**Summary**

**1.1 Target Audience and Motivation**

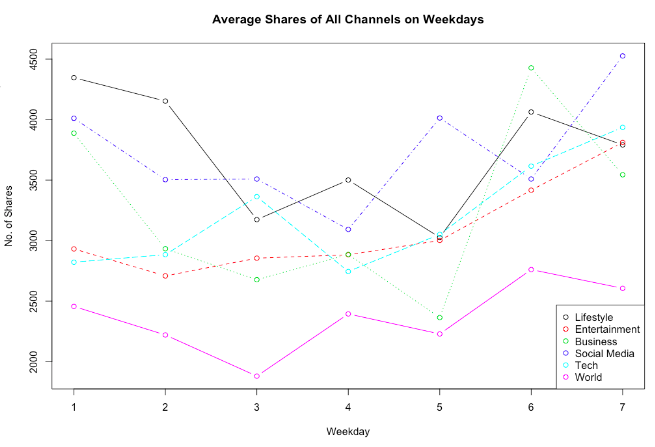
The aim of our analysis is to identify characteristics of an article that significantly affect the number of times it is shared. By identifying these factors, digital media publishing firms will be able to maximize the number of shares on any given article, thereby maximizing popularity of articles. Thus, *Digital Media Companie*s are our target audience for this project.

**1.2 Data Source and Processing**

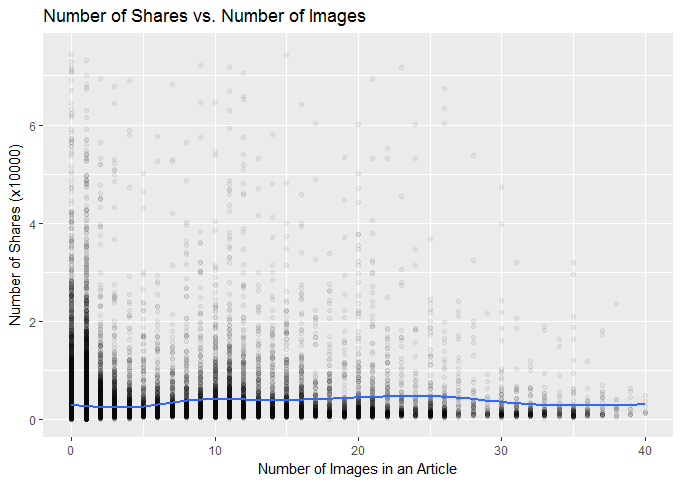
Our data set summarizes features about articles published by Mashable (Lichman 2015), an online news sharing platform, from 2013-2015 The data set has 61 columns and 39,644 rows. We obtained this data set from the archives section of the Machine Learning Repository of University of California, Irvine. Some of the attributes of the data set includes things such as Day Published, Number of Shares, Article Genre (Entertainment, Tech, World etc.) and Number of Words in the Title.

Fortunately, our data set was relatively clean and structured and we did not have issues like missing values or inconsistent data. The data was divided into three subsets containing only the columns required for the three research questions.

