Can we predict your streaming habits?

By Conner Schultz

Overview and Important beginning Factors

- Y- Variable is the number of hours streamed per week within the United States.
- Independent Variables include
  - Numerical-
    - Age, Average Household Income, # Devices Used to Stream, # People in Household.
  - Categorical-
    - Race, Gender, Main device used to stream, Most Streamed Genres, Most Streamed Day, Most Streamed Time
- The project is important in determining the future of consumer streaming habits amongst rapid changes in the market.

A snapshot into the Data (Descriptive Stats)

Age		Average Household Income		Number of Streaming devices		People in Household		Streaming Hours per Week	
Mean	31.84269663	Mean	134987.14	Mean	2.794007491	Mean	1.9101124	Mean	18.082397
Standard Error	0.562319602	Standard Error	4462.3227	Standard Error	0.07388136	Standard Error	0.0533102	Standard Error	0.7565265
Median	30	Median	120000	Median	3	Median	2	Median	16
Mode	26	Mode	70000	Mode	2	Mode	2	Mode	10
Standard Deviation	9.188378009	Standard Deviation	72914.953	Standard Devia	1.207231363	Standard Devi.	0.8710954	Standard Devi.	12.361745
Sample Variance	84.42629045	Sample Variance	5.317E+09	Sample Varian	1.457407564	Sample Varian	0.7588071	Sample Varian	152.81273
Kurtosis	0.305588217	Kurtosis	-0.473311	Kurtosis	-0.886206505	Kurtosis	1.5300377	Kurtosis	0.1191415
Skewness	0.801219875	Skewness	0.5207795	Skewness	0.247373236	Skewness	1.1038135	Skewness	0.6946482
Range	53	Range	349470	Range	4	Range	4	Range	63
Minimum	18	Minimum	530	Minimum	1	Minimum	1	Minimum	1
Maximum	71	Maximum	350000	Maximum	5	Maximum	5	Maximum	64
Sum	8502	Sum	36041567	Sum	746	Sum	510	Sum	4828
Count	267	Count	267	Count	267	Count	267	Count	267
	1		1		1		1		1

	# Streamed hrs/we	DC-S1	# of devices	Race (%)	Genre (%)	Platform (%)	t day stream	Age	usehold Inco	Time (%)	ppl in household (#)
# Streamed hrs/wee	1										
DC-S1	0.002635949	1									
# of devices	0.118828027	-0.112125875	1								
Race (%)	-0.112351757	0.00249066	-0.156226	1							
Genre (%)	-0.029140184	-0.280980074	0.0619669	0.118168607	1						
Platform (%)	-0.034973371	0.03713026	-0.060227	0.116454441	0.073093849	1					
What day streaming	-0.007772749	0.062406489	-0.092126	0.138731606	0.145568682	-0.009941754	1				
Age	0.098262297	0.11323795	0.1534778	-0.207154591	-0.060192584	0.046534601	-0.025471	1			
Household Income	0.165472658	-0.017306268	0.3883205	-0.25917436	-0.059471481	-0.169664348	-0.103391	0.290190727	1		
Time (%)	-0.029832224	0.021280209	-0.121457	-0.05679938	-0.031302585	0.012092149	-0.00447	0.037251126	-0.118026	1	
ppl in household (#)	0.031996058	-0.13905419	0.1341951	-0.13882076	-0.018768885	-0.030292246	-0.010766	0.014375911	0.128036	-0.056	1

Beginning Regression Model

Regression Statistics	
Multiple R	0.834551866
R Square	0.696476817
Adjusted R Squ.	0.68195655
Standard Error	12.36799971
Observations	267

ANOVA

	df	SS	MS	F	Significance F
Regression	10	90208.374	9020.8374	58.97228	1.075E-60
Residual	257	39312.626	152.96742		
Total	267	129521			

	Coefficients	Standard Erro	t Stat	P-value	Lower 95%	Upper 95%	Lower 95.0%	Upper 95.0%
Intercept	0	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A
DC-S1	0.658978653	1.6032603	0.4110241	0.6813974	-2.4982216	3.8161789	-2.4982216	3.8161789
# of devices	0.949169247	0.6829586	1.3897904	0.1657952	-0.3957384	2.2940768	-0.3957384	2.2940768
Race (%)	0.610224604	7.7394522	0.078846	0.9372165	-14.630595	15.851044	-14.630595	15.851044
Genre (%)	2.23479398	9.7210662	0.2298919	0.8183587	-16.908294	21.377882	-16.908294	21.377882
Platform (%)	4.369566667	10.195316	0.4285857	0.6685837	-15.707431	24.446565	-15.707431	24.446565
What day stream	3.566337779	4.2929659	0.8307398	0.4068917	-4.8875316	12.020207	-4.8875316	12.020207
Age	0.123029602	0.0838402	1.4674292	0.143482	-0.0420717	0.2881309	-0.0420717	0.2881309
Household Inco	2.37668E-05	1.193E-05	1.9921243	0.0474154	2.73E-07	4.726E-05	2.73E-07	4.726E-05
Time (%)	3.927928303	4.2405745	0.9262727	0.3551733	-4.42277	12.278627	-4.42277	12.278627
ppl in household	5.206830891	4.5656658	1.1404319	0.2551681	-3.7840495	14.197711	-3.7840495	14.197711

Final Regression Model

SUMMARY OUTPUT								
Regression Statistics								
Multiple R	0.82975861							
R Square	0.68849935							
Adjusted R Square	0.68235161							
Standard Error	12.3622526							
Observations	267							
ANOVA								
	df	*SS*	*MS*	*F*	*Significance F*			
Regression	3	89175.1237	29725.0412	194.503419	1.8778E-66			
Residual	264	40345.8763	152.825289					
Total	267	129521						
	Coefficients	Standard Error	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 95.0%*	*Upper 95.0%*
Intercept	0	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A
Age	0.23956533	0.06787983	3.52925648	0.00049146	0.10591059	0.37322007	0.10591059	0.37322007
Household Income	2.774E-05	1.0917E-05	2.54100202	0.0116265	6.2446E-06	4.9235E-05	6.2446E-06	4.9235E-05
Number of People in Household	11.740431	3.66523976	3.20318226	0.00152594	4.52360872	18.9572532	4.52360872	18.9572532

Interpreting the Prediction Equation

● Based on our final regression model, our prediction equation is below:

Hours Streamed = 0.239(Age) + 0.02(Household income in thousands) + 11.74(Number of People in Household)

Questions?