МОСКОВСКИЙ ГОСУДАРСТВЕННЫЙ ТЕХНИЧЕСКИЙ УНИВЕРСИТЕТ

им. Н.Э. Баумана

Факультет «Информатика и системы управления»

Кафедра «Систем обработки информации и управления»

ОТЧЕТ

**Домашнее задание №\_\_3\_\_**

по дисциплине«Проектирование интеллектуальных систем»

### Тема: «Методы поддержки принятия решений на основе методов вычислительного интеллекта»

ИСПОЛНИТЕЛЬ: Белоусов Е.А.

ФИО

группа ИУ5-71 \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

подпись

"\_\_"\_\_\_\_\_\_\_\_\_2020 г.

ПРЕПОДАВАТЕЛЬ: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

ФИО

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

подпись

"\_\_"\_\_\_\_\_\_\_\_\_2020 г.

Москва - 2020

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

1. **Задание**

**Задание 1.**

Даны три временных ряда, определите какие из них связаны между собой.

Решить задачу, как минимум двумя методами, сравнить методы между собой и обосновать их выбор для решения данной задачи.

**Задание 2.**

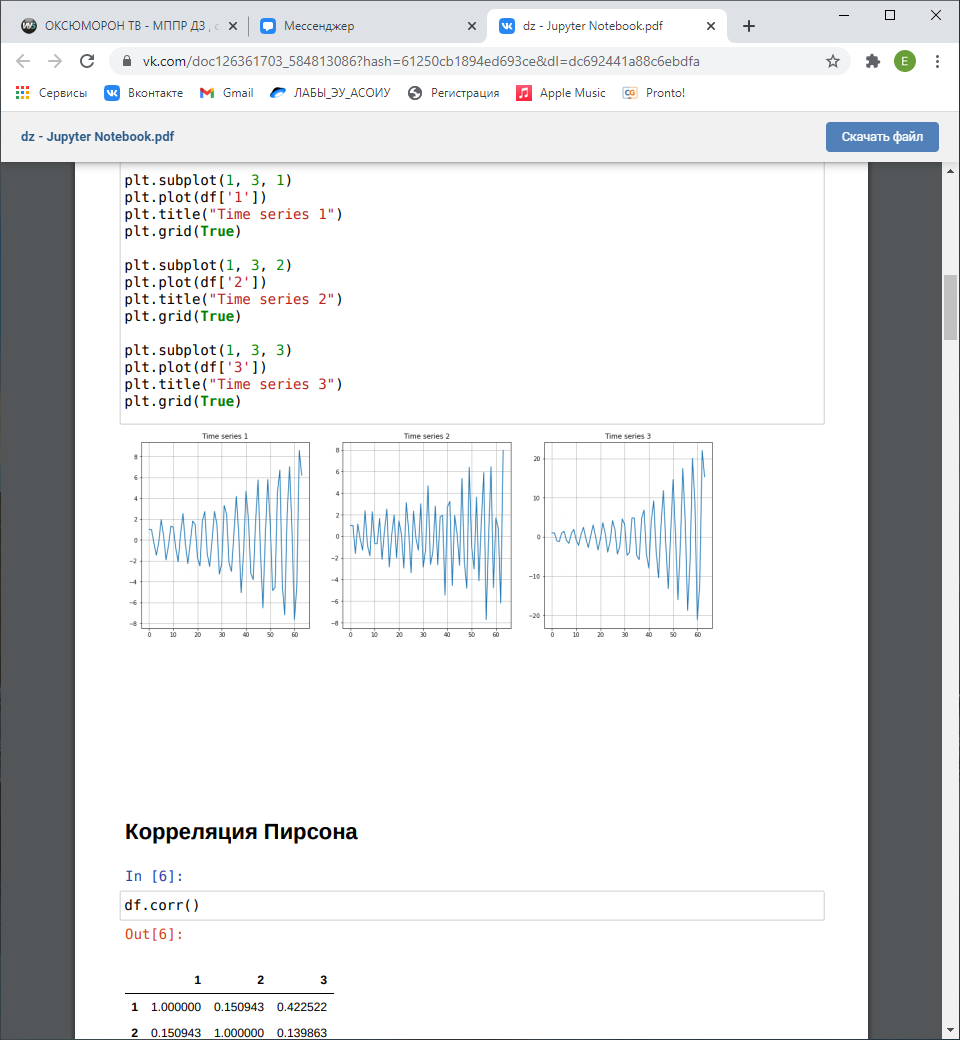
Дано множество пар чисел. Разделить их на 2 класса примерно одинаковой мощности.

