

# 拯救婚姻神機趨勢與情緒分析 - 洗碗機

指導老師：黃三益 教授

組別：第十六組

組員：

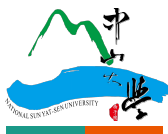
N074220025 陳昱廷

N104020009 蔡雨臻

N104020010 盧貴聰

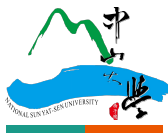
N104220017 胡冠甄





# 目錄

1. 動機和分析目的
2. 資料集描述
3. 資料的分析過程
4. 視覺化的分析結果與解釋
5. 結論



# 動機與目的

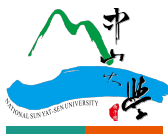
a. 主題: 拯救婚姻神機趨勢與情緒分析- 洗碗機

b. 議題發想動機:

在婚姻中，經常會出現因為家務瑣事的分配問題而產生的矛盾和摩擦，家庭中由誰來負責洗碗的分工議題越吵越兇，日積月累的小爭吵更很有可能會讓我們的婚姻走向結尾，近期討論度高的家電設備「洗碗機」不就正是對於此些問題的完美解答嗎？

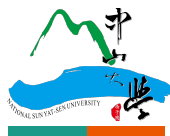
c. 分析目的:

為讓一段婚姻幸福美滿，啟動洗碗機，就可以不用親自洗碗，不僅能夠避免家庭成員之間因為洗碗而發生矛盾，希望能從網友的發言中得到關於洗碗機的評論與推薦品牌，若真如此強大，還真可考慮入坑一台，為我們的幸福婚姻再加一道保險。

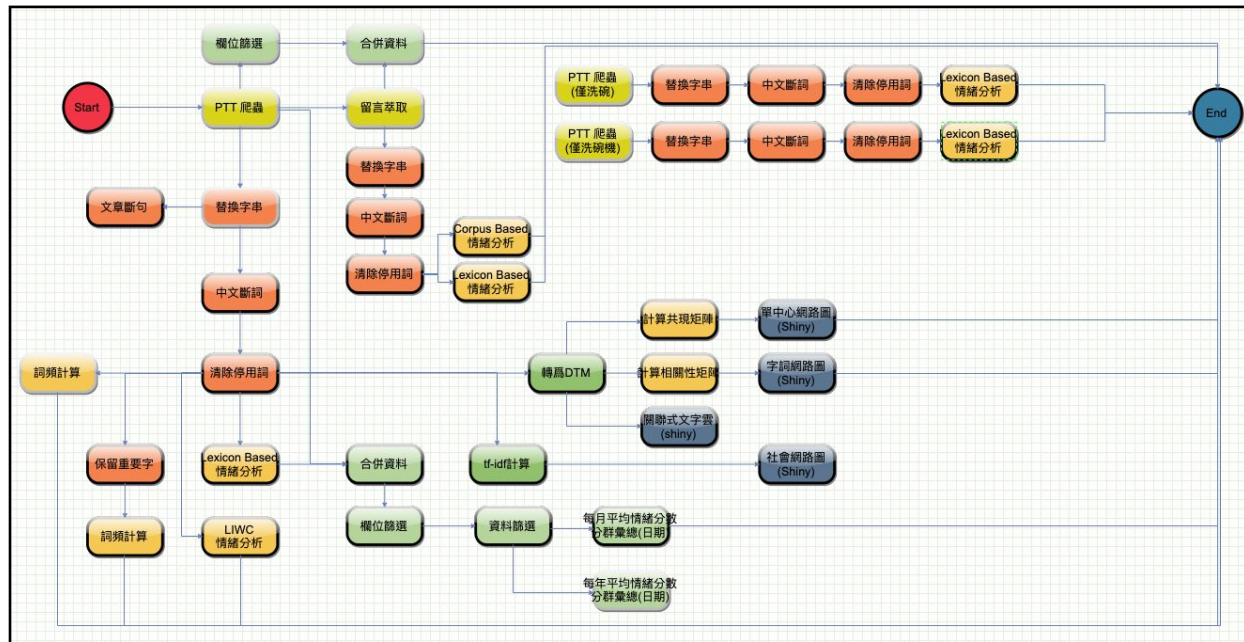


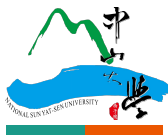
## 資料集描述

- 使用平台: 文字探勘工作流程設計平台
- 資料來源:
  - 主要本文來源: PTT-Marriage、PTT-Lifeismoney
  - 次要本文來源: PTT-Allergy、PTT-DailyArticle、PTT-GetMarry、PTT-HomeMaker
- 資料時間: 2017-01-01 - 2023-04-20
- 關鍵字: 洗碗機、洗碗
- 資料筆數: 1084筆



# 工作流程設計





## 資料前處理 - 替換字串

去除換行符號、網址、html標籤、底線與多餘的空格。

### 替換字串 (7)

參數有做更動，建議重新執行

參數設定

Input - 4

任務結果

選擇處理欄位 \*

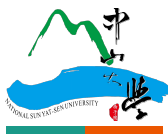
artContent

替換字串設定 ⓘ

```
\n\n>> °  
\n>> ,  
((http|ftp|https://)(([a-zA-Z0-9\._-]+\.[a-zA-Z]{2,6})|  
([0-9]{1,3}\.[0-9]{1,3}\.[0-9]{1,3}\.[0-9]{1,3}))(:[0-9]{1,4})?  
/[a-zA-Z0-9&%_\./~]*?)?>>
```

選擇替換規則檔案 ⓘ

-----請選擇-----



## 資料前處理 - 中文斷詞

使用平台預設字典進行中文斷詞，並將與分析主題洗碗機較為相關的字詞寫入自定義詞彙中，例如：富貴手、解放雙手、解救婚姻、省水等字眼，避免字詞被拆散。

中文斷詞 (9)

