

Итоговый проект по программе «Data-аналитик: старт карьеры»

Поток № ДА-101/2023

Сердюк Наталья Александровна





Описание проекта:

Название проекта:

«Анализ товарного предложения персональных компьютеров в онлайн-магазинах»

Бизнес-цель заказчика:

Выйти на онлайн-рынки Wildberries, Ozon, OnlineTrade с товаром персональные компьютеры (ПК) и увеличить количество продаж на 50% за период 6 мес. Снизить затраты на содержание собственного магазина, доставку и рекламу товаров на 10% за период 6 мес.

Объект исследования:

Персональные компьютеры, которые продаются на онлайн-площадках (маркетплейсах).

Предмет исследования:

Цена, популярность (количество продаж), в зависимости от технических характеристик ПК: процессор, оперативная память, жесткий диск, видеопроцессор.

Цель анализа:

Определить взаимосвязь цены и популярности от технических характеристик, ценовой диапазон, наиболее популярные комплектующие.

2

Этапы исследования:

Разработка аналитического решения:

Определение бизнес-цели и цели анализа, источников данных.

Предварительная обработка данных:

Очистка и преобразование данных для дальнейшего анализа с помощью Python.

Разведочный анализ данных:

Изучение данных и оценка признаков, анализ и визуализация, выявление их взаимосвязей в Python.

Статистическое исследование данных:

Определение типов данных, измерительных шкал. Оценка связи между признаками (корреляционный анализ, таблицы сопряженности, критерии сравнения групп).

Интерпретация анализа:

Заказчику предоставлен аналитический отчет с выводами и рекомендациями.



Результаты анализа:

Гипотеза исследования:

Технические характеристики, такие как:

- процессор (тип и кол-во ядер),
- оперативная память (тип и объем),
- жесткий диск (HDD и SSD),
- видеопроцессор,

оказывают значительное влияние на цену и популярность персональных компьютеров при продаже в онлайн-магазинах.

Методы проверки гипотезы:

1) Построение графиков и таблиц: боксплоты (диаграммы размаха), гистограммы, круговые диаграммы, диаграммы распределения, тепловые карты.

2) Статистические методы:

- Описательные статистики: центральная тенденция, дисперсия и форма распределения набора данных.
- Статистические критерии: о нормальности распределения и выявления взаимосвязей – критерии Спирмена, Х², Шапиро-Уилка, Краскала-Уолиса, Манна-Уитни.

Целевые показатели:

По цене (price) проводился анализ всей совокупности значений в качестве количественного признака.

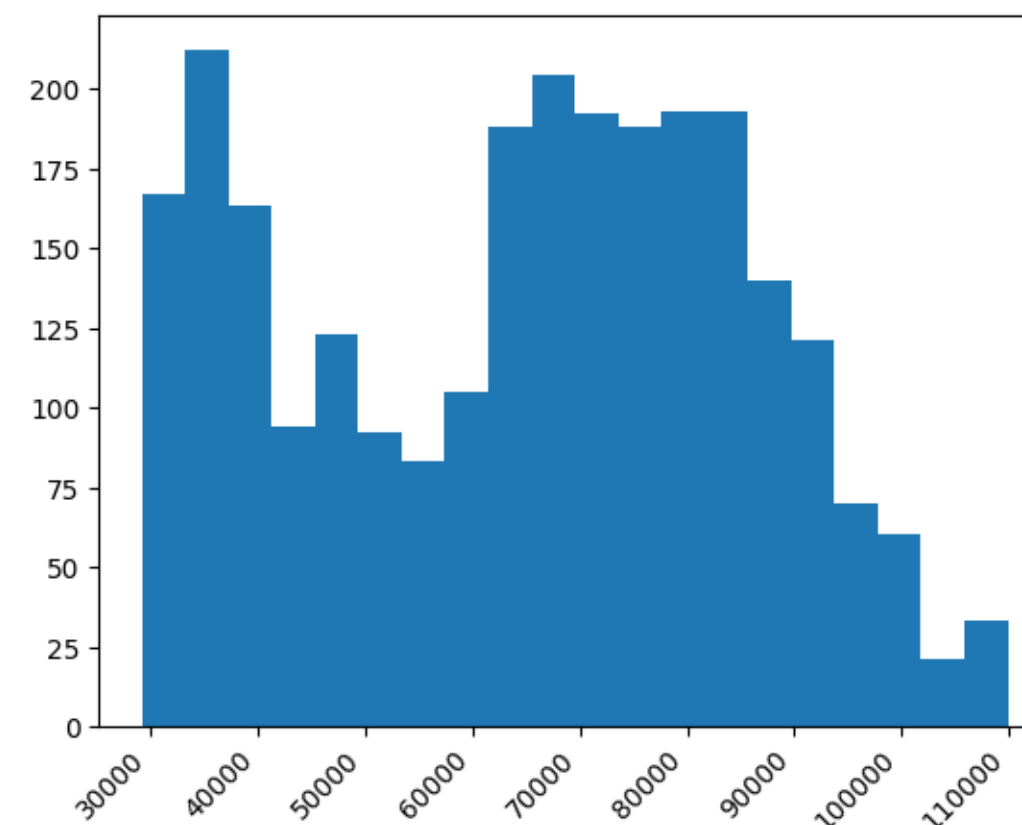
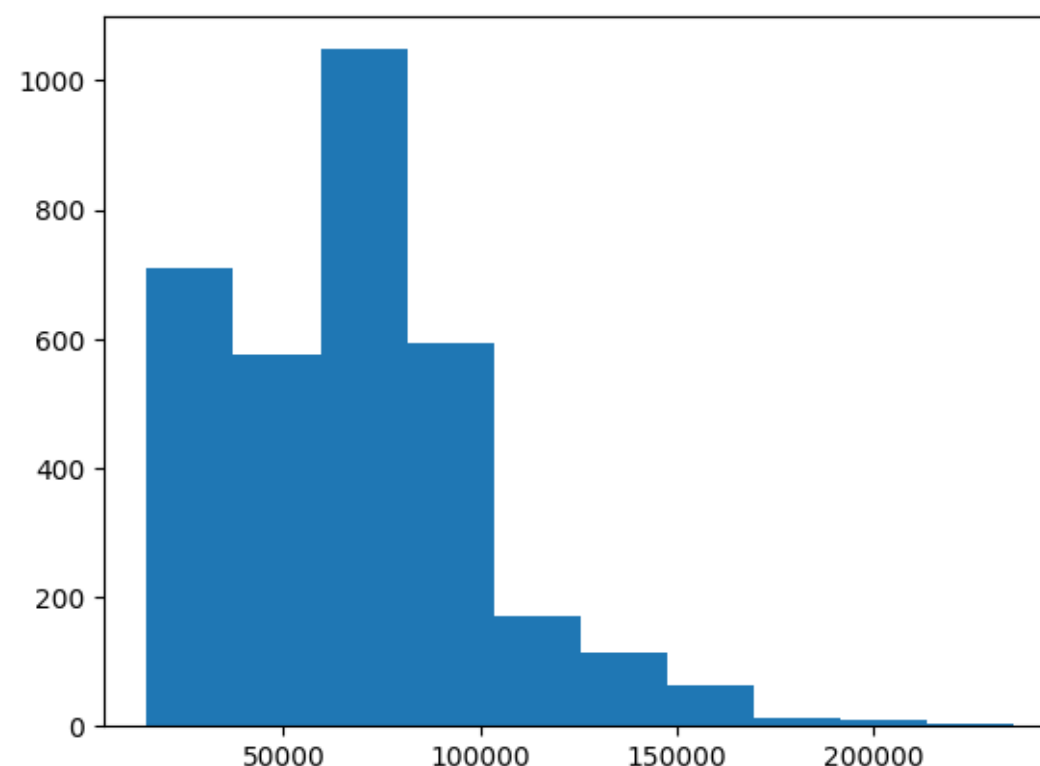
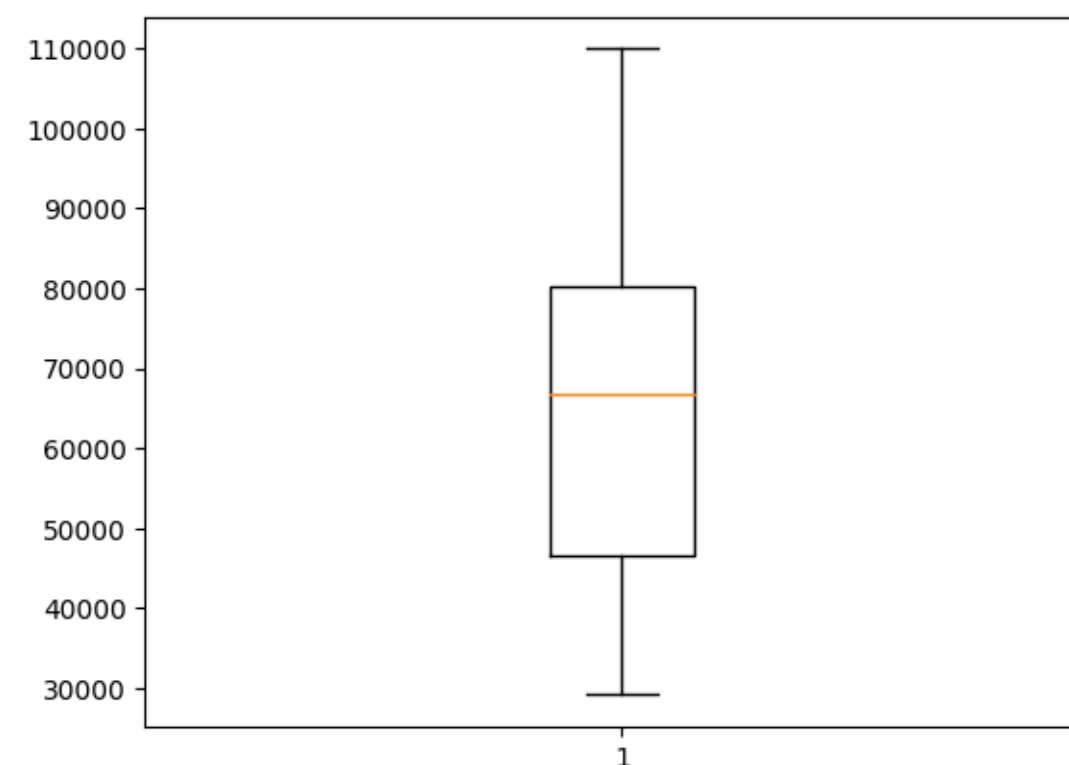
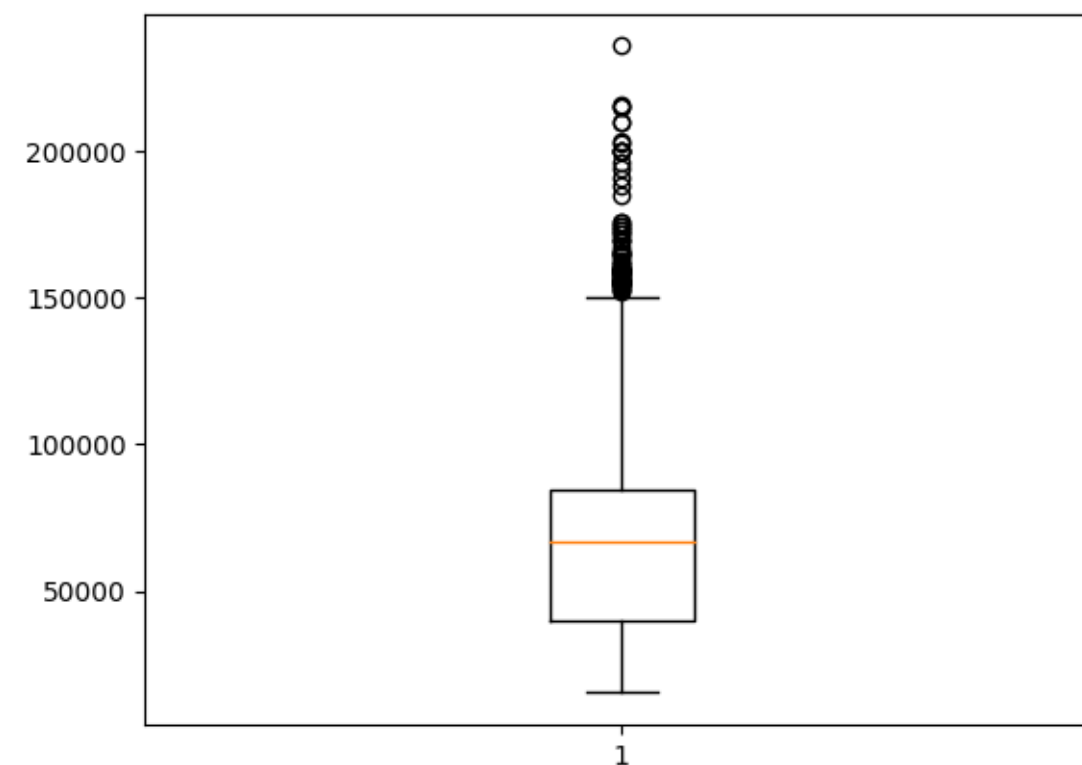
По популярности (кол-во продаж sales) – без учета нулевых продаж, значения разделены на группы: 1 группа 5-50 продаж, 2 группа > 50 продаж, анализ в качестве и категориального и количественного признака.