Оценить качество разделения двумя методами.

Обосновать выбор методов решения задачи.

1. **Задание 1**

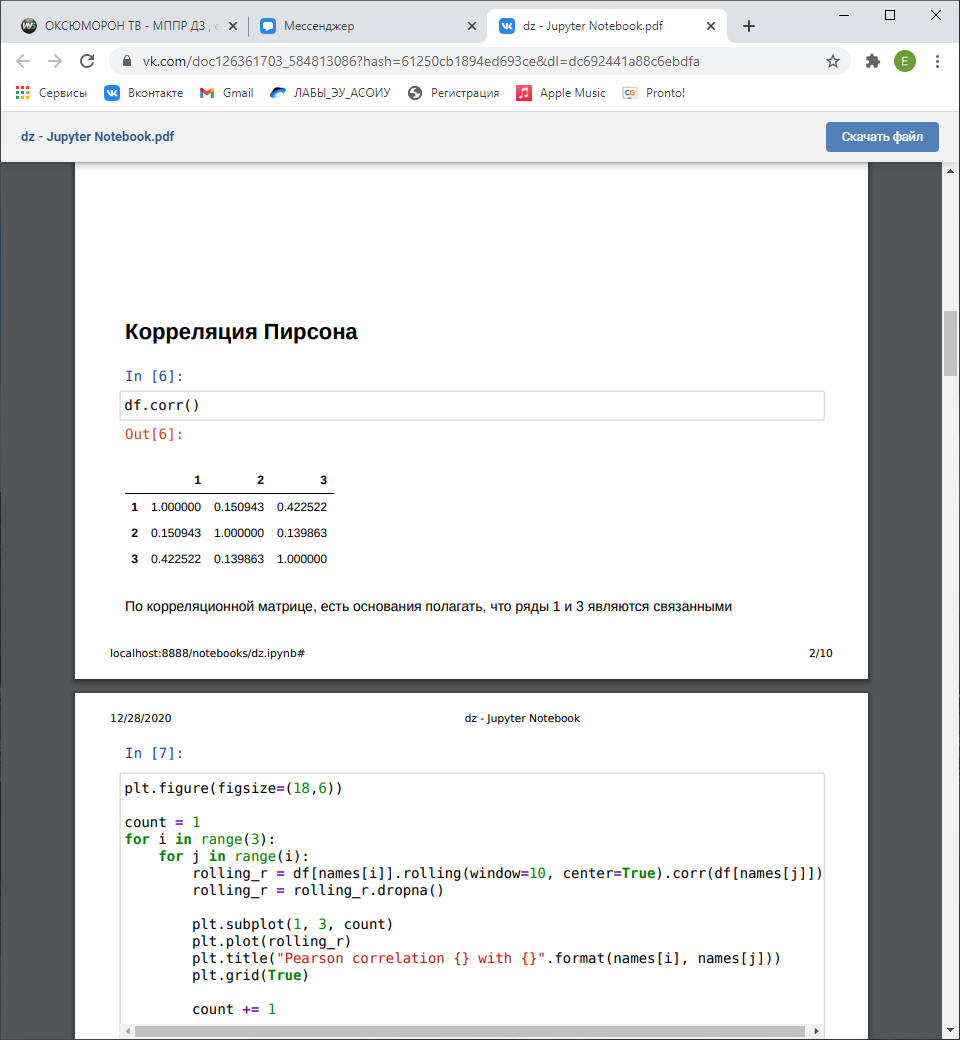
**Даны 3 временных ряда:**



**Корреляция Пирсона**

Коэффициент корреляции Пирсона рассчитывается по формуле:

