Siddhant Mundhe

Data engineering batch 1

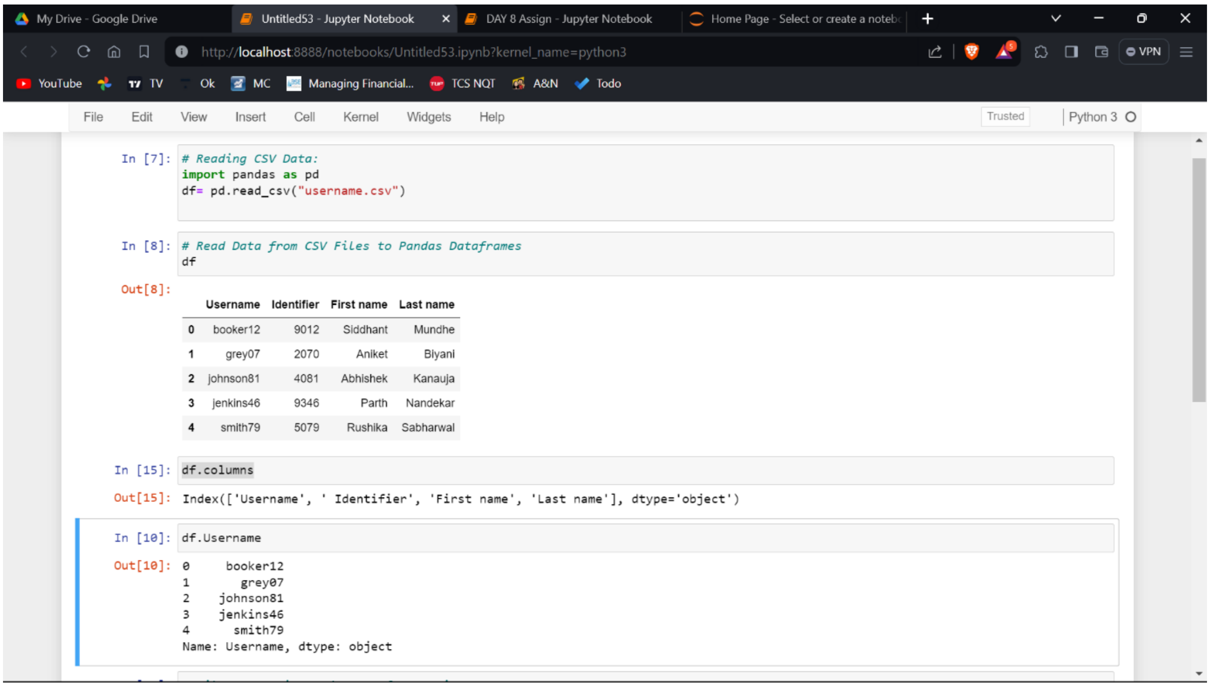
Coding Challenge 2

PANDAS

-Pandas is a popular open-source data manipulation and analysis library for Python.

-It provides data structures like DataFrame for efficient data manipulation with built-in functions for various tasks.

-Assuming you have a CSV file named Username.csv with some sample data, you can read it into a Pandas DataFrame using the read_csv function:


My Drive - Google Drive
Untitled53 - Jupyter Notebook
DAY 8 Assign - Jupyter Notebook
Home Page - Select or create a notebo
http://localhost:8888/notebooks/Untitled53.ipynb?kernel_name=python3
YouTube
TV
Ok
MC
Managing Financial...
TCS NQT
A&N
Todo
VPN
File
Edit
View
Insert
Cell
Kernel
Widgets
Help
Trusted
Python 3

```
In [7]: # Reading CSV Data:
        import pandas as pd
        df= pd.read_csv("username.csv")

In [8]: # Read Data from CSV Files to Pandas Dataframes
        df
Out[8]:
```

	Username	Identifier	First name	Last name
0	booker12	9012	Siddhant	Mundhe
1	grey07	2070	Aniket	Biyani
2	johnson81	4081	Abhishek	Kanauja
3	jenkins46	9346	Parth	Nandekar
4	smith79	5079	Rushika	Sabharwal

