

# POKESTATÍSTICA

Análise exploratória de dados da Pokédex

**Grupo:**

Caio Passos Torkst Ferreira

Vitor Neves dos Santos

Ryan Filipe de Mendonça Borges

Daniel Salvador Motta

João Guilherme de Oliveira Ribeiro Kongevold

– página branca intencional –

# 1. Introdução

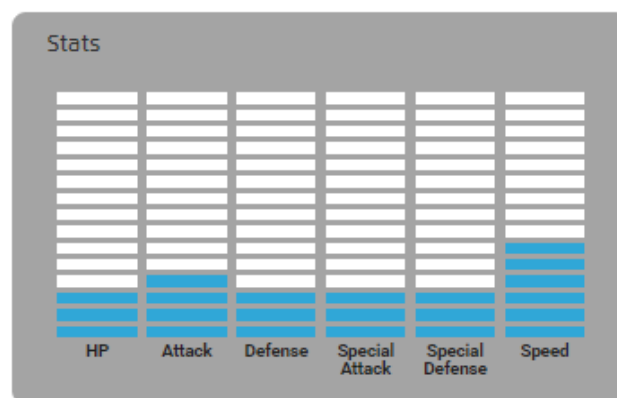
Os jogos da franquia Pokémon são baseados na captura, treinamento e batalhas entre criaturas chamadas Pokémon, cada uma com características únicas.

A *Pokédex* é uma base de dados oficial da Nintendo que contém informações detalhadas sobre cada Pokémon, incluindo seus atributos estatísticos, tipos e habilidades que determinam seu desempenho nas batalhas.

Name	Type	Total	HP	Attack	Defense	Sp. Atk	Sp. Def	Speed
Pikachu	ELECTRIC	320	35	55	40	50	50	90

## 1.1. Principais Estatísticas (atributos)

- **HP (Hit Points)** – Representa a quantidade de vida do Pokémon. Quanto maior, mais dano ele pode suportar antes de ser derrotado.
- **Attack (ATK)** – Determina o dano que o Pokémon causa com ataques **físicos**.
- **Defense (DEF)** – Indica a resistência do Pokémon contra ataques **físicos**.
- **Special Attack (Sp. ATK)** – Mede o poder de ataques especiais (geralmente ataques elementais como fogo e água).
- **Special Defense (Sp. DEF)** – Define a resistência do Pokémon contra ataques especiais.
- **Speed (SPD)** – Afeta a ordem de ataque na batalha; Pokémon mais rápidos atacam primeiro.



## 1.2. Tipos de Pokémon

Cada Pokémon tem um ou dois tipos, como Fogo (Fire), Água (Water), Dragão (Dragon), entre outros. Os tipos influenciam a força e fraqueza contra outros Pokémon (exemplo: Água é forte contra Fogo, mas fraca contra Elétrico). Os tipos dos Pokémon são uma mecânica essencial da franquia, funcionando em um sistema de vantagens e desvantagens semelhante ao jogo "pedra, papel e tesoura".

### 1.2.1 Combinação de Tipos

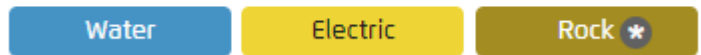
Quando um Pokémon tem **dois tipos**, ele herda características de ambos, tornando-se mais versátil ou vulnerável. Por exemplo:

- *Charizard* (Fogo/Voador)

Type



Weaknesses



**Observação:** Tanto o tipo Fogo como o tipo Voador são fracos contra o tipo Pedra, portanto o *Charizard* possui uma fraqueza que chamamos de 4x, isto é, qualquer dano de pedra será **quadruplicado** contra ele.

### 1.2.2. Quantidade de Combinações

No momento, existem **18 tipos diferentes** de Pokémon. Que podem ser combinados entre si, uma mecânica usada para manipular as forças e fraquezas de um Pokémon.

Como cada pokémon pode ter até dois tipos:

$$C(18, 2) = \frac{18!}{2!(18-2)!} = \frac{18 \times 17}{2} = 153$$

Temos um número total de combinações possíveis de tipos duplos de **153**, e somando os 18 tipos individuais, totalizam **171 combinações de dois tipos**.

O interessante é que, mesmo com mais de 25 anos de gerações com Pokémons novos usando combinações diferentes, **ainda não existe Pokémon** de todas as combinações possíveis.

## 1.3. Gerações de Pokémon

A franquia Pokémon é dividida em diferentes **gerações**, que representam os lançamentos principais dos jogos e a introdução de novos Pokémon, mecânicas e mudanças na jogabilidade. Atualmente, existem **nove gerações**, cada uma com sua própria Pokédex e ambientação dentro do mundo Pokémon.

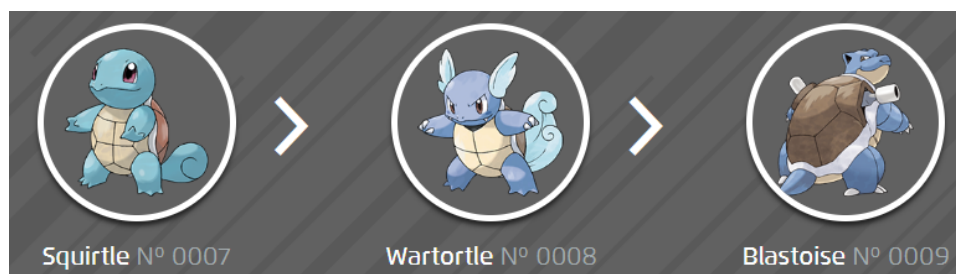
Entretanto, neste trabalho vamos cobrir somente as **sete primeiras gerações**, que já contém **mais de 800 Pokémon diferentes** - mais de 1000 contando com diferentes formas do mesmo Pokémon -.

## 1.4. Estágios de Evolução dos Pokémon

Nos jogos **Pokémon**, muitas espécies passam por **estágios de evolução**, nos quais mudam de forma, ganham novas habilidades e se tornam mais fortes.

Os Pokémon podem passar por até **duas evoluções**, se tornando progressivamente mais poderoso. Exemplo:

- **Squirtle → Wartortle → Blastoise**



Além da evolução tradicional, alguns Pokémon podem mudar de forma:

- **Mega Evolução** - (Charizard → Mega Charizard X/Y)
- **Formas Regionais** - (Meowth → Meowth de Alola)

## 2. Motivos para a Análise da Pokédex

Nesta análise exploratória, nosso objetivo é investigar padrões, correlações e insights dentro dos dados da Pokédex. Exploraremos questões como:

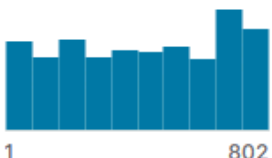
- **Entendimento do balanceamento do jogo**
  - Comparar a distribuição dos atributos (atributos) entre si.
  - Quais são os Pokémon mais fortes em termos de atributos (atributos)?
- **Descoberta de correlações**
  - Existem padrões entre os tipos de Pokémon e seus atributos?
  - ex: Pokémon do tipo Fogo tendem a ter mais Ataque?
- **Insights sobre a distribuição dos Pokémon**
  - Quais são os tipos mais raros e mais comuns?
  - Existe uma relação entre peso, altura e estatísticas de combate?
- **Aplicações em Estratégia Competitiva (Análise futura)**
  - Quais Pokémon têm o melhor custo-benefício para batalhas?
  - Existe um padrão nos atributos dos Pokémon mais usados em torneios?

Com isso, por meio da visualização de dados e análise estatística, buscamos entender melhor a diversidade dos Pokémon e como suas características impactam o jogo.

### 3. Base de Dados

Foi selecionada uma base de dados encontrada no site Kaggle, disponível em: [link](#), colocada no ar pelo usuário Myles O'Neill. Na página existem 7 tabelas, porém somente a “pokemon.csv” foi usada nas análises.

A tabela é composta de 1061 linhas e 26 colunas, sendo cada linha correspondente a uma forma de Pokémon (existem espécies repetidas, porém com formas diferentes).

# index the pokedex number for this pokemon	Δ species the name of this pokemon	Δ forme used to differentiate different forms of a single pokemon species	Δ type1 the first type of this pokemon	Δ type2 the secondary type of this pokemon (if it has one)
	Unown 3% Vivillon 2% Other (1013) 95%	Unown (One form) 3% Minior (Meteor For... 1% Other (1027) 97%	Water 12% Normal 12% Other (805) 76%	[null] 49% Flying 13% Other (398) 38%
1	Bulbasaur	Bulbasaur	Grass	Poison
2	Ivysaur	Ivysaur	Grass	Poison
3	Venusaur	Venusaur	Grass	Poison
4	Charmander	Charmander	Fire	
5	Charmeleon	Charmeleon	Fire	
6	Charizard	Charizard	Fire	Flying
7	Squirtle	Squirtle	Water	
8	Wartortle	Wartortle	Water	

## 4. Tratamento dos dados

A base de dados escolhida, apesar de possuir uma grande variedade de colunas, não possuía todos os dados necessários ou eles estavam de uma forma não desejável. O primeiro problema a ser resolvido foi acerca das colunas de peso(*weight*) e altura(*height*), que estavam representadas, respectivamente, em libras e polegadas.

