

Netflix EDA

程式設計與資料科學導論 期末專案

外文三 B10102102 陳品澄

外文三 B10102092 林貞好

1. 主題和簡介

本報告旨在透過對 Netflix 的數據分析，探究其內容與市場趨勢。動機源於希望研究與日常生活緊密相連的主題。Netflix 作為全球領先的串流媒體服務供應商，其豐富的內容和獨特的經營策略對當代文化產生了深遠影響。我們將分析 Netflix 的內容類型、發行時間、評級和國家製作等方面的趨勢，旨在提供對 Netflix 更全面的了解。

2. 資料來源

為進行這項研究，我們從兩個主要來源收集了數據：Kaggle 和維基百科。Kaggle 提供了詳細的 Netflix 數據集，也是我們本次報告的核心資料來源。而維基百科則為我們提供了有關 Netflix 的歷史背景和重要事件等資料，這兩種數據的結合，使我們能夠從更多角度分析 Netflix 的發展脈絡。

3. 資料視覺化

在本報告中，我們對收集到的數據進行了視覺化處理，以便更直觀地展示 Netflix 的數據特徵。包括 Netflix 內容的種類、發行時間、評級和製作國家等關鍵指標。這種視覺化的方法不僅使數據更易於理解，也有助於揭示數據中的潛

