



LINEAR REGRESSION

`predict_steam_discounts`

Ask: Questions

~~1~~_{/2} If and when a game on Steam would go on sale. — CLASSIFICATION

2_{/2} What determines the magnitude of strategy game discounts

What

on Steam during a regular week, using linear regression modeling.

Where

When

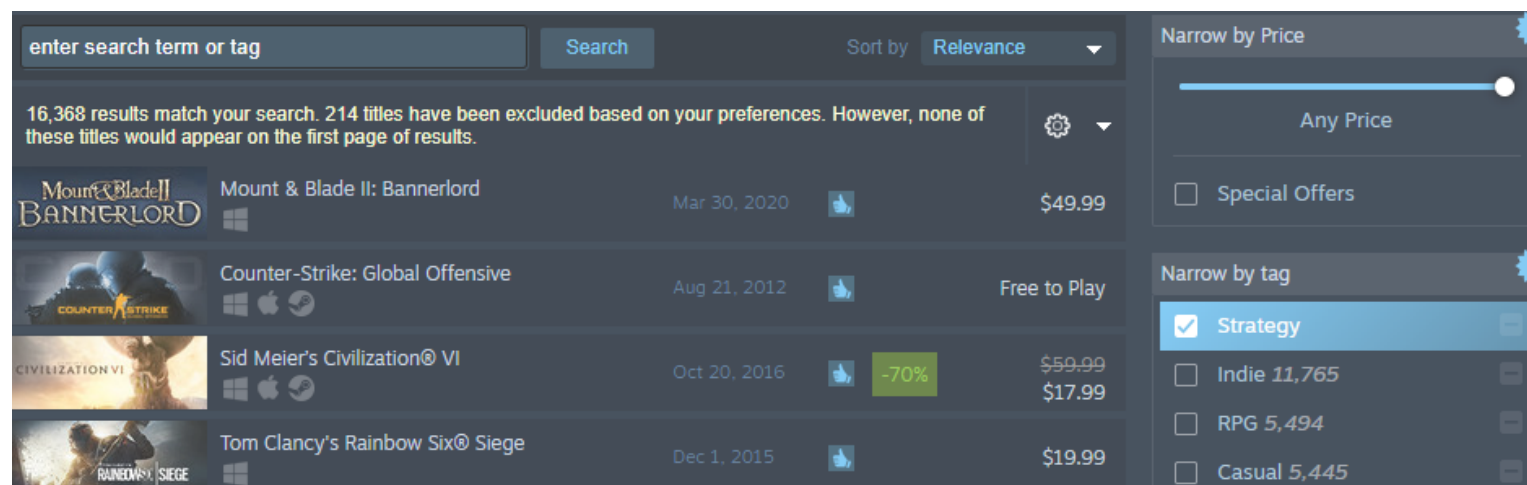
How

Why?

CURIOSITY

Data Scraping

Column: 14,806 Features: 12 →



```
In [4]: df.shape
Out[4]: (14806, 12)
```

Modeling

Effectiveness:

R^2 & MAE

Data Cleaning

Column: 370 Features: 7

Exclusion: Games not on sale & Free games (demos, upcoming)

```
In [29]: df1.shape
Out[29]: (370, 7)
```

EDA

Findings: What matters
& What doesn't

During a regular week on Steam in the strategy genre:

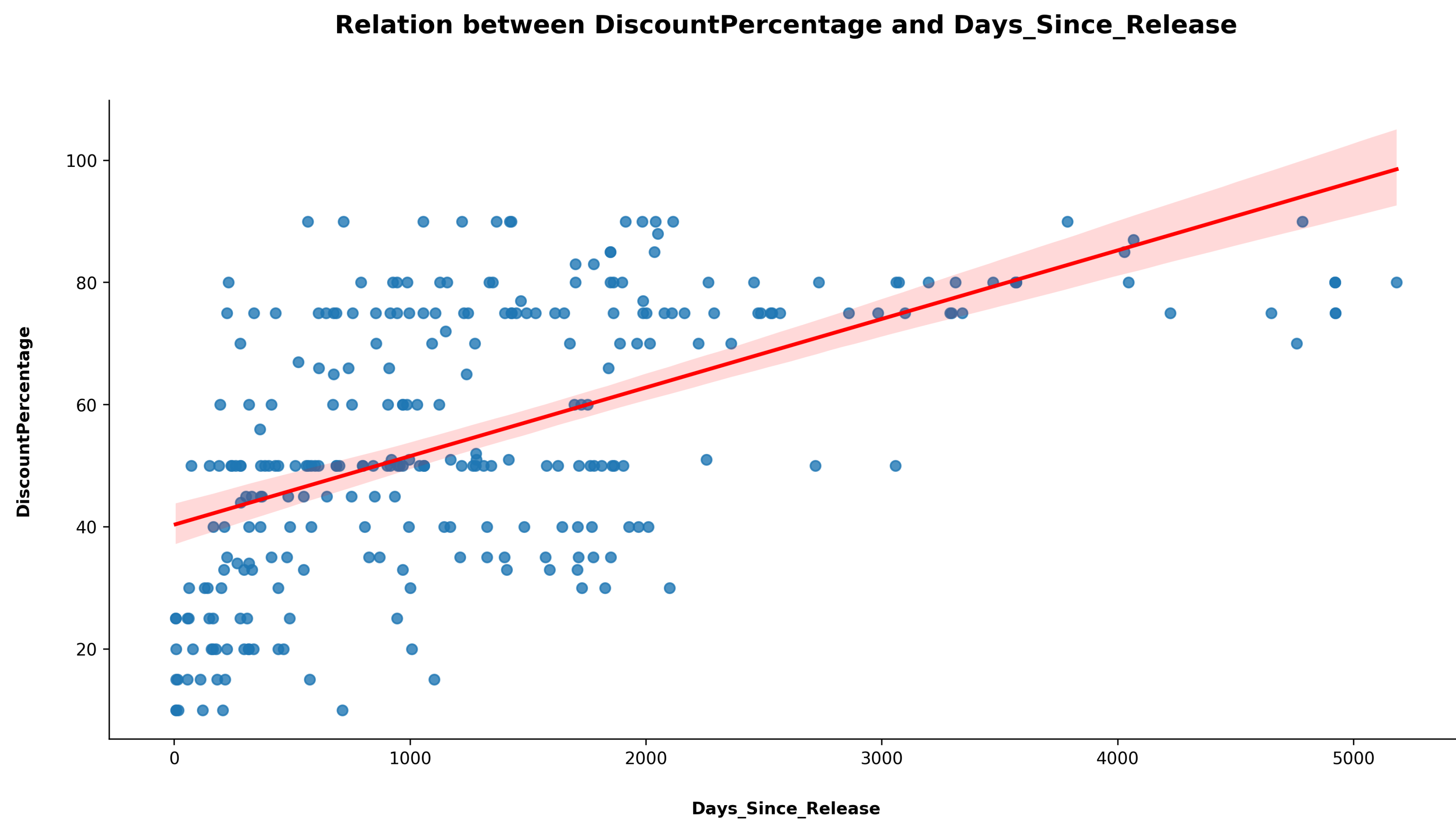
3.5% are pay-to-play games.
[527 / 14806]

