

AI與飛機工程應用

大數據分析

- R/Python/Julia/SQL 程式設計與應用
(R/Python/Julia/SQL Programming and Application)
- 資料視覺化 (Data Visualization)
- 機器學習 (Machine Learning)
- 統計品管 (Statistical Quality Control)
- 最佳化 (Optimization)



李明昌博士

<https://www.youtube.com/@alan9956>

<http://rwepa.blogspot.com/>

alan9956@gmail.com

大綱

1. 資料分析暨視覺化的心法 【[ai_01_apc_method.py](#)】
2. AI與螺旋槳性能最佳化應用 【[ai_02_aeronautical_engineering.py](#)】
3. AI與黃金價格深度學習預測應用(LSTM) 【[ai_03_gold_price.py](#)】
4. 結論

下載: https://github.com/rwepa/DataDemo/tree/master/propeller_design



1. 資料分析暨視覺化的心法

大綱

- 1.1 RWEPA 簡介
- 1.2 資料分析架構暨APC方法
- 1.3 資料分析暨視覺化工具
- 1.4 資料分析與視覺化應用



1.1 RWEPA 簡介

RWEPA簡介 <http://rwepa.blogspot.com/>

- 姓名：李明昌 (ALAN LEE)
- 現職：中華R軟體學會 常務理事
臺灣資料科學與商業應用協會 常務理事
- 學歷：中原大學 工業與系統工程所 博士
- 經歷：
 - 育達科技大學 資訊管理系(所) 兼任助理教授
 - 佛光大學 兼任教師
 - 國立台北商業大學 兼任教師
 - 東吳大學 兼任教師
 - 崇友實業 行銷企劃專員
 - 國航船務代理股份有限公司 海運市場運籌管理員
- 大專院校、資策會、工業技術研究院、國家發展委員會、中央氣象局、公平交易委員會、各縣市政府與日本名古屋產業大學等公民營單位演講達341餘場，3128小時以上。
- 連絡資訊：alan9956@gmail.com

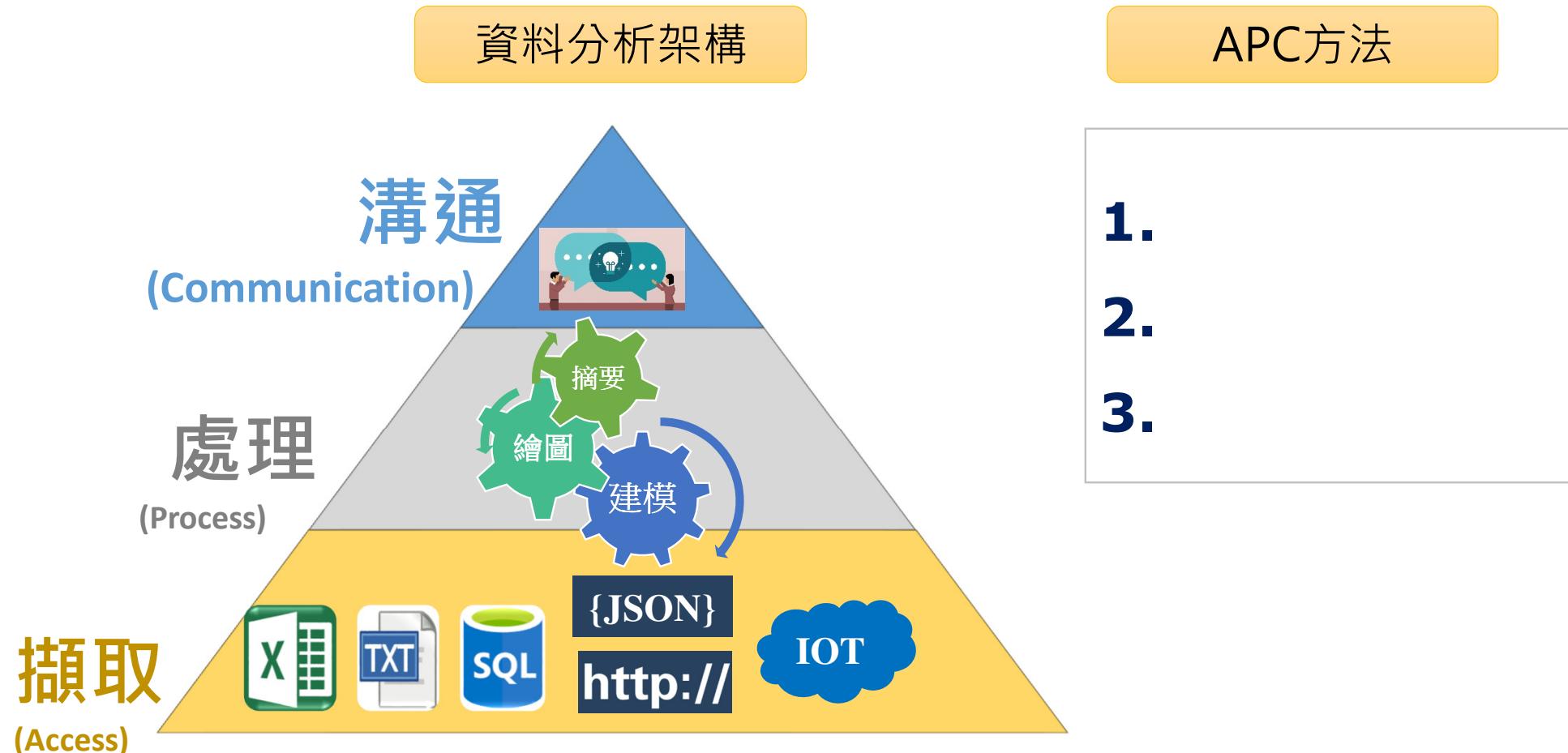


- iPAS 巨量資料分析師 證照推廣
- iPAS 營運智慧分析師 證照推廣



1.2 資料分析架構暨APC方法

★★★資料分析架構暨APC方法





1.3 資料分析暨視覺化工具

資料分析暨視覺化工具

- R - <http://rwepa.blogspot.com/> 【免費】



- Python - <http://rwepa.blogspot.com/2020/02/pythonprogramminglee.html> 【免費】



- Julia - <https://julialang.org/> 【免費】



- PowerBI- <https://powerbi.microsoft.com/zh-tw/> 【免費/付費】



- Tableau - <https://www.tableau.com/> 【免費/付費】



- Excel 【付費】 https://zh.wikipedia.org/zh-tw/Microsoft_Excel



大數據分析工具



- Microsoft Excel 2019: 104萬餘筆資料限制

A	B	C	D	E	F	G	
1	WEEK_END_DATE	STORE_NUM	UPC	UNITS	VISITS	HHS	SPEND
1048572	14-Jan-09	367	1111009477	13	13	13	18.07
1048573	14-Jan-09	367	1111009497	20	18	18	27.8
1048574	14-Jan-09	367	1111009507	14	14	14	19.32
1048575	14-Jan-09	367	1111035398	4	3	3	14
1048576	14-Jan-09	367	1111038078	3	3	3	7.5

1,048,576筆資料限制

大數據分析免費工具



軟體	Python	R	Julia
Released	1991	2000	2012
用途	程式語言 系統結合	統計,繪圖,視覺化 程式語言	科學計算 程式語言
版本	自由軟體 物件導向	自由軟體 物件導向	自由軟體 物件導向
附加功能	免費模組	免費套件	免費模組
Python 完全向下版本相容?			
使用者	工科+ 商管	商管+ 工科	商管+ 工科

如何學習 Python?

- 熟悉教材內容
- 將教材內容的資料集改為工作資料集(企業, 學術)
- 遇到問題時, 想辦法**尋找答案**
- 掌握 APC方法
- 掌握 ①摘要 ②繪圖 ③建模
- 參考網路應用文章 (進階) & 學術論文

```
尋找答案 = {"方法1": "同事,同學,朋友等",  
            "方法2": "Google",  
            "方法3": "alan9956@gmail.com"}
```

WHY!

