Predicting Theater Season Ticket Renewals

Cody Frisby

10/28/2017

The number of performances attended (performances) appears to have the largest influence on the response variable (Ticket Renewal). This makes intuitive sense. Why would you buy season tickets if you never attend the theater? Next largest influencer appears to be the number of different users (users) of the tickets. This sort of makes sense. People may buy season tickets as a group and then share them among the group where only a subset of the group goes to each performance. Next up is the average time before the curtain (timeBeforeCurtain). It appears people who are serious about attending the theater arrive earlier, on average. As a general rule, the more you do something the better you get at it. Season ticket holders likely make an event out of attending. They get ready, know by when they need to eat, and then arrive at the theater before the show begins so that they have time to check their coat, visit the concessions, and use the facilities.

	Estimate	Std. Error	z value	$\Pr(>|z|)$
(Intercept)	-6.9311897	1.7246106	-4.018988	0.0000584
price	-0.0194302	0.0193850	-1.002328	0.3161853
tickets	-0.7626146	0.4718551	-1.616205	0.1060500
users	0.5544902	0.1358542	4.081508	0.0000447
voucher	0.5536790	0.4600864	1.203424	0.2288123
timeBeforeCurtain	0.0263600	0.0097657	2.699258	0.0069494
performances	0.9858401	0.1666560	5.915419	0.0000000

Let's see what the model looks like with predicting on a blind dataset.

Since we cannot represent all variables, I've chosen to exclude time (as this was our least significant out of the significant factors :)) and to include **users**, **timeBeforeCurtain**, and the predicted probabilities.
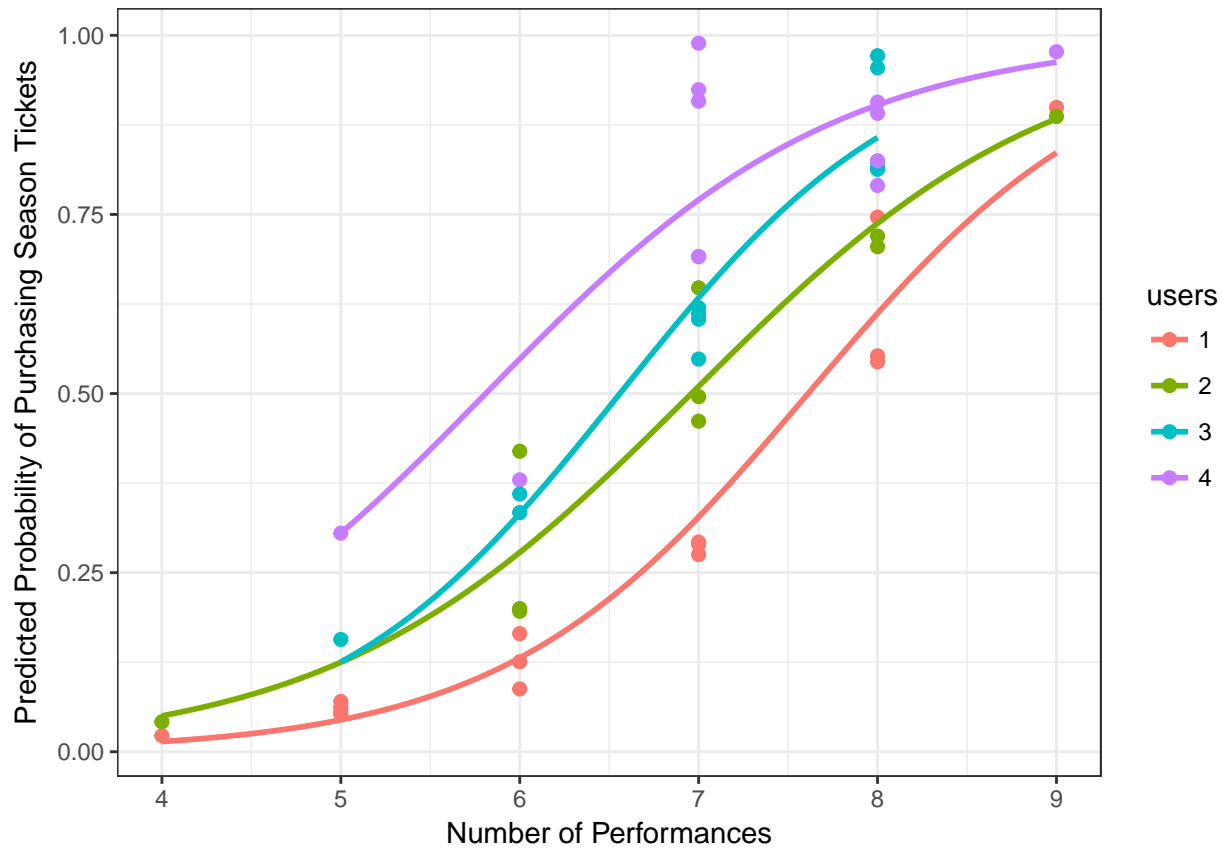
It appears that the model at least appears to be predicting to what intuition would suggest based on the model and that it does reasonably well on the blind dataset. Huzzah!

If we make the decision point (whether or not someone will renew) as $p = 0.5$ then we can predict that **27** will renew, or around **55%** of those in the test dataset.