Movie Analysis pdf

Minh K. Chau

2023-07-22

About this project

In this project, I used the Movies data, provided by a course in Google Data Analytics certificate on Coursera. This data table includes movies titles, director, five cast members, released years, budgets, and revenues released from 2012 - 2016. I address the following questions in my analysis:

- What are the top 10 directors and cast members for movies that generated the most revenue, from 2012 - 2016?
- What is the most profitable genre, by year and of all time?
- Can we predict revenue just from budget for a particular movie?

Data At first glance

At first glance, we have 508 rows with 15 columns. There are no missing values. No Data Cleaning process is necessary.

Top 10 Director and First Cast Member

```
## # A tibble: 10 x 2
##    director         sum.in.millions
##    <chr>                      <dbl>
##  1 Chris Renaud               2044.
##  2 Zack Snyder                1541.
##  3 Francis Lawrence           1409.
##  4 Bryan Singer               1292.
##  5 Steve Martino              1123.
##  6 Ridley Scott               1034.
##  7 Justin Lin                 1032.
##  8 Phil Lord                  1002
##  9 Peter Jackson               956
## 10 David Ayer                  914.
```

The top 10 directors whose movies received the highest revenues returned are shown by the table above. How about Cast Member?

```
## # A tibble: 10 x 2
##    cast1             sum.in.millions
##    <chr>                       <dbl>
##  1 Jennifer Lawrence           2204.
##  2 Tom Cruise                  1914.
##  3 Hugh Jackman                1733.
##  4 Ben Affleck                 1504.
##  5 Adam Sandler                1451.
##  6 Will Smith                  1311
##  7 Matt Damon                  1272.
##  8 Ray Romano                  1245.
##  9 Bradley Cooper              1209.
## 10 Ryan Reynolds               1174
```

The table above showed the top 10 actors/actresses whose movies that were casting on generated the highest sum of revenues.

What genre generates the most profit for each year in 2012 - 2016?

```
## # A tibble: 5 x 3
## # Groups:   year [4]
##    year genre  sum.in.millions
##   <int> <chr>            <dbl>
## 1  2014 Action           4225.
## 2  2016 Action           4215.
## 3  2012 Action           3947.
## 4  2013 Action           3897.
## 5  2012 Comedy           2829.
```

Action movies generated the most profit for: 2012, 2013, 2014, 2016. So the next question is: ***Does that mean Action moves generate the most profit per movie released?***

```
## # A tibble: 5 x 3
## # Groups:   year [3]
##    year genre     avg.in.mil
##   <int> <chr>          <dbl>
## 1  2014 Adventure       352.
## 2  2012 Adventure       331.
## 3  2013 Thriller        298
## 4  2012 Fantasy         282.
## 5  2013 Adventure       246.
```

According to the output, **Action movies do NOT generate the most profit per movie released in any year**.

Even though action movies generated the most profits for each year from 2012 - 2016 (except 2015), each action movie does not generate the most profit on average, in comparing to other genres.

To explain this contradiction, let's look at the number of movies released by year for each genre.

```
## # A tibble: 5 x 3
## # Groups:   year [4]
```

```
##    year genre  num.movies.released
##   <int> <chr>                <int>
## 1  2012 Comedy                  28
## 2  2015 Drama                   28
## 3  2012 Action                  27
## 4  2013 Action                  26
## 5  2014 Action                  26
```

Based on this number, we can see that **Action movies stand in the top 5 in terms of number released (i.e., 2012, 2013, 2014)**. How about In terms of budget and Revenue?

```
movies %>%
  select(year,
         genre,
         budget.mils) %>%
  group_by(year,genre) %>%
  summarize(sum = sum(budget.mils)) %>%
  arrange(desc(sum)) %>%
  head(5)
```

```
## # A tibble: 5 x 3
## # Groups:   year [5]
##    year genre    sum
##   <int> <chr>  <dbl>
## 1  2013 Action 2212
## 2  2016 Action 2202
## 3  2014 Action 2116.
## 4  2012 Action 1958.
## 5  2015 Action 1449
```