參數設定

Input - 7

任務結果

選擇處理欄位 \*

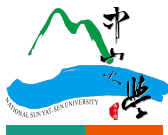
result

定義詞彙 ⓘ

拖曳機器入 1000  
烘衣機 500  
放進洗碗機 500  
富貴手 500  
非天然 500  
純天然 500

選取字典 ⓘ

-----請選擇-----



## 資料前處理 - 清除停用字

清除一些可能干擾分析結果的停用字(例如:本來、這種、超過、出生、發個、拿去、網站、發個、一年、ooo、東西、反正、這幾年、準備好、好像等等)。

### 清除停用詞 ( 11 )

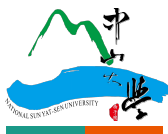
參數設定

Input - 9

任務結果

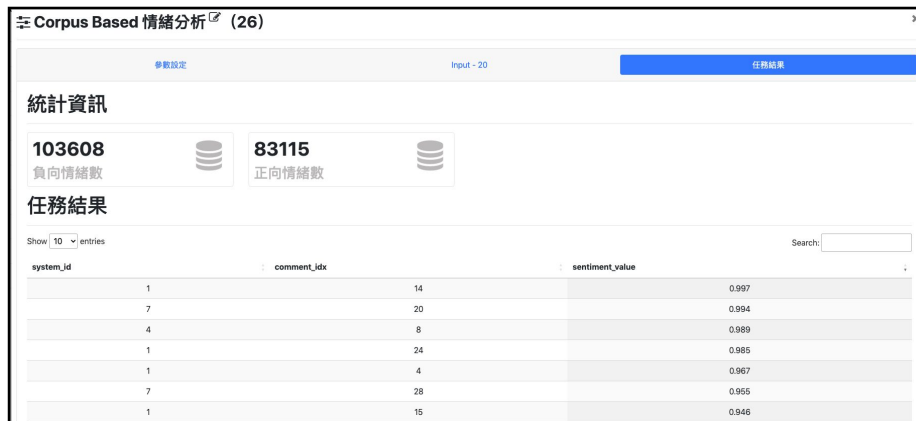
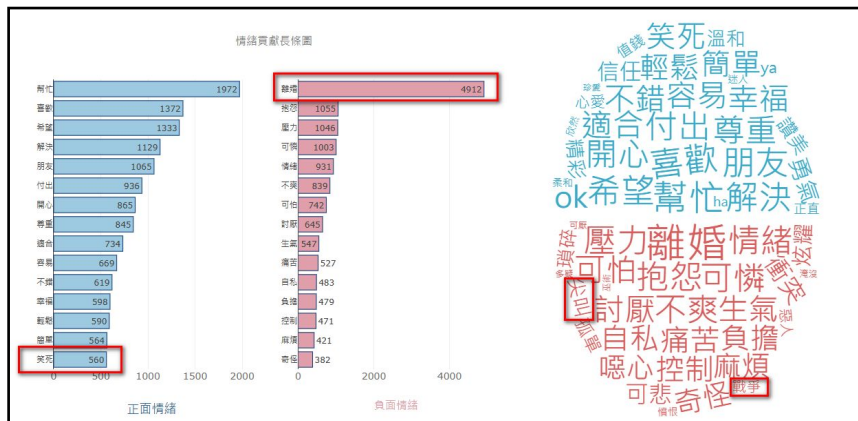
語言 *	使用預設停止詞
<div>Chinese</div>	<div>是</div>
是否清除單字元 ⓘ	是否轉為小寫英文
<div>是</div>	<div>是</div>
清除英文字母 *	清除數字 *
<div>否</div>	<div>是</div>
清除換行符號 *	清除特殊標點符號 *
<div>是</div>	<div>是</div>
清除html tag *	自定義停止詞
<div>是</div>	<div>東西 反正 這幾年 準備好 好像 ☐</div>





## 資料前處理 的發現

留言萃取:原本想納入網友的留言一同分析,但查看情緒詞彙時發現留言的部分有點過度偏激,例如:離婚、尖叫、戰爭等詞彙比例偏高,使用「Corpus Based」進行分析也發現負向言論超出了我們認為的可信度,因此本研究僅採用PTT上的文章來做分析。

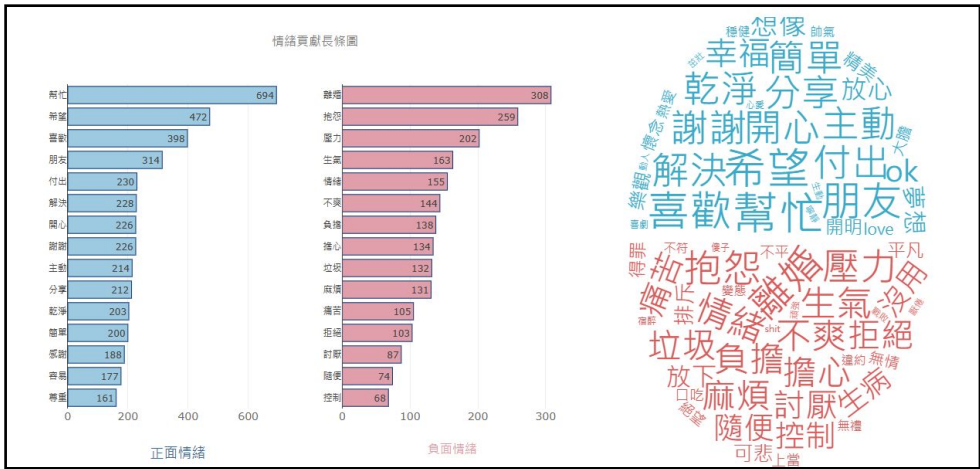


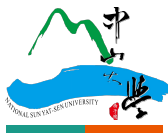


## 情緒分析 - Lexicon Based

情緒貢獻長條圖

正面情緒	負面情緒
幫忙	難堪
祈望	抱怨
喜歡	壓力
朋友	生氣
付出	情緒
解決	不爽
開心	負擔
謝謝	擔心
期望	垃圾
主動	煩惱
ok	拒絕
love	擔心
	隨便
	控制





## 搜尋關鍵字 - 洗碗

網友們於PPT討論版分別對於「洗碗」與「洗碗機」這兩個關鍵字並套用Lexicon Based情緒反應分析, 剖析網友們是洗碗造成婚姻危機呢? 抑或是, 家裏沒有安裝洗碗機造成危機呢?

