

# Отчет

В данном проекте были исследованы данные о клиентах страховой компании, купленных полюсах страхования и убытках компании с целью увеличения прибыли. Во время исследования были произведены следующие операции:

1. Выгрузка и объединения набора данных
2. Повышение качества данных
3. Кластерный анализ
4. A/B – тест и анализ результатов

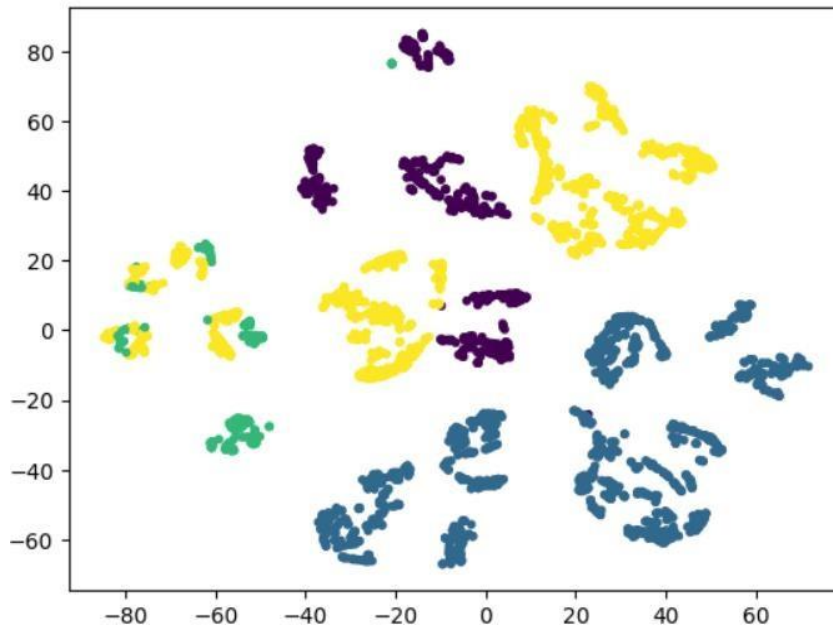
В результате чего были получены неоднозначные результаты:

Люди стали реже оформлять страховку, но средняя цена приобретенного страхования повысилась. Данные оказались не самые подходящие для кластеризации. Повтор эксперимента с увеличением количества наблюдений может быть хорошей идеей.

Подробное описание работы:

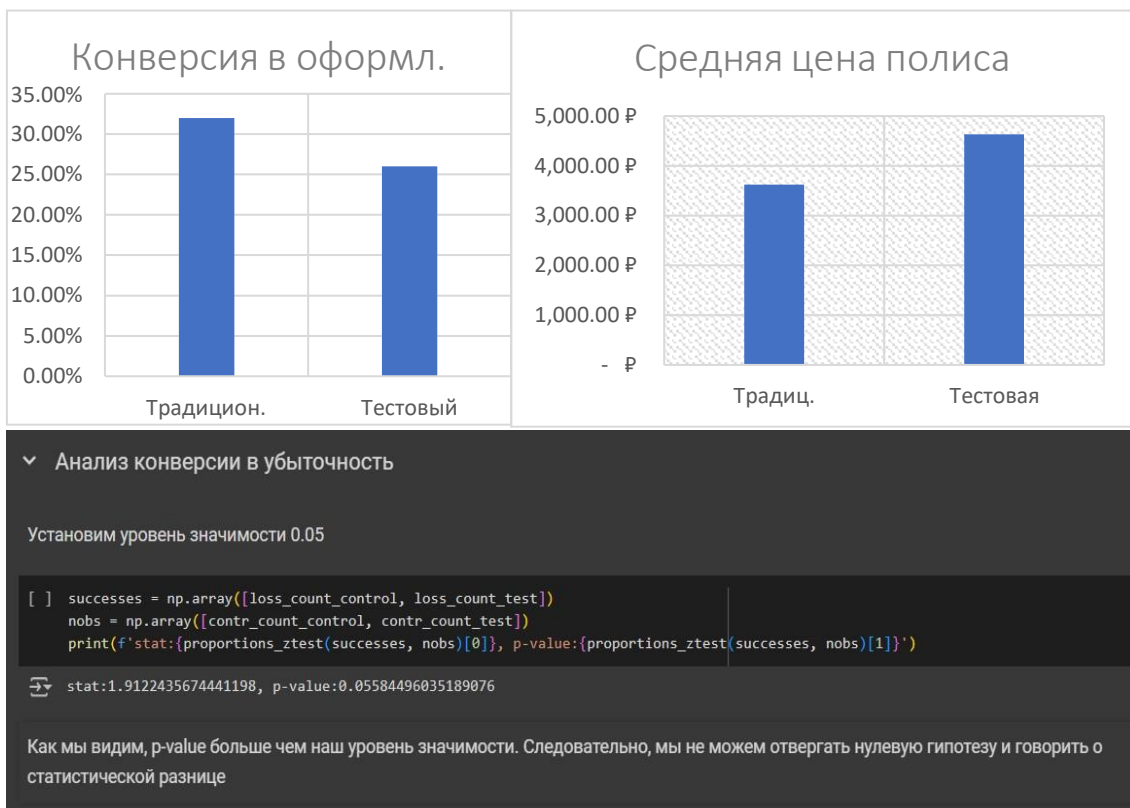
1. Были получены разношерстные данные о страховой компании, ее клиентах, купленных полюсах страхования и убытках. Данные были объединены в один набор данных и была начата чистка данных.
2. В процессе улучшения качества данных выяснилось, что у некоторого количества записей цена была указана в другой валюте. Это могло создать ошибки при анализе и неверную интерпретацию данных, поэтому долларовые были перенесены в отдельный признак. Так же были заполнены пропуски в колонке убытков. Далее данные были закодированы, после чего отмечены аномальные записи. Выполнено снижение размерности методом TSNE.
3. Был начат кластерный анализ. Использовались методы Локтя и silhouette score для выбора оптимального числа кластеров (4 кластера). Модель формировалась методами Иерархической кластеризации,

