Лабораторная работа №2

Выполнил: Поздняков Иван

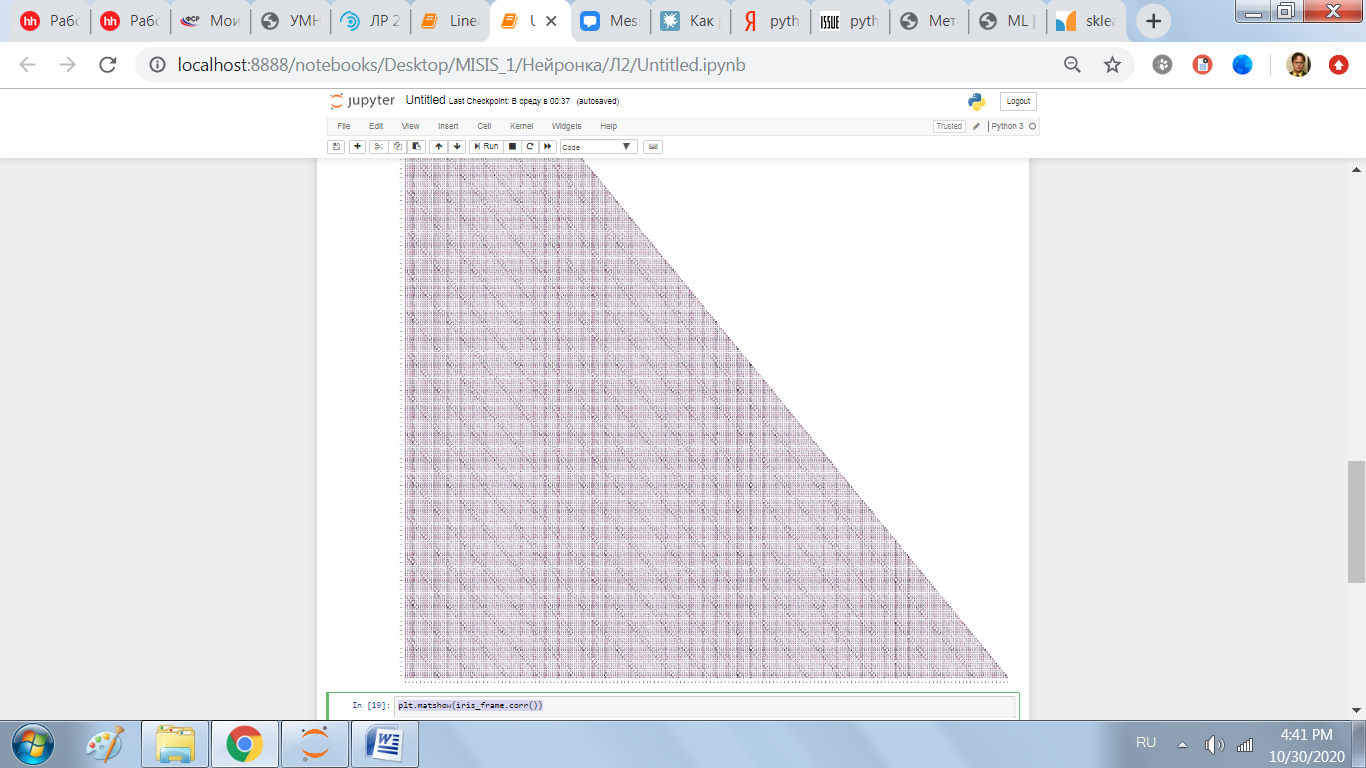
1. я создал две уникальные. В первой два класса и 20 признаков. А также для базы с 1000 признаков, я разработал собственную на основе фотографий. Т.к. их разрешение было велико я их подредактировал. Во-первых, считал из папки все фото, затем сжал 100х100, а затем освободился от RGB, сделав всё чёрно-белым

При построении матриц корреляции возникли трудности, т.к. она получилась большой при использовании sns.heatmap и я попробовал стандартную библиотеку plt.matshow(iris\_frame.corr())

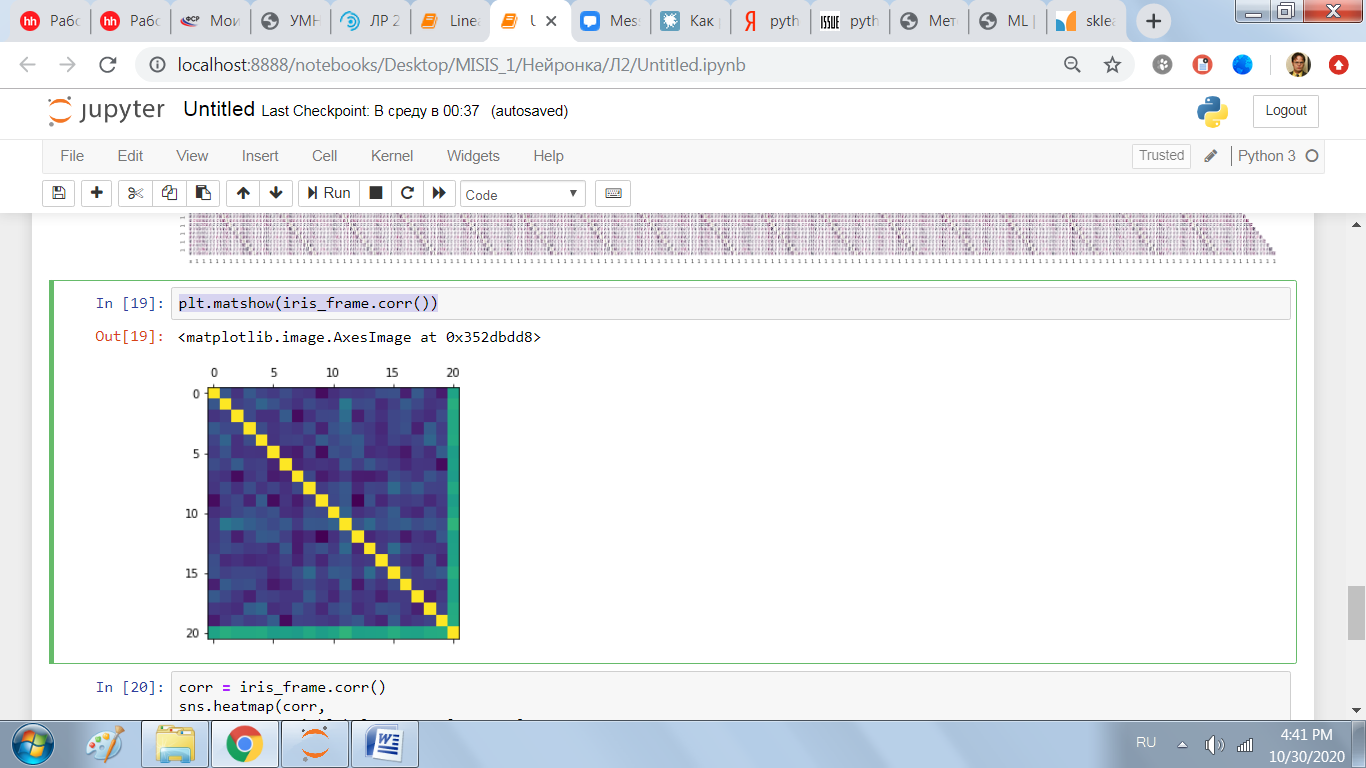
**Датасет 20**

(во всех графиках функция на вход получает дату)