PTT 爬蟲 (僅洗碗) (47)

參數有做更動，建議重新執行

參數設定

任務結果

選擇看板

KoreaDrama(韓劇)

learnyt(自學)

**Lifeismoney(省錢)**

Loan(貸款)

MacShop(蘋果買賣)

MakeUp(美妝)

marriage(婚姻)

搜尋起始日期

2017/01/01

搜尋關鍵字

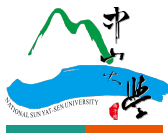
洗碗

排除關鍵字

洗碗機  
餐廳  
交易  
業配  
政治

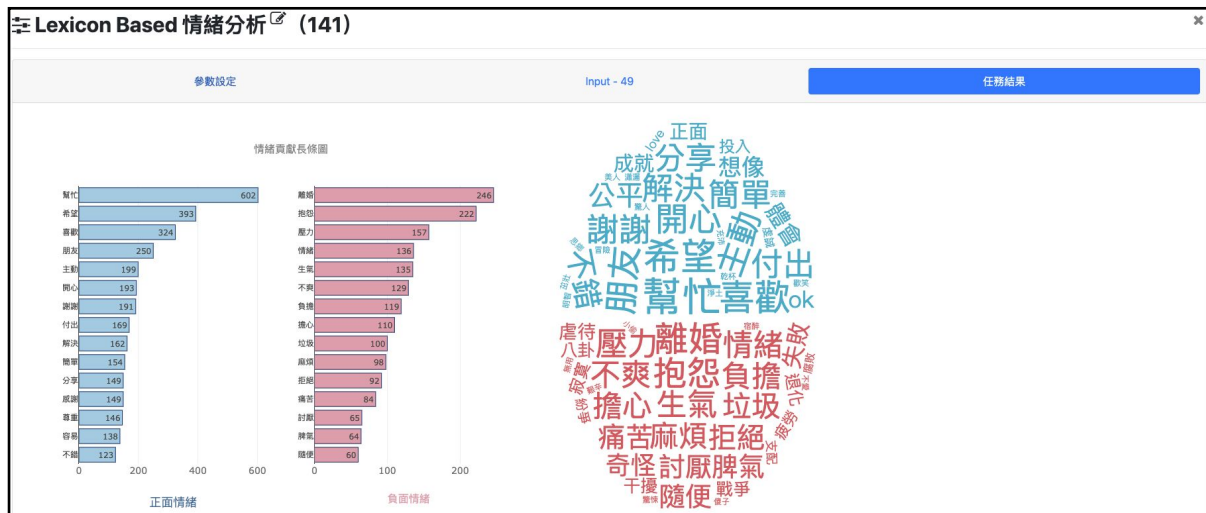
搜尋結束日期

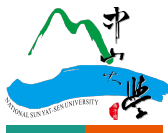
2023/04/20



## 搜尋關鍵字 - 洗碗

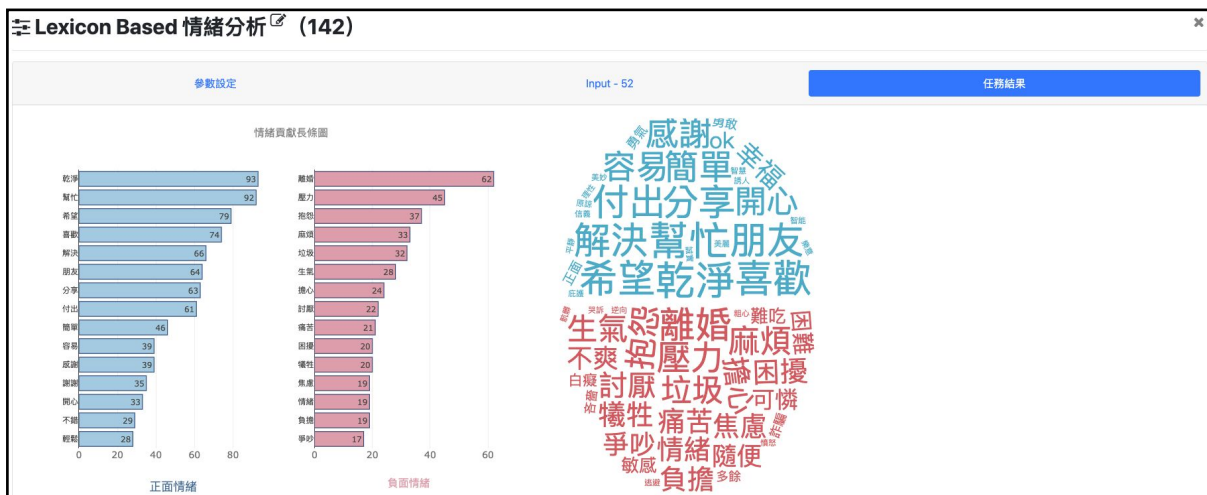
如圖所示結果，洗碗的正向字詞前三名為「幫忙、希望、喜歡」；負向字詞前三名為「離婚、抱怨、壓力」，依結果推斷洗碗這件事在婚姻分工中極具壓力欲請求幫忙。

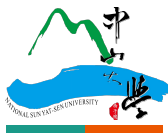




## 搜尋關鍵字 - 洗碗機

如圖所示結果，洗碗機的正向字詞前三名為「乾淨、幫忙、希望」；負向字詞前三名為「離婚、壓力、抱怨」。

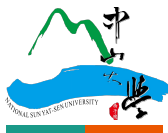




## 搜尋關鍵字 - 洗碗 V.S. 洗碗機

綜合關鍵字洗碗與洗碗機的比對結果，可有效佐證洗碗機於主題探討的連結度，確實與「婚姻」有相當大關聯，有洗碗機的幫忙就能減少洗碗壓力，共同字詞如下：

- ★ 前三名正向詞：幫忙、希望
- ★ 前三名負向詞：離婚、壓力、抱怨



## 情緒分析結果與原文合併 - 合併資料

取得情緒分析結果後，首先進行資料合併，設定「System\_id」做為JOIN欄位。

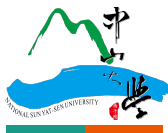
### 合併資料 (77)

參數設定Input - 4Input - 84任務結果

JOIN規則

新增規則刪除規則

left_key	right_key
system_id	system_id
-----請選擇-----	-----請選擇-----



## 情緒分析結果與原文合併 - 欄位篩選

接著進行欄位篩選，設定欲分析欄位有「positive\_count、positive\_words、negative\_count、negative\_words、sentiment\_value」，以便了解正負向字詞為何與資料筆數以及情緒分數。

