

Market basket analysis and association rules based on customer data

Анализ рыночной корзины и ассоциативные правила на основе клиентских данных

Students: Erik Golikov, Dmitrii Karikh

Эрик Голиков, Дмитрий Карих

Supervisor: Natalia Titova

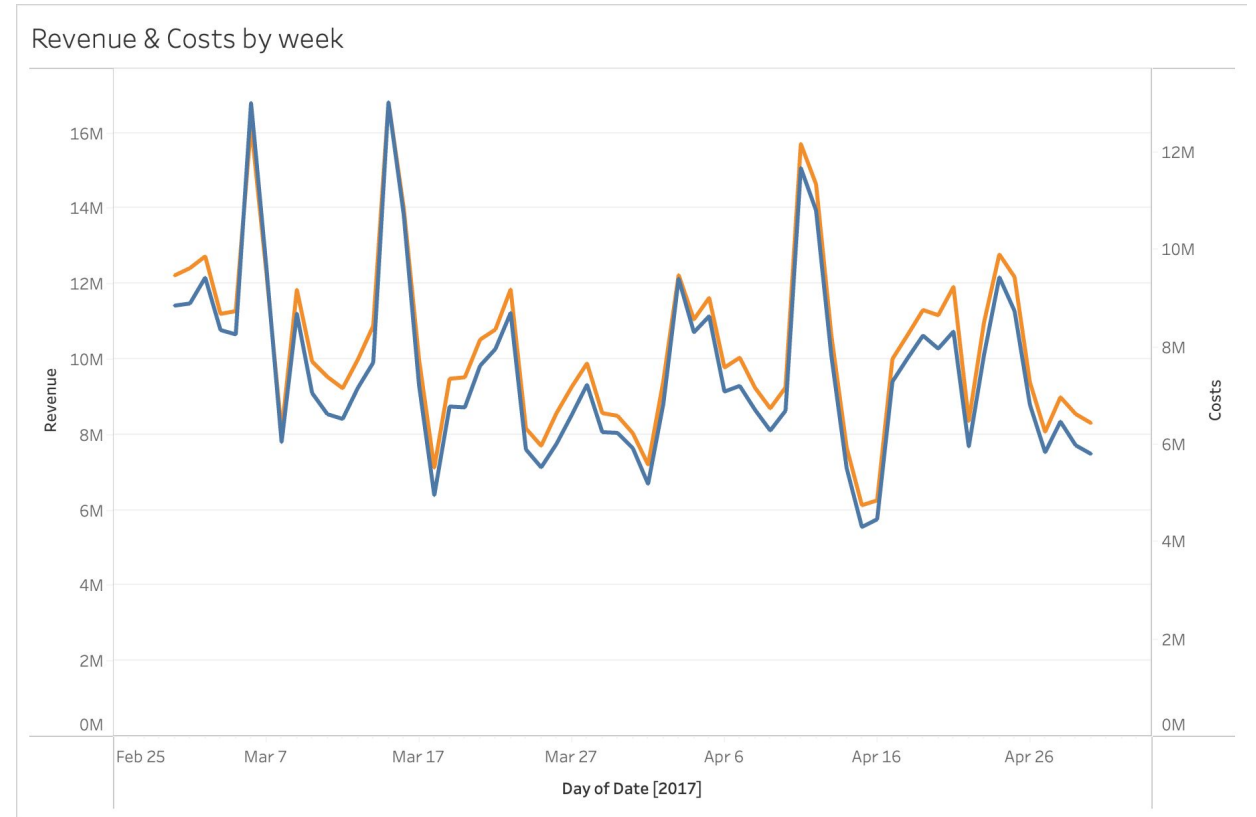
Customer analytics





Data overview

- Data collection: March 6 – 15, 2025 (all channels)
-
- Scope: >600 000 orders for $\approx 50\,000$ unique customers
- Goal: uncover purchase patterns to boost cross-sell/upsell





Data overview: Category

- + Women specials
- + Textile, Baby food, Related products
- Pet products, Diapers
- Nulls

Main_group	Revenue	avg Marg/Price	total Marg/Rev
КРУПНОГАБАРИТНЫЙ ТОВАР	167,019,991	0.30	0.31
ПОДГУЗНИКИ	70,625,380	0.05	0.09
ИГРУШКИ	68,869,602	0.22	0.26
ТЕКСТИЛЬ, ТРИКОТАЖ	56,906,910	0.35	0.36
ДЕТСКОЕ ПИТАНИЕ	43,317,332	0.15	0.29
ОБУВЬ	18,034,957	0.35	0.35
ТОВАРЫ ДЛЯ КОРМЛЕНИЯ	14,815,471	0.17	0.21
КОСМЕТИКА/ГИГИЕНА	10,273,550	0.19	0.29
Null	9,006,311	-	0.00
КАНЦТОВАРЫ, КНИГИ, ДИСКИ	4,412,372	0.23	0.37
ТОВАРЫ ДЛЯ ЖИВОТНЫХ	3,860,297	0.01	0.05
СОПУТСТВУЮЩИЕ ТОВАРЫ	378,898	0.35	0.39
ЖЕНСКИЕ ШТУЧКИ	123,693	0.58	0.55
ТЕХНИКА И ТОВАРЫ ДЛЯ ДОМА	36,931	0.27	0.23
Total / Avg	467681697.6	0.23	0.27





Data overview: SubSubCategory

- + Mashed dishes
- + Bikes/Scooters
- + Baby Cloth

Sub_group	Revenue	avg Marg/Price	total Marg/Rev
АКСЕССУАРЫ ДЛЯ ВОЛОС	21313	0,78	0,79
ТРУСИКИ НЕПРОМОКАЕМЫЕ	442208	0,31	0,74
ОДЕЖДА, ОБУВЬ	52540	0,64	0,69
АКСЕССУАРЫ МАКИЯЖ, МАНИКЮР	3015	0,56	0,65
СОПУТСТВУЮЩИЕ ТОВАРЫ ДЛЯ ОБУВИ	744003,71	0,49	0,59
СПОРТ.ИНВЕНТАРЬ	1712	0,58	0,58
КИСЛОМОЛОЧНЫЕ ПРОДУКТЫ	46956,4	0,18	0,58
АКСЕССУАРЫ ДЛЯ ДЕВОЧЕК	153901,5	0,47	0,55
НАПИТКИ	699748,95	0,21	0,49
ДЕТСКИЕ КОЛГОТКИ И НОСКИ	666975,19	0,46	0,49
СУВЕНИРНАЯ ПРОДУКЦИЯ	98	0,92	0,45
ПЮРЕ	10183744,06	0,21	0,45
ПОСТЕЛЬНОЕ БЕЛЬЕ	5552082,33	0,40	0,43
ВЕЛОСИПЕДЫ/САМОКАТЫ	13834454,57	0,44	0,43
ДЕТСКОЕ БЕЛЬЁ	1541825,58	0,40	0,42
КАТАЛКИ/КАЧАЛКИ	3144370,89	0,42	0,42
ОДЕЖДА ДЛЯ НОВОРОЖДЕННЫХ (0-2 лет)	10790849,95	0,41	0,41
КОРЗИНЫ, ЯЩИКИ ДЛЯ ИГРУШЕК	359947,07	0,40	0,41
КОЛЯСКИ ДЛЯ КУКОЛ	2370918,15	0,41	0,40





