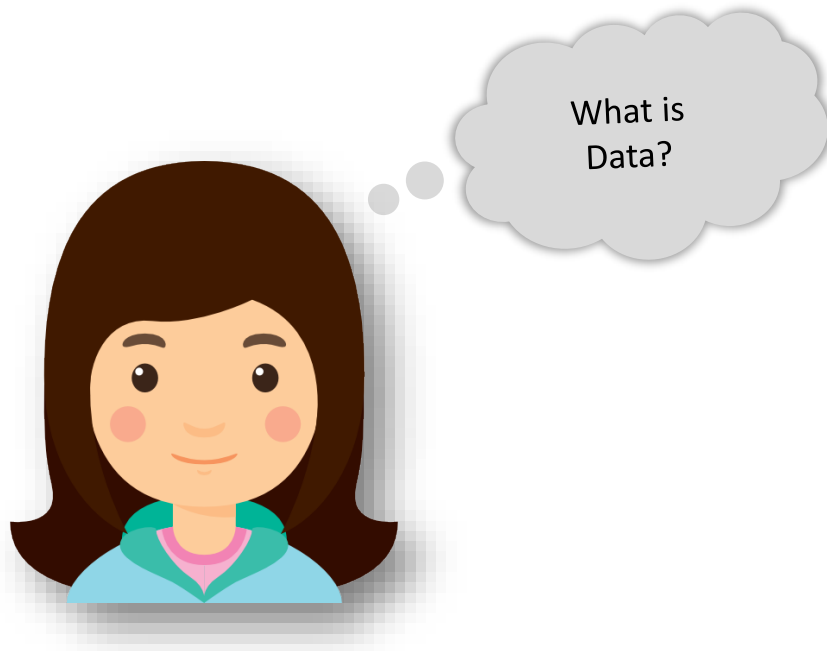


Data Science

What is Data?

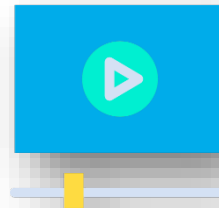


What is Data?

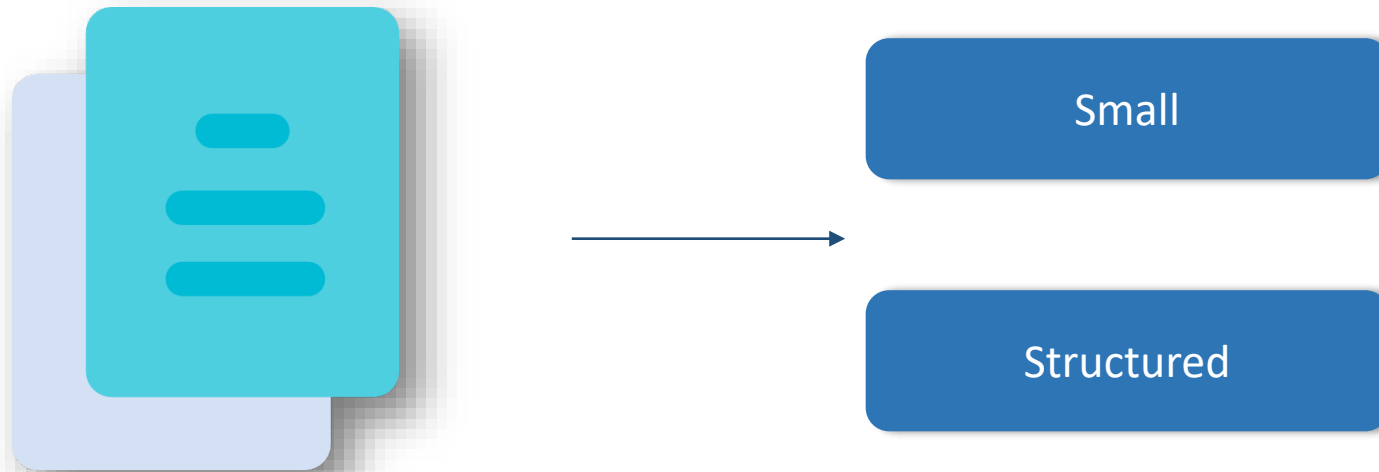
Data is just collection of facts

23, 8000

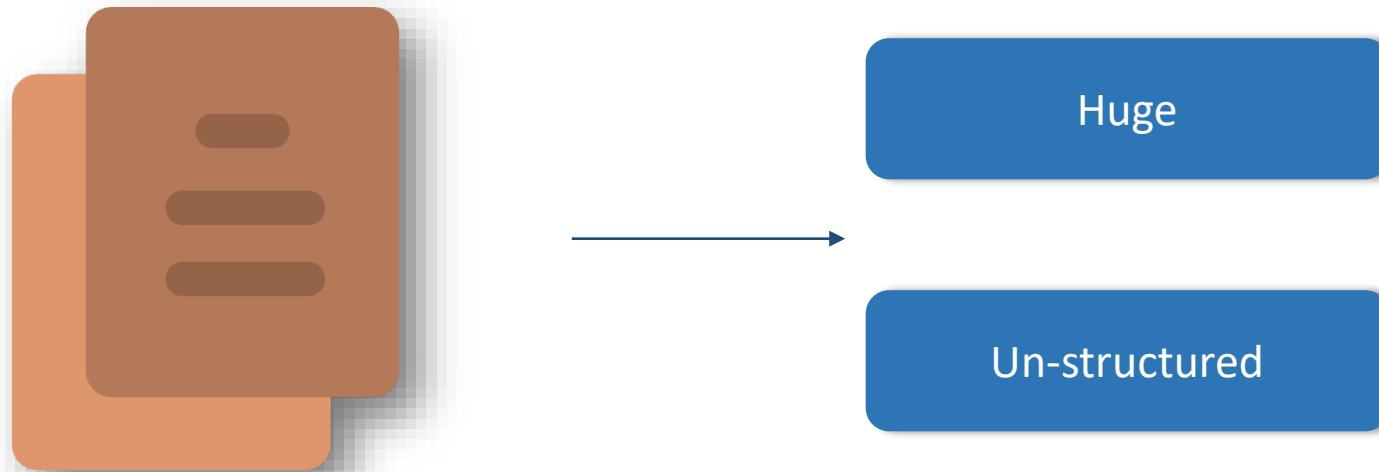
“This is
Sparta”

