Product range analisys

link to the full dashboards and graphs

What data do we have?

We've been given the data with 25900 invoices (=orders), 4372 unique stock codes (=goods), 4070 customers id's (=clients), data range: 2018-11-29 - 2019-12-07.

The main purpose is to find the most promising products and to get rid of the unprofitable ones. And in the end to help business grow loyal clients and increase profit.

Revenue and the number of orders during the year
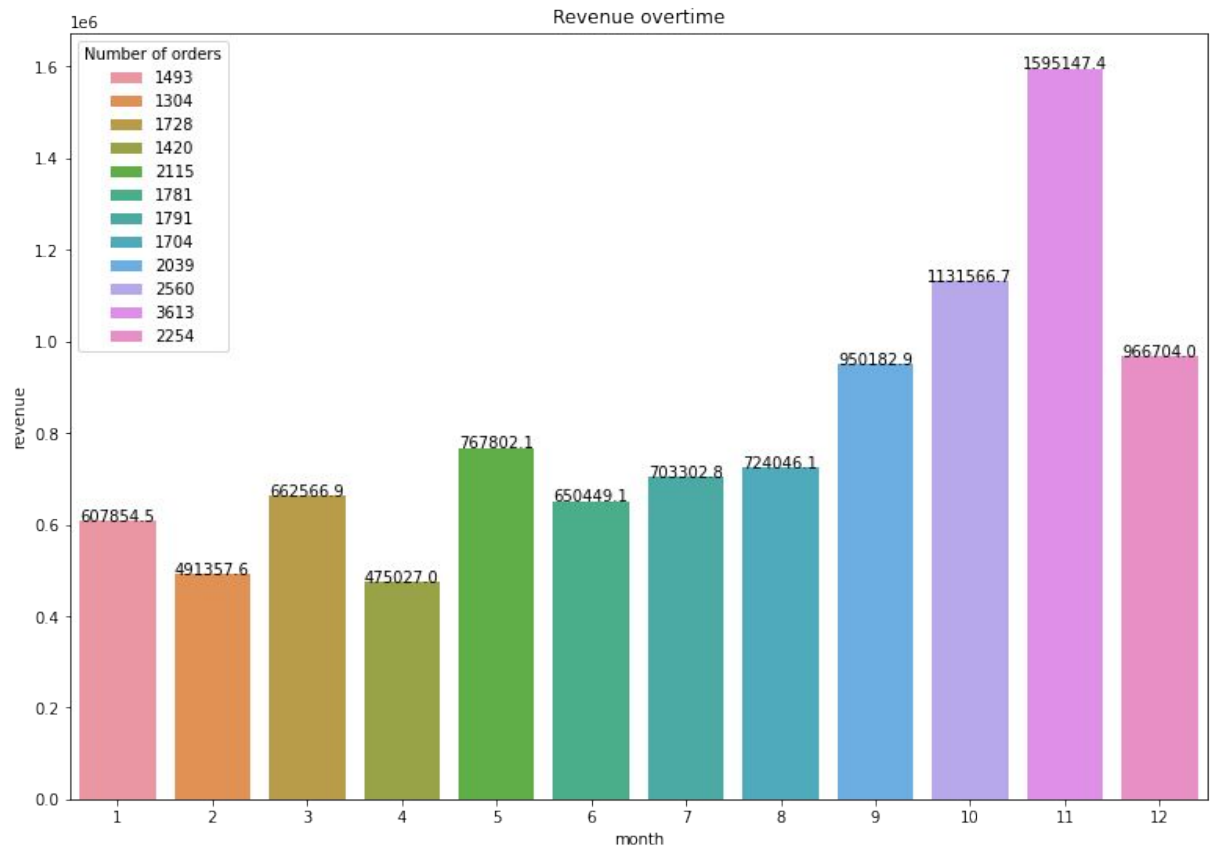
month	revenue	number_of_orders
4	475027.041	1420
2	491357.550	1304
1	607854.450	1493
6	650449.140	1781
3	662566.860	1728
7	703302.801	1791
8	724046.080	1704
5	767802.050	2115
9	950182.892	2039
12	966703.990	2254
10	1131566.720	2560
11	1595147.380	3613

The most profitable month is November with $15 951 147 and 3613 orders.

