



An analysis on competitive PokéMon teams

Generation	A group of core series Pokemon games released around the same time. Abbreviated as "Gen".
Smogon	A competitive Pokemon battling community.
Showdown	A platform that hosts Pokemon battles.
OU	A Smogon battle format where "Over Used" pokemon are allowed.

Datasets Used

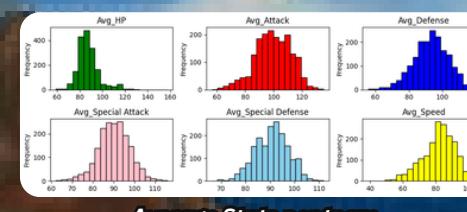


*Battle logs were scraped and processed into tables which generally contain various aspects of the battle. One table contains the first player's name, elo, team composition (6 columns of Pokemon), lead Pokemon, battle turn count, and result (Win/Lose). Another contains the name and team composition of each of the two players.

Exploratory Data Analysis

Base Stats

We identified the average base stats of each team and the average of all lead Pokemon to get an idea of the spread of base stats.



Types

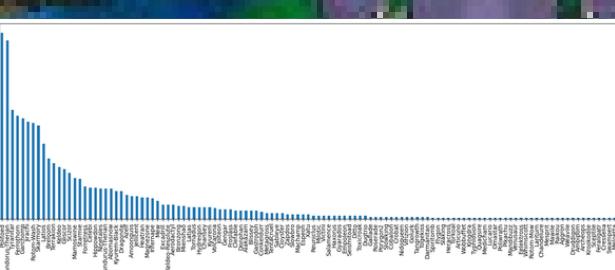
Types are pivotal to Pokemon matchups as such we identified if having a certain type may swing the win to the player.

ELO and Turncount

ELO is the main ladder rating that dictates a player's position in the format and we compared it if there would be any difference in match Turncounts with higher or lower ELOs.

Lead Pokemon Frequency

Players choose a Pokemon to go first and it may set the tempo for the battle. We identified which Pokemon are sent out as leads the most:



Statistical Inferences

Multiple Groups:

Pokemon Usage Across ELO

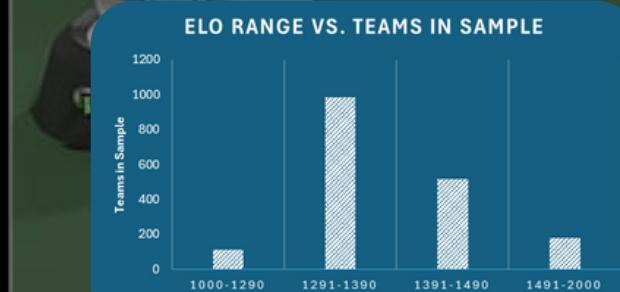
We bracketed players around four different ELO intervals (pictured below) and found that each group's Pokemon usage statistics are mostly the same, save for the usage rates of those in the 1291-1390 and 1491-2000 categories.

	1000-1290	1291-1390	1391-1490	1491-2000
1000-1290	False	False	False	False
1291-1390	False	False	False	True
1391-1490	False	False	False	False
1491-2000	False	True	False	False

Post-hoc Dunn's test results*

*Results generated after passing through Shapiro-Wilk, Levene, and Kruskal-Wallis

We speculate that this is because most of the teams in our sample come from the 1291-1390 range, which may cause a greater variation in Pokemon usage:



Wilcoxon: Pokemon Usage Across Time

We tried to compare Pokemon usage rates between January of 2015 and 2025 and found that there is a significant difference.

*Wilcoxon was used as a Shapiro-Wilk test indicated that the differences between usage rates were not normally distributed.

The Gen 5 OU tier likely faced metagame changes in this duration resulting to the difference we see in the paired observation test.



Some examples of this:
Excadrill was not used at all in 2015, but began being used by around 20% of teams in 2025 and this may be due to its Rapid Spin move removing hazards.
Dugtrio saw little use in 2015 (~4% usage) but its usage declined even more after its Arena Trap ability was banned due to it being a nuisance in the competitive scene of other gens.