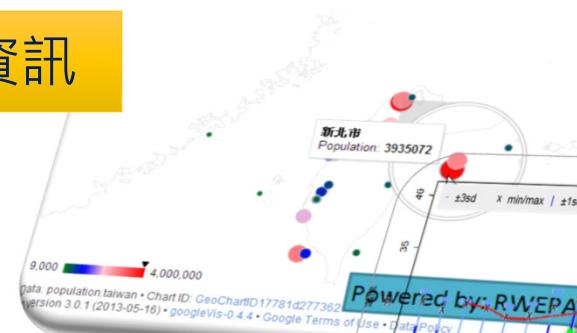


1.4 資料分析與視覺化應用

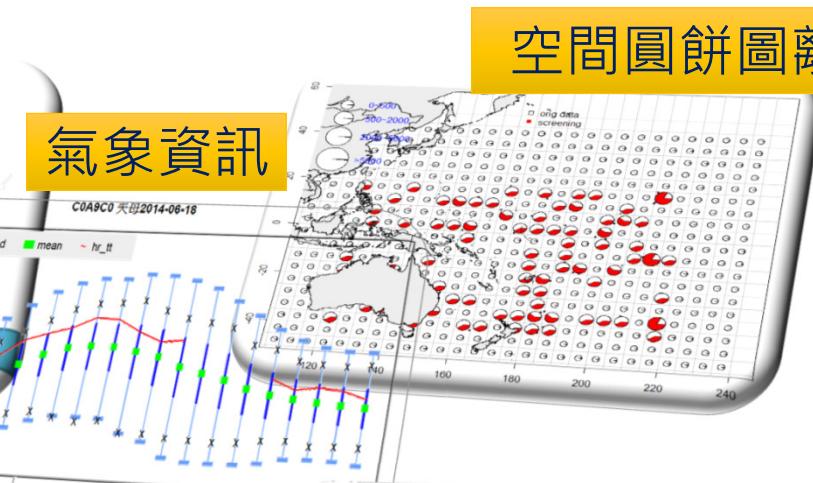
資料分析與視覺化應用

R + shiny, Python + Streamlit → 互動式網頁

地理資訊



氣象資訊

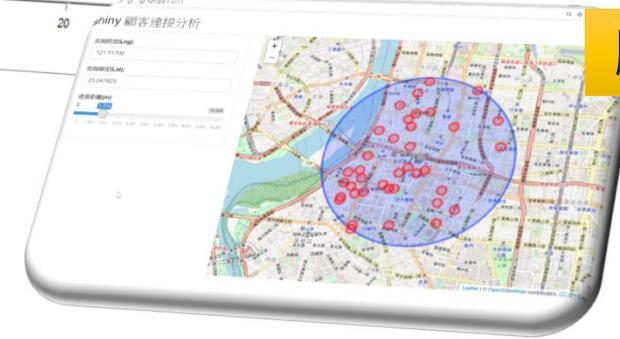


空間圓餅圖離群值分析

保險預測



顧客連結資訊



中央氣象局 1,600萬筆資料(14,328個檔案)

網頁呈現



客製化選單

R統計運算

2001, ... - 內容

一般 自訂

14,328 個檔案, 18 個資料夾

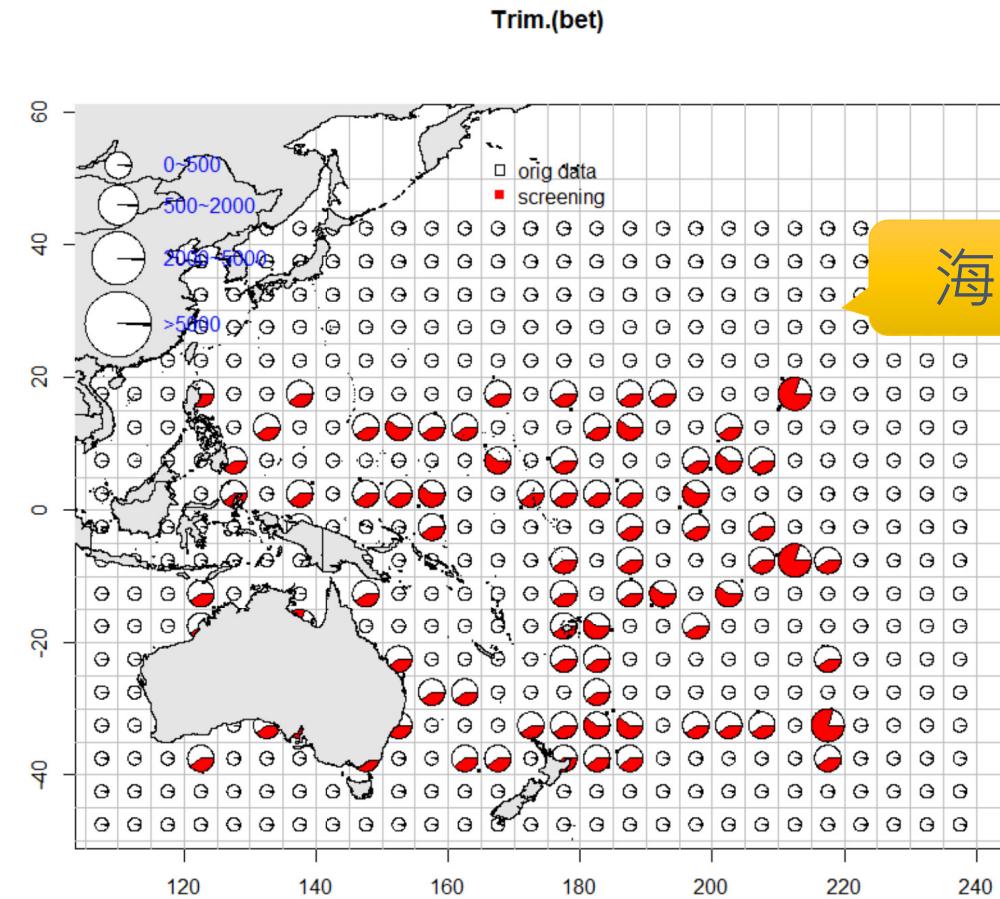
動態繪圖

保險預測模型

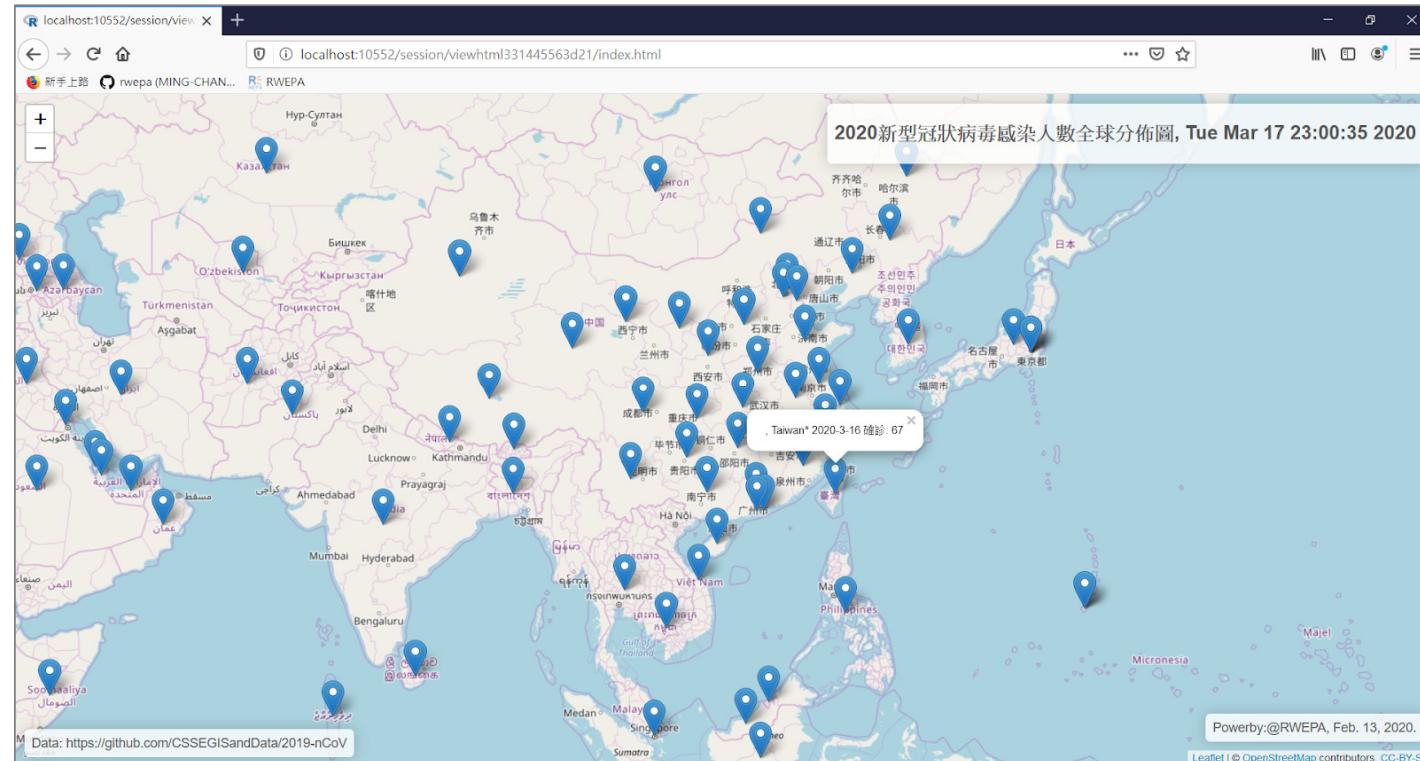
The screenshot shows a web-based data analysis interface for the iinsurance platform. At the top, there's a navigation bar with tabs like '檔案上傳', '資料處理', '統計圖表', '模型評估', and '預測模型'. A red box highlights the '預測模型' tab. Below it, a sub-menu for '檢視結果' is also highlighted with a red box. On the left, a yellow callout box contains the text '機率模型閥值調整' (Probability Model Threshold Adjustment) with a red box around the '機率模型閥值' slider. The main area displays a table of 12 entries with various columns including gender, vehicle type, exposure risk, deductible count, discount rate, insured person age, car age, and prediction probability. The last column indicates if there is a claim ('理賠'). A large yellow speech bubble on the right says '預測結果 {有,無}' (Prediction Result {Yes, No}). Red arrows point from the '檢視結果' button to the table and from the '機率模型閥值' slider to the same area.

