

ТЕЛЕКОМ. ПОИСК КЛИЕНТОВ С НЕОПТИМАЛЬНЫМИ ТАРИФАМИ

Выпускной проект

Серкибаев Аким

ЦЕЛЬ ПРОЕКТА

Определить клиентов с неоптимальными тарифами, предложить рекомендации по смене тарифного плана, чтобы оптимизировать расходы клиентов и минимизировать возможные потери выручки компании при переходе на более выгодные тарифы



ПРЕДОБРАБОТКА ДАННЫХ

Обзор данных

Датасет data:
53902 строки и 9 столбцов
Датасет clients:
732 строки и 3 столбца

Пропуски

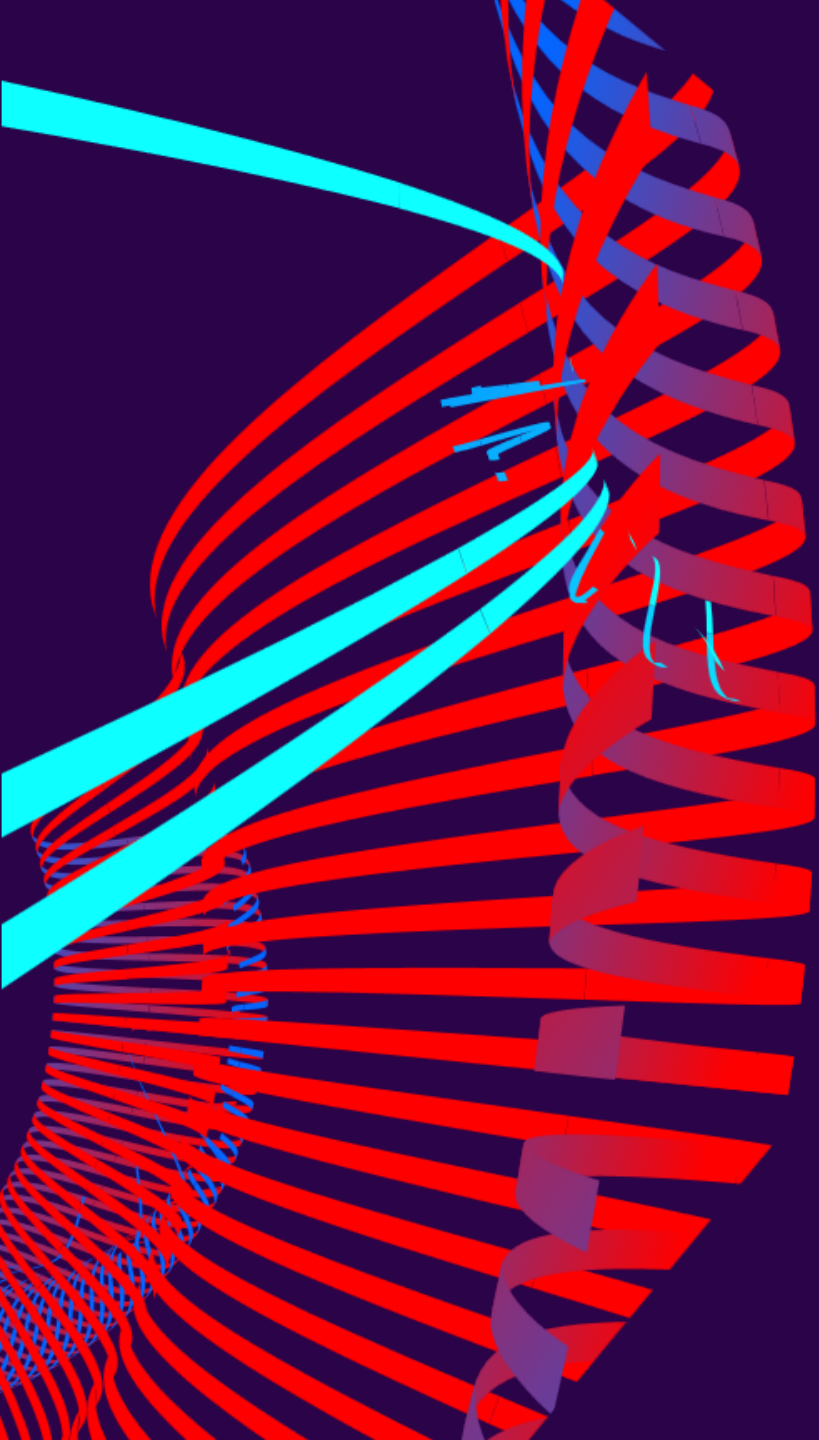
Пропуски в двух столбцах
заполнены на основе анализа
данных, строки не удалялись

