Блок

# Feature Engeneering

Занятие № 2

# Использование pandas и numpy для очистки данных

# Цели занятия

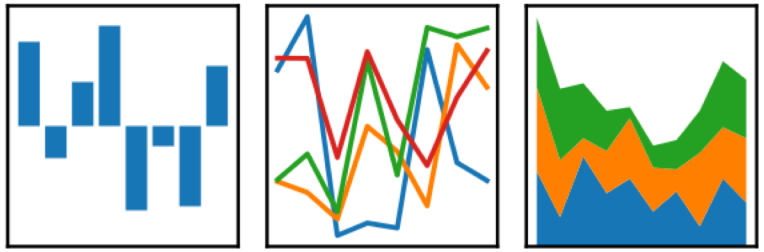
- Получение навыков выявления основных проблем в данных
- Научиться пользоваться средствами pandas и numpy для решения проблем в данных

# Немного о pandas и numpy

# Pandas

pandas

$$y_{it} = \beta' x_{it} + \mu_i + \epsilon_{it}$$



## Основные возможности библиотеки:

1. Объект DataFrame для манипулирования индексированными массивами двумерных данных
2. Инструменты для обмена данными между структурами в памяти и файлами различных форматов
3. Встроенные средства совмещения данных и способы обработки отсутствующей информации
4. Переформатирование наборов данных, в том числе создание сводных таблиц
5. Срез данных по значениям индекса, расширенные возможности индексирования, выборка из больших наборов данных
6. Вставка и удаление столбцов данных
7. и т.д.

https://ru.wikipedia.org/wiki/Pandas

# Numpy

Является базовой библиотекой для научных вычислений с помощью Python. Он имеет следующие возможности:

- Мощный N-размерный массив
- поддержка сложный функций
- инструментарий для интеграции с C/C++ и кодом Fortran
- Линейная алгебра, преобразование Фурье и работа со случайными числами

https://numpy.org/

# ПРАКТИЧЕСКАЯ ЧАСТЬ

# ВОПРОСЫ