Netflix TV Shows: Renewal or Cancellation?

Alison Yao (yy2564)

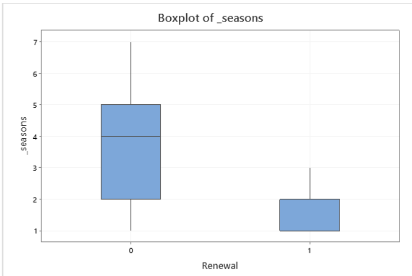
Whether a TV show gets renewed or cancelled is a big decision for the production company and an even bigger decision for the avid viewers. Continuing the topic of television streaming services from the last homework, I will focus on Netflix this time and predict whether a show will be renewed or cancelled given the number of seasons streamed, IMDb ratings and IMDb popularity.

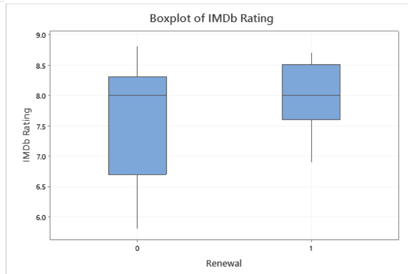
I obtained a retrospective sample of 15 cancelled shows and 15 renewed shows in 2021. The response is whether or not the Netflix show will be renewed after the latest released season; the three numerical predictors are the number of seasons released (**#seasons**), **IMDb Rating** and IMDb **Popularity** (by May 9, 2021). The first few observations are shown below:

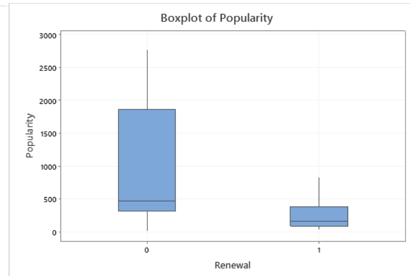
Title	#seasons	IMDb Rating	Popularity	Renewal
Bridgerton	1	7.2	60	1
Cobra Kai	3	8.6	117	1
Elite	3	7.6	286	1
Fate: The Winx Saga	1	6.9	359	1
Locke & Key	1	7.4	385	1

All data are collected from IMDb according to this list by Rotten Tomatoes: https://editorial.rottentomatoes.com/article/renewed-and-cancelled-tv-shows-2021/. The list consists of the renewal information of 452 shows from traditional cable television and streaming platforms, 74 of which are from Netflix. The number of shows being cancelled is much lower than those renewed. To keep the number of shows renewed and cancelled relatively balanced, I chose 15 cancelled shows and 15 renewed shows. The list does not always give me the number of seasons I want. It shows the number of seasons a show is announced to have. For example, Stranger Things have 3 seasons released at this moment and season 4 is renewed, then I will put 3 as the number of seasons instead of 4. For the IMDb rating and popularity, I searched the names of these shows on IMDb one by one and copy-pasted these two values. Popularity is the weekly rank of most viewed page among all TV show pages, so the lower the number is, the more popular a show is. I would like to use viewership data if possible because how many views a show generates is generally considered a key indicator, but I couldn't find any free resources online. And there is a chance that viewership would define my response, so I am happy with my current predictors.

First, let's take a look at the data:







There does not seem to be a distinctive split between renewal and cancellation in IMDb Rating versus response, but we can see that the TV shows that are renewed have fewer seasons, higher IMDb rating and rank higher in Popularity in general. But there is no guarantee that shows which check these three boxes are going to be renewed. Also, the boxplot of Popularity indicates taking natural log, but after trying it out, there is not much difference in taking log or not. So, I will stick to the one without logging. Here are the results from Minitab:

Binary Logistic Regression: Renewal versus _seasons, IMDb Rating, Popularity

* WARNING * When the data are in the Response/Frequency format, the Residuals versus fits plot is unavailable.

