


 ddey117 / Neural_Network_Predicting_Reseller_Success_Ebay

Public

forked from [learn-co-curriculum/dsc-5-capstone-project](#)


 View license


 0 stars  2 forks


☆ Star


👁 Watch ▾


<> Code


 Pull requests


 Actions


 Projects

 Wiki

 Security



 Insights

 Settings

 master ▾

...


This branch is [100 commits ahead](#) of learn-co-curriculum:master.


 Contribute ▾  Sync fork ▾

Dylan Dey create pdf notebooks ...

2 minutes ago 🕒 104

[View code](#)

 README.md



Predicting Resale Value of Knives from a Texas Government Surplus Store

Using Machine Learning to Support an Ebay Store's Financial Success

Project Links

Below is the link for the GitHub project page.

[Github link](#)

Author: Dylan Dey

Overview

[Texas State Surplus Store](#)

[What happens to all those items that get confiscated by the TSA? Some end up in a Texas store.](#)

Texas Surplus Store PDF




Everything that doesn't make it through Texas airports can be found at one Austin store

The Texas Facilities Commission collects left behind possessions, salvage, and surplus from Texas state agencies such as DPS, TXDOT, TCEQ, and Texas Parks & Wildlife. Examples of commonly available items include vehicles, furniture, office equipment and supplies, small electronics, and heavy equipment. The goal of this project is to create a predictive model in order to determine the resale value of knives from the Texas State Surplus Store on eBay to support the financial health of an eBay storefront.

Business Problem

Family Ebay Store Front



texasdave3 (17443 ★)
 100% positive feedback
[Save](#)

[Items for sale](#)
[Visit store](#)
[Contact](#)

Based in United States, texasdave3 has been an eBay member since Nov 18, 1999

Feedback ratings

★★★★★ 1,483	Item as described	<div> <div>2,125</div> <div>Positive</div> </div> <div> <div>3</div> <div>Neutral</div> </div> <div> <div>0</div> <div>Negative</div> </div>	<div> <div>+</div> <div>Good seller.</div> <div>Jun 27, 2022</div> </div>
★★★★★ 1,586	Communication		
★★★★★ 1,577	Shipping time		
★★★★★ 1,593	Shipping charges		

Feedback from the last 12 months

[See all feedback](#)

Texas Dave's Knives

3 Major Painpoints:

1. Selection of an item to sell:

- The item should be readily available in decent condition for the seller to purchase at a low price but not so widely available that the market is saturated with that item.
- there needs to be a demand for the item.
- Be useful to a large number of people even in a used condition.
- Be easy, cost effective and safe to ship.

2. Buying at a cost low enough to make a profit:

- Ebay takes about 13.5% of the sold price for brokering the sale.
- Shipping costs and overhead fees can add up.

3. The cost of excess inventory:

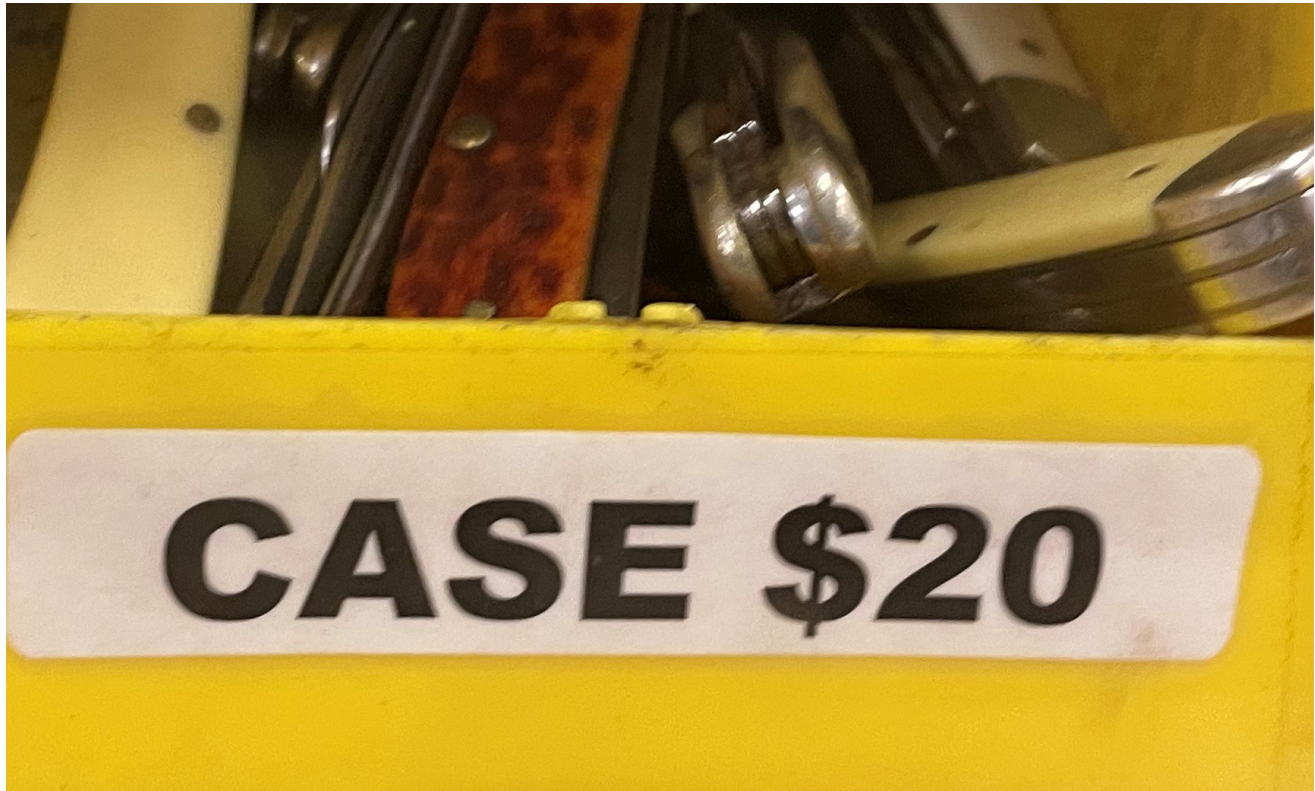
- A seller can obtain quality items at a reasonable cost and then the inventory may sit with no sales, meaning the capital expended is sitting tied up in unwanted items. This inventory carry cost is a drain on profitability.

