

Youtube Analysis

緒論

當今的數位時代中，YouTube已成為世界上非常重要的影片分享平台之一。每天有數以億計的用戶在此觀看和分享各種內容。這種龐大的使用量和影響力使YouTube成為一個非常有價值的研究對象。因此，本研究的動機在於識別、探索和理解YouTube平台上影片特徵與觀看數之間的關聯性，並挖掘其中潛在的價值。在這個背景下，本研究旨在通過分析YouTube影片的相關特徵和觀看數之間的關係，深入瞭解影片在平台上的表現。透過分析三個子主題，進而探討這些特徵與觀看數之間的關聯性，並尋找其中的趨勢。而這項研究的目的是為了為內容創作者、行銷人員和平台管理者提供實際的洞察和建議。通過深入了解影片特徵和觀看數之間的關係，可以提供指導性的建議，例如如何優化影片特徵以吸引更多觀眾、提高觀看數，或是如何選擇最適合的影片類別和標題等。

相關研究

在台灣，對於YouTube的相關研究主要聚焦於評估YouTube影片所帶來的商業效益，例如廣告效益分析和付費訂閱制度等。這些研究結果一致指出透過YouTube影片進行行銷或廣告是非常有效的策略。然而，較少有研究聚焦於探討影片本身特性與觀看數之間的關聯性。目前，我僅找到了一篇由許蔓萍博士發表的研究論文，題為「利用機器學習使用熱門關鍵字預測YouTube影片觀看次數」。然而，由於該論文尚未開放電子全文，無法查閱詳細內容。根據論文摘要的描述，許博士利用不同的模型來預測YouTube影片的觀看次數，並探討了各種不同層面的影片特徵對觀看數的影響。最終，研究結果顯示隨機森林模型在此預測任務中具有較高的解釋力和預測能力。

對於這些研究，我認為它們提供了對於YouTube影片在商業層面上的價值評估，以及透過YouTube進行行銷和廣告的有效性的證明。然而，對於影片本身特性與觀看數之間的關聯性，目前的研究仍然有限。進一步的研究可以進一步探討影片內容、標題、長度等特徵對於觀看數的影響，以提供更深入的洞察和實際的建議。

方法論

本研究旨在透過分析以及探索Youtube 平台上影片特徵與觀看數之間的關聯性，以下是本次研究的流程及方法：

1. 數據收集 → 透過串接 Youtube API 進行資料收集，包含影片的觀看數、留言數、影片長度以及相關數據等等。以下是本次研究主要使用的 Feature

title	The title of the resource primarily associated with the activity.
description	The description of the resource primarily associated with the activity.
viewCount	The total view count of this video
likeCount	The total like count of the video
commentCount	The total comment count of the video
favoriteCount	The total favorite count of the video

categoryId	The type of the video
duration	The duration of the video

- 問題 1 → 因初學 python，而對 API 串接並不熟悉 → 查看官方文件以及技術部落格
- 問題 2 → API 有每日使用 Quota，超過上限後則需等待一天 → 取得資料後先將資料以 CSV 檔存放以便存取

2. 數據清理

- 因 API 回傳格式為 JSON 檔，因此需先將其攤平。
- 而因缺失值較少約佔整體數據的(1%)，因此直接將帶有缺值的 row 直接去除。
- 將 Category mapping 成實際的 Category Name
- 將不正確的資料型態轉型

