Marvel vs DC US Box Office Movie Comparison

OVERVIEW:

This report provides a fun example of data storytelling.

Users can select two movies from Marvel and DC to compare with the slicers at the top of the page. Slicer values update based on selection, so that a movie cannot be compared with itself.

Smart Narratives are used to describe the box office performance in plain English.

Donut Charts allow for a view of Domestic and Worldwide Box Office data as a whole per movie franchise.

DATA GATHERING:

Box Office data was sourced from Kaggle and The Numbers:
https://www.kaggle.com/datasets/davidgdong/marvel-cinematic-universe-box-office-dataset
https://www.the-numbers.com/movies/franchise/Marvel-Cinematic-Universe#tab=summary
https://www.the-numbers.com/movies/franchise/DC-Extended-Universe#tab=summary

RottenTomatoes data was sourced from RottenTomatoes:
https://www.rottentomatoes.com/

Image URL data was sourced from Marvel and DC Official websites:
https://www.marvel.com/movies
https://www.dc.com/movies

IMPORTING DATA:

Marvel Box Office data was originally sourced on Kaggle. It did not include the latest 3 Marvel movies, so I imported the data into Microsoft Excel before transforming the data to include Thor: Love and Thunder and Doctor Strange in the Multiverse of Madness.

DC Box Office data was sourced on The Numbers, and I created a table on Microsoft Excel to store all 11 DC movies.

Image URL data was sourced on Marvel and DC official websites, and I created two separate tables on Microsoft Excel to store the Movie Title to create a relationship with other tables and Image URL.

I loaded the data onto Power BI Desktop using the 'Get Data' option available importing the csv. files.

TRANSFORMING DATA:

Image URL Data:

Using First Row as Headers: The image URL data did not recognise the first row as a header in both tables. This was done by using the 'Transform' tab on the 'Home' page of Power BI Desktop.

Box Office Data:

New Columns: I wanted to present my Tomatometer Rating and Audience Score as Percentages. So, I added new custom columns and divided both columns by 100.

I also added a new column total_box_office. This was done using 'Custom Column' using two columns already in the table as reference([domestic_box_office]+[worldwide_box_office]).

Deleting Columns: I deleted the original tomato_meter and audience_score columns from both the Marvel and DC Box Office Data tables. The Marvel Box Office Data also contained the mcu_phase column which would not be needed for our analysis, so it was deleted.

Changing Data Types: The following columns were changed from whole number's to fixed decimal numbers in both Box Office Data tables: production_budget, opening_weekend, domestic_box_office, worldwide_box_office and total_box_office. The tomato_meter and audience_score columns were changed from whole numbers to percentages in both Box Office Data tables.

Changing Data Type with Locale: The release_date was imported in the US date format mm/dd/yyyy. Although the Box Office is US based, this analysis is for a UK audience. So, I changed the data type with Locale from Text to Date.

Grouping Tables: I grouped all my tables into folders to make them clearer. I had already renamed all tables in Excel.

Change of Currency: Once all the data was loaded, I changed the following columns in both Box Office Data tables: production_budget, opening_weekend, domestic_box_office and worldwide_box_office from £ into $ with the use of Colum Tools. In addition, I also changed the tomato_meter and audience_score into percentages.

DATA MODELLING:

After data was transformed into the Power BI desktop. I double checked all relationships between my tables. Power BI had already established a 1 to 1 relationship between the relevant Box Office Data Tables and the relevant Image URL Tables.

VISUALISING THE DATA:

Creating Measures Using DAX: I created 4 measures, First Week Sales % and Total Sales for both the Marvel and DC data.

Creating New Table/Grouping Tables: I created a new table 'Key Measures' which was used to store all my measures. I grouped all my new measures into this table before deleting the default column1 table.

Slicers: I created 2 slicers using the movie title information for both sets of data. Slicers are a great choice when you want to: display important filters on the report canvas for easier access. I changed the slicer settings to only allow single select. This is a movie comparison report, users should not be able to select more than one movie per franchise.

Smart Narrative Text Boxes: I created 2 smart narrative text boxes, this visualisation allowed me to quickly summarise customised data, dependant on what movie is selected.

Simple Image Visual: I used the 'Get more visuals' option and added 2 'Simple Image' visualisations to my dashboard. This allowed me to illustrate the movie cover art on my dashboard, using the URLs, I imported.

Donut chart: 2 donut charts were added to present the domestic box office sales in comparison to the worldwide box office sales per movie.

Reformatting data: Lastly, I reformatted all the data to make my report more visually appealing.

CONCLUSION:

Although this report does not drive much action, I was happy with the design element of this report. I aimed to make this as visually appealing and as a story-telling dashboard as opposed to an informative dashboard.