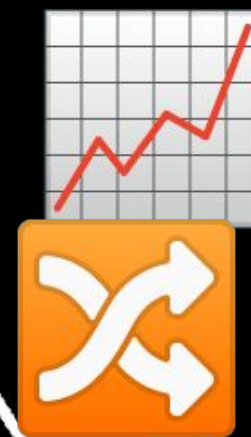


# Стартуем в 19-00

**Есть время чтобы налить кофе, проверить имя в зуме, микрофон и камеру)**

---

# Консультация Спринт 3. Проверяем гипотезы Джойним и мерджим



Горленко Екатерина,  
Наставник, ревьюер факультета DA, Яндекс.Практикум

Яндекс Практикум

# Как настроение?



# Наши договорённости

## Организованность

- Будь вовремя
- Понятное имя в Zoom (и везде)
- Камера включена
- Правило одного включенного микрофона
- Вопросы в чате

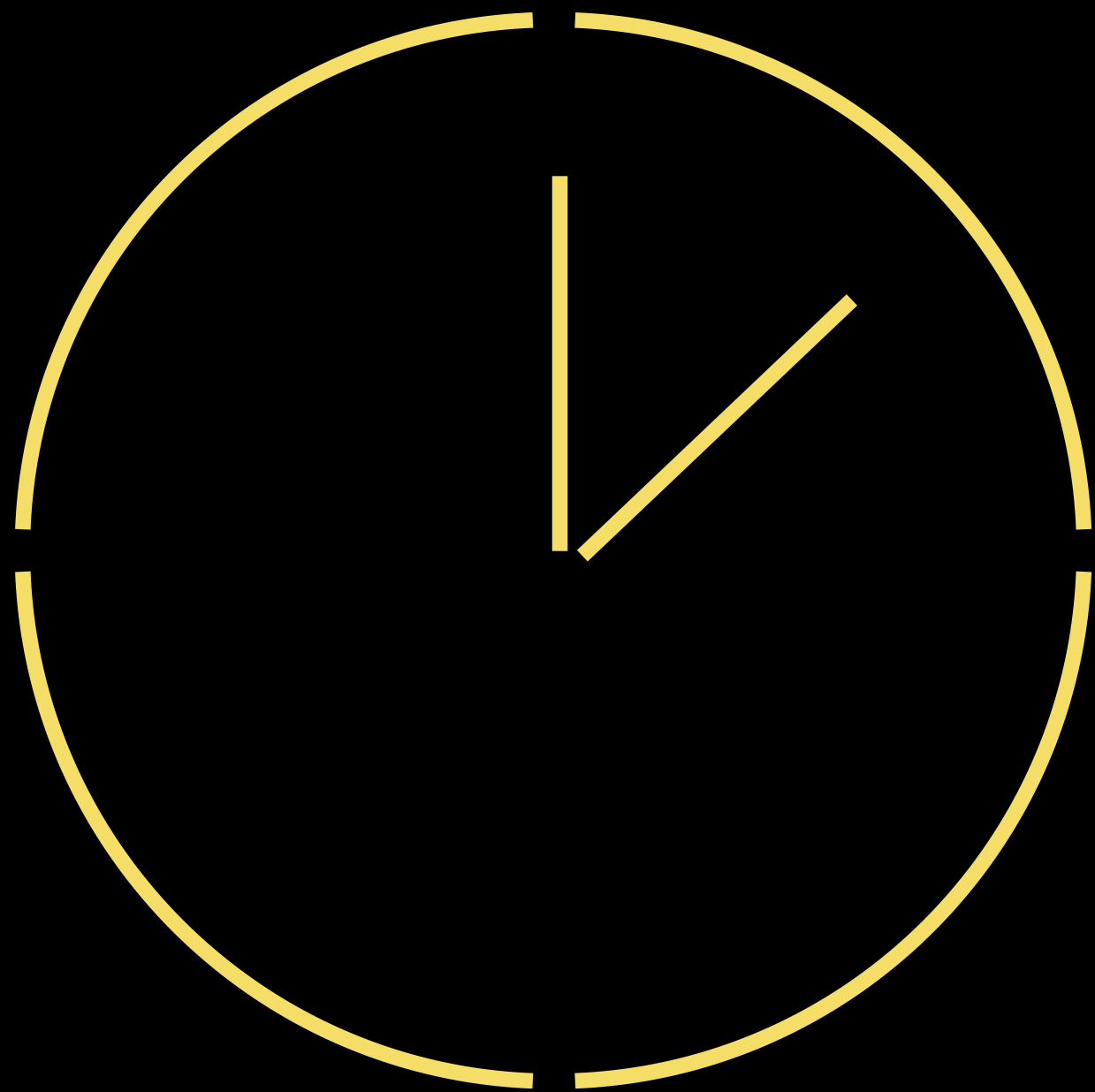
## Комфортная коммуникация

- Общаемся на «ты»
- Вовлеченность и проактивность
- Обучение – это ошибки
- Уважительное отношение
- Критикуешь – предлагай
- Береги общее время
- Запись только для потока

## Цель – научиться

- Мы – разные
- Учимся самостоятельно принимать решения
- Самостоятельность – это поиск и общение, а не одиночество
- Взаимопомощь
- Нет спойлерам

# План встречи



- Статус спринта (5 мин)

