# Project - Progress Report

## Project motivation:

之前在升大一的暑假，為了升級電腦配備花了很多功夫，光是一個 CPU 就有價位、效能和相容的規格等許多必須研究的細項，之後還要去爬文尋找這個 CPU 的評價如何，是否有什麼災情。何況一台電腦並不是只有 CPU，每個組件這樣精挑細選下來就得花上更多時間。所以我想要為正在挑選配備的人打造一個能輕鬆蒐集並整理組件的販售資訊與相關評價的程式。

## Plan description:

那首先，必須先獲取市場上大部分零件的資訊，所以呢，第一步就是先用爬蟲的方式從原價屋的網頁獲取商品的種類、價格、廠牌與規格型號等等，基本上只要在原價屋估價頁面看的到的都先抓下來。

再者是關於評價的部分，我個人在需要查找這類評價時都會優先從 PTT 的 PC_shopping 版下手，所以會用到其他人做好放在 Github 上的 PyPtt 這個 API 套件。主要目標是從 PTT 版上取下相關的討論、評價或是開箱文，可能只做到標題相關，能否內文相關還不確定。

最後，由於原價屋不會附上一些較細節的規格，必須去官網才找的到，而且有時只參考 PTT 版上的討論是不夠的，最後會導入 Google custom search 這項 API 提供使用者更多元豐富的資訊來源，大概顯示 5~10 項結果與其 URL，預期可能的內容有產品官網、其他的購物網站或來自 PTT 以外的評價文章。

## User interaction:

而在程式使用上呢，先要求使用者輸入想查找的商品類型及型號，並從原價

屋取出並顯示所有販售中的品項，接著可再輸入指令在原價屋資訊、PTT 相關討論或 google 搜尋結果三者間切換。在搜尋產品時的輸入必須能無視大小寫，可能會規定特殊的輸入格式，若查無此項商品必須顯示查無此項。切換到 PTT 文章時，若文章大於一定數量則導入翻頁機制，當然，對於錯誤的輸入必須有所回應。

## Update 1:

1. **DONE:**
   爬取原價屋網頁(克服編碼、網頁架構問題)，將資料儲存成可轉換成 json 檔的格式、測試 pyptt 套件能否使用。
2. **CHANGE:**
   pyptt 套件需要有 ptt 帳號才可使用，目前找不到其他可替代的 API，改變計畫移除顯示 ptt 文章的功能。用 regular expression 擷取價格後用 google chart API 顯示此類型商品的價位分布。

## Timeline(updated):

1. 確認上述每項提到的內容能否成功地與網頁互動並抓取資料下來 (Before 12/20)(√)
2. 爬取原價屋的網站，並提取出商品資訊，視進度將其作為第一次 Demo (Before 12/24) (√)
3. 完成 Progress report (Before 12/25) (√)
4. 搜索 PTT 的文章，將所有相關討論抓取並顯示 (X) (Removed)
5. 測試並使用 google chart API，整理價格資訊 (Before 1/7)
6. 透過 google search API 提供其他的相關資訊 (Before 1/7)

7. 完善使用介面，進行 Bug 偵測 (Before 1/10)