

# **Анализ и сегментация клиентов с помощью алгоритмов кластеризации**

Подготовил ст. гр. 2023-ФГиИБ-ПИ-1м  
Савельев М.В.

## Цель работы

Разработать систему сегментации клиентов для розничной компании с использованием алгоритмов кластеризации. Это позволит компании лучше понимать своих клиентов, персонализировать маркетинговые кампании и оптимизировать бизнес-процессы.

Сегментация клиентов с помощью алгоритмов кластеризации позволяет компании глубже понимать свою аудиторию, персонализировать маркетинг, оптимизировать бизнес-процессы и повышать прибыль за счет более эффективного управления ресурсами и улучшения клиентского опыта.

# Методология

1. Сбор и анализ данных о клиентах.
2. Предобработка и подготовка данных для моделирования.
3. Применение различных алгоритмов кластеризации для сегментации клиентов.
4. Оценка качества кластеризации с использованием внутренних и внешних метрик.
5. Интерпретация и визуализация результатов.
6. Формирование рекомендаций для бизнес-стратегии компании на основе полученных сегментов.

# Исходные данные

Этот набор данных Online Retail II содержит все транзакции, произошедшие для зарегистрированной в Великобритании немагазинной онлайн-розницы между 01/12/2009 и 09/12/2011. Компания в основном продает уникальные подарочные изделия на все случаи жизни. Многие клиенты компании являются оптовиками.

## Invoice

Номер счета-фактуры. Номинальный. 6-значный целый номер, уникально присваиваемый каждой транзакции. Если этот код начинается с буквы «с», это указывает на отмену.

## StockCode

Код продукта (товара). Номинальный. Пятизначное целое число, уникально присваиваемое каждому отдельному продукту.

## Description

Наименование товара (предмета). Номинальное.

## Quantity

Количество каждого продукта (товара) за транзакцию. Числовое.

## InvoiceDate

Дата и время выставления счета. Числовой. День и время, когда была создана транзакция.

## Price

Цена за единицу. Числовая. Цена продукта за единицу в фунтах стерлингов.

## Customer ID

Номер клиента. Номинальный. Пятизначный целый номер, уникально присваиваемый каждому клиенту.

## Country

Название страны. Номинал. Название страны, в которой проживает клиент.

# Первичный анализ данных (EDA)



# Первичный анализ данных (EDA)

Частота покупок по клиентам

ID клиента	Количество покупок
17841.0	13097
14911.0	11613
12748.0	7307
14606.0	6709
14096.0	5128
15311.0	4717
14156.0	4130
14646.0	3890
13089.0	3438
16549.0	3255

