

# Goals and Activities in the Machine Learning Process

# After this video you will be able to...

- Explain the goals of each step in the machine learning process
- List key activities in each step in the process

```
graph LR; A[ACQUIRE] --> B[PREPARE]; B --> C[ANALYZE]; C --> D[REPORT]; D --> E[ACT]
```

ACQUIRE

PREPARE

ANALYZE

REPORT

ACT



**ACQUIRE**

PREPARE

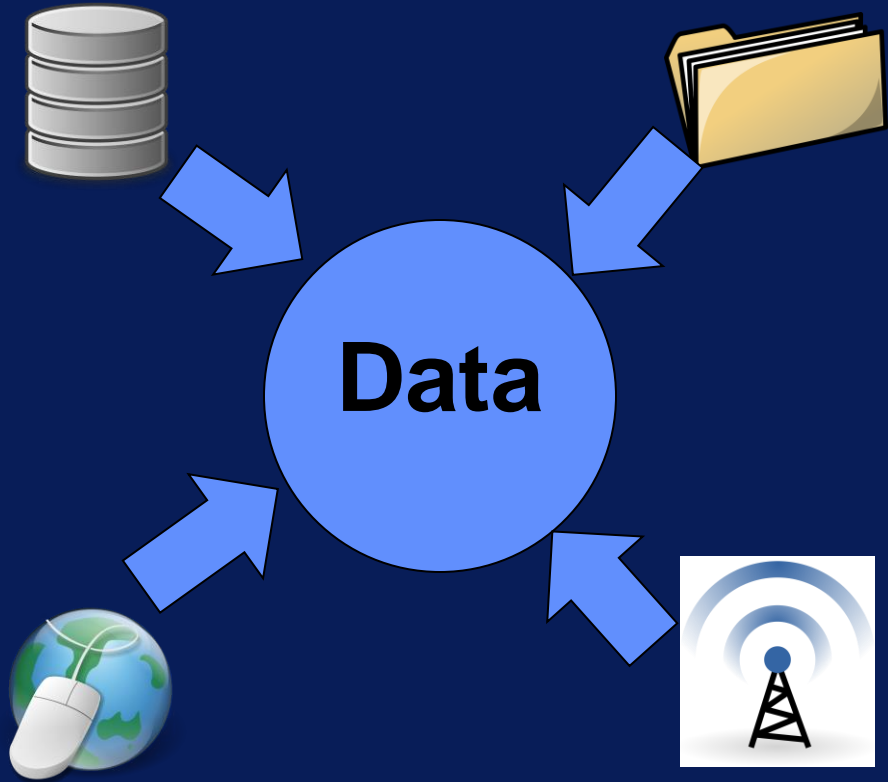
ANALYZE

REPORT

ACT

Goal: Identify and obtain all data  
related to problem

# Acquire Data



Identify data sources  
Collect data  
Integrate data



Step 2-A: Explore

Step 2-B: Pre-process



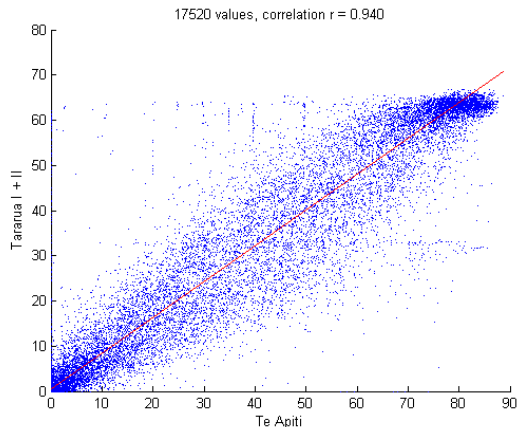
# Why Explore?

