

IMDB Movie Database Case Study Using Pivot Table

Xiaojia Zhang
3/23/2020

1. Excel Pivot Table

- a. Set up a view to show **Budget** by **Title** (as rows), with filters for **Country** (set to *Japan*) and **Language** (set to *English*). How many Japanese movies in the database were produced in English?

Country	Japan	
Language	English	
Row Labels		
		Sum of Budget
One Missed Call	\$27,000,000	
Street Fighter	\$35,000,000	
The Quick and the Dead	\$32,000,000	
Tora! Tora! Tora!	\$25,000,000	
Trainwreck	\$35,000,000	
Grand Total	\$154,000,000	

There are five Japanese movies in the database were produced in English.

- b. Clear the **Language** filter, set the **Country** filter to *Denmark*, and pull in **Gross Revenue** as a second metric. How much gross revenue did “*The Celebration*” generate?

Country	Denmark	
Language	(All)	
Row Labels		
	Sum of Budget	Sum of Gross Revenue
A Lego Brickumentary	\$1,000,000	\$100,240
Dancer in the Dark	\$12,800,000	\$4,157,491
Flame and Citron	\$45,000,000	\$145,109
Manderlay	\$14,200,000	\$74,205
Melancholia	\$7,400,000	\$3,029,870
Only God Forgives	\$4,800,000	\$778,565
Smilla's Sense of Snow	\$35,000,000	\$2,221,994
The Celebration	\$1,300,000	\$1,647,780
The Hunt	\$3,800,000	\$610,968
Grand Total	\$125,300,000	\$12,766,222

“*The Celebration*” generated \$1,647,780 gross revenue.

- c. Use the “Clear All” command in the PivotTable Analyze options to remove all fields from the table, and create a new view showing **Gross Revenue** by **Country** (as rows) and **Genre** (as columns). How much revenue was generated by Comedy films in Finland?

Sum of Gross Revenue	Column Labels	
Row Labels	Comedy	Grand Total
Finland	\$611,709	\$611,709
Grand Total	\$611,709	\$611,709

\$611,709 was generated by Comedy films in Finland.

- d. Remove the **Country** field, move **Genre** to the row labels, and drag in **Rating** as secondary row labels. How much revenue was generated by PG-rated Family films? Double click on the cell to see the exact source data populating the value. Which title drove most of the revenue?

Family	
PG	
E.T. the Extra-Terrestrial	\$434,949,459
Raise Your Voice	\$10,411,980
PG Total	\$445,361,439
Family Total	\$445,361,439

\$445,361,439 revenue was generated by PG-rated Family films.

E.T. the Extra-Terrestrial drove most of the revenue.

2. Excel Pivot Table Formatting

- a. Show **Budget** and **Gross Revenue** by **Title**, and change the **number format** to **currency**, with a dollar sign and no decimal places. What was the budget for “A Passage to India”?

Row Labels	Sum of Gross Revenue	Sum of Budget
A Passage to India	\$26,400,000	\$16,000,000
Grand Total	\$26,400,000	\$16,000,000

The budget for “A Passage to India” is \$16,000,000.

- b. Remove **Budget** and **Title**, show **Gross Revenue** by **Genre** (rows) and **Rating** (columns). Update the PivotTable options to show “\$0” instead of blank values

