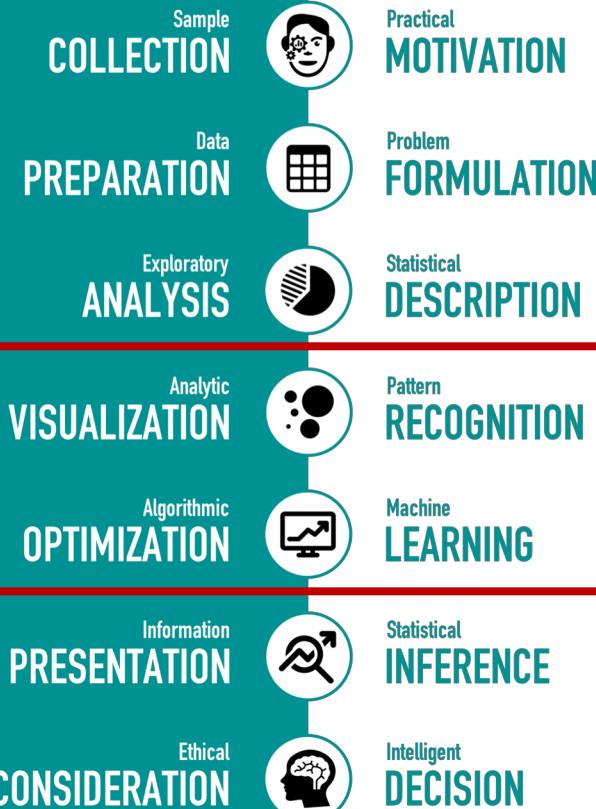


# Anomaly Detection

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





## Data Science Anomaly Detection

### Pattern Recognition

Is there a pattern in the acquired data?  
How to learn the underlying pattern?  
How to exploit the pattern in data?

