

Orange – Association rules

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)



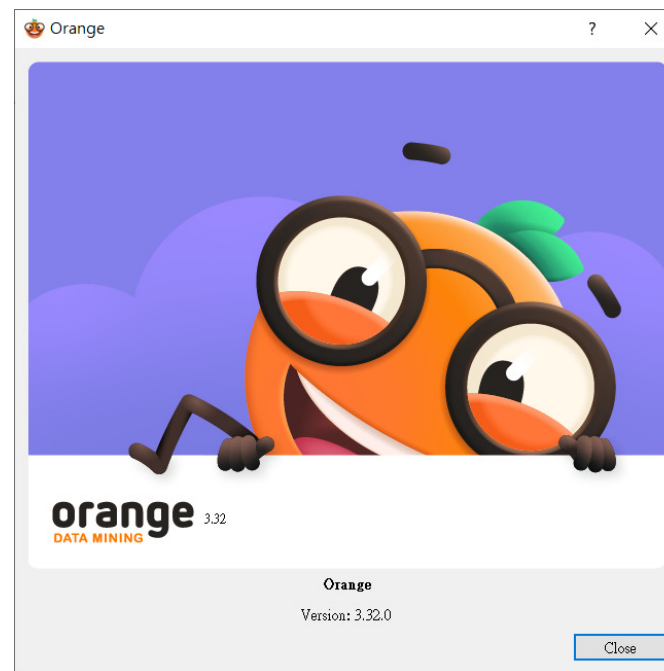
李明昌博士

alan9956@gmail.com

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

大綱

1. Orange 3 簡介
2. Orange 3 安裝
3. 安裝 Add-ons Associate
4. 案例示範
5. Q & A



1. Orange 3 簡介

Orange 3 特性

- University of Ljubljana, Slovenia, 10 October 1996
 - 視覺化程式設計工具(Visual Programming Tools)
 - 執行更加快速(C++)與視覺化操作
 - 提供多種機器學習模組
 - 開放原始碼與跨平台
 - 適用於Python模組
 - 資料預處理
 - 模型訓練
 - 部署
- 官網 <https://orangedatamining.com/>
 - 元件 <https://orangedatamining.com/widget-catalog/>
 - Tutorial <https://orange3.readthedocs.io/projects/orange-data-mining-library/en/latest/>

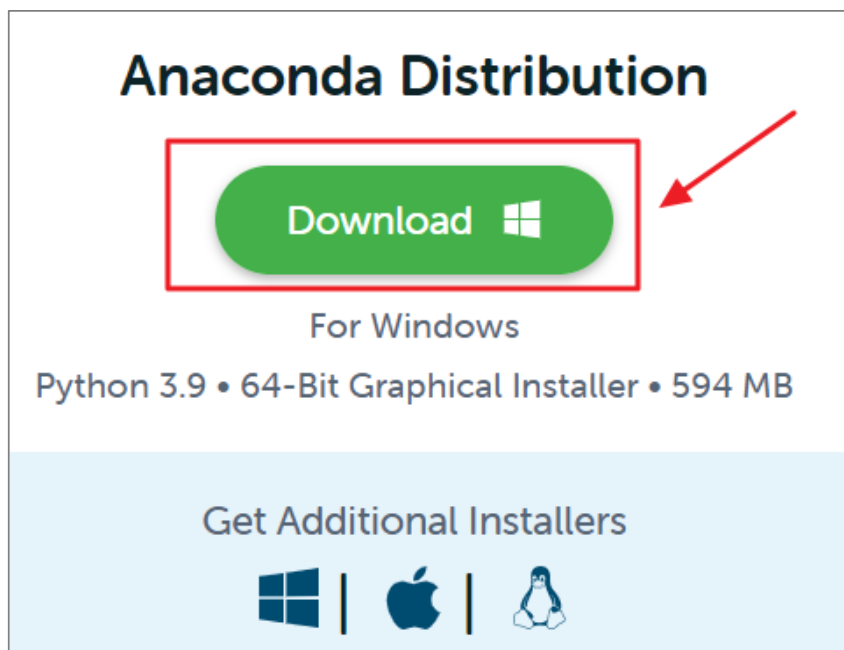
Orange3 add-ons (外掛元件)

1. Orange3-Associate	關聯規則	9. Orange3-SingleCell	單細胞分析
2. Orange3-Bioinformatics	生物資訊	10. Orange-Spectroscopy	光譜分析
3. Orange3-Educational	機器學習教育示範	11. Orange3-Text	文字探勘
4. Orange3-Explain	重要特徵解釋	12. Orange3-Textable	文字探勘, 編碼
5. Orange3-Geo	地理資料視覺化	13. Orange3-Timeseries	時間序列
6. Orange3-ImageAnalytics	影像分析	14. Orange3-Survival-Analysis	存活分析
7. Orange3-Network	網路分析	15. Orange3-WorldHappiness	世界社會經濟指標
8. Orange3-Prototypes	原型		

