IMDb Movie Assignment

You have the data for the 100 top-rated movies from the past decade along with various pieces of information about the movie, its actors, and the voters who have rated these movies online. In this assignment, you will try to find some interesting insights into these movies and their voters, using Python.

Task 1: Reading the data

• ### Subtask 1.1: Read the Movies Data.

Read the movies data file provided and store it in a dataframe movies .

```
In [1]:
            import pandas as pd
           df = pd.read_csv('IMDB_dataset.csv')
           df
                     Title
                           title_year
                                          budget
                                                              actor_1_name actor_2_name
                                                                                             actor_3_name
                                                                                                            actor_1_facebook_likes
                                                                                                                                     actor_2_facebook_likes
Out[1]:
                                                      Gross
            0
              La La Land
                                       30000000
                                                  151101803
                                                                                                                              14000
                                                                                                                                                     19000.0
                                2016
                                                                Ryan Gosling
                                                                                Emma Stone
                                                                                                Amiée Conn
                                                                     Ginnifer
                                                                                                                               2800
                                                                                                                                                     28000.0
            1
                  Zootopia
                                2016
                                      150000000
                                                  341268248
                                                                                                  Idris Elba
                                                                    Goodwin
                                                                                   Bateman
                                                                              Nicole Kidman
            2
                                2016
                                        12000000
                                                   51738905
                                                                   Dev Patel
                                                                                               Rooney Mara
                                                                                                                              33000
                                                                                                                                                     96000.0
                                                                                     Jeremy
                                                                                                     Forest
            3
                    Arrival
                                2016
                                       47000000
                                                  100546139
                                                                 Amy Adams
                                                                                                                              35000
                                                                                                                                                      5300.0
                                                                                     Renner
                                                                                                   Whitaker
               Manchester
                                                                                    Michelle
                                2016
                                         9000000
                                                   47695371
                                                                Casey Affleck
                                                                                              Kyle Chandler
                                                                                                                                518
                                                                                                                                                     71000.0
                by the Sea
                                                                                    Williams
                                                                                    Melissa
                 Whiplash
           95
                                2014
                                         3300000
                                                   13092000
                                                               J.K. Simmons
                                                                                               Chris Mulkey
                                                                                                                              24000
                                                                                                                                                       970 0
                                                                                     Benoist
                                                                     Seamus
                   Before
                                                                                              Athina Rachel
           96
                                2013
                                         3000000
                                                    8114507
                                                                               Ariane Labed
                                                                                                                                140
                                                                                                                                                        63.0
                                                                     Davev-
                  Midnight
                                                                                                   Tsangari
                                                                   Fitzpatrick
                Star Wars:
               Episode VII
                                2015 245000000 936662225
                                                                Doug Walker
                                                                                 Rob Walker
                                                                                                         0
                                                                                                                                131
                                                                                                                                                        12 0
                .
The Force
                 Awakens
                    Harry
                Potter and
               the Deathly
                                2010 150000000 296347721
                                                                 Rupert Grint
                                                                                 Toby Jones
                                                                                               Alfred Enoch
                                                                                                                              10000
                                                                                                                                                      2000.0
                  Hallows
               Tucker and
                                                                     Katrina
                                                                                                    Chelan
                                2010
                                         5000000
                                                      223838
                                                                                Tyler Labine
                                                                                                                                948
                                                                                                                                                       779.0
               Dale vs Evil
                                                                     Bowden
                                                                                                   Simmons
          100 rows × 62 columns
```

• ### Subtask 1.2: Inspect the Dataframe

float64

Votes1000

Inspect the dataframe for dimensions, null-values, and summary of different numeric columns.

```
In [2]:
         # Check the number of rows and columns in the data
         df.shape
Out[2]: (100, 62)
In [3]:
         #check data types of each column
         df.dtypes
Out[3]: Title
                            object
         title_year
                             int64
                             int64
        budget
        Gross
                             int64
        actor 1 name
                            object
```

```
VotesUS
                   float64
                  float64
VotesnUS
content_rating object
Country object
Length: 62, dtype: object
```

In [4]: df.info()