- Основная часть (40 - 50 мин):

- Проверка гипотез
- Джойны и мерджи

Вопросы? (10 мин)

# Статус спринта

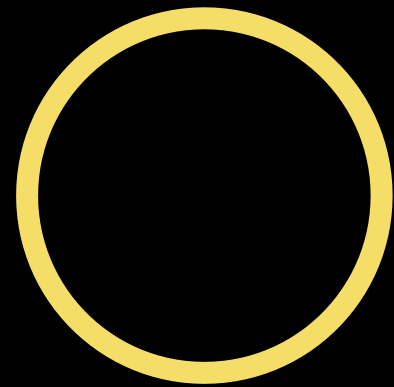
# Основная часть

# Проверка гипотез



# Проверка гипотез

- Зачем гипотезы?
- Формулируем правильно
- Проверяем



1. Вы запустили новую версию лендинга для сервиса доставки еды по подписке. На старой версии на регистрировалось 78% попавших на страницу, на новой - 81%
2. В вашем мобильном приложении 50% пользователей видят зеленую кнопку “купить”, а 50% - красную. На зеленую кнопку нажали 30%, на красную 35%
3. Вы добавили рекламные баннеры на сайт. В среднем пользователи стали проводить на сайте 11 минут, а раньше тусили все 20

Чем отличаются ситуации с точки зрения статистики?

# Проверка гипотез

- Зачем гипотезы?
- Формулируем правильно
- Проверяем

## Еще пример

Две выборки, как сравнить среднее?

- Выборка А ( $N > 1000$ )
- Выборка В ( $N > 1000$ )

