

Discount Percentage Prediction

Linear Regression

TABLE OF CONTENTS

1

INTRODUCTION

Problem description.

2

Data

Source, features, added features.

3

EDA

Visualizing the data.

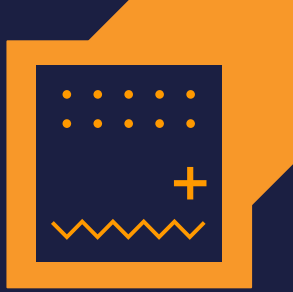
4

Experiments & Results

Discussing different used models and their results.

01

Introduction



**find out what influences the discount
percentage of video games on Steam**



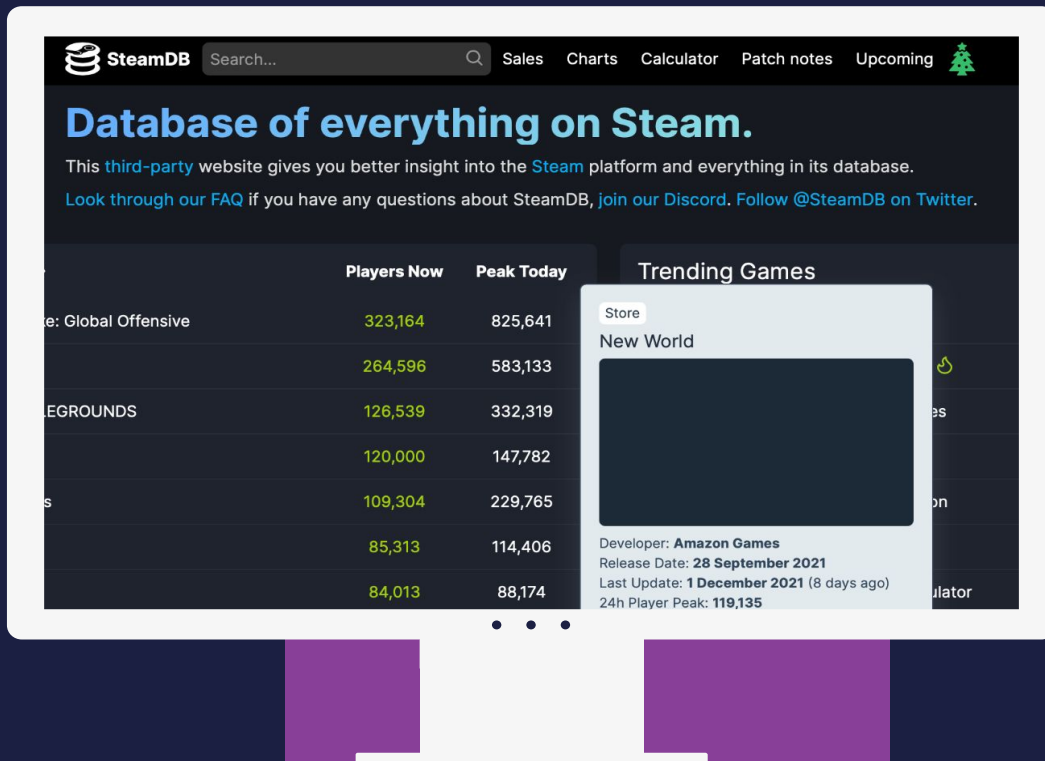
02

Data

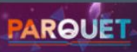











Data Source

SteamDB :
third-party website for
steam stats.