<class 'pandas.core.frame.DataFrame'> RangeIndex: 100 entries, 0 to 99 Data columns (total 62 columns):

	columns (total 62 column		
#	Column	Non-Null Count	Dtype
0	Title	100 non-null	object
1	title year	100 non-null	int64
2	budget	100 non-null	int64
3	Gross	100 non-null	int64
4	actor 1 name	100 non-null	object
5	actor_2_name	100 non-null	object
6	actor_3_name	100 non-null	object
7			int64
	actor_1_facebook_likes	100 non-null	
8	actor_2_facebook_likes	99 non-null	float64
9	actor_3_facebook_likes	98 non-null	float64
10	IMDb_rating	100 non-null	float64
11	genre_1	100 non-null	object
12	genre_2	97 non-null	object
13	genre_3	74 non-null	object
14	MetaCritic	95 non-null	float64
15	Runtime	100 non-null	int64
16	CVotes10	100 non-null	int64
17	CVotes09	100 non-null	int64
18	CVotes08	100 non-null	int64
19	CVotes07	100 non-null	int64
20	CVotes06	100 non-null	int64
21	CVotes05	100 non-null	int64
22	CVotes04	100 non-null	int64
23	CVotes03	100 non-null	int64
24			
	CVotes02	100 non-null	int64
25	CVotes01	100 non-null	int64
26	CVotesMale	100 non-null	int64
27	CVotesFemale	100 non-null	int64
28	CVotesU18	100 non-null	int64
29	CVotesU18M	100 non-null	int64
30	CVotesU18F	100 non-null	int64
31	CVotes1829	100 non-null	int64
32	CVotes1829M	100 non-null	int64
33	CVotes1829F	100 non-null	int64
34	CVotes3044	100 non-null	int64
35	CVotes3044M	100 non-null	int64
36	CVotes3044F	100 non-null	int64
37	CVotes45A	100 non-null	int64
38	CVotes45AM	100 non-null	int64
39	CVotes45AF	100 non-null	int64
40	CVotes1000	100 non-null	int64
41	CVotesUS	100 non-null	int64
42	CVotesnUS	100 non-null	int64
43		100 non-null	
44	VotesM		float64
	VotesF	100 non-null	float64
45	VotesU18	100 non-null	float64
46	VotesU18M	100 non-null	float64
47	VotesU18F	100 non-null	float64
48	Votes1829	100 non-null	float64
49	Votes1829M	100 non-null	float64
50	Votes1829F	100 non-null	float64
51	Votes3044	100 non-null	float64
52	Votes3044M	100 non-null	float64
53	Votes3044F	100 non-null	float64
54	Votes45A	100 non-null	float64
55	Votes45AM	100 non-null	float64
56	Votes45AF	100 non-null	float64
57	Votes1000	100 non-null	float64
58	VotesUS	100 non-null	float64
59	VotesnUS	100 non-null	float64
60			
	content_rating	100 non-null	object
61	Country	100 non-null	object
	es: float64(21), int64(33	2), Object(9)	

memory usage: 48.6+ KB

```
#check for null values
         df.isnull().sum()
Out[5]: Title
                           0
                           0
        title year
        budget
                           0
        Gross
        actor_1_name
                           0
        Votes1000
                           0
        VotesUS
                           0
                           0
        VotesnUS
                           0
        content rating
        Country
                           0
        Length: 62, dtype: int64
In [6]:
         # Check the column-wise info of the dataframe
         df.info()
        <class 'pandas.core.frame.DataFrame'>
        RangeIndex: 100 entries, 0 to 99
        Data columns (total 62 columns):
         # Column
                                      Non-Null Count Dtype
        - - -
         0
             Title
                                      100 non-null
                                                       object
                                      100 non-null
                                                       int64
             title_year
         1
         2
                                     100 non-null
                                                      int64
             budget
                                      100 non-null
         3
             Gross
                                                       int64
         4
             actor_1_name
                                      100 non-null
                                                       object
         5
             actor_2_name
                                      100 non-null
                                                       obiect
         6
            actor_3_name
                                      100 non-null
                                                       object
             actor_1_facebook_likes 100 non-null
actor_2_facebook_likes 99 non-null
         7
                                                       int64
         8
                                                       float64
             actor_3_facebook_likes 98 non-null
         9
                                                       float64
         10 IMDb_rating
                                      100 non-null
                                                       float64
         11 genre 1
                                      100 non-null
                                                       object
```

