


# 天貓商品評論爬蟲

—以ColourPop唇膏的商品頁面為例



報告者：黃意程

2020/03/25

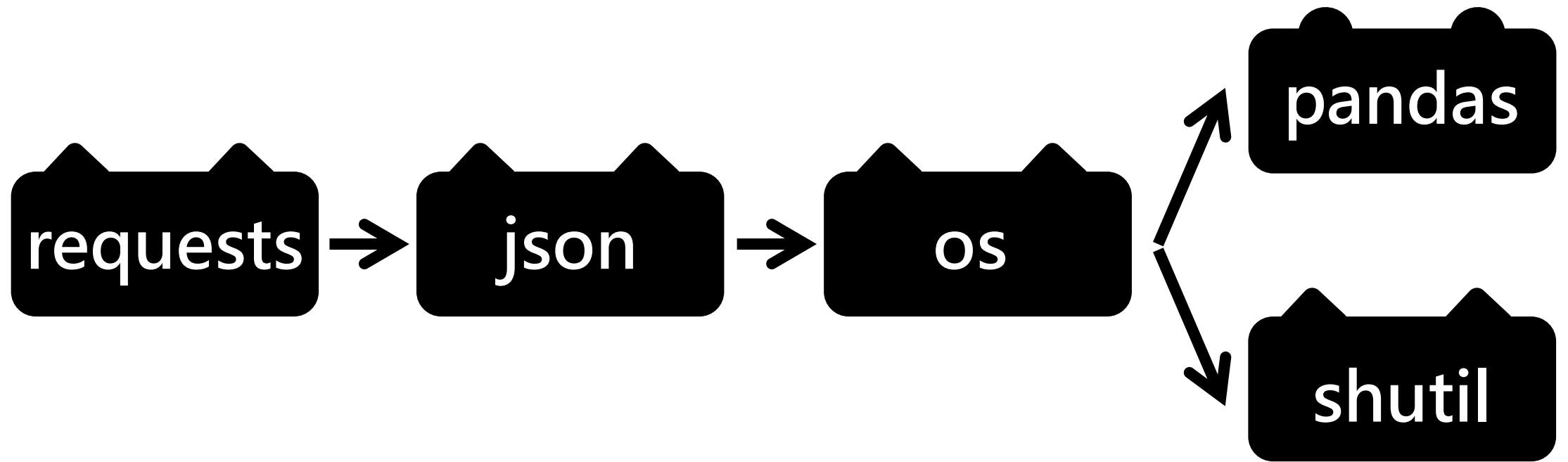


# WHY?

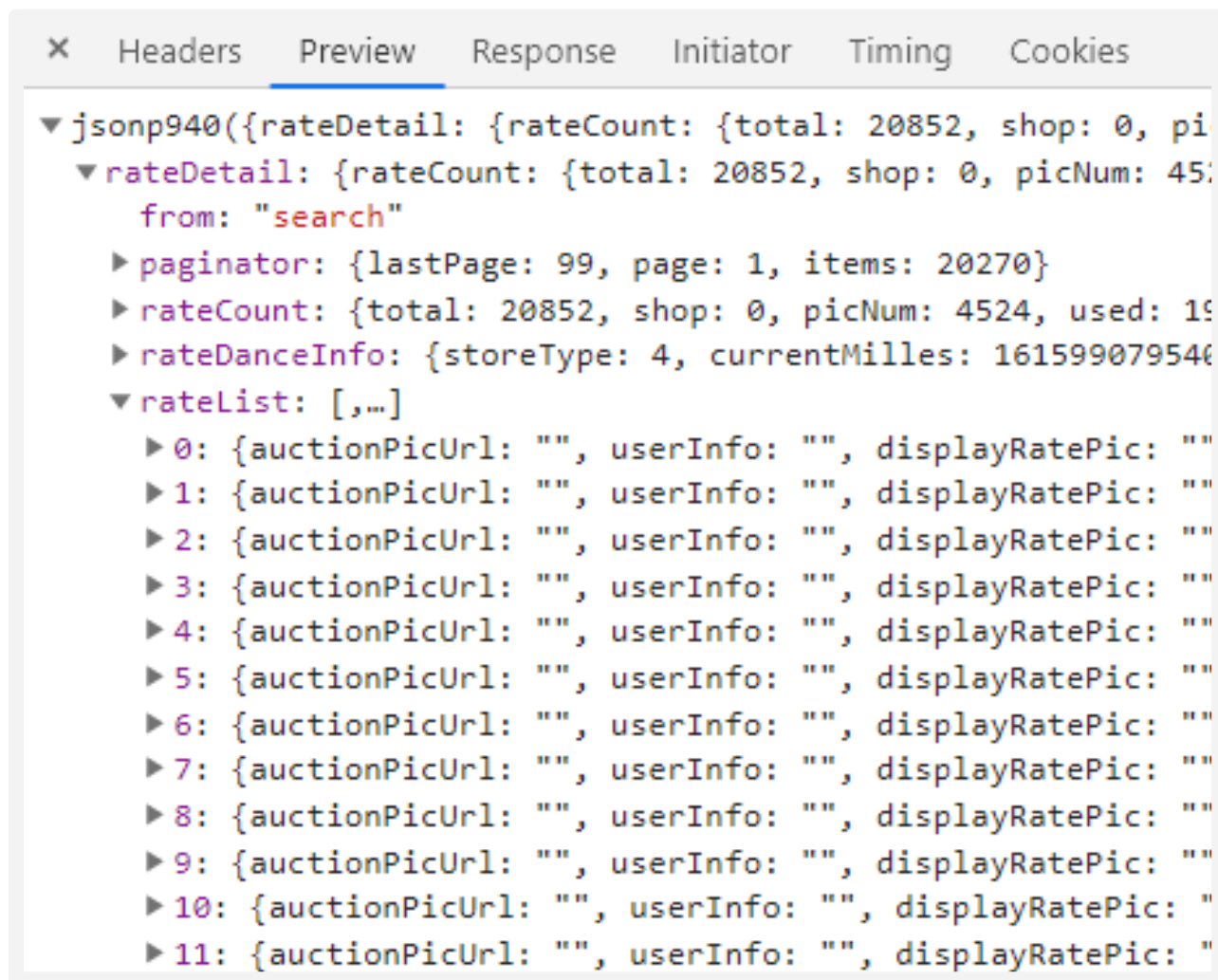
- 評論太多，瀏覽麻煩
- 篩選簡陋，不夠有效



# HOW?



# 問題一：



**content-type:**  
text/html; charset=UTF-8

看起來是Json，但沒辦法成功解碼



畫個圖

~~Jsonp908 (~~

```
rateDetail : {
```

```
    from: "search"
```

```
    paginator: {}
```

```
    rateCount: {}
```

```
    rateDanceInfo: {}
```

```
    rateList: [, ...]
```

```
    searchinfo: ""
```

```
    tags: []
```

```
}
```

Json

~~)~~

# 問題二：

由於淘寶網頁機制，評論上限99頁  
因此，實際爬到的評論僅2900條評論.....

