In [32]:
```python
#Ingesting All_Users_Sentiment file into users
import boto3
import pandas as pd


s3_client = boto3.client("s3")

BUCKET='etoilecinemas'
KEY='Cloud_Computing_Fina_Assignment_BOXOFFICE/ALL_User_With_sentiment.c
sv'

response = s3_client.get_object(Bucket=BUCKET, Key=KEY)
users = pd.read_csv(response.get("Body"))
users.head()
```

Out[32]:

	Unnamed: 0	ID_Movie	User	Score	Date	Review	Helpful	Out_
0	0	1	DemiRonin	7	16/04/2011 00:00	This review contains spoilers, click expand to...	0	
1	1	1	steven	4	04/08/2010 00:00	I don't mean to be a Debbie Downer and I am al...	1	
2	2	3	RayJ.	9	05/11/2004 00:00	Superb.	0	
3	3	3	MichaelV.	9	24/03/2003 00:00	Lillo is so hot!	0	
4	4	3	GilbertMulroneycakesAndFriends	6	06/12/2002 00:00	What the hell is that title all about? I assum...	0	

In [33]:
```
#Checking overall view of dataset
users.info
```

Out[33]: <bound method DataFrame.info of   Unnamed: 0   ID_Movie
         User Score \
         0   0   1   DemiRonin   7
         1   1   1   steven   4
         2   2   3   RayJ.   9
         3   3   3   MichaelV.   9
         4   4   3   GilbertMulroneycakesAndFriends   6
         ...   .
         286354   286354   10781   Piero-   10
         286355   286355   10781   lionelbot   10
         286356   286356   10781   AaronJupiter   0
         286357   286357   10781   Stephanon   0
         286358   286358   10781   angel404   5

                         Date   Re
         view \
         0   16/04/2011 00:00   This review contains spoilers, click expand t
         o...
         1   04/08/2010 00:00   I don't mean to be a Debbie Downer and I am a
         l...
         2   05/11/2004 00:00   Sup
         erb.
         3   24/03/2003 00:00   Lillo is so
         hot!
         4   06/12/2002 00:00   What the hell is that title all about? I assu
         m...
         ...   ...
         ...
         286354   25/02/2021 00:00   It was really beauti
         ful
         286355   26/02/2021 00:00   Impresionante como han plasmado a la perfecci
         ó...
         286356   24/04/2021 00:00   Kate Hudson is a terrible actress and in Musi
         c...
         286357   18/02/2021 00:00   Disgusting and offensive portrayal of autism
         n...
         286358   24/02/2021 00:00   This movie is amazing, it was made with love
         a...

                  Helpful   Out_of   Spoiler   Sentiment
         0   0   0   2855.0   12.0
         1   1   4   0.0   2.0
         2   0   0   0.0   0.0
         3   0   0   0.0   0.0
         4   0   0   0.0   -4.0
         ...   .
         286354   0   1   0.0   1.0
         286355   0   1   0.0   0.0
         286356   0   0   0.0   -3.0
         286357   0   0   1.0   -2.0
         286358   0   0   1.0   3.0

         [286359 rows x 10 columns]>
```

```
In [34]:  #checking the data shape
          users.shape

Out[34]:  (286359, 10)
```

```
In [35]:  #Checking missing values
          users.isna().sum()

Out[35]:  Unnamed: 0        0
          ID_Movie          0
          User              5
          Score             0
          Date              0
          Review          109
          Helpful           0
          Out_of            0
          Spoiler           5
          Sentiment         0
          dtype: int64
```

```
In [36]:  #Checking UNIQUE values of each column of users
          users.nunique()

Out[36]:  Unnamed: 0    286359
          ID_Movie        8259
          User           78910
          Score             11
          Date            7386
          Review        222325
          Helpful          181
          Out_of           236
          Spoiler         1656
          Sentiment        165
          dtype: int64
```

Filtering out Most useful variable for the project's queries

ID_Moviee, User, Score, Review, Sentiment from Users Data Table

In [37]:

```python
#Ingesting Critic Data File into crtix_df

BUCKET='etoilecinemas'
KEY='Cloud_Computing_Fina_Assignment_BOXOFFICE/ALL_CRTIX_FULLL_FIXED.csv'

response = s3_client.get_object(Bucket=BUCKET, Key=KEY)
critics = pd.read_csv(response.get("Body"))
critics.head()
```

/opt/conda/lib/python3.7/site-packages/IPython/core/interactiveshell.py:3063: DtypeWarning: Columns (4) have mixed types.Specify dtype option on import or set low_memory=False.
  interactivity=interactivity, compiler=compiler, result=result)

Out[37]:

	Unnamed: 0	Unnamed: 0.1	ID_Movie	Expert	Score	Date	Review	Sentiment	Unnam
0	0	0	7369	RogerEbert.com	88	Aug 7, 2015	Call Me Lucky will be an especially grueling r...	3	l
1	1	1	7369	New York Daily News	80	Aug 6, 2015	Angry, quixotic, tragic, heroic — Crimmins' li...	1	l
2	2	2	7369	Village Voice	80	Aug 4, 2015	Call Me Lucky is a loving but fair portrait of...	7	l
3	3	3	7369	TheWrap	75	Aug 10, 2015	There should be more Crimmins performance foot...	2	l
4	4	4	7369	Movie Nation	75	Jul 28, 2015	Call Me Lucky is another of those "the funnies...	3	l

In [7]: `critics.shape`

Out[7]: (223825, 9)

In [8]: `critics.info`

Out[8]: `<bound method DataFrame.info of`

	Unnamed: 0	Unnamed: 0.1	ID_Mo vie	Expert Score
0	0	0	7369	RogerEbert.com	88
1	1	1	7369	New York Daily News	80
2	2	2	7369	Village Voice	80
3	3	3	7369	TheWrap	75
4	4	4	7369	Movie Nation	75
...	...	...	...	...	...
223820	15576	15576	7206	Slant Magazine	75
223821	15577	15577	7206	Los Angeles Times	70
223822	15578	15578	7206	The New York Times	70
223823	15579	15579	7206	The Dissolve	70
223824	15580	15580	7206	Village Voice	70

	Date	Review
0	Aug 7, 2015	Call Me Lucky will be an especially grueling r...
1	Aug 6, 2015	Angry, quixotic, tragic, heroic — Crimmins' li...
2	Aug 4, 2015	Call Me Lucky is a loving but fair portrait of...
3	Aug 10, 2015	There should be more Crimmins performance foot...
4	Jul 28, 2015	Call Me Lucky is another of those "the funnies...
...	...	...
223820	Jan 31, 2015	The doc adopts the viewpoint specifically of t...
223821	Feb 12, 2015	The film is enough to prompt soul-searching am...
223822	Feb 5, 2015	In lingering over moody night streets and trem...
223823	Feb 4, 2015	Josue shows perhaps too much restraint, as if ...
223824	Feb 3, 2015	Josue tries to reclaim his narrative with this...

	Sentiment	Unnamed: 8
0	3	NaN
1	1	NaN
2	7	NaN
3	2	NaN
4	3	NaN
...	...	...
223820	1	NaN
223821	4	NaN
223822	−1	NaN
223823	−3	NaN
223824	3	NaN

`[223825 rows x 9 columns]>`

```
In [9]: critics.isna().sum()
```

```
Out[9]: Unnamed: 0                0
        Unnamed: 0.1              0
        ID_Movie                  0
        Expert                    0
        Score                     0
        Date                 113676
        Review                    6
        Sentiment                 0
        Unnamed: 8           223825
        dtype: int64
```

