

SEASONAL PATTERNS IN MOVIES

A research report

<https://github.com/jezebleum/project-vu-object-to-data>

Anne-Sophie van Hulst - 2574603

Sophie van Duin - 2601225

Jesse Blom - 2606653

Digitization: from Object to Data

22/10/2018

Professor: Erika Kuijpers

Introduction

Not everyone has the same Christmas feelings, but a lot of them are very alike. Pretend a cold night, while walking home you see all sorts of houses on the street with Christmas decoration, Christmas trees behind windows and burning wood in the fireplace. It starts snowing a bit, and a car passes with your favorite Christmas song. Isn't this getting you in the festive Christmas mood?

Most people do not really experience this feeling until the first Christmas day. There are also people who prefer Boxing day over Christmas. The feelings are getting less and less after Boxing day, because New Years is around the corner (Beech, 2015). Therefore the time frame of 25-27 December was chosen for this research report.

In this report research is done to formulate an answer to the following research question:

“Are there any seasonal patterns within (re)viewing habits? And if so, what could be the cause of this?”

The relevance for this research question is that it could be very helpful to marketing purposes. Research has shown us that every product has some sort of seasonal marketing (Radas, 1998). It is common knowledge that movies within a certain theme, for example Christmas, will most likely be watched within the month of preference, which is in this case December. It could be possible that being exposed to a product outside of the seasonal preference could have a negative effect on the experience of the product. People working in marketing are able to see this in reviews. If there are a lot of negative reviews that have this cause, sellings could drop. Our question could help prove if there is a seasonal pattern or not.

It is proven that events (like a specific holiday) have strong memories associated with them (Holak, 1992). This means that whenever someone watches a Christmas movie in December, they're most likely going to have a different movie viewing experience as opposed to not having that sort of nostalgia associated with it. There is a possibility that this also influences the rating of a product, and this could be the cause of any seasonal preference.

To make it easier to answer the research question it is very important to create a frame in which our answers can be found. In this report there will be more about the feeling behind Christmas and why people would like to watch movies about this holiday, or with this holiday as its theme. Therefore the following sub questions were formulated:

What is nostalgia? How stand holidays in connection to nostalgia? Are user reviews usable and reliable? Do the ratings of the reviews show similarity with the texts of the reviews?

First we will explain more about nostalgia and user reviews. After that there will follow a sentiment analysis. Then the method and the data curation are discussed, followed by the results, the conclusion and the discussion.

Nostalgia

There is a difference between historical and personal nostalgia. Historical nostalgia refers to feeling attracted to times in the past. You can also experience this nostalgia feelings for a period in which you did not live. Persons today can feel very nostalgic about e.g. the Dutch Golden age time period. The other type of nostalgia is the one with the personal touch. It refers to feeling emotions towards an individual's past they experienced themselves. People could miss for example their childhood memories and they feel very nostalgic when something similar happens in the present. (Parry, 2011)

To make it easier one could say that nostalgia are the warm and welcome feelings and memories from the (personal) past.

A really important factor within nostalgia are relationships. How we are connected to other people helps to identify who we are, and nostalgia helps us to appreciate those relationships.

Relationships are also a very important factor in holidays. People want to be with their family in holiday times, to participate in traditions and be with people who they don't see that often. Think about the song: *I'll be home for Christmas*. This is a song about feeling nostalgic and wanting to go home for the holiday. (Parry, 2011)

Nostalgia also helps people that tend to feel lonely easily, to feel connected, and less lonely. It is like going back in time and remembering the good feelings. This is a feeling that we believe is a very important one. During Christmas you do not want to be alone, you want to spend time with the ones you love. It does not really matter what you do, as long as you are together and feeling the love of Christmas.

User reviews

To decide whether to buy or use a certain product, people often make use of the user reviews from people that have already consumed this product. By knowing how someone else rated the product and what they thought of it, it makes it easier to make the decision whether to purchase or not. This raises the question if these user reviews are usable and reliable. The phenomenon of user reviews also goes by the name e-WOM, electronic Word of Mouth. Duan et. al (2008) established a relationship between the information in e-WOM and offline movie sales, which could mean for the producer that e-WOM has a positive effect on the sales of a product.

Kim & Gupta (2012) researched the effects of the expression of emotion in e-WOM. The results of this research showed that “negative emotions in a negative review lower perceived reviewer rationality, thereby reducing review informative value and, in turn, lead to less negative product evaluations”. On the other hand, when the identity of the reviewer is known, a negative review is assumed to be more conceivable than a positive review (Kusumasondjaja et al. 2012).

