

Эконометрическое исследование зависимости стоимости смартфонов различных брендов и ценовых сегментов от их технических характеристик

Исследование подготовили:

Денисенко Дмитрий Сергеевич

Шестяев Егор Алексеевич

Мусаковский Иван Алексеевич

Введение

Цель исследования: *Определить и количественно оценить влияние технических характеристик смартфонов различных брендов и ценовых сегментов на их рыночную стоимость с помощью эконометрических методов.*

Иными словами, цель состоит в том, чтобы выявить, какие параметры (мощность процессора, объем памяти, количество камер, разрешение экрана, поддержка сетей нового поколения и т.д.) оказывают наибольшее и статистически значимое влияние на цену смартфона, а также понять различия в данном влиянии для разных брендов и классов устройств (бюджетный, средний, премиальный сегменты).

Задачи исследования:

- 1. Провести обзор литературы и различных теоретических подходов** – в первую очередь необходимо изучить существующие исследования по ценообразованию на рынке смартфонов, рассмотреть экономические и маркетинговые модели, описывающие зависимость стоимости от технических характеристик и бренда;
- 2. Определить перечень ключевых технических характеристик, которые потенциально могут оказывать влияние на цену устройства, чтобы в последующем включить их в эконометрическую модель;**
- 3. Провести EDA (разведочный анализ данных)** - рассчитать и проанализировать различные описательные статистики, подготовить визуализацию данных, провести проверку на наличие выбросов и пропусков, избавиться от них;
- 4. Подготовить данные для построения эконометрической модели** - провести категоризацию переменных (*dumtту*), кодирование, возможно, сделать нормализацию (если окажется, что имеет место быть чувствительность модели к масштабу данных);
- 5. Разработать регрессионную модель** - протестировать различные варианты (на основе значений выбранных метрик качества), начиная с простой множественной линейной регрессии и заканчивая более продвинутыми эконометрическими методами для оценки влияния технических характеристик на цену смартфонов;
- 6. Определить содержательные гипотезы, которые могут быть протестированы в ходе исследования;**
- 7. Оценить качество модели и проинтерпретировать полученные результаты** - проверить значимость коэффициентов модели, оценить качество подгонки модели, протестировать модель на адекватность путем постановки диагностических тестов, провести проверку устойчивости результатов. **Определить лучшую модель.**
- 8. Проинтерпретировать полученные коэффициенты, чтобы понять, какие технические характеристики вносят наибольший вклад в формирование цены;**

9. Протестировать сформулированные гипотезы и проанализировать полученные результаты;
10. Сформулировать выводы о том, какие технические параметры оказывают наибольшее влияние на цену устройства.

Источники:

<https://www.counterpointresearch.com/> (раздел “Trending Research”)

<https://ir.uitm.edu.my/id/eprint/98695/1/98695.pdf> (ключевой ресурс для построения моделей)

<https://marketingandmetrics.com/wp-content/uploads/2020/06/12.-Free-to-free.pdf>

<https://www.sciencedirect.com/science/article/abs/pii/S109499681500033X>

Актуальность:

Рынок смартфонов динамичен и крайне конкурентен. Появление новых моделей устройств, инновационных технологий, расширение функционала и рост числа производителей с каждым годом приводят к все большему усложнению структуры ценообразования.

Понимание того, какие технические параметры действительно оказывают влияние на цену и какова роль бренда в формировании конечной стоимости продукта имеют важное значение для всех экономических агентов:

- **Для потребителей:** понимание логики ценообразования на рынке смартфонов поможет пользователям в принятии более осознанных решений при покупке и позволит подобрать для себя наиболее оптимальное устройство с точки зрения соотношения цены и качества;
- **Для производителей и ритейлеров:** исследование зависимости стоимости смартфонов от их характеристик позволит своевременно и точно вносить корректировки в ценовую стратегию, приведет к более точному пониманию, что сейчас нужно рынку, и как позиционировать свой бренд самым наилучшим образом;
- **Для инвесторов и аналитиков:** понимание того, как устроен процесс ценообразования на данном рынке позволит точнее оценивать потенциальную прибыль и риски, связанные с инвестициями в определенный бренд или технологический / ценовой сегмент.

Экономическая модель

Описание использованных данных:

- **Источник данных:**
<https://www.kaggle.com/datasets/agarwalyashhh/smartphones-complete-data>
- **Кем собирались:** Как мы знаем, платформа Kaggle дает возможность всем своим пользователям публиковать в открытый доступ собранные ими датасеты. Наши данные о смартфонах были собраны студентом (<https://www.linkedin.com/in/agarwalyashh2004/>) NIT Kurukshetra (Индия) в рамках его подготовки к написанию дипломной работы по дата-анализу и машинному обучению.
- **Когда собирались:** Данные были собраны около 8-9 месяцев назад, непосредственно перед публикацией датасета на Kaggle.
- **Как собирались:** Датасет собирался путем парсинга данных со специализированного индийского сайта, предназначенного для сравнения стоимости смартфонов разных моделей и ценовых категорий - <https://www.smartprix.com/mobiles>
- **Количество наблюдений в исходном датасете:** 980
- **Количество наблюдений в датасете, очищенном от пропущенных значений:** 924

Объясняющие переменные:

Переменная	Предположительное влияние на зависимую переменную	Обоснование выбора
<i>num_cores</i> (кол-во ядер процессоров)	+	Более мощные процессоры с большим количеством ядер, как правило, обеспечивают лучшую производительность устройства. Это является важным фактором выбора для пользователя и влияет на стоимость устройства
<i>processor_speed</i> (частота процессора, ГГц)	+	Частота процессора напрямую связана с его скоростью выполнения задач, что делает эту переменную важным фактором, влияющим на стоимость и на пользовательский опыт
<i>ram_capacity</i> (объем ОЗУ, ГБ)	+	Большее количество оперативной памяти обеспечивает более плавную работу смартфона и возможность выполнять более сложные задачи, что также влияет на его стоимость и пользовательский опыт
<i>battery_capacity</i> (объем батареи, mAh)	Вероятно, сильное влияние может отсутствовать	Хотя объем батареи и не всегда может напрямую влиять на цену, этот фактор, очевидно, крайне важен для всех потребителей
<i>internal_memory</i> (объем встроенной памяти, Гб)	+	Большой объем встроенной памяти - значительное преимущество для всех групп пользователей. И это, очень вероятно, оказывает влияние на цену
<i>screen_size</i> (диагональ экрана, дюймы)	Вероятно, какое-либо влияние может отсутствовать	Размер экрана важен для пользователей, но отношение к этому параметру может варьироваться в зависимости от предпочтений. Интересно исследовать, как это влияет на цену
<i>refresh_rate</i> (частота обновления экрана, Гц)	+	Высокая частота обновления экрана обеспечивает плавное отображение контента, что становится всё более важным фактором для пользователей и влияет на стоимость устройства. Плюс ко всему, это достаточно

		новая технология, которая, однако, с каждым годом становится все дешевле и появляется во все более бюджетных устройствах
<i>num_rear_cameras</i> (кол-во задних камер)	Вероятно, какое-либо влияние может отсутствовать	Большое количество камер не связано с увеличением качества фотографий, но часто является важным элементом маркетинговой стратегии многих производителей, тоже интересно поисследовать
<i>num_front_cameras</i> (кол-во фронтальных камер)	Вероятно, какое-либо влияние может отсутствовать	Здесь все аналогично кол-ву задних камер выше
<i>has_5g</i> (<i>dummy</i> , поддержка 5G)	Вероятно, сильное влияние может отсутствовать	Поддержка стандарта 5G с одной стороны является современной и востребованной технологией, а с другой - ее внедрение не сопряжено с высокими издержками. Интересно оценить влияние данного параметра на цену
<i>has_nfc</i> (<i>dummy</i> , наличие NFC)	Вероятно, сильное влияние может отсутствовать	Аналогично 5G, NFC - крайне востребованная в современном мире технология, но в то же время недорогая
<i>extended_memory_available</i> (<i>dummy</i> , наличие возможности расширения памяти с помощью SD-карты)	Вероятно, какое-либо влияние может отсутствовать	Наличие слота для SD-карты может быть полезным для пользователей, но не всегда существенно влияет на цену, особенно в премиальном сегменте, где производители часто исключают эту функцию (например, в айфонах никогда не было возможности расширить встроенную память, но это устройство все же относится к классу премиальных)

*целевая переменная **price** (цена) изначально была представлена в индийских рупиях, далее переведена в рубли по курсу на 12.12.24

Формулировка содержательных гипотез:

1. Гипотеза о брендовой наценке - “Даже при прочих равных технических параметрах смартфоны более популярных брендов будут иметь цену существенно выше, чем аналогичные по характеристикам устройства менее известных производителей. То есть коэффициент при переменных, отражающих “бренд”, должен быть положительным и значимым.”:

- **Обоснование:** Согласно теории брендинга и концепции бренд-капитала (Aaker, Keller), известные бренды могут взимать премию за товар независимо от объективных характеристик. Это связано с ассоциацией бренда с качеством, статусом, послепродажной поддержкой, лояльностью покупателей и т.д.
- **Источники:**
https://id.hse.ru/data/2017/11/13/1158257047/Antonova_Patosha_site.pdf
<https://slonrekomenduet.com/note/ot-chego-zavisit-tsena-smartfona.html>
<https://criterionglobal.com/faq/brand-equity-theory-models/>

2. Гипотеза о роли объема внутренней памяти и эффекте убывающей предельной полезности - “Коэффициент при переменной объема внутренней памяти будет положительным, но при включении нелинейной формы появится дополнительный отрицательный коэффициент, отражающий убывающую предельную полезность. Это будет означать, что, например, увеличение памяти с 32 Гб до 64 Гб даст больший прирост цены, чем увеличение со 128 Гб до 256 Гб.”:

