

I. Project Summary

This project attempts to answer three questions using data from the TMDB movie database and the Oscars Awards dataset using Excel techniques to combine and calculate the value needed for them. The Excel file showing how the two datasets are combined for section II can be found [here](#) (The file is too large and need to download). The Excel file showing the steps to produce the answers for section III can be found [here](#).

The first question is, "Does winning or being nominated for Oscars affect the ratings of movies on TMDB?". And the data illustrates that Oscar-winning or nominated movies receive higher TMDB ratings. More Oscar wins mean higher TMDB scores. Movies with a certain number of Oscars have higher ratings than those with nominations only. This trend applies to various award categories. This may be due to the Oscar awards being determined by professionals and experts in the film industry with their own criteria, and the mainstream audience will also have a chance of having the same standards as the experts when rating movies.

The second question is, "Does having casts or crews who won or are nominated for the Oscars awards affect the score on TMDB." The results found that there is a connection between the presence of Oscar-winning cast and crew members and the audience's score on TMDB. The reasons may stem from the TMDB score being the overall quality of the films that the mainstream audience thinks, and the quality and performance of actors, actresses and crews significantly affect the overall quality of the movie.

The Last question is, "Do the Oscars' Best Picture awards reflect a film's financial success?". The data clearly indicates that Best Picture Awards do not reflect a film's financial success, as some movies were not nominated or won the awards that received a significant profit. Even though the first question suggests that they can reflect the overall audience belief of the film's quality; however, having only high quality does not mean most people will watch the movies in cinemas since most of the revenue is calculated from the box office. There are a lot of factors that affect a film's commercial success; for instance, the period when the movies were released in the cinemas (they should not be released at the same time as another famous sequel movie, during the financial crisis or any incidents that decrease a chance of people buying a ticket for the film). How successful are the marketing campaigns that the studio makes which generate awareness and encourage people to watch movies in cinemas?

II. Wrangling Details

A. Dataset details

1. Dataset 1 (Provided) - TMDB 5000 Movie Dataset

The data in this dataset is extracted from the movie database website: [TMDB](#), which is a website that provides information related to movies, TV shows, and actors. The data is in a CSV format and comprises two files: credits.csv and movies.csv, with 4 and 20 columns, respectively. The credit.csv has 19,212 blocks, and movies.csv has 96,120 blocks.

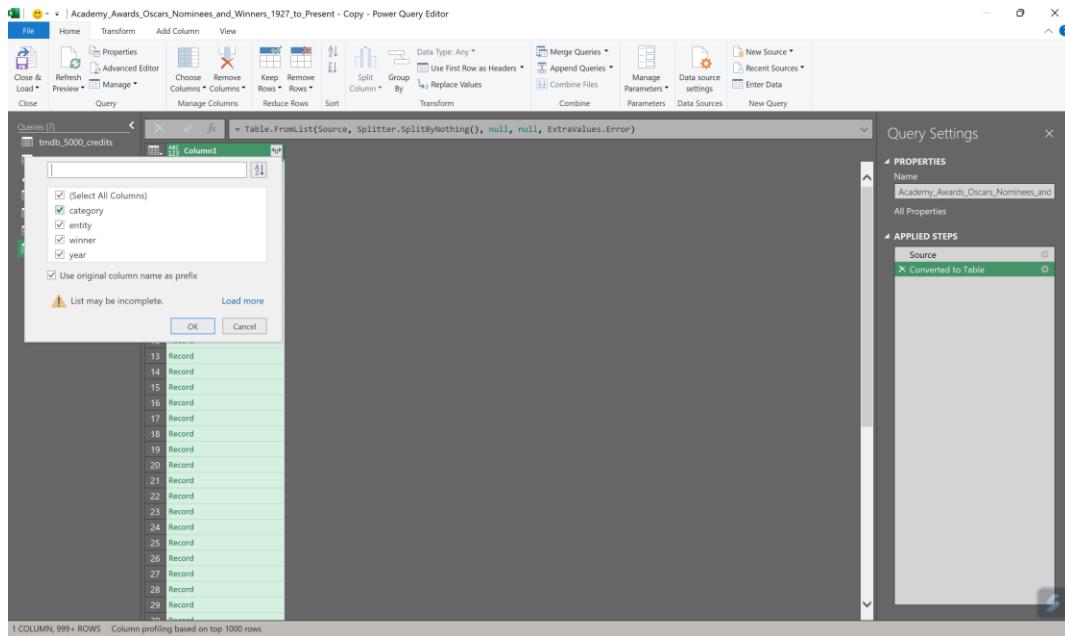
However, there are some problems with this data. Firstly, the data in 7 columns are in JSON format, and two datasets are separated, so a combination of both must be done first before combining with dataset 2, plus two of them do not have the same number of rows with movies.csv has three rows more so we can assume that every rows are not directly connected to each other for two datasets. The movie.csv's columns include budget, genres, homepage, **id**, keywords, original_language, original_title, overview, popularity, production_companies, production_countries, release_date, revenue, runtime, spoken_languages, status, tagline, title, vote_average and vote_count and credits.csv includes **movie_id**, title and cast, crew while they are connected with id and movie_id

2. [Dataset 2](#) - Academy Awards Oscars: Nominees and Winners 1927 to Present.

The data is extracted from the [Oscars](#) website about Academy Awards, honouring excellence in the film industry. The format of this data is JSON, and the data has a very clean format which is objects inside an array because there are a total of 11,058 objects; every object has exactly four properties which are an entity, category, winner and year and values which all of them are single values which includes string, numeric and Boolean.

B. Step to combine the datasets

1. Before combining data sets, the two CSV files are imported into Excel files using From Text/CSV and JSON data is imported using Power Query.



2. A VLOOKUP with an exact match is used to combine movie.csv and credit.csv with this syntax:
`=VLOOKUP([@id],tmdb_5000_credits[#All],3,TRUE)`. This syntax returns the cast for each movie by matching the movie ID. As mentioned previously, movies.csv has three more rows, so after matching, it is found that all of the data in these rows are entered incorrectly and are supposed to belong to the row above (half of row 3031 is empty, and the format of each block is suggesting). Therefore a fix in the CSV files is needed, which can be simply done by merging lines 3031 & 3032, 3367 & 3368, 3554 & 3555 using VSCode (see 2nd Picture)

The error example (it shows 3553 because first two have already been fixed)

3547 8000000, [{"id": 35, "name": "Comedy"}, {"id": 53, "name": "Thriller"}, {"id": 28, "name": "Action"}, {"id": 13173, {"id": 10183, "name": "Independent film"}, {"id": 18, {"id": 35, "name": "Drama"}, {"id": 35, "name": "Comedy"}, {"id": 10769, "name": "Romantic"}, {"id": 53, {"id": 88, "name": "Thriller"}, {"id": 80, {"id": 35, "name": "Crime"}, {"id": 70670, {"id": 13095, {"id": 35, "name": "Science Fiction"}, {"id": 35, {"id": 53, "name": "Comedy"}, {"id": 53, {"id": 18, {"id": 35, "name": "Drama"}}, {"id": 25248, {"id": 6075, "name": "Thriller"}]}]}], 3548 It's the fictional tale of a wayward 9th grader, Ralph (Adam Butcher), who is secretly living on his

3. Redo the second step into TMDB_combined and still use the VlookUp function to ensure that all data matches are correct. Every row is combined successfully, as there is no #N/A value.

4. In Oscar.JSON, switching the column 1 category and column 2 entity because the second column will be used to match with TMDB, and Vlookup can only match the leftmost column.
5. This step will combine TMDB_combined with Oscar.JSON with the movie title. According to the pivot table of data in the Oscars dataset, a movie receives one or more awards, and there are more than 100 distinct award types. Therefore the best way to combine them is to only merge those that received the best award(best Picture) and the total number of the award received except the best Picture.

The screenshot shows a Microsoft Excel spreadsheet with a PivotTable Fields pane open on the right. The PivotTable Fields pane lists fields: Column1.entity (checked), Column1.category (checked), Column1.winner (unchecked), Column1.year (unchecked), and Column1 (unchecked). Below the pane, the text "More Tables..." is visible. The main worksheet area displays a list of movies and their awards. A callout box highlights the entry "12 Monkeys (Column1.entity)" at row 25. The list includes:

- 10 MUSIC (Original Score)
 - MUSIC (Original Song)
- 102 Dalmatians COSTUME DESIGN
- 12 FOREIGN LANGUAGE FILM
- 12 Angry Men BEST MOTION PICTURE
 - DIRECTING
 - WRITING (Screenplay--based on material from another medium)
- 12 Monkeys COSTUME DESIGN
- 12 Years a Slave BEST PICTURE
 - COSTUME DESIGN
 - DIRECTING
 - FILM EDITING
 - PRODUCTION DESIGN
 - WRITING (Adapted Screenplay)

- a. Use pivot table to get a list of movies that were nominated and won the best picture

The screenshot shows two pivot tables side-by-side in Microsoft Excel. Both pivot tables have "Column1.category" and "Column1.winner" as columns. The first pivot table has "BEST_PICTURE" as the filter and shows rows for movies like "12 Years a Slave", "A Beautiful Mind", and "The Hurt Locker". The second pivot table has "BEST_PICTURE" as the filter and shows rows for movies like "All the Presidents' Men", "American Beauty", and "The Godfather". The bottom of the screen shows the Excel ribbon with tabs for "nines", "oscar_vlookup", "tmdb_5000_combined", "cast", "cast extracted", and others.

