Predicting Game Sales

Vitoria Calil

Why predicting game sales?

- Improve relationship between GD Companies and affiliate stores
- Decrease spendings, increase revenue

Data Acquisition
&
Data Processing

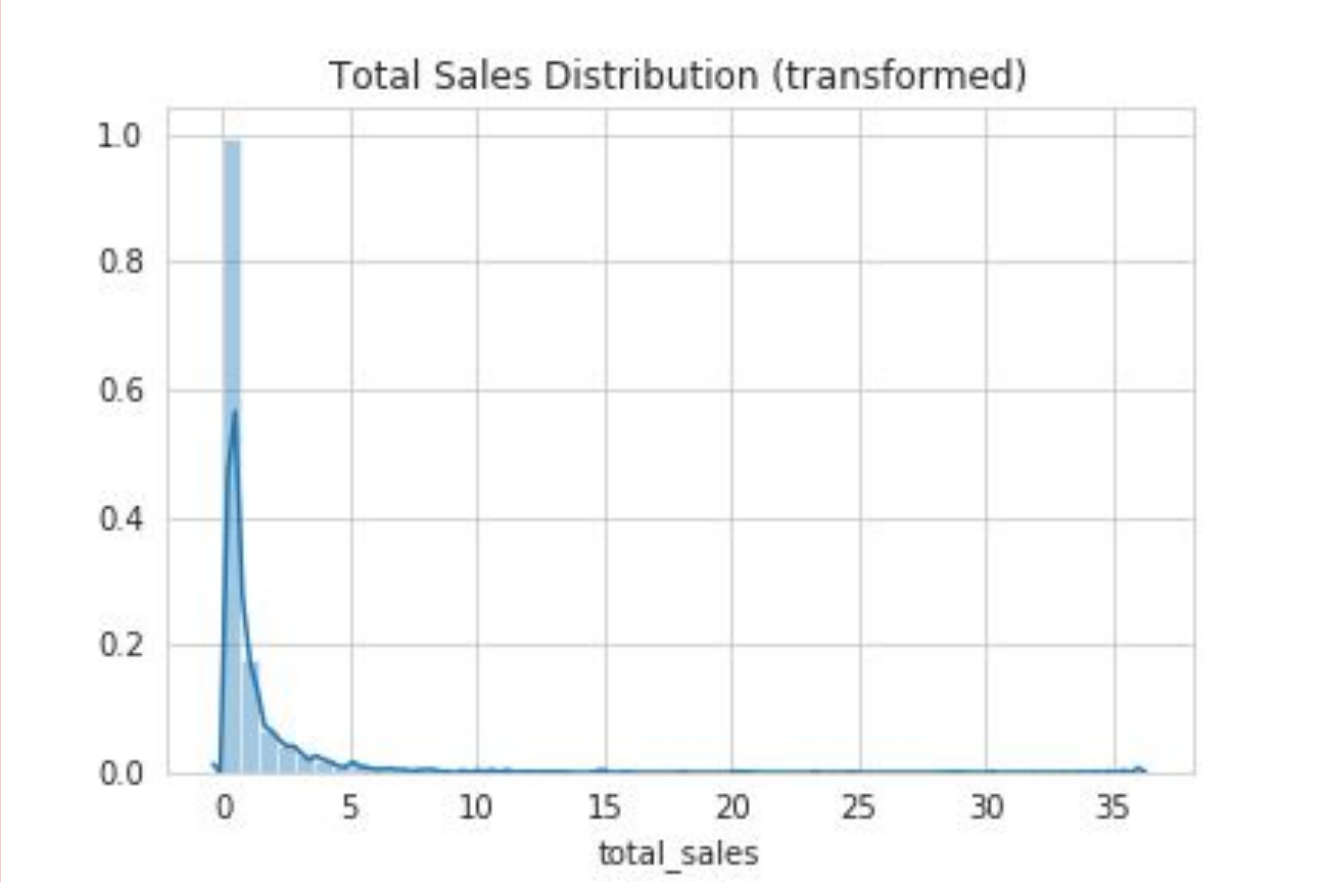
Data Acquired:

Scrapy

- 2369 games
- 27 different consoles
- 190 publishers
- 16 genres

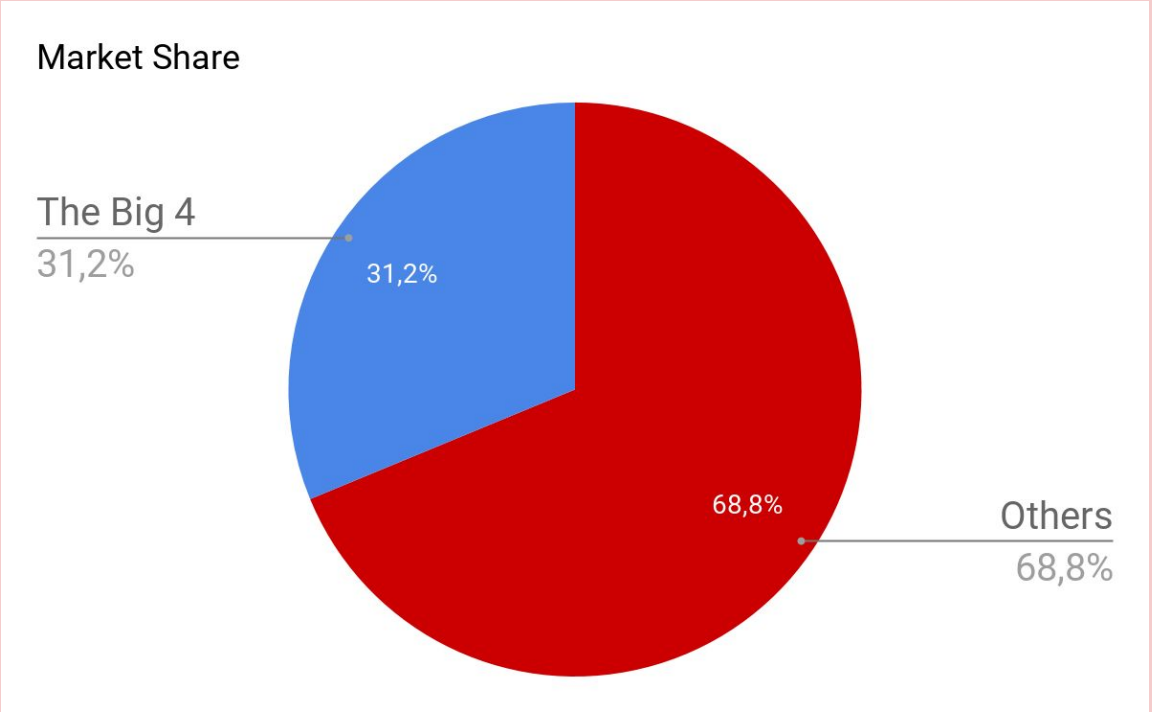
VGCHARTZ

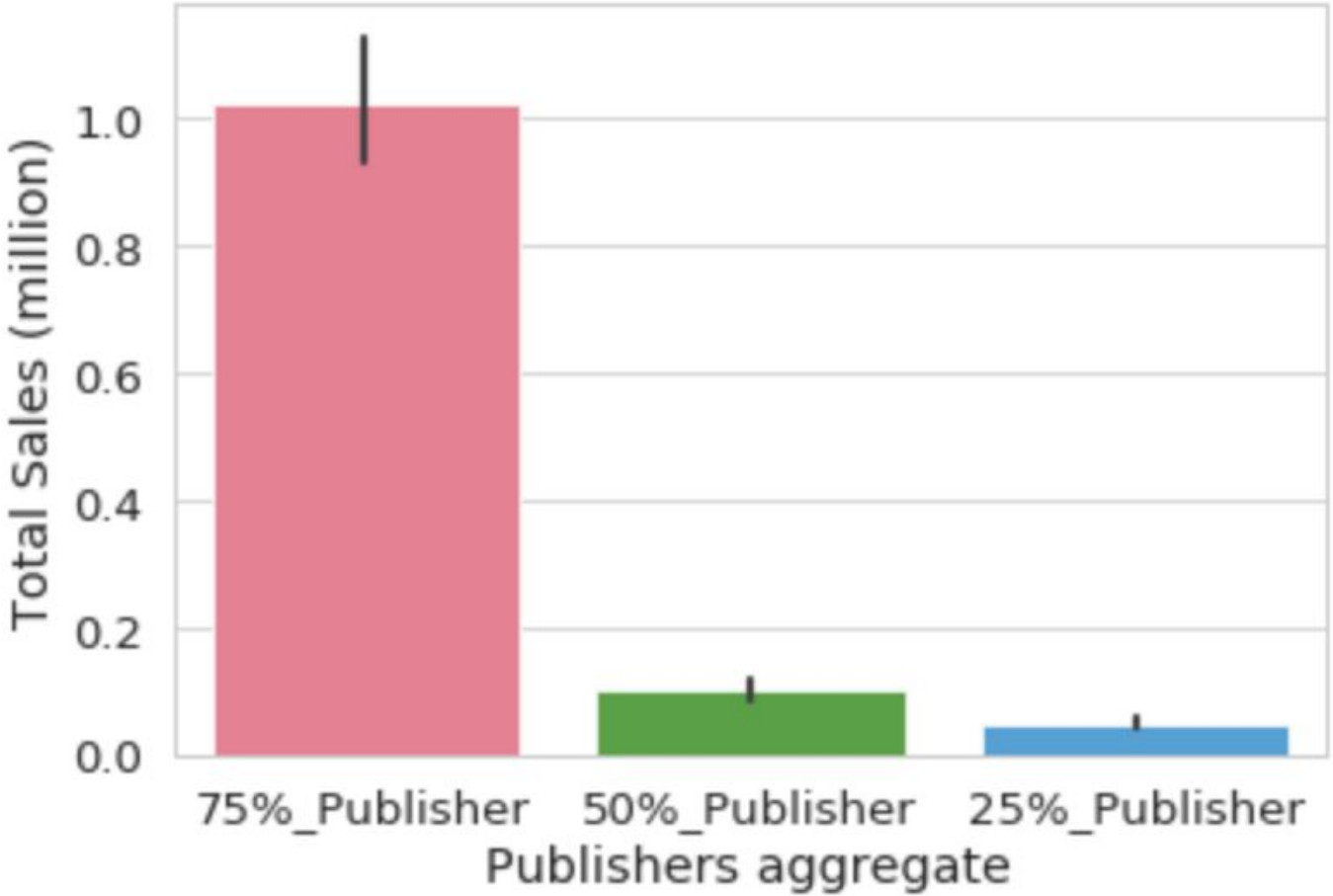
Target distribution



Total Sales Distribution (transformed)
2.0
1.5
1.0
0.5
0.0
0
1
2
3
4
log_total_sales

Games Industry




Total Sales (million)
1.0
0.8
0.6
0.4
0.2
0.0
75%_Publisher
50%_Publisher
25%_Publisher
Publishers aggregate