Sum of Gross Revenue	Column Labels														
Row Labels	PG-13	R	PG	G	Approved	X	Unrated	M	Not Rated	GP	(blank)	Passed	NC-17	Grand Total	
Action	\$40,036,209,554	\$16,092,170,595	\$10,950,373,811	\$96,090,224	\$244,967,035	\$44,793,200	\$269,061	\$22,800,000	\$1,316,558	\$43,800,000	\$8,250,780	\$0	\$0	\$67,540,890,818	
Comedy	\$17,635,413,030	\$11,749,018,050	\$7,847,520,104	\$628,777,083	\$105,815,005	\$22,948,916	\$5,677,168	\$0	\$26,384,674	\$0	\$5,676,960	\$0	\$762,507	\$38,027,993,497	
Adventure	\$7,967,864,978	\$2,045,278,178	\$15,452,844,612	\$5,466,455,165	\$7,970,873	\$0	\$0	\$0	\$686,383	\$0	\$8,033,176	\$22,202,612	\$0	\$30,971,335,977	
Drama	\$9,924,568,486	\$7,686,364,177	\$1,819,043,120	\$360,597,043	\$72,000,000	\$0	\$77,902,509	\$0	\$69,265,971	\$0	\$11,373,871	\$0	\$24,303,265	\$20,045,418,442	
Crime	\$1,588,280,368	\$6,473,562,062	\$0	\$0	\$0	\$0	\$43,977,919	\$0	\$12,644,242	\$0	\$100,609	\$0	\$1,141,829	\$8,119,707,029	
Biography	\$3,337,237,188	\$2,815,992,534	\$993,271,055	\$177,412,152	\$57,750,000	\$0	\$866,313	\$102,308,900	\$484,221	\$0	\$927,107	\$8,000,000	\$653,621	\$7,494,903,091	
Horror	\$804,331,593	\$4,330,823,209	\$21,884,663	\$0	\$0	\$38,273,641	\$3,645,438	\$0	\$12,843	\$0	\$0	\$0	\$0	\$5,198,971,387	
Animation	\$16,855,175	\$63,418,514	\$2,733,967,565	\$727,207,806	\$269,225,485	\$0	\$0	\$0	\$0	\$0	\$0	\$0	\$0	\$3,810,674,545	
Fantasy	\$696,992,331	\$574,631,031	\$106,100,000	\$0	\$0	\$22,168,359	\$0	\$0	\$3,478	\$0	\$0	\$0	\$0	\$1,399,895,199	
Mystery	\$679,235,812	\$416,829,007	\$0	\$0	\$0	\$0	\$31,899,000	\$0	\$0	\$0	\$0	\$0	\$0	\$1,127,963,819	
Documentary	\$40,369,512	\$144,421,497	\$175,109,754	\$180,682,783	\$0	\$0	\$1,443,875	\$0	\$443,645	\$0	\$5,743,976	\$0	\$0	\$548,215,042	
Family	\$0	\$0	\$445,361,439	\$2,119,994	\$0	\$0	\$0	\$0	\$0	\$0	\$0	\$0	\$0	\$447,481,433	
Sci-Fi	\$37,402,542	\$91,585,988	\$79,568,000	\$0	\$0	\$0	\$0	\$0	\$0	\$0	\$18,195	\$0	\$0	\$208,574,725	
Musical	\$181,360,000	\$0	\$0	\$0	\$0	\$0	\$0	\$0	\$0	\$0	\$2,808,000	\$0	\$184,168,000		
Romance	\$62,453,315	\$0	\$0	\$0	\$0	\$0	\$0	\$0	\$0	\$0	\$205,701	\$0	\$0	\$62,659,016	
Western	\$0	\$41,643,768	\$0	\$0	\$6,100,000	\$0	\$0	\$0	\$0	\$0	\$0	\$0	\$0	\$47,743,768	
Thriller	\$0	\$2,468	\$0	\$0	\$0	\$0	\$0	\$0	\$70,071	\$0	\$49,000	\$0	\$0	\$121,539	
Grand Total	\$83,008,423,884	\$52,525,741,078	\$40,625,044,123	\$7,639,342,250	\$763,828,398	\$160,083,116	\$133,782,283	\$125,108,900	\$111,312,086	\$43,800,000	\$40,379,375	\$33,010,612	\$26,861,222	\$185,236,717,327	

- c. Move **Rating** to the row labels (beneath **Genre**), change your table layout to **Outline View**, and Update your column headers from "*Rating*" to "*Film Rating*", and from "*Sum of Gross Revenue*" to "*Gross Revenue*" (**hint:** you may need a trailing space)