Распределение выручки по товарам

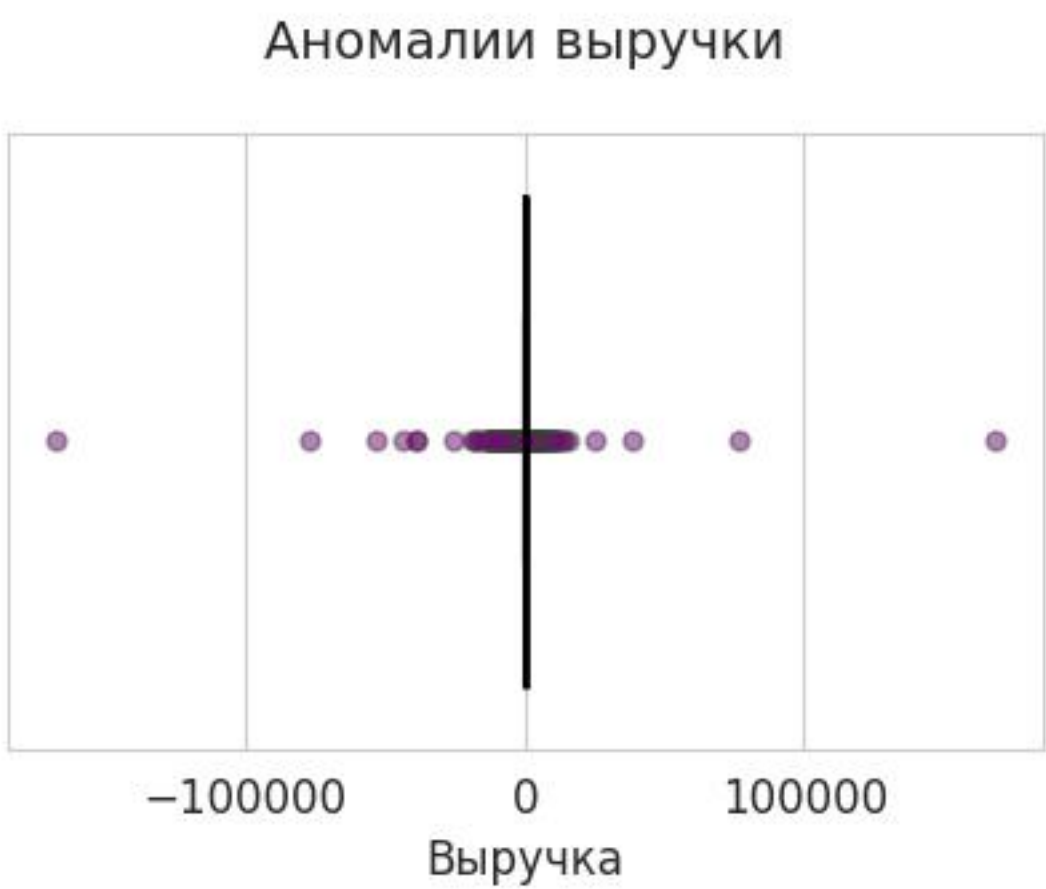
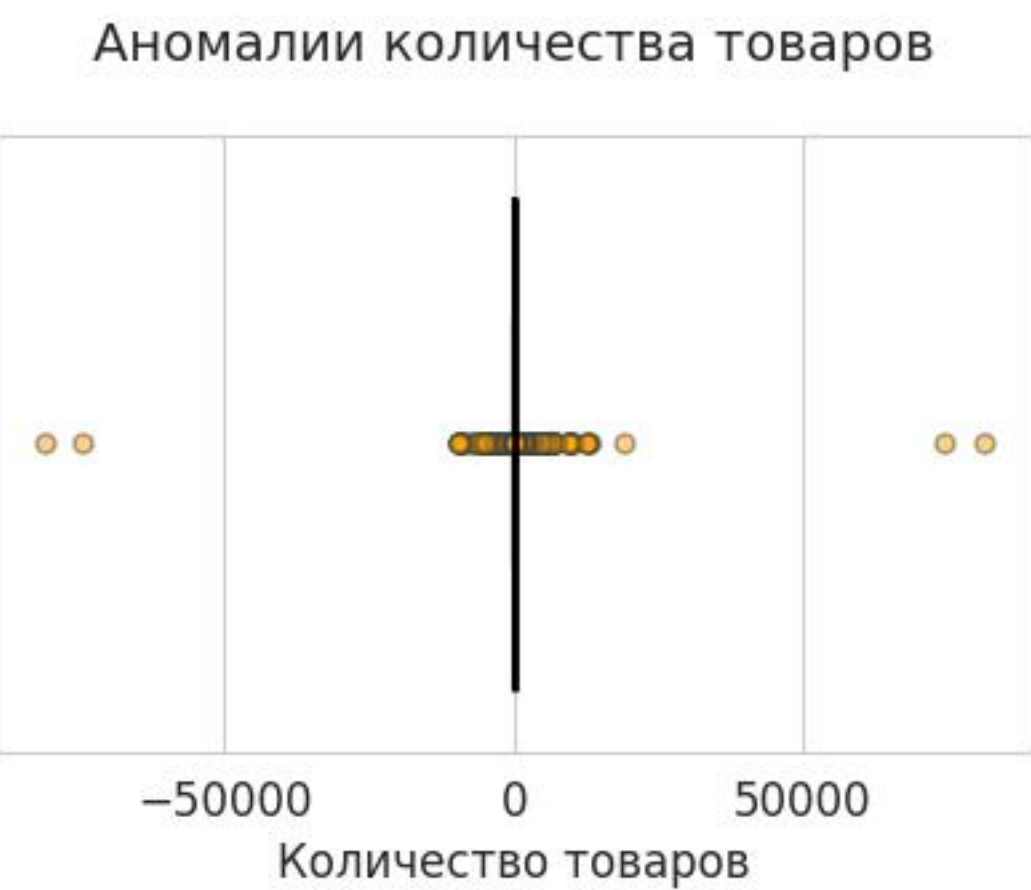
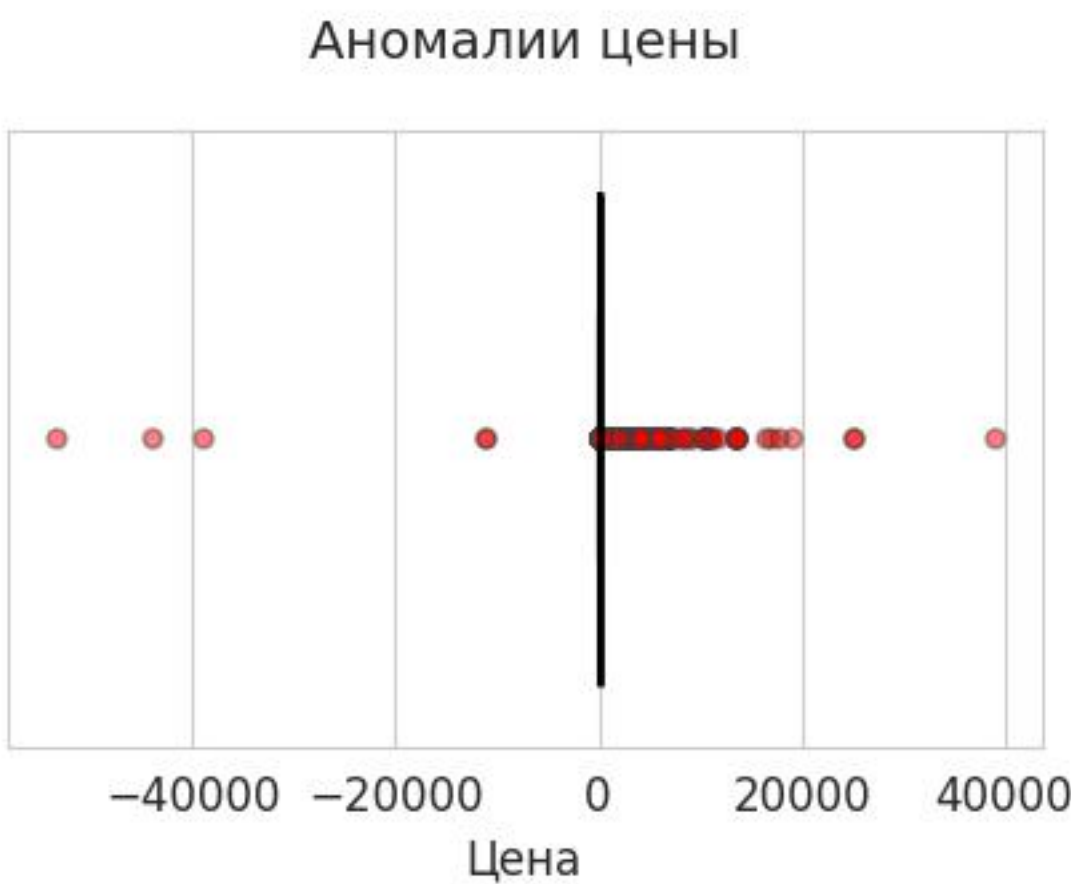
Название товара	Выручка
ПОДСТАВКА ДЛЯ ТОРТА REGENCY 3 ЯРУСА	327813.65
ПОЧТОВАЯ ОПЛАТА DOTCOM	322647.47
БЕЛАЯ ПОДВЕСНАЯ ПОДСВЕТКА В ФОРМЕ СЕРДЦА	257533.90
БОЛЬШОЙ МЕШОК КРАСНЫЙ РЕТРОСПУТ	148800.64
ВЕЧЕРИНКА	147948.50
УКРАШЕНИЕ В ВИДЕ ПТИЦЫ РАЗНЫХ ЦВЕТОВ	131413.85
НАБОР БУМАЖНЫХ ЦЕПЕЙ В СТИЛЕ РОЖДЕСТВА 50-Х	121662.14
ПОЧТОВАЯ ОПЛАТА	112341.00
ОГНИ ЧИЛИ	84854.16
ВРАЩАЮЩАЯСЯ СЕРЕБРЯНАЯ ПОДСВЕТКА В ФОРМЕ АНГЕЛА HLDR	73814.72

# Предобработка данных

Пропущенные данные

Атрибут	Кол-во пропущенных знач.
Invoice	0
StockCode	0
Description	4382
Quantity	0
InvoiceDate	0
Price	0
Customer ID	243007
Country	0

- Строки без названия товара будут удалены
- Покупателям будет присвоен новый ID, если значения отсутствуют



Строки с отрицательными значениями атрибутов цена, количество товара и выручка будут удалены

# Предобработка данных

Создание новых признаков:

- Recency (давность): Время с момента последней покупки.
- Frequency (частота): Количество покупок за определенный период.
- Monetary (сумма): Общая сумма покупок.
- Средний чек
- Предпочтительные категории товаров