- b. Use IF, ISNA and Vlookup function to check if a movie in that row is in the pivot table for best Picture & true or not. If the lookup returns #N/A (not found), then the ISNA function will return true and proceed to the second IF function, but if it returns a value (found), then that cell is equal to 'won'. The second if function has the same meaning but for checking the best Picture & false pivot table (nominated) and then returns 'nominated' if

found. A blank indicates that a movie has not won or nominated Oscars Best Picture award.

c. Create another two pivot tables to count every time a movie receive a nomination or is won in any award simply by counting how many time they appear in the datasets for TRUE(won) and FALSE(nominated) separately.

The screenshot shows a Microsoft Excel spreadsheet titled "Project Combined Data autoRecov(AutoRecovered) most r...". The ribbon menu includes File, Home, Insert, Page Layout, Formulas, Data, Review, View, and Help. The Home tab is selected, displaying various font and style tools. A yellow status bar at the bottom indicates "SECURITY WARNING External Data Connections have been disabled" and "Enable Content".

The main content area displays two pivot tables side-by-side. The left pivot table has columns for "Column1.entity" (values like "winner", "TRUE", and "Row Labels" which list movie titles). The right pivot table also has columns for "Column1.entity" (values like "winner", "FALSE", and "Row Labels" which list movie titles). Both pivot tables include a "Count of Column1.entity" column.

	Column1.entity	Count of Column1.entity		Column1.entity	Count of Column1.entity
1					
2					
3	Column1.winner	TRUE		Column1.winner	FALSE
4					
5	Row Labels	Count of Column1.entity		Row Labels	Count of Column1.entity
6	12 Years a Slave	2		\$1,000 a Minute	1
7	20 Feet from Stardom	1		(A) Torzija [(A) Torsion	1
8	20,000 Leagues under the Sea	2		...And Justice for All	1
9	2001: A Space Odyssey	1		10	2
10	8 Mile	1		102 Dalmatians	1
11	A Beautiful Mind	1		12 Angry Men	1
12	A Boy and His Dog	3		12 Monkeys	3
13	A Chance to Live	1		12 Years a Slave	4
14	A Christmas Carol	1		12:01 PM	1
15	A Close Shave	1		127 Hours	5
16	A Double Life	1		13 Hours: The Secret S	1
17	A Fantastic Woman	1		13th	1
18	A Farewell to Arms	2		140 Days under the W	1
19	A Funny Thing Happened on the Way to the Forum	1		1776	1
20	A Girl in the River: The Price of Forgiveness	1		1949	1
21	A Good Year	1			

d. Use Vlookup function to merge these pivot tables with the tmdb_5000_combined

The screenshot shows a Microsoft Excel spreadsheet with the following details:

- Formula Bar:** =VLOOKUP([@title],Academy Awards Oscars Nominees 'IP5:Q4750,2,FALSE)
- Table Headers:** L: release_date, M: revenue, N: runtime, O: speakers, P: status, Q: title, R: vote_aw, S: vote_cd, T: cast, U: crew, V: best_picture, W: other_awards_nominated, X: other_awards_won, Y: cast_nominated.
- Data:** The table contains over 30 rows of movie information, including titles like 'Enter the Avatar', 'A Plan No Spectre', 'The Legend of Dark Ki', etc., along with their respective release dates, revenues, runtimes, speakers, statuses, titles, votes, casts, crews, and various award counts.

e. Modify the syntax so that they do not count the Best Picture by minus one in other awards nominated if the Best Picture cell is nominated and the same for won.

The screenshot shows a Microsoft Excel spreadsheet with the following details:

- Formula Bar:** =IF([@best_picture] = "nominated", VLOOKUP([@title], Academy Awards Oscars Nominees 'IP5:Q4750,2,FALSE), -1), VLOOKUP([@title], Academy Awards Oscars Nominees 'IP5:Q4750,2,FALSE))
- Table Headers:** I: popularity, J: product, K: product, L: release_date, M: revenue, N: runtime, O: speakers, P: status, Q: title, R: vote_aw, S: vote_cd, T: cast, U: crew, V: best_picture, W: other_awards_nominated, X: other_awards_won, Y: cast_nominated.
- Data:** The table contains over 30 rows of movie information, similar to the previous table, but with modified logic for handling the 'best_picture' column in the VLOOKUP formulas.

f. Modify the syntax so that it returns 0 instead of #N/A by using the IFERROR function

The screenshot shows a Microsoft Excel spreadsheet titled "Project Combined Data autoRecov(AutoRecovered) most recent". The table contains the following columns:

- Y: popularity, production, release_date, revenue, runtime, status, spoken, title, vote_av, vote_ct, cast, crew
- Z: best picture, other awards nominated, other awards won, cast_nom

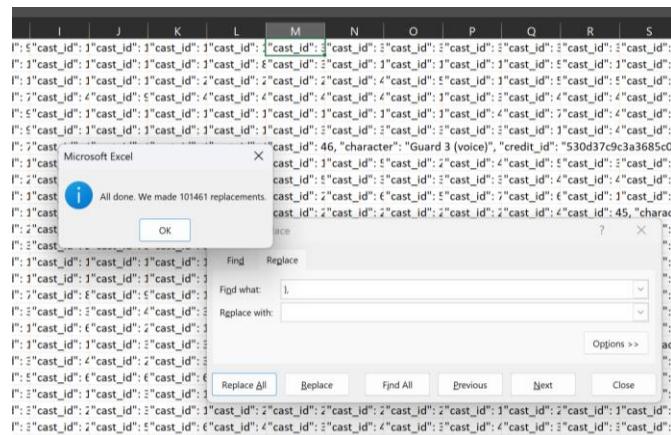
A formula is present in cell Y2: =IFERROR(IF([@best picture] = "won", VLOOKUP([@title], 'Academy Awards Oscars Nominees'!\$M\$2:\$N\$64, 2, FALSE) - 1), (VLOOKUP([@title], 'Academy Awards Oscars Nominees'!\$M\$2:\$N\$64, 2, FALSE)) * 0).

6. Assuming there will be no increase in the number of rows (4804 rows, including headers),

importing credit.csv to MongoDB using this query: `mongoimport --type csv --headerline --db`

`project --collection castcrew --file tmdb_5000_credits.csv`, for separating value in array of object purposes. However, MongoDB does not detect arrays of objects because there are double quotation marks around them. Therefore it is not possible to use Mongodb to separate each object.

7. Therefore, we have to separate each object by using { signs so that every cast is separated into single columns and then use find and replace to delete all of }, and }) signs so that every block have the same format, which is "cast_id": value, "character": value, "credit_id": value, "gender": value, "id": value, "name": value, "order": value



- The next step is to create a value extractor from the object by using the data validation function and vlookup to answer the question and further exploration; the steps are

a. Create an id and the total number of the cast for each movie

b. Use Vlookup to get an object based on the given criteria, which are movie_id and order of object. The movie_id is validated using a list of id, and order is validated using the total object, so when an unidentified movie id is entered, the error message will appear

The screenshot shows a Microsoft Excel interface with the following details:

- File ribbon:** AutoSave, Home, Insert, Page Layout, Formulas, Data, Review, View, Help.
- Data ribbon tab:** Recent Sources, Existing Connections, Refresh, Properties, Sort & Filter, Advanced, Text to Columns, Data Validation, Manage Data Model, Forecast, Analysis, Relationships, Remove Duplicates, Consolidate, Flash Fill, Group, Ungroup, Subtotal, Outline.
- Formula Bar:** B6 contains the formula =VLOOKUP(B2, cast!1:4804, 'cast extracted'!B4+3, FALSE).
- Table Data:** A table with columns A through F. Row 2: movie_id (19995), total object (83). Row 3: order (82). Row 4: object ({"cast_id": 102, "character": "Ambient Room Tech / Troupe", "credit_id": "52fe48019251416c750acb6f", "gender": 1, "id": 42286, "name": "Julene Renee", "order": 82}). Row 5: property ({"name": "Julene Renee"}). Row 6: raw value ({"name": "Julene Renee"}). Row 7: trim space outside text ({"name": "Julene Renee"}). Row 8: value without ("Julene Renee").

