Netflix Userbase Analysis

By Daryl Sangalang

Objective

Utilising Data Analytic tools such as Google Sheets, SQL, and Tableau, we're to explore and analyse the data set **Netflix Userbase** available on Kaggle in order to draw valuable insights and form a data-driven, and impactful business suggestions to assist the business.

Processing

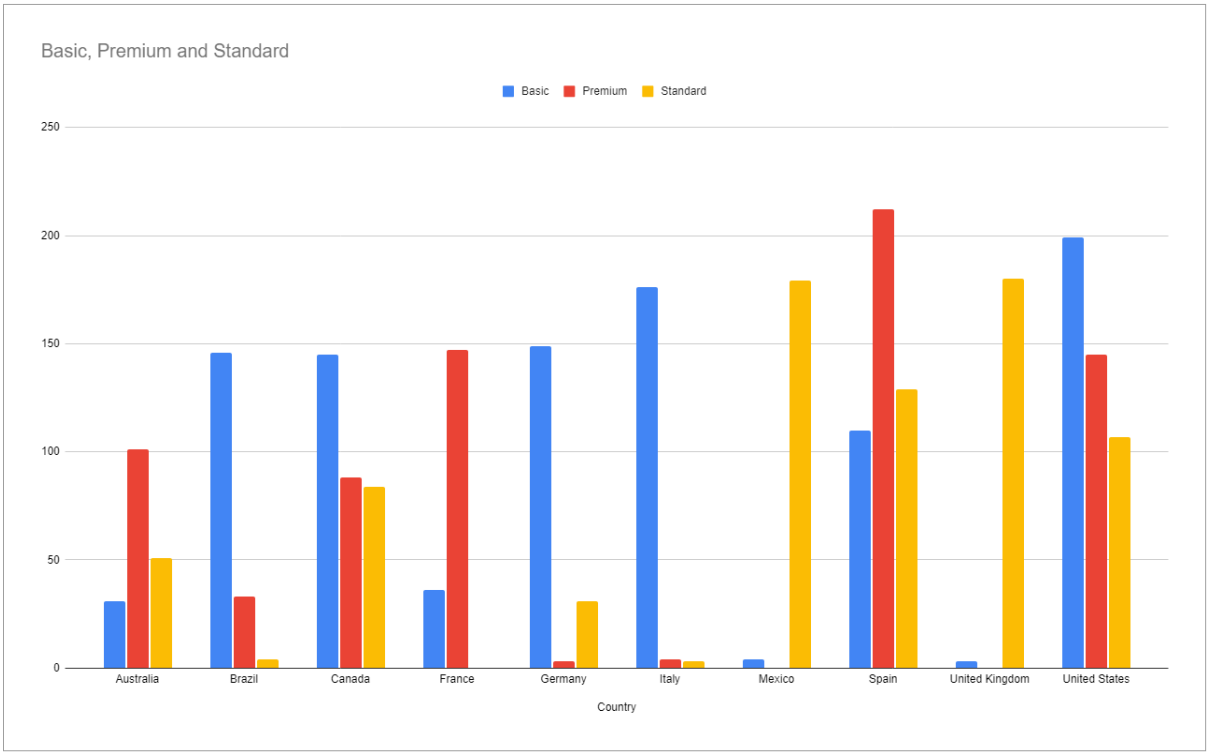
The **Netflix Userbase** dataset is relatively small and neat so it required very little cleaning. Utilising **Google Sheets** for cleaning:

- Removed the "Join_date" column as information stored seems unusable
    - All dates are the 15th day of January
        - Confirmed by using *distinct LEFT(Join_date, 3)*, which returned "15-01".
    - Year value seems random.
- Removed "Last_Payment_Date" column, some of the years are in the future, unsure if the months and days are trustworthy.
- Removed "Plan_Duration" column as it only contains a single value for every entry (1 month).

Making use of Pivot Tables and drawing Bar and Line graphs from the table data, I was able to conduct a preliminary analysis of the relationship of each column, Opening up ideas for starting points for when I conduct a deep-dive analysis using powerful visualisation tools such as Tableau public.

Observation

Google Sheet - Pivot Table



- Certain countries have recorded significantly more subscriptions than others over all.
- United Kingdom and Mexico have not recorded any Premium subscription.
- Certain countries record more Premium Subscriptions than subscription type, or at the least more than Standard Subscription.
- Australia, France, Mexico, and UK has some of the lowest recorded Basic subscription type amongst the countries included.

Tableau

Dashboard link:
https://public.tableau.com/app/profile/daryl.sangalang/viz/Netflix_Userbase_analysis/Dashboard1?publish=yes

Age to Gender Distribution

	Gender1		
Age Range	Female	Male	Grand To..
20 - 29	143	164	307
30 - 39	494	526	1,020
40-49	529	467	996
50+	91	86	177

Analysing the charts we can draw some observations:

- Significantly lower subscribers for 20-29 and 50+ age range in comparison to the rest.
- The majority of older demographic (40+) of subscribers are Females and the 39 and below are majority Males.

Conclusion

To increase the platform's reach and appeal to the minority in each demographic, we can target acquisition of material that appeal to their taste:

Age group specific:

- Introduce more newer romance / RomCom / feminine material to the platform (Movies, Series, Documentaries etc…) to entice more of the younger Female population to subscribe.
- Introduce more classic action / thriller / masculine material to the platform to invite more of the older Male population to subscribe.

Country vs SubType table

	Subscription Type		
Country1	Basic	Standard	Premium
Australia	31	51	101
Brazil	146	4	33
Canada	145	84	88
France	36		147
Germany	149	31	3
Italy	176	3	4
Mexico	4	179	
Spain	110	129	212
United Kingdom	3	180	
United States	199	107	145

Country specific:

- Run targeted promotions for certain subscription types for new users on the following countries:
    - **Australia** - Basic and Standard
    - **Brazil** - Standard and Premium
    - **Canada** - Standard and Premium
    - **France** - Basic and Standard
    - **Germany** - Standard and Premium
    - **Italy** - Standard and Premium
    - **Mexico** - Basic and Premium
    - **Spain** - Basic and Standard
    - **UK** - Basic and Premium
    - **US** - Standard and Premium

Areas for Improvement

There are several areas of improvement in regards to the data that, if addressed, can help substantiate and/or improve the Business Suggestion:

- Re-collect / fix the issue that has caused the “Join_date” column to be compromised and render the data unusable for analysis.
    - Having access to this data, we’ll be able to pinpoint any “hotspots” for subscriptions throughout the year and relate it to any possible causation for the spike in subscription (i.e, promo event, international holiday, specific shows being on-boarded on the platform etc…)
- Re-collect / fix the isse that has caused the “Last_Payment_Date " column to be similar to “Join_Date " column where it’s been compromised to the point of unusability.
    - Access to this information, we’ll be able to identify the “active” status of each subscription, with subscription period being 1 month, any last payment date over this duration can be deemed “inactive”.
    - With this new information, we can track the rate of users unsubscribing to the platform, pinpoint any causation (whether it be declined payments, manual exit etc…) and build a marketing strategy around such trends throughout the year, possibly even create a structure for similar “down” periods for the coming year(s).

With the “Join_date” and “Last_Payment_Date” columns issue being addressed, analysis around the influx in revenue throughout the year will also be possible.

This will open up opportunities to draw valuable insights and provide impactful business suggestions for the Sales and / or Marketing team.