性別	女性	車輛種類	私家車	曝露風險	曝露風險對數	無索償折扣	被保險人年齡	私家車 一車齡 0	私家車 一車齡 1	私家車 一車齡 2	私家車 車齡 0_1_2 組合	車齡 0_1_2 組合	預測機率	理賠		
M	0	A	1	0.9144422	-0.08944106	50	4	1	0	0	1	0	2	0.1069	有	
M	0	A	1	0.8158795	-0.20348856	20	4	0	0	1	1	2	2	0.1441	有	
3	M	0	A	1	0.8377823	-0.17699695	50	3	0	0	1	1	2	2	0.1866	有
4	M	0	A	1	0.4325804	-0.83798702	50	6	0	1	0	1	1	2	0.0944	無
5	M	0	A	1	0.7173169	-0.33223755	50	4	0	0	1	1	2	2	0.1218	有
6	M	0	A	1	0.8377823	-0.17699695	50	4	0	0	1	1	2	2	0.1495	有
7	M	0	A	1	0.8487337	-0.16400975	50	5	0	0	1	1	2	2	0.1422	有
8	F	1	A	1	0.8268309	-0.19015503	10	3	0	0	1	1	2	2	0.1733	有
9	M	0	A	1	0.7145791	-0.33606164	0	5	1	0	0	1	0	2	0.0694	無
10	M	0	A	1	0.3340178	-1.09656101	0	3	0	0	1	1	2	2	0.0783	無

空間圓餅圖離群值分析



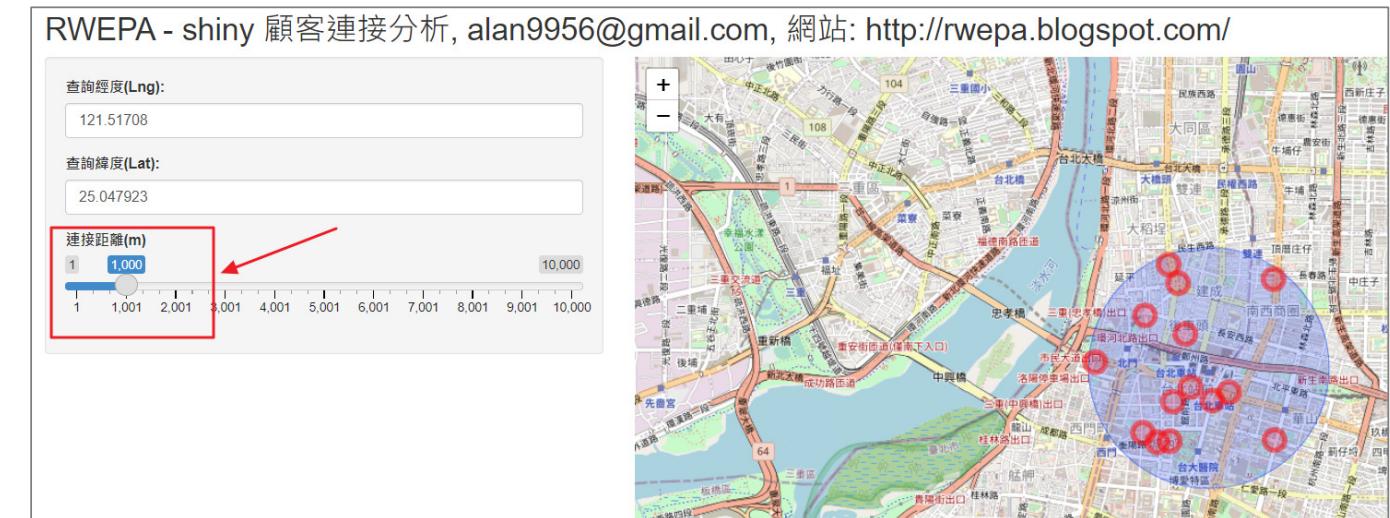
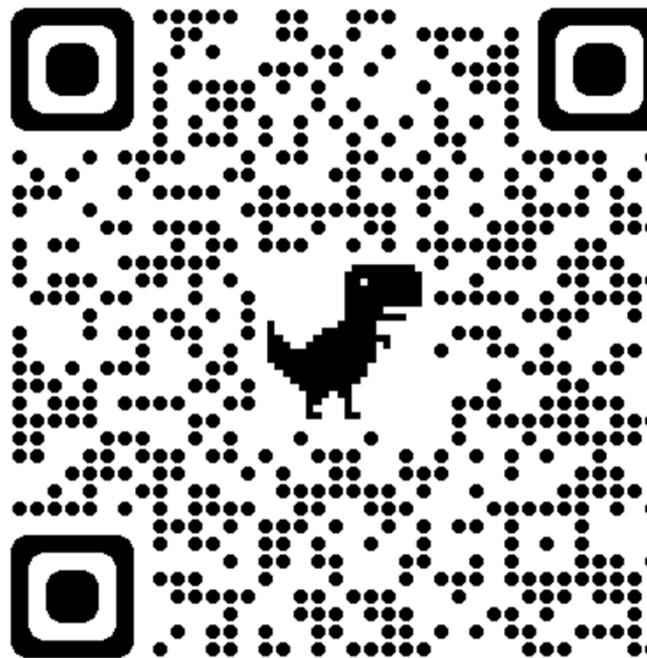
2020新型冠狀病毒視覺化