Antes da conversão das bases, fez-se necessário tratar os valores como texto, visto que eles possuíam caracteres embutidos, os quais indicavam suas bases(lbs e "). Ao separar os valores numéricos foi possível convertê-los para quilogramas e centímetros.

weight	height	weight	height
15.2 lbs.	2'04"	6.9	71.1
28.7 lbs.	3'03"	13.0	99.1
220.5 lbs.	6'07"	100.0	200.7
18.7 lbs.	2'00"	8.5	61.0
41.9 lbs.	3'07"	19.0	109.2
199.5 lbs.	5'07"	90.5	170.2
19.8 lbs.	1'08"	9.0	50.8
49.6 lbs.	3'03"	22.5	99.1

**Tabela 1:** Dados originais(esquerda) e após conversões(direita)

Após a transformação das colunas, foram criadas duas novas colunas, a fim de aprimorar as análises, sendo elas a coluna de Geração e Pós-evolução. A geração de um pokémon diz respeito à qual jogo da série principal ele foi introduzido (Por exemplo, Charizard apareceu em Red & Blue, sendo da primeira geração, Oshawott apareceu em Black&White, pertencente à quinta geração), enquanto a pós-evolução se trata do próximo pokémon na linha de evolução.

A imagem acima apresenta 3 Pokémon da mesma linha evolutiva. Gastly(esquerda) é o primeiro, não possuindo pré-evolução, Haunter(meio) tem Gastly e Gengar como pré-evolução e pós-evolução, respectivamente, e Gengar(direita) tem Haunter como pré-evolução e não possui pós-evolução.





A fim de separar os Pokémon por geração, foi criada uma lista, com valores de 1 a 7, que representa o número da geração que cada Pokémon faz parte. Para a identificação, foi usado o número da pokédex, que, se dentro de um certo limite, indica a geração a qual o indivíduo faz parte. Por exemplo, os Pokémon de número até 151 se enquadram na primeira geração, de 152 até 251 estão na segunda, e assim por diante.

Após isso, foi desenvolvido um código para a criação da coluna de pós-evolução. Para isso, cada linha foi percorrida, e para cada indivíduo sua pré-evolução e seu nome eram guardados. Se o Pokémon tivesse uma pré-evolução, o programa selecionava a linha correspondente ao Pokémon anterior na linha evolutiva e colocava o nome guardado(sua pós-evolução) na nova coluna. Os Pokémon com mais de uma pós-evolução as apresentam separadas com uma ‘/’.

A última etapa de tratamento de dados foi a eliminação de colunas que não seriam usadas nas análises escolhidas pelo grupo, sendo elas as seguintes: Habilidade 1, 2 e h, Dex 1 e 2, porcentagem de machos e fêmeas, grupo do ovo 1 e 2.

ndex	species	pre-evolution	gen	post-evolution
93	Haunter	Gastly	1	Gengar

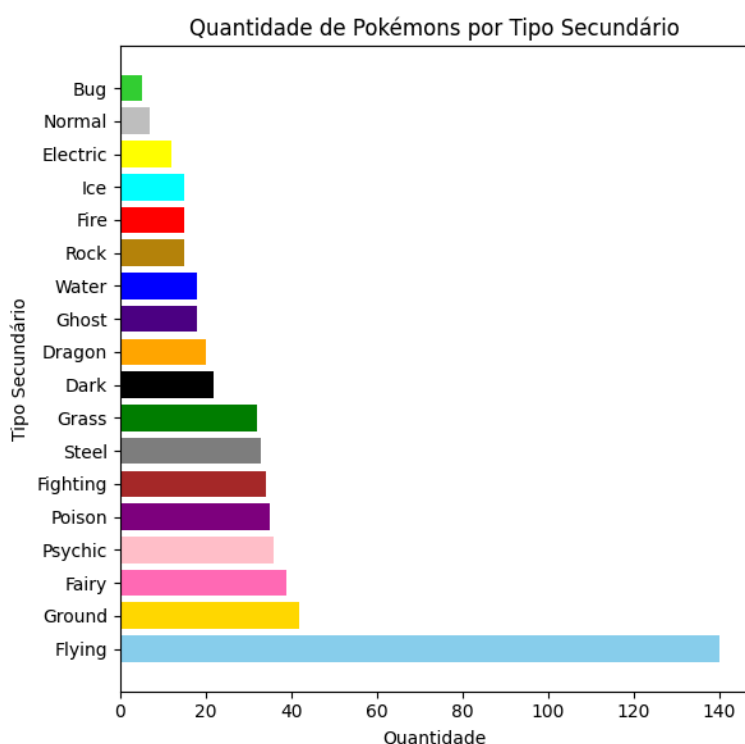
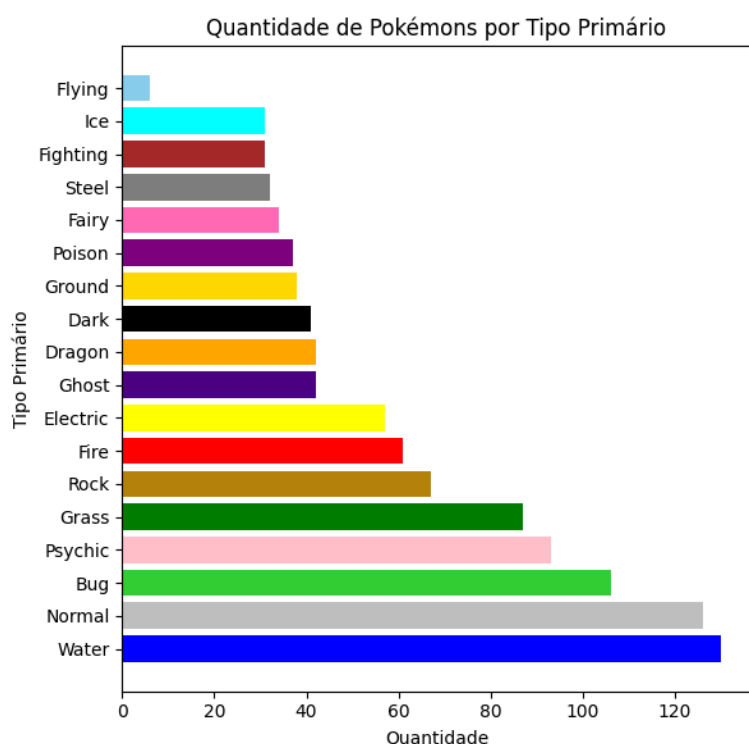
A tabela acima apresenta as colunas novas (*gen* e *post-evolution*) e algumas antigas (*ndex*, *species* e *pre-evolution*)

## 5. Distribuições de Frequência dos Pokémon

### 5.1. Quantidade de Pokémon por tipo primário e secundário.

Esses gráficos trazem e resumem quantos Pokémon possuem cada tipo.

Diferentes tipos possuem vantagem sobre outros, portanto esses dados podem ser utilizados para análise da composição de times.