Sentiment analysis

When writing a review, people want to share their thoughts of a specific product or service with the rest of the world. Mostly a review consists of two parts. The first part of the review is a rating, where the customer can give a rating from 1 to 5. The other part is writing the review, where the customer can tell their opinion and explain what they did or did not like about the product. To find out what emotions the writer of the review experienced when using the product, sentiment analysis is a good method to learn about these emotions. Sentiment analysis was used in this research

to answer the question whether the ratings of the reviews show similarity with the texts of the reviews.

Sentiment analysis or opinion mining is a way of analysing textual data with as goal to find out which emotions were used in the text. Broadly speaking, sentiment analysis intends to determine the attitude of the writer of the text with respect to some topic. To do a sentiment analysis, the affective words of the text need to be classified. One way of classifying these affective words is by dividing them into the five basic emotions that have been established by Ekman (1992). These five basic emotions are happiness, sadness, anger, fear and disgust. By learning about the score in every category, it becomes clear what the attitude of the writer of the text is with respect to the topic. Instead of classifying the emotions that were used in the user reviews into Ekman's five basic emotions, the emotions could be divided into positive and negative emotions. By using polarity instead of using five different categories, it simplifies drawing a conclusion from the results.

As stated by Strapparava & Mihalcea (2010) is the classification of emotions especially complicated. "On the one hand, emotions are not linguistic entities, and thus many of the previously proposed approaches for emotion detection were developed in a variety of other fields, including psychology, sociology, or philosophy". Sentiment analysis is, in case of a big dataset, done by a computer instead of manually. The fact that it is done by a computer decreases the accuracy of the analysis. The text in user reviews can be seen as natural language, and a computer cannot process natural language as good as a human being could.

Method and data curating

Provenance, limitations and promises

The original folder contained 141 csv files, each representing the reviews of a movie on Amazon. Amazon released these files on the web to provide data for further research. These files were formatted in the same way - details about the movie, the user and their rating, the text of their review, the date of the review and whether other users found the review helpful or not. Both users and movies were identified with a unique id, the actual reviews did not have a unique id.

We chose 7 movies by looking at the theme, the story, the characters and how the movie was intended. More about the movies itself with a description why they were chosen are stated in *Appendix 1*.

The dataset is intended to showcase reviews of users to other users, this way Amazon used a new form of marketing that is being created by people who bought their product and whether they advise the product to others, which is a different approach of the data than used in this report. The format that Amazon used at the time showcased simply a rating of a product and the experience around the product, for example the delivery, packaging, costs all influence this rating and the grade given to the product. This limits the research since there is no clear description of what made a user define their rating, however the rating is supported by the review text and the summary of this text. Since the project is on a smaller period of time, there was decided to not dive into this research field on a deeper level.

Users were able to review a product after purchasing, but in this case there is no confirmation whether a user actually watched the movie or not. There is no transparency behind this, the same goes for the review date - the review does not provide insight whether a user watched the date of reviewing, how many hours are in between writing the review and watching and so on. The data can not be altered to provide more information behind these reviews and the moment of watching.

Data analytics

All analysis within this research was executed with the use of Python 3.6.4 and Jupyter Notebook was used to apply this language. From there on a combination of multiple libraries within Python were used to provide the necessary functions to get to the final results that were needed. The most important alterations and important steps within this part of the research is explained in this chapter. The complete notebook can be accessed on the github of the project, including the necessary resources to recreate the alterations and the visualizations.

To be able to answer the questions in this research report the original data had to be altered and compiled into a different format of files - the files had to be divided into a

dataset with Christmas movies and a dataset with movies that do not lie within the spectrum of Christmas. The manual alteration of this process is described earlier on in this report. This combination of files were added to two different dataframes (pandas library) which were then used to apply calculations onto the different rows. The rows that included text had to be altered into text without any html characters within the text - that way sentiment analysis can also be applied to these rows.

To visualize the monthly patterns within the dataset, the data of the ratings row had to be divided into months with the use of regex (re library). The ratings were added to twelve lists, each representing a month. This way averages can be calculated and the length of these lists represent the total of reviews that occurred within these months throughout the years.

To apply any sentiment analysis onto the review texts, the reviews had to be divided into two different sets of lists representing positive (4.0 / 5.0) and negative (1.0 / 2.0) ratings in both dataframes. From there on the actual analysis can be executed with the use of the sentiment function within the TextBlob library.