Goal: Understand your data

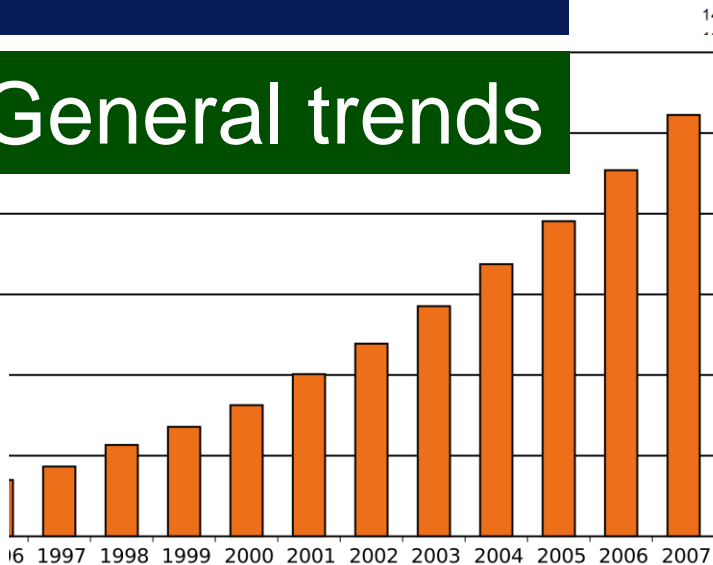


# Why Explore?

## Correlations

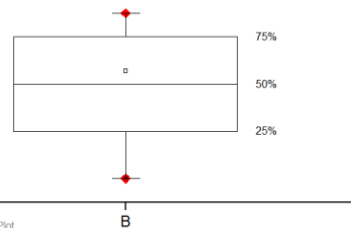


## General trends



## Outliers

$\{-1, 0, 1, 2, 3, 4, 5, 6, 12\}$

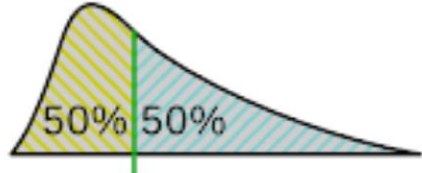




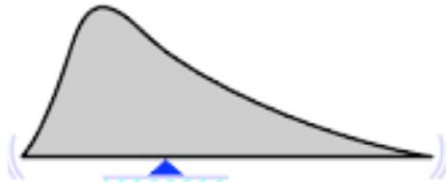
# Describe Your Data



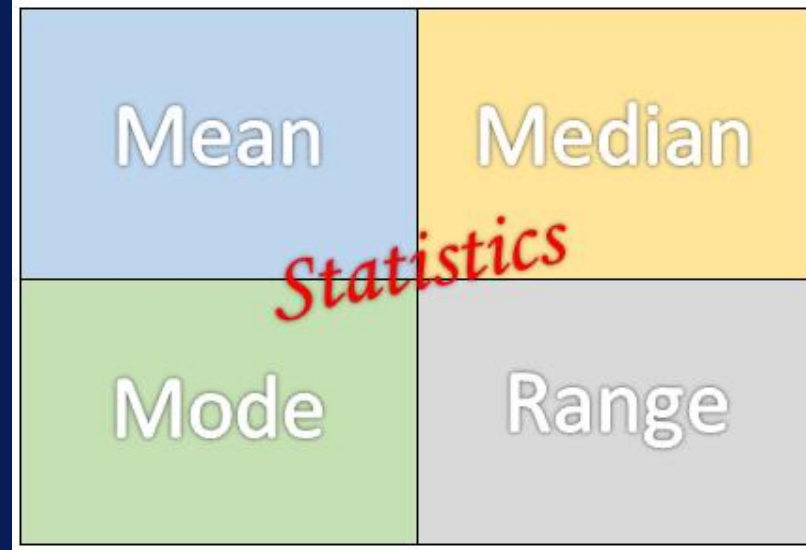
mode



median

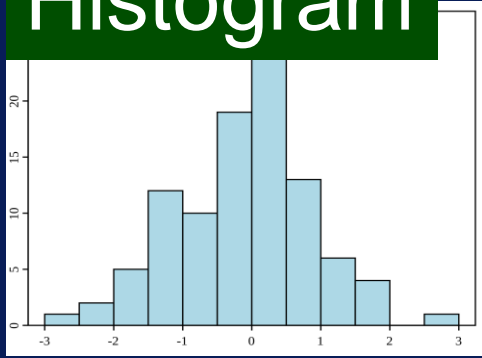


mean

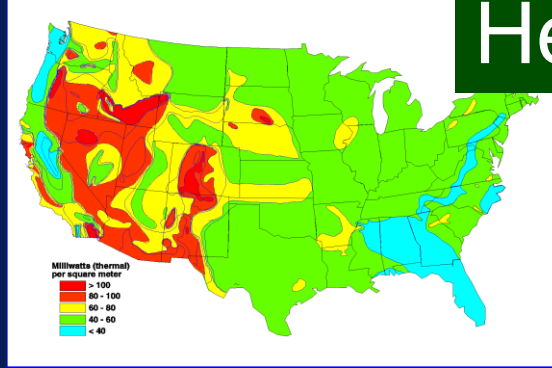


# Visualize Your Data

## Histogram



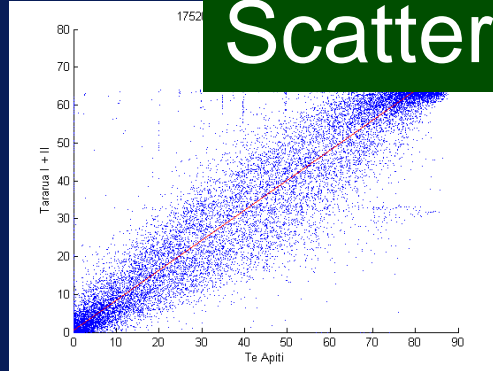
