

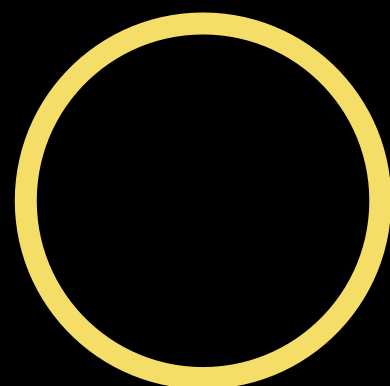
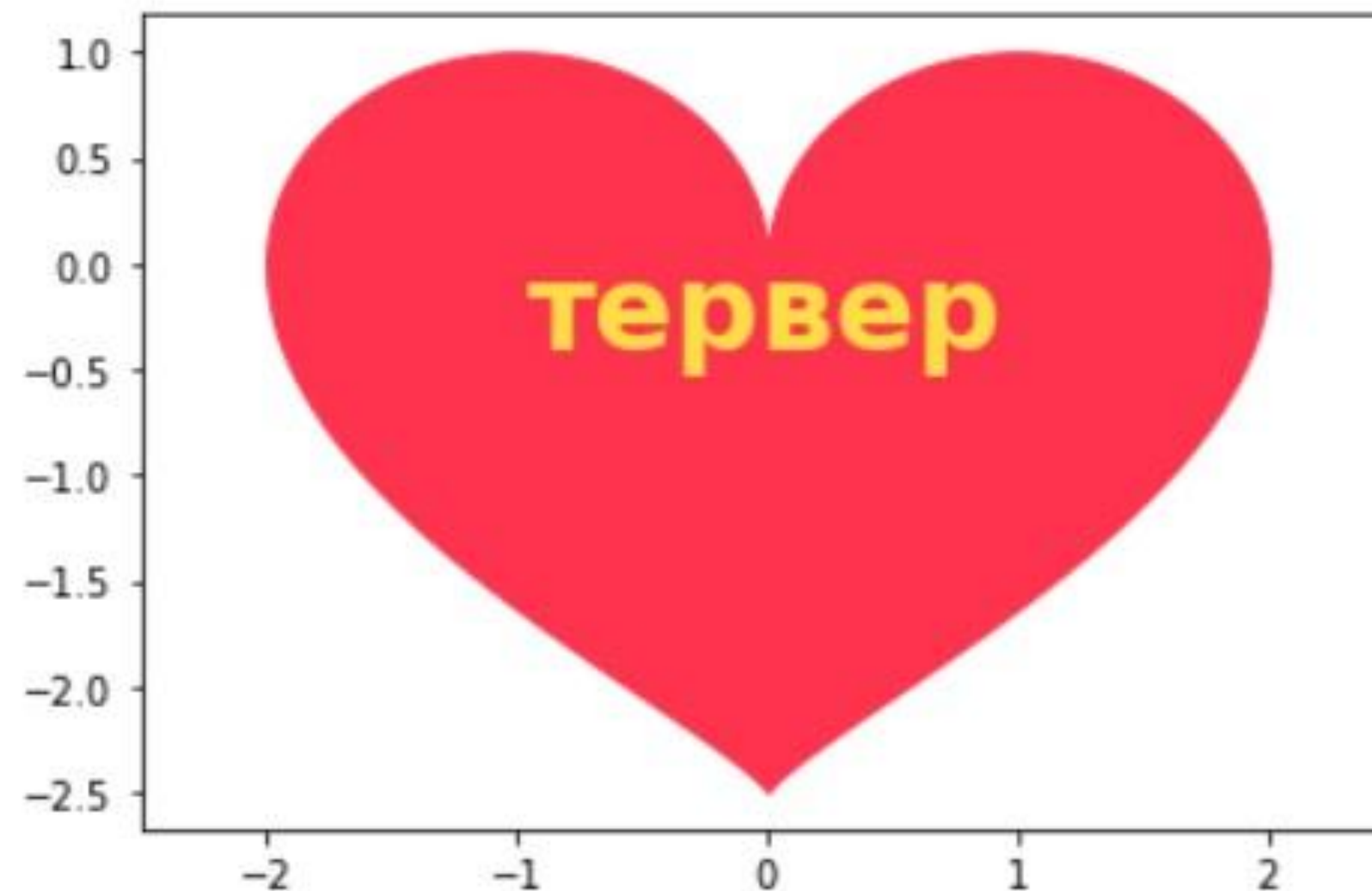
буду в 15.00, а пока:

- пожалуйста, включите камеру,
- назовитесь понятным именем
- и располагайтесь поудобнее

```
x = np.linspace(-2,2,2000)
y1 = np.sqrt(1-(abs(x)-1)**2)
y2 = -2.5*np.sqrt(1-(abs(x)/2)**0.8)
```

```
plt.fill_between(x, y1, color=(1.0,0.2,0.3))
plt.fill_between(x, y2, color=(1.0,0.2,0.3))
plt.xlim([-2.5, 2.5])
plt.text(0, -0.4, 'тервер', fontsize=32,
        fontweight='bold', color='#FFDD44',
        horizontalalignment='center')
```

Text(0, -0.4, 'тервер')



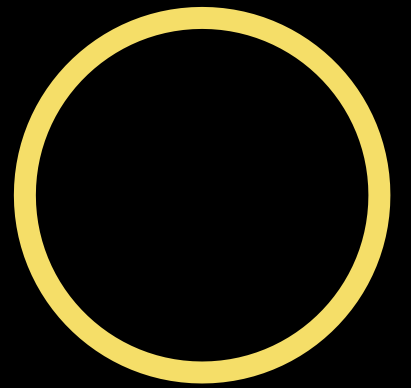
Основы статистики, вероятности и реляционной алгебры. Как самому себе объяснить на пальцах.

Елена Эльзесер,
Наставник DA, Яндекс.Практикум

Яндекс Практикум

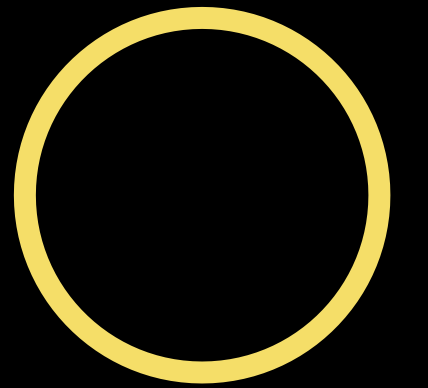
план

1. Варианты
2. Кодим
3. Перерыв 10 минут
4. Рисуем
5. Для чего все это нам?
6. Вопросы?



варианты

1. Пробуем объяснить себе “на пальцах”
2. Отбираем у школьников (khanacademy.org)
3. Проверяем экспериментами (руками, питоном)
4. Рисуем кружочки
5. Знать, для чего учу




варианты

Жило на свете одно племя и был у них обычай - давать разрешение на свадьбу только после испытания...



варианты

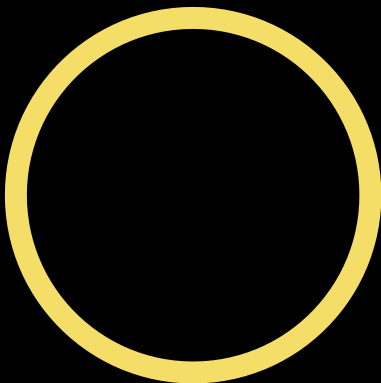
1. Пробуем объяснить себе “на пальцах”
2. Отбираем у школьников (khanacademy.org)
3. Проверяем экспериментами много раз (руками,  питоном)

$$\text{удачные исходы} / \text{возможные исходы} = (4 * 2 * 1) / (5 * 3 * 1) = 8 / 15 \approx 53\%$$

