Project_MS_Movie_Studio

Introduction:

Analysis of the following DATA SETS to launch a movie studio

• Box Office Mojo

• IMDB

• Rotten Tomatoes

• TMDB

Total 9 data sets used for cleaning with total 53 Columns and 2368367 Rows

```
In [6]: imdb_df = pd.read_csv('zippedData/imdb.name.basics.csv.gz')
        imdb_df1 = pd.read_csv('zippedData/imdb.title.akas.csv.gz')
        imdb_df2 = pd.read_csv('zippedData/imdb.title.basics.csv.gz')
        imdb_df3 = pd.read_csv('zippedData/imdb.title.crew.csv.gz')
        imdb_df4 = pd.read_csv('zippedData/imdb.title.principals.csv.gz')
        imdb_df5 = pd.read_csv('zippedData/imdb.title.ratings.csv.gz')
        tmdb_df = pd.read_csv('zippedData/tmdb.movies.csv.gz')
        bom = pd.read_csv('zippedData/bom.movie_gross.csv.gz')
        tn = pd.read_csv('zippedData/tn.movie_budgets.csv.gz')
        executed in 5.03s, finished 23:30:10 2020-09-03

In [7]: print(imdb_df.shape,imdb_df1.shape,imdb_df2.shape,imdb_df3.shape,imdb_df4.shape,imdb_df5.shape,tmdb_df.shape,bom.shape,tn.sh
        executed in 12ms, finished 23:30:10 2020-09-03

        (606648, 6) (331703, 8) (146144, 6) (146144, 3) (1028186, 6) (73856, 3) (26517, 10) (3387, 5) (5782, 6)
```

Shortlisted dataset for visualisation total 39 Columns and 181704 Rows

```
In [49]: tn = pd.read_csv("./zippedData2/tn_no_null")
         executed in 230ms, finished 06:35:32 2020-09-05

In [50]: tmdb = pd.read_csv("./zippedData2/tmdb_df_no_null")
         executed in 205ms, finished 06:35:32 2020-09-05

In [51]: bom = pd.read_csv("./zippedData2/new_bom_no_null")
         executed in 16ms, finished 06:35:32 2020-09-05

In [52]: imdb = pd.read_csv("./zippedData2/imdb_r_no_null")
         executed in 505ms, finished 06:35:33 2020-09-05

In [53]: print(tn.shape,tmdb.shape,bom.shape,imdb.shape)
         executed in 9ms, finished 06:35:33 2020-09-05

         (5782, 11) (26517, 13) (3387, 5) (146018, 10)
```

Assumptions Datasets used and other information pulled from the internet is good at the basic level of project analysis.

Questions that I have attempted to addressed with the research:

Why Should MS Invest in Movie Studio?

What should be the average runtime for maximum viewership ?

Which location the movies should be the focused on?

Why Should MS Invest in Movie Studio?

Current Industry Overview

•Revenue in the Video Streaming segment is projected to reach US$51,617m in 2020.

