

Отсчет по анализу маркетинговых кампаний и прогнозированию склонности клиентов к покупке

Имя докладчика:
Мухитдинов Тимур

Введение

Цель проекта

Задачи:

Провести A/B-тест для оценки эффективности изменений.

Сегментировать клиентов (кластеризация).

Построить модель склонности клиента к покупке.

Данные

Источники данных (таблицы personal, purchases и др.)

Объем данных (количество клиентов, транзакций)

Основные признаки: возраст, пол, город, траты и т.д.

A/B-тест

Формирование контрольной и тестовой групп

Метрики:

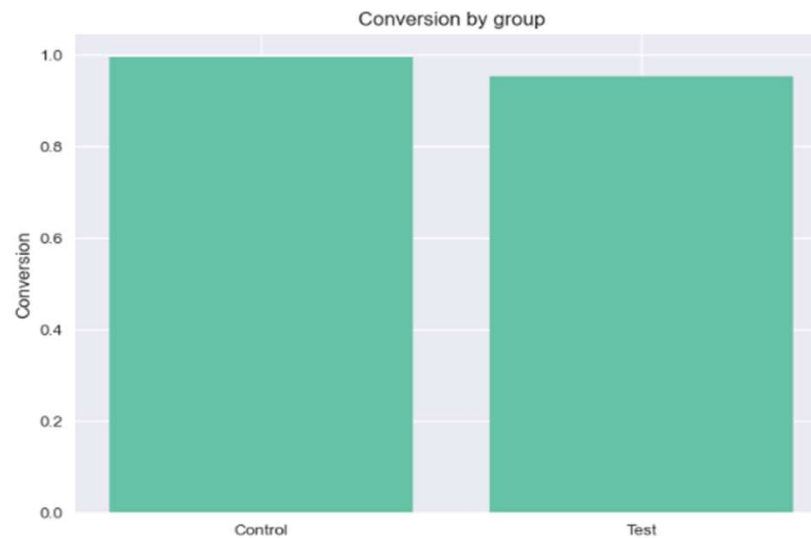
Conversion

ARPU

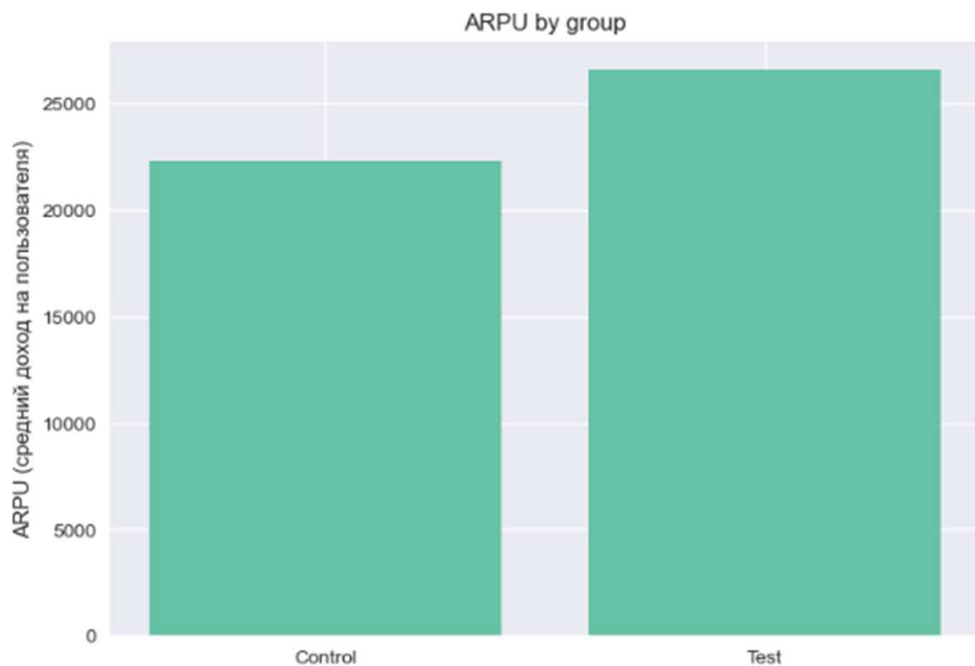
AOV

Результаты теста: таблица + графики (бар-чарты для conversion и ARPU)

