

MOVIE SCHEDULING CASE

To interpret the topics and labels identified from the LDA analysis, we examined the topics and the top 10 terms (refer Exhibit 4) with the highest weights associated with each topic. This gave clarification of the significance of the terms and its grouped topic. Simultaneously, we reviewed the dataset of movies to determine the top 10 movies strongly associated with each topic (weightage). For example, Topic 6, labelled as “Superhero” included top terms like “marvel” and “based on a comic”, with movies such as “The avengers” and “X-Men”. Likewise, Topic 8, labelled as “Survival” showcased terms such as “dystopian”, “post-apocalyptic” and “horror”, with movies like “Limitless” and “Daybreakers”. This interpretation allowed us to align topics with meaningful clusters using term relevance and examples.

To summarize the similarity between *The Maze Runner* and other movies, the Euclidean distance between its 10-dimensional topic score and those of all other films, has been used. Euclidean distance serves as a direct measure of similarity, with smaller values indicating greater resemblance. To validate these results, cross checking was done using cosine similarity (*the higher the number, the greater the similarity*), which measures how directionally aligned the topic distributions are. Based on my analysis, I determined that movies with a Euclidean distance within one and a half to two standard deviation (*of the whole subset i.e. $SD = 0.14$*) from 0 are very similar to *The Maze Runner*. This ensures that only movies with closely aligned topic distributions are considered while filtering out significant outliers.

The ten most similar movies to *The Maze Runner* (refer Exhibit 1) based on this method include *The Twilight Saga: New Moon*, *Daybreakers*, *28 Weeks Later*, *The Conjuring*, *Underworld: Evolution*, *Insidious*, *1408*, *Doomsday*, *The Hunger Games: Catching Fire*, and *Resident Evil: Extinction*.

Many of these films share common themes of dystopia, survival, action, and supernatural elements, reinforcing the model’s validity. However, the inclusion of horror movies like *The Conjuring* and *Insidious* suggests some overlap in suspense-building and thriller aspects rather than strict dystopian storytelling. To further refine the similarity measure, one could apply genre-based filtering, the inclusion of more topics, weighted topic scores, or clustering techniques to better differentiate between dystopian sci-fi and supernatural horror.

To determine the optimal release week for *The Maze Runner*, our team analyzed movies released in 2014 to identify weeks with minimal competition. The objective was to recommend a launch date that avoids direct competition with similar films while aligning with key seasonal trends like spring break and the holiday season. The analysis involved

calculating weekly similarity scores and factoring in residual competition from prior weeks. Weekly groupings were created to ensure accurate time-based comparisons.

Measuring Similarity: (Refer Exhibit 2 and 3) Using the topic model data provided, Euclidean distances were calculated between the 10-dimensional topic distribution of *The Maze Runner* and all other movies. This metric quantified how “similar” other movies were to *The Maze Runner*. Smaller distances indicated higher similarity and, therefore, greater competition. For each week, the following measures were calculated:

- **Average (Same-Week) Similarity:** Representing overall competition from movies that open in the same week.
- **Residual (Prior-Week) Similarity:** Accounting for lingering competition from movies released in the prior week.
- **Next-Week Similarity:** Factoring in potential overshadowing by movies scheduled to open the week immediately following (since these can cannibalize second-week grosses).

Ranking the Weeks: Weeks were ranked based on their **combined similarity scores**, which incorporated **same-week**, **prior-week**, and **next-week** competition. Weightings were assigned to each of these factors to reflect their relative impact on box office. This ensured a balanced evaluation of both direct (same-week) and indirect (ongoing or upcoming) competitors. Weeks with the highest combined similarity (i.e., **least** competition) were ultimately recommended for the release of *The Maze Runner*.

Recommendation:

Top Choice: 7th November, 2014

Second Best Choice: 9th May, 2014

Third Choice: 23th May, 2014

Our analysis identified several key weeks in 2014 with minimal competition for *The Maze Runner*, ranked based on combined similarity scores. These scores account for both direct competition from movies released the same week and residual competition from earlier releases still performing well. While exhibit 2 highlights a Monday as the start date of the week, we prefer releasing the movie on a Thursday or Friday to attract working audiences who may accompany

teens and kids, ensuring stronger opening numbers. Our primary goal is to ensure that teens can attend with their parents during the opening weekend, leveraging strong word-of-mouth to boost performance.

