

Python Fundamental

Dr. Steve Lai

2021/08

Outline for The Four Time Slots

1. Basic python
 - Python fundamental
 - Data visualization modules
 - Data analysis modules
 - Financial graphics
2. API & Machine learning
 - API demo
 - Machine learning introduction and processing
 - Sci-kit learn
3. Financial model Machine learning Project
 - Financial Models
 - Case Studies
 - House Prices: Advanced Regression Techniques
 - Home Credit Default Risk
4. Python crawling
 - beautifulshop
 - selenium
 - ptt crawling demo

Quizzes

- What is the maximum value of Int in python?
 1. $2^{64} - 1$
 2. $2^{32} - 1$
 3. Infinity

Quizzes

- The BIF `id()` return the identity of an object which the identity is the address of the object in memory. Read the following program

```
1 a = 1
2 id1 = id(a)
3 a = 2
4 id2 = id(a)
5 print(id1 == id2)
```

- What is the output?
 1. An address number
 2. True
 3. true
 4. False
 5. false
 6. Cause an error

Quizzes

- What is the output of the following program?

```
1 print = 3
2 print += 3
3 print(2)
```

- 2
- 3
- 6
- Cause an error

Basic Data Type in Python

- int
- float
- bool
- str

Control Flow

- if-elif-else
- Loop
 - while loop
 - foreach loop

Basic Data Structure

- tuple
- list
- dict
- set

Programming Structure

- Function
- Class

Quizzes

- What is benefit of using function?
 - Easy to read
 - Easy to debug
 - Easy to use

Quizzes

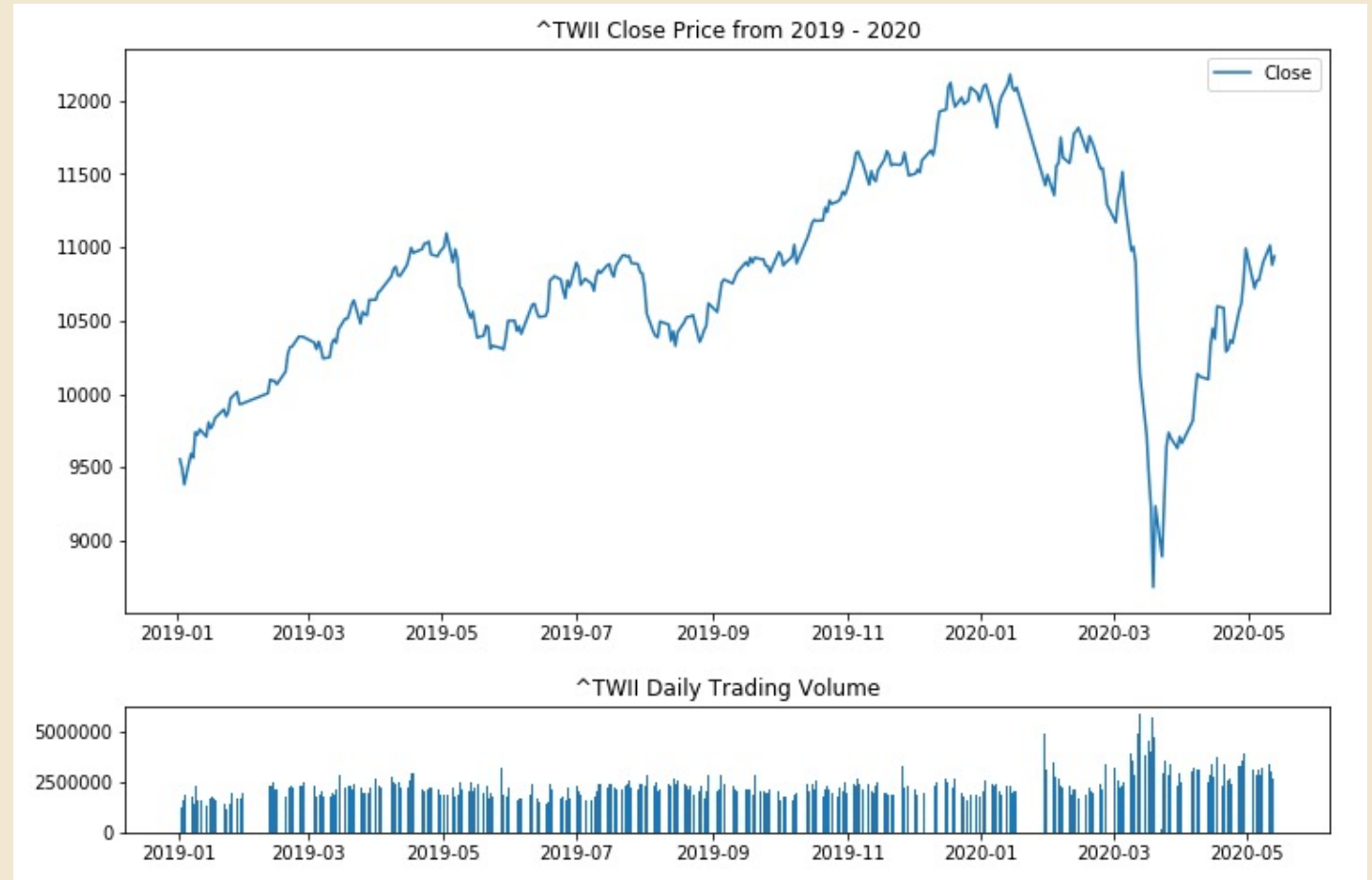
- Which of the following are synonyms?
 1. instance
 2. attributes
 3. method
 4. object
 5. function
 6. variable
 7. class

