

# Circle of Life

## 少子女化保衛戰

組別	指導老師	小組成員
第二組	蔡政廷	蘇彥庭
愛心組	郭惠民	任皓暉
	施丞優	姚信義
		李怡萱



# 小組成員

CEB102 AI/Big Data Team member



蘇彥庭 (組長)



姚信義



任皓暉



李怡萱

指導老師 郭惠民 . 蔡政廷 . 施丞優

# 工作分配表

Work distribution chart

	爬蟲	ETL	環境建置	指標 視覺化分析	模型 預測與分析	輿情系統	WEB	美術設計
蘇彥庭	✓	✓	✓		✓		✓	
任皓暉	✓	✓				✓		
李怡萱	✓			✓				✓
姚信義	✓	✓	✓					

# 目 錄

Contents

01 動機與目的

02 環境建置

03 資料收集

04 指標視覺化分析

05 模型預測與分析

06 輿情系統

07 未來展望



01

## 動機與目的

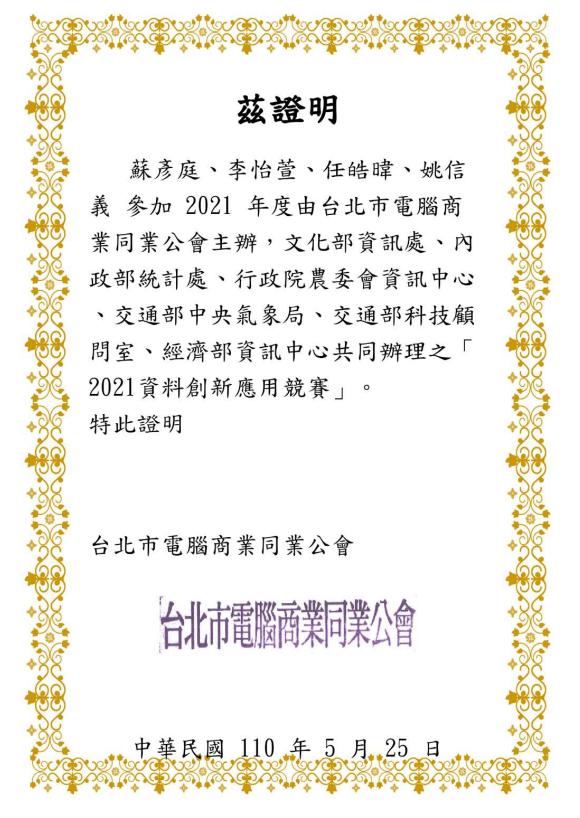
# 2021 資料創新應用競賽

- 目前已晉級**內政黑客松組**決賽
- 內政黑客松組本屆參賽隊伍約60隊，取12隊晉級決賽
- 7/7決賽報告
- 競賽官方網站連結



內政黑客松		
獎項	參賽編號	產品或服務名稱
晉級	OD-41650014	社區護理
晉級	OD-43510314	新住民於人口結構變遷之影響力
晉級	OD-39440140	「老”友”所居」危老重建金融決策支援系統
晉級	OD-20770281	研而有信
晉級	OD-10290216	安心選宅風水師
晉級	OD-12170103	Circle of Life: 少子女化保衛戰

## 2021 資料創新應用競賽參賽證明



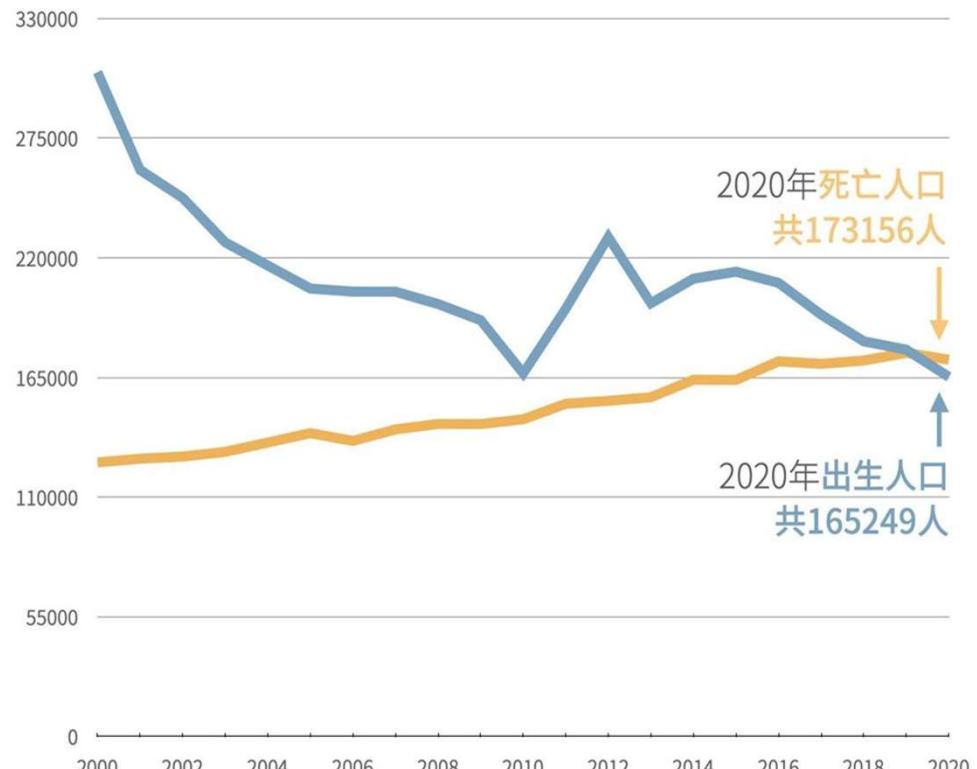
# 我國少子女化現況

 我國近年來少子女化情況日益嚴重，依據內政部2020年的人口數統計，台灣人口數首度出現負成長

 今年1月，新生兒出生數量自統計以來首度單月低於萬人

## 近20年每年出生人口持續下降

— 出生人數變化 — 死亡人數變化



製圖 : If Lin

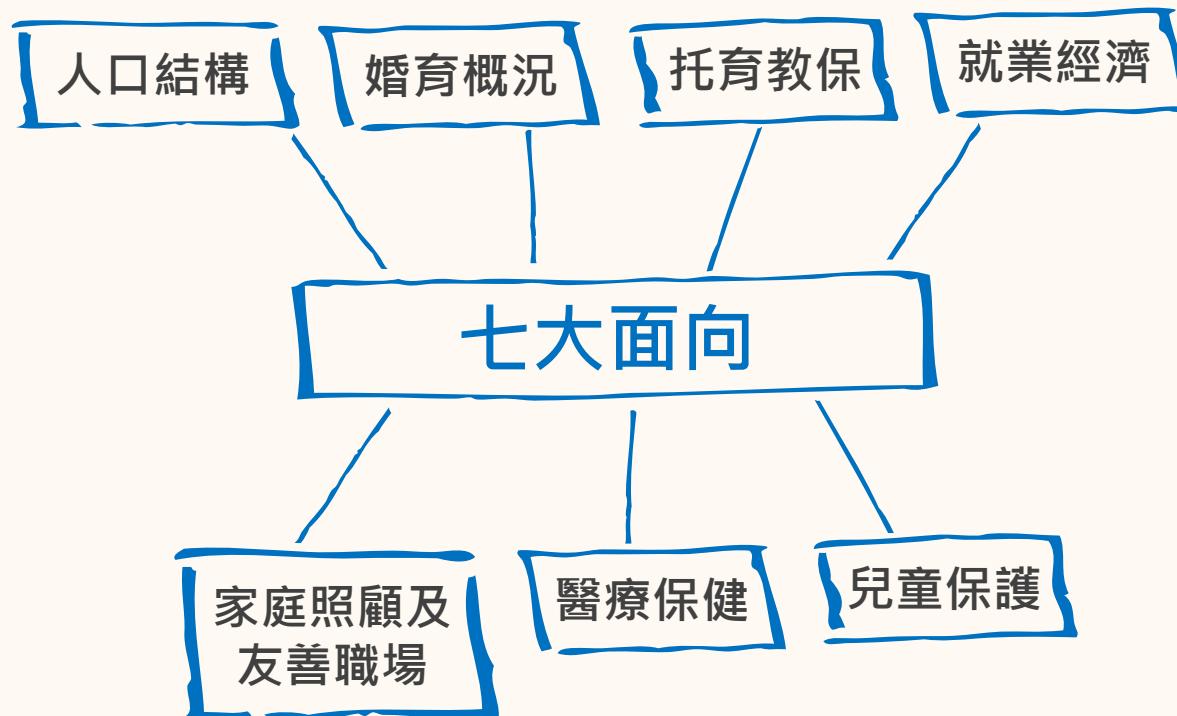
資料來源：內政部戶政司人口統計資料

# 我國少子女化現況



- 行政院-我國少子女化對策計畫(107年-113年)
- 少子女化是嚴重的國安危機
- 少子女化對我國造成四個影響：
  - 出生數持續減少，**加速**人口結構失衡
  - 在學人數**下降**，衝擊教育體系
  - 勞動人口**減少**，影響經濟發展
  - 總扶養比**增加**，青壯年人口的撫養負擔加重

# 台北市政府主計處少子女文化統計與指標項目



台北市政府主計處少子女文化統計與項目網站

[https://dbas.gov.taipei/Content\\_List.aspx?n=DAFE38DC47F17F7D](https://dbas.gov.taipei/Content_List.aspx?n=DAFE38DC47F17F7D)

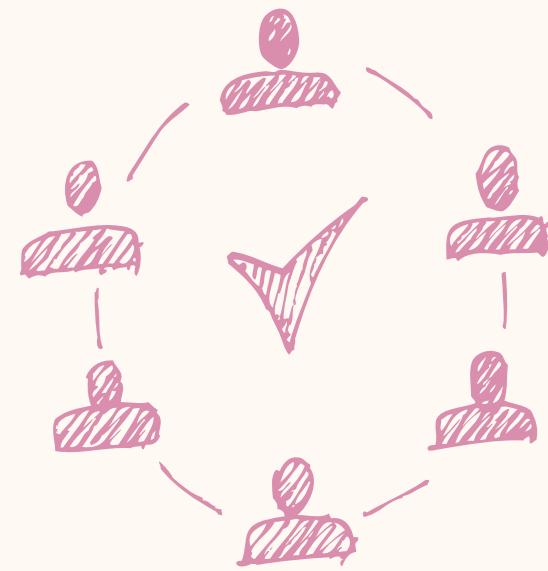
## 本專題自定義少子女化成因五大構面



# 產品服務對象與服務目標

## 本專題服務對象為政府單位

- 協助政府單位掌握各鄉鎮市區影響生育率之因素，在政策實行時能更加精準、因地制宜
- 政府單位能透過本專題開發之少子女化政策輿情系統，在少子女化政策制定或推行時，可以有民意進行參考



# 產品功能

The website interface includes a header with the title 'Circle of Life: 少子女化保衛戰' and a navigation bar with links like '首頁', '指標視覺化分析', '模型預測與分析', and '輿情系統'. Below the header is a large banner image of two children playing with bubbles. The main content area is divided into three columns:

- 指標變化分析**: Describes how low birth rates affect society, mentioning population decline, aging, and economic challenges.
- 模型預測與分析**: Details the use of machine learning models to predict birth rates across townships, mentioning PTT, Dcard, and BabyHome forums.
- 輿情系統**: Focuses on monitoring public sentiment through various media sources.

