

# Разработка моделей оценки бизнес-эффективности рекламных кампаний и определения склонности клиентов к покупкам товаров

## СОДЕРЖАНИЕ:

- Слайд 2. Цели, задачи и обстоятельства исследования
- Слайд 3-4. Витрина данных и рекомендации по их нормализации
- Слайд 5-6. Методы оценки и оценка эффектов рекламных кампаний
- Слайд 7-8. Кластеризация данных (методы и выводы)
- Слайд 9. Прогноз склонности жителей города 1188 к покупкам

Аргентов Сергей

<https://argentov.pro>

2023 г.

# Цели и задачи



## ОБСТОЯТЕЛЬСТВА ПРОВЕДЕНИЯ ИССЛЕДОВАНИЯ

Компания является международной сетью по продаже спортивных товаров, представлена во многих городах разных стран

Проведены две тестовые рекламные кампании по предоставлению скидок с разделением клиентов на тестовые и контрольную группы

При наличии эффекта от рекламных кампаний их планируется повторить в отдельных городах

Цель 1

Статистически значимо определить эффективны ли проведённые рекламные кампании

Цель 2

Научиться предсказывать склонность клиентов к покупкам товаров (на примере города 1188)

Задачи

- Сформировать витрину (хранилище) данных о клиентах и товарах
- Разработать модель машинного обучения для восстановления данных
- Определить бизнес-критерии для сравнения выборок по тестовым и контрольным группам
- Подобрать статистические тесты бизнес-эффективности рекламных кампаний
- Подготовить данные и разработать модель, предсказывающую склонность клиентов к покупкам

### 3. ФОРМИРОВАНИЕ ВИТРИНЫ ДАННЫХ

#### ТЕХНОЛОГИЯ

Данные собраны в легко-настраиваемом хранилище sqlite3

#### БИЗНЕС-ЭФФЕКТИВНОСТЬ ВЫБРАННОЙ ТЕХНОЛОГИИ

Сопряжение хранилища с аналитическими инструментами производится при помощи Python3, то есть не требует расходов

Эксплуатация хранилища будет осуществляться силами аналитиков без привлечения Инженеров по эксплуатации дополнительного оборудования и более сложных систем управления базами данных

## 4. РЕКОМЕНДУЕМАЯ НОРМАЛИЗАЦИЯ ДАННЫХ

### НЕОДНОЗНАЧНОСТЬ ДАННЫХ

Пол потребителя – 40%,  
Цвет товара – 1%

Целесообразно формализовать заполнение при регистрации товара на складе, а также централизовать справочную систему характеристик товаров

Для предсказания склонности к покупкам не критично

### ОТСУТСТВИЕ ДАННЫХ

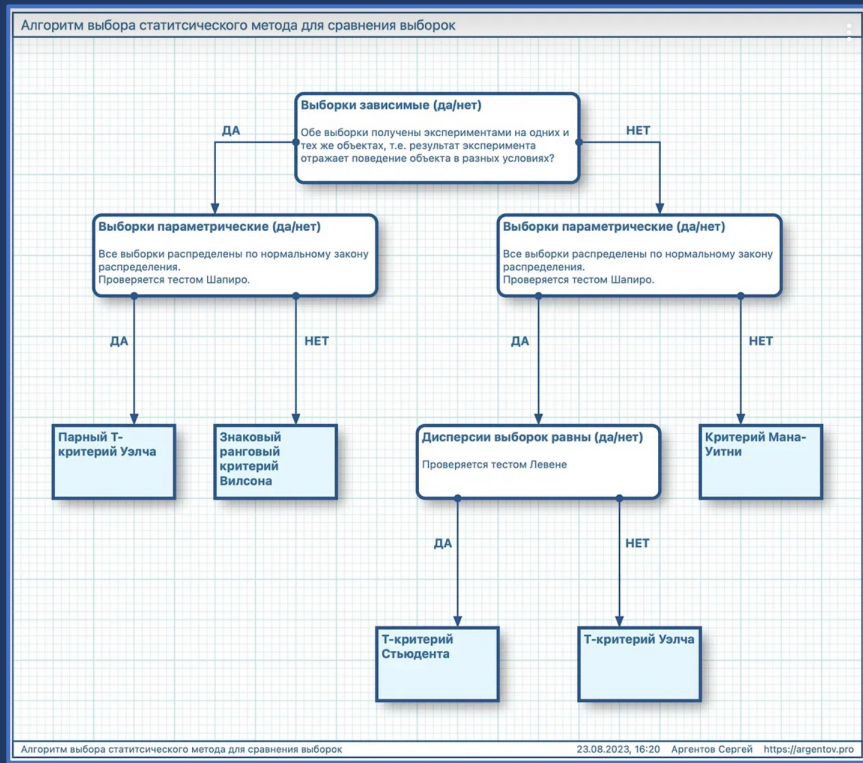
Пол покупателя – 15%  
Цвет товара – 14%

