Case Study Instructions

1. Setting Expectations
    a. The goal of this case studies is to demonstrate how the concepts covered in this course can be applied to IMDB Movie data sets.
    b. Some data may have been altered or fabricated to facilitate the demonstrations.

2. Download Case Study Resources
    a. *PivotTable Case Studies.xlsx* (this workbook contains all of the case study data)

3. Dataset Summary
    a. Movie titles, attributes, budgets and revenues from the Internet Movie Database (IMDB), 1920-2015 (n = 3725)

4. Dimensions &Measures
    a. Dimensions: Title, Release Date, Color/B&W, Genre, Language, Country, Rating, Lead Actor, Director Name
    b. Measure: Lead Actor FB Likes, Cast FB Likes, Director FB Likes, Movie FB Likes, IMDB Score, Total Reviews, Duration (min), Gross Revenue, Budget

5. Concepts covered
    a. Sorting
    b. Filtering
    c. Date Grouping
    d. Value Grouping
    e. Calculated Values
    f. Calculated Field
    g. Conditional Formatting (Highlight Cells, Data Bars)
    h. Value Setting (% of Column, % of Parent, % Difference From, Running Total)
    i. Table Layout
    j. Pivot Charts
    k. Slicers
    l. Show Values As (Rank)

6. Questions
    a. Excel Pivot Table
        i. 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?
        ii. 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?
        iii. 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?

iv. 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?

b. Excel Pivot Table Formatting
   i. 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"?
   ii. Remove **Budget** and **Title**, show **Gross Revenue** by **Genre** (rows) and **Rating** (columns). Update the PivotTable options to show "**$0**" instead of blank values
   iii. 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)
   iv. 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?
   v. 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?

c. Sorting, Filtering & Grouping Data with Excel Pivot Tables
   i. 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)
   ii. 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?
   iii. 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?
   iv. 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?

v. 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?

d. Enriching Data with Pivot Table Calculated Values & Fields
   i. 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)
   ii. 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?
   iii. 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?
   iv. 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)?
   v. 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**. If you wanted to show this field as an average across titles, rather than a sum, how could you accomplish this?

e. Visualizing Data with Excel Pivot Charts
   i. 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**?
   ii. 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?
   iii. Change the chart type to a **Clustered Bar**, change the PivotTable sorting to ascending by # of Titles.
   iv. 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)*
   v. 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.

vi. 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!