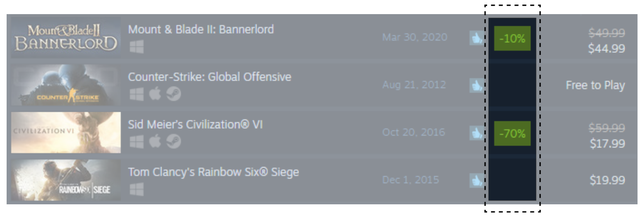
PROGRESS REPORT

Question:

~~IF a strategy game would go on sale~~ (Classification)



The **MAGNITUDE** of discount for strategy games on sale

MVP:

To have a model that has non-negative r^2 score

AKA having a valid model first, regardless of accuracy

```
Simple regression scores:  [-0.8712213661985753, -0.15773507212826177, -0.08655
602179665012, 0.06587732713005157, -0.041567824258621044]
Ridge scores:  [-0.8388331235826887, -0.1545458851858017, -0.08426954757266092,
0.06542175489043889, -0.04008859022967326]

Poly scores:  [-0.4195352918522708, 0.08203208456310784, 0.025578433307165693,
-1.5214338335302637, 0.14432685121285327]

Simple mean cv r^2: -0.218 +- 0.334
Ridge mean cv r^2: -0.210 +- 0.322
Poly mean cv r^2: -0.338 +- 0.624
```

F.E. 1: % of positive review + Multiplatform?


```
In [58]: RSquare(df4, 'DiscountPercentage')

Simple regression scores:  [0.13485468607703377, -1.764963941376421e+17, 0.1013
2572230024285, -0.32058014062391105, -226869547315357.4]
Ridge scores:  [0.16769285908339127, 0.025700881163476352, 0.12350988154052323,
-0.3156494402754604, -0.5765198288777611]

Poly scores:  [-21373.49028996433, -5696385146236974.0, -9593.190456480443, -11
755.209585854009, -28450.58057332444]

Simple mean cv r^2: -353446527369914960.000 +- 70575925396538456.000
Ridge mean cv r^2: -0.115 +- 0.286
Poly mean cv r^2: -1139277029261629.250 +- 2278554058487672.500
lasso_model.alpha_: 1.231550603292826
Lasso cv r^2: 0.200 +- mae: 15.434
```

MVP

20 tags

Mount & Blade II

F.E. 2: Top 5 tags + get_dummies

```
In [58]: RSquare(df4, 'DiscountPercentage')
  tol, rng, random, positive)
F:\Anaconda\lib\site-packages\sklearn\linear_model\_coordinate_descent.py:47
2: ConvergenceWarning: Objective did not converge. You might want to increase
the number of iterations. Duality gap: 45.1735765782214, tolerance: 12.063175
61983471
  tol, rng, random, positive)

Simple regression scores:  [-1.7895274493868927, -2.1547677246119195, -2.1648
66138806862, -2.584081432685239, -1.0965250965601472]
Ridge scores:  [-0.7833693915901674, -0.9821692669544807, -1.285731918120684
5, -1.698851334181954, -0.7586381618634175]

Poly scores:  [-11178415882.912918, -18194240.84463007, -135256751444714.44,
-41988897331.00571, -1134160.4869739907]

Simple mean cv r^2: -1.958 +- 0.499
Ridge mean cv r^2: -1.102 +- 0.353
Poly mean cv r^2: -27061987617265.934 +- 54097384096491.758
lasso_model.alpha_: 0.9329304026284686
Lasso cv r^2: 0.432 +- mae: 12.429
```

ADDENDUM

Latest

```
In [26]: RSquare(df1,'DiscountPercentage')
```

```
Simple regression scores:  [0.30750284610204726, 0.28649837551148893, 0.3457299
77673659, 0.5200963028928389, 0.16658407395263852]
Ridge scores:  [0.3071949423736203, 0.2872291593099917, 0.3455525148563239, 0.5
192973818863691, 0.1682362800753936]

Poly scores:  [0.32367199635228505, 0.27849255601895717, 0.1420455118957622, -
6.952982273669873, 0.1735496073427446]

Simple mean cv r^2: 0.325 +- 0.114
Ridge mean cv r^2: 0.326 +- 0.114
Poly mean cv r^2: -1.207 +- 2.874
lasso_model.alpha_: 1e-10
Lasso cv r^2: 0.461 +- mae: 15.029
```

Same dataset, same steps, different scores?

```
In [36]: RSquare(df2, 'DiscountPercentage')
```

```
Simple regression scores:  [0.21153999178467764, 0.256462693286006, 0.318118916
5969409, 0.3337896352688502, 0.07633544491435473]
Ridge scores:  [0.2111647652816686, 0.2567542764586157, 0.31782351558233135, 0.
3344205993801748, 0.07832442223457492]

Poly scores:  [0.2616916228600896, 0.26966389237869093, 0.15513199463602756, -
5.6017535812781025, 0.10954643716344215]

Simple mean cv r^2: 0.239 +- 0.092
Ridge mean cv r^2: 0.240 +- 0.092
Poly mean cv r^2: -0.961 +- 2.321
lasso_model.alpha_: 0.4659525668664682
Lasso cv r^2: 0.341 +- mae: 16.943
```