Investigate_a_Dataset

March 10, 2019

1 Project: Investigating Movie Data

1.1 Table of Contents

Introduction
Data Wrangling
Exploratory Data Analysis
Conclusions
Introduction I chose the movie dataset since I really love to watch movies. Movies have been advancing year over year due to many factors. We will take a look into the tmdb move data set to explore trends year over year as well as what is popular and how much movies make since the emergence of new technologies.

Few questions I have: 1. Profits change year over year? 2. Are certain genres more popular than the others? 3. Do certain types of movies have longer run times? 4. Does ratings affect the profit the movie makes at the end of the day?

```
In [69]: # Use this cell to set up import statements for all of the packages that you
         #   plan to use.
         import pandas as pd
         import numpy as np
         import matplotlib.pyplot as plt
         # Remember to include a 'magic word' so that your visualizations are plotted
         #   inline with the notebook. See this page for more:
         #   http://ipython.readthedocs.io/en/stable/interactive/magics.html
```

Data Wrangling

```
In [70]: # Load in the movie dataset here using pandas
         df = pd.read_csv('tmdb-movies_dataset.csv')
         df.head()
```

Out[70]:

	id	imdb_id	popularity	budget	revenue	\
0	135397	tt0369610	32.985763	150000000	1513528810	
1	76341	tt1392190	28.419936	150000000	378436354	
2	262500	tt2908446	13.112507	110000000	295238201	
3	140607	tt2488496	11.173104	200000000	2068178225	
4	168259	tt2820852	9.335014	190000000	1506249360	

```
                original_title  \
0                Jurassic World
1            Mad Max: Fury Road
2                      Insurgent
3  Star Wars: The Force Awakens
4                      Furious 7

                                                 cast  \
0  Chris Pratt|Bryce Dallas Howard|Irrfan Khan|Vi...
1  Tom Hardy|Charlize Theron|Hugh Keays-Byrne|Nic...
2  Shailene Woodley|Theo James|Kate Winslet|Ansel...
3  Harrison Ford|Mark Hamill|Carrie Fisher|Adam D...
4  Vin Diesel|Paul Walker|Jason Statham|Michelle ...

                                       homepage          director  \
0                    http://www.jurassicworld.com/   Colin Trevorrow
1                      http://www.madmaxmovie.com/     George Miller
2    http://www.thedivergentseries.movie/#insurgent  Robert Schwentke
3  http://www.starwars.com/films/star-wars-episod...       J.J. Abrams
4                          http://www.furious7.com/         James Wan

                         tagline      ...       \
0             The park is open.      ...
1            What a Lovely Day.      ...
2    One Choice Can Destroy You      ...
3  Every generation has a story.     ...
4           Vengeance Hits Home      ...

                                       overview runtime  \
0  Twenty-two years after the events of Jurassic ...     124
1  An apocalyptic story set in the furthest reach...     120
2  Beatrice Prior must confront her inner demons ...     119
3  Thirty years after defeating the Galactic Empi...     136
4  Deckard Shaw seeks revenge against Dominic Tor...     137

                                    genres  \
0  Action|Adventure|Science Fiction|Thriller
1  Action|Adventure|Science Fiction|Thriller
2         Adventure|Science Fiction|Thriller
3   Action|Adventure|Science Fiction|Fantasy
4                     Action|Crime|Thriller

                              production_companies release_date vote_count  \
0  Universal Studios|Amblin Entertainment|Legenda...       6/9/15       5562
1  Village Roadshow Pictures|Kennedy Miller Produ...      5/13/15       6185
2  Summit Entertainment|Mandeville Films|Red Wago...      3/18/15       2480
3          Lucasfilm|Truenorth Productions|Bad Robot     12/15/15       5292
4  Universal Pictures|Original Film|Media Rights ...       4/1/15       2947
```

```
   vote_average  release_year    budget_adj   revenue_adj
0           6.5          2015  1.379999e+08  1.392446e+09
1           7.1          2015  1.379999e+08  3.481613e+08
2           6.3          2015  1.012000e+08  2.716190e+08
3           7.5          2015  1.839999e+08  1.902723e+09
4           7.3          2015  1.747999e+08  1.385749e+09

[5 rows x 21 columns]
```

By using df.head() we want to see what the variables look like, how the rows are divided and what kind of questions we could ask

```
In [71]: df.info()

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 10866 entries, 0 to 10865
Data columns (total 21 columns):
id                      10866 non-null int64
imdb_id                 10856 non-null object
popularity              10866 non-null float64
budget                  10866 non-null int64
revenue                 10866 non-null int64
original_title          10866 non-null object
cast                    10790 non-null object
homepage                2936 non-null object
director                10822 non-null object
tagline                 8042 non-null object
keywords                9373 non-null object
overview                10862 non-null object
runtime                 10866 non-null int64
genres                  10843 non-null object
production_companies    9836 non-null object
release_date            10866 non-null object
vote_count              10866 non-null int64
vote_average            10866 non-null float64
release_year            10866 non-null int64
budget_adj              10866 non-null float64
revenue_adj             10866 non-null float64
dtypes: float64(4), int64(6), object(11)
memory usage: 1.7+ MB
```

Based on the using the function info we are able to see that we are have 10866 entries. We can see that homepage, tagline, keywords, and production companies are missing a large amount of data. The imdb_id, cast, director, overview, and genres are missing minimal amounts of data. Looking at the cast we can see that each cast is divided by a '|' same with genres and production companies. We should look to make the 'release date' a date and not an object.

I am guessing that the most missing rows are due to release year of the movie. The older the movie the more likely the movie will have missing rows such as a homepage or taglines.

```
In [72]: df.describe()
```

Out[72]:

	id	popularity	budget	revenue	runtime
count	10866.000000	10866.000000	1.086600e+04	1.086600e+04	10866.000000
mean	66064.177434	0.646441	1.462570e+07	3.982332e+07	102.070863
std	92130.136561	1.000185	3.091321e+07	1.170035e+08	31.381405
min	5.000000	0.000065	0.000000e+00	0.000000e+00	0.000000
25%	10596.250000	0.207583	0.000000e+00	0.000000e+00	90.000000
50%	20669.000000	0.383856	0.000000e+00	0.000000e+00	99.000000
75%	75610.000000	0.713817	1.500000e+07	2.400000e+07	111.000000
max	417859.000000	32.985763	4.250000e+08	2.781506e+09	900.000000

	vote_count	vote_average	release_year	budget_adj	revenue_adj
count	10866.000000	10866.000000	10866.000000	1.086600e+04	1.086600e+04
mean	217.389748	5.974922	2001.322658	1.755104e+07	5.136436e+07
std	575.619058	0.935142	12.812941	3.430616e+07	1.446325e+08
min	10.000000	1.500000	1960.000000	0.000000e+00	0.000000e+00
25%	17.000000	5.400000	1995.000000	0.000000e+00	0.000000e+00
50%	38.000000	6.000000	2006.000000	0.000000e+00	0.000000e+00
75%	145.750000	6.600000	2011.000000	2.085325e+07	3.369710e+07
max	9767.000000	9.200000	2015.000000	4.250000e+08	2.827124e+09

I want to see what the minimum values are for release year to see what the oldest movie we could be analyzing

1.1.1 Cleaning and Restructuring Movie Data

Lets clean a bit here. 1. Change release_date to a date format 2. Change genre column to find averages (must create a separate dateframe specifically for genre analysis) 3. Drop rows with missing values 4. Create a new column to figure out the gross profit per movie 5. Insert median revenue and median budget into missing value rows for those columns

```
In [73]: # After discussing the structure of the data and any problems that need to be
         #   cleaned, perform those cleaning steps in the second part of this section.
         # Insert median values for missing cells in revenue_adj and budget_adj (found on stacko
         med_rev_adj = df.query('revenue_adj > 0')['revenue_adj'].median()
         med_budg_adj = df.query('budget_adj > 0')['budget_adj'].median()
         df['revenue_adj']=df['revenue_adj'].replace(0,med_rev_adj)
         df['budget_adj']=df['budget_adj'].replace(0,med_budg_adj)
```

```
In [74]: # Calculate out an adj_gross_profit
         df['adj_gross_profit'] = df['revenue_adj'] - df['budget_adj']
         df['release_date'] =  pd.to_datetime(df['release_date'])
         # Dropping these columns so we drop as minimum NA rows as possible
         df = df.drop('homepage', axis = 1)
         df = df.drop('tagline', axis = 1)
```

```
df = df.drop('keywords', axis = 1)

# Drop na values
df = df.dropna()
df.head()
```

