

**The Immensity that is**



---

Capstone Project - DSI 7

Sarah Berner  
May 16, 2019

# Overview

- Breakdown of Hearthstone
  - Cards, basic rules, why the data is INSANE and there is so much of it
- Problem Statements
  - What are we trying to understand and why?
  - Looking at this from the view of Hearthstone developers
    - Is there a good, balanced meta?
    - Is there consistency in types of cards?
- EDA
- Models, Results, Evaluations
- Conclusion and next steps



# Hearthstone



Pick one of 9 classes

Create a deck of 30 cards, some class specific

Defeat the opponent using your cards

- minions
- spells
- weapons

No ONE strongest deck or class (hopefully!)

Basically a very complicated rock-paper-scissors

# Card Type: Minion



Type	Minion
Name	Beryllium Nullifier
Class	Warrior
Expansion	The Boomsday Project
Rarity	Epic (purple)
Attack	3
Health	8
Race	Mech
Mechanic	Magnetic
Text	"Magnetic Can't be targeted by spells or Hero Powers."

# The Meta



## The “meta”:

- what are the most popular decks right now?
- what cards are seeing a lot of play?
- what are the strongest classes (due to their unique class cards)?
- how could we potentially use this information to build a strong deck? ←- this is a pipedream



# The Data

1. Using HearthstoneAPI, collected data on every individual card
  - a. Type, Class, Text, etc, - Expansion, img files, artist, etc.
  - b. Drop non-collectable cards and non-relevant columns: 2012 cards with 27 features per card
2. Using data provided by metastats.net, collected:
  - a. Top decks archetypes from April 25th to May 2nd
  - b. Top decks archetypes May 3rd to May 10th
  - c. Deck matchup numbers and win percentages
  - d. Top deck lists for the top 23 most popular deck archetypes

Data wrangling / cleaning / creating new dfs for days.

# The Questions

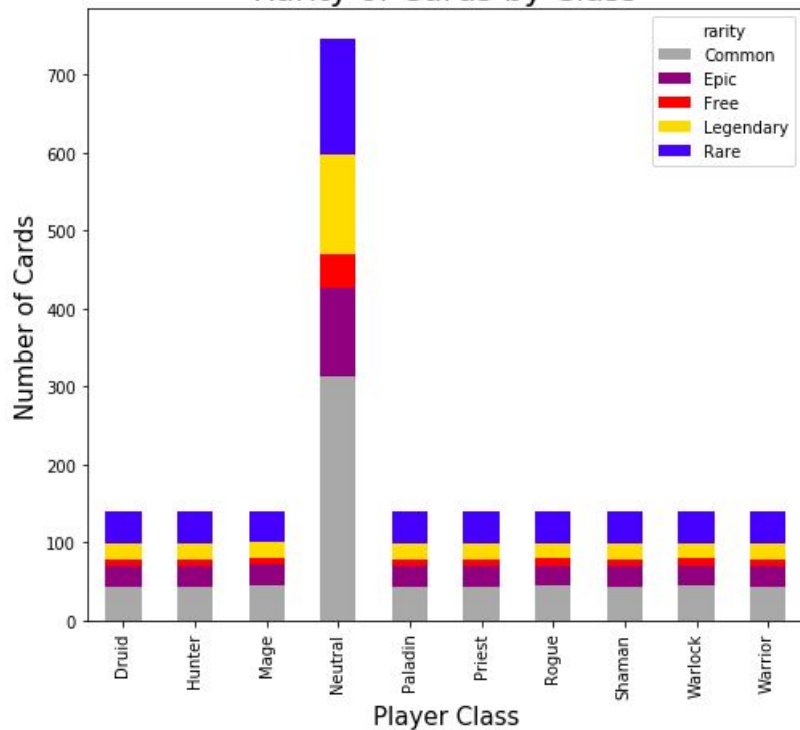
What does the meta look like? What can we learn about the structure of Hearthstone?

Are Legendary Cards PREDICTABLE, or are they inconsistently predicted due to their novelty?

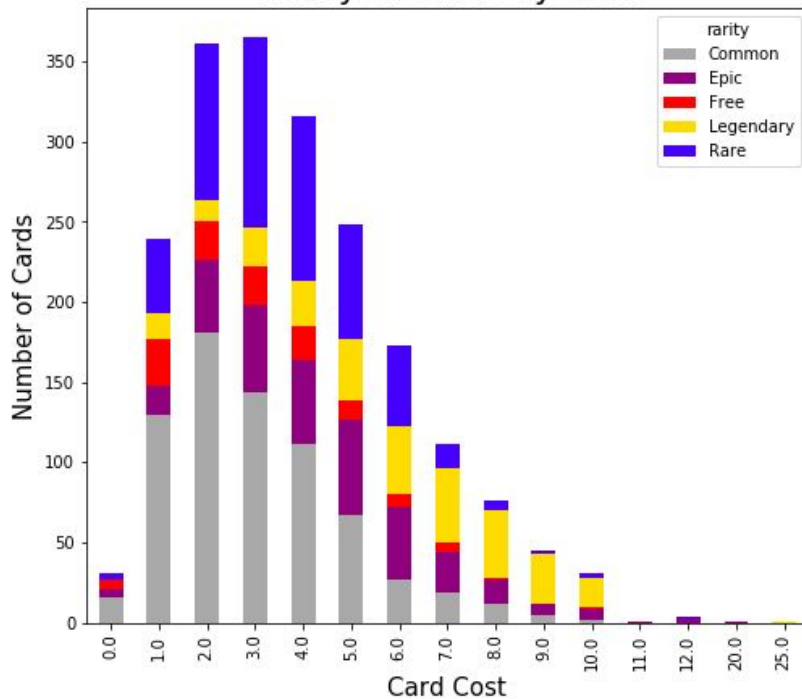


# EDA 1 - The cards

Rarity of Cards by Class



Rarity of Cards by Cost



# EDA 2 - Deck Archetype Matchups

Two data sets - top matchups (~25 deck archetypes) over two weeks Week of April 25th and May 3rd