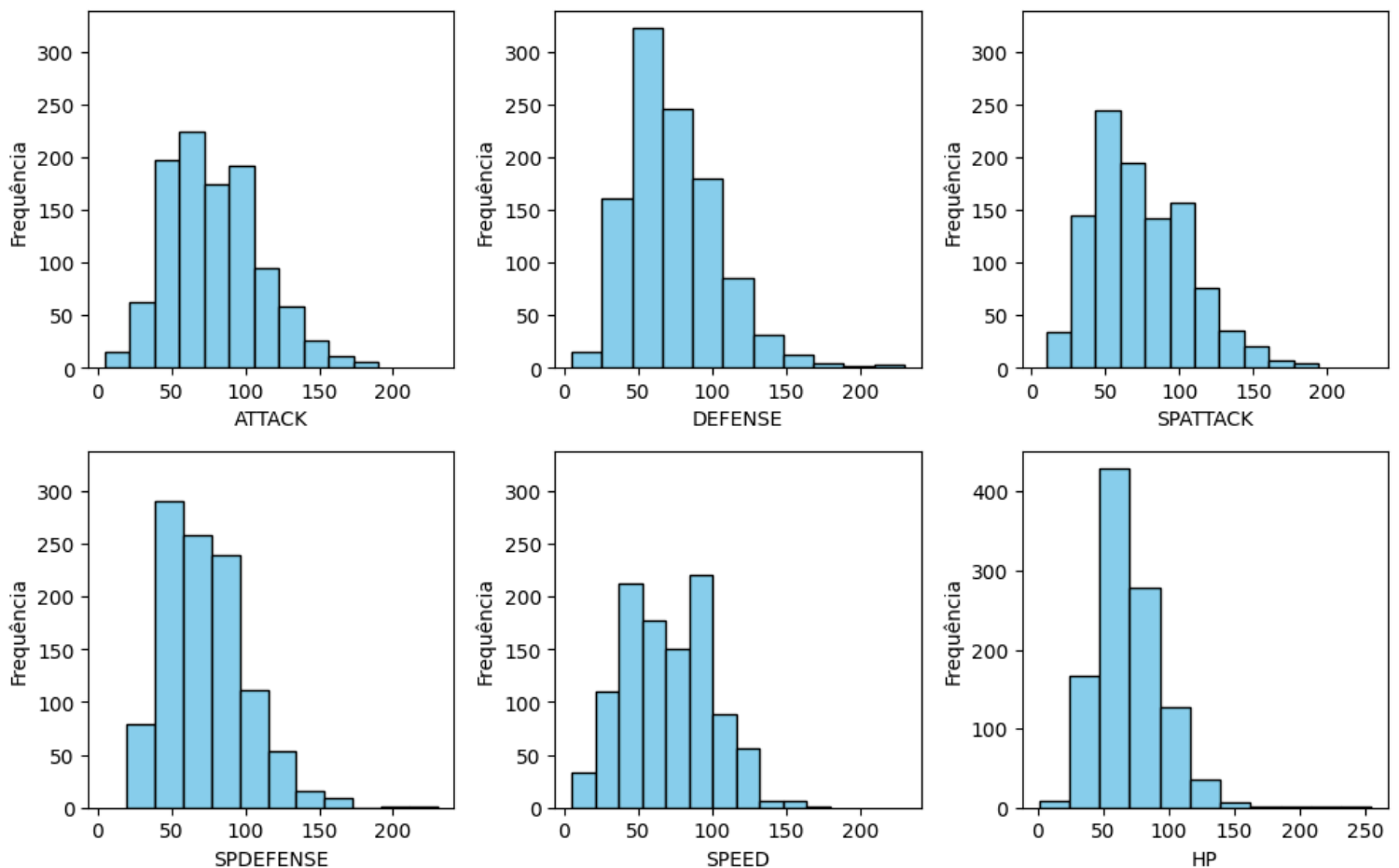


- Interessante notar como o tipo **Voador (Flying)** é o tipo **primário** mais raro, e ao mesmo tempo é o tipo **secundário** mais comum.
- O tipo **Inseto (Bug)** é o inverso, é um dos mais comuns tipos **primários**, porém o mais raro tipo **secundário**.

### 5.3. Histogramas das distribuições por atributo

Os **histogramas** da distribuição de Pokémon por atributo podem revelar **muito** sobre **como os Pokémon estão distribuídos** em termos de suas características e como esses atributos impactam as estratégias de batalha.

Histogramas de Todos os Atributos



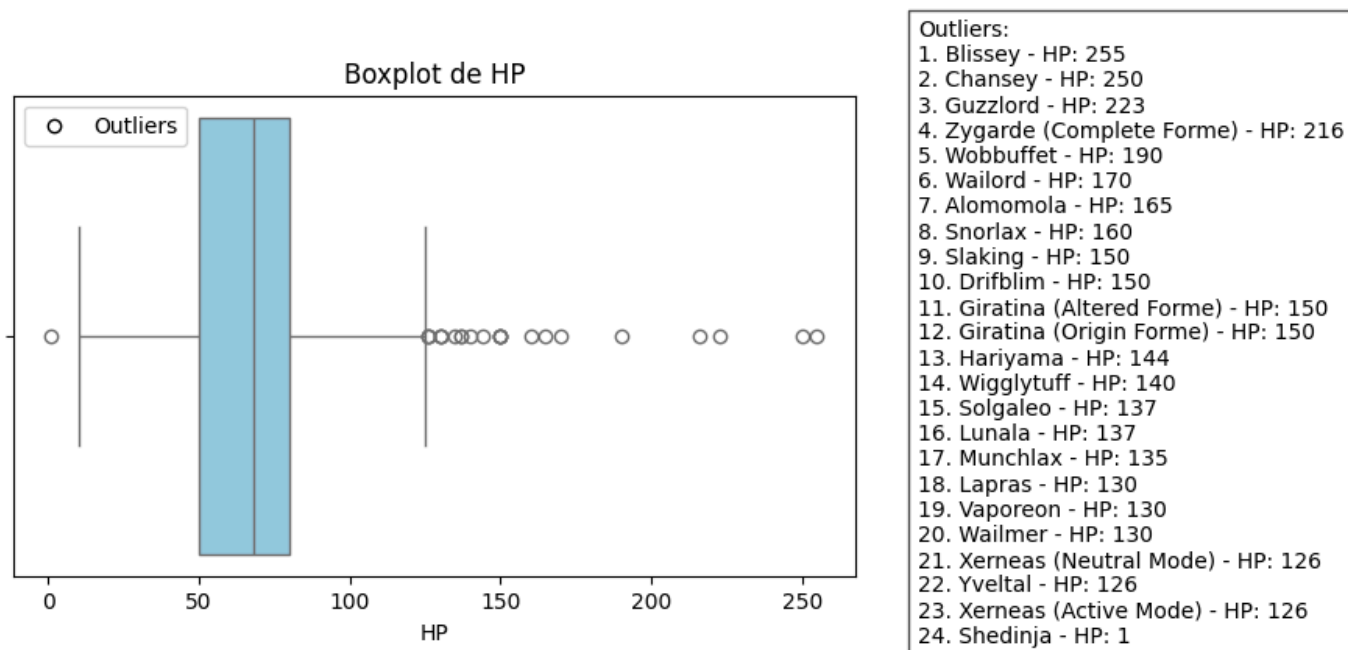
## 5.2. Medidas Descritivas das distribuições por Atributo

	hp	attack	defense	spattack	spdefense	speed	total
count	1061.000000	1061.000000	1061.000000	1061.000000	1061.000000	1061.000000	1061.000000
mean	70.041470	79.602262	73.730443	74.550424	72.911404	70.321395	441.157399
std	25.893508	31.378369	30.394899	31.975146	27.995681	29.328288	121.289505
min	1.000000	5.000000	5.000000	10.000000	20.000000	5.000000	175.000000
25%	50.000000	55.000000	50.000000	50.000000	50.000000	48.000000	336.000000
50%	68.000000	75.000000	70.000000	70.000000	70.000000	68.000000	455.000000
75%	80.000000	100.000000	91.000000	95.000000	90.000000	93.000000	520.000000
max	255.000000	190.000000	230.000000	194.000000	230.000000	180.000000	780.000000

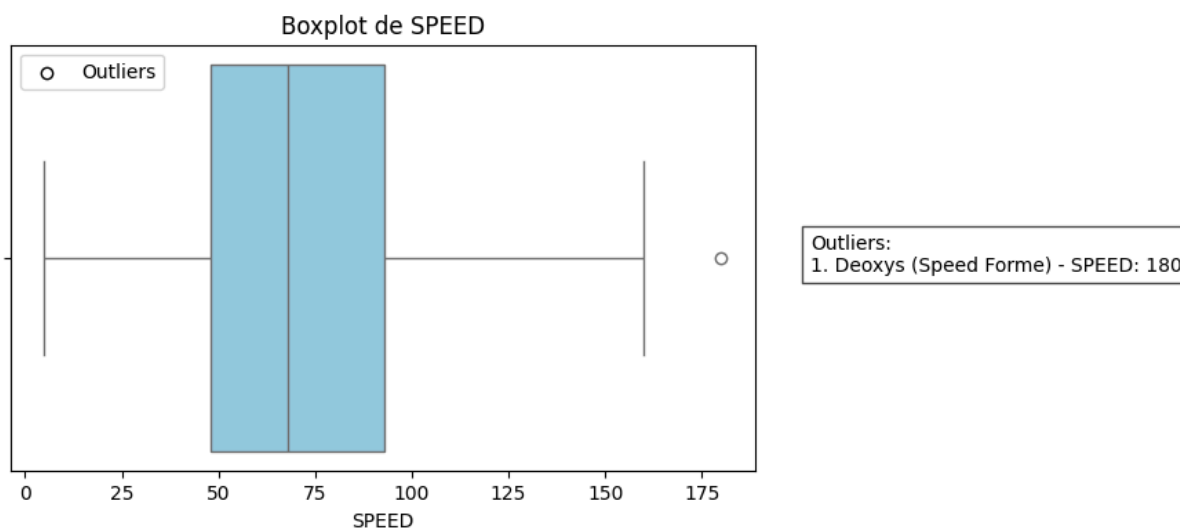
- **HP (Pontos de Vida):** Normalmente, **a grande maioria** dos Pokémon tem valores **baixos/moderados** de HP, com alguns casos excepcionais com valores muito altos. O histograma mostra uma concentração na faixa intermediária.
- **Velocidade:** O atributo Speed geralmente segue uma distribuição **mais ampla**, com muitos Pokémon apresentando velocidades baixas ou médias, e **um pequeno número de Pokémon extremamente rápidos**.
- **Ataque e Defesa:** A maioria dos Pokémon tende a ter Ataques e Defesas **moderadas**, enquanto **uma minoria tem valores extremamente altos ou baixos**. Isso reflete a especialização de certos Pokémon para ofensiva ou defensiva.

## 5.2.2. BoxPlots e Outliers

Podemos identificar **outliers** (valores extremos), como **Pokémon com atributos muito altos ou baixos** (por exemplo, **Shuckle** com Defesa absurdamente alta, ou **Deoxys (Velocidade)** com Velocidade muito alta). Esses outliers são essenciais na formulação de estratégias de batalha, pois podem oferecer vantagens ou desvantagens significativas.