- **Обоснование:** Согласно теории предельной полезности и законам убывающей отдачи каждая дополнительная единица ресурса (в данном случае памяти, Гб) приносит все меньшую дополнительную полезность для потребителя. При низких стартовых значениях уровня памяти ее увеличение существенно повышает ценность (и, как следствие, стоимость) устройства, но при достижении определённого порога дальнейшее увеличение памяти приводит к все меньшему приросту стоимости (“полезности”) смартфона.
- **Источники:**
https://www.econstor.eu/bitstream/10419/23616/1/paper_29.pdf

https://www.researchgate.net/publication/269365608_Hedonic_Price_Analysis_on_Mobile_Phone_Market

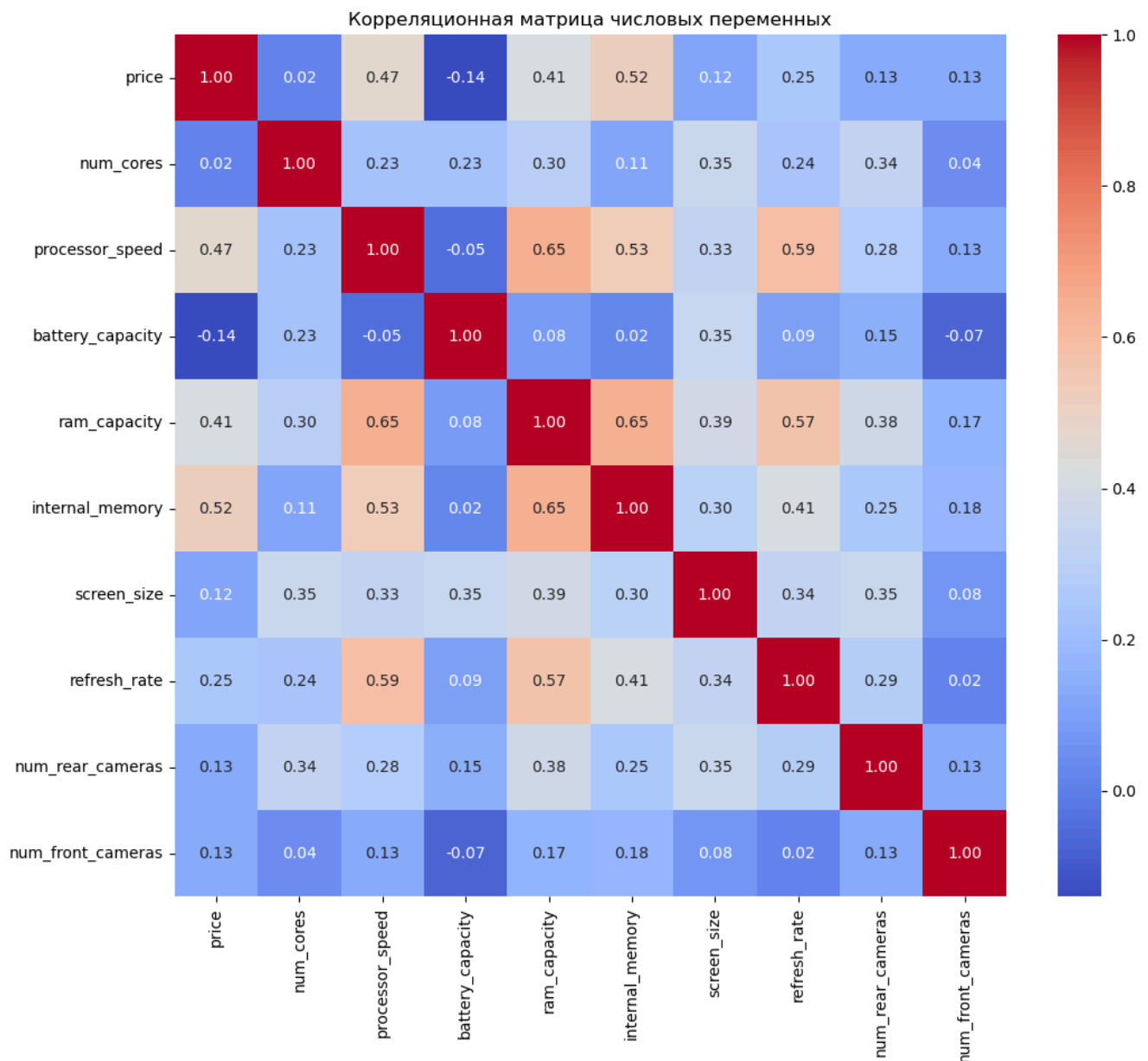
3. Гипотеза об отсутствии различий между группами брендов в рамках одной страны - “Мы можем выбрать подмножество брендов (например, рассмотреть популярных китайских производителей) и проверить гипотезу о том, что их премии/дисконты относительно базового бренда не отличаются друг от друга.”:

- **Обоснование:** Проверка данной гипотезы позволит нам выявить возможные структурные различия в ценовой политике производителей. Например, если китайские бренды придерживаются единой ценовой стратегии, основанной на конкурентоспособных характеристиках, это подтвердит их ориентацию на массовый рынок и стратегию проникновения. В противном случае различия укажут на дифференцированные подходы к ценообразованию. Реализовать это можно путем проведения совместных F-тестов.

- **Источники:**
https://blog.rchss.sinica.edu.tw/FCLai/wp-content/uploads/2016/11/20060209_Shaked-and-Sutton-1982_Relaxing-Price-Competition-Through-Product-Differentiation_The-Review-of-Economic-Studies-491-3-13-Lai.pdf
<https://www.jstor.org/stable/10.1086/383434>

Предварительный анализ данных

	price	num_cores	processor_speed	battery_capacity	ram_capacity	internal_memory	screen_size	refresh_rate	num_rear_cameras	num_front_cameras
mean	37252	7,8	2,42	4828,52	6,57	137,6	6,55	92,54	2,82	1,03
std	46005	0,81	0,46	1018,99	2,75	97,25	0,33	28,83	0,77	0,17
min	4304	4	1,2	1821	1	8	3,54	60	1	1
25%	15989	8	2,05	4500	4	64	6,5	60	2	1
50%	24372	8	2,3	5000	6	128	6,58	90	3	1
75%	40588	8	2,84	5000	8	128	6,67	120	3	1
max	799500	8	3,22	22000	18	1024	8,03	240	4	2

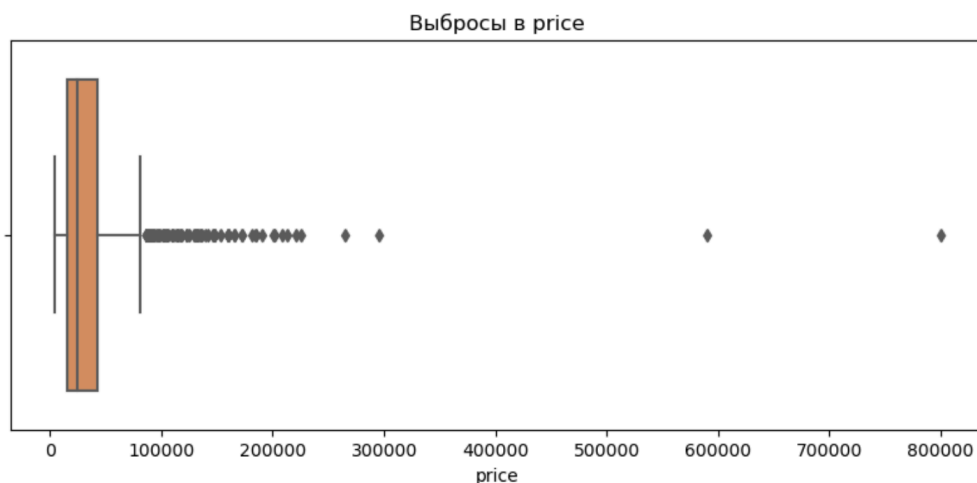
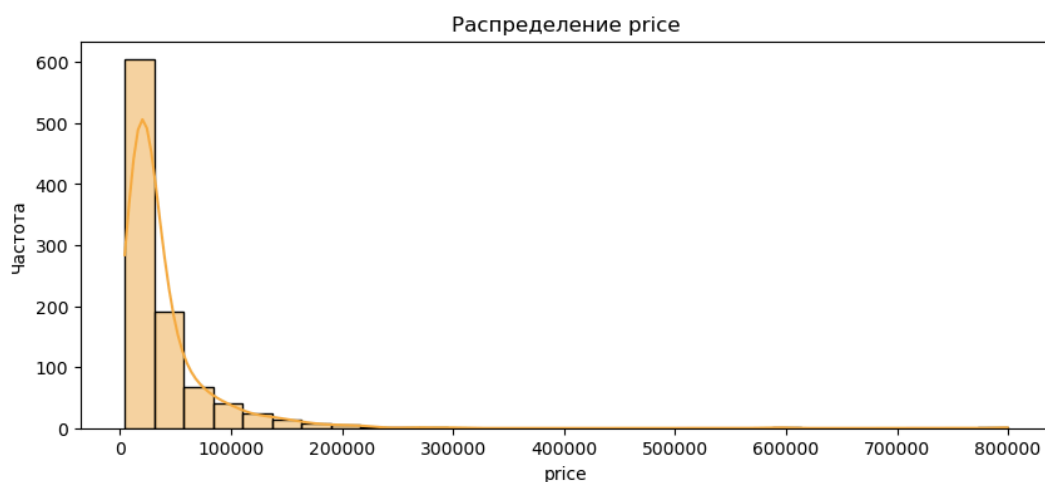


Анализ полученных корреляций между числовыми переменными:

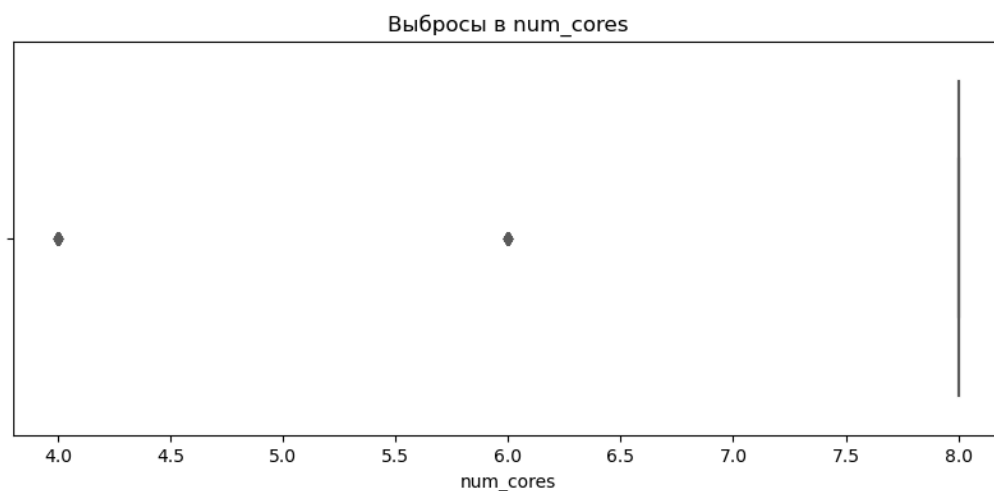
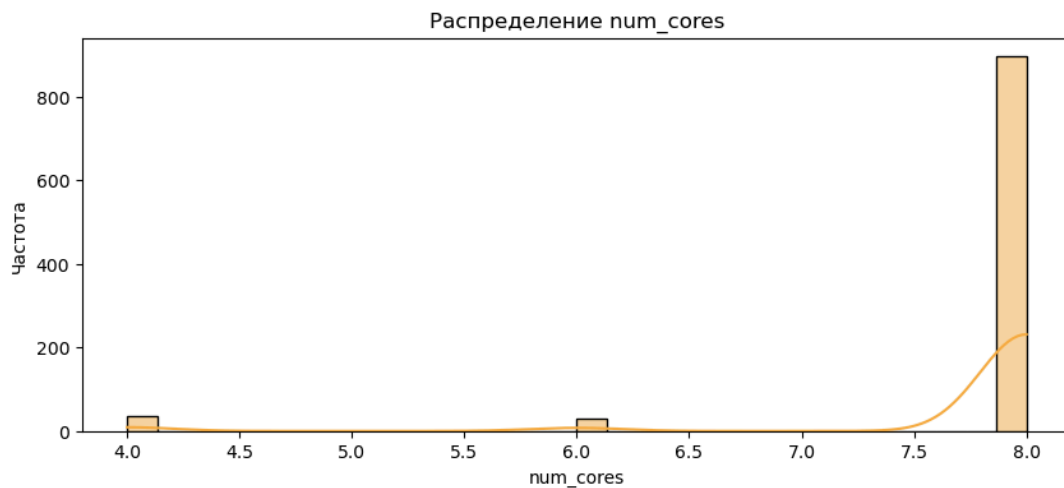
- **Наибольшую корреляцию технических характеристик смартфона с его ценой** демонстрирует встроенная память (*internal_memory*) - она составляет около 0.52. Это означает, что более высокая емкость внутренней памяти, как правило, связана с более дорогими смартфонами. Также относительно высокая положительная корреляция цены наблюдается с оперативной памятью (*ram_capacity*) - около 0.41 и с частотой процессора (*processor_speed*) ~0.47.
- **Наибольшую отрицательную корреляцию с ценой устройства** показывает емкость аккумулятора (*battery_capacity*) - корреляция составляет около -0.14. Хотя это может показаться контринтуитивным, скорее всего, такое значение просто свидетельствует о том, что наличие более емкой батареи не является прямым драйвером роста цены устройства.
- **Что касается корреляции между объясняющими переменными**, между ОЗУ (*ram_capacity*) и встроенной памятью (*internal_memory*) наблюдается довольно высокая корреляция (~0.65). Это может означать, что производители часто увеличивают объем ОЗУ и объем внутренней памяти одновременно. Аналогичный вывод можно сделать и судя по высокой корреляции между *processor_speed* и *ram_capacity* (0.65) - увеличение частоты процессора сопряжено с увеличением ОЗУ. Получается, здесь мы наблюдаем некоторый “кластер” характеристик, связанных с производительностью и объемом памяти. Между частотой обновления экрана (*refresh_rate*) и объемом оперативной памяти (*ram_capacity*) тоже есть заметная корреляция (около 0.57). Это говорит о том, что премиальные устройства оснащаются одновременно и большим объемом RAM, и экранами с более высокой *refresh_rate*. С технической точки зрения это весьма логично - большая частота обновления требует больше ресурсов. Еще интересно, что сильной корреляции между частотой обновления экрана и объемом батареи не наблюдается (0.09), хотя известно, что большая частота обновления экрана приводит к быстрому расходу заряда аккумулятора.

Анализ полученных описательных статистик:

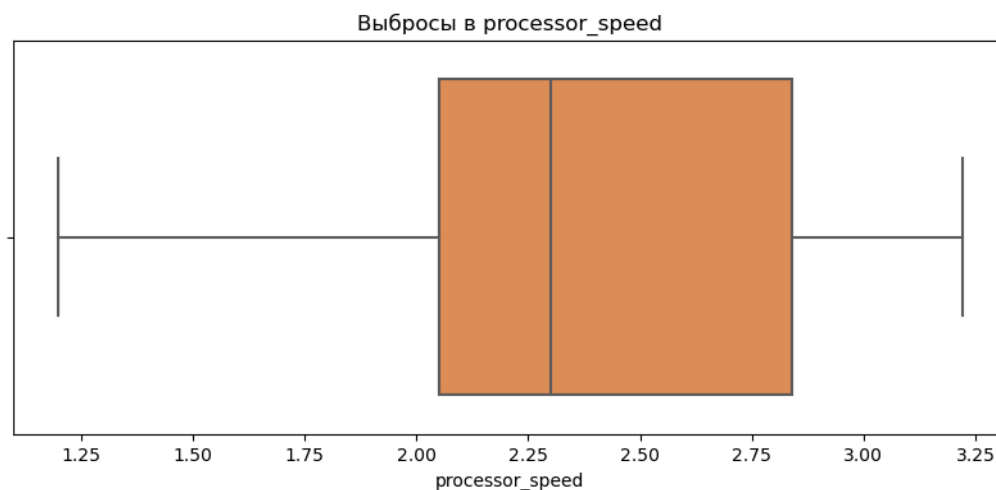
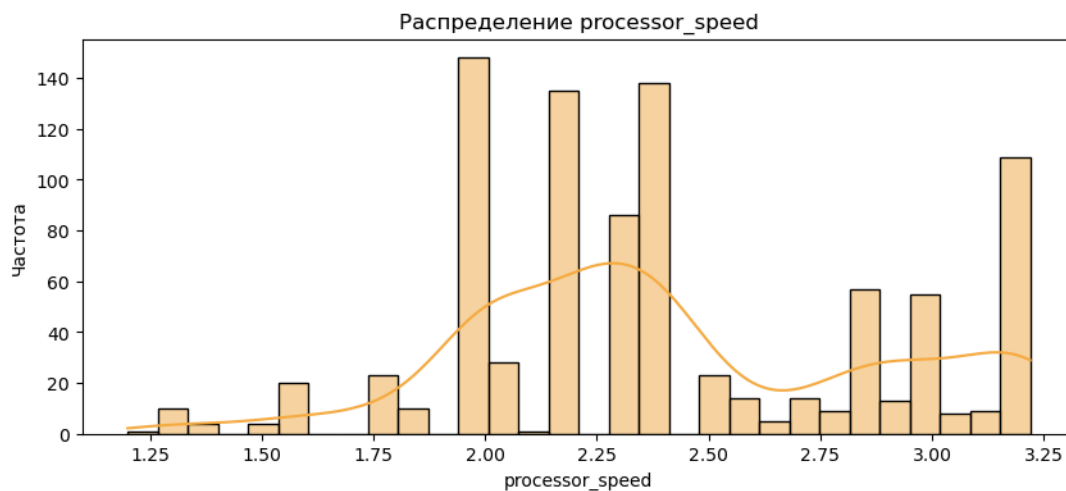
- **Цена устройства (price) в рублях - среднее значение цены** составляет около 37.252 рублей, что соответствует общепринятой градации устройств по ценовому сегменту (средний сегмент), однако мы наблюдаем достаточно высокое **стандартное отклонение** (около 46.005), что указывает на значительную вариацию стоимости смартфонов. **Минимальная цена** чуть более 4.300, а максимальная — практически 800.000. Такая большая разница свидетельствует о наличии крайне дорогих, возможно премиальных устройств, которые существенно поднимают верхний предел. **Медиана** (около 24.372) и **квартильные значения** (25%-кв. ~15.989, 75%-кв. ~40.588) указывают, что значительная часть устройств сосредоточена в диапазоне до 40–45 тыс. Однако наличие «хвоста» из очень дорогих аппаратов формирует сильный правый перекося распределения, делая его «ненормальным».