Типы данных

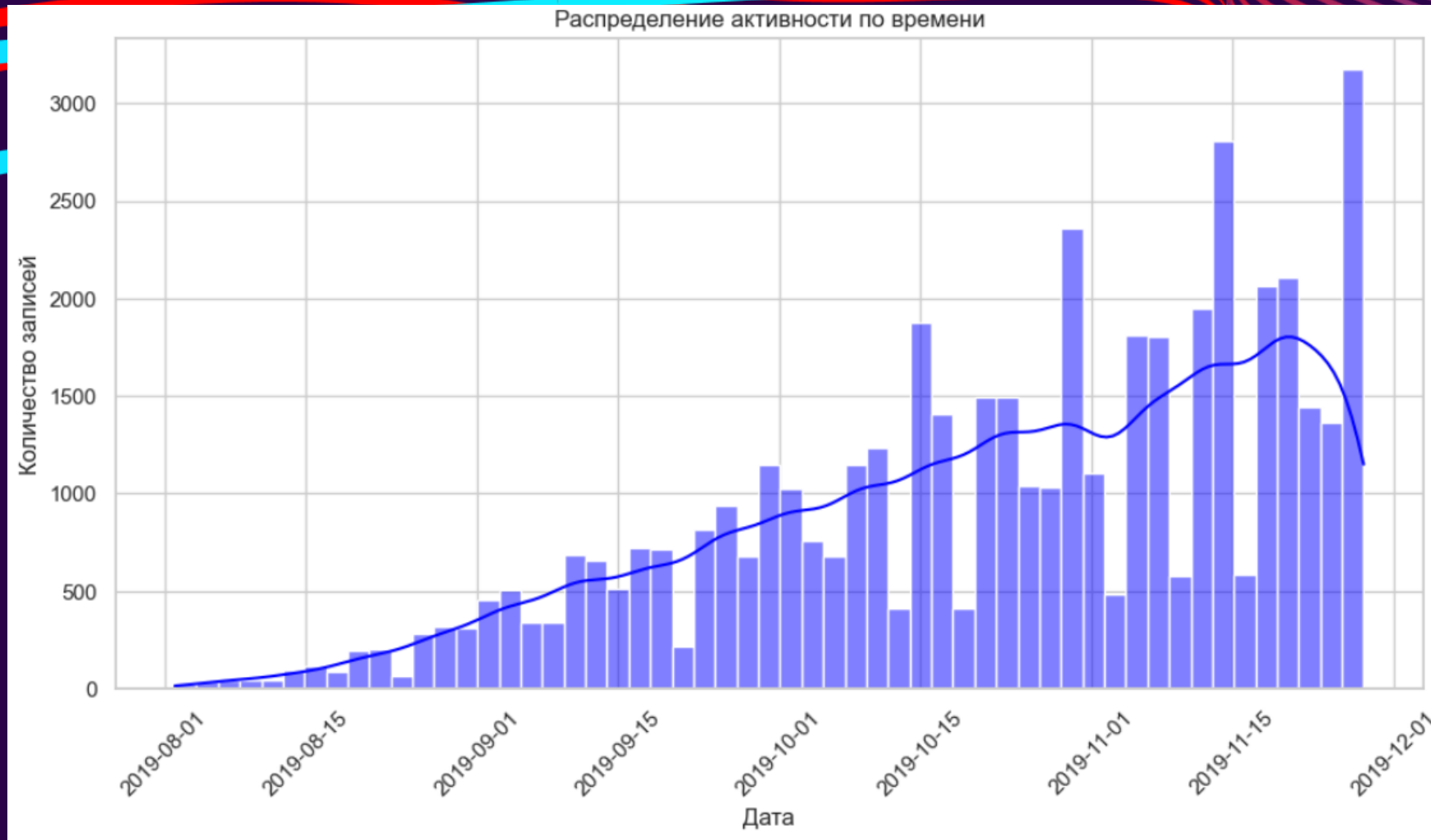
- Столбцы с датой преобразованы в тип datetime.
- Часовой пояс был "обнулён".
- Остальные столбцы проверены на корректность типов