在趨勢。以下將展示四個基本資料最為範例。

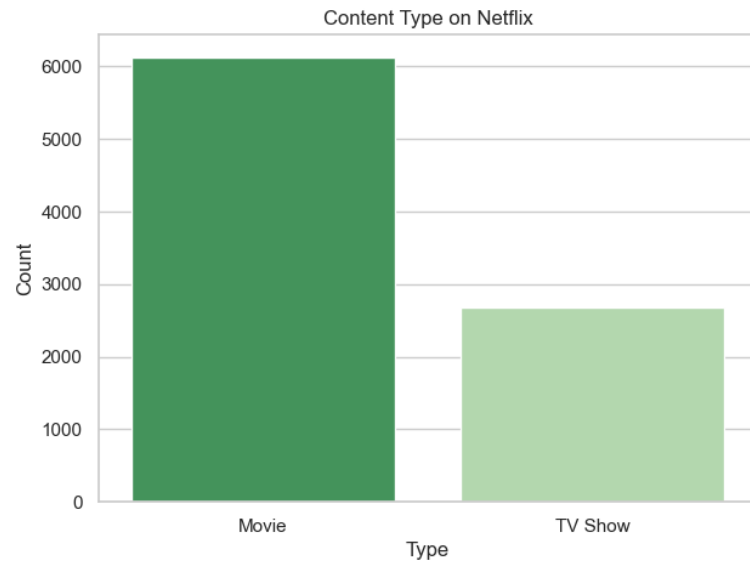


圖 1: 類型分佈

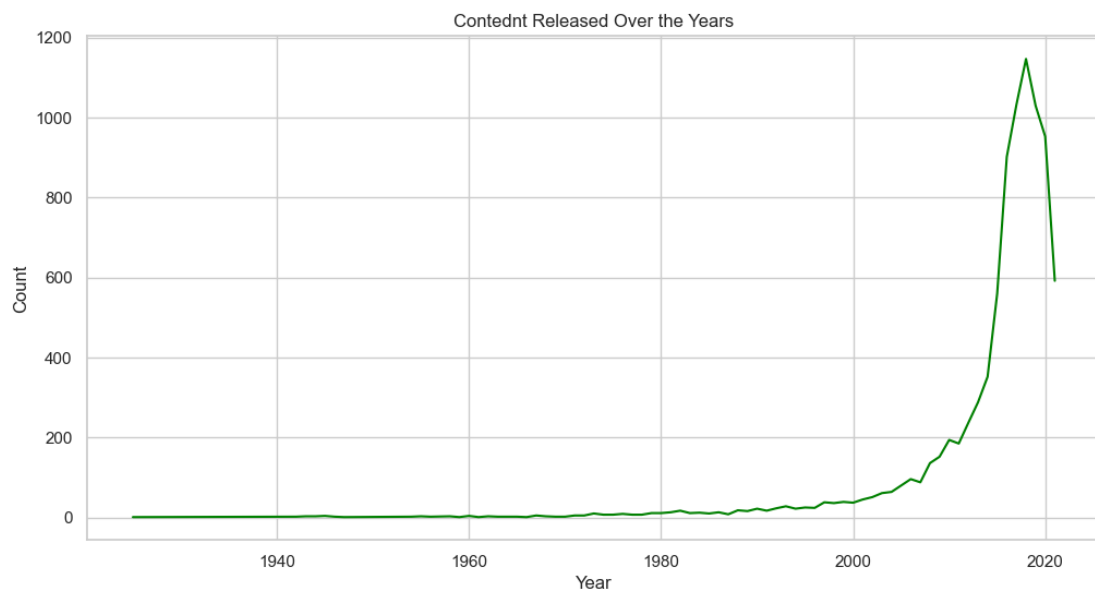


圖 2: 發行時間分佈

在圖 1 中，綠色長條圖顯示 Netflix 上電影數量約為 6000 部，而電視目約 2000 部，表明電影是平台上較為主流的內容。圖 2 展示了從 1940 年到 2020 年 Netflix 發布內容的趨勢。可以看到，直到 2000 年初，每年新增的內容量相對較少。但從 2010 年開始，新增內容量急遽增長，尤其是在 2019 年到 2020 年間達

