



Data Mining Final Project

-Association between Genres and Revenue of Movies-

A1093356 黃冠儒

CATALOG

1.

資料來源與介紹

2.

背景與動機

3.

預算與票房關係

4.

種類與票房關聯

5.

結論



資料來源與介紹

1

IMDB Movies Dataset from Kaggle (10088 obs.)

| 1 | names | date_x | score | genre | overview |
|----|------------------------------|------------|-------|---|---|
| 2 | Creed III | 2023/3/2 | 73 | Drama,Action | After dominating the boxing world, Adonis Creed has been thriving in b |
| 3 | Avatar: The Way of Water | 2022/12/15 | 78 | Science Fiction,Adventure,Action | Set more than a decade after the events of the first film, learn the story of |
| 4 | The Super Mario Bros. Movie | 2023/4/5 | 76 | Animation,Adventure,Family,Fantasy,Comedy | While working underground to fix a water main, Brooklyn plumbers—a |
| 5 | Mummies | 2023/1/5 | 70 | Animation,Comedy,Family,Adventure,Fantasy | Through a series of unfortunate events, three mummies end up in presen |
| 6 | Supercell | 2023/3/17 | 61 | Action | Good-hearted teenager William always lived in hope of following in his |
| 7 | Cocaine Bear | 2023/2/23 | 66 | Thriller,Comedy,Crime | Inspired by a true story, an oddball group of cops, criminals, tourists and |
| 8 | John Wick: Chapter 4 | 2023/3/23 | 80 | Action,Thriller,Crime | With the price on his head ever increasing, John Wick uncovers a path to |
| 9 | Puss in Boots: The Last Wish | 2022/12/26 | 83 | Animation,Family,Fantasy,Adventure,Comedy | Puss in Boots discovers that his passion for adventure has taken its toll: P |
| 10 | Attack on Titan | 2022/9/30 | 59 | Action,Science Fiction | As viable water is depleted on Earth, a mission is sent to Saturn's moon T |

| names | date_x | score | genre | overview |
|-------|--------|--------|-------|----------|
| 電影名稱 | 上映日期 | IMDB評分 | 分類 | 劇情簡述 |

(<https://www.kaggle.com/datasets/ashpalsingh1525/imdb-movies-dataset>)

IMDB Movies Dataset from Kaggle (10088 obs.)

| 1 | crew | orig_title | status | orig_lang | budget_x | revenue | country |
|----|---|-----------------------------|----------|--------------------|-----------|-------------|---------|
| 2 | Michael B. Jordan, Adonis Creed, Tessa Thompson, Bianca Taylor, Jonathan Majors | Creed | Released | English | 75000000 | 271616668 | AU |
| 3 | Sam Worthington, Jake Sully, Zoe Saldana, Neytiri, Sigourney Weaver, Kiri / Dr. C | Avatar: The Way of Water | Released | English | 460000000 | 2316794914 | AU |
| 4 | Chris Pratt, Mario (voice), Anya Taylor-Joy, Princess Peach (voice), Charlie Day, I | The Super Mario Bros. Movie | Released | English | 100000000 | 724459031 | AU |
| 5 | Óscar Barberán, Thut (voice), Ana Esther Alborg, Nefer (voice), Luis Pérez Reina, Momias | Momias | Released | Spanish, Castilian | 12300000 | 34200000 | AU |
| 6 | Skeet Ulrich, Roy Cameron, Anne Heche, Dr Quinn Brody, Daniel Diemer, Williar Supercell | Dr. Quinn, Medicine Woman | Released | English | 77000000 | 340941958.6 | US |
| 7 | Keri Russell, Sari, Alden Ehrenreich, Eddie, O'Shea Jackson Jr., Daveed, Ray Liott | Cocaine Bear | Released | English | 35000000 | 80000000 | AU |
| 8 | Keanu Reeves, John Wick, Donnie Yen, Caine, Bill Skarsgård, Marquis de Gramot | John Wick: Chapter 4 | Released | English | 100000000 | 351349364 | AU |
| 9 | Antonio Banderas, Puss in Boots (voice), Salma Hayek, Kitty Softpaws (voice), Ha Puss in Boots: The Last Wish | The Last Wish | Released | English | 90000000 | 483480577 | AU |
| 10 | Paul Bianchi, Computer (voice), Erin Coker, Allison Quince, Jack Pearson, Max R | Attack on Titan | Released | English | 71000000 | 254946484.2 | US |

| crew | orig_title | status | orig_lang | budget_x | revenue | country |
|------|------------|--------|-----------|----------|---------|---------|
| 演員 | 原著名字 | 是否上映 | 原著語言 | 製片預算 | 票房 | 國家 |

Data Preparation

| 1 | names | date_x | score | genre | overview |
|------|--|------------|-------|-------|--|
| 1563 | Hen | 2023/4/9 | 0 | | When a non-binary Gem attends their cis mate's hen's night, the girls... |
| 1764 | NJPW Sakura Genesis 2023 | 2023/4/8 | 0 | | Professional wrestling event promoted by New Japan Pro-Wrestling (N... |
| 1778 | この小さな手 | 2023/4/8 | 0 | | Based on the manga of the same name by Mamora Goda, a manga artis... |
| 2022 | Yung Libro sa Napanood Ko | 2023/4/8 | 0 | | We don't have an overview translated in English. Help us expand our d... |
| 2146 | ACA 159: Karginov vs. Silva | 2023/4/8 | 0 | | ACA 159: Karginov vs. Silva was a mixed martial arts event that took p... |
| 2269 | Onimanz | 2023/4/8 | 0 | | a spin-off to 2015 ninja action adventure, The Great War of Torakage... |
| 2319 | My Beautiful Man ~eternal~ | 2023/4/7 | 0 | | We don't have an overview translated in English. Help us expand our d... |
| 2433 | From the End of the World | 2023/4/7 | 0 | | Hana is a normal high school student, unsatisfied in her daily life. One... |
| 2553 | The Ballad of Alan Batman | 2023/4/7 | 0 | | A burnt-out sound mixer has spent his entire life listening to others. Bu... |
| 2579 | Circuit | 2023/4/7 | 0 | | Upcoming Marathi Movie |
| 2875 | Calcutta '99 | 2023/4/7 | 0 | | A neo-noir by Jayabrata Das. |
| 3029 | ARES 14 | 2023/4/7 | 0 | | ARES 14 was a mixed martial arts event that took place on Friday, Apr... |
| 3393 | Yeran | 2023/4/7 | 0 | | We don't have an overview translated in English. Help us expand our d... |
| 3723 | Las Vegas | 2023/4/7 | 0 | | Tristan is into fashion design while his parents take care of the rest. Sur... |
| 3789 | Tulliana | 2023/4/7 | 0 | | Two zoologists seeking the extinct Anatolian panther search for its trac... |
| 3858 | Excuse Me | 2023/4/7 | 0 | | A therapist deals with an unruly patient as they navigate the inner work... |
| 3880 | Corona Papers | 2023/4/6 | 0 | | We don't have an overview translated in English. Help us expand our d... |
| 3908 | Hamlet: Bristol Old Vic Live | 2023/4/6 | 0 | | 'Must I remember?' Hamlet's father is dead. His mother has remarried. I... |
| 4110 | De man uit Rome | 2023/4/6 | 0 | | Follows Filippo, a skeptical Catholic priest who becomes attracted and... |
| 4160 | มาดามเตน | 2023/4/6 | 0 | | Plot Unknown |
| 4251 | SECTION 306 IPC | 2023/4/6 | 0 | | We don't have an overview translated in English. Help us expand our d... |
| 4310 | Fanatic | 2023/4/6 | 0 | | Charlie and Gerald reunite their failed 2000's boy band for a competit... |
| 4331 | Euphoria | 2022/11/29 | 0 | | Artist and filmmaker Julian Rosefeldt creates elaborately staged films f... |
| 4347 | Reclaim | 2022/7/29 | 20 | | She is a good woman living a fulfilling life. Or so it seems. A caring m... |
| 4422 | Avengers: Quantum Encounter | 2022/7/14 | 50 | | We don't have an overview translated in English. Help us expand our d... |
| 4428 | Gabriel's Inferno: Part IV | 2022/3/30 | 50 | | We don't have an overview translated in English. Help us expand our d... |
| 4432 | The Sniper | 2021/11/19 | 40 | | The Salvation Society is the most powerful criminal organization in Sov... |
| 4454 | Phantastische Harry Potter Momente - Das große SAT.1 Spezial | 2021/10/31 | 62 | | We don't have an overview translated in English. Help us expand our d... |
| 4494 | The Sword | 2021/10/18 | 53 | | Lord of Darkness Artha was born from human struggle and resentment |
| 4530 | The Legend of Zhao Yun | 2021/1/4 | 20 | | When political tactics and ambitions dominated, Zhao Yun stepped for... |
| 4537 | 丈夫去上班的日子 | 2019/12/18 | 0 | | We don't have an overview translated in English. Help us expand our d... |