2. Orange 3 安裝

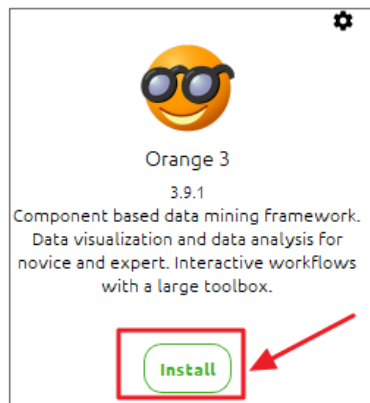
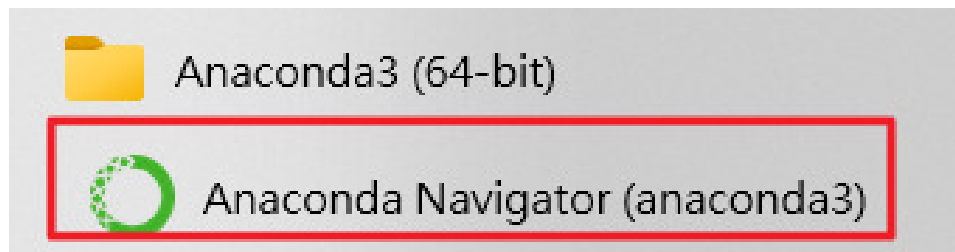
方法1 使用 Anaconda 下載安裝(1/3)

- 下載 Anaconda 安裝程式
- <https://www.anaconda.com/products/distribution>



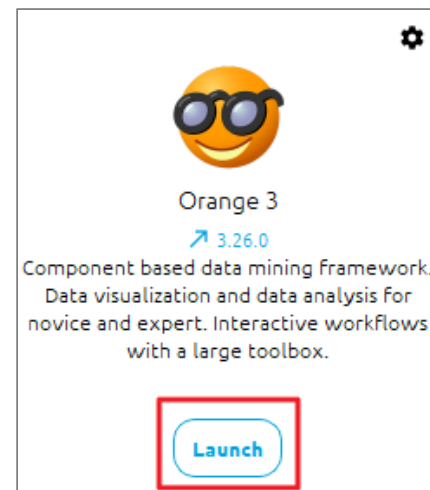
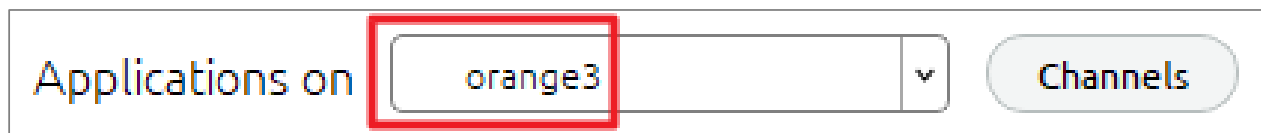
方法1 使用 Anaconda 下載安裝(2/3)

- 開啟 程式集 \ Anaconda Navigator (anaconda3)
- 選取 Orange 3
- 按下 **[Install]** 開始下載並安裝 ...



方法1 使用 Anaconda 下載安裝(3/3)

- 安裝完成後, 選取 orange3 , 按下 [\[Launch\]](#) 啟動 Orange3



方法2 下載 windows 免安裝檔 (1/2)

- <https://orangedatamining.com/download/#windows>
- Orange3-3.32.0.zip (550MB)

Download the latest version for Windows

Download Orange 3.32.0

Standalone installer (default)

Orange3-3.32.0-Miniconda-x86_64.exe (64 bit)

Can be used without administrative privileges.

Portable Orange

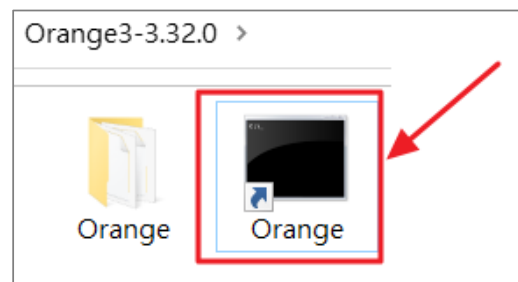
Orange3-3.32.0.zip

No installation needed. Just extract the archive and open the shortcut in the extracted folder.

採用免安裝版本, 直接使用

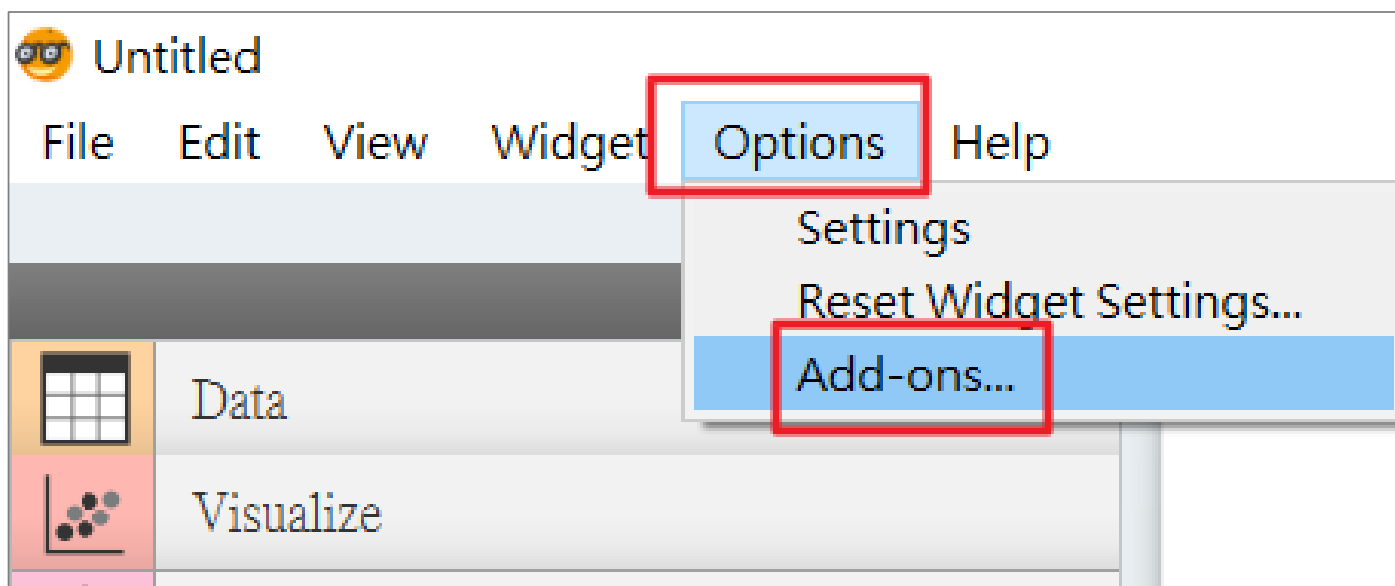
方法2 下載 windows 免安裝檔 (2/2)

