

Дипломная работа по теме: «Анализ расширения торгового направления ниши сейфов, стеллажей и металлической мебели (поиск инсайтов, составление рекомендаций стейкхолдерам)»

Программа: «Разработчик - Data Engineer»

Ласточкина Елена Александровна

Оглавление

Оглавление	2
1. Введение	4
2. Сбор данных, их описание	4
3. Подготовка, преобразование и очистка данных	5
3. 1 Аномалии в данных	7
4. Анализ данных для стейкхолдеров	7
4.1 Анализ по отдельным организациям	7
4.1.1 Организация 1	7
4.1.2 Организация 2	8
4.1.3 Организация 3	9
4.1.4 Организация 4	9
4.1.5 Организация 5	10
4.1.6 Организация 43	11
Выводы по части 4.1	13
Рекомендация стейкхолдерам	13
4.2 Анализ цен по аналогичному ассортименту	13
4.2.1 Верхнеуровневый анализ	14
4.2.2 Совпадения по организации 43	14
4.2.3 Совпадения ассортимента организаций 2 и 3	15
4.2.4 Совпадения ассортимента организаций 1 и 5	16
Выводы по части 4.2	16
Рекомендация стейкхолдерам	17
5. Проверка статистических гипотез	17
5.1 Формулировка нулевой и альтернативной гипотезы	17
Вывод по части 5	17
6. Заключение	17

1. Введение

Цели проекта:

В рамках работы в небольшой компании в городе N., планирующей запустить новое направление деятельности в нише сейфов, стеллажей, металлической и производственной мебели, провести анализ ассортимента конкурентов и потенциального поставщика с целью определения перспективности освоения данной ниши.

Бизнес-Задачи:

1. 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;
2. Анализ предложений конкурентов с предложением поставщика в области ценовой политики на основании сравнения аналогичного ассортимента.

Стейкхолдеры

Основной стейкхолдер: руководитель компании (для чёткого понимания ситуации на местном рынке, а также для оценки перспективности запуска нового направления).

2. Сбор данных, их описание

Данные для исследования были получены путём парсинга официальных сайтов организаций, занимающихся направлением сейфов, стеллажей, металлической и производственной мебели на местном рынке города N. в количестве 5 шт., а также 1 сайта потенциального поставщика.

Таким образом были получены 6 наборов данных, нуждающихся в дополнительной обработке и очистке. Для наборов данных, касающихся конкурентов были присвоены номера 1-5, для поставщика - 43.

Набор данных	Количество строк
1	10 947
2	981
3	900
4	614
5	9 420
43	1 108
Итого	23 970

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	+	-	-	-	-	-

3. Подготовка, преобразование и очистка данных

Над первоначальными данными были произведены следующие преобразования:

- Добавлен столбец “category”, тип данных object. В котором указана категория товара: Сейфы, Стеллажи, Металлическая мебель, Производственная мебель.
- Добавлен столбец “shop”, тип данных int64. С обозначением обезличенного номера конкурента/поставщика.
- В столбце “price” изменён тип данных на float64.
- Из столбца “size” были выделены высота, ширина, глубина и записаны в соответствующие колонки “height”, “width”, “depth” с типом данных float64.

- В столбце “weight” для набора данных №43 изменён тип данных на float64.
- Для наборов данных №1 и №2 были выделены значения “weight” из колонки “description” с типом данных float64.
- Для наборов данных №1 значения колонки “availability” были изменены на 0 - “Нет в наличии” и 1 - “Есть в наличии”. Тип данных int64.
- Для набора №5 были выделены значения “manufacturer” из колонки “description” с типом данных object.
- Колонки “title” были очищены от служебных символов и приведены к верхнему регистру.
- Из набора данных организации 5 был удалён столбец “depth” ввиду своей некорректности (ошибка на официальном сайте организации - при выводе списка товаров в описании габаритов глубина, в большинстве случаев, дублирует ширину товара).
- Для всех наборов данных была добавлена колонка “reduced” с сокращённым названием товаров¹.

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	+	+	+	+	+	+

¹ Только для анализа цен на аналогичный ассортимент.

3. 1 Аномалии в данных

В наборе данных организации 5 было обнаружено аномальное значение высоты для “Шкаф для одежды ТМ 12-60 по ГОСТу” равное 18350мм (в действительности 1830мм). Данное значение было отнесено к выбросам.

4. Анализ данных для стейкхолдеров

Целью данного раздела является содержательное описание данных, а также поиск инсайтов для составления рекомендаций стейкхолдерам.

Общая информация по организациям

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

Организация	Сейфы	Стеллажи	Мет. мебель	Произв. мебель	Всё вместе	Место по разнообр. 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”.

