

“Анализ расширения торгового направления ниши сейфов, стеллажей и металлической мебели (поиск инсайтов, составление рекомендаций стейкхолдерам)”

Дипломная работа по программе «Аналитик данных с нуля до middle»

Вертелецкая А.Ю.

Группа: DAU-33

Москва, 2023



Описание бизнес-задачи, стейкхолдеров отчёта



1

Описание бизнес-задач

- Общий анализ ассортимента по направлению сейфов, стеллажей, металлической и производственной мебели на местном рынке города N., что позволит компании получить цельное представление о текущем положении дел;
- Анализ предложений конкурентов с предложением поставщика в области ценовой политики на основании сравнения аналогичного ассортимента.

Стейкхолдер

Руководитель компании

Ссылки на ноутбуки:

[Сбор и обработка данных](#)

[Анализ и гипотезы](#)



Описание данных и их предобработки



2

Описание данных и их предобработки

- Парсинг официальных сайтов интересующих организаций (5 конкурентов и 1 поставщик)
- Получено 6 наборов данных. Все столбцы имели тип данных object, некоторые содержали лишние служебные символы.
- Каждый набор данных был обработан: очищен, приведён к нужным форматам данных, были созданы новые необходимые столбцы и удалены ненужные.

Данные до преобразований

Имя столбца	Описание	Тип данных	1	2	3	4	5	43
title	Название товара	object	+	+	+	+	+	+
description	Описание кратко	object	+	+	-	-	+	-
size	Габариты	object	+	+	+	+	+	+
weight	Вес в кг	object	-	-	-	-	-	+
price	Цена в руб	object	-	+	+	+	+	+
availability	Наличие	object	+	-	-	-	-	-

Данные после очистки и преобразований

Имя столбца	Описание	Тип данных	1	2	3	4	5	43
title	Название товара	object	+	+	+	+	+	+
weight	Вес в кг	float64	+	+	-	-	-	+
price	Цена в руб	float64	+	+	+	+	+	+
category	Категория	object	+	+	+	+	+	+
shop	Номер организации	int64	+	+	+	+	+	+
height	Высота товара в мм	float64	+	+	+	+	+	+
width	Ширина товара в мм	float64	+	+	+	+	+	+
depth	Глубина товара в мм	float64	+	+	+	+	-	+
manufacturer	Производитель	object	-	-	-	-	+	-
availability	Наличие (0 - нет, 1 - да)	int64	+	-	-	-	-	-
reduced	Сокращённое название	object	+	+	+	+	+	+

Описание проведенного исследования



3

Описание проведенного исследования

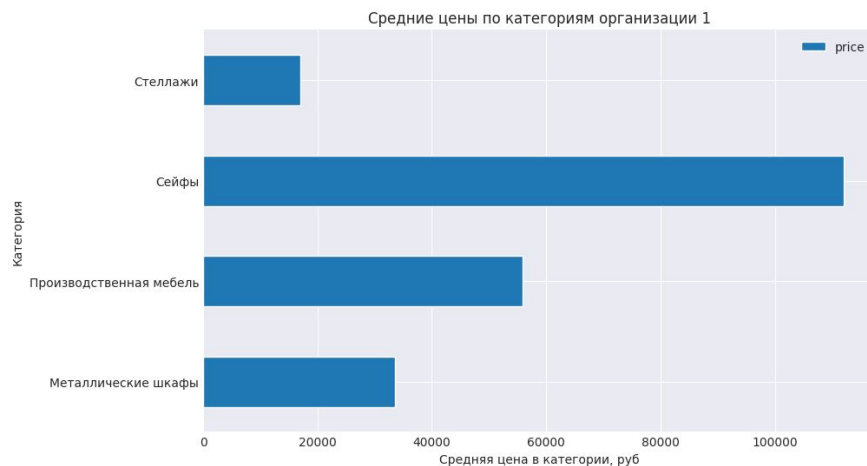
- Полученные данные были проанализированы как по отдельным организациям, так и по совпадающему ассортименту. Было получено общее представление о ситуации на местном рынке.

Табл.1 Количество ассортимента по категориям (с условным форматированием по цвету)

Организация	Сейфы	Стеллажи	Мет. мебель	Произв. мебель	Всё вместе	Место по разнобр. ассорт.
1	1302	6055	1399	2191	10947	1
2	484	112	145	240	981	4
3	482	41	166	211	900	5
4	261	68	176	109	614	6
5	3052	2522	1625	2221	9420	2
43	492	185	216	215	1108	3

Самый широкий ассортимент в категориях “Стеллажи” и “Сейфы”.

↗



Самая дорогая категория - “Сейфы”.