- 解壓縮 zip 檔案
- 開啟 Orange 檔案
- 或是在命令提示字元輸入:
 - `python -m Orange.canvas`

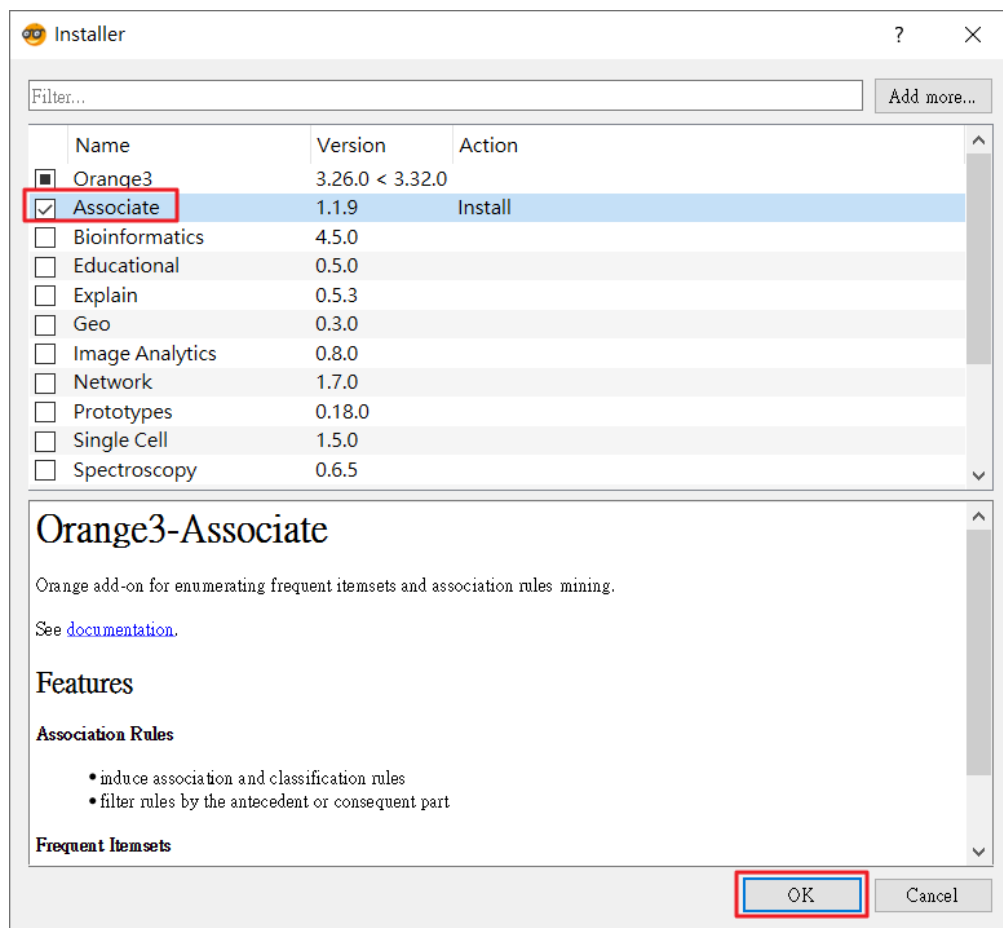


3. 安裝 Add-ons Associate

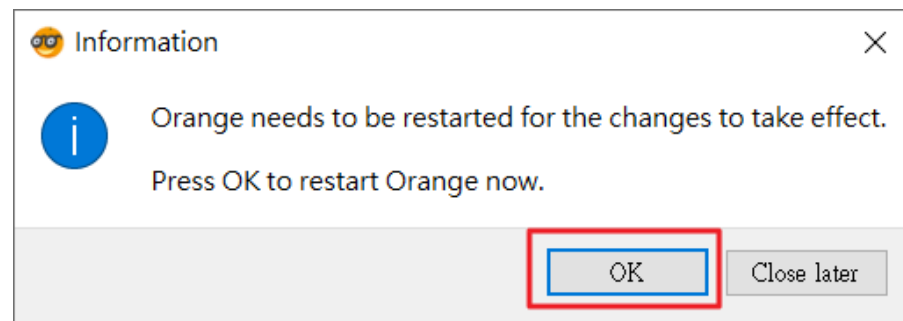