4.1 Анализ по отдельным организациям

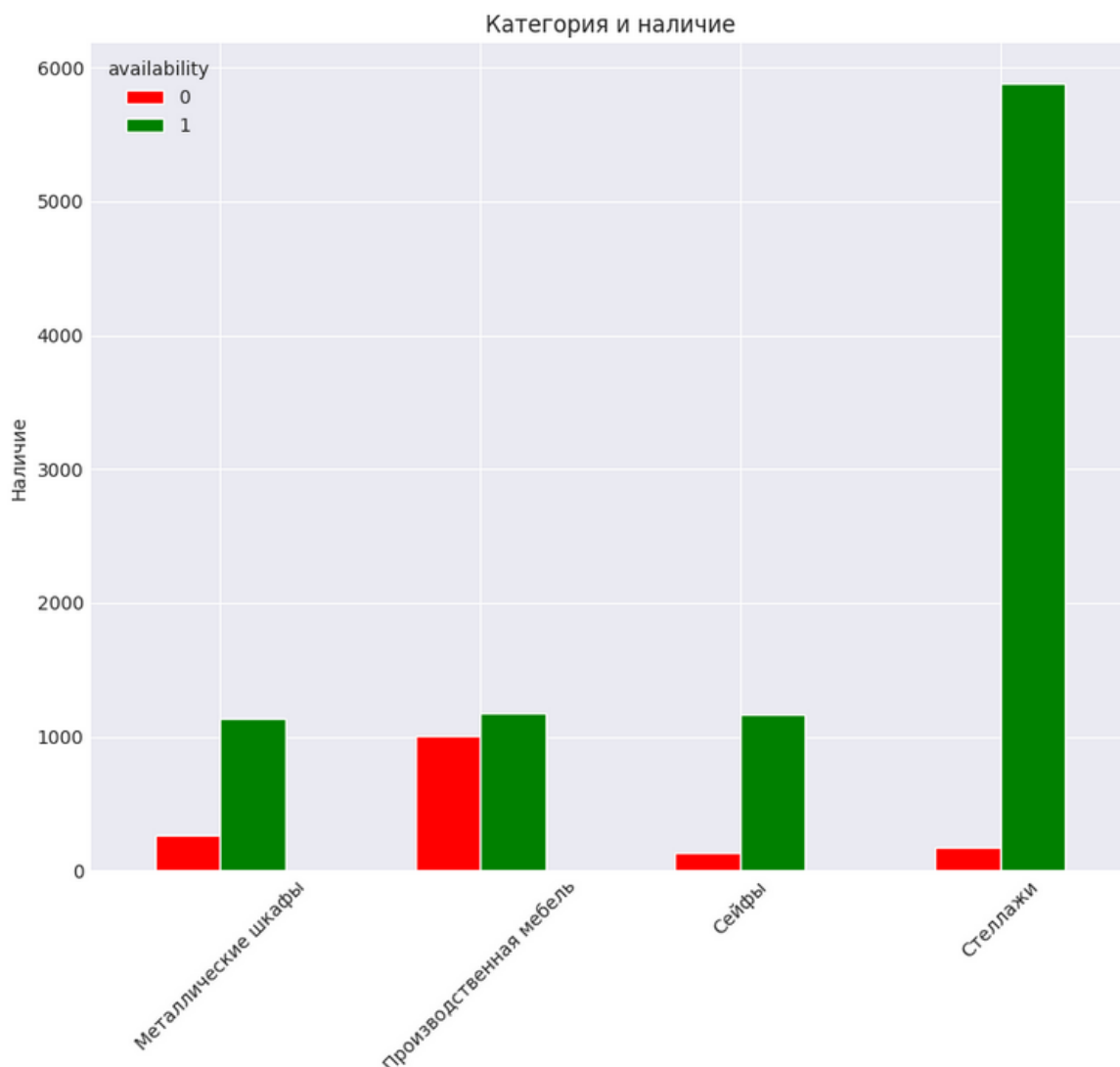
4.1.1 Организация 1

Данный датасет отличается тем, что имеются данные по наличию товара.

	title
availability	
0	1577
1	9370

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

В ассортименте организации 1 больше всего представлен ассортимент категории "Стеллажи" = 55,3%.

