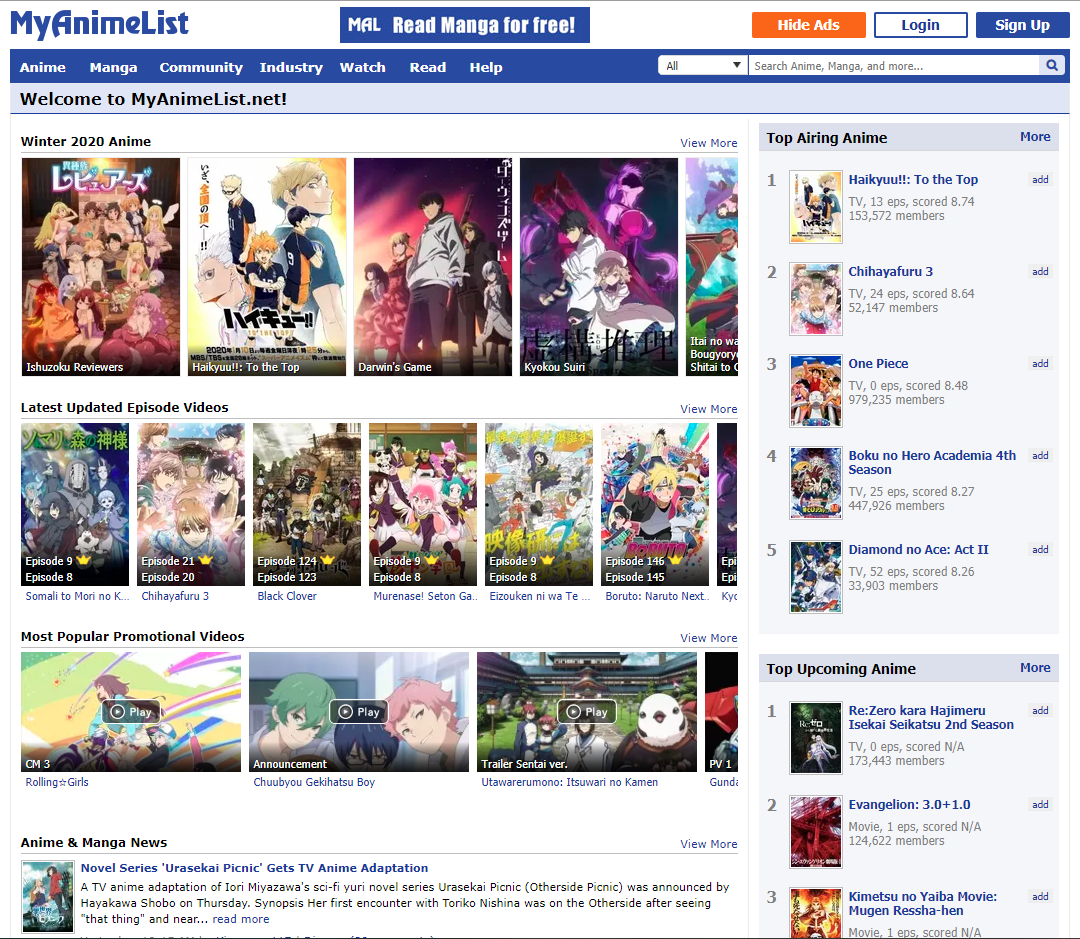
**Anime Analysis**

**สมาชิกกลุ่ม**6110400653 ธนกฤต ญาณนาม   
6110400823 นิธิวัฒน์ อภัยกาวี   
6110400831 ปิยทัช สุลา   
6110400840 วัทธิกร เฟื่องฟู  
6110405949 จิตติณณ์ จินดานรเศรษฐ์

**แหล่งที่มาของข้อมูล**

* Dataset: [*https://www.kaggle.com/azathoth42/myanimelist#anime\_cleaned.csv*](https://www.kaggle.com/azathoth42/myanimelist#anime_cleaned.csv)
* ข้อมูลใน Dataset ถูกเก็บมาจากเว็บไซต์ MyAnimeList *(*[*https://myanimelist.net*](https://myanimelist.net/)*)* ซึ่งเป็นเว็บไซต์ที่ทำการเก็บรวบรวมข้อมูลรายละเอียดของอนิเมะแต่ละเรื่อง เช่น ชื่อเรื่อง สตูดิโอผู้สร้าง   
  ฤดูกาลฉาย หมวดหมู่ เพลงประกอบ รายชื่อตัวละคร นักพากย์ ฯลฯ

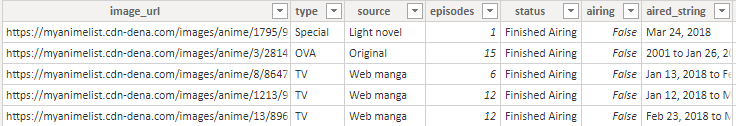


**ไฟล์ anime\_cleaned.csv** ประกอบด้วยข้อมูลจำนวน 33 คอลัมน์ 6668 แถว

**คำอธิบายคอลัมน์**



* anime\_id: Key ของอนิเมะแต่ละเรื่อง เป็นข้อมูลประเภท Integer ที่สามารถอ้างอิงบนเว็บไซต์ได้ เช่น [*https://myanimelist.net/anime/37505/*](https://myanimelist.net/anime/37505/)
* title: ชื่อเรื่องอนิเมะ เป็นข้อมูลประเภท String
* title\_english: ชื่อเรื่องอนิเมะเป็นภาษาอังกฤษ เป็นข้อมูลประเภท String และมีข้อมูลว่าง
* title\_japanese: ชื่อเรื่องอนิเมะเป็นภาษาญี่ปุ่น เป็นข้อมูลประเภท String และมีข้อมูลว่าง
* title\_synonyms: ชื่อเรื่องอนิเมะที่เป็นที่รู้จักกันโดยทั่วไป เป็นข้อมูลแบบ String และมีข้อมูลว่าง



* image\_url: ที่อยู่ของภาพตัวอย่าง หรือหน้าปกของอนิเมะแต่ละเรื่องบนเว็บไซต์ เป็นข้อมูลประเภท String
* image\_url: ที่อยู่ของภาพตัวอย่าง หรือหน้าปกของอนิเมะแต่ละเรื่องบนเว็บไซต์ เป็นข้อมูลประเภท String
* type: ประเภทการฉาย เป็นข้อมูลประเภท String มี 6 ประเภท ดังนี้  
  TV คือ ออกฉายทางโทรทัศน์