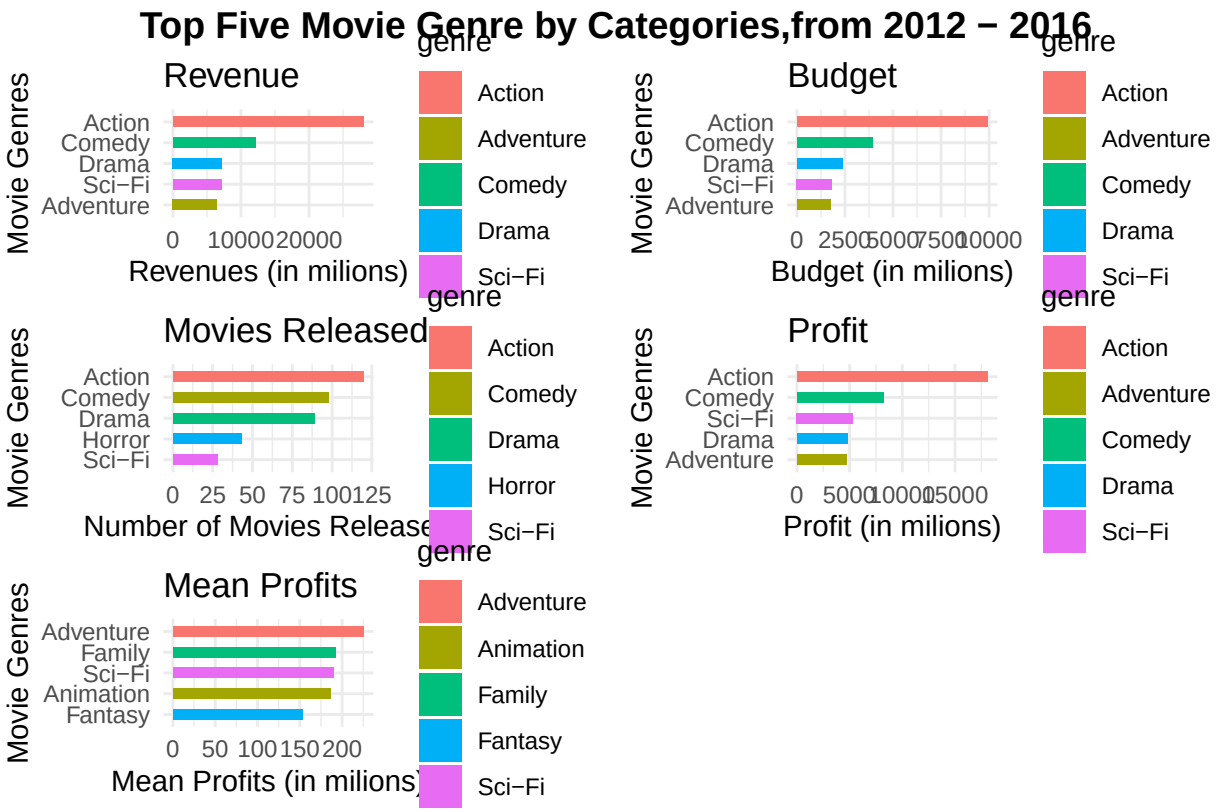
From 2012 - 2016, action movies received the most budget.

```
## # A tibble: 5 x 3
## # Groups:   year [4]
##    year genre  sum.in.millions
##   <int> <chr>            <dbl>
## 1  2016 Action           6417.
## 2  2014 Action           6341
## 3  2013 Action           6109.
## 4  2012 Action           5905.
## 5  2012 Comedy           4151.
```

Except for 2015, Action movies made the most revenues.

so let's look at this trend as a whole.

From 2012 - 2016, what genres generate the most revenues, receive the most budget, release the most movies, generate the most net profit, and average profit/movie?



Results

Action movies do not generate the most profit per movies. But they do generate the most Revenues and tend to be the most popular genre (e.g., with more budgets and revenues).

Let's confirm this finding: ***what are the top 10 movies that generate the most revenues?***

```
## # A tibble: 10 x 4
## # Groups:   genre [5]
##    title                                    genre      year revenue.mils
##    <chr>                                    <chr>     <int>        <dbl>
##  1 Despicable Me 2                          Comedy     2013         971.
##  2 The Hobbit: The Battle of the Five Armies Adventure  2014         956
##  3 Ice Age: Continental Drift               Adventure  2012         877
##  4 Batman v Superman: Dawn of Justice       Action     2016         873.
##  5 The Twilight Saga: Breaking Dawn - Part 2 Fantasy    2012         830.
##  6 Fast & Furious 6                         Action     2013         789.
##  7 Deadpool                                 Action     2016         783.
##  8 The Amazing Spider-Man                   Action     2012         758.
##  9 The Hunger Games: Mockingjay - Part 1    Sci-Fi     2014         755.
## 10 X-Men: Days of Future Past               Action     2014         748.
```

We can see that 5/10 movies with the highest revenues belong to Action movies. This confirm our hypothesis that Action Movies are the most profitable genre.