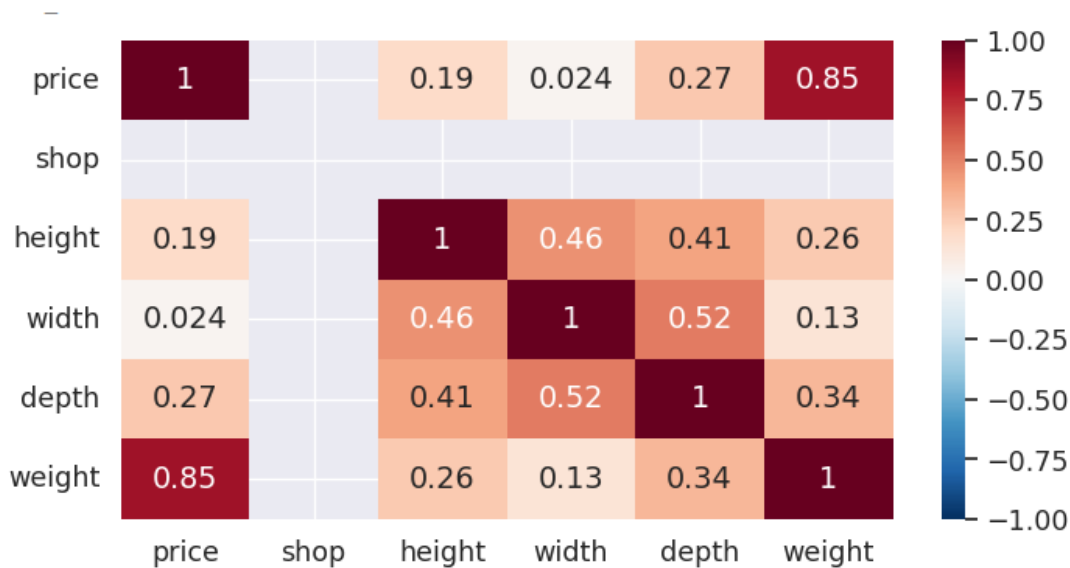


Как отмечалось ранее, самая популярная категория для 1 организации - "Стеллажи". В наличии также больше всего поддерживается ассортимент из данной категории. Меньше всего в наличии производственной мебели. Можно предположить, что заказы на данную продукцию чаще принимаются под заказ для нужд производственных предприятий.

4.1.2 Организация 2

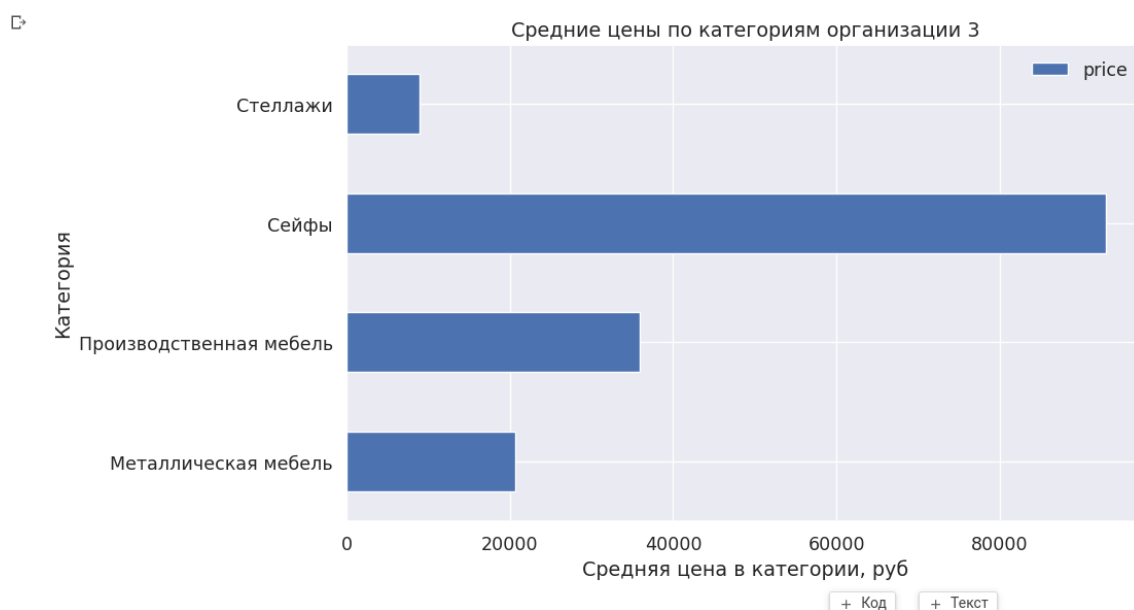
В ассортименте организации 2 больше всего представлен ассортимент категории "Сейфы" = 49,3%.

Из приведённой ниже корреляционной матрицы можно заметить сильную положительную связь между ценой и весом товара - 0,85.



