IMPACT OF CONTENT TYPE ON THE SUBSCRIBER COUNT

Presented by – Team 6
Amit Singh
Paroma Roy
Kuldeepsinh Vaghela
Oreoluwa Akinboboye

NETFLIX

Problem Statement

For a streaming platform such as Netflix, content type would be one of the most important strategic levers to increase its paid subscriber base. ***So, what content should Netflix be adding to increase its subscriber base?*** Concerned about this question, our team is evaluating if there's any relationship between the type of content and the subscriber count of Netflix.

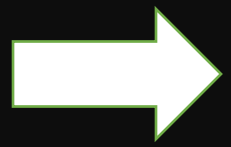
Dataset overview

The dataset used in this analysis was collected as of Q4 of 2021 and is based on the most up-to-date figures available.

It's sourced from the website of Comparitech i.e., https://www.comparitech.com/tv-streaming/netflix-subscribers/ which deals in in-depth tech research, and start.io i.e., https://www.start.io/audience/ which focusses on consumer insights and customer data analytics. The worldwide stats by content type is extracted from the website of Flixwatch i.e., https://www.flixwatch.co/statistics/worldwide/content-type/ which serves the purpose of a global search engine for Netflix.

Key variables

- Total
- Movie
- TV shows
- Documentary
- Reality TV
- Stand-up
- Originals
- Number of Subscribers



- Sample size - 88
- Type - Ratio Data

Dataset overview cont.

Countries	Total	Movie	TV shows	Documentary	Reality TV	Stand-up	Originals	Number of Subscribers
Argentina	5023	2571	1345	634	158	311	2218	5240340
Australia	6150	3279	1649	693	167	357	2202	6622078
Austria	5260	2772	1415	606	167	296	2136	961399
Bangladesh	5508	2703	1647	648	158	347	2186	256421
Belgium	5580	3060	1437	610	157	312	2145	1699092
Brazil	5026	2593	1339	642	161	287	2193	18836193
Bulgaria	6237	3461	1575	684	164	348	2192	177887
Canada	5917	3159	1533	698	163	359	2221	6895241
Chile	5228	2753	1357	642	160	312	2219	978172
Colombia	5310	2811	1375	647	161	312	2222	2823517
Costa Rica	5331	2827	1379	648	161	312	2222	307825
Croatia	6215	3449	1569	680	164	348	2174	119162
Cyprus	5897	3108	1620	653	162	349	2173	51003
Czech Republic	6075	3341	1546	671	164	348	2189	472210
Denmark	4912	2556	1311	600	158	285	2146	951883
Estonia	5049	2400	1514	628	157	345	2178	54382
Finland	4555	2251	1293	577	152	280	2142	762385
France	5278	2801	1412	611	165	286	2137	8650354
Germany	5460	2947	1429	615	168	297	2142	11057366
Greece	4820	2406	1316	616	155	322	2132	535372

Descriptive statistics

Descriptive measure	Total	Movie	TV shows	Documentary	Reality TV	Stand-up	Originals	Number of Subscribers
Mean	5546.272727	2919.511364	1500.215909	638.7840909	162.9886364	320.5795455	2172.25	2412957.739
Standard Error	51.31014127	36.28221191	14.48240969	3.849827612	0.613223815	3.337382138	3.77052897	840839.8431
Median	5492	2875.5	1514	642	162.5	316.5	2175	356137
Mode	5388	3279	1380	648	161	348	2173	#N/A
Standard Deviation	481.3317905	340.3573171	135.8570453	36.1145842	5.752549296	31.30741956	35.37069701	7887776.904
Sample Variance	231680.2926	115843.1033	18457.13676	1304.263192	33.09182341	980.1545193	1251.086207	6.2217E+13
Kurtosis	-0.575548532	-0.125054998	-0.793265234	0.024261038	4.249061825	-1.346516275	-0.768350605	61.83241429
Skewness	0.182028982	0.307744289	0.168407643	-0.147100761	1.471051124	-0.313058806	-0.126908287	7.406430931
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

Kurtosis

- CK < 0 for every content type except Reality TV and documentary. The number of subscribers also has CK > 0

