

АНАЛИТИЧЕСКИЙ ОТЧЁТ

Оптимизация подхода к формированию стоимости полиса ВЗР.

Подготовка данных и кластеризация.

Создание витрины данных.

- Загрузил данные и объединил их в одну таблицу.
- С помощью API сервиса Open Exchange Rates получил актуальный курс валют и конвертировал рубли в доллары США.
- Создал новый столбец с инициалами.
- Оставил данные только по действующим договорам и удалил ненужные столбцы.

	initials	client_id	duration	country	age	sex	price_usd	insurance_amount_usd	loss_payout_amt_usd
0	A. X. A.	10161870404	10	Беларусь	20	M	11	21089	NaN
1	З. X. M.	1017159879	7	Индонезия	55	M	20	52723	NaN
2	Б. В. Щ.	10161883357	10	Беларусь	61	F	11	21089	NaN
3	И. У. Н.	100766461	14	Грузия	25	M	16	21089	NaN
4	З. Ч. Г.	10171604758	7	Аргентина	33	M	20	52723	NaN

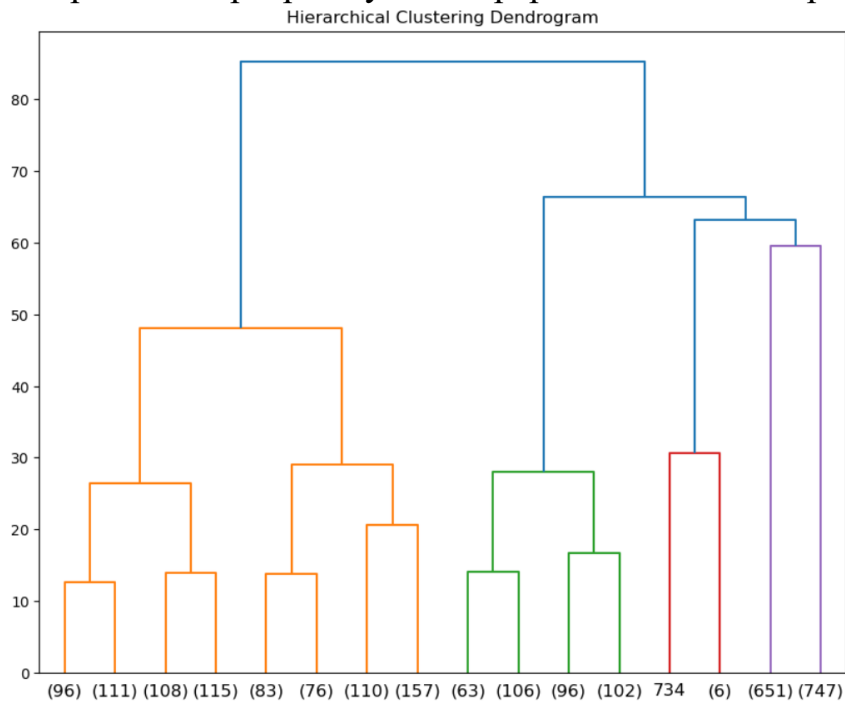
Предобработка данных.

- Заменял пропуски в столбце с потерями на 0 и изменял тип данных.
- Закодировал категориальные признаки с помощью LabelEncoder.
- С помощью IsolationForest нашёл и удалил аномалии.
- Заменял выбросы в данных 95 перцентилем.
- Удалял инициалы и айди клиентов.

