

Практична робота №1

Встановлення Jupyter Notebook. Створення сесії. Імпортування, обробка та друк даних.

Хід роботи

1. Встановити глобально та запустити Jupyter Notebook, для цього в консолі:

- встановити: `$ pip install notebook`

- встановити: `$ pip install ruyspark`

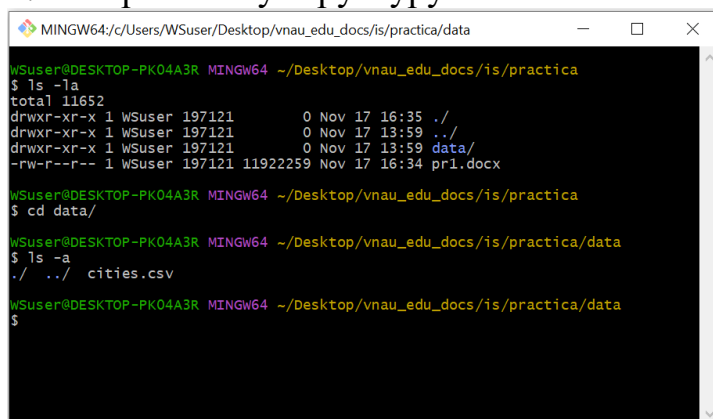
- перевірити: `$ pip list`

(має бути у списку таке – notebook 7.2.2)

- запустити: `$ jupyter notebook`

- підключитися: `http://localhost:8888/tree`

2. Створити таку структуру папок:



```
MINGW64~/Users/WSuser/Desktop/vnau_edu_docs/is/practica/data
wSuser@DESKTOP-PK04A3R MINGW64 ~/Desktop/vnau_edu_docs/is/practica
$ ls -la
total 11652
drwxr-xr-x 1 wSuser 197121  0 Nov 17 16:35 ./
drwxr-xr-x 1 wSuser 197121  0 Nov 17 13:59 ../
drwxr-xr-x 1 wSuser 197121  0 Nov 17 13:59 data/
-rw-r--r-- 1 wSuser 197121 11922259 Nov 17 16:34 pr1.docx

wSuser@DESKTOP-PK04A3R MINGW64 ~/Desktop/vnau_edu_docs/is/practica
$ cd data/

wSuser@DESKTOP-PK04A3R MINGW64 ~/Desktop/vnau_edu_docs/is/practica/data
$ ls -la
./ ../ cities.csv

wSuser@DESKTOP-PK04A3R MINGW64 ~/Desktop/vnau_edu_docs/is/practica/data
$
```

Вміст файлу «cities.csv» такий:

CityName,Region,Population,Area,PopulationDensity,IsCapital,NumberOfDistricts

Vinnytsia,Vinnytsia Oblast,370000,113.2,3267,True,3

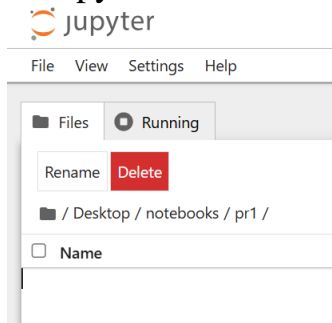
Khmelnyskyi,Vinnytsia Oblast,130000,90.0,1444,False,4

Zhmerinka,Vinnytsia Oblast,35000,40.5,864,False,2

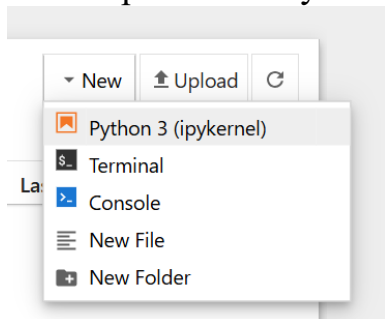
Mogilev-Podolsky,Vinnytsia Oblast,31000,18.2,1704,False,1

Kozyatyn,Vinnytsia Oblast,26000,25.7,1012,False,1

У Jupyter notebook створити таку структуру для даного практичного заняття:



3. Створити та запустити ноутбук:




Files **Running**

Select items to perform actions on them.