At the bottom left of the monitor, there is a credit line: '(Photo by Jacky Zhao On Unsplash)'.



## 地圖指標視覺化分析

提供各鄉鎮市區在五大面向指標的視覺化資訊與生育率比較



## 機器學習模型分析及預測

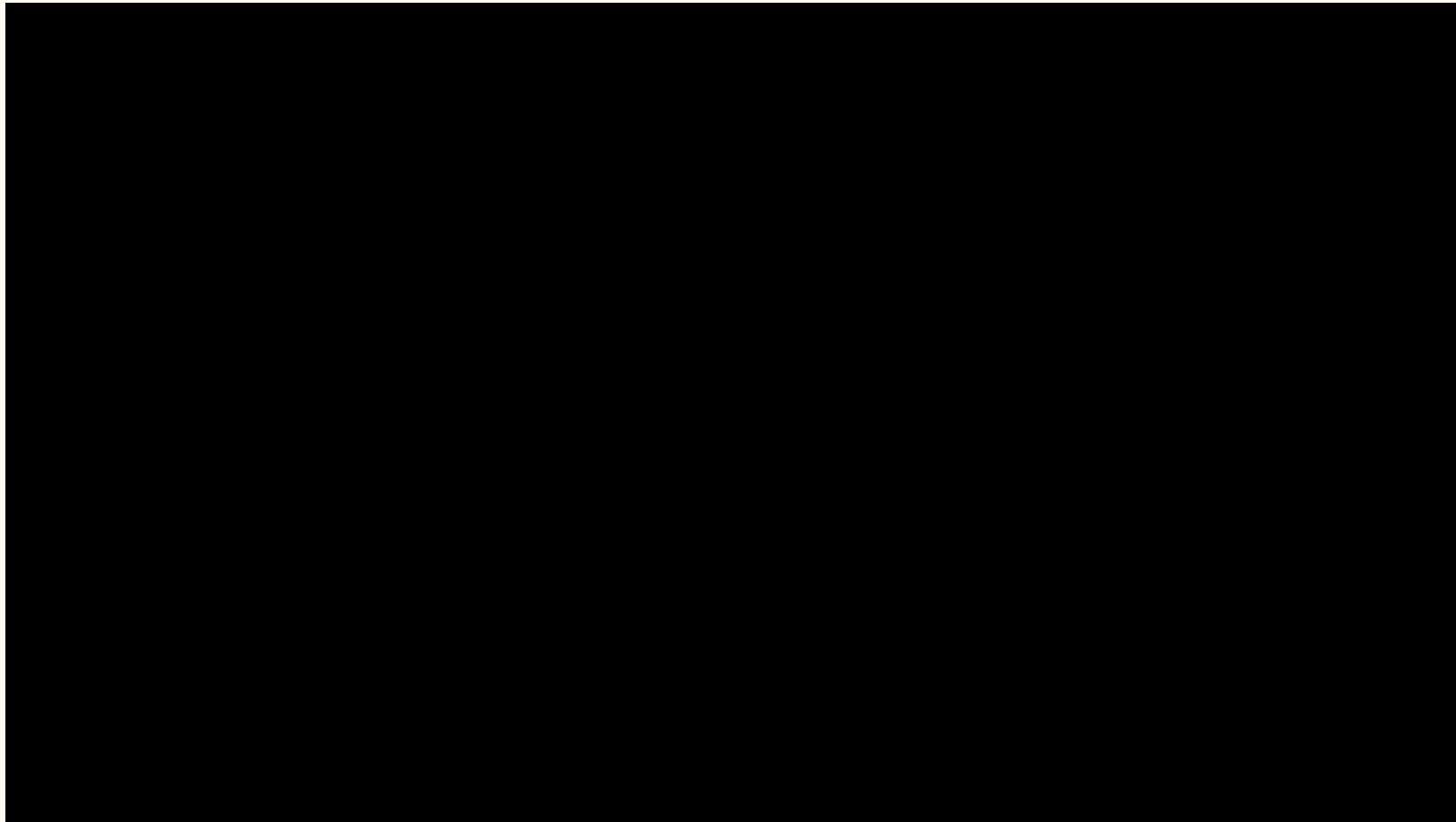
提供各式機器學習模型預測各鄉鎮市區生育率



## 輿情系統

針對少子女化相關議題做聲量及文字探勘分析，協助政府掌握政策執行狀況及聆聽民意

## 產品Demo影片



請掃描我~~  
專題 Demo 網站



Web 網址：  
<http://3.13.171.111/>



02

## 環境建置

# 專題環境

## 開發環境



WSL2(Windows Subsystem for Linux)



Ubuntu 20.04 LTS



mongoDB

## 雲端環境



WEB



Amazon EC2



Ubuntu 20.04 LTS

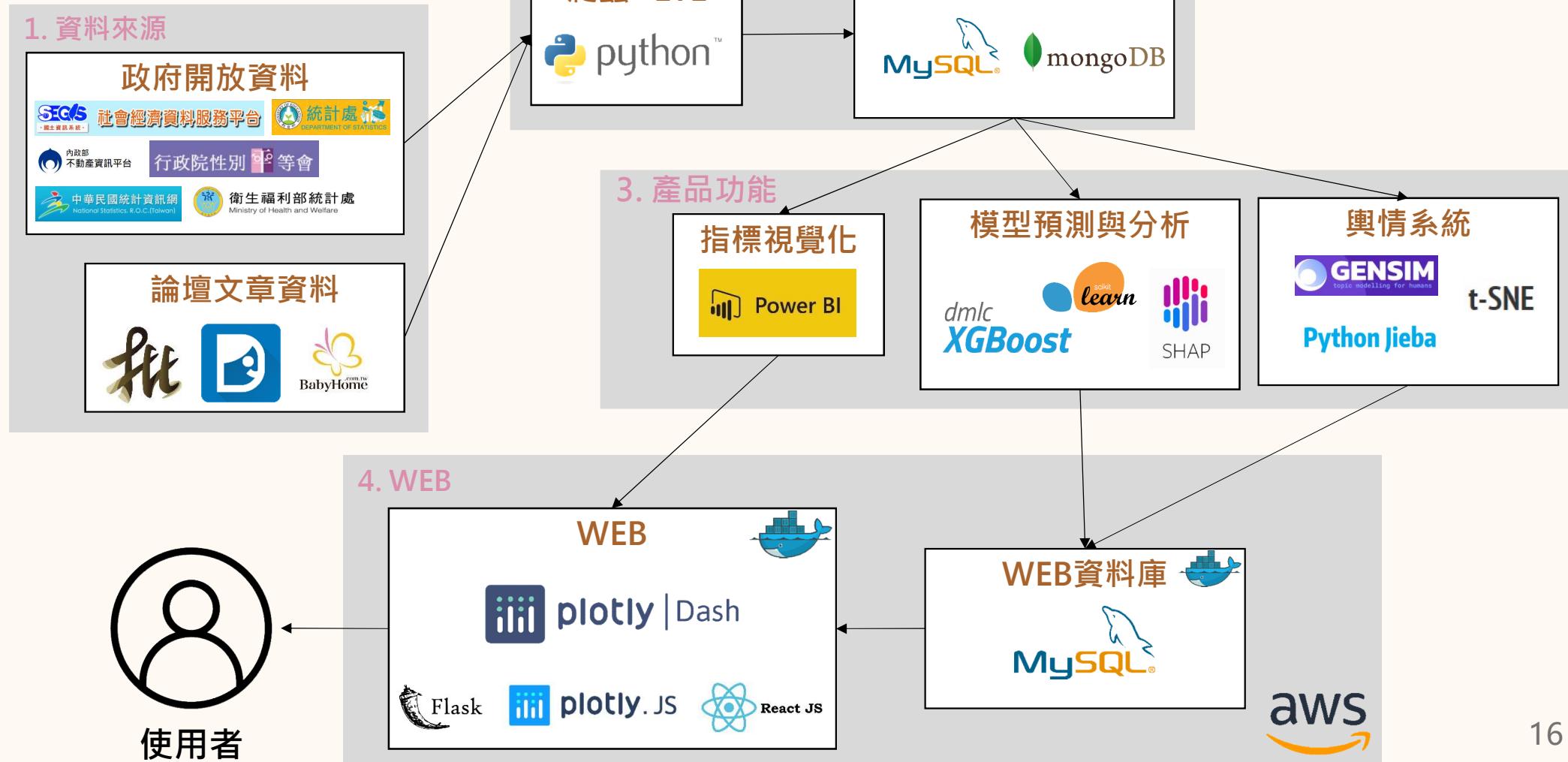
DB



MySQL

MySQL 8.0

# 專題架構



# WEB 說明

- ✓ 以Plotly公司開發的Dash套件搭建而成
- ✓ 能夠直接以Python語言撰寫客製化的數據儀表板
- ✓ Dash底層是以Flask、Plotly.js及React.js構成



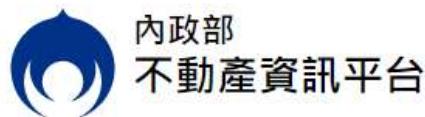


03

## 資料收集

# 資料來源

## 政府開放資料



## 論壇文章資料



# PTT 八卦板與親子板

## PPT文章標題與內文

批踢踢實業坊 > 看板 Gossiping		聯絡資訊	關於我們		
看板	精華區	最舊	上頁	下頁	最新
搜尋文章...					
3 Re: [爆卦] 反核的核三上線中 Qorqios					
		5/20	...		
6 [問卦] 183club? kevin05233		5/20	...		
7 [問卦] 戴口罩、不戴安全帽是什麼想法？？ henryhao		5/20	...		
Re: [問卦] 如果你在防疫記者會上 會問啥問題 KEND0777		5/20	...		
Re: [爆卦] 反核的核三上線中 xzcb2008		5/20	...		