Out[74]:

	id	imdb_id	popularity	budget	revenue	original_title	cast	director	overview	runtime	genres	production_companies	release_date	vote_count
0	135397	tt0369610	32.985763	150000000	1513528810	Jurassic World	Chris Pratt|Bryce Dallas Howard|Irrfan Khan|Vi...	Colin Trevorrow	Twenty-two years after the events of Jurassic ...	124	Action|Adventure|Science Fiction|Thriller	Universal Studios|Amblin Entertainment|Legenda...	2015-06-09	5562
1	76341	tt1392190	28.419936	150000000	378436354	Mad Max: Fury Road	Tom Hardy|Charlize Theron|Hugh Keays-Byrne|Nic...	George Miller	An apocalyptic story set in the furthest reach...	120	Action|Adventure|Science Fiction|Thriller	Village Roadshow Pictures|Kennedy Miller Produ...	2015-05-13	6185
2	262500	tt2908446	13.112507	110000000	295238201	Insurgent	Shailene Woodley|Theo James|Kate Winslet|Ansel...	Robert Schwentke	Beatrice Prior must confront her inner demons ...	119	Adventure|Science Fiction|Thriller	Summit Entertainment|Mandeville Films|Red Wago...	2015-03-18	2480
3	140607	tt2488496	11.173104	200000000	2068178225	Star Wars: The Force Awakens	Harrison Ford|Mark Hamill|Carrie Fisher|Adam D...	J.J. Abrams	Thirty years after defeating the Galactic Empi...	136	Action|Adventure|Science Fiction|Fantasy	Lucasfilm|Truenorth Productions|Bad Robot	2015-12-15	5292
4	168259	tt2820852	9.335014	190000000	1506249360	Furious 7	Vin Diesel|Paul Walker|Jason Statham|Michelle ...	James Wan	Deckard Shaw seeks revenge against Dominic Tor...	137	Action|Crime|Thriller	Universal Pictures|Original Film|Media Rights ...	2015-04-01	2947

	vote_average	release_year	budget_adj	revenue_adj	adj_gross_profit
0	6.5	2015	1.379999e+08	1.392446e+09	1.254446e+09
1	7.1	2015	1.379999e+08	3.481613e+08	2.101614e+08
2	6.3	2015	1.012000e+08	2.716190e+08	1.704191e+08
3	7.5	2015	1.839999e+08	1.902723e+09	1.718723e+09
4	7.3	2015	1.747999e+08	1.385749e+09	1.210949e+09

In [75]:
```
# Checking our dataset on the tail spectrum
df.tail()
```

Out[75]:

	id	imdb_id	popularity	budget	revenue	original_title	cast	director	overview	runtime	genres	production_companies	release_date
10861	21	tt0060371	0.080598	0	0	The Endless Summer	Michael Hynson|Robert August|Lord 'Tally Ho' B...	Bruce Brown	The Endless Summer, by Bruce Brown, is one of ...	95	Documentary	Bruce Brown Films	2066-06-15
10862	20379	tt0060472	0.065543	0	0	Grand Prix	James Garner|Eva Marie Saint|Yves Montand|Tosh...	John Frankenheimer	Grand Prix driver Pete Aron is fired by his te...	176	Action|Adventure|Drama	Cherokee Productions|Joel Productions|Douglas ...	2066-12-21
10863	39768	tt0060161	0.065141	0	0	Beregis Avtomobilya	Innokentiy Smoktunovskiy|Oleg Efremov|Georgi Z...	Eldar Ryazanov	An insurance agent who moonlights as a carthie...	94	Mystery|Comedy		
10864	21449	tt0061177	0.064317	0	0	What's Up, Tiger Lily?	Tatsuya Mihashi|Akiko Wakabayashi|Mie Hama|Joh...	Woody Allen	In comic Woody Allen's film debut, he took the...	80	Action|Comedy		
10865	22293	tt0060666	0.035919	19000	0	Manos: The Hands of Fate	Harold P. Warren|Tom Neyman|John Reynolds|Dian...	Harold P. Warren	A family gets lost on the road and stumbles up...	74	Horror		

```
10863                                          Mosfilm   2066-01-01
10864                          Benedict Pictures Corp.   2066-11-02
10865                                       Norm-Iris    2066-11-15

       vote_count  vote_average  release_year    budget_adj   revenue_adj  \
10861          11           7.4          1966  2.272271e+07  4.392749e+07
10862          20           5.7          1966  2.272271e+07  4.392749e+07
10863          11           6.5          1966  2.272271e+07  4.392749e+07
10864          22           5.4          1966  2.272271e+07  4.392749e+07
10865          15           1.5          1966  1.276423e+05  4.392749e+07

       adj_gross_profit
10861      2.120478e+07
10862      2.120478e+07
10863      2.120478e+07
10864      2.120478e+07
10865      4.379984e+07
```

In [76]:
```
# Checking for missing values and entry count
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 9771 entries, 0 to 10865
Data columns (total 19 columns):
id                      9771 non-null int64
imdb_id                 9771 non-null object
popularity              9771 non-null float64
budget                  9771 non-null int64
revenue                 9771 non-null int64
original_title          9771 non-null object
cast                    9771 non-null object
director                9771 non-null object
overview                9771 non-null object
runtime                 9771 non-null int64
genres                  9771 non-null object
production_companies    9771 non-null object
release_date            9771 non-null datetime64[ns]
vote_count              9771 non-null int64
vote_average            9771 non-null float64
release_year            9771 non-null int64
budget_adj              9771 non-null float64
revenue_adj             9771 non-null float64
adj_gross_profit        9771 non-null float64
dtypes: datetime64[ns](1), float64(5), int64(6), object(7)
memory usage: 1.5+ MB
```

This is to see how many rows we lost from deleting out the variables and dropping rows that are unnecessary for this analysis

```
In [77]: # Check for duplicate values and drop
         sum(df.duplicated())

Out[77]: 1

In [78]: df.drop_duplicates(inplace=True)

In [79]: sum(df.duplicated())

Out[79]: 0

In [80]: # Create a 2015 dataframe
         df_2015 = df.query('release_year == 2015')

In [81]: df_2015.info()

<class 'pandas.core.frame.DataFrame'>
Int64Index: 558 entries, 0 to 625
Data columns (total 19 columns):
id                      558 non-null int64
imdb_id                 558 non-null object
popularity              558 non-null float64
budget                  558 non-null int64
revenue                 558 non-null int64
original_title          558 non-null object
cast                    558 non-null object
director                558 non-null object
overview                558 non-null object
runtime                 558 non-null int64
genres                  558 non-null object
production_companies    558 non-null object
release_date            558 non-null datetime64[ns]
vote_count              558 non-null int64
vote_average            558 non-null float64
release_year            558 non-null int64
budget_adj              558 non-null float64
revenue_adj             558 non-null float64
adj_gross_profit        558 non-null float64
dtypes: datetime64[ns](1), float64(5), int64(6), object(7)
memory usage: 87.2+ KB
```