	Name		%	Price	Rating	Ends in	Started	Release
	PARQUET Daily Deal	lowest price is 25.56 SR at -20%	-15%	27.15 SR	90.58%	in 12 hours	2 days ago	3 months ago
	Anno 1404 - History Edition Daily Deal		-50%	34.64 SR	78.91%	in 4 days	3 days ago	2 years ago
	Call of Cthulhu® Daily Deal	new historical low	-70%	23.70 SR	76.30%	in 2 days	12 hours ago	3 years ago
	Tennis Manager 2021 Daily Deal	new historical low	-50%	31.49 SR	72.13%	in 2 days	12 hours ago	3 months ago
	Anno 2070™ - The Financial Crisis Package Daily Deal		-75%	9.21 SR	68.97%	in 4 days	2 days ago	10 years ago
	Anno 2205™ Daily Deal	lowest price is 42.25 SR at -75%	-75%	46.25 SR	68.19%	in 4 days	3 days ago	6 years ago
	Anno 2070 DLC Complete Pack Daily Deal		-75%	17.32 SR	63.91%	in 4 days	2 days ago	9 years ago
	Anno 2070 - The Eden Complete Package Daily Deal		-75%	9.21 SR	61.95%	in 4 days	2 days ago	10 years ago
	Anno 2070 Complete Edition Daily Deal		-75%	34.75 SR	59.77%	in 4 days	2 days ago	-
	Anno 2070™		-75%	33.83 SR	59.73%	in 4 days	2 days ago	10 years ago





	name	price	rating	dis_ends	dis_start	release	discount
0	Little Inferno\nnew historical low	2.99	91.28	in 3 days	4 hours ago	9 years ago	0.90
1	Human Resource Machine\nnew historical low	7.48	90.11	in 3 days	4 hours ago	6 years ago	0.75
2	HIVESWAP: ACT 1\nnew historical low	6.28	89.56	in 2 days	4 hours ago	4 years ago	0.65
3	7 Billion Humans\nnew historical low	7.48	88.34	in 3 days	4 hours ago	3 years ago	0.75
4	Pesterquest\nnew historical low	8.48	86.06	in 2 days	4 hours ago	2 years ago	0.66





× Calculated Original Price

	name	price	rating	dis_ends	dis_start	release	discount	recent	op
0	Little Inferno	2.99	91.28	in 3 days	4 hours ago	9 years ago	0.90	0	87.01
1	Human Resource Machine	7.48	90.11	in 3 days	4 hours ago	6 years ago	0.75	0	67.52
2	HIVESWAP: ACT 1	6.28	89.56	in 2 days	4 hours ago	4 years ago	0.65	0	58.72
3	7 Billion Humans	7.48	88.34	in 3 days	4 hours ago	3 years ago	0.75	1	67.52
4	Pesterquest	8.48	86.06	in 2 days	4 hours ago	2 years ago	0.66	1	57.52



× Added Recency Feature

	name	price	rating	dis_ends	dis_start	release	discount	recent	op
0	Little Inferno	2.99	91.28	in 3 days	4 hours ago	9 years ago	0.90	0	87.01
1	Human Resource Machine	7.48	90.11	in 3 days	4 hours ago	6 years ago	0.75	0	67.52
2	HIVESWAP: ACT 1	6.28	89.56	in 2 days	4 hours ago	4 years ago	0.65	0	58.72
3	7 Billion Humans	7.48	88.34	in 3 days	4 hours ago	3 years ago	0.75	1	67.52
4	Pesterquest	8.48	86.06	in 2 days	4 hours ago	2 years ago	0.66	1	57.52

