CSE 216 – Programming Abstractions (Fall 2019)
Programming Assignment # 2

In this assignment, you will create a software program that shows a particular statistics requested by a user. The program will be developed in Python language using PyCharm IDE (preferred). The data presented in this assignment is taken from The Movies Dataset[1] made available by Rounak Banik on Kaggle – which is an online community of data scientists. These files contain metadata for about 45,000 movies released on or before July 2017. Some information about movies in production or planned is given as well. We will use a file named movies_metadata from this dataset that contains information on the following:

1) adult: TRUE/FALSE
2) genres: Zero or multiple genres. It can be considered as a list of dictionaries in Python. E.g. [{'id': 80, 'name': 'Crime'}, {'id': 35, 'name': 'Comedy'}]
3) id: ID of the movie in this dataset.
4) Imdb_id: ID of the movie as assigned by IMDB
5) Original_language: Original language of movie as given by two letter code (e.g. es for Spanish)
6) Original_title: Original title of the movie
7) Popularity: Some value showing popularity of the movie – higher is better
8) Production_companies: Dictionary list containing names of production companies. E.g. [{'name': 'Universal Pictures', 'id': 33}, {'name': 'Largo Entertainment', 'id': 1644}, {'name': 'JVC Entertainment Networks', 'id': 4248}]
9) Production_countries: Dictionary list containing names of production countires. E.g. [{'iso_3166_1': 'US', 'name': 'United States of America'}]
10) Release_date: Date of release in MM/DD/YYYY format
11) Revenue: Revenue obtained (in $?)
12) Runtime: length of movie in minutes
13) Spoken_language: Language spoken in the movie
14) Status: One of cancelled, In Production, Planned, Post Production, Released
15) Title: Movie title

Your program will accept a calendar period from the user and should be able to answer all of the following questions:

1) "List most popular top 20 movies in a particular period."
2) "List most popular top 20 movies released in a particular language."
3) "List most popular top 20 movies released in a particular genre type."
4) "List most popular top 20 movies produced in a particular country."
5) "List top 20 movies which earned highest revenue in a particular period."
6) "List top 20 movies according to their runtime in a particular period."
7) "List the production companies of top 20 most popular movies in a particular period."

---

[1] The Movies Dataset, https://www.kaggle.com/rounakbanik/the-movies-dataset.

Your tasks are the following:

A. Download, unzip and open the project folder:

Download and unzip attached zip folder. Open folder in PyCharm. I have provided you a code to read the metadata file line by line as well as the code to display a list of questions. See files **moviedatareader.py**, **constants.py** and **UserInteraction.py (main file to run)**. The data file movies_metadata_edited.csv is also provided and it should be in the same directory as rest of the scripts. An example of parsing genres string is also given in **constants.py**. You need to extend this code to complete desired functionality.

