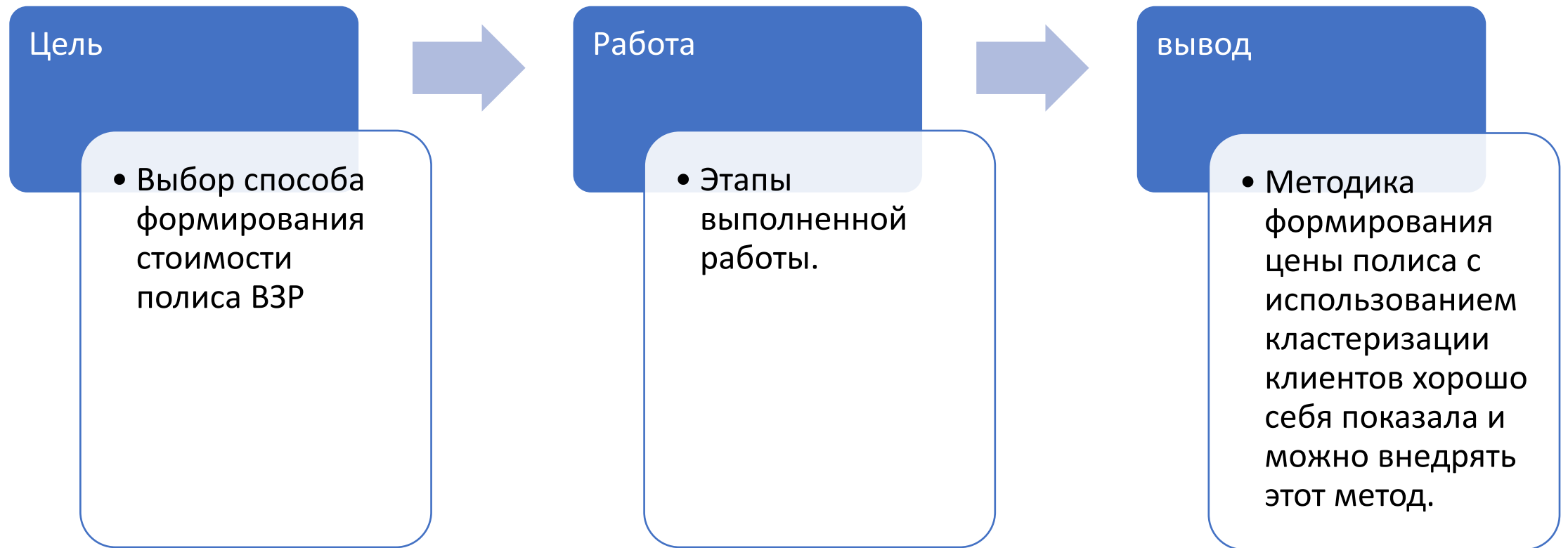


Деятельность страховой компании



Этапы работы

1. Загрузка датафрейма



2. Проверяем пропуски в данных и устраняем их.
3. Кодлируем столбцы, которые имеют не числовой формат.
4. Делаем предсказания аномалий.
5. Получаем датафрейм со столбцом аномалии.



6. Кластеризация клиентов



7. A/B-тестирование



8. Вывод.

Выбор методов при кластеризации

Выбор между методами снижения размерности t-SNE (t-distributed Stochastic Neighbor Embedding) и PCA (Principal Component Analysis) .
Так как не известны , что данные имеют линейные зависимости и метод PCA на больших данных менее производителен , то выбираем метод t-SNE

