



ЗАНЯТИЕ 1.0

Временные ряды

ЦЕЛИ ЗАНЯТИЯ

Сегодня:

- Что такое временной ряд
- Автокорреляция
- Стационарность и как ее достичь
- Авторегрессия
- Тест Дики-Фуллера

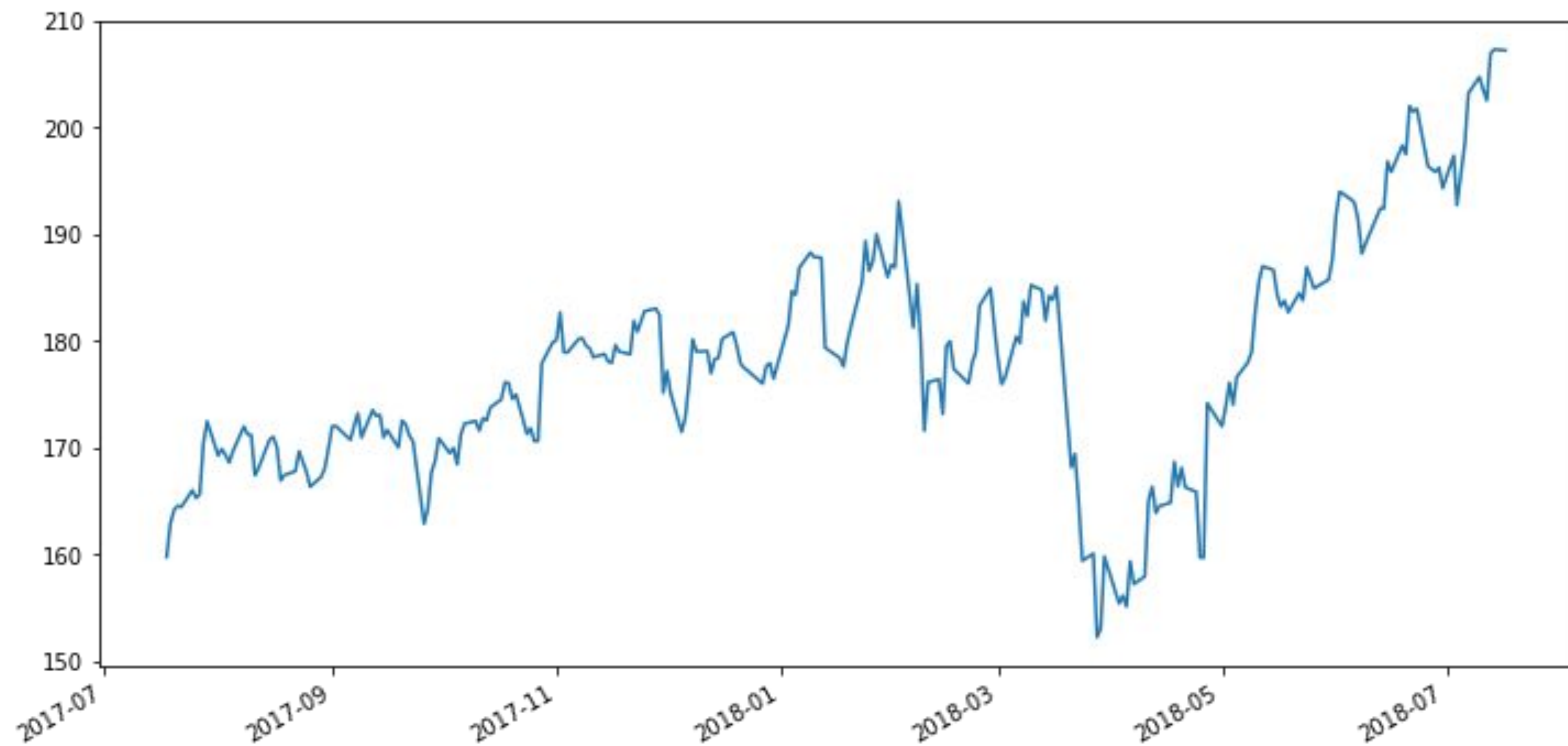
Временной ряд

ВРЕМЕННОЙ РЯД:

y_1, \dots, y_T, \dots , где $y_t \in R$, – значения признака, измеренные через **постоянные** временные интервалы.

ЦЕЛИ ЗАНЯТИЯ

Пример:



Примеры:

- Финансовый сектор
- Показания датчиков
- Объемы продаж/производства
- Телеметрия it-систем (gps/утилизация сри/памяти и тп)
- ...

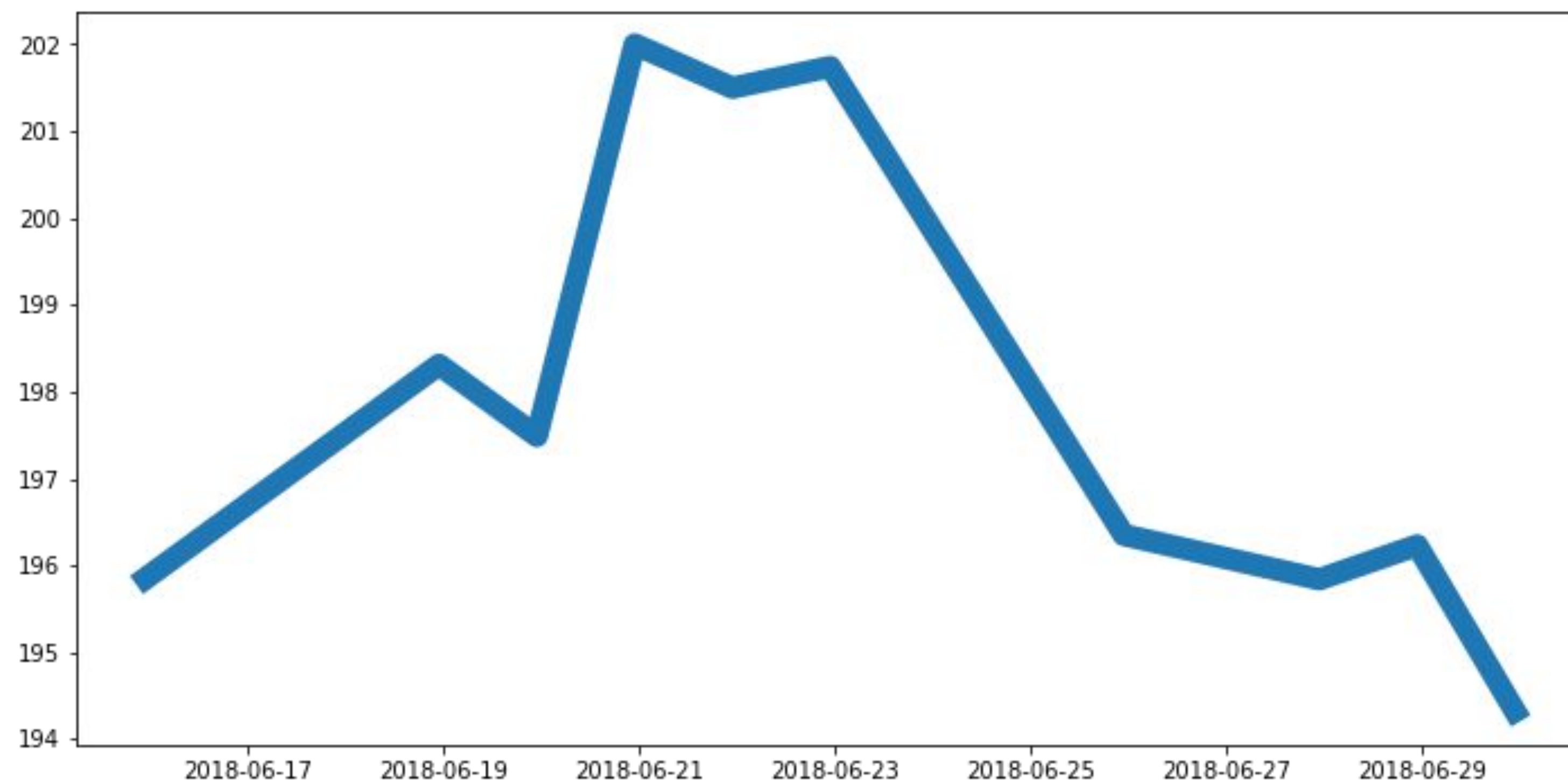
Способы визуализации:

- Простой график
- Ящик с усами (box plot)
- Свечи (candles)

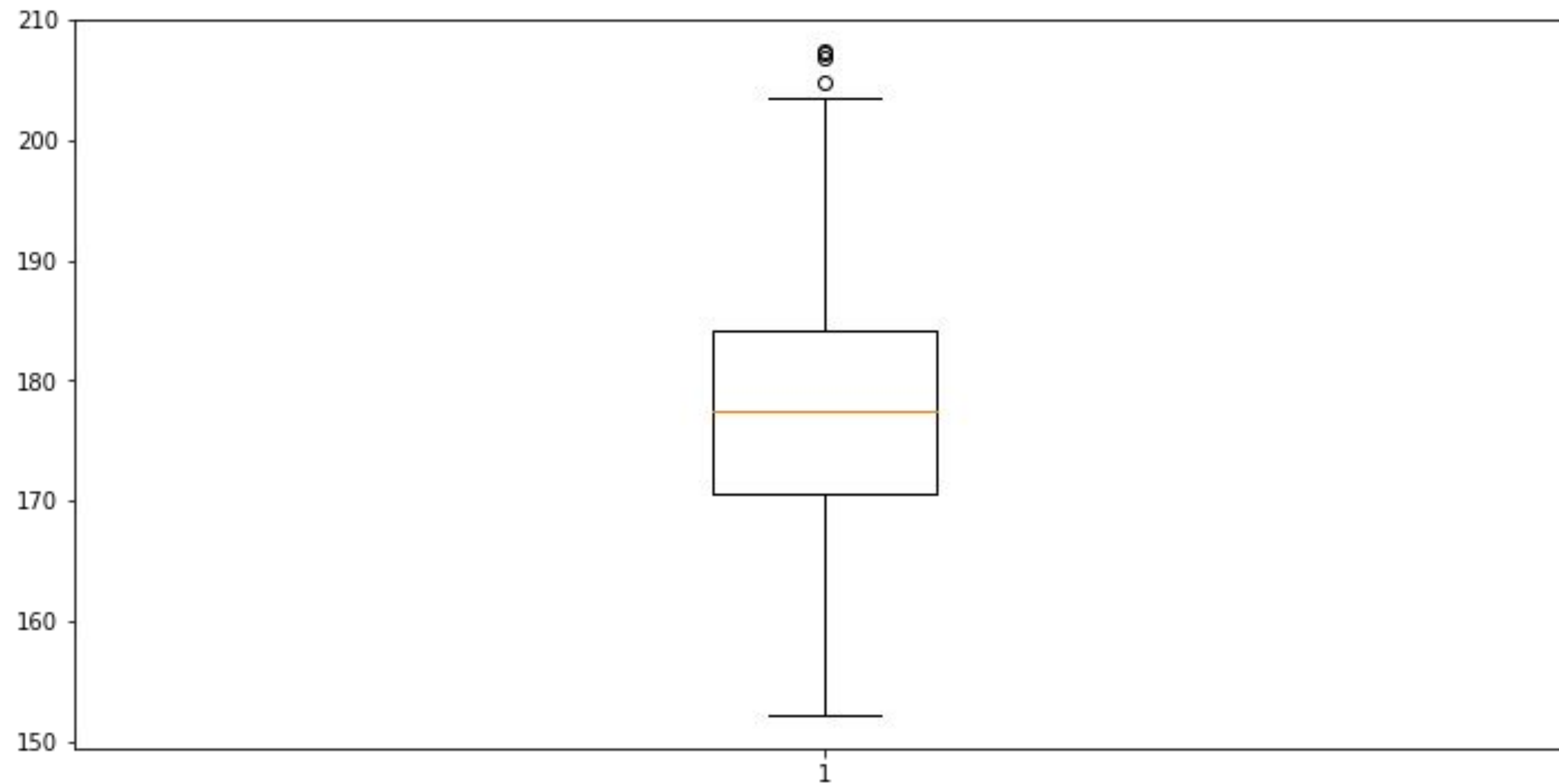
ЦЕЛИ ЗАНЯТИЯ



Простой график:

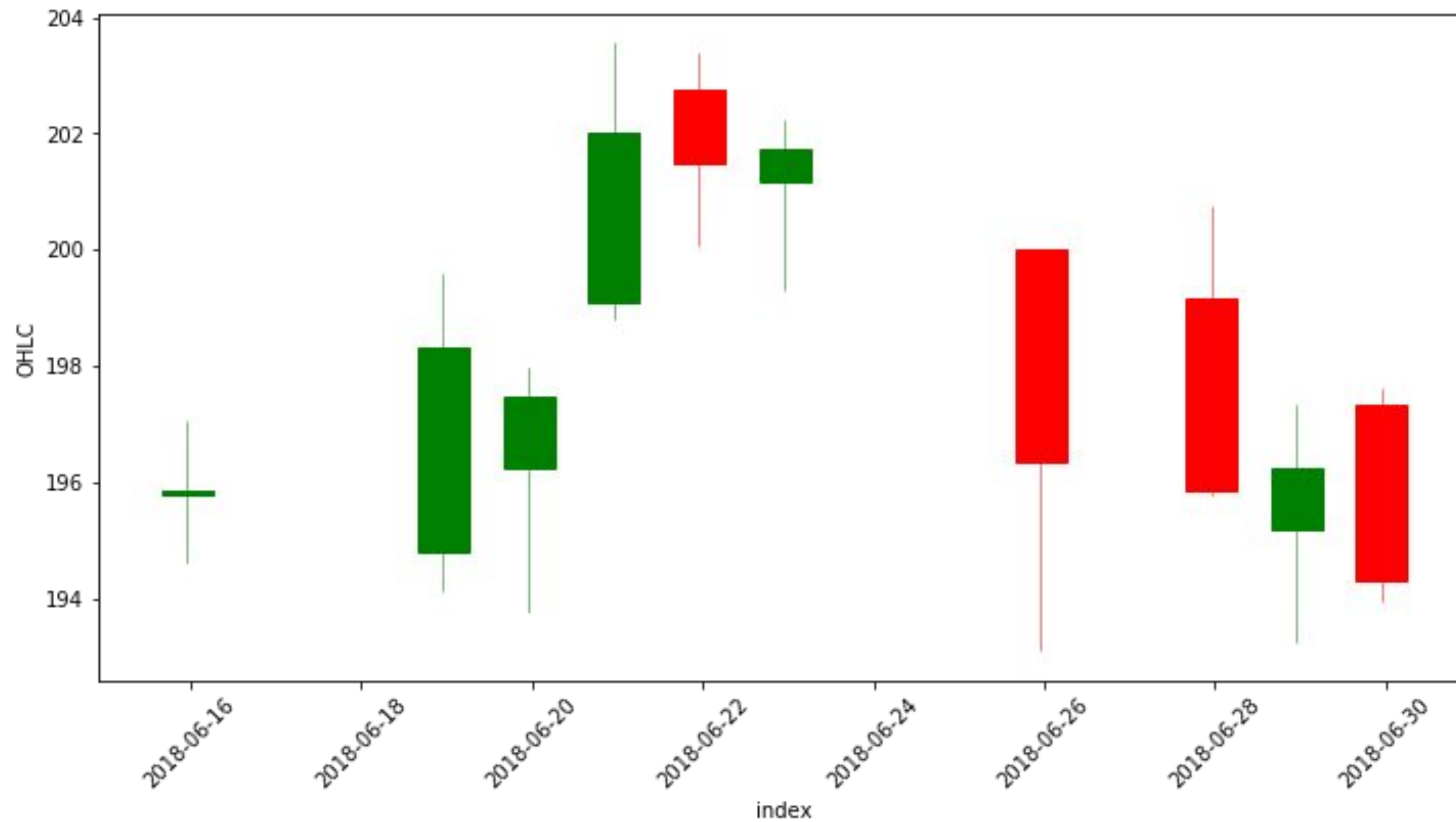


Box plot:



ЦЕЛИ ЗАНЯТИЯ

Candle Plot



Вопросы к временным рядам:

- Анализ (где мы сейчас)?
- Прогноз (куда мы идем)?
- Поиск аномалий (где было что-то необычное)?

Составляющие временного ряда

Составляющие временного ряда

Тренд — плавное долгосрочное изменение уровня ряда

Составляющие временного ряда

Сезонность — циклические изменения уровня ряда с постоянным периодом.

Составляющие временного ряда

Цикл — изменения уровня ряда с переменным периодом (экономические циклы, периоды солнечной активности).

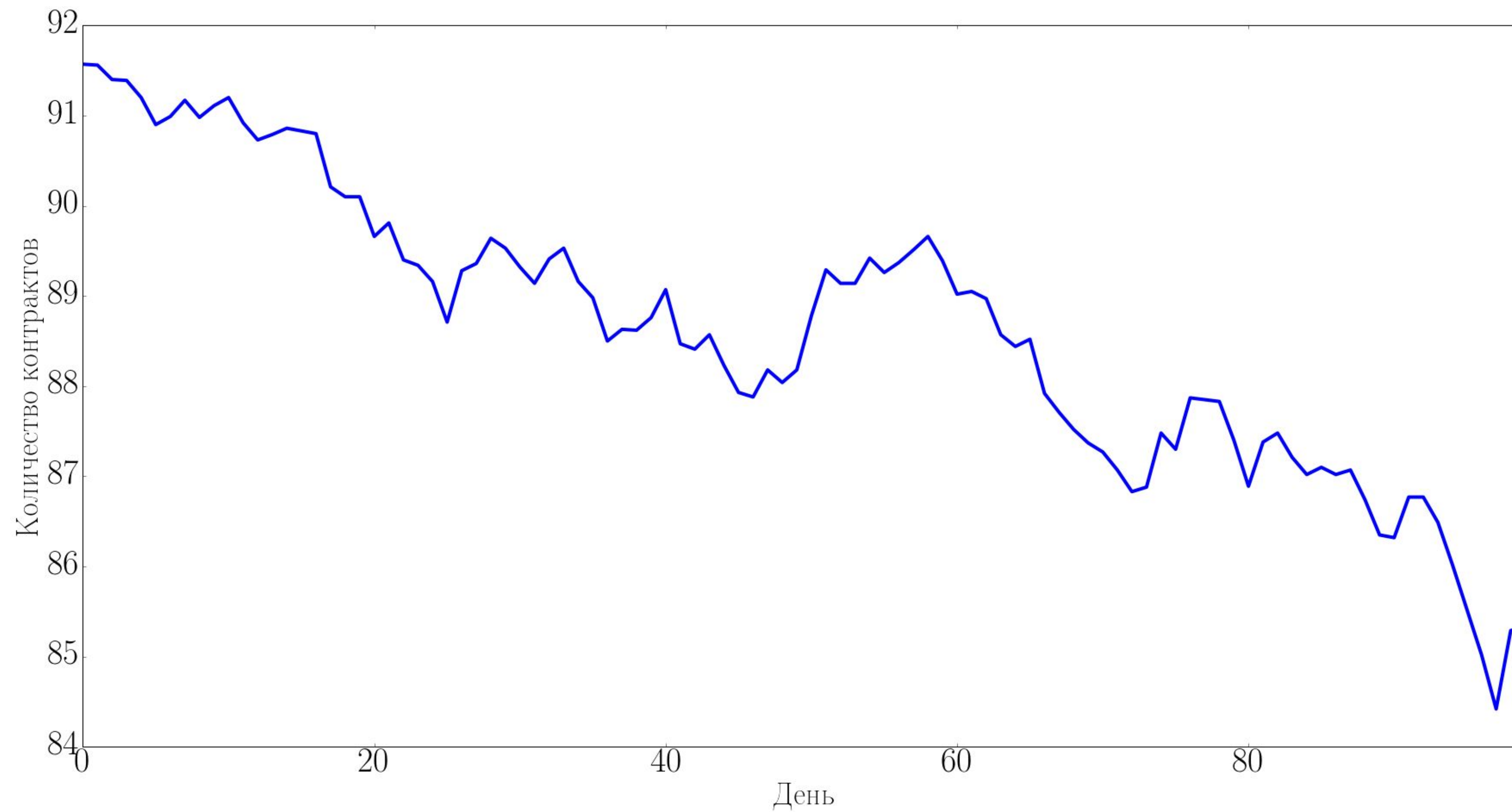
Составляющие временного ряда

Ошибка — непрогнозируемая случайная компонента ряда.

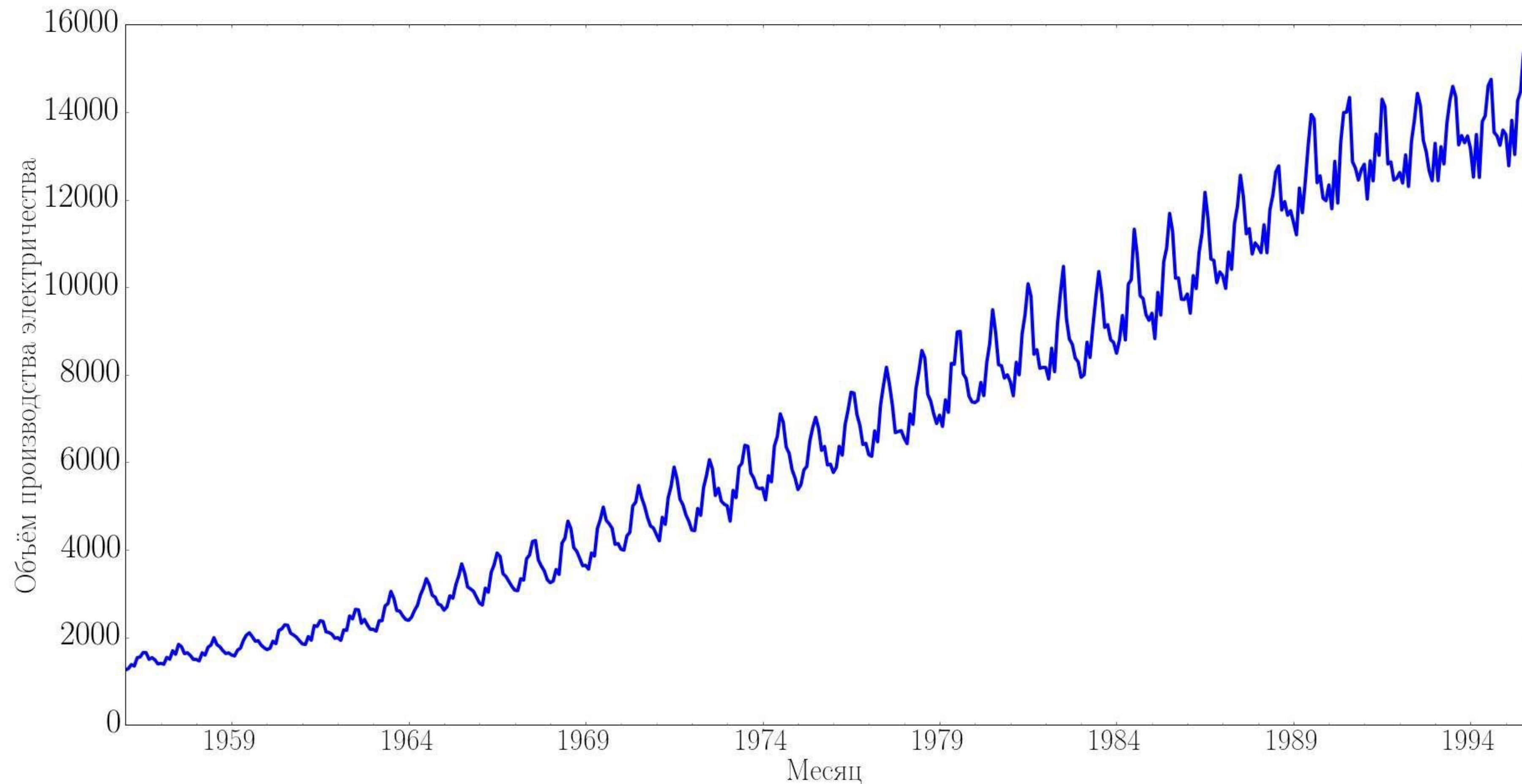
ЦЕЛИ ЗАНЯТИЯ



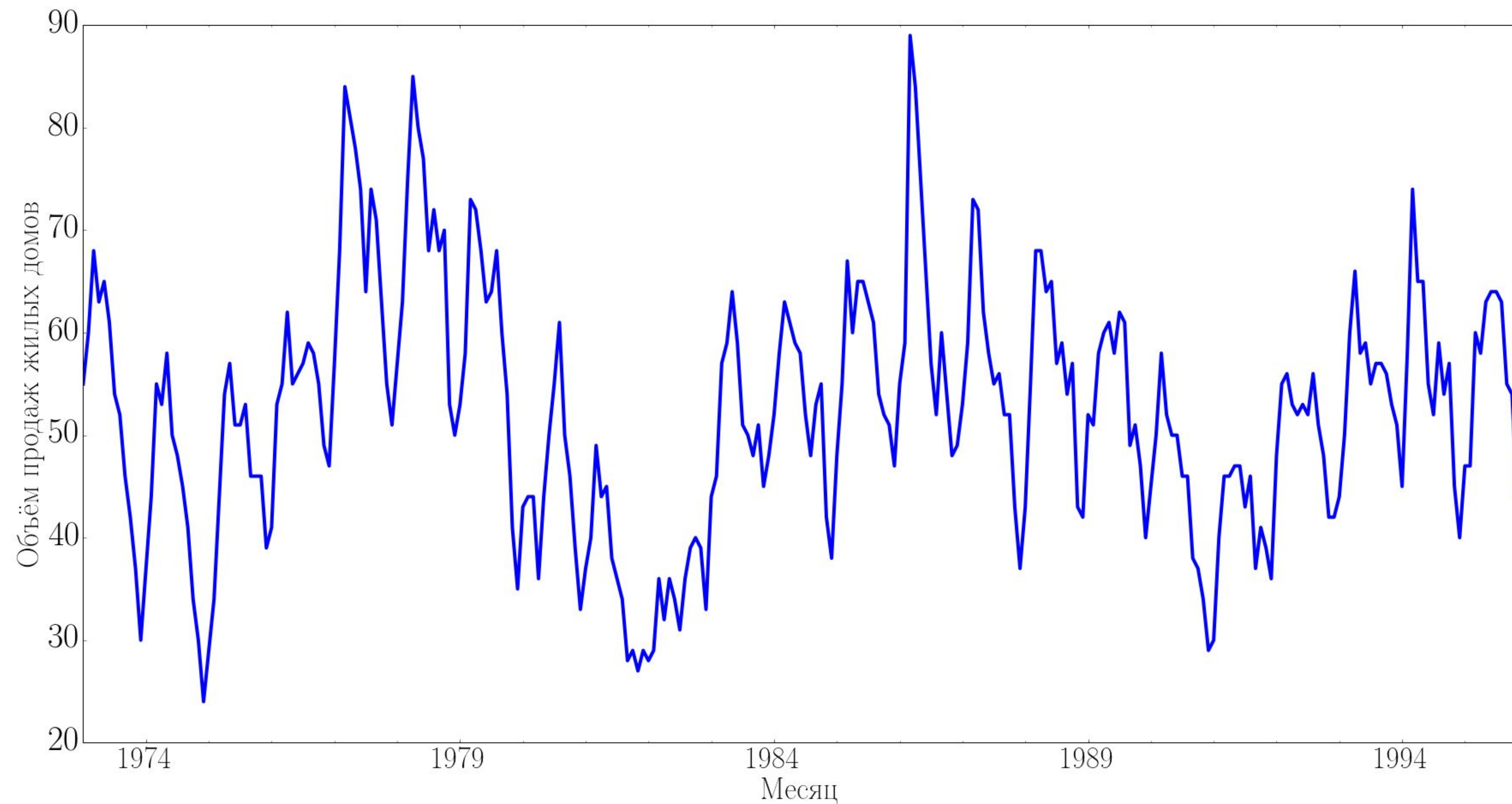
Пример (тренд)