4.1.3 Организация 3

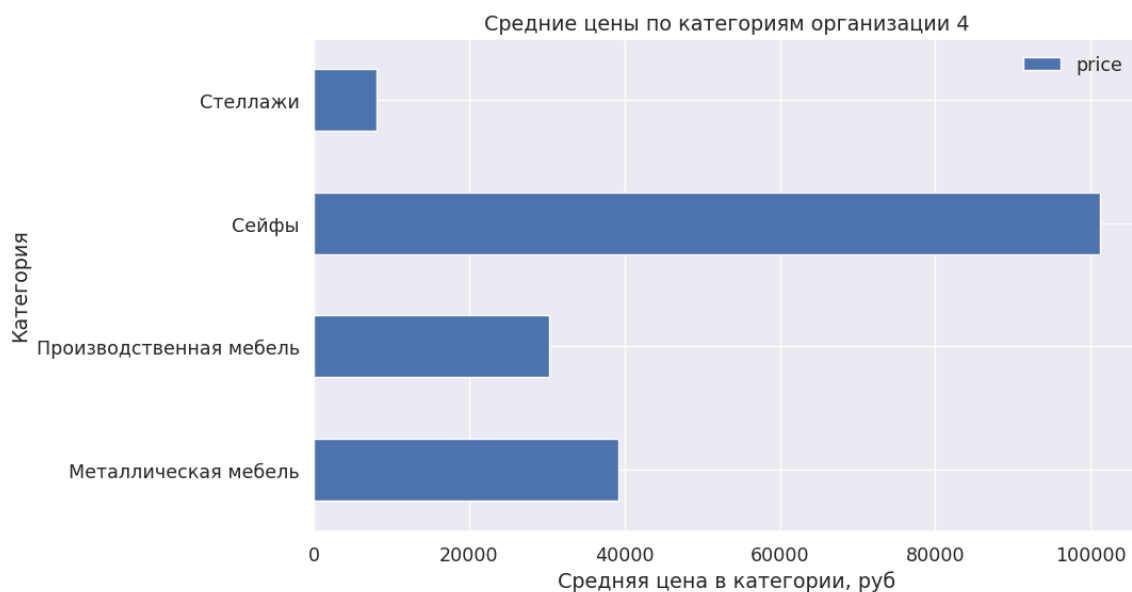
В ассортименте организации 3 больше всего представлен ассортимент категории "Сейфы" = 53,6%.



Подавляющее большинство цен распределены равномерно в интервалах до 74052 руб.

4.1.4 Организация 4

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



Подавляющее большинство цен распределены равномерно в интервалах до 71754 руб.

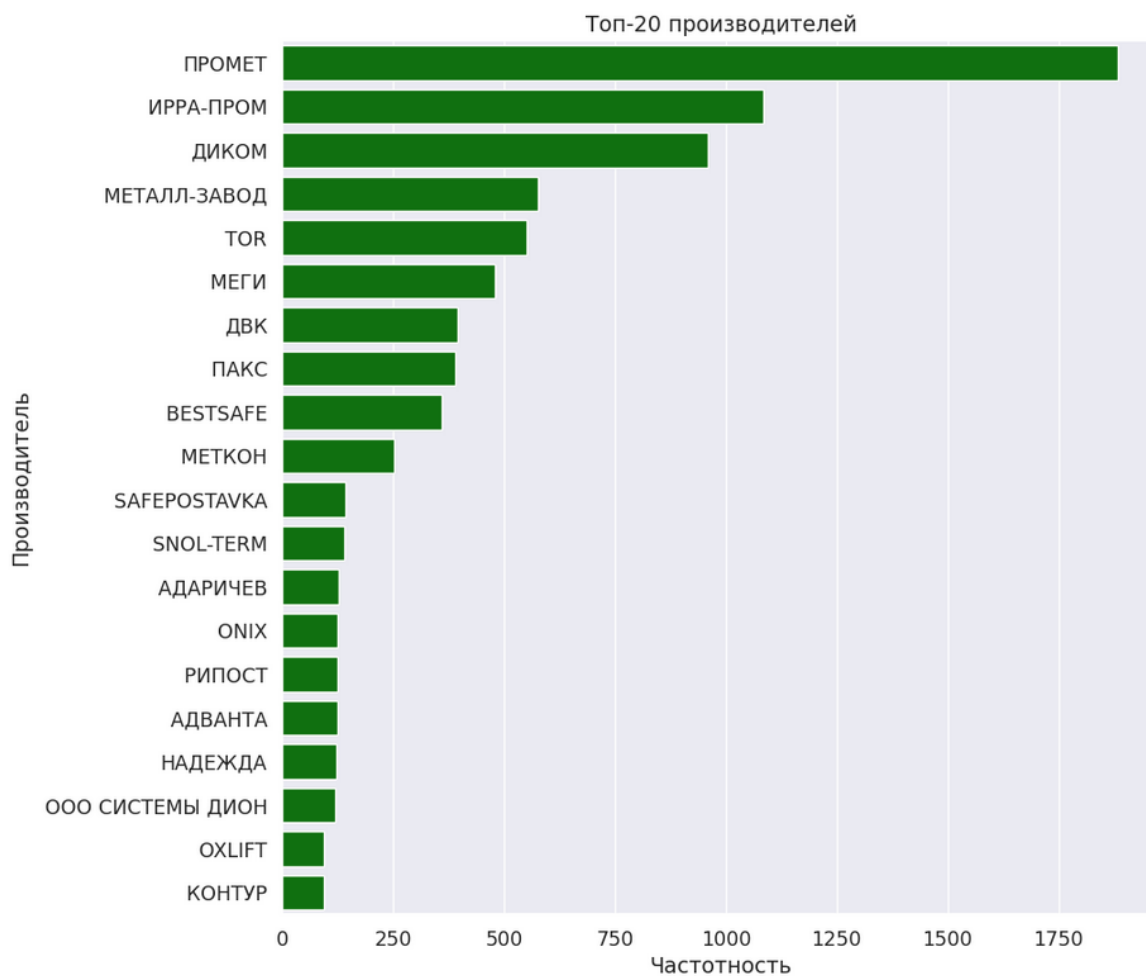
4.1.5 Организация 5

Данный датасет отличается тем, что имеются данные по производителю товара.