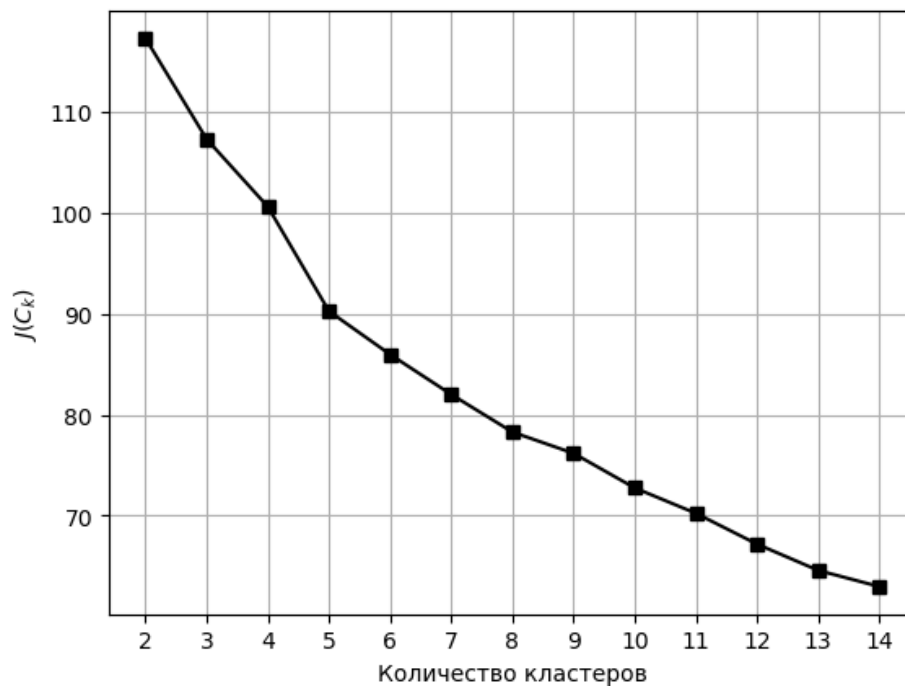
	duration	country	age	sex	price_usd	insurance_amount_usd	loss_payout_amt_usd
0	10	3	20	1	11	21089	0
1	7	9	55	1	20	52723	0
3	14	7	25	1	16	21089	0
5	30	15	33	0	52	31634	0
6	30	15	38	1	52	31634	0

Кластеризация.

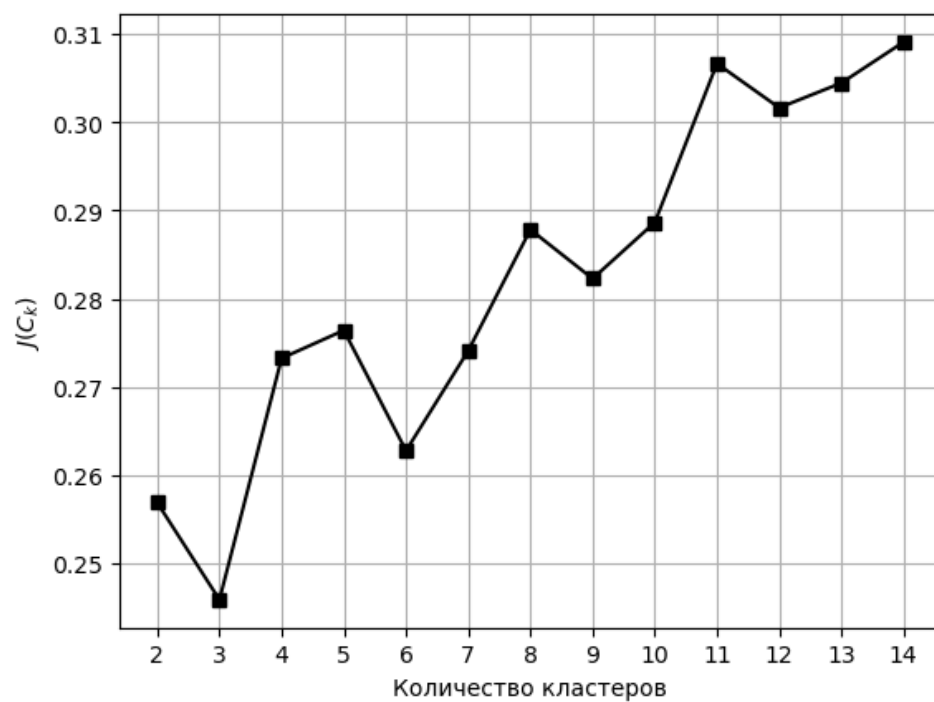
- Стандартизировал данные.
- Построил дендрограмму для иерархической кластеризации.



