資庫管作業五

第二組：

張廷鋒、黃竫霖、凃兆蓮、陳恩宇、吳玨青、陳柏瑞

大綱：

1. 研究目的
2. 組內分工
3. 研究方法
4. 研究過程（SQL、python程式碼解說）
5. 分析、結果發現
6. 結論

* 研究目的

於ptt mobilecomm 看板及dcard 3c 版做觀察、研究，找出今年討論度最高的3c產品，進而對今年討論度最高之手機iphone12, samsung s20 做個方面的分析、比較，得到具有商業價值之資訊。

* 組內分工

張廷鋒：Python、SQL資料分析、書面報告

黃竫霖：SQL資料分析、書面報告

陳恩宇：SQL資料分析

凃兆蓮：資料視覺化、ppt製作、書面報告、錄製影片

吳玨青：查詢網路資料

陳柏瑞：討論場地租借、其他雜項事務

* 研究方法

使用 ptt mobilecomm 及 dcard 3c 版之資料，結合上網搜尋之公開資訊做分析。分析工具使用SQL及python。資料視覺化工具使用excel 及 tableau。最後使用經上述處理之資料分析出具有商業價值之insight。

* 研究過程

我們設計了一個計算討論度的公式，計算方法如下。

**討論度公式=標題提及數 \* 10 + 內文提及數\* 5 + 留言數**

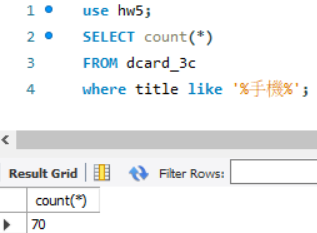
舉例：以手機為例，只要任一篇文章的標題有包含「手機」二字，則手機的討論度加10分；任一篇文章的內文有包含「手機」二字，則手機的討論度加5分；任一篇文章的標題或內文有包含「手機」二字，則該篇文章每則留言都會幫「手機」討論度加1分，若標題跟內文都有提及，則只會加一次留言分數。

根據這個討論度公式，本組設計了10個問題。

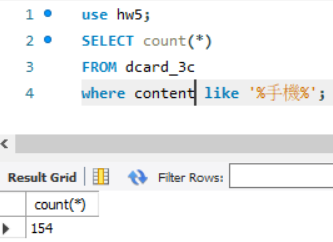
1.從dcard的3c版中找出各種3c產品的討論熱度

作法(SQL):

(1)使用SELECT count(\*)、WHERE title like’%產品%’，找出各產品的標題提及數



(2)使用SELECT count(\*)、WHERE content like’%產品%’，找出各產品的內文提及數

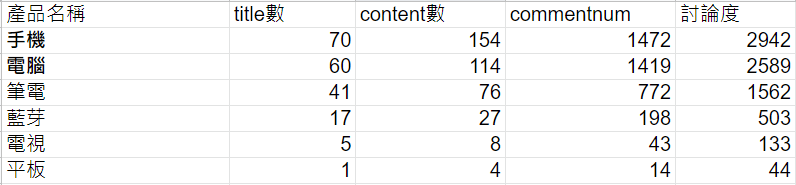


(3)使用SELECT sum(comment\_count)、WHERE title like ‘%手機%’ or content like ‘%手機%’，找出各產品的留言分數



(4)運用excel內的函數運算**討論度公式**

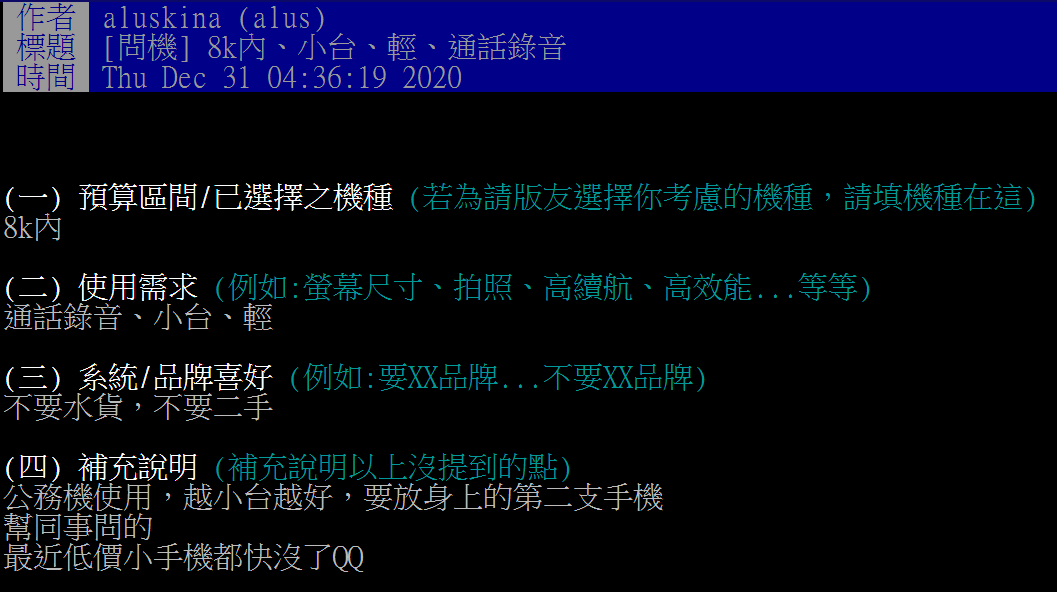
結果:



2.大眾買手機時最注重哪些方面

作法(Python):