## Disclaimers 1

- Each class has a subset of popular TYPES of decks focused around a theme - “archetypes”

Ex: Mech Hunter vs Secret Hunter,

Burn Mage vs Khadgar Mage

Mech

1	Mecharoo	x2
1	Tracking	x2
2	Bomb Toss	x2
2	Fireworks Tech	x2
2	Galvanizer	x2
2	Venomizer	x2
2	Whirliglider	x2
3	Animal Companion	x2
3	Spider Bomb	x2
3	Ursatron	x2
4	Explodinator	x2
4	Replicating	x2
5	Leeroy Jenkins	x1
5	Wargear	x2
5	Zillax	x1
6	Missile Launcher	x2



Secret

1	Secretkeeper	x2
2	Explosive Trap	x1
2	Freezing Trap	x2
2	Rat Trap	x2
2	Snake Trap	x1
2	Snipe	x1
2	Sunreaver Spy	x2
3	Animal Companion	x2
3	Deadly Shot	x1
3	Eaglehorn Bow	x2
3	Kill Command	x2
3	Masked Contender	x2
3	Unleash the Hounds	x1
4	Lifedrinker	x2
4	Marked Shot	x2
4	Balied Arrow	x1
5	Subject 9	x1
5	Zillax	x1
6	Unleash the Hounds	x1
10	Zul'jin	x1

## Disclaimer 2

When collecting individual deck data, there could be literally hundreds of decks of the same archetype, but SLIGHTLY different decks (swapping out one or two or three cards = LOTS of possible combinations. Not all decks in an archetype are perfectly identical)

Mech

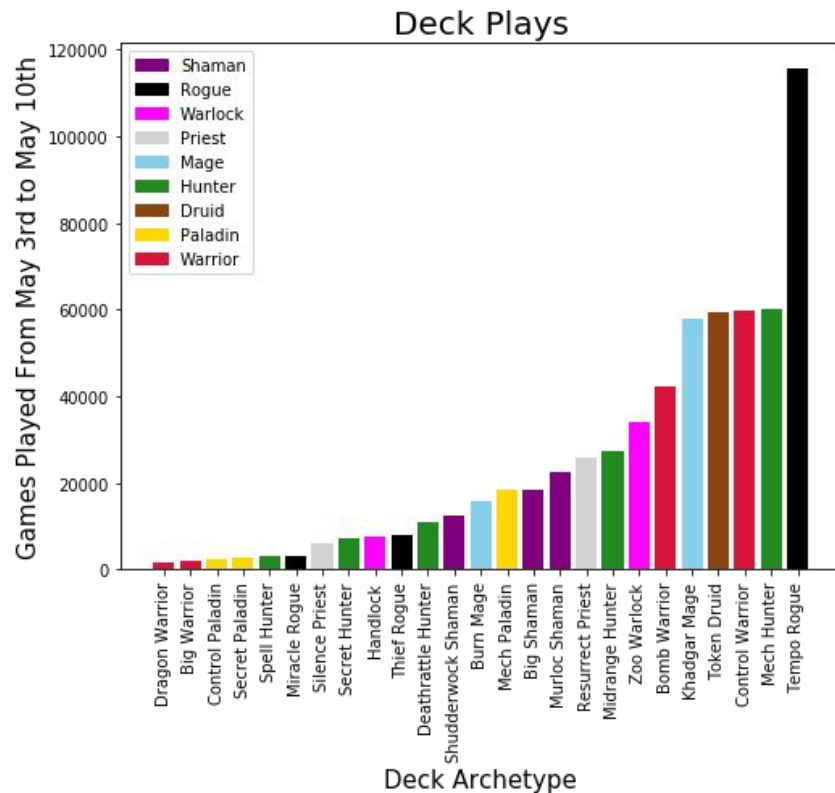
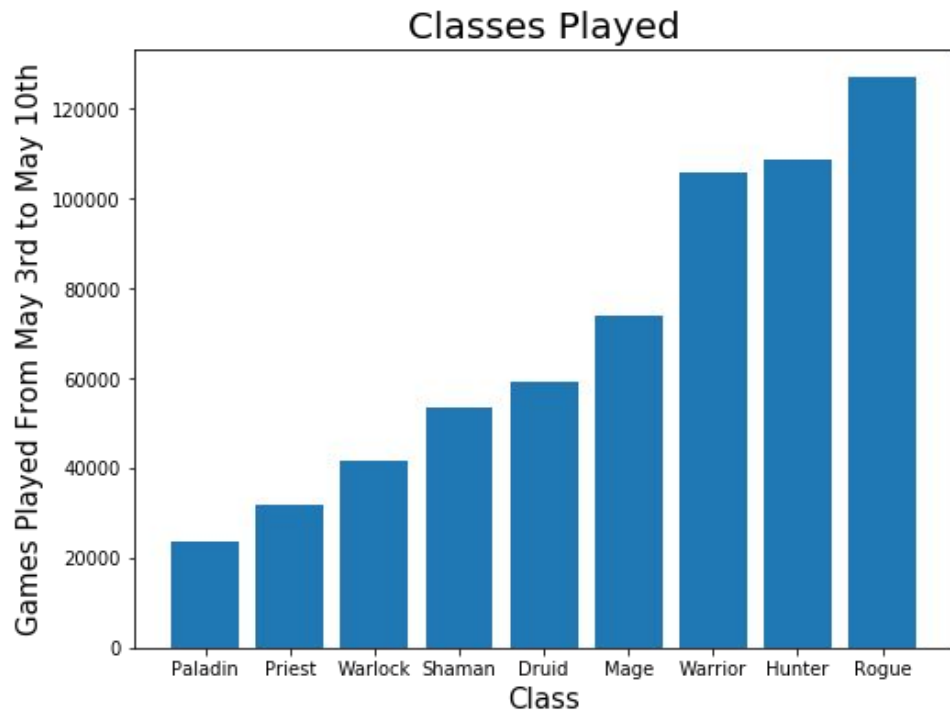
1	Mecharoo	x2
1	Tracking	x2
2	Bomb Toss	x2
2	Fireworks Tech	x2
2	Galvanizer	x2
2	Venomizer	x2
2	Whirliglider	x2
3	Animal Companion	x2
3	Spider Bomb	x2
3	Ursatron	x2
4	Explodinator	x2
4	Replicating	x2
5	Leeroy Jenkins	x1
5	Wargear	x2
5	Zillax	x1
6	Missile Launcher	x2