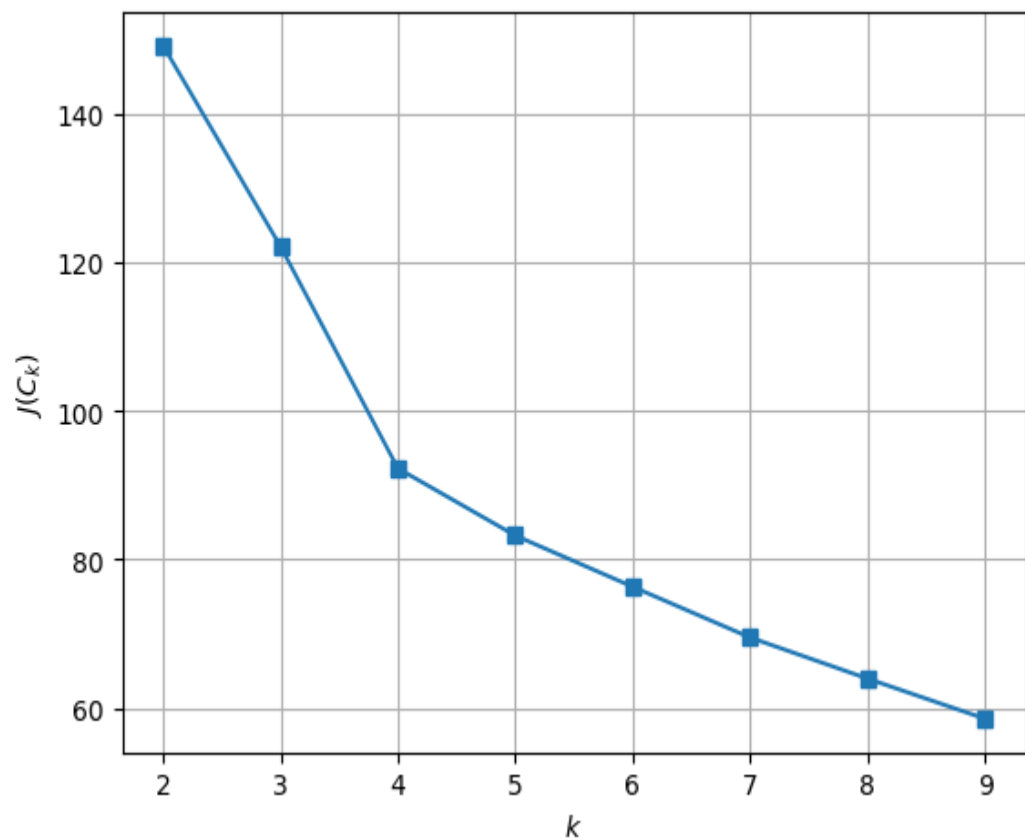
Выбор между методами Agglomerative Clustering и KMeans
Так как мы изначально задаем количество кластеров равное 4-м (известное количество кластеров) , то KMeans является предпочтительным выбором из-за его относительной простоты и скорости.

Кластеризация клиентов

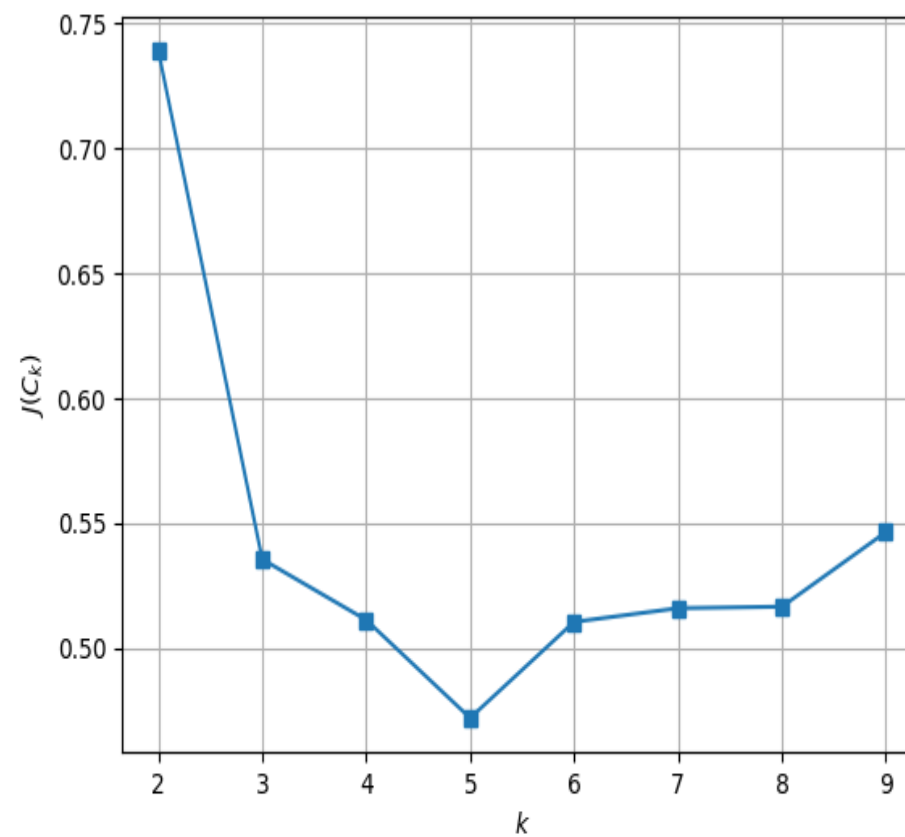


Визуализация

“метод локтя”