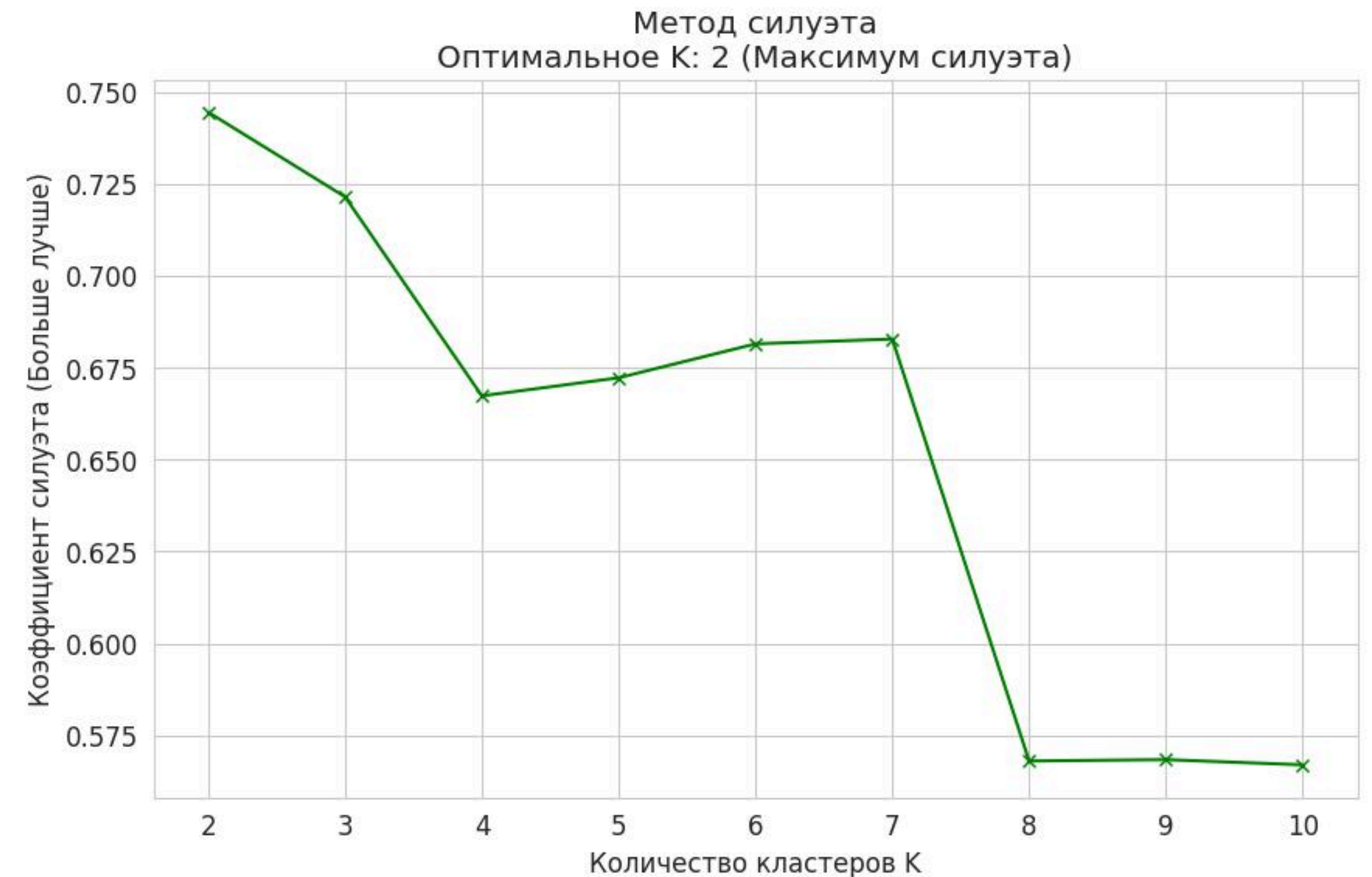
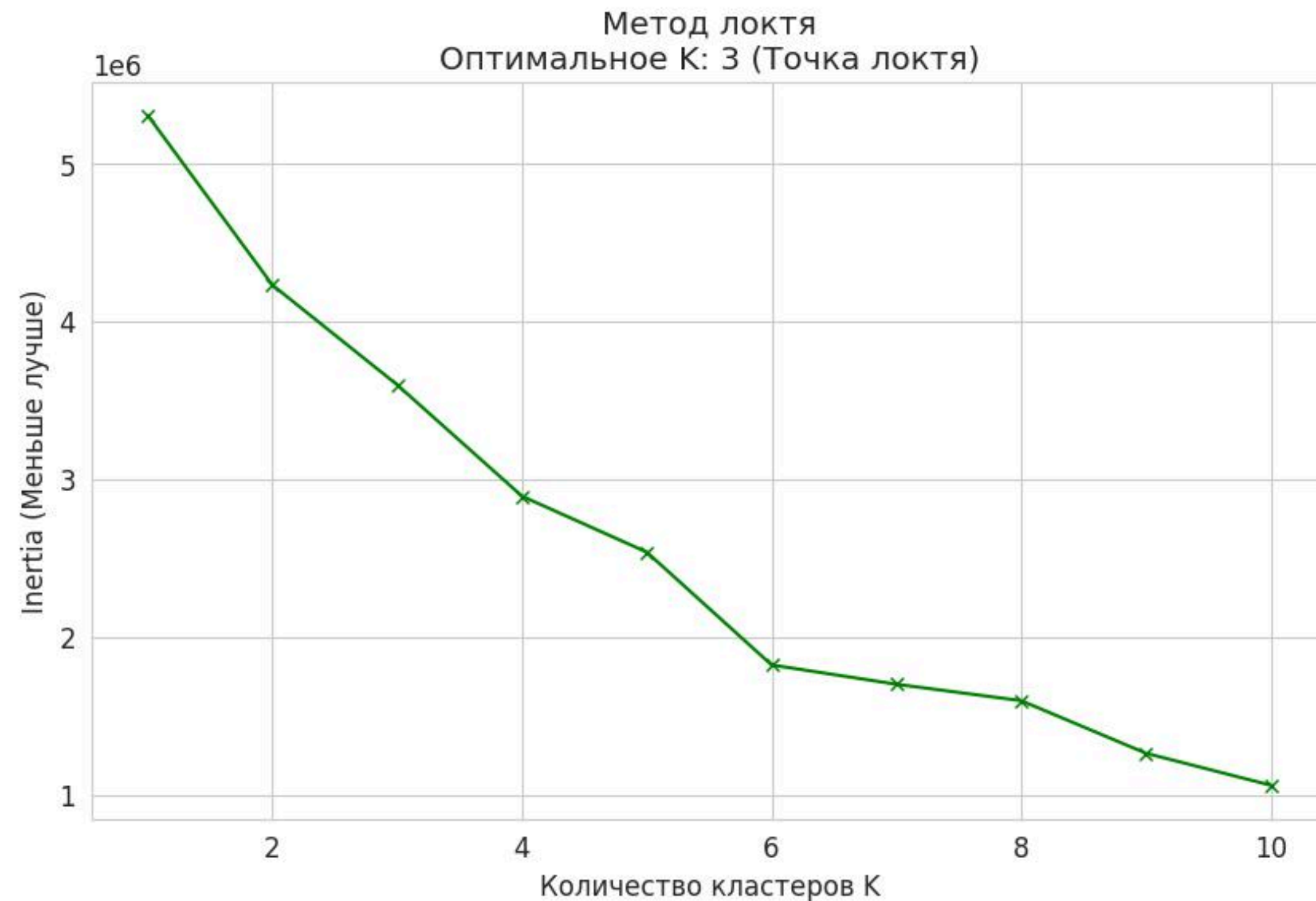
Таблица новых признаков

