



วิเคราะห์ประเภทของภาพยนตร์ใดที่ผู้ชมเลือกดู

เสนอ

ผศ.ดร.นุรีย์ วิวัฒน์วัฒนา

ผศ.ดร.รัตน์ชัยนันท์ ธรรมสุจริต

จัดทำโดย

นายชาคริต ธีรนนท์ รหัสนิสิต 65102010191

คำนำ

ในยุคที่เทคโนโลยีการสตรีมมิ่งเข้ามามีบทบาทสำคัญต่อพฤติกรรมการรับชมสื่อของผู้บริโภค แพลตฟอร์มอย่าง Netflix ได้กลายเป็นหนึ่งในสื่อบันเทิงหลักของผู้คนทั่วโลก การทำความเข้าใจแนวโน้มความนิยมของผู้ชมต่อประเภทของภาพยนตร์จึงเป็นสิ่งจำเป็นอย่างยิ่ง ทั้งในด้านการพัฒนาเนื้อหา การวางแผนกลยุทธ์ทางการตลาด และการปรับปรุงระบบแนะนำคอนเทนต์ให้ตอบสนองความสนใจของผู้ชมได้อย่างตรงจุด

รายงานฉบับนี้มีวัตถุประสงค์เพื่อวิเคราะห์ประเภทของภาพยนตร์ที่ได้รับความนิยมบนแพลตฟอร์ม Netflix โดยใช้กระบวนการวิเคราะห์ข้อมูล (Data Analytics) และเทคนิคเชิงสถิติเบื้องต้น เพื่อค้นหาแนวโน้ม ความสัมพันธ์ และปัจจัยที่ส่งผลต่อความนิยมของภาพยนตร์แต่ละประเภท ซึ่งข้อมูลและข้อค้นพบจากการศึกษานี้จะสามารถนำไปประยุกต์ใช้ในการตัดสินใจเชิงธุรกิจ และพัฒนาประสบการณ์ของผู้ใช้งานได้อย่างมีประสิทธิภาพ

ผู้จัดทำหวังว่ารายงานฉบับนี้จะเป็นประโยชน์ต่อผู้ที่สนใจในการวิเคราะห์ข้อมูล การตลาดดิจิทัล รวมถึงผู้ที่ทำงานด้านการผลิตสื่อและบันเทิงในยุคดิจิทัล

1. Project Canvas

1.1. Problem Statement (ปัญหา)

Netflix มีคอนเทนต์ภาพยนตร์และซีรีส์หลากหลายประเภท (genre) แต่ยังไม่มีความชัดเจนและเป็นระบบว่า ผู้ชมเลือกชมหนังประเภทใดมากที่สุด และแนวโน้มความนิยมของแต่ละประเภทเป็นอย่างไร ซึ่งส่งผลต่อการวางแผนผลิตและจัดหาคอนเทนต์ รวมถึงการทำการตลาดและพัฒนาระบบแนะนำภาพยนตร์

1.2. Business Objectives (วัตถุประสงค์ทางธุรกิจ)

- ระบุประเภทภาพยนตร์ที่ได้รับความนิยมสูงสุดในแพลตฟอร์ม Netflix
- สนับสนุนการตัดสินใจในการลงทุนผลิตหรือจัดหาคอนเทนต์ใหม่ที่ตรงกับความต้องการของผู้ชม
- พัฒนากลยุทธ์การตลาดและระบบแนะนำภาพยนตร์ให้มีประสิทธิภาพมากขึ้น
- เพิ่มความพึงพอใจและการมีส่วนร่วมของผู้ใช้งานบนแพลตฟอร์ม

1.3. Key Stakeholders (ผู้มีส่วนได้ส่วนเสีย)

- ทีมบริหารและฝ่ายกลยุทธ์ Netflix
- ทีมพัฒนาคอนเทนต์ (Content Acquisition & Production)
- ฝ่ายการตลาดและโฆษณา
- ทีมพัฒนาระบบ Recommendation Engine
- ผู้ชมและสมาชิก Netflix

1.4. Success Metrics (ตัวชี้วัดความสำเร็จ)

- ความถูกต้องและความชัดเจนของรายงานวิเคราะห์ genre ที่นำเสนอ

- จำนวนและความหลากหลายของภาพยนตร์ใน genre ที่ได้รับความนิยมเพิ่มขึ้นภายใน 1 ปี
- อัตราการรับชมของคนเอนต์ใน genre ที่ระบุว่าเป็นที่นิยม

1.5. Canvas Sketch

1.5.1. กลุ่มเป้าหมาย (Target Audience)

- ทีมบริหาร Netflix ที่วางกลยุทธ์การผลิตและจัดหาคอนเทนต์
- ฝ่ายการตลาดและโฆษณา
- นักวิเคราะห์ข้อมูลและทีมพัฒนาระบบแนะนำคอนเทนต์
- ผู้ชม Netflix ทั่วไปที่เป็นกลุ่มเป้าหมายของแพลตฟอร์ม