Результаты анализа:

Рассмотрим анализ цены:

Боксплот и гистограмма показывают на большой размах в данных, что подтверждается и показателями описательной статистики (min=15222 и max=235900).



По графикам основная часть выборки находится в диапазоне цен от 30 до 85 тыс. руб.

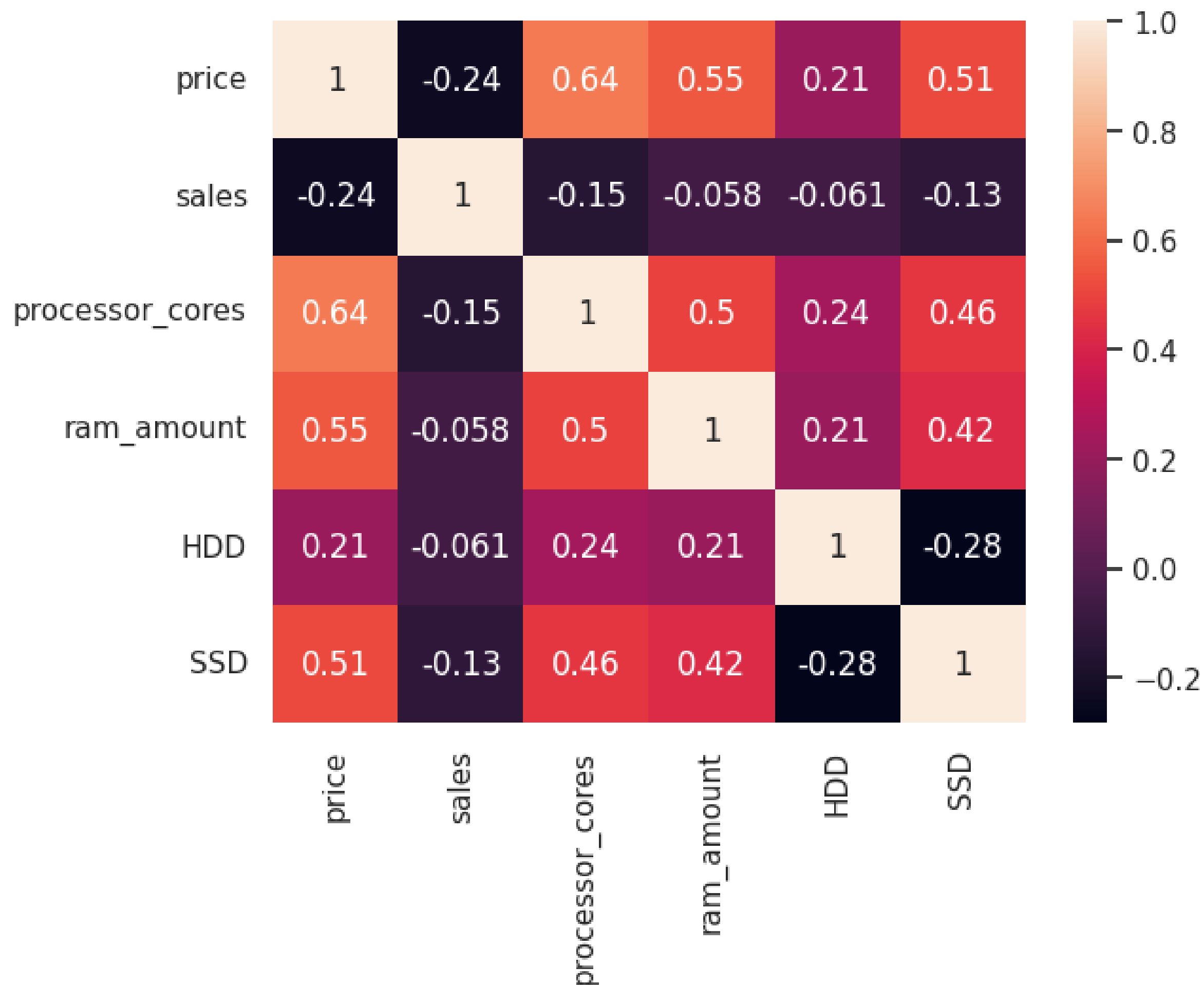
Условно можно выделить ценовые диапазоны:

- эконом-сегмент - 30-40 тыс. руб.
- средний - 40-60 тыс. руб.
- премиум - 60-85 тыс. руб.

Наличие выбросов говорит о более дорогих моделях компьютеров, количество которых уменьшается с ростом цены.



Результаты анализа:



Оценку силы связи цены и количественных факторов можно увидеть с помощью тепловой карты:

- по целевым показателям price и sales - связь обратная, слабая, т.е. при увеличении цены, снижается кол-во продаж;
- price и processor_cores - связь заметная;
- price и ram_amount - связь заметная;
- price и HDD - связь слабая;
- price и SSD - связь заметная.

Расчет коэф-та Спирмена показывает, что связь между признаками есть и она статистически значима.