### How to optimally learn from the Data?



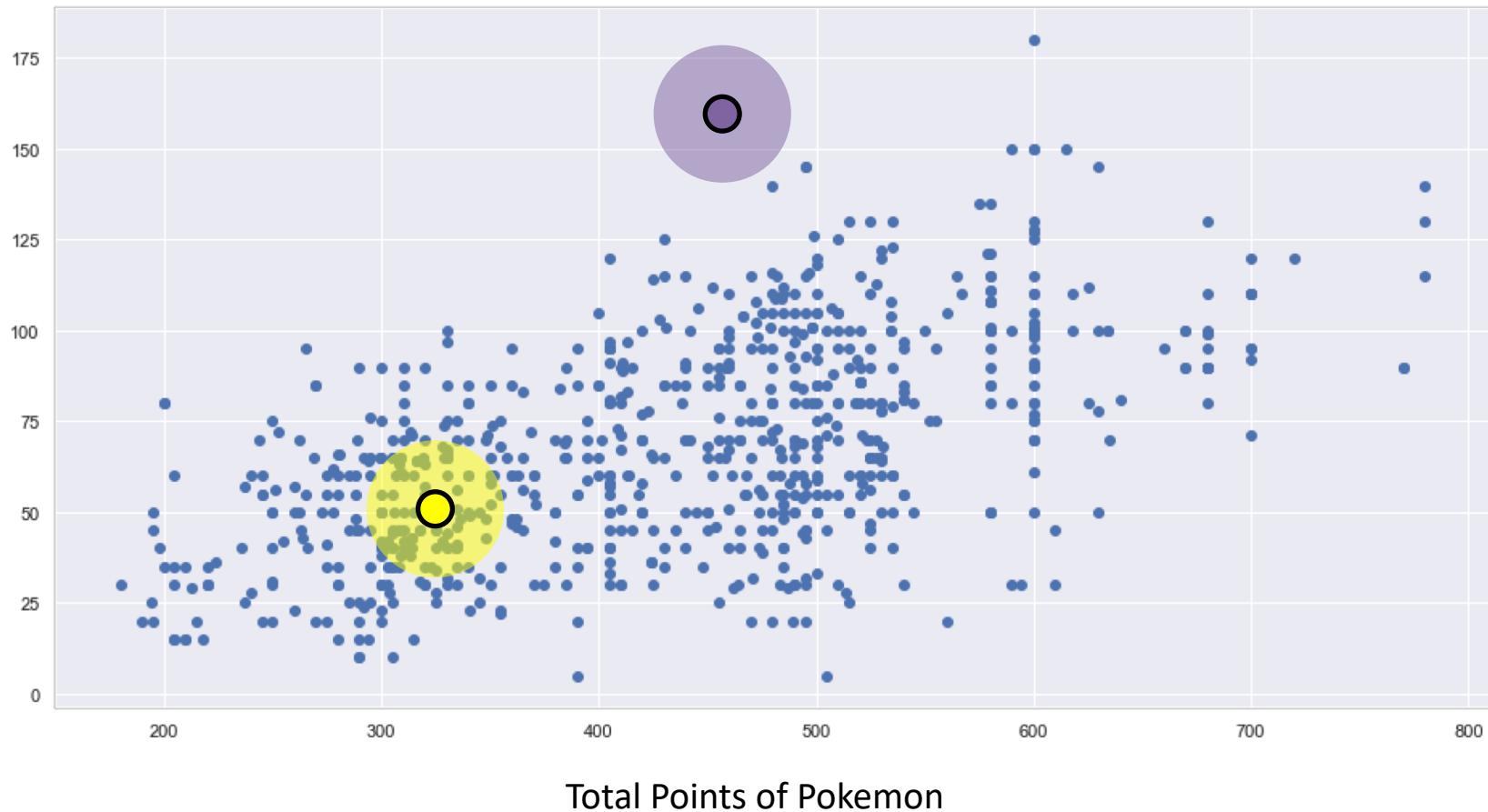
# Data Science

## 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>

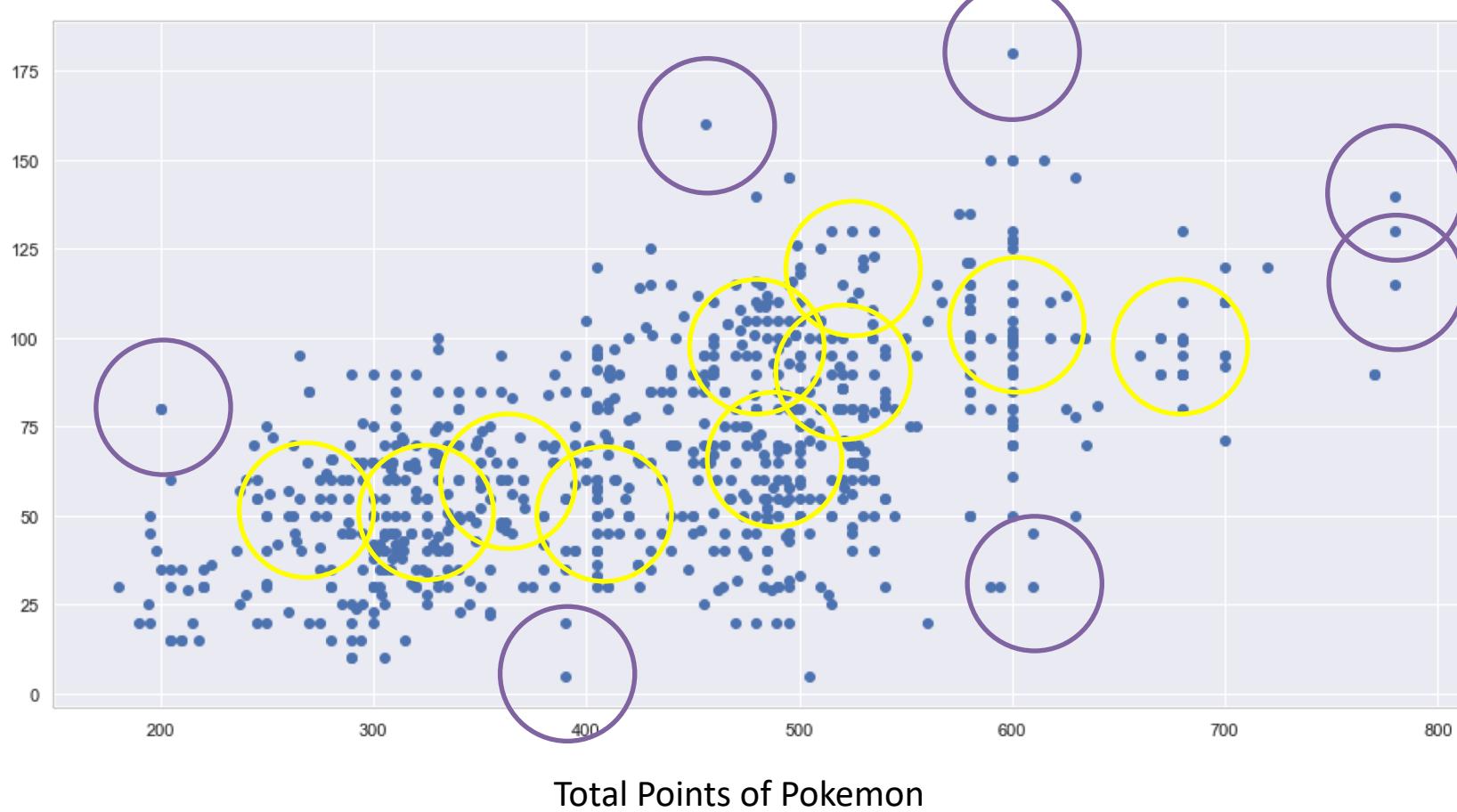
Speed of Pokemon



4



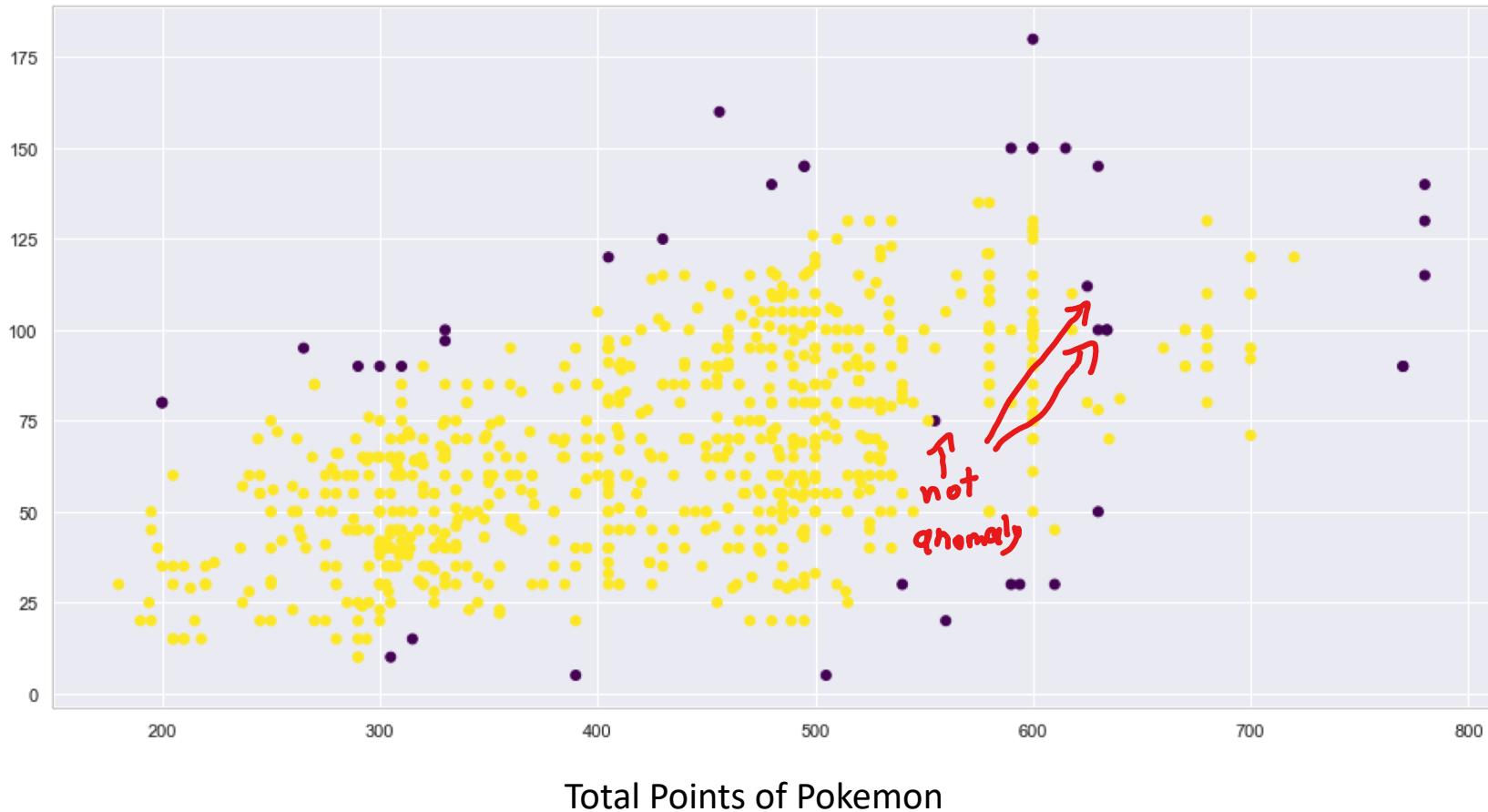
Speed of Pokemon



5

