> **Tip**: Welcome to the Investigate a Dataset project! You will find tips in quoted sections like this to help organize your approach to your investigation. Before submitting your project, it will be a good idea to go back through your report and remove these sections to make the presentation of your work as tidy as possible. First things first, you might want to double-click this Markdown cell and change the title so that it reflects your dataset and investigation.

Project: Investigate a Dataset (TMDb_Movies Dataset)

Table of Contents

Introduction

> I'm going to investigate the (TMDb_Movies Dataset) which I dowloaded from Kaggle web page.
>
> The dataset has information about 10,000 movies and consist of 21 columns such as popularity, budget, revenue, original_title, cast ...etc.
>
> I'm lookingforward to figure out which genres are most popular from year to year? and what kinds of properties are associated with movies that have high revenues?

In [357]:
```
# Use this cell to set up import statements for all of the packages that you
#   plan to use.

# 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
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import matplotlib.patches as mpatches
import seaborn as sns
from collections import Counter
%matplotlib inline
```

Data Wrangling

> **Tip**: In this section of the report, you will load in the data, check for cleanliness, and then trim and clean your dataset for analysis. Make sure that you document your steps carefully and justify your cleaning decisions.

General Properties

In [358]:
```
# Load tmdb-movies.csv dataset file
# Change release date column into date format.
tmdb = pd.read_csv('tmdb-movies.csv' , \
        parse_dates = ['release_date'])
tmdb.head(3)
```

Out[358]:

	id	imdb_id	popularity	budget	revenue	original_title	cast	
0	125336	tt2044056	0.006925	-	-	The Story of Film: An Odyssey	Mark Cousins|Jean-Michel Frodon|Cari Beauchamp...	http://www.cha
1	150004	tt0289830	0.469332	-	-	Taken	Dakota Fanning|Matt Frewer|Eric Close|Emily Be...	
2	331214	tt0185906	0.537593	-	125,000,000	Band of Brothers	Damian Lewis|Ron Livingston|Frank John Hughes|...	http://wv

3 rows × 21 columns

In [359]:
```
# Count number of rows and columns
tmdb.shape
```

Out[359]: (10866, 21)

In [360]:
```
tmdb.describe()
```

Out[360]:

	id	popularity	runtime	vote_count	vote_average	release_year	
count	10866.000000	10866.000000	10866.000000	10866.000000	10866.000000	10866.000000	1.0
mean	66064.177434	0.646441	102.070863	217.389748	5.974922	2001.322658	1.7
std	92130.136561	1.000185	31.381405	575.619058	0.935142	12.812941	3.4
min	5.000000	0.000065	0.000000	10.000000	1.500000	1960.000000	0.0
25%	10596.250000	0.207583	90.000000	17.000000	5.400000	1995.000000	0.0
50%	20669.000000	0.383856	99.000000	38.000000	6.000000	2006.000000	0.0
75%	75610.000000	0.713817	111.000000	145.750000	6.600000	2011.000000	2.0
max	417859.000000	32.985763	900.000000	9767.000000	9.200000	2015.000000	4.2

In [361]:
```
tmdb.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 10866 entries, 0 to 10865
Data columns (total 21 columns):
 #   Column                Non-Null Count  Dtype
---  ------                --------------  -----
 0   id                    10866 non-null  int64
 1   imdb_id               10856 non-null  object
 2   popularity            10866 non-null  float64
 3    budget               10866 non-null  object
 4    revenue              10866 non-null  object
 5   original_title        10866 non-null  object
 6   cast                  10790 non-null  object
 7   homepage              2936 non-null   object
 8   director              10822 non-null  object
 9   tagline               8042 non-null   object
 10  keywords              9373 non-null   object
 11  overview              10862 non-null  object
 12  runtime               10866 non-null  int64
 13  genres                10843 non-null  object
 14  production_companies  9836 non-null   object
 15  release_date          10866 non-null  datetime64[ns]
 16  vote_count            10866 non-null  int64
 17  vote_average          10866 non-null  float64
 18  release_year          10866 non-null  int64
 19  budget_adj            10866 non-null  float64
 20  revenue_adj           10866 non-null  float64
dtypes: datetime64[ns](1), float64(4), int64(4), object(12)
memory usage: 1.7+ MB
```

In [362]:
```
# Check duplicate data
duplicats = tmdb[tmdb.duplicated(keep='last')]
duplicats
```

Out[362]:

	id	imdb_id	popularity	budget	revenue	original_title	cast	homepage
9025	42194	tt0411951	0.59643	30,000,000	967,000	TEKKEN	Jon Foo|Kelly Overton|Cary-Hiroyuki Tagawa|Ian...	NaN

1 rows × 21 columns

In [363]:
```
# Count all zero value in each colunms
(tmdb == 0).sum()
```

Out[363]:
```
id                        0
imdb_id                   0
popularity                0
 budget                   0
 revenue                  0
original_title            0
cast                      0
homepage                  0
director                  0
tagline                   0
keywords                  0
overview                  0
runtime                  31
genres                    0
production_companies      0
release_date              0
vote_count                0
vote_average              0
release_year              0
budget_adj             5696
revenue_adj            6016
dtype: int64
```

In [364]:
```
# Count all null value in each colunms
tmdb.isnull().sum()
```

Out[364]:
```
id                         0
imdb_id                   10
popularity                 0
 budget                    0
 revenue                   0
original_title             0
cast                      76
homepage                7930
director                  44
tagline                 2824
keywords                1493
overview                   4
runtime                    0
genres                    23
production_companies    1030
release_date               0
vote_count                 0
vote_average               0
release_year               0
budget_adj                 0
revenue_adj                0
dtype: int64
```

> **Tip**: You should *not* perform too many operations in each cell. Create cells freely to explore your data. One option that you can take with this project is to do a lot of explorations in an initial notebook. These don't have to be organized, but make sure you use enough comments to understand the purpose of each code cell. Then, after you're done with your analysis, create a duplicate notebook where you will trim the excess and organize your steps so that you have a flowing, cohesive report.
>
> **Tip**: Make sure that you keep your reader informed on the steps that you are taking in your investigation. Follow every code cell, or every set of related code cells, with a markdown cell to describe to the reader what was found in the preceding cell(s). Try to make it so that the reader can then understand what they will be seeing in the following cell(s).

Data Cleaning (Delete unnecessary information)

> **From the data wrangling results I confiermed that there are a unesessary columns and duplicated data and there are some rows withe null data and others with zero budget, zero revenue and zero runtime, so these data need to be cleaned by doing the following steps:**
>
> 1. Delet unesessary columns which are (id, imdb_id, homepage, keywords, overview, production_companies, vote_count)
> 2. Delet duplicated rows.
> 3. Delet all rows which have zero value.
> 4. Replace zero with NAN value.

In [365]:
```
# Delete unesessary columns

del_columns=[ 'id','imdb_id', 'homepage', 'keywords', 'overview', 'production_
companies', 'vote_count']
tmdb_clean_columns= tmdb.drop(del_columns,1)

tmdb_clean_columns.head()
```

Out[365]:

	popularity	budget	revenue	original_title	cast	director	tagline
0	0.006925	-	-	The Story of Film: An Odyssey	Mark Cousins|Jean-Michel Frodon|Cari Beauchamp...	Mark Cousins	NaN
1	0.469332	-	-	Taken	Dakota Fanning|Matt Frewer|Eric Close|Emily Be...	Breck Eisner|FÃ©lix EnrÃquez AlcalÃ¡|John Faw...	Some secrets we keep. Some are kept from us
2	0.537593	-	125,000,000	Band of Brothers	Damian Lewis|Ron Livingston|Frank John Hughes|...	Phil Alden Robinson|Richard Loncraine|Mikael S...	Ordinary men. Extraordinary times.
3	0.147489	-	-	Shoah	Simon Srebnik|Michael Podchlebnik|Motke Zaidl	Claude Lanzmann	NaN
4	0.000065	-	-	North and South, Book I	Patrick Swayze|Philip Casnoff|Kirstie Alley|Ge...	NaN	NaN

In [366]:
```
tmdb.shape
```

Out[366]: (10866, 21)

In [367]:
```
tmdb_clean_columns.shape
```

Out[367]: (10866, 14)

In [368]:
```
# Delete duplicate rows
tmdb_clean_columns.drop_duplicates(keep='last', inplace=True)
tmdb_clean_columns.shape
```

Out[368]: (10865, 14)

In [369]:
```
tmdb_clean_columns.dtypes
```

Out[369]:
```
popularity                float64
 budget                    object
 revenue                   object
original_title             object
cast                       object
director                   object
tagline                    object
runtime                     int64
genres                     object
release_date       datetime64[ns]
vote_average              float64
release_year                int64
budget_adj                float64
revenue_adj               float64
dtype: object
```

In [370]:
```
(tmdb_clean_columns == 0).sum()
```

Out[370]:
```
popularity           0
 budget              0
 revenue             0
original_title       0
cast                 0
director             0
tagline              0
runtime             31
genres               0
release_date         0
vote_average         0
release_year         0
budget_adj        5696
revenue_adj       6016
dtype: int64
```

In [371]:
```
tmdb_clean_columns.isnull().sum()
```

Out[371]:
```
popularity           0
 budget              0
 revenue             0
original_title       0
cast                76
director            44
tagline           2824
runtime              0
genres              23
release_date         0
vote_average         0
release_year         0
budget_adj           0
revenue_adj          0
dtype: int64
```

In [372]:
```
# Delete all rows with zero values.

budget_revenue_runtime = [' budget ', ' revenue ', 'budget_adj', 'revenue_adj'
, 'runtime'  ]

# This will replace all zero values from '0' to NAN.
tmdb_clean_columns[budget_revenue_runtime]= tmdb_clean_columns[budget_revenue_
runtime].replace(0, np.NAN)


# Removing all row which has NaN value in budget_revenue
tmdb_clean_columns.dropna(subset = budget_revenue_runtime, inplace = True)

tmdb_clean_columns.shape
```

Out[372]:
```
(3854, 14)
```

In [373]:
```
(tmdb_clean_columns == 0).sum()
```

Out[373]:
```
popularity        0
 budget           0
 revenue          0
original_title    0
cast              0
director          0
tagline           0
runtime           0
genres            0
release_date      0
vote_average      0
release_year      0
budget_adj        0
revenue_adj       0
dtype: int64
```

In [374]:
```
tmdb_clean_columns=tmdb_clean_columns.fillna(" ")
```

In [375]:
```
tmdb_clean_columns.isnull().sum()
```

Out[375]:
```
popularity        0
 budget           0
 revenue          0
original_title    0
cast              0
director          0
tagline           0
runtime           0
genres            0
release_date      0
vote_average      0
release_year      0
budget_adj        0
revenue_adj       0
dtype: int64
```

In [376]:
```
tmdb_clean_columns
```

Out[376]:

	popularity	budget	revenue	original_title	cast	director	tag
20	9.432768	237,000,000	2,781,505,847	Avatar	Sam Worthington|Zoe Saldana|Sigourney Weaver|S...	James Cameron	Enter Worl Pand
37	11.173104	200,000,000	2,068,178,225	Star Wars: The Force Awakens	Harrison Ford|Mark Hamill|Carrie Fisher|Adam D...	J.J. Abrams	E\ genera has a st
59	4.355219	200,000,000	1,845,034,188	Titanic	Kate Winslet|Leonardo DiCaprio|Frances Fisher|...	James Cameron	Nothing Earth co co betw th
60	7.637767	220,000,000	1,519,557,910	The Avengers	Robert Downey Jr.|Chris Evans|Mark Ruffalo|Chr...	Joss Whedon	So assen requi
61	32.985763	150,000,000	1,513,528,810	Jurassic World	Chris Pratt|Bryce Dallas Howard|Irrfan Khan|Vi...	Colin Trevorrow	The par op
...	...	...	...	...	...	...	
10506	0.578849	15,000,000	5	Bordello of Blood	Dennis Miller|Erika Eleniak|Angie Everhart|Joh...	Gilbert Adler	
10511	0.208637	10	5	Kid's Story	Clayton Watson|Keanu Reeves|Carrie-Anne Moss|K...	Shinichiro Watanabe	
10607	0.352054	200,000	3	Dr. Horrible's Sing-Along Blog	Neil Patrick Harris|Nathan Fillion|Felicia Day...	Joss Whedon	He ha Ph.D horriblene
10642	0.462609	6,000,000	2	Shattered Glass	Hayden Christensen|Peter Sarsgaard|ChloÃ« Sevi...	Billy Ray	
10696	0.552091	6,000,000	2	Mallrats	Jason Lee|Jeremy London|Shannen Doherty|Claire...	Kevin Smith	They're ther sh They're there

3854 rows × 14 columns

Exploratory Data Analysis

Questions that can analysed from this data set:

1. Find the top 10 Highest Runtime Movies?
2. Find the top 10 Highest Revenues Movies?
3. Find the top 10 Highest Budgets Movies?
4. Find the top 10 Highest Rating Movies?
5. Find the top 10 Highest Net Profits Movies?