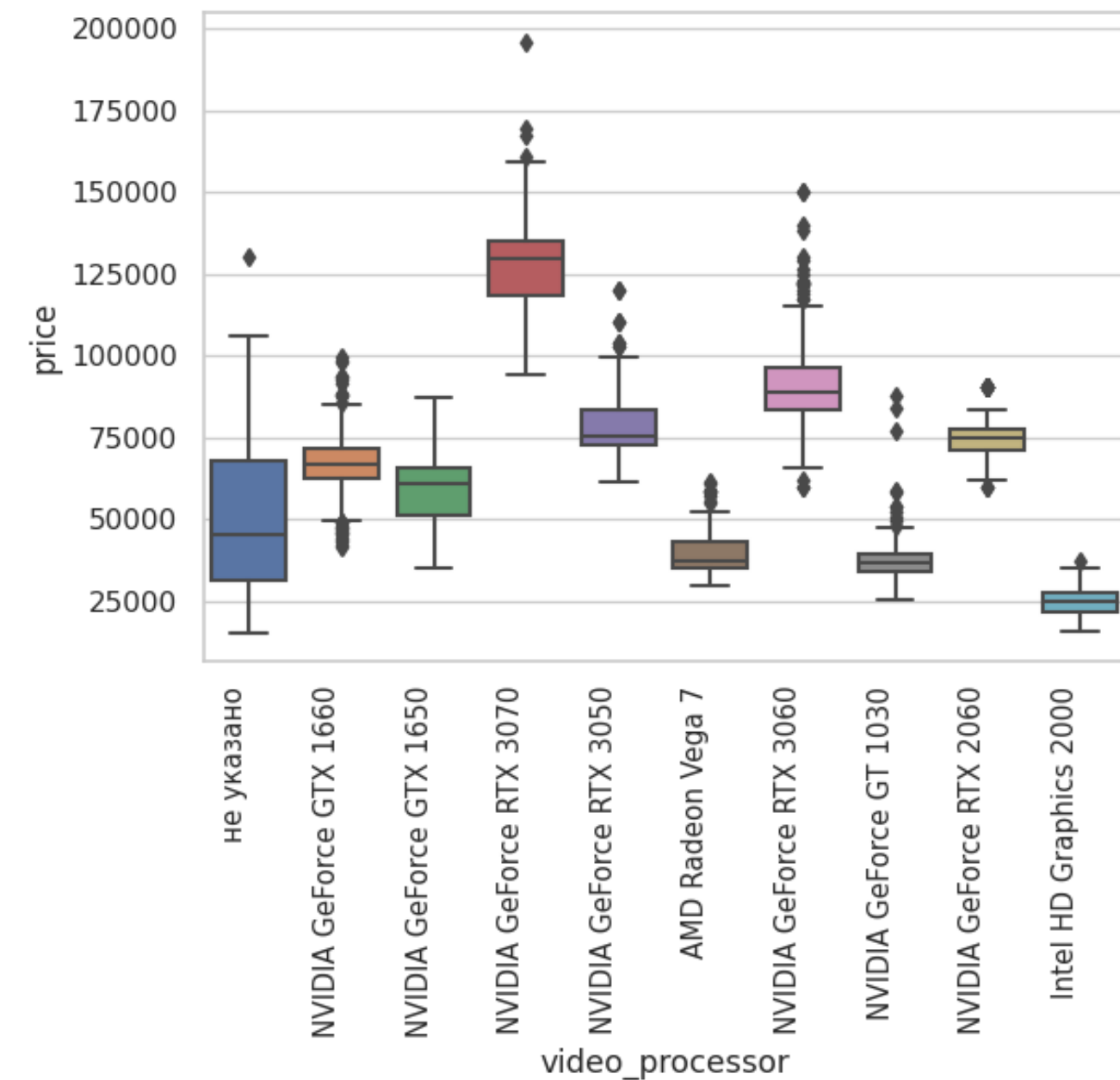
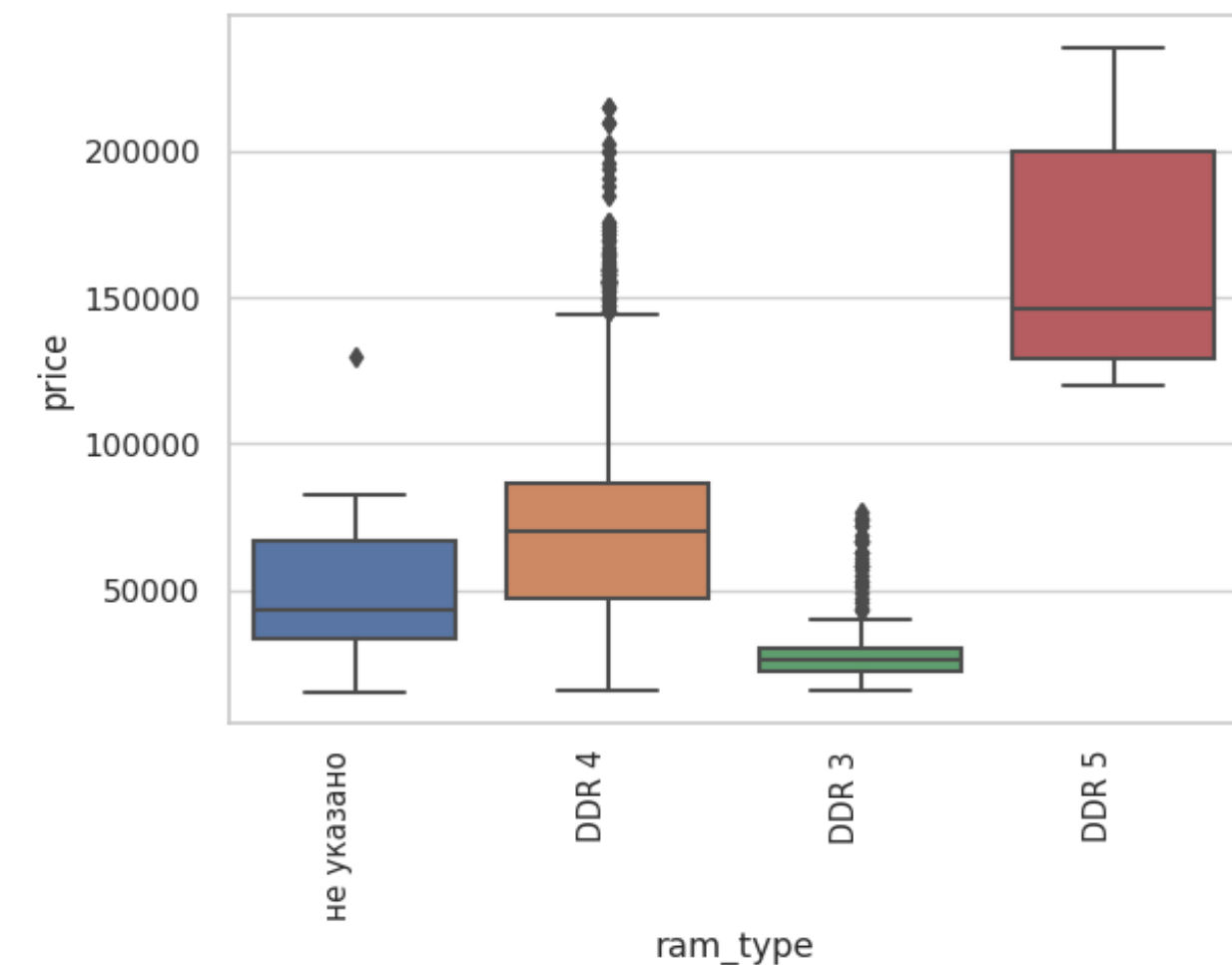
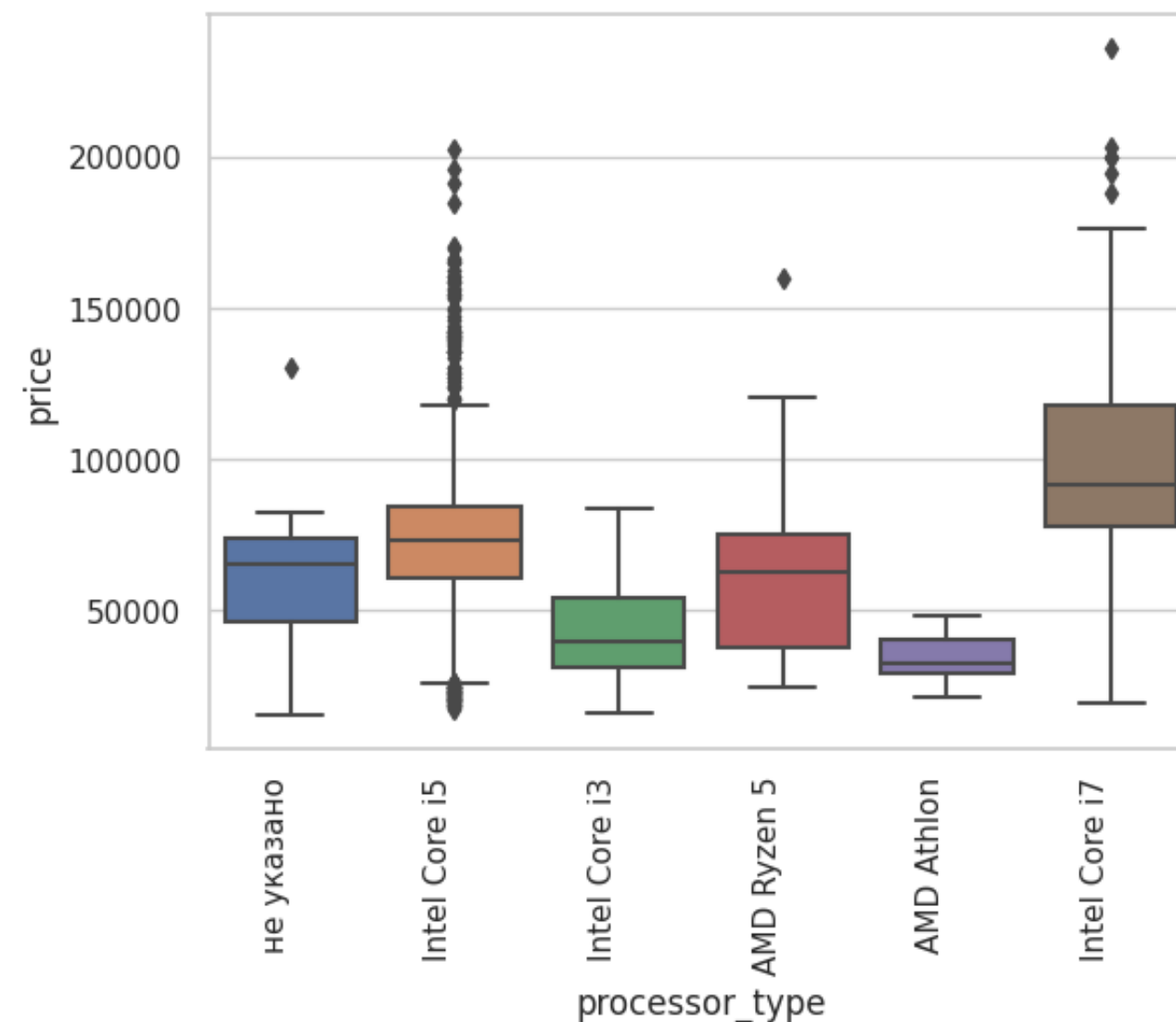
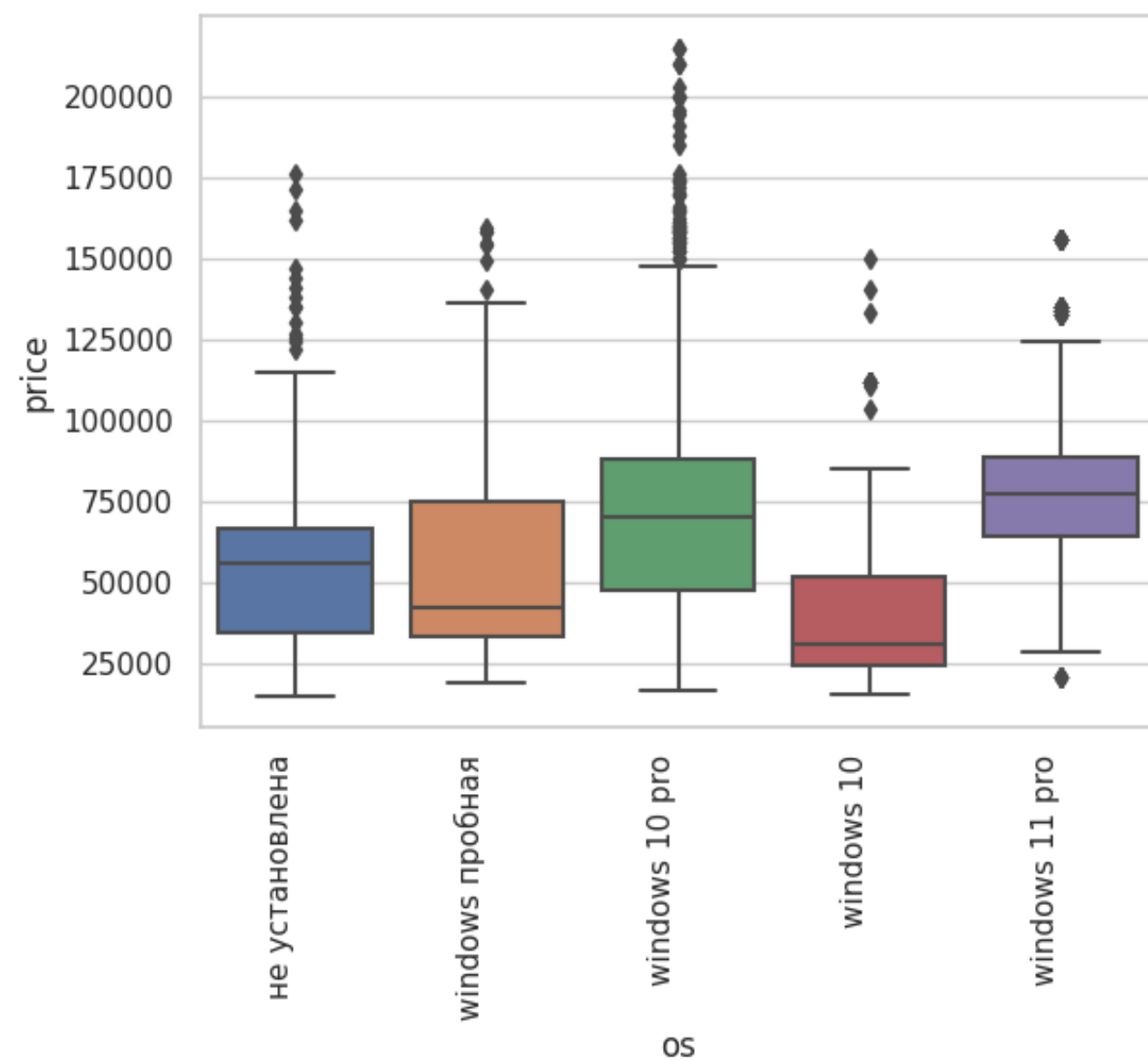
Наибольшее влияние на цену оказывает кол-во ядер.



Результаты анализа:

На боксплотах изображена взаимосвязь цены и категориальных факторов.

По факторам выделены группы ТОП-показателей. Расчет критерия Краскала-Уолиса показывает, что отличия групп статистически значимы. Операционная система, тип процессора, тип оперативной памяти, видеопроцессор влияют на цену ПК.