Дубликаты, выбросы, аномалии

- Удалено 4900 явных дубликатов из data.
- Неявных дубликатов не обнаружено
- Удалены только экстремальные значения, превышающие 99-й перцентиль



ИССЛЕДОВАТЕЛЬСКИЙ АНАЛИЗ ДАННЫХ



Мы проанализировали данные за август–ноябрь 2019 года. Максимальное количество уникальных клиентов (273 человека, ~89%) зарегистрировано в октябре, что делает его оптимальным периодом для дальнейшего исследования

РАСЧЁТ ТЕКУЩИХ РАСХОДОВ КЛИЕНТОВ

Анализ показал, что тариф С имеет максимальные суммарные расходы из-за отсутствия лимитов и высоких цен. Тариф В занимает второе место, предлагая сбалансированное соотношение стоимости и включенных услуг. Тариф А, несмотря на высокую плату, показывает минимальные расходы

	user_id	over_limit_cost	external_call_cost	total_operator_cost	tariff_plan	monthly_fee	total_expenses
0	166377	0.00	40799.5	18600	B	2000	61399.50
1	166391	0.00	0.0	0	C	1000	1000.00
2	166392	0.00	0.0	5400	C	1000	6400.00
3	166405	0.00	141570.0	39150	B	2000	182720.00
4	166406	8.25	36588.5	12900	B	2000	51496.75
...
151	167744	0.00	0.0	300	C	1000	1300.00
152	167747	12.60	2011.8	2700	C	1000	5724.40
153	167756	0.00	16333.8	13800	C	1000	31133.80
154	167781	0.00	0.0	4800	C	1000	5800.00
155	167792	0.00	0.0	1800	C	1000	2800.00

156 rows × 7 columns

	tariff_plan	total_expenses
0	A	225639.5
1	B	1626165.0
2	C	2491471.6

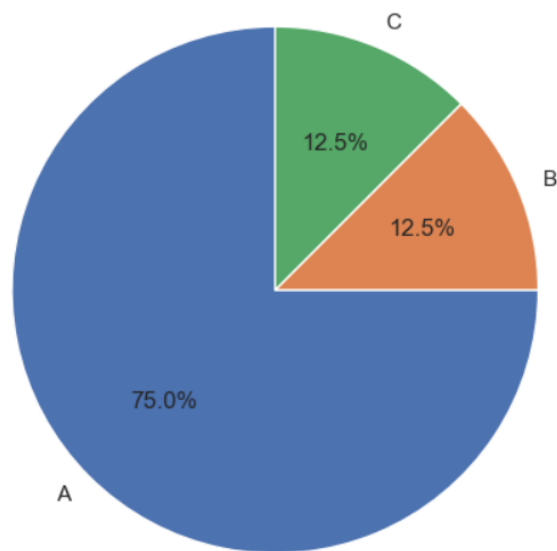
СРАВНЕНИЕ ОПТИМАЛЬНОГО ТАРИФА И ЕГО РАСХОДЫ

	user_id	tariff_plan	total_expenses	total_expenses_A	total_expenses_B	total_expenses_C	optimal_tariff	optimal_expenses	savings
0	166377	B	61399.50	34679.7	61399.50	95319.9	total_expenses_A	34679.7	26719.80
1	166391	C	1000.00	4000.0	2000.00	1000.0	total_expenses_C	1000.0	0.00
2	166392	C	6400.00	4900.0	4700.00	6400.0	total_expenses_B	4700.0	1700.00
3	166405	B	182720.00	101992.0	182720.00	277579.3	total_expenses_A	101992.0	80728.00
4	166406	B	51496.75	30253.1	51496.75	78198.8	total_expenses_A	30253.1	21243.65
...
151	167744	C	1300.00	4050.0	2150.00	1300.0	total_expenses_C	1300.0	0.00
152	167747	C	5724.40	5312.2	4787.00	5724.4	total_expenses_B	4787.0	937.40
153	167756	C	31133.80	13300.2	20567.00	31133.8	total_expenses_A	13300.2	17833.60
154	167781	C	5800.00	4800.0	4400.00	5800.0	total_expenses_B	4400.0	1400.00
155	167792	C	2800.00	4300.0	2900.00	2800.0	total_expenses_C	2800.0	0.00