Data overview: SubCategory

- Diapers
- Dog's products

Sub_group	Revenue	avg Marg/Price	total Marg/Rev
ТОВАРЫ ДЛЯ ХОРЬКОВ	12561	0,09	0,09
ТОВАРЫ ДЛЯ РЫБ	11414,29	0,09	0,09
ПОДГУЗНИКИ	69038358,8	0,05	0,08
ВИТАМИНЫ/БАДЫ	2315	0,09	0,05
Null	9006311	-	0,00
ЗЕРКАЛА	989	-	0,00
ТОВАРЫ ДЛЯ СОБАК	1245050,31	(0,04)	-0,13
ТОВАРЫ ДЛЯ ЧЕРЕПАХ И РЕПТИЛИЙ	20776,25	(0,06)	-0,15





Data overview: Status

- Some statistics on order statuses
- We'll conduct analysis only for delivered orders

Статус	НовыйСтатус	count
В процессе	Принят в ПВЗ	4154
	Скомплектован Регион	3265
	В резерве	285
	Передан в Регион	10
	Отправлен в ПВЗ	7
	Комплектация Регион	1
Возврат	Возврат из ПВЗ	98966
	Частичный возврат	25107
	Возврат	1358
Доставлен	Доставлен	255084





Segmentation: Data preparation

- Deleted missed values
- Deleted irrelevant features (useless & duplicative)
- Deleted incomplete orders
- Deleted outliers using IsolationForest

Overall, we are left with $\approx 200k$ observations

Column	Unique
НомерЗаказаНаСайте	119792
НовыйСтатус	13
СуммаЗаказаНаСайте	12243
СуммаДокумента	12960
ФормаОплаты	2
Регион	457
Группа2	10
Группа3	59
Группа4	359
Тип	4
Номенклатура	57337
ТипТовара	2
Количество	63
Цена	13131
СуммаСтроки	13583
Гео	3
СуммаУслуг	118
СуммаДоставки	118
КоличествоПроданоКлиенту	63
Телефон_new	85512





Segmentation: Features

Телефо н_new	ЧислоЗ аказов	ОбщаяСум маПокупок	ОбщееКоличес твоТоваров	Средн ийЧек	СрКолТовар ов_вЧеке	G2_ДЕ ТСКОЕ ПИТАН ИЕ	G2_ИГ РУШК И	G2_КАНЦ ТОВАРЫ, КНИГИ, ДИСКИ	G2_КОС МЕТИКА/ ГИГИЕНА	G2_КРУПНОГА БАРИТНЫЙ ТОВАР	G2_О БУВЬ	G2_ПОДГ УЗНИКИ	G2_СОПУТС ТВУЮЩИЕ ТОВАРЫ	G2_ТЕК СТИЛЬ, ТРИКОТ АЖ	G2_ТО ВАРЫ ДЛЯ КОРМ ЛЕНИЯ
0	1	1	1039.0	5	1039.0	5.0	0	0	5	0	0	0	0	0	0
1	2	1	1630.0	1	1630.0	1.0	0	1	0	0	0	0	0	0	0
2	3	2	4801.0	12	2400.5	6.0	0	5	0	0	0	0	1	0	5

Final segmentation dataset composition

- Choice “Группа 2” (G2_*) as prime product decomposition because it combined just enough diversity with feasible computational cost (unlike either too generous “Тип” feature or too detailed “Группа 3/4”)
- Aggregated customers by phone number as unique ID
- Added aggregated features





Segmentation: Algorithms

k-means

Pros:

- fast computation given data size
- simple interpretation

Cons:

- sensitive to outliers
- might not capture mixed intercluster behaviour

GMM:

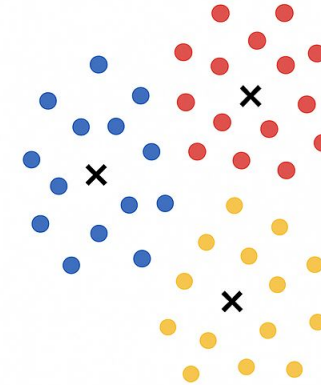
Pros:

- soft probabilistic cluster assignment
- Flexible covariance models feature correlations per segment

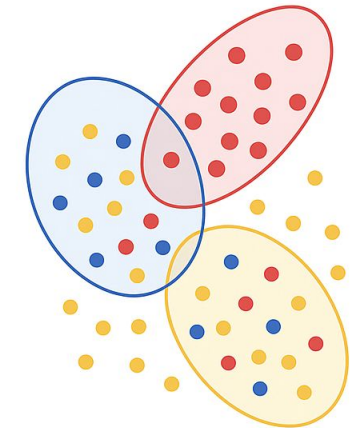
Cons:

- slow
- might not be efficient on sparse data

k-means



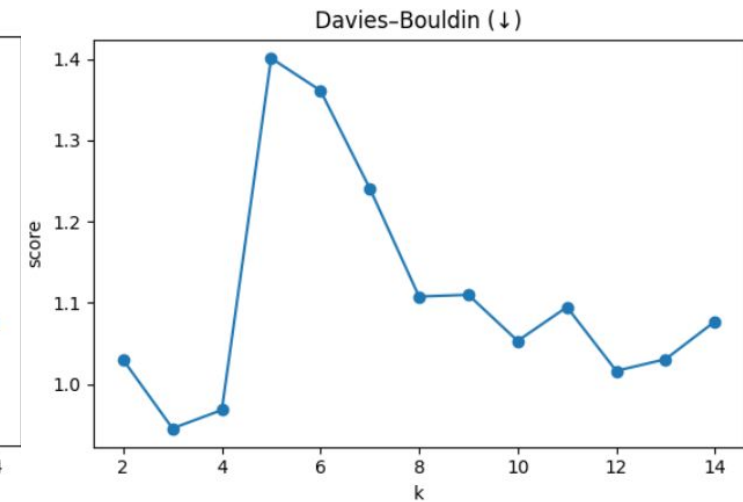
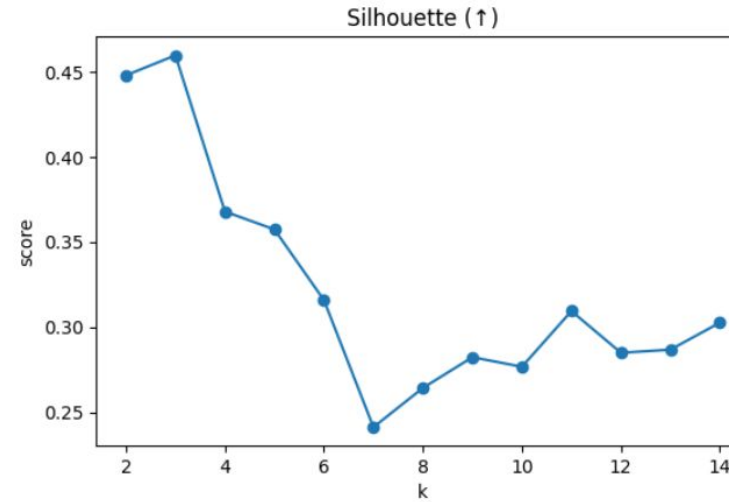
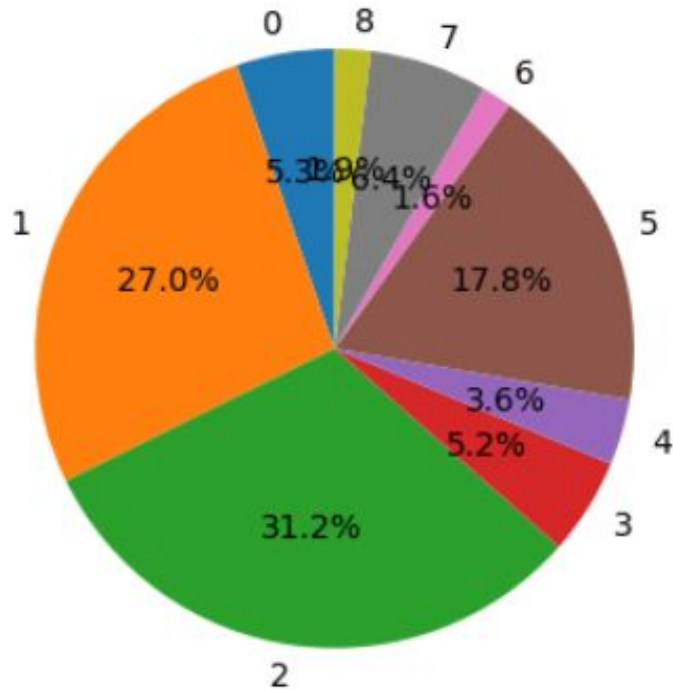
GMM





Segmentation: k-means

Распределение клиентов (k=9)



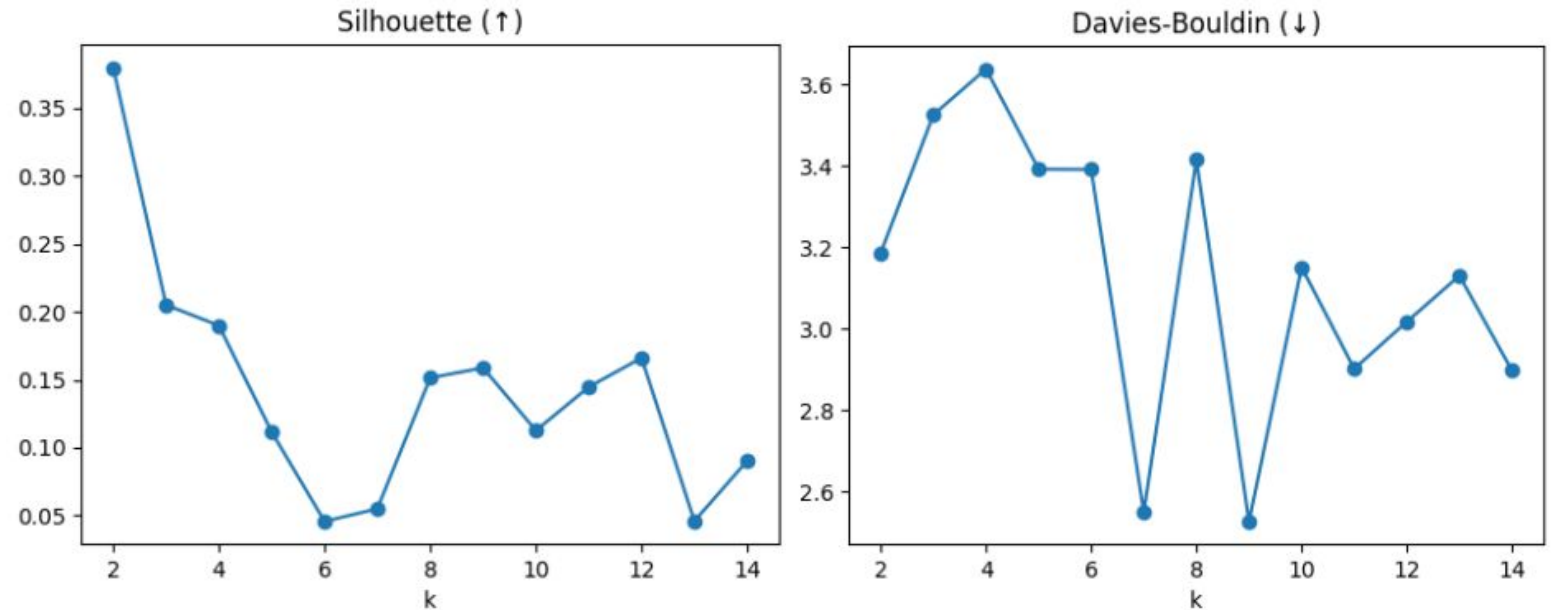
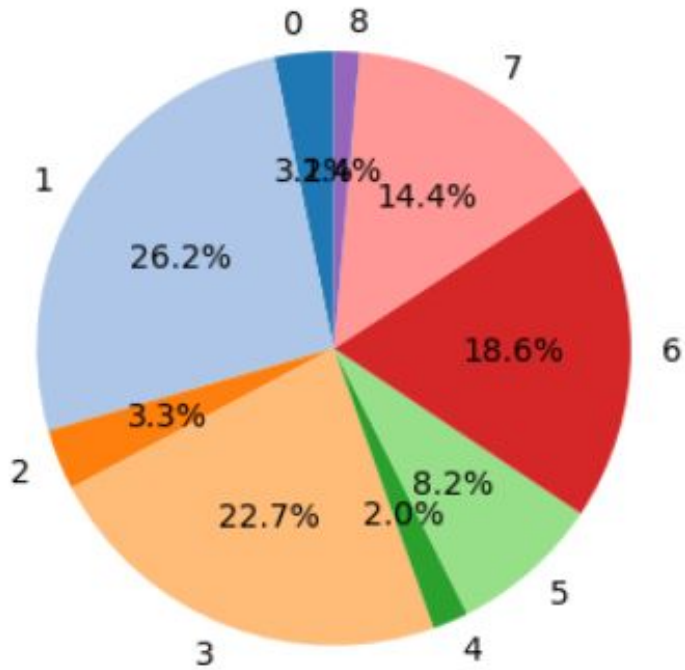
Based on Silhouette score = 0.282 and the future market basket analysis, the best option is to choose 9 customer segments