We made a 2015 data set to answer a question
Exploratory Data Analysis

1.1.2 What was the average profit a movie made per year in 80s - 21st century?

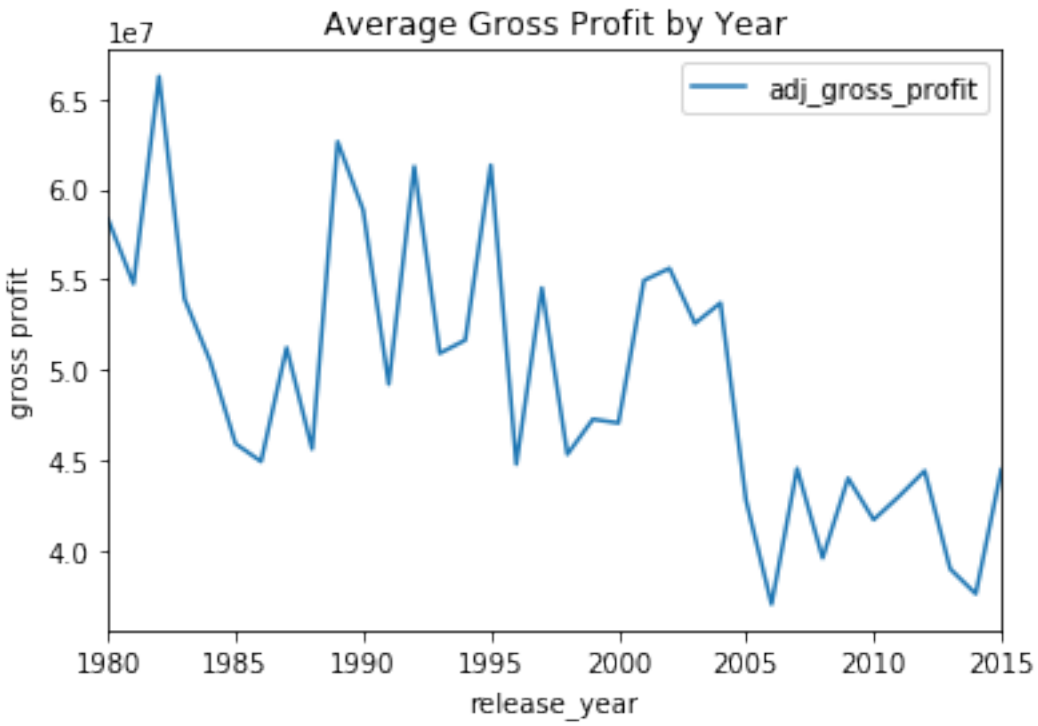
```
In [82]: # Create a 1980s-2015 dataframe
         df_80s =  df.query('release_year >= 1980')
```

```
In [83]: df_mean_profit = df_80s.groupby('release_year')['adj_gross_profit'].mean().reset_index(
         df_mean_profit.head()
```

Out[83]:

	release_year	adj_gross_profit
0	1980	5.835532e+07
1	1981	5.476850e+07
2	1982	6.625652e+07
3	1983	5.392917e+07
4	1984	5.050073e+07

```
In [84]: df_mean_profit.plot('release_year', 'adj_gross_profit', title = 'Average Gross Profit b
         plt.ylabel('gross profit');
```



Based on this table we can see year profits year over year fluctuate, but has been trending down.

1.1.3 What was the highest trending genres in the year 2015?

```
In [85]: # Found on stack overflow so we can segment this dataframe and run the split column met
         a = df['genres']
         df_2015['count_delimiter'] = a.str.count('\|')

         df_2015['count_delimiter'].describe()
```

```
/opt/conda/lib/python3.6/site-packages/ipykernel_launcher.py:3: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
```

```
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: http://pandas.pydata.org/pandas-docs/stable/indexing.html#
  This is separate from the ipykernel package so we can avoid doing imports until
```

```
Out[85]: count    558.000000
         mean       1.243728
         std        1.029941
         min        0.000000
         25%        0.000000
         50%        1.000000
         75%        2.000000
         max        4.000000
         Name: count_delimiter, dtype: float64
```

I used the functions above to see what the max and minimum number of '|' were in the genres column

```
In [86]: # Found on stackoverflow
         df_2015 = df_2015[df_2015['genres'].str.contains('|')]
         df_2015.head()
```

```
Out[86]:        id    imdb_id  popularity     budget     revenue  \
         0  135397  tt0369610   32.985763  150000000  1513528810
         1   76341  tt1392190   28.419936  150000000   378436354
         2  262500  tt2908446   13.112507  110000000   295238201
         3  140607  tt2488496   11.173104  200000000  2068178225
         4  168259  tt2820852    9.335014  190000000  1506249360

                         original_title  \
         0                Jurassic World
         1             Mad Max: Fury Road
         2                      Insurgent
         3  Star Wars: The Force Awakens
         4                      Furious 7

                                                  cast           director  \
         0  Chris Pratt|Bryce Dallas Howard|Irrfan Khan|Vi...   Colin Trevorrow
         1  Tom Hardy|Charlize Theron|Hugh Keays-Byrne|Nic...     George Miller
         2  Shailene Woodley|Theo James|Kate Winslet|Ansel...  Robert Schwentke
         3  Harrison Ford|Mark Hamill|Carrie Fisher|Adam D...       J.J. Abrams
         4  Vin Diesel|Paul Walker|Jason Statham|Michelle ...         James Wan

                                             overview  runtime  \
         0  Twenty-two years after the events of Jurassic ...      124
         1  An apocalyptic story set in the furthest reach...      120
         2  Beatrice Prior must confront her inner demons ...      119
         3  Thirty years after defeating the Galactic Empi...      136
```

```
4  Deckard Shaw seeks revenge against Dominic Tor...      137

                                       genres  \
0  Action|Adventure|Science Fiction|Thriller
1  Action|Adventure|Science Fiction|Thriller
2         Adventure|Science Fiction|Thriller
3   Action|Adventure|Science Fiction|Fantasy
4                      Action|Crime|Thriller

                                production_companies release_date  vote_count  \
0  Universal Studios|Amblin Entertainment|Legenda...   2015-06-09        5562
1  Village Roadshow Pictures|Kennedy Miller Produ...   2015-05-13        6185
2  Summit Entertainment|Mandeville Films|Red Wago...   2015-03-18        2480
3          Lucasfilm|Truenorth Productions|Bad Robot   2015-12-15        5292
4  Universal Pictures|Original Film|Media Rights ...   2015-04-01        2947

   vote_average  release_year    budget_adj   revenue_adj  adj_gross_profit  \
0           6.5          2015  1.379999e+08  1.392446e+09      1.254446e+09
1           7.1          2015  1.379999e+08  3.481613e+08      2.101614e+08
2           6.3          2015  1.012000e+08  2.716190e+08      1.704191e+08
3           7.5          2015  1.839999e+08  1.902723e+09      1.718723e+09
4           7.3          2015  1.747999e+08  1.385749e+09      1.210949e+09

   count_delimiter
0                3
1                3
2                2
3                3
4                2
```

In [87]:
```
# Created this function to break out the multiple genres into separate columns (like a
def separate(data):
    return data.str[0:].str.split('|',expand = True)

genretest = separate(df_2015['genres'])
genretest = genretest.rename(columns = {0: "genre1",
                                        1: "genre2",
                                        2: "genre3",
                                        3: "genre4",
                                        4: "genre5"})
genretest.head()
```

Out[87]:

	genre1	genre2	genre3	genre4	genre5
0	Action	Adventure	Science Fiction	Thriller	None
1	Action	Adventure	Science Fiction	Thriller	None
2	Adventure	Science Fiction	Thriller	None	None
3	Action	Adventure	Science Fiction	Fantasy	None
4	Action	Crime	Thriller	None	None

In [88]:
```
df_2015new = df_2015.join(genretest)
df_2015new.head()
```

Out[88]:

	id	imdb_id	popularity	budget	revenue	original_title	cast	director	overview	runtime	...	release_year	budget_adj	revenue_adj	adj_gross_profit	count_delimiter	genre1	genre2	genre3	genre4	genre5
0	135397	tt0369610	32.985763	150000000	1513528810	Jurassic World	Chris Pratt|Bryce Dallas Howard|Irrfan Khan|Vi...	Colin Trevorrow	Twenty-two years after the events of Jurassic ...	124	...	2015	1.379999e+08	1.392446e+09	1.254446e+09	3	Action	Adventure	Science Fiction	Thriller	None
1	76341	tt1392190	28.419936	150000000	378436354	Mad Max: Fury Road	Tom Hardy|Charlize Theron|Hugh Keays-Byrne|Nic...	George Miller	An apocalyptic story set in the furthest reach...	120	...	2015	1.379999e+08	3.481613e+08	2.101614e+08	3	Action	Adventure	Science Fiction	Thriller	None
2	262500	tt2908446	13.112507	110000000	295238201	Insurgent	Shailene Woodley|Theo James|Kate Winslet|Ansel...	Robert Schwentke	Beatrice Prior must confront her inner demons ...	119	...	2015	1.012000e+08	2.716190e+08	1.704191e+08	2	Adventure	Science Fiction	Thriller	None	None
3	140607	tt2488496	11.173104	200000000	2068178225	Star Wars: The Force Awakens	Harrison Ford|Mark Hamill|Carrie Fisher|Adam D...	J.J. Abrams	Thirty years after defeating the Galactic Empi...	136	...	2015	1.839999e+08	1.902723e+09	1.718723e+09	3	Action	Adventure	Science Fiction	Fantasy	None
4	168259	tt2820852	9.335014	190000000	1506249360	Furious 7	Vin Diesel|Paul Walker|Jason Statham|Michelle ...	James Wan	Deckard Shaw seeks revenge against Dominic Tor...	137	...	2015	1.747999e+08	1.385749e+09	1.210949e+09	2	Action	Crime	Thriller	None	None