In [377]:
```
# Change the dtype fo the fields budget and revenue from object to float64

tmdb_clean_columns[[' budget ', ' revenue ']] = tmdb_clean_columns[[' budget '
, ' revenue ']].apply(pd.to_numeric, errors='coerce')
```

In [378]:
```
tmdb_clean_columns.dtypes
```

Out[378]:
```
popularity               float64
 budget                  float64
 revenue                 float64
original_title            object
cast                      object
director                  object
tagline                   object
runtime                  float64
genres                    object
release_date      datetime64[ns]
vote_average             float64
release_year               int64
budget_adj               float64
revenue_adj              float64
dtype: object
```

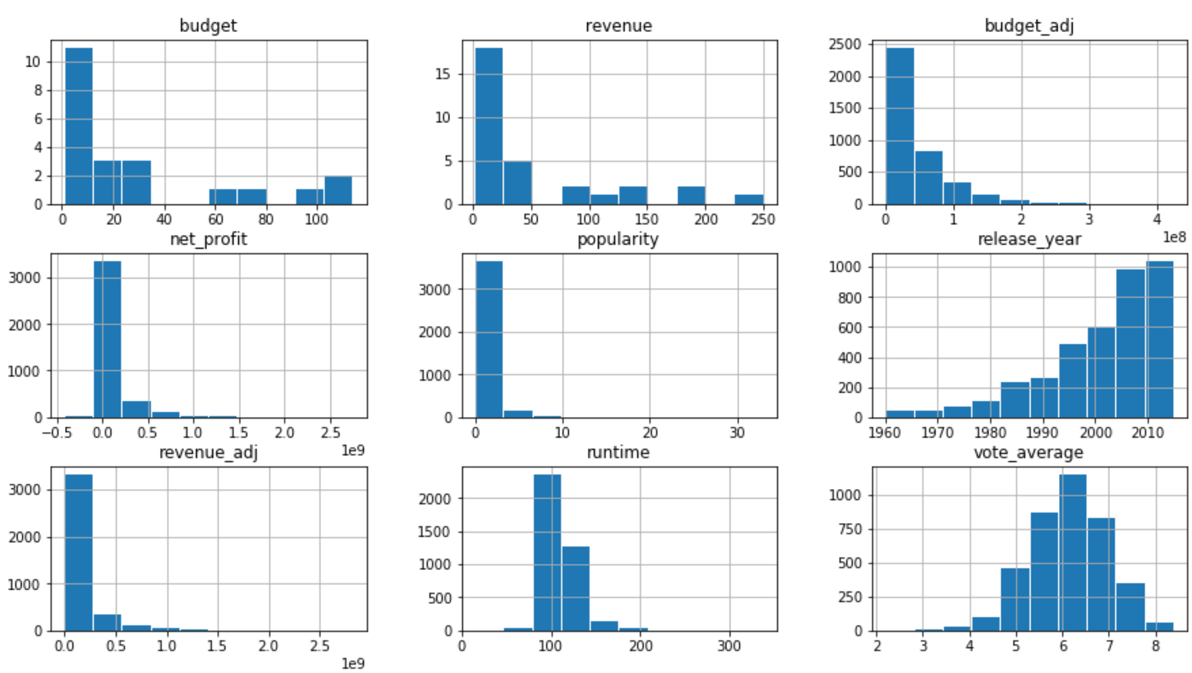
In [379]:
```
# Add new column for the Net Profit of the each movie

tmdb_clean_columns.insert(2,'net_profit',tmdb_clean_columns['revenue_adj'] - t
mdb_clean_columns['budget_adj'])
```

In [380]:
```
tmdb_clean_columns.head()
```

Out[380]:

	popularity	budget	net_profit	revenue	original_title	cast	director	tag
20	9.432768	NaN	2.586237e+09	NaN	Avatar	Sam Worthington|Zoe Saldana|Sigourney Weaver|S...	James Cameron	Enter Worl Pand
37	11.173104	NaN	1.718723e+09	NaN	Star Wars: The Force Awakens	Harrison Ford|Mark Hamill|Carrie Fisher|Adam D...	J.J. Abrams	Ev genera ha st
59	4.355219	NaN	2.234714e+09	NaN	Titanic	Kate Winslet|Leonardo DiCaprio|Frances Fisher|...	James Cameron	Notl on E co co betw th
60	7.637767	NaN	1.234248e+09	NaN	The Avengers	Robert Downey Jr.|Chris Evans|Mark Ruffalo|Chr...	Joss Whedon	Sc assen requi
61	32.985763	NaN	1.254446e+09	NaN	Jurassic World	Chris Pratt|Bryce Dallas Howard|Irrfan Khan|Vi...	Colin Trevorrow	The p is op

In [381]:
```
#tmdb_clean_columns.hist(figsize=(15,8));
tmdb_clean_columns.hist (bins = 10, rwidth = 0.95 , figsize=(15,8));
```



Research Question 01 - Top 10 Highest Runtime Movies

In [382]:
```
by_runtime = tmdb_clean_columns.sort_values(['release_year','runtime'], ascend
ing=[True, False])
```

In [383]:
```
by_runtime.shape
```

Out[383]: (3854, 15)

In [384]:
```
top_by_runtime = by_runtime.groupby('release_year').head().reset_index(drop=Tr
ue)
```

In [385]:
```
top_by_runtime.shape
```

Out[385]: (279, 15)

In [386]:
```
top_by_runtime.head()
```

Out[386]:

	popularity	budget	net_profit	revenue	original_title	cast	director	taglin
0	1.136943	NaN	3.539024e+08	NaN	Spartacus	Kirk Douglas|Laurence Olivier|Jean Simmons|Cha...	Stanley Kubrick	More titanic than any story ever told
1	1.872132	NaN	2.141847e+07	NaN	The Magnificent Seven	Yul Brynner|Eli Wallach|Steve McQueen|Charles ...	John Sturges	They were seven And they fough like seven h..
2	0.947307	NaN	1.622053e+08	NaN	The Apartment	Jack Lemmon|Shirley MacLaine|Fred MacMurray|Ra...	Billy Wilder	Movie wise there has neve been anything like..
3	2.610362	NaN	2.299854e+08	NaN	Psycho	Anthony Perkins|Vera Miles|John Gavin|Janet Le...	Alfred Hitchcock	The master o suspense moves his cameras into ..
4	0.055821	NaN	3.022917e+07	NaN	Cinderfella	Jerry Lewis|Ed Wynn|Judith Anderson|Henry Silv...	Frank Tashlin	

In [387]:
```
runtime_release_year = pd.pivot_table(top_by_runtime, index = 'release_year',
values = 'runtime' )
```

In [388]:
```
runtime_release_year.shape
```
Out[388]:
```
(56, 1)
```

In [389]:
```
runtime_release_year.head()
```
Out[389]:

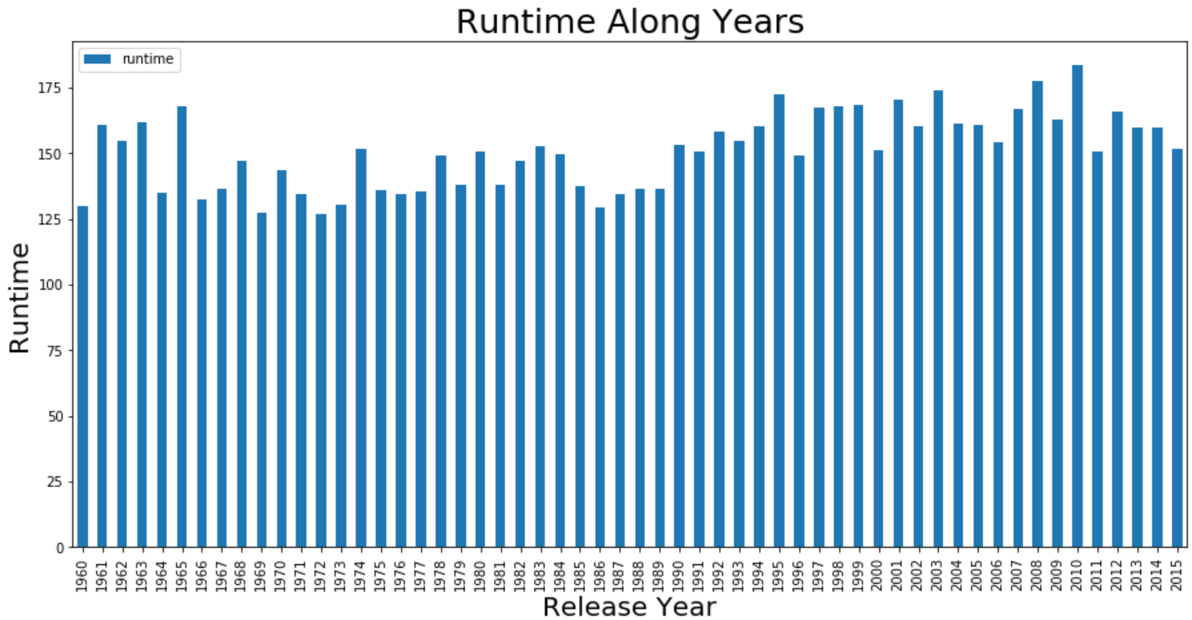
	runtime
release_year	
1960	130.0
1961	161.0
1962	154.6
1963	161.6
1964	134.8

In [390]:
```
runtime_release_year.describe()
```
Out[390]:

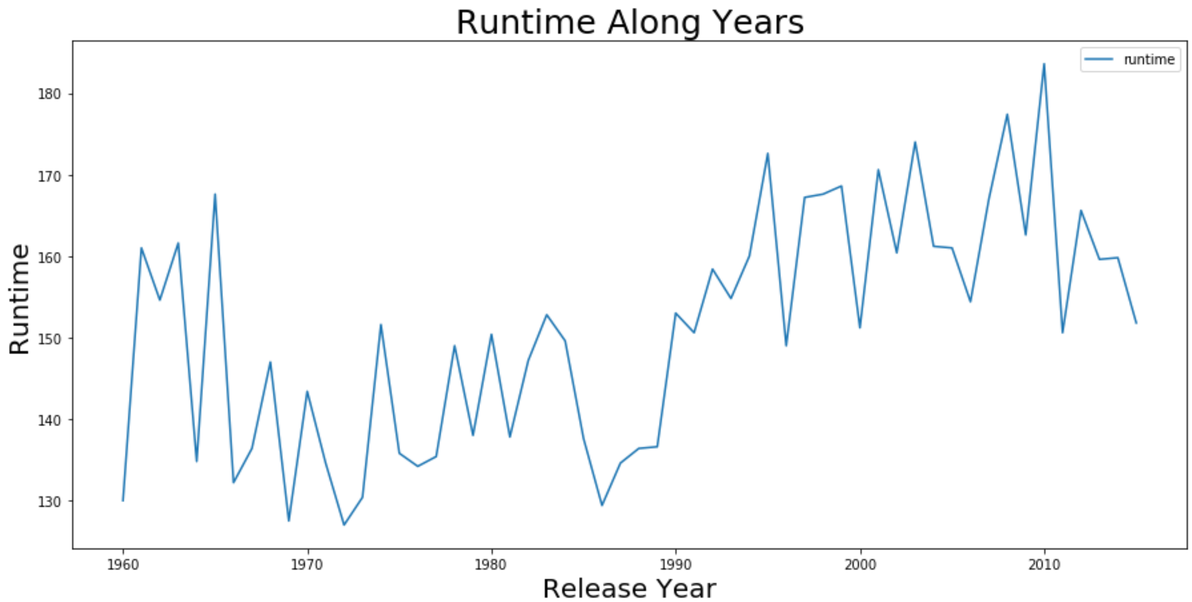
	runtime
count	56.000000
mean	151.019643
std	14.353681
min	127.000000
25%	136.550000
50%	151.400000
75%	161.050000
max	183.600000

In [391]:
```
# Plotting Function

def plotting(DATA, KIND , X_LABEL, Y_LABEL, TITLE):

    DATA.plot(kind = KIND, figsize =(15, 7))
    plt.xlabel(X_LABEL , size =(20))
    plt.ylabel(Y_LABEL , size =(20))
    plt.title(TITLE , size =(25))
    plt.legend ()
```

In [392]:
```
plotting(runtime_release_year, 'bar', 'Release Year', 'Runtime', 'Runtime Alon
g Years')
```



In [393]:
```
plotting(runtime_release_year, 'line', 'Release Year', 'Runtime', 'Runtime Alo
ng Years')
```



NOTE: From the ghraph we can figure out that watching movies were popular between the year of 1960 to the eyear of 195, and then there was a reluctance of watching movies strating from 1967 up to 1987 and then started to populate again and kept growing positively.