Genre	Film Rating	Gross Revenue
Action		\$67,540,890,818
	PG-13	\$40,036,059,554
	R	\$16,092,170,595
	PG	\$10,950,373,811
	Approved	\$244,967,035
	G	\$96,090,224
	X	\$44,793,200
	GP	\$43,800,000
	M	\$22,800,000
	(blank)	\$8,250,780
	Not Rated	\$1,316,558
	Unrated	\$269,061
Comedy		\$38,027,993,497
	PG-13	\$17,635,413,030
	R	\$11,749,018,050
	PG	\$7,847,520,104
	G	\$628,777,083
	Approved	\$105,815,005
	Not Rated	\$26,384,674
	X	\$22,948,916
	Unrated	\$5,677,168
	(blank)	\$5,676,960
	NC-17	\$762,507
Adventure		\$30,971,335,977

(It's a part of the table)

- d. Remove Film Rating from the view, so that you're just viewing **Gross Revenue** by **Genre**. Turn **Grand Totals** off, select the Gross Revenue values, format as *currency* (if they aren't already) and add a **Color Scale** from Green (high) to Red (low). Which Genre produced the most Gross Revenue?

Genre	Gross Revenue
Action	\$67,540,890,818
Comedy	\$38,027,993,497
Adventure	\$30,971,335,977
Drama	\$20,045,418,442
Crime	\$8,119,707,029
Biography	\$7,494,903,091
Horror	\$5,198,971,387
Animation	\$3,810,674,545
Fantasy	\$1,399,895,199
Mystery	\$1,127,963,819
Documentary	\$548,215,042
Family	\$447,481,433
Sci-Fi	\$208,574,725
Musical	\$184,168,000
Romance	\$62,659,016
Western	\$47,743,768
Thriller	\$121,539

Action produced the most Gross Revenue.

- e. Add a second instance of Gross Revenue, format the new column with **Data Bars**. Update the **number format** to make the text invisible, so that only the bars appear. Which Genre produced the second-highest Gross Revenue total in the sample?

Genre	Gross Revenue	Sum of Gross Revenue
Action	\$67,540,890,818	
Comedy	\$38,027,993,497	
Adventure	\$30,971,335,977	
Drama	\$20,045,418,442	
Crime	\$8,119,707,029	
Biography	\$7,494,903,091	
Horror	\$5,198,971,387	
Animation	\$3,810,674,545	
Fantasy	\$1,399,895,199	
Mystery	\$1,127,963,819	
Documentary	\$548,215,042	
Family	\$447,481,433	
Sci-Fi	\$208,574,725	
Musical	\$184,168,000	
Romance	\$62,659,016	
Western	\$47,743,768	
Thriller	\$121,539	

Comedy produced the second-highest Gross Revenue total in the sample.

3. Sorting, Filtering & Grouping Data with Excel Pivot Tables

- a. Create a view showing **Gross Revenue** by **Title**, with a filter for **Year** to only include films released in 2005, 2006, 2007 or 2008, then sort the titles descending by Gross Revenue. What's the top-grossing film released during that 4-year sample? (**Note:** if the Release Dates don't auto-group, you will need to use the "Group" tools in the Analyze tab or create a new column in your raw data to extract the year from the Release Date column)

Years	(Multiple Items)	▼
Title	▼	Gross Revenue
The Dark Knight		\$533,316,061
Pirates of the Caribbean: Dead Man's Chest		\$423,032,628
Star Wars: Episode III - Revenge of the Sith		\$380,262,555
Spider-Man 3		\$336,530,303
Shrek the Third		\$320,706,665
Transformers		\$318,759,914
Iron Man		\$318,298,180
Indiana Jones and the Kingdom of the Crystal Skull		\$317,011,114
Pirates of the Caribbean: At World's End		\$309,404,152
Harry Potter and the Order of the Phoenix		\$292,000,866
The Chronicles of Narnia: The Lion, the Witch and the Wardrobe		\$291,709,845
Harry Potter and the Goblet of Fire		\$289,994,397
I Am Legend		\$256,386,216

(It's a part of the table) *The Dark Knight* is the top-grossing film released during that 4-year sample.

