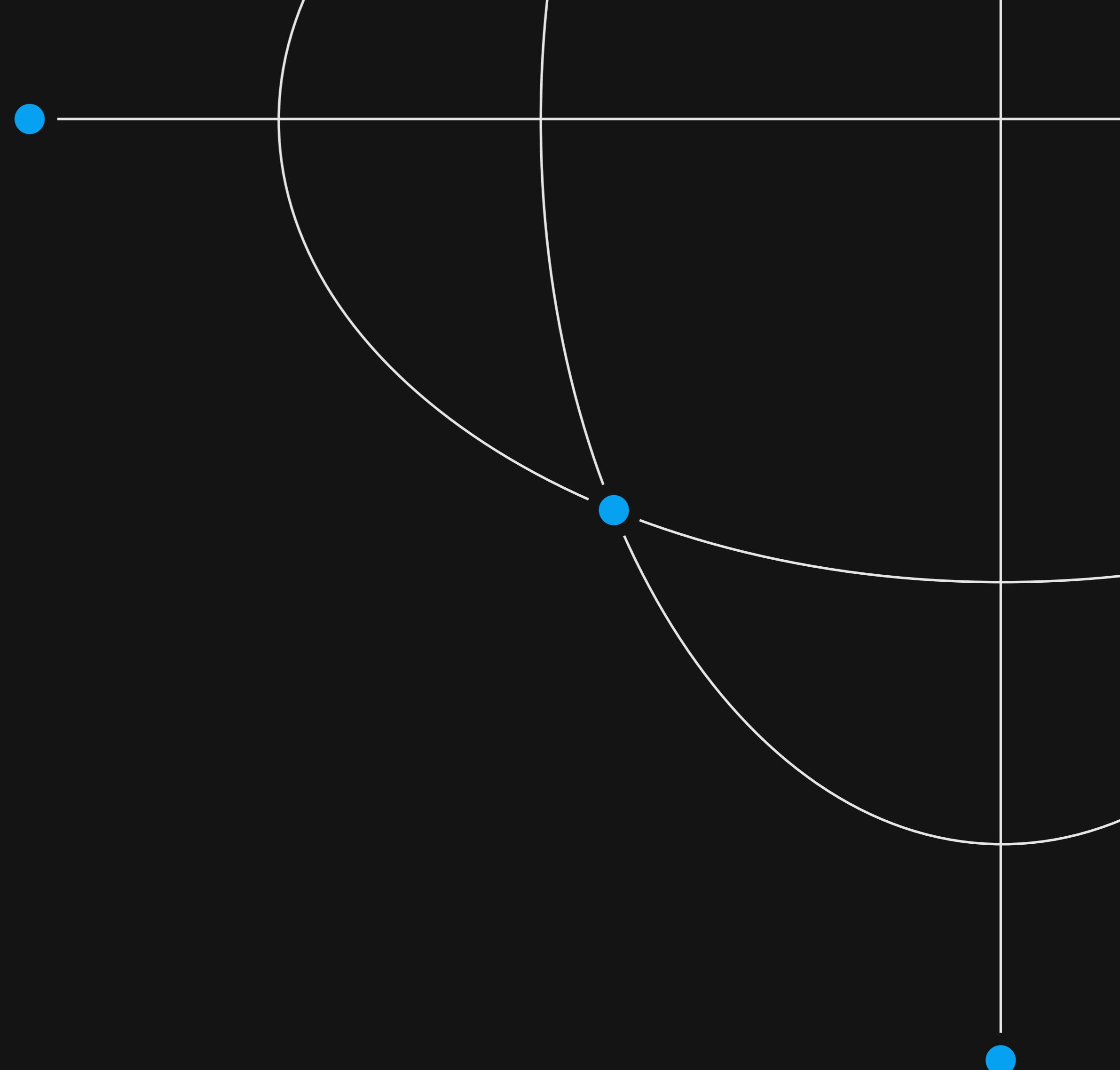


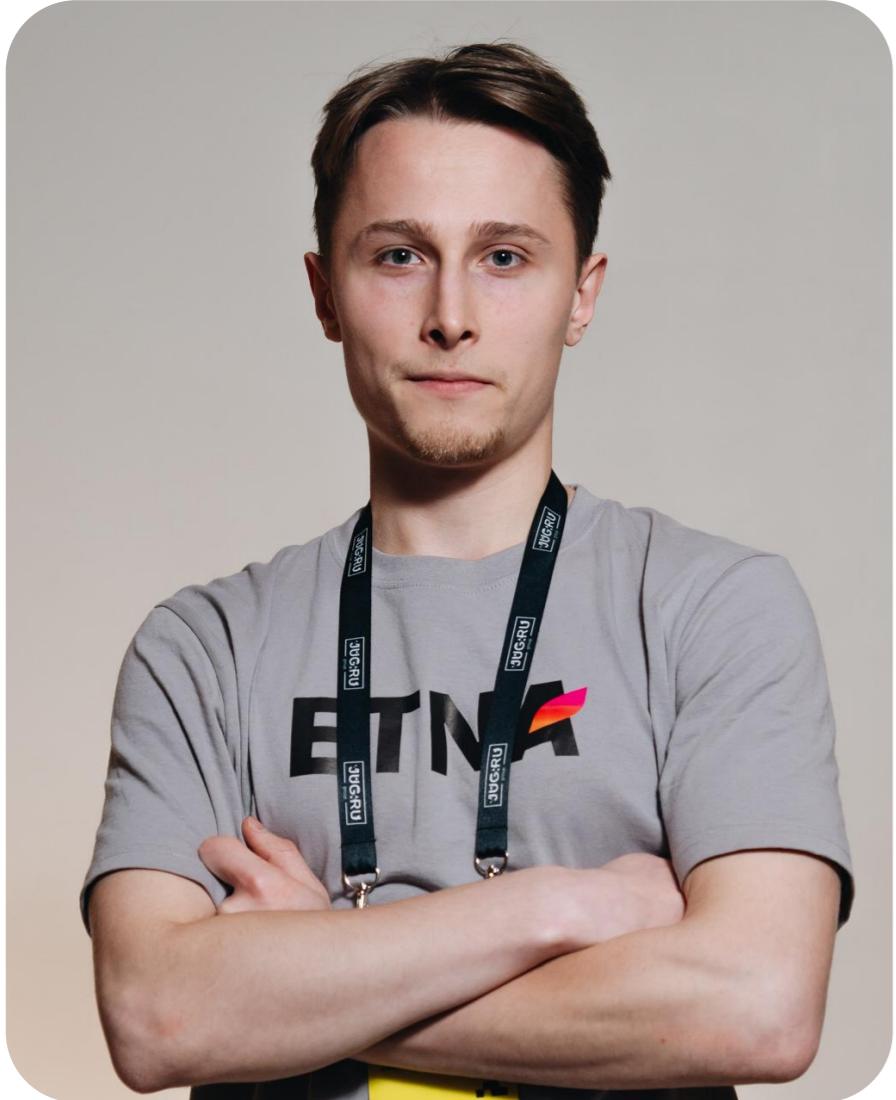
Time Series In Action

День 1



Июль 2025

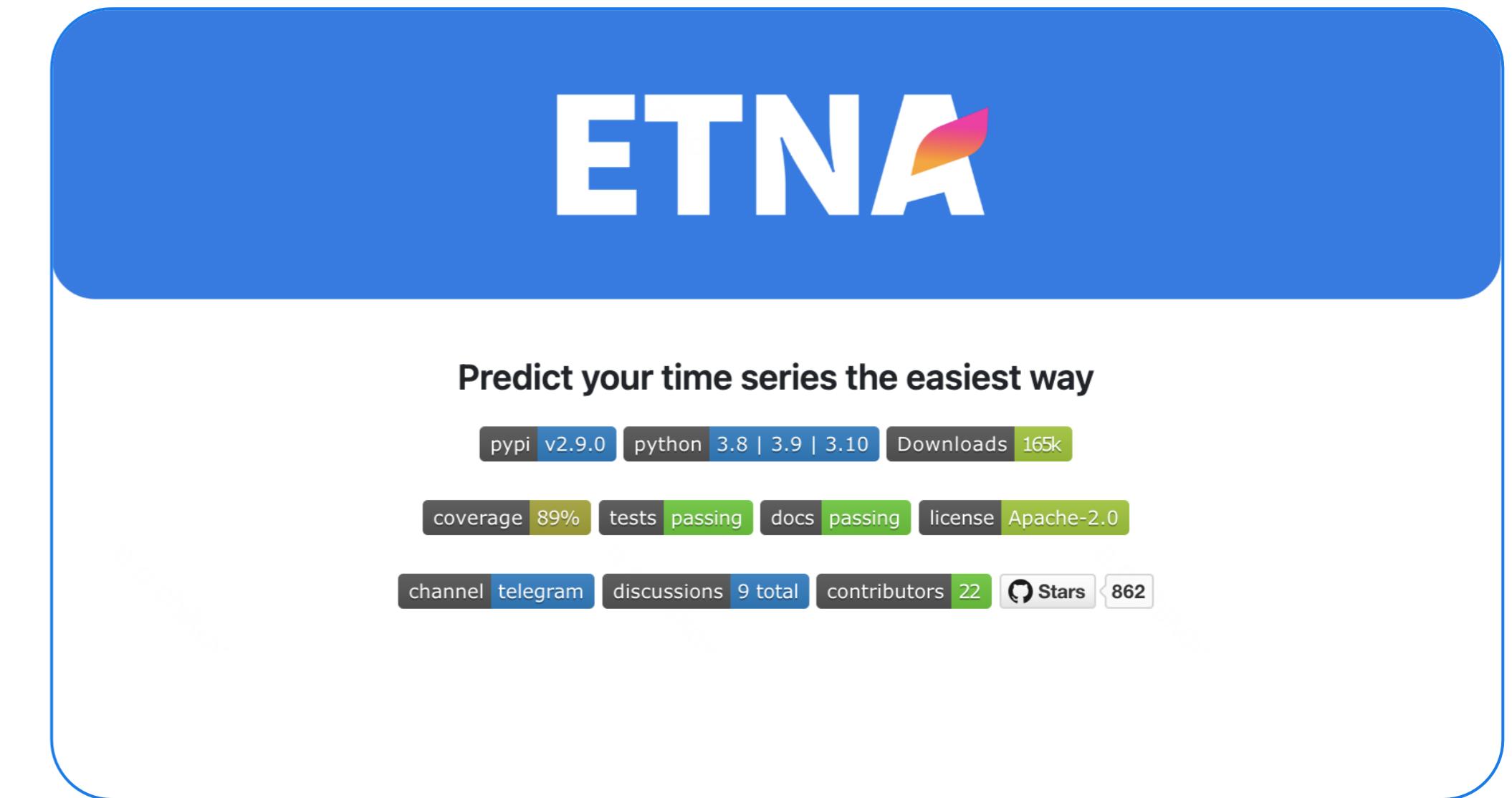
Команда курса



Александр Чиков
«Т-Банк», ех ETNA



Мартин Габдушев
«Т-Банк», ех ETNA



The screenshot shows the GitHub repository page for ETNA. The repository has over 165k downloads and is maintained by 22 contributors. It includes sections for coverage (89%), tests (passing), docs (passing), and license (Apache-2.0). There are also links for the channel (Telegram) and discussions (9 total).

- Разрабатывали open-source-фреймворк для временных рядов в «Т-Банке».
- Выступаем на конференциях про временные ряды.

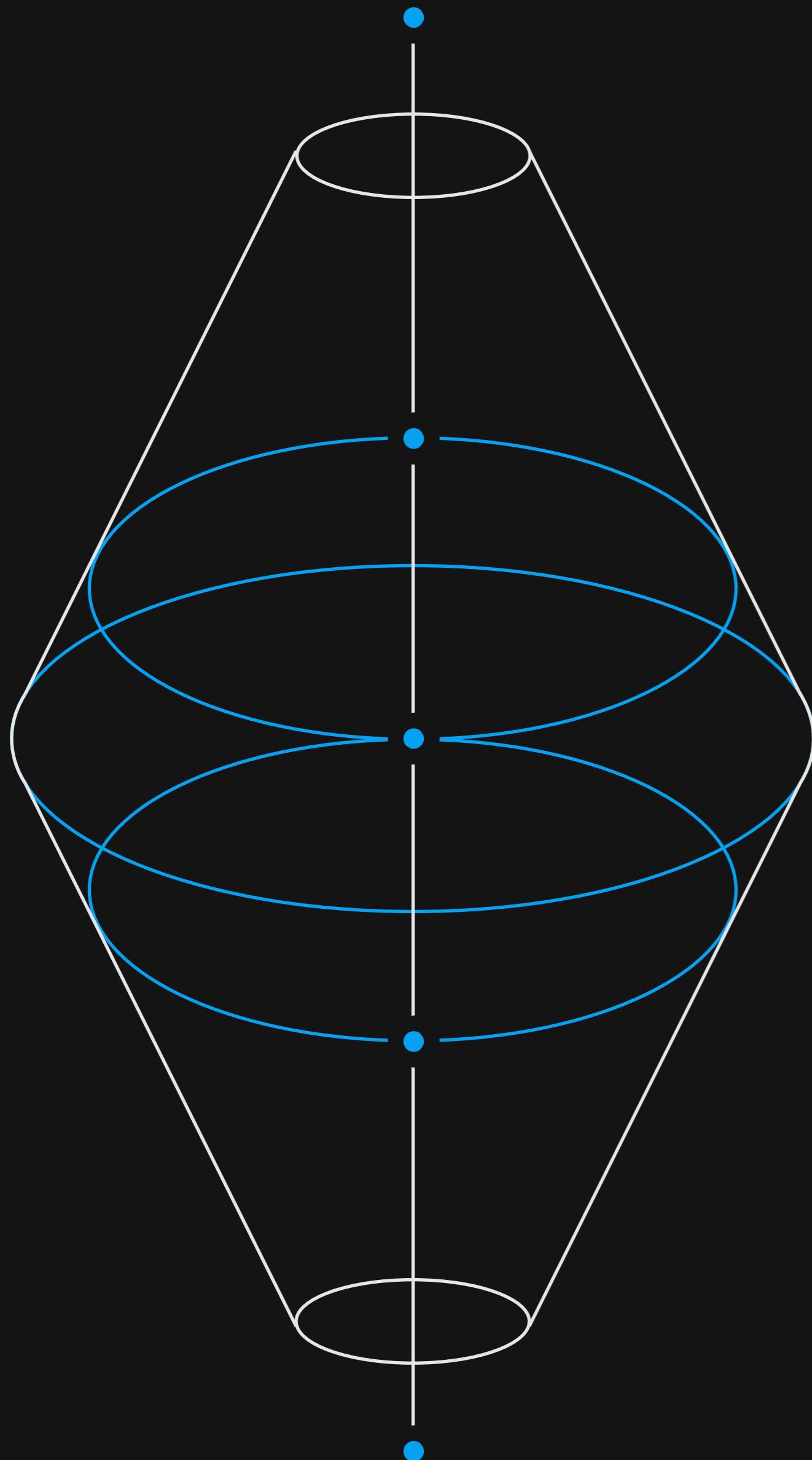
План курса

Part 1. Задача прогнозирования

- 01 Введение в анализ временных рядов.
Классические модели
- 02 Модели машинного обучения.
Предсказательные интервалы
- 03 DL + Foundational Models

Part 2. Другие задачи

- 01 Поиск аномалий и точек смены поведения
- 02 Обучение представлений
- 03 Causal Inference



Формат



Цель: дать общее понимание,
как работать с Time Series.



Теория: чтобы понять, о чём пишут в доке.



Рукомашество: чуть меньше формализации,
чуть больше интуиции.



Может быть ничего непонятно —
это нормально!

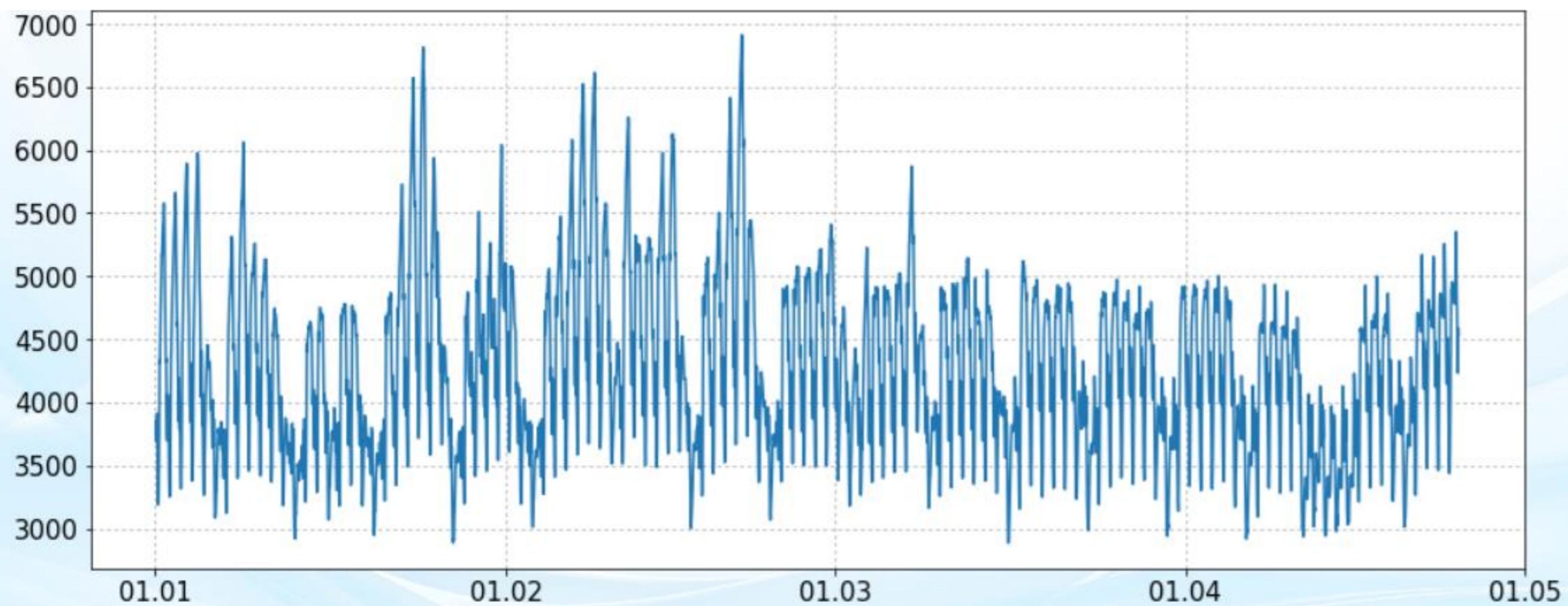
Занятие 1. Введение в анализ временных рядов. Классические модели

Александр Чиков
Исследователь-разработчик, «Т-Банк»
2025

Временные ряды

Временной ряд — последовательность измерений некоторой величины в равноудалённые моменты времени ($y_t, t \in N$).

Electricity consumption in Victoria, Australia

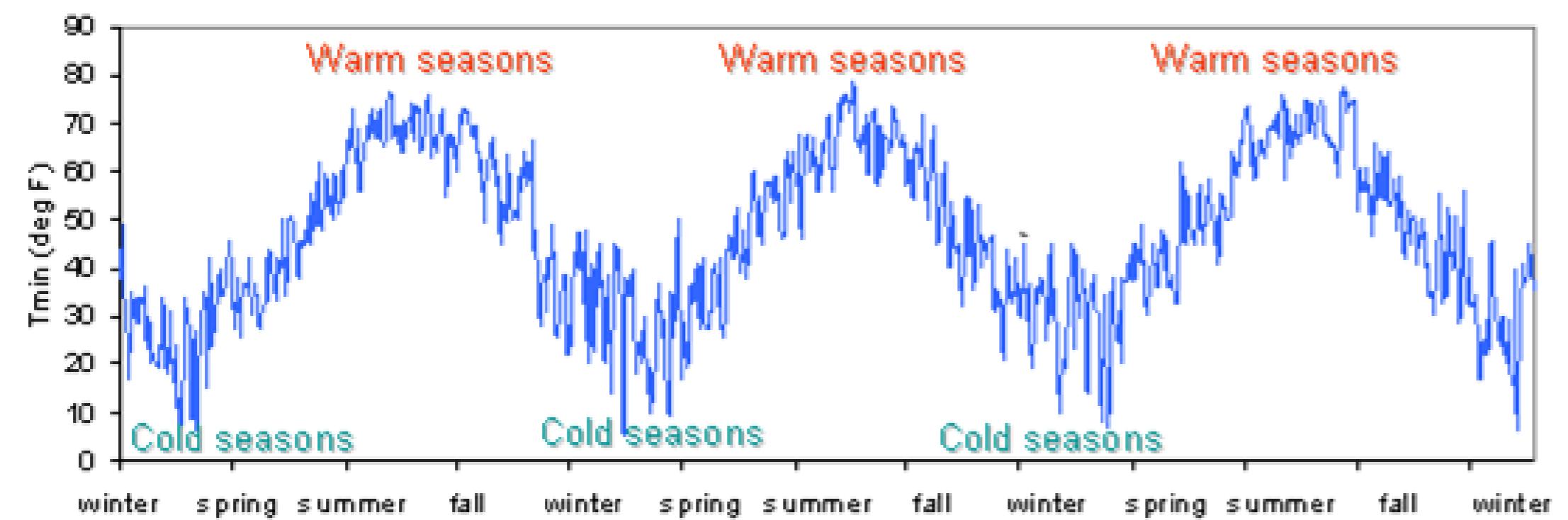


Величина:
потребление электричества.

Моменты времени:
каждый час.

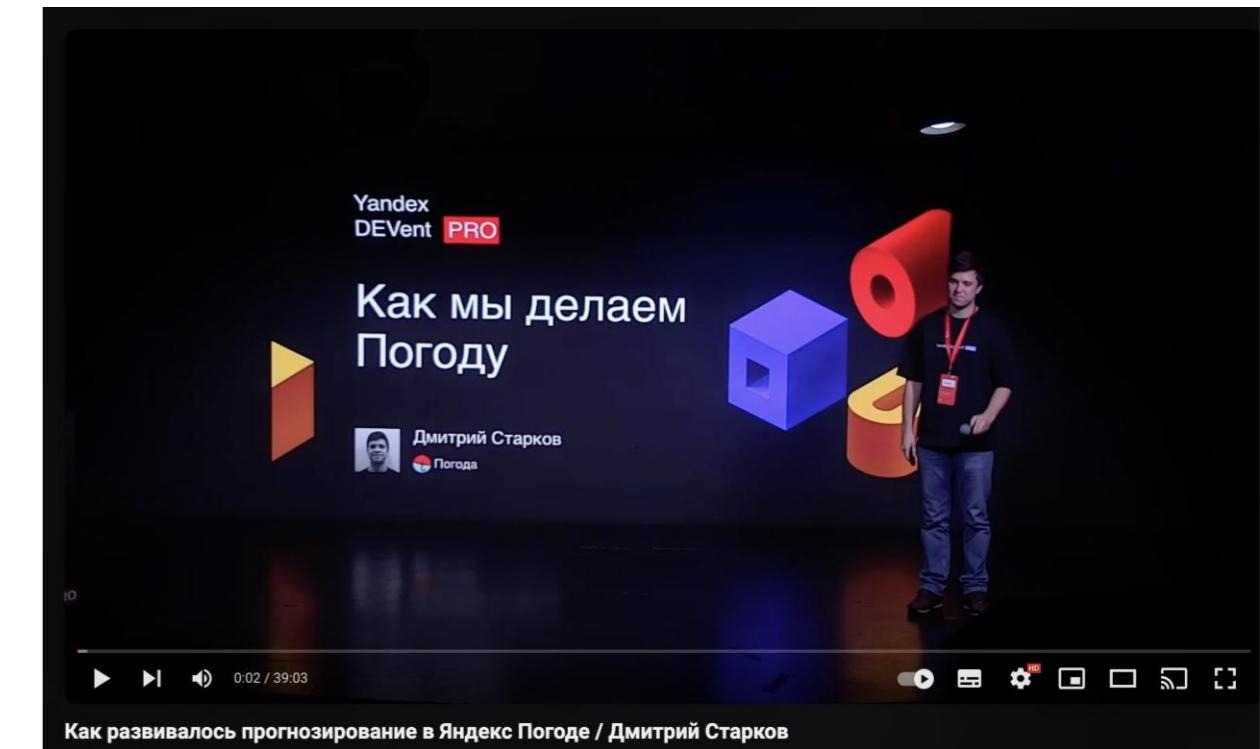
| timestamp | target |
|---------------------|--------|
| 2025-01-01 12:00:00 | 1000 |
| 2025-01-01 13:00:00 | 1500 |
| 2025-01-01 14:00:00 | 950 |

Метеорология



Временные ряды

- Температура воздуха
- Наличие осадков
- Атмосферное давление

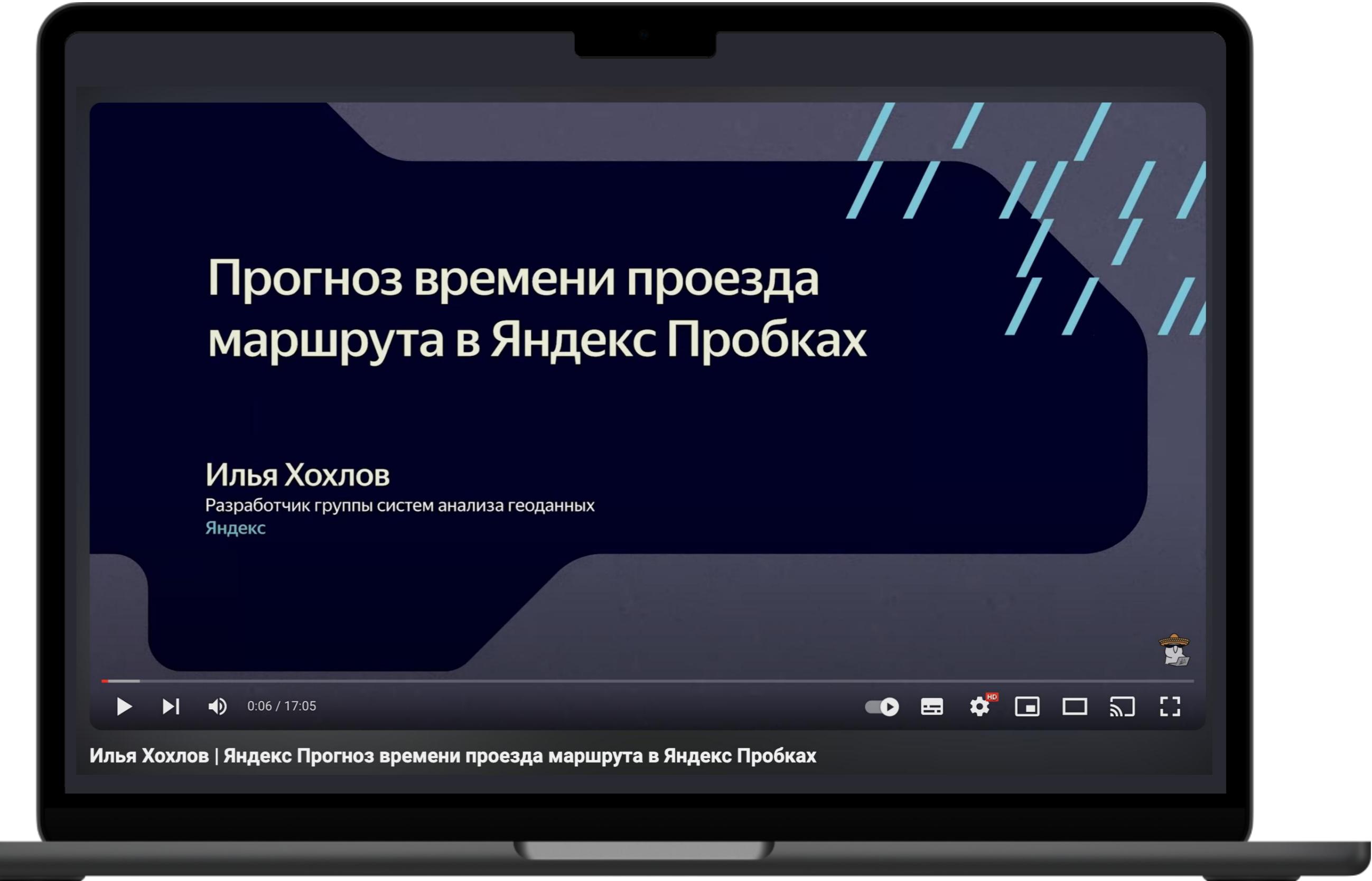


[Дмитрий Старков | Как развивалось прогнозирование в «Яндекс Погоде»](#)

Транспорт



Предсказание трафика
для оптимальной навигации



[Илья Хохлов | «Яндекс Прогноз» времени проезда маршрута в «Яндекс Пробках»](#)

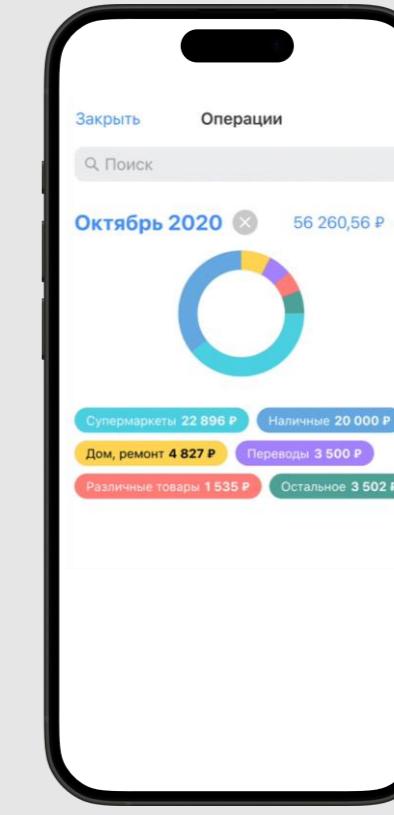
Финтех



Прогнозирование оборота
для планирования
инкассаций



Прогноз цен
на авиабилеты
для планирования поездок



Траты клиентов

- Планирование бюджета
- Детекция фрода
- Оферы по кешбэку

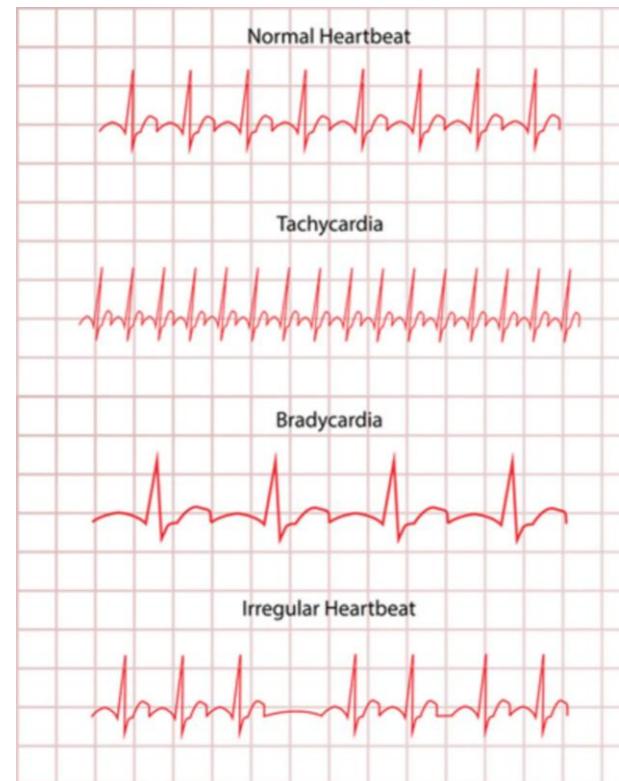
Фондовый рынок

Временные ряды = цены активов

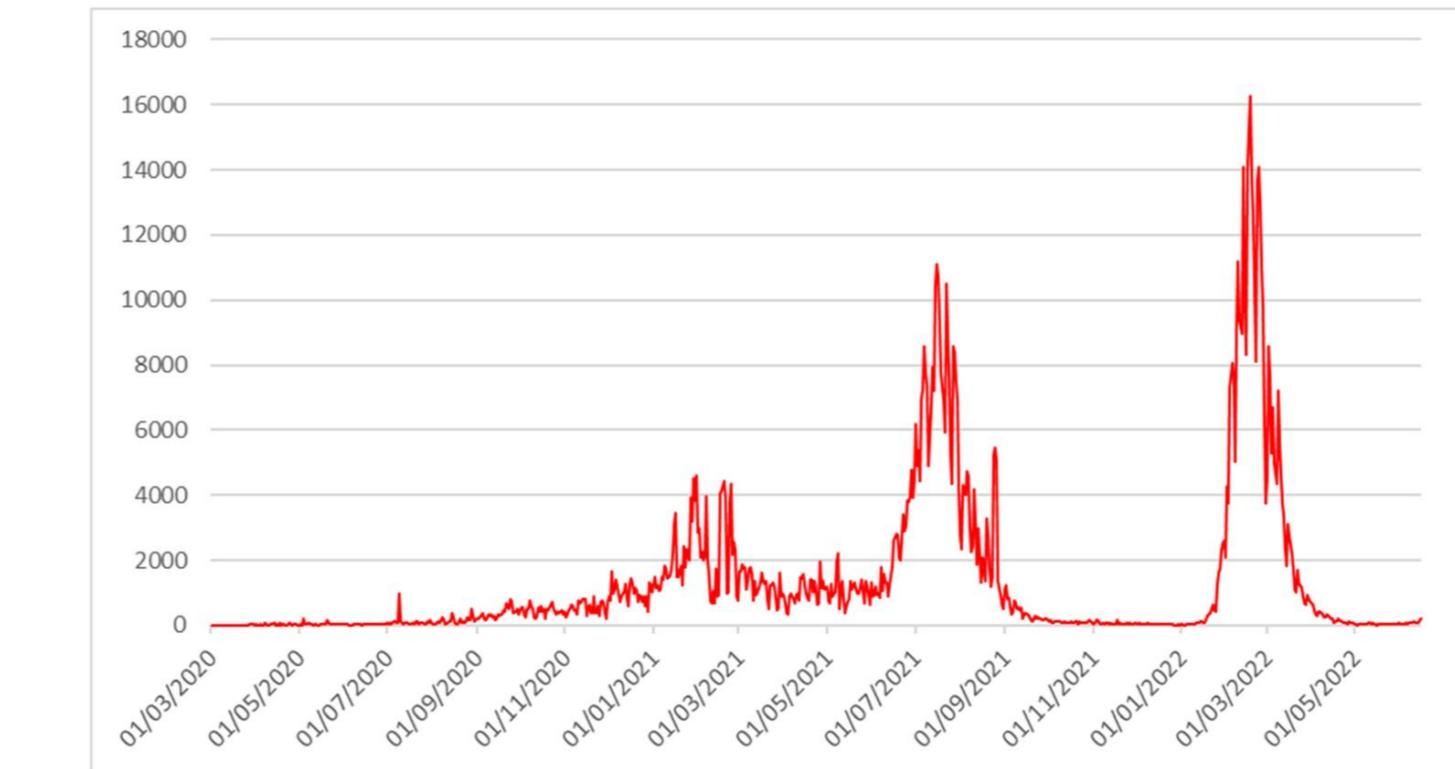


Предсказываем цены: зарабатываем на бирже.

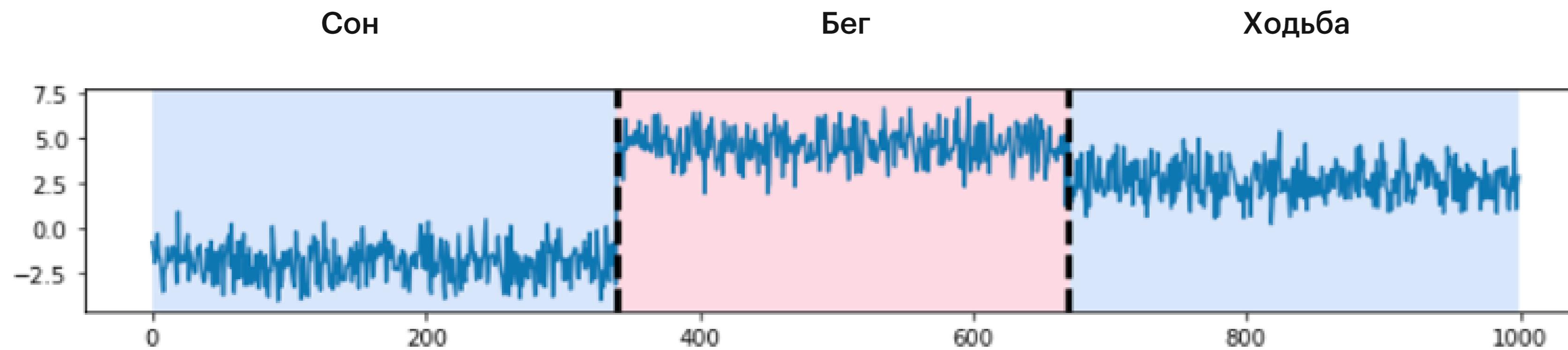
Медицина



Определение
болезни по ЭКГ

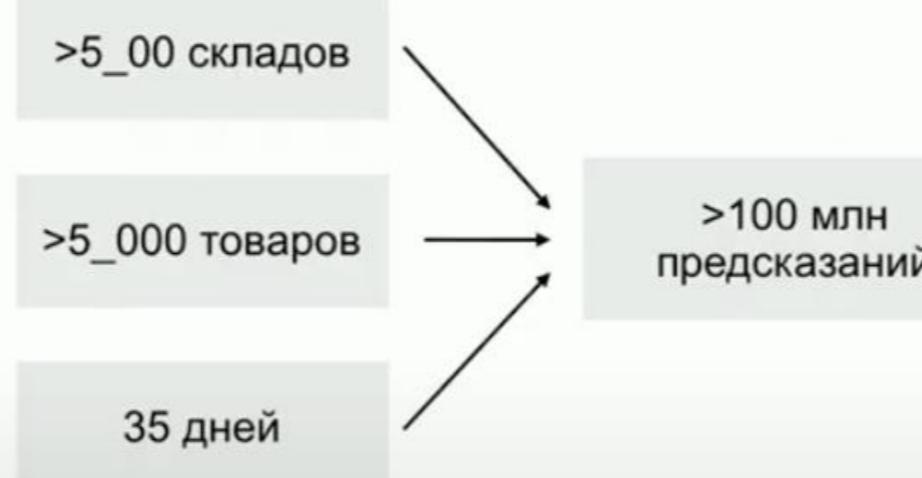


Предсказание
распространения
COVID

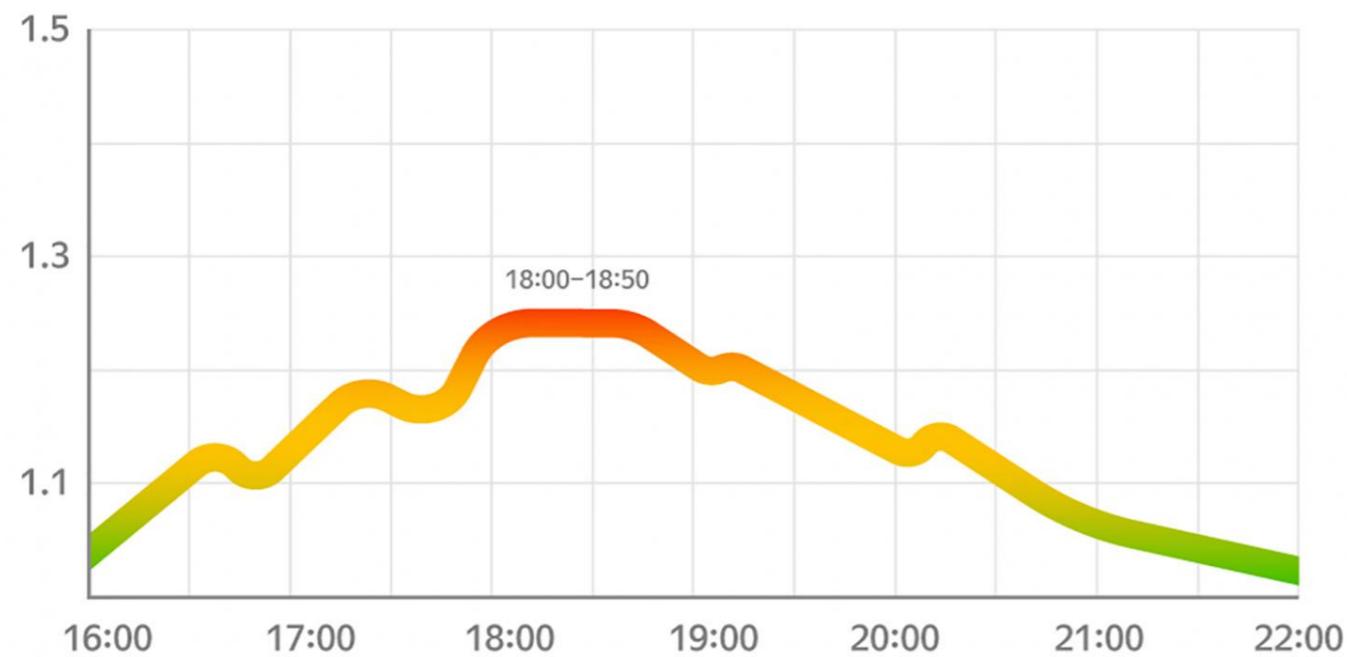


Сегментация
по виду активности
пользователя
в смарт-часах

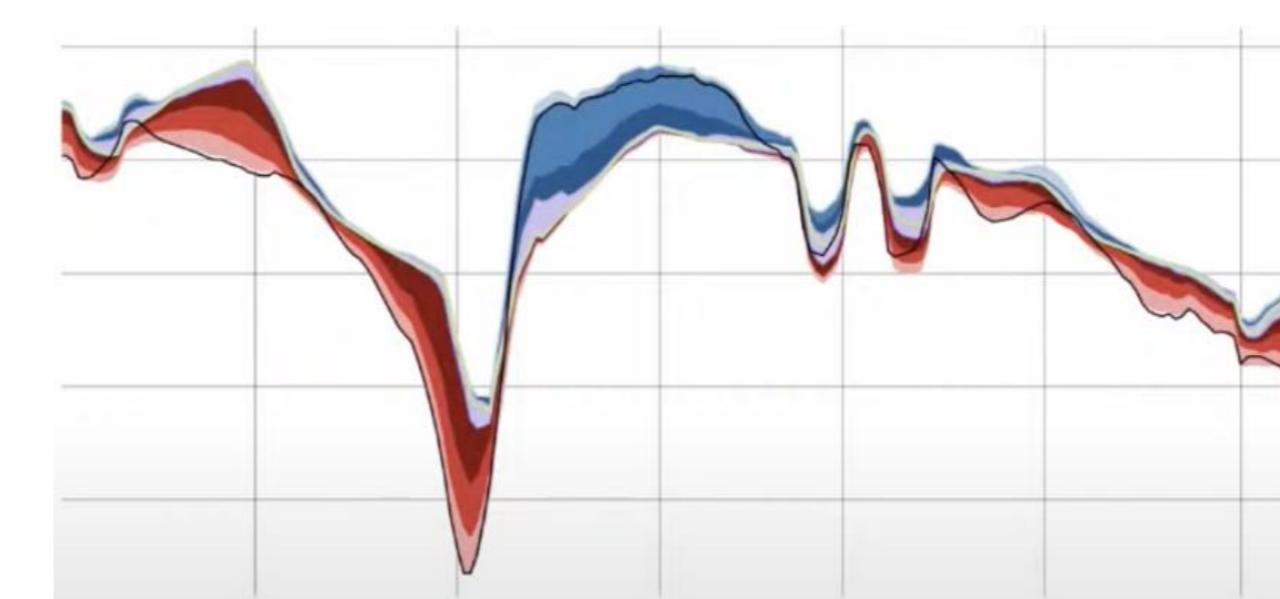
E-commerce



Прогноз спроса на товары
для планирования поставок

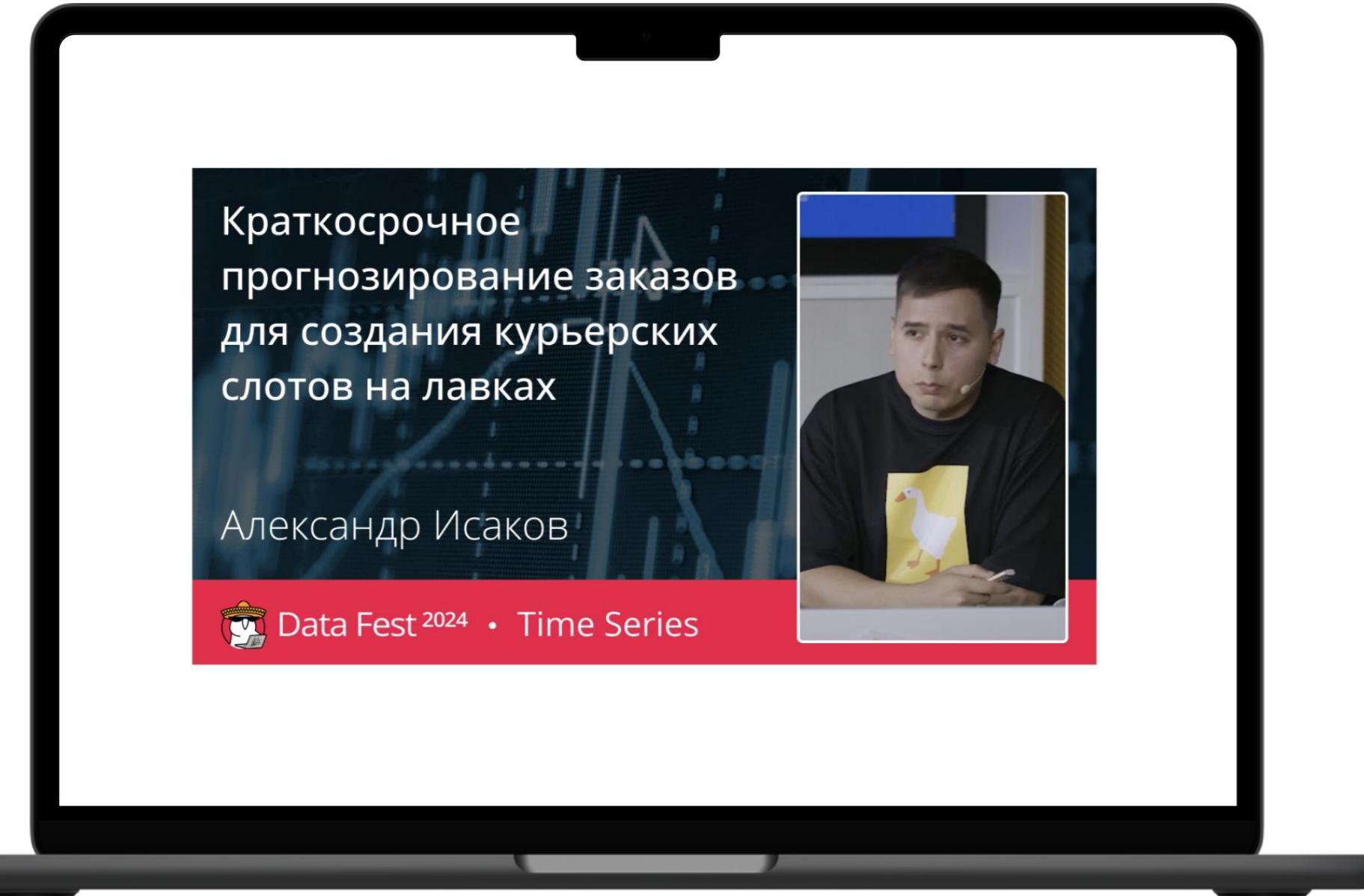


Динамическое
ценообразование

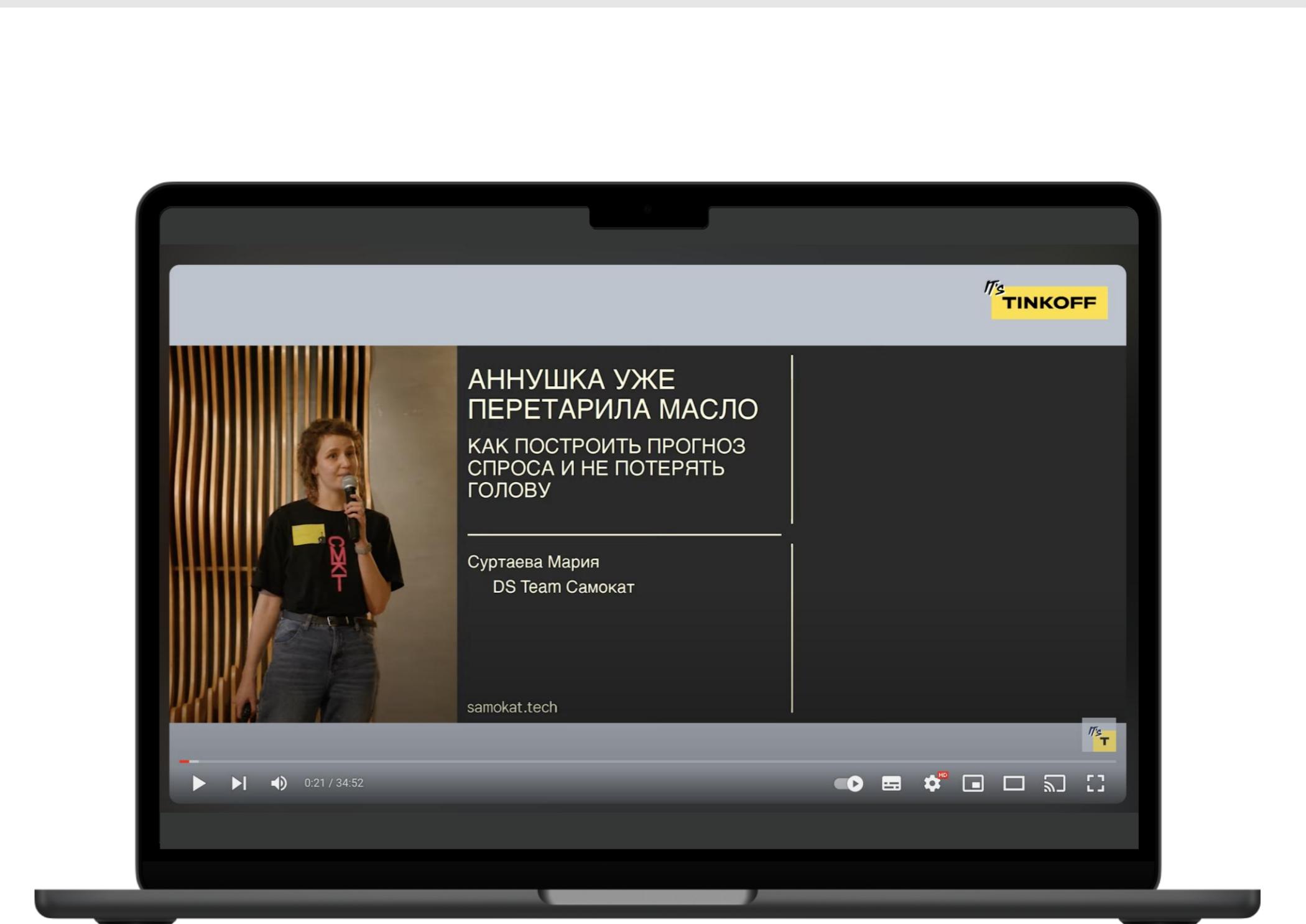


Анализ
эффективности
маркетинговых
компаний

E-commerce

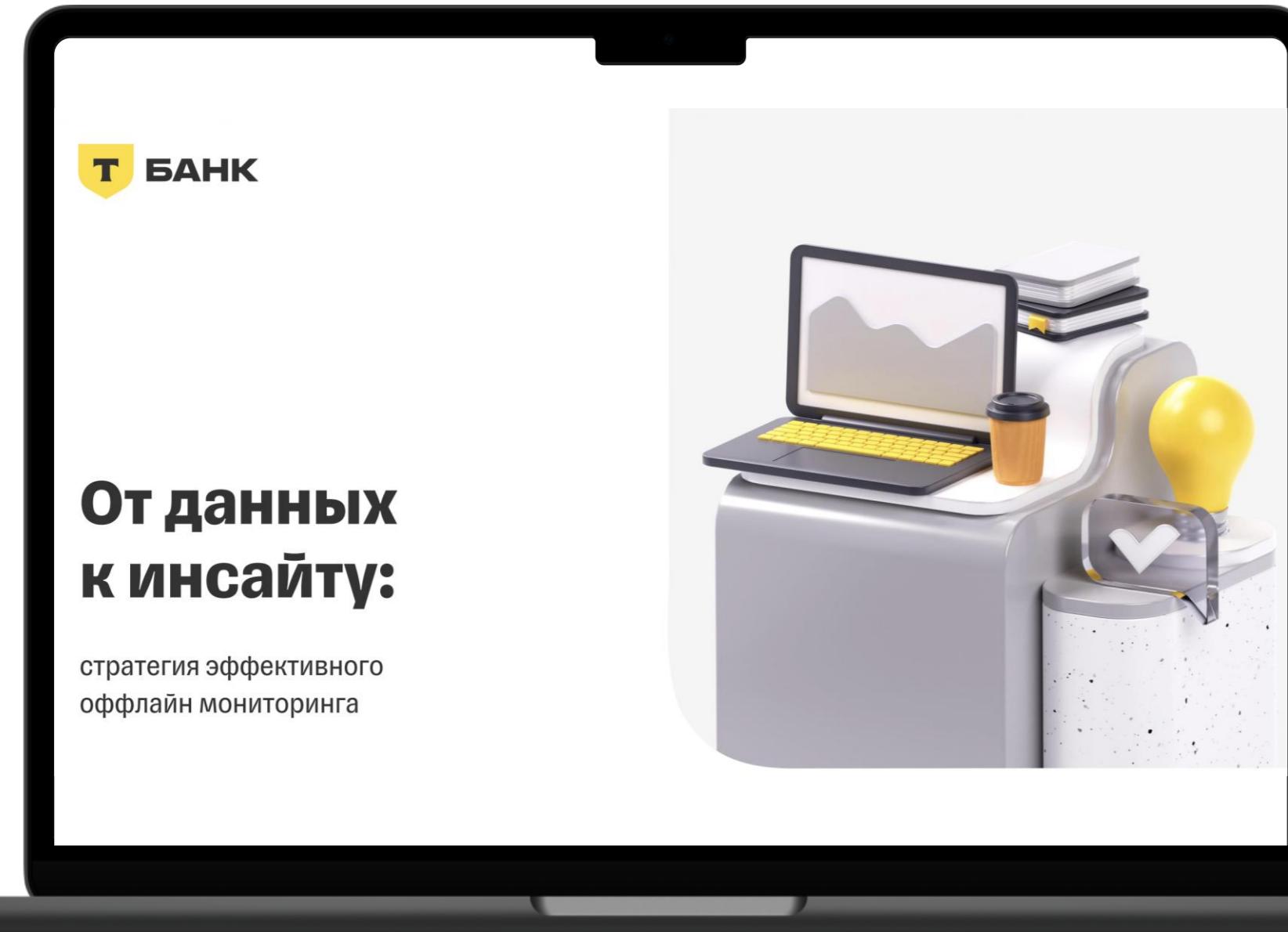


[Александр Исаев | Краткосрочное прогнозирование заказов для создания слотов на лавках](#)

