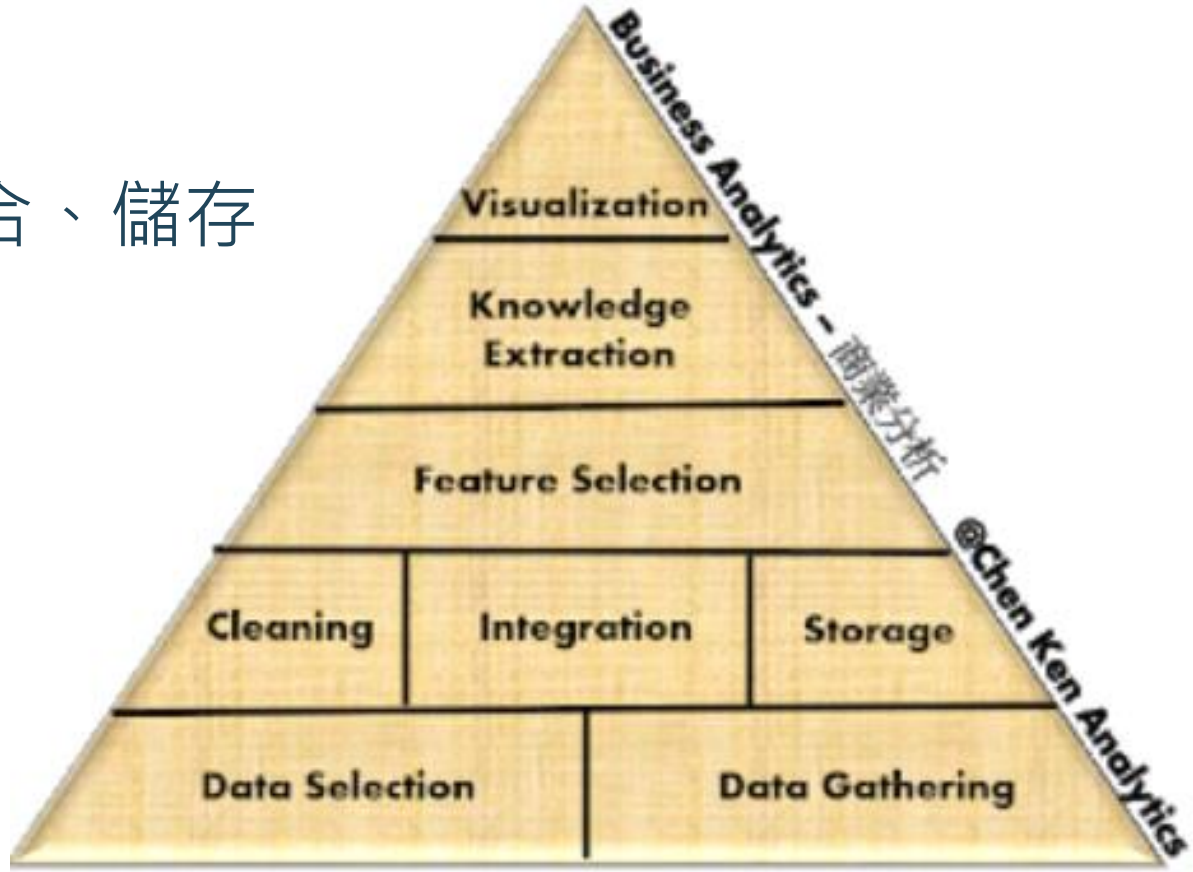
SCU LawTech

Class #09

LAWTECH
learn coding get outstanding

80 – 20 法則

數據分析的 80 % 成本：
資料的選擇、定義、蒐集、清理、整合、儲存

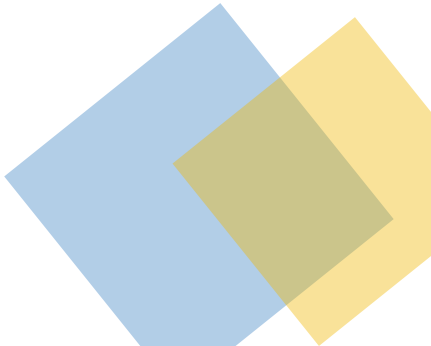


