**IMPACT of content type on subscriber count in netflix**

**TEAM 6:**

* Amit Singh
* Paroma Roy
* Kuldeepsinh Kiransinh Vaghela
* Oreoluwa Akinboboye

**COURSE CODE:** BIA 500-A

**INTRODUCTION**

Being the largest streaming service in the world, Netflix upholds an extraordinary subscriber base, among other remarkable bits and pieces. Let’s suppose it’s a Friday evening and you’re sipping your cup of coffee and scrolling through different titles/genres on Netflix, not realizing an awful lot of time spent on deciding which movie, TV show, documentary, reality TV, etc. you would like to watch. For me at least, I usually search lots of titles by genre and read their description before jumping onto the content. Given the tremendous amount of content choices on Netflix, don’t we all spend a lot of time just figuring out which content to watch? While you and I keep searching for movies, TV shows, documentaries, etc., still unsure which content to pick, our team wondered if the whole content team of Netflix addresses a similar question every day: ***Which content would I pick for increasing my subscriber base?***

Based on the recent statistics, Netflix lost about 200,000 subscribers during the first quarter of this year; reasons being shifting the economic forces, stern competition from other streaming platforms, and the conflict in Ukraine. One of the key reasons was Russia’s invasion of Ukraine which led to Netflix shutting down its service in Russia, ultimately reversing the subscriber growth by a huge amount. Taking these factors into account, the streaming platform would rethink its approach in terms of content strategy.

Amidst all such scenarios and the world being a content-driven market, Netflix needs exhilarating content which attracts the attention of the audience and subsequently influences them to subscribe. Content-type would be one of the most unarguably important tactical pedals to increase the subscriber base of Netflix. So, what content should Netflix be adding to increase its subscriber count?Concerned about this question, our team is evaluating if there’s any relationship between the type of content and the subscriber count of Netflix.

**DATASET EXPLANATION**

Our analysis data was collected as of Q4 of 2021 and is based on the most up-to-date figures available. They are based on the list of genres that Netflix has on its platform. It serves as a tactic to show how Netflix interconnects with the type of content. Using this very type of genre, our team will decode its relationship with the subscriber base.

Our dataset is sourced from the following websites. Since Netflix rarely reveals the number of subscribers outside of the US, we had to dig deep and investigate a bunch of sources like **Comparitech** - With the help of comparitech, we were able to retrieve the total number of subscribers for a majority of countries, **Start.io** **–** It served us in gaining the total number of subscribers for some Latin American and Asian countries, **Flixwatch -** Through Flixwatch, we were able to get hold of the content-wise count in a particular country. It gives us the total number of movies, TV shows, documentaries, reality TV shows, stand-up shows, and original shows for each country mentioned in our dataset.

In summary, the dataset lists the total number of Netflix subscribers in 88 countries. Alongside, this, it also shows the total number of movies, TV shows, documentaries, reality TV shows, stand-up shows, and original shows for each of these countries.

**Key variables:**

* Total - Overall number of content types in a particular country
* Movie - Total number of movies in a particular country
* TV shows - Total number of TV shows in a particular country
* Documentary - Total number of documentaries in a particular country
* Reality TV - Total number of Reality TV shows in a particular country
* Stand-up - Total number of stand-ups in a particular country
* Originals - Total number of original shows in a particular country
* Number of Subscribers - Total number of subscribers in a particular country

All our key variables are of the Ratio data type. Their metrics are based on a continuous scale of measurement. Every variable is a form of quantitative/numeric data.