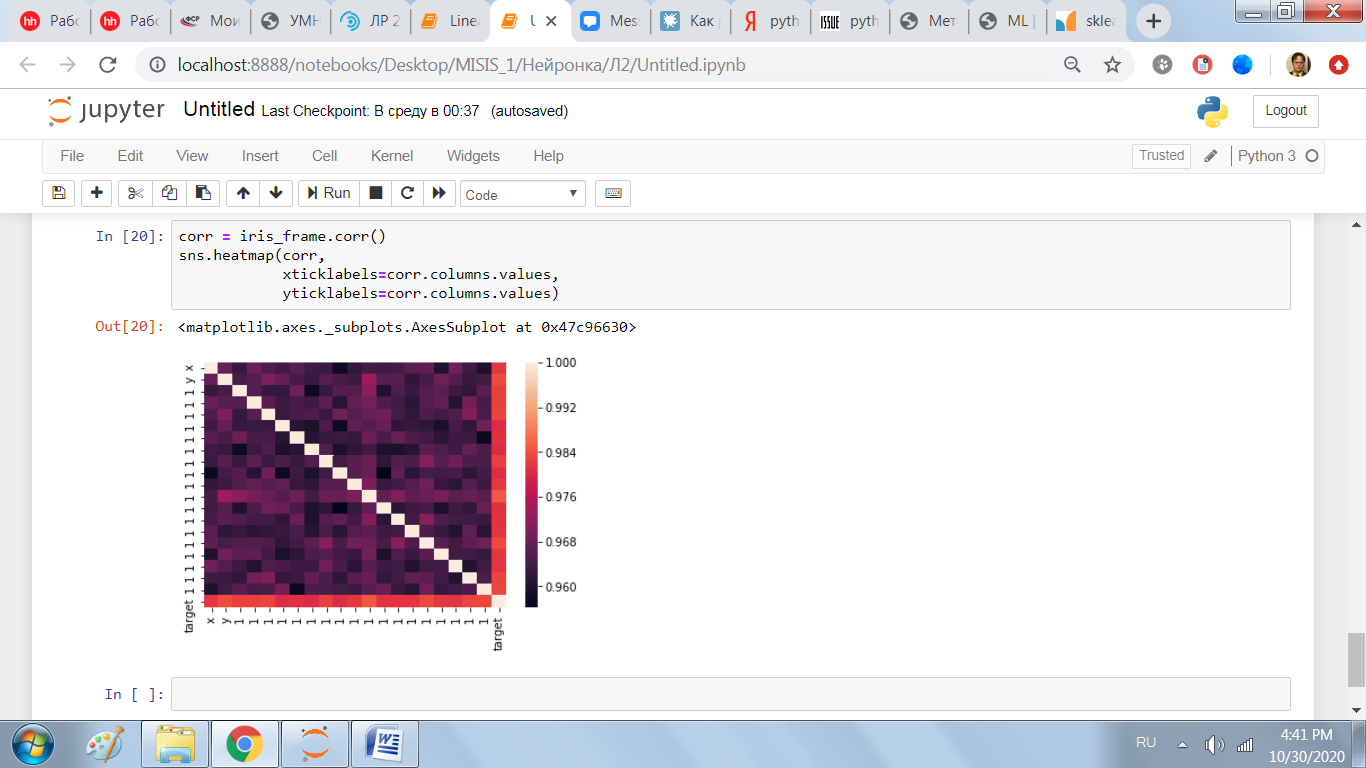
Было



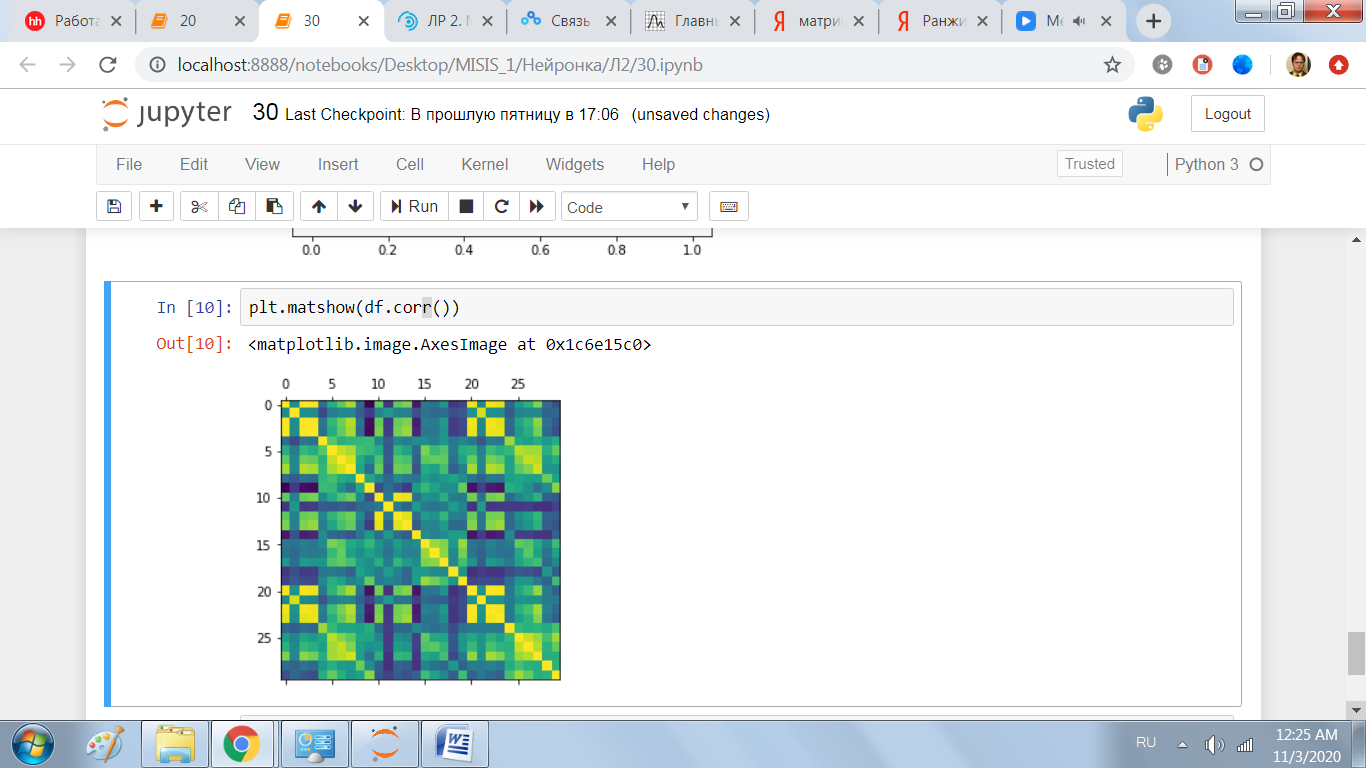
Стало



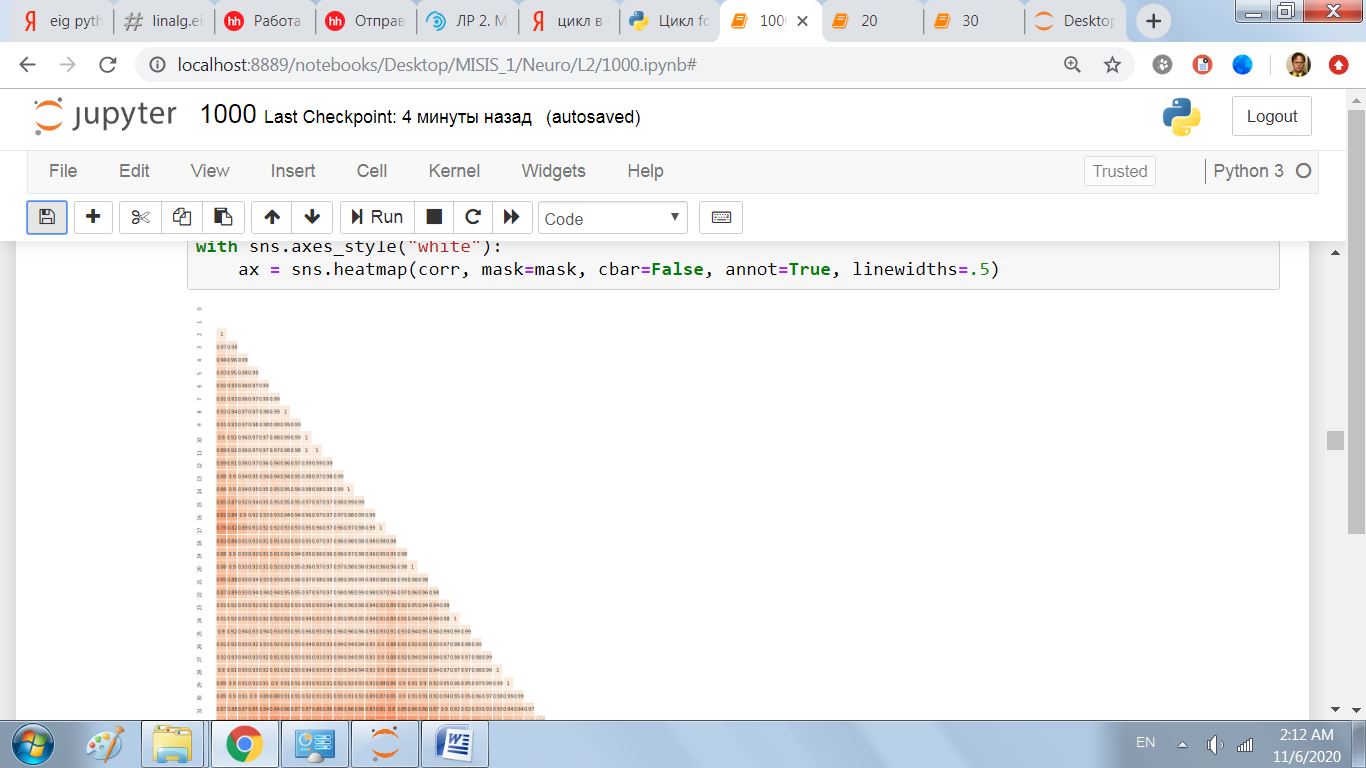
И вот так тоже сделал (по сути тоже самое)