https://aifreeblog.herokuapp.com/posts/10/data_science_105/

EDA（Exploratory Data Analysis）

探索式資料分析，運用視覺化、基本的統計「看」一下資料，以期進行複雜或嚴謹的分析之前，能夠對資料有更多的認識

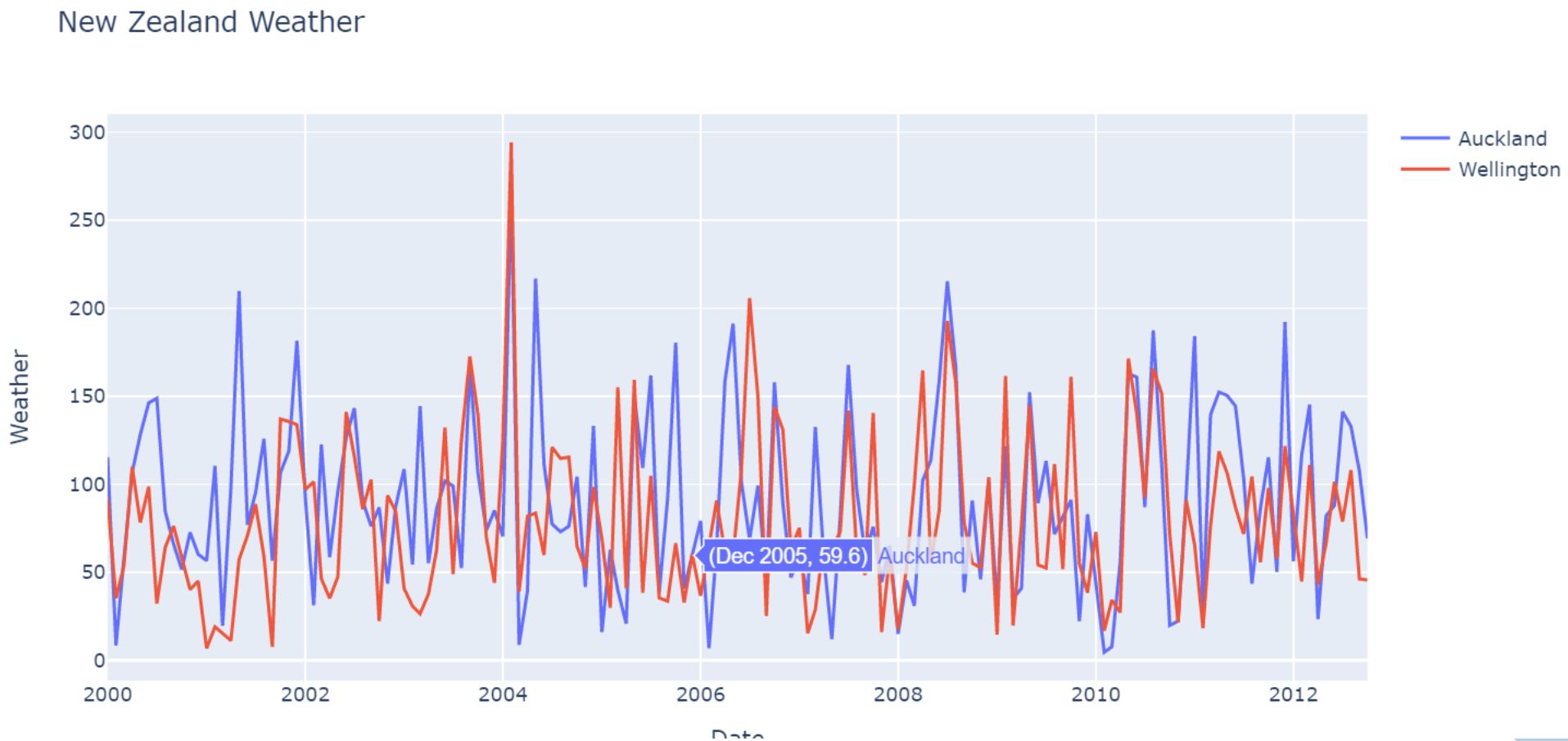
1. 了解資料，獲取資料的資訊、結構和特點
2. 查看資料是否有誤
3. 分析個變數間的關聯性，找出重要的變數



