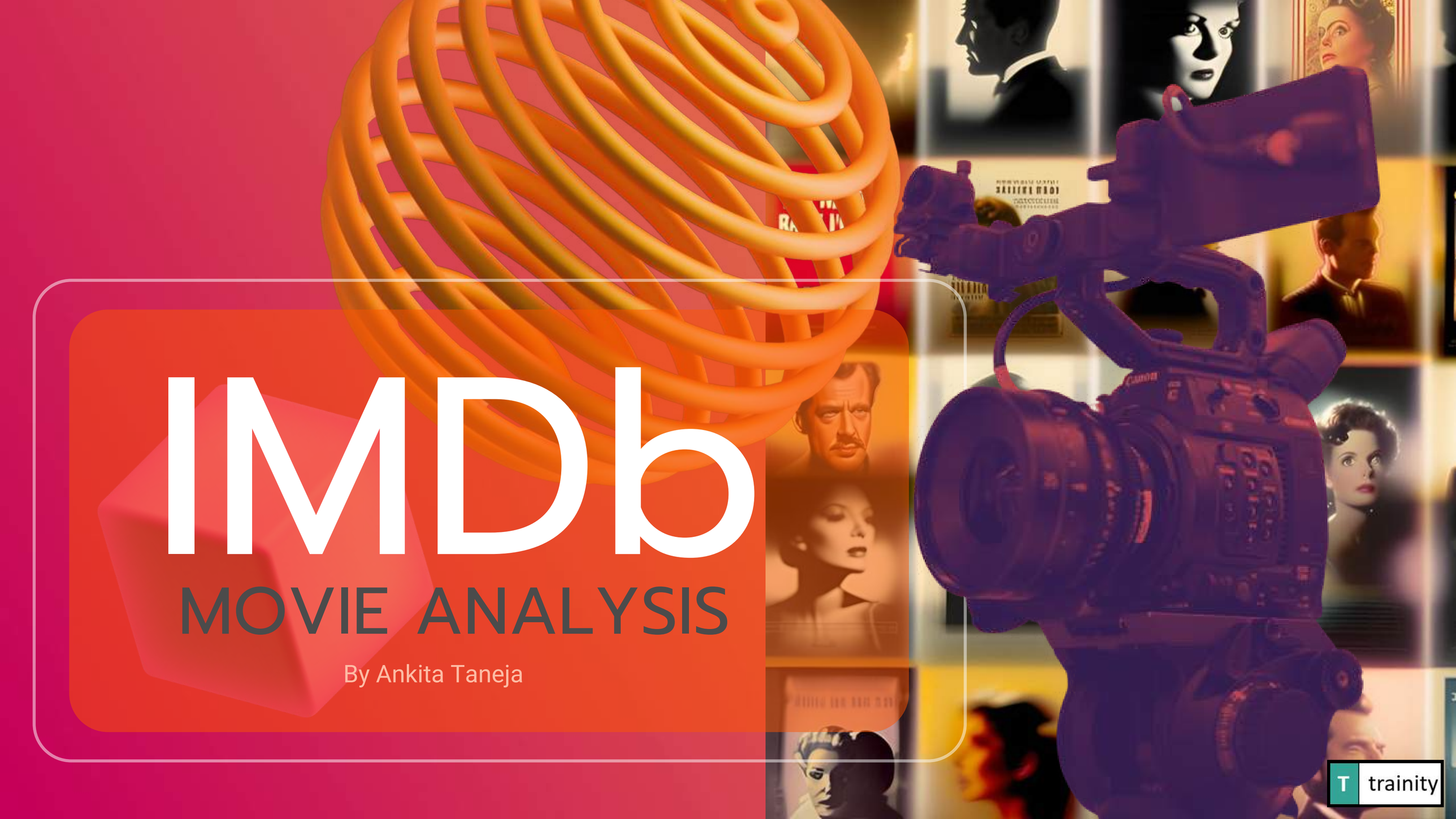
IMDb

MOVIE ANALYSIS

By Ankita Taneja



Contents

01 **Introduction**

Problem Description

02 **Tasks**

Tasks Description

03 **Outputs**

Screenshots of Outputs

Content

Welcome

Thank you for considering my work. I am excited to work with Trainity to create a project that meets the needs and exceeds the expectations.

Introduction

01

PROBLEM DESCRIPTION

Project Description

IMDb MOVIE ANALYSIS

Problem

A potential research question for the provided IMDB Movies dataset could be: "What factors contribute to a movie's success on IMDB, as measured by high ratings?"
Understanding these factors is crucial for movie producers, directors, and investors aiming to identify the key elements that drive audience appreciation and critical acclaim. Insights derived from this analysis can guide future project decisions, optimize resource allocation, and enhance the likelihood of creating successful films.

Data Cleaning

This step focuses on preparing the dataset for analysis by addressing any issues that could affect the accuracy of the results. Tasks include filling or removing missing values, eliminating duplicate records, adjusting data types as needed, and creating new features to enhance the dataset's usefulness.

Data Analysis

In this stage, you'll investigate the dataset to uncover patterns and relationships between variables. For instance, you might analyze how movie ratings are influenced by factors like genre, director, budget, release year, and cast. This step aims to identify key drivers of movie success and provide actionable insights for stakeholders.

Data Cleaning - Removing null/ Redundant data/ Duplicate Data

color	director_name	num_critic_for_reviews	duration	director_facebook_likes	actor_3_facebook_likes	actor_2_name	actor_1_facebook_likes	gross	genres	actor_1_name	movie_title	num_vo
Color	James Cameron	723	178	0	855	Joel David Moore	1000	7.61E+08	Action|Adv	CCH Pounder	AvatarÂ	
Color	Gore Verbinski	302	169	563	1000	Orlando Bloom	40000	3.09E+08	Action|Adv	Johnny Depp	Pirates of the C	
Color	Sam Mendes	602	148	0	161	Rory Kinnear	11000	2E+08	Action|Adv	Christoph Waltz	SpectreÂ	
Color	Christopher Nolan	813	164	22000	23000	Christian Bale	27000	4.48E+08	Action|Thr	Tom Hardy	The Dark Knigh	
	Doug Walker			131		Rob Walker	131		Documentı	Doug Walker	Star Wars: Epi	
Color	Andrew Stanton	462	132	475	530	Samantha Morton	640	73058679	Action|Adv	Daryl Sabara	John CarterÂ	
Color	Sam Raimi	392	156	0	4000	James Franco	24000	3.37E+08	Action|Adv	J.K. Simmons	Spider-Man 3Â	
Color	Nathan Greno	324	100	15	284	Donna Murphy	799	2.01E+08	Adventure|B	rad Garrett	TangledÂ	
Color	Joss Whedon	635	141	0	19000	Robert Downey Jr	26000	4.59E+08	Action|Adv	Chris Hemsworth	Avengers: Age	
Color	David Yates	375	153	282	10000	Daniel Radcliffe	25000	3.02E+08	Adventure|A	lan Rickman	Harry Potter ar	
Color	Zack Snyder	673	183	0	2000	Lauren Cohan	15000	3.3E+08	Action|Adv	Henry Cavill	Batman v Supe	
Color	Bryan Singer	434	169	0								
Color	Marc Forster	403	106	395								
Color	Gore Verbinski	313	151	563								
Color	Gore Verbinski	450	150	563								
Color	Zack Snyder	733	143	0								
Color	Andrew Adamson	258	150	80								
Color	Joss Whedon	703	173	0								
Color	Rob Marshall	448	136	252								
Color	Barry Sonnenfeld	451	106									
Color	Peter Jackson	422	164	0								
Color	Marc Webb	599	153	464								
Color	Ridley Scott	343	156	0								
Color	Peter Jackson	509	186	0								
Color	Chris Weitz	251	113	129								
Color	Peter Jackson	446	201	0								
Color	James Cameron	315	194	0								
Color	Anthony Russo	516	147	94								
Color	Peter Berg	377	131	533								