**DESCRIPTIVE STATISTICS**

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Descriptive measure** | **Total** | **Movie** | **TV shows** | **Documentary** | **Reality TV** | **Stand-up** | **Originals** | **Number of Subscribers** |
| **Mean** | 5546.27 | 2919.511 | 1500.2 | 638.7840909 | 162.99 | 320.58 | 2172.25 | 2412957.74 |
| **Standard Error** | 51.3101 | 36.28221 | 14.482 | 3.849827612 | 0.6132 | 3.3374 | 3.770529 | 840839.843 |
| **Median** | 5492 | 2875.5 | 1514 | 642 | 162.5 | 316.5 | 2175 | 356137 |
| **Mode** | 5388 | 3279 | 1380 | 648 | 161 | 348 | 2173 | #N/A |
| **Standard Deviation** | 481.332 | 340.3573 | 135.86 | 36.1145842 | 5.7525 | 31.307 | 35.3707 | 7887776.9 |
| **Sample Variance** | 231680 | 115843.1 | 18457 | 1304.263192 | 33.092 | 980.15 | 1251.086 | 6.22E+13 |
| **Kurtosis** | -0.57555 | -0.12505 | -0.7933 | 0.024261038 | 4.2491 | -1.347 | -0.76835 | 61.8324143 |
| **Skewness** | 0.18203 | 0.307744 | 0.1684 | -0.147100761 | 1.4711 | -0.313 | -0.12691 | 7.40643093 |
| **Range** | 2035 | 1519 | 603 | 193 | 33 | 108 | 146 | 69535654 |
| **Minimum** | 4555 | 2192 | 1186 | 538 | 152 | 262 | 2083 | 13019 |
| **Maximum** | 6590 | 3711 | 1789 | 731 | 185 | 370 | 2229 | 69548673 |
| **Sum** | 488072 | 256917 | 132019 | 56213 | 14343 | 28211 | 191158 | 212340281 |
| **Count** | 88 | 88 | 88 | 88 | 88 | 88 | 88 | 88 |

* The data that we have selected, contains 88 samples. The first row indicates the sample mean of each variable, where movies have the highest mean and Reality-TV has the least.
* The median and the mean are seen to be having similar values which indicate that the data is symmetrical
* The standard error indicates how accurate the sample mean is to the population data.
* The standard deviation measures the amount of dispersion between the values and their means, and we see that Reality-TV has the least dispersion and originals has the highest.
* For kurtosis, we see that movies, TV shows, stand-up, and originals show that the peakiness is somewhat flat as |CK|<0, whereas for documentary, reality TV and showed peaky curves as |CK|>0
* For skewness, a high degree of skewness for reality TV was seen as |CS| > 1​, Relative symmetrical data for every other variable like movie and TV-show as |CS| < 0.5​, Documentary, stand-up and originals show negative skewness whereas the rest of them are positively skewed​.

**Descriptive statistics inference**

* Using the **VLOOKUP** functions we found out the maximum and minimum number of releases for each content type -

|  |  |  |
| --- | --- | --- |
| Content-Type​ | Countries with the maximum number of the specified content ​ | Countries with a minimum number of the specified content ​ |
| Total#​ | Japan​ | Finland​ |
| Movies​ | Japan​ | Georgia​ |
| TV shows​ | USA​ | Peru​ |
| Documentary​ | USA​ | South Korea​ |
| Reality TV​ | Singapore​ | Sweden​ |
| Stand-up​ | Luxembourg​ | UAE​ |
| Originals​ | USA​ | Taiwan​ |

* Using **Mode** from descriptive statistics we found how many countries have the same number of releases for each content type –

|  |  |
| --- | --- |
| **Content-Type​** | **# of countries with the same number of the specified content ​** |
| **Total#**​ | 2 (Israel, Poland)​ |
| **Movie**​ | 2 (Australia, Mongolia)​ |
| **TV shows**​ | 3 (Guatemala, Venezuela, and Uruguay)​ |
| **Documentary**​ | 5 (Bangladesh, Costa Rica, Guatemala, Venezuela, and Uruguay)​ |
| **Reality TV**​ | 12​ |
| **Stand-up**​ | 14​ |
| **Originals**​ | 4 (Cyprus, Indonesia, Ukraine, and Kyrgyzstan)​ |

**VISUALIZATIONS:** We have used scatterplots to make comparisons and show the trend in the data analyzed and we will discuss the various content types below according to their quantum on Netflix starting from the highest to the lowest. For example, there are more movies on Netflix and the next highest content type is TV shows, followed by documentaries, stand-up comedies, and reality TV.

Chart, scatter chart

Description automatically generated

Movies: There is an overall progressive decrease in the number of subscriptions on Netflix as movie content are increases. This indicates an inverse relationship between movie content and the number of subscriptions. However, we note significant outliers in this trend. This can be explained by how interesting the movies introduced are. For example, in periods where there is a huge spike in a subscription where movie content sits at about 2,600. This may be due to the release of a blockbuster on Netflix. An example of this will be putting a movie like ‘Coming to America’ on Netflix.