The final visualizations were all using the lists mentioned above and were represented into a combination of bar and line charts. These charts were all created with the use of the matplotlib library and the seaborn library was implemented to add a grid to the visualizations.

Results

The Christmas ratings were expected to be positively influenced when compared to the ratings throughout the rest of the year, there is a total of 605 reviews within 25-27 december and 9924 reviews in the rest of the year. The average rating of Christmas movies during Christmas is 4.52 and the rest of the year has an average rating of 4.65. The other dataset has an average rating of 4.36 during both periods of time.

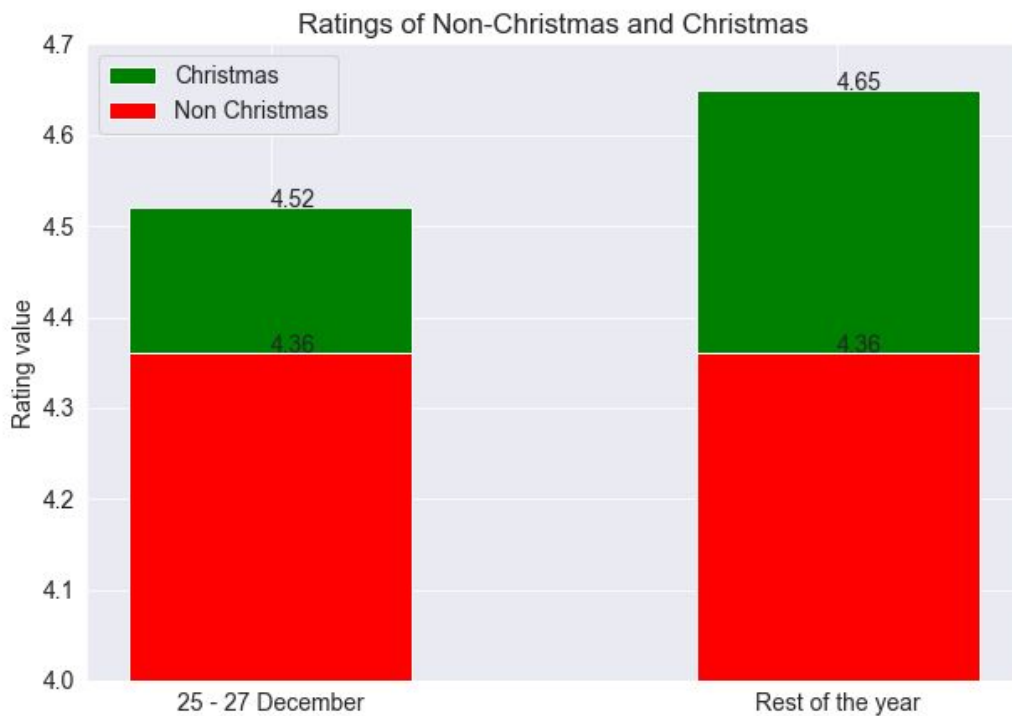


Figure 1: Review ratings during Christmas and the rest of the year in both datasets

To get more insight into the patterns, the reviews got divided into each month of the year. Both datasets were split up into monthly reviews and each month became defined by an average rating. In figure 2 it is visible how the two datasets are different to each other, the Christmas dataset has a more fluctuating rating throughout time and has spikes in August and September. The lower ratings are in April, July and October, which is a different pattern than the normal dataset.



Figure 2: monthly (average) ratings in both datasets

To gain knowledge regarding the popularity a movie gains in a month, the amount of reviews in a month were visualized into two different line charts in figure 3. The reasoning for dividing the two charts was primarily because the non-christmas dataset has over 200.000 entries, while the christmas dataset only had approx. 10.000, which made the combination of the two become skewed by the bigger dataset. The top chart represents the non-christmas dataset and the bottom dataset represents the christmas dataset. It is clear that the Christmas movies gain popularity starting in September and climb in a fast rate up until December, and from there on declines up until February/March. The non-christmas dataset shows a different pattern than the christmas dataset very clearly.