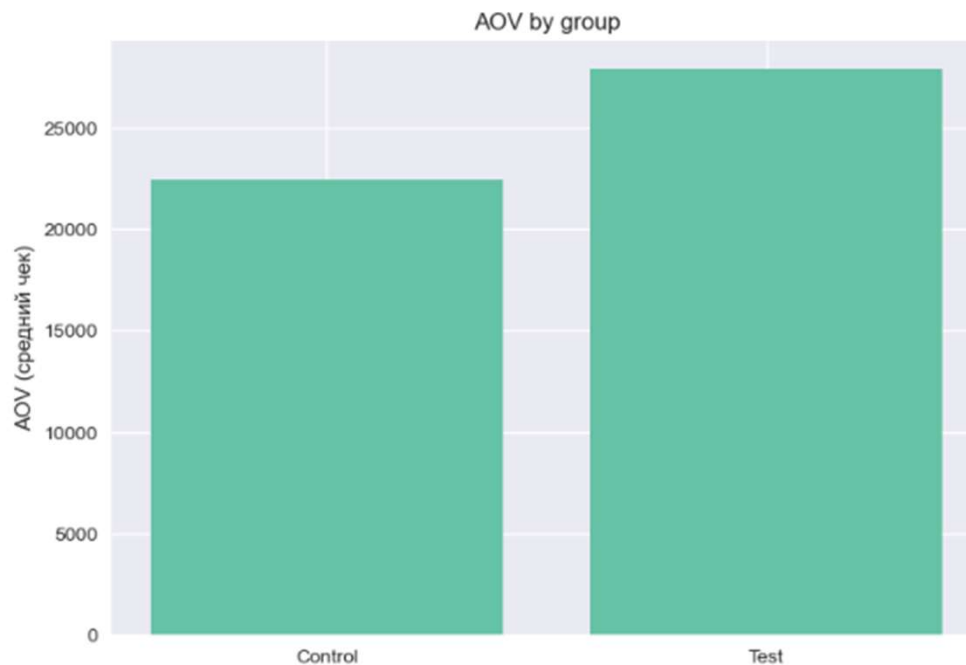
	group	n_users	conversion	arpu	aov
0	Control	5009	0.996007	22293.168097	22382.537382
1	Test	5016	0.954346	26603.402313	27876.053060



Конверсия в группе Control составила ~99,6%, а в Test ~95,4%. Разница статистически значима ($p < 0.001$), то есть предложение персональной скидки снизило вероятность покупки.



Средний доход на пользователя (ARPU) оказался выше в тестовой группе: 26 603 USD против 22 293 USD. Разница значима ($p < 0.001$). Это говорит о том, что клиенты, получившие скидку, тратили больше.



Средний чек (AOV) также выше в тестовой группе — 27 876 USD против 22 383 USD в контрольной. Это подтверждает, что предложение скидки стимулировало клиентов совершать более дорогие покупки

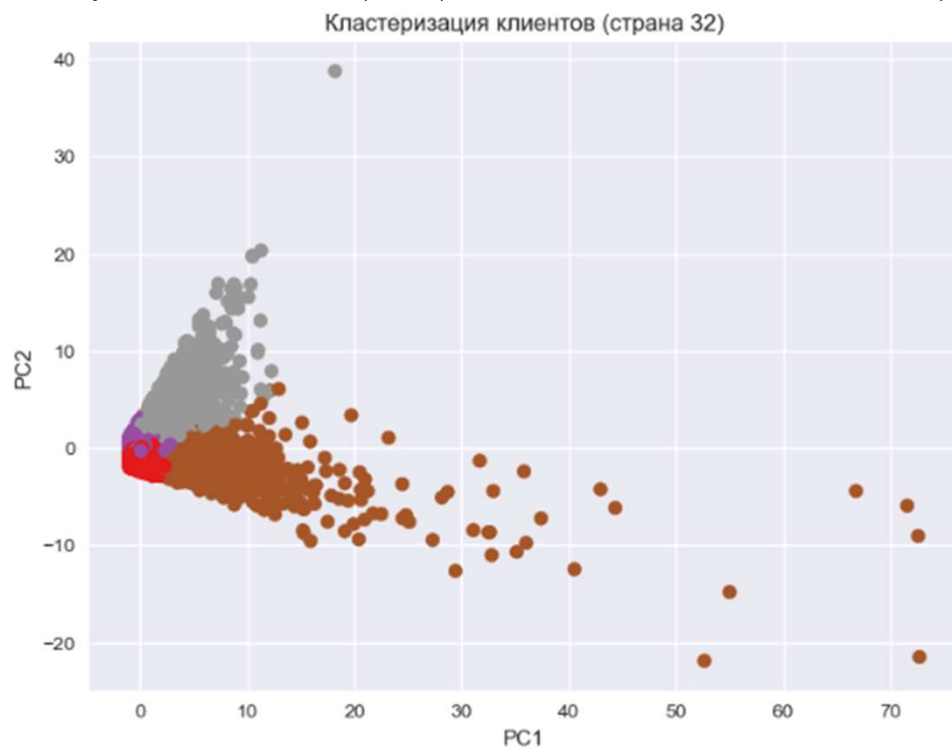
Вывод: тест значимо повлиял на метрики

Кластеризация (страна 32)

Подготовка данных (признаки: возраст, траты, число покупок и др.)