Data Analysis Modules

- Numpy
- Pandas
- Exploratory data analysis

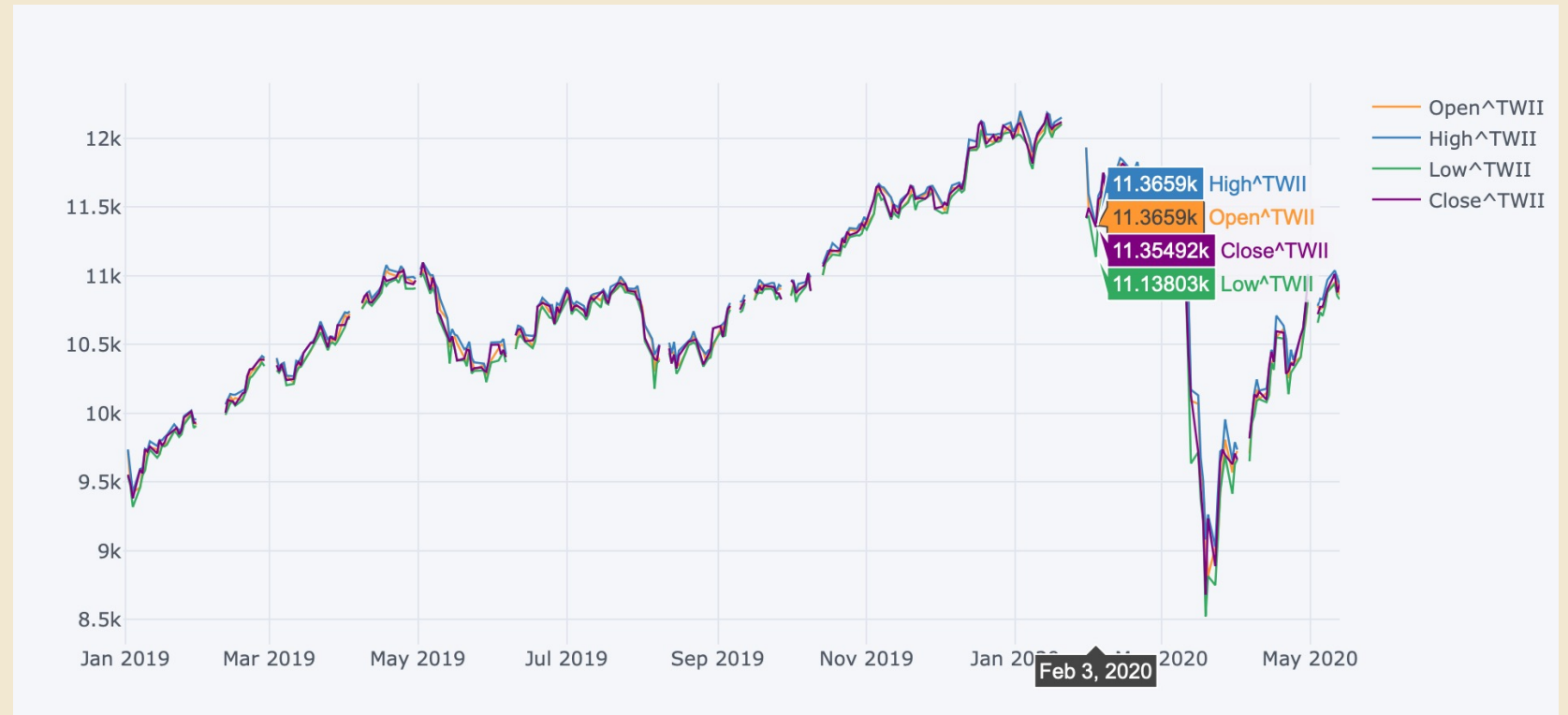
Sample Data Source

- TWSE
- TAIFEX
- Quandl
 - Need sign up an account
- pandas_datareader
 - yfinance



Data Visualization

- Requirement packages
 - Matplotlib
 - Seaborn
 - Plotly
- Graphic Types
 - plot
 - bar
 - histogram
 - scatter
 - pie
 - box
 - polygon



Financial Graphics

- Requirement packages
 - Cufflinks
 - chart_studio
 - mplfinance
- Graphics Types
 - Candlestick chart
 - Bollinger bands
 - Relative strength indicator
 - Moving average convergence divergence



Financial Time Series

- Summary Statistics
 - `df.describe()`
 - mean, min, max, std, quartile
- Time difference
 - `df.diff()`
 - `df.pct_change()`
- Time resample
- Long/Short simple moving averages
 - `df.rolling(window=5, min_periods=5).mean()`
 - `df.rolling(window=30, min_periods=30).mean()`
- Exponential moving averages
 - `df.ewm(span=5).mean()`
 - `df.ewm(span=30).mean()`
- Correlation Analysis
 - `df.pct_change().corr()`