OVA คือ ไม่ได้มีการฉายออกอากาศทางทีวีหรือในโรงภาพยนตร์ แต่มีจำหน่ายเป็นแผ่น

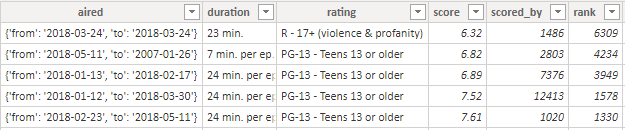
ONA คือ ออกฉายทางอินเตอร์เน็ต

Special คือ จัดทำขึ้นมาในวาระพิเศษ

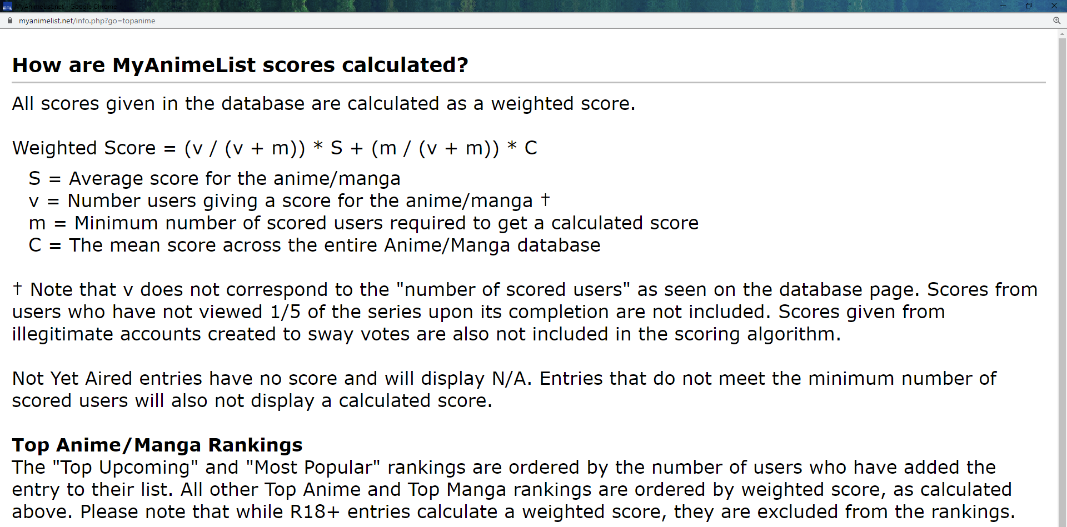
Movie คือ ออกฉายในโรงภาพยนตร์

Music คือ ออกฉายเป็นวิดีโอเพลง

* source: ต้นฉบับเนื้อหาที่มาของอนิเมะแต่ละเรื่อง เป็นข้อมูลประเภท String
* episodes: จำนวนตอนทั้งหมดของอนิเมะแต่ละเรื่องในเว็บไซต์ เป็นข้อมูลประเภท Integer
* status: สถานะการฉายของอนิเมะ เป็นข้อมูลประเภท String
* airing: แสดงว่ากำลังฉายอยู่หรือไม่ เป็นข้อมูลประเภท Boolean
* aired\_string: ช่วงระยะเวลาที่ฉาย เป็นข้อมูลประเภท String



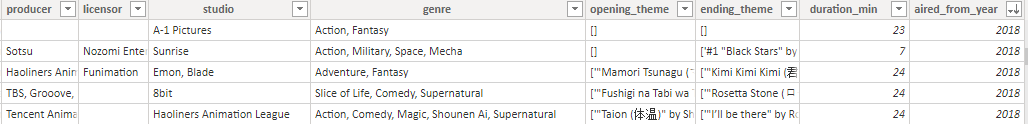
* aired: ช่วงระยะเวลาฉาย เป็นข้อมูลประเภท String
* duration: ความยาวของอนิเมะที่ฉายต่อตอน เป็นข้อมูลประเภท String
* rating: กลุ่มอายุที่เหมาะสมของอนิเมะแต่ละเรื่อง เป็นข้อมูลประเภท String
* score: คะแนนของอนิเมะแต่ละเรื่อง เป็นข้อมูลประเภท Float วิธีการคำนวณคะแนนของเว็บไซต์มีวิธีการดังนี้



* scored\_by: จำนวนคนที่ลงคะแนน เป็นข้อมูลประเภท Integer
* rank: อันดับของอนิเมะบนหน้าเว็บไซต์ เป็นข้อมูลประเภท Integer

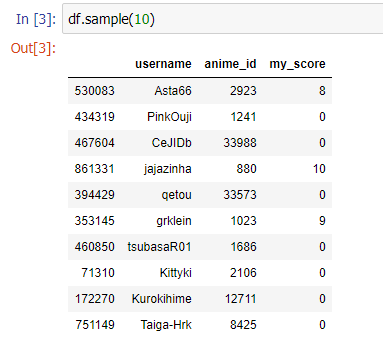


* popularity: อันดับความนิยมบนหน้าเว็บ เป็นข้อมูลประเภท Integer
* members: ยอดจำนวนคนดูอนิเมะแต่ละเรื่องในเว็บไซต์ เป็นข้อมูลประเภท Integer
* favorites: จำนวนคนที่ชอบอนิเมะแต่ละเรื่องบนหน้าเว็บไซต์ เป็นข้อมูลประเภท Integer
* background: ประวัติความเป็นมา เป็นข้อมูลประเภท String
* premiered: ฤดู และ ปีที่ทำการฉาย เป็นข้อมูลประเภท String
* broadcast: วันเวลาออกอากาศ เป็นข้อมูลประเภท String
* related: เชื่อมโยงข้อมูล เป็นข้อมูลประเภท String



* producer: ผู้อำนวยการผลิต เป็นข้อมูลประเภท String
* licensor: ผู้ถือลิขสิทธิ์ เป็นข้อมูลประเภท String
* studio: สตูดิโอที่จัดทำอนิเมะ เป็นข้อมูลประเภท String
* genre: ประเภทเนื้อหาของอนิเมะ เป็นข้อมูลประเภท String
* opening\_theme: ชื่อเพลงเปิดที่ใช้ประกอบในอนิเมะ เป็นข้อมูลประเภท String
* ending\_theme: ชื่อเพลงปิดที่ใช้ประกอบในอนิเมะ เป็นข้อมูลประเภท String
* duration\_min: ความยาวของอนิเมะที่ฉายต่อตอน (นาที) เป็นข้อมูลประเภท Float
* aired\_from\_year: ปีที่เริ่มฉาย เป็นข้อมูลประเภท Integer

