Predicting the Prices of Magic the Gathering Cards

An analysis of Commander format legal cards from MTG

Goal: To perform data analysis on cards in the Commander format and build a machine learning model to predict the price of cards

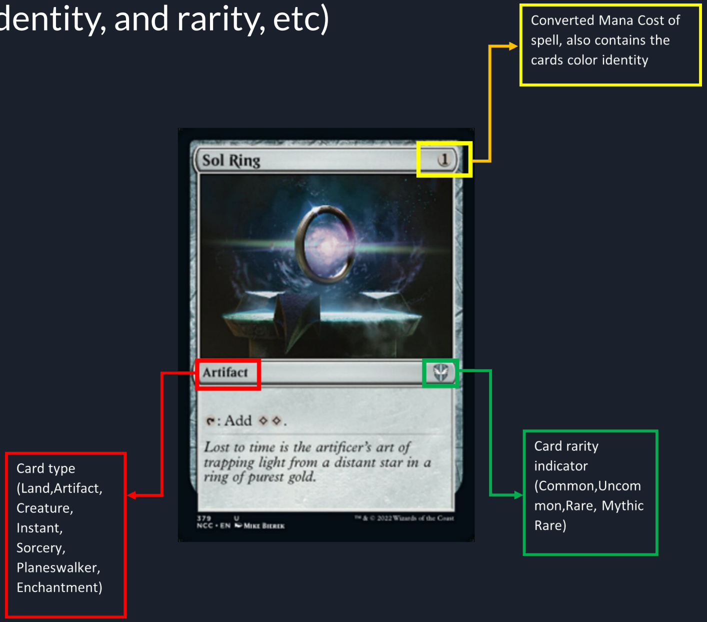
1. Build a webscraping tool that will scan scryfall.com for pertinent card information (i.e. CMC, color identity, rarity, etc.)
2. Organize data into a pandas dataframe using Python and perform data cleaning and exploratory data analysis
3. Use insights from data analysis to test multiple machine learning models and compare their performance in predicting the price of cards based on key features

Method

1. Build a webscraping tool that will scan scryfall.com for pertinent card information (i.e. CMC, color identity, rarity, etc.)
2. Organize data into a pandas dataframe using Python and perform data cleaning and exploratory data analysis
3. Use insights from data analysis to test multiple machine learning models and compare their performance in predicting the price of cards based on key features

Magic the Gathering Basics

- Magic the Gathering is a trading card game the was developed by Wizards of the Coast in the early 1990's
- The core mechanic of the game revolves around using a resource (mana) in order to cast spells
- All cards have a set of key characteristics that define them (i.e. - Converted Mana Cost, Card type, Color identity, and rarity, etc)


Sol Ring
1
Artifact
: Add ◇◇.
Lost to time is the artificer's art of trapping light from a distant star in a ring of purest gold.
379 U
NCC • EN Mike Bierek
™ & © 2022 Wizards of the Coast

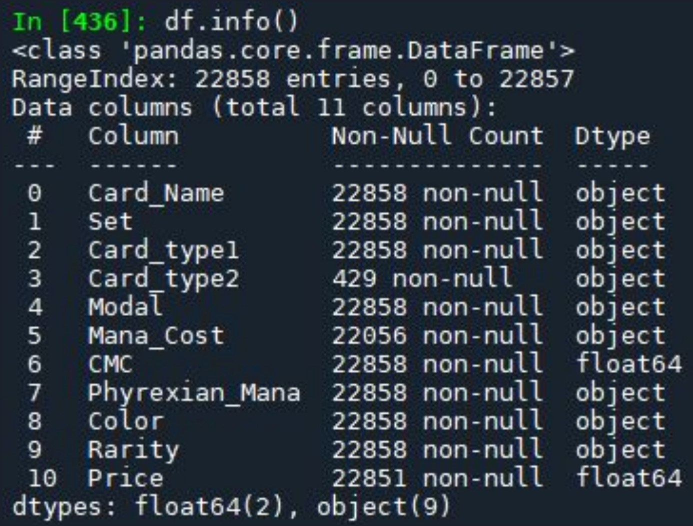
Converted Mana Cost of spell, also contains the cards color identity

Card type (Land,Artifact, Creature, Instant, Sorcery, Planeswalker, Enchantment)