統計圖表

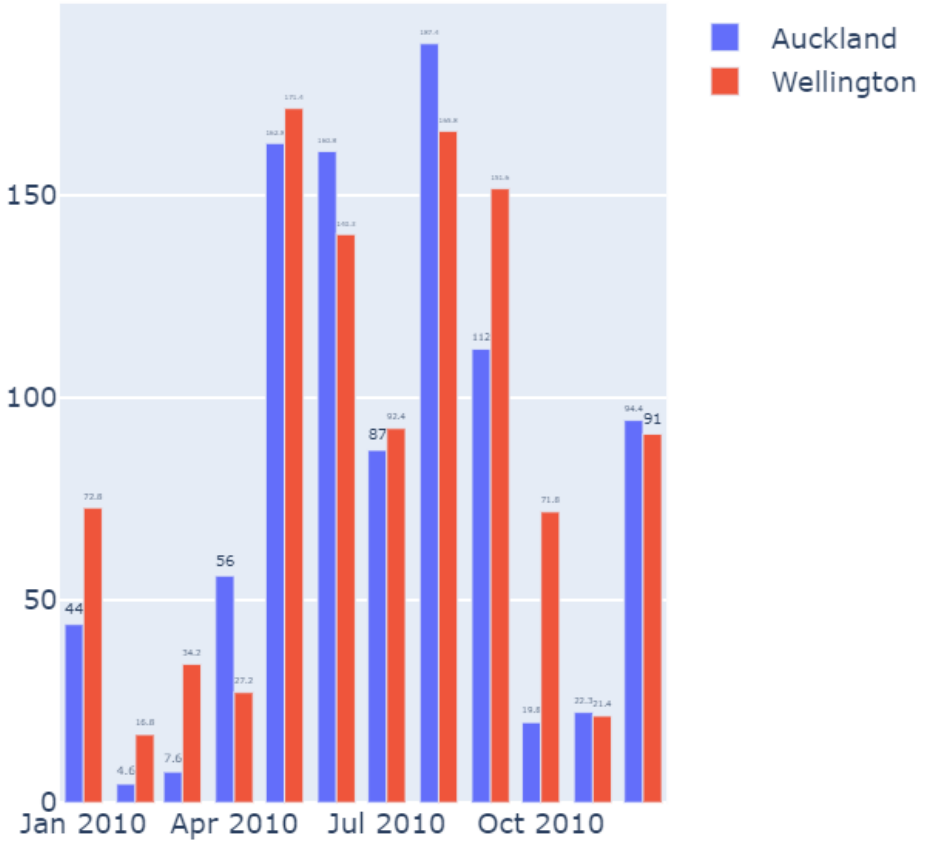
Line Plot

走勢圖，可以表時間性的資料



Bar Plot

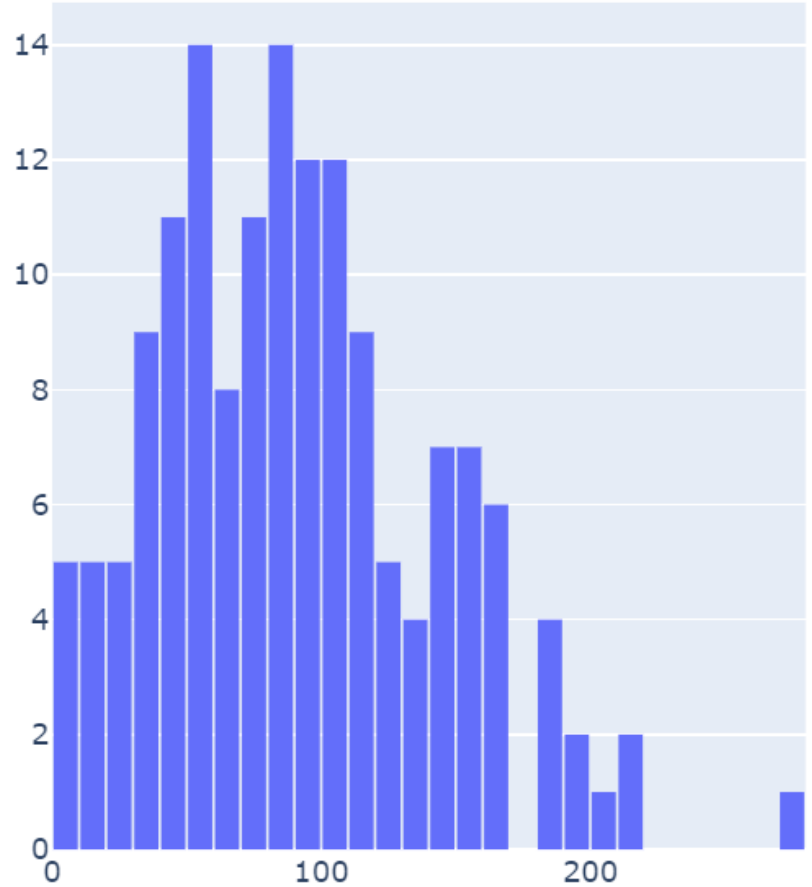
長條圖，以住狀的長短、高度或數值來表示各個類別的次數



Histogram

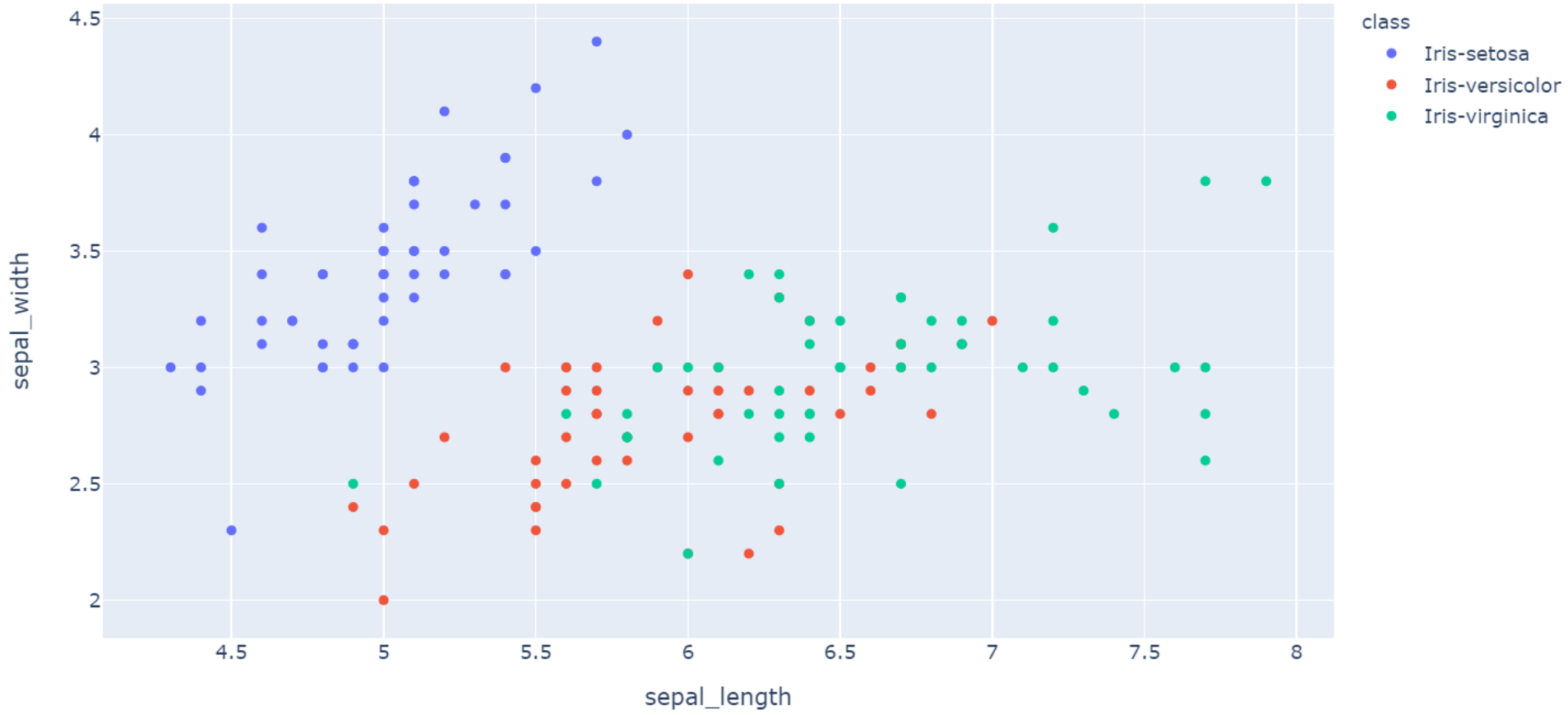
直方圖，表示次數分配的長方形圖

- X 軸為連續型的資料，以組界表示

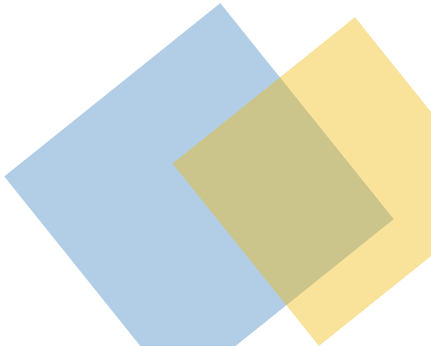


Scatter plot

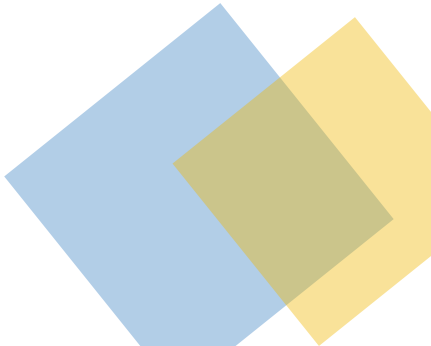
散佈圖，以點表示，可以看出資料在不同維度上的資料



程式碼示範



如果處理的資料是文字、文章，要怎麼下手？



RegEx

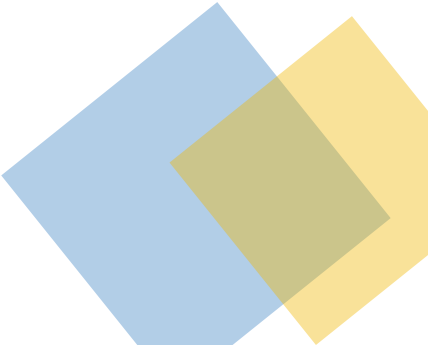
Regular Expression 介紹

正規表示式

是一種表達「具有某種特徵」字串的方式，可以用來完全指定需要加以處理的資料，避免反覆判斷找尋的困擾

Example：抓出地號

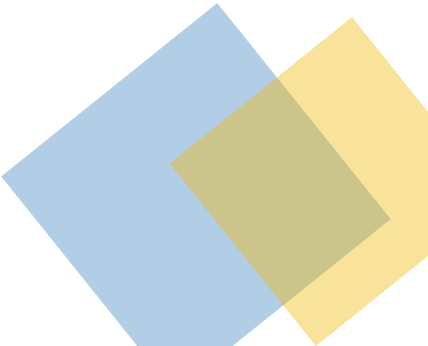
- 桃園市8德區中華段40地號 -> 40
