

WRITE A PYTHON PROGRAM TO DEMONSTRATE THE PANDAS LIBRARY AND its FUNCTIONS.

Pandas First step

Install and import

In [35]:

```
1 #!pip install pandas
2 #!pip install numpy --upgrade
3 !pip install numpy==1.16.5
```

Collecting numpy==1.16.5

Downloading https://files.pythonhosted.org/packages/f4/f6/aa112f76ada64787f677278218738bb895e9642118b1e8db68c7edd66ec2/numpy-1.16.5-cp37-cp37m-win_amd64.whl (https://files.pythonhosted.org/packages/f4/f6/aa112f76ada64787f677278218738bb895e9642118b1e8db68c7edd66ec2/numpy-1.16.5-cp37-cp37m-win_amd64.whl) (11.9MB)

Installing collected packages: numpy

Found existing installation: numpy 1.16.4

Uninstalling numpy-1.16.4:

ERROR: Could not install packages due to an EnvironmentError: [WinError 5] Access is denied: 'c:\\programdata\\anaconda3\\lib\\site-packages\\numpy\\compat\\py3k.py'

Consider using the `--user` option or check the permissions.

In [3]:

```
1 import pandas as pd
2 import numpy as np
3 print(np.__version__)
```

1.16.4

Creating DataFrames from Scratch

In [5]:

```
1 #Creating dataframes from dictionary
2 #Create Dictionary
3
4 data = {
5     'apples': [3,2,0,1],
6     'oranges': [0,3,7,2],
7 }
8 print(data)
```

{'apples': [3, 2, 0, 1], 'oranges': [0, 3, 7, 2]}

In [9]:

```

1 # Lets create DataFrames
2 purchase = pd.DataFrame(data)
3 print(purchase)
4
5 ##we call columns as attributs or features in pandas
6 ##0,1,2,3 are the indexes

```

	apples	oranges
0	3	0
1	2	3
2	0	7
3	1	2

In [11]:

```

1 purchase = pd.DataFrame(data, index = ['Aaditya','Deep','Dipak','Waidehee'])
2 print(purchase)

```

	apples	oranges
Aaditya	3	0
Deep	2	3
Dipak	0	7
Waidehee	1	2

In [12]:

```

1 #I want to find how many apples and oranges
2 # Purchased by dipak
3
4 print("Purchase of dipak \n",purchase.loc['Dipak'])

```

```

Purchase of dipak
apples      0
oranges     7
Name: Dipak, dtype: int64

```

How to read data from file

Reading the data from csv's(comma seperated values)

In [25]:

```
1 df = pd.read_csv('purchases.csv', index_col = 0)
2 df
```

Out[25]:

	apples	oranges
June	3	0
Robert	2	3
Lily	0	7
David	1	2

Reading data from JSON

In [18]:

```
1 df = pd.read_json('purchases.json')
2 df
```

Out[18]:

	apples	oranges
David	1	2
June	3	0
Lily	0	7
Robert	2	3

Reading data from JSON

In [19]:

```
1 !pip install pysqlite3
```

Collecting pysqlite3

Downloading <https://files.pythonhosted.org/packages/15/a1/396fd493b2c0cce6b61bed2ead4f1d0e254e3dbea9ea723e2c042900499f/pysqlite3-0.4.7.tar.gz> (<http://files.pythonhosted.org/packages/15/a1/396fd493b2c0cce6b61bed2ead4f1d0e254e3dbea9ea723e2c042900499f/pysqlite3-0.4.7.tar.gz>) (40kB)

Building wheels for collected packages: pysqlite3

Building wheel for pysqlite3 (setup.py): started

Building wheel for pysqlite3 (setup.py): finished with status 'done'

Running setup.py clean for pysqlite3

Failed to build pysqlite3

Installing collected packages: pysqlite3

Running setup.py install for pysqlite3: started

Running setup.py install for pysqlite3: finished with status 'done'

Successfully installed pysqlite3

WARNING: Legacy build of wheel for 'pysqlite3' created no files.

Command arguments: 'C:\ProgramData\Anaconda3\python.exe' -u -c 'import setuptools, tokenize;__file__='\"C:\\Users\\student\\AppData\\Local\\Temp\\pip-install-46ox6ux1\\pysqlite3\\setup.py\"';f=getattr(tokenize, '\"open\"', open)(__file__);code=f.read().replace('\"\\r\\n\"', '\"\\n\"');f.close();exec(compile(code, __file__, '\"exec\"'))' bdist_wheel -d 'C:\\Users\\student\\AppData\\Local\\Temp\\pip-wheel-xmdbjw3p' --python-tag cp37

Command output: [use --verbose to show]

In [20]:

```
1 # step 1 : Import the sql package
2
3 import sqlite3
4
5 # step 2 : Create connection object
6
7 con = sqlite3.connect('database.db')
```

In [22]:

```
1 # step 3 : Read the data by querying the database
2
3 df = pd.read_sql_query("Select * FROM purchases", con)
4
5 df
```