1.5.2. ทรัพยากร (ข้อมูล/เครื่องมือ)

- Netflix Titles Dataset แหล่งข้อมูลประกอบด้วยข้อมูลภาพยนตร์และซีรีส์ ได้แก่: ชื่อ, ปีที่ออกฉาย (release_year), country, principal country, rating
- เครื่องมือวิเคราะห์ข้อมูล: Python (Pandas, Matplotlib/Seaborn) และ Tableau

1.5.3. สมมติฐาน

- จำนวนรายการในแต่ละ genre สะท้อนถึงความนิยมของผู้ชมในระดับหนึ่ง
- ความนิยมของ genre มีแนวโน้มเปลี่ยนแปลงตามปีที่ออกฉายและเรตติ้ง
- พฤติกรรมการเลือกชมในอดีตสามารถคาดการณ์แนวโน้มในอนาคตได้

1.5.4. ข้อจำกัด

- ไม่มีข้อมูลพฤติกรรมมารับชมจริง (เช่น เวลาชม, จำนวนครั้งที่ดู)
- ความนิยมที่วัดจากจำนวนรายการใน genre อาจไม่สะท้อนพฤติกรรมผู้ชมได้เต็มที่
- ข้อมูลอาจมีบางส่วนไม่สมบูรณ์หรือมีข้อผิดพลาดเล็กน้อย

1.5.5. ไทม์ไลน์ (Timeline)

| วัน | กิจกรรม |
|-----|----------------------------------|
| 1-2 | รวบรวมและทำความสะอาดข้อมูล |
| 3-4 | วิเคราะห์ข้อมูลและสร้างกราฟ |
| 5-6 | สรุปผลการวิเคราะห์และเขียนรายงาน |
| 7 | นำเสนอ |

2. การเตรียมข้อมูล (Data Preparation)

- การทำความสะอาดข้อมูล (Data Cleaning) :

Drop

```
# Drop Columns
data.drop(columns=['show_id', 'cast', 'director', 'description'], inplace=True)

# Replacements
data['country'] = data['country'].fillna(data['country'].mode()[0])

# Drop the rest
data.dropna(inplace=True)
```

บรรทัดที่ 1 : ลบคอลัมน์ที่ไม่จำเป็นออกจาก DataFrame ชื่อ `data` ได้แก่:

- `show_id`
- `cast`
- `director`
- `description`

บรรทัดที่ 2 : กรอกค่าที่หาย (NaN) ในคอลัมน์ `country` ด้วย ค่าที่พบบ่อยที่สุด (mode) ของคอลัมน์นั้น

บรรทัดที่ 3 : ลบแถว (row) ที่ยังมี ค่าหาย (missing values) อยู่ หลังจากขั้นตอนเติมค่าหายแล้ว

- การสร้างคุณลักษณะใหม่ (Feature Engineering):

```
Added Date

[ ] # Change Data Type to DateTime
    data['date_added'] = pd.to_datetime(data['date_added'].str.strip(), format='mixed', errors='coerce')
```

แปลงข้อมูลในคอลัมน์ `date_added` ให้เป็นชนิดวันที่ (datetime)

```
Country Column

[ ] # เก็บไว้เฉพาะประเทศแรกจากแต่ละรายการเท่านั้น
    data['principal_country'] = data['country'].apply(lambda x: x.split(",")[0])
```

ดึงชื่อประเทศแรกจากคอลัมน์ `country` มาเก็บในคอลัมน์ใหม่ชื่อ `principal_country` เพื่อระบุประเทศหลักของแต่ละรายการ

```
ratings_ages = {
    'TV-PG': 'Older Kids',
    'TV-MA': 'Adults',
    'TV-Y7-FV': 'Older Kids',
    'TV-Y7': 'Older Kids',
    'TV-14': 'Teens',
    'R': 'Adults',
    'TV-Y': 'Kids',
    'NR': 'Adults',
    'PG-13': 'Teens',
    'TV-G': 'Kids',
    'PG': 'Older Kids',
    'G': 'Kids',
    'UR': 'Adults',
    'NC-17': 'Adults'
}
```

เปลี่ยนค่าการจัตเรตเป็นกลุ่มอายุที่สอดคล้องกับเรตนั้น

```
data['ages'] = data['rating'].replace(ratings_ages)
```