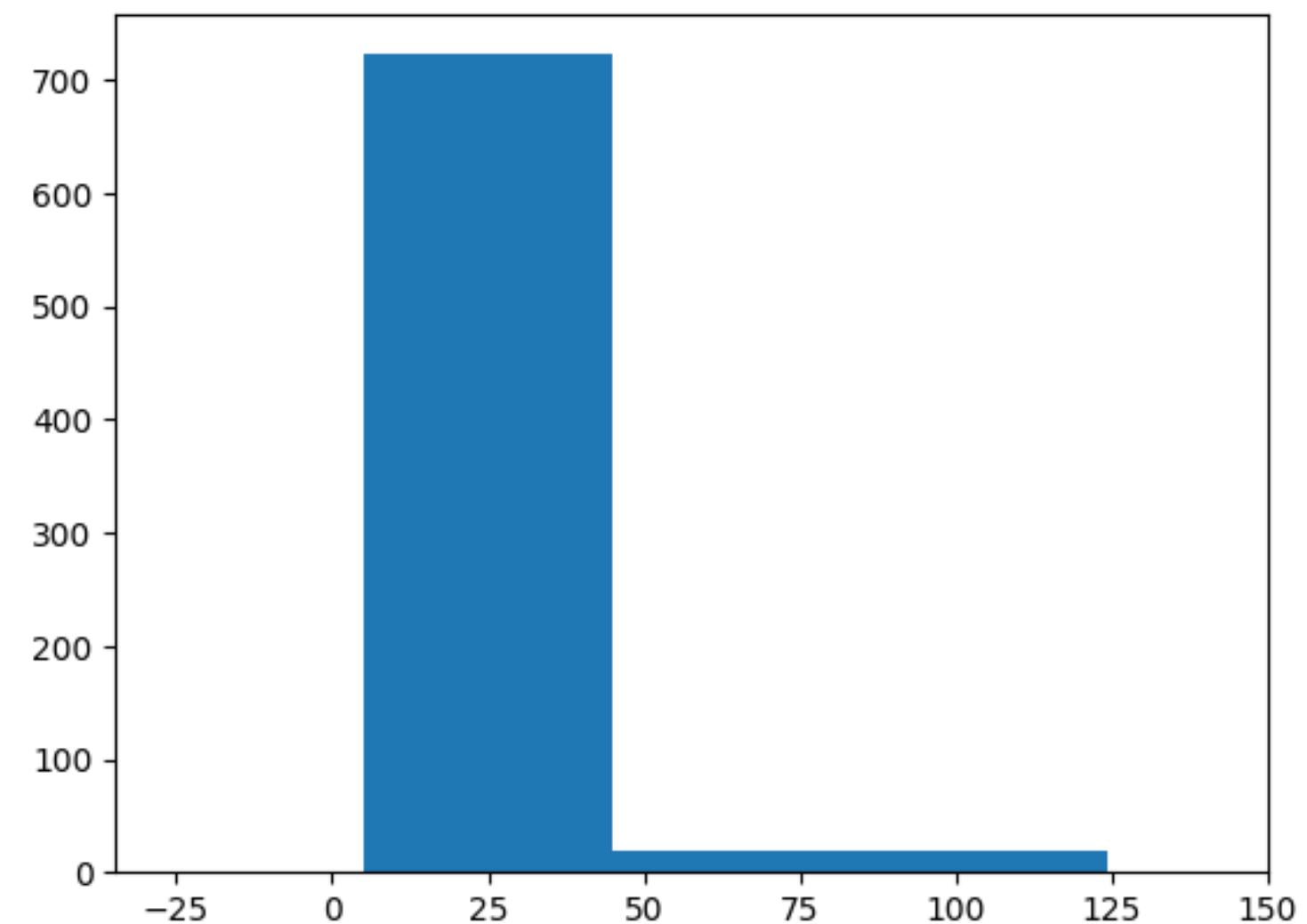
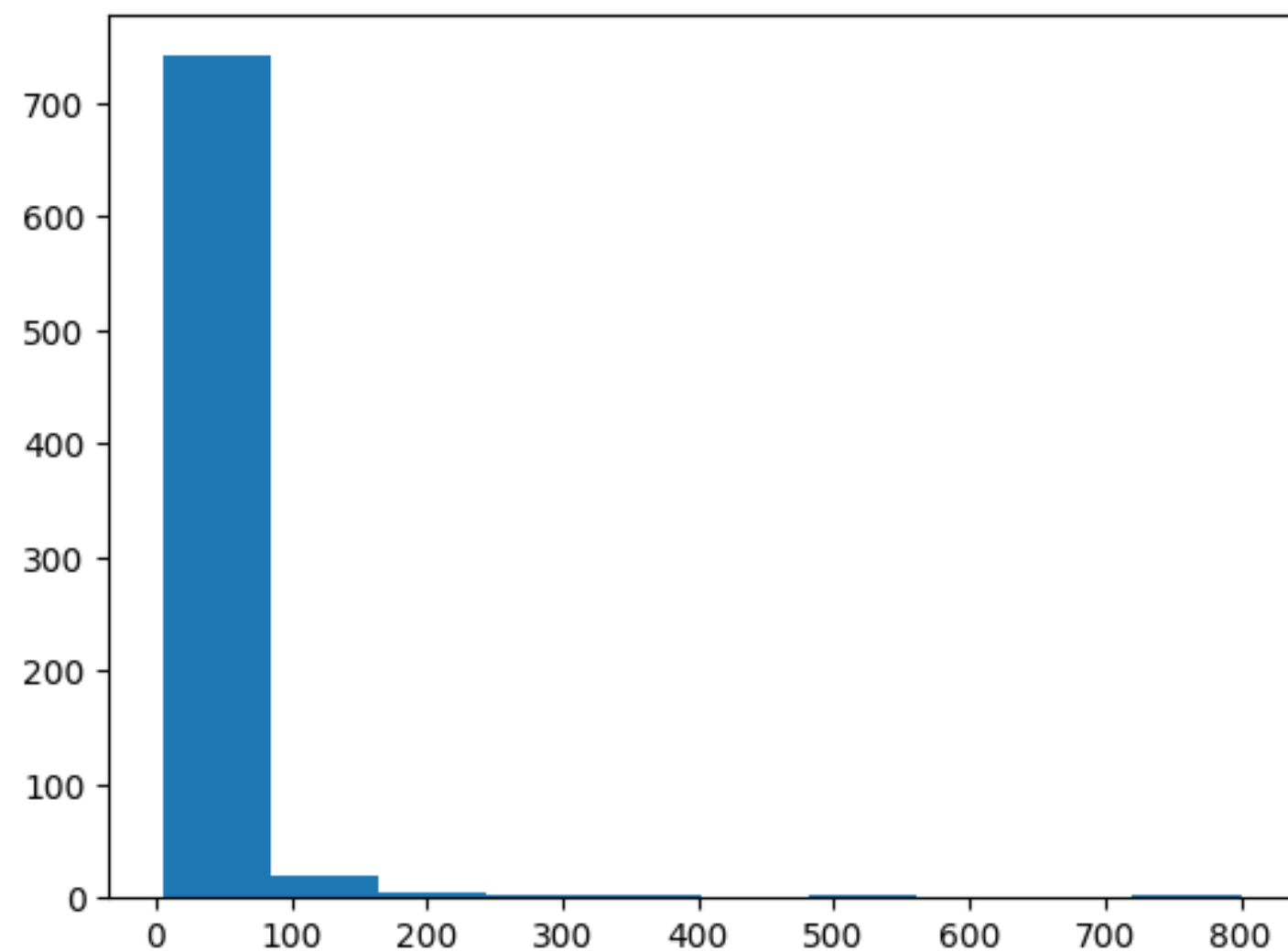




Результаты анализа:

Рассмотрим анализ популярности (кол-ва продаж):

Гистограмма показывает на очень большой размах в данных, что подтверждается и показателями описательной статистики (min=5 и max=800).



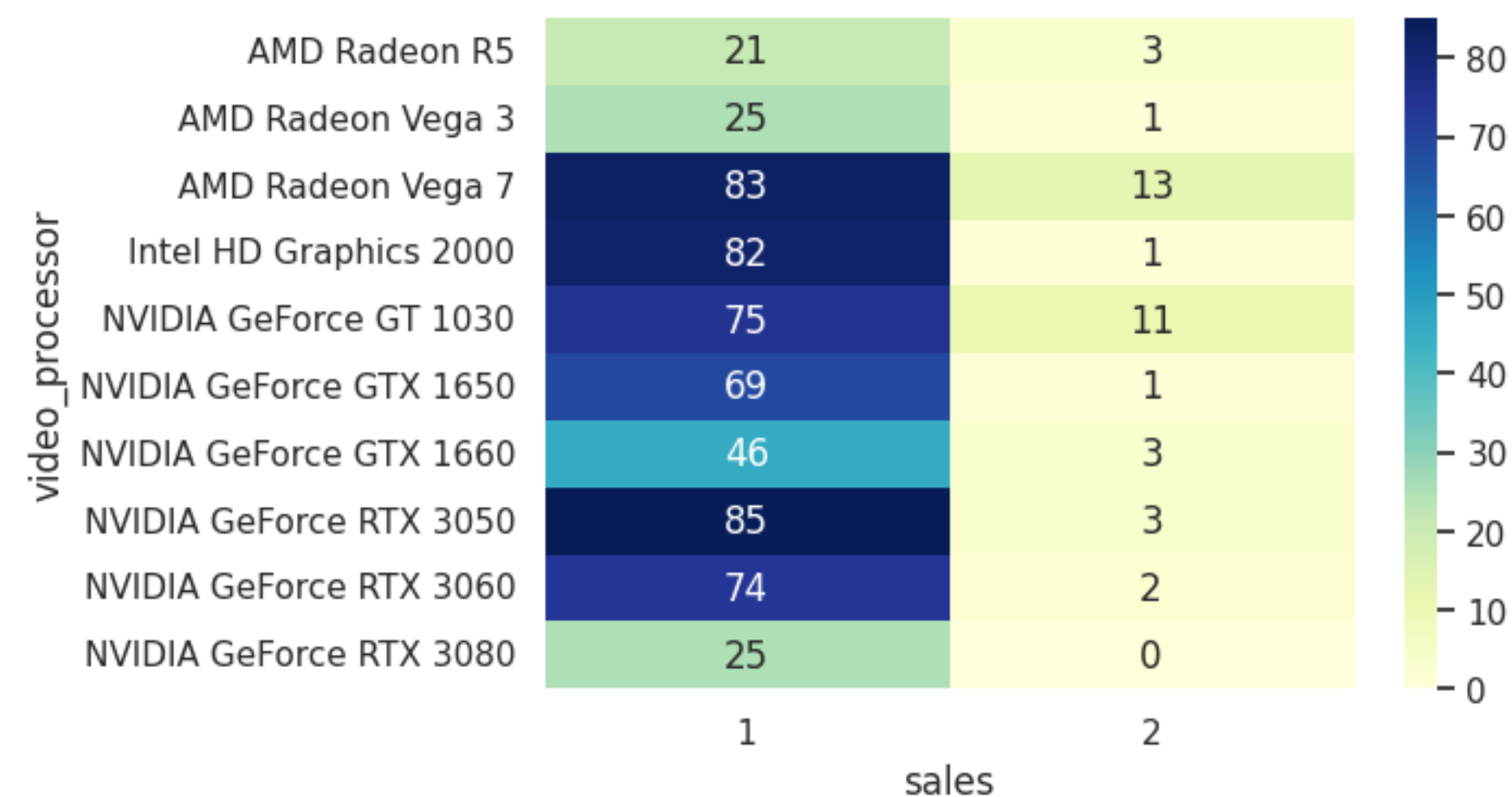
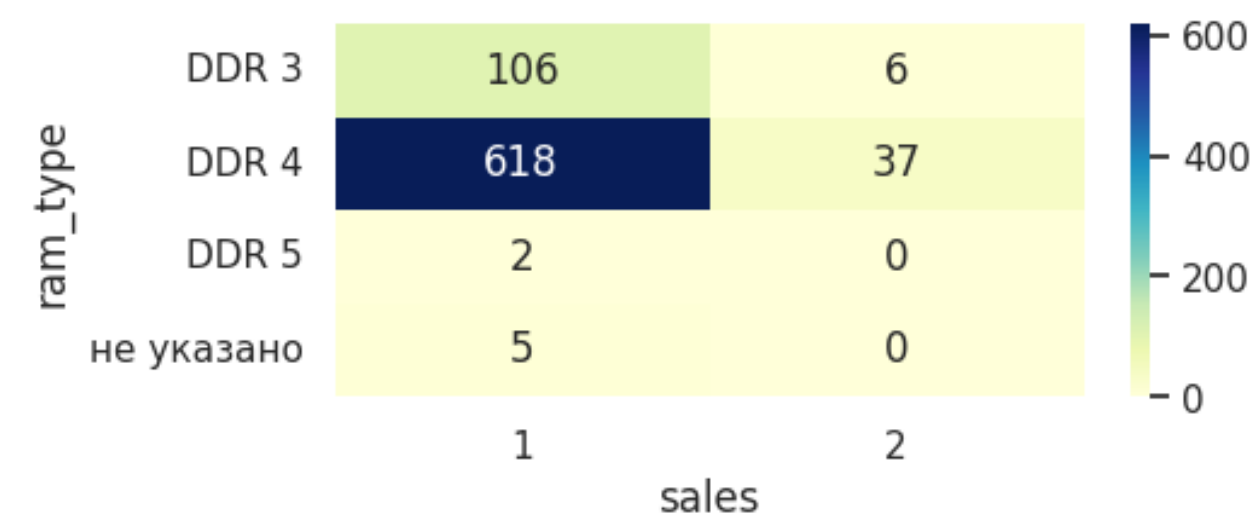
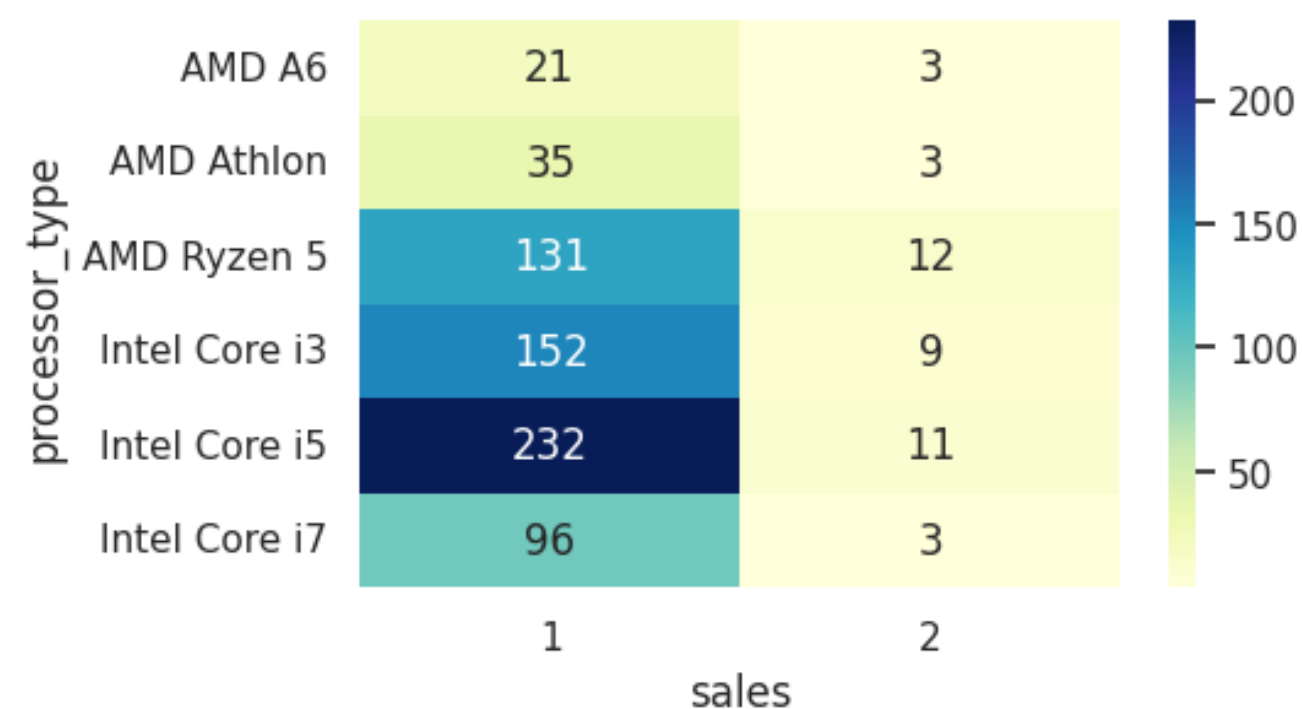
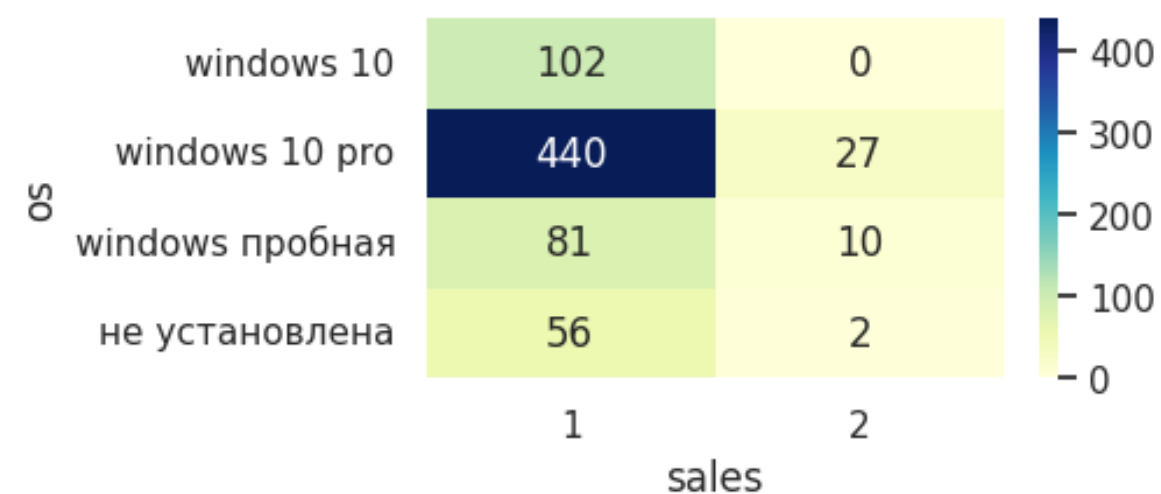
Основная часть выборки находится в диапазоне по кол-ву продаж от 5 до 50 раз.

