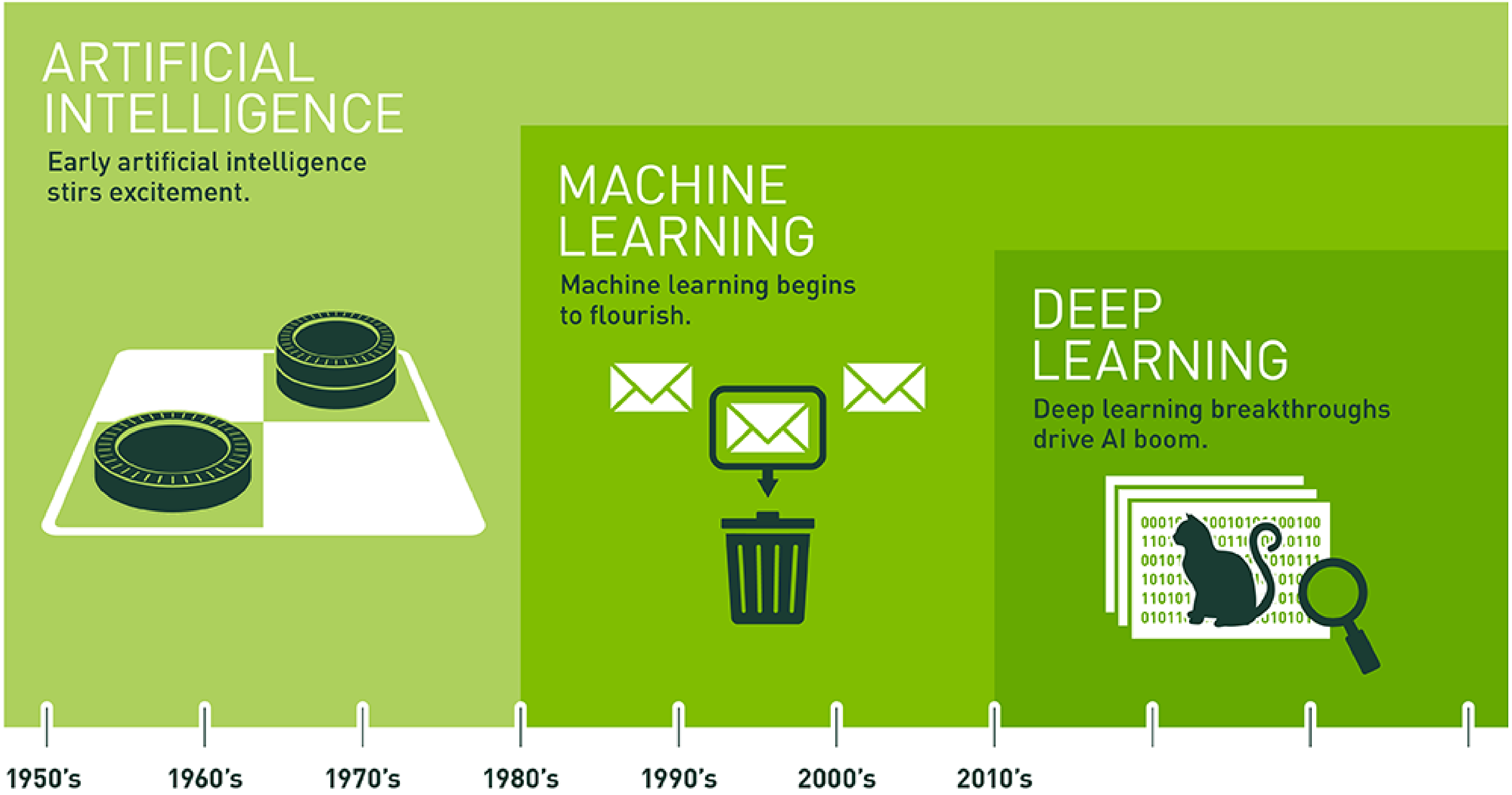
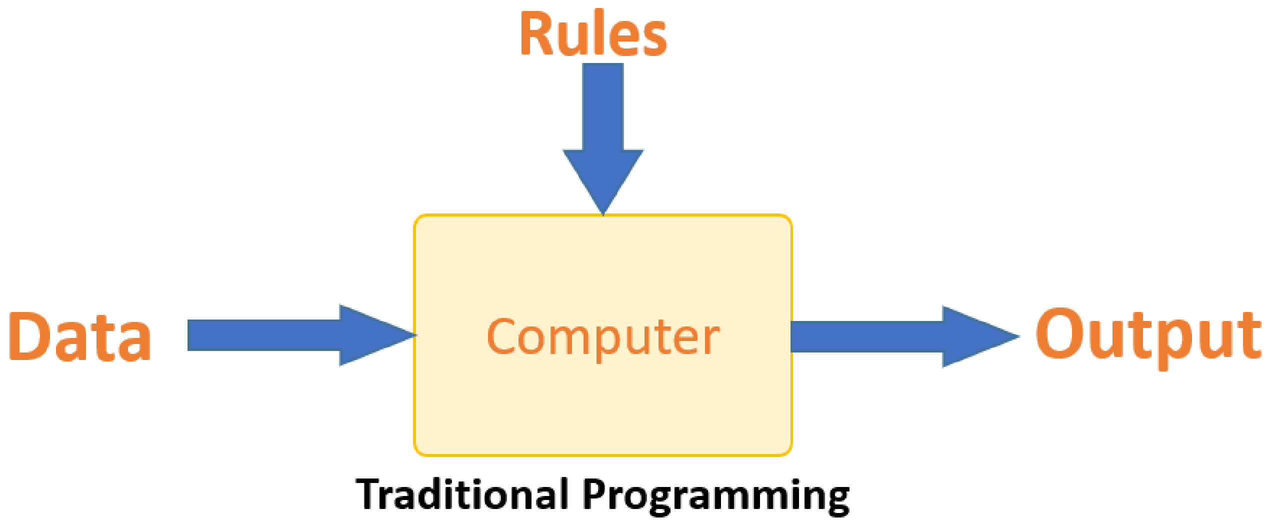
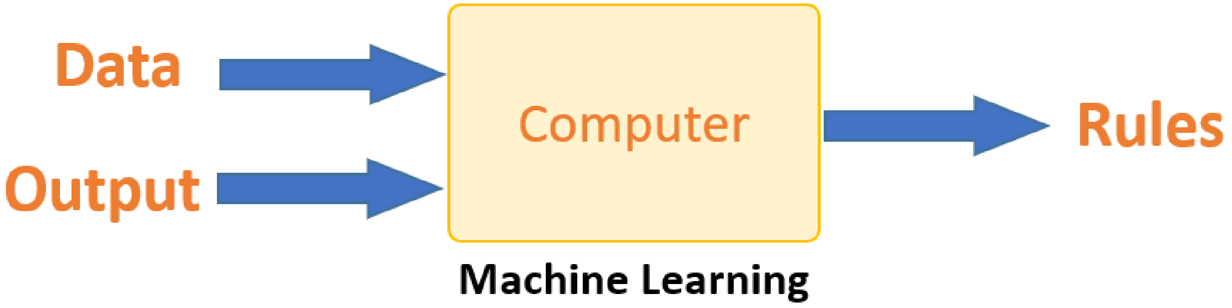
What is AI?



https://blogs.nvidia.com/blog/2016/07/29/whats-difference-artificial-intelligence-machine-learning-deep-learning-ai/

Traditional programming vs AI



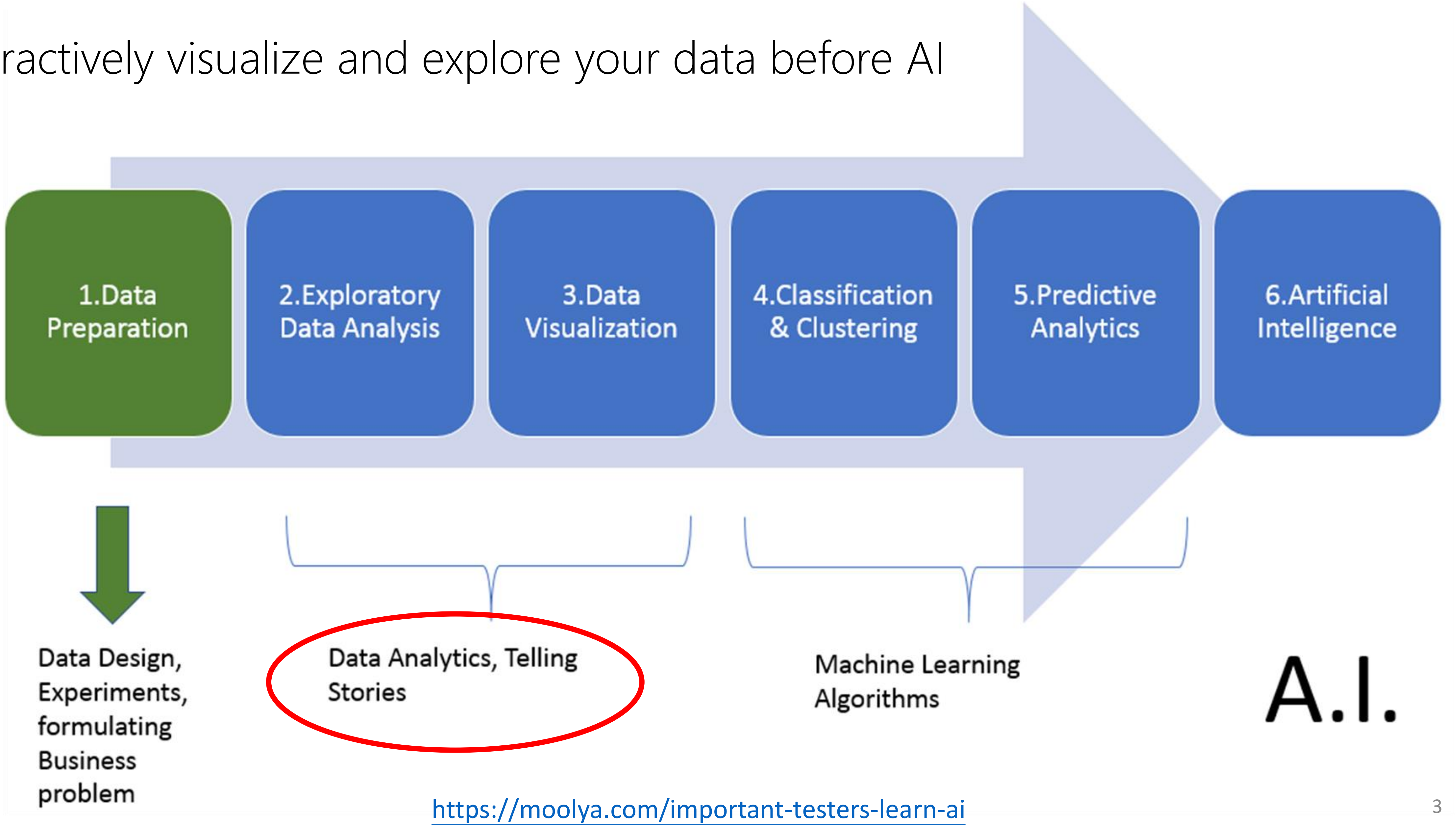


The practice of using algorithms to parse data, learn from it, and then make a determination or prediction about something in the world.