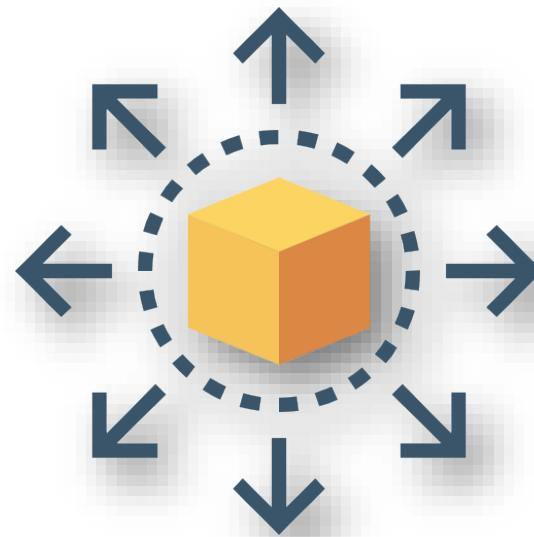


Data Back Then



Data Now





Data Mining Use Case



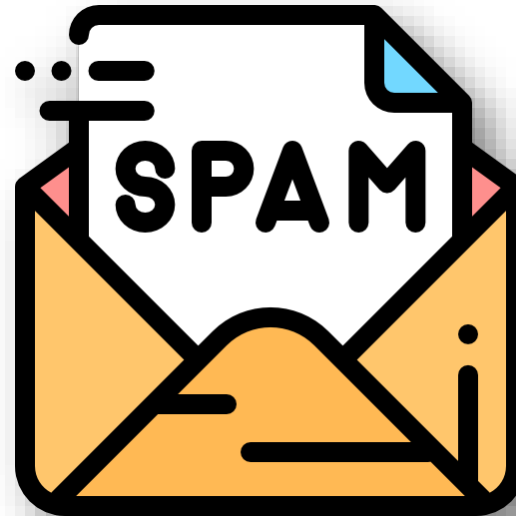
Data Mining can
be used for finding
fraudulent credit
card transactions



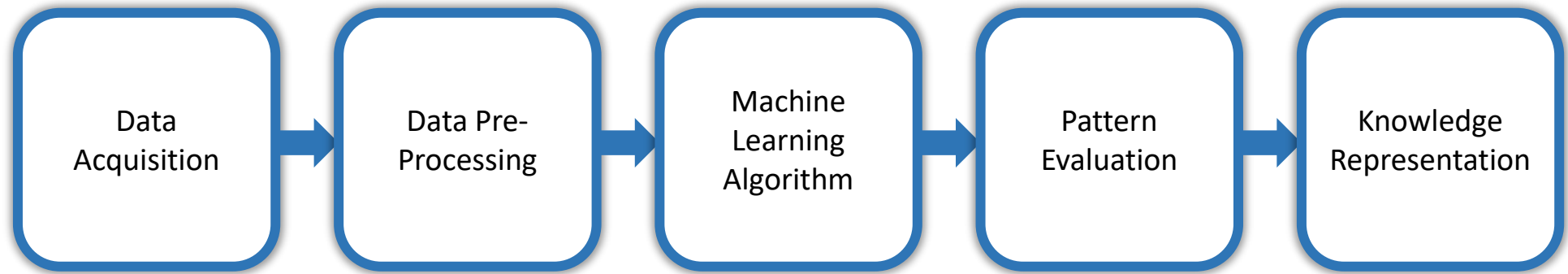
Data Mining Use Case



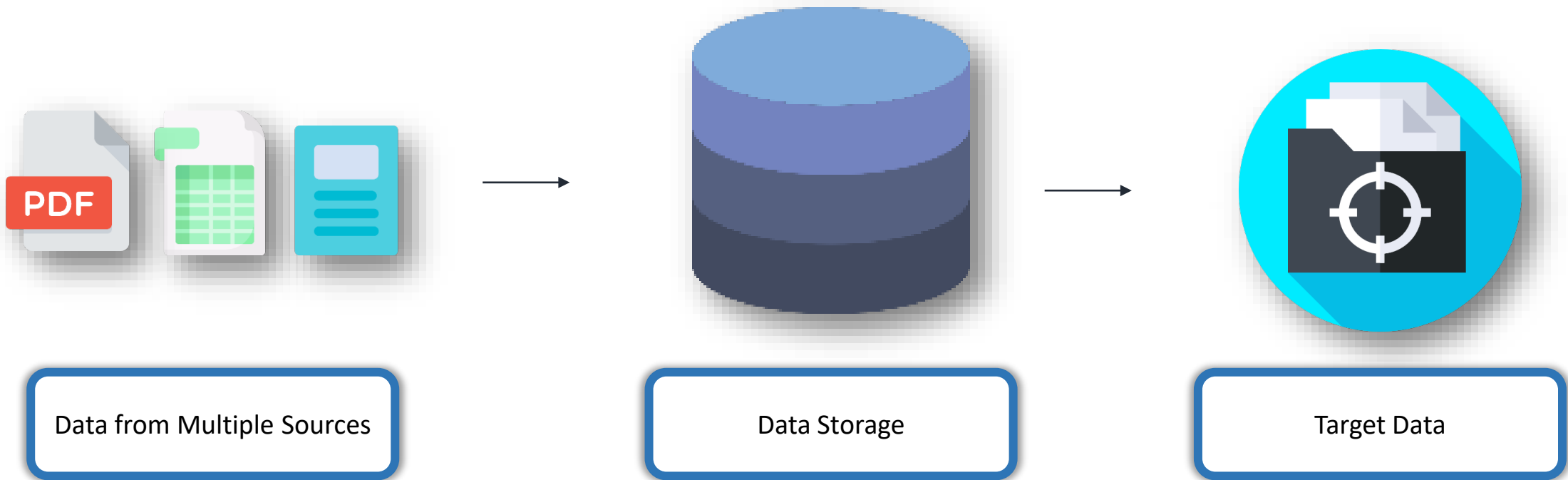
Data Mining can
be used for
tagging a mail as
spam or genuine