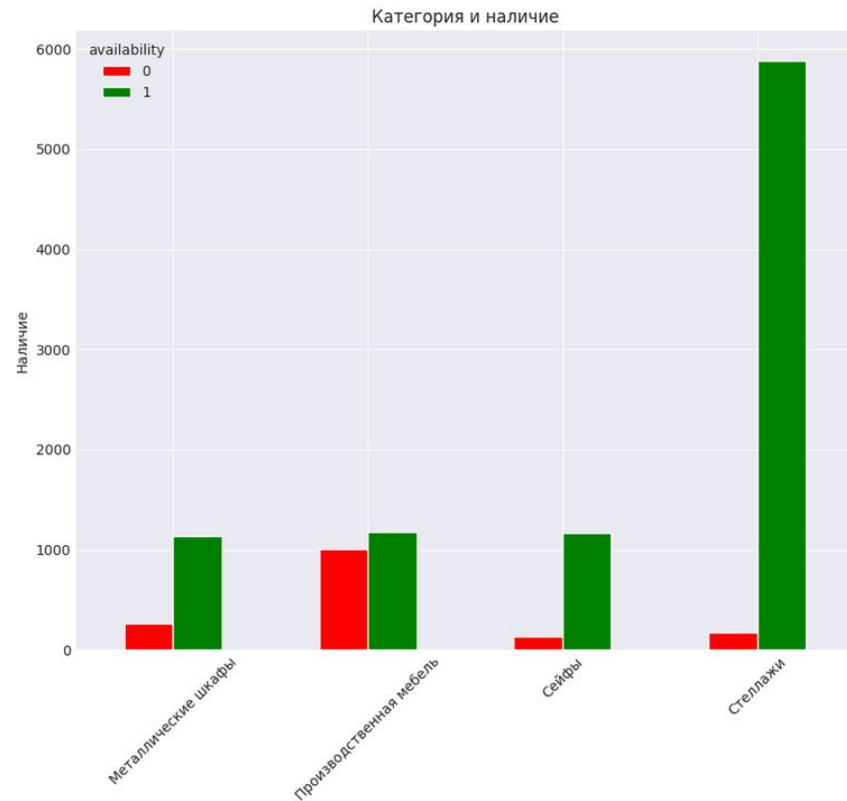
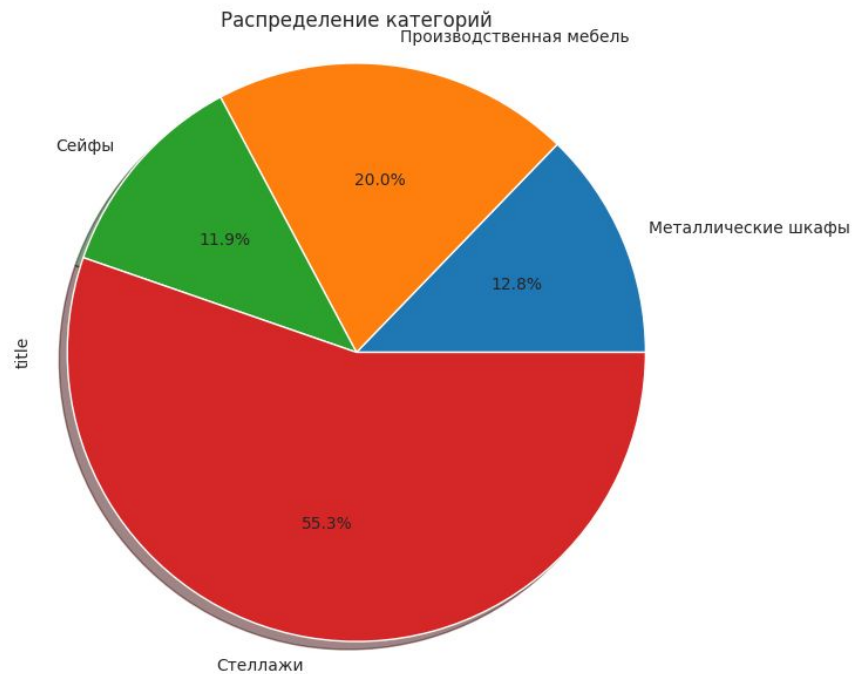
Самая бюджетная категория - “Стеллажи”.