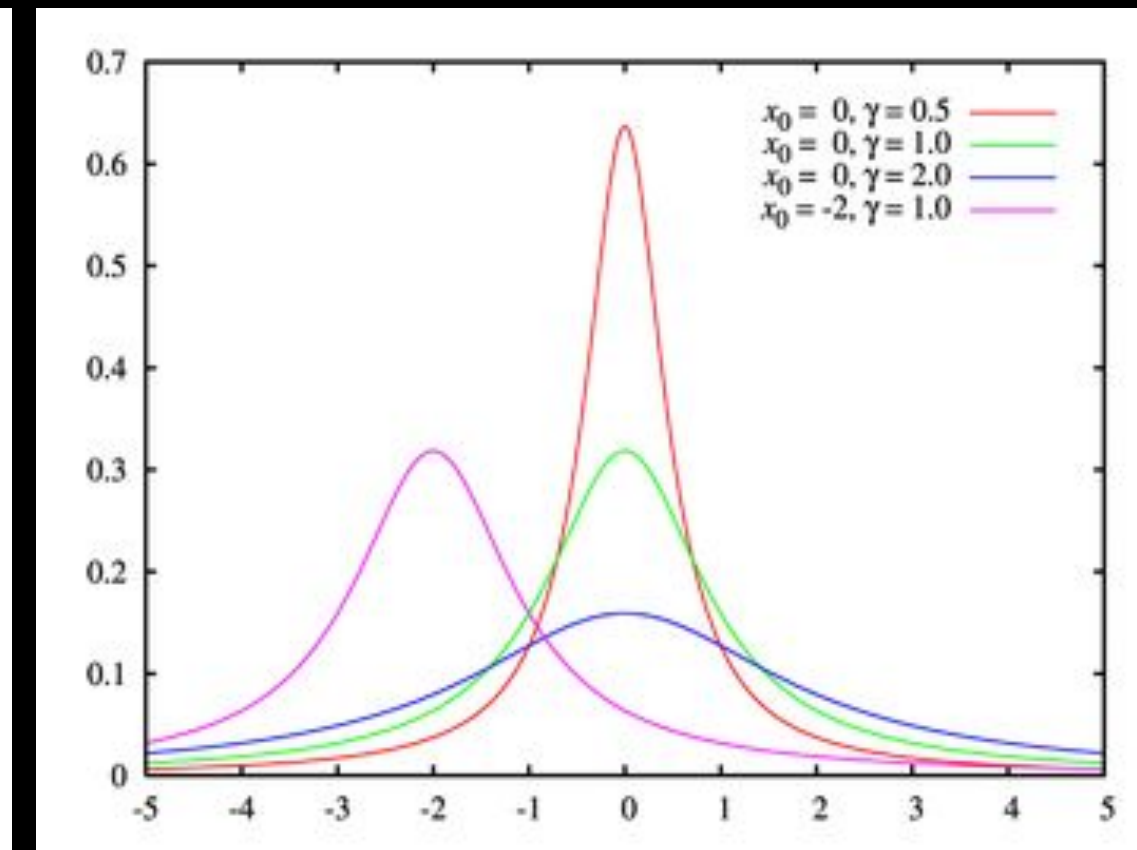
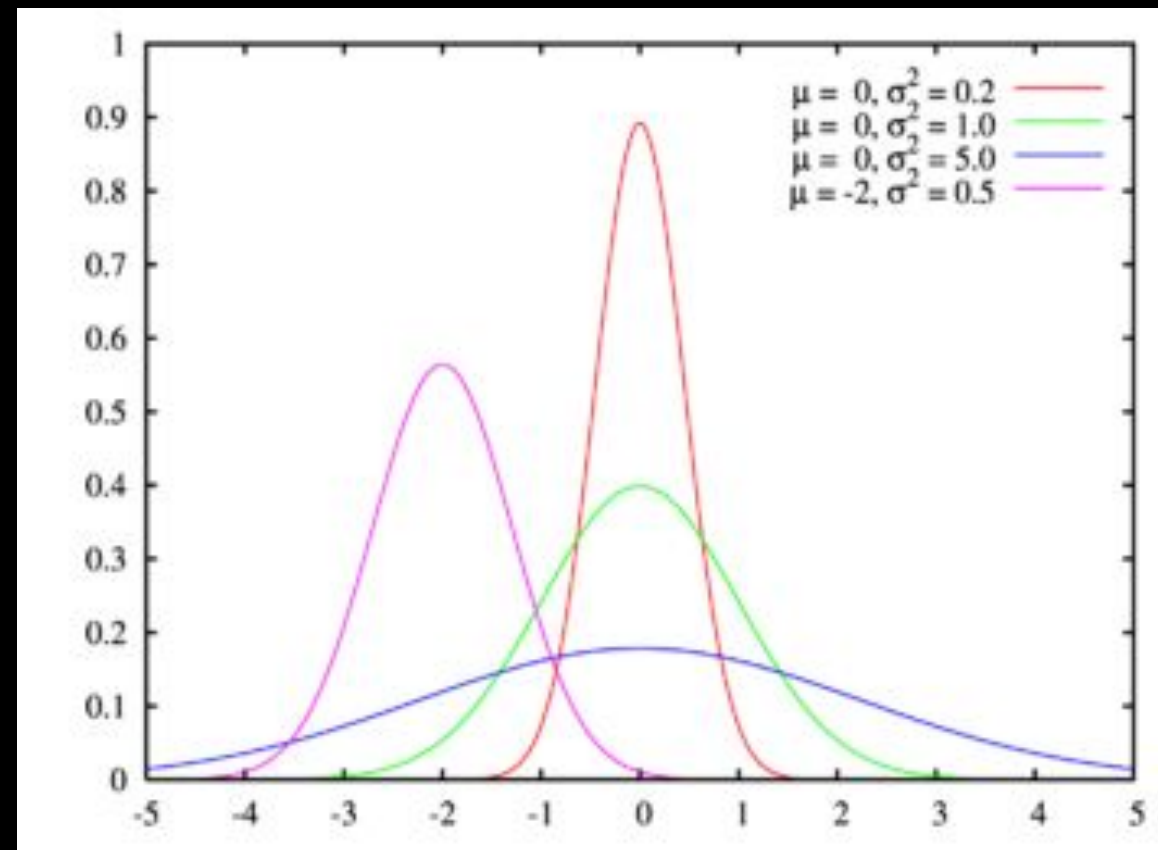
Например, каждое наблюдение – это пользователь. А значение наблюдения – сумма, на которую он совершил покупку.