Segmentation: GMM

Распределение клиентов (GMM, k=9)



Based on Silhouette score = 0.159 and the future market basket analysis, the best option is to choose 9 customer segments





APRIORI

-An algorithm behind
"You may also like"



@HarshaManoj





MBA: Mandatory columns

- Half of clusters contain a unit support
- To see clear picture, we'll specify it and drop before MBA

support	itemsets
1.000000	(G2_ТЕКСТИЛЬ, ТРИКОТАЖ)
0.453782	(G2_ИГРУШКИ)
0.453782	(G2_ИГРУШКИ, G2_ТЕКСТИЛЬ, ТРИКОТАЖ)
0.257983	(G2_ПОДГУЗНИКИ, G2_ТЕКСТИЛЬ, ТРИКОТАЖ)
0.257983	(G2_ПОДГУЗНИКИ)
0.190756	(G2_КОСМЕТИКА/ГИГИЕНА, G2_ТЕКСТИЛЬ, ТРИКОТАЖ)
0.190756	(G2_КОСМЕТИКА/ГИГИЕНА)
0.177311	(G2_ТОВАРЫ ДЛЯ КОРМЛЕНИЯ, G2_ТЕКСТИЛЬ, ТРИКОТАЖ)
0.177311	(G2_ТОВАРЫ ДЛЯ КОРМЛЕНИЯ)
0.161345	(G2_ОБУВЬ)





MBA: Cluster 0

2574 users. Mandatory cat:
textile

Apriori:

- 60% buy toys
- 43% - diapers
- 1/3 buy feeding products or cosmetics

Cluster	Revenue	Avg Revenue
0	8,059,269	3131
1	24,768,135	1891
2	34,735,558	2293
3	6,563,097	2595
4	3,938,704	2281
5	23,749,965	2747
6	1,622,300	2083
7	37,850,558	12230
8	1,686,952	1795

support	itemsets
0.604119	(G2_ИГРУШКИ)
0.437071	(G2_ПОДГУЗНИКИ)
0.372998	(G2_КОСМЕТИКА/ГИГИЕНА)
0.320366	(G2_ТОВАРЫ ДЛЯ КОРМЛЕНИЯ)
0.208238	(G2_ИГРУШКИ, G2_ПОДГУЗНИКИ)
0.203661	(G2_КРУПНОГАБАРИТНЫЙ ТОВАР)
0.167048	(G2_КОСМЕТИКА/ГИГИЕНА, G2_ПОДГУЗНИКИ)
0.164760	(G2_ОБУВЬ)
0.157895	(G2_ИГРУШКИ, G2_ТОВАРЫ ДЛЯ КОРМЛЕНИЯ)
0.137300	(G2_ИГРУШКИ, G2_КОСМЕТИКА/ГИГИЕНА)

Association rules:

- Cosmetics + Diapers
- Toys + breastfeeding

antecedents	consequents	support	lift	confidence
(G2_КОСМЕТИКА/ГИГИЕНА)	(G2_ПОДГУЗНИКИ)	0.167048	1.024668	0.447853
(G2_ПОДГУЗНИКИ)	(G2_КОСМЕТИКА/ГИГИЕНА)	0.167048	1.024668	0.382199
(G2_ИГРУШКИ)	(G2_ТОВАРЫ ДЛЯ КОРМЛЕНИЯ)	0.157895	0.815828	0.261364
(G2_ТОВАРЫ ДЛЯ КОРМЛЕНИЯ)	(G2_ИГРУШКИ)	0.157895	0.815828	0.492857
(G2_КОСМЕТИКА/ГИГИЕНА)	(G2_ТОВАРЫ ДЛЯ КОРМЛЕНИЯ)	0.123570	1.034093	0.331288
(G2_ТОВАРЫ ДЛЯ КОРМЛЕНИЯ)	(G2_КОСМЕТИКА/ГИГИЕНА)	0.123570	1.034093	0.385714
(G2_ИГРУШКИ)	(G2_ОБУВЬ)	0.107551	1.080545	0.178030
(G2_ОБУВЬ)	(G2_ИГРУШКИ)	0.107551	1.080545	0.652778
(G2_ИГРУШКИ)	(G2_КАНЦТОВАРЫ, КНИГИ, ДИСКИ)	0.093822	1.190657	0.155303
(G2_КАНЦТОВАРЫ, КНИГИ, ДИСКИ)	(G2_ИГРУШКИ)	0.093822	1.190657	0.719298





MBA: Cluster 1

13097 users. Mandatory cat: toys

Apriori:

- 39% buy textile
- 1/3 buy diapers, feeding, books and disks

Cluster	Revenue	Avg Revenue
0	8,059,269	3131
1	24,768,135	1891
2	34,735,558	2293
3	6,563,097	2595
4	3,938,704	2281
5	23,749,965	2747
6	1,622,300	2083
7	37,850,558	12230
8	1,686,952	1795

support	itemsets
0.386435	(G2_ТЕКСТИЛЬ, ТРИКОТАЖ)
0.354890	(G2_ПОДГУЗНИКИ)
0.345426	(G2_ТОВАРЫ ДЛЯ КОРМЛЕНИЯ)
0.302839	(G2_КАНЦТОВАРЫ, КНИГИ, ДИСКИ)
0.274448	(G2_КОСМЕТИКА/ГИГИЕНА)
0.247634	(G2_КРУПНОГАБАРИТНЫЙ ТОВАР)
0.173502	(G2_ОБУВЬ)
0.105678	(G2_ТЕКСТИЛЬ, ТРИКОТАЖ, G2_ТОВАРЫ ДЛЯ КОРМЛЕНИЯ)
0.096215	(G2_ПОДГУЗНИКИ, G2_ТОВАРЫ ДЛЯ КОРМЛЕНИЯ)
0.094637	(G2_КОСМЕТИКА/ГИГИЕНА, G2_ПОДГУЗНИКИ)