## 標準解-APP爬蟲



Progress Telerik Fiddler Web Debugger

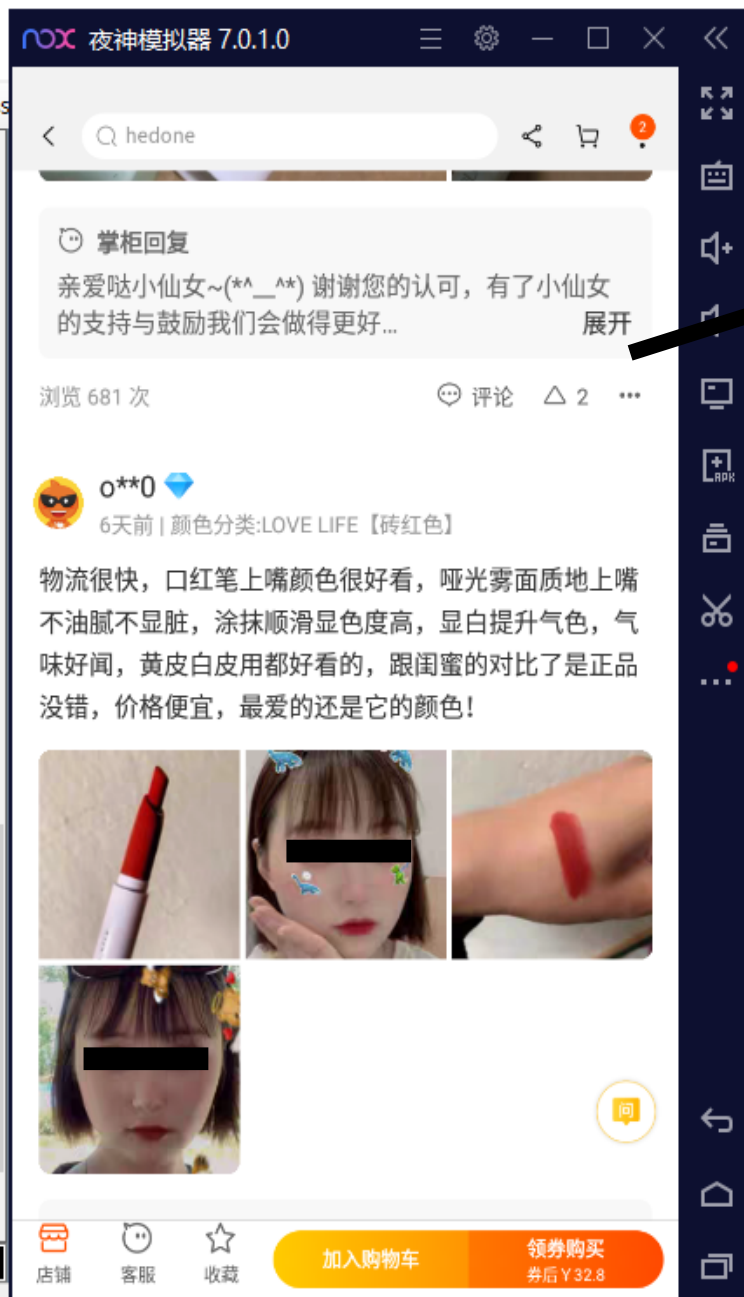
File Edit Rules Tools View Help

WinConfig Replay Go Stream Decode Keep: All sess

#	Result	Proto...	Host	URL	Body	Ca ^
60	200	HTTP	Tunnel to	gw.alicdn.com:443	798	
61	200	HTTP	Tunnel to	gw.alicdn.com:443	798	
62	200	HTTP	Tunnel to	log.mmstat.com:443	0	
63	200	HTTP	Tunnel to	market.m.taobao.com:...	0	
64	200	HTTP	amdc.m.taobao.c...	/amdc/mobileDispatch?...	1,955	no
65	200	HTTP	Tunnel to	zconfig.alibabausercont...	0	
66	200	HTTP	Tunnel to	zconfig.alibabausercont...	0	
67	200	HTTP	Tunnel to	gw.alicdn.com:443	0	
68	200	HTTP	amdc.m.taobao.c...	/amdc/mobileDispatch?...	1,093	no
69	200	HTTP	Tunnel to	dtmall-tel.alicdn.com:443	0	
70	200	HTTP	Tunnel to	v10.events.data.micro...	0	
71	200	HTTP	Tunnel to	update.googleapis.co...	0	
72	200	HTTP	Tunnel to	rhdc-ac.s.taobao.co...	714	
73	200	HTTP	Tunnel to	dtmall-tel.alicdn.com:443	0	
74	200	HTTP	Tunnel to	dtmall-tel.alicdn.com:443	0	
75	200	HTTP	Tunnel to	dtmall-tel.alicdn.com:443	0	
76	200	HTTP	Tunnel to	dtmall-tel.alicdn.com:443	0	
77	200	HTTP	Tunnel to	dtmall-tel.alicdn.com:443	0	