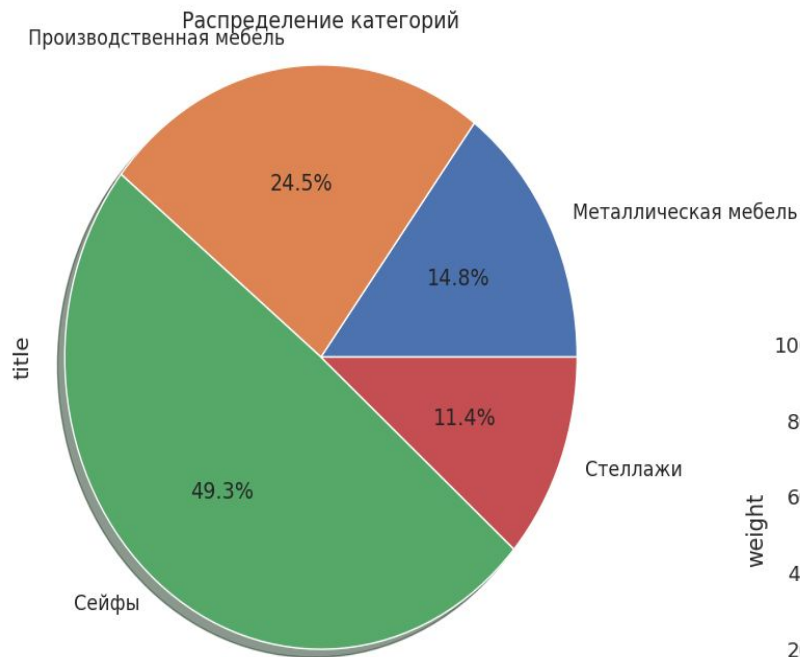
Анализ организации 1

title	
availability	
0	1577
1	9370

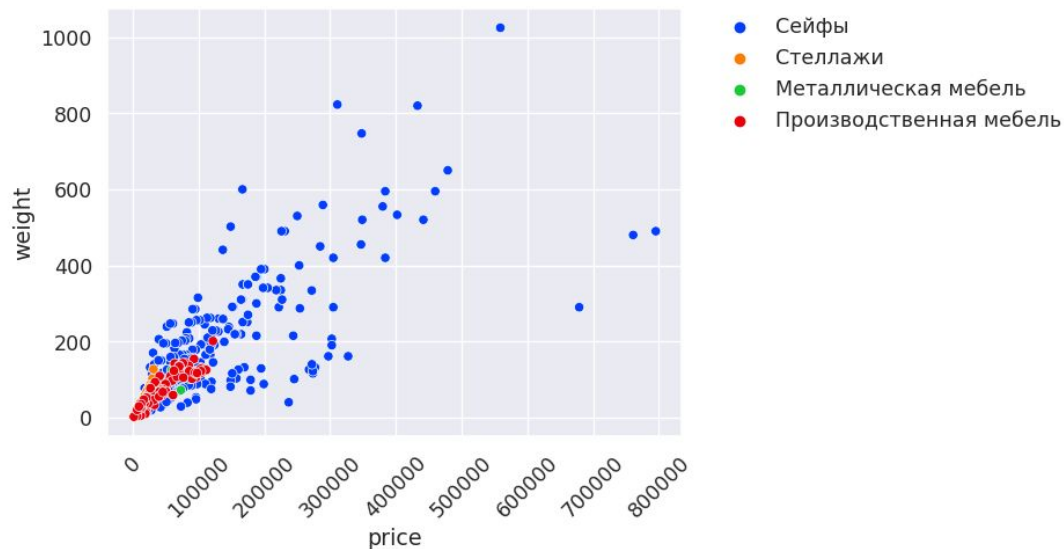
Примерно 85,6% ассортимента есть в наличии, что говорит о хорошей складской программе.



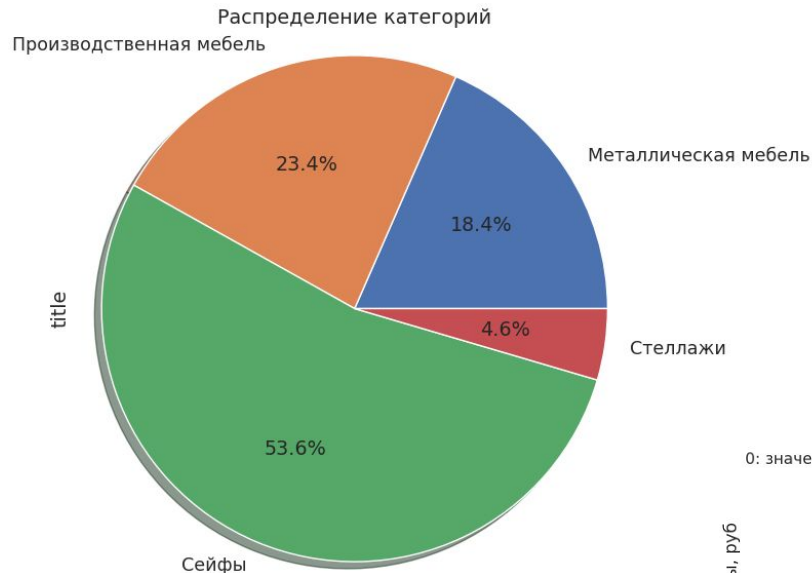
Анализ организации 2



Взаимосвязь цены и веса наиболее заметно сказывается на товарах категории "Сейфы".