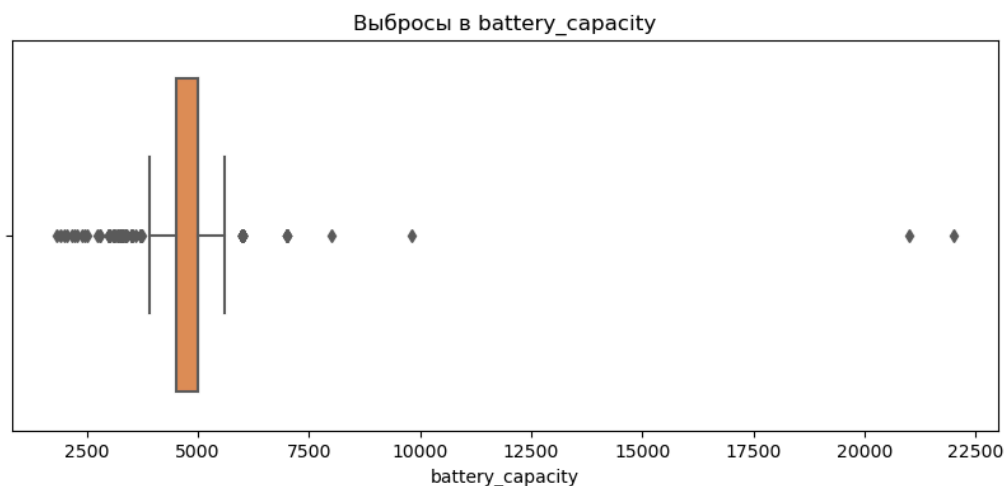
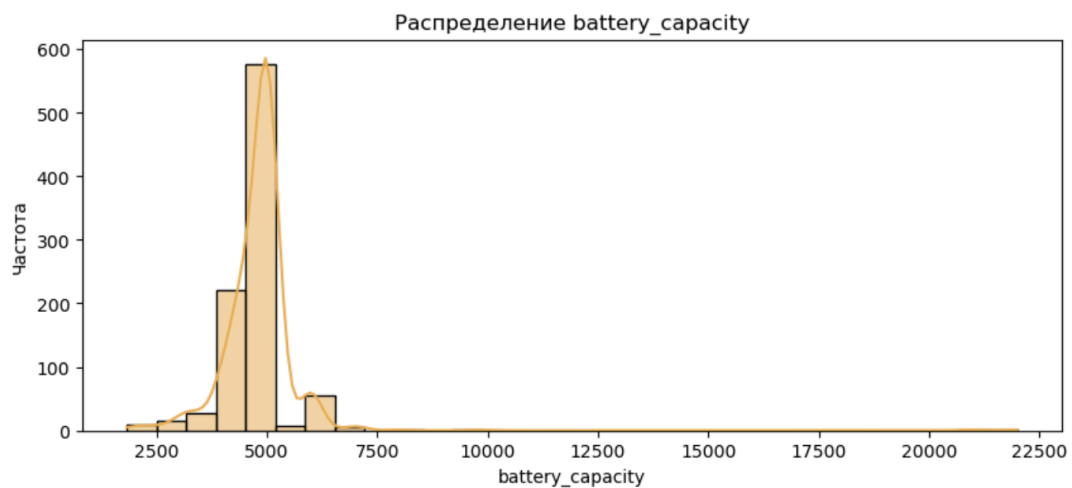
- **Количество ядер процессора (num_cores)** - среднее значение около 7.8, медиана и большинство квартилей равны 8, выбросов наблюдается достаточно мало. Столь низкая вариативность может означать, что простой учет ядер, скорее всего, не оказывает серьезного влияния на цену и стоит сделать упор на оценку более “тонких” аппаратных характеристик процессора. Но мы это еще проверим.



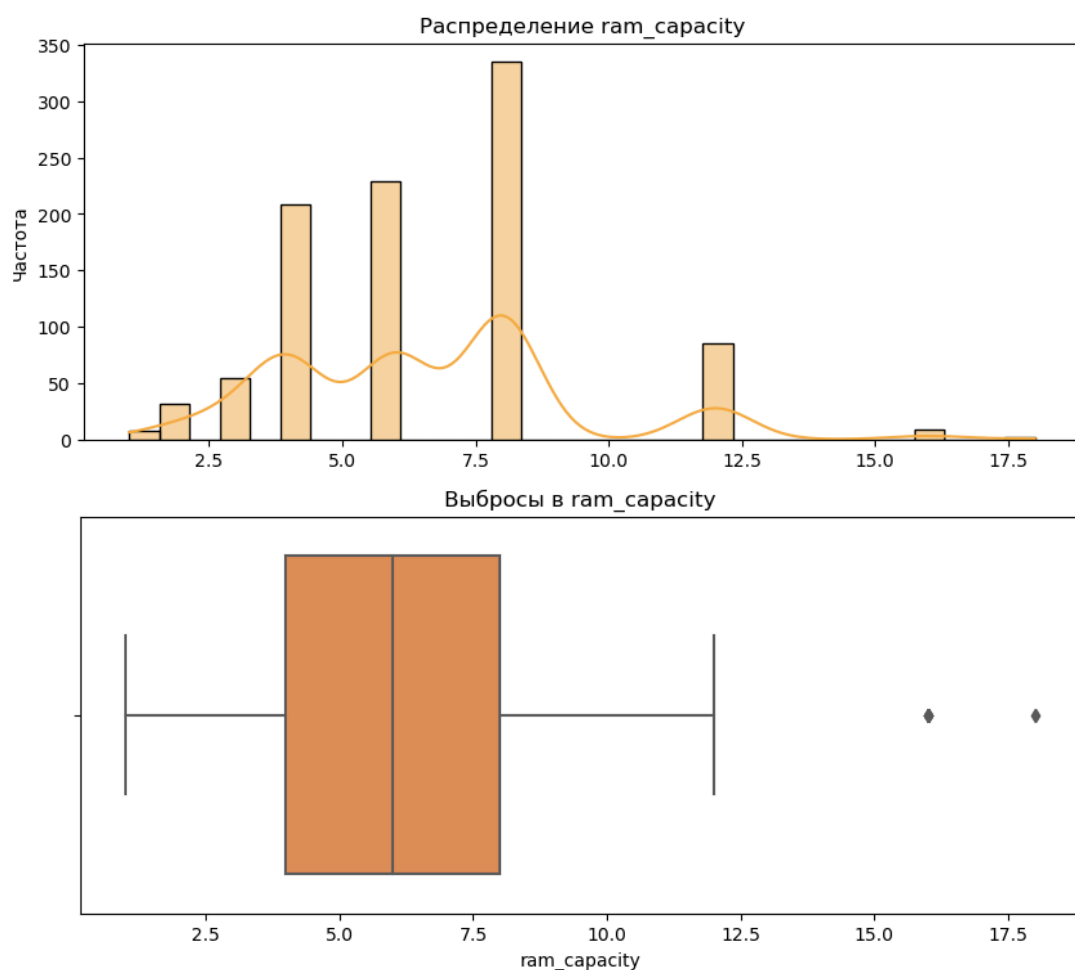
- **Частота процессора в ГГц (*processor_speed*)** - среднее значение составляет около 2.42 ГГц, стандартное отклонение ~0.46, размах от 1.2 ГГц до 3.22 ГГц, медиана ~2.3 ГГц. Такой результат в принципе отражает реальное положение дел на рынке смартфонов - большинство устройств работает в диапазоне от 2 до 3 ГГц, что обеспечивает баланс между энергопотреблением и производительностью. Есть устройства с низкими частотами (бюджетный сегмент, энергоэффективные и недорогие чипы), есть дорогие устройства с флагманскими процессорами.



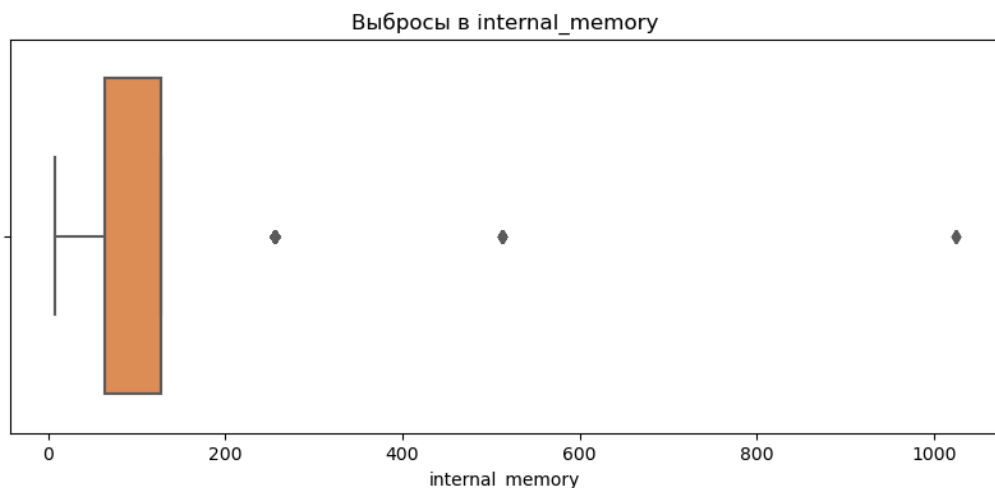
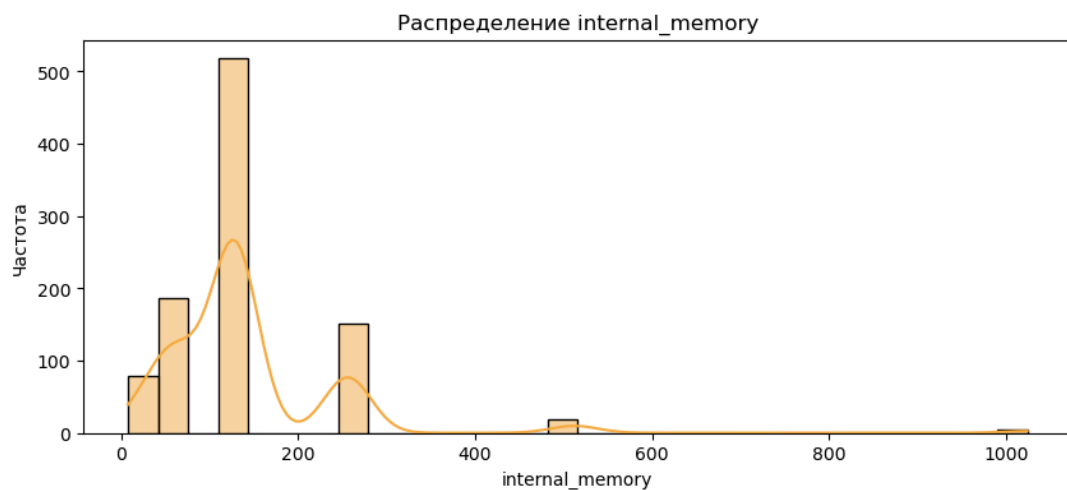
- **Емкость аккумулятора (*battery_capacity*)** - среднее значение составляет около 4 829 mAh, медиана — 5 000 mAh. Большинство устройств сконцентрировано в диапазоне около 4 500–5 000 mAh, что в принципе является современным стандартом рынка смартфонов. Максимальное значение (22 000 mAh) выглядит как выброс или устройство со специфическим позиционированием (например, rugged-смартфон с повышенной емкостью аккумулятора). Данная характеристика вполне может оказывать умеренное влияние на цену, но в пределах типичных значений (4500–5000 mAh) она вряд ли является главным драйвером роста стоимости.