Association rules:

- Cosmetics + diapers 10%
- shoes, textile 8.2%

antecedents	consequents	support	lift	confidence
(G2_КОСМЕТИКА/ГИГИЕНА)	(G2_ПОДГУЗНИКИ)	0.094637	0.971648	0.344828
(G2_ПОДГУЗНИКИ)	(G2_КОСМЕТИКА/ГИГИЕНА)	0.094637	0.971648	0.266667
(G2_ОБУВЬ)	(G2_ТЕКСТИЛЬ, ТРИКОТАЖ)	0.082019	1.223302	0.472727
(G2_ТЕКСТИЛЬ, ТРИКОТАЖ)	(G2_ОБУВЬ)	0.082019	1.223302	0.212245
(G2_КОСМЕТИКА/ГИГИЕНА)	(G2_ТОВАРЫ ДЛЯ КОРМЛЕНИЯ)	0.080442	0.848528	0.293103
(G2_ТОВАРЫ ДЛЯ КОРМЛЕНИЯ)	(G2_КОСМЕТИКА/ГИГИЕНА)	0.080442	0.848528	0.232877
(G2_КРУПНОГАБАРИТНЫЙ ТОВАР)	(G2_КАНЦТОВАРЫ, КНИГИ, ДИСКИ)	0.066246	0.883360	0.267516
(G2_КАНЦТОВАРЫ, КНИГИ, ДИСКИ)	(G2_КРУПНОГАБАРИТНЫЙ ТОВАР)	0.066246	0.883360	0.218750





MBA: Cluster 2 15149 users.

Apriori:

- $\frac{2}{3}$ textile
- 40% large products
- $\frac{1}{3}$ feeding and shoes

Cluster	Revenue	Avg Revenue
0	8,059,269	3131
1	24,768,135	1891
2	34,735,558	2293
3	6,563,097	2595
4	3,938,704	2281
5	23,749,965	2747
6	1,622,300	2083
7	37,850,558	12230
8	1,686,952	1795

support	itemsets
0.632919	(G2_ТЕКСТИЛЬ, ТРИКОТАЖ)
0.400829	(G2_КРУПНОГАБАРИТНЫЙ ТОВАР)
0.343991	(G2_ТОВАРЫ ДЛЯ КОРМЛЕНИЯ)
0.327413	(G2_ОБУВЬ)
0.248076	(G2_ОБУВЬ, G2_ТЕКСТИЛЬ, ТРИКОТАЖ)
0.193606	(G2_КОСМЕТИКА/ГИГИЕНА)
0.187685	(G2_КРУПНОГАБАРИТНЫЙ ТОВАР, G2_ТЕКСТИЛЬ, ТРИКОТАЖ)
0.136767	(G2_ТОВАРЫ ДЛЯ КОРМЛЕНИЯ, G2_ТЕКСТИЛЬ, ТРИКОТАЖ)
0.125518	(G2_ИГРУШКИ)
0.122558	(G2_КРУПНОГАБАРИТНЫЙ ТОВАР, G2_ТОВАРЫ ДЛЯ КОРМЛЕНИЯ)

Association rules:

- shoes + textile
- large + feeding products

antecedents	consequents	support	lift	confidence
(G2_ОБУВЬ)	(G2_ТЕКСТИЛЬ, ТРИКОТАЖ)	0.248076	1.197129	0.757685
(G2_ТЕКСТИЛЬ, ТРИКОТАЖ)	(G2_ОБУВЬ)	0.248076	1.197129	0.391955
(G2_КРУПНОГАБАРИТНЫЙ ТОВАР)	(G2_ТОВАРЫ ДЛЯ КОРМЛЕНИЯ)	0.122558	0.888864	0.305761
(G2_ТОВАРЫ ДЛЯ КОРМЛЕНИЯ)	(G2_КРУПНОГАБАРИТНЫЙ ТОВАР)	0.122558	0.888864	0.356282
(G2_КОСМЕТИКА/ГИГИЕНА)	(G2_ТОВАРЫ ДЛЯ КОРМЛЕНИЯ)	0.101835	1.529094	0.525994
(G2_ТОВАРЫ ДЛЯ КОРМЛЕНИЯ)	(G2_КОСМЕТИКА/ГИГИЕНА)	0.101835	1.529094	0.296041
(G2_ИГРУШКИ)	(G2_ТЕКСТИЛЬ, ТРИКОТАЖ)	0.076377	0.961404	0.608491
(G2_ТЕКСТИЛЬ, ТРИКОТАЖ)	(G2_ИГРУШКИ)	0.076377	0.961404	0.120674
(G2_КРУПНОГАБАРИТНЫЙ ТОВАР)	(G2_ИГРУШКИ)	0.061575	1.223879	0.153619
(G2_ИГРУШКИ)	(G2_КРУПНОГАБАРИТНЫЙ ТОВАР)	0.061575	1.223879	0.490566





MBA: Cluster 3

2529 users. Mandatory cat:
toys

Apriori:

- 41.5% for both books, disks and textile categories

Cluster	Revenue	Avg Revenue
0	8,059,269	3131
1	24,768,135	1891
2	34,735,558	2293
3	6,563,097	2595
4	3,938,704	2281
5	23,749,965	2747
6	1,622,300	2083
7	37,850,558	12230
8	1,686,952	1795

support	itemsets
0.415648	(G2_КАНЦТОВАРЫ, КНИГИ, ДИСКИ)
0.413203	(G2_ТЕКСТИЛЬ, ТРИКОТАЖ)
0.403423	(G2_ПОДГУЗНИКИ)
0.300733	(G2_ТОВАРЫ ДЛЯ КОРМЛЕНИЯ)
0.283619	(G2_КОСМЕТИКА/ГИГИЕНА)
0.215159	(G2_КРУПНОГАБАРИТНЫЙ ТОВАР)
0.151589	(G2_КАНЦТОВАРЫ, КНИГИ, ДИСКИ, G2_ТЕКСТИЛЬ, ТРИКОТАЖ)
0.134474	(G2_КОСМЕТИКА/ГИГИЕНА, G2_ПОДГУЗНИКИ)
0.114914	(G2_ПОДГУЗНИКИ, G2_ТЕКСТИЛЬ, ТРИКОТАЖ)
0.102689	(G2_ОБУВЬ)

