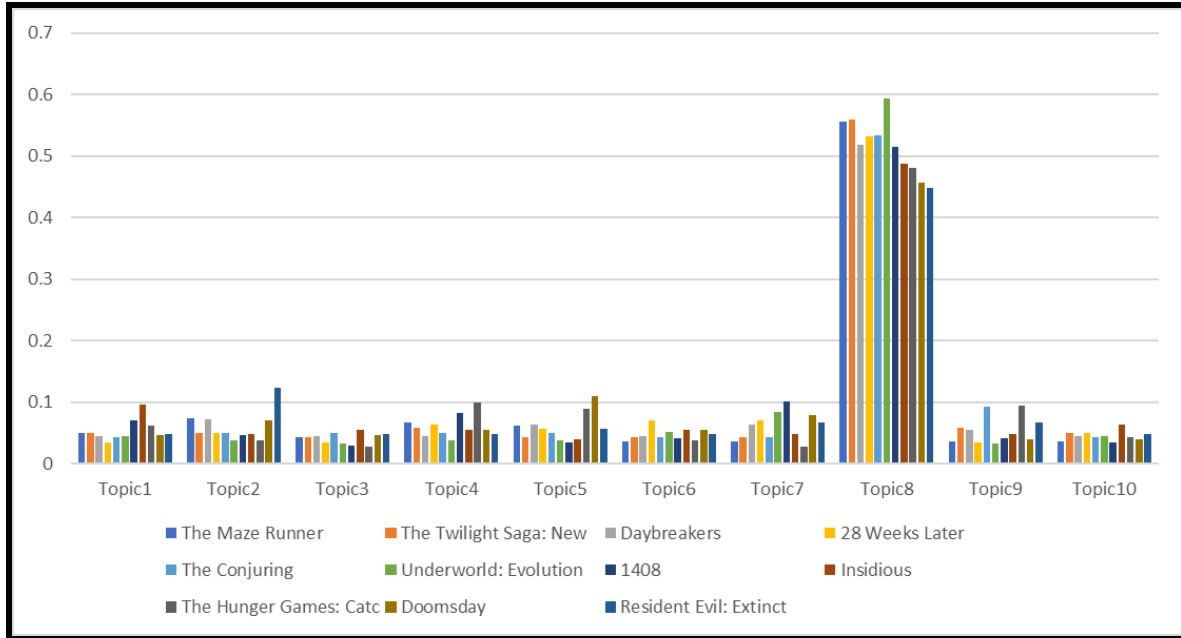


Homework 2: Movie Scheduling

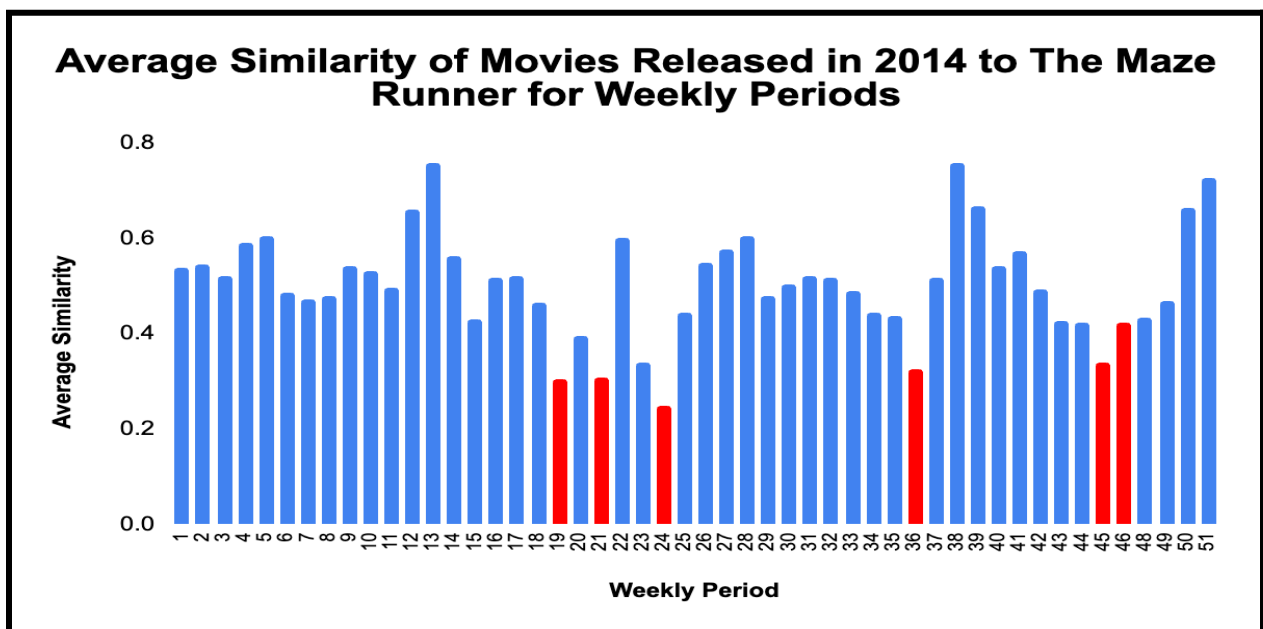
These film clusters span various genres and themes that can be categorized into 10 topics. **Topic 1:** Atmospheric/Action films captivate with complex narratives and high-stakes tension, exemplified by "Inception" and "The Dark Knight." **Topic 2:** Espionage/Action movies focus on high-stakes spy missions, as seen in "Casino Royale" and "Skyfall." **Topic 3:** Animation/Family features heartwarming stories like "Up" and "Finding Nemo." **Topic 4:** Fantasy/Adventure adaptations, including "Harry Potter" and "The Hobbit," offer fantastical journeys. **Topic 5:** Sci-Fi/Action explores futuristic themes in titles like "Avatar" and "Edge of Tomorrow." **Topic 6:** Superhero/Action showcases iconic characters in visually striking tales like "The Avengers" and "Iron Man." **Topic 7:** Revenge/Drama delves into darker narratives like "Inglourious Basterds" and "Django Unchained." **Topic 8:** Dystopia/Survival depicts post-apocalyptic scenarios, as seen in "I Am Legend" and "The Road." **Topic 9:** Romance/Drama covers emotional experiences from "Titanic" to "The Blind Side," while **Topic 10 as** Comedy offers humor in films like "The Hangover" and "Superbad."

