

# A.I. summer program

---

Hands on data manipulation  
with Pandas

Yueh-Lin Tsai  
2019 / 07

# About me

---



- Yueh-Lin Tsai
- Education
  - National Cheng Kung University, M.S., Psychology (2013-2015)
  - National Cheng Kung University, B.S., Psychology (2009-2013)
- Present
  - AI Engineer in Taiwan AI Academy

# What is Artificial Intelligence ?

---



# Artificial Intelligence

---

- Definition : Intelligence demonstrated by machine
- How ?

Expertise system

Explicit rule from human

Machine learning

Get knowledge from data

# Artificial Intelligence

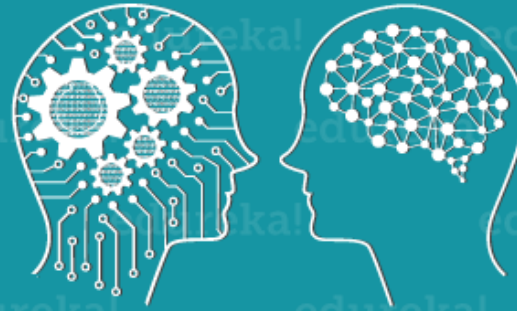
## ARTIFICIAL INTELLIGENCE

Engineering of making Intelligent Machines and Programs



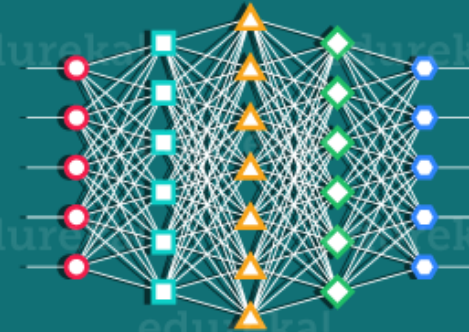
## MACHINE LEARNING

Ability to learn without being explicitly programmed



## DEEP LEARNING

Learning based on Deep Neural Network





+

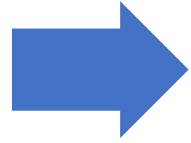


# Work with Python



## Collect Data

- BeautifulSoup
- Lxml
- Requests
- Pandas



## Preprocessing and EDA

- Numpy
- Pandas
- Scikit-learn
- Matplotlib
- NLTK



## Analysis and Modeling

- Statsmodels
- Scikit-learn
- Tensorflow
- Keras
- Pytorch

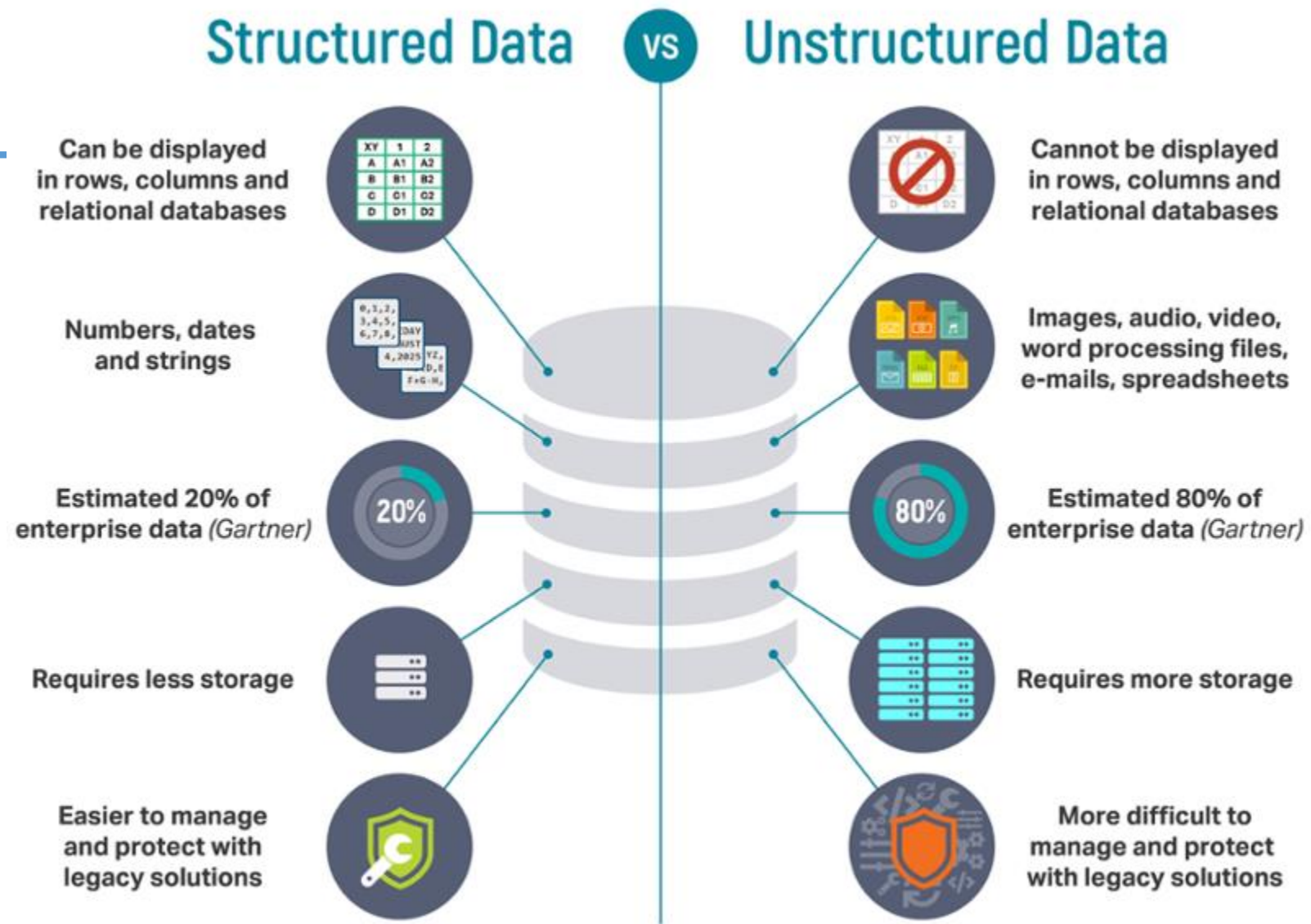
# Before we start to talk about models..