Association rules:

- Books, Disks + Textile
- Diapers + Cosmetics

antecedents	consequents	support	lift	confidence
(G2_КАНЦТОВАРЫ, КНИГИ, ДИСКИ)	(G2_ТЕКСТИЛЬ, ТРИКОТАЖ)	0.151589	0.882631	0.364706
(G2_ТЕКСТИЛЬ, ТРИКОТАЖ)	(G2_КАНЦТОВАРЫ, КНИГИ, ДИСКИ)	0.151589	0.882631	0.366864
(G2_КОСМЕТИКА/ГИГИЕНА)	(G2_ПОДГУЗНИКИ)	0.134474	1.175287	0.474138
(G2_ПОДГУЗНИКИ)	(G2_КОСМЕТИКА/ГИГИЕНА)	0.134474	1.175287	0.333333
(G2_ТОВАРЫ ДЛЯ КОРМЛЕНИЯ)	(G2_КОСМЕТИКА/ГИГИЕНА)	0.100244	1.175287	0.333333
(G2_КОСМЕТИКА/ГИГИЕНА)	(G2_ТОВАРЫ ДЛЯ КОРМЛЕНИЯ)	0.100244	1.175287	0.353448
(G2_ТОВАРЫ ДЛЯ КОРМЛЕНИЯ)	(G2_ПОДГУЗНИКИ)	0.097800	0.806110	0.325203
(G2_ПОДГУЗНИКИ)	(G2_ТОВАРЫ ДЛЯ КОРМЛЕНИЯ)	0.097800	0.806110	0.242424
(G2_КРУПНОГАБАРИТНЫЙ ТОВАР)	(G2_КАНЦТОВАРЫ, КНИГИ, ДИСКИ)	0.085575	0.956885	0.397727
(G2_КАНЦТОВАРЫ, КНИГИ, ДИСКИ)	(G2_КРУПНОГАБАРИТНЫЙ ТОВАР)	0.085575	0.956885	0.205882





MBA: Cluster 4

1727 users. Mandatory cat:
cosmetics

Apriori:

- 2/3 for Diapers
- 55% feeding products

Cluster	Revenue	Avg Revenue
0	8,059,269	3131
1	24,768,135	1891
2	34,735,558	2293
3	6,563,097	2595
4	3,938,704	2281
5	23,749,965	2747
6	1,622,300	2083
7	37,850,558	12230
8	1,686,952	1795

support	itemsets
0.666667	(G2_ПОДГУЗНИКИ)
0.552469	(G2_ТОВАРЫ ДЛЯ КОРМЛЕНИЯ)
0.365741	(G2_ИГРУШКИ)
0.344136	(G2_ПОДГУЗНИКИ, G2_ТОВАРЫ ДЛЯ КОРМЛЕНИЯ)
0.305556	(G2_ТЕКСТИЛЬ, ТРИКОТАЖ)
0.212963	(G2_ИГРУШКИ, G2_ПОДГУЗНИКИ)
0.202160	(G2_КРУПНОГАБАРИТНЫЙ ТОВАР)
0.154321	(G2_ПОДГУЗНИКИ, G2_ТЕКСТИЛЬ, ТРИКОТАЖ)
0.143519	(G2_ДЕТСКОЕ ПИТАНИЕ)
0.131173	(G2_ТЕКСТИЛЬ, ТРИКОТАЖ, G2_ТОВАРЫ ДЛЯ КОРМЛЕНИЯ)

Association rules:

- feeding + diapers
- toys + diapers

	antecedents	consequents	support	lift	confidence
0	(G2_ТОВАРЫ ДЛЯ КОРМЛЕНИЯ)	(G2_ПОДГУЗНИКИ)	0.344136	0.934358	0.622905
1	(G2_ПОДГУЗНИКИ)	(G2_ТОВАРЫ ДЛЯ КОРМЛЕНИЯ)	0.344136	0.934358	0.516204
2	(G2_ПОДГУЗНИКИ)	(G2_ИГРУШКИ)	0.212963	0.873418	0.319444
3	(G2_ИГРУШКИ)	(G2_ПОДГУЗНИКИ)	0.212963	0.873418	0.582278
4	(G2_ТОВАРЫ ДЛЯ КОРМЛЕНИЯ)	(G2_КРУПНОГАБАРИТНЫЙ ТОВАР)	0.106481	0.953388	0.192737
5	(G2_КРУПНОГАБАРИТНЫЙ ТОВАР)	(G2_ТОВАРЫ ДЛЯ КОРМЛЕНИЯ)	0.106481	0.953388	0.526718
6	(G2_ИГРУШКИ)	(G2_ТЕКСТИЛЬ, ТРИКОТАЖ)	0.092593	0.828539	0.253165
7	(G2_ТЕКСТИЛЬ, ТРИКОТАЖ)	(G2_ИГРУШКИ)	0.092593	0.828539	0.303030
8	(G2_ПОДГУЗНИКИ)	(G2_ДЕТСКОЕ ПИТАНИЕ)	0.091049	0.951613	0.136574
9	(G2_ДЕТСКОЕ ПИТАНИЕ)	(G2_ПОДГУЗНИКИ)	0.091049	0.951613	0.634409





MBA: Cluster 5

8646 users. Mandatory cat:

Cluster	Revenue	Avg Revenue
0	8,059,269	3131
1	24,768,135	1891
2	34,735,558	2293
3	6,563,097	2595
4	3,938,704	2281
5	23,749,965	2747
6	1,622,300	2083
7	37,850,558	12230
8	1,686,952	1795

diapers

Apriori:

- every category < 50%
- feeding, cosmetics and toys at the top

ort	itemsets
0.486811	(G2_ТОВАРЫ ДЛЯ КОРМЛЕНИЯ)
0.442446	(G2_КОСМЕТИКА/ГИГИЕНА)
0.442446	(G2_ИГРУШКИ)
0.287770	(G2_ТЕКСТИЛЬ, ТРИКОТАЖ)
0.232614	(G2_КРУПНОГАБАРИТНЫЙ ТОВАР)
0.199041	(G2_КОСМЕТИКА/ГИГИЕНА, G2_ТОВАРЫ ДЛЯ КОРМЛЕНИЯ)
0.163070	(G2_ИГРУШКИ, G2_ТОВАРЫ ДЛЯ КОРМЛЕНИЯ)
0.155875	(G2_ДЕТСКОЕ ПИТАНИЕ)
0.121103	(G2_ИГРУШКИ, G2_КОСМЕТИКА/ГИГИЕНА)
0.103118	(G2_ТЕКСТИЛЬ, ТРИКОТАЖ, G2_ТОВАРЫ ДЛЯ КОРМЛЕНИЯ)