IMDB_Movies_Raw data

	A	B	C	D	E	F	G	H
1	director_name	duration	gross	genres	movie_title	language	budget	imdb_score
2	James Cameron	178	760505847	Action|Adventure|Fantasy|Sci-Fi	AvatarÂ	English	237000000	7.9
3	Gore Verbinski	169	309404152	Action|Adventure|Fantasy	Pirates of the Caribbean: At World's EndÂ	English	300000000	7.1
4	Sam Mendes	148	200074175	Action|Adventure|Thriller	SpectreÂ	English	245000000	6.8
5	Christopher Nolan	164	448130642	Action|Thriller	The Dark Knight RisesÂ	English	250000000	8.5
7	Andrew Stanton	132	73058679	Action|Adventure|Sci-Fi	John CarterÂ	English	263700000	6.6
8	Sam Raimi	156	336530303	Action|Adventure|Romance	Spider-Man 3Â	English	258000000	6.2
9	Nathan Greno	100	200807262	Adventure|Animation|Comedy|Family|Fantasy|Musical|Romance	TangledÂ	English	260000000	7.8
10		141	458991599	Action|Adventure|Sci-Fi	Avengers: Age of UltronÂ	English	250000000	7.5
11	David Yates	153	301956980	Adventure|Family|Fantasy|Mystery	Harry Potter and the Half-Blood PrinceÂ	English	250000000	7.5
12	Zack Snyder	183	330249062	Action|Adventure|Sci-Fi	Batman v Superman: Dawn of JusticeÂ	English	250000000	6.9
13	Bryan Singer	169	200069408	Action|Adventure|Sci-Fi	Superman ReturnsÂ	English	209000000	6.1
14	Marc Forster	106	168368427	Action|Adventure	Quantum of SolaceÂ	English	200000000	6.7
15	Gore Verbinski	151	423032628	Action|Adventure|Fantasy	Pirates of the Caribbean: Dead Man's ChestÂ	English	225000000	7.3
16	Gore Verbinski	150	89289910	Action|Adventure|Western	The Lone RangerÂ	English	215000000	6.5
17	Zack Snyder	143	291021565	Action|Adventure|Fantasy|Sci-Fi	Man of SteelÂ	English	225000000	7.2
18	Andrew Adamson	150	141614023	Action|Adventure|Family|Fantasy	The Chronicles of Narnia: Prince CaspianÂ	English	225000000	6.6
19	Joss Whedon	173	623279547	Action|Adventure|Sci-Fi	The AvengersÂ	English	220000000	8.1
20	Rob Marshall	136	241063875	Action|Adventure|Fantasy	Pirates of the Caribbean: On Stranger TidesÂ	English	250000000	6.7
21	Barry Sonnenfeld	106	179020854	Action|Adventure|Comedy|Family|Fantasy|Sci-Fi	Men in Black 3Â	English	225000000	6.8
22	Peter Jackson	164	255108370	Adventure|Fantasy	The Hobbit: The Battle of the Five ArmiesÂ	English	250000000	7.5
23	Marc Webb	153	262030663	Action|Adventure|Fantasy	The Amazing Spider-ManÂ	English	230000000	7
24	Ridley Scott	156	105219735	Action|Adventure|Drama|History	Robin HoodÂ	English	200000000	6.7
25	Peter Jackson	186	258355354	Adventure|Fantasy	The Hobbit: The Desolation of SmaugÂ	English	225000000	7.9
26	Chris Weitz	113	70083519	Adventure|Family|Fantasy	The Golden CompassÂ	English	180000000	6.1
27	Peter Jackson	201	218051260	Action|Adventure|Drama|Romance	King KongÂ	English	207000000	7.2
28	James Cameron	194	658672302	Drama|Romance	TitanicÂ	English	200000000	7.7
29	Anthony Russo	147	407197282	Action|Adventure|Sci-Fi	Captain America: Civil WarÂ	English	250000000	8.2
30	Peter Berg	131	65173160	Action|Adventure|Sci-Fi|Thriller	BattleshipÂ	English	209000000	5.9
31	Colin Trevorrow	124	652177271	Action|Adventure|Sci-Fi|Thriller	Jurassic WorldÂ	English	150000000	7

IMD8_Movies_Raw data - Copy

Ready 3887 of 5043 records found Accessibility: Unavailable

FIVE WHY'S APPROACH

Problem: "Movies with higher budgets tend to have higher ratings."

WHY's

- **Why do movies with higher budgets tend to have higher ratings?**
  A: Higher budgets allow for better production quality, including visual effects, set design, and sound.
- **Why does better production quality lead to higher ratings?**
  A: High-quality production creates a more immersive and enjoyable experience for the audience.
- **Why does an immersive and enjoyable experience lead to higher ratings?**
  A: Viewers are more inclined to rate the movie positively when entertained or emotionally moved.
- **Why are viewers more inclined to rate a movie positively when they are entertained or emotionally moved?**
  A: Positive experiences resonate with viewers, influencing their opinions and motivating them to leave favorable reviews.
- **Why do positive reviews matter?**
  A: Positive reviews build a movie's reputation, encouraging more people to watch it, which often sustains or boosts its IMDb rating.

FIVE WHY'S APPROACH

Problem: "Movies with well-known actors tend to have higher IMDb ratings."

WHY's

- **Why do movies with well-known actors tend to have higher ratings?**
  A: Well-known actors often deliver strong performances that resonate with audiences.
- **Why do strong performances resonate with audiences?**
  A: Compelling acting helps audiences connect emotionally with the story and characters.
- **Why does emotional connection lead to higher ratings?**
  A: When viewers feel engaged and emotionally invested, they are more likely to appreciate the film and give it a positive rating.
- **Why does audience appreciation result in higher ratings?**
  A: Satisfied viewers are more motivated to share their positive experiences through reviews and high ratings.
- **Why do positive reviews from satisfied viewers matter?**
  A: Positive reviews attract more viewers, creating a cycle of increased visibility and a strong reputation, which keeps the IMDb rating high.

Objective

IMDb is a widely recognized platform for movie and series ratings, providing insights from users and critics worldwide. It offers comprehensive details about movies and series, including ratings, directors, actors, and financial data such as budgets and collections. In this project, we have access to a dataset of movies spanning from 1920 to 2010. This dataset includes information on movies, their casts, directors, budgets, box office collections, and more.

We aim to clean the dataset and use analytical methods, specifically the Five Whys technique, to explore and answer key questions. The analysis will be conducted using Microsoft Office 365 Excel for data cleaning, processing, and insights generation.

Key Takeaway of 5 Why's:

Movies with higher budgets leverage superior resources to deliver exceptional viewer experiences, leading to positive feedback and widespread popularity, which ultimately reflect in higher IMDb ratings. Well-known actors enhance the quality of a movie by delivering memorable performances and fostering emotional connections with the audience. This connection leads to positive experiences and favorable ratings, amplifying the movie's success.