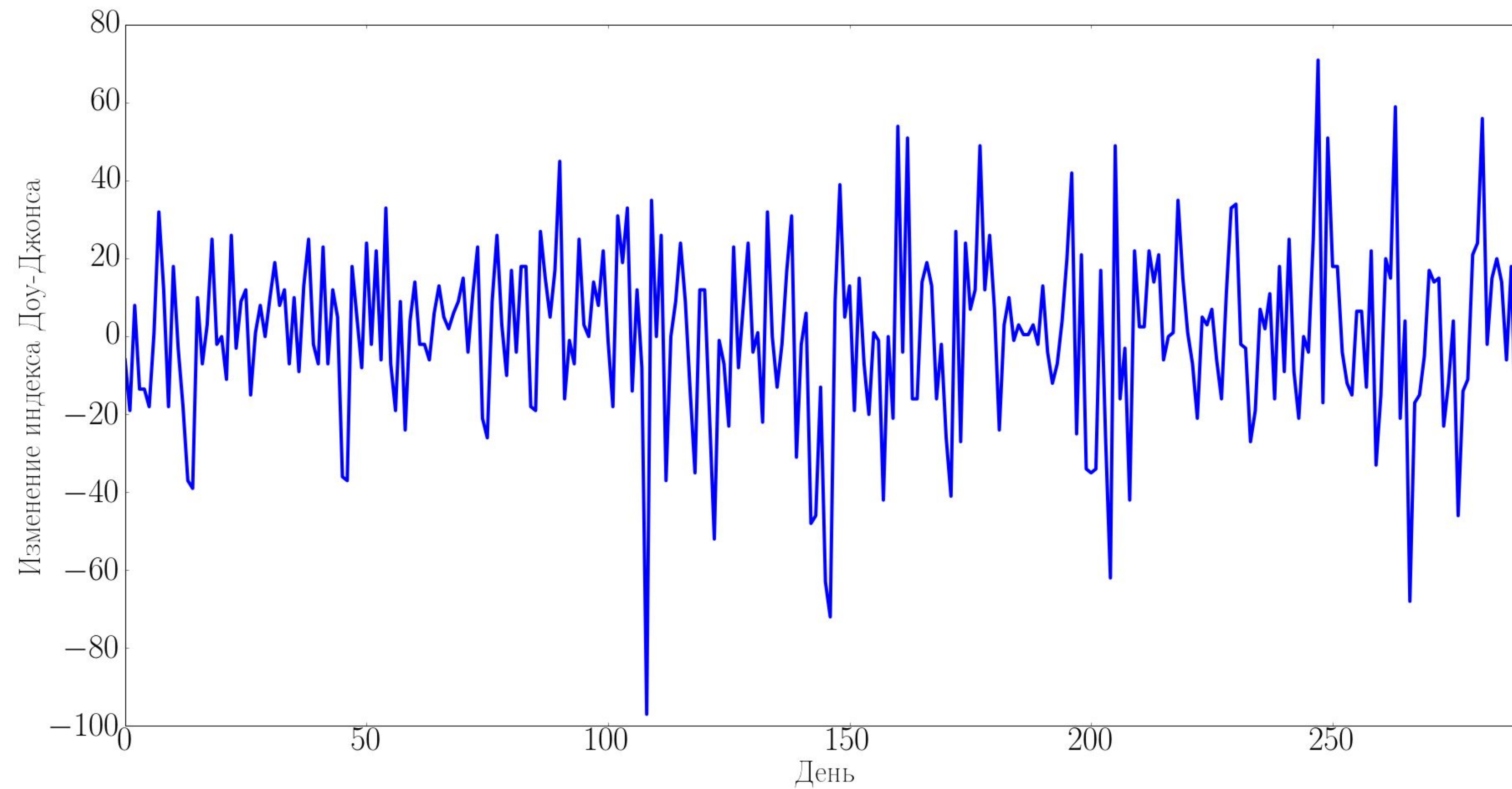
Пример (тренд, сезонность)



Пример (сезонность, циклы)



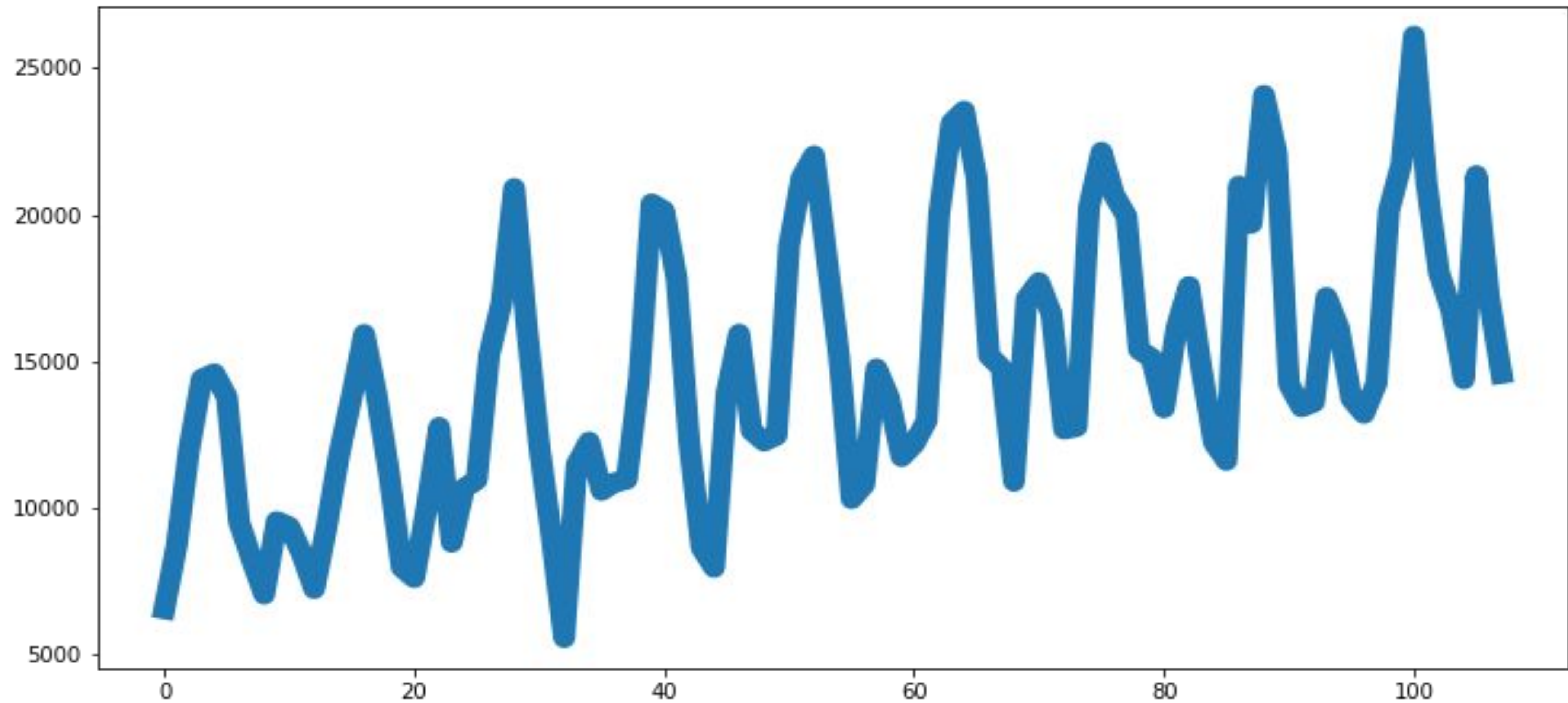
Пример (сезонность, циклы)