In [394]:
```
top_10_runtime = top_by_runtime.nlargest(10, 'runtime')
top_10_runtime
```

Out[394]:

	popularity	budget	net_profit	revenue	original_title	cast	
249	0.534192	NaN	-1.712872e+07	NaN	Carlos	Edgar RamÃ-rez|Alexander Scheer|Fadi Abi Samra...	Olivier
15	0.804533	NaN	1.896460e+08	NaN	Cleopatra	Elizabeth Taylor|Richard Burton|Rex Harrison|R...	J Mankiewicz Mamouli
99	0.418950	NaN	-1.072059e+08	NaN	Heaven's Gate	Kris Kristofferson|Christopher Walken|John Hur...	Michae
10	1.168767	NaN	3.964647e+08	NaN	Lawrence of Arabia	Peter O'Toole|Alec Guinness|Anthony Quinn|Jack...	Da
214	0.469518	NaN	-5.106033e+07	NaN	Gods and Generals	Stephen Lang|Jeff Daniels|Robert Duvall|Kevin ...	Ronald F.
239	0.389554	NaN	4.682315e+06	NaN	Jodhaa Akbar	Hrithik Roshan|Aishwarya Rai Bachchan|Sonu Soo...	/ G
159	0.648937	NaN	2.202038e+07	NaN	Malcolm X	Denzel Washington|Angela Bassett|Albert Hall|A...	S
215	7.122455	NaN	1.214855e+09	NaN	The Lord of the Rings: The Return of the King	Elijah Wood|Ian McKellen|Viggo Mortensen|Liv T...	Peter
69	3.264571	NaN	1.527582e+08	NaN	The Godfather: Part II	Al Pacino|Robert Duvall|Diane Keaton|Robert De...	Frai
25	0.146033	NaN	-5.536451e+07	NaN	The Greatest Story Ever Told	Max von Sydow|Michael Anderson Jr.|Carroll Bak...	George

In [395]:
```
# Top 10 Movies by Runtime

top_10_runtime = pd.pivot_table(top_10_runtime, index = 'original_title', valu
es = 'runtime')
```

In [396]:
```
top_10_runtime.shape
```

Out[396]: (10, 1)

In [397]:
```
top_10_runtime.head()
```

Out[397]:

	runtime
original_title	
Carlos	338.0
Cleopatra	248.0
Gods and Generals	214.0
Heaven's Gate	219.0
Jodhaa Akbar	213.0

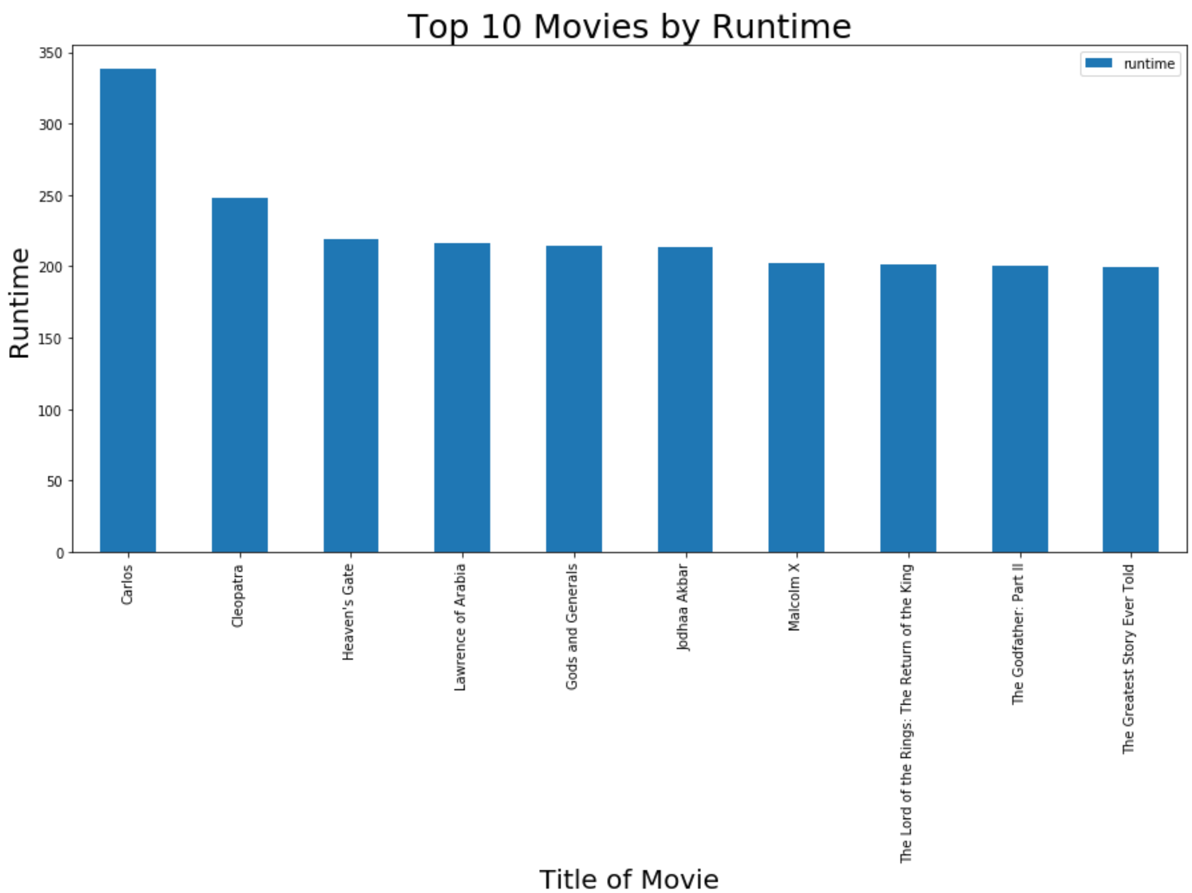
In [398]:
```
top_10_runtime
```

Out[398]:

	runtime
original_title	
Carlos	338.0
Cleopatra	248.0
Gods and Generals	214.0
Heaven's Gate	219.0
Jodhaa Akbar	213.0
Lawrence of Arabia	216.0
Malcolm X	202.0
The Godfather: Part II	200.0
The Greatest Story Ever Told	199.0
The Lord of the Rings: The Return of the King	201.0

In [399]:
```
top_10_runtime = top_10_runtime.runtime.sort_values( ascending=False)
```

In [400]:
```
plotting(top_10_runtime, 'bar', 'Title of Movie', 'Runtime', 'Top 10 Movies by Runtime')
```

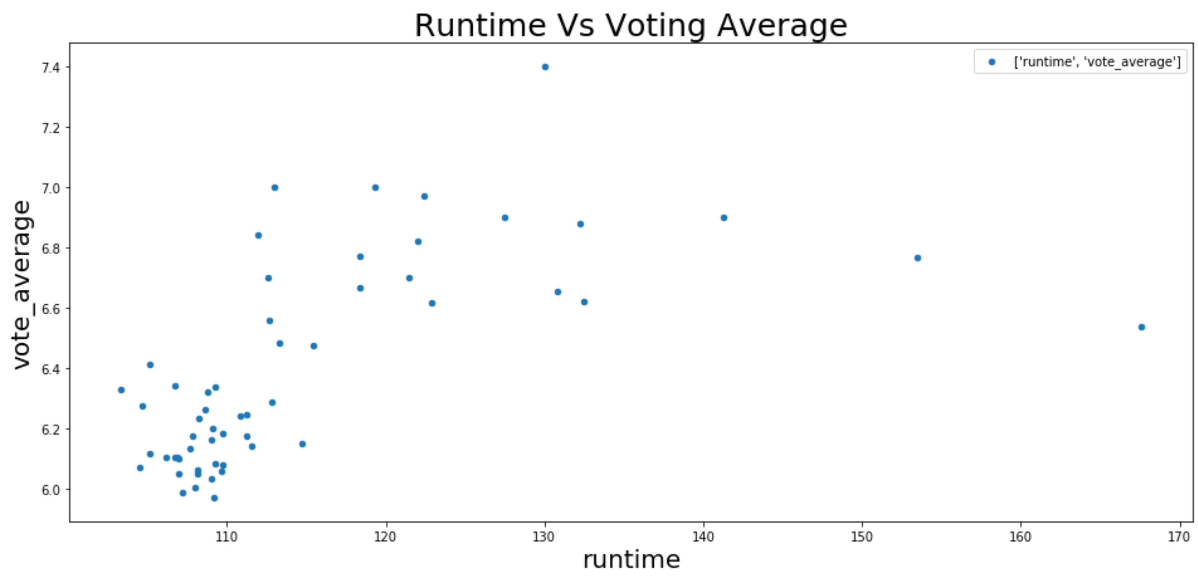


Note: From the ghraph we can figur out that the highest runtime movie is Carlos.

In [401]:
```
runtime_by_vote_average = pd.pivot_table(by_runtime, index = 'release_year', v
alues = ['runtime', 'vote_average'])
```

In [402]:
```
# Plotting 2d Function

def plotting2d (DATA, KIND, X_LABEL, Y_LABEL, TITLE):

    DATA.plot( x = X_LABEL , y = Y_LABEL , kind = KIND, figsize =(16 , 7), lab
el = [X_LABEL , Y_LABEL] )
    plt.xlabel ( X_LABEL , size =(20) )
    plt.ylabel ( Y_LABEL , size =(20) )
    plt.title ( TITLE , size =(25) )
    plt.legend()
```

In [403]:
```
plotting2d (runtime_by_vote_average, 'scatter' , 'runtime' , 'vote_average' , 
'Runtime Vs Voting Average' )
```



NOTE: From the ghraph we can figure out that there is a moderate postive non-liner correlation, and there is anout lier valye.

Research Question 02 - Top 10 Highest Revenue Movies

In [404]:
```
by_revenue = tmdb_clean_columns.sort_values(['release_year','revenue_adj'], as
cending=[True, False])
```

In [405]:
```
by_revenue.shape
```

Out[405]: (3854, 15)

In [406]:
```
by_revenue.describe()
```

Out[406]:

	popularity	budget	net_profit	revenue	runtime	vote_average	release_y
count	3854.000000	22.000000	3.854000e+03	31.000000	3854.000000	3854.000000	3854.000
mean	1.191554	30.090909	9.282470e+07	51.516129	109.220291	6.168163	2001.261
std	1.475162	36.818615	1.940715e+08	67.591356	19.922820	0.794920	11.282
min	0.001117	1.000000	-4.139124e+08	2.000000	15.000000	2.200000	1960.000
25%	0.462368	6.500000	-1.504995e+06	11.000000	95.000000	5.700000	1995.000
50%	0.797511	13.000000	2.737064e+07	16.000000	106.000000	6.200000	2004.000
75%	1.368324	28.750000	1.074548e+08	62.000000	119.000000	6.700000	2010.000
max	32.985763	114.000000	2.750137e+09	250.000000	338.000000	8.400000	2015.000

In [407]:
```
top_by_revenue = by_revenue.groupby('release_year').head().reset_index(drop=True)
```

In [408]:
```
top_by_revenue.shape
```

Out[408]: (279, 15)

In [409]:
```
top_by_revenue.head()
```

Out[409]:

	popularity	budget	net_profit	revenue	original_title	cast	director	tagline
0	1.136943	NaN	3.539024e+08	NaN	Spartacus	Kirk Douglas|Laurence Olivier|Jean Simmons|Cha...	Stanley Kubrick	More titanic than any story ever told
1	2.610362	NaN	2.299854e+08	NaN	Psycho	Anthony Perkins|Vera Miles|John Gavin|Janet Le...	Alfred Hitchcock	The master of suspense moves his cameras into ..
2	0.947307	NaN	1.622053e+08	NaN	The Apartment	Jack Lemmon|Shirley MacLaine|Fred MacMurray|Ra...	Billy Wilder	Movie wise there has never been anything like..
3	0.055821	NaN	3.022917e+07	NaN	Cinderfella	Jerry Lewis|Ed Wynn|Judith Anderson|Henry Silv...	Frank Tashlin	
4	1.872132	NaN	2.141847e+07	NaN	The Magnificent Seven	Yul Brynner|Eli Wallach|Steve McQueen|Charles ...	John Sturges	They were seven And they fought like seven h..

In [410]:
```
revenue_release_year = pd.pivot_table(top_by_revenue, index = 'release_year', values = 'revenue_adj')
```

In [411]:
```
runtime_release_year.shape
```

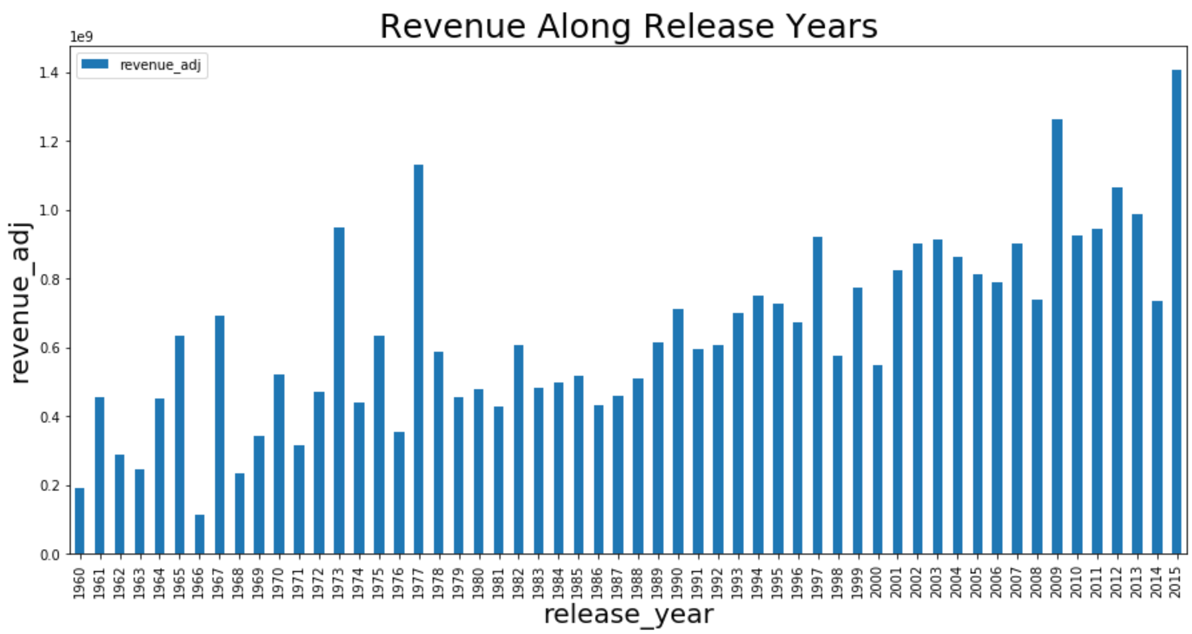
Out[411]: (56, 1)

In [412]:
```
revenue_release_year.head()
```

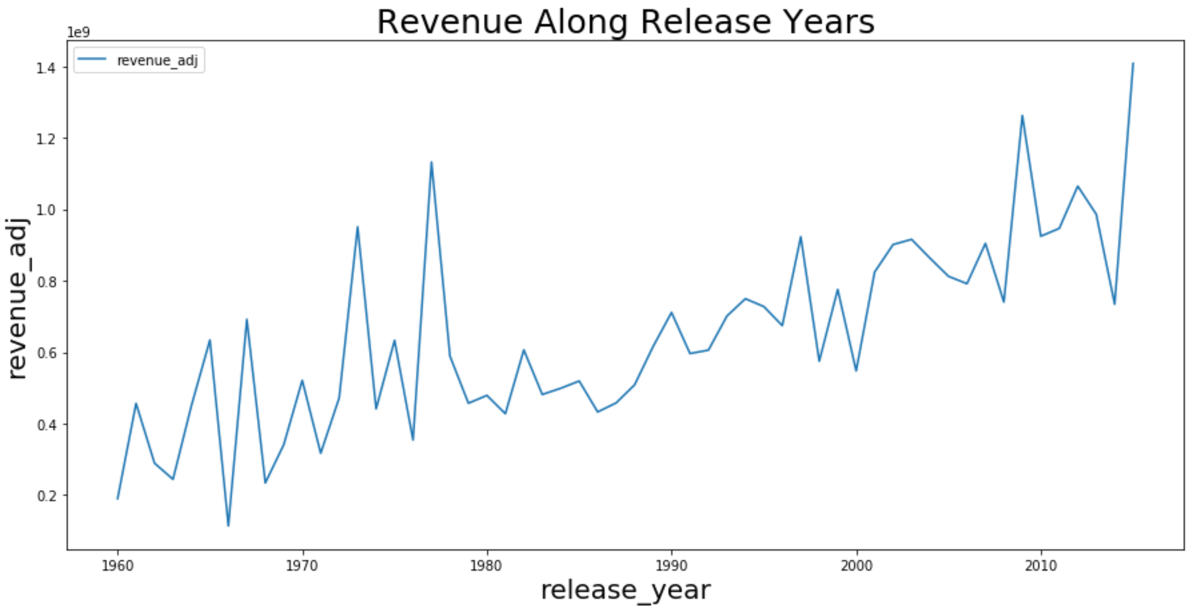
Out[412]:

release_year	revenue_adj
1960	1.902299e+08
1961	4.565419e+08
1962	2.893902e+08
1963	2.442604e+08
1964	4.507990e+08

In [413]:
```
plotting (revenue_release_year, 'bar', 'release_year', 'revenue_adj', 'Revenue
Along Release Years')
```



In [414]:
```
plotting (revenue_release_year, 'line', 'release_year', 'revenue_adj', 'Revenue Along Release Years')
```



NOTE: From the graph we can figure out that there is a moderate linear positive association between revenue and release years as the revenue kept growth.