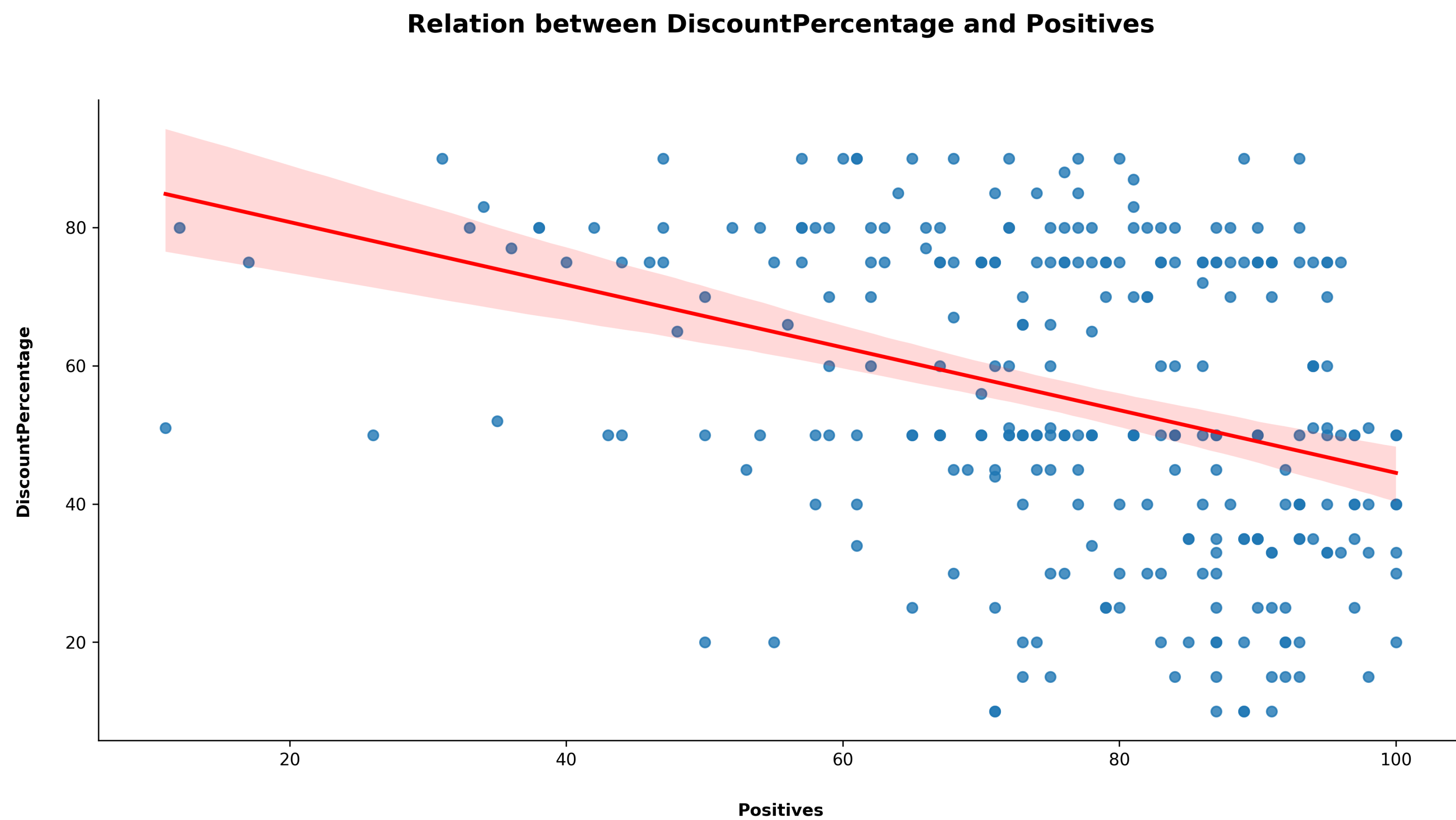
2.4% are on sale in this genre.
[370 / 14806]

64% are on sale among pay-to-play games.
[370 / 572]