到高峰。不過 2020 開始卻呈現下滑，我們認為是因為疫情影響，因為民眾無法前往電影院，導致片商製作意願下降。

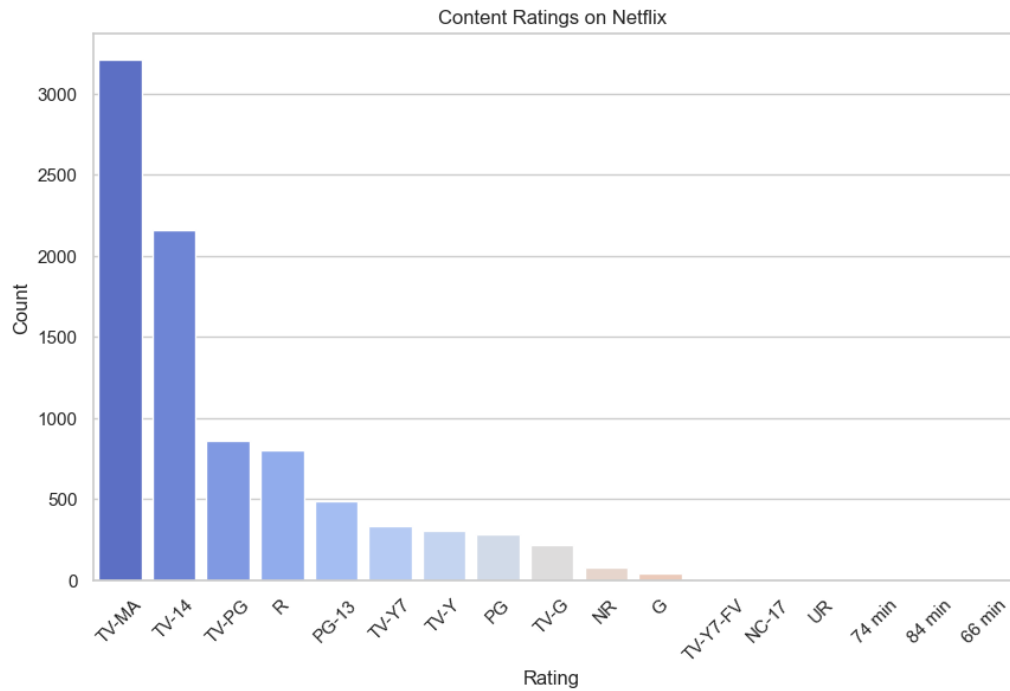


圖 3: 不同評級的節目數量

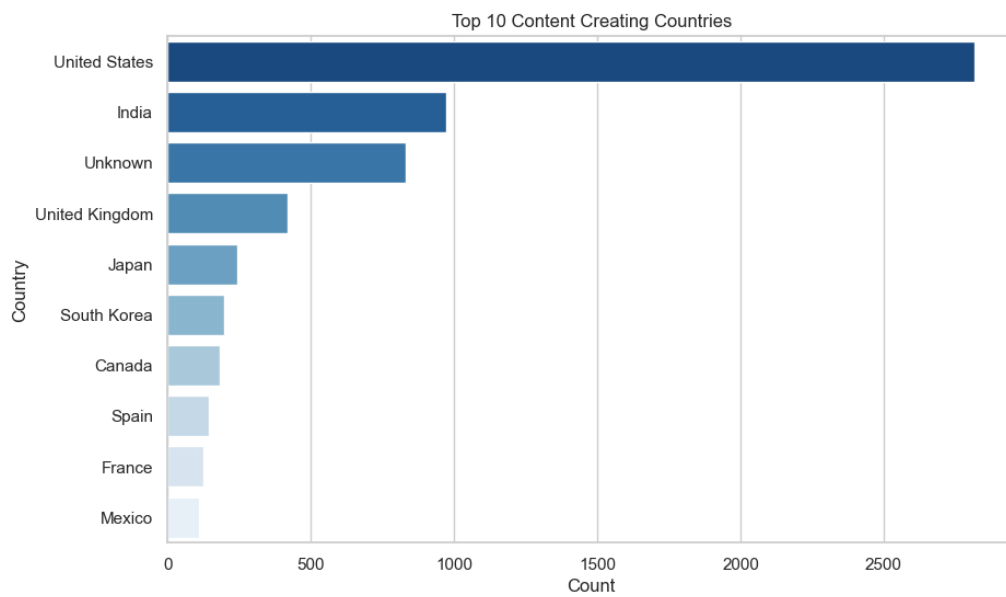


圖 4: 國家分佈

圖 3 的柱狀圖展示了 Netflix 各種評級內容的數量分佈，其中 TV-MA 和 TV-14 評級的內容最多，顯示 Netflix 主要服務對象為成年及成熟青少年觀眾。

圖 4 揭示 Netflix 內容的製作國家分佈，美國製作的內容最多，其次是印度。英國、日本和韓國等國家也位列前十，凸顯了它們在全球娛樂市場中的重要地位。

4. 假說設計

為豐富本報告的內容，我們根據日常生活中的觀察設計了三個假說，目的在於探究 Netflix 的內容和趨勢。接下來，我們將逐一介紹這些假說及其驗證過程。

4-1: 一年之中在 7-8 月以及 11-12 月這兩段時間的電影發布數量是否比較多？

這個想法來自於我們常聽到「暑期強檔」和「年末強檔」的廣告宣傳。為了驗證這一假說，我們將把 Netflix 電影的發布日期數據製作成圖表，以直觀地觀察在這兩個特定時段是否的確有更多影片發布。