<http://rwepa.blogspot.com/2020/02/2019nCoV.html>

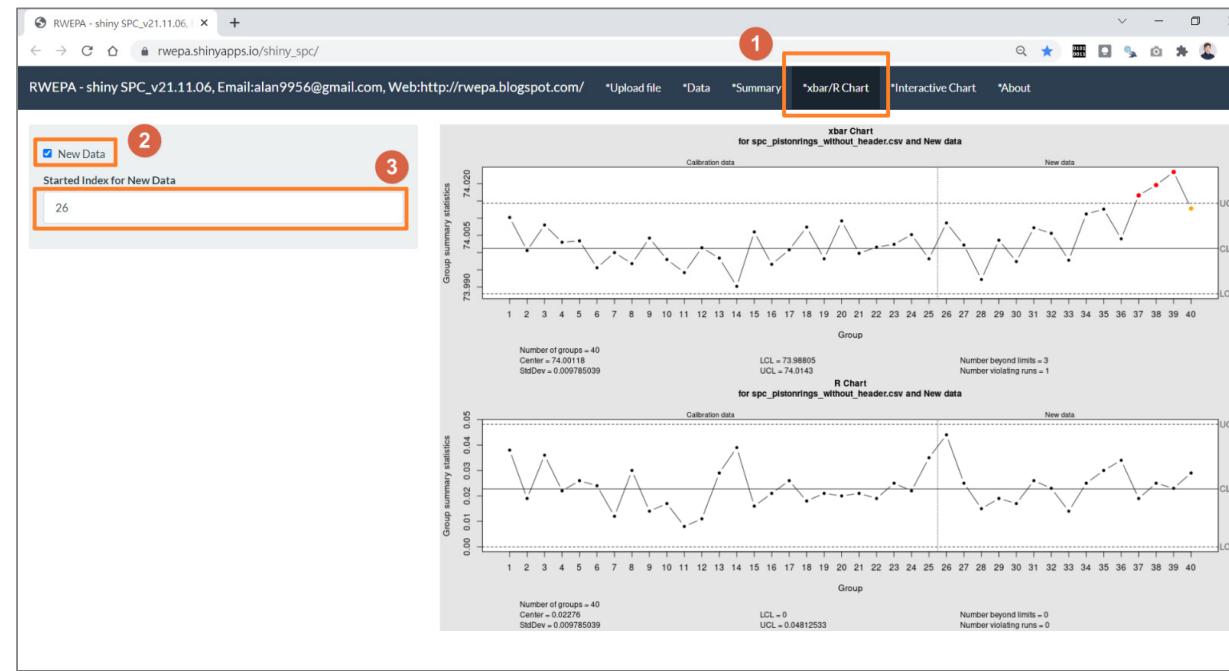
shiny 顧客連接分析

- <https://rwepa.shinyapps.io/shinyCustomerConnect/>



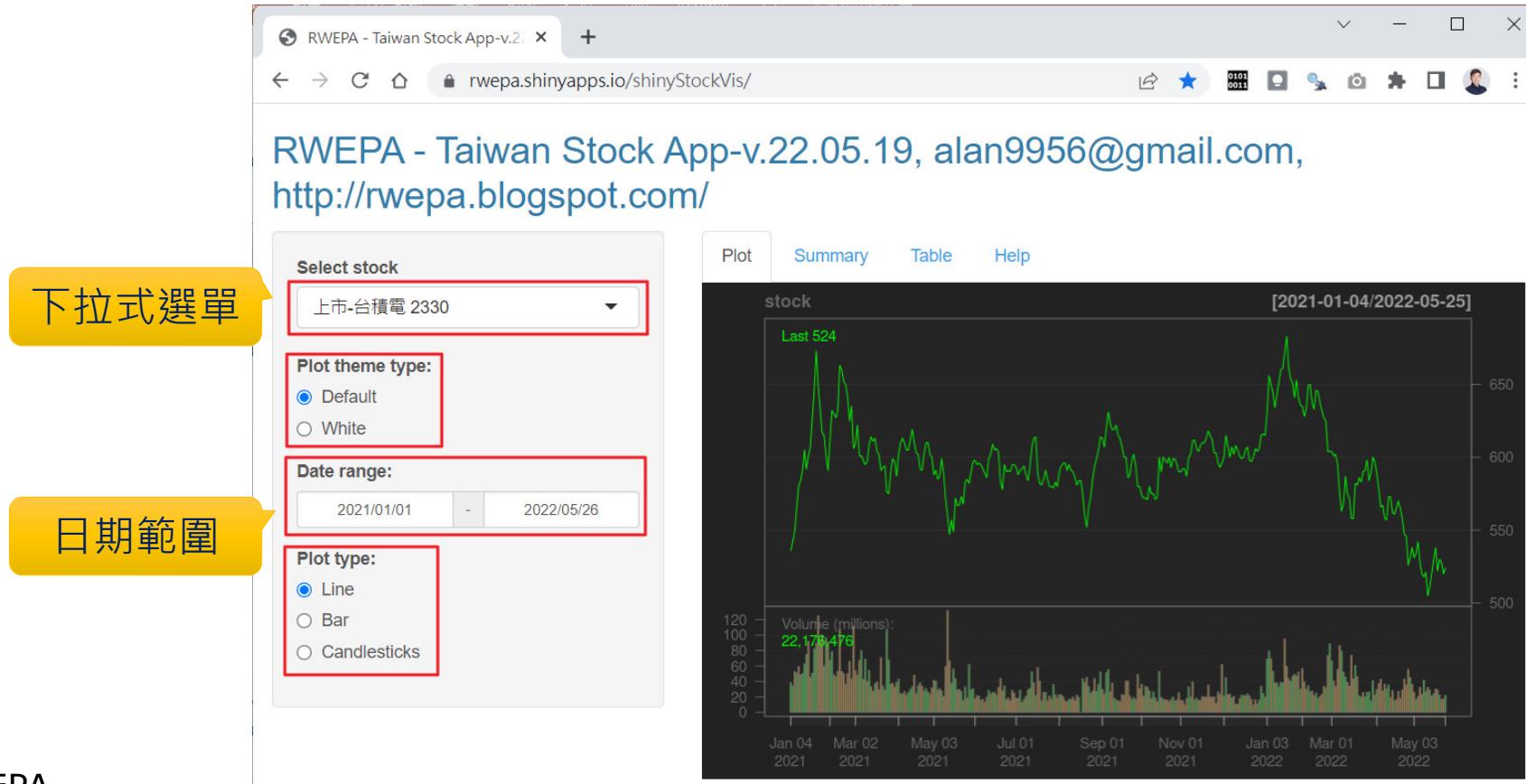
品質管制圖(quality control chart)應用

- 說明: <http://rwepa.blogspot.com/2021/10/r-shiny-quality-control-chart.html>
- 資料1: https://github.com/rwepa/shiny_spc/blob/main/data/spc_wafer_with_header.csv
- 資料2: https://github.com/rwepa/shiny_spc/blob/main/data/spc_pistonrings_without_header.csv
- 線上示範: https://rwepa.shinyapps.io/shiny_spc/