12

13

14

15

16

17

18

19

20

23

24

27

31

32

33

35

36

37

39

40

41

43

44

46 47

51

52

53 54 genre 2

genre 3

Runtime

CVotes10

CVotes09

CVotes08

CVotes07

CVotes06

CVotes03

CVotes02

26 CVotesMale

28 CVotesU18

29 CVotesU18M

30 CVotesU18F

34 CVotes3044

38 CVotes45AM

CVotes1829

CVotes1829M

CVotes1829F

CVotes3044M

CVotes3044F

CVotes45A

CVotes45AF

CVotes1000

CVotesUS

VotesU18M

VotesU18F

Votes3044

Votes3044M

Votes3044F

Votes45A

55 Votes45AM

42 CVotesnUS

VotesM

VotesE

45 VotesU18

48 Votes1829

49 Votes1829M

50 Votes1829F

CVotesFemale

21 CVotes05

22 CVotes04

25 CVotes01

MetaCritic

97 non-null

74 non-null

95 non-null

100 non-null 100 non-null

100 non-null

100 non-null

100 non-null 100 non-null

100 non-null

100 non-null

100 non-null

100 non-null

100 non-null

100 non-null

100 non-null

100 non-null

100 non-null

100 non-null

100 non-null

100 non-null

100 non-null

100 non-null

100 non-null

100 non-null

100 non-null

100 non-null

100 non-null

100 non-null

100 non-null

100 non-null

100 non-null

100 non-null

100 non-null

100 non-null

100 non-null

object

object

int64

float64

```
Votes45AF
                             100 non-null
                                              float64
 57
    Votes1000
                             100 non-null
                                              float64
 58
    VotesUS
                             100 non-null
                                              float64
 59
    VotesnUS
                             100 non-null
                                              float64
 60 content_rating
                             100 non-null
                                              object
 61 Country
                             100 non-null
                                              object
dtypes: float64(21), int64(32), object(9)
memory usage: 48.6+ KB
```

In [7]: # Check the summary for the numeric columns
 df.describe()

title_year budget Gross actor_1_facebook_likes actor_2_facebook_likes actor_3_facebook_likes IMDb_rating MetaCriti count 100.000000 1.000000e+02 1.000000e+02 100.000000 99.000000 98.000000 100.000000 95.00000 mean 2012.820000 7.838400e+07 1.468679e+08 13407.270000 7377.303030 3002.153061 7.883000 78.25263 1.919491 7.445295e+07 1.454004e+08 10649.037862 13471.568216 6940.301133 0.247433 9.12206 std min 2010.000000 3.000000e+06 2.238380e+05 39.000000 12.000000 0.000000 7.500000 62.00000 2011.000000 1.575000e+07 4.199752e+07 1000.000000 580.000000 319.750000 7.700000 72.00000 2013.000000 4.225000e+07 1.070266e+08 13000.000000 1000.000000 626.500000 7.800000 78.00000 20000.000000 11000.000000 2014.000000 1.500000e+08 2.107548e+08 1000.000000 8.100000 83.50000 2016.000000 2.600000e+08 9.366622e+08 35000.000000 96000.000000 46000.000000 8.800000 100.00000

8 rows × 53 columns

In [8]:

Check the summary for the numeric and caretorical columns

Task 2: Data Analysis

Now that we have loaded the dataset and inspected it, we see that most of the data is in place. As of now, no data cleaning is required, so let's start with some data manipulation, analysis, and visualisation to get various insights about the data.

• ### Subtask 2.1: Reduce those Digits!

