****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:  
студент группы ИУ5-22М  
Джин Шуо

Москва — 2024 г.

**1. Цель лабораторной работы**

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

**2. Задание**

* Выбрать набор данных (датасет). Вы можете найти список свободно распространяемых датасетов [здесь.](https://github.com/ugapanyuk/courses_current/wiki/DSLIST)

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. текст программы**

import pandas as pd

import matplotlib.pyplot as plt

# read data

data = {

'rank': [1, 2, 3, 4, 5],

'university': ['Harvard', 'MIT', 'Stanford', 'Cambridge', 'Oxford'],

'overall\_score': [100, 99, 98, 97, 96],

'international\_students\_ratio': [0.2, 0.3, 0.25, 0.4, 0.35],

'international\_research\_network': [80, 85, 70, 90, 75],

'sustainability': [70, 60, 75, 80, 65],

'employer\_reputation': [95, 90, 85, 88, 92]

}

df = pd.DataFrame(data)

df = pd.read\_csv('top 100 world university 2024 new.csv')

# Scatter

plt.figure(figsize=(8, 6))

plt.scatter(df['international\_students\_ratio'], df['international\_research\_network'], color='orange', alpha=0.5)

plt.xlabel('International Students Ratio')

plt.ylabel('International Research Network')

plt.title('Relationship between International Students Ratio and International Research Network')

plt.grid(True)

plt.tight\_layout()

plt.show()

# Line

plt.figure(figsize=(10, 6))

plt.plot(df['rank'], df['sustainability'], marker='o', linestyle='-', color='green')

plt.xlabel('Rank')

plt.ylabel('Sustainability Score')

plt.title('Sustainability Score of Universities')

plt.xticks(rotation=45, ha='right')

plt.grid(True)

plt.tight\_layout()

plt.show()

# Box

plt.figure(figsize=(8, 6))

plt.boxplot(df['employer\_reputation'], vert=False)

plt.xlabel('Employer Reputation')

plt.title('Distribution of Employer Reputation Score')

plt.grid(True)

plt.tight\_layout()

plt.show()

# Pie

plt.figure(figsize=(8, 8))

labels = df['rank']

sizes = df['international\_students\_ratio']

plt.pie(sizes, labels=labels, autopct='%1.1f%%', startangle=140)

plt.axis('equal')

plt.title('Distribution of International Students Ratio among Universities')

plt.tight\_layout()

plt.show()

#bar

plt.figure(figsize=(10, 6))

plt.bar(df['rank'], df['overall\_score'], color='skyblue')

plt.xlabel('Rank')

plt.ylabel('Overall Score')