A potential target that has been selling decently on the eBay store front from random sampling is something that has been overflowing out of bins at the State Surplus Store for years: pocketknives. Does this target answer all of the pain points above?

I have been experimenting with low cost used knives for resale but have not risked a large capital investment in the higher end items. The goal of this project is to attempt to address the pain points to determine if a larger investment would pay off. Can I identify which knives are worth investing in so that I can turn a decent profit and hopefully avoid excess inventory? A data driven approach would help avoid costly mistakes from the "system" currently employed, which seems to be mainly a gambler's approach. By managing resources upfront through a model, I can effectively increase my return on investment with messy data such as pictures and titles.

There are eight buckets of presorted brand knives that I was interested in, specifically. These bins are behind glass, presorted, branded (and therefore have specific characteristics and logos for my model to identify), and priced higher. However, the staff has a very large amount of confiscated items flowing into the facility to list for resale, and when that happens they will not have time to presort them and they end up in huge buckets of unsorted knives for people to dig through. The brands will be priced the same, they are just no longer sorted and harder to find. Expanding the bins to pull inventory from will increase the chance of finding inventory worth reselling, at the cost of time spent sorting the knives.

sorted bucket example



overflow example



Domain Understading: Cost Breakdown

- padded envelopes: \$0.50 per knife
- flatrate shipping: \$4.45 per knife
- brand knife at surplus store: 15, 20, 30, or 45 dollars per knife
- overhead expenses (gas, cleaning suplies, sharpening supplies, etc): \$3.00
- Ebay's comission, with 13% being a reasonable approximation

Brand	Cost at Surplus Store	Cost Adjusted for Expenses
Benchmade	\$45.00	\$52.95 + 13% comission
Buck	\$20.00	\$27.95 + 13% comission
Case	\$20.00	\$27.95 + 13% comission
CRKT	\$15.00	\$22.95 + 13% comission
Kershaw	\$15.00	\$22.95 + 13% comission
SOG	\$15.00	\$22.95 + 13% comission
Spyderco	\$30.00	\$37.95 + 13% comission
Victorinox	\$20.00	\$27.95 + 13% comission

Ebay Developer Website

Ebay has a website for developers to create an account and register an application keyset in order to make API call requests to their live website. By making a `findItemsAdvanced` call to the eBay Finding API Version 1.13.0, I was able to get a large dataset of `category_id=<48818>` knives listed for sale. This data is limited to anything listed within the past 90 days from when the API call was made.

The eBay Finding API Version 1.13.0 `findItemsAdvanced` and The eBay Shopping API Version 1247 `GetMultipleItems` were used to collect data from listed eBay posts.

All of the data gathered from eBay's public API is limited to listed data posted in the past 90 days and doesn't include a "sold" price. Sold data is only available to seller accounts on a webpapp known as Terapeak. The sold data goes back 2 years.

A majority of the data was scraped from Terapeak, as this data goes back 2 years as compared to the API listed data that only goes back 90 days. It is assumed a large enough amount of listed data should approximate sold data well enough to prove useful for this project.

After cleaning the data to remove lots, groups, and most of the listings with multiple knives using regex, there were over 70,000 listings scraped from Terapeak, representing the final sale price of 70K different used knives and their associated titles and images. Around 10K listings were obtained using eBay's API, representing current listed used knives and their asking price. A large enough sample size of knives currently listed for sale should approximate the true value of the knife. A lot more of the sold data was available and is a clear indicator of the value of knives over the past two years.

After combining the two datasets mentioned above, the data was then filtered to remove outliers using IQR filtering. The final dataset includes roughly 76K listings for knives between the 8 brands of interest.

feature creation

- converted price: Combined sold price with shipping in USD.
- profit: Combined sold price with shipping in USD minus the cost of the knife at the Surplus Store, Ebay's commission, flat shipping cost for a 0.5 lb item and cost of envelope, and overhead costs.
- Return On Investment(ROI): calculated by dividing the profit earned on an investment by the cost of that investment, expressed in %USD.