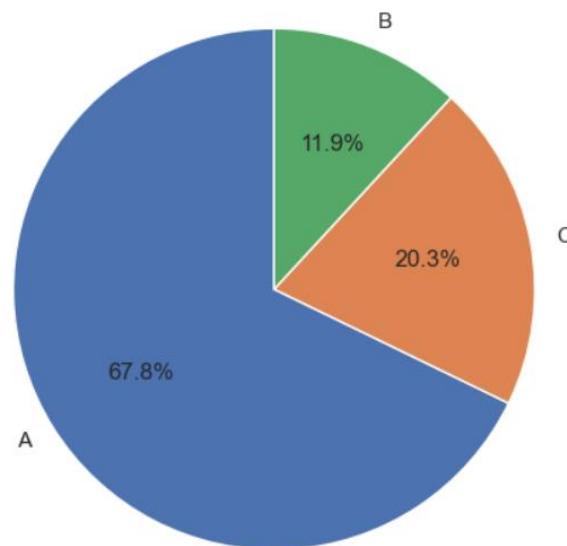
56 rows × 9 columns

РАСПРЕДЕЛЕНИЕ ОПТИМАЛЬНЫХ ТАРИФОВ

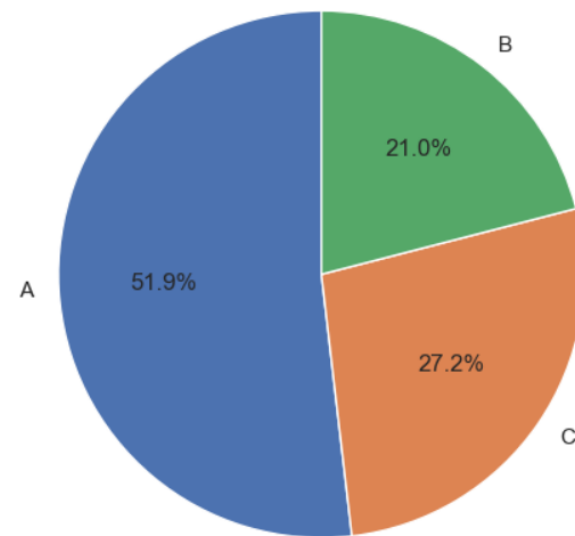
Распределение оптимальных тарифов для текущего тарифа А



