

# Итоговый проект по курсу

А/В-тестирование:  
практическое руководство  
(АВТ-15)

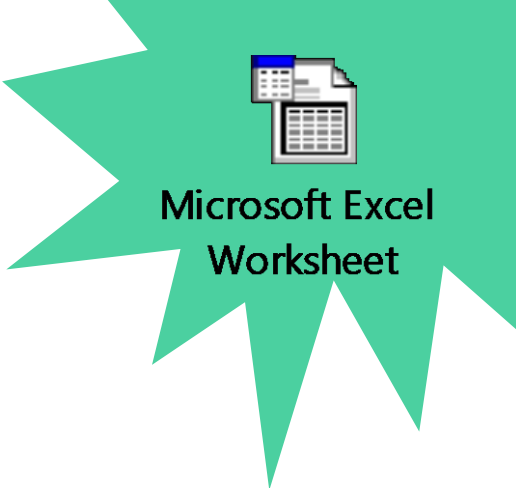
# Анализ исходных данных

## ● Что вы сделали с исходными данными?

Для анализа использовался тестовый дата-сет по сессиям пользователей мобильного приложения Gett. Данные импортировались из файла .xlsx в Google Sheets. Анализ проводился в гугл-таблицах. Ознакомительный анализ дата-сета проводится для понимания общей структуры и логики сбора данных. Посчитано общее количество сессий, количество отмен, а также количество успешных сессий для каждого этапа пути клиента.

app_opened	101 500					FUNNEL
price_seen	91 431		conv_seen	90%		
order_made	74 236		conv_order	81%		
ride_completed	62 967		conv_compl	85%		
user_cancelled	11 269		conv_thru	62%		
city_center_order	58 025					

Дата-сет содержит следующую информацию: основным действием является открытие приложения. В общем случае открытие приложения имеет распределение Бернулли, т.е. принимает значения 0 или 1. В тестовый дата-сет попали положительные исходы указанного действия. Общее количество сессий – 101 500. Далее на каждом этапе результат действия такое же распределение – 0 или 1 – в зависимости от результата на каждом из этапов считается конверсия перехода к следующему этапу. Каждая сессия имеет id-пользователя, время заказа (час), операционную систему устройства, с которого выполняется вход, класс заказа, маркер применения повышенного коэффициента при просмотре стоимости поездки и заказа из центра, а также возраст пользователя, расстояние (предполагаемой) поездки и RFM сегментация клиентов.