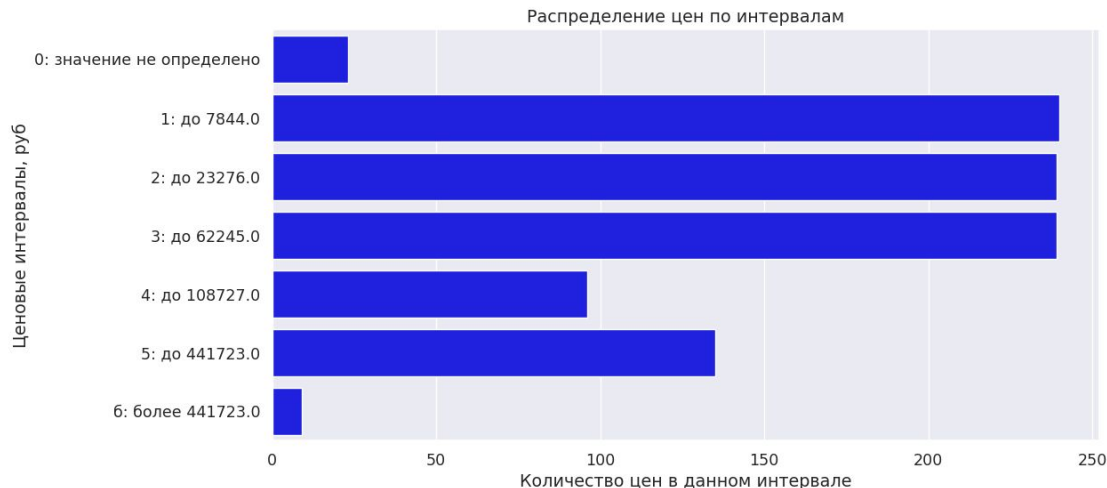


Анализ организации 3



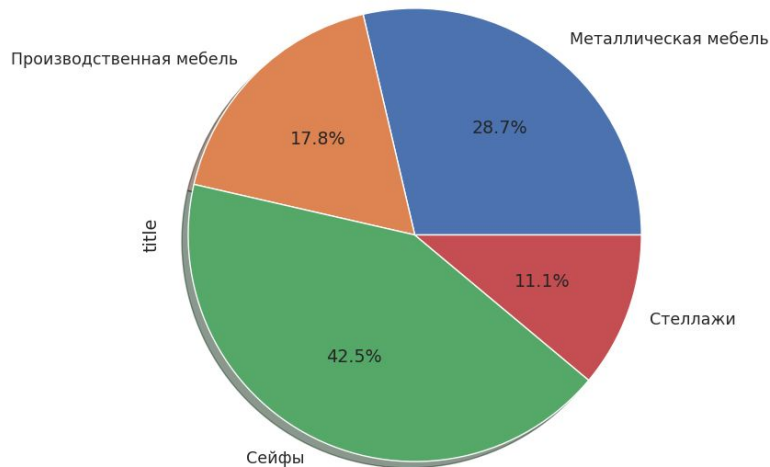
В ассортименте организации 3 больше всего представлен ассортимент категории "Сейфы" = 53,6%.

Подавляющее большинство цен распределены равномерно в интервалах до 74052 руб. - до 75 перцентиля.



Анализ организации 4

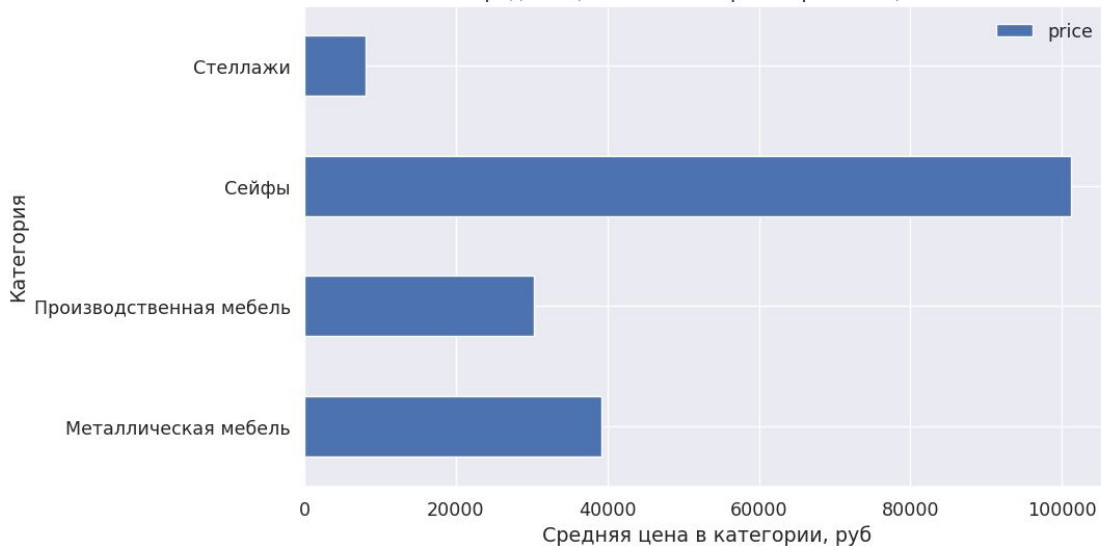
Распределение категорий



В ассортименте организации 4 больше всего представлен ассортимент категории "Сейфы" = 42,5%.