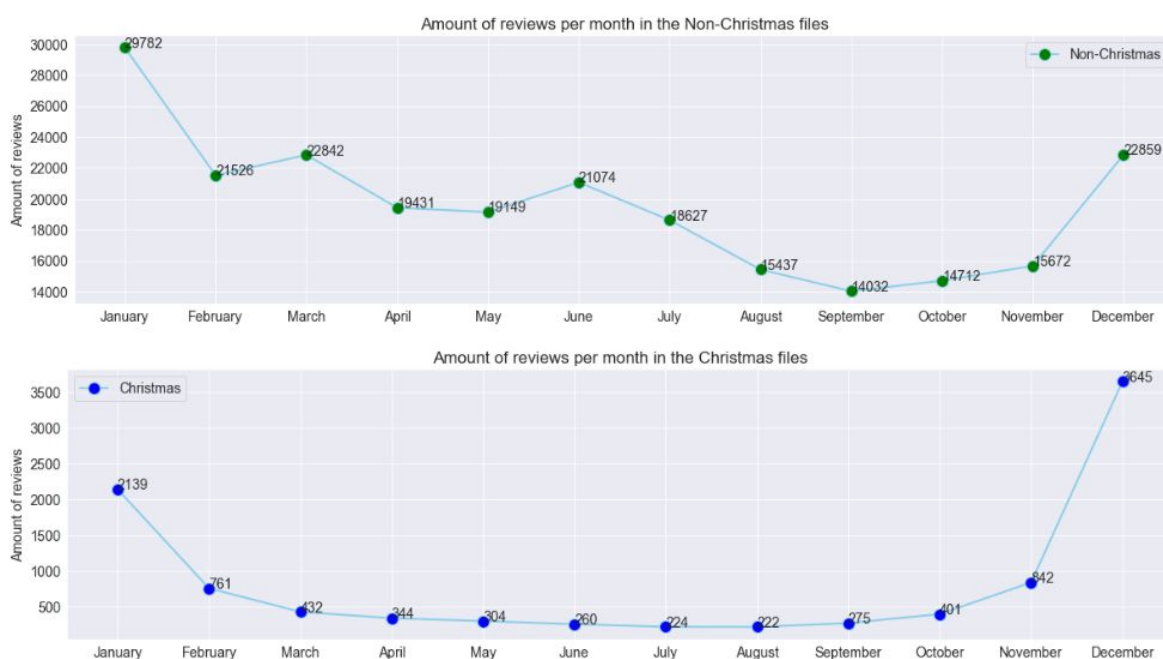


Figure 3: amount of reviews per month in both datasets, top line chart represents the non-christmas dataset and the bottom line chart represents the christmas dataset.

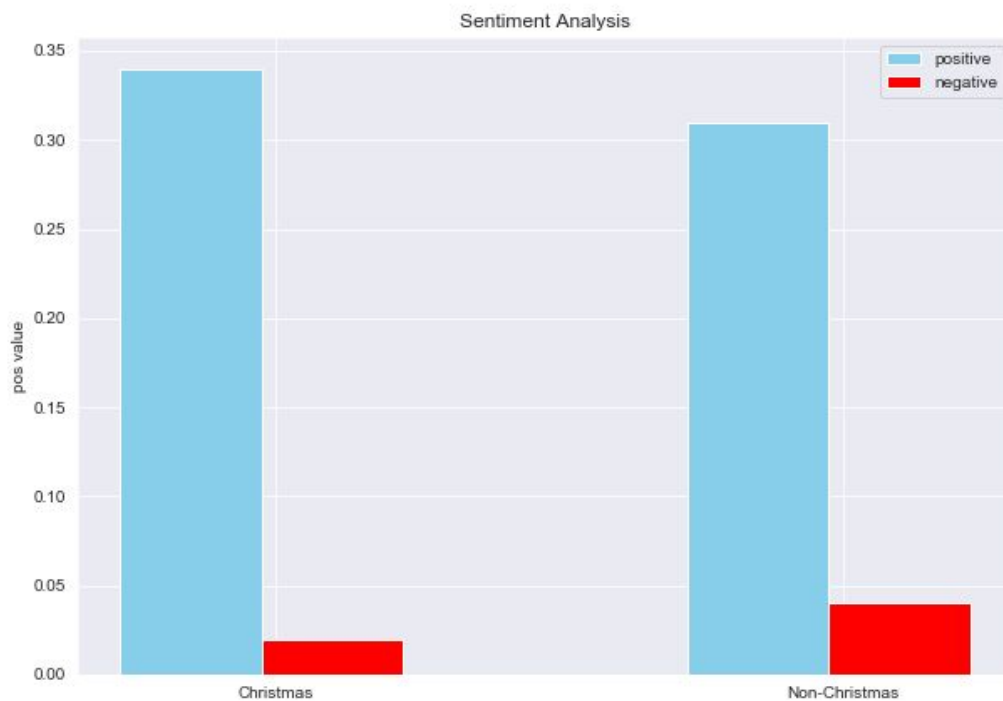


Figure 4: polarity of positively and negatively rated reviews for both Christmas and non-Christmas movies

The sentiment analysis gave the polarity of the positively and negatively rated reviews. The positively rated reviews had a polarity of 0.34, which is reasonably positive. The negatively rated reviews had a polarity of 0.02, which is neutral. These are the results of the Christmas movies. The non-Christmas movies gave similar results, namely a polarity of 0.31 for the positively rated reviews and a polarity of 0.04 for the negatively rated reviews.