- b. Add a **Label Filter** to only include titles that end in "2". How many sequels were released during these years? Which earned the most Gross Revenue?

Years	(Multiple Items)	▼
Title	▼	Gross Revenue
Cheaper by the Dozen 2		\$82,569,532
Big Momma's House 2		\$70,163,652
Transporter 2		\$43,095,600
The Grudge 2		\$39,143,839
Basic Instinct 2		\$5,851,188
Hamlet 2		\$4,881,867
Ong-bak 2		\$102,055

7 sequels were released during these years.

Cheaper by the Dozen 2 earned the most Gross Revenue.

- c. Clear your label filter, add a **Value Filter** to only show titles that earned between \$1,000,000 and \$3,000,000 in Gross Revenue. How many titles fell into this range?

Years	(Multiple Items)
Title	Gross Revenue
Miss Potter	\$2,975,649
Man on Wire	\$2,957,978
Gracie	\$2,955,039
Choke	\$2,926,565
Tsotsi	\$2,912,363
Things We Lost in the Fire	\$2,849,142
How to Lose Friends & Alienate People	\$2,775,593
The History Boys	\$2,706,659
Half Nelson	\$2,694,973
Frozen River	\$2,508,841
The Lost City	\$2,483,955
You Kill Me	\$2,426,851
Redbelt	\$2,344,847
Waltz with Bashir	\$2,283,276
Perfume: The Story of a Murderer	\$2,208,939
Transsiberian	\$2,203,641
The Host	\$2,201,412
The Wackness	\$2,077,046
The Work and the Glory II: American Zion	\$2,024,854
Oliver Twist	\$1,987,287
Shortbus	\$1,984,378
The Proposition	\$1,900,725
A Sound of Thunder	\$1,891,821
Quinceañera	\$1,689,999

(It's a part of the table) 51 titles fell into this range.

- d. Adjust your PivotTable Options to allow multiple filters, then add a label filter to only show movies that start with the letter "M". How many titles are now listed?

Years	(Multiple Items)
Title	Gross Revenue
Miss Potter	\$2,975,649
Man on Wire	\$2,957,978
Murderball	\$1,523,883
Married Life	\$1,506,998
Me and Orson Welles	\$1,186,957

5 titles are listed.

- e. Add a wildcard to your label filter to only show titles that start with the letter "M" and also contain the letter "s", separated by any number of characters. Which titles are returned?

Years	(Multiple Items)
Title	Gross Revenue
Miss Potter	\$2,975,649
Me and Orson Welles	\$1,186,957

Miss Potter and Me and Orson Welles

4. Enriching Data with Pivot Table Calculated Values & Fields

- a. Create a view to show IMDB Score by Title. What happens when you replace Title with Genre? How can you fix this issue? (Hint: look at the summarization type)

Genre	Sum of IMDB Score (1-10)	Genre	Average of IMDB Score (1-10)
Action	5653.3	Action	6.281444444
Adventure	2373.2	Adventure	6.555801105
Animation	304.3	Animation	6.762222222
Biography	1458.5	Biography	7.149509804
Comedy	6269.3	Comedy	6.170570866
Crime	1748.3	Crime	6.937698413
Documentary	298.9	Documentary	6.951162791
Drama	4587.2	Drama	6.826190476
Family	19.5	Family	6.5
Fantasy	218.2	Fantasy	6.234285714
Horror	874.4	Horror	5.790728477
Musical	13.5	Musical	6.75
Mystery	151.5	Mystery	6.586956522
Romance	18.4	Romance	6.133333333
Sci-Fi	52.7	Sci-Fi	6.5875
Thriller	15.9	Thriller	5.3
Western	20.3	Western	6.766666667

It shows the sum of ratings of all the genre.

Change the summarization to average.

