

DAN



ВЫСШАЯ ШКОЛА
ЭКОНОМИКИ



ТИНЬКОФФ



Уральский
федеральный
университет

ВВЕДЕНИЕ

Подписка Tinkoff Pro — платная услуга, в рамках которой клиент получает бонусы по основным продуктам Тинькофф. В первый месяц подписка бесплатна, далее — 199 ₽ в месяц.

Подписка дает множество преимуществ при использовании таких продуктов Тинькофф, как дебетовые и кредитные карты, счета и вклады, Тинькофф Мобайл, Тинькофф Путешествия, Тинькофф Инвестиции, а также специальные условия на партнерские подписки. Основное преимущество подписки заключается в финансовой выгоде. Обладатели подписки меньше платят за обслуживание, пользуются особыми лимитами и условиями начисления кэшбэка, бесплатными оповещениями и другими бонусами.

ОПИСАНИЕ ДАННЫХ

В датасете, который вы получили, представлены данные о клиентах, которые пользовались или не пользовались подпиской Tinkoff Pro. Данные собраны за период с 01.01.2022 по 01.12.2022.

Одной записью представлен уникальный клиент. Всего в базе 20 001 наблюдение.

Доступная вам база данных включает следующие переменные:

subscription_rk	Уникальное поле, идентификатор подписчика
Возраст	Целочисленное поле, отражающее возраст клиента, года
Пол	Индикатор пола клиента: F — женский, M — мужской
Кол-во детей	Целочисленное поле, отражающее количество детей у клиента
Индикатор отказа от коммуникаций	Бинарный индикатор: 1 — отказ от коммуникаций с банком 0 — коммуникации с клиентом возможны
Доход	Численное поле, отражающее доход, указанный клиентом при оформлении карты, рубли
Возраст в банке	Целочисленное поле, отражающее количество дней с момента оформления первого продукта по сегодня
Гео	Город проживания / регион, указанный клиентом при оформлении продукта
Дата подключения	Фактическая дата подключения подписки Tinkoff Pro, в формате YY-MM-DD
Дата отключения	Дата отключения подписки Tinkoff Pro, в формате YY-MM-DD. Если дата отсутствует, это означает, что подписка активна / не отключена
Активных дней	Целочисленное поле, отражающее количество дней с момента подключения подписки до
Выгода	Численное поле, отражающее общую сумму, которую клиент сэкономил/заработал благодаря опциям подписки, рубли
Кол-во опций	Целочисленное поле, отражающее, каким количеством опций в подписке пользовался клиент. Всего опций 13, пользование опцией означает, что была выгода по этой опции

Активная дебетовая карта	Бинарный индикатор: 1 — наличие у клиента дебетовой карты Tinkoff Black; 0 — отсутствие у клиента дебетовой карты Tinkoff Black
Активная кредитка	Бинарный индикатор: 1 — наличие у клиента кредитной карты Tinkoff Platinum; 0 — отсутствие у клиента кредитной карты Tinkoff Platinum
НС (накопительный счет)	Бинарный индикатор: 1 — наличие у клиента накопительного счета 0 — отсутствие у клиента накопительного счета
Активный Мобайл	Бинарный индикатор: 1 — наличие у клиента симки Тинькофф Мобайла 0 — отсутствие у клиента симки Тинькофф Мобайла
Активный брокерский счет	Бинарный индикатор: 1 — наличие у клиента брокерского счета 0 — отсутствие у клиента брокерского счета
Вклад	Бинарный индикатор: 1 — наличие у клиента вклада 0 — отсутствие у клиента вклада
Индикатор оплаты подписки	Бинарный индикатор: 1 — была хотя бы одна оплата подписки; 0 — не было оплат, только триальный период
Кол-во партнерских подписок	Целочисленное поле, отражающее количество активных подписок у партнеров (Ivi, Kaspersky, GFN, Mybook, Whoosh)
Оборот	Численное поле, отражающее общую сумму транзакций клиента по дебетовым или кредитным картам, фактически сумма среднемесячных трат клиента, рубли
Кол-во транзакций	Целочисленное поле, отражающее количество транзакций клиента по дебетовым или кредитным картам, фактически количество среднемесячных трат клиента

Транзакций кино	Целочисленное поле, отражающее количество транзакций клиента по дебетовым или кредитным картам в категории «Кино» (кинотеатры, покупки фильмов)
Транзакций развлечения	Целочисленное поле, отражающее количество транзакций клиента по дебетовым или кредитным картам в категории «Развлечения» (аттракционы, выставки и так далее)
Транзакций фастфуд	Целочисленное поле, отражающее количество транзакций клиента по дебетовым или кредитным картам в категории «Фастфуд» (рестораны быстрого питания)
Транзакций музыка	Целочисленное поле, отражающее количество транзакций клиента по дебетовым или кредитным картам в категории «Музыка» (подписки на музыкальные сервисы, покупки песен и так далее)
Транзакций рестораны	Целочисленное поле, отражающее количество транзакций клиента по дебетовым или кредитным картам в категории «Рестораны» (оплата счетов в ресторанах)
Транзакций в оффере партнеров	Целочисленное поле, отражающее количество транзакций клиента по дебетовым или кредитным картам у партнеров по специальным предложениям с повышенным кэшбэком

* транзакция = покупка

ЗАДАНИЕ

Вам необходимо подготовить исследование, позволяющее понять, каким образом привлечь больше пользователей подписки и/или увеличить срок ее использования.