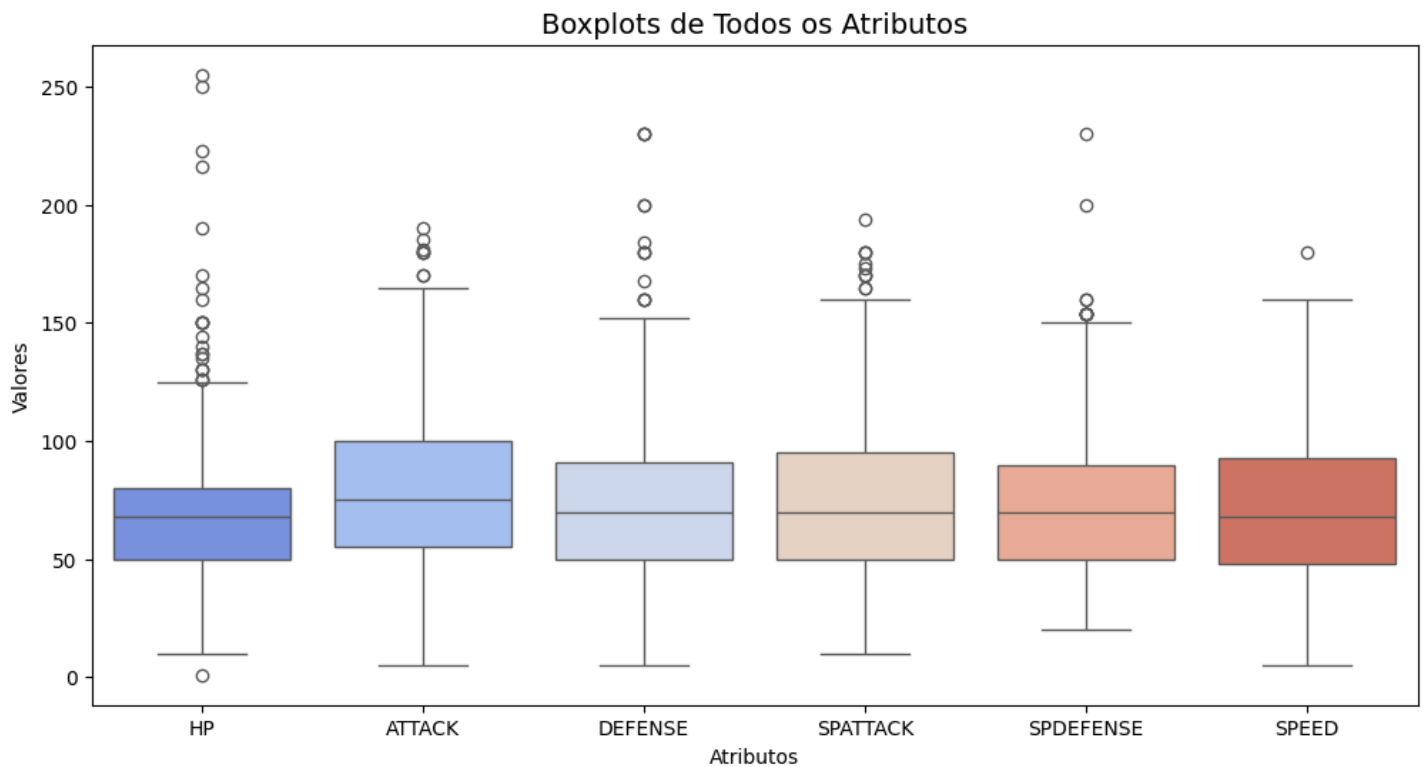


O Boxplot do Stat HP reforça como a grande maioria dos pokémon estão concentrados em torno de uma média baixa de HP, **mesmo com vários Pokémon extremamente acima da média.**

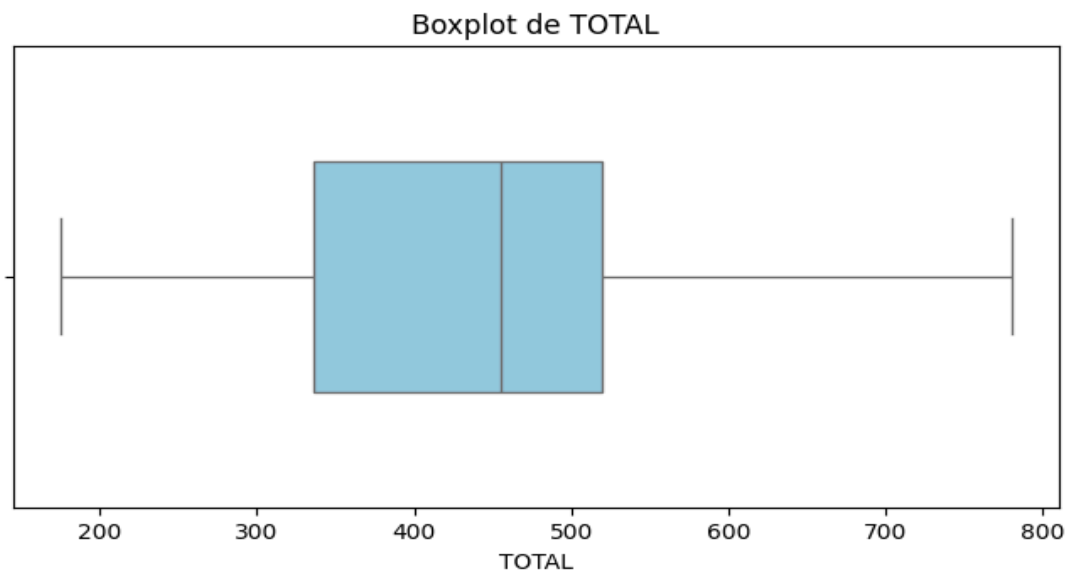


Boxplot de Velocidade e seu único outlier.

Também podemos comparar os Boxplots lado a lado e entender como estão distribuídos:



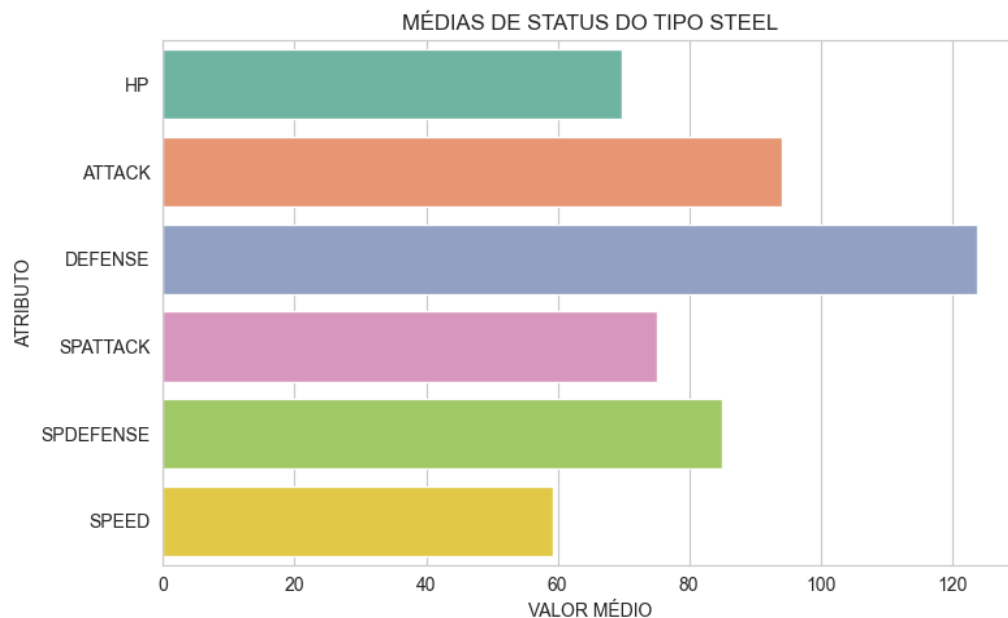
O valor dos atributos também costumam em geral estar em torno das mesmas duas centenas.



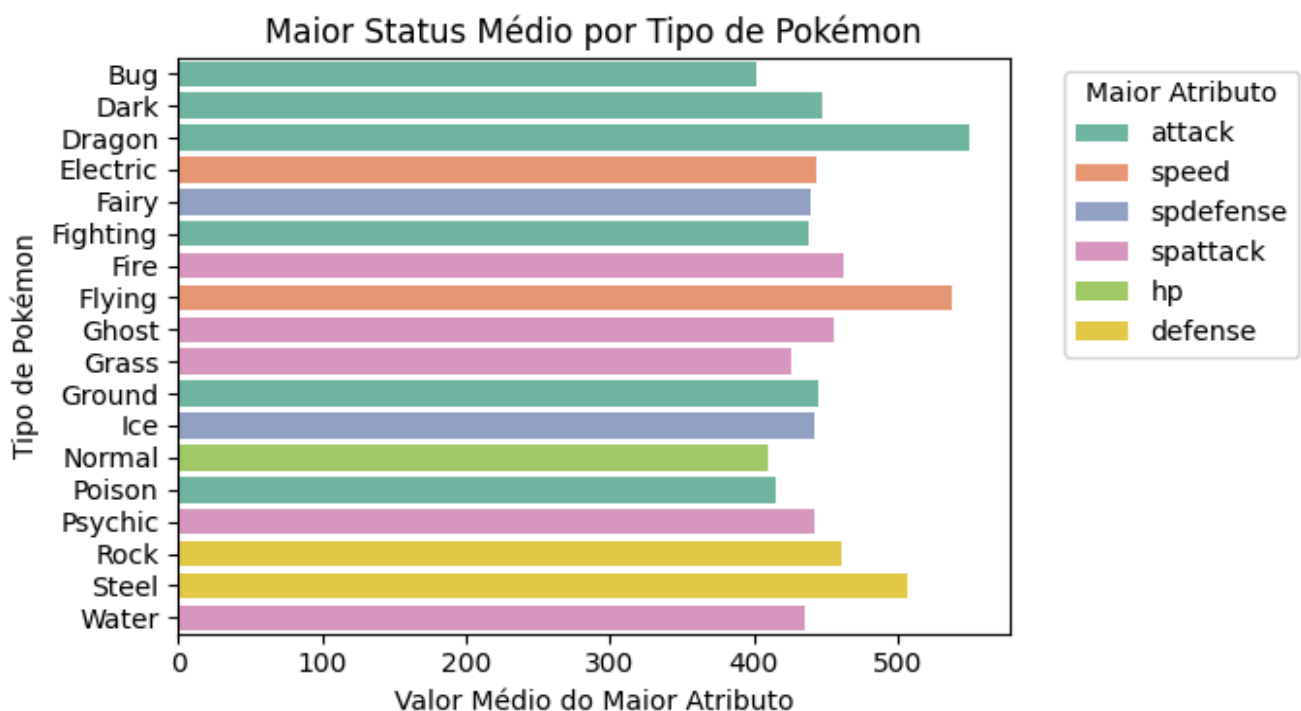
O Boxplot do Total de atributos, não possui **nenhum outlier**. Isso mostra como o **balanceamento funciona** entre os status **Totais** dos Pokémon.