The screenshot shows an Excel spreadsheet titled "Project Combined Data". The "Data" tab is selected. In row 12, column B, there is a formula: =IF(BB=H18,(RIGHT(B6,LEN(B6)-O8+1)),MID(B6,E8,(L8-E8+1))). The cell B12 contains the value "10989". A validation error dialog box is open, stating: "This value doesn't match the data validation restrictions defined for this cell." The dialog has "Retry", "Cancel", and "Help" buttons.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P
1																
2	movie_id	99999		total object	49											
3																
4	order	1														
5																
6	object	"cast_id": 2, "character": "Ron Weasley", "credit_id": "52fe4273c3a36847f801fa6f", "gender": 2, "id": 10989, "name": "Rupert Grint", "order": 1														
7																
8	property	"id"		value starts from	Microsoft Excel											
9																
10	raw value	10989														
11	trim space outside text	10989														
12	value without "	10989														
13																
14																
15																
16																

- c. Next is that the raw values are extracted based on the given property, validated by the cast property list below with the formula shown. Then the extra space before and after the value is removed, and the quotation marks are removed if there are any.

The screenshot shows an Excel spreadsheet titled "Project Combined Data". The "Data" tab is selected. In row 18, column B, there is a formula: =IF(BB=H18,(RIGHT(B6,LEN(B6)-O8+1)),MID(B6,E8,(L8-E8+1))). The cell B18 contains the value "cast property". The cell B19 contains the formula: =IF(BB=H18,(RIGHT(B6,LEN(B6)-O8+1)),MID(B6,E8,(L8-E8+1))). The cell B19 also contains the value "cast_id". The cell C19 contains the formula: =IF(BB=H18,(RIGHT(B6,LEN(B6)-O8+1)),MID(B6,E8,(L8-E8+1))). The cell C19 also contains the value "character". The cell D19 contains the formula: =IF(BB=H18,(RIGHT(B6,LEN(B6)-O8+1)),MID(B6,E8,(L8-E8+1))). The cell D19 also contains the value "credit_id". The cell E19 contains the formula: =IF(BB=H18,(RIGHT(B6,LEN(B6)-O8+1)),MID(B6,E8,(L8-E8+1))). The cell E19 also contains the value "gender". The cell F19 contains the formula: =IF(BB=H18,(RIGHT(B6,LEN(B6)-O8+1)),MID(B6,E8,(L8-E8+1))). Thecell F19 also contains the value "id". The cell G19 contains the formula: =IF(BB=H18,(RIGHT(B6,LEN(B6)-O8+1)),MID(B6,E8,(L8-E8+1))). The cell G19 also contains the value "name". The cell H19 contains the formula: =IF(BB=H18,(RIGHT(B6,LEN(B6)-O8+1)),MID(B6,E8,(L8-E8+1))). The cell H19 also contains the value "order".

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P
1																
2	movie_id	58		total object	33											
3																
4	order	1														
5																
6	object	"cast_id": 3, "character": "Will Turner", "credit_id": "52fe4211c3a36847f8001823", "gender": 2, "id": 114, "name": "Orlando Bloom", "order": 1														
7																
8	property	"credit_id"		value starts from	55			next property: "gender"								
9																
10	raw value	"52fe4211c3a36847f8001823"														
11	trim space outside text	"52fe4211c3a36847f8001823"														
12	value without "	52fe4211c3a36847f8001823														
13																
14																
15																
16																
17																
18	cast property	"cast_id"	"character"	"credit_id"	"gender"	"id"	"name"	"order"								
19																

9. Repeat the 7th step for the crew object so the format is consistent, which is credit_id: "value", "department": "value", "gender": "value", "id": "value", "job": "value", "name": "value". Then repeat the 8th step.

Screenshot of Microsoft Excel showing a table with movie credits. Row 8 highlights the formula in cell C2:

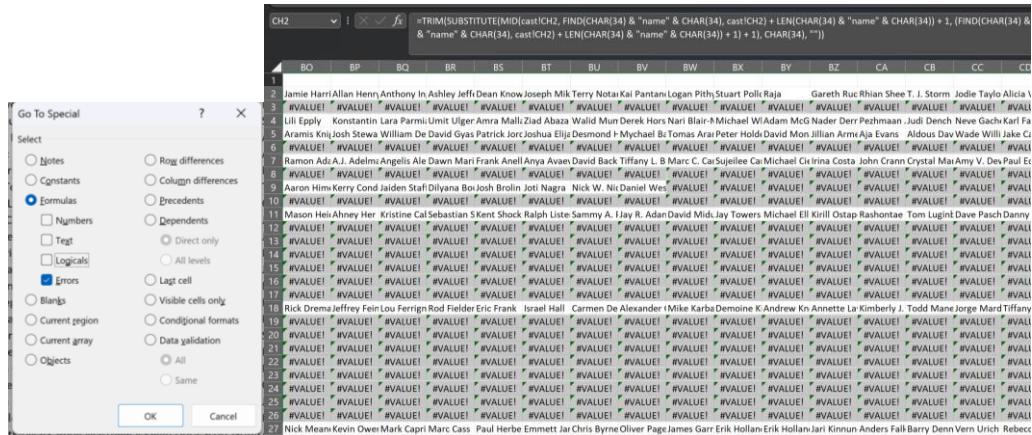
```
=TRIM(MID(cast!C2, FIND(CHAR(34) & "name" & CHAR(34), cast!C2) + LEN(CHAR(34) & "name" & CHAR(34)) + 1, (FIND(CHAR(34) & "order" & CHAR(34), cast!C2) - 3) - (FIND(CHAR(34) & "name" & CHAR(34), cast!C2) + LEN(CHAR(34) & "name" & CHAR(34)) + 1) + 1), CHAR(34), "")
```

10. Nesting all the formulas done in the 8th step to create a new sheet specifically for storing all the cast names for each movie to prepare to match with the Oscars dataset

Screenshot of Microsoft Excel showing a table with movie credits. Row 8 highlights the formula in cell C2:

```
=TRIM(MID(cast!C2, FIND(CHAR(34) & "name" & CHAR(34), cast!C2) + LEN(CHAR(34) & "name" & CHAR(34)) + 1, (FIND(CHAR(34) & "order" & CHAR(34), cast!C2) - 3) - (FIND(CHAR(34) & "name" & CHAR(34), cast!C2) + LEN(CHAR(34) & "name" & CHAR(34)) + 1) + 1), CHAR(34), "")
```

11. Delete all of the #VALUE! Result for the formula because there is no object in that cell, using the Go To Special function to select all of the error results and delete them



12. The next steps will start to combine the Oscars.JSON with all of the cast names.

Unfortunately, the Oscars dataset does not contain information about what movies people were nominated for or won the Oscars belong to. So we need to use the year column to assume that if cast A won Oscars in 2000, the movie released in 2000 that has cast A will be the movie that has cast who won the Oscars. Therefore, for simple matching, we combine the name and year together using concatenate in one cell for both Oscars.JSON and the cast sheet.

The screenshot shows two Microsoft Excel sheets side-by-side. The left sheet, 'imdb_3000_combined', has a formula in cell C2: =CONCATENATE([@Column1],[@year]). The right sheet, 'cast', lists cast members and their years. Both sheets have security warnings at the top.

Category	Column1	Year
ACTION	FALSE	1997 Robert Redford
ACTOR	TRUE	1997 Ed Harris
ADVISOR	FALSE	1997 Edward Norton
ACTRESS	TRUE	1997 Jennifer Lopez
ARTIST	FALSE	1997 Meryl Streep
ART DIRECTION	FALSE	1997 Noomi Rapace
ART DIRECTOR	FALSE	1997 William Campbell
ART DIRECTOR	FALSE	1997 William Campbell
CINEMATOGRAPHY	FALSE	1997 George Burns
CINEMATOGRAPHER	TRUE	1997 John Goodman
CINEMATOGRAPHY	TRUE	1997 Karl Malden
DIRECTOR (Cinematography)	TRUE	1997 Karl Malden
DIRECTOR (Cinematography)	FALSE	1997 Ted White
DIRECTOR (Cinematography)	TRUE	1997 Frank Darabont
DIRECTOR (Cinematography)	FALSE	1997 Paul Haggis
DIRECTOR (Drama/Picture)	FALSE	1997 King Vidor
DIRECTOR (Drama/Picture)	TRUE	1998 Michael Mann
ENGINEERING EFFECTS	TRUE	1998 Michael Mann
ENGINEERING EFFECTS	TRUE	1998 Roy Kinnear
OUTSTANDING PICTURE	FALSE	1998 The Godfather
OUTSTANDING PICTURE	FALSE	1997 The Godfather
OUTSTANDING PICTURE	TRUE	1997 Tom Hanks
UNIQUE AND ARTISTIC PICTURE	TRUE	1997 Amélie
UNIQUE AND ARTISTIC PICTURE	TRUE	1997 Paramount/Mayer
UNIQUE AND ARTISTIC PICTURE	TRUE	1997 Paramount/Famenni Lally

13. Create a new sheet specifically for matching, which contains some data merged together

A screenshot of Microsoft Excel showing a table of Academy Award nominees from 1927 to 1937. The columns include name, year, and category_won. The table shows various awards like Best Actor, Best Supporting Actress, and Best Picture.