資料蒐集期間：一年份的歷史資料

	資料筆數
八卦板 Gossiping	60萬筆
親子板 BabyMother	13580 筆

url_id	ptt_board	authors	datetime_list	websites	ptt_title	pushtext_col	article_content
M.16091...	Gossiping	onepunchgg...	2020-12-28 1...	https://www.ptt.cc/bb...	[問卦] 沒朋友怎麼...	1	欸幹 我不是說死邊緣胖肥宅那種沒朋友 就...
M.16091...	Gossiping	MikaHakkine...	2020-12-28 1...	https://www.ptt.cc/bb...	[新聞] 美豬來台換...		備註請放最後面 違者新聞文章刪除 1.媒體...
M.16091...	Gossiping	kingtama (...	2020-12-28 1...	https://www.ptt.cc/bb...	[問卦] 日本男生的g...	18	大概20年前 大家都嘲笑 日本人小gg 可是...
M.16091...	Gossiping	bignoob (有...	2020-12-28 1...	https://www.ptt.cc/bb...	Re: [問卦] 易利委實...		原爆卦文章導出這個結論蠻怪的 易利委只...
M.16091...	Gossiping	Janice1998...	2020-12-28 1...	https://www.ptt.cc/bb...	[新聞] 吃到飽完勝...	3	三立新聞網 吃到飽完勝無菜單料理？網喊...
M.16091...	Gossiping	brokenlin (...	2020-12-28 1...	https://www.ptt.cc/bb...	[問卦] 有沒有江亞...	2	乳題阿 江亞輪是一艘中華民國籍的輪船 1...
M.16091...	Gossiping	papaganon...	2020-12-28 1...	https://www.ptt.cc/bb...	[問卦] 台北是不是...	3	近年來台北好像除了北門破橋拆掉 似乎都...
M.16091...	Gossiping	pipi1983 (放...	2020-12-28 1...	https://www.ptt.cc/bb...	Re: [新聞] 監委質疑...	1	監委要不要查一下當時飛去日內瓦的閃尿 ...
M.16091...	Gossiping	shan820704...	2020-12-28 1...	https://www.ptt.cc/bb...	Re: [問卦] 為啥台灣...	5	台灣市場這麼小 你要有多少厲害的頻道啦 ...
M.16091...	Gossiping	pauljet (拜...	2020-12-28 1...	https://www.ptt.cc/bb...	Re: [新聞] 中國出版...		連言情小說中國玄幻都不能出版 熟 中國寫...
M.16091...	Gossiping	WizZ (我很...	2020-12-28 1...	https://www.ptt.cc/bb...	[問卦] 在欣葉吃什...	14	欸欸欣葉 老牌吧貴肥宅都知道 但其實也就...
M.16091...	Gossiping	CannedHam...	2020-12-28 1...	https://www.ptt.cc/bb...	[問卦] 日男:台男不...	23	安安 前些日子與日本同事友人們聊天 聊到...
M.16091...	Gossiping	HTC92 (中...	2020-12-28 1...	https://www.ptt.cc/bb...	[新聞] 立院新規遭...	24	1.媒體來源: 新頭殼newtalk 2.記者署名: 周...
M.16091...	Gossiping	skyofyours ...	2020-12-28 1...	https://www.ptt.cc/bb...	[新聞] 境外移入+8...	11	1. 媒體來源：公視新聞 2. 記者署名：Na...
M.16091...	Gossiping	akuuan413 (...)	2020-12-28 1...	https://www.ptt.cc/bb...	[問卦] 綠鬚蜥會是...	5	台灣中南部很多綠鬚蜥 抓不完 殺掉他...

## PPT推文內容與時間

url_id	tags	messages_id	messages	messages_time
M.1609144832.A.B7A	推	alloc	嘻嘻	12/28 16:41
M.1609144832.A.B7A	噓	sunsam	我從國小到研究所都沒有 紹你參考	12/28 16:41
M.1609144832.A.B7A	→	takechance	為何要有朋友，肥宅只需要食物	12/28 16:41
M.1609144832.A.B7A	→	a240daniel	好朋友有一兩個就好 @@	12/28 16:41
M.1609144832.A.B7A	→	a240daniel	你也不會把你的秘密告訴那麼多人吧= =	12/28 16:42
M.1609144832.A.B7A	推	fnm525	我這輩子只有一個朋友還有在聯絡	12/28 16:48
M.1609144832.A.B7A	噓	deepdish	交新朋友阿	12/28 16:57
M.1609144832.A.B7A	推	hvariables	由愛故生憂，由憂故生怖；若離於愛者，無...	12/28 22:18
M.1609144889.A.347	推	i376ers	直接八門全開	12/28 16:41
M.1609144889.A.347	推	jma306	沒有暗盤 就是純粹	12/28 16:42
M.1609144889.A.347	推	Selvage	霸氣 這樣就對了	12/28 16:42
M.1609144889.A.347	推	waloloo	豹頭 不給錢就不算嫖囉	12/28 16:42
M.1609144889.A.347	→	jma306	https://i.imgur.com/4WMgB3Y.jpg	12/28 16:43
M.1609144889.A.347	推	neilisme	看不懂他在說啥 我們現在也有進口非毒豬阿	12/28 16:43
M.1609144889.A.347	推	userpeter		12/28 16:43

# Dcard 時事板與親子板

The screenshot shows a Dcard page for a trending topic. The URL in the address bar is highlighted with a red box: [dcard.tw/f/trending/p/235138672](https://dcard.tw/f/trending/p/235138672). The main content is a news article titled "死亡人數超過出生人數 我人口首度成長 -公視新聞". The sidebar on the left lists various boards such as "感情", "COVID-19", and "心情".

## 使用 Dcard API 2.0 :

取文章ID、標題、內容、回文數、按讚數、網址等資料

- 資料蒐集時間區間為：

2020-12至今日

- 資料量：

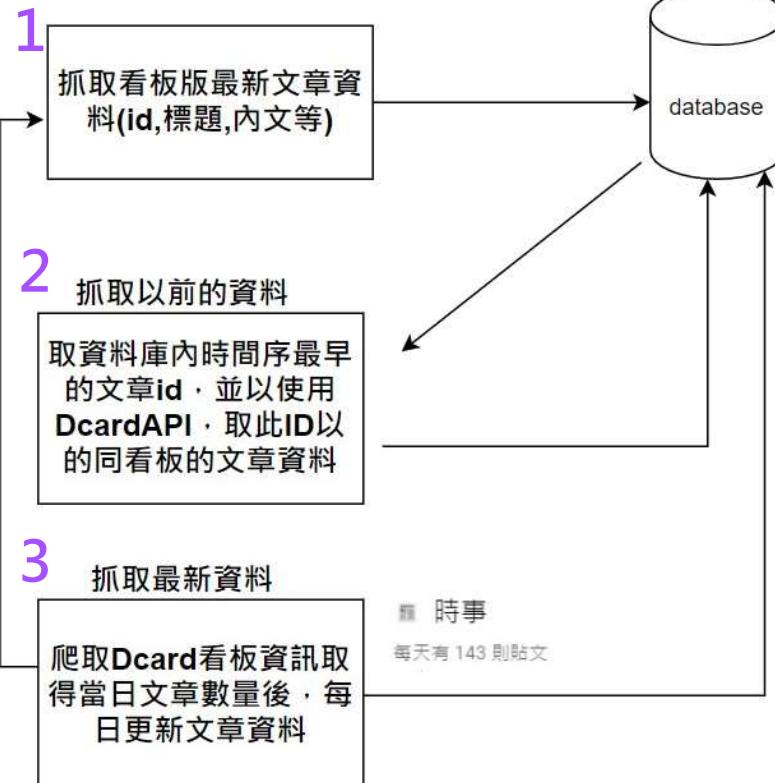
時事版約1萬1千筆

親子版約2000筆

	<b>id</b>	<b>title</b>	<b>content</b>	<b>time</b>	<b>forumna</b>	<b>commre</b>	<b>likecou</b>	<b>url</b>
	235824213	一說生育率和房價有關,好多...	<a href="https://www.storm.mg/article/362...">https://www.storm.mg/article/362...</a>	2021-04-22	時事	3	6	<a href="https://www.dcard.tw/f/trending/p/235824213">https://www.dcard.tw/f/trending/p/235824213</a>
	235820880	總生育率世界最低,台南粗生...	總生育率是計算每個婦女一生會...	2021-04-22	時事	4	4	<a href="https://www.dcard.tw/f/trending/p/235820880">https://www.dcard.tw/f/trending/p/235820880</a>
	235815914	生育率太低是因為寵物太可愛...	訊息來源:視網膜FB 生育率低的...	2021-04-21	時事	8	25	<a href="https://www.dcard.tw/f/trending/p/235815914">https://www.dcard.tw/f/trending/p/235815914</a>
	235799053	全球生育率預測台灣倒數第一	<a href="https://bit.ly/3sp3PA0">https://bit.ly/3sp3PA0</a> 大家也不...	2021-04-19	時事	161	108	<a href="https://www.dcard.tw/f/trending/p/235799053">https://www.dcard.tw/f/trending/p/235799053</a>
	235463005	如何提高生育率	<a href="https://i.imgur.com/eZ2w3uZ.jpg">https://i.imgur.com/eZ2w3uZ.jpg</a> ...	2021-03-02	時事	10	14	<a href="https://www.dcard.tw/f/trending/p/235463005">https://www.dcard.tw/f/trending/p/235463005</a>
	235148731	女權提升導致生育率下降本來...	我不懂這篇各位有什麼好吵的欸...	2021-01-10	時事	46	67	<a href="https://www.dcard.tw/f/trending/p/235148731">https://www.dcard.tw/f/trending/p/235148731</a>
	235138836	台灣女權是亞洲第一，導致台...	大家應該要知道台灣女權是亞洲...	2021-01-08	時事	1045	769	<a href="https://www.dcard.tw/f/trending/p/235138836">https://www.dcard.tw/f/trending/p/235138836</a>
	235009437	(中國)年輕人不想生孩子，民...	<a href="https://i.imgur.com/j6Wfgym.jpg">https://i.imgur.com/j6Wfgym.jpg</a> h...	2020-12-18	時事	8	11	<a href="https://www.dcard.tw/f/trending/p/235009437">https://www.dcard.tw/f/trending/p/235009437</a>

