

Abdulelah Almajed



# STEAM PRICE PREDICTION

Regression, Machine Learning Project

# Content

01 Introduction

02 Dataset

03 Tools

07 Conclusion

04 Data Wrangling

05 EDA

06 Models

# Steam

Steam is a popular video game digital distribution service where gamers view and download games of their interest.

64% of US adults and 70% of US minors are considered gamers.



# Scraped Dataset

```
In [14]: scrape_df.head(5)
```

```
Out[14]:
```

Unnamed: 0		Title	Price	Release Date	Game Rating	Users Reviewed
0	0	Counter-Strike: Global Offensive	Free to Play	21 Aug, 2012	88%	6,038,424
1	1	Apex Legends™	Free to Play	4 Nov, 2020	86%	351,931
2	2	Red Dead Redemption 2	114.50 SR	5 Dec, 2019	86%	197,785
3	3	Farming Simulator 22	179.00 SR	21 Nov, 2021	85%	6,973
4	4	Forza Horizon 5	249.00 SR	8 Nov, 2021	85%	36,986

# Another Dataset

Out[15]:

	english	developer	publisher	platforms	required_age	categories	genres	steamspy_tags	achievements	positive_ratings	negative_ratings	avera
	1	Valve	Valve	windows;mac;linux	0	Multi-player;Online Multi-Player;Local Multi-P...	Action	Action;FPS;Multiplayer	0	124534	3339	
	1	Valve	Valve	windows;mac;linux	0	Multi-player;Online Multi-Player;Local Multi-P...	Action	Action;FPS;Multiplayer	0	3318	633	
	1	Valve	Valve	windows;mac;linux	0	Multi-player;Valve Anti-Cheat enabled	Action	FPS;World War II;Multiplayer	0	3416	398	
	1	Valve	Valve	windows;mac;linux	0	Multi-player;Online Multi-Player;Local Multi-P...	Action	Action;FPS;Multiplayer	0	1273	267	
	1	Gearbox Software	Valve	windows;mac;linux	0	Single-player;Multi-player;Valve Anti-Cheat en...	Action	FPS;Action;Sci-fi	0	5250	288	