В выборке А люди из одного города, в выборке В – из другого. Нас интересует вопрос, различается ли среднее значение между выборками, или нет.

# Проверка гипотез

- Зачем гипотезы?
- Формулируем правильно
- Проверяем

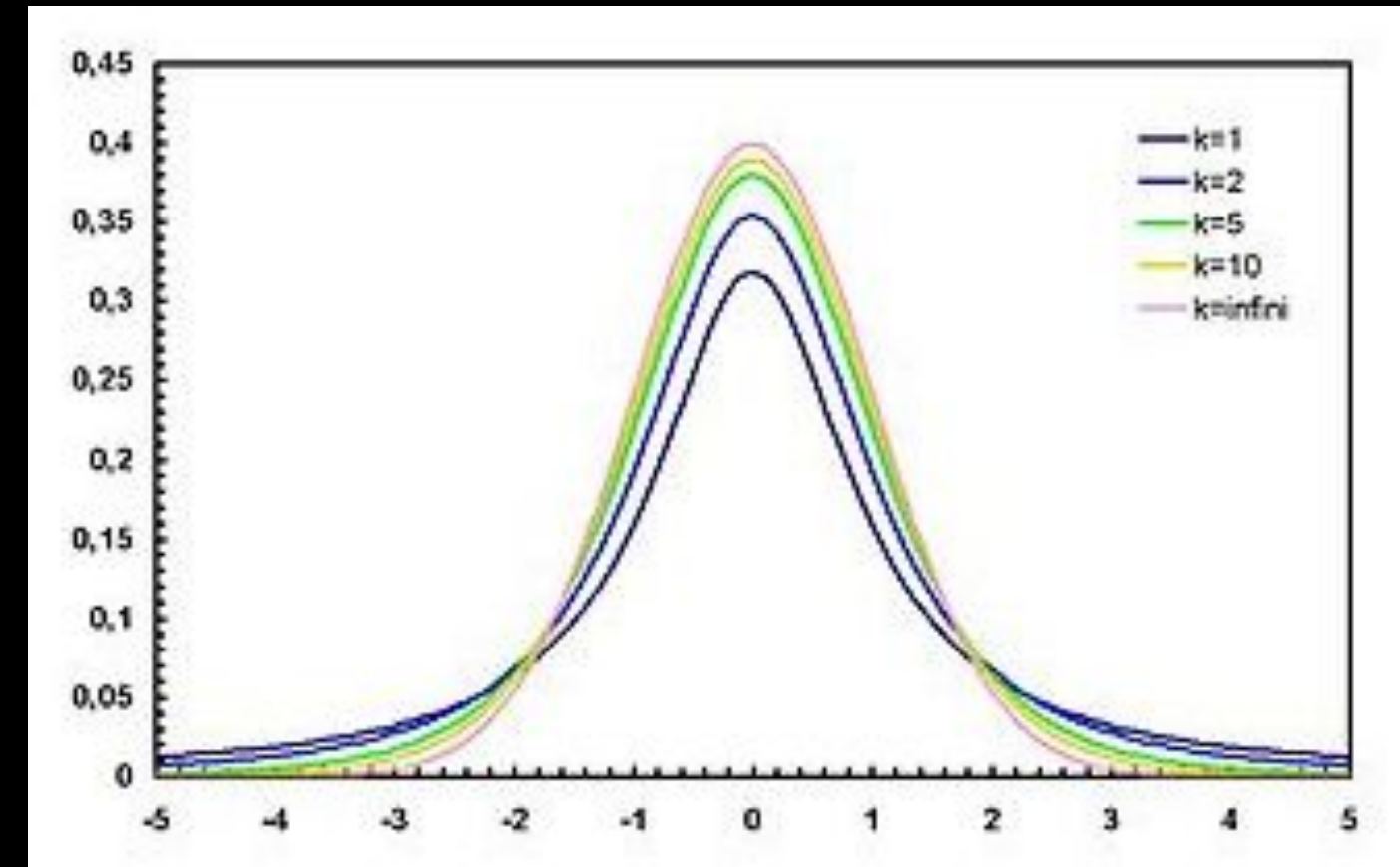
## Про хвосты



# Проверка гипотез

- Зачем гипотезы?
- Формулируем правильно
- Проверяем


## Распределение Стьюдента (t-распределение)



используется для проведения t-теста: теста на равенство матожидания выборки стандартно нормально распределённых случайных величин некоторому значению

Статья про разные распределения <https://habr.com/ru/post/311092/#Gauss>

# Проверка гипотез

- 
- Зачем гипотезы?
  - **Формулируем правильно**
  - Проверяем

## Вспоминаем термины

**Математическое ожидание** - среднее (взвешенное по вероятностям возможных значений) значение случайной величины

**Дисперсия случайной величины** - мера разброса значений случайной величины относительно её математического ожидания

**P-value** - вероятность того, что **H0** (нулевая гипотеза) верна, если мы примем **H1** (альтернативная гипотеза)

# Проверка гипотез

- Зачем гипотезы?
- **Формулируем правильно**
- Проверяем

**H0**: Среднее выборки  $a$  **равно** среднему выборки  $b$

**H1**: Среднее выборки  $a$  **не равно** среднему выборки  $b$

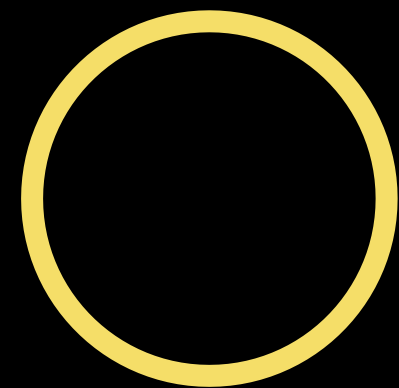


# Проверка гипотез

С чем сравнить p-value?

$\alpha = 0.05\%$  - это достаточно?

- Зачем гипотезы?
- **Формулируем правильно**
- Проверяем



	Истинное (но неизвестное нам) положение дел	
	Верна $H_0$	Верна $H_1$
Мы «приняли» $H_0$	ПРАВИЛЬНО! (чувствительность критерия) $1-\alpha$	$\beta$ ОШИБКА 2-го рода
Мы отвергли $H_0$	ОШИБКА 1-го рода (уровень значимости) $\alpha$	ПРАВИЛЬНО! (мощность критерия) $1-\beta$