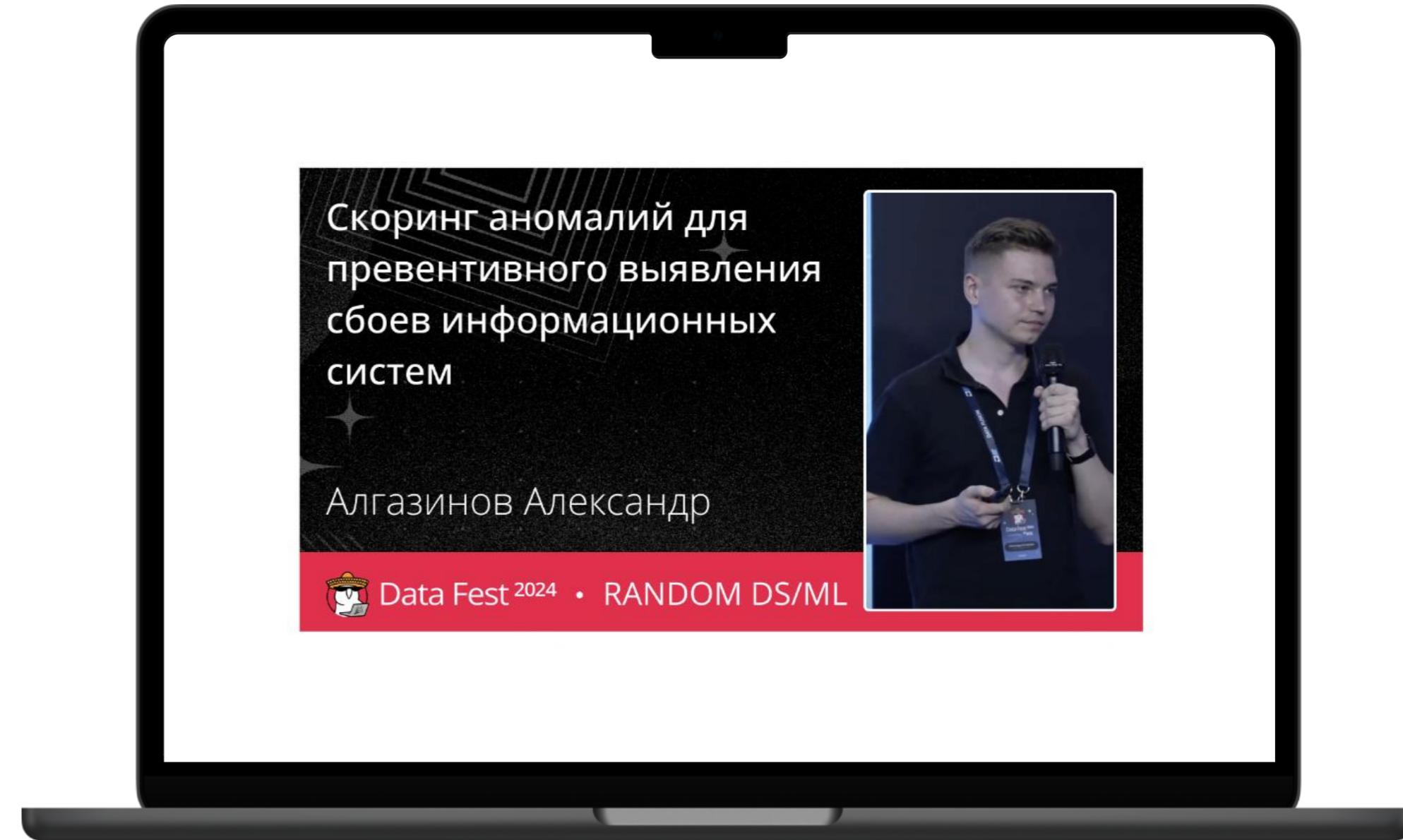


[Мария Суртсева, «Самокат» | Как построить прогноз спроса и не потерять голову](#)

Системы мониторинга



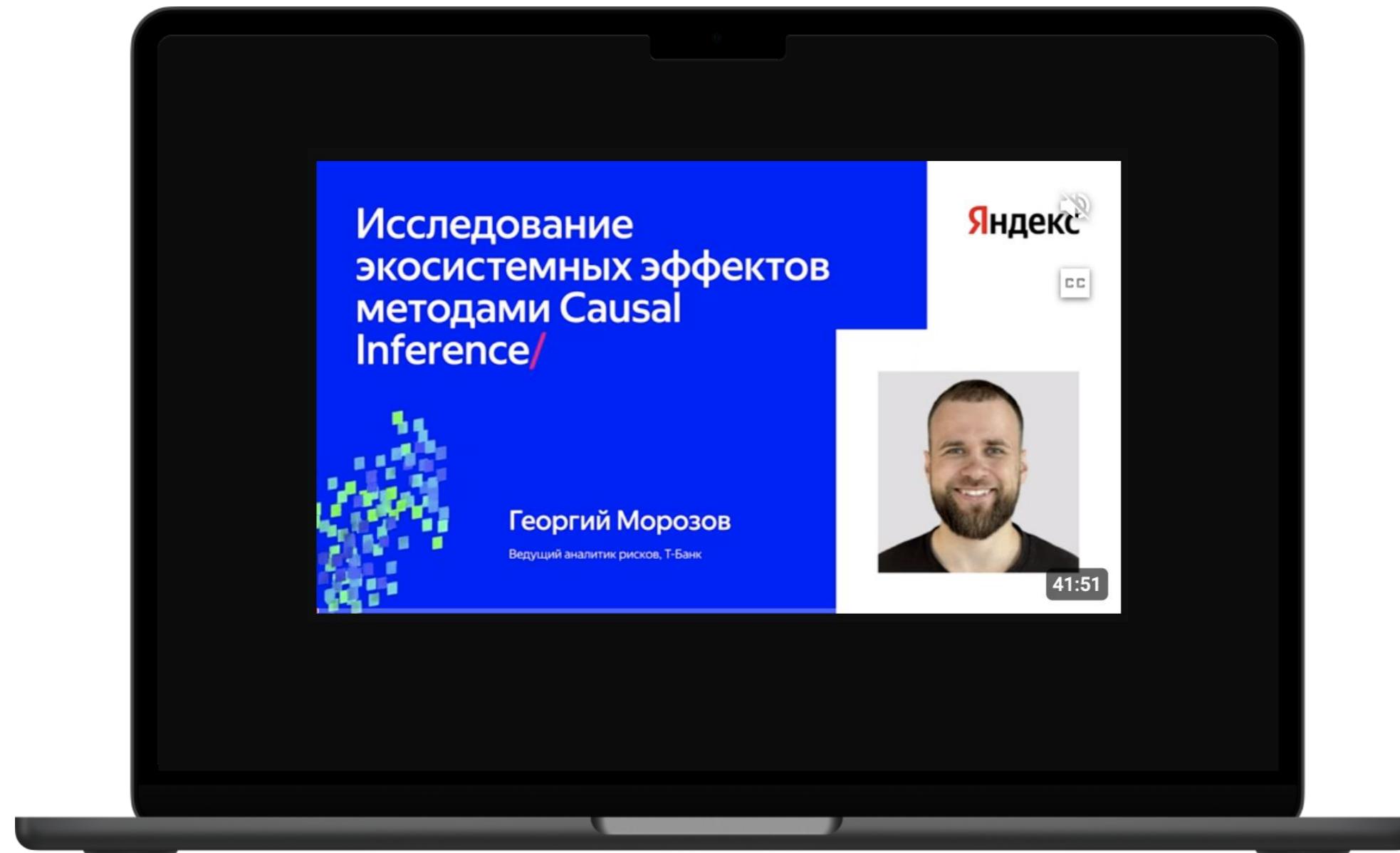
[Максим Жерело | От данных к инсайту:
стратегия эффективного оффлайн-мониторинга](#)



[Александр Алгазинов | Скоринг аномалий для превентивного
выявления сбоев информационных систем](#)

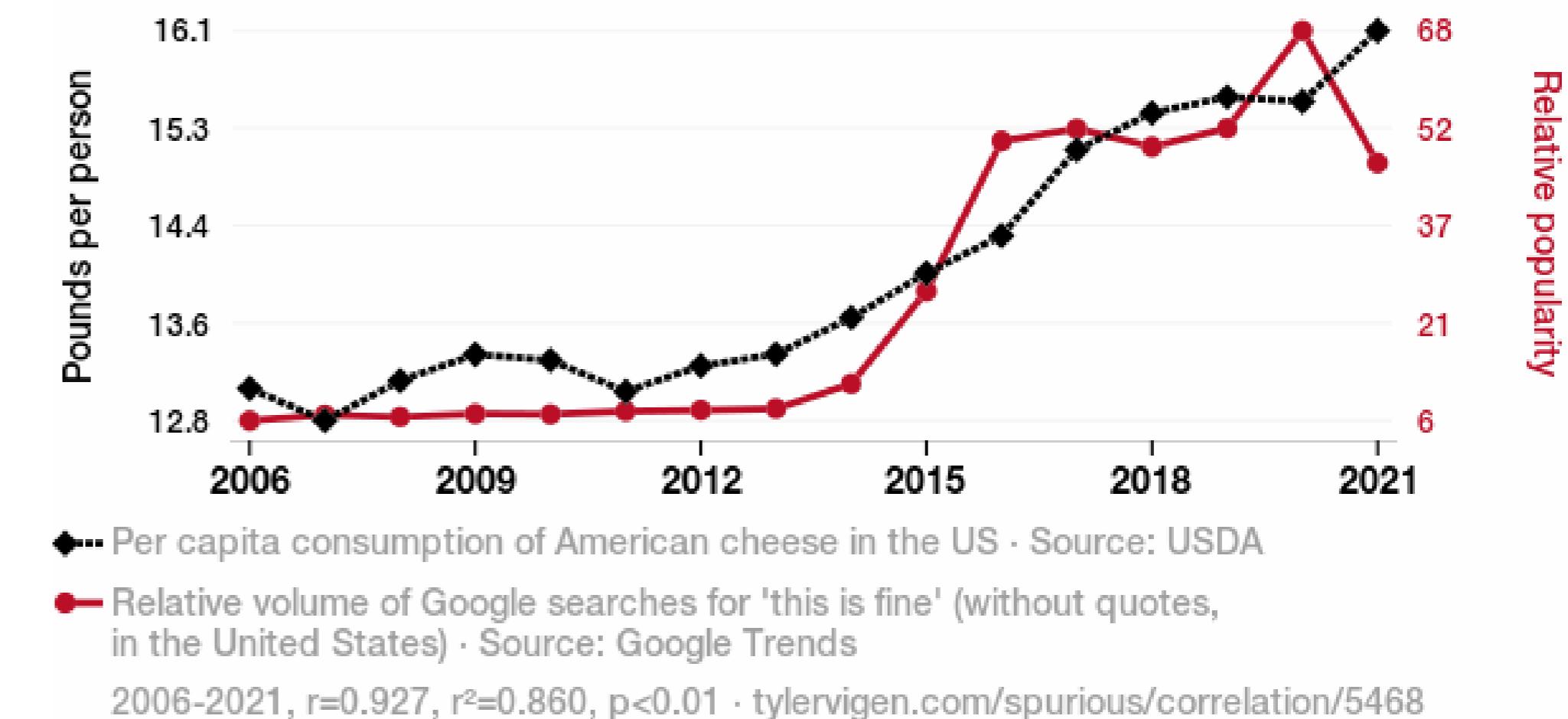
→ Хотим автоматически обнаруживать сбои в различных системах.

Поиск причинностей



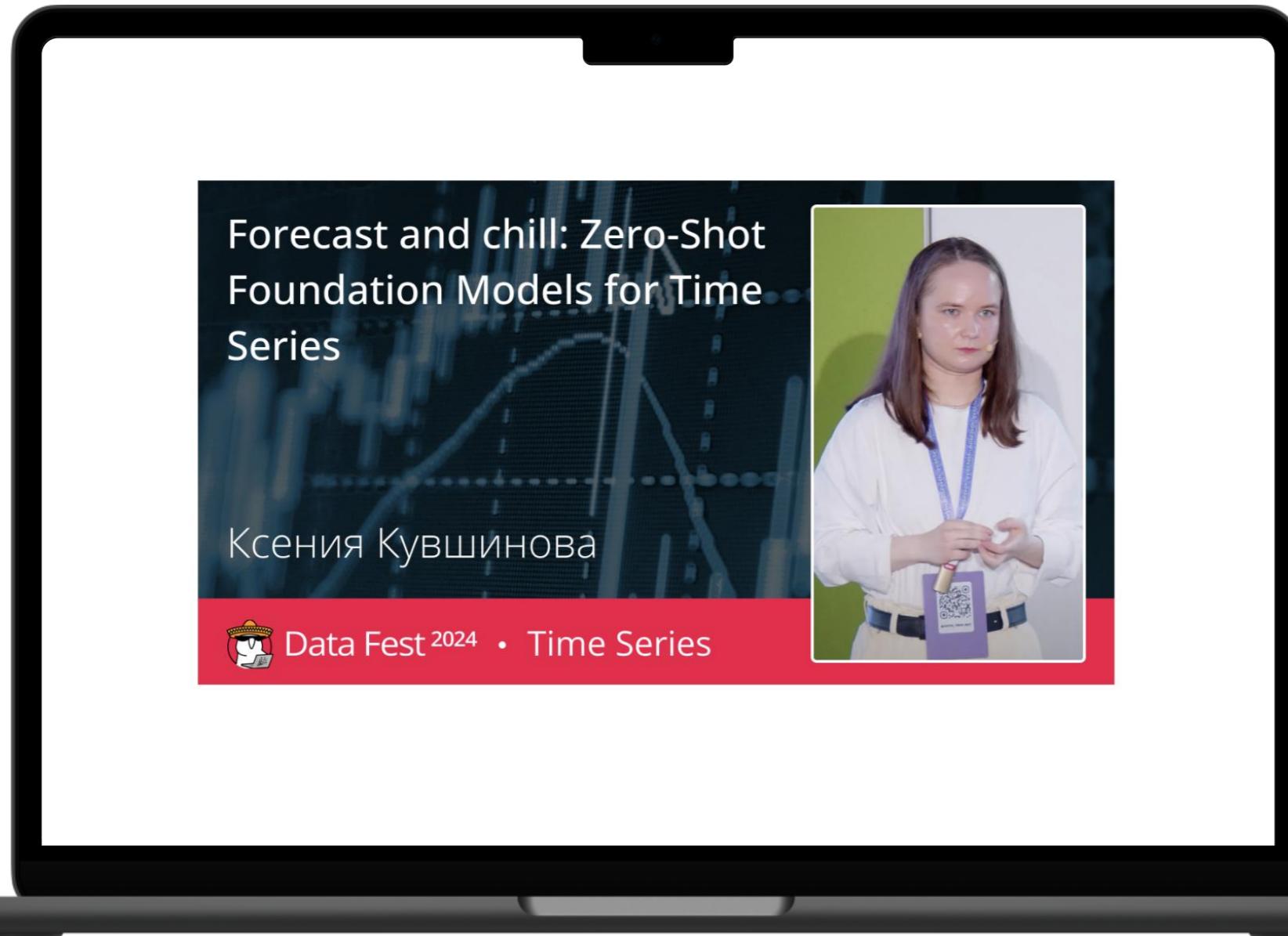
[Максим Жерело | От данных к инсайту:
стратегия эффективного онлайн-мониторинга](#)

American cheese consumption
correlates with
Popularity of the 'this is fine' meme

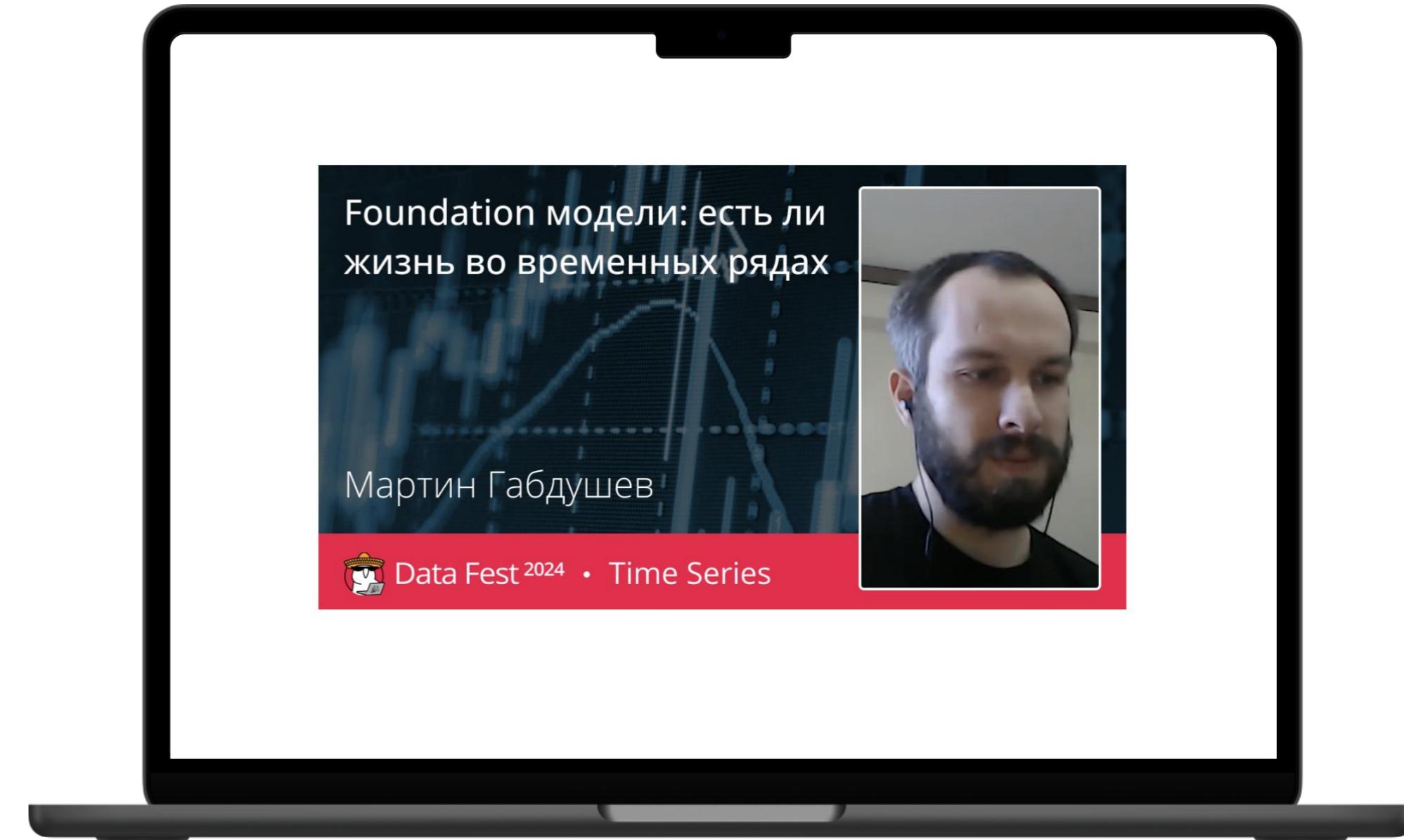


→ Хотим определить наличие причинно-следственной связи
между переменными и измерить величину эффекта.

Foundational Models



[Ксения Кувшинова | Forecast and chill: Zero-Shot Foundational Models for Time Series](#)



[Мартин Габдушев | Foundational-модели: есть ли жизнь во временных рядах](#)

→ Хотим обучить большую модель, которая умеет решать основные задачи в Zero-shot.

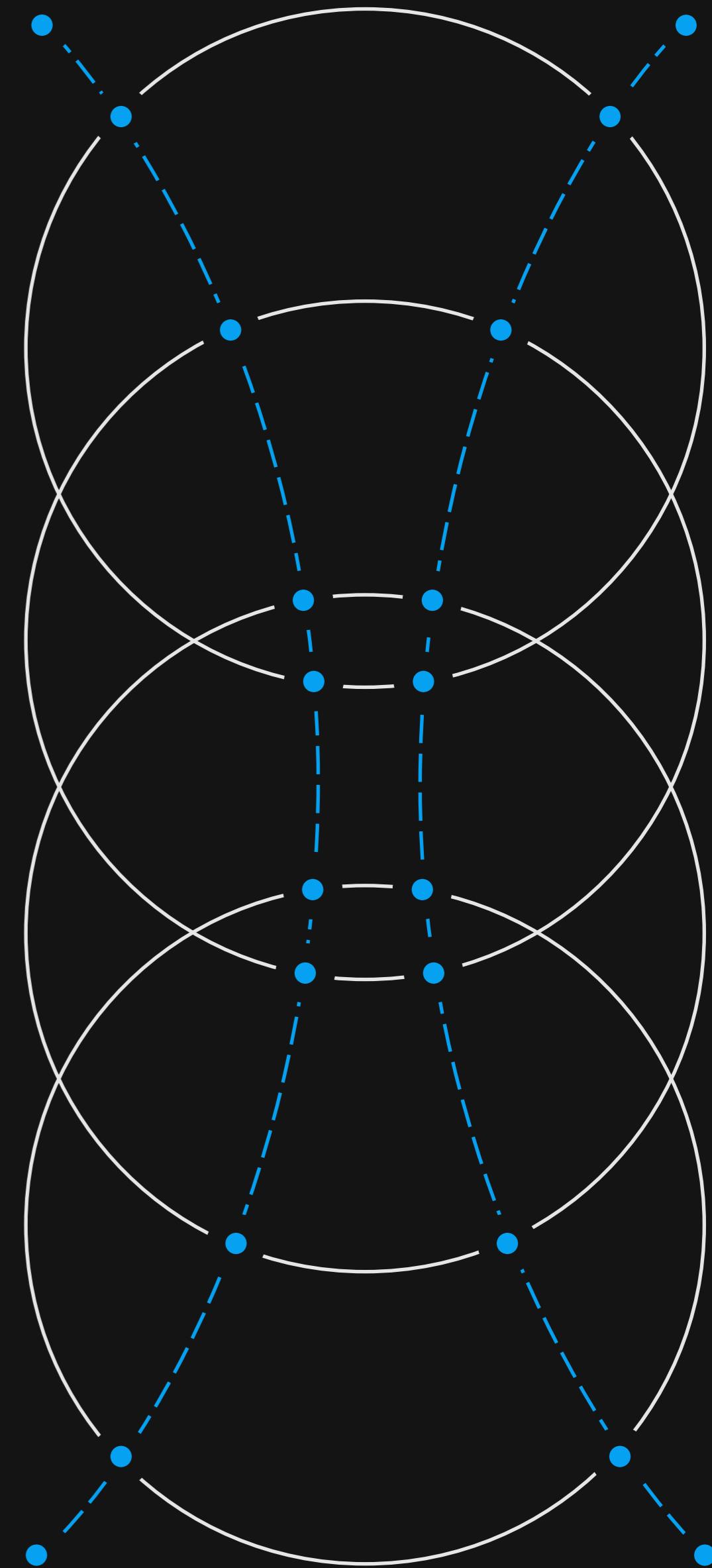
Задачи на временных рядах

- Прогнозирование
- Поиск аномалий
- Обучение представлений
- Классификация
- Causal Inference
- Foundational Models
- Сегментация

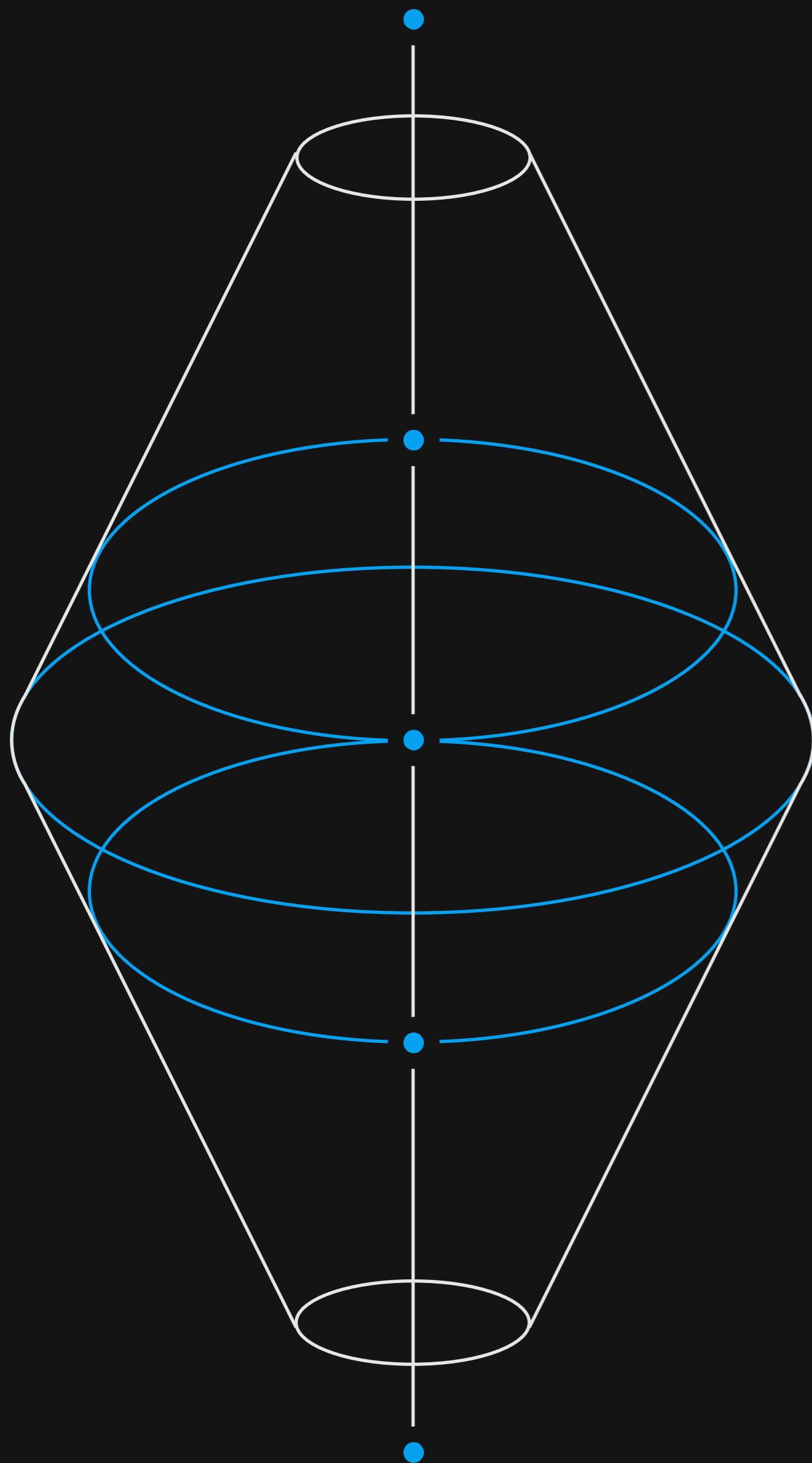


Особенности данных

Что бы мы хотели учитывать в моделях



1. Типы временных рядов



Типы временных рядов



По частоте замеров: регулярность точек на временной оси.

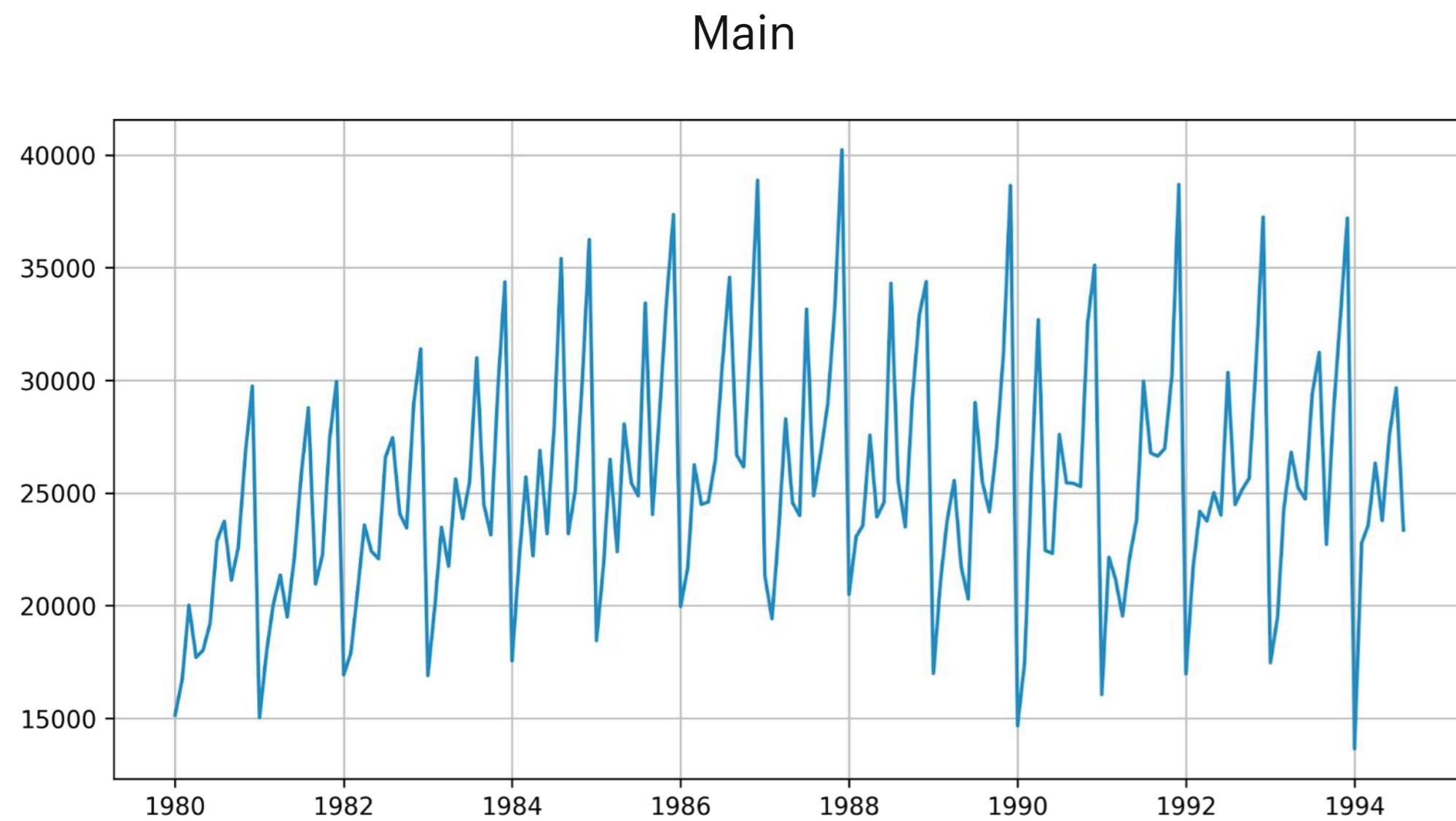


По структуре датасета: количество и связность временных рядов в датасете.



По наличию иерархии: иерархическая структура связи между рядами.

Регулярные ряды

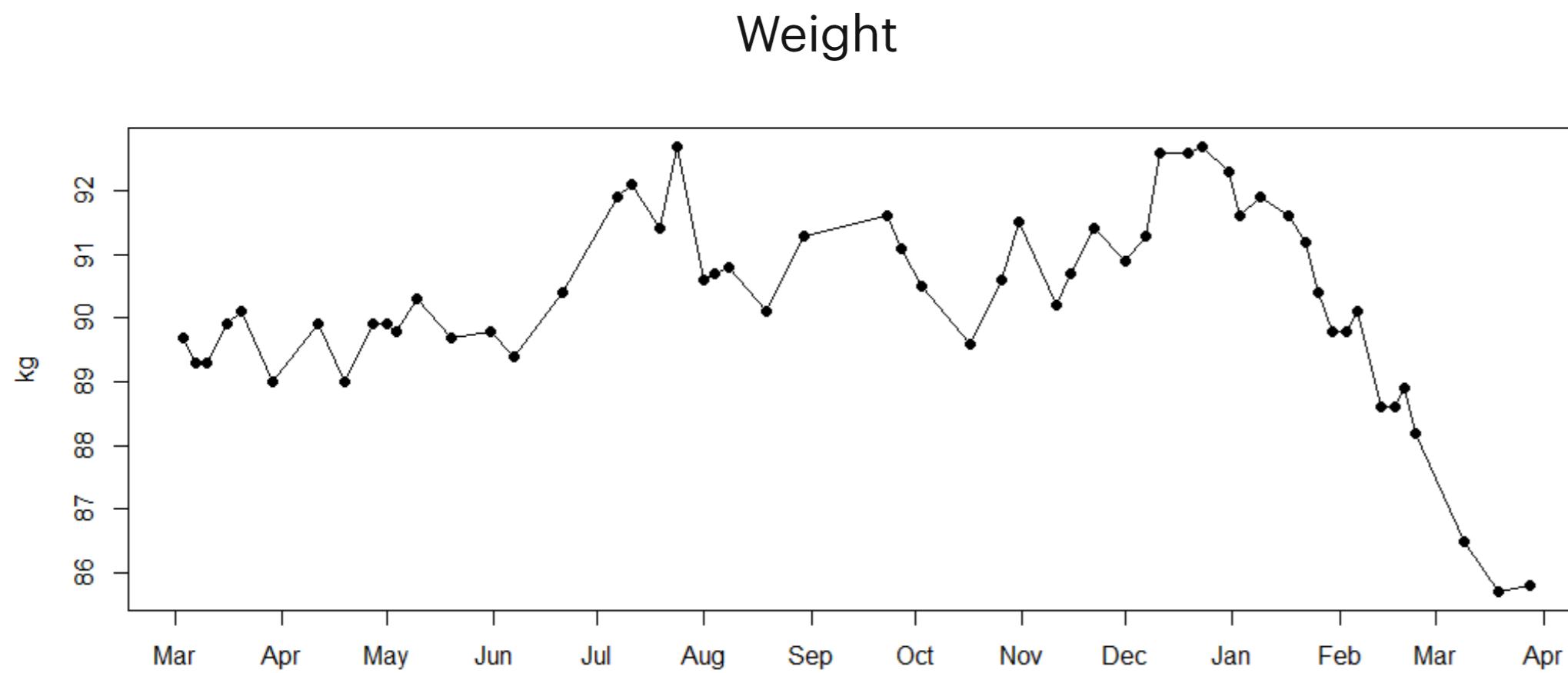


→ Замеры через **равные промежутки времени**

→ **Примеры**

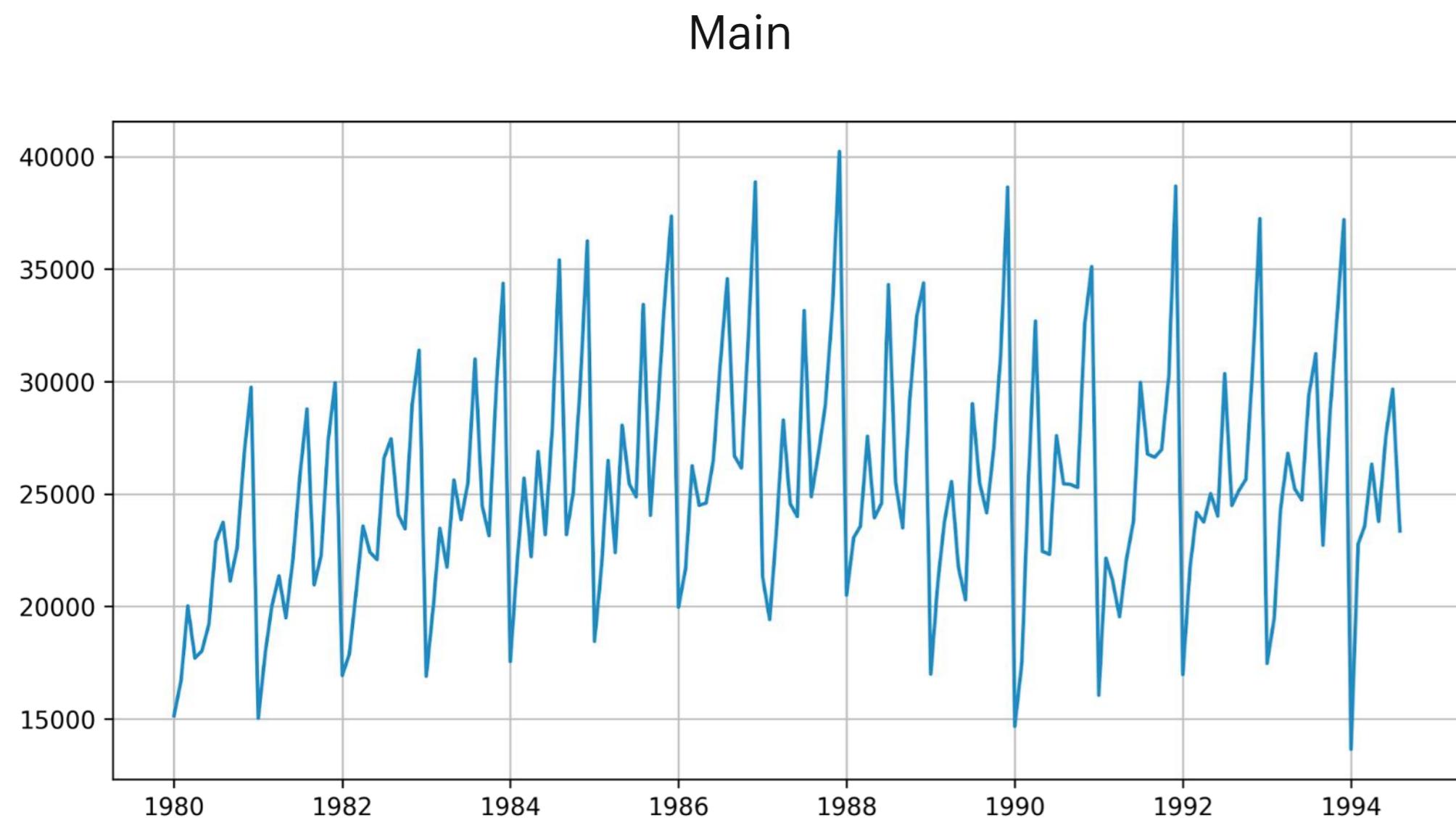
- Продажи вина в месяц
- Среднедневная температура
- Потребление электричества в час

Нерегулярные ряды



- Замеры в произвольные точки времени
- Можно свести к регулярным, уменьшив частотность или перейдя к рядам с пропусками.
- Примеры
 - Замеры своего веса
 - Время ответа сервера на запрос
 - Сумма транзакции

Univariate

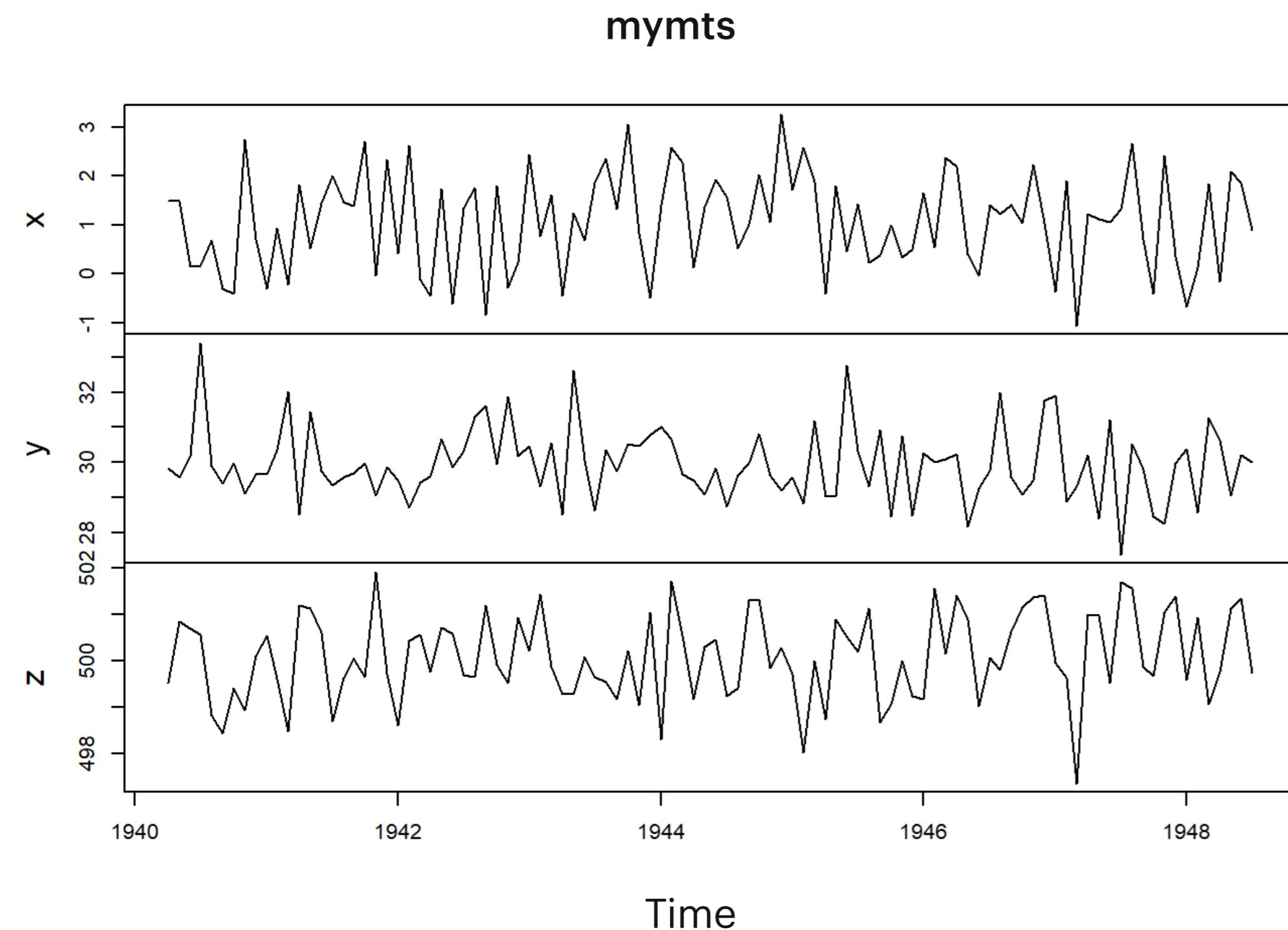


→ Измерение одной величины

→ Примеры

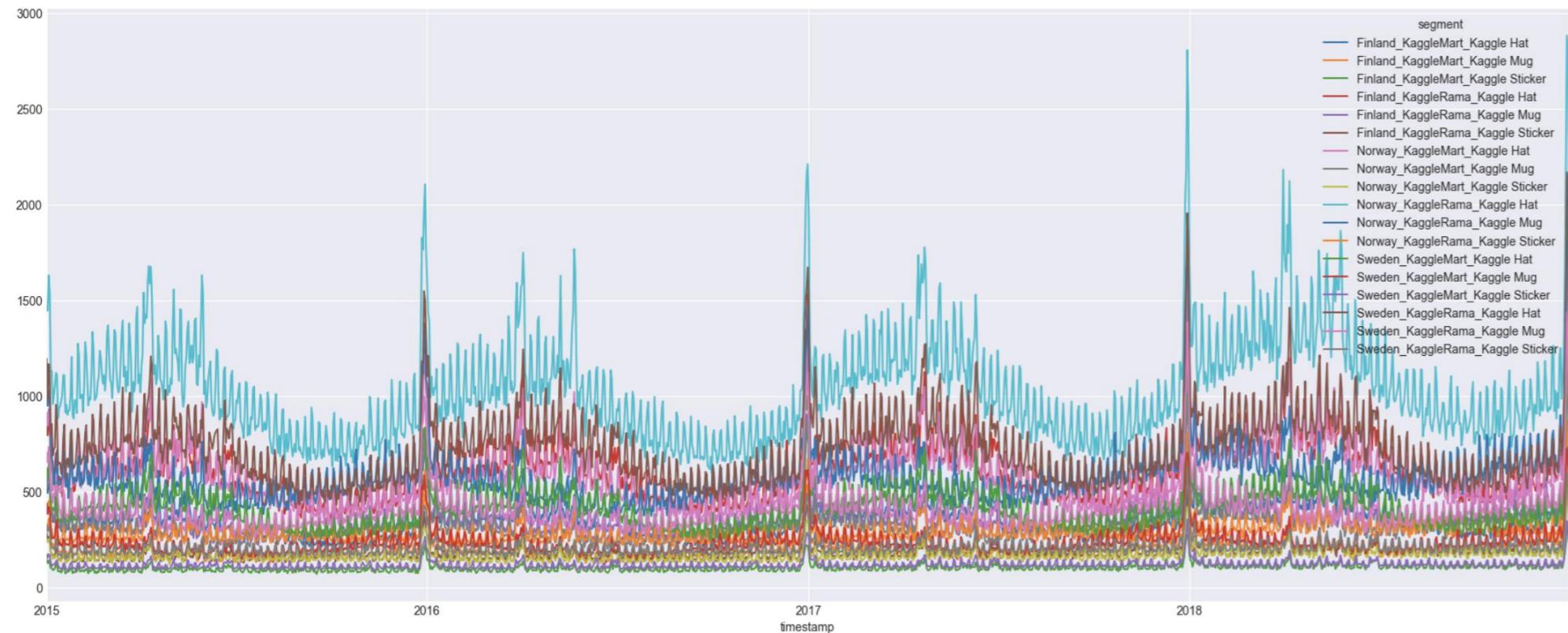
- Продажи кепок в моём ларьке
- Количество пользователей приложения

Multivariate



- Измерения нескольких величин в рамках **одного объекта**
- Значения величин в один момент времени не независимы.
- Примеры
 - Координаты дрона
 - Давление, пульс, температура человека

Panel



→ Изменение значений одной или нескольких переменных для группы объектов

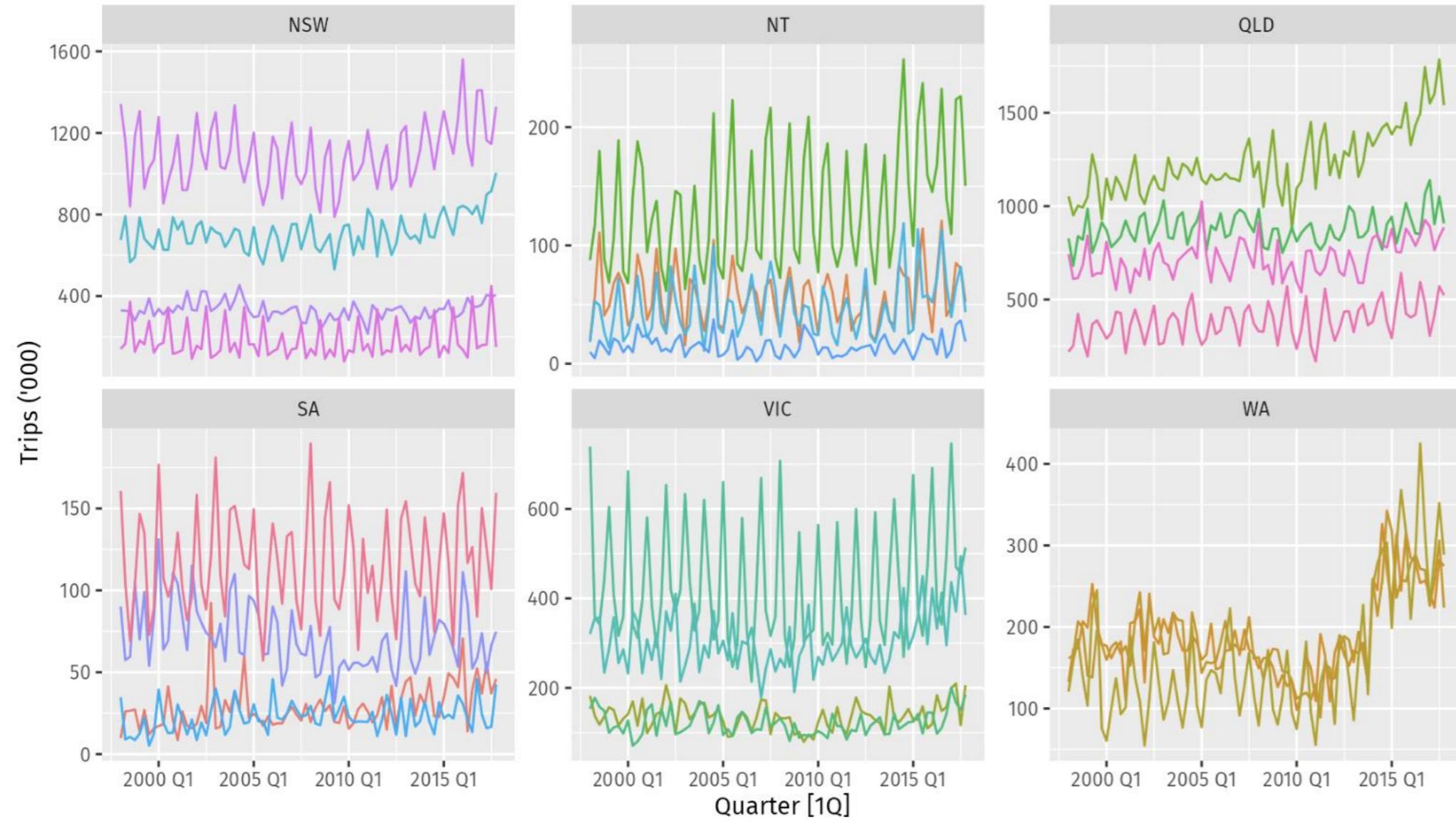
→ Значения величин одних объектов могут влиять на значения у других.

→ Примеры

- Panel Multivariate:
координаты (величины) + дроны (объекты)
- Panel Univariate: цены (величины)
на квартиры (объекты)

Иерархические

Australian tourism: by regions nested within states

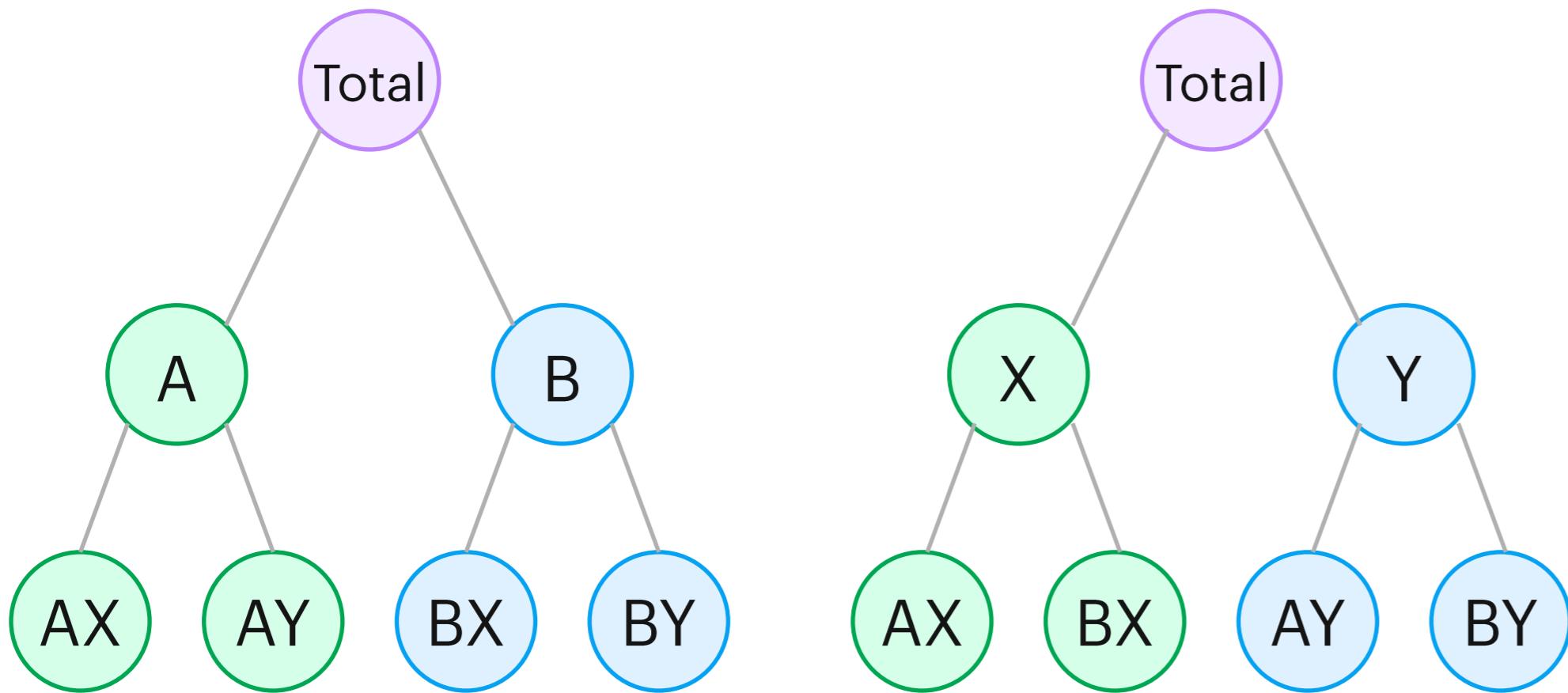


→ Ряды могут быть однозначно декомпозированы по набору факторов.