In [415]:
```
# Top 10 Movies by Revenue

top_10_revenue = top_by_revenue.nlargest(10, 'revenue_adj')
top_10_revenue
```

Out[415]:

	popularity	budget	net_profit	revenue	original_title	cast	directo
244	9.432768	NaN	2.586237e+09	NaN	Avatar	Sam Worthington|Zoe Saldana|Sigourney Weaver|S...	James Cameroı
84	12.037933	NaN	2.750137e+09	NaN	Star Wars	Mark Hamill|Harrison Ford|Carrie Fisher|Peter ...	George Luca:
184	4.355219	NaN	2.234714e+09	NaN	Titanic	Kate Winslet|Leonardo DiCaprio|Frances Fisher|...	James Cameroı
64	2.010733	NaN	2.128036e+09	NaN	The Exorcist	Linda Blair|Max von Sydow|Ellen Burstyn|Jason ...	William Friedkiı
74	2.563191	NaN	1.878643e+09	NaN	Jaws	Roy Scheider|Robert Shaw|Richard Dreyfuss|Lorr...	Steven Spielber(
274	11.173104	NaN	1.718723e+09	NaN	Star Wars: The Force Awakens	Harrison Ford|Mark Hamill|Carrie Fisher|Adam D...	J.J. Abram:
109	2.900556	NaN	1.767968e+09	NaN	E.T. the Extra-Terrestrial	Henry Thomas|Drew Barrymore|Robert MacNaughton...	Steven Spielber(
174	1.136610	NaN	1.551568e+09	NaN	The Net	Sandra Bullock|Jeremy Northam|Dennis Miller|We...	Irwin Winkle
5	2.631987	NaN	1.545635e+09	NaN	One Hundred and One Dalmatians	Rod Taylor|J. Pat O'Malley|Betty Lou Gerson|Ma...	Clyd(Geronimi|Hamiltoı Luske|Wolfgan(Reithermaı
259	7.637767	NaN	1.234248e+09	NaN	The Avengers	Robert Downey Jr.|Chris Evans|Mark Ruffalo|Chr...	Joss Whedoı

In [416]:
```
top_10_revenue = pd.pivot_table(top_10_revenue, index = 'original_title', valu
es = 'revenue_adj')
```

In [417]:
```
top_10_revenue.shape
```

Out[417]: (10, 1)

In [418]:
```
top_10_revenue.head()
```

Out[418]:

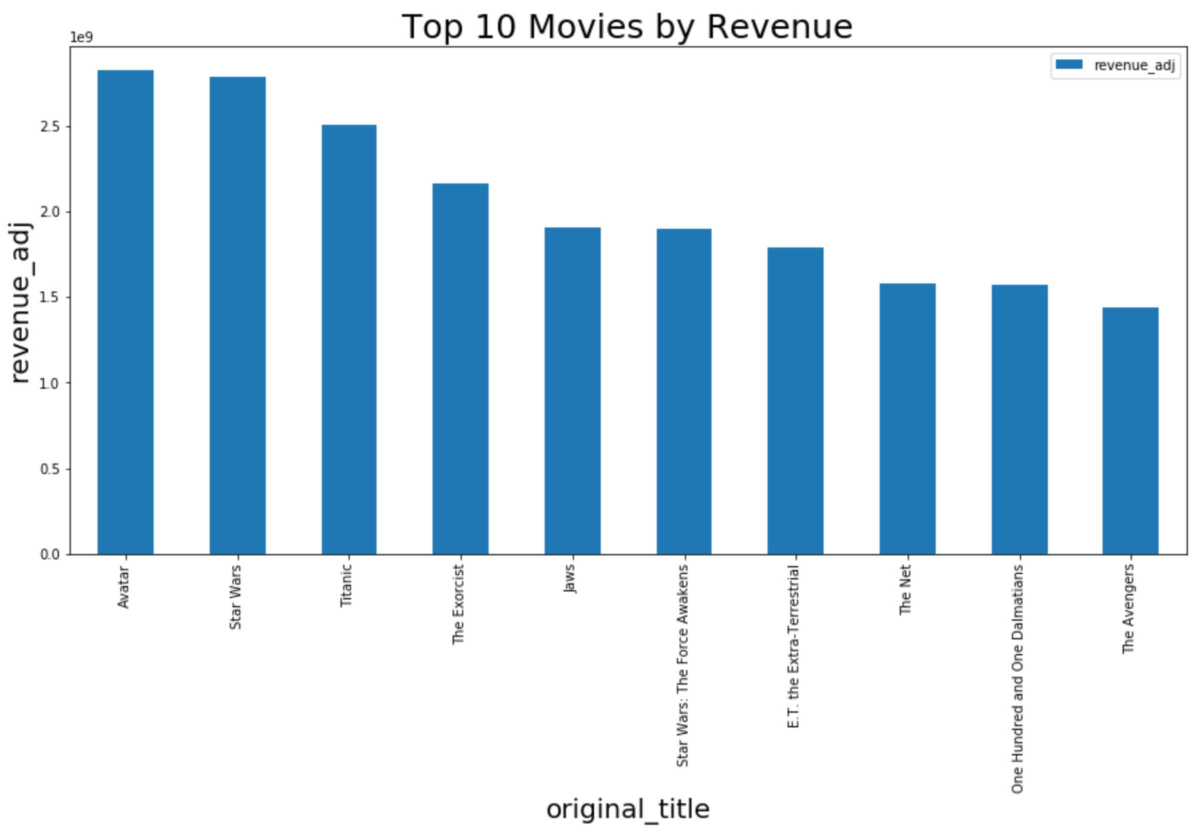
original_title	revenue_adj
Avatar	2.827124e+09
E.T. the Extra-Terrestrial	1.791694e+09
Jaws	1.907006e+09
One Hundred and One Dalmatians	1.574815e+09
Star Wars	2.789712e+09

In [419]:
```
top_10_revenue = top_10_revenue.revenue_adj.sort_values(ascending = False)
```

In [420]:
```
top_10_revenue
```

Out[420]:
```
original_title
Avatar                            2.827124e+09
Star Wars                         2.789712e+09
Titanic                           2.506406e+09
The Exorcist                      2.167325e+09
Jaws                              1.907006e+09
Star Wars: The Force Awakens      1.902723e+09
E.T. the Extra-Terrestrial        1.791694e+09
The Net                           1.583050e+09
One Hundred and One Dalmatians    1.574815e+09
The Avengers                      1.443191e+09
Name: revenue_adj, dtype: float64
```

In [421]:
```
plotting (top_10_revenue, 'bar', 'original_title', 'revenue_adj', 'Top 10 Movies by Revenue')
```



NOTE: From the grapgh we can figure out that the top revenue movie ws Avatar.

Research Question 03 - Top 10 Highest Budget Movies

In [422]:
```
by_budget = tmdb_clean_columns.sort_values(['release_year','budget_adj'], ascending=[True, False])
```

In [423]:
```
by_budget.shape
```

Out[423]:
```
(3854, 15)
```

In [424]:
```
by_budget.describe()
```

Out[424]:

	popularity	budget	net_profit	revenue	runtime	vote_average	release_y
count	3854.000000	22.000000	3.854000e+03	31.000000	3854.000000	3854.000000	3854.000
mean	1.191554	30.090909	9.282470e+07	51.516129	109.220291	6.168163	2001.261
std	1.475162	36.818615	1.940715e+08	67.591356	19.922820	0.794920	11.282
min	0.001117	1.000000	-4.139124e+08	2.000000	15.000000	2.200000	1960.000
25%	0.462368	6.500000	-1.504995e+06	11.000000	95.000000	5.700000	1995.000
50%	0.797511	13.000000	2.737064e+07	16.000000	106.000000	6.200000	2004.000
75%	1.368324	28.750000	1.074548e+08	62.000000	119.000000	6.700000	2010.000
max	32.985763	114.000000	2.750137e+09	250.000000	338.000000	8.400000	2015.000

In [425]:
```
top_by_budget  = by_budget.groupby('release_year').head().reset_index(drop=Tru
e)
```

In [426]:
```
top_by_budget.shape
```

Out[426]: (279, 15)

In [427]:
```
top_by_budget.head()
```

Out[427]:

	popularity	budget	net_profit	revenue	original_title	cast	director	tagline
0	1.136943	NaN	3.539024e+08	NaN	Spartacus	Kirk Douglas|Laurence Olivier|Jean Simmons|Cha...	Stanley Kubrick	More titanic than any story ever told
1	0.947307	NaN	1.622053e+08	NaN	The Apartment	Jack Lemmon|Shirley MacLaine|Fred MacMurray|Ra...	Billy Wilder	Movie wise there has never been anything like..
2	0.055821	NaN	3.022917e+07	NaN	Cinderfella	Jerry Lewis|Ed Wynn|Judith Anderson|Henry Silv...	Frank Tashlin	
3	1.872132	NaN	2.141847e+07	NaN	The Magnificent Seven	Yul Brynner|Eli Wallach|Steve McQueen|Charles ...	John Sturges	They were seven And they fough like seven h..
4	2.610362	NaN	2.299854e+08	NaN	Psycho	Anthony Perkins|Vera Miles|John Gavin|Janet Le...	Alfred Hitchcock	The master o suspense moves his cameras into ..

In [428]:
```
budget_release_year = pd.pivot_table(top_by_budget, index = 'release_year', va
lues = 'budget_adj')
```

In [429]:
```
budget_release_year.shape
```

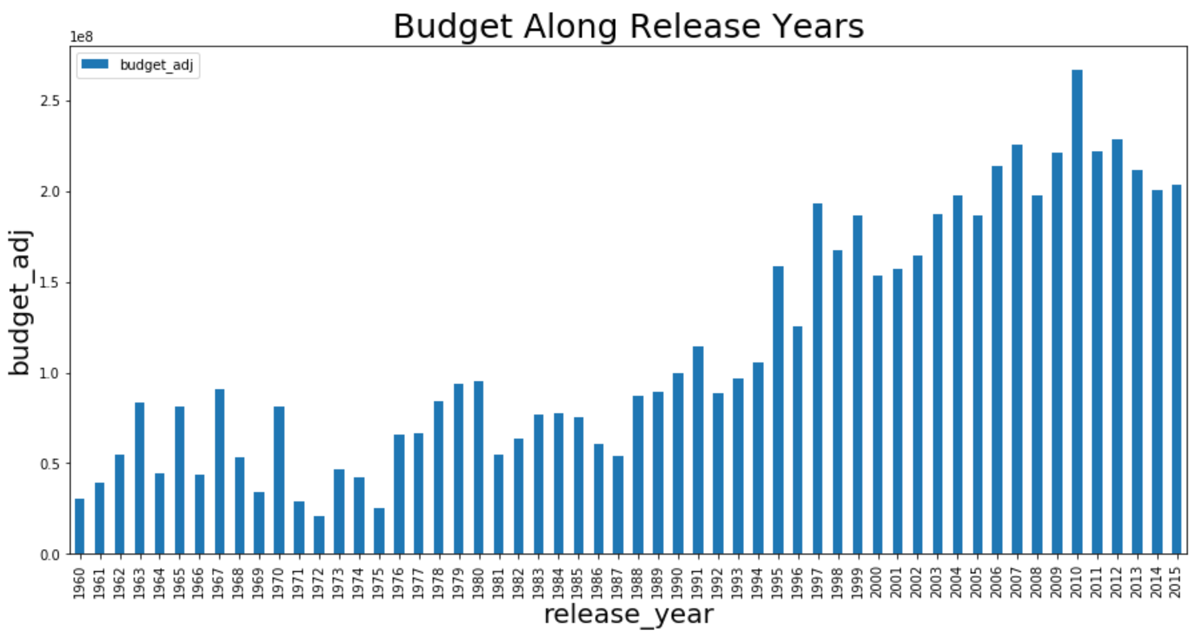
Out[429]: (56, 1)