The week of November 3, 2014, ranks among the top three weeks with the lowest competition, as indicated by a combined similarity score of 0.7719. This makes it one of the most distinct weeks for releasing *The Maze Runner*. Additionally, **the film aligns well with the teen sci-fi/action genre**, and students are typically settled into their school routines, with exams still a while away. The release would also coincide with the holiday season, leading up to Thanksgiving, and take advantage of the “Awards Season,” ensuring the film is remembered as a teen-friendly release to close out the year. Based on these factors, **we recommend releasing the film on November 7, 2014.**

We are avoiding the month of June as some schools/colleges have their exam weeks in that month while some are closed so teens might usually be out on vacations. The most optimal dates as per the similarity score are around the end of June but the release for “Deliver Us from Evil” is scheduled for 2nd July, 2014 and we do not want to compete with the same in the 2nd or 3rd week of the release. We are also not keen on releasing the movie in the month of April as that might be the peak exam season for the students. In the early month of April, we may also face competition from the flare of animated films that release before Easter.

The next best options for the film's release are May 23, 2014, and May 9, 2014, with respective combined similarity scores of 0.6497 and 0.6179. We believe May is an ideal release month since most students have finished their exams, which typically occur in the first week of June. Additionally, many cities have summer breaks during this period, making movie-going a popular recreational activity. Furthermore, as indicated in Exhibit 3, there are no similar films scheduled for release in May, providing a strategic advantage.

LDA results change if we use 15 or 20 topics:

Moving from 10 to 15 topics, additional sub-genres and thematic variations emerged.

At 20 topics, topics became even more specific, with more niche classifications.

10 topics offered a good balance between interpretability and precision, grouping movies in meaningful categories.

15 topics introduced more refined distinctions, such as splitting sci-fi into space-related vs. dystopian.

20 topics increased granularity but reduced interpretability, as probabilities spread across more categories, making direct comparisons harder.

Exhibit 1: Movies with the least Euclidean Distance to The Maze Runner

| Movies | Euclidean_Distance | Cosine_Similarity |
|---------------------------------|---------------------------|--------------------------|
| The Twilight Saga: New Moon | 0.042 | 0.997 |
| Daybreakers | 0.056 | 0.997 |
| 28 Weeks Later | 0.063 | 0.995 |
| The Conjuring | 0.069 | 0.993 |
| Underworld: Evolution | 0.082 | 0.992 |
| 1408 | 0.090 | 0.989 |
| Insidious | 0.098 | 0.990 |
| The Hunger Games: Catching Fire | 0.111 | 0.985 |
| Doomsday | 0.120 | 0.986 |
| Resident Evil: Extinction | 0.129 | 0.984 |

Exhibit 2: Weekly Similarity Table (Most dissimilar to similar weeks)

| Release week | Combined Similarity Score |
|--------------------------------|----------------------------------|
| Monday, 31 March 2014 | 0.842053 |
| Monday, 23 June 2014 | 0.817872 |
| Monday, 3 November 2014 | 0.771964 |
| Monday, 2 June 2014 | 0.758714 |
| Monday, 19 May 2014 | 0.649682 |
| Monday, 5 May 2014 | 0.617855 |
| Monday, 27 October 2014 | 0.590224 |
| Monday, 28 July 2014 | 0.590028 |
| Monday, 25 August 2014 | 0.583306 |
| Monday, 9 June 2014 | 0.576818 |



Exhibit 3: Most Similar Movies (Most similar to dissimilar movies)

| Movie | Release Date | Avg Similarity Score |
|--|----------------------------------|-----------------------------|
| The Maze Runner | Friday, 19 September 2014 | 0 |
| The Hunger Games: Mockingjay - Part 1 | Friday, 21 November 2014 | 0.190416 |
| Divergent | Friday, 21 March 2014 | 0.199044 |
| Noah | Friday, 28 March 2014 | 0.247163 |
| Exodus: Gods and Kings | Friday, 12 December 2014 | 0.250885 |
| Left Behind | Friday, 3 October 2014 | 0.287502 |
| Vampire Academy | Friday, 7 February 2014 | 0.331192 |

| | | |
|-----------------------------|--------------------------------|-----------------|
| I, Frankenstein | Friday, 24 January 2014 | 0.39041 |
| Dracula Untold | Friday, 10 October 2014 | 0.397479 |
| Deliver Us from Evil | Wednesday, 2 July 2014 | 0.412057 |

Exhibit 4: The Top 10 topics and associated labels

| Topic Number | Label |
|---------------------|--------------------|
| Topic 1 | atmospheric |
| Topic 2 | action |
| Topic 3 | animation |
| Topic 4 | fantasy |
| Topic 5 | sci.fi |
| Topic 6 | superhero |
| Topic 7 | revenge |
| Topic 8 | survival |
| Topic 9 | romance |
| Topic 10 | comedy |

Exhibit 5: Topic Distribution