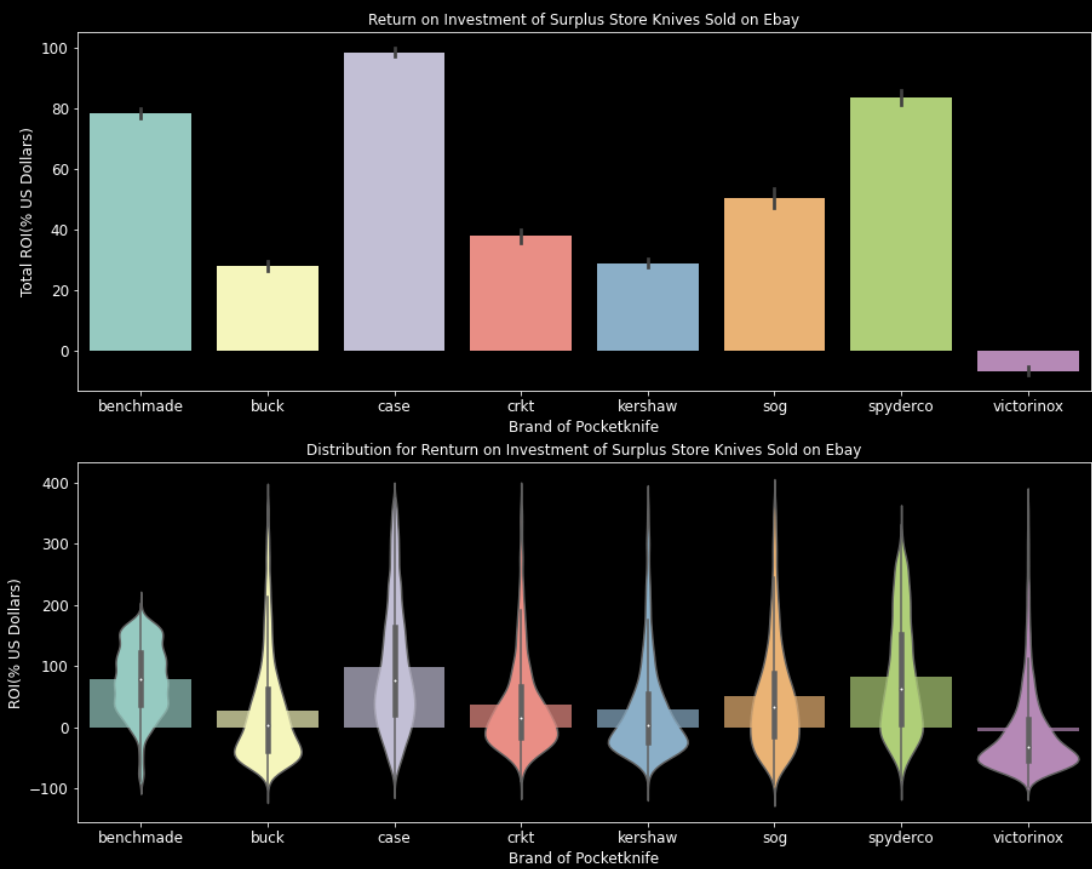
Methods

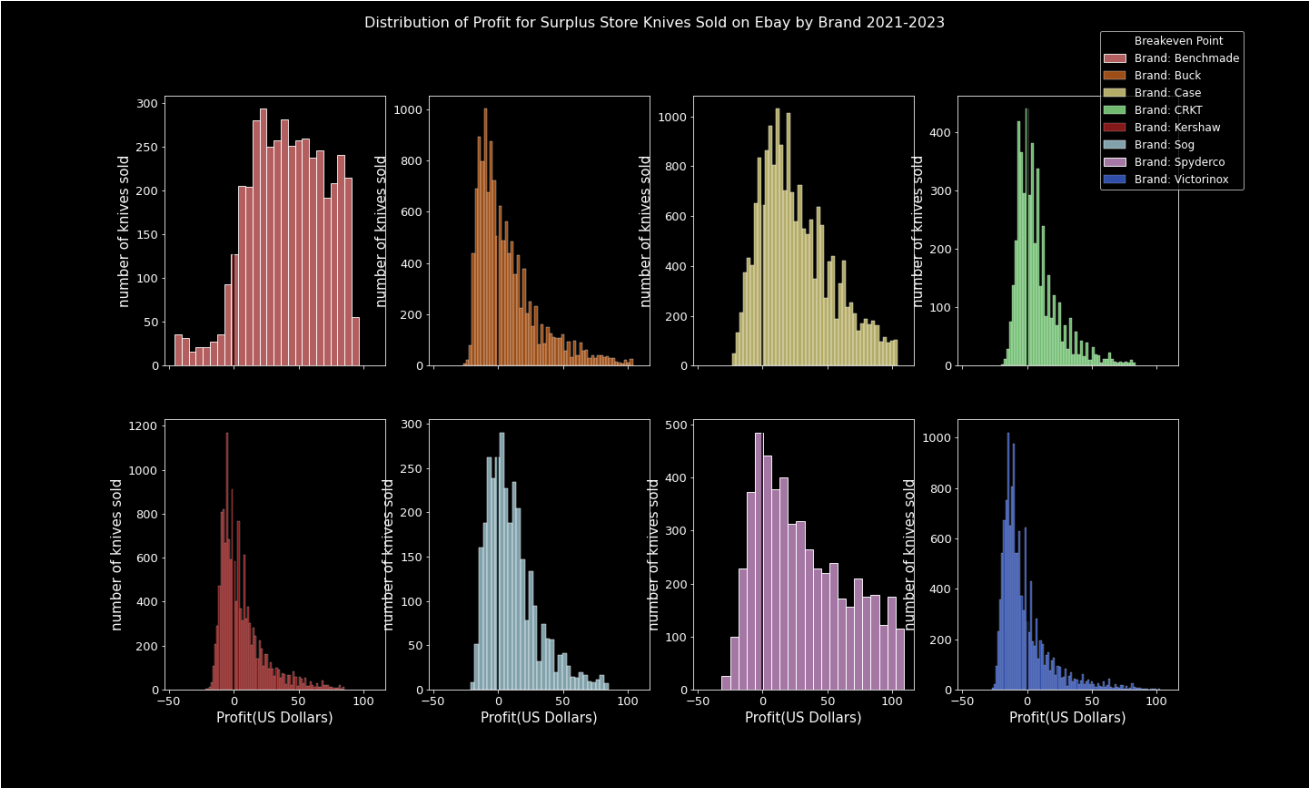
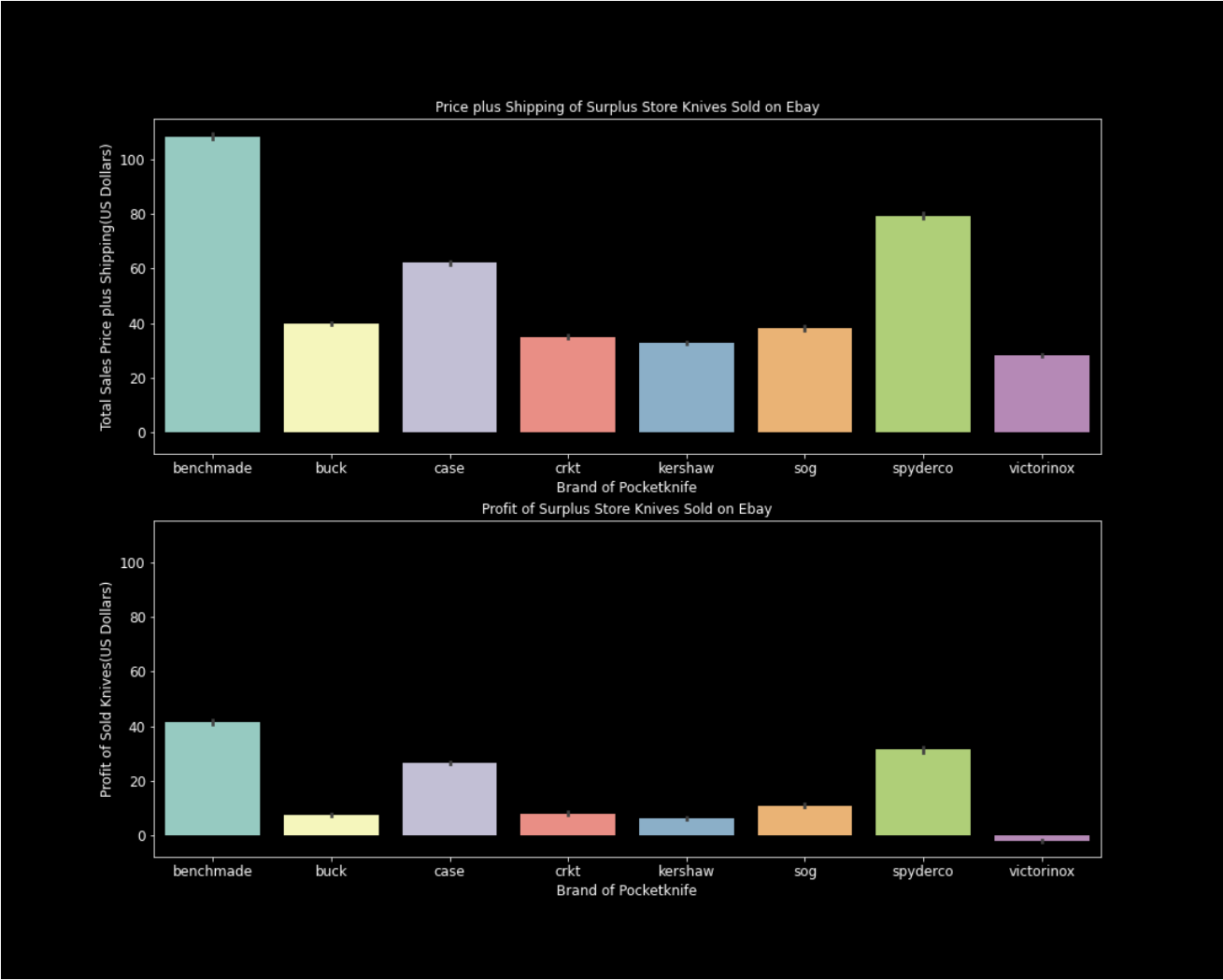
This project utilizes descriptive analysis to compare the profitability and return on investment for the 8 different brands available for sale at the Texas State Surplus store. By examining different measures of centrality for previous prices sold on eBay in the past 2 years, the measure of variability, percentiles and also the construction of tables & graphs, it can be inferred which brands of knives to invest more heavily into and which ones to ignore completely at just a glance.