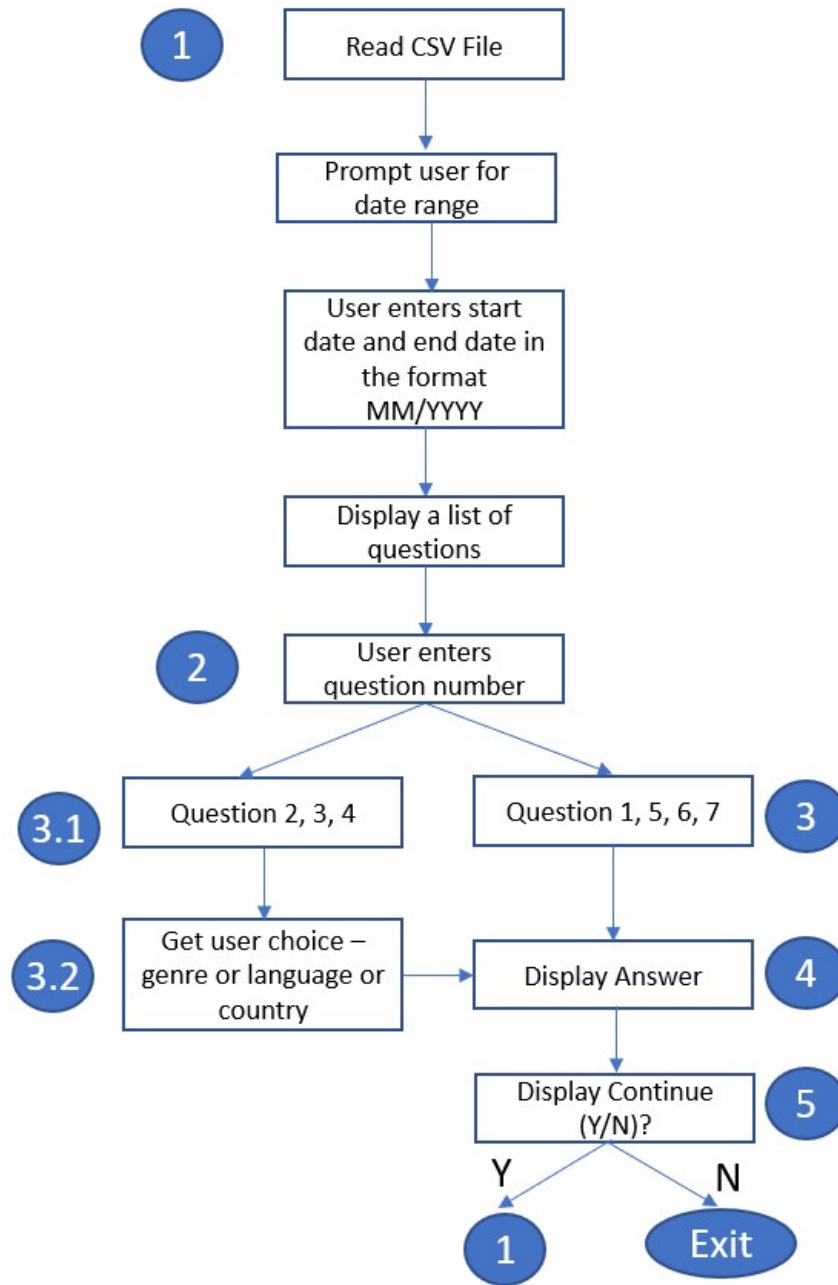B. **User interaction:** You will extend this basic project to provide the following user interaction:

Figure 1: User interaction workflow

C. Submission:

Submit all the Python scripts you will write as separate files. Do NOT submit the csv data file.

D. **Evaluation:** The graduate TA will download and run your program and it will be checked for correctness using a few use-cases. If required, the graduate TA will contact you for clarifications.
   Evaluation Rubric: (Total points: 40)
   1. Proper submission: All necessary Python scripts: 5 pts
   2. Reaching point 2 of user interaction: +5 pts
   3. Reaching point 3.2 of user interaction: +3 pts
   4. Displaying proper information: +3 pts each (21 points total)
   5. Reaching point 5 of user interaction: +3 pts
   6. Restarting from point 1: +3 pts

   Sample output is provided towards the end of this section.

Note:

If your code does not compile, it will not be graded.

Late submissions will not be accepted under any circumstances. Submit whatever you can if you were not able to fully complete the assignment.

To be safe, always, ALWAYS, prepare to submit ahead of time, not exactly AT last moment!

Submission deadline: Sunday 20 October, 11:59 PM

Sample output:

```
"C:\Users\SUNY  Korea  CS\AppData\Local\Programs\Python\Python37\python.exe"
C:/pravinp/SUNYK/Fall2019/CSE216/assignments/assignment2/the-movies-
dataset/UserInteraction.py
Total movie items =  45295
First date:  1900-01-01
Last date:  2020-12-16
Available date interval is from:  1900-01-01 till 2020-12-16
Enter valid interval start date in the format YYYY-MM-DD: 2010-01-01
Enter valid interval end date in the format YYYY-MM-DD: 2015-01-01
You entered:  2010-01-01  -  2015-01-01
Available questions are:
1 :  List most popular top 20 movies in a particular period.
2 :  List most popular top 20 movies released in a particular language.
3 :  List most popular top 20 movies released in a particular genre type.
4 :  List most popular top 20 movies produced in a particular country.
5 :  List top 20 movies which earned highest revenue in a particular period.
6 :  List top 20 movies according to their runtime in a particular period.
7 :  List the production companies of top 20 most popular movies in a particular
period.
Select the number of any of the above question. 1
_____
No.   Popularity  Movie title
_____
```

```
1)    213.849907   Big Hero 6
2)    183.870374   John Wick
3)    154.801009   Gone Girl
4)    147.098006   The Hunger Games: Mockingjay - Part 1
5)    89.887648    The Avengers
6)    76.93789     The Maze Runner
7)    75.385211    Dawn of the Planet of the Apes
8)    64.29999     Whiplash
9)    53.291601    Guardians of the Galaxy
10)   41.613762    Rise of the Planet of the Apes
11)   36.713807    Fury
12)   36.447603    Lucy
13)   34.905447    Thor: The Dark World
14)   34.047399    The Twilight Saga: Eclipse
15)   33.949359    Pacific Rim
16)   32.213481    Interstellar
17)   31.982986    Edge of Tomorrow
18)   31.718977    The Hobbit: The Battle of the Five Armies
19)   31.59594     The Imitation Game
20)   31.102267    The Amazing Spider-Man
_____
Total number of movies produced during the period  2010-01-01  -  2015-01-01 is
8781

_____
Enter [Y/N] to continue: Y
Available date interval is from:  1900-01-01 till 2020-12-16
Enter valid interval start date in the format YYYY-MM-DD: 2010-01-01
Enter valid interval end date in the format YYYY-MM-DD: 2015-01-01
You entered:  2010-01-01  -  2015-01-01
Available questions are:
1 :  List most popular top 20 movies in a particular period.
2 :  List most popular top 20 movies released in a particular language.
3 :  List most popular top 20 movies released in a particular genre type.
4 :  List most popular top 20 movies produced in a particular country.
5 :  List top 20 movies which earned highest revenue in a particular period.
6 :  List top 20 movies according to their runtime in a particular period.
7 :  List the production companies of top 20 most popular movies in a particular
period.
Select the number of any of the above question. 2
Available languages are:

ab
af
am
ar
ay
bg
bm
bn
bo
bs
ca
cn
cs
cy
da
de
```

