

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

Полные материалы проекта размещены на GitHub:  
<https://github.com/ArgentovS/PurchasePropensityModel>

## 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 и классификация данных
- Слайд 9. Модель склонности клиентов 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 тестовую 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 недостающих 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. НОРМАЛИЗАЦИЯ И ВОССТАНОВЛЕНИЕ ДАННЫХ

## НЕ НОРМАЛИЗОВАННЫЕ ДАННЫХ

Пол пользователя – 40%,  
Цвет товара – 1%

Для предсказания склонности к  
покупкам не критично

ПОЛ заполнен значением «унисекс»  
ЦВЕТ заполнен из наименований  
товаров и нормализацией значений

## ОТСУТСТВУЮЩИЕ ДАННЫХ

Цвет товара – 14%  
Пол покупателя – 15%

Для предсказаний склонности к  
покупкам может быть критично

ЦВЕТ заполнен наиболее частыми  
значениями (в многопроцес. режиме)  
ПОЛ заполнен машинным обучением

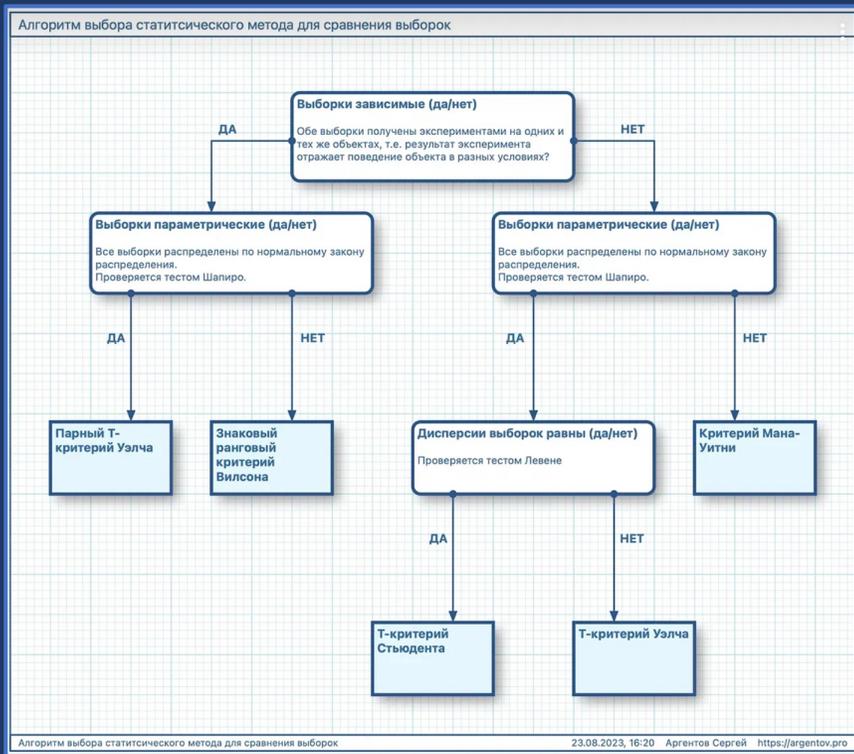
Целесообразно формализовать заполнение данных при регистрации и  
продажах товаров с централизацией системы справочных характеристик