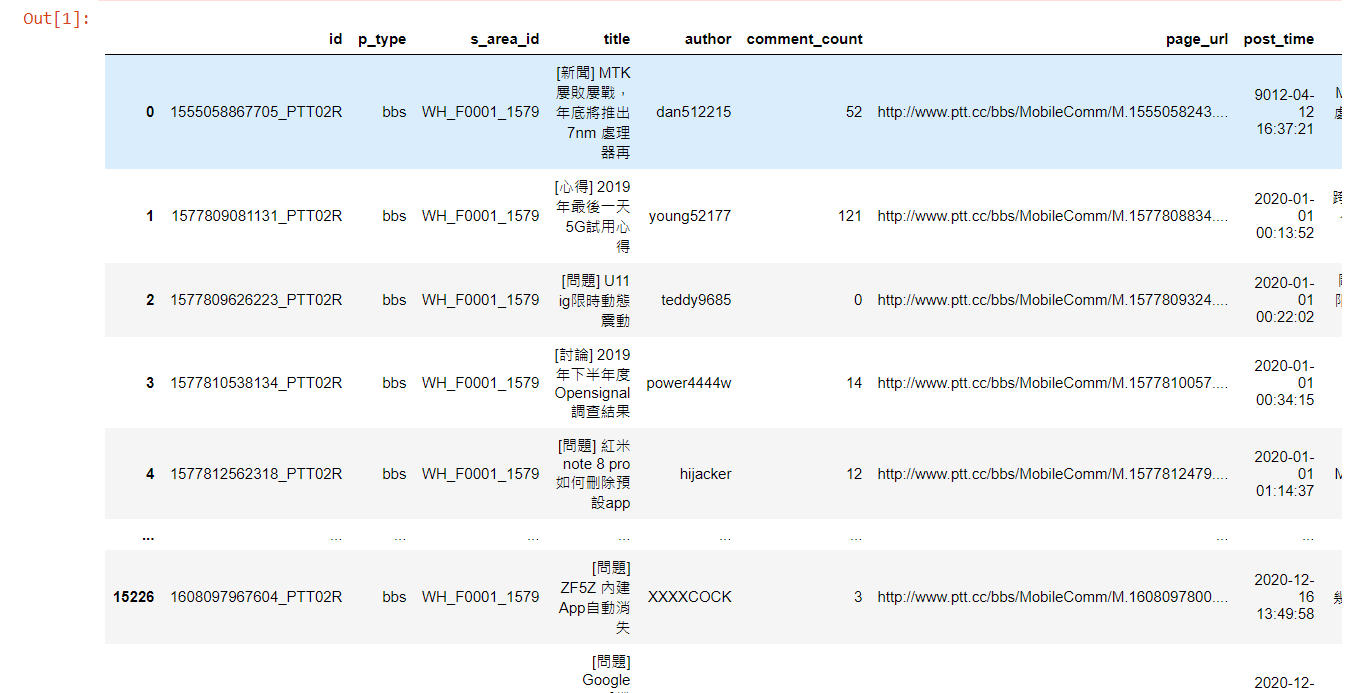
(1)我們透過觀察發現，ptt\_mobilecomm版上有一種發文分類稱為[問機]，**問機文**是大眾買手機不知如何挑選時向ptt鄉民們求助的方式，通常會在標題跟內文打上自己的使用需求、預算、品牌偏好等，所以我們想透過大量分析**問機文**來了解目前手機市場上消費者的偏好



(2)用python3的pandas模組中的read\_csv函數讀取並清理ptt\_mobilecomm版的資料



pandas dataframe



(3)先截取需要的欄位，再從所有文章中將問機文過濾出來



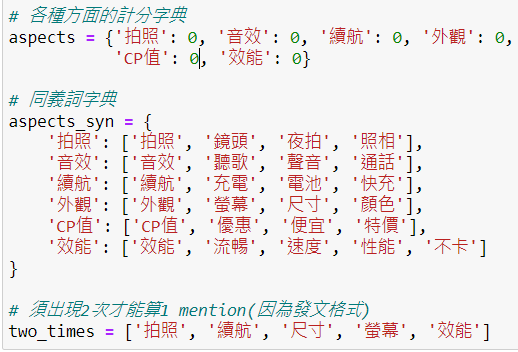
問機文的dataframe，總共有1513篇



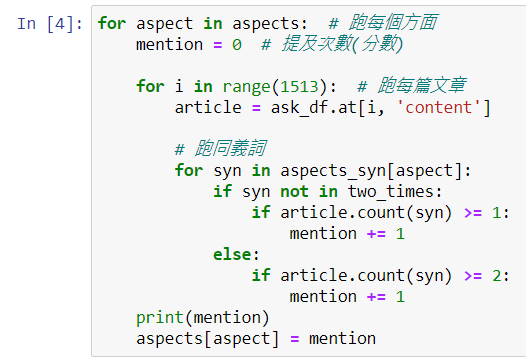
(4)先建構消費者挑選手機的六大面向dataframe，以及六大面向的高相關詞彙dictionary(舉例:拍照的高相關詞彙為鏡頭、夜拍、照相)

**計分方式**:只要該篇文章中含有該面向的高相關詞彙或該面向詞彙，則該面向的重要度加一分

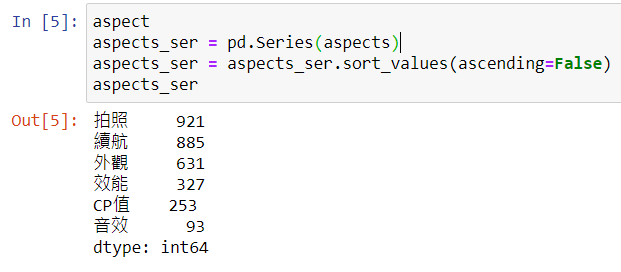
另外，我們觀察到由於ptt版上固定的發文格式，在大部分的文章中，「拍照、續航、尺寸、螢幕、效能」必定至少會出現一次，為了避免影響統計結果，我們設定這些詞彙要至少出現兩次以上才能加一分