**Chart, scatter chart

Description automatically generated**

TV-Shows: There is a mostly stable and slightly progressive or positive relationship between the level of TV shows on Netflix and the number of subscribers. Again, there are very significant outliers and this may be due to the popularity of a particular content introduced on Netflix. For example, there is a spike in subscription levels where the quantum of TV shows sits at about 1600. This may be due to the introduction of popular TV shows on Netflix like ‘The Oprah Winfrey Show’.

Chart, scatter chart

Description automatically generated

Documentary: Again, we see significant outliers and this can be explained based on how popular the documentary content is. For example, where there is a spike in subscriptions with documentary content at a level of about 604. This may be due to the introduction of the popular documentary on Netflix based on scandals like the ‘Tinder Swindler’ or celebrity content like the ‘Oprah + Viola – Netflix special’

**Chart, scatter chart

Description automatically generated**

Reality TV: There is an overall progressive increase in the level of subscription as the reality TV content level increases. This indicates a positive relationship. Again, we note significant spikes and outliers. This may be due to the introduction of very popular reality TV content on Netflix. For example, there is a spike in the level of subscriptions where reality TV content sits at about 162. This may be due to the introduction of popular reality TV shows like ‘Keeping up with the Kardashians’ on Netflix.

**ANALYSIS**

First, we did the multiple regression considering all the variables, and below is the result that we found:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Regression Statistics** | | | | |
| R Square  Adjusted R Square | | | 0.375486752  0.329226511 | |
| **Results** | | | | |
|  | **Coefficients** | **Standard Error** | **t Stat** | **P-value** |
| Intercept | -120460611.7 | 66736439.76 | -1.805020049 | 0.074787609 |
| Movie | -13602.58315 | 3290.375758 | -4.134051596 | 8.6403E-05 |
| TV Shows | -388.0022561 | 9985.62372 | -0.038856086 | 0.969100737 |
| Originals | -10745.0445 | 34040.3197 | -0.315656392 | 0.753075205 |
| Documentary | 178804.8294 | 47057.36486 | 3.799720405 | 0.000279129 |
| Stand-up | -135556.1559 | 53268.3782 | -2.544777229 | 0.01283328 |
| Reality TV | 710161.939 | 159565.339 | 4.450602765 | 2.70718E-05 |

So, the adjusted R square value that we got is 0.33 and only TV Shows and Originals are insignificant as their P-value is greater than the significant value which is 0.05

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Regression Statistics** | | | | |
| R Square  Adjusted R Square | | | 0.264170155  0.219302482 | |
| **Results** | | | | |
|  | **Coefficients** | **Standard Error** | **t Stat** | **P-value** |
| Intercept | -191543327.5 | 69110792.89 | -2.77154 | 0.006902 |
| Movie | -6783.681833 | 2975.31682 | -2.27999 | 0.025205 |
| TV Shows | -5129.455017 | 10688.35646 | -0.47991 | 0.632569 |
| Originals | 43287.70507 | 33365.88487 | 1.297364 | 0.198144 |
| Reality TV | 841429.7603 | 168061.0492 | 5.006691 | 3.12E-06 |
| Stand-up | -30315.70561 | 49088.61932 | -0.61757 | 0.538569 |

Then we removed the Documentary which had a strong correlation with stand-up and did the regression again, below is the result that we received:

So, on removing the Documentary the adjusted R square value dropped to 0.22, and the stand-up which was signed earlier became insignificant. TV Shows and Originals remained insignificant. Then we removed the variables that had a high P-value compared to the significant value, which is 0.05, one at a time, and below is the final result we got:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Regression Statistics** | | | | |
| R Square  Adjusted R Square | | | 0.227892403  0.209725165 | |
| **Results** | | | | |
|  | **Coefficients** | **Standard Error** | **t Stat** | **P-value** |
| Intercept | -103262044.7 | 22334182.37 | -4.6235 | 1.34E-05 |
| Movie | -7948.85791 | 2670.329375 | -2.97673 | 0.003794 |
| Reality TV | 790740.9147 | 157993.6294 | 5.004891 | 2.98E-06 |

So, the adjusted R square dropped to 0.21 and both Movie and Reality TV became significant. This is the final regression model we received when we removed Documentary to remove multicollinearity.