el
en
eo
es
et
eu
fa
fi
fr
fy
gl
he
hi
hr
hu
hy
id
is
it
iu
ja
jv
ka
kk
kn
ko
ku
ky
la
lb
lo
lt
lv
mk
ml
mn
mr
ms
mt
nb
ne
nl
no
pa
pl
ps
pt
qu
ro
ru
rw
sh
si
sk
sl
sm
sq

```
sr
sv
ta
te
tg
th
tl
tr
uk
ur
uz
vi
wo
xx
zh
zu
```
Type any language code from above list: en

_____
No.   Popularity Movie title Movie language
_____
1)   213.849907   Big Hero 6   en
2)   183.870374   John Wick   en
3)   154.801009   Gone Girl   en
4)   147.098006   The Hunger Games: Mockingjay - Part 1   en
5)   89.887648   The Avengers   en
6)   76.93789   The Maze Runner   en
7)   75.385211   Dawn of the Planet of the Apes   en
8)   64.29999   Whiplash   en
9)   53.291601   Guardians of the Galaxy   en
10)   41.613762   Rise of the Planet of the Apes   en
11)   36.713807   Fury   en
12)   36.447603   Lucy   en
13)   34.905447   Thor: The Dark World   en
14)   34.047399   The Twilight Saga: Eclipse   en
15)   33.949359   Pacific Rim   en
16)   32.213481   Interstellar   en
17)   31.982986   Edge of Tomorrow   en
18)   31.718977   The Hobbit: The Battle of the Five Armies   en
19)   31.59594   The Imitation Game   en
20)   31.102267   The Amazing Spider-Man   en

_____
Total number of movies with language   en   during the period 2010-01-01   -
2015-01-01 is 6062

_____
Enter [Y/N] to continue: Y
Available date interval is from: 1900-01-01 till 2020-12-16
Enter valid interval start date in the format YYYY-MM-DD: 2010-01-01
Enter valid interval end date in the format YYYY-MM-DD: 2015-01-01
You entered: 2010-01-01   -   2015-01-01
Available questions are:
1 : List most popular top 20 movies in a particular period.
2 : List most popular top 20 movies released in a particular language.
3 : List most popular top 20 movies released in a particular genre type.
4 : List most popular top 20 movies produced in a particular country.
5 : List top 20 movies which earned highest revenue in a particular period.
6 : List top 20 movies according to their runtime in a particular period.