Teck Stack

Microsoft Excel 365

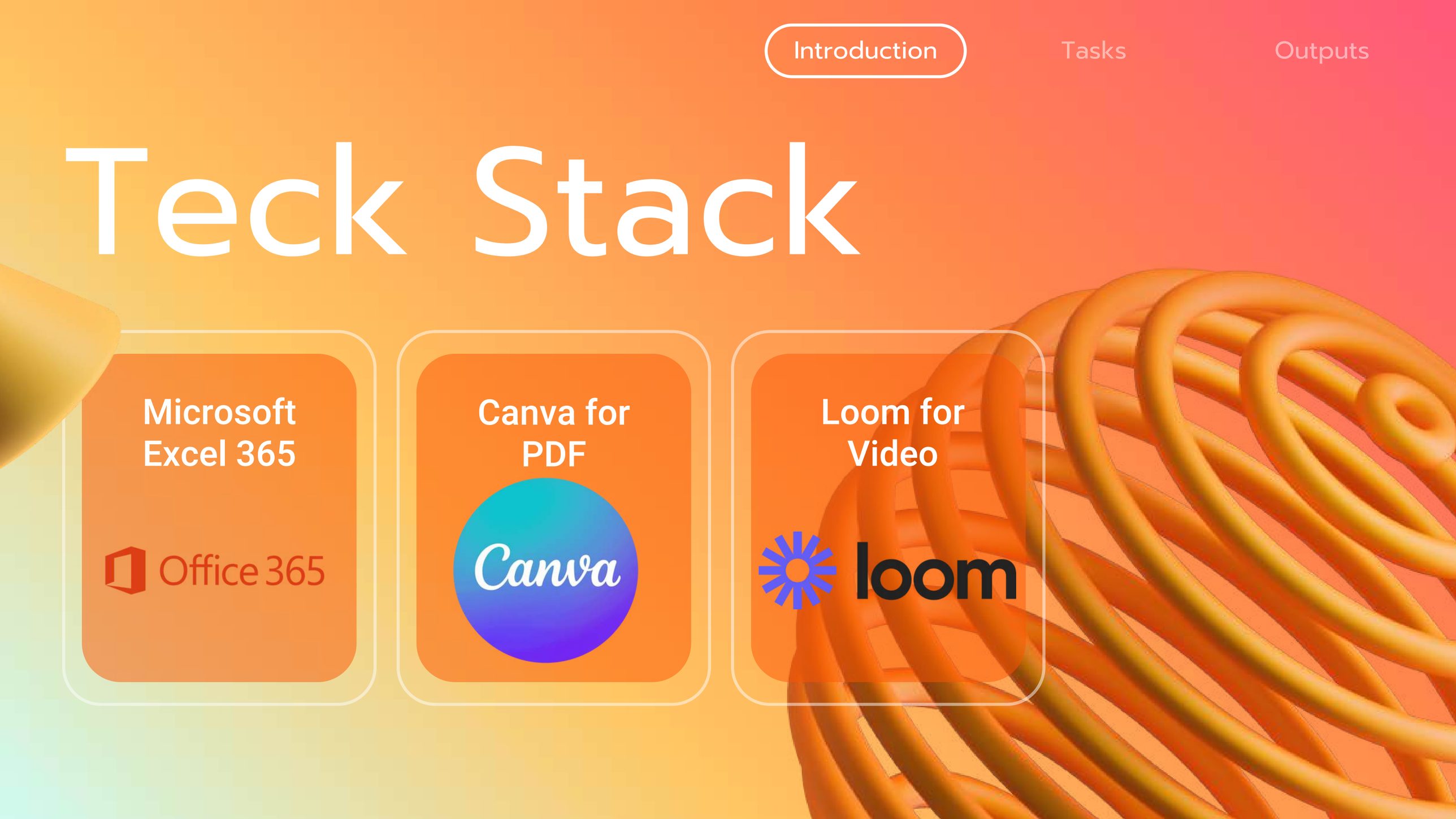
Office 365

Canva for PDF

Canva

Loom for Video

loom



Deliverables

Data Analysis Tasks Description

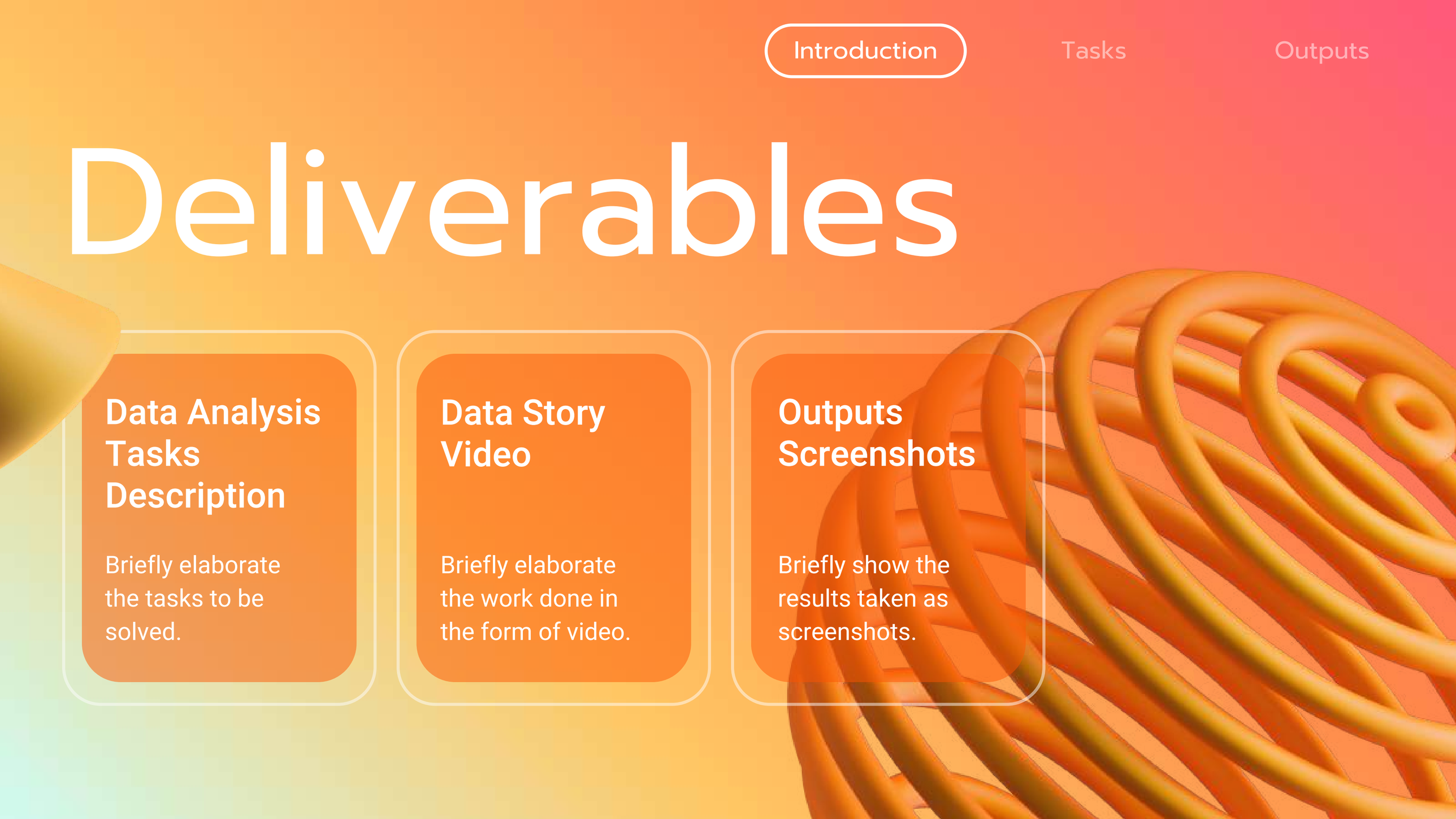
Briefly elaborate the tasks to be solved.

Data Story Video

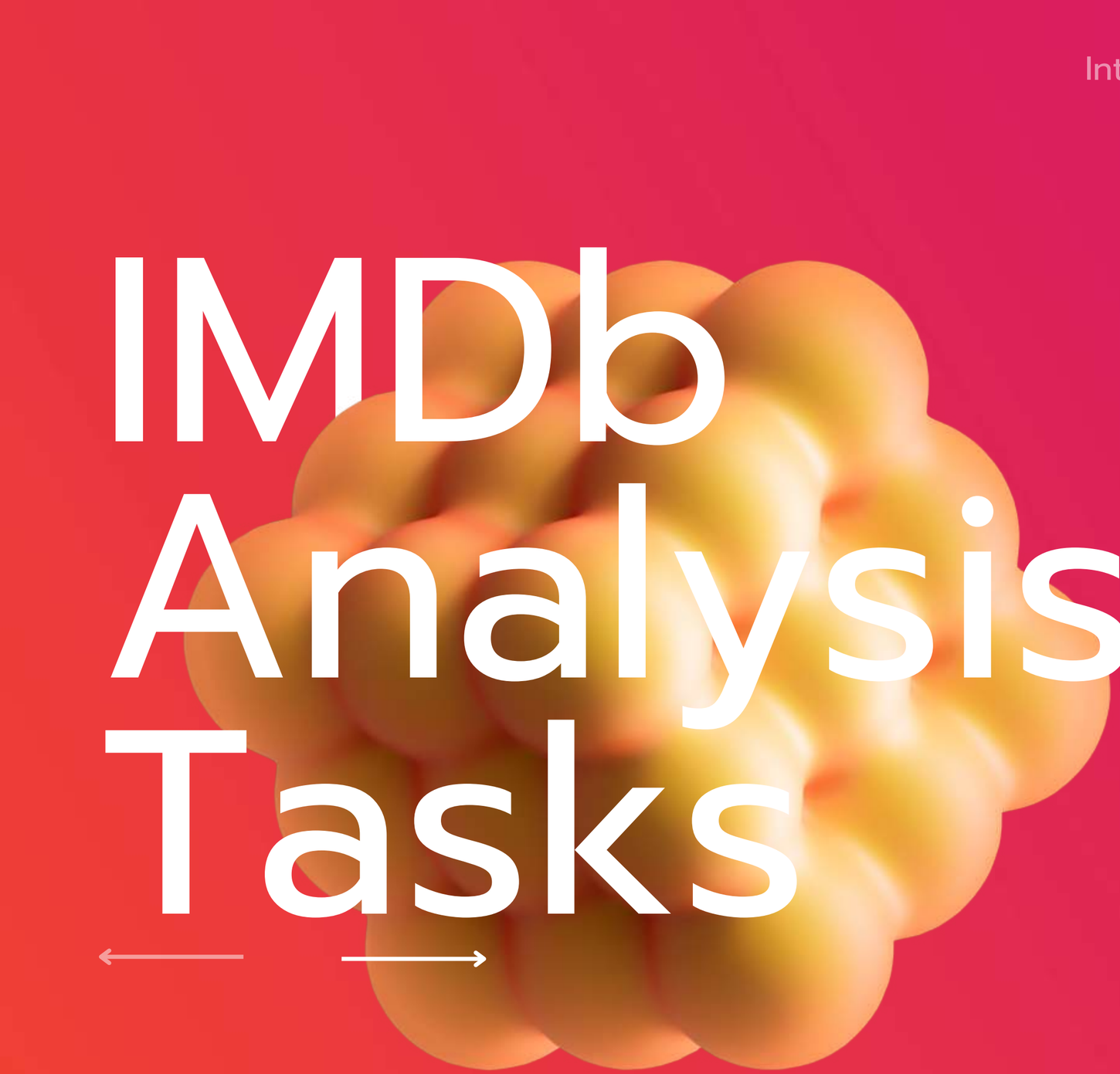
Briefly elaborate the work done in the form of video.

Outputs Screenshots

Briefly show the results taken as screenshots.



