I have had an enjoyable time completing this assignment and showcasing my skills as I have genuinely learnt a few new things about MS Excel and Tableau (such as creating calculated fields) that I can demonstrate and leverage in any analytical role going forward! I have downloaded a dataset of Netflix Titles and associated data in CSV format from Kaggle, and cleaned the file using the Data Cleaning techniques described below, and then created my visualizations and dashboard with KPI’s.

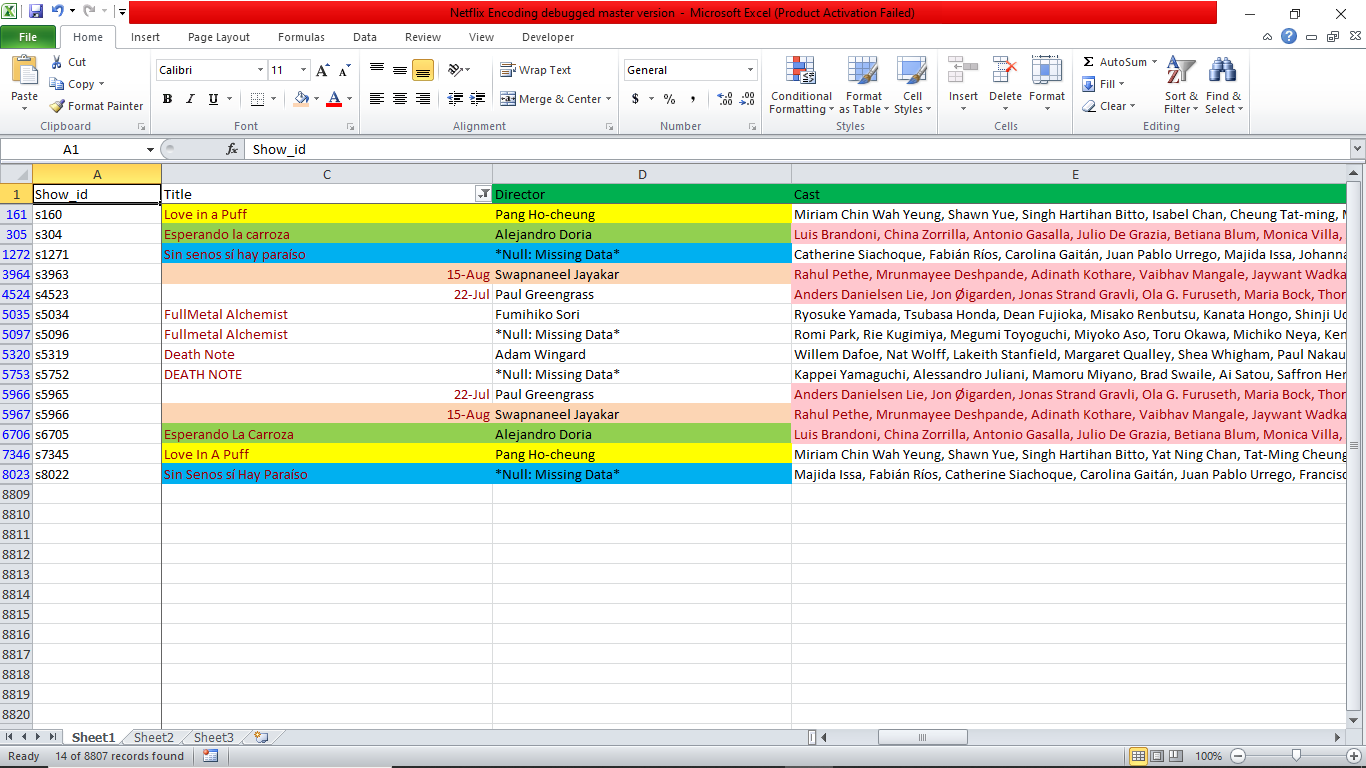
**Data cleaning steps:**

1. My data set contains Director’s names and Titles with accented letters and even in foreign languages altogether. I noticed when I opened the downloaded file that Excel had trouble displaying these characters in their true forms, and instead displayed odd characters in their place. For example, the Mexican documentary, titled ‘Zoé: Panoramas’ was displayed as ‘ZoÃ©: Panoramas’ in the file.

Since I have never faced such encoding issues to do with accented letters and foreign languages in my data, I looked up another way to clean my data on the internet. The solution was to first save the file in .txt type (using ‘Save as’). Then I opened a new Excel workbook and went to the Data tab to select ‘Get External Data from Text file’. For ‘File Origin’ drop-down, I learned I am supposed to change the default to ‘65001: Unicode (UTF-8)’, in order to debug the encoding issues with my data set. I learnt this trick through excelforum.com.

This ‘cleaning technique’ eliminated any time I had to take to replace the broken characters with the right letters or foreign language for multiple affected rows and columns.