These numbers in the budget and gross are too big, compromising its readability. Let's convert the unit of the budget and gross columns from \$ to million \$ first.

```
In [5]:
          df.Gross
               151101803
Out[5]: 0
               341268248
         2
                51738905
         3
               100546139
                47695371
         95
                13092000
         96
                 8114507
         97
               936662225
         98
               296347721
                  223838
        Name: Gross, Length: 100, dtype: int64
```

```
In [6]:
# Divide the 'gross' and 'budget' columns by 1000000 to convert '$' to 'million $'
df['Gross_m_$']=df.Gross/1000000
```

In [7]: df

Title title_year budget Gross actor_1_name actor_2_name actor_3_name actor_1_facebook_likes actor_2_facebook_likes 0 La La Land 19000 0 2016 30000000 151101803 Ryan Gosling Emma Stone Amiée Conn 14000 Ginnifer 2016 150000000 341268248 Idris Elba 2800 28000.0 Zootopia Bateman Goodwin 2 2016 12000000 51738905 Dev Patel Nicole Kidman 33000 96000.0 Lion Rooney Mara

	3	Arrival	2016	47000000	100546139	Amy Adams	Jeremy Renne			5300.0
	4	Manchester by the Sea		9000000) 47695371	Casey Affleck	Michelle Williams	Kyle Chandle	- 518	71000.0
	95	Whiplash	2014	3300000	13092000		Benois		24000	970.0
	96	Before Midnight		3000000	8114507	Seamus Davey- Fitzpatrick	- Ariane Labec	Athina Rache Tsangar		63.0
	97	Star Wars: Episode VII - The Force Awakens	2015	5 245000000	936662225	Doug Walker	r Rob Walke	r (131	12.0
	98	Harry Potter and the Deathly Hallows: Part I	2010 2010	150000000) 296347721	Rupert Grin	t Toby Jones	Alfred Enoch	10000	2000.0
	99	Tucker and Dale vs Evil		5000000	223838	Katrina Bowder		Chelar Simmons		779.0
	100	rows × 63	columns							
	4									>
In [10]:	#				ing in yea o_rating.m					
					_ 3					
Out[10]:	8.	2								
In [11]:	#				udget till []['Title					
Out[11]:		Tangle me: Title		object						
In [12]:	#			pent in 20 2013].budg	013 get.mean()					
Out[12]:	60	588235.294	4117644							
In [11]:	#	How many df.actor_3	Action m B_name=='	novies Sca 'Scarlett	arlett Joh Johansson	ansson did? ']) (df.ac	tor_3_name==	'Scarlett Jo	hansson']) (df.acto	or_3_name=='Scarlett 3
Out[11]:		Title	title_year	budget	Gross	actor_1_name	actor_2_name	actor_3_name	actor_1_facebook_likes	actor_2_facebook_likes act
	11	The Avengers	2012	220000000	623279547	Chris Hemsworth	Robert Downey Jr.	Scarlett Johansson	26000	21000.0
	_	ws × 63 col	lumns							
	4									þ.
In [14]:	#	Plot top f.head(10)		es						
Out[14]:		Title	title_year	budget	Gross	actor_1_name	actor_2_name	actor_3_name	actor_1_facebook_likes	actor_2_facebook_likes ac
	0	La La Land	2016	30000000	151101803	Ryan Gosling	Emma Stone	Amiée Conn	14000	19000.0
	1	Zootopia	2016	150000000	341268248	Ginnifer Goodwin	Jason Bateman	Idris Elba	2800	28000.0
	2	Lion	2016	12000000	51738905	Dev Patel	Nicole Kidman	Rooney Mara	33000	96000.0
	3	Arrival	2016	47000000	100546139	Amy Adams	Jeremy Renner	Forest Whitaker	35000	5300.0