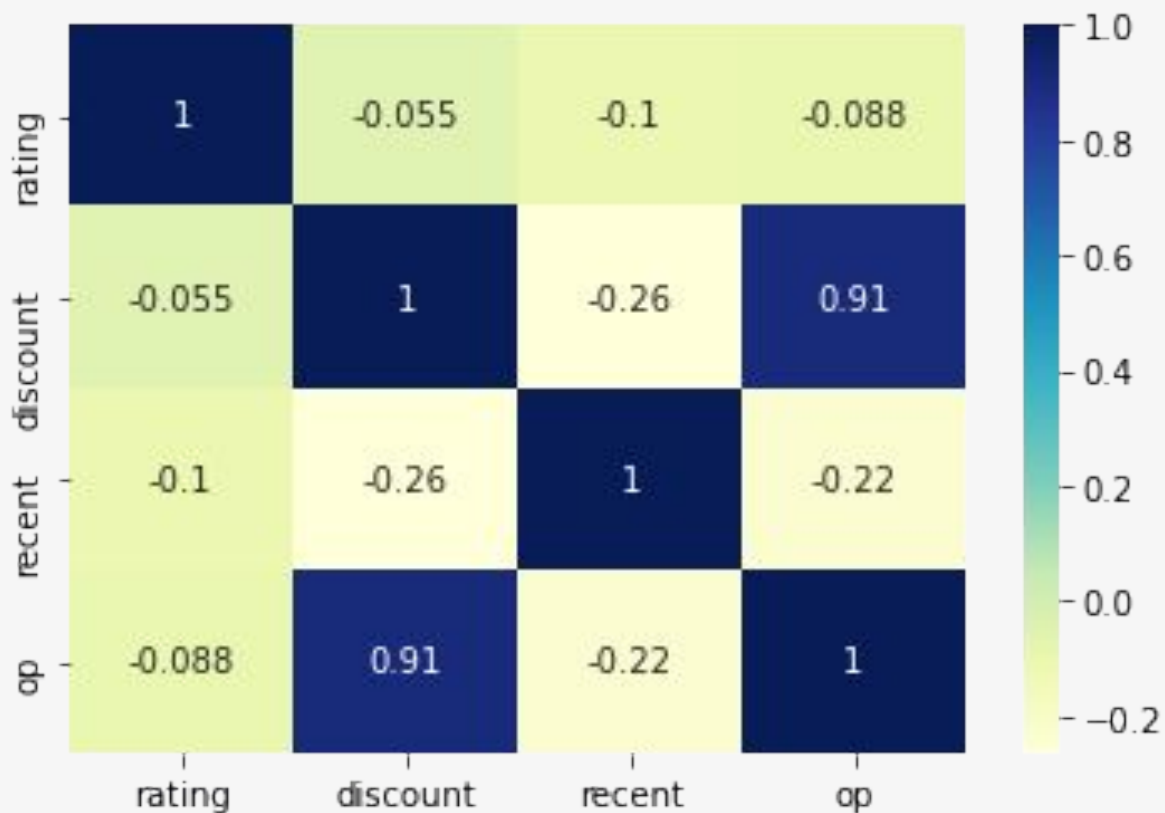


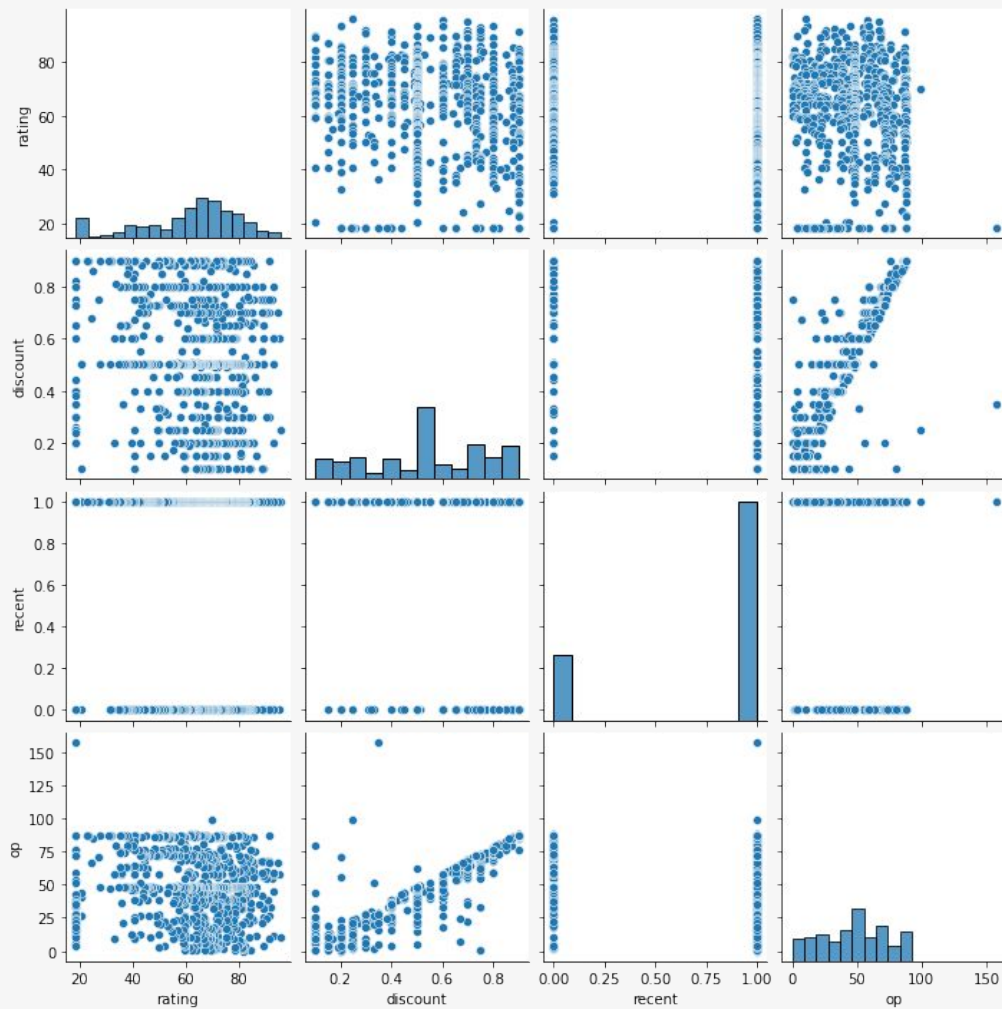
8 Features

978 Row

03

EDA





04

Experiments & Results



BaseLine Model



Numerical Only

- Original Price.