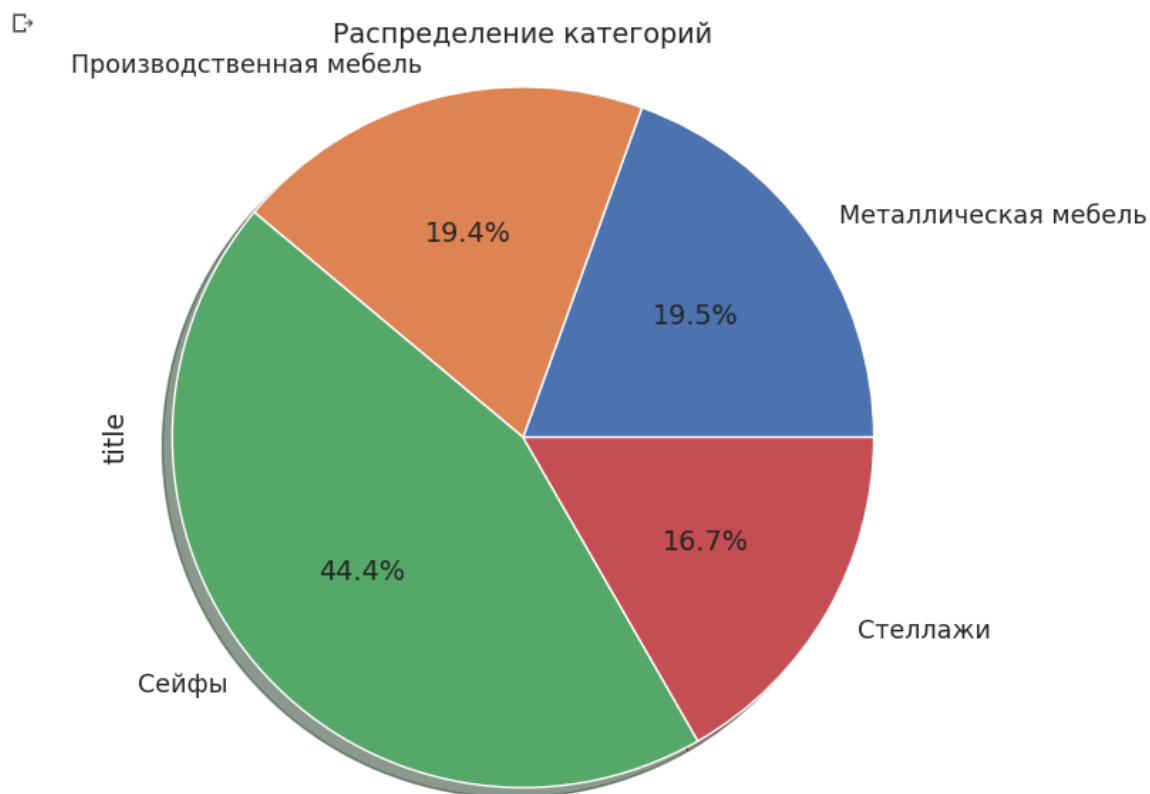
В ассортименте организации 5 больше всего представлен ассортимент категорий "Сейфы" = 32,4% и "Стеллажи" = 26,8%. Наиболее равномерное распределение по сравнению с остальными организациями.

Больше всего товаров представлено от таких производителей как ПРОМЕТ, ИРРА-ПРОМ, ДИКОМ.