**ไฟล์ userAmineList-Drop.csv** ประกอบด้วยข้อมูลจำนวน 3 คอลัมน์ 1,048,575 แถว เป็นไฟล์ที่ดัดแปลงมาจากไฟล์ userAnimeList.csv

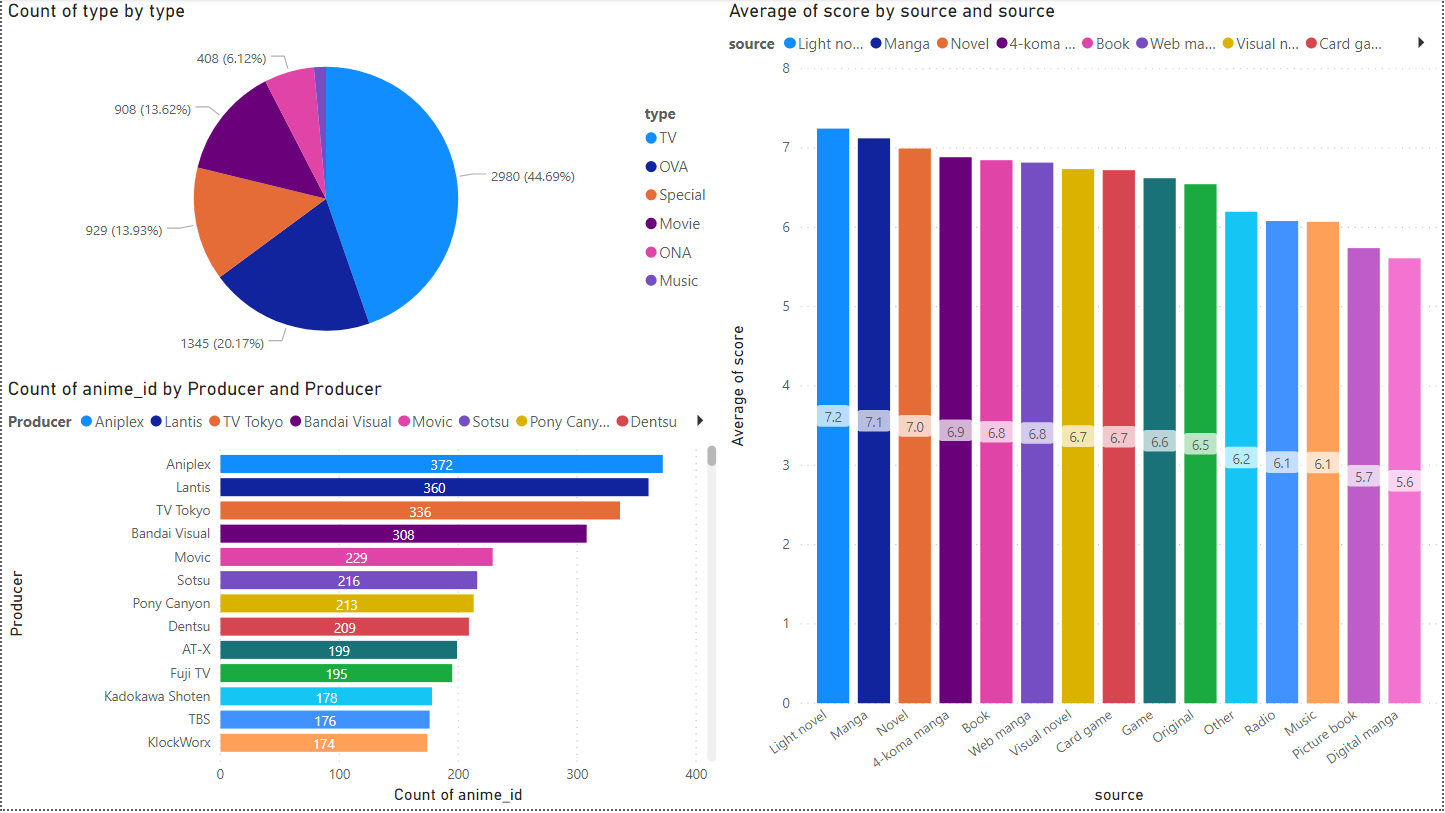
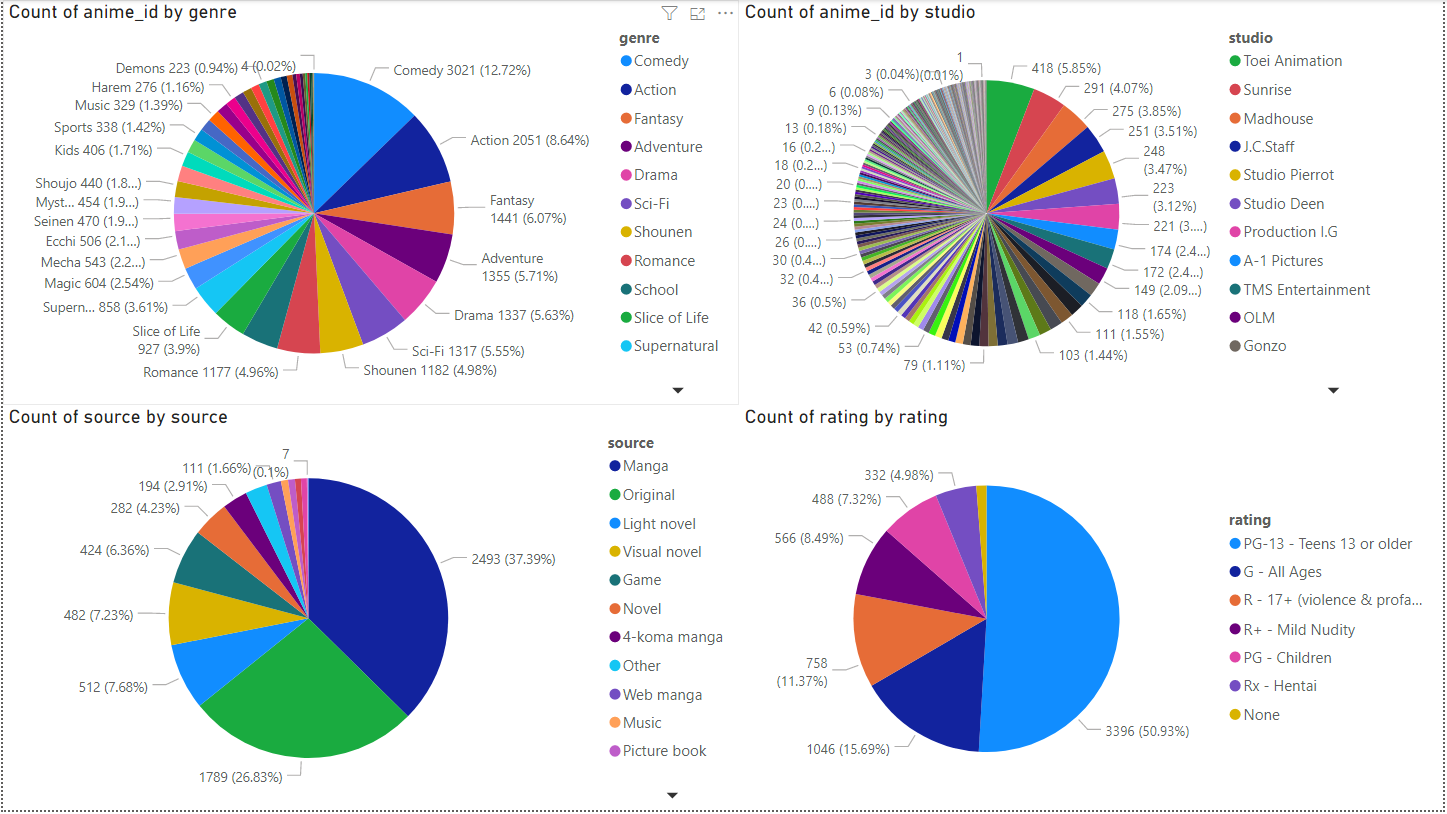