# 5. МЕТОДЫ ОЦЕНКИ 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

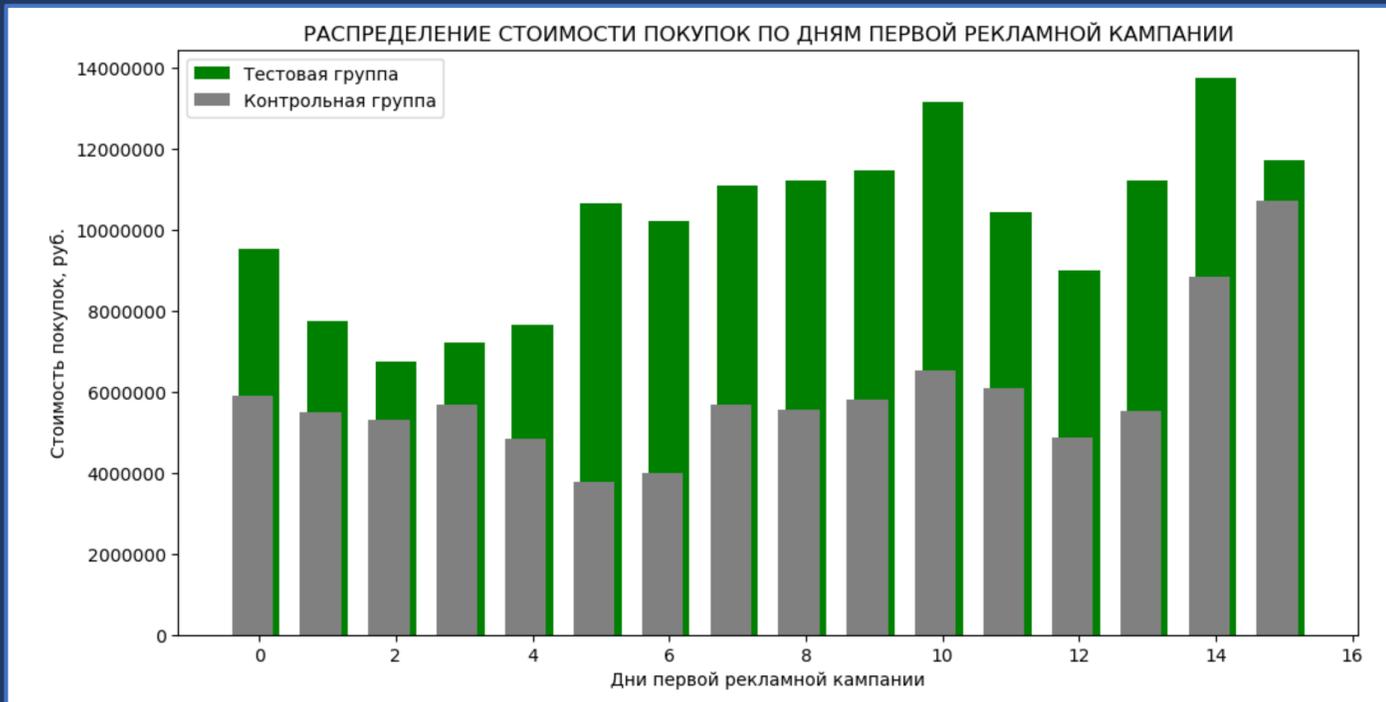
ВТОРАЯ  
КАМПАНИЯ

Не эффективна  
по группе метрик А

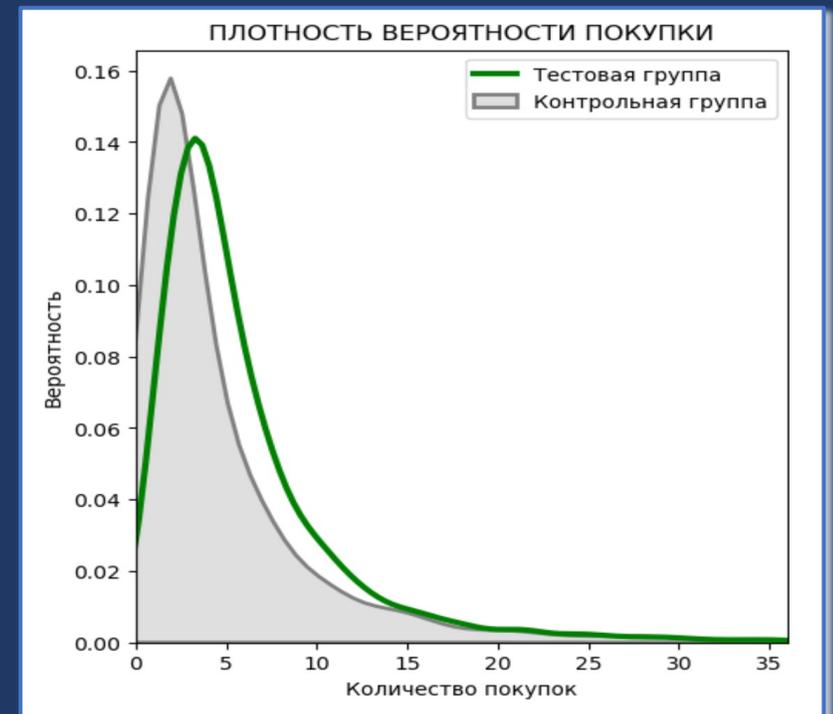
Не эффективна  
по группе метрик Б

# 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 И КЛАССИФИКАЦИЯ ДАННЫХ

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

2. ВЫРАНЫ 4 КЛАСТЕРА, ПО КОТОРЫМ ПОЛУЧЕНО ТОП-10 ТОВАРОВ И ПРОВЕДЕНА КЛАССИФИКАЦИЯ С ВЫДЕЛЕНИЕМ ВАЖНЕЙШИХ ПРИЗНАКОВ ОТНЕСЕНИЯ К КАЖДОМУ КЛАСТЕРУ

Лиф женский Joss  
Плавки женские Joss  
Футболка женская Demix  
Футболка мужская Demix  
Кресло кемпинговое Out  
Ветровка женская Outve  
Брюки женские Outventu  
Балетки женские Demix  
Футболка женская FILA  
Сабо Crocs Crocband