### 5.2.3. Tendências e Padrões por Tipo dos Pokémon:

É importante observar padrões entre os tipos para montar um bom time. Por exemplo, **Pokémon de tipo Fogo** frequentemente têm um **alto Ataque e Ataque Especial**, enquanto os **tipo Steel** possuem **altíssima Defesa**.

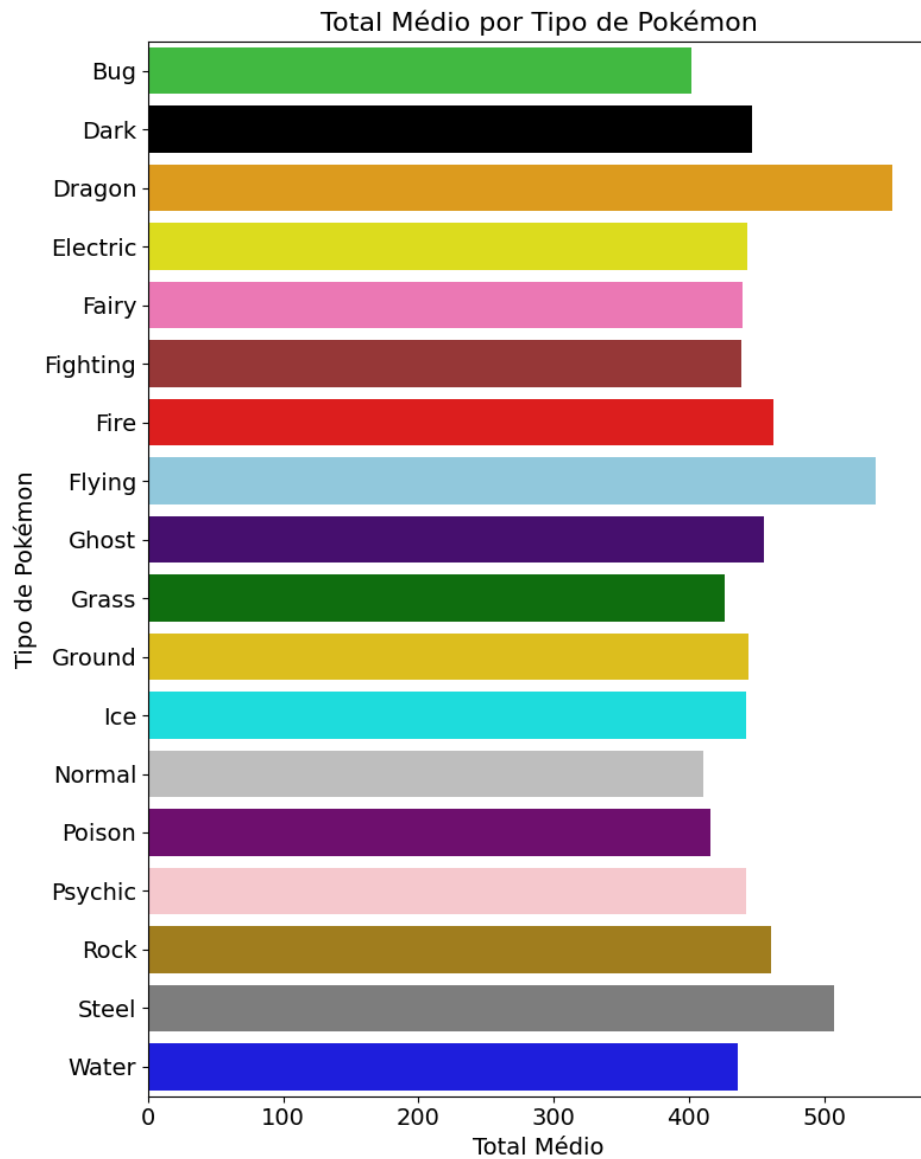


Podemos também, para facilitar essa análise, grafar qual maior atributo de cada tipo:



Podemos também analisar quais tipos possuem a maior média dos atributos totais, os que seriam hipoteticamente mais fortes - caso não existissem **Forças e Fraquezas** entre tipos

~.

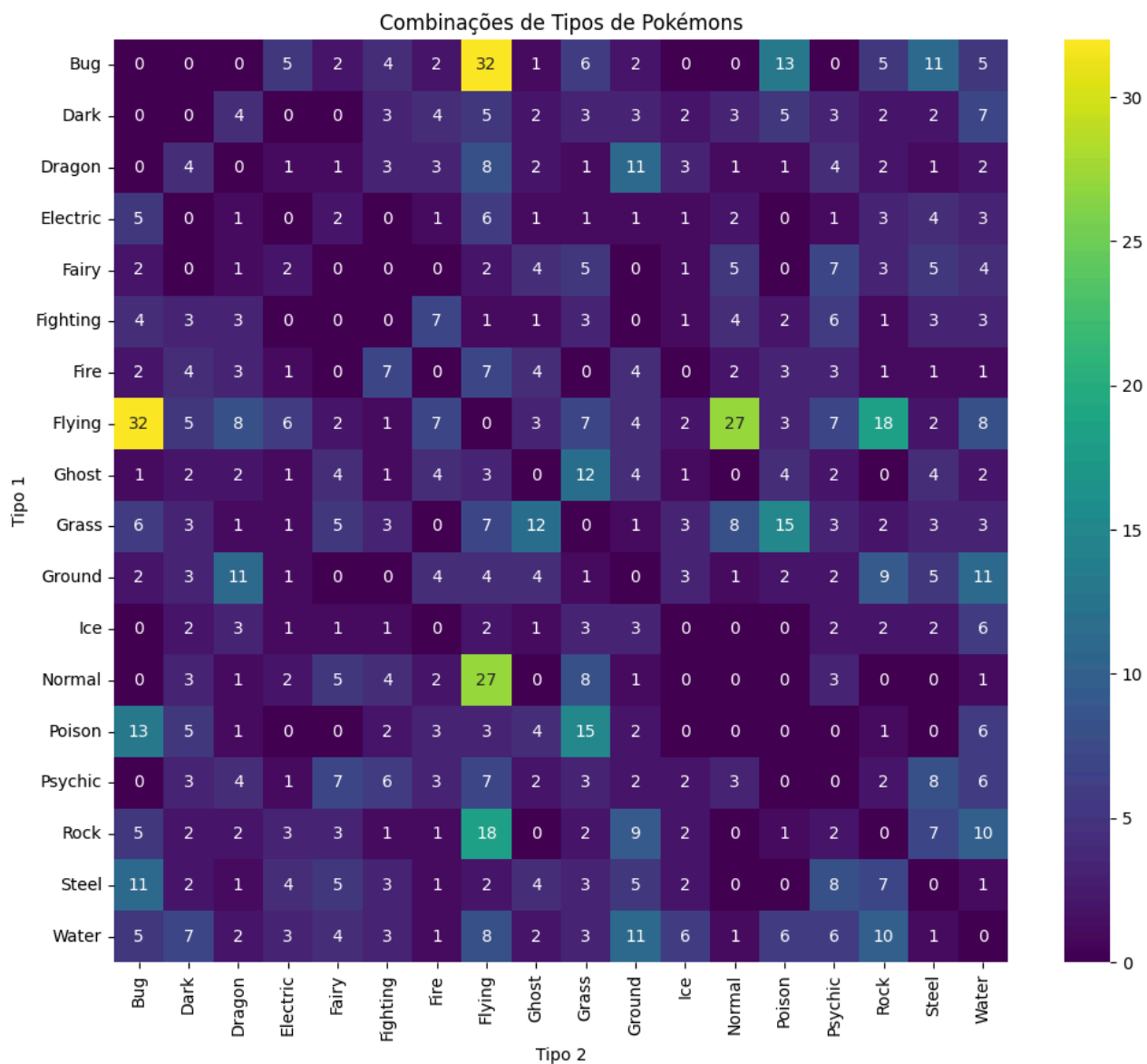




## 5.3. Combinação de Tipos dos Pokémon

### 5.3.1. Quantidade de Pokémon de cada combinação:

A quantidade de Pokémon em cada combinação de tipos pode revelar **padrões e tendências** na distribuição das espécies. Algumas combinações são **mais comuns**, enquanto outras são **raríssimas ou até inexistentes**.