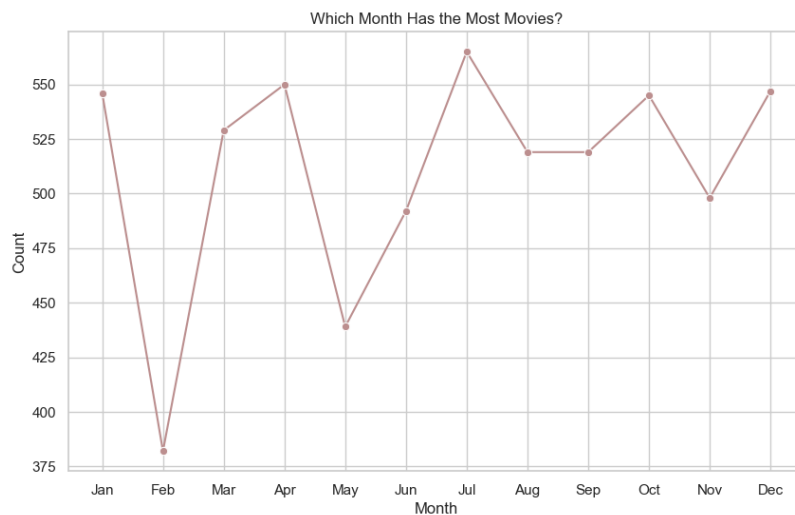


圖 5: 月份統計電影數量

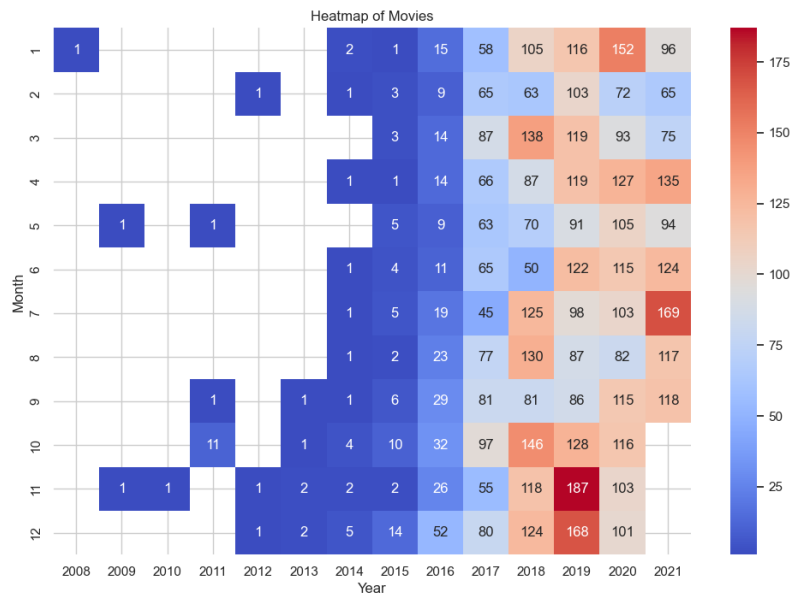


圖 6: 月份統計電影數量

圖 5 的折線圖以月份為 x 軸，數量為 y 軸，顯示了各月份的發行數量。可見在 7 月份，其數量達到了明顯的高峰。圖 6 的熱力圖則以年份為 x 軸，月份為 y 軸，數字與顏色反映了相應位置的發行數量，色彩越接近深紅表示數量越多，我們觀察到 7 月和 12 月都呈現出明顯的高數量區域。因此，我們的初步假說得到確認，即在 7 月和年末時段，影片發行量確實不少於其他月份。

4-2: 電影的平均時長是否隨年代增加而越來越長？

自從幾年前的漫威電影--終局之戰出現後，我們發現有許多電影時常似乎越來越長，例如蝙蝠俠、跟最近很紅的奧本海默都是超過兩小時的長片，所以我們猜測電影在近年來有越做越長的趨勢。以下我們會先以兩張圖來說明電影及電視劇的平均長度。

