

# Воронка авиабилетов



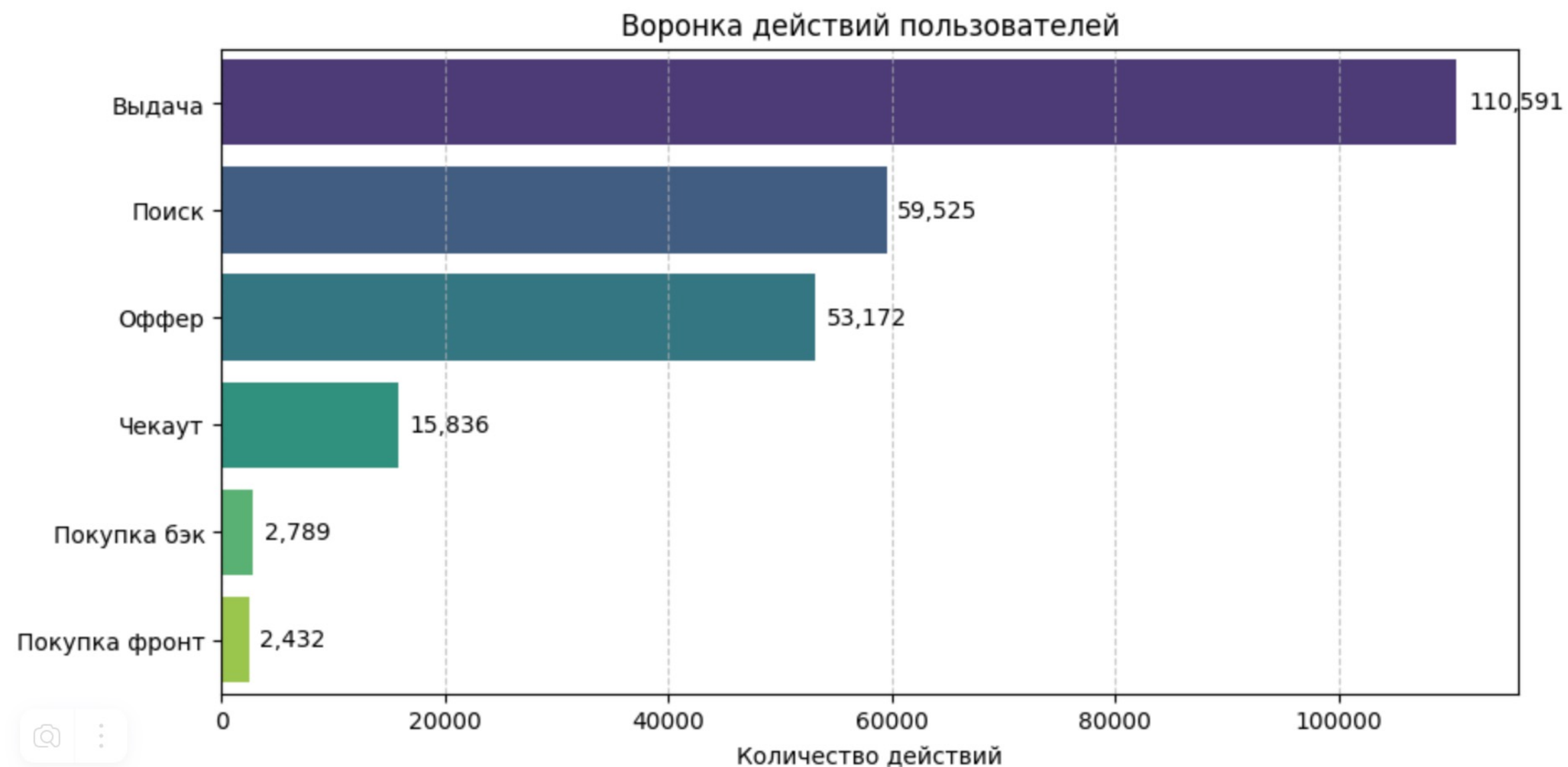
Рыбаков Даниил



Каждая запись в датасете - действие пользователя в рамках отдельной сессии.

Экран Выдача является основным этапом, где пользователь принимает решение и совершает больше всего действий.

