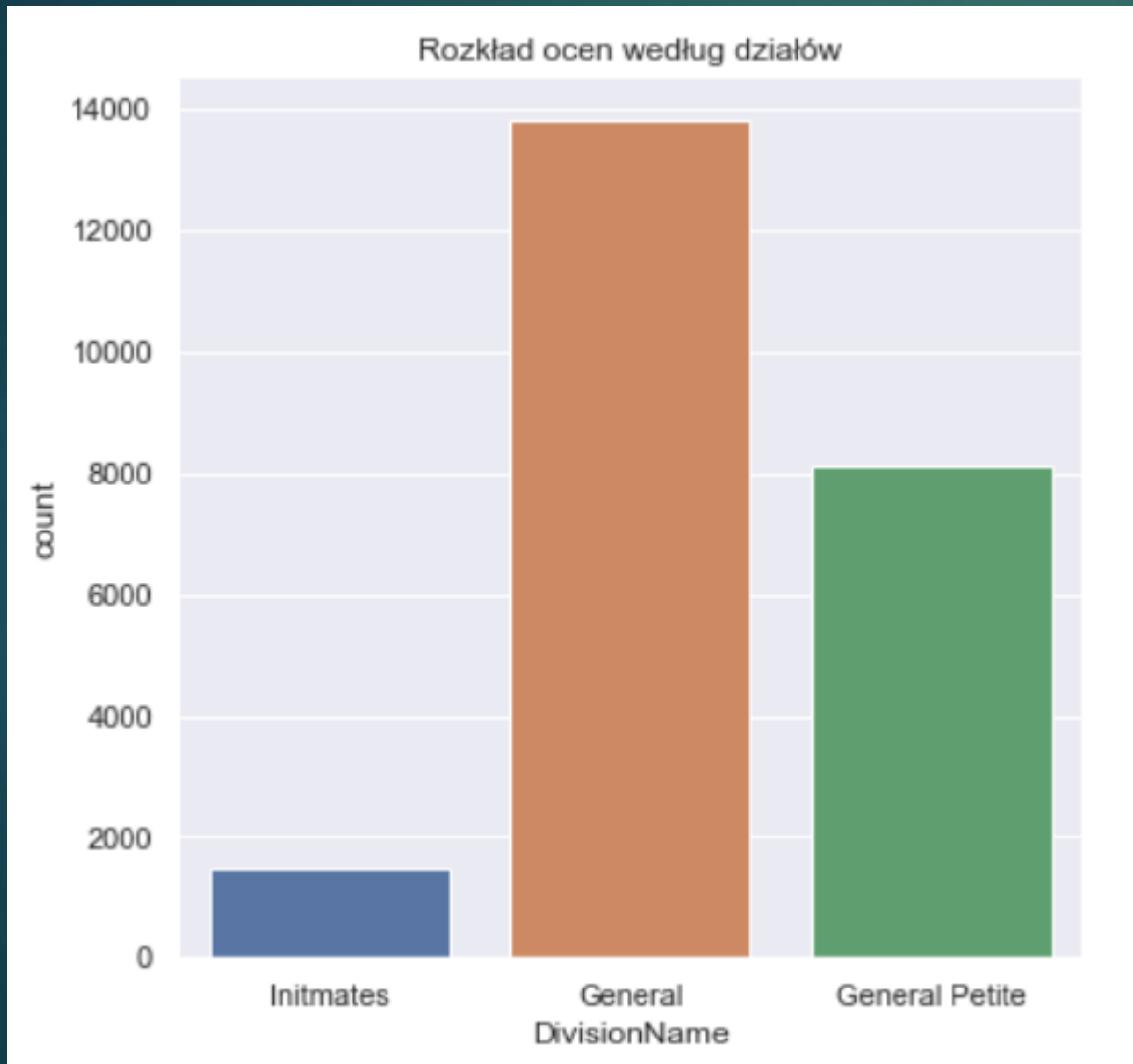
Trend	0	26.050420
	1	73.949580

Rozkład opinii według struktury produktów w sklepie:

- DZIAŁ
- DEPARTAMENT
- KLASA ODZIEŻY

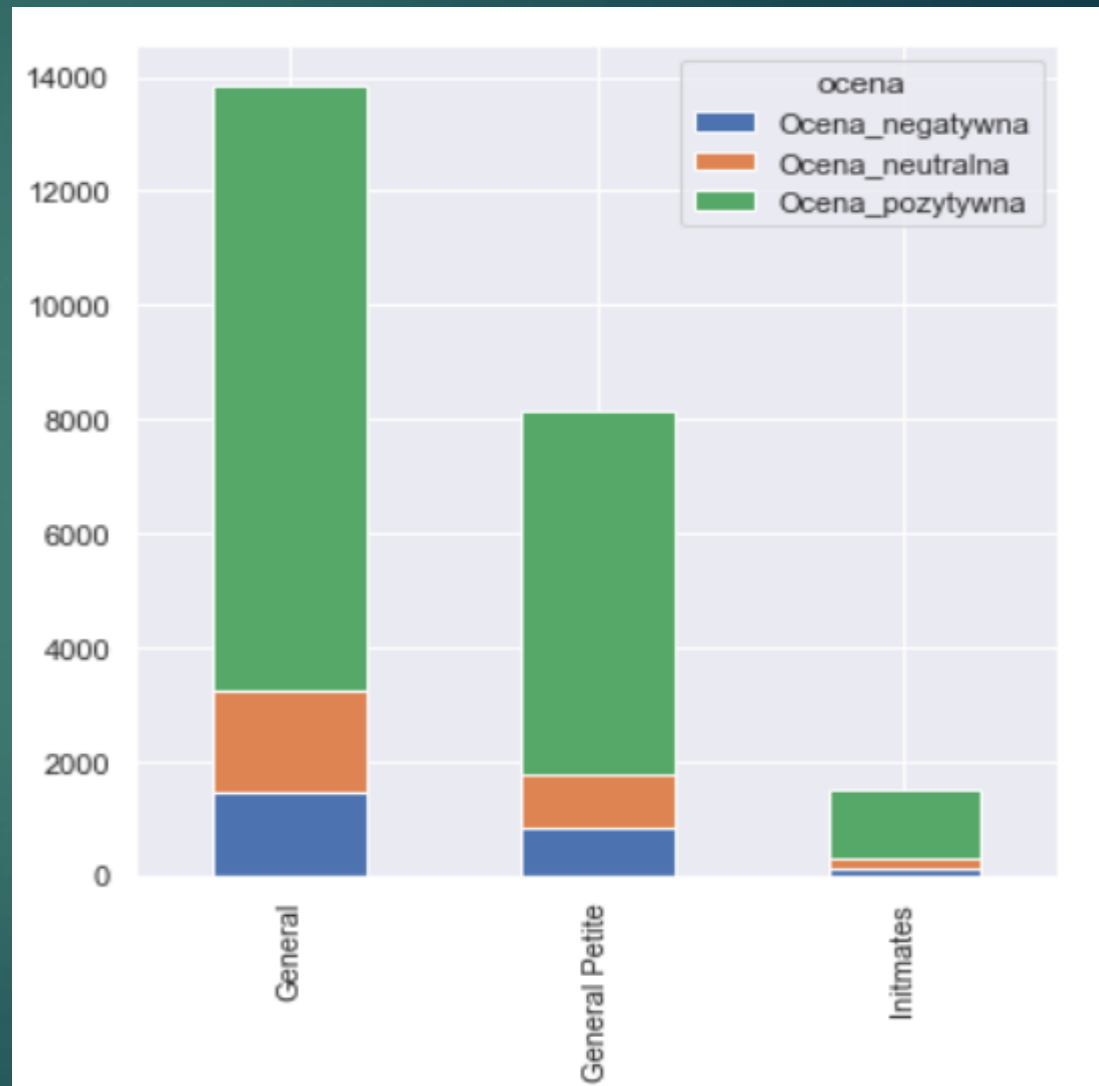
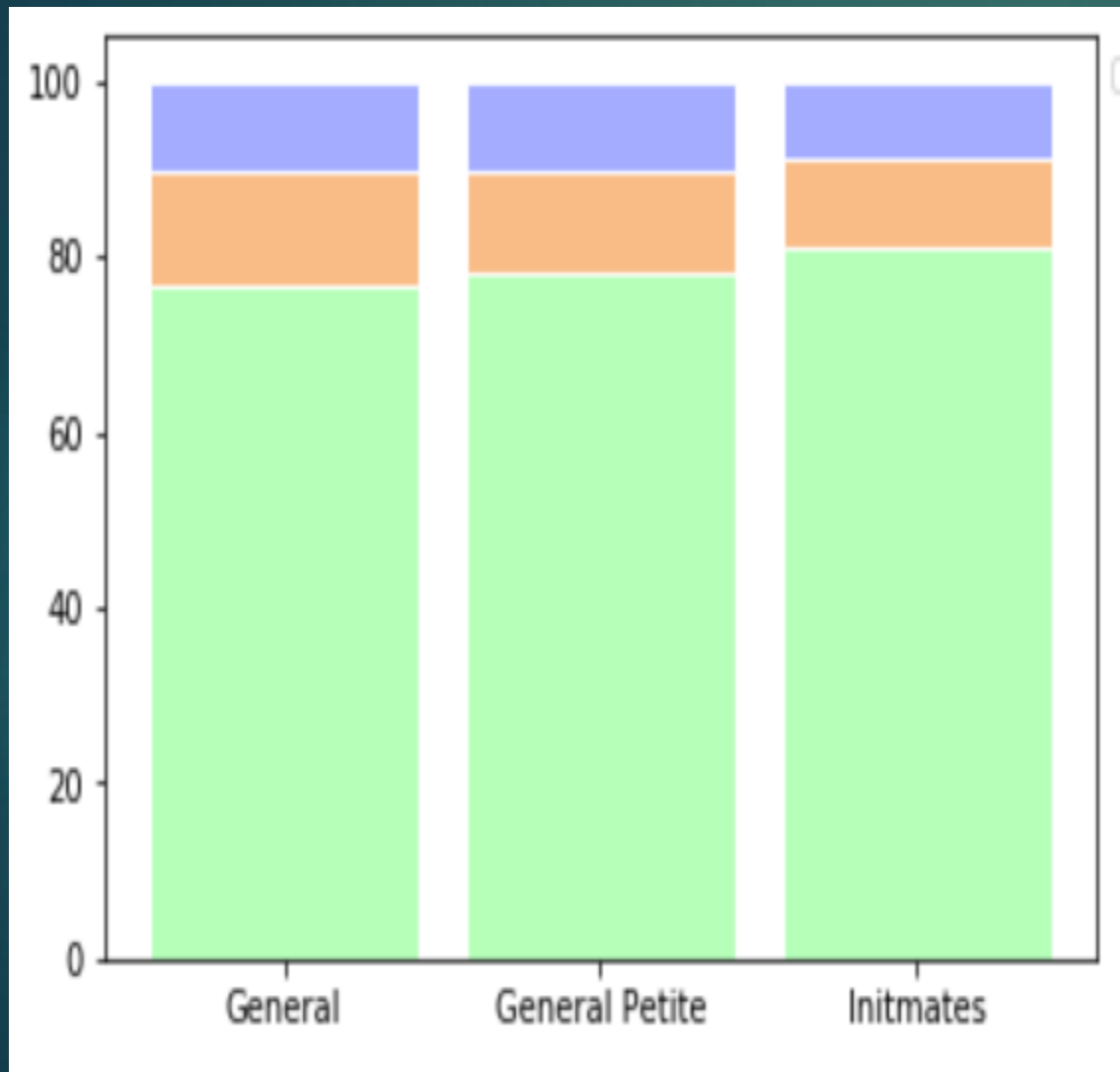
Rozkład opinii według struktury produktów w sklepie