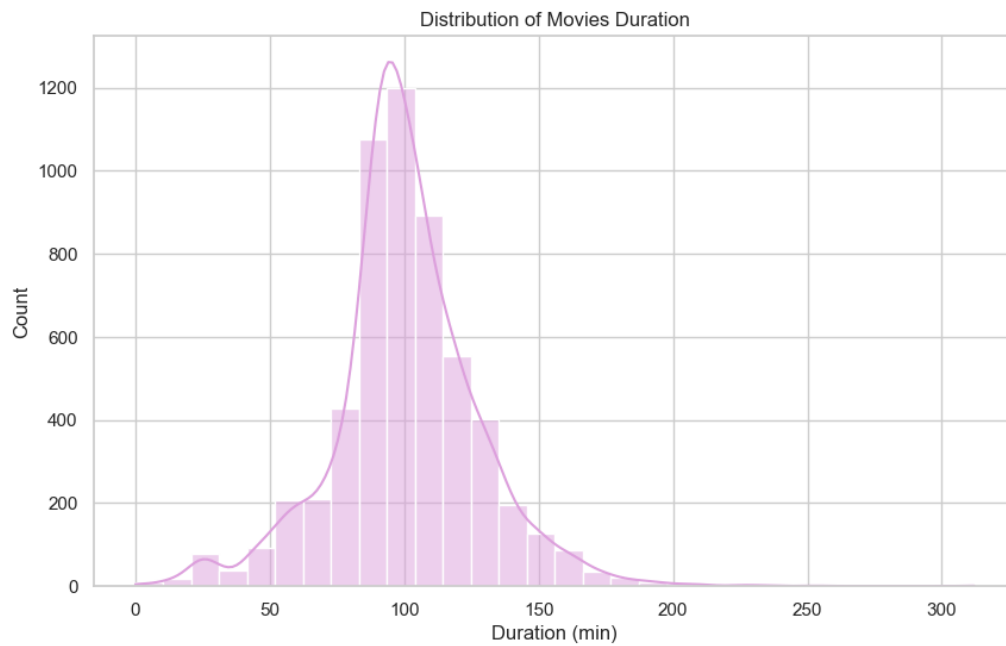


圖 7: 電影時長分佈圖

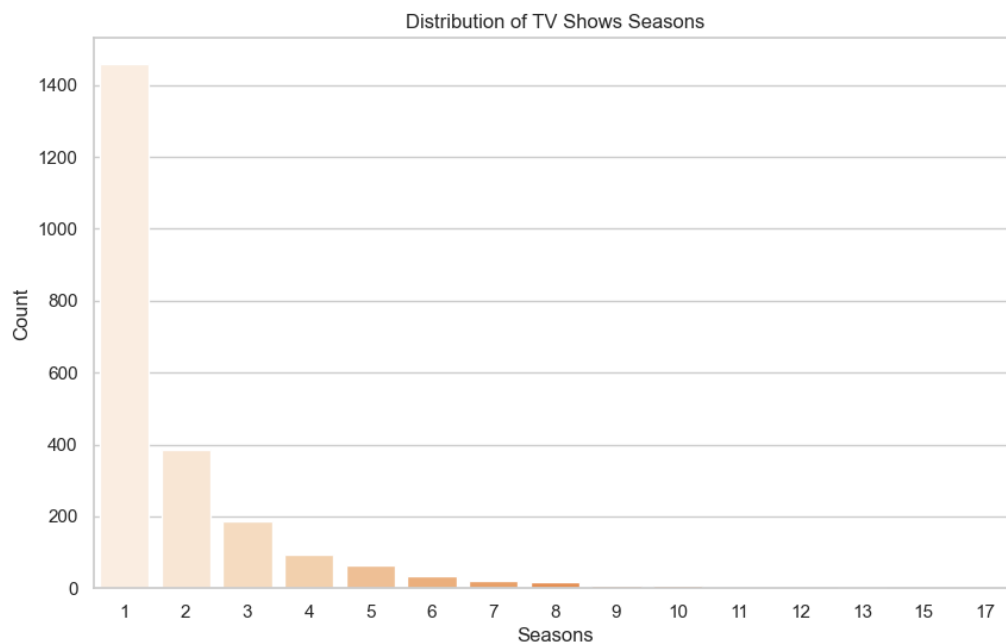


圖 8: 電視劇季數分佈圖

首先，圖 7 是電影的平均時長分佈圖。該圖呈現大致的常態分佈，其中以約 100 分鐘的時長為峰值，表示有最多數量的電影接近這個時長。隨著時長向

兩側遞減，數量逐漸減少，形成了一個相對均衡的分佈。其次，圖 8 是電視劇的平均季數分佈圖，該圖顯示了一季的電視劇佔據了最多的數量，而隨著季數增加，圖表向右遞減，表明有更多季度的電視劇數量逐漸減少，可以觀察到出現第三季的電視劇數量已經相對較少。以下將呈現每個年份的電影平均時長趨勢圖來驗證假說。

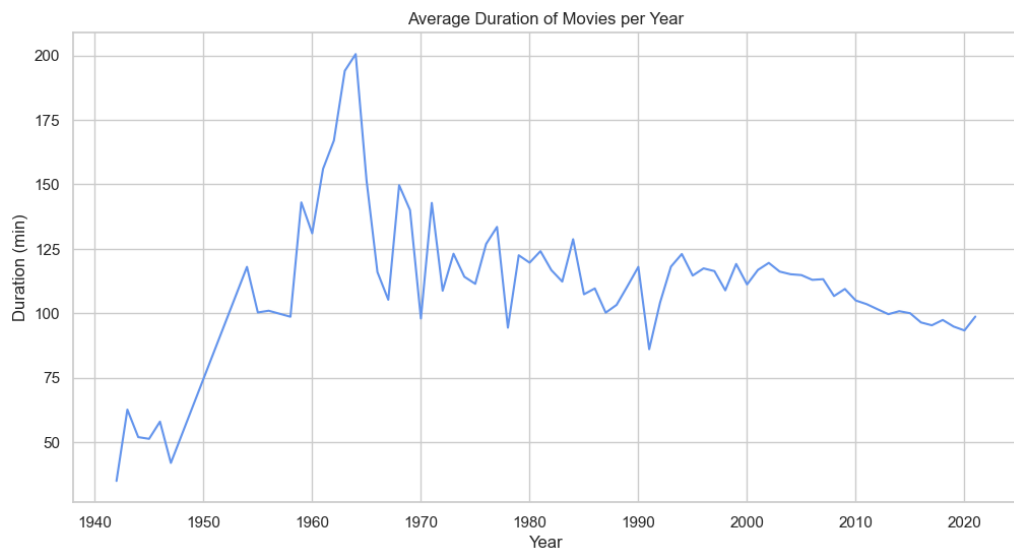


圖 9: 每個年份的電影平均時長

圖 9 呈現每個年份的電影平均時長趨勢的圖表，其中 x 軸表示年份，y 軸表示時長。值得注意的是，在 1960 年到 1965 年之間，平均時長達到了一個高峰，而在此後的年份，時長呈現較大的波動。特別是自 2000 年以來，我們還觀察到整體時長呈現下降的趨勢。基於以上的數據趨勢，得出第二個假說的結論，即電影的平均時長並未隨著年代的增加而逐漸延長，與最初的預期有所不同。