14. Use Vlookup with an exact match to find if the cast+ year matches with the data in Oscars.json. Then use go to special to delete all #N/A

A screenshot of Microsoft Excel showing two tables side-by-side. The left table is 'cast_nomineewon' and the right table is 'cast_nominated'. Both tables have columns for id, cast_name, and category_won. The right table includes additional columns for cast_name+year and category_won+year.

15. Count the total number of cast nominated and won for each movie by using countif, if it ends with a false count for nominated and ends with a true count for won

A screenshot of Microsoft Excel showing a table with columns for id, cast_nomineewon, and cast_won. The cast_won column contains counts of 0 or 1. A formula =COUNTIF(D3:XF93, "FALSE") is shown in cell B3, which counts the number of 0s in the range D3:XF93. Another formula =COUNTIF(D3:XF93, "TRUE") is shown in cell C3, which counts the number of 1s in the range D3:XF93.

16. Replicating since step 10 for the crew. Firstly, extracting the name from the array of the object of the crew by nesting function, done in step 9. Then, concatenating crew name and year for matching. Then, use Vlookup to match the category and if win for crew

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	
1	id	total	object																	
2	19995	153	Stephen E.	Rick Carter	Christopher	Christophe	Mali Finn	James Horr	James Cam	James Cam	James Cam	James Cam	James Cam	Andrew M.	Jill Brooks	Margery S.	Kevin Ishio	Dick Bern	Shannon M.	
3	285	32	Dariusz W.	Gore Verbinski	Bruckheimer	Ted Elliott	Terry Rossi	Stephene	E. Craig	Woo	Hans Zim	Mike Stens	Eric McLeo	Chad Omar	Peter Kohn	Pat Sandst	Denise Cha	Rich Heinri	John Dext	Priscilla Jol
4	206647	155	Thomas Ne	Manda Anna	Pinnic John	Logan John	Dennis Gas	Ian Fleming	Lee Smith	Bill Berne	Craig Jany	Temin Debbie	McBarbara Bri	Roberto M.	Peter Clark	Christophe Per	Halber Nicole	Schr Nic	Michael Us	Benjamin J.Jm Palmer Karen Blynn
5	49026	217	Hans Zimmer	Charles Ro	Christopher	Christophe	Christopher Jonathan	N Emma	Thoi Wally	Pifst John	Papiss David S.	David G.	Sam Kane	Lee Smith	Michael Us	Benjamin J.Jm Palmer Karen Blynn	Linc	Michael Us	Benjamin J.Jm Palmer Karen Blynn	
6	49529	132	Andrew St.	Andrew Stanton	John Lasset	Colin Wilso	Gail Steven T.	Johns Eric	Zumbr Michael	Sil Juan	Pearl Gooch	Paki Ed	Catnall Steve	Jobs Bill	Corso Robert	Wo Nathan	Crc Michael Gi	Daniel Min Ste	Michael Gi	Daniel Min Ste
7	559	56	Francine M	Joseph M.	Sam Raini	Ramni Stan	Lee Stan	Lee Stee	Deit Bob	Muraw Laura	Ziski Avi	Avi Arad	Spikes Aschak	Ach Alvin	Kevin Feige	Bill Pope	Grant Curti	Ivan Raimi	Christophe Kat	
8	38757	79	John Lasset	Jacob Grin	Wilhelm Grin	Haks Douglas	Ro Alan	Menk Alen	Menk	Fogeln Roy	Conil	Mark Kenn	Dan Coop	Tim Mert	Kevin Klies	Byron How	Dean A.	Zu David E.	Fil Nathan Gre Jamie Spar	
9	99861	74	Danny Elf	Christopher	Christophe Sarah	Finn Brian	Yster Stan	Lee Stan	Lee Kevin	Feige Richard	Ro Jerry	Bruck Joss	Whealloss Hedi	Alexandra	Feige For	Denis	David Kirby	Jon Favre	Charles W.	
10	767	27	Bruno Delph	Stephanie I	Stuart Craij	J.K. Rowlin	Steve Kov	David Heyw	Andrew Ac Gary	Tomk Jany	Fiona Weir	Alastair Bu	David Yate	Nicholas H.	Mark Day	David Barr	Tim Lewis Lionel	Wig Molly	Hug Tine	
11	209112	111	Hans Zimmer	Charles Ro	Christopher Emma	Thoi David S.	Gc David S.	Gc Frank Mille	Frank Mille	Lora Kenne Bob	Kane	Patrick Tat	Beat Frutig	Michael Pei	Debbie McBarbara	Barbara Bri	Roberto M.	Michael Us	Benjamin J.Jm Palmer Karen Blynn	
12	1452	42	Roger Mus	Peter Diane	MacGilbert Ad	Brayn Sing	Bryan Sing	Bryan Sing	Bryan Sing	Bryan Sing	Bryan Sing	Bryan Sing	Bryan Sing	Bryan Sing	Bryan Sing	Bryan Sing	Bryan Sing	Bryan Sing	Bryan Sing	
13	10764	17	Paul Hagg	Dennis Gas	Ian Fleming Louise	Frog David	Ari Richard	Pei Debbie	McBarbara Bri	Anthony W.	Mar Gregg	Philippa Bo	Andrew	Joe Jerry Siegel	Joe Shuster	Chris Lee Thomas Tu William	Fai Elliot Grabs	Scot	Scot	
14	58	23	Dariusz W.	Gore Verbi	Gore Verbi Jerry	Bruck Ted Elliott	Elliot	Terry Rossi	Terry Rossi Stephen	E. Craig	Woo	Hans Zim	Bruce Henc	Mike Stens	Eric McLeo	Chad Omar	Peter Kohn	Pat Sandst	Denise Cha	Rich Heinri
15	57201	35	Gore Verbi	Gore Verbi Jerry	Bruck Ted Elliott	Elliot	Terry Rossi	Terry Rossi Johnny	De Craig	Woo	Hans Zim	Mike Stens	Eric McLeo	Chad Oma	Denise Cha	Christopher Christophe Gary Ryd	St Cheryl	Cara Pen	Cara Pen	
16	49521	48	Hans Zimmer	Charles Ro	Lloyd Phillip	Christopher	Christophe Emma	Thor Alex McD	McDo G.	Gc David S.	Gc Lora	Kenne Jon	Peters Michael W.	Anne Bulk	David Br	James Ach	Kim Sinclair	Zack Synde	Kristy Carls Jerr	
17	2454	83	Liz Mullane	Gall Steven	David Mink	Matthew S Mark	Johns Andrew	Ad Andrew Ad	Ac D.C. Lewis	Philip Stev	Sim Evan-J	Glougs	Drugs	Mooy	Cooper K.	Hoden Roger	Ford Kerrie Bro	Browls	Isus Muser Christopher Ste	
18	24428	147	Alan Silvest	Christopher	Christophe Seamus	Mc Sarah	Seamus	Mc Sarah	Seamus	Mc Sarah	Seamus	Mc Sarah	Seamus	Mc Sarah	Seamus	Mc Sarah	Seamus	Mc Sarah	Seamus	
19	1865	39	Dariusz W.	Jerry Bruck	Hans Zim	Charles Ro	Christopher	Christopher Penny	Rey Bob	David Braden	David Braden	David Braden	David Braden	David Braden	David Braden	David Braden	David Braden	David Braden	David Braden	
20	41154	21	Stephanie Spa	Danny Elf	Elfin Ellen Chen	Walter F.	P.B. Welch	Barry Son	Laurell Cun Mary	E. V.	Wayne Wa	Laurell Cun Mary	E. V.	Wayne Wa	Laurell Cun Mary	E. V.	Wayne Wa	Laurell Cun Mary	E. V.	
21	122917	127	Howard Sh.	Christopher Peter Jacks	Peter Jacks	J.R.R. Tolkin	Fran Walsh	Philippe Bo	Andrew Le	Alan Lee	John Howe	Richard Ta Victoria	Bo John Hubb	Iz Mullan	Dan Henna	Leanne Joba	Leanne Joba	Leanne Joba	Leanne Joba	
22	1930	106	Francine Maisler	John Schw	Pietro Scall James	Horr Kym	Brett Star	Lee Stan	Lee Steve	Ditk Fran Walsh	Philippe Bo	Andrew Le	Le Alan Lee	Leanne Joba	Leanne Joba	Leanne Joba	Leanne Joba	Leanne Joba	Leanne Joba	
23	20662	18	Brian Grazer	Jina Jay	Ridley Scott	Rey Bob	David Cro John	Mathi	Yanties Arthur	Max Pietro Scall	James	Horr Kym	Bole David	Lei David	Lei David	Lei David	Lei David	Lei David	Lei David	
24	57158	108	Howard Sh.	Peter Jacks	Peter Jacks	J.R.R. Tolkin	Fran Walsh	Philippe Bo	Andrew Le	Le John Howe	Richard Ta Victoria	Bo John Hubb	Iz Mullan	Dan Henna	Jabz Olse	Brent Burg	David Farm	Car	Car	
25	2268	32	Anna Pinnic	Mark Orde	Michael Li	Dennis Gas	Alexandre	I Chris	Weitz Chris	Weitz Paul	Robert Sha	Anna V.	Co Richard L.	Jandy Niche	Fiona Weir	Ruth Myer	Toby Emmr	Philip Pulin	Bill Carraro Deborah Fc Her	
26	254	38	James Newton	Howard Peter	Peter Jackson	2005	Peter Jackson	Peter Jackson	Peter Jackson	Peter Jackson	Peter Jackson	Peter Jackson	Peter Jackson	Peter Jackson	Peter Jackson	Peter Jackson	Peter Jackson	Peter Jackson	Peter Jackson	