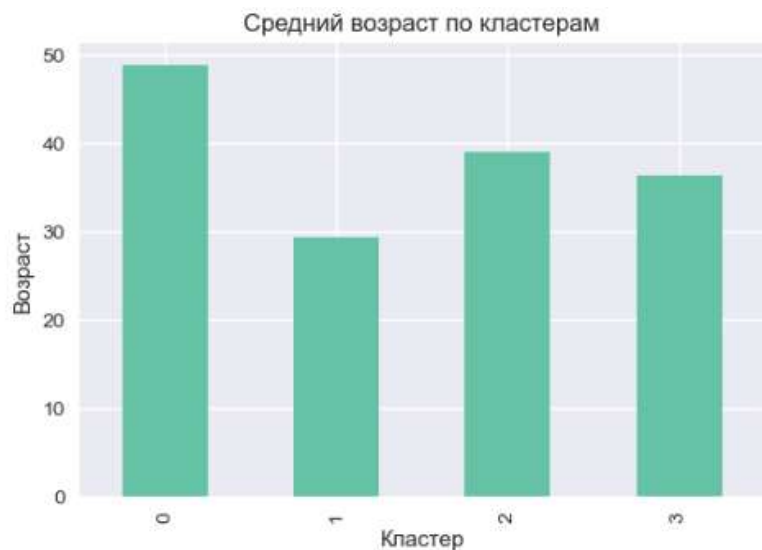
Метод: KMeans, $k=3$ (или выбранное число кластеров)

2D-визуализация PCA (распределение клиентов по кластерам)



На графике видно, что клиенты разбиваются на 4 чётко выраженные группы. Это подтверждает наличие различных сегментов внутри аудитории.

Средние показатели по кластерам: Средний возраст



На представленном графике показан средний возраст клиентов для каждого из четырех кластеров.

- **Кластер 0** имеет самый высокий средний возраст, близкий к 50 годам.
- **Кластер 1** является самым молодым, со средним возрастом около 30 лет.
- **Кластеры 2 и 3** находятся в среднем возрастном диапазоне, примерно от 35 до 40 лет

Средний чек

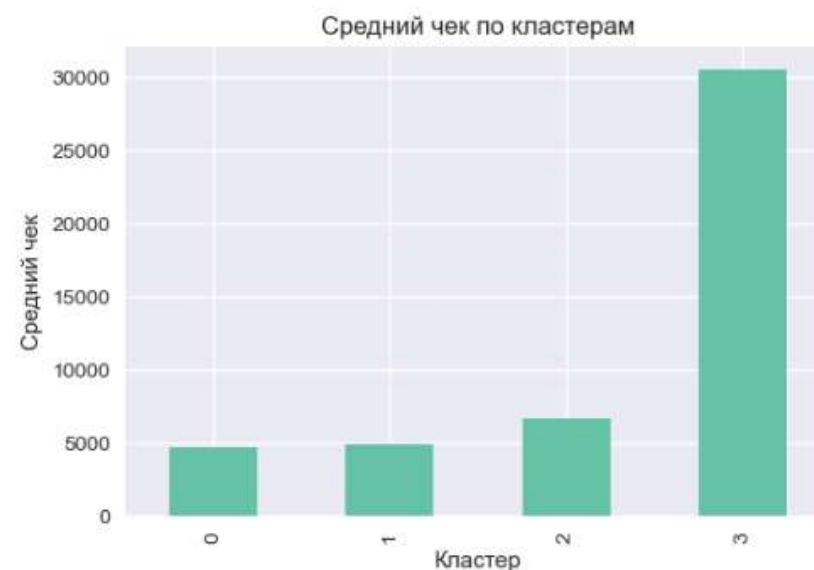


График отражает среднюю сумму одной покупки. Клиенты из Кластера 3 имеют самый высокий средний чек, что говорит о том, что их покупки, хотя и редкие (как показано на третьем графике), очень дорогие. У Кластера 2 средний чек самый низкий, несмотря на то что их общие траты очень высоки. Это указывает на то, что они совершают много мелких покупок

