



МИИГАиК

МОСКОВСКИЙ ГОСУДАРСТВЕННЫЙ
УНИВЕРСИТЕТ ГЕОДЕЗИИ И КАРТОГРАФИИ

Анализ и сегментация клиентов с помощью алгоритмов кластеризации

Подготовил: студент
группы 2023-ФГиИБ-ПИ-1м
Грищенко Н.Н.

Цель

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.

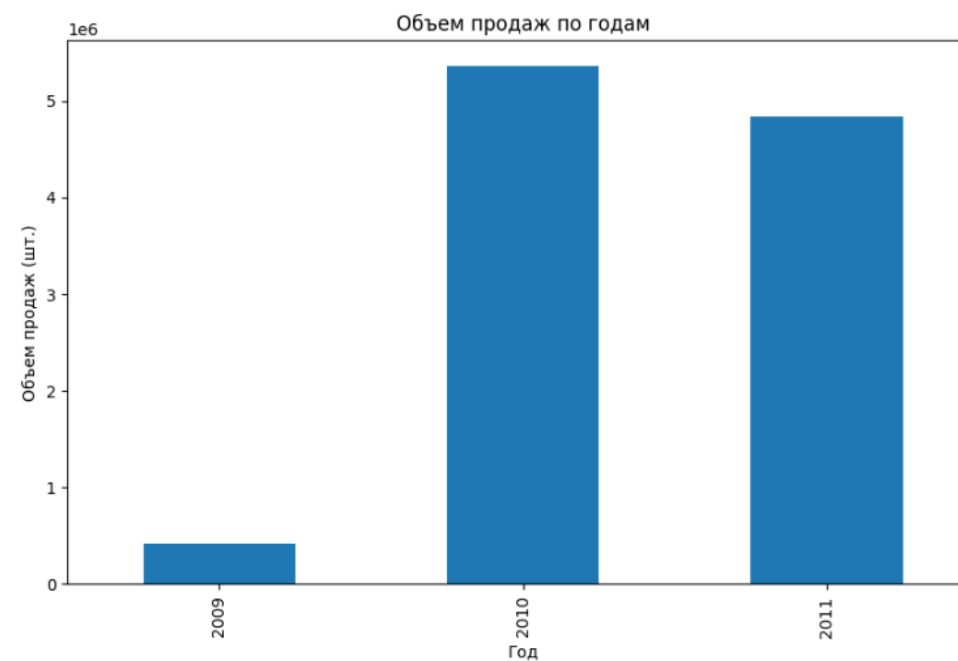
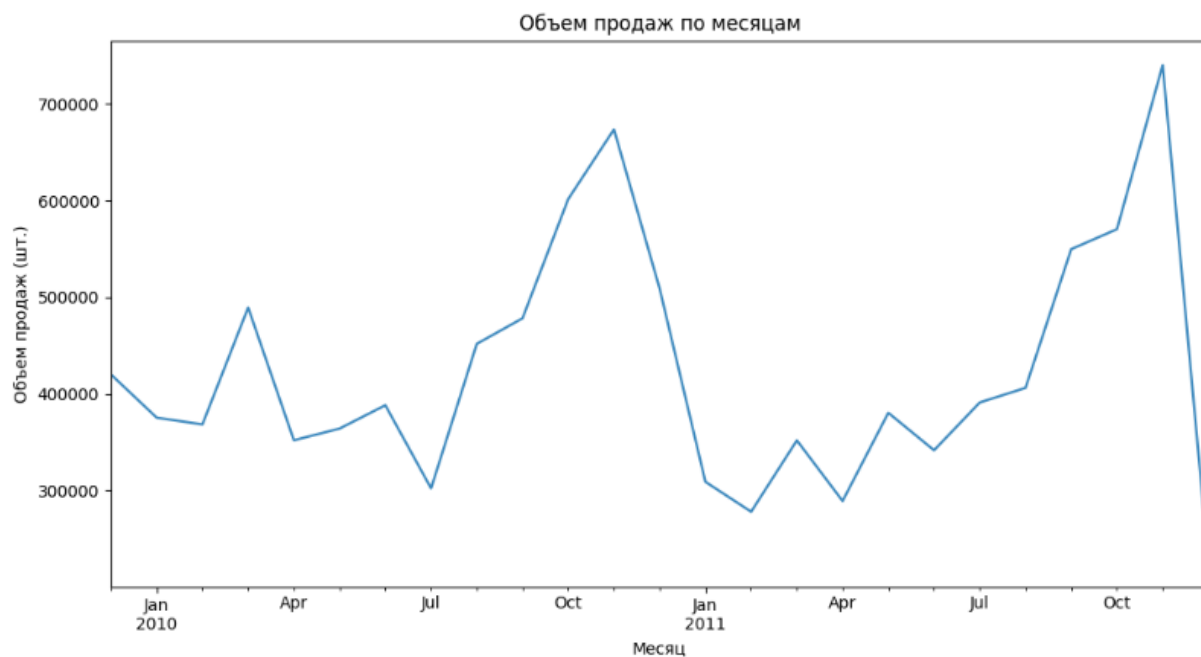
Customer ID

