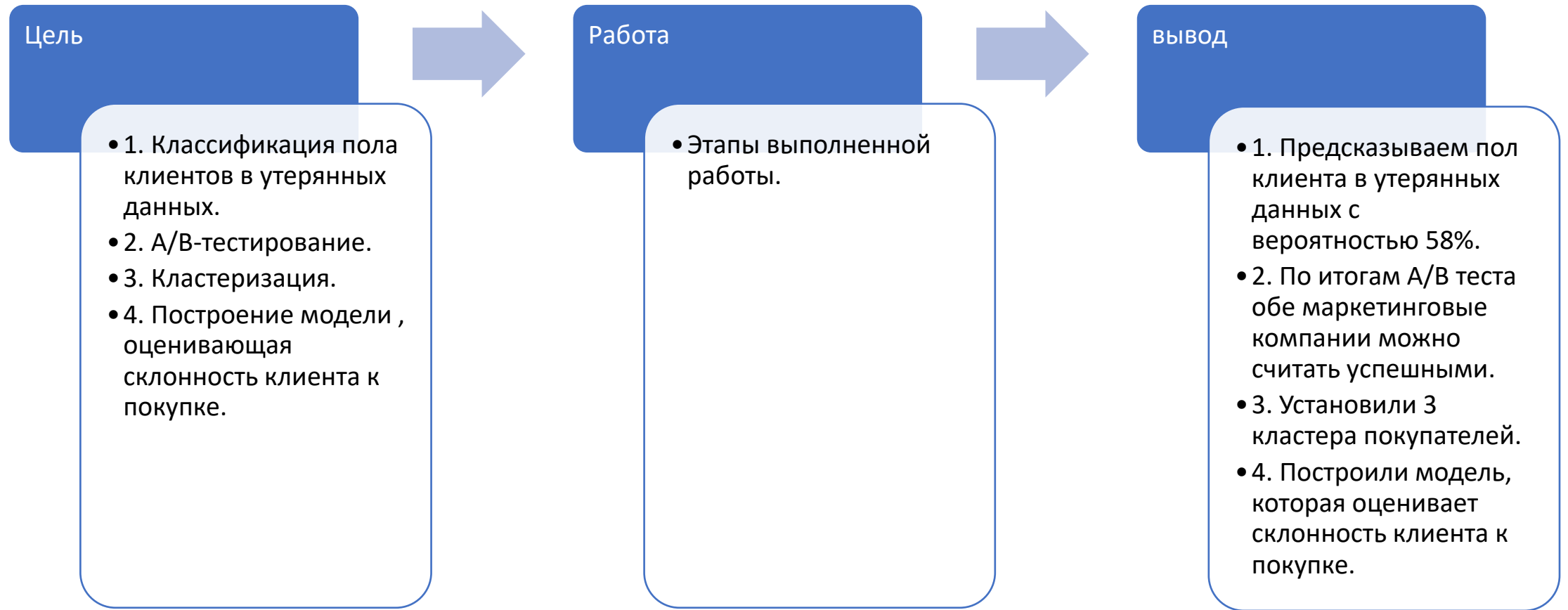


Выполнение итогового проекта



Этапы работы

1. Подготовка данных для выявления данных кто участвовал в рекламной компании.
2. Загружаем данные.
3. Удаление пропусков и приведение название продукта к одному формату.

Визуализация исходных данных.



4. Модель для прогноза пола клиентов.



5. A/B-тестирование.



6. Кластеризация



7. Вывод.

Визуализация исходных данных.



Модель для прогноза пола клиентов

| Модель | Ассурасу на трейне | Ассурасу на тесте |
|-------------------------|--------------------|--------------------|
| логистическая регрессия | 0.5734936287379138 | 0.5681470137825421 |
| Случайный лес | 0.5753505794967023 | 0.5673999925297875 |
| многослойный персептрон | 0.5773195876288659 | 0.5676614499682516 |

Вывод: используем для предсказания логистическую регрессию, но точность предсказания не высокая из за исходных данных.

A/B-тестирование.

1-ая маркетинговая компания.

| метрика | участвующих в акции | не участвующих в акции |
|----------------------------------|---------------------|------------------------|
| средний чек на одного покупателя | 16271 | 15473 |
| общая сумма покупок покупателей | 48 424 990 | 43 589 803 |
| количество покупателей | 2976 | 2817 |

2-ая маркетинговая компания.

| метрика | участвующих в акции | не участвующих в акции |
|----------------------------------|---------------------|------------------------|
| средний чек на одного покупателя | 9497 | 10220 |
| общая сумма покупок покупателей | 18 614 778 | 10 220 941 |
| количество покупателей | 1960 | 1000 |

Вывод : обе компании можно признать удавшейся

Кластеризация покупателей.