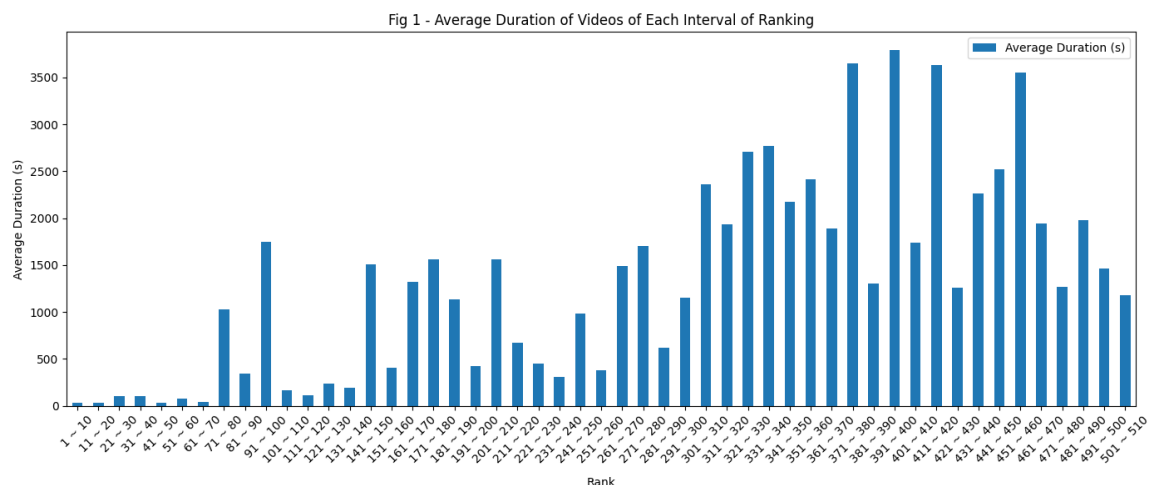
3. 數據分析以及資料視覺化 → 透過對影片的三個特徵，時間長度、影片類別、影片標題分別進行數據分析以及視覺化

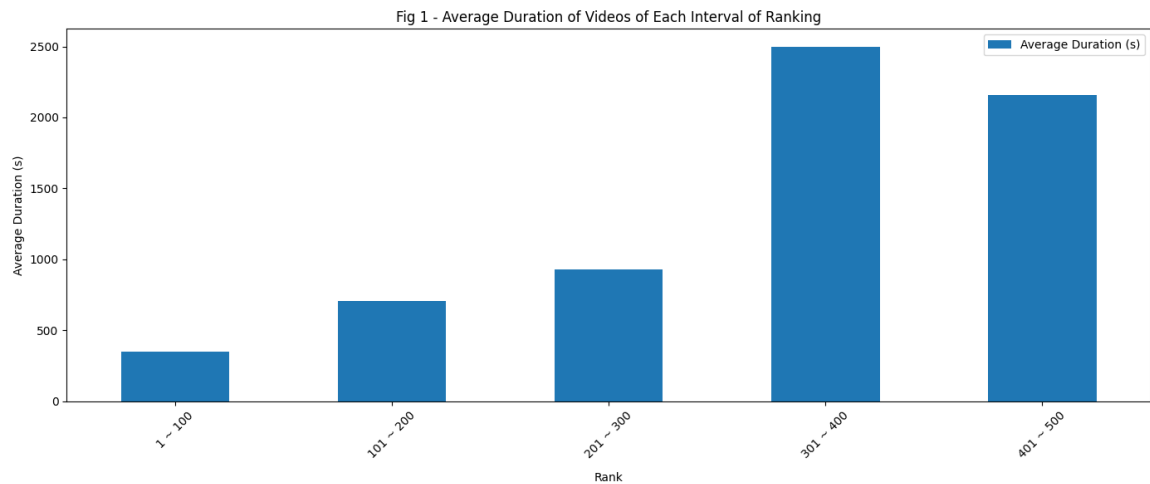
- 問題 1 → 因 Feature 數量相對少，可分析面向較少 → 透過學習使用更多不同套件，對資料進行不同方面的分析

研究結果

• Duration Analysis

- 平均值與標準差：我們觀察到影片的平均時長約為1345秒（約22分鐘），標準差則約為2692。這意味著影片的時長在整體數據中存在一定的變化和差異。
- 關聯性：透過數據觀察到 Duration 與 viewCount 之間存在一定的負相關。具體而言，時間的長度越長，觀看數往往較低。而越短的影片通常有更好的觀看數。
- 回歸分析：透過回歸分析得到了回歸係數約為 -569.96。這表示在此數據中，每增加一秒的時常，觀看次數將減少約 569 次