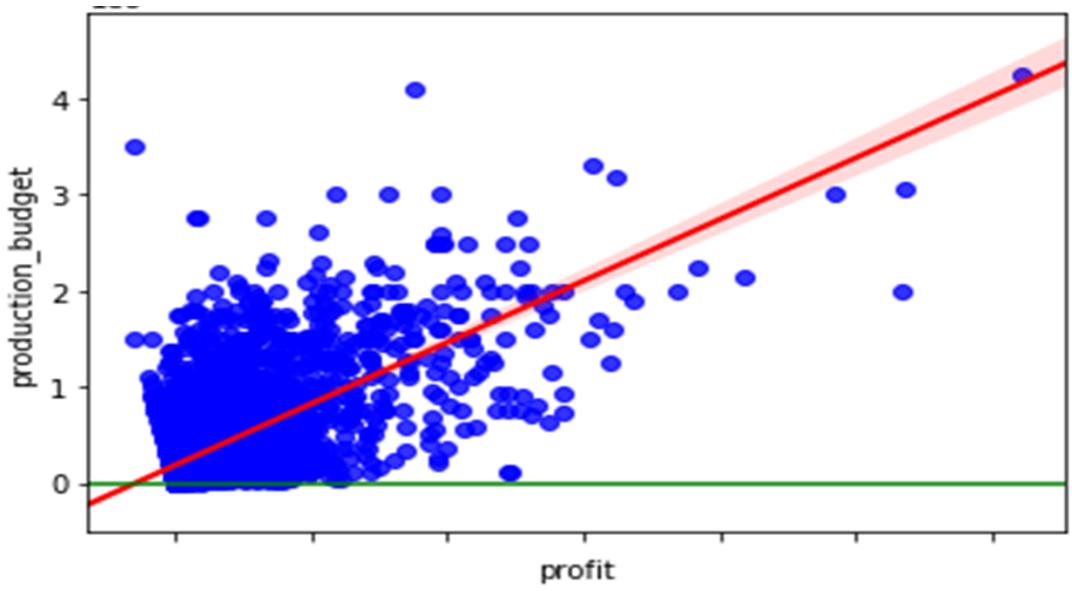
•Revenue is expected to show an annual growth rate (CAGR 2020-2025) of 10.7%, resulting in a projected market volume of US$85,735m by 2025.

•User penetration will be 11.9% in 2020 and is expected to hit 17.2% by 2025.

•The average revenue per user (ARPU) is expected to amount to US$58.52.