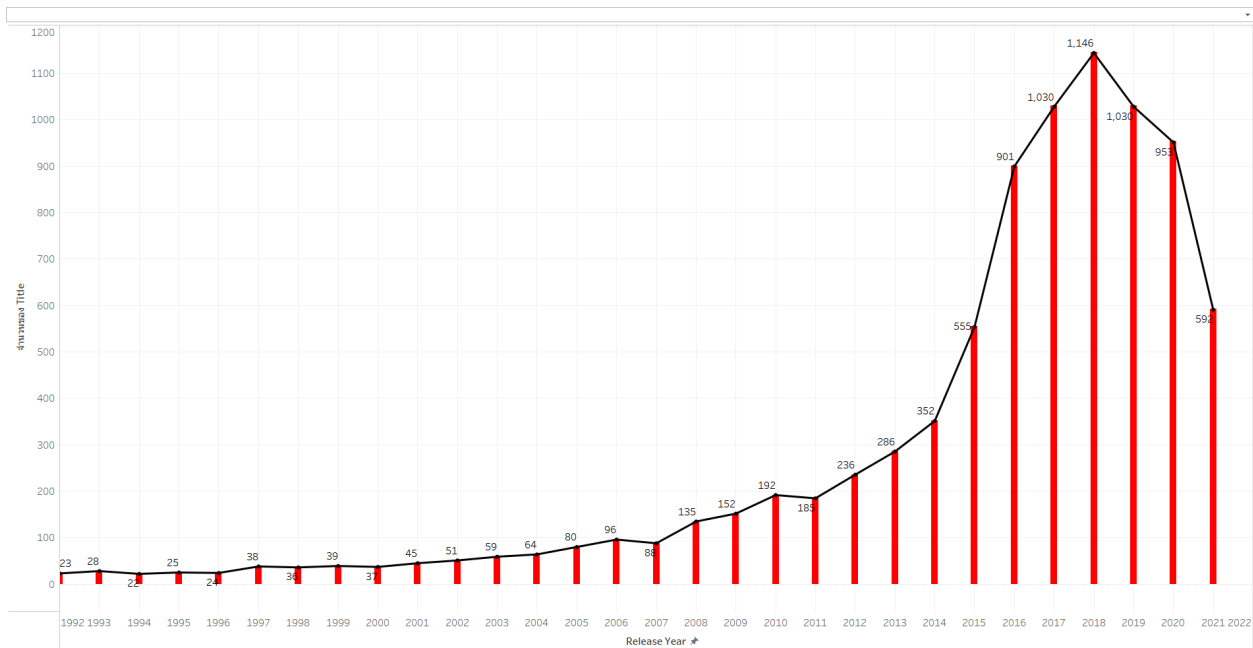
```
data['ages'].unique()
```

เปลี่ยนข้อมูลการจัดเรตเป็นกลุ่มอายุที่เข้าใจง่ายขึ้น

3. การสำรวจข้อมูลเบื้องต้น (Exploratory Data Analysis, EDA)

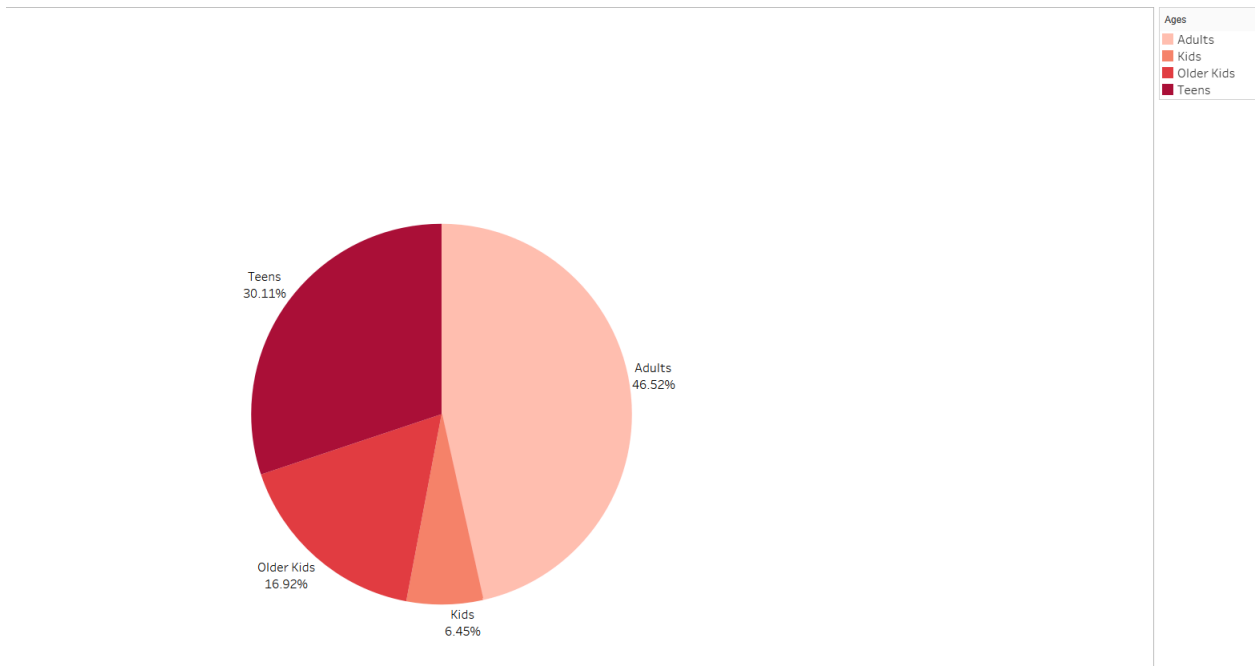
ประเทศใดมีจำนวนเนื้อหามากที่สุด?