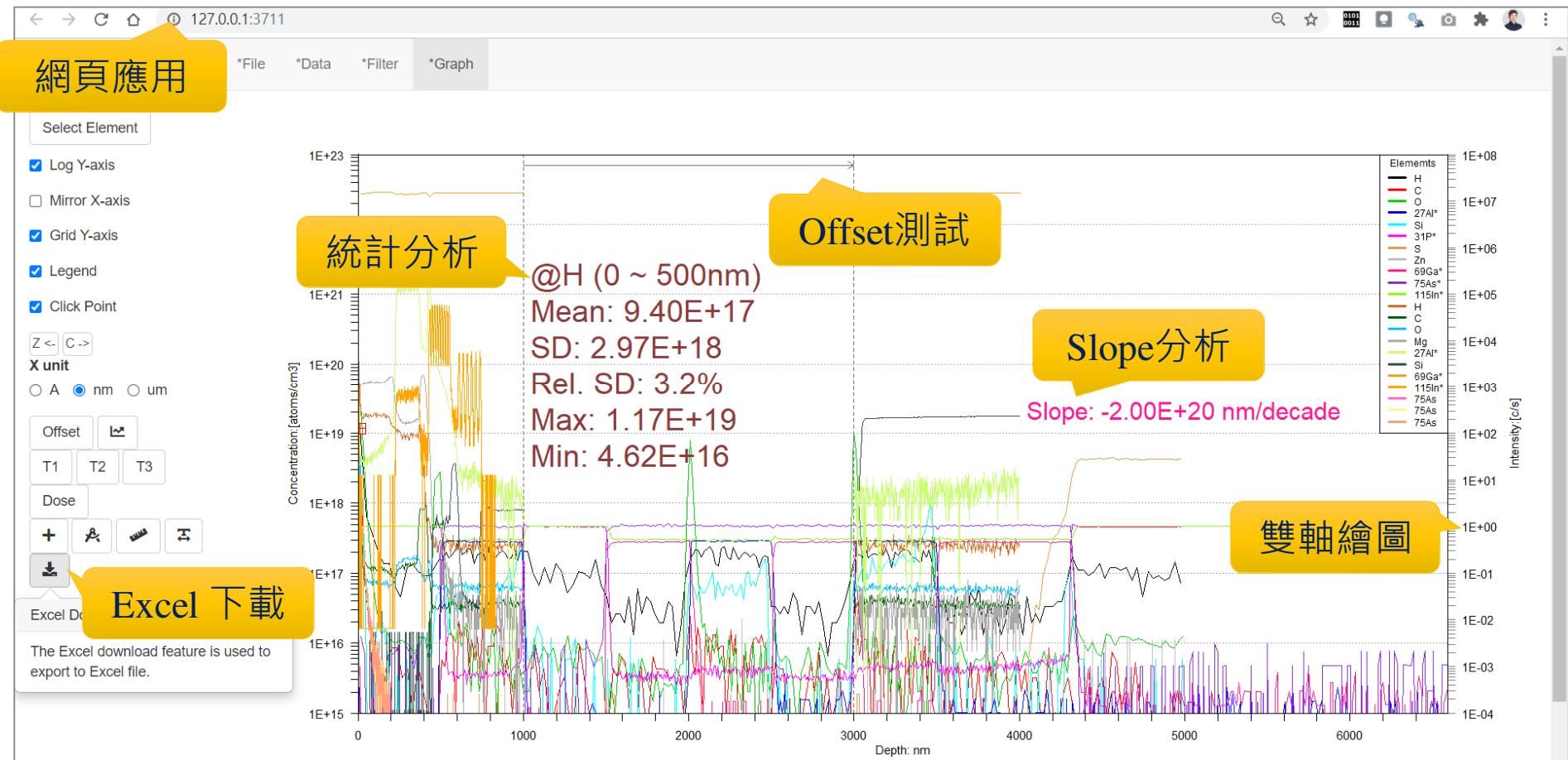


Taiwan Stock App

- <https://rwepa.shinyapps.io/shinyStockVis/>

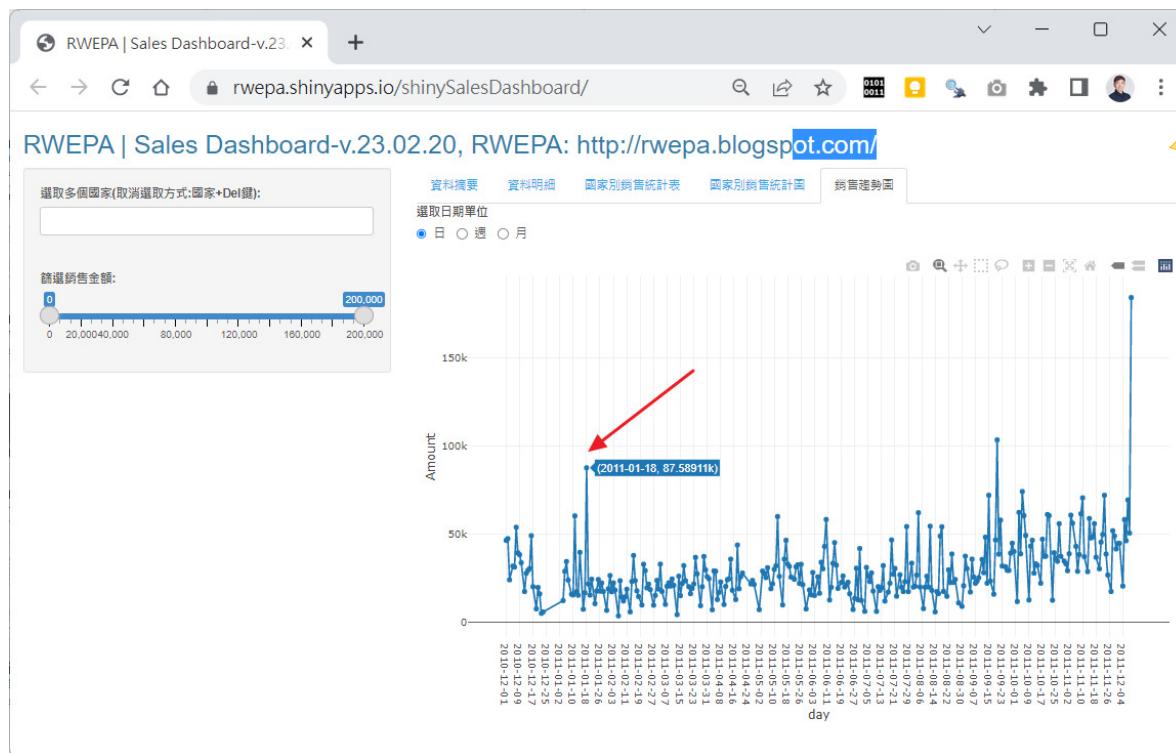


離子資料分析與視覺化應用



RWEPA | shiny企業實務應用 第4集-shiny銷售儀表板

- Shiny: <https://rwepa.shinyapps.io/shinySalesDashboard/>
- YouTube: <https://youtu.be/4GgZlf8heQk>



謝謝 ^_ ^

訂閱、讚、開啟小鈴鐺

shiny企業實務應用 第6集-小明算命師(下) - 第1季完結篇

- Ubuntu Shiny Server: <https://shiny.rwepa.net/shiny-hr-teller/>
- YouTube: <https://youtu.be/rrD6KV3eV-w>



Power BI - RFM分析

- YouTube : <https://youtu.be/Lkr9HmzLTtg>
- <http://rwepa.blogspot.com/2023/07/rwepa-rfm-analysis-using-power-bi.html>

Customer Segmentation Using RFM Analysis, 2023



Calendar



最近消費 (recency) :
顧客上次消費時間愈近，用戶價值愈大。

消費頻率 (frequency) :
顧客在一段時間中，總購買次數，購買頻率愈高，用戶價值愈大。

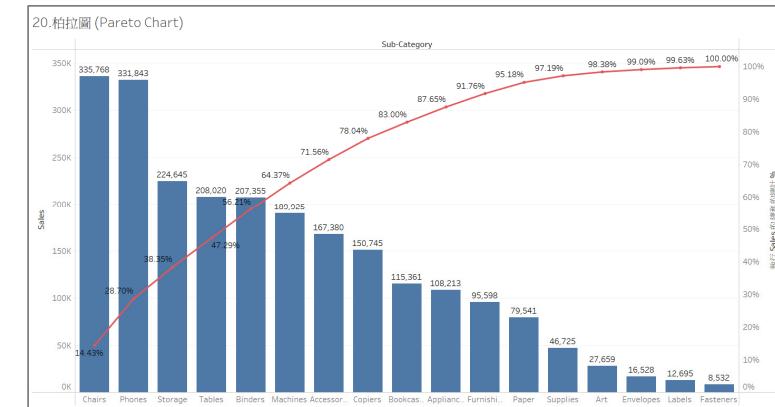
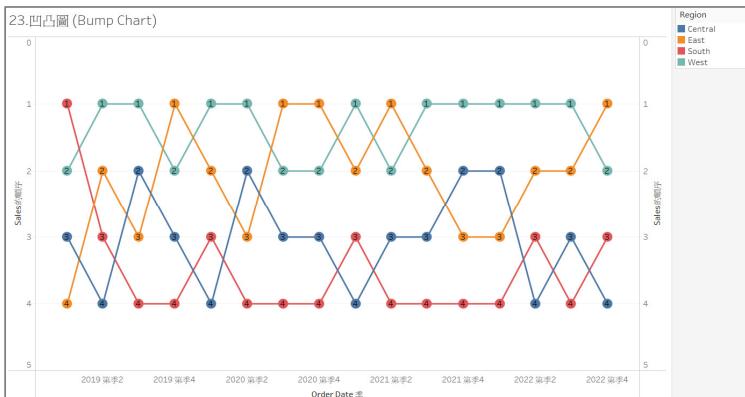
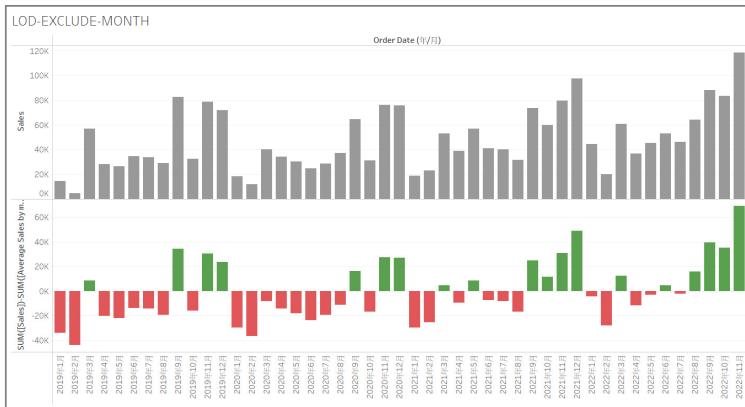
消費金額 (monetary) :
顧客總消費金額，消費金額愈高，用戶價值愈大。

Author : Ming-Chang Lee
YouTube : <https://www.youtube.com/@alan9956>
RWEPA : <http://rwepa.blogspot.tw/>
GitHub : <https://github.com/rwepa>
Email : alan9956@gmail.com

RFM分析 x RFM視覺化分析 RECENTY FREQUENCY Monetary +

Tableau - 智慧製造應用

- <https://github.com/rwepa/Talks>
- <https://public.tableau.com/app/profile/ming.chang.lee/vizzes>