- b. Update your view to show Average IMDB Score by Genre (primary row labels) and Year (secondary row labels), for 2011-2014. Drag in a second instance of IMDB Score, change the summarization to Average, and show the values as a Rank (large to small) based on the year. Which year in the 4-year sample saw the highest-rated Biography films on average? The lowest?

Genre	Years	Average of IMDb Score (1-10)	Average of IMDb Score (1-10)2
Action		6.399428571	
	2011	6.255319149	4
	2012	6.346341463	3
	2013	6.56	1
	2014	6.453191489	2
Adventure		6.643835616	
	2011	6.4	4
	2012	6.763157895	2
	2013	6.595454545	3
	2014	6.8125	1
Animation		6.7375	
	2011	5.35	4
	2012	7.333333333	1
	2013	7	3
	2014	7.2	2
Biography		7.088372093	
	2011	6.811111111	4
	2012	7.1875	2
	2013	7.491666667	1
	2014	6.864285714	3

2013 is the highest-rated Biography film and 2011 is the lowest.

- c. Add in a column for **Gross Revenue**, show the values as the **% Difference From** the previous year. By what percentage did Action movie revenue grow in 2014?

Genre	Years	Average of IMDb Score	Average of IMDb Score	Sum of Gross Revenue
Action		6.399428571		
	2011	6.255319149	4	
	2012	6.346341463	3	4.65%
	2013	6.56	1	-13.37%
	2014	6.453191489	2	51.70%
Adventure		6.643835616		

51.7% for Action movie grew in 2014.

- d. Create two new **calculated fields** named "**Profit**" (*Gross Revenue - Budget*), and "**Profit Margin**" (*Profit / Gross Revenue*). Update the view to show both new fields by **Title**. Which Title generated the strongest Profit Margin in the entire sample (across all years)?

Title	Sum of Profit	Sum of Profit Margin
Paranormal Activity	\$107,902,283	99.99%
Tarnation	\$591,796	99.96%
The Blair Witch Project	\$140,470,114	99.96%
The Brothers McMullen	\$10,221,600	99.76%
The Texas Chain Saw Massacre	\$30,775,468	99.73%
El Mariachi	\$2,033,920	99.66%
The Gallows	\$22,657,819	99.56%
Super Size Me	\$11,464,368	99.44%
Halloween	\$46,700,000	99.36%
American Graffiti	\$114,223,000	99.32%

Paranormal Activity generated the strongest Profit Margin in the entire sample across all years.

- e. Create a new calculated field for "**Cast + Director Likes**" (*Cast FB Likes + Director FB Likes*), update the view to show Cast + Director Likes by **Genre**. What are the Cast FB Likes and Director FB Likes of **Documentary** and **Western** films?

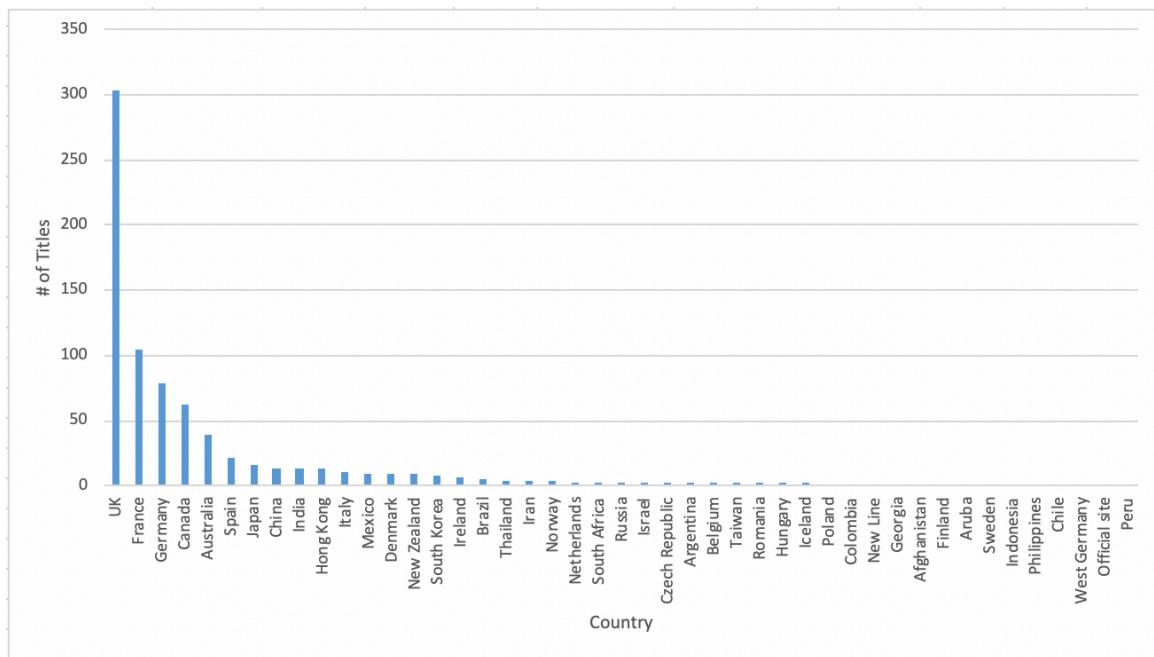
Genre	Sum of Cast+Director
Documentary	1006
Western	16459