In [430]:
```
budget_release_year.head()
```

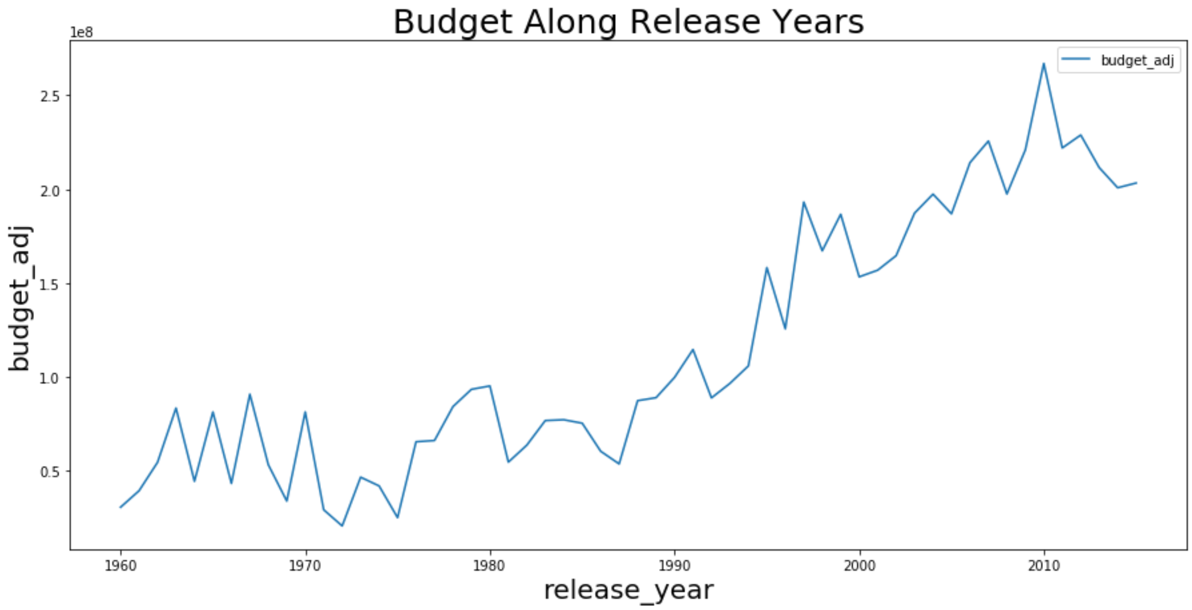
Out[430]:

	budget_adj
release_year	
1960	3.068179e+07
1961	3.944623e+07
1962	5.456796e+07
1963	8.346986e+07
1964	4.448073e+07

In [431]:
```
plotting (budget_release_year, 'bar', 'release_year', 'budget_adj', 'Budget Al
ong Release Years')
```



In [432]:
```
plotting (budget_release_year, 'line', 'release_year', 'budget_adj', 'Budget A
long Release Years')
```



NOTE: From the graph we can figure out that budgeting of movies were increasing by the years and the highest budgeting were in the recent years after the year 2000.

In [433]:
```
# Top 10 Movies by Budget

top_10_budget = top_by_budget.nlargest(10, 'budget_adj')
top_10_budget
```

Out[433]:

	popularity	budget	net_profit	revenue	original_title	cast	director	
249	0.250540	NaN	-4.139124e+08	NaN	The Warrior's Way	Kate Bosworth|Jang Dong-gun|Geoffrey Rush|Dann...	Sngmoo Lee	As L
254	4.955130	NaN	6.220462e+08	NaN	Pirates of the Caribbean: On Stranger Tides	Johnny Depp|PenÃ©lope Cruz|Geoffrey Rush|Ian M...	Rob Marshall	I
234	4.965391	NaN	6.951529e+08	NaN	Pirates of the Caribbean: At World's End	Johnny Depp|Orlando Bloom|Keira Knightley|Geof...	Gore Verbinski	en wo ad
229	1.957331	NaN	1.309698e+08	NaN	Superman Returns	Brandon Routh|Kevin Spacey|Kate Bosworth|James...	Bryan Singer	
184	4.355219	NaN	2.234714e+09	NaN	Titanic	Kate Winslet|Leonardo DiCaprio|Frances Fisher|...	James Cameron	I o b
235	2.520912	NaN	6.655712e+08	NaN	Spider-Man 3	Tobey Maguire|Kirsten Dunst|James Franco|Thoma...	Sam Raimi	Th
250	2.865684	NaN	3.317949e+08	NaN	Tangled	Zachary Levi|Mandy Moore|Donna Murphy|Ron Perl...	Nathan Greno|Byron Howard	ad I
274	5.944927	NaN	1.035032e+09	NaN	Avengers: Age of Ultron	Robert Downey Jr.|Chris Hemsworth|Mark Ruffalo...	Joss Whedon	A
244	5.076472	NaN	6.951764e+08	NaN	Harry Potter and the Half-Blood Prince	Daniel Radcliffe|Rupert Grint|Emma Watson|Tom ...	David Yates	Re
174	1.232098	NaN	1.276683e+08	NaN	Waterworld	Kevin Costner|Chaim Girafi|Rick Aviles|R. D. C...	Kevin Reynolds	I s

In [434]:
```
top_10_budget = pd.pivot_table(top_10_budget, index = 'original_title', values
= 'budget_adj')
```

In [435]:
```
top_10_budget.shape
```

Out[435]:
```
(10, 1)
```

In [436]:
```
top_10_budget.head()
```

Out[436]:

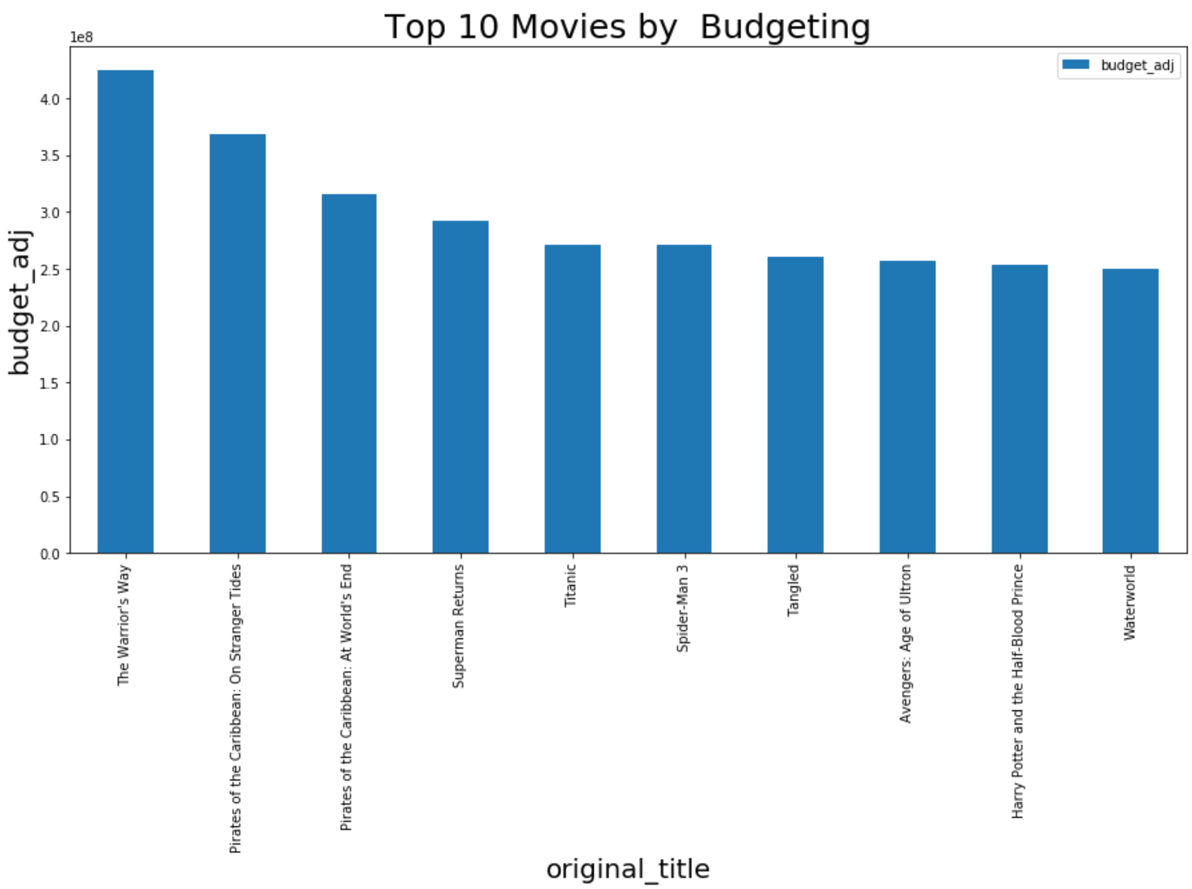
original_title	budget_adj
Avengers: Age of Ultron	257599886.7
Harry Potter and the Half-Blood Prince	254100108.5
Pirates of the Caribbean: At World's End	315500574.8
Pirates of the Caribbean: On Stranger Tides	368371256.2
Spider-Man 3	271330494.3

In [437]:
```
top_10_budget = top_10_budget.budget_adj.sort_values(ascending = False)
```

In [438]:
```
top_10_budget
```

Out[438]:
```
original_title
The Warrior's Way                                 425000000.0
Pirates of the Caribbean: On Stranger Tides       368371256.2
Pirates of the Caribbean: At World's End          315500574.8
Superman Returns                                  292050672.7
Titanic                                           271692064.2
Spider-Man 3                                      271330494.3
Tangled                                           260000000.0
Avengers: Age of Ultron                           257599886.7
Harry Potter and the Half-Blood Prince            254100108.5
Waterworld                                        250419201.7
Name: budget_adj, dtype: float64
```

In [439]:
```
plotting (top_10_budget, 'bar', 'original_title', 'budget_adj', 'Top 10 Movies by  Budgeting')
```



NOTE: From the graph we can figur out that the highest budget of movie was The Warrior's Way.

Research Question 04 - Top 10 Highest Rating Movies

In [440]:
```
by_vote = tmdb_clean_columns.sort_values(['release_year','vote_average'], ascending=[True, False])
```

In [441]:
```
by_vote.shape
```

Out[441]: (3854, 15)

In [442]:
```
top_by_vote  = by_vote.groupby('release_year').head().reset_index(drop=True)
```

In [443]:
```
top_by_vote.shape
```

Out[443]: (279, 15)

In [444]:
```
top_by_vote.head()
```

Out[444]:

	popularity	budget	net_profit	revenue	original_title	cast	director	tagline
0	2.610362	NaN	2.299854e+08	NaN	Psycho	Anthony Perkins|Vera Miles|John Gavin|Janet Le...	Alfred Hitchcock	The master o suspense moves his cameras into ..
1	0.947307	NaN	1.622053e+08	NaN	The Apartment	Jack Lemmon|Shirley MacLaine|Fred MacMurray|Ra...	Billy Wilder	Movie wise there has neve beer anything like..
2	0.055821	NaN	3.022917e+07	NaN	Cinderfella	Jerry Lewis|Ed Wynn|Judith Anderson|Henry Silv...	Frank Tashlin	
3	1.872132	NaN	2.141847e+07	NaN	The Magnificent Seven	Yul Brynner|Eli Wallach|Steve McQueen|Charles ...	John Sturges	They were seven And they fough like sever h..
4	1.136943	NaN	3.539024e+08	NaN	Spartacus	Kirk Douglas|Laurence Olivier|Jean Simmons|Cha...	Stanley Kubrick	More titanic than any story ever told

In [445]:
```
vote_release_year = pd.pivot_table(top_by_vote, index = 'release_year', values
= 'vote_average')
```

In [446]:
```
vote_release_year.shape
```

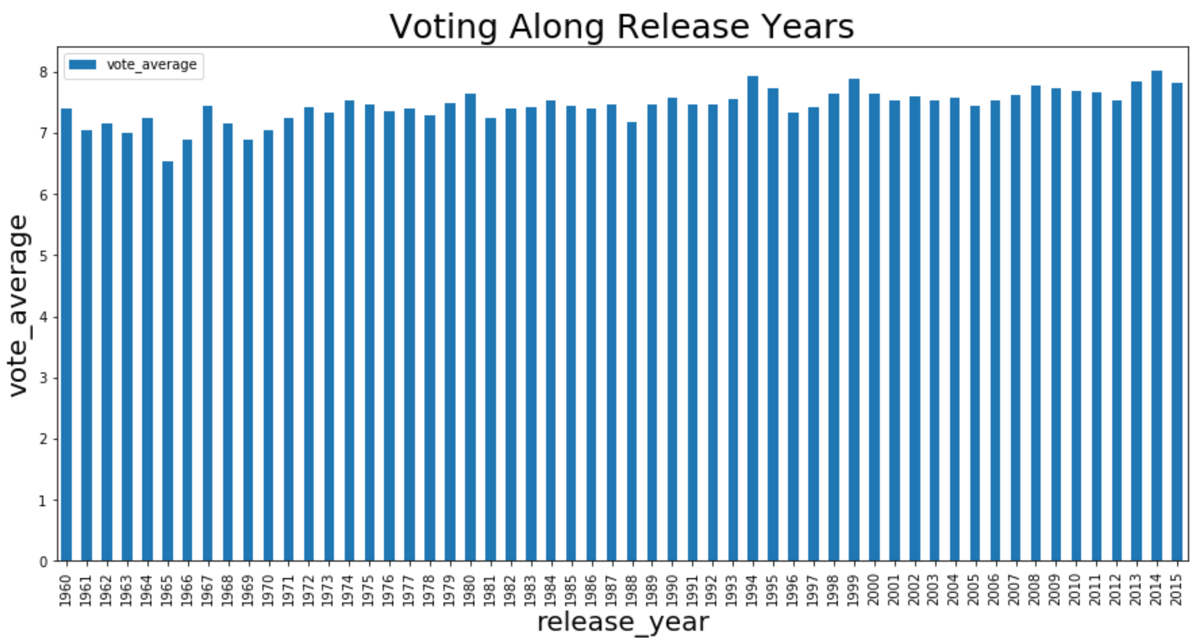
Out[446]: (56, 1)