Options \ Add-ons



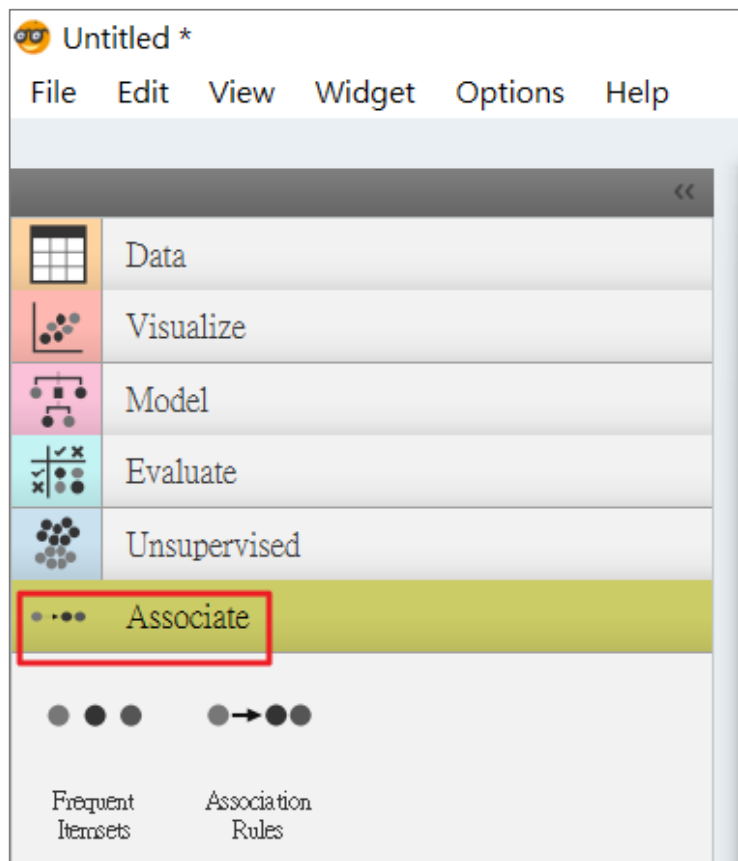
Associate 打勾 \ OK



重新啟動 Orange



Associate 安裝完成

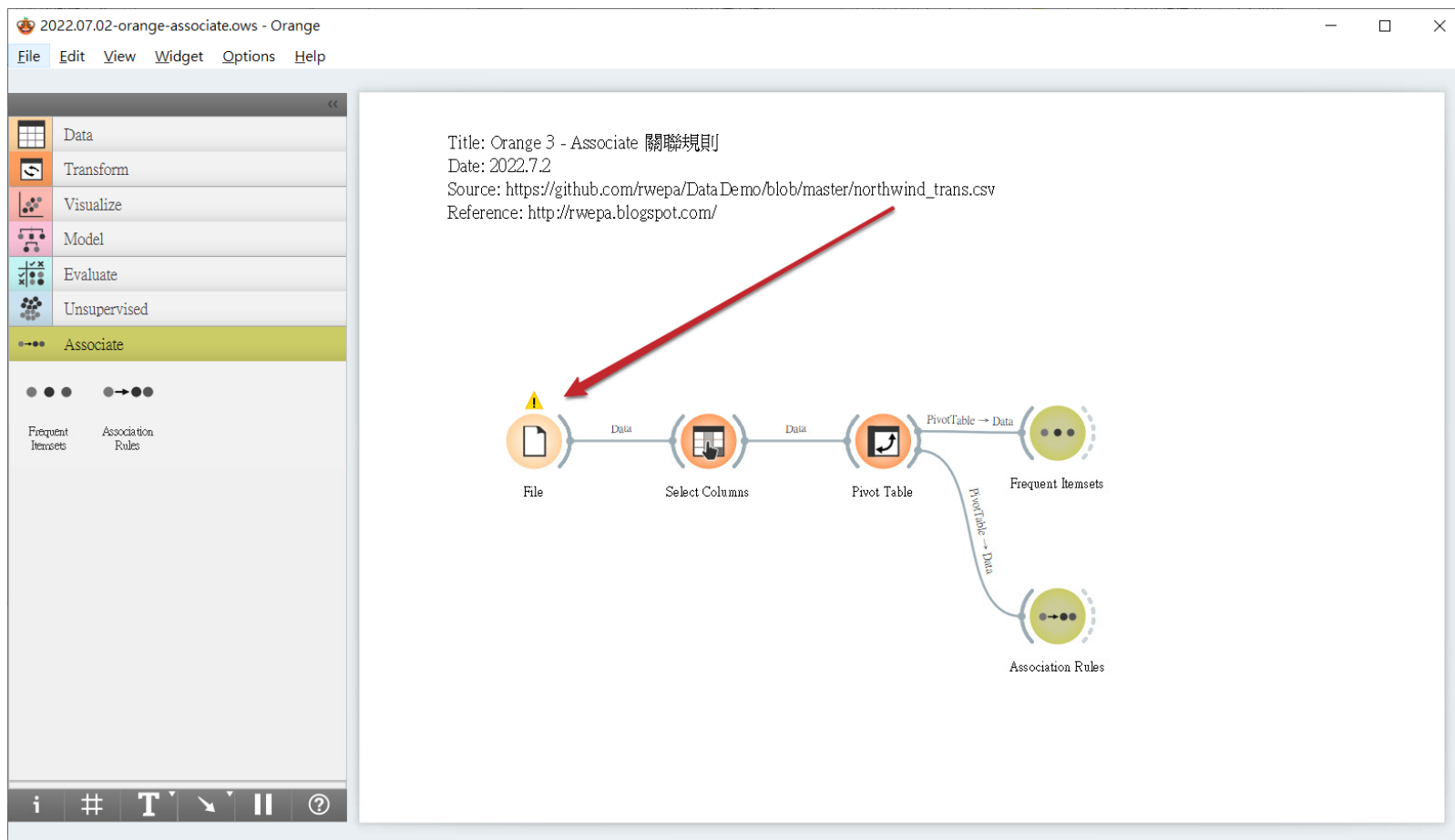


二大元件:

1. Frequent Itemsets
頻繁項目集
2. Association Rules
關聯規則

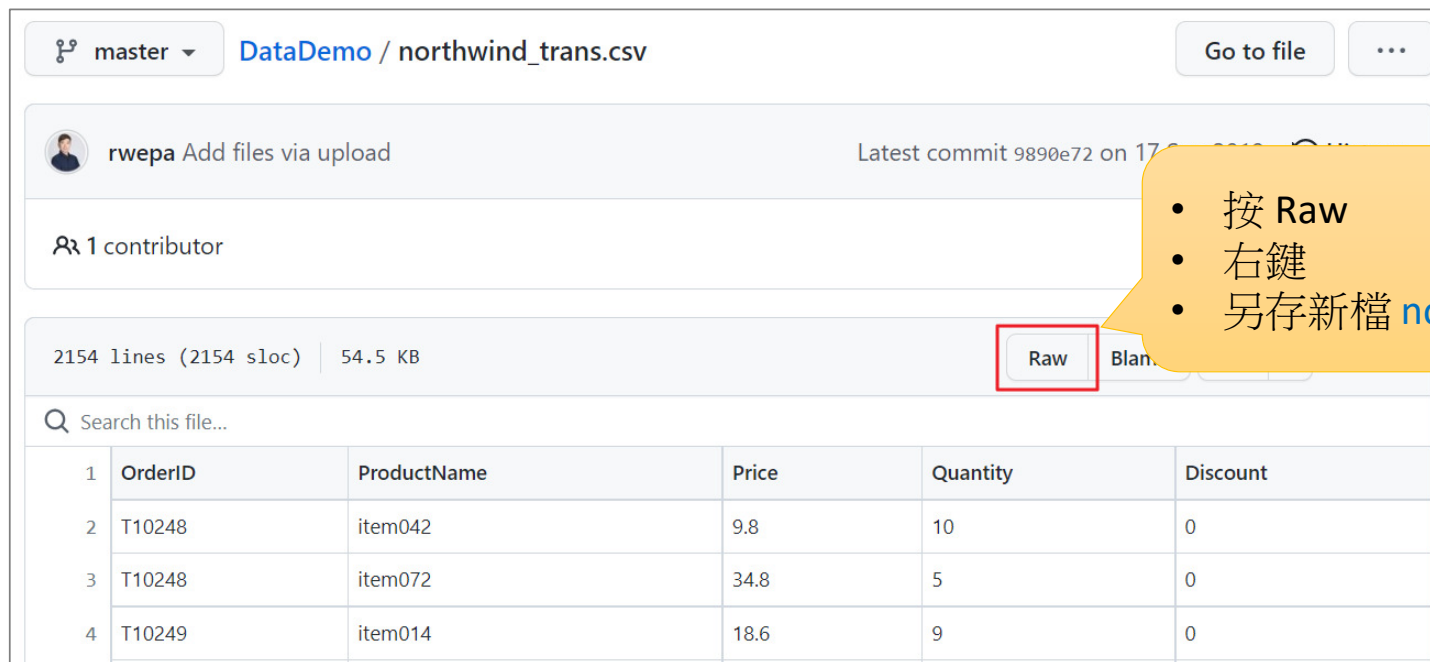
4. 案例示範

Orange demo



northwind_trans.csv

資料: https://github.com/rwepa/DataDemo/blob/master/northwind_trans.csv



master ▾ DataDemo / northwind_trans.csv

rwepa Add files via upload Latest commit 9890e72 on 17...