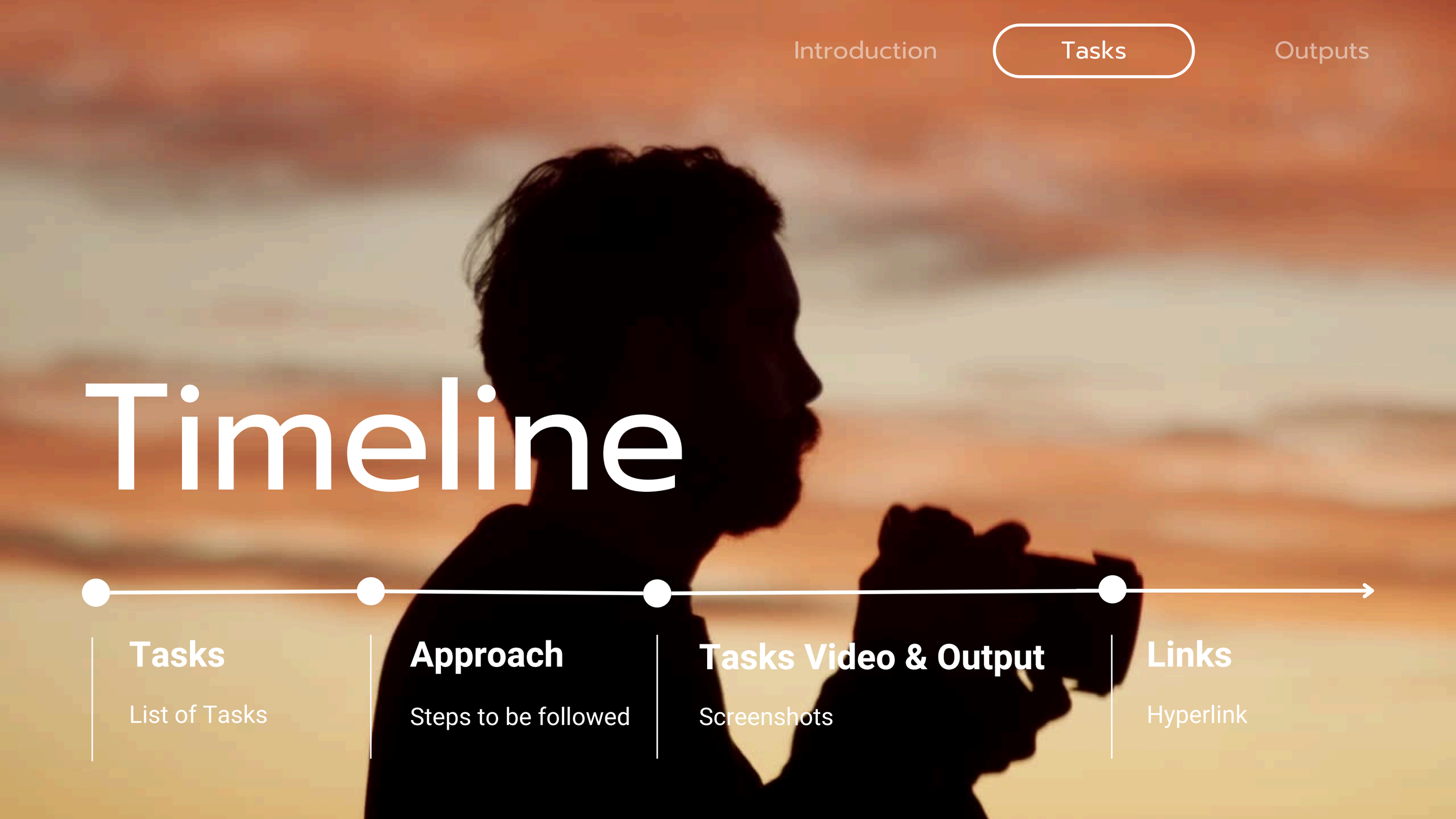
IMDb Analysis Tasks



02

01

Elaborate on what you want to discuss.



Timeline

Tasks

List of Tasks

Approach

Steps to be followed

Tasks Video & Output

Screenshots

Links

Hyperlink

Movie Genre Analysis | Examine the prevalence of different movie genres and how they influence IMDb ratings.

Movie Duration Analysis | Study how the variation in movie durations affects IMDb ratings.

Language Analysis | Analyze how movies are distributed across different languages.

Director Analysis | Analyzing the impact of directors on movie ratings.

Budget Analysis: | Investigate how movie budgets correlate with their box office performance.

Approach

Data Collection

The process of collecting the data is used to be processed. In this project, the given data is collected.

Data Cleaning

The process of cleaning the data by deleting redundant & unnecessary data is data cleaning process.

Visualisation

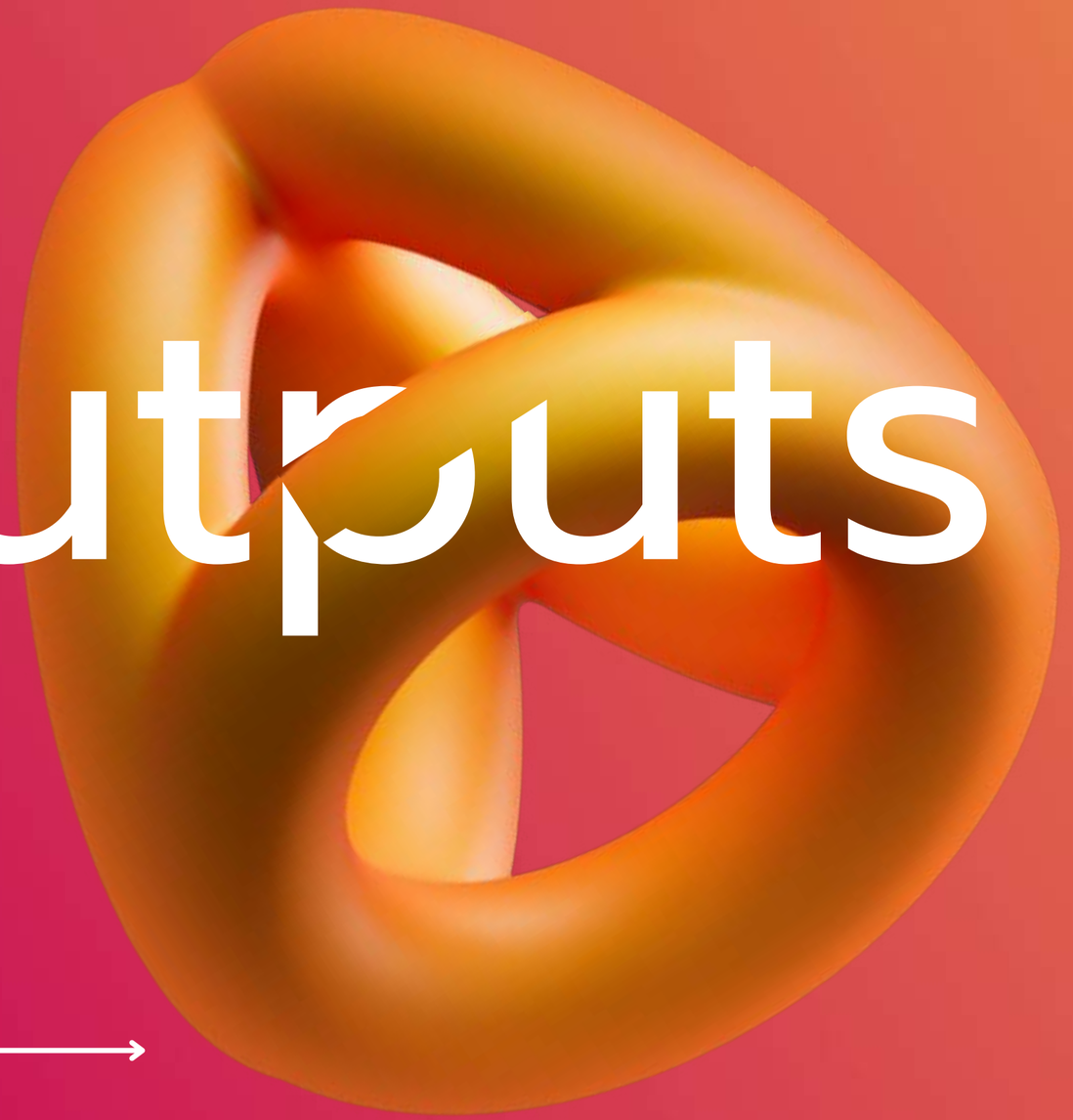
The cleaned data is represented in the forms of charts, bar graph or tables.

PPT/PDF & Video Prep.

Using canva and loom, the particulars are created and represented.


IMDb movies analysis shows that behind every great film is a story shaped by trends, talent, and the audience's heart.

Outputs



03

02

Elaborate on what you want to discuss.

TASK 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.

`=MEDIAN(K1:K17)`

`=MIN(K1:K17)`

`=MAX(K1:K17)`

`=VAR(K1:K17)`

`=AVERAGE(K1:K17)`

`=MODE(K1:K17)`

`=STDEV(K1:K17)`

Using COUNTIF FUNCTION - 5 TOP GENRES

`=COUNTIF(A2:A5040,"action")`

`=COUNTIF(A2:A51040,"drama")`

`=COUNTIF(A2:A51040,"romance")`

`=COUNTIF(A2:A51040,"thriller")`

`=COUNTIF(A2:A5040,"Comedy")`

Column1
Action
Action
Action
Action
Documentary
Action
Action
Adventure
Action
Adventure
Action
Action
Action
Action
Action
Action
Action
Action
Adventure
Action
Action
Adventure
Adventure
Action
Drama
Action
Action
Action
Action
Action
Action
Adventure
Action
Adventure
Action

Row Labels	Count of Column1
Animation	242
Biography	293
Comedy	1872
Crime	889
Documentary	121
Drama	2594
Family	546
Fantasy	610
Film-Noir	6
Game-Show	1
History	207
Horror	565
Music	214
Musical	132
Mystery	500
News	3
Reality-TV	2
Romance	1106
Sci-Fi	612
Short	5
Action	1153
Sport	182
Thriller	1411
War	212
Western	97
Grand Total	**14498**

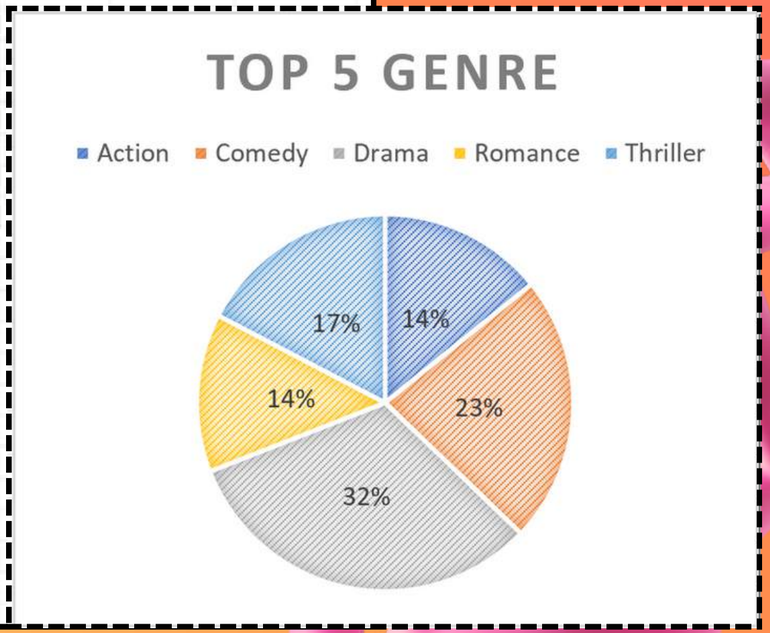
Column1	Column2
Median	267.5
Average	557.61538
Mode	#N/A
Max	2594
Min	1
Var	412055.53
StdDev	641.91551

M	N
Function	Value
Median	43
Mode	3
Max	1026
Min	2
Var	113218.5
Stdev	336.4795
Average	222.7059

