

МГТУ им. Н. Э. Баумана, кафедра ИУ5
курс “Технологии машинного обучения”

Лабораторная работа №1

«Разведочный анализ данных. Исследование и
визуализация данных»

ВЫПОЛНИЛ:

Ерохин И.А.

Группа: ИУ5-61Б

ПРОВЕРИЛ:

Гапанюк Ю.Е.

Москва 2020

Цель лабораторной работы: изучение различных методов визуализации данных.

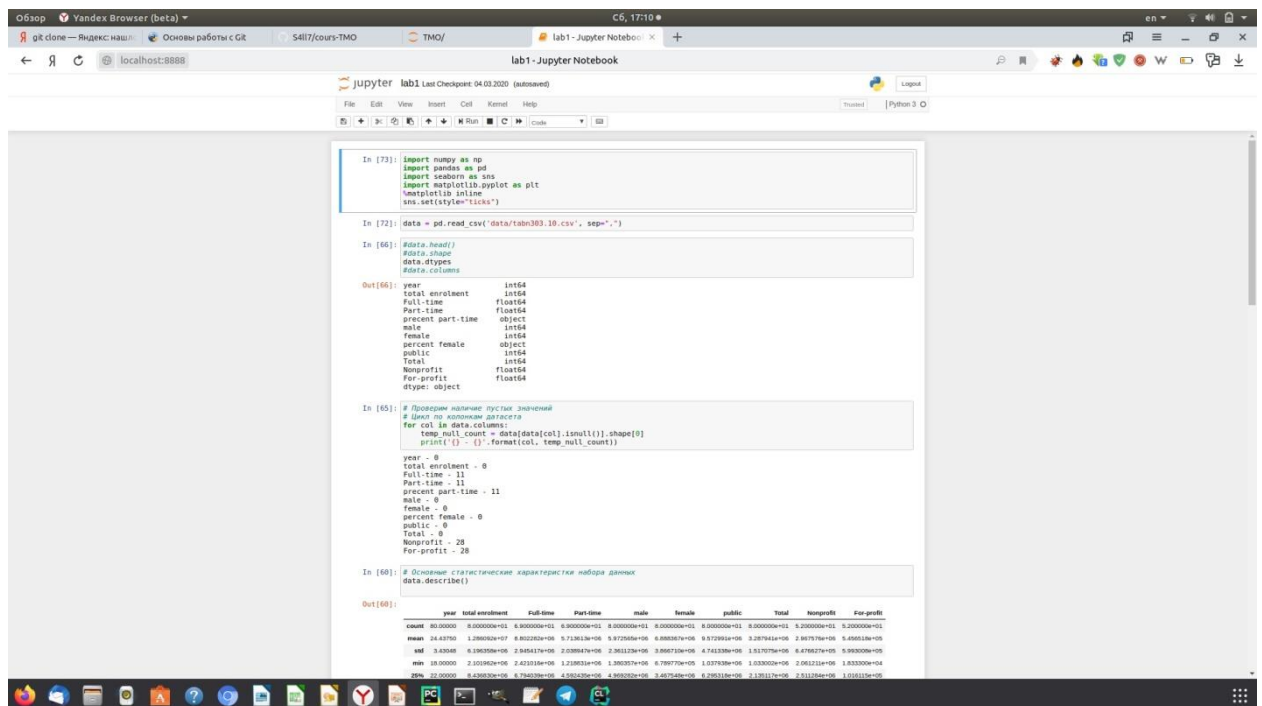
Задание:

- Выбрать набор данных (датасет).

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.

Текст программы и экранные формы с примерами выполнения программы:



```
In [73]: import numpy as np
import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt
matplotlib inline
sns.set(style='ticks')
```

```
In [72]: data = pd.read_csv('data/tabs303.10.csv', sep=',')
```

```
In [65]: #data.head()
#data.shape
data.dtypes
#data.columns
```

```
Out[65]: year                int64
total enrollment            int64
Full-time                  float64
Part-time                  float64
percent part-time          object
male                       int64
female                     int64
percent female             object
public                     int64
Total                      int64
Nonprofit                  float64
For-profit                  float64
dtype: object
```

```
In [65]: # Проверим наличие пустых значений
# (здесь мы используем isna)
for col in data.columns:
    temp_null_count = data[data[col].isnull()].shape[0]
    print('{} - {}'.format(col, temp_null_count))
```

```
year - 0
total enrollment - 0
Full-time - 11
Part-time - 11
percent part-time - 11
male - 0
female - 0
percent female - 0
public - 0
Total - 0
Nonprofit - 29
For-profit - 28
```

```
In [60]: # Основные статистические характеристики набора данных
data.describe()
```

```
Out[60]:
```

	year	total enrollment	Full-time	Part-time	male	female	public	Total	Nonprofit	For-profit
count	30.00000	3.00000e+01	8.80000e+00	8.80000e+00	8.80000e+00	8.80000e+00	8.80000e+00	3.00000e+01	3.00000e+00	3.00000e+00
mean	24.43750	1.20000e+07	6.802302e+06	5.713613e+06	5.972506e+06	6.888367e+06	6.572893e+06	5.287841e+06	5.967579e+06	5.498515e+06
std	3.42048	6.138039e+06	2.945417e+06	2.038847e+06	2.361123e+06	3.886713e+06	4.742339e+06	5.517077e+06	6.476627e+06	5.993009e+06
min	18.00000	2.00000e+06	2.423318e+06	1.218831e+06	1.385357e+06	6.789770e+05	1.027839e+06	1.033002e+06	2.061211e+06	1.833300e+06
max	27.00000	8.428033e+06	6.782094e+06	4.652354e+06	4.905702e+06	8.487640e+06	8.295338e+06	7.131174e+06	7.111264e+06	1.030104e+06

