Product range analysis

Final project

Project description

A large retail chain have recently acquired a small local store.

It wants to keep it as business, but improve the efficiency of sales. The project's main business goal is to study the assortment of the store :
- what does the store sell and how it's sold
- what are popular items? What are items are not popular?
- what items generate the biggest revenue? What items are sold together and etc.

Based on our research as a team we need to make recommendations to the head managers on how to change/improve product range.

Product range at a glance

About 4000 -   number of stock units

3.3 USD -   average price unit, with minimum of 0.04 USD and maximum of 649.5 USD per unit

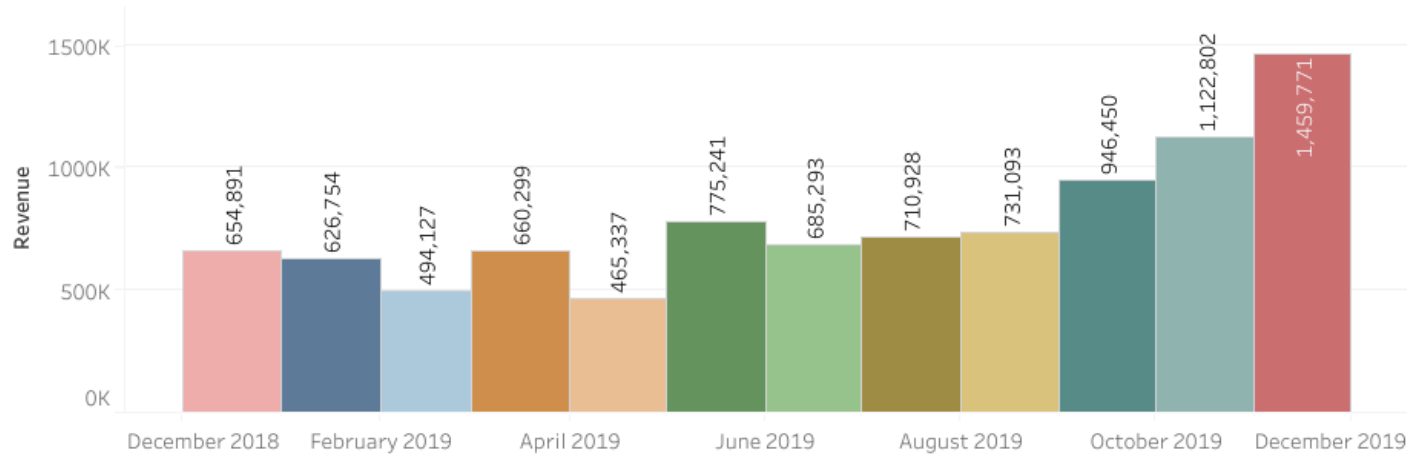**5.17 million -**   total quantity of items sold

504 USD -   average check in the year analyzed, with total number of checks of 18929

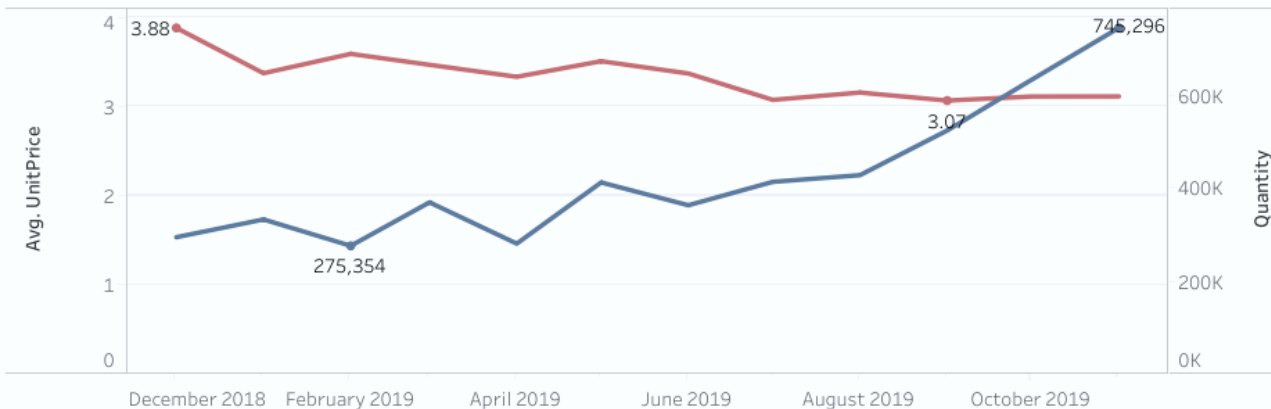**26 items -**   average number of items in one check

2023

What are the trends?

Revenue by month

Quantity and price by month

- Revenue for the year from 1 December 2018 to 30 November 2019 was **9.33 million USD** (after corrections and returns).

- The revenue rose from 655 thousand in Dec 2018 to 1.46 million in Nov 2019, a 123% increase, with most of the growth occurring in **Sep-Nov 2019**.

- The growth was mostly driven by increase in number of invoices and quantity of items sold, while change in average check was not so significant.