Also Mech

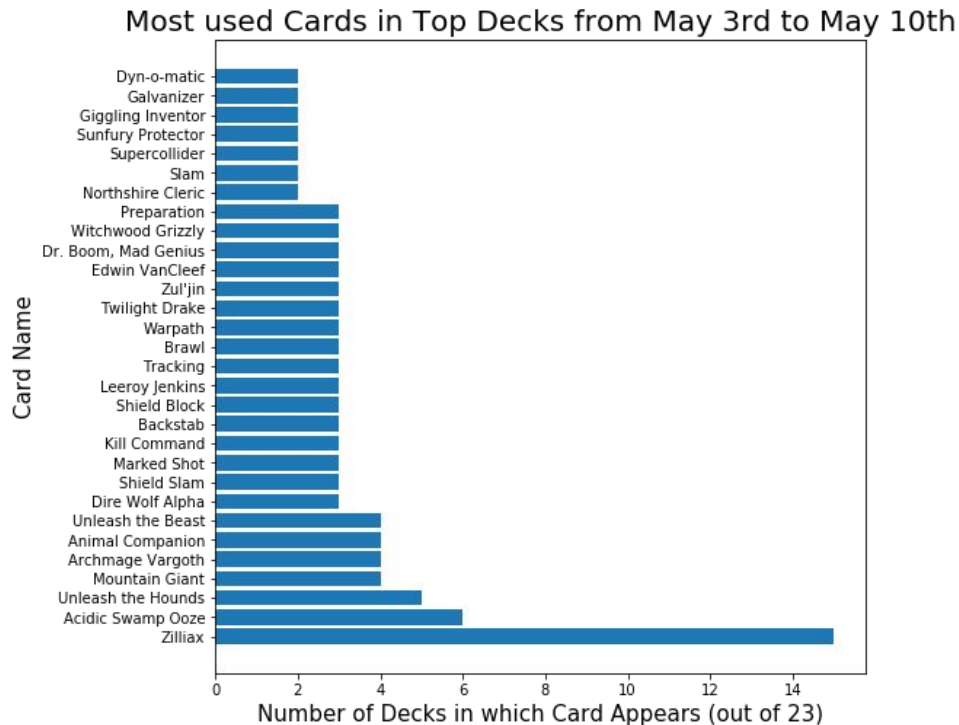
1	Mecharoo	x2
2	Bomb Toss	x2
2	Dire Wolf Alpha	x2
2	Fireworks Tech	x2
2	Galvanizer	x2
2	Upgradeable FrameBot	x1
2	Venomizer	x2
3	Spider Bomb	x2
3	Unleash the Hounds	x2
3	Ursatron	x2
4	Explodinator	x2
4	Replicating	x2
5	Leeroy Jenkins	x1
5	Wargear	x2
5	Zillax	x1
6	Missile Launcher	x2
7	Boommaster	x1



# Deck Archetype Matchups - Week of May 3



# The Most Popular Cards

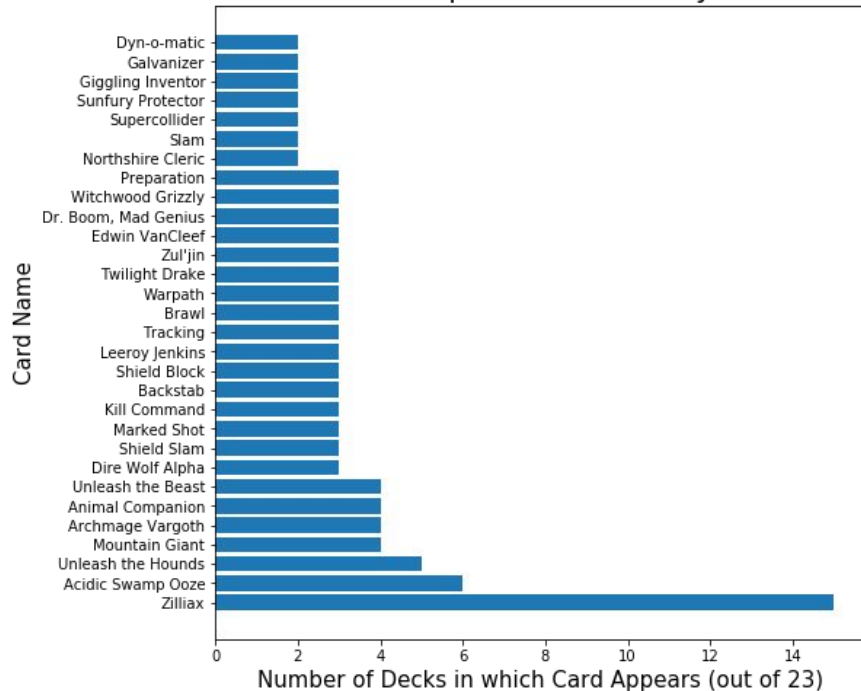


Unsurprisingly, the 3 out of 4 top cards are Neutral cards (these can be played in any deck).

Surprisingly, of the 23 most common archetypes, only ~300 of the 913 available cards in *Standard* are currently used in the meta!