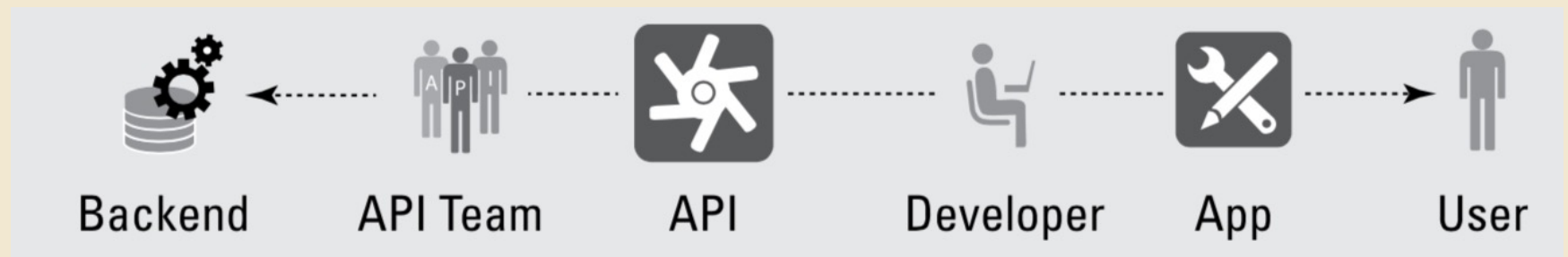
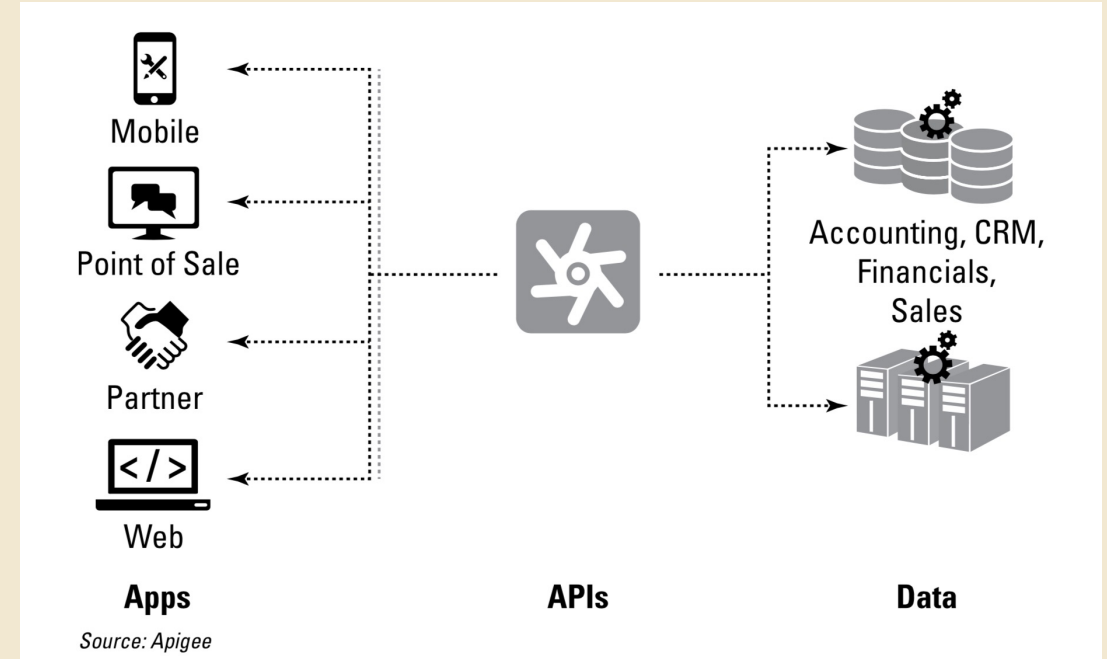
Python with API

Dr. Steve Lai

2021/08

Connecting to The Application

- API
 - Application Programming Interface
 - Libraries, CSV, JSON, XML, ...



JSON

- What is the following similar with JSON Object in Java?
 1. Dictionary
 2. Set
 3. Tuple
 4. List
- What is the following similar with JSON Array in Java?
 1. Array
 2. Set
 3. Map
 4. StringBuilder

What is JSON?

- JSON
 - JavaScript Object Notation
- JSON Object
 - Key-Value pairs
- JSON Array
 - Sequential data with order
- `import json`
 - `load`, `loads`
 - `dump`, `dumps`