Invoices

Scatter plot for price of the orders
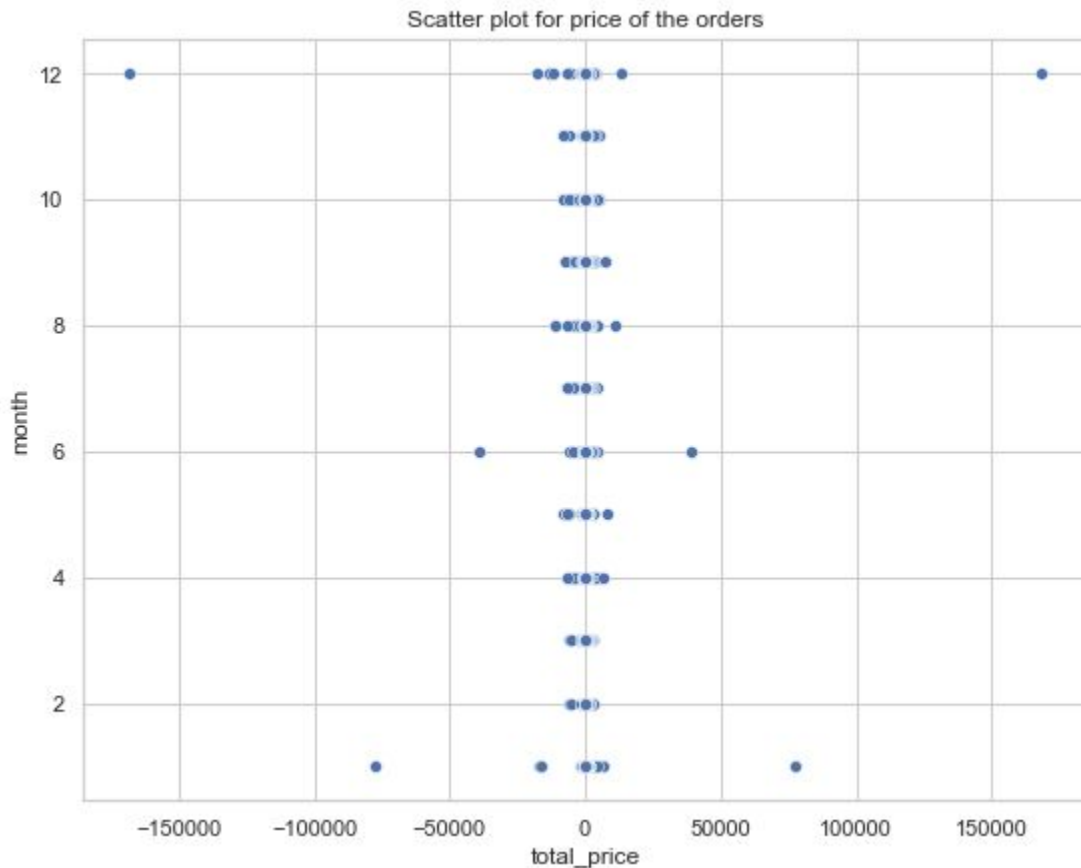
There were 25 900 orders in total.

- Average invoice is $18,2,
- Mode invoice - $15,
- Median invoice - $9,9.

Some orders (including abnormally big like the order in December $168469.6) were cancelled or returned and purchased again.

In the next slides we we'll see numbers for the data without extremely small and exrtemely big unreturned orders.

Revenue and the number of orders without outliers

Revenue overtime without outliers
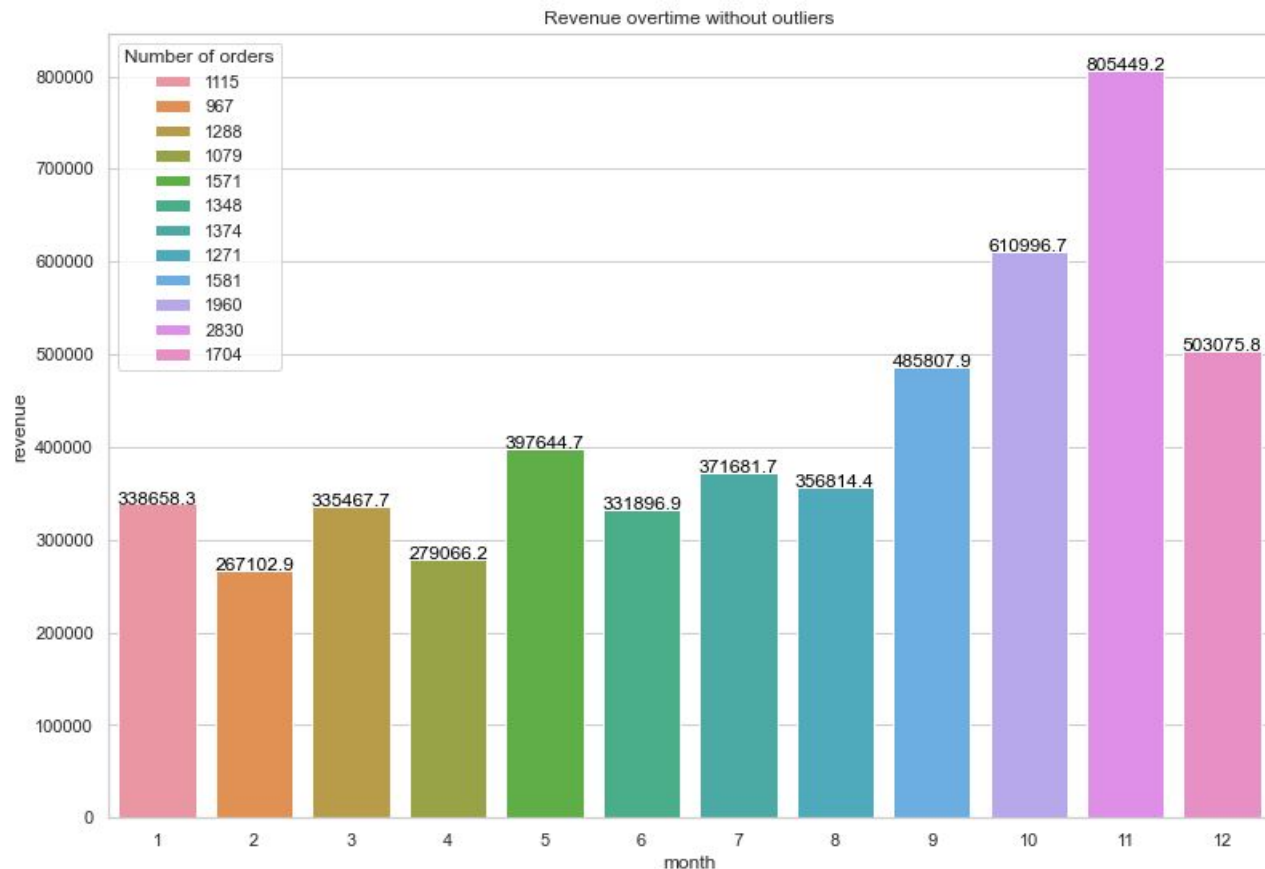
month	revenue	number_of_orders
2	267102.890	967
4	279066.181	1079
6	331896.880	1348
3	335467.680	1288
1	338658.340	1115
8	356814.420	1271
7	371681.741	1374
5	397644.670	1571
9	485807.862	1581
12	503075.820	1704
10	610996.740	1960
11	805449.220	2830

