



L OVELY
P ROFESSIONAL
U NIVERSITY

Transforming Education Transforming India

Introduction to Data Management PROJECT REPORT

(Project Semester August-December 2021)

PROJECT REPORT

ON

Analysis of IMDB data from 2006 to 2016

Submitted by

Baggam Sai Karthik

11906091

Programme: Bachelor of

TechnologySection: KM039

Course Code: INT217

Under the Guidance of

Tanima Thakur:23532

Assistant Professor

Discipline of CSE/IT

Lovely School of Computer Science & Engineering

Lovely Professional University, Phagwara

DECLARATION

I, Baggam Sai Karthik, student of Computer Science & Engineering under CSE/IT Discipline at, Lovely Professional University, Punjab, hereby declare that all the information furnished in this project report is based on my own intensive work and is genuine.

Date: 8/12/2021

Baggam Sai Karthik

Registration No: 11906091

Signature

B. Sai Karthik

ACKNOWLEDGEMENT

A project work is a combination of views, ideas, suggestions and contribution of many people. Thus, one of the pleasant parts of writing the report is to thank those who have contributed towards its fulfilment.

I consider it as great privilege to have esteemed Lecturer **Tanima Thakur** as my project guide. I take this opportunity to express my sincere gratitude to her through constant advice and constructive criticism nourished my interest in the subject and provided a free and pleasant atmosphere to work against all odd situations. I avail this opportunity to extend my heart full thanks and deep respect to faculty member for their able guidance during this project.

My gratitude to all those, who responded to my questionnaire in a well-defined manner and helped me acquiring knowledge.

I would like to communicate a deep sense of gratitude to all these people without whom my project would not have been such a great learning experience.

Baggam Sai Karthik

11906091

KM039

Lovely Professional University

LIST OF CONTENTS

Sr No.	Title	Page No.
1	Cover page	1
2	Declaration	2
3	Acknowledgement	3
4	List of Contents	4
5	Introduction	5
6	Objectives/Scope of the Analysis	6
7	Source of Dataset	6
8	Content of data	7
9	ETL Process	8
10	Analysis of Dataset	13
11	List of Analysis with results	19
12	Final Dashboard	28
13	Bibliography	29

INTRODUCTION

- Data management is The Very important because the data your organization creates is a very valuable resource.
- The last thing you want to do is spend time and resources collecting data and business intelligence, only to lose or misplace that information.
- In that case, you would then have to spend time and resources again to get that same business intelligence you already had.
- And on that data analysis is carried out which show visualization of our problems in efficient way.
- Data Analysis is a process of inspecting, cleansing, transforming, and modeling data with the goal of discovering useful information, informing conclusions, and supporting decision- making.
- This project is based on data analysis on IMDb data from 2006 to 2016.
- The Full form of IMDb is internet movie database .It is an online database of information related to films, television programs, home videos, video games, and streaming content online including cast, production crew and personal biographies, plot summaries, trivia, ratings, and fan and critical reviews.
- This IMDb is the world's most popular and authoritative source for movie,TV and celebrity content.
- This IMDb dataset contains 12 data fields.

OBJECTIVES/SCOPE OF ANALYSIS

After analysis of the dataset, the aim of this project is to give answer of given objectives in easy way:

- Analysis of Number of Movies Released year wise
- Analysis of Top 10 IMDb rating Movies and Meta Score Movies.
- Analysis of Top 10 Longest Runtime Movies From entire Dataset
- Analysis of Total Revenue Earned by movies year wise
- Analysis of Top 10 revenue movies from the entire dataset
- Analysis of most voted director from the dataset

SOURCE OF DATASET:

Source of dataset: <https://www.kaggle.com/PromptCloudHQ/imdb-data>

CONTENT IN DATASET

This dataset is based on 1000 popular movies on IMDb from 2006 to 2016. It contains 12 columns. The columns included in the dataset are given below:

1. Rank: Movie rank order
2. Title: The title of the film
3. Genre: Category of the film
4. Description: Summary of the film
5. Director: The name of the film's director
6. Actors: Name of the Cast
7. Year: Released year of the film
8. Runtime: Duration of film in minutes
9. Rating: IMDB rating given by the user
10. Votes: Number of votes
11. Revenue: Total revenues in millions
12. Metascore: An aggregated average of critic scores. Values are between 0 and 100. Higher scores represent positive reviews.

These were the 12 columns that are present in the data set of IMDb and each column is specified clearly so the person can see the rating and the meta score of the movie and choose the movie they like.

ETL PROCESS:

- ETL is a process that extracts the data from different source systems, then transforms the data (like applying calculations, concatenations, etc.) and finally loads the data into the Data Warehouse system.
- Full form of ETL is Extract, Transform and Load.
- The triple combination of ETL provides crucial functions that are many times combined into a single application or suite of tools
- So by using the ETL process we can make the changes in the raw dataset as the raw data can be filtered the null values can be removed by using the ETL process and the unwanted columns can be removed in this method

Sample of dataset with data fields is given below:

	A	B	C	D	E	F	G	H	I	J	K	L	M	
1	Rank	Title	Genre	Description	Director	Actors	Year	Runtime (Minutes)	Rating	Votes	Revenue (Millions)	Metascore		
2	1	Guardians of the Galaxy	Action,Adventure,Sci-Fi	A group of intergalactic criminals	James Gunn	Chris Pratt, Vin Diesel, Bradley Pitt	2014	121	8.1	757074	333.13	76		
3	2	Prometheus	Adventure,Mystery,Sci-Fi	Following clues to the origin of	Ridley Scott	Noomi Rapace, Logan Marshall-Green	2012	124	7	485820	126.46	65		
4	3	Split	Horror,Thriller	Three girls are kidnapped by a	Night Shyamalan	James McAvoy, Anya Taylor-Joy	2016	117	7.3	157606	138.12	62		
5	4	Sing	Animation,Comedy,Family	In a city of humanoid animals	Christophe Lourdelet	Matthew McConaughey, Ryan Murphy	2016	108	7.2	60545	270.32	59		
6	5	Suicide Squad	Action,Adventure,Fantasy	A secret government agency	David Ayer	Will Smith, Jared Leto, Maya Erskine	2016	123	6.2	393727	325.02	40		
7	6	The Great Wall	Action,Adventure,Fantasy	European mercenaries seek	Yimou Zhang	Matt Damon, Tian Jing, Wen Jiang	2016	103	6.1	56036	45.13	42		
8	7	La La Land	Comedy,Drama,Music	A jazz pianist falls for an	Damien Chazelle	Ryan Gosling, Emma Stone	2016	128	8.3	258682	151.06	93		
9	8	Mindhorn	Comedy	A has-been actor best known	Sean Foley	Essie Davis, Andrea Riseborough	2016	89	6.4	2490		71		
10	9	The Lost City of Z	Action,Adventure,Biography	A true-life drama, centering	James Gray	Charlie Hunnam, Robert Pattinson	2016	141	7.1	7188	8.01	78		
11	10	Passengers	Adventure,Drama,Romance	A spacecraft traveling to a	Morten Tyldum	Jennifer Lawrence, Chris Pratt	2016	116	7	192177	100.01	41		
12	11	Fantastic Beasts and Where to Find Them	Adventure,Family,Fantasy	The adventures of writer	N David Yates	Eddie Redmayne, Katherine Waterston	2016	133	7.5	232072	234.02	66		
13	12	Hidden Figures	Biography,Drama,History	The story of a team of female	Theodore Melfi	Taraji P. Henson, Octavia Spencer	2016	127	7.8	93103	169.27	74		
14	13	Rogue One	Action,Adventure,Sci-Fi	The Rebel Alliance makes a	Gareth Edwards	Felicity Jones, Diego Luna, Ben Mendelsohn	2016	133	7.9	323118	532.17	65		
15	14	Moana	Animation,Adventure,Comedy	In Ancient Polynesia, when	Ron Clements	Auli'i Cravalho, Dwayne Johnson	2016	107	7.7	118151	248.75	81		
16	15	Colossal	Action,Comedy,Drama	Gloria is an out-of-work	p Nacho Vigalondo	Anne Hathaway, Jason Sudeikis	2016	109	6.4	8612	2.87	70		
17	16	The Secret Life of Pets	Animation,Adventure,Comedy	The quiet life of a terrier	n Chris Renaud	Louis C.K., Eric Stonestreet	2016	87	6.6	120259	368.31	61		
18	17	Hacksaw Ridge	Biography,Drama,History	WWII American Army Med	Mel Gibson	Andrew Garfield, Sam Worthington	2016	139	8.2	211760	67.12	71		
19	18	Jason Bourne	Action,Thriller	The CIA's most dangerous	Paul Greengrass	Matt Damon, Tommy Lee Jones	2016	123	6.7	150823	162.16	58		
20	19	Lion	Biography,Drama	A five-year-old Indian boy	Garth Davis	Dev Patel, Nicole Kidman, Li Cui	2016	118	8.1	102061	51.69	69		
21	20	Arrival	Drama,Mystery,Sci-Fi	When twelve mysterious	discoveries	Villeneuve	Amy Adams, Jeremy Renne	2016	116	8	340798	100.5	81	
22	21	Gold	Adventure,Drama,Thriller	Kenny Wells, a prospector	Stephen Gaghan	Matthew McConaughey, Eddie Redmayne	2016	120	6.7	19053	7.22	49		
23	22	Manchester by the Sea	Drama	A depressed uncle is asked	Kenneth Lonergan	Casey Affleck, Michelle Williams	2016	137	7.9	134213	47.7	96		
24	23	Hounds of Love	Crime,Drama,Horror	A cold-blooded predatory	r Ben Young	Emma Booth, Ashleigh Crum	2016	108	6.7	1115		72		
25	24	Trolls	Animation,Adventure,Comedy	After the Bergens invade	Ti Walt Dohrn	Anna Kendrick, Justin Timberlake	2016	92	6.5	38552	153.69	56		
26	25	Independence Day: Resur	Action,Adventure,Sci-Fi	Two decades after the first	Roland Emmerich	Liam Hemsworth, Jeff Goldblum	2016	120	5.3	127553	103.14	32		
27	26	Darjeeling Night	Comedy	Fiona visits Darjeeling for	the first time	Dominique		83	6.8	777				