Below are the top 10 similar movies to “The Maze Runner” based on Euclidean Distance-

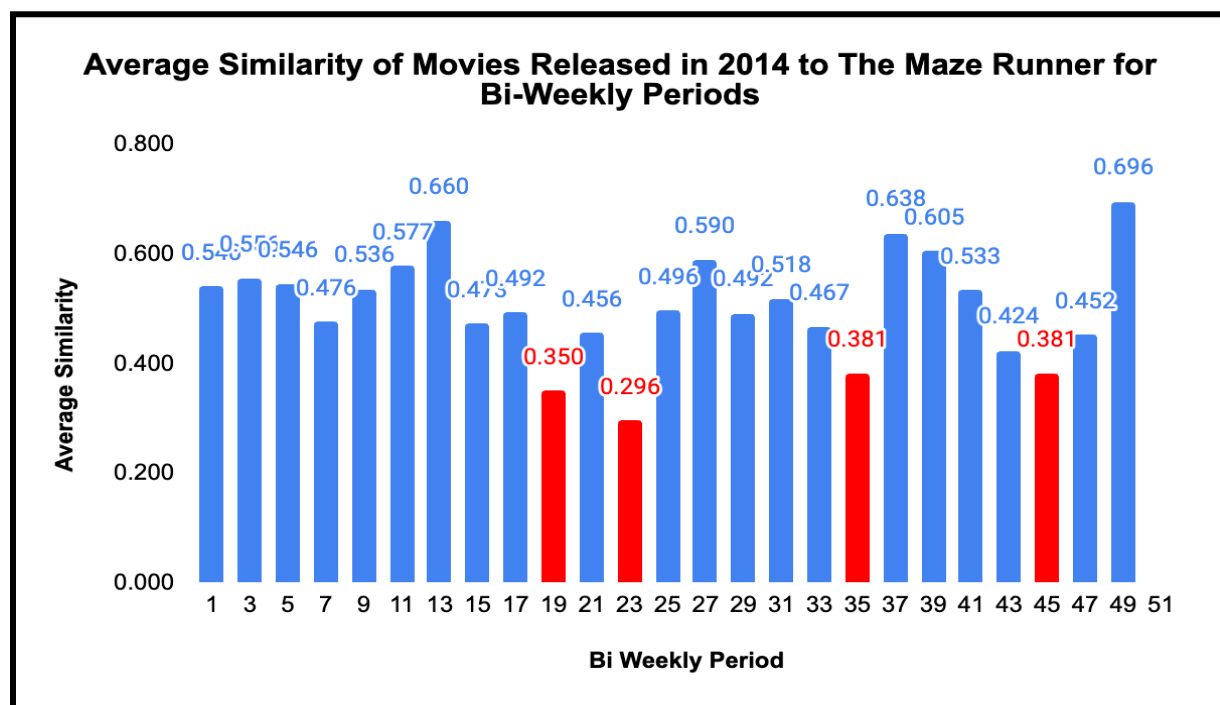
Movie	Euclidean Distance	Genre	Notes on Validation
The Twilight Saga: New Moon	0.0418	Fantasy, Romance	Diverges in theme; lacks dystopian/sci-fi elements.
Daybreakers	0.0555	Sci-Fi and horror	Shares dystopian elements; they differ in core focus.
28 Weeks Later	0.0625	Horror and sci-fi	Aligns well with dystopian, survival themes.
The Conjuring	0.0690	Horror, Supernatural	Lacks alignment with dystopian/sci-fi themes.
Underworld: Evolution	0.0819	Action, Fantasy	Focuses on fantasy, and vampires; lacks dystopian feel.
1408	0.0896	Horror, Psychological	Psychological thriller, not aligned with dystopia.
Insidious	0.0976	Horror, Supernatural	Supernatural theme; not aligned with dystopia.
The Hunger Games: Catching	0.1112	Sci-Fi, Dystopian	Close alignment with dystopian, sci-fi themes.
Doomsday	0.1200	Action, Sci-Fi	Shares post-apocalyptic, dystopian elements.
Resident Evil: Extinction	0.1287	Action, Sci-Fi, Horror	Shares post-apocalyptic theme; more action-focused.