The most profitable month is November with $805 449 and 2830 orders.

Orders and clients

Clients who purchased the most often

customer_id

14911.0	248
12748.0	224
17841.0	169
14606.0	128
13089.0	118
15311.0	118
12971.0	89
14527.0	86
13408.0	81
14646.0	77
16029.0	76
16422.0	75
14156.0	66
13798.0	63
18102.0	62

Some statistics
- average number of orders per client_id is 5,
- mode number of orders per client_id is 1,
- maximum number of orders is 248,
- the majority of clients made between 1 and 5 orders.

Distribution of orders per client

Number of orders overtime
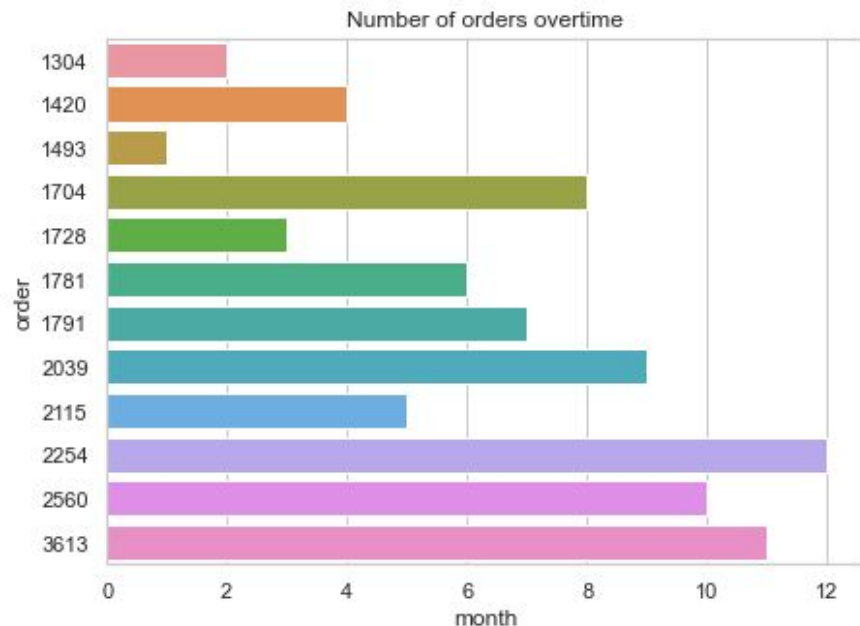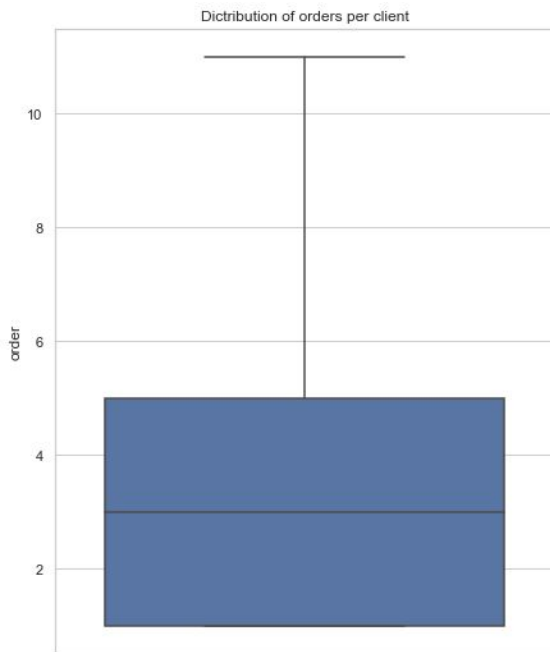
Number of items per order
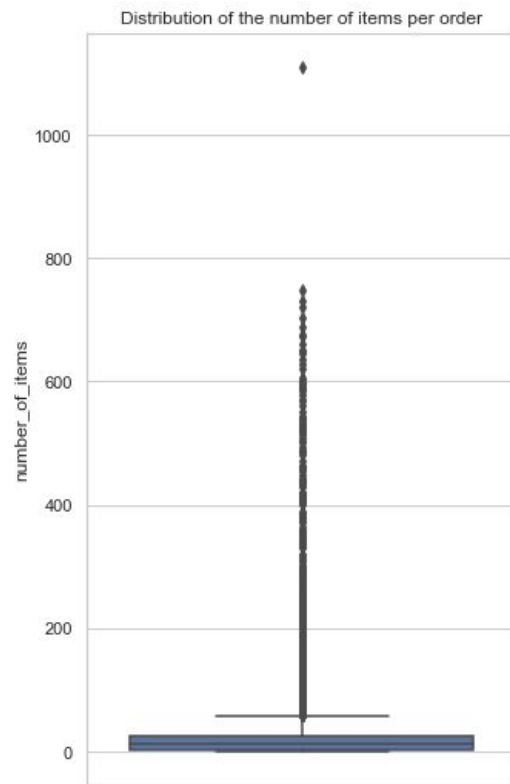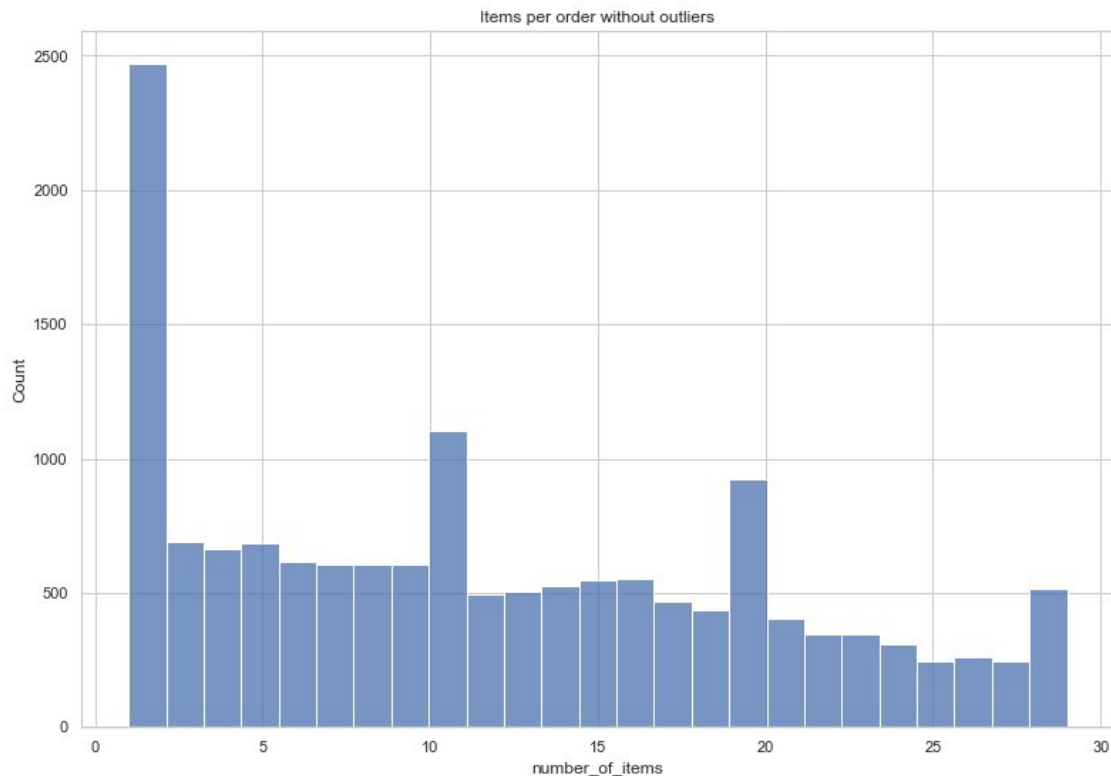
Majority of the orders contains only 1 item. Average order contains 26 items.

Categorizing items

What do we have?

We don't have any given hierarchy for the assortment

One item could fall into two or more categories

Number of unique items - 4042

To understand product range we have categorized them. The assortment is big, we don't have any information, so, the reasonable solution is two put a tag to each item.

So, we used categories from the competitors and with some common sense we've come to this tags (=categories):
- decor,
- accessories,
- kitchenware,
- furniture,
- stationary,
- christmas,
- toys_games.

Categories revenue

Categories that brought the most revenue is kitchenware and decor.