* anime\_id: Key ของอนิเมะแต่ละเรื่อง เป็นข้อมูลประเภท Integer ที่สามารถอ้างอิงบนเว็บไซต์ได้ เช่น [*https://myanimelist.net/anime/37505/*](https://myanimelist.net/anime/37505/)
* username: ชื่อสมาชิกผู้ใช้งานในเว็บไซต์ เป็นข้อมูลประเภท string
* my\_score: คะแนนที่สมาชิกแต่ละคนลงให้ในแต่ละเรื่อง โดยคะแนนจะมีตั้งแต่ 0 ถึง 10 เป็นข้อมูลประเภท Integer

**สาเหตุที่มาของการเลือกข้อมูลชุดนี้**

เนื่องจากสมาชิกภายในกลุ่มชอบดูอนิเมะ จึงอยากทำการวิเคราะห์ข้อมูลต่าง ๆ ของอนิเมะ ไม่ว่าจะเป็นความนิยมของผู้ชม ความนิยมของผู้สร้าง ความหลากหลายของเนื้อหา ฯลฯ

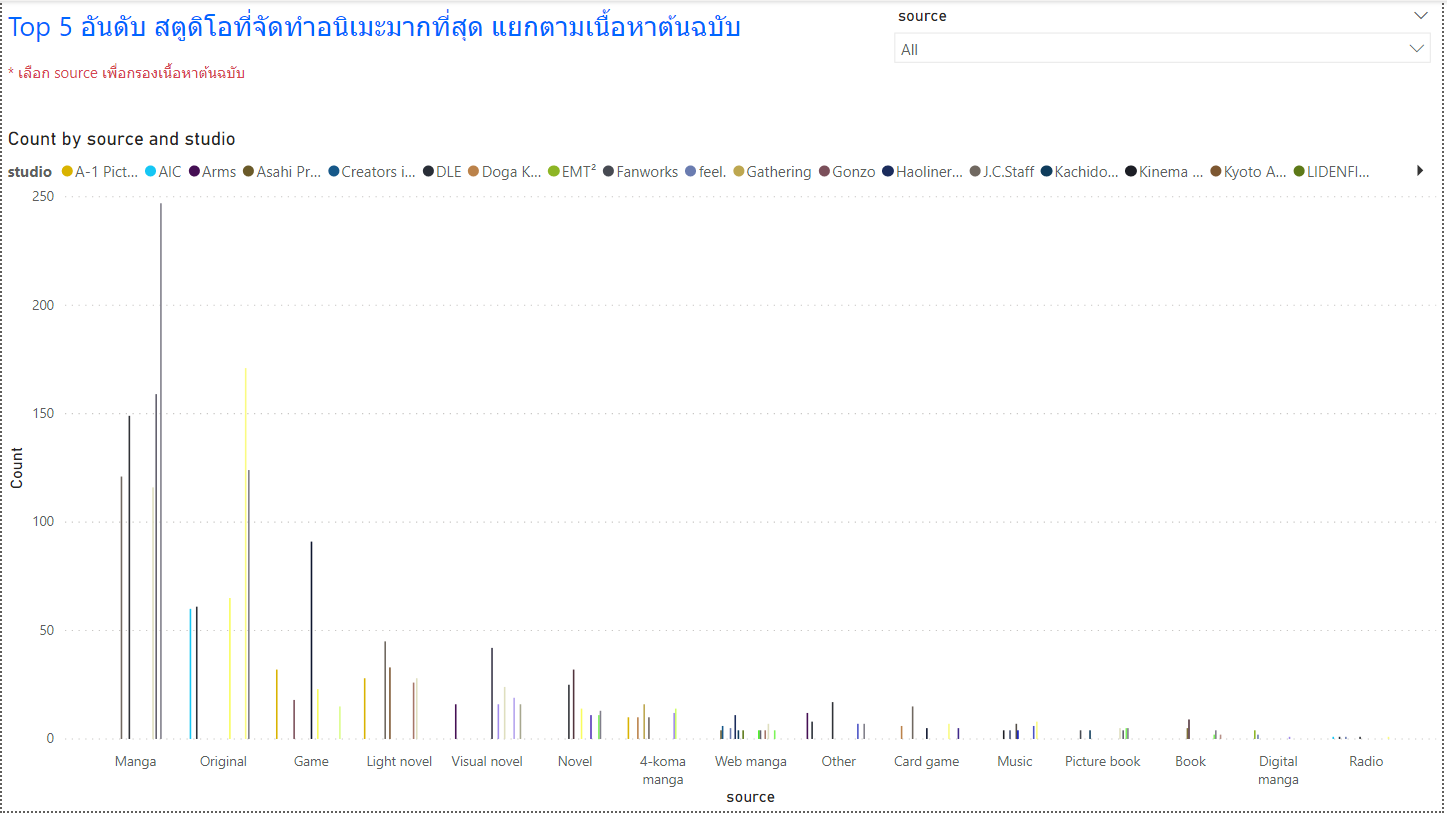
**สิ่งที่น่าสนใจจากการวิเคราะห์ข้อมูลเบื้องต้น**