Воронка действий пользователей не отражает реальное количество продажной конверсии. А лишь отражает движение пользователя по экранам.



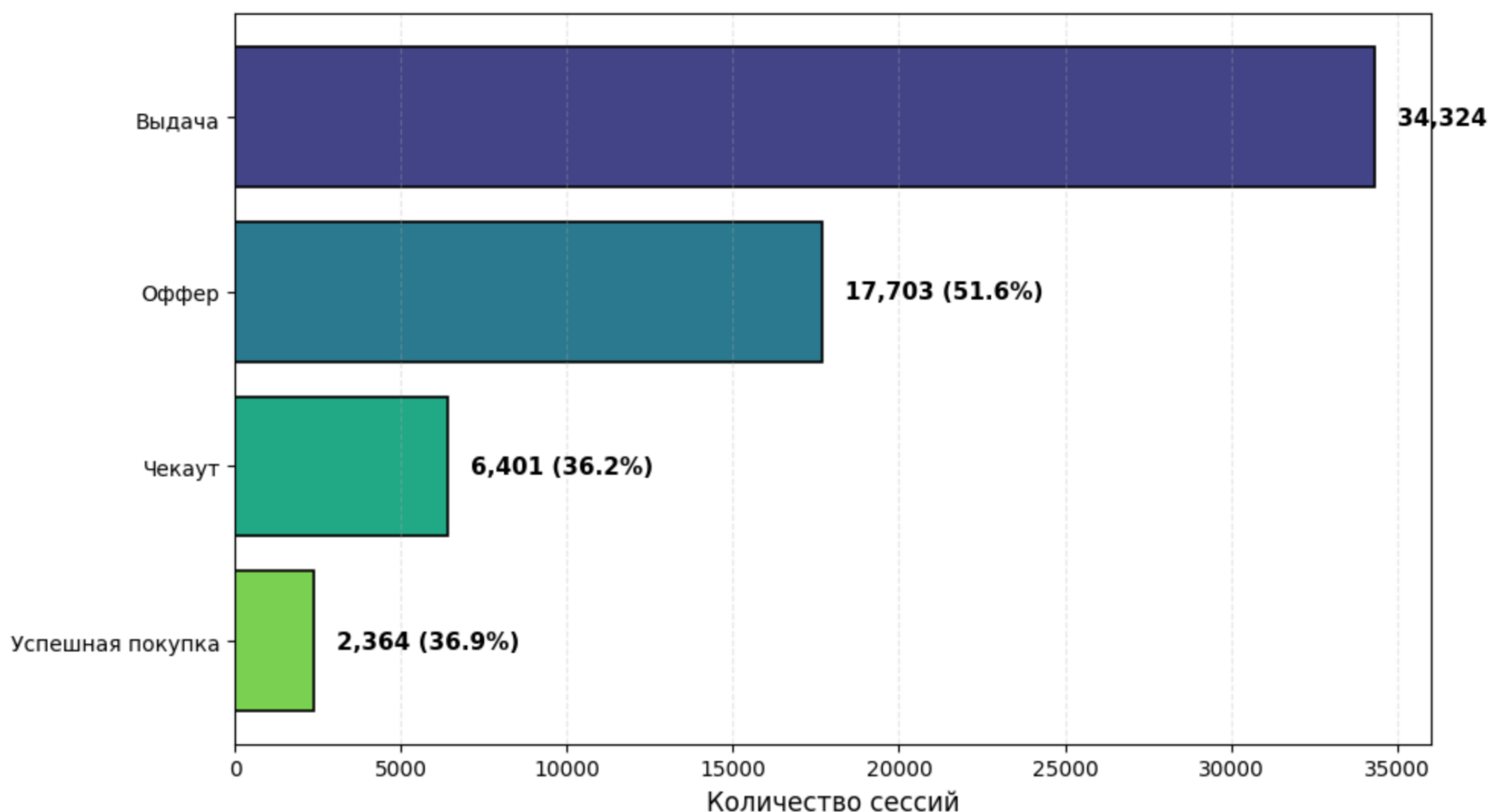
Поэтому можем рассмотреть пользовательский сценарий, который будет предполагать в себе Выдача->Оффер->Чекаут->Количество успешных статусов покупки.