: จากแผนที่ USA มากที่สุด

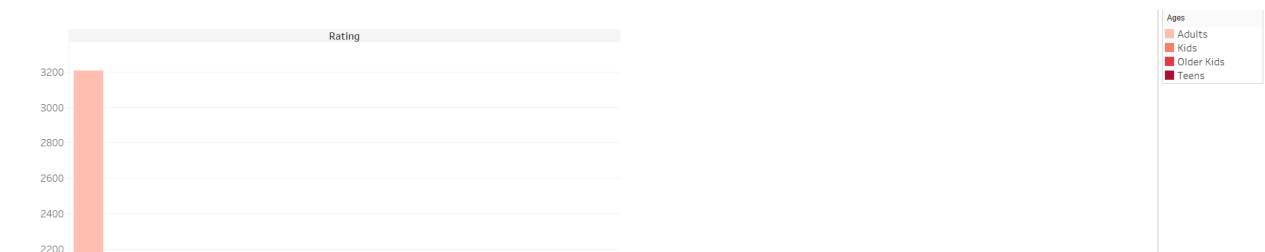
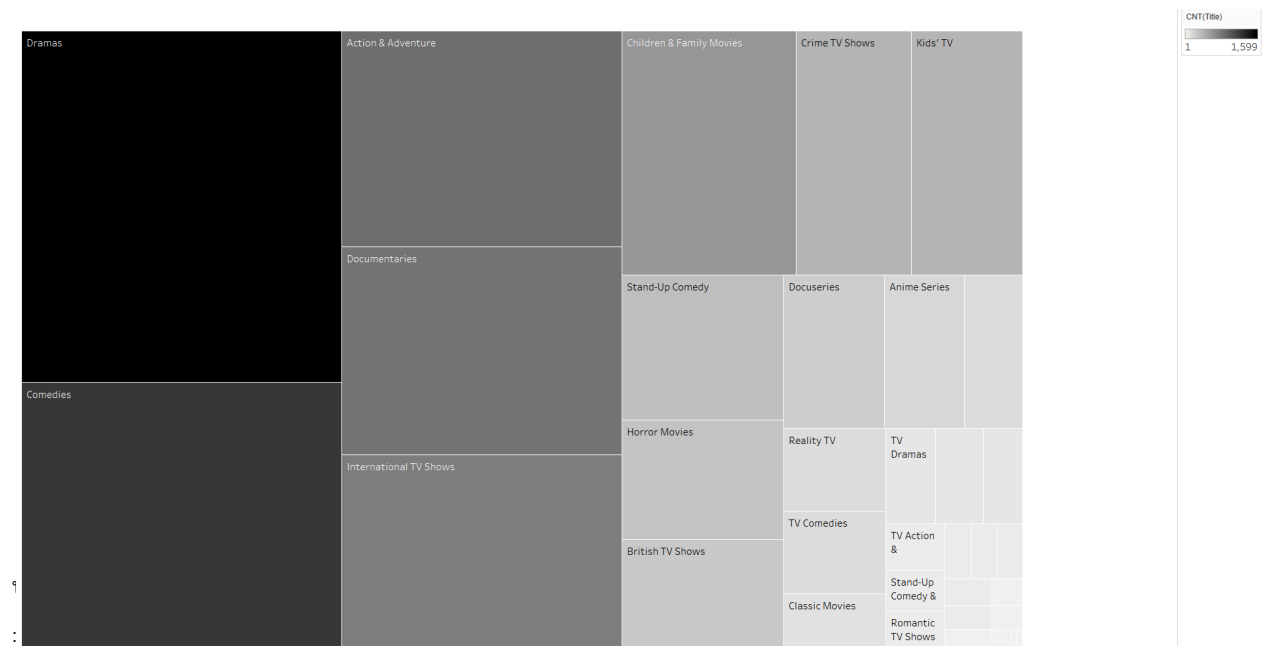


แนวโน้มจำนวนรายการตามปีที่เผยแพร่

: จากกราฟได้โชว์ว่าในช่วงปี 1992-2007 นั้นมีการผลิตภาพยนตร์ ออกมาน้อยมากแต่พอเข้าในช่วง 2007- 2021 นั้นก็ได้มีการสร้างภาพยนตร์ออกมาเพิ่มมากขึ้นแบบก้าวกระโดด