Почитать про p-value  
<https://habr.com/ru/company/stepic/blog/250527/>

Яндекс.Практикум

Нельзя одновременно сделать обе ошибки

# Проверка гипотез

- Зачем гипотезы?
- Формулируем правильно
- **Проверяем**

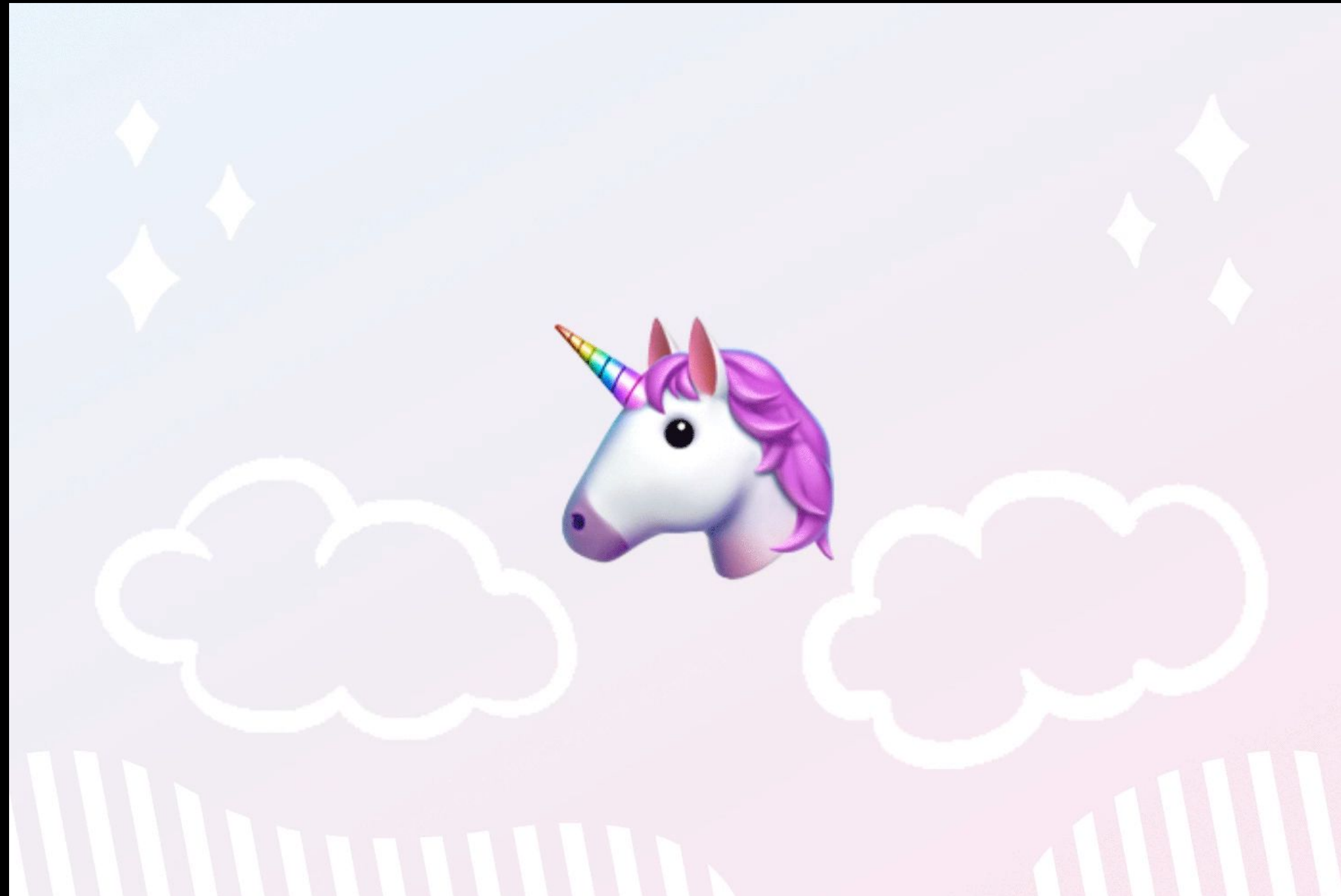
1. Импортировать `numpy` и `scipy.stats`
2. Пара строчек кода
3. ...
4. Profit!

(Пример в ноутбуке)



**Merge & join**  
а еще concat & append

**Тут живут идеальные датасеты**



# Merge & join

- Помехи

- Порядок таблиц

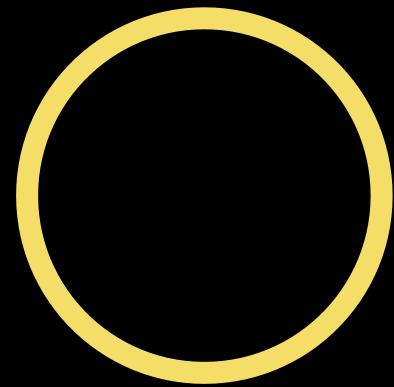
- Проверка

## Враги мерджа 🤖

1. “Кривые заголовки”. Лучше:
  - a. маленькими латинскими буквами с нижними подчеркиваниями
  - b. понятные, стандартные
2. Разные регистры. `str.lower()`
3. Пропуски. `how='outer'`
4. Типы данных. Подобное с подобным
5. Ошибка “Dead kernel” (большие датасеты)
6. Дубликаты. Множат сущности
7. “Проблема справочников”
8. Временные зоны
9. Курсы валют

# Merge & join

- Помехи
- Порядок таблиц
- Проверка



Как можно обогатить данные?