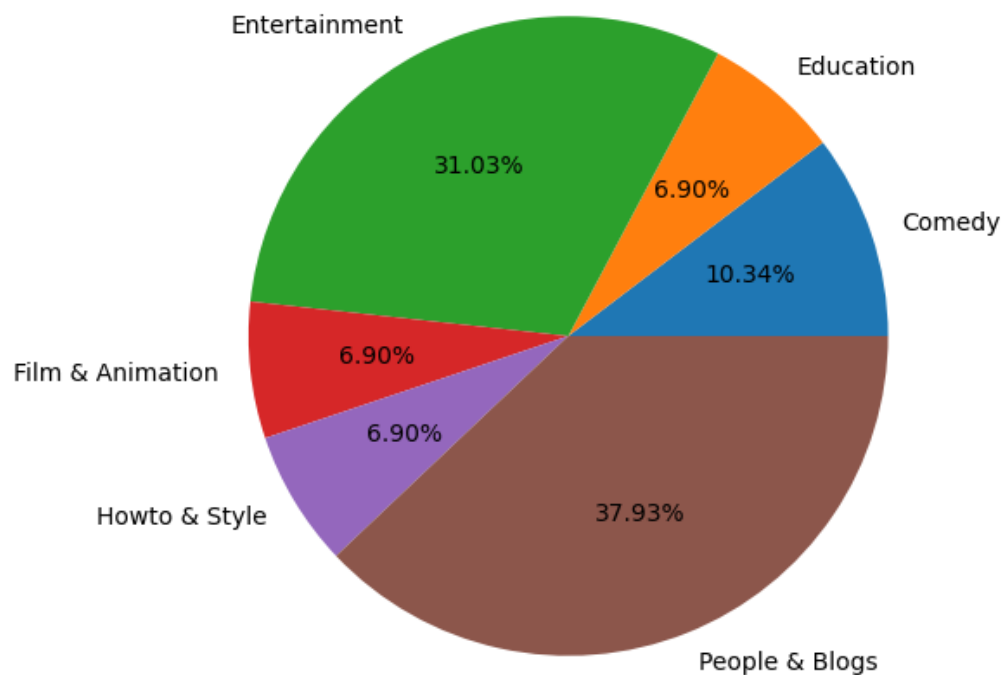




• Category Analysis

- 我們觀察到在此資料集中，觀看數大於100000的影片大多都是 People & Blogs 以及 Entertainment 相關的影片，兩者分別佔了 31.03%, 37.93%。這說明雖然影片類型有極多種，但是在 Youtube 上，仍有幾個影片類別是相對受大眾歡迎的。