Michelle Williams

Jeff Bridges

Kyle Chandler

Ben Foster

518

71000.0

12000.0

Casey Affleck

Chris Pine

Manchester by the Sea

5 Hell or High Water

2016

2016

9000000

12000000

47695371

27007844

```
The Dark
                                                                                                                                                                                                  Joseph
                                                              2012 250000000 448130642
                                                                                                                                                                                                                                                  27000
                                                                                                                                                                                                                                                                                               23000.0
                        8
                                       Knight
                                                                                                                               Tom Hardy
                                                                                                                                                        Christian Bale
                                                                                                                                                                                       Gordon-Levitt
                                        Rises
                                     Captain
                                                                                                                                       Robert
                                                                                                                                                                   Scarlett
                                                             2016 250000000 407197282
                                                                                                                                                                                                                                                                                               19000 0
                                   America:
                                                                                                                                                                                         Chris Evans
                                                                                                                                                                                                                                                  21000
                                                                                                                               Downey Jr.
                                                                                                                                                              Johansson
                                   Civil War
                      10 rows × 62 columns
In [15]:
                         # List out movies with Sci-fi genre
                         df[(df.genre 1=="Sci-Fi")|(df.genre 2=="Sci-Fi")|(df.genre 3 =="Sci-Fi")]['Title']
Out[15]: 3
                                                                                                            Arrival
                                                             Captain America: Civil War
                       9
                        11
                                                                                               The Avengers
                                                            X-Men: Days of Future Past
                        14
                                                                    Star Trek Into Darkness
                        17
                                                                                     Edge of Tomorrow
                        19
                                                                    Guardians of the Galaxy
                                   Captain America: The Winter Soldier
                        20
                        26
                                                                                               Interstellar
                        27
                                                                                                       Inception
                        28
                                                                                 X-Men: First Class
                                                                                Mad Max: Fury Road
                       29
                        33
                                                                                                  The Martian
                       34
                                                                                                            Gravity
                        37
                                                  Rise of the Planet of the Apes
                       67
                                                                                                                     Her
                       80
                                                                                                     Ex Machina
                       Name: Title, dtype: object
In [16]:
                         # What is maximum rating of Action movies
                         df[(df.genre_1=="Action") |(df.genre_2=="Action") |(df.genre_3=="Action")].IMDb_rating.max()
Out[16]: 8.8
In [17]:
                         #List out X man series (means all movies under name - Xman)
                         df[(df.Title=='X-Men: Days of Future Past') | (df.Title=='X-Men: First Class')]
                                   Title title_year
                                                                                                  Gross \ \ actor\_1\_name \ \ actor\_2\_name \ \ actor\_3\_name \ \ actor\_1\_facebook\_likes \ \ actor\_2\_facebook\_likes \ \ actor\_3\_name \ \ actor\_3\_
Out[17]:
                                                                         budget
                                       X-
                                  Men:
                                  Days
                                                                                                                              Jennifer
                                                                                                                                                                                                                                                                                        22000.0
                                                       2014 200000000 233914986
                                                                                                                                               Peter Dinklage Hugh Jackman
                                                                                                                                                                                                                                           34000
                                       of
                                                                                                                           Lawrence
                                Future
                                   Past
                                      X-
                                                                                                                              Jennifer
                                                                                                                                                            Michael
                                  Men:
                                                       2011 160000000 146405371
                                                                                                                                                                                    Oliver Platt
                                                                                                                                                                                                                                            34000
                                                                                                                                                                                                                                                                                        13000.0
                                   First
                                                                                                                           Lawrence
                                                                                                                                                     Fassbender
                                 Class
                      2 rows × 62 columns
In [18]:
                         # List out actors and actress played role in movie 127 Hours
                         df[df["Title"]=="127 Hours"]
                                  Title title_year
Out[18]:
                                                                      budget
                                                                                             Gross
                                                                                                            actor_1_name actor_2_name actor_3_name actor_1_facebook_likes actor_2_facebook_likes actor_3_
                                   127
                       73 Hours
                                                                                                                                                                                                                                      11000
                                                                                                                                                                                                                                                                                       642.0
                                                      2010 18000000 18329466
                                                                                                              James Franco
                                                                                                                                           Treat Williams
                                                                                                                                                                              Kate Burton
                      1 rows × 62 columns
```

Benedict

Brad Garrett Donna Murphy

Cumberbatch

Chiwetel

Ejiofor

Rachel

McAdams

M.C. Gainey

19000

799

NaN

553.0

Doctor

Strange

Tangled

2016 165000000 232641920

2010 260000000 200807262

- 1. Create a new column called profit which contains the difference of the two columns: gross and budget .
- 2. Sort the dataframe using the profit column as reference.