DZIAŁY



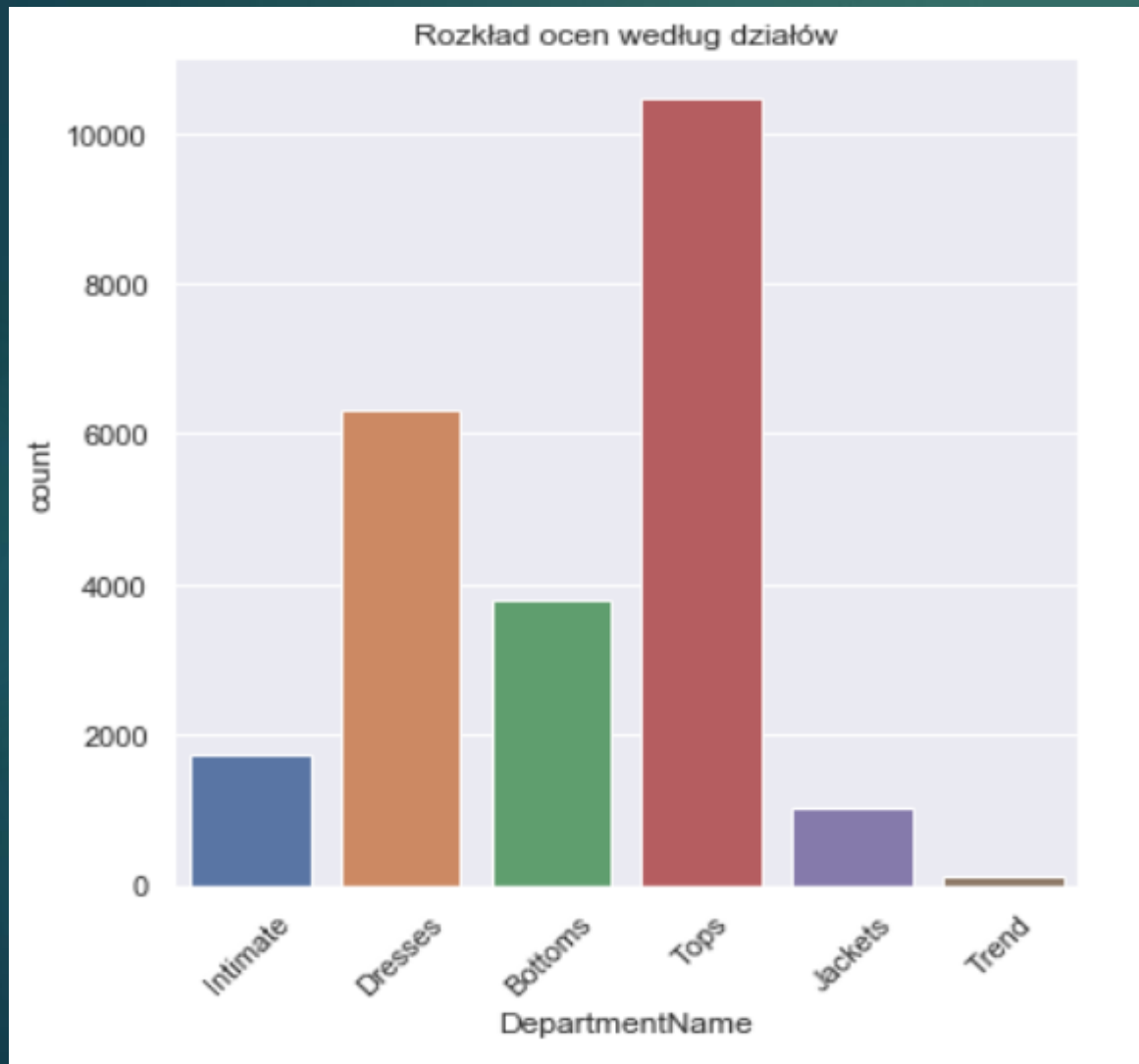
Rozkład opinii według struktury produktów w sklepie