[5 rows x 25 columns]

In [89]:
```
#Created new dataframes to append the genre list on the new column Main_Genre then drop
df1 = df_2015new.drop(['genre2','genre3','genre4','genre5'], 1)
```

```
df2 = df_2015new.drop(['genre1','genre3','genre4','genre5'], 1)
df3 = df_2015new.drop(['genre1','genre2','genre4','genre5'], 1)
df4 = df_2015new.drop(['genre1','genre2','genre3','genre5'], 1)
df5 = df_2015new.drop(['genre1','genre2','genre3','genre4'], 1)

df1 = df1.rename(columns = {"genre1": "Main_Genre"})
df2 = df2.rename(columns = {"genre2": "Main_Genre"})
df3 = df3.rename(columns = {"genre3": "Main_Genre"})
df4 = df4.rename(columns = {"genre4": "Main_Genre"})
df5 = df5.rename(columns = {"genre5": "Main_Genre"})

newrow = df1.append(df2)
newrow = newrow.append(df3)
newrow = newrow.append(df4)
newrow = newrow.append(df5)

newrow = newrow.dropna()
newrow.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 1252 entries, 0 to 216
Data columns (total 21 columns):
id                      1252 non-null int64
imdb_id                 1252 non-null object
popularity              1252 non-null float64
budget                  1252 non-null int64
revenue                 1252 non-null int64
original_title          1252 non-null object
cast                    1252 non-null object
director                1252 non-null object
overview                1252 non-null object
runtime                 1252 non-null int64
genres                  1252 non-null object
production_companies    1252 non-null object
release_date            1252 non-null datetime64[ns]
vote_count              1252 non-null int64
vote_average            1252 non-null float64
release_year            1252 non-null int64
budget_adj              1252 non-null float64
revenue_adj             1252 non-null float64
adj_gross_profit        1252 non-null float64
count_delimiter         1252 non-null int64
Main_Genre              1252 non-null object
dtypes: datetime64[ns](1), float64(5), int64(7), object(8)
memory usage: 215.2+ KB
```

After we appended all the rows we can see that we have 1252 data points for our genre analysis

```
In [90]: newrow.sort_values(by=['id'])
```

Out[90]:

	id	imdb_id	popularity	budget	revenue	\
186	10317	tt1018765	0.795762	28000000	7002261	
186	10317	tt1018765	0.795762	28000000	7002261	
325	57876	tt1519461	0.369529	5000000	0	
325	57876	tt1519461	0.369529	5000000	0	
325	57876	tt1519461	0.369529	5000000	0	
488	75861	tt1850418	0.253949	3400000	0	
488	75861	tt1850418	0.253949	3400000	0	
1	76341	tt1392190	28.419936	150000000	378436354	
1	76341	tt1392190	28.419936	150000000	378436354	
1	76341	tt1392190	28.419936	150000000	378436354	
1	76341	tt1392190	28.419936	150000000	378436354	
11	76757	tt1617661	6.189369	176000003	183987723	
11	76757	tt1617661	6.189369	176000003	183987723	
11	76757	tt1617661	6.189369	176000003	183987723	
11	76757	tt1617661	6.189369	176000003	183987723	
357	79698	tt1321869	0.329109	27000000	0	
357	79698	tt1321869	0.329109	27000000	0	
357	79698	tt1321869	0.329109	27000000	0	
357	79698	tt1321869	0.329109	27000000	0	
119	86828	tt1727770	1.360827	0	5189646	
119	86828	tt1727770	1.360827	0	5189646	
212	86835	tt2101383	0.654335	0	0	
212	86835	tt2101383	0.654335	0	0	
6	87101	tt1340138	8.654359	155000000	440603537	
6	87101	tt1340138	8.654359	155000000	440603537	
6	87101	tt1340138	8.654359	155000000	440603537	
6	87101	tt1340138	8.654359	155000000	440603537	
313	94365	tt1883367	0.386529	0	0	
14	99861	tt2395427	5.944927	280000000	1405035767	
14	99861	tt2395427	5.944927	280000000	1405035767	
..	...	...	...	...	...	
625	367735	tt5069564	0.017050	0	0	
400	368256	tt3619102	0.272995	0	0	
92	370687	tt3608646	1.876037	0	0	
92	370687	tt3608646	1.876037	0	0	
92	370687	tt3608646	1.876037	0	0	
430	371442	tt3482062	0.239558	0	0	
540	371758	tt3581932	0.112284	0	0	
540	371758	tt3581932	0.112284	0	0	
553	371759	tt4701546	0.068741	0	0	
553	371759	tt4701546	0.068741	0	0	
553	371759	tt4701546	0.068741	0	0	
569	371833	tt4974584	0.088764	0	0	
569	371833	tt4974584	0.088764	0	0	
455	373558	tt5297750	0.209036	0	0	

```
455  373558  tt5297750    0.209036          0           0
143  378373  tt3532278    1.128081          0           0
143  378373  tt3532278    1.128081          0           0
143  378373  tt3532278    1.128081          0           0
190  382517  tt4938374    0.773909          0           0
190  382517  tt4938374    0.773909          0           0
190  382517  tt4938374    0.773909          0           0
190  382517  tt4938374    0.773909          0           0
458  386501  tt3881680    0.207366          0           0
458  386501  tt3881680    0.207366          0           0
515  395560  tt3108244    0.142759    1300000           0
515  395560  tt3108244    0.142759    1300000           0
515  395560  tt3108244    0.142759    1300000           0
515  395560  tt3108244    0.142759    1300000           0
352  395883  tt4073890    0.333656          0           0
352  395883  tt4073890    0.333656          0           0

                                 original_title  \
186                         Our Brand Is Crisis
186                         Our Brand Is Crisis
325                                     Area 51
325                                     Area 51
325                                     Area 51
488                  To Write Love on Her Arms
488                  To Write Love on Her Arms
1                           Mad Max: Fury Road
1                           Mad Max: Fury Road
1                           Mad Max: Fury Road
1                           Mad Max: Fury Road
11                           Jupiter Ascending
11                           Jupiter Ascending
11                           Jupiter Ascending
11                           Jupiter Ascending
357                                 The Lovers
357                                 The Lovers
357                                 The Lovers
357                                 The Lovers
119                         Absolutely Anything
119                         Absolutely Anything
212                              Knight of Cups
212                              Knight of Cups
6                            Terminator Genisys
6                            Terminator Genisys
6                            Terminator Genisys
6                            Terminator Genisys
313    The Human Centipede 3 (Final Sequence)
14                       Avengers: Age of Ultron
14                       Avengers: Age of Ultron
```

```
..                                        ...
625            John Mulaney: The Comeback Kid
400                                  Condemned
92                    Mythica: The Necromancer
92                    Mythica: The Necromancer
92                    Mythica: The Necromancer
430                                Bad Roomies
540                   And Then There Were None
540                   And Then There Were None
553    Doctor Who: The Husbands of River Song
553    Doctor Who: The Husbands of River Song
553    Doctor Who: The Husbands of River Song
569                  Harry Price: Ghost Hunter
569                  Harry Price: Ghost Hunter
455  Taylor Swift: The 1989 World Tour - Live
455  Taylor Swift: The 1989 World Tour - Live
143                       Brothers of the Wind
143                       Brothers of the Wind
143                       Brothers of the Wind
190                  Open Season: Scared Silly
190                  Open Season: Scared Silly
190                  Open Season: Scared Silly
190                  Open Season: Scared Silly
458                               Waffle Street
458                               Waffle Street
515                                     Capsule
515                                     Capsule
515                                     Capsule
515                                     Capsule
352                                      Andron
352                                      Andron

                                                    cast  \
186  Sandra Bullock|Anthony Mackie|Billy Bob Thornt...
186  Sandra Bullock|Anthony Mackie|Billy Bob Thornt...
325  Reid Warner|Darrin Bragg|Ben Rovner|Jelena Nik...
325  Reid Warner|Darrin Bragg|Ben Rovner|Jelena Nik...
325  Reid Warner|Darrin Bragg|Ben Rovner|Jelena Nik...
488  Kat Dennings|Chad Michael Murray|Rupert Friend...
488  Kat Dennings|Chad Michael Murray|Rupert Friend...
1    Tom Hardy|Charlize Theron|Hugh Keays-Byrne|Nic...
1    Tom Hardy|Charlize Theron|Hugh Keays-Byrne|Nic...
1    Tom Hardy|Charlize Theron|Hugh Keays-Byrne|Nic...
1    Tom Hardy|Charlize Theron|Hugh Keays-Byrne|Nic...
11   Mila Kunis|Channing Tatum|Sean Bean|Eddie Redm...
11   Mila Kunis|Channing Tatum|Sean Bean|Eddie Redm...
11   Mila Kunis|Channing Tatum|Sean Bean|Eddie Redm...
11   Mila Kunis|Channing Tatum|Sean Bean|Eddie Redm...
```

```
357  Josh Hartnett|Simone Kessell|Tamsin Egerton|Al...
357  Josh Hartnett|Simone Kessell|Tamsin Egerton|Al...
357  Josh Hartnett|Simone Kessell|Tamsin Egerton|Al...
357  Josh Hartnett|Simone Kessell|Tamsin Egerton|Al...
119  Simon Pegg|Kate Beckinsale|Rob Riggle|Sanjeev ...
119  Simon Pegg|Kate Beckinsale|Rob Riggle|Sanjeev ...
212  Christian Bale|Cate Blanchett|Natalie Portman|...
212  Christian Bale|Cate Blanchett|Natalie Portman|...
6    Arnold Schwarzenegger|Jason Clarke|Emilia Clar...
6    Arnold Schwarzenegger|Jason Clarke|Emilia Clar...
6    Arnold Schwarzenegger|Jason Clarke|Emilia Clar...
6    Arnold Schwarzenegger|Jason Clarke|Emilia Clar...
313  Dieter Laser|Laurence R. Harvey|Robert LaSardo...
14   Robert Downey Jr.|Chris Hemsworth|Mark Ruffalo...
14   Robert Downey Jr.|Chris Hemsworth|Mark Ruffalo...
..                                                 ...
625                                      John Mulaney
400  Johnny Messner|Michael DeMello|Jordan Gelber|T...
92   Melanie Stone|Adam Johnson|Kevin Sorbo|Nicola ...
92   Melanie Stone|Adam Johnson|Kevin Sorbo|Nicola ...
92   Melanie Stone|Adam Johnson|Kevin Sorbo|Nicola ...
430  Patrick Renna|Tommy Savas|Annie Monroe|Jackie ...
540  Maeve Dermody|Aidan Turner|Charles Dance|Dougl...
540  Maeve Dermody|Aidan Turner|Charles Dance|Dougl...
553  Peter Capaldi|Alex Kingston|Greg Davies|Matt L...
553  Peter Capaldi|Alex Kingston|Greg Davies|Matt L...
553  Peter Capaldi|Alex Kingston|Greg Davies|Matt L...
569  Rafe Spall|Cara Theobold|Tom Ward|Sophie Stant...
569  Rafe Spall|Cara Theobold|Tom Ward|Sophie Stant...
455  Taylor Swift|Ellen DeGeneres|Cara Delevingne|L...
455  Taylor Swift|Ellen DeGeneres|Cara Delevingne|L...
143   Manuel Camacho|Jean Reno|Tobias Moretti|Eva Kuen
143   Manuel Camacho|Jean Reno|Tobias Moretti|Eva Kuen
143   Manuel Camacho|Jean Reno|Tobias Moretti|Eva Kuen
190  Garry Chalk|Kathleen Barr|Willa Townsend|Melis...
190  Garry Chalk|Kathleen Barr|Willa Townsend|Melis...
190  Garry Chalk|Kathleen Barr|Willa Townsend|Melis...
190  Garry Chalk|Kathleen Barr|Willa Townsend|Melis...
458  James Lafferty|Danny Glover|Julie Gonzalo|Dale...
458  James Lafferty|Danny Glover|Julie Gonzalo|Dale...
515  Edmund Kingsley|David Wayman|Nigel Barber|Lisa...
515  Edmund Kingsley|David Wayman|Nigel Barber|Lisa...
515  Edmund Kingsley|David Wayman|Nigel Barber|Lisa...
515  Edmund Kingsley|David Wayman|Nigel Barber|Lisa...
352  Alec Baldwin|Michelle Ryan|Danny Glover|Gale H...
352  Alec Baldwin|Michelle Ryan|Danny Glover|Gale H...

                          director  \
```

```
186              David Gordon Green
186              David Gordon Green
325                       Oren Peli
325                       Oren Peli
325                       Oren Peli
488               Nathan Frankowski
488               Nathan Frankowski
1                     George Miller
1                     George Miller
1                     George Miller
1                     George Miller
11   Lana Wachowski|Lilly Wachowski
11   Lana Wachowski|Lilly Wachowski
11   Lana Wachowski|Lilly Wachowski
11   Lana Wachowski|Lilly Wachowski
357                   Roland JoffÃľ
357                   Roland JoffÃľ
357                   Roland JoffÃľ
357                   Roland JoffÃľ
119                     Terry Jones
119                     Terry Jones
212                 Terrence Malick
212                 Terrence Malick
6                     Alan Taylor
6                     Alan Taylor
6                     Alan Taylor
6                     Alan Taylor
313                       Tom Six
14                    Joss Whedon
14                    Joss Whedon
..                             ...
625                     Rhys Thomas
400               Eli Morgan Gesner
92                    A. Todd Smith
92                    A. Todd Smith
92                    A. Todd Smith
430                   Jason Schnell
540                  Craig Viveiros
540                  Craig Viveiros
553               Douglas Mackinnon
553               Douglas Mackinnon
553               Douglas Mackinnon
569                     Alex Pillai
569                     Alex Pillai
455                 Jonas Ãkerlund
455                 Jonas Ãkerlund
143    Gerado Olivares|Otmar Penker
143    Gerado Olivares|Otmar Penker
```

```
143    Gerado Olivares|Otmar Penker
190                      David Feiss
190                      David Feiss
190                      David Feiss
190                      David Feiss
458            Eshom Nelms|Ian Nelms
458            Eshom Nelms|Ian Nelms
515                    Andrew Martin
515                    Andrew Martin
515                    Andrew Martin
515                    Andrew Martin
352             Francesco Cinquemani
352             Francesco Cinquemani

                                                  overview  runtime  \
186  A feature film based on the documentary "Our B...       108
186  A feature film based on the documentary "Our B...       108
325  Three young conspiracy theorists attempt to un...        91
325  Three young conspiracy theorists attempt to un...        91
325  Three young conspiracy theorists attempt to un...        91
488  The story follows 19-year-old Renee who has al...       118
488  The story follows 19-year-old Renee who has al...       118
1    An apocalyptic story set in the furthest reach...       120
1    An apocalyptic story set in the furthest reach...       120
1    An apocalyptic story set in the furthest reach...       120
1    An apocalyptic story set in the furthest reach...       120
11   In a universe where human genetic material is ...       124
11   In a universe where human genetic material is ...       124
11   In a universe where human genetic material is ...       124
11   In a universe where human genetic material is ...       124
357  The Lovers is an epic romance time travel adve...       109
357  The Lovers is an epic romance time travel adve...       109
357  The Lovers is an epic romance time travel adve...       109
357  The Lovers is an epic romance time travel adve...       109
119  Eccentric aliens give a man the power to do an...        85
119  Eccentric aliens give a man the power to do an...        85
212  Once there was a young prince whose father, th...       118
212  Once there was a young prince whose father, th...       118
6    The year is 2029. John Connor, leader of the r...       125
6    The year is 2029. John Connor, leader of the r...       125
6    The year is 2029. John Connor, leader of the r...       125
6    The year is 2029. John Connor, leader of the r...       125
313  Taking inspiration from The Human Centipede fi...       103
14   When Tony Stark tries to jumpstart a dormant p...       141
14   When Tony Stark tries to jumpstart a dormant p...       141
..                                                 ...       ...
625  Armed with boyish charm and a sharp wit, the f...        62
400  Fed up with her parents' bickering, poor-littl...        83
```

```
92   Mallister takes Thane prisoner and forces Mare...        0
92   Mallister takes Thane prisoner and forces Mare...        0
92   Mallister takes Thane prisoner and forces Mare...        0
430  Two guys find a beautiful young woman to take ...       93
540  Ten strangers, drawn away from their normal li...      168
540  Ten strangers, drawn away from their normal li...      168
553  Itâs Christmas Day on a remote human colony ...       60
553  Itâs Christmas Day on a remote human colony ...       60
553  Itâs Christmas Day on a remote human colony ...       60
569  When MP's wife Grace Goodwin is found naked on...       90
569  When MP's wife Grace Goodwin is found naked on...       90
455  Taylor delivers the concert event of the year...      132
455  Taylor delivers the concert event of the year...      132
143  The way of the eagle is to raise two chicks. T...       98
143  The way of the eagle is to raise two chicks. T...       98
143  The way of the eagle is to raise two chicks. T...       98
190  The humans and animals believe a werewolf is o...       85
190  The humans and animals believe a werewolf is o...       85
190  The humans and animals believe a werewolf is o...       85
190  The humans and animals believe a werewolf is o...       85
458  WAFFLE STREET is the true story of Jimmy Adams...       86
458  WAFFLE STREET is the true story of Jimmy Adams...       86
515  Guy is an experienced British fighter pilot wh...       91
515  Guy is an experienced British fighter pilot wh...       91
515  Guy is an experienced British fighter pilot wh...       91
515  Guy is an experienced British fighter pilot wh...       91
352  A group of people are plunged into a dark, cla...      100
352  A group of people are plunged into a dark, cla...      100

          ...                                    production_companies  \
186       ...                    Participant Media|Smokehouse Pictures
186       ...                    Participant Media|Smokehouse Pictures
325       ...                                       Insurge Pictures
325       ...                                       Insurge Pictures
325       ...                                       Insurge Pictures
488       ...        Two Streets Entertainment|Birchwood Pictures|N...
488       ...        Two Streets Entertainment|Birchwood Pictures|N...
1         ...        Village Roadshow Pictures|Kennedy Miller Produ...
1         ...        Village Roadshow Pictures|Kennedy Miller Produ...
1         ...        Village Roadshow Pictures|Kennedy Miller Produ...
1         ...        Village Roadshow Pictures|Kennedy Miller Produ...
11        ...        Village Roadshow Pictures|Dune Entertainment|A...
11        ...        Village Roadshow Pictures|Dune Entertainment|A...
11        ...        Village Roadshow Pictures|Dune Entertainment|A...
11        ...        Village Roadshow Pictures|Dune Entertainment|A...
357       ...        Corsan|Bliss Media|Limelight International Med...
357       ...        Corsan|Bliss Media|Limelight International Med...
357       ...        Corsan|Bliss Media|Limelight International Med...
```

```
357       ...          Corsan|Bliss Media|Limelight International Med...
119       ...          Premiere Picture|Bill and Ben Productions|GFM ...
119       ...          Premiere Picture|Bill and Ben Productions|GFM ...
212       ...                      Waypoint Entertainment|Dogwood Films
212       ...                      Waypoint Entertainment|Dogwood Films
6         ...                    Paramount Pictures|Skydance Productions
6         ...                    Paramount Pictures|Skydance Productions
6         ...                    Paramount Pictures|Skydance Productions
6         ...                    Paramount Pictures|Skydance Productions
313       ...                                         Six Entertainment
14        ...          Marvel Studios|Prime Focus|Revolution Sun Studios
14        ...          Marvel Studios|Prime Focus|Revolution Sun Studios
..        ...                                                       ...
625       ...                   3 Arts Entertainment|Irwin Entertainment
400       ...                                       Caliber Media Company
92        ...          Arrowstorm Entertainment|Camera 40 Productions...
92        ...          Arrowstorm Entertainment|Camera 40 Productions...
92        ...          Arrowstorm Entertainment|Camera 40 Productions...
430       ...                                          Eastside Films
540       ...          British Broadcasting Corporation (BBC)|Mammoth...
540       ...          British Broadcasting Corporation (BBC)|Mammoth...
553       ...                    British Broadcasting Corporation (BBC)
553       ...                    British Broadcasting Corporation (BBC)
553       ...                    British Broadcasting Corporation (BBC)
569       ...                                    Bentley Productions
569       ...                                    Bentley Productions
455       ...                                             Apple Inc.
455       ...                                             Apple Inc.
143       ...                               Terra Mater Factual Studios
143       ...                               Terra Mater Factual Studios
143       ...                               Terra Mater Factual Studios
190       ...                                   Sony Pictures Animation
190       ...                                   Sony Pictures Animation
190       ...                                   Sony Pictures Animation
190       ...                                   Sony Pictures Animation
458       ...                                      Side Gig Productions
458       ...                                      Side Gig Productions
515       ...            Ecaveo Capital Partners|Hermes Space Industries
515       ...            Ecaveo Capital Partners|Hermes Space Industries
515       ...            Ecaveo Capital Partners|Hermes Space Industries
515       ...            Ecaveo Capital Partners|Hermes Space Industries
352       ...                                           Ambi Pictures
352       ...                                           Ambi Pictures

    release_date vote_count  vote_average  release_year    budget_adj  \
186   2015-09-11        122           5.7          2015  2.575999e+07
186   2015-09-11        122           5.7          2015  2.575999e+07
325   2015-05-15         82           4.4          2015  4.599998e+06
```