Out[20]:

- 3. Extract the top ten profiting movies in descending order and store them in a new dataframe top10.
- 4. Plot a scatter between the columns budget and profit and write a few words on what you observed.
- 5. Extract the movies with a negative profit and store them in a new dataframe neg profit

In [19]:
Create the new column named 'profit' by subtracting the 'budget' column from the 'gross' column
df['profit']=df['budget']-df['Gross']
df

actor_1_name actor_2_name actor_3_name actor_1_facebook_likes actor_2_facebook_likes a Out[19]: Title title_year budget Gross 0 La La Land 2016 30000000 151101803 Ryan Gosling Emma Stone Amiée Conn 14000 19000.0 Ginnifer Jason 341268248 Zootopia 2016 150000000 Idris Elba 2800 28000.0 Goodwin Bateman 2 Lion 2016 12000000 51738905 Dev Patel Nicole Kidman Rooney Mara 33000 96000.0 Forest Jeremy 3 Arrival 2016 47000000 35000 5300.0 100546139 Amy Adams Whitaker Renner Manchester Michelle 4 2016 9000000 47695371 Casey Affleck Kyle Chandler 518 71000.0 by the Sea Williams Melissa Whiplash 2014 3300000 13092000 J.K. Simmons Chris Mulkey 24000 970.0 95 Benoist Seamus Before Athina Rachel 96 2013 3000000 8114507 Davey-Ariane Labed 140 63.0 Midnight Tsangari Fitzpatrick Star Wars: Episode VII 2015 245000000 936662225 Rob Walker O 131 12 0 Doug Walker The Force Awakens Harry Potter and 98 the Deathly 2010 150000000 296347721 Rupert Grint **Toby Jones** Alfred Enoch 10000 2000.0 Hallows: Part I Tucker and Katrina Chelan 779.0 99 2010 5000000 223838 948 Tyler Labine Dale vs Evil Bowden Simmons 100 rows × 63 columns

Sort the dataframe with the 'profit' column as reference using the 'sort_values' function. Make sure to set the
#'ascending' to 'False'
df.sort_values(["Title","profit"],axis=0,ascending=True,inplace=True)
df

	Title	title_year	budget	Gross	actor_1_name	actor_2_name	actor_3_name	actor_1_facebook_likes	actor_2_facebook_likes a
69	12 Years a Slave	2013	20000000	56667870	Quvenzhané Wallis	Scoot McNairy	Taran Killam	2000	660.0
73	127 Hours	2010	18000000	18329466	James Franco	Treat Williams	Kate Burton	11000	642.0
91	50/50	2011	8000000	34963967	Joseph Gordon-Levitt	Anna Kendrick	Bryce Dallas Howard	23000	10000.0
88	About Time	2013	12000000	15294553	Tom Hughes	Tom Hollander	Lindsay Duncan	565	555.0
89	Amour	2012	8900000	225377	Isabelle Huppert	Emmanuelle Riva	Jean-Louis Trintignant	678	432.0
95	Whiplash	2014	3300000	13092000	J.K. Simmons	Melissa Benoist	Chris Mulkey	24000	970.0
25	Wreck-It Ralph	2012	165000000	189412677	Jack McBrayer	Sarah Silverman	Joe Lo Truglio	975	931.0
14	X-Men: Days of Future Past	2014	200000000	233914986	Jennifer Lawrence	Peter Dinklage	Hugh Jackman	34000	22000.0
28	X-Men: First Class	2011	160000000	146405371	Jennifer Lawrence	Michael Fassbender	Oliver Platt	34000	13000.0
1	Zootopia	2016	150000000	341268248	Ginnifer Goodwin	Jason Bateman	Idris Elba	2800	28000.0

In [21]:

Get the top 10 profitable movies by using position based indexing. Specify the rows till 10 (0-9) df.head(10)