→ **Пример:** количество туристов в городах Австралии, иерархия:

- страна,
- штат,
- город.

Групповые



Иерархия (1):

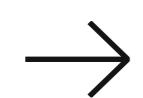
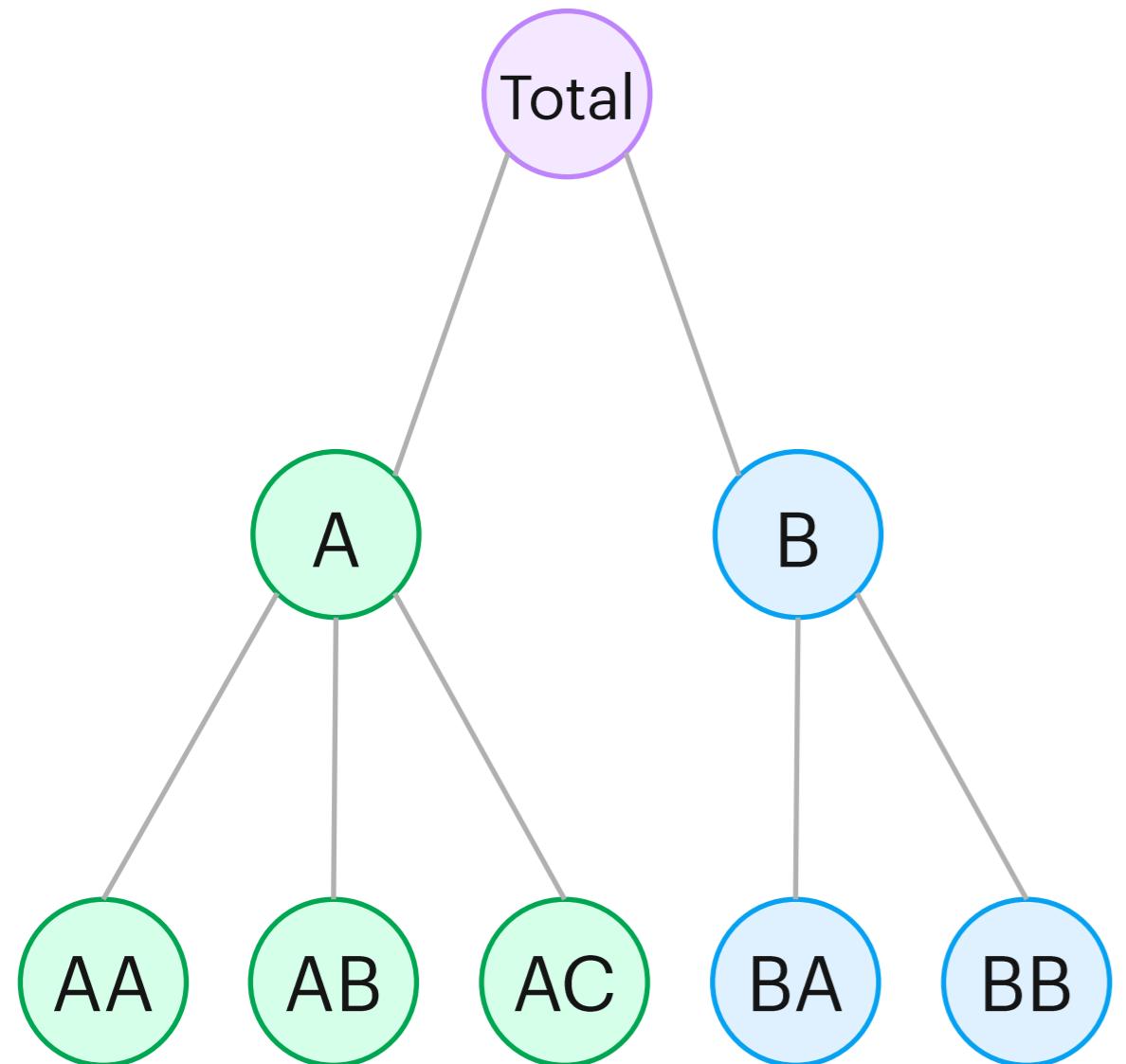
- грузоперевозчик,
- тип груза,
- станция назначения.

Иерархия (2):

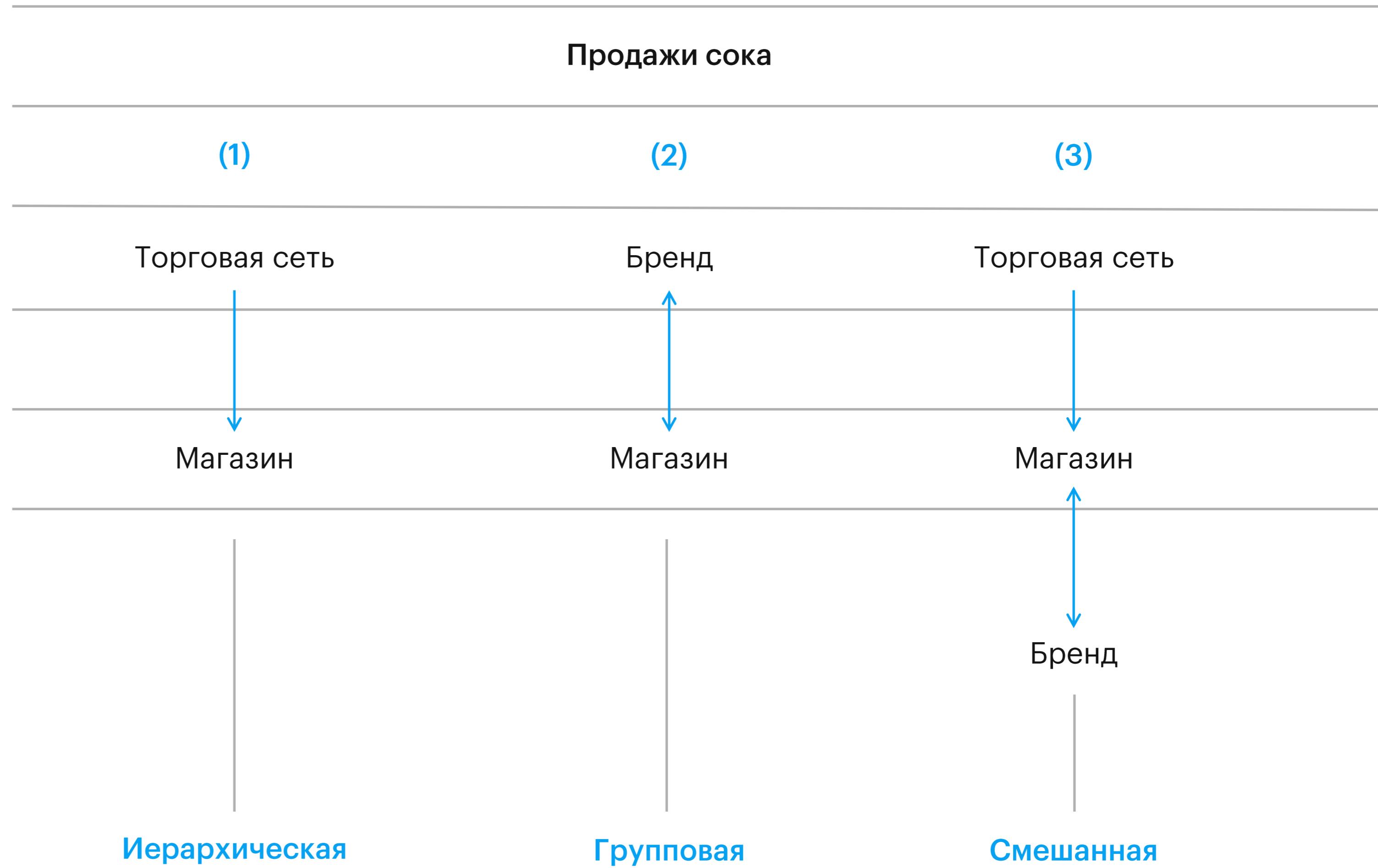
- тип груза,
- грузоперевозчик,
- станция назначения.

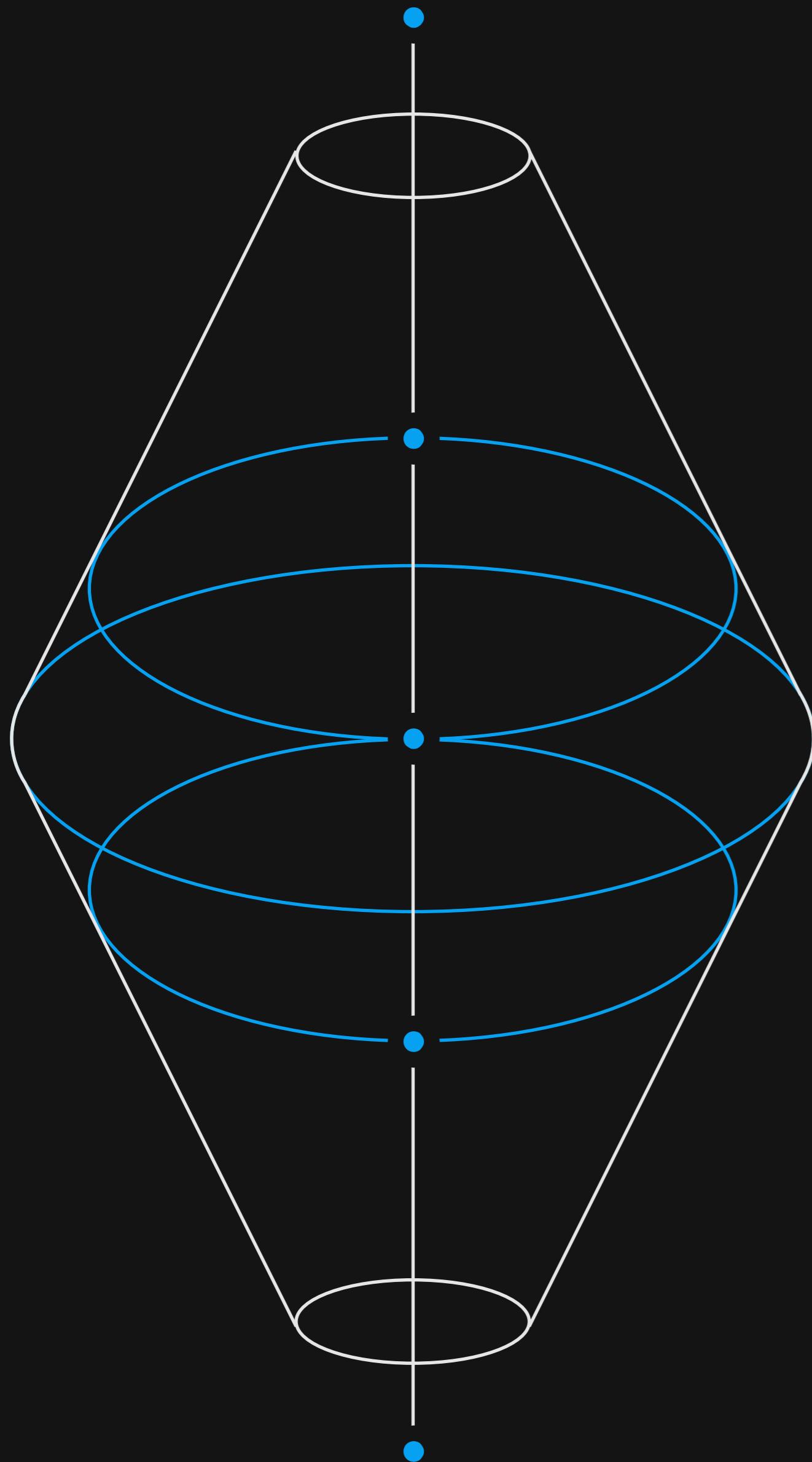
- Ряды могут быть декомпозированы по набору факторов несколькими способами.
- Пример: количество тонн перевезённого груза в день.

Смешанная



Можем переставлять уровни
в иерархии.

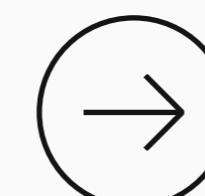




В чём суть иерархии?



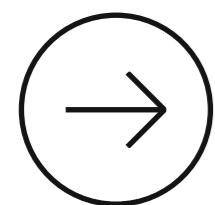
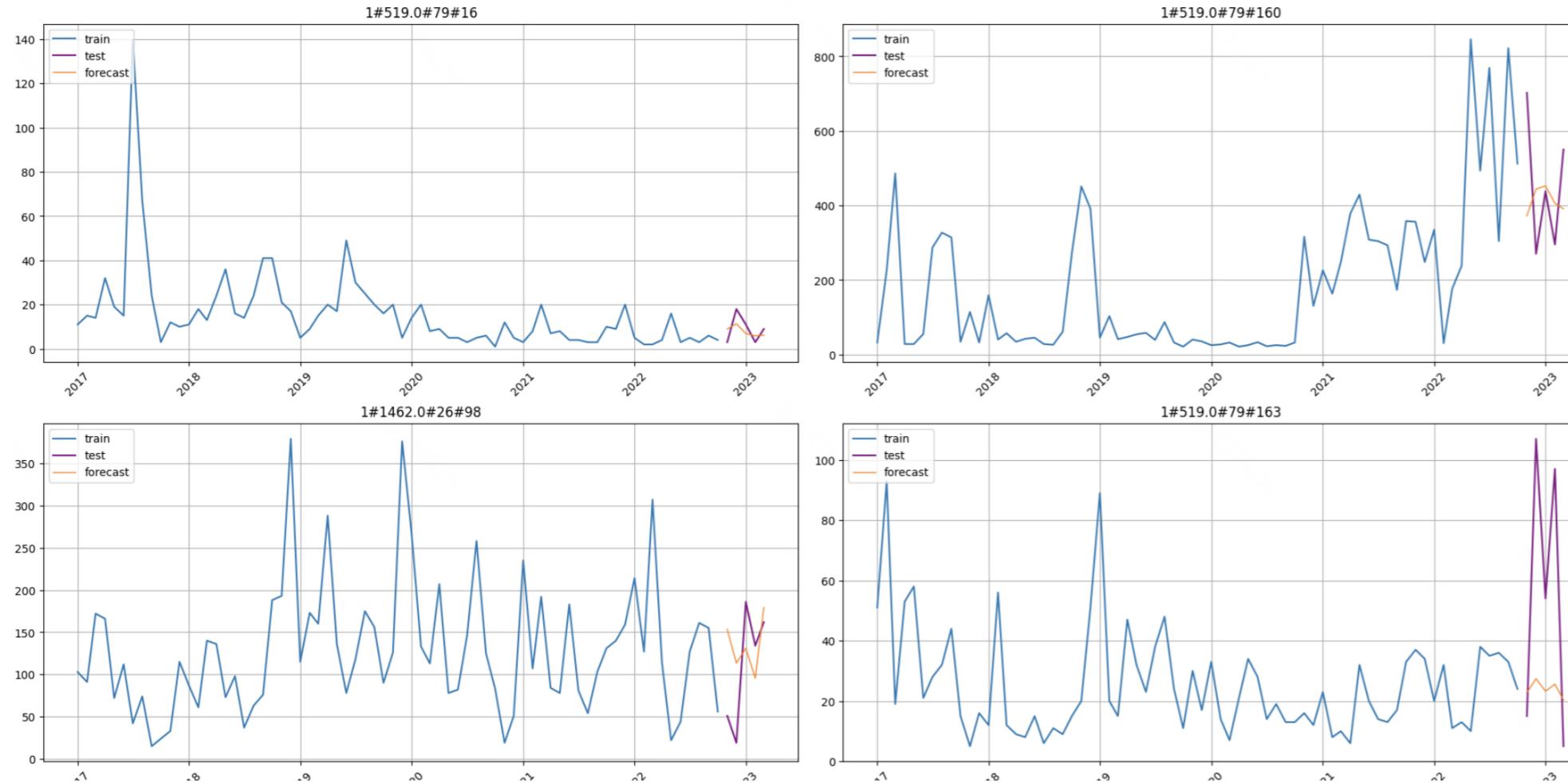
Важен прогноз на определённом уровне иерархии.



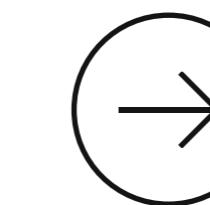
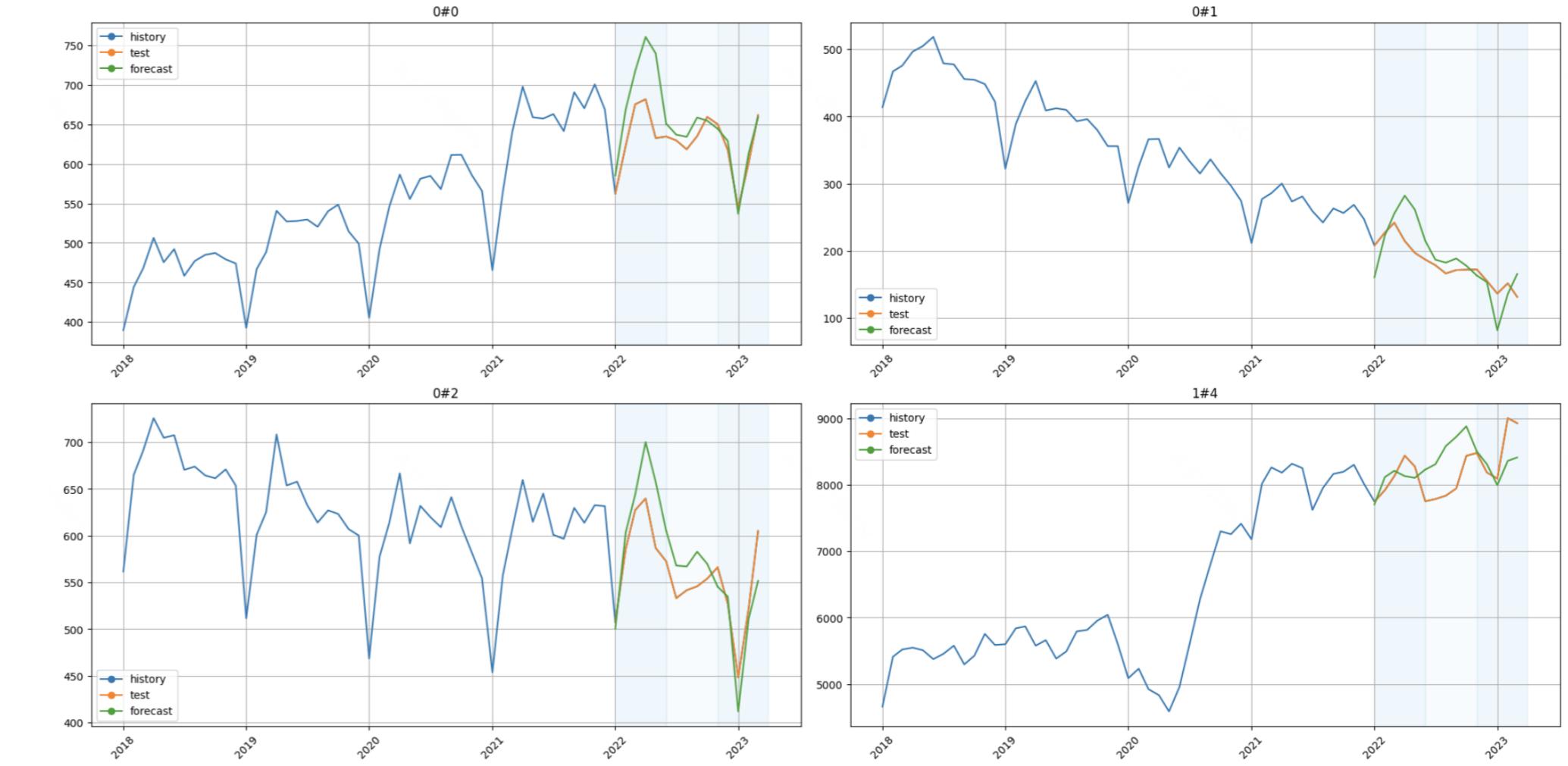
Ряды на других уровнях:

- содержат дополнительную информацию,
- лучше прогнозируются.

В чём суть иерархии?



Самый нижний уровень:
груз + перевозчик + станции
отправления/прибытия.



Верхний уровень:
тип груза.

Типы временных рядов

По частоте замеров

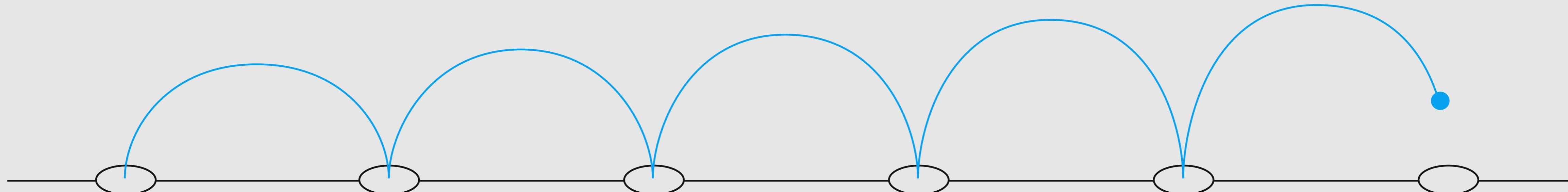
- Регулярные
- Нерегулярные

По структуре датасета

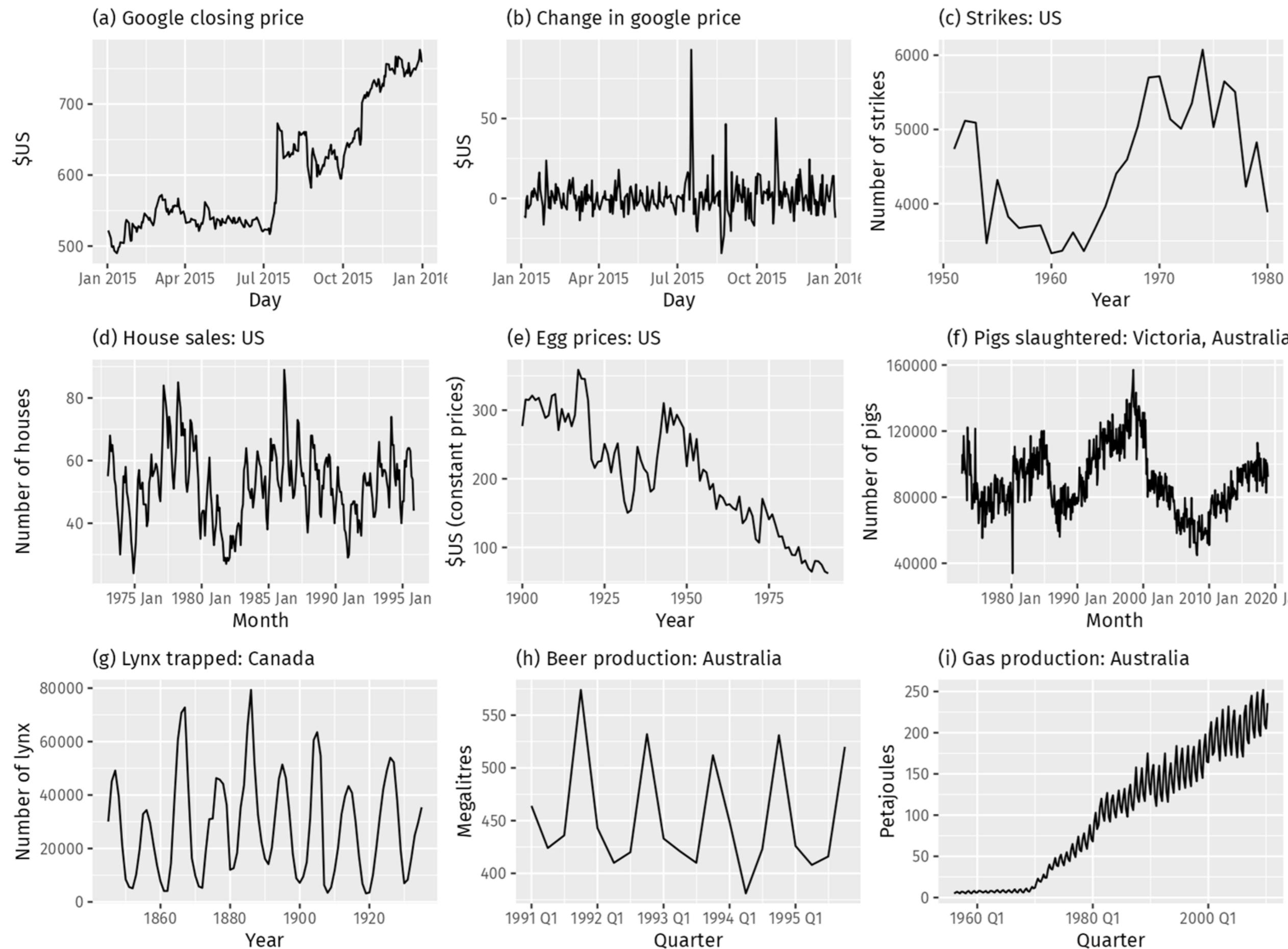
- Univariate
- Multivariate
- Panel

По наличию иерархии

- Иерархические
- Групповые
- Смешанные



Картина = главный инструмент анализа



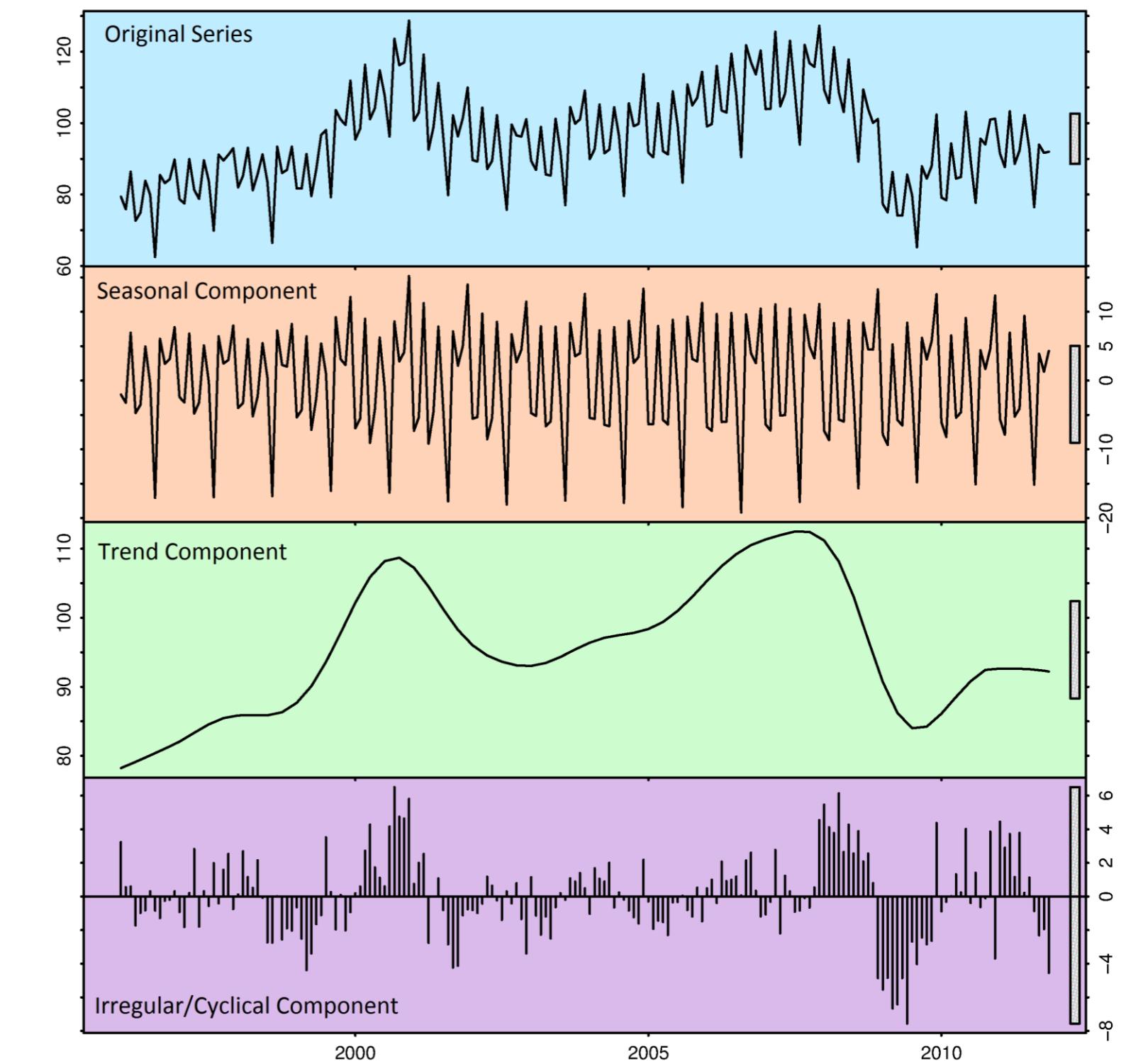
Что можно искать:

- ТИП РЯДОВ,
- КОМПОНЕНТЫ РЯДА,
- АВТОКОРРЕЛЯЦИЮ,
- СВЯЗАННОСТЬ РЯДОВ,
- АНОМАЛИИ,
- ПРАЗДНИКИ,
- ПРОПУСКИ,
- СМЕНУ РАСПРЕДЕЛЕНИЯ.

2. Компоненты временного ряда

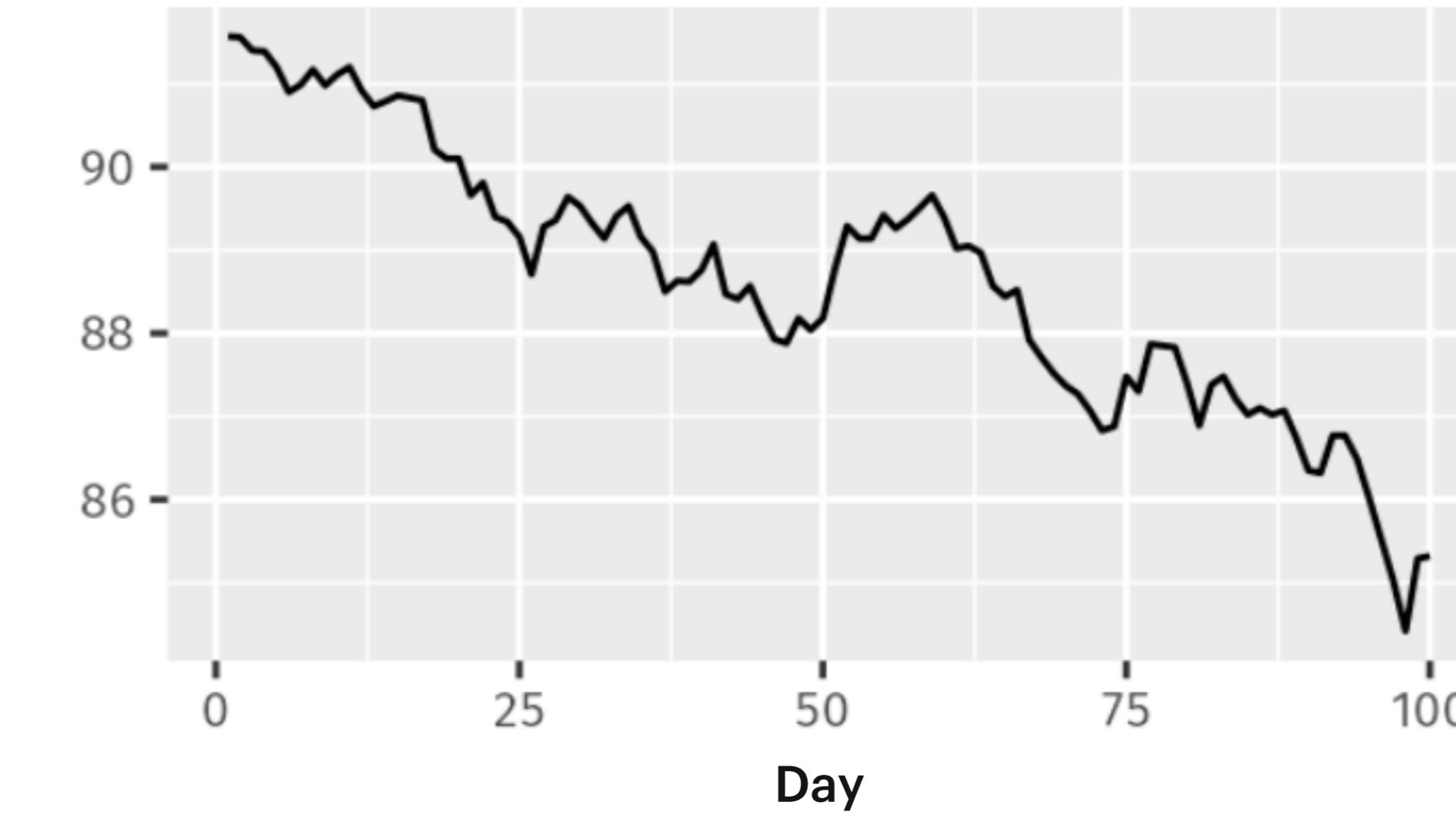
Компоненты временного ряда

- Тренд — плавное долгосрочное изменение уровня ряда.
- Сезонность — циклическое изменение уровня ряда с **постоянным** периодом.
- Цикличность — циклическое изменение уровня ряда с **переменным** периодом.
- Ошибка — непрогнозируемая случайная компонента ряда.

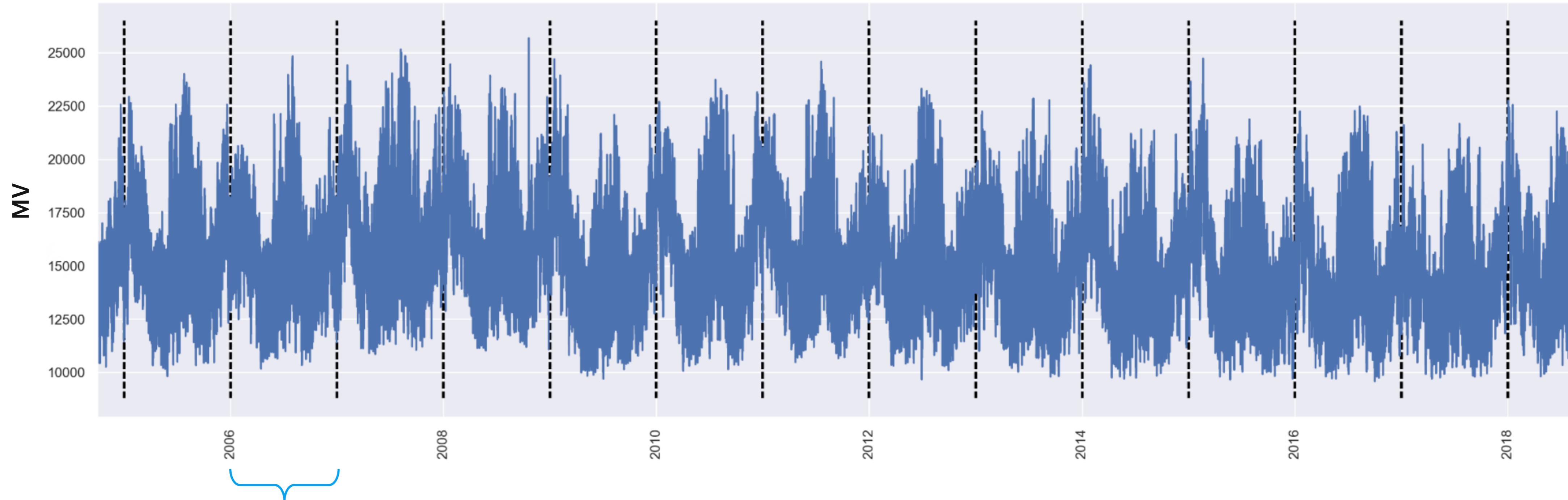


Тренд

US treasury bill contracts



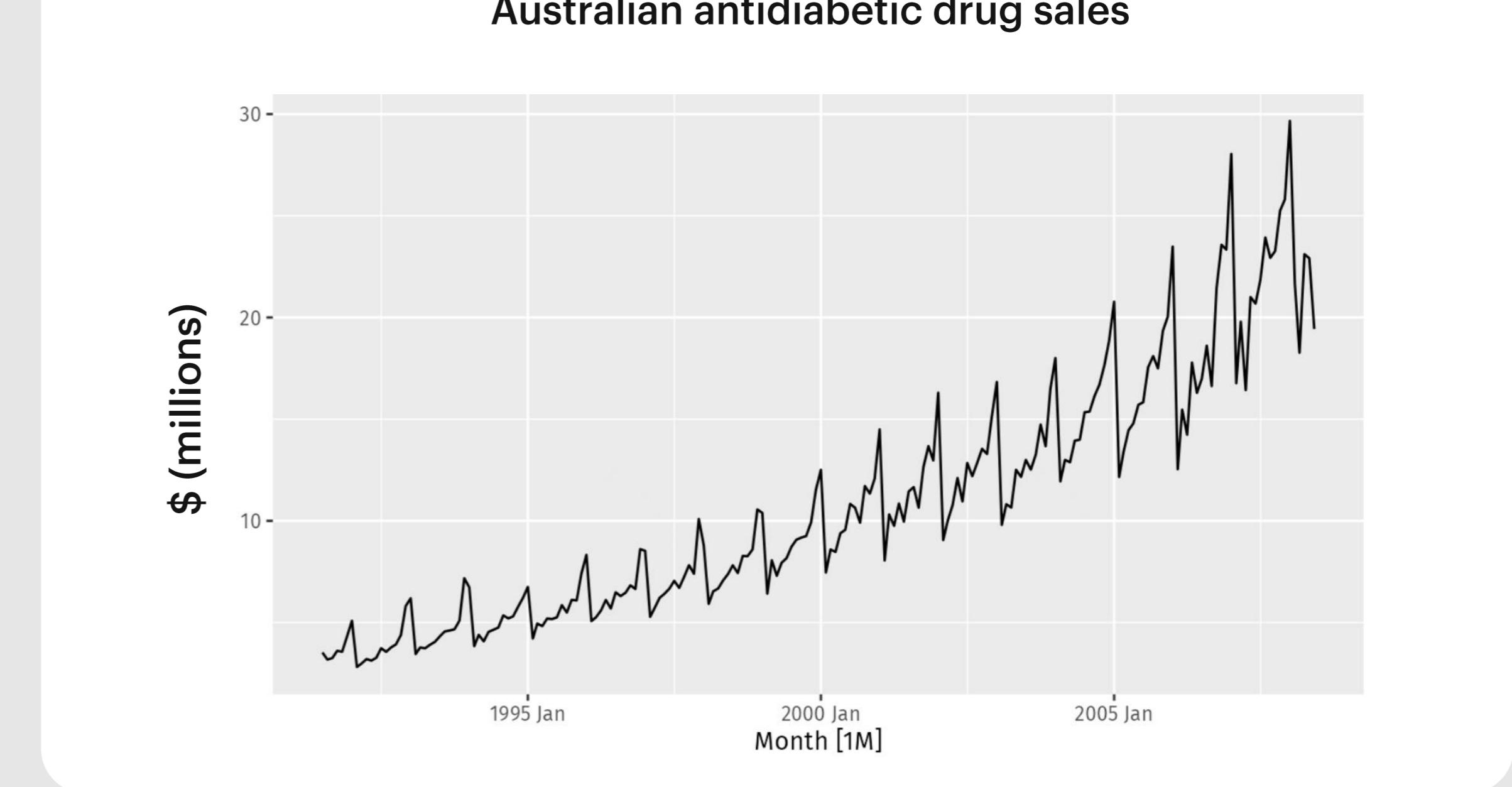
Сезонность



Период сезонности — расстояние между соответствующими повторяющимися пиками ([тут равен году](#)).

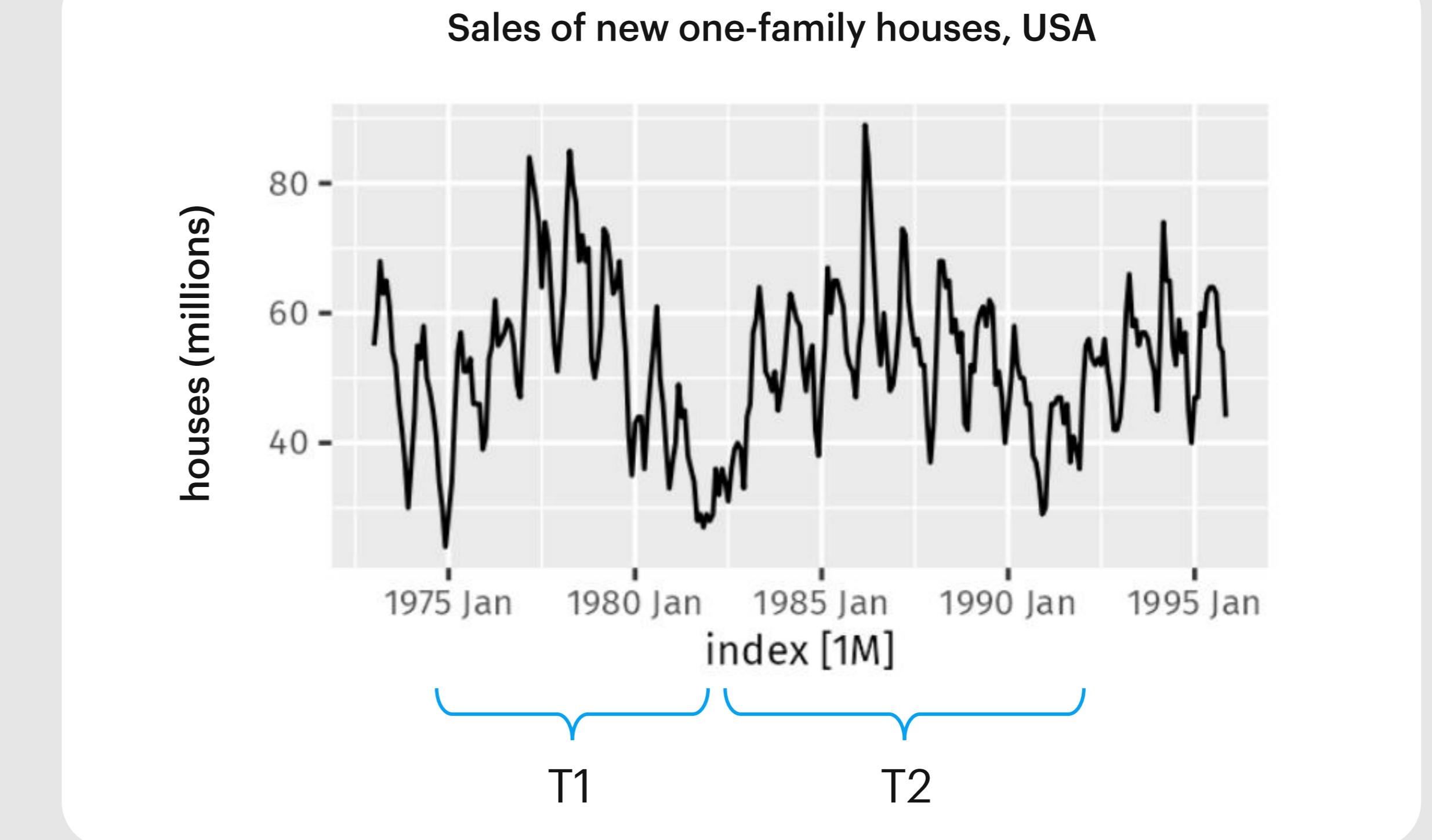
Их может быть несколько!

Тренд + сезонность



→ Паттерн повторяется, но дисперсия меняется.

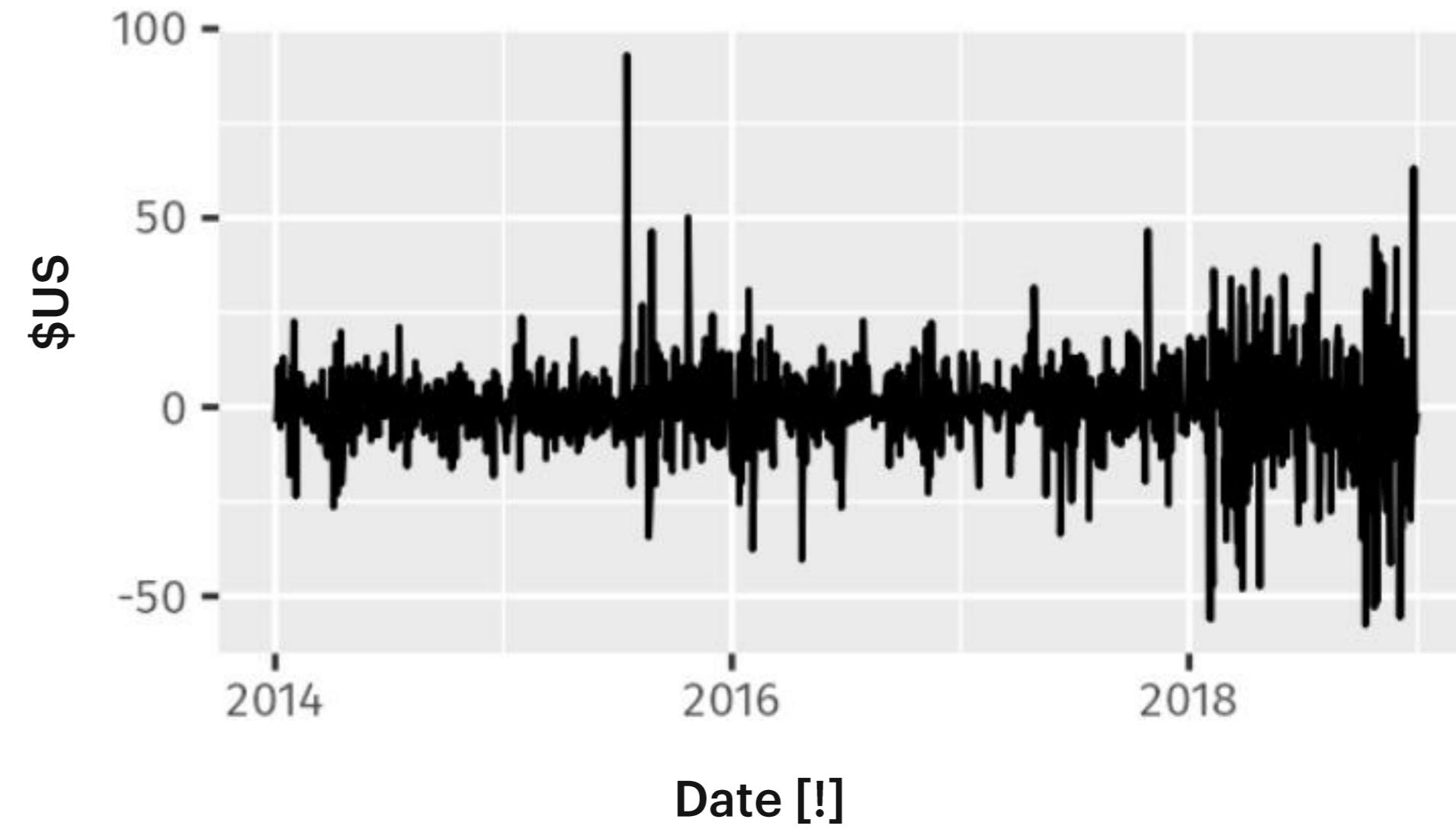
Цикличность + сезонность



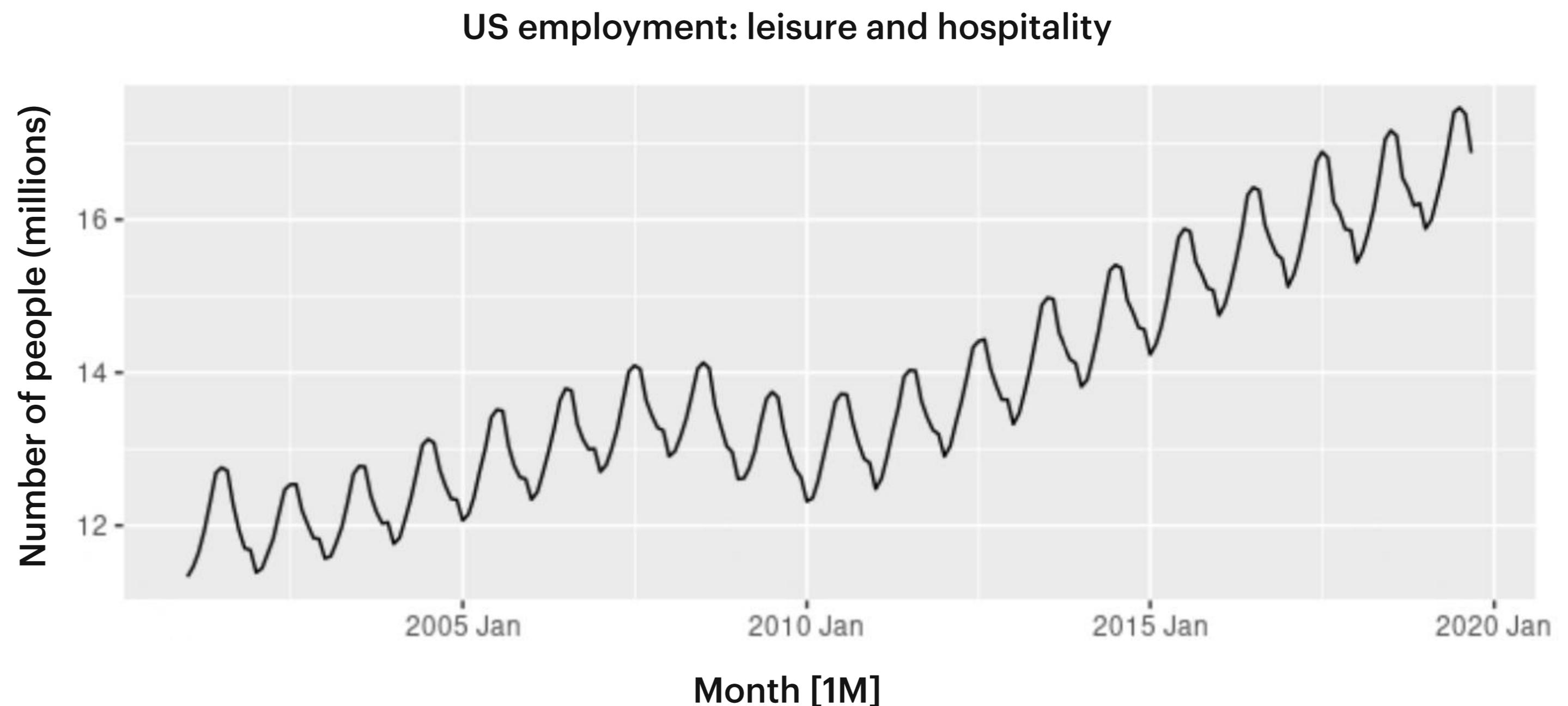
→ Цикличность — это когда $T1 \neq T2$.

