ЛЕКЦИЯ 1

**Библиотека pandas
для работы с данными.
Знакомство с
JupyterNotebook и Markdown**

Пример .ipynb файла: https://github.com/saba070201/test_repo_suai/blob/main/module2/pr1/algorithms%20(1).ipynb

ОСОБЕННОСТИ .IPYNB ФАЙЛОВ

- 1 ячейка - 1 вывод
- Позволяет использовать Markdown и LaTeX
- Экспортируется в pdf и другие форматы

Для работы с .ipynb файлами существует несколько решений:

- VSCode(https://code.visualstudio.com/)
- Jupyter-Notebook (https://jupyter.org/)
- Google Collab (https://colab.research.google.com/)

УСТАНОВКА JUPYTER-NOTEBOOK

1. pip install notebook(установка)
2. jupyter-notebook (запуск)

Для работы с .ipynb в VsCode можно просто создать .ipynb файл в любой понравившейся директории.

Отличным примером Markdown файла служат README.md файлы в git репозиториях (пример: https://github.com/pandas-dev/pandas/blob/main/README.md).

https://rosstat.gov.ru/opendata/7708234640-mpn2015744/data-20161226-structure-20161225.csv
https://github.com/pandas-dev/pandas/blob/main/doc/data/titanic.csv

Пример .csv файла.

data-20161226-structure-20161225 (1)

value	area	urban	gender
130565	Российская Федерация	Городское и сельское население	Оба пола
68554	Российская Федерация	Городское и сельское население	Мужчины
62011	Российская Федерация	Городское и сельское население	Женщины
61188	Российская Федерация	Городское население	Оба пола
31561	Российская Федерация	Городское население	Мужчины
29627	Российская Федерация	Городское население	Женщины
69377	Российская Федерация	Сельское население	Оба пола
36993	Российская Федерация	Сельское население	Мужчины
32384	Российская Федерация	Сельское население	Женщины
26173	Центральный федеральный округ	Городское и сельское население	Оба пола
13868	Центральный федеральный округ	Городское и сельское население	Мужчины
12305	Центральный федеральный округ	Городское и сельское население	Женщины
13090	Центральный федеральный округ	Городское население	Оба пола
6782	Центральный федеральный округ	Городское население	Мужчины
6308	Центральный федеральный округ	Городское население	Женщины
13083	Центральный федеральный округ	Сельское население	Оба пола
7086	Центральный федеральный округ	Сельское население	Мужчины
5997	Центральный федеральный округ	Сельское население	Женщины
1121	Белгородская область	Городское и сельское население	Оба пола
558	Белгородская область	Городское и сельское население	Мужчины
563	Белгородская область	Городское и сельское население	Женщины
411	Белгородская область	Городское население	Оба пола
201	Белгородская область	Городское население	Мужчины
210	Белгородская область	Городское население	Женщины

Pandas - это быстрый, мощный, гибкий и простой в использовании инструмент для анализа и манипуляции данных с открытым исходным кодом, построенный на основе языка программирования Python.

```
pip install pandas
```

Series

Создание объеката Series

In [54]:
```
s1 = pd.Series({'ages': [10, 20, 30]})
s2=pd.Series([10, 20, 30],name='ages')
```

In [55]:
```
s1
```

Out[55]:
```
ages    [10, 20, 30]
dtype: object
```

In [56]:
```
s2
```

Out[56]:
```
0    10
1    20
2    30
Name: ages, dtype: int64
```

Операции над объектом Series

Переопределенные арифметические операции

In [57]:
```
s2=s2*2
s2
```

Out[57]:
```
0    20
1    40
2    60
Name: ages, dtype: int64
```

Некоторые методы объекта series

Нахождение максимума

```
In [58]: s2.max()
Out[58]: 60
```

Нахождение среднего

```
In [59]: s2.mean()
Out[59]: 40.0
```

Нахождение медианного

```
In [60]: s2.median()
Out[60]: 40.0
```

DataFrame

Представление датасета(в нашем случае .csv файла) в библиотеке pandas

Создание DataFrame в ручную

In [61]:
```
my_first_data_frame=pd.DataFrame({'names':['Misha','Vasya','Nikita'],'ages':s2})
my_first_data_frame
```

Out[61]:

	names	ages
0	Misha	20
1	Vasya	40
2	Nikita	60

In [62]:
```
type(my_first_data_frame)
```

Out[62]:
```
pandas.core.frame.DataFrame
```

Импорт DataFrame из .csv файла

In [63]:
```
titanic_data=pd.read_csv('titanic.csv')
titanic_data
```

Out[63]:

	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Cabin	Embarked
0	1	0	3	Braund, Mr. Owen Harris	male	22.0	1	0	A/5 21171	7.2500	NaN	S
1	2	1	1	Cumings, Mrs. John Bradley (Florence Briggs Th...	female	38.0	1	0	PC 17599	71.2833	C85	C
2	3	1	3	Heikkinen, Miss. Laina	female	26.0	0	0	STON/O2. 3101282	7.9250	NaN	S
3	4	1	1	Futrelle, Mrs. Jacques Heath (Lily May Peel)	female	35.0	1	0	113803	53.1000	C123	S
4	5	0	3	Allen, Mr. William Henry	male	35.0	0	0	373450	8.0500	NaN	S
...	...	...	...	...	...	...	...	...	...	...	...	...
886	887	0	2	Montvila, Rev. Juozas	male	27.0	0	0	211536	13.0000	NaN	S
887	888	1	1	Graham, Miss. Margaret Edith	female	19.0	0	0	112053	30.0000	B42	S
888	889	0	3	Johnston, Miss. Catherine Helen "Carrie"	female	NaN	1	2	W./C. 6607	23.4500	NaN	S
889	890	1	1	Behr, Mr. Karl Howell	male	26.0	0	0	111369	30.0000	C148	C

In [64]:
```
type(titanic_data)
```

Out[64]:
```
pandas.core.frame.DataFrame
```

Работа с столбцами

In [65]:
```
titanic_data['Name']
```

Out[65]:
```
0                                Braund, Mr. Owen Harris
1      Cumings, Mrs. John Bradley (Florence Briggs Th...
2                                 Heikkinen, Miss. Laina
3           Futrelle, Mrs. Jacques Heath (Lily May Peel)
4                               Allen, Mr. William Henry
                             ...
886                                Montvila, Rev. Juozas
887                         Graham, Miss. Margaret Edith
888             Johnston, Miss. Catherine Helen "Carrie"
889                                Behr, Mr. Karl Howell
890                                  Dooley, Mr. Patrick
Name: Name, Length: 891, dtype: object
```

панель мониторинга

Выбор определенных данных

Выберем всех пассажиров титаника 1 класса, возраст которых < 50.

1 способ

In [66]:
```
mask=titanic_data.eval('Pclass==1 & Age<50')
```

In [67]:
```
titanic_data[mask]
```

Out[67]:

	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Cabin	Embarked
1	2	1	1	Cumings, Mrs. John Bradley (Florence Briggs Th...	female	38.0	1	0	PC 17599	71.2833	C85	C
3	4	1	1	Futrelle, Mrs. Jacques Heath (Lily May Peel)	female	35.0	1	0	113803	53.1000	C123	S
23	24	1	1	Sloper, Mr. William Thompson	male	28.0	0	0	113788	35.5000	A6	S
27	28	0	1	Fortune, Mr. Charles Alexander	male	19.0	3	2	19950	263.0000	C23 C25 C27	S
30	31	0	1	Uruchurtu, Don. Manuel E	male	40.0	0	0	PC 17601	27.7208	NaN	C
...	...	...	...	...	...	...	...	...	...	...	...	...
867	868	0	1	Roebling, Mr. Washington Augustus II	male	31.0	0	0	PC 17590	50.4958	A24	S
871	872	1	1	Beckwith, Mrs. Richard Leonard (Sallie Monypeny)	female	47.0	1	1	11751	52.5542	D35	S
872	873	0	1	Carlsson, Mr. Frans Olof	male	33.0	0	0	695	5.0000	B51 B53 B55	S
887	888	1	1	Graham, Miss. Margaret Edith	female	19.0	0	0	112053	30.0000	B42	S
889	890	1	1	Behr, Mr. Karl Howell	male	26.0	0	0	111369	30.0000	C148	C

142 rows × 12 columns

2 способ

In [68]:
```
titanic_data[(titanic_data['Age']<50) & (titanic_data['Pclass']==1)]
```

Out[68]:

	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Cabin	Embarked
1	2	1	1	Cumings, Mrs. John Bradley (Florence Briggs Th...	female	38.0	1	0	PC 17599	71.2833	C85	C
3	4	1	1	Futrelle, Mrs. Jacques Heath (Lily May Peel)	female	35.0	1	0	113803	53.1000	C123	S
23	24	1	1	Sloper, Mr. William Thompson	male	28.0	0	0	113788	35.5000	A6	S
27	28	0	1	Fortune, Mr. Charles Alexander	male	19.0	3	2	19950	263.0000	C23 C25 C27	S
30	31	0	1	Uruchurtu, Don. Manuel E	male	40.0	0	0	PC 17601	27.7208	NaN	C
...	...	...	...	...	...	...	...	...	...	...	...	...
867	868	0	1	Roebling, Mr. Washington Augustus II	male	31.0	0	0	PC 17590	50.4958	A24	S
871	872	1	1	Beckwith, Mrs. Richard Leonard (Sallie Monypeny)	female	47.0	1	1	11751	52.5542	D35	S
872	873	0	1	Carlsson, Mr. Frans Olof	male	33.0	0	0	695	5.0000	B51 B53 B55	S
887	888	1	1	Graham, Miss. Margaret Edith	female	19.0	0	0	112053	30.0000	B42	S
889	890	1	1	Behr, Mr. Karl Howell	male	26.0	0	0	111369	30.0000	C148	C

142 rows × 12 columns

ДОМАШНЕЕ ЗАДАНИЕ

- Установить JupyterNotebook
- Создать свой DataFrame (достаточно 3 столбцов по 5 значений в каждом)
- С помощью метода to_csv(«path/to/save») сохранить ваш DataFrame
- Оформить README.md файл в вашем репозитории на GitHub

КОНТРОЛЬНЫЕ ВОПРОСЫ