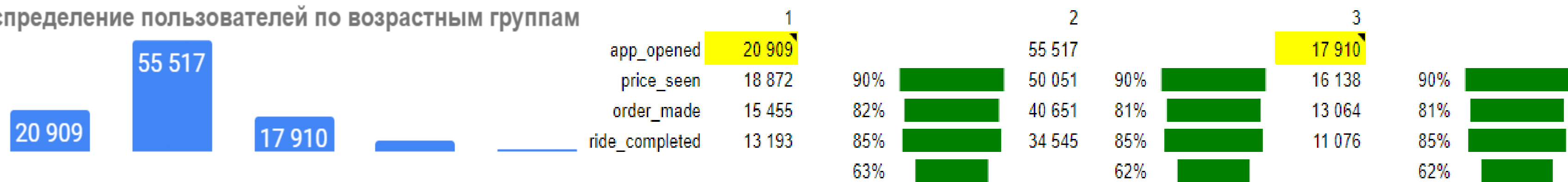


# Анализ исходных данных

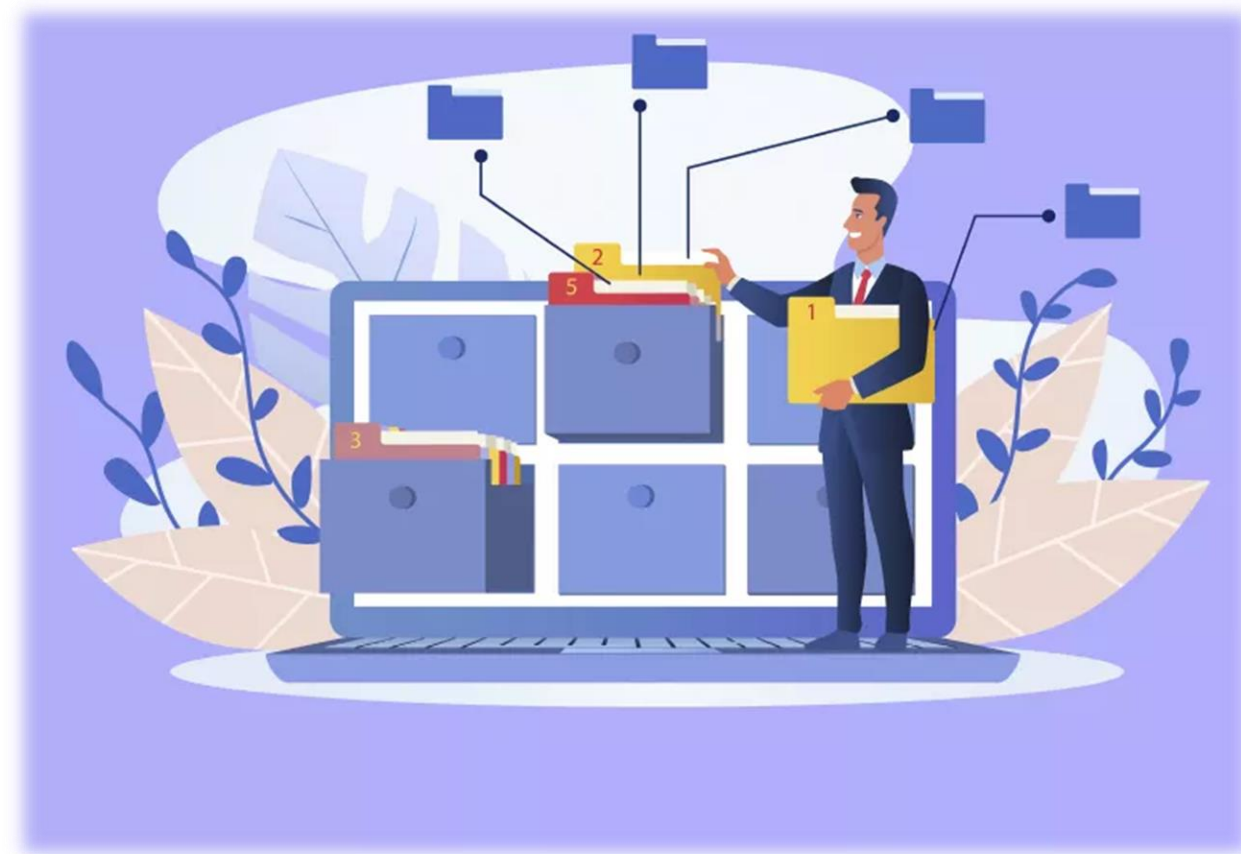
## ● Какие проблемы нашли? Как их устранили?

В ходе анализа дата-сета значительных проблем непосредственно с данными выявлено не было. Основным направлением анализа текущей ситуации было выбрано построение и детальное рассмотрение конверсий перехода по основным этапам для различных срезов пользователей. Например, были выделены возрастные группы (6 групп: 18-20 лет, далее с интервалом по 10 лет – (20+ .. 60+)

Распределение пользователей по возрастным группам



Для каждой из трех самых активных групп были построены воронки конверсий перехода по этапам клиентского пути в приложении. В аналогичном ключе были сделаны срезы по параметрам iOS/Android, surge/no\_surge, economy/comfort/business, а также по времени заказа такси для различных групп пользователей. В каждой группе было выявлено снижение конверсии на том или ином этапе. На основании выявленных проблемных участков была составлена база для построению продуктовых гипотез.



# Анализ исходных данных

## ● Инсайты, к которым пришли в результате анализа данных

Выбраны следующие проблемные участки:

- Конверсия в заказ для пользователей эконом-класса на Android в период 7-9 часов на 16 п.п., чем средняя за день.
- Конверсия в заказ для пользователей комфорт-класса на Android ниже в период 19-21 час на 13 п.п., чем средняя за день.
- Конверсия в заказ для пользователей Android в период “no surge” ниже на 7 п.п., чем у iOS при прочих равных условиях (no surge).
- Конверсия в заказ для пользователей Android в период «surge» ниже на 15 п.п., чем в «no surge» у пользователей Android.
- Конверсия в заказ для пользователей iOS в период «surge» ниже на 12 п.п., чем в «no surge» у пользователей iOS.
- Возрастные группы «до 20 лет» и «30-40 лет» пользуются приложением сравнимо реже, чем пользователи «20-30 лет» в среднем. Пользователи старше 40 лет пользуются приложением несравнимо реже.



