Task : IMDB Movie Analysis

Analysis done on the following points:-

Case Study 1 : Job Data

A. Movie Genre Analysis: Analyze the distribution of movie genres and their impact on the IMDB score.
Task: Determine the most common genres of movies in the dataset. Then, for each genre, calculate descriptive statistics (mean, median, mode, range, variance, standard deviation) of the IMDB scores.
B. Movie Duration Analysis: Analyze the distribution of movie durations and its impact on the IMDB score.
Task: Analyze the distribution of movie durations and identify the relationship between movie duration and IMDB score.
C. Language Analysis: Situation: Examine the distribution of movies based on their language.
Task: Determine the most common languages used in movies and analyze their impact on the IMDB score using descriptive statistics.
D. Director Analysis: Influence of directors on movie ratings.
Task: Identify the top directors based on their average IMDB score and analyze their contribution to the success of movies using percentile calculations.
E. Budget Analysis: Explore the relationship between movie budgets and their financial success.
Task: Analyze the correlation between movie budgets and gross earnings, and identify the movies with the highest profit margin.

Software used : Microsoft Excel

Cleaning the data
Your task: Clean the data

The most difficult and crucial step of any Data analysis project.
The process included in this step varies from question to question and Dataset to Dataset. To clean the dataset we will be:-
1. First dropping the columns which have no use for the analysis that we will be doing
2. Columns like 'Color', 'director_facebook_likes', 'actor_3_facebook_likes,'actor_2_name', 'actor_1_facebook_likes', 'cast_total_facebook_likes, 'actor_3_name', 'facenumber_in_posts', 'plot_keywords', 'movie_imdb_link', 'content_rating' ,'actor_2_facebook_likes', 'aspect_ratio', 'movie_facebook_likes' are the columns containing irrelevant data for the analysis tasks provided. So, these columns needs to be dropped.
3. After dropping the irrelevant columns now we need to remove the rows from the dataset having anyone of its column value as blank/NULL
4. Then we need to get rid off the duplicate values in the dataset which can be achieved by using the 'Remove Duplicate Values/Cells' available in the 'Data' tab.
5. After cleaning data we are ready for our analysis.

Movie Genre Analysis

A. Movie Genre Analysis: Analyze the distribution of movie genres and their impact on the IMDB score.
Task: Determine the most common genres of movies in the dataset. Then, for each genre, calculate descriptive statistics (mean, median, mode, range, variance, standard deviation) of the IMDB scores.

To determine the most common genres of movies:
1. First we will create pivot table .
2. Then in rows section we will select genres column.
3. In values section we will select imdb column and will change value field setting as per our need .
4. For median and mode values we will create formulas by adding measures in pivot table.

Top 5 common genres of movies :

Genre	Count of genres	Average of imdb_score	median	Max of imdb_score	Min of imdb_score	Varp of imdb_score	StdDevp of imdb_score	MODE
Drama	153	7.041830065	7.2	8.8	3.4	0.682563971	0.826174298	6.35
Comedy|Drama|Romance	151	6.494701987	6.5	8	4.3	0.559044779	0.747692971	6.2
Comedy|Drama	147	6.583673469	6.7	8.8	3.3	0.729801472	0.854284187	6.4
Comedy	145	5.840689655	6	8	1.9	1.471654697	1.213117759	6.8
Comedy|Romance	135	5.896296296	6	8.4	2.7	0.762578875	0.873257622	5.9
Drama|Romance	119	6.952941176	7.1	8.1	4.1	0.553415719	0.743919162	6.1