Ничего

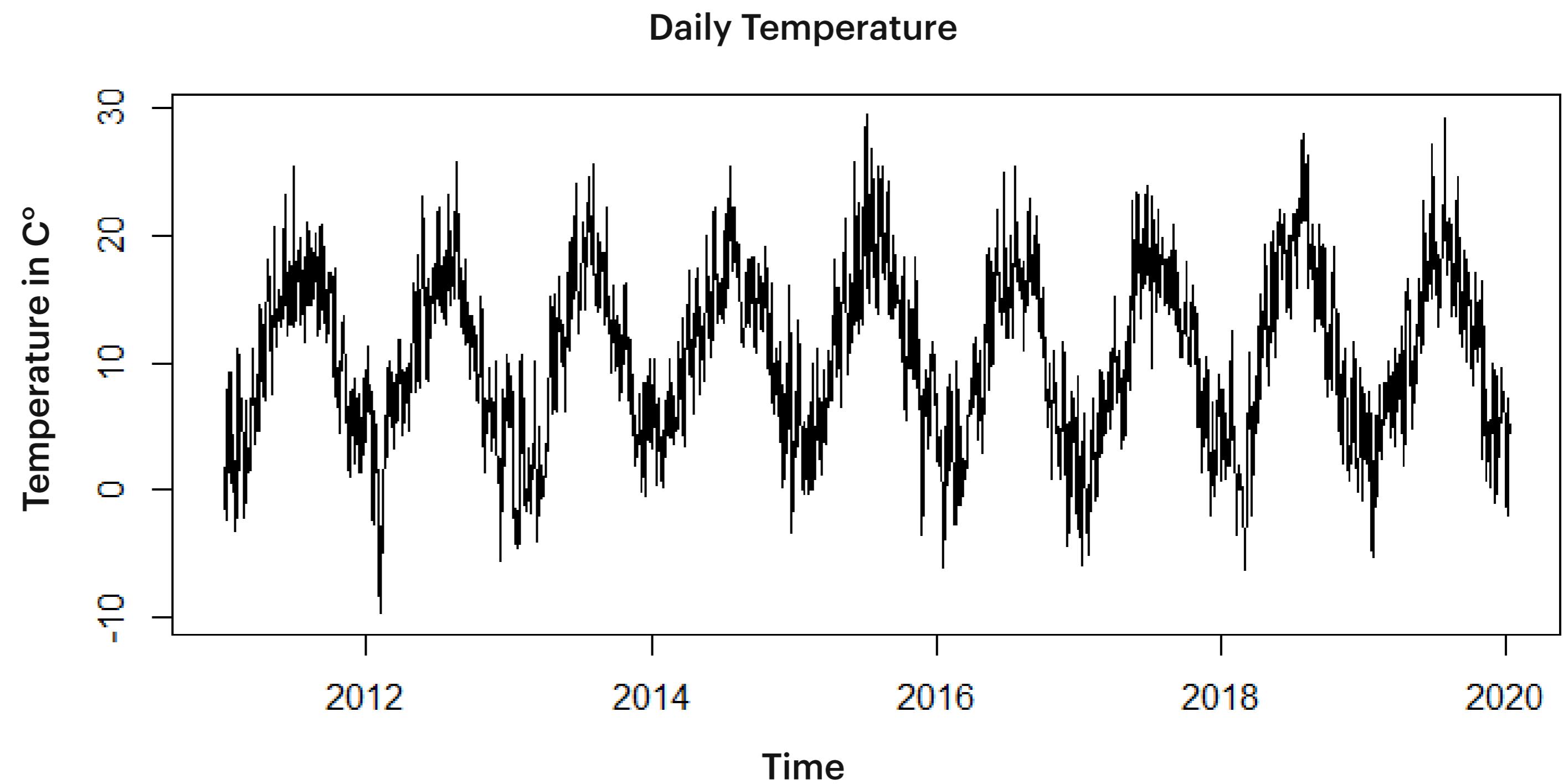
Daily changes in Google closing stock price



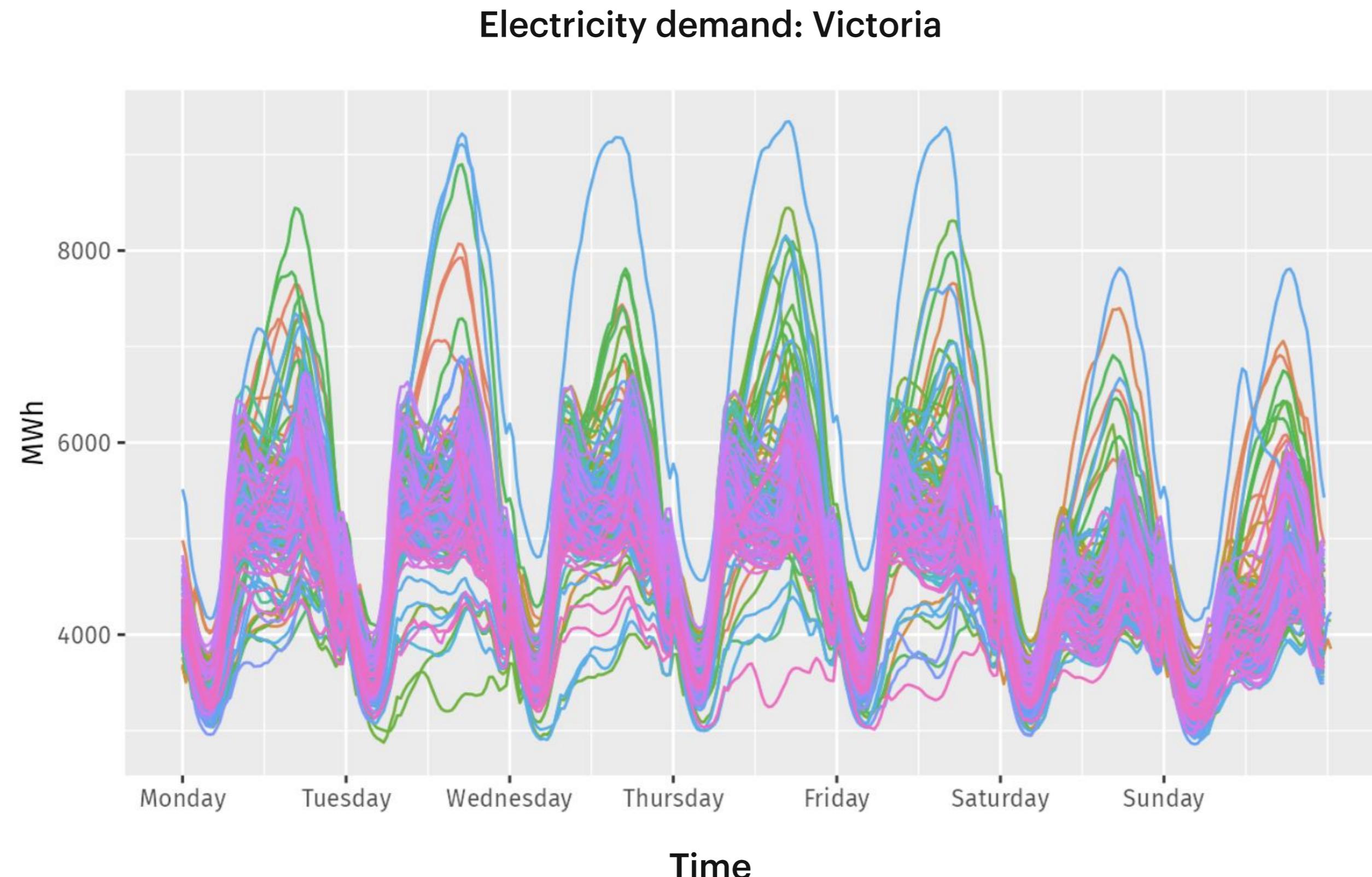
Что тут?



Что тут?



Seasonal plot



Что это?

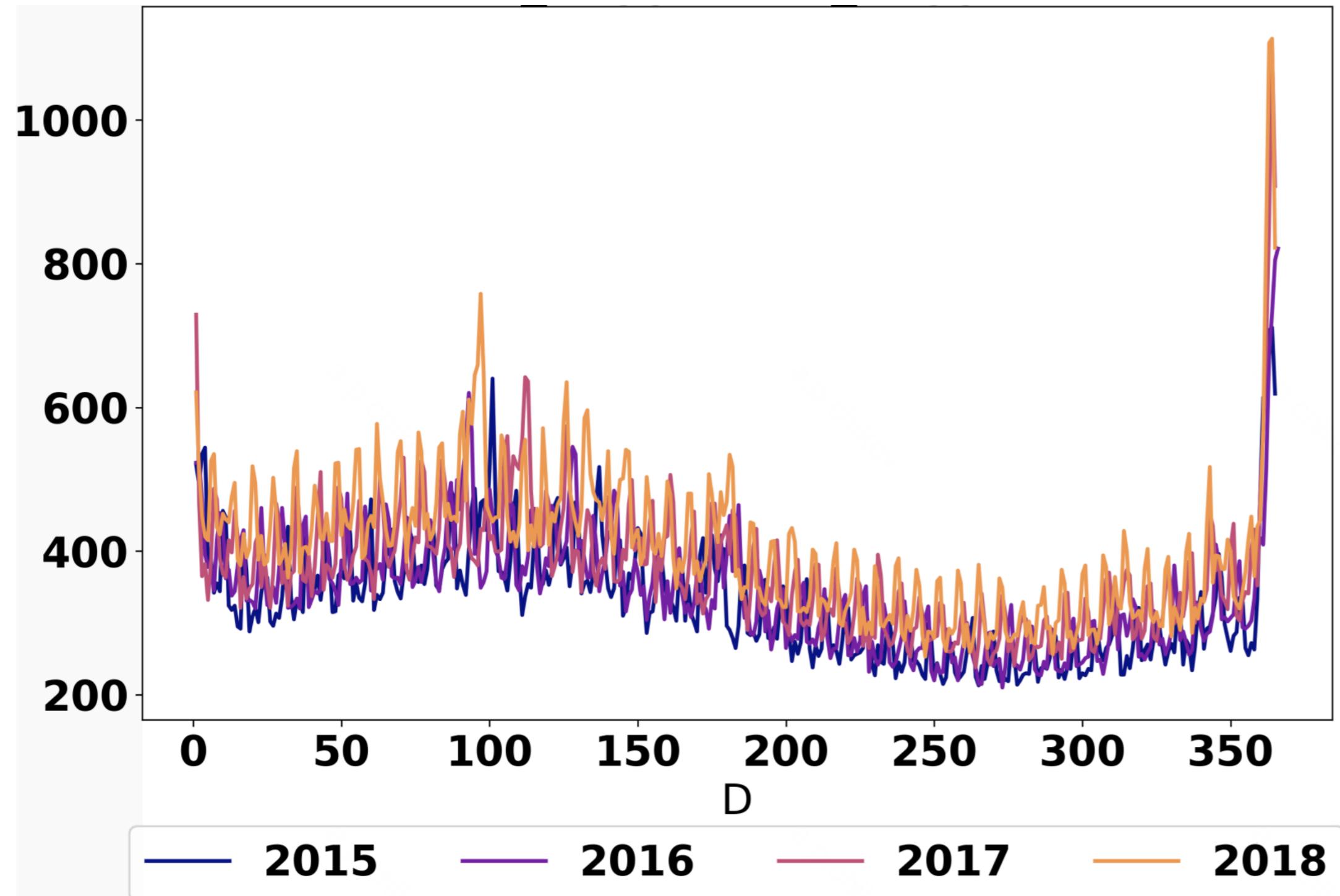
Накладываем отрезки ряда, равные периоду сезонности.

Зачем?

- Проверяем, что сезонность есть.
- Ловим смену паттерна.
- Изучаем паттерны для разных рядов.

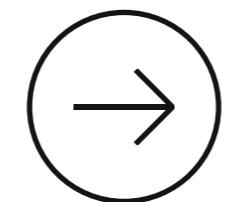
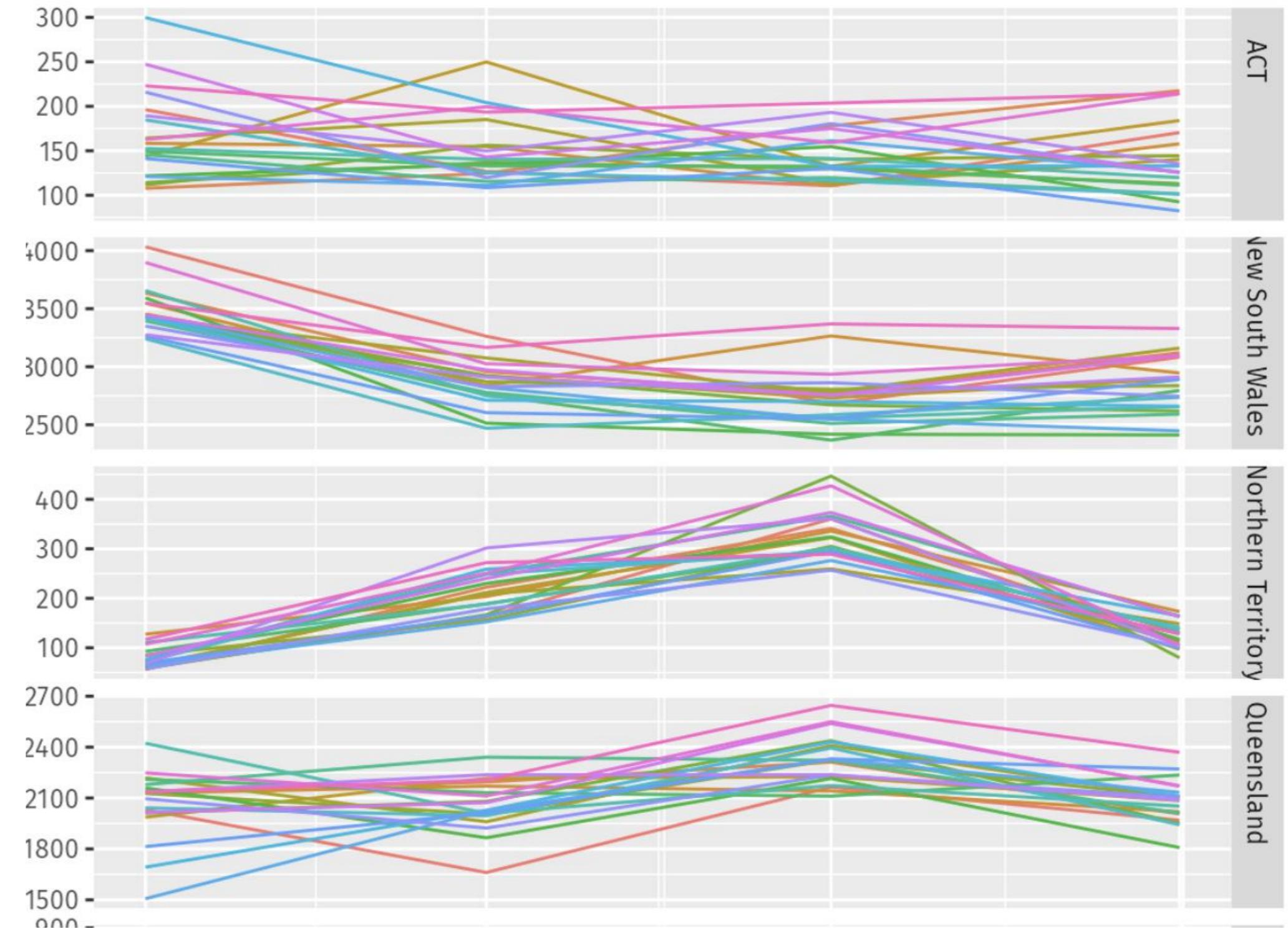
Что видим?

Finland_KaggleMart_Kaggle Hat



1. Значения с каждым годом всё больше —> есть тренд.
2. Паттерн немного съезжает в 2018 —> високосный год.

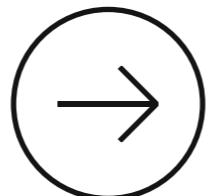
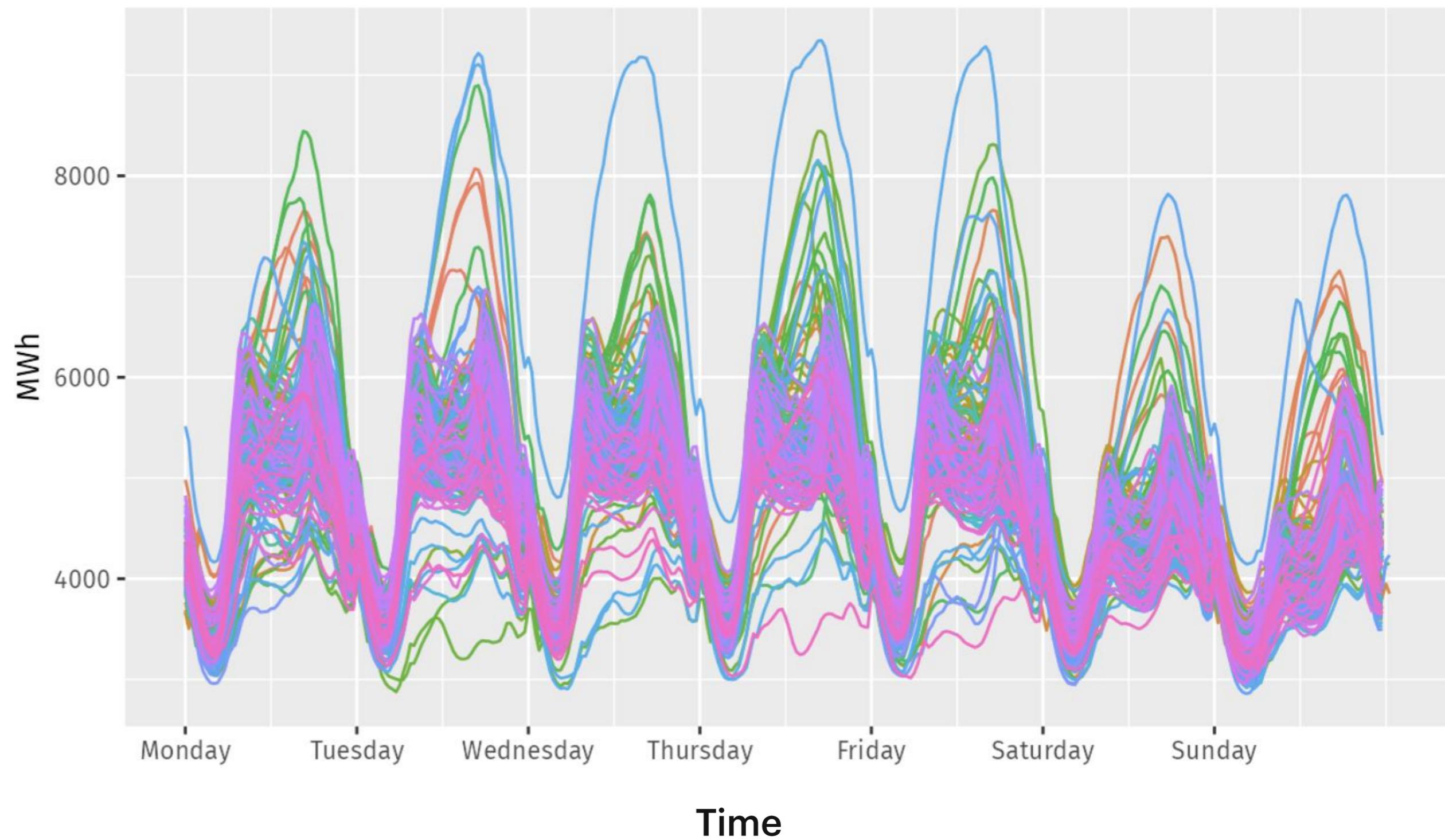
Что видим?



Разный паттерн недельной сезонности —
для разных рядов.

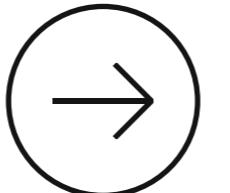
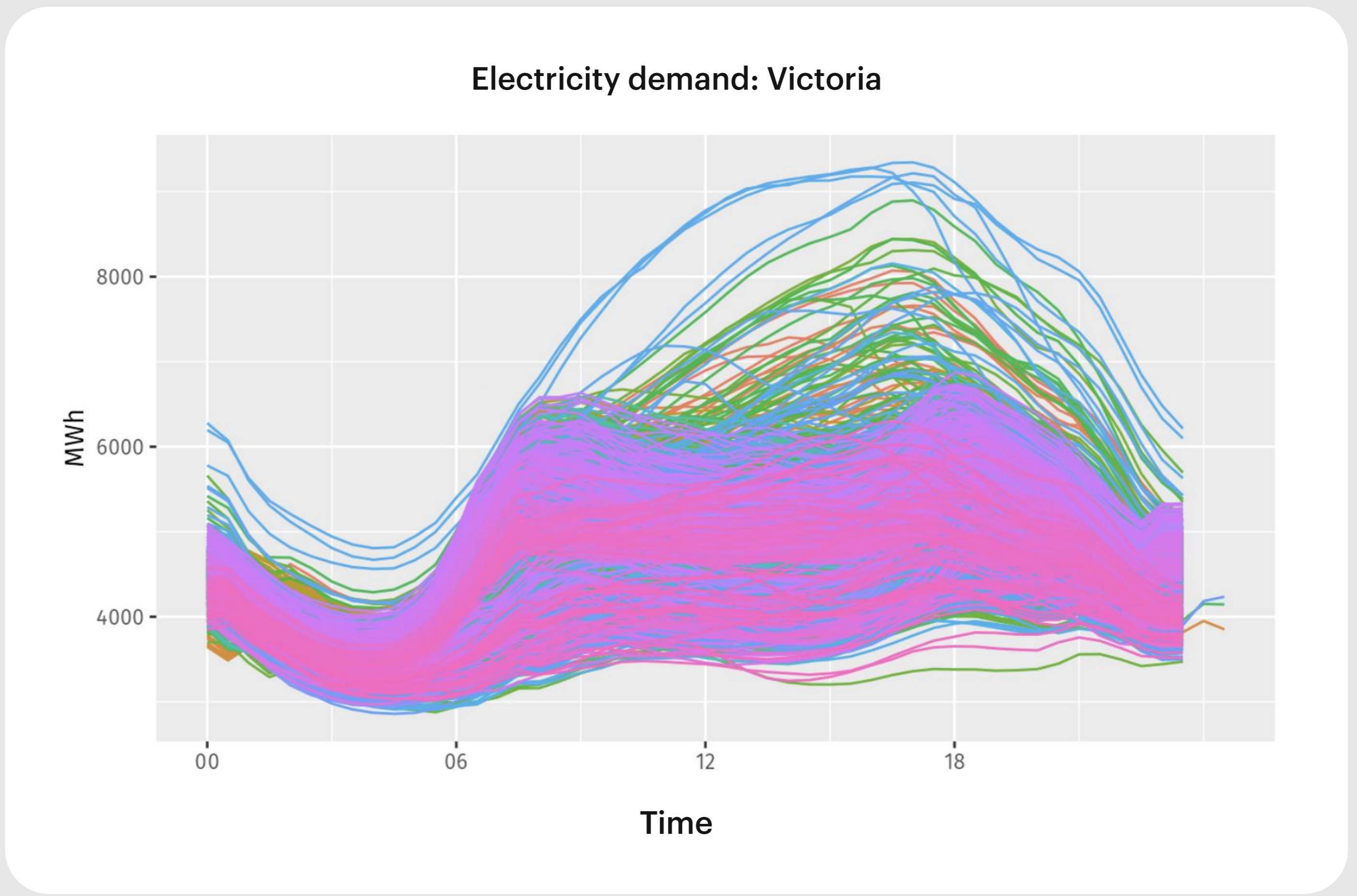
Что видим?

Electricity demand: Victoria



У ряда несколько сезонностей —
недельная + дневная.

Что видим?



Потребление ночью ниже, чем потребление днём.

3. Автокорреляция

Классический ML

01

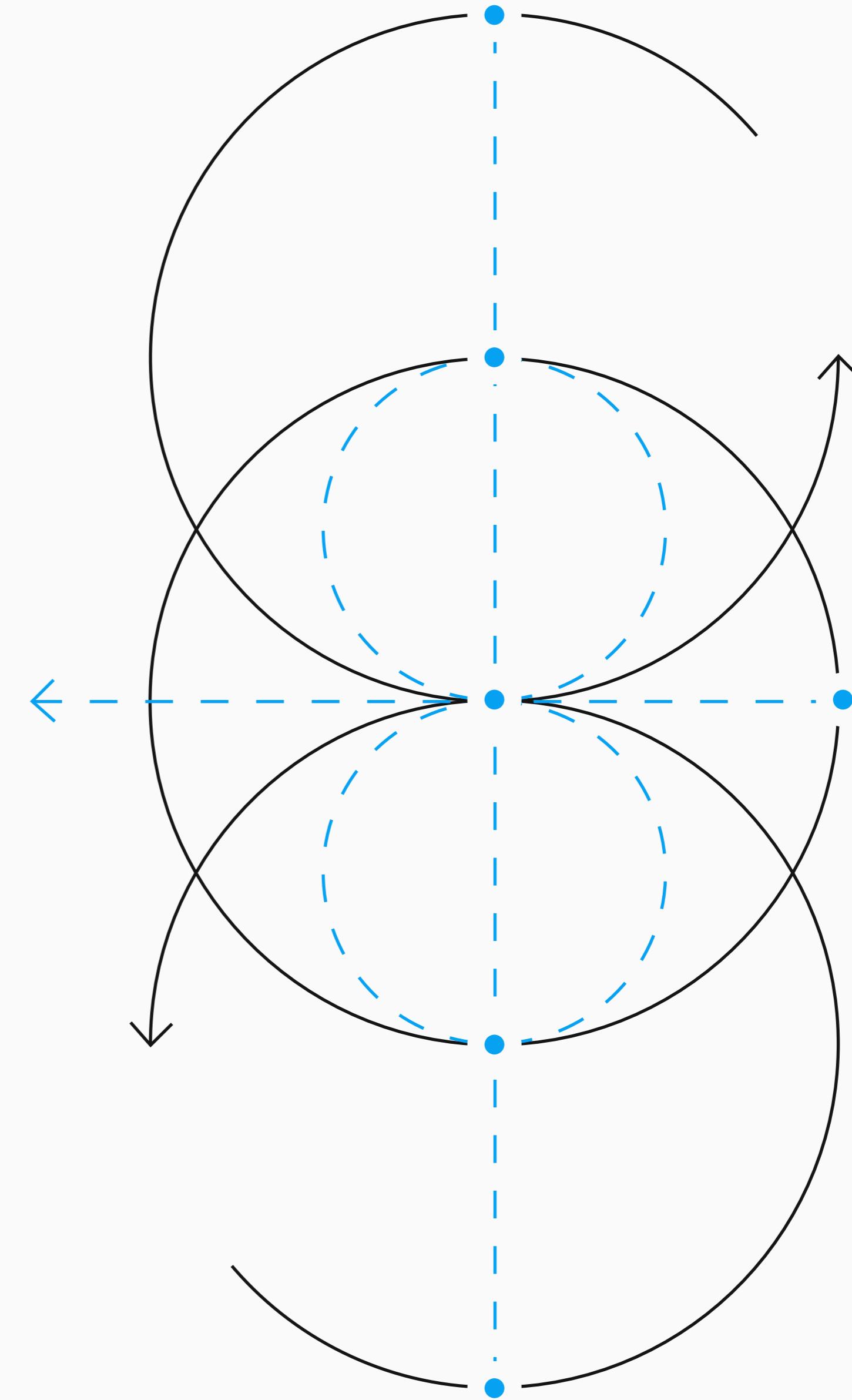
Выборка состоит из независимых наблюдений (Y):

$$y_1, y_2, \dots, y_n \sim IID.$$

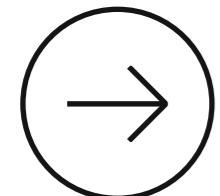
02

Значения наблюдений могут зависеть от других переменных (X):

$$(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n) \sim IID.$$

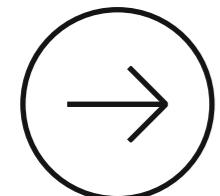


Автокорреляция



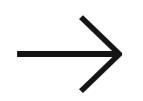
Корреляция — степень линейной связи между двумя переменными (X и Y).

$$\mathbf{r}_{XY} = \frac{\mathbf{cov}_{XY}}{\sigma_X \sigma_Y} = \frac{\sum (X - \bar{X})(Y - \bar{Y})}{\sqrt{\sum (X - \bar{X})^2 \sum (Y - \bar{Y})^2}}$$



Автокорреляция (ACF(k)) — степень линейной связи между запаздывающими значениями временного ряда ($X = Y(t)$, $Y = Y(t - k)$).

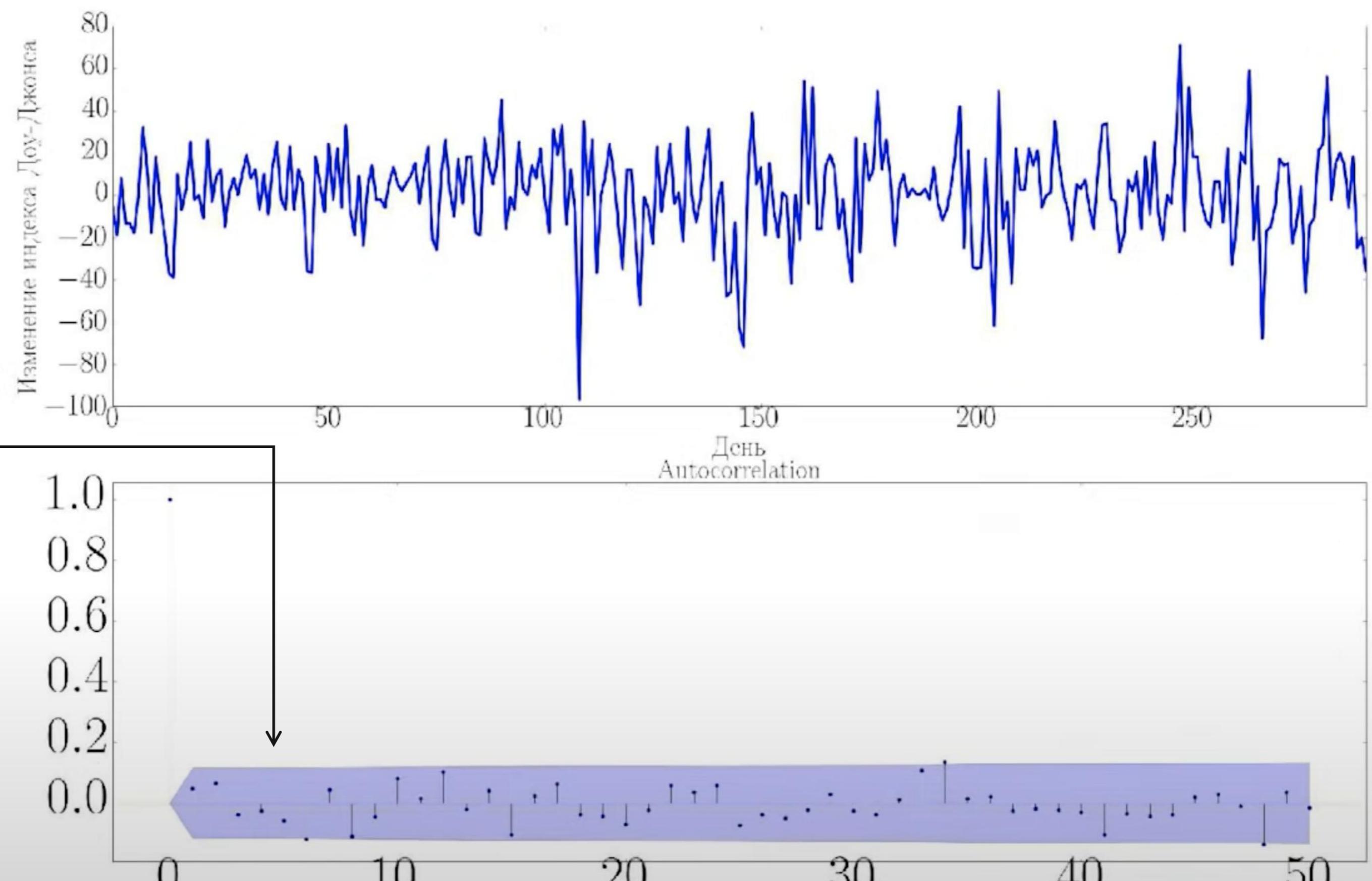
$$c_k = \frac{1}{n} \sum_{t=k+1}^n (y_t - \bar{y})(y_{t-k} - \bar{y}),$$
$$r_k = \frac{c_k}{c_0}$$



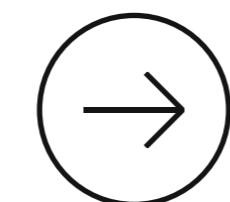
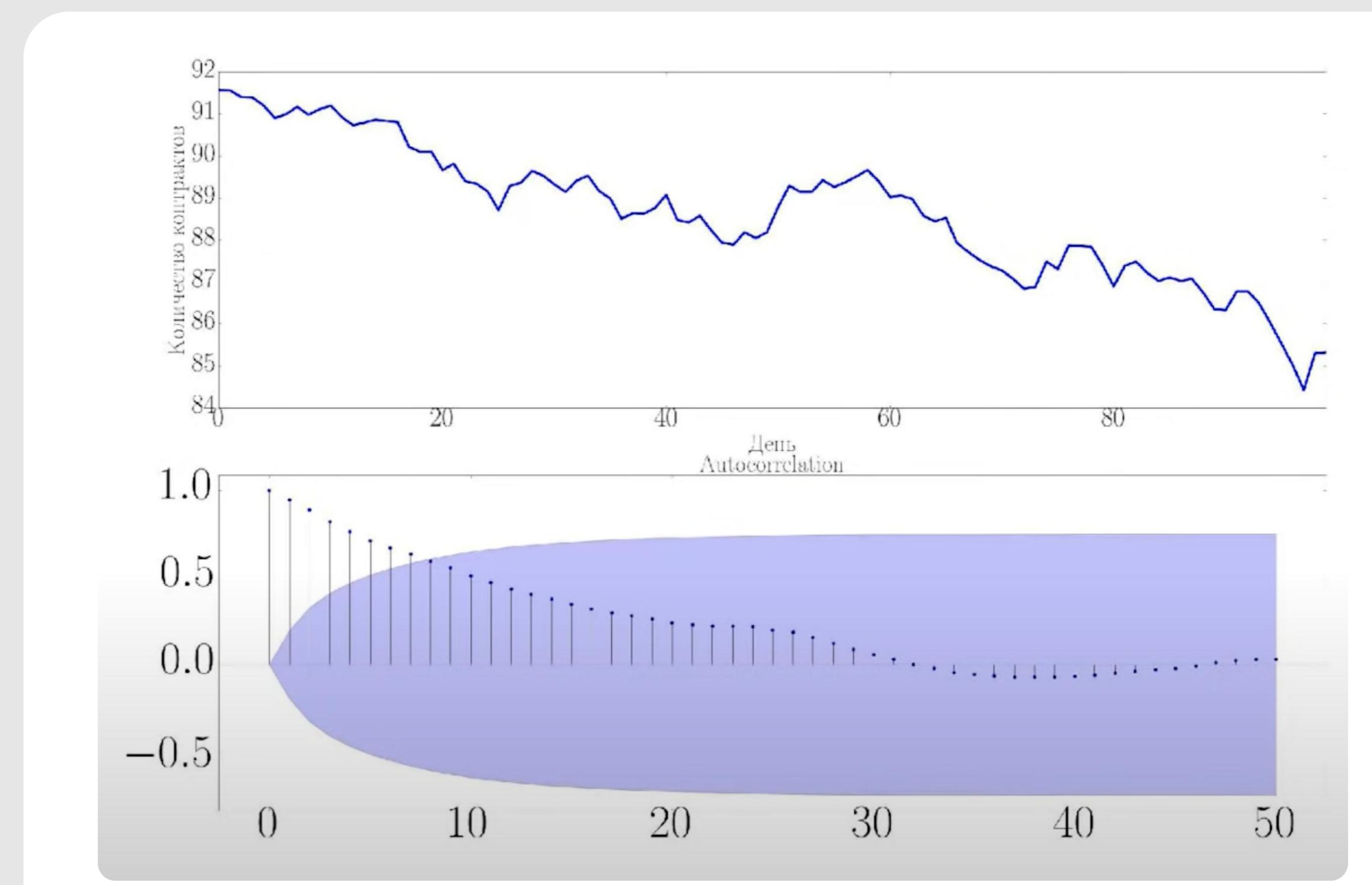
Значение $y_{(t - k)}$ называется **лагом порядка k** .

ACF (ничего)

Гипотеза о равенстве
автокорреляции нулю

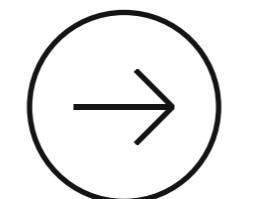
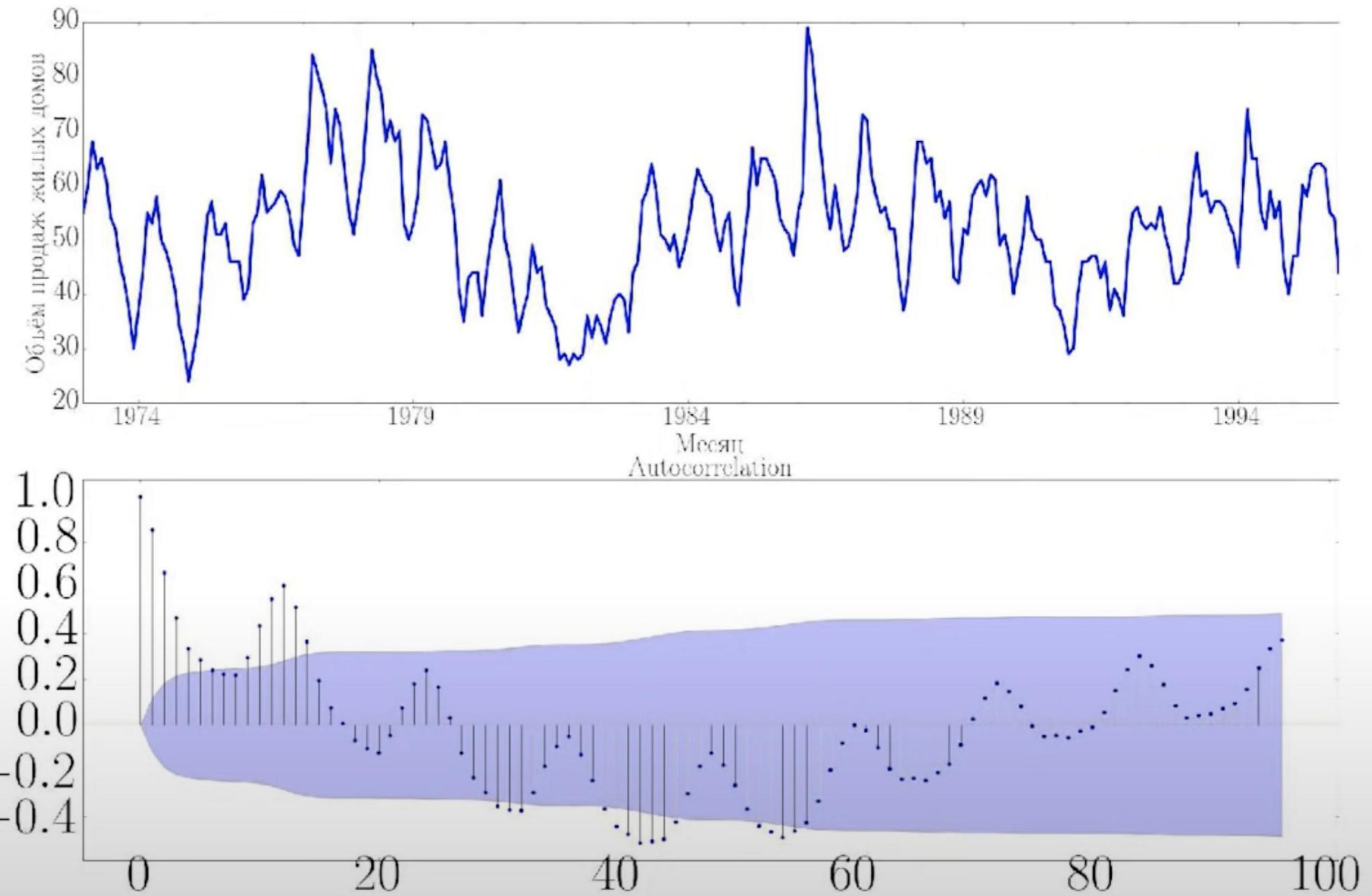


ACF (тренд)



Автокорреляции для малых лагов большие и положительные.

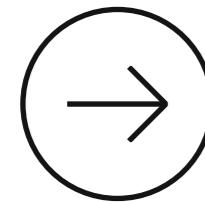
ACF (сезонность)



Автокорреляции больше на сезонных лагах
(т. е. на частотах, кратных сезону).

4. Понятие стационарности

Стационарность

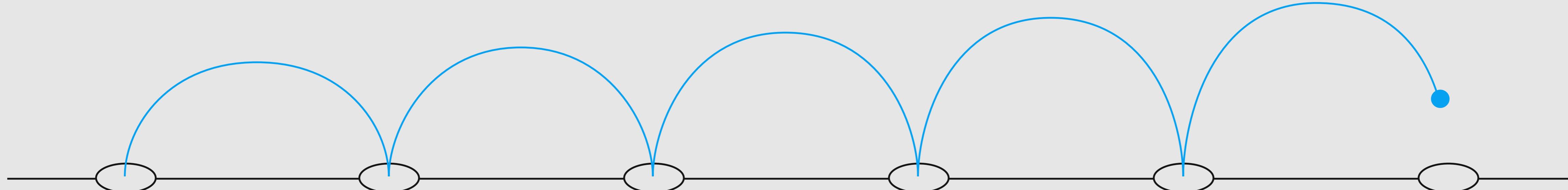


Временной ряд является **стационарным**, если его статистические характеристики не меняются во времени.

$$E[y_t] = \mu$$

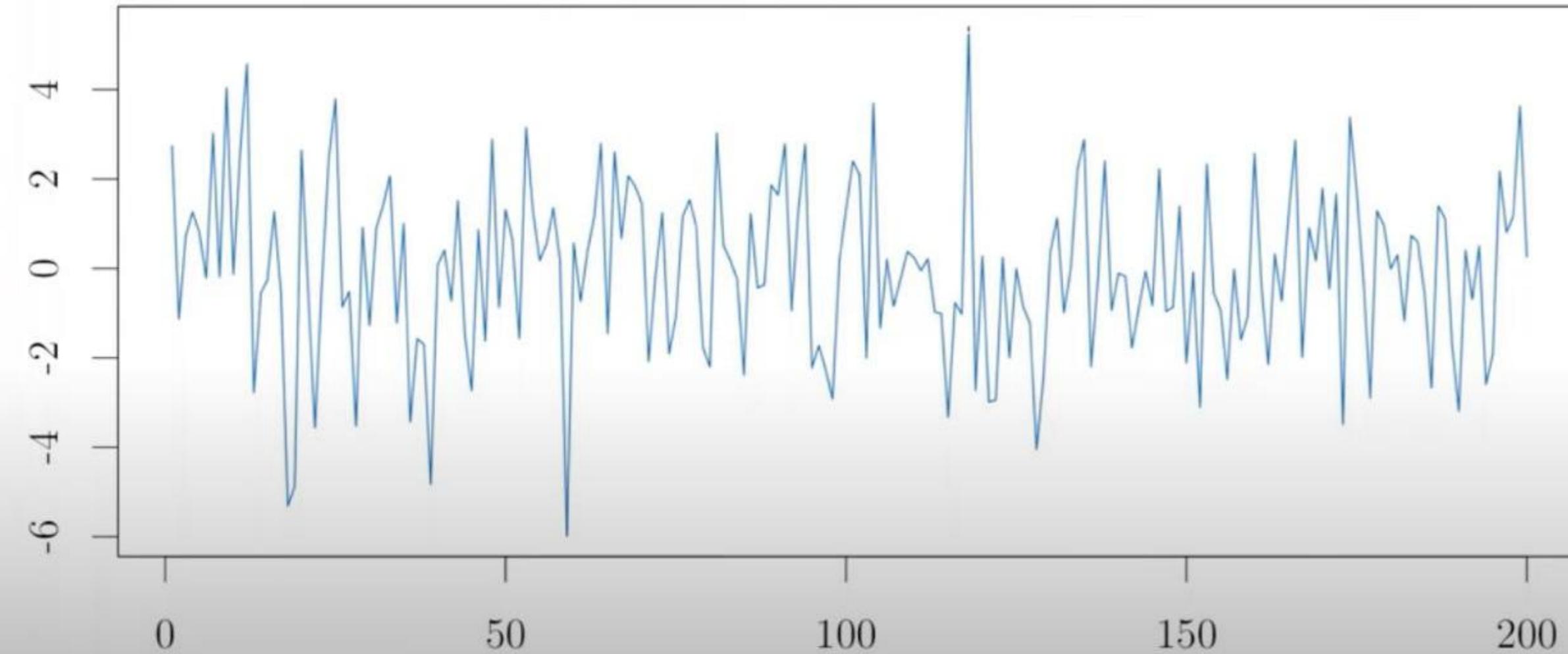
$$D[y_t] = \sigma^2$$

$$Cov(y_t, y_{t+k}) = \gamma_k$$



Временной ряд = случайный процесс

Белый шум, $y_t = \varepsilon_t \sim N(0, 4)$



Стационарный

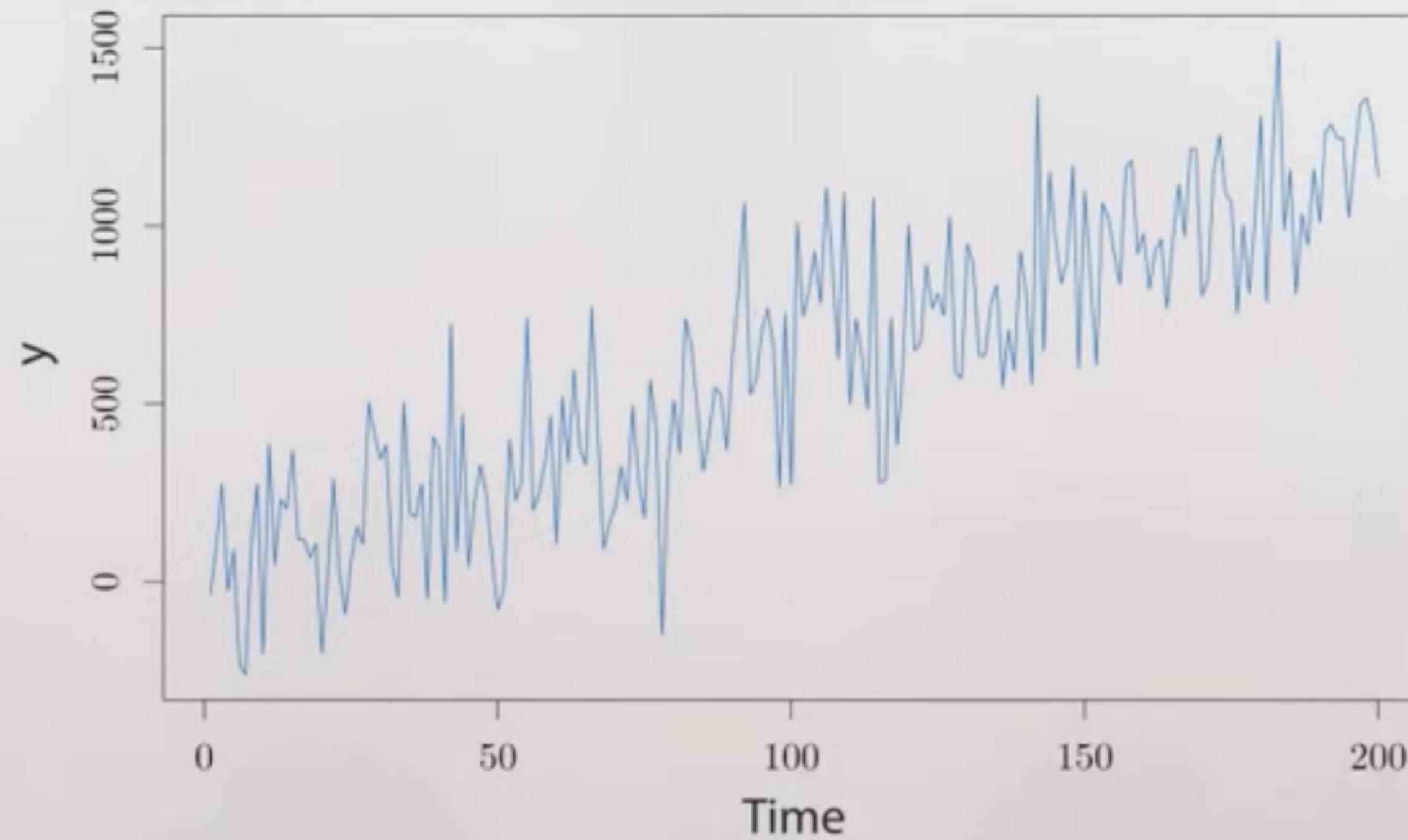
$$E[y_t] = 0$$

$$\text{Var}[y_t] = 4$$

$$\text{Cov}[y_t, y_{t+k}] = 0$$

Временной ряд = случайный процесс

Процесс с трендом, $y_t = 5 + 6t + \varepsilon_t$, $\varepsilon_t \sim N(0, 200^2)$

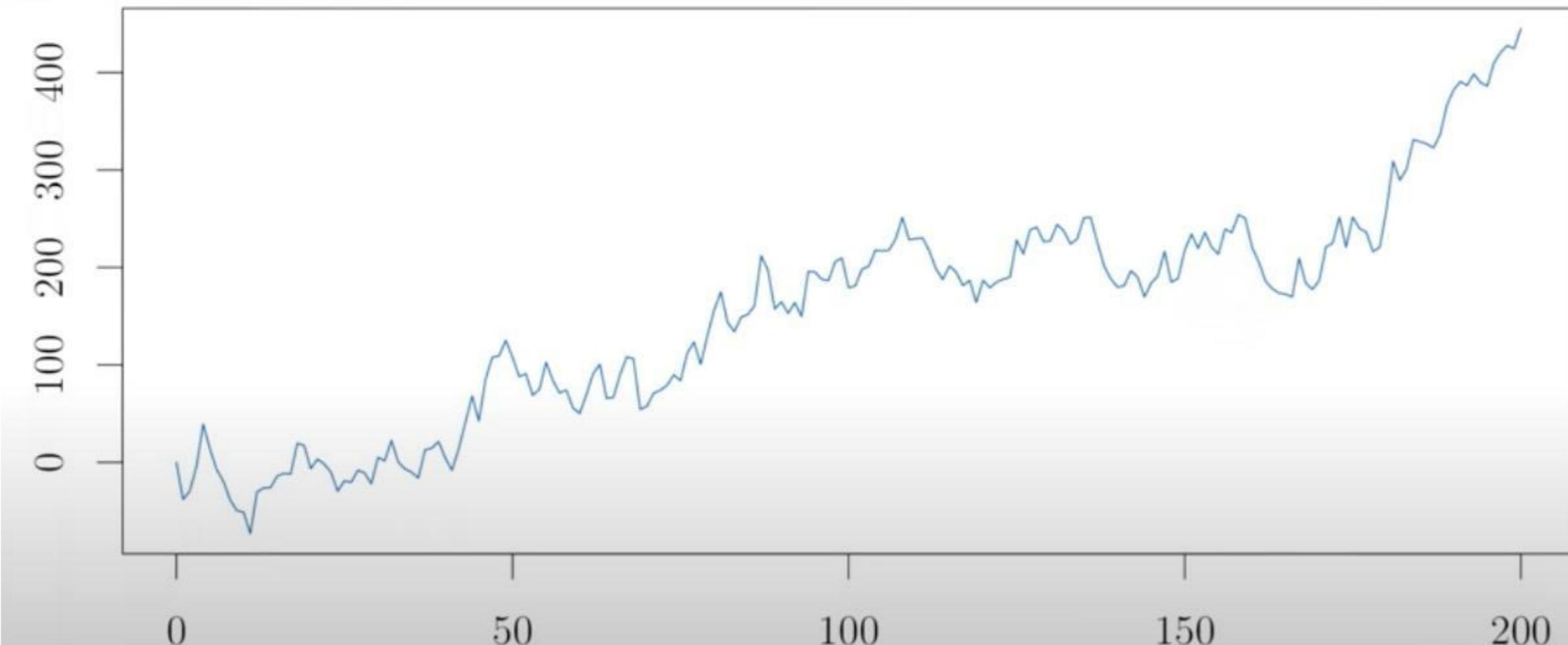


Нестационарный

$$E[y_t] = 5 + 6t$$

Временной ряд = случайный процесс

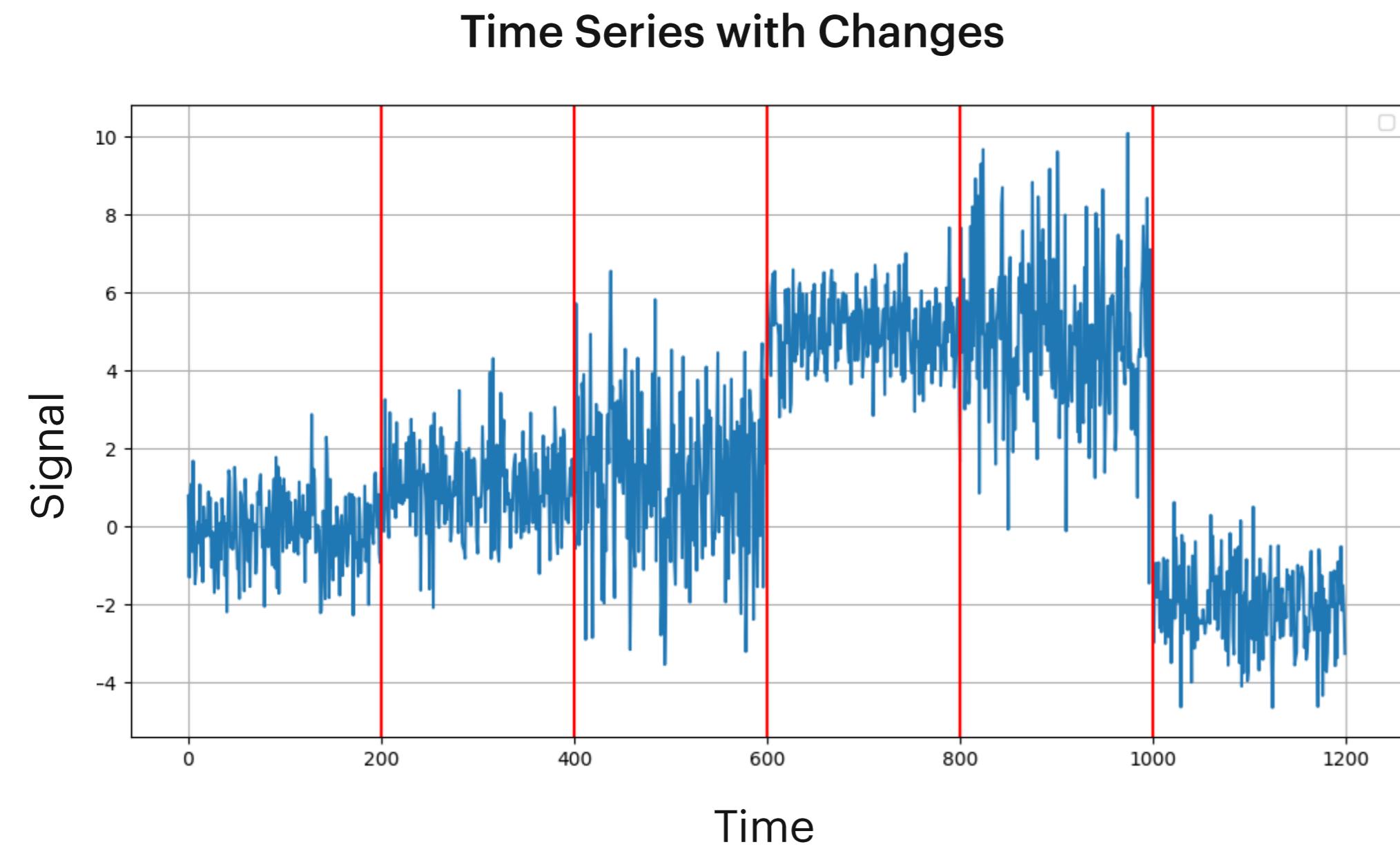
Случайное блуждание, $y_t = y_{t-1} + 2 + \varepsilon_t$



Нестационарный

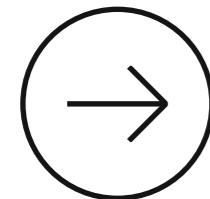
$$\text{Var}[y_t] = t\sigma^2$$

На практике



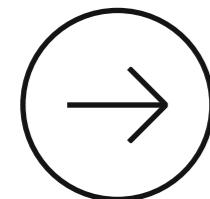
- Изначально не знаем порождающую модель.
- Признаки нестационарности:
 - тренд,
 - сезонность,
 - смена уровня,
 - изменение дисперсии и др.
- Можно использовать статистические тесты.
- Нестационарность затрудняет моделирование.