Shares of different categories
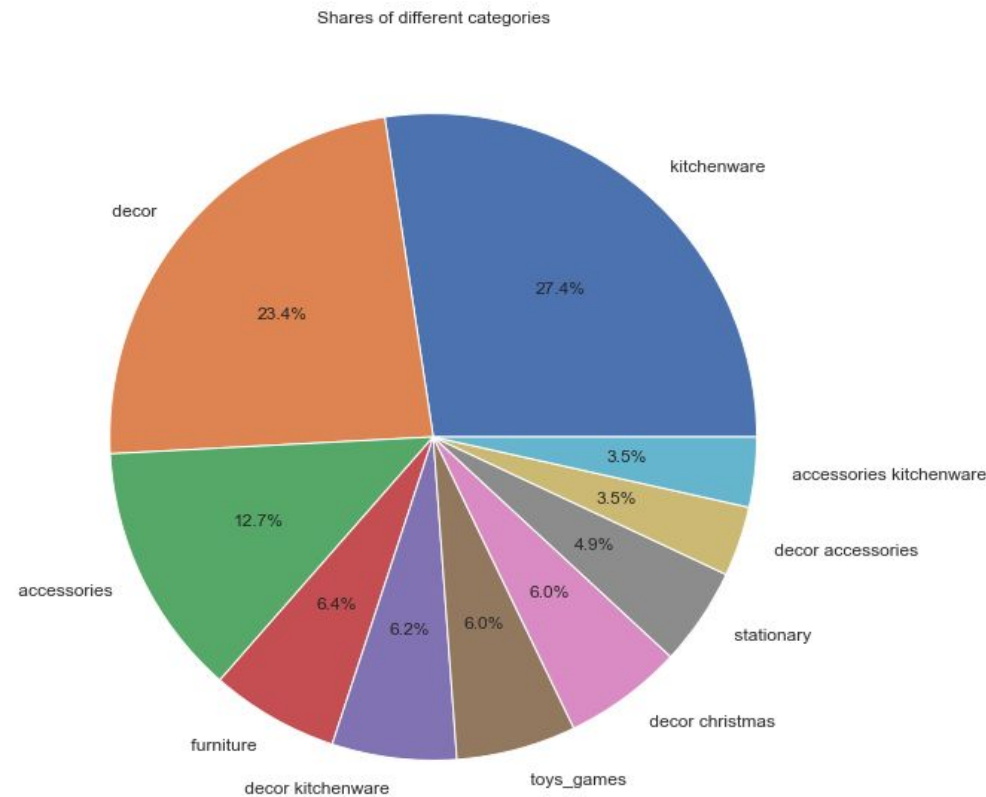
tags	revenue
kitchenware	1962871.2
decor	1542480.6
accessories	819858.1
furniture	516188.3
decor kitchenware	356819.8
toys_games	330866.7
stationary	273601.4
accessories kitchenware	253938.5
decor christmas	237534.3
decor accessories	198241.7

