

# The Dataset

**Sourav SEN GUPTA**  
Lecturer, SCSE, NTU







## Data Science The Pokemon Dataset

### Structured Data : Mixed

Highly Organized Data  
Clearly Defined Variables  
Easy to Mine and Analyze  
Numeric and Categorical

Source : Kaggle Datasets  
**Pokemon with stats** by Alberto Barradas  
<https://www.kaggle.com/abcsds/pokemon>



# Data Science

## The Pokemon Dataset

The screenshot shows the Kaggle website interface for a dataset titled "Pokemon with stats". The dataset was created by Alberto Barradas and updated 2 years ago (Version 2). It contains 721 entries and includes stats and types. The page features a large background image of a dark brown surface with a red circle and a black leaf. A sidebar on the right shows "656 voters" and a "share" button. The main navigation bar includes links for Data, Overview, Kernels, Discussion, Activity, Download (15 KB), and New Kernel. Below the navigation, there's a "Data (15 KB)" section with an API download link and a "Download All" button. The "Data Sources" section lists "Pokemon.csv" as an 800 x 13 file. The "About this file" section describes it as a "Main Database With Generation and Legendary flag." The "Columns" section lists two columns: "# # PokeDex index number" and "Name Name of the Pokemon".

Source : Kaggle Datasets | **Pokemon with stats** by Alberto Barradas | <https://www.kaggle.com/abcsds/pokemon>



# Data Science

## The Pokemon Dataset

#	Name	Type 1	Type 2	Total	HP	Attack	Defense	Sp. Atk	Sp. Def	Speed	Generation	Legendary
430	Honchkrow	Dark	Flying	505	100	125	52	105	52	71	4	False
338	Solrock	Rock	Psychic	440	70	95	85	55	65	70	3	False
32	Nidoran♂	Poison	NaN	273	46	57	40	40	40	50	1	False
442	Spiritomb	Ghost	Dark	485	50	92	108	92	108	35	4	False
480	Uxie	Psychic	NaN	580	75	75	130	75	130	95	4	True
536	Palpitoad	Water	Ground	384	75	65	55	65	55	69	5	False
360	Wynaut	Psychic	NaN	260	95	23	48	23	48	23	3	False
478	Froslass	Ice	Ghost	480	70	80	70	80	70	110	4	False
76	Golem	Rock	Ground	495	80	120	130	55	65	45	1	False
177	Natu	Psychic	Flying	320	40	50	45	70	45	70	2	False

Source : Kaggle Datasets | [Pokemon with stats](#) by Alberto Barradas | <https://www.kaggle.com/abcsds/pokemon>



# Data Science

## The Pokemon Dataset

#	ID for each Pokemon (runs from 1 to 721)
Name	Name of each Pokemon
Type 1	Each Pokemon has a basic Type, this determines weakness/resistance to attacks
Type 2	Some Pokemons are dual type and have a Type 2 value (set to NaN otherwise)
Total	Sum of all stats of a Pokemon, a general guide to how strong a Pokemon is
HP	Hit Points, defines how much damage a Pokemon can withstand before fainting
Attack	The base modifier for normal attacks by the Pokemon (e.g., scratch, punch etc.)
Defense	The base damage resistance of the Pokemon against normal attacks
SP Atk	Special Attack, the base modifier for special attacks (e.g. fire blast, bubble beam)
SP Def	Special Defense, the base damage resistance against special attacks
Speed	Determines which Pokemon attacks first each round
Generation	Each Pokemon belongs to a certain Generation
Legendary	Legendary Pokemons are powerful, rare, and hard to catch

Source : Kaggle Datasets | **Pokemon with stats** by Alberto Barradas | <https://www.kaggle.com/abcsds/pokemon>



# Data Science

## The Pokemon Dataset

#	Index (Numeric)
Name	String (neither Numeric nor Factor)
Type 1	Categorical (one of 18 Types)
Type 2	Categorical (one of 18 Types, or NaN)
Total	Numeric
HP	Numeric
Attack	Numeric
Defense	Numeric
SP Atk	Numeric
SP Def	Numeric
Speed	Numeric
Generation	Categorical 1 to 6
Legendary	Categorical (True or False)

#	int64
Name	object
Type 1	object
Type 2	object
Total	int64
HP	int64
Attack	int64
Defense	int64
Sp. Atk	int64
Sp. Def	int64
Speed	int64
Generation	int64
Legendary	bool

Source : Kaggle Datasets | **Pokemon with stats** by Alberto Barradas | <https://www.kaggle.com/abcsds/pokemon>



## Data Science Pipeline **Exploratory Analysis**

What are the Variables in the Data?  
How to characterize the Variables?  
How to find relation between them?

**How to intelligently  
explore acquired Data?**