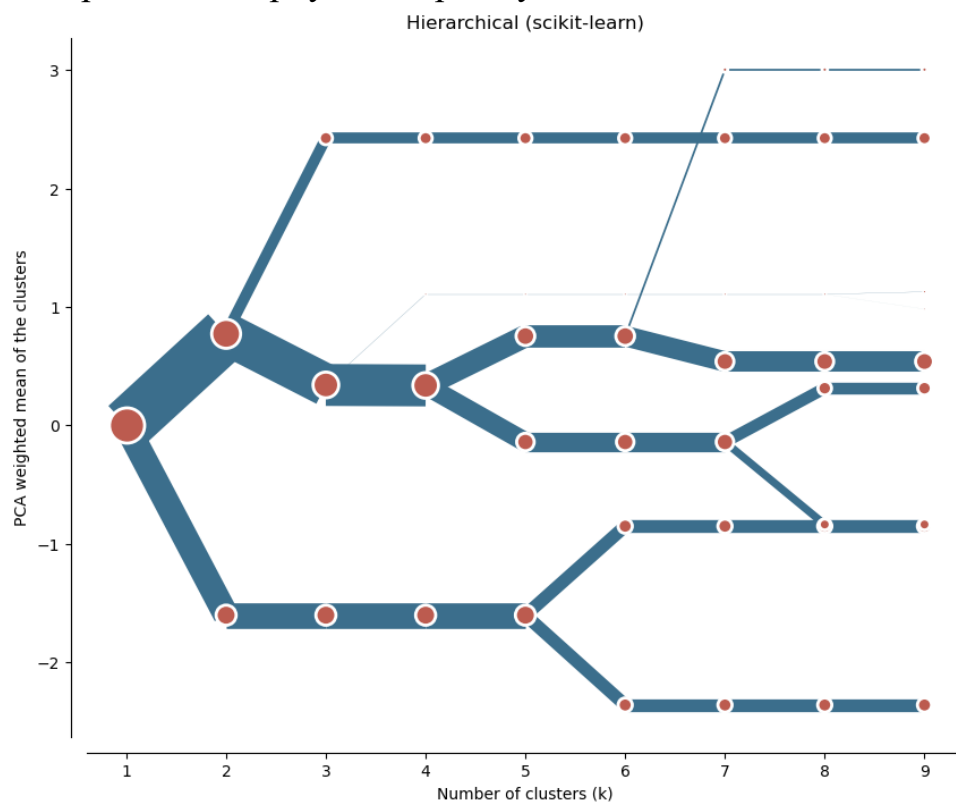
- Вычислил значения инерции и силуэта для различного количества кластеров.
- График инерции:



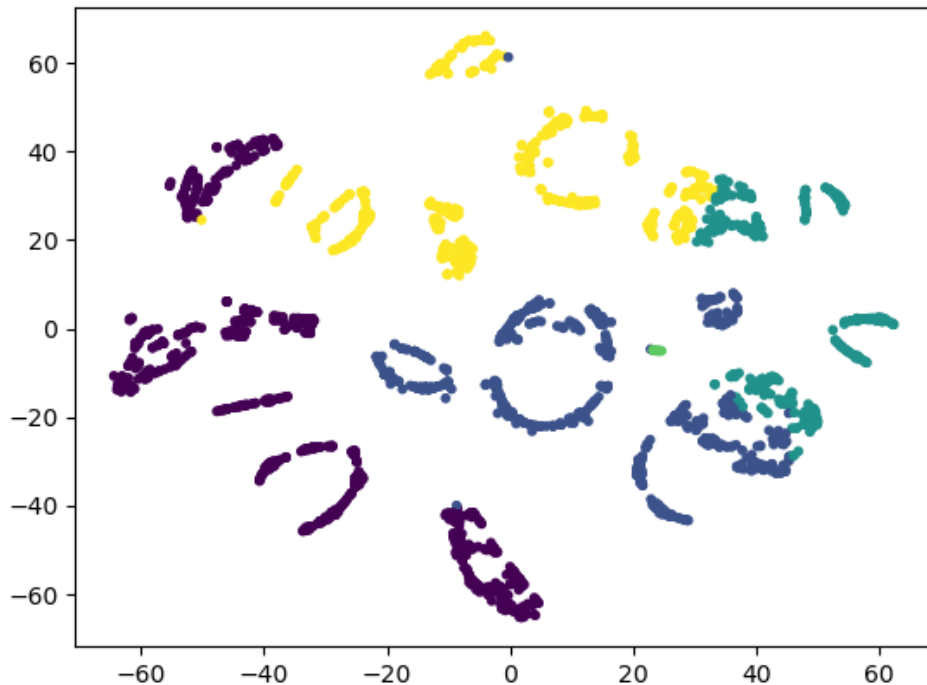
- График силуэта:



- Построил кластерную диаграмму.



- Визуализировал разбиение на 5 кластеров с помощью Агломеративной кластеризации.



- Решил выбрать разбиение на 5 кластеров как самое оптимальное.

Анализ А/В-теста.

Подготовка данных.

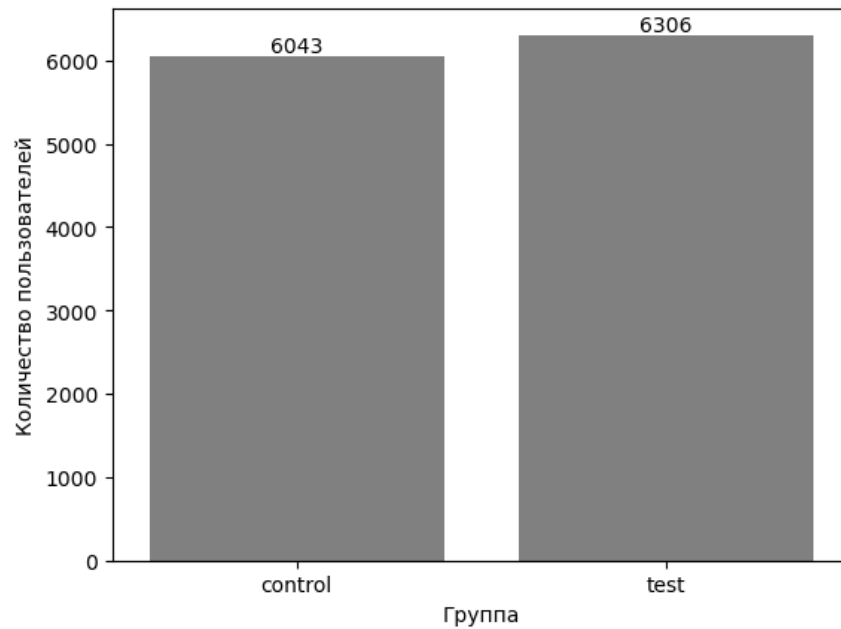
- Загрузил и объединил дата сеты.
- Конвертировал рубли в доллары.
- Создал флаги для потерь и конверсий.
- Заменял выбросы 95 перцентилем.
- Удалил лишние столбцы.
- Удалил данные, не относящиеся к контрольной или тестовой группе.
- Заполним пропуски в числовых признаках значением 0 и изменим тип данных.
- Сгруппировал данные.

	group	client_id	price_usd	insurance_amount_usd	loss_payout_amt_usd	loss_cnt	converted	conversion_rate	avg_price	avg_insurance_amount	avg_loss_payout_amt
0	control	6043	65064	92337500	79750	30	1949	0.322522	33.383273	47376.859928	2658.333333
1	test	6306	57982	64662500	92375	14	1667	0.264351	34.782244	38789.742052	6598.214286

Проверка гипотез.