78	200	HTTP	Tunnel to	rhdc-ac.s.taobao.co...	714	
79	200	HTTP	Tunnel to	rhdc-ac.s.taobao.co...	714	
80	200	HTTP	Tunnel to	rhdc-ac.s.taobao.co...	0	
81	200	HTTP	Tunnel to	zconfig.alibabausercont...	0	
82	200	HTTP	Tunnel to	zconfig.alibabausercont...	0	
83	200	HTTP	amdc.m.taobao.c...	/amdc/mobileDispatch?...	1,877	no
84	200	HTTP	Tunnel to	rhdc-ac.s.taobao.co...	0	
85	200	HTTP	api-new.bignox.c...	/common/input_method	230	
86	200	HTTP	amdc.m.taobao.c...	/amdc/mobileDispatch?...	1,065	no
87	200	HTTP	Tunnel to	policy.ccs.mcafee.com:...	0	
88	200	HTTP	api-new.bignox.c...	/common/input_method	230	
89	-	HTTP	Tunnel to	zconfig.alibabausercont...	-1	
90	200	HTTP	Tunnel to	zconfig.alibabausercont...	0	

[QuickExec] ALT+Q > type HELP to learn more

Capturing All Processes 90



學藝不精  
APP爬蟲失敗

其他解-誤打誤撞

改看看真實網  
址裡的各種參  
數.....

append=1



# 成功爬到3953條評論 (2019/07~2021/03)

tao > itemId=599188374892

名稱

- BRINK
- CHI CHI
- COOKIE
- FLY FI
- GOAL DIGGER
- GOLDIE
- I&ACUTE;M UPSET
- KOI
- LOVE LIFE
- PARACHUTE
- POISON
- PRIMER
- SIKE
- SKY WALKING
- TRUST ME
- WHO RUN THIS
- ZIGGIE