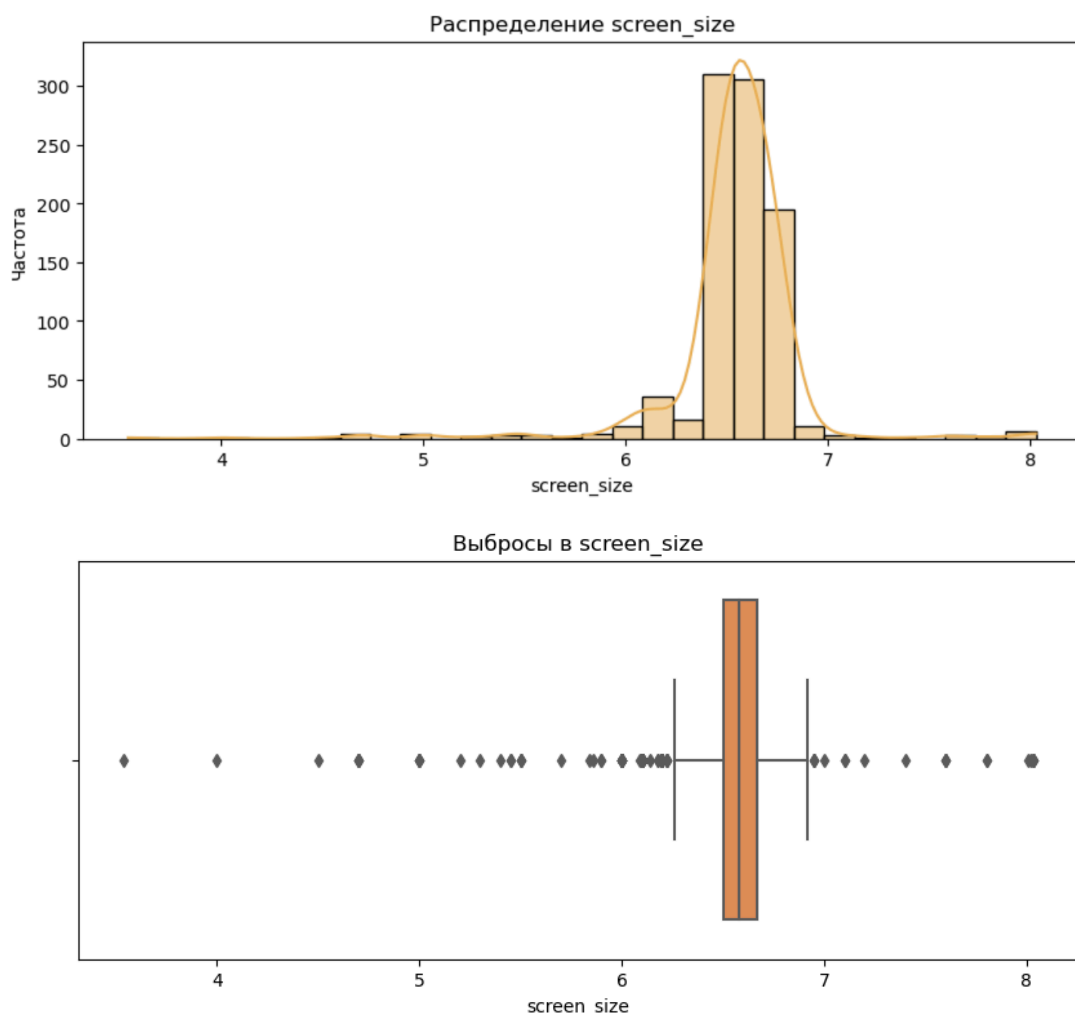
- **Оперативная память (ram_capacity)** - мы видим, что *среднее значение* составляет около 6,57 Гб, *медиана* — 6 Гб, *стандартное отклонение* - 2.75. Также стоит обратить особое внимание на то, что *диапазон значений* от 1 до 18 Гб ОЗУ выглядит очень широким. Это говорит о существенной дифференциации устройств по объему оперативной памяти - модели с большим объемом ОЗУ могут заметно отличаться в цене и, вероятно, объем RAM будет одним из сильных факторов, оказывающих наибольшее влияние на стоимость устройства.



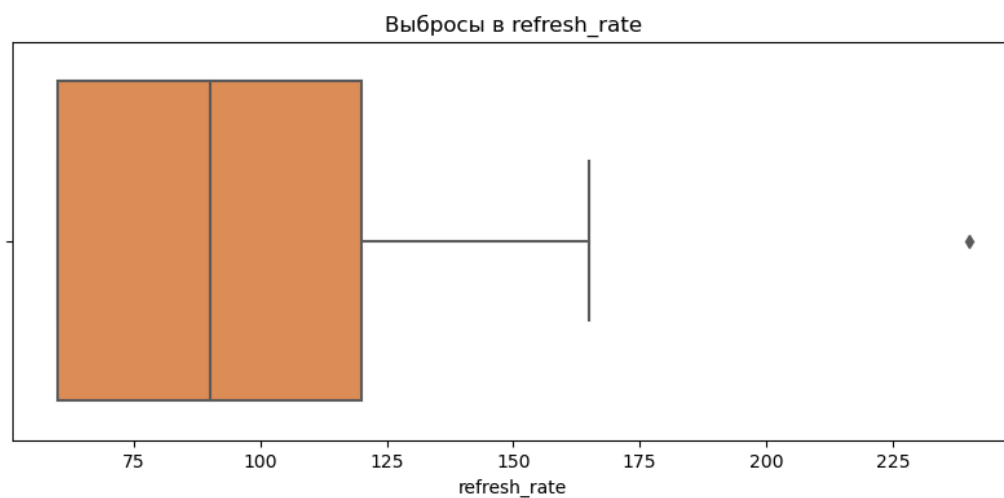
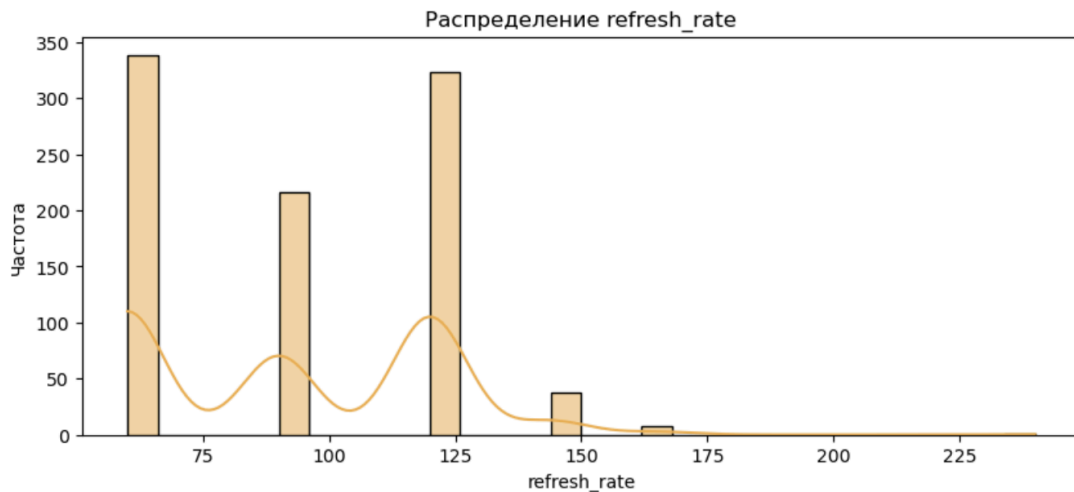
- **Встроенная память (*internal_memory*) - выбросы** практически отсутствуют, **среднее значение** составляет около 138 Гб, **медиана** - 128 Гб. В данном случае с точки зрения интерпретации лучше всего будет ориентироваться именно на медианное значение (так как что встроенная память, что ОЗУ всегда являются степенями двойки). Здесь мы также наблюдаем очень широкий диапазон значений от 8 Гб до 1024 Гб (1 Тб). Это свидетельствует о том, что объем встроенной памяти может оказывать высокое влияние на цену и являться одним из ключевых признаков сегментации рынка - бюджетные модели с 64–128 ГБ встроенной памяти, среднеценовые – 128–256 ГБ, премиум – 512 ГБ и выше.



