

Skillbox

Итоговый проект курса:
«Профессия Data Analyst»

**Анализ маркетинговых компаний
магазина спортивных товаров**

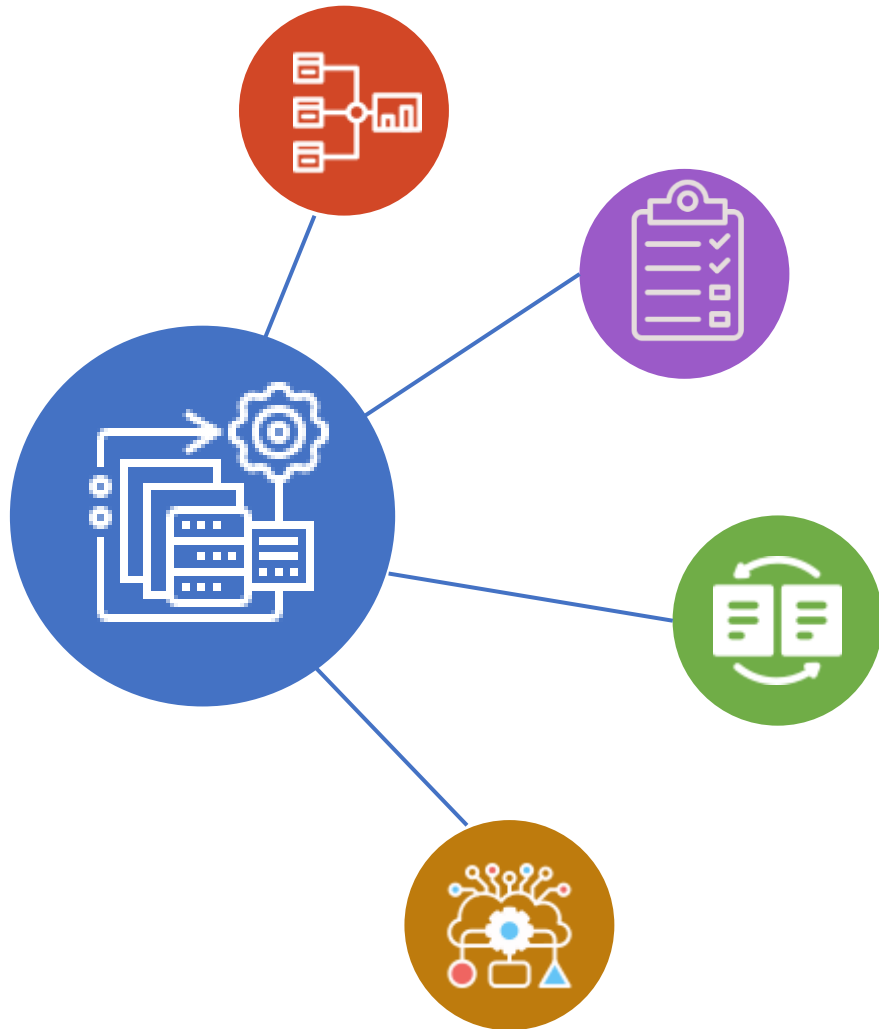


Этапы исследования

- 1 Предобработка данных. Восстановление утерянных значений (пол клиента)
- 2 Оценка эффективности уже проведённой ранее маркетинговой кампании (A/B-тестирование)
- 3 Кластерный анализ предоставленных данных, и выработка методов работы с каждым кластером
- 4 Построение модели склонности клиента к покупке, для планируемой маркетинговой компании.