DZIAŁY



Rozkład opinii według struktury produktów w sklepie

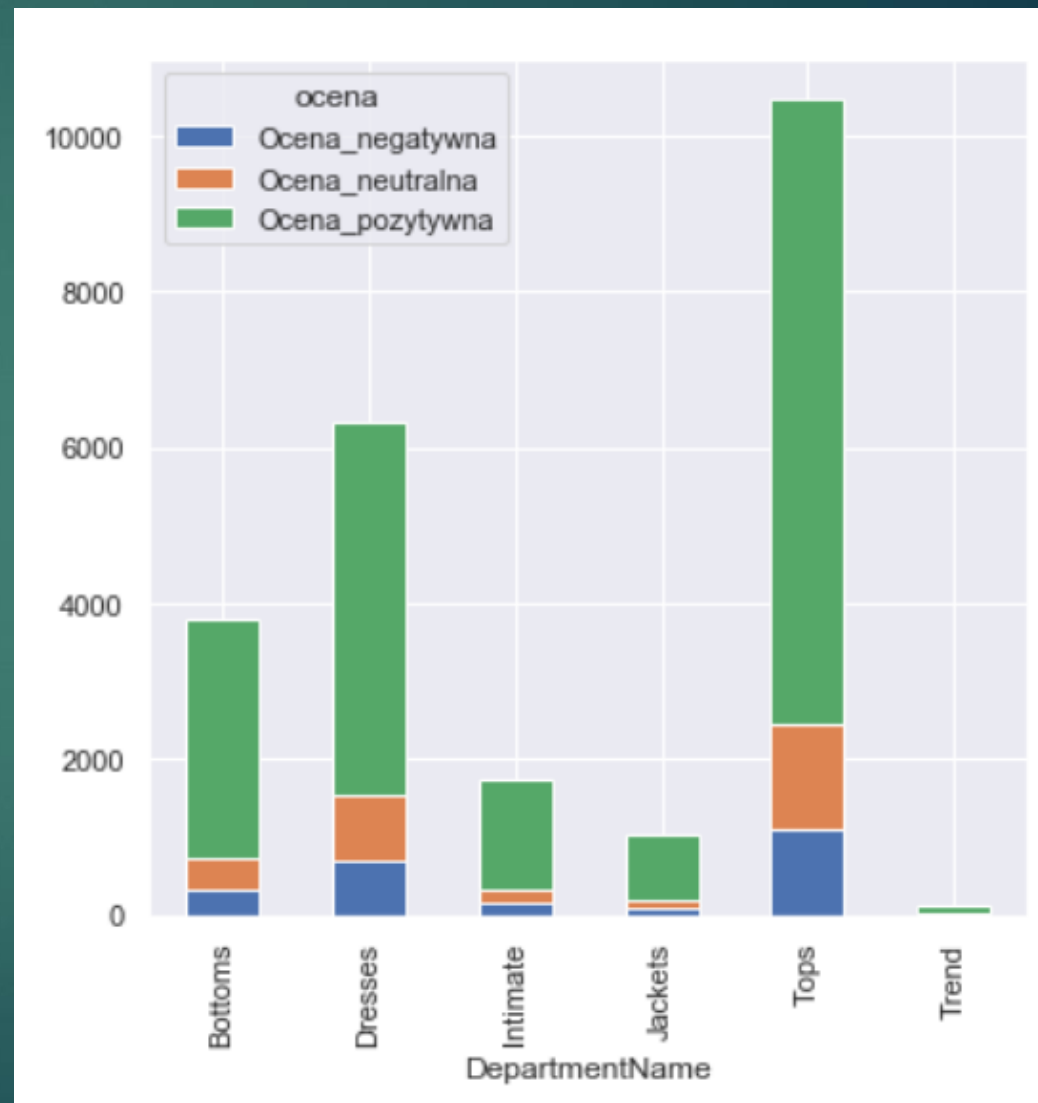
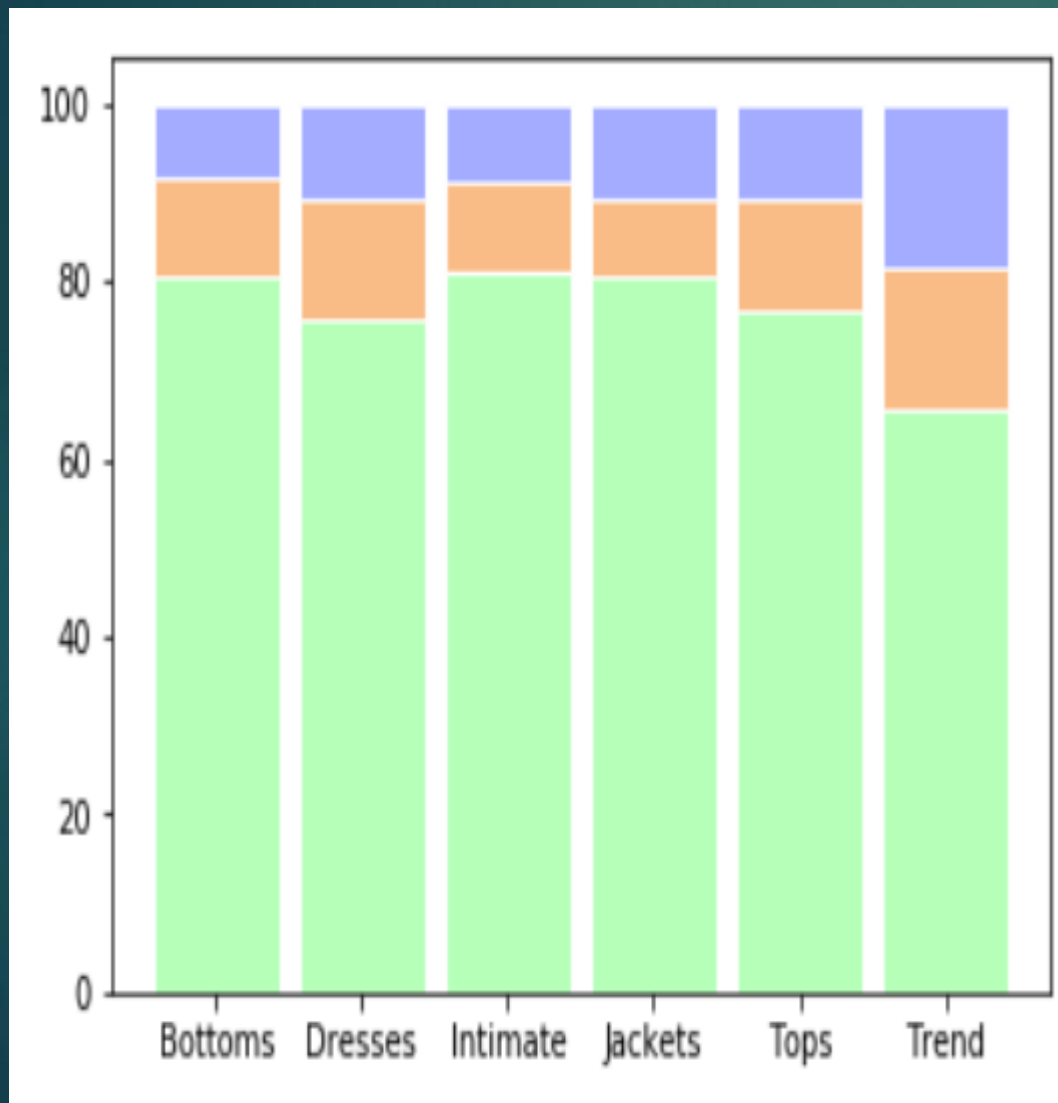
DEPARTAMENTY



	DepartmentName	Rating	Percentage
0	Bottoms	3799	16.185242
1	Dresses	6319	26.921438
2	Intimate	1735	7.391786
3	Jackets	1032	4.396728
4	Tops	10468	44.597819
5	Trend	119	0.506987