# Dcard 時事板與親子板程式架構

設定要抓取資料的看板(ex:時事版)



id	title	content	time	forumna	com	likec	u
236037307	停電怪天氣景氣太熱?	<a href="https://youtu.be/ebb9uq5kkZY">https://youtu.be/ebb9uq5kkZY</a> 全...	2021-05-19	時事	4	1	ht
236036887	大爆跌。虛擬貨幣	各位船長 該入手了嗎？哈哈哈...	2021-05-19	時事	0	1	ht
236036926	復水大約要多久時間	有台中力捧油嗎 想請問一般公...	2021-05-19	時事	3	3	ht
236036968	新道結衣♡星野源	Gakki要結婚了！感覺女神級的妹...	2021-05-19	時事	0	4	ht
236036986	我是崩潰的老師	昨天突然發佈停課後，真的是不...	2021-05-19	時事	5	13	ht
236036989	為什麼這個時後還會有人...	台灣疫情失控已經炸掉好幾天了...	2021-05-19	時事	1	1	ht
236037048	中部拉長停水時間的原因	因為都拿去做發電啦，為了我們...	2021-05-19	時事	3	11	ht
236037058	疫情走向	各位覺得到月底 大專院校會恢...	2021-05-19	時事	5	3	ht
236037119	實名制紀錄	<a href="https://i.imgur.com/3jpoZwi.jpg">https://i.imgur.com/3jpoZwi.jpg</a> 現...	2021-05-19	時事	0	1	ht
236037125	台灣缺電全解析（轉自PT...	[爆卦] 台灣缺電全解析 <a href="http://w...">http://w...</a>	2021-05-19	時事	5	11	ht
236037135	持續更新！公告遠距至「...	△ 鑑於近日疫情尚未趨緩，已...	2021-05-19	時事	7	11	ht
236037151	實聯制真的有用嗎？商家...	我住在台北市，家裡的一樓是商...	2021-05-19	時事	0	1	ht
236037161	相信政府相信黨，1100床...	<a href="https://i.imgur.com/cmg6H1f.jpg">https://i.imgur.com/cmg6H1f.jpg</a> ...	2021-05-19	時事	13	9	ht
236036818	國民黨民進黨人才差別	以下是看人家整理的 1. 國民黨...	2021-05-19	時事	30	22	ht
236037369	火災因素統計	消防局統計，菸蒂佔致傷亡火災...	2021-05-19	時事	0	4	ht
236037461	彭博：自滿心理戳破台灣...	<a href="https://udn.com/news/story/1217...">https://udn.com/news/story/1217...</a>	2021-05-19	時事	8	5	ht
236037638	疫情帶給大眾運動業的悲哀	第一次這麼有感 對在客運業也...	2021-05-19	時事	4	1	ht

設定工作排程：每日取各版最新資料，同時更新資料庫內容

# 政府開放資料

資料名稱	資料期間	資料筆數	資料來源
內政大數據資料應用組競賽用資料集_村里	104年-109年	42,342筆*166欄位	 <b>社會經濟資料服務平台</b> ·國土資訊系統·
內政大數據模擬資料	108年-109年	4,195,812筆*30欄位	 <b>社會經濟資料服務平台</b> ·國土資訊系統·
各縣市每季房價所得比、貸款負擔率	100年Q1-109年Q4	840筆*7欄位	 內政部 不動產資訊平台
台灣地區縣市別失業率	104年-109年	150筆*5欄位	 中華民國統計資訊網 National Statistics, R.O.C.(Taiwan)
各縣市各年齡層女性勞動參與率、首次生產婦女之平均年齡	102年-109年	200筆*26欄位	行政院性別平等會
幼兒(稚)園概況表	102年-109年	200筆*6欄位	 統計處 DEPARTMENT OF STATISTICS
各縣市托嬰中心所數及收托人數	104年-109年	132筆*9欄位	 衛生福利部統計處 Ministry of Health and Welfare

# 處理政府資料難點

以教育部統計處幼兒(稚)園概況表(80~109 學年度)為例

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V		
幼兒園概況-學校所在地區別																							
Summary of Preschool -By Location of Schools																							
109學年度 SY2020-2021																							
單位: 所 ; 人 (Unit : School ; Person)																							
縣市別 District		園數 No. of Schools				園長人數 No. of Principals				教師數 No. of Teachers		教保員數 No. of Educare Givers		助理教保員數 No. of Educare Assistants		服務人員數 No. of Care Givers		職員數 No. of Staffs		幼生數 No. of Children			
		公立 Public	分班 Branches	非營利 Non-profit preschools	準公共 Quasi-public Educare			公立 Public			公立 Public			公立 Public			公立 Public			公立 Public			
總 計 Grand Total	A	6,447	2,104	241	(373)	737	1,262	4,342	204	14,203	8,670	34,148	6,170	4,074	173	4	- 18,306	4,410	583,406	160,363			
	B																						
	C	580,911																			158,152		
	D	89,211																			27,181		
	E	58,312																			20,204		
	F	61,674																			12,430		
幼兒園概況-學校所在地區別																							
Summary of Preschool -By Location of Schools																							
108學年度 SY2019-2020																							
單位: 所 ; 人 (Unit : School ; Person)																							
縣市別 District		園數 No. of Schools				園長人數 No. of Principals				教師數 No. of Teachers		教保員數 No. of Educare Givers		助理教保員數 No. of Educare Assistants		服務人員數 No. of Care Givers		職員數 No. of Staffs		幼生數 No. of Children			
		公立 Public	分班 Branches			公立 Public			公立 Public			公立 Public			公立 Public			公立 Public			公立 Public		
總 計 Grand Total	A	6,384	2,081	259	(381)	4,271	H	205	13,698	8,238	31,906	6,095	3,866	172	6	- 17,411	4,338	564,545	158,210				
	B																						
	C	156,038																			26,955		
	D	19,980																			12,467		
幼兒園概況-學校所在地區別																							
Summary of Preschool -By Location of Schools																							
102學年度 SY2013-2014																							
單位: 園 ; 人 (Unit : School ; Person)																							
縣市別 District		園數 No. of Schools				園長人數 No. of Principals				教師數 No. of Teachers		教保員數 No. of Educare Givers		助理教保員數 No. of Educare Assistants		服務人員數 No. of Care Givers		職員數 No. of Staffs		幼生數 No. of Children			
		公立 Public	分班 Branches			公立 Public			公立 Public			公立 Public			公立 Public			公立 Public			公立 Public		
總 計 Grand Total	A	6,560	1,919	(492)	4,612	236	12,359	6,155	25,216	5,716	3,101	352	8	- 14,552	3,682	448,189	131,910						
	B																						
	C	6,528	1,804	(407)	4,604	235	12,245	6,041	25,172	5,645	3,005	357	8	- 14,408	3,637	446,212	130,117						

資料欄位不一致

專有名詞的意義

專設? 非營利? 準公共?

# 處理政府資料難點

以內政部戶政司育齡婦女之年齡別生育率資料為例

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S
報表產製時間: 2021/04/18 13:42																		
單位: %																		
縣市別	統計																	
年齡別		15-19歲	20-24歲	25-29歲	30-34歲	35-39歲	40-44歲	45-49歲	15-19歲	新北市								
84年		17	86	148	82	20	2	0-	-		-	-	-	-	-	14	72	140
85年		17	83	145	84	21	2	0-	-		-	-	-	-	-	13	68	131
86年		15	80	147	87	22	3	0-	-		-	-	-	-	-	11	65	131
87年		14	66	116	73	21	3	0-	-		-	-	-	-	-	11	54	102
88年		13	66	126	82	21	3	0-	-		-	-	-	-	-	10	53	109
89年		14	72	133	90	24	3	0-	-		-	-	-	-	-	11	58	118
90年		13	62	106	75	21	3	0-	-		-	-	-	-	-	10	50	95
91年		13	57	102	73	20	3	0-	-		-	-	-	-	-	10	47	88
92年		11	52	92	69	20	3	0-	-		-	-	-	-	-	8	41	82



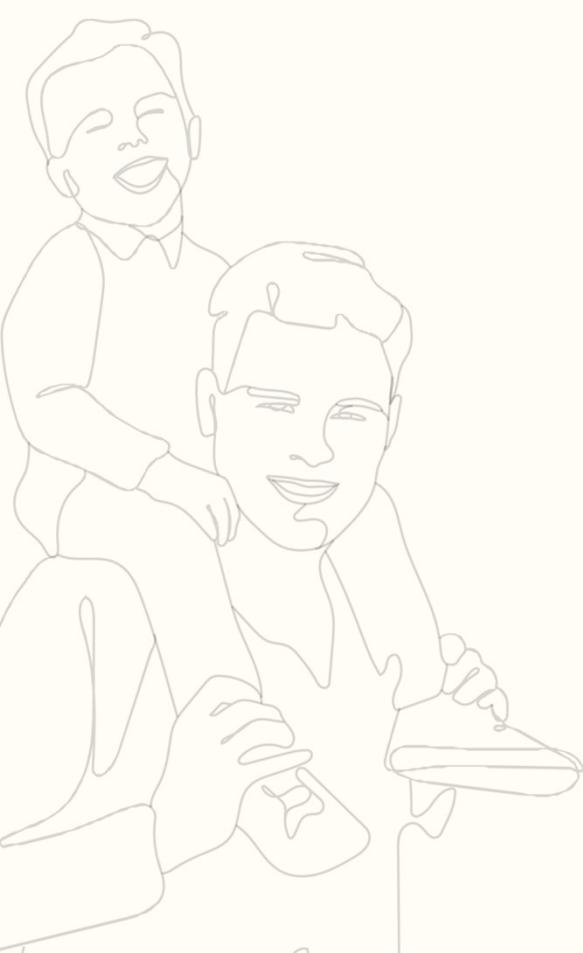
年份	縣市別	15-19歲 生育率	20-24歲 生育率	25-29歲 生育率	30-34歲 生育率	40-44歲 生育率	45-49歲 生育率
101	台北市						
101	新北市						
101	桃園市						



資料不符合正規化



直轄市改制、縣市合併升格問題



# 04

## 指標視覺化分析

# 架構流程圖



# Power BI 五大構面指標

## 人口結構

- 出生率
- 自然增加率
- 社會增加率
- 老化指數

## 婚育概況

- 結婚率
- 生母生育平均年齡
- 生母第一胎平均年齡

## 托育教保

- 公幼生佔比
- 公幼家數佔比
- 公辦民營托嬰中心佔比

## 就業經濟

- 失業率
- 可支配所得中位數(元)
- 各年齡層女性勞動參與率

## 房市狀況

- 房價所得比
- 房貸負擔率(%)
- 住宅自有率(%)

