

# Анализ временных рядов



образование

# Определение и основные задачи

Под **временным рядом** понимаются последовательно измеренные через некоторые (зачастую равные) промежутки времени данные.

## **Анализ временных рядов**

Объединяет методы изучения временных рядов, как пытающиеся понять природу точек данных (откуда они взялись? что их породило?), так и пытающиеся построить прогноз.

## **Прогнозирование временных рядов**

Заключается в построении модели для предсказания будущих событий основываясь на известных событиях прошлого, предсказания будущих данных до того как они будут измерены. Типичный пример — предсказание цены открытия биржи основываясь на предыдущей её деятельности.

# Исследование временных рядов

Существует несколько методов анализа данных, применимых для временных рядов.

## **Общее исследование**

- Визуальное изучение графических представлений временных рядов
- Автокорреляционный анализ для изучения зависимостей
- Спектральный анализ для изучения циклического поведения, не связанного с сезонностью

## **Описание**

- Разделение компонент: тренд, сезонность, медленно и быстро меняющиеся компоненты, циклическая нерегулярность
- Простейшие свойства частных распределений
- Прогнозирование и предсказание
  - Полноценные статистические модели при стохастическом моделировании для создания альтернативных версий временных рядов, показывающих, что могло бы случиться на произвольных отрезках времени в будущем при изменении определенных условий (предсказание)
  - Упрощённые статистические модели для описания вероятных значений временного ряда в ближайшем будущем при известных последних значениях при тех же условиях (прогноз)

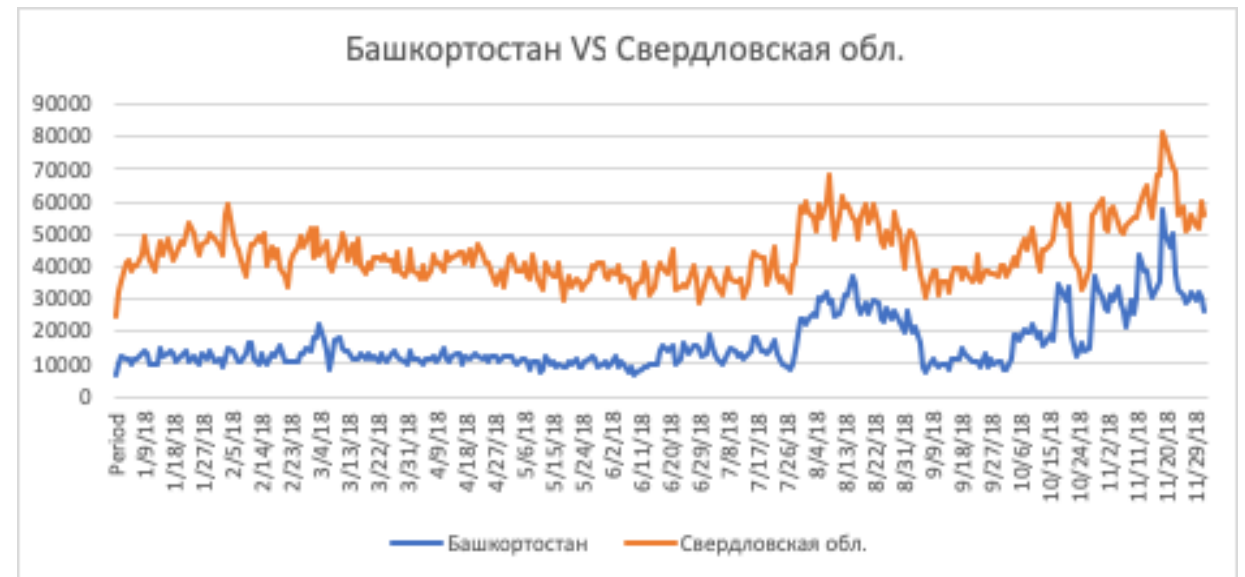
# Causal Impact (R)

