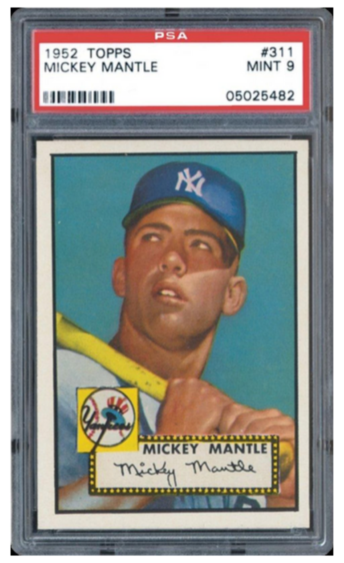
ALTERNATIVE ASSETS

Predicting auction prices for sports memorabilia

INTRODUCTION

Alternative assets (AA) are defined as less traditional and more unexpected forms of investing. They can range anywhere from rare paintings by famous artists to the valued sports memorabilia of legendary athletes. The realm of alternative assets is notorious for its high barriers to entry, since the purchase of a single item often requires a considerable amount of capital. Although more risky, the purchase of an alternative asset can yield monumental returns.



The popularity of sports memorabilia, specifically baseball, basketball, and football cards, has been increasing over the past few years. As this market grows, it has become increasingly profitable to trade these assets. In addition, a number of companies are have been making alternative assets more accessible to small individual investors.

GOAL

Due to the profitability of trading sports memorabilia and increasing access to this form of investing, our goal was to create a predictive model for buying and selling sports memorabilia through auctions and auction sites. We decided to focus on sports cards because that was the subset of sports memorabilia that had the least variability between individual assets.

Purpose of Our Model: Predict the current price of alternative assets by looking at previous selling points in addition to a number of other relevant factors, such as the perceived significance of the athlete on the card.

Use Case of Our Model: Since valuable sports cards are traded infrequently and their prices are determined by auction houses at the time of the auction, it is hard for buyers and sellers to determine the "real" value of an alternative asset. With our model being able to estimate the current price of a sports card, we have the knowledge to **evaluate whether or not the sports memorabilia is overvalued or undervalued and buy or sell based on when we deem most fit.**

DATA

Our data came from the **Professional Sports Authenticator (PSA)** website and we retrieved information on the baseball, football, and basketball cards sold on their site. By combining that with statistics of the athlete (which we got from **sports-reference.com**), we had a dataset that contained the history of prices at which a card was sold and the significance of the player. We had to remove items that were only sold once since we were not able to create a proper history of prices with one sale. In addition, cards that represented more than one player also had to be removed since we did not have an accurate way of combining the players' statistics into one overall metric.

METHODOLOGY

Multiple Regression

We used multiple regression for our prediction model, since our ultimate goal was to predict the price of a card if it were to be sold today given a multitude of variables that are likely to influence price. After a considerable amount of research, we chose to use the 15 independent variables are listed below as a starting point:

Variable Explanations

0: Year the card was created
1: Current grade of the card (quality of the card from 0.0 - 10.0)
2: Is Hall of Fame (0 - was not Hall of Fame, 1 - was Hall of Fame)
3: Average Price of the card
4: Percent Price Change per Year of the card
5: Highest Price at which it was sold
6: Lowest Price at which it was sold
7: Count of number of times sold
8: The number of awards the player on the card won
9: The Oldest Price of the card
10: Oldest Trade Date of the card
11: Most Recent Trade Date of the card
12: Important Player Stat #1 (dependent on the sport)
13: Important Player Stat #2 (dependent on the sport)
14: Important Player Stat #3 (dependent on the sport)

Explanation for Variables 12-14: Since the three sports that we looked at for predicting card prices have wildly different key statistics, we decided to collect different statistics for each sport
Baseball: Wins over Replacement (WAR), Homeruns (HR), and Runs Batted In (RBI)
Basketball: Total Games Played, Points per Game (PPG), and Assists per Game (APG)
Football: Omitted due to huge difference in statistics based on a player's position

We ran the regression separately for each sport due to the differences in the statistics collected for each sport. We determined training and testing data using an 80/20 train/test split.