Клиент:

1. Мейл
2. Телефон
3. 4 цифры карты
4. Устройство

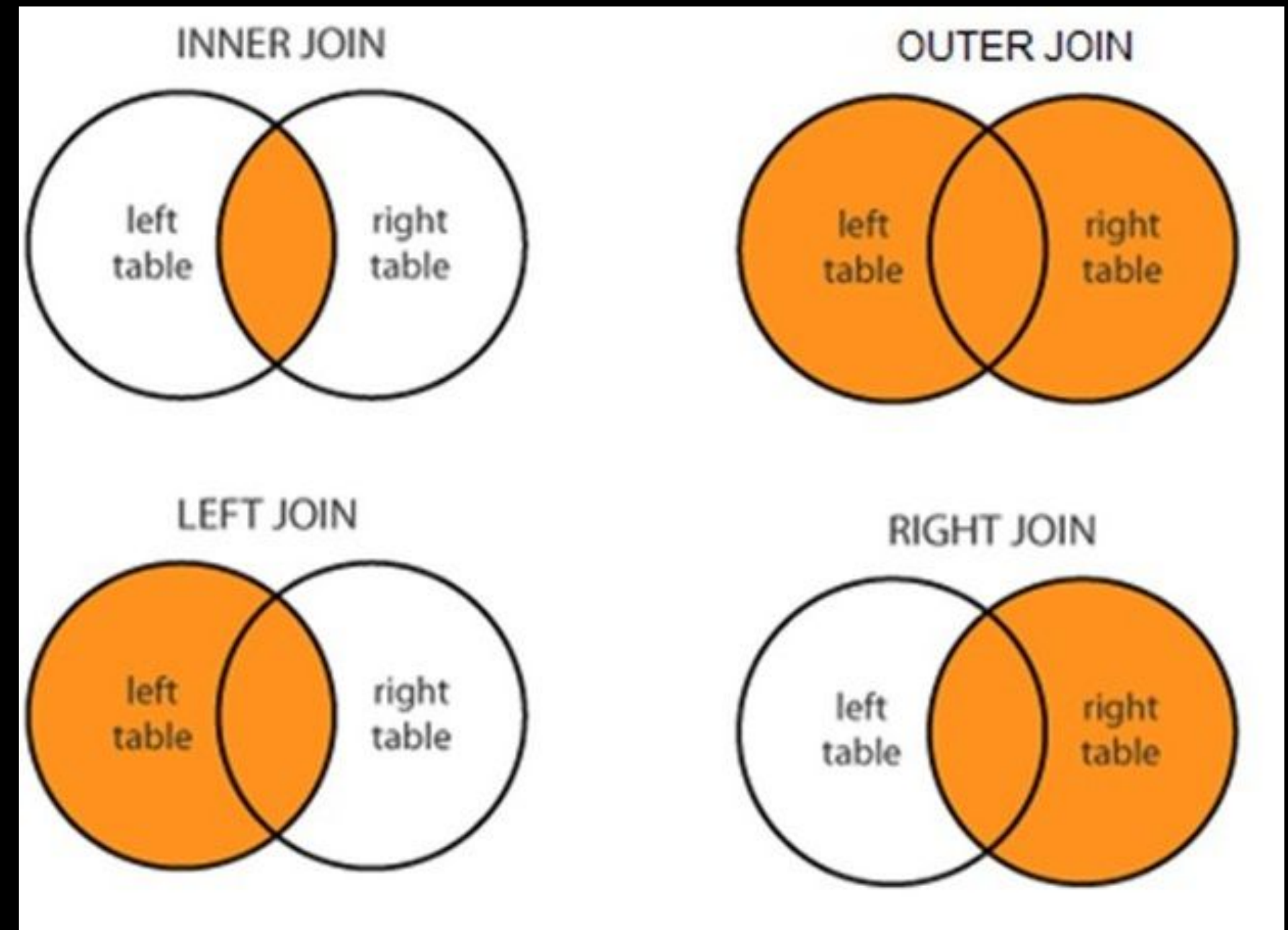
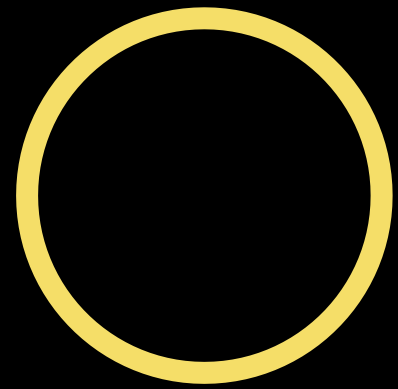
Событие:

1. Локация
2. Погода
3. Связанные события

ваши варианты в чатик...