Skewness

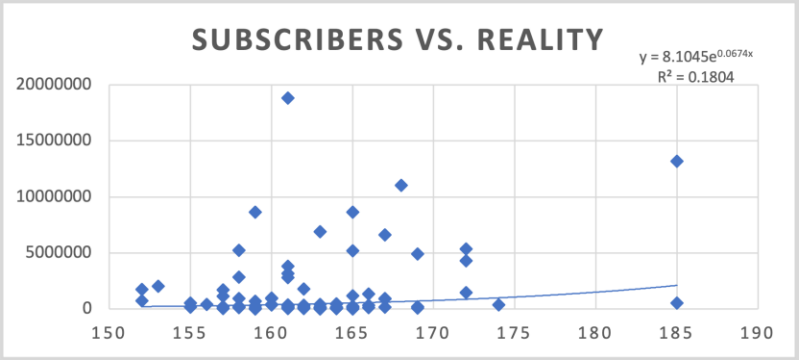
- High degree of skewness for number of subscribers and reality TV, wherein |CS| > 1
- Relative symmetrical data for every other variable, wherein |CS| < 0.5
- Documentary, stand-up and originals show negative skewness whereas the rest of them are positively skewed

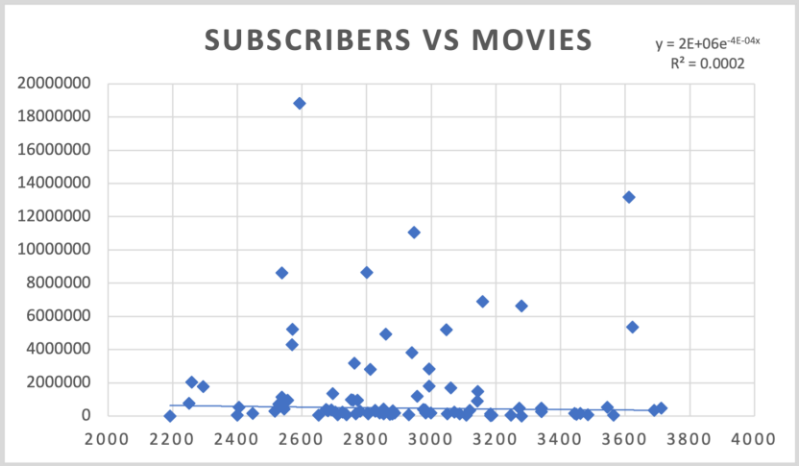
NETFLIX

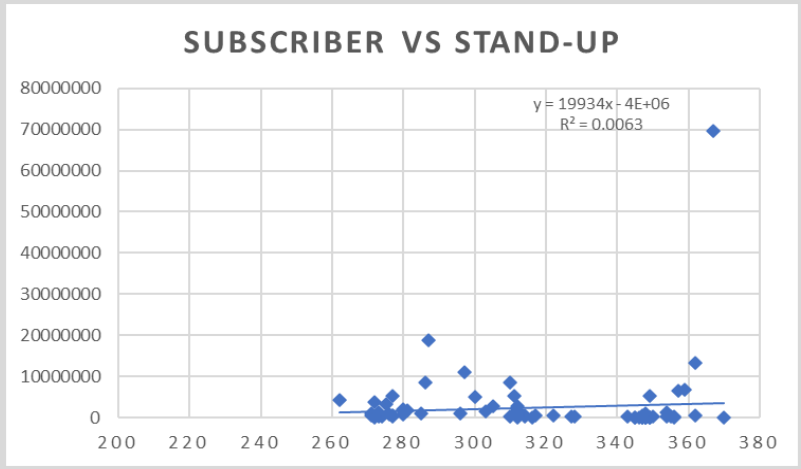
Descriptive statistics inferences

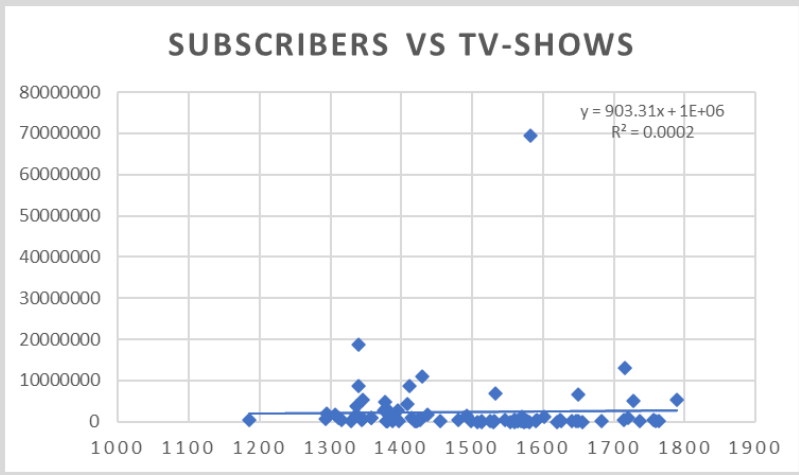
Content type	Countries with maximum number of the specified content	Countries with minimum number of the specified content