/ Desktop / notebooks / pr1 /

Name

 import_cities.ipynb

Select Kernel


Select kernel for: "import_cities.ipynb"

Python 3 (ipykernel)

Always start the preferred kernel **No Kernel** **Select**

jupyter import_cities Last Checkpoint: 1 hour ago

File Edit View Run Kernel Settings Help

 Code

[]: |

4. Увести код та отримати результат

jupyter import_cities Last Checkpoint: 2 minutes ago

File Edit View Run Kernel Settings Help

Trusted

JupyterLab Python 3 (ipykernel)

```
[1]: from pyspark.sql import SparkSession
```

```
[2]: from pyspark.sql import functions as f
```

```
[3]: spark = SparkSession.builder.master("local").appName("PR1").getOrCreate()
```

```
[9]: citiDF = spark.read.format("csv") \
      .option("header", "true") \
      .option("inferSchema", "true") \
      .load("C:/Users/WUser/Desktop/vnau_edu_docs/is/practica/data/cities.csv")
      citiDF.cache()
      citiDF.printSchema()

root
 |-- CityName: string (nullable = true)
 |-- Region: string (nullable = true)
 |-- Population: integer (nullable = true)
 |-- Area: double (nullable = true)
 |-- PopulationDensity: integer (nullable = true)
 |-- IsCapital: boolean (nullable = true)
 |-- NumberOfDistricts: integer (nullable = true)
```

```
[10]: citiDF.show(3)
```

CityName	Region	Population	Area	PopulationDensity	IsCapital	NumberOfDistricts
Vinnitsia	Vinnitsia Oblast	370000	113.2	3267	true	3
Khmelnytskyi	Vinnitsia Oblast	130000	90.0	1444	false	4
Zhmerinka	Vinnitsia Oblast	35000	40.5	864	false	2

only showing top 3 rows

[]: |

5.* Завдання для індивідуальної роботи – знайти та розказати викладачу як прибрати зайві пробіли на початку та у кінці рядків. Розказати до яких небезпечних ситуацій може привести наявність пробілів.

6.* Індивідуальна задача. Вибрати тему та згенерувати джерело даних по типу як у поточній практичній роботі у форматі *.csv та спробувати завантажити, задати структуру та вивести на екран:

- Автомобілі (модель, виробник, рік випуску, тип пального, ціна)
- Кліматичні дані (дата, температура, вологість, опади, швидкість вітру)
- Курси валют (дата, валюта, курс до USD, курс до EUR, зміна відсотків)
- Книги (назва, автор, рік видання, жанр, кількість сторінок)
- Фільми (назва, режисер, рік випуску, жанр, рейтинг IMDb)

- Технологічні продукти (назва, категорія, ціна, дата випуску, виробник)
- Ресторани (назва, місцезнаходження, кухня, середній чек, рейтинг)
- Університети (назва, країна, рейтинг, кількість студентів, рік заснування)
- Спортивні команди (назва, місто, вид спорту, кількість чемпіонатів, рік заснування)
- Музичні альбоми (назва, виконавець, жанр, рік випуску, тривалість)
- Туристичні пакети (назва, місцезнаходження, ціна, тривалість, включені послуги)
- Телефонні контакти (ім'я, номер телефону, електронна пошта, адреса, день народження)
- Лікарські засоби (назва, виробник, тип, вартість за одиницю, дата випуску)
- Погодні станції (назва, місцезнаходження, висота над рівнем моря, тип сенсорів, рік заснування)
- Електронні пристрої (модель, виробник, тип пристрою, ціна, рік випуску)
- Історичні події (дата, подія, місцезнаходження, учасники, наслідки)
- Аеропорти (назва, місто, країна, кількість рейсів на добу, рік заснування)
- Медичні заклади (назва, тип закладу, адреса, кількість лікарів, спеціалізація)
- Наукові дослідження (назва, галузь, рік початку, основні відкриття, керівник проекту)
- Політичні партії (назва, країна, лідер, рік заснування, ідеологія)

Висновки

Замість висновків, ви усі молодці!