itemId=599188374892

# 圖片檔

## 依型號分類

colorpop唇寫.xlsx - Excel

檔案 常用 插入 版面配置 公式 資料 校閱 檢視 開發人員 說明 自訂擴充指令 設計 查詢 告訴我您想做什麼 共用

貼上 剪貼簿 字型 對齊方式 數值 樣式 儲存格 編輯

F47 這個顏色很气场哈哈喜欢就是有点干

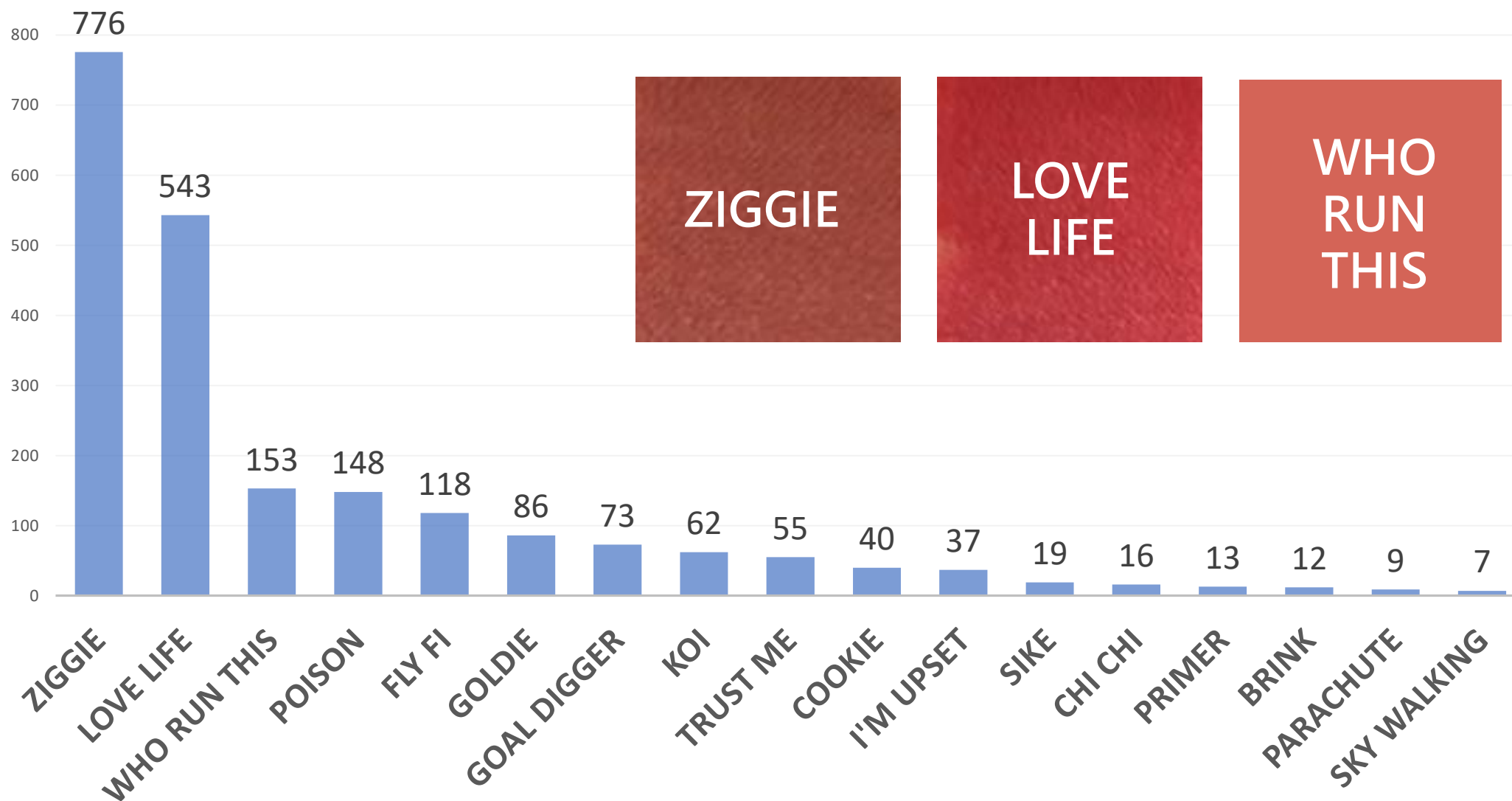
	A	B	C	D	E	F
1	id	product	1st_time	1st_comment	2nd_time	2nd_comment
2	1124256311128	LOVE LIFE	2021/3/2 10:54	<p>哑光的口红带有一点润唇的效果还不黏腻            不过嘴唇比较干的上嘴可能会有卡纹，建议先打底。上妆后的气色就好了好多真的喜欢            然后涂了感觉不油也不干很舒服以后就你们家了            我是从抖音种草过来的，这个是黄皮最爱颜色也好看拿到手就是一种高级的感觉颜色真的很好看值得入手”</p>		
3	1125291775997	LOVE LIFE	2021/3/13 17:06	<p>口红笔上嘴不拔干感觉很柔和的感觉效果很好看            实物与描述相符，着色很好，薄涂的话，素颜完全OK，厚涂的话，气场立马就上来了，最好还是化妆用，滋润度还可以吧嗯，但是如果你的嘴唇比较干的话，最好还是涂一点唇膏，我个人不用打底的，图片是我直接上嘴的颜色，不错的颜色，爱啦</p>		
4	1124369633196	LOVE LIFE	2021/3/12 14:56	<p>在佳琪直播间被种草的，colourpop家的口红真心不错，上嘴颜色超级好看~~~~哑光雾面质地而且不干，价钱还很便宜，这个砖红色太太推荐了！谁用谁好看，口红必入的一个色号。颜色真的真的好看，旗舰店是品牌直营正品保障的，喜欢的姐妹可以放心买的！！！！！！！！</p>		