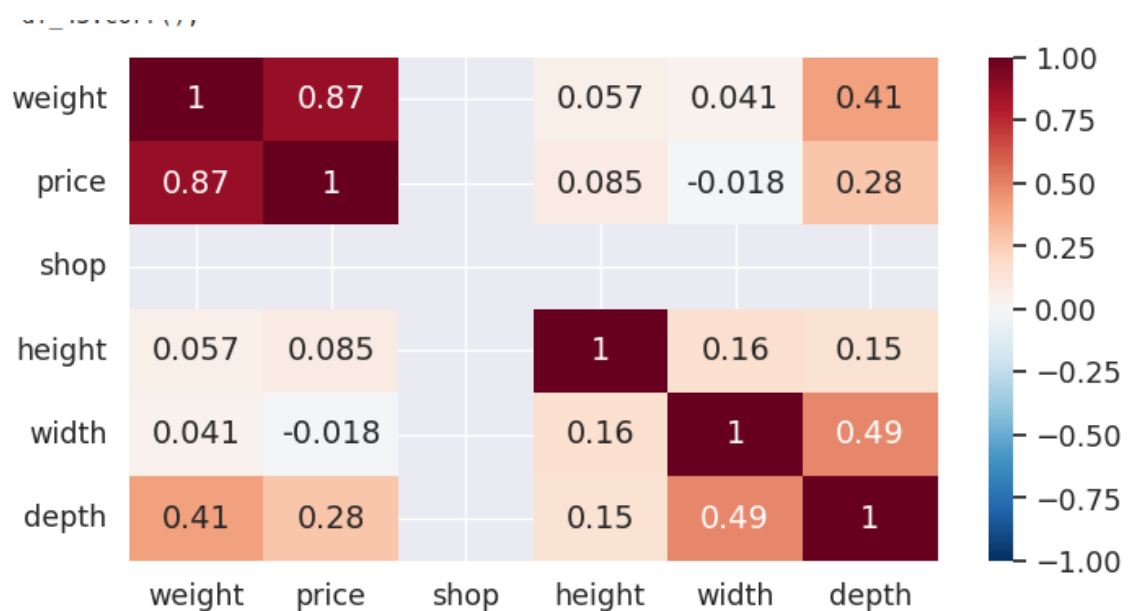


4.1.6 Организация 43

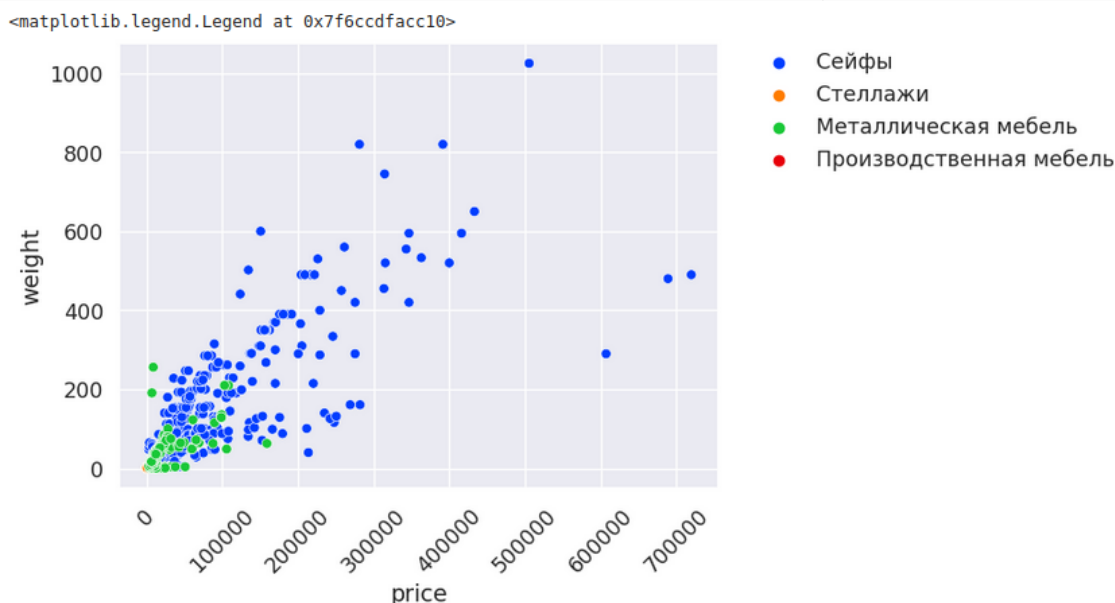
В ассортименте организации 43 больше всего представлен ассортимент категории "Сейфы" = 44,4%. Остальные категории представлены приблизительно равномерно.



Из приведённой ниже корреляционной матрицы можно заметить сильную положительную связь между ценой и весом товара - 0,87.



Взаимосвязь цены и веса наиболее заметно сказывается на товарах категории "Сейфы". Также заметно подобное влияние в категории "Металлическая мебель"



Выводы по части 4.1

На местном рынке города N. направление сейфов, стеллажей и металлической мебели представлено всеми перечисленными категориями. Однако наибольший уклон идёт на категорию “Сейфы” - именно она идёт в приоритете у анализируемых организаций.

Разнообразие ассортимента организации 43 (поставщика) конкурентоспособно по сравнению с другими организациями. 3-е условное место в рейтинге (см. Табл.1) - 1108 наименований. А также достаточно равномерное распределение между категориями.

Рекомендация стейкхолдерам

Перспективы запуска нового направления в компании есть, особенно, учитывая фактор роста производственных предприятий в городе. Однако есть смысл уделить больше внимания на менее популярные категории, чем у конкурентов - на стеллажи и металлическую мебель.

4.2 Анализ цен по аналогичному ассортименту

Наиболее интересными представляются данные по цене и по ассортименту организаций, поэтому данные всех 6 датасетов были дополнены колонкой “reduced” - сокращённым названием товара. Для данной операции предварительно все символы “title” были приведены к верхнему регистру, а также были найдены и удалены наименее уникальные слова, характеризующие модель товара. После чего все наборы данных были объединены в 1 датасет по столбцу “reduced”, также были добавлены “title” и “price” от каждой

организации. Также данные были обогащены минимальной, максимальной и средней ценой за товар (столбцы “min_price”, “max_price”, “mean_price” соответственно).