1 contributor

2154 lines (2154 sloc) | 54.5 KB

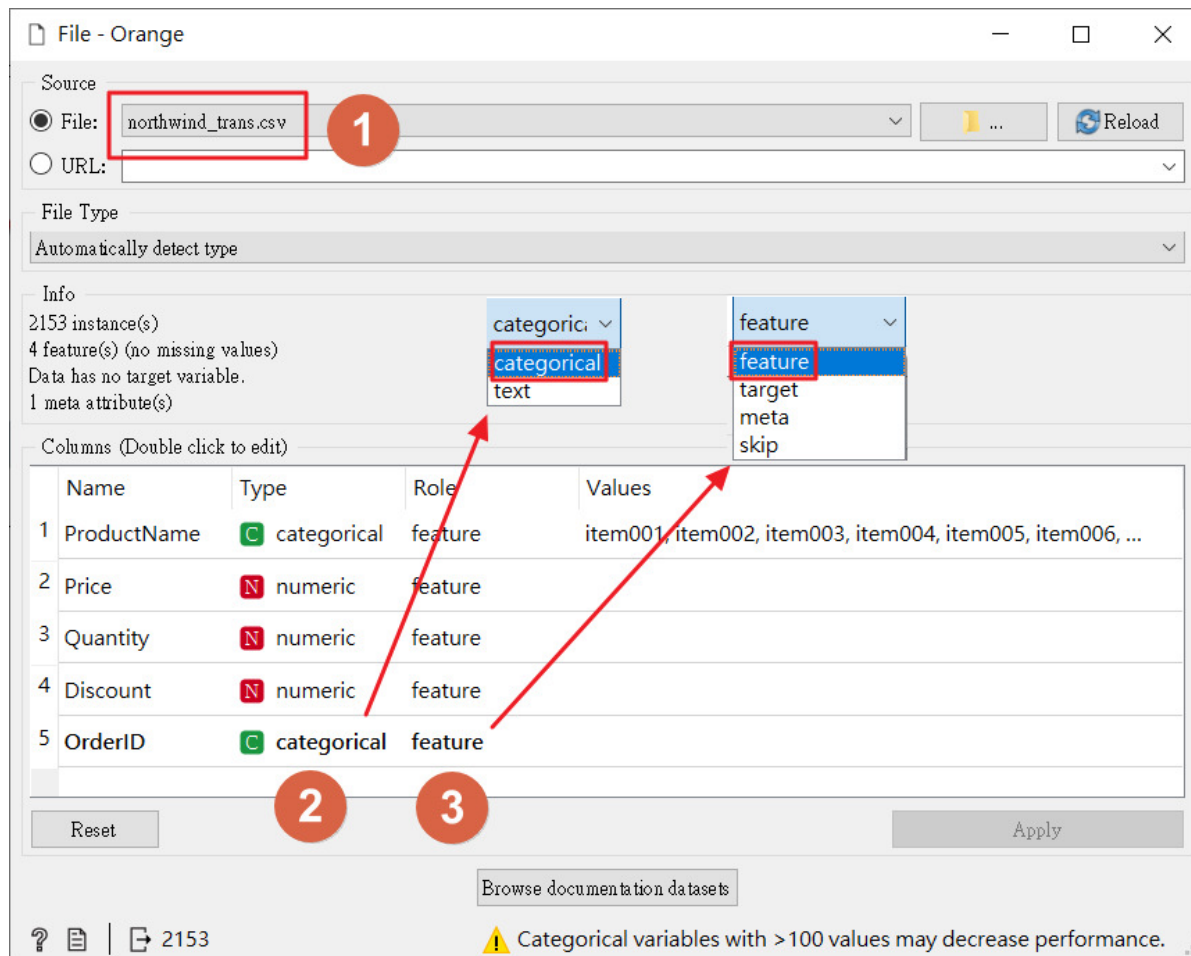
Raw Blame

Search this file...

	OrderID	ProductName	Price	Quantity	Discount
1					
2	T10248	item042	9.8	10	0
3	T10248	item072	34.8	5	0
4	T10249	item014	18.6	9	0

- 按 Raw
- 右鍵
- 另存新檔 [northwind_trans.csv](#)

File - Orange



The screenshot shows the 'File - Orange' widget interface. A red box labeled '1' highlights the 'File:' source selection. Below it, the 'File Type' is set to 'Automatically detect type'. The 'Info' section shows 2153 instances, 4 features, and no target variable. The 'Columns' table lists features: ProductName (categorical), Price (numeric), Quantity (numeric), Discount (numeric), and OrderID (categorical). Red arrows point from the 'categorical' and 'feature' columns to dropdown menus. The 'categorical' dropdown is labeled '2' and shows 'categorical' and 'text' options. The 'feature' dropdown is labeled '3' and shows 'feature', 'target', 'meta', and 'skip' options. The 'Reset' and 'Apply' buttons are at the bottom.

Source

☒ File: northwind_trans.csv **1**

☐ URL:

File Type

Automatically detect type

Info

2153 instance(s)
4 feature(s) (no missing values)
Data has no target variable.
1 meta attribute(s)

Columns (Double click to edit)

	Name	Type	Role	Values
1	ProductName	C categorical	feature	item001, item002, item003, item004, item005, item006, ...
2	Price	N numeric	feature	
3	Quantity	N numeric	feature	
4	Discount	N numeric	feature	
5	OrderID	C categorical	feature	

