uc3m | Universidad Carlos III de Madrid

Final project: Movies Dataset

Advanced Knowledge of Spreadsheets

Group 96

Aya Ben Hriz and Alejandro Bermejo Gordillo

Index

Introduction

It has been a while since the Lumière brothers publicly screened ten of their short films in 1895. The film industry has now become a gigantic industry all over the world. The USA is the country that we mainly associate it with.

What makes American films that great? We also wonder if there is any country that might dethrone Hollywood as the Mecca of the cinematography industry. Additionally, we want to discover if the film industry is profitable or not.

We found Pivot Tables to be the best way to answer the questions we posed. We present our results below:

EDA

1st pivot table: Movies produced by country

Thanks to the pivot tables, we first created one with all the countries and number of movies produced so that we know who are the pioneers.We discovered that those USA, UK, France and Canada.

We will focus on those countries since the gap between them and the others is considerable and the data is more interesting for our top 4.

2nd pivot table: Max IMDB score

We wanted to check next if producing more movies leads to have better chances to have a high imdb_score so we used a pivot table with top 4 countries as columns, years on which countries made the highest imdb_score >=9 of all the history.
We found out that the ranking is still the same and that our hypothesis is confirmed.

3rd pivot table: Total budget used

We then wanted to check another hypothesis. That is, are countries with higher imdb_score investing more money and have higher budget? Or does simple movies with small budget may lead to success too?

For that we use a pivot table with same columns and rows as the second pivot table and sum of budget as value.

We conclude that the ranking is still the same, our 1st hypothesis is confirmed , the higher the budget the higher chances we have to be top1.

Solver:

We are interested now after our latest discovery, in predicting the budget that the other countries behind USA should've invest to beat USA and be the top 1.

Dashboard

The dashboard will allow us to compare key metrics by country in an intuitive manner.
We utilized a combo box drop down list for the selection of countries. The list values were obtained from a pivot table that just had the countries as rows. A selection from the list will invoke the execution of the following macro:

```
Sub DropDown()
    Dim ws As Worksheet
    Dim dd As DropDown
    Dim countryFilter As String

    Set ws = ActiveSheet
    Set dd = ws.Shapes("Drop Down 4").OLEFormat.Object

    countryFilter = dd.List(dd.ListIndex)

    Worksheets("DB - Pivot Tables").PivotTables("PivotTable5").PivotFields("country").CurrentPage = countryFilter
    Worksheets("DB - Pivot Tables").PivotTables("PivotTable1").PivotFields("country").CurrentPage = countryFilter
    Worksheets("DB - Pivot Tables").PivotTables("PivotTable2").PivotFields("country").CurrentPage = countryFilter
    Worksheets("DB - Pivot Tables").PivotTables("PivotTable3").PivotFields("country").CurrentPage = countryFilter
End Sub
```

The program first obtains the value selected and then filters each pivot table accordingly. Our pivot tables serve us for two purposes:

Statistics:

We obtain the values from the pivot chart at the top since the cell's positions don't change after filtering.

- **Number of movies**: it is really useful to put other metrics into context and understand the size of the industry in that region.

- **Average duration**: it can be related to other fields like budget and quality. Using `TIME()` and `TRUNC()` we converted it into time value and formatted the cell as such.
- **IMDB score**: it is a good estimator of quality. The standard deviation allows us to determine if the quality levels are stable for all films.
- **Year**: the first year with movies serves as an indicator of the maturity of the industry. The comparison with the average year of films tells us if they produce more films before or now. We applied `INT()` to the average year to make it easier to see and understand.
- **Profit**: with this metric we can see whether the movie industry is beneficial to the country's economy or not. It's a computed pivot table field of `Revenue-Budget.`

Pivot Charts:

Using time series, we create charts that show the evolution of the industry by country.

- **Quality and budget size**: this allows us to visualize the relation between IMDB score and Budget on a per-country basis. So it is not biased by other factors in a global market such as cultural differences.
- **Number of movies per year**: this shows us the periods when the industry grows and shrinks the output of movies.
- **Percentage of profit from the budget**: We created this computed pivot field (`Revenue/gross-1`) that yield us the percentage of the budget that has been a profit. It allows us to evaluate if the profit stream is stable and if the trend is for the industry to thrive or become stagnant with low or negative profit.

Some of the charts don't present themselves as we saved them as. These known issues are the existence of labels in the movies/year chart and the bar color for negative values in the % profit chart, which is supposed be red.

Self-Assessment Form

ID	Applicability
1	Discover the film industry superpowers. Get insight of what characteristics define a country's film industry and its evolution.
2	Which country is the largest movie producer? Are budget and quality (IMDB score) related? How much should a country invest to become the biggest movie producer? Is a country's film industry profitable?
3	We have selected the following indicators: Duration, Gross Revenue, Title, Language, Country, Budget, Year and IMDB Score
4	Data obtained from the example datasets in the course's Aula Global

5	Data is contained in an .xls file. We defined the dataset table in the first worksheet. We set two worksheets for our pivot tables for each of our objectives. We created one sheet for our data analysis, solver and dashboard.
6	Each movie has: ● Text data ○ Title ○ Categorical ■ Language ■ Country ● Numerical data ○ Duration (integer in minutes) ○ Gross Revenue (does not consider inflation) ○ Budget (does not consider inflation) ○ Year ○ IMDB Score (10-point scale)
7	We will remove all the null values since we display all of them in the dashboard
8	We have almost 4000 movies from over 90 years, so we expect our data to lead us to interesting conclusions. There are anomalies in the country column.
9	We removed entries which had any blank value using Special Find & Select. We also deleted rows which had non-country values in the country field.
10	Answered above.
11	Not applicable.
12	✓

Conclusion

We then conclude that Hollywood is not only considered the oldest film industry where earliest film studios and production companies emerged, and the birthplace of various genres of cinema—among them comedy, drama, action, the musical, romance, horror, science fiction, and the war epic, but also the film industry symbol of prosperity , diversity and wealth.