

## **Практическая работа №6**

### **Работа с облачными системами хранения и визуализации данных**

**Цель занятия:** приобрести практические навыки в области создания интеллектуальных систем с использованием современных средств разработки.

### **Пояснения к работе**

В данной работе взаимодействие источниками данных будет проходить в Jupyter Notebook на базе Google Colab (см. Методические указания к Практической работе №3, <https://colab.research.google.com/notebooks/intro.ipynb>, <https://github.com/deepmipt/dlschl/wiki/Инструкция-по-работе-с-Google-Colab>).

Эта лабораторная работа в значительной степени полагается на библиотеки pandas и NumPy для обработки, анализа и визуализации данных. Основная библиотека для построения графиков, которую мы исследуем в этой лабораторной работе - это Folium.

Folium - это мощная библиотека Python, которая помогает создавать несколько типов карт. Тот факт, что результаты Folium интерактивны, делает эту библиотеку очень полезной для создания информационных панелей. Folium опирается на сильные стороны экосистемы Python и возможности сопоставления библиотеки Leaflet.js. Folium позволяет легко визуализировать данные, обработанные в Python, на интерактивной карте Leaflet. Библиотека имеет ряд встроенных наборов фрагментов из OpenStreetMap, Mapbox и Stamen и поддерживает настраиваемые наборы фрагментов с ключами API Mapbox или Cloudmade. Folium поддерживает наложения как GeoJSON, так и TopoJSON, а также привязку данных к этим наложениям для создания карт с цветовыми схемами.

Наборы данных:

Инциденты полицейского управления Сан-Франциско за 2016 год - инциденты полицейского управления из портала публичных данных Сан-Франциско. Инциденты получены из системы сообщений о преступлениях Департамента полиции Сан-Франциско (SFPD). Обновляется ежедневно, отображая данные за весь 2016 год.

## Программа работы

1. Переходим к созданному в рамках П.р. №4 аккаунту в Google Colab и создаем там новый блокнот.
2. Импортируем необходимые для работы библиотеки:

```
import numpy as np # useful for many scientific computing in Python
import pandas as pd # primary data structure library
```

2. Устанавливаем библиотеку folium:

```
!pip install folium==0.5.0
import folium

print('Folium installed and imported!')
```

3. Создаем карты мира в Folium. Создается объект Folium Map и затем запускается на отображение. Карты интерактивны, поэтому вы можете увеличивать любую интересующую область, несмотря на начальный уровень масштабирования.

```
# define the world map
world_map = folium.Map()

# display world map
world_map
```

4. Создаем карту с центром в Канаде и изменяем уровень масштабирования, чтобы увидеть, как он влияет на визуализированную карту.

```
# define the world map centered around Canada with a low zoom level
world_map = folium.Map(location=[56.130, -106.35], zoom_start=4)

# display world map
world_map
```

5. Еще одна интересная особенность Folium - вы можете создавать карты разных стилей. Попробуйте различные варианты [[https://python-visualization.github.io/folium/latest/user\\_guide/raster\\_layers/tiles.html](https://python-visualization.github.io/folium/latest/user_guide/raster_layers/tiles.html)] и выберите тот, который по Вашему мнению лучше всего.

```
# create a Stamen Toner map of the world centered around Canada
world_map = folium.Map(location=[56.130, -106.35], zoom_start=4, tiles='OpenStreetMap')

# display map
world_map
```

```
# create a Stamen Toner map of the world centered around Canada
world_map = folium.Map(location=[56.130, -106.35], zoom_start=4, tiles='cartodbpositron')

# display map
world_map
```

```
# create a world map with a Mapbox Bright style.
world_map = folium.Map(tiles='Cartodb dark_matter')

# display the map
world_map
```

## 6. Загружаем датасет и запрашиваем обзор его содержимого.

`'https://github.com/shihao-wen/IBM-Data-Science-Professional-Certificate/blob/master/6.%20Data%20Visualization/Final%20Assignment/Police_Department_Incidents_-_Previous_Year__2016_.csv?raw=true'`

```
df_incidents = pd.read_csv('https://github.com/shihao-wen/IBM-Data-Science-Professional-
Certificate/blob/master/6.%20Data%20Visualization/Final%20Assignment/Police_Department_Incidents_-_
_Previous_Year__2016_.csv?raw=true')

print('Dataset downloaded and read into a pandas dataframe!')
df_incidents.head()
```

## 7. Объем датасета слишком большим, поэтому ограничиваем его до первых 100 записей.

```
# get the first 100 crimes in the df_incidents dataframe
limit = 100
df_incidents = df_incidents.iloc[0:limit, :]
```

8. Визуализируем, где преступления имели место в городе Сан-Франциско. Для примера использован стиль по умолчанию и начальный уровень масштабирования до 12 (студенту необходимо выбрать эти параметры самостоятельно).

```
# San Francisco latitude and longitude values
latitude = 37.77
longitude = -122.42
# create map and display it
sanfran_map = folium.Map(location=[latitude, longitude], zoom_start=12)

incidents = folium.map.FeatureGroup()

# loop through the 100 crimes and add each to the incidents feature group
for lat, lng, in zip(df_incidents.Y, df_incidents.X):
    incidents.add_child(
        folium.features.CircleMarker(
            [lat, lng],
            radius=5, # define how big you want the circle markers to be
```

```
        color='yellow',  
        fill=True,  
        fill_color='blue',  
        fill_opacity=0.6  
    )  
)  
  
# add incidents to map  
sanfran_map.add_child(incidents)
```

9. Измените вид карты и представьте результаты в отчете.

## КОНТРОЛЬНЫЕ ВОПРОСЫ

1. Что такое Folium?