- 彰化縣花壇鄉花壇段1418-0000地號 -> 1418-0000
- 臺中市南屯區埔興段35-12地號 -> 35-12
- 桃園市蘆竹區內興段32地號 -> 32
- 桃園市楊梅區大金山下段月眉山下小段1地號 -> 1



Regular Expression 組成

為了表達「特徵」需要定義範本（Pattern）

- 普通字元（ASCII）
- 特殊字元（Metacharacter）
- 數量定義詞（Quantifier）

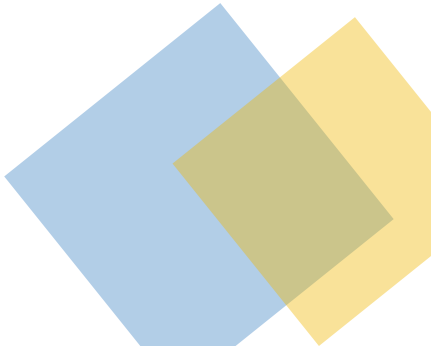


Regular Expression 組成

普通字元（ASCII）

字元本身、字串

e.g. “A” 對應到字串就是” A” 或” a”



Regular Expression 組成

數量定義詞 (Quantifier)

定義前一個字元的數量

Char	Description
?	一個字元或沒有
*	任意數目的字元或沒有
+	一個字元或以上的字元
{N}	N個字元
{N,}	至少N個字元
{N, M}	至少N個字元至多M個字元

Regular Expression 組成

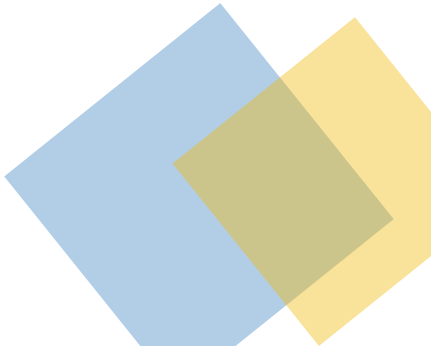
特殊字元(Metacharacter)

Char	Description
.	代表任一個字元