- The average item price dropped 20% **from 3.88 in Dec 2018 to 3.11 in Nov 2019**, but this decrease was offset by a 152% rise in units sold, from 300 thousand in Dec 2018 to 758 thousand in Nov 2019.

What are the key items?

Top10 Items

Description or Other	Revenue	% of Total ..
REGENCY CAKESTAND 3 TIER	155,809	1.67%
PARTY BUNTING	97,627	1.05%
WHITE HANGING HEART T-LIGHT HOLDER	95,814	1.03%
JUMBO BAG RED RETROSPOT	90,146	0.97%
RABBIT NIGHT LIGHT	59,083	0.63%
PAPER CHAIN KIT 50'S CHRISTMAS	57,113	0.61%
ASSORTED COLOUR BIRD ORNAMENT	56,390	0.60%
CHILLI LIGHTS	50,685	0.54%
SPOTTY BUNTING	41,426	0.44%
JUMBO BAG PINK POLKADOT	41,166	0.44%
Other	8,587,727	92.01%
Grand Total	9,332,986	100.00%

Top10 keywords

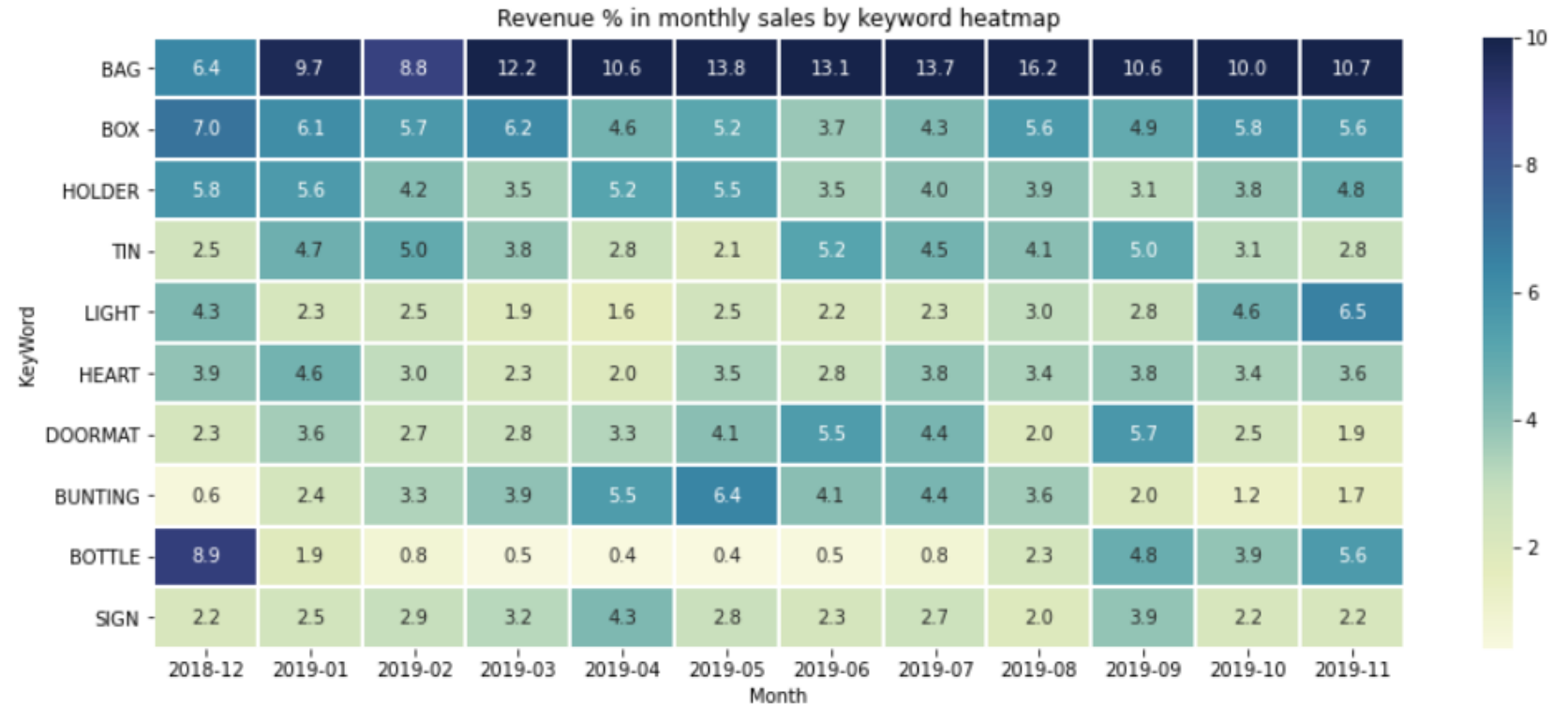| Keyword or ..	Revenue	% of Total ..
BAG	1,062,611	11.39%
BOX	502,867	5.39%
HOLDER	398,396	4.27%
TIN	346,750	3.72%
LIGHT	322,406	3.45%
HEART	320,454	3.43%
DOORMAT	307,773	3.30%
BUNTING	279,094	2.99%
BOTTLE	278,977	2.99%
SIGN	250,417	2.68%
Other	5,263,241	56.39%
Grand Total	9,332,986	100.00%