Genres 的欄位有85筆空值(剛上映、地區的小眾電影為主)

Data Preparation

| names | date_x | score | genre | overview |
|--|------------|-------|-------|---|
| 5028 DPMI-019 キャンギャル狂想曲 美月恋 | 2017/6/25 | 0 | | We don't have an overview translated in English. Help us expand our |
| 5423 Oro y Polvo | 2016/4/19 | 68 | | We don't have an overview translated in English. Help us expand our |
| 5639 Female War: Doggie's Uprising | 2016/3/24 | 0 | | There is a visitor for local herb gatherer Doggie who has no particula |
| 5687 Dream to be a Wife | 2015/9/9 | 0 | | We don't have an overview translated in English. Help us expand our |
| 5757 Housewife Sex Slaves: Hatano Yui | 2015/1/9 | 0 | | We don't have an overview translated in English. Help us expand our |
| 5923 Lu-To | 2013/11/15 | 53 | | Luisa and Tomas try to save their relationship by avoiding their differ |
| 5927 Inko fuchi: Hentai useri shiri | 2013/6/11 | 0 | | Honoka, a daughter who takes care of her self-proclaimed depressed |
| 6159 Super Mario Bros. Mix | 2009/9/8 | 0 | | Mario Bros. Mix is a DVD with episodes from The Super Mario Bros. |
| 6186 Secrets of a Desperate Housewife | 2008/1/1 | 85 | | When she learns that her husband's company is in danger of going un |
| 6216 Halo 3 Essentials | 2007/9/25 | 0 | | Completely remastered, high resolution versions of the Halo and Halo |
| 6460 Moses, Prince of Egypt | 2000/1/1 | 56 | | At birth, Moses, a Hebrew baby is abandoned on the Nile by his moth |
| 6560 High School Girl Story Indecent Fruit | 1997/10/10 | 68 | | We don't have an overview translated in English. Help us expand our |
| 6597 Tokyooshi Nikki 3 Himerareta Sei | 1997/6/11 | 30 | | We don't have an overview translated in English. Help us expand our |
| 6646 Tokyooshi Nikki 2 Abakareta Sei | 1996/6/13 | 0 | | We don't have an overview translated in English. Help us expand our |
| 6674 Musume Bugyo On-na Ro Hisho | 1995/3/24 | 25 | | A woman calling herself Princess Oryu enters the woman's prison. Sh |
| 6948 The Wicked Reporter ♀: The Rebirth of Horserace Betting | 1994/11/19 | 90 | | The continuing adventures of Matsuzaki (Eij Okuda), a racing tipster |
| 7046 権道記者? 馬券転生篇 | 1994/11/19 | 0 | | We don't have an overview translated in English. Help us expand our |
| 7422 The Shoga (Glass and Gas) Company | 1990/9/7 | 37 | | "I have not been very active as a social filmmaker anymore after the r |
| 7547 XXX | 1989/1/1 | 65 | | We don't have an overview translated in English. Help us expand our |
| 8038 Za gyakutai: Nyotai ikedori-jen | 1987/7/18 | 65 | | We don't have an overview translated in English. Help us expand our |
| 8694 Double Rope Torture | 1985/1/15 | 20 | | Two sisters are captured by a bounty hunter for an S&M nightclub . |
| 8832 Shikijō porno: Hageshiku ugoite | 1982/1/01 | 80 | | Pinku film from 1982, distributed by Nikkatsu. |
| 8842 Porno document: Toroku tokkyū bin | 1982/2/25 | 100 | | Pinku from 1982. |
| 8957 Office Lady Rope Slave | 1981/1/23 | 20 | | Two assistants to an S&M photographer decide to practise their maste |
| 9195 Cars | 1979/1/1 | 65 | | "I don't drive, but I know people who" "I drive 100 metres to go t |
| 9628 Female Convict 101: Suck | 1977/8/6 | 50 | | Nikkatsu Films sensation Naomi Tani (FLOWER AND SNAKE) stars |
| 9735 Barbie | 1977/1/1 | 20 | | Barbie comes home from shopping. She takes her groceries out of the |
| 10013 Onsen porno chitai | 1972/1/8 | 60 | | Nikkatsu Roman Porno |
| 10027 Hentai-ehatsu-ehi: Shikiyoku no kyōen | 1971/12/22 | 90 | | Nikkatsu Roman Porno |

補完後仍有44筆空值 ➡ 匯入後移除

PROCESS





背景與動機

2

Movies Produced per Year Since 1903-2023

● 初期1895-1927

1895年盧米埃爾兄弟在巴黎舉辦了首場商業放映<離開工廠>

1908年電影專利公司 (MPPC)