Предобработка данных



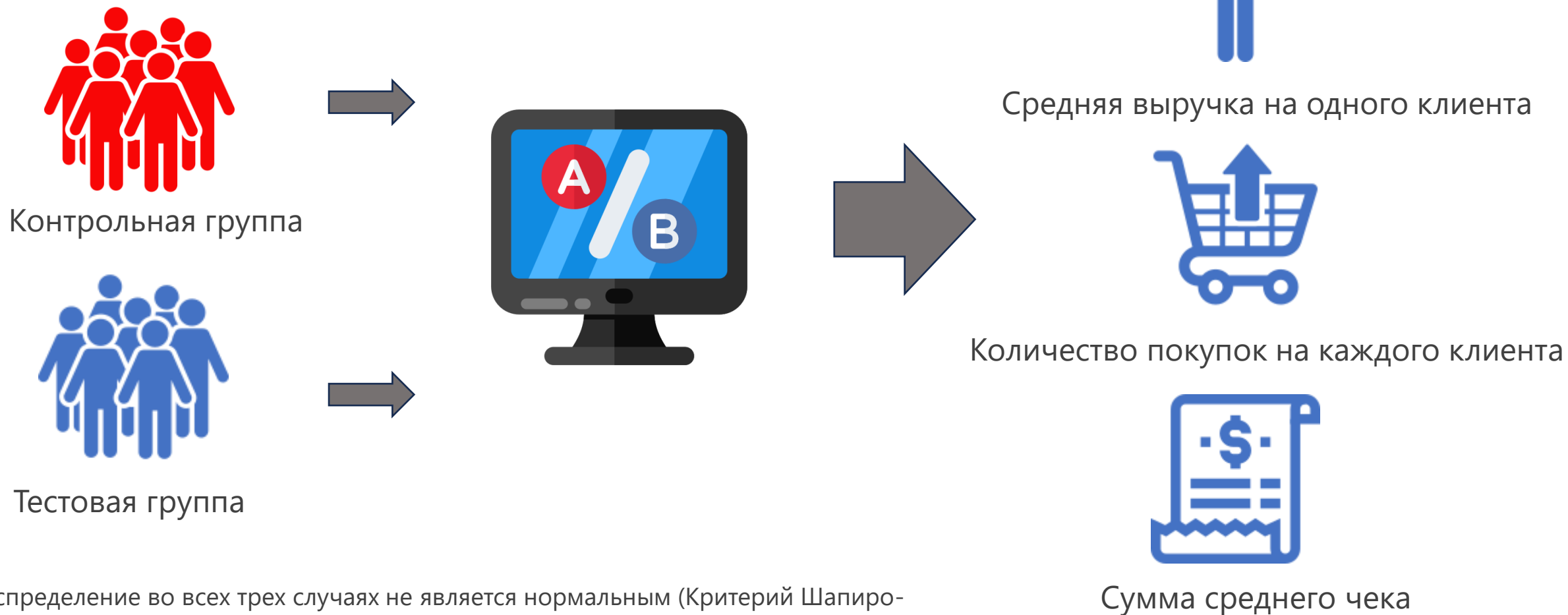
Объединение данных из различных источников

Заполнение пропусков

Приведение различных вариантов описания значений к единому виду

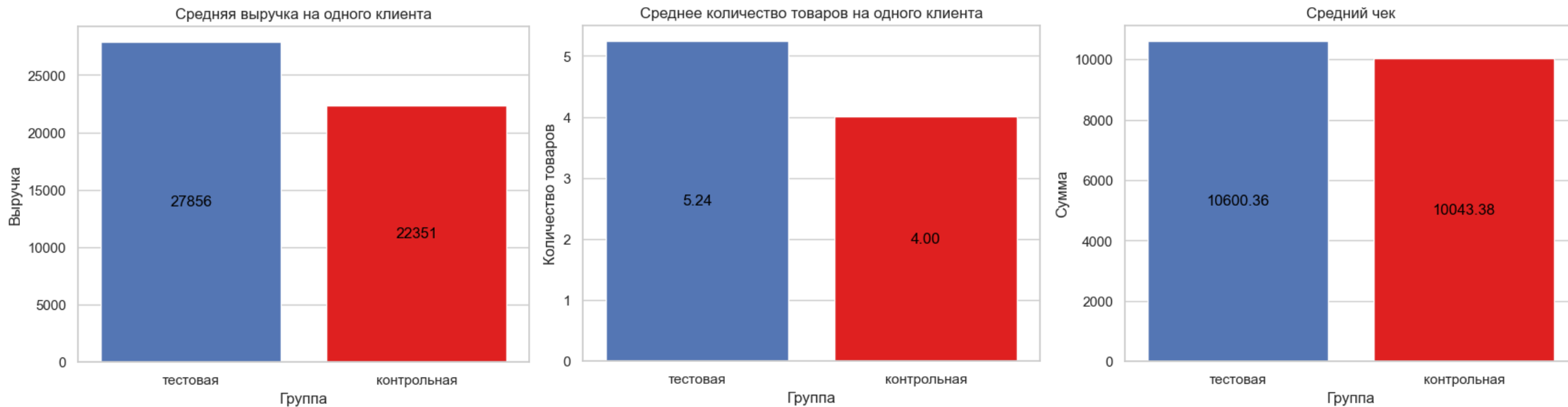
Бинарная классификация пропущенных значений пола клиента при помощи алгоритмов машинного обучения (случайный лес деревьев)