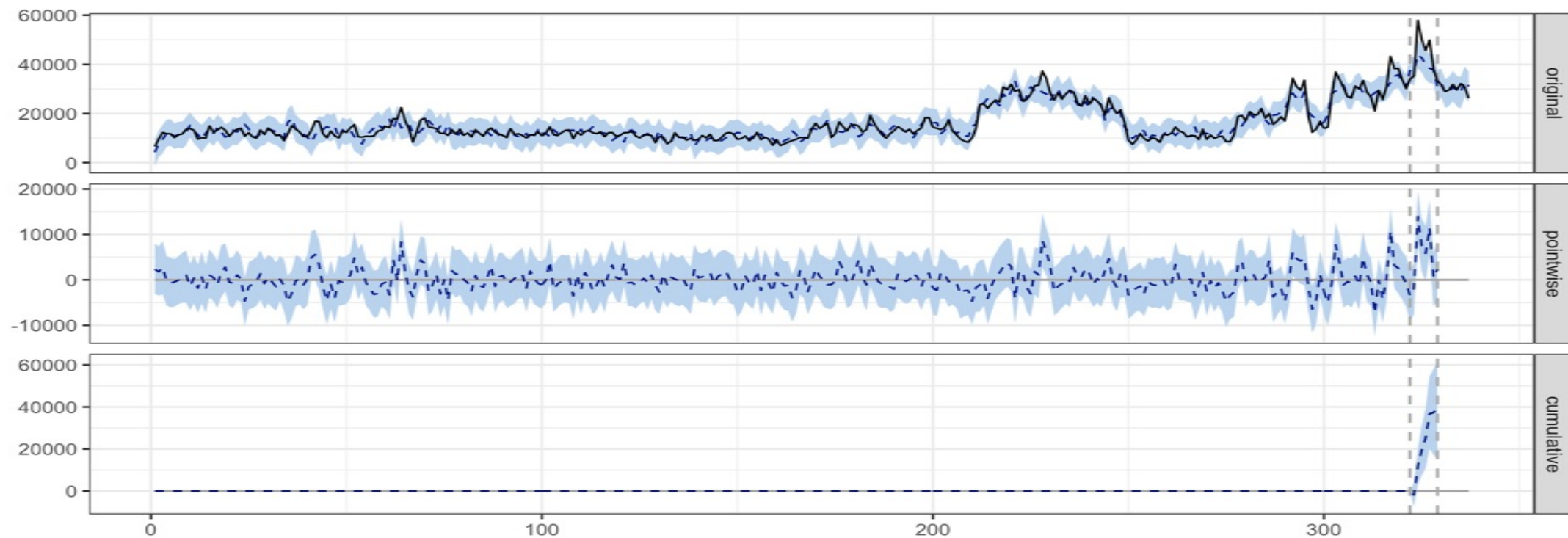
## Библиотека на R

Выбираем 2 схожих ряда (эвристика – например, соседние регионы со схожими трендами, расчёт коэффициента корреляции или PCA-анализ);

Остальное делает Causal Impact.

## Прямые заходы: Башкортостан (событие) VS Свердловская область





Среднее фактическое значение: 44 343;

Среднее модельное: 38 689;

Диапазон модельных значений в рамках 95% доверительного интервала: [35346, 41948]

С 95% точностью можно сказать, что эффект от события в прямых составляет от 6.2% до 23%

# STL-Разложение

## Библиотеки на Python

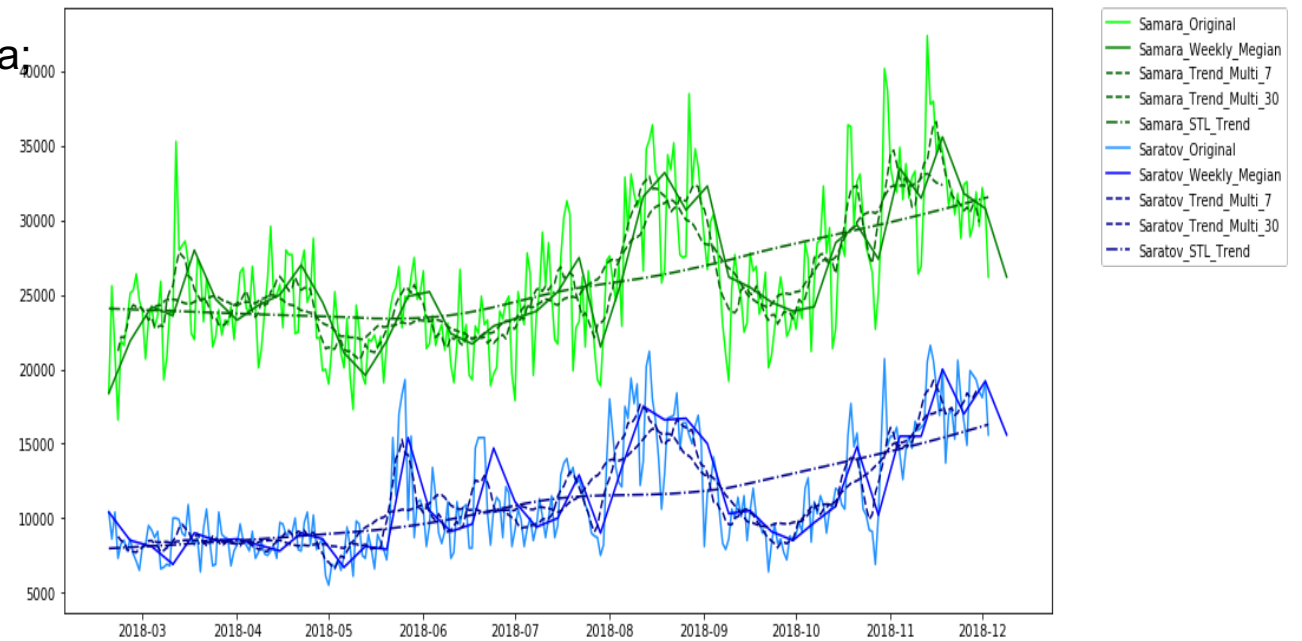
`statsmodels.tsa.seasonal` – выделение тренда (только MA), сезонной составляющей и шума;