● 中期1927-1960

1927有聲電影<爵士歌手>

1930-1940年代西部片成熟

二戰後通膨及電視的普及影響

● 中後期1960-1980

新自由主義與反戰意識

《野蠻人》、《2001 太空漫遊》

● 近代1980-現今

電腦特效的與低成本電影的發展

第一部動畫電影1995<玩具總動員>

Count

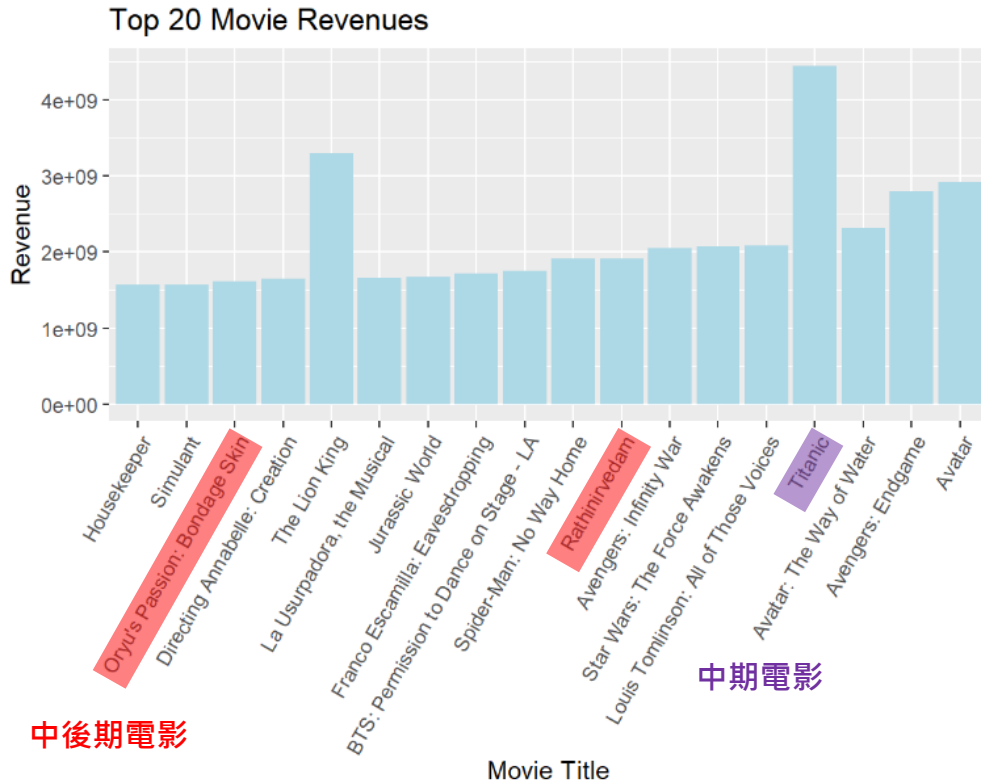
1920

1960

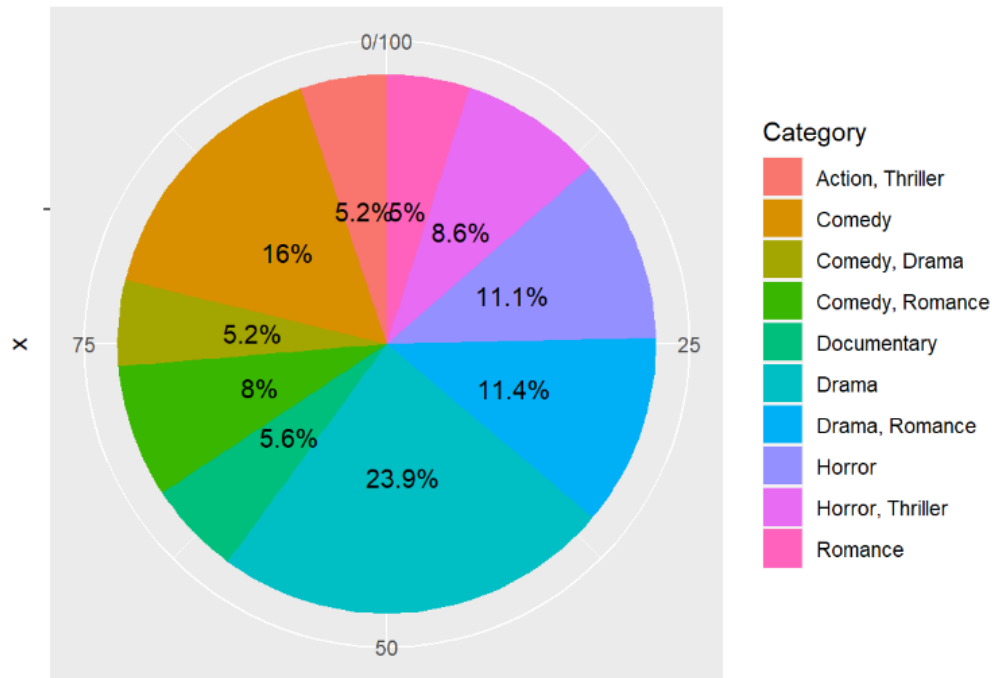
2000

Year

票房前20的電影包含了多種類型



Top 10 Genre of Movies Since 1903-2023



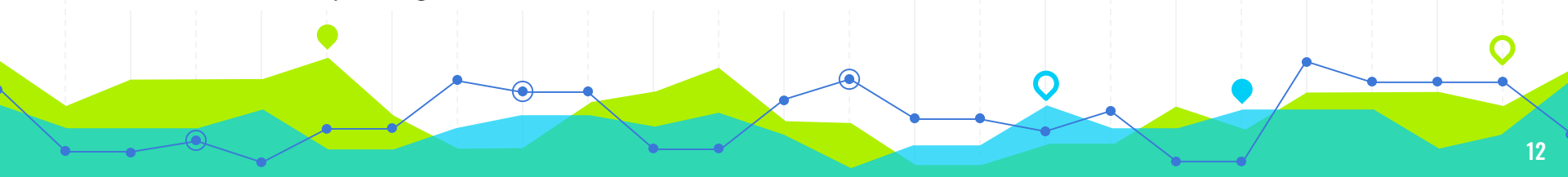
動機

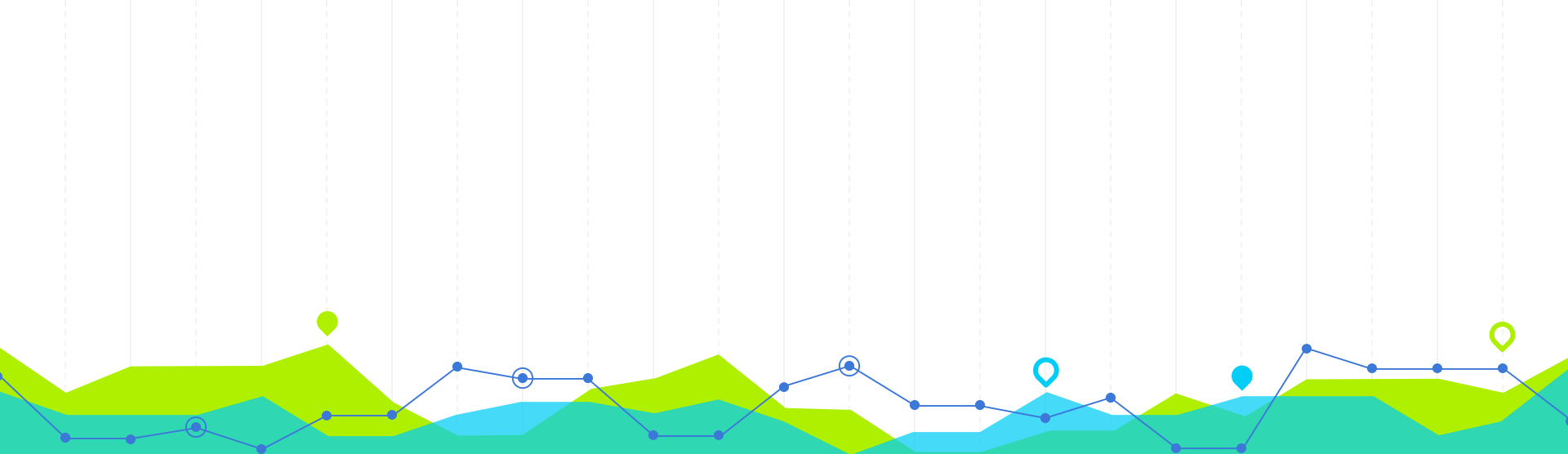
1. 電影的發展反映了各階段社會的變遷
2. 不同種類的電影都有取得高票房的機會