1. The next cleaning technique was using the ‘Filter’ function in Excel. I applied it to each column to get a get a summary view of the different records through the menu drop-down next to the column name. For example, I found that there were show duration times in the Rating column, and promptly copied and pasted those few values to the Duration column. This saved time in having to go through each row for thousands of rows to find incorrectly entered data.
2. Thirdly, I used the ‘Text to Columns’ feature in the Data tab to reformat the dates in my data set into a single, uniform format. In the Date\_added column, I found multiple date formats, such as ‘21-02-2012’ and ‘September 16, 2018’. The ‘Text to Columns’ feature allowed me to change the date format to the MDY convention we use in N. America. I then went to the ‘Number’ section on the Home tab to display dates in the long format: ‘month dd, year’, and all this ensured there was uniformity in appearance.
3. I had a few duplicate entries in my data, so I used the Conditional Formatting option to highlight duplicate Titles with red-colour text. I then used the ‘Filter’ feature to ‘Filter by text colour’ and return only these titles (instead of having to sift through thousands of rows to find the few in red-coloured text), and deleted the duplicate entries.



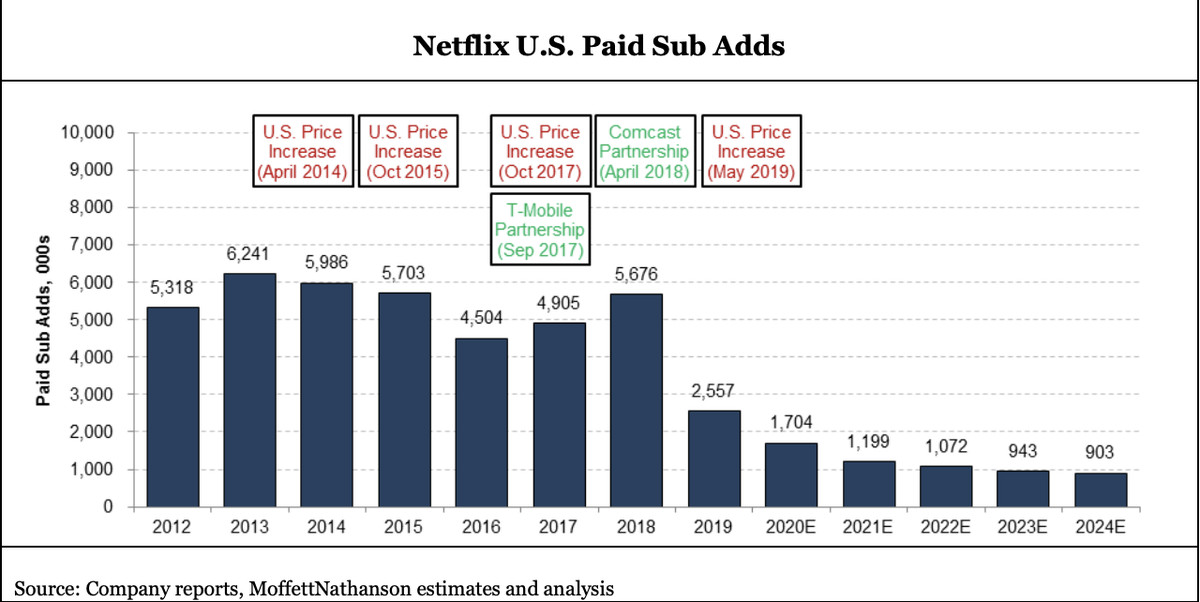
In the above screenshot, you will see that all the duplicate Titles in the database are in red-colour text. This allowed me to quickly delete the rows that were really duplicates – for example, I deleted one of the “Love in a Puff” rows, but I kept both “Fullmetal Alchemist” titles as one was a movie and the other a TV show released by different Director’s and Cast (I confirmed this was the case on imdb.com).

**Insights decision-making:**

Based on the dashboard visualizations, there are a number of decisions an Executive at Netflix may consider, but I have suggested the most important. Noting that 3,210 of the 8,801 titles (Approx. 36.5%) are targeted towards the American audience (demonstrated by the Map visualization – “Number of Netflix Titles by Country”), Netflix must push for more diverse content that appeal to local tastes in different countries around the world.

Partnering with local/international superstar Actors/Actresses from in various regions or with up-and-coming talent to direct and produce exclusive, Netflix-only content found nowhere else. Indeed, a year-old article from The Verge claims U.S. growth has slowed precipitously from 2018 as it has already cornered and dominated that market, even in the face of recent competition from Disney Plus and Amazon Prime. An Analyst describes Netflix’s current and future U.S. growth prospects,

*“Nathanson adds that over the next five years, he expects Netflix to slow down to around 1 million US subscribers each year. It’s quite a departure from the 5 million average subscribers Netflix was adding in the US between 2012 and 2018, as seen in the chart below.”* [*https://www.theverge.com/2020/2/20/21142065/netflix-streaming-wars-subscriber-growth-disney-international*](https://www.theverge.com/2020/2/20/21142065/netflix-streaming-wars-subscriber-growth-disney-international)



There are other highly populous regions such as India, China, and the African continent, each containing more than a Billion people that Netflix has yet to completely embrace and compete for. Further, these regions are experiencing faster population and economic growth than the U.S., and these two trends are expected to continue in the next few years at the very least. Netflix has the opportunity here to take advantage of these two, rare phenomena and propel the company to new heights and further success.