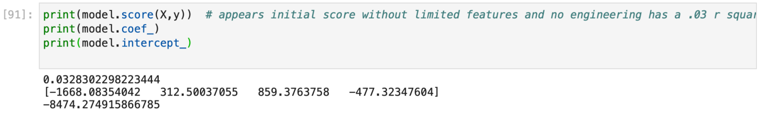
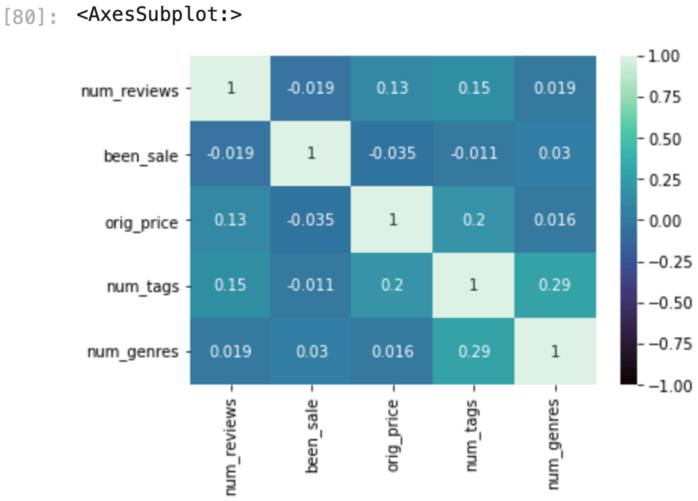
MVP

Data was able to be successfully scraped from 16k + game pages. Of these pages, 5k+ contains total number of reviews. Using only these pages, I created a preliminary data frame with only the numeric features scraped: if the game has been on sale (1 being yes 0 being no), the original price for the game, the number of user-assigned descriptive tags, and the numbers of genres associated with the game. Using both sklearn and stats models, I fit this into a linear regression model. Both r sqared values match which is a great sanity check. Unfortunately the score is only .03. Checking the heatmap and pair plot also shows that, without the categorical data or other feature engineering, there are no strong correlations with our target and our features. There are several more features of categorical nature that are will be included in the final model which should improve it. In the next couple of days complexity and regularization will be added to help improve the model.

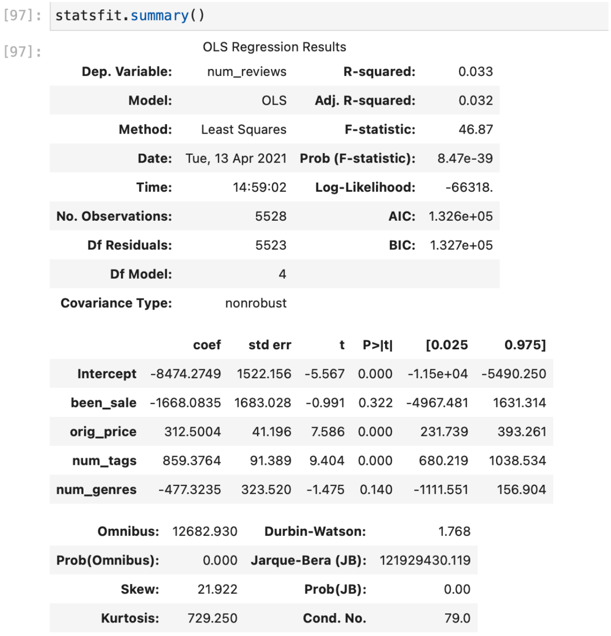


```
[91]: print(model.score(X,y))  # appears initial score without limited features and no engineering has a .03 r squa
      print(model.coef_)
      print(model.intercept_)

0.0328302298223444
[-1668.08354042   312.50037055   859.3763758   -477.32347604]
-8474.274915866785
```



```
[80]: <AxesSubplot:>
```



```
[97]: statsfit.summary()
```

[97]:

OLS Regression Results

Dep. Variable:	num_reviews	**R-squared:**	0.033
Model:	OLS	**Adj. R-squared:**	0.032
Method:	Least Squares	**F-statistic:**	46.87
Date:	Tue, 13 Apr 2021	**Prob (F-statistic):**	8.47e-39
Time:	14:59:02	**Log-Likelihood:**	-66318.
No. Observations:	5528	**AIC:**	1.326e+05
Df Residuals:	5523	**BIC:**	1.327e+05
Df Model:	4		
Covariance Type:	nonrobust		

	coef	**std err**	**t**	**P>|t|**	**[0.025**	**0.975]**
Intercept	-8474.2749	1522.156	-5.567	0.000	-1.15e+04	-5490.250
been_sale	-1668.0835	1683.028	-0.991	0.322	-4967.481	1631.314
orig_price	312.5004	41.196	7.586	0.000	231.739	393.261
num_tags	859.3764	91.389	9.404	0.000	680.219	1038.534
num_genres	-477.3235	323.520	-1.475	0.140	-1111.551	156.904

Omnibus:	12682.930	**Durbin-Watson:**	1.768
Prob(Omnibus):	0.000	**Jarque-Bera (JB):**	121929430.119
Skew:	21.922	**Prob(JB):**	0.00
Kurtosis:	729.250	**Cond. No.**	79.0