Распределение оптимальных тарифов для текущего тарифа В



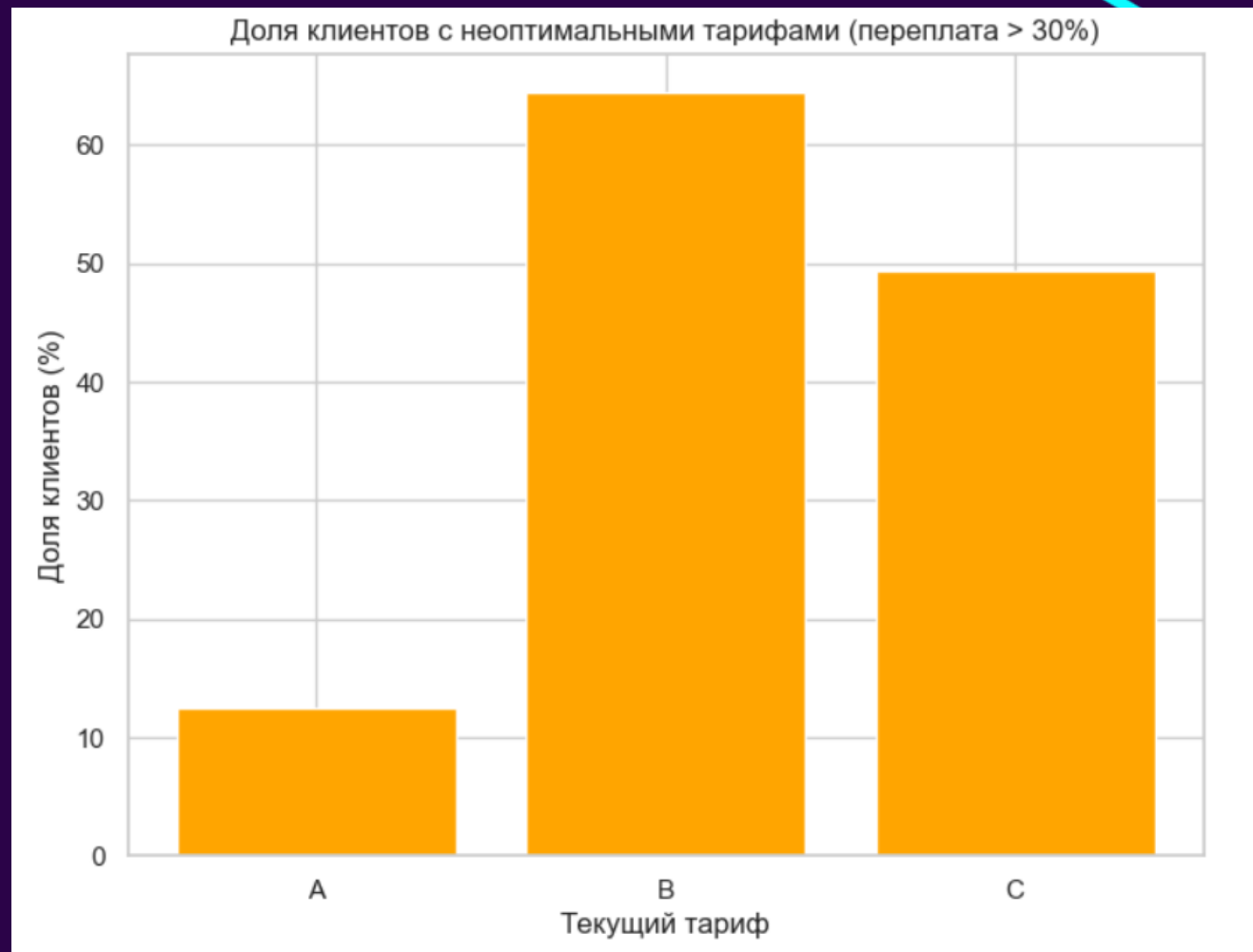
Распределение оптимальных тарифов для текущего тарифа С



Анализ показал, что многие клиенты могли бы снизить расходы, перейдя на оптимальные тарифы. Например, 67,8% клиентов тарифа В предпочли бы тариф А для экономии. Это подчеркивает необходимость пересмотра тарифной структуры для лучшего удовлетворения потребностей клиентов