https://www.guru99.com/machine-learning-tutorial.html

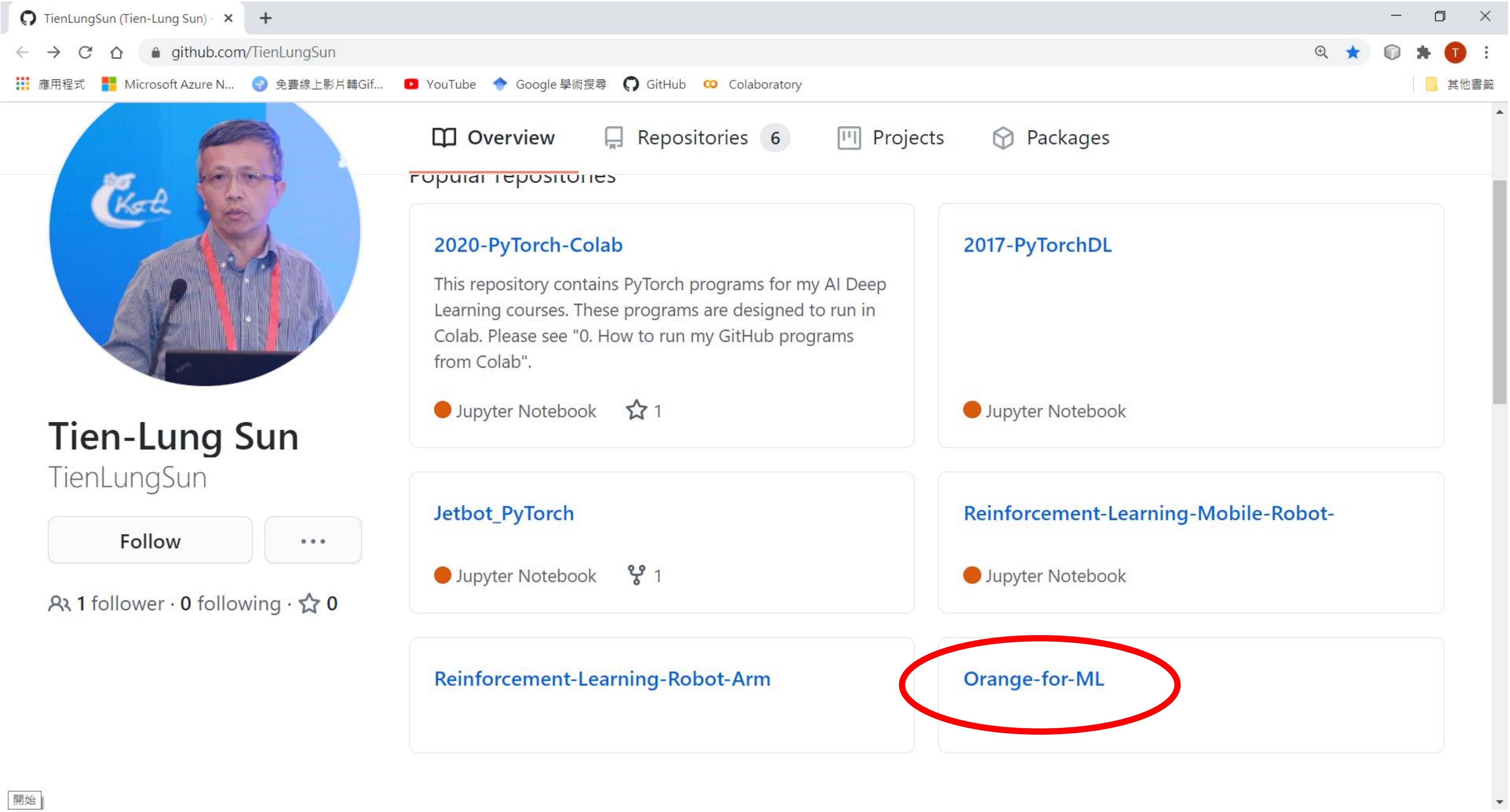
HI before AI

Interactively visualize and explore your data before AI



https://moolya.com/important-testers-learn-ai

Download data files from my GitHub

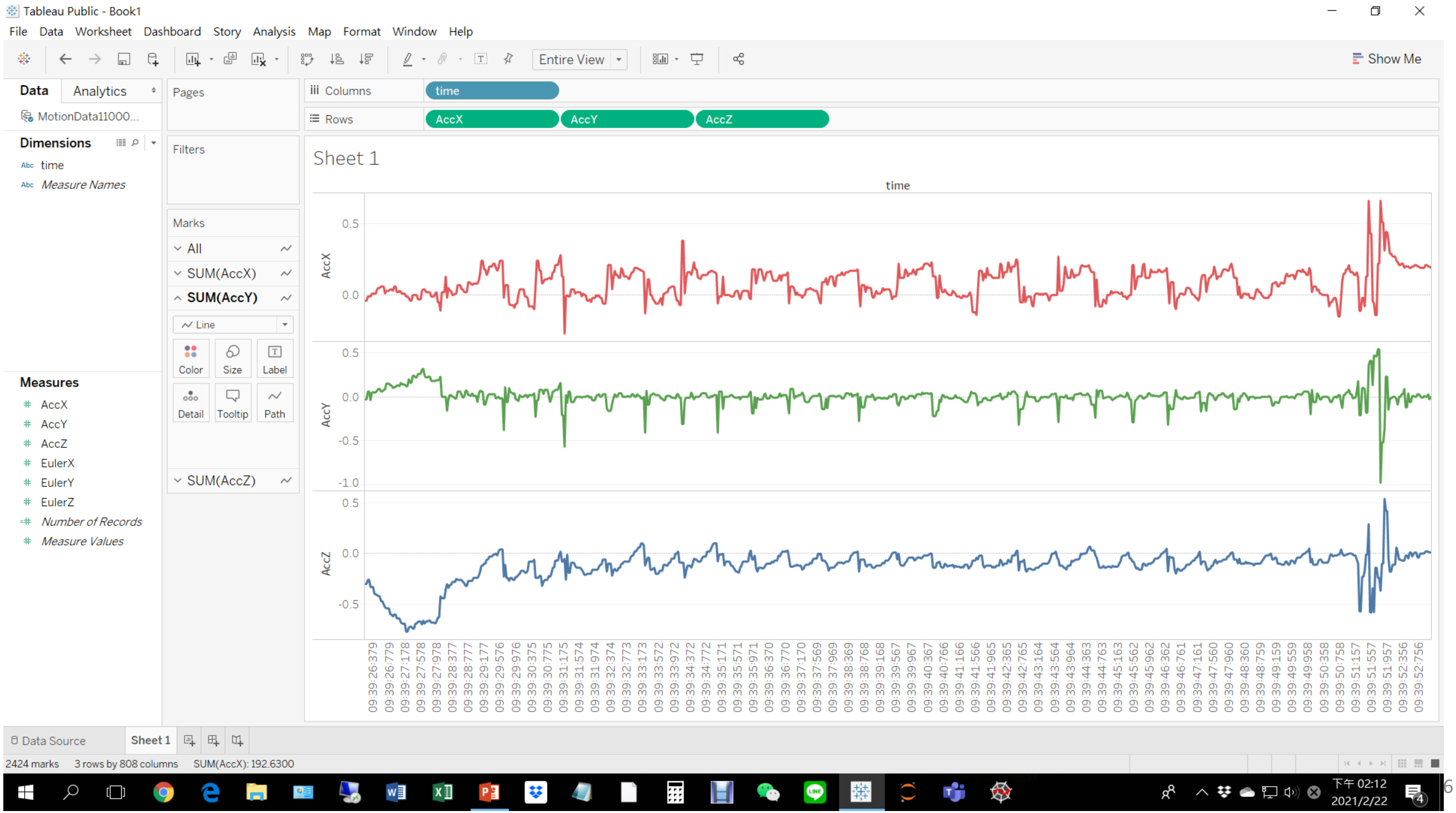


https://github.com/TienLungSun

Interactively visual exploration – (1) Tableau Public

https://public.tableau.com/s/

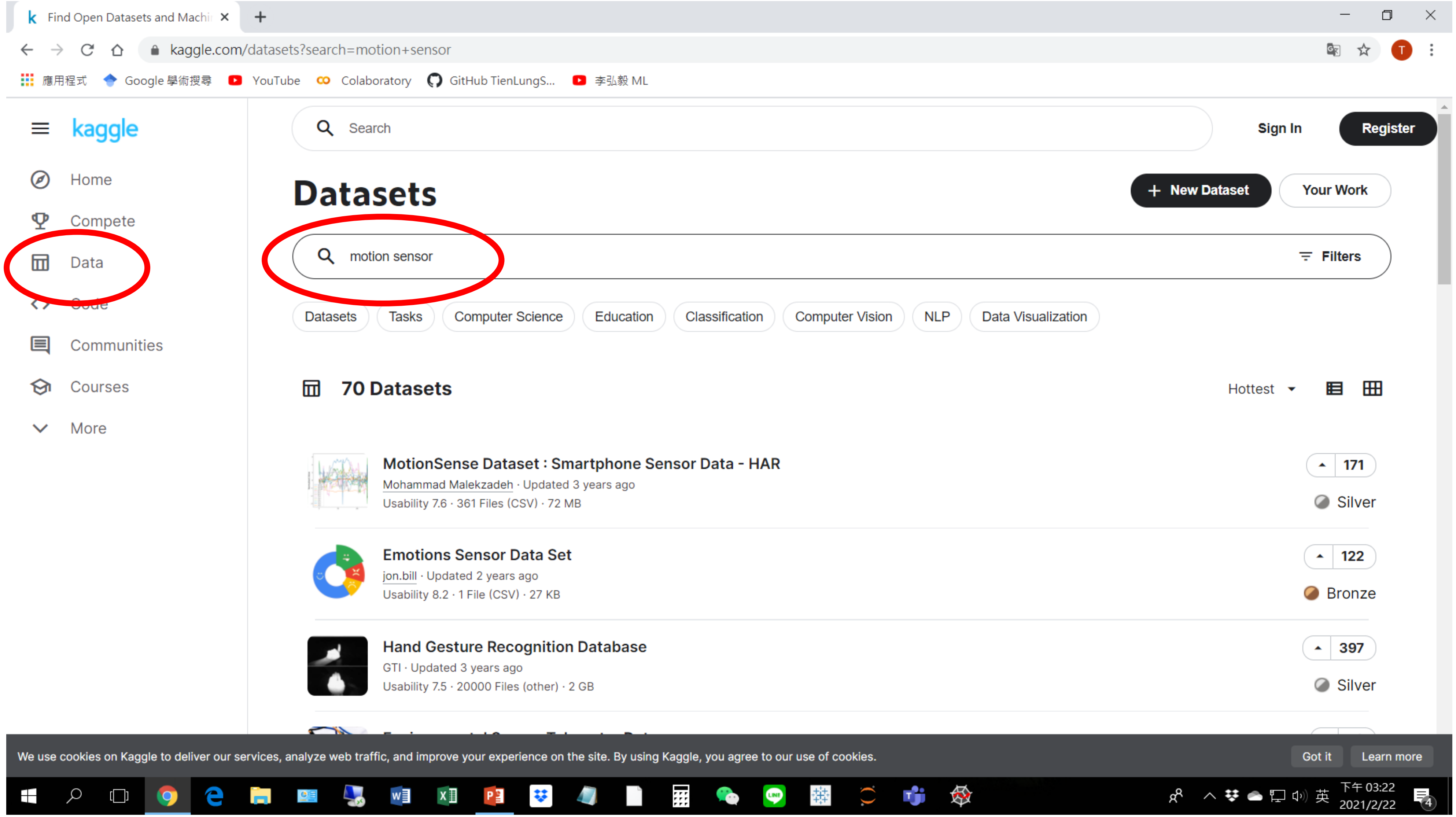
Tableau Public



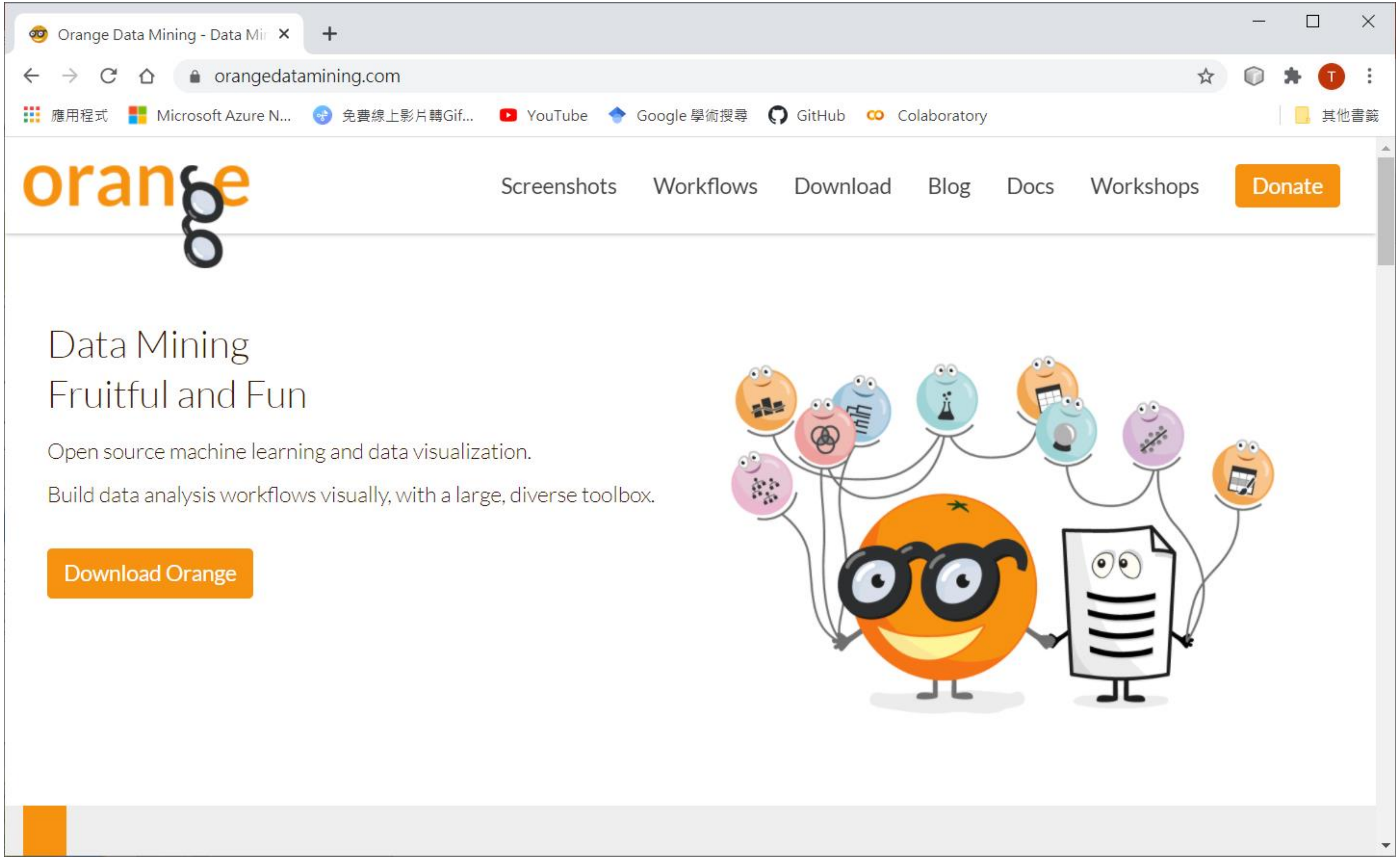
Practice – Tableau public

1. Download and install Tableau Public
2. Visualize the motion data file
3. Search Kaggle (https://www.kaggle.com/) to find a sensor data file (see next slide)
4. Use Tableau public to visualize the data file

Kaggle

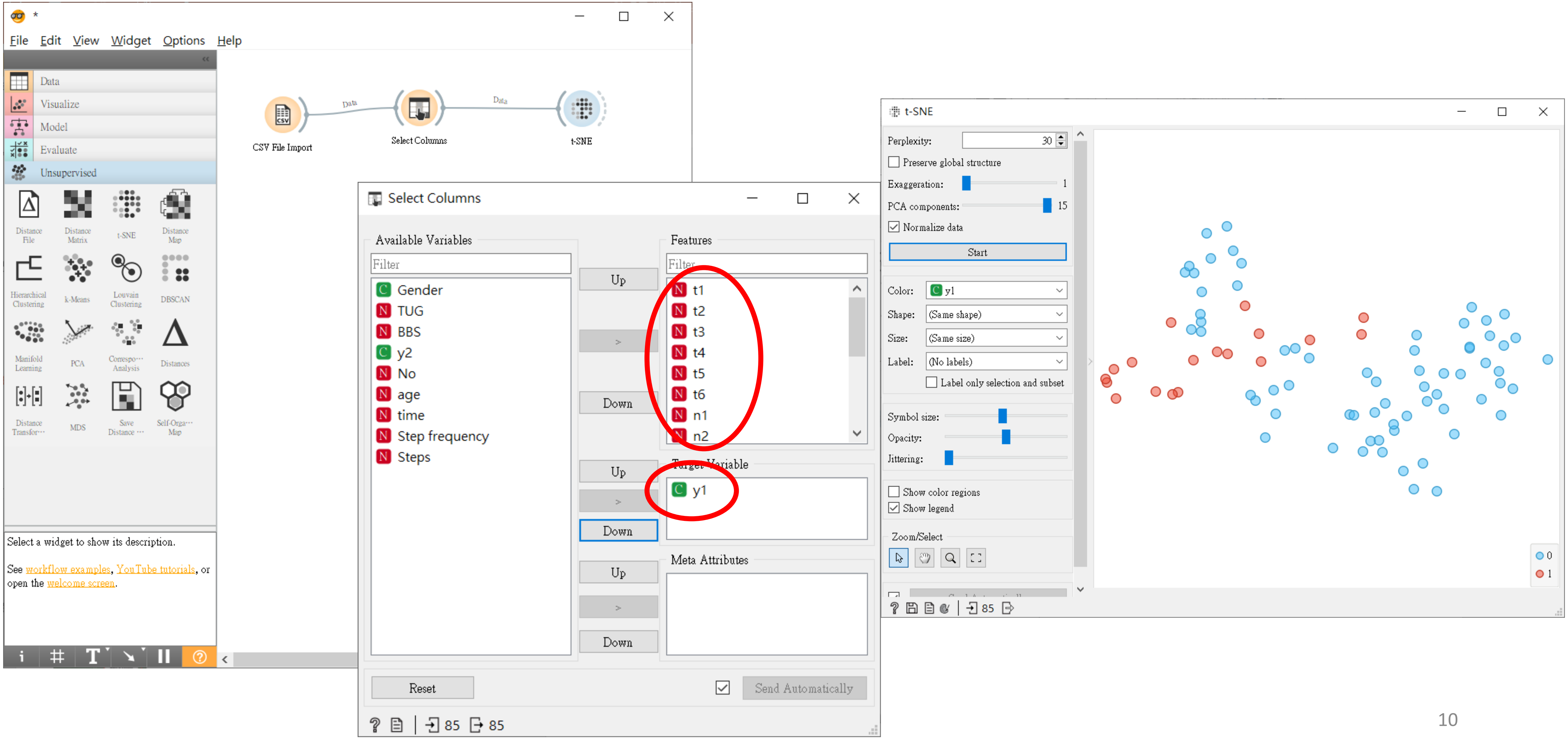

Find Open Datasets and Machi
kaggle.com/datasets?search=motion+sensor
應用程式
Google 學術搜尋
YouTube
Colaboratory
GitHub TienLungS...
李弘毅 ML
kaggle
Home
Compete
Data
Code
Communities
Courses
More
Search
Sign In
Register
Datasets
New Dataset
Your Work
motion sensor
Filters
Datasets
Tasks
Computer Science
Education
Classification
Computer Vision
NLP
Data Visualization
70 Datasets
Hottest
MotionSense Dataset : Smartphone Sensor Data - HAR
Mohammad Malekzadeh · Updated 3 years ago
Usability 7.6 · 361 Files (CSV) · 72 MB
171
Silver
Emotions Sensor Data Set
jon.bill · Updated 2 years ago
Usability 8.2 · 1 File (CSV) · 27 KB
122
Bronze
Hand Gesture Recognition Database
GTI · Updated 3 years ago
Usability 7.5 · 20000 Files (other) · 2 GB
397
Silver
We use cookies on Kaggle to deliver our services, analyze web traffic, and improve your experience on the site. By using Kaggle, you agree to our use of cookies.
Got it
Learn more
下午 03:22
2021/2/22