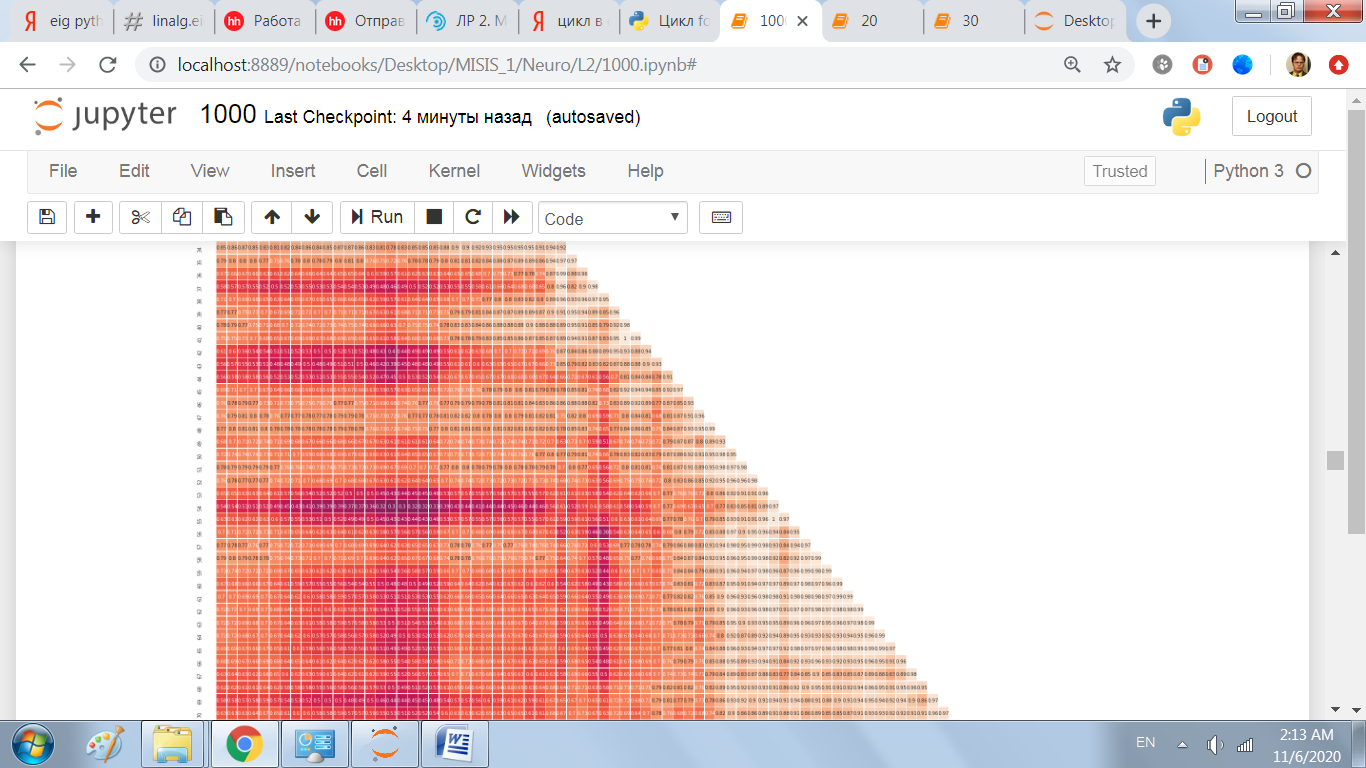


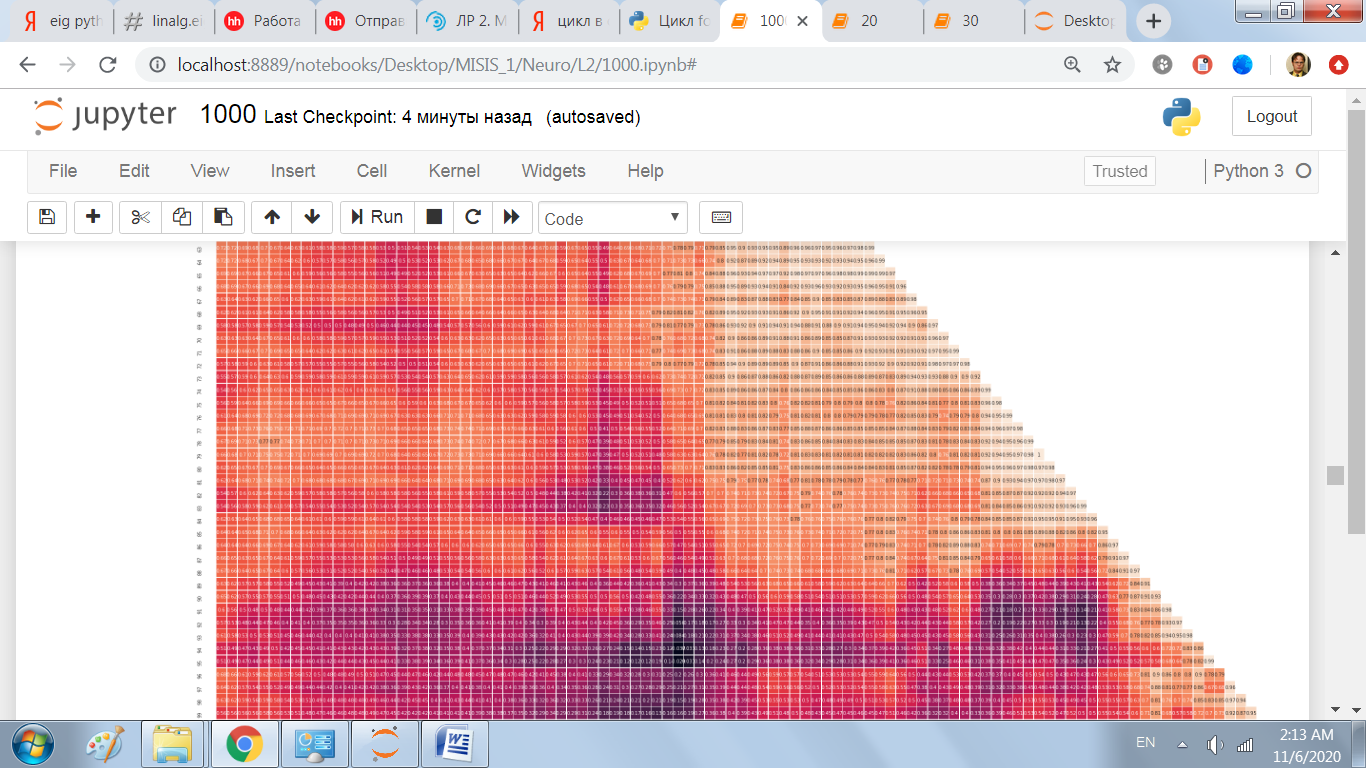
**Датасет 30**



**Датасет 10 000**





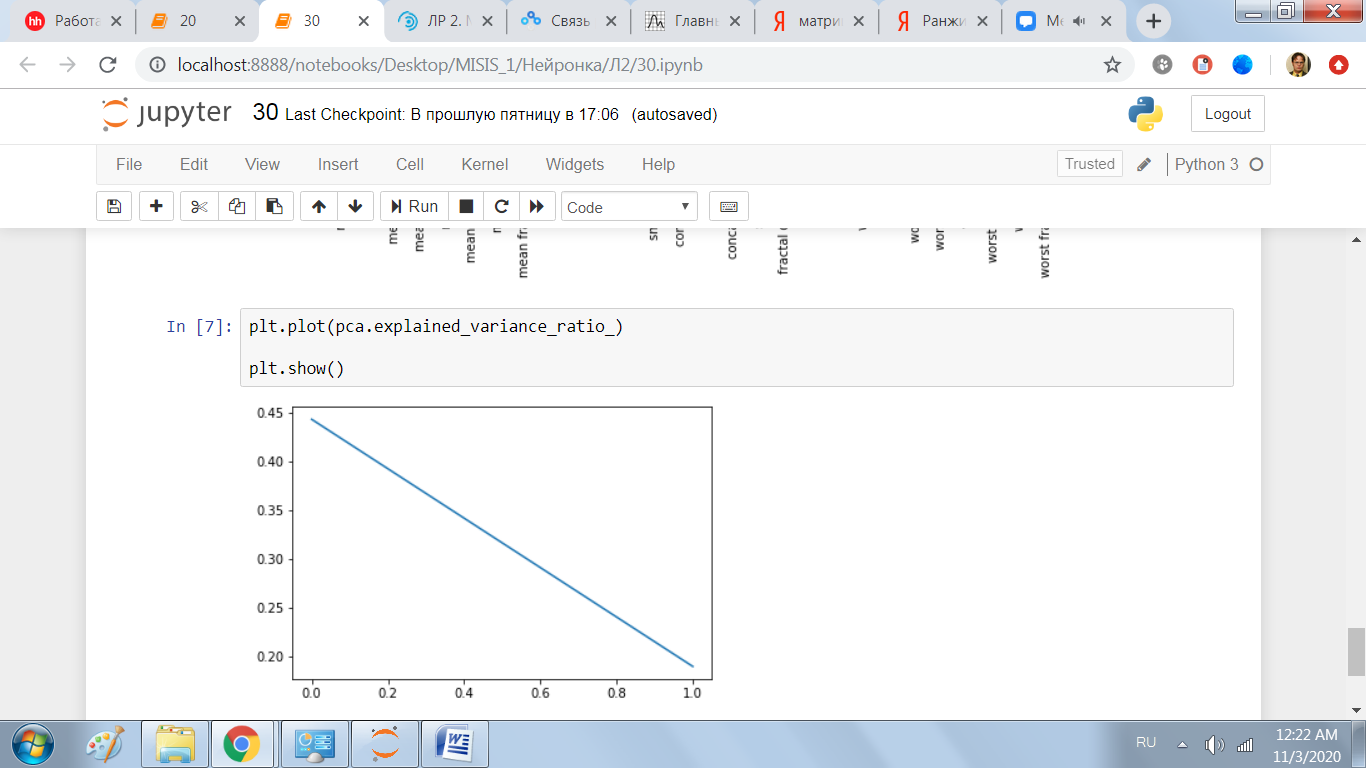


1. Вторым шагом я избавился от признаков, которые плохо коррелируются.
