

Маркетинг, продажи и клиенты: аналитика в действии

Курс: ***Data Analyst Junior, Skillbox***

Автор проекта: ***Лебедева Евгения***



1. Обработка данных

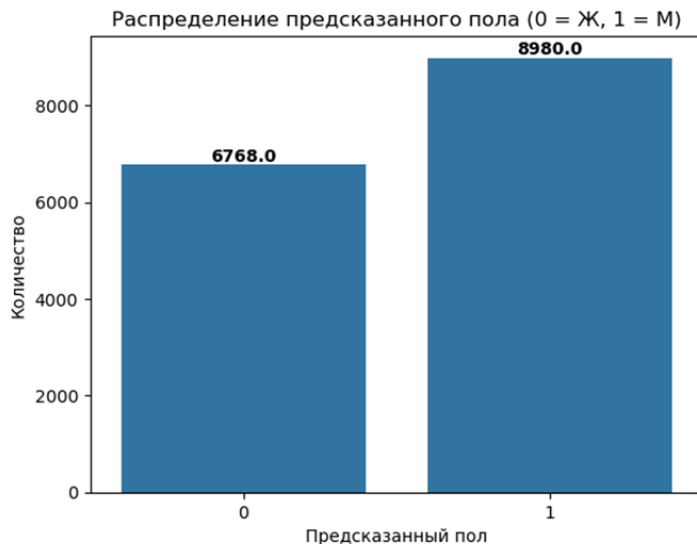
- Данные объединены из SQL и CSV
- Отфильтрованы клиенты страны с кодом 32
- Очистка: пропуски, цвета, названия товаров

2. Восстановление пола клиента

Количество потерянных клиентов = 15748 человек

Модель: RandomForestClassifier

- Значимые признаки: возраст, город, образование, personal_coef



3. Результаты A/B-теста кампании

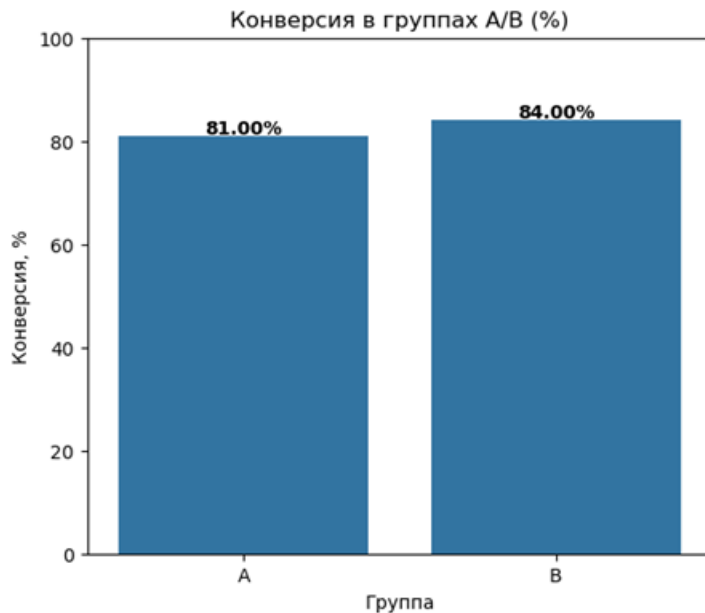
- Объединили группы: позитив (А) и негатив (В) = 10 044 чел.

Группа А – была скидка,
Группа В – не было скидки

H_0 (нулевая гипотеза):
Конверсии в группах А и В равны;
 H_1 (альтернатива):
Конверсии различаются.