## 五大構面(橫斷面+時間序列)

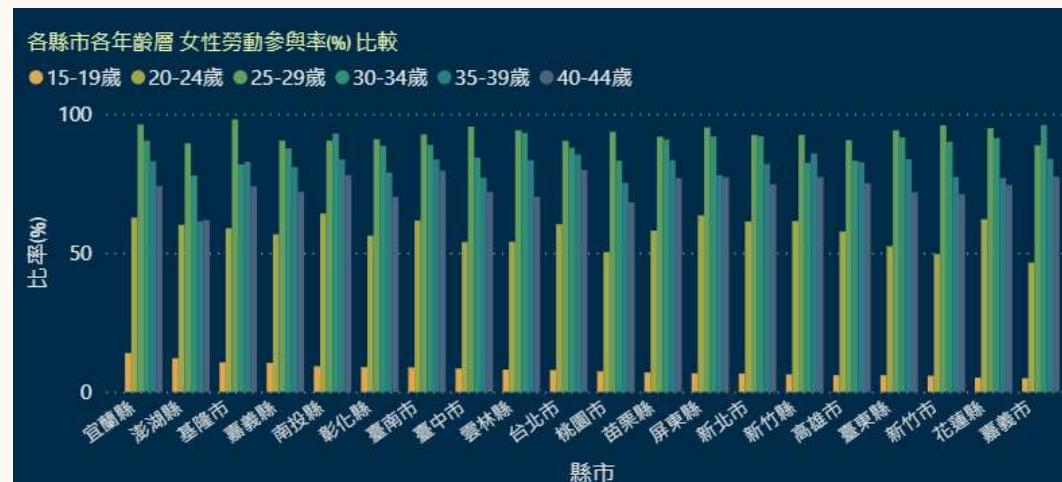


連結 MySQL 資料庫 → 取得資料

## 運用 Power BI 製作數據儀錶板



# Power BI 各指標繪圖呈現



# Power BI 成果展示

## 指標視覺化分析 網站展示





05

## 模型預測與分析

# 預測目標



**108年已婚適育年齡女性是否有小孩？兩分類問題**



透過預測已婚適育年齡女性是否有小孩，來分析影響生小孩之因素



以「有小孩」為1分類，「沒小孩」為0分類

# 資料基底

以內政大數據模擬資料集的**108年人+建物+地理資訊模擬資料**為資料基底

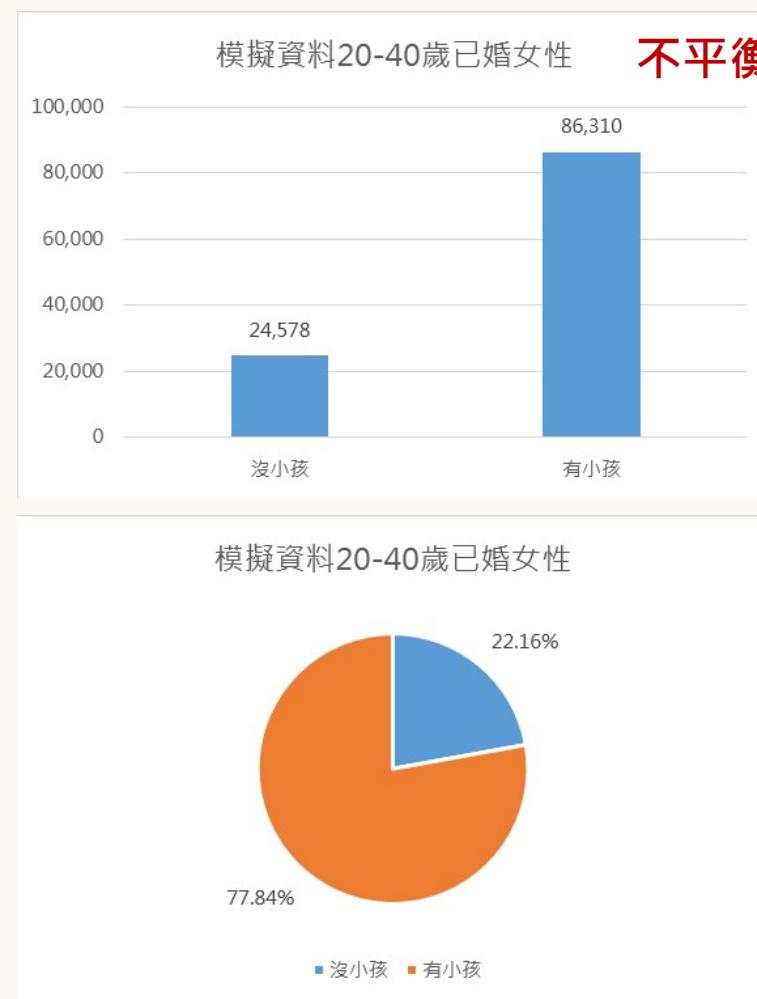
母體資料共有23,596,098筆(即全台人口資料) · 依據母體結構模擬約10%之樣本資料 · 共計有2,359,385筆資料

此模擬資料包含

- 個人資訊：如戶籍資訊、性別、教育程度、婚姻狀況及有無殼等
- 建物：如建物層數、建物用途、建物面積及屋齡等
- 地理：如斷層帶、土壤液化及順逆向坡等



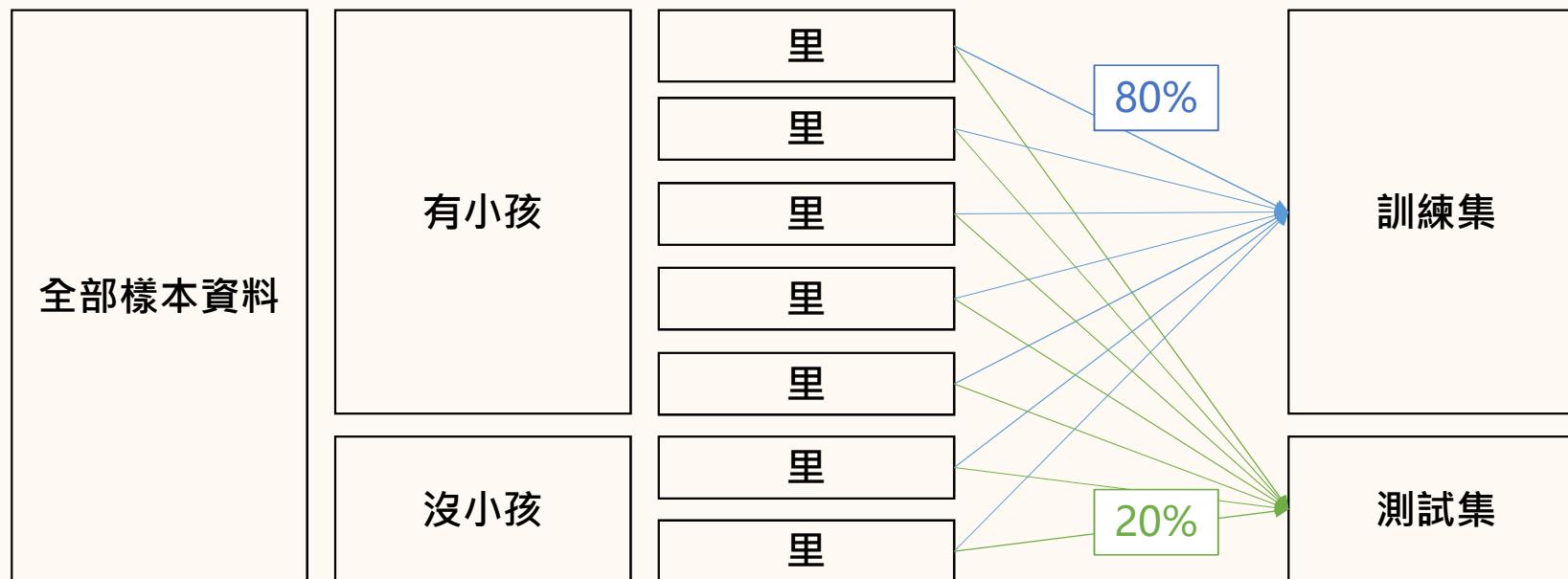
# 樣本篩選



註1：若樣本在教育程度、出生子女數及有殼分類欄位中有任一缺值，則刪除該樣本

# 訓練集/測試集切割

對110,888筆樣本採分層抽樣，依相同分類及相同戶籍里的樣本組，隨機切分80/20做為訓練集及測試集



# 訓練集/測試集切割

資料類別	分類	樣本數	比率
訓練集	有小孩	62,001	77.61%
	沒小孩	17,891	22.39%
	總計	79,892	100%
測試集	有小孩	18,220	78.81%
	沒小孩	4,900	21.19%
	總計	23,120	100%

# 特徵資料說明

特徵編號	特徵名稱	資料來源	資料層集	提供單位	備註
1	年紀	人+建物+地理資訊模擬資料	個人	內政部	
2	教育程度	人+建物+地理資訊模擬資料	個人	內政部	
3	是否為原住民身分	人+建物+地理資訊模擬資料	個人	內政部	
4	是否為身心障礙	人+建物+地理資訊模擬資料	個人	內政部	
5	是否為中低收入戶	人+建物+地理資訊模擬資料	個人	內政部	
6	是否有殼	人+建物+地理資訊模擬資料	個人	內政部	
7	人口密度	SEGIS行政區資料	里	內政部	
8	老化指數	SEGIS行政區資料	里	內政部	
9	社會增加率	SEGIS行政區資料	里	內政部	
10	房價所得比	內政部不動產資訊平台	縣市	內政部	
11	失業率	內政部不動產資訊平台	縣市	主計總處	
12	公幼生佔比	教育部統計處	縣市	教育部	計算方式: 縣市公幼生人數/縣市幼兒園學生總人數
13	托嬰比率	衛福部統計處	縣市	衛福部	計算方式: 縣市托嬰總人數/縣市0-2歲人口數
14	各年齡區間女性勞動參與率	中華民國統計資訊網	縣市	主計總處	

# 特徵相關係數矩陣

	年紀	教育程度	原住民身分	身心障礙	中低收入戶	是否有殼	人口密度	老化指數	社會增加率	房價所得比	失業率	公幼生佔比	托嬰比率	女性勞動參與率	預測目標Y
年紀	1.00	-0.08	-0.00	0.02	0.01	-0.00	0.00	0.01	0.00	-0.01	0.00	0.01	0.00	0.37	0.04
教育程度	-0.08	1.00	-0.08	-0.07	-0.04	0.14	0.10	-0.04	-0.00	0.14	-0.00	-0.01	0.10	0.03	-0.14
原住民身分	-0.00	-0.08	1.00	0.01	0.03	-0.07	-0.05	-0.01	-0.01	-0.05	-0.03	0.09	-0.06	-0.01	0.03
身心障礙	0.02	-0.07	0.01	1.00	0.01	-0.02	-0.02	0.02	-0.00	-0.01	-0.01	0.01	-0.02	-0.00	-0.00
中低收入戶	0.01	-0.04	0.03	0.01	1.00	-0.04	-0.02	0.02	-0.02	-0.02	-0.03	0.01	-0.04	0.00	0.02
是否有殼	-0.00	0.14	-0.07	-0.02	-0.04	1.00	0.04	-0.06	0.03	0.01	0.08	-0.06	0.05	-0.02	-0.04
人口密度	0.00	0.10	-0.05	-0.02	-0.02	0.04	1.00	-0.11	-0.07	0.48	0.29	0.09	0.43	0.04	-0.05
老化指數	0.01	-0.04	-0.01	0.02	0.02	-0.06	-0.11	1.00	-0.21	-0.08	-0.15	0.22	-0.17	0.04	-0.01
社會增加率	0.00	-0.00	-0.01	-0.00	-0.02	0.03	-0.07	-0.21	1.00	-0.11	0.13	-0.11	0.01	-0.02	0.00
房價所得比	-0.01	0.14	-0.05	-0.01	-0.02	0.01	0.48	-0.08	-0.11	1.00	0.16	0.25	0.69	0.09	-0.06
失業率	0.00	-0.00	-0.03	-0.01	-0.03	0.08	0.29	-0.15	0.13	0.16	1.00	-0.23	0.45	-0.03	-0.02
公幼生佔比	0.01	-0.01	0.09	0.01	0.01	-0.06	0.09	0.22	-0.11	0.25	-0.23	1.00	-0.12	0.08	-0.02
托嬰比率	0.00	0.10	-0.06	-0.02	-0.04	0.05	0.43	-0.17	0.01	0.69	0.45	-0.12	1.00	0.06	-0.05
女性勞動參與率	0.37	0.03	-0.01	-0.00	0.00	-0.02	0.04	0.04	-0.02	0.09	-0.03	0.08	0.06	1.00	-0.01
預測目標Y	0.04	-0.14	0.03	-0.00	0.02	-0.04	-0.05	-0.01	0.00	-0.06	-0.02	-0.02	-0.05	-0.01	1.00