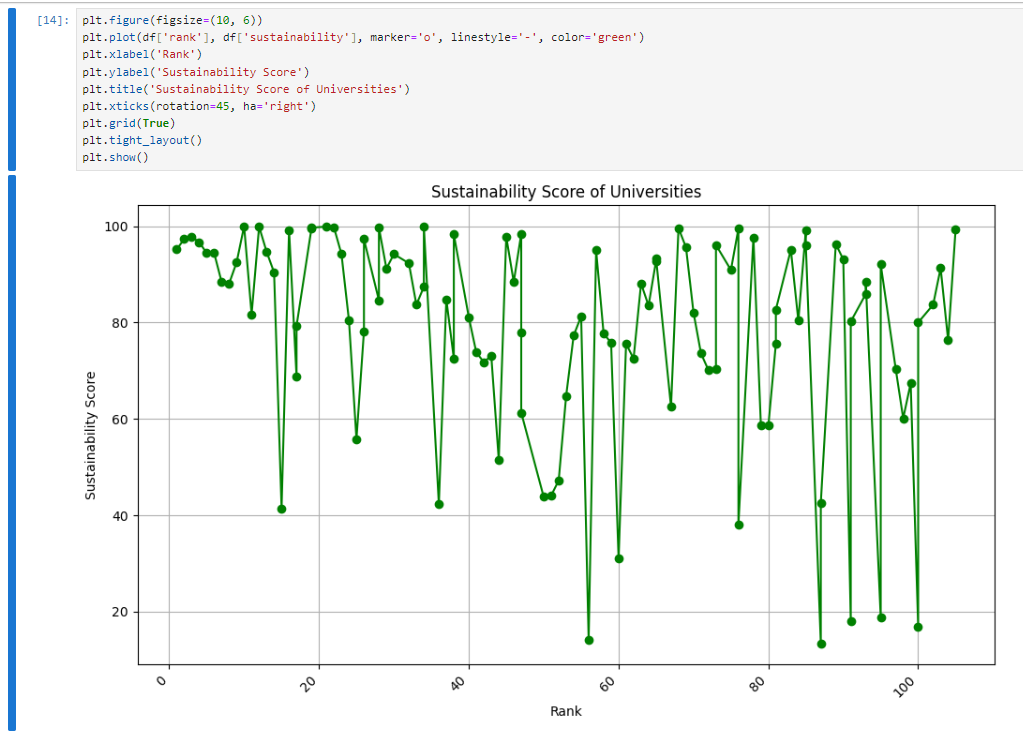
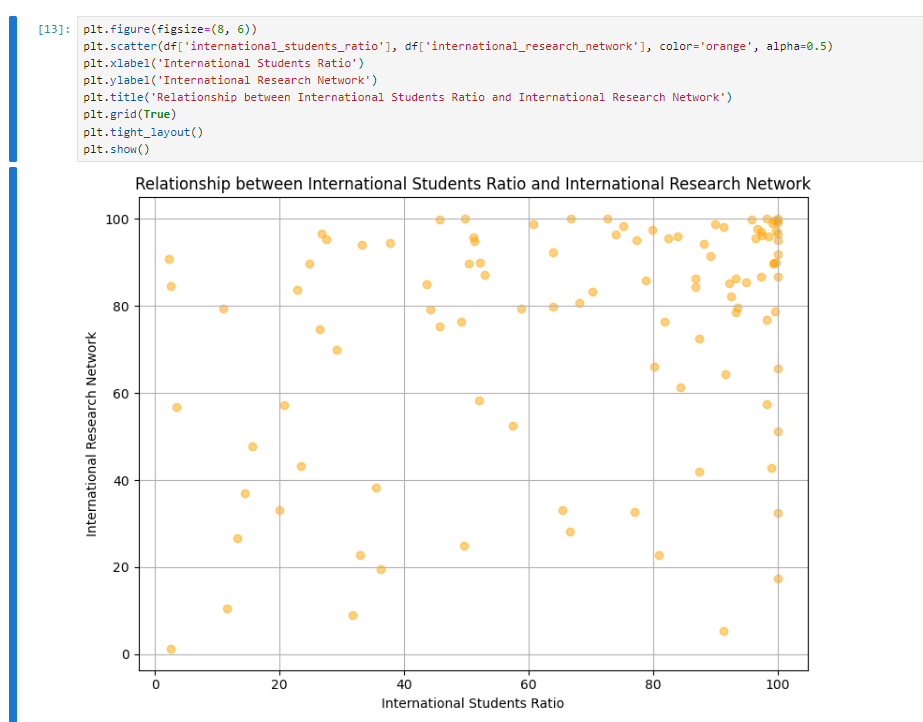
plt.title('Overall Score of Universities')

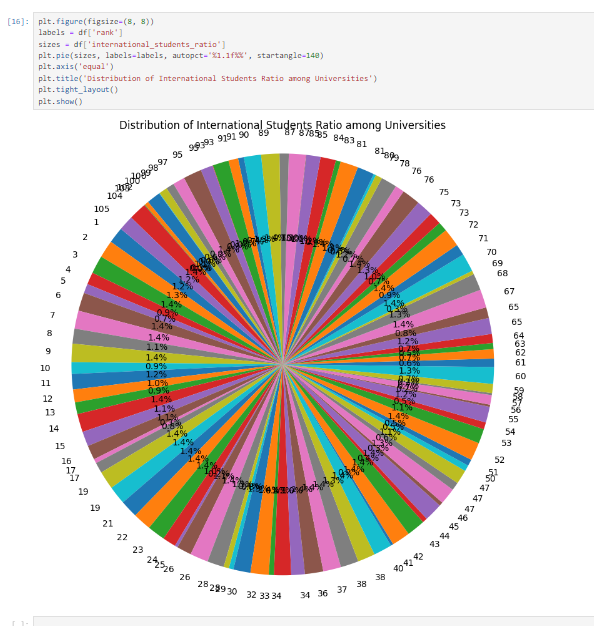
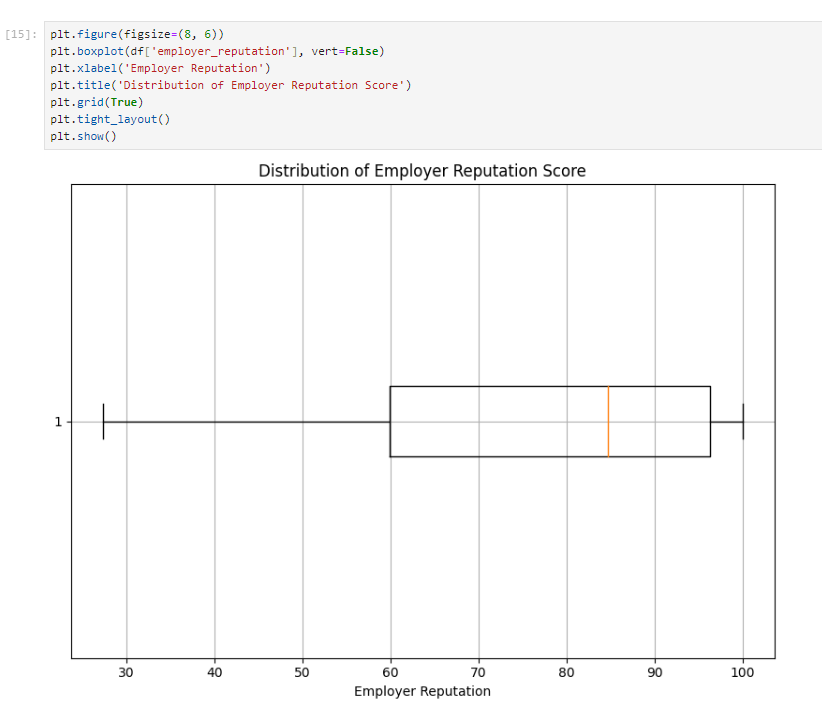
plt.xticks(rotation=45, ha='right')