Modeling

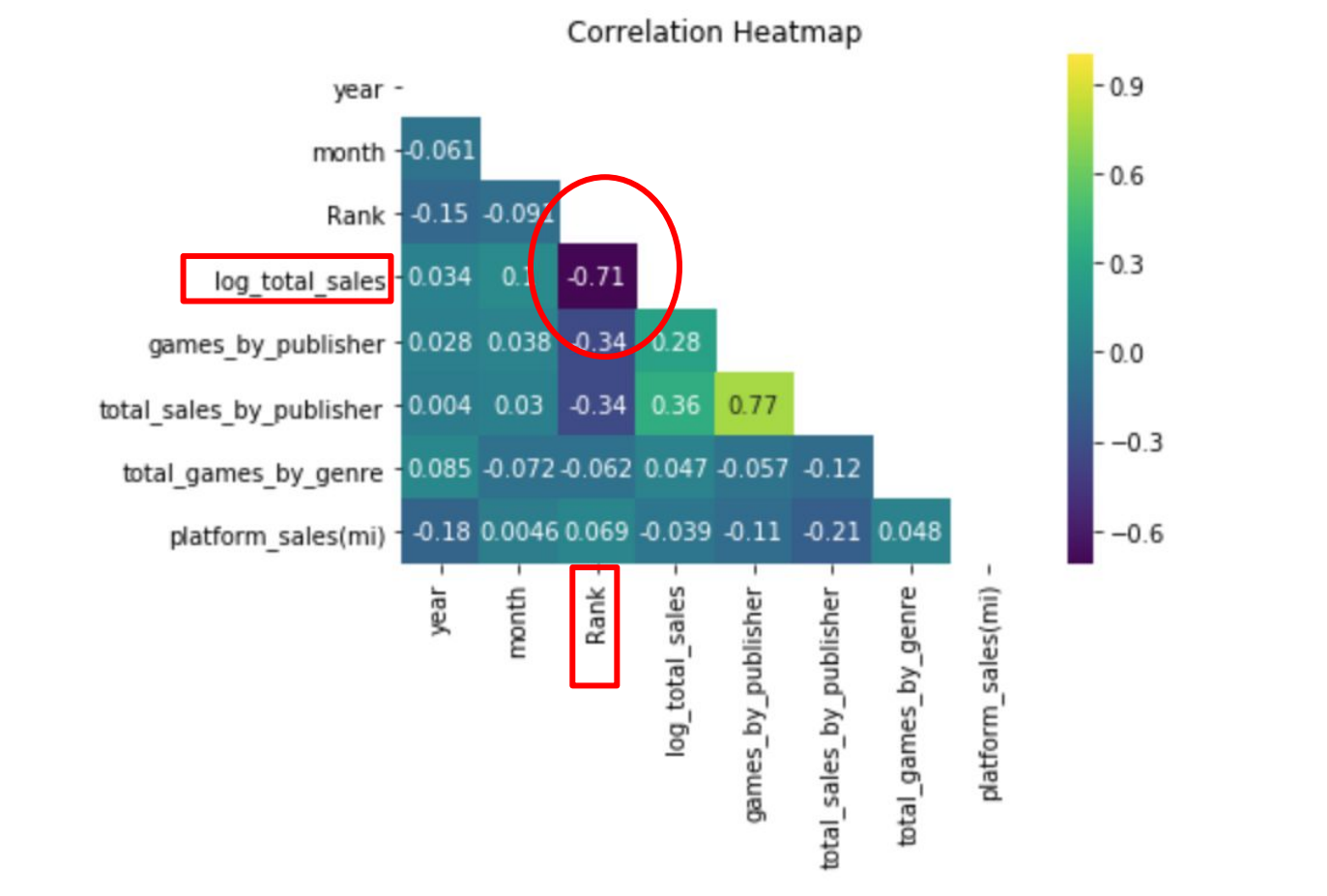
Features

VG Chartz:

- Rank position
- Console
- Release Date
- Game Publisher
- Ratings

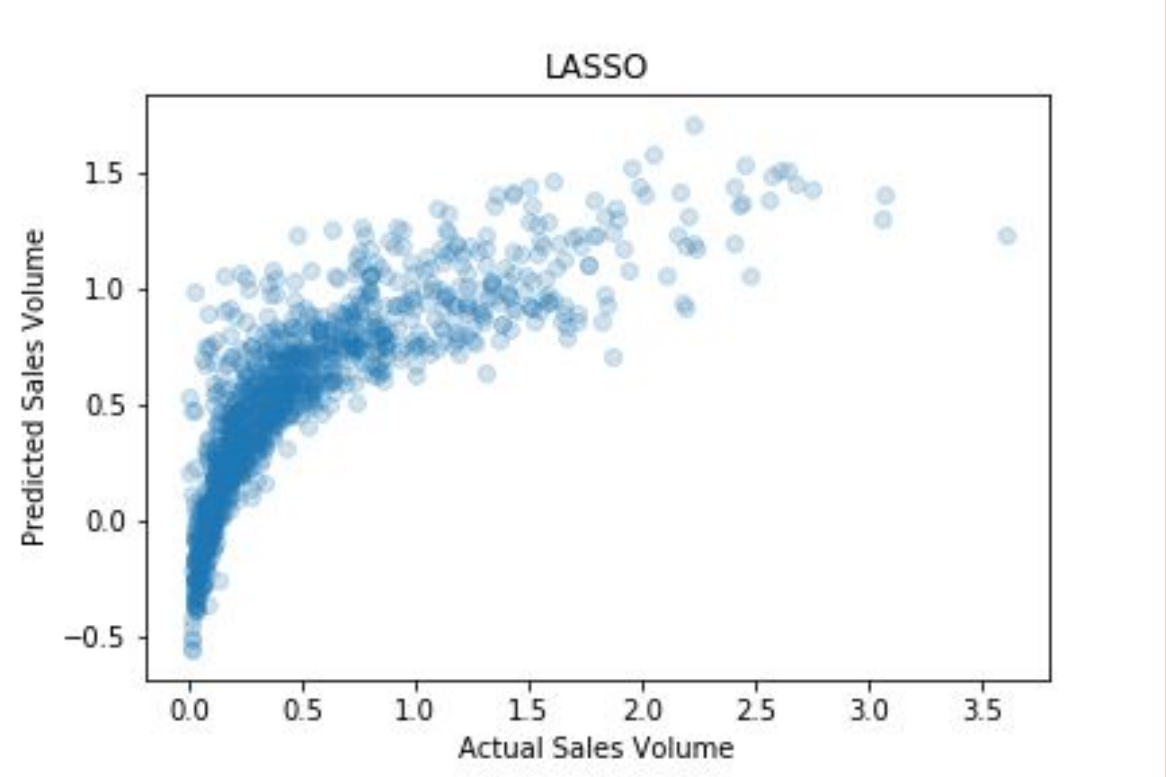
Other sources:

- Games by Publisher
- Total Sales by Publisher
- Games by Genre
- Console sales


Correlation Heatmap
year
month
Rank
log_total_sales
games_by_publisher
total_sales_by_publisher
total_games_by_genre
platform_sales(mi)
-0.061
-0.15
-0.71
0.034
0.028
0.038
-0.34
0.28
0.004
0.03
-0.34
0.36
0.77
0.085
-0.072
-0.062
0.047
-0.057
-0.12
-0.18
0.0046
0.069
-0.039
-0.11
-0.21
0.048
0.9
0.6
0.3
0.0
−0.3
−0.6

How good am I in precting sales volume?

How good am I in precting sales volume?

