****Московский государственный технический университет им. Н.Э. Баумана  
Кафедра «Системы обработки информации и управления»

Лабораторная работа №1  
по дисциплине  
«Методы машинного обучения»  
на тему

## «Создание "истории о данных" (Data Storytelling)»

Выполнил: Ван Пэй  
Группа: ИУ5-22М

Москва — 2021 г.

**1. Цель лабораторной работы**

Изучение различных методов визуализация данных и создание истории на основе данных.

**2. Задание**

2.1 Выбрать набор данных (датасет).

2.2 Создать "историю о данных" в виде юпитер-ноутбука, с учетом следующих требований:

История должна содержать не менее 5 шагов (где 5 - рекомендуемое количество шагов). Каждый шаг содержит график и его текстовую интерпретацию.

На каждом шаге наряду с удачным итоговым графиком рекомендуется в юпитер-ноутбуке оставлять результаты предварительных "неудачных" графиков.

Не рекомендуется повторять виды графиков, желательно создать 5 графиков различных видов.

Выбор графиков должен быть обоснован использованием методологии data-to-viz. Рекомендуется учитывать типичные ошибки построения выбранного вида графика по методологии data-to-viz. Если методология Вами отвергается, то просьба обосновать Ваше решение по выбору графика.

История должна содержать итоговые выводы. В реальных "историях о данных" именно эти выводы представляют собой основную ценность для предприятия.

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

**3. Ход выполнения работы  
3.1. Текстовое описание набора данных**

Доступ к безопасной питьевой воде имеет важное значение для здоровья, является основным правом человека и компонентом эффективной политики по охране здоровья. Это важно как вопрос здоровья и развития на национальном, региональном и местном уровнях. В некоторых регионах было показано, что инвестиции в водоснабжение и санитарию могут принести чистую экономическую выгоду, поскольку сокращение неблагоприятных последствий для здоровья и затрат на здравоохранение перевешивают затраты на проведение мероприятий.

Цель работы: проверка безопасности питьевой воды по:

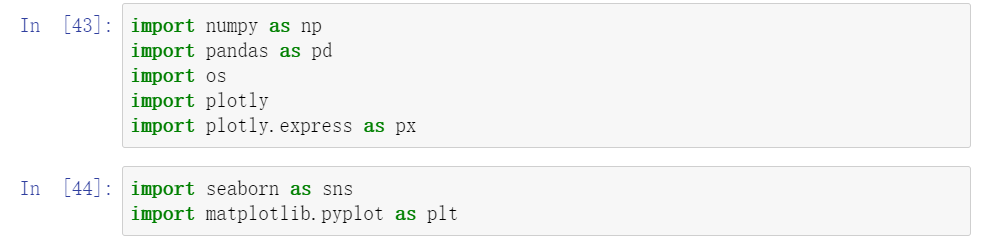
1) PH;

2) Жесткость воды；

3) Твердые вещества (общее количество растворенных твердых веществ - TDS)；

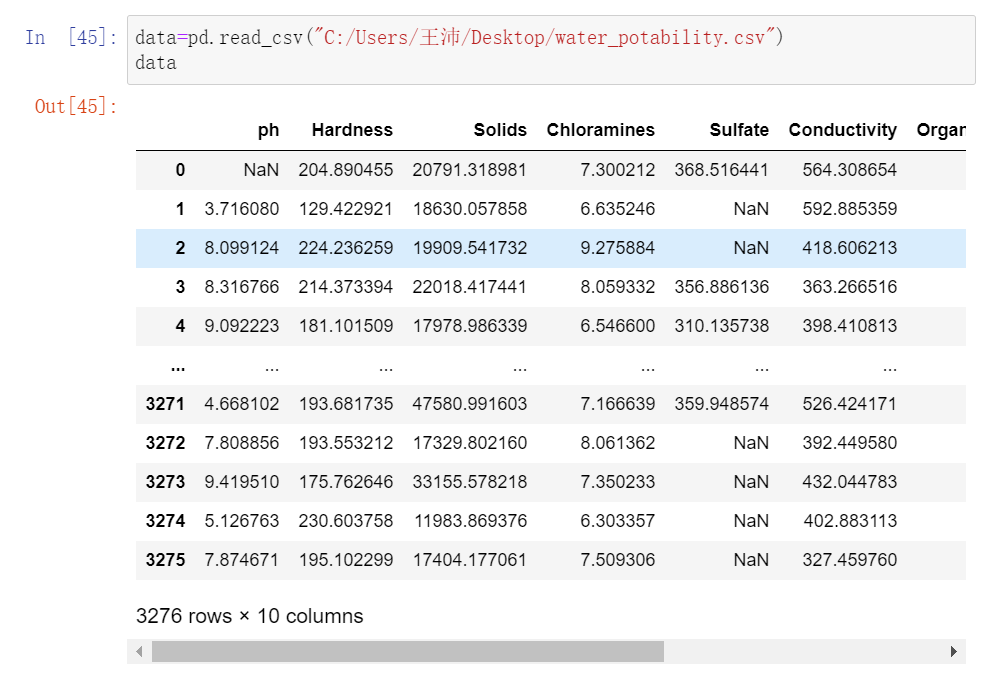
4) Мутность и т.д.