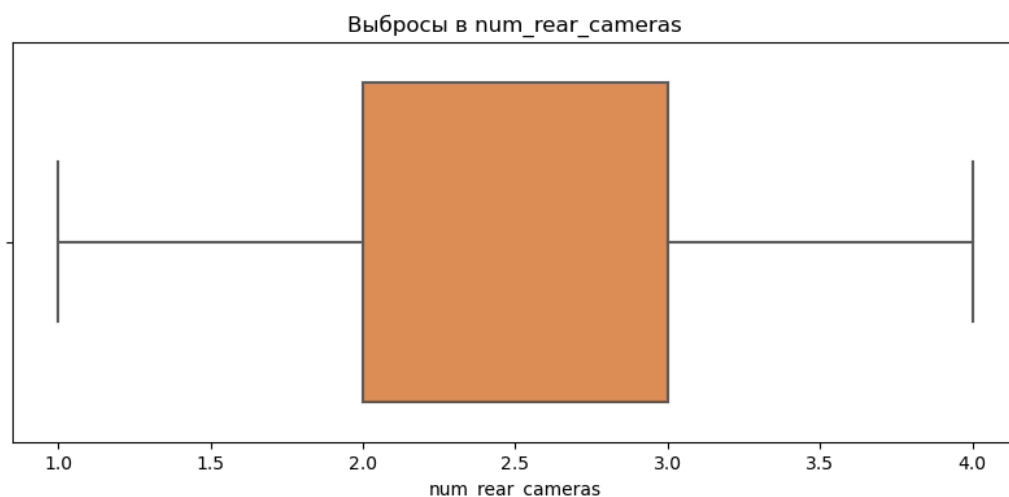
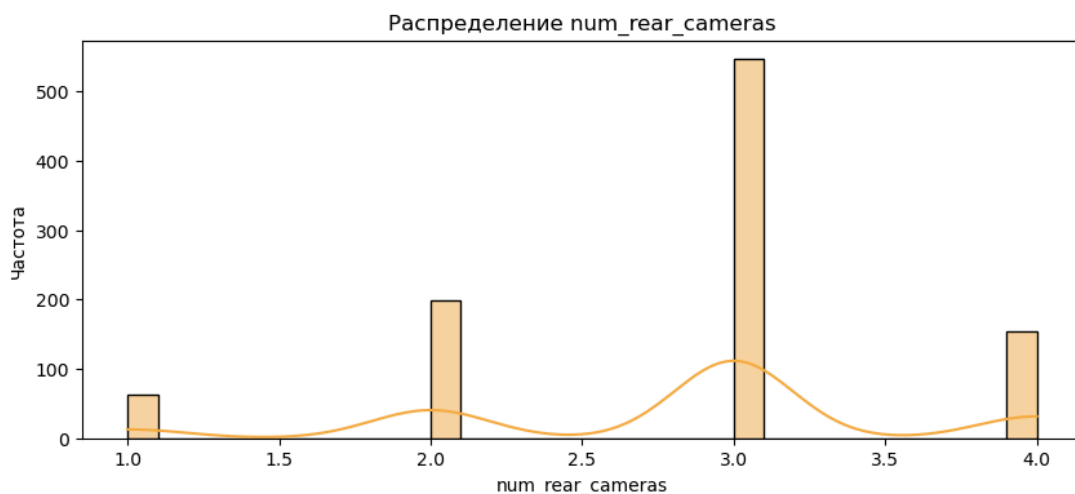
- **Диагональ экрана (screen_size)** - мы имеем **среднее значение** около 6.55 дюйма при относительно небольшой дисперсии (**ст. отклонение** около 0,33). Большинство современных смартфонов имеют размер экрана в районе 6,5 дюйма, что стало де-факто стандартом. Сильного влияния на цену лишь за счет диагонали ожидать, наверное, не стоит, но устройства с очень большим (7+ дюймов) или очень маленьким экраном могут стоить дороже или дешевле в силу особенностей, связанных с принадлежностью к той или иной нише рынка (например, есть компактные смартфоны, предназначенные прежде всего для детского сегмента, которые стоят достаточно дешево).



- **Частота обновления экрана (*refresh_rate*)** - среднее значение 92,5 Гц, *std* ~29, минимальное значение - 60 Гц, максимальное - 240 Гц. Судя по диапазону, в наших данных наблюдается значительный разброс по частоте обновления экрана, что может быть важным фактором ценообразования, поскольку высокие частоты (120 Гц и более), как правило, типичны для более дорогих, премиальных моделей.



- **Количество основных и фронтальных камер (num_rear_cameras и num_front_cameras) - основных камер в среднем около 2.8, медианное значение - 3 камеры, максимум - 4.** Многомодульные системы задних камер в современном мире уже стали своего рода стандартом или даже “трендом” для всех сегментов рынка. Увеличение числа камер безусловно может оказывать положительное влияние на стоимость устройства, но в условиях, когда почти у всех представленных моделей 2-3 камеры, это влияние может быть снижено. Фронтальная камера на устройстве **в среднем всего 1, максимальное значение - 2.** Варьирование в данном случае минимально и существенного вклада в стоимость устройства num_front_cameras, скорее всего, вносить не будет.



Построение регрессии

Для начала мы решили построить **классическую линейную регрессию с константой** и получили следующее уравнение регрессии:

$$\begin{aligned} \text{price} = & 21189.5327 - 1157.2523 \cdot \text{num_cores} + 14082.4214 \cdot \text{processor_speed} \\ & + 118.4243 \cdot \text{ram_capacity} - 3.2655 \cdot \text{battery_capacity} \\ & + 139.7945 \cdot \text{internal_memory} - 2046.5917 \cdot \text{screen_size} \\ & - 96.1218 \cdot \text{refresh_rate} + 155.4768 \cdot \text{num_rear_cameras} \\ & + 3652.1856 \cdot \text{num_front_cameras} - 5292.0927 \cdot \text{has_5g} \\ & + 15344.7121 \cdot \text{has_nfc} - 7781.7349 \cdot \text{extended_memory_available} \end{aligned}$$

Однако результаты работы модели оказались весьма плачевными:

OLS Regression Results						
<hr/>						
Dep. Variable:	price	R-squared:	0.381			
Model:	OLS	Adj. R-squared:	0.373			
Method:	Least Squares	F-statistic:	46.68			
Date:	Fri, 13 Dec 2024	Prob (F-statistic):	2.11e-86			
Time:	01:41:11	Log-Likelihood:	-10818.			
No. Observations:	924	AIC:	2.166e+04			
Df Residuals:	911	BIC:	2.173e+04			
Df Model:	12					
Covariance Type:	nonrobust					
<hr/>						
	coef	std err	t	P> t	[0.025	0.975]
<hr/>						
const	2.119e+04	2.21e+04	0.957	0.339	-2.23e+04	6.46e+04
num_cores	-1157.2523	1390.726	-0.832	0.406	-3886.651	1572.147
processor_speed	1.408e+04	3626.562	3.883	0.000	6965.035	2.12e+04
ram_capacity	118.4243	587.399	0.202	0.840	-1034.387	1271.236
battery_capacity	-3.2655	1.068	-3.059	0.002	-5.361	-1.170
internal_memory	139.7945	13.754	10.164	0.000	112.801	166.788
screen_size	-2046.5917	3620.538	-0.565	0.572	-9152.156	5058.973
refresh_rate	-96.1218	48.718	-1.973	0.049	-191.735	-0.508
num_rear_cameras	155.4768	1454.278	0.107	0.915	-2698.647	3009.600
num_front_cameras	3652.1856	5795.066	0.630	0.529	-7721.045	1.5e+04
has_5g	-5292.0927	2777.671	-1.905	0.057	-1.07e+04	159.285
has_nfc	1.534e+04	2669.379	5.748	0.000	1.01e+04	2.06e+04
extended_memory_available	-7781.7349	3015.381	-2.581	0.010	-1.37e+04	-1863.833
<hr/>						
Omnibus:	1721.385	Durbin-Watson:	1.902			
Prob(Omnibus):	0.000	Jarque-Bera (JB):	2230466.107			
Skew:	12.947	Prob(JB):	0.00			
Kurtosis:	242.298	Cond. No.	1.13e+05			
<hr/>						

- *Модель объясняет всего лишь около 38% вариации целевой переменной - в принципе, это умеренный показатель, но модель улавливает лишь часть факторов, оказывающих влияние на цену.*
- *Значение F-Stat (46.68) и ее p-value (~0.000...) указывают на то, что модель в целом значима.*

- *Теперь перейдем к интерпретации отдельных переменных:*

1. ***processor_speed (частота процессора)*** - коэффициент при ней составляет примерно 14.080, *p-value* ~0.000... Переменная статистически значима и положительно связана с ценой, увеличение частоты процессора связано с повышением цены. Это логично, так как более быстрая аппаратная платформа обычно стоит дороже;
2. ***battery_capacity (емкость батареи)*** - коэффициент составил около -3.2655, *p-value*: 0.002. Странно, что коэффициент при *battery_capacity* отрицательный, корреляция переменной с таргетом также была отрицательной (-0.14). Возможно, в наборе данных устройства с большой батареей могут чаще встречаться в более бюджетном сегменте. Или наличие других характеристик важнее и компенсируют влияние на цену емкости аккумулятора. Также вероятно, что тут присутствует мультиколлинеарность, а классическая линейная регрессия подобное побороть, к сожалению, неспособна;
3. ***internal_memory (встроенная память)*** - коэффициент: ~139.8, *p-value* близко к нулю. Регрессор значим и положительно влияет на цену, тут все вполне логично: больше памяти - выше издержки производителя, и, следовательно, выше конечная цена для потребителя;
4. ***refresh_rate (частота обновления экрана)*** - значение коэффициента составило примерно -96.12, а *p-value*: 0.049 (прямо на грани статистической значимости). Эффект отрицательный и слабозначимый. Это может указывать на сложную и неочевидную связь - возможно, наличие высокочастотных экранов не всегда коррелирует с более дорогими устройствами в данных или же у нас опять имеет место коллинеарность с другими признаками;
5. ***has_nfc (наличие NFC)*** - коэффициент составил примерно 15340, *p-value* стремится к нулю. Наблюдаем очень значимый и положительный эффект - смартфоны с NFC в среднем ощутимо дороже;

6. *extended_memory_available* (возможность расширения памяти) - коэффициент при регрессоре составил -7781.73, *p-value* около 0.01. Видим значимый отрицательный эффект - наличие слота для карты памяти ассоциируется с более низкой ценой. Это может отражать стратегию производителей: премиальные модели часто не дают возможности расширения памяти, полагаясь на большой внутренний объем, а более бюджетные — дают возможность установки карты памяти, зато стоят дешевле;
7. Остальные переменные: *num_cores* (число ядер), *ram_capacity* (объем ОЗУ), *screen_size* (диагональ экрана), *num_rear_cameras* (число задних камер), *num_front_cameras* (число фронтальных камер), *has_5g* не имеют статистически значимого эффекта в данной модели при уровне значимости 5%. Вряд ли такой результат может быть связан с реальным отсутствием влияния данных регрессоров на цену. Скорее всего дело в мультиколлинеарности, нелинейности или в общей неполноте модели, о чем нам намекает высокий *Condition Number* (113000).

Общий вывод о работе модели:

Модель в целом статистически значима и все-таки улавливает часть факторов, влияющих на цену. Некоторые результаты ожидаемы (частота процессора, внутренняя память и NFC увеличивают цену, например), другие — менее интуитивны (большая батарея и расширяемая память связаны со снижением цены, а наличие модуля 5G не оказывает значимого влияния на цену).