Method

Link function	Logit
Residuals for diagnostics	Pearson
Rows used	30

Response Information

Variable	Value	Count
Renewal	1	15 (Event)
	0	15
	Total	30

Regression Equation

P(1) = exp(Y')/(1 + exp(Y'))

Y' = -22.6 - 4.10 _seasons + 4.56 IMDb Rating - 0.00646 Popularity

Coefficients

Term	Coef	SE Coef	Z-Value	P-Value	VIF
Constant	-22.6	13.1	-1.73	0.084	
_seasons	-4.10	1.98	-2.07	0.038	5.06
IMDb Rating	4.56	2.42	1.88	0.060	5.66
Popularity	-0.00646	0.00643	-1.01	0.315	1.52

Odds Ratios for Continuous Predictors

	Odds Ratio	95% CI
_seasons	0.0166	(0.0003, 0.8032)
IMDb Rating	95.2254	(0.8305, 10918.0697)
Popularity	0.9936	(0.9811, 1.0062)

Model Summary

Deviance R-Sq	Deviance R-Sq(adj)	AIC	AICc	BIC	Area Under ROC Curve
81.95%	74.73%	15.51	17.11	21.11	0.9822

Goodness-of-Fit Tests

Test	DF	Chi-Square	P-Value
Deviance	26	7.51	1.000
Pearson	26	10.07	0.998
Hosmer-Lemeshow	8	7.21	0.515

Analysis of Variance

				Likelihood Ratio	
Source	DF	Adj Dev	Adj Mean	Chi-Square	P-Value
Regression	3	34.080	11.3600	34.08	0.000
_seasons	1	25.538	25.5379	25.54	0.000
IMDb Rating	1	9.429	9.4287	9.43	0.002
Popularity	1	5.845	5.8448	5.84	0.016
Error	26	7.509	0.2888		
Total	29	41.589			

Checking the likelihood ratio tests, the Chi-Square of the regression is 34.08, with a p-value zero out to three digits, so we strongly reject the null hypothesis that there is no relationship, which indicates the strength of the relationship is strong. #seasons and IMDb Rating are both highly statistically significant; Popularity is slightly weaker, but is still statistically significant.

The coefficient of #seasons means that holding everything else fixed, adding one season to the existing number of seasons is associated with multiplying the odds of a Netflix TV show being renewed by 0.0166, which equals a decrease in the odds of being renewed by 98.34%. The coefficient of IMDb Rating means that holding everything else fixed, an increase of 0.1 in the IMDb Rating is associated with multiplying the odds of renewal by e^(0.1 * 4.56) = 1.578, which equals a 57.8% higher odds of renewal. The coefficient of Popularity says that adding one to the rank of Popularity (a show is slightly less popular) is associated with 1-0.9936 = 0.64% lower odds of renewal. The VIF scores are low, so we do not need to worry about collinearity.

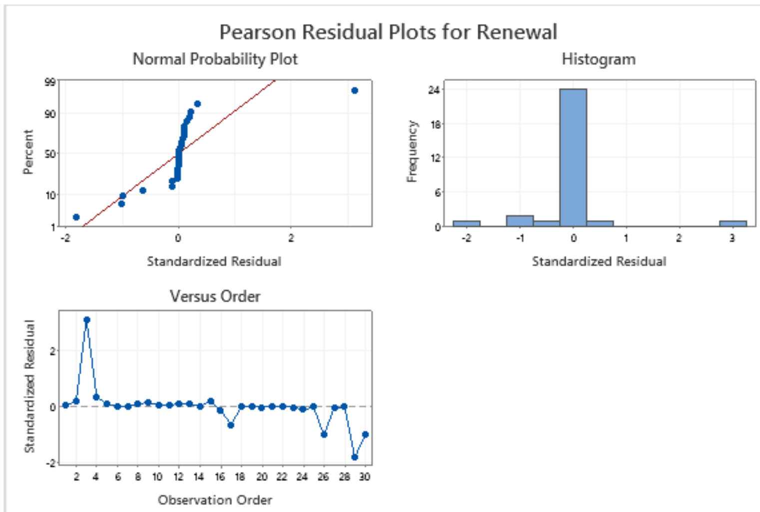
The Hosmer-Lemeshow test has a p-value of 0.515, so we fail to reject the null hypothesis that the model fits the data.