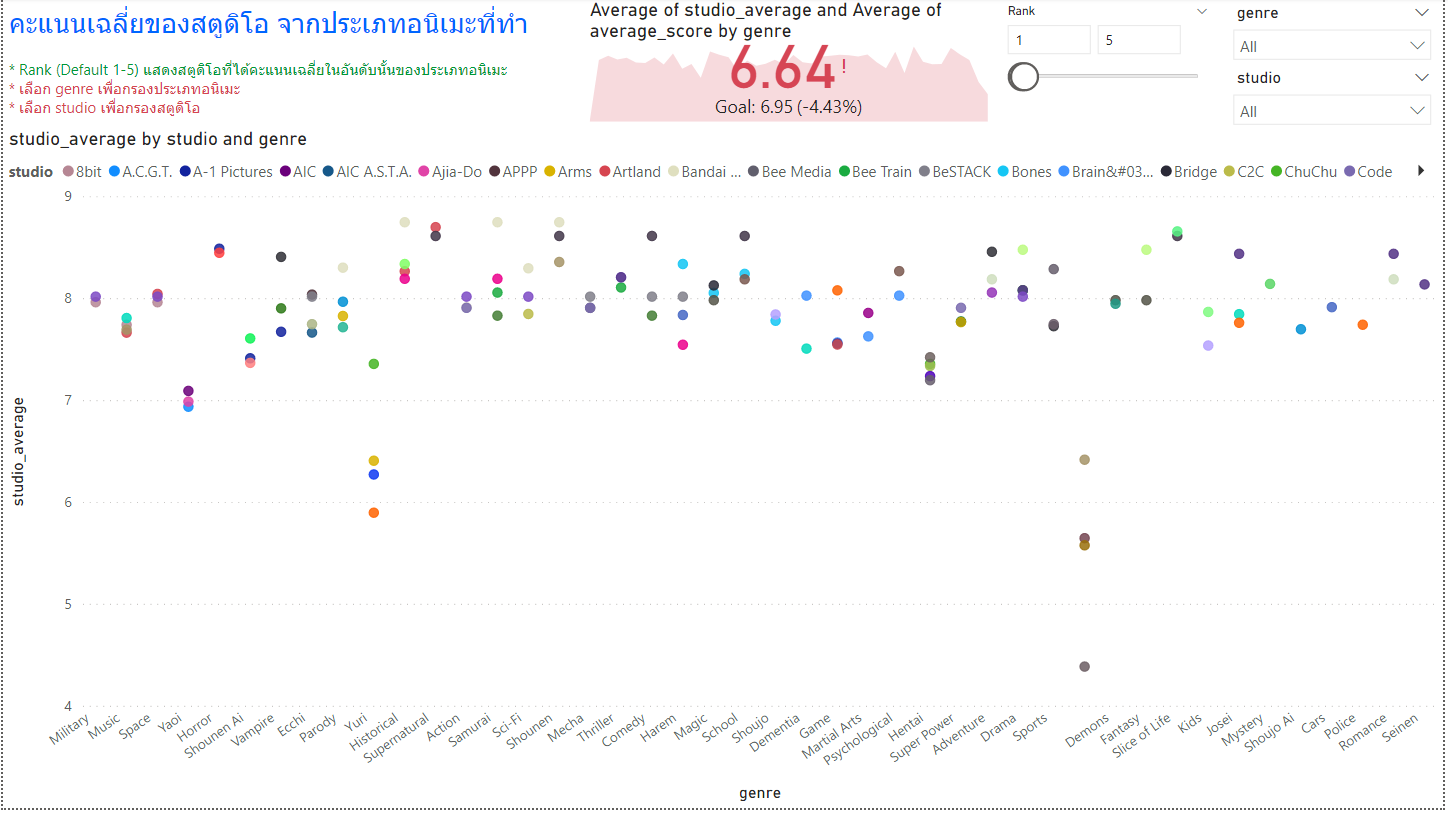
**โจทย์ที่ต้องการหาองค์ความรู้เชิงลึก**

1. สตูดิโอจัดทำอะไรเป็นหลัก และผลงานออกมาเป็นอย่างไร
   1. เนื้อหาต้นฉบับแต่ละประเภท สตูดิโอไหนเป็นผู้จัดทำเป็นส่วนใหญ่