CSV

- `import csv`

APIs

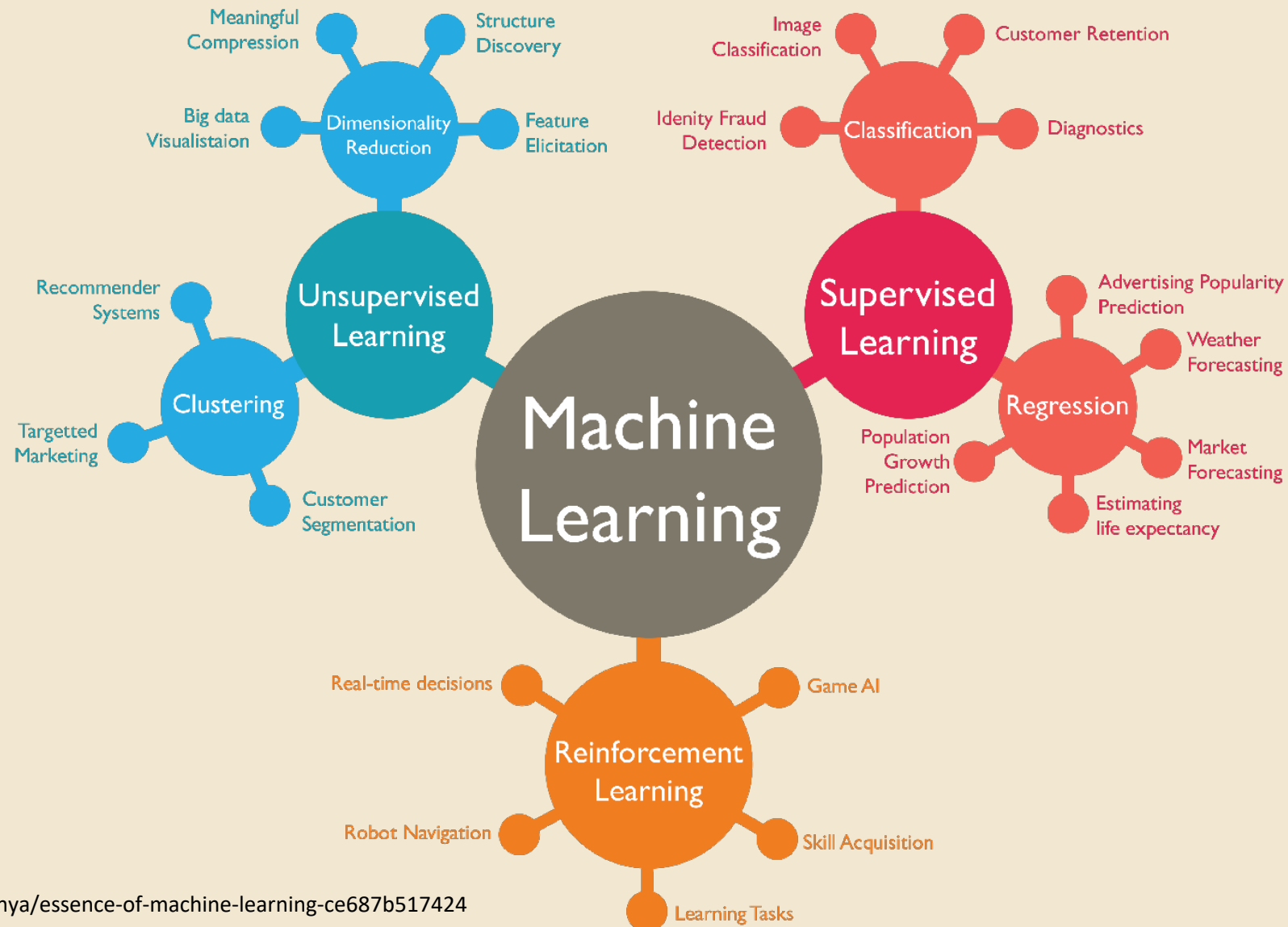
- <https://api.github.com/search/repositories?q=language:python&sort=stars>
- <https://www.travel.taipei/open-api/zh-tw/Attractions/All?categoryIds=12&page=1>