Measures of Association

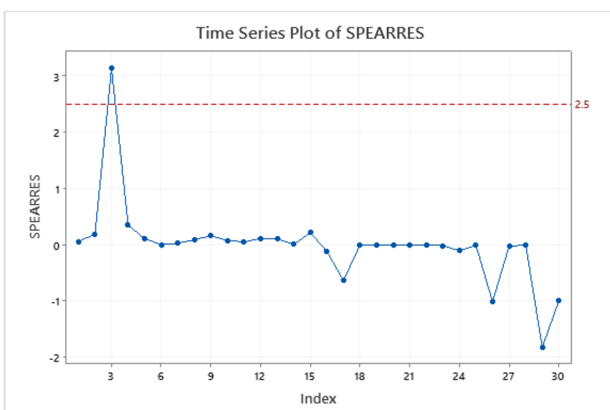
Pairs	Number	Percent	Summary Measures	Value
Concordant	221	98.2	Somers' D	0.96
Discordant	4	1.8	Goodman-Kruskal Gamma	0.96
Ties	0	0.0	Kendall's Tau-a	0.50
Total	225	100.0		

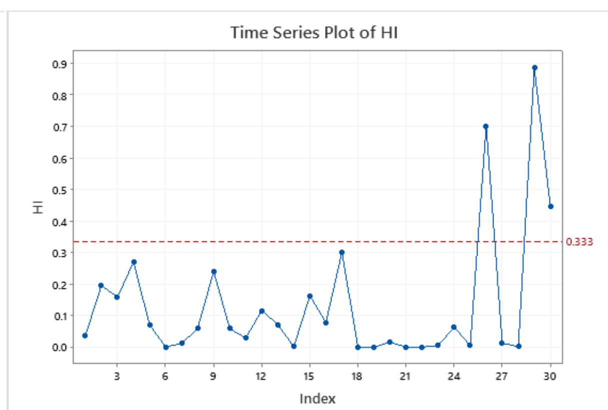
Association is between the response variable and predicted probabilities

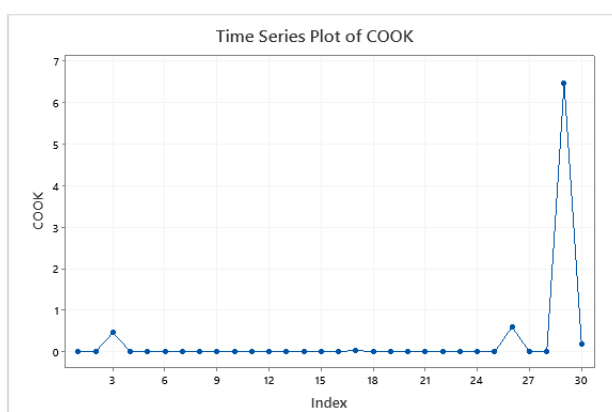
The Somer's D of 0.96 and the Area Under ROC Curve of 0.9822 are both very high, showing an extremely good separation based on the 3 predictors between the renewed shows and the cancelled shows. There are 98.2% of concordant pairs and only 1.8% of discordant pairs, which is very good.



There is clear evidence that we have an outlier here and potentially more unusual points. Let's look at the approximate diagnostics:







Elite is definitely an outlier, while *Ozark*, *The Irregulars* and *The Kominsky Method* are leverage points. Only *The Irregulars* shows up way bigger than 1 in the COOK's distance plot, but *Elite* and *Ozark* also have relatively high COOK's distance values compared to other points. Although these diagnostics are approximate, I am going to consider these 4 as unusual points and deal with them later. It is quite difficult to explain why a Netflix TV show has the existing number of seasons and the current IMDb rating, so the stories of these unusual points are ambiguous. What I do know is that Popularity fluctuates a lot due to media exposure and propaganda.

There is no strong indication that a simpler model might be any better, so let's save the best subset for later and exclude the 4 unusual observations right away.

Binary Logistic Regression: Renewal versus _seasons, IMDb Rating, Popularity

* WARNING * When the data are in the Response/Frequency format, the Residuals versus fits plot is unavailable.

* WARNING * The model could not be fit properly. Maximum likelihood estimates of parameters do not exist due to complete separation of data points. The results are not reliable. Please refer to help for more information about complete separation.