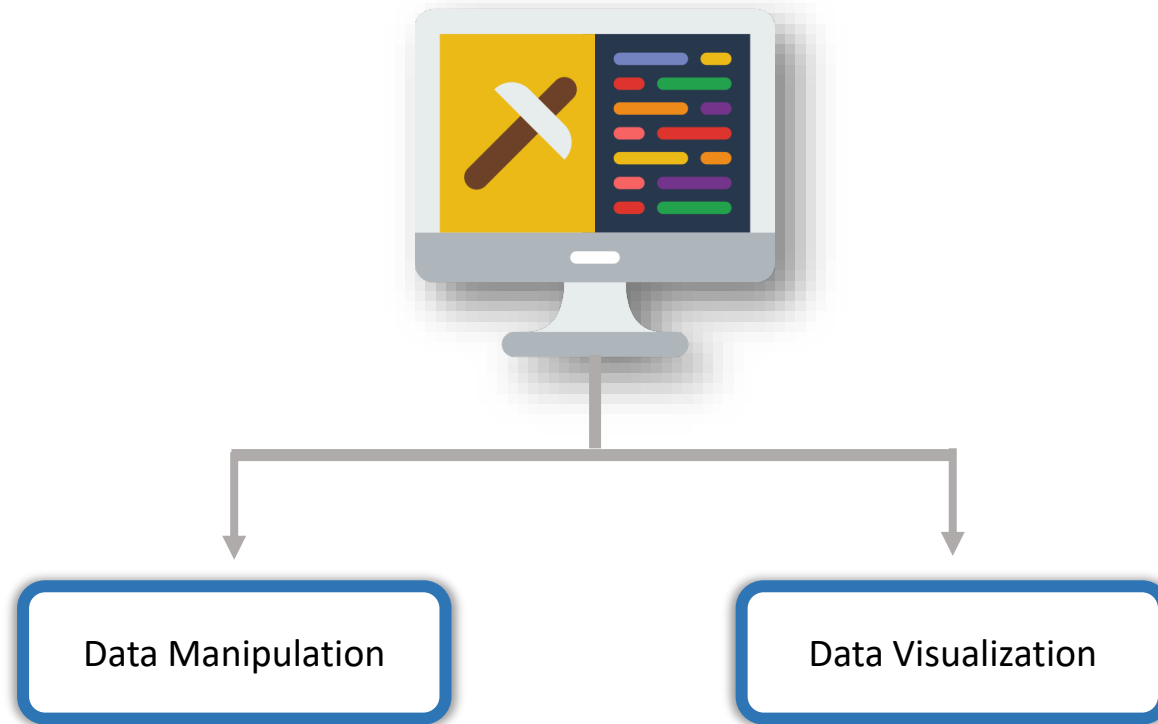
Data Mining Life Cycle



Data Acquisition

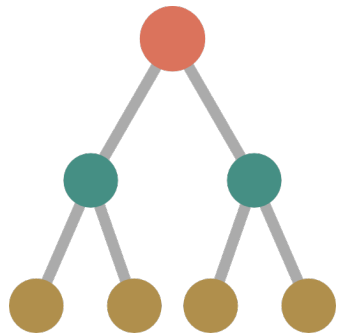


Data Pre-processing

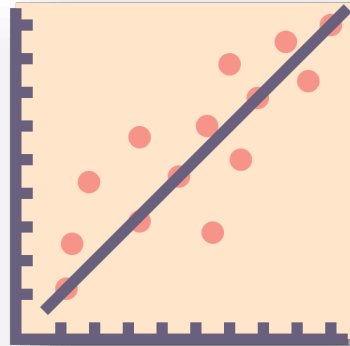


Machine Learning

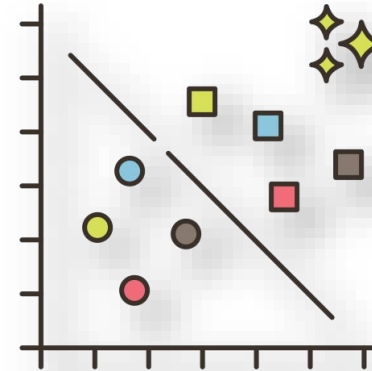
Application of intelligent algorithms to find meaningful information from the data