Мы видим явный потенциал для улучшения модели за счет разного рода трансформаций, нелинейных эффектов, взаимодействий переменных и добавления иных факторов.

Попробуем построить новую модель - прологарифмируем переменные: и *price* (нашу зависимую переменную), и независимые тоже - *processor_speed*, *internal_memory*. Так мы, возможно, сможем немного видоизменить структуру связей между переменными (может даже уменьшить степень линейной зависимости между признаками) и, главное, сможем научить модель улавливать не просто линейный характер влияния, но и мультипликативный. А логарифмирование цены позволит нам стабилизировать вариацию, сделать распределение более нормальным и ослабить влияние выбросов, коих немало. Также добавим в нашу модель квадратичный член частоты обновления экрана смартфона (*refresh_rate_sq*). Такое решение позволит нам учесть возможное наличие нелинейной зависимости: например, цена может расти при увеличении частоты обновления до определенного уровня, а затем эффект может изменяться (усиливаться или ослабевать). А еще сделаем вот что - создадим набор фиктивных переменных по количеству брендов минус 1 (у нас всего есть 48 уникальных брендов). В качестве базового бренда определим “Apple” (высокоуровневая точка отсчета, относительно премиальный ценовой сегмент устройств). Ну и заодно добавим нелинейную формулу для регрессора встроенной памяти устройства (*internal_memory*). Это пригодится нам при проверке сформулированных ранее гипотез.

Результаты получились следующие:

OLS Regression Results						
Dep. Variable:	log_price	R-squared:	0.880			
Model:	OLS	Adj. R-squared:	0.872			
Method:	Least Squares	F-statistic:	103.8			
Date:	Fri, 13 Dec 2024	Prob (F-statistic):	0.00			
Time:	16:25:00	Log-Likelihood:	-66.330			
No. Observations:	924	AIC:	256.7			
Df Residuals:	862	BIC:	556.0			
Df Model:	61					
Covariance Type:	nonrobust					
	coef	std err	t	P> t	[0.025	0.975]
Intercept	8.3731	0.282	29.654	0.000	7.819	8.927
C(brand)[T.Asus]	-0.8780	0.147	-5.984	0.000	-1.166	-0.590
C(brand)[T.BLU]	-1.5119	0.284	-5.321	0.000	-2.070	-0.954
C(brand)[T.Blackview]	-1.3087	0.282	-4.641	0.000	-1.862	-0.755
C(brand)[T.CAT]	0.2636	0.291	0.905	0.366	-0.308	0.835
C(brand)[T.Cola]	-0.5337	0.282	-1.891	0.059	-1.087	0.020
C(brand)[T.Dooqee]	-0.2219	0.251	-0.884	0.377	-0.714	0.271
C(brand)[T.Duoqin]	-0.3289	0.299	-1.101	0.271	-0.915	0.257
C(brand)[T.Gionee]	-1.3547	0.174	-7.784	0.000	-1.696	-1.013
C(brand)[T.Google]	-0.5723	0.117	-4.906	0.000	-0.801	-0.343
C(brand)[T.Honor]	-1.0256	0.110	-9.359	0.000	-1.241	-0.811
C(brand)[T.Huawei]	-0.4770	0.108	-4.430	0.000	-0.688	-0.266
C(brand)[T.Infinix]	-0.9658	0.099	-9.802	0.000	-1.159	-0.772
C(brand)[T.Itel]	-1.0101	0.209	-4.829	0.000	-1.421	-0.600
C(brand)[T.Jio]	-0.8923	0.157	-5.688	0.000	-1.200	-0.584
C(brand)[T.LG]	-0.1398	0.177	-0.791	0.429	-0.486	0.207
C(brand)[T.Lava]	-1.2120	0.155	-7.799	0.000	-1.517	-0.907
C(brand)[T.LeEco]	-1.3118	0.286	-4.595	0.000	-1.872	-0.751
C(brand)[T.Leitz]	-0.5504	0.439	-1.254	0.210	-1.412	0.311
C(brand)[T.Lenovo]	-1.1663	0.213	-5.488	0.000	-1.583	-0.749
C(brand)[T.Letv]	-1.7643	0.210	-8.396	0.000	-2.177	-1.352
C(brand)[T.Lyf]	-1.2450	0.213	-5.856	0.000	-1.662	-0.828
C(brand)[T.Micromax]	-1.1199	0.180	-6.229	0.000	-1.473	-0.767
C(brand)[T.Motorola]	-0.9621	0.091	-10.612	0.000	-1.140	-0.784
C(brand)[T.Nokia]	-0.6328	0.116	-5.461	0.000	-0.860	-0.405
C(brand)[T.Nothing]	-0.9956	0.145	-6.860	0.000	-1.280	-0.711
C(brand)[T.Nubia]	-0.8421	0.139	-6.067	0.000	-1.114	-0.570
C(brand)[T.OPPO]	-0.8538	0.088	-9.743	0.000	-1.026	-0.682
C(brand)[T.OnePlus]	-0.8446	0.090	-9.359	0.000	-1.022	-0.668
C(brand)[T.Oppo]	-0.8180	0.096	-8.555	0.000	-1.006	-0.630
C(brand)[T.Oukitel]	-0.1753	0.217	-0.808	0.420	-0.601	0.251
C(brand)[T.POCO]	-1.0525	0.110	-9.541	0.000	-1.269	-0.836
C(brand)[T.Poco]	-1.0630	0.096	-11.052	0.000	-1.252	-0.874
C(brand)[T.Realme]	-0.9375	0.087	-10.831	0.000	-1.107	-0.768
C(brand)[T.Redmi]	-1.0596	0.177	-5.991	0.000	-1.407	-0.712
C(brand)[T.Royole]	0.4078	0.216	1.887	0.059	-0.016	0.832
C(brand)[T.Samsung]	-0.6619	0.084	-7.921	0.000	-0.826	-0.498
C(brand)[T.Sharp]	-0.5243	0.283	-1.851	0.065	-1.080	0.032
C(brand)[T.Sony]	-0.3448	0.119	-2.889	0.004	-0.579	-0.111
C(brand)[T.TCL]	-0.7985	0.282	-2.832	0.005	-1.352	-0.245
C(brand)[T.Tecno]	-1.0709	0.096	-11.127	0.000	-1.260	-0.882
C(brand)[T.Vertu]	3.6442	0.290	12.566	0.000	3.075	4.213
C(brand)[T.Vivo]	-0.7725	0.084	-9.145	0.000	-0.938	-0.607
C(brand)[T.Xiaomi]	-0.9135	0.084	-10.914	0.000	-1.078	-0.749
C(brand)[T.ZTE]	-0.9535	0.207	-4.605	0.000	-1.360	-0.547
C(brand)[T.iKall]	-0.9635	0.183	-5.264	0.000	-1.323	-0.604
C(brand)[T.iQOO]	-1.0008	0.094	-10.656	0.000	-1.185	-0.816
C(brand)[T.itel]	-1.0163	0.134	-7.603	0.000	-1.279	-0.754
num_cores	-0.0411	0.017	-2.424	0.016	-0.074	-0.008
log_processor_speed	0.8884	0.089	9.929	0.000	0.713	1.064
ram_capacity	0.0553	0.007	8.172	0.000	0.042	0.069
battery_capacity	-6.91e-05	1.41e-05	-4.889	0.000	-9.68e-05	-4.14e-05
internal_memory	0.0021	0.000	6.668	0.000	0.001	0.003
internal_memory_sq	-1.266e-06	3.43e-07	-3.692	0.000	-1.94e-06	-5.93e-07
screen_size	0.2203	0.045	4.911	0.000	0.132	0.308
refresh_rate	-0.0035	0.003	-1.203	0.229	-0.009	0.002
refresh_rate_sq	2.561e-05	1.55e-05	1.649	0.099	-4.87e-06	5.61e-05
num_rear_cameras	0.0845	0.014	5.848	0.000	0.056	0.113
num_front_cameras	0.0594	0.054	1.095	0.274	-0.047	0.166
has_5g	0.1171	0.028	4.244	0.000	0.063	0.171
has_nfc	0.1940	0.028	6.898	0.000	0.139	0.249
extended_memory_available	-0.1275	0.031	-4.173	0.000	-0.188	-0.068
Omnibus:	393.648	Durbin-Watson:	1.871			
Prob(Omnibus):	0.000	Jarque-Bera (JB):	6008.431			
Skew:	1.527	Prob(JB):	0.00			
Kurtosis:	15.113	Cond. No.	4.93e+06			

- **Коэффициент детерминации** в сравнении с предыдущей “простой” моделью **значительно вырос** - теперь он составляет целых 0.88 (против 0.38 в предыдущей модели), что является очень хорошим показателем.
- Значение F-статистики и ее p-value указывают на то, что **модель в целом значима**.
- “Брендовые коэффициенты” показывают то, насколько log-price смартфона отличается от log-price смартфона Apple при условии одинаковых технических характеристик - большинство брендов имеет отрицательный и статистически значимый коэффициент. Это означает, что их смартфоны в среднем дешевле, чем смартфоны Apple (например, коэффициент при BLKview -1.3087, при Samsung: -0.6619). Очень сильно из общего ряда выбивается бренд Vertu (коэффициент при данной переменной значим и составляет целых 3.64). Это вполне логично - бренд Vertu позиционирует свои устройства как нечто невероятно премиальное, недоступное “простым смертным” (они производят смартфоны из драгоценных металлов, инкрустируют драгоценными камнями и т.д.). Стоит отметить, что некоторые бренды (например, Lg и Oukitel) имеют отрицательные, но статистически незначимые коэффициенты - значит, мы не можем уверенно сказать, отличаются ли они от Apple по цене (хотя интуиция подсказывает, что отличия в меньшую сторону явно есть - эти производители не славятся своими премиальными устройствами, способными конкурировать с айфонами).
- **Поговорим об остальных (небрендовых) коэффициентах модели:**
 1. **num_cores (число ядер процессора)** - коэффициент составляет ~0.04, p-value: 0.016, что делает число ядер процессора статистически значимым - при прочих равных увеличение num_cores немного повышает стоимость устройства, в предыдущей нашей модели мы получили иной результат: число ядер процессора не оказывало значимого влияния на стоимость устройства;
 2. **log_processor_speed (логарифм частоты процессора)** - коэффициент ~0.8884, p-value стремится к нулю. Статистически значимо и положительно - устройства с