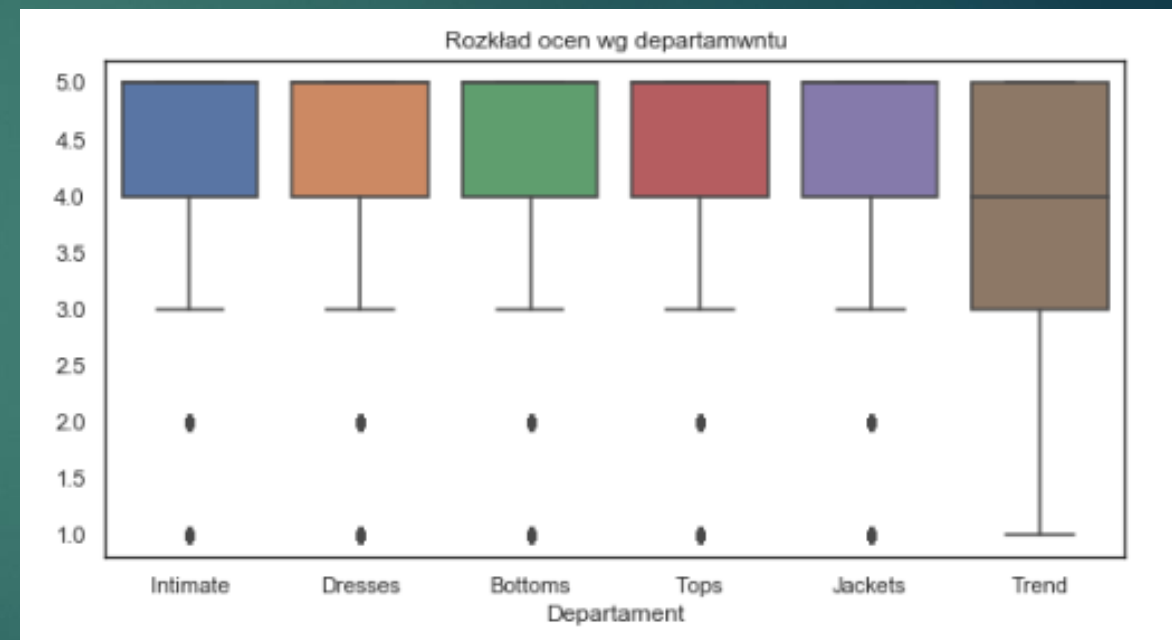
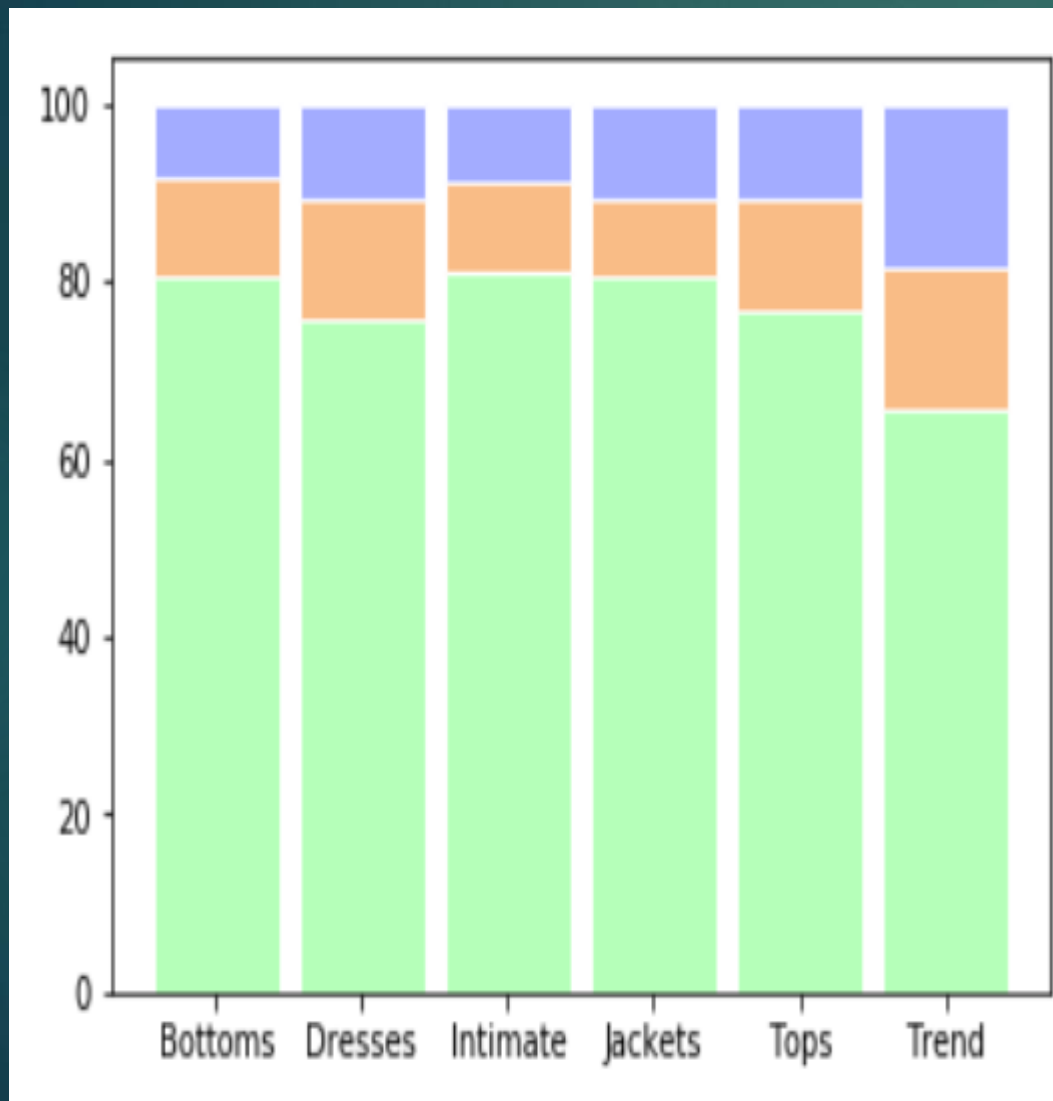
Rozkład opinii według struktury produktów w sklepie

DEPARTAMENTY



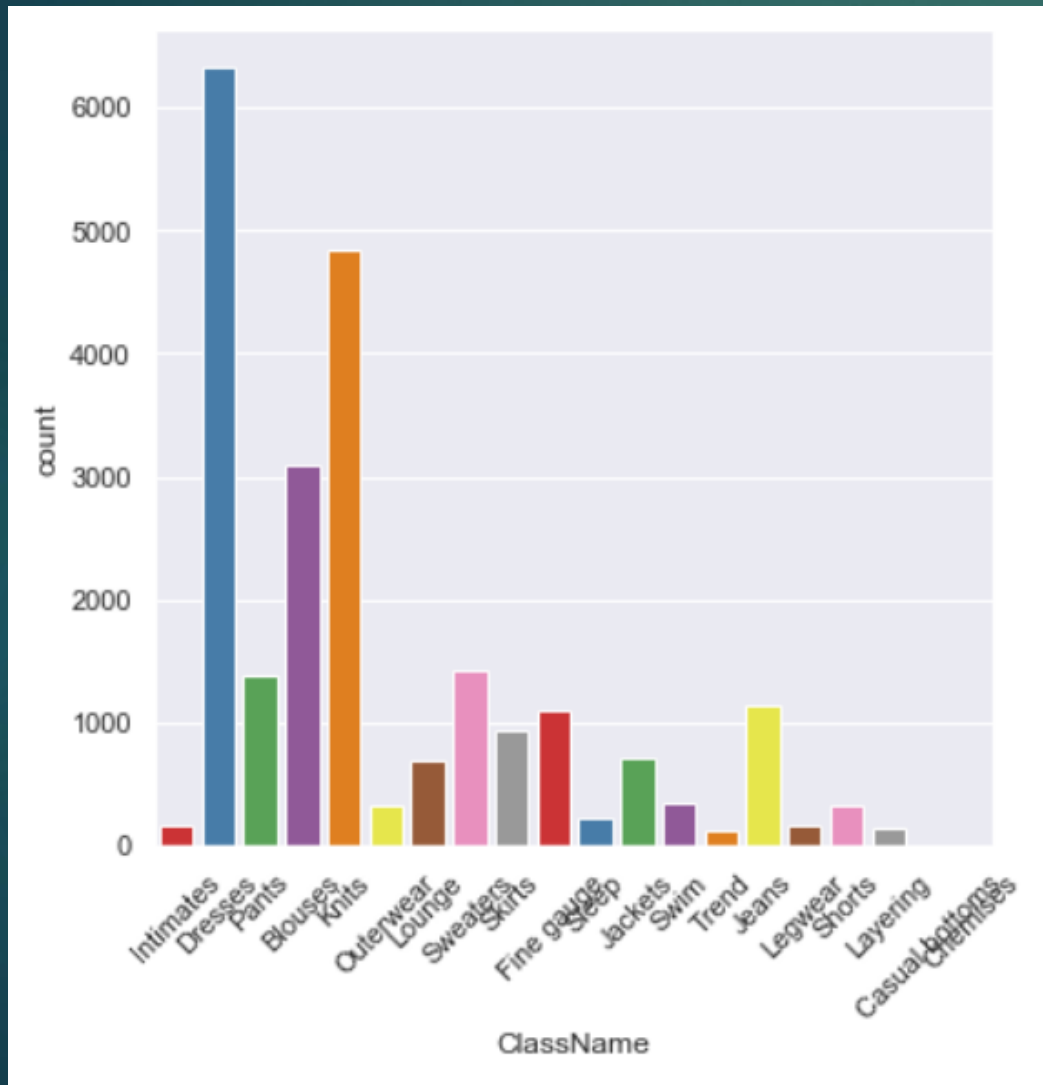
Rozkład opinii według struktury produktów w sklepie