Algumas combinações de tipos são muito comuns, provavelmente porque são **balanceadas** ou refletem características populares de design nos jogos.

**Exemplos de combinações populares:**

- **Inseto/Voador** (Beedrill, Butterfree, Beautifly)
- **Água/Psíquico** (Starmie, Slowbro, Golduck)

### 5.3.2. Combinações Raras ou Exclusivas:

Algumas combinações são extremamente raras e podem aparecer em **apenas um Pokémon** ou em muito poucos. Isso pode ser porque são difíceis de equilibrar ou porque os jogos ainda não exploraram tanto essa variação.

#### Exemplos:

- **Fantasma/Lutador** → Marshadow (único Pokémon com essa combinação)
- **Fada/Fantasma** → Nenhum (da Gen 1 a Gen 7)

Se uma combinação é rara, isso pode significar que ela tem um **potencial estratégico único**, já que seus pontos fortes e fracos são menos explorados nos jogos.

### 5.3.3. Relação com Força e Competitividade:

Algumas combinações de tipos aparecem mais em **Pokémon fortes ou lendários**, enquanto outras são mais comuns em Pokémon iniciais ou intermediários.

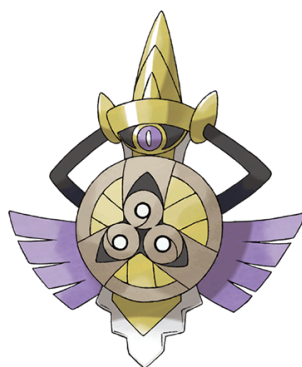
#### Exemplos:

- O tipo **Dragão** raramente aparece sozinho ou combinado com **Aço (Dialga)** ou **Voador (Salamence, Dragonite)**. Normalmente é reservado para Pokémon poderosos.
- **Tipos defensivos**, como **Aço**, tendem a aparecer como secundários para dar resistência extra (exemplo: Metagross é **Psíquico/Aço**).

### 5.3.4. Implicações Estratégicas

Algumas combinações oferecem **excelente cobertura** contra fraquezas.

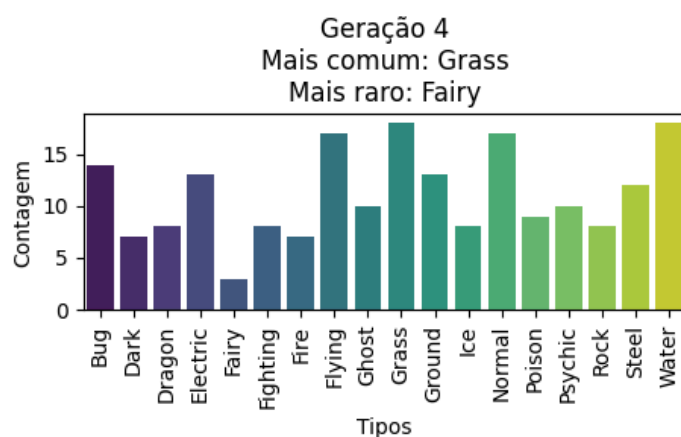
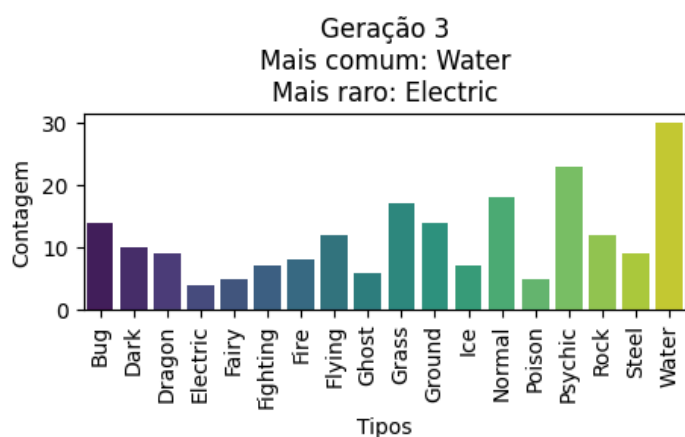
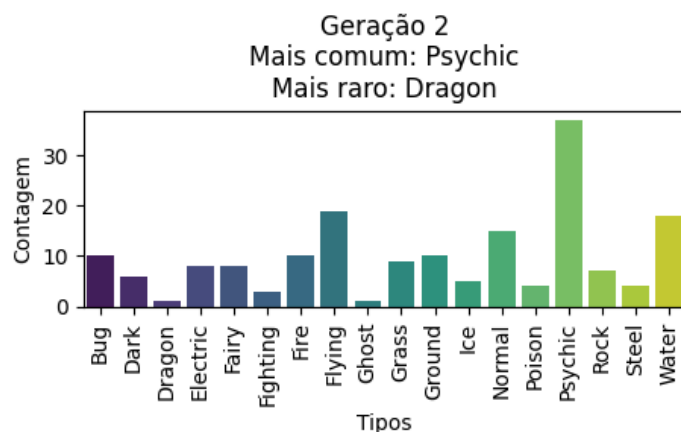
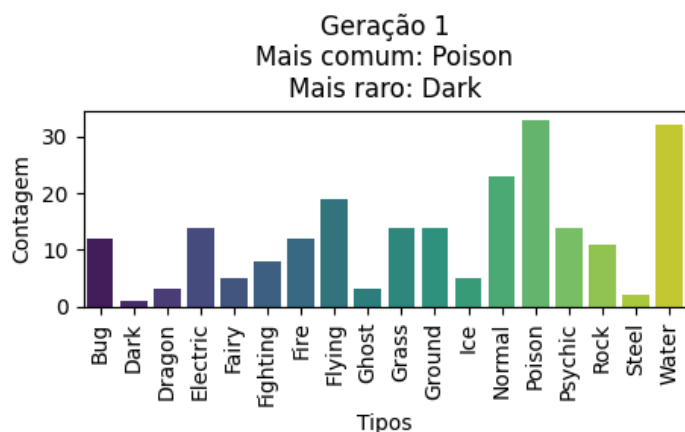
**Exemplo:** Aegislash, pokémon do tipo Metal/Fantasma, que apresenta 9 resistências (Voador, Pedra, Psíquico, Aço, Grama, Gelo, Dragão, Fada, Bug) e imunidades (Normal, Lutador, Veneno).



## 5.4. Análise Geracional dos Pokémon

Cada nova geração introduz uma quantidade variável de Pokémon, de diferentes combinações de tipos e atributos.

A **Geração 1** começou com **151 Pokémon** - a **Geração 9** já ultrapassa **1.000 Pokémon** -.



### 5.4.1. Tipos Mais Frequentes por Geração:

Algumas gerações introduzem **mais Pokémon de certos tipos** do que outras.

A **Geração 1** tinha muitos **Normais, Venenosos e Psíquicos**. Já a **Geração 3** trouxe mais Pokémon **Água**, já que a temática era oceânica (Hoenn), e a **Geração 6** introduziu um novo tipo: **Fada**, então novos Pokémon foram adicionados para equilibrar o jogo.

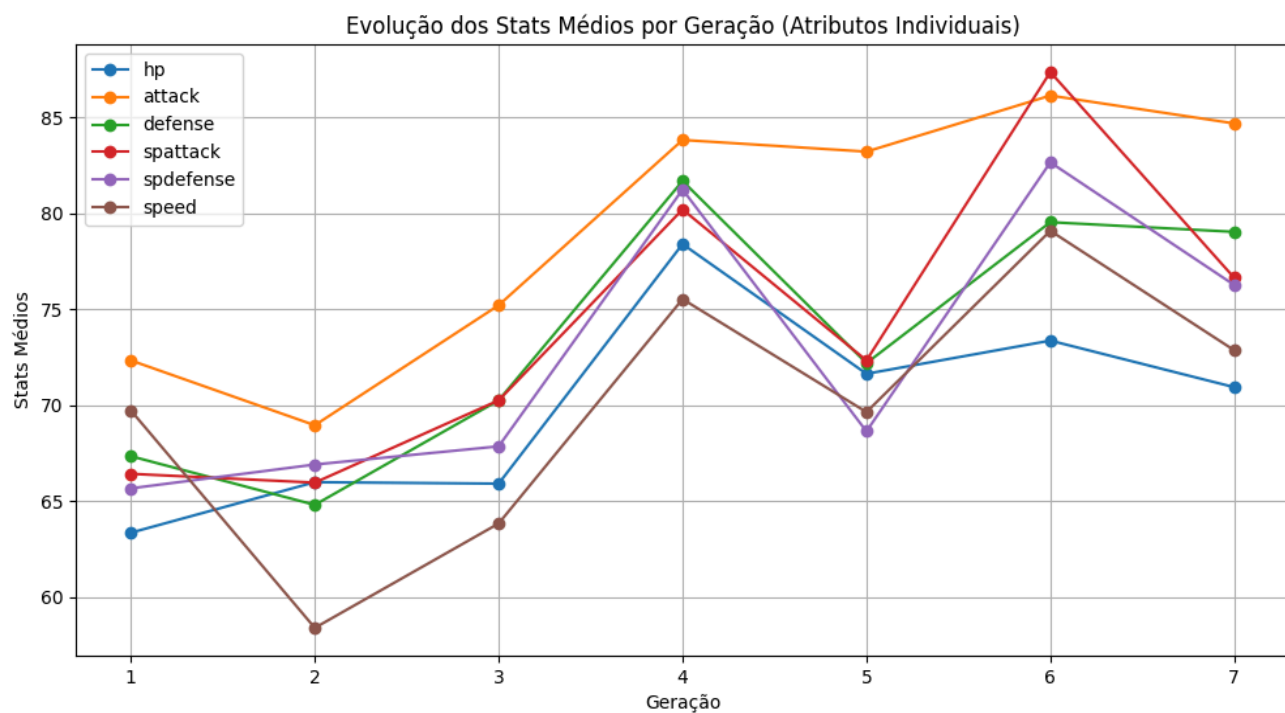
Algumas gerações inovaram ao introduzir **combinações de tipos inéditas**, exemplos:

- **Geração 4:** Dragão/Fantasma (Giratina)
- **Geração 6:** Fantasma/Planta (Trevenant)

## 5.5. Evolução dos Atributos ao Longo das Gerações

Os stats dos Pokémon variam **de acordo com a geração**, pois os desenvolvedores os ajustam a fim de manter o equilíbrio das batalhas.