Традиционно - самая дорогая категория "Сейфы", и самая бюджетная "Стеллажи".

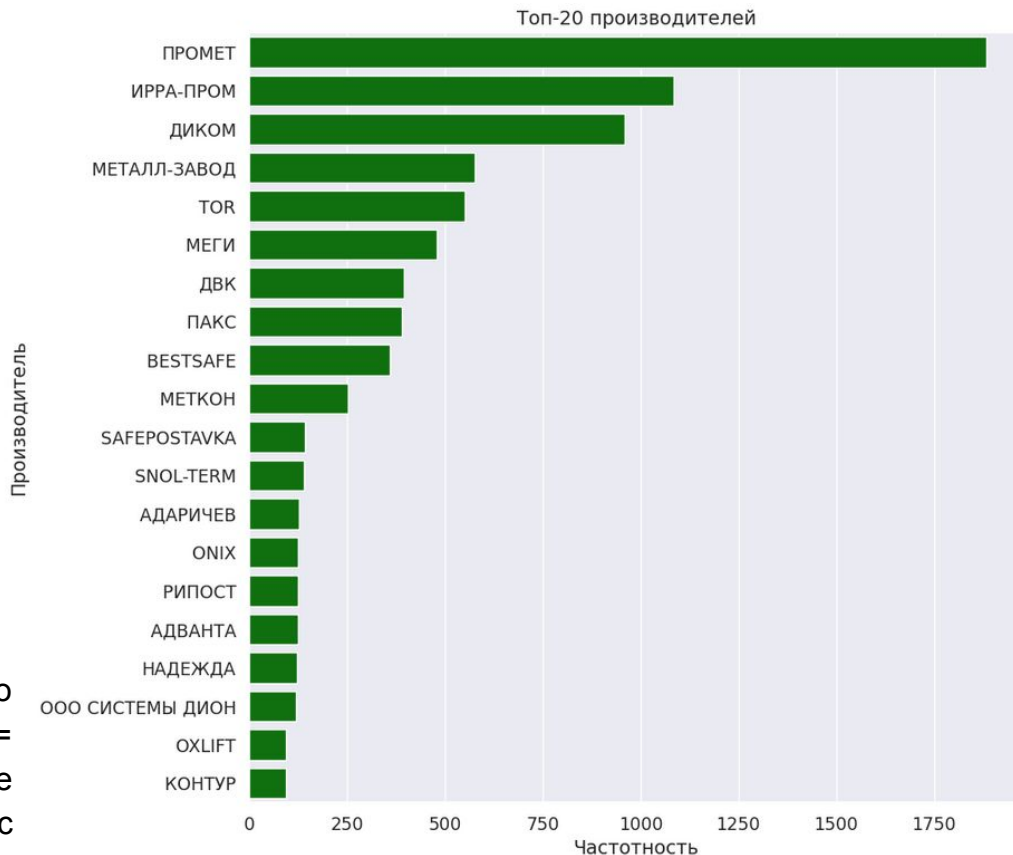
Средние цены по категориям организации 4



Анализ организации 5

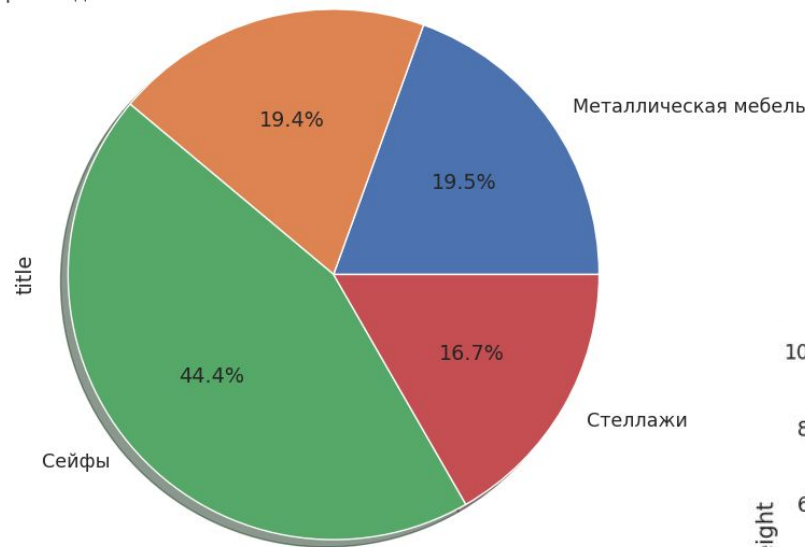


В ассортименте организации 5 больше всего представлен ассортимент категорий "Сейфы" = 32,4% и "Стеллажи" = 26,8%. Наиболее равномерное распределение по сравнению с остальными организациями.