Steps taken to clean dataset thorough ETL process

Step 1:

- Firstly, open blank excel file and go to Data from Tab and select New query and entered data csv file from it.

Rank	Title	Genre	Description	Director	Actors
1	Guardians of the Galaxy	Action,Adventure,Sci-Fi	A group of intergalactic criminals are forced to work together to stop a...	James Gunn	Chris Pratt, Vin Diesel, Bradley Cooper...
2	Prometheus	Adventure,Mystery,Sci-Fi	Following clues to the origin of mankind, a team finds a structure on a...	Ridley Scott	Noomi Rapace, Logan Marshall-Gree...
3	Split	Horror,Thriller	Three girls are kidnapped by a man with a diagnosed 23 distinct perso...	M. Night Shyamalan	James McAvoy, Anya Taylor-Joy, Ha...
4	Sing	Animation,Comedy,Family	In a city of humanoid animals, a hustling theater impresario's attempt...	Christopher Lourdelet	Matthew McConaughey,Reese Wit...
5	Suicide Squad	Action,Adventure,Fantasy	A secret government agency recruits some of the most dangerous inca...	David Ayer	Will Smith, Jared Leto, Margot Rob...
6	The Great Wall	Action,Adventure,Fantasy	European mercenaries searching for black powder become embroiled ...	Yimou Zhang	Matt Damon, Tian Jing, Willem Da...
7	La La Land	Comedy,Drama,Music	A jazz pianist falls for an aspiring actress in Los Angeles.	Damien Chazelle	Ryan Gosling, Emma Stone, Rosem...
8	Mindhunters	Comedy	A has-been actor best known for playing the title character in the 1980...	Sean Foley	Essie Davis, Andrea Riseborough, J...
9	The Lost City of Z	Action,Adventure,Biography	A true-life drama, centering on British explorer Col. Percival Fawcett, ...	James Gray	Charlie Hunnam, Robert Pattinson...
10	Passengers	Adventure,Drama,Romance	A spacecraft traveling to a distant colony planet and transporting thou...	Morten Tyldum	Jennifer Lawrence, Chris Pratt, Mic...
11	Fantastic Beasts and Where to Find Them	Adventure,Family,Fantasy	The adventures of writer Newt Scamander in New York's secret comm...	David Yates	Eddie Redmayne, Katherine Water...
12	Hidden Figures	Biography,Drama,History	The story of a team of female African-American mathematicians who s...	Theodore Melfi	Taraji P. Henson, Octavia Spencer...
13	Rogue One	Action,Adventure,Sci-Fi	The Rebel Alliance makes a risky move to steal the plans for the Death...	Gareth Edwards	Felicity Jones, Diego Luna, Alan T...
14	Moana	Animation,Adventure,Comedy	In Ancient Polynesia, when a terrible curse incurred by the Demigod M...	Ron Clements	Auli'i Cravalho, Dwayne Johnson, R...
15	Colossal	Action,Comedy,Drama	Gloria is an out-of-work party girl forced to leave her life in New York...	Nacho Vigalondo	Anne Hathaway, Jason Sudeikis, Au...
16	The Secret Life of Pets	Animation,Adventure,Comedy	The quiet life of a terrier named Max is upended when his owner take...	Chris Renaud	Louis C.K., Eric Stonestreet, Kevin H...
17	Hacksaw Ridge	Biography,Drama,History	WWII American Army Medic Desmond T. Doss, who served during the war...	Mel Gibson	Andrew Garfield, Sam Worthington...
18	Jason Bourne	Action,Thriller	The CIA's most dangerous former operative is drawn out of hiding to u...	Paul Greengrass	Matt Damon, Tommy Lee Jones, Ali...
19	Lion	Biography,Drama	A five-year-old Indian boy gets lost on the streets of Calcutta, thousa...	Garth Davis	Dev Patel, Nicole Kidman, Rooney F...
20	Arrival	Drama,Mystery,Sci-Fi	When twelve mysterious spacecraft appear around the world, linguist...	Denis Villeneuve	Amy Adams, Jeremy Renner, Forest...
21	Gold	Adventure,Drama,Thriller	Kenny Wells, a prospector desperate for a lucky break, teams up with...	Stephen Gaghan	Matthew McConaughey, Edgar Ran...
22	Manchester by the Sea	Drama	A depressed uncle is asked to take care of his teenage nephew after th...	Kenneth Lonergan	Casey Affleck, Michelle Williams, K...
23	Hounds of Love	Crime,Drama,Horror	A cold-blooded predatory couple while cruising the streets in search of...	Ben Young	Emma Booth, Ashleigh Cummings...
24	Trolls	Animation,Adventure,Comedy	After the Bergens invade Troll Village, Poppy, the happiest Troll ever b...	Walt Doherty	Anna Kendrick, Justin Timberlake, J...
25	Independence Day: Resurgence	Action,Adventure,Sci-Fi	Two decades after the first Independence Day invasion, Earth is faced...	Roland Emmerich	Liam Hemsworth, Jeff Goldblum, Bi...
26	Paris pieds nus	Comedy	Fiona visits Paris for the first time to assist her myopic Aunt Martha. C...	Dominique Abel	Fiona Gordon, Dominique Abel,Emi...
27	Bahubali: The Beginning	Action,Adventure,Drama	In ancient India, an adventurous and daring man becomes involved in ...	S.S. Rajamouli	Prabhas, Rana Daggubati, Anushka...
28	Dead Awake	Horror,Thriller	A young woman must save herself and her friends from an ancient evil...	Philip Guzman	Jocelin Donahue, Jesse Bradford, Je...