# The Most Popular Cards

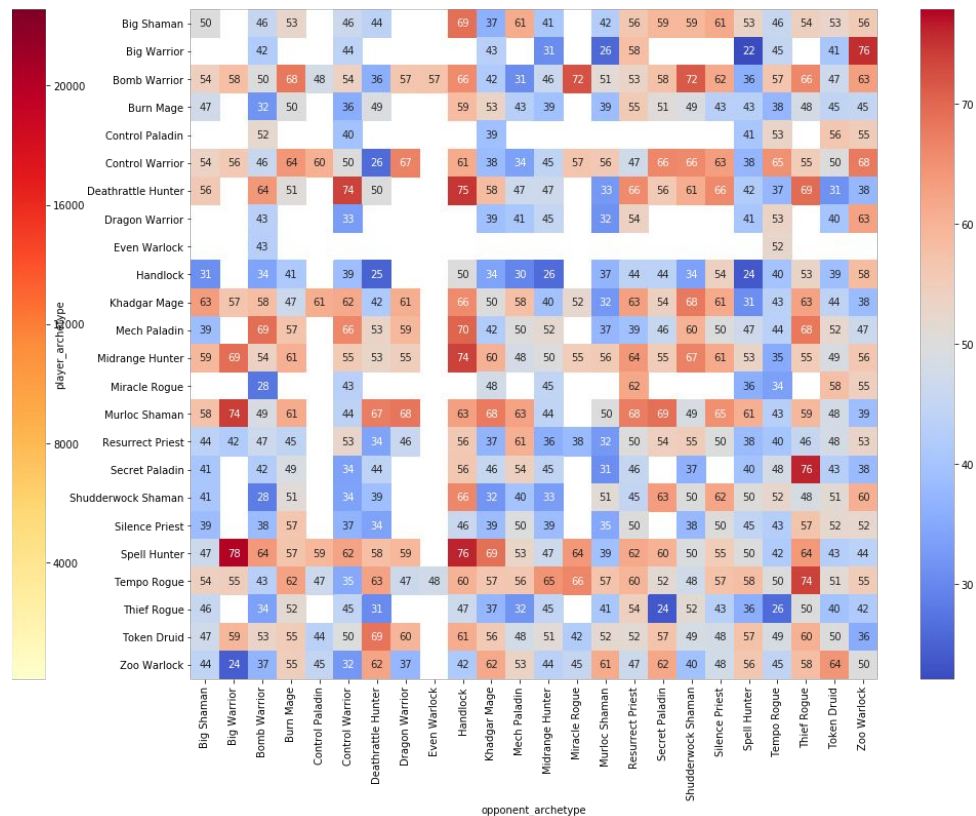
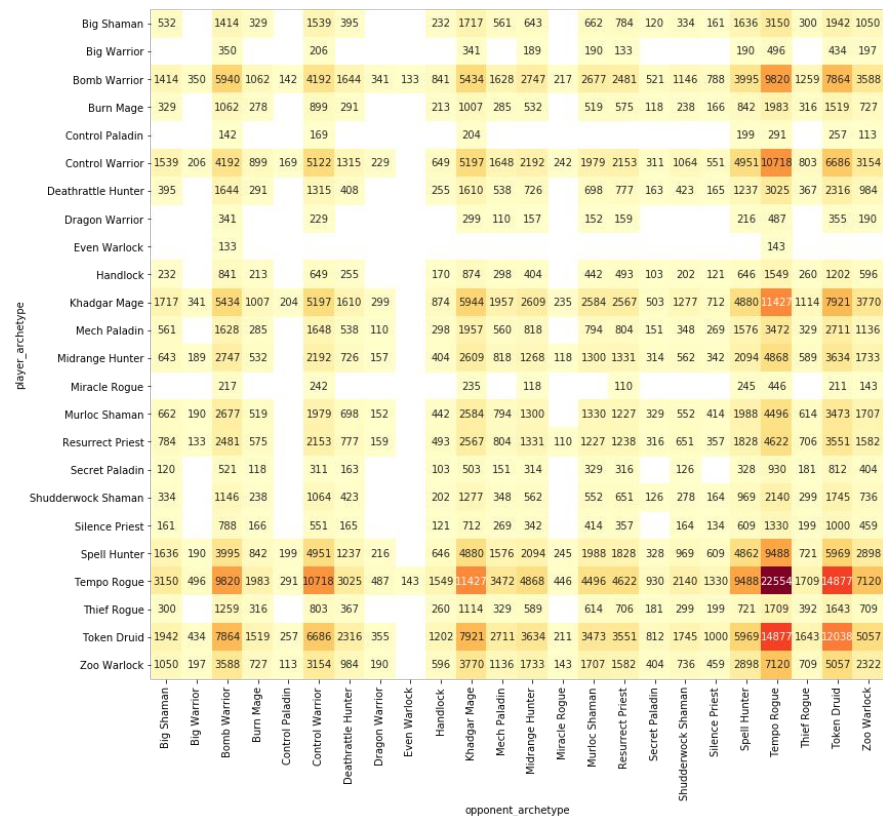
Most used Cards in Top Decks from May 3rd to May 10th



“Zilliax” - this ridiculous card seems like it’s in every deck, and just so you know, it basically is.