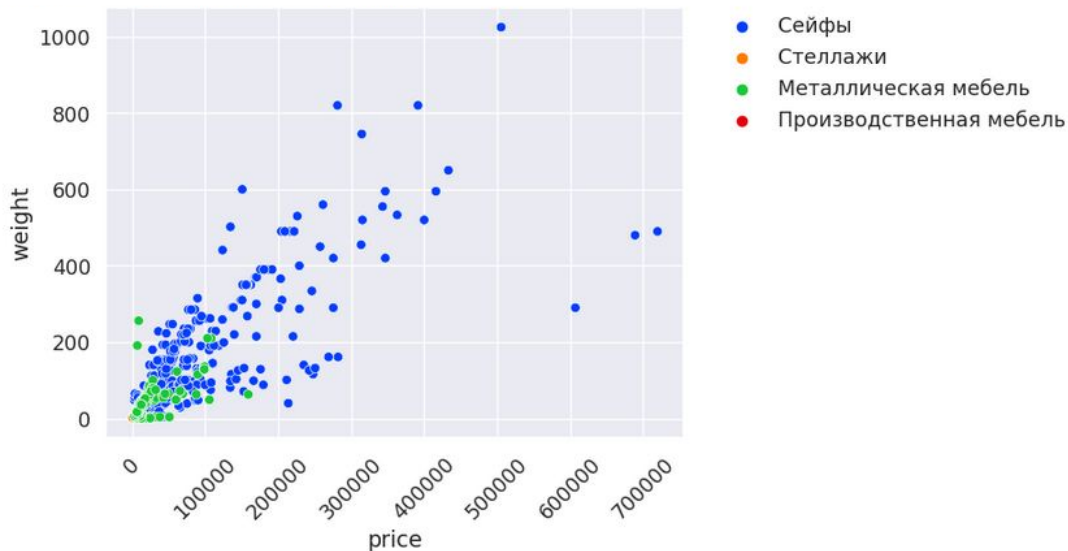
Анализ организации 43

Распределение категорий
Производственная мебель



В ассортименте организации 43 больше всего представлен ассортимент категории "Сейфы" = 44,4%. Остальные категории представлены приблизительно равномерно.

Из корреляционной матрицы можно заметить сильную положительную связь между ценой и весом товара - 0,87. Особенно в отношении сейфов и мет. мебели



Анализ цен по аналогичному ассортименту

- Данные обогащены колонкой “reduced” с сокращённым наименованием
- Объединены в 1 набор данных по “reduced”, добавлены минимальная, максимальная и средняя цены на товар.

3

	price_1	price_2	price_3	price_4	price_5	price_43	min_price	max_price	mean_price
count	742.000000	889.000000	855.000000	243.000000	554.000000	141.000000	1498.000000	1498.000000	1498.000000
mean	49236.994609	63097.084364	65642.748538	61355.016461	31042.617329	109366.063830	52331.750334	59770.785047	56037.969359
std	67101.220978	95826.496269	93325.286499	80849.659793	33513.573457	140706.030105	78015.973049	88976.417346	83335.788171
min	1100.000000	80.000000	70.000000	150.000000	293.000000	71.000000	70.000000	80.000000	75.400000
25%	10655.000000	10696.000000	12911.000000	16236.000000	8608.250000	10010.000000	9773.250000	11172.500000	10592.875000
50%	25520.000000	27590.000000	33142.000000	30498.000000	19500.000000	40700.000000	24163.000000	28920.000000	26907.500000
75%	58065.000000	70072.000000	76302.000000	72672.500000	41147.750000	166155.000000	56942.250000	67592.000000	61739.050000
max	559000.000000	795182.000000	815674.000000	488580.000000	251461.000000	719625.000000	719625.000000	815674.000000	760161.000000

Больше всего пропущенных значений у организации 43 - 90,6%. Можно предположить, что ассортимент организации 43 (поставщика) более уникальный, чем тот, что представлен на местном рынке.