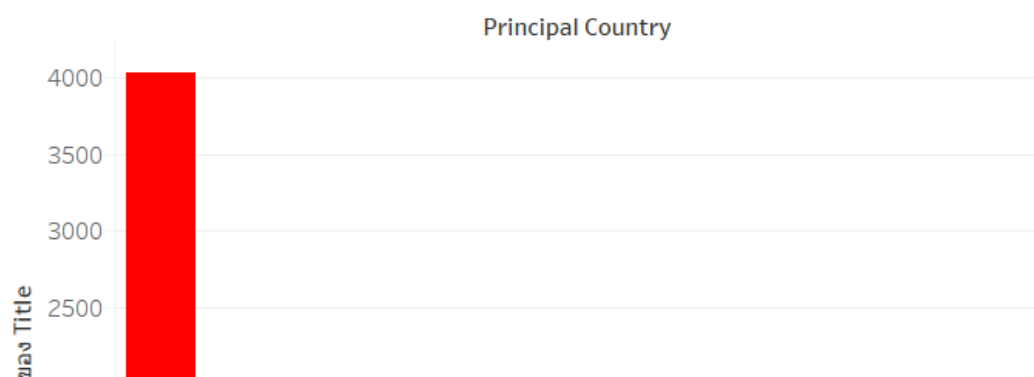
สัดส่วนของช่วงอายุในการรับชม



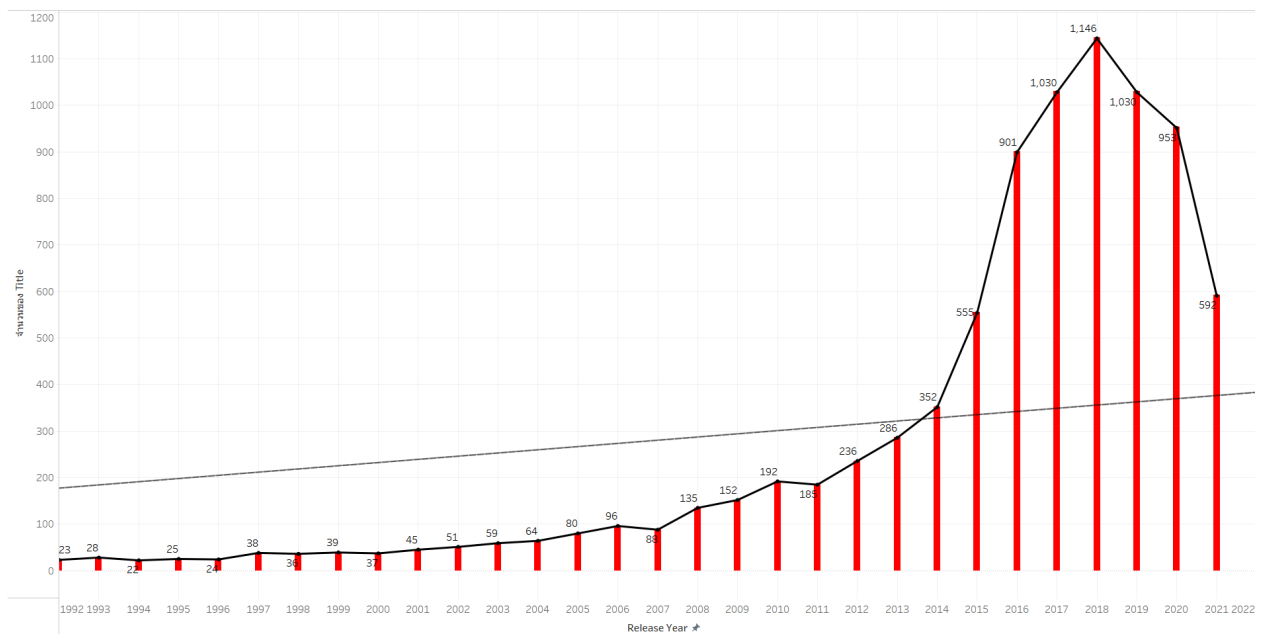
เรตอายุของรายการเป็นเท่าไรบ้าง?

: เรตส่วนใหญ่ที่ผู้ชมเลือกดูจะเป็น TV-MA เพราะเป็นเรตสำหรับ Adults และ Adults มีประมาณมากที่สุด รองลงมาจะเป็น เรตรายการสำหรับ Teens และ Olders Kids

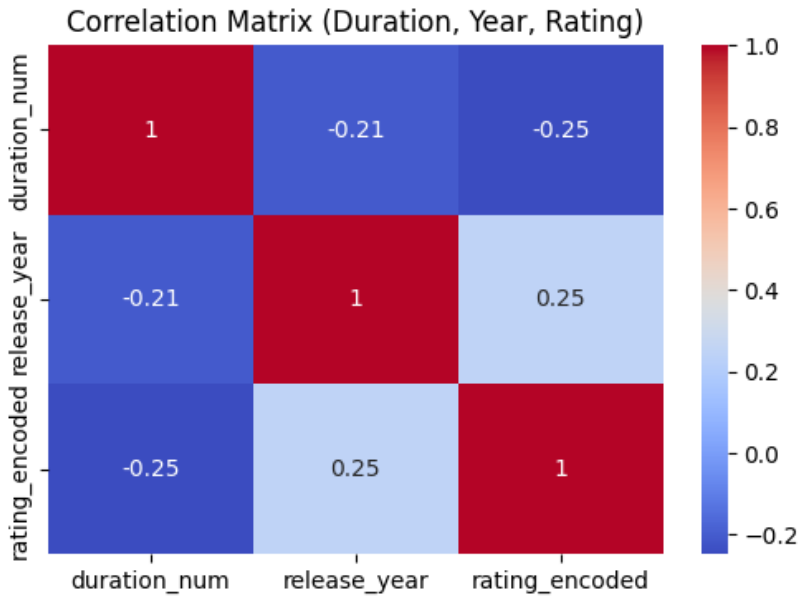
4. การวิเคราะห์เชิงลึก (In-Depth Analysis)



