****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".»

Выполнил:  
студент группы ИУ5-22М  
Лун Сыхань

Москва — 2024г.

**1. Цель лабораторной работы**

Изучить различные методы визуализации данных [1].

**2. Задание**

Выбрать набор данных (датасет). Вы можете найти список свободно распространяемых датасетов здесь.

Для лабораторных работ не рекомендуется выбирать датасеты очень большого размера.

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

1. История должна содержать не менее 5 шагов (где 5 - рекомендуемое количество шагов). Каждый шаг содержит график и его текстовую интерпретацию.
2. На каждом шаге наряду с удачным итоговым графиком рекомендуется в юпитер-ноутбуке оставлять результаты предварительных "неудачных" графиков.
3. Не рекомендуется повторять виды графиков, желательно создать 5 графиков различных видов.
4. 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.
5. История должна содержать итоговые выводы. В реальных "историях о данных" именно эти выводы представляют собой основную ценность для предприятия.

Сформировать отчет и разместить его в своем репозитории на github.

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

В этой тетради я буду использовать графики для визуализации взаимосвязи между переменными в наборе данных "Фильмы Marvel".

**The dataset includes the following columns:**  
Movie Title

Release Date (USA)

Phase

Genre

Movie Rating

Lead Role

Runtime (min)

Production Budget

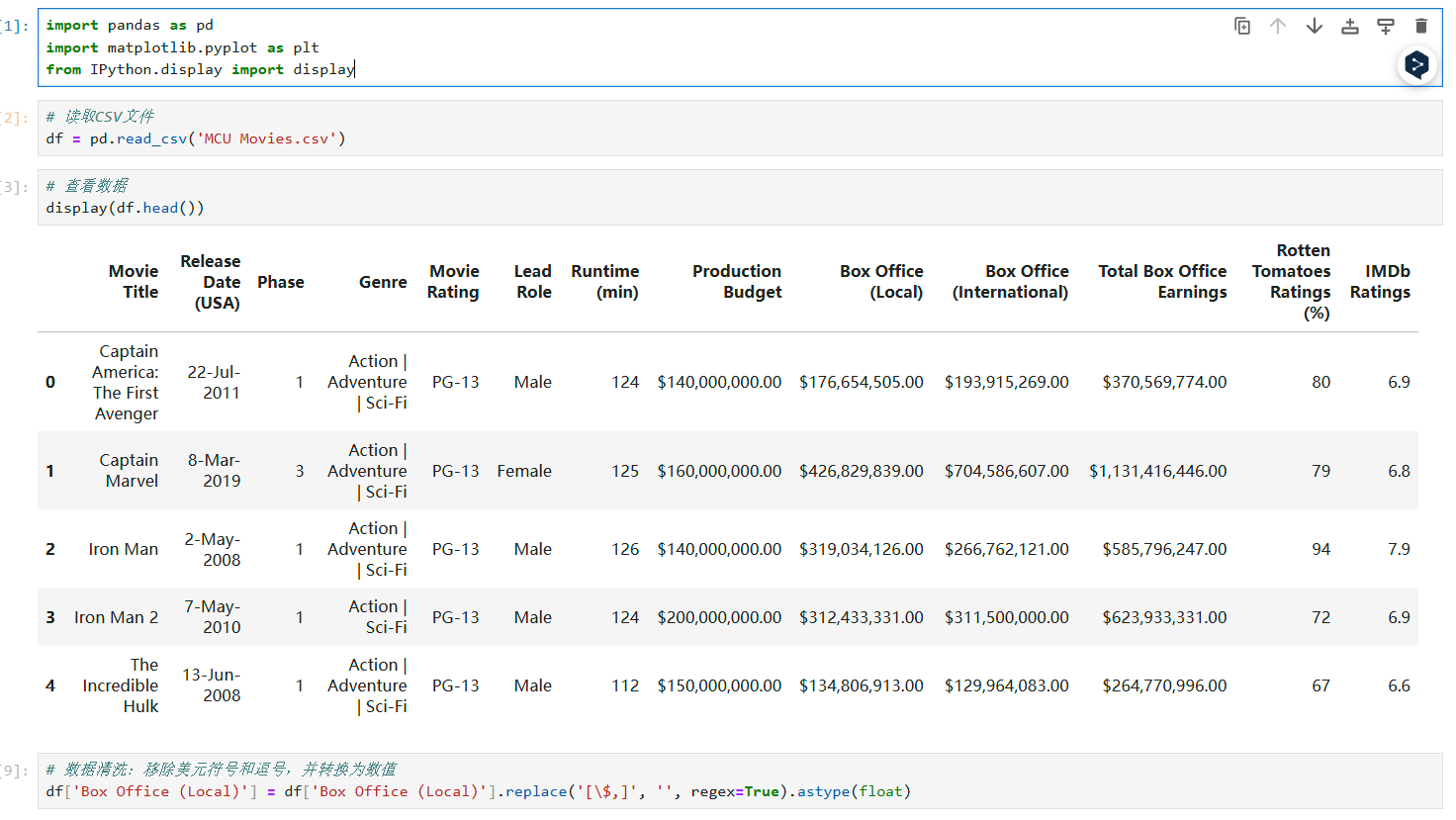
Box Office (Local)

Box Office (International)

Total Box Office Earnings

Rotten Tomatoes Ratings (%)

IMDb Ratings

**3.2. Основные характеристики набора данных**  


图形用户界面, 应用程序

描述已自动生成 图表, 直方图

描述已自动生成 图表, 散点图

描述已自动生成 图表, 条形图

描述已自动生成 图表, 散点图

描述已自动生成

**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>