DEPARTAMENTY



Rozkład opinii według struktury produktów w sklepie

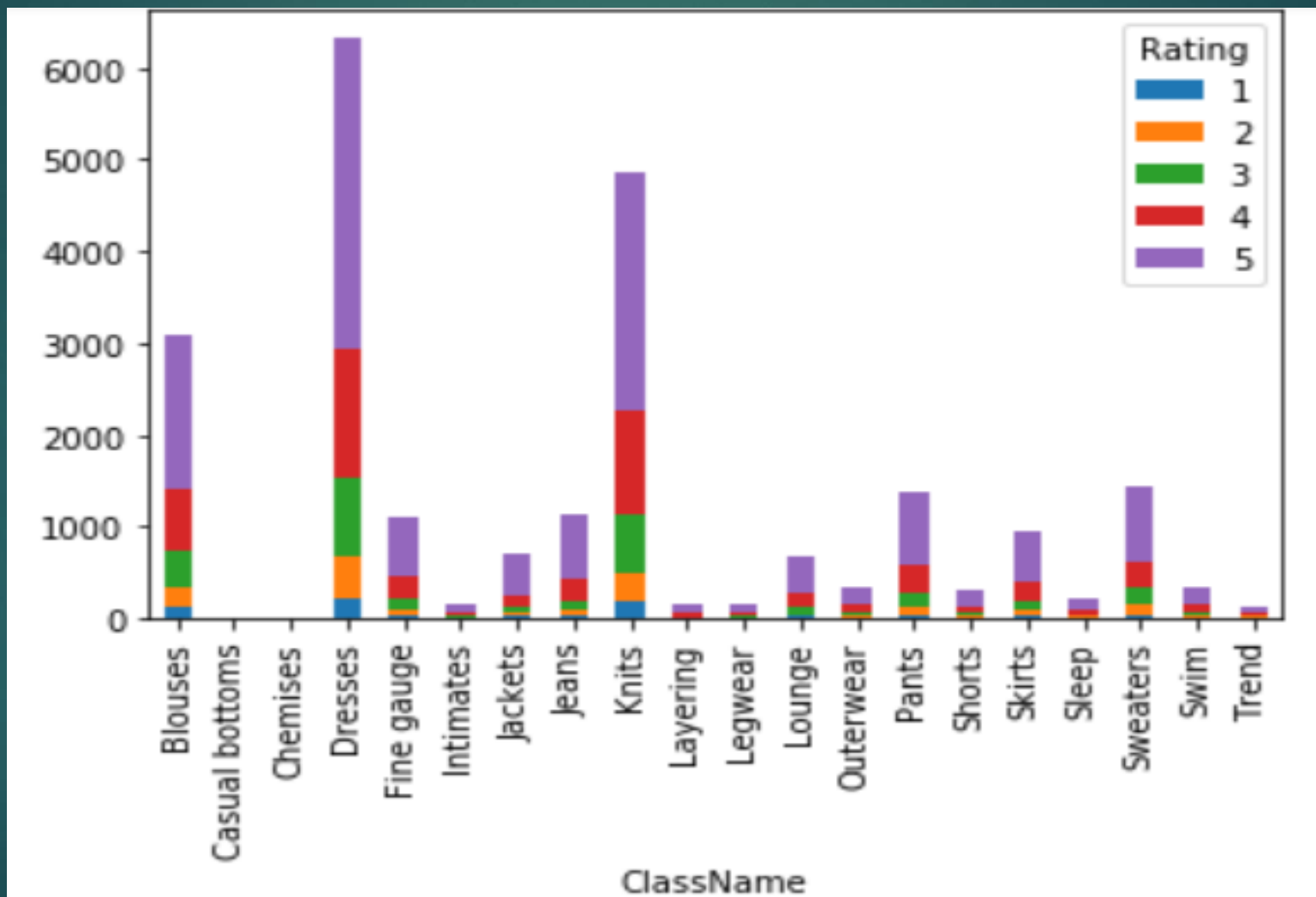
KLASY UBRAŃ



	ClassName	Rating	Percentage
0	Blouses	3097	13.194444
1	Casual bottoms	2	0.008521
2	Chemises	1	0.004260
3	Dresses	6319	26.921438
4	Fine gauge	1100	4.686435
5	Intimates	154	0.656101
6	Jackets	704	2.999318
7	Jeans	1147	4.886673
8	Knits	4843	20.633095
9	Layering	146	0.622018
10	Legwear	165	0.702965
11	Lounge	691	2.943933
12	Outerwear	328	1.397410
13	Pants	1388	5.913429
14	Shorts	317	1.350545
15	Skirts	945	4.026074
16	Sleep	228	0.971370
17	Sweaters	1428	6.083845
18	Swim	350	1.491138
19	Trend	119	0.506987