主題

透過關聯式分析各時期獲得高票房的電影種類

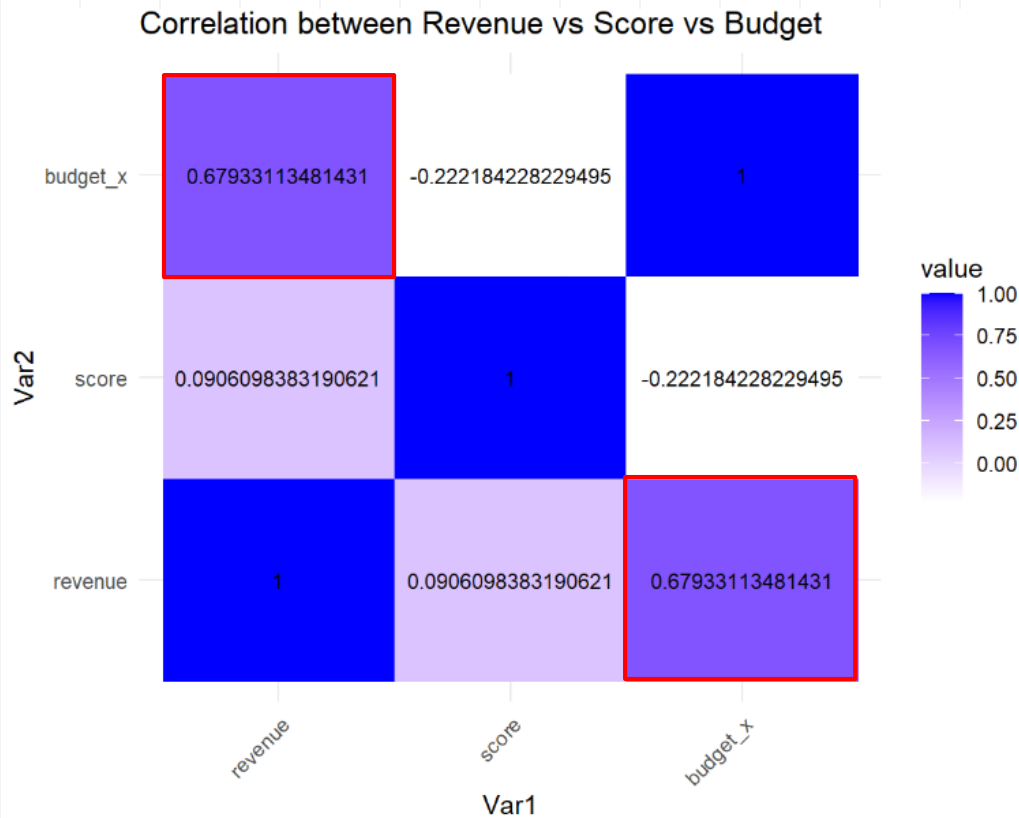




預算與票房關係

3

製片預算與票房有一定的相關性



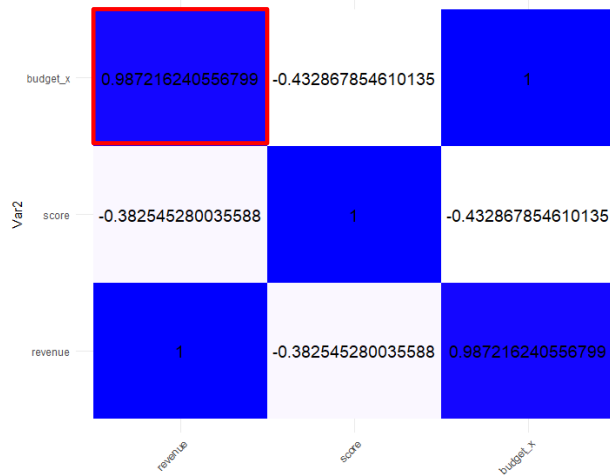
各時期狀況

相關係數逐漸下降

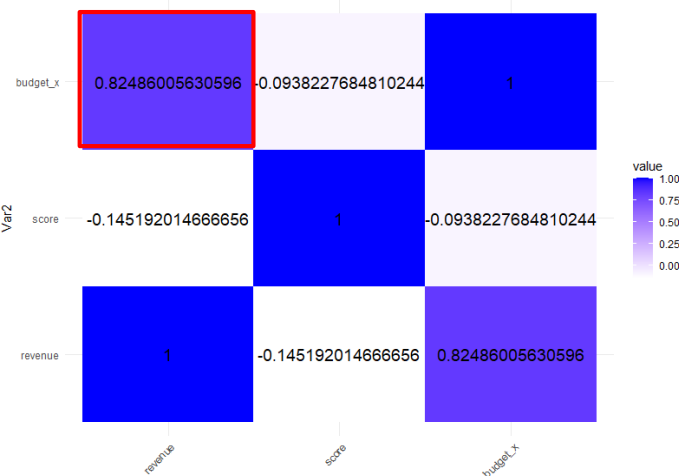


過去預算與票房
有很緊密的關聯，
現代這種現象變
得沒那麼顯著

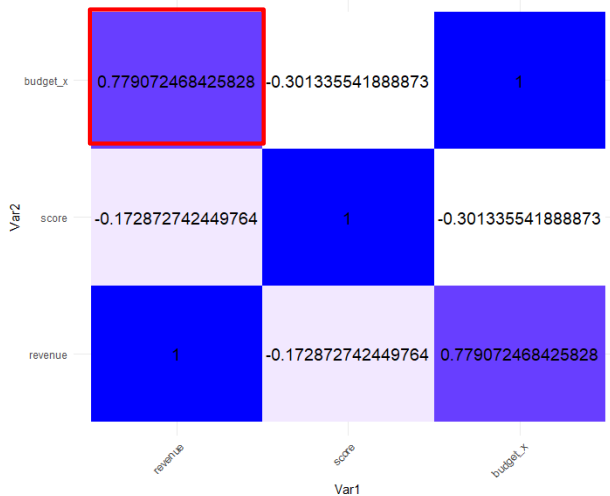
Correlation between Revenue vs Score vs Budget(1895-1927)