4. Рисуем кружочки (после перерыва)
5. Знать, для чего учу



ПОМОЩНИКИ



смотри Jupyter Notebook



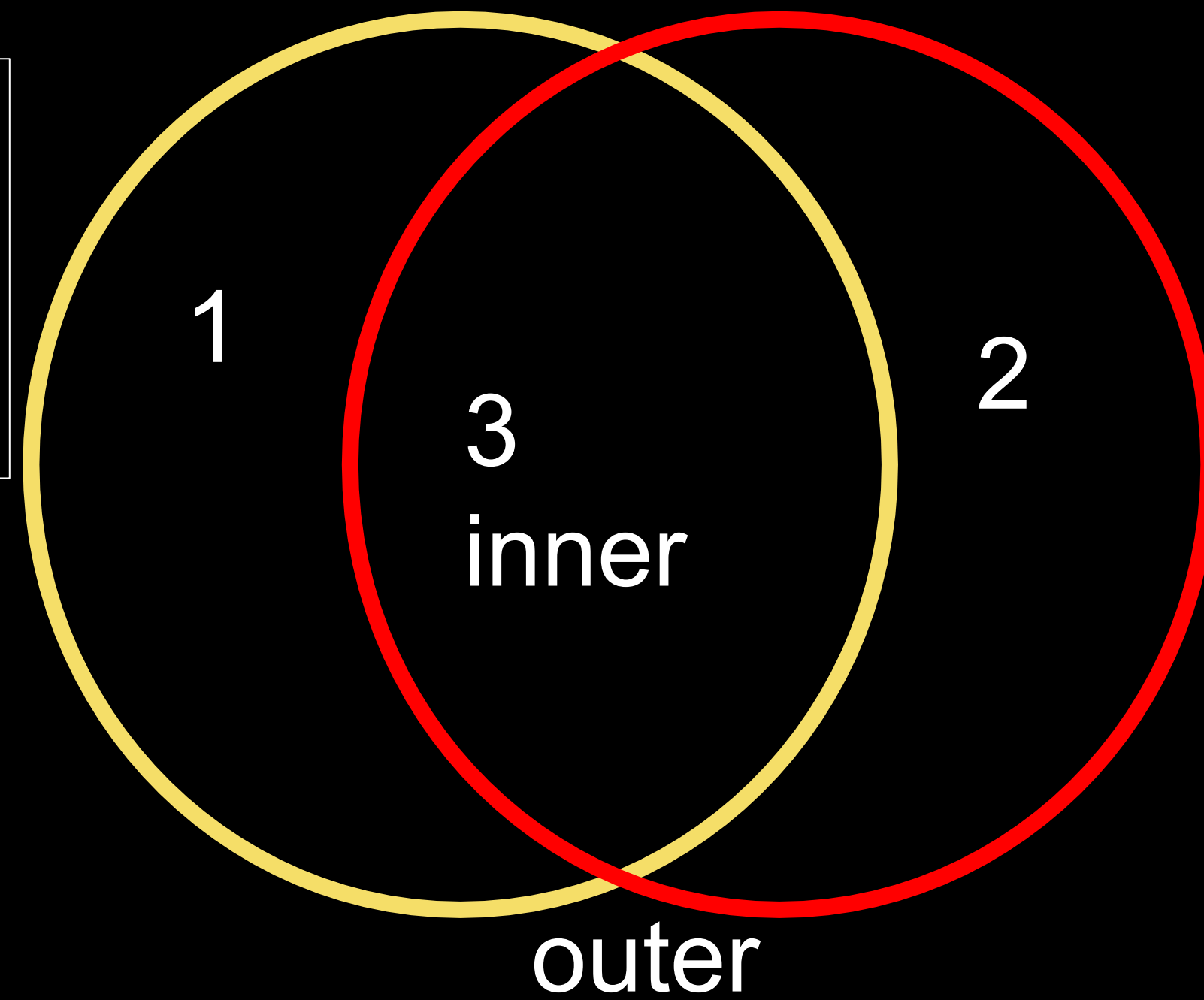
how

собачники left

fio
has_dog

кошатники right

fio
has_cat



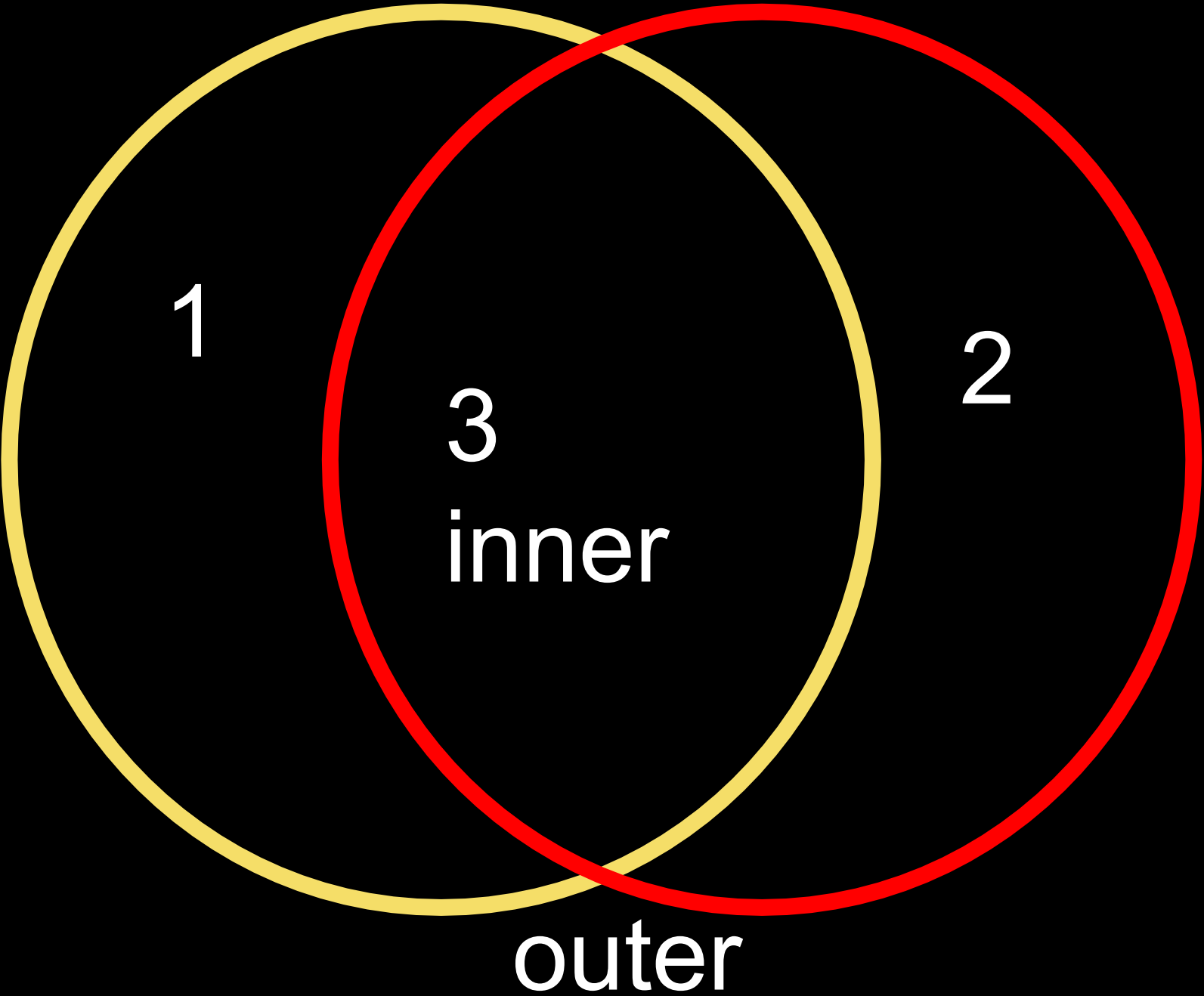
В каком году?

собачники left

кошатники right

fio
has_dog
year

fio
has_cat
year



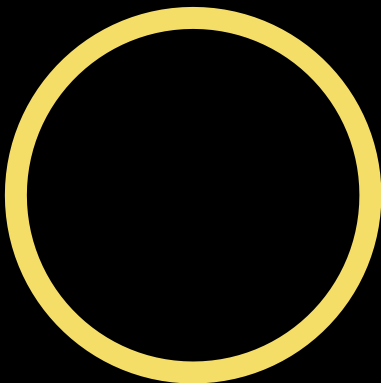
как себя проверить



1. Нарисовать, что ты хочешь на выходе
2. Или представить в виде плоской таблицы
3. Посчитать количество сущностей : сколько собачников у меня было, никого не потеряли?
4. Посчитать количество строк: 5 собачников и 5 кошатников не могут дать больше 10 (кроме декарта)
5. Ключ объединения точно достаточен?
6. Не положу ли я кластер?



перерыв



10 минут



как себя проверить (какие могут быть проблемы)

