**Exercise3**

\*\* Click on following link to get all the datasets required for this exercises

[**https://github.com/PurpleGrad/2203/blob/main/Dataset%20%20for%20Exercise%203.zip**](https://github.com/PurpleGrad/2203/blob/main/Dataset%20%20for%20Exercise%203.zip)

**1)**Create a new Power BI report, and load data from the **Films.csv**file in the above folder:



*The only two columns you'll need to do this exercise are the film title and its length in minutes, so you could if you like use****Query Editor****to remove all of the other columns (this is optional).*

Create a calculated column to show the number of minutes for each film using this formula:

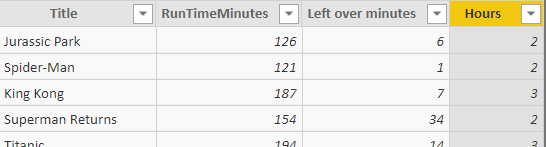
Left over minutes = MOD([RunTimeMinutes],60)

This divides the film length by 60 and gives the remaining minutes (so **Jurassic Park**above should show 6, since it lasts 2 hours and 6 minutes).  Bear with ...

Add another calculated column to give the number of hours (this should equal the film length minus the left over minutes, divided by 60) using this formula:

Hours = ([RunTimeMinutes] - [Left over minutes])/60

You should now have something like this:



*What your table should now look like.*

Use the **&** symbol to concatenate or join together columns to create one more column:



*The final column, giving the film length in hours and minutes.*

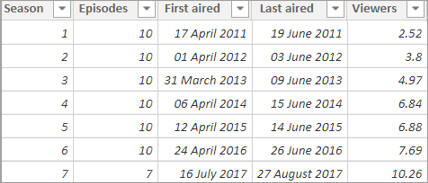
Save this report as **But what about Spider-Man**, then close it down.

**2)**Create a new Power BI report, and load data from the **Series**worksheet in the **Episode** Excel workbook in the above folder:



*You only need to load this worksheet's data.*

Use **Query Editor**to tidy up the Game of Thrones series data:



*Remove the row which has TBA for the number of viewers, remove the extra episodes column and change the data type and name of the viewers column to get something like the above.*

Back in Power BI, use the **DATEDIFF** function to work out the difference in days for each series between when it first and last aired:

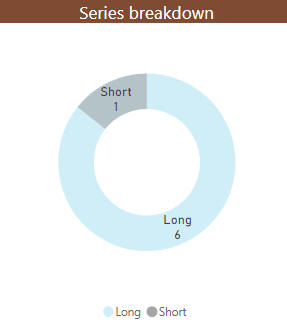


*Don't worry about the arguments for the function - Power BI should prompt you to complete them.*

Now use an **IF** function to show for each series which category it belongs to using these rules:

|  |  |
| --- | --- |
| **Duration** | **Category** |
| Less than 60 days | **Short** |
| Otherwise | **Long** |

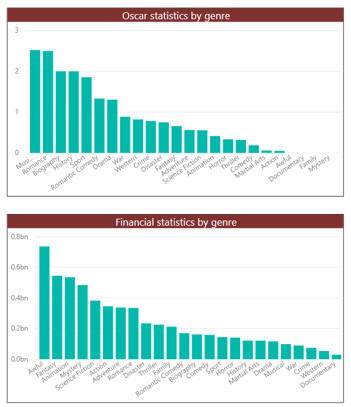
Use this to create a pie chart showing the number of series for each category:



*You don't have to add so much formatting - the person doing this exercise obviously had too much time on his or her hands!*

Save your report as **All too long really**, then close it down.

**3)** Open the Power BI report in the above folder.  It currently shows two charts:



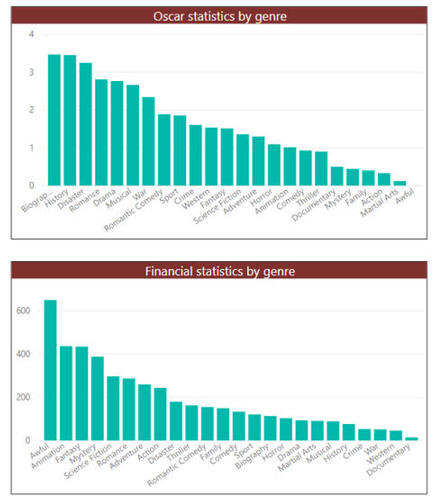
*The charts compare the average Oscars won and the average box-office takings for films by genre.*

In the **Films** table, add two calculated columns:

|  |  |
| --- | --- |
| **Column name** | **What it should show** |
| **Disappointment** | The difference between the number of Oscars a film was nominated for and the number it actually won. |
| **Profit** | The difference between the box office takings and the budget for a film (divided by a million, so you can show the result in millions of dollars rather than in dollars). |

*Note that some films - particularly the ones early on in the table - may not display figures for every row, since either the budget or the box office takings values may be blank.*

Change your two charts to show your newly calculated statistics instead:



*The new charts should show that****Biographical****films have the highest average disappointment factor, but****Awful****ones have the highest average profitability.*

Save this report as **Titanic struggle**, then close it down.