Machine Learning

Dr. Steve Lai

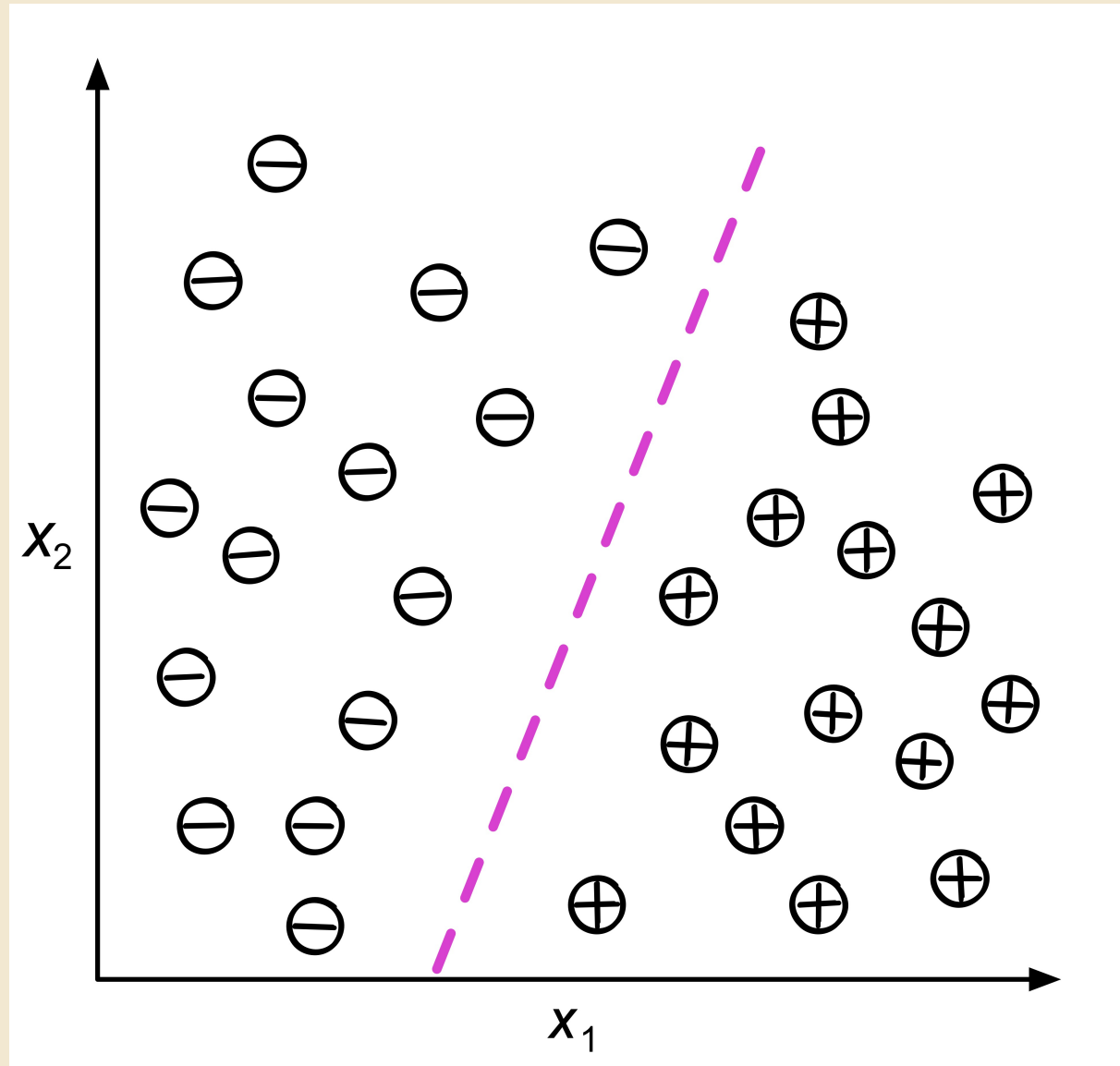
2021/08

Machine Learning

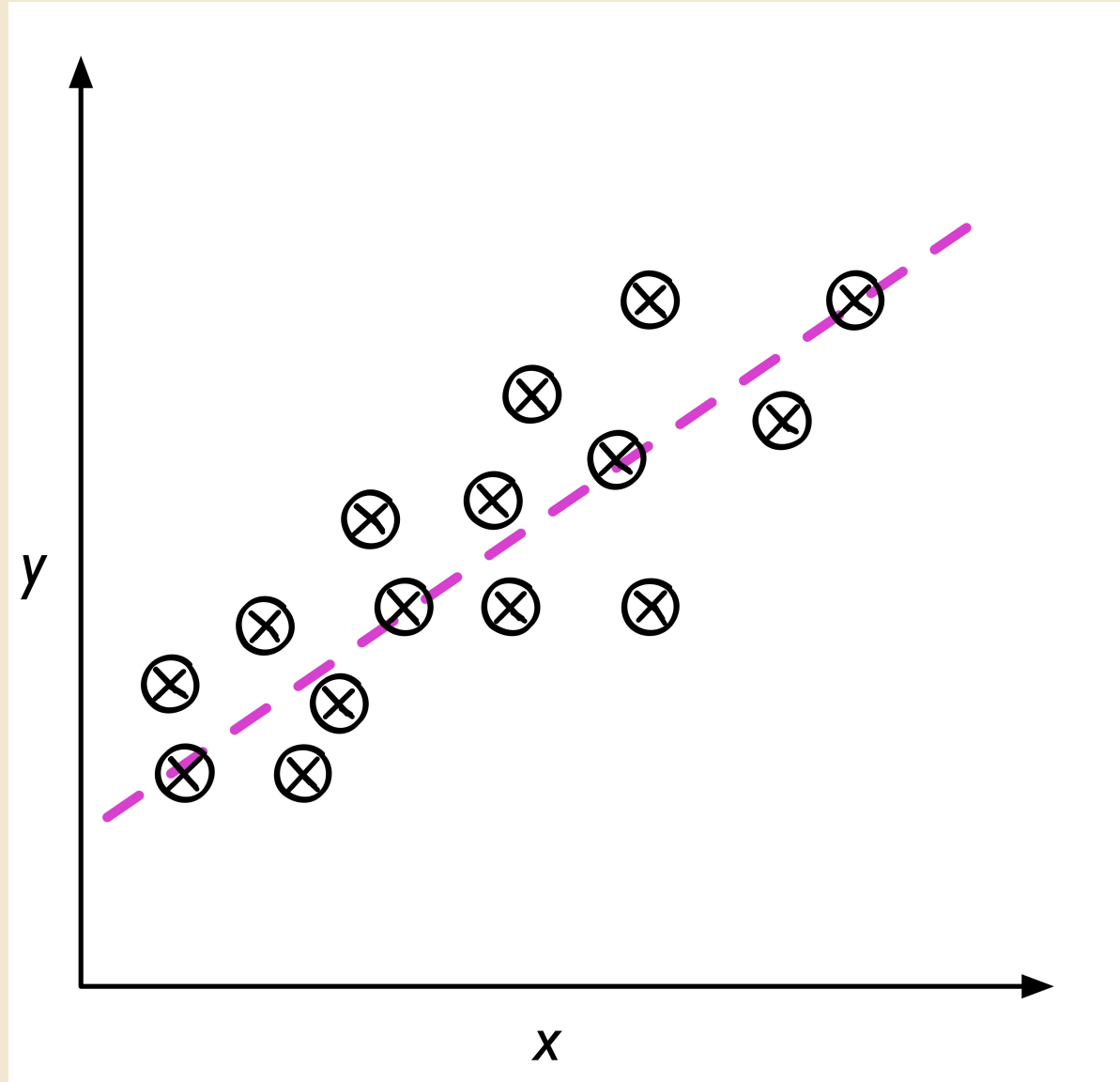


<https://medium.com/analytics-vidhya/essence-of-machine-learning-ce687b517424>

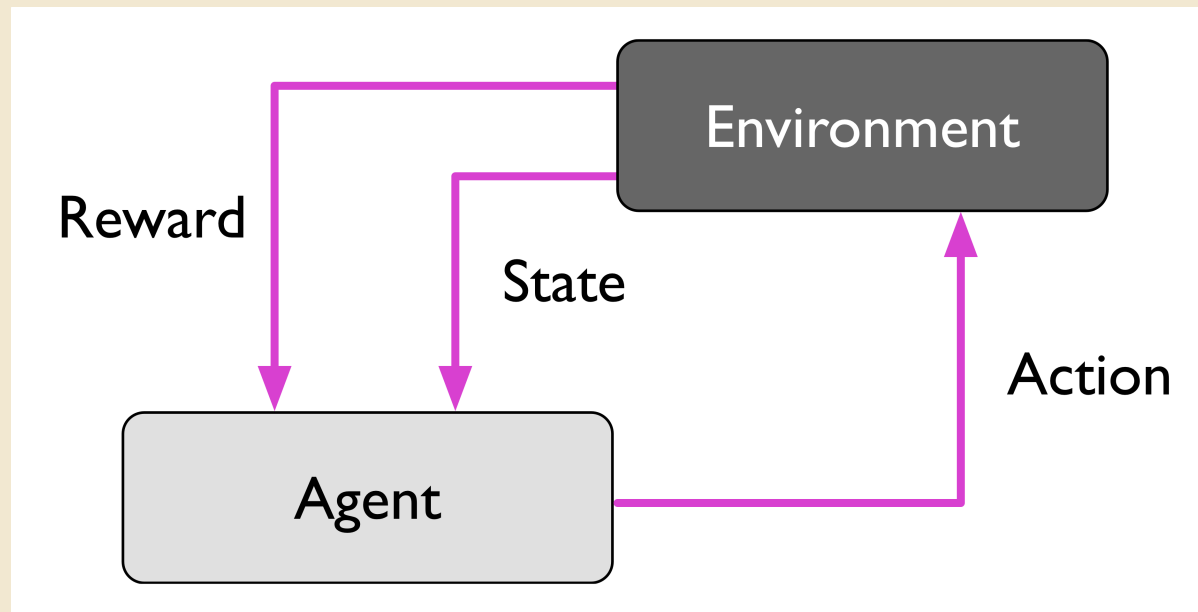
Classification



Regression



Reinforcement Learning



7 Steps to Machine Learning



[Google Cloud Platform](https://www.youtube.com/watch?v=nKW8Ndu7Mjw)

<https://www.youtube.com/watch?v=nKW8Ndu7Mjw>

The 7 steps of machine learning (AI Adventures)

<https://medium.com/dataseries/7-steps-to-machine-learning-how-to-prepare-for-an-automated-future-78c7918cb35d>