2. Применяя метод главных компонент надо было указать их количество, я выбрал два на каждый класс датасета а также в зависимости от результатов анализов от трёх указанных функций.

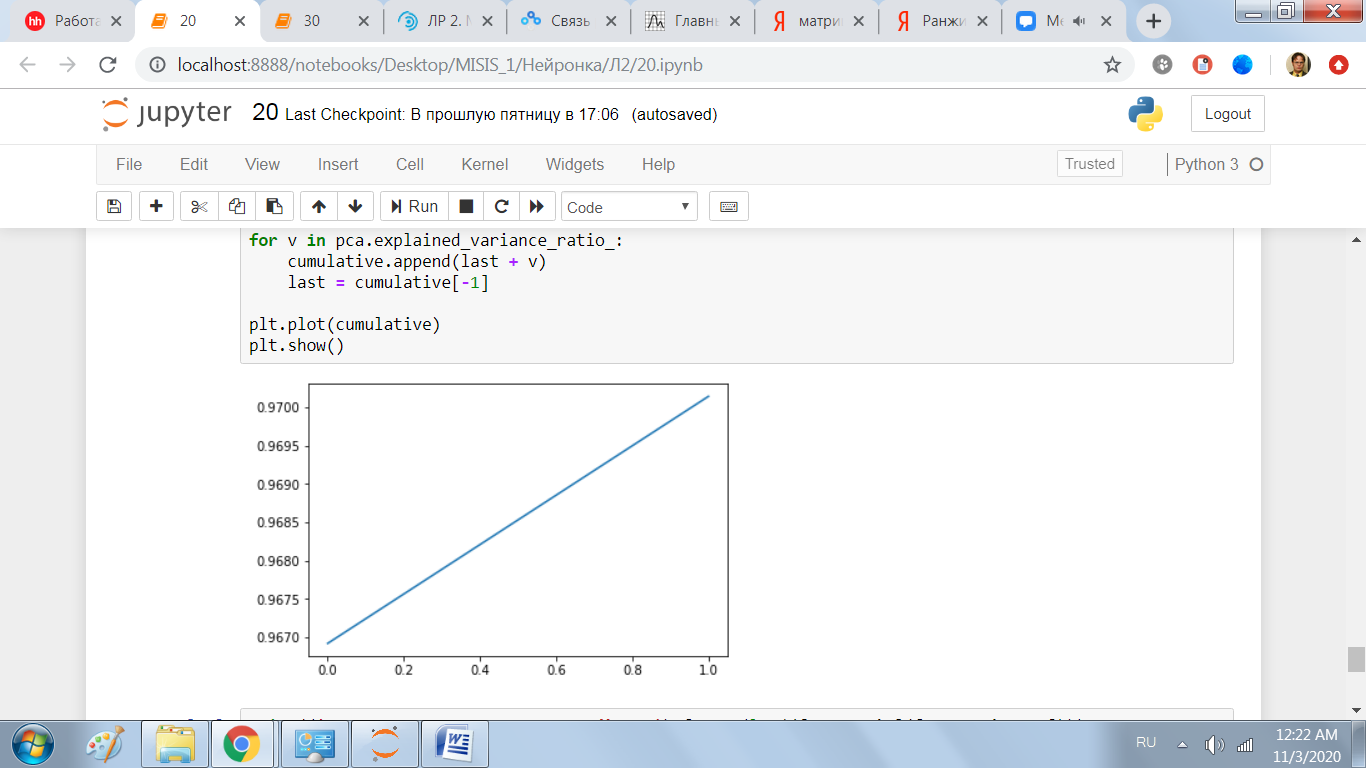
После применения удалось сократить размерность с 30 до 2.