Fig 2 - Categories of the Most Watched Videos



- 然而，影片類別是在創作者上傳時自行選取，因此可能會看到內容與類別毫無相關的相關影片，如以下在 Education 類別中的影片：

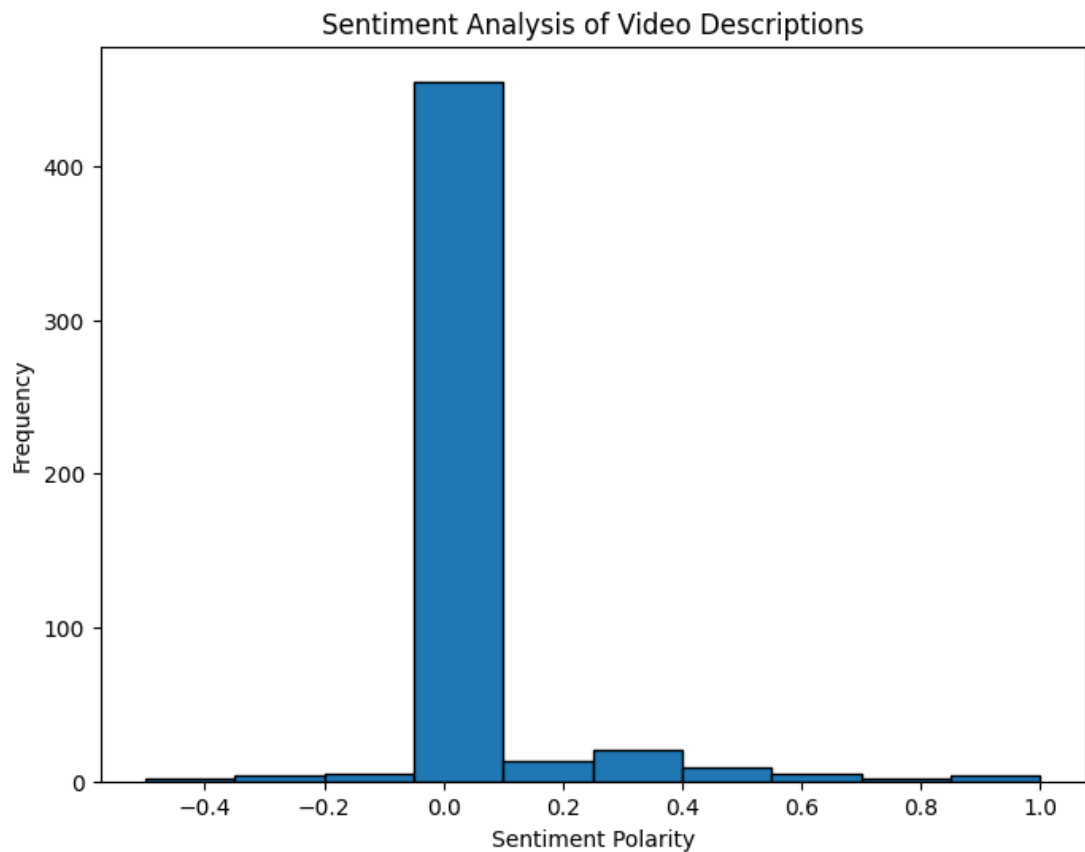
- Top videos in the Education category:

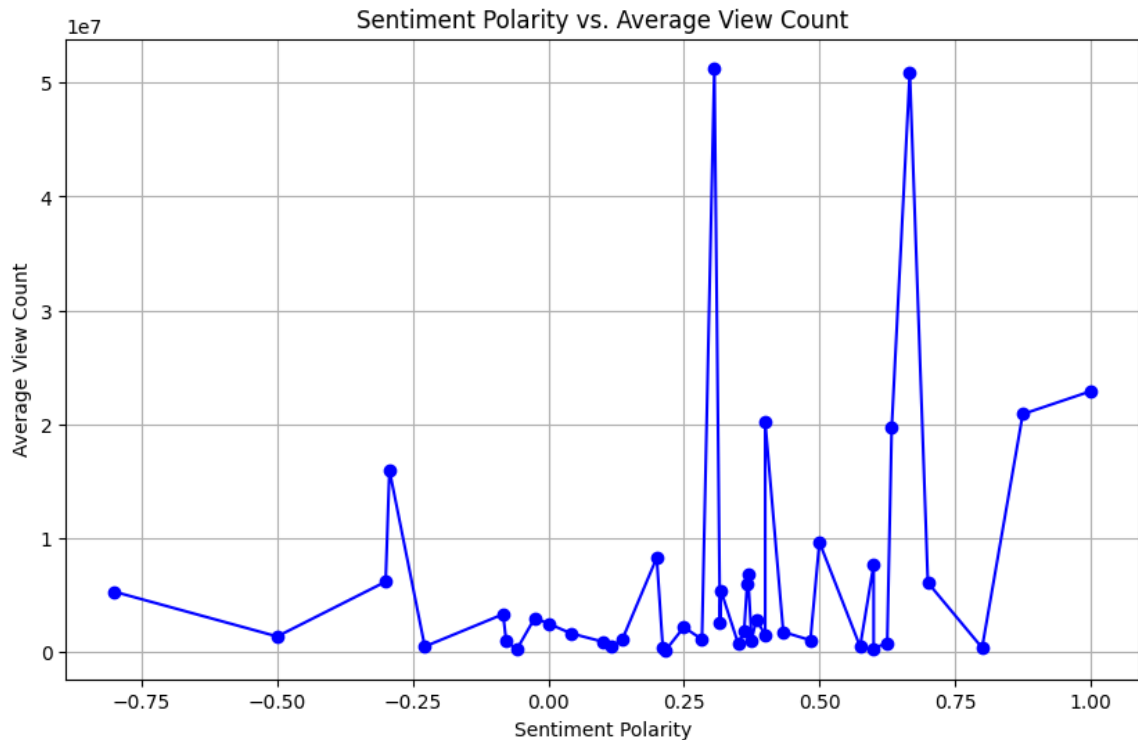


title	duration	viewCount	likeCount	commentCount
ऐसे Coincidence जो लाखों करोड़ों में एक बार होत...	35.0	16363062	952593	358
couples goals 🥰 #cutebaby #cute #cutebabystatus...	10.0	12541173	313425	568
2023 高露潔全效 抗牙菌斑牙膏 30s	31.0	1227533	7	1
10個北韓最可怕的 變態規定！觸犯一 條比死刑更慘，第 一名是所有女兵的 噩夢！ #世界之最 top ...	719.0	756378	5833	679

• Sentiment Analysis

- 平均值與標準差：我們觀察到影片的標題情感 (-1 為最負面、1 為最正面) 平均約為 0.04，說明此數據中大部分標題情感皆為中性的。
- 關聯性：標題情感越接近中性（接近0），觀看數通常也越接近平均水平。然而在一些特殊情況下，極度負面（-0.8）或極度正面（1.0）的標題情感，這些情感極端的標題可能在觀看數方面表現出較高的平均水平。





研究討論 / 結論

• 結論

1. 影片時長對於觀看數有一定的影響。研究觀察到，**較短**的影片通常具有較高的觀看數，而**長時間**的影片則相對觀看數較低。每增加一秒的影片時長，觀看次數平均減少約569次。
2. 在不同的影片類別中，**People & Blogs** 和 **Entertainment** 相關的影片受到觀眾的歡迎，佔據了觀看數超過10萬的影片中的大部分份額。然而，也存在一些與類別不相關的影片在特定類別中出現。
3. 影片標題情感對觀看數也有一定的影響。大部分影片的標題情感都是中性的，且觀看數也皆趨向於平均。然而，在一些特殊情況下，**極度負面或極度正面**的標題情感可能導致觀看數較高。

根據研究結果，短時間的影片、People & Blogs 和 Entertainment 類別的內容以及中性情感的標題可能有助於提高觀看數。然而，這些結果僅是初步的觀察，仍需要進一步的研究來探索更多因素對於 YouTube 影片表現的影響。

• 研究討論

- 此研究藉由 Youtube 資料集來探討觀看數與影片三項特徵的關係來進一步討論 Youtube 創作者成功的因素。透過這個分析我們可以從不同面向、因素去揭露到底什麼才是讓影片變得受歡迎的關鍵。而這可能有助於對內容創作者、營運人員更甚是廣告公司去更好地理解以及預測影片的表現，從而制定更有效的製作或是廣告策略。

- 此外，這項研究的結果也可以用在其他相關領域的研究，例如社交媒體行銷、用戶行為分析等等。這些領域也存在類似的核心問題：「如何最大程度的吸引觀眾的注意力並提升內容的影響力」

- **方法限制、未來研究方向**

- 這項研究有一些限制需要注意。
 - 首先，此研究僅考慮了影片時長、標題情感和觀看次數這三個因素，然而還有其他許多因素（例如內容品質、標籤、上傳時間等）可能對影片的表現產生影響。因此，未來的研究可以擴展分析的範圍，考慮更多因素的影響，以獲得更全面的結果。
 - 其次，這項研究主要基於統計分析，並未深入探討背後的原因和機制。未來的研究可以進一步進行質性研究，例如訪談內容創作者和觀眾，以探討他們對於持續時間、標題情感和觀看次數之間關係的感知和理解。
- 而未來的研究則可以擴大樣本規模和觀察時間範圍，包括更多不同類型的影片和更長的觀察期，或是可以透過建立相關模型來預測影片觀看數，以獲得更具代表性和穩定的結果。

總結而言，這項研究提供了一個初步的框架，用於理解YouTube影片的表現和觀眾行為，對於內容創作者和行銷人員制定策略具有實際價值。然而，仍有許多未來的研究方向可以進一步擴展和深入探索，以提供更具體且實用的結果。