Но есть и выбросы, которые информативны - это отдельные компьютеры и несколько моделей, которые очень популярны и их купили более 50 раз.



Результаты анализа:

Рассмотрим взаимосвязь популярности как категориального признака с категориальными факторами. По факторам выделены группы ТОП-показателей.

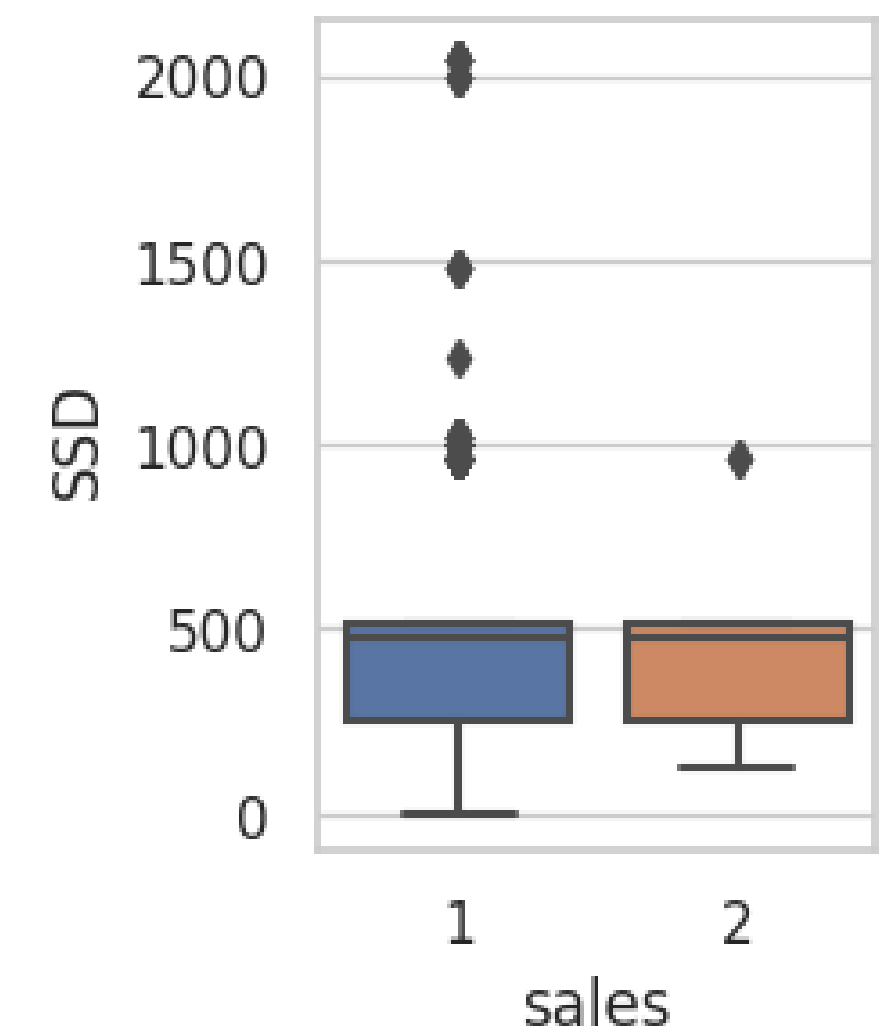
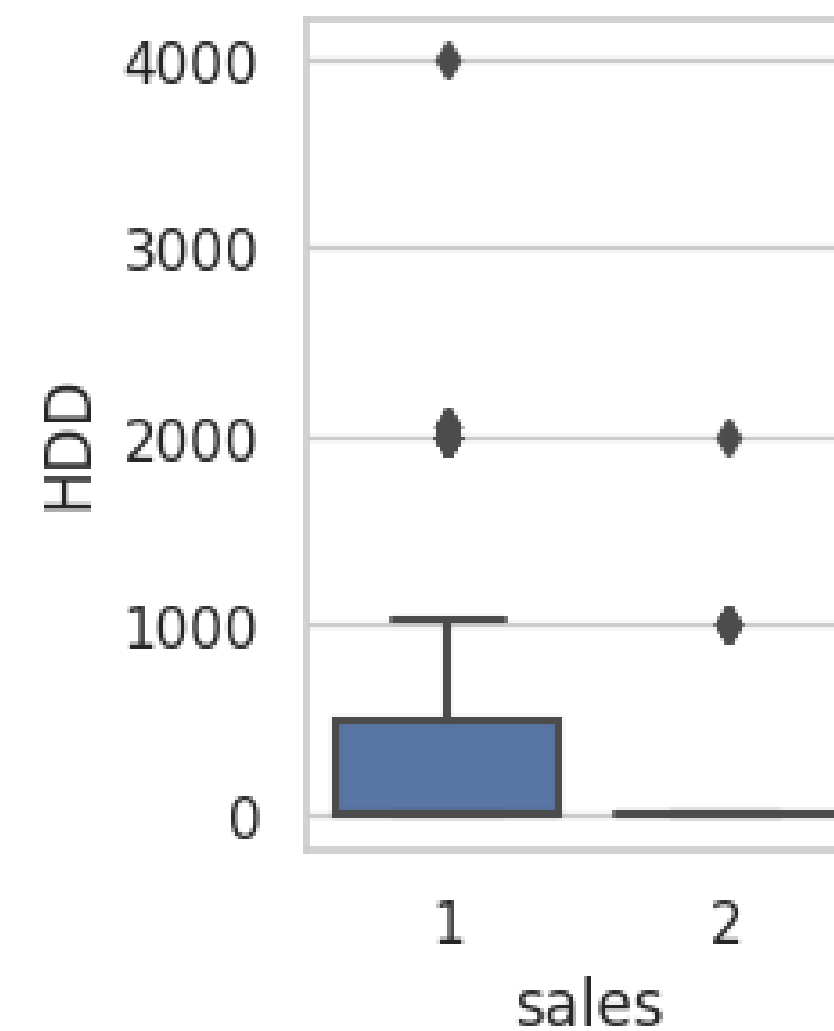
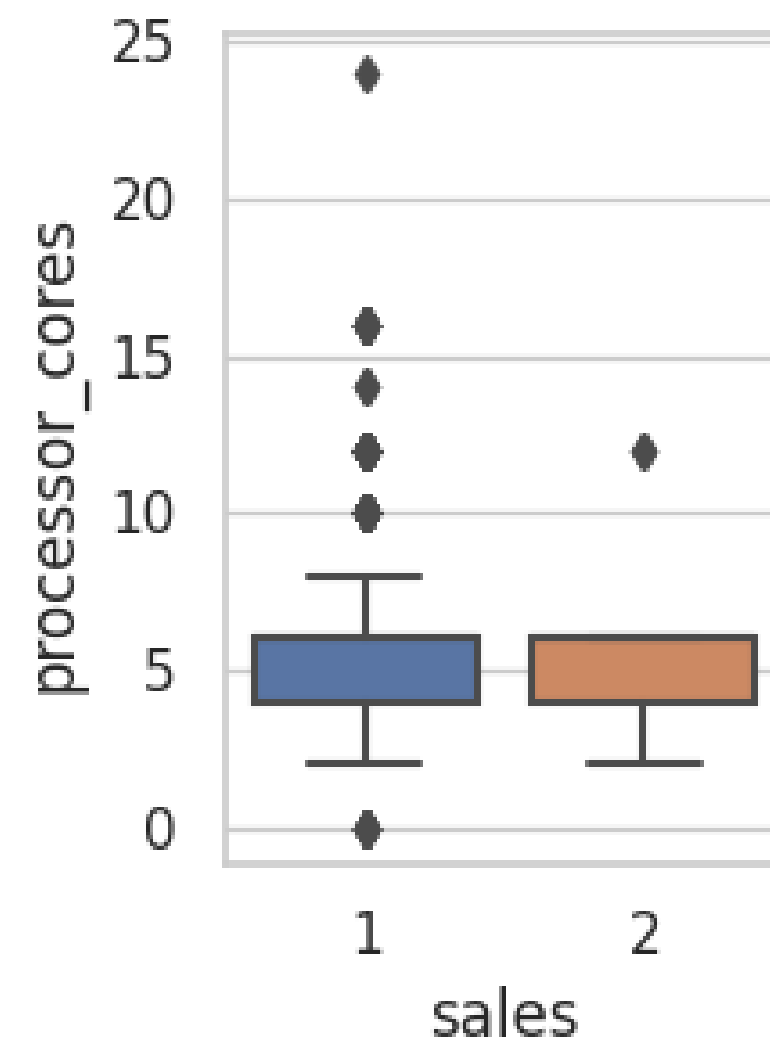
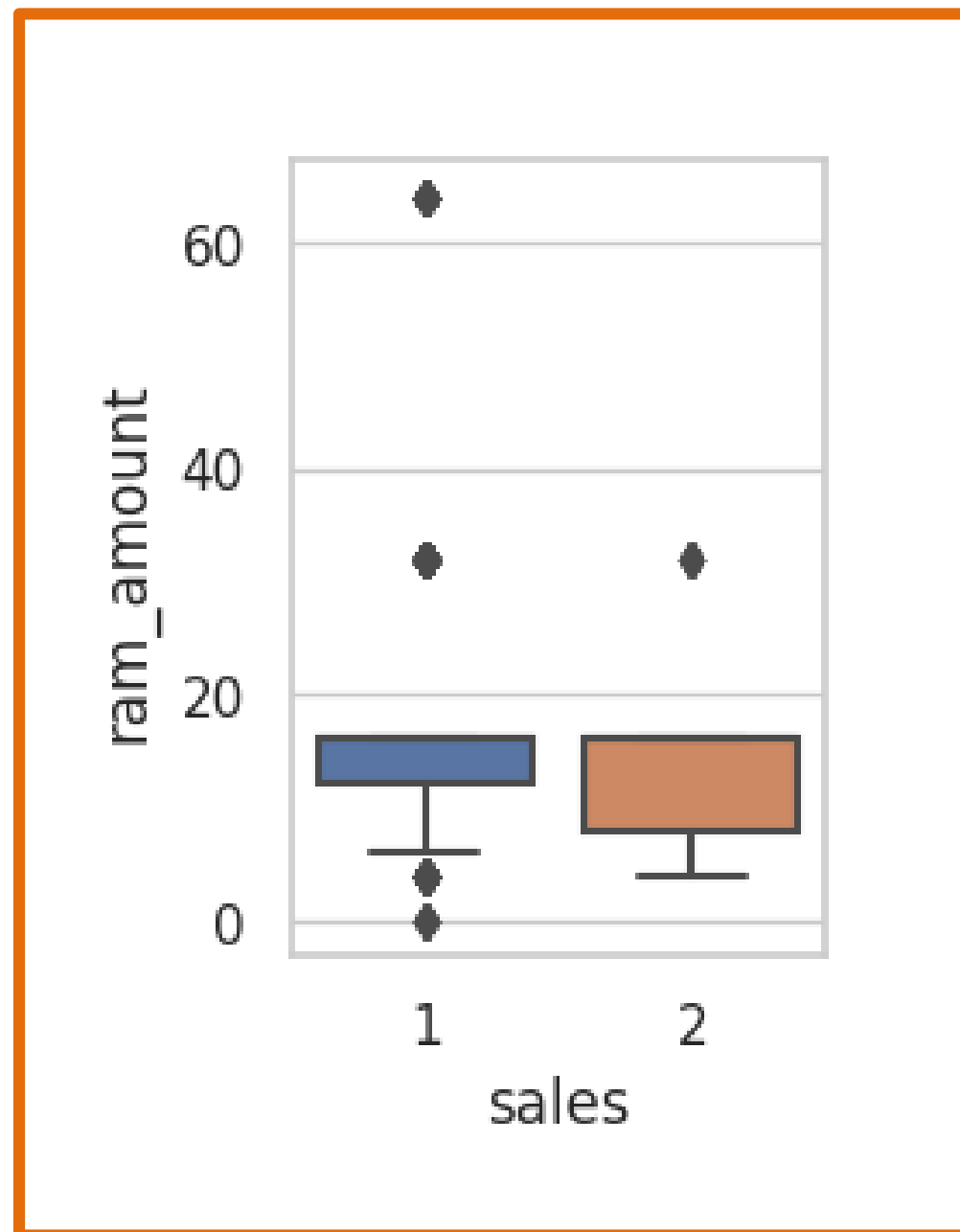