此外，我們還補充了一張每個電影類型的平均時長圖表（圖 10），以提供更豐富的資訊。

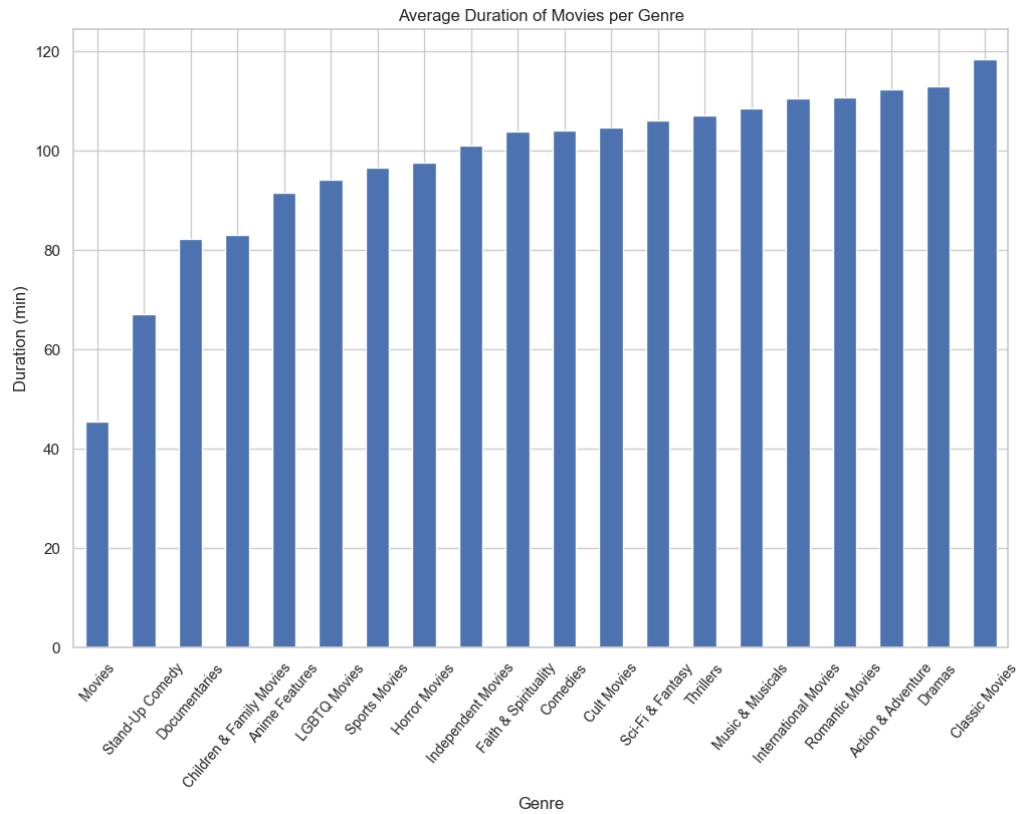


圖 10: 每個電影類型的平均時長

4-3: 國家出產影片數量的趨勢，尤其以出產影片量前五名的國家為主，是否有符合總數量的趨勢？

Netflix 在 2016 年經歷了顯著的影片數量成長。基於這個觀察，我們想了解這些國家的影片生產數量是否與整體趨勢相符。

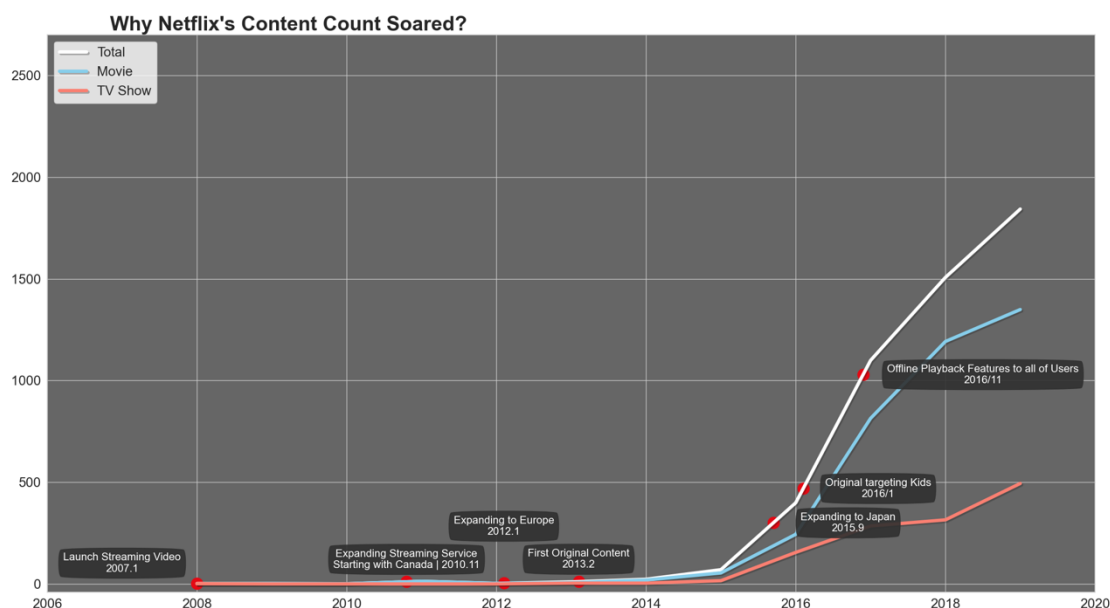