Using Countif fn for each genre

Action	1153	
Adventure	453	
Animation	61	
Biography	252	
Comedy	1327	1106
Crime	348	612
Documentary	0	5
Drama	2594	182
Family	546	1411
Fantasy	610	212
Film-Noir	0	97
Game-Show	1	1153
History	207	
Horror	565	
Music	214	
Musical	132	
Mystery	500	
News	3	
Reality-TV	2	
Romance	1106	
Sci-Fi	612	
Short	5	
Sport	182	
Thriller	1411	
War	212	

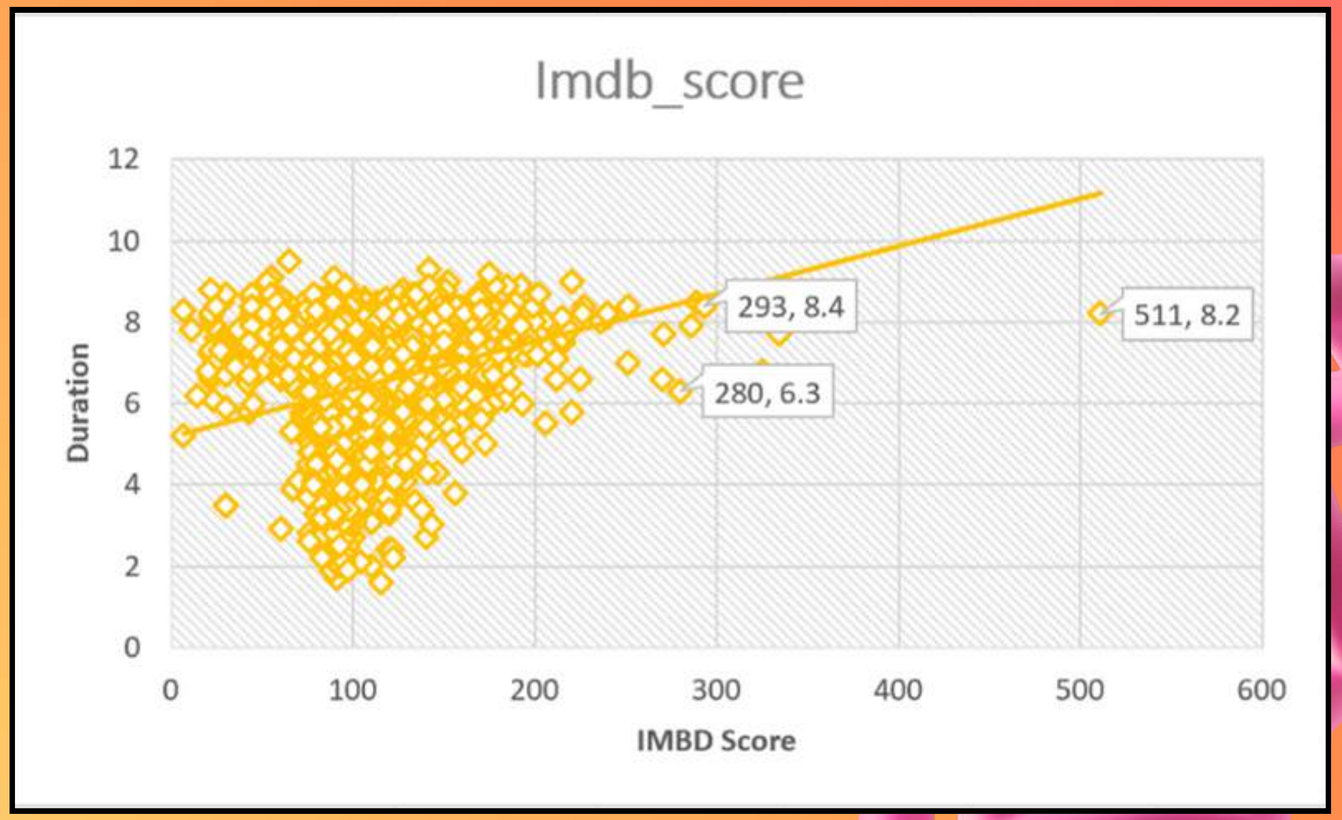
923
293
1872
889
121
2594
546
610
6
1
207
565
214
132
500
3
2
1106
612
5
182
1411
212
97
1153



TASK B - Movie Duration Analysis

Analyze the distribution of movie durations and its impact on the IMDB score.

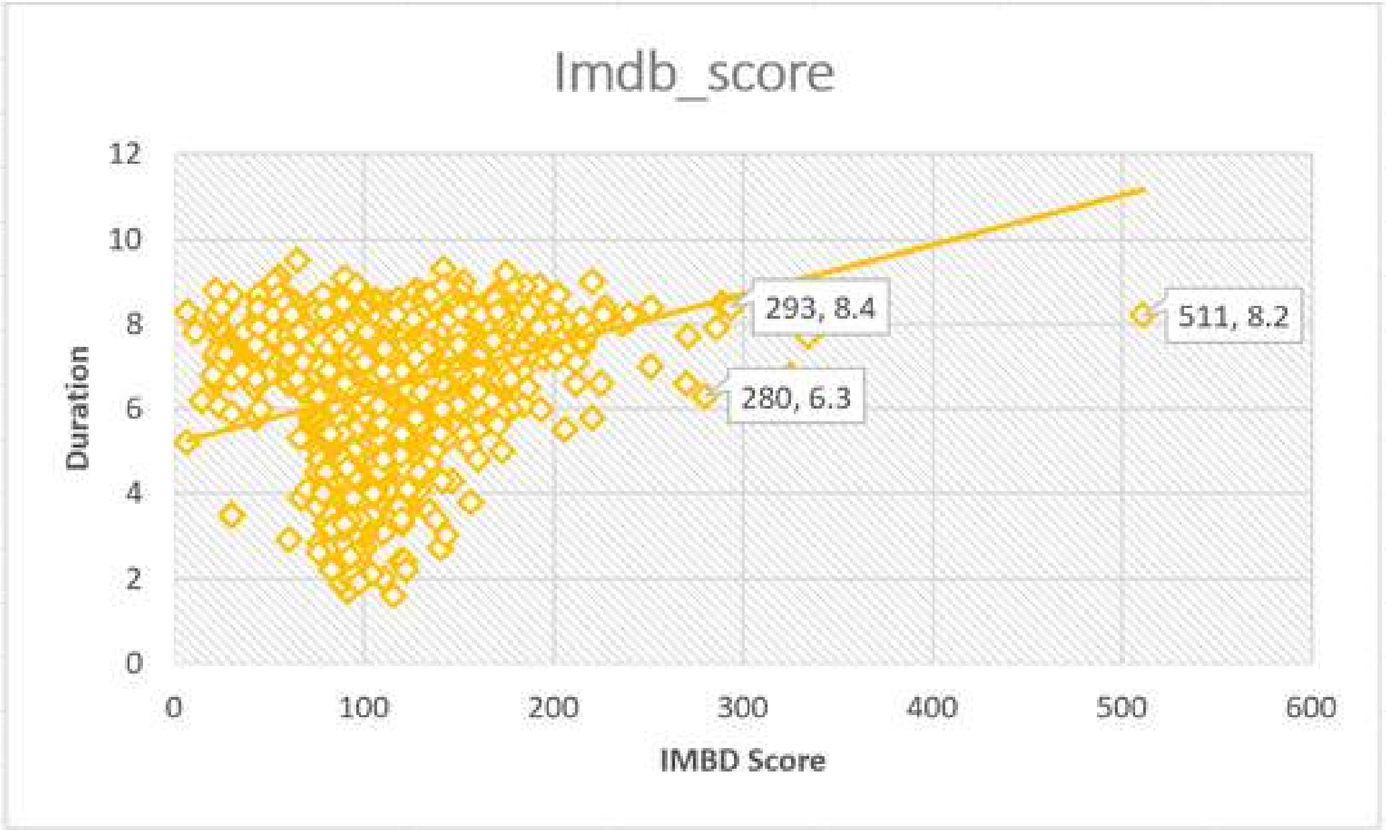
Task: Analyze the distribution of movie durations and identify the relationship between movie duration and IMDB score.



As we can see the plotting touched the gridlines, therefore, we can say that the maximum number of Scattered IMDb score is from 6 to 9.5 duration of time.

duration	imdb_score
178	7.9
169	7.1
148	6.8
164	8.5
132	6.6
156	6.2
100	7.8
141	7.5
153	7.5
183	6.9
169	6.1
106	6.7
151	7.3
150	6.5
143	7.2
150	6.6
173	8.1
136	6.7
106	6.8
164	7.5
153	7
156	6.7

Function (Duration)	Value
Average	107.20107
Median	103
Standard Deviation	25.197441
IMBD SCORE AVG	6.4421376



D	E
Function (Duration)	Value
Average	107.20107
Median	103
Standard Deviation	25.197441
IMBD SCORE AVG	6.4421376