Выбор первоначального количества кластеров ,
используя метод 'локтя' и 'силуэта'



Окончательный выбор количества кластеров.



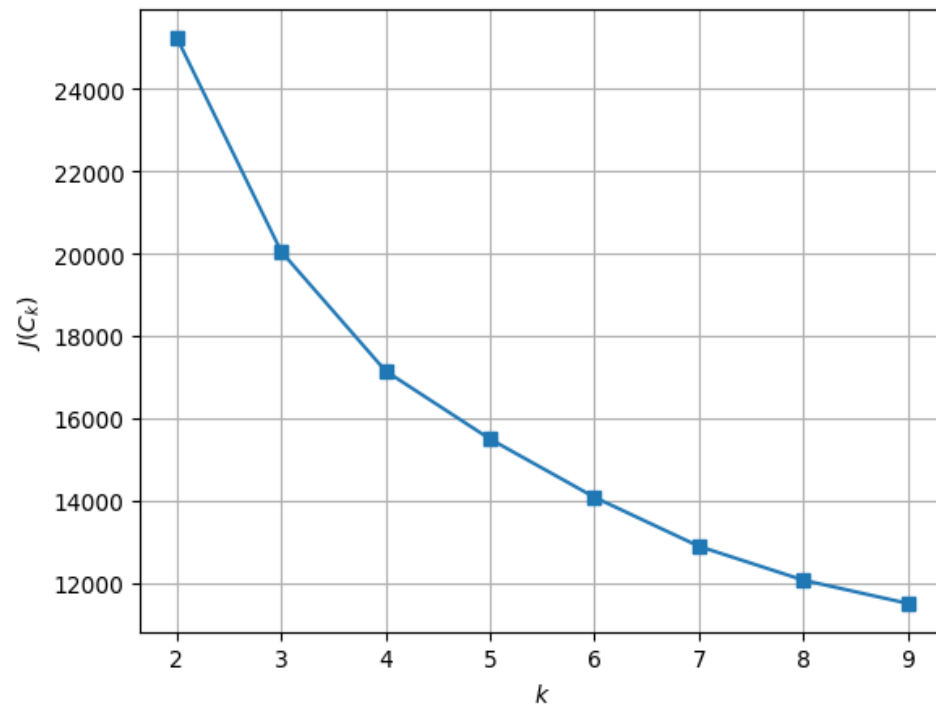
Определение наименования товаров по кластерам.