# Tools

Jupyter Notebook

Pandas

Numpy

Beautiful Soup & Selenium

Matplotlib & Seaborn

sklearn

# Data Wrangling

Data Arranging, Data Cleaning

## Step 1

Merge two Dataset

## Step 2

remove Null Values  
& replace null values  
with mode

## Step 3

Create Dummy Variables

# Dataframe info before dummies

669 Observations

15 Columns

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 669 entries, 0 to 696
Data columns (total 15 columns):
#   Column                Non-Null Count  Dtype
---  -
0   name                  669 non-null   object
1   price                 669 non-null   float64
2   release_year          669 non-null   int64
3   game_rating           669 non-null   float64
4   platforms             669 non-null   object
5   users_reviewed        669 non-null   int64
6   achievements          669 non-null   int64
7   average_playtime      669 non-null   int64
8   median_playtime       669 non-null   int64
9   genres                669 non-null   object
10  owners                669 non-null   object
11  required_age          669 non-null   int64
12  english               669 non-null   int64
13  category1            669 non-null   object
14  genre                 669 non-null   object
dtypes: float64(2), int64(7), object(6)
memory usage: 83.6+ KB
```



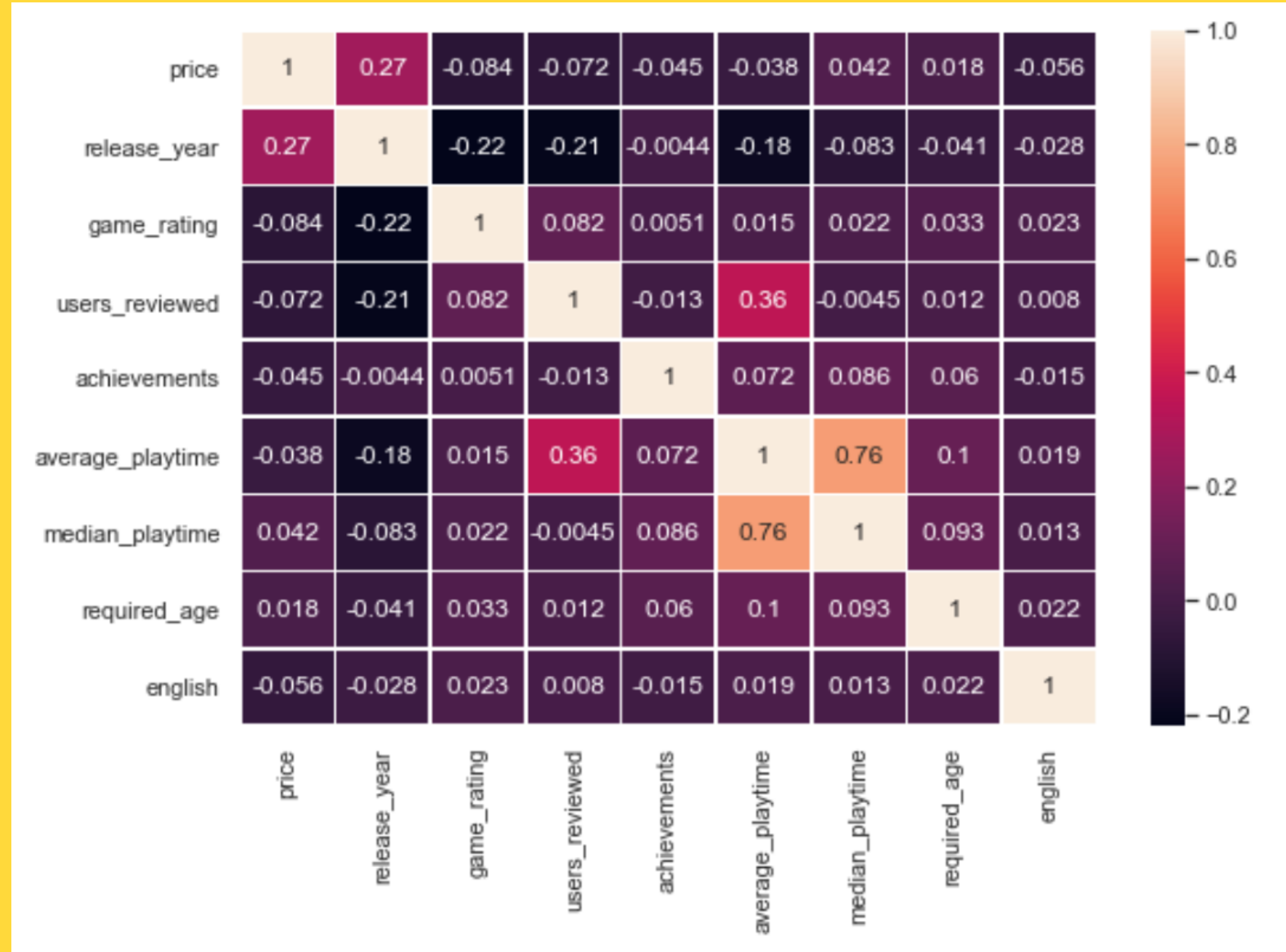
# Dataframe info after dummies

669 Observations

60 Columns

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 669 entries, 0 to 696
Data columns (total 60 columns):
#   Column                                                                 Non-Null Count  Dtype
---  -
0   name                                                                    669 non-null    object
1   price                                                                    669 non-null    float64
2   release_year                                                            669 non-null    int64
3   game_rating                                                             669 non-null    float64
4   users_reviewed                                                          669 non-null    int64
5   achievements                                                            669 non-null    int64
6   average_playtime                                                       669 non-null    int64
7   median_playtime                                                        669 non-null    int64
8   genres                                                                  669 non-null    object
9   required_age                                                            669 non-null    int64
10  english                                                                  669 non-null    int64
11  platforms_windows                                                       669 non-null    int64
12  platforms_windows;linux                                                 669 non-null    int64
13  platforms_windows;mac                                                  669 non-null    int64
14  platforms_windows;mac;linux                                            669 non-null    int64
15  category1_Co-op                                                         669 non-null    int64
16  category1_Cross-Platform Multiplayer  669 non-null    int64
17  category1_Full controller support  669 non-null    int64
18  category1_Local Co-op                                                  669 non-null    int64
19  category1_Local Multi-Player                                           669 non-null    int64
20  category1_MMO                                                           669 non-null    int64