7 : List the production companies of top 20 most popular movies in a particular period.
Select the number of any of the above question. 3
Available genres are:
Action
Adventure
Animation
Comedy
Crime
Documentary
Drama
Family
Fantasy
Foreign
History
Horror
Music
Mystery
Romance
Science Fiction
TV Movie
Thriller
War
Western
Type any genre from above list: Romance

_____
No.   Popularity Movie title Genres
_____
1)   34.047399   The Twilight Saga: Eclipse   ['Adventure',   'Fantasy', 'Drama', 'Romance']
2)   26.080995   The Twilight Saga: Breaking Dawn - Part 2 ['Adventure', 'Fantasy', 'Drama', 'Romance']
3)   25.9725   The Twilight Saga: Breaking Dawn - Part 1 ['Adventure', 'Fantasy', 'Romance']
4)   20.485203   Beautiful Creatures   ['Fantasy', 'Drama', 'Romance']
5)   19.628834   Beauty and the Beast   ['Fantasy', 'Romance']
6)   19.467404   Maleficent   ['Fantasy',   'Adventure',   'Action', 'Family', 'Romance']
7)   18.459955   Prince of Persia: The Sands of Time   ['Adventure', 'Fantasy', 'Action', 'Romance']
8)   17.598936   The Great Gatsby   ['Drama', 'Romance']
9)   16.274653   The Fault in Our Stars   ['Romance', 'Drama']
10)   15.280219   Last Night   ['Drama', 'Romance']
11)   14.566103   Love, Rosie   ['Comedy', 'Romance']
12)   14.488111   Silver Linings Playbook   ['Drama', 'Comedy', 'Romance']
13)   14.347931   Underdogs   ['Animation', 'Adventure', 'Romance']
14)   14.336187   The Mortal Instruments: City of Bones   ['Action', 'Adventure', 'Drama', 'Mystery', 'Romance', 'Fantasy']
15)   14.331454   Hysteria   ['Comedy', 'Romance']
16)   14.061194   Beloved Sisters   ['Romance', 'Drama', 'History']
17)   13.869345   The Oranges   ['Comedy', 'Romance', 'Drama']
18)   13.829515   Her   ['Romance', 'Science Fiction', 'Drama']
19)   13.621581   Desire   ['Drama', 'Romance']
20)   13.594992   Only Lovers Left Alive   ['Drama', 'Romance']
_____
Total number of movies with Genre Romance during the period 2010-01-01 - 2015-01-01 is 1014

```
_____
Enter [Y/N] to continue: Y
Available date interval is from:  1900-01-01 till 2020-12-16
Enter valid interval start date in the format YYYY-MM-DD: 2010-01-01
Enter valid interval end date in the format YYYY-MM-DD: 2015-01-01
You entered:  2010-01-01  -  2015-01-01
Available questions are:
1 :  List most popular top 20 movies in a particular period.
2 :  List most popular top 20 movies released in a particular language.
3 :  List most popular top 20 movies released in a particular genre type.
4 :  List most popular top 20 movies produced in a particular country.
5 :  List top 20 movies which earned highest revenue in a particular period.
6 :  List top 20 movies according to their runtime in a particular period.
7 :  List the production companies of top 20 most popular movies in a particular
period.
Select the number of any of the above question. 4
Available countries are:
AE
AF
AL
AM
AN
AO
AQ
AR
AT
AU
AW
AZ
BA
BB
BD
BE
BF
BG
BM
BN
BO
BR
BS
BT
BW
BY
CA
CD
CG
CH
CI
CL
CM
CN
CO
CR
CS
CU
CY
CZ
```

DE
DK
DO
DZ
EC
EE
EG
ES
ET
FI
FR
GB
GE
GH
GI
GN
GR
GT
HK
HN
HR
HU
ID
IE
IL
IN
IQ
IR
IS
IT
JM
JO
JP
KE
KG
KH
KP
KR
KW
KY
KZ
LA
LB
LI
LK
LR
LT
LU
LV
LY
MA
MC
MD
ME
MG
MK
ML

