Anime Analysis

สมาชิกกลุ่ม

6110400653 ธนกฤต ญาณนาม

6110400823 นิธิวัฒน์ อภัยกาวี่

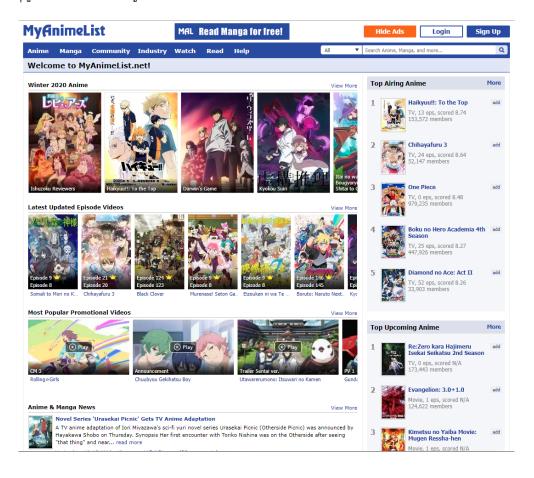
6110400831 ปียทัช สุลา

6110400840 วัทธิกร เฟื่องฟู

6110405949 จิตติณณ์ จินดานรเศรษฐ์

แหล่งที่มาของข้อมูล

- Dataset: https://www.kaggle.com/azathoth42/myanimelist#anime_cleaned.csv
- ข้อมูลใน Dataset ถูกเก็บมาจากเว็บไซต์ MyAnimeList (https://myanimelist.net) ซึ่งเป็นเว็บไซต์ ที่ทำการเก็บรวบรวมข้อมูลรายละเอียดของอนิเมะแต่ละเรื่อง เช่น ชื่อเรื่อง สตูดิโอผู้สร้าง ฤดูกาลฉาย หมวดหมู่ เพลงประกอบ รายชื่อตัวละคร นักพากย์ ฯลฯ



ไฟล์ anime_cleaned.csv ประกอบด้วยข้อมูลจำนวน 33 คอลัมน์ 6668 แถว

คำอธิบายคอลัมน์

anime_id 🔻	title	title_english 🔻	title_japanese 🔻	title_synonyms 🔻
37505	Grancrest Senki Tsuisou	Record of Grancre	グランクレスト戦記 追想	Grancrest Senki Episc
3288	Gundam Evolve	Gundam Evolve	ガンダムEVOLVE	Mobile Suit Gundam
35757	Gin no Guardian 2nd Season	The Silver Guardia	銀の墓守り 第2期	Shirogane no Guardia
35828	Miira no Kaikata	How to Keep a Mu	ミイラの飼い方	How to keep a mumr
36522	Ling Qi 2	Spiritpact 2	灵契 2	Spirit Pact: Yomi no (

- anime_id: Key ของอนิเมะแต่ละเรื่อง เป็นข้อมูลประเภท Integer ที่สามารถอ้างอิงบนเว็บไซต์ได้ เช่น 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	type 💌	source 🔻	episodes 💌	status 🔻	airing 🔻	aired_string 🔻
https://myanimelist.cdn-dena.com/images/anime/1795/9	Special	Light novel	1	Finished Airing	False	Mar 24, 2018
https://myanimelist.cdn-dena.com/images/anime/3/2814	OVA	Original	15	Finished Airing	False	2001 to Jan 26, 20
https://myanimelist.cdn-dena.com/images/anime/8/8647	TV	Web manga	6	Finished Airing	False	Jan 13, 2018 to F∈
https://myanimelist.cdn-dena.com/images/anime/1213/9	TV	Web manga	12	Finished Airing	False	Jan 12, 2018 to N
https://myanimelist.cdn-dena.com/images/anime/13/896	TV	Web manga	12	Finished Airing	False	Feb 23, 2018 to N

- image_url: ที่อยู่ของภาพตัวอย่าง หรือหน้าปกของอนิเมะแต่ละเรื่องบนเว็บไซต์ เป็นข้อมูลประเภท String
- image_url: ที่อยู่ของภาพตัวอย่าง หรือหน้าปกของอนิเมะแต่ละเรื่องบนเว็บไซต์ เป็นข้อมูล
 ประเภท 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 🔻	duration 🔻	rating	score 🔻	scored_by 🔻	rank 💌
{'from': '2018-03-24', 'to': '2018-03-24'}	23 min.	R - 17+ (violence & profanity)	6.32	1486	6309
{'from': '2018-05-11', 'to': '2007-01-26'}	7 min. per ep.	PG-13 - Teens 13 or older	6.82	2803	4234
{'from': '2018-01-13', 'to': '2018-02-17'}	24 min. per ep	PG-13 - Teens 13 or older	6.89	7376	3949
{'from': '2018-01-12', 'to': '2018-03-30'}	24 min. per ep	PG-13 - Teens 13 or older	7.52	12413	1578
{'from': '2018-02-23', 'to': '2018-05-11'}	24 min. per ep	PG-13 - Teens 13 or older	7.61	1020	1330