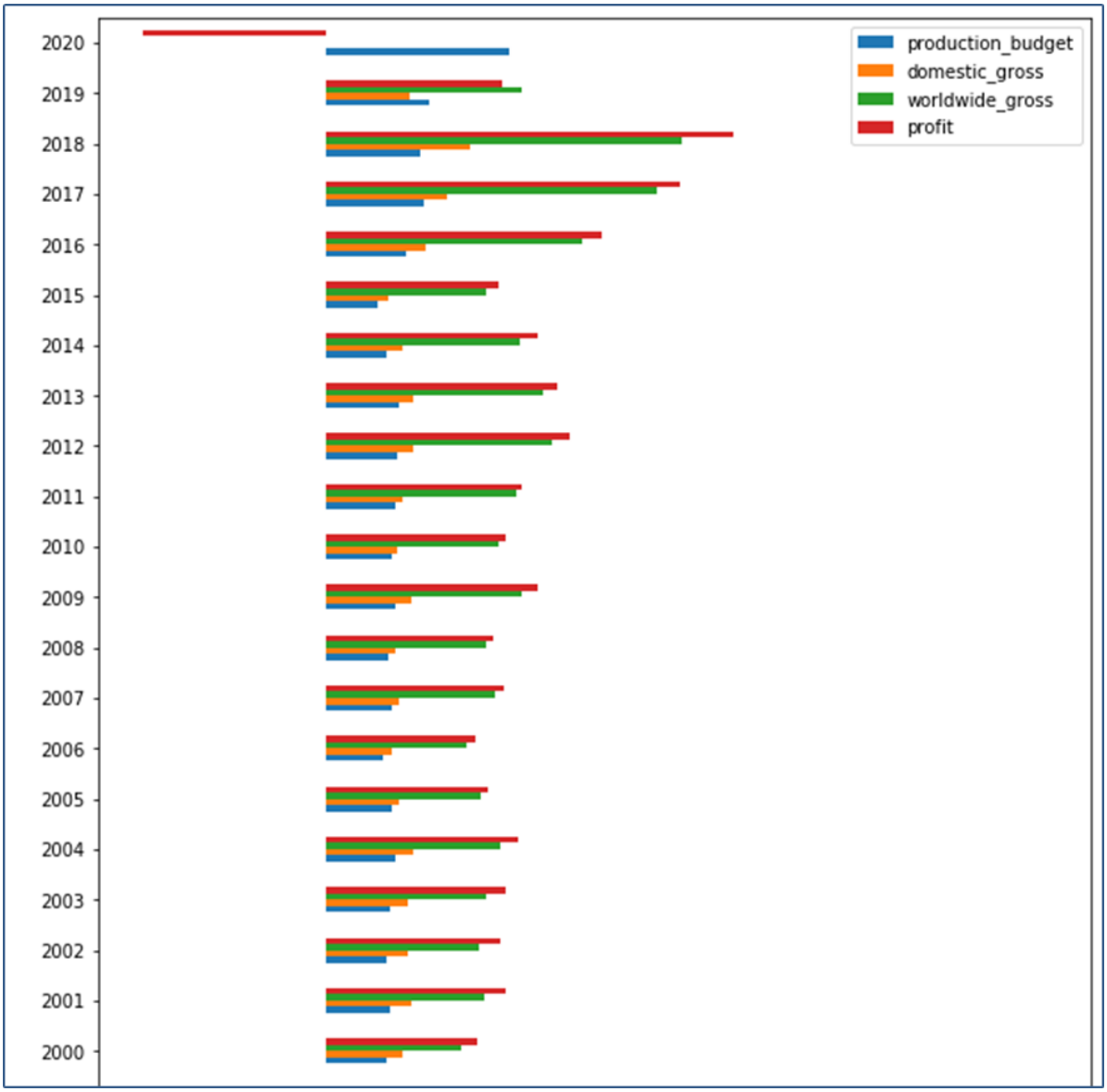
Production Budget vs Profit

The current industry overview is suggesting a huge potential to get into the market but essentially that cant be the only factor to take a decision so if we were to take into consideration the analysis based on the data the chances of making a loss is the minimal or not there at all. The gap between the production budget and the profit is really good when worldwide gross is seen.



The New Normal

We all know COVID19 has changed the outlook all , with the new normal coming into picture and people having to stay at home the source of entertainment has become limited and Internet penetration in % Estimated share of the total population in any selected region using the internet is seeing an sharp increase. To Top it all the 5G technology is pretty close to reality giving a huge boost to the untapped market as well.



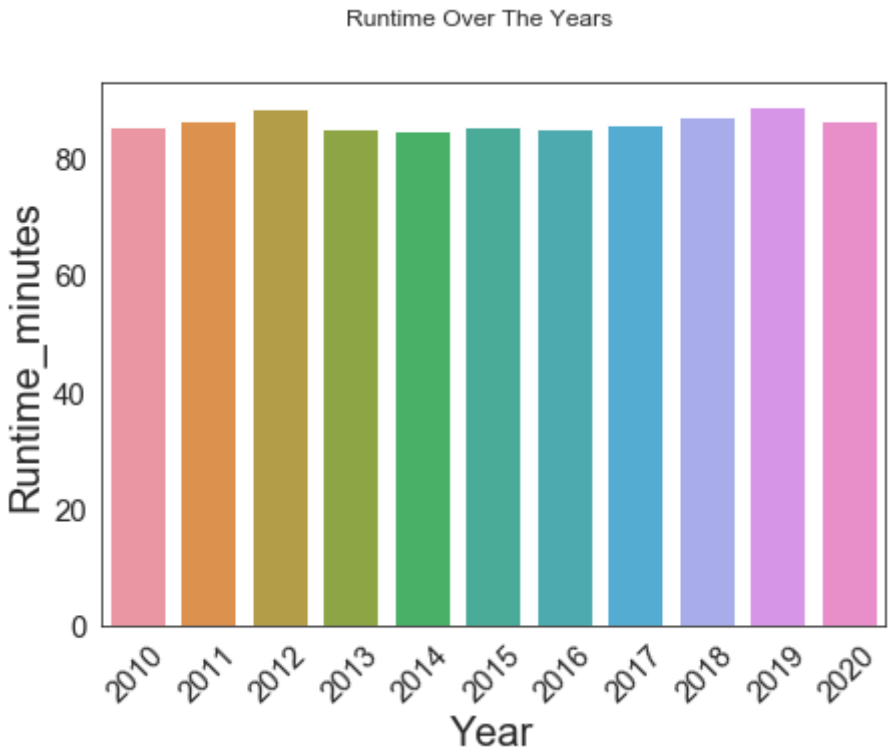
What should be the average runtime for maximum viewership ?