НЕОПТИМАЛЬНЫЕ ТАРИФЫ

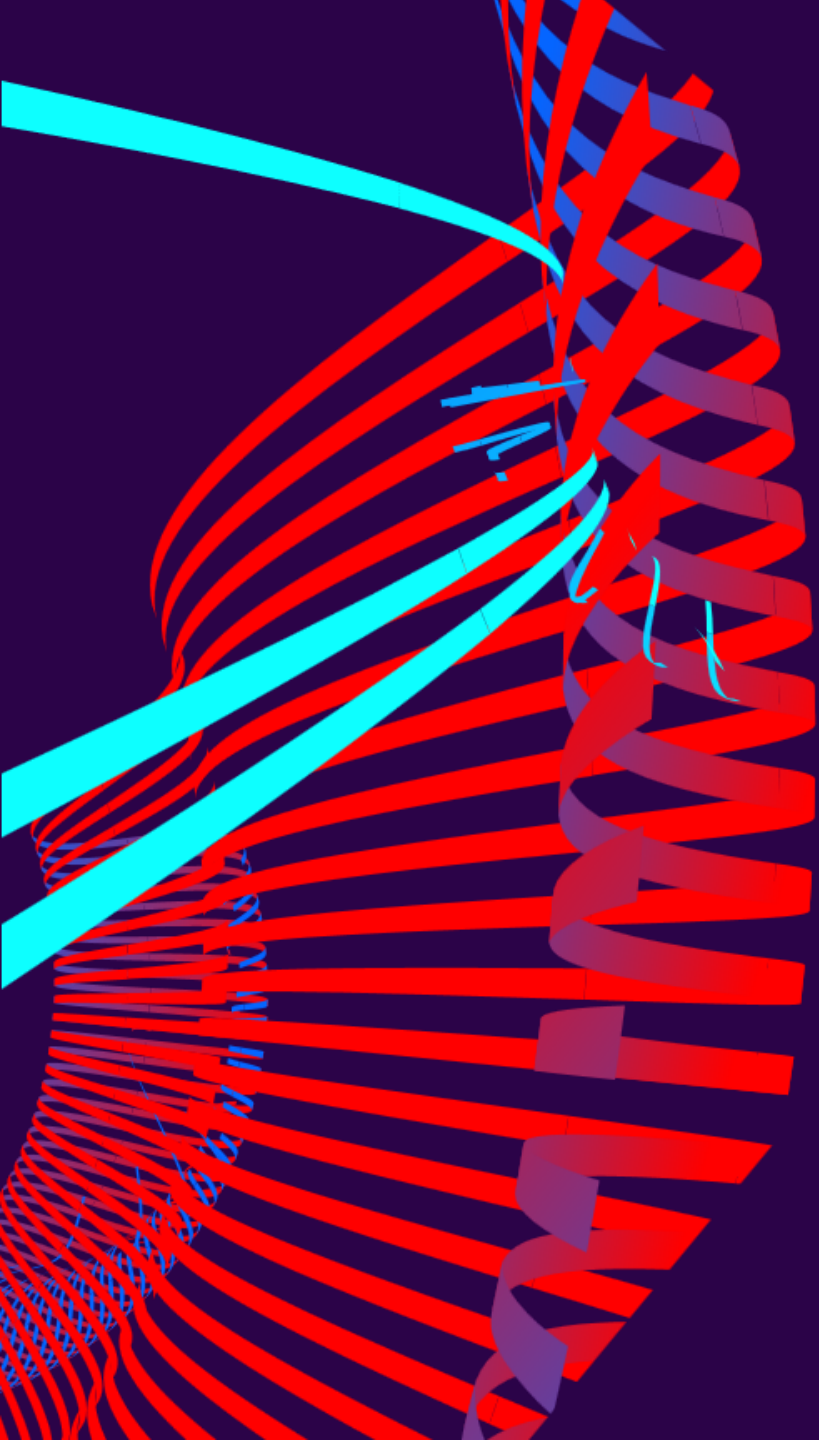
Для определения неоптимальных тарифов выберем порог превышения фиксированного процента, установив его на уровне 30%. Это решение учитывает цель проекта — предложить клиентам оптимальные тарифы без значительных потерь выручки



КЛИЕНТЫ, СУЩЕСТВЕННО СНИЖАЮЩИЕ ВЫРУЧКУ КОМПАНИИ

	tariff_plan	total_expenses	optimal_expenses	savings	client_count	avg_savings_per_client
0	A	97059.4	97059.4	0.0	2	0.000000
1	B	1049196.5	569841.2	479355.3	11	43577.754545
2	C	1908941.8	643441.8	1265500.0	18	70305.555556

Анализ показал, что тариф С обеспечивает максимальную экономию, превышающую 1,2 млн в общей сумме, с наибольшей средней экономией на клиента. Тариф В занимает второе место, что указывает на значительный потенциал снижения выручки при переходе клиентов. Тариф А демонстрирует минимальную экономию, подтверждая его сбалансированность и соответствие потребностям клиентов



СТАТИСТИЧЕСКИЙ АНАЛИЗ ДАННЫХ

ПРОВЕРКА ГИПОТЕЗЫ ИЗ ТЗ

Нулевая гипотеза (H_0): Затраты клиентов на услуги, включая абонентскую плату и дополнительные расходы, одинаковы для тарифных планов А и В

Альтернативная гипотеза (H_1): Затраты клиентов на услуги, включая абонентскую плату и дополнительные расходы, различаются между тарифами А и В