Interactively visual exploration – (2) Orange



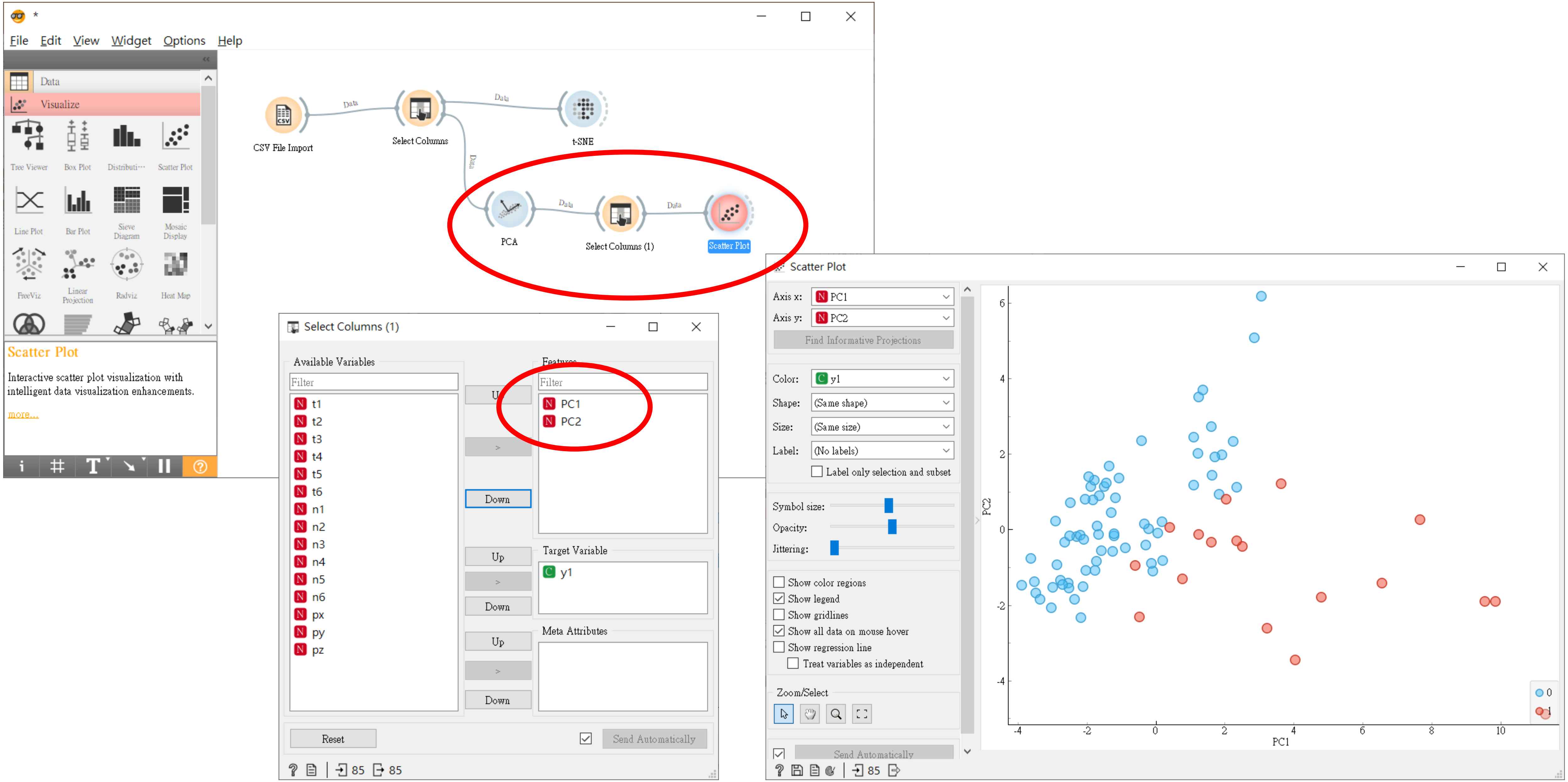
https://orangedatamining.com/

Orange



File Edit View Widget Options Help
Data
Visualize
Model
Evaluate
Unsupervised
Distance File
Distance Matrix
t-SNE
Distance Map
Hierarchical Clustering
k-Means
Louvain Clustering
DBSCAN
Manifold Learning
PCA
Correspo··· Analysis
Distances
Distance Transfor···
MDS
Save Distance ···
Self-Orga··· Map
Select a widget to show its description.
See workflow examples, YouTube tutorials, or open the welcome screen.
CSV File Import
Data
Select Columns
Data
t-SNE
Select Columns
Available Variables
Filter
Gender
TUG
BBS
y2
No
age
time
Step frequency
Steps
Up
Down
Features
Filter
t1
t2
t3
t4
t5
t6
n1
n2
Target Variable
y1
Meta Attributes
Reset
Send Automatically
85
85
t-SNE
Perplexity: 30
Preserve global structure
Exaggeration: 1
PCA components: 15
Normalize data
Start
Color: y1
Shape: (Same shape)
Size: (Same size)
Label: (No labels)
Label only selection and subset
Symbol size:
Opacity:
Jittering:
Show color regions
Show legend
Zoom/Select
85
0
1

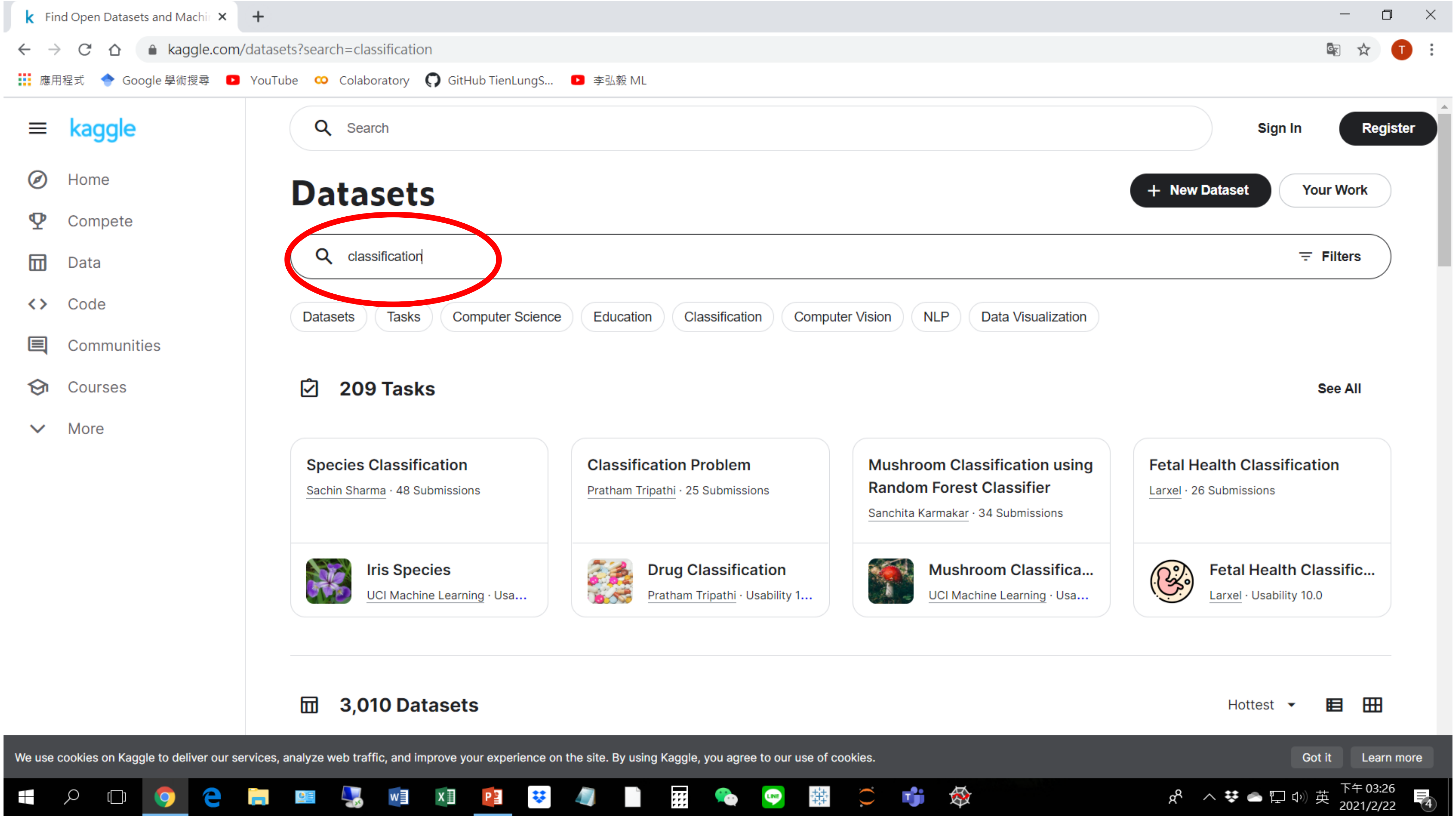
Orange


File Edit View Widget Options Help
Data
Visualize
Tree Viewer
Box Plot
Distributi…
Scatter Plot
Line Plot
Bar Plot
Sieve Diagram
Mosaic Display
FreeViz
Linear Projection
Radviz
Heat Map
Scatter Plot
Interactive scatter plot visualization with intelligent data visualization enhancements.
more...
CSV File Import
Data
Select Columns
Data
t-SNE
Data
PCA
Data
Select Columns (1)
Data
Scatter Plot
Select Columns (1)
Available Variables
Filter
t1
t2
t3
t4
t5
t6
n1
n2
n3
n4
n5
n6
px
py
pz
Features
Filter
PC1
PC2
Down
Up
Down
Up
Down
Target Variable
y1
Meta Attributes
Reset
Send Automatically
85
85
Scatter Plot
Axis x: PC1
Axis y: PC2
Find Informative Projections
Color: y1
Shape: (Same shape)
Size: (Same size)
Label: (No labels)
Label only selection and subset
Symbol size:
Opacity:
Jittering:
Show color regions
Show legend
Show gridlines
Show all data on mouse hover
Show regression line
Treat variables as independent
Zoom/Select
Send Automatically
85
PC2
PC1
0
1