Based on the exploration done on the distribution movie runtimes the following observation and suggestion can be made. The distribution of movie runtimes have a skewed distribution, with median of mean of 86.18 minutes. 50% of the movies again have running time of 86.18 minutes, while 25% of the movies have 75 minutes. We can arrive at a conclusion is that a 90 minute runtime will be a safe bet for maximum viewership

```
In [240]: imdb_r['runtime_minutes'].describe()
          executed in 404ms, finished 07:57:31 2020-09-04

Out[240]: count    146018.000000
          mean         86.186338
          std         147.254673
          min           1.000000
          25%          75.000000
          50%          86.187247
          75%          95.000000
          max       51420.000000
          Name: runtime_minutes, dtype: float64
```

As mentioned above the same thing can be observed when viewing the below figure of runtime over the last ten years



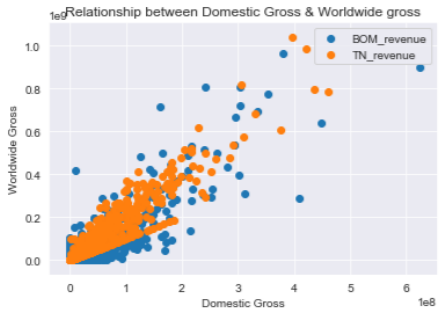
Which location the movie should be the focus on?

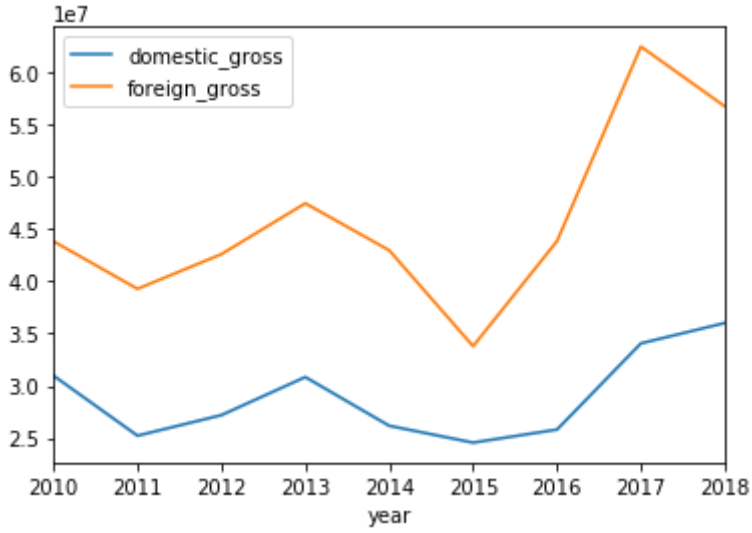
The data analysis suggest that the trend is in the worldwide market and if we were to follow the data it also suggested that a movie which has the viewership of worldwide is either at par with the production budget or has a higher profit. Also one more observation that comes with different dataset is the worldwide gross is higher than the domestic gross. The availability of the technology and the global audience clearly indicates towards making films considering the global market.

Comparing foreign gross and domestic gross from new_bom & tn DFs

In [233]:
```
plt.scatter(new_bom['domestic_gross'][1:1000],new_bom['foreign_gross'][1:1000], label='BOM_revenue')
plt.scatter(tn['domestic_gross'][1:1000],tn['worldwide_gross'][1:1000], label='TN_revenue')
plt.title('Relationship between Domestic Gross & Worldwide gross')
plt.xlabel('Domestic Gross')
plt.ylabel('Worldwide Gross')
plt.legend()
plt.show()
```

executed in 436ms, finished 23:36:39 2020-09-03





Conclusion:

1. This is the right time to get into movie studio
2. A good 90 minute movie has been the trend for viewership
3. There is increase in the population that demands good quality movies at their disposal any where from the globe.
4. The internet penetration and 5G being a closer dream global market is the way to go and get into the untapped market.
5. The New Normal has increased the demand for Video Streaming

Email: ksis1st@hotmail.com

Github:https://github.com/ksis1st/Project_MS_Studio