Статистические тесты



Тест Дики – Фуллера

нулевая гипотеза: H_0 : ряд нестационарен;
альтернатива: H_1 : ряд стационарен;

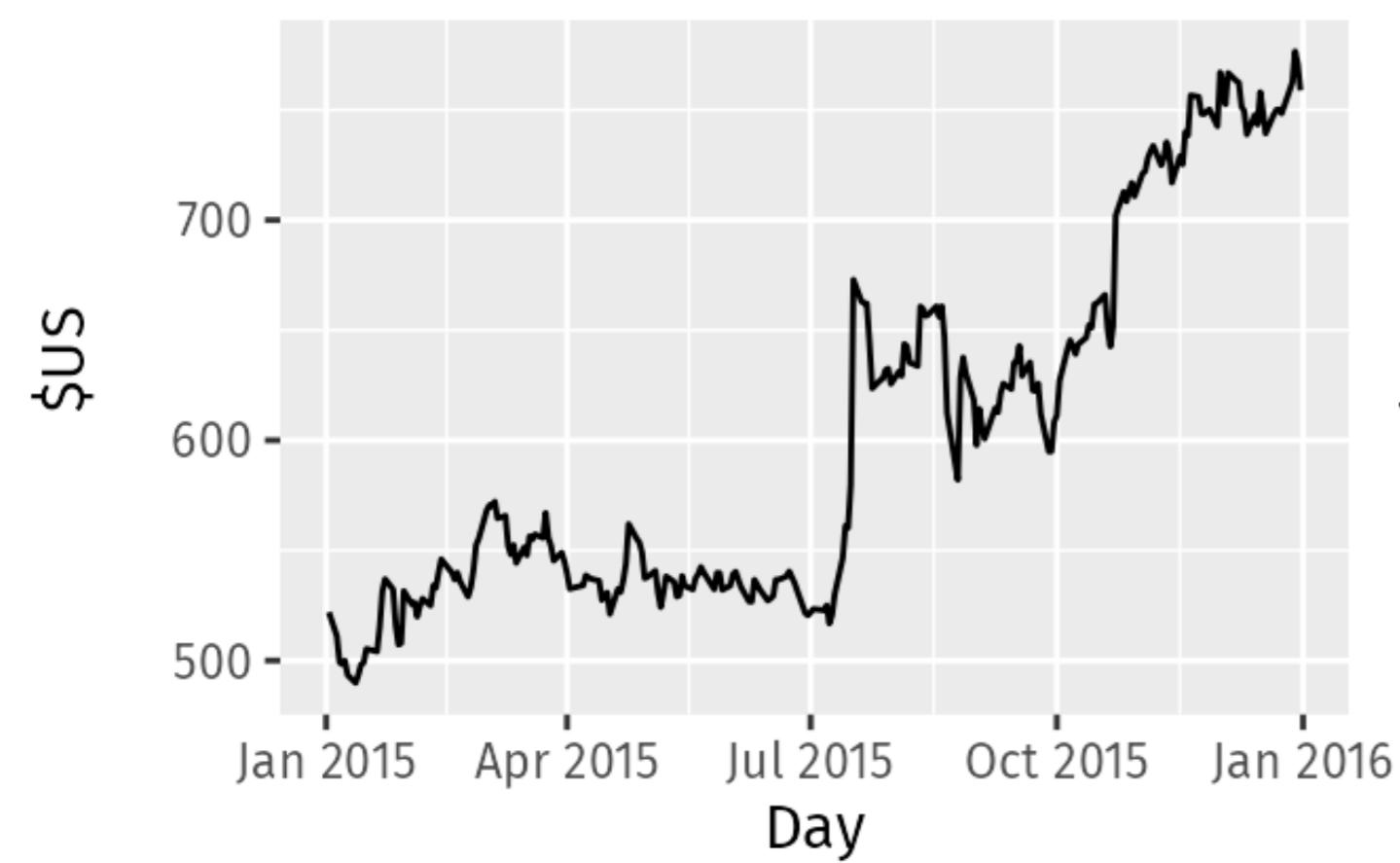


KPSS

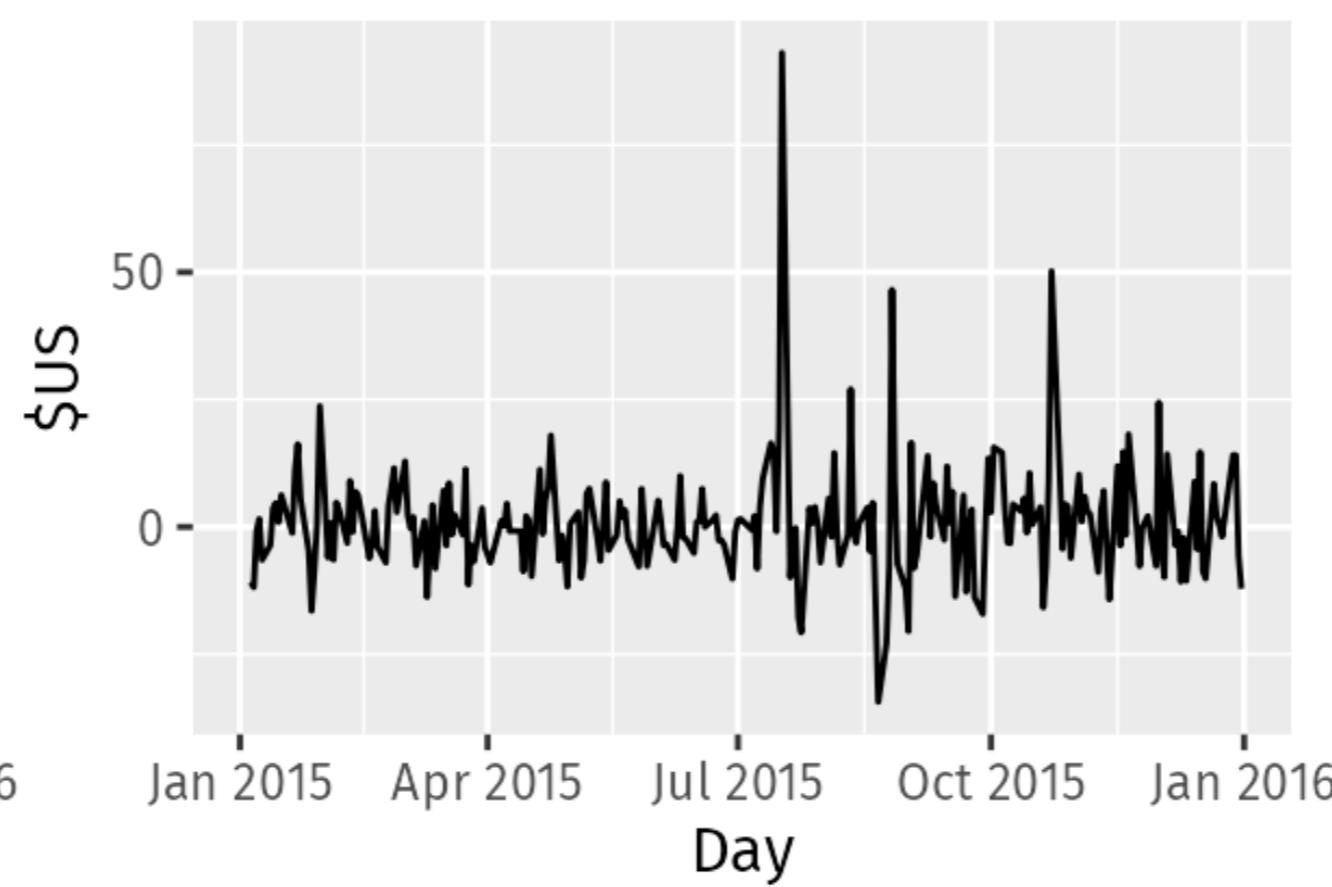
нулевая гипотеза: H_0 : ряд y стационарен;
альтернатива: H_1 : ряд y описывается моделью вида $y_t = \alpha y_{t-1}$;
статистика: $KPSS(y) = \frac{1}{T^2} \sum_{i=1}^T \left(\sum_{t=1}^i y_t \right)^2 / \lambda^2$;
 $KPSS(y)$ при H_0 имеет табличное распределение.

Стационарные?

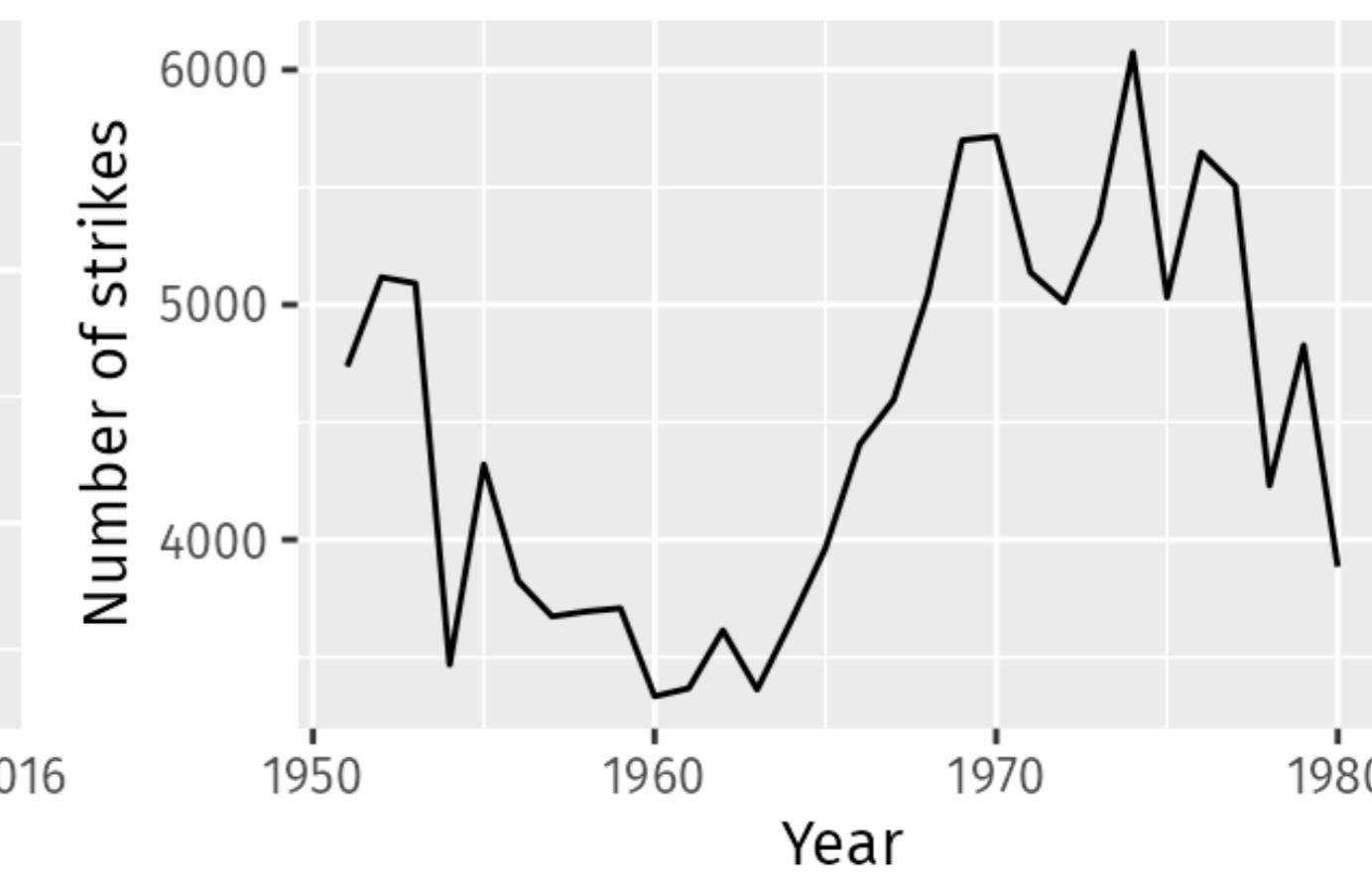
(a) Google closing price



(b) Change in google price



(c) Strikes: US



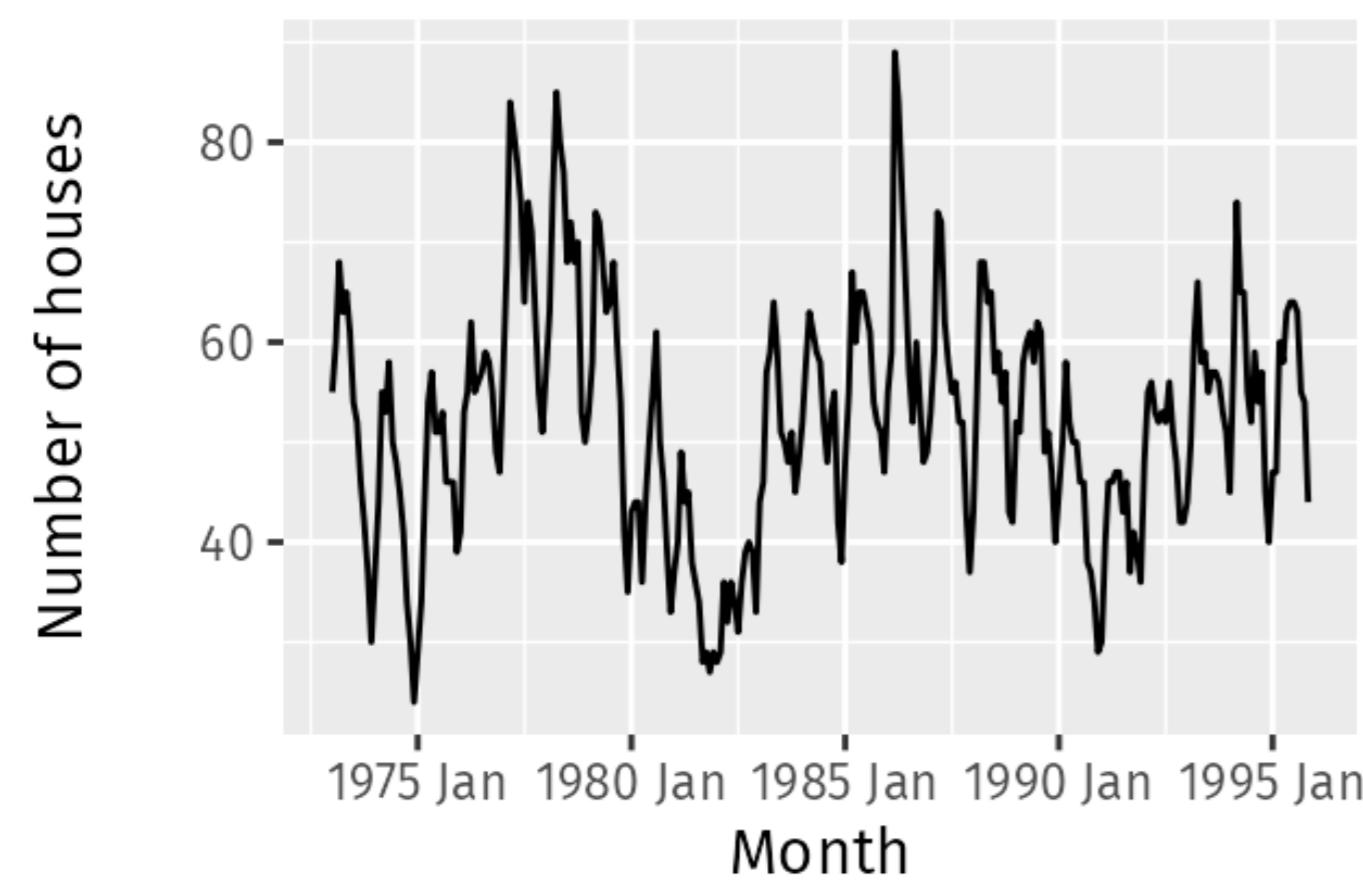
Тренд



Смена уровня

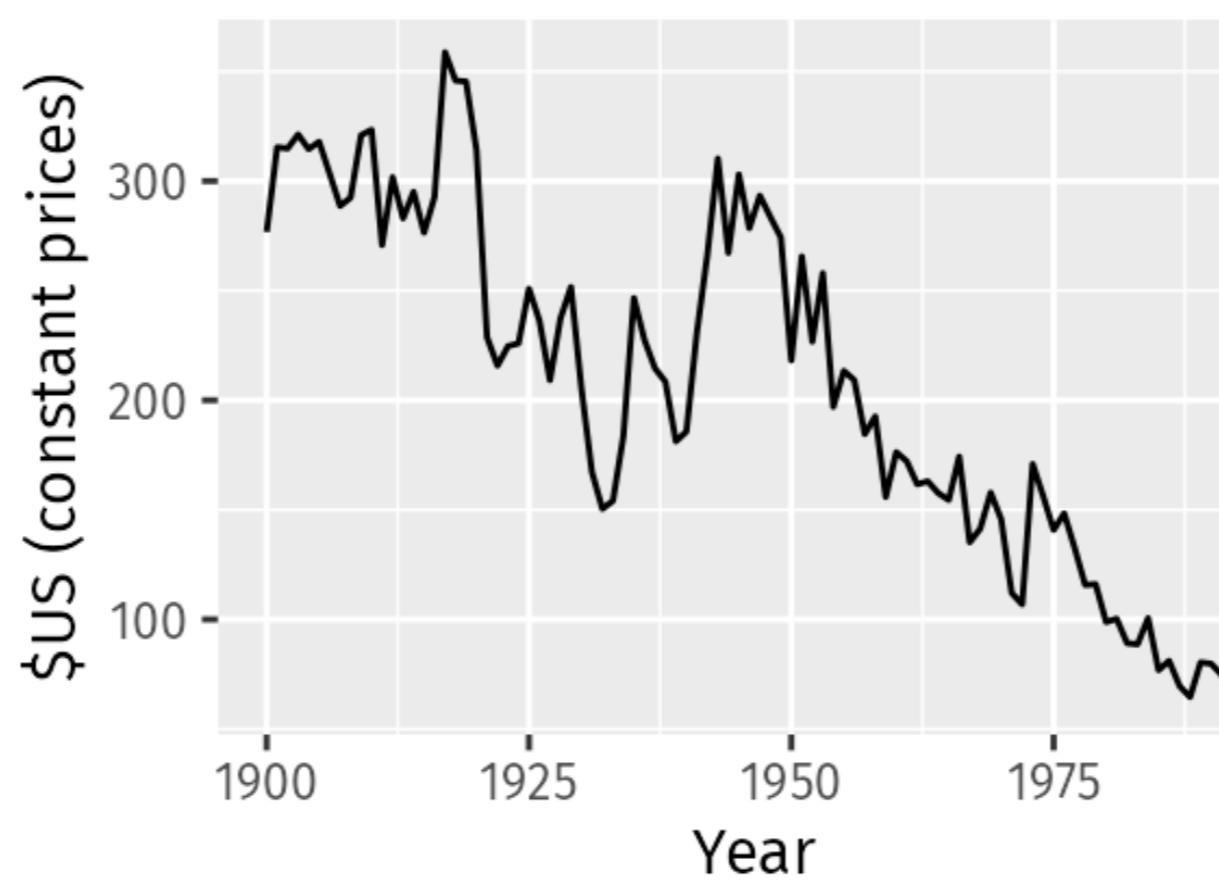
Стационарные?

(d) House sales: US



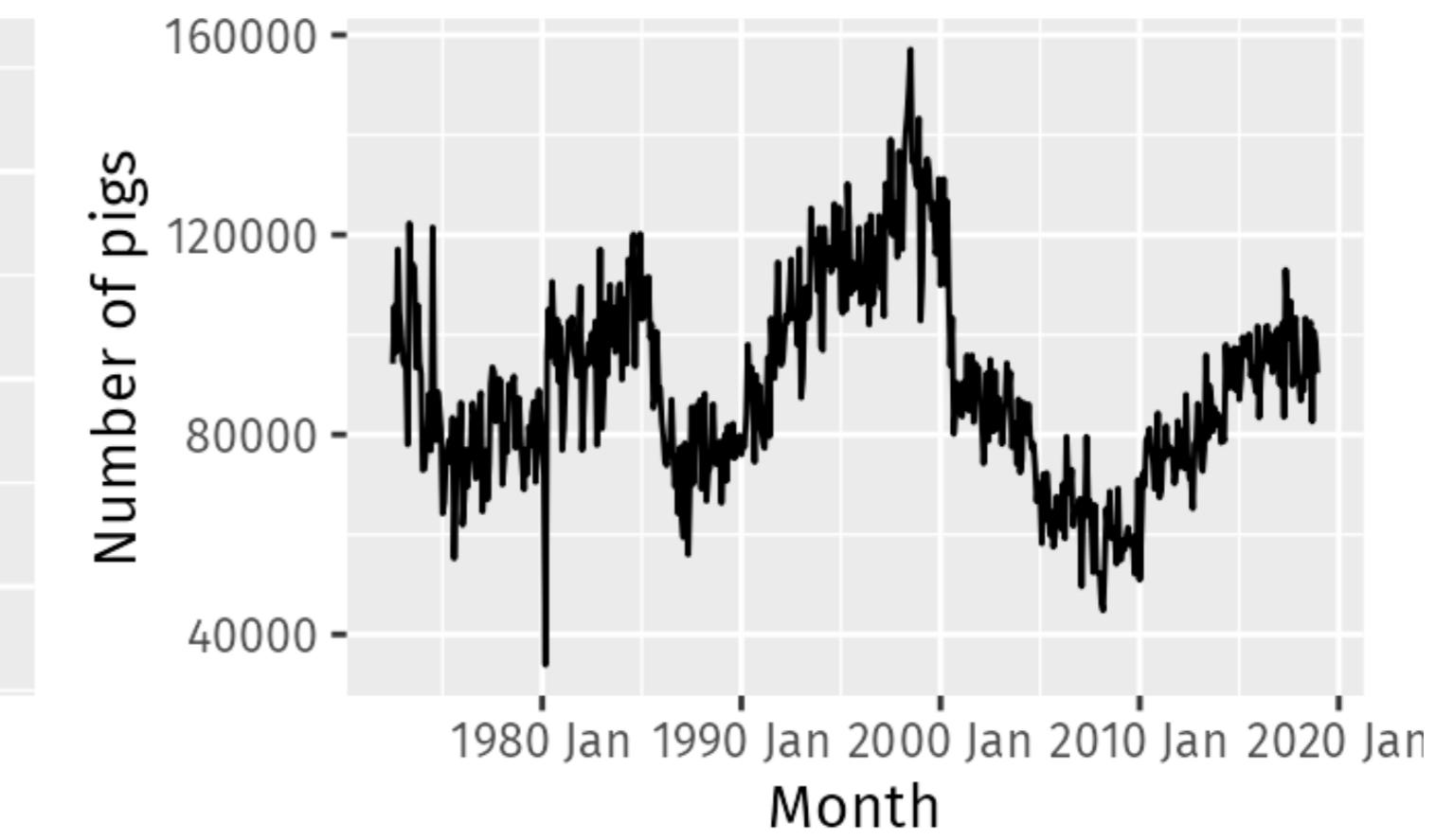
Сезонность

(e) Egg prices: US



Тренд

(f) Pigs slaughtered: Victoria, Australia



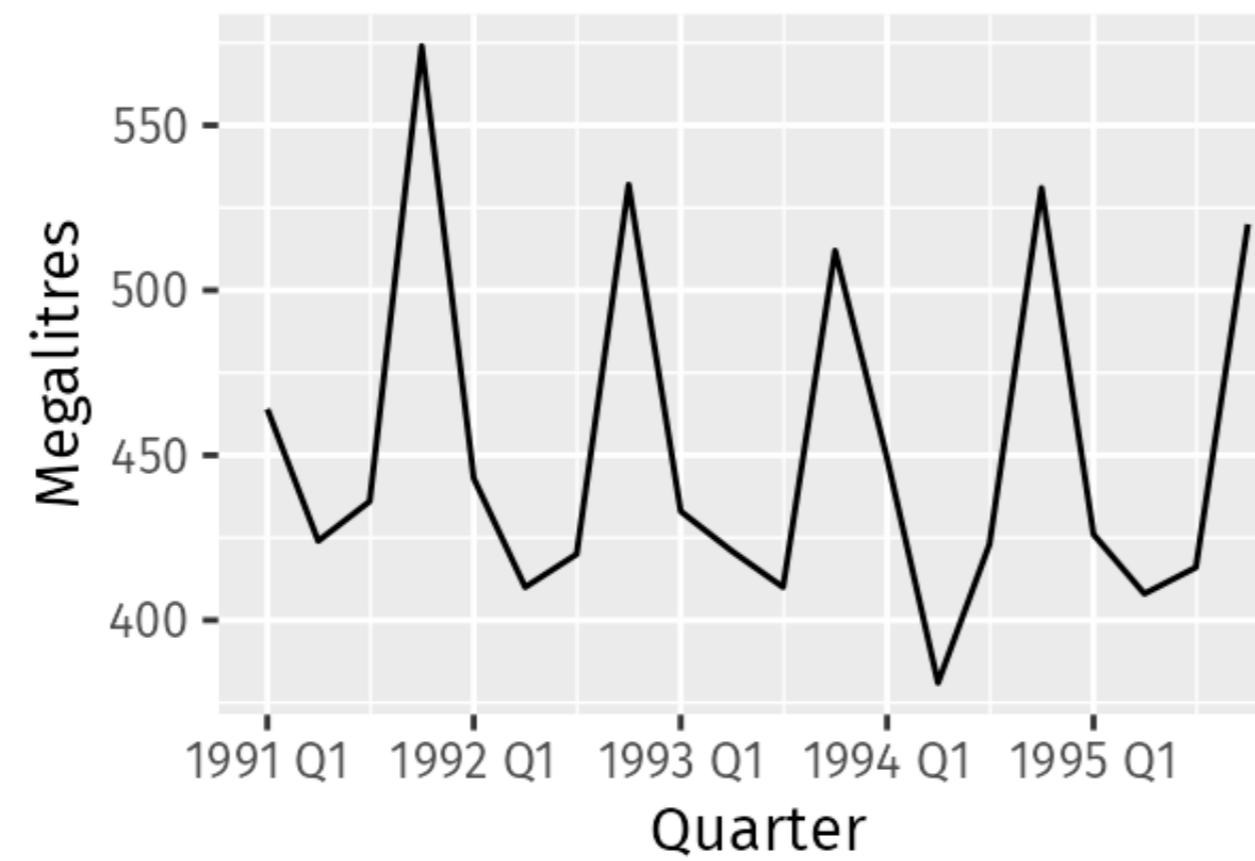
Смена уровня

Стационарные?



Year

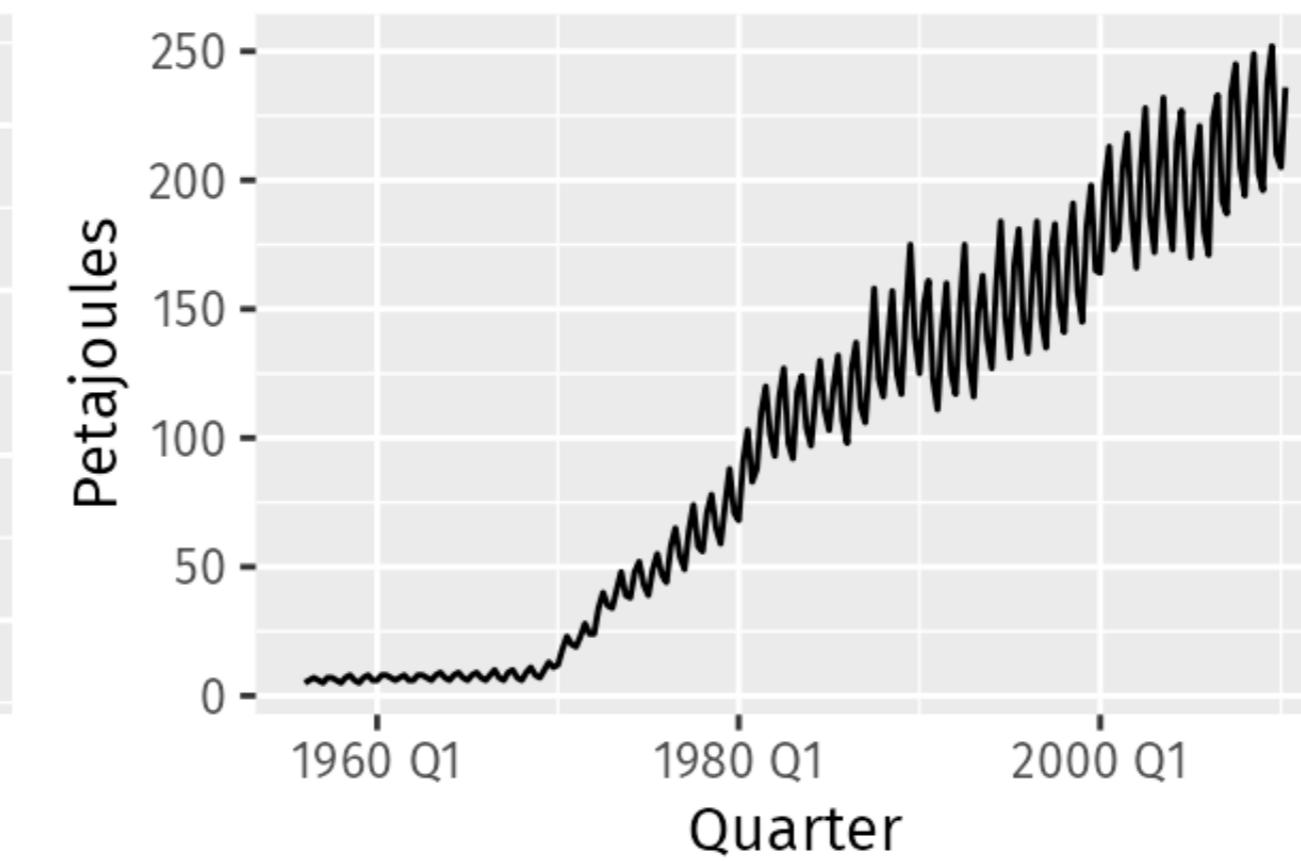
(h) Beer production: Australia



Сезонность

Month

(i) Gas production: Australia

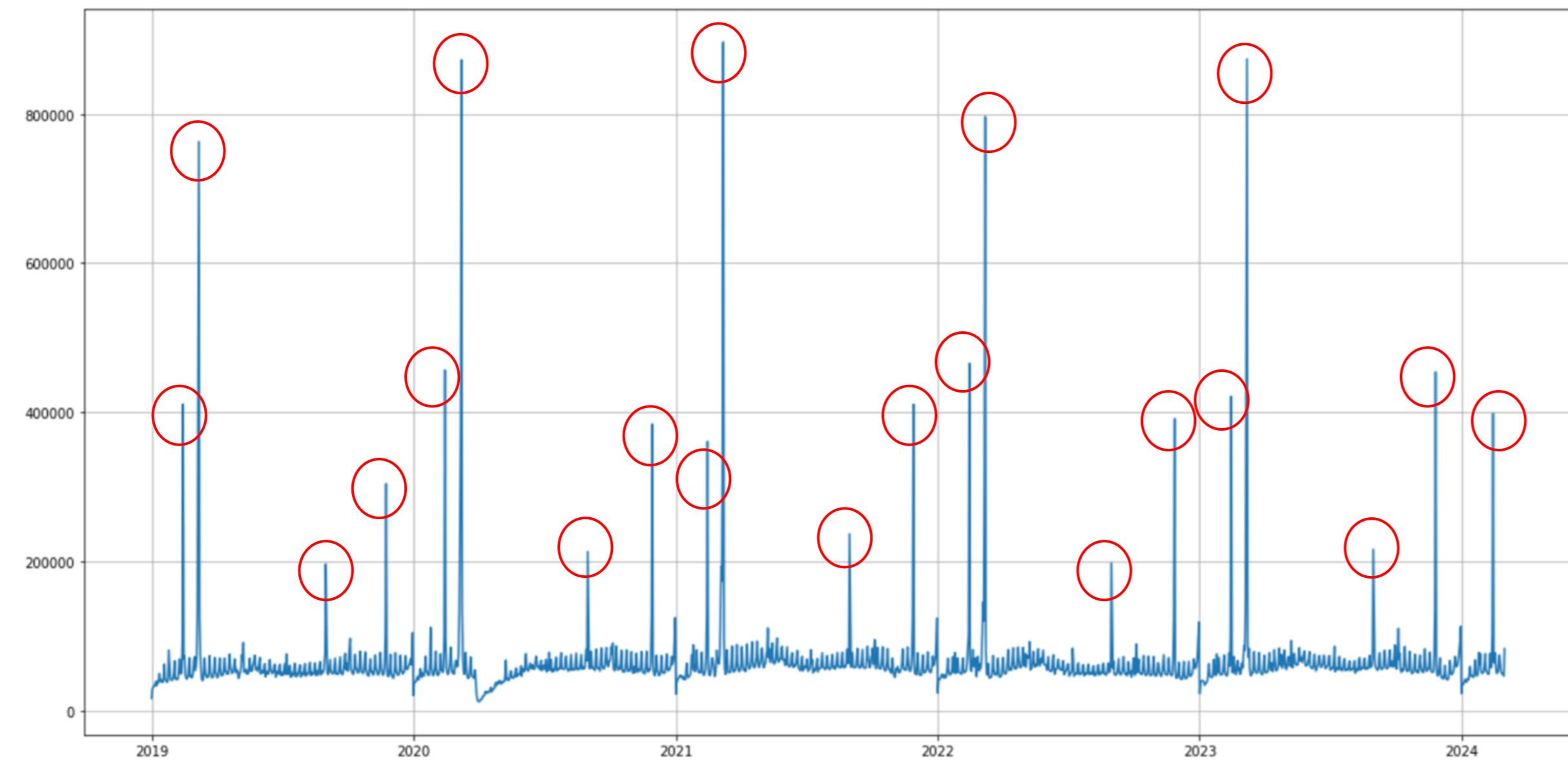


Тренд

5. Аномалии и праздники

Праздники

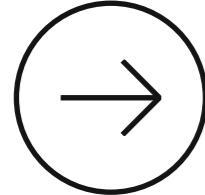
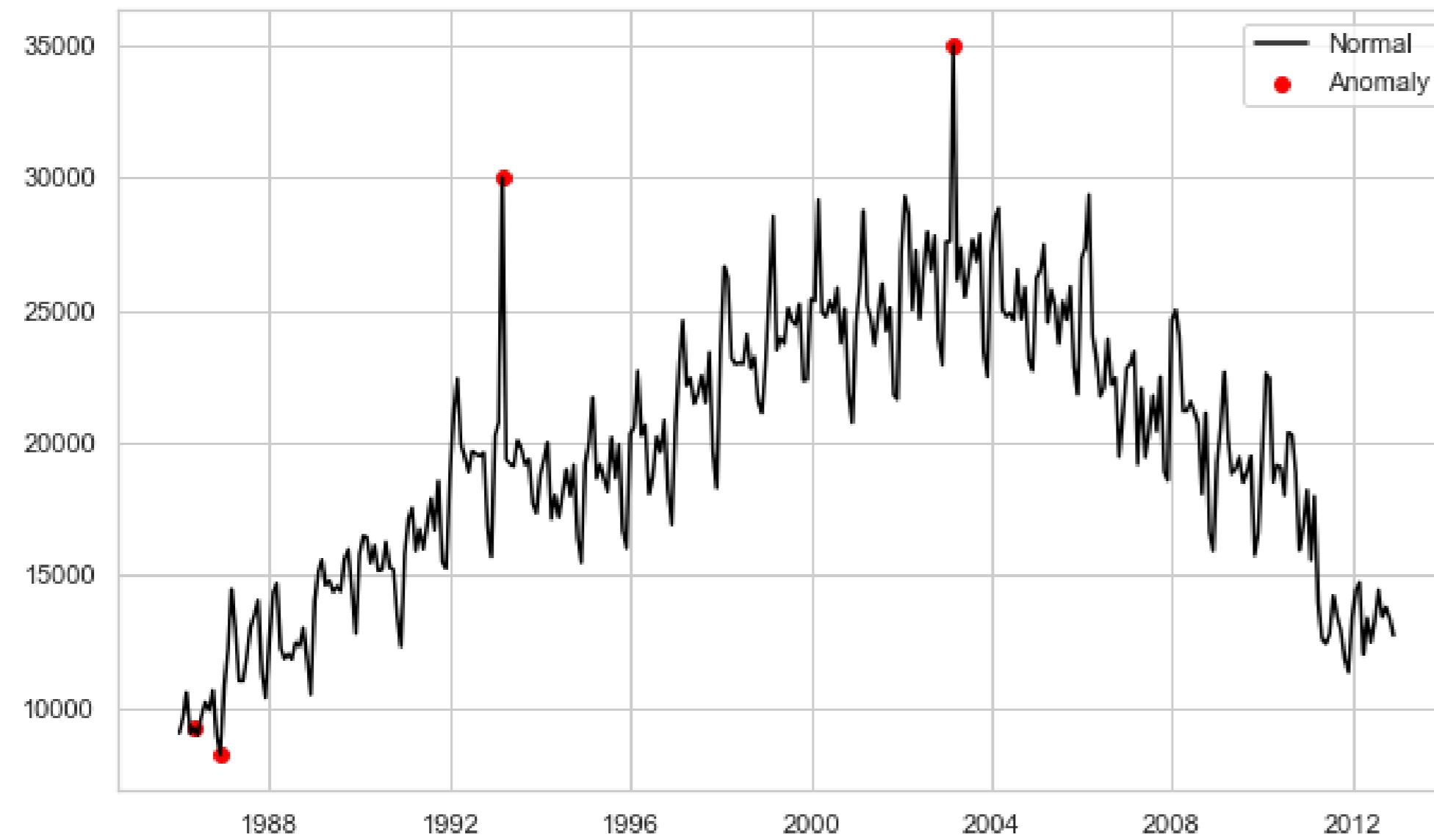
Цветы



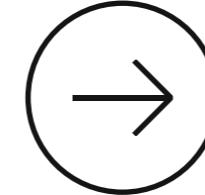
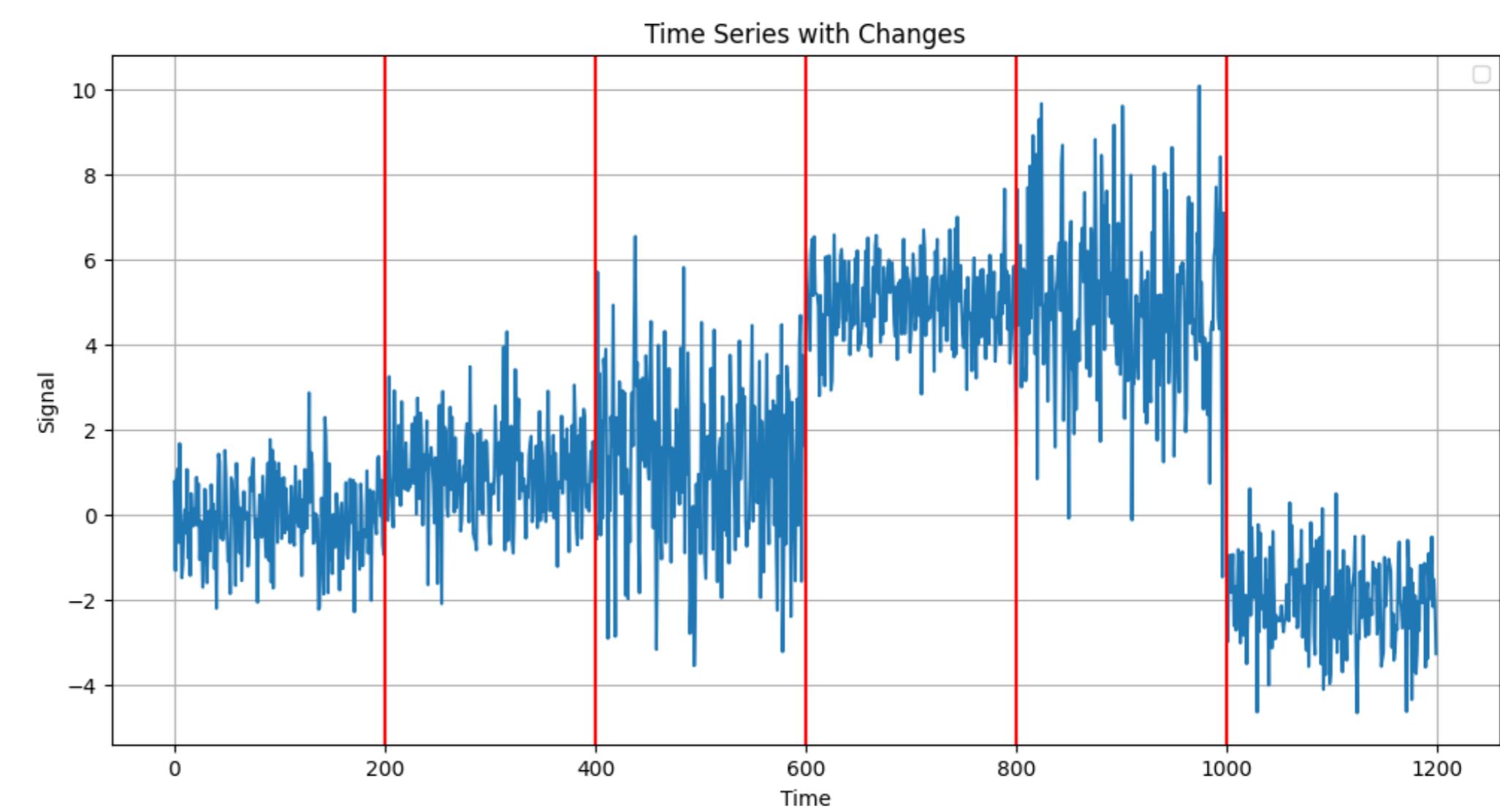
Важно: подсветите праздничные дни на графике, прежде чем искать аномалии.

Зачем: праздники нужно праздновать моделировать, а не выкидывать.

Аномальное поведение

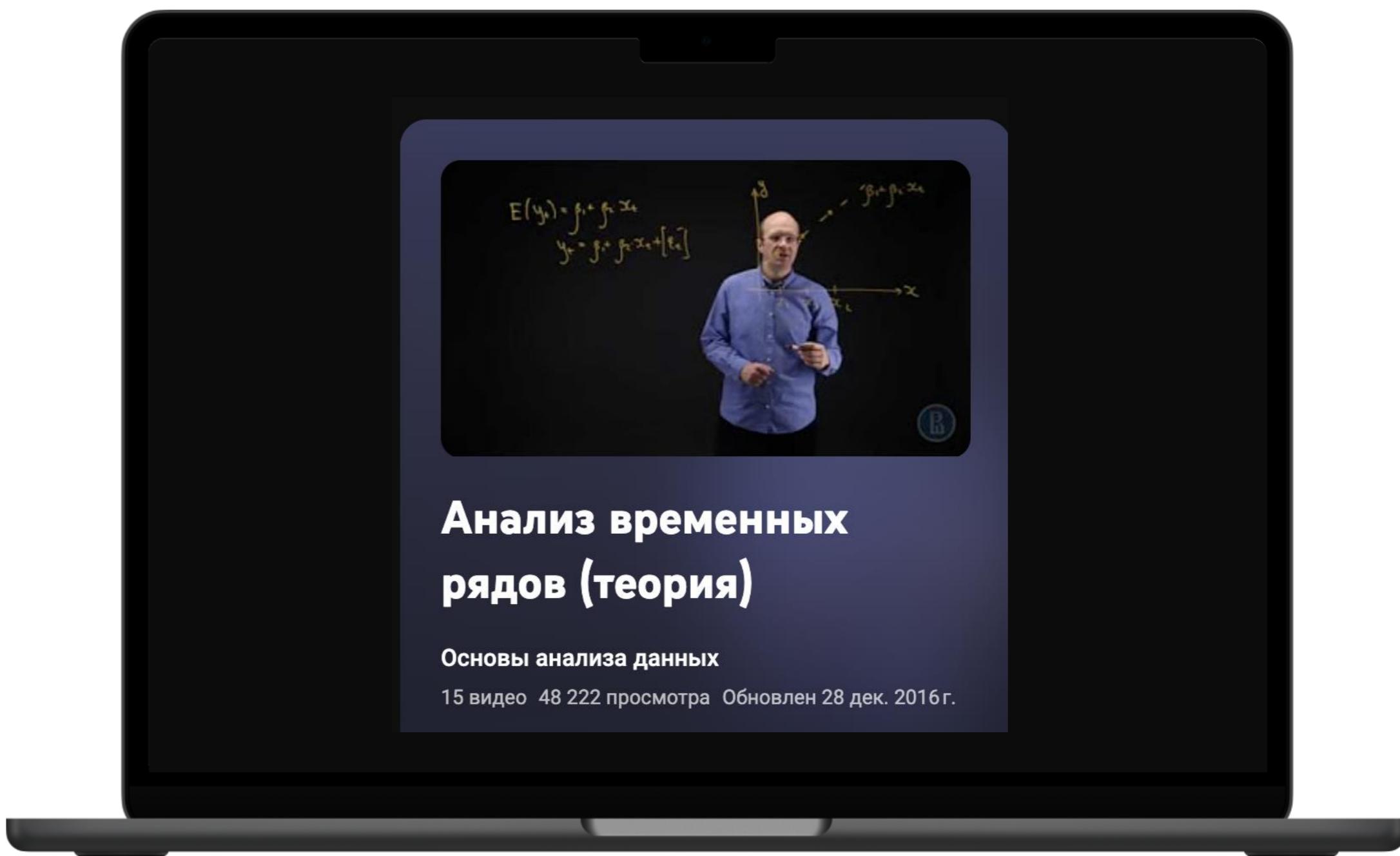


Аномальные точки:
произошло что-то непонятное
единоразово.



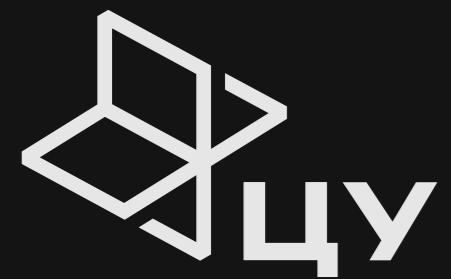
Точки смены поведения:
произошло что-то непонятное,
и ряд начал вести себя по-другому.

Полезные ссылки



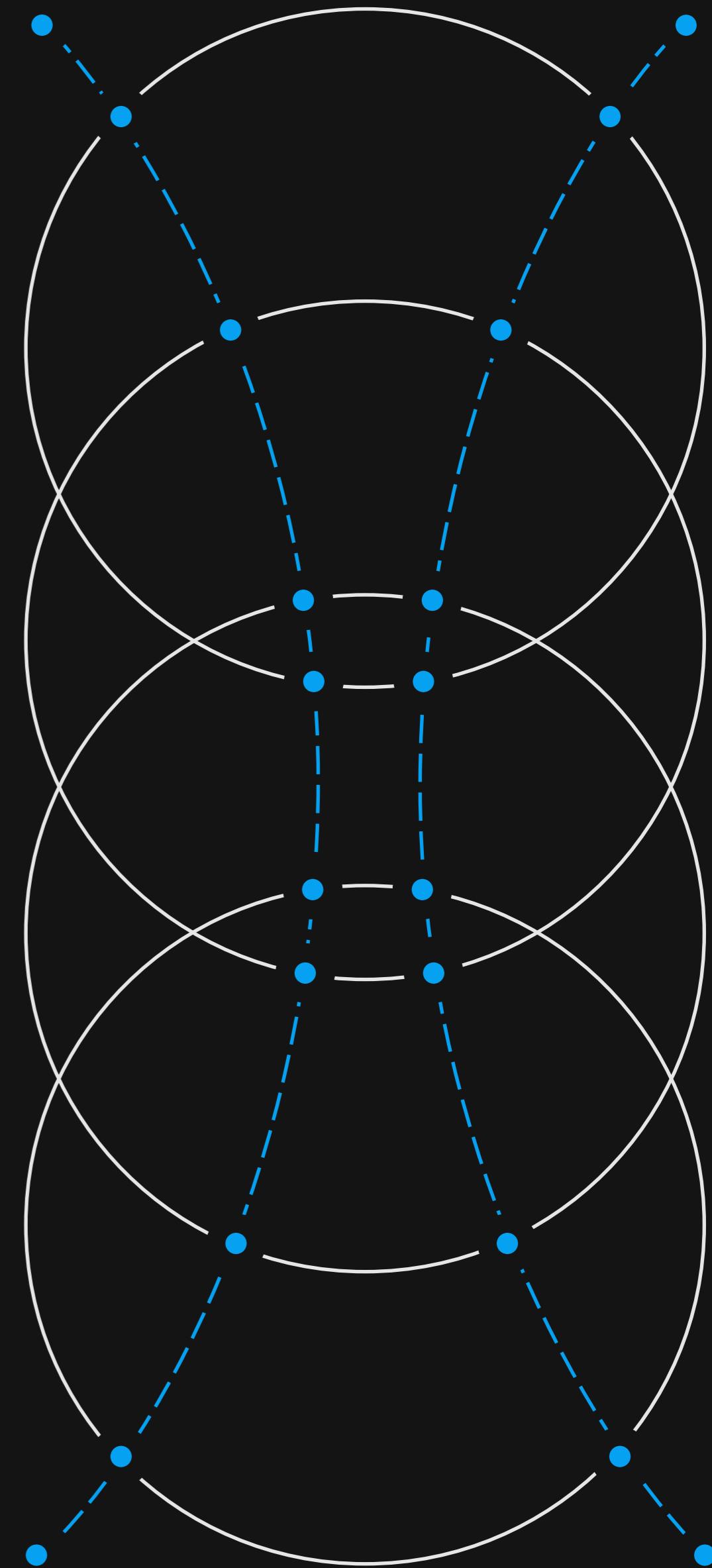
[Борис Демешев \(ВШЭ\)](#)

- Много формализации про ACF / стационарность / AR- и MA-процессы.
- Можно и другие видео поискать от него, формализация — это про него.



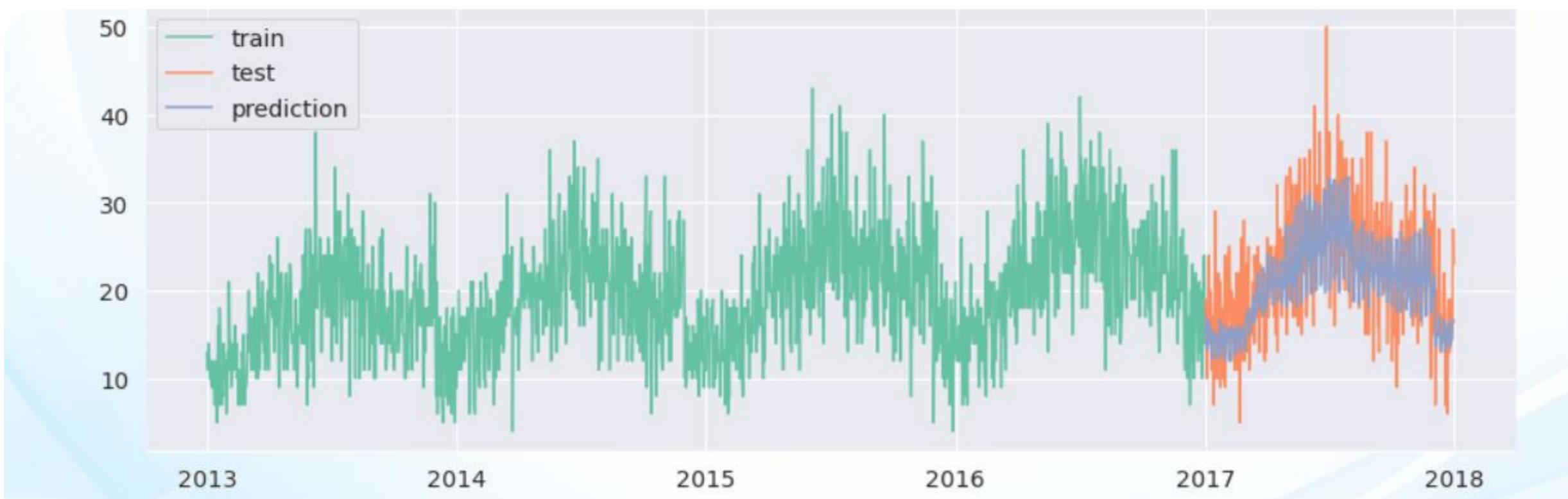
Задача прогнозирования

Как смотреть в будущее уже сегодня



Формализация

Дневные продажи объекта 1 в магазине 1



1. Знаем значения ряда (**зеленые**) до момента времени t .
2. Хотим предсказать (**синее**) будущие значения ряда (**красные**).

$$Y_t = F(y_{t-1}, y_{t-2}, \dots, X, t)$$

y_t — исторические значения;

X — признаки ряда;

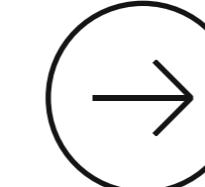
t — время.