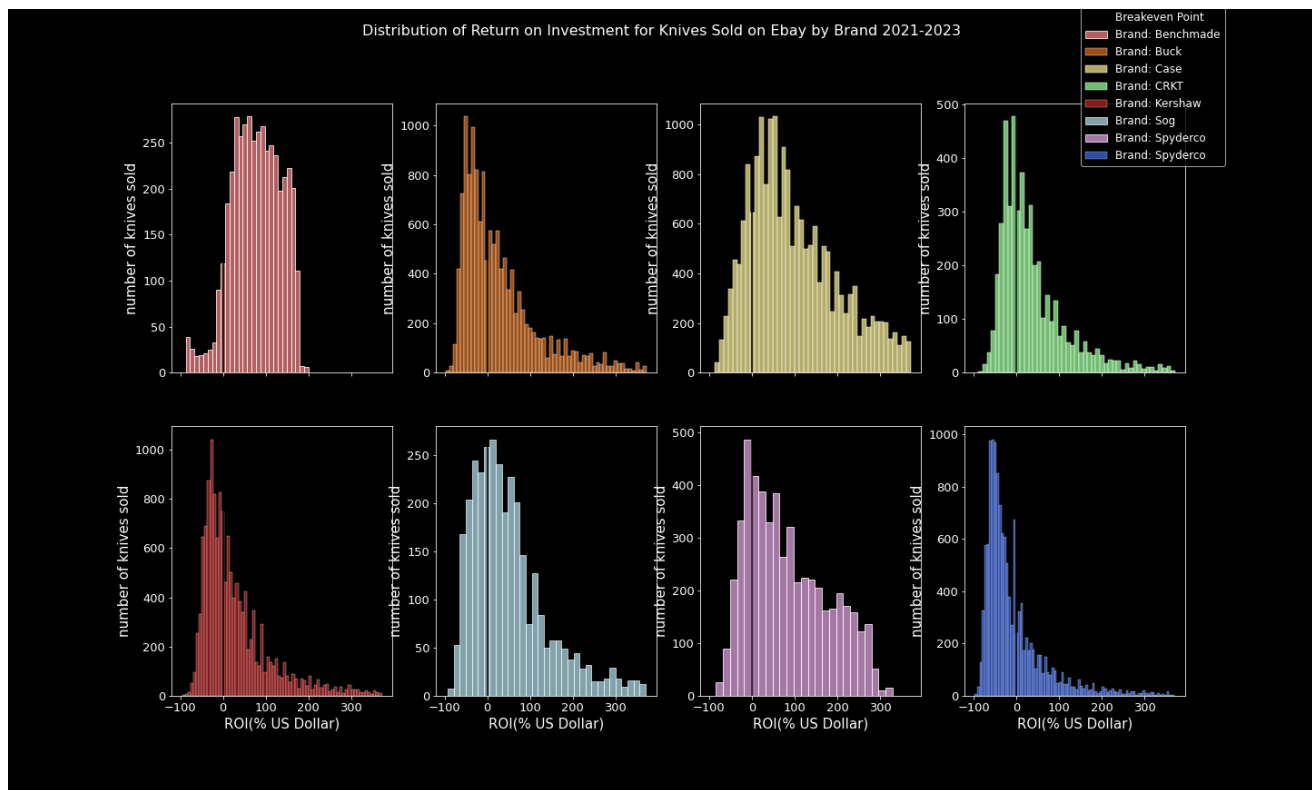
On top of descriptive analysis, the data will be used to train two different Neural Network models to predict the resale value of knives. The target feature for the model to predict is the total price (shipping included) that a knife should be listed on eBay. One model will use an RNN on titles in order to find potential listings that are undervalued and could be worth investing in. Another model will accept only images as input, as this is an input that can easily be obtained in person at the store. This model will use past sold data of knives on eBay in order to determine within an acceptable amount of error the price it will resell for on eBay (shipping included) using only an image.