By examining the bars representing "The Maze Runner" across topics 1 to 10, we notice that they have distinct heights. We anticipate that movies with similar height patterns across these topics would be the closest matches when considering Euclidean distance. The bar chart visually demonstrates that these movies share a comparable pattern with "The Maze Runner" in terms of their topic distributions, **validating their similarity**.



To determine the ideal release date for "The Maze Runner," we conducted **an analysis of the 2014** movie release schedule, computing the **average similarity across 52 weeks**. The similarity metric considered various factors including **genre, target audience, and themes**. Furthermore, we delved deeper into a **bi-weekly schedule** to pinpoint the **optimal two-week window** for maximum performance, ensuring no similar genre movies coincide. In analyzing the release schedule, we carefully considered the potential competition from movies released in the preceding week. Leveraging our expertise, we recognized the **primary audience demographic aged 13-25** and strategically considered the **summer break period** for the release. By aligning with the summer vacation period, we aimed to capitalize on increased leisure time and moviegoing activity among the target demographic. The Bar-plot illustrates the average weekly and bi-weekly similarities between the movies vs "The Maze Runner," revealing five optimal weeks where the least similar movie is being released - **Week 24, Week 19, Week 45, Week 36, Week 46**.



Below are the recommended release dates for the movie-

Week 24 (Release Date 6/12/2014): The two films scheduled for release on this date span different genres, with one falling under animation and the other under comedy. Additionally, the movies slated for release in the following week are unrelated in terms of genre.