Совпадения по организации 43

		title_43	price_1	price_2	price_3	price_4	price_5	Row_Total
Column_Total		NaN	75.0	36.0	37.0	7.0	28.0	183.0
98	СТЕЛЛАЖ ES 150KD/75X30/4		0.0	1.0	1.0	1.0	0.0	3.0
104	ШКАФ PROFI WD		0.0	1.0	1.0	0.0	1.0	3.0
111	ШКАФ ДЛЯ РАЗДЕВАЛОК ПРАКТИК УСИЛЕННЫЙ ML 21-60...		1.0	1.0	1.0	0.0	0.0	3.0
31	ИНСТРУМЕНТАЛЬНАЯ ТУМБА WDS-9		1.0	0.0	1.0	0.0	1.0	3.0
45	ОГНЕВЗЛОМОСТОЙКИЙ СЕЙФ VALBERG ГАРАНТ EBPO 67T		0.0	1.0	1.0	0.0	0.0	2.0
29	ДЕПОЗИТНЫЙ СЕЙФ VALBERG DSC 67 ЕК		0.0	1.0	1.0	0.0	0.0	2.0
30	ДЕПОЗИТНЫЙ СЕЙФ VALBERG DSC 67 КК		0.0	1.0	1.0	0.0	0.0	2.0
37	КАРТОТЕКА AFC-07		0.0	0.0	0.0	1.0	1.0	2.0
38	ЛОЖЕМЕНТ 1 СТВОЛ		0.0	1.0	1.0	0.0	0.0	2.0

Количество товаров на которые цена организации 43 равняется минимальной - составляет 90.07% (127 из 141 шт.). Таким образом, цены организации 43 (поставщика) являются самыми низкими, среди совпадающих.

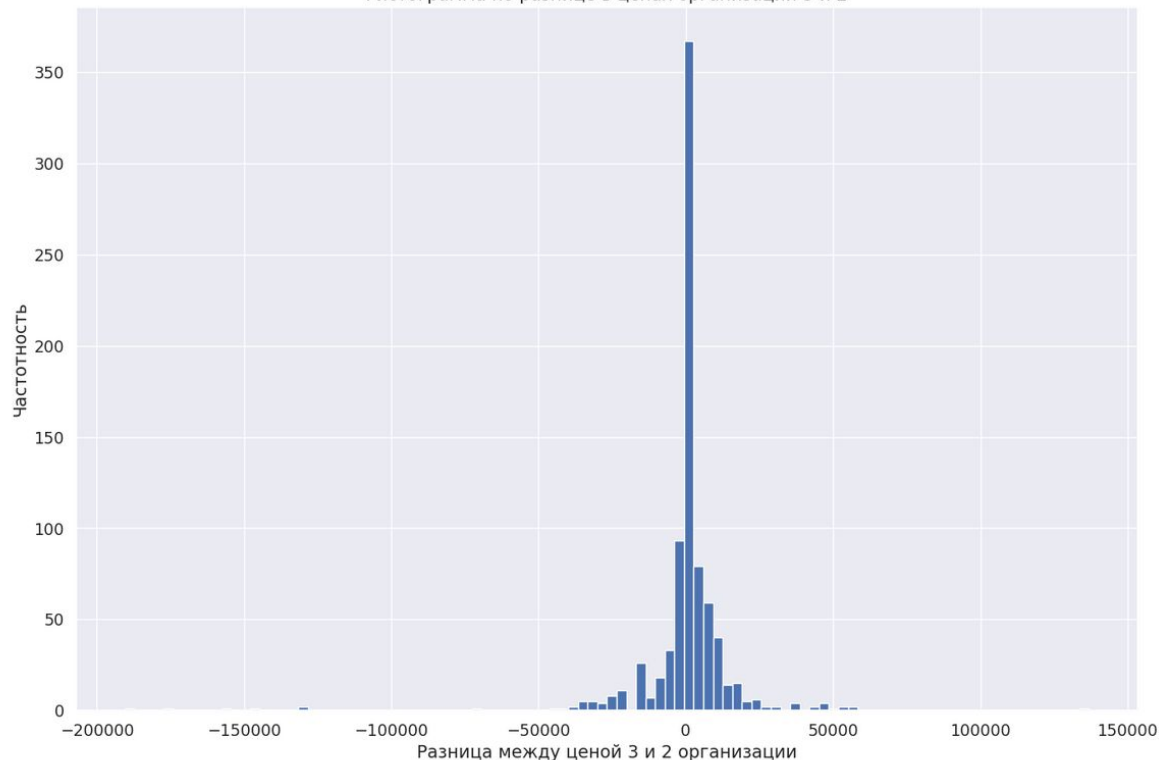
Количество совпадений товаров с организацией 43 составляет:

- 0.69% для организации 1,
- 3.67% для организации 2,
- 4.11% для организации 3,
- 1.14% для организации 4,
- 0.3% для организации 5.