Method

Link function	Logit
Residuals for diagnostics	Pearson
Rows used	26

Response Information

Variable	Value	Count
Renewal	1	14 (Event)
	0	12
	Total	26

Regression Equation

P(1) = exp(Y')/(1 + exp(Y'))

Y' = -48 - 12.0 _seasons + 10.8 IMDb Rating - 0.0182 Popularity

Coefficients

Term	Coef	SE Coef	Z-Value	P-Value	VIF
Constant	-48	323	-0.15	0.881	
_seasons	-12.0	29.4	-0.41	0.684	4.83
IMDb Rating	10.8	46.0	0.23	0.815	6.76
Popularity	-0.0182	0.0843	-0.22	0.829	2.67

Odds Ratios for Continuous Predictors

	Odds Ratio	95% CI
_seasons	0.0000	(0.0000, 6.65559E+19)
IMDb Rating	48075.5920	(0.0000, 6.84977E+43)
Popularity	0.9820	(0.8324, 1.1585)

Model Summary

Deviance R-Sq	Deviance R-Sq(adj)	AIC	AICc	BIC	Area Under ROC Curve
99.97%	91.61%	8.01	9.92	13.04	1.0000

Goodness-of-Fit Tests

Test	DF	Chi-Square	P-Value
Deviance	22	0.01	1.000
Pearson	22	0.01	1.000
Hosmer-Lemeshow	8	0.00	1.000

Analysis of Variance

				Likelihood Ratio	
Source	DF	Adj Dev	Adj Mean	Chi-Square	P-Value
Regression	3	35.8785	11.9595	35.88	0.000
_seasons	1	23.8002	23.8002	23.80	0.000
IMDb Rating	1	*	*	*	*
Popularity	1	*	*	*	*
Error	22	0.0112	0.0005		
Total	25	35.8897			

Measures of Association

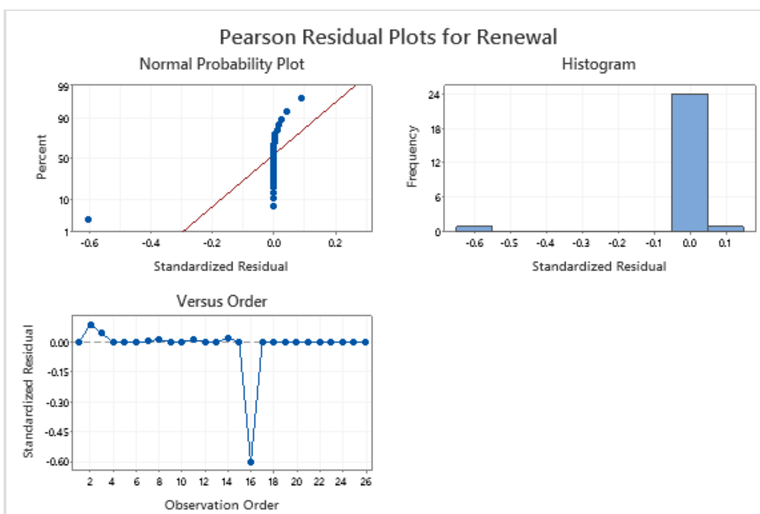
Pairs	Number	Percent	Summary Measures	Value
Concordant	168	100.0	Somers' D	1.00
Discordant	0	0.0	Goodman-Kruskal Gamma	1.00
Ties	0	0.0	Kendall's Tau-a	0.52
Total	168	100.0		

Association is between the response variable and predicted probabilities

Fits and Diagnostics for Unusual Observations

Obs	Observed Probability	Fit	Resid	Std Resid	
2	1.0000	0.9984	0.0405	0.09	X
3	1.0000	0.9995	0.0218	0.04	X
8	1.0000	1.0000	0.0070	0.01	X
16	0.0000	0.0029	-0.0536	-0.60	X

X Unusual X



The results show a clear sign of complete separation. Therefore, I will then try best subsets in an attempt to find a simpler model.

Response is Renewal

Vars	R-Sq	R-Sq (adj)	R-Sq (pred)	Mallows Cp	S	_seasons	IMDb Rating	Popularity
1	38.7	36.1	30.0	43.9	0.40628	X		
1	30.8	28.0	22.3	52.4	0.43150			X
2	79.1	77.3	74.8	2.5	0.24225	X		X
2	66.0	63.0	54.4	16.6	0.30917	X	X	
3	79.5	76.7	71.3	4.0	0.24515	X	X	X

The results of best subsets show us that choosing a 2-predictor model with #seasons and Popularity is probably better because the Cp is smaller. Then, let's try fitting the simpler model:

Binary Logistic Regression: Renewal versus _seasons, Popularity

* WARNING * When the data are in the Response/Frequency format, the Residuals versus fits plot is unavailable.