[...]	代表字元集中的任一字元, 例如 [abc] 可對應 a, b 或 c 連續字元的定義可用"-", 例如 [a-d] = [abcd]
[^...]	代表非字元集中的任一字元, 例如 [^abc] 將不對應 a, b 或 c

Regular Expression 組成

特殊字元(Metacharacter) – 匹配位置的

Char	Description
^	代表字串的開頭
$	代表字串的結尾



Regular Expression 組成

特殊字元(Metacharacter)

```
'最高法院刑事判決\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\
0\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u3000\u30
察官黃騰耀\n上\u3000訴\u3000人\u3000\n（被\u3000告）\u3000王宗立\n選
林宗憲律師\n上\u3000訴\u3000人\u3000王安石（原名王宗貞）\n（被\u3000
0訴\u3000人\u3000\n（被\u3000告）\u3000劉玉增\n選任辯護人\u3000丁中
\u3000\u3000告\u3000楊善博（原名楊偉祥）\n\u3000\u3000\u3000\u3000\
上訴人等因被告等違反證券交易法等罪案件，不服臺灣高等\n法院中華民國108年
號，起訴案號：臺灣臺北地方檢察署\n90年度偵字第21111、23982 號，91年度
年度偵字第500\n、10184、10344、13340號，94年度偵字第840、841、842、7
安石部分由其原審辯護人代為提起上訴），本院判決如下：\n      主   文\n原判
分；及劉玉增有罪部分，均撤銷，發回臺灣高等\n法院。\n其他上訴駁回。\n
事實欄（下稱事實\n      欄）一、二㈠至㈢如其附表（下稱附表）一㈠編號一及
一、二㈡如附表一㈡編號一犯行；上訴人即被告劉玉增有事\n      實欄二㈡犯行
定後，改判就王宗立附表一㈠編號\n      一依修正前刑法想像競合犯規定從一重
第1項第1\n      款共同犯法人行為負責人詐偽出售有價證券罪刑、就附表一\n
125條第1項前段、第3 項法人行為負責人，犯銀行\n      法第125條第1項前段之
像競合犯規定從一重論處\n      王安石如附表一㈡編號一所示修正前證券交易法
\n      券罪刑；就劉玉增依修正前刑法想像競合犯規定從一重論處\n      修正前
出售有價證券罪刑，及均諭知相關沒收。固\n      非無見。\n二、惟按：\n ㈠
券交易法第2條、第6條規定，有價證券之募集、\n      發行、買賣，其管理、監
行」之公司股票、公司\n      債券及經財政部核定之其他有價證券。但89年7 月
```