Анализ продаж машин в Квебеке

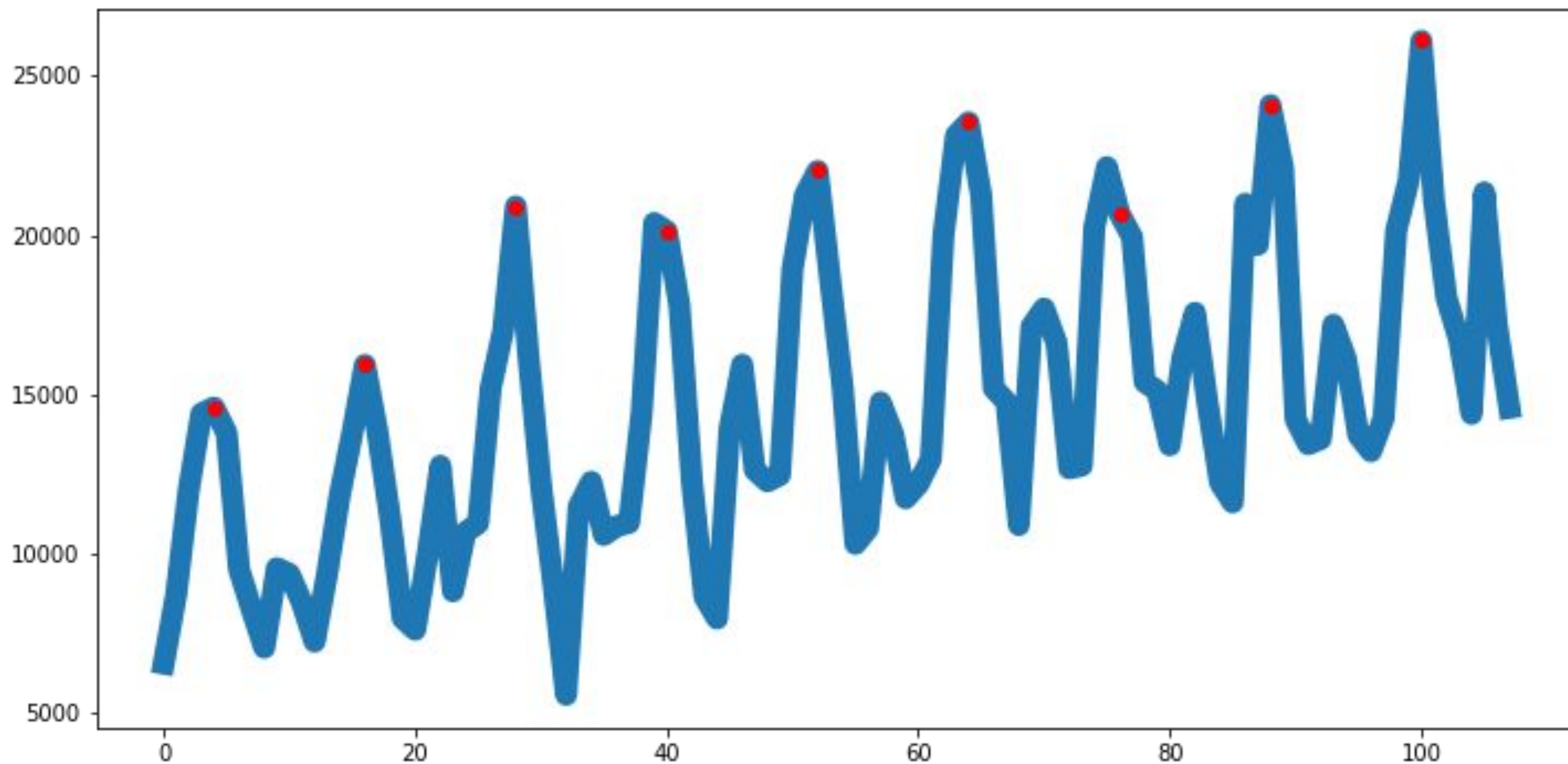
ЦЕЛИ ЗАНЯТИЯ

Продажи



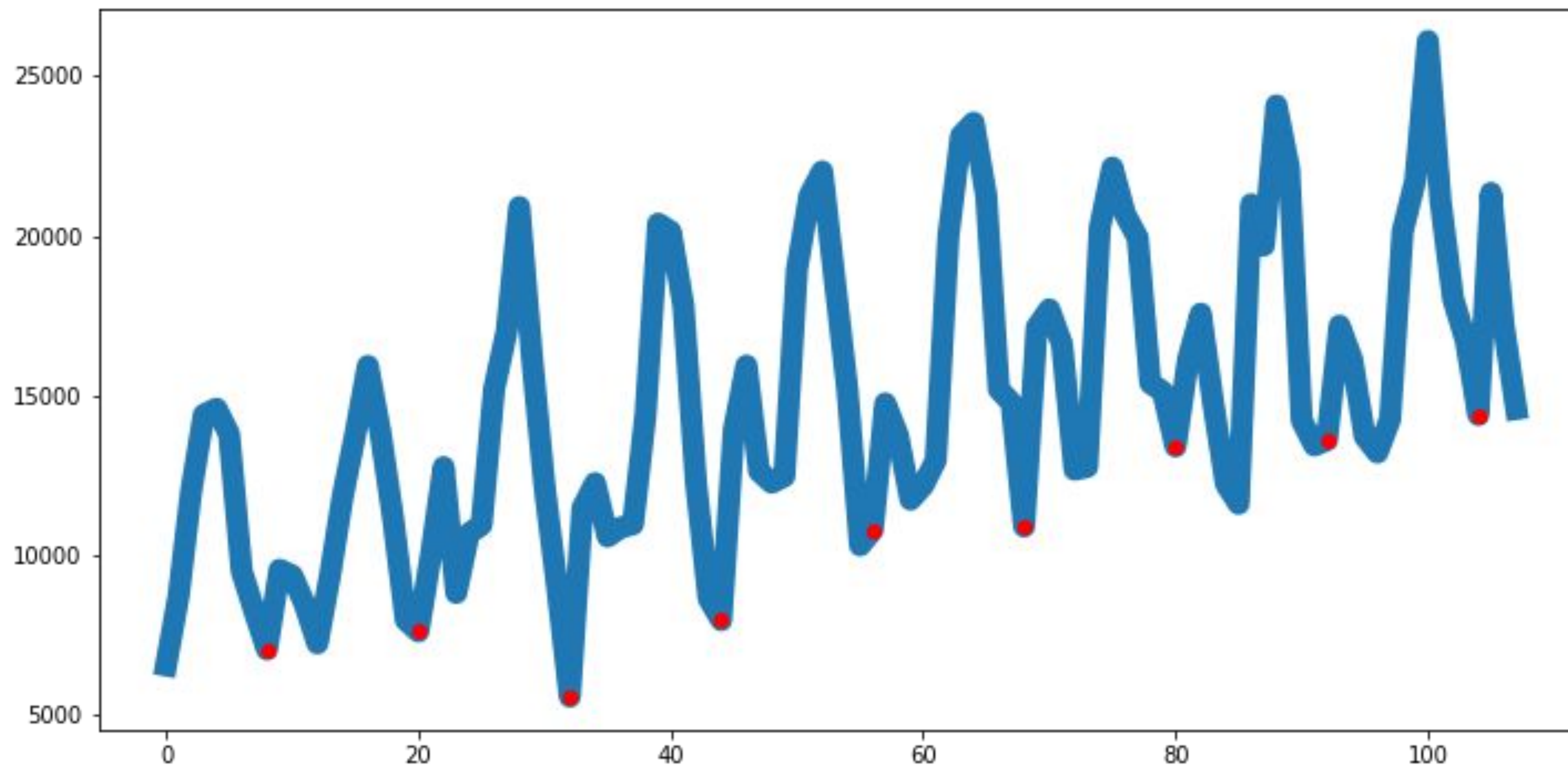
ЦЕЛИ ЗАНЯТИЯ

Каждый июнь продажи растут



ЦЕЛИ ЗАНЯТИЯ

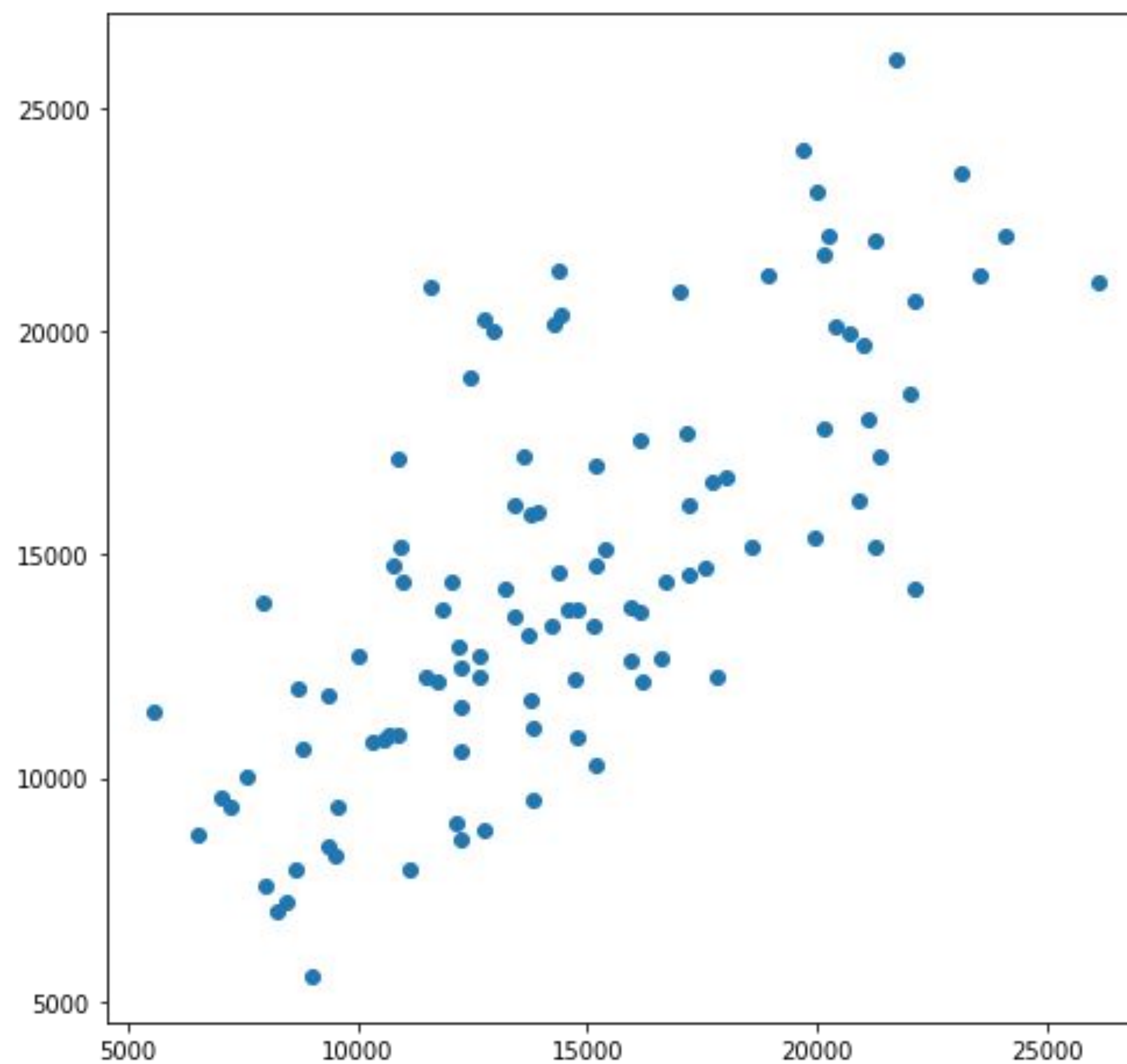
Каждый сентябрь продажи падают



**Посмотрим как связана
продажа в текущем
месяце и в следующем**

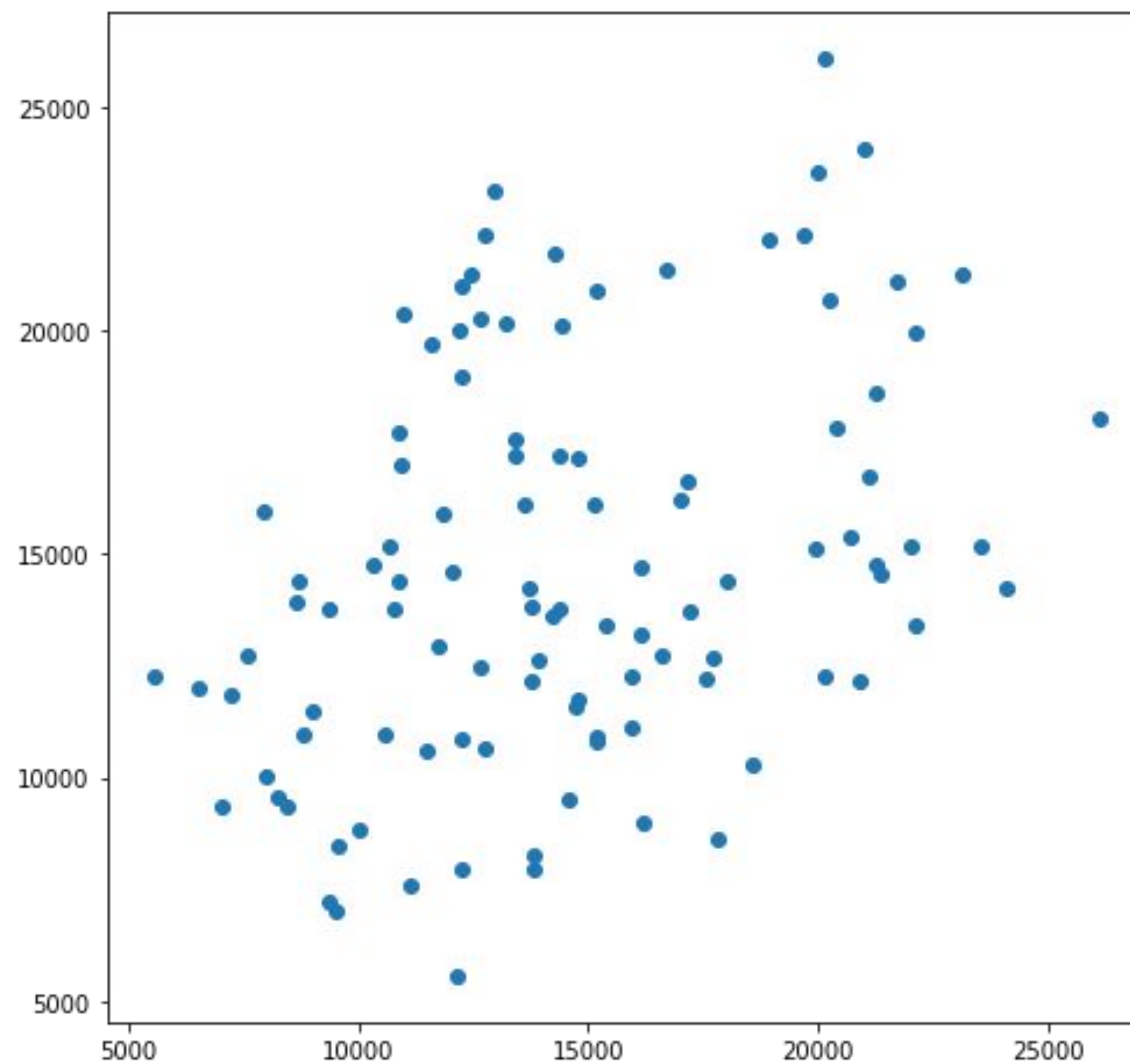
ЦЕЛИ ЗАНЯТИЯ

Текущий к следующему



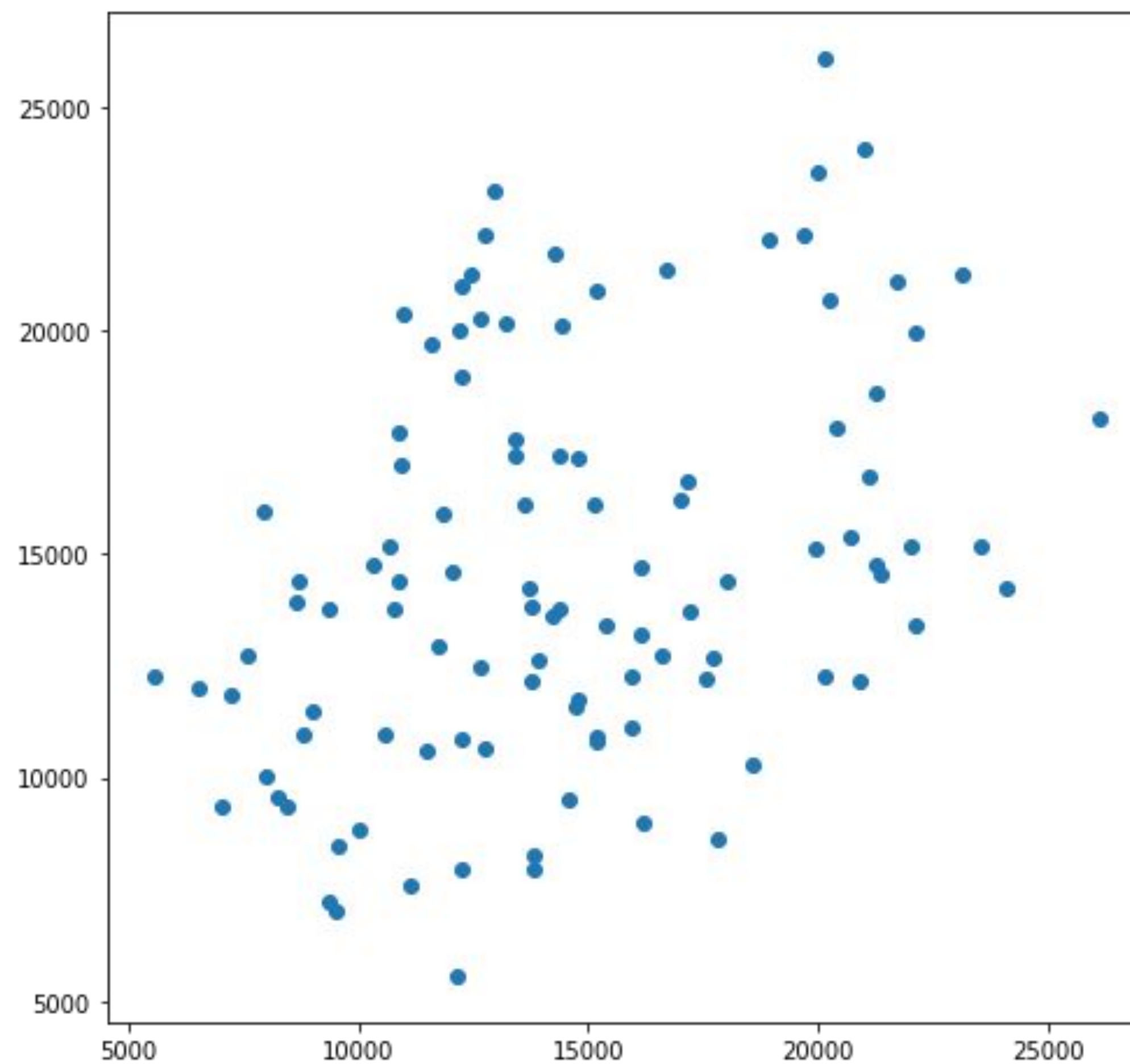
ЦЕЛИ ЗАНЯТИЯ

Текущий к $t + 2$



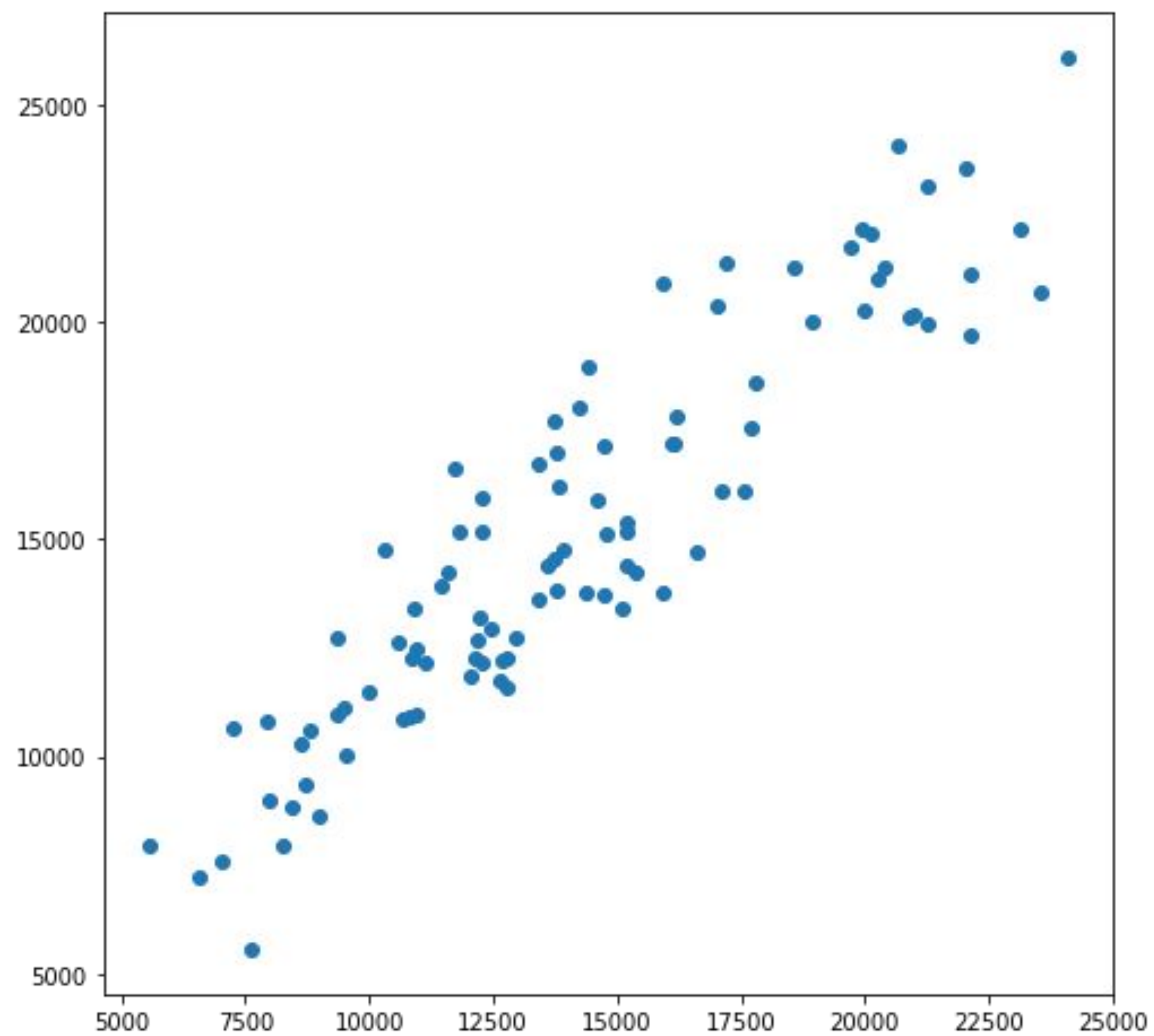
ЦЕЛИ ЗАНЯТИЯ

Текущий к $t + 2$



ЦЕЛИ ЗАНЯТИЯ

Текущий к $t + 12$



Автокорреляция

Автокорреляция:

Статистическая взаимосвязь (коэффициент корреляции) между последовательностями величин одного ряда, взятыми со сдвигом

Автокорреляция:

$$R(\tau) = \frac{E((y_t - Ey)(y_{t+\tau} - Ey))}{Dy}$$

$R(\tau) \in [-1, 1]$, $\tau \in \mathbb{N}$ — лаг автокорреляции.

Автокорреляция для выборки:

$$R(\tau) = \frac{\sum_{t=1}^{T-\tau} (y_t - \bar{y})(y_{t+\tau} - \bar{y})}{\sum_{t=1}^T (y_t - \bar{y})^2}$$

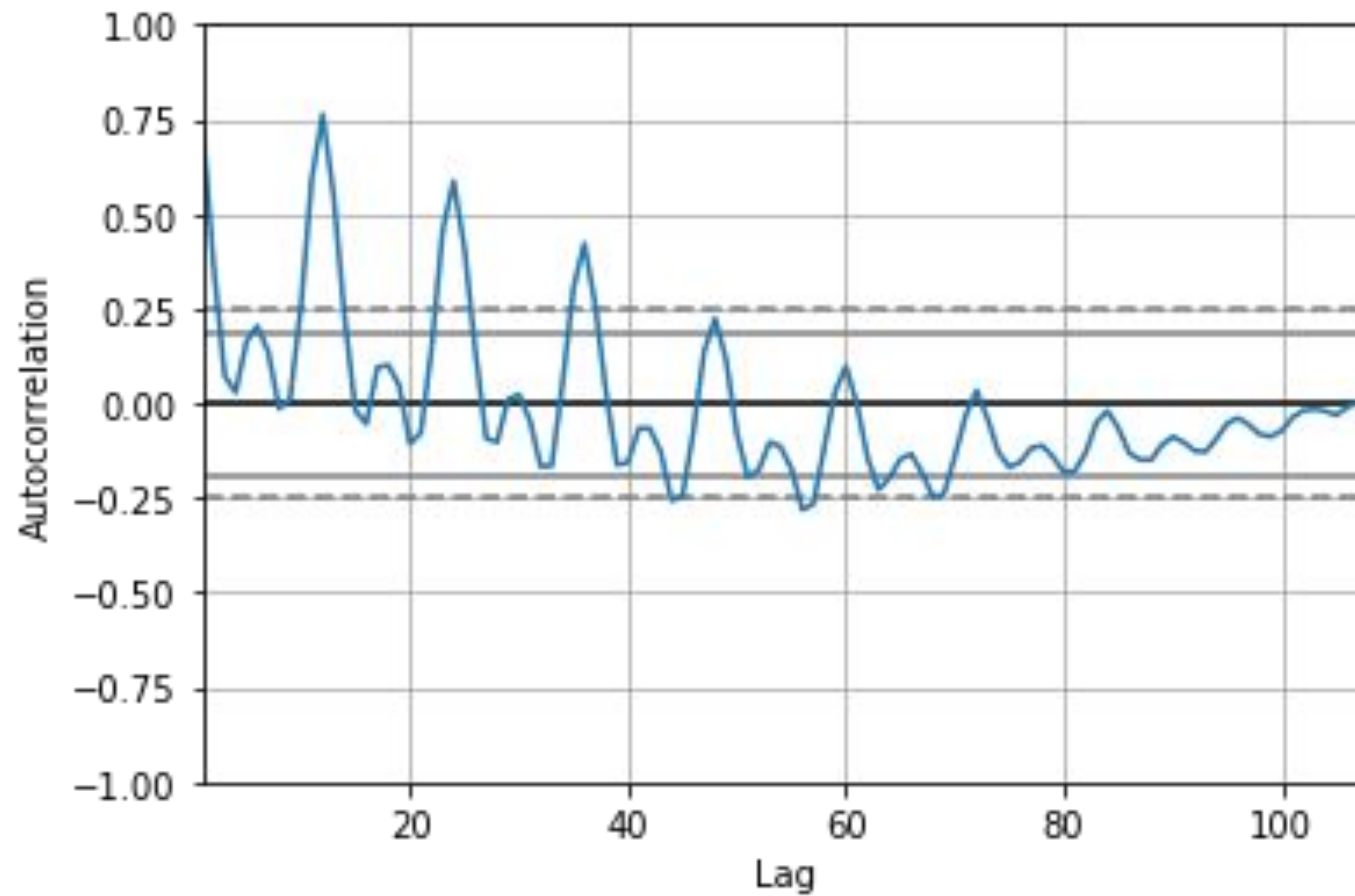
Для нашего ряда:

τ	R
1	0.728
2	0.367
8	-0.011
12	0.919

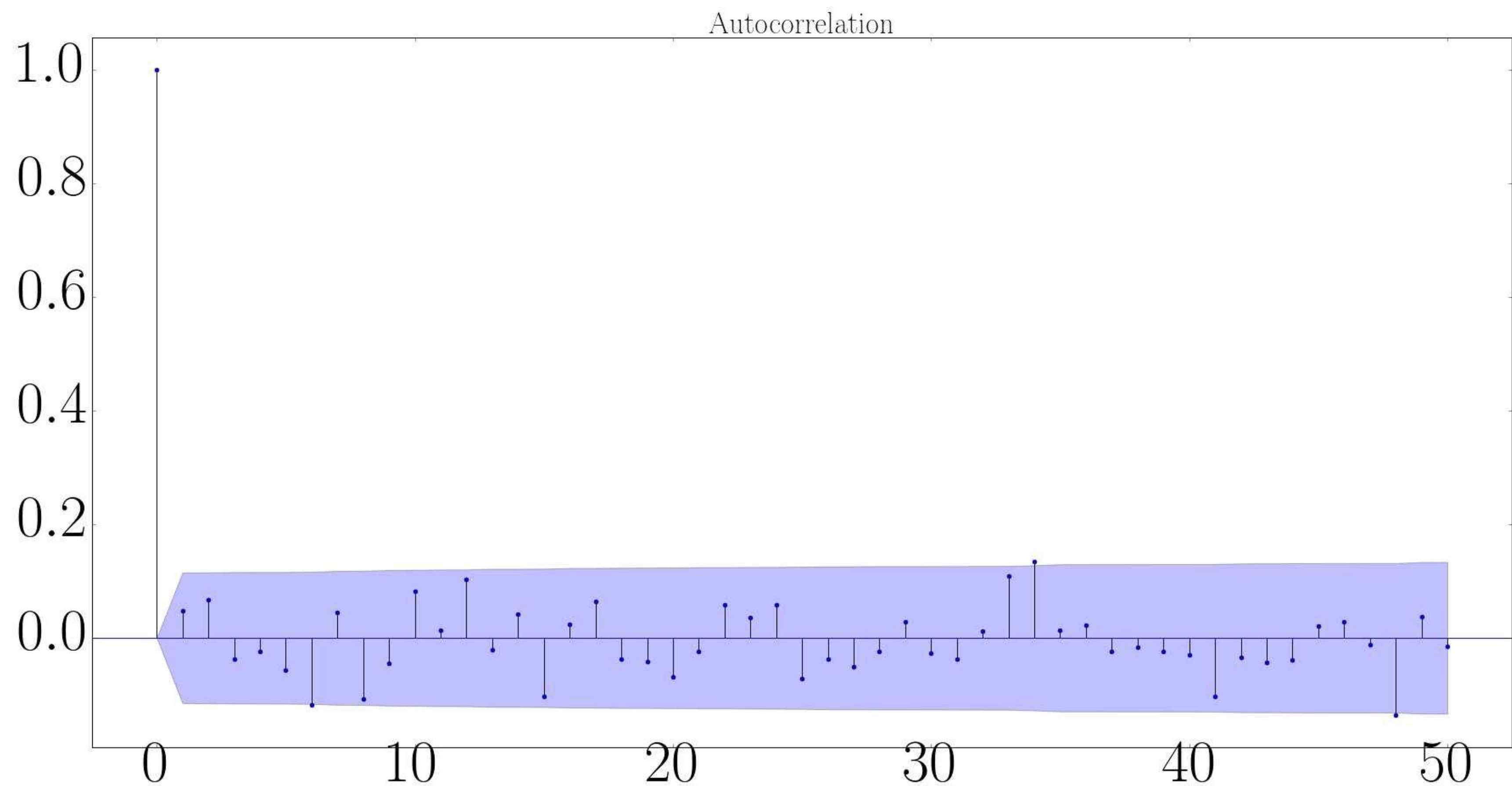
Коррелограмма:

График автокорреляций для разных лагов

Коррелограмма:



Коррелограмма:



Автокорреляция:

Значение 0.9 – хорошее? А 0.1? А 0.4?