А/В тестирование:



Распределение во всех трех случаях не является нормальным (Критерий Шапиро-Уилка), для проверки равенства двух средних применяем критерий Манна-Уитни.

А/В тестирование:



По результатам А/В тестов выявлена значимая разница во всех выбранных для оценки эффективности метриках:



23.8%



30.4%



5.14%

Можно сделать вывод, что маркетинговая компания была эффективна.

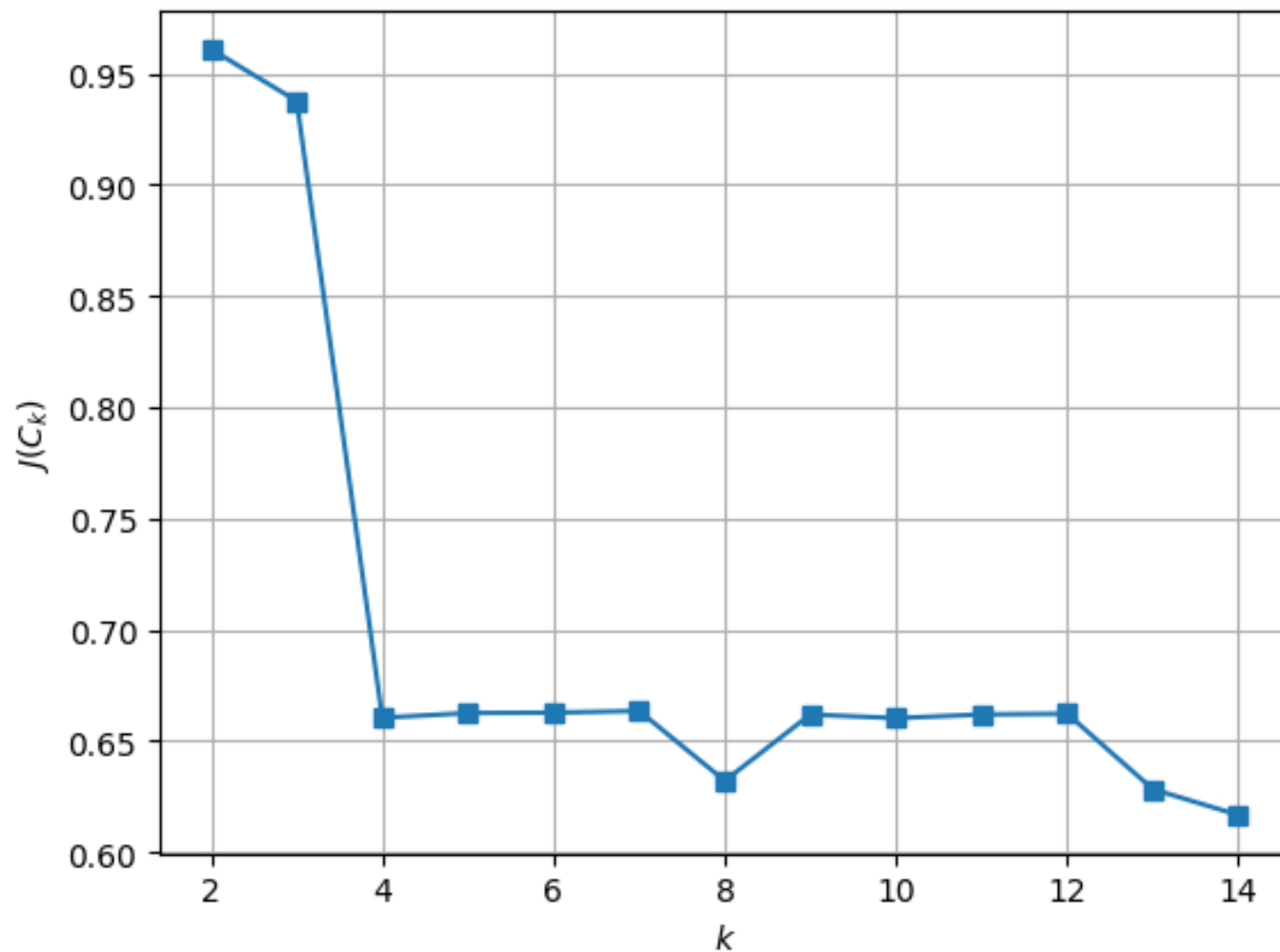
Кластерный анализ

- 1 Пол клиента
- 2 Возраст клиента
- 3 Уровень образования
- 4 Гендер товара
- 5 Сумма расходов
- 6 Наличие скидки у клиента