圖 11: Netflix 重要記事表

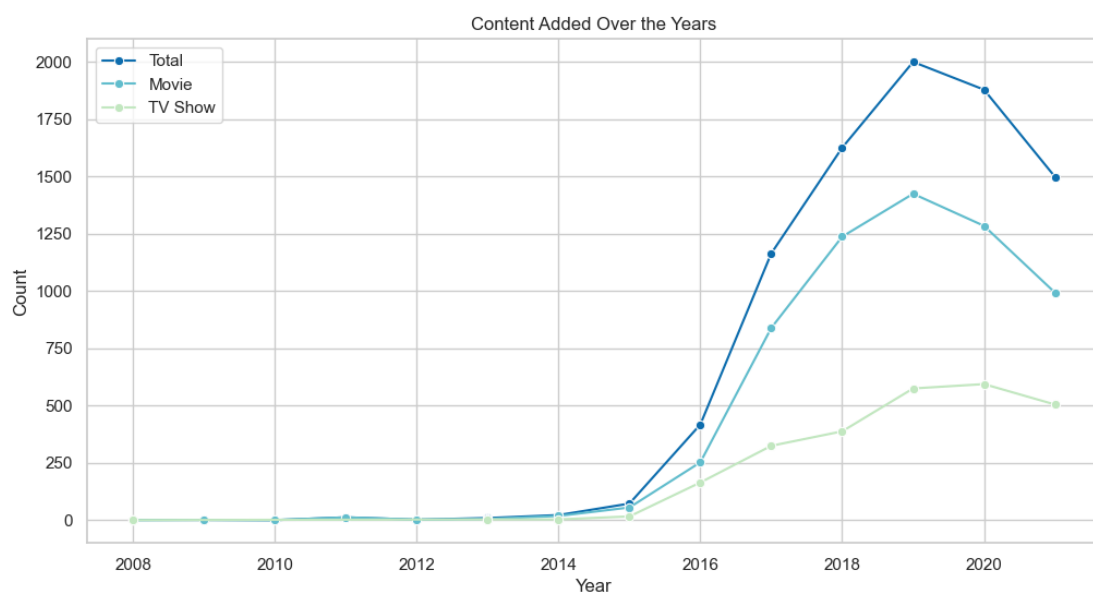


圖 12: 每年新增的總數、電影數和電視節目數

圖 11 展示了 Netflix 的重要時間線，包括 2007 年推出串流影音服務、2012 年擴展至歐洲市場，以及 2016 年推出離線播放功能等重要里程碑。從該圖的折線趨勢中，我們觀察到 2016 年左右 Netflix 的成長呈現指數型增長。而圖 12 則