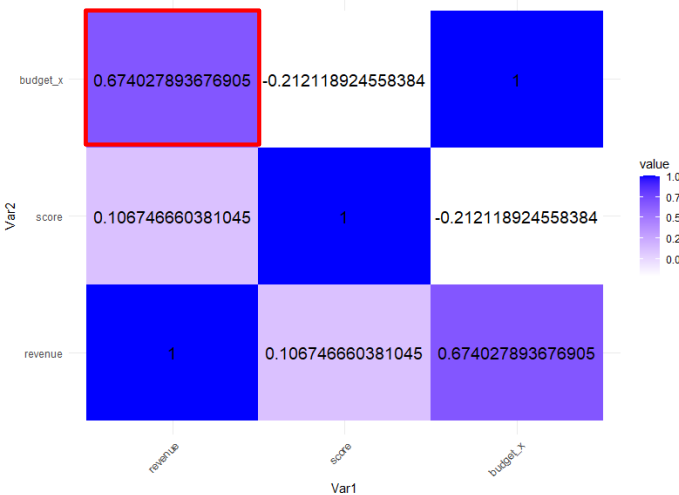
Correlation between Revenue vs Score vs Budget(1927-1960)

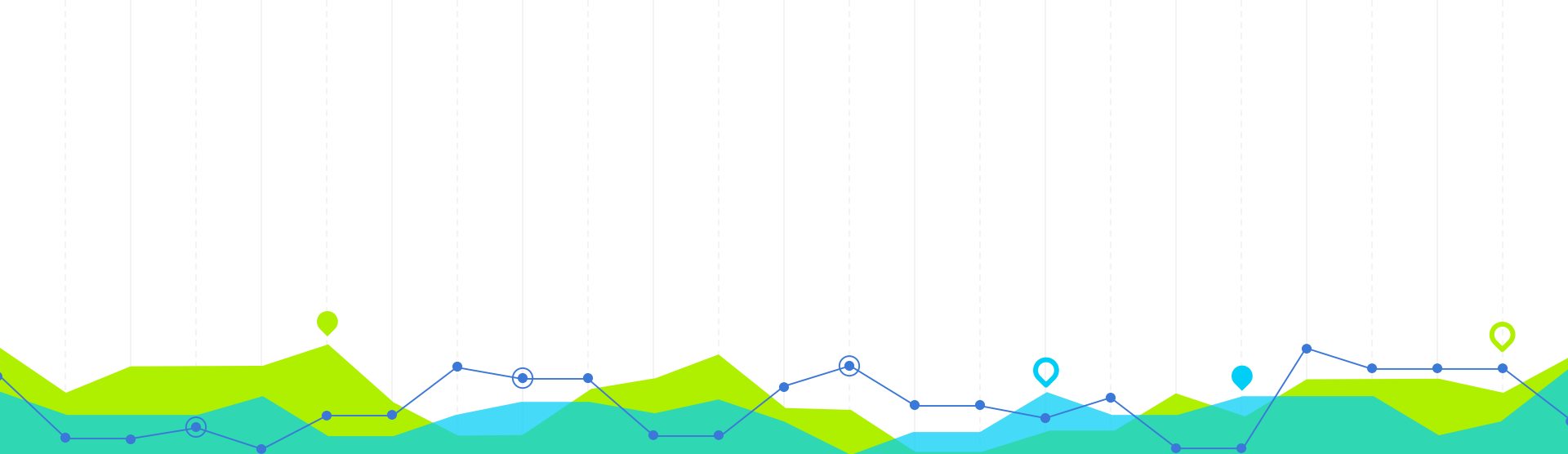


Correlation between Revenue vs Score vs Budget(1960-1980)



Correlation between Revenue vs Score vs Budget(1980-2023)





種類與票房關係

4

時期與票房高低定義

四個時期

早期(1895-1927)

中期(1927-1960)

中後期(1960-1980)

近期(1980-2023)

三段收入區間

```
> summary(dataset$revenue)
```

| Min. | 1st Qu. | Median | Mean | 3rd Qu. | Max. |
|-----------|-----------|-----------|-----------|-----------|-----------|
| 0.000e+00 | 2.797e+07 | 1.504e+08 | 2.517e+08 | 4.164e+08 | 2.924e+09 |

低(<2.797e+07)

中(<4.164e+08 & >2.797e+07)

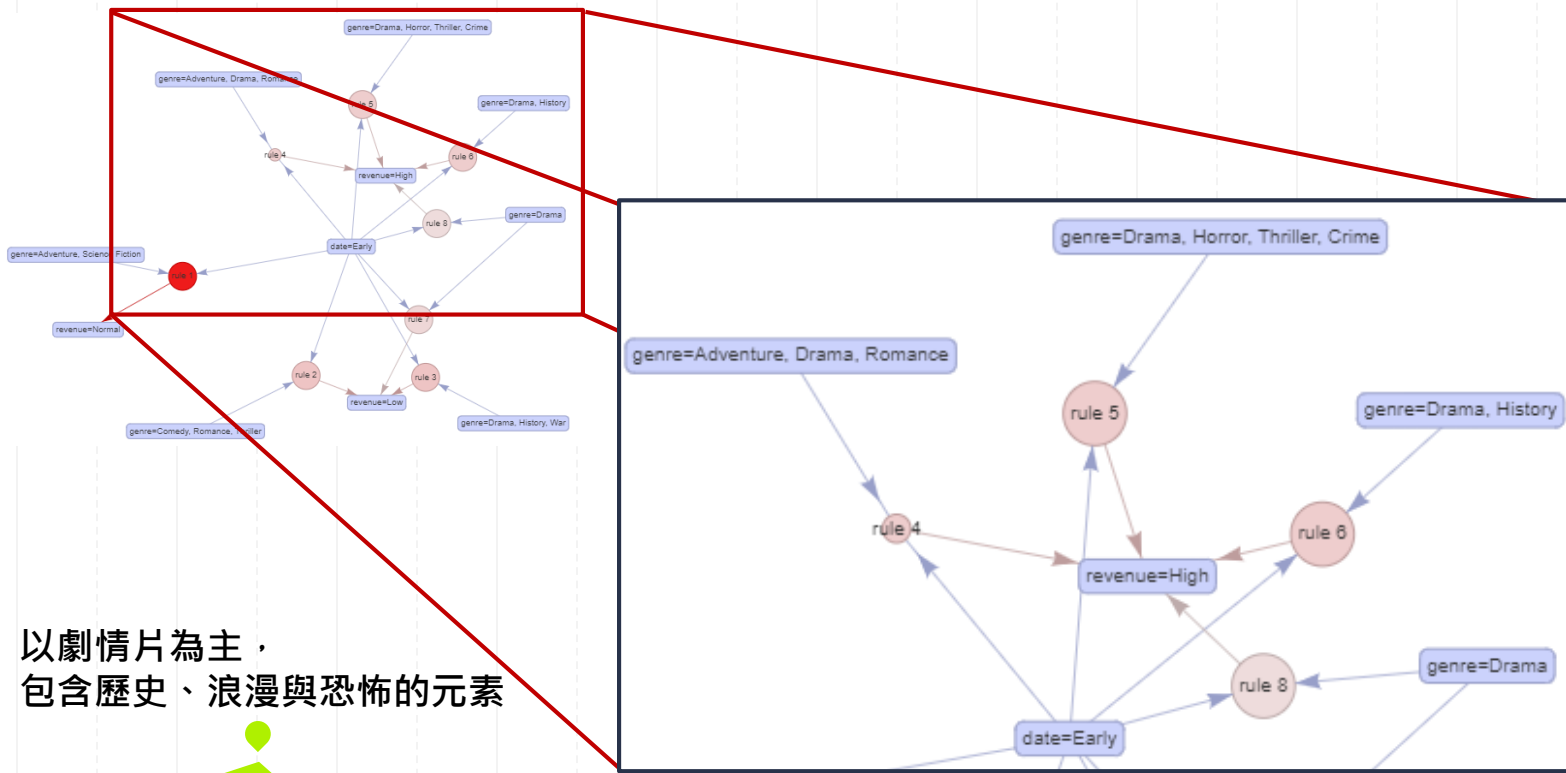
中(>4.164e+08)