การเติบโตของภาพยนตร์ในประเทศ 10 อันดับแรกนั้นจะเห็นได้ว่า USA นั้นมีจำนวนเนื้อหามากที่สุดรองลงมาเป็น India , UK , Canada , Japan , France , South Korea , Mexico และ Australia



ทำให้สามารถวิเคราะห์เพิ่มเติมได้ว่าในปี 2014-2018 นั้นเป็นปีที่วงการภาพยนตร์กำลังเติบโตทำให้สามารถสร้างหนังได้ออกมาได้เป็นจำนวนมาก แต่ก็ค่อยๆลดหลังจากปี 2018 ทันทีอาจมีผลมาจากปัจจัยต่างๆ เช่น การมาของโรค Covid-19 และ อื่นๆ



ความสัมพันธ์ :

duration_num vs release_year มีความสัมพันธ์เชิงลบอ่อน ๆ → ยิ่งหนังใหม่ ยิ่งสั้นลงเล็กน้อย

duration_num vs rating_encoded มีความสัมพันธ์เชิงลบอ่อน ๆ → ช่วงอายุเรตบางประเภทมีแนวโน้มสั้นกว่า

release_year vs rating_encoded มีความสัมพันธ์เชิงลบอ่อน ๆ → หนังยุคใหม่มีช่วงอายุเรตต่างไปเล็กน้อย

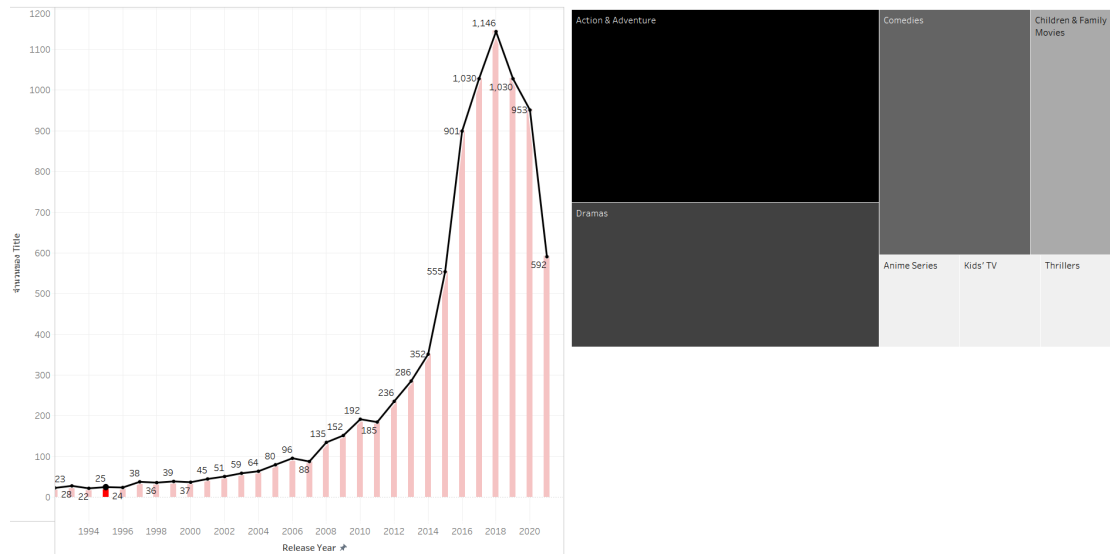
สรุปได้ว่า

- ไม่มีตัวแปรใดที่มีความสัมพันธ์สูงกับความยาวของหนัง
- มีความสัมพันธ์เชิงลบระดับอ่อนระหว่าง duration_num กับ release_year และ rating_encoded
ซึ่งอาจแปลได้ว่า ภาพยนตร์ใหม่ในยุคหลังมีแนวโน้มสั้นลง และหนังบางเรตติ้งมีความยาวเฉลี่ยน้อยกว่าประเภทอื่น

5. ข้อค้นพบและข้อเสนอแนะ (Insights & Recommendations)

5.1. หนังแนว Dramas และ Comedies มีจำนวนผู้ชมติด top 5 ของทุกปี

จากการสำรวจตั้งแต่ปี 1995 เป็นต้นไปยอดดูของ Dramas และ Comedies ไม่เคยหลุดลงจาก top 5 บ่งบอกว่าถ้าจะทำภาพยนตร์ออกนั้นควร ทำเป็นประเภท Dramas หรือ Comedies