RESULTS AND ANALYSIS

Figure 1: Correlation Matrix of All Sports Card Attributes Collected

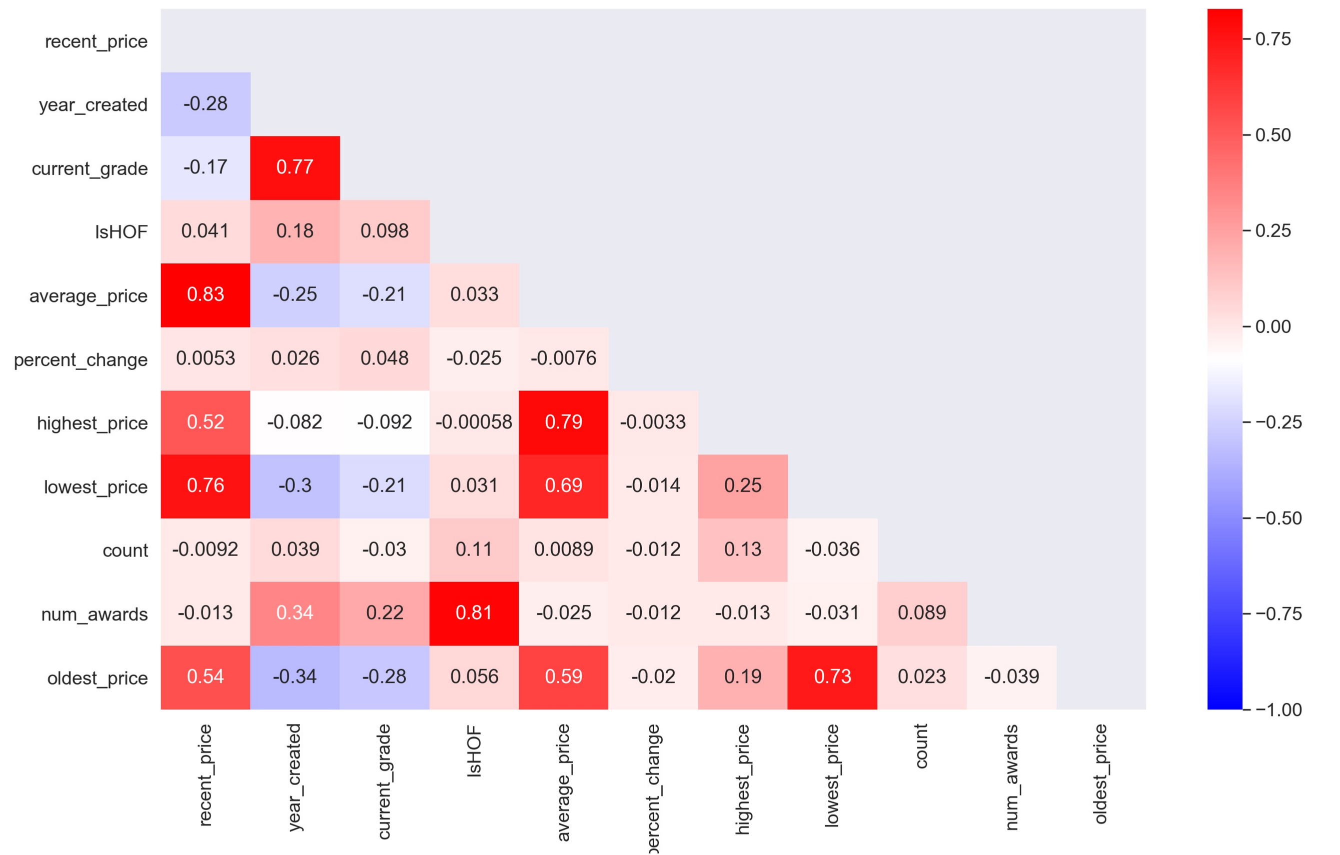


Figure 1 Analysis: The most correlated variables with respect to recent price and their respective correlation values are: **average_price** (0.83), **lowest_price** (0.76), **highest_price** (0.52), and **oldest_price** (0.54). The least correlated are **year_created** (-0.28) and **current_grade** (-0.17).

Figure 2: Multivariable Regression Results Summary

Variables	R^2	Test MSE	Train MSE
Top 9 Efficient	0.5867034709210321	2,930,908.429045171	2,441,127.635897760
Top 4 Correlated	0.5724917544208122	3,255,110.709511859	2,328,575.38211199

Figure 2 Analysis: When evaluating the results of the correlation matrix, we wanted to look at the combination that produced results that most efficiently increased our model's R^2 value and most efficiently decreased training and testing MSE as well as the discrepancy between the two.