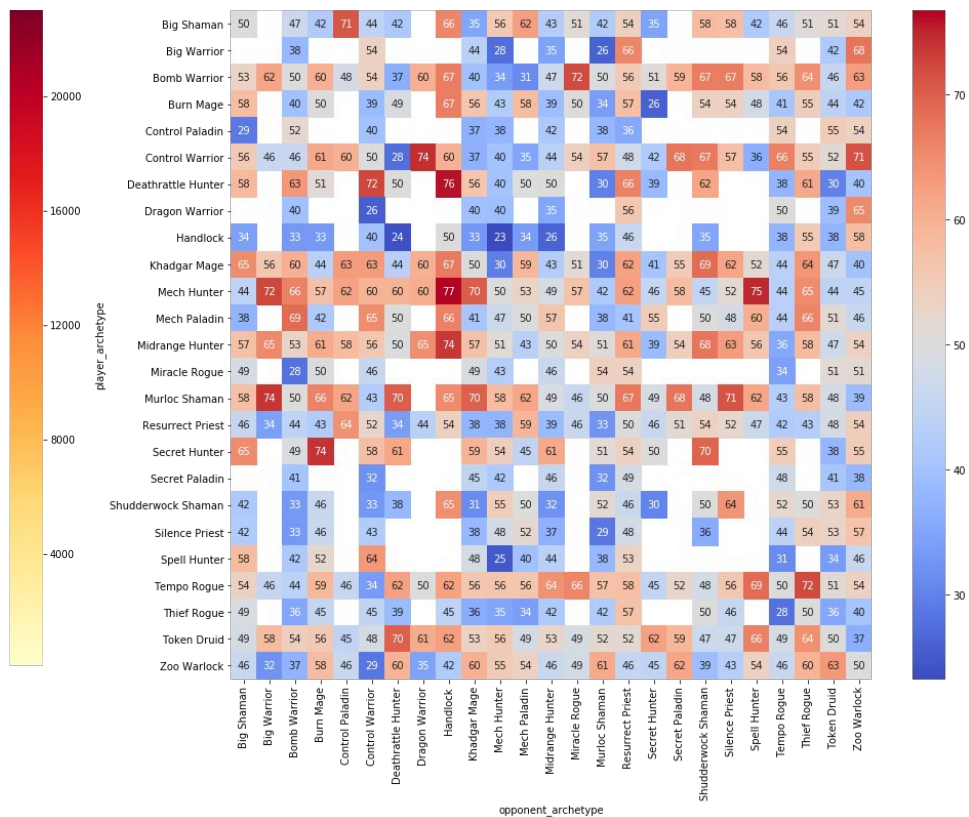


# Deck Archetype Matchups - May 2nd



# Deck Archetype Matchups - May 10th

player_archetype	Big Shaman	576	1234	443	124	1810	296		222	1684	1900	528	742	141	673	815	233		391	172	126	3421	246	1587	992	
	Big Warrior		169			171				255	200		153		123	114						403		234	137	
	Bomb Warrior	1234	169	3076	1076	203	3811	733	168	542	3787	4052	1266	1934	268	1581	1669	425	214	858	469	213	7409	636	4233	2357
	Burn Mage	443		1076	568		1552	304		228	1395	1521	386	712	129	585	705	306		369	177	129	2757	253	1423	856
	Control Paladin	124		203			249				297	298		151		118	123						488		318	164
	Control Warrior	1810	171	3811	1552	249	6212	1054	191	593	5400	5989	1706	2408	400	1922	2314	695	251	1153	458	318	11706	687	5448	3182
	Deathrattle Hunter	296		733	304		1054	188		150	1066	1107	369	468		390	497	114		232			1960	178	1097	584
	Dragon Warrior			168			191				203	215		125			114						338		227	107
	Handlock	222		542	228		593	150		140	769	705	241	331		299	392			196			1238	173	746	497
	Khadgar Mage	1684	255	3787	1395	297	5400	1066	203	769	5364	5403	1811	2391	322	2053	2295	696	294	1174	548	355	10951	797	5404	3221
	Mech Hunter	1900	200	4052	1521	298	5989	1107	215	705	5403	6166	1747	2610	333	2090	2130	757	293	1206	628	287	11220	689	5497	3157
	Mech Paladin	528		1266	386		1706	369		241	1811	1747	512	809		681	800	174		333	216	118	3372	280	1906	1021
	Midrange Hunter	742	153	1934	712	151	2408	468	125	331	2391	2610	809	1266	164	1102	1246	329	160	546	275	139	4752	375	2546	1526
	Miracle Rogue	141		268	129		400				322	333		164		120	145						632		266	202
	Murloc Shaman	673	123	1581	585	118	1922	390		299	2053	2090	681	1102	120	878	962	232	164	462	310	151	3959	341	2231	1240
	Resurrect Priest	815	114	1669	705	123	2314	497	114	392	2295	2130	800	1246	145	962	1126	263	197	603	308	188	4402	510	2508	1505
	Secret Hunter	233		425	306		695	114			696	757	174	329		232	263	278		217			1558		516	329
	Secret Paladin			214			251				294	293		160		164	197						533		375	216
	Shudderwork Shaman	391		858	369		1153	232		196	1174	1206	333	546		462	603	217		252	116		2246	208	1178	643
	Silence Priest	172		469	177		458				548	628	216	275		310	308			116			1172	111	617	378
	Spell Hunter	126		213	129		318				355	287	118	139		151	188						596		300	195
	Tempo Rogue	3421	403	7409	2757	488	11706	1960	338	1238	10951	11220	3372	4752	632	3959	4402	1558	533	2246	1172	596	12301	1094	10369	6186
	Thief Rogue	246		636	253		687	178		173	797	689	280	375		341	510			208	111		1094	156	901	423
	Token Druid	1587	234	4233	1423	318	5448	1097	227	746	5404	5497	1906	2546	266	2231	2508	516	375	1178	617	300	10369	901	6080	3223
	Zoo Warlock	992	137	2357	856	164	3182	584	107	497	3221	3157	1021	1526	202	1240	1505	329	216	643	378	195	6186	423	3223	1796
Big Shaman																										
Big Warrior																										
Bomb Warrior																										
Burn Mage																										
Control Paladin																										
Control Warrior																										
Deathrattle Hunter																										
Dragon Warrior																										
Handlock																										
Khadgar Mage																										
Mech Hunter																										
Mech Paladin																										
Midrange Hunter																										
Miracle Rogue																										
Murloc Shaman																										
Resurrect Priest																										
Secret Hunter																										