Classification



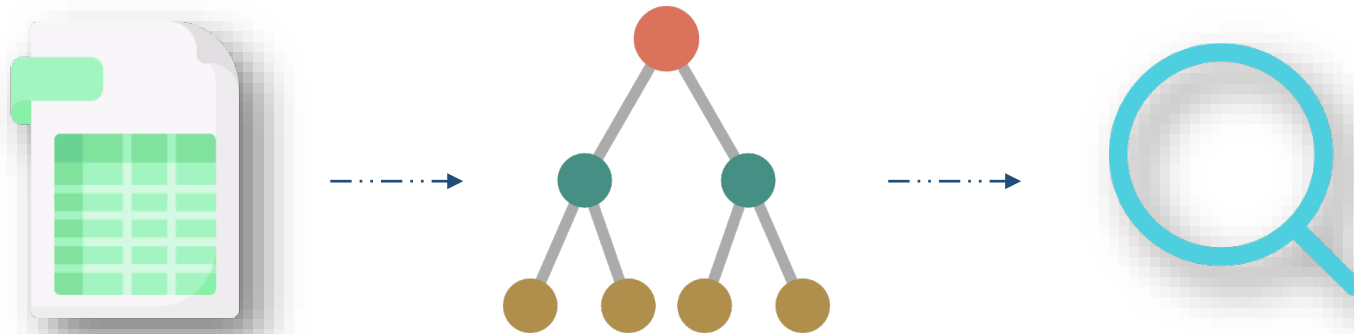
Regression



Clustering

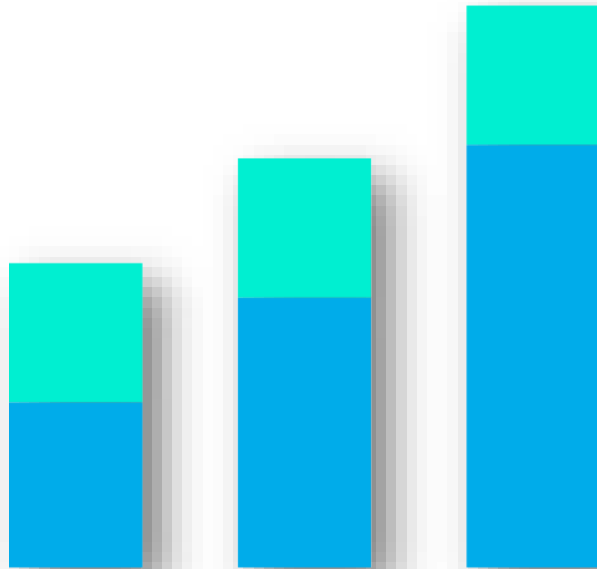
Pattern Evaluation

Once Data Mining techniques have been applied, the results have to be evaluated



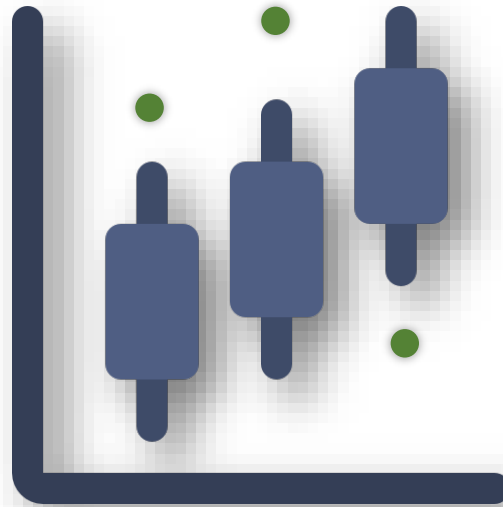
Knowledge Representation

The identified patterns must be represented using simple, aesthetic graphs



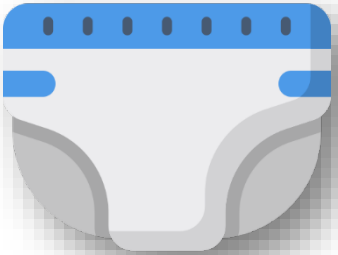
Data Mining Tasks–(Anomaly Detection)

Identification of unusual patterns, outliers helps in understanding the variation in data



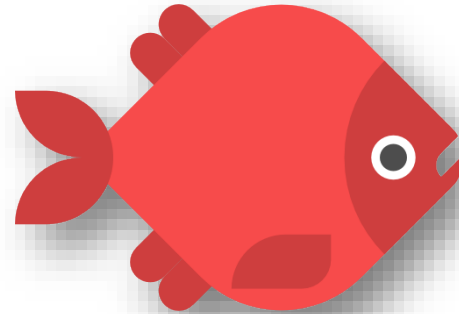
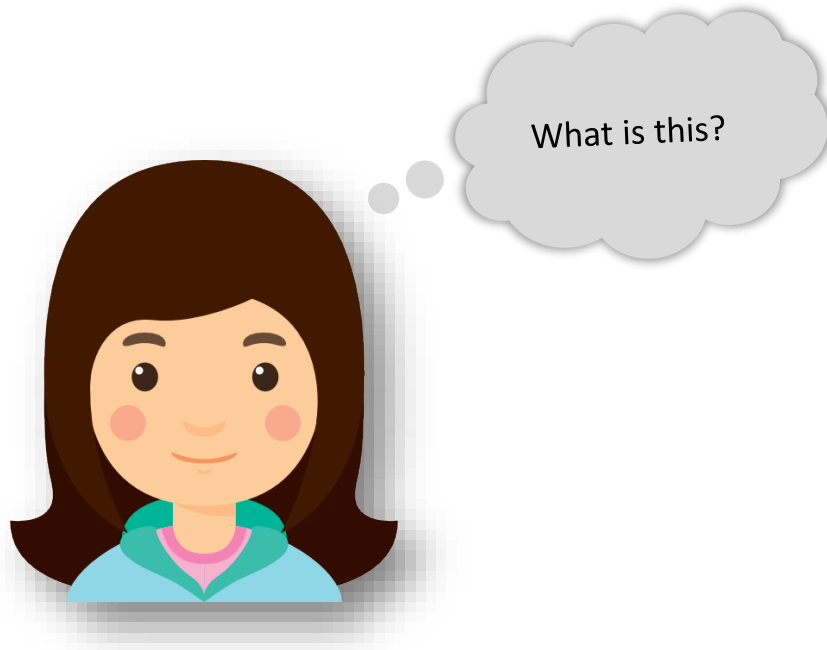
Data Mining Tasks–(Association Rule Mining)

This method is used for finding interesting association amongst different entities