- aired: ช่วงระยะเวลาฉาย เป็นข้อมูลประเภท String
- duration: ความยาวของอนิเมะที่ฉายต่อตอน เป็นข้อมูลประเภท String
- rating: กลุ่มอายุที่เหมาะสมของอนิเมะแต่ละเรื่อง เป็นข้อมูลประเภท String
- score: คะแนนของอนิเมะแต่ละเรื่อง เป็นข้อมูลประเภท Float วิธีการคำนวณคะแนนของเว็บไซต์
 มีวิธีการดังนี้

How are MyAnimeList scores calculated? All scores given in the database are calculated as a weighted score. Weighted Score = (v / (v + m)) * S + (m / (v + m)) * C S = Average score for the anime/manga v = Number users giving a score for the anime/manga † m = Minimum number of scored users required to get a calculated score C = The mean score across the entire Anime/Manga database † Note that v does not correspond to the "number of scored users" as seen on the database page. Scores from users who have not viewed 1/5 of the series upon its completion are not included. Scores given from illegitimate accounts created to sway votes are also not included in the scoring algorithm. Not Yet Aired entries have no score and will display N/A. Entries that do not meet the minimum number of scored users will also not display a calculated score. Top Anime/Manga Rankings The "Top Upcoming" and "Most Popular" rankings are ordered by the number of users who have added the entry to their list. All other Top Anime and Top Manga rankings are ordered by weighted score, as calculated above. Please note that while R18+ entries calculate a weighted score, they are excluded from the rankings.

- scored_by: จำนวนคนที่ลงคะแนน เป็นข้อมูลประเภท Integer
- rank: อันดับของอนิเมะบนหน้าเว็บไซต์ เป็นข้อมูลประเภท Integer

popularity 🔻	members 🔻	favorites 🔻	background 🔻	premiered 🔻	broadcast 💌	related
5120	4442	5				{'Full story': [{'mal_id': 34279, 'type': 'anime', 'url': 'https:/,
4224	7233	16				$\label{eq:continuous} \mbox{\ensuremath{\texttt{'Parent story':}} \texttt{[\{'mal_id': 80, 'type': 'anime', 'url': 'https://anime', 'url': 'https://a$
2368	24865	57		Winter 2018	Saturdays at 21	$\label{eq:continuous} \mbox{\em 'Prequel': [{'mal_id': 32936, 'type': 'anime', 'url': 'https://r} \em between the property of the prop$
1790	39616	166		Winter 2018	Fridays at 01:58	{'Adaptation': [{'mal_id': 99256, 'type': 'manga', 'url': 'http:
3861	8970	51		Winter 2018	Saturdays at 21	$\label{eq:continuous} \mbox{ 'Prequel': [{'mal_id': 33605, 'type': 'anime', 'url': 'https://r} \mbox$

- 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 💌	licensor 💌	studio	genre	opening_theme	ending_theme 💌	duration_min 🔻	aired_from_year 荰
		A-1 Pictures	Action, Fantasy	0		23	2018
Sotsu	Nozomi Enter	Sunrise	Action, Military, Space, Mecha	0	['#1 "Black Stars" by	7	2018
Haoliners Anin	Funimation	Emon, Blade	Adventure, Fantasy	["'Mamori Tsunagu (["Kimi Kimi Kimi (君	24	2018
TBS, Grooove,		8bit	Slice of Life, Comedy, Supernatural	["Fushigi na Tabi wa	["Rosetta Stone (🗆	24	2018
Tencent Anima		Haoliners Animation League	Action, Comedy, Magic, Shounen Ai, Supernatural	["Taion (体温)" by Sh	["I'll be there" by Ro	24	2018

- 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 เป็นไฟล์ที่ดัดแปลง มาจากไฟล์ userAnimeList.csv