* WARNING * The model could not be fit properly. Maximum likelihood estimates of parameters do not exist due to complete separation of data points. The results are not reliable. Please refer to help for more information about complete separation.

Method

Link function	Logit
Residuals for diagnostics	Pearson
Rows used	26

Response Information

Variable	Value	Count
Renewal	1	14 (Event)
	0	12
	Total	26

Regression Equation

P(1) = exp(Y')/(1 + exp(Y'))

Y' = 50.6 - 13.6 _seasons - 0.0362 Popularity

Analysis of Variance

				Likelihood Ratio	
Source	DF	Adj Dev	Adj Mean	Chi-Square	P-Value
Regression	2	35.8723	17.9362	35.87	0.000
_seasons	1	24.8950	24.8950	24.90	0.000
Popularity	1	23.3850	23.3850	23.39	0.000
Error	23	0.0173	0.0008		
Total	25	35.8897			

Measures of Association

Pairs	Number	Percent	Summary Measures	Value
Concordant	168	100.0	Somers' D	1.00
Discordant	0	0.0	Goodman-Kruskal Gamma	1.00
Ties	0	0.0	Kendall's Tau-a	0.52
Total	168	100.0		

Association is between the response variable and predicted probabilities

Fits and Diagnostics for Unusual Observations

Obs	Observed Probability	Fit	Resid	Std Resid	
2	1.0000	0.9965	0.0591	0.16	X
8	1.0000	0.9992	0.0288	0.11	X
16	0.0000	0.0027	-0.0524	-0.09	X
23	0.0000	0.0013	-0.0358	-0.05	X

X Unusual X

Coefficients

Term	Coef	SE Coef	Z-Value	P-Value	VIF
Constant	50.6	54.6	0.93	0.353	
_seasons	-13.6	15.5	-0.87	0.382	1.96
Popularity	-0.0362	0.0416	-0.87	0.384	1.96

Odds Ratios for Continuous Predictors

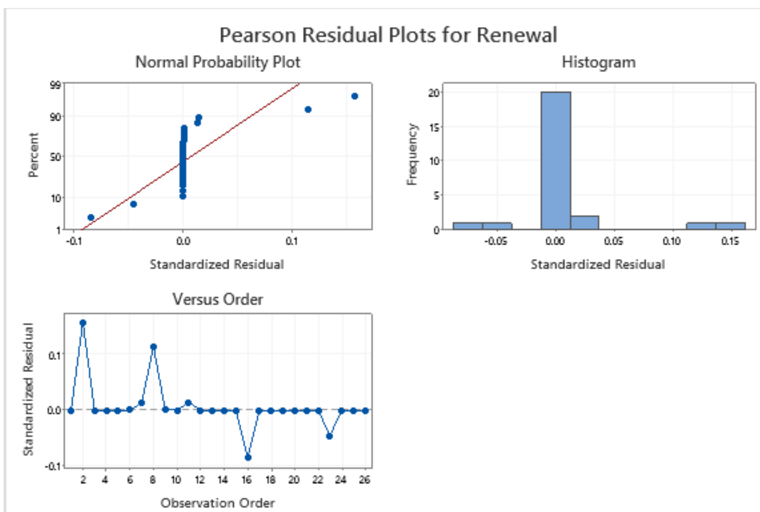
	Odds Ratio	95% CI
_seasons	0.0000	(0.0000, 2.12499E+07)
Popularity	0.9644	(0.8889, 1.0463)

Model Summary

Deviance R-Sq	Deviance R-Sq(adj)	AIC	AICc	BIC	Area Under ROC Curve
99.95%	94.38%	6.02	7.11	9.79	1.0000

Goodness-of-Fit Tests

Test	DF	Chi-Square	P-Value
Deviance	23	0.02	1.000
Pearson	23	0.01	1.000
Hosmer-Lemeshow	8	0.00	1.000



Again, there is complete separation, so I will check the three 1-predictor models to see if their performance can match up.

Binary Logistic Regression: Renewal versus _seasons

* WARNING * When the data are in the Response/Frequency format, the Residuals versus fits plot is unavailable.