Independent Paired Test: Dugtrio Usage

To further investigate the drop in usage for Dugtrio, an independent paired test is utilized to see if the usage of Dugtrio is significantly greater before the ban of arena trap compared to after the ban of arena trap. A Mann-Whitney U-Test is used due to one of the groups being not normally distributed. We found that, as expected, the usage of Dugtrio before the ban on Arena Trap is greater than its usage afterwards.

Chi Square: Weather Usage Across ELO

We compared the usage of weather teams across the different ELO intervals (same ranges as the Multiple Groups) through a Chi Square test and found that weather usage and ELO rating are independent of each other. This may indicate that weather remains a key mechanic during competitive play despite elo.

ELO Range	Observed Counts:		Expected Counts:	
	IsWeather	NotWeather	IsWeather	NotWeather
1000-1290	85	164	85.3572	163.6428
1291-1390	753	1598	803.1884	1539.8196
1391-1490	684	1867	572.8188	1898.1812
1491-2088	272	485	252.6436	484.3564

The definition of weather team is based on our DBSCAN Clustering and domain knowledge. Specifically, our two biggest clusters contain Pokemon teams that synergize well on weather effects – the biggest cluster focusing on setting sand, and the second cluster focusing on setting rain. It is specifically Tyranitar and Politoed that are prominent sand and rain setters respectively, with certain Pokemon working well with these two.



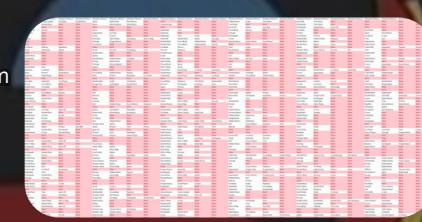
Conclusion

- Emphasis on weather-based teams
 - Teams are built around weather effects, such as sandstorm and rain. It can be argued that Pokemon that both cause those weather effects like Tyranitar (Sandstorm) and Politoed (Rain) are prevalent within the meta. Pokemon that can be built around those weather effects are chosen and are associated with one another.
 - Other archetypes are also uncovered, such as Smurf and DragMag, which match Gen5 OU metagame discussions.
- Observations on ELO and the Metagame
 - ELO was observed as having mixed correlations with either Pokemon usage and weather usage.
 - The distributions of the Pokemon may not differ greatly across ELOs, save for the 1291-1390 ELO vs. 1491-2000 ELO players; this may be due to the high amount of players in the former set.
 - Weather-based teams was not observed to be dependent on ELO.
- Metagame still develops over-time
 - Excadrill and Dugtrio experienced a change in their usage stats, possibly due to a changing perception towards its moves and abilities.
 - Changes in usage stats, however small, still reflect a changing metagame.

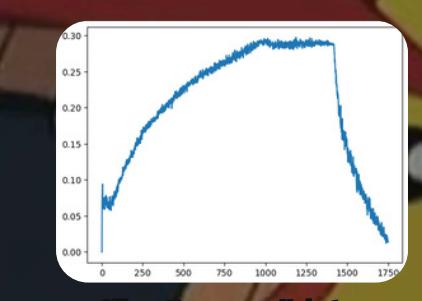
"Failed" Attempts

Exploring Pokemon Moves

We wanted to explore Pokemon moves further from the Showdown battle logs, but there was not enough data: only around 20% of Pokemon in the battle logs had all (four) of their moves recorded.



Red cells indicate missing moves



Silhouette score vs. K clusters

K-medoids clustering

We tried clustered Pokemon teams into "archetypes" using K-medoids clustering, with Jaccard distance as a metric. A K=500 provided a good balance between getting a good Silhouette score and number of clusters... but the clusters generated showed a lot of noise. We soon discovered that DBSCAN Clustering yielded fewer and better clusters that relate to our expected archetypes.

Goodness of fit

We tried to test whether the Smogon usage statistics aligned with the teams used in the scraped Showdown battle logs. We abandoned it because the Showdown data set isn't big enough to form any meaningful conclusion related to the population of battles.

Bayesian Inference

We applied Bayesian Inference to the Pokemon that influenced our Chi-Square results the most. However, the resulting Bayes Factors provided only anecdotal support, offering no strong or conclusive evidence even after updating our prior beliefs with new data.