Movie Duration Analysis

B. Movie Duration Analysis: Analyze the distribution of movie durations and its impact on the IMDB score.
Task: Analyze the distribution of movie durations and identify the relationship between movie duration and IMDB score.
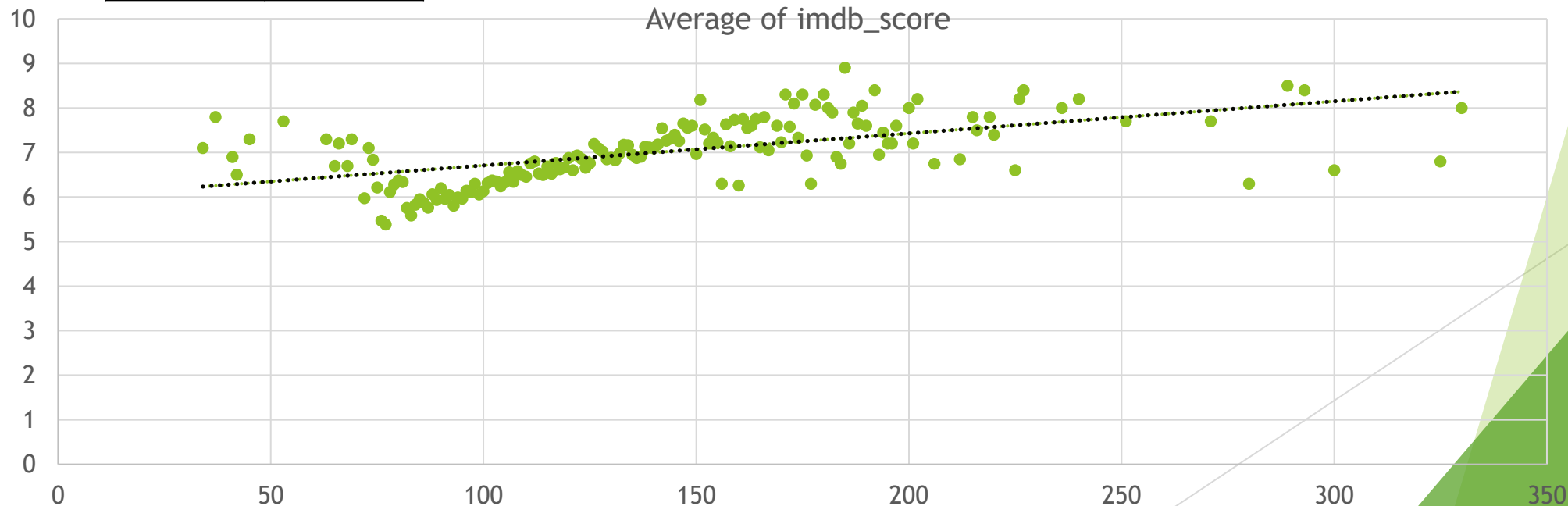
To determine movie durations and its impact on the IMDB score :
1. First we will create pivot table .
2. Then in rows section we will select duration column.
3. In values section we will select imdb column and will change value field setting to Average and we will get Average of imdb_score for each duration

Result :

Avg	143.4359
Median	138.5
Std	58.23314

of movie duration.

Average of imdb_score

Language Analysis

C. Language Analysis: Situation: Examine the distribution of movies based on their language.
Task: Determine the most common languages used in movies and analyze their impact on the IMDB score using descriptive statistics.

To determine movies based on language:
1. First we will create pivot table .
2. Then in rows section we will select language column.
3. In values section we have multiple option to select i.e count of either movie title or language .we can select any one.
4. For imdb impact on language we wil select imdb column and will change value field setting as per our need .

Top 5 languages used in movies:

Language	Number of movie for each language	StdDevp of imdb_score	median	Average of imdb_score
English	3669	1.048553076	6.5	6.423684928
French	37	0.553691378	7.2	7.286486486
Spanish	26	0.810151933	7.15	7.05
Mandarin	14	0.737930089	7.25	7.021428571
German	13	0.615769111	7.7	7.692307692

Director Analysis

D. Director Analysis: Influence of directors on movie ratings.
Task: Identify the top directors based on their average IMDB score and analyze their contribution to the success of movies using percentile calculations.

To top directors based on their average IMDB score :
First we will create pivot table .
1. Then in rows section we will select language column.
2. In values section we will select imdb column and will change value field setting as per our need .
3. For percentile we will find percentile for each director on basis of average imdb score.

Top 5 directors based on their average IMDB score :

Director	Average of imdb_score in percentage	Average of imdb_score
Charles Chaplin	133%	8.6
Tony Kaye	133%	8.6
Alfred Hitchcock	131%	8.5
Damien Chazelle	131%	8.5
Majid Majidi	131%	8.5

Budget Analysis

E. Budget Analysis: Explore the relationship between movie budgets and their financial success.
Task: Analyze the correlation between movie budgets and gross earnings, and identify the movies with the highest profit margin.

To identify the movies with the highest profit margin:
First we will create profit column to find profit of each movies by subtracting budget from gross earning and then we will create pivot table .
1. Then in rows section we will select movie_title column.
2. In values section we will select profit column
3. After that we will sort data in descending order on basis of profit.

Correlation between movie budgets and gross earnings: 0.100880803

Top 5 movies with the highest profit margin:

Row Labels	Sum of Profit
AvatarÂ	523505847
Jurassic WorldÂ	502177271
TitanicÂ	458672302
Star Wars: Episode IV - A New HopeÂ	449935665
E.T. the Extra-TerrestrialÂ	424449459

Movie Genre Analysis

Hence, all the questions given as a part of Data Analytics Trainee Task 5: IMDB Movie Analysis (Final Project - 1) have been provided with answers .

In this task all the concepts regarding to Excel and statistics like sort, filter, pivot-table,etc. have been implemented using Microsoft Excel.

From this project , I came to know that how import is pivot table for Analytics.

Excel Link IMDB_Movies