Method

Link function	Logit
Residuals for diagnostics	Pearson
Rows used	26

Response Information

Variable	Value	Count	
Renewal	1	14	(Event)
	0	12	
	Total	26	

Regression Equation

P(1) = exp(Y')/(1 + exp(Y'))

Y' = 3.13 - 1.207 _seasons

Coefficients

Term	Coef	SE Coef	Z-Value	P-Value	VIF
Constant	3.13	1.20	2.60	0.009	
_seasons	-1.207	0.474	-2.55	0.011	1.00

Odds Ratios for Continuous Predictors

	Odds Ratio	95% CI
_seasons	0.2992	(0.1182, 0.7575)

Model Summary

Deviance R-Sq	Deviance R-Sq(adj)	AIC	AICc	BIC	Area Under ROC Curve
34.79%	32.01%	27.40	27.92	29.92	0.8363

Goodness-of-Fit Tests

Test	DF	Chi-Square	P-Value
Deviance	24	23.40	0.496
Pearson	24	24.39	0.440
Hosmer-Lemeshow	3	2.34	0.504

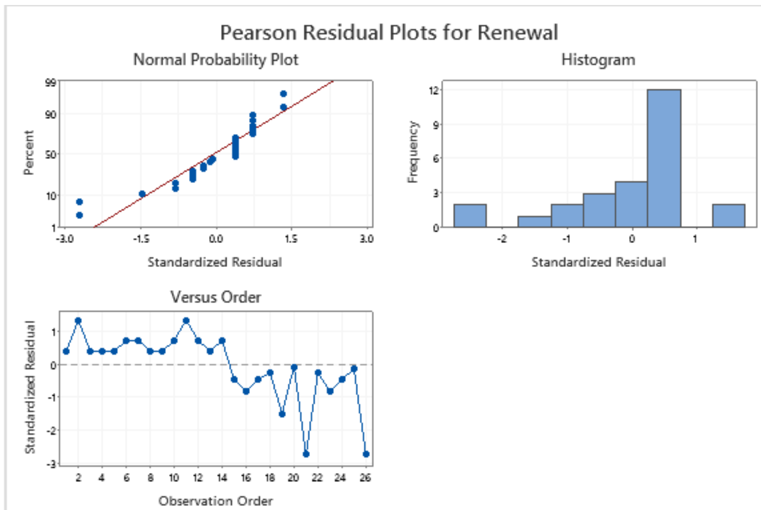
Analysis of Variance

				Likelihood Ratio	
Source	DF	Adj Dev	Adj Mean	Chi-Square	P-Value
Regression	1	12.49	12.4873	12.49	0.000
_seasons	1	12.49	12.4873	12.49	0.000
Error	24	23.40	0.9751		
Total	25	35.89			

Measures of Association

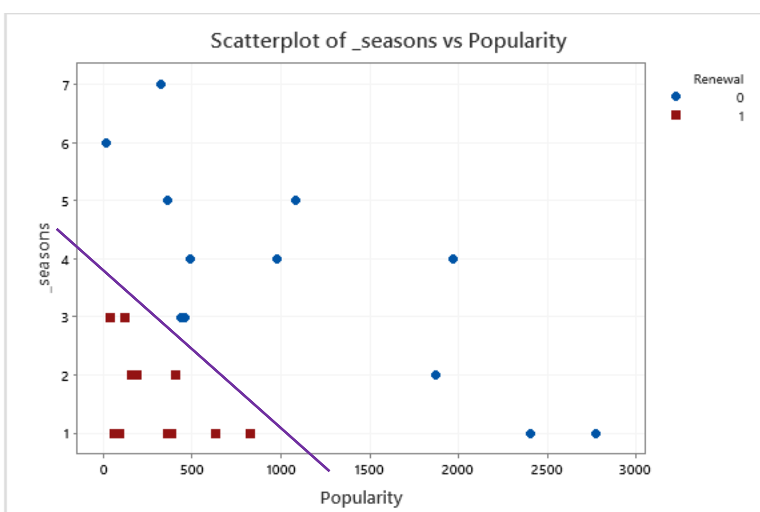
Pairs	Number	Percent	Summary Measures	Value
Concordant	129	76.8	Somers' D	0.67
Discordant	16	9.5	Goodman-Kruskal Gamma	0.78
Ties	23	13.7	Kendall's Tau-a	0.35
Total	168	100.0		

Association is between the response variable and predicted probabilities



We can see that the fit is much worse because Somers' D is only 67% now and the AUC is only about 83.6%. The other two 1-predictor models fit even worse, so I am not going to show the results here. There doesn't seem to be a simpler model that fits as well as, or almost as well as, the 2-predictor model with #seasons and Popularity as the predictors.