Step 2:

- After entering the query editor, data fields title will promote into header

Step 3:

- Changed data type will change the type of data fields such as metascores converted from text into number, Revenue converted from text into currency

Director	Actors	Year	Runtime (Minutes)	Rating	Votes	Revenue (Millions)	Metascore
Gunn	Chris Pratt, Vin Diesel, Bradley Cooper, Zoe Saldana	2014	121	8.1	757074	333.13	
Scott	Noomi Rapace, Logan Marshall-Green, Michael Fassbender, Charlize T...	2012	124	7	485820	126.46	
Shyamalan	James McAvoy, Anya Taylor-Joy, Haley Lu Richardson, Jessica Sula	2016	117	7.3	157606	188.12	
Lourdelet	Matthew McConaughey,Reese Witherspoon, Seth MacFarlane, Scarlet...	2016	108	7.2	60545	270.32	
Ayer	Will Smith, Jared Leto, Margot Robbie, Viola Davis	2016	123	6.2	893727	125.02	
Zhang	Matt Damon, Tian Jing, Willem Dafoe, Andy Lau	2016	103	6.1	56036	45.13	
Chazelle	Ryan Gosling, Emma Stone, Rosemarie DeWitt, J.K. Simmons	2016	128	8.3	258882	251.06	
Foley	Essie Davis, Andrea Riseborough, Julian Barratt,Kenneth Branagh	2016	89	6.4	2490	null	
Gray	Charlie Hunnam, Robert Pattinson, Sienna Miller, Tom Holland	2016	141	7.1	7188	8.01	
Tyldum	Jennifer Lawrence, Chris Pratt, Michael Sheen,Laurence Fishburne	2016	116	7	192177	100.01	
Yates	Eddie Redmayne, Katherine Waterston, Allison Sudol,Dan Fogler	2016	133	7.5	232072	234.02	
Melfi	Taraji P. Henson, Octavia Spencer, Janelle Monáe,Kevin Costner	2016	127	7.8	83103	189.27	
Edwards	Felicity Jones, Diego Luna, Alan Tudyk, Donnie Yen	2016	133	7.9	323118	532.17	
Clements	Auli'i Cravalho, Dwayne Johnson, Rachel House, Temuera Morrison	2016	107	7.7	118151	248.75	
Vigalondo	Anne Hathaway, Jason Sudeikis, Austin Stowell,Tim Blake Nelson	2016	109	6.4	8612	2.87	
Renaud	Louis C.K., Eric Stonestreet, Kevin Hart, Lake Bell	2016	87	6.6	120259	368.31	
Gibson	Andrew Garfield, Sam Worthington, Luke Bracey,Teresa Palmer	2016	139	8.2	211760	67.12	
Greengrass	Matt Damon, Tommy Lee Jones, Alicia Vikander,Vincent Cassel	2016	123	6.7	150823	162.16	
Davis	Dev Patel, Nicole Kidman, Rooney Mara, Sunny Pawar	2016	118	8.1	102061	51.69	
Villeneuve	Amy Adams, Jeremy Renner, Forest Whitaker,Michael Stuhlbarg	2016	116	8	340798	100.50	
Gaghan	Matthew McConaughey, Edgar Ramirez, Bryce Dallas Howard, Corey S...	2016	120	6.7	29053	7.22	
Lonergan	Casey Affleck, Michelle Williams, Kyle Chandler, Lucas Hedges	2016	137	7.9	114213	47.70	
Young	Emma Booth, Ashleigh Cummings, Stephen Curry,Susie Porter	2016	108	6.7	1115	null	
John	Anna Kendrick, Justin Timberlake,Zoey Deschanel, Christopher Mintz...	2016	92	6.5	38552	153.89	
Emmerich	Liam Hemsworth, Jeff Goldblum, Bill Pullman,Mala Morone	2016	120	5.3	127553	109.14	
Abel	Fiona Gordon, Dominique Abel,Loumanuelle Riva, Pierre Richard	2016	82	6.8	222	null	
Rajamouli	Prabhas, Rana Daggubati, Anushka Shetty,Tamannaah Bhatia	2015	159	8.3	76193	6.50	
Guzman	Jocelin Donahue, Jesse Bradford, Jesse Borrego,Lori Petty	2016	99	4.7	523	0.01	

Step 4:

- Now to remove null values in the rows of meta score revenue and few other columns by going to the filter option and click on remove empty fields so those rows having null values will be removed

Table: SelectRows("#Filtered Rows4", each [Year] <> null and [Year] <> "")

Actor	Year	Runtime (Minutes)	Rating	Votes	Revenue (Millions)	Metascore
Chris Pratt, Vin Diesel, Bradley Cooper, Zoe Saldana	2014	121	8.1	757074	333.13	76
Noomi Rapace, Logan Marshall-Green, Michael Fassbender, Charlize T...	2012	124	7	485820	126.46	65
James McAvoy, Anya Taylor-Joy, Haley Lu Richardson, Jessica Sula	2016	117	7.3	157606	138.12	62
Matthew McConaughey, Reese Witherspoon, Seth MacFarlane, Scarlett...	2016	108	7.2	60545	270.32	59
Will Smith, Jared Leto, Margot Robbie, Viola Davis	2016	123	6.2	393727	325.02	40
Matt Damon, Tian Jing, Willem Dafoe, Andy Lau	2016	103	6.1	36036	45.13	42
Ryan Gosling, Emma Stone, Rosemarie DeWitt, J.K. Simmons	2016	128	8.3	258682	151.06	93
Charlie Hunnam, Robert Pattinson, Sienna Miller, Tom Holland	2016	141	7.1	7188	8.01	78
Jennifer Lawrence, Chris Pratt, Michael Sheen, Laurence Fishburne	2016	116	7	192177	100.01	41
Eddie Redmayne, Katherine Waterston, Allison Sudol, Dan Fogler	2016	133	7.5	232072	234.02	66
Taraji P. Henson, Octavia Spencer, Janelle Monáe, Kevin Costner	2016	127	7.8	93103	169.27	74
Felicity Jones, Diego Luna, Alan Tudyk, Donnie Yen	2016	133	7.9	923118	532.17	65
Auli'i Cravalho, Dwayne Johnson, Rachel House, Temuera Morrison	2016	107	7.7	118151	248.75	81
Anne Hathaway, Jason Sudeikis, Austin Stowell, Tim Blake Nelson	2016	109	6.4	8612	2.87	70
Louis C.K., Eric Stonestreet, Kevin Hart, Lake Bell	2016	87	6.6	120239	368.31	61
Andrew Garfield, Sam Worthington, Luke Bracey, Teresa Palmer	2016	139	8.2	211760	67.12	71
Matt Damon, Tommy Lee Jones, Alicia Vikander, Vincent Cassel	2016	123	6.7	150823	162.16	58
Dev Patel, Nicole Kidman, Rooney Mara, Sunny Pawar	2016	118	8.1	102061	51.69	69
Amy Adams, Jeremy Renner, Forest Whitaker, Michael Stuhlbarg	2016	116	8	340798	100.50	81
Matthew McConaughey, Edgar Ramirez, Bryce Dallas Howard, Corey S...	2016	120	6.7	19053	7.22	49
Cassey Affleck, Michelle Williams, Kyle Chandler, Lucas Hedges	2016	137	7.9	134213	47.70	96
Anna Kendrick, Justin Timberlake, Zoëy Deschanel, Christopher Mintz...	2016	92	6.5	38552	153.69	56
Liam Hemsworth, Jeff Goldblum, Bill Pullman, Malika Monroe	2016	120	5.3	127553	103.14	32
Mila Kunis, Kathryn Hahn, Kristen Bell, Christina Applegate	2016	100	6.2	46540	113.08	60
Michael Fassbender, Marion Cotillard, Jeremy Irons, Brendan Gleeson	2016	115	5.9	112813	54.65	36
Zoe Deutch, James Franco, Tanglee Ambrose, Cedric the Entertainer	2016	111	6.3	48123	60.21	39
Amy Adams, Jake Gyllenhaal, Michael Shannon, Aaron Taylor-Johnson	2016	116	7.5	126030	10.64	67
James McAvoy, Michael Fassbender, Jennifer Lawrence, Nicholas Hoult	2016	144	7.1	275510	155.33	52