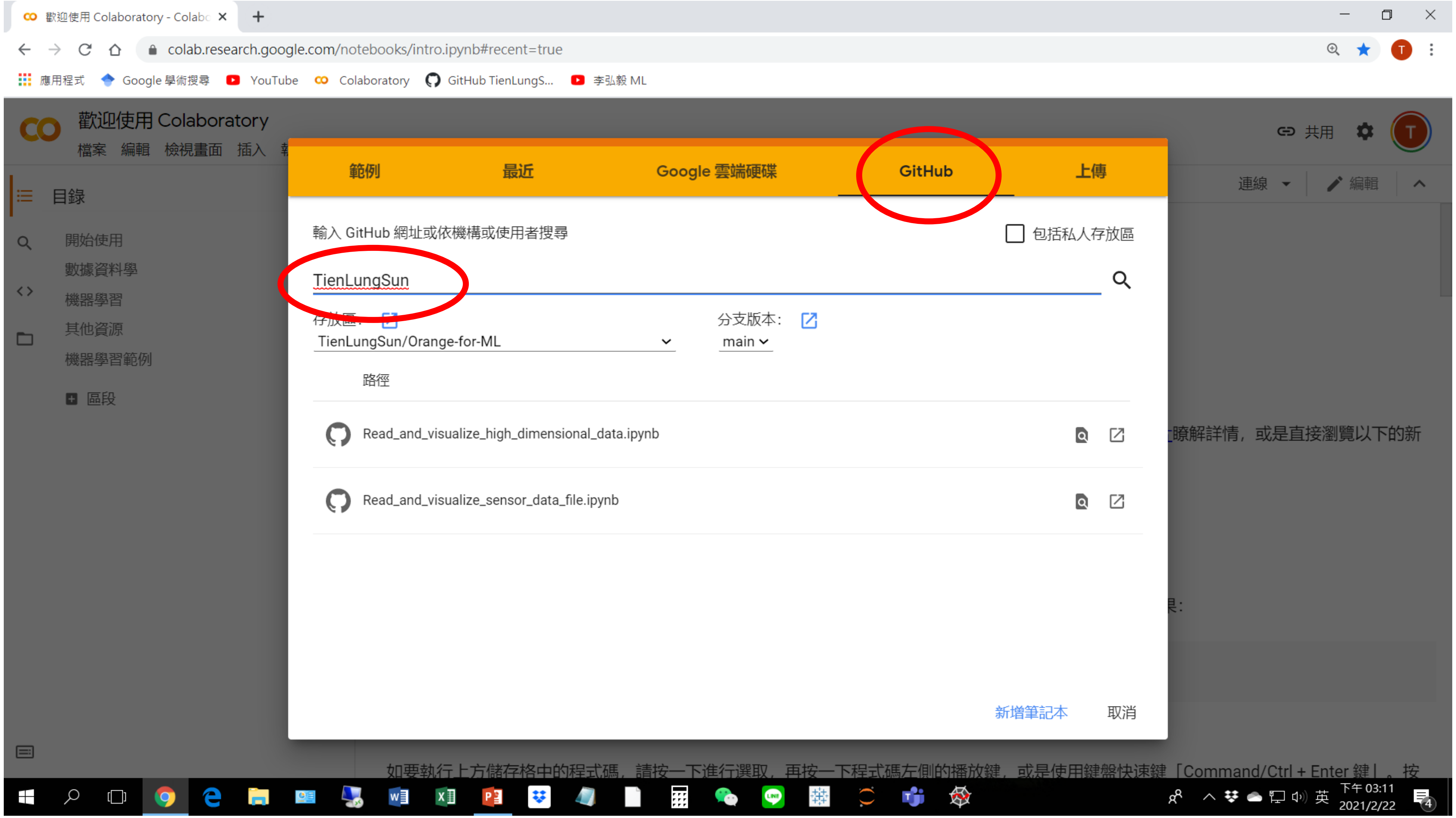
Practice – Orange

1. Download and install Orange
2. Visualize the 3M TUG data file
3. Search Kaggle to find a classification data file
4. Use Orange to visualize the high dimensional data

Kaggle

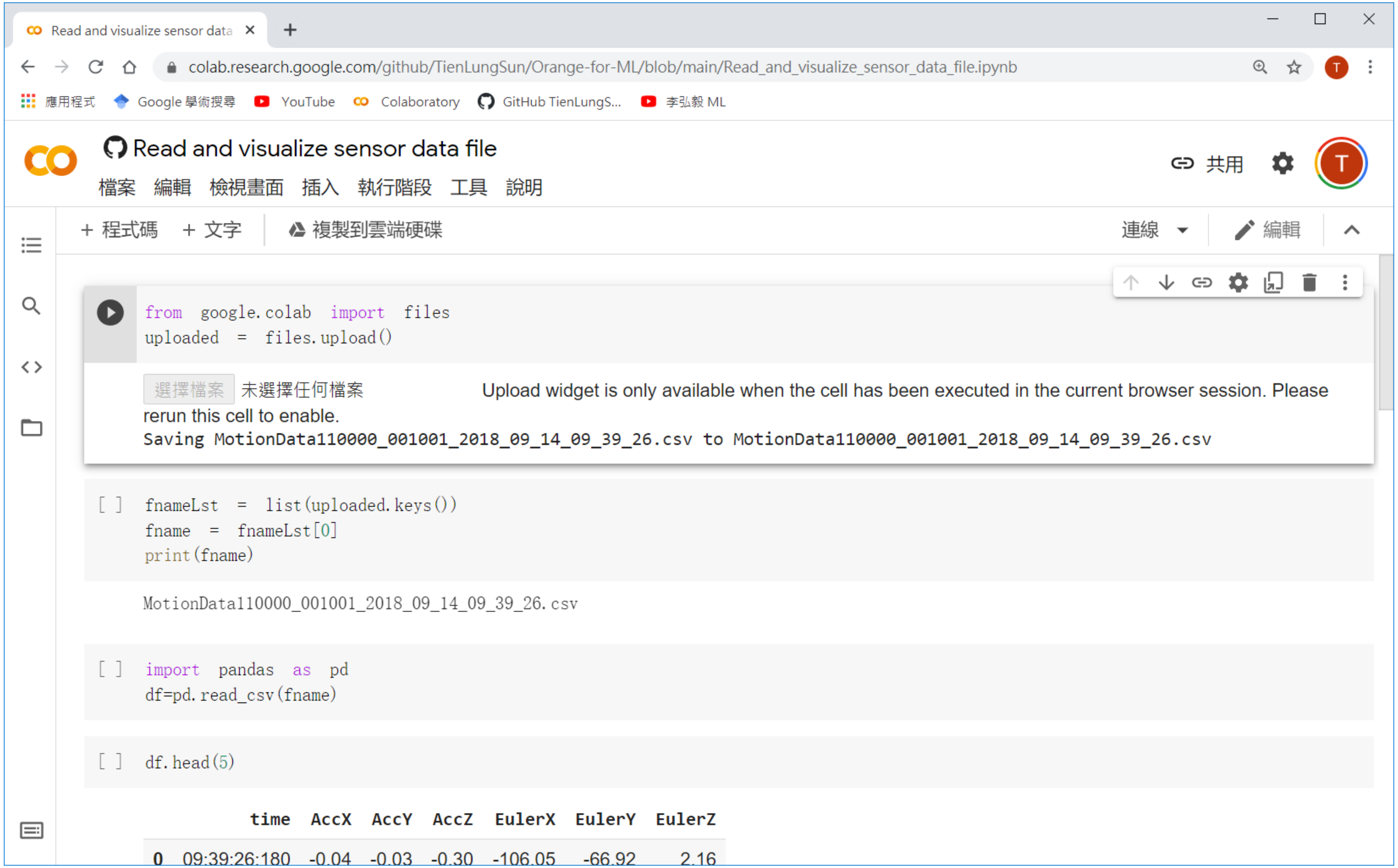

Find Open Datasets and Machi
kaggle.com/datasets?search=classification
應用程式 Google 學術搜尋 YouTube Colaboratory GitHub TienLungS... 李弘毅 ML
kaggle
Search
Sign In
Register
Home
Compete
Data
Code
Communities
Courses
More
Datasets
New Dataset
Your Work
classification
Filters
Datasets
Tasks
Computer Science
Education
Classification
Computer Vision
NLP
Data Visualization
209 Tasks
See All
Species Classification
Sachin Sharma · 48 Submissions
Iris Species
UCI Machine Learning · Usa...
Classification Problem
Pratham Tripathi · 25 Submissions
Drug Classification
Pratham Tripathi · Usability 1...
Mushroom Classification using Random Forest Classifier
Sanchita Karmakar · 34 Submissions
Mushroom Classifica...
UCI Machine Learning · Usa...
Fetal Health Classification
Larxel · 26 Submissions
Fetal Health Classific...
Larxel · Usability 10.0
3,010 Datasets
Hottest
We use cookies on Kaggle to deliver our services, analyze web traffic, and improve your experience on the site. By using Kaggle, you agree to our use of cookies.
Got it
Learn more
下午 03:26
2021/2/22