## Heat map



## Line plot



## Scatter plot





Step 2-A: Explore

Step 2-B: Pre-process



Step 2-A: Explore

Goal: Create data for analysis

Step 2-B: Pre-process

Clean

Select

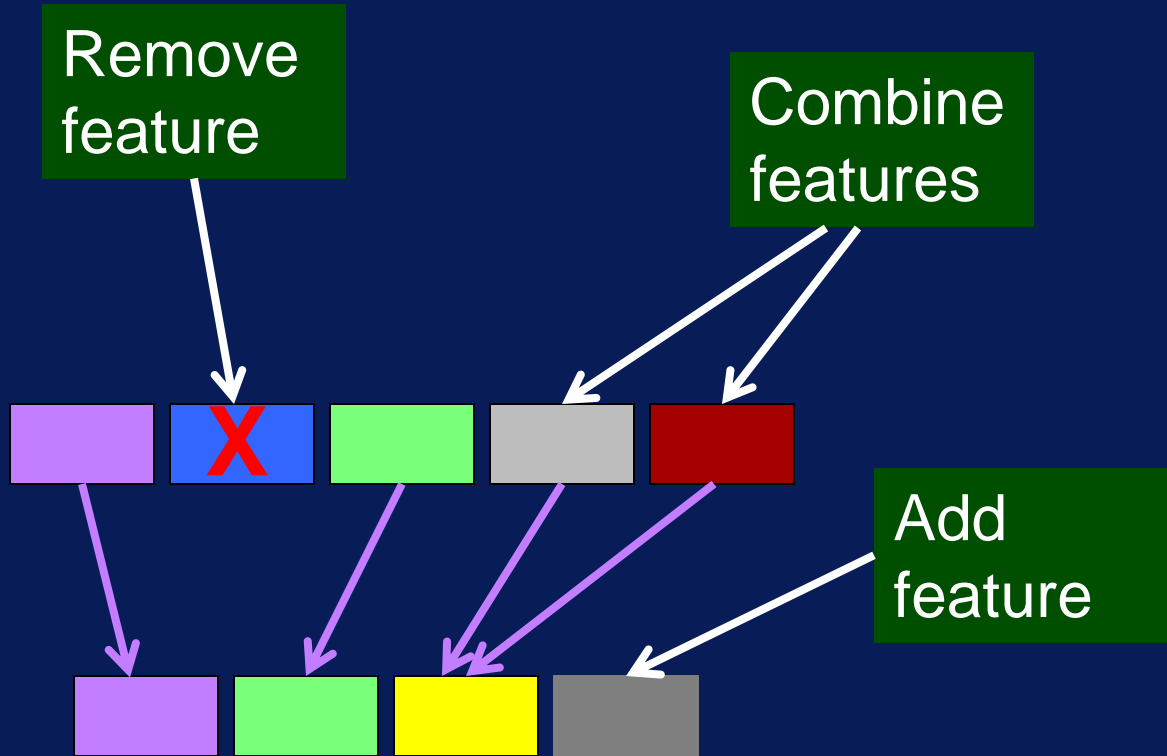
Transform

# Data Cleaning

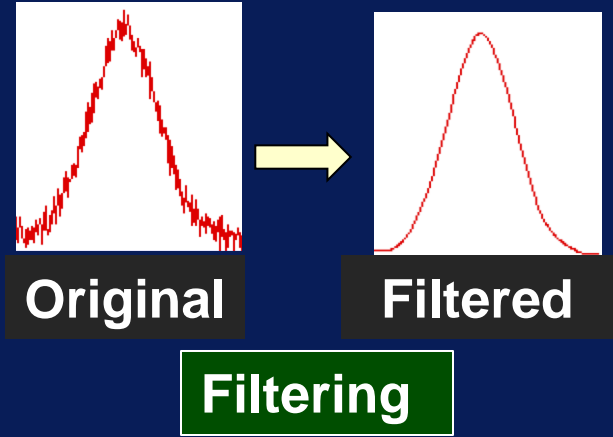
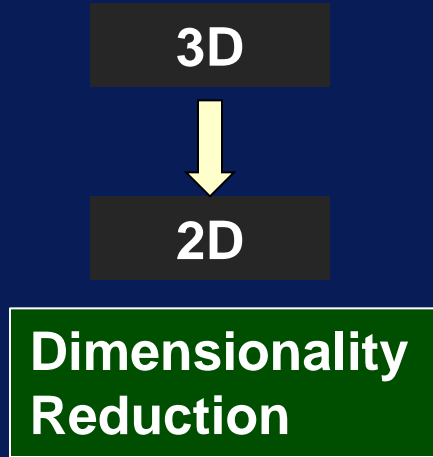
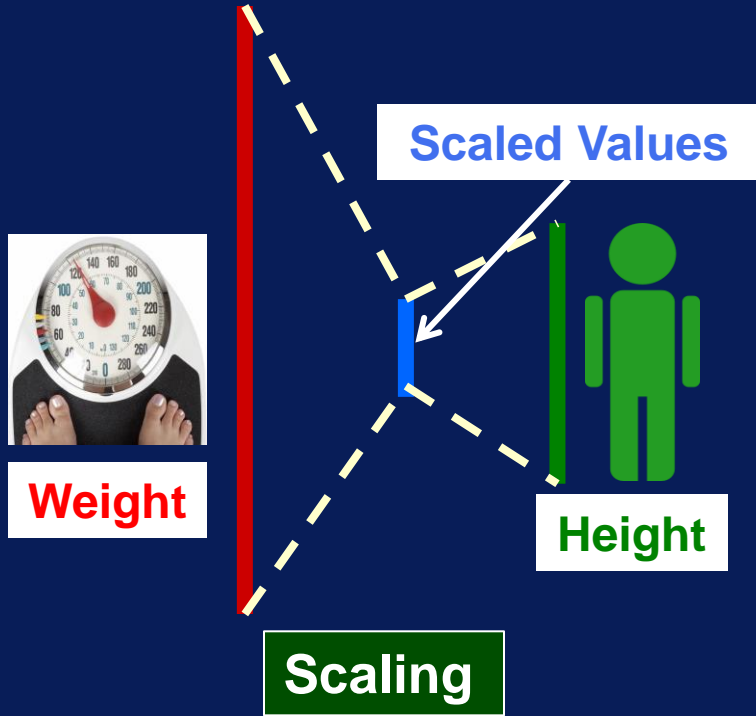
- Missing values
- Duplicate data
- Inconsistent data
- Noise
- Outliers



# Feature Selection



# Feature Transformation





## Goals:

- Build model
- Evaluate results



# Analyze

Select technique

Build model

Evaluate