A	B	C	D	E	F	G	H	I	J	K		
1	id											
2	19995	Stephen E.	Rick Carter	2009	Christopher	Boyes	2009	Christopher	Boyes	2009		
3	285	Dariusz W.	Gore Verbinski	2007	Jerry Bruckheimer	2007	Jerry Bruck	2007	Terry Rossi	2007		
4	206647	Thomas Newma	15	Sam Mendes	2015	Anna Pinnock	2015	John Logan	2015	John Logan	2015	
5	49026	Hans Zimmer	2012	Charles Roven	2012	Christopher Nolan	2012	Christopher Nolan	2012	Christopher Nolan	2012	
6	49529	Andrew Stanton	2012	Andrew Stanton	2012	John Lasseter	2012	John Lasseter	2012	John Lasseter	2012	
7	559	Francine Maisler	2007	Joseph M. Caracilo	2007	Sam Raimi	2007	Gail Stevens	2012	Gail Stevens	2012	
8	38757	John Lasseter	2010	Jacob Grimm	2010	Wilhelm Grimm	2010	John Kahrns	2010	John Kahrns	2010	
9	99861	Danny Elfman	2015	Christopher Boyes	2015	Christopher Boyes	2015	Sarah Finn	2015	Sarah Finn	2015	
10	767	Bruno Delbonnel	2009	Stephanie McMillan	2009	Stuart Craig	2009	J.K. Rowling	2009	J.K. Rowling	2009	
11	209112	Hans Zimmer	2016	Charles Roven	2016	Christopher Nolan	2016	Christopher Nolan	2016	Christopher Nolan	2016	
12	1452	Roger Mussen	2006	Jon Peters	2006	Diane Macke	2006	Gilbert Adler	2006	Bryan Singer	2006	
13	10764	Paul Hagg	2006	Dennis Gassner	2008	Ian Fleming	2008	Louise Frogley	2006	Bryan Singer	2006	
14	58	Dariusz W.	Gore Verbinski	2006	Jerry Bruckheimer	2006	Jerry Bruckheimer	2006	Ted Elliott	2006	Ted Elliott	2006
15	57201	Gore Verbinski	2013	Gore Verbinski	2013	Jerry Bruckheimer	2013	Ted Elliott	2013	Ted Elliott	2013	
16	49521	Hans Zimmer	2013	Charles Roven	2013	Lloyd Phillips	2013	David Minkowski	2008	David Minkowski	2008	
17	2454	Liz Mullane	2006	Gail Stevens	2008	Christopher Boyes	2012	Tell Elliott	2011	Tell Elliott	2011	
18	24428	Alan Silvestri	2012	Christopher Boyes	2012	Christopher Boyes	2012	Christopher Boyes	2012	Christopher Boyes	2012	
19	1865	Dariusz W.	Jerry Bruckheimer	2011	Jerry Bruckheimer	2011	Tell Elliott	2011	Terry Rossi	2011	Terry Rossi	2011
20	41154	Steven Spielberg	2012	Danny Elfman	2012	Ellen Chenoweth	2012	Walter F. Parkes	2012	Bo Welch	2012	
21	122917	Howard Shore	2014	Christopher Boyes	2014	Peter Jackson	2014	Peter Jackson	2014	Peter Jackson	2014	
22	1930	Francine Maisler	2012	John Schwartzman	2012	Pietro Scalia	2012	James Horner	2012	Kym Barrett	2012	
23	20662	Brian Grazer	2010	Jina Jay	2010	Ridley Scott	2010	Ridley Scott	2010	Russell Crowe	2010	
24	57158	Howard Shore	2013	Peter Jackson	2013	Peter Jackson	2013	Peter Jackson	2013	Peter Jackson	2013	
25	2268	Anna Pinnock	2007	Mark Ordlesky	2007	Michael Lynne	2007	Dennis Gassner	2007	Alexandre Desplat	2007	
26	254	James Newton	Howard Peter	Peter Jackson	2005	Peter Jackson	2005	Peter Jackson	2005	Fran Walsh	2005	

The screenshot shows an Excel spreadsheet with the following details:

- Cell C3916:** Contains the formula `=VLOOKUP("crew_name+year",D3916:Table4[#All],2,FALSE)`.
- Data Range:** The range D3916:Table4[#All] is highlighted.
- Column Headers:** The first few columns have headers like "A", "B", "C", "D", "E", "F", "G", "H", "I", "J", "K", "L", "M", "N", "O", "P", "Q", "R", "S", "T".
- Content:** The data includes rows such as "59961 1 0 ACTOR IN A LEADING ROLE FALSE", "22954 1 0 ACTOR IN A LEADING ROLE FALSE", and "58431 1 0 ACTRESS IN A LEADING ROLE FALSE".
- Notes:** A yellow warning bar at the top says "SECURITY WARNING External Data Connections have been disabled".

17. However, most of the categories are actors and actresses, which may suggest that some casts are also part of the crew; therefore, deletion of these cast categories is required to not skew the result by modifying countif function to not contain actress or actor

The screenshot shows an Excel spreadsheet with the following details:

- Cell C3916:** Contains the formula `=COUNTIFS(D3916:XFD3916,"*TRUE",D3916:XFD3916,"<>ACTOR*",D3916:XFD3916,"<>ACTRESS")`.
- Data Range:** The range D3916:XFD3916 is highlighted.
- Column Headers:** The first few columns have headers like "A", "B", "C", "D", "E", "F", "G", "H", "I", "J", "K", "L", "M", "N", "O", "P", "Q", "R", "S", "T", "U", "V", "W", "X".
- Content:** The data includes rows such as "IRVING G. THALBERG MEMORIAL AWARD TRUE", "IRVING G. THALBERG MEMORIAL AWARD TRUE", and "IRVING G. THALBERG MEMORIAL AWARD TRUE".
- Notes:** A yellow warning bar at the top says "SECURITY WARNING External Data Connections have been disabled".

File Home Insert Page Layout Formulas Data Review View Help

Font Alignment Number Styles Cells Editing

Comments Share

C3916 : fx =COUNTIFS(D3916:XFD3916,"*TRUE",D3916:XFD3916,">ACTOR*",D3916:XFD3916,">ACTRESS*")