```
In [10]: critics.nunique()
```

```
Out[10]: Unnamed: 0        208244
         Unnamed: 0.1      203135
         ID_Movie           10510
         Expert               201
         Score                184
         Date                3697
         Review            220492
         Sentiment             43
         Unnamed: 8             0
         dtype: int64
```

In [23]:
```python
#Ingesting Movie Products into movies_df

BUCKET='etoilecinemas'
KEY='Cloud_Computing_Fina_Assignment_BOXOFFICE/ALL_FIXED_Product_Movie.csv'

response = s3_client.get_object(Bucket=BUCKET, Key=KEY)
movies = pd.read_csv(response.get("Body"))
movies.head()
```

Out[23]:

	Unnamed: 0	ID	Title	Publisher	Release_Date	Summary	Director	Starring
0	0	1	9.99	Regent Releasing	unknown	Have you ever wondered "What is the meaning of...	Tatia Rosenthal	Anthony LaPaglia, Barry Otto, Ben Mendelsohn, ...
1	1	2	$pent	Regent Releasing	unknown	This comic drama examines the relationships an...	Gil Cates Jr.	Barbara Barrie, Bassia Loebel, Charlie Spradli...
2	2	3	'R Xmas	Pathfinder Pictures	08-Nov-02	It's a few days before Christmas, and a Latin ...	Abel Ferrara	Andrew Fiscella, Anne Ackerman, Denia Brache, ...
3	3	4	(500) Days of Summer	Fox Searchlight Pictures	17-Jul-09	After it looks as if she's left his life for g...	Marc Webb	Joseph Gordon-Levitt, Zooey Deschanel, Charles...	Drama
4	4	5	1	IFC Midnight	unknown	NaN	NaN	NaN

5 rows × 22 columns

In [12]: `movies.info`

```
Out[12]: <bound method DataFrame.info of      Unnamed: 0    ID
         Title  \
         0             0      1                        9.99
         1             1      2                      $pent
         2             2      3                    'R Xmas
         3             3      4       (500) Days of Summer
         4             4      5                          1
         ...         ...    ...                        ...
         10776     10776  10777             Blithe Spirit
         10777     10777  10778   Breaking News in Yuba County
         10778     10778  10779             American Skin
         10779     10779  10780               The Virtuoso
         10780     10780  10781                     Music

                                          Publisher    Release_Date  \
         0                          Regent Releasing         unknown
         1                          Regent Releasing         unknown
         2                        Pathfinder Pictures       08-Nov-02
         3                   Fox Searchlight Pictures       17-Jul-09
         4                              IFC Midnight         unknown
         ...                                   ...             ...
         10776                             IFC Films  February 19, 2021
         10777  American International Pictures (AIP)  February 12, 2021
         10778                                   NaN  January 15, 2021
         10779                             Lionsgate     April 30, 2021
         10780                                  IMAX  February 12, 2021

                                            Summary        Direct
         or  \
         0      Have you ever wondered "What is the meaning of...  Tatia Rosenth
         al
         1      This comic drama examines the relationships an...    Gil Cates J
         r.
         2      It's a few days before Christmas, and a Latin ...      Abel Ferra
         ra
         3      After it looks as if she's left his life for g...       Marc We
         bb
         4                                                NaN              N
         aN
         ...                                                ...
         ...
         10776  Best-selling crime novelist Charles (Dan Steve...     Edward Ha
         ll
         10777  After her husband goes missing, Sue Buttons (A...     Tate Tayl
         or
         10778  This is a MUST SEE for all white people and es...     Nate Park
         er
         10779  Danger, deception, and murder descend upon a s...   Nick Staglia
         no
         10780  Zu (Kate Hudson) is a newly sober drug dealer ...            S
         ia

                                            Starring  \
         0      Anthony LaPaglia, Barry Otto, Ben Mendelsohn, ...
         1      Barbara Barrie, Bassia Loebel, Charlie Spradli...
         2      Andrew Fiscella, Anne Ackerman, Denia Brache, ...
         3      Joseph Gordon-Levitt, Zooey Deschanel, Charles...
```

```
4                                                          NaN
...                                                        ...
10776   Adil Ray, Aimee-Ffion Edwards, Dan Stevens, Em...
10777   Allison Janney, Awkwafina, Chris Lowell, Clift...
10778   Beau Knapp, Dominic Bogart, Larry Sullivan, Mi...
10779   Abbie Cornish, Anson Mount, Anthony Hopkins, B...
10780   Angelina Capozzoli, Ben Schwartz, Beto Calvill...