### 5.5.1. Stat Médios por Geração:



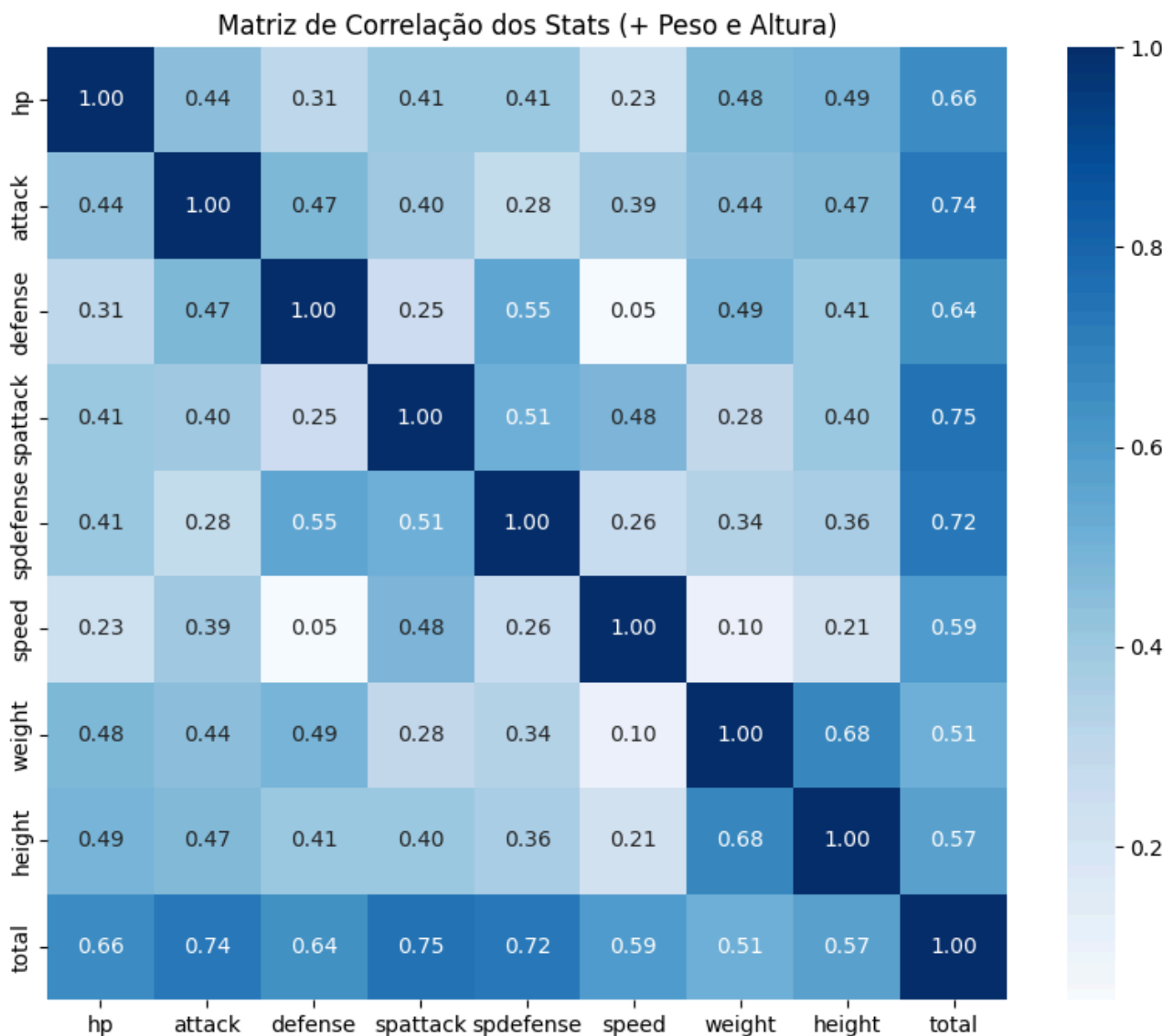
## 5.6. Sobre as Gerações

- **Geração 1:** Stats básicos, com alguns picos em Pokémon Lendários e pseudo-lendários (Dragonite).
- **Geração 2:** A Velocidade e o Total caíram em relação à primeira. Foi uma tentativa de mudança da jogabilidade, focado mais na sobrevivência.
- **Geração 4:** Aumento significativo de todos os Stats, aumentou o padrão. Deu um resultado muito positivo, é uma das mais populares.
- **Geração 5:** Pokémon continuaram muito ofensivos, mas decaiu em todo o resto por questões de balanceamento.
- **Geração 6 e 7:** A adição das Mega Evoluções e outras novas mecânicas subiu o Total novamente, principalmente alguns outliers absurdos.

## 6. Correlação dos Atributos

A correlação dos atributos dos Pokémon ajuda a entender **quais stats estão relacionados** e como **diferentes arquétipos de Pokémon** são formados. Isso pode revelar padrões como:

- Pokémon mais rápidos tendem a ter menos defesa?
- Pokémon mais pesados tendem a ter mais HP?



## 6.1. Insights

### 6.1.1. Correlação entre os Status Ofensivos e Defensivos

- A. Ataque e Defesa possuem uma correlação **moderada** (0.47), indicando que Pokémon com alto Ataque físico podem ter uma tendência a possuir uma Defesa elevada, embora isso não seja uma regra absoluta.
- B. Ataque Especial e Defesa Especial também apresentam uma correlação **moderada** (0.51), sugerindo que Pokémon que se destacam no ataque especial costumam ter uma Defesa Especial razoável.

Isso reforça a ideia de que os Pokémon tendem a se especializar em dois grupos principais: físicos (Ataque e Defesa) e especiais (Ataque Especial e Defesa Especial).

### 6.1.2. Velocidade e Outros Atributos

A velocidade tem **baixa correlação** com os demais atributos, com o maior valor sendo (0.48) em relação à Defesa Especial.

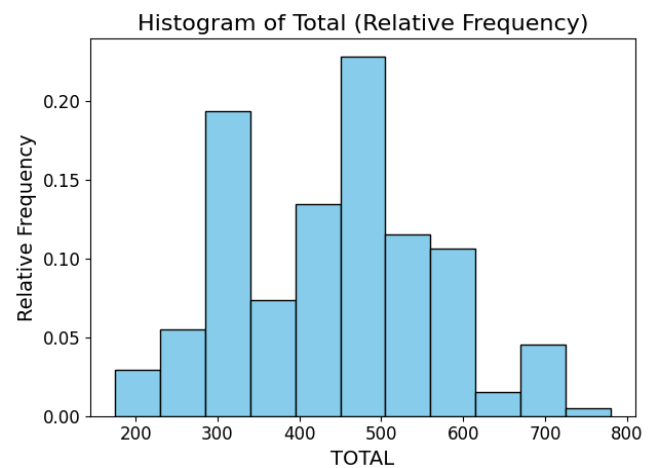
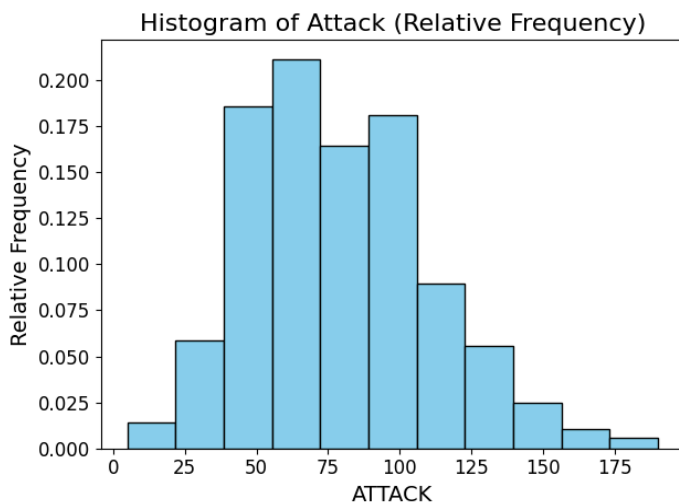
Isso indica que Pokémon rápidos podem variar bastante em seus outros atributos, **não havendo um padrão claro de que sejam mais ofensivos ou defensivos.**

### 6.1.3. Peso e Altura em Relação aos Status

- A. O Peso tem boa correlação com Defesa (0.49) e HP (0.48), **sugerindo que Pokémon mais pesados geralmente são mais resistentes.**
- B. Já a Altura tem uma correlação relevante com HP (0.49) e Defesa Especial (0.36), **sugerindo que Pokémon mais altos também tendem a ser mais resistentes.**

## 7. Normalização e Probabilidade dos atributos

Normalizando os histogramas de cada atributo obtemos uma distribuição que pode ser interpretada como a **probabilidade** de um pokémon selecionado aleatoriamente estar dentro de um intervalo de atributos.