2 **3**

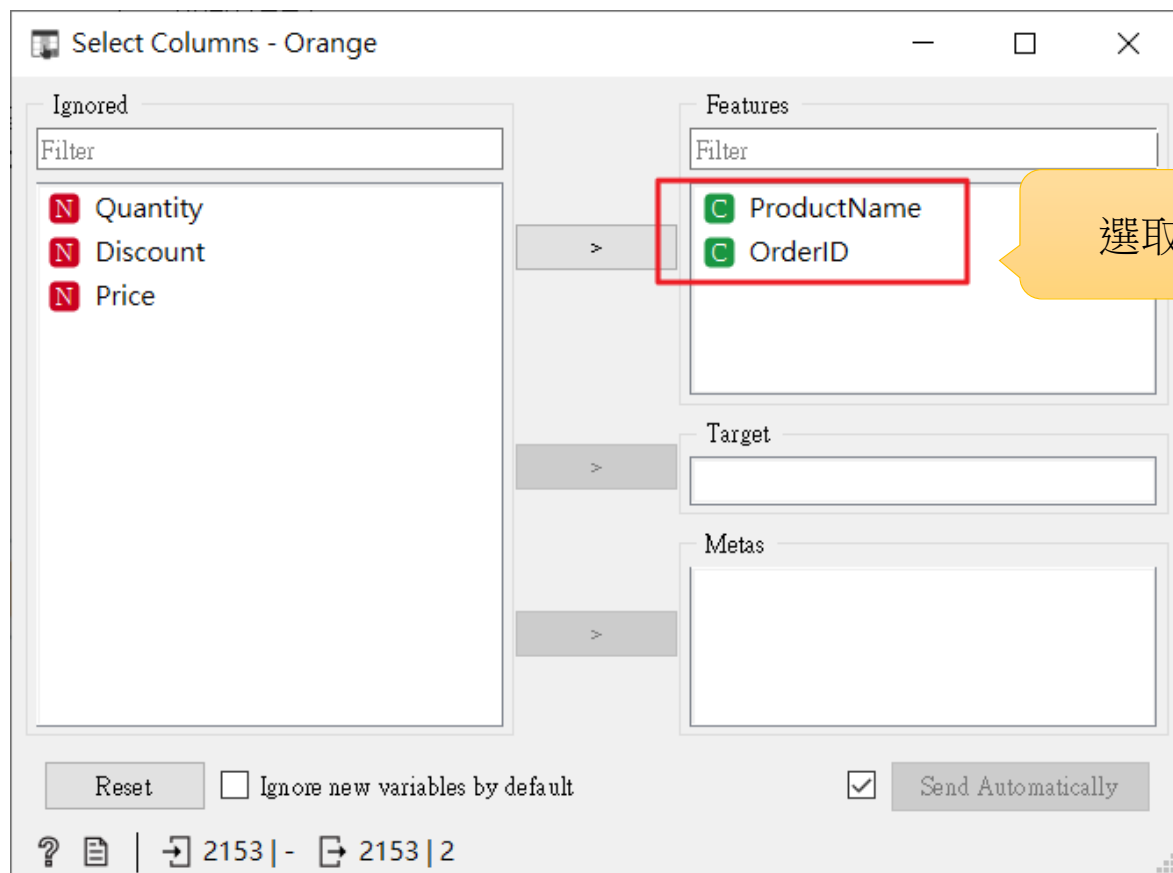
Reset Apply

Browse documentation datasets

? | 2153

⚠ Categorical variables with >100 values may decrease performance.

Select Columns – Orange (選取行)



選取2個特徵

Pivot Table – Orange (樞紐分析)

1. 列變數
2. 行變數
3. 計算值
4. 計算函數

Pivot Table - Orange

Rows: OrderID (1)

Columns: ProductName (2)

Values: ProductName (3)

Aggregations: ☒ Count (4), ☐ Count defined, ☐ Sum, ☐ Mode, ☐ Mean, ☐ Min, ☐ Var, ☐ Max, ☐ Median, ☐ Majority

☒ Apply Automatically

Count	item001	item002	item003	item004	item005	item006
T10248	0.0	0.0	0.0	0.0	0.0	0.0
T10249	0.0	0.0	0.0	0.0	0.0	0.0
T10250	0.0	0.0	0.0	0.0	0.0	0.0
T10251	0.0	0.0	0.0	0.0	0.0	0.0
T10252	0.0	0.0	0.0	0.0	0.0	0.0
T10253	0.0	0.0	0.0	0.0	0.0	0.0
T10254	0.0	0.0	0.0	0.0	0.0	0.0
T10255	0.0	1.0	0.0	0.0	0.0	0.0
T10256	0.0	0.0	0.0	0.0	0.0	0.0
T10257	0.0	0.0	0.0	0.0	0.0	0.0

? | 2153 | 830 | - | 63.9k

Frequent Itemsets – Orange (頻繁項目集)

*** Frequent Itemsets - Orange

Info
Number of itemsets: 10000
Selected itemsets: 1
Selected examples: 813

Find itemsets
Minimal support: 35%
Max. number of itemsets: 10000
☐

Filter itemsets
Contains:
Min. items: Max. items:
☒ Apply these filters in search

☒ Send Selection Automatically

? | 830 813

Itemsets	Support	%
> Aggregate=Count	830	100
✓ item009=0.0	825	99.4
✓ item015=0.0	819	98.67
> item037=0.0	813	97.95
✓ item048=0.0	813	97.95
> item066=0.0	805	96.99
> item050=0.0	803	96.75
item067=0.0	803	96.75
item074=0.0	800	96.39
> item066=0.0	811	97.71
> item027=0.0	811	97.71
> item050=0.0	809	97.47
item067=0.0	809	97.47
item074=0.0	806	97.11
> item037=0.0	819	98.67
> item048=0.0	819	98.67
> item066=0.0	817	98.43
> item027=0.0	817	98.43
> item050=0.0	815	98.19
item067=0.0	815	98.19
item074=0.0	812	97.83
> item015=0.0	824	99.28
> item037=0.0	824	99.28
> item048=0.0	824	99.28
> item066=0.0	822	98.94

Association Rules – Orange (關聯規則)

*** Association Rules - Orange

Info
Rules: 10000 (shown 10000)

Find association rules

Min. supp.: 10 %

Min. conf.: 90 %

Max. rules: 10k

☐ Induce only classification rules

☒ Restrict search by below filters

Find Rules

Filter by Antecedent

Contains:

Items, min: 2 max: 999

Filter by Consequent

Contains:

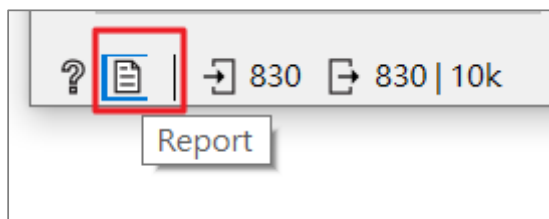
Items, min: 1 max: 999

☒ Send selection

Supp	Conf	Covr	Strg	Lift	Levr	Antecedent	Consequent
0.973	0.996	0.977	1.017	1.002	0.002	item005=0.0, item027=0.0	item009=0.0
0.973	0.979	0.994	0.983	1.002	0.002	Aggregate=Count, item009=0.0	item005=0.0, item027=0.0
0.973	0.996	0.977	1.017	1.002	0.002	Aggregate=Count, item005=0.0, ...	item009=0.0
0.973	0.996	0.977	1.017	1.002	0.002	item005=0.0, item027=0.0	Aggregate=Count, item009=0.0
0.966	0.996	0.970	1.025	1.002	0.002	item005=0.0, item015=0.0, item027=0.0	item009=0.0
0.966	0.972	0.994	0.976	1.002	0.002	Aggregate=Count, item009=0.0	item005=0.0, item015=0.0, item027=0.0
0.966	0.996	0.970	1.025	1.002	0.002	Aggregate=Count, item005=0.0, ...	item009=0.0
0.966	0.996	0.970	1.025	1.002	0.002	item005=0.0, item015=0.0, item027=0.0	Aggregate=Count, item009=0.0
0.966	0.996	0.970	1.025	1.002	0.002	item005=0.0, item027=0.0, item037=0.0	item009=0.0
0.966	0.972	0.994	0.976	1.002	0.002	Aggregate=Count, item009=0.0	item005=0.0, item027=0.0, item037=0.0
0.966	0.996	0.970	1.025	1.002	0.002	Aggregate=Count, item005=0.0, ...	item009=0.0
0.966	0.996	0.970	1.025	1.002	0.002	item005=0.0, item027=0.0, item037=0.0	Aggregate=Count, item009=0.0
0.966	0.996	0.970	1.025	1.002	0.002	item005=0.0, item027=0.0, item048=0.0	item009=0.0
0.966	0.972	0.994	0.976	1.002	0.002	Aggregate=Count, item009=0.0	item005=0.0, item027=0.0, item048=0.0
0.966	0.996	0.970	1.025	1.002	0.002	Aggregate=Count, item005=0.0, ...	item009=0.0
0.966	0.996	0.970	1.025	1.002	0.002	item005=0.0, item027=0.0, item048=0.0	Aggregate=Count, item009=0.0
0.964	0.996	0.967	1.027	1.002	0.002	item005=0.0, item027=0.0, item066=0.0	item009=0.0
0.964	0.970	0.994	0.973	1.002	0.002	Aggregate=Count, item009=0.0	item005=0.0, item027=0.0, item066=0.0
0.964	0.996	0.967	1.027	1.002	0.002	Aggregate=Count, item005=0.0, ...	item009=0.0
0.964	0.996	0.967	1.027	1.002	0.002	item005=0.0, item027=0.0, item066=0.0	Aggregate=Count, item009=0.0
0.961	0.996	0.965	1.030	1.002	0.002	item005=0.0, item027=0.0, item050=0.0	item009=0.0

830 | 10k

Report – Orange (報表)



Orange Report

檔案 | C:/rdata/rules.html

Association Rules Sat Jul 02 22, 14:43:00

Number of rules: 10000
Selected rules: 10000
Covered examples: 830

Rules

Supp	Conf	Covr	Strg	Lift	Levr	Antecedent	Consequent
0.973	0.996	0.977	1.017	1.002	0.002	item005=0.0, item027=0.0	→ item009=0.0
0.973	0.979	0.994	0.983	1.002	0.002	Aggregate=Count, item009=0.0	→ item005=0.0, item027=0.0
0.973	0.996	0.977	1.017	1.002	0.002	Aggregate=Count, item005=0.0, item027=0.0	→ item009=0.0
0.973	0.996	0.977	1.017	1.002	0.002	item005=0.0, item027=0.0	→ Aggregate=Count, item009=0.0
0.966	0.996	0.970	1.025	1.002	0.002	item005=0.0, item015=0.0, item027=0.0	→ item009=0.0
0.966	0.972	0.994	0.976	1.002	0.002	Aggregate=Count, item009=0.0	→ item005=0.0, item015=0.0, item027=0.0
0.966	0.996	0.970	1.025	1.002	0.002	Aggregate=Count, item005=0.0, item015=0.0, item027=0.0	→ item009=0.0

參考資料

- https://github.com/rwepa/orange3_associate
- R: arules package - 提供資料探勘中關聯規則apriori algorithm
 - <http://rwepa.blogspot.com/2013/01/arules-apriori-algorithm.html>
- R: arules package - 北風資料庫操作篇
 - <http://rwepa.blogspot.com/2013/01/arules-package.html>
- 第 9 章 關聯規則應用, Python 程式設計-李明昌 免費電子書
 - <http://rwepa.blogspot.com/2020/02/pythonprogramminglee.html>
- RWEPA
 - <http://rwepa.blogspot.com/>

謝謝您的聆聽

Q & A



李明昌

alan9956@gmail.com

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