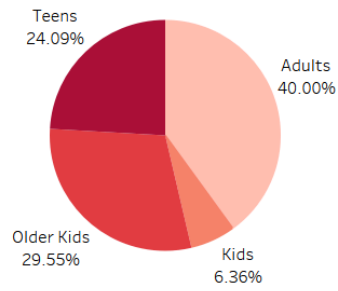
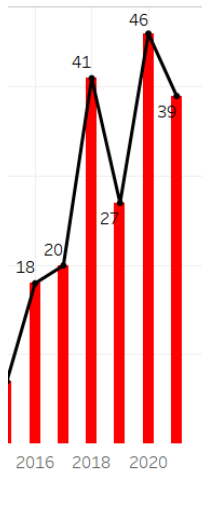
5.2. ทำไมภาพยนตร์ ประเภท Dramas และ Comedies ผู้ชมถึงเยอะ

เป็นเพราะ ภาพยนตร์ ประเภท Dramas และ Comedies ผู้ชมส่วนใหญ่เป็น กลุ่ม Adults นอกจากนี้ยังเป็นการชี้ถึงว่าถ้าต้องการจะทำภาพยนตร์ ที่จะเน้นไปที่แนว Dramas และ Comedies

นั่นถ้าอยากให้มีฐานคนดูเพิ่มมากขึ้น ควรตีกลุ่มตลาดของ Adults จะทำให้ได้ประสิทธิภาพสูงสุด

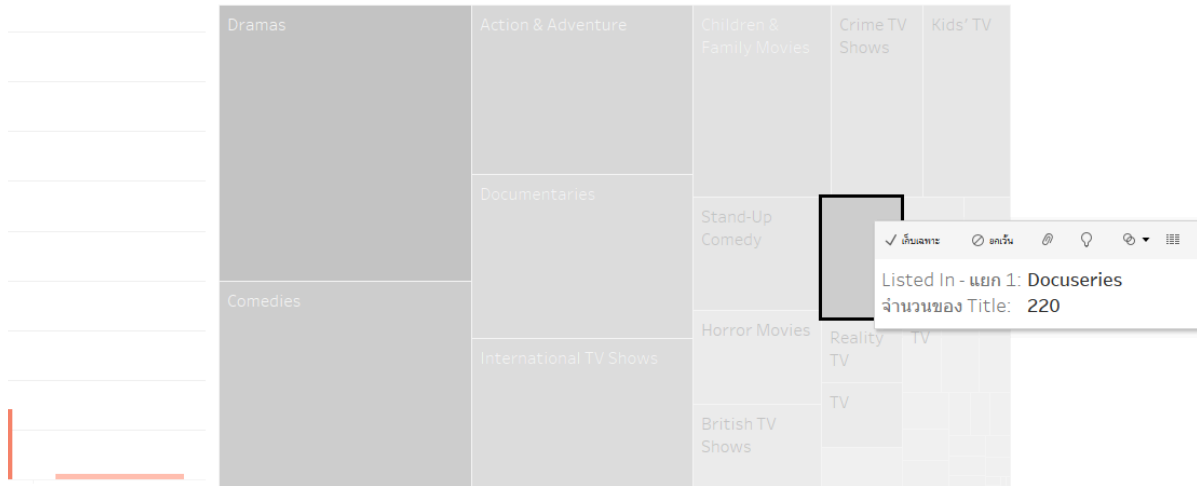
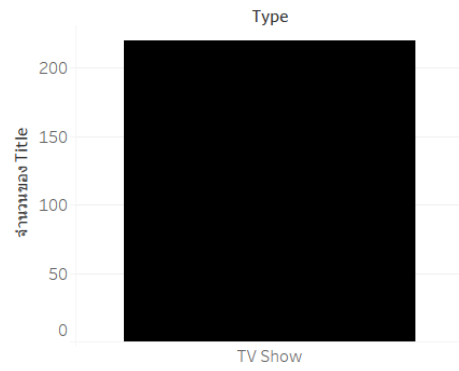


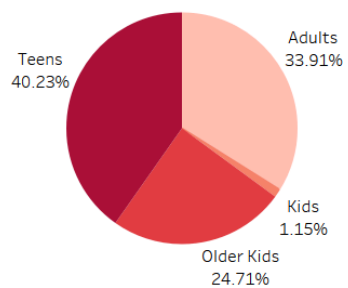
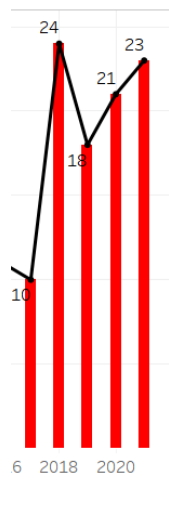
แนวโน้มเพิ่มขึ้น อาจเป็นโอกาสในการตีตลาดกลุ่มผู้ดูใหม่ๆ เพื่อให้เข้ามารับชมมากขึ้น