Z-статистика: -3.6
P-value: < 0.0001

Разница статистически значима — компания А показала результат хуже, чем контрольная группа



Вывод:

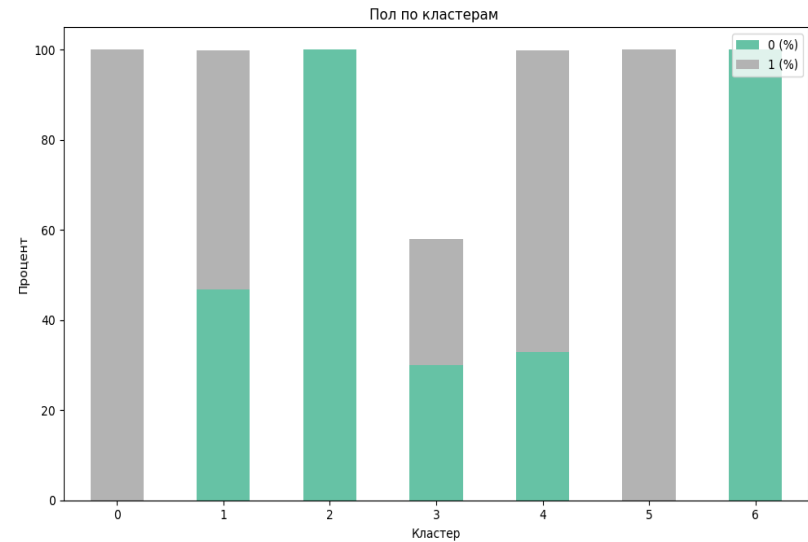
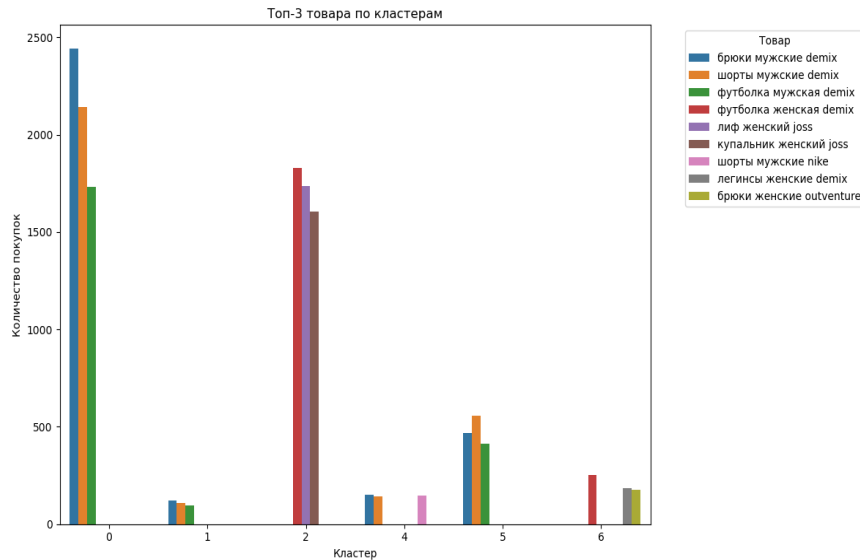
- Предоставление персональной скидки не повысило, а снизило вероятность покупки.
- С высокой долей вероятности — скидка воспринималась как лишняя, неуместная или подозрительная.
- Кампания оказалась неэффективной и, возможно, даже вредной для выручки.

Рекомендации:

- Остановить повторение email-скидок в текущем формате.
- Протестировать другие виды стимулов: Бонусы, подарки, лотереи;
- Информирование, а не навязывание скидки.
- Провести кластеризацию клиентов: Выявить группы, которым действительно нужны скидки; Предлагать персонализированные стимулы по сегментам.

4. Распределение по Кластерам:

Набор признаков: возраст, пол, образование, город, персональный коэффициент.
Методом Kmeans нашли подходящее число кластеров = 7 (max silhouette_score)