# Merge & join

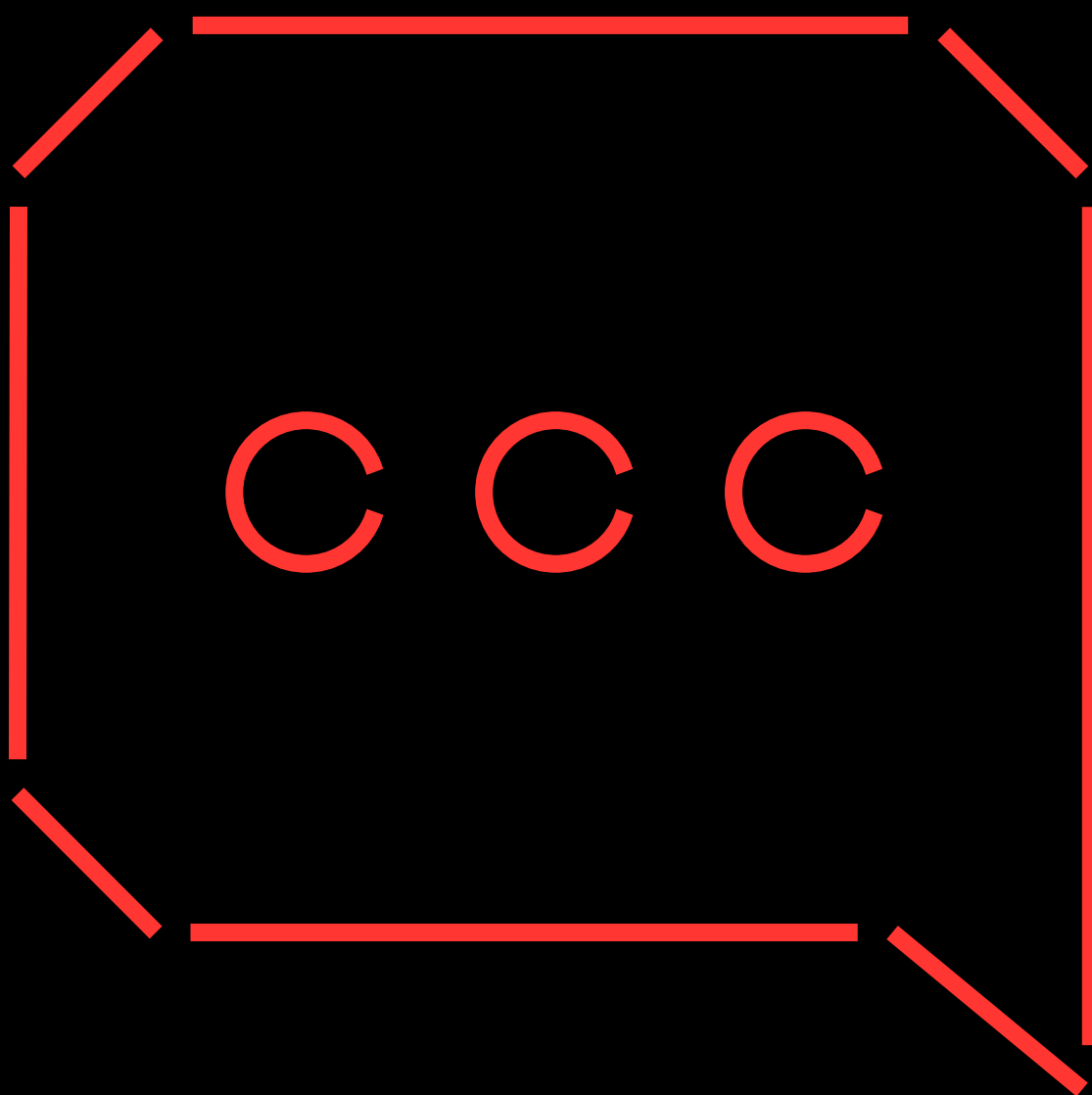
- Помехи
- Порядок таблиц
- Проверка



# Merge & join

- Помехи
- Порядок таблиц
- Проверка

1. Пойми, как будешь объединять(left, inner, outer)
2. Выбери метод
3. Агрегируй, потом объединяй
4. Сохраняй исходные данные
5. Проверяй итог
6. Находи инсайты в нестыковках



Вопросы?

# Обратная связь

Напишите в чат впечатления от консультации