Средние траты

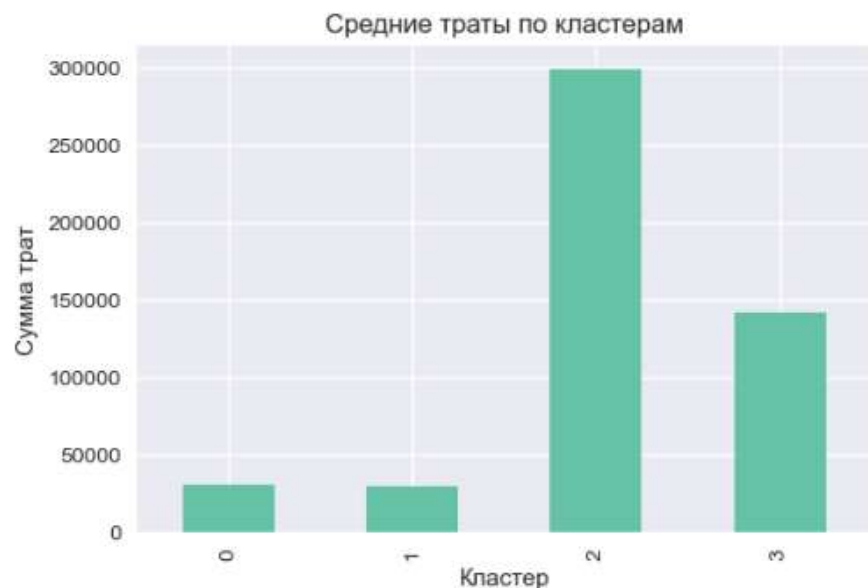
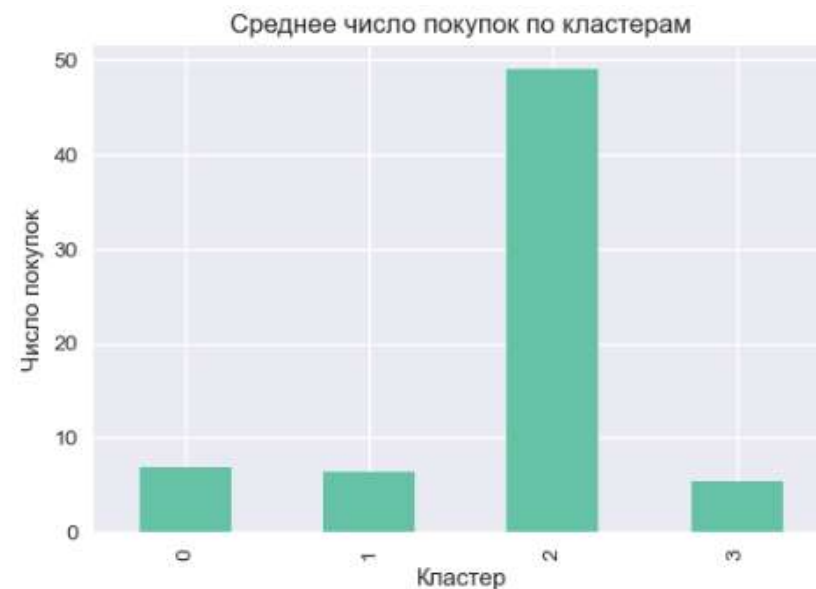


График показывает, сколько в среднем тратят клиенты из каждого кластера. Самые высокие средние траты наблюдаются у **Кластера 2**, а самые низкие — у **Кластера 0** и **Кластера 1**, которые тратят примерно одинаково мало. Средние траты **Кластера 3** значительно выше, чем у первых двух, но ниже, чем у Кластера 2

Число покупок



Графике видно, сколько покупок в среднем совершает клиент из каждого кластера. Кластер 2 совершает значительно больше покупок, чем все остальные. Это объясняет, почему их общие траты высоки, несмотря на низкий средний чек. Кластеры 0, 1 и 3 совершают примерно одинаково мало покупок

Модель склонности к покупке (город 1188)

Цель: предсказать вероятность покупки

Признаки: возраст, образование, коэффициенты и др.

Метод: Logistic Regression

Проблема: таргет оказался несбалансированным (все = 1) → модель не обучается

```
purchase_flag  
1    12438  
Name: count, dtype: int64
```



Вывод:

Все клиенты из города 1188 имеют purchase_flag = 1. Это означает, что в выборке отсутствует контрольная группа (нет клиентов без покупок). В таких условиях модель машинного обучения построить невозможно, так как нет различия классов



Заключение

А/В-тест: Предложение персональной скидки снизило конверсию, но привело к росту среднего дохода на пользователя (ARPU) и среднего чека (AOV), так как клиенты, которые совершили покупку, тратили больше.

Кластеризация: Были выделены четыре сегмента клиентов со схожим поведением. Это позволит разработать индивидуальные маркетинговые стратегии для каждого из них.

Модель склонности к покупке: Построить модель не удалось из-за отсутствия данных о клиентах, которые не совершали покупок. Это является ключевым ограничением для дальнейшего прогнозирования.