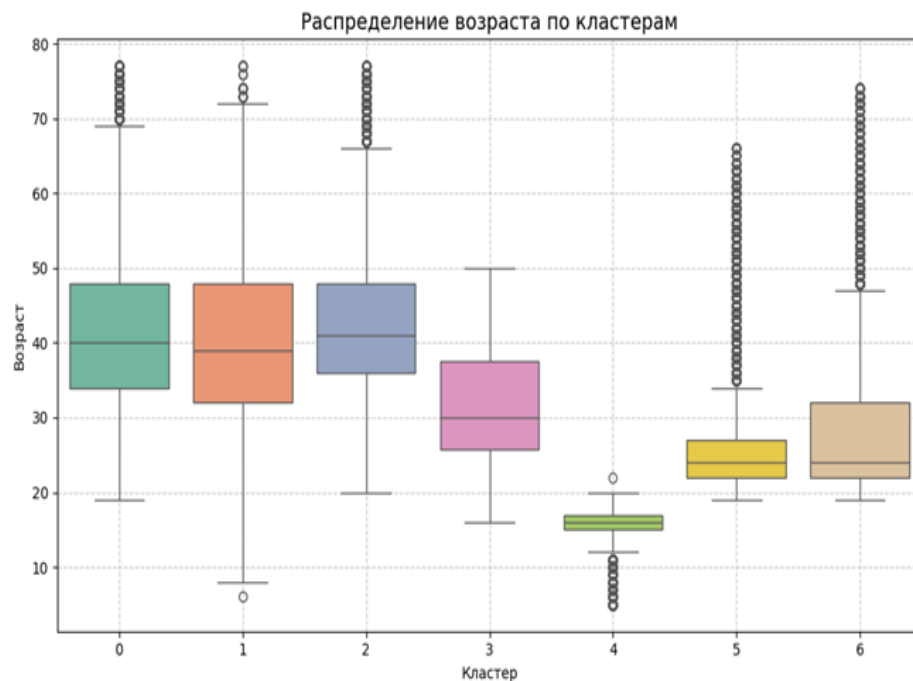
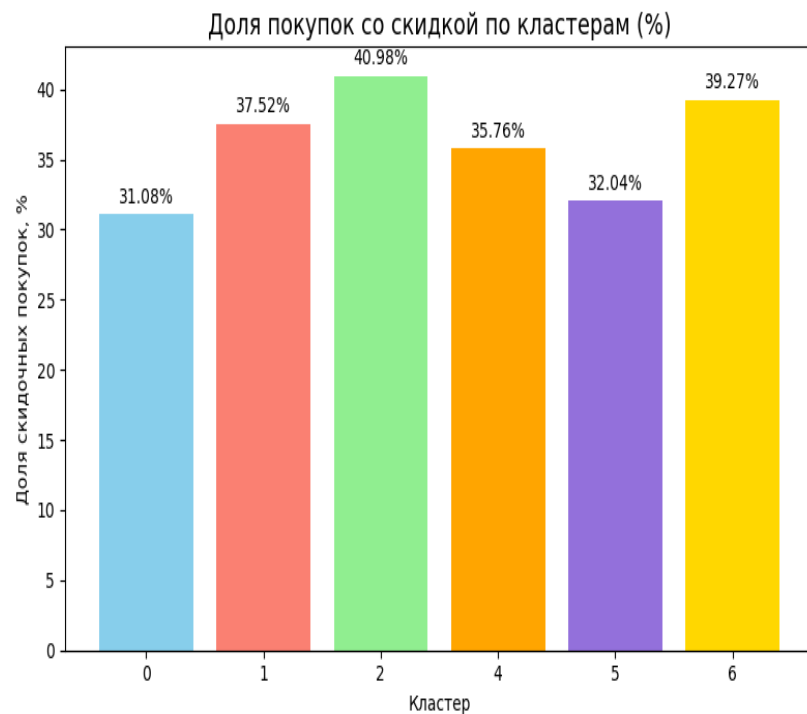
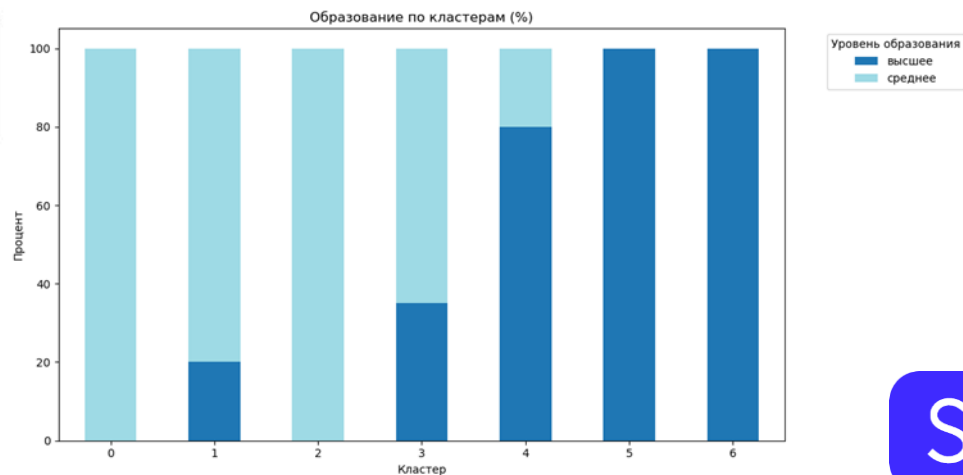
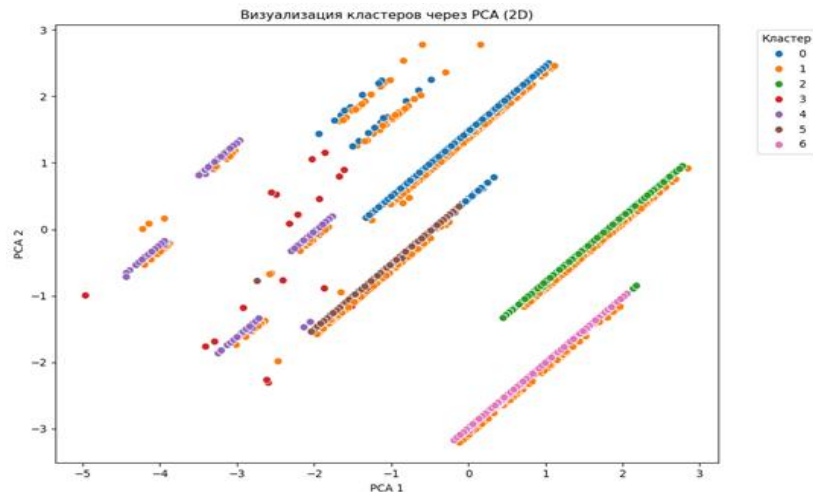