1 crew_nominal < crew_won
907 1 0
1389 1 0
4805 4805

MUSIC (Original Musical or

Costume Design FALSE

18. Then merge the total number of award cast and crew received with the combined sheet.

File Home Insert Page Layout Formulas Data Review View Help

Font Alignment Number Styles Cells Editing

Comments Share

AE18 : fx

Q	R	S	T	U	V	W	X	Y	Z	AA	AB	AC	AD	AE	AF	AG	AH	AI	AJ	AK	A
1	Tagline	Title	Vote_Aw	Vote_Crew	Cast	Crew	best picture	other aw	other awards	cast nominated	cast won	crew nominated	crew won								
2	Enter the V Avatar	7.2	11800	"cast_ic([{"credit_id": "52fe48009251416c7"}, {"nominated": 5}], [{"credit_id": "52fe4232ca36847f8"}])"	0	0	0	0	0	0	0	0	0								
3	At the end Pirates of t	6.9	4500	"cast_ic([{"credit_id": "52fe4232ca36847f8"}])"	0	0	0	0	0	0	0	0	0								
4	A Pan No Speach	6.3	4466	"cast_ic([{"credit_id": "54805967ca36847f8"}])"	0	0	0	0	0	0	0	0	0								
5	The Dark Knight	7.0	9300	"cast_ic([{"credit_id": "52fe4232ca36847f8"}])"	0	0	0	0	0	0	0	0	0								
6	Lost in our John Carte	6.1	2124	"cast_ic([{"credit_id": "52fe4232ca36847f8"}])"	0	0	0	0	0	0	0	0	0								
7	The battle Spider-Mai	5.9	3576	"cast_ic([{"credit_id": "52fe4252ca36847f8"}])"	0	0	0	0	0	0	0	0	0								
8	They're takTangled	7.4	3330	"cast_ic([{"credit_id": "52fe46d69251416c9"}])"	0	0	0	0	0	0	0	0	0								
9	A New Age Avengers: i	7.3	6767	"cast_ic([{"credit_id": "55d517dc4a36847f7"}])"	0	0	0	0	1	0	0	0	0								
10	Dans Secret Iron Pott	7.4	5293	"cast_ic([{"credit_id": "52fe42723ca36847f8"}])"	0	0	0	0	0	0	0	0	0								
11	Justice or r Batman v S	5.7	7004	"cast_ic([{"credit_id": "553b123092514135d"}])"	0	0	0	0	1	0	0	0	0								
12	Superman	5.4	1402	"cast_ic([{"credit_id": "553b3fe69251416c7"}])"	0	0	0	0	0	0	0	0	0								
13	For love, fc Quantent	6.1	2965	"cast_ic([{"credit_id": "52fe46d69251416c7"}])"	0	0	0	0	0	0	0	0	0								
14	It's a Star War of t	7.0	5524	"cast_ic([{"credit_id": "52fe4232ca36847f8"}])"	0	0	0	0	0	0	0	0	0								
15	Never Take The Lone R	5.9	2311	"cast_ic([{"credit_id": "52fe4232ca36847f8"}])"	0	0	0	0	0	0	0	0	0								
16	You will be Man of Ste	6.5	6359	"cast_ic([{"credit_id": "52fe47092ca36847f8"}])"	0	0	0	0	1	0	0	0	0								
17	Hope has a The Chroni	6.3	1630	"cast_ic([{"credit_id": "55a239e6925141297"}])"	0	0	0	0	0	0	0	0	0								
18	Some assse The Avengy	7.4	11776	"cast_ic([{"credit_id": "52fe4495ca368484"}])"	0	0	0	0	0	0	0	0	0								
19	Live FirePre Pirates of t	6.4	4948	"cast_ic([{"credit_id": "566b4f54ca3683f5e"}])"	0	0	0	0	0	0	0	0	0								
20	They are b Men in Bla	6.2	4160	"cast_ic([{"credit_id": "52fe455b7ca36847f8"}])"	0	0	0	0	1	0	0	0	0								
21	Witness th The Hobbit	7.1	4760	"cast_ic([{"credit_id": "54805967ca36847f8"}])"	0	0	0	1	0	0	0	0	0								
22	The untold The Amazin	6.5	6586	"cast_ic([{"credit_id": "539543dca3684841"}])"	0	0	0	0	1	0	0	0	0								
23	Beyond the Hobbit	6.0	1393	"cast_ic([{"credit_id": "52fe4232ca36847f8"}])"	0	0	0	0	0	0	0	0	0								
24	Beyond the Hobbit	7.6	4524	"cast_ic([{"credit_id": "52fe4232ca36847f8"}])"	0	0	0	0	0	0	1	0	0								
25	There are t The Golder	5.8	1303	"cast_ic([{"credit_id": "52fe4248ca36847f8"}])"	0	0	0	0	0	0	0	0	0								
26	The eighth King Kong	6.6	2337	"cast_ic([{"credit_id": "52fe4226ca36847f8"}])"	0	0	0	0	0	0	0	0	0								
27	Nothing on Titanic	7.5	7562	"cast_ic([{"credit_id": "52fe4252ca36847f8"}])"	0	0	0	0	2	11	0	0	0								
28	Divided W/Captain Ar	7.1	7241	"cast_ic([{"credit_id": "569443d59251414b6"}])"	0	0	0	0	0	0	0	0	0								
29	Battle Battleship	5.5	2114	"cast_ic([{"credit_id": "52fe4696ca36847f8"}])"	0	0	0	0	0	0	0	0	0								
30	The park is Jurassic W	6.5	8662	"cast_ic([{"credit_id": "52fe4bf7ca368484e"}])"	0	0	0	0	0	0	0	0	0								
31	Think on y Skyll	6.9	7604	"cast_ic([{"credit_id": "52fe466b9251416c9"}])"	0	0	0	0	3	2	0	0	0								
32	There's a Star Wars	6.3	4232	"cast_ic([{"credit_id": "52fe4232ca36847f8"}])"	0	0	0	0	2	1	0	0	0								
33	Unleash th Iron Man 3	6.8	8806	"cast_ic([{"credit_id": "52fe4717ca3684841"}])"	0	0	0	0	1	0	0	0	0								
34	Unleash th Iron Man 3	6.4	4545	"cast_ic([{"credit_id": "52fe44d69251416c7"}])"	0	0	0	0	2	2	1	0	0								
35	They're Invil Alice in Wc	6.4	4545	"cast_ic([{"credit_id": "52fe44d69251416c7"}])"	0	0	0	0	0	0	0	0	0								
36	Calvin & Stu Y Man Th	7.3	2575	"cast_ic([{"credit_id": "52fe43320ca36847f8"}])"	0	0	0	0	0	0	0	0	0								

Academy Awards Oscars Nominees oscars_vlookup tmdb_5000_combined cast cast extracted ... + : 76%

Ready Accessibility: Investigate 16°C Sunny 347 pm 25/05/2023

19. Copy the whole sheet into the new file as a final storage format to perform tasks for the next section

III. Question and Answer

1. Question:

Is there a discernible correlation between the attainment of Academy Awards (Oscars, movie awards only) victories or nominations and the attainment of higher overall ratings on TMDB from mainstream audiences for movies released between the years 1927 and 2017, considering that the movie data in both TMDB and Oscars datasets correspond to the year of release? (using year outside this range can skew result as Oscars datasets do not have a movie before 1927 and after 2017)

Steps performed to produce the answers:

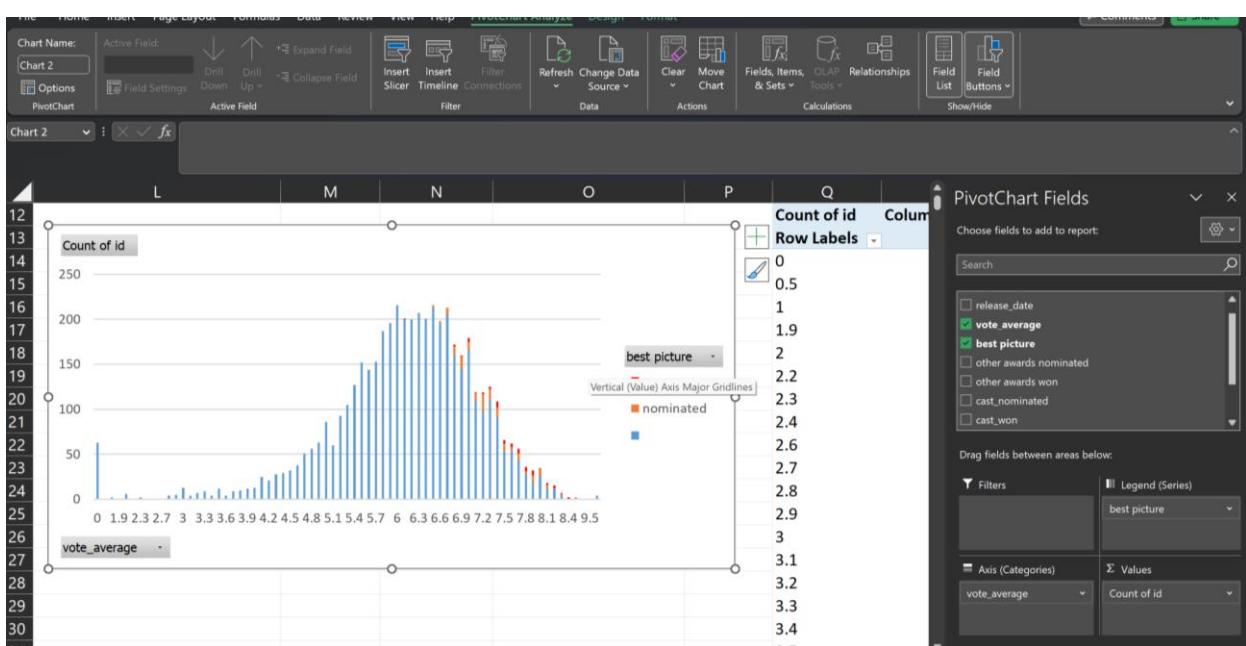
- Only imported the needed column into a new sheet, including id, release date, vote_average, best Picture, other awards nominated, other awards won, cast_nominated, cast_won, crew_nominated, and crew_won.