顯示了每年新增的總數、電影數和電視節目數。這張圖延續了重要記事的數據，但特別指出 2019 年後數量出現下降趨勢，我們認為這可能與新冠疫情的影響有關。

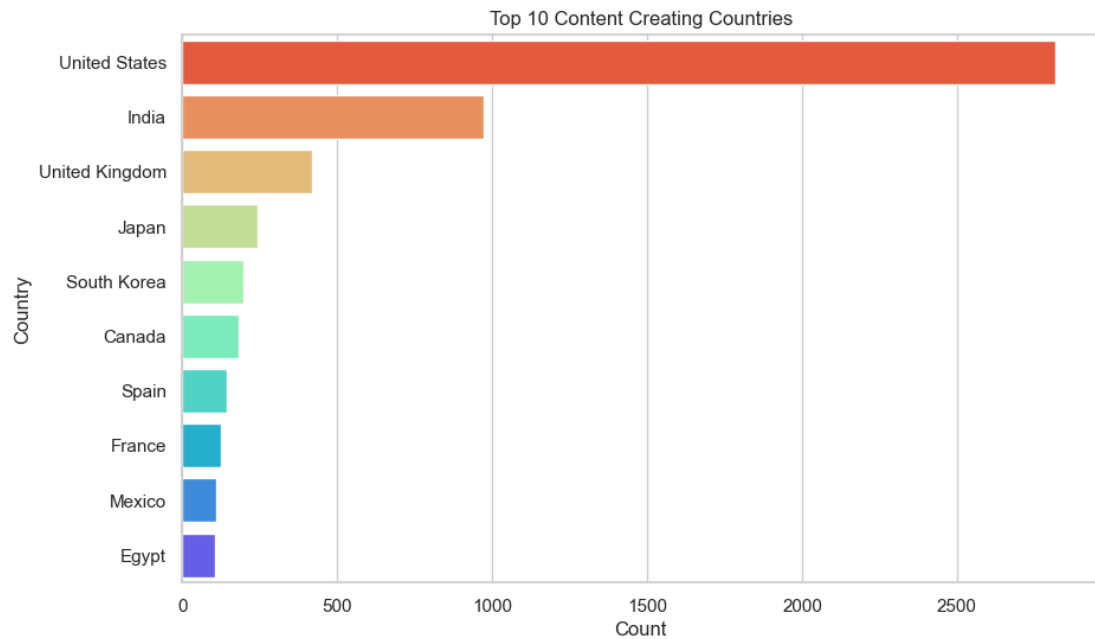


圖 13: 去除 unknown 後各個國家製作的節目數量

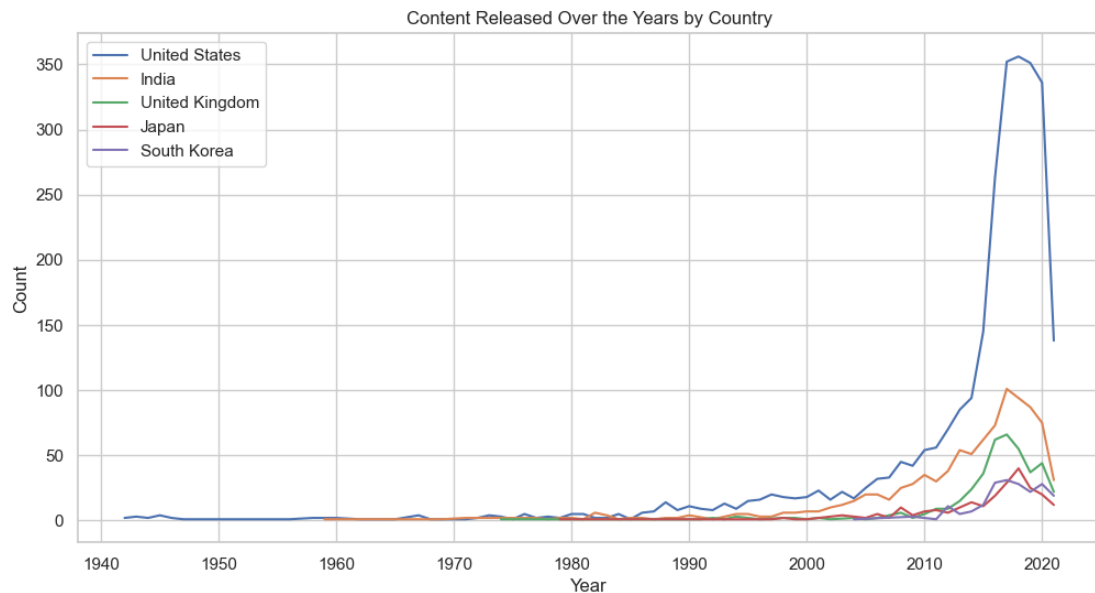


圖 14: 出產影片量前五名國家節目量

從資料中整理出圖 13 展示了各國製作的節目數量（排除未知國家），顯示出美國、印度、英國、日本和南韓是前五大製作國。而圖 14 則在同一折線圖上比較了這五個國家的節目生產趨勢。從圖中可以看出，從 2014 年開始，這些國家的節目數量均出現了顯著增長。特別是美國從 2014 年開始增長迅速，印度和英國的成長趨勢也顯著，而日本和南韓自 2014 年起也呈現明顯增長。這些數據表明，這五個國家在約 2016 年左右的節目產量都有顯著上升。透過這張圖的趨勢成功驗證了我們的假說。

5. 總結

在這次的 Netflix EDA 報告中，我們透過對 Netflix 數據的深入分析，探討其內容和市場趨勢。報告整體架構涵蓋主題和簡介、資料來源、資料視覺化、假說設計等多個面向。以下是報告的主要總結：

- a) 在 7-8 月以及 11-12 月這兩段時間的影片發布數量比較多。
- b) 電影的平均時長沒有隨年代增加而越來越長。
- c) 前五大影片出產國，出產的影片數量趨勢有符合總數量的趨勢。

分工表

	陳品澄	林貞妤
程式碼	60%	40%
投影片製作	40%	60%
書面報告	60%	40%
口頭報告	50%	50%