Conclusion 1

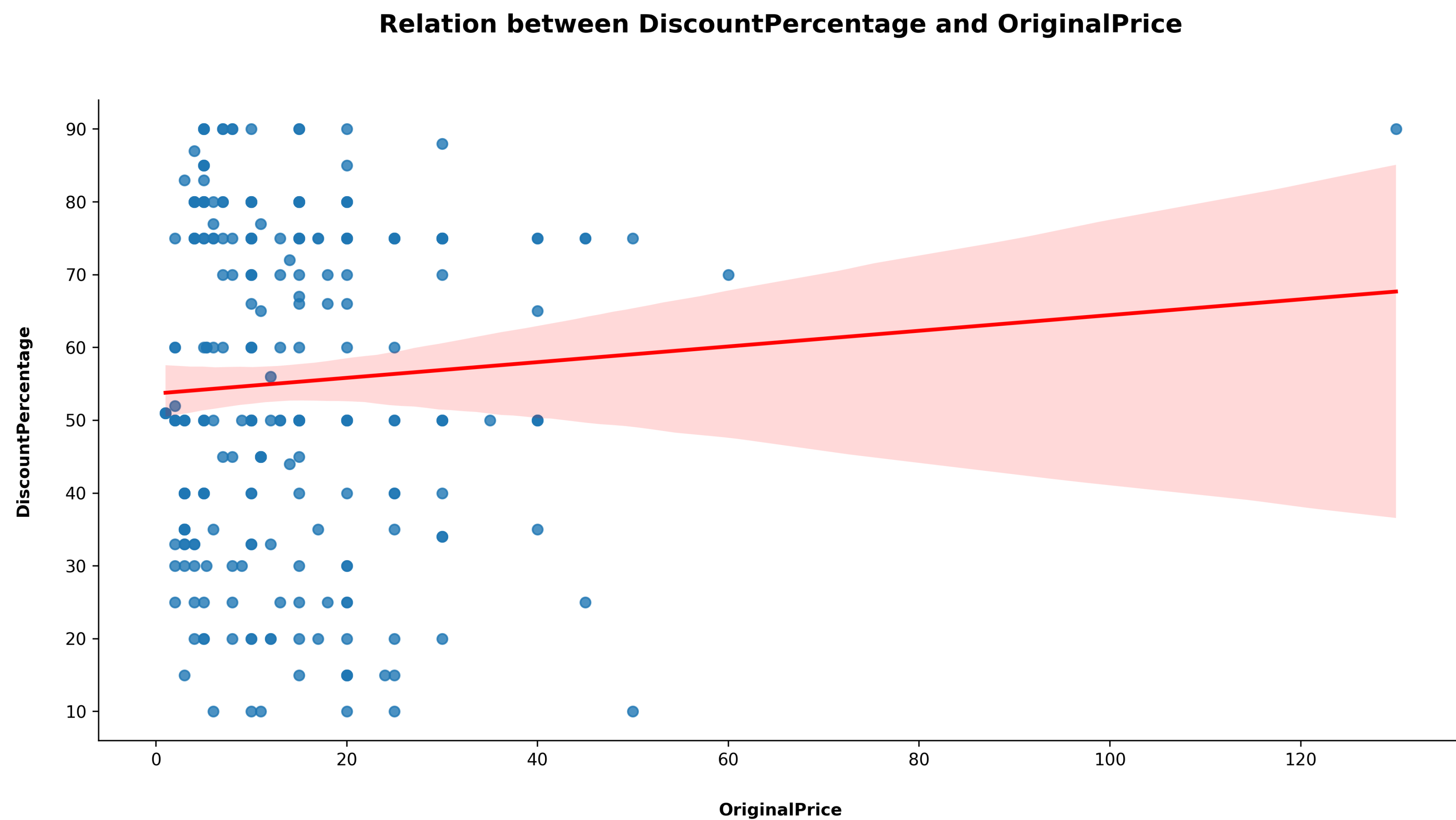
EDA: **What matters?**



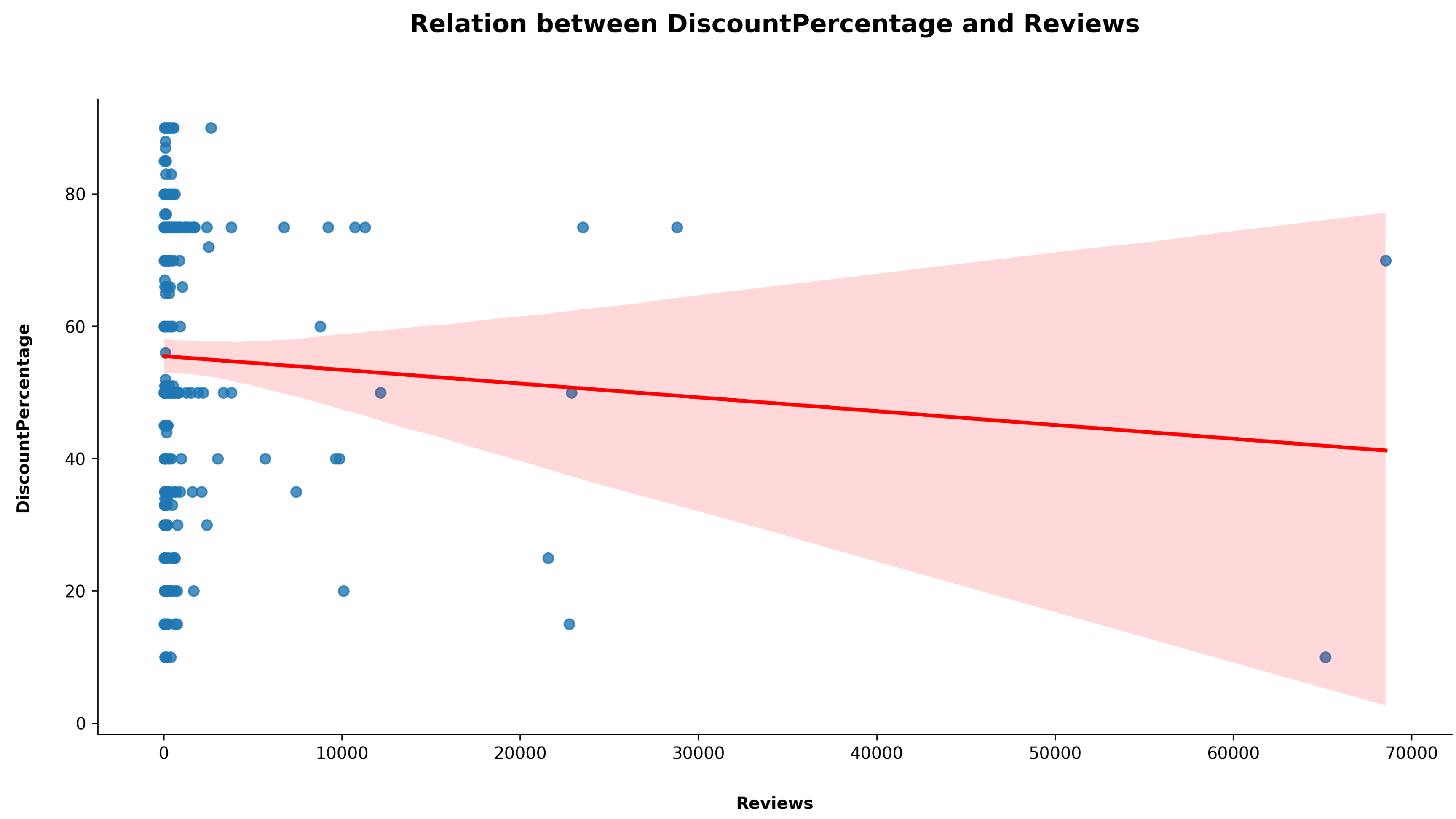


EDA: **What doesn't matter?**

What doesn't
Original price doesn't affect discount



What doesn't
Amount of reviews doesn't affect discount

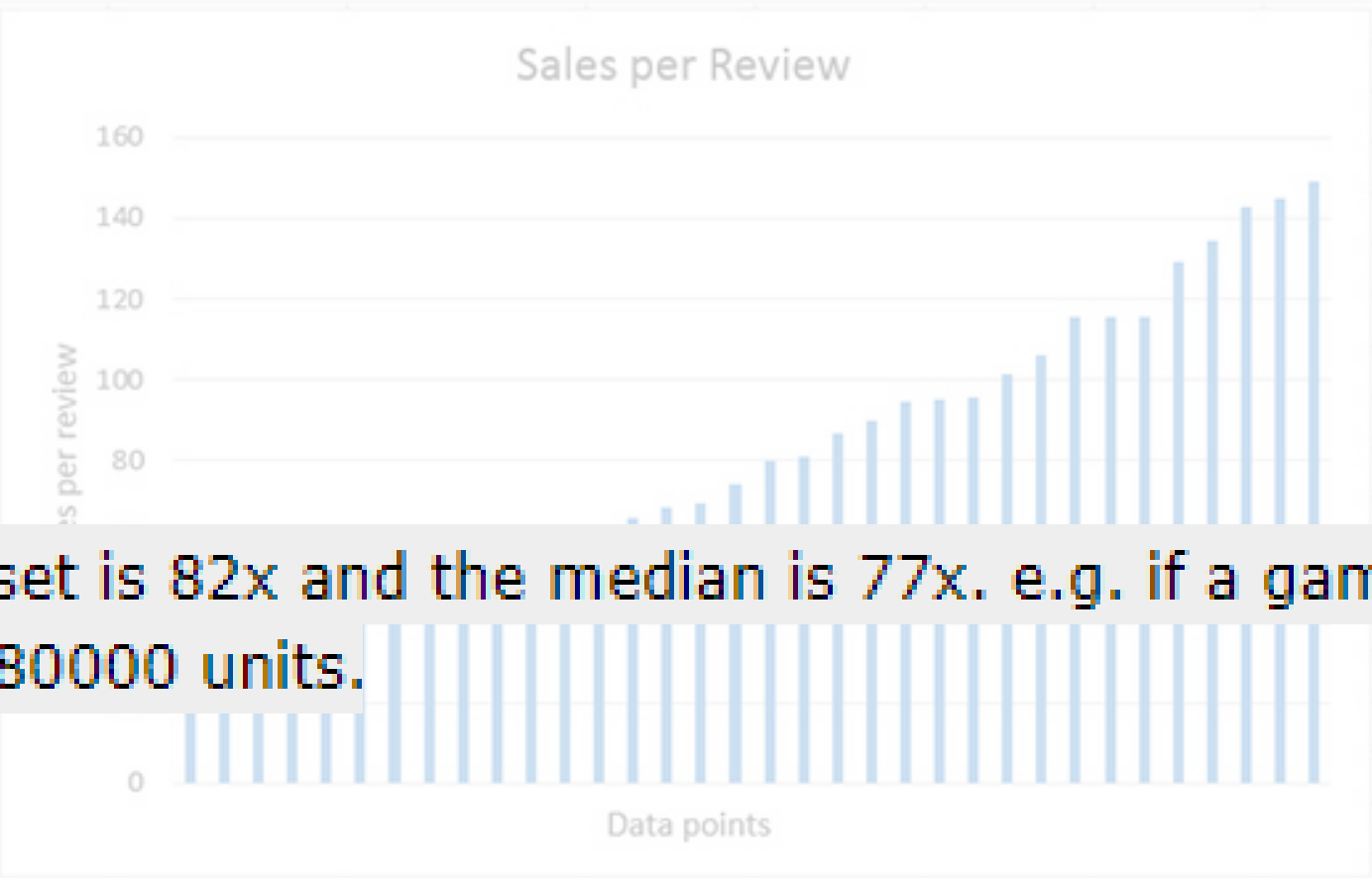


Using Steam reviews to estimate sales

by Jake Birkett on 05/04/18 04:03:00 am Expert Blogger Featured Post

5 comments   Share 

The following blog post, unless otherwise noted, was written by a member of Gamasutra's community. The thoughts and opinions expressed are those of the writer and not Gamasutra or its parent company.



The average from my dataset is 82x and the median is 77x. e.g. if a game has 1000 reviews you can estimate it has sold about 80000 units.

Back in 2014, before SteamSpy, I wrote [this article](#) about using Steam's review count to estimate sales. Then SteamSpy came along and we all used that instead.

However, now that SteamSpy is unable to show the number of owners due to Valve's privacy policy changes, the good old review count method has become valid again!

I decided it was worth getting an updated "Boxleiter" number (how much you should multiply a game's review count by in order to estimate units sold). So as per usual I asked a bunch of indies to supply accurate data. This time I got over 50 data points. Read on to discover my conclusions!

The Headline

The average from my dataset is 82x and the median is 77x. e.g. if a game has 1000 reviews you can estimate it has sold about 80000 units. But please keep reading as there are some important things to note.