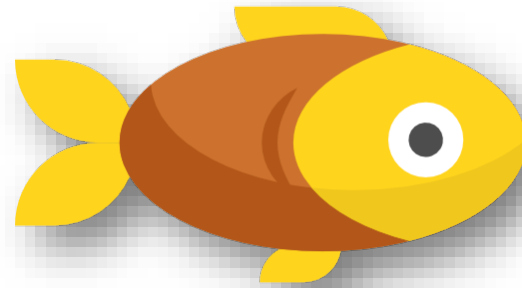


Beer Diaper Syndrome

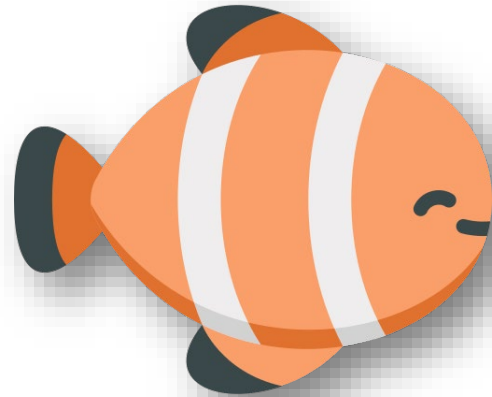
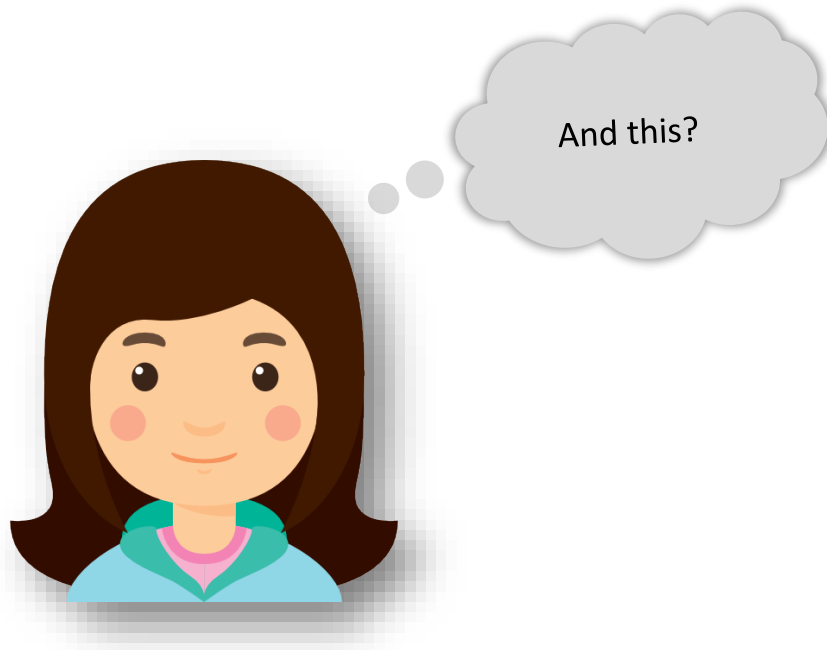
What is Machine Learning?



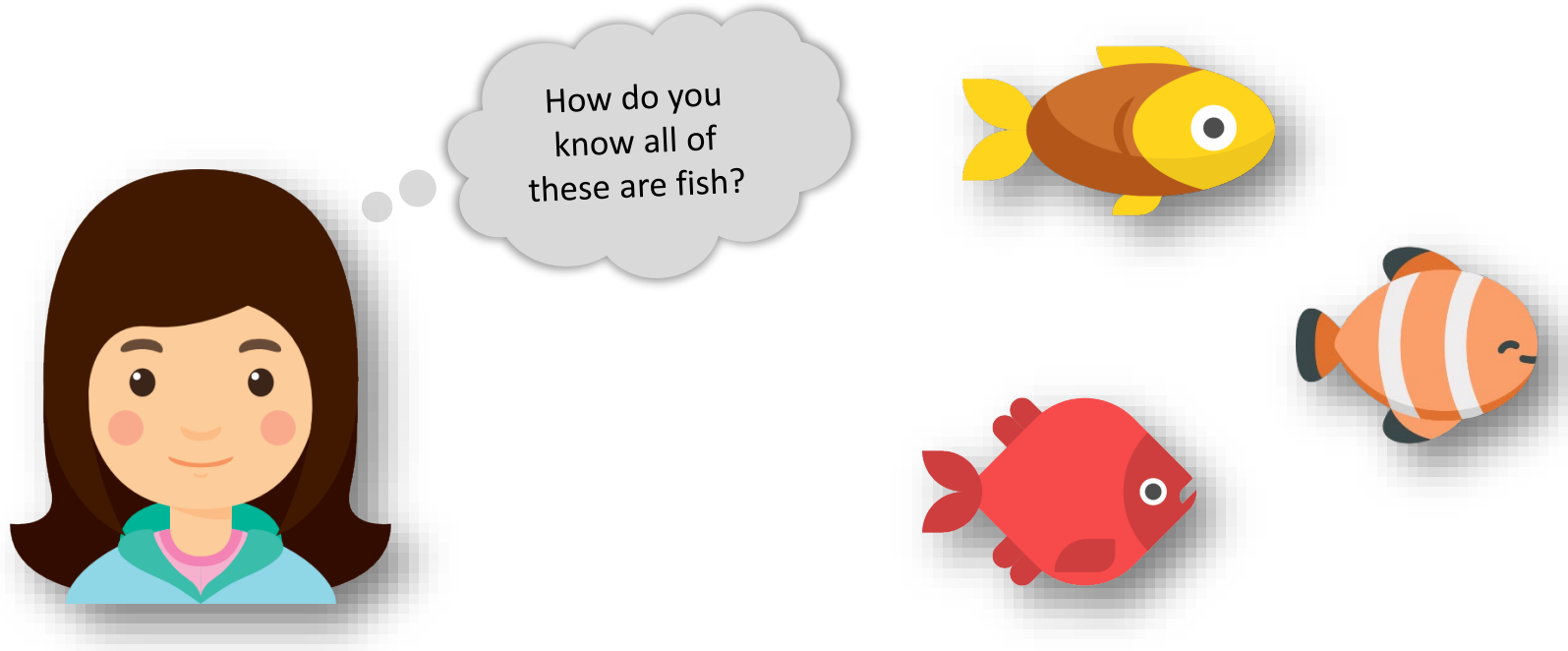
What is Machine Learning?