# Формирование гипотез

## Как осуществляли экспресс-анализ и сколько гипотез в итоге отметили?

Все сформулированные гипотезы сведены в таблицу, каждой из гипотез проставлен личный рейтинг. Гипотеза N°1 (см. таблицу) проверена с использованием фреймворка ICE. Так выполнялась приоритизация.

№	Гипотеза	Источник	Почему не стоит проверять	Ваш личный рейтинг 1-10	Почему поставили такую оценку?
1	Если сделать предложение по повышению класса поездки для пользователей Android класса "эконом" в период 7-9 часов, то конверсия совершения заказа (вызов машины) изменится на 7 процентных пунктов (увеличится), потому что при получении услуги более высокого качества за те же деньги пользователи будут более активнее завершать заказ с положительным результатом. .	дата-сет	Возможно временной разброс в 3 часа слишком большой, может нужно сузить до одного	10	низкие затраты на реализацию
2	Если сделать предложение по повышению класса поездки для пользователей Android класса "комфорт" в период 19-21, то конверсия совершения заказа (вызов машины) изменится на 7 процентных пунктов (увеличится), потому что при получении услуги более высокого качества за те же деньги пользователи будут более активнее завершать заказ с положительным результатом.	дата-сет	Возможно временной разброс в 3 часа слишком большой, может нужно сузить до одного	8	низкие затраты на реализацию
3	Если сделать предложение со скидкой 5% на следующую поездку при совершении заказа "сейчас" для пользователей Android в период по surge, то конверсия в заказ (вызов машины) изменится на 5 процентных пунктов (увеличится), потому что 50% пользователей указанной группы имеют RFM - high и medium, и ,возможно, для них важна экономия (андроид как-никак) в дальнейшем использовании сервиса	дата-сет	надо проверять	9	1. в выборке большой процент активных пользователей. 2. обещанная скидка определенно не будет использована в полном (выданном) объеме, но заказ "сейчас" будет сделан.
4	Если сделать стоимость поездки ниже на 30% от удорожания в связи с surge (т.е. в период удорожания увеличивать стоимость на 70% расчетного удорожания) для пользователей Android в период surge, то конверсия в заказ (вызов машины) изменится на 10 процентных пунктов (увеличится), потому что именно конверсия в заказ для указанной группы пользователей после просмотра цены значительно ниже, чем для "surge-iOS" и "no surge - Android"	дата-сет	возможно это будет экономически не выгодно, надо проверить размер скидки	9	чтобы опровергнуть предположение, что владельцы Android менее платежеспособны (*распаковывая флагман Samsung)
5	Если сделать стоимость поездки ниже на 30% от удорожания в связи с surge (т.е. в период удорожания увеличивать стоимость на 70% расчетного удорожания) для пользователей iOS в период surge, то конверсия в заказ (вызов машины) изменится на 10 процентных пунктов (увеличится), потому что более дешевые поездки с большей вероятностью будут завершены с положительным результатом	дата-сет	возможно это будет экономически не выгодно, надо проверить размер скидки	10	вся надежда на пользователей iOS
6	Если сделать периодическое напоминание из приложения о возможности заказать такси, о программах лояльности и/или возможности оплачивать часть поездки бонусами для для пользователей возрастной группы "до 20 лет" в 3-4 раза в неделю во время повышенного спроса среди указанной возрастной категории, то app_opened изменится на 20%, потому что иногда напоминание будет оказываться полезным, к месту и ко времени. Порядка 50% пользователей имеют RFM - high и medium.	дата-сет	возможен обратный эффект от частых напоминаний	7	сложные (для меня) механизмы маркетинга (что написать, когда и как часто написать). Больше интуитивное предположение





# Формирование гипотез

## Как осуществляли экспресс-анализ и сколько гипотез в итоге отметили?

Для максимально оцененных гипотез составлено краткое описание АБ-теста с указанием тестируемого изменения, генеральной совокупности, а также первичных и вторичных метрик, необходимый для корректного выполнения теста.