                                                       Genre  \
0                                            Drama;Animation
1           Genre:                                      ...
2                                                Drama;Crime
3                                       Drama;Comedy;Romance
4                                                       NULL
...                                                      ...
10776                             Fantasy,Comedy,Romance
10777                       Drama,Thriller,Comedy,Crime
10778                                              Drama
10779                             Action,Thriller,Crime
10780                                      Drama,Musical

                                                      Rating  ... Meta_Pos_
Count  \
0       Rated R for language and brief sexuality and n...  ...
12.0
1                                                      NaN  ...
1.0
2       Rated R for strong language, drug content and ...  ...
4.0
3            Rated PG-13 for sexual material and language.  ...
33.0
4                                                      NaN  ...
3.0
...                                                      ...  ...
...
10776                                                   NR  ...
0.0
10777                                                    R  ...
0.0
10778                                                  NaN  ...
1.0
10779                                                    R  ...
0.0
10780                                                PG-13  ...
0.0

        Meta_Neut_Count  Meta_Neg_Count  User_Score  User_Pos_Count  \
0                   3.0             0.0         7.1               1
1                   3.0             4.0         tbd               0
2                   5.0             2.0         7.1               4
3                   3.0             0.0         7.9             161
4                   2.0             1.0           8               0
...                 ...             ...         ...             ...
10776               8.0            11.0         5.8               2
10777               2.0             5.0         3.8               0
10778               2.0             7.0         8.2              20
10779               3.0             4.0         tbd               1
```

```
10780                6.0              12.0            6.8                      114

          User_Neut_Count User_Neg_Count   \
0                  1.0                0
1                  0.0                0
2                  0.0                0
3                  7.0               24
4                  0.0                0
...                ...              ...
10776              8.0               11
10777              2.0                5
10778              2.0                7
10779              3.0                4
10780              6.0               12

                                                    URevUrl   \
0                                                       NaN
1                                                       NaN
2                                                       NaN
3                                                       NaN
4                                                       NaN
...                                                     ...
10776                                                   NaN
10777  /movie/breaking-news-in-yuba-county/user-reviews
10778                  /movie/american-skin/user-reviews
10779                  /movie/the-virtuoso/user-reviews
10780                         /movie/music/user-reviews

                                                    CRevUrl   \
0                                                       NaN
1                                                       NaN
2                                                       NaN
3                                                       NaN
4                                                       NaN
...                                                     ...
10776                  /movie/blithe-spirit/critic-reviews
10777  /movie/breaking-news-in-yuba-county/critic-rev...
10778                  /movie/american-skin/critic-reviews
10779                  /movie/the-virtuoso/critic-reviews
10780                         /movie/music/critic-reviews

                                       URL
0                                      NaN
1                                      NaN
2                                      NaN
3                                      NaN
4                                      NaN
...                                    ...
10776                  /movie/blithe-spirit
10777  /movie/breaking-news-in-yuba-county
10778                  /movie/american-skin
10779                  /movie/the-virtuoso
10780                         /movie/music