1. Равенство пропорций

- H_0 : Наблюдаемые пропорции выборок соответствуют ожидаемым теоретическим пропорциям.
- H_1 : Наблюдаемые пропорции выборок не соответствуют ожидаемым теоретическим пропорциям.



С помощью критерия хи-квадрат сравнил выборки.

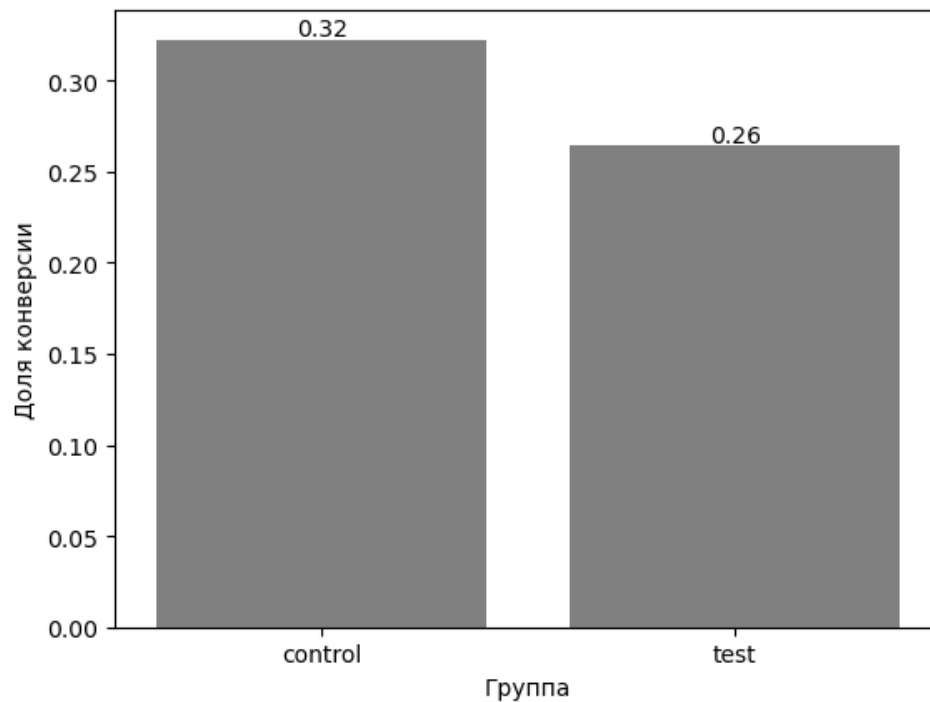
p-value равно 0.017948361469529415

На уровне значимости 0.01 принимаем нулевую гипотезу.

Пропорции клиентов в группах не отличаются.

2. Конверсия.

- H_0 : Конверсия контрольной группы равна конверсии тестовой группы.
- H_1 : Конверсия контрольной группы не равна конверсии тестовой группы.



Провёл z-тест на равенство пропорций.