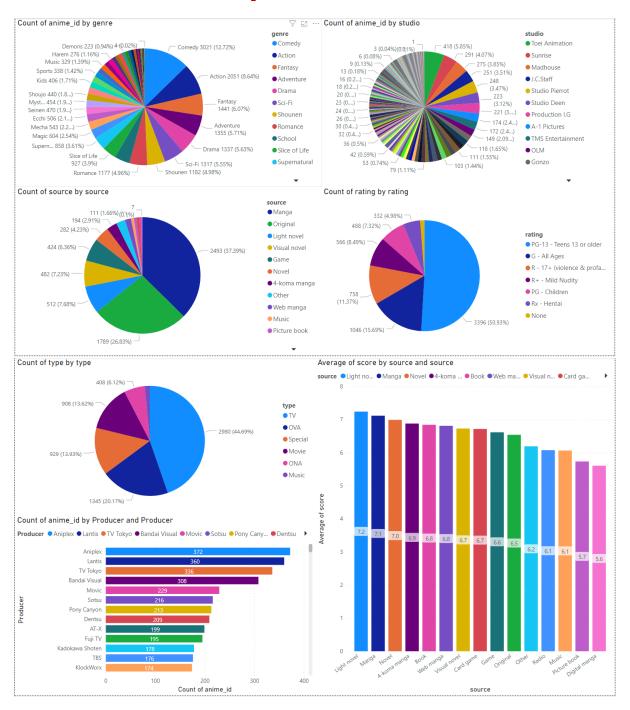
In [3]:	df.sampl	e(10)		
Out[3]:		username	anime_id	my_score
	530083	Asta66	2923	8
	434319	PinkOuji	1241	0
	467604	CeJIDb	33988	0
	861331	jajazinha	880	10
	394429	qetou	33573	0
	353145	grklein	1023	9
	460850	tsubasaR01	1686	0
	71310	Kittyki	2106	0
	172270	Kurokihime	12711	0
	751149	Taiga-Hrk	8425	0

- anime_id: Key ของอนิเมะแต่ละเรื่อง เป็นข้อมูลประเภท Integer ที่สามารถอ้างอิงบนเว็บไซต์ได้ เช่น <u>https://myanimelist.net/anime/37505/</u>
- username: ชื่อสมาชิกผู้ใช้งานในเว็บไซต์ เป็นข้อมูลประเภท string
- my_score: คะแนนที่สมาชิกแต่ละคนลงให้ในแต่ละเรื่อง โดยคะแนนจะมีตั้งแต่ 0 ถึง 10 เป็น ข้อมูลประเภท Integer

สาเหตุที่มาของการเลือกข้อมูลชุดนี้

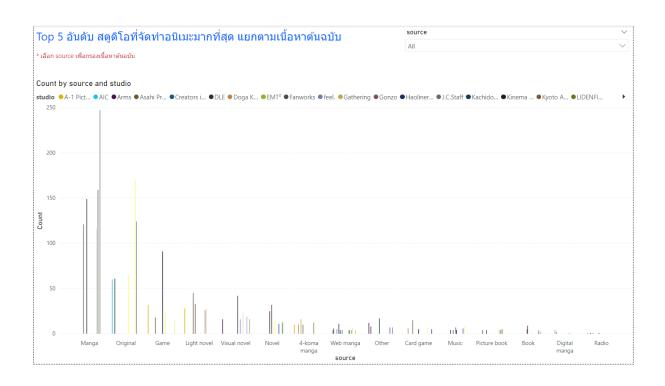
เนื่องจากสมาชิกภายในกลุ่มชอบดูอนิเมะ จึงอยากทำการวิเคราะห์ข้อมูลต่าง ๆ ของอนิเมะ ไม่ว่า จะเป็นความนิยมของผู้ชม ความนิยมของผู้สร้าง ความหลากหลายของเนื้อหา ฯลฯ