Results

Descriptive Analysis

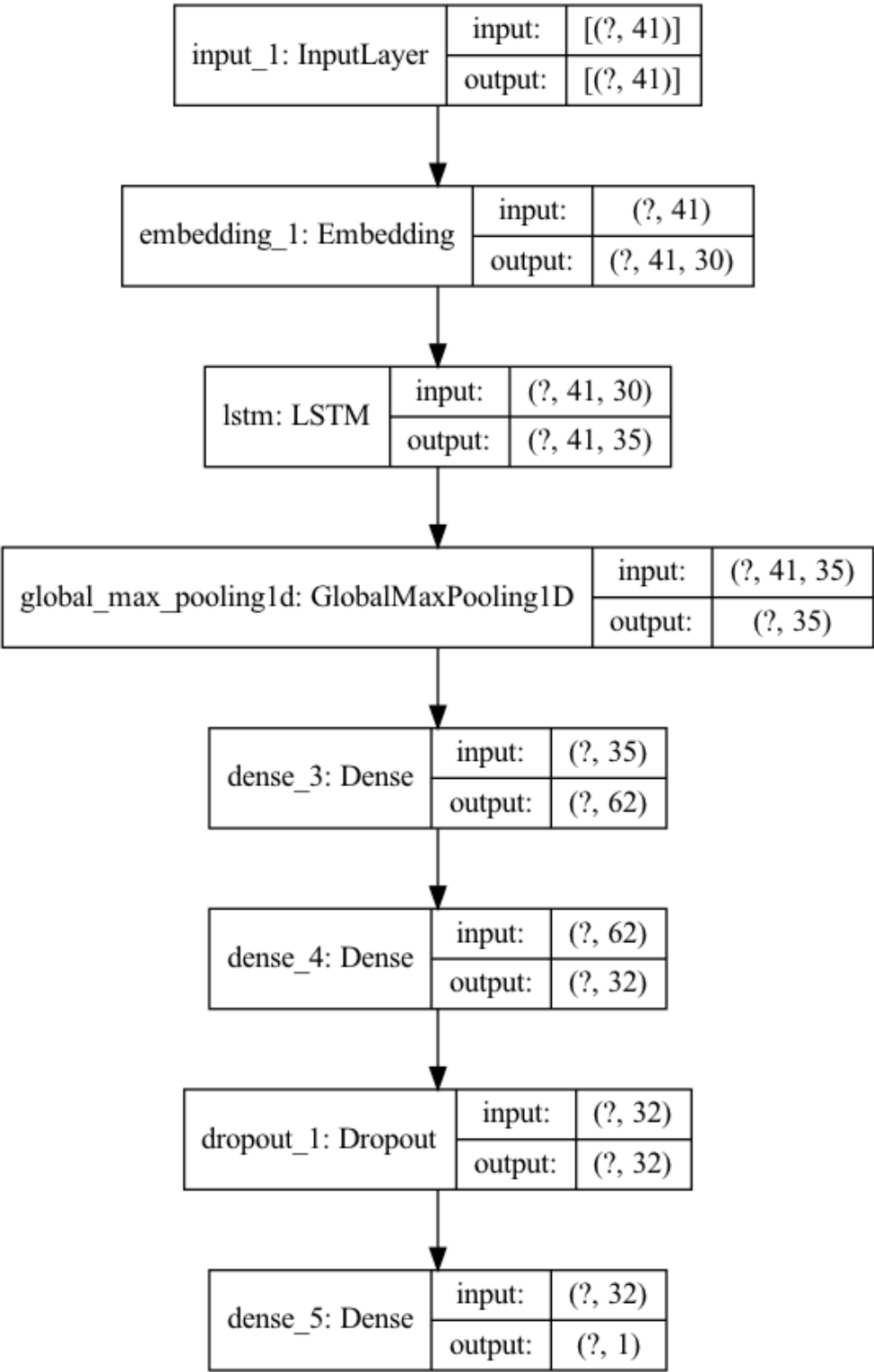