Genres' association with Revenue(1895-1927)

| lhs | rhs | support | confidence | coverage | lift | count |
|---|---------------------|---------|------------|----------|-------|-------|
| [1] {date=Early, genre=Adventure, Science Fiction} | => {revenue=Normal} | 0.1 | 1.0 | 0.1 | 10.00 | 1 |
| [2] {date=Early, genre=Adventure, Drama, Romance} | => {revenue=High} | 0.1 | 1.0 | 0.1 | 2.00 | 1 |
| [3] {date=Early, genre=Drama, Horror, Thriller, Crime} | => {revenue=High} | 0.1 | 1.0 | 0.1 | 2.00 | 1 |
| [4] {date=Early, genre=Comedy, Romance, Thriller} | => {revenue=Low} | 0.1 | 1.0 | 0.1 | 2.50 | 1 |
| [5] {date=Early, genre=Drama, History} | => {revenue=High} | 0.2 | 1.0 | 0.2 | 2.00 | 2 |
| [6] {date=Early, genre=Drama, History, War} | => {revenue=Low} | 0.2 | 1.0 | 0.2 | 2.50 | 2 |
| [7] {date=Early, genre=Drama} | => {revenue=Low} | 0.1 | 0.5 | 0.2 | 1.25 | 1 |
| [8] {date=Early, genre=Drama} | => {revenue=High} | 0.1 | 0.5 | 0.2 | 1.00 | 1 |

Lift=1,
該關聯規則
表示無關

Genres' association with Revenue(1895-1927)

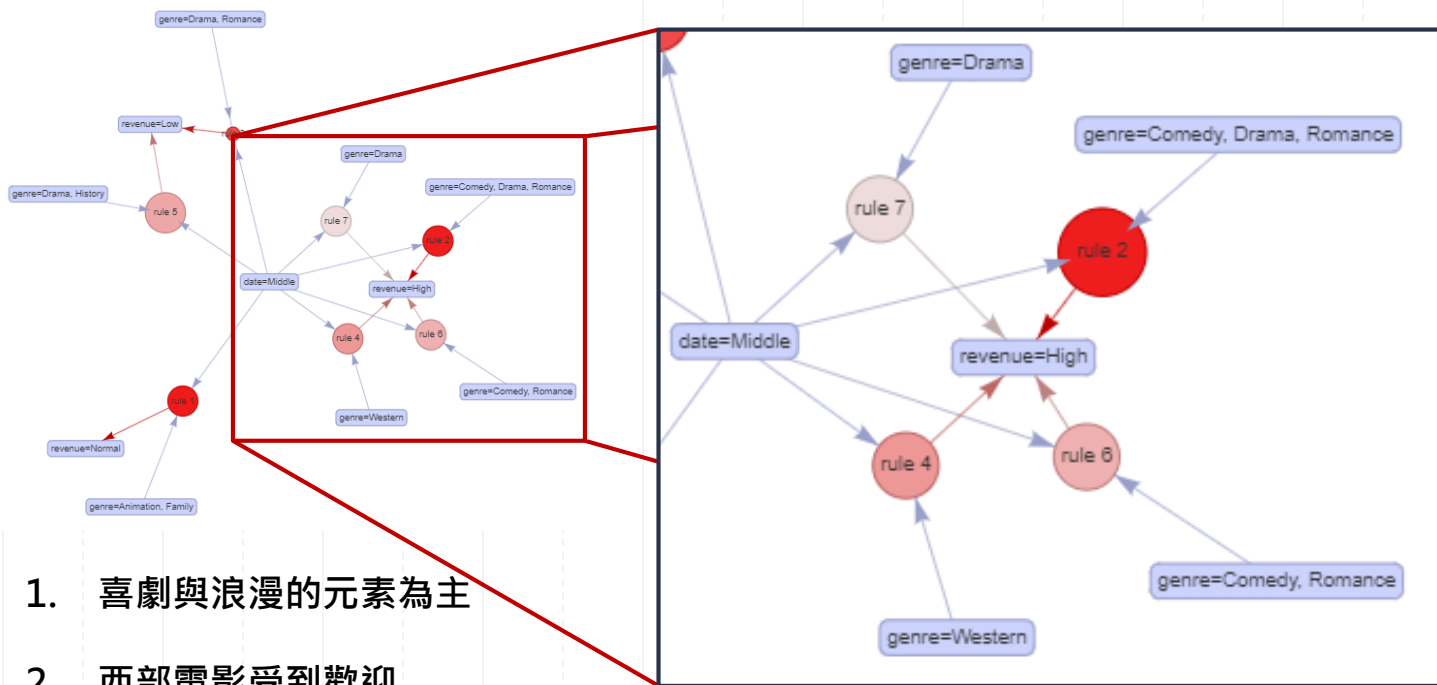


以劇情片為主，
包含歷史、浪漫與恐怖的元素

Genres' association with Revenue(1927-1960)

| lhs | rhs | support | confidence | coverage | lift | count |
|--|---------------------|------------|------------|------------|----------|-------|
| [1] {date=Middle, genre=Animation, Family} | => {revenue=Normal} | 0.02173913 | 0.6000000 | 0.03623188 | 2.435294 | 3 |
| [2] {date=Middle, genre=Comedy, Drama, Romance} | => {revenue=High} | 0.02898551 | 1.0000000 | 0.02898551 | 2.421053 | 4 |
| [3] {date=Middle, genre=Drama, Romance} | => {revenue=Low} | 0.02173913 | 0.7500000 | 0.02898551 | 2.202128 | 3 |
| [4] {date=Middle, genre=Western} | => {revenue=High} | 0.02173913 | 0.7500000 | 0.02898551 | 1.815789 | 3 |
| [5] {date=Middle, genre=Drama, History} | => {revenue=Low} | 0.02898551 | 0.5714286 | 0.05072464 | 1.677812 | 4 |
| [6] {date=Middle, genre=Comedy, Romance} | => {revenue=High} | 0.02898551 | 0.6666667 | 0.04347826 | 1.614035 | 4 |
| [7] {date=Middle, genre=Drama} | => {revenue=High} | 0.02898551 | 0.5000000 | 0.05797101 | 1.210526 | 4 |