สิ่งที่น่าสนใจจากการวิเคราะห์ข้อมูลเบื้องต้น



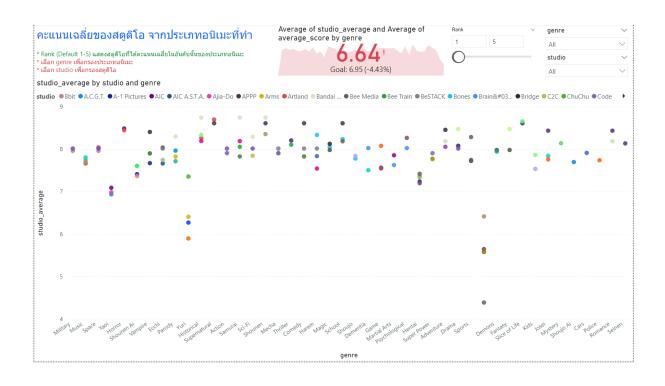
ใจทย์ที่ต้องการหาองค์ความรู้เชิงลึก

- 1. สตูดิโอจัดทำอะไรเป็นหลัก และผลงานออกมาเป็นอย่างไร
 - 1.1 เนื้อหาต้นฉบับแต่ละประเภท สตูดิโอไหนเป็นผู้จัดทำเป็นส่วนใหญ่
 - **สิ่งที่น่าสนใจ**: ทำให้รู้ว่าสตูดิโอไหนมีศักยภาพในการผลิตผลงานเป็นจำนวนมาก หรือถือ ครองลิขสิทธิ์เป็นจำนวนมาก ในแต่ละเนื้อหาต้นฉบับ
 - **กระบวนการ**: น้ำ source จากไฟล์ anime_cleaned มา group by ร่วมกับ studio เพื่อ ทำการนับจำนวนอนิเมะที่จัดทำ และจัด rank ของ studio ภายในแต่ละ source
 - ผลลัพธ์: ได้รู้ว่าสตูดิโอไหนมักเป็นผู้จัดทำเนื้อหาต้นฉบับประเภทนั้น ซึ่งจากกราฟที่ได้เรา จะพบว่าบางสตูดิโอนั้น มักผลิตเนื้อหาต้นฉบับบางประเภทอย่างเยอะเป็นพิเศษ เนื่องจากมีความเกี่ยวข้องกับบริษัทที่ผลิตของเล่น การ์ตูน เกม ในประเทศญี่ปุ่น หรือมี ความชำนาญในเนื้อหาต้นฉบับประเภทนั้น



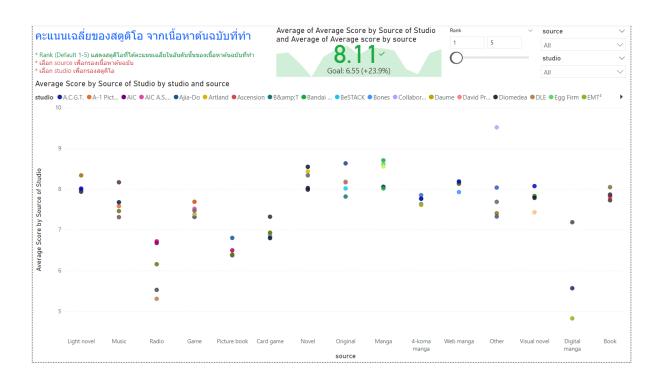
1.2 คะแนนเฉลี่ยของสตูดิโอ จากประเภทอนิเมะที่จัดทำ

- **สิ่งที่น่าสนใจ**: ทำให้รู้ว่าสตูดิโอผลิตผลงานที่มีคุณภาพ คะแนนเฉลี่ยสูง ในแต่ละ ประเภทอนิเมะ ซึ่งหากจะผลิตอนิเมะสักเรื่อง สามารถดูประเภทของเนื้อหาอนิเมะที่ ต้องการผลิต นำมาใช้ในการเลือกสตูดิโอที่เหมาะสำหรับการจัดทำได้
- **กระบวนการ**: นำ genre จากไฟล์ anime_cleaned มา group by ร่วมกับ studio เพื่อทำ การหา score เฉลี่ยของอนิเมะที่จัดทำ และจัด rank ของ studio ภายในแต่ละ genre
- ผลลัพธ์: ได้รู้ว่าสตูดิโอไหนมีศักยภาพในการจัดทำอนิเมะประเภทนั้น ซึ่งจากกราฟที่ได้ เราจะพบว่าแต่ละสตูดิโอ มักมีความถนัดในแต่ละประเภทที่ไม่เหมือนกัน