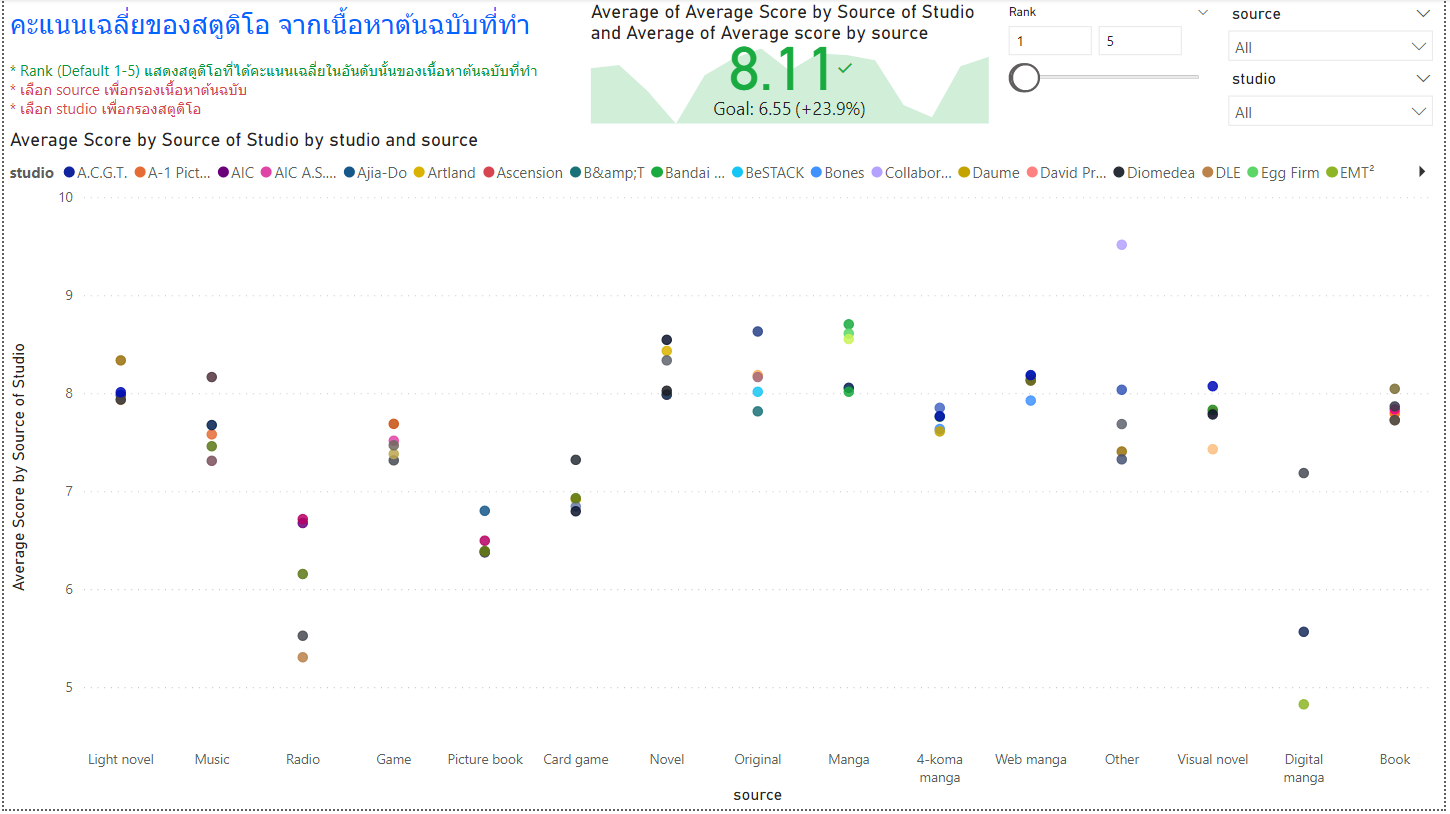
* **สิ่งที่น่าสนใจ:** ทำให้รู้ว่าสตูดิโอไหนมีศักยภาพในการผลิตผลงานเป็นจำนวนมาก หรือถือครองลิขสิทธิ์เป็นจำนวนมาก ในแต่ละเนื้อหาต้นฉบับ
* **กระบวนการ:** นำ source จากไฟล์ anime\_cleaned มา group by ร่วมกับ studio เพื่อทำการนับจำนวนอนิเมะที่จัดทำ และจัด rank ของ studio ภายในแต่ละ source
* **ผลลัพธ์:** ได้รู้ว่าสตูดิโอไหนมักเป็นผู้จัดทำเนื้อหาต้นฉบับประเภทนั้น ซึ่งจากกราฟที่ได้เราจะพบว่าบางสตูดิโอนั้น มักผลิตเนื้อหาต้นฉบับบางประเภทอย่างเยอะเป็นพิเศษ เนื่องจากมีความเกี่ยวข้องกับบริษัทที่ผลิตของเล่น การ์ตูน เกม ในประเทศญี่ปุ่น หรือมีความชำนาญในเนื้อหาต้นฉบับประเภทนั้น



* 1. คะแนนเฉลี่ยของสตูดิโอ จากประเภทอนิเมะที่จัดทำ
* **สิ่งที่น่าสนใจ:** ทำให้รู้ว่าสตูดิโอผลิตผลงานที่มีคุณภาพ คะแนนเฉลี่ยสูง ในแต่ละประเภทอนิเมะ ซึ่งหากจะผลิตอนิเมะสักเรื่อง สามารถดูประเภทของเนื้อหาอนิเมะที่ต้องการผลิต นำมาใช้ในการเลือกสตูดิโอที่เหมาะสำหรับการจัดทำได้
* **กระบวนการ:** นำ genre จากไฟล์ anime\_cleaned มา group by ร่วมกับ studio เพื่อทำการหา score เฉลี่ยของอนิเมะที่จัดทำ และจัด rank ของ studio ภายในแต่ละ genre
* **ผลลัพธ์:** ได้รู้ว่าสตูดิโอไหนมีศักยภาพในการจัดทำอนิเมะประเภทนั้น ซึ่งจากกราฟที่ได้เราจะพบว่าแต่ละสตูดิโอ มักมีความถนัดในแต่ละประเภทที่ไม่เหมือนกัน