Titles
220

Ages
Adults
Kids
Older Kids
Teens
จำนวนของ Titl
1





Titles

174

Ages

Adults

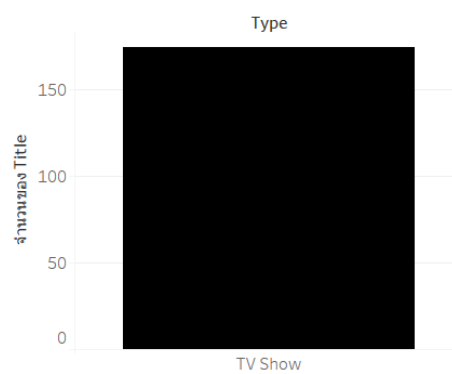
Kids

Older Kids

Teens

จำนวนของ Title

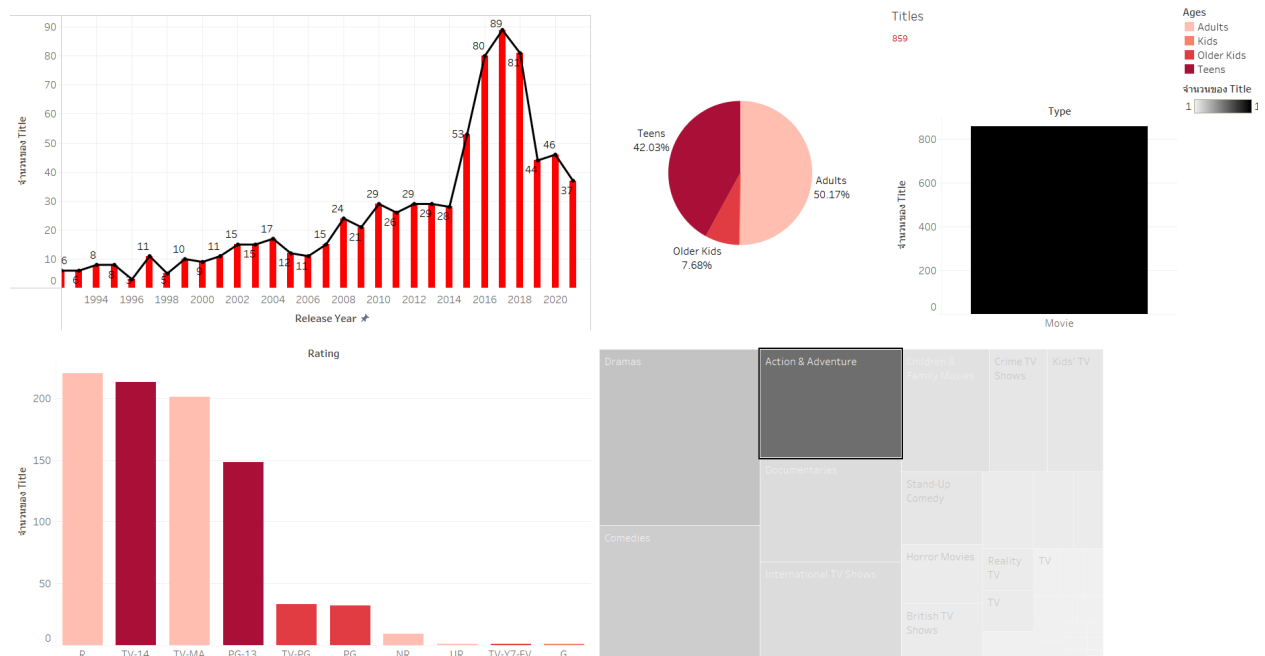
1 1,5



TV-V

5.4 การลดลงของภาพยนตร์

ในบรรดาภาพยนตร์ต่างๆ ส่วนใหญ่ใน ปี 2018 จะมีอัตราที่ลดลงไม่มากนักน้อยแต่ ประเภท Action & Adventure นั้นมีอัตราที่ลดมามากผิดปกติเป็นการบ่งบอกถึง สถานการณ์ของการสร้างภาพยนตร์ประเภท Action & Adventure ว่าอาจจะไม่เป็นที่นิยมแล้วในปัจจุบัน เพราะตั้งแต่ 1992 Action & Adventure ก็ครองยอดภาพยนตร์มาตลอดจนกระทั่ง 2019 ที่ด้วยปัจจัยต่างๆทำให้ยอดภาพยนตร์ ลดจาก 81 > 44 หายไปเกือบครึ่งหนึ่งและยังคงลดต่ำกว่าอีกในปี 2021



สรุปได้ว่า ถ้าอยากทำภาพยนตร์ให้มีคนดูเยอะๆนั้นควรที่จะ ทำเป็นประเภท Dramas หรือ Comedies และอยู่ในเรท ของ Adults และ Teens เพราะเข้าถึงได้ง่าย