Firstly, I create a new column in the raw data ("# of Titles") equal to 1 for each row, then update my calculated field formula to: $(\text{Cast FB Likes} + \text{Director FB Likes}) / \# \text{ of Titles}$.

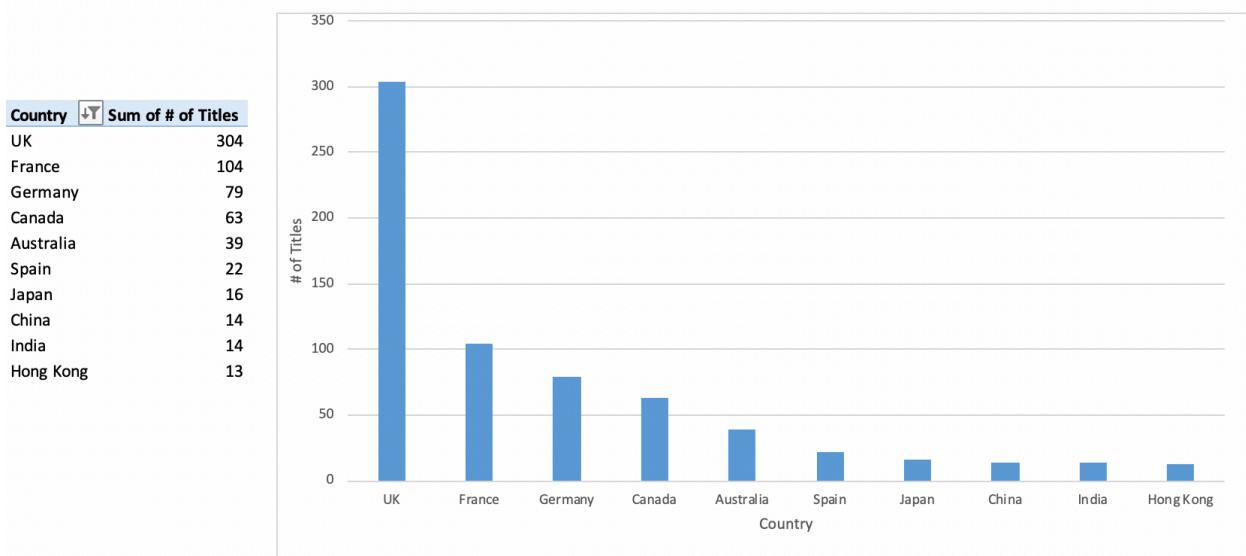
Cast FB Likes and Director FB Likes of **Documentary** and **Western** films are 1006 and 16459.