Customer ID	Recency	Frequency	Monetary	average_bill	preferred_category
12346.0	325	12	77556.46	6463.038333	MEDIUM CERAMIC TOP STORAGE JAR
12347.0	2	8	5633.32	704.165000	ICE CREAM SUNDAE LIP GLOSS
12348.0	75	5	2019.40	403.880000	DOUGHNUT LIP GLOSS
12349.0	18	4	4428.69	1107.172500	SWEETHEART CERAMIC TRINKET BOX
12350.0	310	1	334.40	334.400000	TEA BAG PLATE RED RETROSPOT

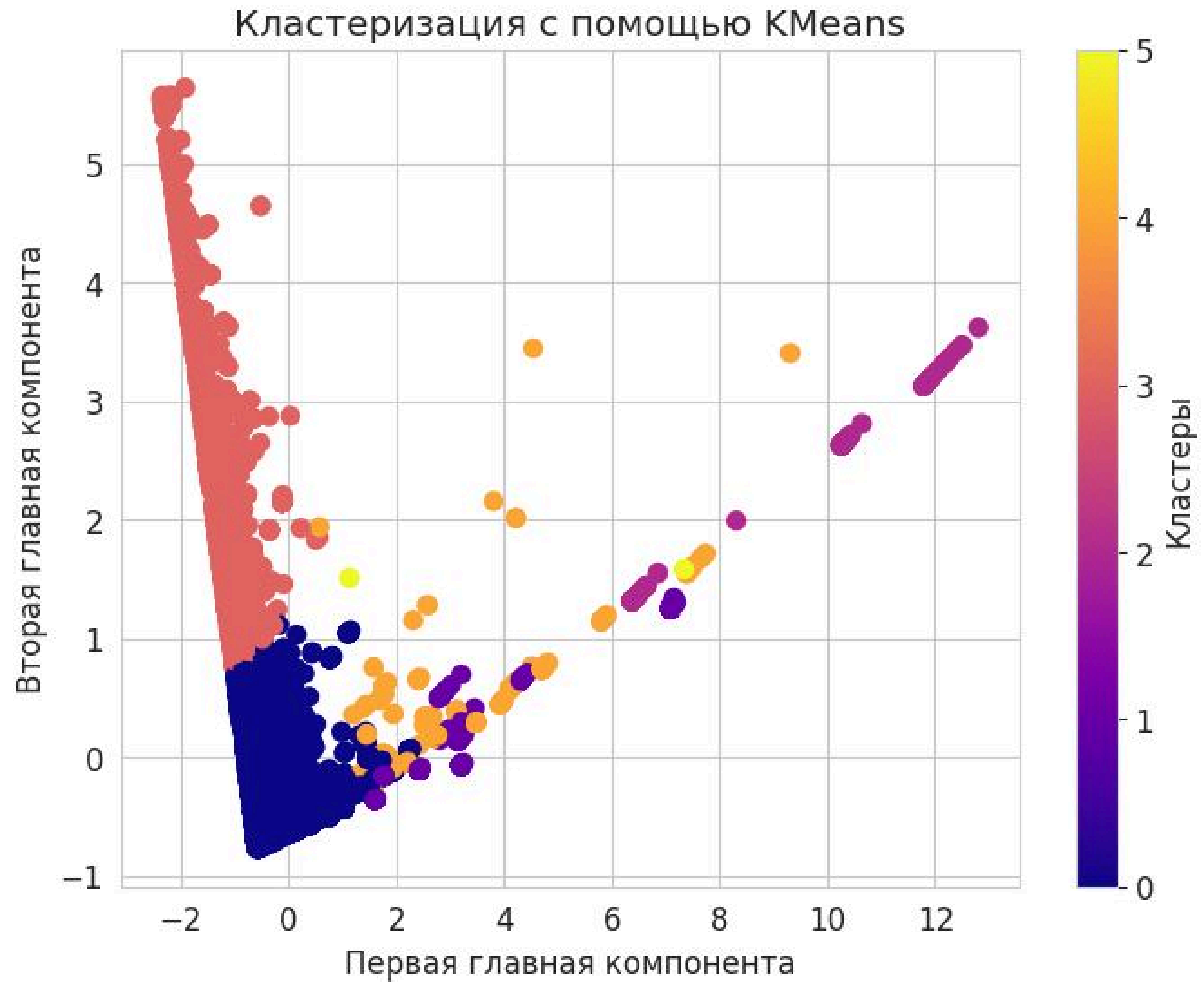


# Применение алгоритмов кластеризации

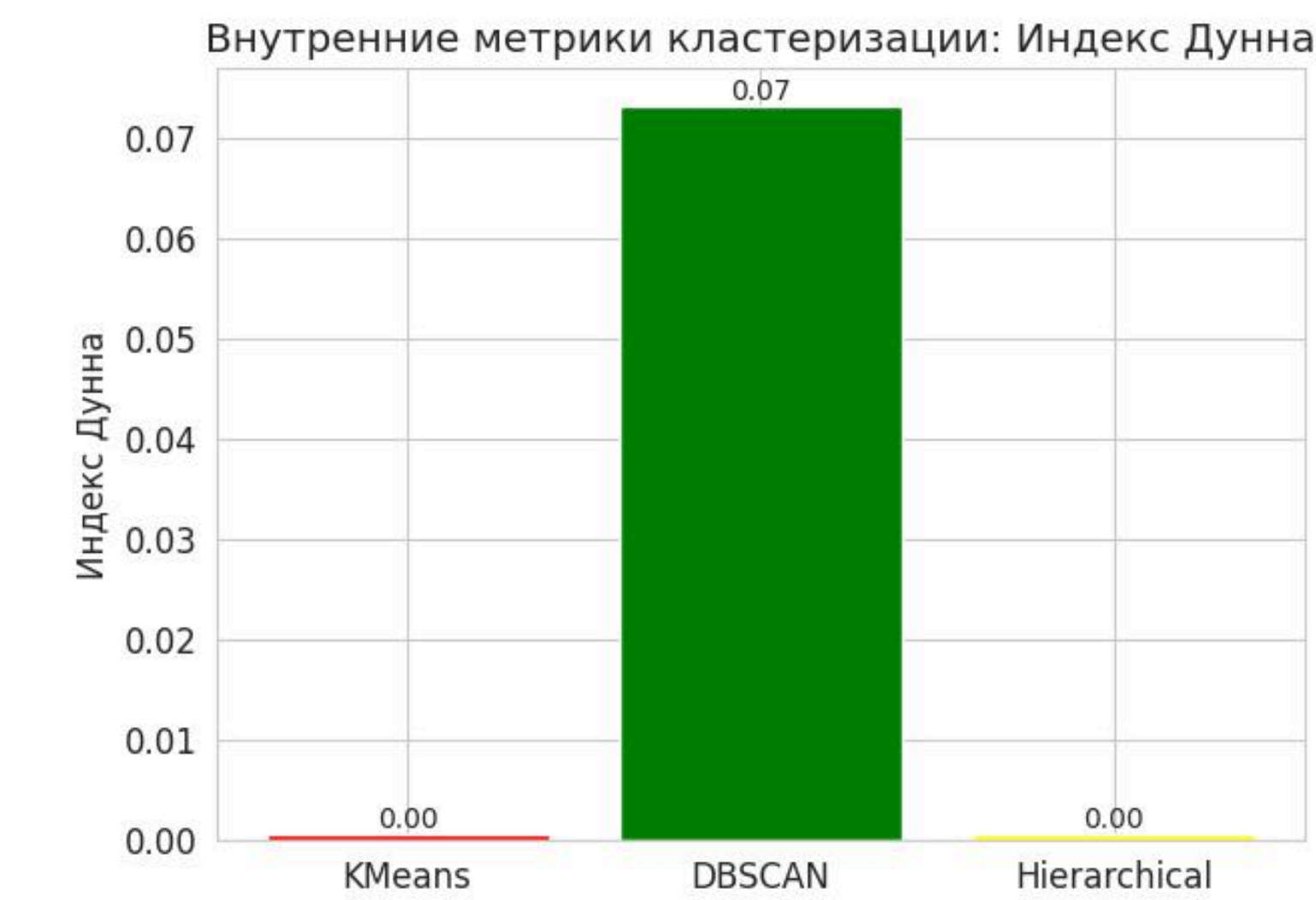
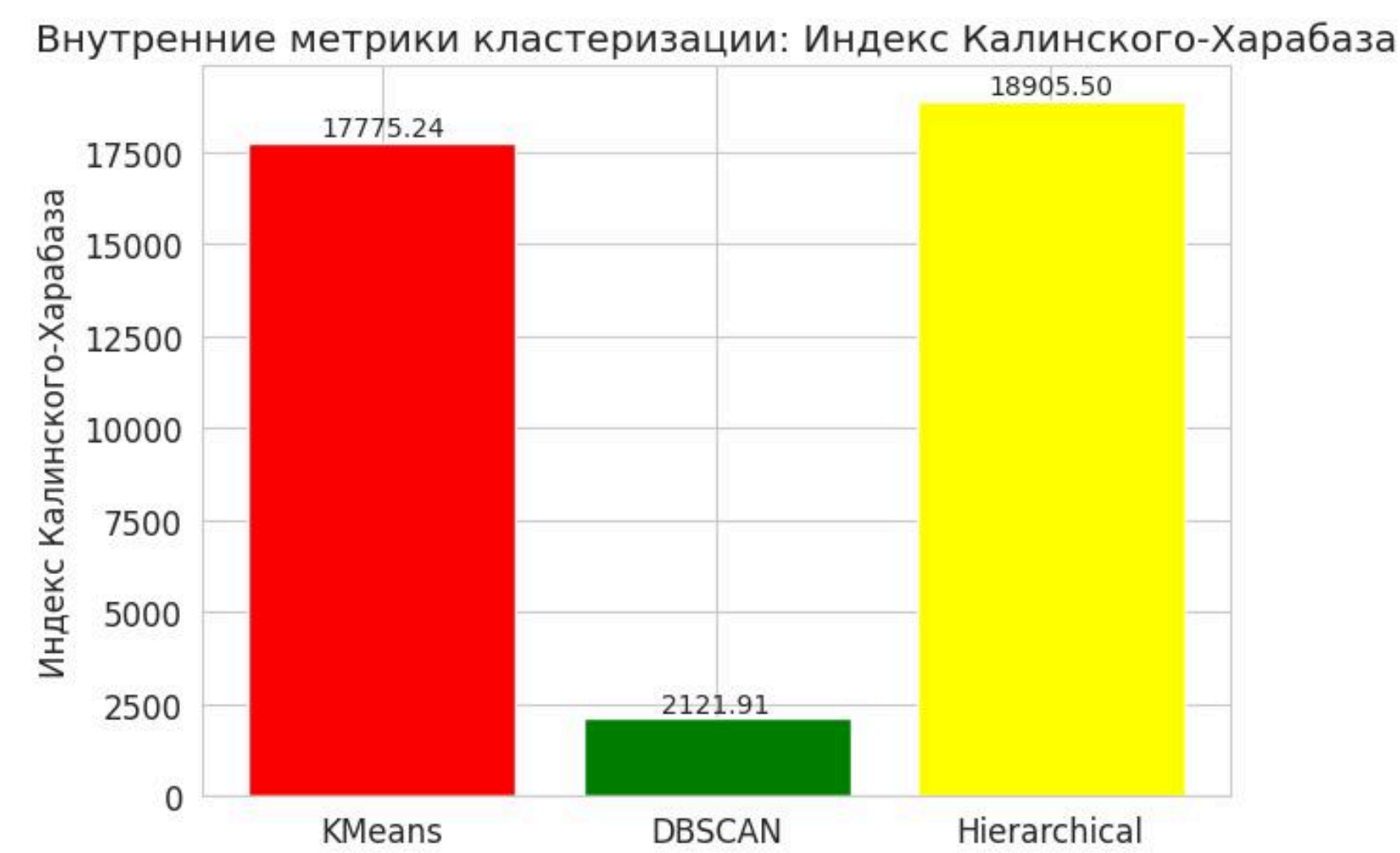
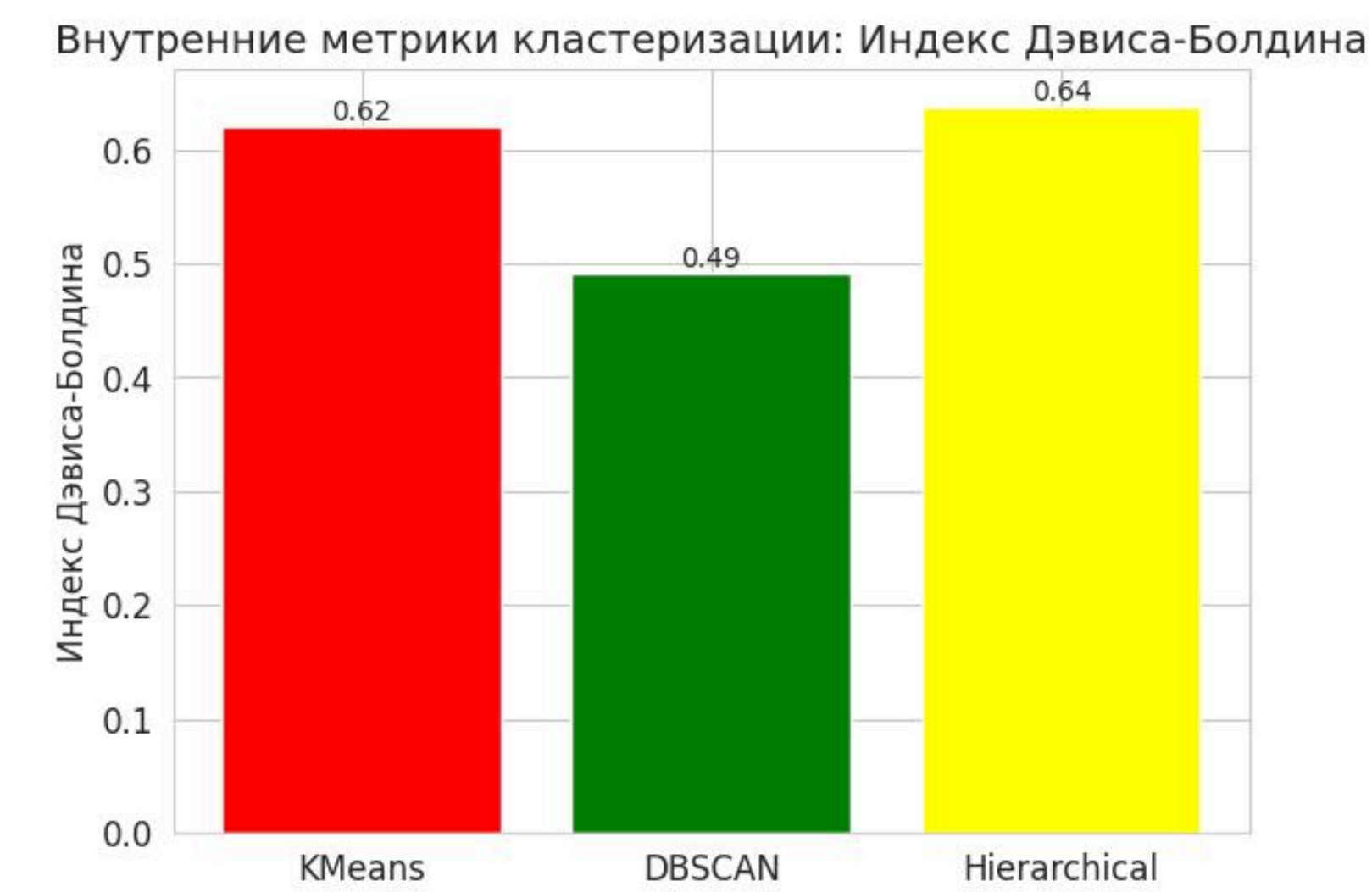