325	2015-05-15	82	4.4	2015	4.599998e+06
325	2015-05-15	82	4.4	2015	4.599998e+06
488	2015-03-13	32	6.9	2015	3.127999e+06
488	2015-03-13	32	6.9	2015	3.127999e+06
1	2015-05-13	6185	7.1	2015	1.379999e+08
1	2015-05-13	6185	7.1	2015	1.379999e+08
1	2015-05-13	6185	7.1	2015	1.379999e+08
1	2015-05-13	6185	7.1	2015	1.379999e+08
11	2015-02-04	1937	5.2	2015	1.619199e+08
11	2015-02-04	1937	5.2	2015	1.619199e+08
11	2015-02-04	1937	5.2	2015	1.619199e+08
11	2015-02-04	1937	5.2	2015	1.619199e+08
357	2015-02-13	22	4.6	2015	2.483999e+07
357	2015-02-13	22	4.6	2015	2.483999e+07
357	2015-02-13	22	4.6	2015	2.483999e+07
357	2015-02-13	22	4.6	2015	2.483999e+07
119	2015-06-26	199	5.8	2015	2.272271e+07
119	2015-06-26	199	5.8	2015	2.272271e+07
212	2015-09-10	101	5.7	2015	2.272271e+07
212	2015-09-10	101	5.7	2015	2.272271e+07
6	2015-06-23	2598	5.8	2015	1.425999e+08
6	2015-06-23	2598	5.8	2015	1.425999e+08
6	2015-06-23	2598	5.8	2015	1.425999e+08
6	2015-06-23	2598	5.8	2015	1.425999e+08
313	2015-05-22	75	3.5	2015	2.272271e+07
14	2015-04-22	4304	7.4	2015	2.575999e+08
14	2015-04-22	4304	7.4	2015	2.575999e+08
..	...	...	...	...	...
625	2015-11-13	19	6.7	2015	2.272271e+07
400	2015-11-13	16	3.7	2015	2.272271e+07
92	2015-12-19	11	5.4	2015	2.272271e+07
92	2015-12-19	11	5.4	2015	2.272271e+07
92	2015-12-19	11	5.4	2015	2.272271e+07
430	2015-12-01	16	5.1	2015	2.272271e+07
540	2015-12-26	37	7.7	2015	2.272271e+07
540	2015-12-26	37	7.7	2015	2.272271e+07
553	2015-12-25	31	8.0	2015	2.272271e+07
553	2015-12-25	31	8.0	2015	2.272271e+07
553	2015-12-25	31	8.0	2015	2.272271e+07
569	2015-12-27	10	5.3	2015	2.272271e+07
569	2015-12-27	10	5.3	2015	2.272271e+07
455	2015-12-20	15	7.6	2015	2.272271e+07
455	2015-12-20	15	7.6	2015	2.272271e+07
143	2015-12-24	11	7.5	2015	2.272271e+07
143	2015-12-24	11	7.5	2015	2.272271e+07
143	2015-12-24	11	7.5	2015	2.272271e+07
190	2015-12-31	33	5.6	2015	2.272271e+07
190	2015-12-31	33	5.6	2015	2.272271e+07