По тепловым картам видно, что частоты в группах сильно различаются. Но по тесту на независимость групп X2 эти частоты можно считать примерно равными, пропорции примерно одинаковыми, поэтому эти факторы не зависимы, связи нет.



Результаты анализа:

Рассмотрим взаимосвязь популярности как категориального признака с количественными факторами. Также видны различия групп. Но согласно критерию Манна-Уитни отличия групп статистически не значимы. На кол-во продаж влияет только объем оперативной памяти, остальные факторы нет.



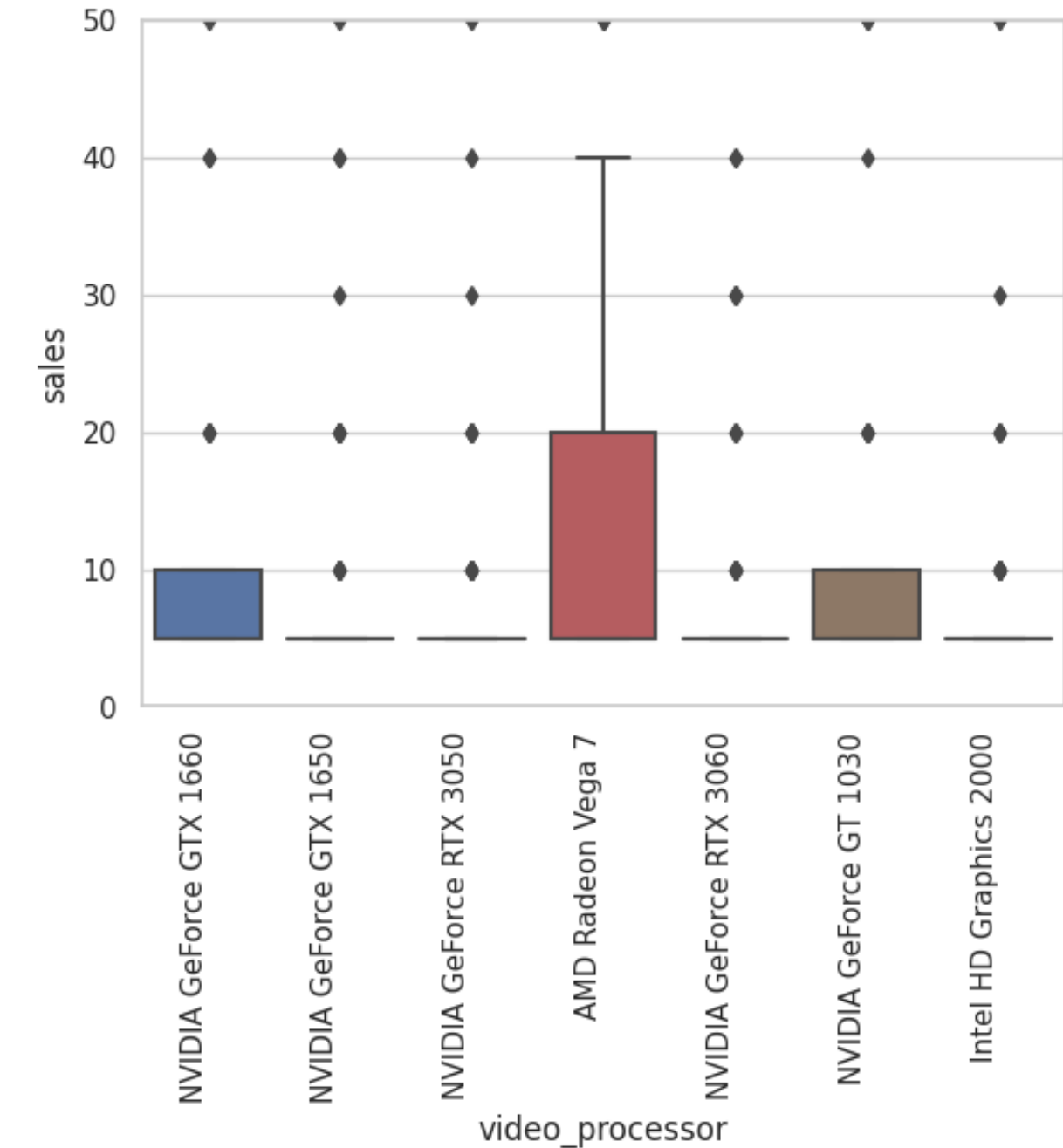
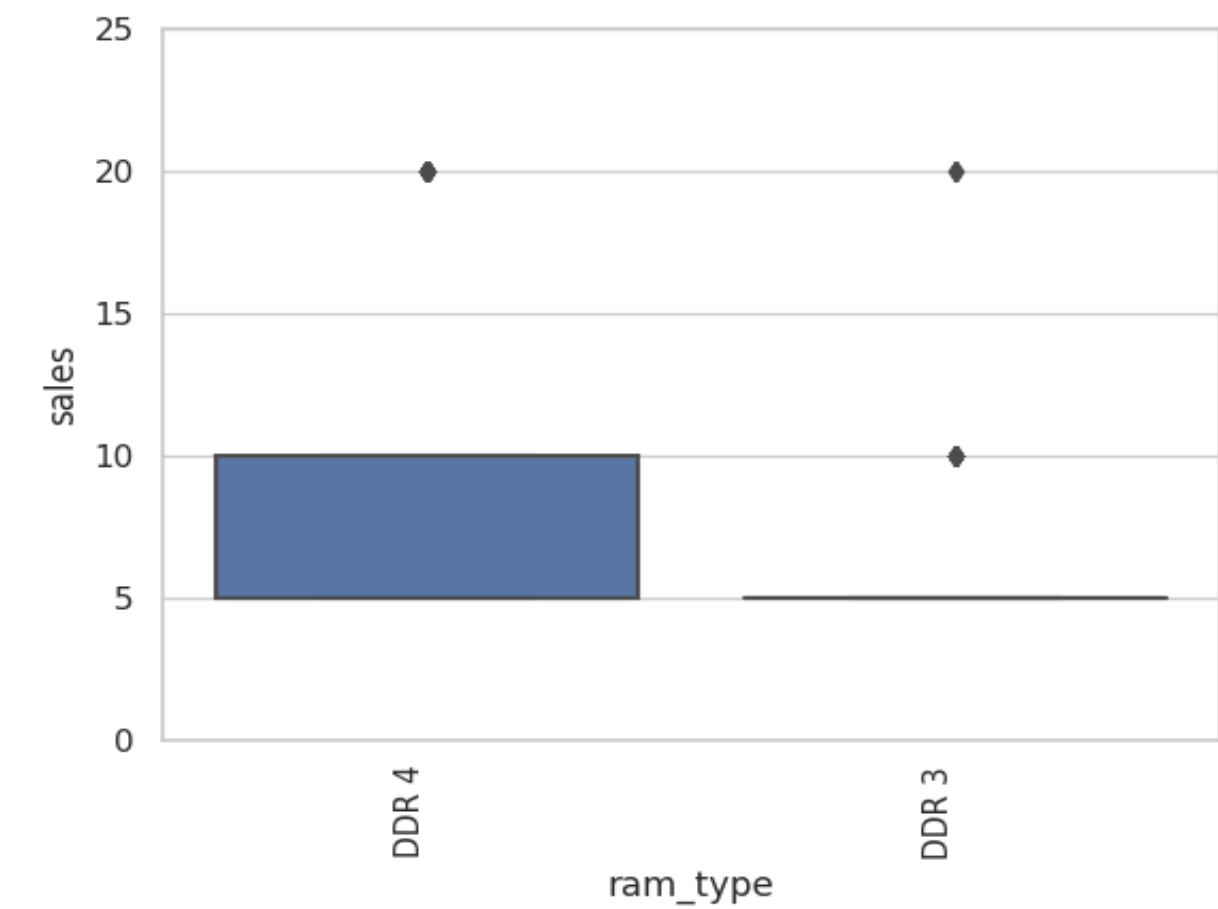