Financial Models

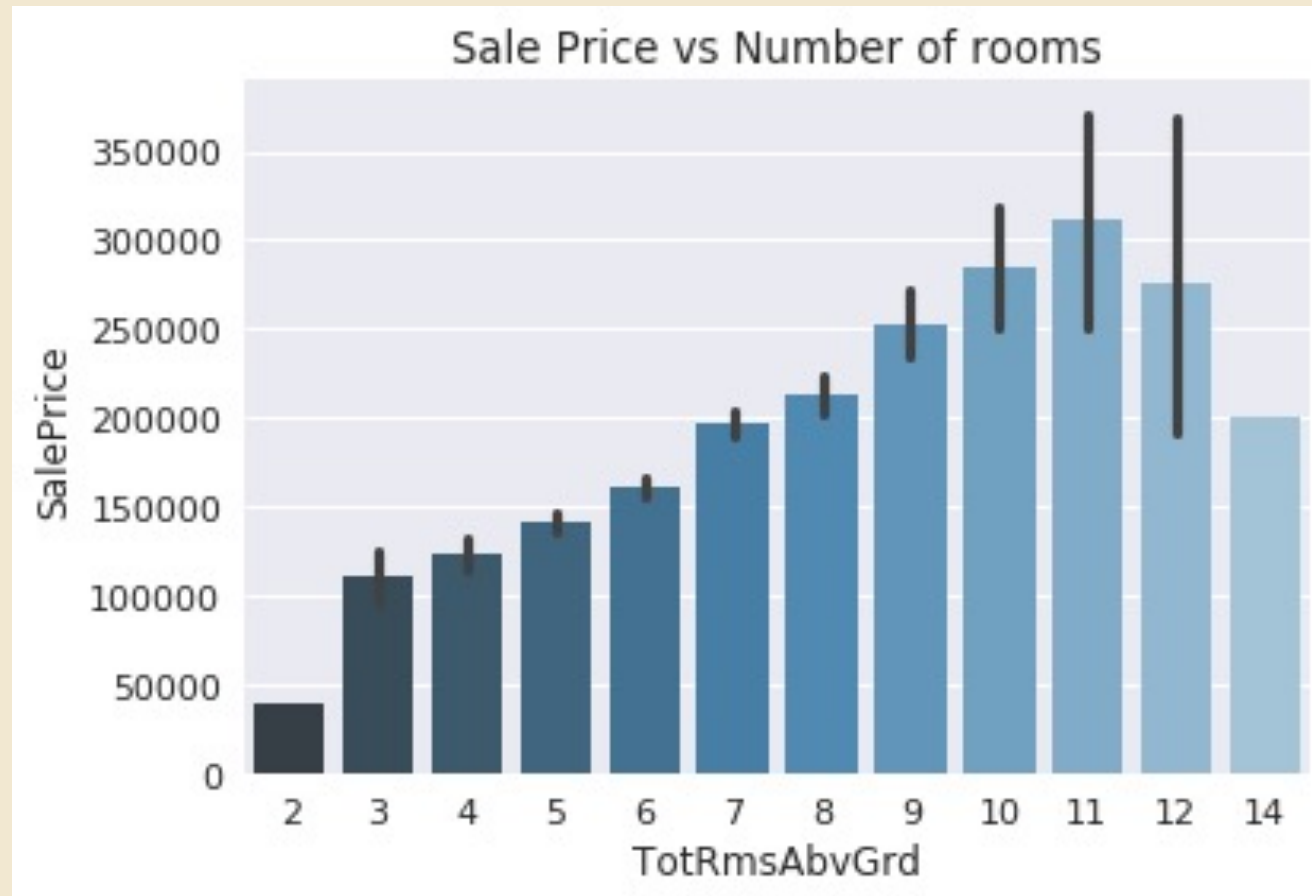
- Linear models
- Nonlinear models
 - Implied volatility model
 - Markov switching model
 - Threshold model
 - Smooth transition model

Minimizing Risks for Loan Investments

- Higher credit scores (more trustworthy and less risky) get lower interest rates for their loans
 - lower credit scores (less trustworthy and more risky) get higher rates.
- However, the loans with higher interest rates are more attractive because they provide higher return on investment (ROI)
 - They pose risks of being not returned at all.
- The machine learning model that could predict which of the high interest loans are more likely to be returned, would bring added value by minimizing the associated risks.
- <https://www.kaggle.com/wordsforthewise/lending-club>