# Проблематика

При переходе от воронки действий пользователей к воронке продаж наглядно видно, на каком этапе уходит большинство пользователей.

Воронка продаж авиабилетов (уникальные сессии)



Практически наибольшее снижение конверсии происходит между этапами Оффер и Чекаут.

При этом пользователи, дошедшие до этапа Оффера, уже проявили высокий интерес к покупке, но не конвертируются в оплату.

Заметим, что с этапа Выдачи до этапа Чекаут доходит  $\sim 1/6$  часть сессий, поэтому предположим гипотезы, которые могут повлиять на продажную конверсию и конверсию между определенными этапами.

## Гипотеза №1

- Если пользователю на этапе Выдачи предложить два специальных предложения - самый дешевый рейс и самый быстрый рейс на выбранные даты, то конверсия из этапа Оффера в этап Чекаут увеличится, за счёт увеличения качества Офферов и упрощения выбора.

## Гипотеза №2

- Если пользователя на этапе Оффера информировать о более выгодных датах перелёта в ближайший период(+/-2-3 дня), то конверсия из этапа Выдачи в этап Успешной покупки увеличится, за счёт удержания пользователя на этапе Оффера.

## Гипотеза №3

- Если пользователю на этапе Оффера сообщать о том, что "Осталось мало билетов"("Билеты скоро кончатся"), то конверсия из этапа Оффера в этап Чекаут увеличится, за счёт уменьшения неопределённости при выборе.



# Дизайн А/В теста

## Этапы дизайна

1. Формулирование бизнес-гипотезы
2. Выбор метрик
3. Формулирование статистической гипотезы
4. Выбор статистического критерия
5. Определение MDE, длительности теста и размера выборки
6. Применение поправок
7. Сплитование трафика

## Формулирование бизнес-гипотезы

**Гипотеза №1 является наиболее перспективной**, так как она направлена на главный этап пользовательского сценария - экран Выдачи, где пользователь принимает основное решение и совершает наибольшее количество действий.

Предложение двух удобных и контрастных вариантов упрощает процесс выбора, за счёт чего должно уменьшится количество действий связанных экраном Выдачи в следствие чего, улучшится качество Офферов.

При этом гипотеза не несёт высоких продуктовых рисков.





## Выбор метрик

- **Ключевая метрика:** конверсия из этапа Оффер -> Чекаут.
- **Защитные метрики:** продажная конверсия, средний GMV.
- **Прокси-метрики:** Количество сессий, остановившихся на этапе Оффер, время до перехода в Чекаут.

## Формулирование статистической гипотезы

- **H0:** конверсия Оффер -> Чекаут в контрольной и тестовой группах не отличается.
- **H1:** конверсия Оффер -> Чекаут в тестовой группе выше, чем в контрольной. (односторонняя гипотеза, так как ожидается рост метрики).

# Дизайн A/B теста

## Выбор статистического критерия

- Для проведения эксперимента выбирается **Z-тест для пропорций**, так как мы работаем с метрикой пропорции и имеем достаточно большую выборку.
- Также выбирается **уровень значимости  $\alpha=0.05$**  и значение параметра  **$\beta=0.2$** , **мощность теста =  $(1-\beta)$** .



## Определение MDE, длительности теста и размера выборки

- **Размер выборки = 1900**(значение всей генеральной совокупности), чтобы снизить уровень **MDE = X**(довольно большое), и **время проведения теста = X**(довольно большое).

## Применение поправок

- Для контроля ошибки первого рода используется **поправка Бенферони**. С учётом количества метрик  $m = 5$ , уровень значимости  $\alpha' = \alpha/m$ . Следовательно уровень значимости равен  $\alpha' = 0.01$ .

## Дизайн A/B теста

### Сплитование трафика

- **Деление трафика** в пропорции 50/50.
- **Способ сплитования** трафика был выбран стратификация пользователей по их покупательной способности, то есть по сегментам RFM.
- **Контроль однородности** групп, по среднему GMV группы и количеству пользователей по RFM сегментах.

**Спасибо за внимание!**