F — ?

Методы прогнозирования

Статистические модели

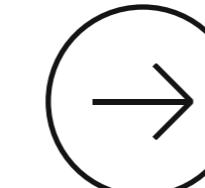
- Prophet
- SARIMA
- ETS



$$y(t) = \text{trend}(t) + \text{seasonality}(t) + \text{noise}(t)$$

ML

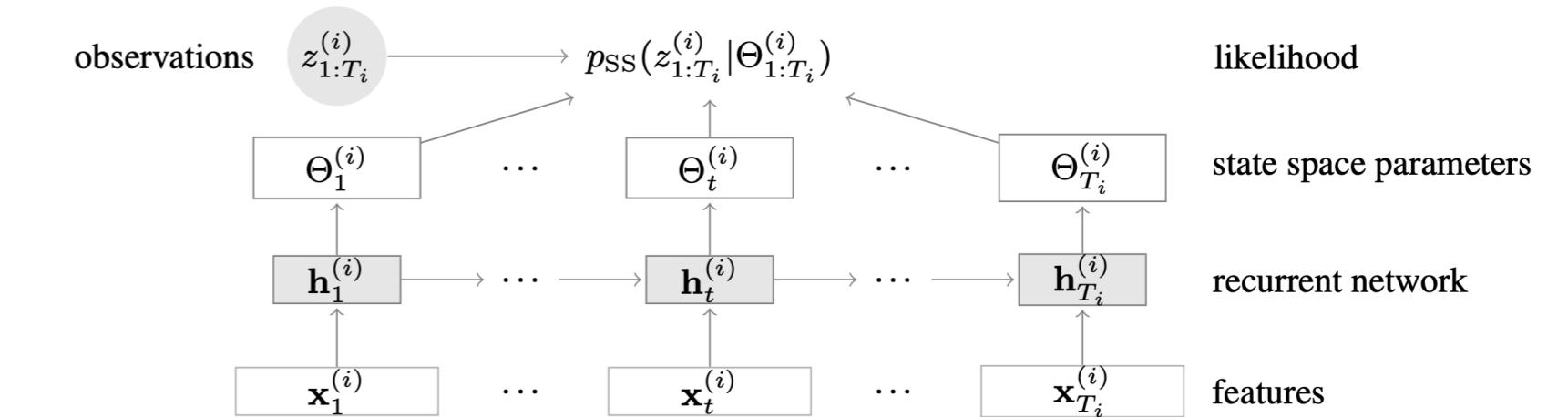
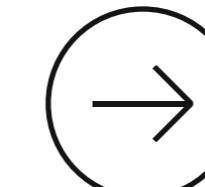
- Catboost
- Linear
- RF



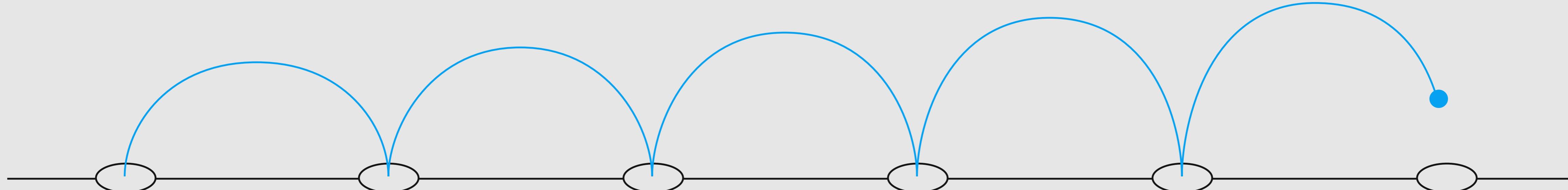
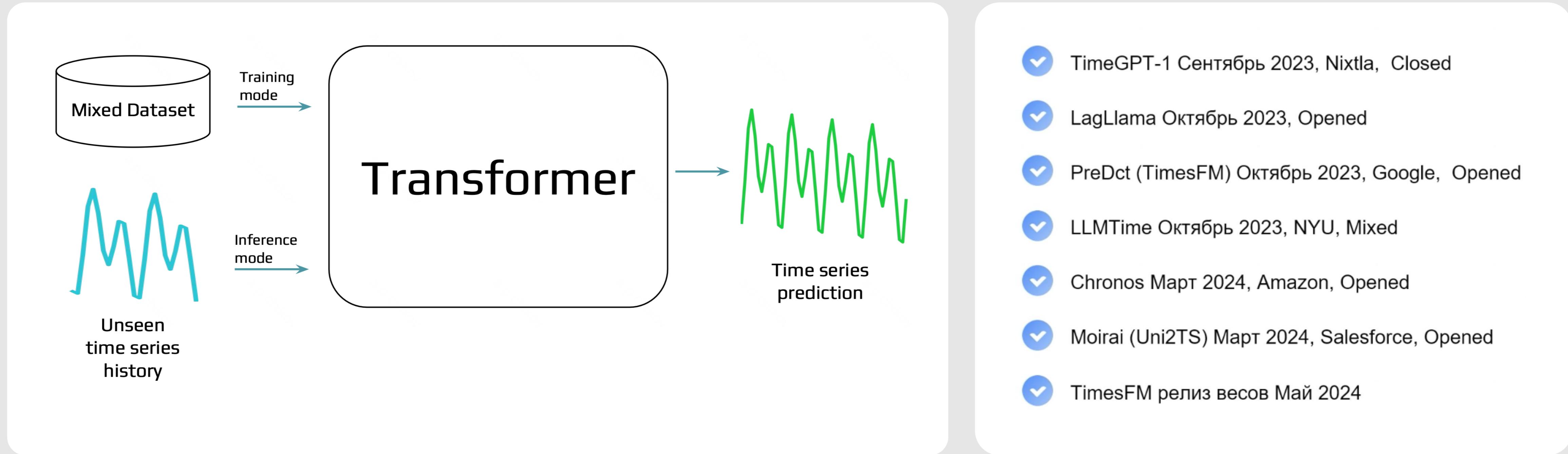
| Dates | Store ID | Product ID | Product name | # of Sales |
|------------|----------|------------|---------------|------------|
| 02-12-2021 | 001 | RP01 | Almond milk | 21 |
| 10-12-2021 | 005 | RS21 | Oat milk | 15 |
| 18-01-2022 | 004 | RK32 | Hazelnut milk | 9 |

DL

- DeepAR
- TFT
- DeepState



Foundational Models (Zero-shot-предсказания)



Где найти реализации?



Статистические модели



DL-модели

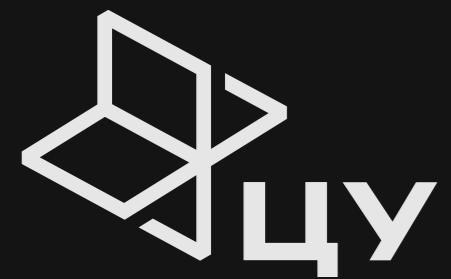


Neural  Forecast

Это мы!

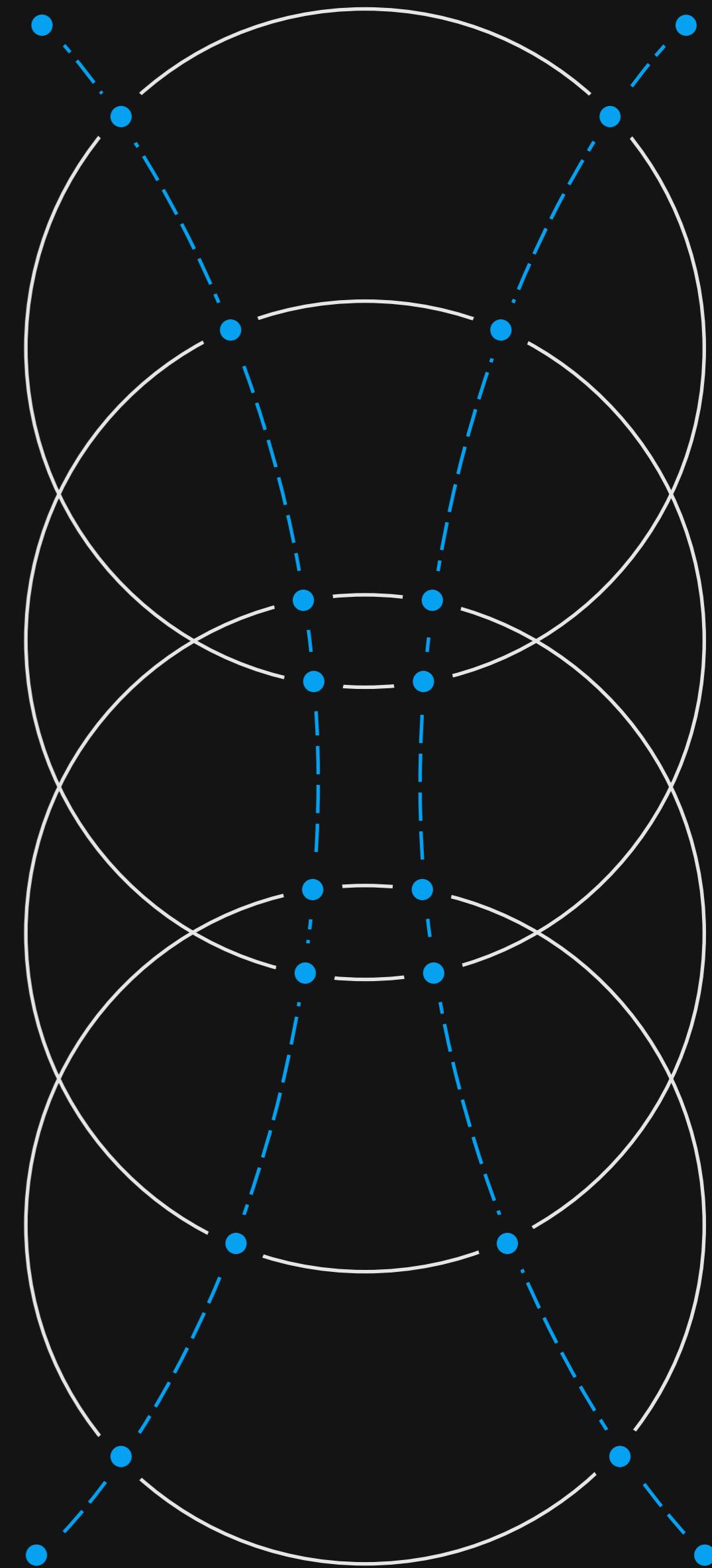


Фреймворки



Статистические подходы

SMA, SARIMAX, ETS, Prophet & Co



1. Seasonal Moving Average

Seasonal Moving Average

```
1 SeasonalMovingAverageModel(  
2     window: int,  
3     seasonality: int  
4 )
```

$$\hat{y}_t = \frac{1}{w} \sum_{n=1}^w y_{t-n*s}$$

Seasonal Moving Average
(window = 3, seasonality = 7)

$$\hat{y}_t = \frac{1}{3} (y_{t-7} + y_{t-14} + y_{t-21})$$

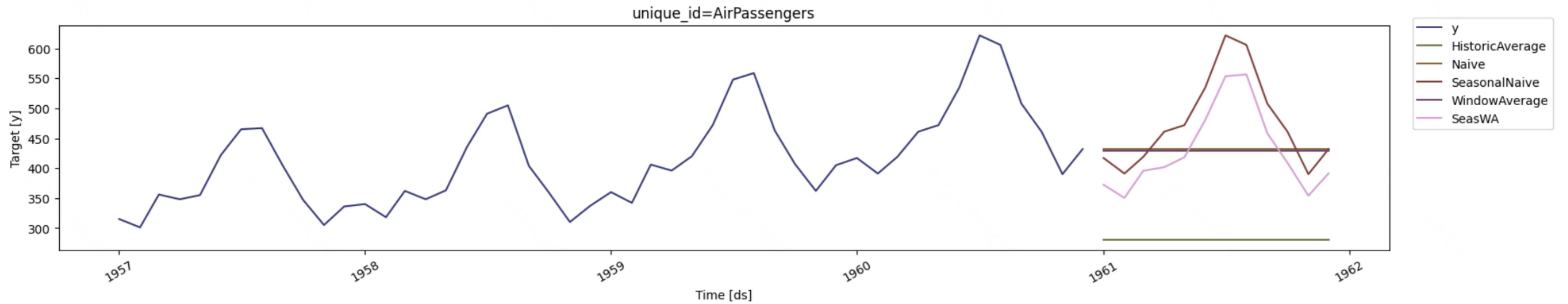
→ Особенности:

- сильный бейзлайн;
- «сегодня будет так же, как вчера»;
- можно всячески модифицировать.

→ Когда использовать:

- ряды непредсказуемы;
- поведение слабо меняется.

Seasonal Moving Average (вариации)



window

1

1

inf

N

N

seasonality

1

S

1

1

M

Naive

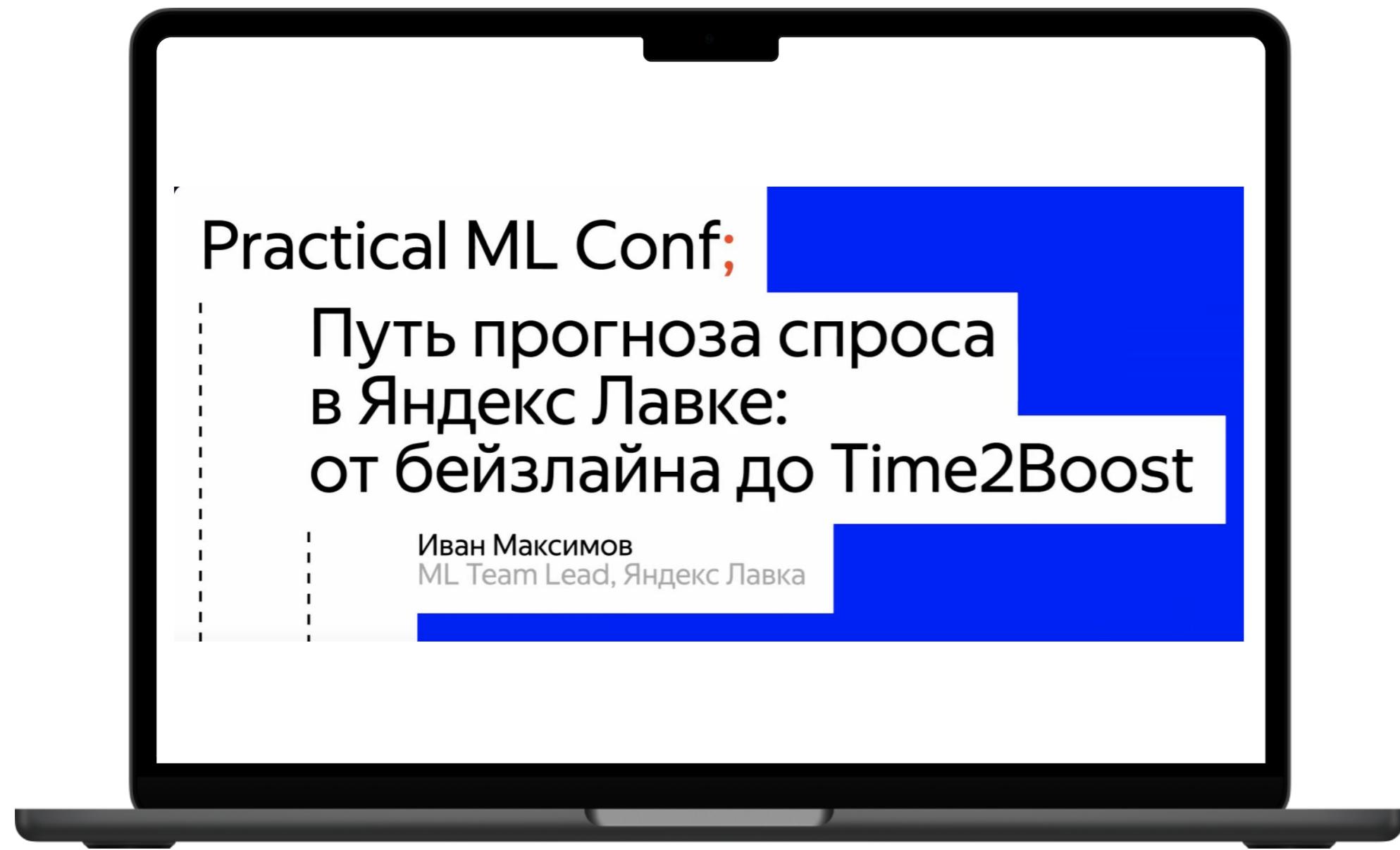
Seasonal Naive

Historic Average

Window Average

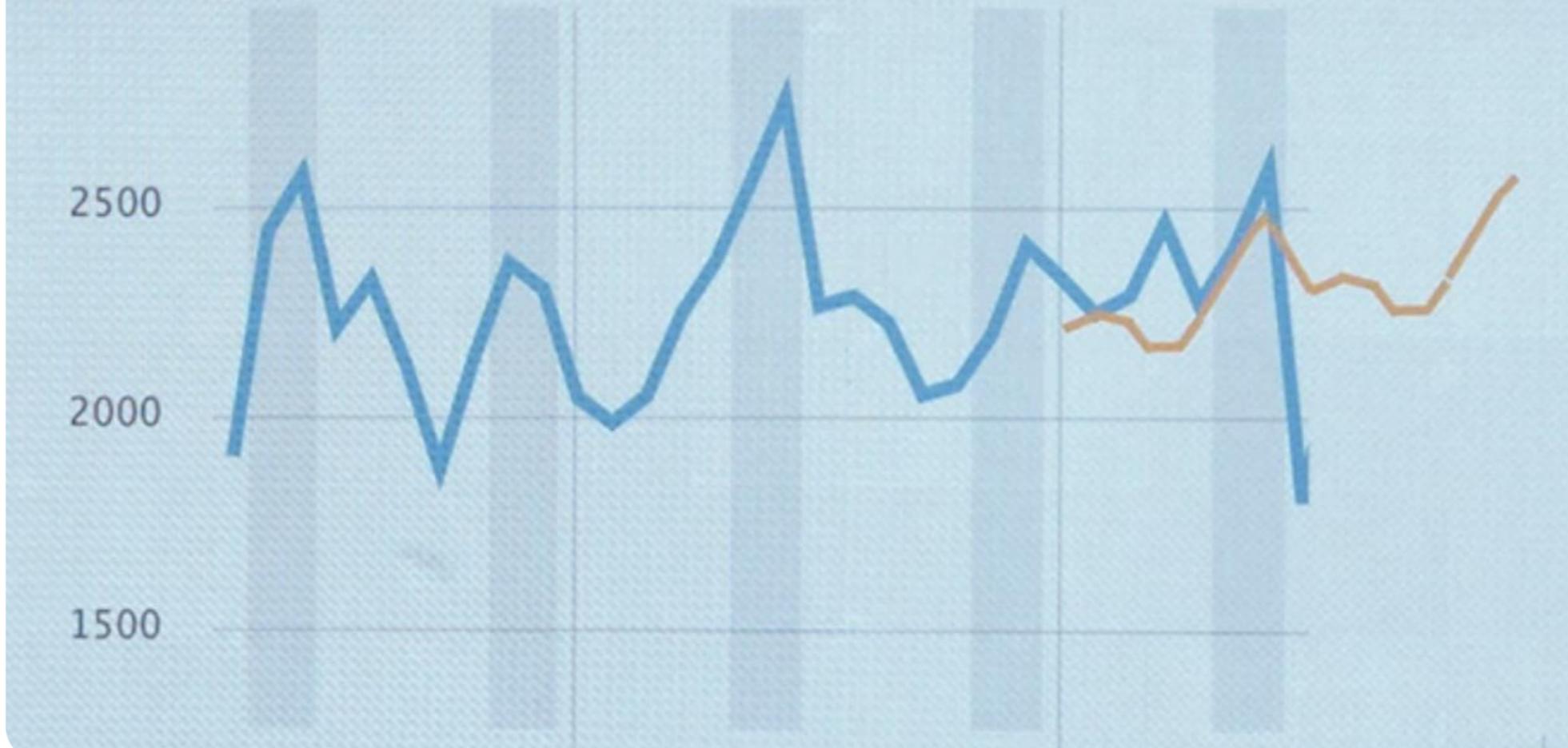
Seasonal Window Average

Seasonal Moving Average



[Путь прогноза спроса в «Яндекс Лавке»:
от бейзлайна до Time2Boost](#)

Прогноз =
[clip(Робастное среднее спроса за 4 понедельника, 30%)
+ clip(Тренд неделя-к-неделе, 10%)] * Holiday_coef



2. ETS (Error, Trend, Seasonal)

ETS (Error, Trend, Seasonal)

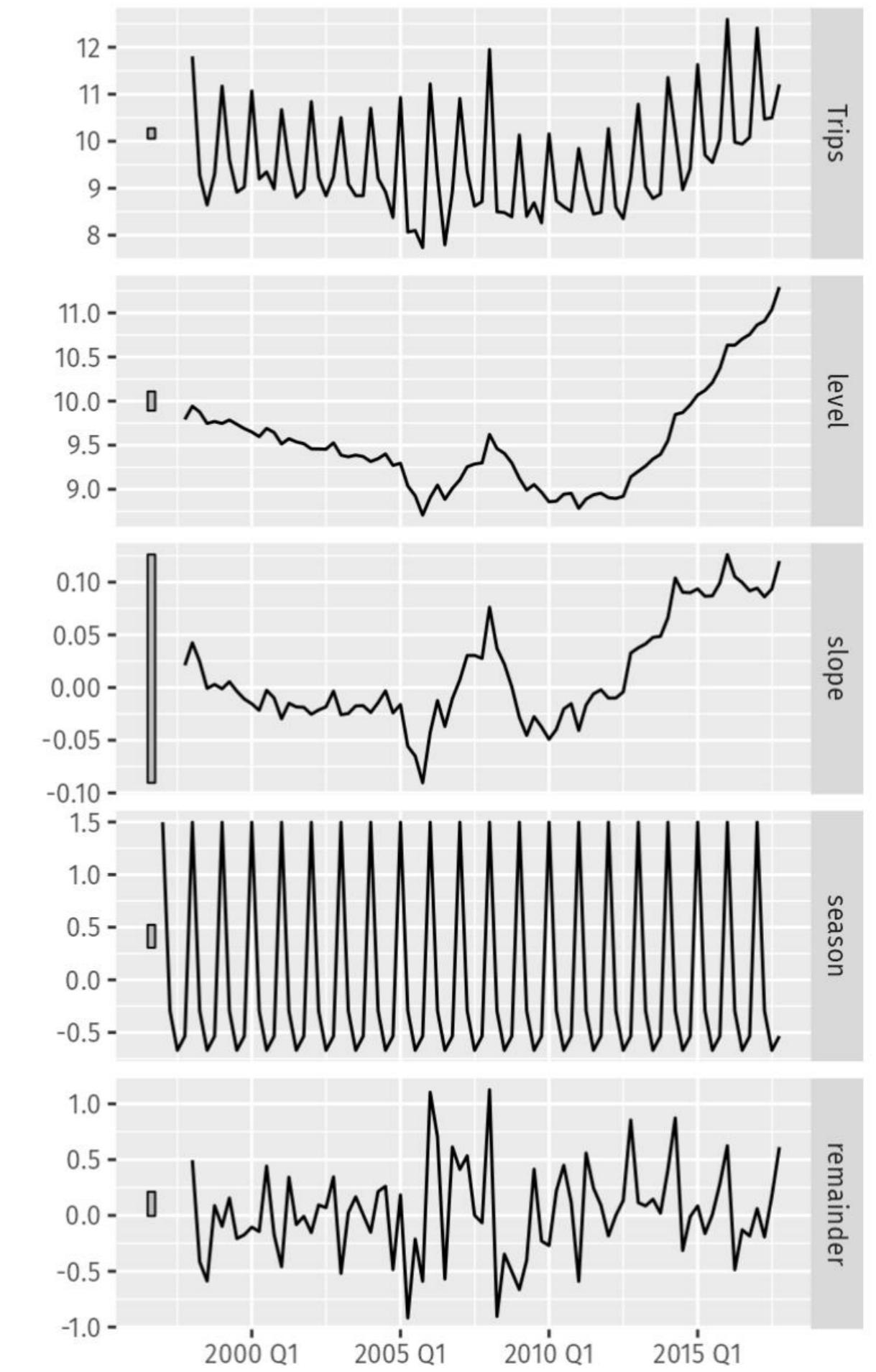
```

1 HoltWintersModel(
2     trend: str,
3     damped_trend: bool,
4     seasonal: str,
5     seasonal_periods: int
6 )

```

$$\begin{aligned}
 y_t &= \ell_{t-1} + \phi b_{t-1} + s_{t-m} + \varepsilon_t \\
 \ell_t &= \ell_{t-1} + \phi b_{t-1} + \alpha \varepsilon_t \\
 b_t &= \phi b_{t-1} + \beta \varepsilon_t \\
 s_t &= s_{t-m} + \gamma \varepsilon_t
 \end{aligned}$$

$$N(0, \sigma)$$



ETS (Error, Trend, Seasonal)

```
1 HoltWintersModel(  
2     trend: str,  
3     damped_trend: bool,  
4     seasonal: str,  
5     seasonal_periods: int  
6 )
```

Про терминологию:

Y — наблюдаемая переменная;
 l, b, s — скрытые состояния.

Гиперпараметры:

- тип модели (далее);
- тип тренда;
- период сезонности.

Обучаемые параметры

$$\begin{aligned}y_t &= \ell_{t-1} + \phi b_{t-1} + s_{t-m} + \varepsilon_t \\ \ell_t &= \ell_{t-1} + \phi b_{t-1} + \alpha \varepsilon_t \\ b_t &= \phi b_{t-1} + \beta \varepsilon_t \\ s_t &= s_{t-m} + \gamma \varepsilon_t\end{aligned}$$

$l_0, b_0, s_0, \dots, s_{m-1}$

Filtering and Forecasting equations

→ **Filtering** — получение скрытых состояний по наблюдаемому (итеративно с $t = 1$).

$$\epsilon_t = y_t - (l_{t-1} + b_{t-1} + s_{t-m})$$

$$l_t = l_{t-1} + b_{t-1} + \alpha \epsilon_t$$

$$b_t = b_{t-1} + \beta \epsilon_t$$

$$s_t = s_{t-m} + \gamma \epsilon_t$$

$$t \in [1, T]$$

→ **Forecasting** — предсказание наблюдаемого состояния как эволюции скрытых.

$$\hat{y}_{t+h|t} = l_t + h b_t + s_{t+h-m}$$

$$l_t = l_{t-1} + b_{t-1}$$

$$b_t = b_{t-1}$$

$$s_t = s_{t-m}$$

$$t \in [T + 1, T + H]$$

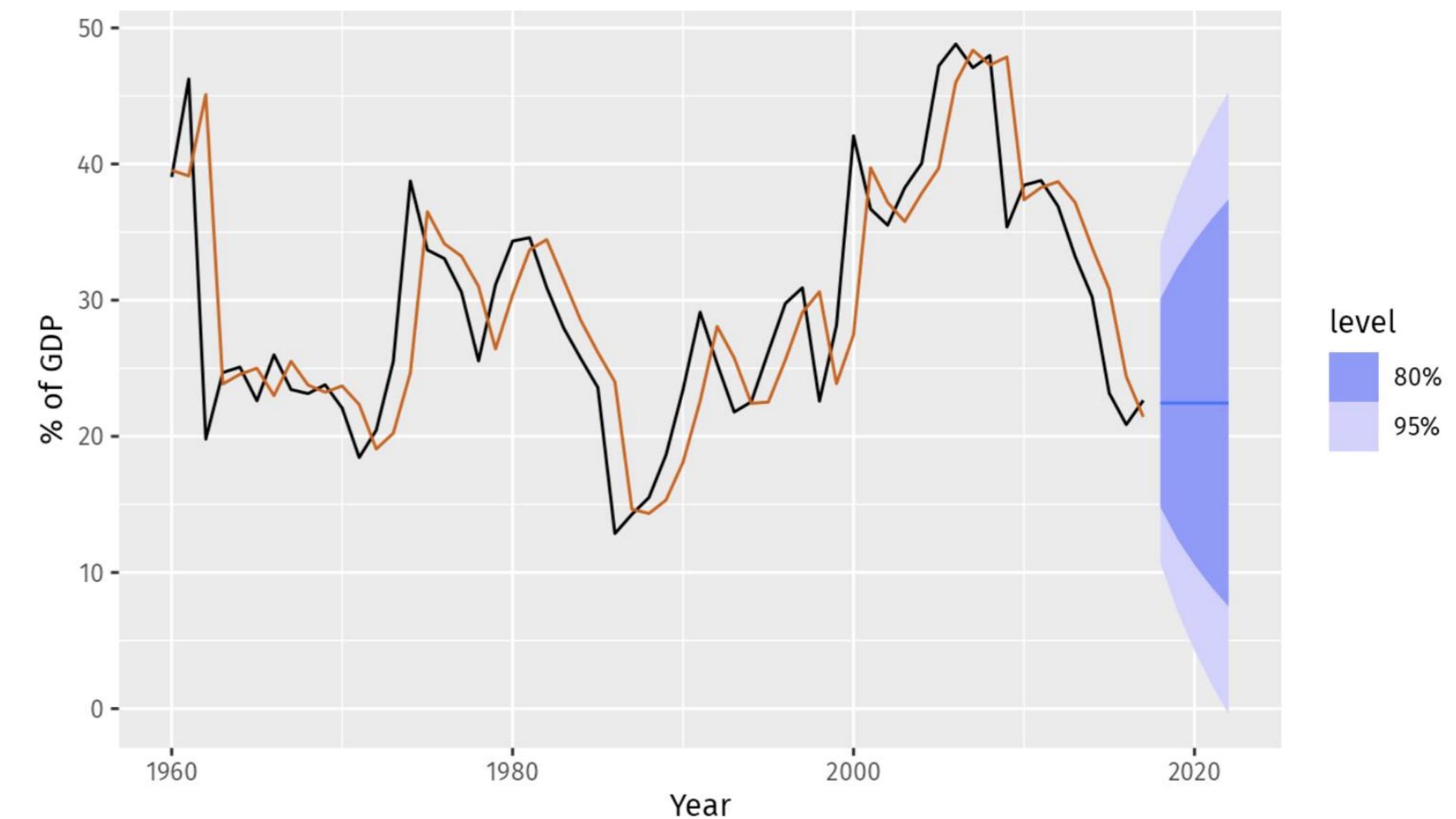
ETS (Simple Exponential Smoothing)

```
1 SimpleExpSmoothingModel(  
2     ... # Нет параметров  
3 )
```

→ Особенности:

- прогноз = константа
(последний известный уровень ряда);
- используем для рядов без тренда
и сезонности.

$$\hat{y}_{t+h|t} = l_t$$
$$l_t = l_{t-1}$$



ETS (Holt)

```

1 HoltModel(
2     exponential: bool,
3     damped_trend: bool
4 )

```

→ **damped_trend**

False: $\hat{y}_{t+h|t} = \ell_t + hb_t$

True: $\hat{y}_{t+h|t} = \ell_t + (\phi + \phi^2 + \dots + \phi^h)b_t$

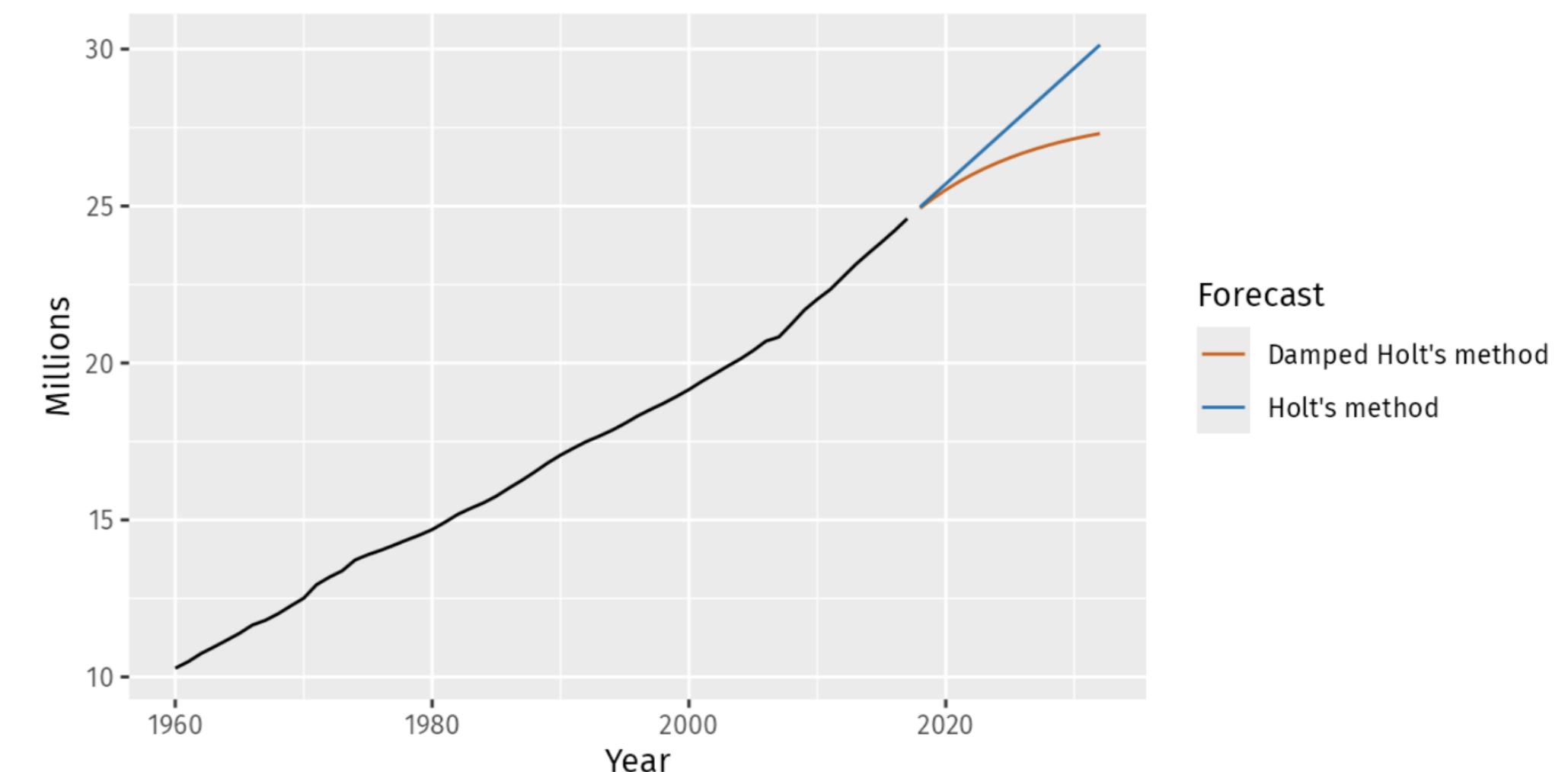
→ **Особенности**

1. Прогноз = линейная функция от горизонта (h).
2. Используем для рядов с трендом.

$$\hat{y}_{t+h|t} = l_t + hb_t$$

$$l_t = l_{t-1} + b_{t-1}$$

$$b_t = b_{t-1}$$



ETS (HoltWinters)

```

1 HoltWintersModel(
2     trend: str,
3     damped_trend: bool,
4     seasonal: str,
5     seasonal_periods: int
6 )

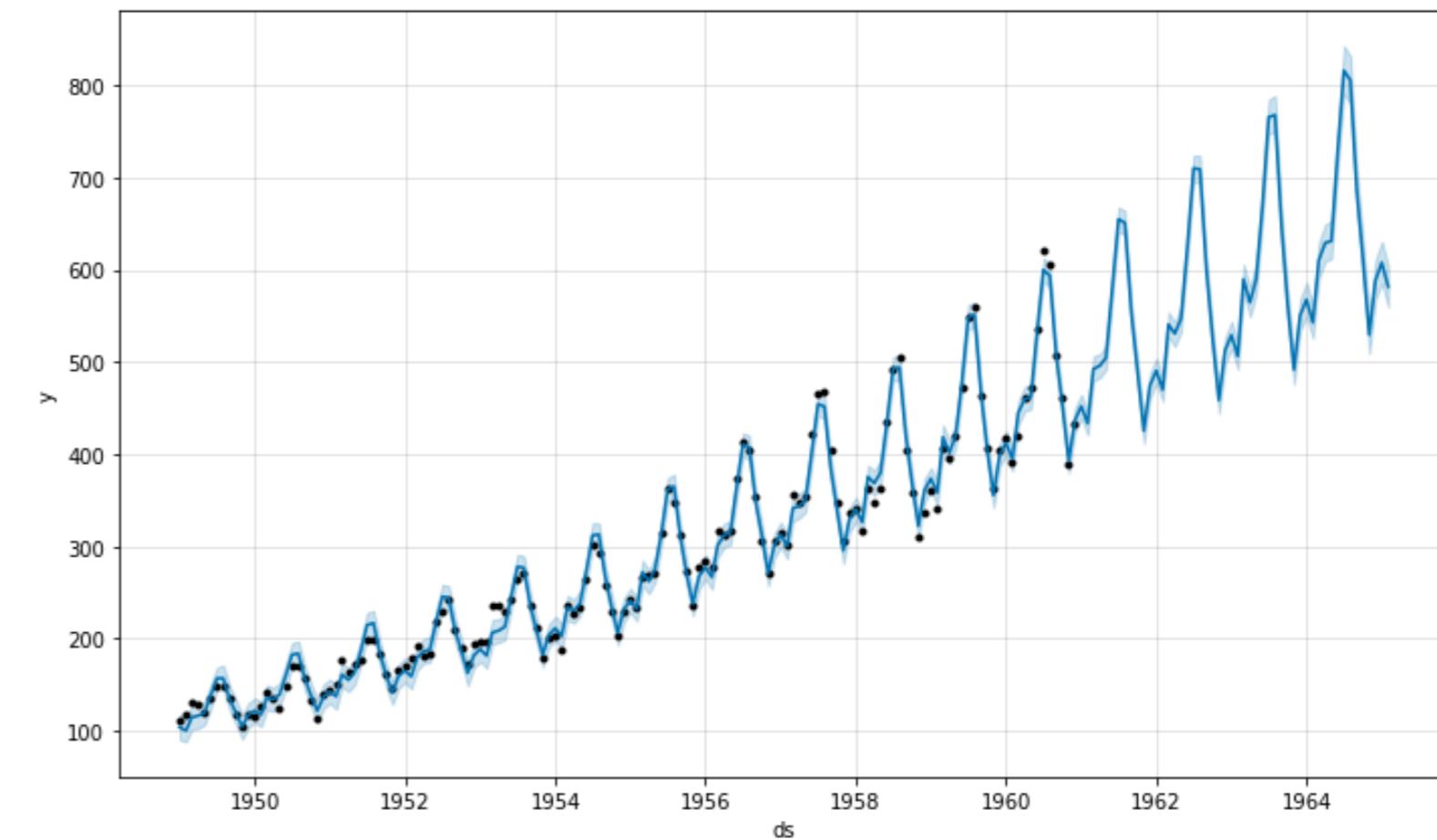
```

→ **trend/seasonal**

add: $\hat{y}_{t+h|t} = \ell_t + hb_t + s_{t+h-m(k+1)}$

mult: $\hat{y}_{t+h|t} = (\ell_t + \phi_h b_t) s_{t+h-m(k+1)}$

$$\begin{aligned}\hat{y}_{t+h|t} &= l_t + hb_t + s_{t+h-m} & b_t &= b_{t-1} \\ l_t &= l_{t-1} + b_{t-1} & s_t &= s_{t-m}\end{aligned}$$



→ **mult:** когда тренд/сезонность зависят от масштаба ряда.

ETS (как подбирать тип)

Семейство моделей

ADDITIVE ERROR MODELS

| Trend | Seasonal | | |
|----------------|--|--|--|
| | N | A | M |
| N | $y_t = \ell_{t-1} + \varepsilon_t$ $\ell_t = \ell_{t-1} + \alpha \varepsilon_t$ | $y_t = \ell_{t-1} + s_{t-m} + \varepsilon_t$ $\ell_t = \ell_{t-1} + \alpha \varepsilon_t$ $s_t = s_{t-m} + \gamma \varepsilon_t$ | $y_t = \ell_{t-1} s_{t-m} + \varepsilon_t$ $\ell_t = \ell_{t-1} + \alpha \varepsilon_t / s_{t-m}$ $s_t = s_{t-m} + \gamma \varepsilon_t / \ell_{t-1}$ |
| A | $y_t = \ell_{t-1} + b_{t-1} + \varepsilon_t$ $\ell_t = \ell_{t-1} + b_{t-1} + \alpha \varepsilon_t$ $b_t = b_{t-1} + \beta \varepsilon_t$ | $y_t = \ell_{t-1} + b_{t-1} + s_{t-m} + \varepsilon_t$ $\ell_t = \ell_{t-1} + b_{t-1} + \alpha \varepsilon_t$ $b_t = b_{t-1} + \beta \varepsilon_t$ $s_t = s_{t-m} + \gamma \varepsilon_t$ | $y_t = (\ell_{t-1} + b_{t-1}) s_{t-m} + \varepsilon_t$ $\ell_t = \ell_{t-1} + b_{t-1} + \alpha \varepsilon_t / s_{t-m}$ $b_t = b_{t-1} + \beta \varepsilon_t / s_{t-m}$ $s_t = s_{t-m} + \gamma \varepsilon_t / (\ell_{t-1} + b_{t-1})$ |
| A _d | $y_t = \ell_{t-1} + \phi b_{t-1} + \varepsilon_t$ $\ell_t = \ell_{t-1} + \phi b_{t-1} + \alpha \varepsilon_t$ $b_t = \phi b_{t-1} + \beta \varepsilon_t$ | $y_t = \ell_{t-1} + \phi b_{t-1} + s_{t-m} + \varepsilon_t$ $\ell_t = \ell_{t-1} + \phi b_{t-1} + \alpha \varepsilon_t$ $b_t = \phi b_{t-1} + \beta \varepsilon_t$ $s_t = s_{t-m} + \gamma \varepsilon_t$ | $y_t = (\ell_{t-1} + \phi b_{t-1}) s_{t-m} + \varepsilon_t$ $\ell_t = \ell_{t-1} + \phi b_{t-1} + \alpha \varepsilon_t / s_{t-m}$ $b_t = \phi b_{t-1} + \beta \varepsilon_t / s_{t-m}$ $s_t = s_{t-m} + \gamma \varepsilon_t / (\ell_{t-1} + \phi b_{t-1})$ |

MULTIPLICATIVE ERROR MODELS

| Trend | Seasonal | | |
|----------------|--|---|---|
| | N | A | M |
| N | $y_t = \ell_{t-1}(1 + \varepsilon_t)$ $\ell_t = \ell_{t-1}(1 + \alpha \varepsilon_t)$ | $y_t = (\ell_{t-1} + s_{t-m})(1 + \varepsilon_t)$ $\ell_t = \ell_{t-1} + \alpha(\ell_{t-1} + s_{t-m})\varepsilon_t$ $s_t = s_{t-m} + \gamma(\ell_{t-1} + s_{t-m})\varepsilon_t$ | $y_t = \ell_{t-1} s_{t-m}(1 + \varepsilon_t)$ $\ell_t = \ell_{t-1}(1 + \alpha \varepsilon_t)$ $s_t = s_{t-m}(1 + \gamma \varepsilon_t)$ |
| A | $y_t = (\ell_{t-1} + b_{t-1})(1 + \varepsilon_t)$ $\ell_t = (\ell_{t-1} + b_{t-1})(1 + \alpha \varepsilon_t)$ $b_t = b_{t-1} + \beta(\ell_{t-1} + b_{t-1})\varepsilon_t$ | $y_t = (\ell_{t-1} + b_{t-1} + s_{t-m})(1 + \varepsilon_t)$ $\ell_t = \ell_{t-1} + b_{t-1} + \alpha(\ell_{t-1} + b_{t-1} + s_{t-m})\varepsilon_t$ $b_t = b_{t-1} + \beta(\ell_{t-1} + b_{t-1} + s_{t-m})\varepsilon_t$ $s_t = s_{t-m} + \gamma(\ell_{t-1} + b_{t-1} + s_{t-m})\varepsilon_t$ | $y_t = (\ell_{t-1} + b_{t-1}) s_{t-m}(1 + \varepsilon_t)$ $\ell_t = (\ell_{t-1} + b_{t-1})(1 + \alpha \varepsilon_t)$ $b_t = b_{t-1} + \beta(\ell_{t-1} + b_{t-1})\varepsilon_t$ $s_t = s_{t-m}(1 + \gamma \varepsilon_t)$ |
| A _d | $y_t = (\ell_{t-1} + \phi b_{t-1})(1 + \varepsilon_t)$ $\ell_t = (\ell_{t-1} + \phi b_{t-1})(1 + \alpha \varepsilon_t)$ $b_t = \phi b_{t-1} + \beta(\ell_{t-1} + \phi b_{t-1})\varepsilon_t$ | $y_t = (\ell_{t-1} + \phi b_{t-1} + s_{t-m})(1 + \varepsilon_t)$ $\ell_t = \ell_{t-1} + \phi b_{t-1} + \alpha(\ell_{t-1} + \phi b_{t-1} + s_{t-m})\varepsilon_t$ $b_t = \phi b_{t-1} + \beta(\ell_{t-1} + \phi b_{t-1} + s_{t-m})\varepsilon_t$ $s_t = s_{t-m} + \gamma(\ell_{t-1} + \phi b_{t-1} + s_{t-m})\varepsilon_t$ | $y_t = (\ell_{t-1} + \phi b_{t-1}) s_{t-m}(1 + \varepsilon_t)$ $\ell_t = (\ell_{t-1} + \phi b_{t-1})(1 + \alpha \varepsilon_t)$ $b_t = \phi b_{t-1} + \beta(\ell_{t-1} + \phi b_{t-1})\varepsilon_t$ $s_t = s_{t-m}(1 + \gamma \varepsilon_t)$ |

Вопрос: какую использовать?

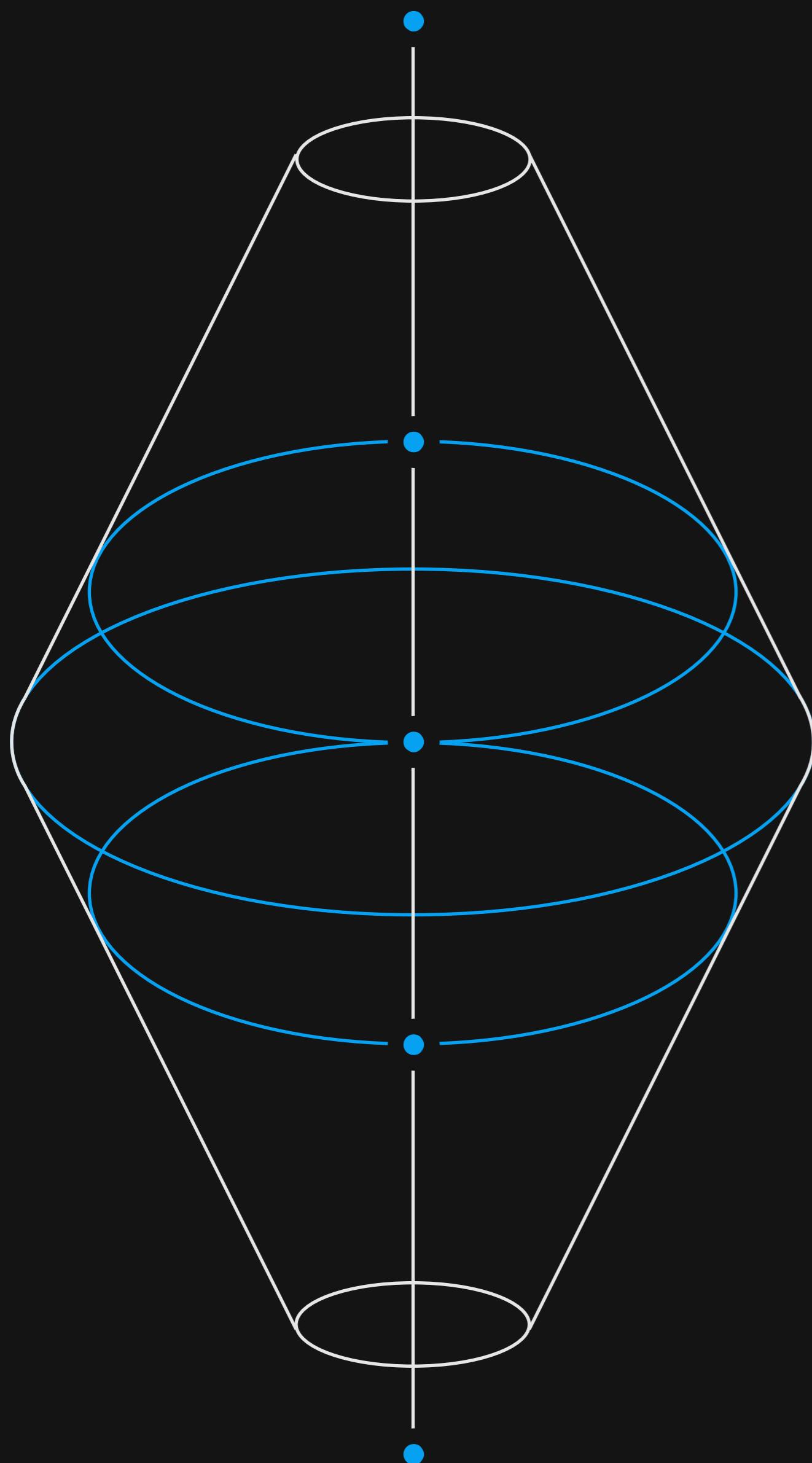
Ответ: подберём автоматически.

```
1 StatsForecastAutoETSModel(
2     season_length: int
3     model: str,
4     damped: bool
5 )
```

$$AIC = -2 \log(L) + 2k,$$

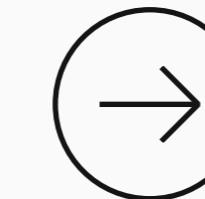
L — функция правдоподобия;
 k — количество параметров.

ETS (итог)



Когда использовать:

- месячные/квартальные данные;
- есть выраженный тренд/сезонность.



Как использовать:

- прогнозы (Forecasting);
- разложение на компоненты (Filtering).

3. Prophet

Prophet

$$y(t) = g(t) + s(t) + h(t) + \epsilon_t.$$

$g(t)$ — тренд;

$s(t)$ — сезонность;

$h(t)$ — праздники;

ϵ_t — ошибки.

Выглядит как ETS, но:

- автоматическое определение точек смены поведения;
- любое количество сезонностей;
- дополнительные регрессоры + праздники;
- тонкая настройка компонент.