Step 5:

- Now, from query editor select Home tab and remove column which are not in required for analysis. Column which are removed from dataset are given below:

- Rank
- Genre
- Description
- Actor

Table.RemoveColumns(#"Filtered Rows5",{"Rank", "Genre", "Description", "Actors"})

	Title	Director	Year	Runtime (Minutes)	Rating	Votes	Revenue (Millions)	Metascore
1	Guardians of the Galaxy	James Gunn	2014	121	8.1	757074	333.13	76
2	Prometheus	Ridley Scott	2012	124	7	485820	126.46	65
3	Split	M. Night Shyamalan	2016	117	7.3	157606	138.12	62
4	Sing	Christophe Lourdelet	2016	108	7.2	60545	270.32	59
5	Suicide Squad	David Ayer	2016	123	6.2	393727	325.02	40
6	The Great Wall	Yimou Zhang	2016	103	6.1	56036	45.13	42
7	La La Land	Damien Chazelle	2016	128	8.3	258682	151.06	93
8	The Lost City of Z	James Gray	2016	141	7.1	7188	8.01	78
9	Passengers	Morten Tyldum	2016	116	7	192177	100.01	41
10	Fantastic Beasts and Where to Find Them	David Yates	2016	133	7.5	232072	234.02	66
11	Hidden Figures	Theodore Melfi	2016	127	7.8	93103	169.27	74
12	Rogue One	Gareth Edwards	2016	133	7.9	323118	532.17	65
13	Moana	Ron Clements	2016	107	7.7	118151	248.75	81
14	Colossal	Nacho Vigalondo	2016	109	6.4	8612	2.87	70
15	The Secret Life of Pets	Chris Renaud	2016	87	6.6	120259	368.31	61
16	Hacksaw Ridge	Mel Gibson	2016	139	8.2	211760	67.12	71
17	Jason Bourne	Paul Greengrass	2016	123	6.7	150823	162.16	58
18	Lion	Garth Davis	2016	118	8.1	102061	51.69	69
19	Arrival	Denis Villeneuve	2016	116	8	340798	100.50	81
20	Gold	Stephen Gaghan	2016	120	6.7	19053	7.22	49
21	Manchester by the Sea	Kenneth Lonergan	2016	137	7.9	134213	47.70	96
22	Trolls	Walt Dohrn	2016	92	6.5	38552	153.69	56
23	Independence Day: Resurgence	Roland Emmerich	2016	120	5.3	127553	103.14	32
24	Bad Moms	Jon Lucas	2016	100	6.2	66540	113.08	60
25	Assassin's Creed	Justin Kurzel	2016	115	5.9	112813	54.65	36
26	Why Him?	John Hamburg	2016	111	6.3	48123	60.31	39
27	Nocturnal Animals	Tom Ford	2016	116	7.5	126030	10.64	67
28	X-Men: Apocalypse	Bryan Singer	2016	144	7.1	275510	155.33	52
29	Deadpool	Tim Miller	2016	108	8	627797	363.02	65

8 COLUMNS, 836 ROWS

PREVIEW DOWNLOADED AT 9

Step 6:

- Finally, sort the dataset through filter of rating data fields

Step 7:

- Remove duplicates from the data using duplicate option from home tab

Table.Distinct(#"Removed Columns1",{"Title"})

	Title	Director	Year	Runtime (Minutes)	Rating	Votes	Revenue (Millions)	Metascore
1	Guardians of the Galaxy	James Gunn	2014	121	8.1	757074	333.13	76
2	Prometheus	Ridley Scott	2012	124	7	485820	126.46	65
3	Split	M. Night Shyamalan	2016	117	7.3	157606	138.12	62
4	Sing	Christophe Lourdelet	2016	108	7.2	60545	270.32	59
5	Suicide Squad	David Ayer	2016	123	6.2	393727	325.02	40
6	The Great Wall	Yimou Zhang	2016	103	6.1	56036	45.13	42
7	La La Land	Damien Chazelle	2016	128	8.3	258682	151.06	93
8	The Lost City of Z	James Gray	2016	141	7.1	7188	8.01	78
9	Passengers	Morten Tyldum	2016	116	7	192177	100.01	41
10	Fantastic Beasts and Where to Find Them	David Yates	2016	133	7.5	232072	234.02	66
11	Hidden Figures	Theodore Melfi	2016	127	7.8	93103	169.27	74
12	Rogue One	Gareth Edwards	2016	133	7.9	323118	532.17	65
13	Moana	Ron Clements	2016	107	7.7	118151	248.75	81
14	Colossal	Nacho Vigalondo	2016	109	6.4	8612	2.87	70
15	The Secret Life of Pets	Chris Renaud	2016	87	6.6	120259	368.31	61
16	Hacksaw Ridge	Mel Gibson	2016	139	8.2	211760	67.12	71
17	Jason Bourne	Paul Greengrass	2016	123	6.7	150823	162.16	58
18	Lion	Garth Davis	2016	118	8.1	102061	51.69	69
19	Arrival	Denis Villeneuve	2016	116	8	340798	100.50	81
20	Gold	Stephen Gaghan	2016	120	6.7	19053	7.22	49
21	Manchester by the Sea	Kenneth Lonergan	2016	137	7.9	134213	47.70	96
22	Trolls	Walt Dohrn	2016	92	6.5	38552	153.69	56
23	Independence Day: Resurgence	Roland Emmerich	2016	120	5.3	127553	103.14	32
24	Bad Moms	Jon Lucas	2016	100	6.2	66540	113.08	60
25	Assassin's Creed	Justin Kurzel	2016	115	5.9	112813	54.65	36
26	Why Him?	John Hamburg	2016	111	6.3	48123	60.31	39
27	Nocturnal Animals	Tom Ford	2016	116	7.5	126030	10.64	67
28	X-Men: Apocalypse	Bryan Singer	2016	144	7.1	275510	155.33	52
29	Deadpool	Tim Miller	2016	108	8	627797	363.02	65

8 COLUMNS, 836 ROWS

PREVIEW DOWNLOADED AT 900 P

Finally, after cleaning the data, the final dataset sample is shown below:

	Title	Director	Year	Runtime (Minutes)	Rating	Votes	Revenue (Millions)	Metascore
1	The Dark Knight	Christopher Nolan	2008	152	9	1791916	\$533.32	82
2	Inception	Christopher Nolan	2010	148	8.8	1583625	\$292.57	74
3	Interstellar	Christopher Nolan	2014	169	8.6	1047747	\$187.99	74
4	The Intouchables	Olivier Nakache	2011	112	8.6	557965	\$13.18	57
5	Kimi no na wa	Makoto Shinkai	2016	106	8.6	34110	\$4.68	79
6	The Dark Knight Rises	Christopher Nolan	2012	164	8.5	1222645	\$448.13	78
7	The Departed	Martin Scorsese	2006	151	8.5	937414	\$132.37	85
8	The Prestige	Christopher Nolan	2006	130	8.5	913152	\$53.08	66
9	Whiplash	Damien Chazelle	2014	107	8.5	477276	\$13.09	88
10	The Lives of Others	Florian Henckel von Donnersmarck	2006	137	8.5	278103	\$11.28	89
11	Taare Zameen Par	Aamir Khan	2007	165	8.5	102697	\$1.20	42
12	Django Unchained	Quentin Tarantino	2012	165	8.4	1039115	\$162.80	81
13	3 Idiots	Rajkumar Hirani	2009	170	8.4	238789	\$6.52	67
14	Toy Story 3	Lee Unkrich	2010	103	8.3	586669	\$414.98	92
15	Up	Pete Docter	2009	96	8.3	722203	\$292.98	88
16	La La Land	Damien Chazelle	2016	128	8.3	258682	\$151.06	93
17	Inglourious Basterds	Quentin Tarantino	2009	153	8.3	959065	\$120.52	69
18	Jagten	Thomas Vinterberg	2012	115	8.3	192263	\$0.61	76
19	Inside Out	Pete Docter	2015	95	8.2	416689	\$356.45	94
20	The Wolf of Wall Street	Martin Scorsese	2013	180	8.2	865134	\$116.87	75
21	Hacksaw Ridge	Mel Gibson	2016	139	8.2	211760	\$67.12	71
22	Pan's Labyrinth	Guillermo del Toro	2006	118	8.2	498879	\$37.62	98
23	El secreto de sus ojos	Juan José Campanella	2009	129	8.2	144524	\$20.17	80
24	Room	Lenny Abrahamson	2015	118	8.2	224132	\$14.68	86
25	Warrior	Gavin O'Connor	2011	140	8.2	355722	\$13.65	71
26	Ok	Rajkumar Hirani	2014	153	8.2	103370	\$10.52	51

Analysis on dataset

1.Total revenues(millions)Earned by Movie year-wise

Introduction:

By performing this analysis, we will Total revenues earn by movies from 2006 to 2016.

Description:

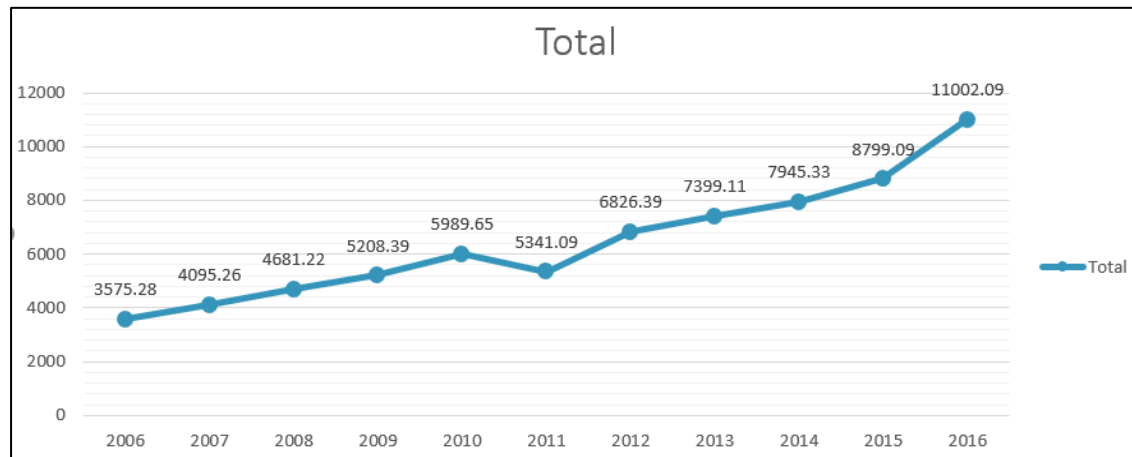
The analysis based on the year, revenue of the dataset

Specific requirements, functions and formulas:

1. Pivot table is used for the analysis.
2. Sum function is used in pivot table for the sum of the revenues in the pivot table.
3. Formatting of background colour is used in the columns.

Year	Sum of Revenue (Millions)
2006	3575.28
2007	4095.26
2008	4681.22
2009	5208.39
2010	5989.65
2011	5341.09
2012	6826.39
2013	7399.11
2014	7945.33
2015	8799.09
2016	11002.09
Grand Total	70862.9

Visualization



2. Analysis of Number of Movies release Year-wise

Introduction:

By performing this analysis, we will get Number of movies released in each year

Description:

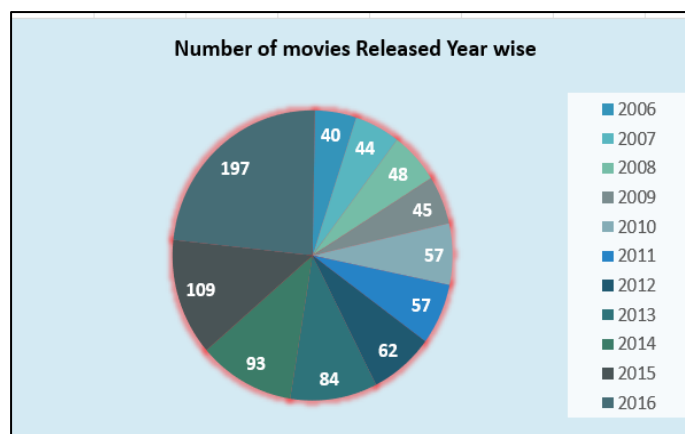
The analysis based on the year, title of the dataset

Specific requirements, functions and formulas

1. Pivot table is used for the analysis.
2. Count function is used in pivot table for the counting of the title in the pivot table in their respective year

Year	Sum of Revenue (Millions)
2006	3575.28
2007	4095.26
2008	4681.22
2009	5208.39
2010	5989.65
2011	5341.09
2012	6826.39
2013	7399.11
2014	7945.33
2015	8799.09
2016	11002.09
Grand Total	70862.9

Visualization



3.Top 10 longest movies from the entire dataset year wise

Introduction:

By performing this analysis, we will get Top 10 longest movies of all time.

Description:

The analysis is based on title, runtimes of the dataset

Filter is used to show top 10 movies

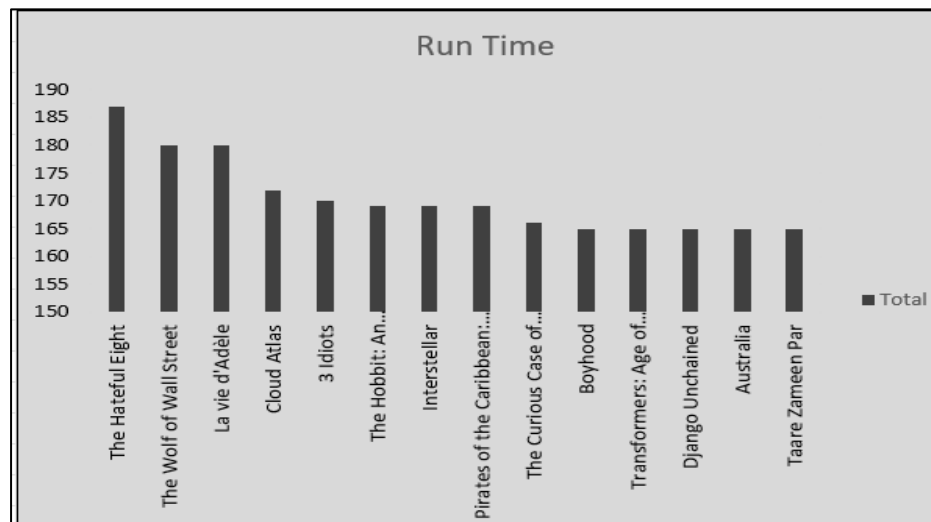
Slicer is connected with the graph to get each year top 10 run time movies

Specific requirements, functions and formulas:

1. Pivot table is used for the analysis.
2. max function is used in pivot table for the max of runtimes of the title
3. A slicer is used to select the year and analyze the data.