4.2.1 Верхнеуровневый анализ

	Missing Values	% of Total Values
title_43	1358	90.6
price_43	1358	90.6
price_4	1256	83.8
title_4	1254	83.7
price_5	945	63.0
title_5	923	61.6
price_1	757	50.5
title_1	693	46.2
price_3	644	43.0
title_3	643	42.9
price_2	610	40.7
title_2	609	40.6
min_price	1	0.1
max_price	1	0.1
mean_price	1	0.1

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.

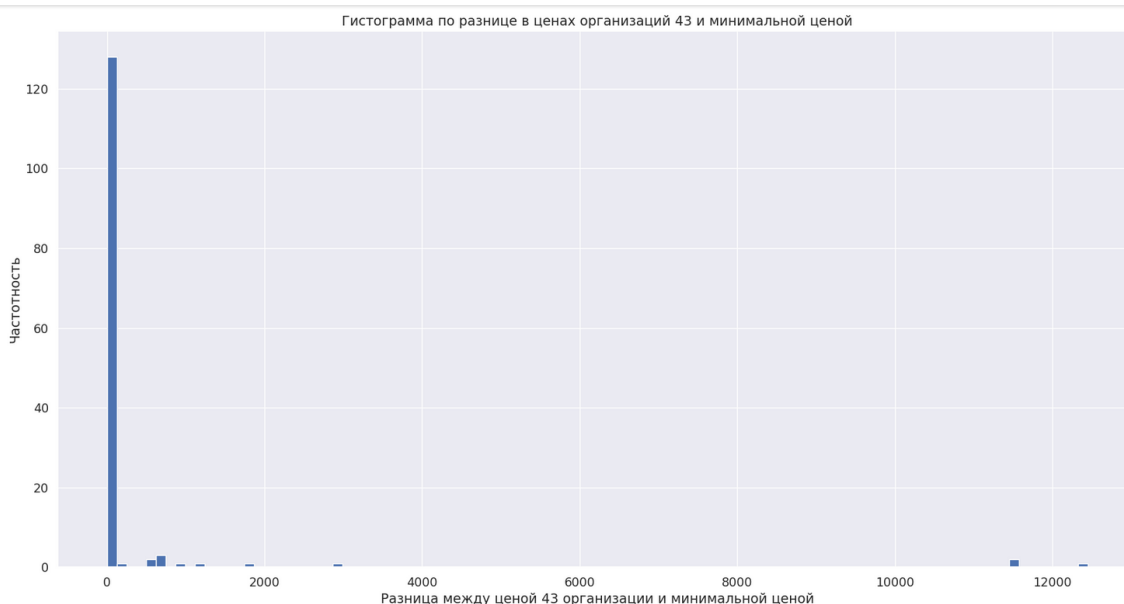
4.2.2 Совпадения по организации 43

Были добавлены результирующие строка и столбец с количеством совпадений “Column_Total” и “Row_Total” соответственно.

		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	0.0	3.0
104	ШКАФ PROFI WD	0.0	1.0	1.0	0.0	1.0	0.0	3.0
111	ШКАФ ДЛЯ РАЗДЕВАЛОК ПРАКТИК УСИЛЕННЫЙ ML 21-60...	1.0	1.0	1.0	0.0	0.0	0.0	3.0
31	ИНСТРУМЕНТАЛЬНАЯ ТУМБА WDS-9	1.0	0.0	1.0	0.0	1.0	0.0	3.0
45	ОГНЕВЗЛОМОСТОЙКИЙ СЕЙФ VALBERG ГАРАНТ EBPO 67T	0.0	1.0	1.0	0.0	0.0	0.0	2.0
29	ДЕПОЗИТНЫЙ СЕЙФ VALBERG DSC 67 EK	0.0	1.0	1.0	0.0	0.0	0.0	2.0
30	ДЕПОЗИТНЫЙ СЕЙФ VALBERG DSC 67 KK	0.0	1.0	1.0	0.0	0.0	0.0	2.0
37	КАРТотека AFC-07	0.0	0.0	0.0	1.0	1.0	0.0	2.0
38	ЛОЖЕМЕНТ 1 СТВОЛ	0.0	1.0	1.0	0.0	0.0	0.0	2.0

Таким образом 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.



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.