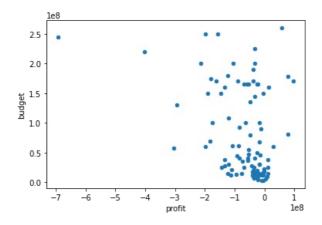
Title	title_year	budget	Gross	actor_1_name	actor_2_name	actor_3_name	actor_1_facebook_likes	actor_2_facebook_likes a
12 Years a Slave	2013	20000000	56667870	Quvenzhané Wallis	Scoot McNairy	Taran Killam	2000	660.0
127 Hours	2010	18000000	18329466	James Franco	Treat Williams	Kate Burton	11000	642.0
50/50	2011	8000000	34963967	Joseph Gordon-Levitt	Anna Kendrick	Bryce Dallas Howard	23000	10000.0
About Time	2013	12000000	15294553	Tom Hughes	Tom Hollander	Lindsay Duncan	565	555.0
Amour	2012	8900000	225377	Isabelle Huppert	Emmanuelle Riva	Jean-Louis Trintignant	678	432.0
Argo	2012	44500000	136019448	Clea DuVall	Scoot McNairy	Tate Donovan	1000	660.0
Arrival	2016	47000000	100546139	Amy Adams	Jeremy Renner	Forest Whitaker	35000	5300.0
Before Midnight	2013	3000000	8114507	Seamus Davey- Fitzpatrick	Ariane Labed	Athina Rachel Tsangari	140	63.0
Big Hero 6	2014	165000000	222487711	Damon Wayans Jr.	Daniel Henney	Abraham Benrubi	756	719.0
Birdman or (The Unexpected Virtue of Ignorance)	2014	18000000	42335698	Emma Stone	Naomi Watts	Merritt Wever	15000	6000.0
	12 Years a Slave 127 Hours 50/50 About Time Amour Argo Arrival Before Midnight Big Hero 6 Birdman or (The Unexpected Virtue of	12 Years a Slave 2013 127 Hours 2010 50/50 2011 About Time 2013 Amour 2012 Argo 2012 Arrival 2016 Before Midnight 2013 Big Hero 6 2014 Birdman or (The Unexpected Virtue of 2014	12 Years a Slave 2013 20000000 127 Hours 2010 18000000 50/50 2011 8000000 About Time 2013 12000000 Amour 2012 8900000 Argo 2012 44500000 Arrival 2016 47000000 Before Midnight 2013 3000000 Big Hero 6 2014 165000000 Birdman or (The Unexpected Virtue of 2014 18000000	12 Years a Slave 2013 20000000 56667870 127 Hours 2010 18000000 18329466 50/50 2011 8000000 34963967 About Time 2013 12000000 15294553 Amour 2012 8900000 225377 Argo 2012 44500000 136019448 Arrival 2016 47000000 100546139 Before Midnight 2013 3000000 8114507 Big Hero 6 2014 165000000 222487711 Birdman or (The Unexpected Virtue of 2014 18000000 42335698	12 Years a Slave 2013 20000000 56667870 QuvenzhanĀ@ Wallis 127 Hours 2010 18000000 18329466 James Franco 50/50 2011 8000000 34963967 Gordon-Levitt About Time 2013 12000000 15294553 Tom Hughes Amour 2012 8900000 225377 Isabelle Huppert Argo 2012 44500000 136019448 Clea DuVall Arrival 2016 47000000 100546139 Amy Adams Before Midnight 2013 3000000 8114507 Seamus Davey-Fitzpatrick Big Hero 6 2014 165000000 222487711 Damon Wayans Jr. Birdman or (The Unexpected Virtue of 2014 18000000 42335698 Emma Stone	12 Years a Slave 2013 20000000 56667870 Quvenzhané Wallis Wallis Scoot McNairy 127 Hours 2010 18000000 18329466 James Franco Treat Williams 50/50 2011 8000000 34963967 Joseph Gordon-Levitt Anna Kendrick About Time 2013 12000000 15294553 Tom Hughes Tom Hollander Amour 2012 8900000 225377 Isabelle Huppert Riva Emmanuelle Riva Argo 2012 44500000 136019448 Clea DuVall Scoot McNairy Arrival 2016 47000000 100546139 Amy Adams Peremy Renner Before Midnight 2013 3000000 8114507 Seamus Davey-Fitzpatrick Ariane Labed Birdman or (The Unexpected Virtue of 2014 18000000 42335698 Emma Stone Naomi Watts	12 Years a Slave 2013 20000000 56667870 Quvenzhané Wallis Scoot McNairy Taran Killam 127 Hours 2010 18000000 18329466 James Franco Treat Williams Kate Burton 50/50 2011 