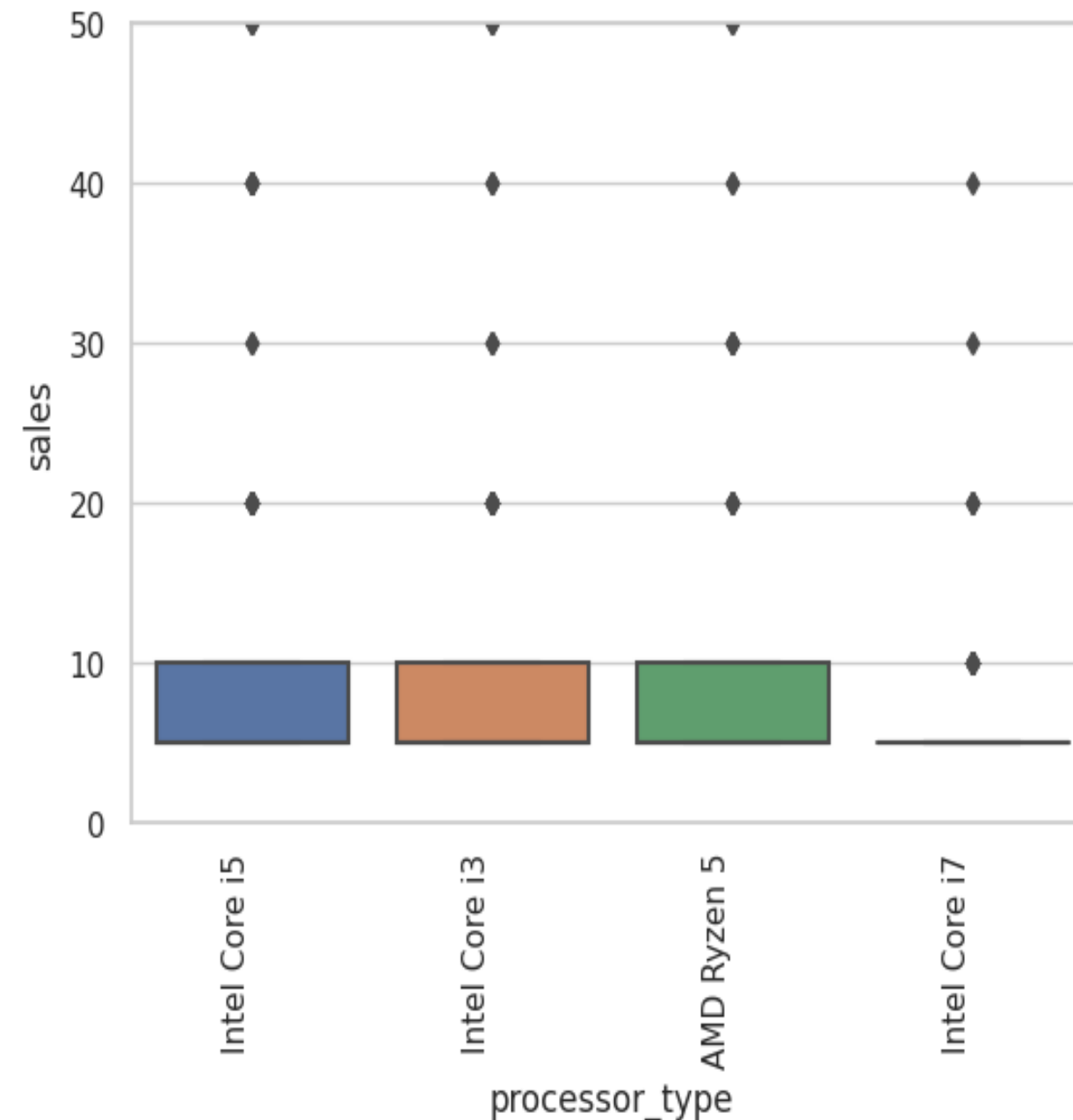
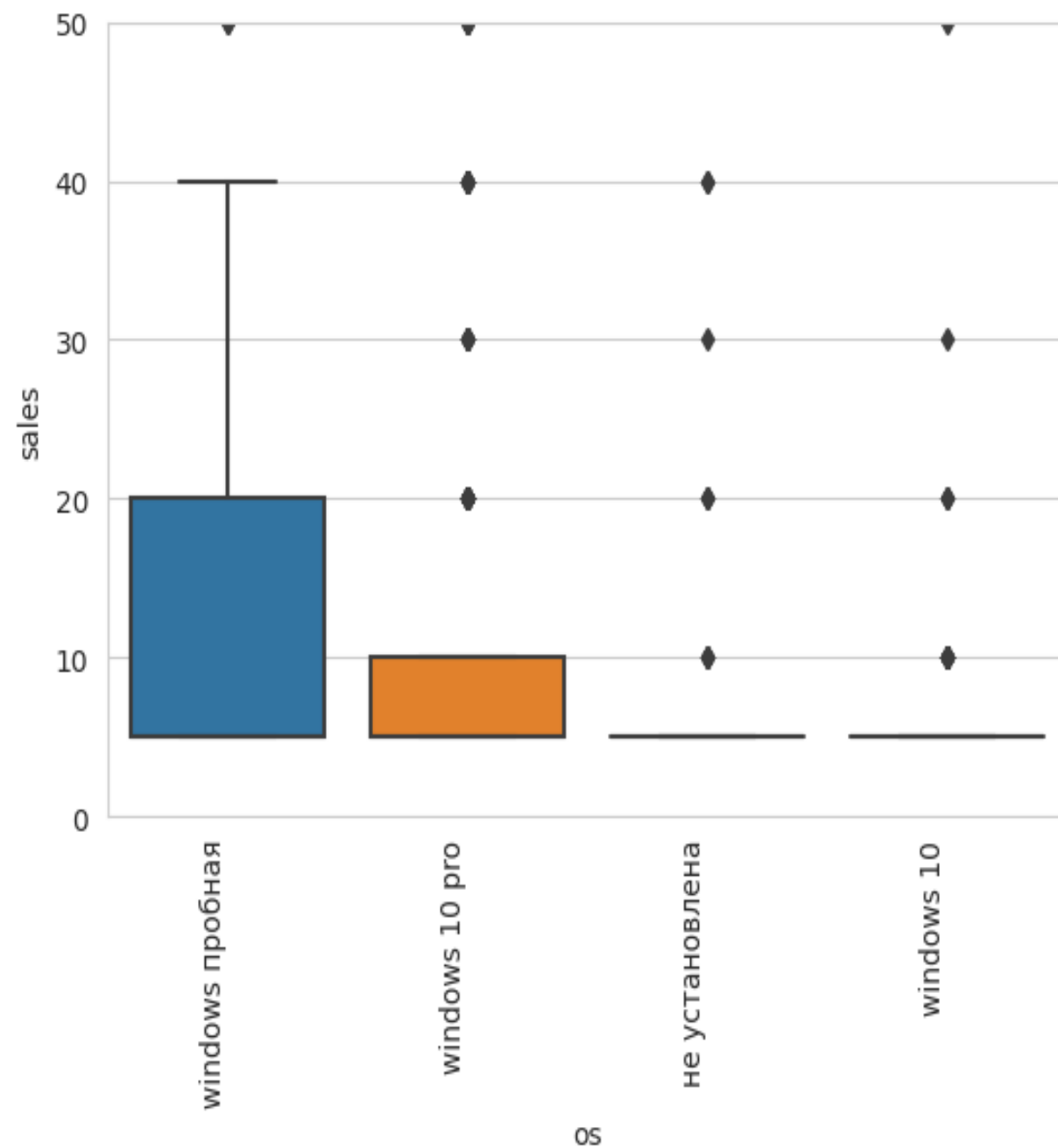


Результаты анализа:

Рассмотрим взаимосвязь популярности как количественного признака с категориальными факторами.

Также по факторам выделены группы ТОП-показателей.

Отличия групп заметны. И согласно критерию Краскала-Уолиса отличия групп статистически значимы. Операционная система, тип процессора, видеопроцессор влияют на кол-во продаж ПК, тип оперативной памяти нет.

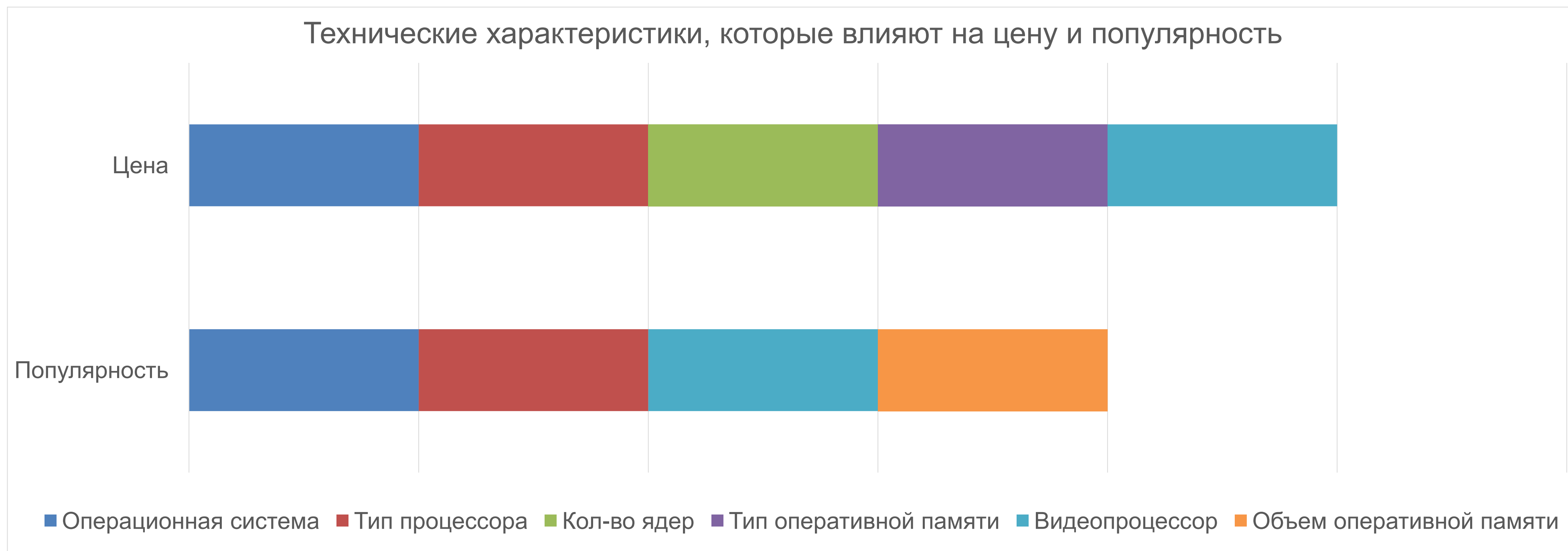




Интерпретация данных:

Резюмируем:

На гистограмме изображены технические характеристики, влияние которых было обнаружено в результате анализа.