Genres' association with Revenue(1927-1960)

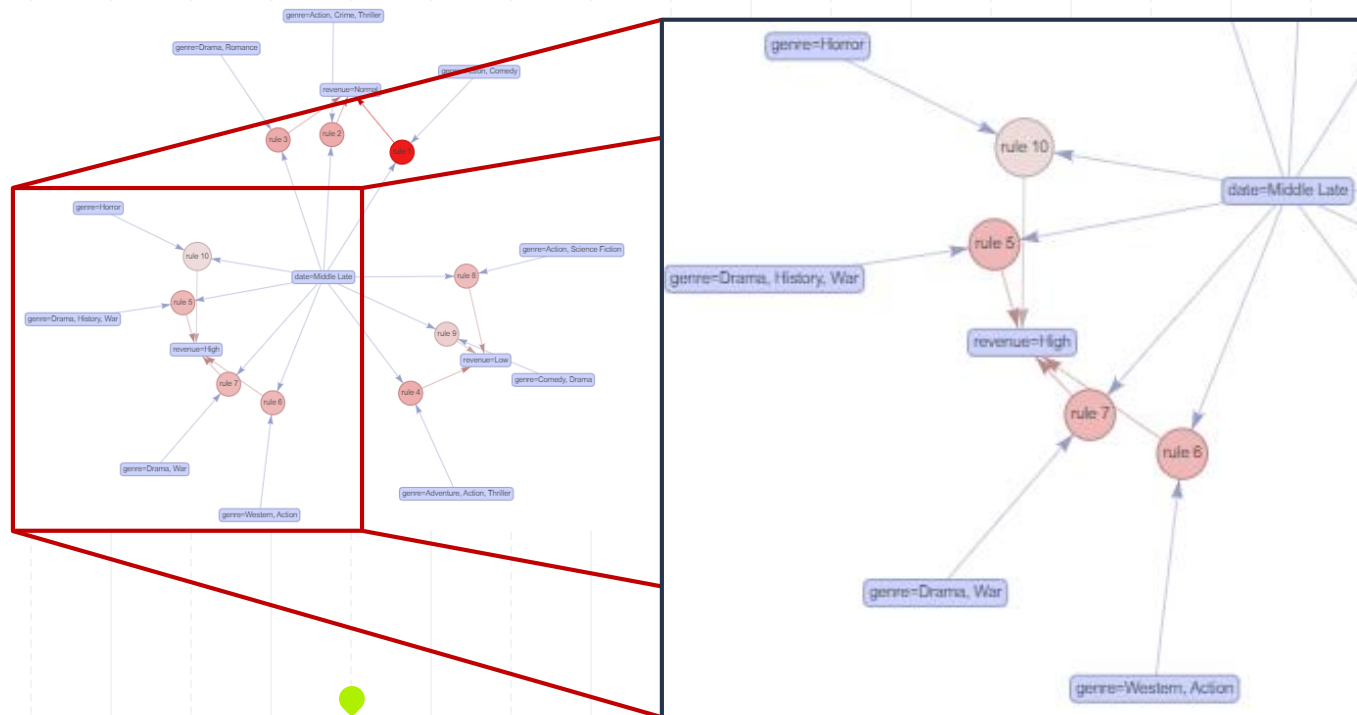


1. 喜劇與浪漫的元素為主
2. 西部電影受到歡迎

Genres' association with Revenue(1960-1980)

| | lhs | rhs | support | confidence | coverage | lift | count |
|------|--|---------------------|-------------|------------|-------------|----------|-------|
| [1] | {date=Middle Late, genre=Action, Comedy} | => {revenue=Normal} | 0.007142857 | 0.7500000 | 0.009523810 | 3.351064 | 3 |
| [2] | {date=Middle Late, genre=Action, Crime, Thriller} | => {revenue=Normal} | 0.007142857 | 0.5000000 | 0.014285714 | 2.234043 | 3 |
| [3] | {date=Middle Late, genre=Drama, Romance} | => {revenue=Normal} | 0.019047619 | 0.5000000 | 0.038095238 | 2.234043 | 8 |
| [4] | {date=Middle Late, genre=Adventure, Action, Thriller} | => {revenue=Low} | 0.011904762 | 0.6250000 | 0.019047619 | 2.187500 | 5 |
| [5] | {date=Middle Late, genre=Drama, History, War} | => {revenue=High} | 0.007142857 | 1.0000000 | 0.007142857 | 2.038835 | 3 |
| [6] | {date=Middle Late, genre=Western, Action} | => {revenue=High} | 0.007142857 | 1.0000000 | 0.007142857 | 2.038835 | 3 |
| [7] | {date=Middle Late, genre=Drama, War} | => {revenue=High} | 0.009523810 | 1.0000000 | 0.009523810 | 2.038835 | 4 |
| [8] | {date=Middle Late, genre=Action, Science Fiction} | => {revenue=Low} | 0.009523810 | 0.5714286 | 0.016666667 | 2.000000 | 4 |
| [9] | {date=Middle Late, genre=Comedy, Drama} | => {revenue=Low} | 0.011904762 | 0.5000000 | 0.023809524 | 1.750000 | 5 |
| [10] | {date=Middle Late, genre=Horror} | => {revenue=High} | 0.016666667 | 0.7777778 | 0.021428571 | 1.585761 | 7 |

Genres' association with Revenue(1960-1980)