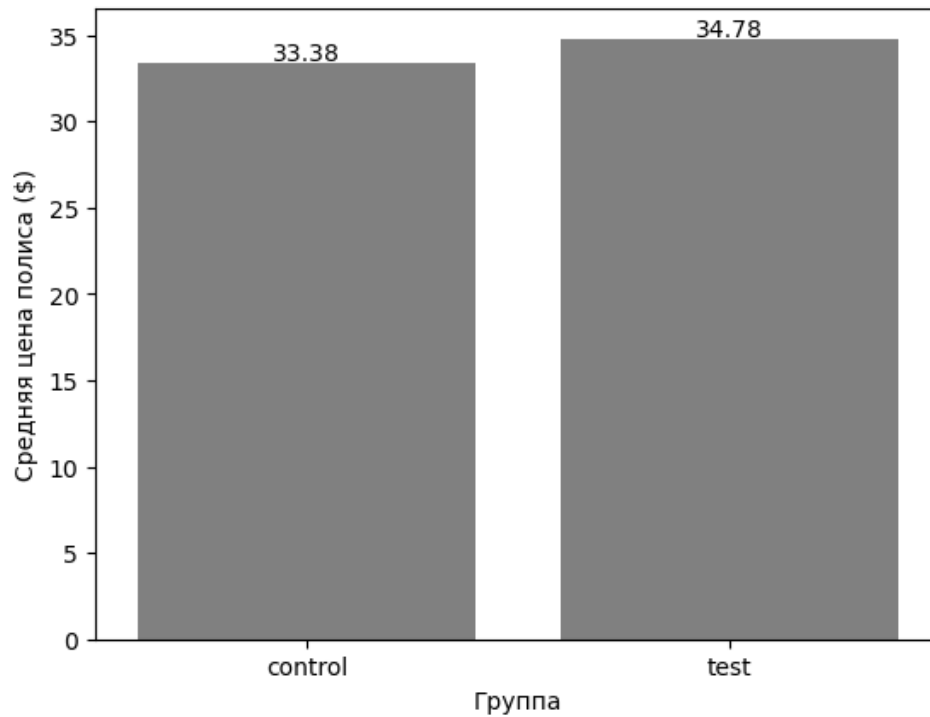
p-value равно $1.2376025494738044 \times 10^{-12}$

На уровне значимости 0.01 отвергаем нулевую гипотезу.

Конверсия в группах различается.

3. Цена полиса.

- H_0 : Средняя цена полиса для контрольной и тестовой группы не отличается.
- H_1 : Средняя цена полиса для контрольной и тестовой группы отличается.



С помощью t-теста сравнил цены.