→ [Статья](#)

→ [Документация](#)

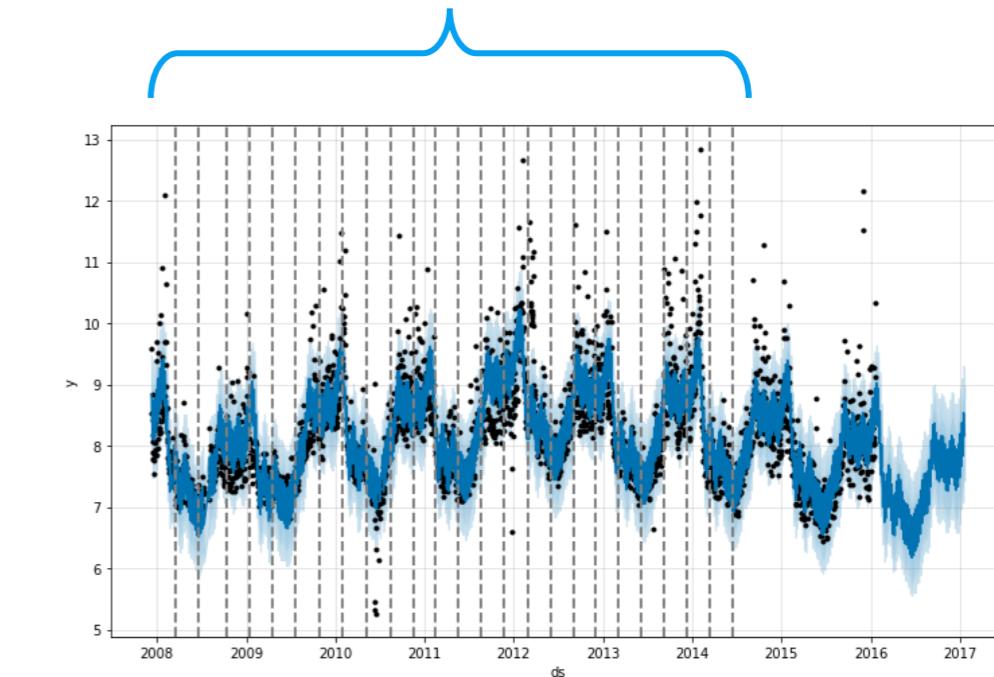
Prophet (тренд)

```

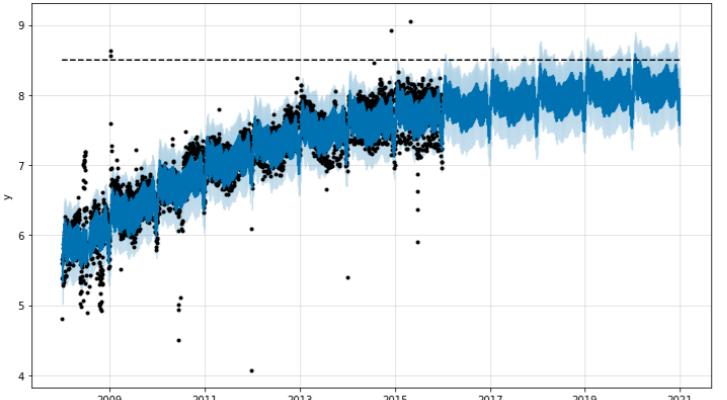
1 ProphetModel(
2     growth: str = "linear",
3     changepoints: Optional[List[datetime]] = None,
4     n_changepoints: int = 25,
5     changepoint_range: float = 0.8,
6     changepoint_prior_scale: float = 0.05,
7 )

```

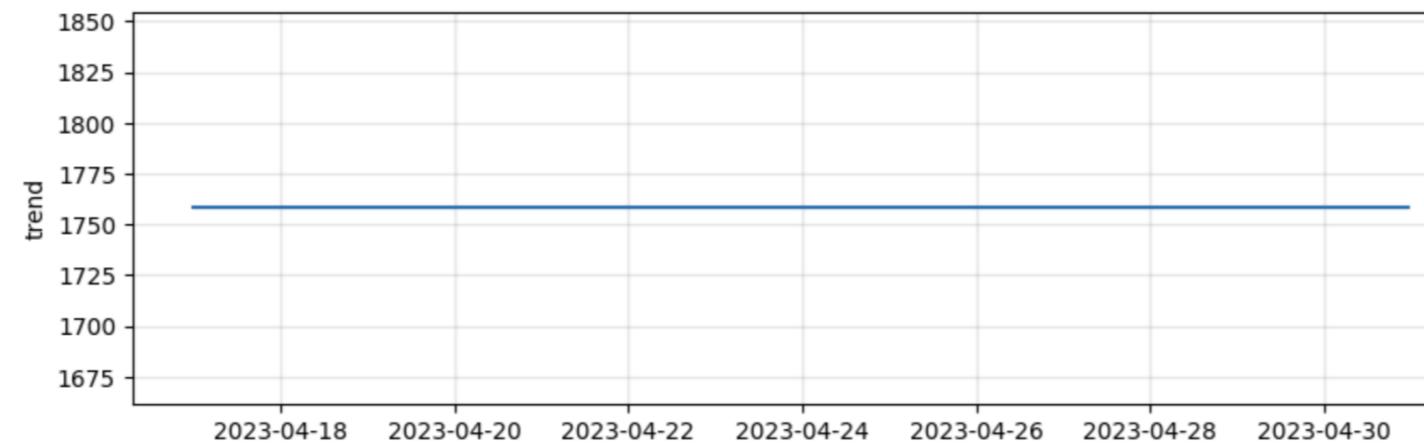
`change_point_range`



1. Равномерно разбрасываем `n_changepoints`.



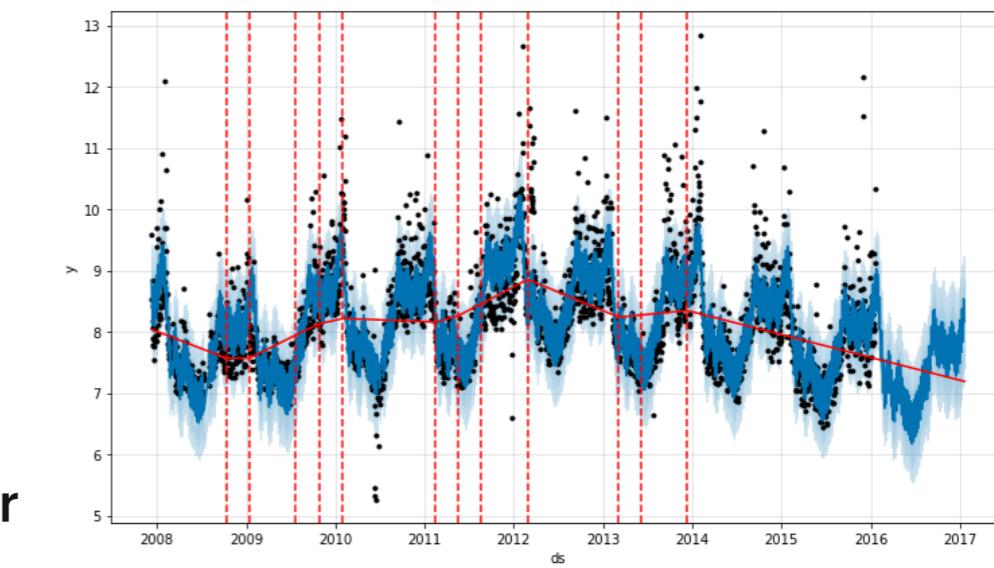
logistic



flat

growth

linear



2. Количество выбранных точек зависит от `changepoint_prior_scale`.

Prophet (сезонность)

```

1 ProphetModel(
2     yearly_seasonality: Union[str, bool] = "auto",
3     weekly_seasonality: Union[str, bool] = "auto",
4     daily_seasonality: Union[str, bool] = "auto",
5     seasonality_mode: str = "additive",
6     seasonality_prior_scale: float = 10.0,
7 )

```

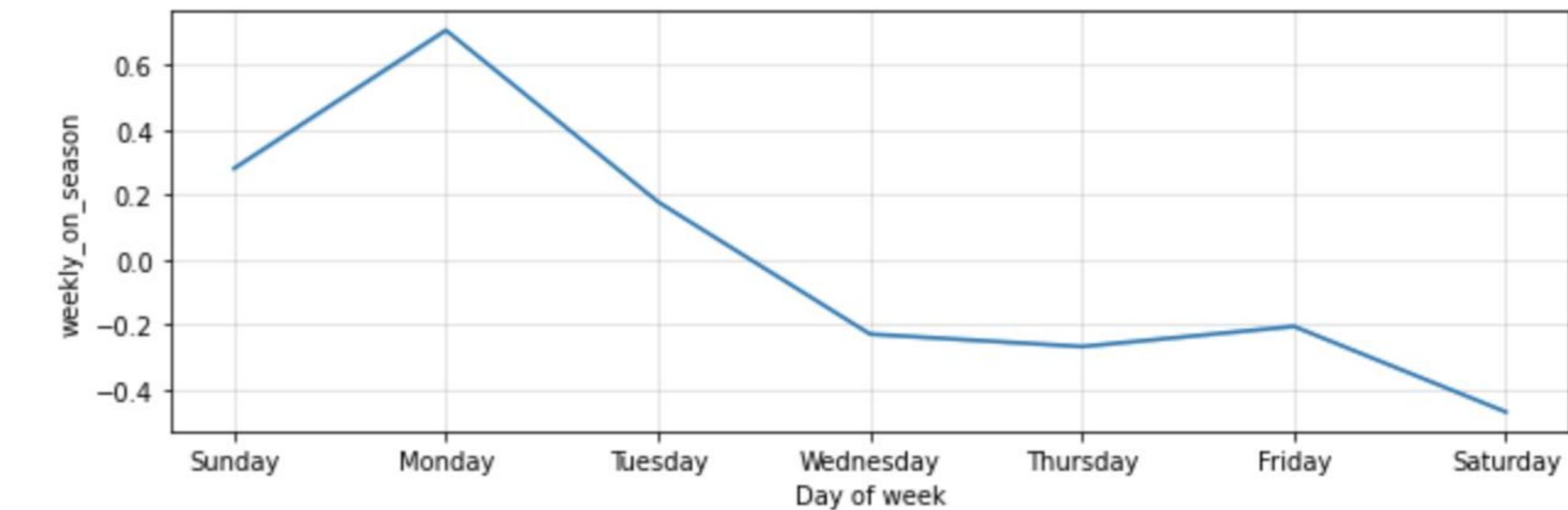
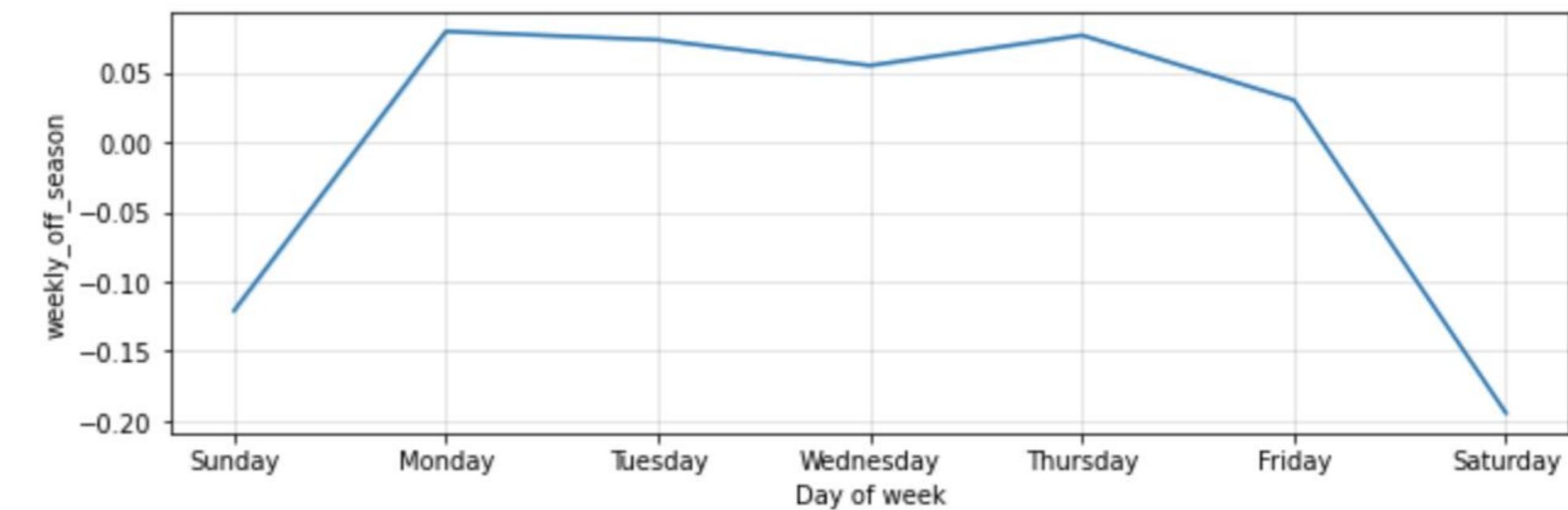
Моделируем через ряд Фурье

$$s(t) = \sum_{n=1}^N \left(a_n \cos\left(\frac{2\pi n t}{P}\right) + b_n \sin\left(\frac{2\pi n t}{P}\right) \right)$$

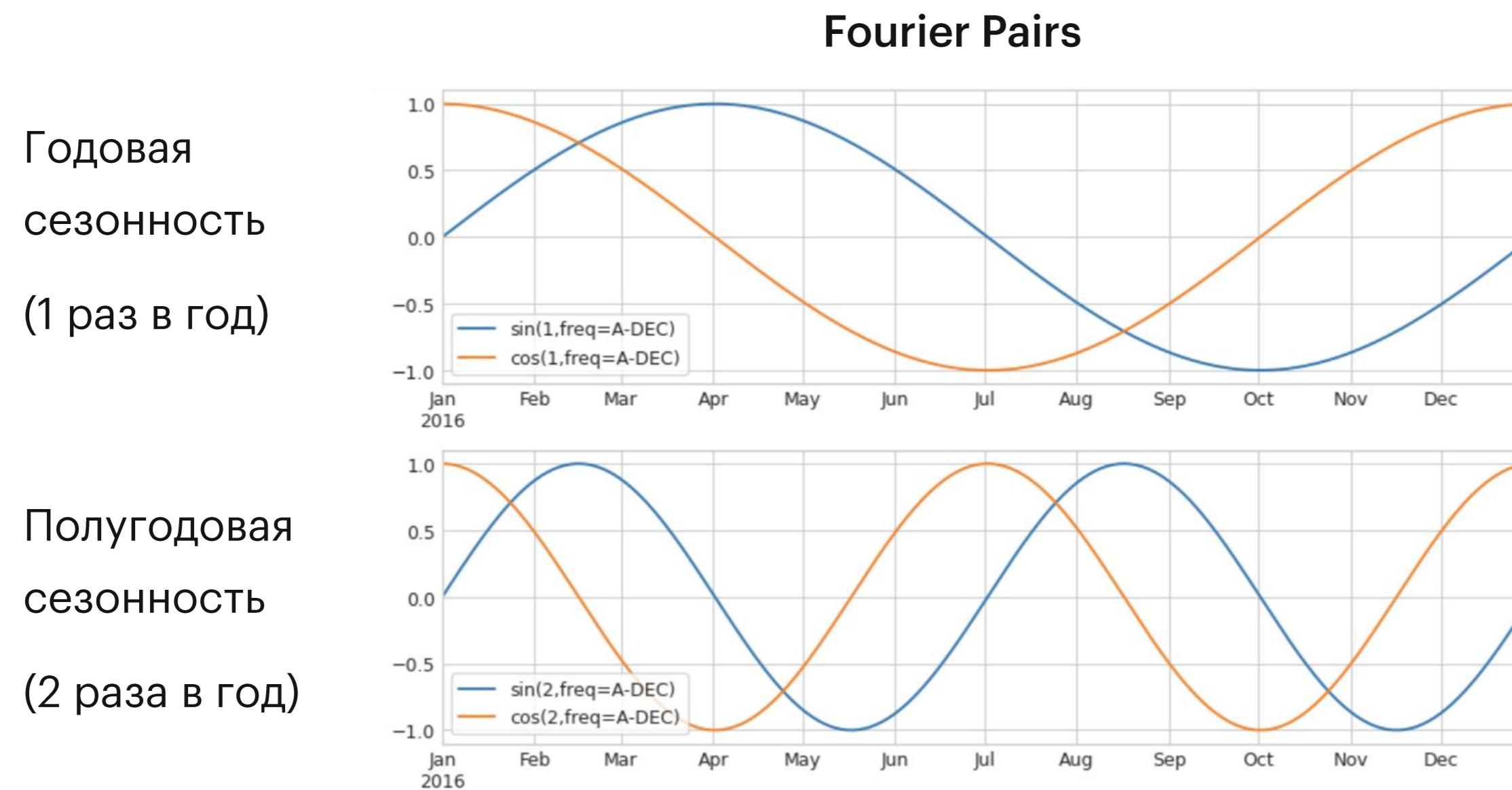
- **daily_seasonality** — количество членов разложения (больше значение, больше оверфит).
- **seasonality_mode** — аддитивная/мультипликативная.
- **seasonality_prior_scale** — регуляризация на сезонную компоненту.

Можно:

- сразу много сезонностей;
- сезонность любого периода;
- условная сезонность (летом и зимой — разный недельный паттерн для мороженого).



Prophet (сезонность)



$$s(t) = \sum_{n=1}^N \left(a_n \cos\left(\frac{2\pi n t}{P}\right) + b_n \sin\left(\frac{2\pi n t}{P}\right) \right) = X(t)\beta$$

$$X(t) = \left[\cos\left(\frac{2\pi t \cdot 1}{365.25}\right), \dots, \sin\left(\frac{2\pi t \cdot 10}{365.25}\right) \right]$$

Prophet (праздники + регрессоры)

```

1 ProphetModel(
2     holidays: Optional[pd.DataFrame] = None,
3     holidays_prior_scale: float = 10.0,
4 )

```

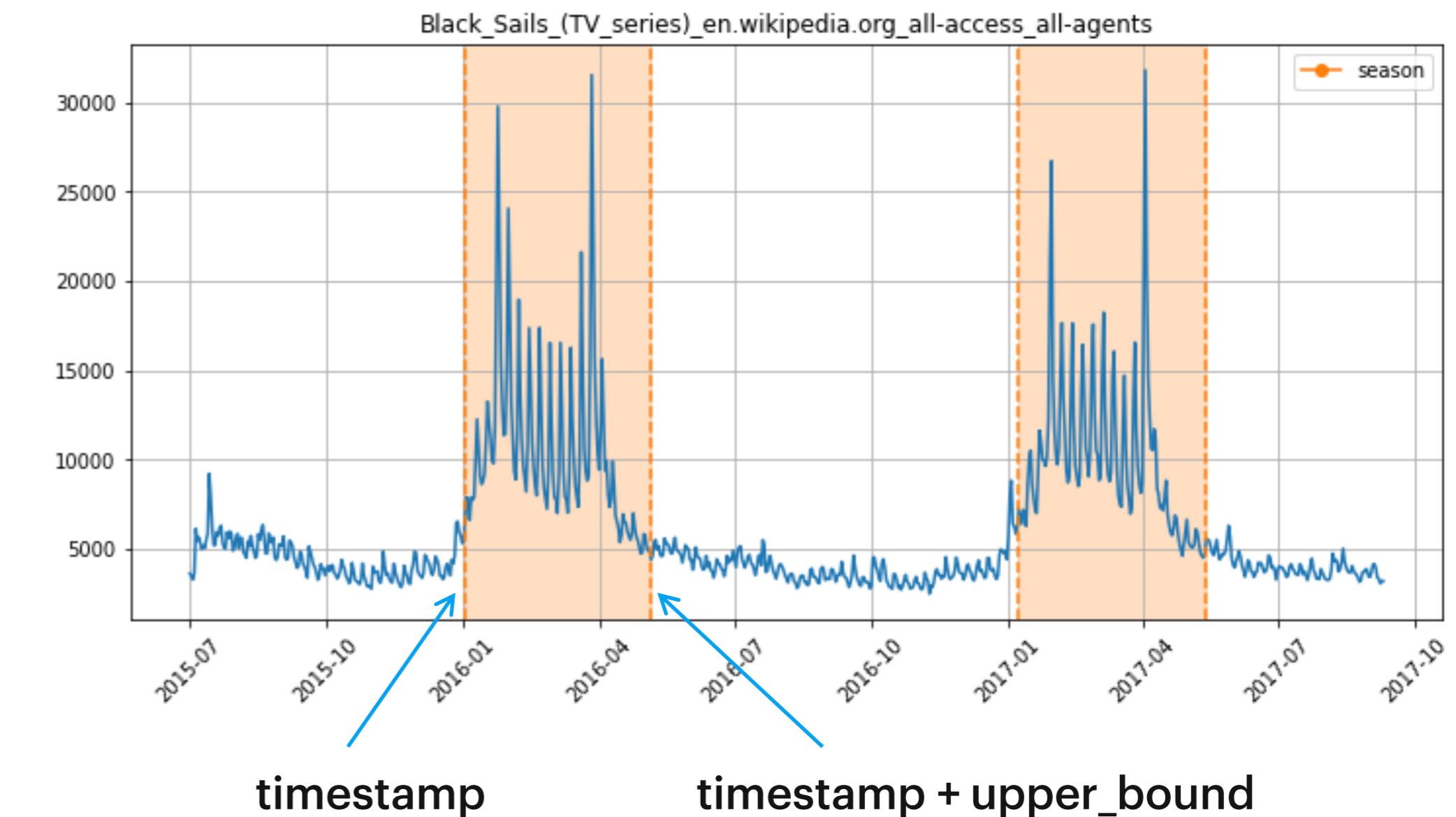
- **holidays** — описание праздников в формате (holiday, timestamp, lower_bound, upper_bound).
- **holidays_prior_scale** — регуляризация на компоненту праздников.

Сюда же можно добавить другие регрессоры.

$$h(t) = \text{holidays}(t) + x_1(t) + x_2(t) + \dots$$

holidays(t) — бинарный вектор наличия праздника.

Событие: выход нового сезона

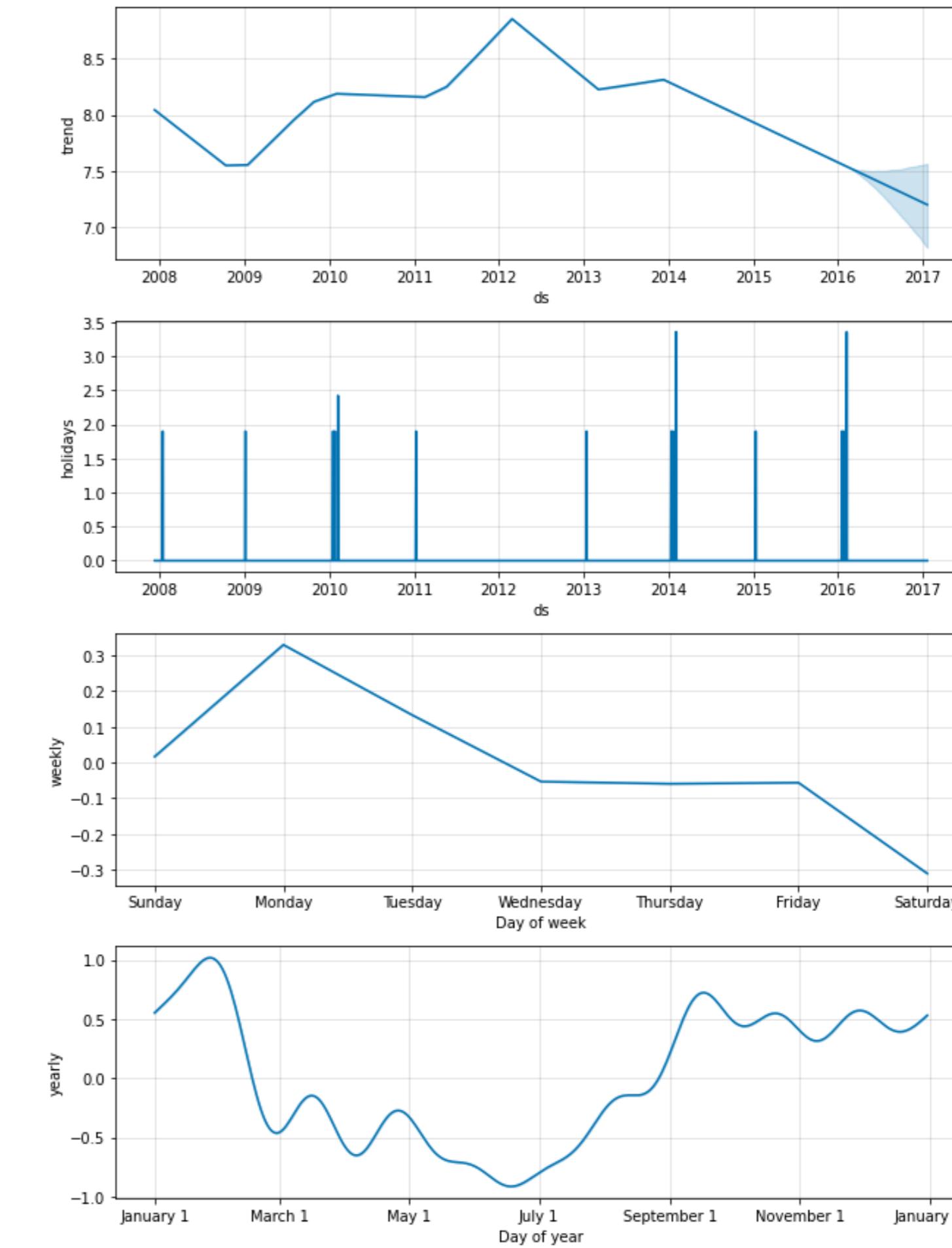
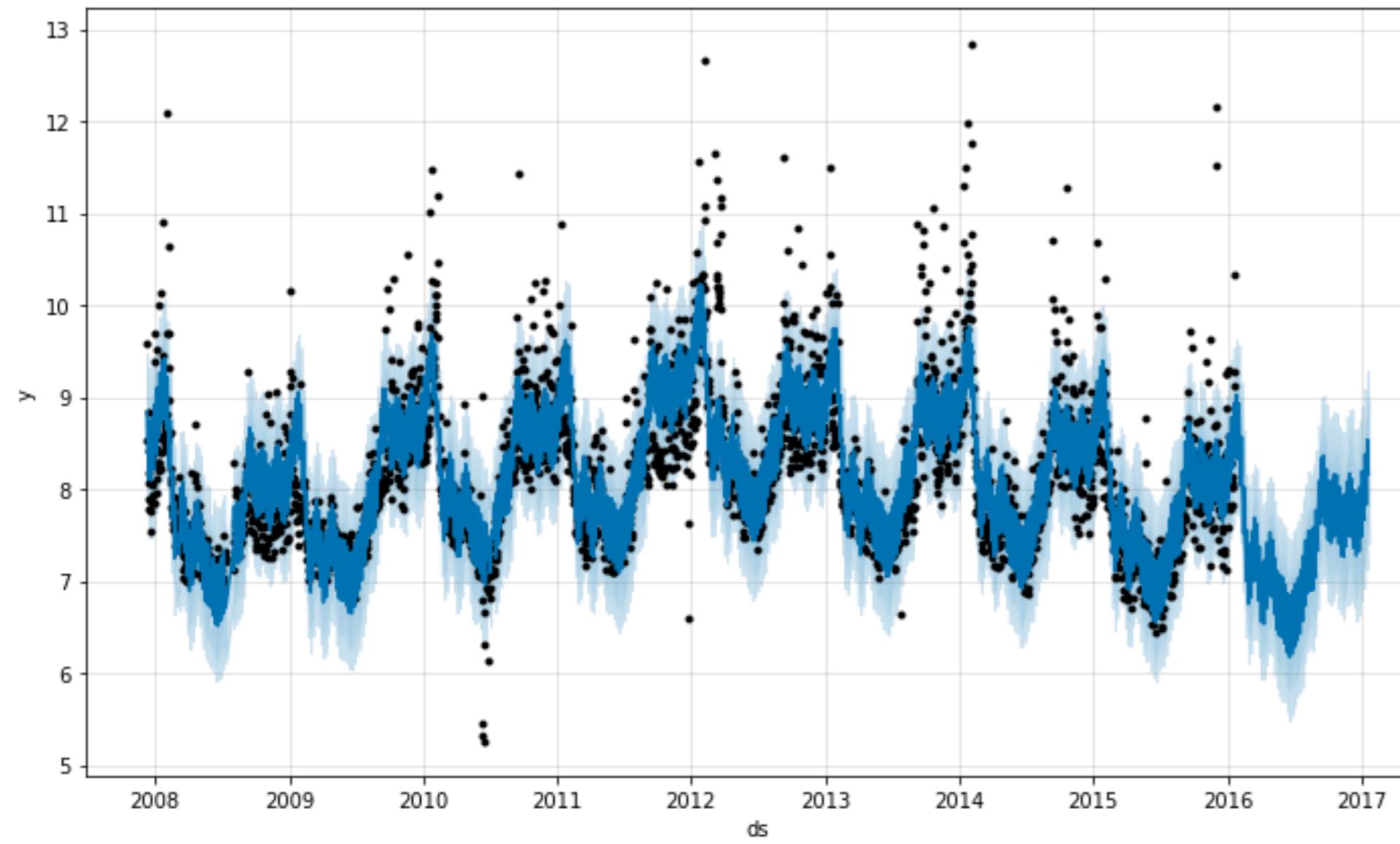


Prophet (итоги)



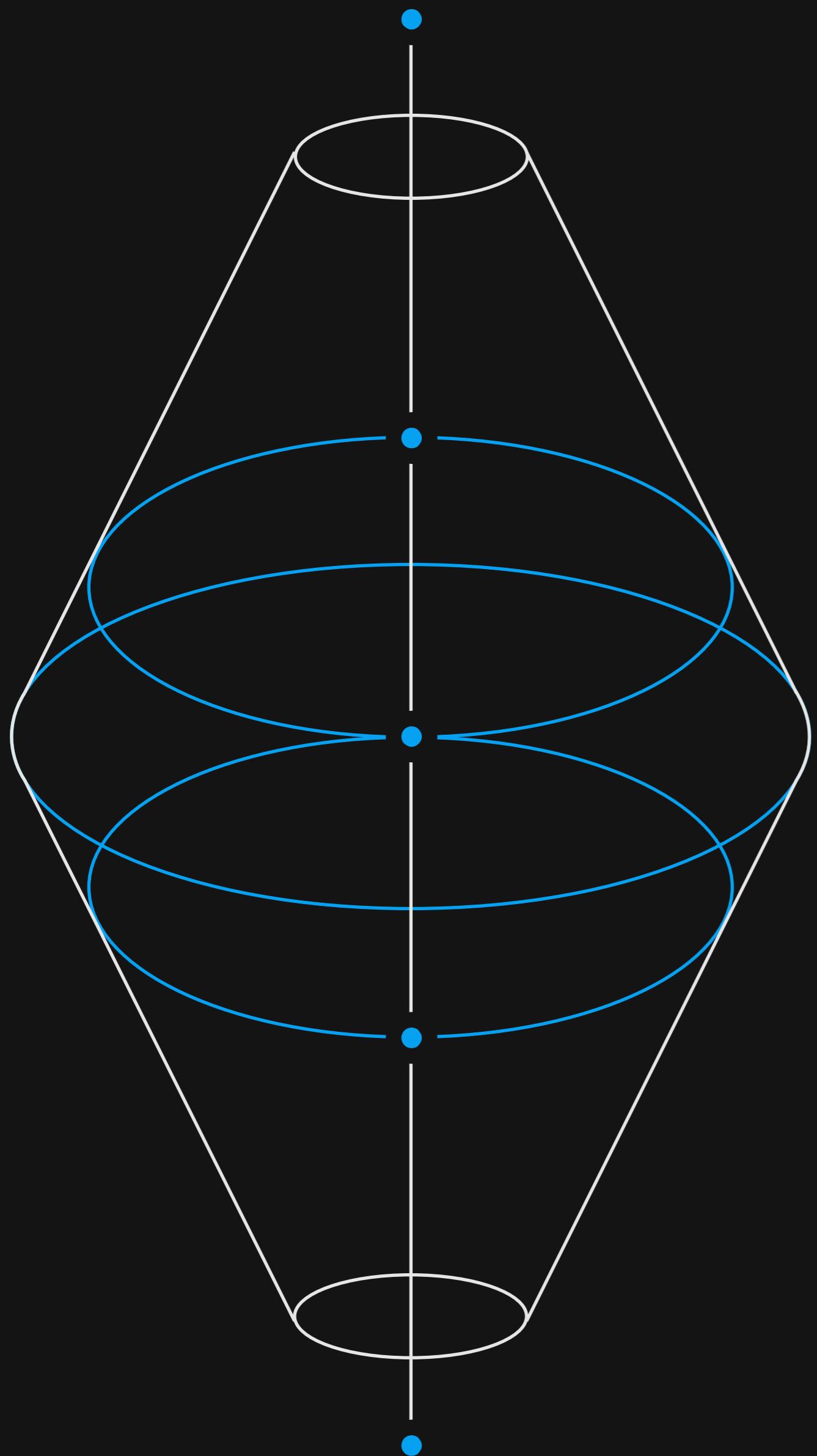
Когда использовать:

- дневные данные;
- есть точки смены поведения;
- есть несколько сезонностей;
- есть влияние праздников и других регрессоров;
- важна интерпретируемость.



4. SARIMAX

SARIMAX



- **S** = Seasonal
- **AR** = AutoRegressive
- **I** = Integrated
- **MA** = Moving average
- **X** = with eXogenous regressors

ARIMA

AR(p) (Autoregression) - модель авторегрессии порядка p .

$$y_t = c + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \cdots + \phi_p y_{t-p} + \varepsilon_t$$

Если объединим две модели, то получим **ARMA(p, q)** модель:

$$y_t = c + \sum_{i=1}^p \phi_i y_{t-i} + \varepsilon_t + \sum_{j=1}^q \theta_j \varepsilon_{t-j}$$

Теоретическая мотивация: любой стационарный в широком смысле процесс можно аппроксимировать с заданной точностью, выбрав необходимые p и q .

MA(q) (Moving average) - модель скользящего среднего порядка q :

$$y_t = c + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \cdots + \theta_q \varepsilon_{t-q}$$

ARIMA(p, d, q) является расширением модели ARMA на нестационарные ряды:

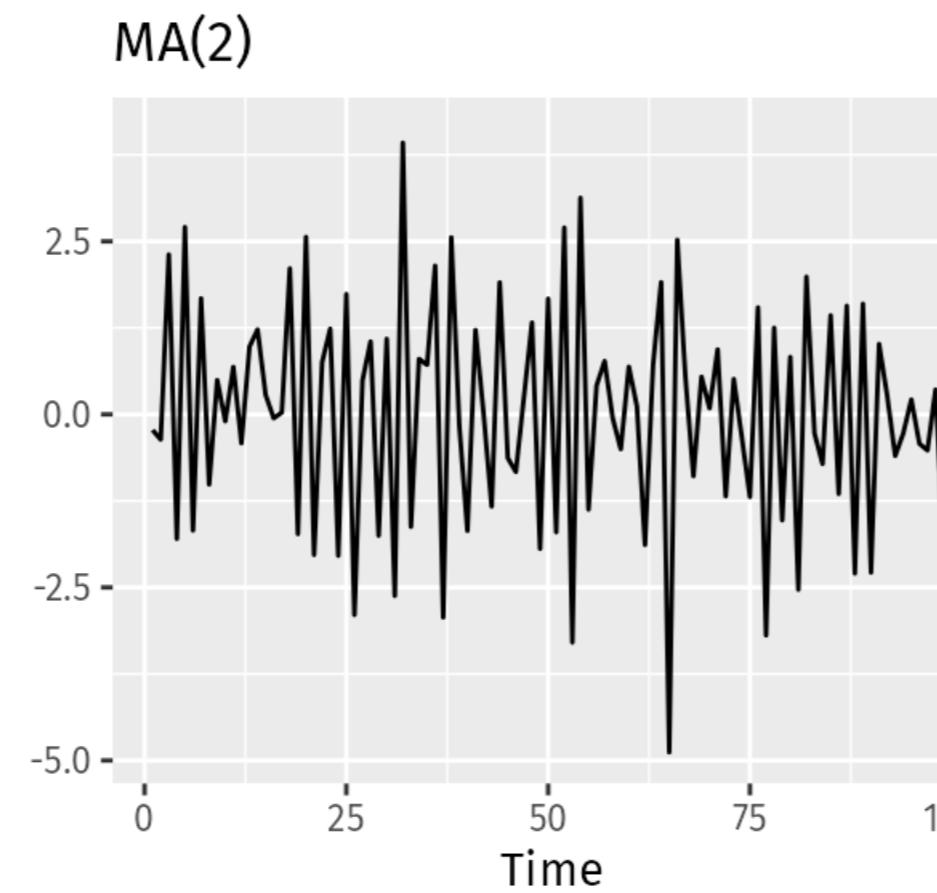
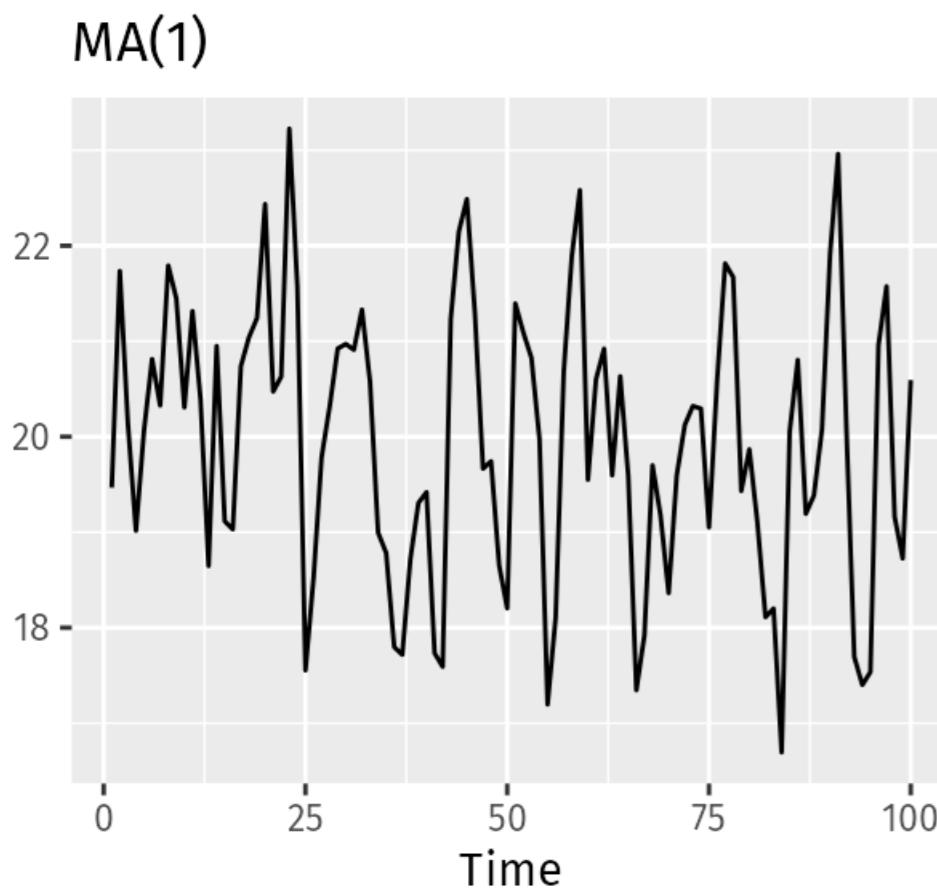
$$\Delta^d y_i = c + \sum_{j=1}^p \phi_j \Delta^d y_{i-j} + \varepsilon_i + \sum_{j=1}^q \theta_j \varepsilon_{i-j}$$

где d - порядок дифференцирования для приведения ряда к стационарному

Тут случайность

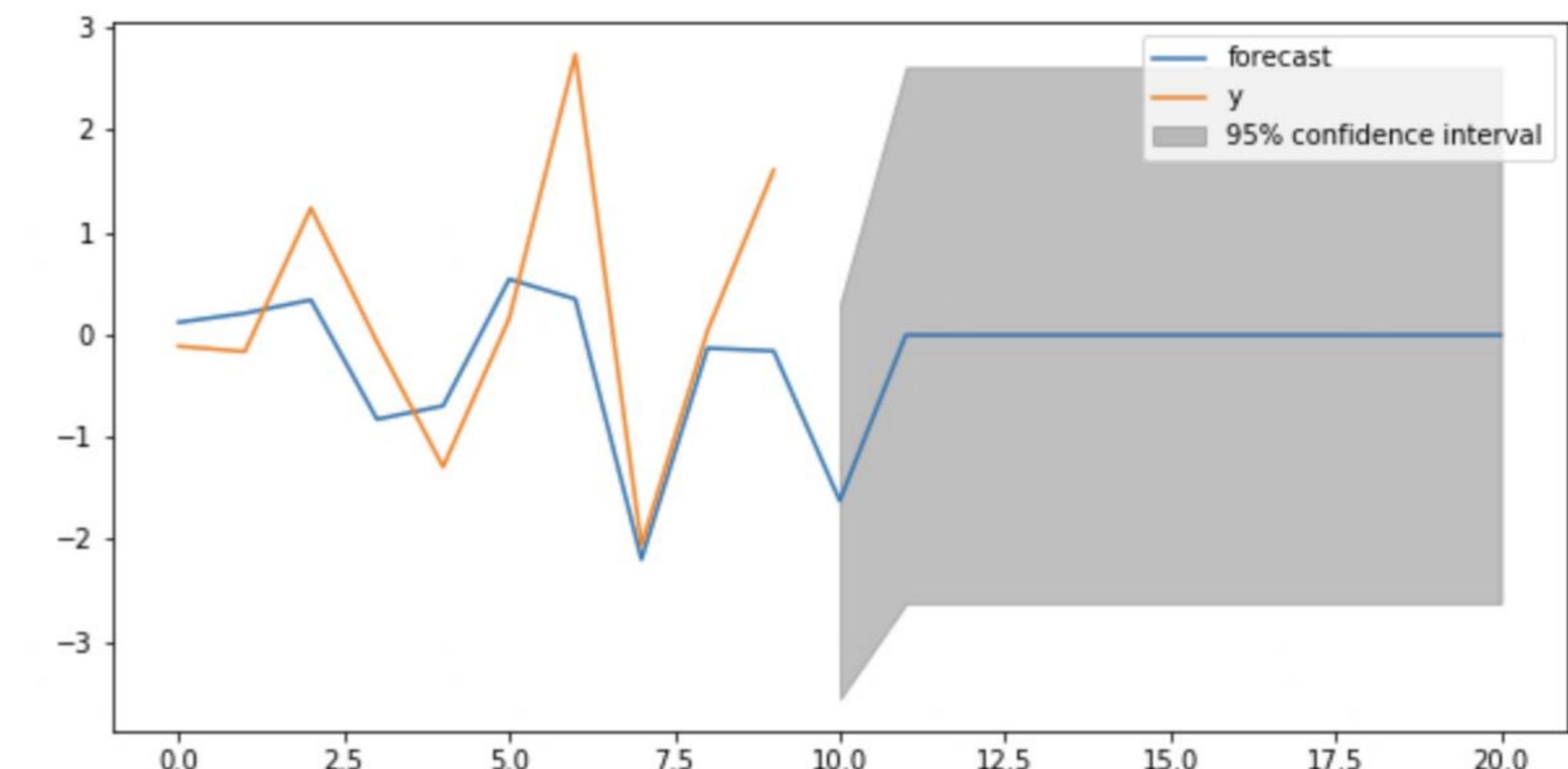


$$y_t = c + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \dots + \theta_q \varepsilon_{t-q},$$



Как прогнозировать

1. Оцениваем все известные ошибки прогноза.
2. Неизвестные зануляем.

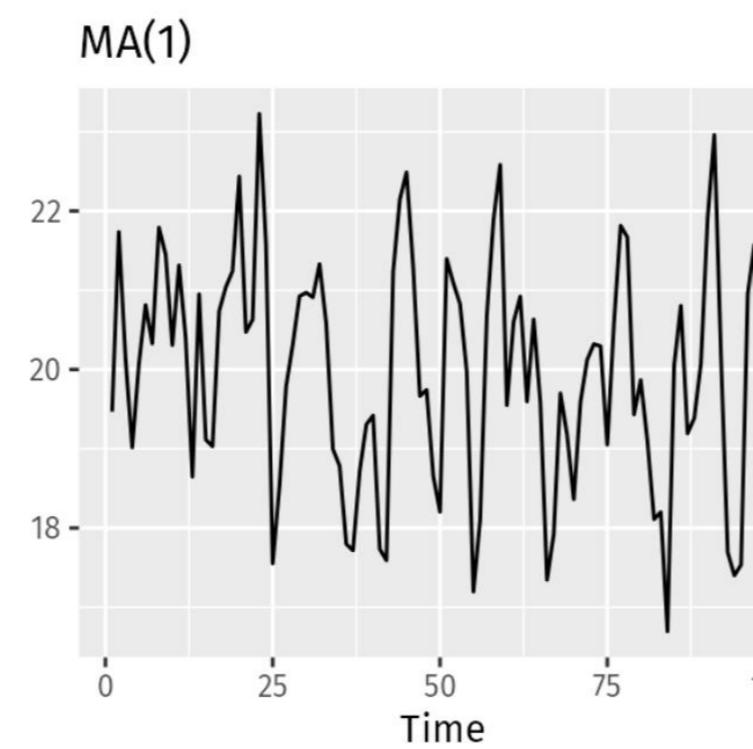
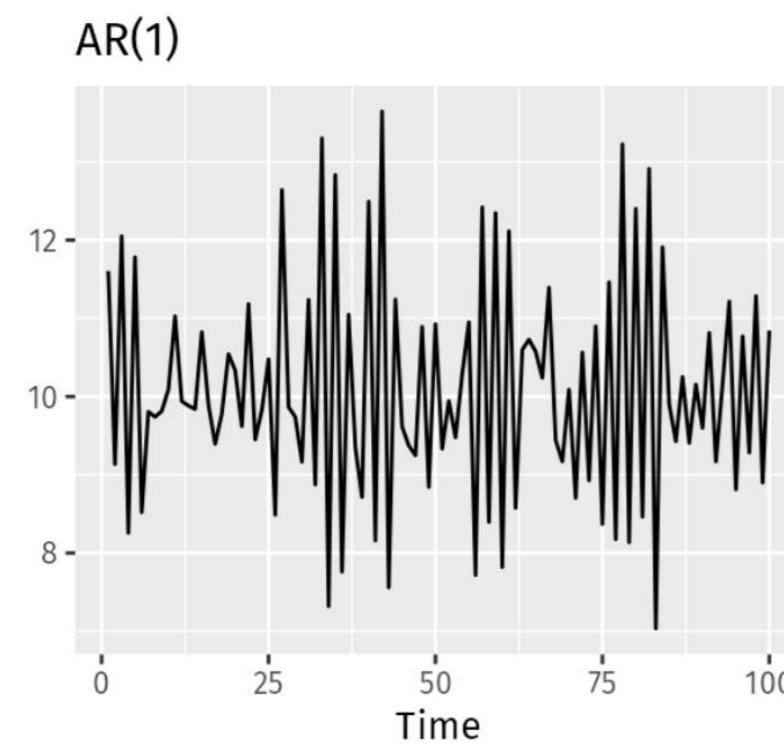


МА(1): всё дальше одного шага в будущее — константа.

SARIMAX

$$y_t = c + \underbrace{\sum_{n=1}^p \alpha_n y_{t-n}}_{\text{AR}(p)} + \underbrace{\sum_{n=1}^q \theta_n \epsilon_{t-n}}_{\text{MA}(q)} + \underbrace{\sum_{n=1}^P \phi_n y_{t-sn}}_{\text{AR}(P)} + \underbrace{\sum_{n=1}^Q \eta_n \epsilon_{t-sn}}_{\text{MA}(Q)} + \epsilon_t + X_t$$

S = Seasonal X = другие фичи



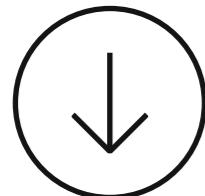
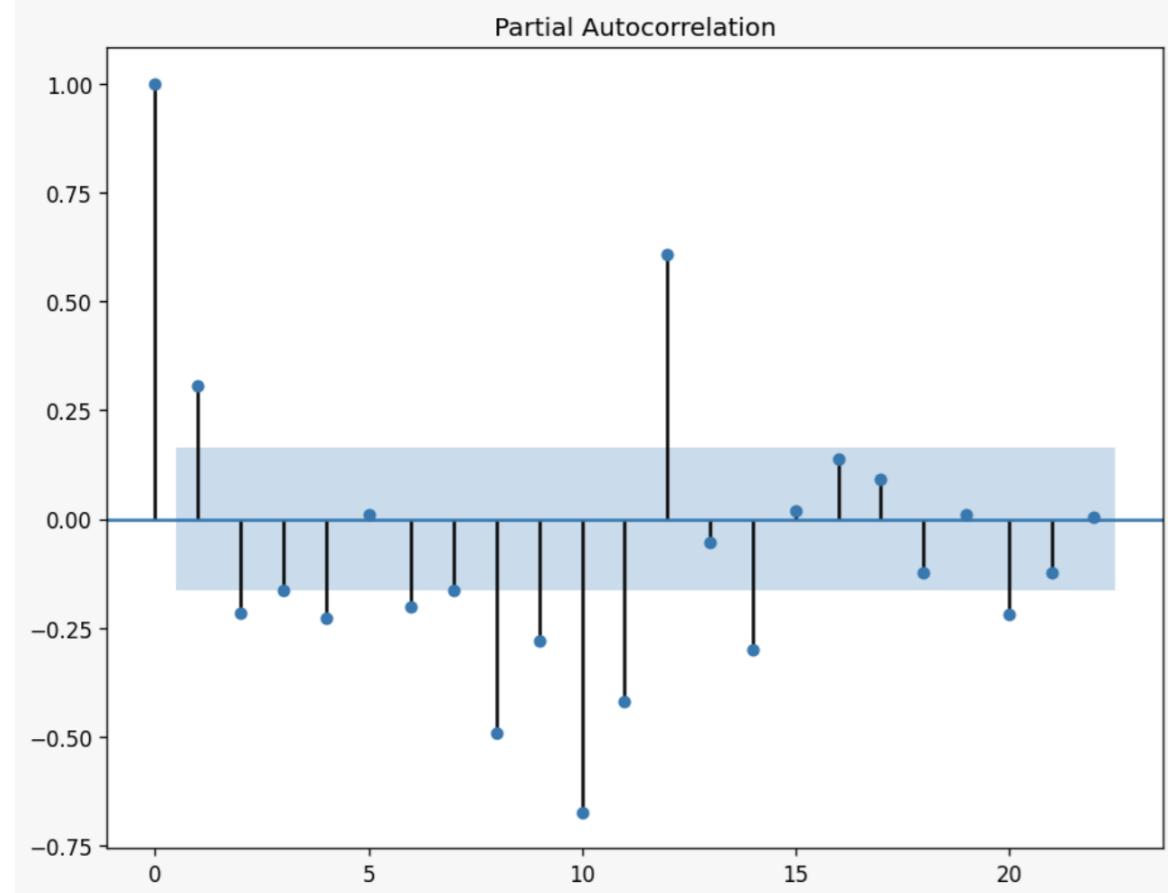
Важна стационарность!

Делаем замену. d раз $\hat{y}_t = y_t - y_{t-1}$

D раз $\hat{y}_t = y_t - y_{t-s}$

SARIMAX (как подбирать параметры)

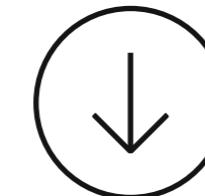
ACF(q)/PACF(p)



Берём последний значимый.

Автоматически

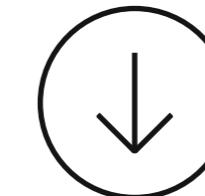
```
1 StatsForecastAutoARIMAModel(  
2     max_p: int = 5,  
3     max_q: int = 5,  
4     max_P: int = 2,  
5     max_Q: int = 2,  
6     max_order: int = 5,  
7     max_d: int = 2,  
8     max_D: int = 1,  
9     season_length: int = 1,  
10    )
```



Выбираем по критерию.
 $AIC = -2 \log(L) + 2(p + q + k + 1)$.

Кросс-валидация

$(1,1,1)(1,1,0,7) \rightarrow 10$ – берём эти.
 $(1,2,1)(2,1,0,7) \rightarrow 20$.
 $(1,1,3)(1,0,0,7) \rightarrow 15$.

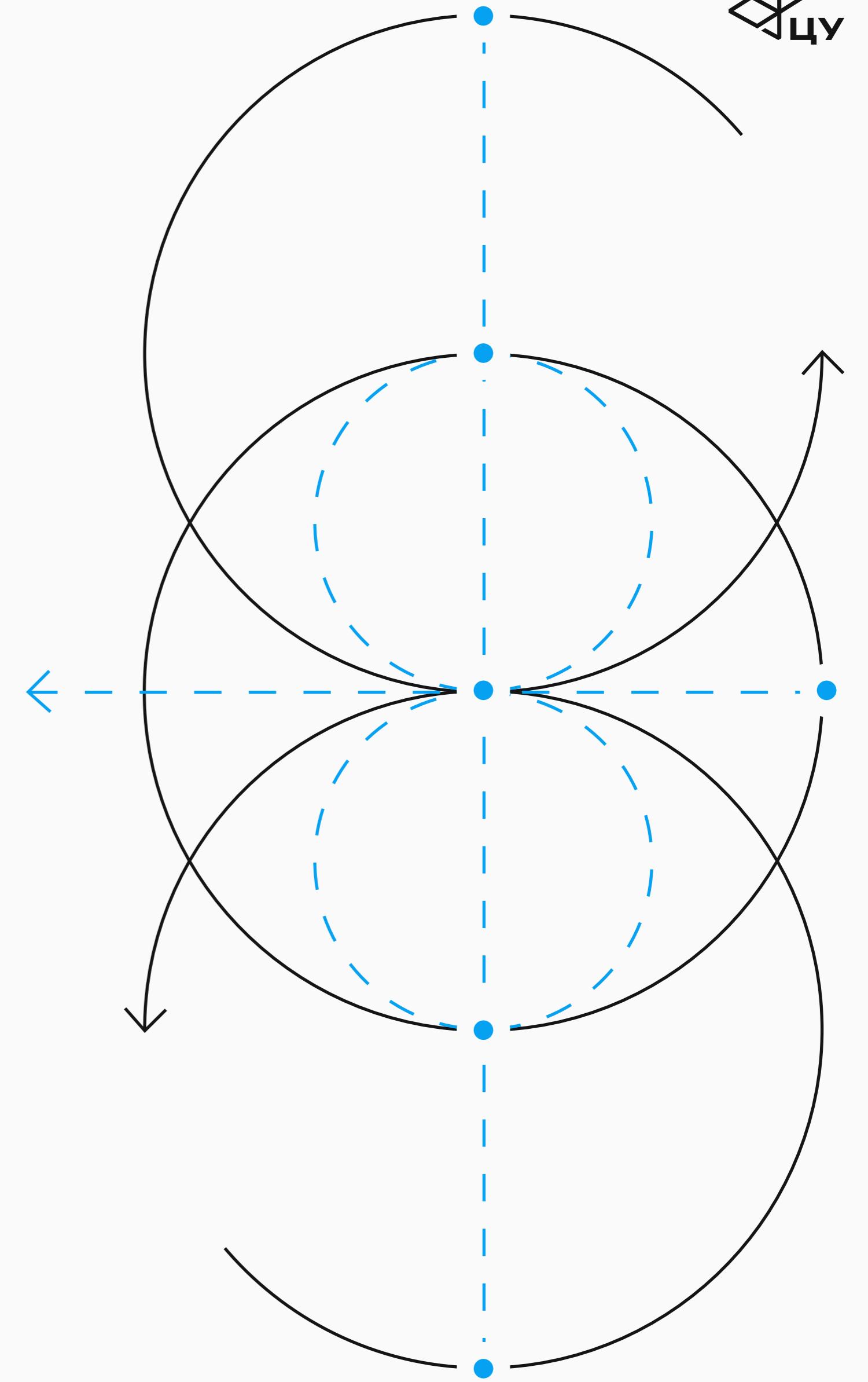


Выбираем по метрикам на валидации.