р-значение: 0.7759723666827334

Не получилось отвергнуть нулевую гипотезу, вывод о различии сделать нельзя

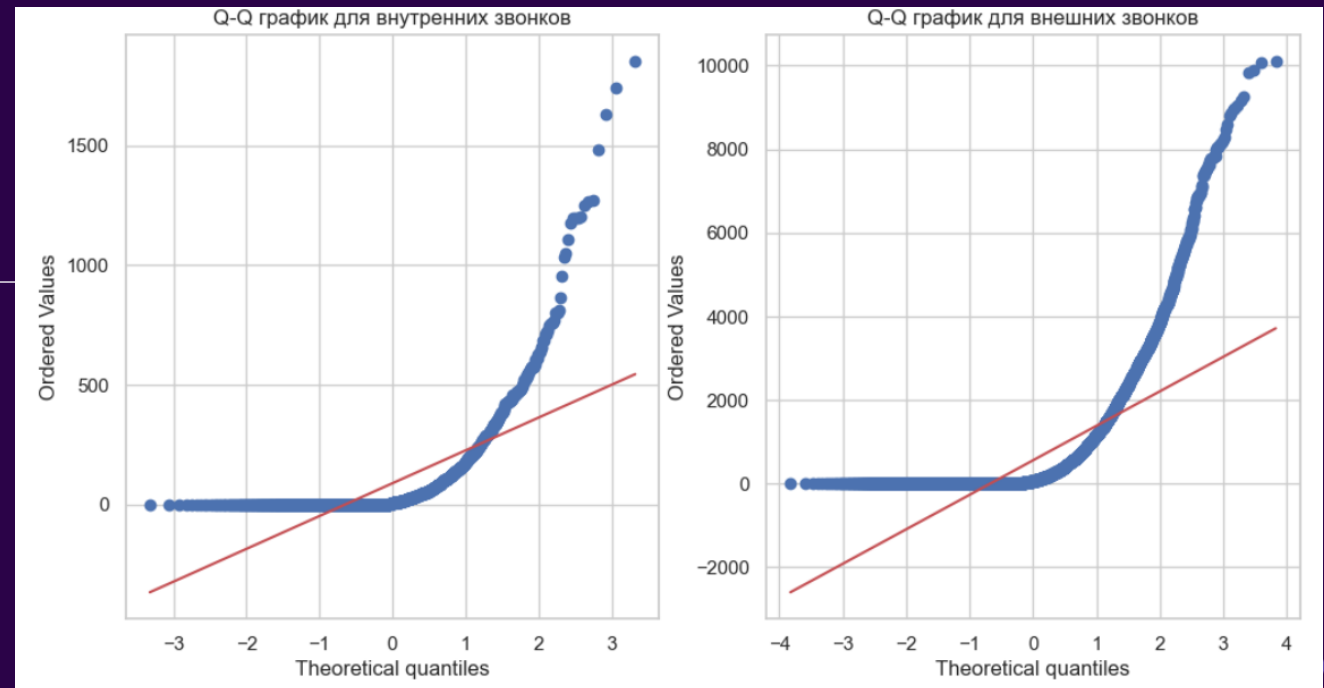
Данные проверены на нормальность - распределение не нормально
Р-значение по тесту Манна-Уитни составляет 0.776, что значительно превышает стандартный уровень значимости 0.05. Это означает, что у нас нет достаточных оснований для отклонения нулевой гипотезы. Различия в затратах клиентов между тарифами А и В статистически незначимы

ФОРМУЛИРОВКА СОБСТВЕННОЙ ГИПОТЕЗЫ

Нулевая гипотеза (H_0): Длительность звонков (call_duration) внутри сети и вне сети в среднем одинакова

Альтернативная гипотеза (H_1): Длительность звонков (call_duration) внутри сети и вне сети в среднем различается

Данные не нормальны (Q-Q плот).
Тест Манна-Уитни, который показал p-значение ($1.06e-53$), меньше стандартного уровня значимости 0.05, что позволяет отвергнуть нулевую гипотезу и сделать вывод, что средняя длительность звонков внутри сети и вне сети статистически значимо различается





ИТОГОВЫЙ ВЫВОД

Проведён анализ данных, включающий предобработку, исследовательский и статистический анализ. Выявлены ключевые особенности тарифных планов, активности клиентов и влияния изменений тарифов на выручку компании.

Разработаны рекомендации по оптимизации тарифной структуры для снижения потерь выручки. Статистический анализ подтвердил равенство затрат клиентов на услуги для тарифных планов А и В, а также значимые различия в средней длительности звонков внутри и вне сети.

**БЛАГОДАРЮ
ЗА ВНИМАНИЕ!**

ТЕЛЕКОМ. ПОИСК КЛИЕНТОВ С НЕОПТИМАЛЬНЫМИ ТАРИФАМИ

Выпускной проект

Серкибаев Аким