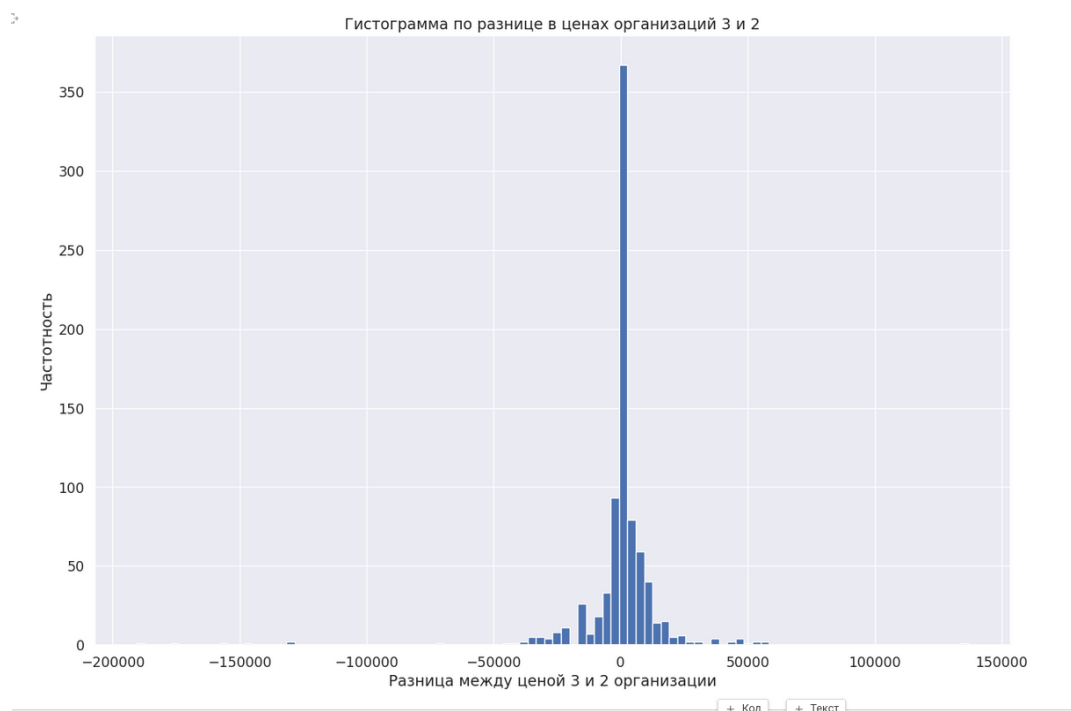
4.2.3 Совпадения ассортимента организаций 2 и 3

Был замечен высокий уровень совпадений товаров данных организаций, поэтому имеет смысл обратить на них отдельное внимание.

В организациях 2 и 3 общее количество товаров 981 и 900 соответственно.

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.



4.2.4 Совпадения ассортимента организаций 1 и 5

Данные организации имеют самый широкий ассортимент из всех.

Общее кол-во совпадений в целом по товарам было 1499.

457 совпадений 1 и 5 организаций составляет порядка 1/3 от числа всех совпадений.

В организациях 1 и 5 общее количество товаров 10947 и 9420 соответственно.

Количество совпадений составляет 4.17% для организации 1 и 4.85% для организации 5. В целом показатель совпадений не высокий.

В целом цены на аналогичные товары ниже в организации 5.

Выводы по части 4.2

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.

Рекомендация стейкхолдерам

Руководителю и менеджерам компании следует обратить внимание на популярный в регионе ассортимент, и, проанализировав его характеристики, выдвинуть собственное отличное от имеющегося предложение потенциальным заказчикам.

5. Проверка статистических гипотез

Целью данного раздела является проверка конкурентоспособности цен от организации 43 (потенциального поставщика) по данным на аналогичные товары с другими организациями по критерию минимальной цены.

Уровень значимости определен в 5%.

5.1 Формулировка нулевой и альтернативной гипотезы

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

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

Вывод по части 5

Распределение цен не является нормальным. Поэтому для проверки гипотез использовался тест Манна-Уитни

Были получены следующие результаты:

Статистика = 9975.50000, $p = 0.95981$

Не отклоняем нулевую гипотезу, распределения, вероятно, одинаковые. Цены товаров организации 43 от минимальных цен статистически значимо не отличаются и являются конкурентоспособными для рассматриваемого региона.

6. Заключение

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

По бизнес-задачам:

1. Был проведён общий анализ ассортимента по направлению сейфов, стеллажей, металлической и производственной мебели на местном рынке города N. по специализированным организациям, имеющим оффлайн точки продаж, что позволило компании получить цельное представление о текущем положении дел;
2. Был проведён сравнительный анализ предложений конкурентов (организации 1-5) с предложением потенциального поставщика (организация 43) в области ценовой политики на основании сравнения аналогичного ассортимента.

Общие выводы по разделам:

1. На местном рынке города N. направление сейфов, стеллажей и металлической мебели представлено всеми перечисленными категориями. Однако наибольший уклон идёт на категорию “Сейфы” - именно она идёт в приоритете у анализируемых организаций. Разнообразие ассортимента организации 43 (поставщика) конкурентоспособно по сравнению с другими организациями. 3-е условное место в рейтинге (см. Табл.1) - 1108 наименований. А также достаточно равномерное распределение между категориями.
2. 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.