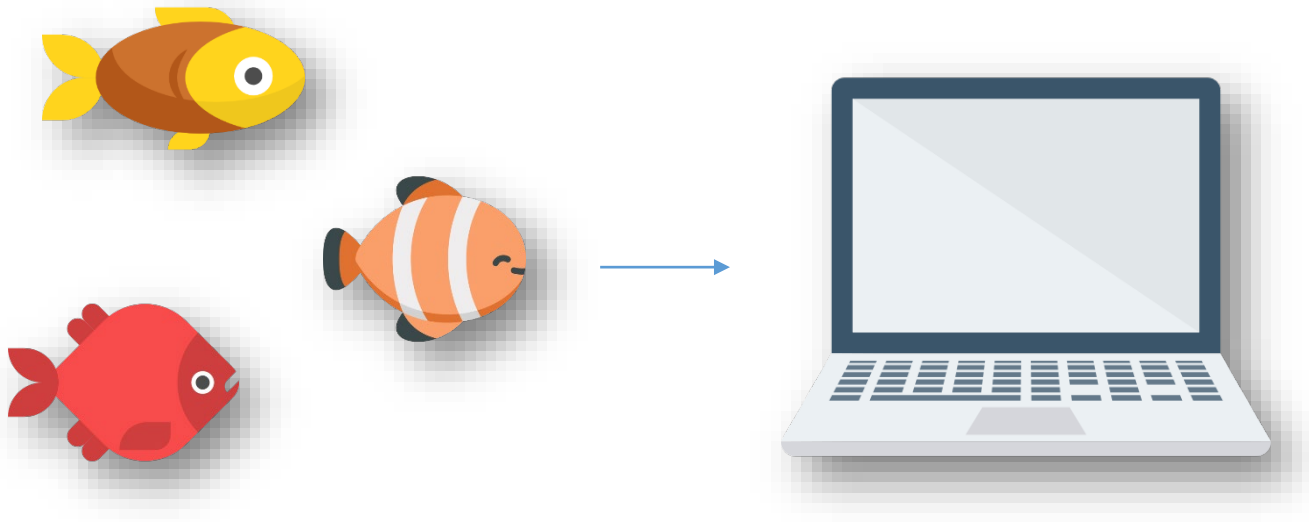
What is Machine Learning?



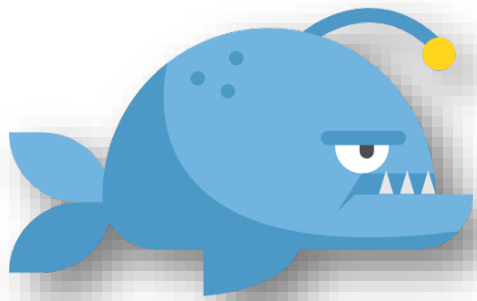
What is Machine Learning?



What is Machine Learning?

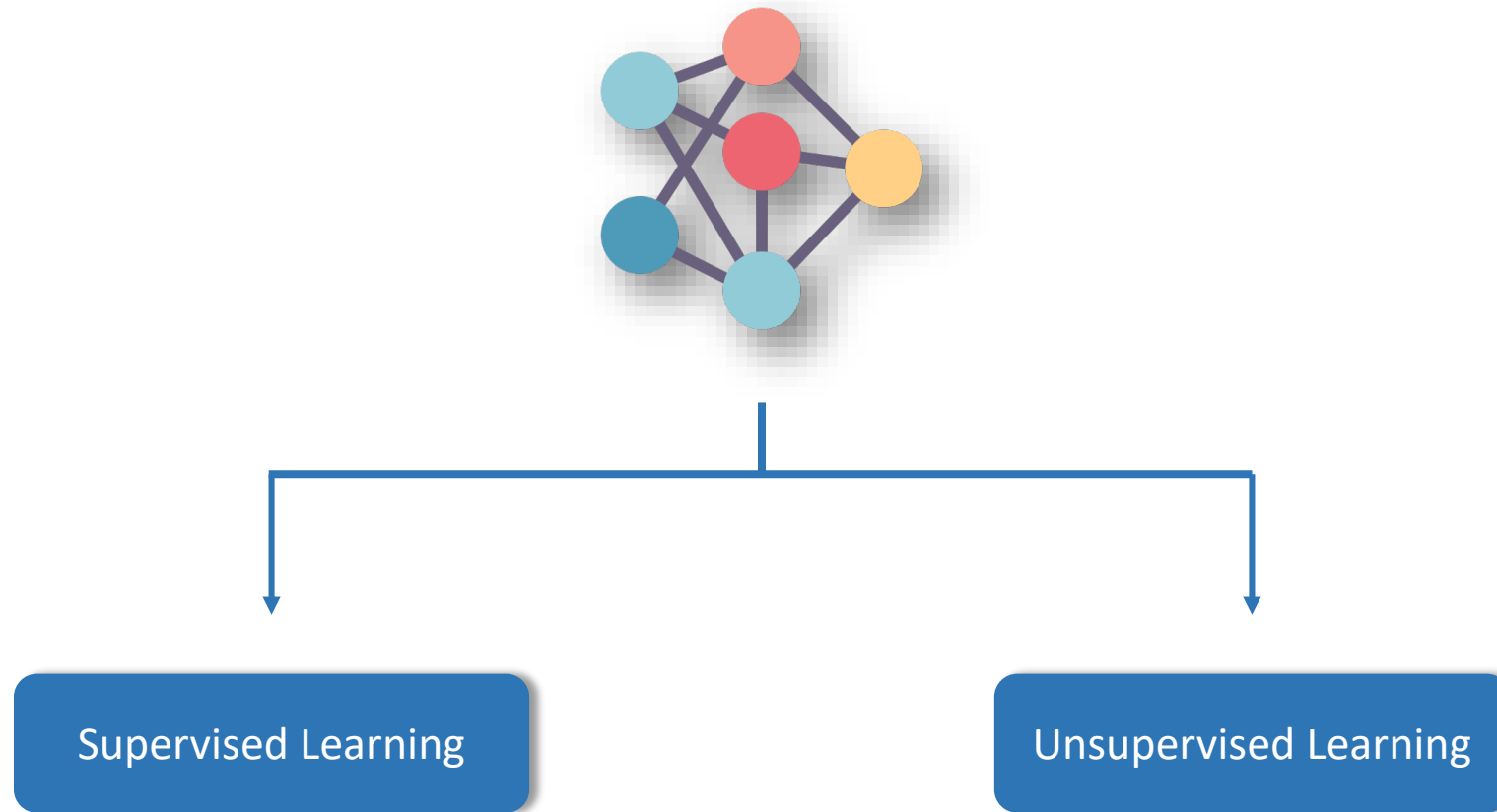


What is Machine Learning?

