UTS Open Assessment Task

The goal of this assignment is to employ the concepts and skills you have learned throughout the course to analyse a real-world dataset, and to represent your findings in a report.

**Task**

Your tasks include:

* understanding the specifics of the dataset
* extracting information about some of the attributes, and other specifics of the dataset.

The tasks in the assignment are specified below.

**A. Initial data exploration**

1. Identify the type of the attributes {show\_id, type, title, …} (nominal, ordinal, interval or ratio). Also, please justify why you choose the type.
2. To assess the quality of data, calculate number of blank cells for each attribute (or the number of cells with values if you prefer).
3. Highlight the movie or show, whose title contains “Transformers” (or some other key word you are interested in). Tips: to check if a cell contains specific text, you can use the SEARCH function together with the ISNUMBER function: =[ISNUMBER](https://exceljet.net/excel-functions/excel-isnumber-function)([SEARCH](https://exceljet.net/excel-functions/excel-search-function)(substring,text))
4. Sort the table according to the year of release.
5. Identify the values of the summarising properties for theattribute **duration** including medians, means, interquartile range, variances, etc. (Tips: divide the dataset into two spreadsheets according to the type, one for movie and the other for TV show. Then calculate those values for movies and for TV shows.

**B. Data analysis & visualisation**

1. Are there more TV shows or movies on Netflix? Visualise and discuss the result.
2. Which rating the largest number of titles are made with? Visualise and discuss the count of titles according to rating.
3. What is the frequency distribution of duration? Visualise and discuss the result. You will need to create one chart for movies, and the other for TV shows.
4. (Optional) Which country creates most content? Visualise and discuss the result. When a title is produced by multiple countries, use the first one.   
   Tips: this formula extracts the country before the first comma:  
   =IF(ISNUMBER(SEARCH(",", F2)), LEFT(F2, FIND(",",F2,1)-1), F2)
5. Any other questions you would like to answer.

**C. S**ummary

At the end of the report include a summary section in which you summarise your findings. The summary is not a narrative of what you have done, but a condensed informative section of what you have found about the data.

### Dataset

You will be using the dataset of Netflix Movies and TV Shows, consisting of tv shows and movies available on Netflix as of 2019. The dataset is collected from Flixable which is a third-party Netflix search engine.

****

Here is the description of the attributes.

## 

## **Submission**

The deliverables include:

* A report, which structure should follow the tasks of the assignment, and
* An Excel workbook file with individual spreadsheets for each task (spreadsheets should be labelled according to the task names, for example, "1A"). Each of the results of parts (a) through (d) in task 1B should be presented in a separate spreadsheet (and respectively table in the assignment report).

**Report:** In the report include a section (starting with a section title) for each of the tasks in this assignment.

Your report will likely be between 6-10 pages in length using 11 or 12 point Times or Arial fonts, including title page and graphs. On average you will require between 8 and 12 hours to complete this assignment.

Use the filename xxxx.pdf or xxxx.doc for the report and xxxx.xls for the spreadsheet, where xxxx is your full name.

### Due

Sunday 15th November 11.55pm

## **Criteria**

We are looking for your ability to:

* Understand the dataset.
* Analyse and present the findings of the dataset.

See the document below for further detail on the marking criteria for this assessment task.



## **Further Information**

You will be notified when your assignment has been marked and you will be able to view your mark and feedback in the Marks section in the left-hand navigation.