In [447]:
```
vote_release_year.head()
```

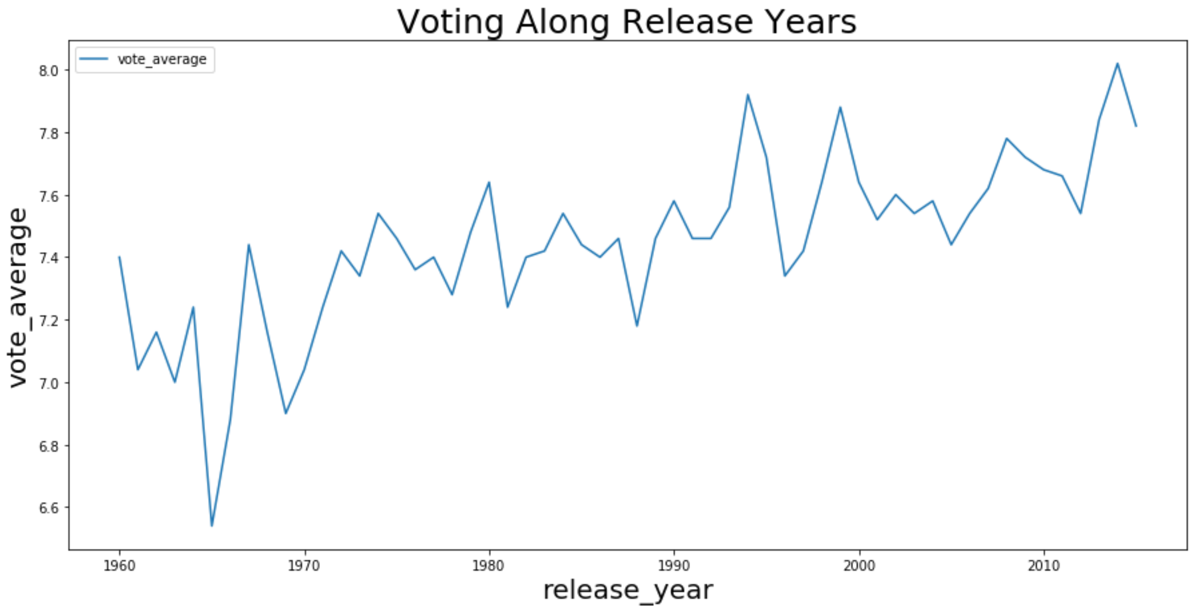
Out[447]:

release_year	vote_average
1960	7.40
1961	7.04
1962	7.16
1963	7.00
1964	7.24

In [448]:
```
plotting (vote_release_year, 'bar', 'release_year', 'vote_average', 'Voting Al
ong Release Years')
```



In [449]:
```
plotting (vote_release_year, 'line', 'release_year', 'vote_average', 'Voting A
long Release Years')
```



NOTES: From the graph we can figure out that voting was moderat increasing along release years except the sharp decrease between the year 1964 and the year 1967.

In [450]:
```
# Top Movies by Rating

top_10_vote = top_by_vote.nlargest(10, 'vote_average')
top_10_vote
```

Out[450]:

	popularity	budget	net_profit	revenue	original_title	cast	director	ta
119	0.283191	NaN	7.932116e+06	NaN	Stop Making Sense	David Byrne|Tina Weymouth|Chris Frantz|Jerry H...	Jonathan Demme	Why ma se W m Why
169	7.192039	NaN	4.915674e+06	NaN	The Shawshank Redemption	Tim Robbins|Morgan Freeman|Bob Gunton|William ...	Frank Darabont	Fea hol pris Hope set yo
59	5.738034	NaN	1.246626e+09	NaN	The Godfather	Marlon Brando|Al Pacino|James Caan|Richard S. ...	Francis Ford Coppola	An you re
269	4.780419	NaN	9.849312e+06	NaN	Whiplash	Miles Teller|J.K. Simmons|Melissa Benoist|Aust...	Damien Chazelle	The great can you t e
69	3.264571	NaN	1.527582e+08	NaN	The Godfather: Part II	Al Pacino|Robert Duvall|Diane Keaton|Robert De...	Francis Ford Coppola	I don' I ha every T
164	2.377288	NaN	4.517327e+08	NaN	Schindler's List	Liam Neeson|Ben Kingsley|Ralph Fiennes|Carolin...	Steven Spielberg	Who saves life, s the e
170	6.715966	NaN	9.164222e+08	NaN	Forrest Gump	Tom Hanks|Robin Wright|Gary Sinise|Mykelti Wil...	Robert Zemeckis	The will b s you'
171	8.093754	NaN	3.029442e+08	NaN	Pulp Fiction	John Travolta|Samuel L. Jackson|Uma Thurman|Br...	Quentin Tarantino	bec you char do me
194	8.947905	NaN	4.955256e+07	NaN	Fight Club	Edward Norton|Brad Pitt|Meat Loaf|Jared Leto|H...	David Fincher	mucl you a your you
239	8.466668	NaN	8.273675e+08	NaN	The Dark Knight	Christian Bale|Michael Caine|Heath Ledger|Aaro...	Christopher Nolan	Wh Seri

In [451]:
```
top_10_vote  = pd.pivot_table(top_10_vote, index = 'original_title', values = 
'vote_average')
```

In [452]:
```
top_10_vote.shape
```

Out[452]: (10, 1)

In [453]:
```
top_10_vote.head()
```

Out[453]:

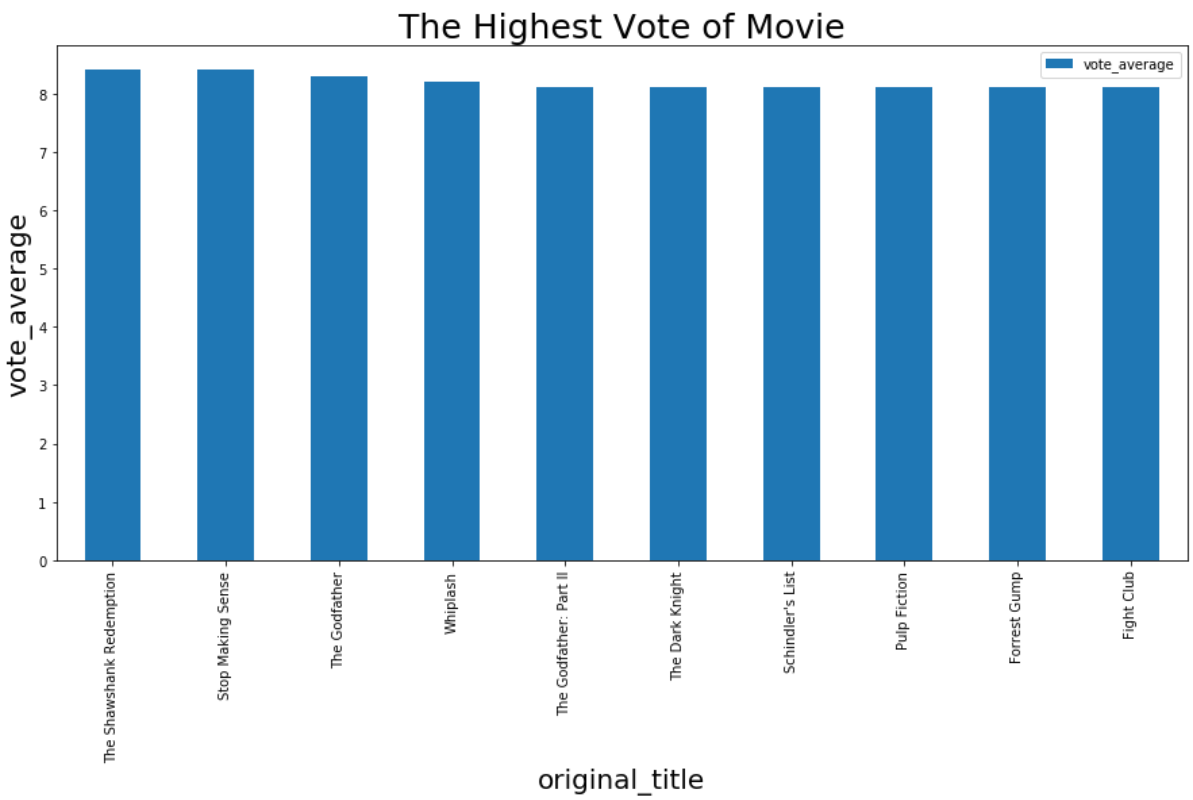
	vote_average
original_title	
Fight Club	8.1
Forrest Gump	8.1
Pulp Fiction	8.1
Schindler's List	8.1
Stop Making Sense	8.4

In [454]:
```
top_10_vote = top_10_vote.vote_average.sort_values(ascending = False)
```

In [455]:
```
top_10_vote
```

Out[455]:
```
original_title
The Shawshank Redemption    8.4
Stop Making Sense           8.4
The Godfather               8.3
Whiplash                    8.2
The Godfather: Part II      8.1
The Dark Knight             8.1
Schindler's List            8.1
Pulp Fiction                8.1
Forrest Gump                8.1
Fight Club                  8.1
Name: vote_average, dtype: float64
```

In [456]:
```
plotting (top_10_vote, 'bar', 'original_title', 'vote_average', 'The Highest V
ote of Movie')
```



NOTE: from the graph we can figur out that highest vote of movie was The Shawshank Redemption.

Research Question 05 - Top 10 Highest Net Profit Movies

In [457]:
```
by_net_profit = tmdb_clean_columns.sort_values(['release_year','net_profit'],
ascending=[True, False])
```

In [458]:
```
by_net_profit.shape
```

Out[458]: (3854, 15)

In [459]:
```
top_by_net_profit = by_net_profit.groupby('release_year').head().reset_index(d
rop=True)
```

In [460]:
```
top_by_net_profit.shape
```

Out[460]: (279, 15)

In [461]:
```
top_by_net_profit.head()
```

Out[461]:

	popularity	budget	net_profit	revenue	original_title	cast	director	tagline
0	1.136943	NaN	3.539024e+08	NaN	Spartacus	Kirk Douglas|Laurence Olivier|Jean Simmons|Cha...	Stanley Kubrick	More titanic than any story ever told
1	2.610362	NaN	2.299854e+08	NaN	Psycho	Anthony Perkins|Vera Miles|John Gavin|Janet Le...	Alfred Hitchcock	The master of suspense moves his cameras into ..
2	0.947307	NaN	1.622053e+08	NaN	The Apartment	Jack Lemmon|Shirley MacLaine|Fred MacMurray|Ra...	Billy Wilder	Movie wise there has never been anything like..
3	0.055821	NaN	3.022917e+07	NaN	Cinderfella	Jerry Lewis|Ed Wynn|Judith Anderson|Henry Silv...	Frank Tashlin	
4	1.872132	NaN	2.141847e+07	NaN	The Magnificent Seven	Yul Brynner|Eli Wallach|Steve McQueen|Charles ...	John Sturges	They were seven And they fought like seven h..

In [462]:
```
net_profit_release_year = pd.pivot_table(top_by_net_profit, index = 'release_year', values = 'net_profit')
```

In [463]:
```
net_profit_release_year.shape
```

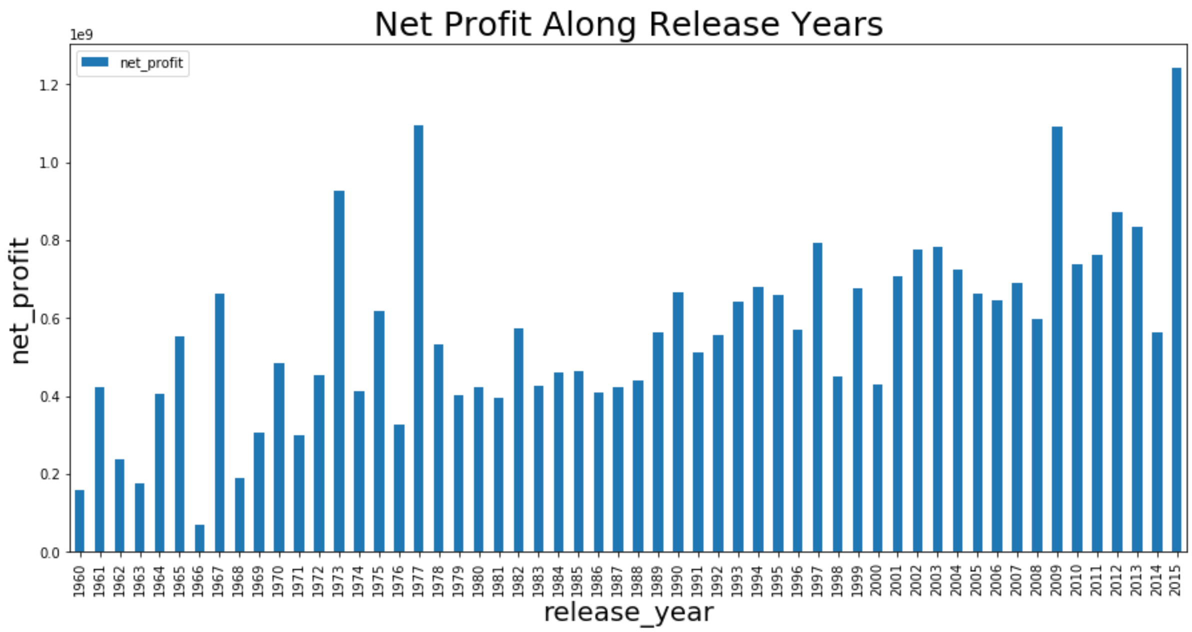
Out[463]: (56, 1)