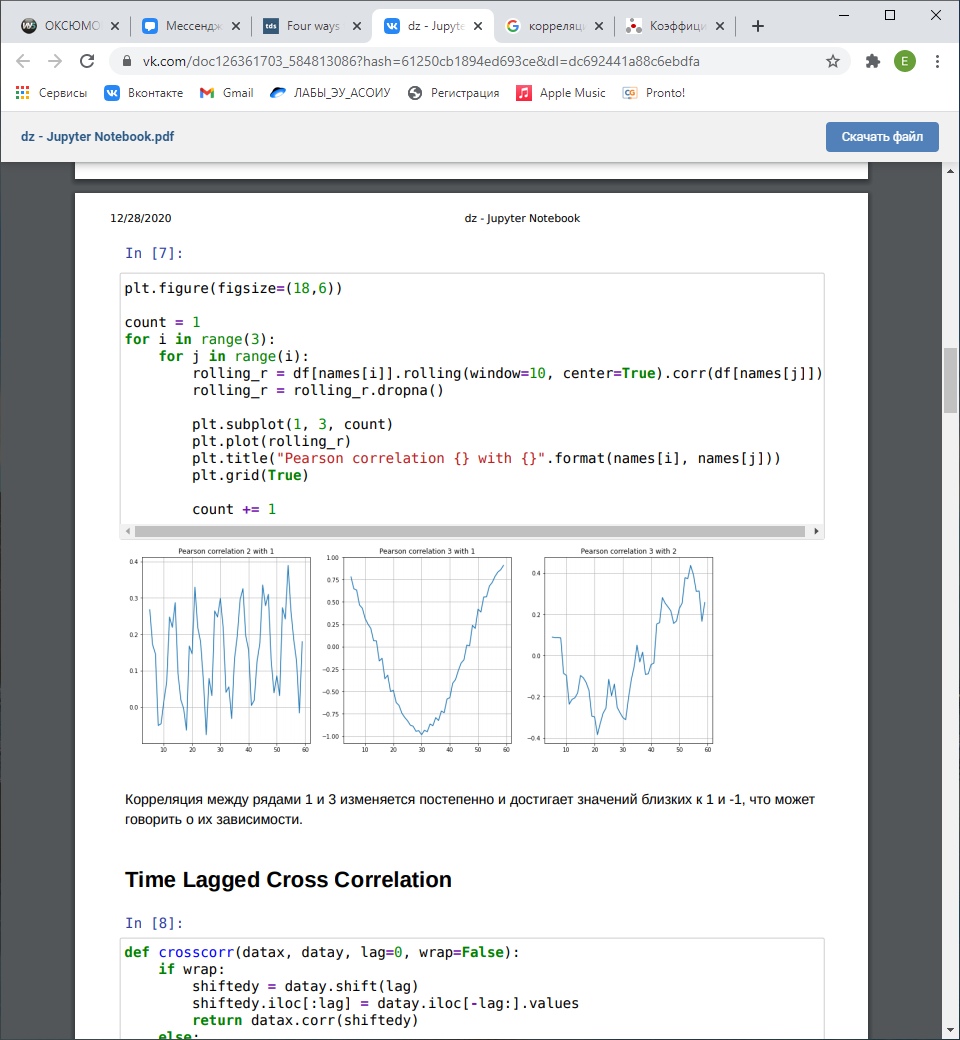
Результат:



По корреляционной матрице, есть основания полагать, что ряды 1 и 3 являются связанными, однако связь не является ярко выраженной.