(5)用for迴圈搭配find()函數計算分數



(6)排序後輸出結果



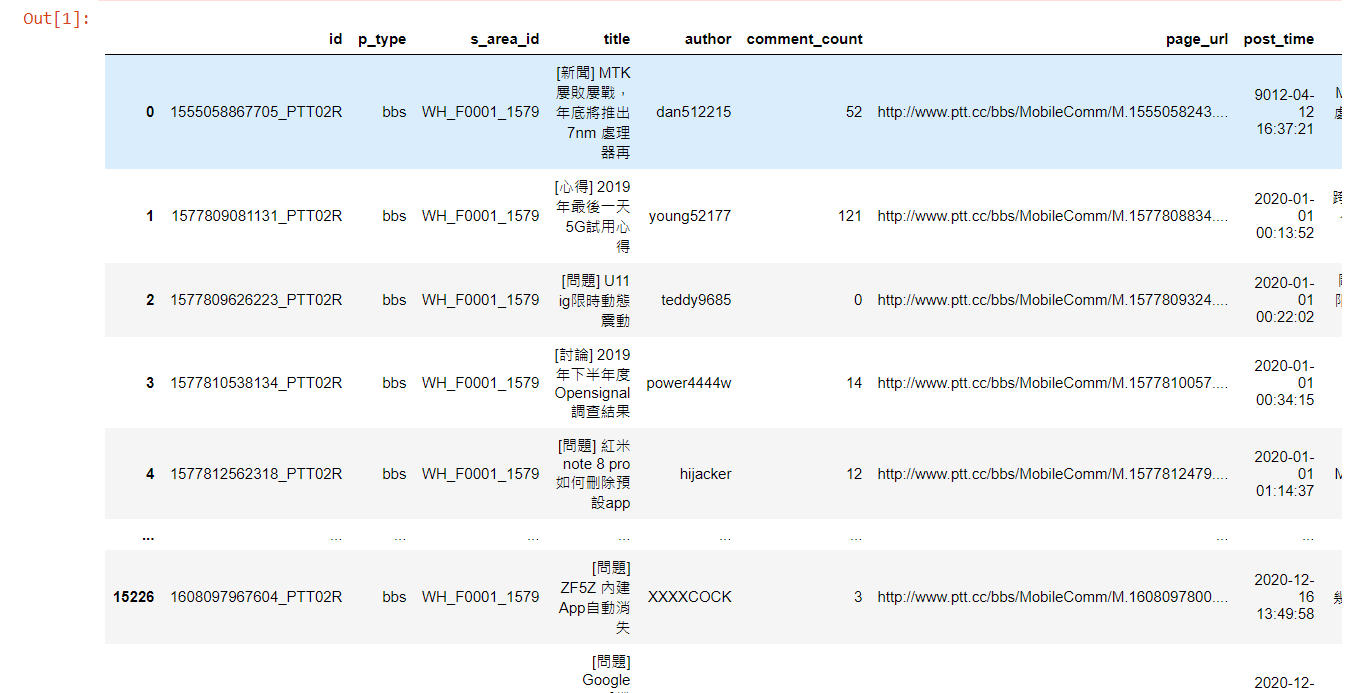
3.各種機型討論度比較

作法(Python)

(1)用python3的pandas模組中的read\_csv函數讀取並清理ptt\_mobilecomm版的資料



pandas dataframe



(2)建立計分dataframe和各機型高相關詞彙

**高相關詞彙:**

舉例來說，一般人稱iPhone 12有非常多種稱呼，i12、iphone12、iPHONE12都有可能被鄉民拿來指「iPhone 12」這支手機。

一般python和sql在進行文字比對時並沒有辦法直接將以上幾種稱呼視為同一支手機。為了避免別稱影響統計結果，我們決定建立dictionary並調整計算方式，讓演算法自動將以上幾種別稱都視為iPhone12。

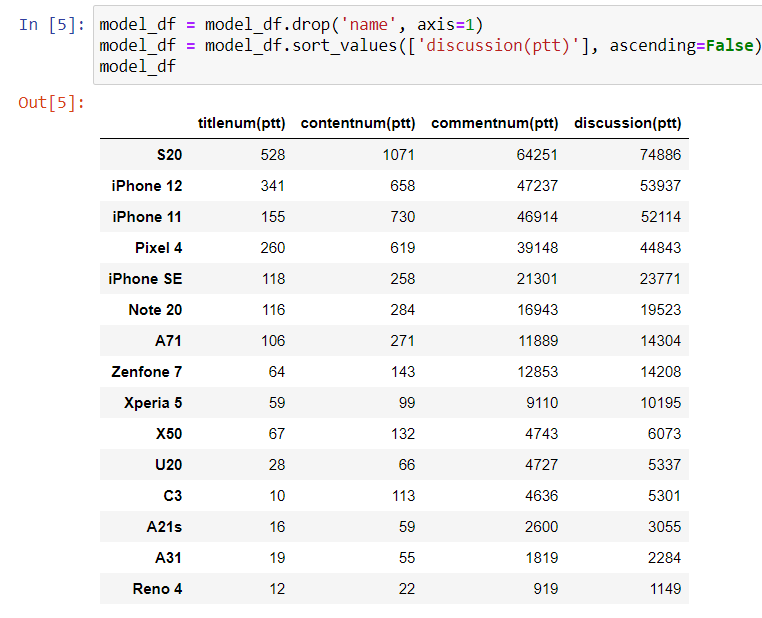
而這種狀況又以華碩的ZenFone 7最為明顯，幾乎全部文章都將這支手機稱作「ZF7」而非他正式全名。



(3)用演算法計算各機型**討論度**得分



(4)再次清理資料、排序後輸出結果



4. iPhone 12 、Samsung S20各面向常被提及的優缺點

作法(Python):

(1)建立需要的dataframe和dictionary



(2)用演算法計算i12的優缺點

**被提及優缺點**：分為六個面向及其高相關詞彙，當某個面向的高相關詞彙和iPhone 12的高相關詞彙被包含在同一篇文章時，若該文章情緒為P則視為優點；情緒為N則視為缺點。



(3)對S20做同樣的事



(4)在excel中計算結果



5.iPhone 12 的競品分析

作法(Python):

(1)先建構所需dataframe和清洗資料



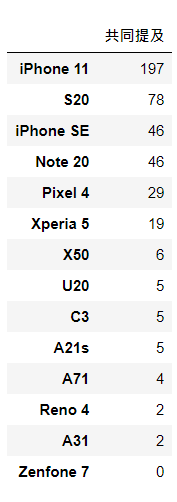
(2)使用演算法找出競品

**競品**：該手機和i12在一篇文章之中同時被提及時，競品分數+1



(3)排序並輸出結果，找出最常被共同提及的競品

https://lh3.googleusercontent.com/bEUAy15BRa8-CVfB9dFU0kQMksBB7__FuawaGxmuYKrsDmIaEqj1gwAjvkyWAv1t70T0jF5OnxpzqhSpOxDt_PUrHVzzkvOZyQTBDtHkLfy4Xe_8mIcnTrUDqisOQGExsF3W6Jvd