Next, we removed Stand-up and kept all other variables, and did the regression, below is the result we received:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Regression Statistics** | | | | |
| R Square  Adjusted R Square | | | 0.325557371  0.284432821 | |
| **Results** | | | | |
|  | **Coefficients** | **Standard Error** | **t Stat** | **P-value** |
| Intercept | -46476139.07 | 62044198.2 | -0.74908 | 0.455952 |
| Movie | -10208.8265 | 3106.808721 | -3.28595 | 0.001497 |
| TV Shows | -17713.46077 | 7544.792693 | -2.34777 | 0.021296 |
| Documentary | 116540.5516 | 41516.69307 | 2.807077 | 0.006245 |
| Reality TV | 767006.1178 | 163184.1897 | 4.700248 | 1.04E-05 |
| Originals | -43360.40238 | 32571.28964 | -1.33125 | 0.186797 |

So, when we removed Stand-up the adjusted R square value is more compared to what we got when we removed Documentary. Also, only Originals is insignificant variable rest are significant. So, we can say that this is a better model compared to the earlier one. Next, we removed Originals and did the regression again and below is the result we received:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Regression Statistics** | | | | |
| R Square  Adjusted R Square | | | 0.310981052  0.27777532 | |
| **Results** | | | | |
|  | **Coefficients** | **Standard Error** | **t Stat** | **P-value** |
| Intercept | -123582309.6 | 22345568.79 | -5.53051 | 3.62E-07 |
| Movie | -9003.253505 | 2985.686934 | -3.01547 | 0.003404 |
| TV Shows | -14556.15247 | 7195.570895 | -2.02293 | 0.046299 |
| Documentary | 73691.33876 | 26344.09201 | 2.797262 | 0.006404 |
| Reality TV | 779470.8131 | 163671.4601 | 4.762411 | 8.02E-06 |

Above is the final regression model with all the variables having P-values less than 0.05 (at a 95% significance level) and with an adjusted R square value of 0.28 which states that 28% of the variation in subscribers count of NETFLIX can be explained by these variables. The final equation we got is:

***Subscribers = -123582309.6 - 9003.253505(Movie) - 14556.15247(TV Shows) + 73691.33876(Documentary) + 779470.8131(Reality TV)***

**RESULTS AND RECOMMENDATIONS**

The P-value of Movie, TV Shows, Documentary, and Reality TV is less than 0.05 (at a 95% significance level) so they are the variables that are statistically significant to explain the change in subscriber count of NETFLIX. The coefficients of Movie and TV Shows are negative which implies a negative correlation between them and the subscriber count of NETFLIX. On the other hand, the coefficients of Documentary and Reality TV are positive which implies a positive correlation between them and the subscriber count of NETFLIX.

From our analysis, we concluded that adding a Movie or TV Show hurts the subscriber count of the NETFLIX maybe because after the covid pandemic people are more interested in going out and watching a movie in a movie theater rather than at home on a small screen. Seems like people are less interested in TV Shows and more interested in Documentaries and Reality TV shows.

Most of the Documentaries and Reality TV shows published by NETFLIX are owned by them which means they do not have to pay any royalty to story writers or producers and from our analysis, we found that both Documentaries and Reality TV have a positive effect on the subscriber count so our recommendation to NETFLIX is that they can focus more on Documentaries and Reality TV to increase their revenue and subscriber base.

**References:**

1. **Comparitech** - It’s a pro-consumer website that deals with in-depth tech research to provide users with information, reviews, and comparisons on a wide range of topics. The website is <https://www.comparitech.com/tv-streaming/netflix-subscribers/>. With the help of comparitech, we were able to retrieve the total number of subscribers for a majority of countries.
2. **start.io** **-** It’s a use case website that focuses on consumer insights and customer data analytics, with access to loads of user segments and target market demographics. The website is <https://www.start.io/audience/>. start.io served us in gaining the total number of subscribers for some Latin American and Asian countries.
3. **Flixwatch -** The worldwide stats by content type are extracted from flixwatch. It serves the purpose of a global search engine for Netflix. The website is <https://www.flixwatch.co/statistics/worldwide/content-type/>. Through Flixwatch, we were able to get hold of the content-wise count in a particular country. It gives us the total number of movies, TV shows, documentaries, reality TV shows, stand-up shows, and original shows for each country mentioned in our dataset.