**Windowed cross correlation**

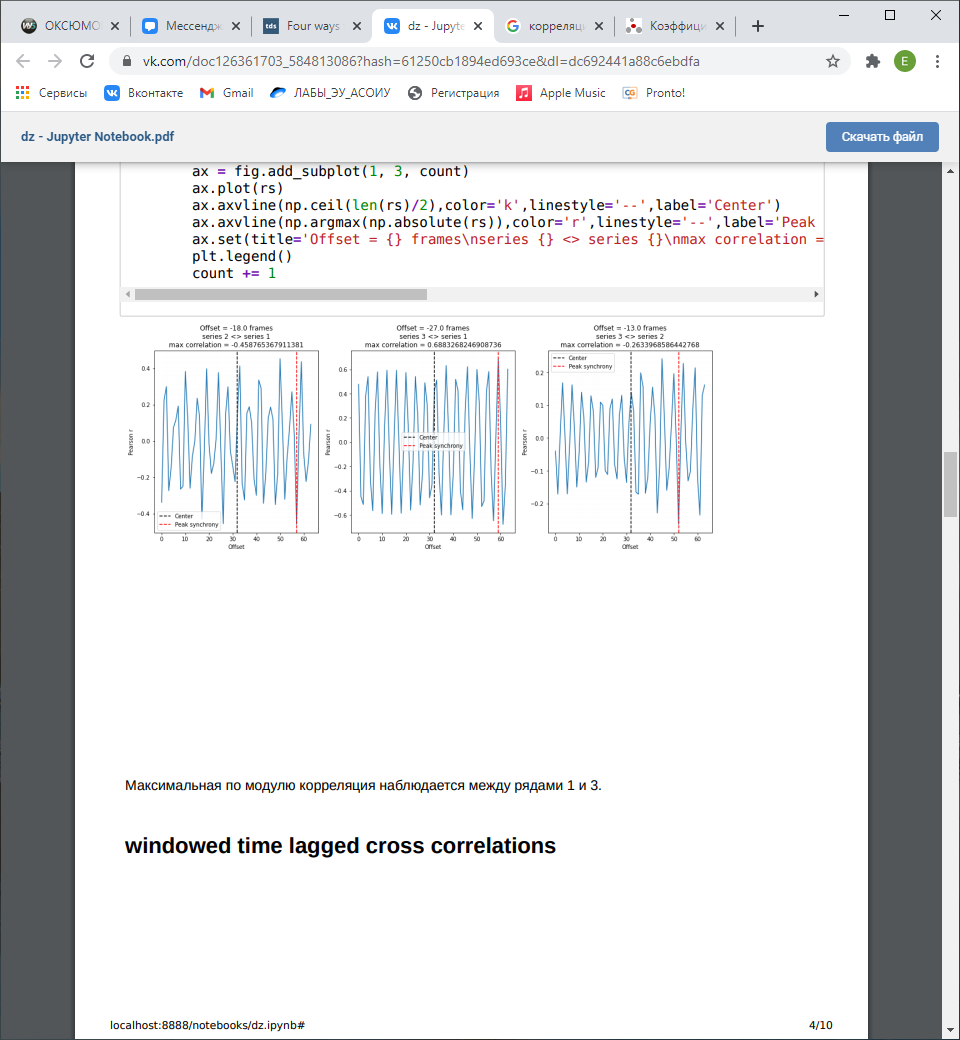
В данном методе рассчитывается коэффициент корреляции в узком окне (в нашем случае в 10 измерений). Метод позволяет выявить участки, на которых связь между временными рядами становится более выраженной.



По графикам, видно, что корреляция между 1 и 3 рядами изменяется плавно и достигает значений близких 1 и -1.