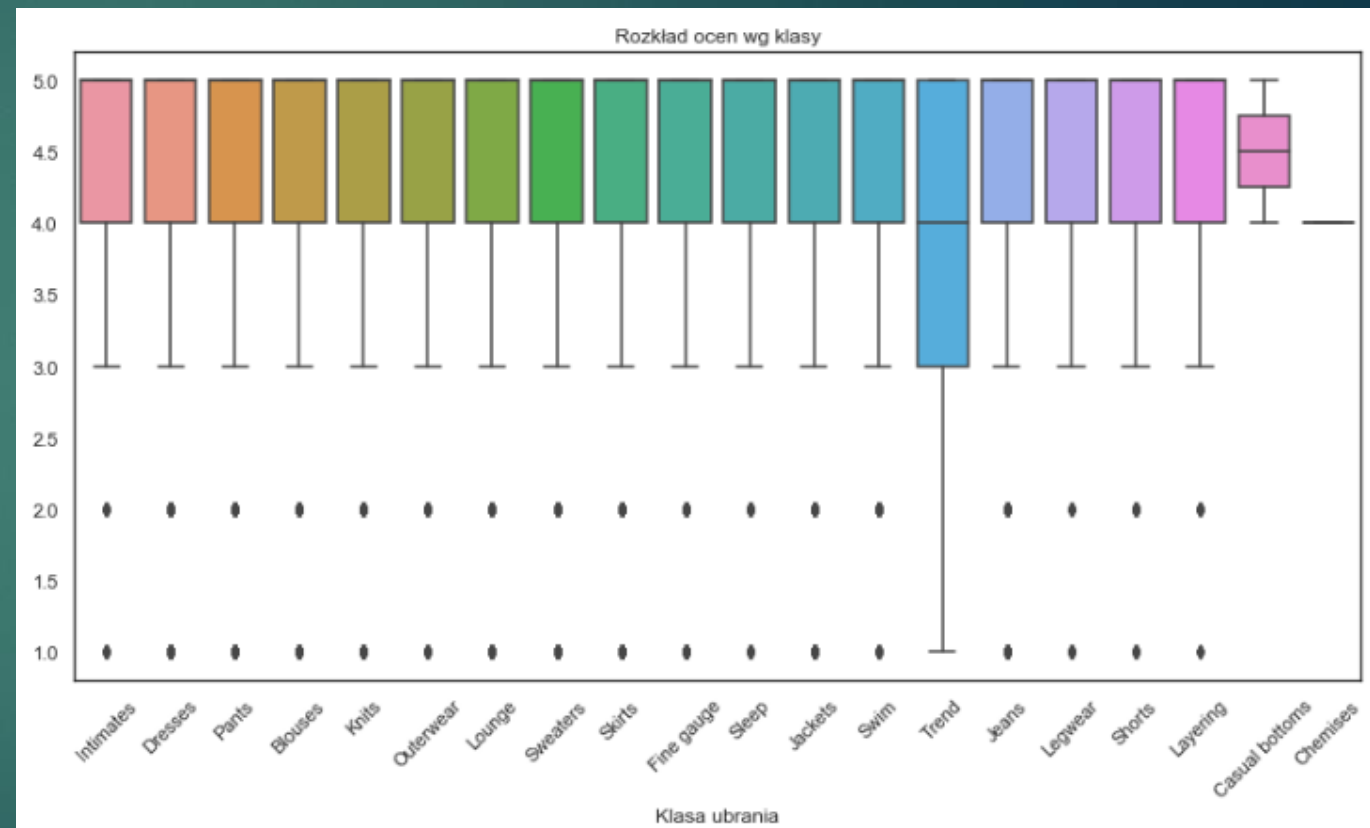
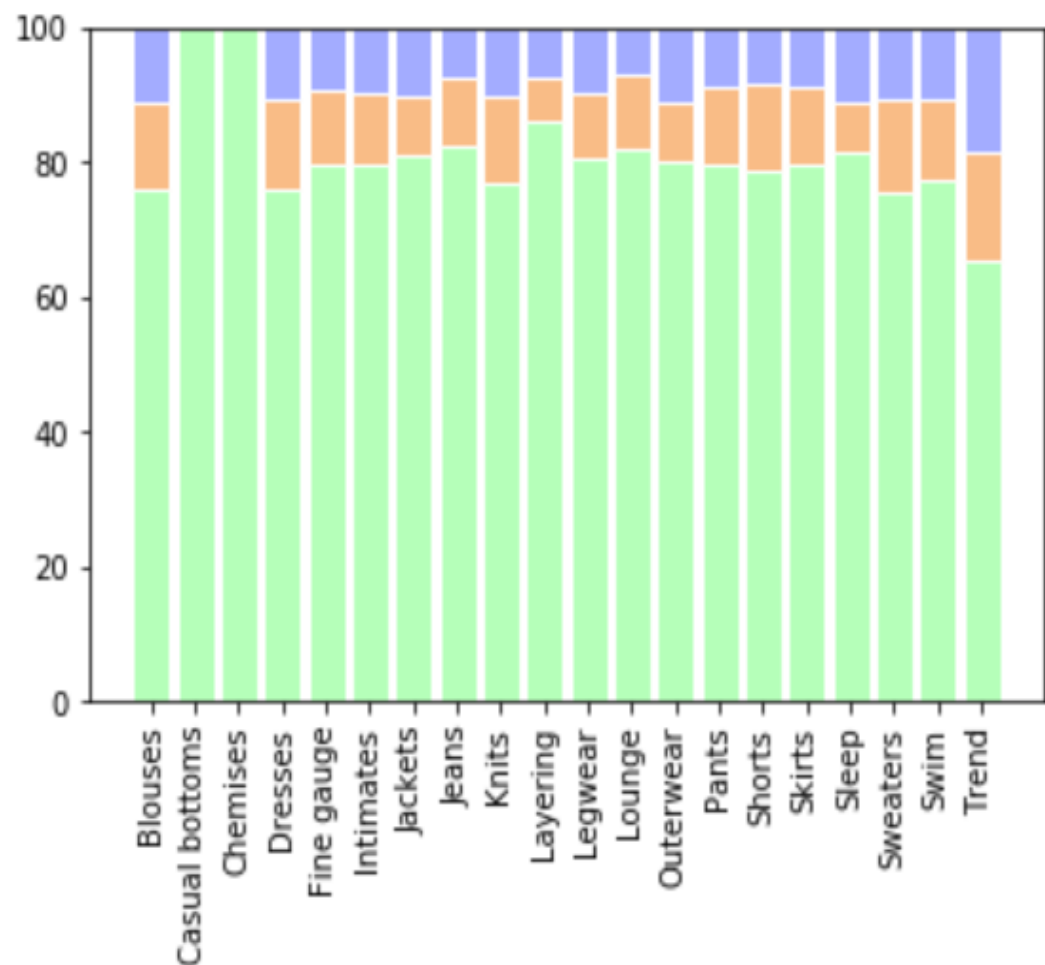
Rozkład opinii według struktury produktów w sklepie