Therefore, let's take a closer look at our best model. We can plot the two predictors with groups:



We can easily find a line that perfectly separates the two groups. The scatterplot shows that if the number of existing seasons is smaller enough and the popularity rank is high enough, the Netflix TV show is predicted to be renewed. Otherwise, our model predicts it to be cancelled.

However, we got this perfect separation now because we excluded unusual points earlier, so I am going to put the 4 unusual points back to evaluate the best model we have. Not all leverage points are predicted wrongly, and the confusion matrix is as follows:

Rows: Renewal   Columns: Predict

	0	1	All
0	14	1	15
	46.67	3.33	50.00
1	1	14	15
	3.33	46.67	50.00
All	15	15	30
	50.00	50.00	100.00

Cell Contents
Count
% of Total

The outlier *Elite* is renewed but misclassified as cancelled; the leverage point *The Irregulars* is cancelled but mis-predicted as renewed. Other than that, 93.34% of the Netflix TV shows are correctly classified, which is much higher than:

$$C_{pro} = 1.25[0.5 * 0.5 + 0.5 * 0.5] = 62.5\% \text{ and } C_{max} = 50\%.$$

Therefore, the 2-predictor model using #seasons and Popularity is doing a very good job at classifying Netflix TV shows into renewal and cancellation. I do not really have any test data because I have already exhausted all of the cancelled shows on the list.

Since this model fits the retrospective data, we can adjust the constant term of the regression using prior probabilities of renewal to estimate prospective probabilities of renewal. I will use 75% as the prior of renewal. Then, the adjusted intercept is:

$$\widetilde{\beta_0} = \widehat{\beta_0} + \ln\left(\frac{75\% * 15}{25\% * 15}\right) = 6.26 - 1.099 = 5.151$$

Regression Equation

P(1) = exp(Y')/(1 + exp(Y'))

Y' = 6.26 - 1.759 _seasons - 0.00386 Popularity

Here are the estimated adjusted probabilities:

Row	Title	newprob
1	Bridgerton	0.959818

2	Cobra Kai	0.362434
3	Elite	0.228431
4	Fate: The Winx Saga	0.882789
5	Locke & Key	0.871999
6	Love, Death & Robots	0.955436
7	Mindhunter	0.718991
8	Narcos: Mexico	0.520612
9	Never Have I Ever	0.552935
10	Russian Doll	0.724947
11	Sex Education	0.735826
12	Stranger Things	0.438247
13	The Umbrella Academy	0.741045
14	The Witcher	0.958915
15	You	0.735075
16	Atypical	0.022929
17	Dead to Me	0.142311
18	Dear White People	0.000077
19	F is for Family	0.000403
20	Feel Good	0.003815
21	Grace and Frankie	0.000227
22	Hoops	0.000673
23	Kim's Convenience	0.006505
24	Lost in Space	0.132701
25	On My Block	0.003582
26	Ozark	0.104316
27	Peaky Blinders	0.004253
28	The Duchess	0.002769
29	The Irregulars	0.95203
30	The Kominsky Method	0.125752

The Irregulars still has an estimated probability of over 95% of being renewed, which only drop by about 3%, so it will continue to be misclassified.

In conclusion, the best model so far is the 2-predictor model with #seasons and IMDb Popularity as the predictors. Given the small sample of observations, it can classify the renewal of a Netflix TV show very well. The model is consistent with my expectation that shows with fewer seasons and higher popularity rank are more likely to be renewed. This model will be a useful tool for me to check if my favorite shows will be renewed or cancelled. But of course, there are more complicated forces at play. It would be nice if I can include viewership data as another predictor or genre of the TV show as a categorical predictor in the future. Some genres such as horror generally do not have a long life expectancy. It would be interesting to see how these predictors can improve the model.