- Rating

```
regressor.score(X_train, y_train)
```

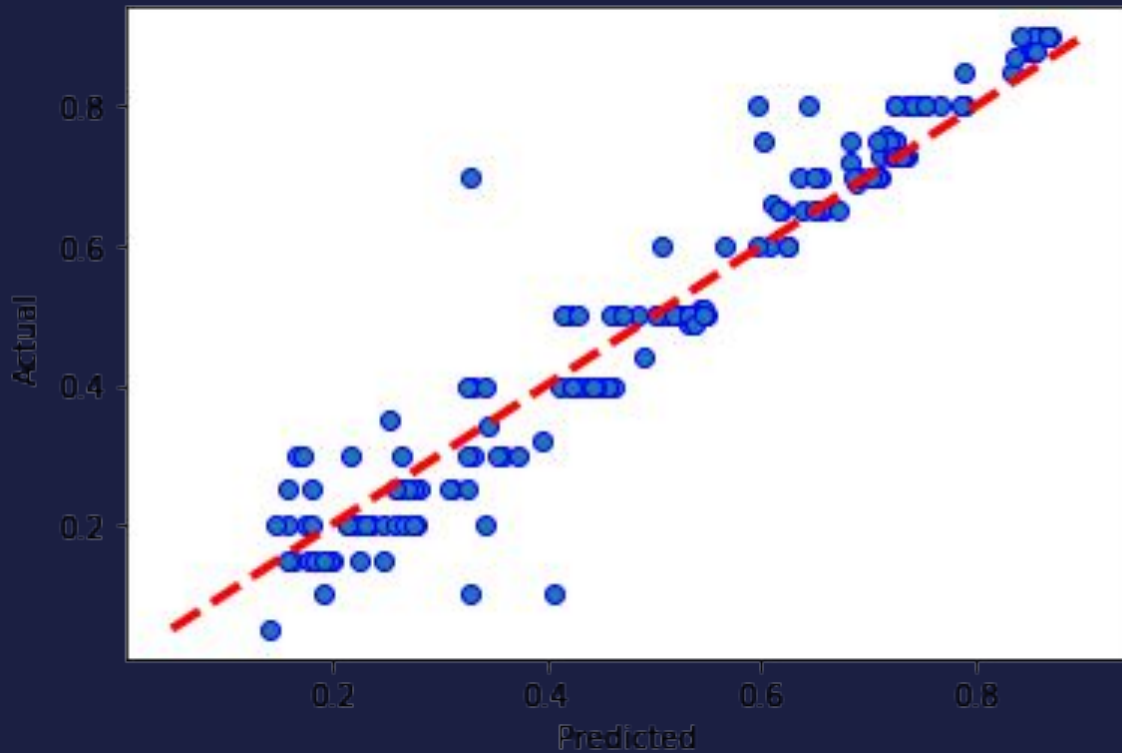
```
0.8194440052825537
```

```
regressor.score(X_val, y_val)
```

```
0.8197733391363717
```