Revenue from the most profitable categories per month
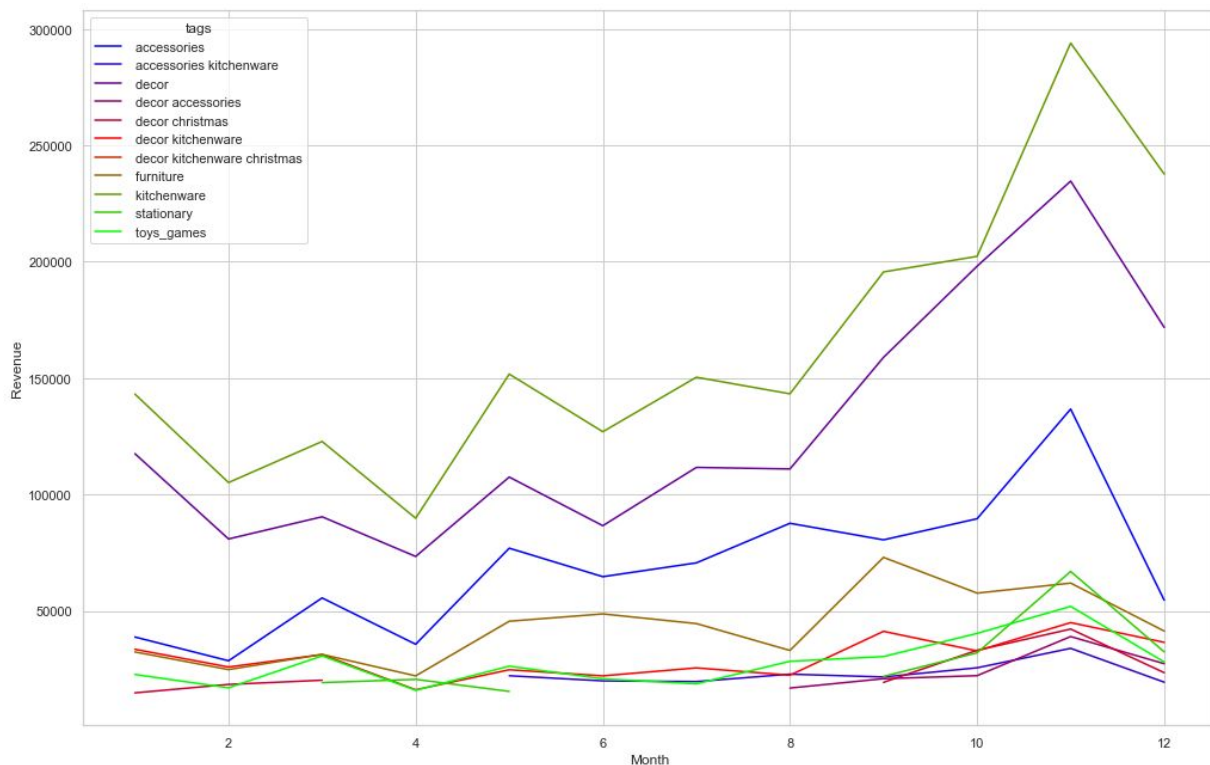
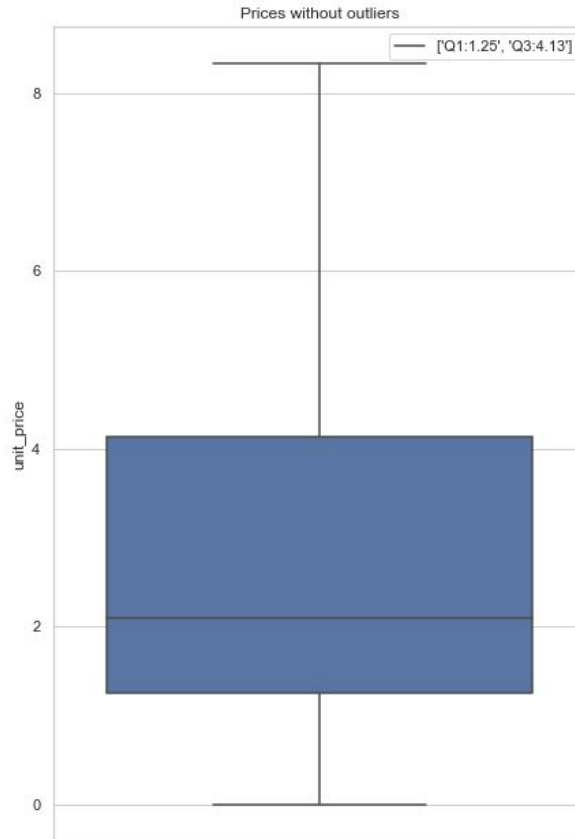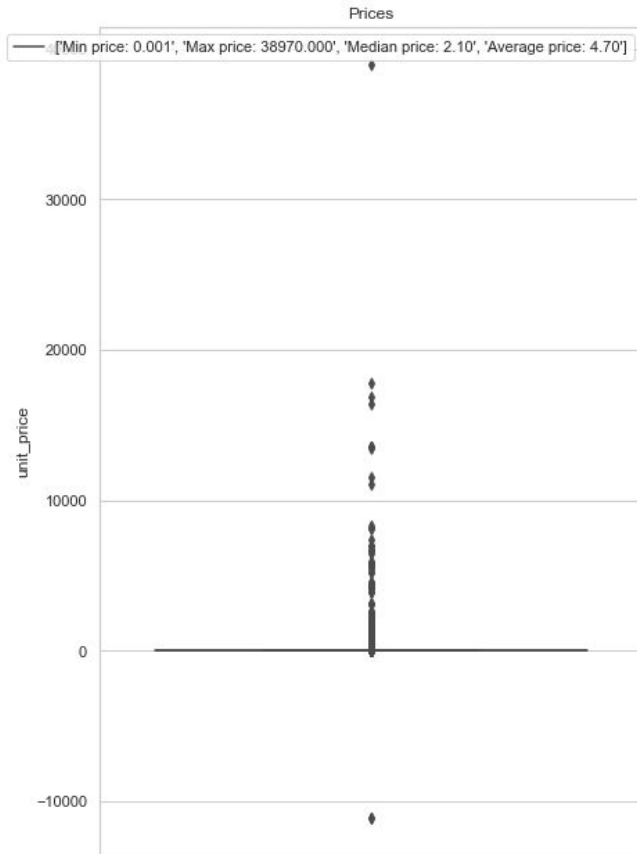| month	accessories	decor	furniture	kitchenware	stationary	toys_games
1	38768.49	117540.590	32355.51	143138.14	NaN	22666.71
2	28564.14	80898.700	24773.67	105138.17	NaN	16940.99
3	55565.30	90440.270	31298.89	122794.58	19185.42	30658.88
4	35657.66	73386.641	22071.16	89816.48	20574.65	15889.23
5	76940.47	107500.530	45550.53	151747.48	15411.74	26252.49
6	64691.69	86596.870	48703.48	126978.07	NaN	20805.09
7	70634.52	111647.241	44540.97	150390.66	NaN	18652.10
8	87640.16	110978.230	32995.10	143264.37	NaN	28341.10
9	80519.36	158929.171	72998.61	195621.36	21854.94	30358.41
10	89587.64	198098.210	57636.48	202331.39	31976.57	40320.63
11	136711.86	234670.970	61938.89	293973.25	66949.27	51931.80
12	54576.82	171793.220	41325.00	237677.27	32367.91	28049.30

Kitchenware, decor and accessories brings the most revenue during the year. Kitchenware and decor are growing. In the next year we shold pay attention to them and look more carefully what happened with accessories, because this category went down too much by the end.