Revenue by keyword
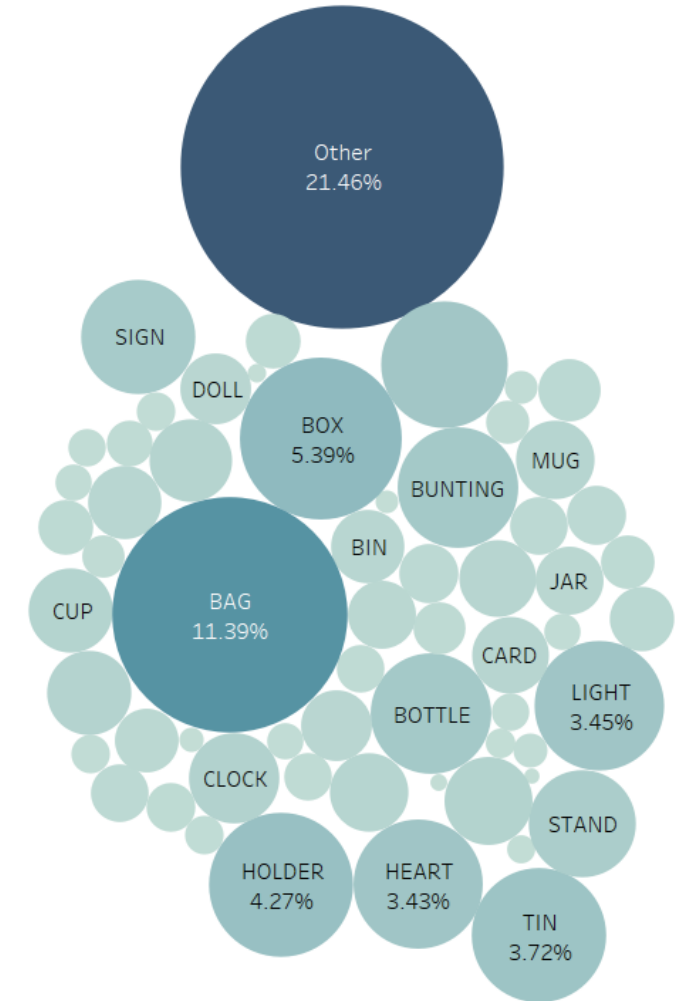
- Individual item's revenue comprise very small % of total sales , with the most valuable item contributing only **1.67%** to the overall revenue.

- To address this, we applied lemmatization and grouped items based on the most frequently occurring nouns – keywords – in item's description, and grouped items by this keyword). This resulted in **60 keywords** that account for **78.5% of sales**, with the remaining 21.5% classified as "Other".

- The table above presents revenue on 10 keyword categories, which covers **43.6% of total revenue.**

Key items fluctuations

On the left there is a heatmap with keyword groups sales monthly shares in total revenue:

- Bags were the top-selling item, with a sales share ranging from **6.4% in December 2018 to 16.2% in August 2019.**
- Boxes and holders had a sales share of 3.7% to 7% and 3.1% to 5.8% respectively. The sales of Tins varied monthly, from 2.1% to 5%.
- Lights, hearts, doormats and signs had sales shares ranging from 2% to 4% with occasional peaks of 5.5% to 6.5%.
- Bunting sales increased from March to August with a **peak of 6.4% in May.**
- Bottle sales were significant in November 2018 **with 8.9% of total** revenue, dropped from February to July 2019 (less than 1% monthly), and then increased from September to November 2019 (3.9% to 5.6% of total revenue).