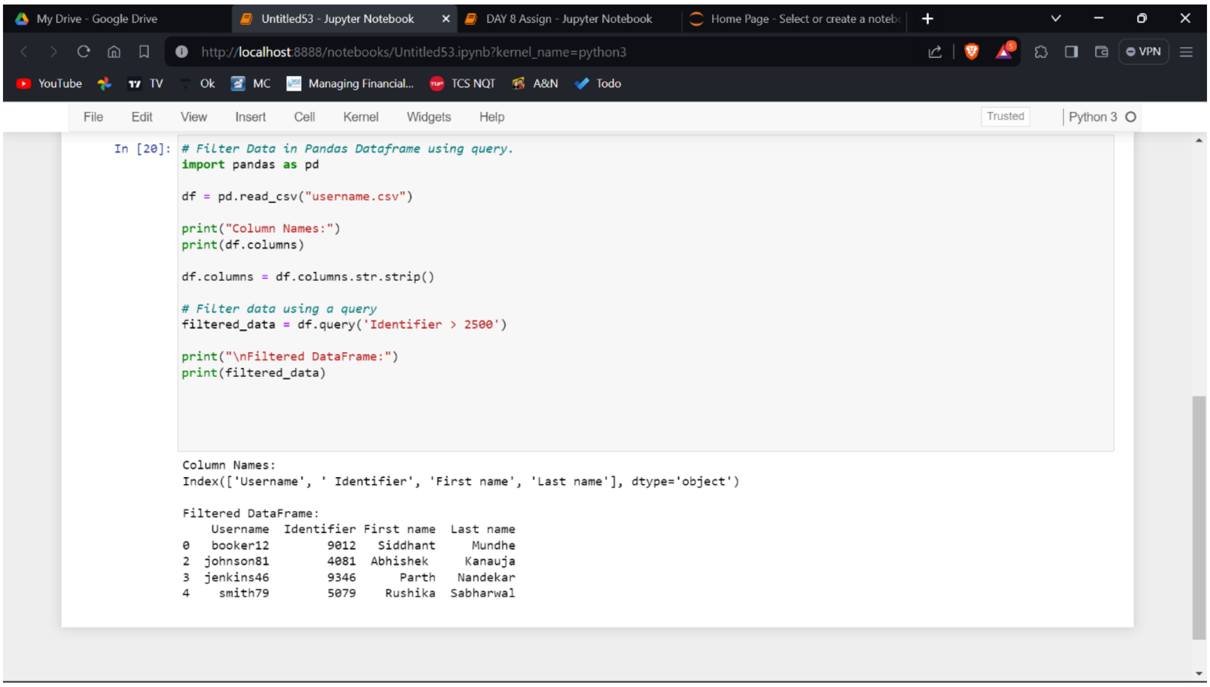
```
In [15]: df.columns
Out[15]: Index(['Username', ' Identifier', 'First name', 'Last name'], dtype='object')

In [10]: df.Username
Out[10]: 0     booker12
         1       grey07
         2    johnson81
         3    jenkins46
         4      smith79
         Name: Username, dtype: object
```

Filtering Data using Query:

-Pandas provides the query method for filtering data based on a query expression.

- To filter pandas DataFrame all you need to do is specify the condition within query function.

For example, let's say you want to filter rows where the 'Identifier > 2500':


My Drive - Google Drive
Untitled53 - Jupyter Notebook
DAY 8 Assign - Jupyter Notebook
Home Page - Select or create a noteb
http://localhost:8888/notebooks/Untitled53.ipynb?kernel_name=python3
YouTube
TV
Ok
MC
Managing Financial...
TCS NQT
A&N
Todo
File
Edit
View
Insert
Cell
Kernel
Widgets
Help
Trusted
Python 3

```
In [20]: # Filter Data in Pandas Dataframe using query.
         import pandas as pd

         df = pd.read_csv("username.csv")

         print("Column Names:")
         print(df.columns)

         df.columns = df.columns.str.strip()

         # Filter data using a query
         filtered_data = df.query('Identifier > 2500')

         print("\nFiltered DataFrame:")
         print(filtered_data)
```