KLASY UBRAŃ

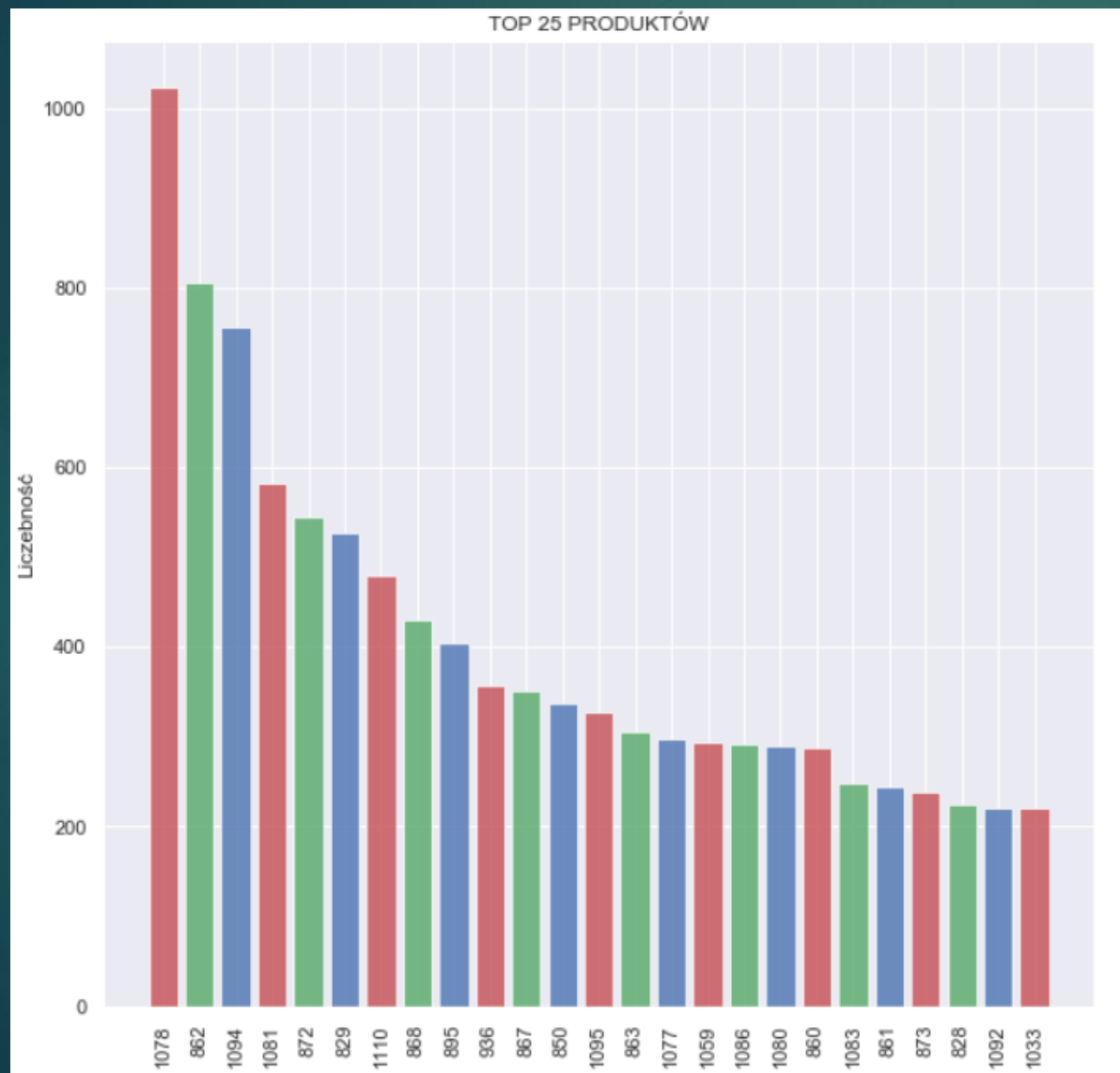


Rozkład opinii według struktury produktów w sklepie

KLASY UBRAŃ



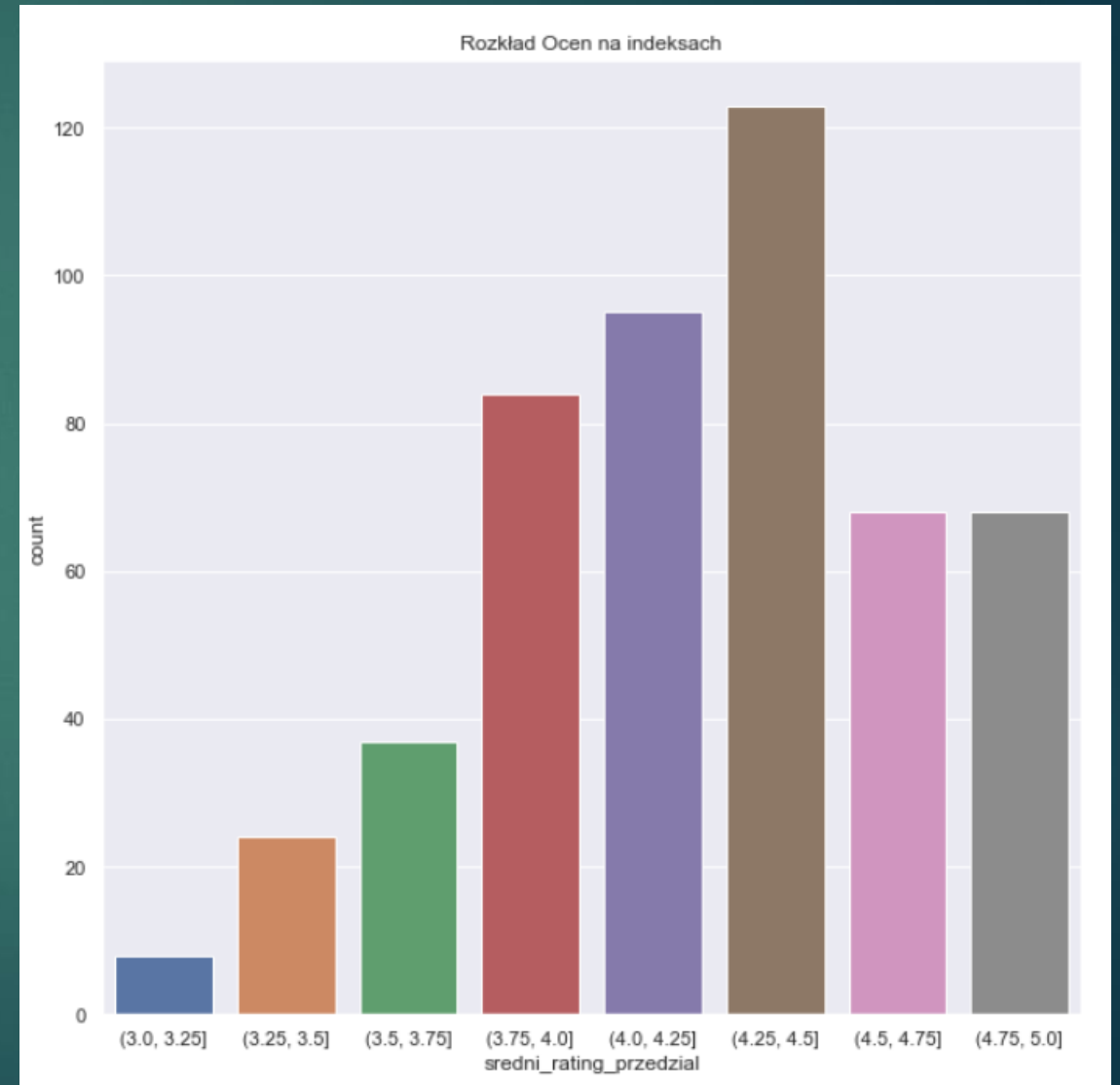
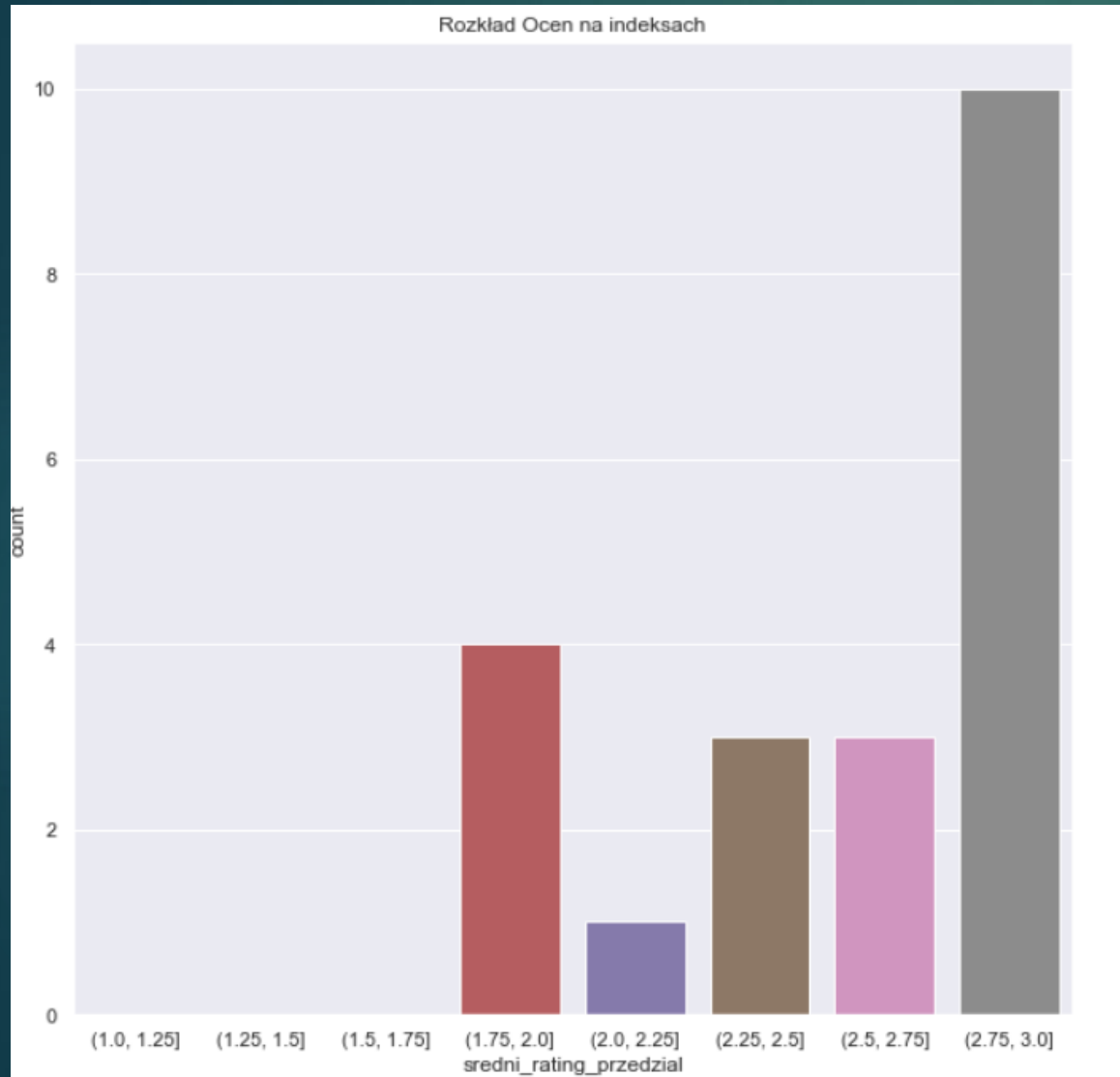
Clothing ID



	suma_rating	count	sredni_rating
count	1206.000000	1206.000000	1206.000000
mean	81.714760	19.474295	4.198540
std	291.167303	69.009764	0.849258
min	1.000000	1.000000	1.000000
25%	5.000000	1.000000	4.000000
50%	9.000000	2.000000	4.333333
75%	26.000000	6.750000	5.000000
max	4305.000000	1024.000000	5.000000