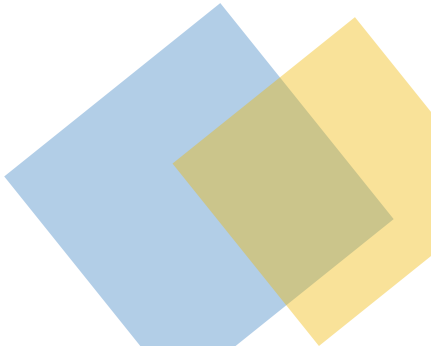
Char	Description
\d	對應0-9的任一數字(= [0-9])
\D	對應非數字的任一字元(=[^0-9])
\f	對應換頁字元
\n	對應換行字元
\r	對應歸正字元
\s	對應空白字元，對等於 [\f\n\r\t\v]
\S	對應非空白字元，對等於 [^ \f\n\r\t\v]
\t	對應 tab字元
\v	對應垂直 tab字元
\w	對應任何文數字元包括"_"，對等於 [a-zA-Z0-9_]
\W	對應任何非文數字元，對等於 [^a-zA-Z0-9_]

Regular Expression 組成

特殊字元(Metacharacter)

Char	Description
|	邏輯 "Or"
(pattern)	使用括號將pattern分組並提供記憶的功能，提供往後運算時再存取被括住的運算式功能。當有許多括號在pattern中使用時，被括住的運算式由左至右，可依序用$1、$2…$9存取。例如，"(a(bc)(d))" 運算式，被括號的運算式將有如下的對應$1="abcd"，$2="bc"，$3="d"。
$1 .. $9	依序對應pattern運算式中被括號刮住的部分

程式碼示範



動手試試看

https://regexone.com/

RegexOne
Learn Regular Expressions with simple, interactive exercises.

Interactiv

Lesson 1: An Introduction, and the ABCs

Regular expressions are extremely useful in extracting information from text such as code, log files, spreadsheets, or even documents. And while there is a lot of theory behind formal languages, the following lessons and examples will explore the more practical uses of regular expressions so that you can use them as quickly as possible.

The first thing to recognize when using regular expressions is that **everything is essentially a character**, and we are writing patterns to match a specific sequence of characters (also known as a string). Most patterns use normal ASCII, which includes letters, digits, punctuation and other symbols on your keyboard like %#$@!, but unicode characters can also be used to match any type of international text.

Below are a couple lines of text, notice how the text changes to highlight the matching characters on each line as you type in the input field below. To continue to the next lesson, you will need to use the new syntax and concept introduced in each lesson to write a pattern that matches all the lines provided.

Go ahead and try writing a pattern that matches all three rows, **it may be as simple as the common letters on each line**.

作業

繳交期限：12/26

- 自行選取 10 篇判決書
- 抓出判決書中出現的金額
- 將判決書中的金額轉為數字
- 結合判決書上的其他資訊畫圖