A	B	C	D	E	F	G	H	I	J
id	release_date	vote_average	best_picture	other_awards_nominated	other_awards_won	cast_nominated	cast_won	crew_nominated	cast_won3
19995	10/12/2009	7.2	nominated	5	3	0	0	0	0
285	19/05/2007	6.9		2	0	1	0	0	0
206647	26/10/2015	6.3		0	1	0	0	0	0
49026	16/07/2012	7.6		0	0	0	1	0	0
49529	7/03/2012	6.1		0	0	0	0	0	0
559	1/05/2007	5.9		0	0	0	0	0	0
38757	24/11/2010	7.4		1	0	0	0	0	0
99861	22/04/2015	7.3		0	0	1	0	0	0
767	7/07/2009	7.4		1	0	0	0	0	0
209112	23/03/2016	5.7		0	0	1	0	0	0
1452	28/06/2006	5.4		1	0	0	0	0	0
10764	30/10/2008	6.1		0	0	0	0	0	0
58	20/06/2006	7		3	1	0	0	0	0
57201	3/07/2013	5.9		2	0	0	0	0	0
49521	12/06/2013	6.5		0	0	1	0	0	0
2454	15/05/2008	6.3		0	0	0	0	0	0
24428	25/04/2012	7.4		0	0	0	0	0	0
1865	14/05/2011	6.4		0	0	0	0	0	0
41154	23/05/2012	6.2		0	0	1	0	0	0
122917	10/12/2014	7.1		1	0	1	0	0	0
1930	27/06/2012	6.5		0	0	1	0	0	0
20662	12/05/2010	6.2		1	0	0	0	0	0
57158	11/12/2013	7.6		3	0	0	1	0	0
2268	4/12/2007	5.8		1	1	0	0	0	0
254	14/12/2005	6.6		3	4	0	0	0	0
597	18/11/1997	7.5	won	2	11	2	0	0	0
271110	27/04/2016	7.1		0	0	0	0	0	0

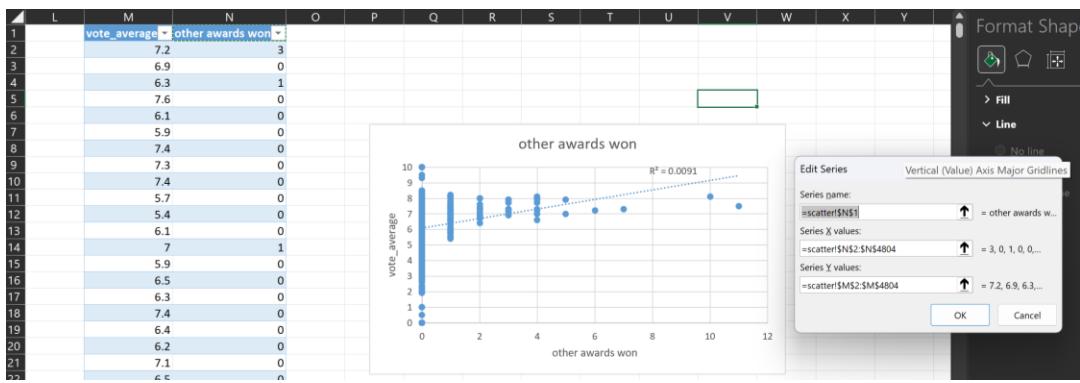
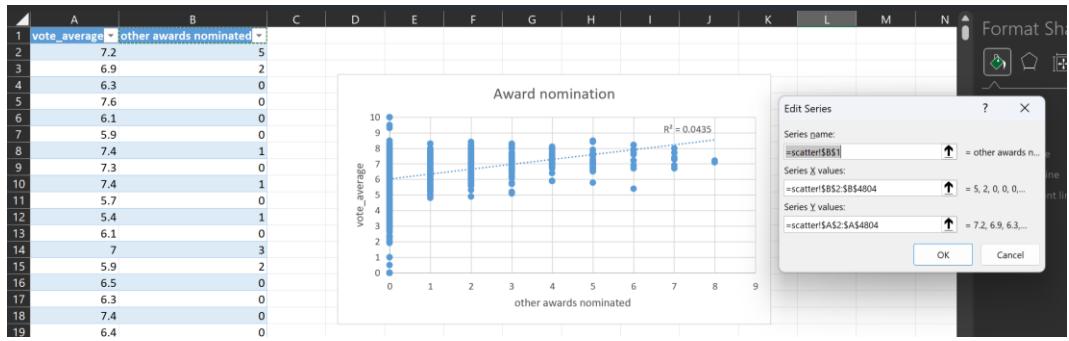
- Create a pivot table that filters out movies outside the 1927 and 2017 range, and calculate average ratings for movies that have not been nominated or won the best Picture, movies that are nominated and movies that won the Oscars.

A screenshot of a Microsoft Excel spreadsheet showing a PivotTable Fields ribbon. The table has columns for 'release_date' (Multiple Items), 'Row Labels' (Average of vote_average, Count of id), 'nominated' (7.252597403, 154), 'won' (7.534693878, 49), and 'Grand Total' (6.092979167, 4800). The PivotTable Fields ribbon shows fields like id, release_date, vote average, best picture, other awards nominated, other awards won, and cast nominated.

- c. Create a pivot chart that shows every rating movie received based on the best picture category.



- d. Create two scatter plots to find the correlation between the vote average and the number of awards won and nominated. In the scatter plot, y-axis is the average rating, and x-axis is the number of awards nominated and awards won, respectively,



Answer:

Based on the available data analysis, there appears to be a positive correlation between the number of Oscars won or nominated by movies and the user ratings they receive on TMDB. Specifically, the data indicate that movies with a higher number of Oscar wins tend to receive higher scores on TMDB compared to movies with fewer wins. Furthermore, on average, movies that have won a certain number of Oscars tend to have higher TMDB scores than those that have received Oscar nominations. For example, the pivot table in b indicates that a movie that has not won any Oscars and was only nominated for Best Picture has an average rating of 6.04. In contrast, movies that were nominated for Best Picture have an average rating of 7.25, while movies that actually won the Best Picture award have an average rating of 7.53. Additionally, this trend is also observed in other award categories. Both the scatter plot of other award nominations and TMDB scores, as well as the scatter plot of other award wins and TMDB scores, show a weak positive relationship. However, it's worth noting that the correlation between nominations and TMDB scores is stronger, but this may be due to the small size of the data.

2. Question:

To what extent does the recognition of cast and crew with Oscar awards in the same year as the release of their respective movies influence the audience score on TMDB, for movies released between the years 1927 and 2017?

Steps performed to produce the answers:

- Copy the columns that are required for this question: id, release_date, vote_average, cast_nominated, cast_won, crew_nominated, crew_won

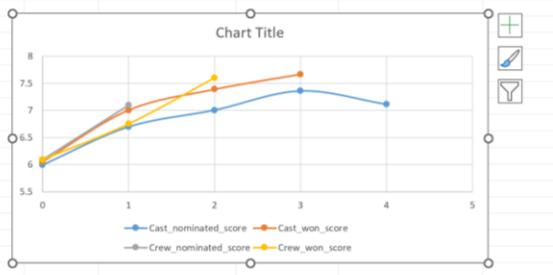
A screenshot of Microsoft Excel showing a table of data. The table has columns labeled A through S. The data includes columns for 'release_date' (e.g., 10/12/2009, 19/05/2007), 'vote_average' (e.g., 7.2, 6.9), and 'cast_nominated' (e.g., 0, 1). The table has approximately 30 rows of data.

- Create the pivot table that shows an average of vote_average for every number of cast and crew nominated and won an Oscars award

A screenshot of Microsoft Excel showing two pivot tables. The top pivot table has columns for 'Cast nominated' (0, 1, 2, 3, 4) and 'Grand Total'. The values for 'Cast nominated' are 5.996132859, 6.696583144, 7.004504505, 7.357692308, and 7.111111111. The 'Grand Total' is 6.092979167. The bottom pivot table has columns for 'Crew nominated' (0, 1) and 'Grand Total'. The values for 'Crew nominated' are 6.0925594 and 7.1. The 'Grand Total' is 6.092979167. Both tables have 'release_date' as a column header.

- For a visualisation, copy the result from the pivot table in a separate row and create a line chart based on the information.

Cast&Crew	Cast_nominated_score	Cast_won_score	Crew_nominated_score	Crew_won_score
0	5.996132859	6.051825293	6.0925594	6.091292545
1	6.696583144	7.004519774	7.1	6.75
2	7.004504505	7.388888889		7.6
3	7.357692308	7.666666667		
4	7.111111111			



Answer:

The data suggests that there is a clear connection between the number of cast and crew members who receive Oscar nominations or awards in the same year as a movie's release and the audience's score on TMDB. This conclusion is supported by evidence showing the average scores for different levels of cast and crew recognition at the Oscars. Therefore, we can confidently say that there is a correlation between the presence of Oscar-winning cast and crew members in a film's release year and the audience's score on TMDB.

3. Question:

To what extent can the Oscars best picture awards be considered indicative of a film's financial success for those released between the years 1927 and 2017?