Revenue % in monthly sales by keyword heatmap

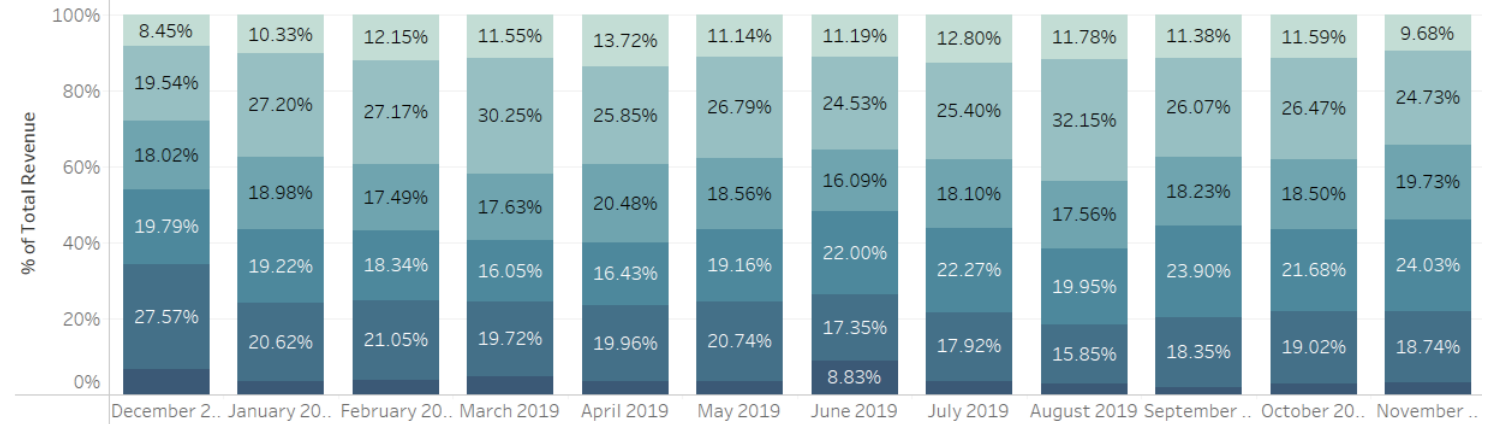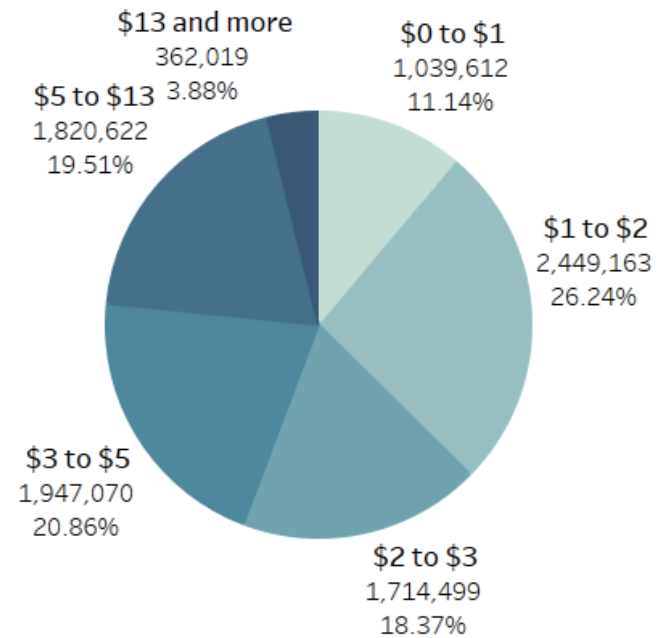| KeyWord	2018-12	2019-01	2019-02	2019-03	2019-04	2019-05	2019-06	2019-07	2019-08	2019-09	2019-10	2019-11
BAG	6.4	9.7	8.8	12.2	10.6	13.8	13.1	13.7	16.2	10.6	10.0	10.7
BOX	7.0	6.1	5.7	6.2	4.6	5.2	3.7	4.3	5.6	4.9	5.8	5.6
HOLDER	5.8	5.6	4.2	3.5	5.2	5.5	3.5	4.0	3.9	3.1	3.8	4.8
TIN	2.5	4.7	5.0	3.8	2.8	2.1	5.2	4.5	4.1	5.0	3.1	2.8
LIGHT	4.3	2.3	2.5	1.9	1.6	2.5	2.2	2.3	3.0	2.8	4.6	6.5
HEART	3.9	4.6	3.0	2.3	2.0	3.5	2.8	3.8	3.4	3.8	3.4	3.6
DOORMAT	2.3	3.6	2.7	2.8	3.3	4.1	5.5	4.4	2.0	5.7	2.5	1.9
BUNTING	0.6	2.4	3.3	3.9	5.5	6.4	4.1	4.4	3.6	2.0	1.2	1.7
BOTTLE	8.9	1.9	0.8	0.5	0.4	0.4	0.5	0.8	2.3	4.8	3.9	5.6
SIGN	2.2	2.5	2.9	3.2	4.3	2.8	2.3	2.7	2.0	3.9	2.2	2.2

Month

The overall sales structure in December 2018 was different from January to November 2018, which may have been affected by seasonal sales. It would be valuable to compare full month-to-month data with December 2019 when it becomes available.

Based on the sales graph by month, we can see that the increases in sales in March and May 2019 were driven by the rise in sales of bags, boxes and bunting. The rise in sales from September to November 2019 cannot be attributed to a significant increase in one or several categories, suggesting it is due to a general increase in sales.

What are the prices of goods?

Revenue by unit price groups:

$13 and more
362,019
3.88%

$5 to $13
1,820,622
19.51%

$0 to $1
1,039,612
11.14%

$1 to $2
2,449,163
26.24%

$3 to $5
1,947,070
20.86%

$2 to $3
1,714,499
18.37%

We split goods into categories, according to their prices and analyzed share of each group in total sales.

More than half of the revenue (55%) comes from items priced below 3 USD. Items priced between 3 to 5 USD and 5 to 13 USD contribute **about 20%** each to the revenue, while items priced over 13 USD make up **about 4%**.

A monthly analysis of sales by price groups reveals fluctuations throughout the year. However, in the last three months of the period, when revenue was growing, the distribution of goods between price groups remained relatively stable.

This indicates that growth in sales was observed for goods in all price groups.

RFM analysis to categorize goods

Top 10 RFM segments by revenue

RFM_segment	RFM_code	Revenue	Items	%_revenue	%_of_items
recent / top seller / best value	111	6532171.72	692	68.35	17.30
recent / popular / good value	122	428344.30	355	4.48	8.88
less recent / top seller / best value	211	409858.38	61	4.29	1.53
old / top seller / best value	311	294603.11	46	3.08	1.15
recent / popular / best value	121	272236.20	86	2.85	2.15
recent / top seller / good value	112	225009.18	159	2.35	3.98
less recent / popular / good value	222	187929.91	150	1.97	3.75
less recent / popular / best value	221	116040.84	36	1.21	0.90
old / popular / good value	322	101041.88	88	1.06	2.20
old / popular / best value	321	75831.02	25	0.79	0.63

The table present top 10 RFM segments by revenue. Dominant part of revenue (68%) is generated by items fall into the best category for all three parameters (recency, frequency, monetary).