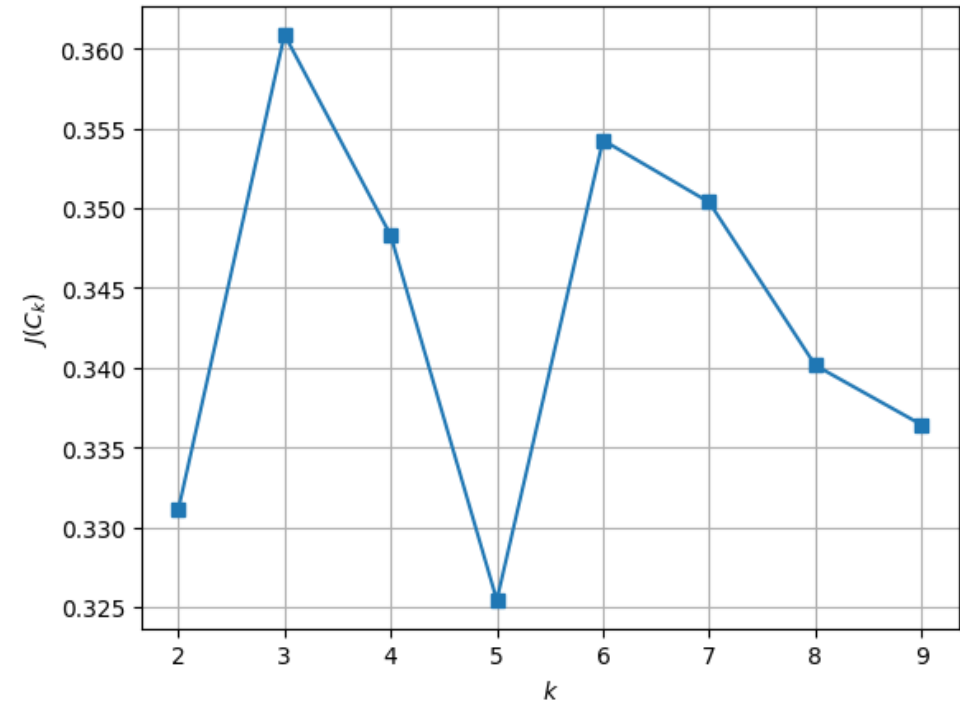
Влияние наличие скидки

Выбор первоначального количества кластеров , используя метод 'локтя' и 'силуэта'

Метод 'локтя'.

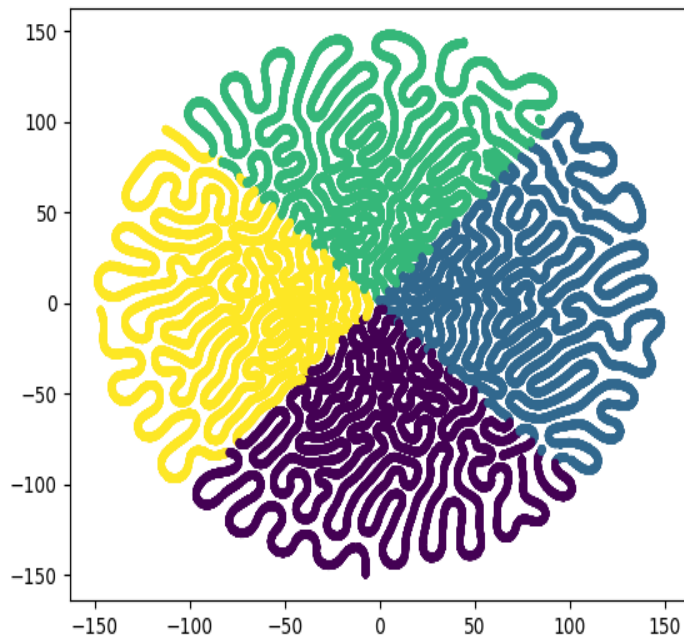


Метод 'силуэта'.