# 建立預測模型

*dmrc*  
**XGBoost**

- 預測模型採用極限梯度提升模型(eXtreme Gradient Boosting; XGBoost)
- 由於資料屬於不平衡資料(Imbalanced Data) · 此處設定XGBoost正樣本權重(scale\_pos\_weight)參數為0.2886(即訓練集預測分類為0的樣本數/分類為1的樣本數)
- 採用交叉驗證方式 · 隨機對訓練集資料切5折 · 以AUC為準則挑選最佳超參數組合 · 超參數組合共計有64組：

超參數中文名稱	超參數英文名稱	超參數選擇範圍
學習比率	learning_rate	0.01, 0.1
最大深度	max_depth	3, 4, 5, 6, 7, 8, 9, 10
特徵隨機抽取比率	colsample_bytree	0.7, 0.9
樣本隨機抽取比率	subsample	0.7, 0.9

# 建立預測模型

*dmrc*  
**XGBoost**

- 經交叉驗證篩選後，最佳超參數組合為：

超參數中文名稱	超參數英文名稱	超參數選擇範圍
學習比率	learning_rate	0.1
最大深度	max_depth	10
特徵隨機抽取比率	colsample_bytree	0.9
樣本隨機抽取比率	subsample	0.7
最佳疊代數	num_boost_round	22

- 訓練集以此最佳參數組合重新訓練，得出最後預測模型。

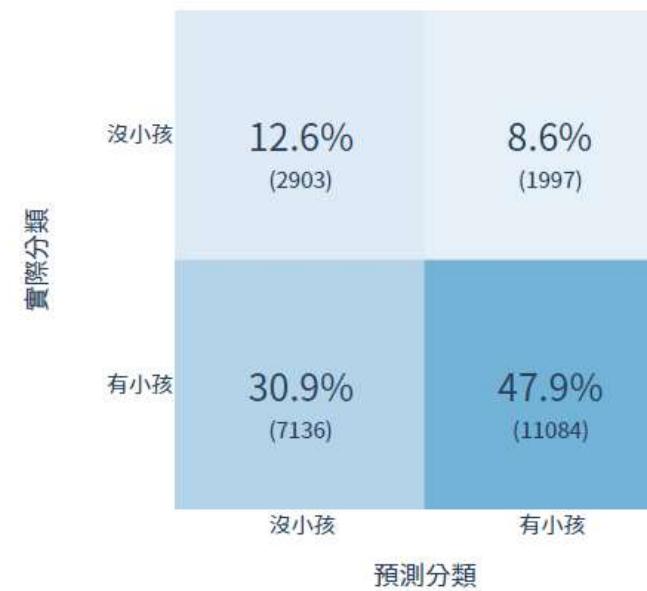
# 預測模型績效

*dmrc*  
**XGBoost**

指標	訓練集	測試集
accuracy	0.6176	0.6050
precision	0.8815	0.8473
recall	0.5861	0.6083
f1	0.7040	0.7082
roc_auc	0.7231	0.6313

## 測試集

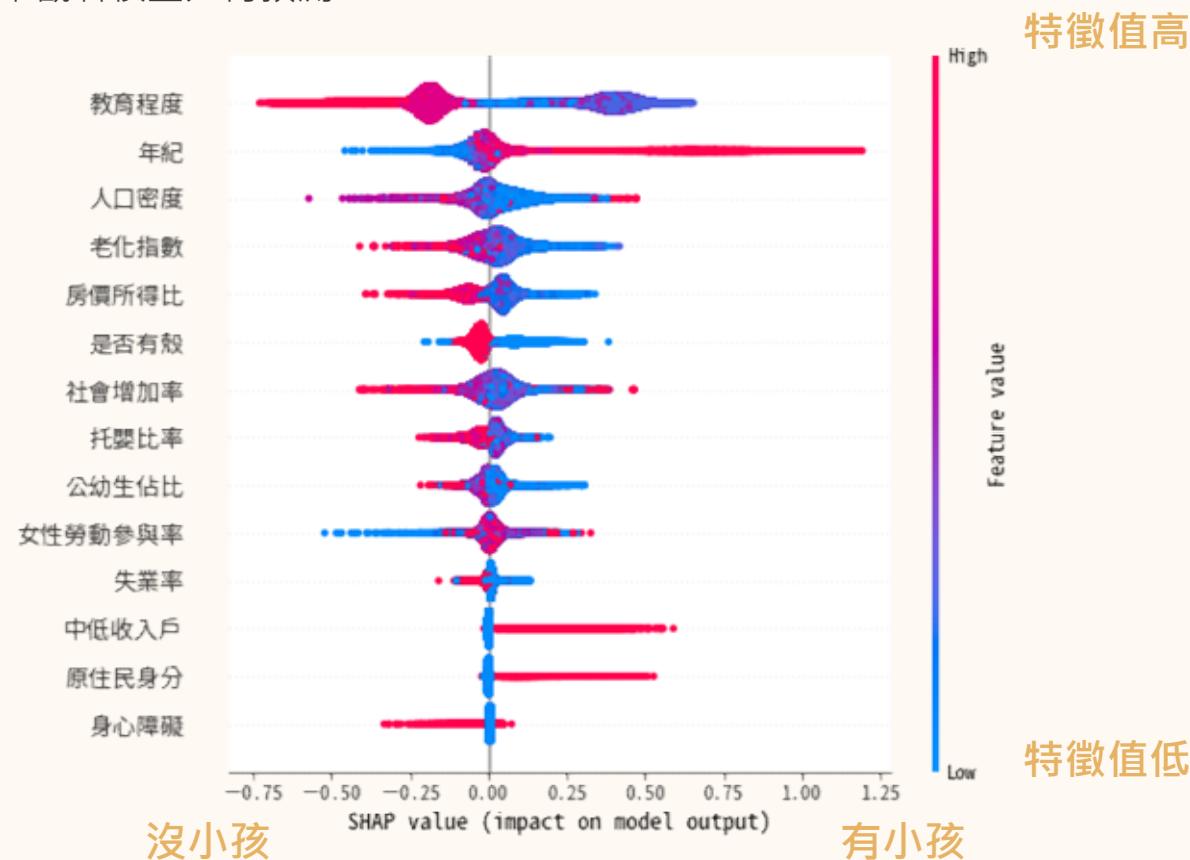
混淆矩陣(Confusion Matrix)



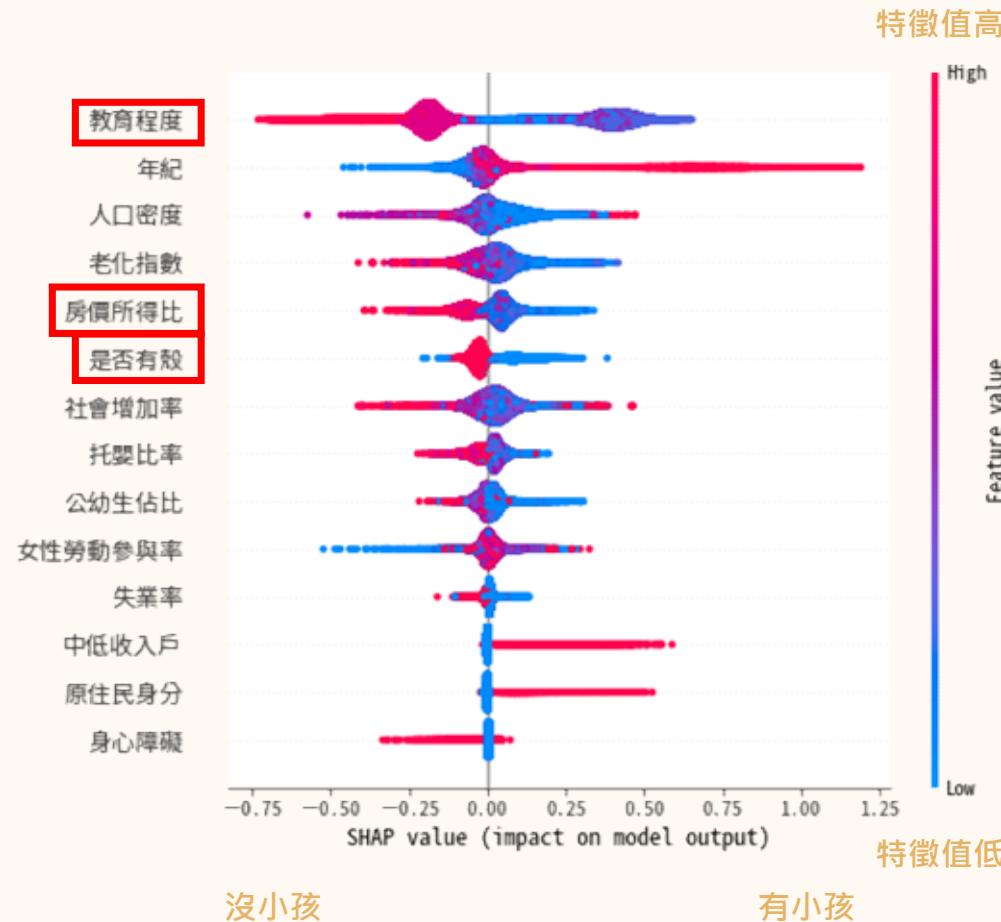
# 模型解釋分析

- 模型解釋分析採用SHAP (SHapley Additive exPlanations)方法
- 由訓練集所有樣本來觀看模型如何預測

特徵重要程度  
由高至低排序



# 模型解釋分析



由**SHAP**模型可以觀察到模型如何預測  
當已婚適育年齡婦女...