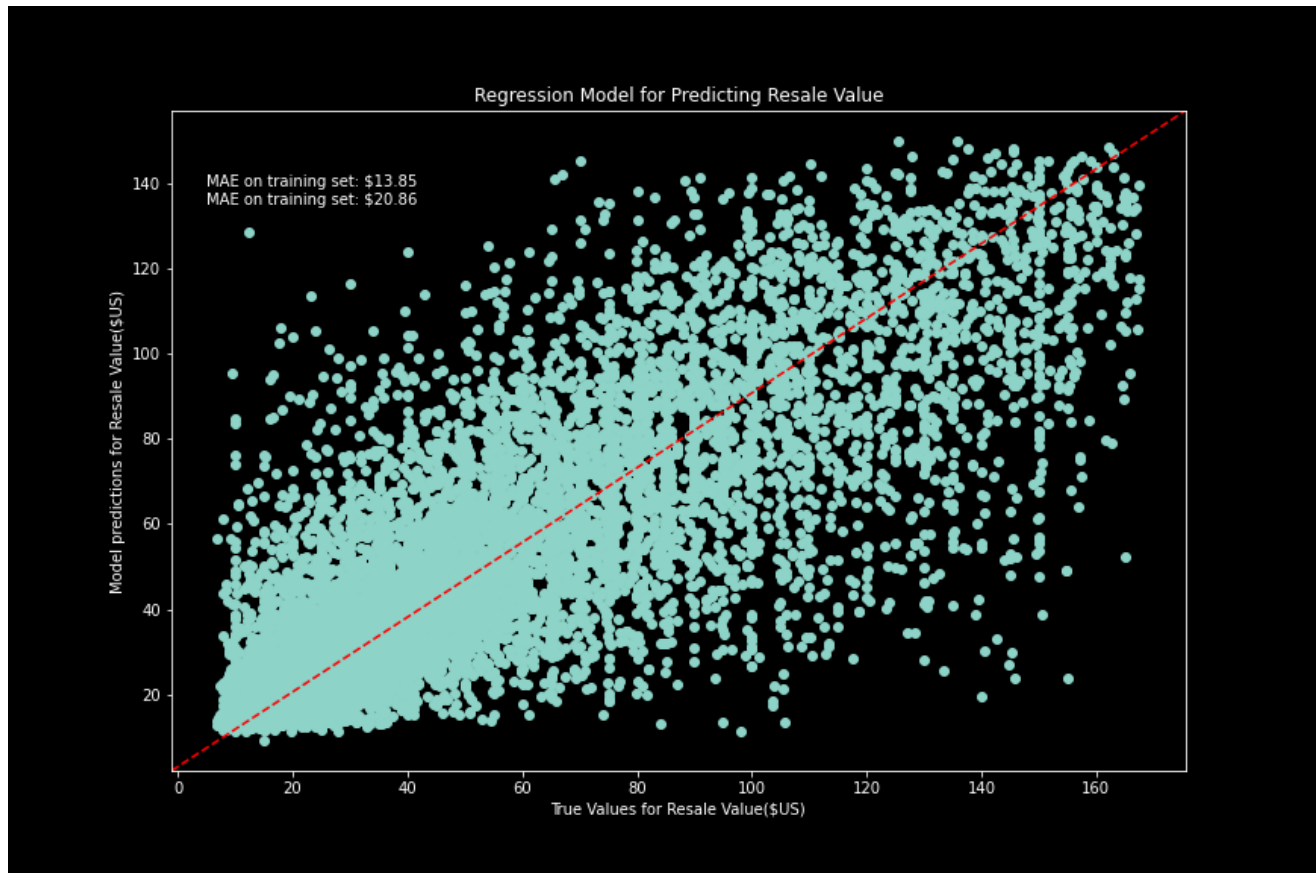
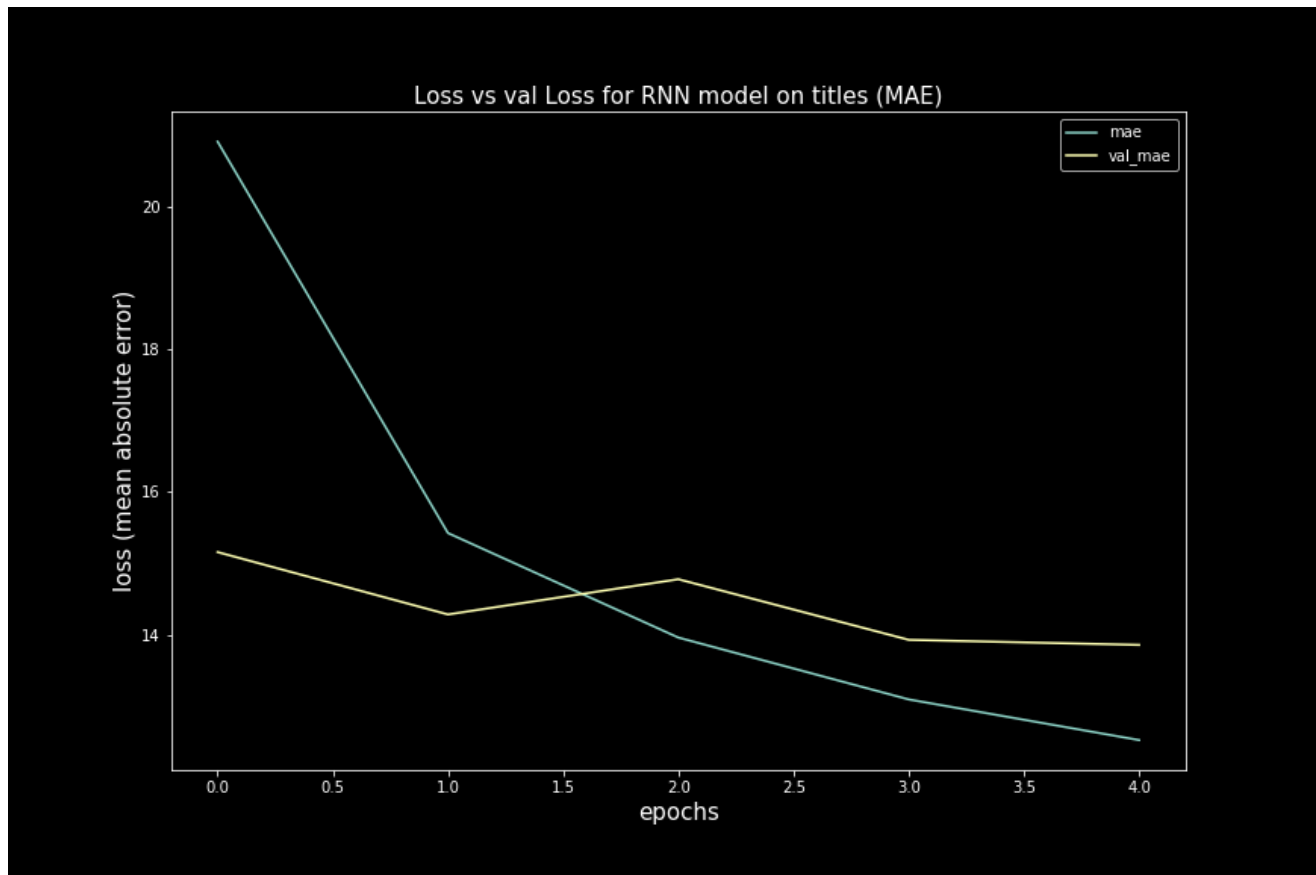