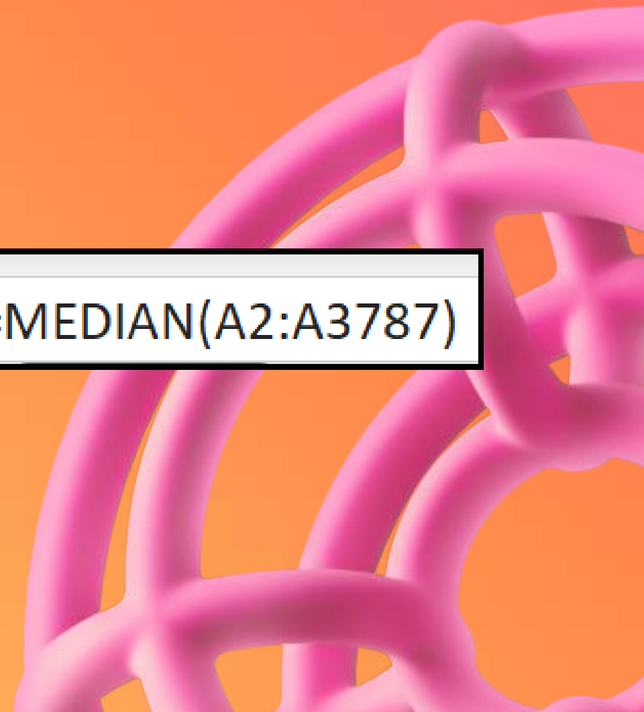
FORMULAE USED

fx =AVERAGE(A2:A3787)

fx =STDEV(A2:A3787)

fx =MEDIAN(A2:A3787)

fx =AVERAGE(Table2[imdb_score])



TASK C - Language Analysis

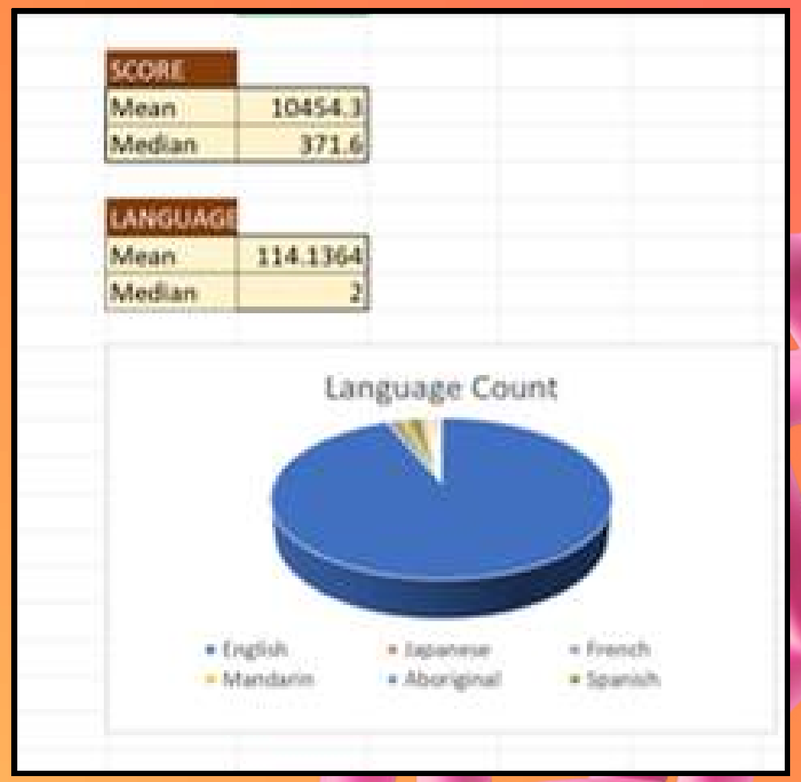
Situation: Examine the distribution of movies based on their language.

Task: Determine the most common languages used in movies and analyze their impact on the IMDB score using descriptive statistics.

```
=SUM(K:K)/COUNT(K:K)
```

Most Common Language
English, French, Hindi, Spanish & Mandarin

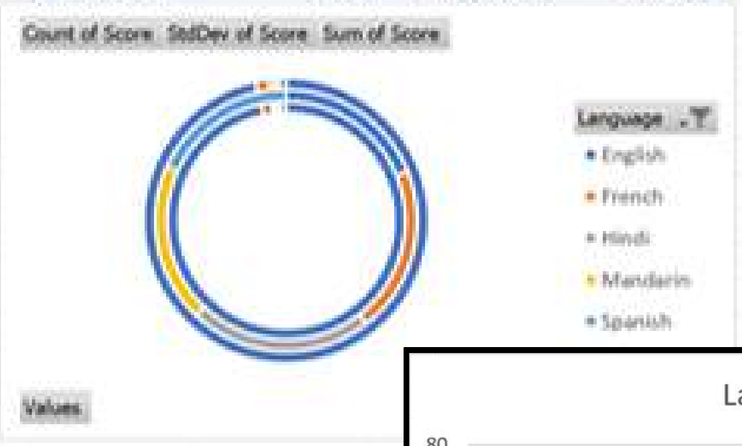
```
=MEDIAN(K3:K5044)
```

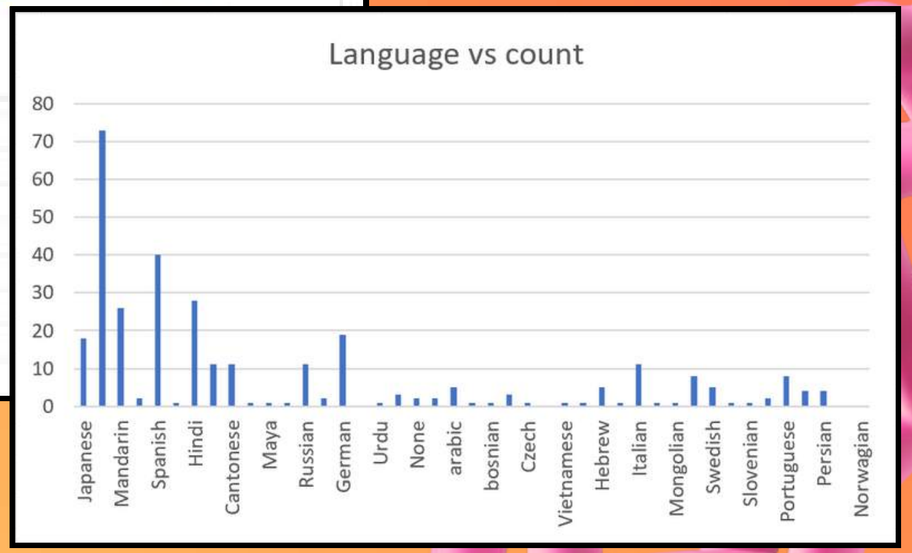


Language	LANG	Count of LANG
English	English	4704
English	Japanese	18
English	French	73
English	Mandarin	26
English	Aboriginal	2
English	Spanish	40
English	Filipino	1
English	Hindi	28
English	Russian	11
English	Cantonese	11
English	Kazakh	1
English	Maya	1
English	Telugu	1
English	Russian	11
English	Icelandic	2
English	German	19
English	Zuku	0
English	Urdu	1
English	Thai	3
English	None	2
English	Aboriginal	2
English	arabic	5
English	aramaic	1
English	bosnian	1
English	Chinese	3
English	Czech	1
English	Dzongkha	0
English	Vietnamese	1

Language	Score
English	7.9
English	7.1
English	6.8
English	8.5
English	7.1
English	6.6
English	6.2
English	7.8
English	7.5
English	7.5
English	6.9
English	6.1
English	6.7
English	7.3
English	6.5
English	7.2
English	6.6
English	8.1
English	6.7
English	6.8
English	7.5
English	7
English	6.7
English	7.9
English	6.1
English	7.2
English	7.7
English	8.2

Row Labels	Count of Score	StdDev of Score	Sum of Score
English	4704	1.125716901	30273.7
French	73	1.230259653	477.4
Hindi	28	1.347346991	185.2
Mandarin	26	1.179726174	160.8
Spanish	40	1.085321647	265.8
Grand Total	**4871**	**1.12864289**	**31362.9**





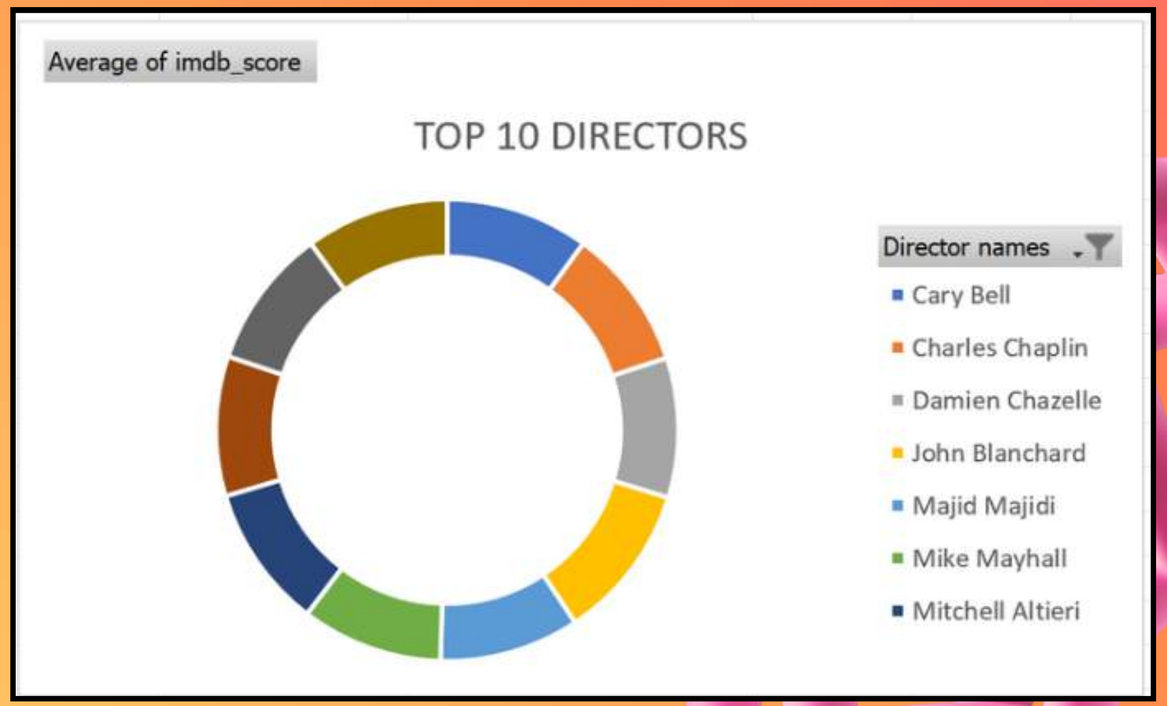
TASK D - Director Analysis

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.