---





- Structured Data
  - table
- Unstructured Data
  - Text
  - Image
  - Voice





# Pandas

Sources for pandas-learner

- [Pandas tutorials](#)
- [Pandas exercises](#)

# What we do with Pandas

---

- Read data
- View data
- Change data types
- Handle outlier
- Select column
- Filter data
- Sorting
- Delete data
- Descriptive stats
- Group
- Concatenate
- Apply
- Cross-table
- Plotting
- Export data

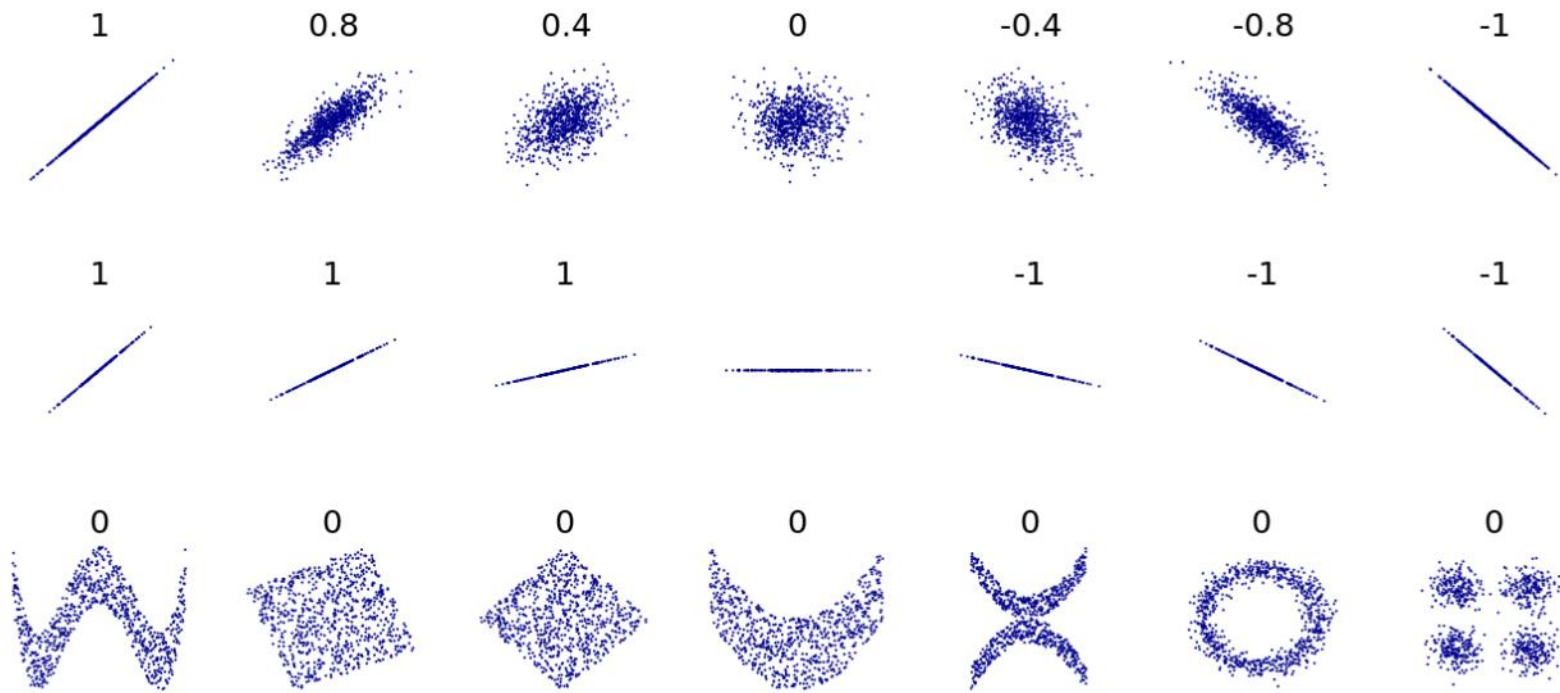
Supplementary

# Data types (structure data)

Source : [Wikipedia](#)

Category	example	Characteristic	Dtype in pandas
Nominal	Country	Elementary arithmetic	'Object' 'Category'
Ordinal	Rank	Sort	
Interval	Celsius	Add, Subtract	'float' 'int'
Ratio	Weight	Multiply, Division	

# Correlation



[Source](#)