**3.2. Основные характеристики набора данных**  
Подключим все необходимые библиотеки:

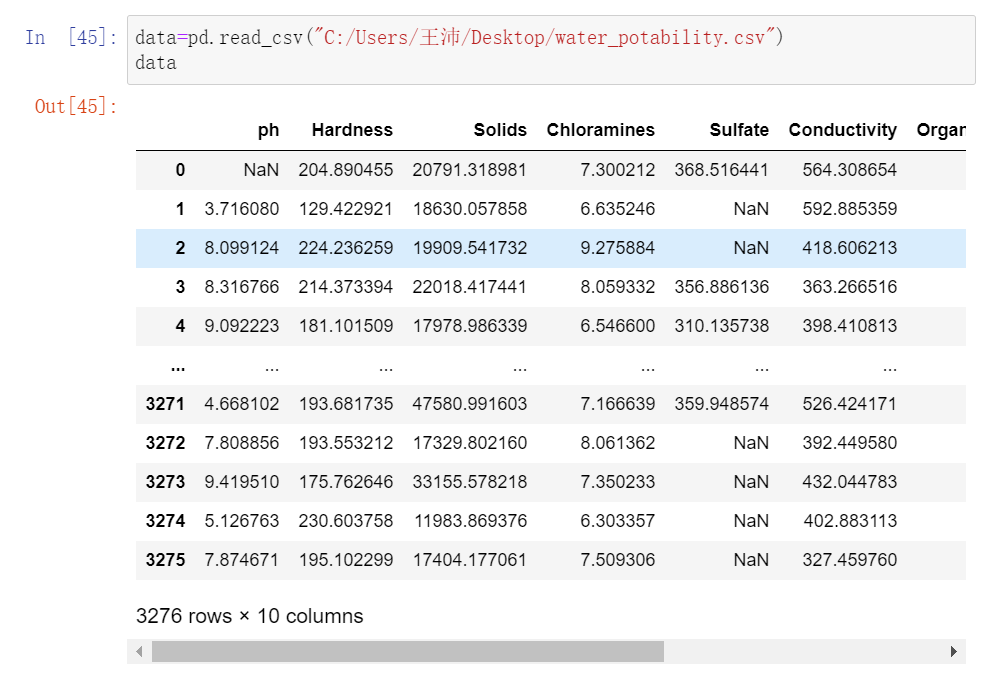


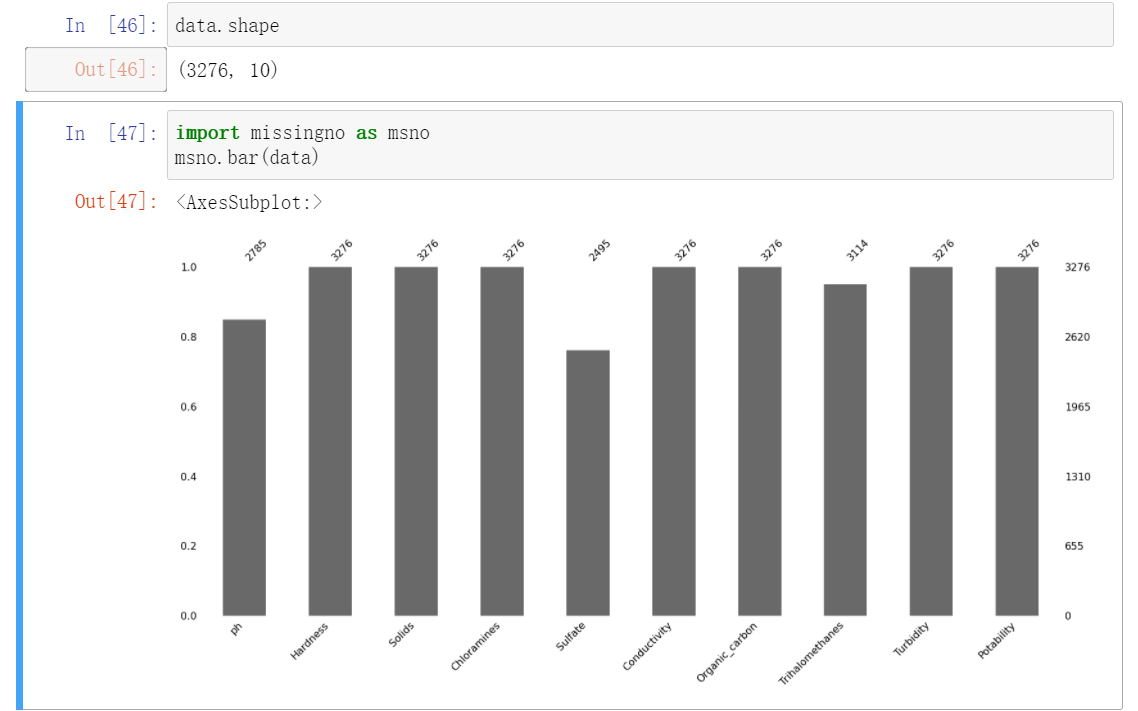


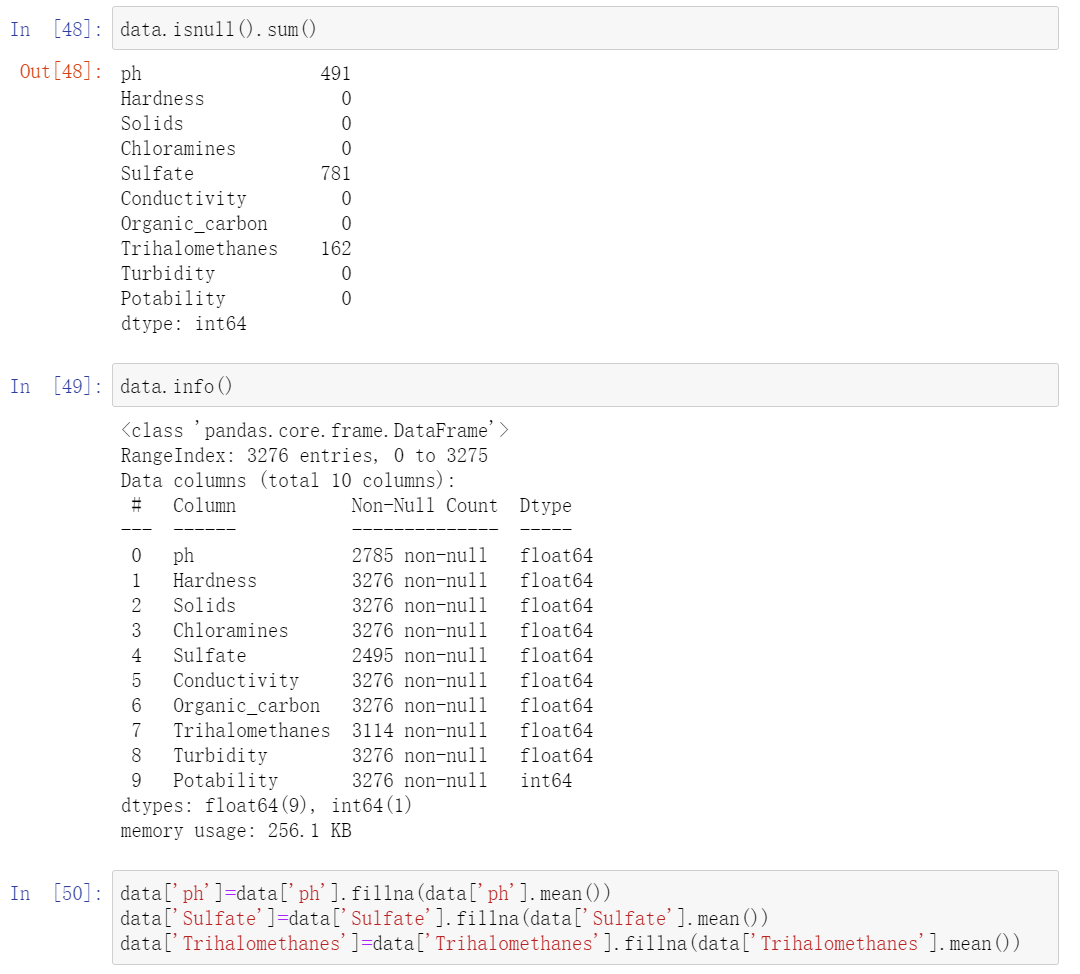
Загрузим непосредственно данные:

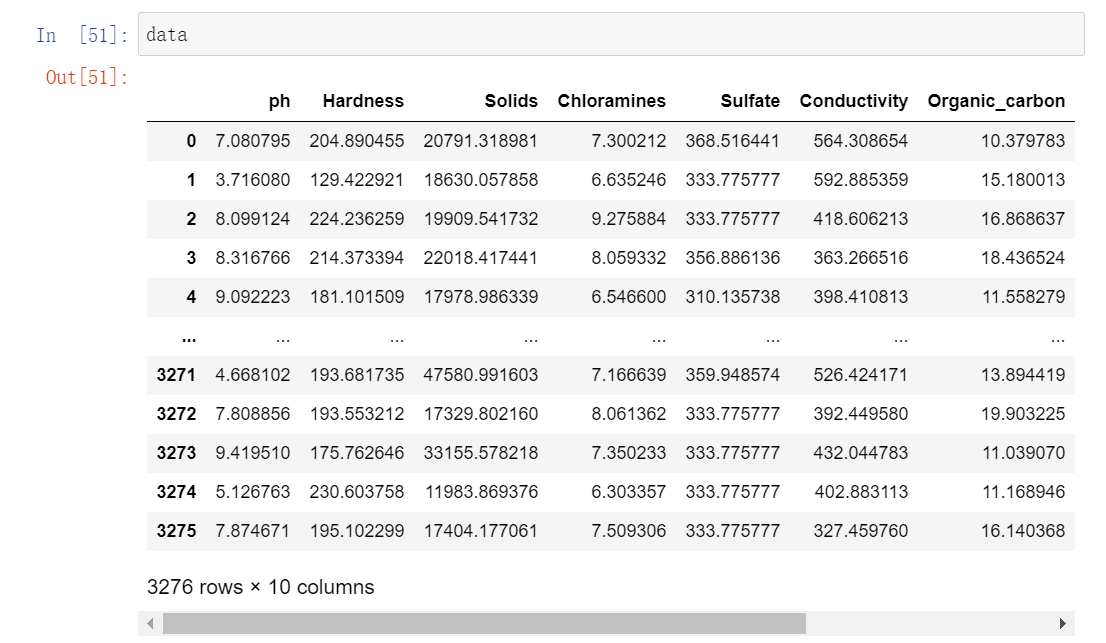


Показание информаций о температуре суши и океана на земле за последние несколько лет:

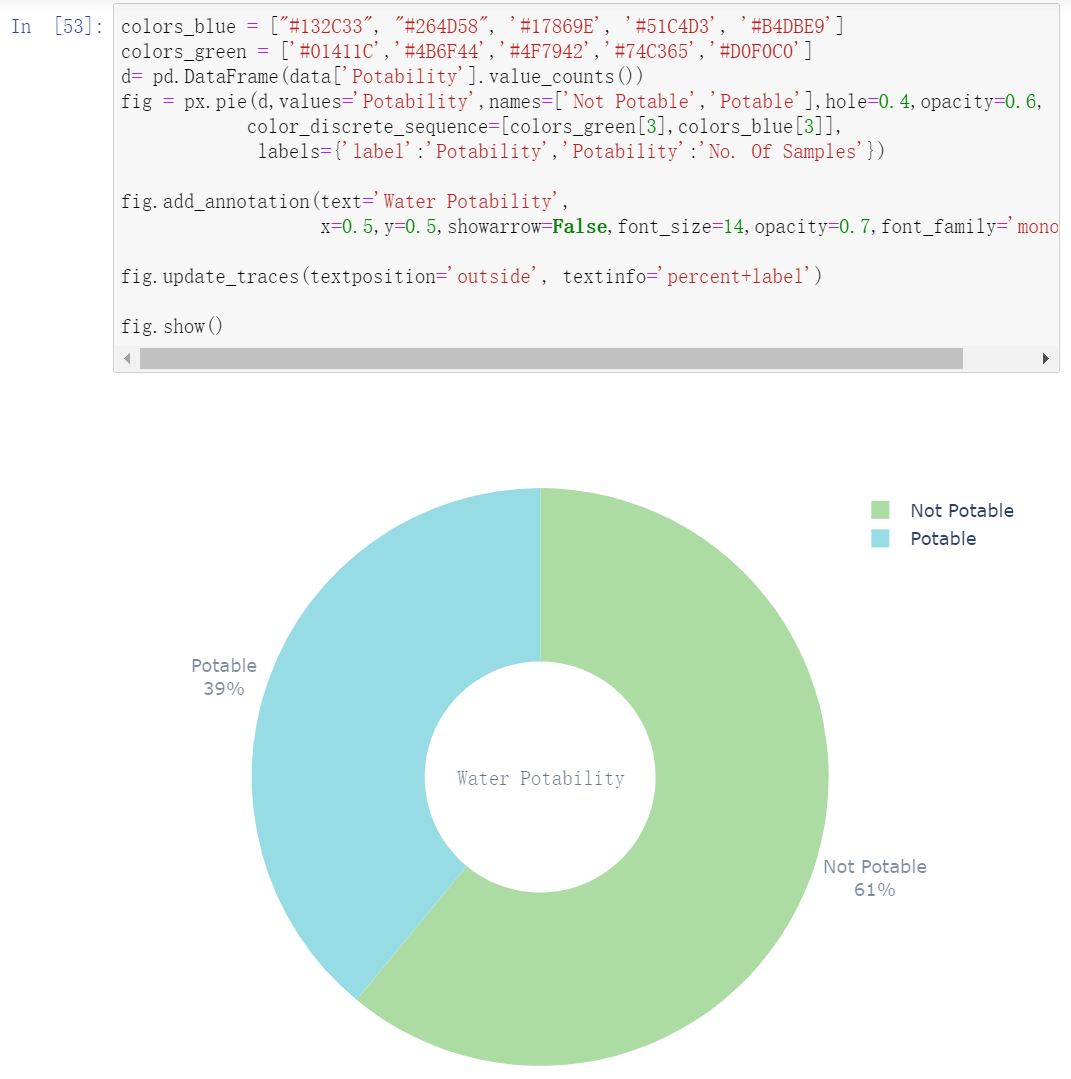


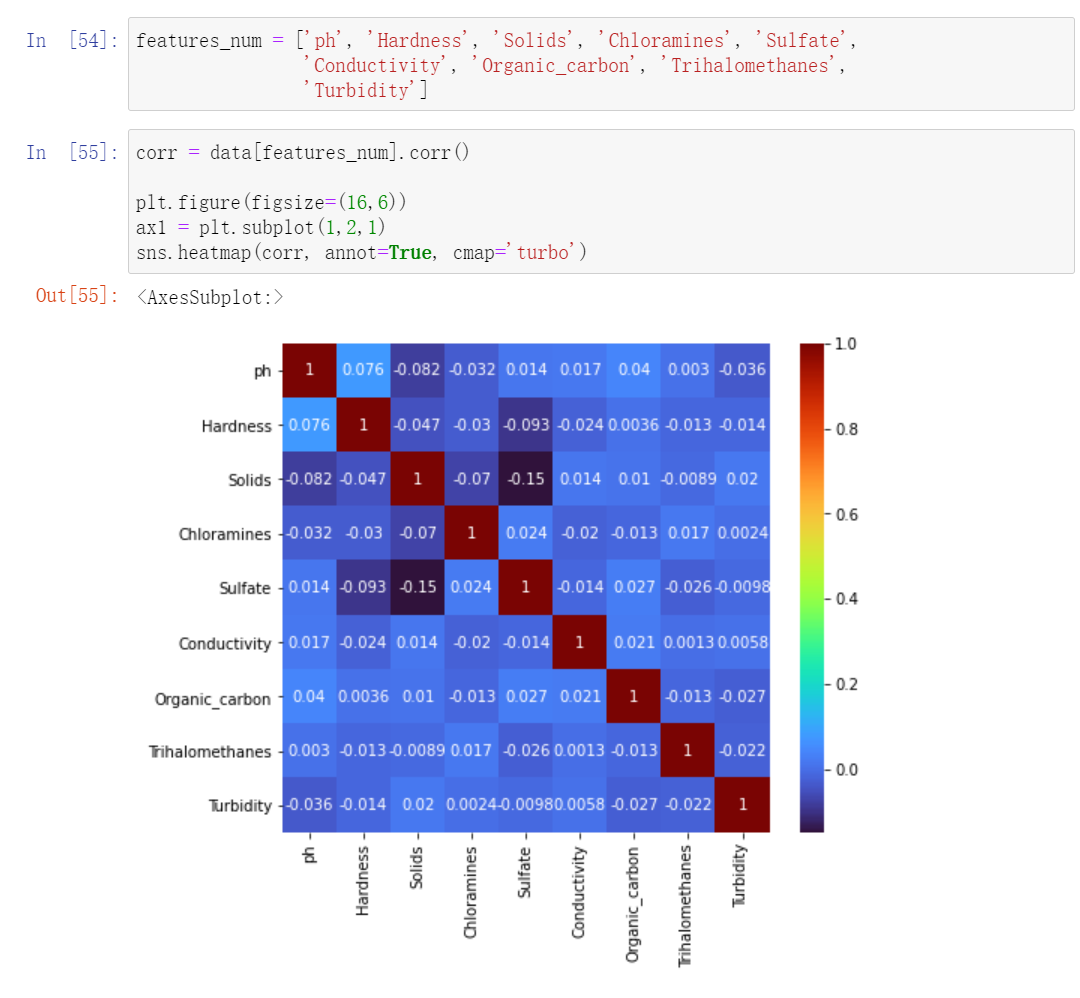


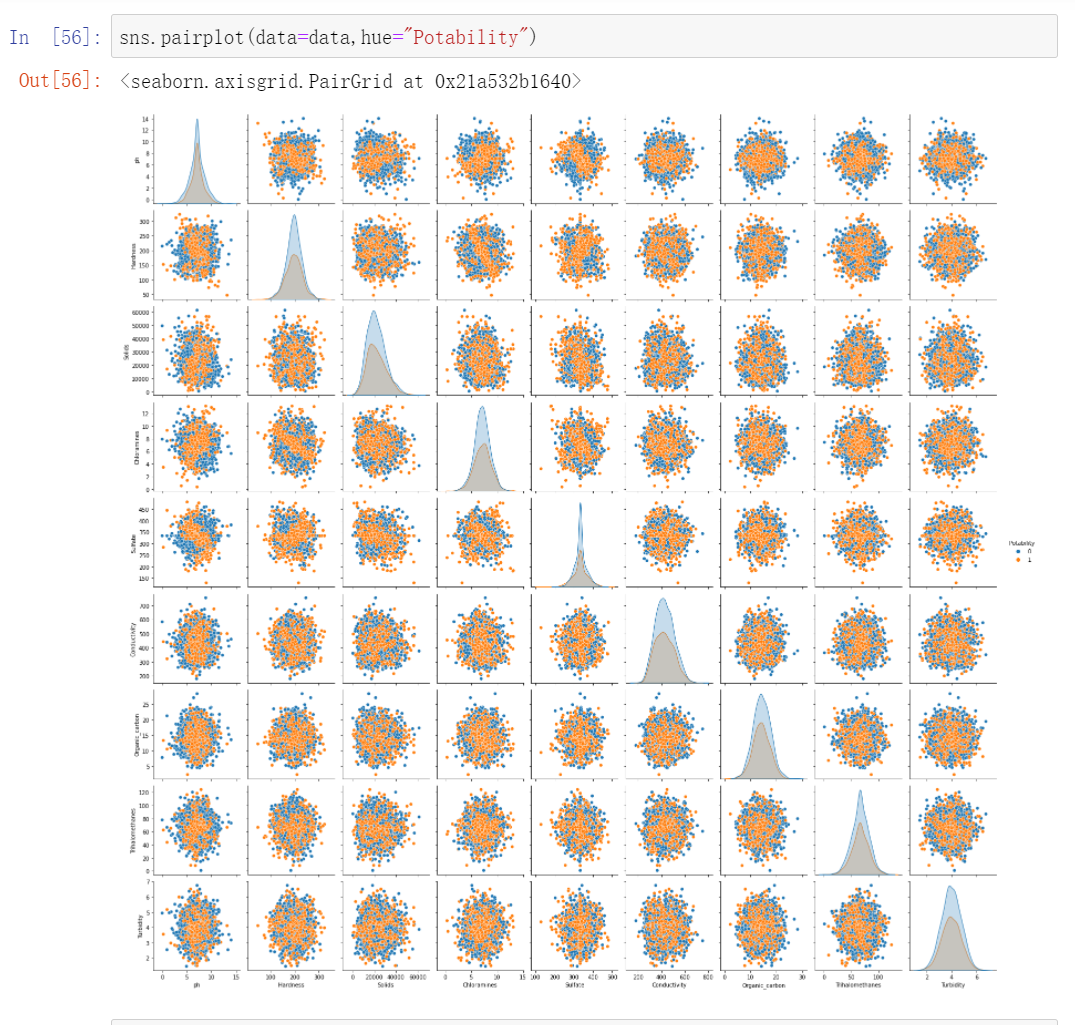