Целесообразно формализовать заполнение, т.к. пустоты могут обозначать не только «отсутствие данных», но и «технические неисправности при выгрузках»

Для предсказаний склонности к покупкам может быть критично

# 5. ОПРЕДЕЛЕНИЕ ЭФФЕКТИВНОСТИ КАМПАНИЙ

Разработан класс для автоматического выбора статистического теста различия двух выборок



Сформулированы наборы метрик эффективности:

- А. по сумме и частоте покупок
- Б. по ориентации на клиентов с наибольшим доходом

ПЕРВАЯ  
КАМПАНИЯ

Эффективна  
по группе метрик А

Не эффективна  
по группе метрик Б

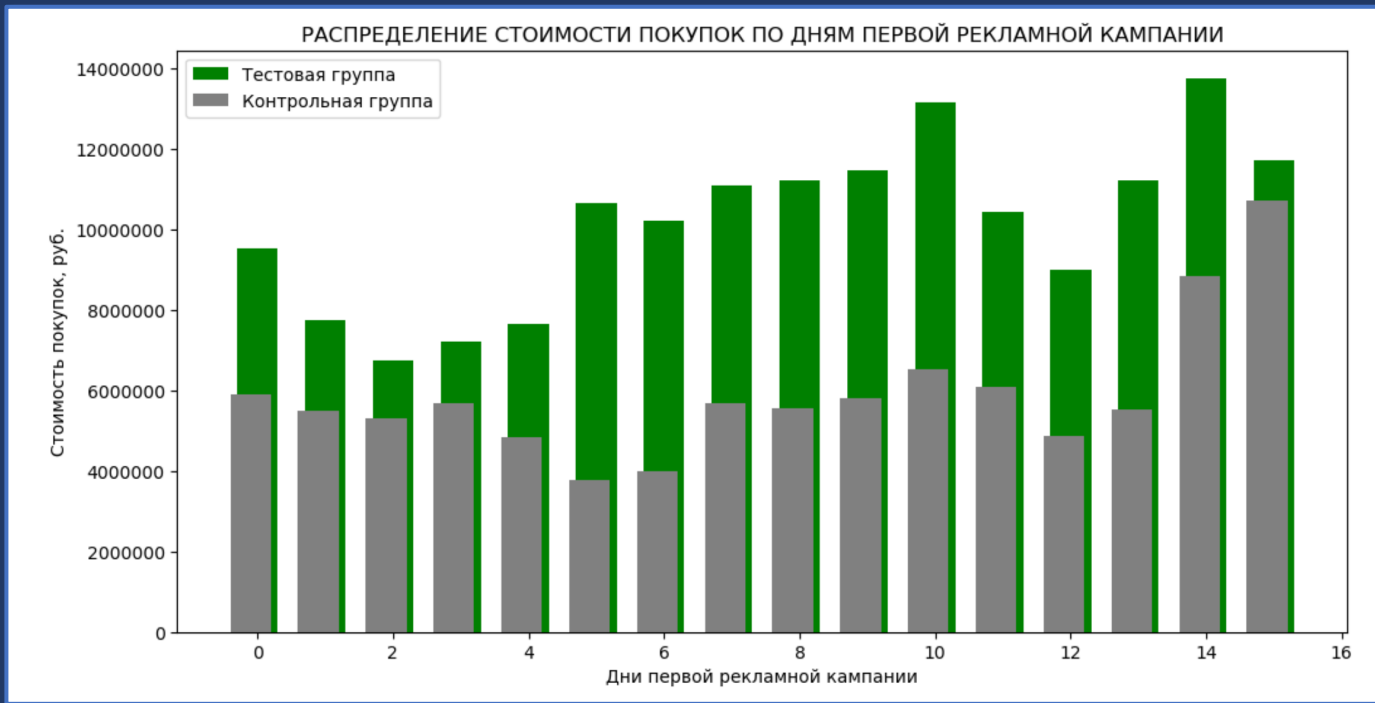
ВТОРАЯ  
КАМПАНИЯ

Не эффективна  
по группе метрик А

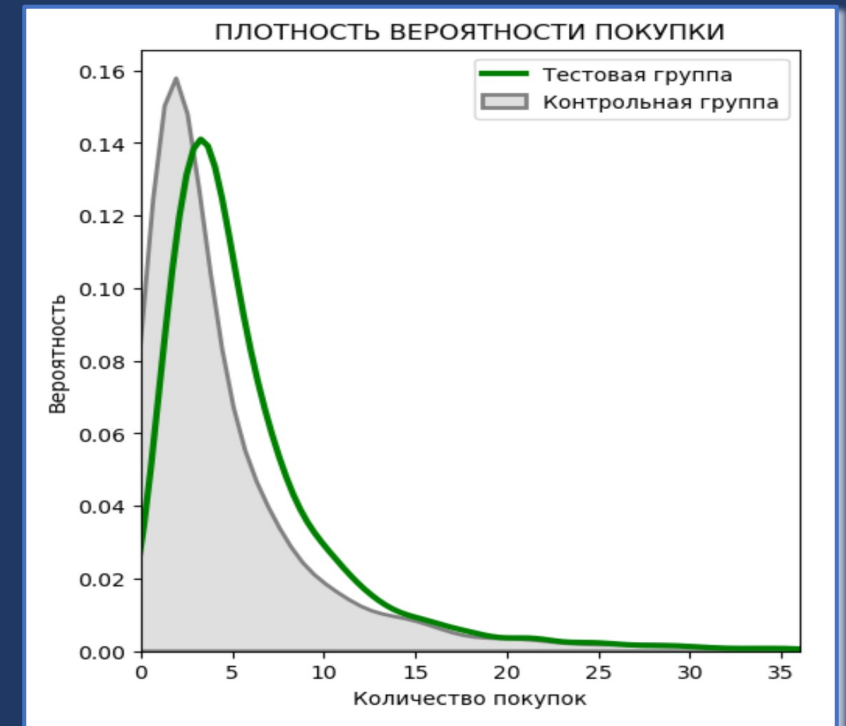
Не эффективна  
по группе метрик Б

# 6. РЕЗУЛЬТАТЫ ЭФФЕКТИВНОЙ РЕКЛАМНОЙ КАМПАНИИ

Увеличилась стоимость покупок



Увеличилась частота покупок



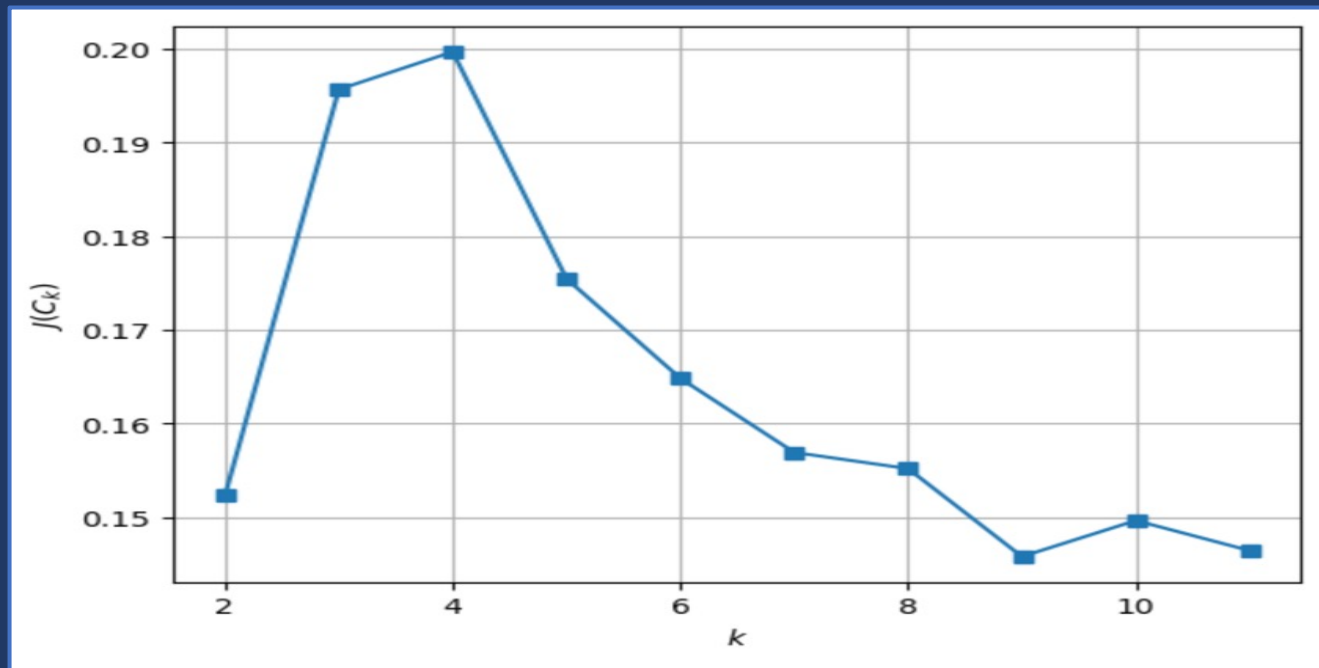
Дополнительной ресурсной зоной для повышения эффективности рекламных кампаний является разделение клиентов по доходам возрастным группам:

<https://blog.domclick.ru/ipoteka/post/kazhduyu-desyatuyu-ipoteku-v-sbere-berut-molodye-lyudi-analitika-domklik>

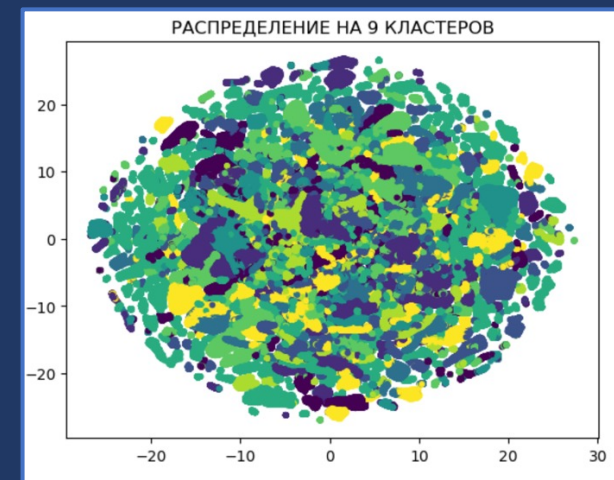


# 7. КЛАСТЕРИЗАЦИЯ

1. ПО ГРАФИЧЕСКОЙ  
КРИВОЙ «КОЛЕН»  
ОПРЕДЕЛЕНА  
ОПТИМАЛЬНЫМИ 4 ИЛИ 9  
КЛАСТЕРОВ



2. ВЫБРАНО 4 КЛАСТЕРА  
ИЗ-ЗА ОТСУТСТВИЯ  
ВИЗУАЛЬНОГО РАЗЛИЧИЯ  
УЛУЧШЕНИЯ РАЗДЕЛЕНИЯ  
ТОЧЕК ПРОСТРАНСТВА  
МЕЖДУ 4 И 9 КЛАСТЕРАМИ



## 8. ВЫВОДЫ ПО КЛАСТЕРИЗАЦИИ

### 1. КЛАССИФИКАЦИЕЙ ДЛЯ КАЖДОГО КЛАСТЕРА ОПРЕДЕЛЕНА НАИБОЛЕЕ ВАЖНЫЕ ПРИЗНАКИ

В каждом кластере присутствует влияние признаков **age**, **education**, **personal\_coef** – значит при разработке кампаний вместе с персональным коэффициентом целесообразно, в первую очередь, акцентировать внимание на конкретные возрастные группы, например:

- клиентам от 31 до 61 года целесообразно предлагать наиболее дорогие и большие по комплекту покупки,
- при наличии ресурсов целесообразно провести кампании для каждой возрастной группы,
- при недостатке ресурсов целесообразно провести кампании для групп с более высокими доходами.

### 2. ПРИЗНАК GENDER ЯВЛЯЕТСЯ ОПРЕДЕЛЯЮЩИМ ДЛЯ ОДНОГО ИЗ КЛАСТЕРОВ

Целесообразно разработать отдельные кампании для женщин и мужчин, проводя их в разное время для оптимизации работы консультантов-продавцов.

### 3. ПЕРЕД КОНТАТОМ С КЛИЕНТОМ ЦЕЛЕСООБРАЗНО ОПРЕДЕЛИТЬ ЕГО КЛАСТЕР

Для каждого кластера определены: влияние скидки и топ-10 самых востребованных товаров, которые можно предлагать клиенту к совместной покупке с другими товарами.



## 9. ПРОГНОЗ СКЛОННОСТИ КЛИЕНТОВ К ПОКУПКАМ

### 1. МОДЕЛЬ ПРОГНОЗА ПОСТРОЕНА НА АЛГОРИТМЕ «СЛУЧАЙНЫЙ ЛЕС»

- в качестве целевой группы выбраны клиенты, участвовавшие в первой кампании и купившие более одного товара,
- из обучающих данных исключены 21% аномалий,
- полученная точность модели при тестировании 96,5%.

### 2. МОДЕЛЬ ПРЕДСКАЗАЛА БОЛЬШУЮ СКЛОННОСТЬ ЖИТЕЛЕЙ ГОРОДА 1188 К ПОКУПКАМ, ЧЕМ ПОКУПАТЕЛЕЙ В РАНЕЕ ПРОВЕДЕННОЙ КАМПАНИИ

- 66% участников первой рекламной кампании купили более одного товара,
- 72% (спрогнозировано моделью) покупателей-жителей города 1188 купят более одного товара в случае проведения рекламной кампании.