Out[22]:

	index	apples	oranges
0	June	3	0
1	Robert	2	3
2	Lily	0	7
3	David	1	2

In [21]:

```
1 con
```

Out[21]:

```
<sqlite3.Connection at 0x12e8102f810>
```

In [23]:

```
1 #step 4 : set the index for dataframes
2
3 df = df.set_index('index')
4 df
```

Out[23]:

	apples	oranges
index		
June	3	0
Robert	2	3
Lily	0	7
David	1	2

Converting back to a CSV,JSON,OR SQL

In [24]:

```
1 df.to_csv('new_purchases.csv')
2 df.to_json('new_purchases.csv')
3 df.to_sql('new_purchases.csv', con)
```

MOST important DataFrame operations

In [27]:

```
1 movies_df = pd.read_csv("IMDB-Movie-Data.csv", index_col="Title")
```

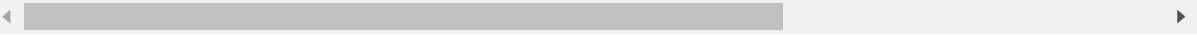
Viewing Your Data

In [32]:

```
1 #whole data
2 #movies_df
3
4 #First five rows of data
5 movies_df.head(5)
```

Out[32]:

Rank		Genre	Description	Director	Actors	Year
Title						
Guardians of the Galaxy	1	Action,Adventure,Sci-Fi	A group of intergalactic criminals are forced ...	James Gunn	Chris Pratt, Vin Diesel, Bradley Cooper, Zoe S...	2014
Prometheus	2	Adventure,Mystery,Sci-Fi	Following clues to the origin of mankind, a te...	Ridley Scott	Noomi Rapace, Logan Marshall-Green, Michael Fa...	2012
Split	3	Horror,Thriller	Three girls are kidnapped by a man with a diag...	M. Night Shyamalan	James McAvoy, Anya Taylor-Joy, Haley Lu Richar...	2016
Sing	4	Animation,Comedy,Family	In a city of humanoid animals, a hustling thea...	Christophe Lourdelet	Matthew McConaughey,Reese Witherspoon, Seth Ma...	2016
Suicide Squad	5	Action,Adventure,Fantasy	A secret government agency recruits some of th...	David Ayer	Will Smith, Jared Leto, Margot Robbie, Viola D...	2016

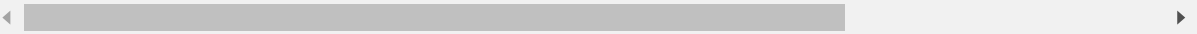


In [33]:

```
1 #Last five rows of DataFrames
2 movies_df.tail(5)
```

Out[33]:

	Rank	Genre	Description	Director	Actors	Year	Runtime (Minutes)
Title							
Secret in Their Eyes	996	Crime,Drama,Mystery	A tight-knit team of rising investigators, alo...	Billy Ray	Chiwetel Ejiofor, Nicole Kidman, Julia Roberts...	2015	111
Hostel: Part II	997	Horror	Three American college students studying abroa...	Eli Roth	Lauren German, Heather Matarazzo, Bijou Philli...	2007	94
Step Up 2: The Streets	998	Drama,Music,Romance	Romantic sparks occur between two dance studen...	Jon M. Chu	Robert Hoffman, Briana Evigan, Cassie Ventura,...	2008	98
Search Party	999	Adventure,Comedy	A pair of friends embark on a mission to reuni...	Scot Armstrong	Adam Pally, T.J. Miller, Thomas Middleditch,Sh...	2014	93
Nine Lives	1000	Comedy,Family,Fantasy	A stuffy businessman finds himself trapped ins...	Barry Sonnenfeld	Kevin Spacey, Jennifer Garner, Robbie Amell,Ch...	2016	87



Getting info about your dataset or dataframe

In [34]:

```
1 movies_df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Index: 1000 entries, Guardians of the Galaxy to Nine Lives
Data columns (total 11 columns):
Rank                1000 non-null int64
Genre               1000 non-null object
Description          1000 non-null object
Director            1000 non-null object
Actors              1000 non-null object
Year               1000 non-null int64
Runtime (Minutes)   1000 non-null int64
Rating              1000 non-null float64
Votes              1000 non-null int64
Revenue (Millions)  872 non-null float64
Metascore           936 non-null float64
dtypes: float64(3), int64(4), object(4)
memory usage: 93.8+ KB
```

In [36]:

```
1 # To find the shape of dataset
2
3 movies_df.shape
```

Out[36]:

```
(1000, 11)
```

Handling duplicate data

In [37]:

```
1 temp_df = movies_df.append(movies_df)
2 temp_df.shape
```

Out[37]:

```
(2000, 11)
```

In [38]:

```
1 temp_df.drop_duplicates(inplace=True)
```

In [39]:

```
1 temp_df.shape
```

Out[39]:

```
(1000, 11)
```

In [40]:

```
1 temp_df.drop_duplicates(inplace=True, keep=False)
```


In [41]:

```
1 temp_df.shape
```

Out[41]:

```
(1000, 11)
```

In []:

```
1
```