Classification  
Regression  
Cluster Analysis  
Association  
Analysis

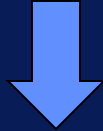


# Analyze

**Select technique**



**Build model**



**Evaluate results**



Goal: Communicate results and recommend actions

# Present

# Report



## with



## using

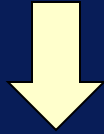




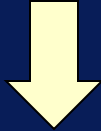
Goal: Determine actions based on insights

# Act

Determine action



Implement



Assess impact

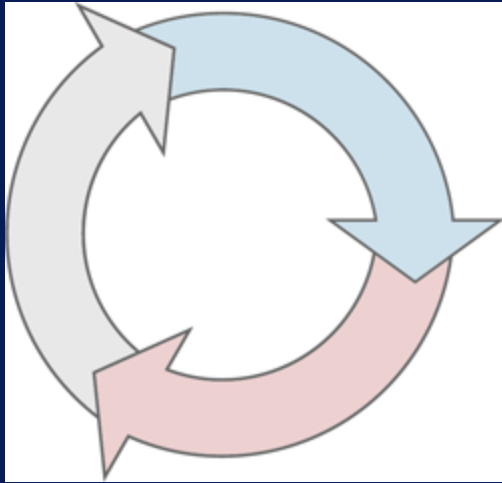
ACQUIRE

PREPARE

ANALYZE

REPORT

ACT



**Iterative process**