190	2015-12-31	33	5.6	2015	2.272271e+07
190	2015-12-31	33	5.6	2015	2.272271e+07
458	2015-09-24	25	6.4	2015	2.272271e+07
458	2015-09-24	25	6.4	2015	2.272271e+07
515	2015-12-23	11	5.3	2015	1.195999e+06
515	2015-12-23	11	5.3	2015	1.195999e+06
515	2015-12-23	11	5.3	2015	1.195999e+06
515	2015-12-23	11	5.3	2015	1.195999e+06
352	2015-11-08	24	4.7	2015	2.272271e+07
352	2015-11-08	24	4.7	2015	2.272271e+07

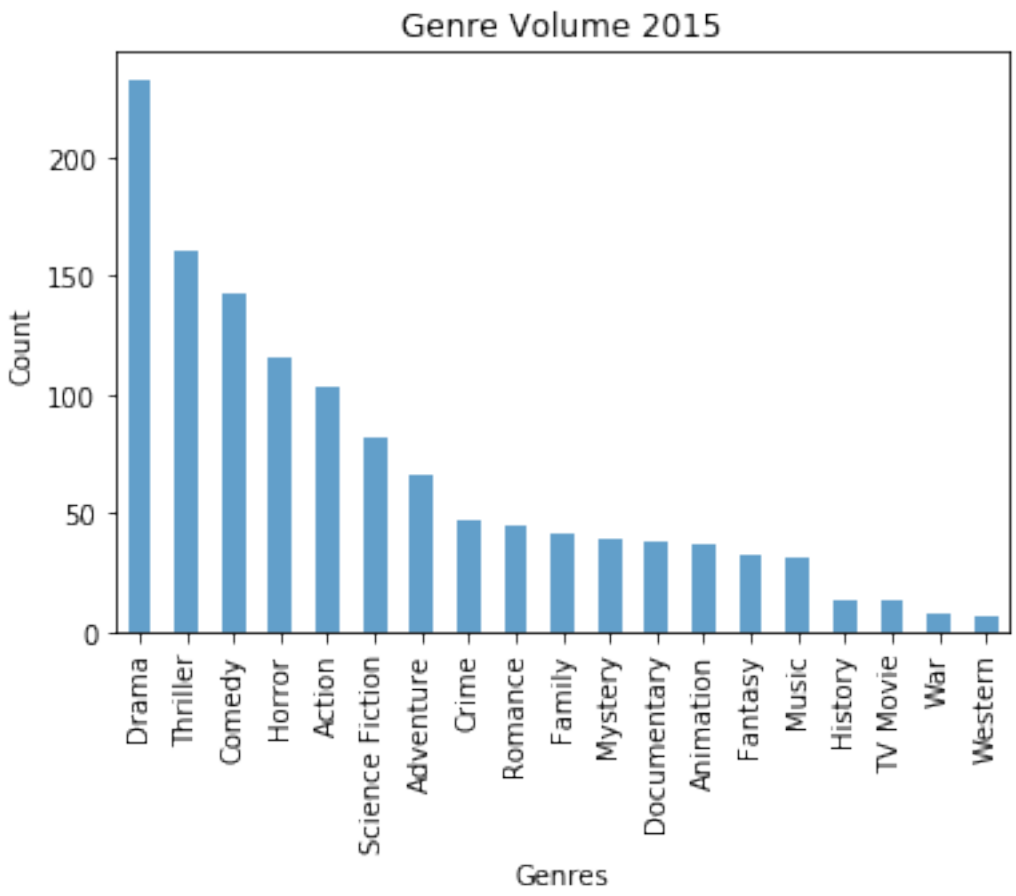
	revenue_adj	adj_gross_profit	count_delimiter	Main_Genre
186	6.442077e+06	-1.931791e+07	1	Comedy
186	6.442077e+06	-1.931791e+07	1	Drama
325	4.392749e+07	3.932749e+07	2	Science Fiction
325	4.392749e+07	3.932749e+07	2	Horror
325	4.392749e+07	3.932749e+07	2	Thriller
488	4.392749e+07	4.079949e+07	1	Drama
488	4.392749e+07	4.079949e+07	1	Music
1	3.481613e+08	2.101614e+08	3	Action
1	3.481613e+08	2.101614e+08	3	Science Fiction
1	3.481613e+08	2.101614e+08	3	Thriller
1	3.481613e+08	2.101614e+08	3	Adventure
11	1.692686e+08	7.348699e+06	3	Adventure
11	1.692686e+08	7.348699e+06	3	Fantasy
11	1.692686e+08	7.348699e+06	3	Action
11	1.692686e+08	7.348699e+06	3	Science Fiction
357	4.392749e+07	1.908750e+07	3	Romance
357	4.392749e+07	1.908750e+07	3	Science Fiction
357	4.392749e+07	1.908750e+07	3	Action
357	4.392749e+07	1.908750e+07	3	Adventure
119	4.774472e+06	-1.794824e+07	1	Science Fiction
119	4.774472e+06	-1.794824e+07	1	Comedy
212	4.392749e+07	2.120478e+07	1	Romance
212	4.392749e+07	2.120478e+07	1	Drama
6	4.053551e+08	2.627551e+08	3	Adventure
6	4.053551e+08	2.627551e+08	3	Science Fiction
6	4.053551e+08	2.627551e+08	3	Action
6	4.053551e+08	2.627551e+08	3	Thriller
313	4.392749e+07	2.120478e+07	0	Horror
14	1.292632e+09	1.035032e+09	2	Adventure
14	1.292632e+09	1.035032e+09	2	Action
..	...	...	...	...
625	4.392749e+07	2.120478e+07	0	Comedy
400	4.392749e+07	2.120478e+07	0	Horror
92	4.392749e+07	2.120478e+07	2	Adventure
92	4.392749e+07	2.120478e+07	2	Action
92	4.392749e+07	2.120478e+07	2	Fantasy

```
430  4.392749e+07      2.120478e+07                0           Comedy
540  4.392749e+07      2.120478e+07                1          Mystery
540  4.392749e+07      2.120478e+07                1            Drama
553  4.392749e+07      2.120478e+07                2            Drama
553  4.392749e+07      2.120478e+07                2  Science Fiction
553  4.392749e+07      2.120478e+07                2           Comedy
569  4.392749e+07      2.120478e+07                1           Horror
569  4.392749e+07      2.120478e+07                1            Drama
455  4.392749e+07      2.120478e+07                1      Documentary
455  4.392749e+07      2.120478e+07                1            Music
143  4.392749e+07      2.120478e+07                2            Drama
143  4.392749e+07      2.120478e+07                2        Adventure
143  4.392749e+07      2.120478e+07                2           Family
190  4.392749e+07      2.120478e+07                3        Adventure
190  4.392749e+07      2.120478e+07                3           Family
190  4.392749e+07      2.120478e+07                3        Animation
190  4.392749e+07      2.120478e+07                3           Comedy
458  4.392749e+07      2.120478e+07                1            Drama
458  4.392749e+07      2.120478e+07                1           Comedy
515  4.392749e+07      4.273149e+07                3            Drama
515  4.392749e+07      4.273149e+07                3  Science Fiction
515  4.392749e+07      4.273149e+07                3         Thriller
515  4.392749e+07      4.273149e+07                3          History
352  4.392749e+07      2.120478e+07                1  Science Fiction
352  4.392749e+07      2.120478e+07                1           Action

[1252 rows x 21 columns]
```

We sorted values here so that we can QA the rows and see that we have made duplicates of them based on the number of genres associated to them. This way we can bucket a movie into multiple genres.

