

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



Лабораторная работа №1
по дисциплине «Методы машинного обучения»
«Создание истории о данных»

ИСПОЛНИТЕЛЬ:

Цветкова Алена
Группа ИУ5-21М

ПРЕПОДАВАТЕЛЬ:

Гапанюк Ю.Е.

Цель лабораторной работы: изучение различных методов визуализация данных и создание истории на основе данных.

В качестве исходных данных был выбран датасет COVID-19 World Vaccination Progress. В нем содержится информация о том, сколько человек сделали прививку в разбивке по странам и дням.

```
data_from_csv = pd.read_csv("country_vaccinations.csv")
data_from_csv = data_from_csv[['country', 'date', 'daily_vaccinations']]
```

```
data_from_csv.head()
```

	country	date	daily_vaccinations
0	Albania	2021-01-10	NaN
1	Albania	2021-01-11	64.0
2	Albania	2021-01-12	64.0
3	Albania	2021-01-13	63.0
4	Albania	2021-01-14	66.0

Для анализа были добавлены колонки с количеством привитых в сумме за день и в сумме во всей стране.

```
data_sum_by_country = data_from_csv.groupby(['country'], as_index=False)['daily_vaccinations'].agg('sum')
data_sum_by_dt = data_from_csv.groupby(['date'], as_index=False)['daily_vaccinations'].agg('sum')
```

```
df = pd.merge(data_from_csv, data_sum_by_country, on='country')
df = pd.merge(df, data_sum_by_dt, on='date')
```

```
df.head()
```

	country	date	daily_vaccinations_x	daily_vaccinations_y	daily_vaccinations
0	Albania	2021-01-10	NaN	1071.0	1891340.0
1	Argentina	2021-01-10	13365.0	613982.0	1891340.0
2	Austria	2021-01-10	NaN	337798.0	1891340.0
3	Bahrain	2021-01-10	4080.0	213122.0	1891340.0
4	Belgium	2021-01-10	3757.0	520240.0	1891340.0

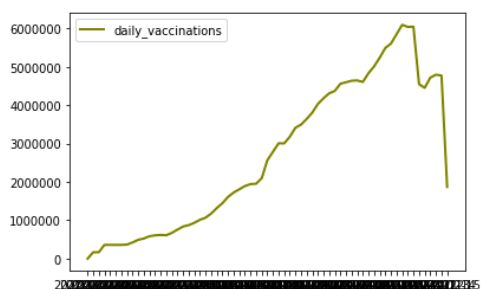
Для наглядного представления прогресса был построен график с количеством привитых по дням.

Распределение по дням

```
df = df.sort_values(by='date')
```

```
plt.plot('date', 'daily_vaccinations', data=df, marker='', color='olive', linewidth=2)
plt.legend()
```

<matplotlib.legend.Legend at 0x120005ac8>



Была построена гистограмма для стран, в которых больше всего привитых.

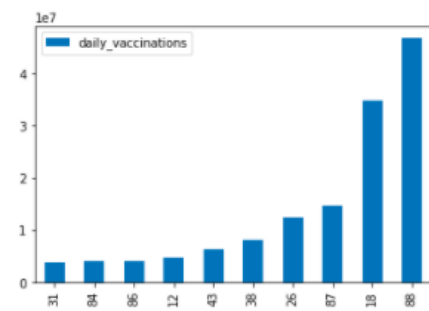
Распределение по странам

```
data_sum_by_country = pd.DataFrame(data=data_sum_by_country)
data_sum_by_country_top = data_sum_by_country.sort_values(by='daily_vaccinations')
data_sum_by_country_top = data_sum_by_country_top.tail(10)
```

data_sum_by_country_top

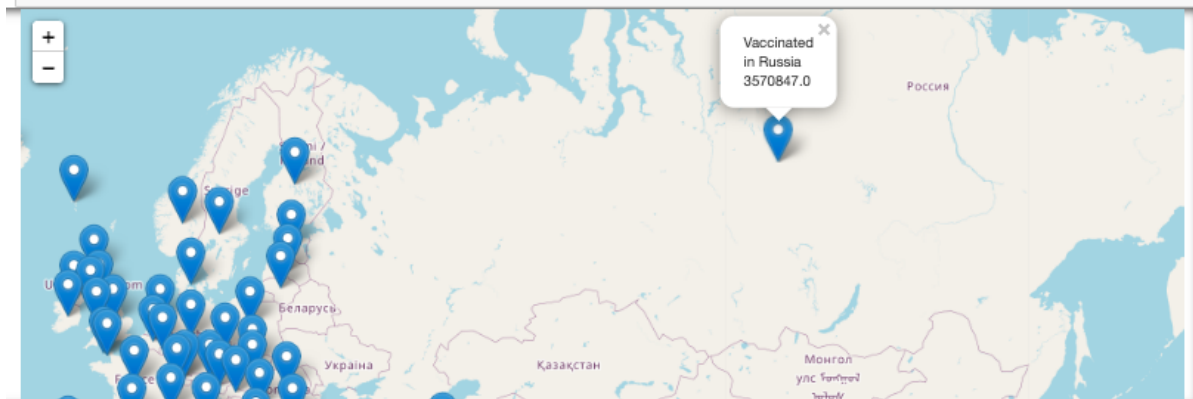
	country	daily_vaccinations
31	Germany	3882758.0
84	Turkey	4091063.0
86	United Arab Emirates	4152403.0
12	Brazil	4690767.0
43	Israel	6306012.0
38	India	8241026.0
26	England	12488524.0
87	United Kingdom	14715317.0
18	China	34922496.0
88	United States	46807045.0

```
ax = data_sum_by_country_top.plot.bar(y='daily_vaccinations', rot=90)
```



Также для визуализации разбивки привитых по странам была построена карта.

```
import folium
world_map = folium.Map(location=[20, 0], zoom_start=4)
for i in range(0, len(data_sum_by_country)):
    folium.Marker([data_sum_by_country.iloc[i]['Latitude'], data_sum_by_country.iloc[i]['Longitude']], popup = "Vaccinated in Russia 3570847.0")
world_map
```



Вывод: в процессе выполнения лабораторной работы были изучены различные методы визуализации данных.