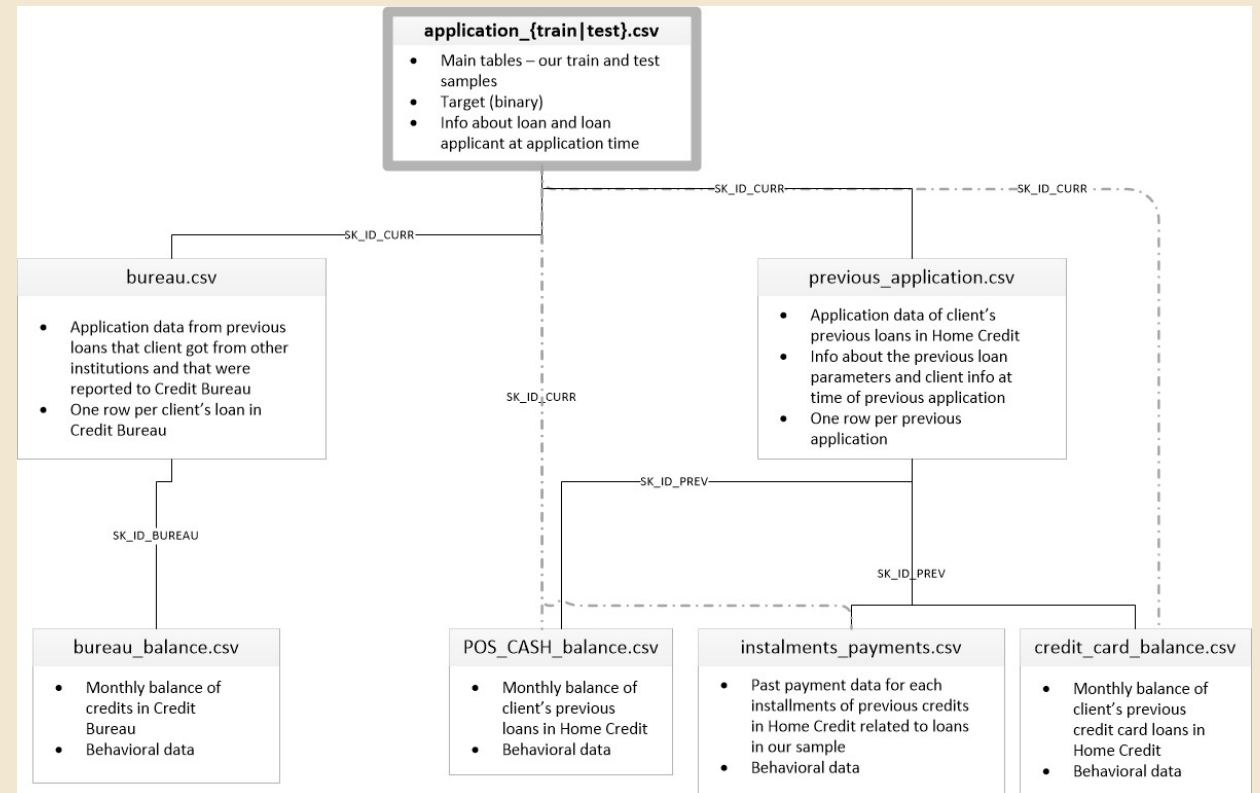
Case Study: House Prices

- Kaggle
 - House Prices: Advanced Regression Techniques
 - <https://www.kaggle.com/c/house-prices-advanced-regression-techniques>



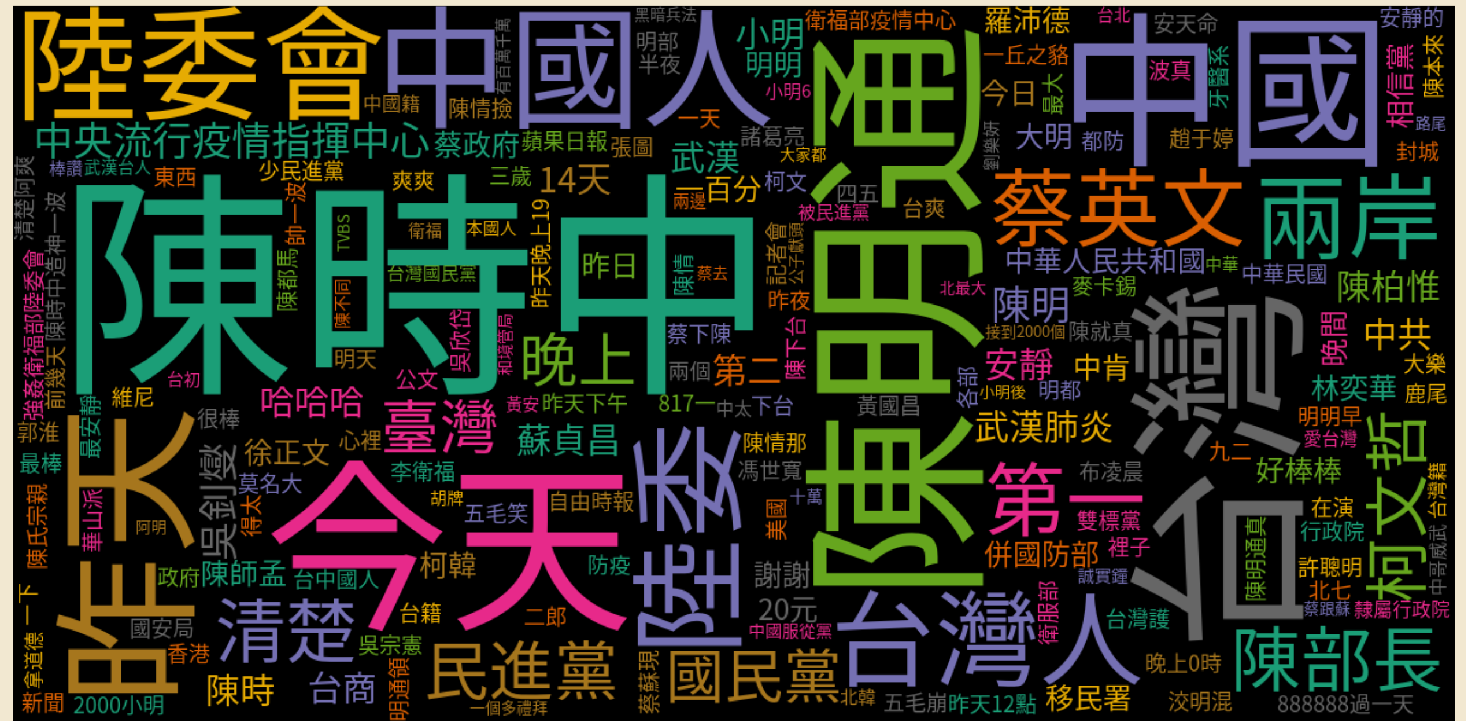
Case Study: Home Credit Default Risk

- Kaggle
 - Home Credit Default Risk
 - <https://www.kaggle.com/c/home-credit-default-risk>



PTT Demo

- PTT crawling
- NLP



Code Reference

- https://github.com/lzrong0203/fin_ios_python