Association rules:

- cosmetics + feeding
- toys + textile

antecedents	consequents	support	lift	confidence
(G2_КОСМЕТИКА/ГИГИЕНА)	(G2_ТОВАРЫ ДЛЯ КОРМЛЕНИЯ)	0.199041	0.924106	0.449864
(G2_ТОВАРЫ ДЛЯ КОРМЛЕНИЯ)	(G2_КОСМЕТИКА/ГИГИЕНА)	0.199041	0.924106	0.408867
(G2_ИГРУШКИ)	(G2_ТЕКСТИЛЬ, ТРИКОТАЖ)	0.101918	0.800474	0.230352
(G2_ТЕКСТИЛЬ, ТРИКОТАЖ)	(G2_ИГРУШКИ)	0.101918	0.800474	0.354167
(G2_КОСМЕТИКА/ГИГИЕНА)	(G2_ДЕТСКОЕ ПИТАНИЕ)	0.070743	1.025766	0.159892
(G2_ДЕТСКОЕ ПИТАНИЕ)	(G2_КОСМЕТИКА/ГИГИЕНА)	0.070743	1.025766	0.453846
(G2_ИГРУШКИ)	(G2_КАНЦТОВАРЫ, КНИГИ, ДИСКИ)	0.050360	1.171936	0.113821
(G2_КАНЦТОВАРЫ, КНИГИ, ДИСКИ)	(G2_ИГРУШКИ)	0.050360	1.171936	0.518519





MBA: Cluster 6

779 users. Mandatory cat:
baby food

Apriori:

- 65%! for diapers
- 44% cosmetics
- 38% toys

Cluster	Revenue	Avg Revenue
0	8,059,269	3131
1	24,768,135	1891
2	34,735,558	2293
3	6,563,097	2595
4	3,938,704	2281
5	23,749,965	2747
6	1,622,300	2083
7	37,850,558	12230
8	1,686,952	1795

support	itemsets
0.650206	(G2_ПОДГУЗНИКИ)
0.440329	(G2_КОСМЕТИКА/ГИГИЕНА)
0.386831	(G2_ИГРУШКИ)
0.341564	(G2_ТОВАРЫ ДЛЯ КОРМЛЕНИЯ)
0.275720	(G2_КОСМЕТИКА/ГИГИЕНА, G2_ПОДГУЗНИКИ)
0.226337	(G2_ТЕКСТИЛЬ, ТРИКОТАЖ)
0.205761	(G2_ИГРУШКИ, G2_ПОДГУЗНИКИ)
0.185185	(G2_ПОДГУЗНИКИ, G2_ТОВАРЫ ДЛЯ КОРМЛЕНИЯ)
0.115226	(G2_ИГРУШКИ, G2_КОСМЕТИКА/ГИГИЕНА)
0.094650	(G2_КАНЦТОВАРЫ, КНИГИ, ДИСКИ)

Association rules:

- Cosmetics + diapers
- toys + diapers

antecedents	consequents	support	lift	confidence
(G2_КОСМЕТИКА/ГИГИЕНА)	(G2_ПОДГУЗНИКИ)	0.275720	0.963031	0.626168
(G2_ПОДГУЗНИКИ)	(G2_КОСМЕТИКА/ГИГИЕНА)	0.275720	0.963031	0.424051
(G2_ИГРУШКИ)	(G2_ПОДГУЗНИКИ)	0.205761	0.818072	0.531915
(G2_ПОДГУЗНИКИ)	(G2_ИГРУШКИ)	0.205761	0.818072	0.316456
(G2_ПОДГУЗНИКИ)	(G2_ТОВАРЫ ДЛЯ КОРМЛЕНИЯ)	0.185185	0.833842	0.284810
(G2_ТОВАРЫ ДЛЯ КОРМЛЕНИЯ)	(G2_ПОДГУЗНИКИ)	0.185185	0.833842	0.542169
(G2_ТЕКСТИЛЬ, ТРИКОТАЖ)	(G2_ИГРУШКИ)	0.094650	1.081044	0.418182
(G2_ИГРУШКИ)	(G2_ТЕКСТИЛЬ, ТРИКОТАЖ)	0.094650	1.081044	0.244681
(G2_КАНЦТОВАРЫ, КНИГИ, ДИСКИ)	(G2_ИГРУШКИ)	0.057613	1.573543	0.608696
(G2_ИГРУШКИ)	(G2_КАНЦТОВАРЫ, КНИГИ, ДИСКИ)	0.057613	1.573543	0.148936





MBA: Cluster 7 3095 users.

Apriori:

- 95% large products
- others < 33%

Association rules:

- Textile
- Toys
- Diapers

Cluster	Revenue	Avg Revenue
0	8,059,269	3131
1	24,768,135	1891
2	34,735,558	2293
3	6,563,097	2595
4	3,938,704	2281
5	23,749,965	2747
6	1,622,300	2083
7	37,850,558	12230
8	1,686,952	1795

support	itemsets
0.947692	(G2_КРУПНОГАБАРИТНЫЙ ТОВАР)
0.332308	(G2_ТЕКСТИЛЬ, ТРИКОТАЖ)
0.326154	(G2_ИГРУШКИ)
0.304615	(G2_КРУПНОГАБАРИТНЫЙ ТОВАР, G2_ТЕКСТИЛЬ, ТРИКОТАЖ)
0.292308	(G2_ИГРУШКИ, G2_КРУПНОГАБАРИТНЫЙ ТОВАР)
0.289231	(G2_ПОДГУЗНИКИ)
0.270769	(G2_КРУПНОГАБАРИТНЫЙ ТОВАР, G2_ПОДГУЗНИКИ)
0.190769	(G2_ТОВАРЫ ДЛЯ КОРМЛЕНИЯ)
0.172308	(G2_КРУПНОГАБАРИТНЫЙ ТОВАР, G2_ТОВАРЫ ДЛЯ КОРМЛЕНИЯ)
0.095385	(G2_КОСМЕТИКА/ГИГИЕНА)