- 教育程度愈高，有小孩的機率愈低
- 所在縣市的房價所得比愈高，有小孩的機率愈低
- 有殼(房)，有小孩的機率愈低

# 模型解釋分析

- 我們也可以觀察特定樣本模型是如何預測

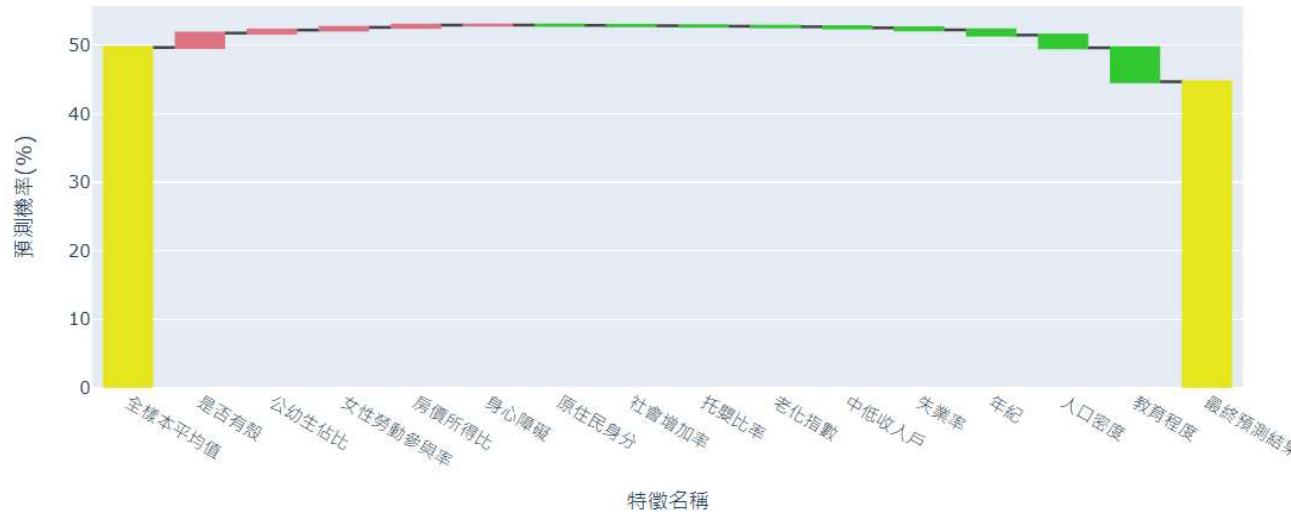
## 特徵貢獻圖

每個特徵如何影響預測結果?

### 調整特徵值觀察模型預測過程

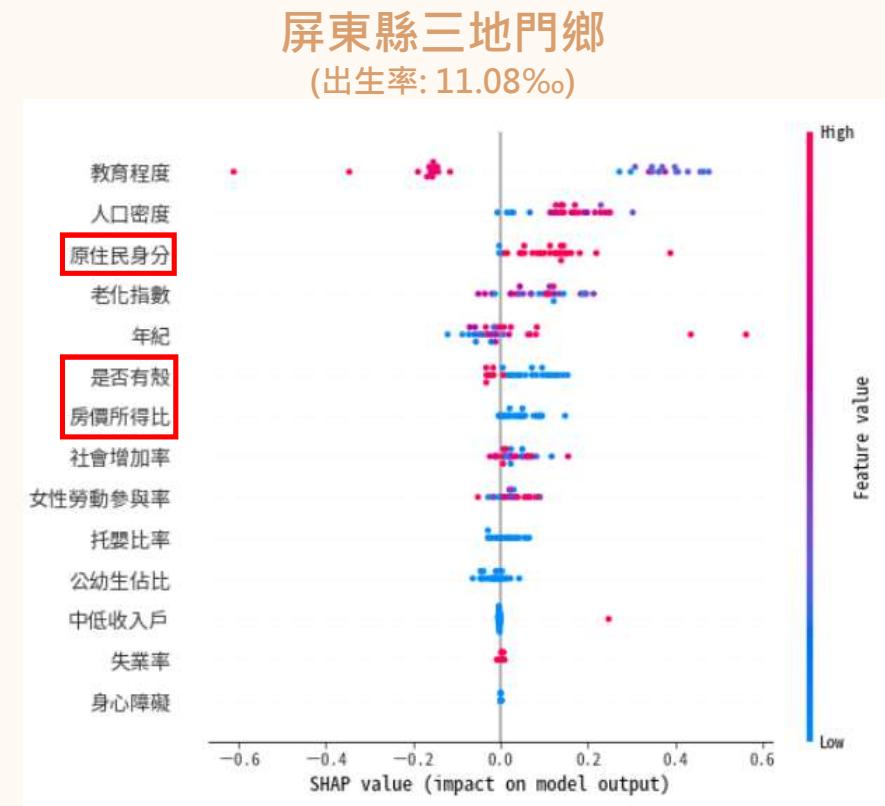
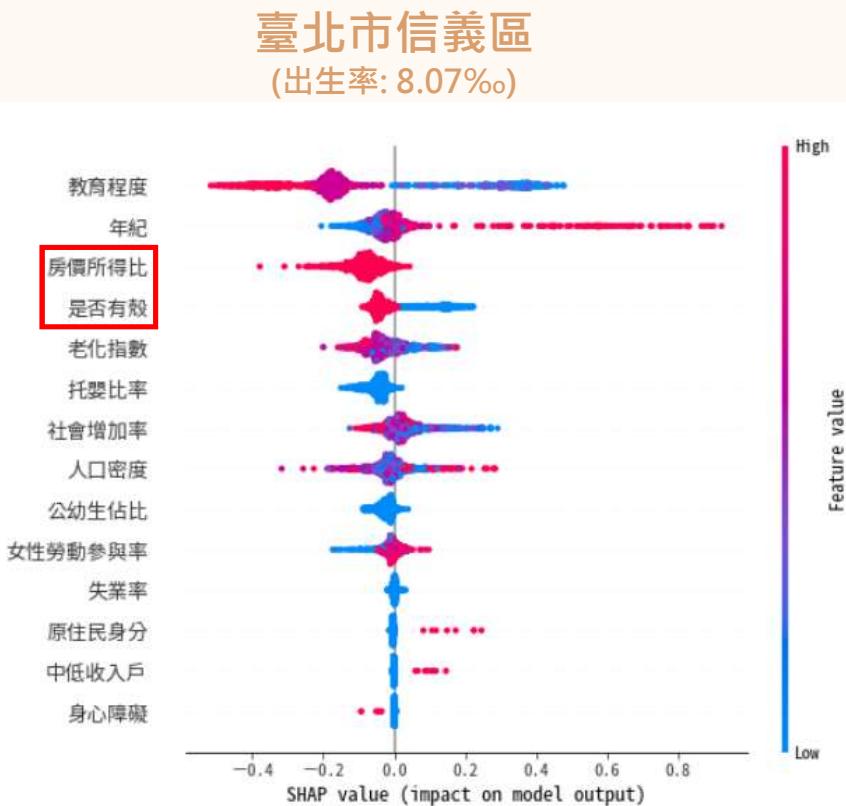
請選取縣市：	桃園市	請選取鄉鎮市區：	中壢區	請選取村里：	中央里
請輸入年紀：	25	請選擇教育程度：	大學	是否有殼：	<input checked="" type="radio"/> 有 <input type="radio"/> 否
是否為身心障礙者：	<input type="radio"/> 有 <input checked="" type="radio"/> 否	是否為中低收入戶：	<input type="radio"/> 有 <input checked="" type="radio"/> 否	是否有原住民身份：	<input type="radio"/> 有 <input checked="" type="radio"/> 否

特徵貢獻預測機率瀑布圖 預測有小孩機率 = 44.63%



# 模型解釋分析

- 針對特定的鄉鎮市區樣本群來進行分析，觀察哪些特徵影響各鄉鎮市區生小孩的意願





06

## 輿情系統

## 輿論系統功能

1. 議題聲量分析
2. 熱門討論文章
3. 文字雲
4. 關鍵字向量圖

少子女化

婚姻概況

議題觀察期間  
30, 90, 180天

公幼+公托

經濟概況

福利補助

# 議題聲量分析

## 輿情分析

最近大家都在討論什麼少子女化議題？

請選取議題：

少子女化

請選取觀察期間(天)：

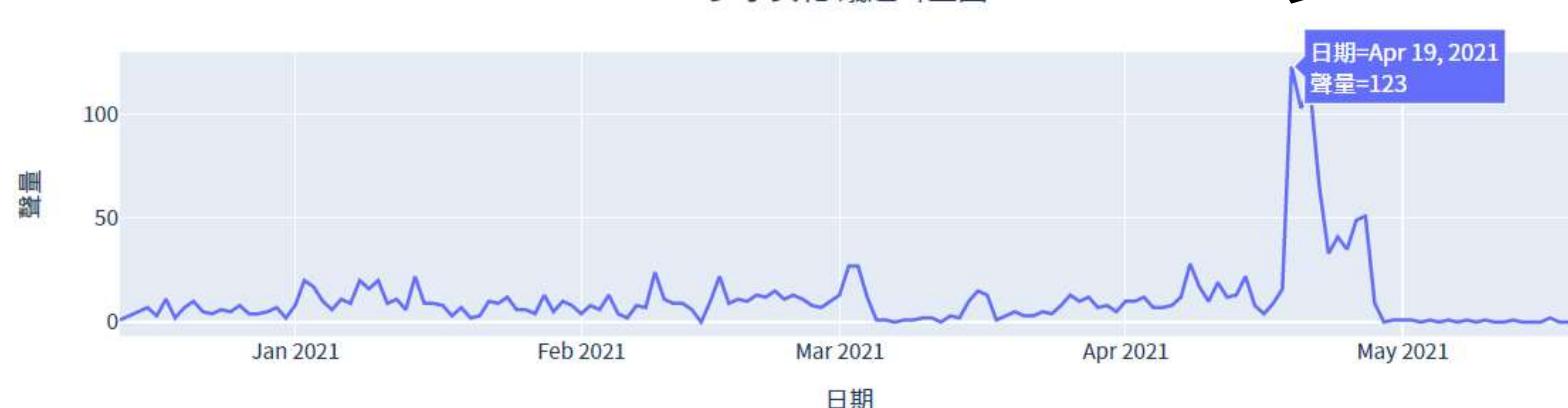
30

## 議題聲量分析

最近這個議題常被討論度嗎？

討論度最高  
CIA:全球生育率預測，  
台灣倒數第一

少子女化 議題聲量圖



## 議題聲量分析

少子女化議題討論度最高文章

CIA: 全球生育率預測，台灣倒數第一

2021-04-19

批踢踢實業坊 > 看板 Gossiping

作者 kko5566 (kko)  
標題 [新聞] CIA全球生育率預測 台灣倒數第一  
時間 Mon Apr 19 11:25:47 2021

1. 媒體來源:  
三立新聞網

2. 記者署名:  
生活中心／游雅嵐報導

3. 完整新聞標題:  
CIA全球生育率預測 台灣倒數第一

4. 完整新聞內文:  
美國中情局（CIA）公布2021全年全球人口生育率預測，在227個國家或地區中，墊底的分別是香港、澳門、新加坡、南韓，台灣則敬陪末座，在生育率預測中成為世界最低。  
台灣少子化現象嚴重。（示意圖／資料照）  
美國中情局預測全球227個國家或地區生育率，其中生育率最高的5個國家都在非洲，西非國家尼日（Niger）拿下第1名，第2名是中非國家為安哥拉（Angola），第3名為中非的剛果（Congo），第4名為西非的馬利（Mali），第5則為中非的查德（Chad）。  
生育率預測最低的倒數5名都在亞洲，依序為香港、澳門、新加坡、南韓、台灣。