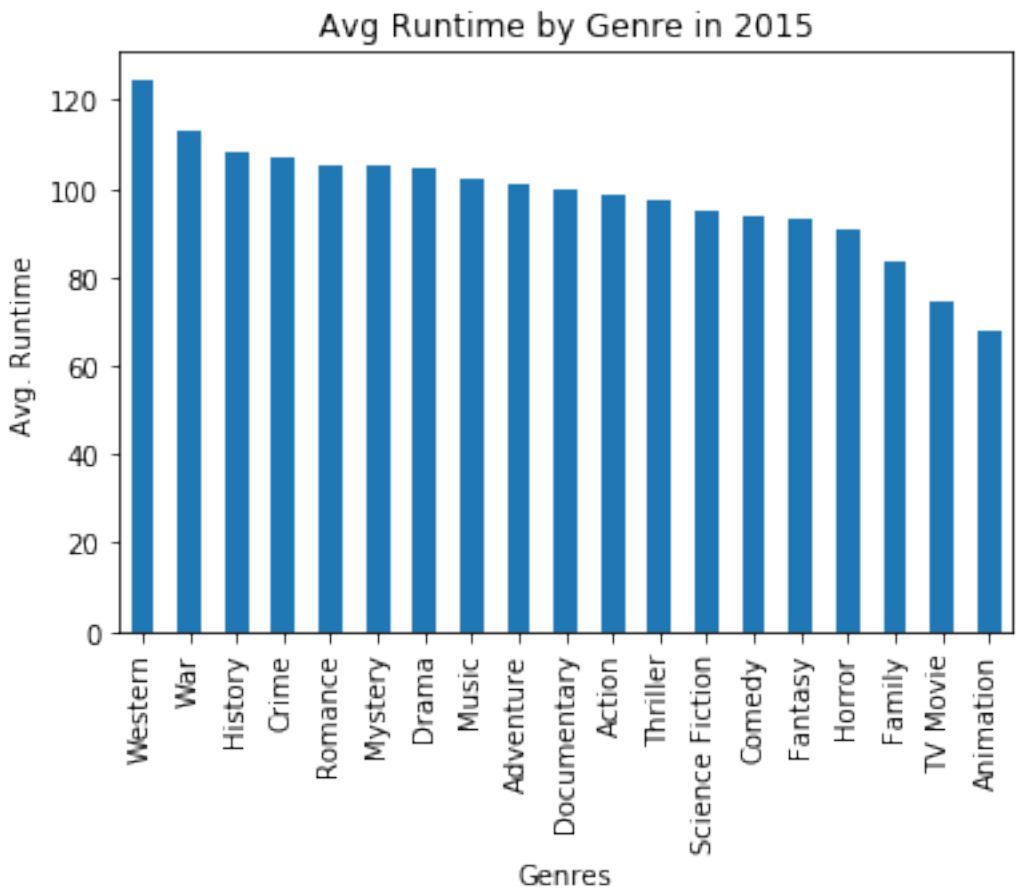
```
In [91]: # Look at volume of titles in each genre to see whats most popular
         newrow.pivot_table(index = 'Main_Genre', aggfunc = 'size').sort_values(ascending = Fals
         plt.xlabel('Genres')
         plt.ylabel('Count');
```



It looks as is Drama is the most popular movie genre in the year 2015. Western and War movies look to not be as popular in 2015

1.1.4 What was the average movie runtime per genre 2015?

```
In [92]: # Create bar chart to see what the average movie runtime is per genre in 2015
         newrow.groupby('Main_Genre')['runtime'].mean().sort_values(ascending = False).plot(kind
         plt.xlabel('Genres')
         plt.ylabel('Avg. Runtime');
```



Western movies look to have the longest average run time. Animations are the shortest.

```
In [93]: newrow.describe()
```

Out[93]:

	id	popularity	budget	revenue	runtime
count	1252.000000	1252.000000	1.252000e+03	1.252000e+03	1252.000000
mean	287705.307508	1.413758	1.838142e+07	6.561915e+07	97.511981
std	63449.604040	2.945139	4.198648e+07	2.166936e+08	23.472977
min	10317.000000	0.017050	0.000000e+00	0.000000e+00	0.000000
25%	257088.000000	0.265010	0.000000e+00	0.000000e+00	88.000000
50%	301728.000000	0.486351	0.000000e+00	0.000000e+00	96.000000
75%	330544.000000	1.380320	1.200000e+07	2.235457e+07	110.000000
max	395883.000000	32.985763	2.800000e+08	2.068178e+09	240.000000

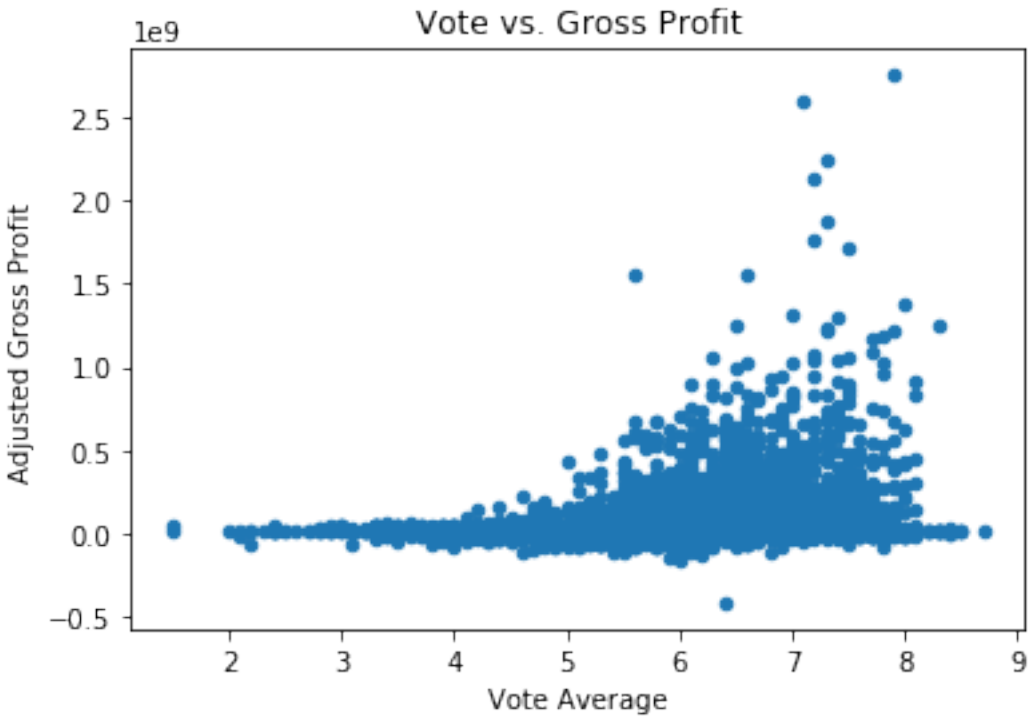
	vote_count	vote_average	release_year	budget_adj	revenue_adj
count	1252.000000	1252.000000	1252.0	1.252000e+03	1.252000e+03
mean	416.205272	5.857668	2015.0	3.028680e+07	8.626295e+07
std	885.974861	1.036850	0.0	3.412404e+07	1.925671e+08
min	10.000000	2.400000	2015.0	9.199996e+00	4.088478e+03
25%	22.000000	5.200000	2015.0	2.272271e+07	4.392749e+07

```
50%      69.000000      5.900000        2015.0  2.272271e+07  4.392749e+07
75%     341.000000      6.600000        2015.0  2.272271e+07  4.392749e+07
max    6185.000000      8.400000        2015.0  2.575999e+08  1.902723e+09

       adj_gross_profit  count_delimiter
count      1.252000e+03      1252.000000
mean       5.597615e+07         1.715655
std        1.684147e+08         1.029063
min       -4.806727e+07         0.000000
25%        2.120478e+07         1.000000
50%        2.120478e+07         2.000000
75%        2.648895e+07         2.000000
max        1.718723e+09         4.000000
```

1.1.5 Is there a relationship between Vote Rating and Adjusted Profit

```
In [94]: # Plot the vote average to adj gross profit
         df.plot('vote_average', 'adj_gross_profit', kind = 'scatter', title = 'Vote vs. Gross P
         plt.xlabel('Vote Average')
         plt.ylabel('Adjusted Gross Profit');
```



It looks like there seems to be some type of relationship between the vote_average and the amount of gross profit made on a movie. It would make sense that a movie rated higher would bring in more money since people are persuaded by reviews to see a movie.

2 Conclusion

2.1 Final Thoughts

1. Adjusted gross profit made on movies looks like it is trending down from 1980-2015
2. Dramas look to be the most popular movie genre in the year 2015
3. Western movies look to have the longest run time with greater than 120 minutes on average
4. There looks to be some type of relationship with the vote average vs the adjusted gross profit. We could potentially run some type of regression analysis to predict adj gross profit based on vote average

2.2 Limitations

1. As we continued to explore the data there were a lot of missing cells such as revenue and budget which had to be replaced by the medians of the variable
2. Casts and Genres had multiple categories associated in them which made it difficult to analyze without breaking them out into their own rows
3. Noticed some of the release date was formatted strangely which also makes it difficult to do any day over day analysis. If I did day over day analysis I would create a function to break out the dates separately by year day and month and then concatenate a new date with the correct year

```
In [95]: from subprocess import call
         call(['python', '-m', 'nbconvert', 'Investigate_a_Dataset.ipynb'])
```

```
Out[95]: 0
```

```
In [ ]:
```