

D A N



ВЫСШАЯ ШКОЛА
ЭКОНОМИКИ



ТИНЬКОФФ



АДЫГЕЙСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ
ADYGHE STATE UNIVERSITY

ЗДРАВСТВУЙТЕ, УВАЖАЕМЫЕ УЧАСТНИКИ!

Сегодня мы предлагаем вам поближе познакомиться с миром анализа данных и поработать над реальной задачей в команде с такими же юными исследователями.

Исследование — это сложный процесс, поэтому для того, чтобы вы могли справиться со всеми возникающими трудностями и вопросами, к каждой команде будет прикреплен ментор.



Кто такой ментор?

Кто такой ментор? Это человек, который имеет опыт работы над исследовательским проектом, не раз сталкивался со всеми трудностями при решении аналитических задач, а теперь готов помочь вам сделать качественный проект и победить! Не стесняйтесь обсуждать с ментором любые свои идеи, предлагать ему различные способы решения поставленной задачи и спрашивать совета в ходе работы. Каждый ментор много знает и умеет, не стоит игнорировать его помощь, однако не надейтесь, что он сделает всю работу за вас! :)

РАЗВИТИЕ ЭЛЕКТРОННОЙ КОММЕРЦИИ И ПАНДЕМИЯ COVID-19

Важнейшей долгосрочной экономической тенденцией во всем мире является развитие рынка электронной коммерции. Все большая часть покупок происходит онлайн, без физического посещения магазинов. Этот процесс особенно ускорился с наступлением пандемии COVID-19: стремясь минимизировать социальные контакты, многие пользователи отказывались от привычных способов покупать товары и услуги и экспериментировали с онлайн-покупками.

В ваших руках уникальная база данных о транзакциях, осуществленных по картам Tinkoff за период с 15 января 2020 до 31 июля 2020 года. Это период начала пандемии COVID-19 в Россию. Вам предстоит выявить ключевые тенденции развития рынка электронной коммерции в этот период.

Прежде чем мы опишем базу данных, мы бы хотели кратко напомнить вам ключевые события, происходящие в этот период. Первые двое заболевших COVID-19 в России были зарегистрированы 31 января, но быстрое распространение болезни началось только в марте. К 31 марту число заболевших достигло 2,3 тыс. человек. Быстрое развитие пандемии в марте заставило правительство пойти на ограничительные меры. Школы были закрыты начиная с 14 марта, обязательное ношение масок введено с 12 мая. Наиболее радикальной ограничительной мерой стали так называемые «нерабочие дни», которые начались с 30 марта и продолжались до 11 мая. Первая волна COVID-19 достигла пика 12 мая, когда число заболевших составило 11 тыс. К концу месяца число заболевших упало до 8,7 тыс., ограничительные меры начали отменять. Число заболевших продолжало снижаться до начала августа.

Примечание: использование дополнительных данных не является обязательным. В то же время, если вы считаете нужным обогатить датасет какой-либо дополнительной информацией (например, данными уровня регионов: численность населения, уровень доходов и т.п.), это допустимо.

База включает 6 переменных:

| | |
|------------------|--|
| data | Число в формате день.месяц.год, например 15.01.2020 |
| category_name | Категория магазинов |
| region_name | Название региона |
| online_rub | Объем онлайн покупок в рублях |
| offline_rub | Объем оффлайн покупок в рублях |
| stringency_index | Индекс жесткости ограничительных мер, о нем подробнее чуть ниже. |

Таким образом, для каждого региона Российской Федерации в каждый день доступного временного периода по каждой категории магазинов у нас есть информация об объеме онлайн и оффлайн покупок.