**Time lagged cross correlation**

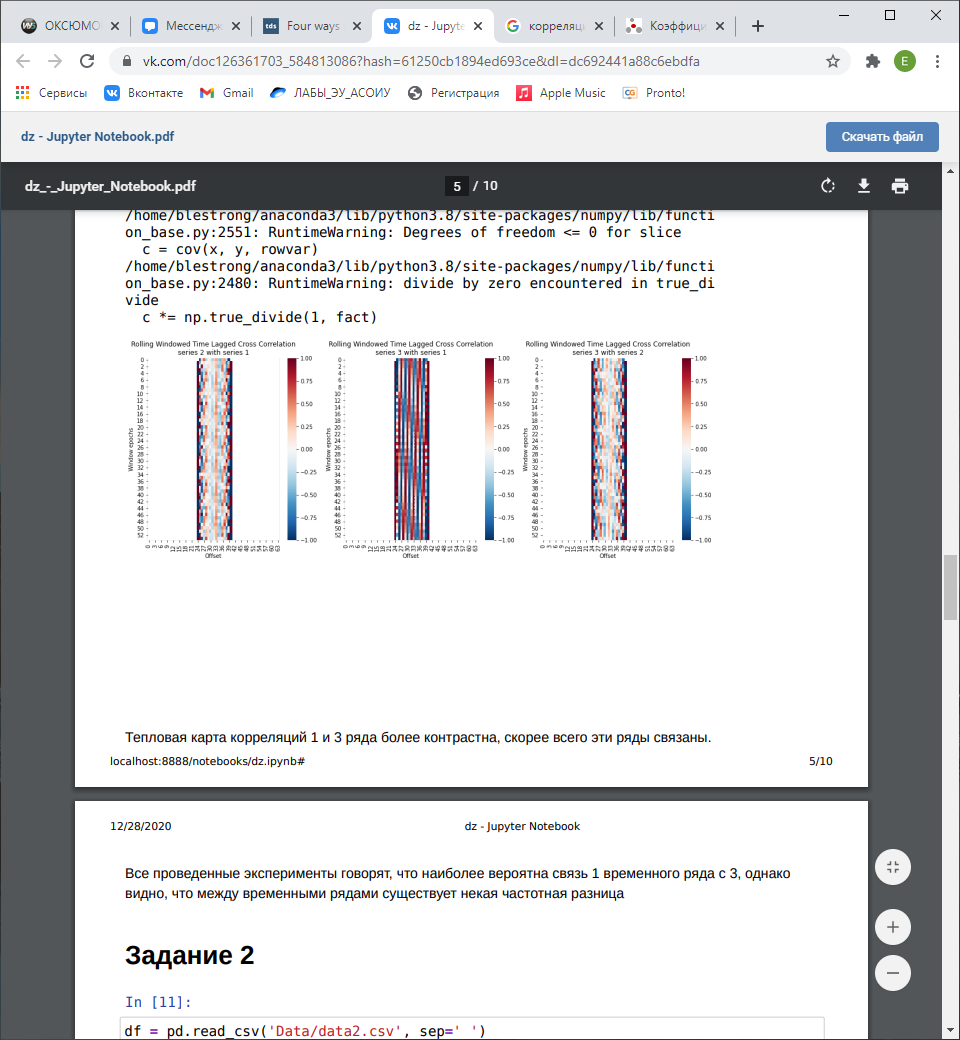
Данный метод производит подсчет корреляции между двумя рядами при сдвиге одного ряда относительно другого. Это может оказаться полезно, в случае, если зависимый сигнал немного запаздывает.



Опять же, в графике отображающем зависимость корреляции 1 и 3 временных рядов, видим наибольшие по модулю значения корреляции.

**Windowed time lagged cross correlations**

Объединим Time lagged cross correlation и Windowed cross correlation между собой и выведем объемный по своей природе график в виде тепловой карты, на которой можно наблюдать как соотносятся между собой небольшие участки зависимостей при сдвиге.