25 najczęściej
ocenianych detali ma
łącznie 10089 ocen
czyli 42% wszystkich
ocen

Clothing ID



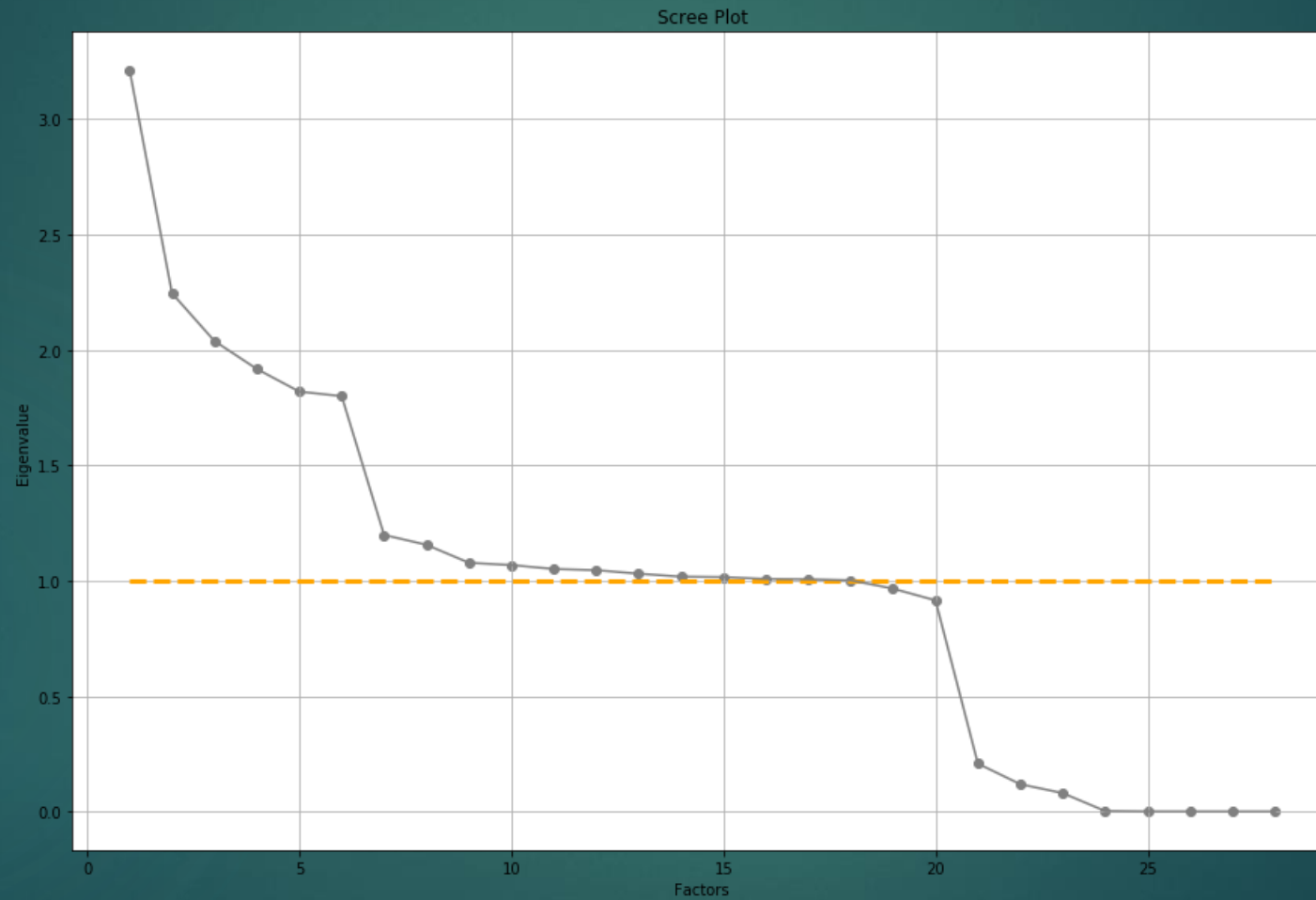
Analiza komentarzy

Analiza czynnikowa

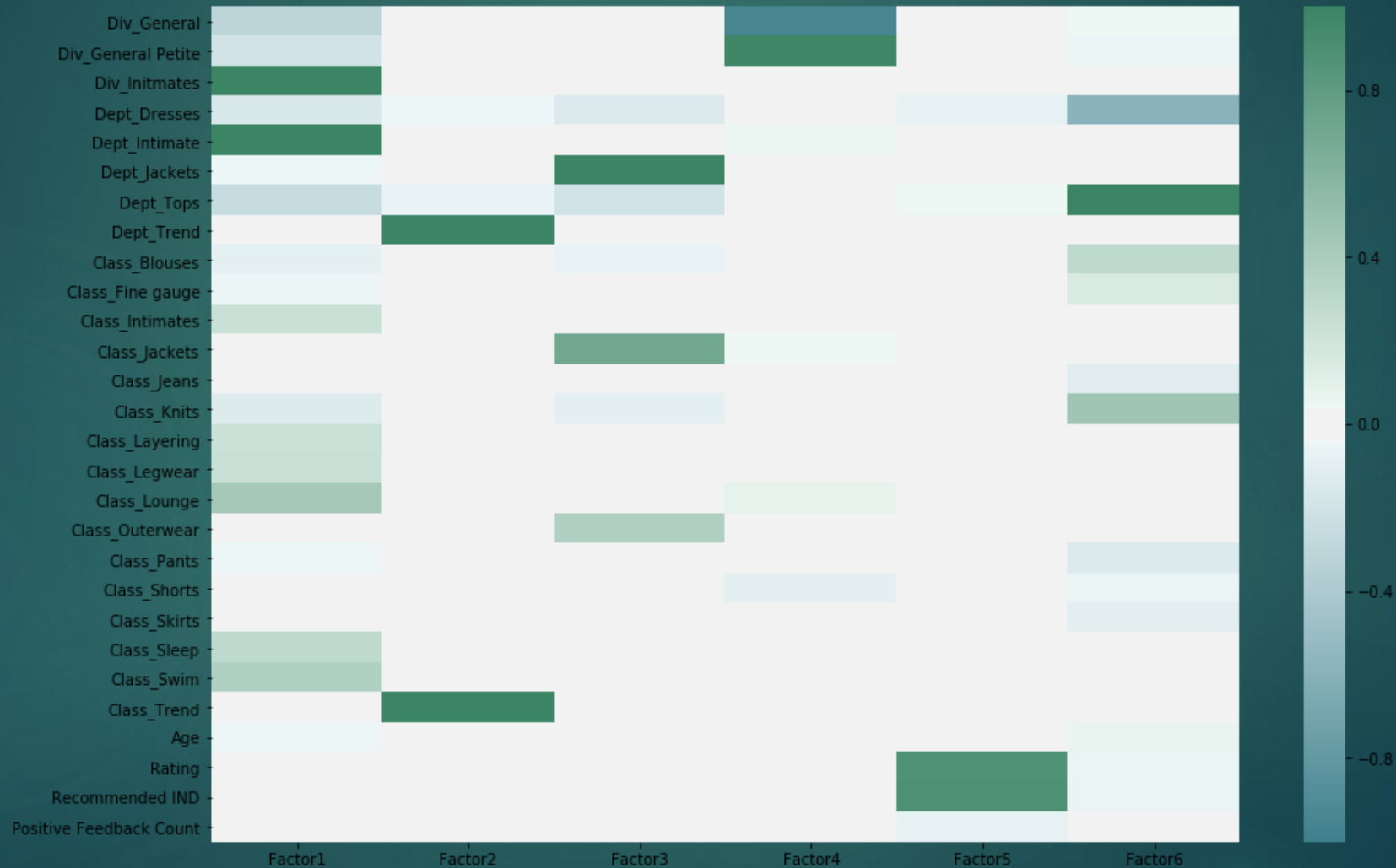
Przygotowanie danych

- Usunięcie danych z komentarzem oceny
- Stworzenie *dummy variables* ze zmiennych kategorycznych
- Test Bartlett'a: p-value = 0.0
- KMO Test: 0.12

Liczba czynników



Analiza czynników



Wariancja

	Factor1	Factor2	Factor3	Factor4	Factor5	Factor6
SS Loadings	2.881451	2.002965	1.804096	1.879699	1.597059	1.733448
Proportion Var	0.102909	0.071534	0.064432	0.067132	0.057038	0.061909
Cumulative Var	0.102909	0.174443	0.238875	0.306008	0.363045	0.424954