5. Visualizing Data with Excel Pivot Charts

- a. Create a view show # of Titles by Country, excluding the USA, for the entire sample. Name the PivotTable “Titles by Country”, then use a **PivotChart** to visualize this view as a **Clustered Column Chart**

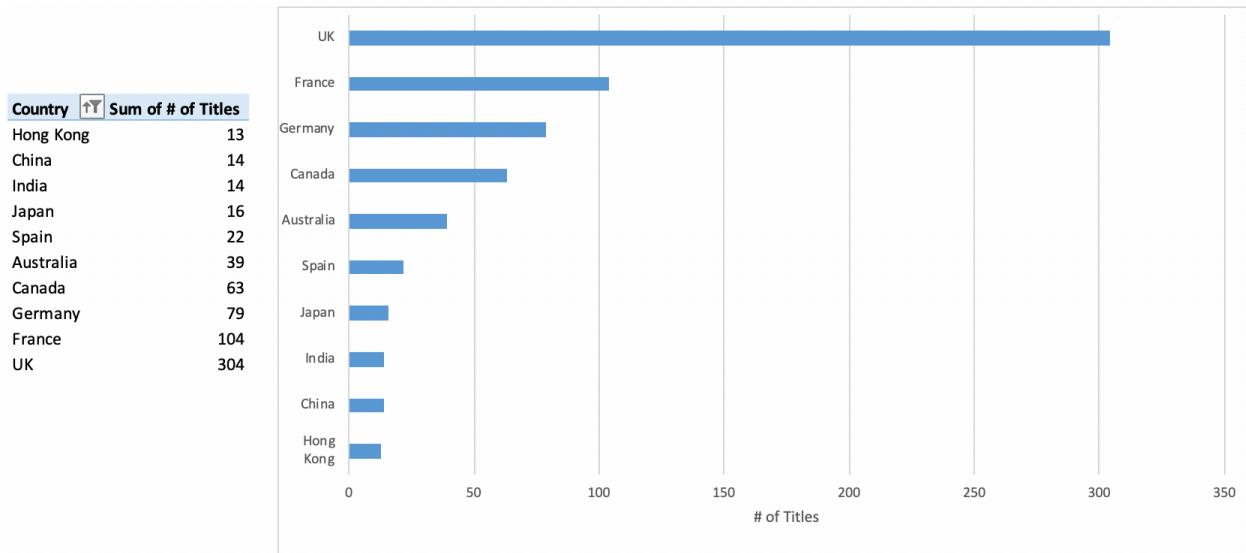


- b. Hide the **Field Buttons** from the PivotChart, then apply a value filter to only show the top 10 countries by # of Titles (*hint: you may need to enable multiple filters*). Which country is #2?