In [464]:
```
net_profit_release_year.head()
```

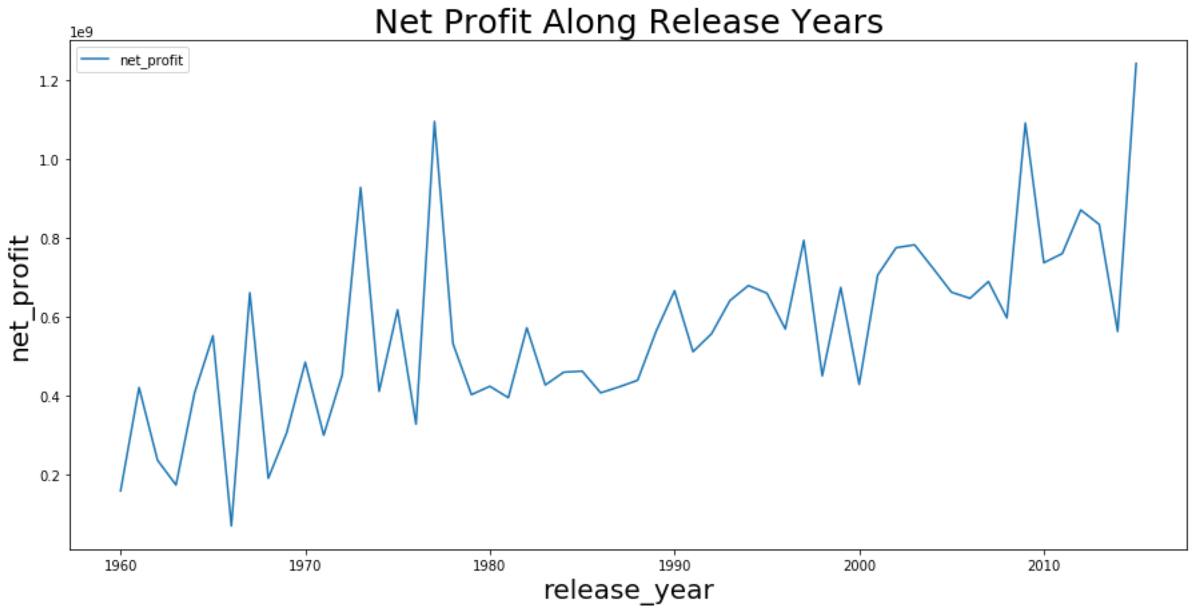
Out[464]:

release_year	net_profit
1960	1.595481e+08
1961	4.214726e+08
1962	2.370568e+08
1963	1.743189e+08
1964	4.067895e+08

In [465]:
```
plotting (net_profit_release_year, 'bar', 'release_year', 'net_profit', 'Net P
rofit Along Release Years')
```



In [466]:
```
plotting (net_profit_release_year, 'line', 'release_year', 'net_profit', 'Net
 Profit Along Release Years')
```



NOTE: From the graph we can figure out that the net profit was increasing along release years and we can notice also there was a sharp increase in some years such as 1973, 1978, 2010 and also the movies after the year 2014.

In [467]:
```
# Top Movies by net_profit

top_10_net_profit = top_by_vote.nlargest(10, 'net_profit')
top_10_net_profit
```

Out[467]:

	popularity	budget	net_profit	revenue	original_title	cast	director
84	12.037933	NaN	2.750137e+09	NaN	Star Wars	Mark Hamill|Harrison Ford|Carrie Fisher|Peter ...	George Lucas
186	4.355219	NaN	2.234714e+09	NaN	Titanic	Kate Winslet|Leonardo DiCaprio|Frances Fisher|...	James Cameron
67	2.010733	NaN	2.128036e+09	NaN	The Exorcist	Linda Blair|Max von Sydow|Ellen Burstyn|Jason ...	William Friedkin
76	2.563191	NaN	1.878643e+09	NaN	Jaws	Roy Scheider|Robert Shaw|Richard Dreyfuss|Lorr...	Steven Spielberg
113	2.900556	NaN	1.767968e+09	NaN	E.T. the Extra-Terrestrial	Henry Thomas|Drew Barrymore|Robert MacNaughton...	Steven Spielberg
9	2.631987	NaN	1.545635e+09	NaN	One Hundred and One Dalmatians	Rod Taylor|J. Pat O'Malley|Betty Lou Gerson|Ma...	Clyde Geronimi|Hamilton Luske|Wolfgang Reitherman
99	5.488441	NaN	1.376998e+09	NaN	The Empire Strikes Back	Mark Hamill|Harrison Ford|Carrie Fisher|Billy ...	Irvin Kershner
166	2.204926	NaN	1.293767e+09	NaN	Jurassic Park	Sam Neill|Laura Dern|Jeff Goldblum|Richard Att...	Steven Spielberg
59	5.738034	NaN	1.246626e+09	NaN	The Godfather	Marlon Brando|Al Pacino|James Caan|Richard S. ...	Francis Ford Coppola
263	7.637767	NaN	1.234248e+09	NaN	The Avengers	Robert Downey Jr.|Chris Evans|Mark Ruffalo|Chr...	Joss Whedon

In [468]:
```
top_10_net_profit  = pd.pivot_table(top_10_net_profit, index = 'original_titl
e', values = 'net_profit')
```

In [469]:
```
top_10_net_profit.shape
```

Out[469]: (10, 1)

In [470]:
```
top_10_net_profit.head()
```

Out[470]:

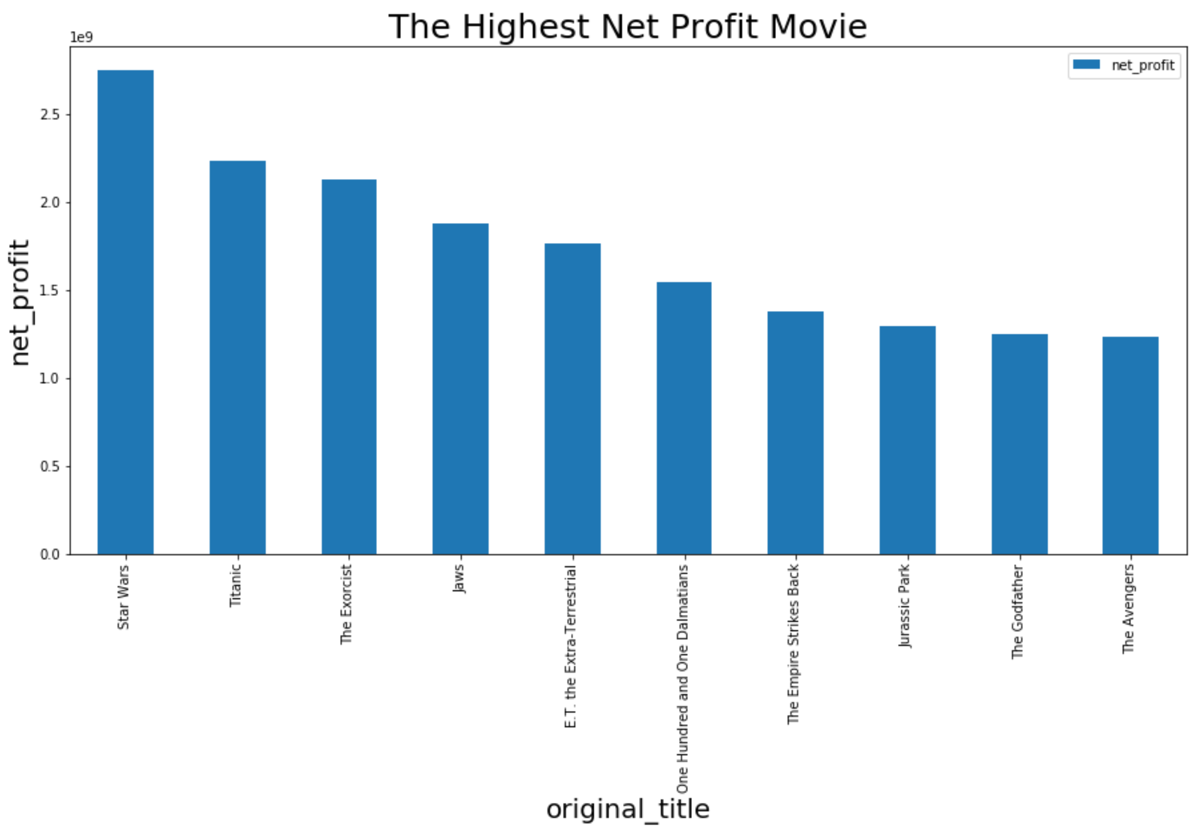
original_title	net_profit
E.T. the Extra-Terrestrial	1.767968e+09
Jaws	1.878643e+09
Jurassic Park	1.293767e+09
One Hundred and One Dalmatians	1.545635e+09
Star Wars	2.750137e+09

In [471]:
```
top_10_net_profit = top_10_net_profit.net_profit.sort_values(ascending = False
)
```

In [472]:
```
top_10_net_profit
```

Out[472]:
```
original_title
Star Wars                         2.750137e+09
Titanic                           2.234714e+09
The Exorcist                      2.128036e+09
Jaws                              1.878643e+09
E.T. the Extra-Terrestrial        1.767968e+09
One Hundred and One Dalmatians    1.545635e+09
The Empire Strikes Back           1.376998e+09
Jurassic Park                     1.293767e+09
The Godfather                     1.246626e+09
The Avengers                      1.234248e+09
Name: net_profit, dtype: float64
```

In [473]:
```
plotting (top_10_net_profit, 'bar', 'original_title', 'net_profit', 'The Highe
st Net Profit Movie')
```



NOTE: From the graph we can figure out that the highest net profit movie was Star Wars.

In [474]:
```
by_net_profit.head()
```

Out[474]:

	popularity	budget	net_profit	revenue	original_title	cast	director	tag
3190	1.136943	NaN	3.539024e+08	NaN	Spartacus	Kirk Douglas|Laurence Olivier|Jean Simmons|Cha...	Stanley Kubrick	M ti than s ever
4520	2.610362	NaN	2.299854e+08	NaN	Psycho	Anthony Perkins|Vera Miles|John Gavin|Janet Le...	Alfred Hitchcock	mast suspe m cam in
5029	0.947307	NaN	1.622053e+08	NaN	The Apartment	Jack Lemmon|Shirley MacLaine|Fred MacMurray|Ra...	Billy Wilder	Mo v there n b anyt li
7363	0.055821	NaN	3.022917e+07	NaN	Cinderfella	Jerry Lewis|Ed Wynn|Judith Anderson|Henry Silv...	Frank Tashlin	
7679	1.872132	NaN	2.141847e+07	NaN	The Magnificent Seven	Yul Brynner|Eli Wallach|Steve McQueen|Charles ...	John Sturges	T v sev And fo se

In [475]:
```
by_net_profit.shape
```

Out[475]: (3854, 15)

In [476]:
```
by_net_profit.describe()
```

Out[476]:

	popularity	budget	net_profit	revenue	runtime	vote_average	release_y
count	3854.000000	22.000000	3.854000e+03	31.000000	3854.000000	3854.000000	3854.000
mean	1.191554	30.090909	9.282470e+07	51.516129	109.220291	6.168163	2001.261
std	1.475162	36.818615	1.940715e+08	67.591356	19.922820	0.794920	11.282
min	0.001117	1.000000	-4.139124e+08	2.000000	15.000000	2.200000	1960.000
25%	0.462368	6.500000	-1.504995e+06	11.000000	95.000000	5.700000	1995.000
50%	0.797511	13.000000	2.737064e+07	16.000000	106.000000	6.200000	2004.000
75%	1.368324	28.750000	1.074548e+08	62.000000	119.000000	6.700000	2010.000
max	32.985763	114.000000	2.750137e+09	250.000000	338.000000	8.400000	2015.000

In [477]:
```
top_10_genres = by_net_profit.nlargest(10, 'net_profit')
```

In [478]:
```
top_10_genres.head()
```

Out[478]:

	popularity	budget	net_profit	revenue	original_title	cast	director	
201	12.037933	NaN	2.750137e+09	NaN	Star Wars	Mark Hamill|Harrison Ford|Carrie Fisher|Peter ...	George Lucas	A ago ir far, f
20	9.432768	NaN	2.586237e+09	NaN	Avatar	Sam Worthington|Zoe Saldana|Sigourney Weaver|S...	James Cameron	
59	4.355219	NaN	2.234714e+09	NaN	Titanic	Kate Winslet|Leonardo DiCaprio|Frances Fisher|...	James Cameron	N Ea come
410	2.010733	NaN	2.128036e+09	NaN	The Exorcist	Linda Blair|Max von Sydow|Ellen Burstyn|Jason ...	William Friedkin	S almo: compr is
385	2.563191	NaN	1.878643e+09	NaN	Jaws	Roy Scheider|Robert Shaw|Richard Dreyfuss|Lorr...	Steven Spielberg	Don't

In [479]:
```
Geners_Net_Profit = pd.pivot_table(top_10_genres, index = 'genres', values = 
'net_profit')
```

In [480]:
```
Geners_Net_Profit
```

Out[480]:

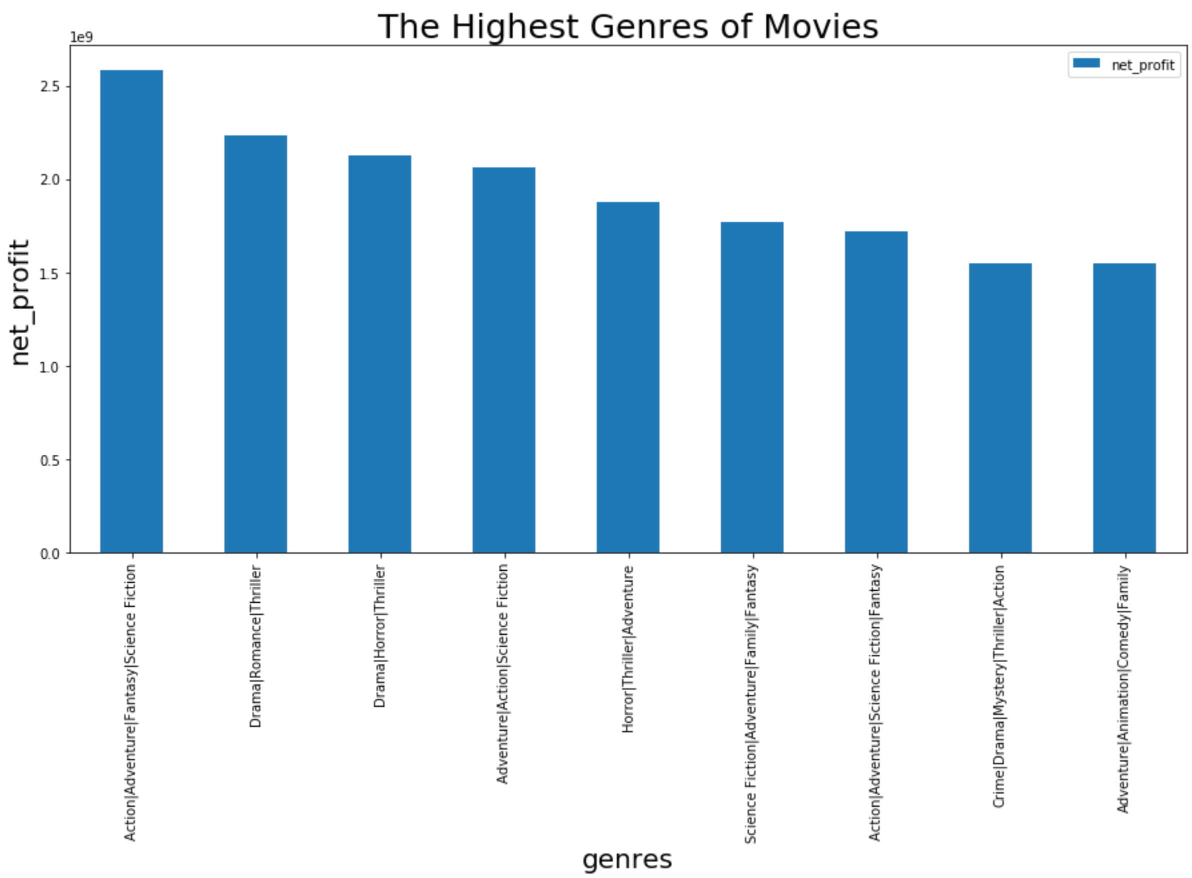
genres	net_profit
Action|Adventure|Fantasy|Science Fiction	2.586237e+09
Action|Adventure|Science Fiction|Fantasy	1.718723e+09
Adventure|Action|Science Fiction	2.063567e+09
Adventure|Animation|Comedy|Family	1.545635e+09
Crime|Drama|Mystery|Thriller|Action	1.551568e+09
Drama|Horror|Thriller	2.128036e+09
Drama|Romance|Thriller	2.234714e+09
Horror|Thriller|Adventure	1.878643e+09
Science Fiction|Adventure|Family|Fantasy	1.767968e+09

In [481]:
```
Geners_Net_Profit = Geners_Net_Profit.net_profit.sort_values(ascending = False
)
```

In [482]:
```
Geners_Net_Profit
```

Out[482]:
```
genres
Action|Adventure|Fantasy|Science Fiction    2.586237e+09
Drama|Romance|Thriller                      2.234714e+09
Drama|Horror|Thriller                       2.128036e+09
Adventure|Action|Science Fiction            2.063567e+09
Horror|Thriller|Adventure                   1.878643e+09
Science Fiction|Adventure|Family|Fantasy    1.767968e+09
Action|Adventure|Science Fiction|Fantasy    1.718723e+09
Crime|Drama|Mystery|Thriller|Action         1.551568e+09
Adventure|Animation|Comedy|Family           1.545635e+09
Name: net_profit, dtype: float64
```

In [483]:
```
plotting(Geners_Net_Profit, 'bar', 'genres','net_profit' , 'The Highest Genres of Movies')
```



NOTE: From the graph we can figure out the highest genres of movies were Action, Adventure, Fantasy and Science Fiction.

In [484]:
```
explorations  = pd.pivot_table(by_net_profit, index = 'release_year', values = ['net_profit', 'runtime', 'budget_adj', 'revenue_adj', 'vote_average', 'popularity', 'genres', 'cast'])
```

In [485]:
```
explorations.describe()
```

Out[485]:

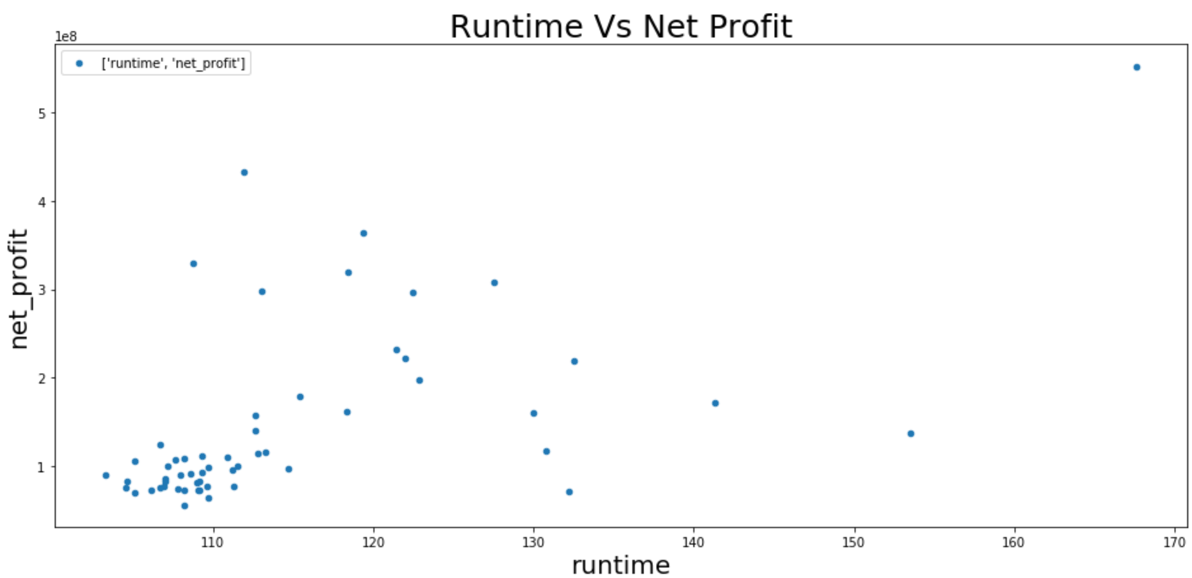
	budget_adj	net_profit	popularity	revenue_adj	runtime	vote_average
count	5.600000e+01	5.600000e+01	56.000000	5.600000e+01	56.000000	56.000000
mean	3.991923e+07	1.456689e+08	1.059704	1.855881e+08	114.847356	6.391060
std	1.312820e+07	1.032232e+08	0.377869	1.026257e+08	12.203451	0.339090
min	1.542484e+07	5.585189e+07	0.395168	9.509591e+07	103.304348	5.971698
25%	3.048337e+07	8.051874e+07	0.889946	1.265249e+08	107.932782	6.104961
50%	4.117684e+07	1.026524e+08	0.999123	1.453704e+08	109.718896	6.271167
75%	4.698083e+07	1.635509e+08	1.159475	2.102314e+08	118.346154	6.658974
max	8.138583e+07	5.526511e+08	2.856943	6.340369e+08	167.600000	7.400000

In [486]:
```
explorations.shape
```
Out[486]:
```
(56, 6)
```

In [487]:
```
# Explorations array was created to explore some associations
explorations.head()
```
Out[487]:

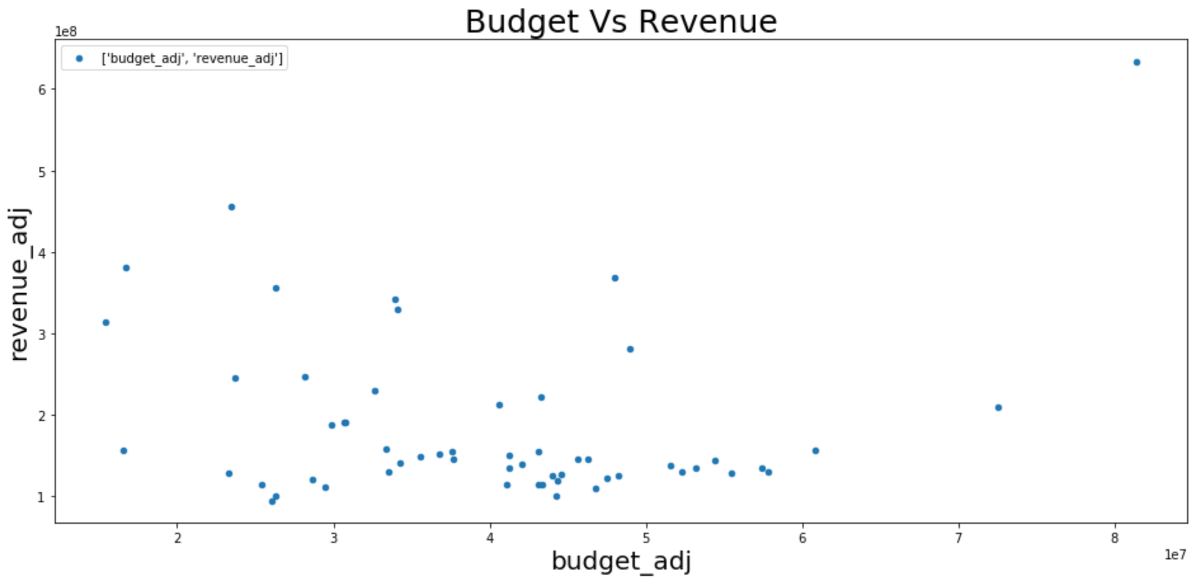
	budget_adj	net_profit	popularity	revenue_adj	runtime	vote_average
release_year						
1960	3.068179e+07	1.595481e+08	1.324513	1.902299e+08	130.000000	7.400000
1961	2.818516e+07	2.181770e+08	0.787718	2.463622e+08	132.500000	6.620000
1962	4.062476e+07	1.718493e+08	0.983485	2.124740e+08	141.285714	6.900000
1963	7.252496e+07	1.369589e+08	1.040612	2.094838e+08	153.500000	6.766667
1964	3.408189e+07	2.959526e+08	1.377790	3.300344e+08	122.428571	6.971429

In [488]:
```
# Top explore the corelation between the Runtime and the Net Profit

plotting2d (explorations, 'scatter', 'runtime', 'net_profit', 'Runtime Vs Net
 Profit')
```



NOTE: From graph we can figure out that there is a moderate positive changing correlation between runtime and net profit, and there are some outlier values.

In [489]:
```
# Top explore the corelation between the Budget and the Revenue

plotting2d (explorations, 'scatter', 'budget_adj', 'revenue_adj', 'Budget Vs R
evenue')
```



NOTE: From the graph we can figure out that there is a week negative changing correlation between budget and revenue and there are some outlier values.

Conclusions

> **From this investigation we can summarize the findings as follow:**
>
> - There is a moderate postive non-liner correlation, and there is anout lier valye.
> - there is a moderate positive changing correlation between runtime and net profit, and there are some outlier values.
> - there is a week negative changing correlation between budget and revenue and there are some outlier values.
>
> **The creteria of the successful movies and could genrate average revenue about 137 million dollar as follow:**
>
> 1. The average runtime is 151 minutes
> 2. The average budget is 44 million dollar.
> 3. The genrs should be adventure, action and science fiction.
>
> **Limitations**
>
> This report was done depending on the provided dataset which which has a missing information, also we don't know the information acuracy included in this dataset or if it is up to date or no. So, droping the rows with missing information may have affetted the analysis results in this report. On the other hand the remaining complete information if we asum they are acurate we can consider the analysis result positively whic can we depend on.

Answer_Q1:

The highest runtimes movies were from the year 2008 to 2010.

The top 10 highest runtimes movies startimg from the minimum runtimes are: [The Greatest Story Ever Told, The Godfather: Part II, The Lord of the Rings: The Return of the King, Malcolm X, Jodhaa Akbar, Gods and Generals, Lawrence of Arabia, Heaven's Gate, Cleopatra and Carlos.]

Answer_Q2:

The highest revenues movies were from the year 2009 to 2015.

The top 10 highest revenue movies startimg from the minimum revenues are: [The Avengers , One Hundred and One Dalmatians, The Net, E.T. the Extra-Terrestrial, Star Wars: The Force Awakens, Jaws, The Exorcist, Titanic, Star Wars, Avatar.]

Answer_Q3:

The highest budgets movies were from the year 2006 to 2013.

The top 10 highest budget movies startimg from the minimum budgets are: [Waterworld, Harry Potter and the Half-Blood Prince, Avengers: Age of Ultron, Tangled, Spider-Man 3, Titanic, Superman, Returns, Pirates of the Caribbean: At World's End, Pirates of the Caribbean: On Stranger Tides, The Warrior's Way.]

Answer_Q4:

The highest rating movies were from the year 2013 to 2015.

The top 10 highest rating movies startimg from the minimum rating are: [original_title, Fight Club, Forrest Gump, Pulp Fiction, Schindler's List, The Dark Knight, Godfather: Part II, Whiplash, The Godfather, Stop Making Sense and the The Shawshank Redemption.]

Answer_Q5:

The highest net profit movies were at the year 1973, 1977 and from the year 2009 to 2015 except the year 2014.

The top 10 highest net profit movies startimg from the minimum net profit are: [The Avengers, The Godfather, Jurassic Park, The Empire Strikes Back, One Hundred and One Dalmatians, E.T. the Extra-Terrestrial, Jaws, The Exorcist, Titanic and the Star Wars.]

End of the investigation.