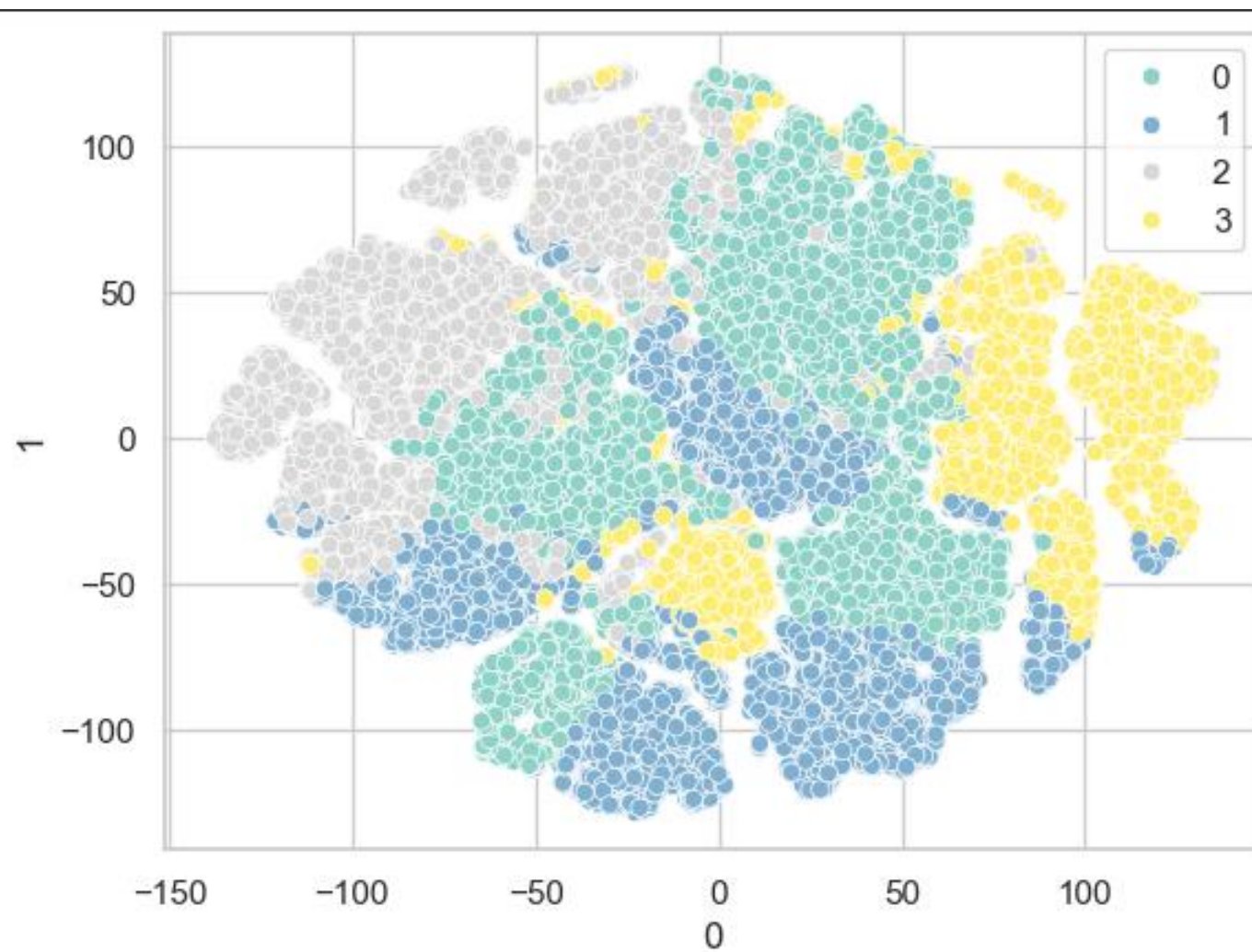


Кластерный анализ

Метод локтя



Кластерный анализ



Кластерный анализ

	Кластер 1	Кластер 2	Кластер 3	Кластер 4
Всего покупателей	35 419 (33,9%)	23 318 (22,33%)	26 183 (25,07%)	19 517 (18,69%)
Сумма покупок	1 881 млн. (44,57%)	663 млн. (15,73%)	814 млн. (19,28%)	862 млн. (20,42%)
Пол клиента	68% мужчины	63% мужчины	65% женщины	75% мужчины
Возраст клиента	43	39	41	21
Уровень образования	96% среднее	63% среднее	93% среднее	85,7% высшее
Средняя сумма покупок	53 129	28 475	31 098	44 184
Доля товаров, купленных со скидкой	24,1%	18,94%	61,39%	34,92%

Кластерный анализ

Кластер 1	Кластер 2	Кластер 3	Кластер 4
кроссовки мужские брюки мужские полуботинки мужские кроссовки женские футболка мужская шорты мужские кеды мужские рюкзак бейсболка брюки женские	кроссовки мужские кроссовки женские брюки мужские полуботинки мужские футболка мужская кеды мужские шорты мужские рюкзак велосипед горный сандалии женские	кроссовки женские кроссовки мужские брюки женские футболка женская брюки мужские полуботинки мужские футболка мужская кеды женские сандалии женские ветровка женская	кроссовки мужские кеды мужские кроссовки женские футболка мужская брюки мужские шорты мужские полуботинки мужские кеды женские рюкзак брюки женские

- Программа лояльности для часто покупающих.
- Увеличение информированности покупателя о товарах в магазине и предложение сопутствующих товаров.
- Рассылка email предложений с акционными товарами.
- Маркетинговая акция со скидками.

Модель склонности клиента к покупкам товара



Были взяты данные из двух ранее проведенных маркетинговых компаний (пол , возраст , образование, кластер, суммарное количество покупок каждого пользователя)

Пользователю в зависимости от количества покупок присвоенная переменная :

- Купил менее 3 товаров – 0
- Купил от 3 до 7 товаров – 1
- Купил более 7 товаров – 2

С помощью RandomForestClassifier была построена модель для предсказания количества товаров, которые приобретут клиенты из города 1188:

- менее 3 товаров – 9666 клиентов
- от 3 до 7 товаров – 237 клиентов
- более 7 товаров – 97 клиентов

Точность модели 65%

Заключение

В результате исследования было выявлено:

1

Требуется провести унификацию данных.

2

Маркетинговая компания на 5000 клиентов показала свою эффективности , можно проводить дальнейшее масштабирование.

3

Аудитория клиентов разбивается на 4 кластера, к каждому из которых можно применить целевой подход для увеличения прибыли.