Total#	Japan	Finland
Movies	Japan	Georgia
TV shows	USA	Peru
Documentary	USA	South Korea
Reality TV	Singapore	Sweden
Stand-up	Luxembourg	UAE
Originals	USA	Taiwan

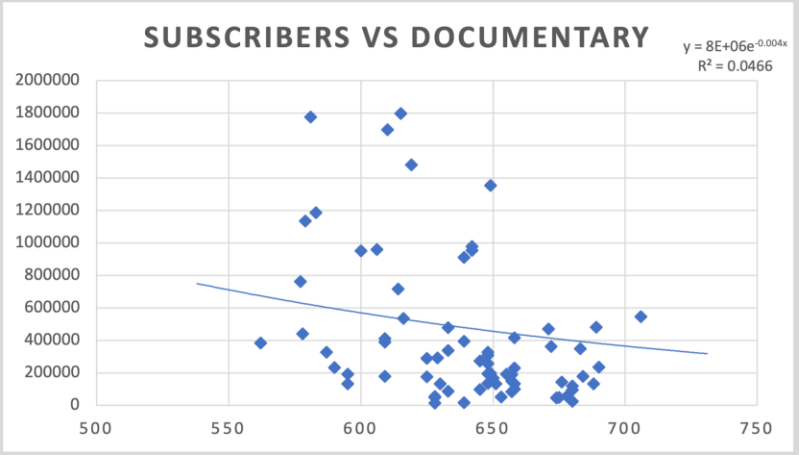
Content type	# of countries with same number of the specified content
Total#	2 (Israel, Poland)
Movie	2 (Australia, Mongolia)
TV shows	3 (Guatemala, Venezuela and Uruguay)
Documentary	5 (Bangladesh, Costa Rica, Guatemala, Venezuela and Uruguay)
Reality TV	12
Stand-up	14
Originals	4 (Cyprus, Indonesia,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