Использовал функцию в sklearn для графика дисперсий

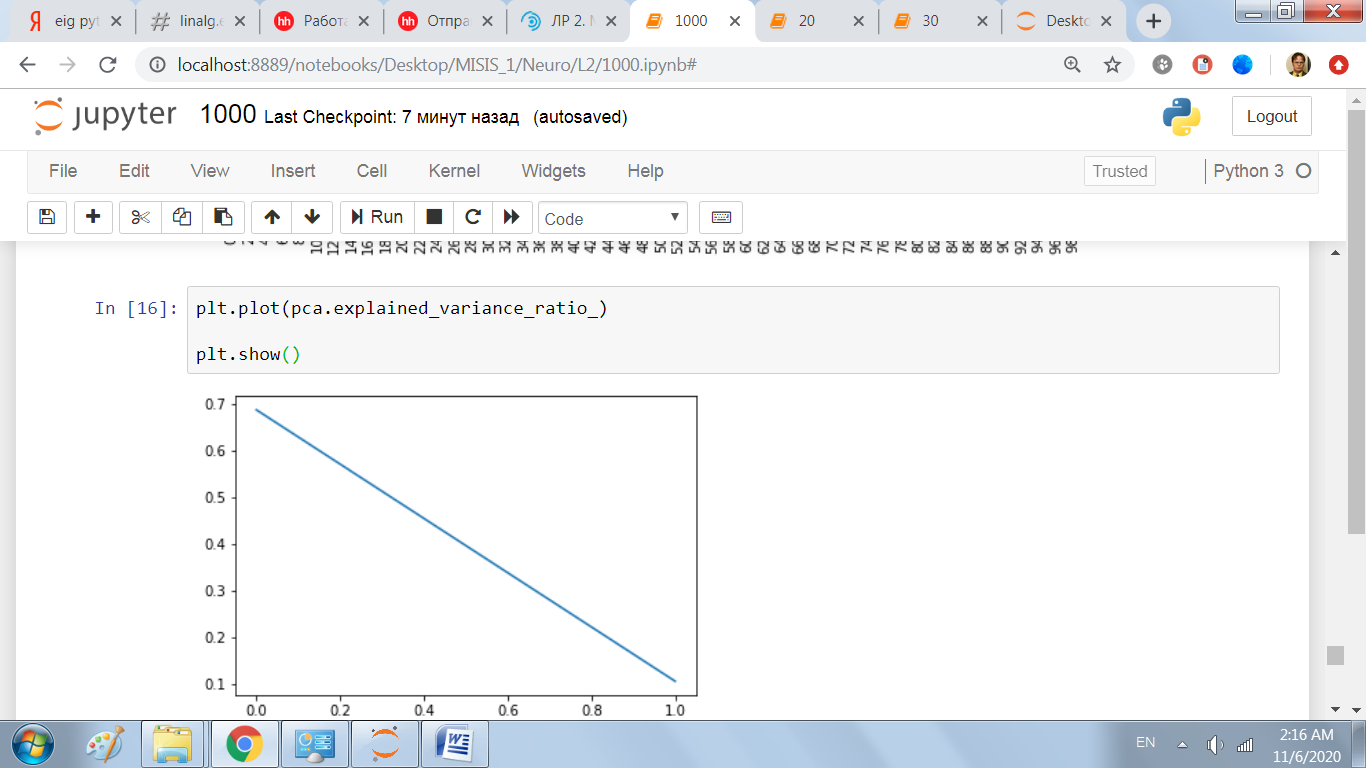
**Датасет 20**



**Датасет 30**

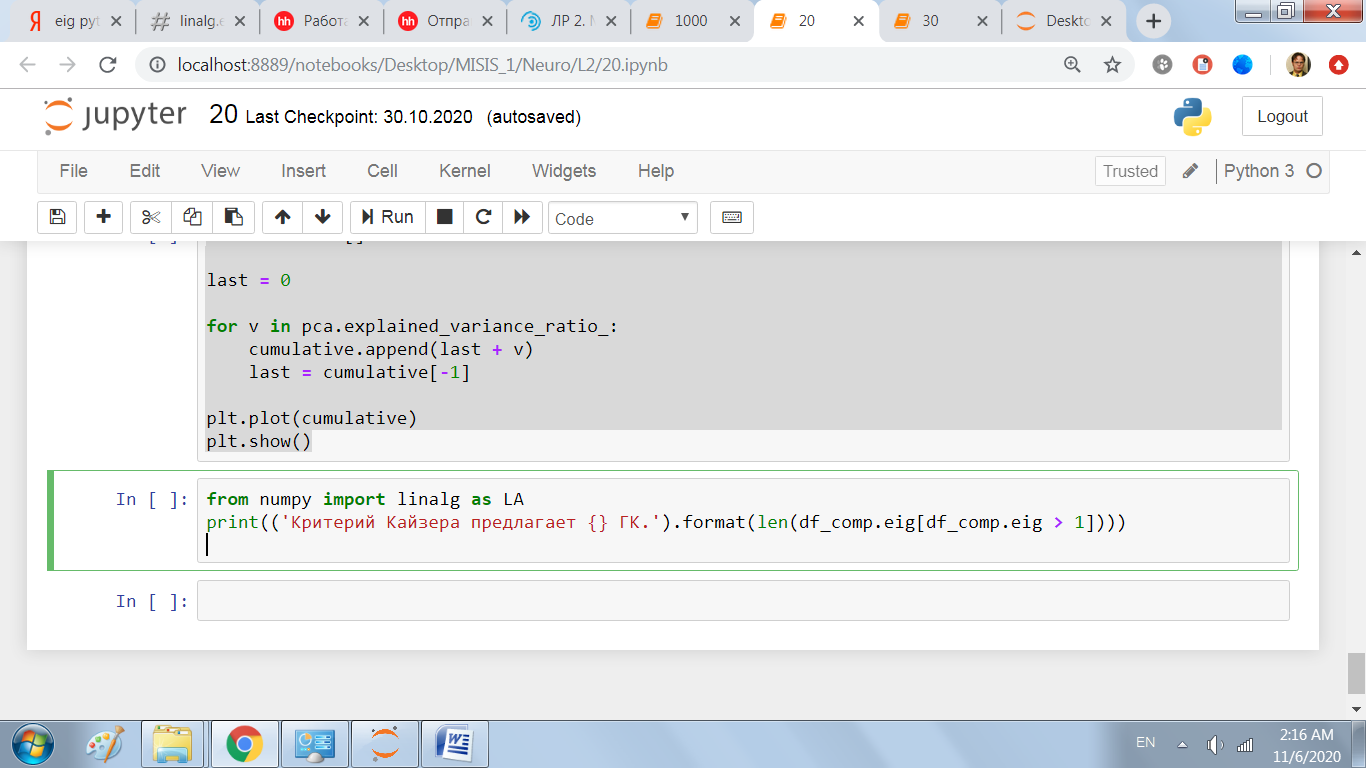


**Датасет 10 000**



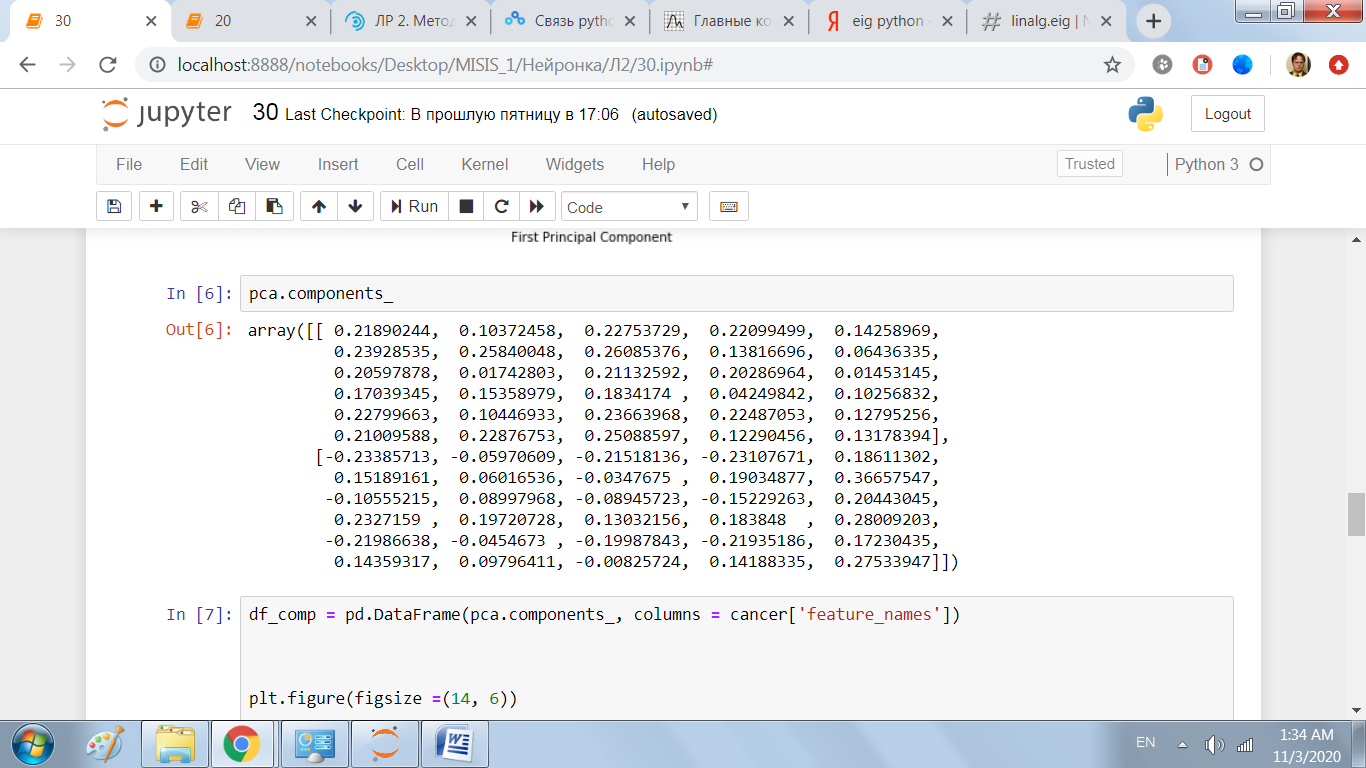
1. Критерий кайзера

Необходимо отобрать только факторы, с собственными значениями, большими *1*



