

楊承燁 E24056572 電機三乙

問題一：從 Kaggle 中找一個資料集，描述此資料相關資訊，

資料集標題：谷歌應用程式商店 app

資料集說明：

While many public datasets (on Kaggle and the like) provide Apple App Store data, there are not many counterpart datasets available for Google Play Store apps anywhere on the web. On digging deeper, I found out that iTunes App Store page deploys a nicely indexed appendix-like structure to allow for simple and easy web scraping. On the other hand, Google Play Store uses sophisticated modern-day techniques (like dynamic page load) using JQuery making scraping more challenging.

資料集內容：

包含了收錄的 app 的名稱、分類 1、排行、大小、下載數、查看率、付費或免費、費用、內容分級、分類 2、最後更新時間、現行版本號與支援的 Android 版本。

分類 1 為主要的分類，只有一個，而分類 2 為較大的分類，一個應用程式可以屬於一個以上的分類。例如：UNICORN - Color By Number & Pixel Art Coloring 這款 APP 的分類 1 為藝術設計，分類 2 為藝術設計與創造力。

資料集更新狀態：

資料集自 2018 年 9 月 5 日發布，截至 2019 年 2 月 3 日更新至版本 6。目前(2019 年 7 月 2 日)為 Kaggle 上投票排名第 4 的資料集。

Kaggle 連結：<https://www.kaggle.com/lava18/google-play-store-apps>

問題定義：

這個資料集並非從 Kaggle Competitions 提供，而是 Kaggle Dataset，所以並沒有存在所謂的問題，但是由 Dataset 的 Kernel 可以發現大部分的人使用這些數據集都是再進行探索性數據分析，探討如何在 Play 商店上獲得較高的排名。

潛在問題：

除了如何獲得較高的排名以外，透過分析這個數據集可以提供開發商預測新應用程式的生命曲線，用以評估需要投入多少的資本與其預期的收益。

分析與預測難度：

資料分析難度並不高，從這項數據集，用戶可以很簡單的統計出哪一種類型的應用程式在哪些時期有比較高的人氣，但是預測難度可能會比分析高得多，近期 Android 生態系上有一個顯著的現象是用戶傾向使用已經熟悉的、以前就經常使用的軟體，儘管換了一隻手機，用戶仍然傾向安裝老軟體，而少了在應用程式商店多加瀏覽的動力，而在短期內吸引大量安裝的軟體通常都是突發性流行，事前的預兆並不明顯，再者，單單觀察此資料集也可以發現，這個資料集在時間軸上的資料比較少，比較多都是截至某一個時間點的資料，因此我們很難預測在下一季流行的應用程式會是什麼。

價值：

如果可以成功預測未來的流行趨勢，那對於小型的應用程式開發商應該會相當有利，小的程式開發商不像大公司往往握有許多老牌用戶，因此獲益的機會大都在於如何開發生命週期短但是又剛好命中當時流行的應用程式，這也正是其價值所在。

問題二：描述一個在你學習領域的資料或是日常生活中可取得的資料，說明可以用來產生何種價值

資料：

手勢資料，藉由分析人手的運動資料可以進行分類，藉此判斷人的手勢

獲取方式：

穿戴式裝置，在手環、手錶或是手套上利用陀螺儀、加速計蒐集慣性資料，除了穿戴式裝置以外，我們亦可以記錄人們日常書寫工具的運動資料。

利用方式：

空中書寫辨識

價值：

目前的手寫辨識大都侷限於圖像辨識的領域，人們在一個預先被規畫好的面板上進行書寫，在現階段人的互動裝置大多是一塊固定的面板，然而我相信空中的書寫辨識有其發展性，從未來的觀點來看空中書寫在可能沒有實體面板的將來來說會成為人機互動方式之一，現階段來說他也對於啞人士有一定助益，此外如果裝置並不是侷限在穿戴式裝置，而是在日用品上例如筆，那麼這個裝置便可以提供用戶一個在任何地方都可以對電腦書寫的利器。