6.品牌討論度

各品牌篩選的關鍵字

蘋果、apple、iphone

三星、samsung

Oppo

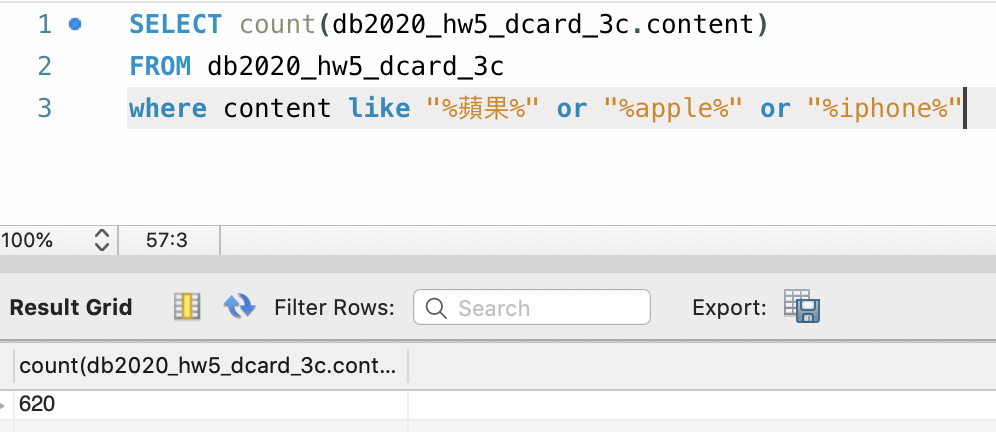
vivo

Realme

Sony

華碩、Asus、zenfone

（內容中被提及的數量，其他標題和留言量方法一樣，再各自加權加起來）

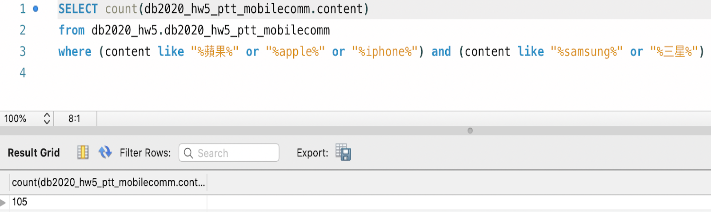


7. 各品牌形象分數（正向、負向）

分兩種算法，一種是正向扣掉負向，另一種是正向除以負向（避免因為討論度的差距造成數值大小的不同）

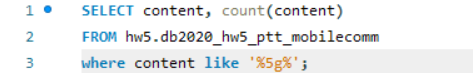


8.蘋果競牌分析

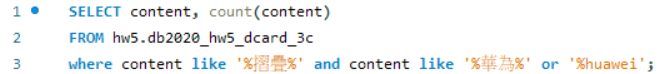


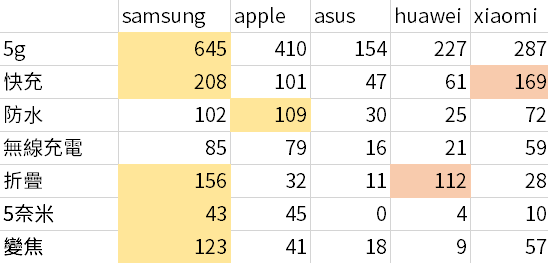
9.2020手機科技趨勢—找出最常出現關鍵字

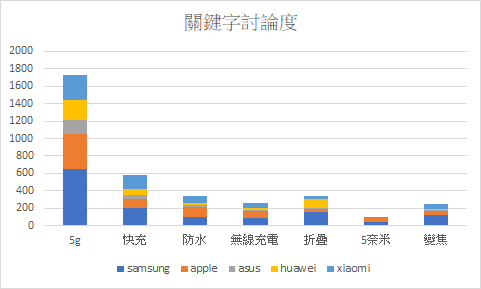
5G(3488則)、快充(839)、防水(648)、無線充電(432則)、變焦(294)、全螢幕(292)、折疊(174則)、臉部辨識(134)、5奈米(128則)、多鏡頭(47則)、螢幕下鏡頭(28)、無邊框(26)



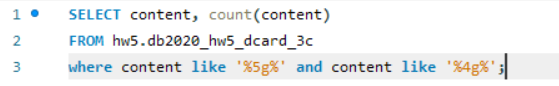
10.找出各種技術中討論度最高的是哪個廠牌







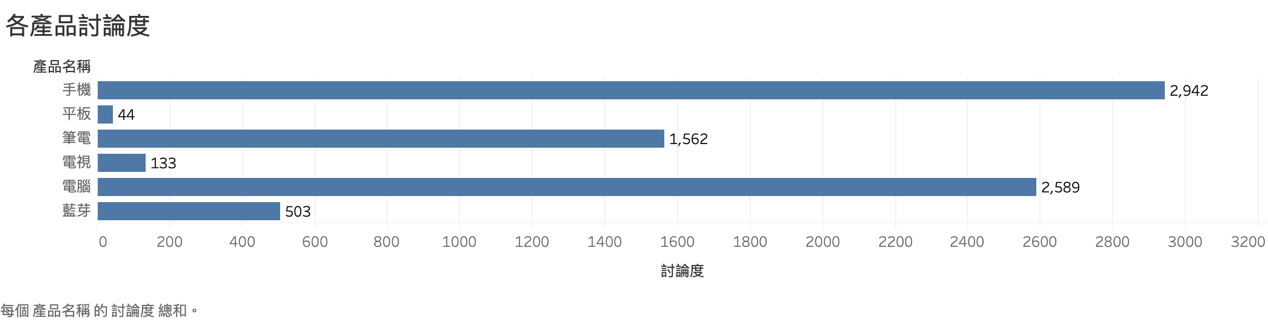
11.找出大眾對5G關注度最高的面向



https://lh4.googleusercontent.com/Oqkeilf1XDGS60Qy7SRTR2ZoiVfRxGoY8zu3bGGcOIYV8gYU0uW-PmfT_JWcrE7nF0VPtNJ-KNuZ0LNB8AT6xp94RWHud8h9XM1dp90OUzDWhNpM-ji40Luuk10MZ0_2QRktFkxn