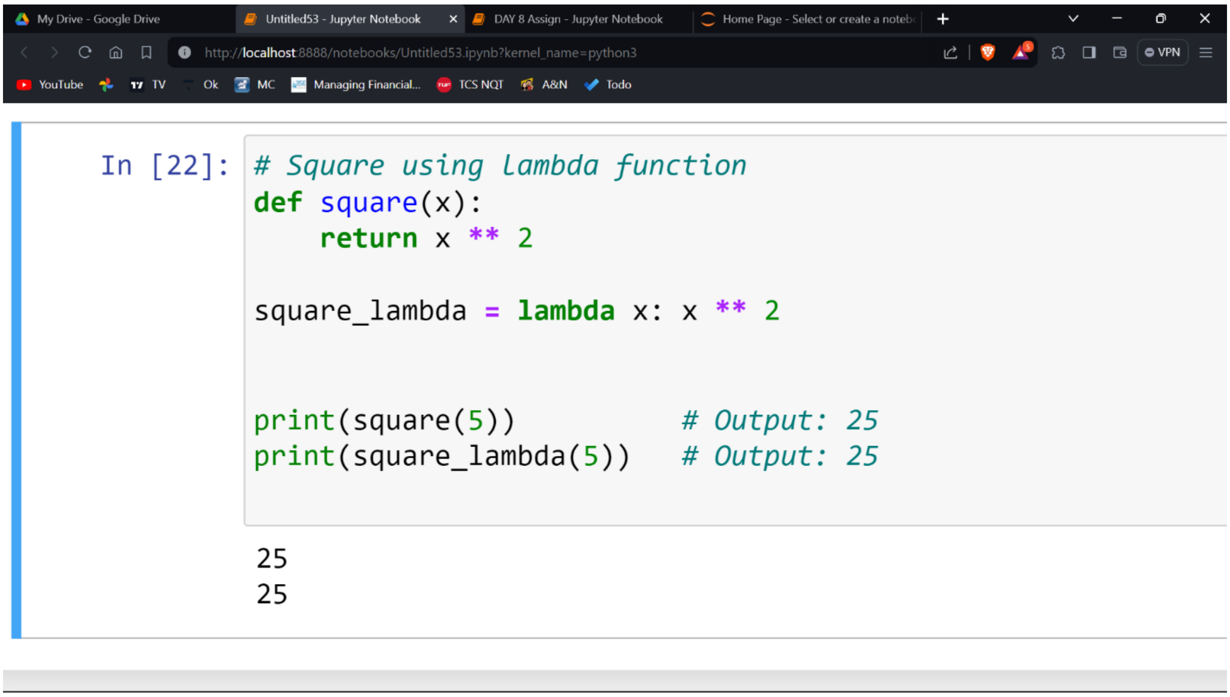
```
Column Names:
Index(['Username', ' Identifier', 'First name', 'Last name'], dtype='object')

Filtered DataFrame:
    Username  Identifier First name  Last name
0   booker12        9012   Siddhant     Mundhe
2  johnson81        4081  Abhishek     Kanauja
3  jenkins46        9346      Parth   Nandekar
4    smith79        5079    Rushika  Sabharwal
```

LAMBDA FUNCTION:

-In Python, a lambda function is a concise way to create anonymous functions. Lambda functions are often used for short, simple operations and are defined using the lambda keyword.

-The syntax for a lambda function is:

lambda arguments: expression


My Drive - Google Drive
Untitled53 - Jupyter Notebook
DAY 8 Assign - Jupyter Notebook
Home Page - Select or create a noteb
http://localhost:8888/notebooks/Untitled53.ipynb?kernel_name=python3
YouTube
TV
Ok
MC
Managing Financial...
TCS NQT
A&N
Todo

```
In [22]: # Square using lambda function
         def square(x):
             return x ** 2

         square_lambda = lambda x: x ** 2


         print(square(5))          # Output: 25
         print(square_lambda(5))   # Output: 25
```