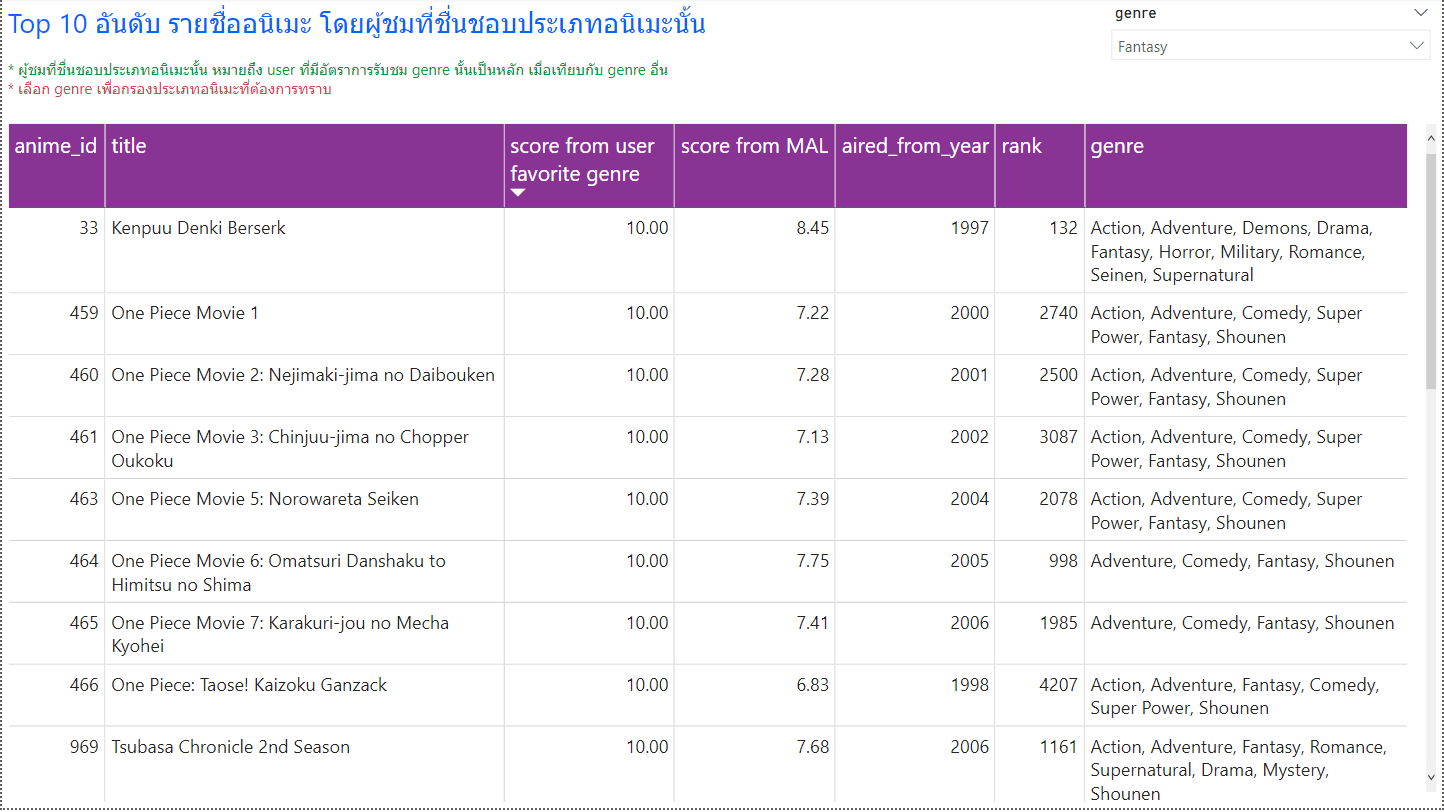


* 1. คะแนนเฉลี่ยของสตูดิโอ จากเนื้อหาต้นฉบับที่จัดทำ
* **สิ่งที่น่าสนใจ:** ทำให้รู้ว่าสตูดิโอผลิตผลงานที่มีคุณภาพ คะแนนเฉลี่ยสูง ในแต่ละประเภทเนื้อหาต้นฉบับอนิเมะ ซึ่งหากจะผลิตอนิเมะสักเรื่อง สามารถดูประเภทของเนื้อหาต้นฉบับอนิเมะที่ต้องการผลิต นำมาใช้ในการเลือกสตูดิโอที่เหมาะสำหรับการจัดทำได้
* **กระบวนการ:** นำ source จากไฟล์ anime\_cleaned มา group by ร่วมกับ studio เพื่อทำการหา score เฉลี่ยของอนิเมะที่จัดทำ และจัด rank ของ studio ภายในแต่ละ source
* **ผลลัพธ์:** ได้รู้ว่าสตูดิโอไหนมีศักยภาพในการจัดทำอนิเมะ จากเนื้อหาต้นฉบับประเภทนั้น ซึ่งจากกราฟที่ได้เราจะพบว่ามีบางสตูดิโอปรากฏชื่อในหลายประเภทเนื้อหาต้นฉบับ ที่มีความใกล้เคียงกัน เช่น Novel, Light novel, Visual novel



1. รายชื่ออนิเมะ 10 อันดับแรก เรียงตามคะแนนจากมากไปน้อย โดยสมาชิกเว็บไซต์ที่ชื่นชอบอนิเมะประเภทนั้น ซึ่งมีสัดส่วนการดูอนิเมะประเภทนั้นเป็นหลัก

* **สิ่งที่น่าสนใจ:** หากเราวิเคราะห์กันแค่เฉพาะกลุ่มผู้ชมที่ดูอนิเมะประเภทนั้นเป็นหลัก นำมาลงคะแนนให้อนิเมะประเภทนั้น ผลลัพธ์ที่ได้นั้นจะน่าเชื่อถือขนาดไหน อนิเมะที่ติดอันดับนั้น จะเป็นอนิเมะที่ยอดเยี่ยมใช่หรือไม่
* **กระบวนการ:** ทำการวิเคราะห์ username จากไฟล์ UserAnimeList-Drop เทียบ anime\_id ว่าเป็น genre อะไรในไฟล์ anime\_cleaned เพื่อหาสัดส่วนว่า username นั้นชอบดูอนิเมะ genre ไหนเป็นหลัก หลังจากทราบแล้ว จึงนำคะแนน my\_score ที่ username นั้นลงคนแนนให้ anime\_id นั้นมาคำนวณ หาก anime\_id นั้น genre ตรงกับสัดส่วน genre หลักที่ username นั้นรับชม ทำการหาคะแนนเฉลี่ย และจัดอันดับของคะแนนภายในแต่ละ genre
* **ผลลัพธ์:** แต่ละ genre ส่วนใหญ่ คะแนนไม่สามารถอ้างอิงกับความเป็นจริง ที่กลุ่มผู้ชมทั้งหมดช่วยกันลงคะแนนให้ได้ เมื่อเทียบ score from user favorite genre กับ score from MAL เราจะเห็นว่าอนิเมะหลายเรื่องนั้นได้คะแนนไม่สมเหตุสมผล มีทั้งอวยจนให้เต็ม 10 คะแนน หรือไม่ชอบจนกดเหลือ 0 คะแนน และเมื่อดู rank ที่ถูกจัดบนเว็บไซต์เทียบด้วย จะยิ่งเห็นว่าหลายเรื่องไม่ใช่อนิเมะที่ได้รับความนิยมอันดับต้น ๆ ของเว็บไซต์