### 欄位篩選 (80)

參數設定

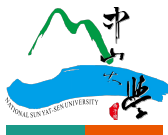
Input - 77

任務結果

選擇要保留的欄位(按住ctrl(Windows)或command(MAC)可以複選) \*

e.jp  
insertedDate  
dataSource  
positive\_count  
positive\_words  
negative\_count  
negative\_words  
sentiment\_value





## 情緒分析結果與原文合併 - 資料篩選

最後，將資料篩選條件設定為「sentiment\_value > 0」與「sentiment\_value < 0」，可有效找出正/負向情緒分數的有953筆資料。

### 資料篩選 (82)

參數設定Input - 80任務結果

#### 統計資訊

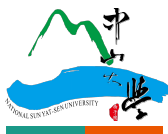
131  
移除數量

953  
保留數量

#### 任務結果

Show 10 entriesSearch:

system_id	artTitle	artDate	artPoster	artCategory	artContent	artComment	positive_count	positive_words	negative_count	negative_words	se
2	[新聞] 汗疱疹又冒水泡！醫點名「禁忌食物」	2022-05-19 23:31:55	ognillin	allergy	新聞來源：\nhttps://reurl.cc/moray1\n汗疱疹手部瘙癢不具傳染性\n\n汗疱疹是春季轉夏季時容易發生的疾病，屬於「急性濕疹」的一種，症狀以手掌劇烈瘙癢為主，不具有傳染性，其成因不明，可能與本身有異位性皮膚炎體質、接觸過敏原、壓力、多汗症、遺傳有關。 \n\n調整生活作息 避免症狀嚴重\n\n汗疱疹患者在生活上，應保持「睡眠」、「壓力」、「飲食」三者的平衡，汗疱疹因壓力大而變嚴重，所以此時睡眠和飲食很重要，無法減少壓力，至少顧好睡眠和飲食以免汗疱疹瘙癢加重。 \n\n保持手腳乾爽 鹽、黏及甜食少碰\n\n而手腳避免接觸肥皂、洗潔精、刺激性的化學物品，保持手腳的乾爽；在飲	[]	1	[容易]	4	[壓力, 壓力, 壓力]	



## 情緒分析結果與原文合併 - 分群彙總 日期 (平均情緒分數)

依照上圖所保留的953筆資料, 其時間維度設定為「每月」、「每年」進行分群彙總, 計算正向數、負向數、情緒值, 保留日期欄位, 以彙總平均情緒分數。

### 每月平均情緒分數 分群彙總(日期) (88)

參數設定

Input - 82

任務結果

選擇日期欄位 \*

artDate

選擇日期類型 \*

月

計算欄位(按住ctrl(Windows)或command(MAC)可以複選) \*

artCategory

artContent

artComment

positive\_count

positive\_words

negative\_count

negative\_words

sentiment\_value

日期格式 \*

%Y-%m-%d %H:%M:%S%z

匯總函數 \*

mean

保留欄位(按住ctrl(Windows)或command(MAC)可以複選) \*

system\_id

artTitle

artDate

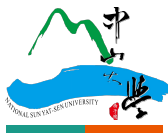
artPoster

artCategory

artContent

artComment

positive\_count



## 情緒分析結果與原文合併 - 分群彙總 日期 (平均情緒分數)

★ 每月

以月份拆分資料，結果如下有76個月的資料，可查看每月平均情緒分數。

參數設定


Input - 82

任務結果

統計資訊

76

群組數量

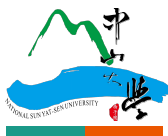


任務結果

Show 10 entries

Search:

artDate	positive_count	negative_count	sentiment_value
2019-02	16.571429	8.500000	8.071429
2017-12	10.636364	4.000000	6.636364
2019-09	12.000000	5.600000	6.400000
2018-07	14.071429	7.928571	6.142857
2017-11	12.400000	6.400000	6.000000
2021-04	8.555556	2.777778	5.777778
2022-04	11.235294	5.470588	5.764706



## 情緒分析結果與原文合併 - 分群彙總 日期 (平均情緒分數)

### ★ 每年

以年份拆分資料，結果如下有7年的資料，可查看每年平均情緒分數。尤其以年度結果，雖維持在正向分數，但可看出網友對於洗碗機的討論熱度其正向情緒逐年遞減。

每年平均情緒分數 分群彙總(日期) (129) [參數有做更動，建議重新執行](#)

參數設定 input - 82 任務結果

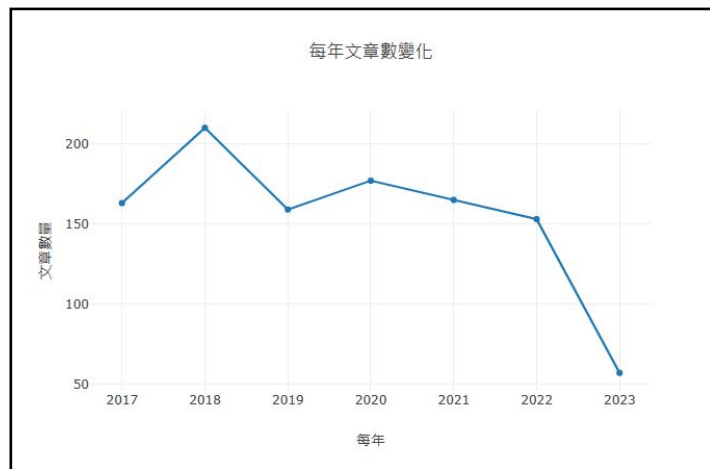
### 統計資訊

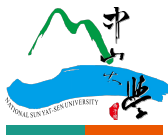
7 群組數量

### 任務結果

Show 10 entries Search:

artDate	positive_count	negative_count	sentiment_value
2017	9.482517	5.972028	3.510490
2018	10.508108	7.508108	3.000000
2019	11.317241	7.627586	3.689655
2020	7.320000	5.320000	2.000000
2021	8.932432	5.648649	3.283784
2022	9.351145	6.312977	3.038168
2023	8.509804	6.705882	1.803922