Conclusion

There are clear patterns in the (re)viewing of users within movies, for example there is a visible pattern in the moment of time when a movie is watched, a big percentage of reviews were written in December and January, supposedly this shows that there is a form of seasonality in movies in general. On the other hand, the average rating

came out negative as opposed to the rest of the year, which is unexpected since the literature on for example nostalgia should positively influence this rating. Based on the research done in this project it is clear that nostalgia is not a variable in the process of rating a movie, but there is no clear definition on what the rating is telling. Users could give a negative rating on a movie by other reasoning than the actual movie, for example delivery time, the costs and so on. Further research should not be done on this dataset, but on a dataset which clearly divides the all parts of a rating, that way it is more valid to do such research. Based on this dataset there is a negative pattern in the reviewing of users based on the seasonality of the movie. This would tell us that if a (seasonal) movie tries to get a high review rating, perhaps the release should be done in a different moment of time, however if the revenue is more important, the movie should be released in its seasonal theme (which is in this case around Christmas).

The sentiment analysis gave results on the polarity of the positively and the negatively rated reviews. On the one hand, the positively rated reviews had a positive polarity, which means that there was used a fair amount of positive emotions when in these reviews. On the other hand, the negatively rated reviews had a neutral polarity, which means that there was not made use of many affective words in the reviews. There was no difference in the polarity of Christmas movies and non-Christmas movies, thus this is not specific for Christmas movies. A reason for this phenomenon could be that people have some kind of fear of expressing negative emotions and rating a movie badly is easier than actually expressing this in words. Another reason could be the fact that negative reviews cause a decreased perceiving of the reviewer's rationality and thus lower the amount of negative reviews (Kim & Gupta, 2012).

Possibilities for future work

By doing this research, several possibilities showed up for future research. Firstly, it might be interesting to include more Christmas movies and to extend the research on what makes a Christmas movie and whether the fact that it is a Christmas movie influences the way people review these. Secondly, in this research it is not clear how

a rating was established. There was not made use of the review text in these measurements. Further research could create a framework in which all the classifications, of the rating as well as the text are taken into account. Thirdly, for future research one could think about setting up a rating for the reviews to establish whether these reviews are usable and reliable.

Discussion

As stated before in the paragraph on limitations of the data, the data does not provide information if the reviewer actually watched the movie or not. There is no transparency behind this, neither behind the review date. We do not know how much time there is between the upload of the review and the moment the reviewer finished watching the movie. Is the information still fresh, and how valid is the review if it is not written within an acceptable amount of time?

It is clear that the results are coming from 7 fairly old movies. The oldest movie is from 1951, and the newest from 2005. Average is 1976. This is over 40 years ago. When compared to all the other Christmas movies ever made, the used selection is a very old selection (except for Narnia). There was not have another option for Christmas movies, since these were the only movies passing the criteria that were set.

There is no usage of any sort of linked data in this research project, the reason for this is that there is no reason to based on the research purpose. However, it could have helped with the reasoning on why a Christmas movie is considered a Christmas movie - a lot of variables could have been added to each title and based on this information the reasoning could be more detailed. But besides this, there is no other reason on why it should be added. Although it seems like the reasoning in this project is accurate and does not need to be altered with a form of linked data.

It became clear that transparency and communication are very important in a group project. Working in interdisciplinary and intercontinental groups

comes with different approaches and other points of view, which makes working as a group very difficult to nearly not working at all.

Bibliography

Beech, S. (2015). When does the 'festive feeling' really start to kick in? *Mirror*.

Duan, W., Gu, B. & Whinston A. B. (2008). Do online reviews matter? – An empirical investigation of panel data. *Decision Support System*. 45(4). 1007-1016.
Amsterdam: Elsevier

Ekman, P. (1992). Are there basic emotions? *Psychological reviews*, (99)3: 550-3

Holak, S. L., & Havlena, W. J. (1992). Nostalgia: An exploratory study of themes and emotions in the nostalgic experience. *ACR North American Advances*.