Результатом исследования должны быть рекомендации для компании по привлечению клиентов. Результатом работы может быть выделение групп клиентов, отличающихся своими характеристиками и/или поведением, и рекомендации для каждой такой группы.

Исследование должно содержать:

- Постановку вопроса и гипотезы. Должно быть высказано предположение, которое будет проверяться в рамках исследования, и дано обоснование адекватности выдвинутой гипотезы — это достигается через механизм гипотезы (более подробно — в комментариях к заданию).
- Анализ ситуации. Нужно дать понять слушателям, с какими данными вы работали, какие в них есть особенности (наличие выбросов, асимметричность распределений, пропуски, неравномерность по подгруппам, в том числе пустые категории, и тому подобное) и с чем они связаны
- Проверку гипотезы (см. подробнее в комментариях к заданию). При помощи статистических инструментов нужно сделать вывод, подтвердилась гипотеза или нет.
- Вывод и рекомендации. Исследование должно быть ценно не само по себе, а из-за призыва к действию, полученного в результате.
- Перспективы развития. Нужно критически оценить результаты исследования, подсветить ограничения с вариантами улучшения и прокомментировать обобщаемость полученных выводов.

КОММЕНТАРИИ К ЗАДАНИЮ

1. Постановка вопроса, гипотезы и механизмы. Подумайте о том, что вам интересно исследовать и почему это может быть интересно кому-то еще. На первом этапе обсудите вопрос, на который вы хотите получить ответ с использованием данных. Например, можете изучить, как наличие подписки отражается на интенсивности использования разных сервисов клиентом Тинькофф.

Что влияет на длительность подписки? Какие преимущества подписки наиболее привлекательны для клиентов? От каких характеристик клиента зависит использование подписки? Как сделать подписку более привлекательной для разных целевых аудиторий?

Какие факторы наиболее привлекательны для молодых потребителей? Вы можете сформировать рекомендации по привлечению какой-то определенной целевой группы или нескольких различных групп потребителей.

2. Выбор переменных. Выберите из представленного набора данных факторы, влияние которых вы считаете важным. Сформулируйте гипотезу о том, положительно или отрицательно эти факторы будут влиять на использование подписки. Расскажите подробно, почему ожидаете именно такую зависимость. Обратите внимание на единицы измерения выбранных переменных.
3. Разведывательный анализ. Определитесь, что будет единицей наблюдения в вашем анализе. Проиллюстрируйте с помощью графиков распределение переменных, на которых вы решили сфокусироваться. Посчитайте основные описательные статистики для них. Какие особенности вы наблюдаете? Как их можно проинтерпретировать, то есть объяснить? Охарактеризуйте форму распределений (скошенность, симметричность/асимметричность, вариативность).

В ходе разведывательного анализа учитывайте наличие выбросов и пропусков в выбранных вами переменных. Стоит обратить внимание на неоднородность данных: в разных категориях распределения одной и той же переменной могут различаться.

Не забывайте, что анализ можно проводить не только по исходным переменным, но и по преобразованным данным. Например, вместо количества транзакций можно взять количество транзакций за определенный период.

Примечание: если вам сложно сразу сформулировать исследовательский вопрос, можете начать работу над заданием с этого пункта, чтобы разобраться с предоставленным датасетом.

4. Анализ взаимосвязей. Посмотрев на переменные отдельно, можно переходить к проверке гипотезы. Отрадите графически взаимосвязь между выбранными переменными и проведите статистический анализ. Например, можно использовать таблицы сопряженности, вычислить корреляции, построить точечные диаграммы. Обратите внимание на типы переменных: для разных типов инструменты различаются. Согласуется ли результат с поставленной гипотезой? Если нет, объясните, с чем это может быть связано.
5. Проверьте, устойчивы ли результаты при анализе по разным подгруппам, определяющим характеристики клиента, например по регионам, полу, количеству детей, возрасту. Получаются ли похожие результаты по разным регионам? Схожи ли результаты для разных групп посетителей? Чем это объясняется? Различаются ли результаты по характеристикам подписки, например по длительности? Возможно, пользователи, которые подключены раньше, пользуются подпиской как-то иначе.

Кроме того, имеет смысл рассматривать отдельно тех, у кого есть подписка. Выгода будет зависеть от наличия подписки. В рамках подписки клиент может пользоваться такими сервисами, как кино, развлечения, рестораны. Учитывайте эти факторы.

Проанализируйте взаимосвязь из предыдущего пункта, поделив выборку по какому-либо фактору. Объясните важность этого фактора. Как можно проинтерпретировать полученные результаты?

6. Выводы и ограничения. Объясните, как компания может использовать полученные вами результаты. Идеальных исследований не бывает. Какие важные факторы не были вами учтены — из представленного набора или среди факторов, которые сложно объективно измерить? Что можно сказать о выборке пользователей и о ее репрезентативности? Можно ли распространять полученные выводы на всех пользователей России, других стран? Почему?

Наконец, обсудите, возможна ли обратная причинно-следственная связь в вашей гипотезе. Если да, объясните почему, а если нет — придумайте экономический механизм для обратного влияния.