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

Country

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

Первичный анализ данных



Первичный анализ данных

| ID клиента | Количество проданных товаров (шт.) |
|------------|------------------------------------|
| 17841 | 13097 |
| 14911 | 11613 |
| 12748 | 7307 |
| 14606 | 6709 |
| 14096 | 5128 |
| 15311 | 4717 |
| 14156 | 4130 |
| 14646 | 3890 |
| 13089 | 3438 |
| 16549 | 3255 |

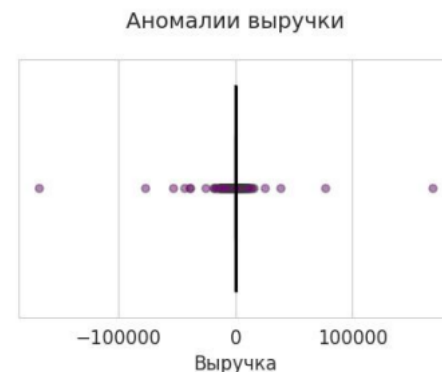
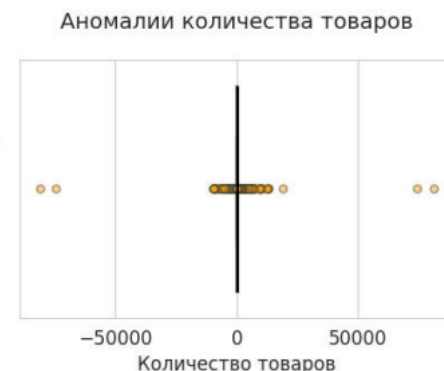
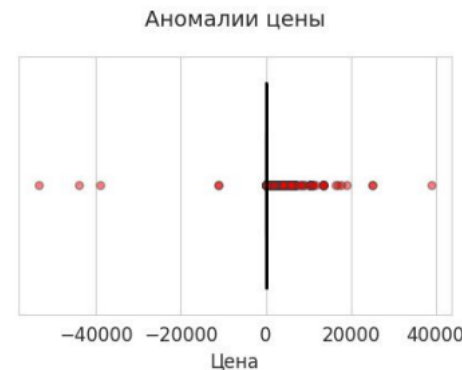
| Наименование товара | Выручка |
|---|------------|
| ПОДСТАВКА ДЛЯ ТОРТА 3-ЯРУСНАЯ "REGENCY" | 327 813,65 |
| ПОЧТОВЫЕ РАСХОДЫ "DOTCOM" | 322 647,47 |
| ПОДВЕСНОЙ СВЕЧНИК "БЕЛОЕ СЕРДЦЕ" | 257 533,90 |
| БОЛЬШАЯ КРАСНАЯ СУМКА "RETROSPOT" | 148 800,64 |
| ГИРЛЯНДА "PARTY BUNTING" | 147 948,50 |
| РАЗНОЦВЕТНЫЕ ПТИЧКИ-УКРАШЕНИЯ | 131 413,85 |
| НАБОР ДЛЯ ГИРЛЯНДЫ "Рождество 50-х" | 121 662,14 |
| ПОЧТОВЫЕ РАСХОДЫ | 112 341,00 |
| ГИРЛЯНДА "ОГНЕННЫЙ ПЕРЕЦ" | 84 854,16 |
| ВРАЩАЮЩИЕСЯ СЕРЕБРЯНЫЕ АНГЕЛЫ-СВЕЧНИКИ | 73 814,72 |