Prices

Prices

['Min price: 0.001', 'Max price: 38970.000', 'Median price: 2.10', 'Average price: 4.70']

Prices without outliers

['Q1:1.25', 'Q3:4.13']

- The majority of items cost 1,25-4,14
- The most expensive item costs 38 970
- The cheapest price - 0,001
- Mean price is 4,7

Items with normal prices (not too expensive) brought 974 403, the most expensive items brought 3727.

Market basket analysis

We implemented apriori algorithm and found association rules for the orders. Here is an examples of the items (antecedents) which were frequently bought together with other items (consequents). This could be used for creating or improving recommendation system for the web-site.

antecedents	consequents
(12 PENCILS SMALL TUBE RED RETROSPOT)	(12 PENCILS SMALL TUBE SKULL)
(12 PENCILS SMALL TUBE SKULL)	(12 PENCILS SMALL TUBE RED RETROSPOT)
(3 PIECE SPACEBOY COOKIE CUTTER SET)	(SET OF 3 BUTTERFLY COOKIE CUTTERS)
(3 STRIPEY MICE FELTCRAFT)	(FELTCRAFT 6 FLOWER FRIENDS)
(FELTCRAFT BUTTERFLY HEARTS)	(3 STRIPEY MICE FELTCRAFT)
(3 STRIPEY MICE FELTCRAFT)	(FELTCRAFT BUTTERFLY HEARTS)
(TRADITIONAL WOODEN CATCH CUP GAME)	(4 TRADITIONAL SPINNING TOPS)
(4 TRADITIONAL SPINNING TOPS)	(TRADITIONAL WOODEN CATCH CUP GAME)
(TRADITIONAL WOODEN SKIPPING ROPE)	(4 TRADITIONAL SPINNING TOPS)
(4 TRADITIONAL SPINNING TOPS)	(TRADITIONAL WOODEN SKIPPING ROPE)
(6 GIFT TAGS 50'S CHRISTMAS)	(6 GIFT TAGS VINTAGE CHRISTMAS)
(6 GIFT TAGS VINTAGE CHRISTMAS)	(6 GIFT TAGS 50'S CHRISTMAS)
(ROLL WRAP 50'S CHRISTMAS)	(6 GIFT TAGS 50'S CHRISTMAS)
(6 GIFT TAGS 50'S CHRISTMAS)	(ROLL WRAP 50'S CHRISTMAS)
(ROLL WRAP VINTAGE CHRISTMAS)	(6 GIFT TAGS VINTAGE CHRISTMAS)
(6 GIFT TAGS VINTAGE CHRISTMAS)	(ROLL WRAP VINTAGE CHRISTMAS)
(CHOCOLATE BOX RIBBONS)	(6 RIBBONS RUSTIC CHARM)

antecedents	consequents
(AIRLINE BAG VINTAGE JET SET RED)	(AIRLINE BAG VINTAGE TOKYO 78)
(AIRLINE BAG VINTAGE TOKYO 78)	(AIRLINE BAG VINTAGE JET SET RED)
(6 GIFT TAGS 50'S CHRISTMAS)	(6 GIFT TAGS VINTAGE CHRISTMAS)
(6 GIFT TAGS VINTAGE CHRISTMAS)	(6 GIFT TAGS 50'S CHRISTMAS)
(ROLL WRAP 50'S CHRISTMAS)	(6 GIFT TAGS 50'S CHRISTMAS)
(6 GIFT TAGS 50'S CHRISTMAS)	(ROLL WRAP 50'S CHRISTMAS)
(6 GIFT TAGS VINTAGE CHRISTMAS)	(ROLL WRAP VINTAGE CHRISTMAS)
(ROLL WRAP VINTAGE CHRISTMAS)	(6 GIFT TAGS VINTAGE CHRISTMAS)
(12 PENCILS SMALL TUBE RED RETROSPOT)	(12 PENCILS SMALL TUBE SKULL)
(12 PENCILS SMALL TUBE SKULL)	(12 PENCILS SMALL TUBE RED RETROSPOT)
(4 TRADITIONAL SPINNING TOPS)	(TRADITIONAL WOODEN CATCH CUP GAME)
(TRADITIONAL WOODEN CATCH CUP GAME)	(4 TRADITIONAL SPINNING TOPS)
(60 CAKE CASES DOLLY GIRL DESIGN)	(PACK OF 60 SPACEBOY CAKE CASES)
(PACK OF 60 SPACEBOY CAKE CASES)	(60 CAKE CASES DOLLY GIRL DESIGN)
(4 TRADITIONAL SPINNING TOPS)	(TRADITIONAL WOODEN SKIPPING ROPE)
(TRADITIONAL WOODEN SKIPPING ROPE)	(4 TRADITIONAL SPINNING TOPS)
(SET OF 20 VINTAGE CHRISTMAS NAPKINS)	(60 CAKE CASES VINTAGE CHRISTMAS)
(60 CAKE CASES VINTAGE CHRISTMAS)	(SET OF 20 VINTAGE CHRISTMAS NAPKINS)
(AIRLINE BAG VINTAGE TOKYO 78)	(SUKI SHOULDER BAG)