[10781 rows x 22 columns]>
```

```
In [29]:  movies.isna().sum()
```

```
Out[29]:  Unnamed: 0             0
          ID                    0
          Title                 0
          Publisher           288
          Release_Date         24
          Summary             526
          Director            614
          Starring           1855
          Genre                29
          Rating             1951
          Runtime             786
          Metascore            24
          Meta_Pos_Count       30
          Meta_Neut_Count      30
          Meta_Neg_Count       30
          User_Score           24
          User_Pos_Count      263
          User_Neut_Count     409
          User_Neg_Count     1113
          URevUrl            7371
          CRevUrl            6446
          URL                6422
          dtype: int64
```

```
In [30]:  movies.nunique()
```

```
Out[30]:  Unnamed: 0         10781
          ID                 10781
          Title              10597
          Publisher           1027
          Release_Date        2305
          Summary            10254
          Director            5324
          Starring            8920
          Genre               1652
          Rating              3555
          Runtime              177
          Metascore             98
          Meta_Pos_Count        58
          Meta_Neut_Count       36
          Meta_Neg_Count        30
          User_Score           101
          User_Pos_Count       486
          User_Neut_Count       72
          User_Neg_Count       149
          URevUrl             3410
          CRevUrl             4335
          URL                 4359
          dtype: int64
```

```
In [31]:  #Informative variables are Title, Rating, Metascore, Meta_Pos_Count, Met
          a_Neut_Count, Meta_Neg_Count,
          #User_Score, User_Pos_count, User_Neut_Count, User_Neg_Count.
```

In [40]:
```python
#Ingesting Sales Data into sales

BUCKET='etoilecinemas'
KEY='Cloud_Computing_Fina_Assignment_BOXOFFICE/Sales_FinalMovieData2000u
ntil2010.csv'

response = s3_client.get_object(Bucket=BUCKET, Key=KEY)
sales = pd.read_csv(response.get("Body"))
sales.head()
```

Out[40]:

	WeekData	MovieName	Rank_data	PreviousWeekRank	GrossW	Theaters
0	January 7-13, 2000	Stuart Little	1	1	13,012,299.00	2,979
1	January 7-13, 2000	The Green Mile	2	3	12,521,303.00	2,678
2	January 7-13, 2000	The Talented Mr. Ripley	3	2	11,780,319.00	2,316
3	January 7-13, 2000	Any Given Sunday	4	4	10,971,011.00	2,505
4	January 7-13, 2000	Galaxy Quest	5	6	9,784,389.00	2,450

In [42]:
```python
sales.shape
```

Out[42]: (118831, 6)

In [44]: `sales.info`

Out[44]:
```
<bound method DataFrame.info of                              WeekData
        MovieName   Rank_data  \
        0               January 7-13, 2000          Stuart Little            1
        1               January 7-13, 2000          The Green Mile           2
        2               January 7-13, 2000   The Talented Mr. Ripley         3
        3               January 7-13, 2000          Any Given Sunday         4
        4               January 7-13, 2000          Galaxy Quest             5
        ...                            ...                     ...         ...
        118826  December 27-January 2, 2020               Stuffed           75
        118827  December 27-January 2, 2020        A Bigger Splash          76
        118828  December 27-January 2, 2020        Give Me Liberty          77
        118829  December 27-January 2, 2020        Chinese Portrait         78
        118830  December 27-January 2, 2020        Light from Light         79

                PreviousWeekRank        GrossW Theaters
        0                      1  13,012,299.00    2,979
        1                      3  12,521,303.00    2,678
        2                      2  11,780,319.00    2,316
        3                      4  10,971,011.00    2,505
        4                      6   9,784,389.00    2,450
        ...                  ...            ...      ...
        118826                81            320        1
        118827                 -            246        1
        118828                67            246        1
        118829                77            160        1
        118830                 -             89        1

        [118831 rows x 6 columns]>
```

In [45]: `sales.nunique()`

Out[45]:
```
WeekData            1040
MovieName          11557
Rank_data            164
PreviousWeekRank     161
GrossW             78275
Theaters            4005
dtype: int64
```

In [47]: *#Useful variables for the project are MovieName, Rank_data, GrossW*