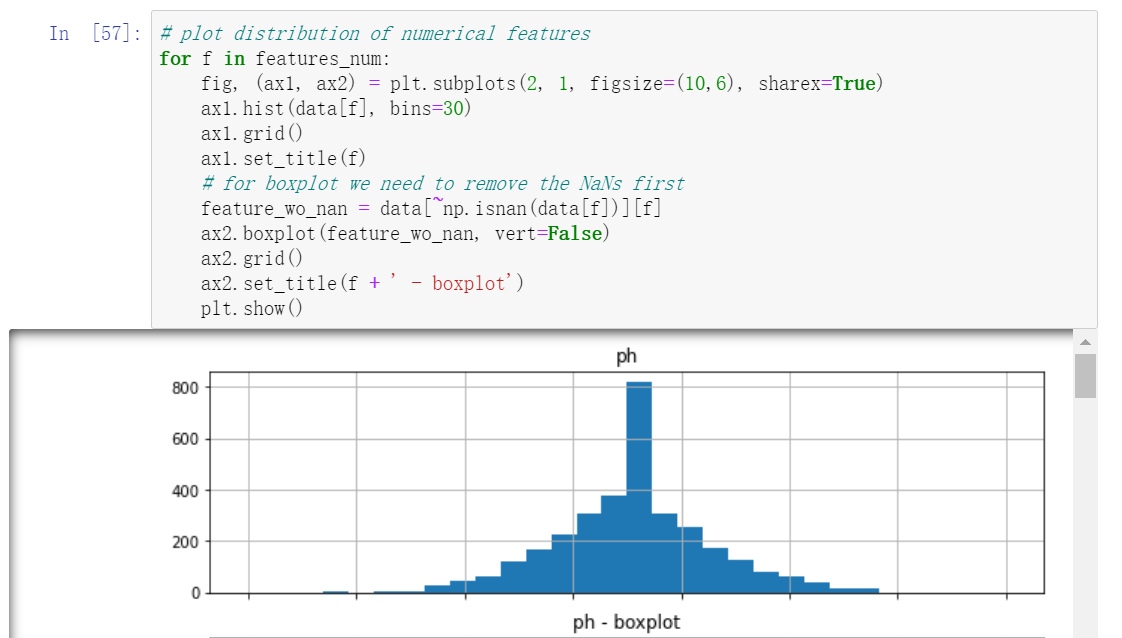


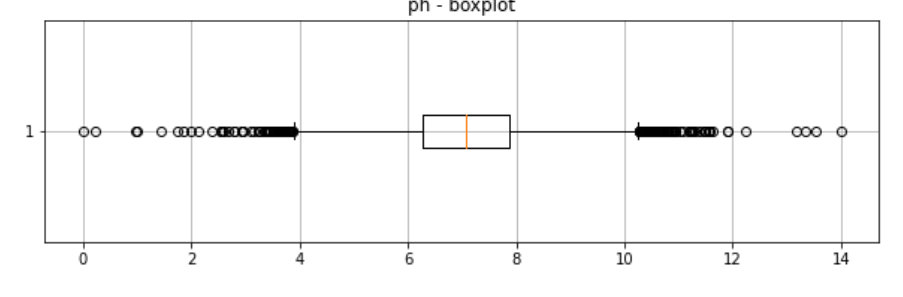
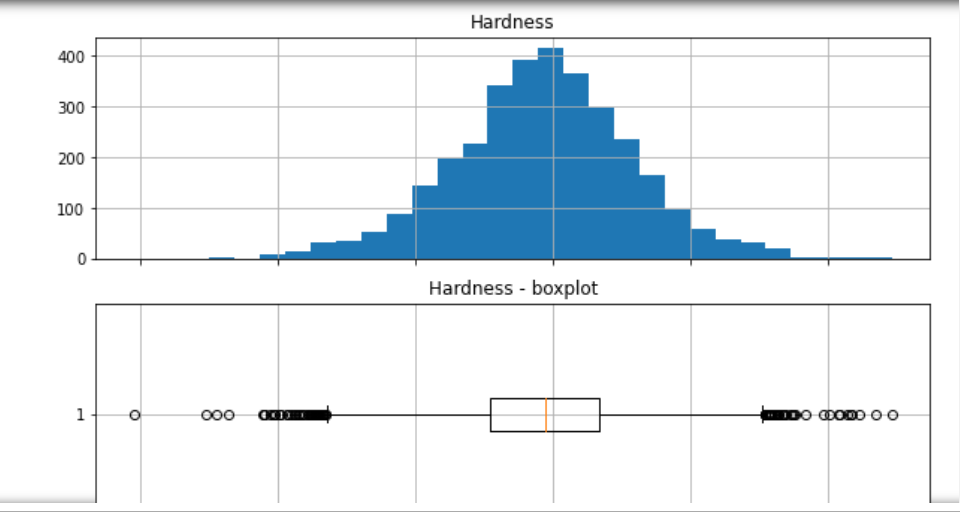