#	Гипотеза	Тестируемое изменение	Бизнес-цель	Генеральная совокупность	Первичные метрики	Вторичные метрики	Метод агрегации метрик
1	Если сделать предложение по повышению класса поездки для пользователей Android класса "эконом" в период 7-9 часов, то конверсия совершения заказа (вызов машины) изменится на 7 процентных пунктов (увеличится), потому что при получении услуги более высокого качества за те же деньги пользователи будут более активнее завершать заказ с положительным результатом. .	Предложение по улучшению класса поездки	Увеличение конверсии в заказ. Увеличение прибыли.	Пользователи андроида класса "Эконом" после просмотра цены поездки	Конверсия в заказ	Конверсия в просмотр цены, конверсия в заказ для клиентов, которые по умолчанию используют предлагаемый более высокий класс обслуживания	Общее значение метрики для выбранной генеральной совокупности
2	Если сделать предложение по повышению класса поездки для пользователей Android класса "комфорт" в период 19-21, то конверсия совершения заказа (вызов машины) изменится на 7 процентных пунктов (увеличится), потому что при получении услуги более высокого качества за те же деньги пользователи будут более активнее завершать заказ с положительным результатом.	Предложение по улучшению класса поездки	Увеличение конверсии в заказ. Увеличение прибыли.	Пользователи андроида класса "Комфорт" после просмотра цены поездки	Конверсия в заказ	Конверсия в просмотр цены, конверсия в заказ для клиентов, которые по умолчанию используют предлагаемый более высокий класс обслуживания	Общее значение метрики для выбранной генеральной совокупности
3	Если сделать предложение со скидкой 5% на следующую поездку при совершении заказа "сейчас" для пользователей Android в период по surge, то конверсия в заказ (вызов машины) изменится на 5 процентных пунктов (увеличится), потому что 50% пользователей указанной группы имеют RFM - high и medium, и ,возможно, для них важна экономия (андроид как-никак) в дальнейшем использовании сервиса	Предложение скидки 5% на следующую поездку при совершении текущей поездки	Увеличение конверсии в заказ. Увеличение прибыли.	Пользователи андроида в период <b>surge</b>	Конверсия в заказ	Мониторинг частоты использования сервиса, объем использования предложенной скидки, выручка в целом для рассматриваемой группы	Общее значение метрики для выбранной генеральной совокупности
4	Если сделать стоимость поездки ниже на 30% от удорожания в связи с surge (т.е. в период удорожания увеличивать стоимость на 70% расчетного удорожания) для пользователей Android в период surge, то конверсия в заказ (вызов машины) изменится на 10 процентных пунктов (увеличится), потому что именно конверсия в заказ для указанной группы пользователей после просмотра цены значительно ниже, чем для "surge-iOS" и "no surge - Android"	Использование 70% удорожания в период surge вместо расчетного 100%	Увеличение конверсии в заказ. Увеличение прибыли.	Пользователи андроида в период <b>surge</b>	Конверсия в заказ	Выручка в целом для рассматриваемой группы пользователей	Общее значение метрики для выбранной генеральной совокупности



# Формирование гипотез

## ● Как выбрали гипотезу для А/Б? Почему ее?

Так как для анализа использовался тестовый дата-сет, то гипотеза для АБ-теста выбиралась исходя из выгрузки с результатами, полученными по итогам проведения тестирования (с разбиением Test/Control).

Была выбрана следующая фича из предложенных :

Понижение коэффициента surge, чтобы увеличить price-to-order конверсию (во время surge)

Ожидаем: выше price-to-order конверсия во время surge



# Подготовка к А/Б-тестированию

## ● Каким образом сформировали выборку и выполнили сплит?

Первоначально тренировочный сплит выполнен на полученном дата-сете. Из исходных данных исключены не участвующие в тестировании подгруппы (например, по surge, пользователи iOS или пользователи классов комфорт и бизнес, в зависимости от выбранной гипотезы и соответствующей ей генеральной совокупности). Далее данные агрегировались по user\_id, затем случайным образом промаркированы:  $\text{=СЛЧИС()} > 0,5$  или  $\text{=СЛЧИС()} \leq 0,5$ .