Матрица факторной нагрузки

Дата 30



Дата 20

***Ранжирование*** – это процесс расположения факторов в порядке

возрастания или убывания какого-либо присущего им

свойства, т.е., упрощенно, по их относительной

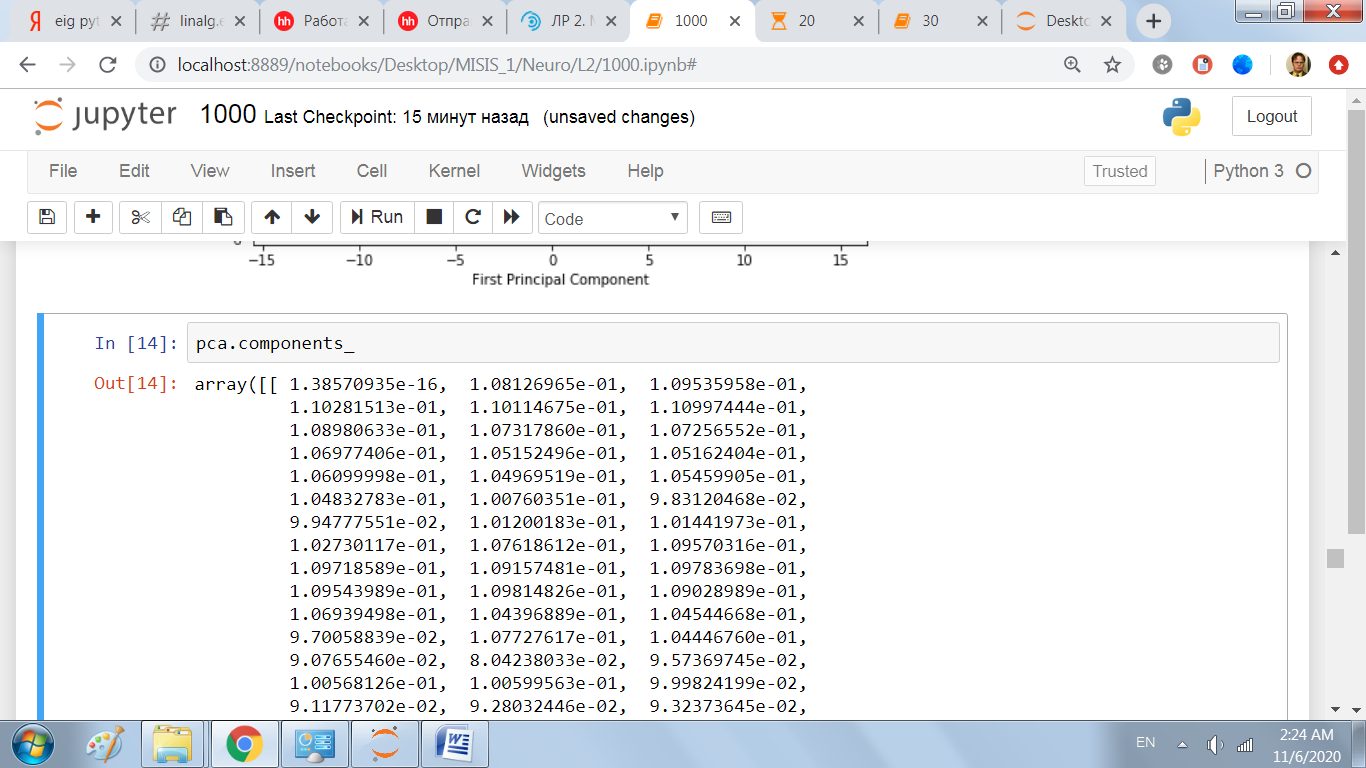
значимости (важности) в решении поставленной проблемы.