Как понять какой коэффициент значимый?

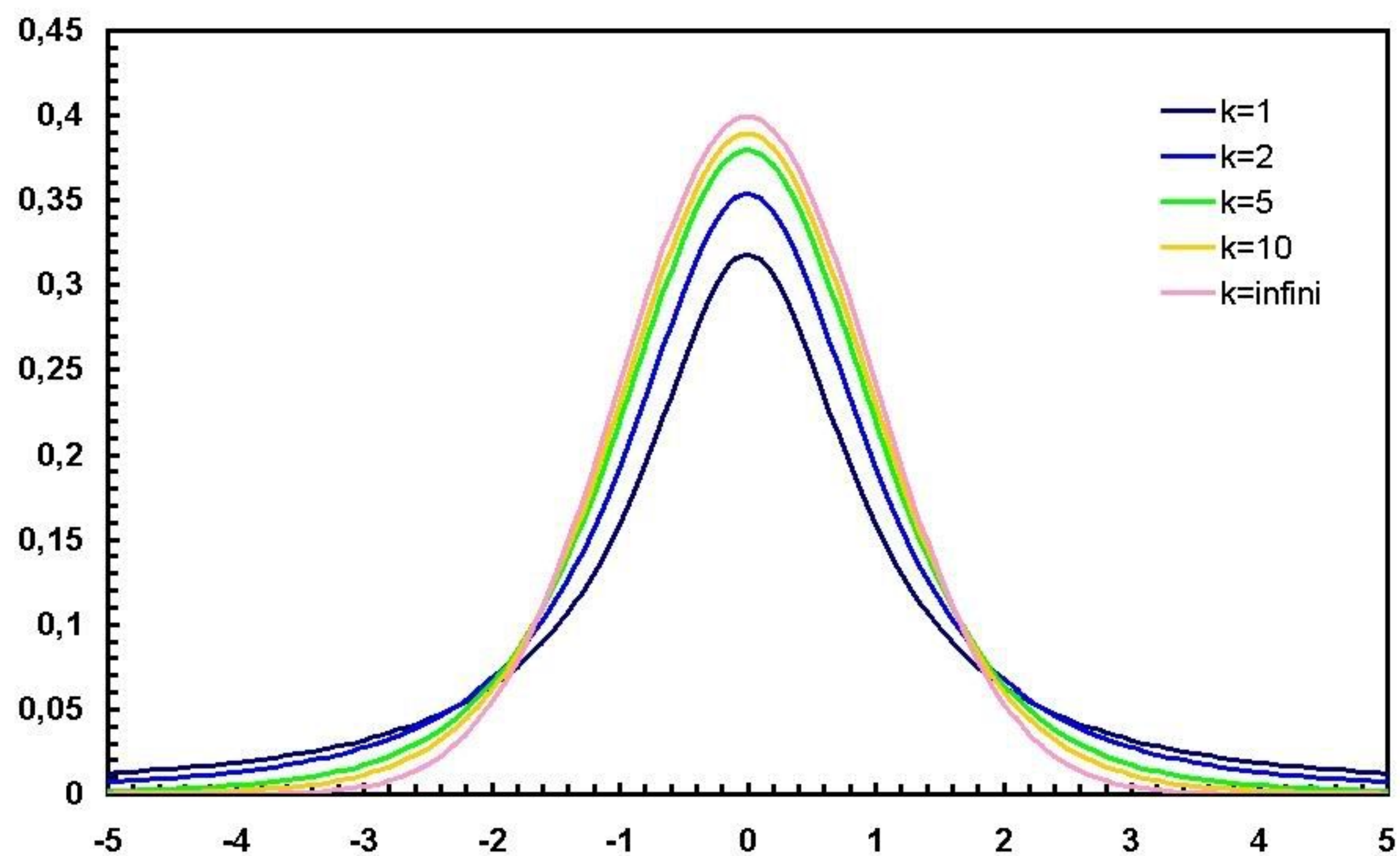
Автокорреляция:

Временной ряд: $y(t)$

Нулевая гипотеза: $H_0: r_\tau = 0$

Альтернативная гипотеза: $H_0: r_\tau \neq 0$

$St(k)$:



Стационарность

Строгая стационарность:

Ряд $y(t)$ называется строго стационарным если для любого s совместное распределение вероятностей $y(t), y(t+1), \dots, y(t+k)$ совпадает с совместным распределением вероятностей $y(t+s), y(t+s+1), \dots, y(t+s+k)$

Слабая стационарность:

Ряд $y(t)$ называется слабо стационарным если $E(y(t)) = \text{const}$ и $\text{cov}(y(t), y(t+s))$ зависит только от s

Стационарность:

Всякий строго стационарный ряд так же и слабо стационарен

Стационарность:

1. Стационарный ряд легче предсказывать. Можно предположить, что статистические свойства в будущем не изменятся
2. В большинство моделей рядов закладывается гипотеза о стационарности. Это означает, что получаемые оценки будут надежными только в случае стационарности ряда.

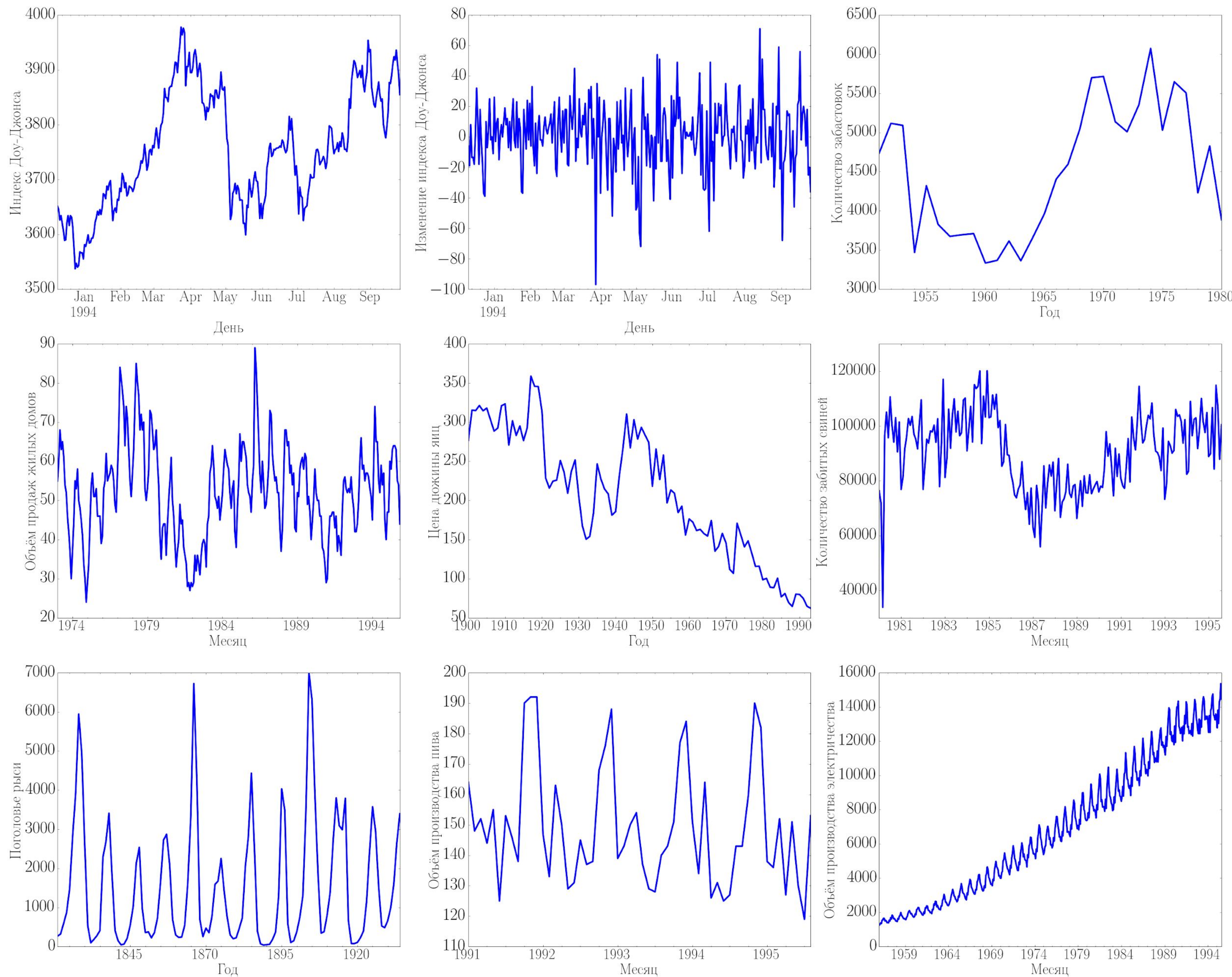
Стационарность:

Тренд \Rightarrow нестационарность

Сезонность \Rightarrow нестационарность

Цикл \Rightarrow ?