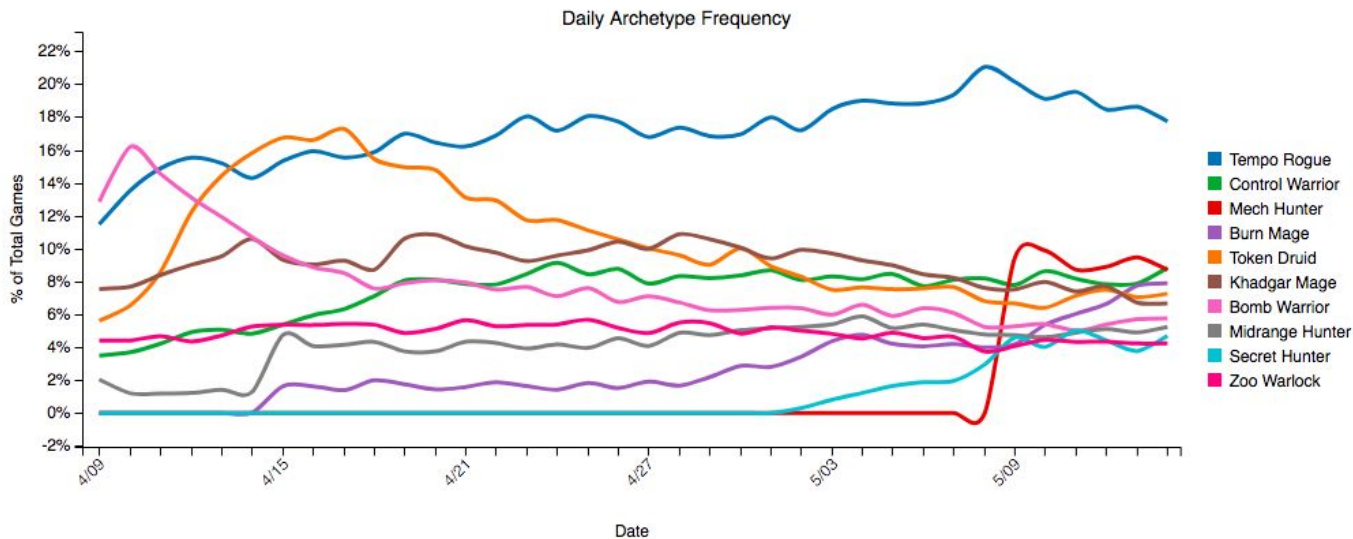
# Meta changes over time

The Meta changes quickly:

- loss of archetype “Even Warrior”
- Fast rise and addition of “Mech Hunter”, “Secret Hunter” between the week window of collected data

But not quickly enough:

- Tempo Rogue still gained, dominating the meta right now



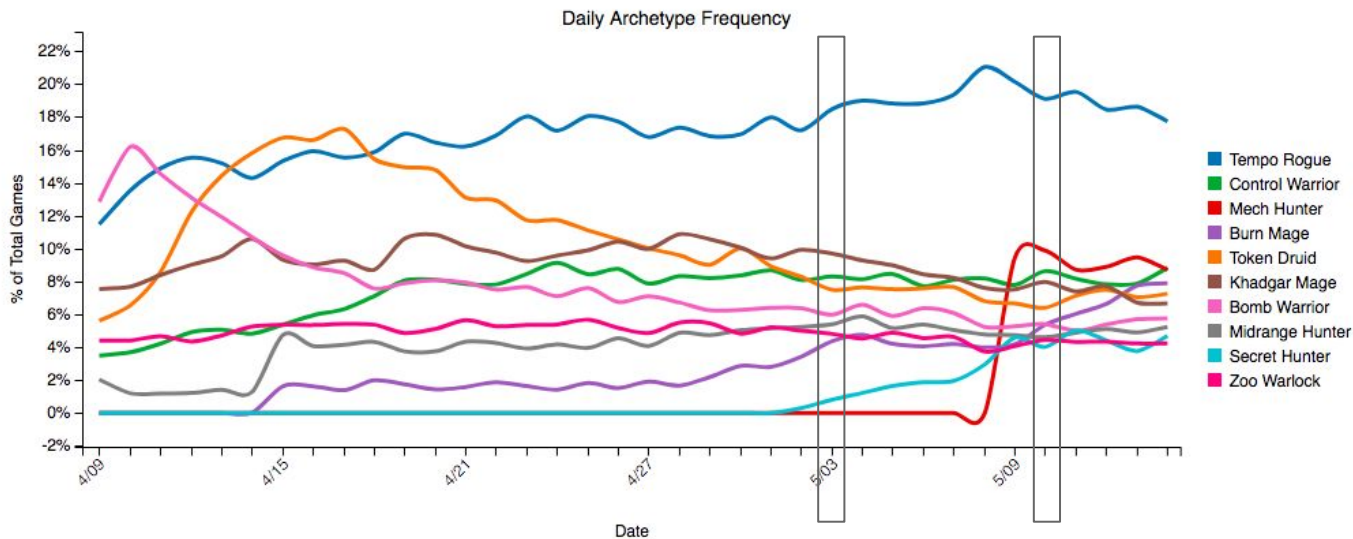
# Meta changes over time

The Meta changes quickly:

- loss of archetype “Even Warrior”
- Fast rise and addition of “Mech Hunter”, “Secret Hunter”

But not quickly enough:

- Tempo Rogue still gained, dominating the meta right now



# Rarity consistency

Obnoxiously Good Legendary



Hot Garbage Legendary



Is it possible to predict the rarity of a card based on card features?

