



文本分析與程式設計



I have no idea ●

2020.12.22

What and Why

社交媒體上 一定時間範圍內的常用食材

- 證實“常見”又“常用”
- 稍稍了解飲食趨勢或習慣
- 觀察可能的迷思或熱潮

可能的應用：營養或飲食媒體





Target Problem and Solution

食材使用率
名字出現的頻率

- 是否包含特定品牌、商家
- 可能的或普遍的烹飪方式
- 可能的形容詞
- 潛在的熱潮迷思

Data

Instagram：食譜部落客、
example: betty.bento

Why instagram: 年齡層、

其他可能性
Facebook:
Youtube



betty.bento

Follow

1,112 posts

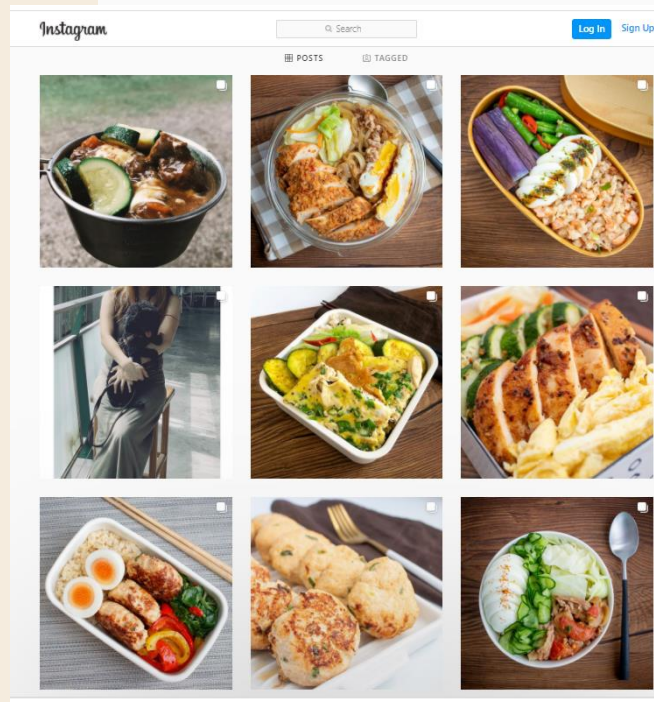
306k followers

609 following

貝蒂做便當

我寫了三本瘦身料理書👍

開啟輕鬆下廚、健康瘦身生活。



新出版：《愛妻瘦身便當【減醣成功三部曲】》

2018年出版《#愛妻瘦身便當》

2019年出版《愛妻無壓力瘦身便當》

❤️各大網路書店、實體書店都可以買到喔～

inktr.ee/betty0620

Procedure

文本類別

- 分類文字/圖片/影像
- Pytesseract
- 識別的正確率

文本分析

- 抓出每篇食材
- 統計結果

1

2

3

4

篩選目標

- 選帳號
- 標註# hashtags

前處理

- 每篇資料語言
- 斷詞

questions

- 要怎麼挑選分析對象?
 1. 資料上：香港人
 2. 可信度評估依據：追蹤數？按讚數？
- 要怎麼有效率的看每一篇貼文？
- 要怎麼處理中英雙語食譜？
- 要怎麼總整處理結果？





Reference

圖片變文字：pytesseract <https://reurl.cc/VXrxnb>