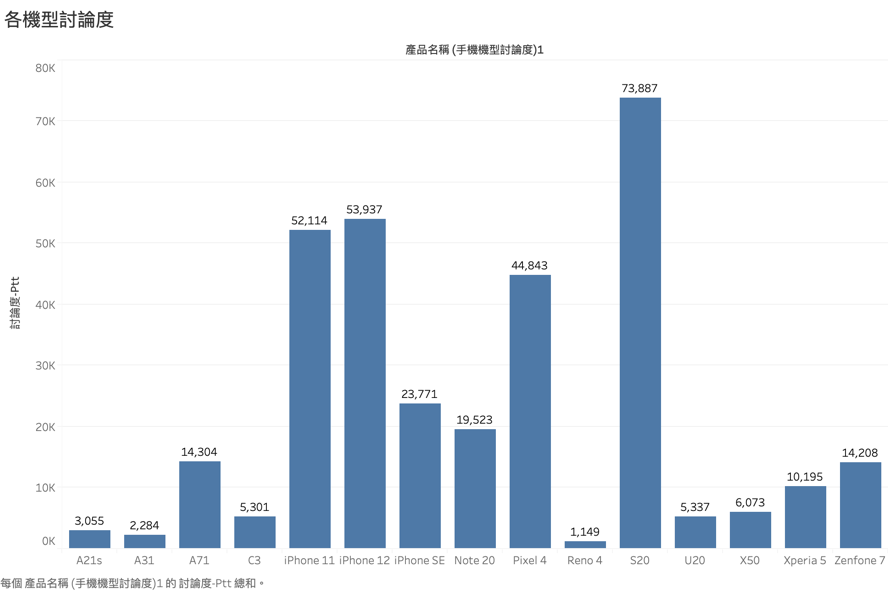
* 分析、結果發現

我們首先在dcard3c板上按照討論度公式，找出2020年**討論度最高的產品為手機。**

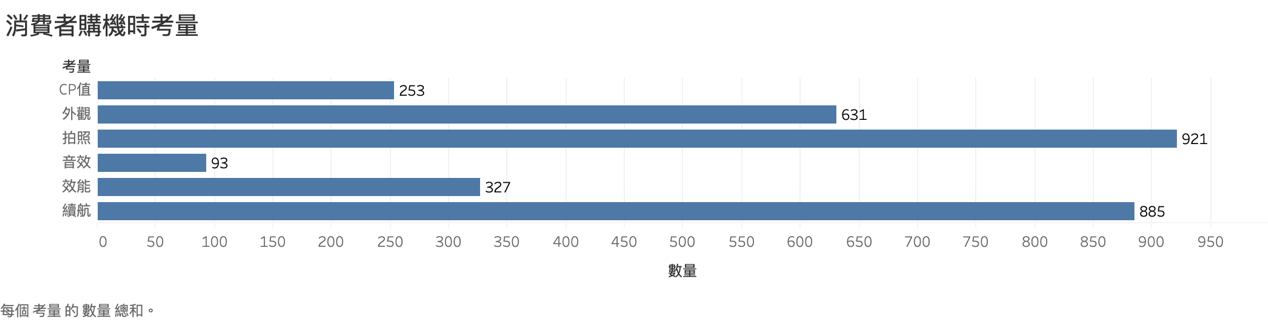


接著再進一步看討論度最高的手機機型為何者，從圖中可以看出**討論度最高的為Samsung S20（討論度73887）** 第二及第三名分別為iPhone12（討論度53937）及iPhone11 (討論度52114)。

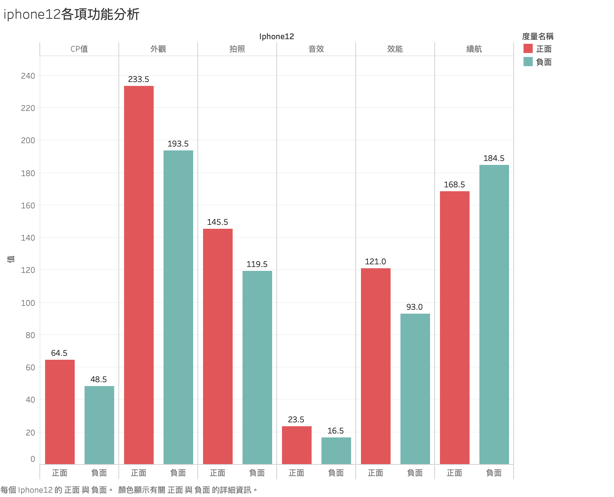
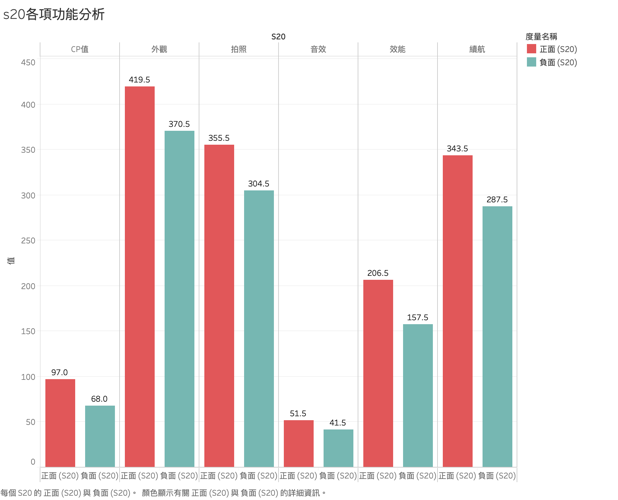
其中Samsung s20的上市月份為三月，2020平均每月討論度為7388.7，在所有機型裡排名第二 ; 而 iphone12 上市月份為10月，2020平均每月討論度為17979，是所有機型裡最高。



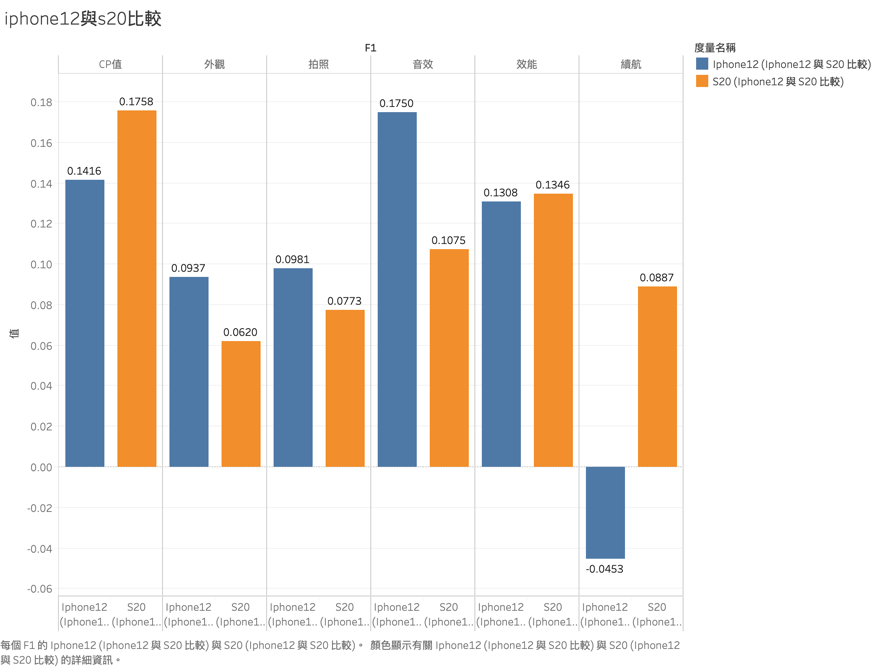
下一步為了瞭解Samsung S20及iPhone12是否有符合當今消費者買手機時的主要考量，對此項目進行分析，由圖中可看出拍照、續航、外觀為消費者最注重的要素前三名。



分析消費者對於S20及iPhone12各項功能的評價，iPhone12大多功能正評大於負評，但在續航部分負評是大於正評的，相較之下S20各個面向皆是正評大於負評。