- Do Not Invest In Kershaw, Buck, CRKT, or especially Victorinox knives (only brand with negative centrality for profit).
- Case Knives appear to be the lowest risk knife to invest in with the highest return on investment.
- Benchmade knives appear to generate more profit but at higher risk with a higher cost to invest.
- Spyderco knives appear to be a great balance between initial cost, profit, and return on Investment.
- Spyderco, Benchmade, and Case knives are all great choices for investing in for an eBay storefront.

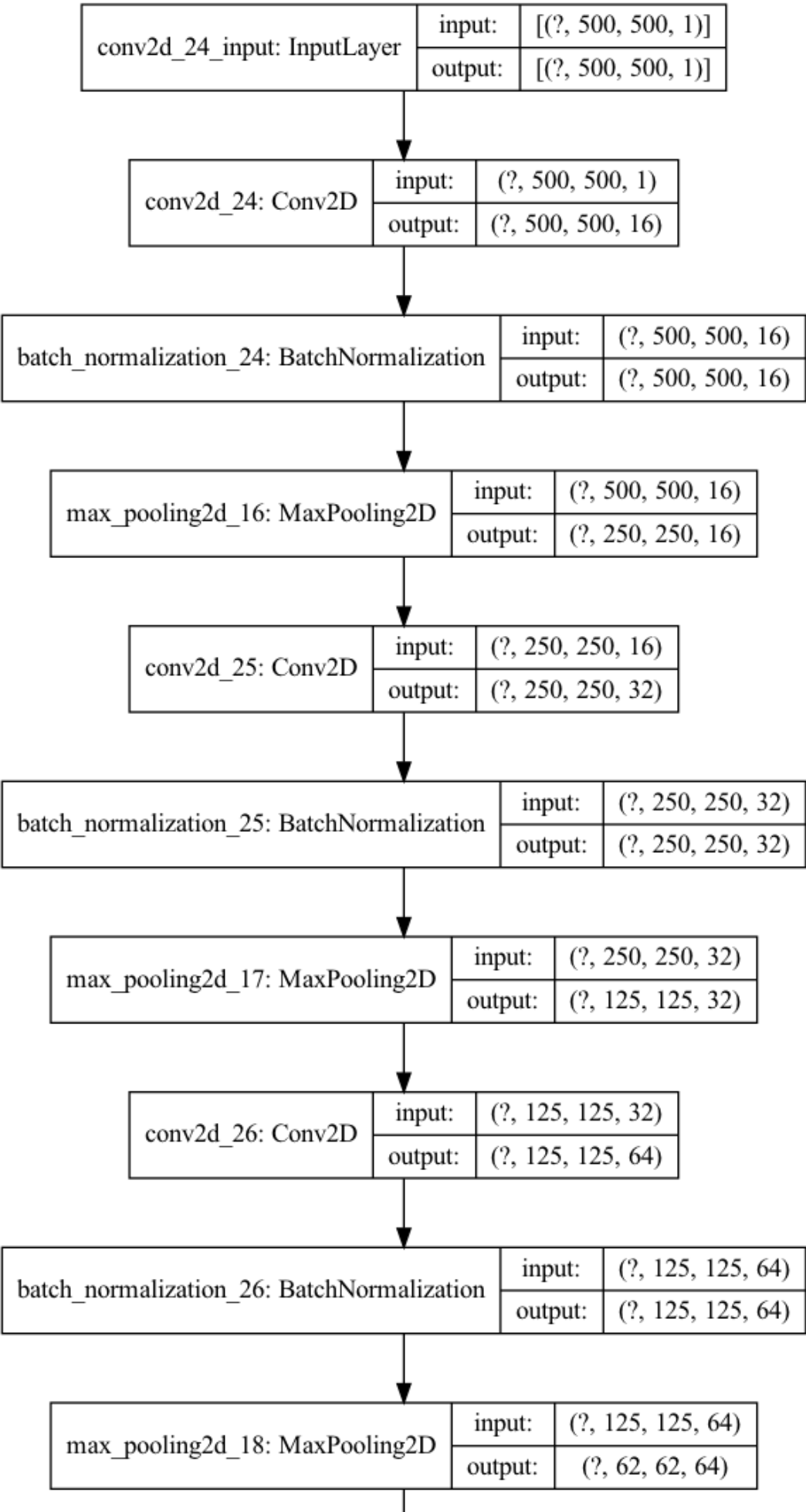
Recurrent Neural Network (Long Short Term Memory)