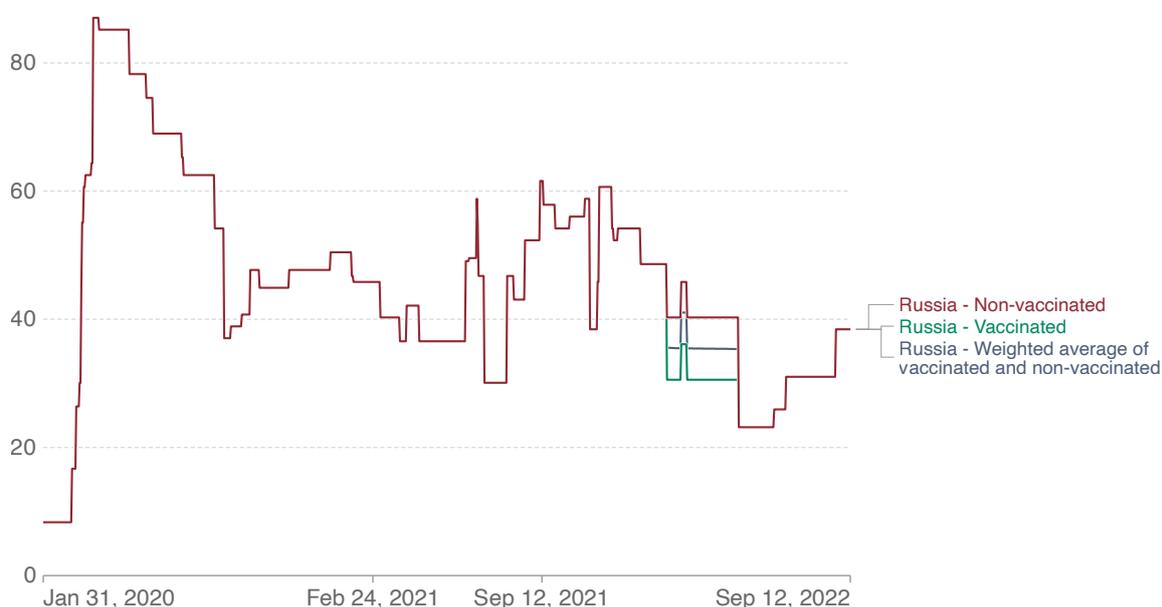
Справочно, категории магазинов: DUTY FREE, АВИАБИЛЕТЫ, АВТО УСЛУГИ, АПТЕКИ, АРЕНДА АВТО, ГОС. СБОРЫ, ДОМ, РЕМОНТ, Ж/Д БИЛЕТЫ, ЖИВОТНЫЕ, ИСКУССТВО, КИНО, КНИГИ, КРАСОТА, МЕД. УСЛУГИ, МУЗЫКА, НКО, ОБРАЗОВАНИЕ, ОДЕЖДА, ОБУВЬ, ОТЕЛИ, РАЗВЛЕЧЕНИЯ, РАЗНЫЕ ТОВАРЫ, РЕСТОРАНЫ, СВЯЗЬ, ТЕЛЕКОМ, СЕРВИС. УСЛУГИ, СПОРТТОВАРЫ, СУВЕНИРЫ, СУПЕРМАРКЕТЫ, ТОПЛИВО, ТРАНСПОРТ, ТУРАГЕНТСТВА, ФАСТФУД, ФОТО/ВИДЕО, ЦВЕТЫ, ЧАСТНЫЕ УСЛУГИ.

Stringency_index — это данные, добавленные к изначальному дата-сету транзакций банка Tinkoff. Мы бы хотели узнать, насколько распространение онлайн покупок связано с ограничительными мерами, принятыми в связи с распространением COVID-19. Сбор таких данных очень трудоемок, в условиях хакатона у вас не было бы достаточно времени. Удобство «[COVID-19 Government Response Stringency Index \(Stringency Index\)](#)», рассчитанного университетом Оксфорда на основе открытых данных для большинства стран мира, в том, что он предоставляет большой объем информации в одной переменной, находящейся в интервале от 0 до 100.

COVID-19: Stringency Index

The stringency index is a composite measure based on nine response indicators including school closures, workplace closures, and travel bans, rescaled to a value from 0 to 100 (100 = strictest).

Our World
in Data



Source: Hale, T., Angrist, N., Goldszmidt, R. et al. A global panel database of pandemic policies (Oxford COVID-19 Government Response Tracker). Nat Hum Behav 5, 529–538 (2021). <https://doi.org/10.1038/s41562-021-01079-8>
CC BY

Рис. 1. Динамика Stringency index для России в рассматриваемый период

Подробное описание методики расчета индекса (на английском языке) вы можете найти по [ссылке](#), но ее суть в том, что оценивается степень жесткости мер по следующим показателям: закрытие школ, закрытие рабочих мест, отмена публичных мероприятий, ограничения на собрания, ограничения работы общественного транспорта, публичная кампания информирования, требования оставаться дома, ограничения на поездки внутри страны, ограничения на международные поездки, обязательно тестирование, отслеживание контактов заболевших, обязательное ношение масок.