методом К-средних и KPrototypes.



На графике  
изображен  
результат  
кластериза  
ции

4. После этого было проведено А/В тестирование старого и нового подходов к формированию стоимости полиса ВЗР, где в контрольной группе были данные с традиционным подходом ценообразования, а в тестовой - подход с использованием результатов кластеризации. Главными метриками были выбраны:
  - Конверсия в оформленность полиса
  - Конверсия в выплату по страховке
  - Средняя цена полиса
5. Было проведено сравнение вышеперечисленных метрик. Получены следующие результаты: в новой модели (тестовой группе) статистически значимо меньше конверсия в оформление, но, зато, больше средняя цена приобретенного страхования, и разница статистически значима. Конверсия в убыток же не дала статистически значимых результатов.



## Интерпретация результатов:

Люди стали реже оформлять страховку, но средняя цена приобретенного страхования повысилась, так что можно сказать, что новая модель ценообразования дает статистические плоды и справляется со своей функцией, пусть и немного ухудшает другие метрики.

Важно отметить, что данные довольно сложные и плохо поддаются кластеризации, с этим и связаны такие двоякие результаты. Вероятнее всего, это происходит из-за того, что люди разных возрастов из разных стран покупают страховки, но далеко не все калечатся и получают выплаты. Тут даже интуитивно сложно выделить кластеры, т.к. не люди выбирают - получить страховку или нет.

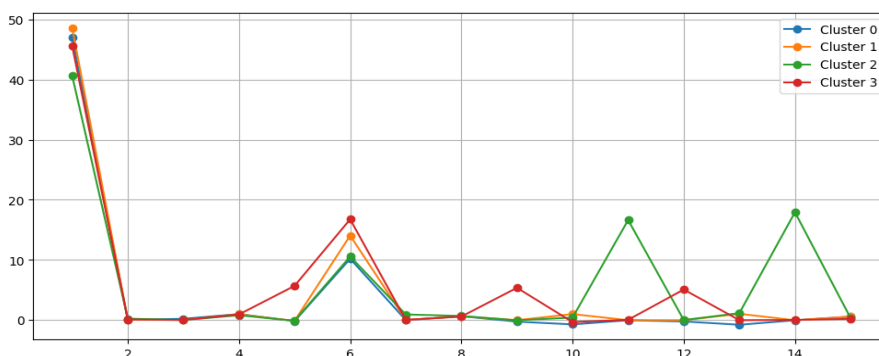


График показывает, сколь мало отличаются кластеры друг от друга (каждая линия – отдельный кластер)

## Рекомендации:

Целесообразно было бы попробовать увеличить размер выборки и повторить эксперимент, так как выплат по страховке мало, следовательно сложно определить насколько статистически значимы различия в конверсии в убыток.

#### Заключение:

Была разработана модель с помощью кластеризации и полный анализ результатов с помощью A/B тестирования. Были выявлены следующие результаты:

- Конверсия в оформленность при традиц. методе – 32%, при экспериментальн. методе – 26%.
- Средняя цена полиса при традиц. методе – 3623 руб., при экспериментальн. методе - 4636 руб.
- Конверсия в выплату по страховке статистически не изменилась.
- Данные плохо подходят для разбиения на кластеры.
- Стоит провести повторный эксперимент на большей выборке.