8000000 34963967 Joseph Gordon-Levitt Anna Kendrick Bryce Dallas Howard About Time 2013 12000000 15294553 Tom Hughes Tom Hollander Lindsay Duncan Amour 2012 8900000 225377 Isabelle Huppert Emmanuelle Riva Jean-Louis Trintignant Argo 2012 44500000 136019448 Clea DuVall Scoot McNairy Tate Donovan Arrival 2016 47000000 100546139 Amy Adams Jeremy Renner Forest Whitaker Before Midnight 2013 3000000 8114507 Seamus Davey-Fitzpatrick Ariane Labed Athina Rachel Tsangari Birdman or (The Unexpected Virtue of 2014 18000000 42335698 Emma Stone Naomi Watts Merritt Wever <th>12 Years a Slave 2013 20000000 56667870 Quvenzhané Wallis Wallis Scoot McNairy Taran Killam 2000 127 Hours 2010 18000000 18329466 James Franco Treat Williams Kate Burton 11000 50/50 2011 8000000 34963967 Joseph Gordon-Levitt Anna Kendrick Bryce Dallas Howard 23000 About Time 2013 12000000 15294553 Tom Hughes Tom Hollander Lindsay Duncan 565 Amour 2012 8900000 225377 Isabelle Huppert Emmanuelle Riva Trintignant 1678 Argo 2012 44500000 136019448 Clea DuVall Scoot McNairy Tate Donovan 1000 Arrival 2016 4700000 100546139 Amy Adams Jeremy Renner Forest Whitaker 35000 Before Midnight 2013 3000000 8114507 Seamus Davey-Fitzpatrick Ariane Labed Athina Rachel Tsangari 140 Big Hero 6 2014 16500000 222487711 Damon Wayans Jr.</th>	12 Years a Slave 2013 20000000 56667870 Quvenzhané Wallis Wallis Scoot McNairy Taran Killam 2000 127 Hours 2010 18000000 18329466 James Franco Treat Williams Kate Burton 11000 50/50 2011 8000000 34963967 Joseph Gordon-Levitt Anna Kendrick Bryce Dallas Howard 23000 About Time 2013 12000000 15294553 Tom Hughes Tom Hollander Lindsay Duncan 565 Amour 2012 8900000 225377 Isabelle Huppert Emmanuelle Riva Trintignant 1678 Argo 2012 44500000 136019448 Clea DuVall Scoot McNairy Tate Donovan 1000 Arrival 2016 4700000 100546139 Amy Adams Jeremy Renner Forest Whitaker 35000 Before Midnight 2013 3000000 8114507 Seamus Davey-Fitzpatrick Ariane Labed Athina Rachel Tsangari 140 Big Hero 6 2014 16500000 222487711 Damon Wayans Jr.

10 rows × 63 columns

```
In [22]:
```

```
#Plot profit vs budget
import matplotlib.pyplot as plt
df.plot.scatter(x='profit',y='budget')
```

Out[22]: <AxesSubplot:xlabel='profit', ylabel='budget'>



My Observations: Movies with higher budgets are not necessarily profitable

The dataset contains the 100 best performing movies from the year 2010 to 2016. However, the scatter plot tells a different story. You can notice that there are some movies with negative profit. Although good movies do incur losses, but there appear to be quite a few movie with losses. What can be the reason behind this? Lets have a closer look at this by finding the movies with negative profit.

```
In [27]:
           #Find the movies with negative profit
          df[df['profit'] < 0]['Title']</pre>
                           12 Years a Slave
Out[27]: 69
```

```
73
                         127 Hours
91
                              50/50
88
                        About Time
51
                               Argo
```

True Grit

95 Whiplash

25 Wreck-It Ralph

14 X-Men: Days of Future Past

1 Zootopia

Name: Title, Length: 89, dtype: object

In [24]: # Create the dataframe df_by_genre which will contain genre info only

Loading [MathJax]/jax/output/CommonHTML/fonts/TeX/fontdata.js