4 Интерпретация данных:

По наиболее популярным комплектующим в разрезе кол-ва продаж сформирована таблица:

Характеристики	Наименование	Кол-во продаж
Операционная система	windows 10 pro	467
	windows 10	102
	windows пробная	91
	не установлена	58
Тип процессора	Intel Core i5	243
	Intel Core i3	161
	AMD Ryzen 5	143
	Intel Core i7	99
Видеопроцессор	AMD Radeon Vega 7	96
	NVIDIA GeForce RTX 3050	88
	NVIDIA GeForce GT 1030	86
	Intel HD Graphics 2000	83
	NVIDIA GeForce RTX 3060	76
	NVIDIA GeForce GTX 1650	70
	NVIDIA GeForce GTX 1660	49
Объем оперативной памяти Гб	16	405
	8	162
	32	154



Интерпретация данных:

Для удобства восприятия результатов анализа, сформирован отдельный датасет по выборке без учета нулевых продаж.

	title	seller	os	processor_type	video_processor	sales	price
354	Компьютер Raskat (Intel Pentium G6400/4Gb/120G...	не указано	не установлена	Intel Pentium	Intel HD Graphics 610	800	36990.0
304	Игровой компьютер i7-2600 16Gb SSD 512Gb GT 73...	ALFA-SYSTEM	windows	Intel Core i7	NVIDIA GeForce GT 730	800	32900.0
120	Игровой Компьютер Robotcomp Старт V1	Robotcomp	windows пробная	Intel Core i3	NVIDIA GeForce GT 1030	595	31731.4
82	Игровой Компьютер Robotcomp Аллигатор V3	Robotcomp	windows 10 pro	AMD Ryzen 5	AMD Radeon Vega 7	595	35496.0
80	Игровой Компьютер Robotcomp Аллигатор V2	Robotcomp	windows 10 pro	AMD Ryzen 5	AMD Radeon Vega 7	505	32617.0
...
223	Игровой компьютер Mist X2	Buchok	windows 10 pro	Intel Core i5	NVIDIA GeForce RTX 3050	5	86104.0
222	Игровой компьютер Life Tech K1v2БК	ABS-TECHNO.RU	windows 10 pro	Intel Core i3	NVIDIA GeForce GT 1030	5	35810.0
221	Игровой компьютер Life Tech Game B V2	ABS-TECHNO.RU	windows 10 pro	AMD A6	AMD Radeon R5	5	22180.0
218	Игровой компьютер Life Tech Bastion	ABS-TECHNO.RU	windows 10 pro	Intel Core i3	AMD Radeon RX 580	5	42000.0
518	Системный блок компьютер офисный 4 ядра 16 гиг...	MERKA	windows 11 pro (trial)	Intel Core i3	Intel UHD Graphics 630	5	27943.0

519 rows × 7 columns

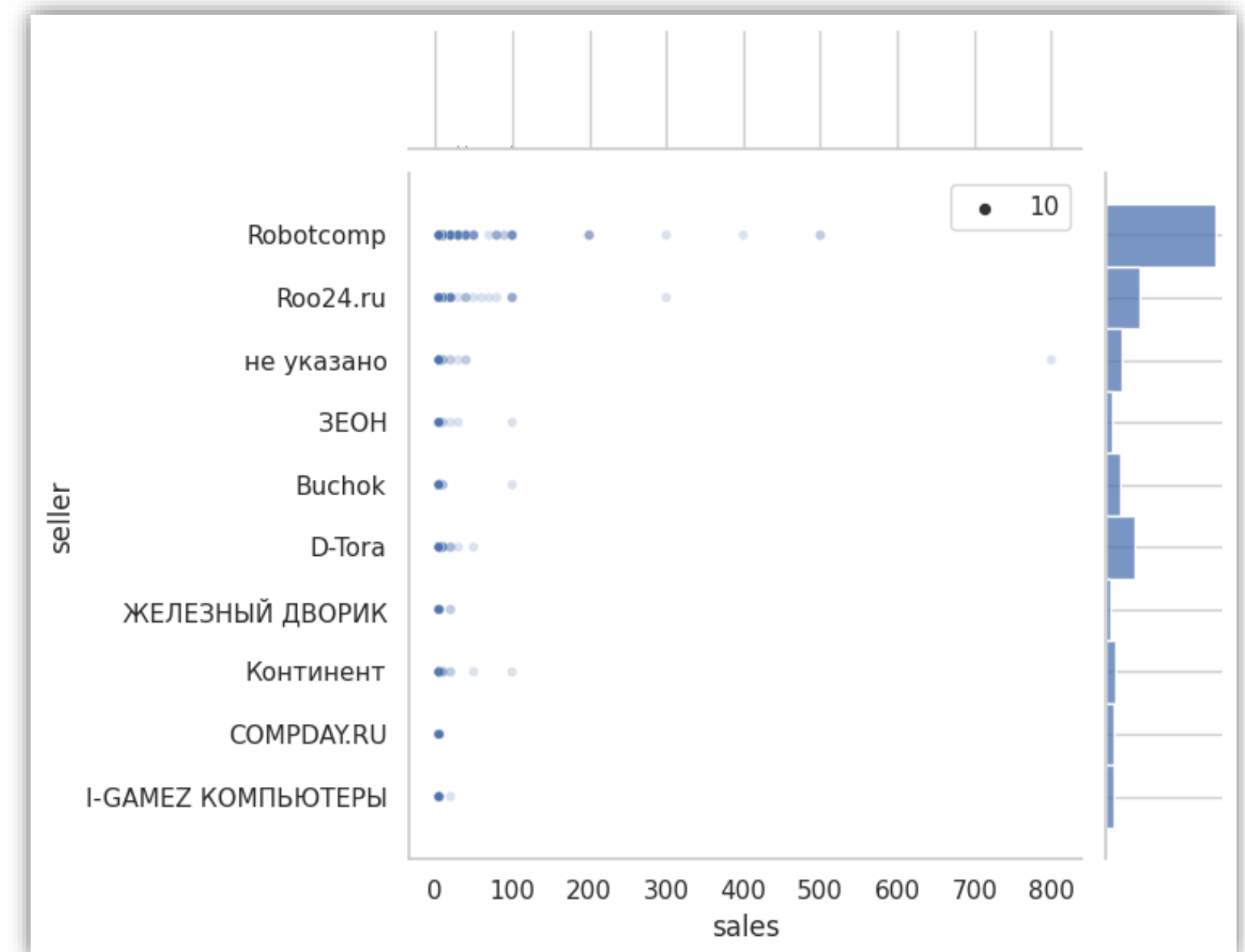
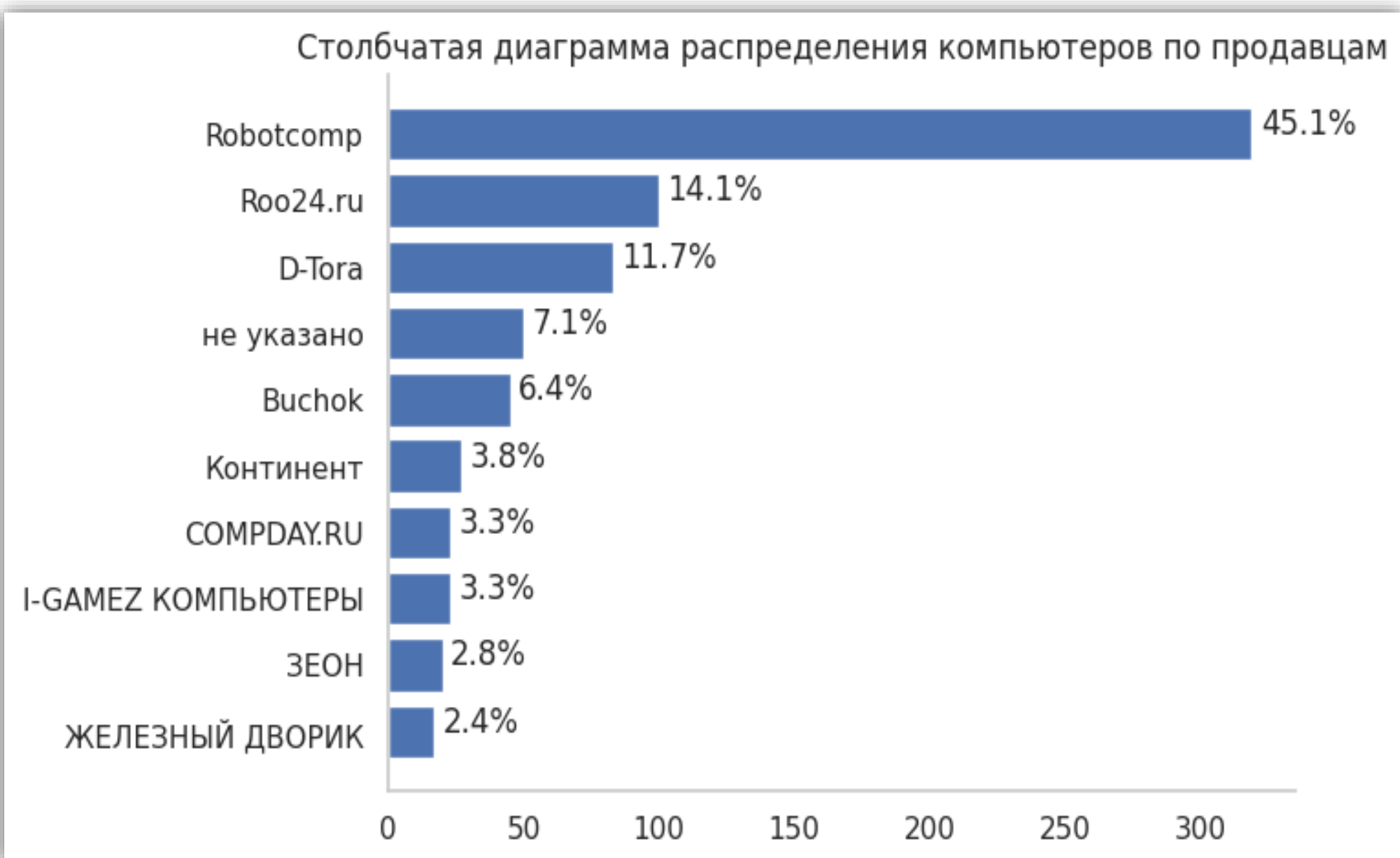
5

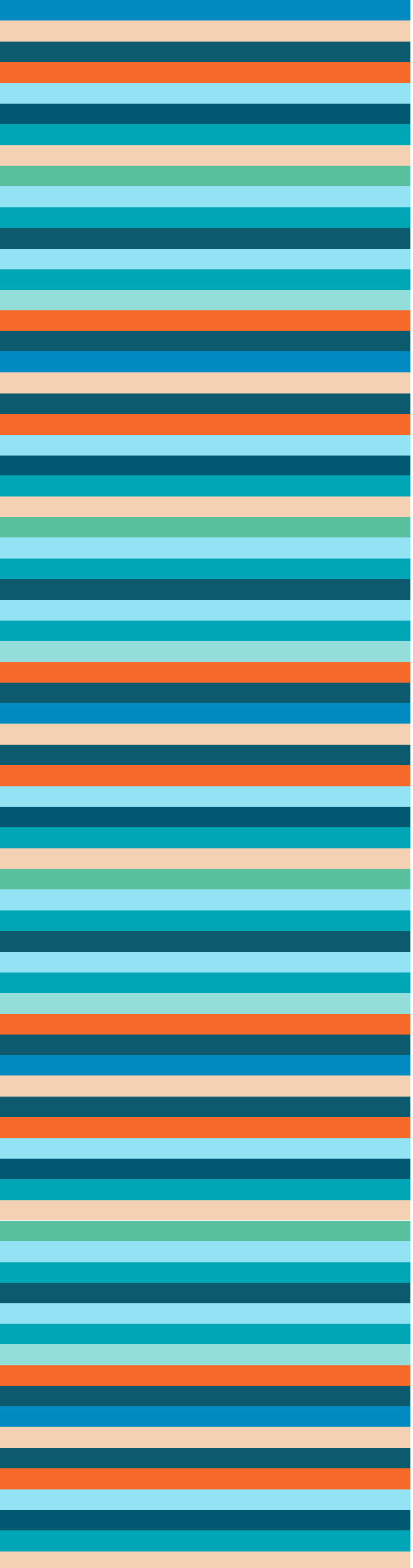
Интерпретация данных:

Рекомендации:

На столбчатой диаграмме и диаграмме распределения представлены продавцы, так как в условиях рынка на формирование цены и популярность могут влиять не только технические характеристики, но и внешние факторы, поэтому даны следующие рекомендации:

- провести анализ конкурентов; изучить их маркетинг, ВАУ-фишки продаж;
- исследовать их продуктовую линейку, торговое предложение: цены, акции, скидки, преимущества, положение на рынке, клиентский сервис, рекламные кампании;
- провести CustDev-исследование целевой аудитории.





Таким образом, анализ проведен - результат проверки гипотезы таков, что технические характеристики действительно влияют на цену и популярность.
Заказчику предоставлен отчет и даны рекомендации.

**Спасибо
за внимание!**

Сердюк Наталья Александровна