1.3 คะแนนเฉลี่ยของสตูดิโอ จากเนื้อหาต้นฉบับที่จัดทำ

- **สิ่งที่น่าสนใจ**: ทำให้รู้ว่าสตูดิโอผลิตผลงานที่มีคุณภาพ คะแนนเฉลี่ยสูง ในแต่ละประเภท เนื้อหาต้นฉบับอนิเมะ ซึ่งหากจะผลิตอนิเมะสักเรื่อง สามารถดูประเภทของเนื้อหา ต้นฉบับอนิเมะที่ต้องการผลิต นำมาใช้ในการเลือกสตูดิโอที่เหมาะสำหรับการจัดทำได้
- **กระบวนการ**: นำ source จากไฟล์ anime_cleaned มา group by ร่วมกับ studio เพื่อ ทำการหา score เฉลี่ยของอนิเมะที่จัดทำ และจัด rank ของ studio ภายในแต่ละ source
- ผลลัพธ์: ได้รู้ว่าสตูดิโอไหนมีศักยภาพในการจัดทำอนิเมะ จากเนื้อหาต้นฉบับประเภทนั้น ซึ่งจากกราฟที่ได้เราจะพบว่ามีบางสตูดิโอปรากฏชื่อในหลายประเภทเนื้อหาต้นฉบับ ที่มี ความใกล้เคียงกัน เช่น Novel, Light novel, Visual novel



- 2. รายชื่ออนิเมะ 10 อันดับแรก เรียงตามคะแนนจากมากไปน้อย โดยสมาชิกเว็บไซต์ที่ชื่น ชอบอนิเมะประเภทนั้น ซึ่งมีสัดส่วนการดูอนิเมะประเภทนั้นเป็นหลัก
 - **สิ่งที่น่าสนใจ**: หากเราวิเคราะห์กันแค่เฉพาะกลุ่มผู้ชมที่ดูอนิเมะประเภทนั้นเป็นหลัก นำมาลงคะแนนให้อนิเมะประเภทนั้น ผลลัพธ์ที่ได้นั้นจะน่าเชื่อถือขนาดไหน อนิเมะที่ติด อันดับนั้น จะเป็นอนิเมะที่ยอดเยี่ยมใช่หรือไม่
 - **กระบวนการ**: ทำการวิเคราะห์ username จากไฟล์ UserAnimeList-Drop เทียบ anime_id ว่าเป็น genre อะไรในไฟล์ anime_cleaned เพื่อหาสัดส่วนว่า username นั้น ชอบดูอนิเมะ genre ไหนเป็นหลัก หลังจากทราบแล้ว จึงนำคะแนน my_score ที่ username นั้นลงคนแนนให้ anime_id นั้นมาคำนวณ หาก anime_id นั้น genre ตรงกับ สัดส่วน genre หลักที่ username นั้นรับชม ทำการหาคะแนนเฉลี่ย และจัดอันดับของ คะแนนภายในแต่ละ genre
 - ผลลัพธ์: แต่ละ genre ส่วนใหญ่ คะแนนไม่สามารถอ้างอิงกับความเป็นจริง ที่กลุ่มผู้ชม ทั้งหมดช่วยกันลงคะแนนให้ได้ เมื่อเทียบ score from user favorite genre กับ score from MAL เราจะเห็นว่าอนิเมะหลายเรื่องนั้นได้คะแนนไม่สมเหตุสมผล มีทั้งอวยจนให้ เต็ม 10 คะแนน หรือไม่ชอบจนกดเหลือ 0 คะแนน และเมื่อดู rank ที่ถูกจัดบนเว็บไซต์ เทียบด้วย จะยิ่งเห็นว่าหลายเรื่องไม่ใช่อนิเมะที่ได้รับความนิยมอันดับต้น ๆ ของเว็บไซต์