Data visualization – (3) Python coding



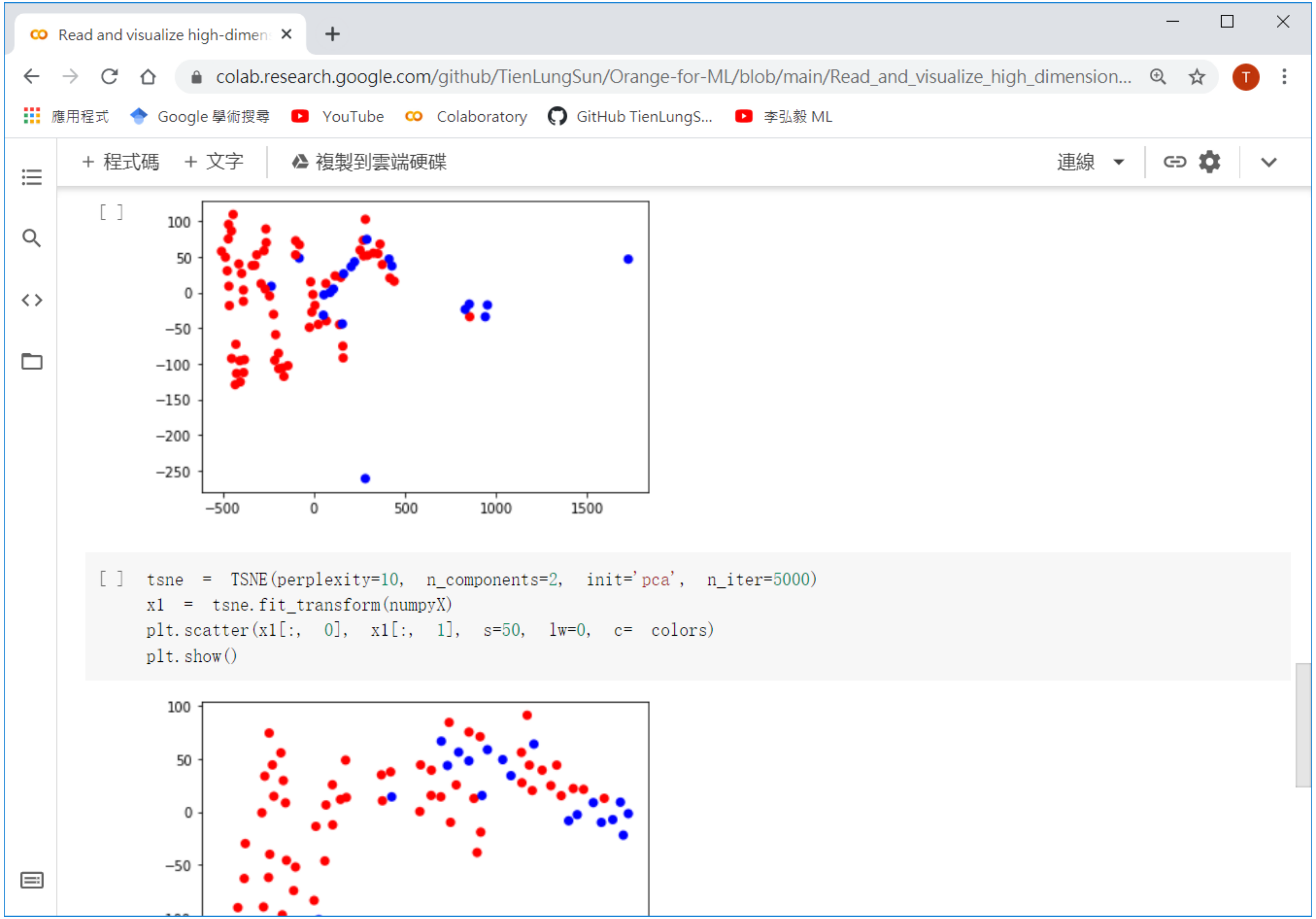
https://colab.research.google.com/

Data visualization with Python



Data visualization with Python

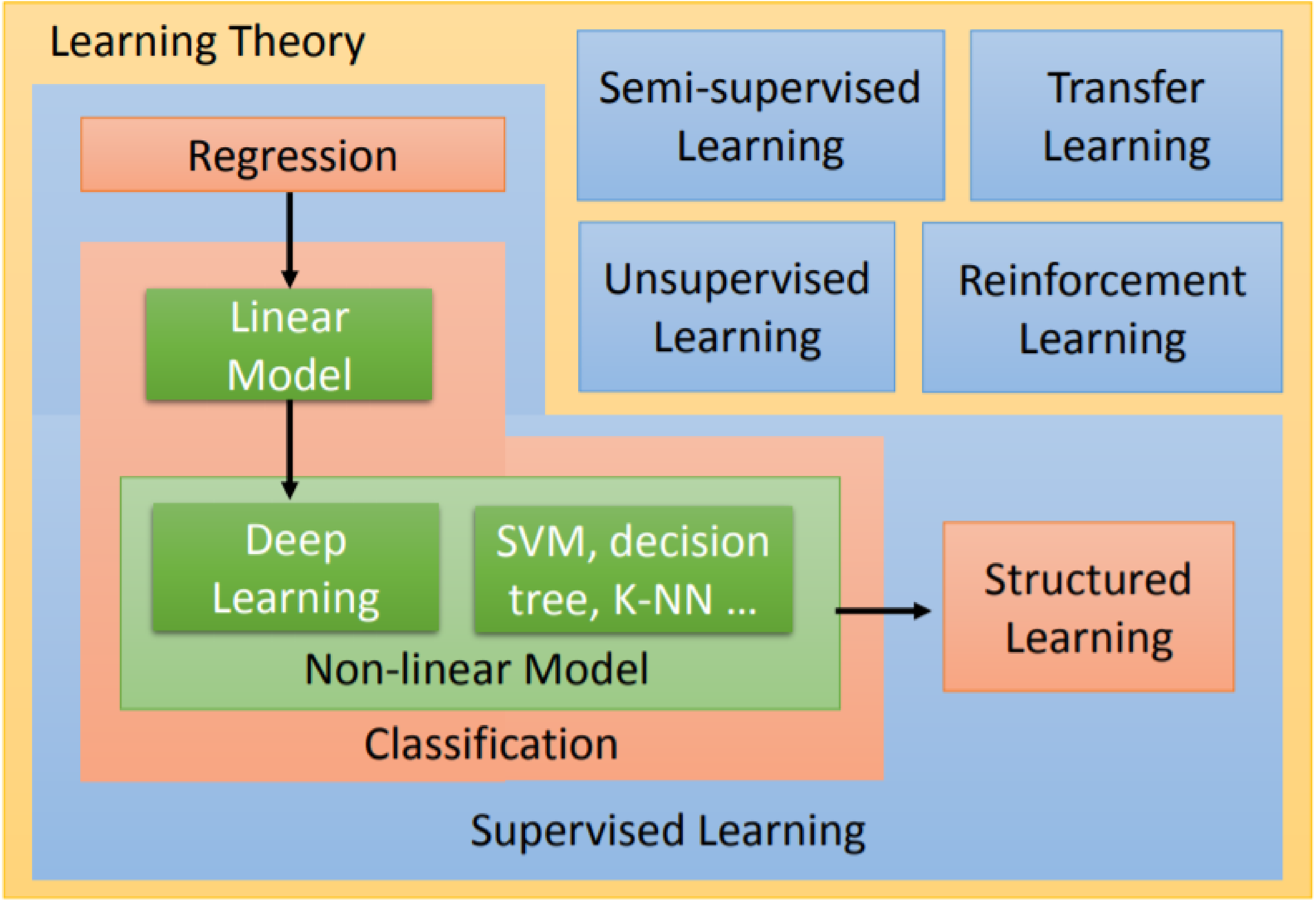
Data visualization with Python



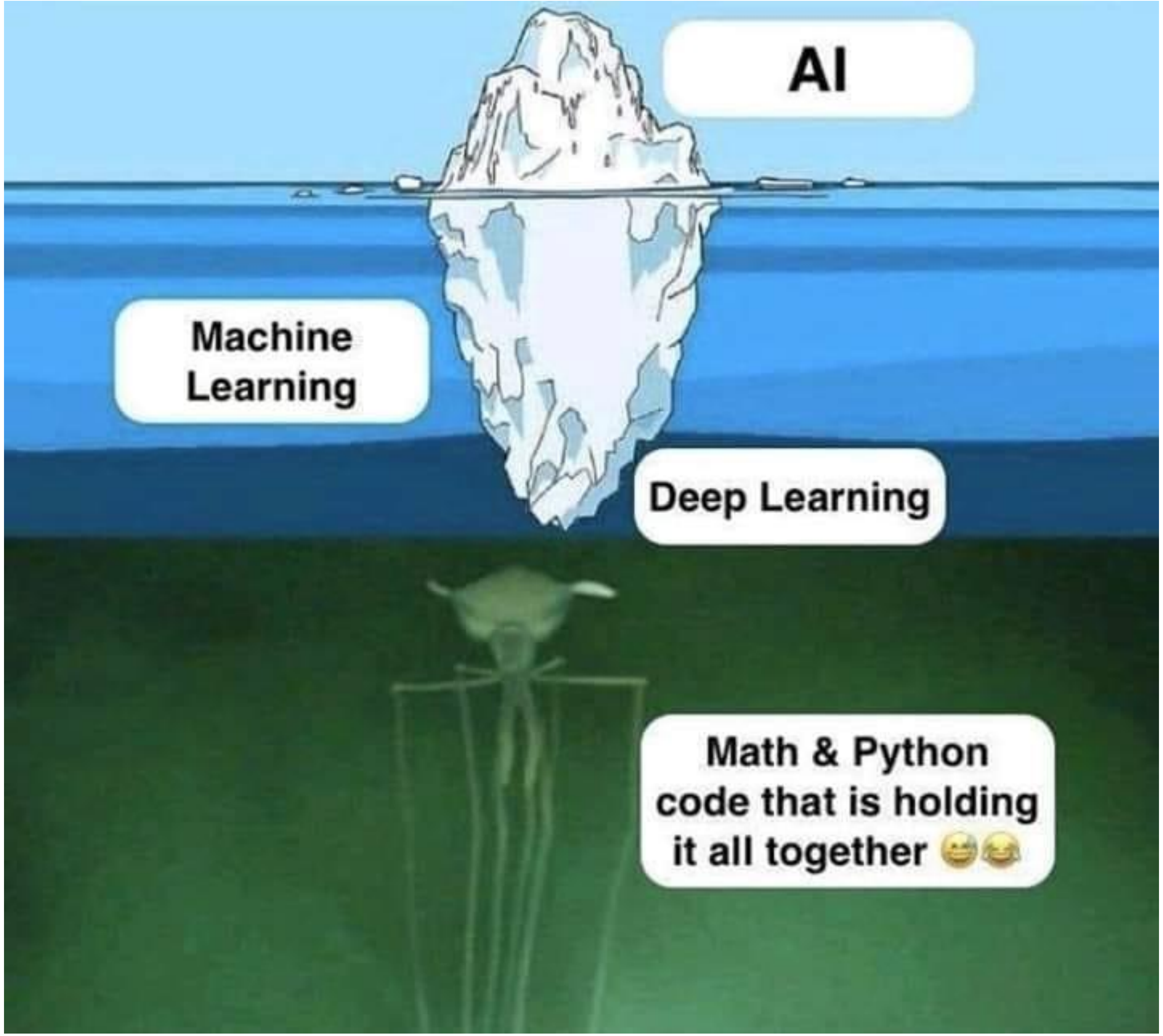
Practice – Python coding

1. Log in to Colab
2. Run python code to visualize the motion sensor data file
3. Run python code to visualize the 3M TUG data file
4. Run python code to visualize the two data files you download from Kaggle