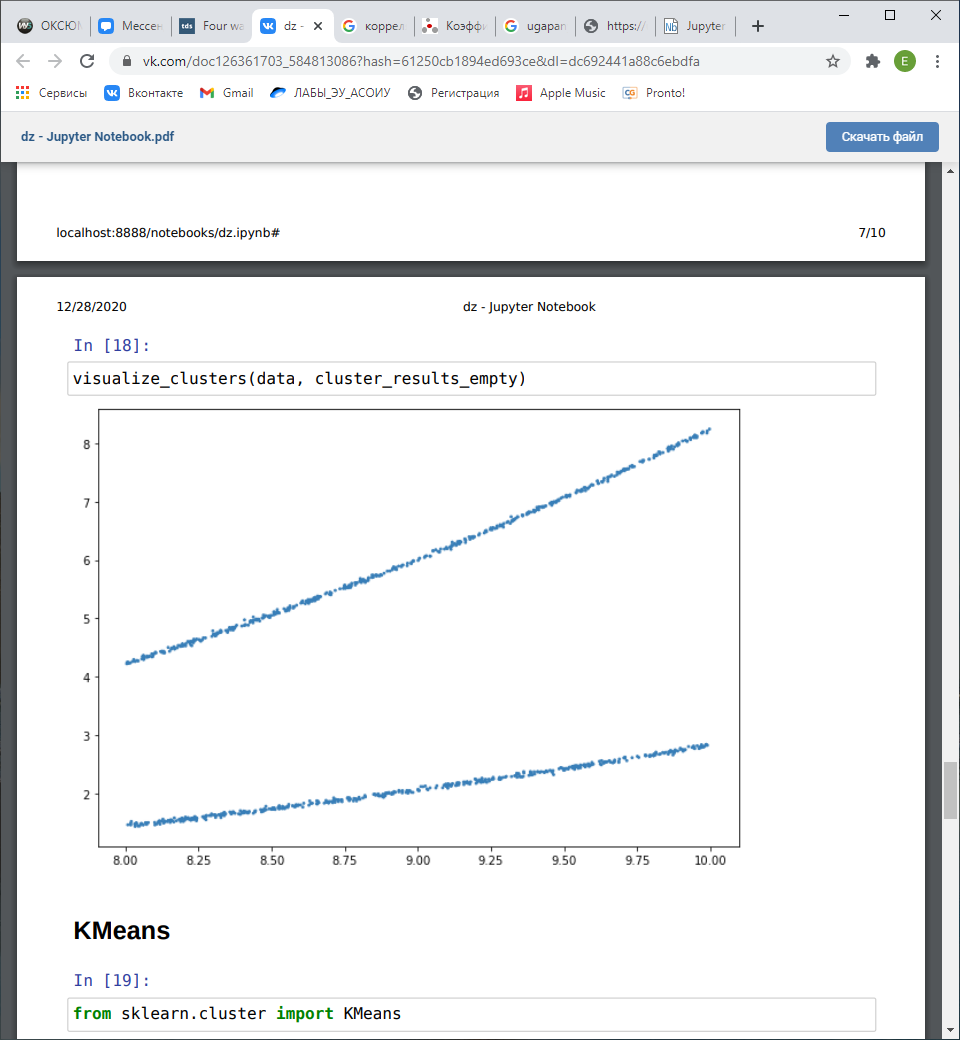
На данных графиках видно, что зависимость временных рядов 1 и 3 является коррелированной, однако, при каждый столбец меняет свой цвет с красного на синий и наоборот. Остальные ряды слабее зависят друг от друга, хотя, стоит заметить, что при смещении +8 и -8 все ряды показывают сильную корреляцию, которая меняется по знаку.

Данные методы различаются между собой простотой и количеством отображаемой информации. В случае дефицита ресурсов и невысокой важности результата следует пользоваться корреляционной матрицей. Когда же имеется неограниченный ресурс или важность результата очень высока, надо использовать Windowed time lagged cross correlations.

# Также, интересным видом анализа является  Instantaneous phase synchrony, который не приведен в работе из-за дефицита ресурсов.