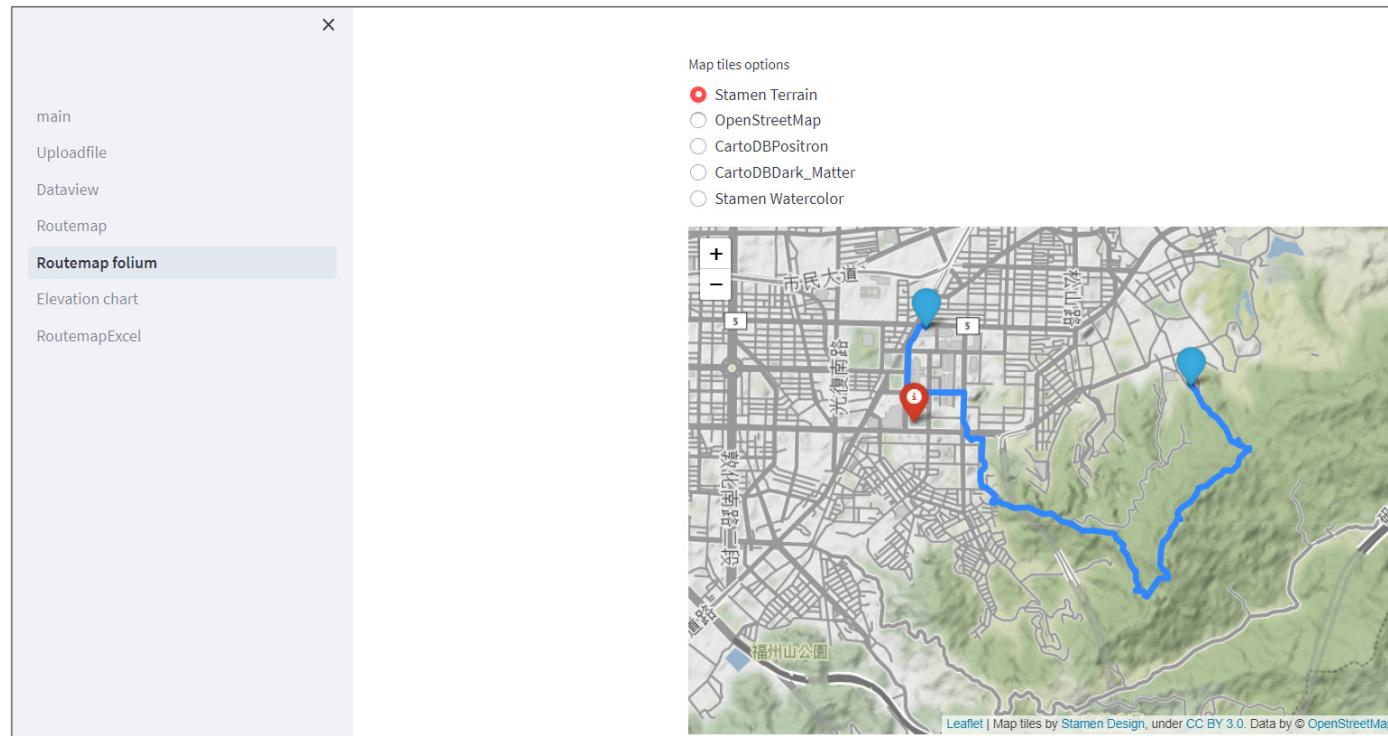


A Market Basket Analysis table showing the frequency of items purchased together. The columns represent different categories like Accessories, Appliances, Art, etc., and the rows represent individual items.

Sub-Categ..	Access..	Applianc..	Art	Binders	Bookcas..	Chairs	Copiers	Envole..	Fasten..	Furnishin..	Machin..	Paper	Phones	Storage	Supplies	Tables
Accessories	718	60	89	161	29	64	6	32	27	115	47	22	153	118	25	44
Appliances	60	459	63	135	13	49	9	21	23	82	21	9	110	78	56	24
Art	89	63	756	159	29	48	7	29	30	105	47	12	152	124	100	41
Binders	161	135	159	1,339	56	126	19	54	62	200	82	30	276	199	201	43
Bookcas..	29	13	34	56	228	21	2	11	7	30	21	6	48	42	37	10
Chairs	64	49	87	126	21	591	10	29	28	107	36	19	133	91	85	19
Copiers	6	9	7	19	2	10	70	4	5	12	6	2	20	11	14	6
Envole..	32	21	29	54	11	29	4	251	11	32	11	6	59	41	38	5
Fasteners	27	23	30	62	7	28	5	11	226	48	15	8	59	39	39	13
Furnishin..	115	82	105	200	30	107	12	32	48	919	57	29	181	154	140	31
Labels	47	21	47	82	21	36	6	11	15	57	348	8	80	58	61	18
Machin..	22	9	12	30	6	19	2	6	8	29	8	114	28	22	21	1
Paper	153	110	152	276	48	133	20	59	59	181	80	28	1,205	179	178	45
Phones	116	78	124	199	42	91	11	41	39	154	58	22	179	826	117	22
Storage	103	56	100	201	37	95	14	38	39	140	61	21	178	117	797	33
Supplies	25	24	29	43	10	19	6	5	13	31	18	1	45	22	33	189
Tables	44	32	41	74	9	36	4	12	14	46	17	6	54	54	50	13

登山路線視覺化分析平台 (Python + Streamlit)

- YouTube : https://youtu.be/-_zghs2qrlg
- 系統展示 <https://rwepa-climb.streamlit.app/>



R 入門資料分析與視覺化應用(7小時28分鐘)

- <https://mastertalks.tw/products/r?ref=MCLEE>

課程提供教學範例的原始程式檔案與資料集 +中文字幕



- **主題**
 1. R, RStudio簡介與套件使用
 2. 認識資料物件
 3. 資料處理與分析
 4. 資料視覺化應用
- **特色**
 1. 資料分析的**關鍵八步**
 2. 提供必備**ggplot2**套件的應用知識與使用情境
 3. 提供日期時間**zoo, xts**套件的整合應用操作
 4. 提供**人力資源**資料與**銷售資料**，強化**實務資料**操作能力

R 商業預測應用(8小時53分鐘)

- <https://mastertalks.tw/products/r-2?ref=MCLEE>



- **主題**

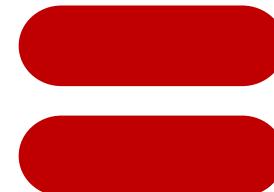
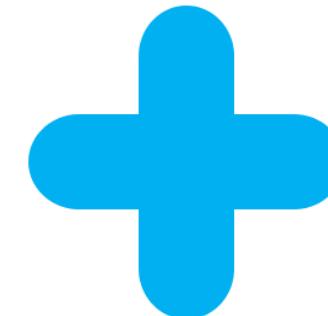
1. R · RStudio工具操作
2. 非監督式學習商業預測
3. 監督式學習商業預測
4. 財金資料預測應用

- **特色**

1. 採用**最有效率**方式學習大數據R語言，並應用於**職場資料分析**與**商業預測應用**
2. 提供**多元線性迴歸**的必備知識
3. 提供**財金資料商業預測應用**的基礎與進階必學技能
4. 提供學員人力資源資料與**台指期tick資料**預測演練

課程提供教學範例的原始程式檔案與資料集 +中文字幕

學習目標



2.AI與螺旋槳性能最佳化應用



大綱

- 2.1 螺旋槳資料集簡介
- 2.2 螺旋槳資料集下載
- 2.3 安裝模組
- 2.4 載入模組
- 2.5 匯入螺旋槳資料集
- 2.6 計算 solidity
- 2.7 資料視覺化
- 2.8 Solidity 遺漏值, 使用 KNN 填補法