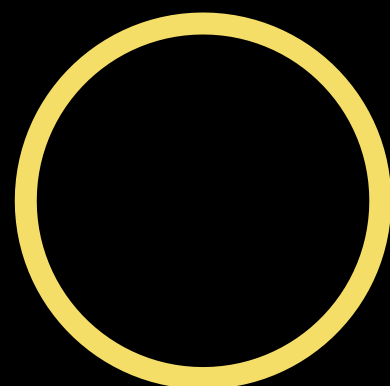
1. Случайная выборка? Точно?!
 2. Объем достаточный?
 3. Форма какая (гистограмма распределения) А почему?
 4. Дисперсия/стандартное отклонение: большое? одинаковая?
0. Про предобработку данных (см. методичку в projects) + смысл



для чего все это нам?

0. Собеседование

1. Тервер - дедукция (какая-то вермодель принимается в качестве истинной и на ее основе делаются выводы)
2. Статистика - индукция (наблюдаем результаты, делаем выводы)
3. “Таблицы смертности”: страхование, гарантийный срок
4. Событие в течении определенного периода времени: кол-во покупателей, посетителей сайта, звонков в хелпдеск, в играх монстры
5. Аномалии: что выбивается из распределения?
6. было 2000 стало 2100 - внедряем?
7. Что я засовываю в машинное обучение?



вопросы?



Спасибо!