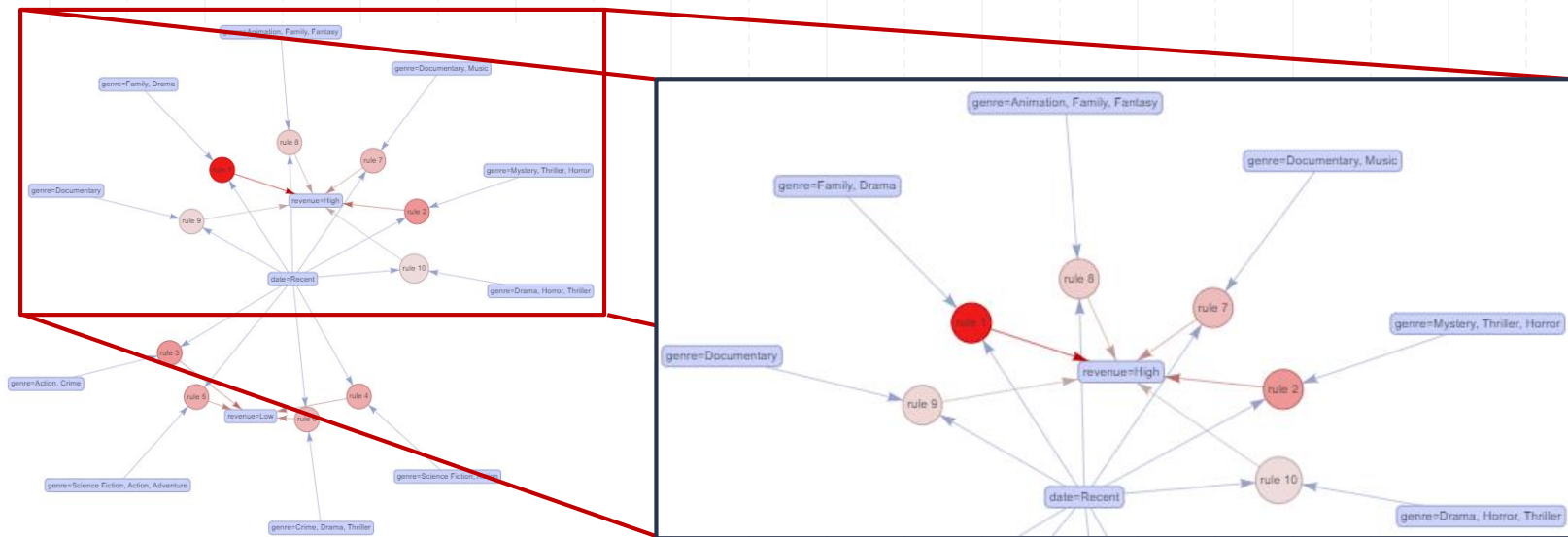


1. 戰爭片取代喜劇與浪漫
2. 西部電影仍受到歡迎

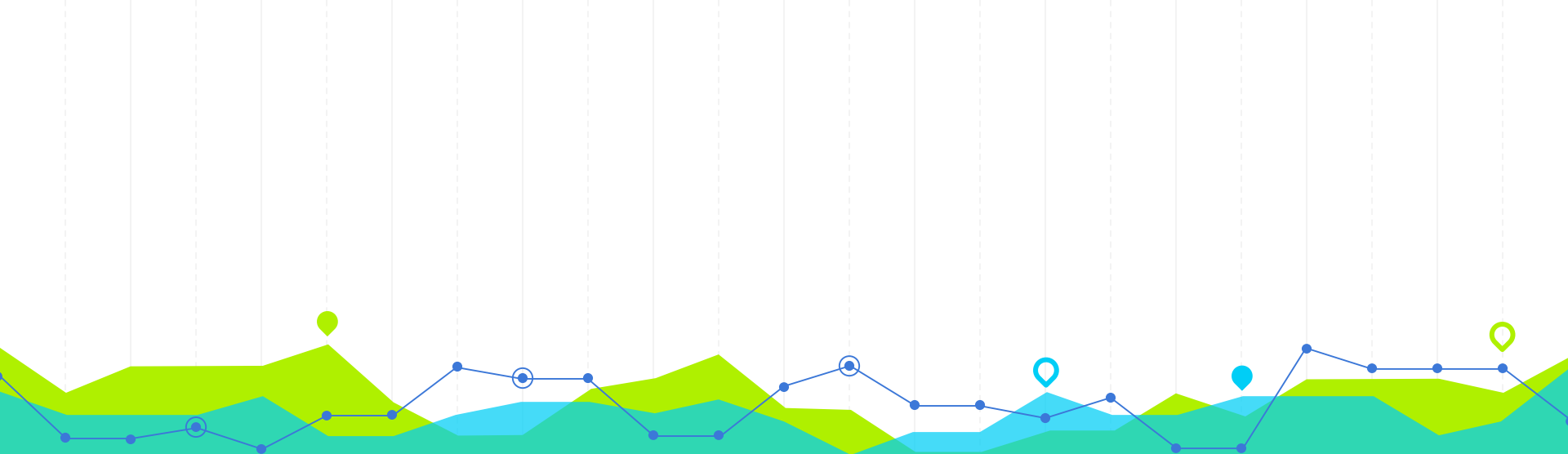
Genres' association with Revenue(1980-2023)

| | lhs | rhs | support | confidence | coverage | lift | count |
|------|--|-------------------|-------------|------------|-------------|----------|-------|
| [1] | {date=Recent, genre=Family, Drama} | => {revenue=High} | 0.001155462 | 0.9166667 | 0.001260504 | 2.074814 | 11 |
| [2] | {date=Recent, genre=Mystery, Thriller, Horror} | => {revenue=High} | 0.001155462 | 0.7857143 | 0.001470588 | 1.778412 | 11 |
| [3] | {date=Recent, genre=Action, Crime} | => {revenue=Low} | 0.001785714 | 0.5483871 | 0.003256303 | 1.767314 | 17 |
| [4] | {date=Recent, genre=Science Fiction, Action} | => {revenue=Low} | 0.001155462 | 0.5238095 | 0.002205882 | 1.688107 | 11 |
| [5] | {date=Recent, genre=Science Fiction, Action, Adventure} | => {revenue=Low} | 0.001155462 | 0.5238095 | 0.002205882 | 1.688107 | 11 |
| [6] | {date=Recent, genre=Crime, Drama, Thriller} | => {revenue=Low} | 0.002941176 | 0.5090909 | 0.005777311 | 1.640672 | 28 |
| [7] | {date=Recent, genre=Documentary, Music} | => {revenue=High} | 0.001050420 | 0.7142857 | 0.001470588 | 1.616738 | 10 |
| [8] | {date=Recent, genre=Animation, Family, Fantasy} | => {revenue=High} | 0.001575630 | 0.6818182 | 0.002310924 | 1.543250 | 15 |
| [9] | {date=Recent, genre=Documentary} | => {revenue=High} | 0.009033613 | 0.6615385 | 0.013655462 | 1.497348 | 86 |
| [10] | {date=Recent, genre=Drama, Horror, Thriller} | => {revenue=High} | 0.001785714 | 0.6538462 | 0.002731092 | 1.479937 | 17 |

Genres' association with Revenue(1980-2023)



1. 電影種類漸漸多元，家庭的元素受到重視
2. 動畫電影、紀錄片、恐怖片等類型都有不錯的票房



結論

5

各時期的製片預算與票房都有一定程度的關聯，隨著科技發展有越來越不明顯的趨勢

1.

分析結果與預期的動作與科幻片不同，也有可能是因為種類太多及技術參差不齊

2.

劇情片元素在每個時期與高票房都有關聯

3.

多元化、視覺效果和技術創新更加受到重視，也會成為未來的發展趨勢

4.

參考資料

1. LIBRARIES : <https://open.lib.umn.edu/mediaandculture/chapter/8-2-the-history-of-movies/>
2. 國家電影及是聽文化中心 : <https://edumovie-tfai.org.tw/article/content/409>
3. Kaggle dataset : <https://www.kaggle.com/datasets/ashpalsingh1525/imdb-movies-dataset>

THANKS!