Almost 1/3 of items in the items list (1173 stock units out of 4000) are unpopular, has low value and last time sold long time ago (RFM score 444). Revenue generated by these items is only 144 thousand (1.5% of total revenue).

RFM analysis is a marketing technique used to quantitatively rank and group customers based on the recency, frequency and monetary total of their recent transactions to identify the best customers and perform targeted marketing campaigns. In this project we applied this technique to make segmentations of product range.
RFM analysis scores goods on each of the three main factors, the score ranges from 1 (the best) to 4 (the worst).

Significant portion of our product range (more than a half of items) can not be included into one of the groups above ("bestsellers" / items to get rid of). They can be analyzed further, using interactive dashboards in Tableau.
Areas for further analysis (beyond the scope of this project):
- Items that have bad R score only, could be analyzed item by item to take into account seasonality factor.
- Items, that have bad F score only, could be investigated further to identify the reasons of rare sales.
- As for the items, that have bad M score only, I would not recommend to get rid of them, as they can be complementary to sales of pricy goods.
- Items that have bad F and M scores could be considered for removal, despite the good R score.

What items are returned most often?

Returns monthly
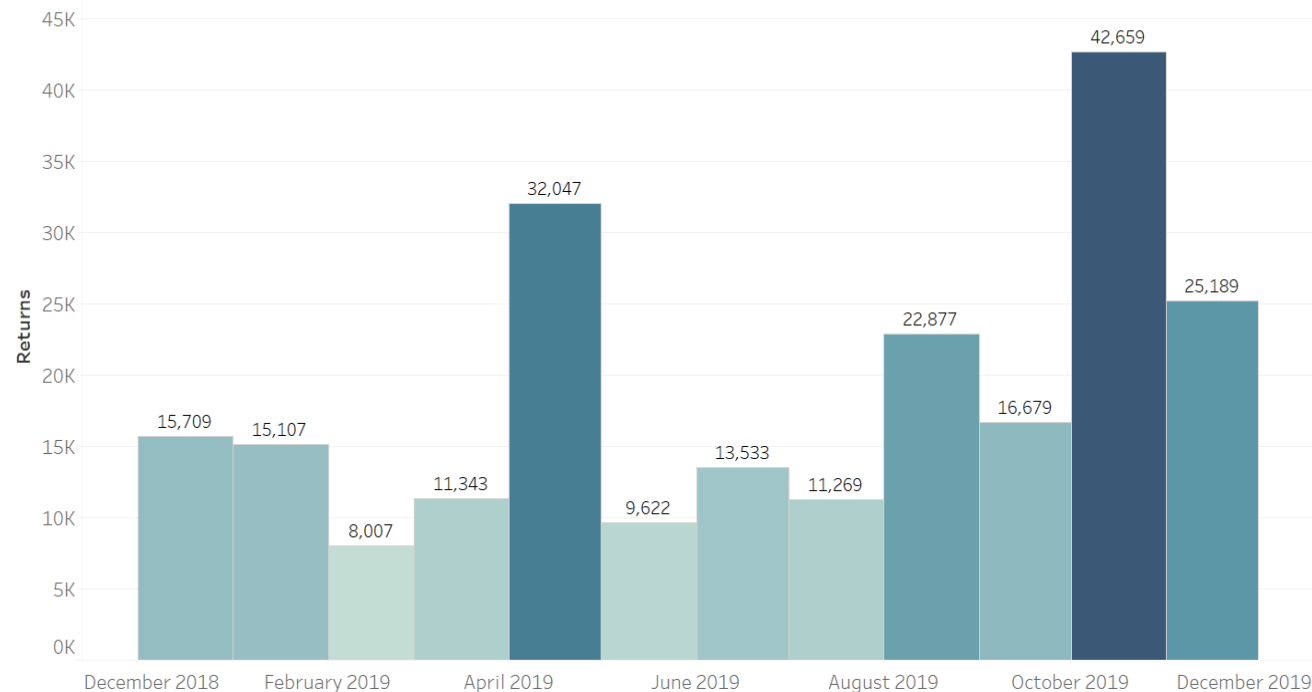
Returns fluctuates significantly from month to month, which means that returns of individual high-value items can increase returns significantly.