文字材料 存成CSV

# 文字檔 存成CSV



# 2020年的銷售量：



評論文字雲：





+ 程式碼 + 文字



編輯



```
1 pd.set_option('display.max_rows', None)
2 fake_df = raw_df[raw_df['1st_comment'].str.contains('假货') | raw_df['2nd_comment'].str.contains('假货')]
3 fake_df
```



119	1123336025654	LOVE LIFE	2021-02-28 14:24:11	怎么说呢，本来以为收到了假货，因为打开之后把口红拧出来的时候很松，质量不行，但是呢，口红颜色...	2021-03-14 14:38:04	我的天，今天想起来用，结果揣到兜里，来到公司就这样了
544	1118649729752	GOLDIE	2021-01-11 22:56:12	壳的质感也太差了吧！我感觉好像买了9.9的假货一样....	NaN	NaN
627	1119719052699	ZIGGIE	2021-01-24 15:50:28	假货	2021-01-27 23:29:29	假货，千万别买
941	1116513686263	ZIGGIE	2020-10-06 14:40:14	此用户没有填写评论!	2020-12-23 10:28:10	假货！假旗舰店！辣鸡！！
1103	1114005686643	PRIMER	2020-12-05 13:22:57	很失望 跟我之前从官网买的一点都不一样 官网的膏体有colourpop贵有的巧克力味 这个唇...	NaN	NaN
1118	1113728285230	ZIGGIE	2020-11-13 16:02:39	此用户没有填写评论!	2020-12-06 11:19:46	原来是假货啊，我同事告诉我才知道的，**，整个口红的质感完全不一样，知道是假的我就不买了，哭...
1148	1113501269523	ZIGGIE	2020-12-04 14:17:30	买之前不知道是假货，没有下次了，再见	NaN	NaN

COLOURPOP

ColourPop V 

+关注





7月23日 23:53 来自 ColourPop超话 已编辑


 置顶 大家好，这里是ColourPop官方微博。


再次声明，**ColourPop在国内没有淘宝天猫旗舰店，没有京东旗舰店** 没有微信微店，没有中文网站，请大家通过以下官方渠道购买：

 官网  网页链接

 小红书旗舰店  网页链接

官网所有产品都已备案完毕，可以使用支付宝付款，顺丰直邮中国！注意填写收件信息时，请用中文汉字填写与身份证对应的姓名，用拼音或英文填写详细地址（如图二）。

超简单的购买流程请看下图，也可以点击  网页链接进入官网查看哦！

等订单寄出后复制rrd快递单号到  网页链接查看物流信息，找到顺丰单号。有关订单和物流的问题，请通过  网页链接联系客服咨询。

欢迎大家来我们官方群玩  ColourPop今天上新了吗

#colourpop#

商標蟑螂  
中國廠商搶註商  
標.....

# 延伸&應用：

## 延伸

### 建立一個評論 等級分類器

- 天貓評論沒有標明好評、中評、差評。
- 爬到資料之後還要手動過濾麻煩。
- 分類後，可以直接撈差評出來看

## 應用

### 分析商品評論 文本內容


- 捕捉顧客對於產品的體驗，解析顧客需求
- 以客戶的使用回饋去改善商品品質
- 研究你的競爭對手





# 天貓商品評論爬蟲

—以ColourPop唇膏的商品頁面為例



報告者：黃意程

2020/03/25