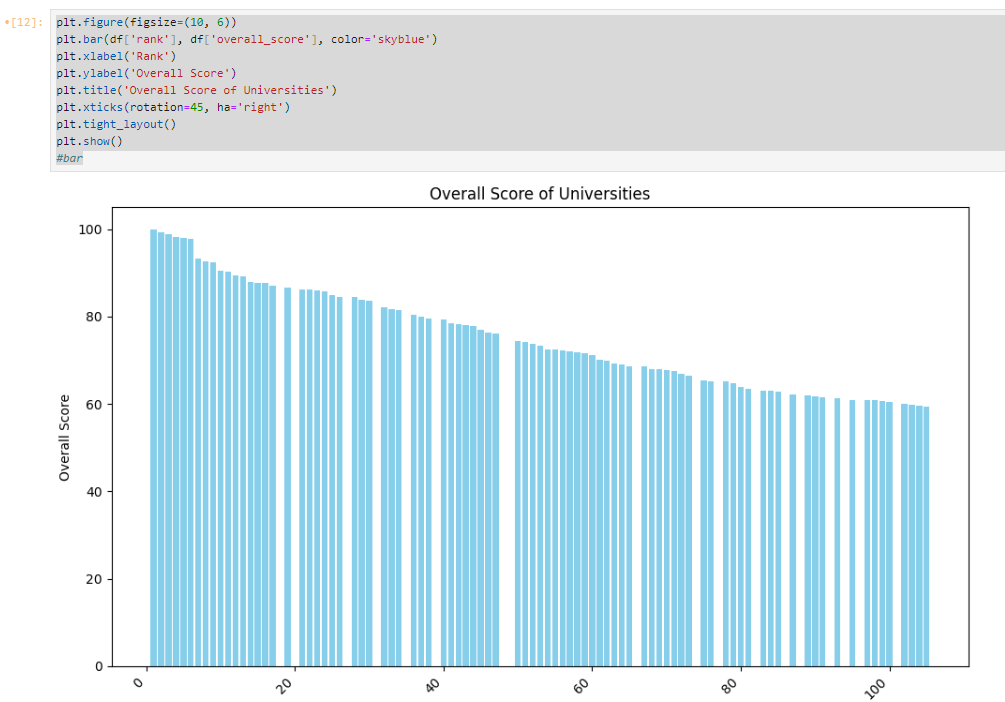
plt.tight\_layout()

plt.show()

**4.экранные формы с примерами выполнения программы.**







**Список литературы**

[1] Гапанюк Ю. Е. LAB\_MMO\_\_DATA\_STORYЛабораторная работа №1Создание "истории о данных" (Data Storytelling)// GitHub. –– 2024. –– Режим доступа:https://github.com/ugapanyuk/courses\_current/wiki/LAB\_MMO\_\_DATA\_STORY#%D0%BB%D0%B0%D0%B1%D0%BE%D1%80%D0%B0%D1%82%D0%BE%D1%80%D0%BD%D0%B0%D1%8F-%D1%80%D0%B0%D0%B1%D0%BE%D1%82%D0%B0-1

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