ЦЕЛИ ЗАНЯТИЯ



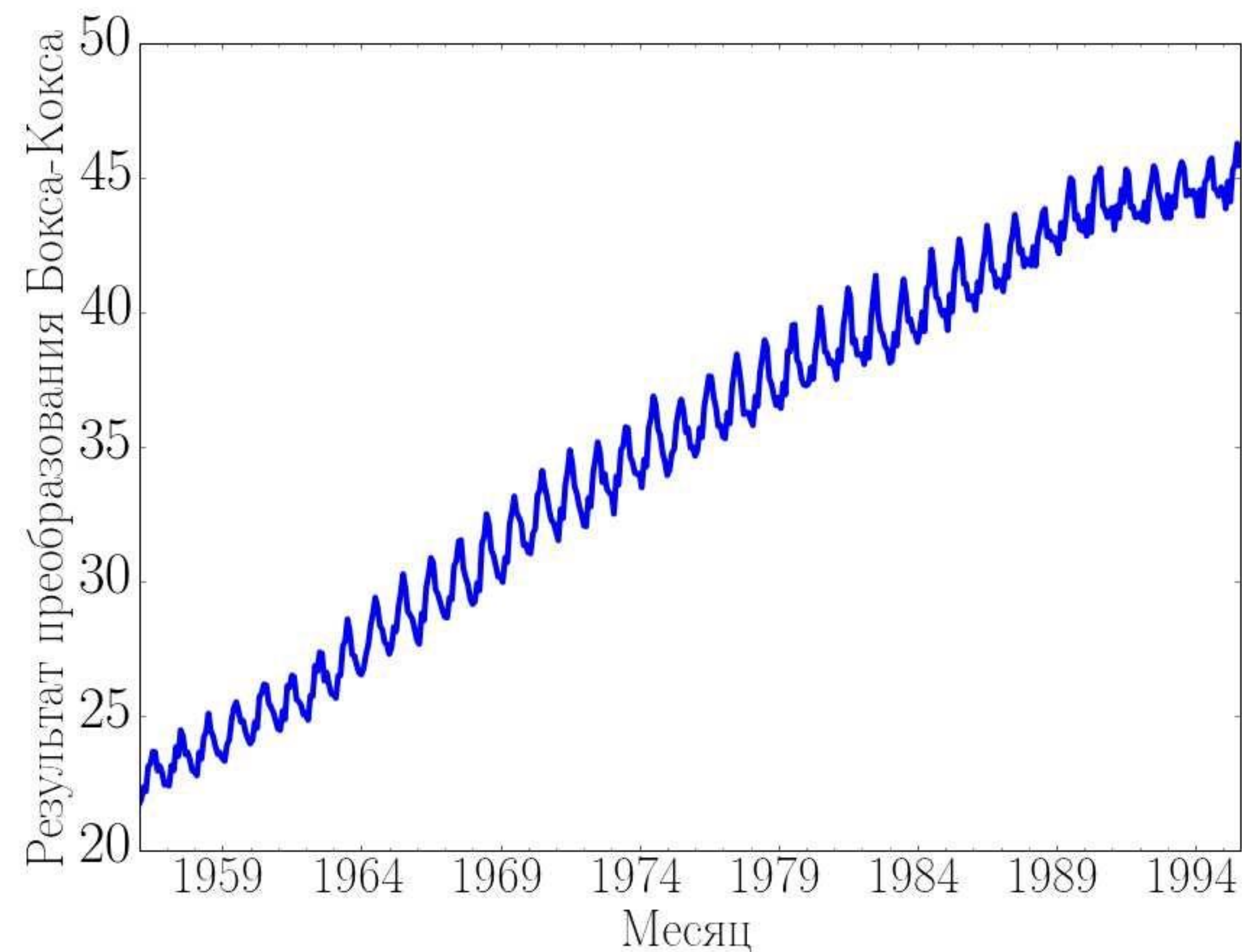
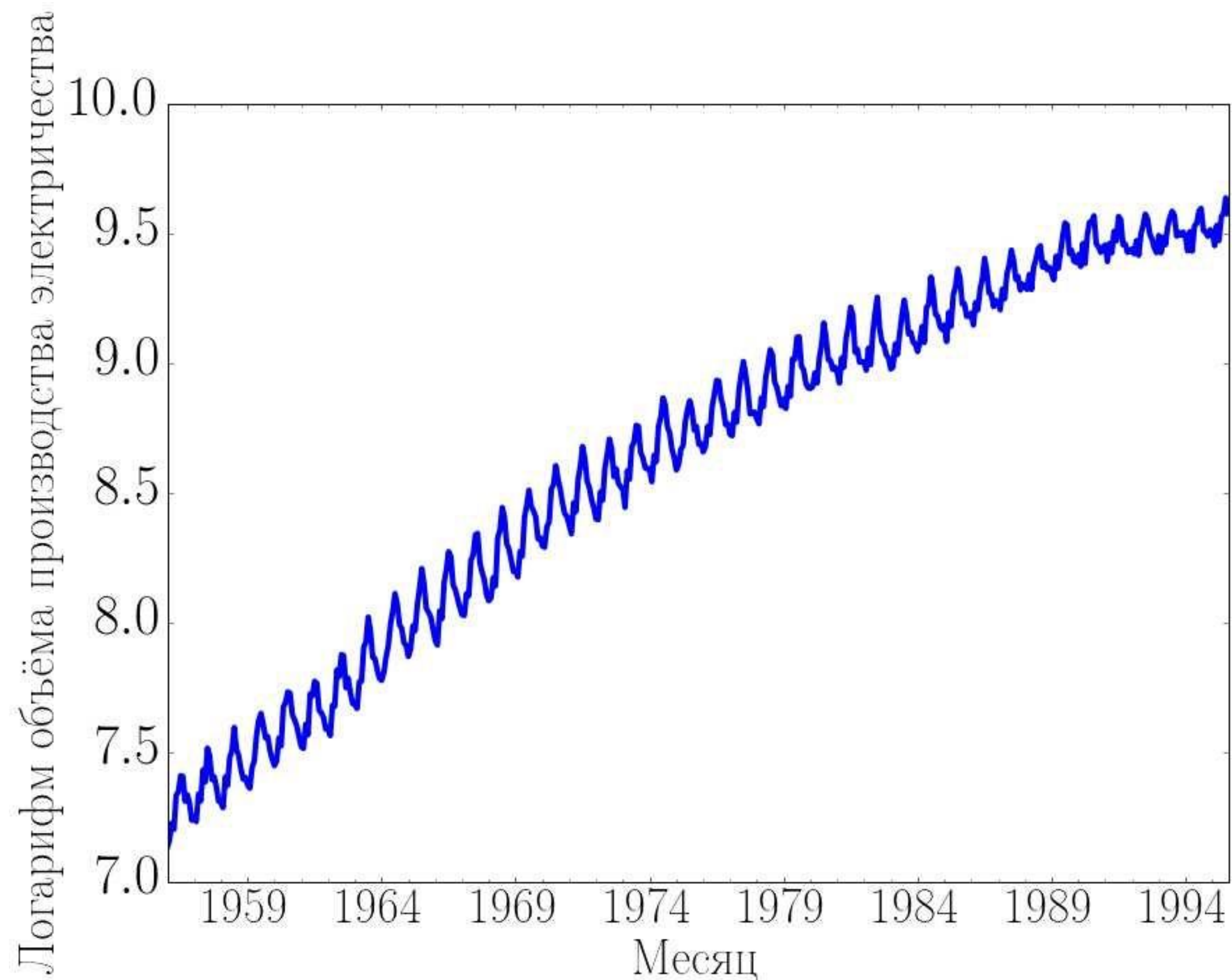
Как сделать ряд стационарным

Как сделать стационарный ряд?

Для рядов с монотонно меняющейся дисперсией – логарифмирование

ЦЕЛИ ЗАНЯТИЯ

Как сделать стационарный ряд?



Как сделать стационарный ряд?

Для прогноза нужно применять обратное преобразование

Лямбда - не очень строгий параметр, его можно спокойно округлять

Как сделать стационарный ряд?

Дифференцирование – переход к попарным разностям. Позволяет избавиться от тренда

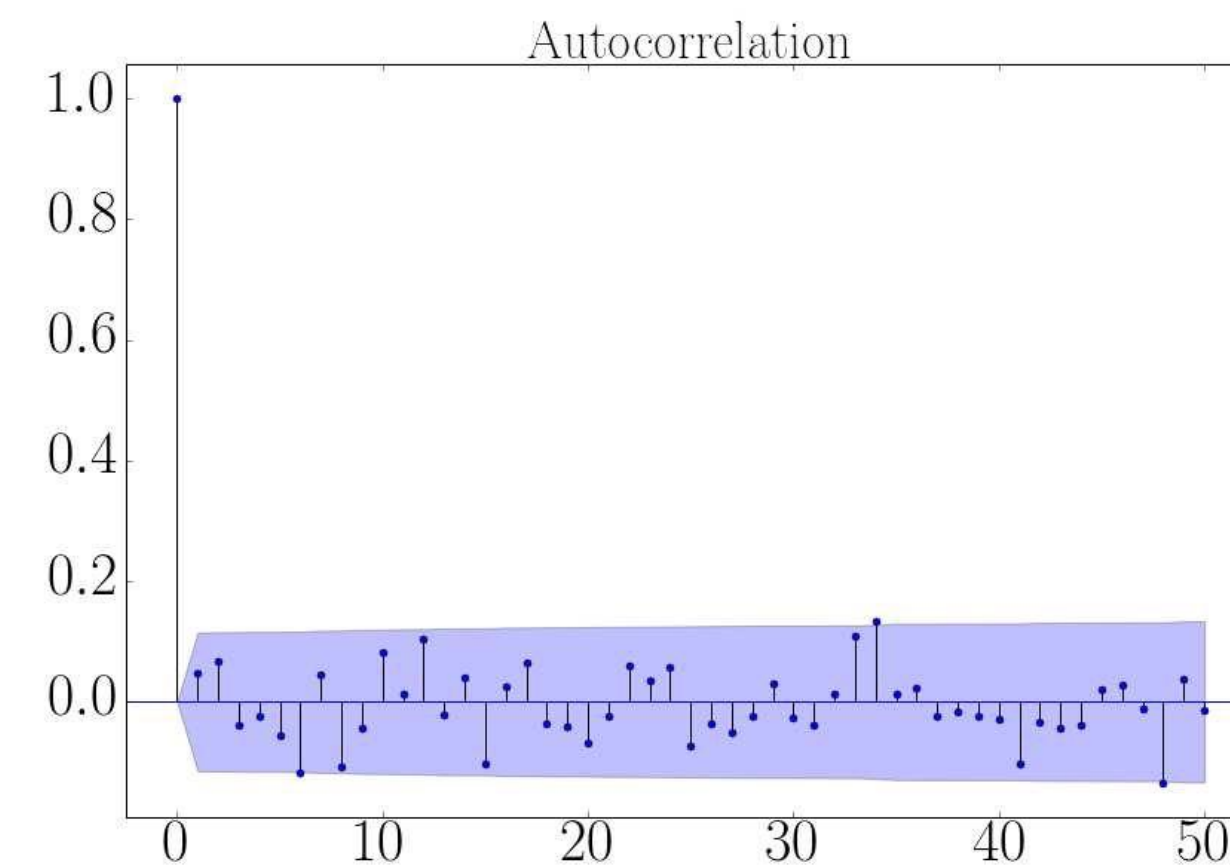
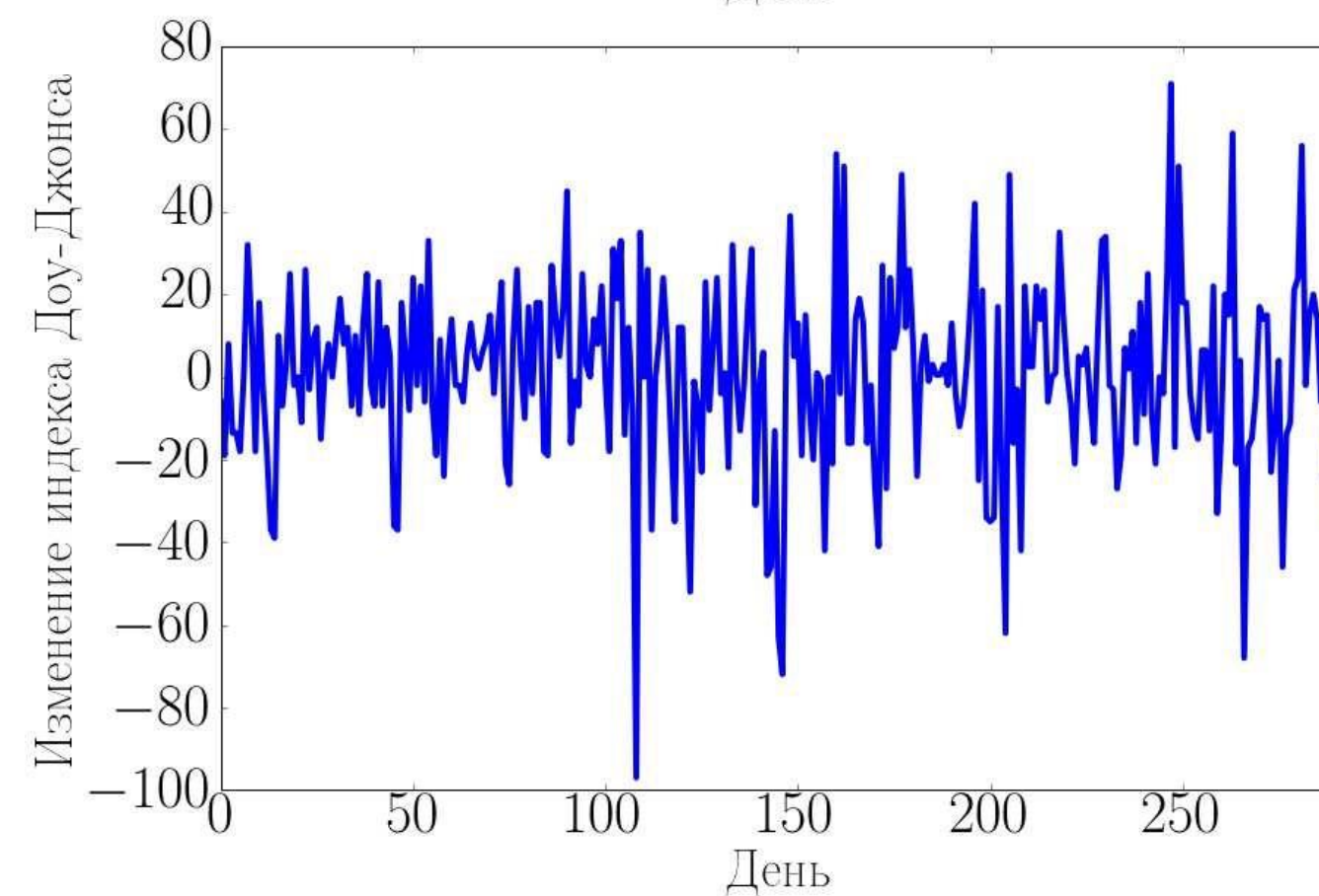
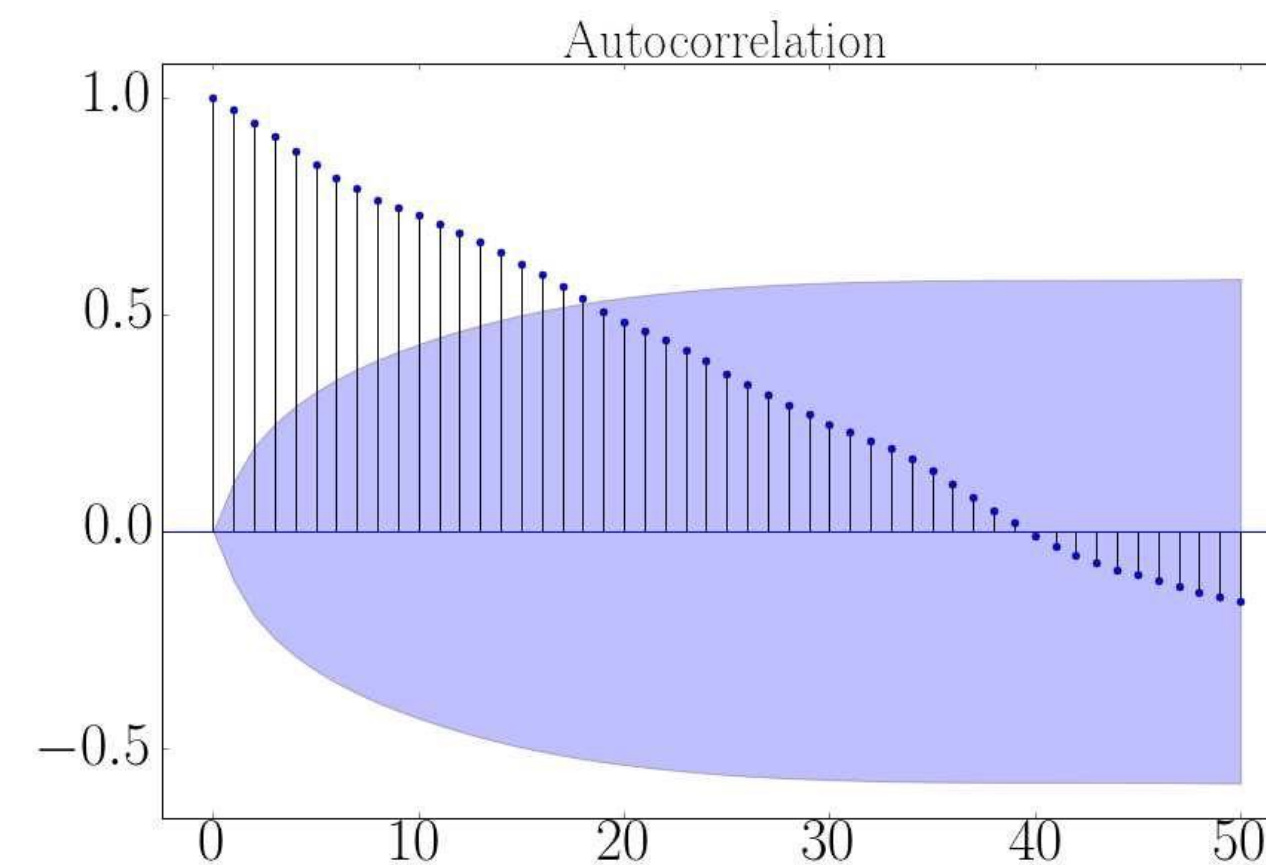
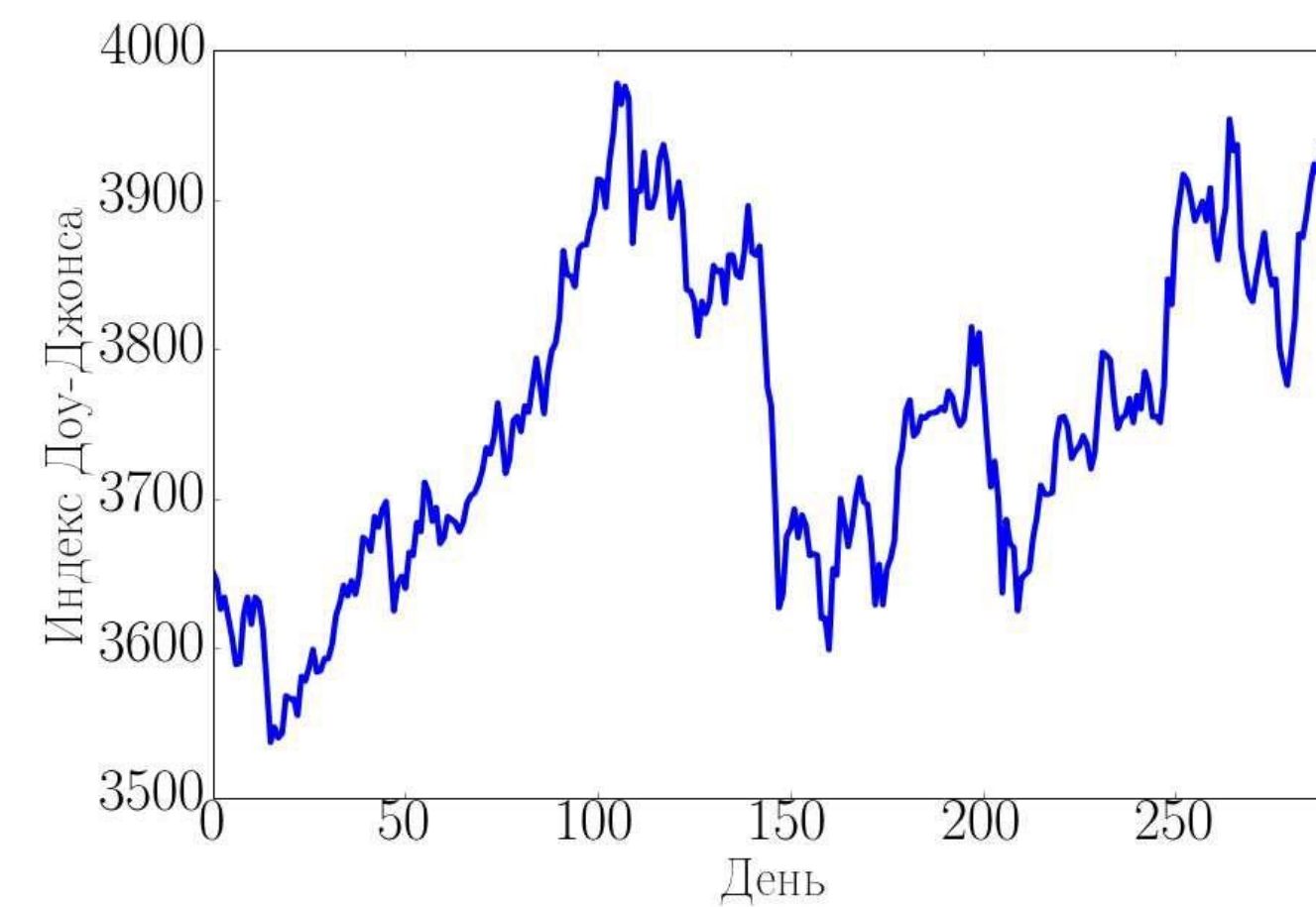
Как сделать стационарный ряд?