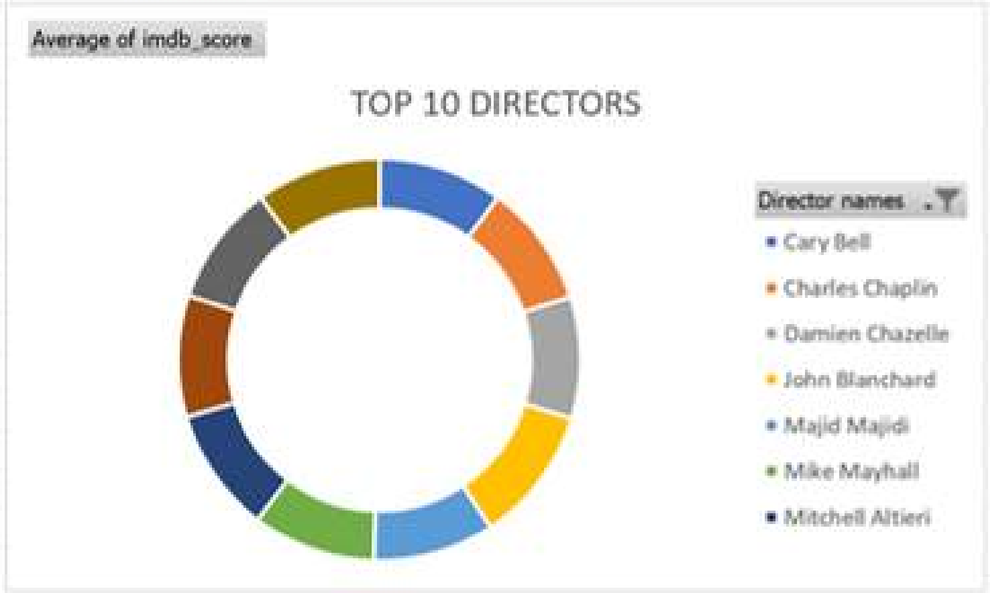
```
=PERCENTILE(B:B,0.9)
```

```
=COUNTIF(B:B,">=7.5")
```



Director names	imdb_score	90th Percentile	TOP DIRECTORS
John Blanchard	9.5	7.8	887
Frank Darabont	9.3		
Francis Ford Coppola	9.2		
John Stockwell	9.1		
Christopher Nolan	9		
Francis Ford Coppola	9		
Peter Jackson	8.9		
Steven Spielberg	8.9		
Quentin Tarantino	8.9		
Sergio Leone	8.9		
Sidney Lumet	8.9		
Christopher Nolan	8.8		
Peter Jackson	8.8		
David Fincher	8.8		
Robert Zemeckis	8.8		
Irvin Kershner	8.8		
Peter Jackson	8.7		
Lana Wachowski	8.7		
Martin Scorsese	8.7		
George Lucas	8.7		
Milos Forman	8.7		
Fernando Meirelles	8.7		
Mitchell Altieri	8.7		
Sadyk Sher-Niyaz	8.7		
Akira Kurosawa	8.7		
Cary Bell	8.7		
Christopher Nolan	8.6		
Steven Spielberg	8.6		

Row Labels	Average of imdb_score
Cary Bell	8.7
Charles Chaplin	8.6
Damien Chazelle	8.5
John Blanchard	9.5
Majid Majidi	8.5
Mike Mayhall	8.6
Mitchell Altieri	8.7
Raja Menon	8.5
Ron Fricke	8.5
Sadyk Sher-Niyaz	8.7
Grand Total	**8.68**



TASK 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.

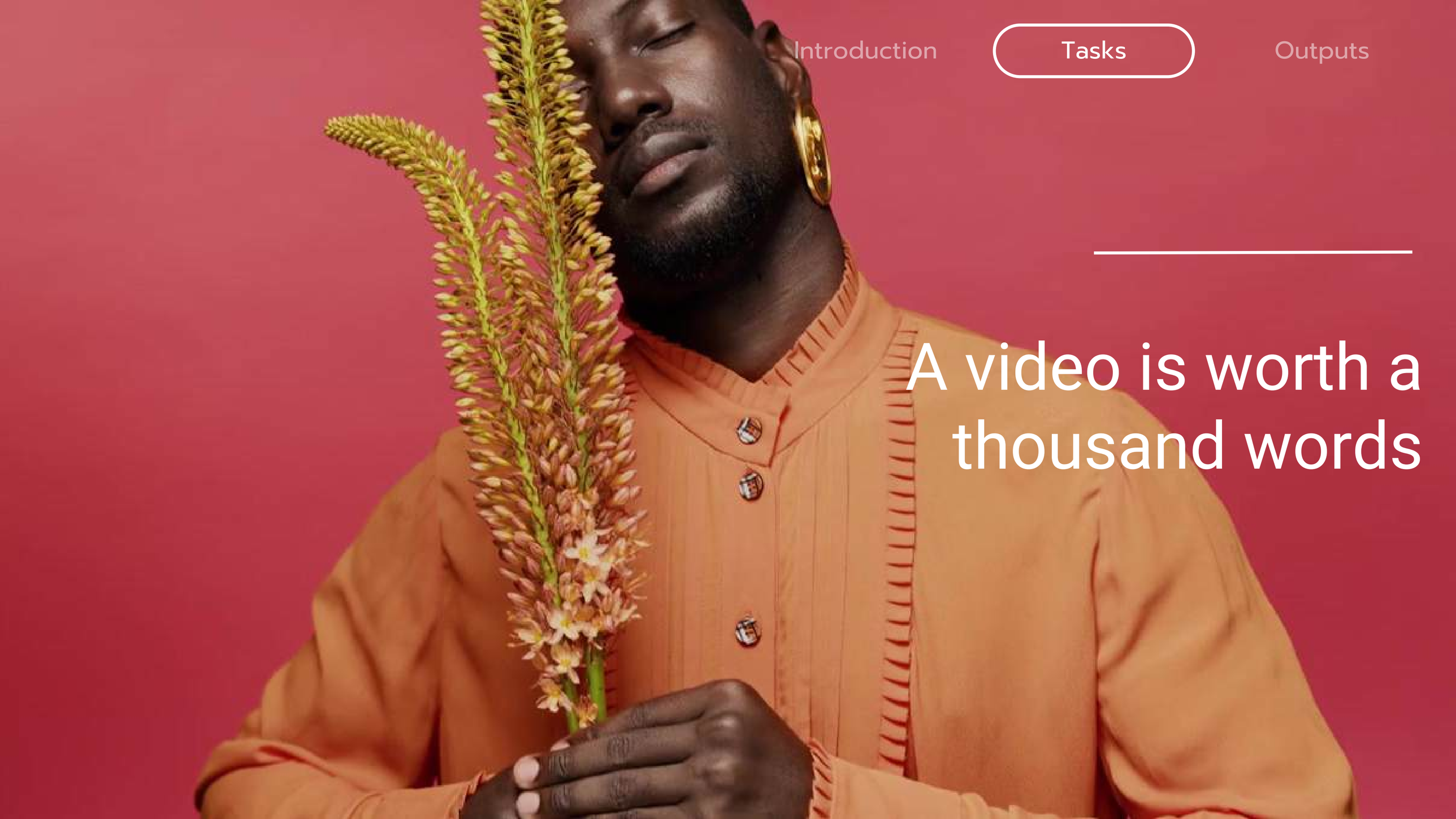
```
=MAX(D:D)
```

```
=[@gross]-[@budget]
```

```
=CORREL([gross],[budget])
```

Column1	gross	budget	Profit	MAXIMUM PROFIT	CORREL
BattleshipÂ	760505847	237000000	523505847	523505847	0.102179454
King KongÂ	652177271	150000000	502177271		
GremlinsÂ	658672302	200000000	458672302		
2001: A Space OdysseyÂ	460935665	11000000	449935665		
The Chronicles of Narnia: Prince CaspianÂ	434949459	10500000	424449459		
Event HorizonÂ	623279547	220000000	403279547		
A Beautiful MindÂ	623279547	220000000	403279547		
The WolverineÂ	422783777	45000000	377783777		
X-Men: ApocalypseÂ	474544677	115000000	359544677		
Mission: ImpossibleÂ	533316061	185000000	348316061		
The Great RaidÂ	407999255	78000000	329999255		
Ant-ManÂ	363024263	58000000	305024263		
Astro BoyÂ	424645577	130000000	294645577		
The League of Extraordinary GentlemenÂ	356784000	63000000	293784000		
Holy ManÂ	368049635	76000000	292049635		
The Lovely BonesÂ	350123553	58800000	291323553		
SignsÂ	380838870	94000000	286838870		
Finding NemoÂ	436471036	150000000	286471036		
Pushing TinÂ	377019252	94000000	283019252		
Walking TallÂ	309125409	32500000	276625409		
The WaterboyÂ	329691196	55000000	274691196		
Beverly Hills CopÂ	290158751	18000000	272158751		
OblivionÂ	285761243	18000000	267761243		
Star TrekÂ	380262555	113000000	267262555		
UnbreakableÂ	403706375	139000000	264706375		
The MedallionÂ	336029560	74000000	262029560		
The ExorcistÂ	293501675	40000000	253501675		
The Matrix RevolutionsÂ	260000000	8000000	252000000		