Goods with highest rate of returns

Description	Correction	Revenue	% of correction
SET OF 3 BABUSHKA STACKING TINS	-1258.05	1336.50	94.13
PANTRY CHOPPING BOARD	-4803.06	5817.22	82.57
FLOWERS CHANDELIER T-LIGHT HOLDER	-1859.07	2879.37	64.57
CINDERELLA CHANDELIER	-1514.05	3717.56	40.73
FAIRY CAKE FLANNEL ASSORTED COLOUR	-6591.42	17721.40	37.19
FELTCRAFT DOLL MOLLY	-3500.85	11822.30	29.61
TEA TIME PARTY BUNTING	-3692.95	12889.75	28.65
IVORY DINER WALL CLOCK	-2653.70	11648.89	22.78
VINTAGE RED KITCHEN CABINET	-1730.50	7830.00	22.10

- Almost half of items in the items list was returned at least once, however amount of returns in total revenue is small - 2.4%.
- We analyzed top 30 of returned items sorted by amount (covers 35% of total amount of returns).
- There are items with % of returns varying from 1 to 10% that represent normal course of sales. However for a number of items, correction comprises significant portion of item's revenue (up to 94%). Items with return rate over 20% are listed in the table above.

Conclusions and recommendations

Trends

- The revenue rose from 655 thousand in Dec 2018 to 1.46 million in Nov 2019, a 123% increase, with most of the growth occurring in Sep-Nov 2019.

- The growth was mostly driven by increase in number of invoices and quantity of items sold, and is not associated with particular group or goods, suggesting it is due to a general increase in sales.

Product range features

- Due to large number of stock units (4000), individual item's revenue comprise very small % of total sales. In order to perform analysis we categorized items by keywords, price group and RFM segment.

- More than half of the revenue (55%) comes from items priced below 3 USD. Items priced between 3 to 5 USD and 5 to 13 USD contribute about 20% each to the revenue, while items priced over 13 USD make up about 4%.

Optimization of product range

Based on the analysis performed we can recommend to exclude from product range the following items:

- 1173 with RFM score 444 that generated only 144 thousand of revenue (1.5% of total).
- Items with return rate over 20% (listed on the slide with analysis of returns)

Further optimization steps can be made on the basis of analysis and detailed investigation of items via dashboards.

What's next?

Dashboards:

3 Tableau dashboards, prepared for this project, allow to group good using multiple criteria's to extend analysis to smaller groups and individual items, if required.
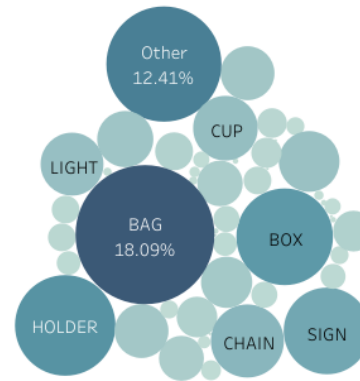
Product range analysis

Keywords dashboard

List of items, can be filtered using various parameters

Filters

Itmes grouped by keyword

KeyWord	Description	Revenue	% of To..
BAG	JUMBO BAG RED RETROSPOT	23,617	8.77%
	CHARLOTTE BAG SUKI DESIGN	19,474	7.23%
	JUMBO BAG VINTAGE LEAF	12,418	4.61%
	JUMBO BAG ALPHABET	11,945	4.44%
	JUMBO BAG APPLES	11,660	4.33%
	JUMBO BAG 50'S CHRISTMAS	11,378	4.22%
	JUMBO BAG VINTAGE DOILY	11,190	4.16%
	JUMBO BAG PINK POLKADOT	10,554	3.92%
	RED RETROSPOT PICNIC BAG	10,121	3.76%
	RED RETROSPOT CHARLOTTE BAG	9,662	3.59%
	JUMBO STORAGE BAG SUKI	8,848	3.29%
	WOODLAND CHARLOTTE BAG	7,894	2.93%
	JUMBO BAG PEARS	7,596	2.82%

Revenue by keyword

Other 12.41%
CUP
LIGHT
BAG 18.09%
BOX
HOLDER
CHAIN
SIGN

KeyWord		UnitPrice Group	
✔	BAG		$0 to $1
	BASKET		$1 to $2
✔	BELL	✔	$2 to $3
✔	BIN		$3 to $5
✔	BOARD		$5 to $13
✔	BOOK		$13 and more
✔	BOTTLE		
✔	BOWL		
✔	BOX	Recency	
✔	BRACELET	✔ 1	
✔	BUNTING	✔ 2	
	CANDLE	3	
✔	CARD	4	
✔	CHAIN		
✔	CHALKBOARD		
✔	CHARM		

Recency: ✔1 ✔2 3 4
Frequency: ✔1 ✔2 3 4
Monetary: ✔1 ✔2 3 4

Revenue by month

- December 2018: 100,617
- 90,030
- February 2019: 70,068
- 89,704
- April 2019: 72,191
- 124,700
- June 2019: 96,161
- 110,986
- August 2019: 117,494
- 158,720
- October 2019: 189,639
- December 2019: 268,598