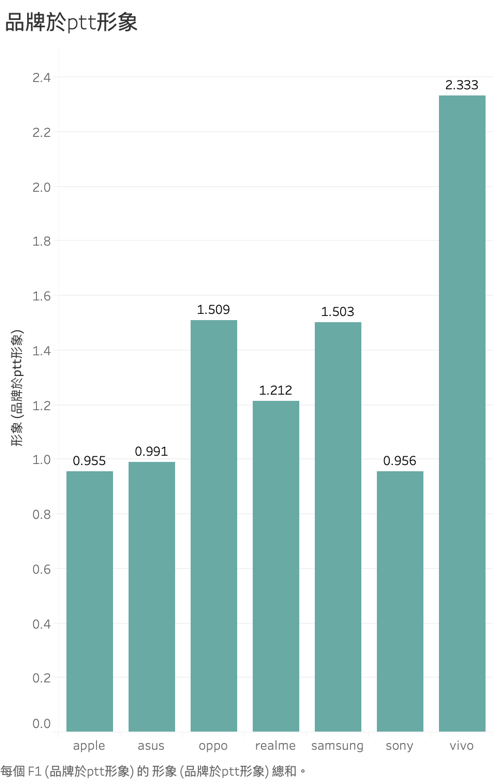
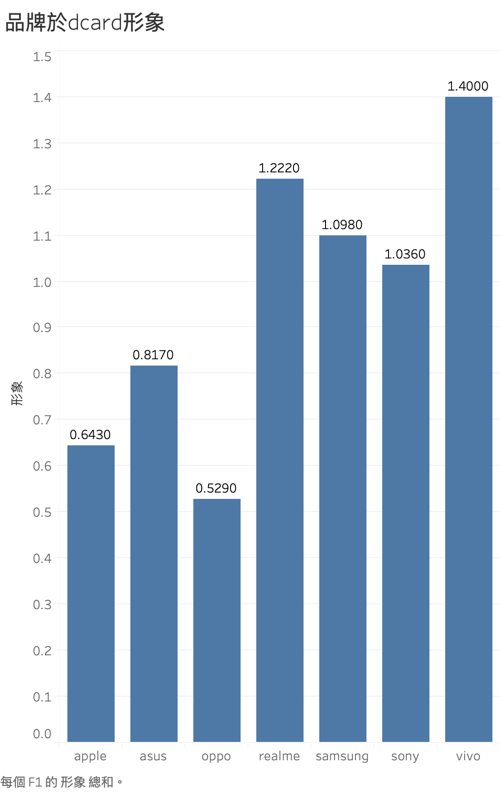
將評價數據標準化之後兩種機型進行比較，可以從下圖中看出iPhone12在外觀、拍照、音效上勝出；S20在cp值、效能、續航勝出。



接著找到兩種手機2020年的銷售量旁行名次，發現S20雖然討論度高，銷售量卻不盡理想，主要因為上市時間正逢疫情肆虐時期，景氣、銷量皆受到重創；iPhone12則是在上市之後銷售量一直獨占鰲頭。我們也由此發現銷售量與討論度未必呈正相關。

|  |  |  |
| --- | --- | --- |
|  | samsung s20 | iphone12 |
| 3月 | 20 | 未上市 |
| 4月 | 18 | 未上市 |
| 5月 | 未進入前20 | 未上市 |
| 6月 | 未進入前20 | 未上市 |
| 7月 | 未進入前20 | 未上市 |
| 8月 | 未進入前20 | 未上市 |
| 9月 | 未進入前20 | 未上市 |
| 10月 | 未進入前20 | 1 |
| 11月 | 未進入前20 | 1 |
| 12月 | 未進入前20 | 1 |

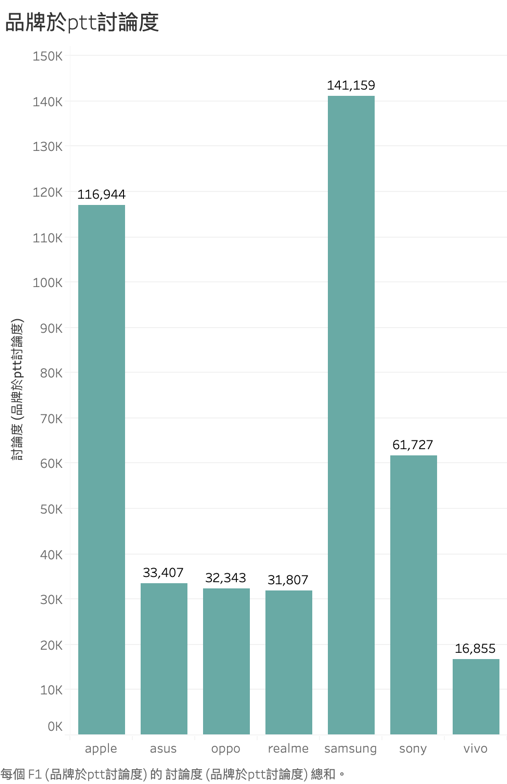
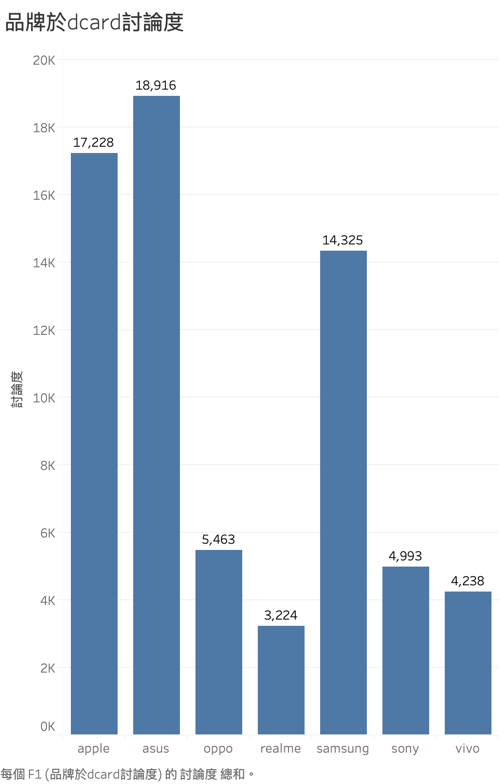
雖然iPhone12討論度沒有S20高，卻能在銷售的部分由極佳的表現，我們希望能探討這背後是否是因為蘋果品牌方面的關係，而帶動iPhone12的銷售表現，因此我們接下來觀察ptt\_mobilecomm和dcard 3c版，針對2020年手機市佔率前七名的品牌形象進行分析。結果發現，vivo在這兩個看板中皆有最高的形象分數，而蘋果在ptt中排名最後，在dcard中也位列倒數第二，由於在dcard中，蘋果的負面評論裡沒提到iphone的佔 132/154，推論蘋果的形象可能被其他產品拉低。