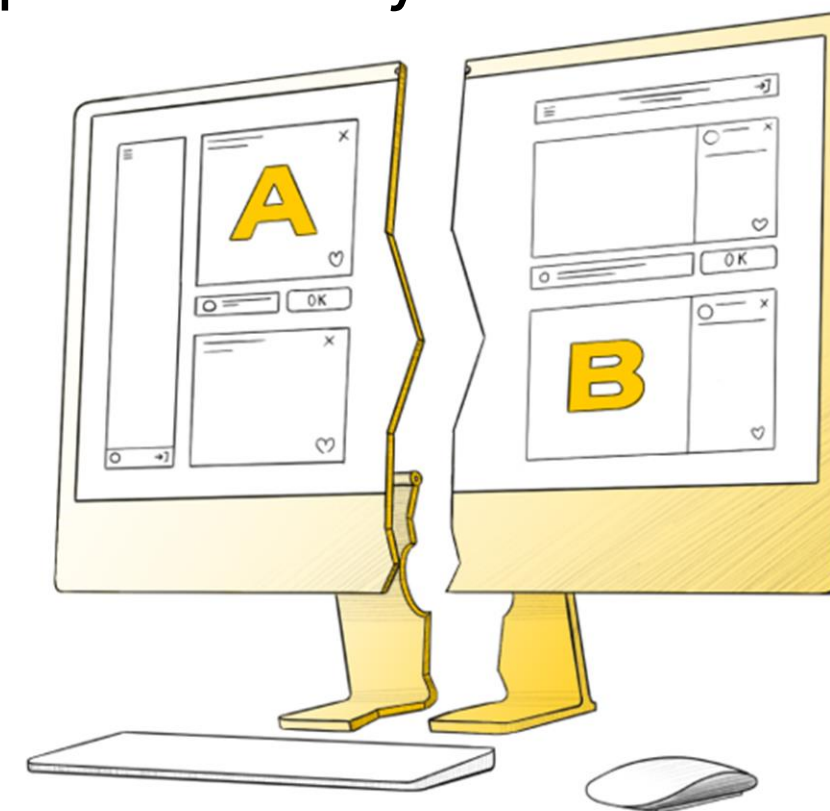
Для анализа предоставлен дата-сет с результатами, полученными по итогам проведения тестирования (с разбиением Test/Control). Сплит выполнен случайным образом, при несоответствии количества сессий в тестовой и контрольной выборках вручную выравняются до отношения 50/50.

## ● Каковы результаты проверки на однородность?

Проверка на однородность выполняется с помощью t-test. Выдвигаются нулевая гипотеза (среднее в двух выборках одинаково) и альтернативная гипотеза (среднее – разное).

При tails = 2 (двустороннее распределение, тип теста – двухвыборочный с равными дисперсиями, вычисляется p-value. В случае  $p\text{-value} > \alpha$  5% результат считается статистически значимым и принимается нулевая гипотеза.

Выборки считаем однородными, можно проводить АБ-тест.





# Проведение А/Б-теста



## ● Что показали результаты А/Б-теста? Как они были получены?

Результаты проведенного АБ-теста получены в виде дата-сета (.csv) с разделением всех сессий на тестовую и контрольную группы.

Данные импортировались из файла .csv в Google Sheets. Анализ проводился в гугл-таблицах.

Произведено разделение выборок тестовой и контрольной группы, сессии агрегированы по user\_id, выровнено количество элементов в каждой из групп для соблюдения отношения сплита 50/50. Затем данные сгруппированы по датам проведения тестирования (семь дней) и для каждой группы вычислены среднее значение и p-value.