Card rarity indicator (Common,Uncommon,Rare, Mythic Rare)

Gathering the data

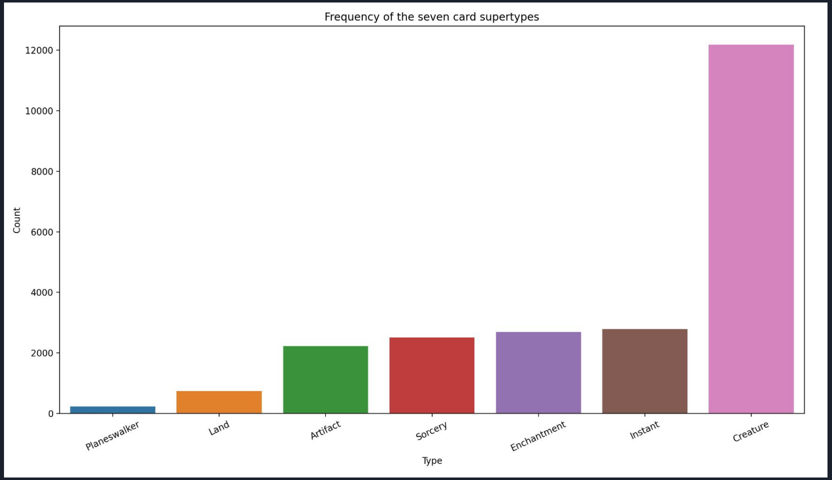
- In order to gather the data a webscraping tool using Python with the BeautifulSoup library was implemented to parse and read the HTML base of webpages containing the pertinent cards and their characteristics
- The resulting information was stored in a dataframe and exported to a csv file which was then used for data cleaning and analysis



```
In [436]: df.info()
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 22858 entries, 0 to 22857
Data columns (total 11 columns):
 #   Column          Non-Null Count  Dtype  
---  ------          --------------  -----  
 0   Card_Name       22858 non-null  object 
 1   Set             22858 non-null  object 
 2   Card_type1      22858 non-null  object 
 3   Card_type2      429 non-null    object 
 4   Modal           22858 non-null  object 
 5   Mana_Cost       22056 non-null  object 
 6   CMC             22858 non-null  float64
 7   Phyrexian_Mana  22858 non-null  object 
 8   Color           22858 non-null  object 
 9   Rarity          22858 non-null  object 
 10  Price           22851 non-null  float64
dtypes: float64(2), object(9)
```

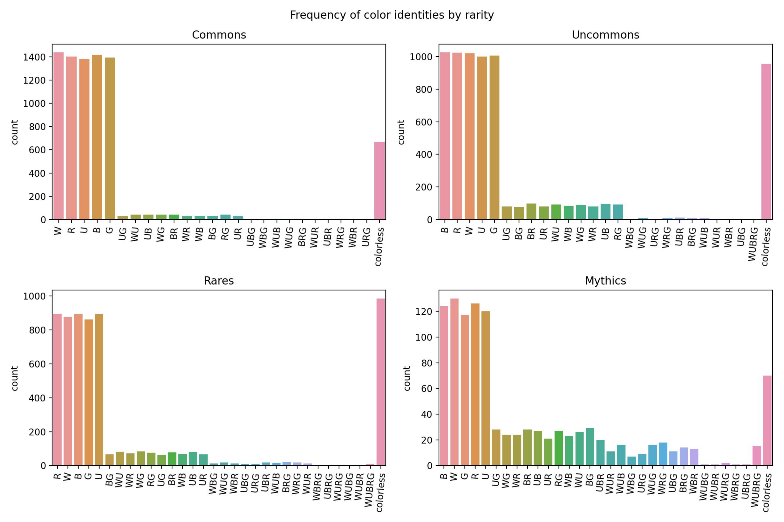
Data Analysis

- To get a broad idea of the data the first feature explored was the frequency of each card type among the pool of Commander legal cards
- Seen below it is clear that the card type creature is far more common than any of the other types possible, while artifacts, enchantments, instants and sorceries all have fairly even appearance across all the cards