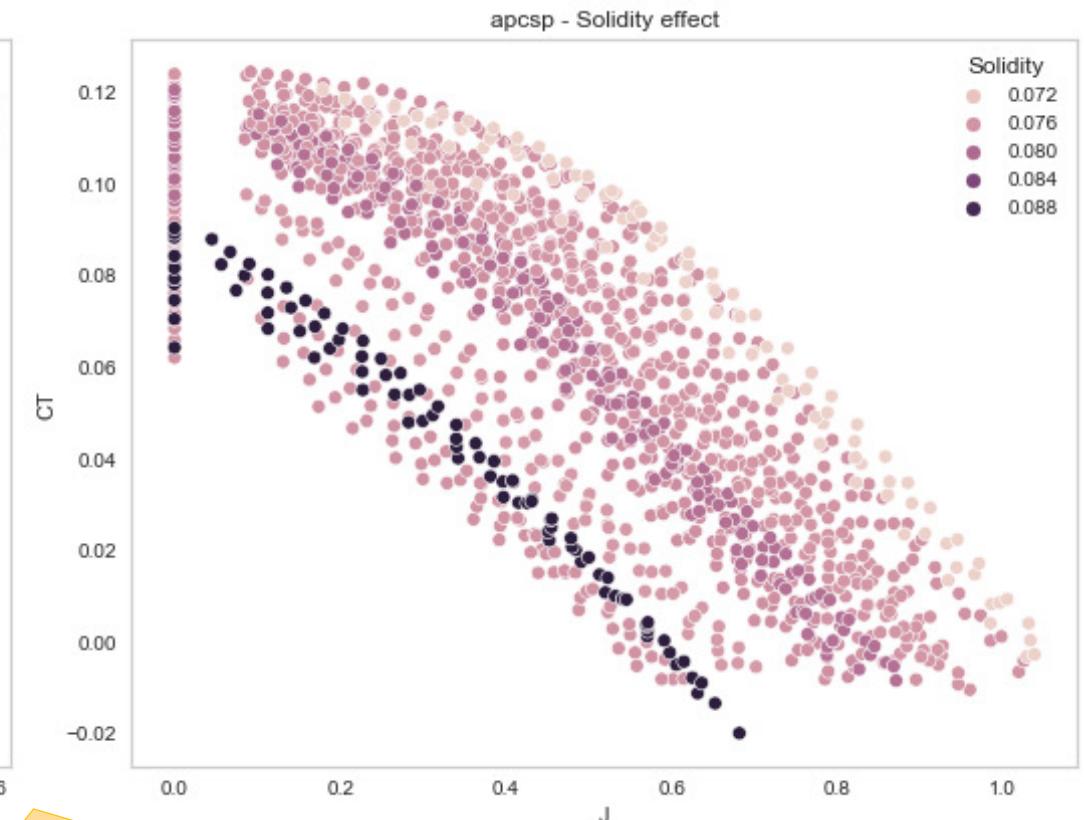
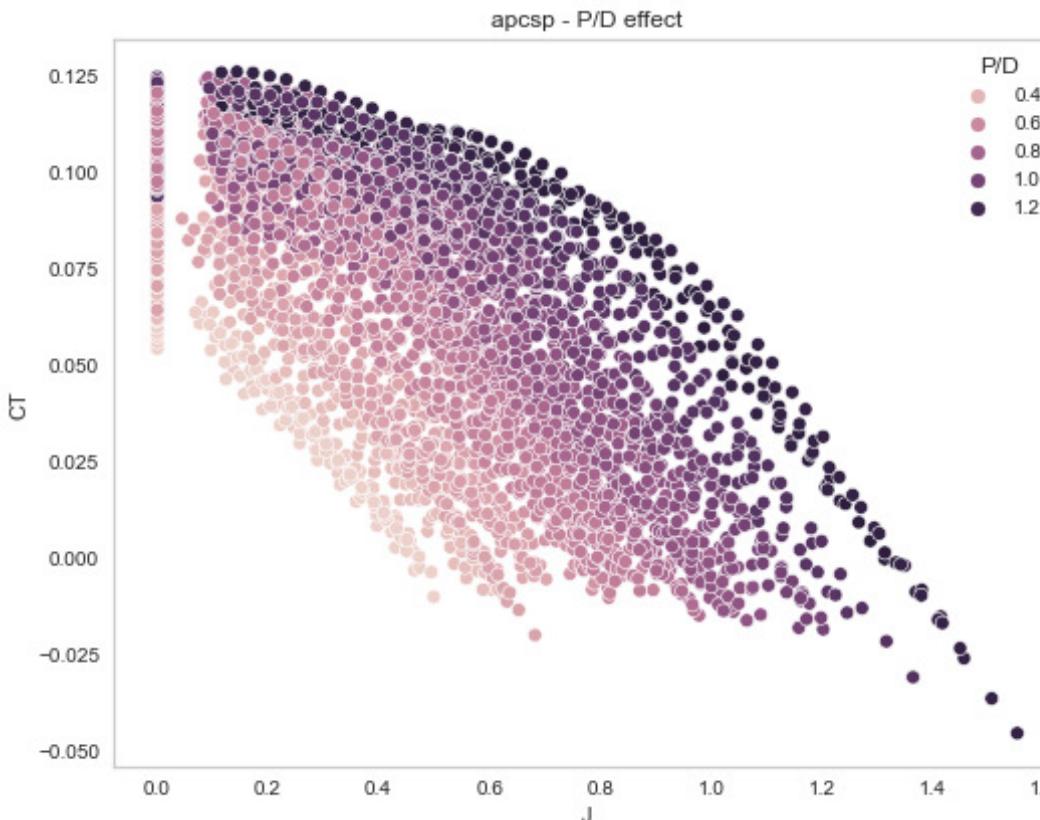


<https://simpleflying.com/how-an-airplane-propeller-works/>

2.1 螺旋槳資料集簡介

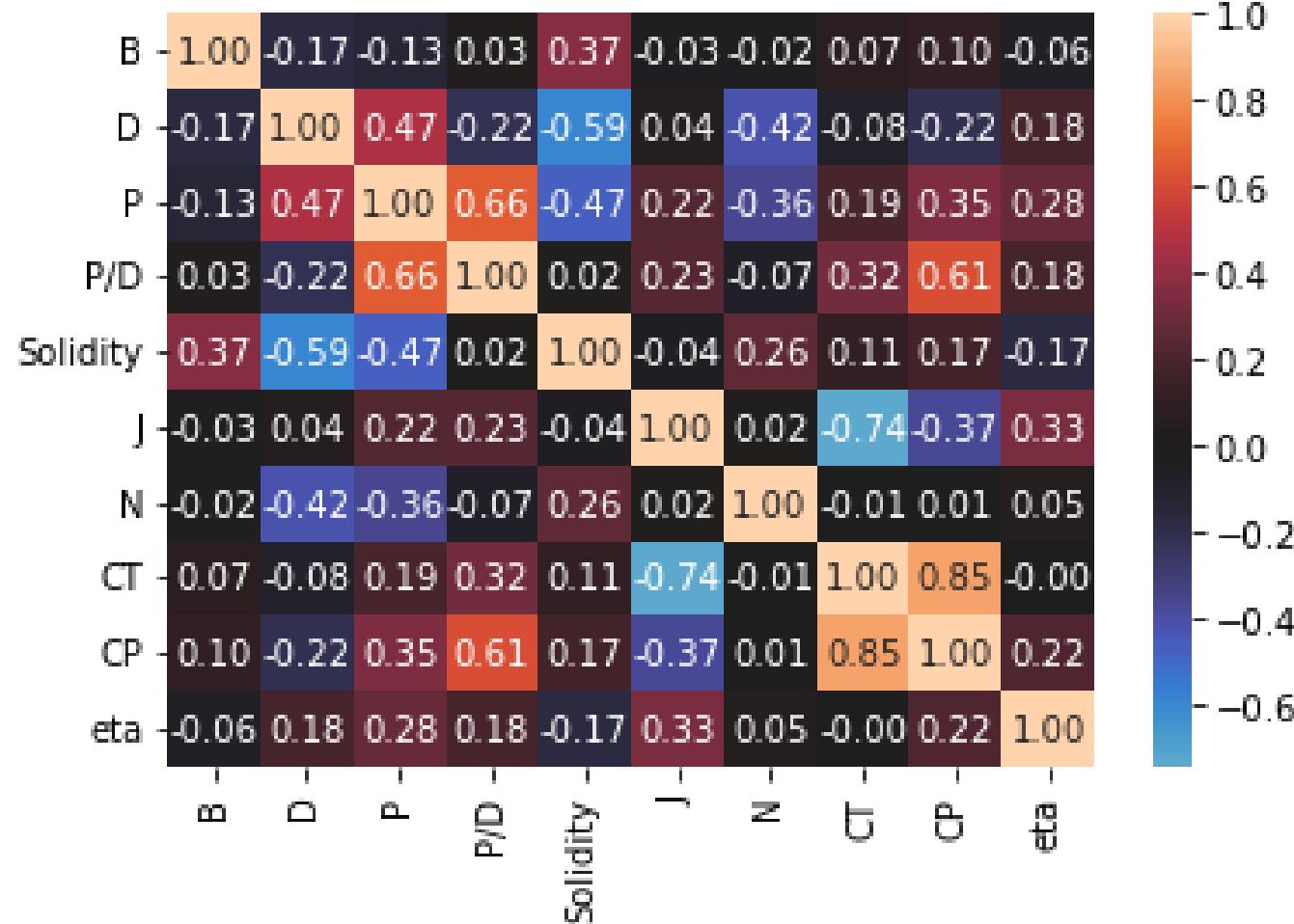
- UIUC螺旋槳資料集: UIUC小型無人機和模型飛機上使用的螺旋槳的風洞測量
- UIUC: 伊利諾大學厄巴納 - 香檳分校 (University of Illinois Urbana-Champaign)
- UIUC螺旋槳資料集下載(RWEPA GitHub)
- https://github.com/rwepa/DataDemo/tree/master/propeller_design
- ai_02_aeronautical_engineering.py