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

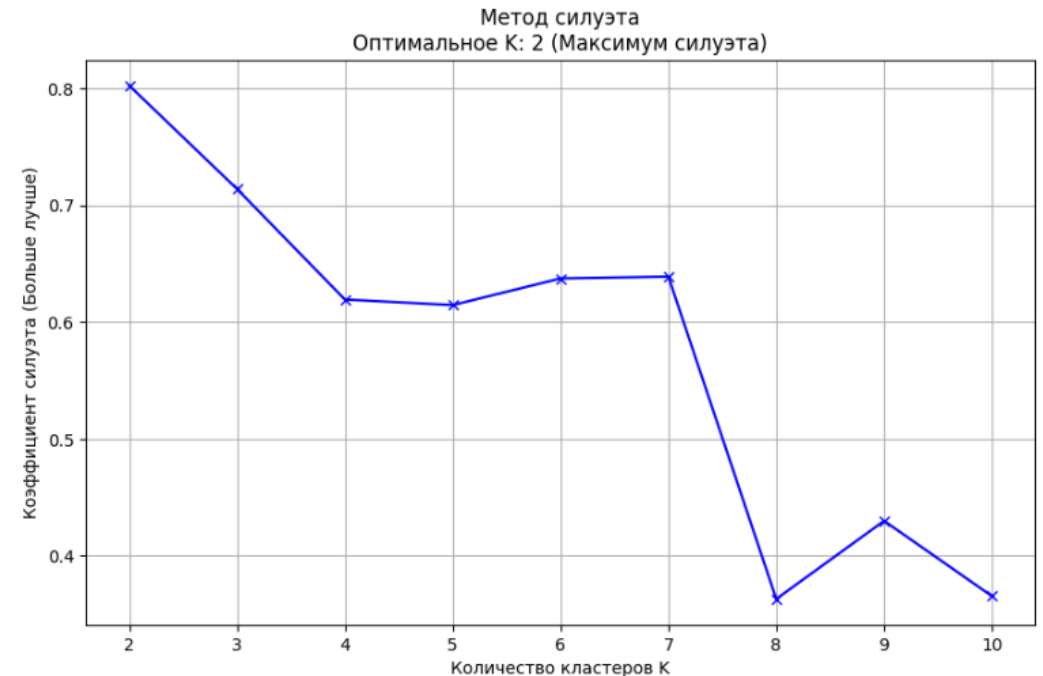
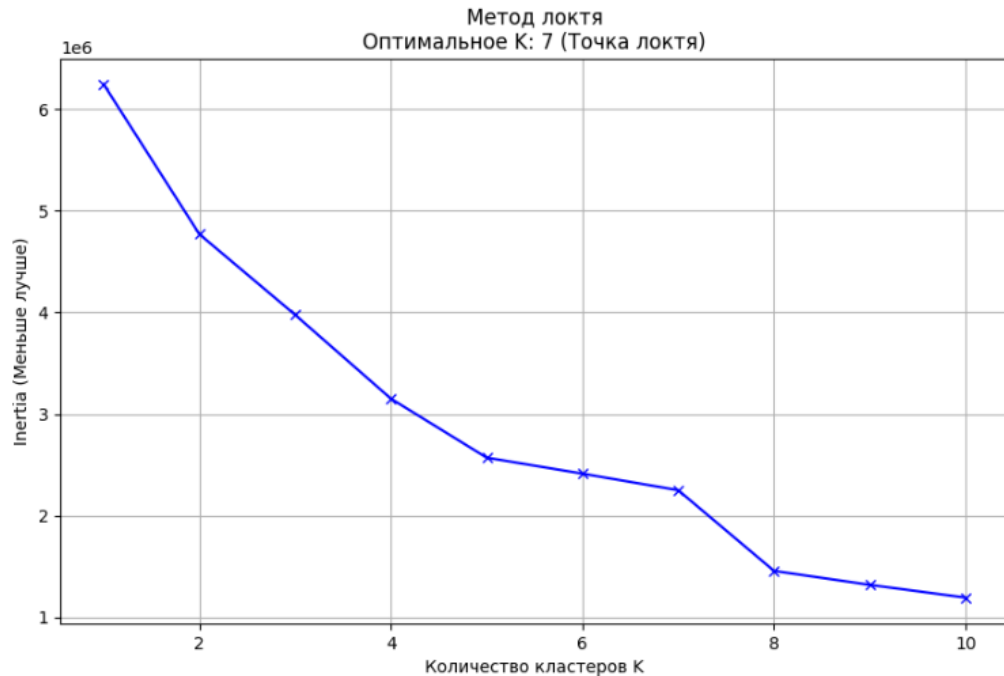
Создание новых признаков:

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

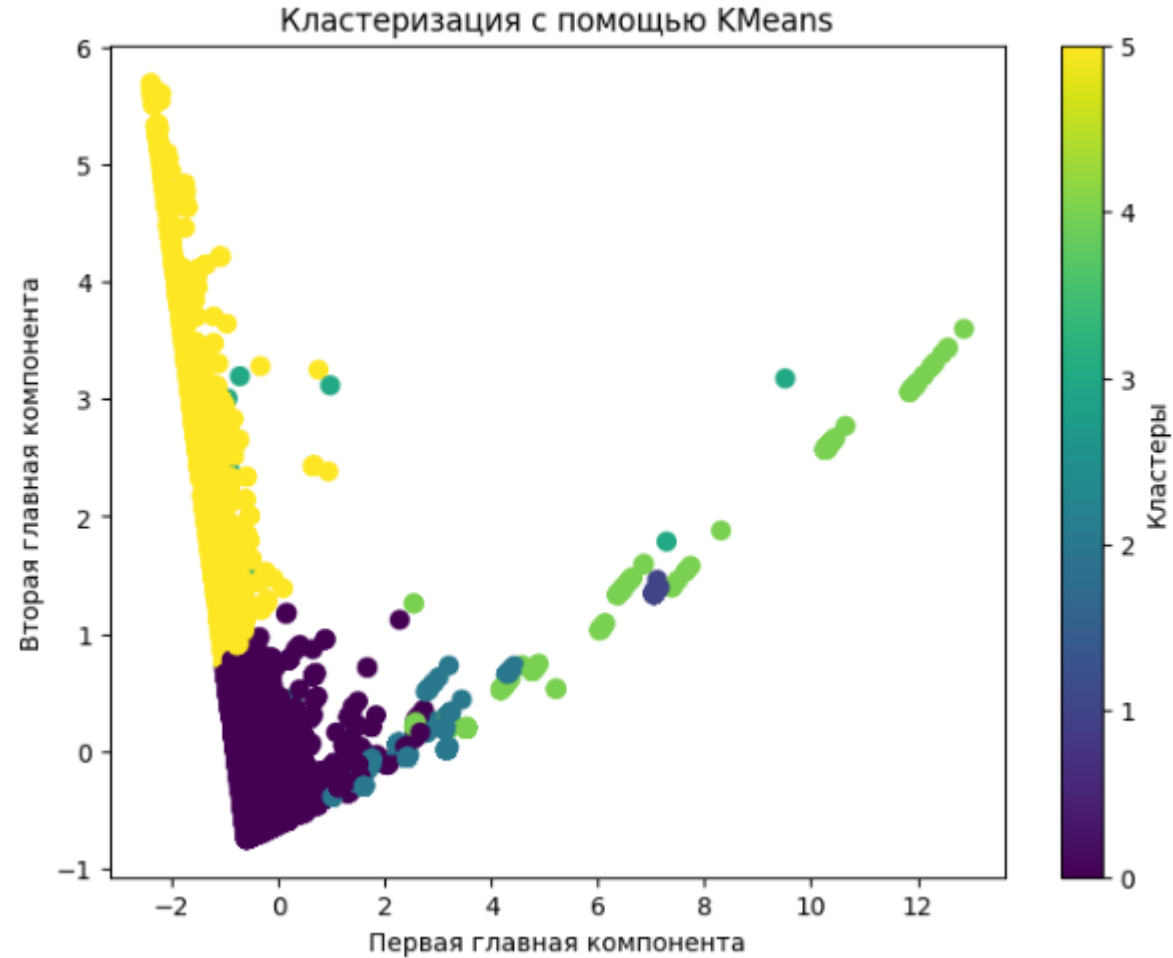
| Customer ID | Recency | Frequency | Monetary | average_bill | preferred_category |
|-------------|---------|-----------|----------|--------------|--|
| 10000.0 | 28 | 7 | 87.77 | 12.54 | Набор для гирлянды "Винтажное Рождество" |
| 10001.0 | 11 | 1 | 0.83 | 0.83 | Набор 12 детских цветных мелков |
| 10002.0 | 184 | 4 | 9.91 | 2.48 | Лента для ограждения "День Рождения" |
| 10003.0 | 462 | 2 | 18.37 | 9.19 | Губка для ванны "Арбуз" |
| 10004.0 | 3 | 5 | 9.93 | 1.99 | Чашка в синий горошек |
| 10005.0 | 114 | 4 | 37.20 | 9.30 | Коврик для двери "Красные точки" |