Movie name	Max of Runtime (Minutes)
Superman Returns	154
The Departed	151
Pirates of the Caribbean: Dead Man's Chest	151
The Da Vinci Code	149
Perfume: The Story of a Murderer	147
Casino Royale	144
Blood Diamond	143
Babel	143
Apocalypto	139
The Lives of Others	137
Maximum one among the 10	154

Visualization



4.Top 10 Revenue Movies from the entire dataset

Introduction:

By performing this analysis, we will get Top 10 revenue movies of all time.

Description:

1. The analysis is based on title, revenue of the dataset
2. Filter is used to show top 10 movies with max revenue.

Specific requirements, functions and formulas:

1. Pivot table is used for the analysis.
2. max function is used in pivot table for the maximum of revenue of the title.

Movie Title	Max of Revenue (Millions)
Star Wars: Episode VII - The Force Awakens	936.63
Avatar	760.51
Jurassic World	652.18
The Avengers	623.28
The Dark Knight	533.32
Rogue One	532.17
Finding Dory	486.29
Avengers: Age of Ultron	458.99
The Dark Knight Rises	448.13
The Hunger Games: Catching Fire	424.65
Highest revenue	936.63

Visualization



5. Most Voted Director

Introduction:

By performing this analysis, we will get Top 10 Most voted director from the dataset

Description:

The analysis is based on Director, Votes of the dataset

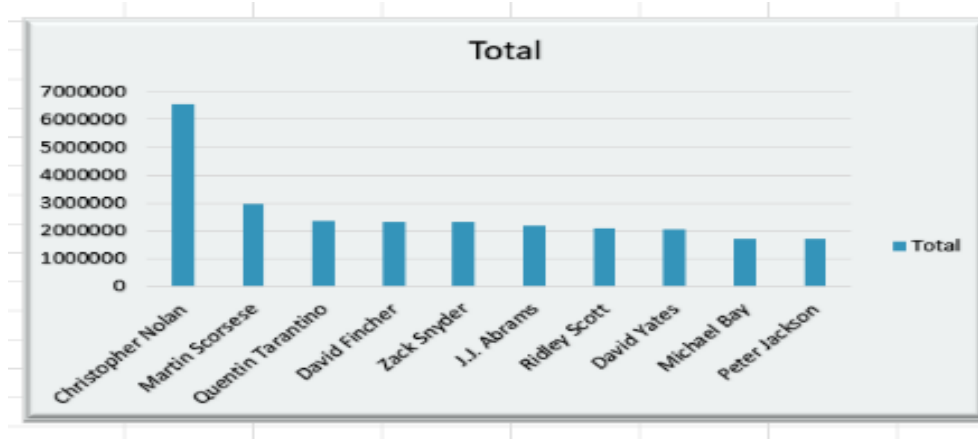
Filter is used to show top 10 most voted director

Specific requirements, functions and formulas

1. Pivot table is used for the analysis.
2. Sum function is used in pivot table for the sum of votes for the correspondence director.

Row Labels	Sum of Votes
Christopher Nolan	6559085
Martin Scorsese	2966524
Quentin Tarantino	2339350
David Fincher	2309652
Zack Snyder	2301544
J.J. Abrams	2174937
Ridley Scott	2080074
David Yates	2033854
Michael Bay	1706531
Peter Jackson	1698695
Total Votes of Top 10 Directors	26170246

Visualization



6.Top 10 rating movies and Meta scores movies

Introduction:

By performing this analysis, we will get top 10 Movies with highest IMDb rating as well as top 10 movies with highest meta scores.

Description:

The analysis is based on title, rating, meta score of the movies.

Specific requirements, functions and formulas

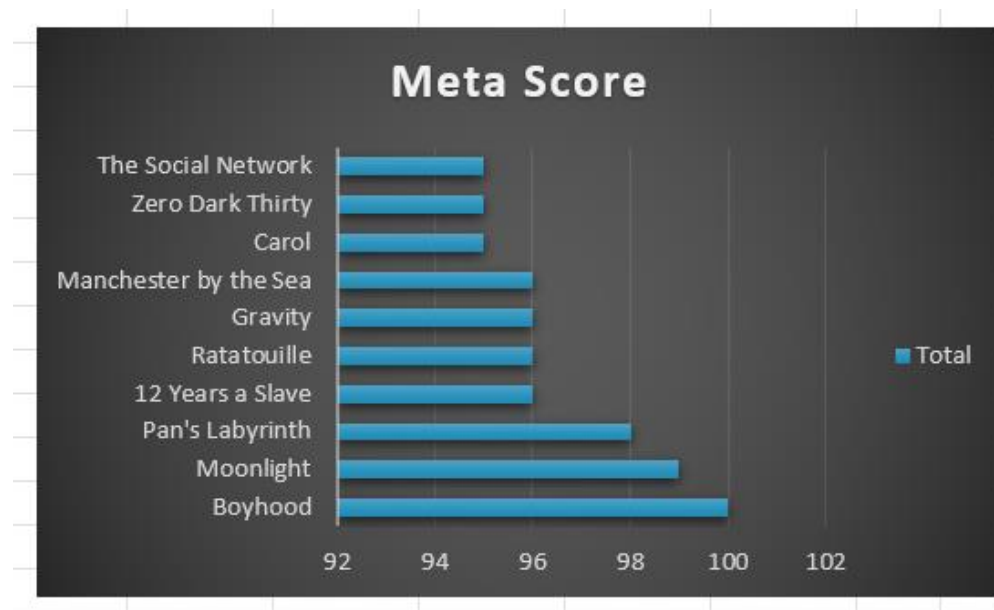
1. Pivot table is used for the analysis.

2. max function is used in pivot table for the maximum rating of the movie and the meta score is also filtered and everything is sorted in descending order. Pivot chart representation is done for it.

Row Labels	Max of Rating	Row Labels	Max of Metascore
The Dark Knight	9	Boyhood	100
Inception	8.8	Moonlight	99
Kimi no na wa	8.6	Pan's Labyrinth	98
The Intouchables	8.6	12 Years a Slave	96
Interstellar	8.6	Ratatouille	96
The Lives of Others	8.5	Gravity	96
Taare Zameen Par	8.5	Manchester by the Sea	96
The Prestige	8.5	Carol	95
The Departed	8.5	Zero Dark Thirty	95
Whiplash	8.5	The Social Network	95
The Dark Knight Rises	8.5		
Top most Movie	9	Top most Movie	100

Visualization

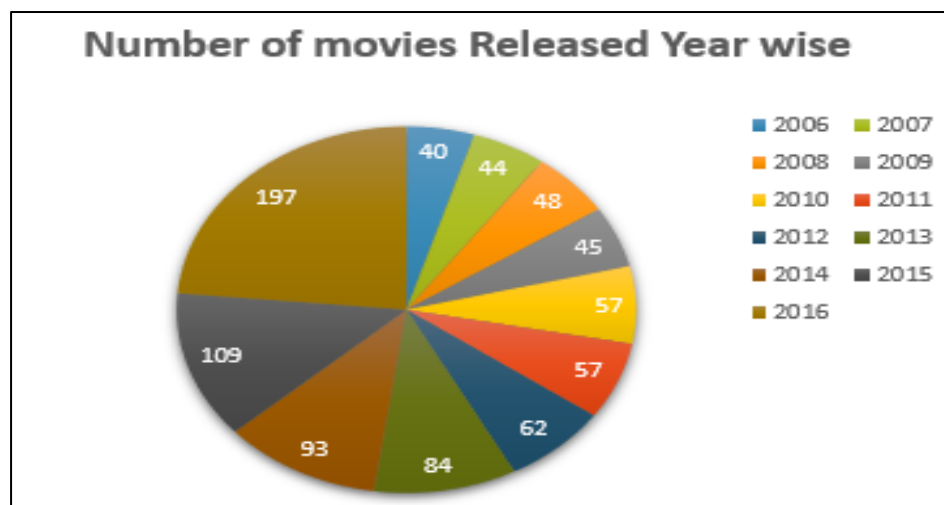




List of Analysis with results

1. Analysis of number of movies released year-wise

- The highest number of movies were released in 2016 with total 198 movies.
- The lowest number of movies were released in 2006 with total 40 movies.