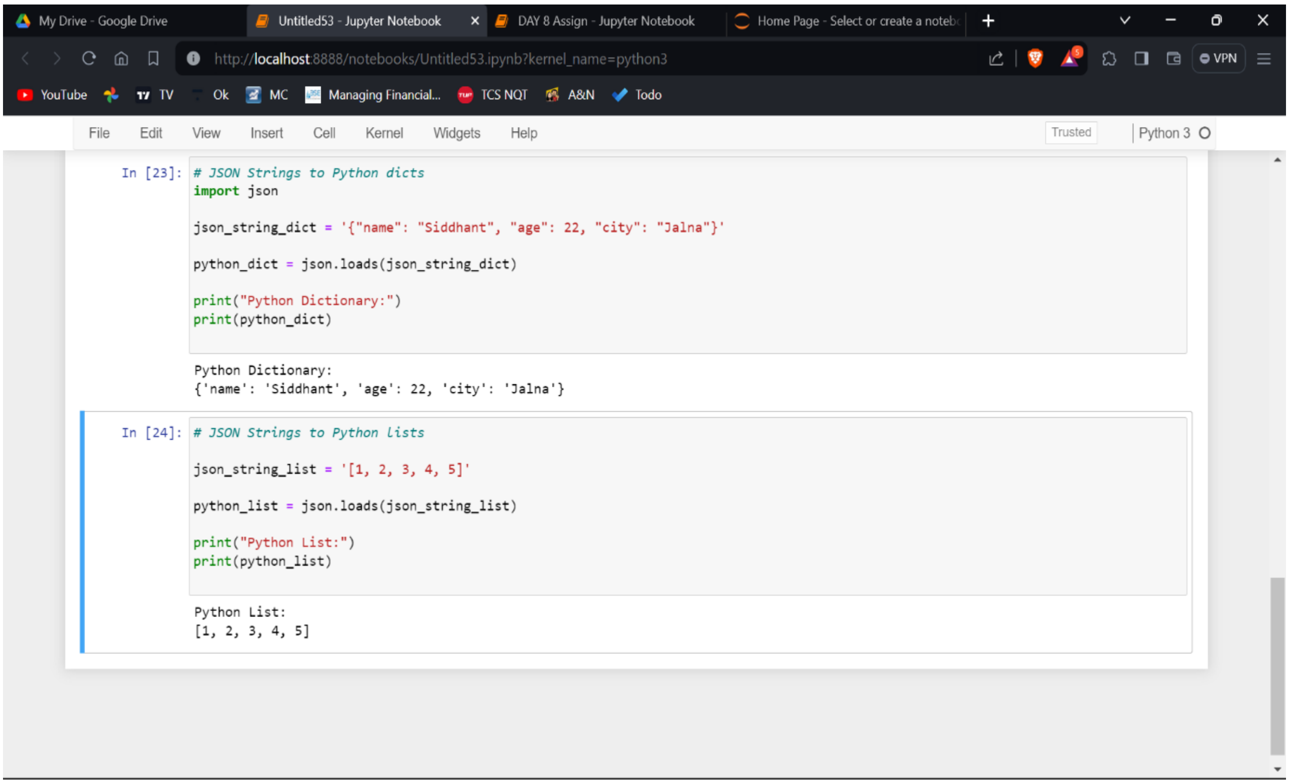
```
25
25
```

In this example, we have a regular function square that squares a given number. The equivalent lambda function square_lambda accomplishes the same task. The lambda function is particularly useful for short, one-line operations.

-In Python, you can convert **JSON (JavaScript Object Notation) strings to Python dictionaries or lists** using the json module.

- JSON is a lightweight data interchange format that is easy for humans to read and write and easy for machines to parse and generate.

- If you have a JSON string representing a dictionary and list, you can use the json.loads (load string) function to convert it into a Python dictionary or lists.


My Drive - Google Drive
Untitled53 - Jupyter Notebook
DAY 8 Assign - Jupyter Notebook
Home Page - Select or create a noteb
http://localhost:8888/notebooks/Untitled53.ipynb?kernel_name=python3
YouTube
TV
Ok
MC
Managing Financial...
TCS NQT
A&N
Todo
File
Edit
View
Insert
Cell
Kernel
Widgets
Help
Trusted
Python 3

In [23]:
```
# JSON Strings to Python dicts
import json

json_string_dict = '{"name": "Siddhant", "age": 22, "city": "Jalna"}'

python_dict = json.loads(json_string_dict)

print("Python Dictionary:")
print(python_dict)
```

```
Python Dictionary:
{'name': 'Siddhant', 'age': 22, 'city': 'Jalna'}
```

In [24]:
```
# JSON Strings to Python Lists

json_string_list = '[1, 2, 3, 4, 5]'

python_list = json.loads(json_string_list)

print("Python List:")
print(python_list)
```

```
Python List:
[1, 2, 3, 4, 5]
```