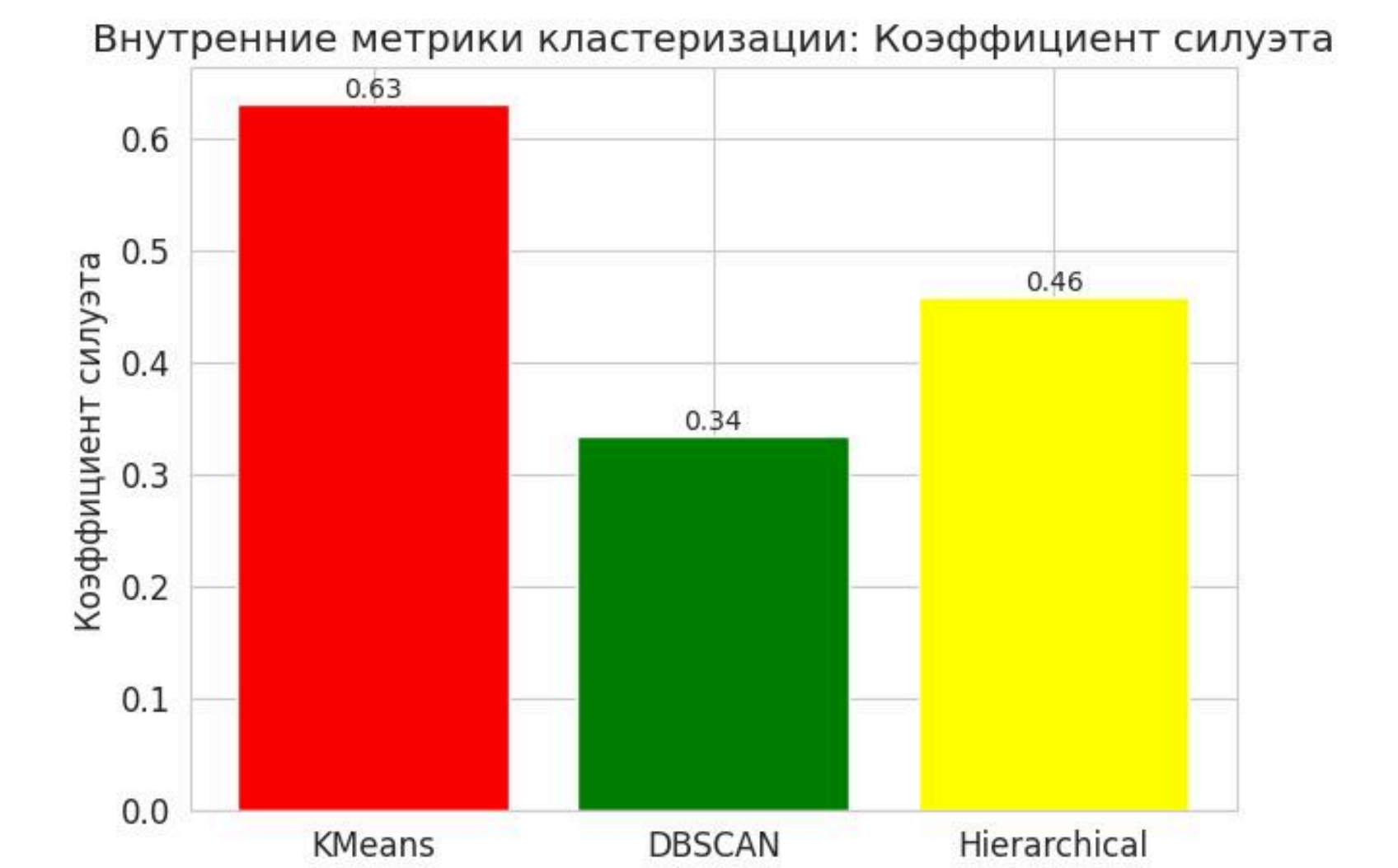
- К-средних (K-Means): Для разбиения данных на  $k$  кластеров на основе эвристического подхода.
- Иерархическая кластеризация: Для выявления вложенной структуры кластеров.
- DBSCAN и OPTICS: Для обнаружения кластеров произвольной формы и выявления выбросов.