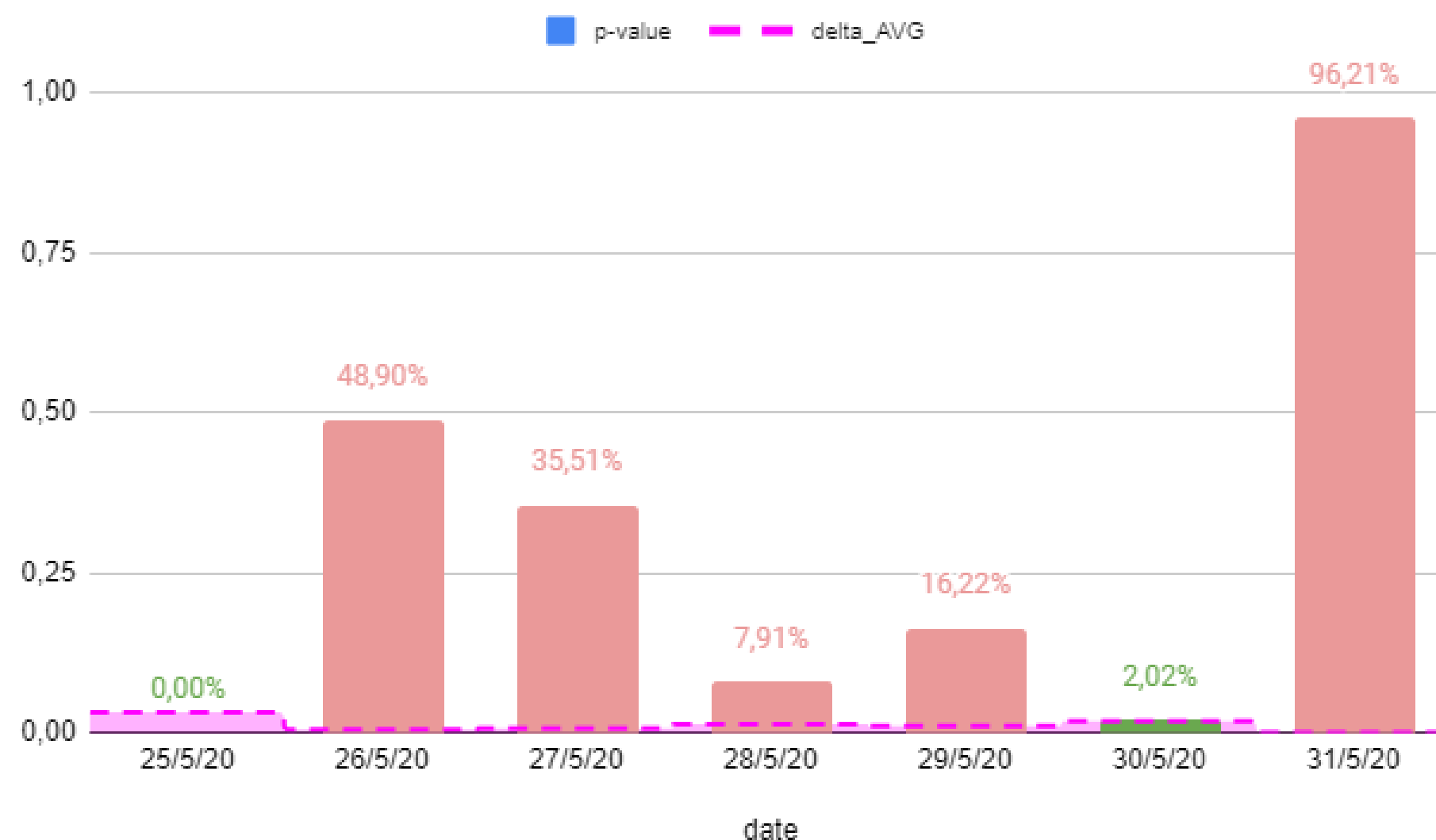
Строим диаграмму p-value по дням тестирования.



# Анализ результатов А/Б-теста

## ● Какой вывод следует из теста?

p-value по дням



В данном случае принимаем нулевую гипотезу  $H_0$  при  $p\text{-value} < 5\%$ .

Большинство полученных результатов по дням статистически не значимы, поэтому на основании анализа p-value в отдельные дни тестирования решение о масштабировании фичи отрицательное. Предлагается рассмотреть p-value накопительным итогом.