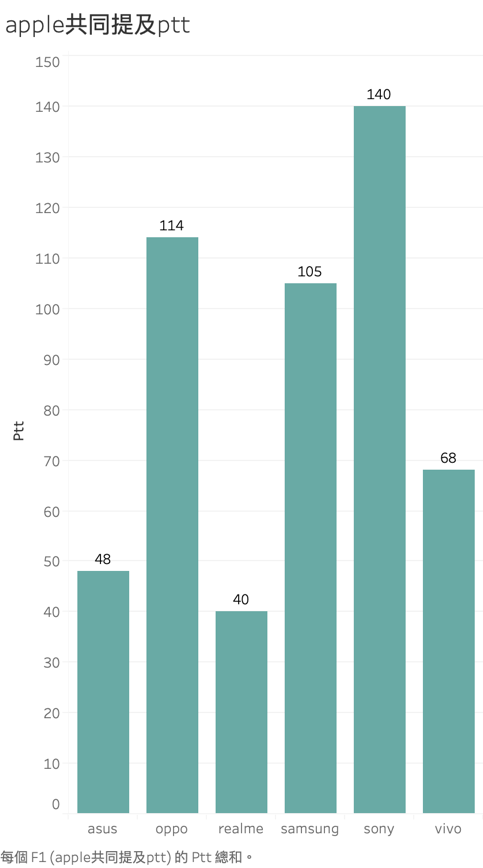
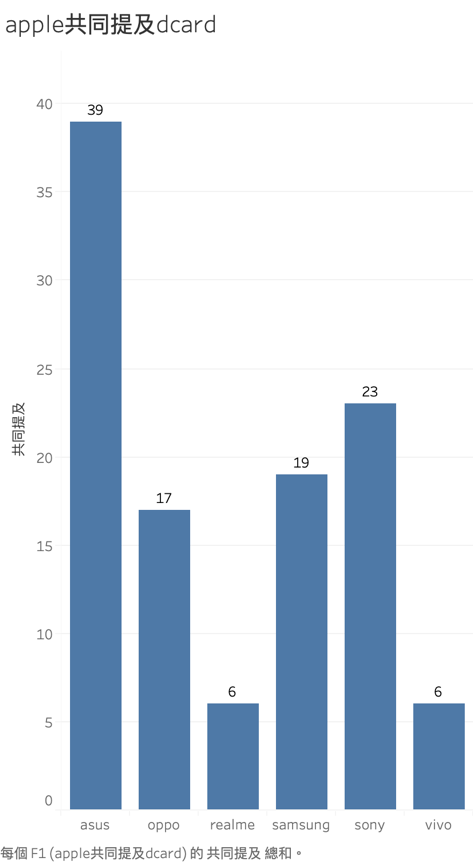
接下來是品牌的討論度，蘋果、samsung兩大廠牌在各版均有頗高討論度。

dcard上討論度前三名為 asus、蘋果、samsung；品牌於 ptt 討論度前三名則是 samsung、蘋果、sony，其中asus在dcard中特別高的原因是因為其中有299則

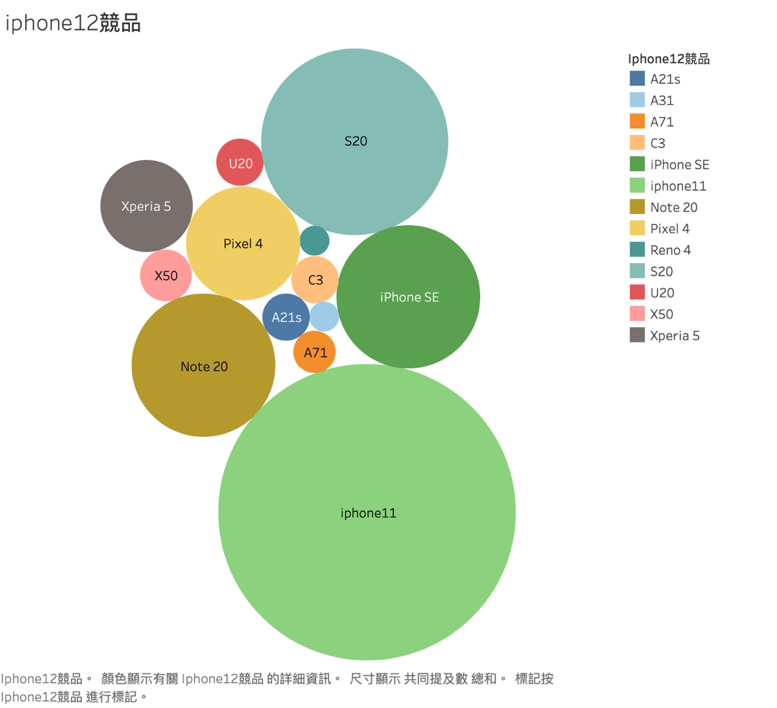
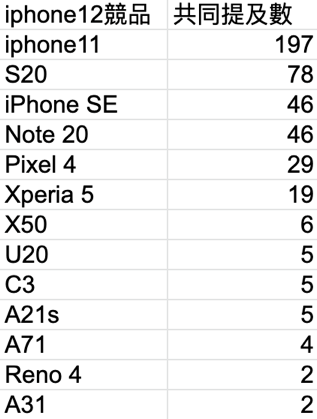
在討論電腦，因此其討論度甚至超越蘋果及samsung。



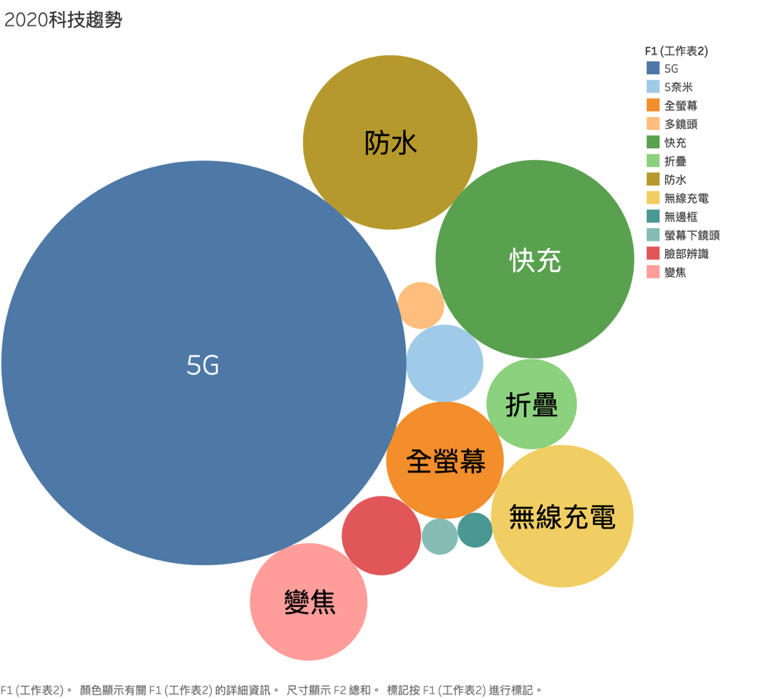
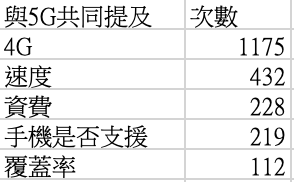
接下來觀察蘋果最常共同被提及的品牌，在dcard中蘋果最常和asus一起被提起，可能也和asus在dcard討論度相當高有關或是asus的電腦亦常常被一同和蘋果的電腦被提起；在ptt中，sony最常和蘋果一起被提起，從前像的分析結果中可以得知消費者相當注重手機的拍照功能，而sony在鏡頭、感光元件等部分相較其他品牌表現較為優異，因次在以手機為主的ptt\_mobilecomm板上較常被和蘋果一起被提起。



