

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

Pandas First step

Install and import

In [1]:

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

In [2]:

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

1.16.4

Creating DataFrames from Scratch

In [3]:

```
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 [4]:

```
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 [5]:

```
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 [6]:

```
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 [7]:

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

Out[7]:

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

Reading data from JSON

In [8]:

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

Out[8]:

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

Reading data from JSON

In [9]:

```
1 !pip install pysqlite3
```

Collecting pysqlite3

Using cached <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>)

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-iq0olsvh\\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-3btr9z4j' --python-tag cp37

Command output: [use --verbose to show]

In [10]:

```
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 [11]:

```

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

```

Out[11]:

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

In [12]:

```

1 con

```

Out[12]:

```
<sqlite3.Connection at 0x22636d6dd50>
```

In [13]:

```

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

```

Out[13]:

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

Converting back to a CSV,JSON,OR SQL

In [14]:

```

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 [15]:

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

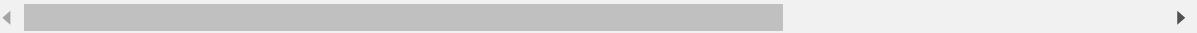
Viewing Your Data

In [16]:

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

Out[16]:

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 [17]:

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

Out[17]:

	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 [18]:

```
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 [19]:

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

Out[19]:

```
(1000, 11)
```

Handling duplicate data

In [20]:

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

Out[20]:

```
(2000, 11)
```

In [21]:

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

In [22]:

```
1 temp_df.shape
```

Out[22]:

```
(1000, 11)
```

In [23]:

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

In [24]:

```
1 temp_df.shape
```

Out[24]:

```
(1000, 11)
```

Column cleanup

In [25]:

```
1 #Step 1: to get the names of columns
2 movies_df.columns
```

Out[25]:

```
Index(['Rank', 'Genre', 'Description', 'Director', 'Actors', 'Year',
       'Runtime (Minutes)', 'Rating', 'Votes', 'Revenue (Millions)',
       'Metascore'],
      dtype='object')
```

In [27]:

```
1 #step 2: find which columns are not in format
2 #step 3: rename it
3 movies_df.rename(columns={
4     'Runtime (Minutes)': 'Runtime', 'Revenue (Millions)' : 'Revenue_Millions'
5 }, inplace=True)
6 movies_df.columns
```

Out[27]:

```
Index(['Rank', 'Genre', 'Description', 'Director', 'Actors', 'Year', 'Runtime',
       'Rating', 'Votes', 'Revenue_Millions', 'Metascore'],
      dtype='object')
```

In [28]:

```
1 #step 4: Make each column name in lower case
2 movies_df.columns = [col.lower() for col in movies_df]
3 movies_df.columns
```

Out[28]:

```
Index(['rank', 'genre', 'description', 'director', 'actors', 'year', 'runtime',
       'rating', 'votes', 'revenue_millions', 'metascore'],
      dtype='object')
```

How to work with missing values

In [30]:

```

1 #Is there any null value in dataframe
2
3 movies_df.isnull()

```

Out[30]:

	rank	genre	description	director	actors	year	runtime	rating	votes	revenu
Title										
Guardians of the Galaxy	False	False	False	False	False	False	False	False	False	
Prometheus	False	False	False	False	False	False	False	False	False	
Split	False	False	False	False	False	False	False	False	False	
Sing	False	False	False	False	False	False	False	False	False	
Suicide Squad	False	False	False	False	False	False	False	False	False	
The Great Wall	False	False	False	False	False	False	False	False	False	
La La Land	False	False	False	False	False	False	False	False	False	
Mindhorn	False	False	False	False	False	False	False	False	False	
The Lost City of Z	False	False	False	False	False	False	False	False	False	
Passengers	False	False	False	False	False	False	False	False	False	
Fantastic Beasts and Where to Find Them	False	False	False	False	False	False	False	False	False	
Hidden Figures	False	False	False	False	False	False	False	False	False	
Rogue One	False	False	False	False	False	False	False	False	False	
Moana	False	False	False	False	False	False	False	False	False	
Colossal	False	False	False	False	False	False	False	False	False	
The Secret Life of Pets	False	False	False	False	False	False	False	False	False	
Hacksaw Ridge	False	False	False	False	False	False	False	False	False	
Jason Bourne	False	False	False	False	False	False	False	False	False	
Lion	False	False	False	False	False	False	False	False	False	
Arrival	False	False	False	False	False	False	False	False	False	
Gold	False	False	False	False	False	False	False	False	False	
Manchester by the Sea	False	False	False	False	False	False	False	False	False	
Hounds of Love	False	False	False	False	False	False	False	False	False	
Trolls	False	False	False	False	False	False	False	False	False	
Independence Day: Resurgence	False	False	False	False	False	False	False	False	False	

	rank	genre	description	director	actors	year	runtime	rating	votes	revenu
Title										
Paris pieds nus	False	False	False	False	False	False	False	False	False	False
Bahubali: The Beginning	False	False	False	False	False	False	False	False	False	False
Dead Awake	False	False	False	False	False	False	False	False	False	False
Bad Moms	False	False	False	False	False	False	False	False	False	False
Assassin's Creed	False	False	False	False	False	False	False	False	False	False
...
Texas Chainsaw 3D	False	False	False	False	False	False	False	False	False	False
Disturbia	False	False	False	False	False	False	False	False	False	False
Rock of Ages	False	False	False	False	False	False	False	False	False	False
Scream 4	False	False	False	False	False	False	False	False	False	False
Queen of Katwe	False	False	False	False	False	False	False	False	False	False
My Big Fat Greek Wedding 2	False	False	False	False	False	False	False	False	False	False
Dark Places	False	False	False	False	False	False	False	False	False	False
Amateur Night	False	False	False	False	False	False	False	False	False	False
It's Only the End of the World	False	False	False	False	False	False	False	False	False	False
The Skin I Live In	False	False	False	False	False	False	False	False	False	False
Miracles from Heaven	False	False	False	False	False	False	False	False	False	False
Annie	False	False	False	False	False	False	False	False	False	False
Across the Universe	False	False	False	False	False	False	False	False	False	False
Let's Be Cops	False	False	False	False	False	False	False	False	False	False
Max	False	False	False	False	False	False	False	False	False	False
Your Highness	False	False	False	False	False	False	False	False	False	False
Final Destination 5	False	False	False	False	False	False	False	False	False	False
Endless Love	False	False	False	False	False	False	False	False	False	False
Martyrs	False	False	False	False	False	False	False	False	False	False
Selma	False	False	False	False	False	False	False	False	False	False
Underworld: Rise of the Lycans	False	False	False	False	False	False	False	False	False	False
Taare Zameen Par	False	False	False	False	False	False	False	False	False	False

	rank	genre	description	director	actors	year	runtime	rating	votes	revenu
Title										
Take Me Home Tonight	False	False	False	False	False	False	False	False	False	
Resident Evil: Afterlife	False	False	False	False	False	False	False	False	False	
Project X	False	False	False	False	False	False	False	False	False	
Secret in Their Eyes	False	False	False	False	False	False	False	False	False	
Hostel: Part II	False	False	False	False	False	False	False	False	False	
Step Up 2: The Streets	False	False	False	False	False	False	False	False	False	
Search Party	False	False	False	False	False	False	False	False	False	
Nine Lives	False	False	False	False	False	False	False	False	False	

1000 rows × 11 columns



In [31]:

```
1 # How many null values are there and which column
2 movies_df.isnull().sum()
```

Out[31]:

```
rank                0
genre               0
description         0
director            0
actors              0
year                0
runtime             0
rating              0
votes              0
revenue_millions   128
metascore           64
dtype: int64
```

Removing Null values

Method 1

In [33]:

```
1 # Drop the rows which contain null values
2 movies_df.dropna()
```

The Other Boleyn Girl	963	Biography,Drama,History	Two sisters contend for the affection of King ...	Justin Chadwick	Natalie Portman, Scarlett Johansson, Eric Bana...	2008	115
I Spit on Your Grave	964	Crime,Horror,Thriller	A writer who is brutalized during her cabin re...	Steven R. Monroe	Sarah Butler, Jeff Branson, Andrew Howard,Dani...	2010	100
Texas Chainsaw 3D	971	Horror,Thriller	A young woman travels to Texas to collect an i...	John Luessenhop	Alexandra Daddario, Tania Raymonde, Scott East...	2013	90
Rock of Ages	973	Comedy,Drama,Musical	A small town girl and a city boy meet on the S...	Adam Shankman	Julianne Hough, Diego Boneta, Tom Cruise, Alec...	2012	120
Scream 4	974	Horror,Mystery	Ten years have passed, and Sidney	Wes Craven	Neve Campbell, Courteney Cox,	2011	110

In [35]:

```
1 # Drop the rows which contain null values
2 movies_df.dropna(axis = 1)
```

Out[35]:

	rank	genre	description	director	actors	year	runtime
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	120
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	120
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	110

Imputation : Replacing null values by Non-null values

Perform for revenue_millions

In [39]:

```
1 # Columns with na
2 revenue = movies_df['revenue_millions']
```

In [40]:

```
1 revenue.head()
```

Out[40]:

```
Title
Guardians of the Galaxy    333.13
Prometheus                 126.46
Split                     138.12
Sing                      270.32
Suicide Squad             325.02
Name: revenue_millions, dtype: float64
```

In [41]:

```
1 # mean() finds average
2 revenue_mean = revenue.mean()
3 revenue_mean
```

Out[41]:

```
82.95637614678897
```

In [42]:

```
1 # fill na with mean()
2 revenue.fillna(revenue_mean, inplace=True)
```

In [43]:

```
1 movies_df.isnull().sum()
```

Out[43]:

```
rank          0
genre         0
description   0
director      0
actors        0
year          0
runtime       0
rating        0
votes         0
revenue_millions  0
metascore     64
dtype: int64
```

Perform for metascore

In [45]:

```
1 meta = movies_df['metascore']
```

In [46]:

```
1 meta.head()
```

Out[46]:

```
Title
Guardians of the Galaxy    76.0
Prometheus                 65.0
Split                     62.0
Sing                      59.0
Suicide Squad              40.0
Name: metascore, dtype: float64
```

In [47]:

```
1 meta_mean = meta.mean()
2 meta_mean
```

Out[47]:

```
58.98504273504273
```

In [48]:

```
1 # Fill na with mean()
2 meta.fillna(meta_mean, inplace=True)
```

In [49]:

```
1 movies_df.isnull().sum()
```

Out[49]:

```
rank          0
genre         0
description   0
director      0
actors        0
year          0
runtime       0
rating        0
votes         0
revenue_millions  0
metascore     0
dtype: int64
```

In []:

```
1
```