Kim, J. & Gupta, P. (2012). Emotional expressions in online user reviews: How they influence consumers' product evaluations. *Journal of Business Research*. 65(7). 985-992. Amsterdam: Elsevier.

Kusumasondjaja, S., Shanka, T. & Marchegiani, C. (2012). Credibility of online reviews and initial trust: The roles of reviewer's identity and review valence. *Journal of Vacation Marketing*. 18(3). 185-195.

IMDb. We used the movie descriptions from IMDb to describe the movies and why we chose them.

Parry, W. (2011). Why We Feel Nostalgic During the Holidays. *Holiday Psychology & Traditions*.

Radas, S., & Shugan, S. M. (1998). Seasonal marketing and timing new product

introductions. *Journal of Marketing Research*, 296-315.

Strapparava C., Mihalcea R. (2010) Annotating and identifying emotions in text.

Intelligent Information Access. 301. 22-38. Berlin: Springer.

Appendix 1

Description per movie why we selected it for our research.

- A Christmas Carol (1951)

Main character Scrooge is awakened on Christmas Eve by 3 ghosts (ghost of Christmas past, ghost of Christmas present and ghost of Christmas future) who reveal to him his own miserable existence, what opportunities he wasted in his youth, his current cruelties, and the dire fate that awaits him if he does not change his ways. Scrooge is faced with his own story of growing bitterness and meanness, and must decide what his own future will hold: death or redemption.

The film takes place during Christmas and it has this Christmas feeling while watching.

- A Christmas Story (1983)

Christmas is approaching and 9 year-old Ralphie wants only one thing: a Red Ryder Range 200 Shot BB gun. When the big day arrives he gets lots of present under the tree including a lovely gift from his aunt that his mother just adores. But what about the BB gun?

The film takes place during Christmas and it has this Christmas feeling while watching. The characters are singing Christmas songs, it is snowing, the street and

the house is with warm lights and in their house they have a beautiful tree with presents under it. The typical Christmas image.

- Miracle on 34th Street (1959)

A kindly old gentleman is working as a Macy's department store Santa when he causes a commotion when he claims to be the real St. Nick. Forced to prove his sanity, he is taken to court where he must convince the judge and his friends that he actually is Santa Claus. The faith of one skeptical little girl and a Christmas miracle are the keys to his true identity.

This is a Christmas movie because the main character tries to prove he is Santa Claus, who is the person that brings the presents around at Christmas night (what the little kids believe).

- National Lampoon's Christmas Vacation (1989)

This is a comedy Christmas movie about a family preparing for Christmas while everything that could go wrong, goes wrong. Clark's (the father of the family) continual bad luck is worsened by his obnoxious family guests, but he manages to keep going knowing that his Christmas bonus is due soon.

It is a Christmas movie because the theme is all about Christmas preparations, getting in the mood for the festive celebration.

- Chronicles of Narnia - The Lion, the Witch, and the Wardrobe (2005)

Narnia was never intended to be a Christmas movie, but still it feels like one, for various reasons. Narnia was supposed to be a fantasy movie about four kids in a magic land where an evil witch rules, causing it to be always winter and never Christmas. This is exactly the reason why Narnia is a Christmas movie. Narnia is about bringing back Christmas, there is even a character Father Christmas (Santa Claus).

- The Nightmare before Christmas (1993)

Jack Skellington, the pumpkin king of Halloween Town, is bored with doing the same thing every year for Halloween. One day he stumbles by accident into Christmas

Town, and he likes the idea of Christmas immediately. Therefore he tries to get the resident bats, ghouls, and goblins of Halloween Town to help him put on Christmas instead of Halloween, but this causes a lot of confusion at home.

This is a Christmas movie because the theme Christmas is a very important one. The main character tries to bring Christmas to his own town and wants to share this holiday with the others living in his town, but this doesn't really work out.

- White Christmas (1954)

Bob Wallace and Phil Davis team up to become a top song-and-dance act. Davis plays matchmaker and introduces Wallace to a pair of beautiful sisters (Betty and Judy) who also have a song-and-dance act. When Betty and Judy travel to a Vermont lodge to perform a Christmas show, Wallace and Davis follow and become romantically involved. This is a Christmas movie because the show is a Christmas show, the movie takes place during Christmas and the release of the movie was on 24 of December 1954. It was intended as a Christmas movie.