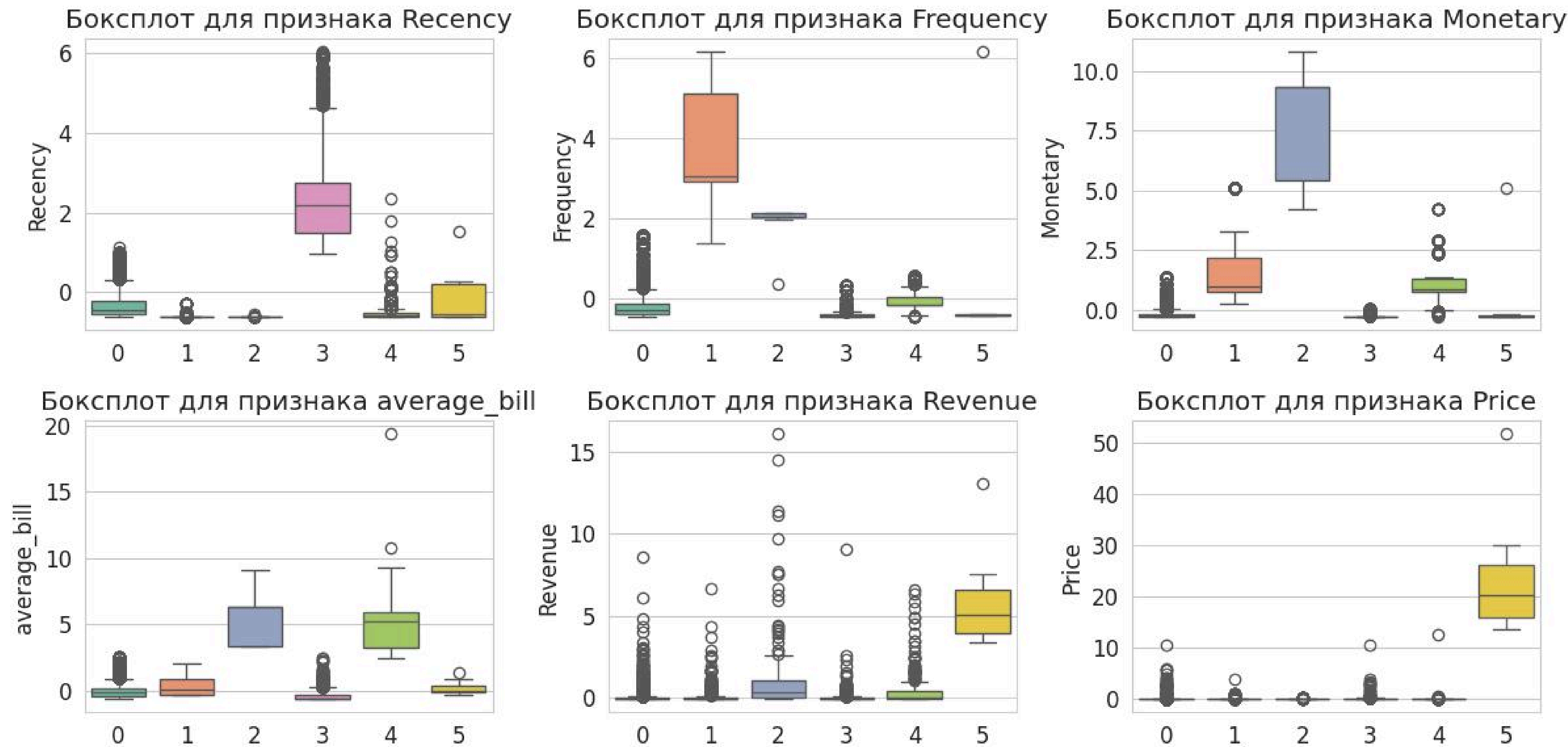
# Применение алгоритмов кластеризации



# Оценка качества кластеризации с использованием внутренних метрик



# Интерпретация и визуализация результатов



# Результаты

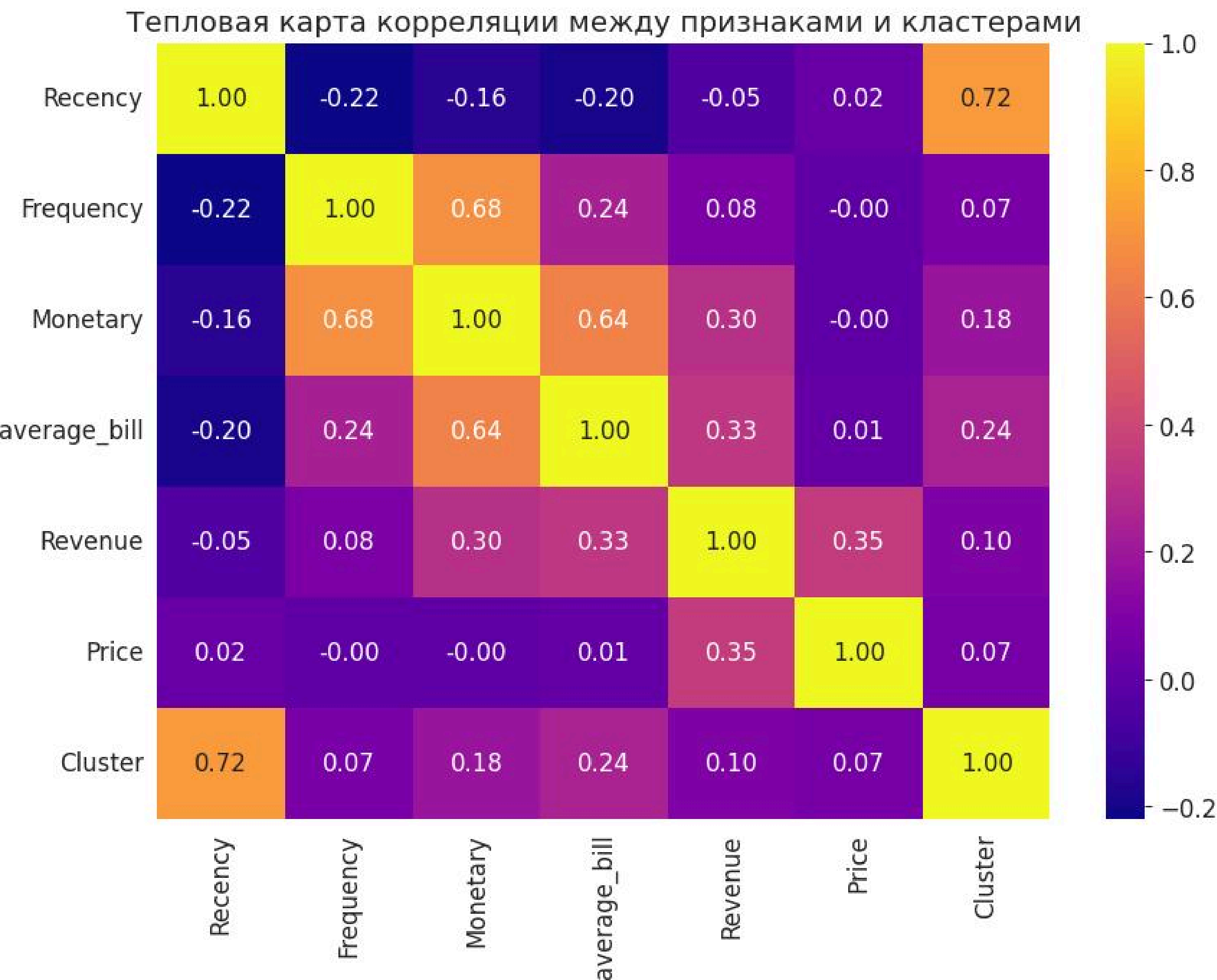
Кластер	Время с последней покупки	Частота покупок	Сумма покупок	Средний чек	Предпочитаемая категория	Топ товаров
1	34.8	16.8	6102.1	277.6	ПЛАНЕРЫ ВТОРОЙ МИРОВОЙ ВОЙНЫ ASSTD ДИЗАЙНЫ	БЕЛЫЙ ПОДВЕСНОЙ ДЕРЖАТЕЛЬ ДЛЯ СВЕТИЛЬНИКА ПОДСТАВКА ДЛЯ ТОРТА REGENCY 3 ЯРУСА УКРАШЕНИЕ В ВИДЕ ПТИЦЫ РАЗНЫХ ЦВЕТОВ
2	1.6	261.5	125187.9	462.3	ДИСКО-ШАР РОЖДЕСТВЕНСКОЕ УКРАШЕНИЕ	ПОДСТАВКА ДЛЯ ТОРТА REGENCY 3 ЯРУСА БЕЛЫЙ ПОДВЕСНОЙ ДЕРЖАТЕЛЬ ДЛЯ СВЕТИЛЬНИКА ШКАФ ДЛЯ ДЕРЕВА S/3, ОТДЕЛКА БЕЛАЯ АНТИКВАРИАТ
3	0.9	153.8	443483.9	2906.9	Ланч-бокс для девочки	СВЕТОДИОДНЫЙ НОЧНИК С ПОГАНКОЙ НАБОР ИЗ 4 КРУГЛЫХ КОРОБОК ДЛЯ ЗАКУСОК WOODLAND 20 КОЛЫШЕК ДЛЯ ТЕЛЕЖЕК RETROSPOT
4	324.1	4.9	989.8	119.3	КОЛЬЦО В ВИДЕ ЖЕВАТЕЛЬНОЙ ЖВАЧКИ ASSTD DESIGN	БЕЛЫЙ ПОДВЕСНОЙ ДЕРЖАТЕЛЬ ДЛЯ СВЕТИЛЬНИКА ПОДСТАВКА ДЛЯ ТОРТА REGENCY 3 ЯРУСА БУТЫЛКА ДЛЯ ГРЕЮЩЕЙ ВОДЫ SCOTTIE DOG
5	15.9	28.3	79949.1	2886.2	ПОМАДА-РУЧКА КРАСНАЯ	УПАКОВКА ИЗ 60 РОЗОВЫХ КОРОБОК ДЛЯ ТОРТА PAISLEY СУМКА CHARLOTTE В РОЗОВЫЙ ГОРОШЕК КОРОБКА ДЛЯ ЛАНЧА DOLLY GIRL
6	62.3	61.0	43938.4	439.4	60 КОРОБОК ДЛЯ ТОРТОВ ВИНТАЖНОЕ РОЖДЕСТВО	ПОЧТОВАЯ ОПЛАТА DOTCOM РУКОВОДСТВО

# Обсуждение

Результаты кластеризации показали, что покупатели разделились на шесть групп, каждая из которых обладает своими уникальными характеристиками. Однако при анализе важно учитывать не только особенности этих групп, но и взаимосвязи между ключевыми показателями.

Например, частота покупок напрямую влияет на общий объем расходов клиента: чем чаще совершаются покупки, тем выше общая сумма затрат. Также наблюдается зависимость между средним чеком и выручкой: рост среднего чека способствует увеличению выручки.

Наиболее значимыми факторами, влияющими на разделение клиентов по кластерам, являются давность последней покупки, частота покупок, сумма корзины и средний чек.





# Рекомендации

Как повысить эффективность каждого кластера:

1. Кластер 1 (Покупатели по интересам):  
Внедрить акции, стимулирующие повторные покупки, и разработать программу лояльности. Например, предложить бонусы или скидки за каждую последующую покупку, чтобы увеличить частоту взаимодействия с клиентами.
2. Кластер 2 (Среднеэффективные покупатели):  
Для клиентов, совершающих дорогие или специальные покупки, предложить эксклюзивные товары или услуги повышенного качества. Настроить рекламные кампании на привлечение этого сегмента, чтобы увеличить средний чек и повысить их вовлечённость.
3. Кластер 3 (Неэффективные покупатели):  
Разработать целевые скидки и персонализированные рассылки с предложениями, которые могут заинтересовать эту группу. Стимулировать покупки через ограниченные по времени акции или специальные условия, чтобы повысить их активность.
4. Кластер 4 (Сезонные покупатели):  
Создать сезонные предложения и акции, учитывающие интересы этой группы. Учитывать их предпочтения в определённое время года, чтобы повысить их вовлечённость и частоту покупок.
5. Кластер 5 (Высокоэффективные покупатели):  
Предоставить специальные привилегии, такие как доступ к эксклюзивным продуктам или услугам. Поддерживать высокий уровень обслуживания и предлагать индивидуальные условия для сохранения их лояльности.
6. Кластер 6 (Лояльные покупатели):  
Разрабатывать персонализированные предложения, соответствующие их интересам. Предлагать новые товары или услуги, которые могут их заинтересовать, чтобы поддерживать их вовлечённость и удовлетворённость.

# Заключение

Таким образом, анализ и сегментация клиентов с использованием алгоритмов кластеризации позволяют глубже изучить поведение покупателей, выделить ключевые группы с разным уровнем вовлечённости, интересами и потребностями. Такой подход способствует разработке стратегий, направленных на повышение эффективности продаж, привлечение новых клиентов и расширение ассортимента. Учёт взаимосвязей между ключевыми показателями, такими как частота покупок, средний чек и выручка, помогает увеличить доходы компании и обеспечить устойчивый спрос на её товары и услуги.