Применение алгоритмов кластеризации

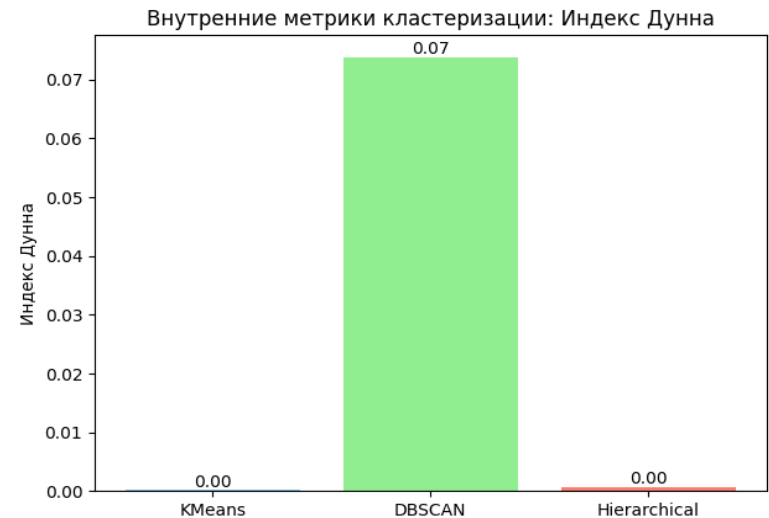
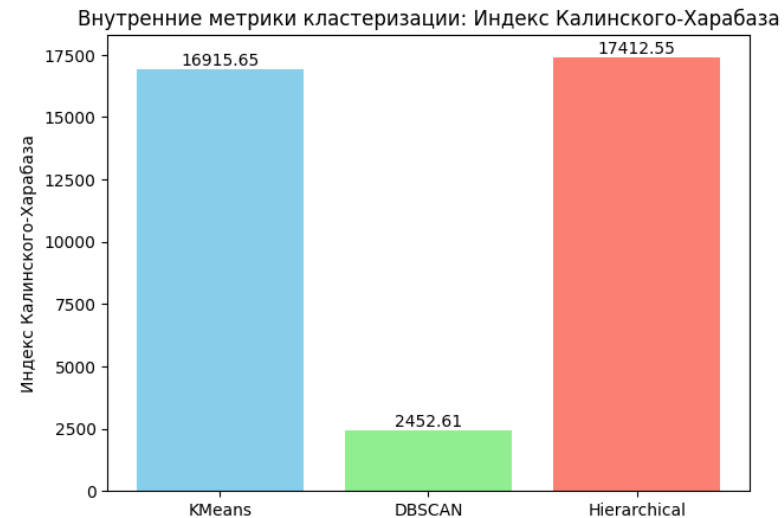
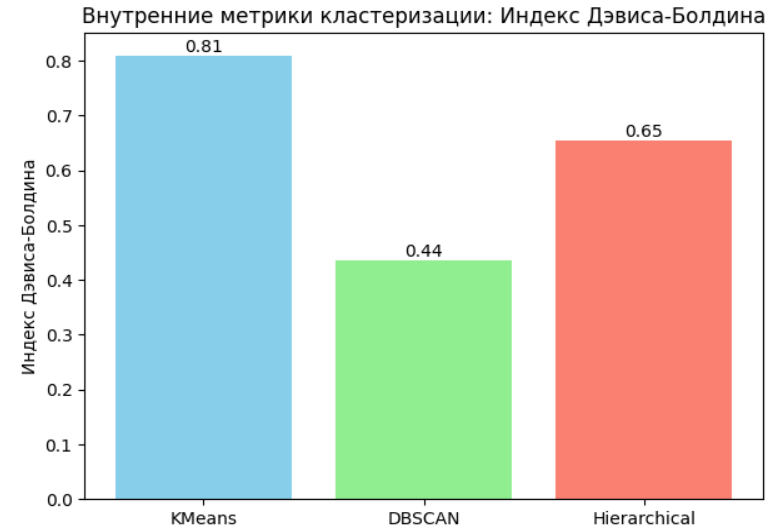
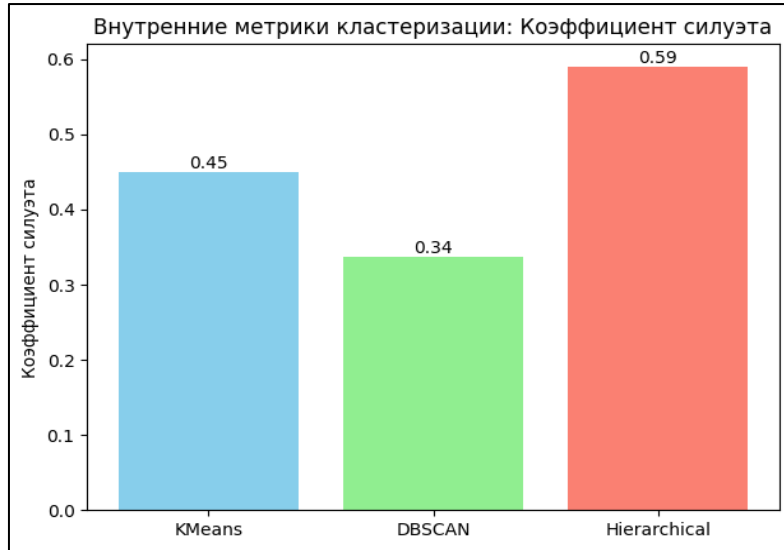
- К-средних (K-Means): Для разбиения данных на k кластеров на основе эвристического подхода.
- Иерархическая кластеризация: Для выявления вложенной структуры кластеров.
- DBSCAN и OPTICS: Для обнаружения кластеров произвольной формы и выявления выбросов.