The most popular products

stock_code	description	total_price	quantity
23084	RABBIT NIGHT LIGHT	34499.58	14800
22197	POPCORN HOLDER	14914.23	13330
84826	ASSTD DESIGN 3D PAPER STICKERS	60.35	12611
22086	PAPER CHAIN KIT 50'S CHRISTMAS	30734.28	8476
84077	WORLD WAR 2 GLIDERS ASSTD DESIGNS	1985.60	8276
84077	WORLD WAR 2 GLIDERS ASSTD DESIGNS	1779.21	7839
84077	WORLD WAR 2 GLIDERS ASSTD DESIGNS	1939.83	7829
23084	RABBIT NIGHT LIGHT	14119.11	7275
85099B	JUMBO BAG RED RETROSPOT	13050.44	6723
22197	SMALL POPCORN HOLDER	5054.18	6690
84879	ASSORTED COLOUR BIRD ORNAMENT	9878.48	6544
22197	POPCORN HOLDER	5115.52	6278
84879	ASSORTED COLOUR BIRD ORNAMENT	9794.52	6116
21915	RED HARMONICA IN BOX	7670.20	5982
85123A	WHITE HANGING HEART T-LIGHT HOLDER	16285.07	5931

The most profitable products

stock_code	description	total_price	quantity
22502	PICNIC BASKET WICKER 60 PIECES	39619.50	61
23084	RABBIT NIGHT LIGHT	34499.58	14800
22086	PAPER CHAIN KIT 50'S CHRISTMAS	30734.28	8476
22423	REGENCY CAKESTAND 3 TIER	28387.07	2087
47566	PARTY BUNTING	19744.01	3691
22423	REGENCY CAKESTAND 3 TIER	16662.96	1295
85123A	WHITE HANGING HEART T-LIGHT HOLDER	16285.07	5931
85123A	WHITE HANGING HEART T-LIGHT HOLDER	16243.88	5704
47566	PARTY BUNTING	15483.11	2853
22197	POPCORN HOLDER	14914.23	13330
22423	REGENCY CAKESTAND 3 TIER	14605.29	1119
23084	RABBIT NIGHT LIGHT	14119.11	7275
22423	REGENCY CAKESTAND 3 TIER	14026.50	1138
22423	REGENCY CAKESTAND 3 TIER	14025.90	1132
79321	CHILLI LIGHTS	13449.94	2551

Which clients are more profitable

We had a hypothesis about clients with different number of purchases:

Null hypothesis: Is the average revenue from the clients who purchased once more than the average revenue from the clients who made multiple orders?

After testing it with Mann-Whitney test we reject the null hypothesis. So, the average revenue from the clients who purchased once isn't bigger than the average revenue from the clients who made multiple orders.

Statistical tests

We had two hypothesis.
1. About clients with different number of purchases:

Null hypothesis: Is the average revenue from the clients who purchased once more than the average revenue from the clients who made multiple orders?

After testing it with Mann-Whitney test we reject the null hypothesis. So, the average revenue from the clients who purchased once isn't bigger than the average revenue from the clients who made multiple orders.

2. About driver of higher sales. Is there a difference between the average revenue from the orders with at least one kitchenware item and the average revenue from the orders with at least one decor item.

After testing it with Mann-Whitney test we couldn't reject the null hypothesis. So, there is no difference between revenue.

Conclusions and recommendations

We analyzed data about sales and products during the year.

1. The most profitable month is November. May be because of the season - Thanksgiving Day, Christmas, New Year. May be because of some promotions the store made.
2. There are some symmetrical orders: same itemsets, but with negative and positive prices. May be the orders were returned, may be it was technical issues or something else.
3. The store has 4042 unique items. We categorized them.
4. Kitchenware ($1 962 871.2) and decor ($1 542 480.6) are he biggest and the most profitable kind of items. They showed stable growth during the year.
5. We created the list of items for recommendation system.
6. We analyzed correlation between frequently bought items and season of the year. There is a correlation between some of them, so we can use it in marketing events and promotions.
7. We proved that the average revenue from clients who purchased once isn't bigger than revenue from the clients with two or more orders. And we proved that there is a difference between the average revenue from the orders with at least one kitchenware item and the average revenue from the orders with at least one decor item.
   This will help to determine marketing budget and prevent unnecessary efforts.

   Note: Unfortunately we don't have any information about costs, so we can't make strong recommendations about profits.