Quantity and price by month

- February 2019: 2.58
- 27,779
- October 2019: 110,212 / 2.45
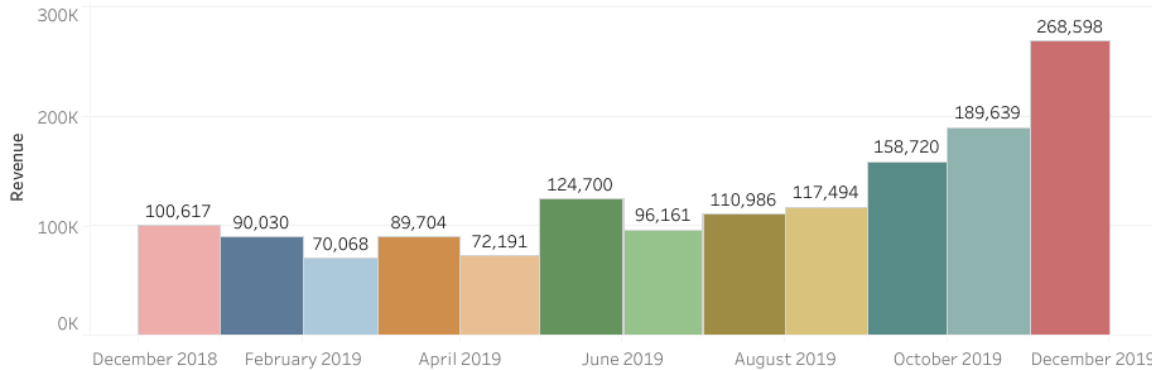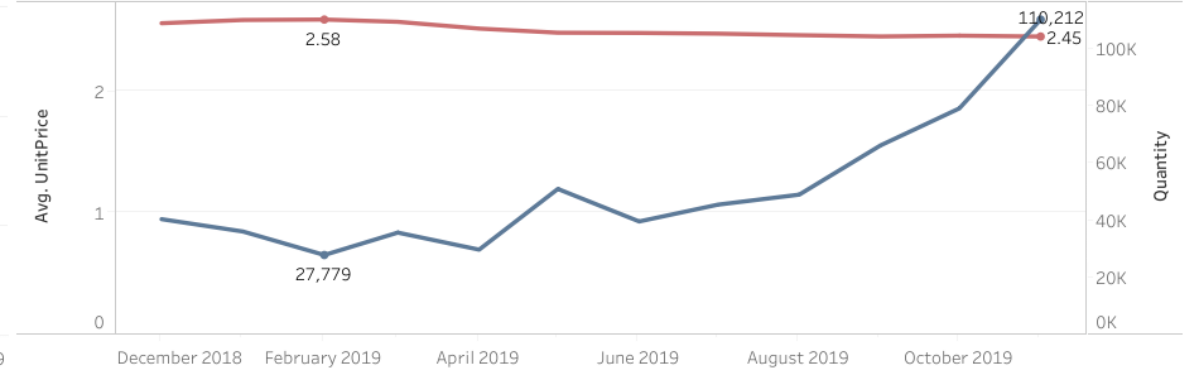
Interactive visualizations of item's data.

RFM dashboard

RFM filters

List of goods grouped by RFM segments

Recency
- ✔ 1
- ✔ 2
- ✔ 3
- ✔ 4

Frequency
- ✔ 1
- ✔ 2
- ✔ 3
- ✔ 4

Monetary
- ✔ 1
- ✔ 2
- ✔ 3
- ✔ 4

RFM segments

RFM_code	RFM_segment	Number of items (distinct)
111	recent / top seller / best value	686
444	oldest / rare / cheap	455
122	recent / popular / good value	357
344	old / rare / cheap	247
233	less recent / unpopular / low value	205
133	recent / unpopular / low value	177
333	old / unpopular / low value	155
112	recent / top seller / good value	155

Items grouped by RFM

RFM_segment	Description	Revenue	% of Total Rev..
recent / top seller / best value	REGENCY CAKESTAND 3 TIER	155,809	1.67%
	PARTY BUNTING	97,627	1.05%
	WHITE HANGING HEART T-LIGHT HOLDER	95,814	1.03%
	JUMBO BAG RED RETROSPOT	90,146	0.97%
	RABBIT NIGHT LIGHT	59,083	0.63%
	PAPER CHAIN KIT 50'S CHRISTMAS	57,113	0.61%
	ASSORTED COLOUR BIRD ORNAMENT	56,390	0.60%
	CHILLI LIGHTS	50,685	0.54%
	SPOTTY BUNTING	41,426	0.44%
	JUMBO BAG PINK POLKADOT	41,166	0.44%
	SET OF 3 CAKE TINS PANTRY DESIGN	36,527	0.39%
	DOORMAT KEEP CALM AND COME IN	35,283	0.38%
	JAM MAKING SET WITH JARS	34,576	0.37%
	LUNCH BAG RED RETROSPOT	34,062	0.36%
	WOOD BLACK BOARD ANT WHITE FINISH	33,858	0.36%
	BLACK RECORD COVER FRAME	33,155	0.36%
	JUMBO BAG STRAWBERRY	32,086	0.34%
	HEART OF WICKER SMALL	31,351	0.34%
	VICTORIAN GLASS HANGING T-LIGHT	30,840	0.33%
	POPCORN HOLDER	29,875	0.32%
	CHOCOLATE HOT WATER BOTTLE	29,622	0.32%
	JUMBO STORAGE BAG SUKI	29,539	0.32%
	HEART OF WICKER LARGE	29,527	0.32%
	SET OF 3 REGENCY CAKE TINS	29,417	0.32%
	PAPER CHAIN KIT VINTAGE CHRISTMAS	29,036	0.31%
	HOT WATER BOTTLE TEA AND SYMPATHY	29,024	0.31%
	JUMBO BAG APPLES	28,712	0.31%
	CHARLOTTE BAG SUKI DESIGN	27,212	0.29%