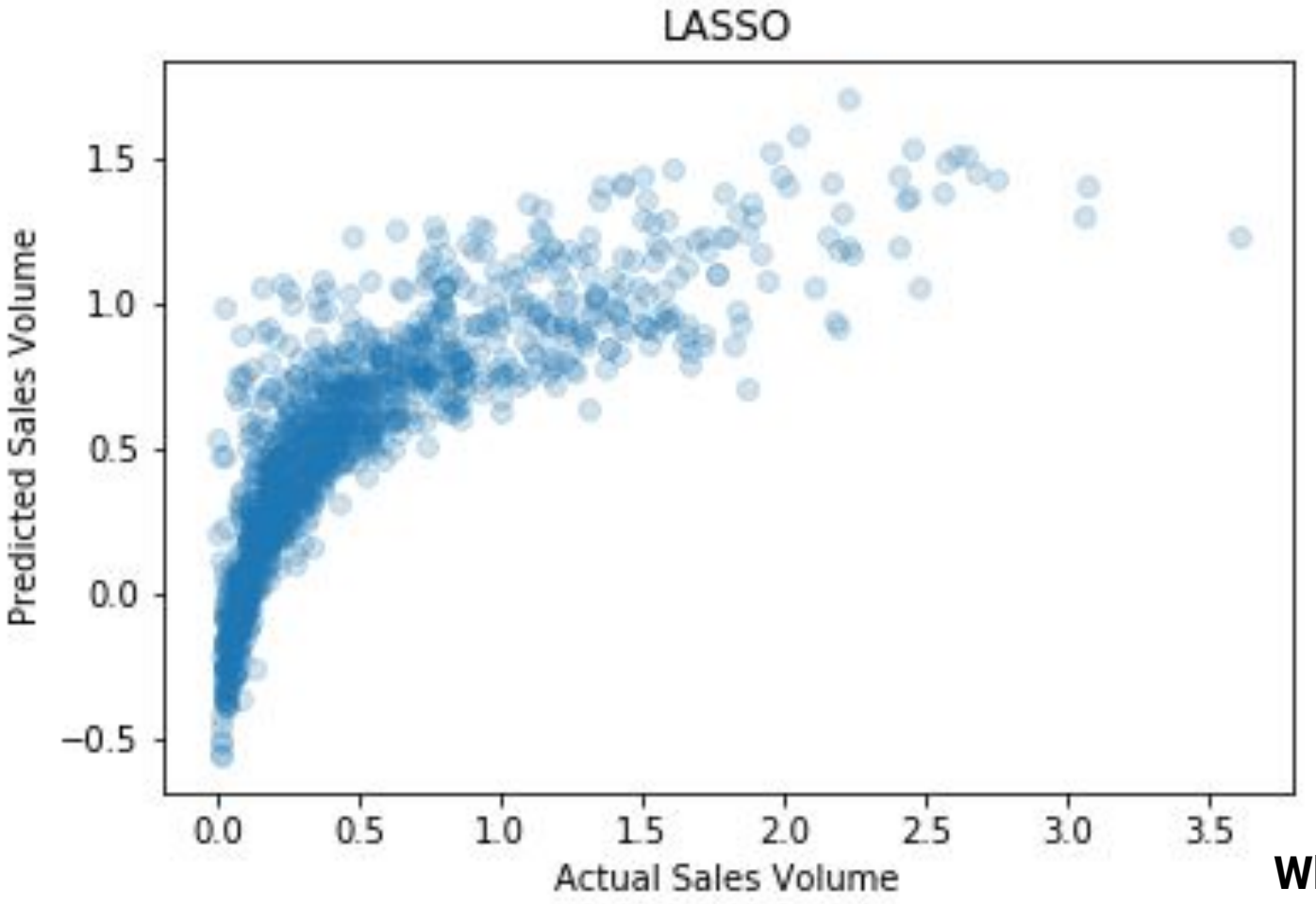


Linear Model
LASSO Regulariztion

Training set:
R2 - 58.88%

Test set:
R2 - 56.03%

Holdout set:
R2 - 46.69%



Why Keeping outliers?

Most relevant features

Publishers:

- Nintendo
- Take-Two Interactive
- Sony Ent.
- Activision

Others:

- Ranking
- Ratings MATURE
- Platform PS4

Least relevant features

- Small publishers
- Platform PSP

Next steps...

- Improve the model to the top 10% publishers
- More features (Sales by region, Marketing Budget, Theme of the game)
- Use advanced models

IT'S FRIDAY!!!

SLOTHILDA.COM