Visualizations

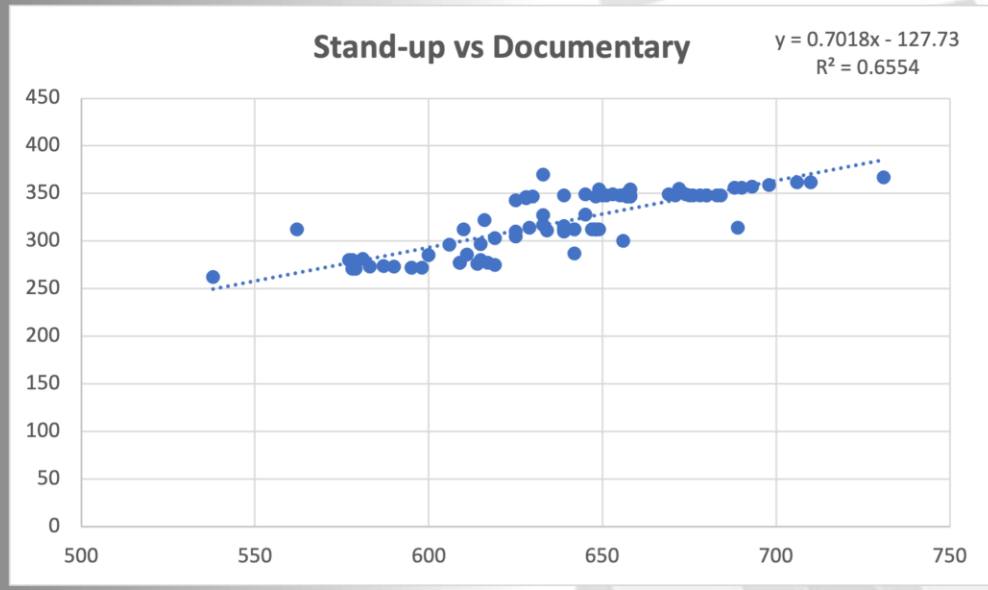
Analytics

- Conducted a Regression model analysis using the described key variables
  - Dependent variable – Subscribers
  - Predictors – Movie, TV shows, Documentary, Reality TV, Stand-up and Originals
- Made a correlation matrix
- Ran the Regression model after getting rid of multicollinearity

Findings Experiment 2

Correlation

Multicollinearity can be seen with Stand-up and Documentary



	Movie	TV shows	Originals	Documentary	Stand-up	Reality TV	Subscribers (Estimated)
Movie	1						
TV shows	0.602819985	1					
Originals	0.139425875	0.075828306	1				
Documentary	0.596474135	0.538173769	0.674726019	1			
Stand-up	0.372806371	0.653096904	0.620529042	0.809570923	1		
Reality TV	0.561981573	0.56094642	0.167621962	0.516964325	0.400586364	1	
Subscribers (Estimated)	-0.018905658	0.015558308	0.174797049	0.264626279	0.079121327	0.383930984	1

NETFLIX

Findings Experiment 1

Regression #1

Regression Statistics	
Multiple R	0.612769738
R Square	0.375486752
Adjusted R Square	0.329226511
Standard Error	6460149.424
Observations	88

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	6	2.03247E+15	3.38744E+14	8.116835251	7.05958E-07
Residual	81	3.38042E+15	4.17335E+13		
Total	87	5.41288E+15			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 95.0%*	*Upper 95.0%*
Intercept	-120460611.7	66736439.76	-1.805020049	0.074787609	-253245193.2	12323969.72	-253245193.2	12323969.72
Movie	-13602.58315	3290.375758	-4.134051596	8.6403E-05	-20149.3988	-7055.767508	-20149.3988	-7055.767508
TV shows	-388.0022561	9985.62372	-0.038856086	0.969100737	-20256.2612	19480.25669	-20256.2612	19480.25669
Originals	-10745.0445	34040.3197	-0.315656392	0.753075205	-78474.60306	56984.51406	-78474.60306	56984.51406
Documentary	178804.8294	47057.36486	3.799720405	0.000279129	85175.43417	272434.2247	85175.43417	272434.2247
Stand-up	-135556.1559	53268.3782	-2.544777229	0.01283328	-241543.5194	-29568.79228	-241543.5194	-29568.79228
Reality TV	710161.939	159565.339	4.450602765	2.70718E-05	392676.9663	1027646.912	392676.9663	1027646.912

Key Takeaways -

- ***Every predictor is considered***
- ***Multicollinearity exists***
- ***Adjusted R Square is 0.32***
- ***Increased p-values of TV shows and Originals***

NETFLIX

Findings Experiment 3

Regression #2

Regression Statistics	
Multiple R	0.513974859
R Square	0.264170155
Adjusted R Square	0.219302482
Standard Error	6969410.061
Observations	88