21  category1_Multi-player                                                 669 non-null    int64
22  category1_Online Co-op                                                 669 non-null    int64
23  category1_Online Multi-Player                                           669 non-null    int64
24  category1_Partial Controller Support  669 non-null    int64
25  category1_Shared/Split Screen  669 non-null    int64
26  category1_Steam Achievements  669 non-null    int64
27  category1_Steam Cloud                                                  669 non-null    int64
28  category1_Steam Leaderboards  669 non-null    int64
29  category1_Steam Trading Cards  669 non-null    int64
30  category1_Steam Workshop                                              669 non-null    int64
31  genre_Action                                                            669 non-null    int64
32  genre_Adventure                                                         669 non-null    int64
33  genre_Casual                                                            669 non-null    int64
34  genre_Design & Illustration  669 non-null    int64
35  genre_Early Access                                                     669 non-null    int64
36  genre_Free to Play                                                     669 non-null    int64
37  genre_Gore                                                              669 non-null    int64
38  genre_Indie                                                             669 non-null    int64
39  genre_Massively Multiplayer  669 non-null    int64
40  genre_Nudity                                                            669 non-null    int64
41  genre_RPG                                                              669 non-null    int64
42  genre_Racing                                                            669 non-null    int64
43  genre_Simulation                                                       669 non-null    int64
44  genre_Sports                                                            669 non-null    int64
45  genre_Strategy                                                         669 non-null    int64
46  genre_Violent                                                           669 non-null    int64
47  genre_Web Publishing                                                    669 non-null    int64
48  owners_0-20000                                                         669 non-null    int64
49  owners_100000-200000                                                   669 non-null    int64
50  owners_1000000-2000000                                                 669 non-null    int64
51  owners_10000000-20000000                                               669 non-null    int64
52  owners_100000000-200000000                                             669 non-null    int64
53  owners_20000-50000                                                     669 non-null    int64
54  owners_200000-500000                                                   669 non-null    int64
55  owners_2000000-5000000                                                 669 non-null    int64
56  owners_20000000-50000000                                               669 non-null    int64
57  owners_50000-100000                                                    669 non-null    int64
58  owners_500000-1000000                                                  669 non-null    int64
59  owners_5000000-10000000                                                669 non-null    int64
dtypes: float64(2), int64(56), object(2)
memory usage: 318.8+ KB
```