MM
MN
MO
MQ
MR
MT
MX
MY
NA
NG
NI
NL
NO
NP
NZ
PA
PE
PF
PG
PH
PK
PL
PR
PS
PT
PY
QA
RO
RS
RU
RW
SA
SE
SG
SI
SK
SN
SO
SU
SV
SY
TD
TF
TH
TJ
TN
TR
TT
TW
TZ
UA
UG
UM
US
UY
UZ
VE

```
VN
WS
XC
XG
YU
ZA
ZW
Type any country code from above list: US
_____
No.    Popularity  Movie title Production countries
_____
1)     213.849907   Big Hero 6        ['US']
2)     183.870374   John Wick    ['CA', 'CN', 'US']
3)     154.801009   Gone Girl    ['US']
4)     147.098006   The Hunger Games: Mockingjay - Part 1     ['US']
5)     89.887648    The Avengers      ['US']
6)     76.93789     The Maze Runner    ['US']
7)     75.385211    Dawn of the Planet of the Apes      ['US']
8)     64.29999     Whiplash    ['US']
9)     53.291601    Guardians of the Galaxy       ['GB', 'US']
10)    41.613762    Rise of the Planet of the Apes      ['US']
11)    36.713807    Fury        ['GB', 'US', 'CN']
12)    36.447603    Lucy        ['FR', 'US']
13)    34.905447    Thor: The Dark World     ['US']
14)    34.047399    The Twilight Saga: Eclipse     ['US']
15)    33.949359    Pacific Rim       ['US', 'CA']
16)    32.213481    Interstellar      ['CA', 'US', 'GB']
17)    31.982986    Edge of Tomorrow        ['AU', 'US']
18)    31.718977    The Hobbit: The Battle of the Five Armies      ['NZ', 'US']
19)    31.102267    The Amazing Spider-Man       ['US']
20)    30.316249    12 Years a Slave        ['US', 'GB']
_____
Total number of movies produced in country  US  during the period  2010-01-01
-  2015-01-01 is  3433
_____
Enter [Y/N] to continue: Y
Available date interval is from:  1900-01-01 till 2020-12-16
Enter valid interval start date in the format YYYY-MM-DD: 2010-01-01
Enter valid interval end date in the format YYYY-MM-DD: 2015-01-01
You entered:  2010-01-01  -  2015-01-01
Available questions are:
1 :  List most popular top 20 movies in a particular period.
2 :  List most popular top 20 movies released in a particular language.
3 :  List most popular top 20 movies released in a particular genre type.
4 :  List most popular top 20 movies produced in a particular country.
5 :  List top 20 movies which earned highest revenue in a particular period.
6 :  List top 20 movies according to their runtime in a particular period.
7 :  List the production companies of top 20 most popular movies in a particular
period.
Select the number of any of the above question. 5
_____
No.    Revenue     Movie title
_____
1 )    1519557910       The Avengers
2 )    1342000000       Harry Potter and the Deathly Hallows: Part 2
3 )    1274219009       Frozen
4 )    1215439994       Iron Man 3
```

```
5  )    1123746996       Transformers: Dark of the Moon
6  )    1108561013       Skyfall
7  )    1091405097       Transformers: Age of Extinction
8  )    1084939099       The Dark Knight Rises
9  )    1066969703       Toy Story 3
10 )    1045713802       Pirates of the Caribbean: On Stranger Tides
11 )    1025491110       Alice in Wonderland
12 )    1021103568       The Hobbit: An Unexpected Journey
13 )    970761885   Despicable Me 2
14 )    958400000   The Hobbit: The Desolation of Smaug
15 )    956019788   The Hobbit: The Battle of the Five Armies
16 )    954305868   Harry Potter and the Deathly Hallows: Part 1
17 )    877244782   Ice Age: Continental Drift
18 )    847423452   The Hunger Games: Catching Fire
19 )    829000000   The Twilight Saga: Breaking Dawn - Part 2
20 )    825532764   Inception
_____
Total number of movies produced during the period  2010-01-01  -  2015-01-01 is
8781

_____
Enter [Y/N] to continue: Y
Available date interval is from:  1900-01-01 till 2020-12-16
Enter valid interval start date in the format YYYY-MM-DD: 2010-01-01
Enter valid interval end date in the format YYYY-MM-DD: 2015-01-01
You entered:  2010-01-01  -  2015-01-01
Available questions are:
1 :  List most popular top 20 movies in a particular period.
2 :  List most popular top 20 movies released in a particular language.
3 :  List most popular top 20 movies released in a particular genre type.
4 :  List most popular top 20 movies produced in a particular country.
5 :  List top 20 movies which earned highest revenue in a particular period.
6 :  List top 20 movies according to their runtime in a particular period.
7 :  List the production companies of top 20 most popular movies in a particular
period.
Select the number of any of the above question. 6

_____
No.    Runtime     Movie title

_____
1)     900     The Story of Film: An Odyssey
2)     840     The Roosevelts: An Intimate History
3)     585     The Untold History of the United States
4)     543     Long Way Down
5)     540     The Pacific
6)     468     The Sixties
7)     440     The Bible
8)     400     Crystal Lake Memories: The Complete History of Friday the 13th
9)     392     Frozen Planet
10)    389     World Without End
11)    363     Dancing on the Edge
12)    360     The Body Farm
13)    360     Century of Birthing
14)    360     The Men Who Built America
15)    350     Adolf Hitler: The Greatest Story Never Told
16)    338     Carlos
17)    336     Mildred Pierce
18)    330     Prohibition
19)    320     Gangs of Wasseypur
```

```
20)    315    Great Migrations
```
_____
Total number of movies produced during the period 2010-01-01 - 2015-01-01 is 8781