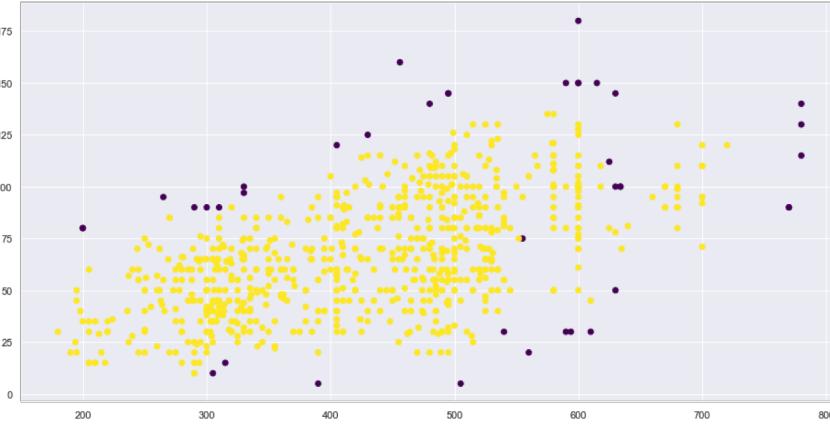


Speed of Pokemon



6





# Data Science Anomaly Detection

## Nearest Neighbor

**Total** Total Points of Pokemon  
**Speed** Speed of Pokemon

### Local Outlier Factor

Choose  $K$  – the total number of neighbors

parameter

Choose  $d$  – fraction of anomalies in data