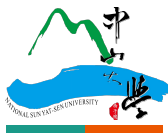


## 詞頻與關聯性 - 詞頻計算

為了解PTT討論版中的最常使用的字詞數量，故計算詞頻以文字雲呈現前200名的詞彙權重。詞頻前三名為「小孩、老公、老婆」都是角色，然第四名才是主角「洗碗」。由此研判引發洗碗機討論度的原由來自於夫妻及小孩，也因此衍生洗碗機的需求。



詞頻計算 (131) 參數有效更動，建議重新執行		
Input - 11		任務結果
任務結果		
Show 10 entries Search:		
n	Term	
3238	小孩	小孩
2479	老公	老公
1661	老婆	老婆
1198	洗碗	洗碗
1186	婆婆	婆婆
1120	公公	公公
1083	家事	家事
999	工作	工作
827	媽媽	媽媽
722	孩子	孩子



## 保留重要字 - 洗碗機品牌

特別好奇在洗碗機的討論中，哪個洗碗機品牌討論頻率最高，故設定保留字以歐、日、台系品牌為主。

保留重要字 (133)

參數設定

Input - 11

任務結果

設定保留詞彙 ⓘ

Bosch  
ASKO  
Frigidaire  
HERAN  
BEST

選取字典

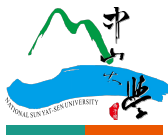
-----請選擇-----

字典欄位

name

設定保留洗碗機品牌字詞後，接續計算詞頻以文字雲顯示前 200名詞彙，目的為找出最高討論的品牌為何。排序由高至低為：伊萊克斯（瑞典）> 博世（德國）> 林內（台灣）= 禾聯（台灣）= 富及第（美國），由上述結果顯示，可解讀 網友最愛洗碗機品牌為歐系為大宗，後三者則權重一致無法分辨網友的偏好程度。





## 分析結果解釋 - 轉為DTM

為提取資料中的特徵故轉為DTM, 計算共現矩陣與相關性矩陣, 並以Shiny網路圖呈現結果。  
取1084篇文章, 篩選最常出現的前200個字詞在各篇文章中出現的次數, 拆解結果如下所示。

三轉為DTM (92)

參數設定

Input - 11

任務結果

統計資訊

201  
字數

1084  
文章數

任務結果

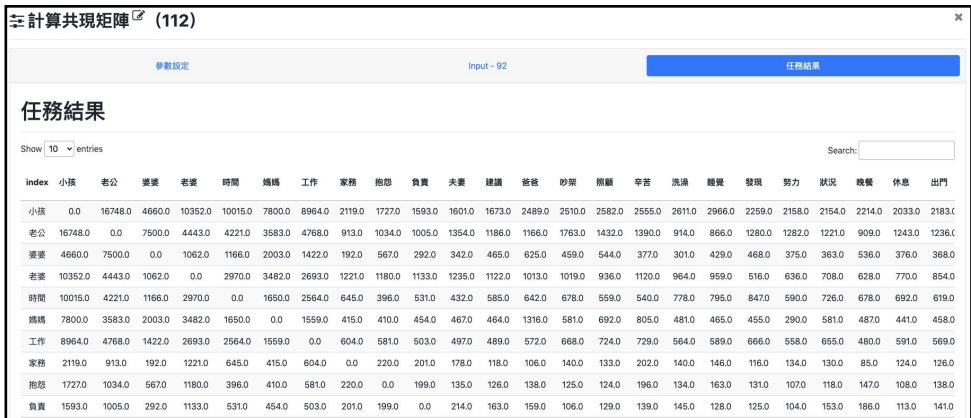
Show 10 entries

Search:

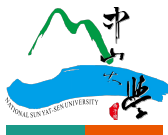
system_id	no	上班	下班	不想	不爽	不用	不錯	主動	主婦	乾淨	交往	人生	付出	休息	個性	假日	傳統	兒子	兩人	公司	公司	公婆	出門	分享	分配	分鐘	努力	原因	台灣	吃完	吃飯	吵架
1		0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
2		0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
3		0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
4		0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
5		0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
6		0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	5.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0



熱圖



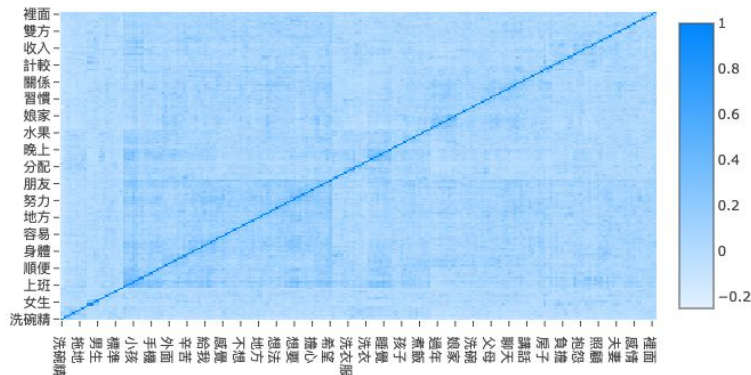




## 分析結果解釋 - 計算相關性矩陣

針對前200名詞彙結果，以熱圖查看兩種維度之間的關聯性，其結果如下。

熱圖



計算相關性矩陣 (113)

參數設定

Input - 92

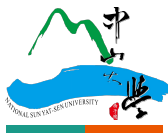
任務結果

任務結果

Show 10 entries

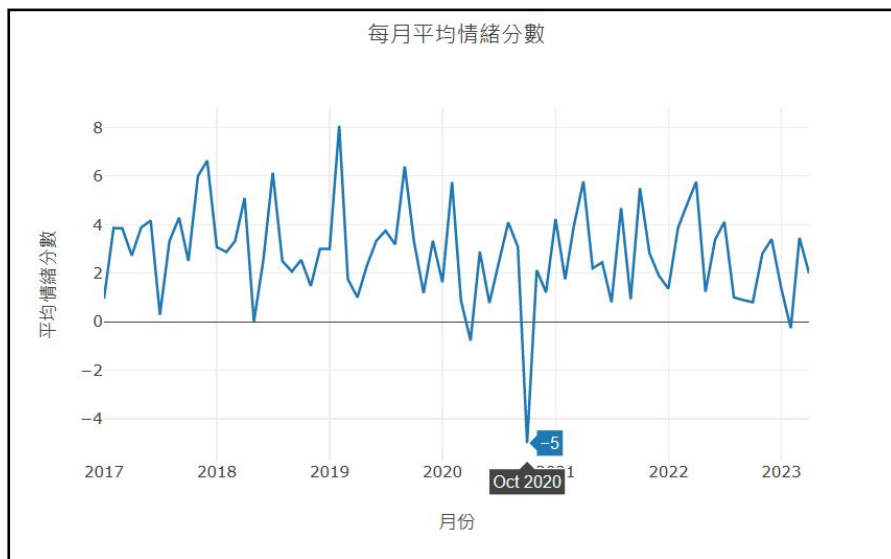
Search:

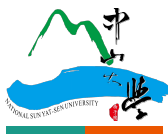
index	洗碗精	洗碗機	碗盤	乾淨	清潔	地板	拖地	分鐘	洗完	女方	男方	女生	男生	外食	小朋友	不錯	空間	方法
洗碗精	1.000000	-0.088917	-0.019984	0.092579	0.063860	0.046364	-0.052779	-0.045134	0.023884	-0.046077	-0.044790	-0.065738	-0.052428	-0.076650	-0.036372	-0.018881	-0.030592	0.0
洗碗機	-0.088917	1.000000	0.361503	0.157759	0.024845	0.000806	0.069075	0.072322	0.085543	-0.031327	-0.032015	-0.051401	-0.030950	-0.008918	-0.033268	-0.047633	0.099822	-0.0
碗盤	-0.019984	0.361503	1.000000	0.246201	0.075415	0.006044	0.038336	0.119895	0.203024	-0.040650	-0.039164	-0.049771	-0.039189	0.012768	-0.030496	-0.015228	0.057083	0.0
乾淨	0.092579	0.157759	0.246201	1.000000	0.260033	0.196140	0.213871	0.139231	0.167041	-0.020983	-0.014806	0.000196	0.011003	0.008512	-0.026526	0.022305	0.082575	0.0
清潔	0.063860	0.024845	0.075415	0.260033	1.000000	0.219636	0.190627	0.074232	0.093086	-0.028835	-0.034385	-0.017627	-0.009748	-0.007905	0.057515	0.011034	0.052299	0.1
地板	0.046364	0.000806	0.006044	0.196140	0.219636	1.000000	0.182057	0.041432	0.044855	-0.019696	-0.029628	-0.008132	0.001963	0.012991	0.055531	0.063348	-0.015212	0.0
拖地	-0.052779	0.069075	0.038336	0.213871	0.190627	0.182057	1.000000	0.013128	0.054636	0.022895	0.001827	-0.008395	0.016997	0.026752	0.033198	0.033799	0.023594	0.0
分鐘	-0.045134	0.072322	0.119895	0.139231	0.074232	0.041432	0.013128	1.000000	0.090514	0.006870	-0.000996	0.002831	0.032290	-0.045072	0.056080	-0.036914	0.002522	-0.0
洗完	0.023884	0.085543	0.203024	0.167041	0.093086	0.044855	0.054636	0.090514	1.000000	-0.035079	-0.034049	-0.034942	-0.037068	-0.036521	0.011727	0.034077	0.019413	0.1
女方	-0.046077	-0.031327	-0.040650	-0.020983	-0.028835	-0.019696	0.022895	0.006870	-0.035079	1.000000	0.795740	0.139385	0.155299	-0.027763	-0.023840	0.002707	-0.014351	-0.0



## 視覺化分析結果 - 每月平均情緒分數

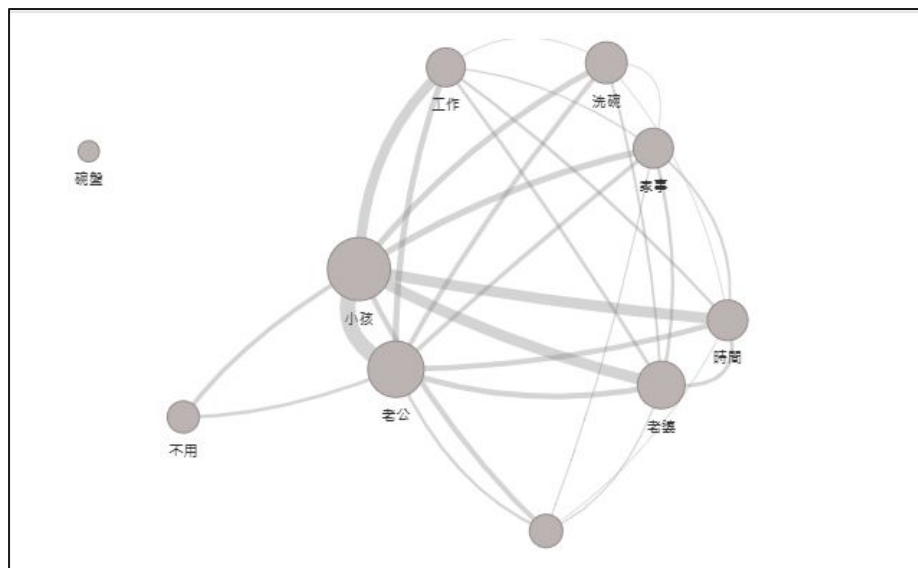
可以看到僅有2020/5月份負向情緒非常飽滿，查詢文章後發現當時有篇文章正討論著疫情爆發後居家帶小孩帶來的家事負擔，此篇文章也被不同的網友們轉發並討論，由此我們能推論，因為洗碗這件事很容易帶給家庭爭吵與負面情緒，洗碗機或許真是解救婚姻神器。

