



The Sneaker Game: Datafied

By Marcos Dominguez, Data Scientist
and Sneakerhead





Sneaker culture

- What is a sneakerhead?
- Sports history
- Community
- Swag

The sneaker market is like no other

- What's with the crazy prices?
- Tired of getting ripped off?
- Is there a better way to price sneakers?



Have no fear, linear regression is here!



I have a better way of pricing sneakers to ensure you pay a reasonable price!

Features used to predict prices

- Age of sneaker (in years)
- Sales Volume
- Price over Premium
- Volatility
- Brand Name
 - Jordan, Nike, Adidas, or Other

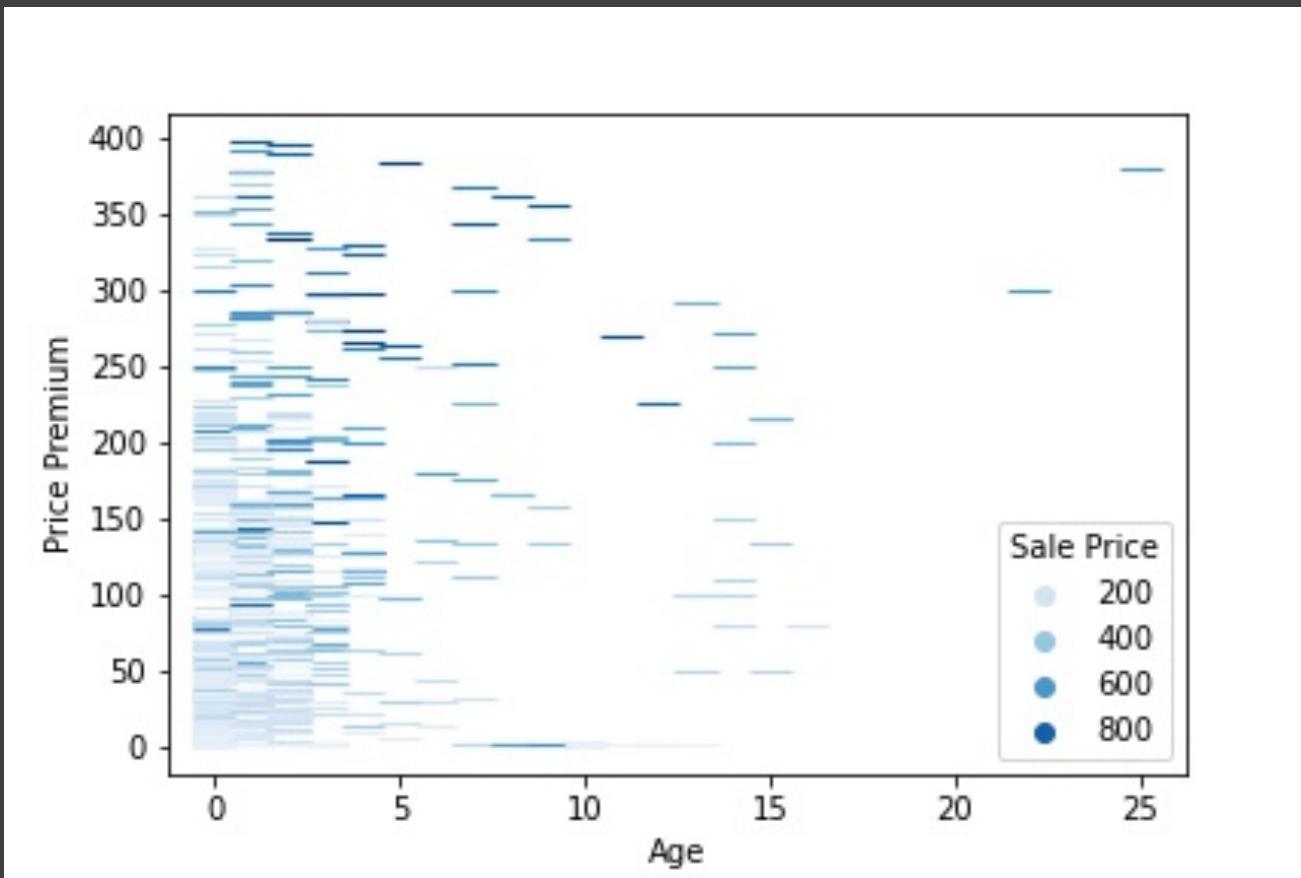


Method: Linear Regression

- Tested 2 models
 - OLS
 - Polynomial
- Two variations
 - Without Ridge regularization
 - With Ridge regularization
- Feature engineering to correct overfitting



Feature Engineering



Multiplicative Interaction: Price Premium & Age of sneaker

Results before feature engineering

Model	Regularization	RMSE	Overfit?
OLS Method	None	\$176	Yes
	Ridge CV	\$176	Yes
Polynomial	None	\$163	Yes
	Ridge CV	\$166	Yes

*dataset was sampled with 5-fold cross validation for all models

Results after feature engineering

Model	Regularization	RMSE	Overfit?
OLS Method	None	\$179	Yes
	Ridge CV	\$178	Yes
Polynomial	None	\$159	No
	Ridge CV	\$164	Yes

*dataset was sampled with 5-fold cross validation for all models

And the winner is...

After feature engineering

Model	Regularization	RMSE	Overfit?
OLS Method	None	\$179	Yes
	Ridge CV	\$178	Yes
Polynomial	None	\$159	No
	Ridge CV	\$164	Yes

*dataset was sampled with 5-fold cross validation for all models

Real World Test



- Predict sale price on a new data point
- Jordan 12 Flu Game Size 10
- Recall :
 - $y = \beta_0 + \beta_1 x_1 + \dots + \beta_k x_k + \epsilon$

Real World Test



Age (Yrs)	Sales Volume	Volatility (%)	Price Premium (%)	Jordan	Predicted Price (\$)
4	265	3.3%	110%	1	\$283

Real World Test



- Predicted price = \$283
- Avg actual price = \$345
- Difference = \$(62)
- NOT BAD!

Conclusion

- Predicting human behavior is extremely difficult
- The model with some complexity tends to be optimal
- This can be a great tool for bargaining a reduced price

