1.Create long form

2.Create the column: cum_label, which is the cumsum of all labeled apps in the category.

	stop	start	Label	app_category	app_name	app_type	id	rank_on_14_Dec	cum_label
0	1.0	0.0	0	1	Bible Verses: Daily Devotional	free	0	58	1
1	1.0	0.0	0	1	Daily Bible Inspirations	free	1	52	1
2	1.0	0.0	0	1	Daily Bible Verse & Motivation	free	2	85	1
3	1.0	0.0	0	1	Daily Devotional For Women App	gross	3	70	1
4	1.0	0.0	0	1	Dictionary.com: English Words	gross	4	28	1
...	...	...	...	...	...	...	...	...	...
9151	26.0	25.0	0	87	Gardenscapes	gross	465	26	12
9152	27.0	26.0	0	87	Gardenscapes	gross	465	26	12
9153	28.0	27.0	0	87	Gardenscapes	gross	465	26	12
9154	29.0	28.0	0	87	Gardenscapes	gross	465	26	12
9155	30.0	29.0	0	87	Gardenscapes	gross	465	26	12

9156 rows × 9 columns

3.Fit CoxTimeVaring Model

	coef	exp(coef)	se(coef)	coef lower 95%	coef upper 95%	exp(coef) lower 95%	exp(coef) upper 95%	cmp to	z	p	−log2(p)
rank_on_14_Dec	-0.01	0.99	0.00	-0.01	-0.01	0.99	0.99	0.00	-4.97	<0.005	20.54
cum_label	0.04	1.04	0.01	0.02	0.05	1.02	1.06	0.00	4.11	<0.005	14.65
app_type_gross	0.27	1.31	0.11	0.06	0.48	1.06	1.62	0.00	2.50	0.01	6.32
app_type_paid	-0.12	0.89	0.14	-0.40	0.16	0.67	1.17	0.00	-0.85	0.39	1.34

Before this, the coefficients were -0.01 for rank, 0.1 for app_type_gross, -0.23 for app_type_paid. After adding the covariate number of labeled apps in the same category, the coefficients that are different than before are 0.04 for cum_label, 0.27 for gross app, -0.12 for paid app. The cum_label shows a positive coefficient, which means that one labeled app increase in the same category as my app, the app has a 4% increase in hazard of getting labeled.