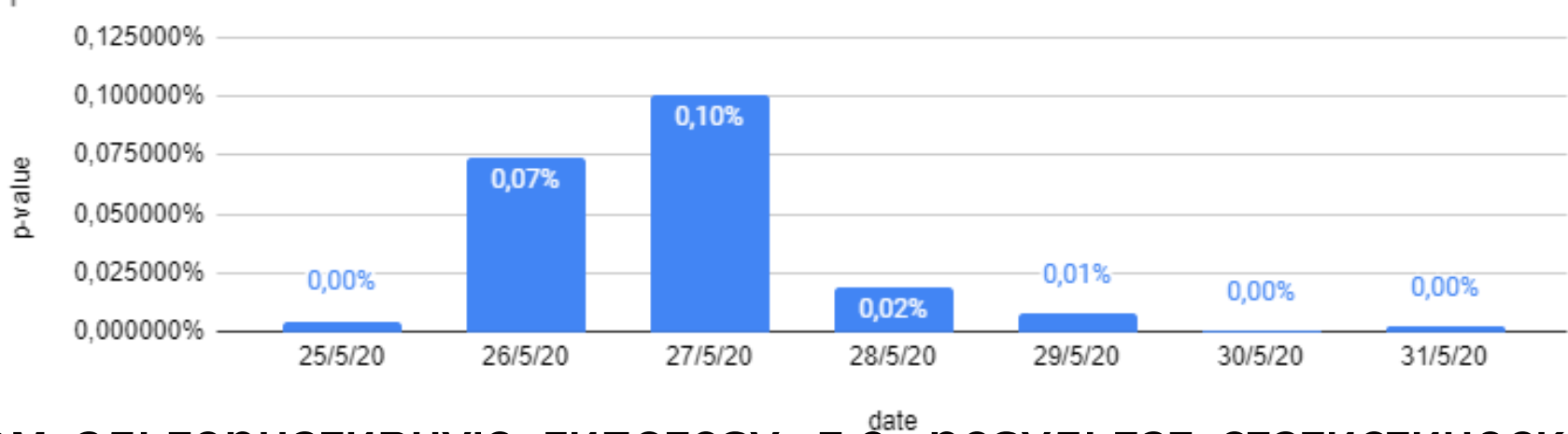
# Анализ результатов А/Б-теста

## ● Какой вывод следует из теста?

По итогам накопления p-value по дням тестирования получаем следующие результаты:

date	Qnt	p-value	delta_AVG				
25/5/20	5 117	0,004628%	3,19%		0,0046%	<	5%
26/5/20	10 152	0,074029%	1,88%		0,0740%	<	5%
27/5/20	15 219	0,100868%	1,50%		0,1009%	<	5%
28/5/20	20 297	0,019509%	1,47%		0,0195%	<	5%
29/5/20	25 414	0,007647%	1,40%		0,0076%	<	5%
30/5/20	30 520	0,000516%	1,47%		0,0005%	<	5%
31/5/20	35 631	0,002288%	1,26%		0,0023%	<	5%

p-value накопительным итогом



Так как во всех случаях p-value меньше 5% то принимаем альтернативную гипотезу, т.е. результат статистически значимый.

Полученные результаты при рассмотрении накопительного p-value действительно показывают увеличение конверсии в заказ при предложении скидки 5% через 20 секунд после просмотра стоимости поездки. Однако, ожидаемый эффект – увеличение на 10 п.п., – очевидно, не подтвердился.

Предлагается оценить по ICE экономическую целесообразность фичи при увеличении конверсии всего на 2 п.п., и, в зависимости от результатов оценки, масштабировать фичу на всю группу пользователей или отклонять.

По ICE: с учетом допущений получено

$ICE = 572\,767 - N$ , где  $N$  – Effort – стоимость внедрения фичи и проведения тестирования.

Выглядит так, будто бы тестируемая фича при 2 п.п. не целесообразна.

Решение: не применять на всех пользователей.

Возвращаемся к списку гипотез и начинаем прорабатывать следующую или выдвигаем новую гипотезу.



# Инсайты курса

***Я раньше даже не знал о существовании АВ-тестирования,  
а теперь и сам смогу (наверное).***

**A**



**B**



***Я раньше не думал, что статистика такая интересная наука,  
а теперь узнаю каждый день что-то новое.***