J vs. CT 散佈圖資料視覺化



Python demo

相關係數圖-所有變數-視覺化



Python demo

3.AI與黃金價格深度學習預測應用(LSTM)



大綱

- 3.1 CRISP-DM 六大步驟
- 3.2 黃金價格深度學習預測應用(LSTM)
- 3.3 台灣股市,ETF下載

3.1 CRISP-DM 六大步驟

ai_03_gold_price.py

資料探勘生命週期 - CRISP-DM

- 跨產業資料探勘標準作業流程
(CRoss Industry Standard Process for Data Mining)
- CRISP-DM是於1990年起，由SPSS以及NCR兩大廠商在合作戴姆克萊斯勒-賓士(Daimler Benz)的資料倉儲以及資料探勘過程中發展出來的。

CRISP-DM 資料探勘流程(續)

- 步驟 1：商業理解
- 步驟 2：資料理解
- 步驟 3：資料準備
- 步驟 4：模式建立
- 步驟 5：評估與測試
- 步驟 6：佈署應用

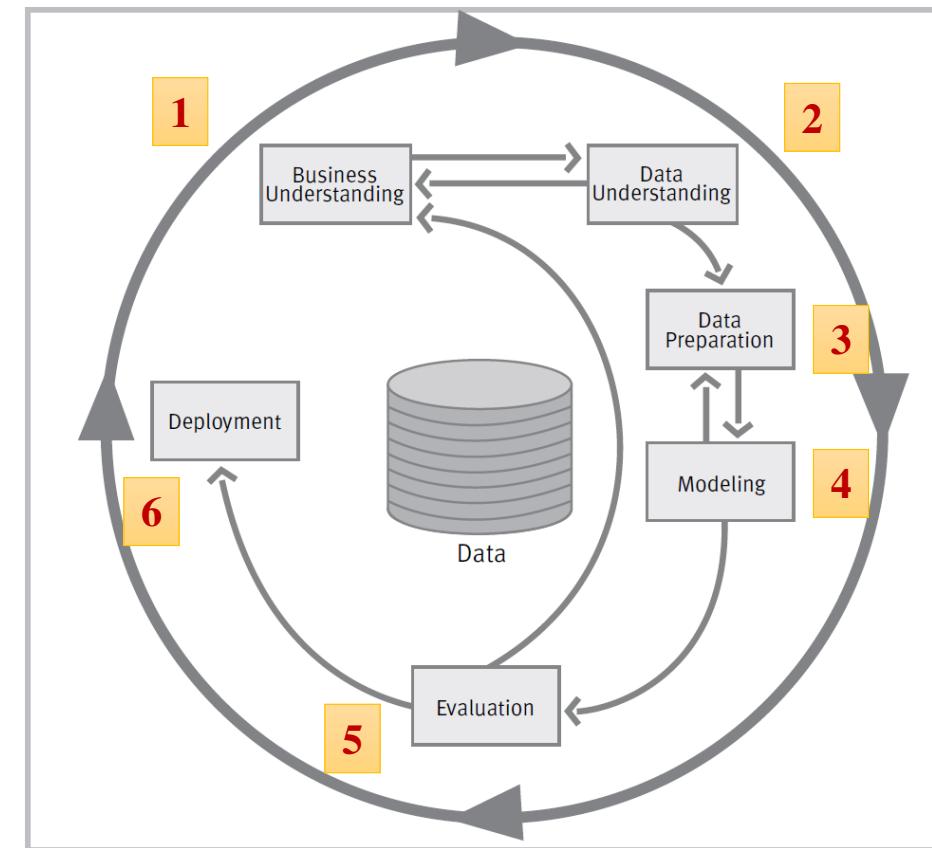
佔整專案時間的~80%

步驟3 資料準備

將資料隨機區分為二大類：

- 訓練資料70% (較大)
- 測試資料30% (較小)

CRISP-DM 資料探勘流程(續)



參考 https://en.wikipedia.org/wiki/Cross_Industry_Standard_Process_for_Data_Mining

數值模型績效指標

- 不可直接使用誤差的算術平均!

$$\cancel{\text{Total error}} = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)$$

- 均方誤差 (Mean Squared Error, MSE)

$$MSE = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2$$

- 均方根誤差 (Root Mean Squared Error, RMSE)

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2}$$

- 平均絕對誤差 (Mean Absolute Error, MAE)

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i|$$

類別模型績效指標 - 混淆矩陣

- <http://rwepa.blogspot.com/2013/01/rocr-roc-curve.html>

```
#           | 真實P類別 真實N類別
# ****|*****|*****|*****
# 預測P類別 | TP真陽數 FP假陽數
# 預測N類別 | FN假陰數 TN真陰數
# ****|*****|*****|*****
#           | P          N

# 1.TPR(True positive rate) 真陽性率, 愈大愈好 -----
# =TP/ (TP+FN)
# =TP/P
# =Sensitivity 積敏度
# =Recall 召回率
# =Probability of detection
# =Power
# 實際為陽性的樣本中, 判斷為陽性的比例。
# 例如真正有生病的人中, 被醫院判斷為有生病者的比例。
```



3.2 黃金價格深度學習預測應用(LSTM)

Yahoo 財金網站

- <https://tw.stock.yahoo.com/>

- 左上角輸入 GC=F
- 上漲金額 USD.13.20
- 上漲比例 0.64%



圖1. GC價格走勢圖(2013-2023年)



模型訓練, 正確率

```
Epoch 1/150
WARNING:tensorflow:From C:\Users\asus\anaconda3\Lib\site-packages\keras\src\utils\tf_utils.py:492:
The name tf.ragged.RaggedTensorValue is deprecated. Please use
tf.compat.v1.ragged.RaggedTensorValue instead.

70/70 [=====] - 13s 64ms/step - loss: 0.0459 - val_loss: 0.0705
Epoch 2/150
70/70 [=====] - 3s 41ms/step - loss: 0.0123 - val_loss: 0.0325
Epoch 3/150
70/70 [=====] - 3s 42ms/step - loss: 0.0065 - val_loss: 0.0164
Epoch 4/150
70/70 [=====] - 3s 42ms/step - loss: 0.0039 - val_loss: 0.0103
Epoch 5/150
70/70 [=====] - 3s 46ms/step - loss: 0.0028 - val_loss: 0.0075
Epoch 6/150
70/70 [=====] - 5s 73ms/step - loss: 0.0023 - val_loss: 0.0074
```

```
In [44]:
...: print("Test Loss:", result)
...: print("Test MAPE:", MAPE)
...: print("Test Accuracy:", Accuracy)
Test Loss: 0.0006855355459265411
Test MAPE: 0.021935985195150547
Test Accuracy: 0.9780640148048495
```

正確率: 97.8%

Python demo

實際與預測比較圖

圖4. Model Performance on Gold Price Prediction



3.3 台灣股市,ETF下載

元大台灣50ETF走勢圖





4.結論



學習心得

- 資料分析暨視覺化的心法- APC方法
- 螺旋槳性能最佳化應用 – 資料分析群組、視覺化
- 黃金價格深度學習預測應用(LSTM)、ETF下載與視覺化

Python 常用模組

模組	功能	
Numpy	Large, multi-dimensional arrays and matrices	
Scipy	Optimization, linear algebra, integration, FFT, signal	
Pandas	DataFrame object for data manipulation	
Matplotlib	Static, animated, and interactive visualizations	
Statsmodels	Statistical models	
Scikit-learn	Machine learning library	
Tensorflow, PyTorch	Deep learning	 
Biopython	Biological computation	
Scanpy	Single-cell analysis	
Django, Flask, Streamlit	Web	  
Plotly, dash, bokeh	Interactive visualization	  

參考資料

- RWEPA
 - <http://rwepa.blogspot.com/>
- Python 程式設計-李明昌 <免費電子書>
 - <http://rwepa.blogspot.com/2020/02/pythonprogramminglee.html>
- iPAS Python programming <免費教材>
 - https://github.com/rwepa/ipas_bda/blob/main/ipas-python-program.py
- RWEPA | 登山路線視覺化分析平台 (Python + Streamlit) 【中文字幕】
 - YouTube: https://youtu.be/-_zghs2qrlq
 - Link: <https://rwepa.blogspot.com/2023/08/visualization-climbing-routes-with.html>
- 投影片: https://github.com/rwepa/DataDemo/tree/master/propeller_design

謝謝您的聆聽

Q & A



李明昌

alan9956@gmail.com

<http://rwepa.blogspot.tw/>