## 熱門討論文章

### 輿情分析

最近大家都在討論什麼內容？

PTT-八卦板

PTT-親子板

Dcard-時事板

Dcard-親子板

BabyHome寶貝家庭親子網

日期 文章標題

2021-05-10 準公共化幼兒園收費資料看到的混亂情況

2021-04-15 從太魯閣出軌事件來看台灣交通 以及佳龍部長

2021-01-27 台南少子化問題的解方

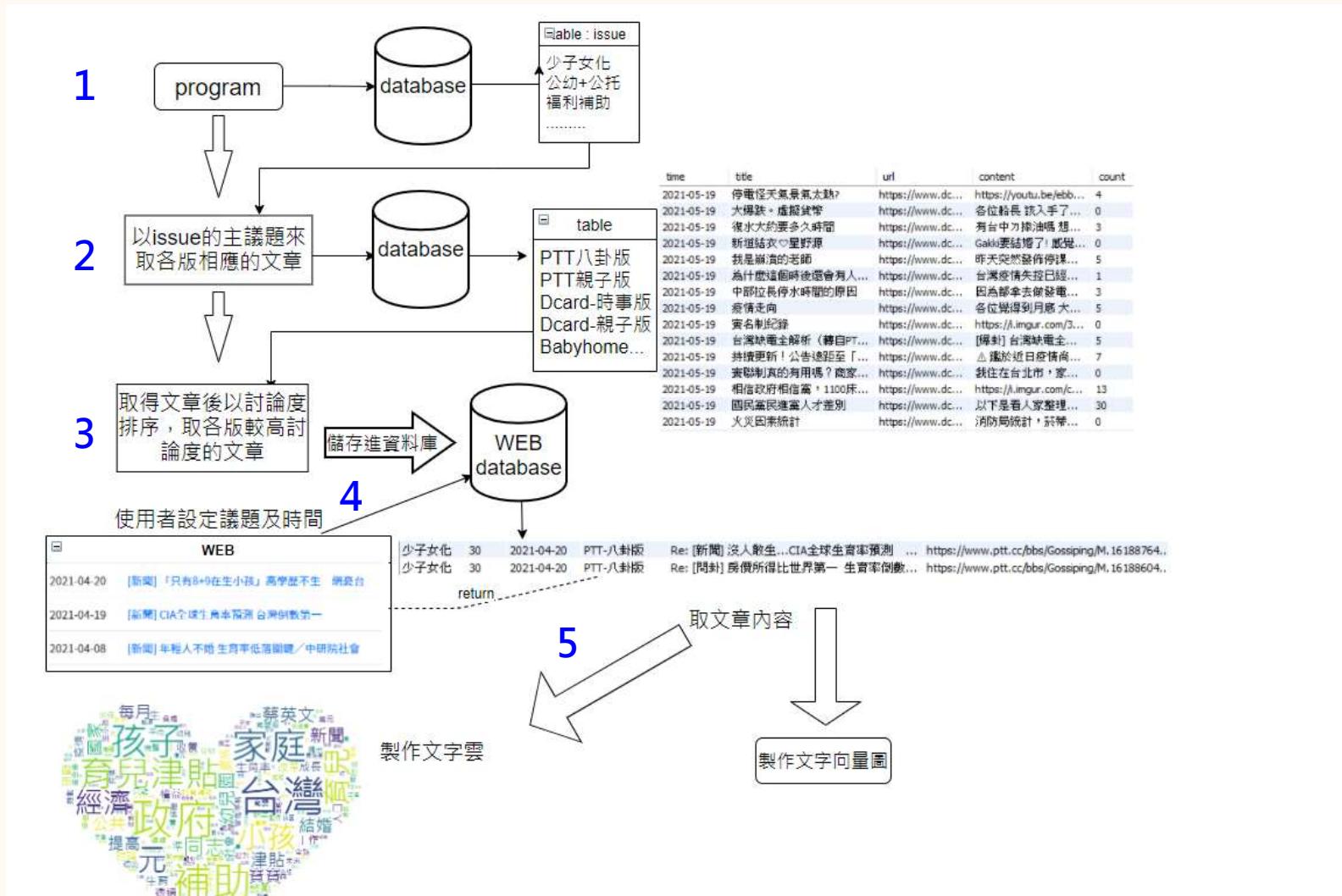
2021-01-08 人口負成長 林萬億示警：出生數階梯式下滑

2021-01-06 【情報】獨／5寶爸4萬養9口開銷大公開 祕訣是..網一看全跪了

# 熱門討論文章文字雲



# 熱門討論文章與文字雲程式架構



## 關鍵字向量圖

關鍵字向量圖

觀察關鍵字的相近程度(採用word2Vec及TSNE模型)

○2D圖

03D

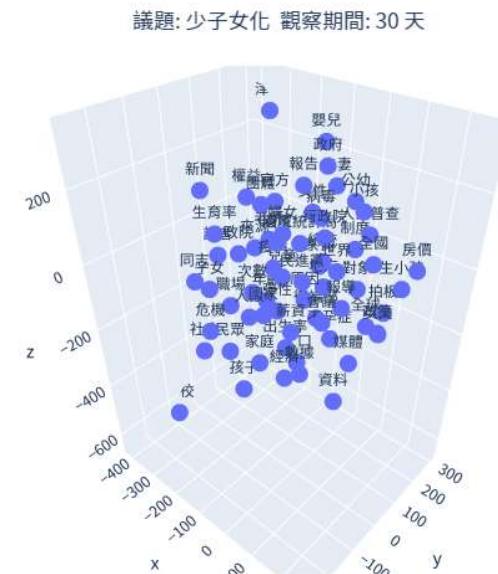


關鍵字向量圖

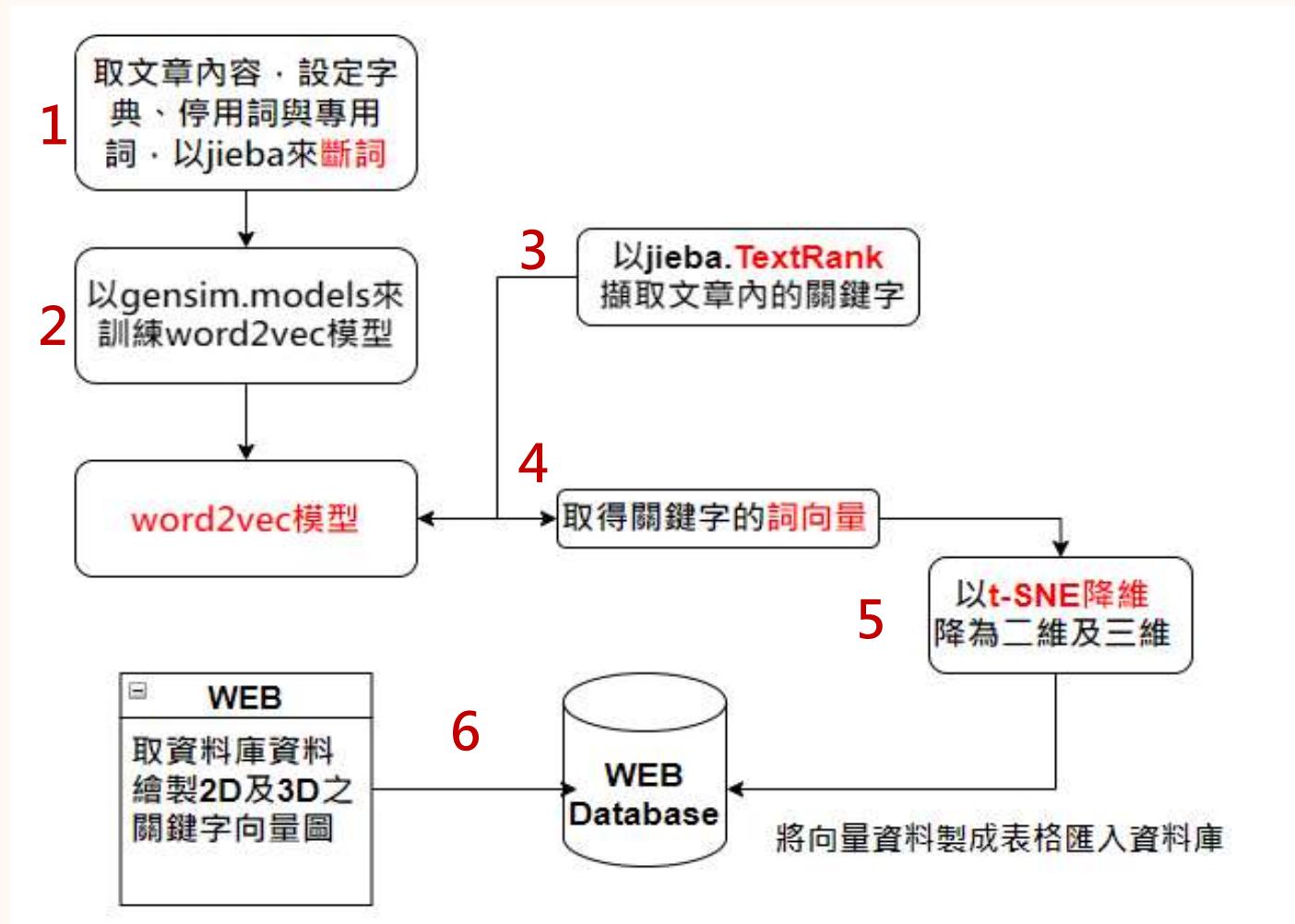
觀察關鍵字的相近程度(採用word2Vec及TSNE模型)

O2D

03D



## 關鍵字向量圖程式架構





07

未來展望

# 未來展望



我們以XGBoost模型加上SHAP方法，分析哪些因素會影響我國已婚適育年齡的婦女是否有孩子。我們也可將預測目標轉換為其他問題來進行分析，例如：

- 在有一個以上小孩的婦女樣本條件下，以模型預測是否有兩個以上的孩子，來分析影響婦女多生小孩的因素。
- 預測適婚年齡的女生是否有結婚，分析影響結婚的關鍵因素。



在輿情系統部分，斷字優化、增加議題與時間段，未來亦可做文章情緒分析以及優質文章的篩選，並可導入Hadoop及Spark做分散式計算，加速計算過程。



簡報結束

謝謝您的聆聽

The briefing is over, Thank you for listening .

