МГТУ имени Баумана

Факультет «Информатика и Системы управления»

Кафедра «Системы обработки информации и управления»

Дисциплина «Теория машинного обучения»

Отчет по лабораторной работе №1

## «Разведочный анализ данных. Исследование и визуализация данных»

Выполнил: Преподаватель:

Студент группы ИУ5-61Б Гапанюк Ю.Е.

Гапчук Л.Д.

Москва, 2020г.

**Цель лабораторной работы:** изучение различных методов визуализация данных.

Задание:

* Выбрать набор данных (датасет). Вы можете найти список свободно распространяемых датасетов [здесь.](https://github.com/ugapanyuk/ml_course_2021/wiki/DSLIST)
* Для первой лабораторной работы рекомендуется использовать датасет без пропусков в данных, например из [Scikit-learn.](https://scikit-learn.org/stable/datasets/toy_dataset.html)
* Пример преобразования датасетов Scikit-learn в Pandas Dataframe можно посмотреть [здесь.](https://github.com/ugapanyuk/ml_course_2021/blob/main/common/notebooks/ds/sklearn_datasets.ipynb)

Для лабораторных работ не рекомендуется выбирать датасеты большого размера.

* Создать ноутбук, который содержит следующие разделы:

1. Текстовое описание выбранного Вами набора данных.
2. Основные характеристики датасета.
3. Визуальное исследование датасета.
4. Информация о корреляции признаков.

* Сформировать отчет и разместить его в своем репозитории на github.





