BaseLine Model





LR Model



features:

- Original Price.
- Rating
- Recent

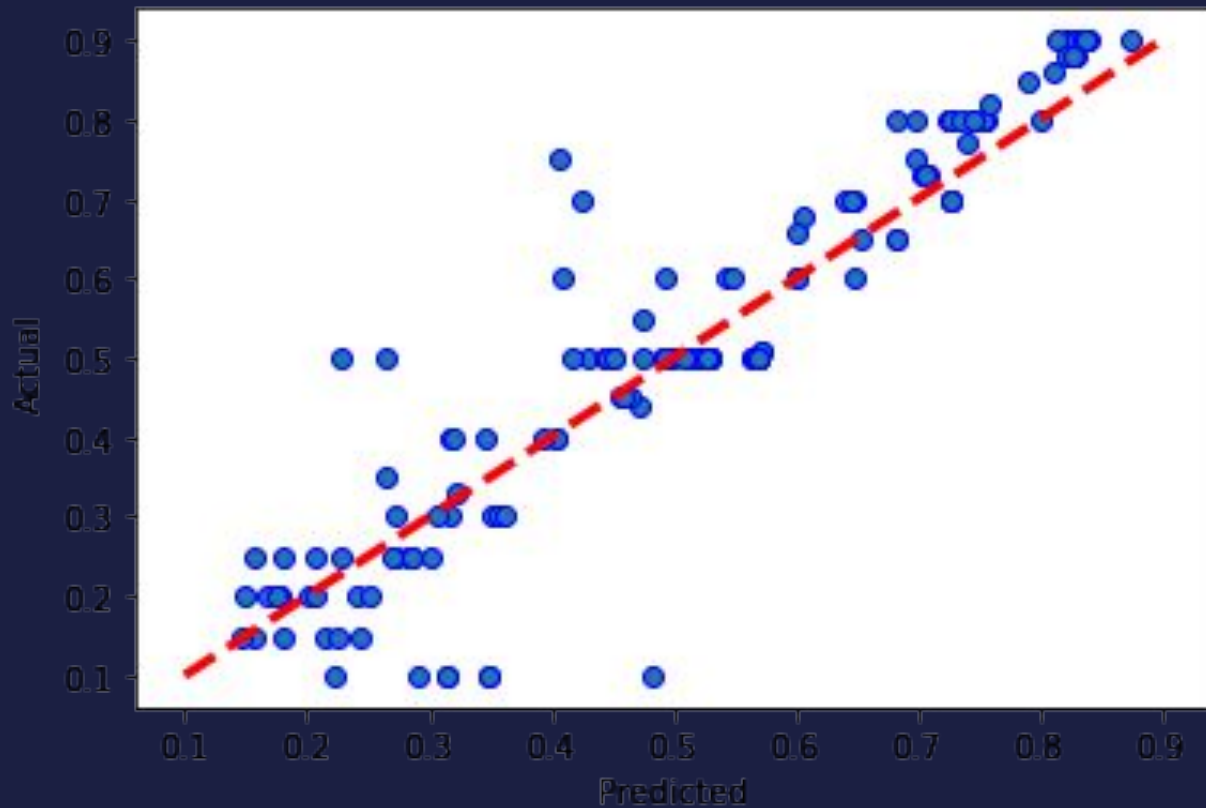
```
regressor.score(X_val, y_val)
```

```
0.8501689426936717
```

```
regressor.score(X_test, y_test)
```

```
0.8713686592317529
```