parameter

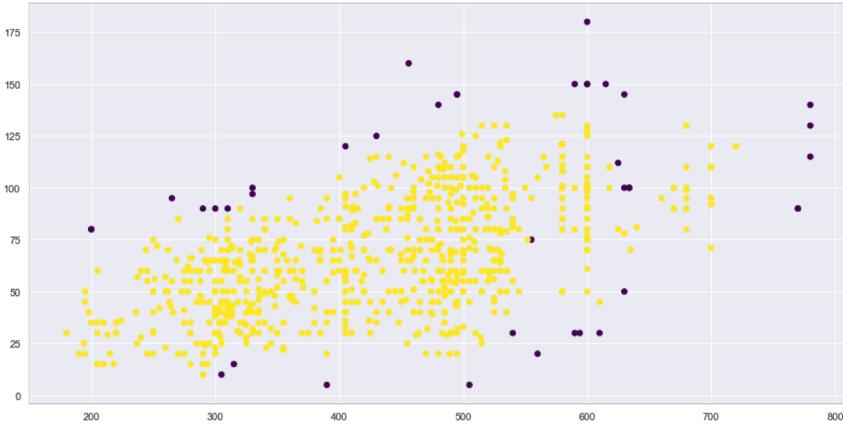
for each point in the dataset

Find the  $K$  nearest neighbors in data

Compute if the density is high enough

### Machine Learning Questions

- How many Anomalies are “visible”?
- Can we identify those Anomalies?
- What do the Anomalies signify?