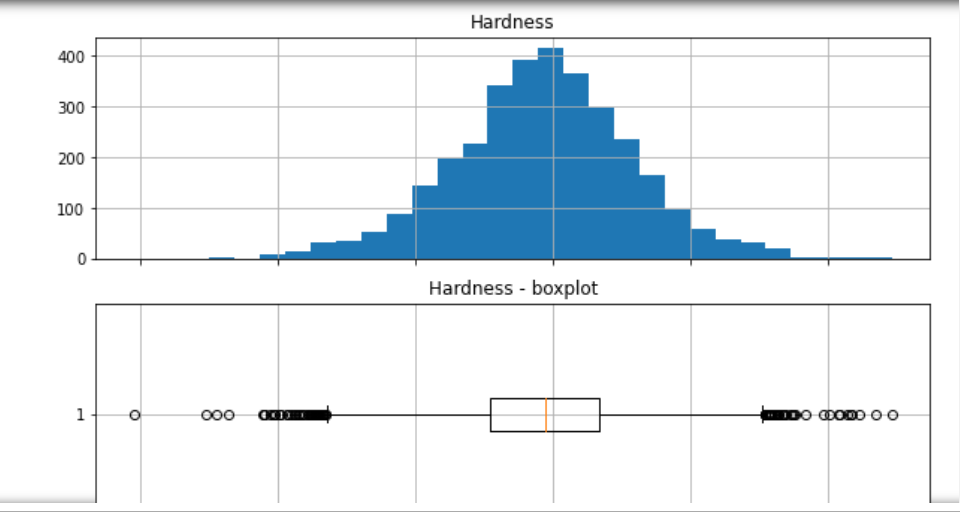


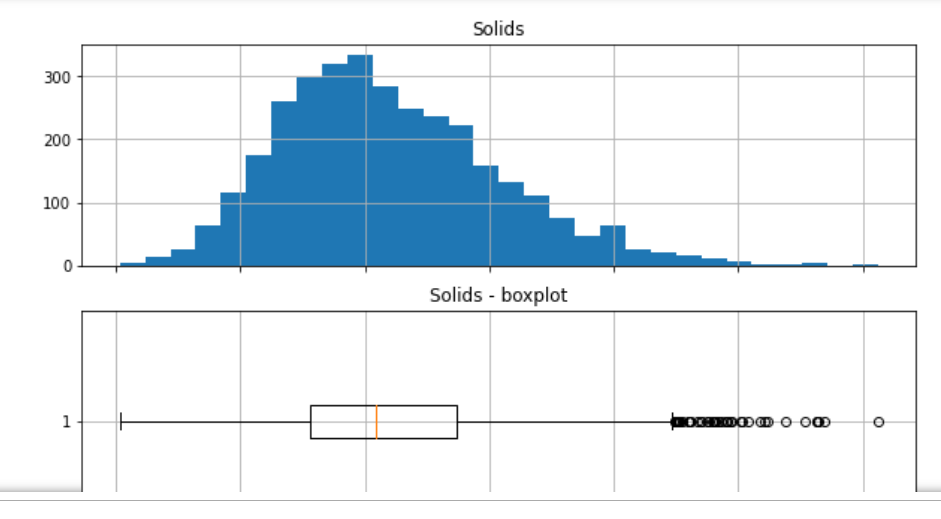


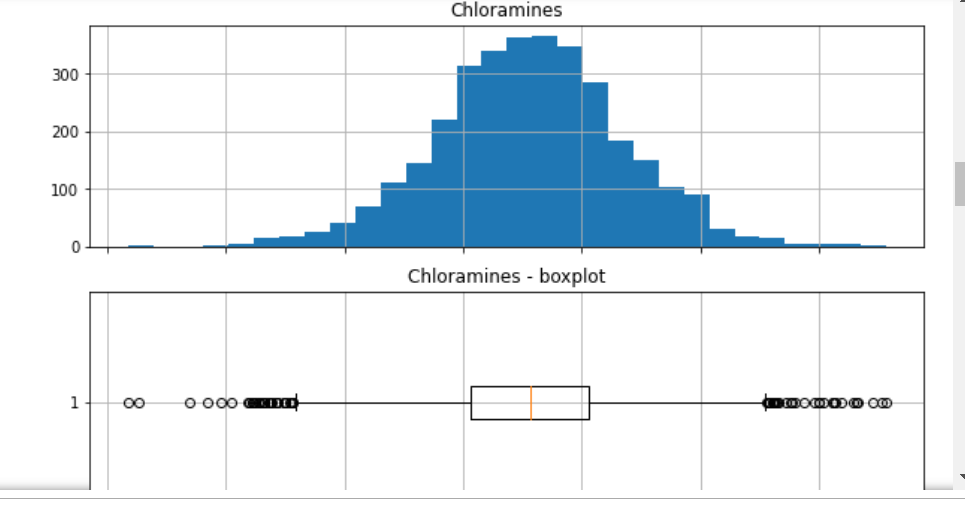


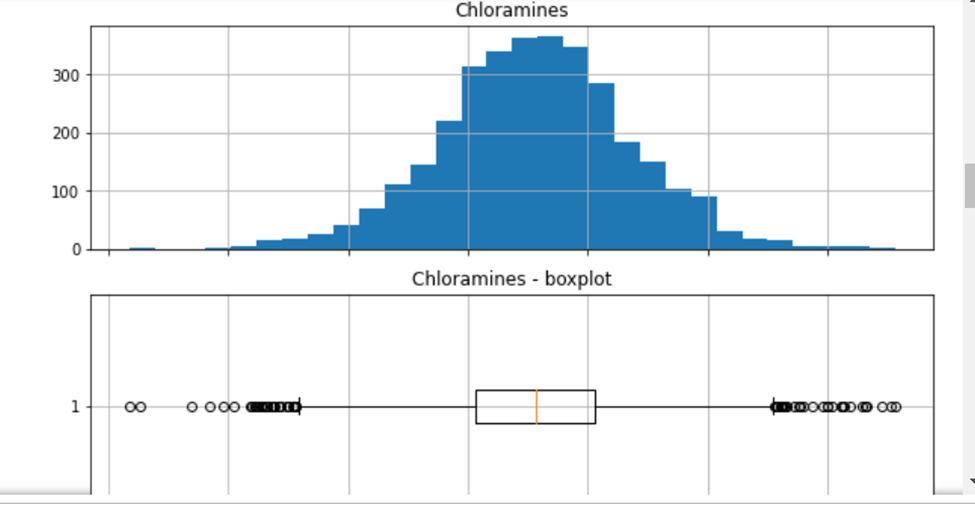


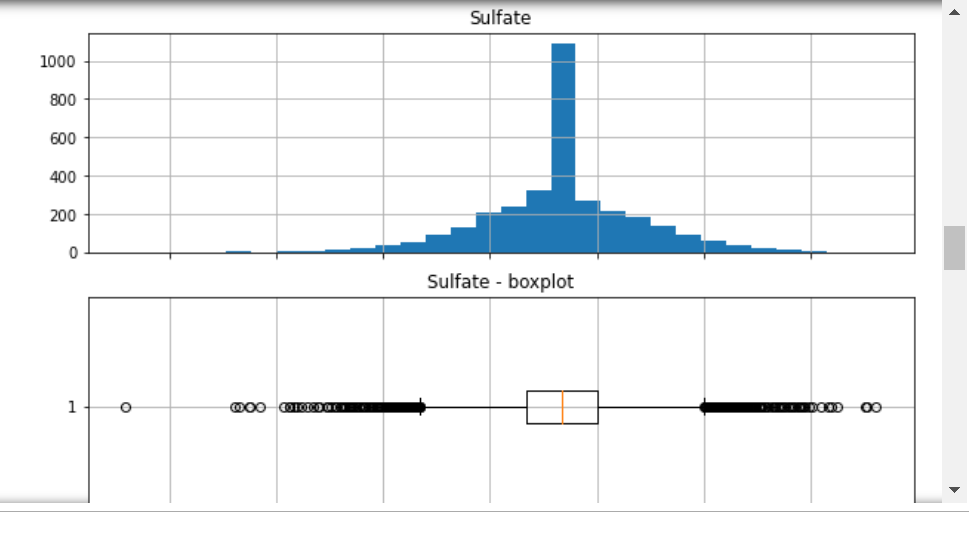


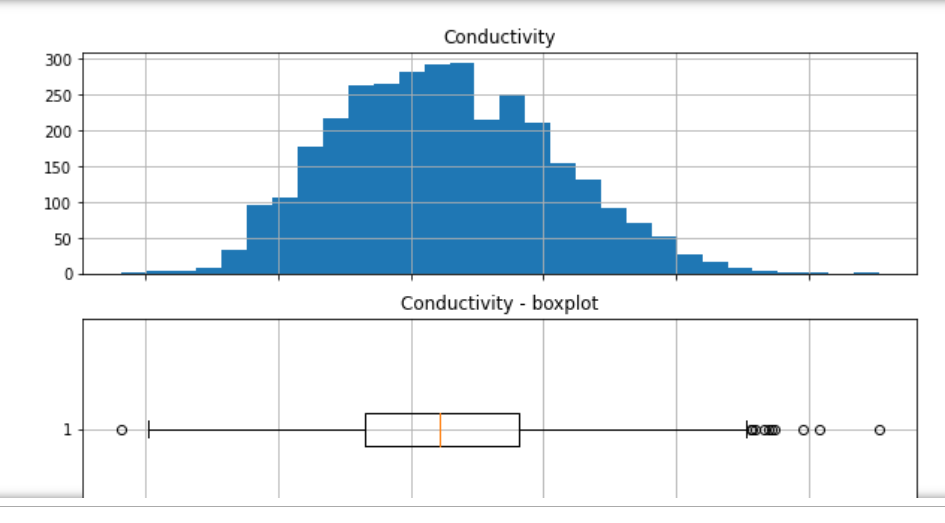