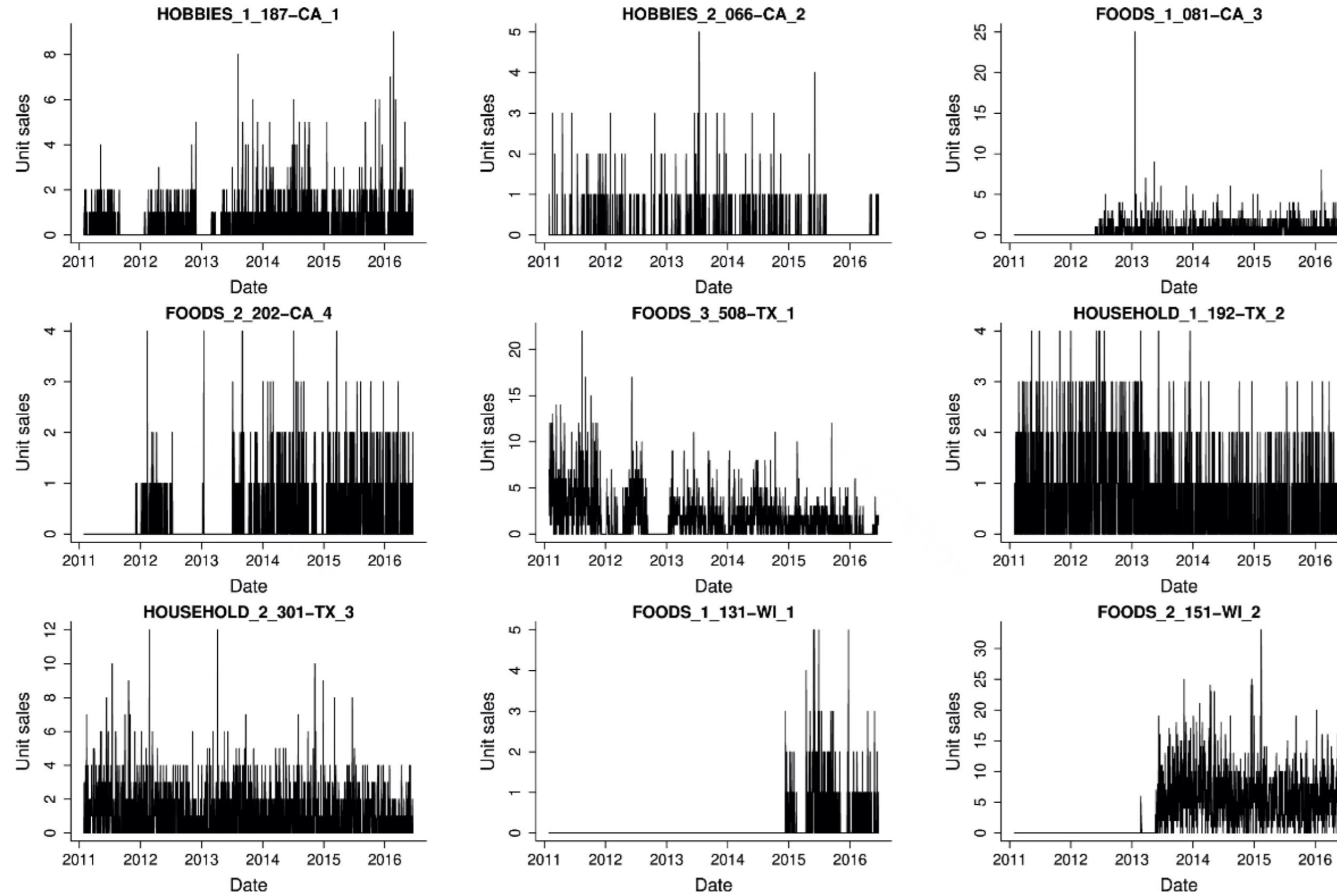
И другие



- **MFLES** (Multiple Frequency Long Exponential Smoothing).
- **CES** (Complex Exponential Smoothing) – модификация ETS с использованием теории комплексных чисел.
- **TBATS** (Trigonometric, Box-Cox transform, ARMA errors, Trend, Seasonal) – модификация ETS с несколькими сезонностями (тригонометрическими) + ошибками из ARMA.
- **Theta** – отдельно моделируются долгосрочные и краткосрочные эффекты.
- **(M)STL** (Multiple Seasonal-Trend Decomposition using LOESS) – декомпозиция ряда на компоненты с выделением нескольких сезонностей.
- **ADIDA/Croston/...** – модели для прерывистого спроса.



Классический датасет из ритейла



- Рядов много.
- Разреженность данных.
- Различие в длине рядов в рамках датасета.
- Поведение рядов разное.

Проблемы «статистических моделей»

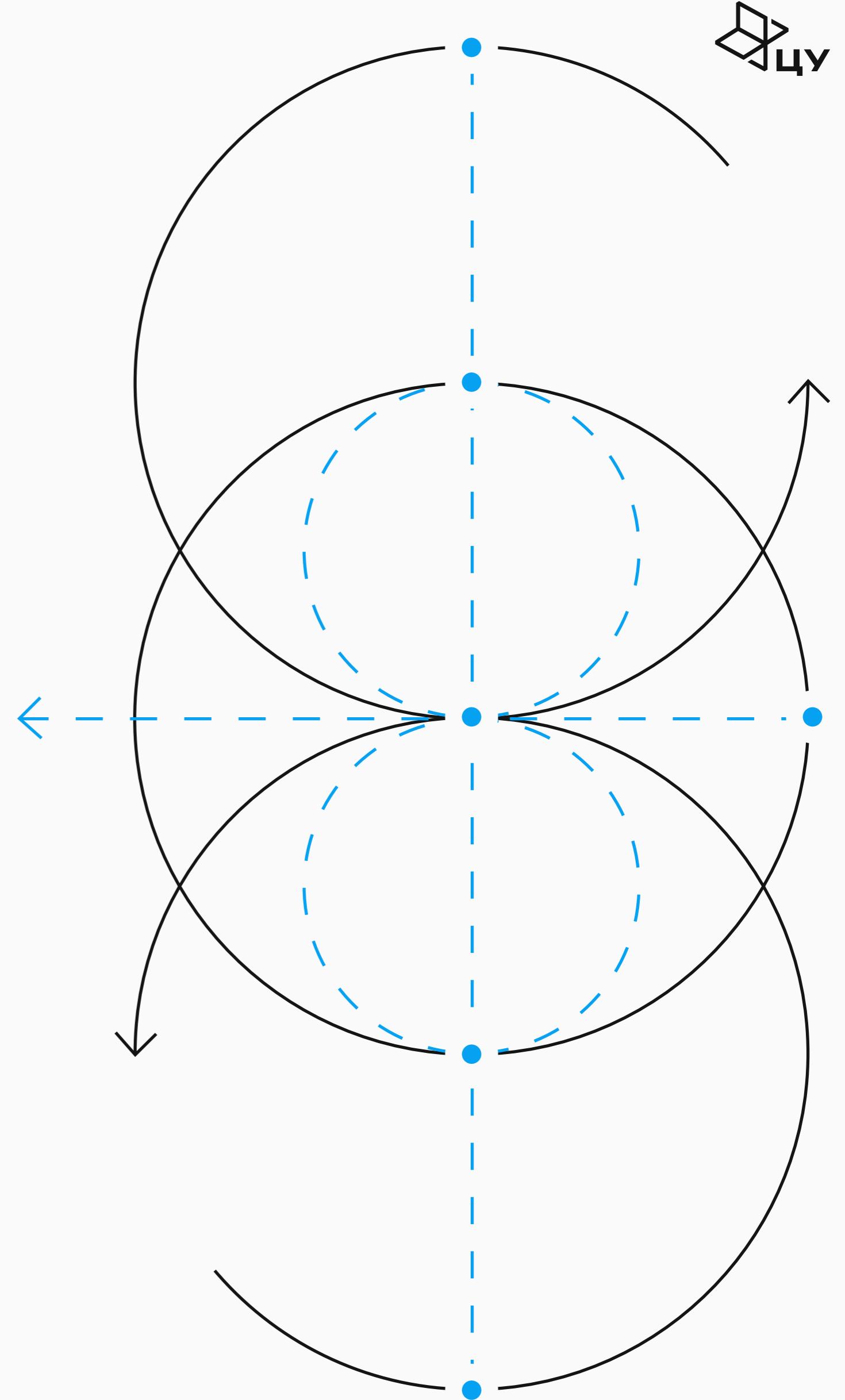


Достоинства

- Можно хорошо обучиться под конкретный домен или ряд.
- Понятно, как параллельно обучать.

Недостатки

- Невозможно работать на больших датасетах.
- Дополнительные данные не используются (не для всех).
- Разреженность данных (в целом и под это подходы есть).
- Различие в длине рядов в рамках датасета.
- Больше требований к количеству и качеству данных.



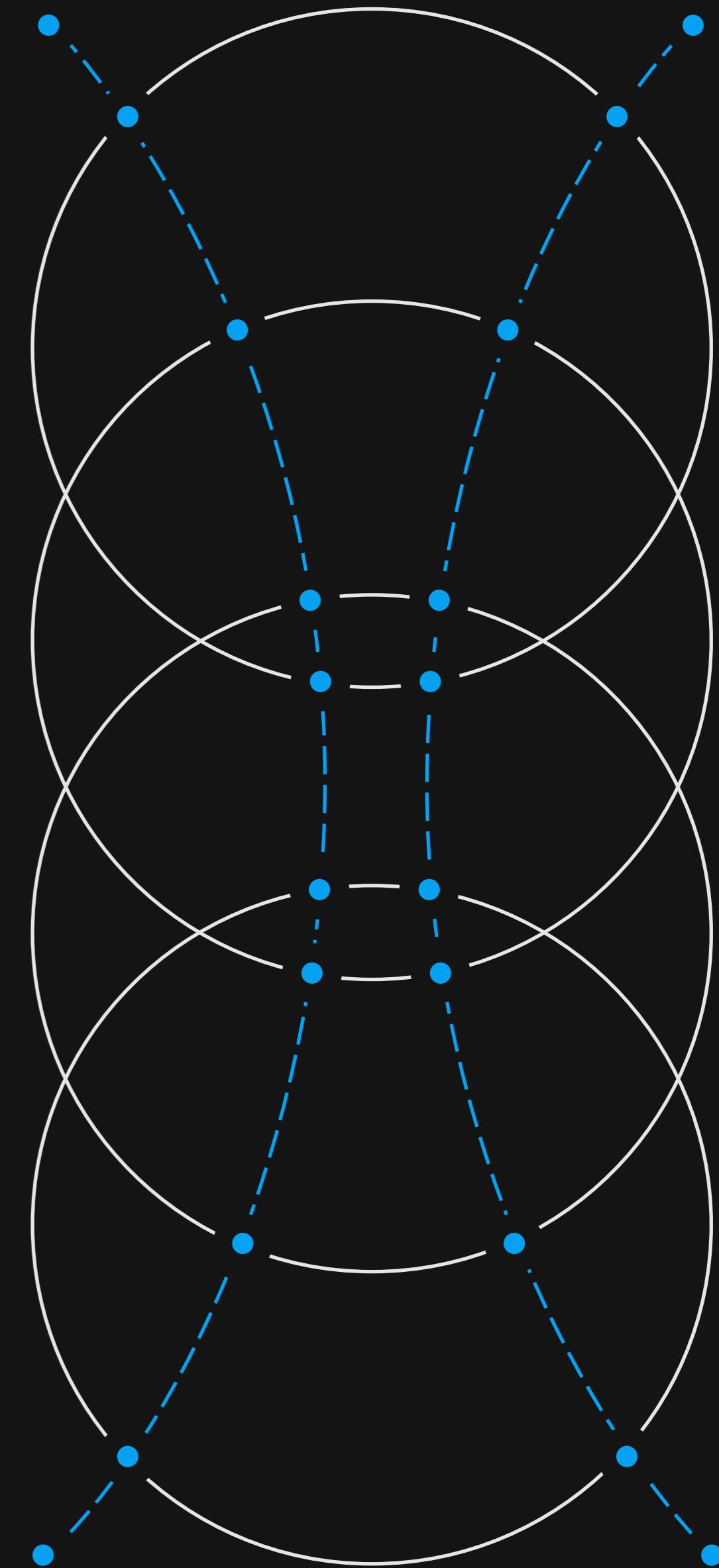
Как же быть?

ML в студию!





Валидация на временных рядах



1. Метрики

Метрики

01

Scale-Dependent

MSE (Mean Squared Error)

$$MSE = \frac{1}{n} \sum_{t=1}^n (y_t - \hat{y}_t)^2$$

MAE (Mean Absolute Error)

$$MAE = \frac{1}{n} \sum_{t=1}^n |y_t - \hat{y}_t|$$

MdAE (Median Absolute Error)

$$MdAE = median(|y_t - \hat{y}_t|)$$

02

Scale-Independent

MAPE (Mean Absolute Percentage Error)

$$MAPE = 100 \times \frac{1}{n} \sum_{t=1}^n \left| \frac{y_t - \hat{y}_t}{y_t} \right|$$

SMAPE (Symmetric Mean Absolute Percentage Error)

$$SMAPE = 100 \times \frac{1}{n} \sum_{t=1}^n \frac{|y_t - \hat{y}_t|}{|y_t| + |\hat{y}_t|}$$

msMAPE (Modified SMAPE)

$$mSMAPE = 100 \times \frac{1}{n} \sum_{t=1}^n \frac{|y_t - \hat{y}_t|}{\max(\frac{|y_t| + |\hat{y}_t| + \epsilon}{2}, 0.5 + \epsilon)}$$

03

Relative

RelRMSE (Relative RMSE)

$$RelRMSE = \sqrt{\frac{MSE}{MSE_{baseline}}}$$

RelMAE (Relative MAE)

$$RelMAE = \frac{MAE}{MAE_{baseline}}$$

MASE (Mean Absolute Scaled Error)

$$MASE = \frac{MAE}{\frac{n}{m-s} \sum_{k=s+1}^m |y_k - y_{k-s}|}$$

Scale Dependent

- Сохраняют масштаб исходных данных (ошибки в штуках).
- Проблемы при агрегации метрик на датасетах с рядами разного масштаба.

01

Scale-Dependent

MSE (Mean Squared Error)

$$MSE = \frac{1}{n} \sum_{t=1}^n (y_t - \hat{y}_t)^2$$

MAE (Mean Absolute Error)

$$MAE = \frac{1}{n} \sum_{t=1}^n |y_t - \hat{y}_t|$$

MdAE (Median Absolute Error)

$$MdAE = median(|y_t - \hat{y}_t|)$$

Scale Independent

- Можно сравнивать качество на рядах разного масштаба.
- Нестабильное поведение в случае близости к нулю.
- Ассиметричные относительно ошибок.

02

Scale-Independent

MAPE (Mean Absolute Percentage Error)

$$MAPE = 100 \times \frac{1}{n} \sum_{t=1}^n \left| \frac{y_t - \hat{y}_t}{y_t} \right|$$

SMAPE (Symmetric Mean Absolute Percentage Error)

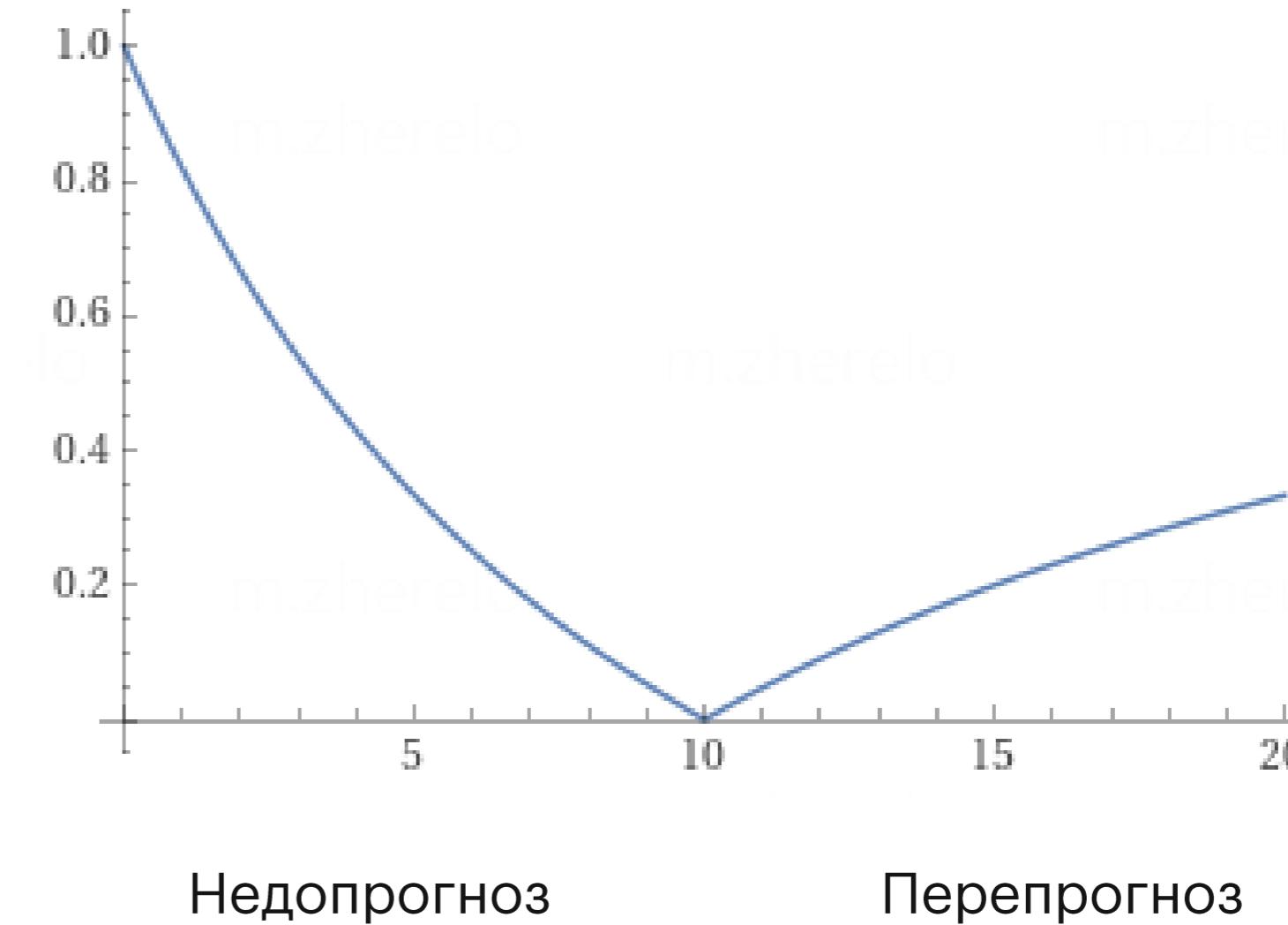
$$SMAPE = 100 \times \frac{1}{n} \sum_{t=1}^n \frac{|y_t - \hat{y}_t|}{\frac{|y_t| + |\hat{y}_t|}{2}}$$

Симметричность

→ **MAPE**: несимметричность относительно аргументов.

MAPE (3,5) ≠ MAPE (5,3)

→ **SMAPE**: недопрогноз наказывается сильнее перепрогноза.



Relative

- Масштабирование на бейзлайн-прогнозе.
- Сложнее в расчёте.
 - RelMAE: масштабирует на test-часть ряда.
 - MASE: требует train-данные.
Для сравнения они должны совпадать.
- Ассиметричные относительно ошибок.

03

Relative

RelRMSE (Relative RMSE)

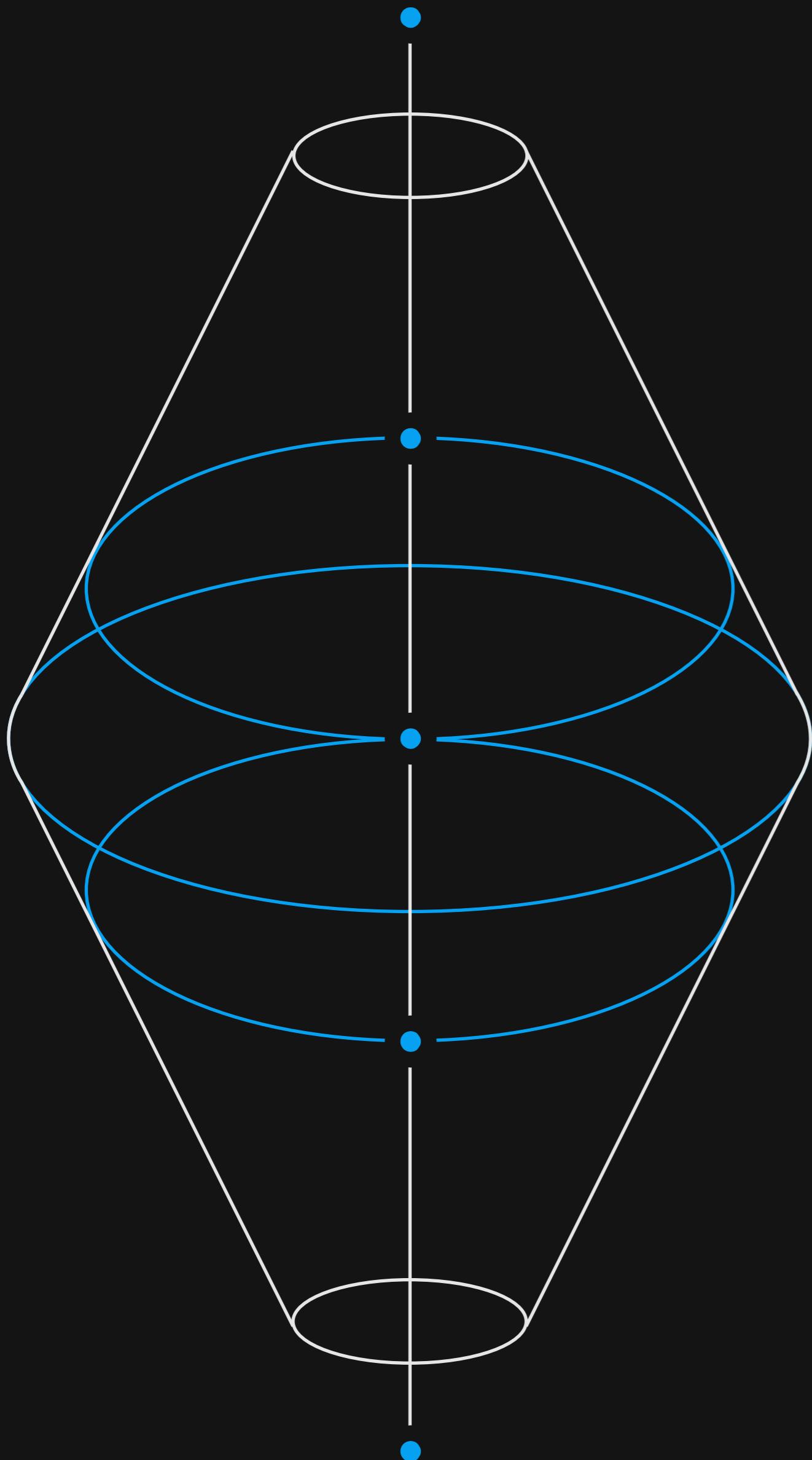
$$\text{RelRMSE} = \sqrt{\frac{MSE}{MSE_{baseline}}}$$

RelMAE (Relative MAE)

$$\text{RelMAE} = \frac{MAE}{MAE_{baseline}}$$

MASE (Mean Absolute Scaled Error)

$$MASE = \frac{MAE}{\frac{n}{m-s} \sum_{k=s+1}^m |y_k - \hat{y}_k|}$$

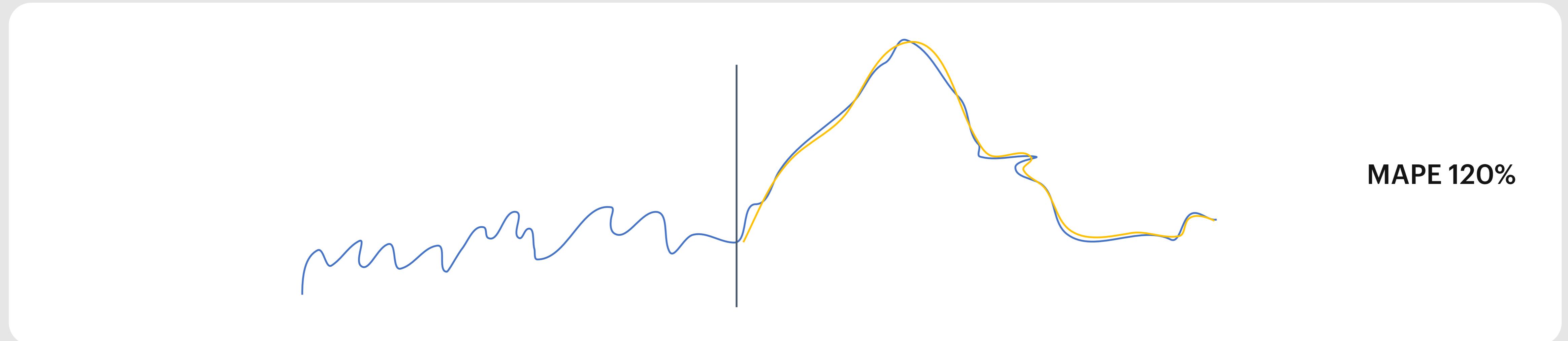
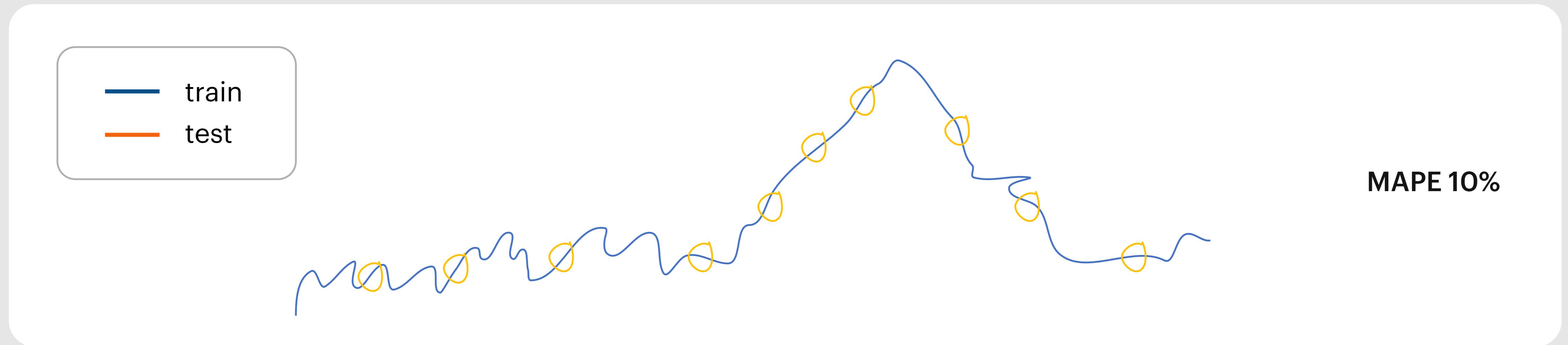


Выводы

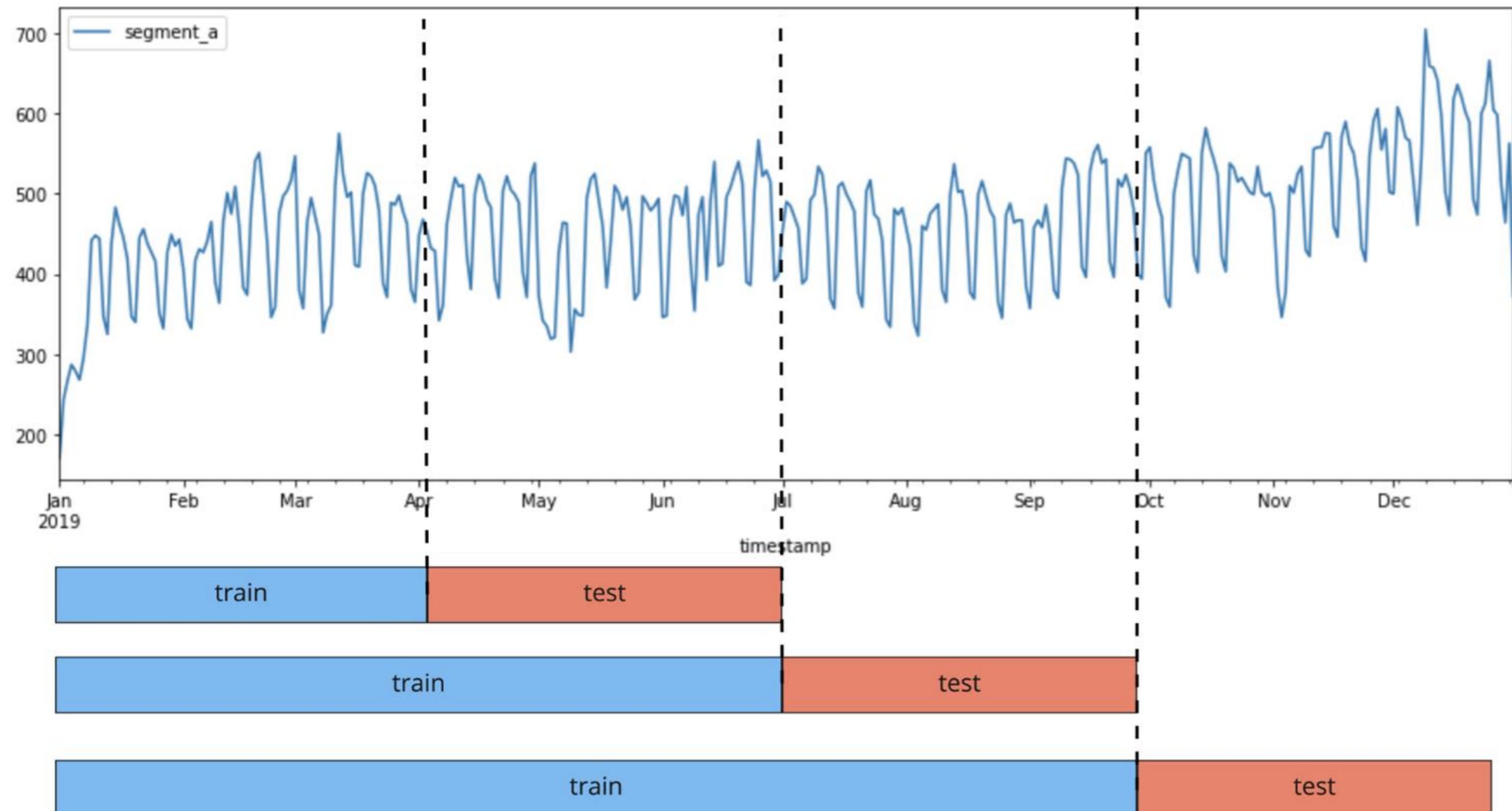
- Чаще всего метрики агрегируют по временной оси, а затем — по рядам в датасете.
- Нет единой «правильной» метрики: всё зависит от задачи.
- В зависимости от выбора метрики может меняться оптимальная модель.

2. Стратегии валидации

Чем плохо случайное разбиение

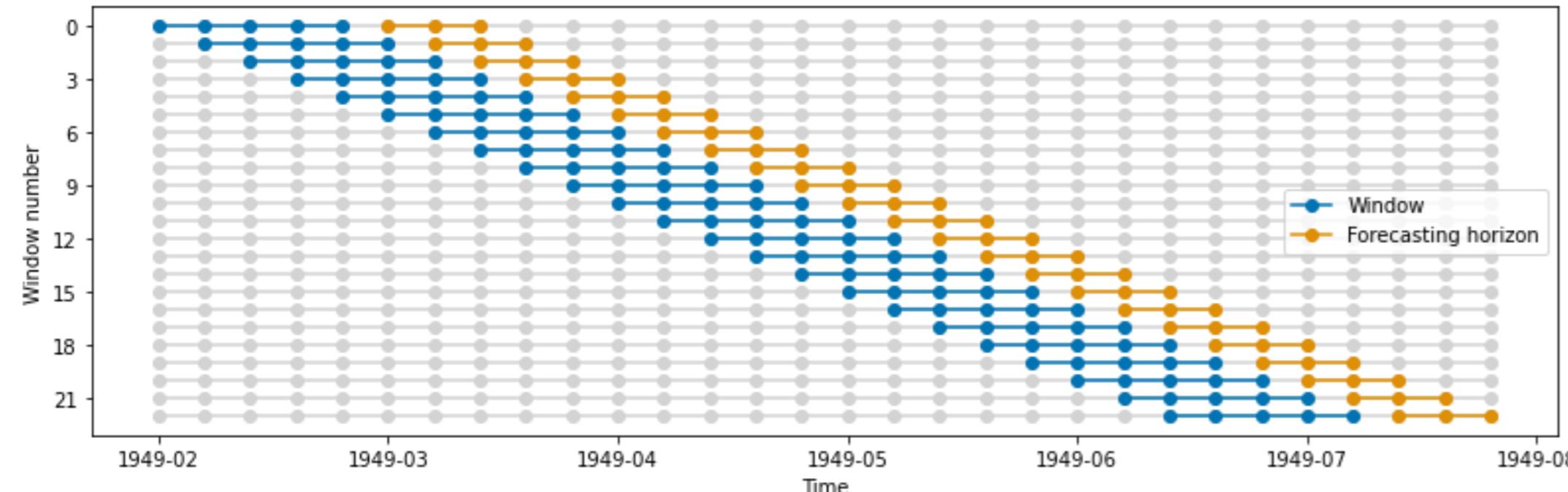


Общий принцип



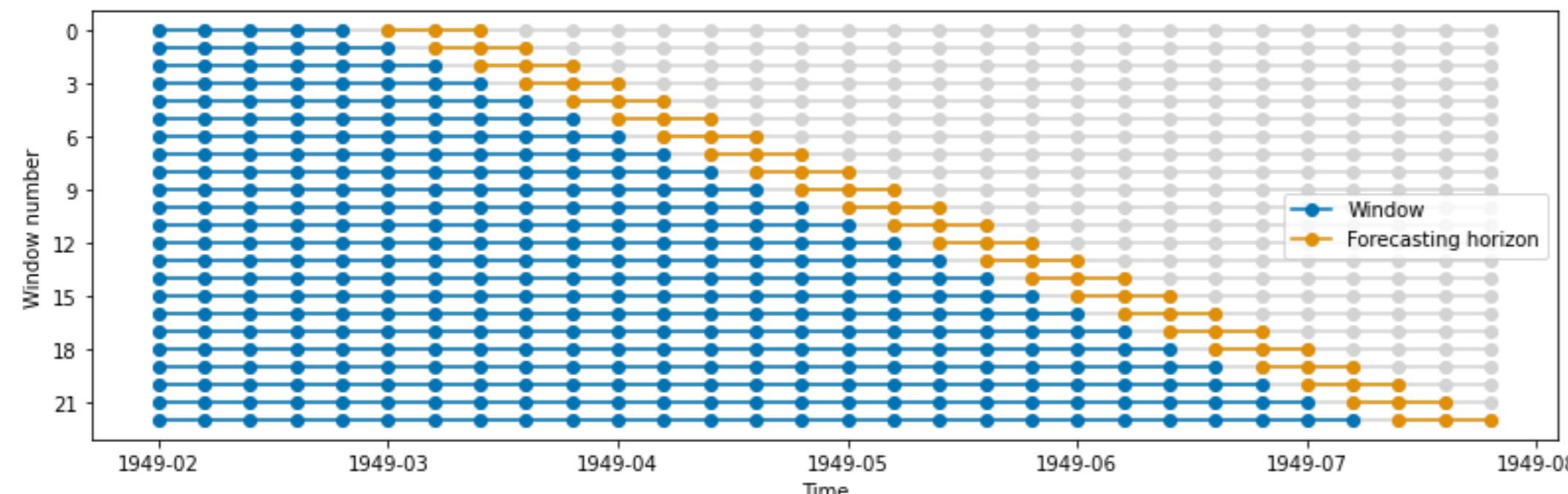
**Не подглядываем
в будущее:**
разделяем на фолды
по временной шкале.

Разбиение по времени



Constant window

- Распределение таргета меняется.
- Доступны только свежие данные.



Expanding window

- Распределение таргета не меняется.
- Чем больше данных, тем лучше.

Разбиение по времени и сегментам

| | | | | | | | | | | |
|----|----|----|----|----|----|----|----|----|----|----|
| s1 | | | | | | | | | | |
| s2 | | | | | | | | | | |
| s3 | | | | | | | | | | |
| s4 | | | | | | | | | | |
| | t0 | t1 | t2 | t3 | t4 | t5 | t6 | t7 | t8 | t9 |

Temporal + Segment

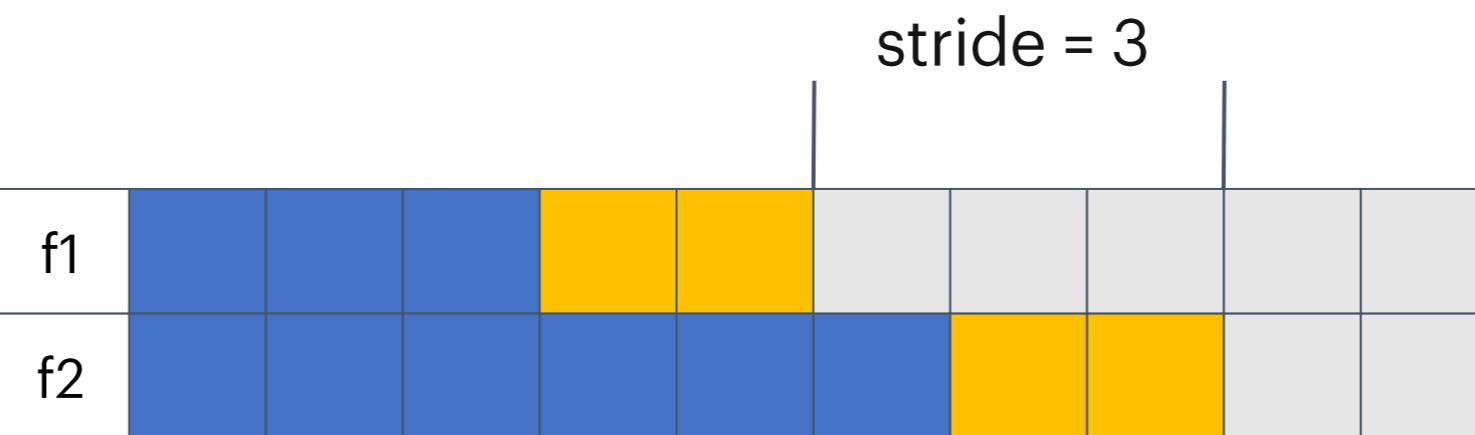


- Оцениваем обобщающую способность **глобальных моделей***.
- Хотим обучать модель на подвыборке рядов.

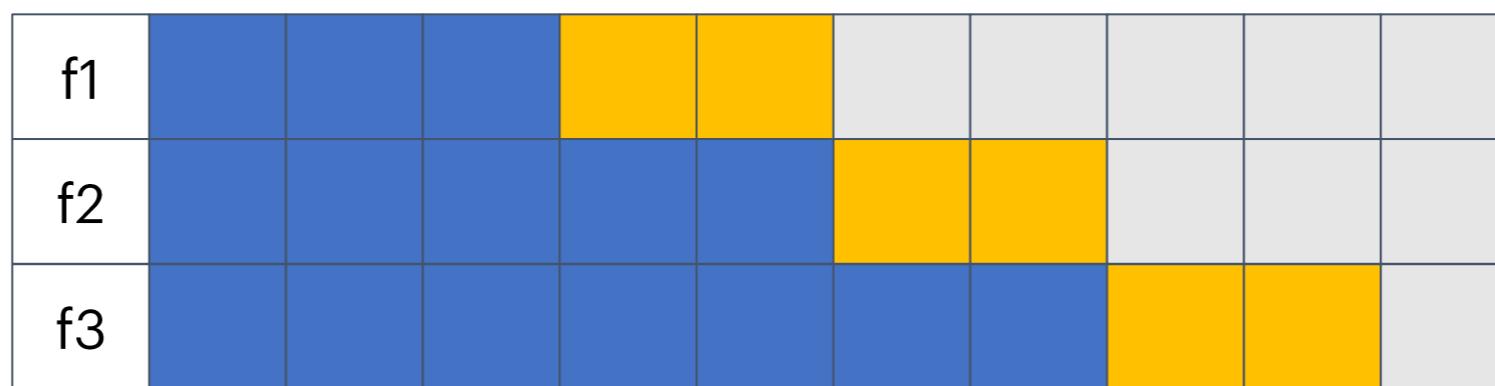
*Глобальная модель — одна модель, предсказывающая все сегменты.

Другие вариации

Шаг между фолдами
(stride)



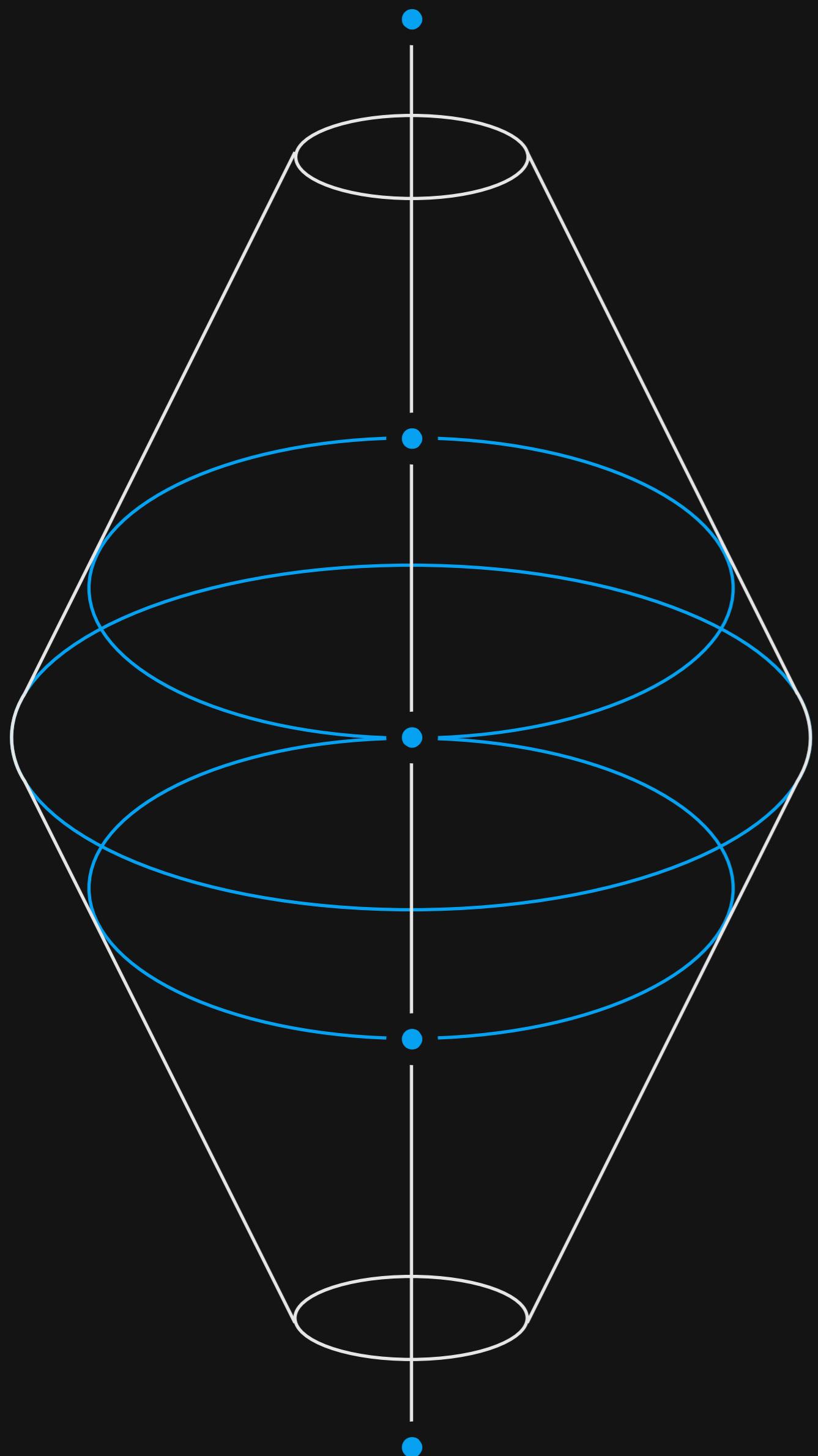
Стратегия переобучения
(refit)



fit(on) → model_1
fit(off) → model_1
fit(on) → model_2

Маскирование теста



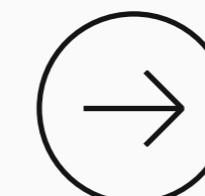


Примеры



Продажа мерча в нескольких магазинах

- Метрика: SMAPE
- Разбиение: «Time Series K-Fold»



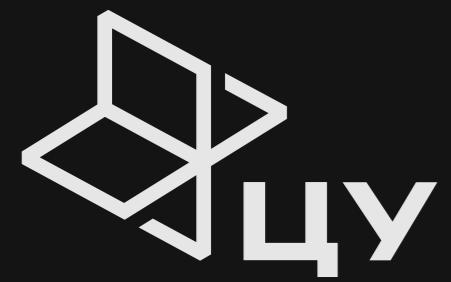
Продажа цветов на праздники

- Метрика: MAE
- Разбиение: «Маскирование теста»

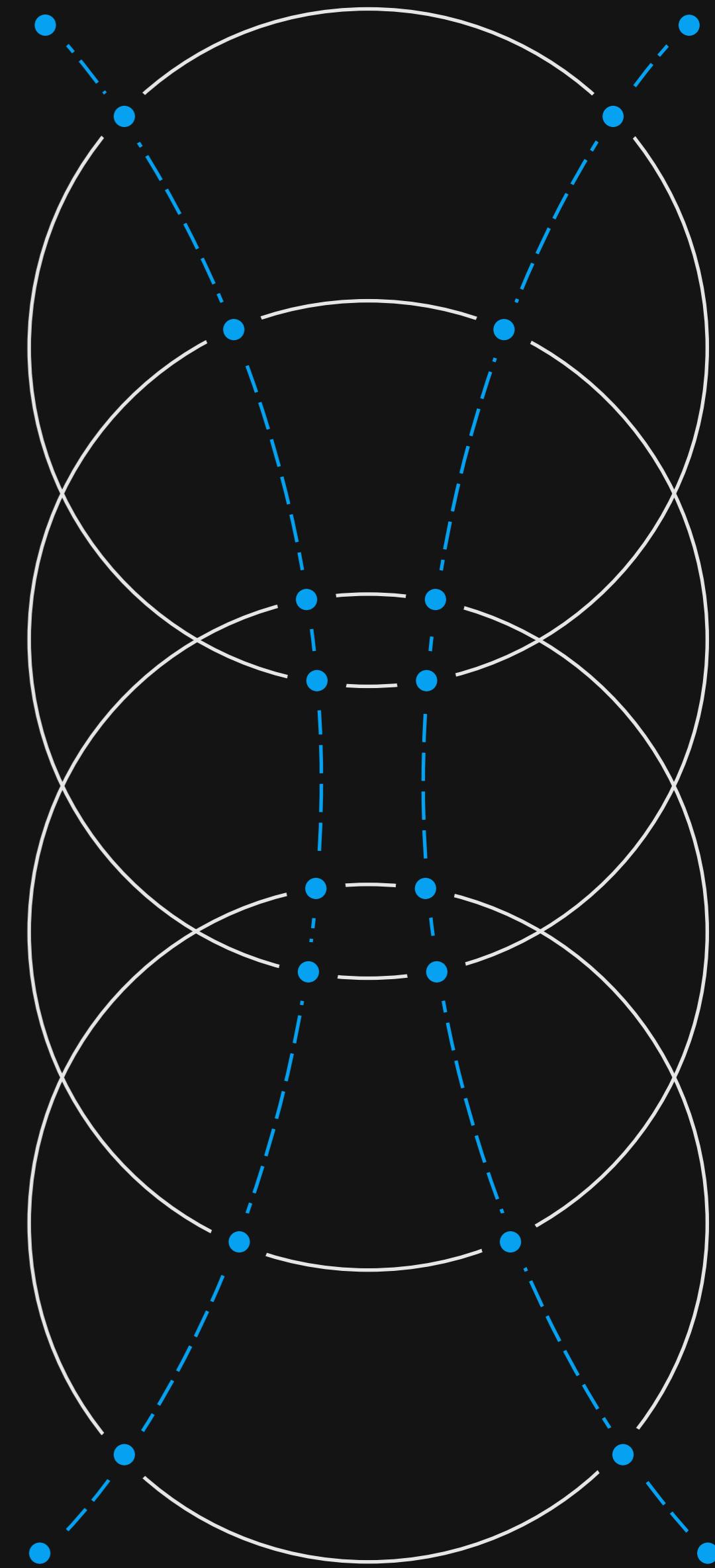


Прогноз спроса на маркетплейсе

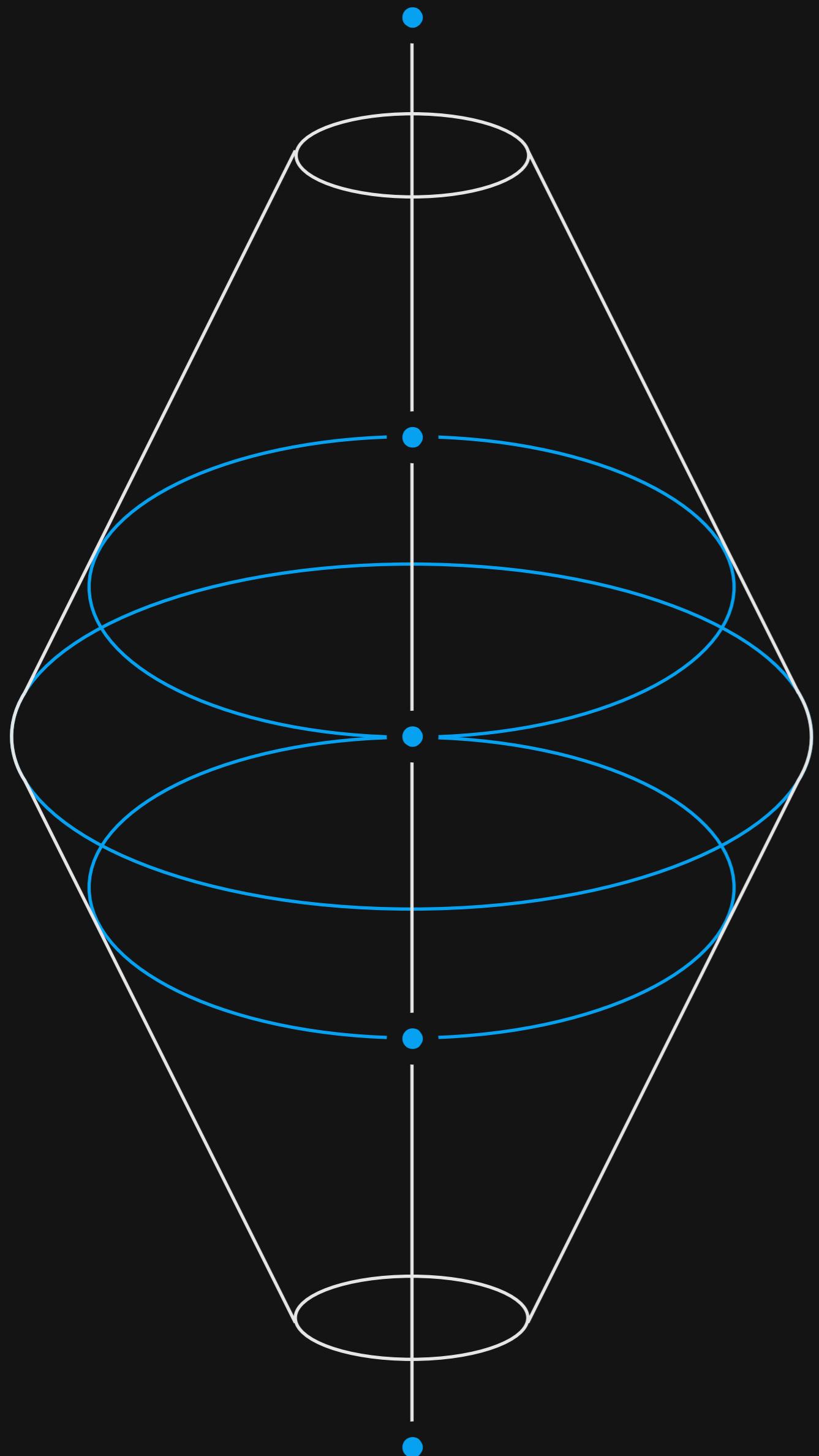
- Метрика: SMAPE
- Разбиение: «Temporal + Segment Split»



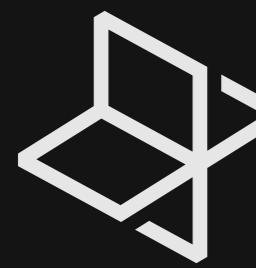
Заключение



Про что поговорили



- Примеры задач на временных рядах в индустрии
- Особенности данных, которые хотелось бы учитывать в моделях
- Классические модели (SMA/ETS/Prophet/SARIMAX)
- Метрики и валидация на временных рядах



ЦЕНТРАЛЬНЫЙ
УНИВЕРСИТЕТ

Вопросы