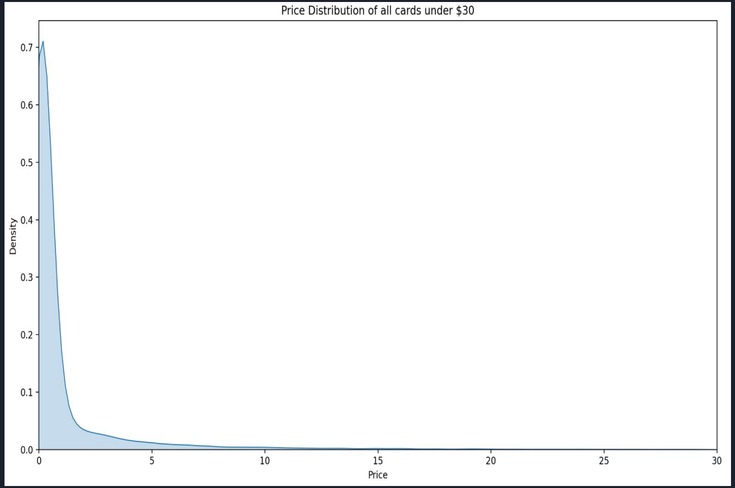
Color identity distributions

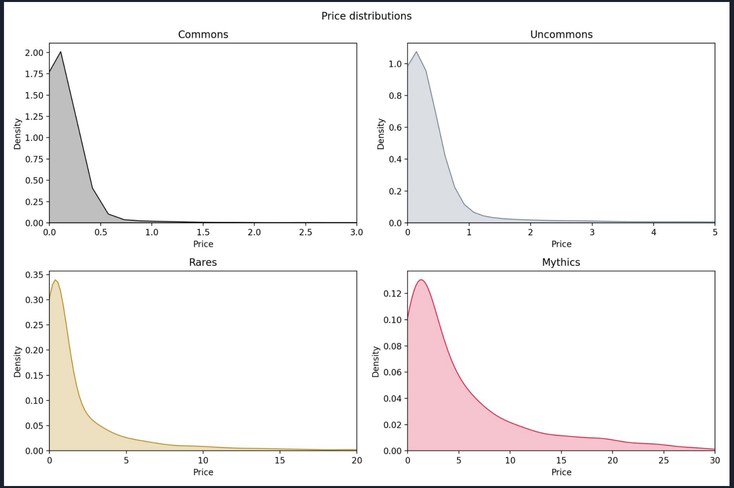
- Pictured below is a distribution of the 32 possible color identity variants for a card in MTG separated by rarity
- While mono-colored cards occupy and overwhelming majority of the cards, 2 and 3 color cards become more prevalent as the card rarity increases, with 2 color cards being the most common within the mythic rare category



Price distributions

- Pictured below are the price distributions of cards (for all cards and separated by rarity)
- The vast majority of cards fall below the $1 price point
- Rare and Mythic Rares shift their distributions to the right slightly, with Mythic Rares having a nontrivial amount of cards in the $10-25 range
- The huge dominance of cards less than $1 does hold weight in how it will affect price prediction performance as the data set is heavily over saturated with cards at those prices





Price Predictions: Machine Learning

- In order to predict the price of a given MTG card, various machine learning algorithms were implemented and compared to each other
- Multiple models were employed using both sci-kit learn and TensorFlow
- All models used implement a standard train test split methodology and a grid searches where possible to find optimum hyperparameters
- Model 1: Linear regression - The first model used was a linear regression model, specifically ridge regression after determining regularization from an elastic net model
- Model 2: Random Forest - A random forest regressor was implemented as the second model

Summary and results

- All models tested were able to make their price predictions within an error of $3 or less
- Overall the neural network model performed the best with an MAE just over $1 and a RMSE of $2.78

Model	MAE	RMSE
Linear Regression	1.43	2.84
Support Vector Regression	0.93	2.92
Random Forest Regression	1.20	2.78
Neural Network Regression	1.10	2.78

Considerations and thoughts

Future changes that could be made to improve model performance:

- Train the models to more aggressively adjust for the large abundance of cards $1 and less in order to more accurately predict more expensive cards
- Include more features to train on (i.e. - length of card rules text, whether certain keywords or abilities are present, number of printings the cards has, what type of set the card was printed in, etc.)

Final thoughts: This analysis was an excellent chance to utilize machine learning and data analysis skills to take a deeper look at one of the most popular trading card games in the world. The analysis not only gave insight into how card prices may depend on certain aspects of the cards, but also gave me a deeper understanding of the foundational makeup of the game.