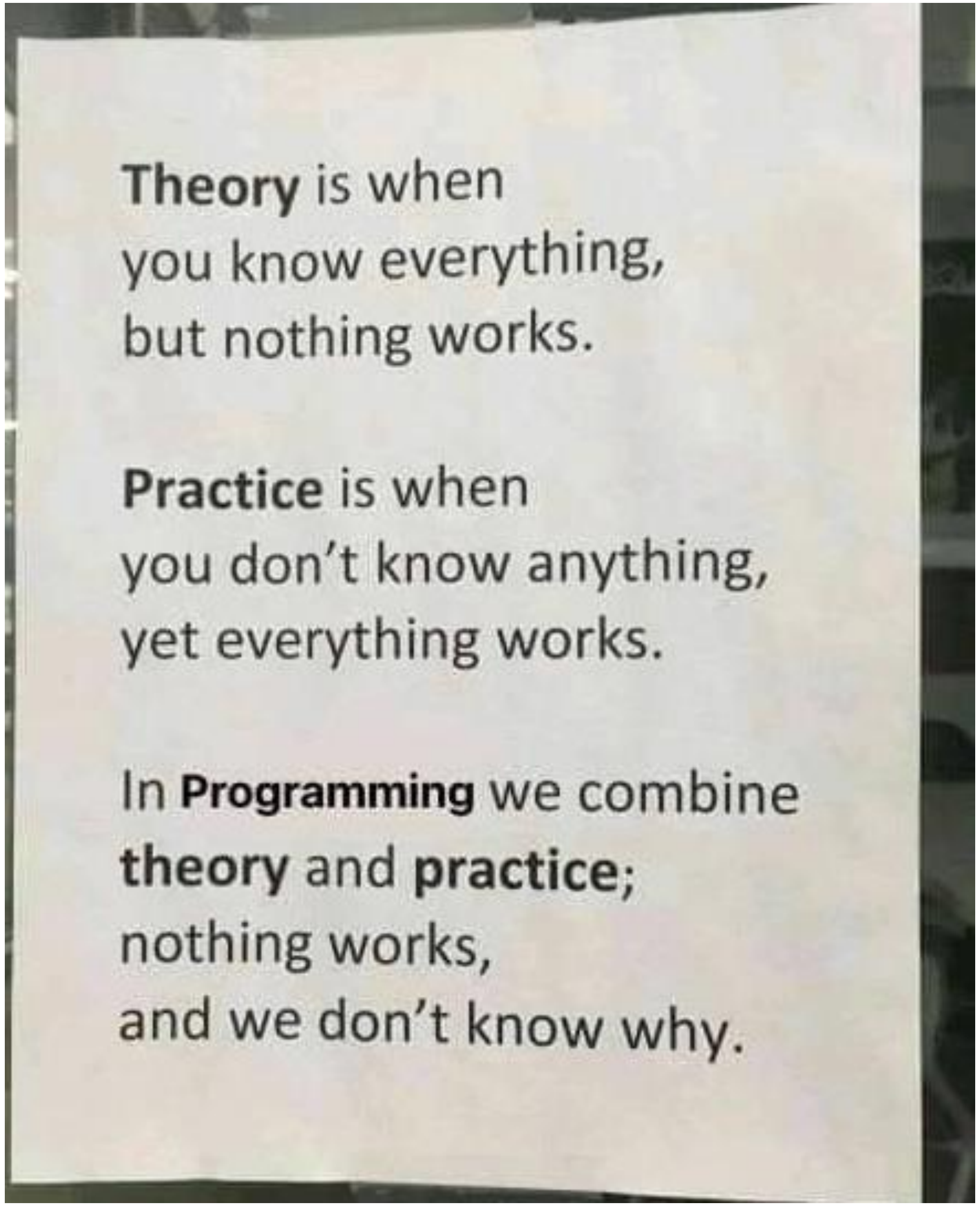
AI Learning map



Reference: 李弘毅 ML Lecture 0-1 https://youtu.be/CXgbekl66jc



https://www.facebook.com/257118421832544/photos/a.320314752179577/540486140162436/



https://www.facebook.com/113246317168557/photos/a.113263613833494/219006239925897/

Python development tools

Google colab

PYTORCH

jupyter

spyder

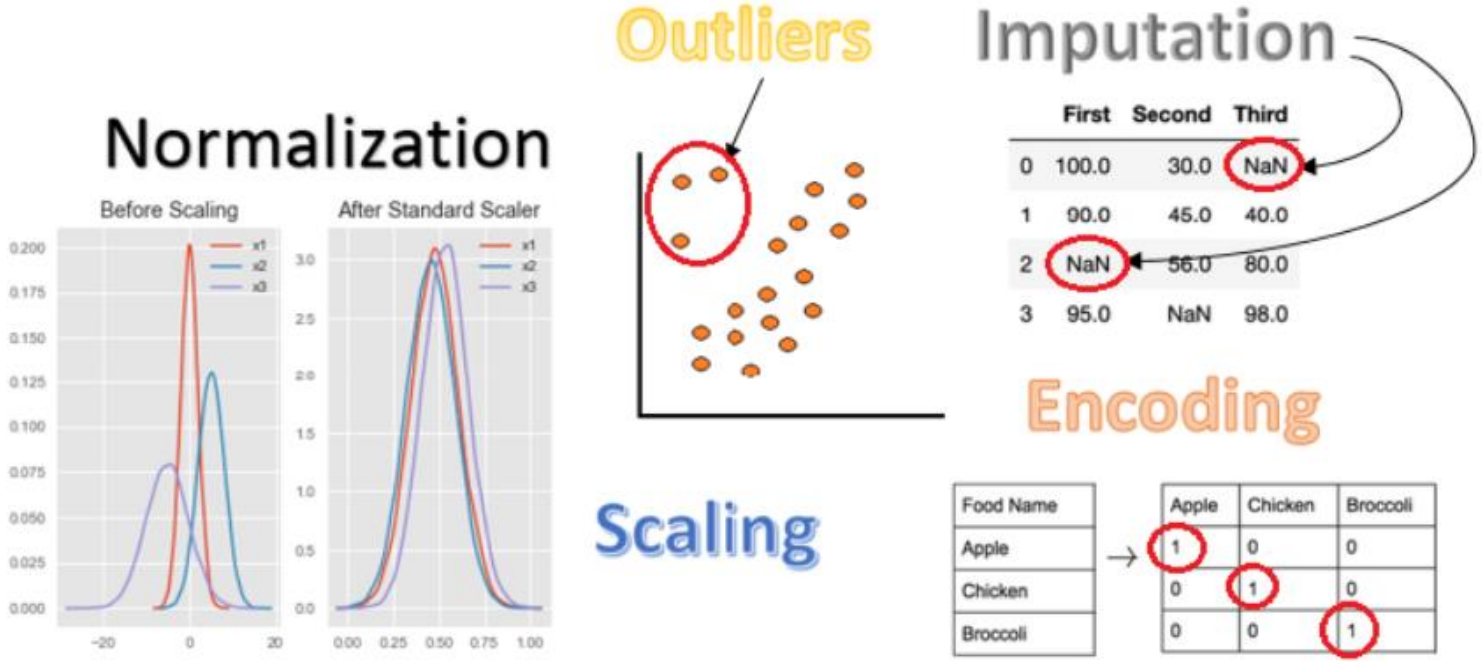
ANACONDA®

python™

Data preprocessing

Topics to be covered:

1. Standardization
2. Scaling with sparse data and outliers
3. Normalization
4. Categorical Encoding
5. Imputation



https://pub.towardsai.net/data-preprocessing-concepts-with-python-b93c63f14bb6

Data preprocessing in Python

Steps involved in data preprocessing :

1. Importing the required Libraries
2. Importing the data set
3. Handling the Missing Data.
4. Encoding Categorical Data.
5. Splitting the data set into test set and training set.
6. Feature Scaling.

https://aaaanchakure.medium.com/data-preprocessing-3cd01eefd438

Steps in Data Preprocessing in Machine Learning

1. Acquire the dataset
2. Import all the crucial libraries
3. Import the dataset
4. Identifying and handling the missing values
5. Encoding the categorical data
6. Splitting the dataset
7. Feature scaling

https://www.upgrad.com/blog/data-preprocessing-in-machine-learning/

Why Data Preprocessing?

- **Data in the real world is not clean**
  - incomplete: lacking attribute values, lacking certain attributes of interest, or containing only aggregate data, e.g., occupation=“ ”
  - noisy: containing errors or outliers, e.g., Salary=“-10”
  - inconsistent: containing discrepancies in codes or names, e.g., Age=“42” Birthday=“03/07/1997”
- **No quality data, no quality mining results!**
- **Quality decisions must be based on quality data**
  - Duplicates or missing data may cause incorrect or misleading analyses.

Yang, H. (2018). Data preprocessing.

Data Preprocessing: Major Tasks

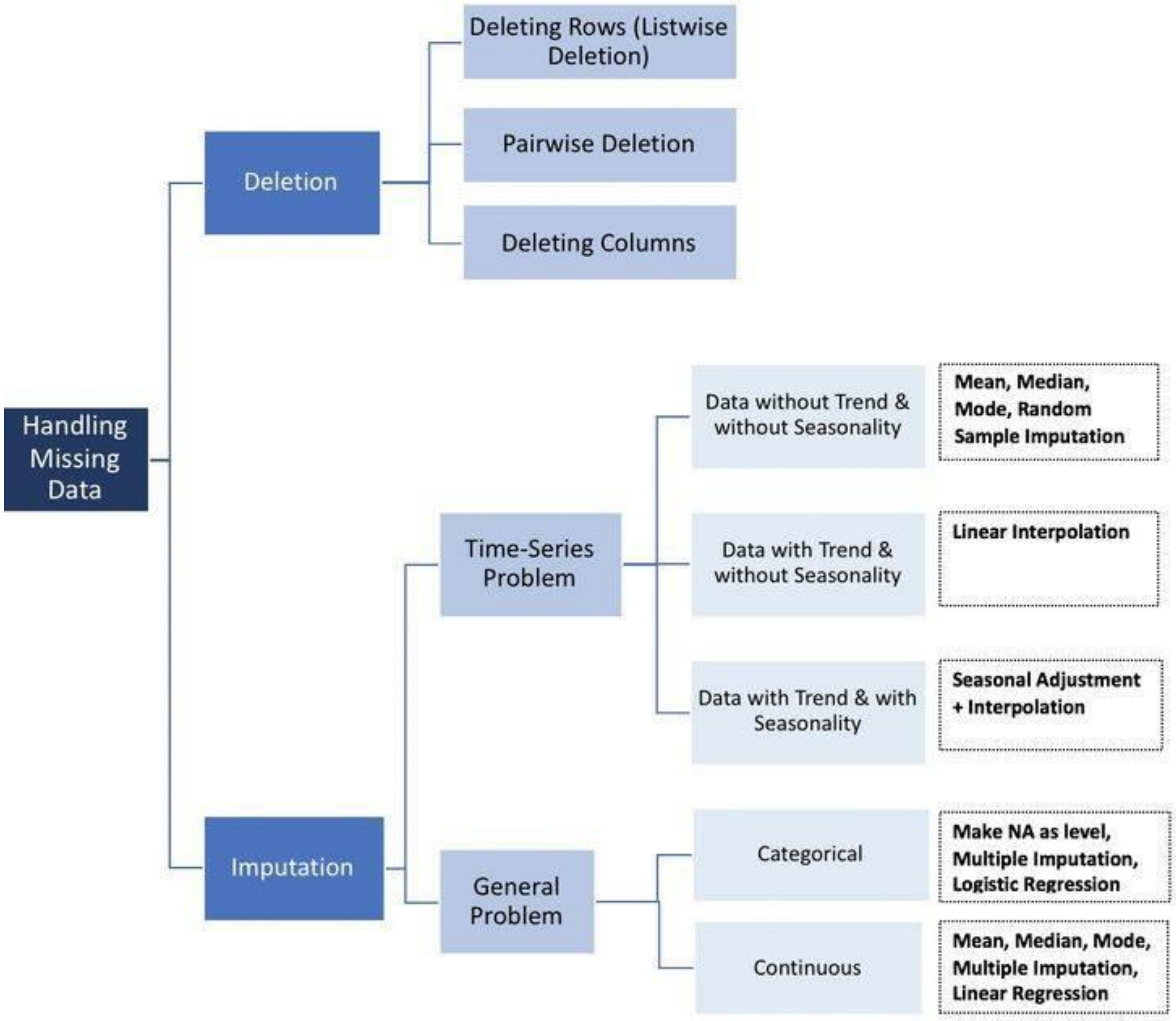
- **Data cleaning**
  - Fill in missing values, smooth noisy data, identify or remove outliers, and resolve inconsistencies
- **Data integration**
  - Integration of multiple databases, data cubes, or files
- **Data transformation**
  - Normalization and aggregation
- **Data reduction**
  - Obtains reduced representation in volume but produces the same or similar analytical results
- **Data discretization**
  - Part of data reduction but with particular importance, especially for numerical data

Yang, H. (2018). Data preprocessing.