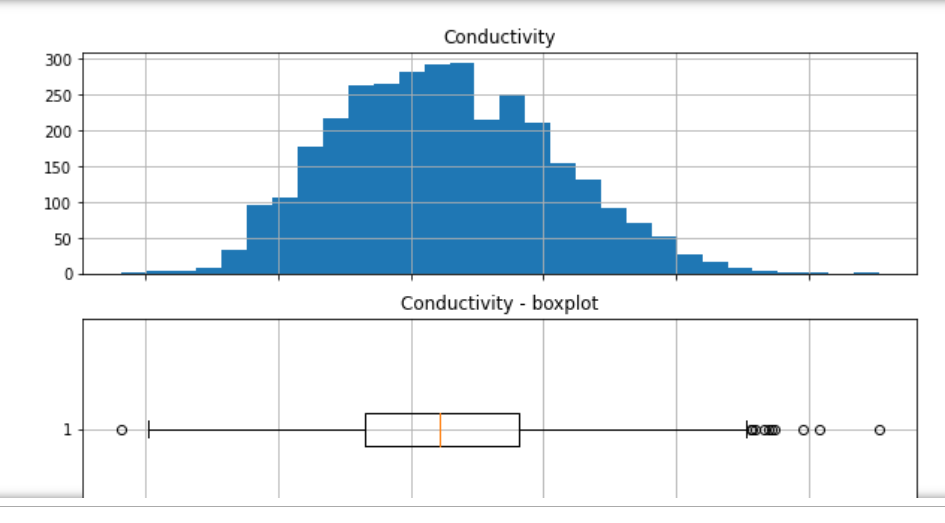


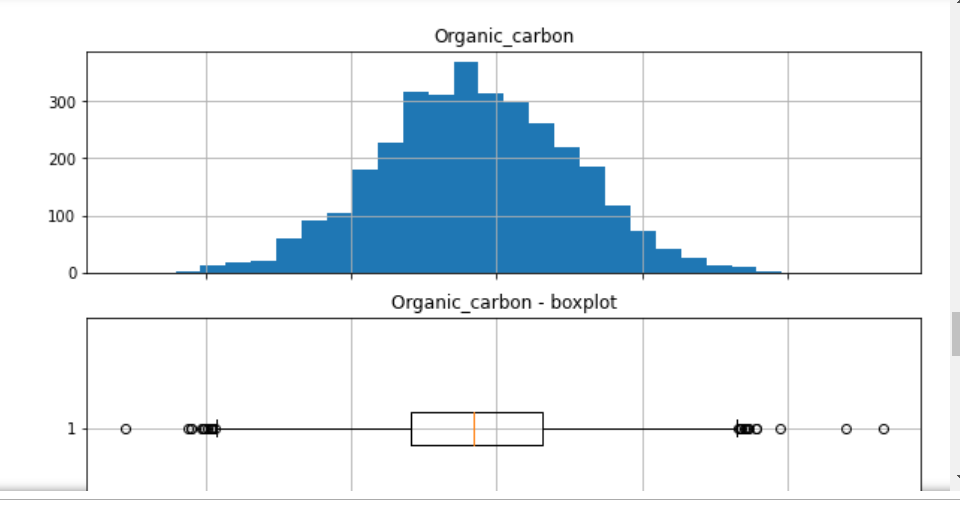


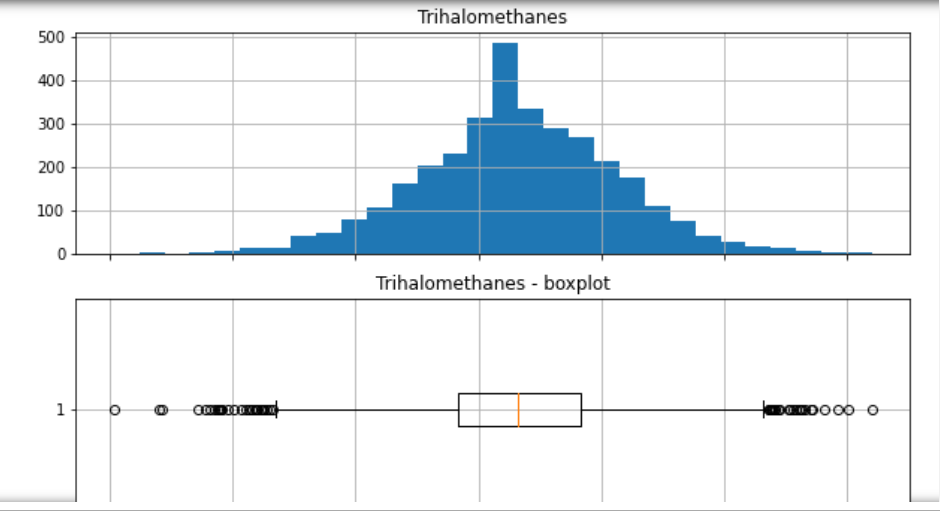
****

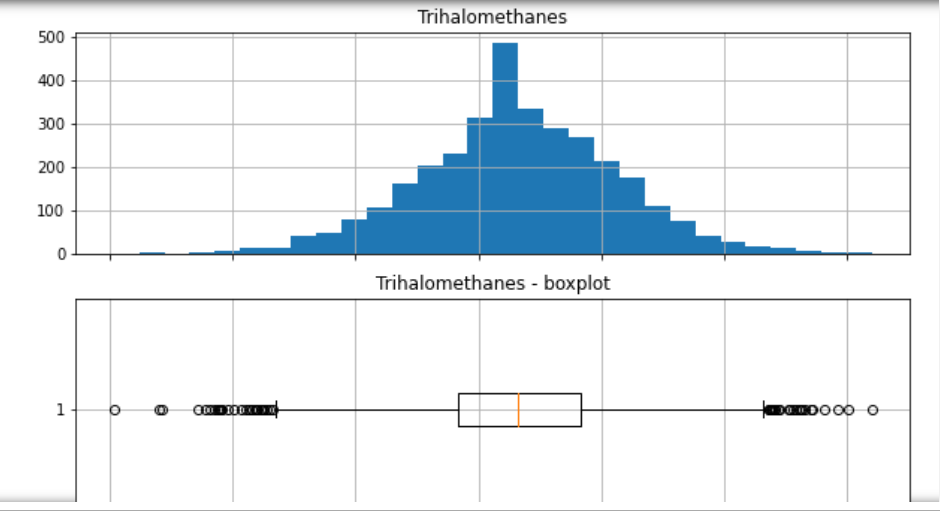


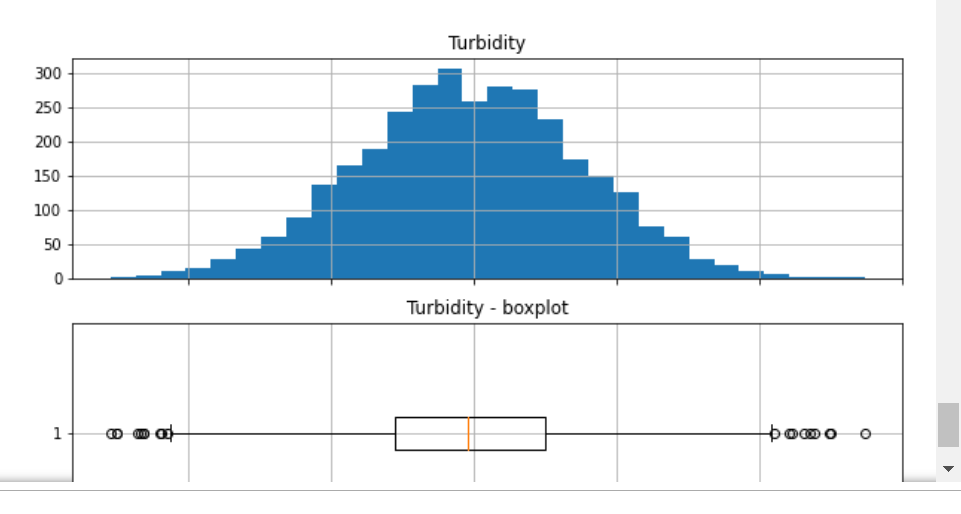












**Список литературы**

[1] Гапанюк Ю. Е. Лабораторная работа «Разведочный анализ данных. Исследование  
и визуализация данных» [Электронный ресурс] // GitHub. –– 2019. –– Режим доступа: https://github.com/ugapanyuk/ml\_course/wiki/LAB\_EDA\_VISUALIZATION (дата  
обращения: 13.02.2019)

[2] <https://www.kaggle.com/datasets>