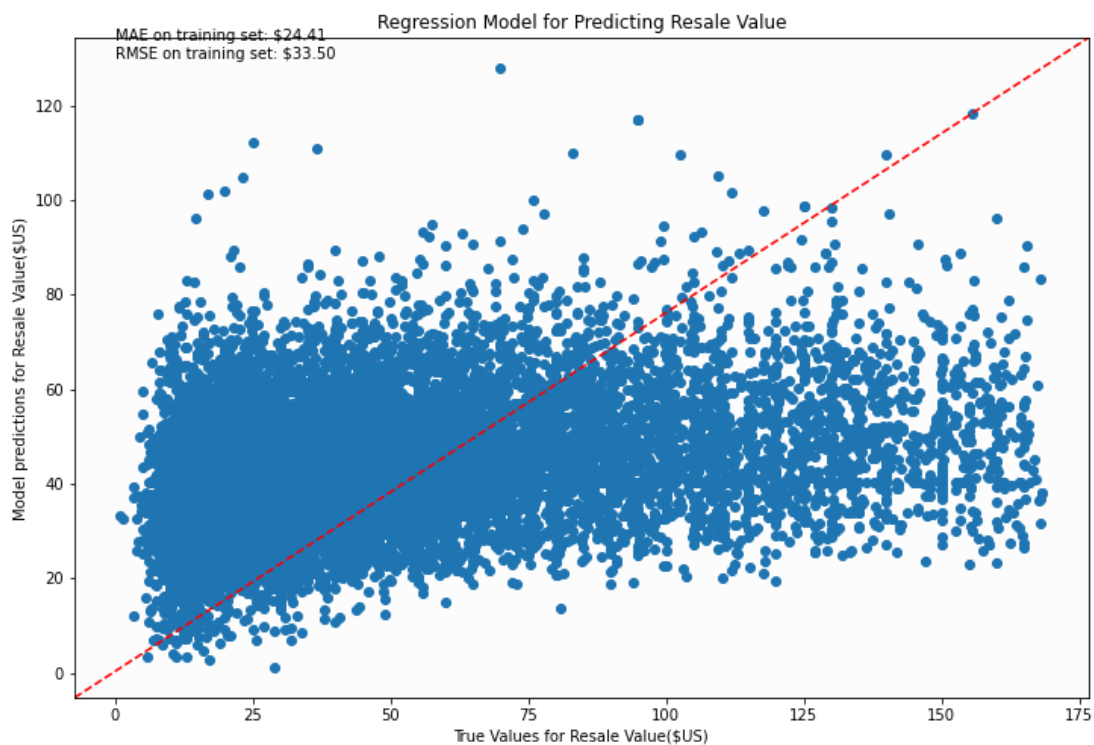
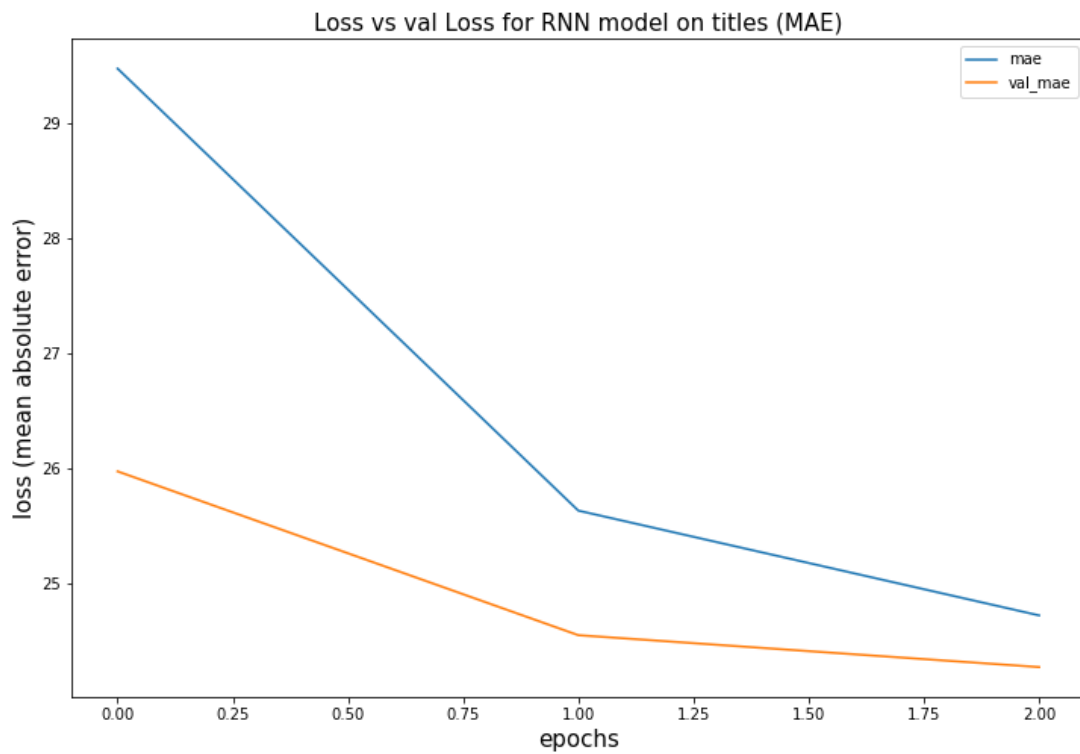




- The mean price of the 8 brands of knives sold on ebay is around \$50.00. A mean absolute error of about plus or minus \$13.80 is acceptable.

Convolutud Neural Network on Grayscale Images





- The RMSE for the CNN was roughly \$33.00, and the MAE was roughly \$25.00. That is an error of plus or minus about 50% of the mean price of knives sold. Not acceptable yet as compared to the RNN with titles. Will address in future work.

Future Work

- Expand data to include other products readily purchasable at the Surplus Store.
- Explode nested list of images as form of data augmentation to increase image input size by about 5-fold. Attempt other data augmentation techniques.
- Obtain more aspect related data for knives to understand most important cost altering features more thoroughly. Enough of this data could use a predictive model that isn't a black box.

THANK YOU

[Blog](#) | [GitHub](#) | [PreProcess Github](#)

Dylan Dey

email: ddey2985@gmail.com

— README.md	<— The top-level README for reviewers of this project
— Data_Sourcing_Notebook.ipynb	<— Notebook for obtaining and processing Ebay data
— Data_Exploration.ipynb	<— Descriptive Analysis of data
— Model_Interpret.ipynb	<— modeling notebook
— Project_Presentation.pdf	<— project presentation pdf
— listed_data	<— Both sourced externally and generated from code
— sold_data	<— Both sourced externally and generated from code
— images	<— Graphs and visualization/various helpful images

Releases

No releases published

[Create a new release](#)

Packages

No packages published

[Publish your first package](#)

Languages

● Jupyter Notebook 100.0%