MORINGA
Discover · Grow · Transform

Phase 1 - Project

Anthony Githinji Ngatia

INtroduction

Microsoft sees all the big companies creating original video content and they want to get in on the fun. They have decided to create a new movie studio, but they don’t know anything about creating movies.

I was charged with exploring what types of films are currently doing the best at the box office. I then translated those findings into actionable insights that the head of Microsoft's new movie studio can use to help decide what type of films to create.

Objectives

- To determine the highest and lowest grossing movies and their ratings based on the available datasets.
- To determine the most and least produced genres of films.
- To determine the highest and lowest grossing movie genres.
- To determine the average runtime and period to produce a movie.
- To determine the relationship between the relationship between the average rating of customers based on several factors such as: the average production time, the release year of the movie and the runtime.

Agenda

1. Import and understand the datasets
2. Data Cleaning and Visualization
3. Conclusions and Recommendations

Dataset Structure

The datasets contain different datasets but on combining them into one, the final table resembles as shown

title	primary_title	start_year	runtime_minutes	genres	averagerating	numvotes	studio	domestic_gross	foreign_gross	year
'71	'71	2014	99.0	Action,Drama,Thriller	7.2	46103.0	RAtt.	1300000.0	355000.0	2015
1,000 Times Good Night	1,000 Times Good Night	2013	117.0	Drama,War	7.1	6848.0	FM	53900.0	18700000.0	2014
10 Cloverfield Lane	10 Cloverfield Lane	2016	103.0	Drama,Horror, Mystery	7.2	260383.0	Par.	72100000.0	38100000.0	2016
10 Years	10 Years	2011	100.0	Comedy,Drama, Romance	6.1	22484.0	Anch.	203000.0	18700000.0	2012
1001 Grams	1001 Grams	2014	93.0	Drama	6.3	1301.0	KL	11000.0	18700000.0	2015