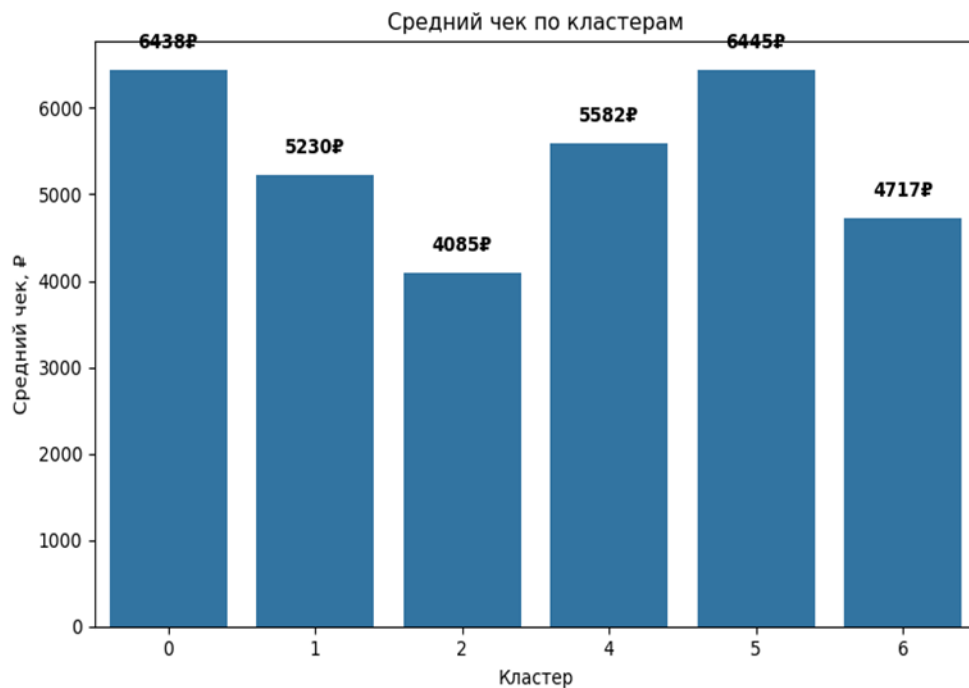
Сравнивая набор товаров и распределение полов то мы видим, что:

- ✓ Кластеры 0 и 5 — полностью мужские,
- ✓ Кластеры 2 и 6 — полностью женские,
- ✓ Кластер 1 — наиболее сбалансированный,
- ✓ Кластер 4 — с преобладанием мужчин.