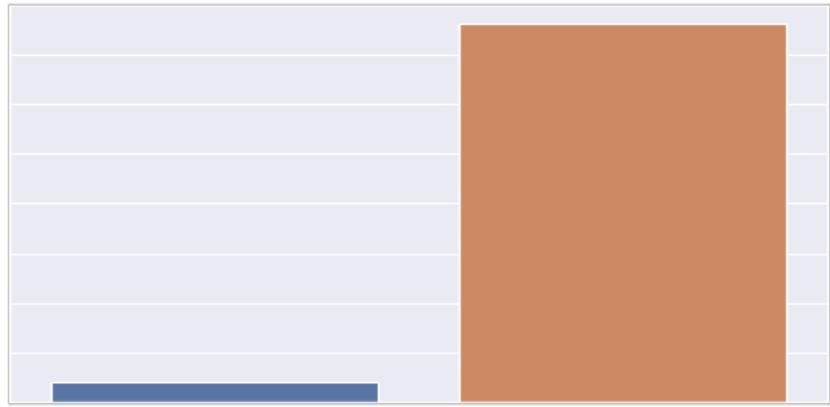
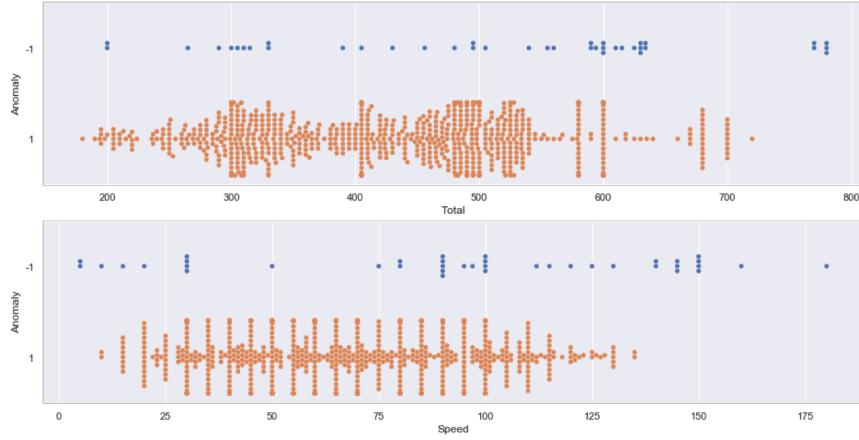