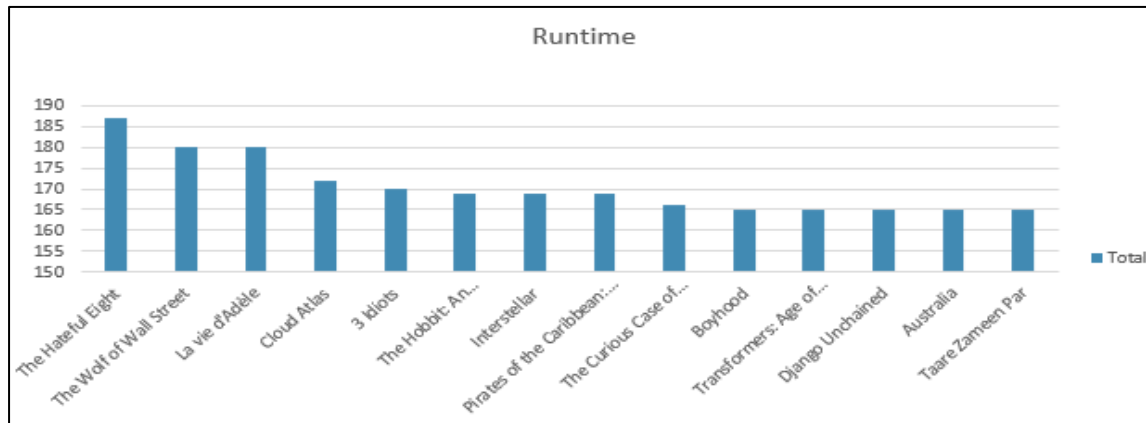


1. Analysis of Top 10 longest movies from the dataset

1. The Hateful Eight
2. La vie d'Adèle
3. The Wolf of Wall Street
4. Cloud Atlas
5. 3 Idiots
6. Pirates of the Caribbean: At World's End
7. The Hobbit: An Unexpected Journey
8. Interstellar
9. The Curious Case of Benjamin Button

10.Transformers: Age of Extinction

- The Hateful eight is the movie with 187 minutes which is the longest movie in the entire dataset
- Ma vie de Courgette is the movie with 66 minutes as a run time is the shortest movies in the entire dataset

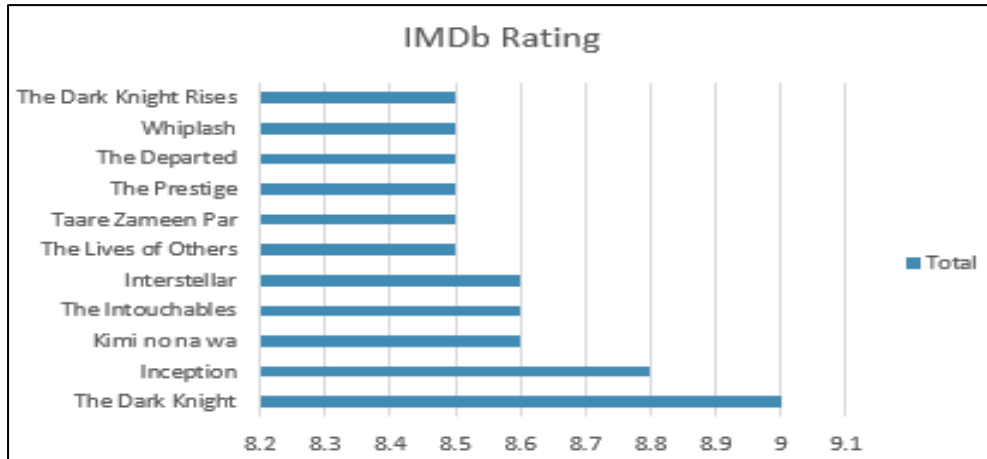


2.Analysis of Top 10 IMDb rating movies and Metascore movies

1. Top 10 IMDb rating movies:

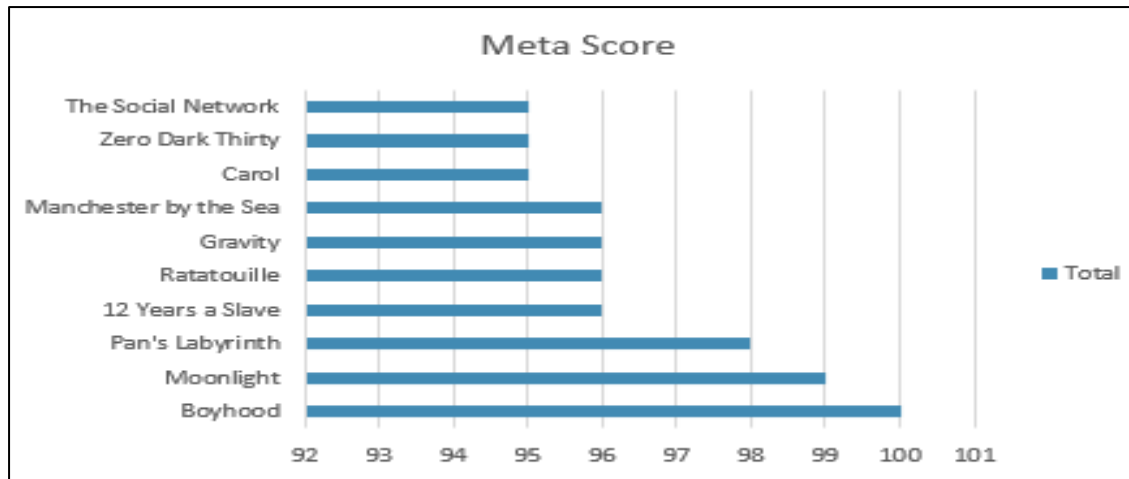
1. The Dark Knight
2. Inception
3. Kimi no na wa
4. The Intouchables
5. Interstellar
6. The Lives of Others
7. Taare Zameen Par
8. The Prestige
9. The Departed
10. Whiplash

- The Dark knight is the movie with highest IMDb rating in the dataset which is of 9.
- Disaster movie is the movie with lowest IMDb rating in the dataset which is of 1.9.



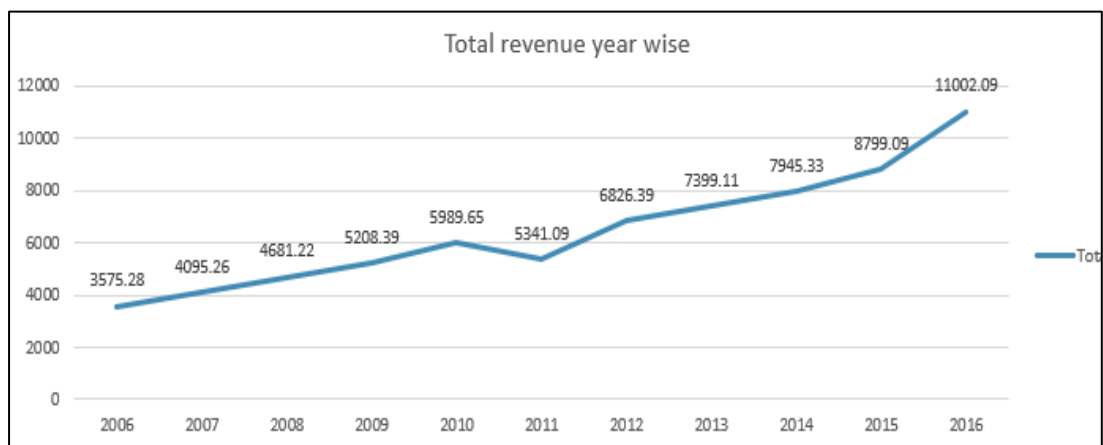
1. Top 10 Metascore movies:

1. Boyhood
 2. Moonlight
 3. Pan's Labyrinth
 4. Gravity
 5. Ratatouille
 6. 12 Years a Slave
 7. Manchester by the Sea
 8. Carol
 9. The Social Network
 10. Zero Dark Thirty
- Boyhood is the movie from the entire dataset having the highest value of meta score which is of 100
 - Nine lives is the movie from the entire dataset having the lowest value of meta score which is of 11.