Linear regression: **The Model**

After **8** rounds of feature engineering:

```
Lasso cv r^2: 0.419 +- 0.073
```

```
mae: 12.76497338985648
```

Accounts for **42%** of variance with $\pm 12\%$ in % discount

Practicality: Use Case

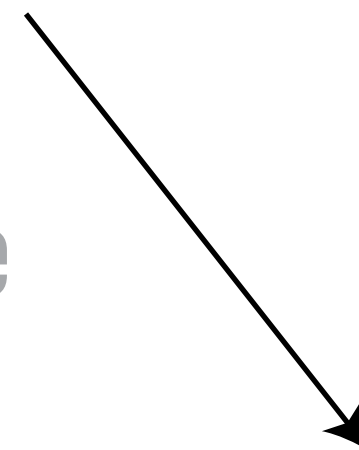
Video game resellers



Find big discounts
Buy in bulk



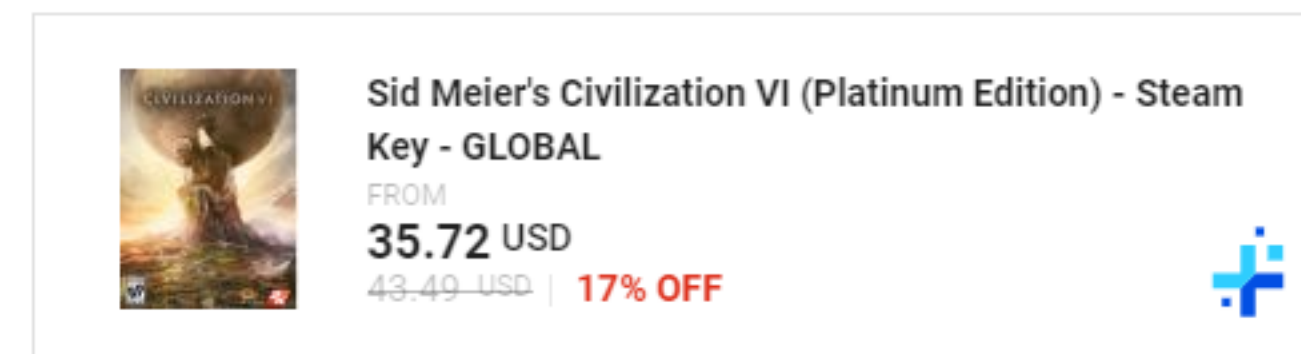
Wait for it to go off sale



Gamer buys cheap for \$36,
and reseller just earned \$18



Win - Win



List on their website for a little cheaper

Useful?
For online video game resellers

1_{/2} If and when a strategy game would go on sale.

2_{/2} What determines the magnitude of strategy game discounts
on Steam during a regular week, using linear regression modeling.

ANSWERED BUT NEED IMPROVEMENTS

THANK YOU

During a regular week on Steam in the strategy genre:

Stats:

3.5% are pay-to-play games.
[527 / 14806]

2.4% are on sale in this genre.
[370 / 14806]

64% are on sale among pay-to-play games.
[370 / 572]

What affects discount:

Older games get more discount.

Lower pos. review % mean more discount.

What doesn't:

Amount of reviews (sales) doesn't affect discount.

Original price doesn't affect discount.

This model is useful for:

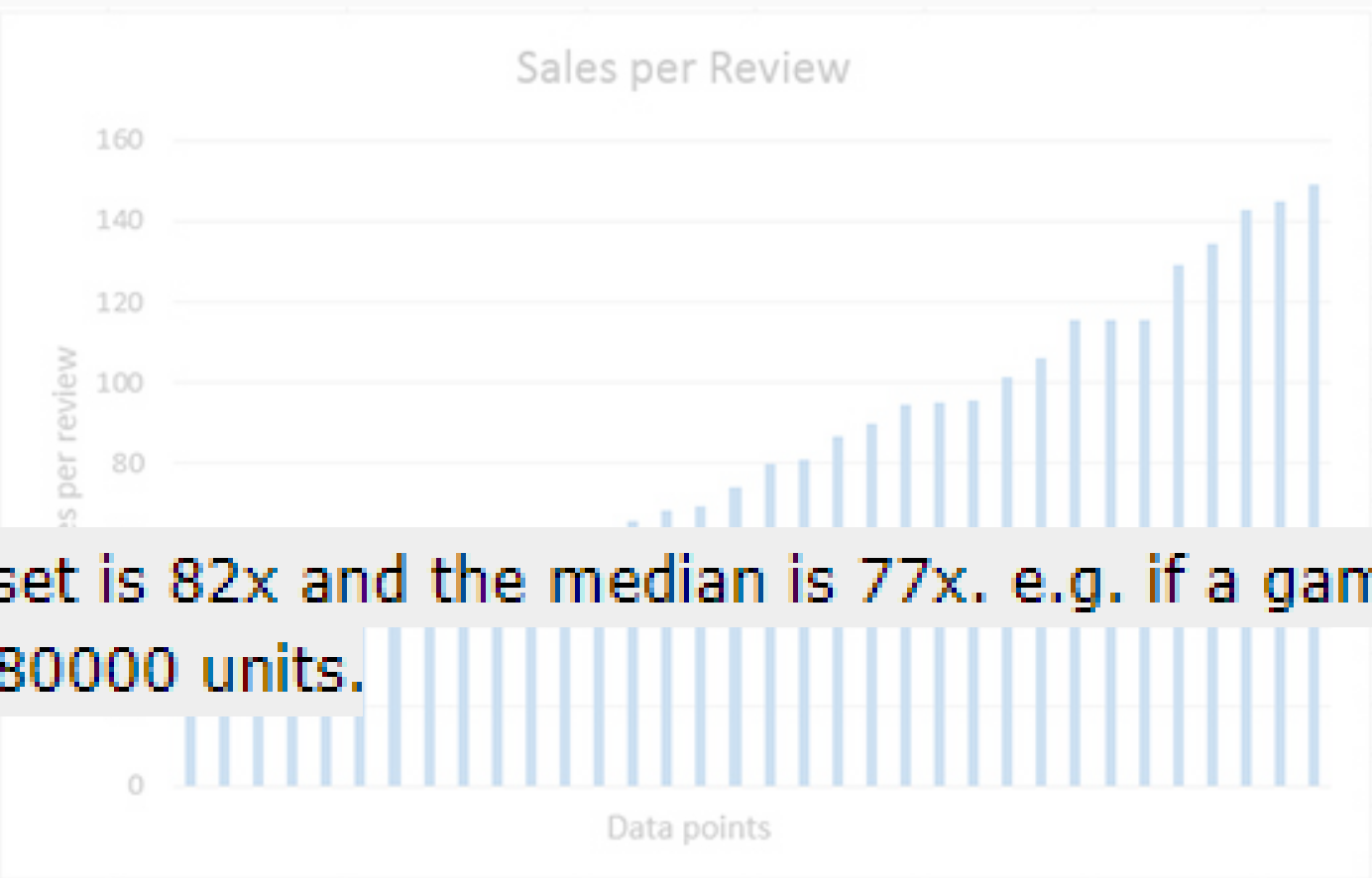
Online video game resellers to increase their profit margin and more efficient resource allocation.

Using Steam reviews to estimate sales

by Jake Birkett on 05/04/18 04:03:00 am Expert Blogger Featured Post

5 comments   Share 

The following blog post, unless otherwise noted, was written by a member of Gamasutra's community. The thoughts and opinions expressed are those of the writer and not Gamasutra or its parent company.



The average from my dataset is 82x and the median is 77x. e.g. if a game has 1000 reviews you can estimate it has sold about 80000 units.

Back in 2014, before SteamSpy, I wrote [this article](#) about using Steam's review count to estimate sales. Then SteamSpy came along and we all used that instead.

However, now that SteamSpy is unable to show the number of owners due to Valve's privacy policy changes, the good old review count method has become valid again!

I decided it was worth getting an updated "Boxleiter" number (how much you should multiply a game's review count by in order to estimate units sold). So as per usual I asked a bunch of indies to supply accurate data. This time I got over 50 data points. Read on to discover my conclusions!

The Headline

The average from my dataset is 82x and the median is 77x. e.g. if a game has 1000 reviews you can estimate it has sold about 80000 units. But please keep reading as there are some important things to note.



Jim Lahey ▾ Nov 9, 2016 @ 7:33pm

Who sets the price of games and determines sales? Valve or developer / publisher?

Lets be very clear

Publishers/Developers set prices

If you see a price on Steam, the publisher/developer set that.

Devs Have Control Over Sale Prices

Devs control the sale price of their games