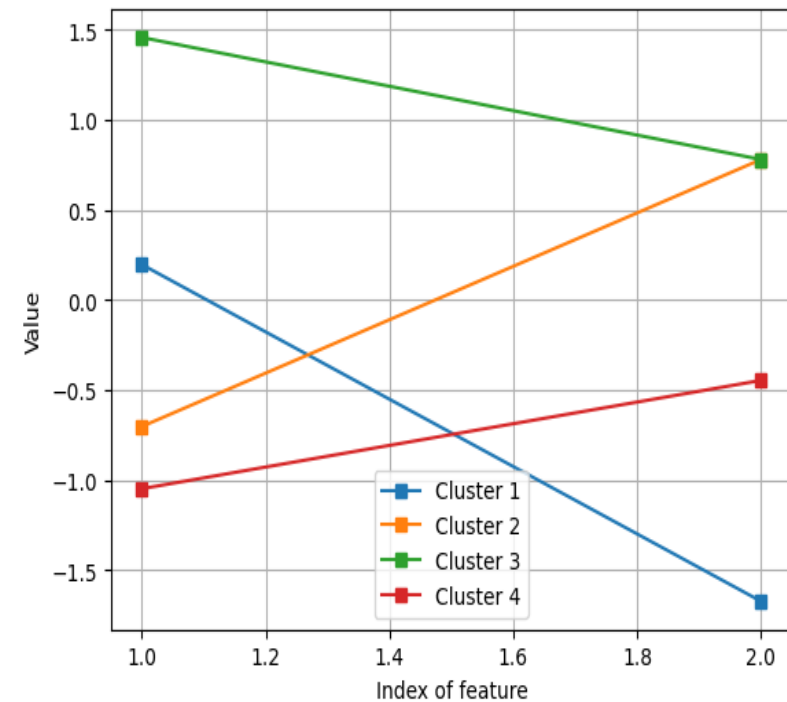


Выбор количества кластеров

4-е кластера

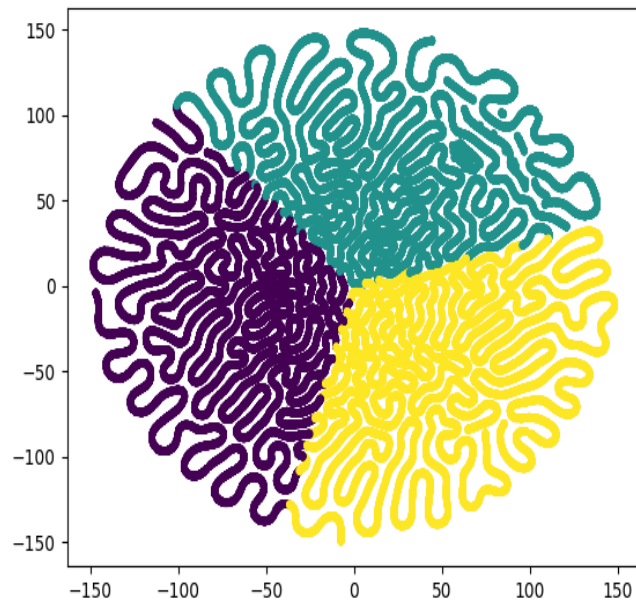


Распределение кластеров по признакам

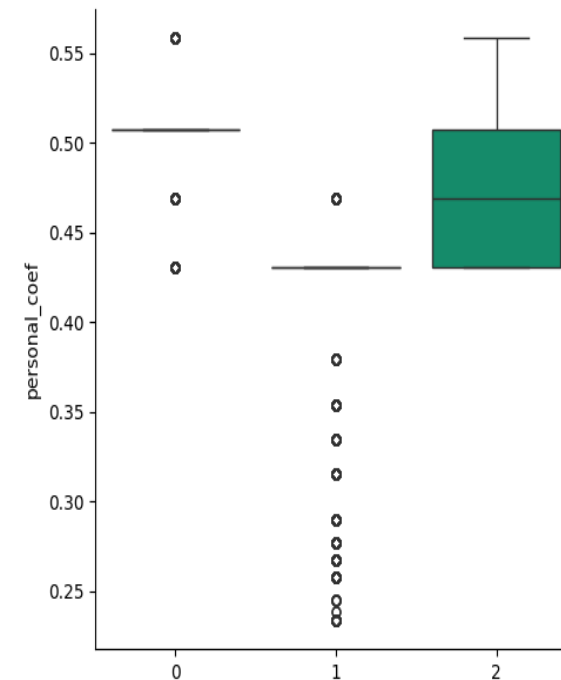


Итоговое количество кластеров и персональный коэффициент по кластерам.

3-и кластера

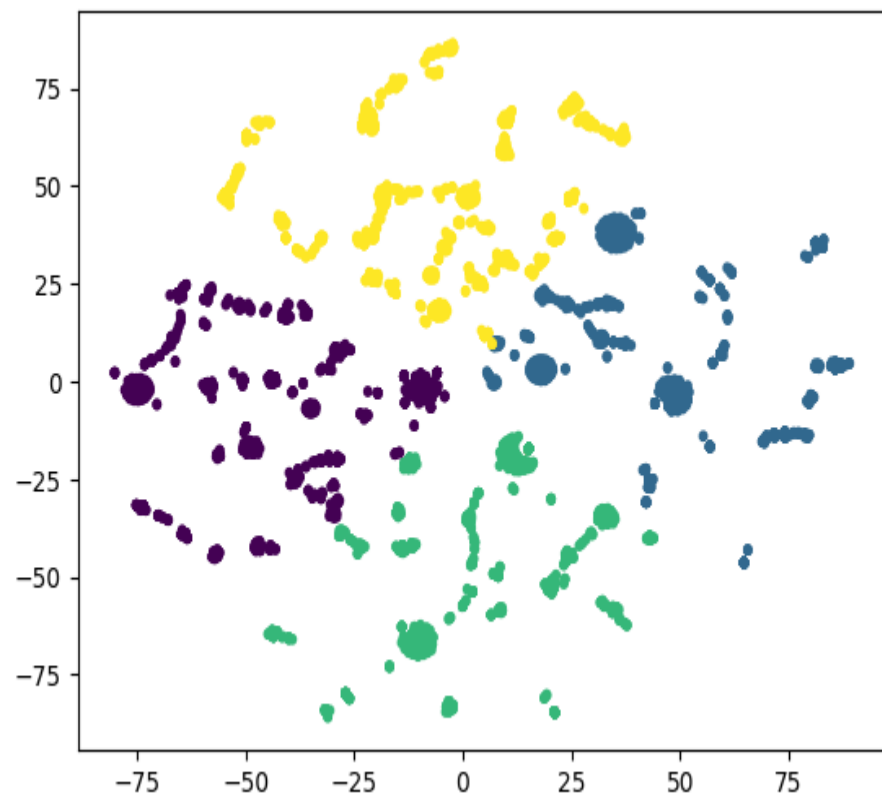


Распределение персонального коэф. по кластерам.