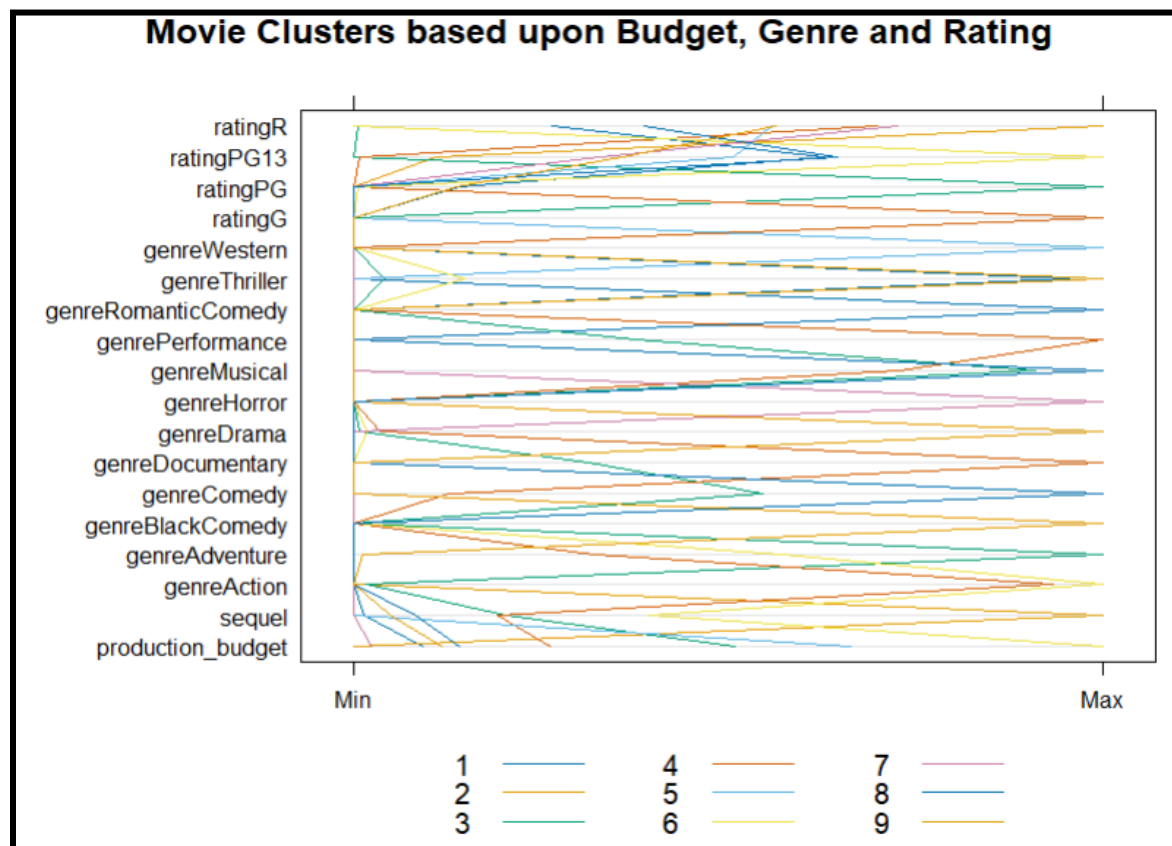
Week 19 (Release Date 5/9/2014): Likewise, the films set for release on this date do not share the same genre. Despite being a week after the release of a major film like "Spider-Man," we believe that the summer vacation period enhances the potential for "Maze Runner" to succeed.

Week 45 (Release Date 11/07/2014): Likewise, the films set for release on this date do not share the same genre. Despite being a week after the release of a major film like "Spider-Man," we believe that the summer vacation period enhances the potential for "Maze Runner" to succeed.

Week 36 (Release Date 9/5/2014): Given that only one movie, "The Identical," is scheduled for release this week, we consider it as another potential option for launching our movie. Additionally, "The Identical" does not belong to a related genre, further supporting the suitability of this week for our movie's release.

Week 46 (Release Date 11/14/2014): Lastly, Week 46 does not provide a strong complement to "The Maze Runner" as the movies do not share the same topic. However, this remains a viable option as Week 47 sees the release of "Hunger Games," potentially leading to a decline in ticket sales for our movie.

KMeans Using Movie Characteristics



Parallel plot represents the clustering analysis for 9 clusters against the selected columns. On performing the k-means method using 9 clusters, “The Maze Runner” was part of Cluster 3. Looking at the results as generated by the code, KMeans preferred dates for launching The Maze Runner are - **April 18, April 25, August 8, and May 9.**

While k-means effectively partition the data into clusters based on similarity, it does not inherently capture the semantic relationships or thematic coherence present in the text data. This means that the resulting clusters may not represent coherent topics or themes, making interpretation challenging. As we know K-means isn't naturally suited for categorical data like movie genres. If genres were simply encoded as numeric data without accounting for the categorical nature, k-means might yield clusters that don't make sense in terms of genre grouping. The interpretation of k-means clusters is more straightforward in terms of centroid-based similarity, but less nuanced in capturing the underlying distribution of features as LDA does. Also, another limitation of k-means is the lack of inherent semantic meaning within the clusters generated by k-means. K-means also requires specifying the number of clusters beforehand, which can be arbitrary and may not reflect the true underlying structure of the data. Additionally, k-means operates on vector representations of the text data, which may not adequately capture the complex semantic relationships inherent in natural language. As a result, the clusters generated by k-means may lack interpretability and may not accurately reflect the underlying topics present in the text corpus.