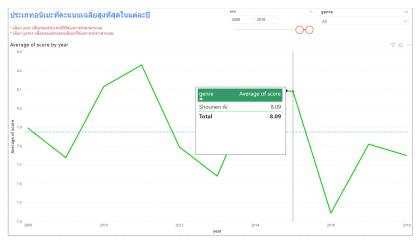
on 10 a	วันดับ รายชื่ออนิเมะ โดยผู้ชมที่ชื่นชอ	ນປຣະເຄທລານນະ	าั๊าเ			genre
	ประเภทอนิเมะนั้น หมายถึง user ที่มีอัดราการรับชม genre นั้นเป็นห		10 10			Fantasy
	บระเภทอนเมะนน หมายถึง user ทุมยตร in iรรบบม geine นนเบนข พื้อกรองประเภทอนิเมะที่ต้องการทราบ					
anime_id	title	score from user favorite genre	score from MAL	aired_from_year	rank	genre
33	Kenpuu Denki Berserk	10.00	8.45	1997	132	Action, Adventure, Demons, Drama, Fantasy, Horror, Military, Romance, Seinen, Supernatural
459	One Piece Movie 1	10.00	7.22	2000	2740	Action, Adventure, Comedy, Super Power, Fantasy, Shounen
460	One Piece Movie 2: Nejimaki-jima no Daibouken	10.00	7.28	2001	2500	Action, Adventure, Comedy, Super Power, Fantasy, Shounen
461	One Piece Movie 3: Chinjuu-jima no Chopper Oukoku	10.00	7.13	2002	3087	Action, Adventure, Comedy, Super Power, Fantasy, Shounen
463	One Piece Movie 5: Norowareta Seiken	10.00	7.39	2004	2078	Action, Adventure, Comedy, Super Power, Fantasy, Shounen
464	One Piece Movie 6: Omatsuri Danshaku to Himitsu no Shima	10.00	7.75	2005	998	Adventure, Comedy, Fantasy, Shoune
465	One Piece Movie 7: Karakuri-jou no Mecha Kyohei	10.00	7.41	2006	1985	Adventure, Comedy, Fantasy, Shoune
466	One Piece: Taose! Kaizoku Ganzack	10.00	6.83	1998	4207	Action, Adventure, Fantasy, Comedy, Super Power, Shounen
969	Tsubasa Chronicle 2nd Season	10.00	7.68	2006	1161	Action, Adventure, Fantasy, Romance, Supernatural, Drama, Mystery, Shounen

3. ประเภทอนิเมะ ปีและฤดูกาล

3.1 ประเภทอนิเมะที่คะแนนเฉลี่ยสูงที่สุดในแต่ละปีและฤดูกาล

- **สิ่งที่น่าสนใจ**: ทำให้รู้ว่าในแต่ละปีและฤดูกาล อนิเมะประเภทไหนได้รับความนิยมสูง โดยดูจากคะแนนเฉลี่ย ซึ่งสามารถนำมากำหนดแผนการฉายอนิเมะได้ ในช่วงเวลาที่ ใกล้เคียงกัน ควรจะมีการเปลี่ยนประเภทของอนิเมะ เพื่อให้ผู้ชมไม่เบื่อหน่ายในการรับชม
- **กระบวนการ**: ทำการวิเคราะห์ค่าเฉลี่ยของ score ในแต่ละ genre จากไฟล์ anime_cleaned โดยการ group by หาค่าเฉลี่ยในแต่ละปีและฤดูกาลออกมา จากนั้น แสดง genre ที่มีค่าเฉลี่ยของ score สูงที่สุด
- ผลลัพธ์: ทำให้เราทราบถึงความนิยมของประเภทของอนิเมะในแต่ละปีหรือฤดูกาล ซึ่ง จากการดูไทม์ไลน์แล้วจะพบว่า ในช่วงเวลาที่ติดกัน ประเภทอนิเมะที่ได้รับความนิยม มักจะต่างกัน

3.1.1 จำแนกตามปี



3.1.2 จำแนกตามปีและฤดู



- 3.2 คะแนนเฉลี่ยอนิเมะในแต่ละปีและฤดู และประเภทอนิเมะที่นิยมจัดทำมากสุด
 - **สิ่งที่น่าสนใจ**: ทำให้รู้ว่าในแต่ละปีและฤดูกาล มีค่าคะแนนเฉลี่ยเป็นอย่างไร และในแต่ ละปีและฤดูกาลอนิเมะแนวไหนที่มีจำนวนการจัดทำสูงสุด เพื่อเป็นแนวทางในการ ตัดสินใจว่าจะสร้างอนิเมะเรื่องถัดไปให้เป็นประเภทไหน
 - **กระบวนการ**: นำข้อมูลจาก anime_cleaned มาแยกตามปีและฤดูกาล จากนั้นก็ทำการ แยก genre ออกมาเพื่อหาว่า genre ไหนในแต่ละปีและฤดู ที่มีจำนวนมากสุด และนำ score มาคิดเฉลี่ยตามปีและฤดู
 - ผลลัพธ์: ทำให้เราทราบถึงว่าในช่วงปีและฤดูที่ใกล้เคียง มีอนิเมะประเภทไหนที่มีมีการ จัดทำเยอะสุด จากข้อมูลยังพบว่าในช่วงปีหลัง ๆ อนิเมะประเภท Comedy มีจำนวนมาก ติดต่อกันหลายเรื่อง หลายปี และหลายฤดู