более быстрой частотой процессора ощутимо дороже, что вполне логично;

3. **ram_capacity (объем ОЗУ)** - коэффициент примерно равен 0.0553 и статистически значим: больший объем ОЗУ ожидаемо ведет к повышению цены устройства, вывод отличен от предыдущей модели;
4. **battery_capacity (емкость аккумулятора)** - коэффициент -0.0000691 , *p-value* стремится к нулю. Наблюдаем значимый и отрицательный эффект (так же, как и в простой линейной регрессии). Как вариант, это может отражать особенности образца (например, производители часто ставят в дешевые модели большие батареи, но не повышают цену);
5. **internal_memory (внутренняя память)** - коэффициент ~ 0.021 , *p-value* стремится к нулю. Наблюдаем положительный и значимый эффект, все логично - чем больше памяти, тем выше цена устройства;
6. **internal_memory_sq (квадрат объема внутренней памяти)** - коэффициент при объясняющей переменной примерно равен -0.000001266 , *p-value* стремится к нулю. Отрицательный коэффициент при квадратичном члене говорит о убывающей предельной отдаче: на маленьких объемах добавление памяти сильно повышает цену, но по мере роста памяти эффект на цену снижается;
7. **screen_size (размер экрана)** - коэффициент ~ 0.2203 , *p-value* примерно равно нулю. В сравнении с предыдущей моделью здесь мы уже можем наблюдать положительное влияние диагонали экрана на цену, что вполне логично и интуитивно понятно;
8. **refresh_rate (частота обновления экрана)** - *p-value* составляет 0.229, что говорит нам об отсутствии статистической значимости данного коэффициента в модели. Такая же ситуация наблюдалась и в предыдущей модели - это странно, возможно, мы не все-таки не до конца поборол мультиколлинеарность или еще что-то;
9. **refresh_rate_sq (квадрат частоты обновления экрана)** - коэффициент при регрессоре составил ~ 0.00002561 , *p-value* ~ 0.099 . Наблюдаем легкий намек на наличие нелинейной

взаимосвязи, но, возможно, эта связь описывается неквадратично;

10. **num_rear_cameras (число задних камер)** - коэффициент ~ 0.0845 , *p-value* стремится к нулю. Значимо и положительно: больше задних камер - выше цена. Вполне разумно, ведь чем больше камер, тем выше издержки производителя на аппаратное оснащение устройства (более логичный результат в отличие от предыдущей нашей модели, где количество задних камер не оказывало значимого влияния на цену устройства);
11. **num_front_cameras (число фронтальных камер)** - коэффициент ~ 0.0594 , *p-value* = 0.274. Незначимо. Число фронтальных камер не оказывает статистически доказуемого влияния на целевую переменную. Вероятно, это связано с тем, что абсолютное большинство объектов в выборке имеют 1 фронтальную камеру, поэтому влияние их количества на цену отследить невозможно;
12. **has_5g (наличие модуля 5G)** - коэффициент ~ 0.1171 , *p-value* стремится к нулю. Значимо и положительно: смартфоны с 5G дороже, что вполне разумно и логично. Данный стандарт связи - новая и крайне востребованная в современном мире технология, которая высоко ценится пользователями и не может не влиять на стоимость;
13. **has_nfc (наличие NFC)** - коэффициент ~ 0.1940 , *p-value* практически нулевое. Значимо и положительно - наличие модуля NFC повышает цену. Здесь все аналогично технологии 5G: NFC востребовано и несет для производителей дополнительные издержки, что ведет к повышению конечной стоимости устройства;
14. **extended_memory_available (наличие слота для расширения памяти)** - коэффициент примерно равен -0.1275 , *p-value* стремится к нулю. Устройства с возможностью расширения памяти дешевле, что может быть связано с тем, что премиальные модели чаще обходятся без слотов для SD-карт, предлагая большой встроенный объем памяти "из коробки". Аналогичная интерпретация следовала и из предыдущей нашей модели.

Общий вывод о работе модели:

Как мы можем заметить, новая более “умная” модель показала себя куда лучше простой линейной регрессии по всем параметрам: ее R^2 кратно выше, чем у предыдущей модели (0.88 против 0.38), статистическая значимость многих коэффициентов в этот раз была доказана, их интерпретация вполне интуитивна (наличие модуля 5G оказывает значимое влияние на цену, ровно так же, как число задних камер и объем оперативной памяти). Странности с убывающей стоимостью устройства по мере увеличения объема аккумулятора сохранились. Также до сих пор не очень понятно, почему частота обновления экрана не оказывает значимого влияния на конечную цену устройства, ведь внедрение данной технологии, как минимум, несет дополнительные издержки для производителя. Мы связываем такой результат прежде всего с наличием какой-то крайне сложной взаимосвязи (нелинейной и неквадратичной) с ценой или с весьма сложным и необычным распределением (см. раздел “Описательные статистики”).

Проверка сформулированных ранее гипотез

- **Гипотеза о брендовой наценке** - “Даже при прочих равных технических параметрах смартфоны более известных брендов будут иметь цену существенно выше, чем аналогичные по характеристикам устройства менее известных производителей. То есть коэффициент при переменных, отражающих “бренд”, должен быть положительным и значимым.”:

Во вторую нашу модель мы выключили категориальную переменную $C(\text{brand})$, где Apple стал базовым брендом, а для остальных производителей были оценены соответствующие фиктивные переменные.

Результаты работы модели показывают, что для абсолютного большинства производителей коэффициенты отрицательны и значимы, что интерпретируется как дисконт к цене в сравнении с базовым брендом. Некоторые производители (например, Vertu, iQOO) имеют положительные и значимые коэффициенты, что говорит о еще более высокой цене на их устройства при прочих равных. Связано это с их еще более элитарным позиционированием, чем у Apple.

Таким образом, данные подтверждают наличие "брендовой" премии, гипотеза о брендовой наценке подтверждается: известные и "премиальные" бренды имеют значимо более высокую цену при одинаковых характеристиках, чем менее известные или более массовые производители.

- **Гипотеза о роли объема внутренней памяти и эффекте убывающей предельной полезности** - “Коэффициент при переменной объема внутренней памяти будет положительным, но при включении нелинейной формы появится дополнительный отрицательный коэффициент, отражающий убывающую предельную полезность.”:

В новую “умную” модель были включены переменные internal_memory (объем внутренней памяти) и $\text{internal_memory_sq}$ (квадрат объема внутренней памяти). Результаты показали, что

линейный коэффициент при *internal_memory* положительный и значимый, а при *internal_memory_sq* - отрицательный и тоже значимый.

Это означает, что при относительно низких значениях внутренней памяти добавление дополнительных гигабайт памяти существенно повышает цену, однако по мере роста объема встроенной памяти предельный эффект от увеличения памяти снижается - коэффициент при квадратичном члене отрицателен. То есть при переходе от очень маленького объема к умеренно большому цена растёт быстро, но при дальнейшем увеличении памяти прирост цены замедляется.

Гипотеза о роли объема внутренней памяти и убывающей предельной полезности подтвердилась.

- **Гипотеза об отсутствии различий между группами брендов в рамках одной страны** - “Мы можем выбрать подмножество брендов (например, рассмотреть популярных китайских производителей) и проверить гипотезу о том, что их премии/дисконты относительно базового бренда не отличаются друг от друга.” (была сформулирована после подтверждения Гипотезы о наличии брендовой наценки, так как они в некоторой степени взаимосвязаны):

Мы проверили данную гипотезу при помощи совместного F-теста, выбрав несколько известных китайских брендов смартфонов: Huawei, Xiaomi, Oppo и OnePlus. F-Stat оказалась равной примерно 9.96, что довольно много и часто указывает на отклонение от нулевой гипотезы о том, что коэффициенты выбранных нами для анализа брендов равны между собой. P-value составило 0.0000018454.

Следовательно **на всех разумных уровнях значимости H_0 отвергается** - по крайней мере один из брендов существенно отличается от других в терминах ценовой премии или дисконта и предлагает свою продукцию по значимо другой цене относительно Apple, чем остальные. Вероятно, такой результат связан с тем, что все рассмотренные нами бренды присутствуют в разных ценовых сегментах (и в бюджетном, и в средне-бюджетном, и в среднем, и в премиальном).

Заключение

Наше исследование показало, что **стоимость смартфонов может зависеть от целого множества факторов**: не только от “сырой”

производительности устройства (частоты процессора, количества ядер в нем, ОЗУ, встроенной памяти и др.), но и от имиджа того или иного бренда, наличия современных стандартов связи и передачи данных (5G и NFC-модулей).

Две из трех выдвинутых нами в самом начале гипотез были подтверждены - смартфоны от более известных производителей действительно имеют брендовую наценку, а объем внутренней памяти обладает убывающей предельной полезностью. **Гипотеза об отсутствии различий между группами брендов в рамках одной страны не подтвердилась** - возможно, стоило рассмотреть больше различных страновых групп, может мы не учли фактор международной репутации выбранных нами для анализа брендов, а может их премии/дисконты относительно базового бренда действительно отличаются друг от друга и каждый производитель самостоятельно определяет свою стратегию ценообразования, в чем тоже есть своя логика.

Результаты нашего исследования будут интересны как производителям и ритейлерам, так и потребителям. Первые смогут выстраивать более эффективную стратегию ценообразования, а вторые - принимать более взвешенные решения при выборе подходящего для себя устройства.

Но, наверное, наибольшую пользу наша работа несет для инвесторов и аналитиков. Влияние бренда на ценообразование устройств - крайне интересная область для последующих исследований. Роль этого фактора безусловно важна и очевидна, но насколько - вопрос все еще открытый.

Спасибо за внимание!