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	5	1.42992E+15	2.85984E+14	5.8877614	0.00010671
Residual	82	3.98296E+15	4.85727E+13		
Total	87	5.41288E+15			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 95.0%*	*Upper 95.0%*
Intercept	-191543327.5	69110792.89	-2.771540008	0.0069018	-329026710.3	-54059944.68	-329026710.3	-54059944.68
Movie	-6783.681833	2975.31682	-2.279986382	0.0252052	-12702.5348	-864.8288694	-12702.5348	-864.8288694
TV shows	-5129.455017	10688.35646	-0.479910549	0.6325687	-26392.00088	16133.09084	-26392.00088	16133.09084
Originals	43287.70507	33365.88487	1.297364216	0.1981441	-23087.66889	109663.079	-23087.66889	109663.079
Reality TV	841429.7603	168061.0492	5.006691107	3.118E-06	507102.7965	1175756.724	507102.7965	1175756.724
Stand-up	-30315.70561	49088.61932	-0.617570957	0.5385695	-127968.607	67337.19576	-127968.607	67337.19576

Key Takeaways -

- ***Getting rid of multicollinearity***
- ***Removed Documentary***
- ***Adjusted R Square changes to 0.21***
- ***Increased p-values of Originals, TV shows and Stand-up***

NETFLIX

Findings Experiment 3

Regression #3

Regression Statistics	
Multiple R	0.511960373
R Square	0.262103423
Adjusted R Square	0.226542143
Standard Error	6937019.998
Observations	88

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	4	1.41873E+15	3.54684E+14	7.370472	3.94939E-05
Residual	83	3.99415E+15	4.81222E+13		
Total	87	5.41288E+15			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 95.0%*	*Upper 95.0%*
Intercept	-208800999.6	58744074.26	-3.554418078	0.0006283	-325640600.6	-91961398.53	-325640600.6	-91961398.53
Movie	-7367.539296	2702.598123	-2.726095025	0.0078164	-12742.89826	-1992.180328	-12742.89826	-1992.180328
Originals	52559.20591	27077.42639	1.941070955	0.0556436	-1296.70613	106415.118	-1296.70613	106415.118
Reality TV	820714.4504	161668.37	5.076530742	2.318E-06	499162.5648	1142266.336	499162.5648	1142266.336
Stand-up	-47461.83087	33506.88791	-1.416479829	0.1603773	-114105.6833	19182.02154	-114105.6833	19182.02154

Key Takeaways -

- ***Regression #2 doesn't yield better results***
- ***Removed TV shows due to its highest p-value***
- ***Adjusted R Square changes to 0.22***
- ***Increased p-value of Stand-up***

NETFLIX

Findings Experiment 3

Regression #4

Regression Statistics	
Multiple R	0.494232486
R Square	0.244265751
Adjusted R Square	0.217275242
Standard Error	6978452.941
Observations	88

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	3	1.32218E+15	4.40727E+14	9.050061	2.93918E-05
Residual	84	4.0907E+15	4.86988E+13		
Total	87	5.41288E+15			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 95.0%*	*Upper 95.0%*
Intercept	-161862048.1	48794923.98	-3.317190291	0.0013451	-258896105.6	-64827990.57	-258896105.6	-64827990.57
Movie	-8147.998657	2661.639591	-3.061270461	0.0029595	-13440.96088	-2855.036436	-13440.96088	-2855.036436
Originals	28989.26927	21488.86271	1.349036925	0.1809505	-13743.69085	71722.22938	-13743.69085	71722.22938
Reality TV	767484.4279	158179.332	4.85198931	5.564E-06	452927.4792	1082041.377	452927.4792	1082041.377

Key Takeaways -

- ***Removed Stand-up due to its higher p-value***
- ***Adjusted R Square changes to 0.21***
- ***Increased p-value of Originals***

NETFLIX

Findings Experiment 3

Regression #5

Regression Statistics	
Multiple R	0.477380773
R Square	0.227892403
Adjusted R Square	0.209725165
Standard Error	7012028.861
Observations	88

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	2	1.23355E+15	6.16777E+14	12.544142	1.68387E-05
Residual	85	4.17933E+15	4.91685E+13		
Total	87	5.41288E+15			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 95.0%*	*Upper 95.0%*
Intercept	-103262044.7	22334182.37	-4.623497872	1.336E-05	-147668382.7	-58855706.62	-147668382.7	-58855706.62
Movie	-7948.85791	2670.329375	-2.976733127	0.0037941	-13258.18811	-2639.527711	-13258.18811	-2639.527711
Reality TV	790740.9147	157993.6294	5.00489113	2.983E-06	476607.2517	1104874.578	476607.2517	1104874.578