_____
Enter [Y/N] to continue: Y
Available date interval is from: 1900-01-01 till 2020-12-16
Enter valid interval start date in the format YYYY-MM-DD: 2010-01-01
Enter valid interval end date in the format YYYY-MM-DD: 2015-01-01
You entered: 2010-01-01 - 2015-01-01
Available questions are:
1 : List most popular top 20 movies in a particular period.
2 : List most popular top 20 movies released in a particular language.
3 : List most popular top 20 movies released in a particular genre type.
4 : List most popular top 20 movies produced in a particular country.
5 : List top 20 movies which earned highest revenue in a particular period.
6 : List top 20 movies according to their runtime in a particular period.
7 : List the production companies of top 20 most popular movies in a particular period.
Select the number of any of the above question. 7

_____
No.   Popularity Movie title Production companies
_____
1)   213.849907   Big Hero 6   ['Walt Disney Pictures', 'Walt Disney Animation Studios']
2)   183.870374   John Wick   ['Summit   Entertainment',   'Warner   Bros.', '87Eleven', 'DefyNite Films', 'MJW Films']
3)   154.801009   Gone Girl   ['Twentieth Century Fox Film Corporation', 'Regency Enterprises', 'New Regency Pictures', 'Pacific Standard', 'TSG Entertainment', 'Artemple - Hollywood']
4)   147.098006   The Hunger Games: Mockingjay - Part 1   ['Lionsgate', 'Color Force']
5)   89.887648   The Avengers   ['Paramount Pictures', 'Marvel Studios']
6)   76.93789   The Maze Runner   ['Ingenious Media', 'Twentieth Century Fox Film Corporation', 'Gotham Group', 'Dayday Films', 'Temple Hill Entertainment', 'TSG Entertainment']
7)   75.385211   Dawn of the Planet of the Apes   ['Ingenious   Media', 'Chernin Entertainment', 'TSG Entertainment']
8)   64.29999   Whiplash   ['Bold Films', 'Blumhouse Productions', 'Right of Way Films']
9)   53.291601   Guardians of the Galaxy   ['Marvel   Studios',   'Moving Picture Company (MPC)', 'Bulletproof Cupid', 'Revolution Sun Studios']
10)   41.613762   Rise of the Planet of the Apes   ['Ingenious   Film Partners', 'Ingenious Media', 'Twentieth Century Fox Film Corporation', 'Dune Entertainment', 'Dune Entertainment III', 'Chernin Entertainment', 'Big Screen Productions']
11)   36.713807   Fury   ['Columbia   Pictures',   'QED   International', 'Crave Films', 'LStar Capital', 'Huayi Brothers Media', 'Le Grisbi Productions']
12)   36.447603   Lucy   ['Universal Pictures', 'TF1 Films Production', 'Canal+', 'EuropaCorp', 'Ciné+']
13)   34.905447   Thor: The Dark World   ['Marvel Studios']
14)   34.047399   The Twilight Saga: Eclipse   ['Summit   Entertainment', 'Maverick Films', 'Imprint Entertainment', 'Sunswept Entertainment', 'Temple Hill Entertainment']
15)   33.949359   Pacific Rim   ['Legendary Pictures', 'Warner Bros.', 'Disney Double Dare You (DDY)', 'Indochina Productions']

16) 32.213481   Interstellar   ['Paramount   Pictures',   'Legendary Pictures', 'Warner Bros.', 'Syncopy', 'Lynda Obst Productions']
17) 31.982986   Edge of Tomorrow   ['Village   Roadshow   Pictures', 'Warner Bros.', 'Viz Media', 'Province of British Columbia Production Services Tax Credit', '3 Arts Entertainment', 'RatPac-Dune Entertainment']
18) 31.718977   The Hobbit: The Battle of the Five Armies   ['WingNut Films', 'New Line Cinema', 'Warner Bros. Pictures', '3Foot7', 'Metro-Goldwyn-Mayer (MGM)']
19) 31.59594   The Imitation Game   ['Black   Bear   Pictures',   'Bristol Automotive']
20) 31.102267   The Amazing Spider-Man   ['Columbia Pictures', 'Laura Ziskin Productions', 'Marvel Entertainment']

_____
Total number of movies produced during the period 2010-01-01 - 2015-01-01 is 8781

_____
Enter [Y/N] to continue: N
Thank you. Exiting!!