Описание 7 кластеров





Кластер	Количество клиентов	Процент от общего числа
0	274563	40.35
2	247430	36.36
5	68608	10.08
4	36445	5.36
6	34302	5.04
1	19045	2.80
3	20	0.00

Рекомендации:

- **Фокус на кластеры 0, 2 и 5:** они дают основной объём продаж.
- **Рост выручки** возможен за счёт увеличения среднего чека в кластере 2.
- **Кастомизированные стратегии** по каждому кластеру помогут повысить эффективность маркетинга



Сравнение кластеров: поведение, скидки и рекомендации

Кластер	Целевая аудитория	Поведение	Топ-3 товара	Скидки	Средний чек	Рекомендации
0	Мужчины ~35 лет, среднее образование	Рациональны, покупают без акций	брюки, шорты, футболки Demix	~31%	6 438 ₽	Ремаркетинг, уведомления о новинках
1	Молодые (28–32), сбалансированный пол	Ориентированы на скидки	брюки, шорты, футболки Demix	~37.5%	5 229 ₽	Тестирования бюджетных предложений и push-рассылок
2	Женщины 38+, среднее образование	Одежда для отдыха, акции важны	футболки, купальники Joss/Demix	~41%	4 084 ₽	Офлайн-акции, рассылки
3	Малочисленный сегмент	Поведение не выявлено	-	-	-	Требует анализа
4	Мужчины и женщины ~33, высшее образование	Предпочтение брендам	брюки, шорты Nike и Demix	~36%	5 582 ₽	Персональные предложения
5	Мужчины ~40+, высшее образование	Меньше чувств. к скидкам	шорты, брюки, футболки Demix	~32%	6 445 ₽	YouTube, Telegram
6	Женщины 30–40 лет	Спорт, чувств. к акциям	футболка, легинсы Demix	~39%	4 716 ₽	Инфлюенсеры, подборки

5. Модель склонности к покупке

Самостоятельно сформировали отрицательные примеры непокупок
Обучили модель – RandomForestClassifier

Метрика	0 — не покупка	1 — покупка
Precision	0.80	0.58
Recall	0.78	0.61
F1-score	0.79	0.59
Accuracy: 0.72		

Интерпретация:

- Модель достаточно хорошо распознаёт тех, кто не покупает товар (класс 0) — точность 80%.
- Для класса 1 (тех, кто покупает товар) модель работает удовлетворительно, но не идеально:
- F1-score = 0.59 — это компромисс между точностью и полнотой.
- Recall (60%) — модель ловит 60% потенциальных покупателей, что может быть приемлемо для маркетинга.

Общая точность — 72%, что для задачи рекомендации товаров — вполне рабочий результат.

Выводы и рекомендации:

- ✓ Модель можно использовать для приоритизации рассылок и товарных предложений: запускать кампании только на клиентов с высокой вероятностью покупки.
- ✓ Чтобы улучшить качество: учесть повторные покупки, добавить данные о прошлых реакциях на акции (например, по `base_sale`), использовать признаки категорий товара или бренда.
- ✓ F1-score в районе 0.59 — это хороший старт для рекомендательной модели без поведенческих меток (например, кликов, просмотров).



Бизнес-рекомендации

- **Итоговый вывод:**

Анализ позволил выделить сегменты клиентов и оценить эффективность кампаний.

Построена модель склонности к покупке, которую можно использовать для персонализированных коммуникаций.

- ✦ **Рекомендации:**

- Использовать сегментированный подход для разных кластеров
- Продолжить А/В-тестирование с чёткой методологией
- Применять модель склонности для повышения отклика
- Развивать аналитику: LTV, отток, поведенческие признаки
- Учесть высокую чувствительность некоторых сегментов к скидкам



Спасибо за внимание

Проект выполнен в рамках курса Skillbox
Data Analyst Junior, 2025

