



Python Programming – crawler to social media 「DCARD」

專案目標：

這份爬蟲的目標是主要是想要調查關於 Dcard 版上熱門文章排名背景以及曝光效益，圍繞四個問題。

1. 「男女生誰比較喜歡上網討拍？－談熱門文章排行之性別比。」
2. 「卡到龍穴是否一切飛黃騰達？－談排行前面的文章是否曝光度會較高？」
3. 「卡到龍穴是否一切門庭若市？－談排行前面的文章是否互動率會較高？」
4. 「越長越好？數大真的便是美嗎？－談標題文案長度與排行之相關性」

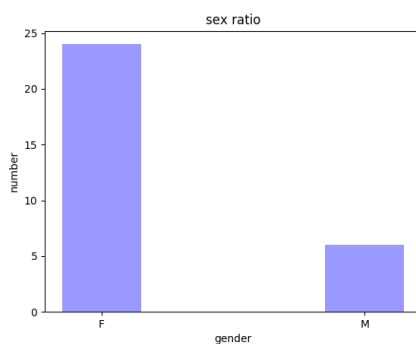
專案動機：

透過調查熱門文章的效應，想要了解排名的版位是否會影響文章的日後發酵，究竟內容為王，還是版位為王。除此之外，也想調查一項都市傳說，「品名越長的商品越容易在菜單上被點選」那是否標題越長的文章越容易登上熱門？

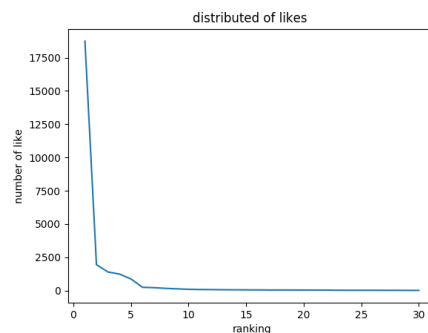
後續用途：

透過排行的效益差異，對於日後想要針對特定族群做行銷推廣可以更有方向，也更有效率，原因在於能夠選擇主要流量集中的排行，捨去後面的長尾，將資源與時間投注在特定的大群即可。除此之外，也想要尋找壹些如何登上熱門的獨門訣竅。

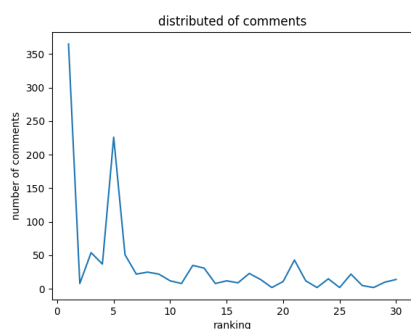
圖表解說：



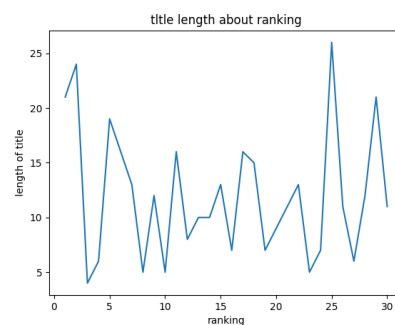
(熱門排行前 30 名之性別比)



(熱門排行前 30 名之 like 分佈)



(熱門排行前 30 名之留言分佈)



(熱門排行前 30 名之標題字數)

現象觀察：

從各個數據圖我們可以來一一檢驗先前所提出的四個問題，首先從性別比的分佈圖來看，我們可以得知在心情板上女性發文數量較男性多（以熱門前三十名為樣本），從此得知，女生朋友在心情不好時，最重要的需求與最好的陪伴就是「聆聽。」再者，從第二三張圖來看，熱門前 30 名之愛心數與留言分佈，有絕大部分的曝光與互動量都集中在前 5 名，5 名之後不論名次前後，曝光與互動率皆所差無幾，從此得知，若想要讓文章的倡導議題，或是分享內容讓更廣的大眾所接觸到，必須努力讓文章擠進排行前五名。而最後，從標題字數的分佈圖來說，上下起伏的波動非常大，從此可知，都市傳說的數大便是美在社群平台「Dcard」的熱門文章上不成立。