Совпадения по организациям 2 и 3

Гистограмма по разнице в ценах организаций 3 и 2



+ Код + Текст

	price_2	price_3
count	826.000000	826.000000
mean	66012.101695	66140.736077
std	98409.470153	94530.465740
min	80.000000	70.000000
25%	11507.250000	12906.500000
50%	29114.000000	32862.000000
75%	72946.000000	76730.500000
max	795182.000000	815674.000000

Количество совпадений составляет 84.2% для организации 2 и 91.78% для организации 3. Очень высокий показатель совпадений.

В целом цены в организации 3 выше, чем цены на аналогичные товары в организации 2.

Проверка статистических гипотез

H0: Цены товаров организации 43 от минимальных цен статистически значимо не отличаются.

H1: Цены товаров организации 43 от минимальных цен статистически значимо отличаются.

Уровень значимости определен в 5%.

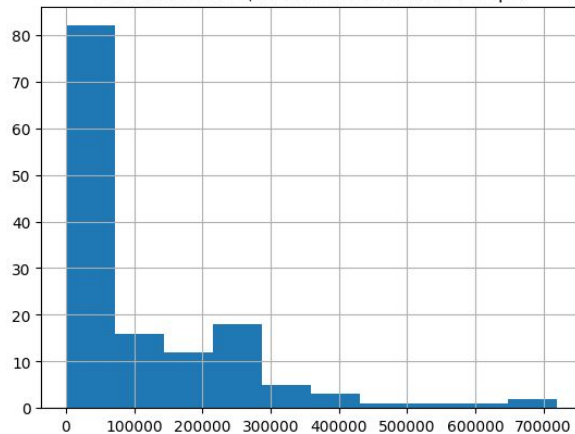
Распределение цен не является нормальным. Воспользуемся тестом Манна-Уитни

```
[ ] stat, p = st.mannwhitneyu(df_match_with_43['price_43'], df_match_with_43['min_price'])  
  
print(f'Статистика = {stat: .5f}, p = {p: .5f}')  
  
if p > 0.05:  
    print('Не отклоняем нулевую гипотезу, распределения, вероятно, одинаковые')  
else:  
    print('Отклоняем нулевую гипотезу, распределения, вероятно, различаются')
```

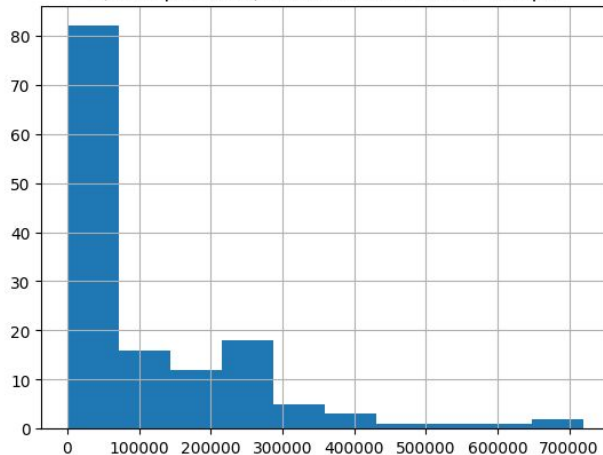
Статистика = 9975.50000, p = 0.95981

Не отклоняем нулевую гипотезу, распределения, вероятно, одинаковые

Минимальные цены на аналогичные товары



Цены организации 43 на аналогичные товары



Выводы и рекомендации



4

Выводы:



В настоящее время местный рынок города N. направления сейфы, стеллажи, металлическая мебель является перспективным в плане освоения. Особенно это актуально с учётом роста производства.



Ценовое предложение организации 43 (поставщика) является конкурентоспособным.

Кроме того, на местном рынке сложился очень близкий ассортимент у организаций 2 и 3. Поэтому предложение с отличными модификациями товаров в рассматриваемых категориях также могут заинтересовать потенциальных заказчиков.

Рекомендации:

Для дальнейшего погружения в нишу сейфов, стеллажей и металлической мебели стоит собрать и проанализировать данные по топовым производителям оборудования.

При запуске нового товарного направления рекомендуется вести учёт и анализ собственных продаж, а также ассортимента и цен конкурентов для своевременной корректировки.



Трудности в процессе работы и как их решали



5

Трудности в процессе работы и как их решали

- Разнородность данных, т.е у разных организаций на сайтах, с которых собирались данные, информация представлена в разном формате и объёме.
Выход: в приоритете собирала максимально обобщённую информацию по наименованию и цене, плюс то, что могло быть интересным, например, наличие.
- На сайте организации 5 большое количество ошибок, которые не сразу бросались в глаза.
- Сопоставление данных по аналогичным товарам, т.к. один и тот же товар имеет зачастую совершенно разное написание. Например, “Шкаф мебельный ШМ-4” и “ШКАФ БУХГАЛТЕРСКИЙ ШМ-4” это на самом деле один и тот же шкаф.
Выход: приведение к верхнему регистру, поиск и удаление неуникальных значений. Так и появился столбец “reduced”.



Спасибо за внимание!