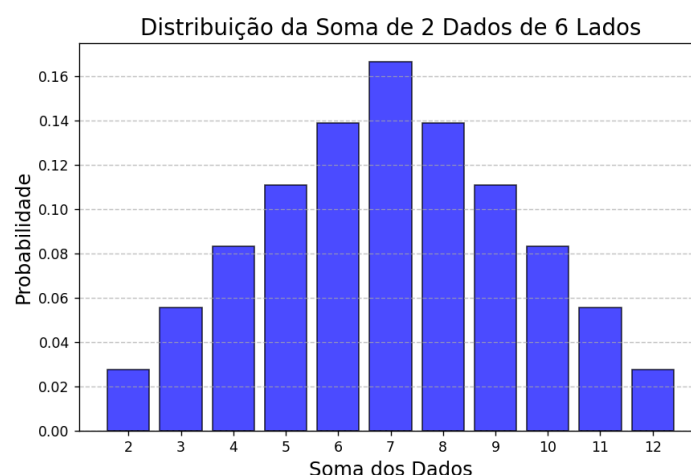
### 7.1. Convolução

A Convolução é um operador linear que recebe duas funções e resulta em uma terceira, e pode ser definida como:

$$(f * g)(k) = h(k) = \sum_{j=-\infty}^{\infty} f(j) \cdot g(k - j)$$

Na estatística ela é utilizada para calcular a **distribuição de probabilidade da soma de dois eventos aleatórios**. Por exemplo : a distribuição de probabilidades da soma de dois dados justos.

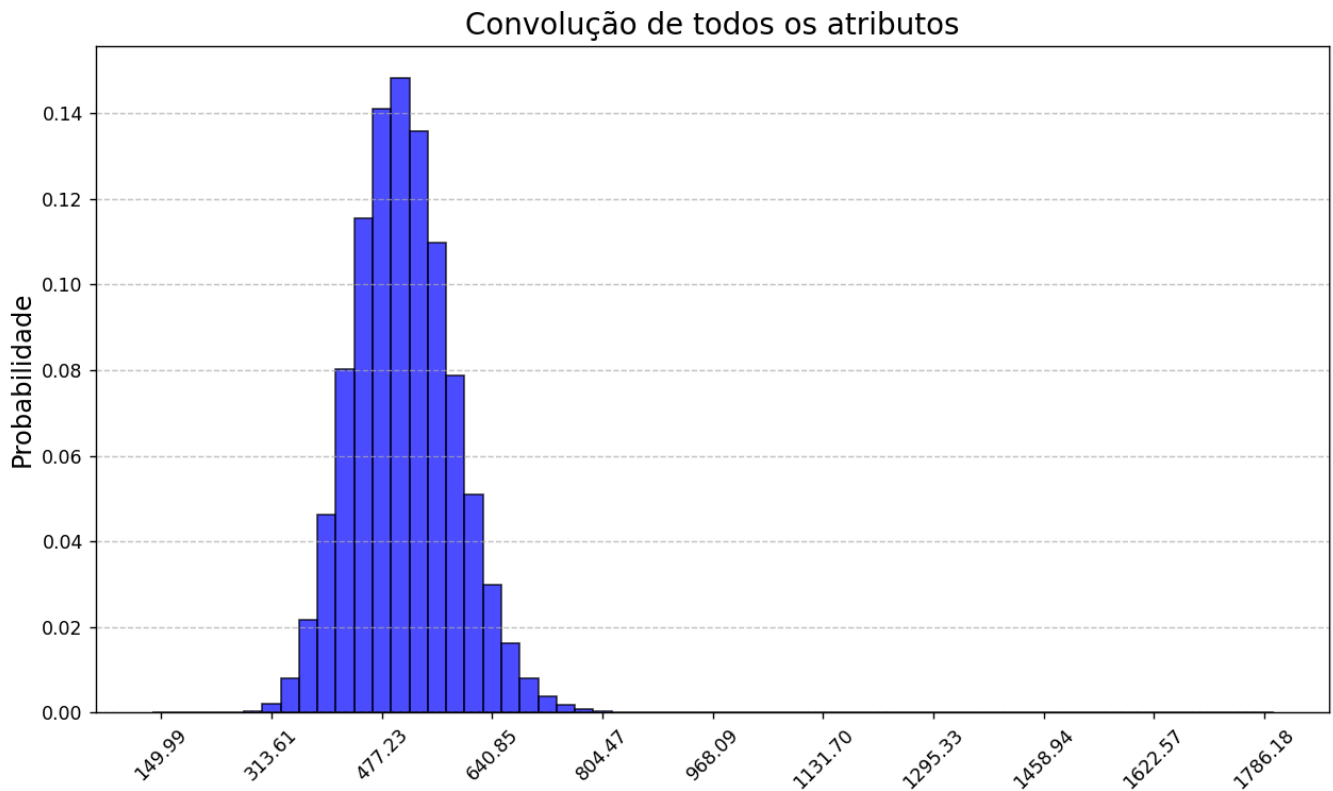
A convolução da distribuição de cada dado é:





## 7.2. Convolução dos Atributos dos Pokémon

Utilizando o conceito da **convolução** podemos analisar como as distribuições de todos os atributos (Stats) se comportam:



Ao analisar profundamente, mesmo com a tendência da convolução para uma curva normal (o valor esperado dos dados é parecido) :

**No nosso caso, a convolução não é uma representação fiel do Total verdadeiro.**

A razão disso é que o evento: “*valor estar dentro de um intervalo  $[a, b]$  de um atributo*” **não é independente**, pois os atributos tem **correlação positiva** entre si.

## 8. Conclusão

Nossa análise permitiu identificar padrões estatísticos relevantes sobre os atributos dos Pokémon, sua distribuição e suas relações ao longo das gerações. O refinamento dos dados possibilitaram uma compreensão mais detalhada dessas informações, abrindo caminho para estudos ainda mais aprofundados.

A análise estatística dos dados presentes em nosso repositório oferece diversas oportunidades para aprimorar estratégias competitivas no universo Pokémon. A seguir, exploramos aplicações futuras desses dados:

### 8.1. Expansão da análise

Além do ambiente de batalhas, pode-se utilizar este banco de dados para uma série de análises, desde aplicações do uso de atributos do tipo de Pokémon em ambiente competitivo até tendências de desenvolvimento dos Pokémons ao decorrer das gerações.

Como já mostrado há uma gama de possíveis novas análises ainda a serem feitas com o banco de dados, o que abre margem para **diversas novas correlações** e tendências a serem analisadas.

Essas descobertas contribuem para um entendimento mais amplo do universo Pokémon, permitindo explorar sua evolução, balanceamento e padrões ocultos ao longo das gerações.

## 9. Análise Futura - ¡PokéOtimizador!

Com tais dados coletados podemos auxiliar na criação de **modelos de otimização de montagem de equipes**, aprimorando a montagem de equipes **equilibradas**, visando estratégias defensivas e ofensivas.

Algoritmos podem ser desenvolvidos, para ajudar na formação de equipes ideais levando em consideração os atributos Pokémons, ressaltando os pontos fortes e fraquezas de cada tipo, o que auxiliaria o usuário na decisão de escolha de seu time estatisticamente.

Um exemplo de otimização para montar uma equipe de 6 Pokémon é:

### Minimizar a quantidade de fraquezas por tipo do time

A seguir foi modelado um problema de *programação linear* visando minimizar o número de fraquezas totais do time **considerando apenas Pokémon com um único tipo e não é possível repetir os tipos**:

$$\min Z = \sum_{i=0}^{17} \sum_{j=0}^{17} X_i \cdot F_{ji} \quad \begin{cases} \sum_{i=0}^{17} X_i = 6 \\ X_i \in [0, 1] \end{cases}$$

- a variável  $X_i$  é uma escolha binária (usar ou não usar) o Tipo  $i$ ,
- A matriz  $F(i, j)$  representa o tipo  $i$  ter vantagem sobre o tipo  $j$ .

### 9.1. Solução Trivial:

A **solução** para esse problema pode ser encontrada utilizando o algoritmo SIMPLEX.

O resultado é um time de 6 Pokémon com os tipos :

[Elétrico, Fada, Fantasma, Normal, Veneno, Água] com um total de **10 fraquezas**.

### 9.2. Possíveis melhorias:

Essa análise poderia ser melhorada levando em conta **tipos duplos** e a **distribuição de tipos** de pokémon, **dando mais peso para fraquezas a tipos que aparecem mais**, é possível também otimizar um time para uma geração específica, utilizando a distribuição de tipos por geração para a análise.