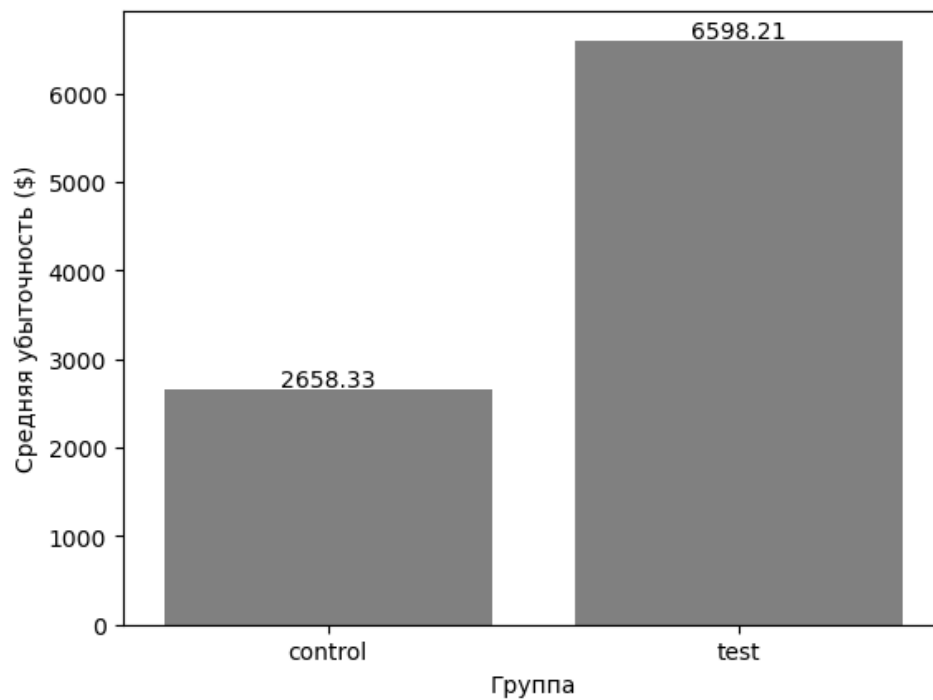
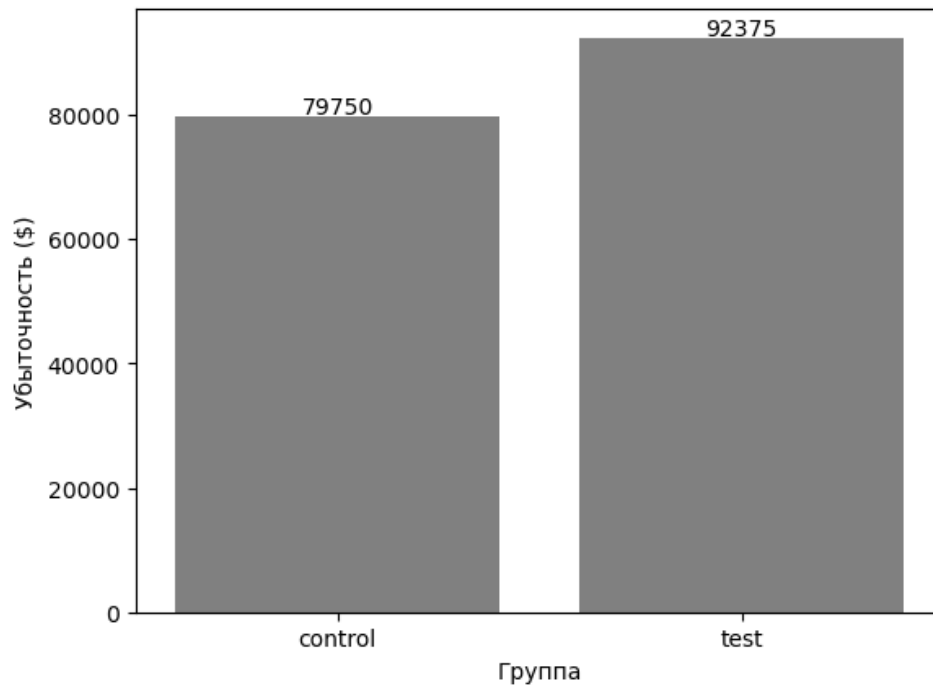
p-value равно 0.08169860997398731

На уровне значимости 0.01 принимаем нулевую гипотезу.

Цены полиса в группах не отличаются.

4. Убыточность.

- H_0 : Убыточность контрольной и тестовой группы не отличается.
- H_1 : Убыточность контрольной и тестовой группы отличается.



С помощью t-теста сравнил убыточность в группах.

p-value равно 0.6156632574348682

На уровне значимости 0.01 принимаем нулевую гипотезу.

Тест не позволяет однозначно утверждать, что убыточность в группах различается.

Выводы:

- Конверсия в оформлениях в тестовой группе на 6.0% ниже, чем в контрольной группе.
- Тест не показал статистически значимой разницы цен между группами.
- Тест не показал статистически значимой разницы убыточности между группами.

Рекомендация:

Новый кластерный подход, похоже, не дает каких-либо существенных преимуществ по сравнению с традиционным методом оценки рисков. Учитывая снижение коэффициента конверсии и отсутствие явных улучшений в ценообразовании и убытках, переход на кластерный подход может оказать негативное влияние на бизнес, потенциально снижая доходы и доверие клиентов. Поэтому рекомендуется продолжать использовать традиционный подход к оценке рисков для определения стоимости полисов ВЗР. Если компания по-прежнему заинтересована в изучении кластерного подхода, потребуются дальнейшие исследования и уточнения, прежде чем рассматривать его внедрение.