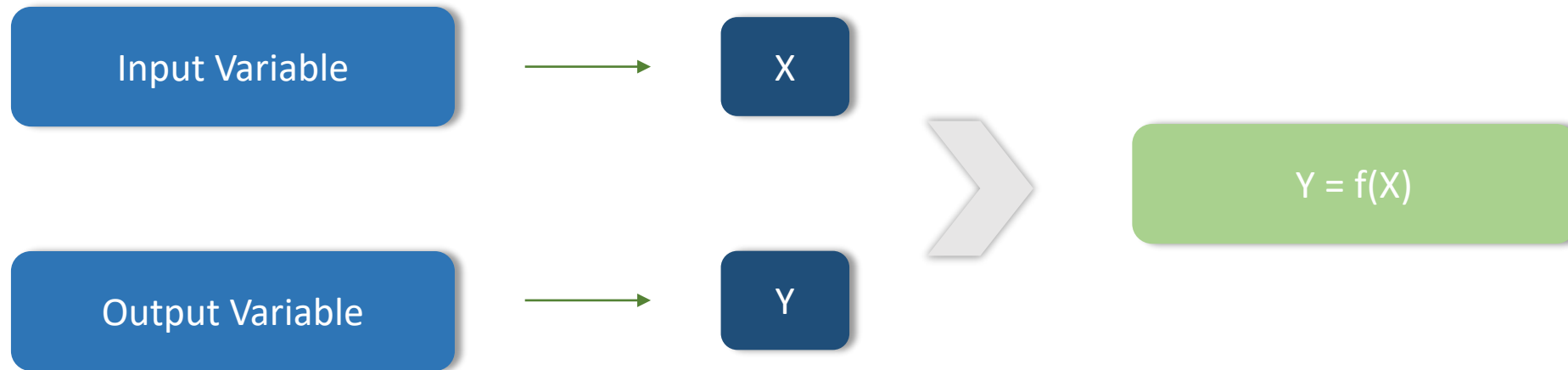


Fish

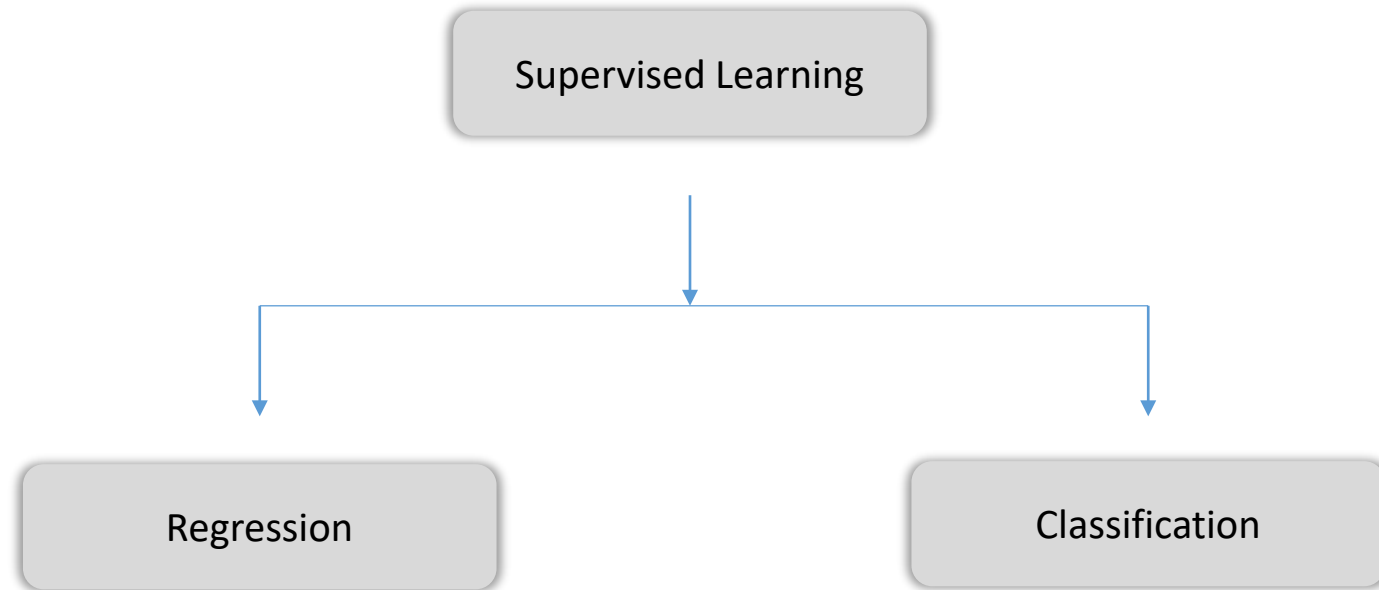
Categories of Machine Learning



Supervised Learning

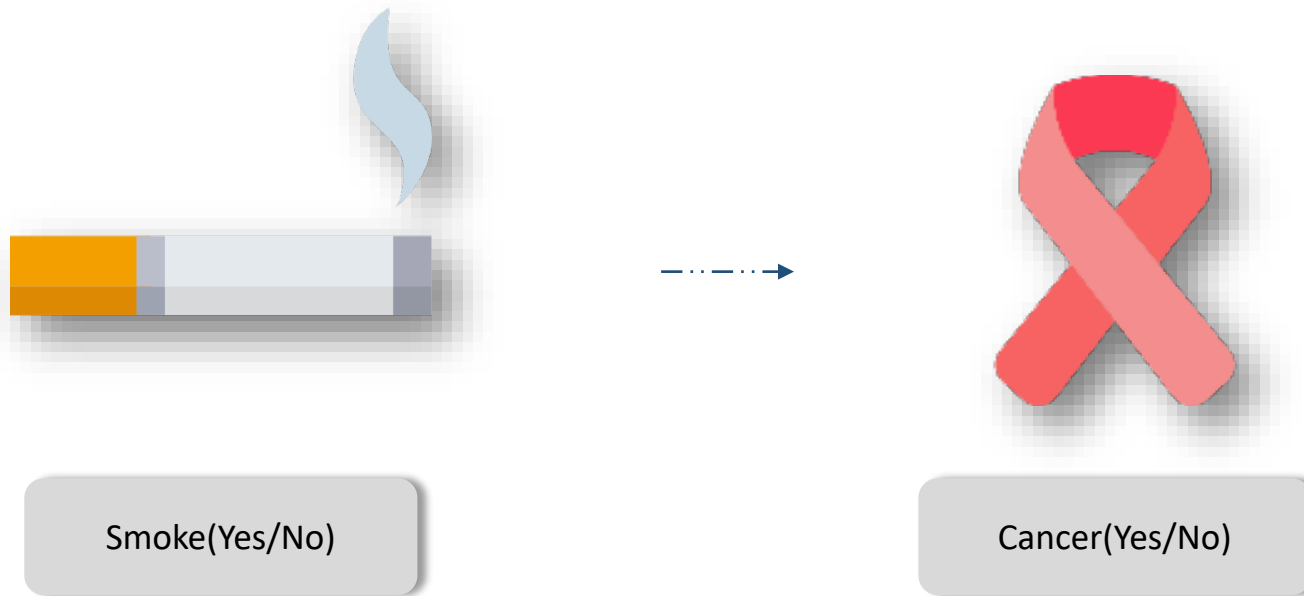


Categories of Supervised Learning



Classification

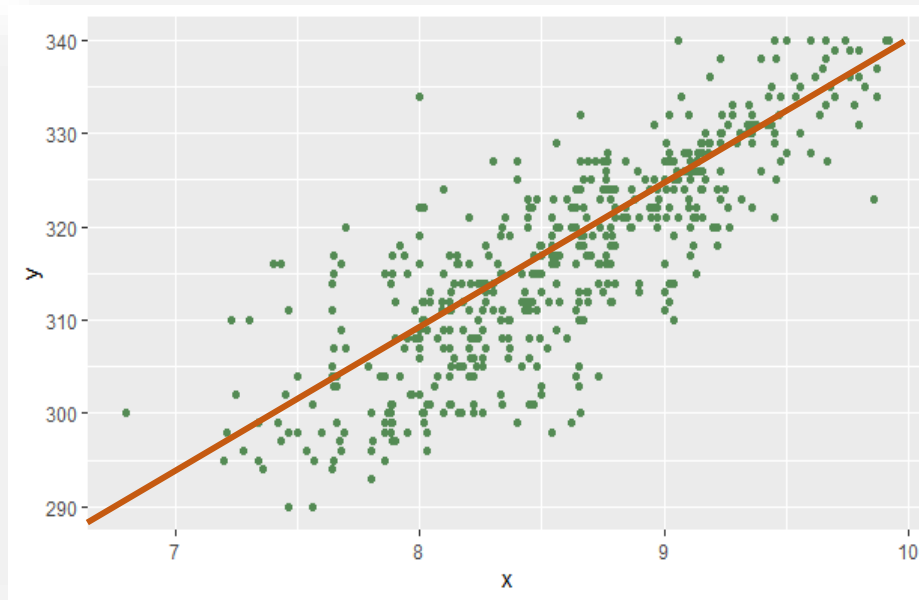
Classification is the process of predicting the class of a new variable