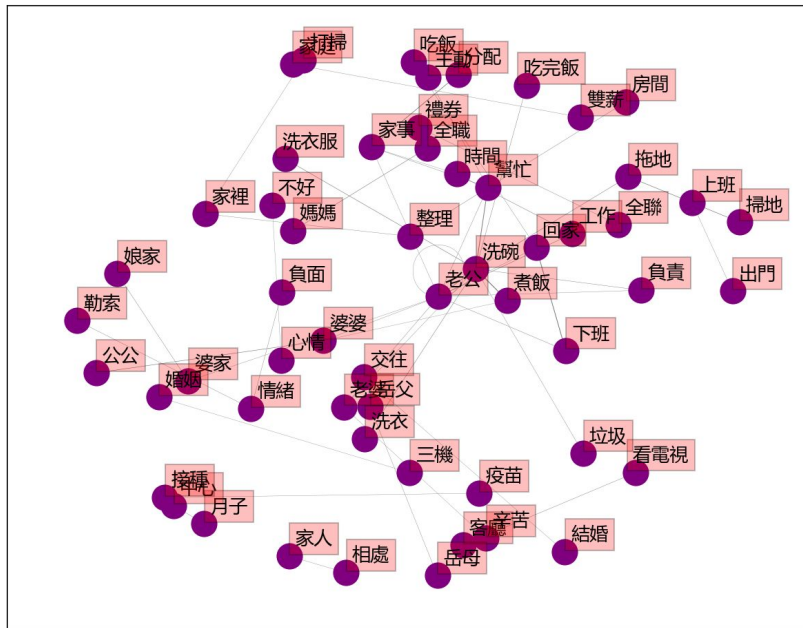


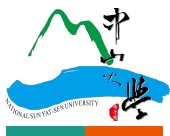


## 視覺化分析結果 - 相關性單中心網路圖

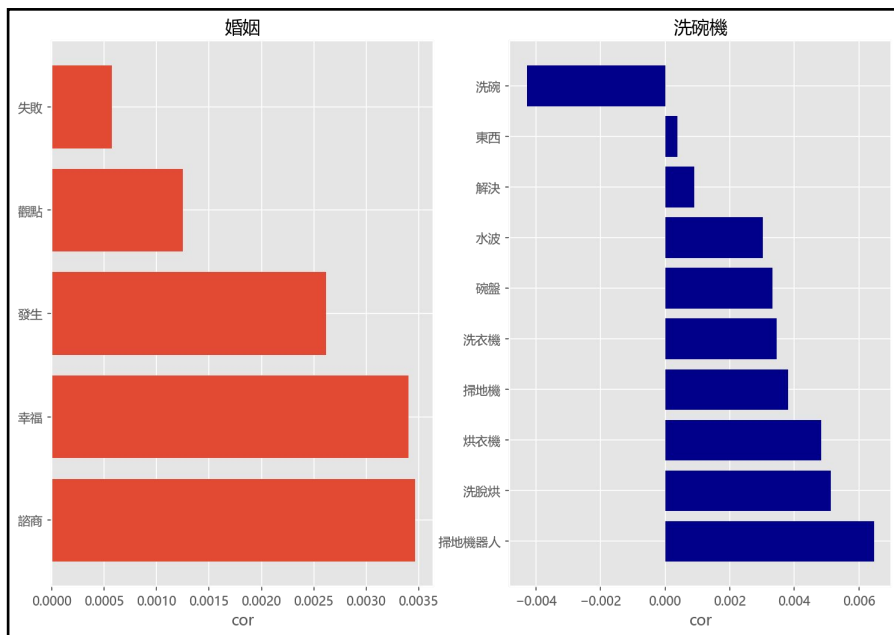
將字詞進行相關性分析，以洗碗與洗碗機分別找出最相關的字詞，可以發現議題總是圍繞在家庭問題上，包含小孩或時間等等，可見此項家事所帶來的困擾影響的不僅是個人。