Steps performed to produce the answers:

- Copy only the columns needed, which are id, release_date, budget, revenue, and best Picture, then Calculate the profit by assuming that it is revenue minus budget

	A	B	C	D	E	F	G	H
1	Id	release_date	budget	revenue	Profit	best picture		
2	19995	10/12/2009	237000000	2787965087	2550965087	nominated		
3	285	19/05/2007	300000000	961000000	661000000			
4	206647	26/10/2015	245000000	880674600	635674600			
5	49026	16/07/2012	250000000	1084939099	834939099			
6	49529	7/03/2012	260000000	284139100	24139100			
7	559	3/05/2007	258000000	890871626	632871626			
8	38757	24/11/2010	260000000	591794936	331794936			
9	99861	22/04/2015	280000000	1405403694	1125403694			
10	767	7/07/2008	250000000	933959197	683959197			
11	209112	23/03/2016	250000000	873260194	623260194			
12	1452	28/06/2006	270000000	391081192	121081192			
13	10764	30/10/2008	200000000	586090727	386090727			
14	58	26/06/2006	200000000	1065659812	865659812			
15	57201	3/07/2013	255000000	89289910	-165710090			
16	49521	12/06/2013	225000000	662845518	437845518			
17	2454	15/05/2008	225000000	419651413	194651413			
18	24428	25/04/2012	220000000	1519557910	1299557910			
19	1865	14/05/2011	380000000	1045713802	665713802			
20	41154	23/05/2012	225000000	624026776	399026776			
21	122917	10/12/2014	250000000	956019780	706019788			
22	1930	27/06/2012	215000000	752215857	537215857			
23	20662	12/05/2010	200000000	310669540	110669540			
24	57158	11/12/2013	250000000	958400000	708400000			
25	2268	4/12/2007	180000000	372234864	192234864			
26	254	14/12/2005	207000000	550000000	343000000			
27	597	18/11/1997	200000000	1845034180	1645034188	won		
28	271110	27/04/2016	250000000	11533040495	9033040495			

- b. While creating a pivot table, it is found that 891 movies have 0 profit, most of which results from having both 0 budget and revenue; thus, excluding this movie is necessary not to skew the result.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
1	Id	release_date	budget	revenue	Profit	best picture									
1807	2779	5/08/2001	0	0	0						24892047	1			
1826	254024	6/09/2014	0	0	0						25696770	1			
1845	175555	13/12/2013	0	0	0						26999213	1			
1856	9416	22/08/1997	0	0	0						27483764	1			
1859	20694	13/06/2003	0	0	0						27663038	1			
1867	10735	27/03/2003	0	0	0						28405313	1			
1882	21724	22/05/2009	0	0	0						28850000	1			
1901	17170	1/04/1988	0	0	0						30055439	1			
1902	18975	26/07/1996	0	0	0						30100000	1			
1910	12770	13/09/2002	0	0	0						30345360	1			
1921	133698	29/10/2012	0	0	0						30891409	1			
1932	21338	17/05/1991	0	0	0						31416464	1			
1935	19803	2/09/2005	0	0	0						31489558	1			
1939	27360	22/04/2005	0	0	0						31726644	1			
1949	49953	3/08/2010	0	0	0						32175038	1			
1951	39845	30/06/2010	0	0	0						32300000	1			
1952	25353	1/04/2009	0	0	0						32302796	1			
1975	18828	12/01/2001	0	0	0						33719388	1			
1987	45881	1/01/2004	0	0	0						34192128	1			
2003	19150	21/08/2000	0	0	0						34956140	1			
2006	9819	18/12/1996	0	0	0						35065672	1			
2007	13579	22/08/2008	0	0	0						35099643	1			
2011	11156	22/04/2009	0	0	0						35489265	1			
2040	10279	2/07/1999	0	0	0						37656822	1			
2049	30379	7/09/2001	0	0	0						38034460	1			
2050	42586	1/01/1970	0	0	0						38100000	1			
2053	7548	16/09/2004	0	0	0						38369434	1			
2057	12085	17/01/2008	0	0	0						38715192	1			

- c. Create a pivot table that filters out movies outside the year range, 0 budget and revenue, which indicate the average, max and min profit for each best picture type movie has achieved.

	Average of Profit	Min of Profit	Max of Profit
release_date	73581356.8805675	-165710090	1363528810
budget	(Multiple Items)	(Multiple Items)	(Multiple Items)
revenue	(Multiple Items)	(Multiple Items)	(Multiple Items)
Row Labels	Average of Profit	Min of Profit	Max of Profit
nominated	180796562.1554050	-46604061	2550965087
won	216824147.3541670	20984230	1645034188
Grand Total	80629235.51	-165710090	2550965087

- d. The pivot table in (c.) may not be enough to conclude that Oscar's best Picture can be a financial success indicator for films. Therefore, it is crucial to consider if the best movies not in the nominated or won category are not as financially successful as those in. Firstly, get a list of the top 10 highest-profit movies for each best picture category using the filter function and copy the value into a new sheet.

A	B	C	D	E	F
id	release_date	budget	revenue	Profit	best picture
19995	10/12/2009	237000000	2787965087	2550965087	nominated
10193	16/06/2010	200000000	1066969703	866969703	nominated
121	18/12/2002	79000000	926287400	847287400	nominated
601	3/04/1982	10500000	792910554	782410554	nominated
120	18/12/2001	93000000	871368364	778368364	nominated
11	25/05/1977	11000000	775398007	764398007	nominated
27205	14/07/2010	160000000	825532764	665532764	nominated
745	6/08/1999	40000000	672806292	632806292	nominated
49047	27/09/2013	105000000	716392705	611392705	nominated
14160	13/05/2009	175000000	735099082	560099082	nominated
597	18/11/1997	200000000	1845034188	1645034188	won
122	1/12/2003	94000000	1118888979	1024888979	won
13	6/07/1994	55000000	677945399	622945399	won
581	9/11/1990	22000000	424208848	402208848	won
45269	6/09/2010	15000000	414211549	399211549	won
380	11/12/1988	25000000	412800000	387800000	won
12405	12/05/2008	15000000	377910544	362910544	won
98	1/05/2000	103000000	457640427	354640427	won
14	15/09/1999	15000000	356296601	341296601	won
424	29/11/1993	22000000	321365567	299365567	won
135397	9/06/2015	150000000	1513528810	1363528810	won
168259	1/04/2015	190000000	150624936	131624936	won
24428	25/04/2012	220000000	1519557910	1299557910	won
99861	22/04/2015	280000000	1405403694	1125403694	won
109445	27/11/2013	150000000	1274219009	1124219009	won
211672	17/06/2015	74000000	1156730962	1082730962	won
68723	18/04/2013	200000000	1215439994	1015439994	won

- e. Create a pivot table the same way as (c.)

The screenshot shows a Microsoft Excel spreadsheet titled "Section 3 Q&A". The PivotTable Fields pane on the right indicates that the data includes fields such as id, release_date, budget, revenue, Profit, and best_picture. The main table displays the following data:

	Row Labels	Sum of Profit4	Average of Profit	Max of Profit2	Min of Profit3
7	nominated	11067742243.00	1106774224.30	1363528810.00	903304495.00
8	won	9060229958.00	906022995.80	2550965087.00	560099082.00
9	Grand Total	25968274303	865609143.4	2550965087	299365567

Answer:

The data in (c.) shows that the average profit for movies nominated or won Oscars Best Picture awards is significantly higher than those not. However, the data in (d.) suggests that the average profit from the top 10 highest-profit films in won and nominated categories is less than movies that are not in these categories, which indicates that films that were not nominated or won Oscar Best Picture could also be financially successful and even more than those that achieved the award. To conclude, the Oscar Best Picture cannot be a movie's financial success indicator.

IV. Individual Reflection

A situation when you applied what you had already learned in Data Management to reason, think, and decide

(I did not do the Data Management paper as I took Database Systems as a prerequisite for this course.)

I use some of the concepts about ERD relationships. For example, when I combined TMDB movies with the Oscars, I initially used Vlookup to return the type of awards films achieved. After reviewing the result, I found that the movie I'm sure won Best Picture returned other awards instead. So I re-audited the datasets and saw that the TMDB dataset has an optional one-to-many relationship with the Oscar dataset, where a movie can have zero to many Oscars. Then I had to change the method. I combined them to count how many Awards each film achieved.

A situation when you applied self-management strategies to complete the project tasks.

The first thing I did for this assignment was planned what section to be done first, which is Section II → III → I → IV. Then I chose to prioritise what section based on the mark allocated, which in this project, I spent most of my time doing Section II. Then, I set goals for starting and completing each task based on my timetable, workload of each section and assignment due dates from other courses to not fall behind or be too late to change something that I need to inform the teaching staff.

A situation when you applied resilience strategies to complete the project tasks.

One task that I found too complicated is writing an Excel function that extracts just the name of cast&crew from the object where a cell can only contain one name. Therefore, I overcame the challenge by, instead of writing from scratch, I broke down and simplified the task by doing a function one by one (8th and 9th steps), such as finding when the value starts and ends and how long it is, then nested all the functions together in one cell so that I can apply it to other cells without do it one by one for every cell.