2. Analysis of Total revenue earned by movies year wise

- The Highest Revenue was recorded in 2016 with total \$11002.09 million.
- The lowest Revenue was recorded in 2006 with total \$3575.28 million.

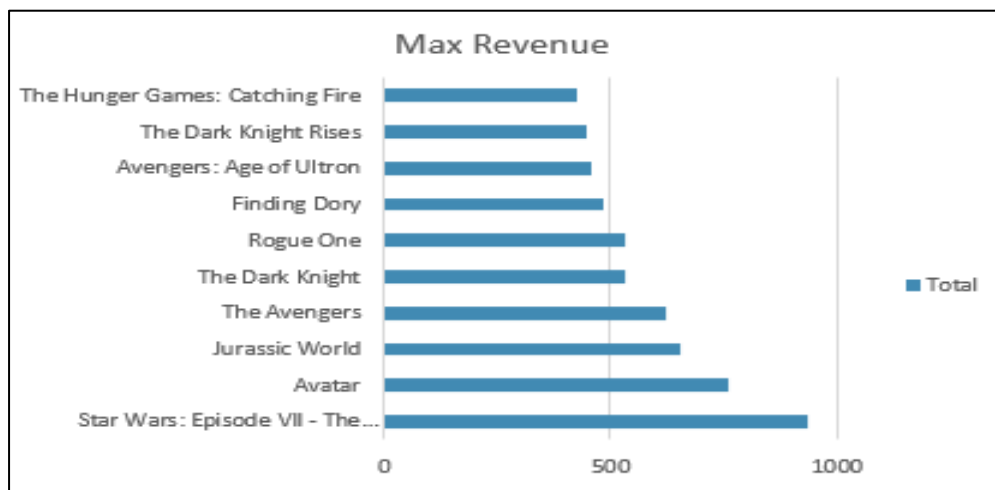


2. Analysis of Top 10 revenue movies from the dataset

1. Star Wars: Episode VII - The Force Awakens
2. Avatar

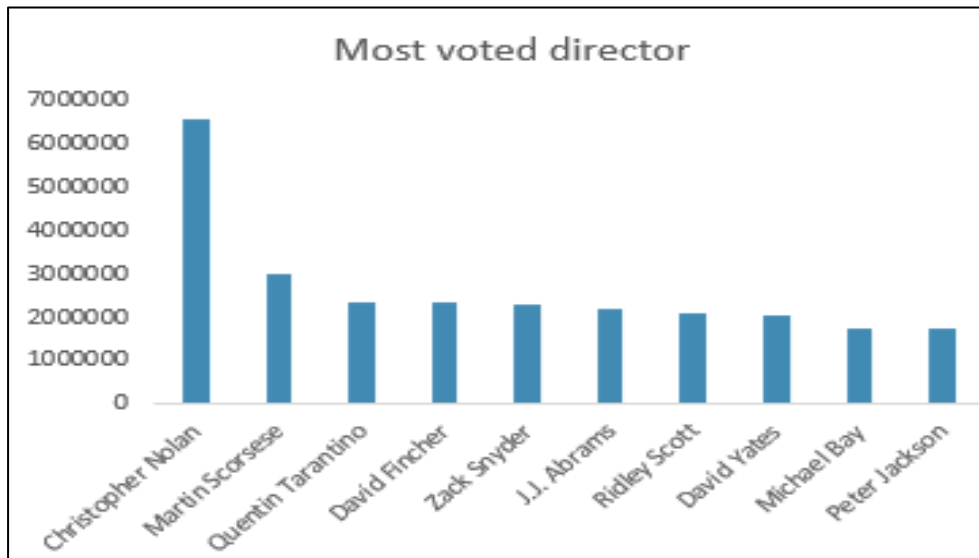
3. Jurassic World
4. The Avengers
5. The Dark Knight
6. Rogue One
7. Finding Dory
8. Avengers: Age of Ultron
9. The Dark Knight Rises
10. The Hunger Games: Catching Fire

- Star Wars: Episode VII - The Force Awakens is the movie from the entire dataset having highest revenue movie which is of 936.63 million dollars.
- Love, Rosie is the movie from the entire dataset having the lowest revenue movie which is of 0.01 million dollars.

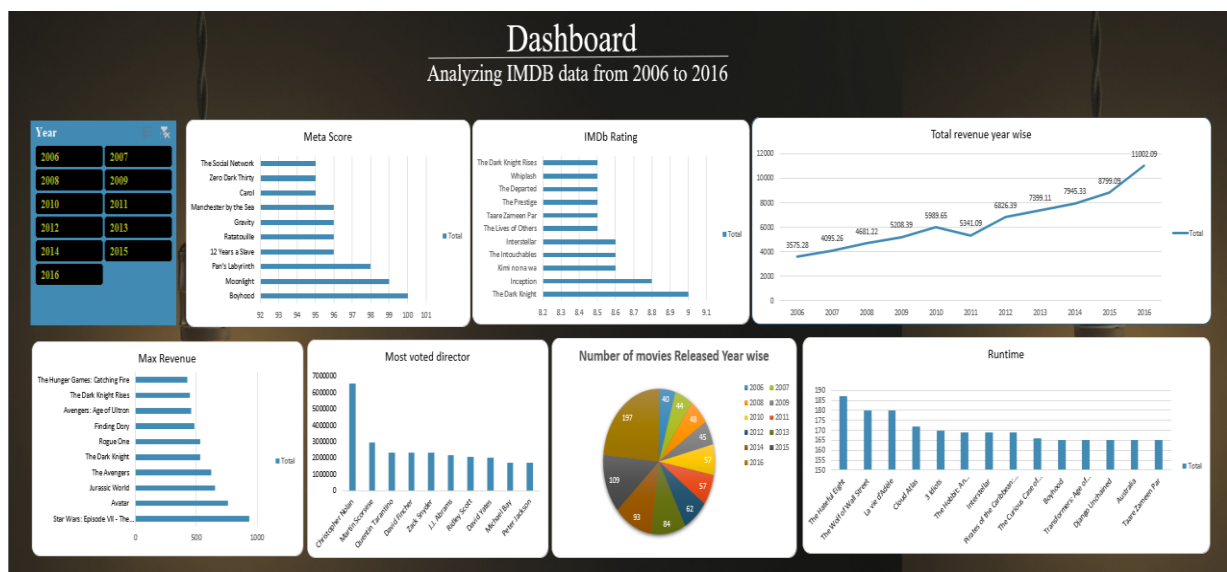


2. Analysis of Top 10 most voted directors from the dataset.

- The Most Voted Director in the dataset is Christopher Nolan. There is a speciality about him that he got the maximum number of votes for a film and also the total number of votes is also more for him.
- The second most voted Director is Martin Scorsese.



FINAL DASHBOARD:



BIBLIOGRAPHY:

Dataset source:

<https://www.kaggle.com/PromptCloudHQ/imdb-data>

Information about Data Management:

<https://www.blue-pencil.ca/what-is-data-management-and-why-it-is-important/>