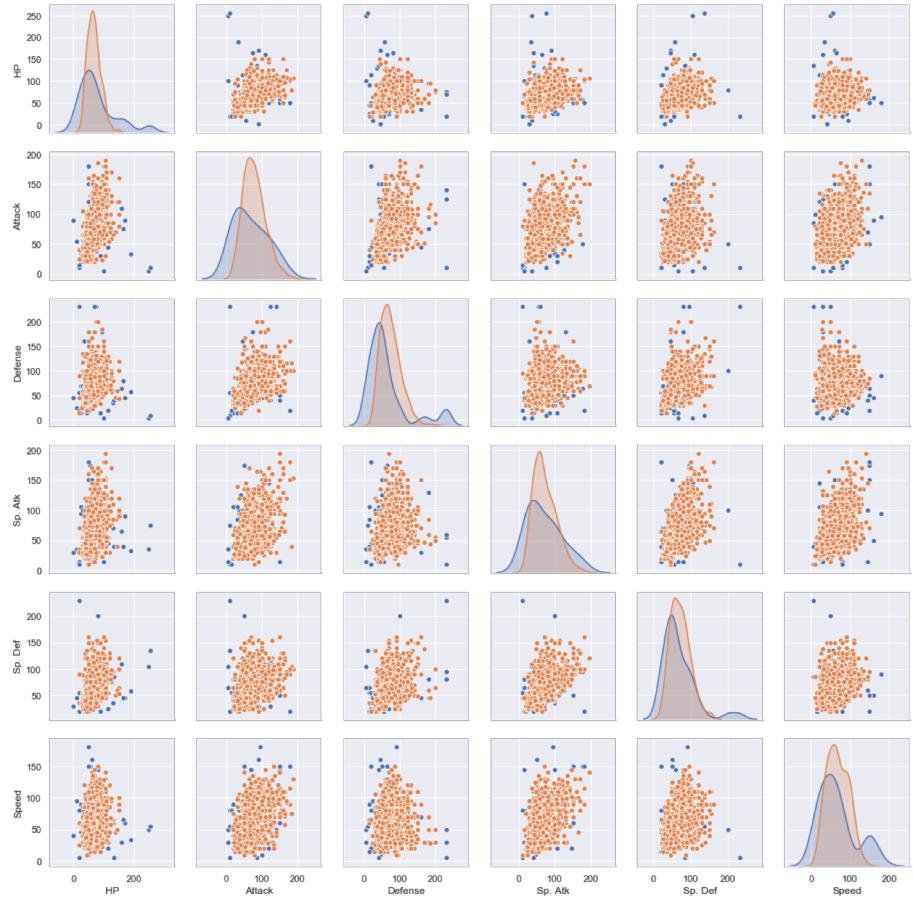
# Data Science

## Anomaly Detection

### Nearest Neighbor

**Total**      Total Points of Pokemon  
**Speed**      Speed of Pokemon

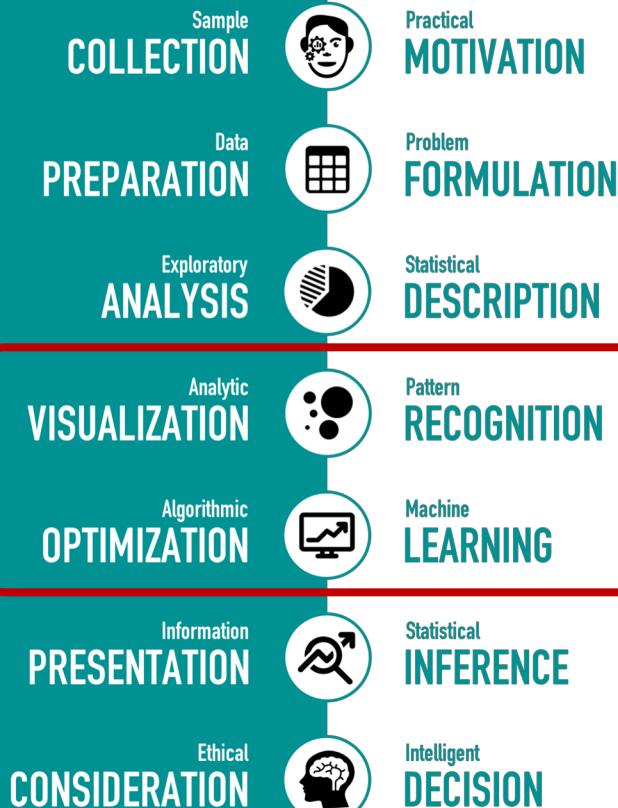




# Data Science Anomaly Detection

## Nearest Neighbor

<b>HP</b>	Hit Points
<b>Attack</b>	Attack Points
<b>Defense</b>	Defense Points
<b>Sp. Atk</b>	Special Attack
<b>Sp. Def</b>	Special Defense
<b>Speed</b>	Speed of Pokemon



## Data Science Pipeline Pattern Recognition

How to learn from the acquired Data?  
How to find pattern in acquired Data?  
How to utilize the pattern in the Data?

## How to optimally learn from the Data?