Нужно отсортировать??

3. Продемонстрировать различное количество главных компонент по различным критериям

Дата 10 000

При 4

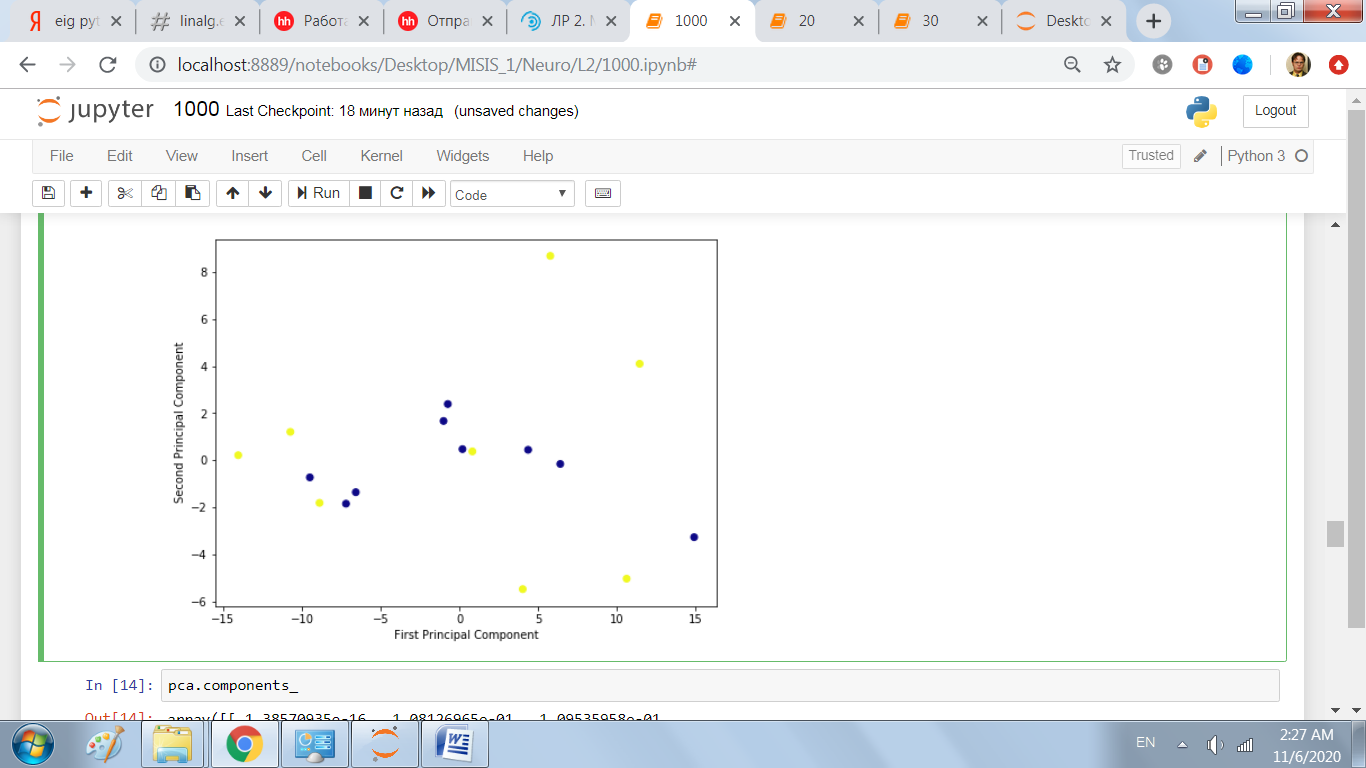


4. Продемонстрировать "хороший" и "плохой" случаи применения МГК. Пояснить, что Вы считаете "плохим", а что "хорошим" случаем работы МГК

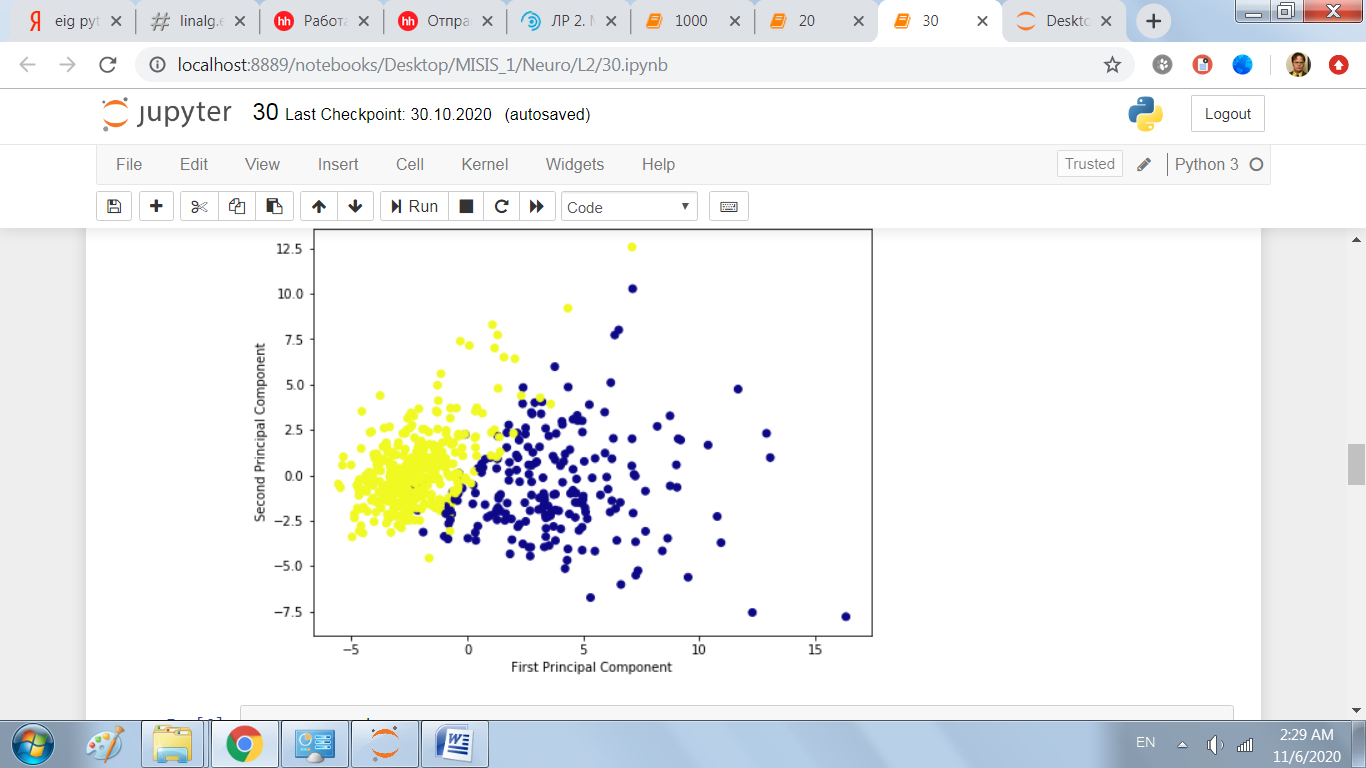
Ответ: Я считаю, что хороший случай отображает выборку с максимальной точностью, чтобы можно было восстановить данные по компонентам.