# Predicting Legendaries

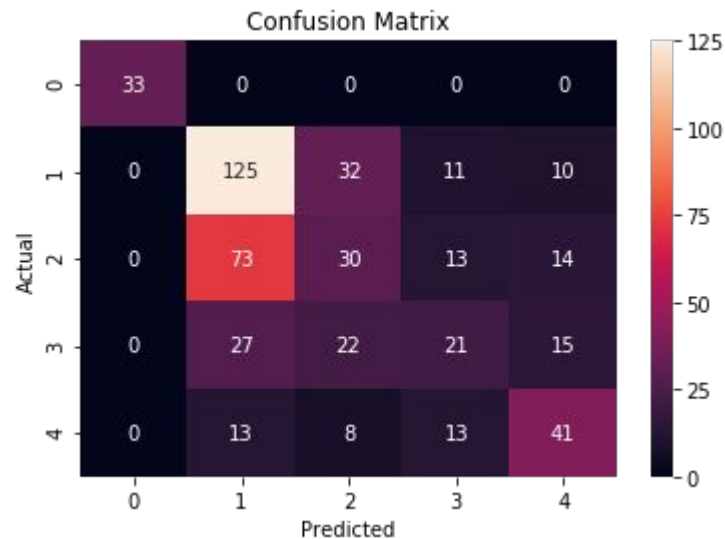
- All numerical features (attack, mana cost, etc)
- Dummies for all categorical features
- Multiclass Prediction of 5 types of rarities

LogReg	Rarity	Count
0	Free	133
1	Common	713
2	Rare	519
3	Epic	339
4	Legendary	299

# Predicting Legendaries

- All numerical features (attack, mana cost, etc)
- Dummies for all categorical features
- Multiclass Prediction of 5 types of rarities

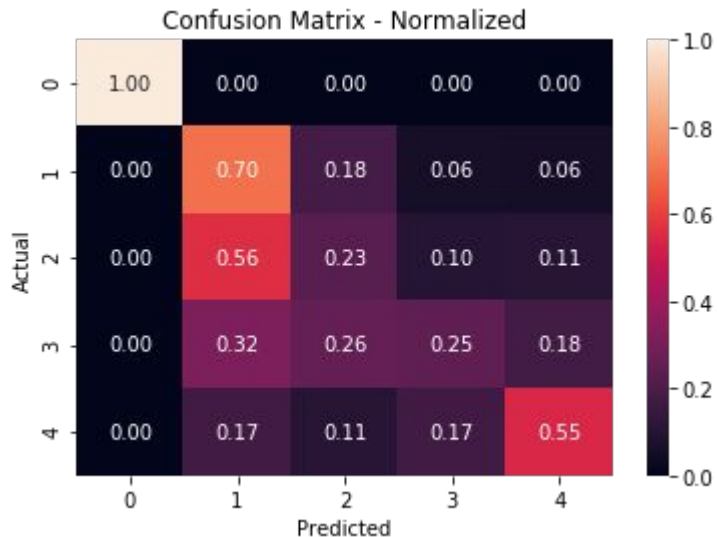
LogReg	Rarity	Count
0	Free	133
1	Common	713
2	Rare	519
3	Epic	339
4	Legendary	299





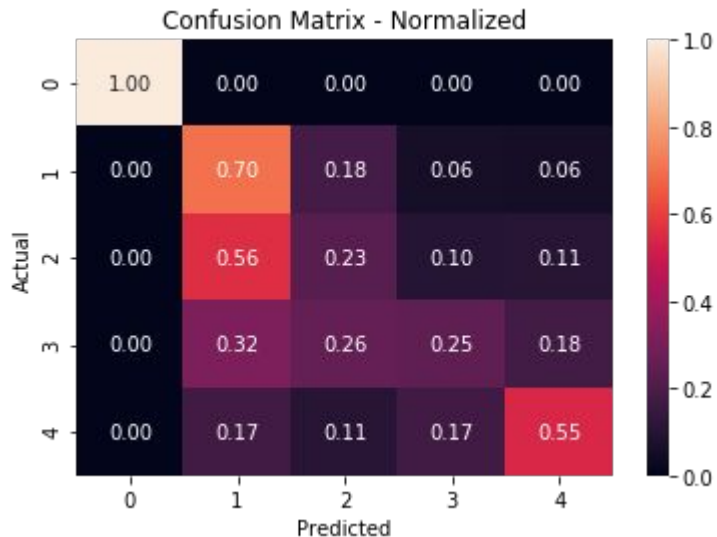
# Predicting Legendaries

- Train score: 0.57, Test score: 0.50
- Why so good?!



# Predicting Legendaries

- Train score: 0.57, Test score: 0.50
- Why so good?!



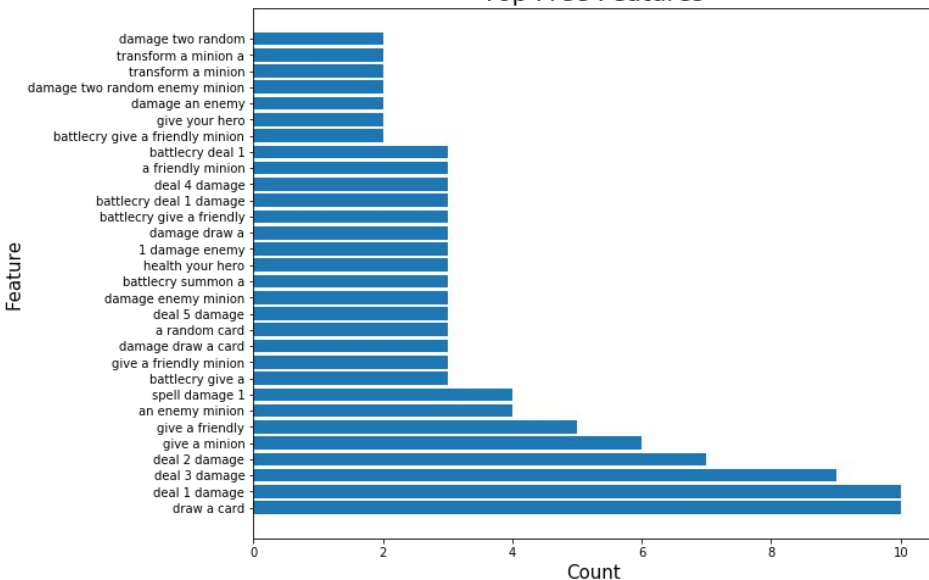
Looking at coefficients:  
....oh.