LR Model



mean	0.528227
std	0.231439
min	0.050000
25%	0.400000
50%	0.500000
75%	0.730000
max	0.900000





LR Model



The model performance for testing set

MAE is 0.060509529749583284

MSE is 0.007502251015179731





Polynomial Regression Model



features:

- Original Price.
- Rating
- Recent

```
polyreg_scaled.score(X_train, y_train)
```

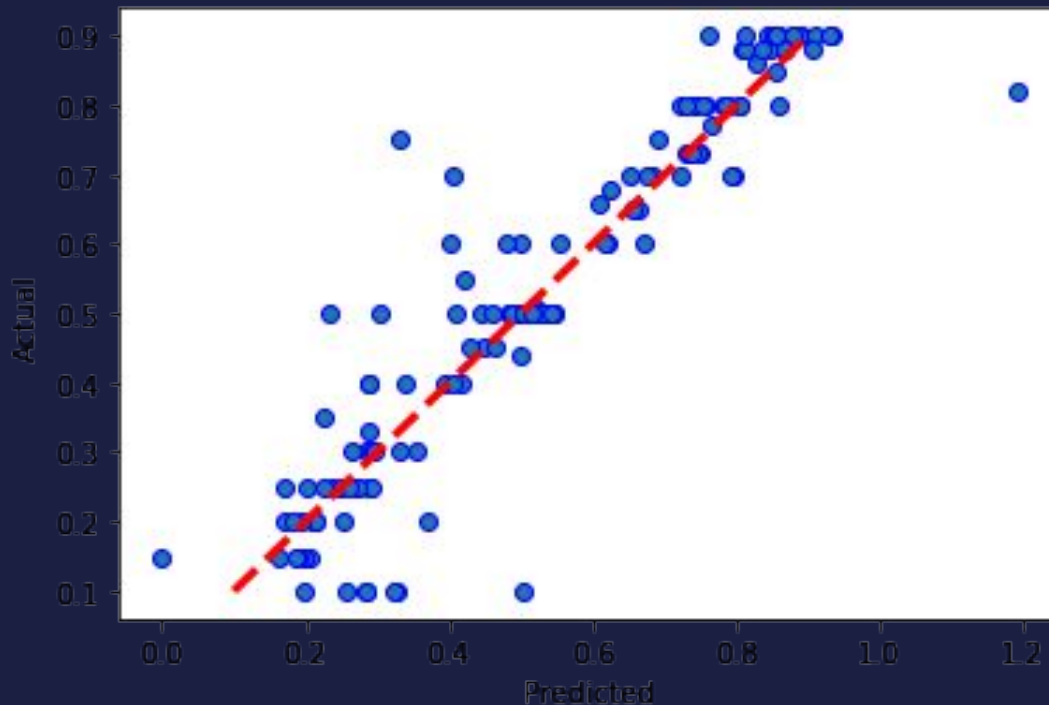
```
0.8831594087907316
```

```
polyreg_scaled.score(X_val, y_val)
```

```
0.8352089022974507
```



Polynomial Regression Model





Polynomial Regression Model

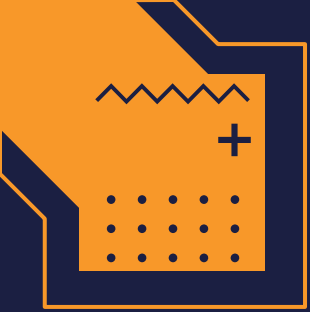


The model performance for testing set

MAE is 0.05377827247125458

MSE is 0.008112178227070834





Thank You!

Any Questions?