1. **Задание 2**

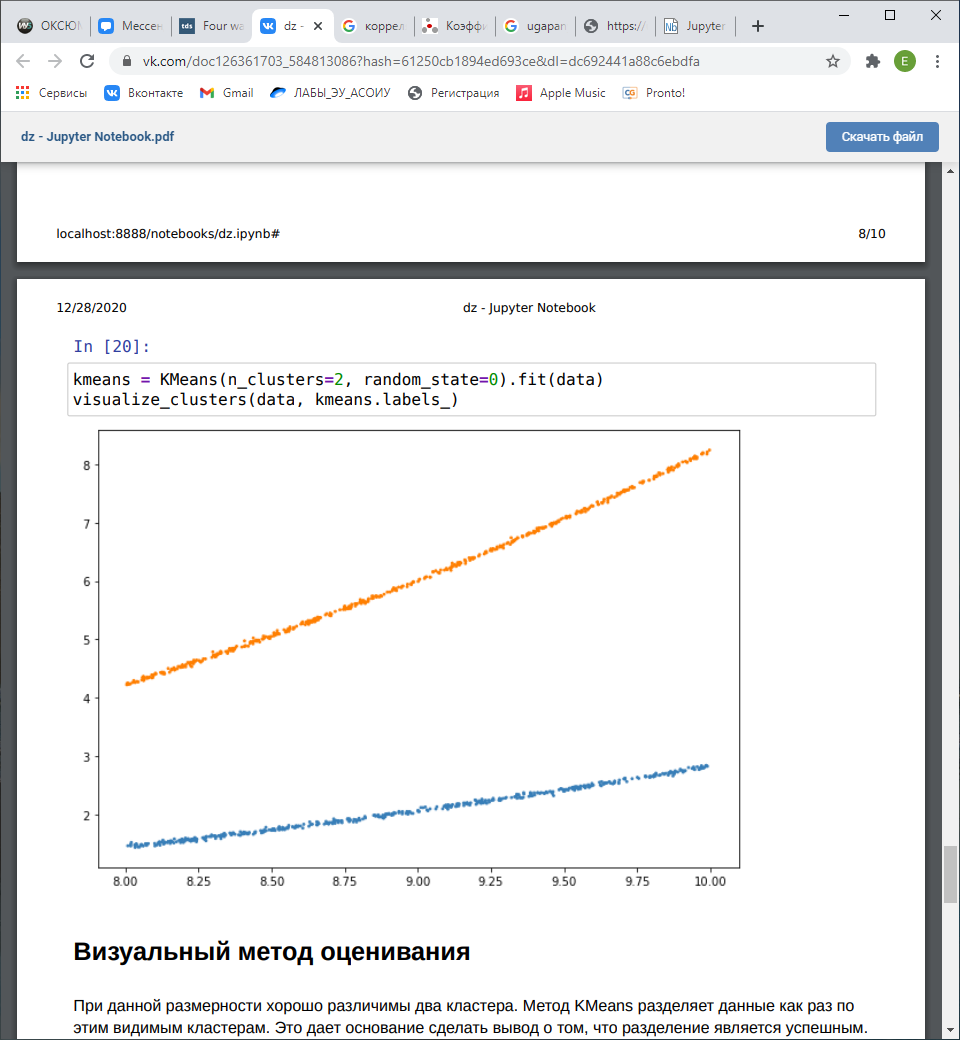
Необходимо разделить исходные данные на два кластера и оценить работу алгоритма двумя методами. Данные приведены на графике ниже.



**KMeans**

Один из самых популярных методов кластеризации – метод k средних, для его работы необходимо знать количество кластеров, которое в данной задаче известно и равно 2.

Результат работы алгоритма:



**Визуальный метод оценивания**

При данной размерности среди данных, невооруженным глазом различимы два кластера. Метод KMeans разделяет данные как раз по этим видимым кластерам. Это дает основание сделать вывод о том, что разделение является успешным.

Если данные не являются хорошо различимыми, можно попробовать изменить их размерность, однако в данной конкретной задаче это не требуется.

Недостатком метода является невозможность точной численной интерпретации результата.

**Коэффициент силуэта**

Данный метод не требует знания истинных значений меток кластеров.

Пусть:

* *a* - среднее расстояние между текущей точкой и другими точками этого же кластера.
* *b* - среднее расстояние между текущей точкой и другими точками следующего ближайшего кластера.

Тогда коэффициент силуэта для точки (объекта) определяется как:

Силуэтом выборки называется средняя величина силуэта объектов данной выборки. Таким образом, силуэт показывает, насколько среднее расстояние до объектов своего кластера отличается от среднего расстояния до объектов других кластеров. Данная величина лежит в диапазоне [-1;1]. Значения, близкие к -1, соответствуют плохим (разрозненным) кластеризациям, значения, близкие к нулю, говорят о том, что кластеры пересекаются и накладываются друг на друга, значения, близкие к 1, соответствуют "плотным" четко выделенным кластерам. Таким образом, чем больше силуэт, тем более четко выделены кластеры, и они представляют собой компактные, плотно сгруппированные облака точек.

Для кластеризации наших данных алгоритмом KMeans значение коэффициента силуэта = 0,7, что говорит о хорошем разделении данных.

1. **Литература**
   1. <https://towardsdatascience.com/four-ways-to-quantify-synchrony-between-time-series-data-b99136c4a9c9>
   2. <https://stats.stackexchange.com/questions/133155/how-to-use-pearson-correlation-correctly-with-time-series>
   3. Конспект лекций ТМО 2020
   4. <https://scikit-learn.org/stable/modules/generated/sklearn.metrics.silhouette_score.html>
2. **Вывод**

При выполнении данного домашнего задания, я познакомился с методами определения связанности между временными рядами и вспомнил методы машинного обучения без учителя.