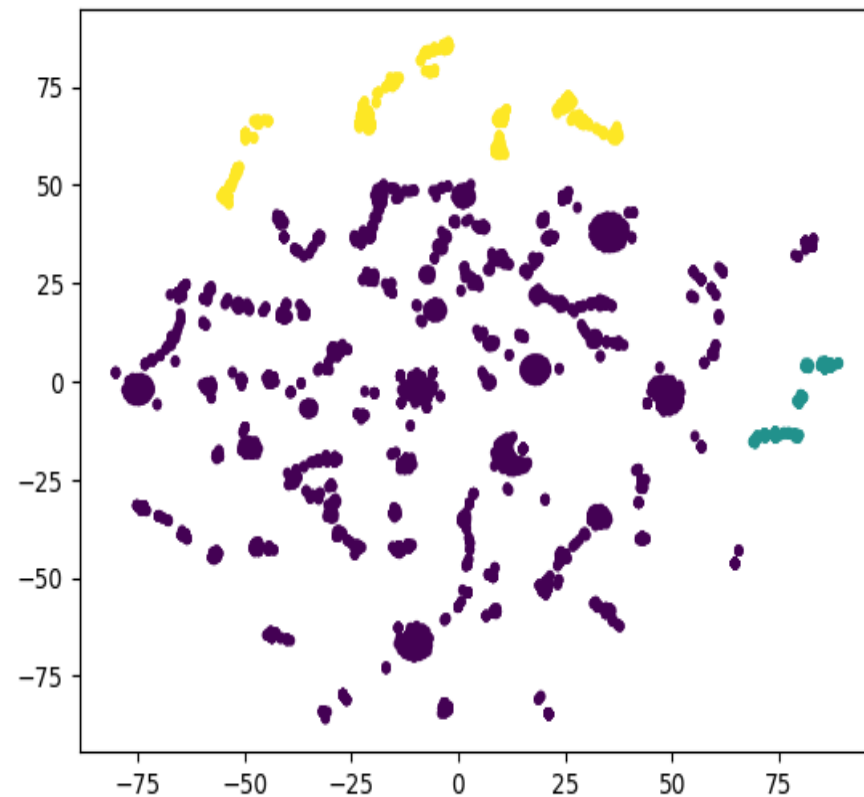


Визуализация кластеров

4-е кластера



3-и кластера



Товары по кластерам.

1 кластер (personal_coef > 0,5)

| наименование | количество, шт. |
|--------------|-----------------|
| Кроссовки | 8568 |
| Рюкзак | 4662 |
| Сандалии | 4544 |
| Футболка | 3562 |
| Бейсболка | 3405 |
| Носки | 3141 |
| Сумка | 3121 |

2 кластер (personal_coef < 0,5 и > 0, 45)

| наименование | количество, шт. |
|--------------|-----------------|
| Велосипед | 1927 |
| Мяч | 1213 |
| Рюкзак | 1199 |
| Носки | 1115 |
| Сумка | 1084 |
| Бейсболка | 1074 |
| Кроссовки | 861 |

3 кластер (personal_coef < 0,45)

| наименование | количество, шт. |
|--------------|-----------------|
| Велосипед | 7509 |
| Рюкзак | 5310 |
| Бейсболка | 4726 |
| Носки | 4223 |
| Кроссовки | 4174 |
| Сумка | 4071 |
| Палатка | 3999 |

дисконт не повлиял ни на общую сумму продаж ни на число покупателей, во всех кластерах покупателей

Модель склонности клиента к покупке

| модель | Ассурасу на трейне | Ассурасу на тесте |
|-------------------------|--------------------|--------------------|
| логистическая регрессия | 0.8615009151921904 | 0.8614581685908588 |
| случайный лес | 0.863399091587011 | 0.8611418630396963 |

Так как модели предсказывают на данных которые не видели так же как на данных для обучения, то можно сделать вывод что модели не переобучились и предсказывают корректно.