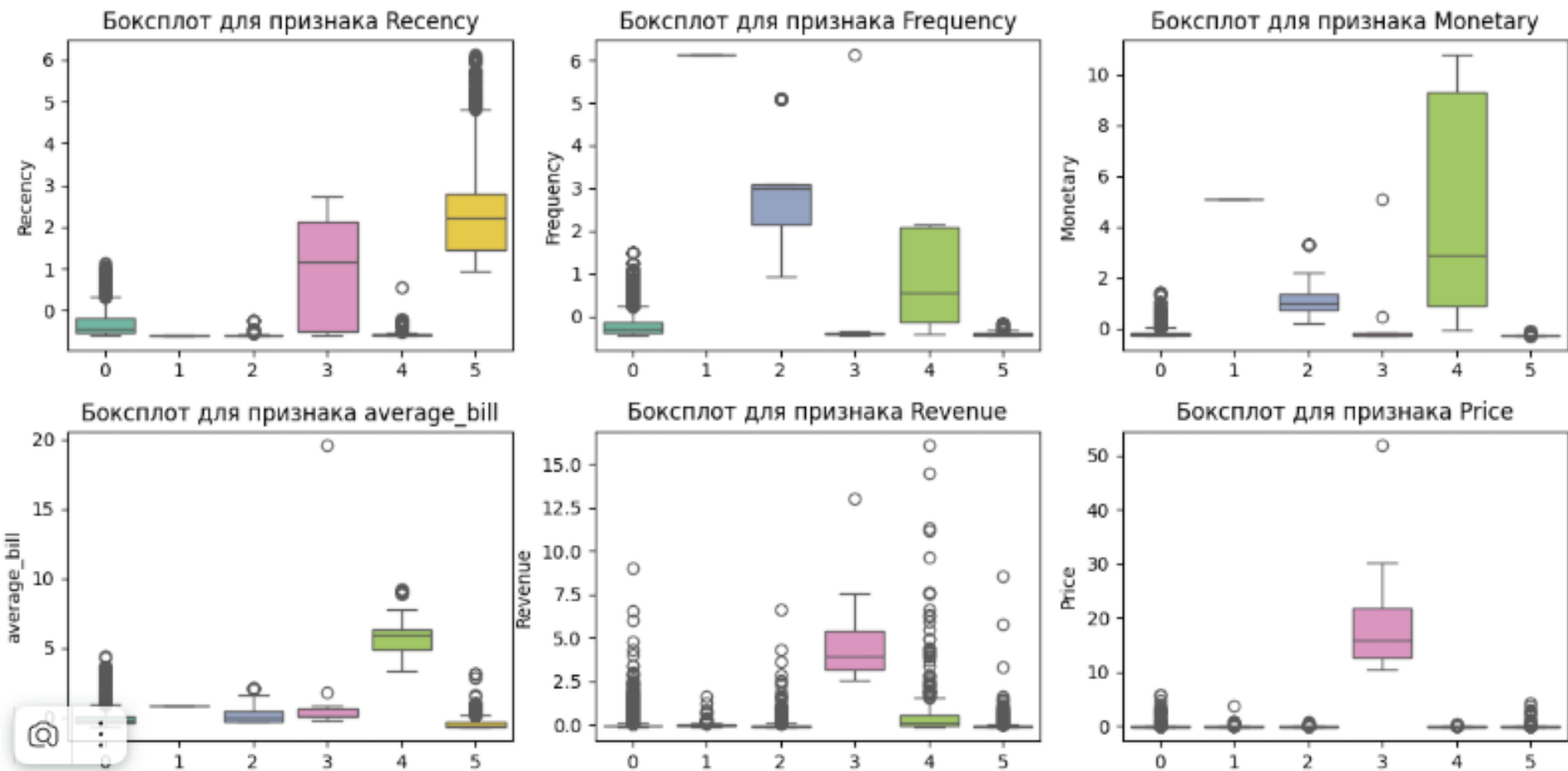
Применение алгоритмов кластеризации



Оценка качества кластеризации с использованием внутренних метрик



Интерпретация и визуализация результатов



Результаты

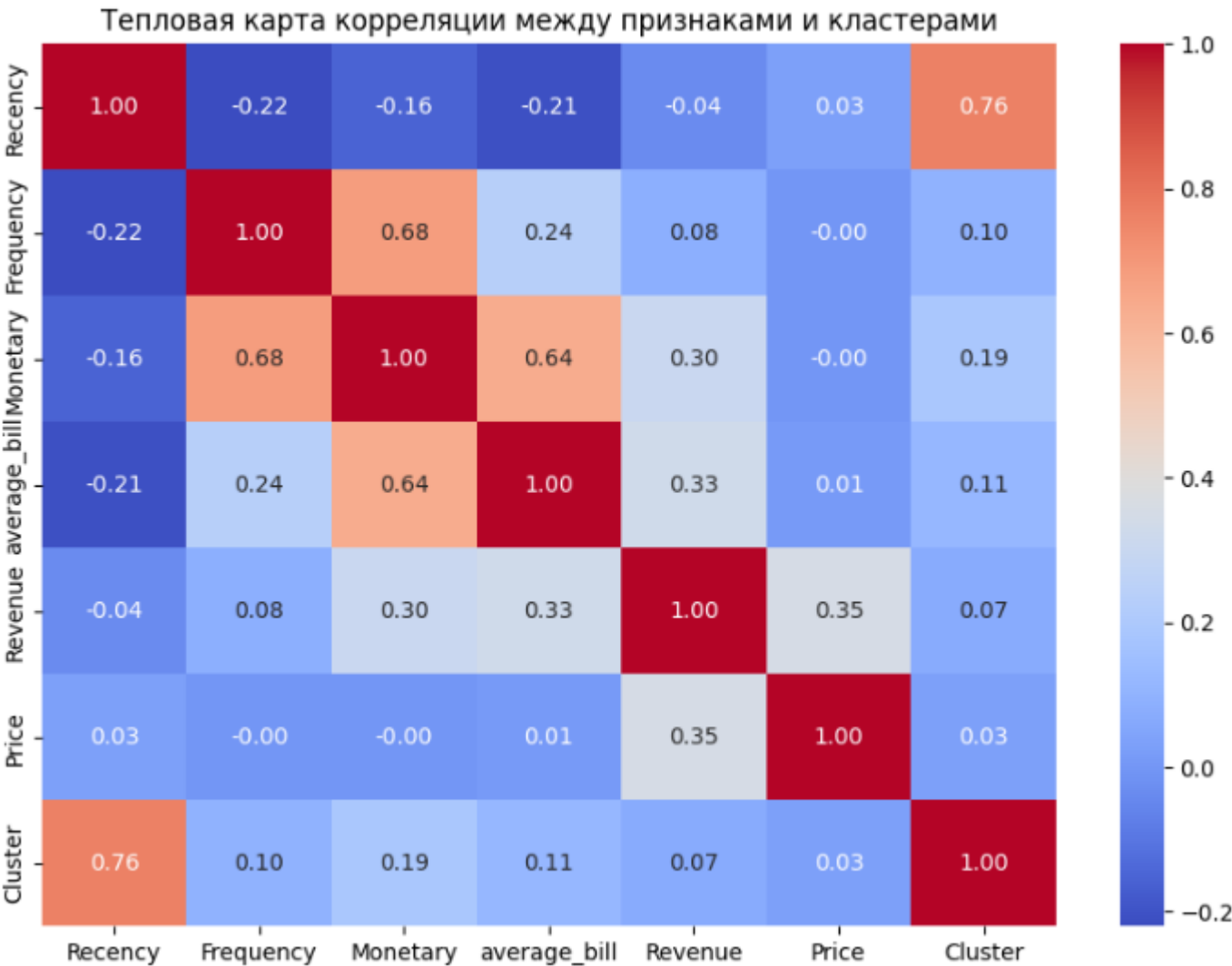
| Кластер | Время с последней покупки (дни) | Частота покупок | Сумма покупок | Средний чек | Предпочитаемая категория | Топ товаров |
|---------|---------------------------------|-----------------|---------------|-------------|---|---|
| 1 | 34.8 | 16.8 | 6 102.1 | 277.6 | Планеры Второй Мировой войны | Белый подвесной держатель для светильника Подставка для торта Regency 3 яруса Украшение в виде птицы разных цветов |
| 2 | 1.6 | 261.5 | 125 187.9 | 462.3 | Диско-шар рождественское украшение | Подставка для торта Regency 3 яруса Белый подвесной держатель для светильника Шкаф для дерева S/3, отделка белая антиквар |
| 3 | 0.9 | 153.8 | 443 483.9 | 2 906.9 | Ланч-бокс для девочки | Светодиодный ночник с поганкой Набор из 4 круглых коробок для закусок Woodland 20 колышек для тележек Retrosport |
| 4 | 324.1 | 4.9 | 989.8 | 119.3 | Кольцо в виде жевательной жвачки | Белый подвесной держатель для светильника Подставка для торта Regency 3 яруса Бутылка для греющей воды Scottie Dog |
| 5 | 15.9 | 28.3 | 79 949.1 | 2 886.2 | 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 | 43 938.4 | 439.4 | 60 коробок для тортов "Винтажное Рождество" | Почтовые расходы 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.



Примечание:
Recency – давность, average_bill – средний чек, Frequency – частота, Revenue – выручка, Monetary – сумма корзины, Price – цена

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.