antecedents	consequents	support	lift	confidence
(G2_ТЕКСТИЛЬ, ТРИКОТАЖ)	(G2_КРУПНОГАБАРИТНЫЙ ТОВАР)	0.304615	0.967262	0.916667
(G2_КРУПНОГАБАРИТНЫЙ ТОВАР)	(G2_ТЕКСТИЛЬ, ТРИКОТАЖ)	0.304615	0.967262	0.321429
(G2_КРУПНОГАБАРИТНЫЙ ТОВАР)	(G2_ИГРУШКИ)	0.292308	0.945693	0.308442
(G2_ИГРУШКИ)	(G2_КРУПНОГАБАРИТНЫЙ ТОВАР)	0.292308	0.945693	0.896226
(G2_КРУПНОГАБАРИТНЫЙ ТОВАР)	(G2_ПОДГУЗНИКИ)	0.270769	0.987842	0.285714
(G2_ПОДГУЗНИКИ)	(G2_КРУПНОГАБАРИТНЫЙ ТОВАР)	0.270769	0.987842	0.936170
(G2_ТОВАРЫ ДЛЯ КОРМЛЕНИЯ)	(G2_КРУПНОГАБАРИТНЫЙ ТОВАР)	0.172308	0.953079	0.903226
(G2_КРУПНОГАБАРИТНЫЙ ТОВАР)	(G2_ТОВАРЫ ДЛЯ КОРМЛЕНИЯ)	0.172308	0.953079	0.181818
(G2_КОСМЕТИКА/ГИГИЕНА)	(G2_КРУПНОГАБАРИТНЫЙ ТОВАР)	0.086154	0.953079	0.903226
(G2_КРУПНОГАБАРИТНЫЙ ТОВАР)	(G2_КОСМЕТИКА/ГИГИЕНА)	0.086154	0.953079	0.090909





MBA: Cluster 8

940 users. Mandatory cat:

books, disks

Apriori:

- 84% buy toys
- 34% - diapers
- 1/3 buy clothing

Cluster	Revenue	Avg Revenue
0	8,059,269	3131
1	24,768,135	1891
2	34,735,558	2293
3	6,563,097	2595
4	3,938,704	2281
5	23,749,965	2747
6	1,622,300	2083
7	37,850,558	12230
8	1,686,952	1795

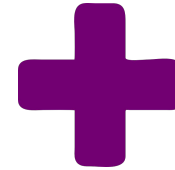
support	itemsets
0.844444	(G2_ИГРУШКИ)
0.348148	(G2_ПОДГУЗНИКИ)
0.337037	(G2_ТЕКСТИЛЬ, ТРИКОТАЖ)
0.277778	(G2_ИГРУШКИ, G2_ТЕКСТИЛЬ, ТРИКОТАЖ)
0.251852	(G2_ИГРУШКИ, G2_ПОДГУЗНИКИ)
0.200000	(G2_ТОВАРЫ ДЛЯ КОРМЛЕНИЯ)
0.192593	(G2_КРУПНОГАБАРИТНЫЙ ТОВАР)
0.185185	(G2_КОСМЕТИКА/ГИГИЕНА)
0.155556	(G2_ИГРУШКИ, G2_ТОВАРЫ ДЛЯ КОРМЛЕНИЯ)
0.151852	(G2_ИГРУШКИ, G2_КРУПНОГАБАРИТНЫЙ ТОВАР)

antecedents	consequents	support	lift	confidence
(G2_ТЕКСТИЛЬ, ТРИКОТАЖ)	(G2_ИГРУШКИ)	0.277778	0.975998	0.824176
(G2_ИГРУШКИ)	(G2_ТЕКСТИЛЬ, ТРИКОТАЖ)	0.277778	0.975998	0.328947
(G2_ПОДГУЗНИКИ)	(G2_ИГРУШКИ)	0.251852	0.856663	0.723404
(G2_ИГРУШКИ)	(G2_ПОДГУЗНИКИ)	0.251852	0.856663	0.298246
(G2_ИГРУШКИ)	(G2_ТОВАРЫ ДЛЯ КОРМЛЕНИЯ)	0.155556	0.921053	0.184211
(G2_ТОВАРЫ ДЛЯ КОРМЛЕНИЯ)	(G2_ИГРУШКИ)	0.155556	0.921053	0.777778
(G2_КРУПНОГАБАРИТНЫЙ ТОВАР)	(G2_ИГРУШКИ)	0.151852	0.933704	0.788462
(G2_ИГРУШКИ)	(G2_КРУПНОГАБАРИТНЫЙ ТОВАР)	0.151852	0.933704	0.179825
(G2_ИГРУШКИ)	(G2_КОСМЕТИКА/ГИГИЕНА)	0.144444	0.923684	0.171053
(G2_КОСМЕТИКА/ГИГИЕНА)	(G2_ИГРУШКИ)	0.144444	0.923684	0.780000





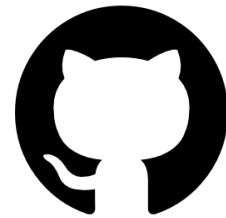
MBA: summary





Project: Stack

- Tableau for primary analysis
- MySQL at DataGrip for aggregation
- Excel to analyze data from SQL
- Pycharm for clustering and MBA
- GitHub as version control and coworking



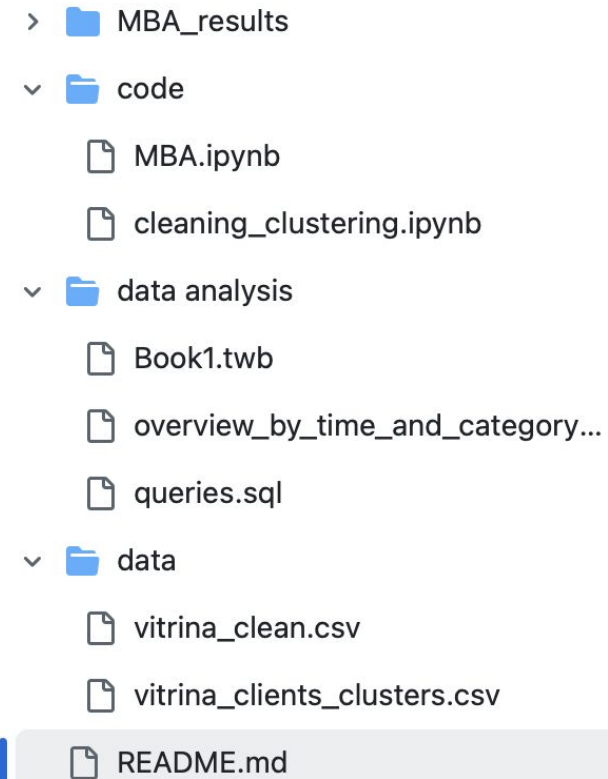
MySQL™





Project: Structure

- results of MBA
- code notebooks
- additional analysis instruments and results
- Витрины с данными



FCS