Брюки мужские Demix  
Брюки мужские Outventu  
Брюки мужские FILA  
Купальник женский Joss  
Брюки женские FILA  
Лиф женский Termit  
Легинсы женские Demix  
Брюки женские Demix  
Брюки женские Outventu  
Купальник женский FILA

Шорты мужские Demix  
Футболка мужская Demix  
Рюкзак Demix  
Сумка FILA  
Футболка женская Demix  
Толстовка мужская Demix  
Шорты для мальчиков Demix  
Рюкзак FILA  
Сандалии женские Skechers  
Солнцезащитные очки K

Шорты мужские Demix  
Брюки мужские Demix  
Футболка мужская Demix  
Брюки мужские FILA  
Сабо Crocs Crocband  
Шорты мужские Nike  
Футболка женская Demix  
Брюки мужские Outventu  
Брюки мужские Карра  
Футболка мужская Termit

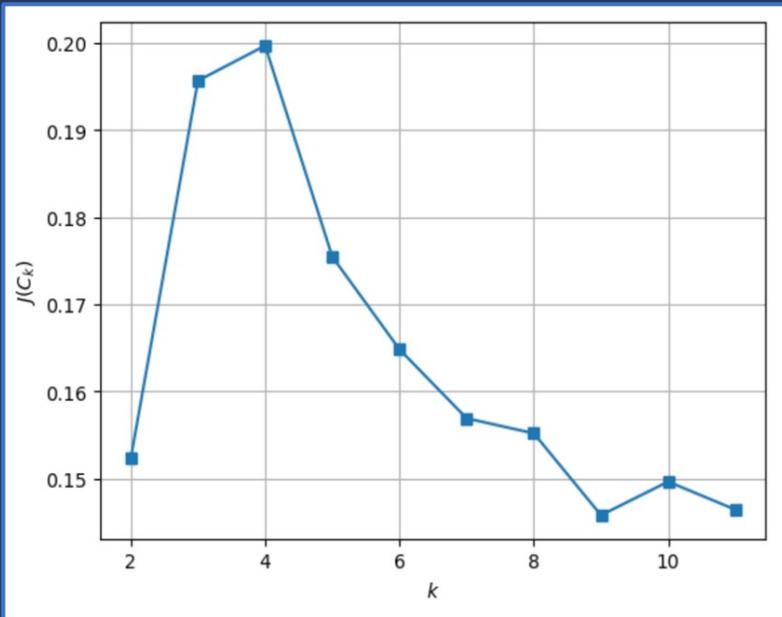
product: 0.39  
cost: 0.23  
age: 0.16  
city: 0.14

product: 0.61  
cost: 0.13  
age: 0.10

product: 0.61  
cost: 0.11  
age: 0.10

education: 0.53  
age: 0.31  
pers\_coef: 0.15

КЛАССИФИКАЦИЯ Дополнительно подтвердила что ресурсной зоной для повышения эффективности рекламных кампаний является разделение клиентов по возрастным группам (в т.ч. по доходам возрастных групп)



## 8. ВЫВОДЫ ПО КЛАСТЕРИЗАЦИИ

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

Для каждого кластера определены: топ-10 самых востребованных товаров, которые можно предлагать клиенту к совместной покупке с другими товарами. Скидка влияет на 1, 3, 4 кластер

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

В каждом кластере присутствует влияние признаков **age** – значит при разработке кампаний, в первую очередь, целесообразно акцентировать внимание на конкретные возрастные группы, учитывая статистические данные о различии доходов возрастных групп, например:

- клиентам от 31 до 50 лет целесообразно предлагать наиболее дорогие и большие по комплекту покупки,
- при наличии ресурсов целесообразно провести кампании отдельно для каждой возрастной группы,
- при недостатке ресурсов целесообразно провести кампании для групп с более высокими доходами.

### 3. ПРИЗНАК EDUCATION ЯВЛЯЕТСЯ ОПРЕДЕЛЯЮЩИМ ДЛЯ ОДНОГО ИЗ КЛАСТЕРОВ

Целесообразно разработать отдельные кампании для, например, сотрудников офисов и предприятий, проводя их в разное время для оптимизации работы консультантов-продавцов.

## 9. ПРОГНОЗ СКЛОННОСТИ КЛИЕНТОВ К ПОКУПКАМ

### 1. МОДЕЛЬ ПРОГНОЗА ПОСТРОЕНА НА АЛГОРИТМЕ НЕЙРОСЕТИ «ПЕРЦЕПТРОН»

- для обучения выбраны данные о клиентах, участвовавших в первой кампании (таргетом является покупка более одного товара),
- из обучающих данных исключены 21% аномалий,
- полученная точность модели при тренинге 97,0%.

### 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,
- 77% (спрогнозировано моделью) покупателей-жителей города 1188 купят более одного товара в случае проведения рекламной кампании.