ЗАДАНИЕ

Вам нужно подготовить исследование изменений рынка онлайн-платежей за период карантина. Результатом исследования должны быть рекомендации (~призыв к действию) бизнесу либо государству.



Исследование должно содержать:

- Анализ ситуации. Нужно дать понять слушателям, с какими данными вы работали, какие там есть особенности и с чем они связаны
- Постановку гипотезы. Должно быть высказано предположение, которое будет проверяться в рамках исследования, и дано обоснование адекватности выдвинутой гипотезы — это достигается через механизм гипотезы
- Проверку гипотезы. При помощи статистических инструментов нужно сделать вывод о подтверждении (или нет) гипотезы
- Вывод и рекомендации. Исследование должно быть ценно не само по себе, а из-за призыва к действию, полученного в результате
- Перспективы развития. Результаты исследования должны быть критически оценены, подсвечены ограничения с вариантами улучшения и прокомментирована обобщаемость полученных выводов

НАВОДЯЩИЕ ВОПРОСЫ

Представленные ниже вопросы призваны помочь вам построить исследование. Рекомендуется ответить на них, но они не являются обязательными

1. Мы хотим оценить степень распространения онлайн-платежей. Для этого нам необходим показатель, который будет показывать долю онлайн платежей во всех платежах. Рассчитайте такой показатель. Обратите внимание, что для некоторых наблюдений в графе `online_rub` или `offline_rub` стоит пропуск. Это означает, что в этот день в этом регионе по этой категории товаров не было онлайн или оффлайн покупок (иными словами, мы можем заменить пропуск нулем). Подумайте, как лучше поступить с такими категориями.
2. Распространение онлайн платежей очень отличается в зависимости от региона и категории магазинов. Давайте начнем с регионов. Рассчитайте степень распространения онлайн платежей для каждого из регионов. Представьте рейтинг регионов по популярности онлайн покупок. Соответствуют ли полученные результаты тому, что вы ожидали? Как вы считаете, чем объясняется разная степень распространенности онлайн платежей в различных регионах? Влияет ли на это различный доход в разных регионах? Различная численность населения?
3. Теперь перейдем к различиям между категориями магазинов. Рассчитайте степень распространения онлайн платежей для каждой из категорий магазинов. Представьте рейтинг категорий по популярности онлайн покупок. Соответствуют ли полученные результаты тому, что вы ожидали? Как вы считаете, чем объясняется разная степень распространенности онлайн платежей по различным категориям магазинов?
4. Пришло время проанализировать динамику — изменение показатель со временем. Насколько сильно увеличилась распространенность онлайн платежей за рассматриваемый период? В каких регионах рост был самым большим? По каким категориям магазинов? Соответствуют ли полученные результаты тому, что вы ожидали? Чем можно объяснить полученные результаты?
5. Как на развитие онлайн-торговли повлияла пандемия? На этот вопрос тяжело ответить, так как рост онлайн торговли происходил и без пандемии, чтобы точно оценить эффект пандемии нам необходимо знать «чтобы было бы, если бы не пандемия». Тем не менее мы можем оценить, как связана жесткость ограничений и распространенность онлайн-покупок. Каким образом можно оценить эту связь? Придумайте способ и оцените. Обратите внимание, что доступный вам временной период включает как период увеличение жесткости ограничительных мер, так и период смягчения ограничений. Сопровождался ли этот период сокращением распространенности онлайн-покупок?

6. Помимо распространенности онлайн-покупок нас интересует абсолютный объем покупок в рассматриваемый период. Как потребители реагировали на наступление пандемии в рассматриваемый период? От каких категорий вы ожидаете рост или, напротив, падение спроса? Выберите несколько категорий магазинов (3–5), которые кажутся вам наиболее интересными, и проанализируйте их динамику. Совпадают ли результаты анализа с вашими ожиданиями? Чем вы можете объяснить динамику этих товаров?

7. Подведите итог проведенному исследованию. Что именно вы хотели выяснить? Какие результаты были получены? Как эти результаты могут быть использованы на практике? Какие ограничения своего исследования вы видите? Если бы у вас было больше времени и не было ограничений, связанных с доступностью данных, как вы могли бы дополнить и улучшить свое исследование?