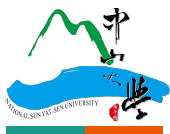




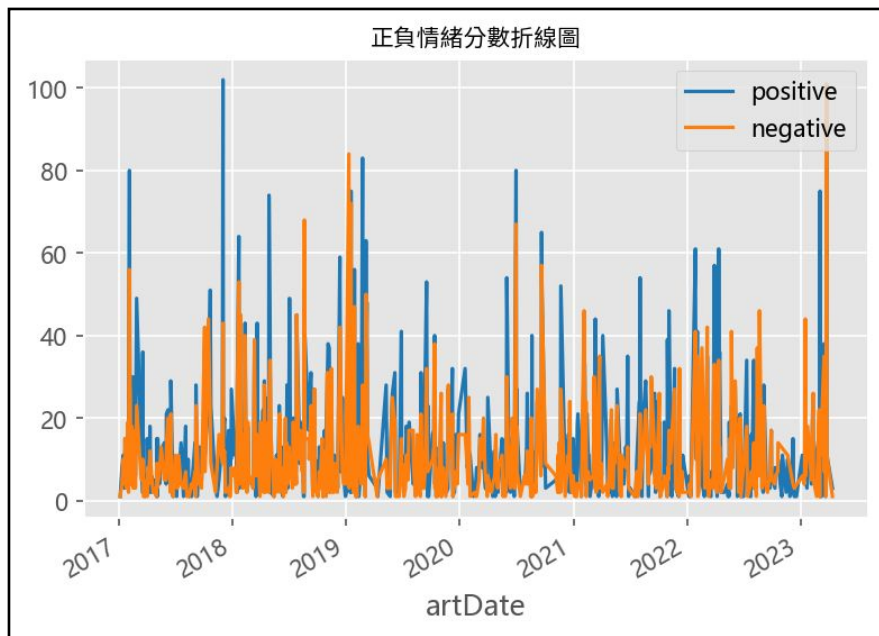


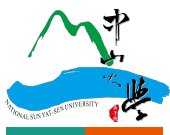
## Python視覺化 - 與洗碗機&婚姻相關性最高的10個詞彙



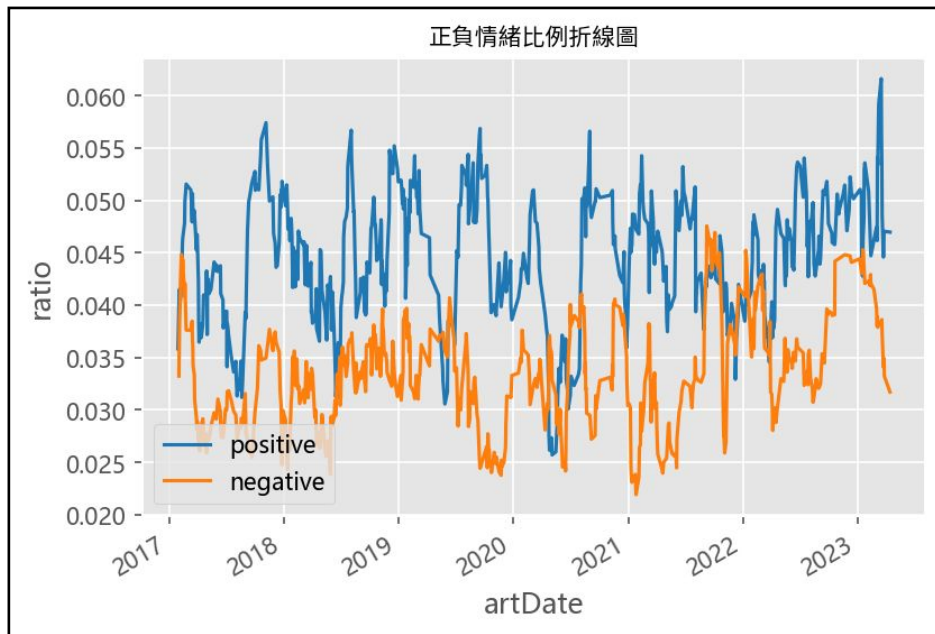


## Python視覺化 - 正負情緒分數折線圖





## Python視覺化 - 正負情緒比例折線圖

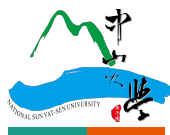




# Python視覚化 - 2017-11-30文字雲



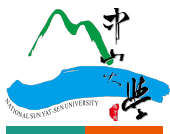




# Python視覺化 - 2017-11-30文字雲

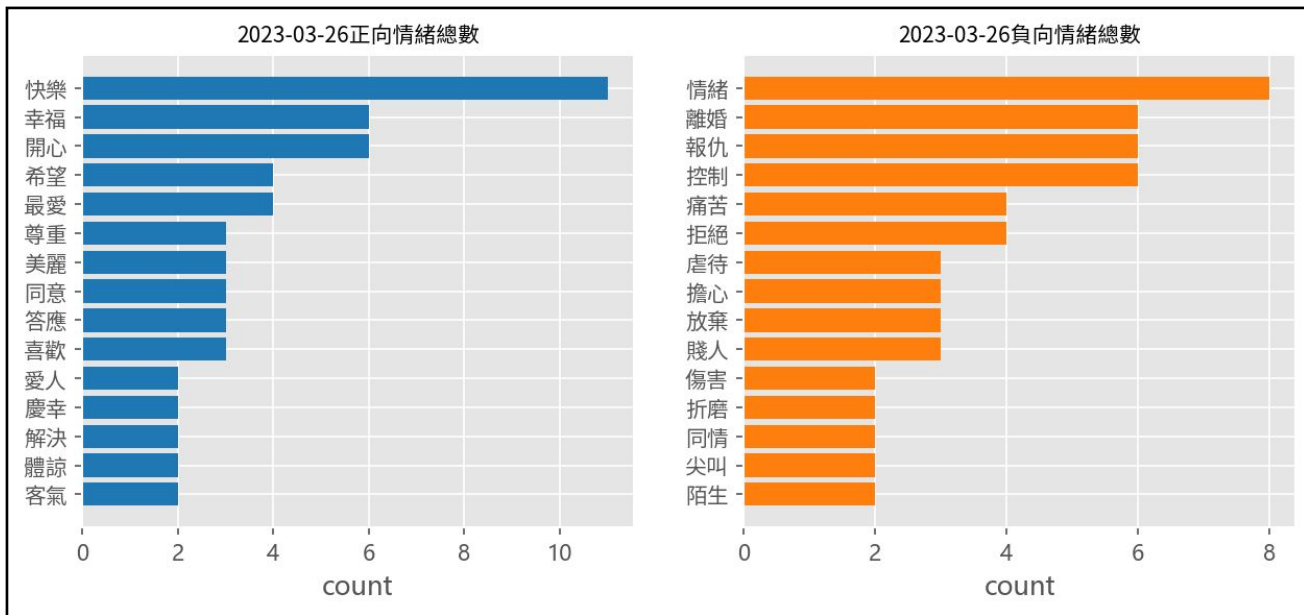
	artDate	sentiments	size	ratio
561	2017-11-30	positive	102	0.077273
3845	2023-03-26	positive	92	0.058008
1544	2019-02-22	positive	83	0.073909
2286	2020-06-30	positive	80	0.047676
85	2017-02-02	positive	80	0.065898

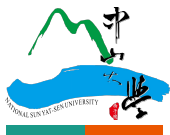




	artDate	sentiments	size	ratio
3844	2023-03-26	negative	101	0.063682
1417	2019-01-09	negative	84	0.047700
1432	2019-01-16	negative	72	0.052440
1151	2018-08-19	negative	68	0.058570
2285	2020-06-30	negative	67	0.039928

## Python視覺化 - 2023-03-26 正/負向情緒總數

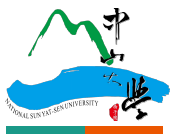




## 結論

在分析當中，常看見時間、麻煩或壓力等負向情緒字詞，尤其現已雙薪家庭為主流，現代人忙碌的工作回到家吃飽飯後，面對洗碗這項家事所帶來的厭惡感真讓人力不從心，但對於充滿油污的碗盤又不能不處理，做家事所日積月累的抱怨情緒讓[離婚]這一詞彙成為了我們分析文章中最頻繁出現的字詞，又以疫情後須長時間待在家更為明顯，不管事多深厚的感情基礎都可以被此消磨殆盡。

分析可看出洗碗機的出現確實解救了一些家庭的危機，但對於洗碗機是否能夠真正地做到解放雙手或許還仍待更多的技術文章來佐證，但起碼洗碗機的出現不僅能夠幫助節省時間和精力，還能夠提高洗碗的效率和質量，尤其對於那些沒有太多時間或者經常需要處理大量碗盤的家庭來說更加重要。



報告結束，謝謝閱讀



*Happy wash machine Happy Marriage.*