	coef	feature
37	2.585011	cardset_Basic
56	0.058437	faction_Horde
80	0.039245	type_Spell
11	0.030682	mech_Charge
75	0.025265	race_Murloc

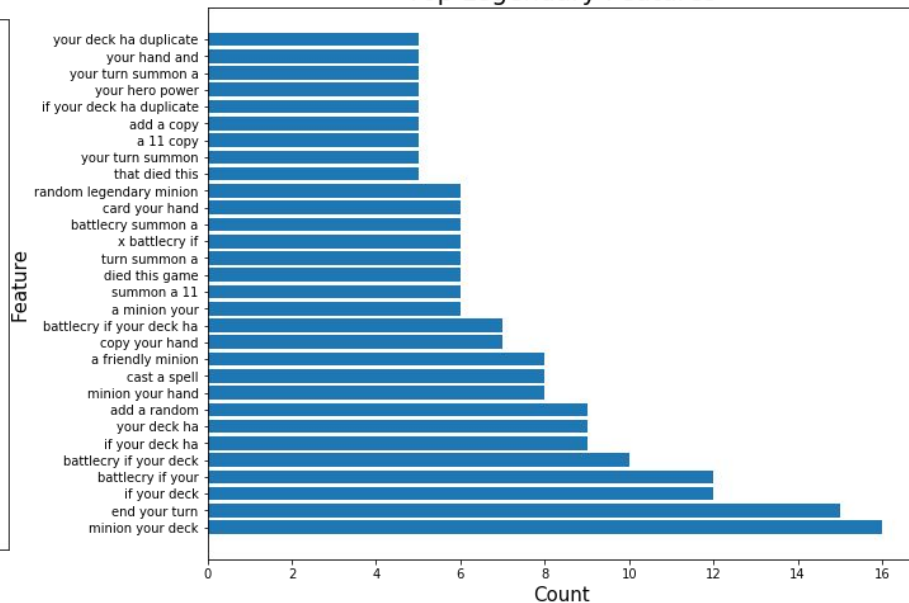
# Predicting Legendaries

- NLP!
- Visually, ngram length (3,5) most interesting!

Top Free Features



Top Legendary Features



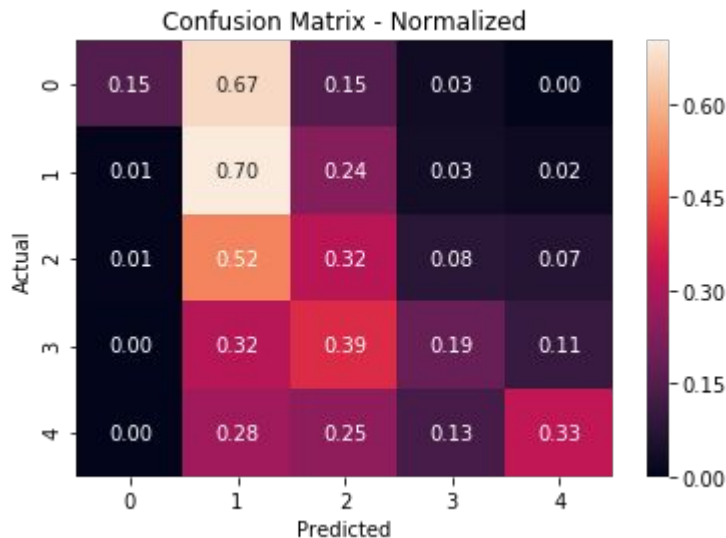
# Predicting Legendaries

Best Model:

- Train score: 0.58, Test Score: 0.42
- Ngram length (1,2) - surprising!

Definitely less help on classifying  
'Free' cards without the card set.

We can do more to try to predict the rarity  
of a card - but it seems inconsistent (ramping).



# Reflection and Conclusion

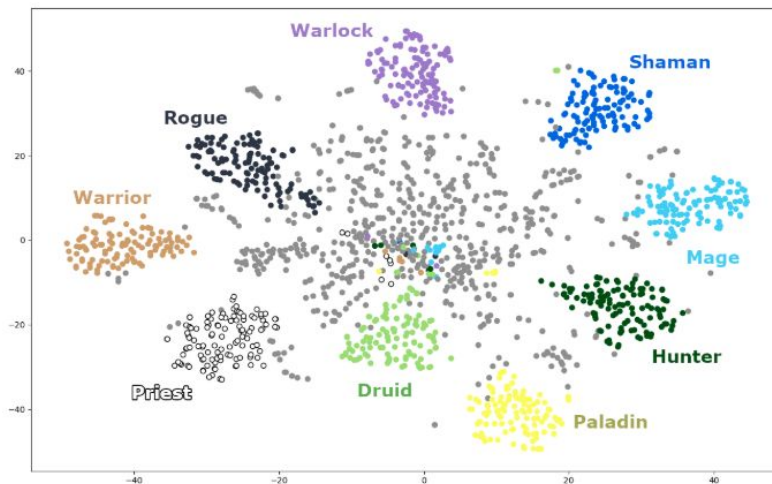
- First reaction: SO MUCH DATA to look at and try - spread myself too thin
- All the data I did not use (not collected):
  - Top decks of other expansions
  - Standard ladder vs Wild ladder
  - Different Tier Ranks
  - Other gameplay modes (Arena, Tavern Brawl, Solo play)
  - The other 2000 NON-collectible cards
  - NOT EVEN AN ATTEMPT TO LOOK AT THE ACTUAL GAMEPLAY RULES
- Entire companies exist just to work on Hearthstone Analytics
  - HearthSim, MetaStats, HearthArena





# Stuff I'm still excited to work on:

- Predicting Win Rates of a Deck
  - doc2vec: input is deck as document, cards are words
  - Use model.predict\_proba to win/lose (binary outcome)
  - Especially with legendaries (since some are basically useless, novelty only)
  - Interactions between cards!
- Unsupervised Learning: Class/Deck clustering using word2vec / LDA



**Thank You!**



**Questions?**