Сезонное дифференцирование – вычитаем не предыдущее значение, а сдвинутое на сезон

Как сделать стационарный ряд?

Дифференцировать можно несколько раз подряд

Как сделать стационарный ряд?



Модель авторегрессии

Модель авторегрессии:

Цель – пытаемся предсказать ряд по его же наблюдениям в прошлом

Модель авторегрессии:

$$y_t = a + \sum_{i=1}^p b_i y_{t-i} + \varepsilon_t$$

a — свободный член

Модель авторегрессии:

$$AR(1): y_t = a + by_{t-1} + \varepsilon_t$$

ТЕСТ ДИКИ-ФУЛЛЕРА

Тест Дики Фуллера:

Проверяем гипотезу не стационарности
против стационарности

Техническая идея теста ДФ:

$$y_t = a + by_{t-1} + \varepsilon_t$$

$$y_t - y_{t-1} = a + by_{t-1} - y_{t-1} + \varepsilon_t$$

$$\Delta y_t = a + (b - 1)y_{t-1} + \varepsilon_t$$

Единичный корень:

Говорят, что $y(t)$ имеет единичный корень, если ряд $y(t) - y(t-1)$ стационарен

Если $y(t)$ имеет единичный корень, то ряд не стационарен

Виды ДФ:

Без константы: $\Delta y_t = (b - 1)y_{t-1} + \varepsilon_t$

С константой: $\Delta y_t = a + (b - 1)y_{t-1} + \varepsilon_t$

С трендом: $\Delta y_t = a + (b - 1)y_{t-1} + ct + \varepsilon_t$

Расширенный тест ДФ:

Является модификацией теста Дики-Фуллера, необходимость введения связана с тем, что рассматриваемый процесс может являться авторегрессией не первого, а более высокого порядка

Расширенный тест ДФ:

В каждое уравнение теста вводятся авторегрессионные переменные (лаги) переменной разности для коррекции возможной коррелированности случайных отклонений тестируемой модели

Расширенный тест ДФ:

Без константы: $\Delta y_t = (b - 1)y_{t-1} + \sum_{i=1}^k d_i \Delta y_{t-i} + \varepsilon_t$

С константой: $\Delta y_t = a + (b - 1)y_{t-1} + \sum_{i=1}^k d_i \Delta y_{t-i} + \varepsilon_t$

С трендом: $\Delta y_t = a + (b - 1)y_{t-1} + ct + \sum_{i=1}^k d_i \Delta y_{t-i} + \varepsilon_t$

ВИЗУАЛЬНАЯ ПРОВЕРКА

График автокорреляций (ACF):

Стационарный ряд: Независимо от значений в первом лаге, быстро убывает после несколько первых значений.

Нестационарный ряд: Значения в первом лаге, т.е. $ACF(1)$, близки к единице, а затем медленно убывает по угасающей экспоненте(синусоиде)

Частичная автокорреляция(PACF):

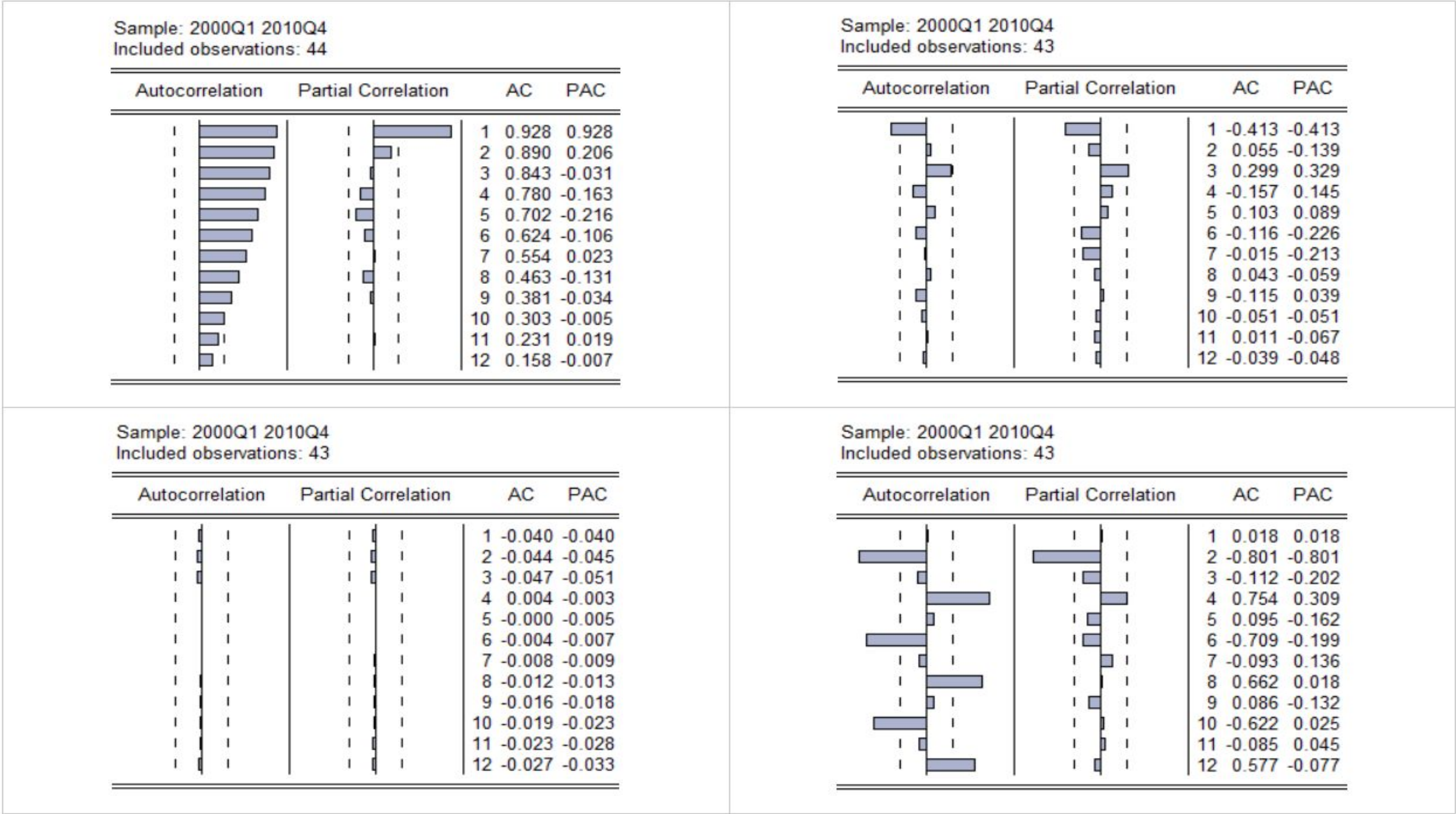
определяется как значение коэффициента корреляции между случайными величинами X_t и X_{t+k} , очищенными от влияния “промежуточных” случайных величин $X_{t+1}, \dots, X_{t+k-1}$.

Частичная автокорреляция(PACF):

Стационарный ряд: Также быстро убывает после несколько первых значений

Нестационарный ряд: Значение в первом лаге $ACF(1)=PACF(1)$ близко к единице, однако остальные значения коэффициентов корреляции статистически незначимы, т.е. значения функций не выходят за пределы доверительного интервала.

Визуальная проверка:



**Для проверки на
стационарность
рекомендуется
комбинировать ДФ и
визуальные методы**

ПОЛЕЗНЫЕ МАТЕРИАЛЫ

1. http://economy.bsu.by/wp-content/uploads/2014/10/econometr_zao_1_2.pdf
2. <https://www.youtube.com/watch?v=u433nrxdf5k&frags=pl%2Cwn>
3. <http://people.duke.edu/~rnau/whatuse.htm>
4. <https://machinelearningmastery.com/decompose-time-series-data-trend-seasonality/>



НЕТОЛОГИЯ
групп

Спасибо за внимание!

Сапрыкин Артур