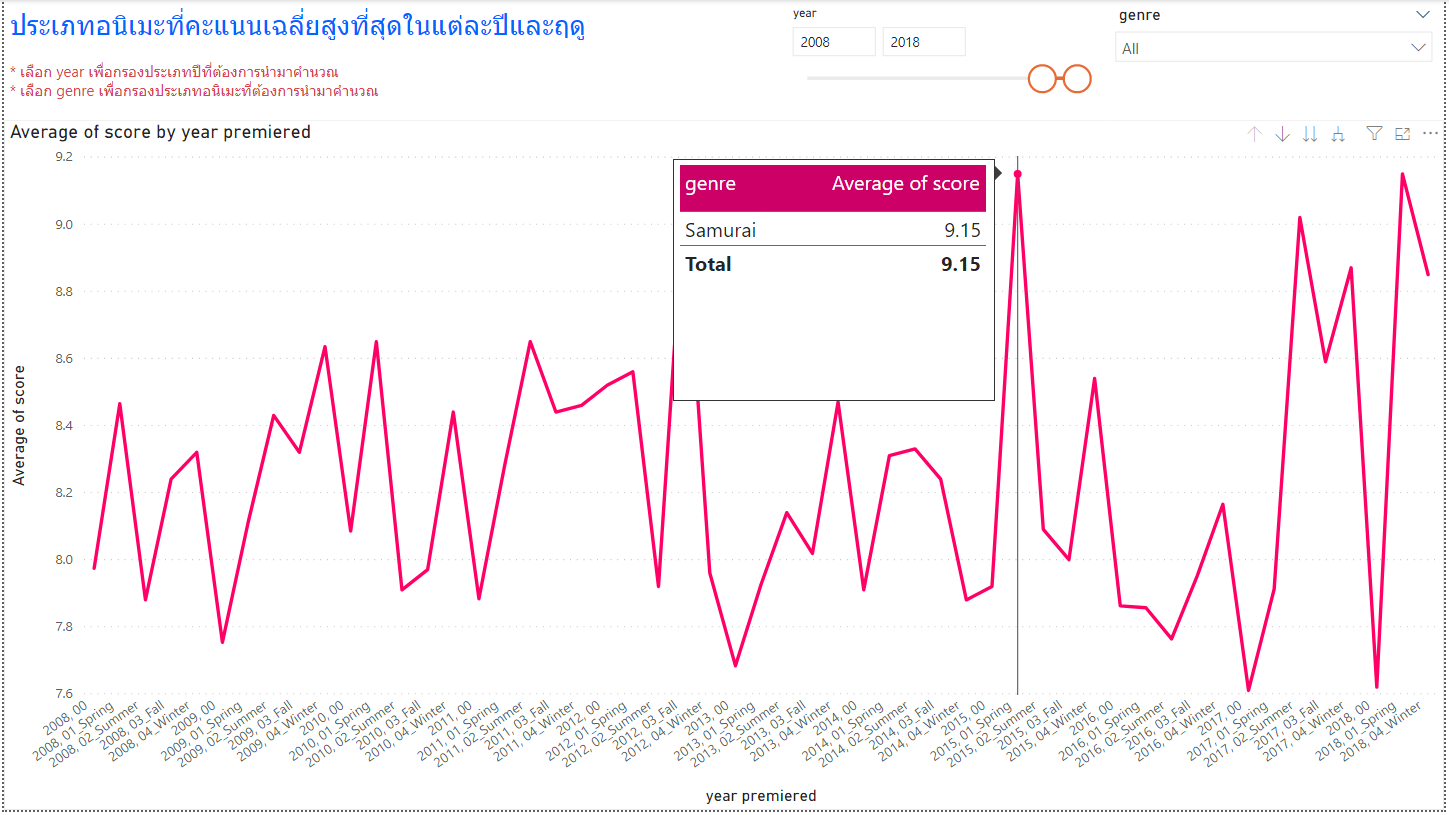
1. ประเภทอนิเมะ ปีและฤดูกาล

3.1 ประเภทอนิเมะที่คะแนนเฉลี่ยสูงที่สุดในแต่ละปีและฤดูกาล

* **สิ่งที่น่าสนใจ:** ทำให้รู้ว่าในแต่ละปีและฤดูกาล อนิเมะประเภทไหนได้รับความนิยมสูง โดยดูจากคะแนนเฉลี่ย ซึ่งสามารถนำมากำหนดแผนการฉายอนิเมะได้ ในช่วงเวลาที่ใกล้เคียงกัน ควรจะมีการเปลี่ยนประเภทของอนิเมะ เพื่อให้ผู้ชมไม่เบื่อหน่ายในการรับชม
* **กระบวนการ:** ทำการวิเคราะห์ค่าเฉลี่ยของ score ในแต่ละ genre จากไฟล์ anime\_cleaned โดยการ group by หาค่าเฉลี่ยในแต่ละปีและฤดูกาลออกมา จากนั้นแสดง genre ที่มีค่าเฉลี่ยของ score สูงที่สุด
* **ผลลัพธ์:** ทำให้เราทราบถึงความนิยมของประเภทของอนิเมะในแต่ละปีหรือฤดูกาล ซึ่งจากการดูไทม์ไลน์แล้วจะพบว่า ในช่วงเวลาที่ติดกัน ประเภทอนิเมะที่ได้รับความนิยมมักจะต่างกัน  
  + 1. จำแนกตามปี



* + 1. จำแนกตามปีและฤดู



* 1. คะแนนเฉลี่ยอนิเมะในแต่ละปีและฤดู และประเภทอนิเมะที่นิยมจัดทำมากสุด
* **สิ่งที่น่าสนใจ:** ทำให้รู้ว่าในแต่ละปีและฤดูกาล มีค่าคะแนนเฉลี่ยเป็นอย่างไร และในแต่ละปีและฤดูกาลอนิเมะแนวไหนที่มีจำนวนการจัดทำสูงสุด เพื่อเป็นแนวทางในการตัดสินใจว่าจะสร้างอนิเมะเรื่องถัดไปให้เป็นประเภทไหน
* **กระบวนการ:** นำข้อมูลจาก anime\_cleaned มาแยกตามปีและฤดูกาล จากนั้นก็ทำการแยก genre ออกมาเพื่อหาว่า genre ไหนในแต่ละปีและฤดู ที่มีจำนวนมากสุด และนำ score มาคิดเฉลี่ยตามปีและฤดู
* **ผลลัพธ์:** ทำให้เราทราบถึงว่าในช่วงปีและฤดูที่ใกล้เคียง มีอนิเมะประเภทไหนที่มีมีการจัดทำเยอะสุด จากข้อมูลยังพบว่าในช่วงปีหลัง ๆ อนิเมะประเภท Comedy มีจำนวนมากติดต่อกันหลายเรื่อง หลายปี และหลายฤดู