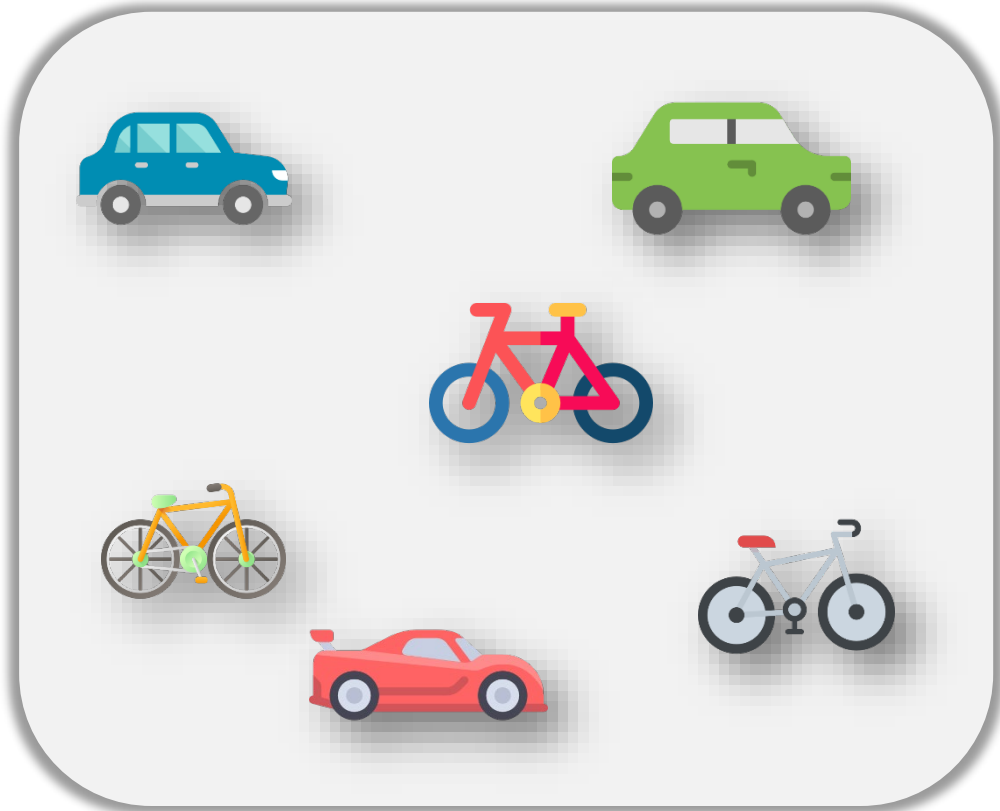
Regression

This method is used to estimate the relationship between different entities

Dependent Variable \leftarrow $Y=f(x)$ \rightarrow Independent Variable



Unsupervised Learning (Clustering)



Input Data with no
class labels

Unsupervised Learning (Clustering)