Revenue RFM Treemap

recent / top seller / best value **6,375,895** 68.32%	recent / popular / good value 423,750	recent / top seller / good
		less recent /
	less recent / top seller / best value 389,988	old /
		old /
	recent / popular / best value	
	old / top seller / best value	

Visualizations

Daily sales dashboard

Period filters

Summary information on selected data

List of goods sold with indication of revenue, price and quantity

Month
- ✔ December 2018
- January 2019
- February 2019
- March 2019
- April 2019
- May 2019
- June 2019
- July 2019
- August 2019
- September 2019
- October 2019
- November 2019

Day
- 1
- 2
- 3
- 4
- 5
- 6
- 7
- 8
- 9
- 10
- 11
- 12
- 13
- 14
- ✔ 15
- ✔ 16
- ✔ 17
- ✔ 18
- ✔ 19

Revenue:	Average check:
129,219	**541.2**

No of customers:	No of invoices:
175	**247**

Revenue by unit price groups:

- **$13 and more** 11,348 8.78%
- **$0 to $1** 11,848 9.17%
- **$1 to $2** 21,580 16.70%
- **$2 to $3** 17,570 13.60%
- **$3 to $5** 25,534 19.76%
- **$5 to $13** 41,338 31.99%
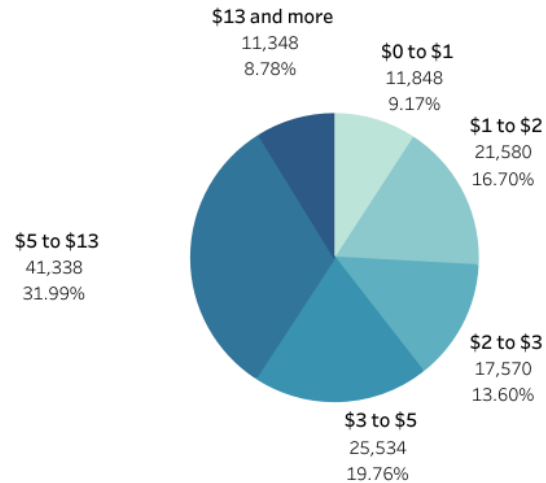
Goods sold

Description	KeyWord	Revenue	Avg. UnitPrice	Quantity
REGENCY CAKESTAND 3 TIER	STAND	4,444	15.93	265
WHITE SKULL HOT WATER BOTTLE	BOTTLE	2,286	4.23	346
KNITTED UNION FLAG HOT WATER BOTTLE	BOTTLE	2,225	5.07	420
HOT WATER BOTTLE TEA AND SYMPATHY	BOTTLE	2,007	7.53	220
RED WOOLLY HOTTIE WHITE HEART.	HEART	1,884	7.94	223
CHOCOLATE HOT WATER BOTTLE	BOTTLE	1,666	7.50	251
SCOTTIE DOG HOT WATER BOTTLE	BOTTLE	1,423	5.57	280
ROTATING SILVER ANGELS T-LIGHT HLDR	LIGHT	1,064	4.22	209
FAWN BLUE HOT WATER BOTTLE	BOTTLE	1,037	3.90	307
CHARLOTTE BAG SUKI DESIGN	BAG	1,014	1.65	613
HOT WATER BOTTLE I AM SO POORLY	BOTTLE	998	6.66	166
CHICK GREY HOT WATER BOTTLE	BOTTLE	969	5.45	198
HOT WATER BOTTLE BABUSHKA	BOTTLE	925	6.56	187
PAPER CHAIN KIT 50'S CHRISTMAS	CHAIN	919	5.15	180
SET OF 3 CAKE TINS PANTRY DESIGN	TIN	782	6.35	120
RETROSPOT HEART HOT WATER BOTTLE	BOTTLE	759	6.89	129
WOODEN FRAME ANTIQUE WHITE	FRAME	747	4.98	163
PACK OF 72 RETROSPOT CAKE CASES	Other	739	0.97	1,612
HAND WARMER BABUSHKA DESIGN	WARMER	727	1.93	602
ROUND SNACK BOXES SET OF4 WOODLAND	BOX	715	4.22	250
DELUXE SEWING KIT	Other	713	9.55	68
RED TOADSTOOL LED NIGHT LIGHT	LIGHT	697	2.21	449
PAPER CHAIN KIT VINTAGE CHRISTMAS	CHAIN	654	3.85	196
CABIN BAG VINTAGE RETROSPOT	BAG	650	12.75	51
VICTORIAN SEWING BOX LARGE	BOX	609	20.69	36
SET 7 BABUSHKA NESTING BOXES	BOX	597	12.62	68
ROSES REGENCY TEACUP AND SAUCER	CUP	578	4.36	181
ANTIQUE SILVER TEA GLASS ETCHED	Other	562	2.09	269

This presentation is a part of a project made during study on course «Data Analyst» by Practicum

Prepared by Anna Halamaizer

(972) 53-341-3332

anna.halamaizer@gmail.com