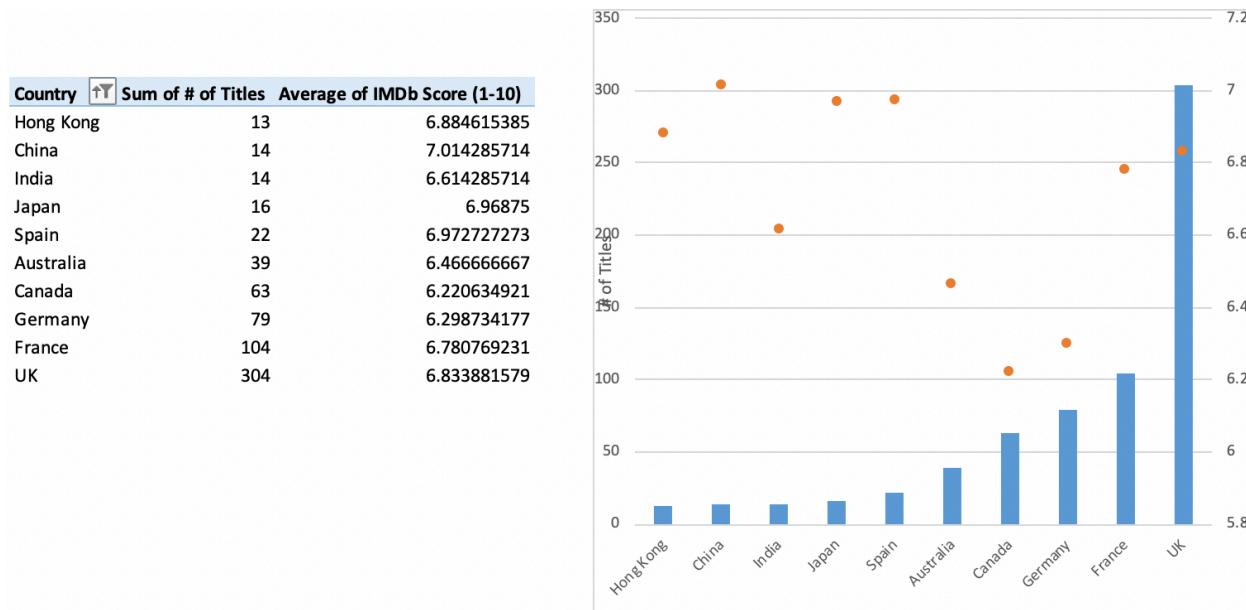


France is #2 of titles.

- c. Change the chart type to a **Clustered Bar**, change the PivotTable sorting to ascending by # of Titles.

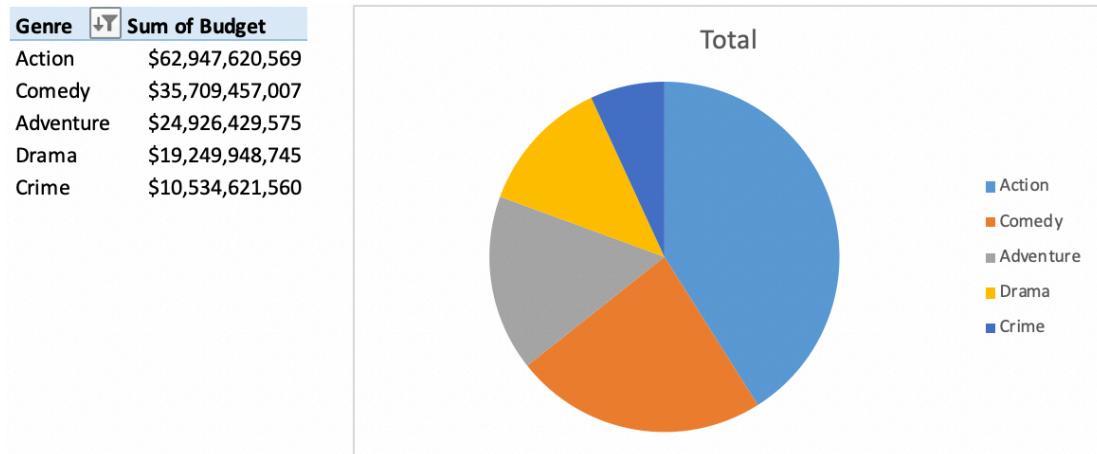


- d. Pull in IMDb Score as a second series, summarize values by **Average**. Change your PivotChart type to **Combo**, with # of Titles as a **Clustered Column** and IMDb Score as a **Line with Markers**, on the **Secondary Axis**. Which of the 10 countries generated the lowest average IMDb scores? (**Bonus:** Format the IMDb series in the chart to only show the markers, with no line)



Canada generated the lowest average IMDB scores.

- e. Copy the existing pivot and create a second view below the combo chart to show **Budget by Genre**, with a Top 5 filter applied. Name the table "Budget by Genre", then visualize this view with a **Pie chart**, with hidden field buttons.



- f. Insert a **Slicer** for Genre, enable **multi-select**, then connect it to both PivotTables. Create a simple dashboard by hiding the columns of your raw PivotTable views, disabling gridlines, and aligning/formatting the Pivot Charts and Slicer as you see fit. Practice adjusting slicer selections to see how the dashboard updates!