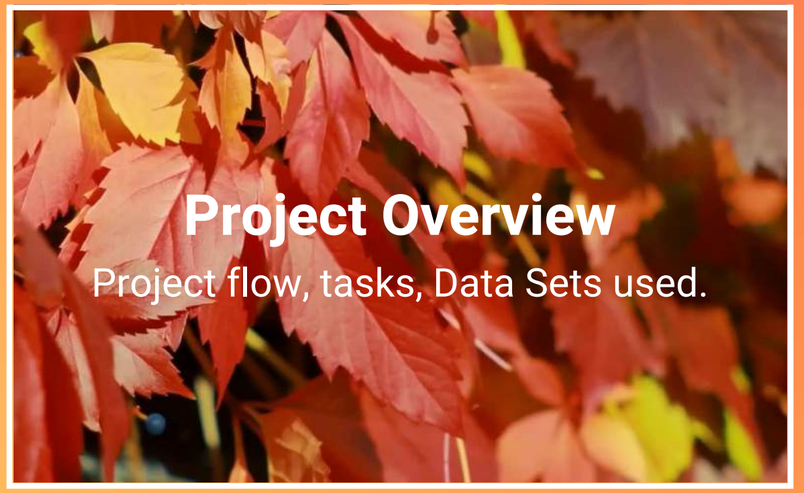
movies	Profit
BattleshipÂ	523505847
King KongÂ	502177271
GremlinsÂ	458672302
2001: A Space OdysseyÂ	449935665
The Chronicles of Narnia: Prince CaspianÂ	424449459
Event HorizonÂ	403279547
A Beautiful MindÂ	403279547
The WolverineÂ	377783777
X-Men: ApocalypseÂ	359544677
Mission: ImpossibleÂ	348316061



A video is worth a thousand words

Introduction

Objective with Tools used



Project Overview

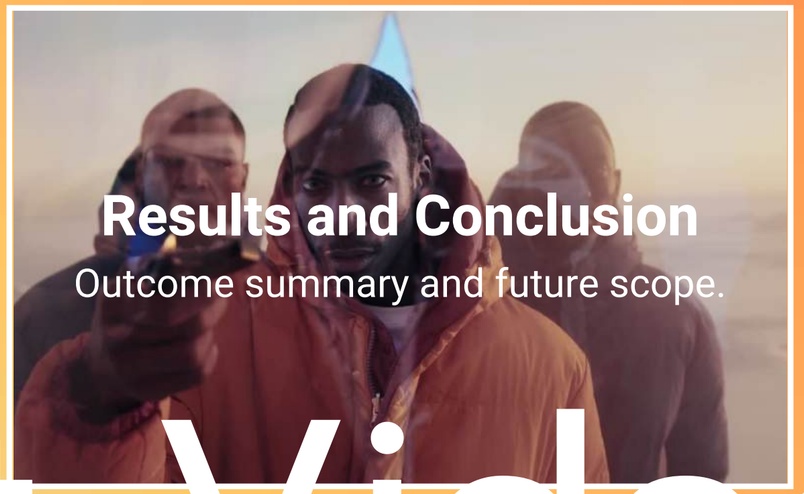
Project flow, tasks, Data Sets used.

Data Analysis and Insights

Project Outputs Description

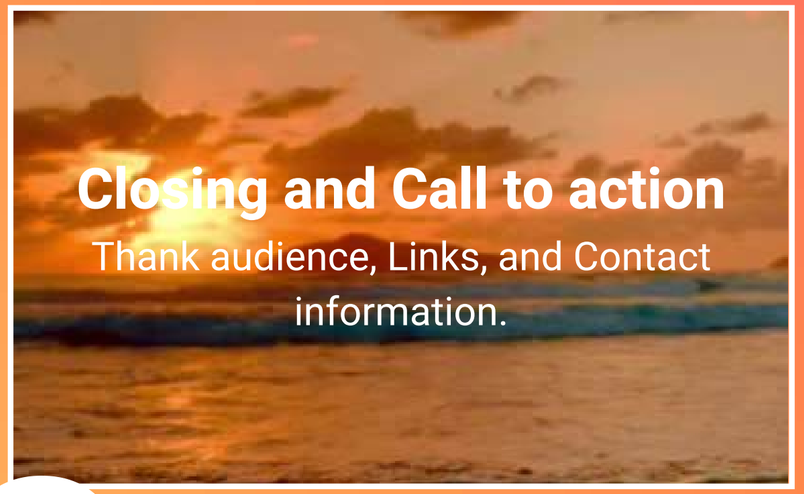
Code Demonstration

Code Snippet Description.



Results and Conclusion

Outcome summary and future scope.

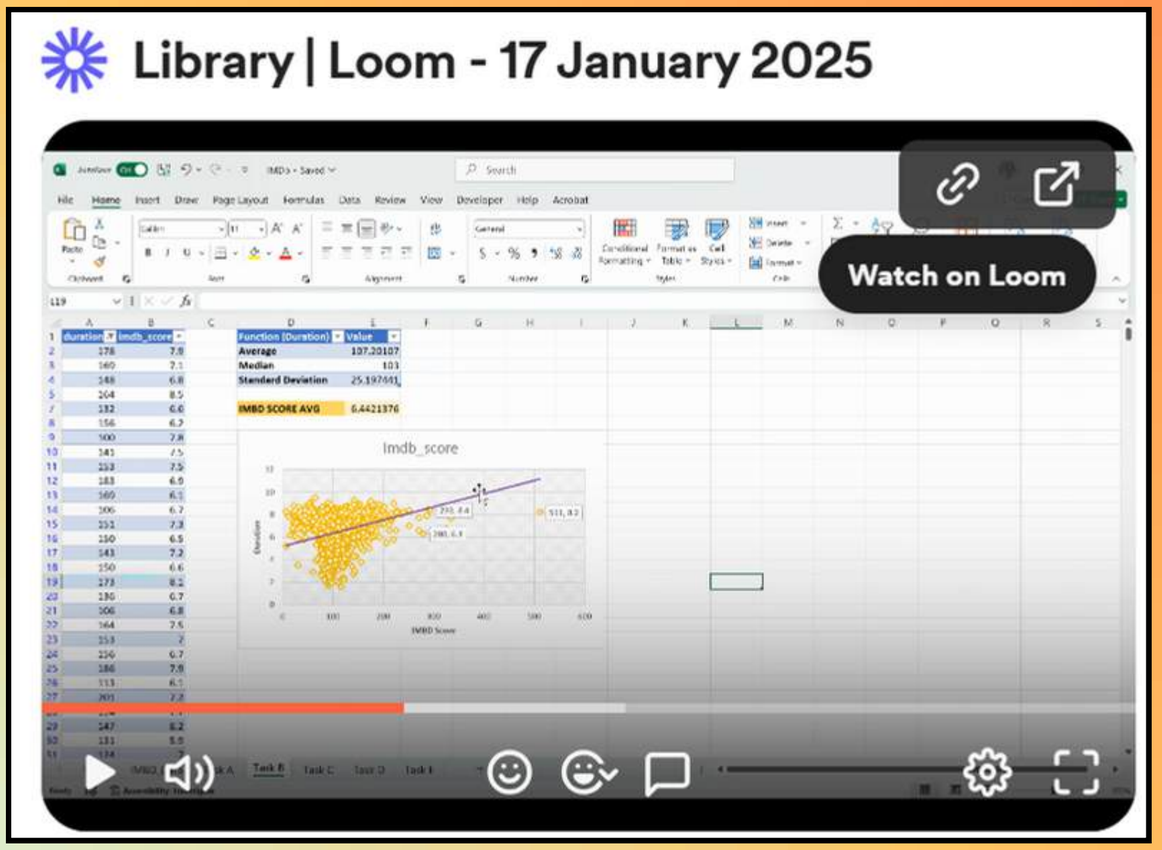


Closing and Call to action

Thank audience, Links, and Contact information.

Project Video

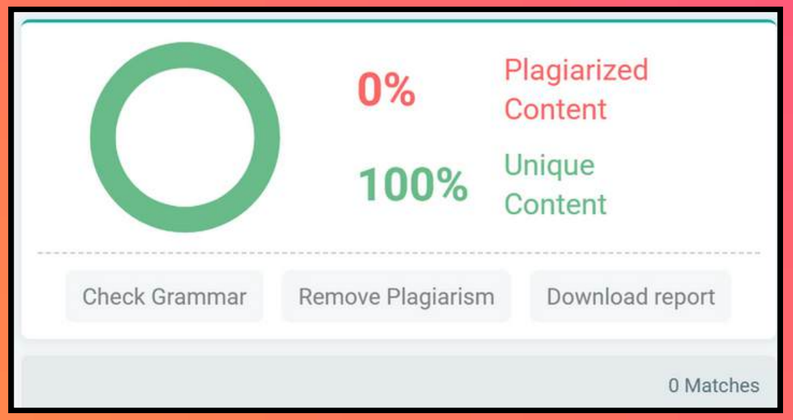
Project



Video Link

Video

DRIVE


0%
Plagiarized Content
100%
Unique Content
Check Grammar
Remove Plagiarism
Download report
0 Matches


PDF


Video


EXCEL
XLS

LINK

The Team

Ankita Taneja

Presentation

Ankita Taneja

Cinematographer

Ankita Taneja

Producer

Ankita Taneja

Editor

Contact Us

Gurugram
(Haryana), India.

70-11-33-40-48

Venusgirlatwork
@gmail.com

@ig_she_has_
no_idea

Linkedin

Thank you!

An IMDb movies analysis reveals that the magic of cinema lies at the crossroads of data, creativity, and audience connection.