

Практическая работа №2 – Обработка данных в pandas (2)

Выполнить задание №1 (текст ниже)

1. Выберите датасет на сайте <https://www.kaggle.com/> при выборе датасета убедитесь в следующем:

- набор данных представлен в формате CSV;
- в наборе данных имеются переменных разных типов: численные, категориальные, временные;
- в наборе присутствуют пропущенные значения.

2. Проведите анализ данных используя Jupyter Notebook применяя методы, используемые на первом занятии. При анализе помимо строк кода, используйте Markdown разметку для описания процесса.

3. Итоговый файл отправьте на электронную почту artem@golubnichij.ru с темой письма **Big Data 10_23**

Выполнить задание №2 (текст ниже)

4. Построить визуализацию данных (гистограммы – численные данные, столбчатые диаграммы – категориальным данным, временные данные в виде линии, диаграмма рассеивания – численные данные двух переменных).

5. Обработка пропущенных значений (создание и обработка).

6. Оформить подробно Markdown (сделать анализ и написать основные выводы).

Дополнительное задание (объединение датафреймов) (не обязательно)

Отсмотреть набор данных (из репозитория на гитхаб) и провести объединение датафреймов (один к одному, один ко многим).