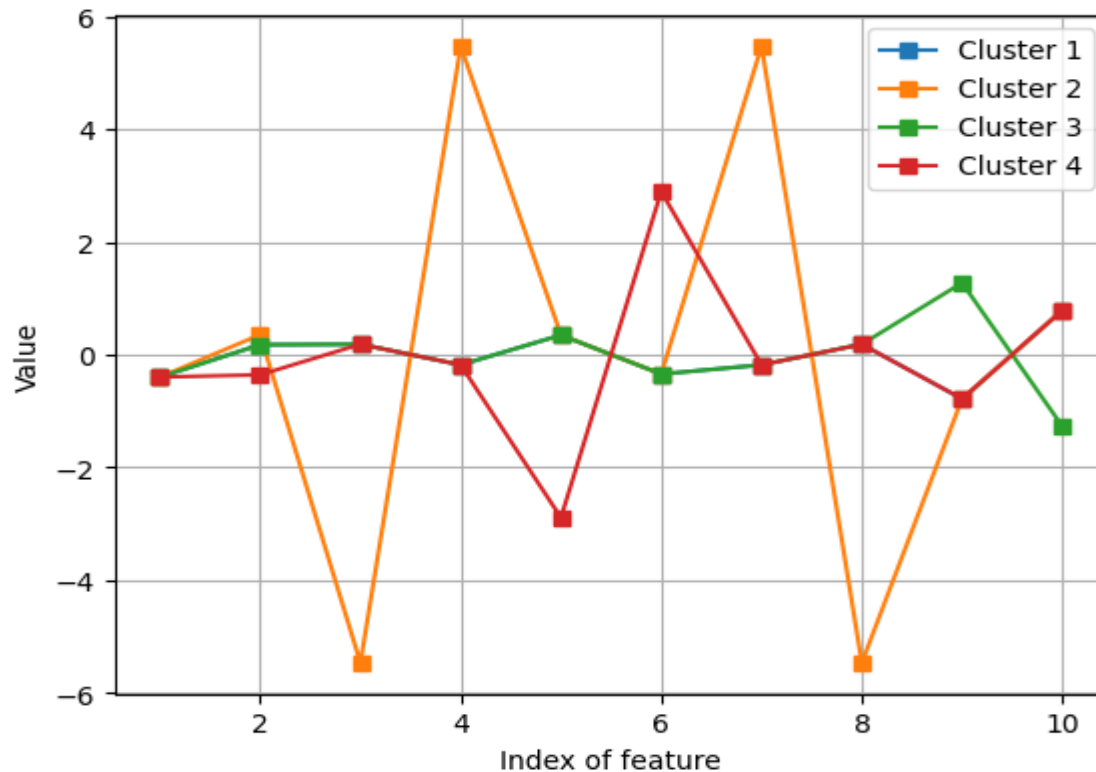


Коэффициент “силуэта”



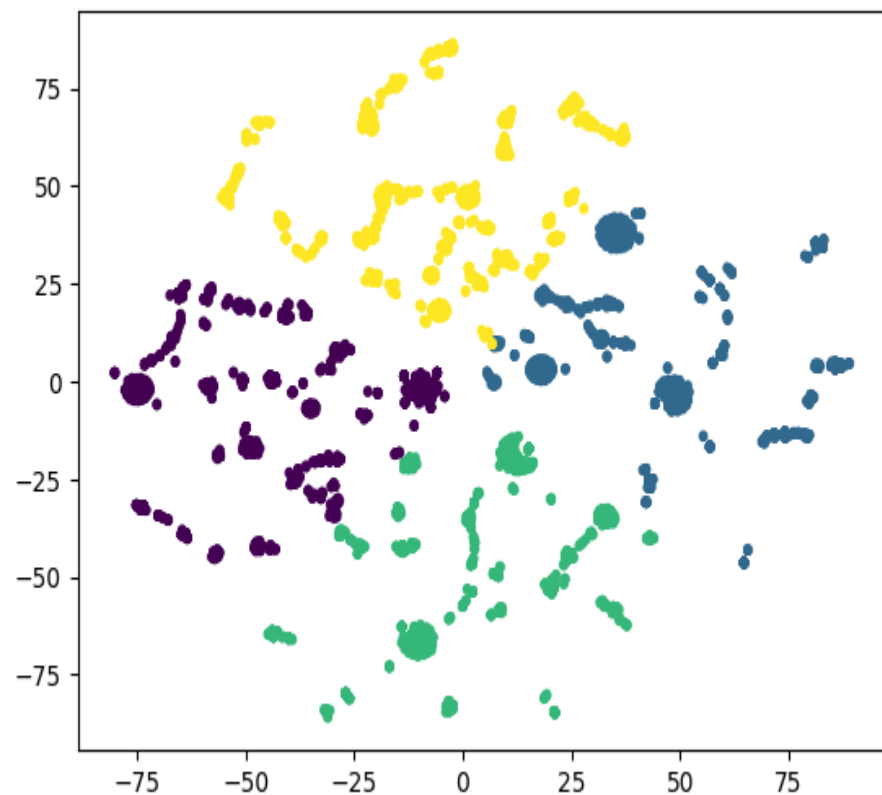
Визуализация.