`stldecompose` - полиномиальный тренд "из коробки";

`sklearn.preprocessing` – полиномиальная регрессионная модель выделения тренда;

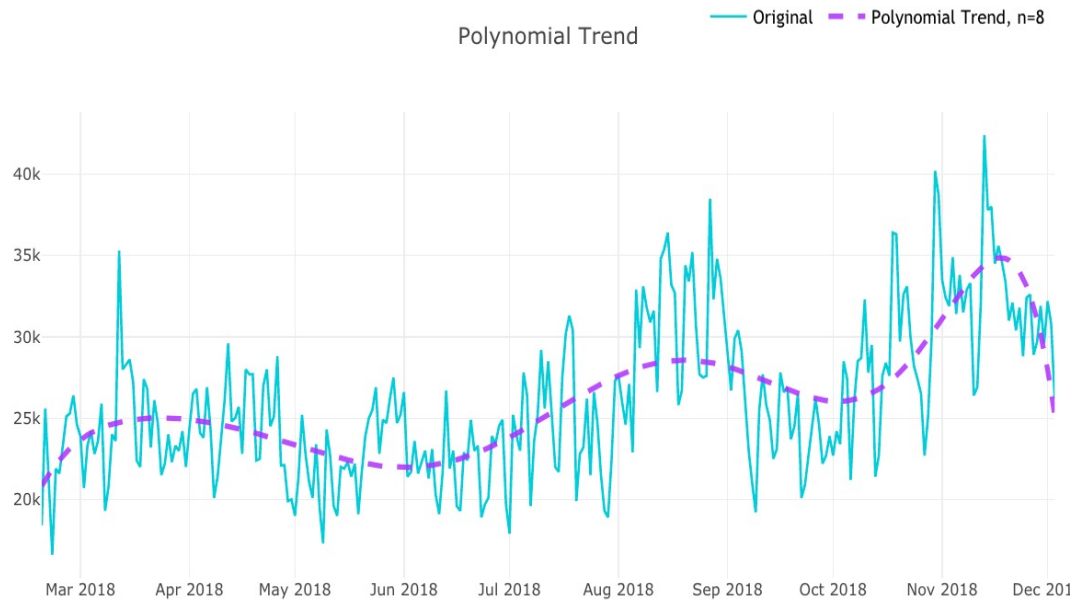
MA, Median Smoothing (for some range)

## Графики трендов (различные подходы):

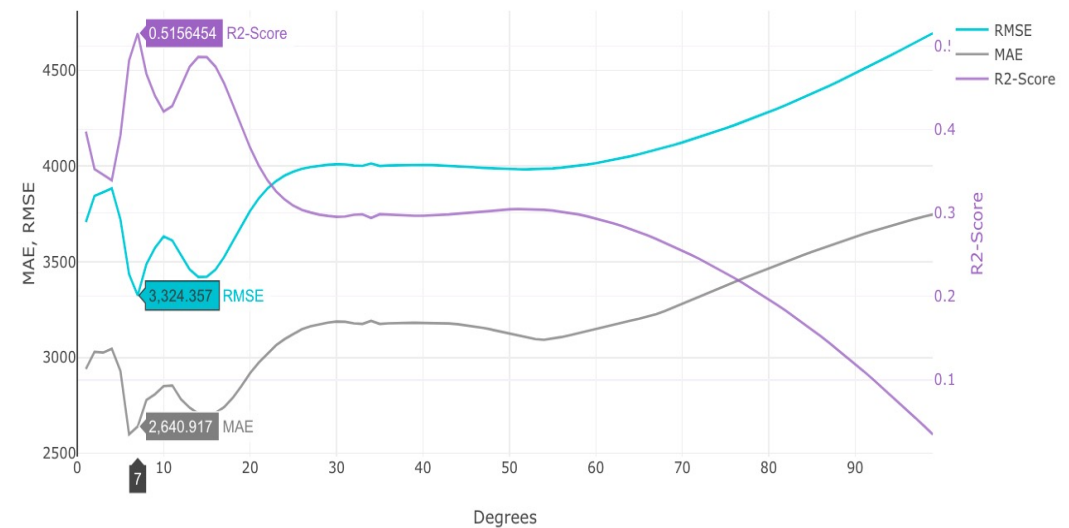


# Подбор степени полинома по метрикам R2, MAE, RMSE:

Polynomial Trend



Polynomial Model Metrics



# ARIMA (p, d, q):

Авторегрессионная модель:

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

Стационарный ряд: математическое ожидание и дисперсия постоянны (т. е. среднее значение временного ряда, вокруг которого изменяются уровни, является величиной постоянной)

Проверка: критерий Дикки-Фуллера (DF-тест)

ADF: -0.037

p-value: 0.95

Critical values: {

'5%': -2.87;

'1%': -3.45;

'10%': -2.57}

5%: ADF > Critical value => Единичных корней нет, ряд стационарен



# ARIMA (p, d, q):

