



Using linear regression to predict **Magic: the Gathering** card prices



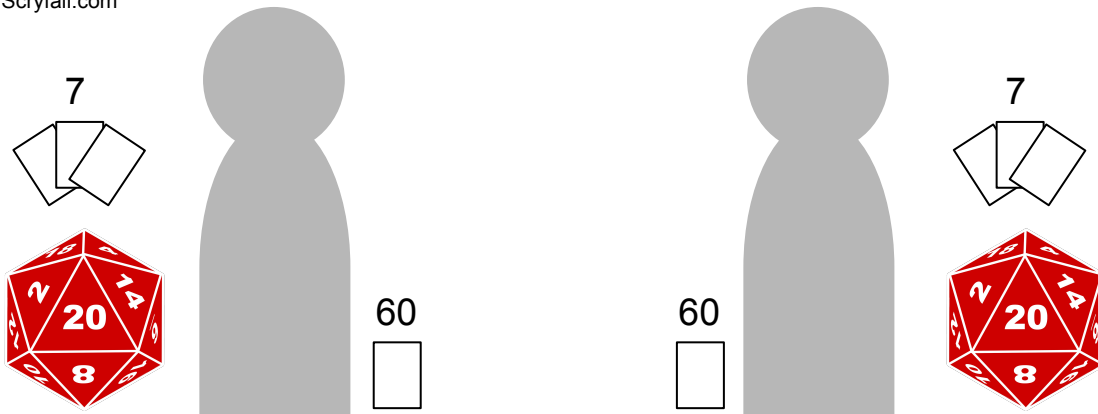
Noah Jaffe
Project 2
Ridge cohort



What is Magic: the Gathering?

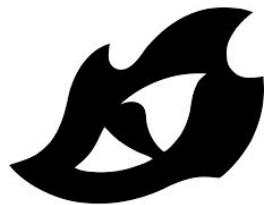


All card images: Scryfall.com



Card supply and playability may affect price

- Quarterly expansions of approx 300 new cards
 - Cards may be reprinted, affecting supply
- Multiple MTG formats:
 - Focus is 'Modern' format (2004 - present)



Which features might influence price?



Which features might influence price?



Rarity

Which features might influence price?



Rarity

Number of printings

Which features might influence price?



Rarity

Number of printings

Card color(s)

Which features might influence price?



Rarity

Number of printings

Card color(s)

Converted mana cost

Data scraped from CardKingdom.com

Beautifulsoup, Selenium

The screenshot shows the CardKingdom.com website interface. At the top is a navigation bar with the site logo, search bar, and links for 'Advanced Search', 'Sign In', and 'Register'. Below the navigation bar, the page title is 'MTG Tourn Format / Modern Format'. On the left is a sidebar with filters for 'Card Name', 'Edition', 'Card Color', 'Color Identity', 'Rarity', 'Type', 'Power/Tough', 'Converted Cost', 'Price', 'Oracle Text', and 'Mana Production'. The main content area displays search results for 'Gideon of the Trials'. A red box highlights the card's details, including its image, name, expansion, rarity, cost, CMC, color identity, and rules text. The card is priced at \$6.99. Below it, 'Force of Negation' is also visible, priced at \$54.99.



Pandas

	name	expansion	rarity	cost	cmc	color_identity	rules_text	price
0	Gideon of the Trials	Amonkhet	M	1ww	3	w	[+1]: Until your next turn, prevent all damage...	6.99
1	Force of Negation	Modern Horizons	R	1uu	3	u	If it's not your turn, you may exile a blue ca...	54.99
2	Polluted Delta	Khans of Tarkir	R		0	colorless	, Pay 1 life, Sacrifice Polluted Delta: Search...	32.99
3	Cavern of Souls	Avacyn Restored	R		0	colorless	As Cavern of Souls enters the battlefield, cho...	79.99
4	Wooded Foothills	Khans of Tarkir	R		0	colorless	, Pay 1 life, Sacrifice Wooded Foothills: Sear...	29.99

Target

Feature engineering

- Rarity → 1-4
- Presence/absence binary for card color
- 'Time since release' column for card age
- Rules text length for complexity

is_w	is_u	is_b	is_r	is_g	rarity_num	rules_text_len	num_printings	rel_date	time_since_release
1	0	0	0	0	4	401	1	2017-04-28	1259.0

Feature engineering

- Rarity → 1-4
- Presence/absence binary for card color
- 'Time since release' column for card age
- Rules text length for complexity

is_w	is_u	is_b	is_r	is_g	rarity_num	rules_text_len	num_printings	rel_date	time_since_release
1	0	0	0	0	4	401	1	2017-04-28	1259.0

Feature engineering

- Rarity → 1-4
- Presence/absence binary for card color
- 'Time since release' column for card age
- Rules text length for complexity

is_w	is_u	is_b	is_r	is_g	rarity_num	rules_text_len	num_printings	rel_date	time_since_release
1	0	0	0	0	4	401	1	2017-04-28	1259.0

Feature engineering

- Rarity → 1-4
- Presence/absence binary for card color
- 'Time since release' column for card age
- Rules text length for complexity

is_w	is_u	is_b	is_r	is_g	rarity_num	rules_text_len	num_printings	rel_date	time_since_release
1	0	0	0	0	4	401	1	2017-04-28	1259.0

Ten model features

- Rarity 0-4
- Number of printings
- CMC
- Rules text length
- Time since release
- Is_white
- Is_blue
- Is_black
- Is_red
- Is_green

Model testing workflow

- **EDA**, feature examination
- 5-fold **train-val-test** on 6 models
 - Selected **LassoCV**
- Examine residuals, r^2 , RMSE, MAE



Weak but significant feature effects

rarity_num	1.158
num_printings	0.313
time_since_release	0.292
cmc	-0.432
is_w is_r	0.309
is_r is_g	0.276

$p < 0.05$

Weak but significant feature effects

rarity_num	1.158
num_printings	0.313
time_since_release	0.292
cmc	-0.432
is_w is_r	0.309
is_r is_g	0.276

$p < 0.05$

Weak but significant feature effects

rarity_num	1.158
num_printings	0.313
time_since_release	0.292
cmc	-0.432
is_w is_r	0.309
is_r is_g	0.276

$p < 0.05$

Weak but significant feature effects

rarity_num	1.158
num_printings	0.313
time_since_release	0.292
cmc	-0.432
is_w is_r	0.309
is_r is_g	0.276

$p < 0.05$

Strongest significant feature effects

rarity_num	1.158
num_printings	0.313
time_since_release	0.292
cmc	-0.432
is_w is_r	0.309
is_r is_g	0.276

$p < 0.05$

Results

- r^2 of **0.296** on final test set
- RMSE of **\$3.99**
- MAE of **\$1.50**



Takeaways

- Low r^2 suggests model poorly explains price variance
- However, model predicts within a few dollars



Limitations and future directions

- Only Modern format data
- Other card features
 - Card type, specific abilities
- Covid has affected in-person Magic



Thank you!