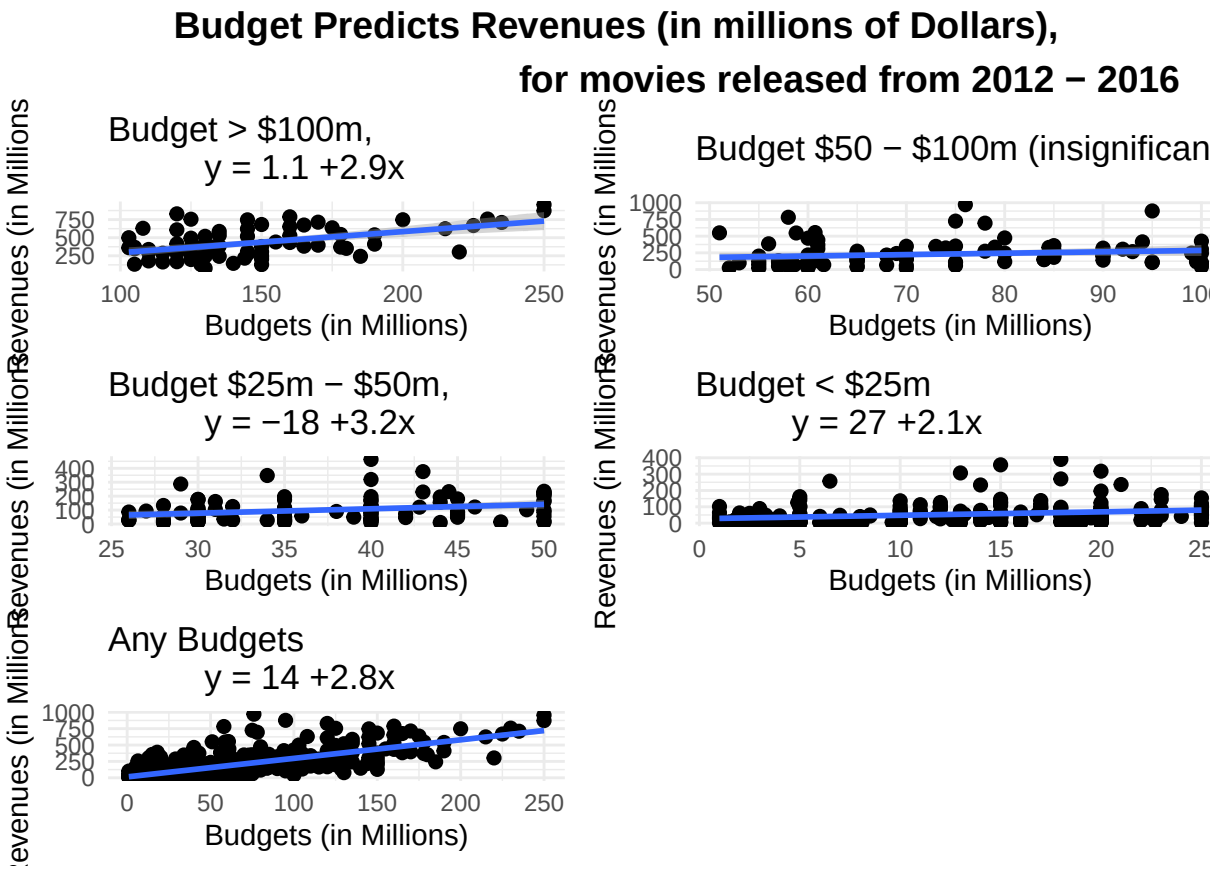
Can we predict Revenue based on Budget?

For this question, I use linear regression test to predict Revenue by Budget for each movies, regardless of genre, for:

1. All movies released
2. Movies with revenue greater than $100 millions
3. Movies with revenue Between $100 millions and $50 millions
4. Movies with revenue between $50 millions and $25 millions
5. Movies with revenue less or equal to $25 millions

The following insights were identified:

1. **All movies, on average:** for every $1 million increase in budget, revenue is expected to increase by $2.83 million.
2. **Movies with revenue greater than $100 millions:** for every $1 million increase in budget, revenue is expected to increase by $2.92 millions.
3. **Movies with revenue Between $100 millions and $50 millions**: There exists no relationship between budgets and revenue returned (in million of dollars).
4. **Movies with revenue between $50 millions and $25 millions**: for every $1 million increase in budget, revenue is expected to increase by $3.17 millions.
5. **Movies with revenue less or equal to $25 millions:** For every $1 million increase in budget, revenue is expected to increase by $2.11 millions.



Put it together

When analyzed by different budget groups, budget significantly predict revenues returned by each movies, across genre, from 2012 - 2016. This regression coefficient does not apply to movies with budget between $50 - $100 however.

TAKE-HOME MESSAGE

- Action movies generated the most profits and revenues. With lots of budgets, it is the most popular genre.
- It is best to have a movie with a budget around $25 - $50 millions, or $100 millions or more, for those movies to generate profits.