Также были построены графики по компонентам

**Дата 10000**



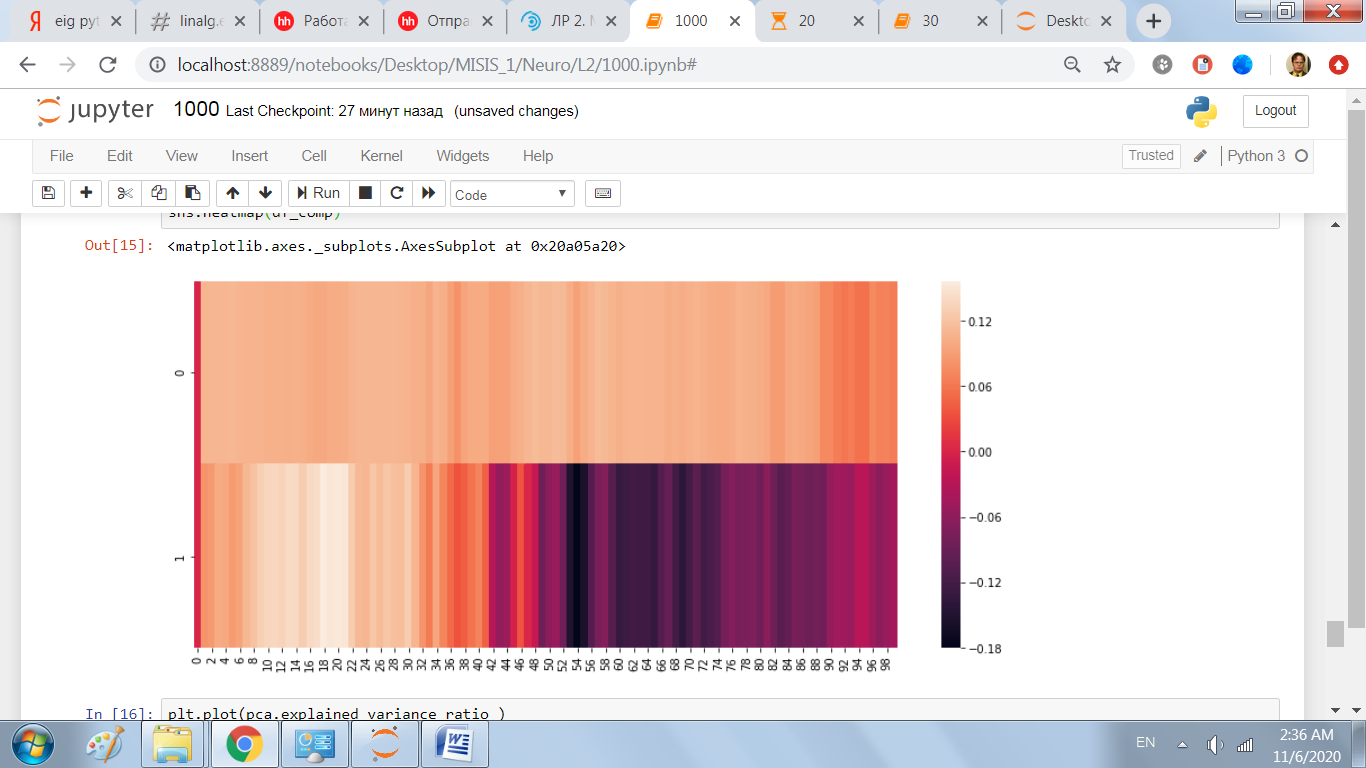
Дата 30



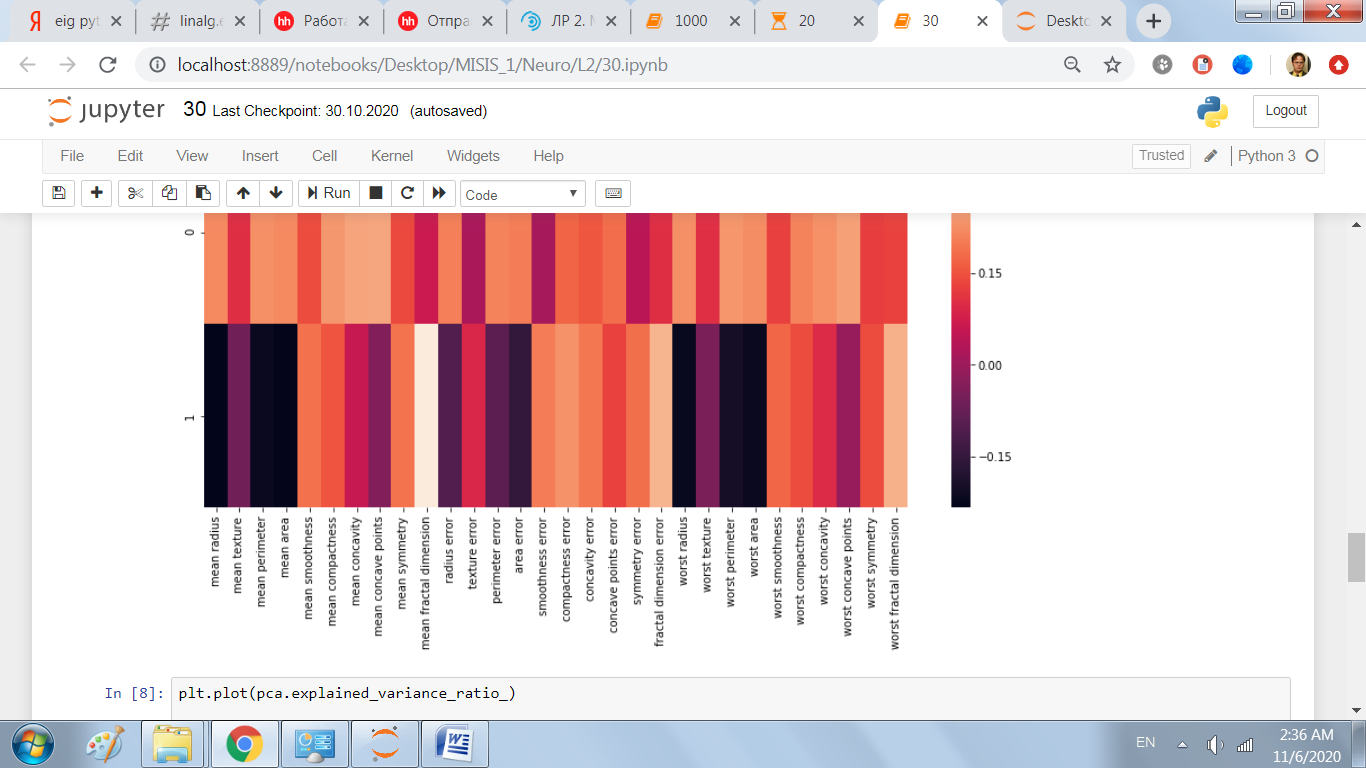
**Дата 20**

**Также были другие версии для матриц**

**1000**

****

**30**

****