Key Takeaways -

- ***Removed Originals due to its higher p-value***
- ***Adjusted R Square changes to 0.20, which is comparatively less***
- ***P-values looks good***

NETFLIX

Findings Experiment 3

Regression #6

Regression Statistics	
Multiple R	0.57057635
R Square	0.325557371
Adjusted R Square	0.284432821
Standard Error	6672365.451
Observations	88

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	5	1.7622E+15	3.52441E+14	7.9163752	4.00383E-06
Residual	82	3.65068E+15	4.45205E+13		
Total	87	5.41288E+15			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 95.0%*	*Upper 95.0%*
Intercept	-46476139.07	62044198.2	-0.749081146	0.4559518	-169901813.9	76949535.78	-169901813.9	76949535.78
Movie	-10208.8265	3106.808721	-3.285952699	0.0014969	-16389.25875	-4028.39426	-16389.25875	-4028.394261
TV shows	-17713.46077	7544.792693	-2.347773025	0.0212957	-32722.45688	-2704.46465	-32722.45688	-2704.46465
Documentary	116540.5516	41516.69307	2.807076936	0.0062446	33950.62363	199130.4795	33950.62363	199130.4795
Reality TV	767006.1178	163184.1897	4.700247735	1.035E-05	442380.7811	1091631.454	442380.7811	1091631.454
Originals	-43360.40238	32571.28964	-1.331246102	0.1867974	-108155.0733	21434.26853	-108155.0733	21434.26853

Key Takeaways -

- ***Circling back to Regression #2 wherein Adjusted R Square was more, and multicollinearity was also handled***
- ***Added Documentary, removed Stand-up predictor instead***
- ***Adjusted R Square changes to 0.28***
- ***Increased p-value of Originals***

NETFLIX

Findings Experiment 3

Regression #7

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

Regression Statistics	
Multiple R	0.557656751
R Square	0.310981052
Adjusted R Square	0.27777532
Standard Error	6703332.797
Observations	88

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	4	1.6833E+15	4.20826E+14	9.3652821	2.69021E-06
Residual	83	3.72958E+15	4.49347E+13		
Total	87	5.41288E+15			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 95.0%*	*Upper 95.0%*
Intercept	-123582309.6	22345568.79	-5.530506328	3.62E-07	-168026747.2	-79137872	-168026747.2	-79137872
Movie	-9003.253505	2985.686934	-3.015471381	0.0034039	-14941.66473	-3064.84228	-14941.66473	-3064.842276
TV shows	-14556.15247	7195.570895	-2.022932257	0.0462992	-28867.85358	-244.451366	-28867.85358	-244.4513657
Documentary	73691.33876	26344.09201	2.797262427	0.0064041	21293.99931	126088.6782	21293.99931	126088.6782
Reality TV	779470.8131	163671.4601	4.762411311	8.02E-06	453934.8617	1105006.764	453934.8617	1105006.764

Key Takeaways -

- ***Removed Originals due to its higher p-value***
- ***Adjusted R Square changes to 0.27***
- ***P-values looks good***
- ***Perfect model***

NETFLIX

FINDINGS

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

- **Increasing the number of Documentary and Reality TV shows in a particular country would mark an increase in Netflix's subscriber base**
  - If documentary related content is added, Netflix will gain subscribers at an approximate rate of 73691
  - If Reality TV shows are added, Netflix will gain subscribers at an approximate rate of 779470
- **Number of movie and TV shows are indirectly proportional to the increasing subscriber count**
  - If movie related content is added, Netflix will lose subscribers at an approximate rate of 9003
  - If TV shows are added, Netflix will lose subscribers at an approximate rate of 14556

NETFLIX

CONCLUSION

- For NETFLIX adding a 'Documentary' or 'Reality TV' is profitable as it will increase the subscriber count
- For NETFLIX adding a 'TV Show' or 'Movie' is not profitable as it will decrease the subscriber count

QUESTIONS?

NETFLIX

THANKS