- Визуализация кластеров по признакам.

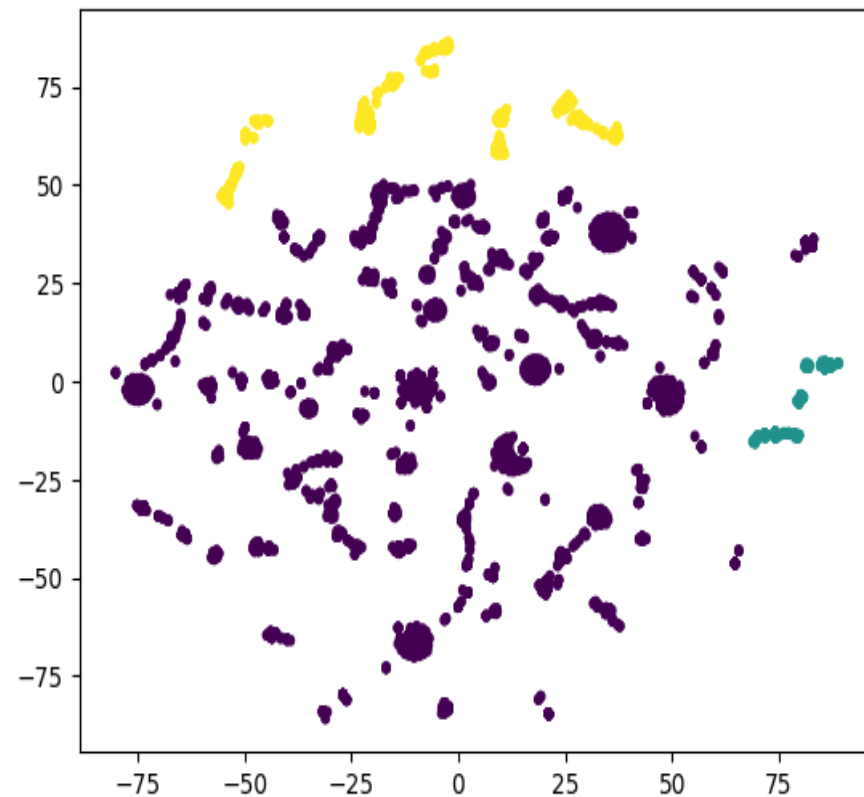


Визуализация кластеров

4-е кластера



3-и кластера



Выводы по кластерам

Важные признаки для категориальных признаков contract_num (номер контракта) и country (страна проживания клиента) , для числовых признаков duration (длительность действия полиса) и age (возраст клиента)

A/B тестирование.

1. Цель теста это найти ответ на вопрос какой подход лучше: традиционная оценка рисков или второй — использование результатов кластеризации.



2. Основные влияющие факторы: цена полиса, конверсия в оформления и убыточность.



3. Результаты теста

Результаты А/В теста

- для контрольной группы средняя цена полиса 51.93\$, а наиболее частая цена 47.95\$. для тестовой группы средняя цена полиса 65.34\$, а наиболее частая цена 13.7\$. Что говорит у тестовой группы цена полиса более дешевая для большинства клиентов, а у выборочных клиентов цена высокая. Несмотря , что количество клиентов в тестовой группе на 282 меньше, сумма за полисы на 7720\$ больше.
- конверсия контрольной группы 22.8%
- конверсия тестовой группы 26.4%
- убыток контрольной группы -21465\$
- убыток тестовой группы -16560\$

Вывод по всей работе.

- По всем вышеперечисленным признакам можно сделать вывод , что методика формирования цены полиса с использованием кластеризации клиентов хорошо себя показала и можно внедрять этот метод.