由於上述分析未能找出iPhone12銷量勝過S20的原因，因此我們對iPhone12進行競品分析，從圖中可以看出iPhone11為iPhone12的最大競品。由於兩者上市時間僅差一年且功能上並沒有太大的差別，因此容易被拿來比較且iphone12 推出時，iphone11多會降價出售，增加消費者購買意願，導致同品牌的iPhone11成為iPhone12的最大競品，而非其他品牌之手機。

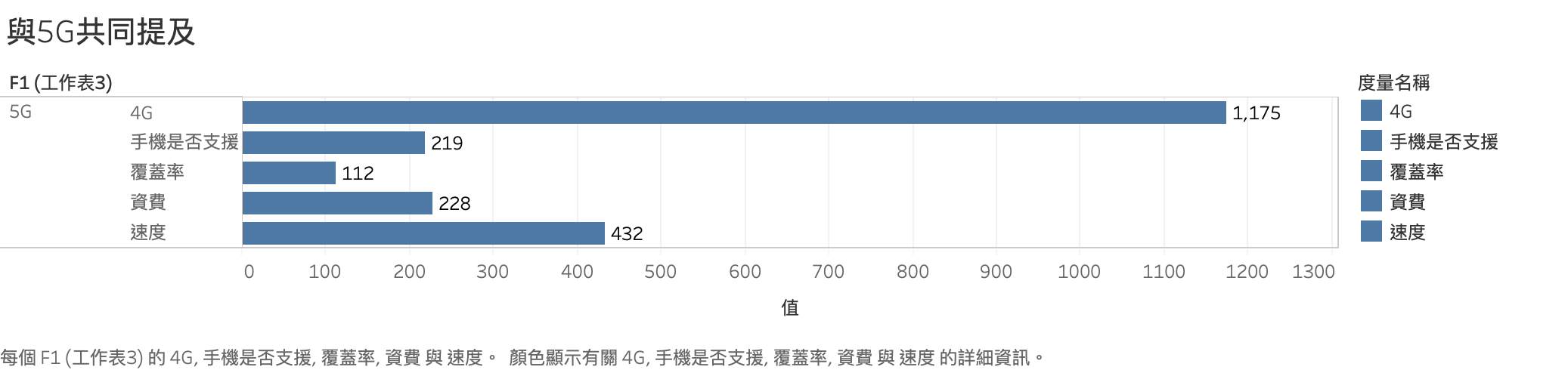


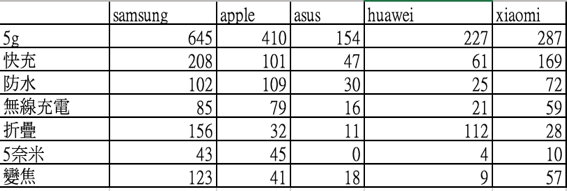
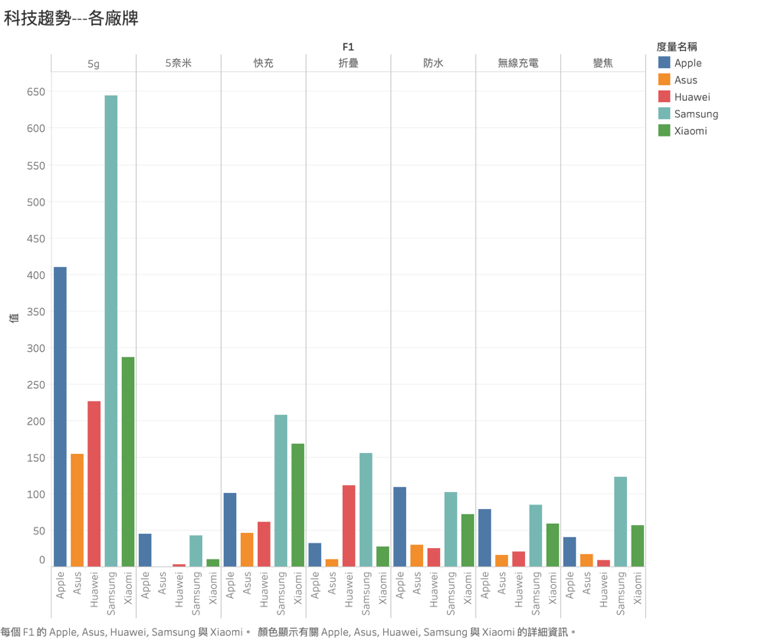
下一個部分我們將探討2020的最新科技趨勢，其中5G為今年大家最關注之科技趨勢，其他依序為快充、防水、無線充電。



最常和5G一同被提及的關鍵字是4G，再來是速度、資費及手機是否支援。

而覆蓋率相較之下較少被拿出來討論。



比較各種技術和各廠牌的討論情形，可以發現samsung在各個項目的討論度皆領先其他品牌；小米的快充相較於其他技術有較高討論度；防水的部分，apple具有最高的討論度；而討論度最高的5G中，討論度依序為samsung、apple、小米、華為、asus。

* 結論

最後總結來看，Iphone12 於看板上有高討論度，且銷售量表現佳。然而各項功能，尤其在cp值、效能上未有突出的表現，續航力甚至是負評居多。其銷售量可能來自於高討論度及高品牌忠誠度。然而，推測非「果粉」對iphone 印象較差，導致品牌形象分數低，建議未來可以延長出新機時間，並且在花心力在效能等功能上有更高的突破；而Samsung s20 在看板上有高討論度，各項功能均正評大於負評，尤其在cp值、效能、續航有不錯的表現。Samsung品牌形象多排名中上。然而 s20 以上優點卻上市時機不佳，沒有反映在銷售量上。在科技趨勢，samsung 在各方面均有領先的技術及討論度，品牌前景看好，期待其產品未來有更多技術突破。