# Heatmap



# Dummies variables

Categories

was about 16  
column

Genres

was about 17  
column

Platforms

was about 4  
column

Owners

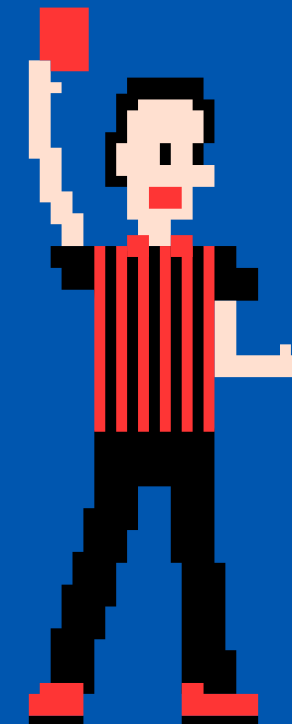
was about 12  
column

# Linear Regression

Training and testing  $R^2$

Training  $R^2$  was : 0.18

Testing  $R^2$  was: 0.045

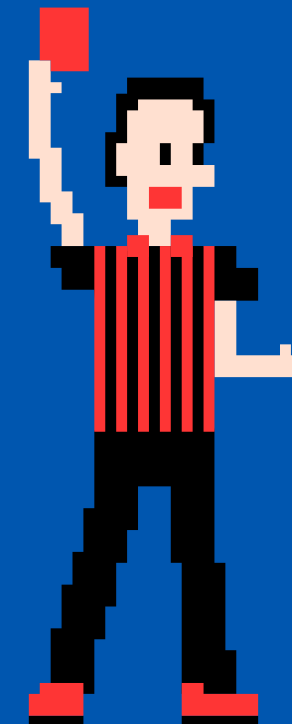


# Ridge Regression

Training and testing  $R^2$

Training  $R^2$  was : 0.024

Testing  $R^2$  was: 0.013



# Gradient Boosting Regression

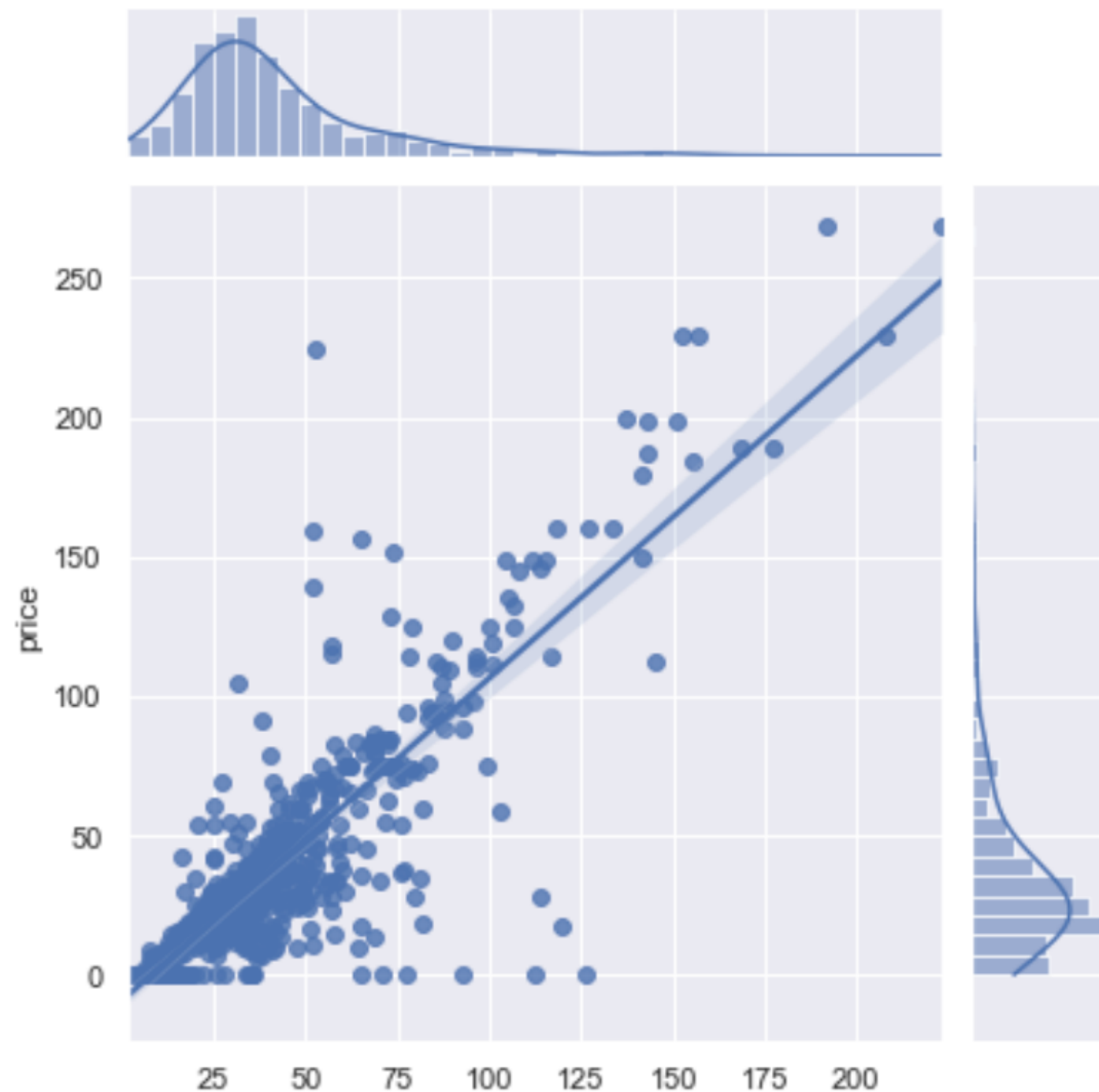
Training and testing  $R^2$

Training  $R^2$  was : 0.67

Testing  $R^2$  was: 0.91



# Best Model: Random Forest Regression



Training  $R^2 = 0.89$

Testing  $R^2 = 0.88$

Mean Square Error = 1385.0406909248802

Root Mean Square Error = 37.21613481979127

Mean Absoluter Error= 26.2020583333333326

THANK YOU!