Data Cleaning Tasks

- **Fill in missing values**
- **Identify outliers and smooth out noisy data**
- **Correct inconsistent data**
- **Resolve redundancy caused by data integration**

Yang, H. (2018). Data preprocessing.



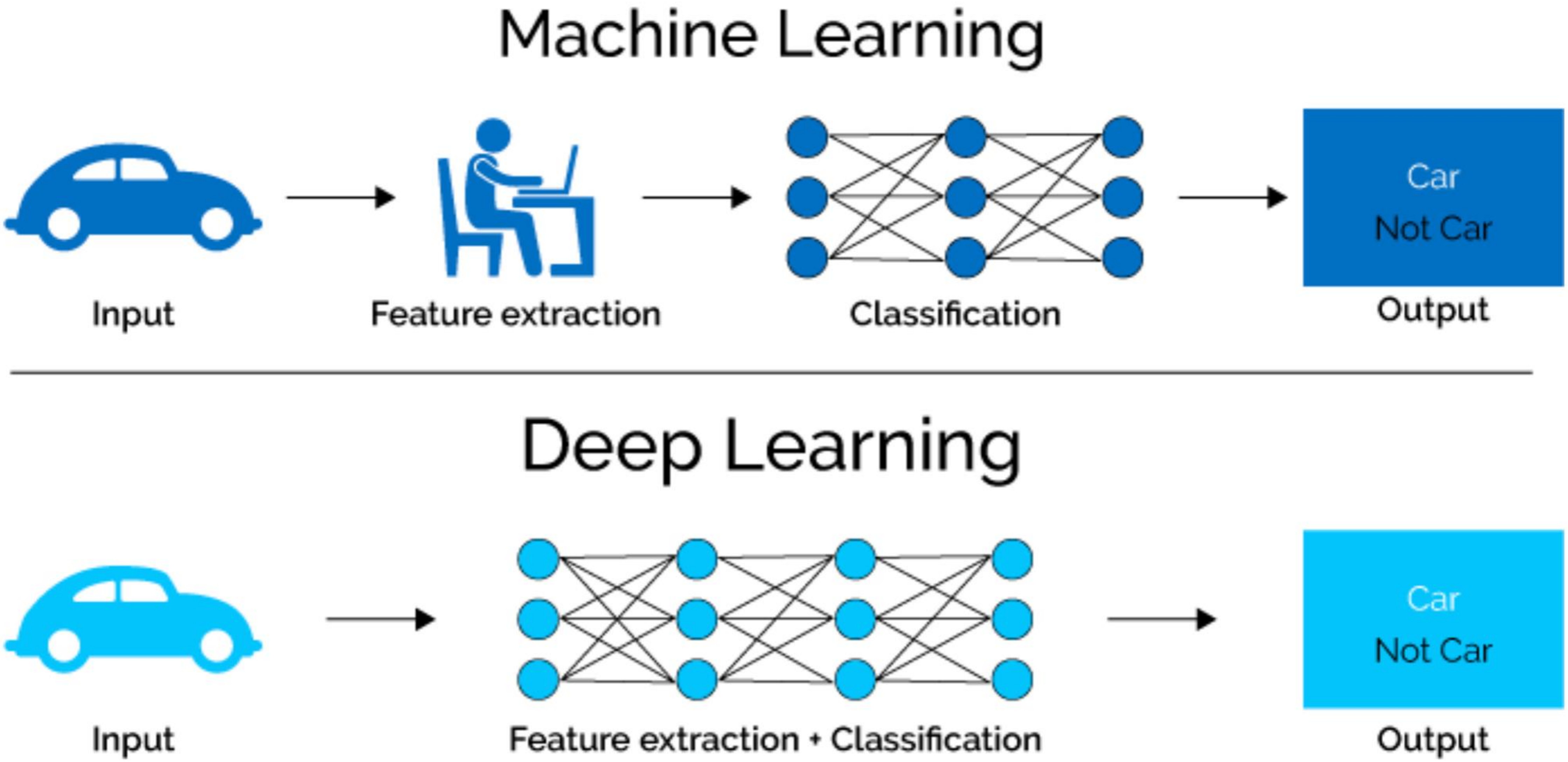
https://www.facebook.com/buggyprogrammers/photos/a.112917697219820/212476683930587/

Data Transformation

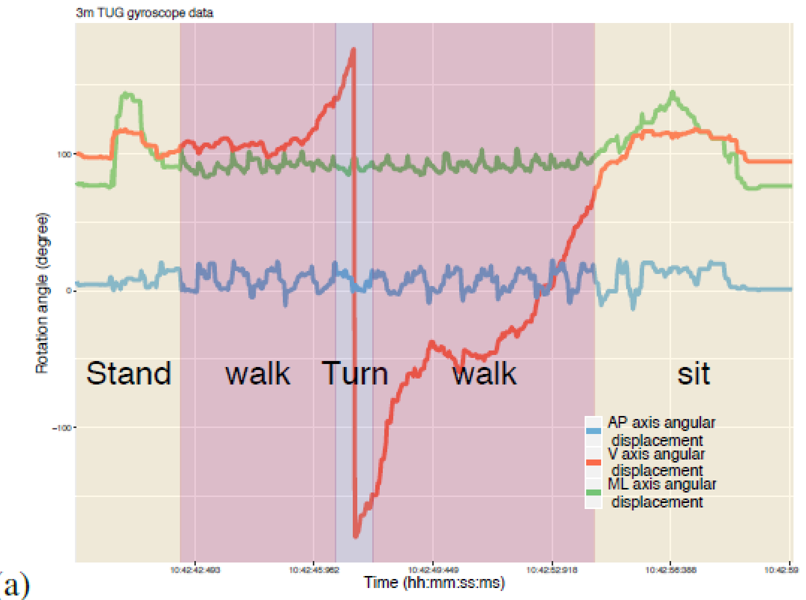
- **Smoothing: remove noise from data**
- **Aggregation: summarization, data cube construction**
- **Generalization: concept hierarchy climbing**
- **Normalization: scaled to fall within a small, specified range**
  - min-max normalization
  - z-score normalization
  - normalization by decimal scaling
- **Attribute/feature construction**
  - New attributes constructed from the given ones

Yang, H. (2018). Data preprocessing.

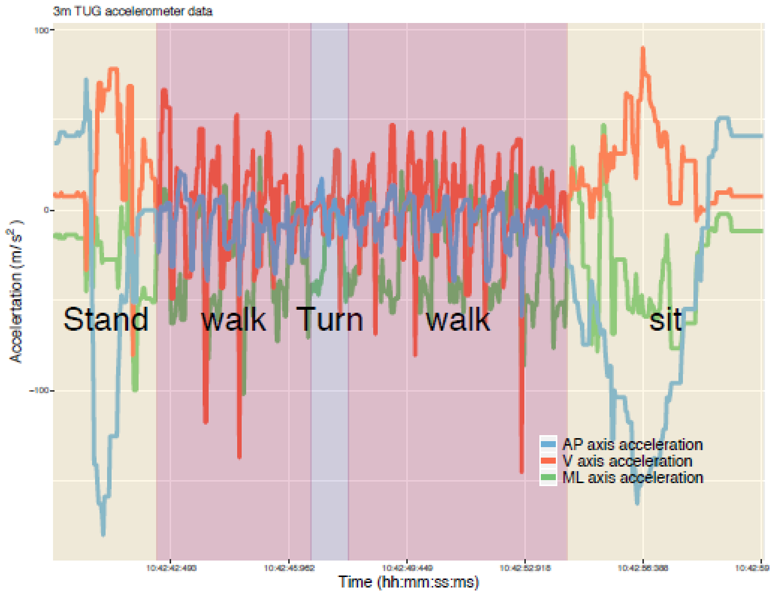
Feature engineering



https://www.facebook.com/ntirawenDL/photos/a.2280155882223083/2588834621355206



(a)



(b)

TABLE II
SUMMARY OF FEATURES

TUG sensor-based features		
Feature name	**Direction**	**Unit**
Sit to stand		
Angle range	ML	deg
Acceleration range	V, AP	m/s^2
Completion time		s
Walking		
Acceleration CV	V, AP, ML	
Acceleration range	V, AP, ML	m/s^2
Acceleration median	V, AP, ML	m/s^2
Acceleration RMS	V, AP, ML	m/s^2
Angular velocity CV	V, AP, ML	
Angular velocity range	V, AP, ML	deg/s
Angular velocity median	V, AP, ML	deg/s
Angular velocity RMS	V, AP, ML	deg/s
Speed		m/s
Turning		
Acceleration CV	V, AP, ML	
Acceleration range	V, AP, ML	m/s^2
Acceleration median	V, AP, ML	m/s^2
Acceleration RMS	V, AP, ML	m/s^2
Angular velocity CV	V, AP, ML	

Hsu, Y. C., Zhao, Y., Huang, K. H., Wu, Y. T., Cabrera, J., Sun, T. L., & Tsui, K. L. (2020). A novel approach for fall risk prediction using the inertial sensor data from the timed-up-and-go test in a community setting. IEEE Sensors Journal, 20(16), 9339-9350.

- Supervised Feature Selection Methods
  - Wrapper Feature Selection Methods
  - Filter Feature Selection Methods
  - Embedded or Intrinsic Feature Selection Methods
- Feature Selection with Statistical Measures
  - Univariate Feature Selection
- Feature Selection Strategies
  - Selection Method
  - Transform Variables
- Which Feature Selection Method is the Best?
- Feature Selection Implementations
  - Feature Selection For Regression models
  - Classification Feature Selection

https://dataaspirant.com/feature-selection-methods-machine-learning/?fbclid=IwAR3H0DE-No4A8R_wDY6gIaTQ00uaa7RVZdcsdI1nU1UhqvsrqkveSugvbFA