By doing so, we discovered that **the 9 variables highlighted to the left were the ones that best optimized our model.**

Figure 3: Cumulative Importance's of Baseball Memorabilia Features

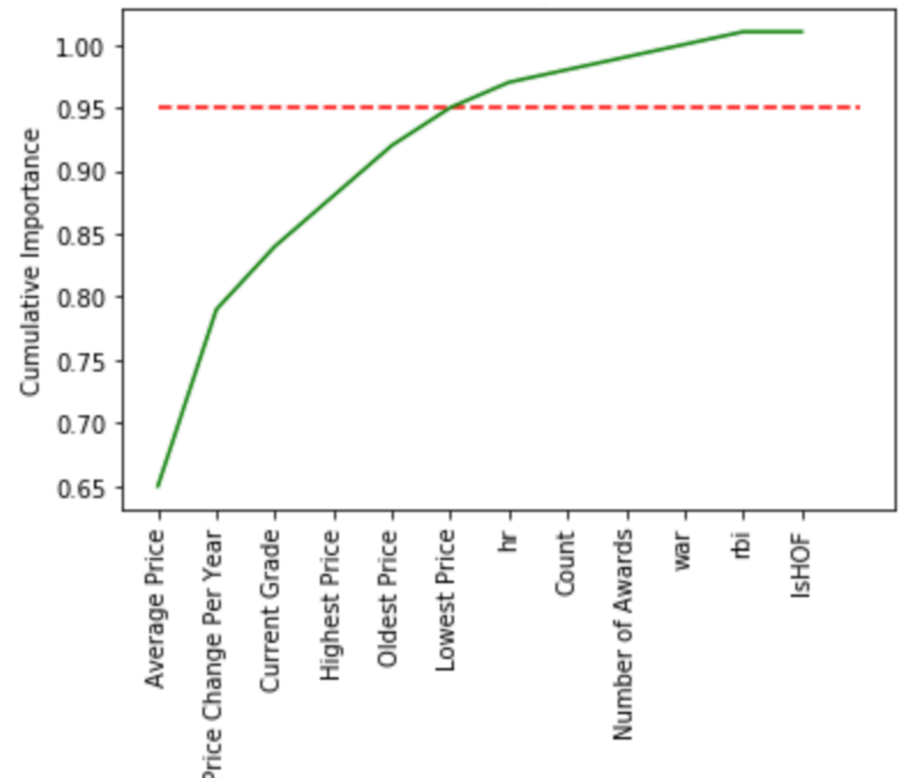


Figure 4: Normalized Errors for Baseball Testing Data
(Lines represent (y_test − y_prediction / (y_test))

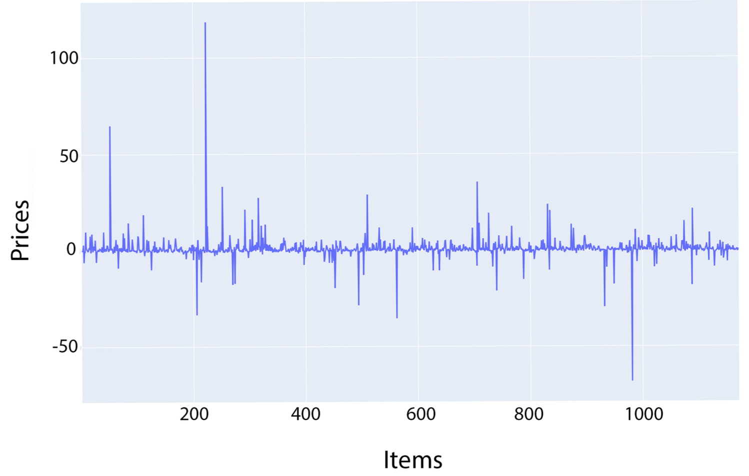


Figure 4 Analysis: Most predictions have a low normalized error (between -$5 and $5). However, there are a few outliers where the prediction is considerably off.

SIGNIFICANCE / LIMITATIONS

Significance: Our model is significant to the field of alternative assets because predicting the value of sports memorabilia using machine learning has the potential to be a very effective method of investing in AAs. By bringing some sense of predictability into a inherently unpredictable field, models like this can encourage people to explore new methods of investment in a safer and more profitable way.

Limitations: We were unable to take into account current events in our model, something that would make our model's prediction more accurate and up to date. (After Kobe Bryant's death, the value of his sports cards increased significantly.) We also wanted to explore how to predict the prices of new assets for players, but were not able to get accurate results. Lastly, we wanted to enhance our model by better incorporating the popularity of each player using Google search data, but could not find a suitable API.

A csci1951a Final Project by Ben Hurd, Rohit Jawle, Dylan Garcia, and Kei Nawa