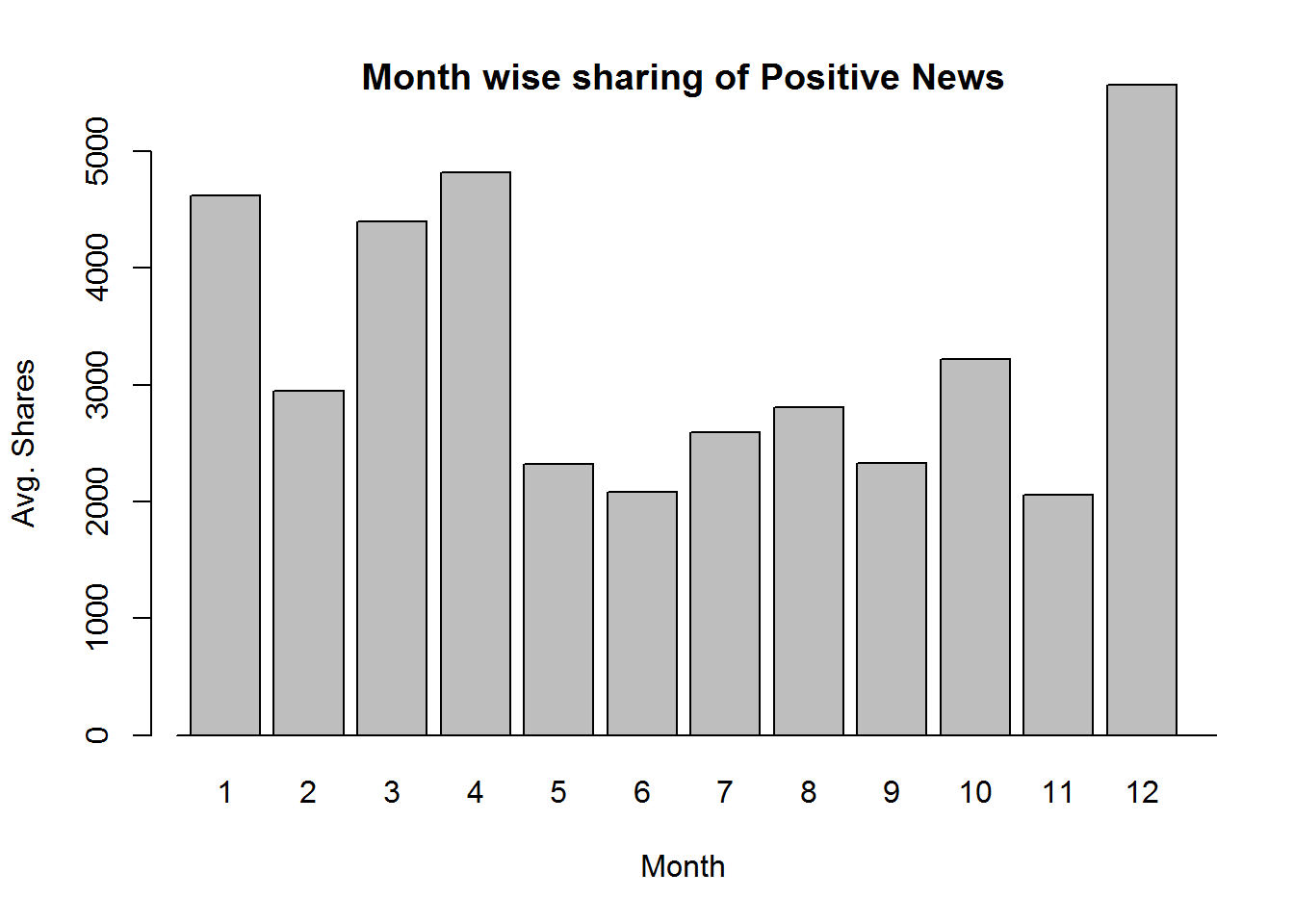
**1.3 Plots**



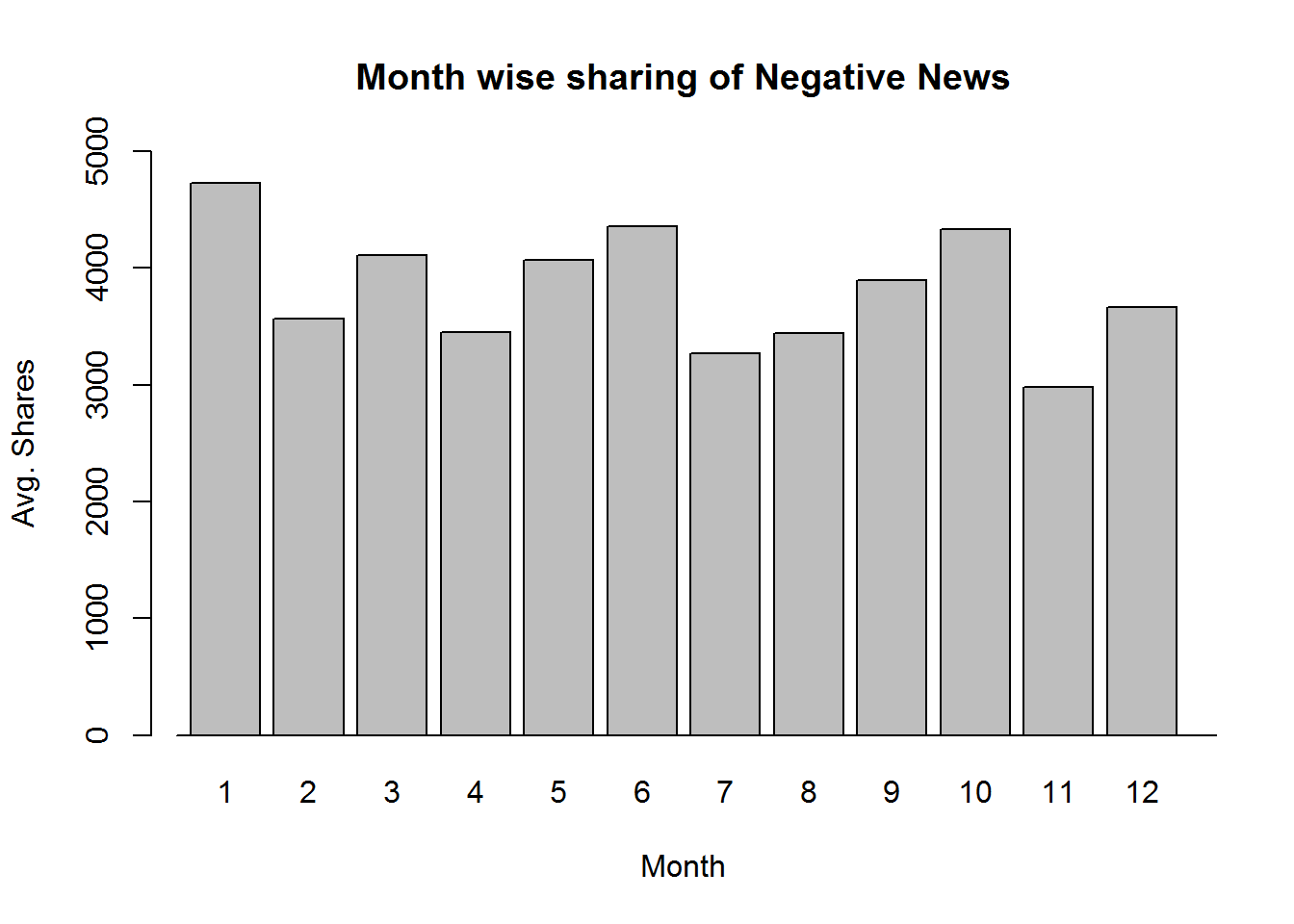
**Figure 1: Best Day to Publish**



**Figure 2 : Optimal number of images**



**Figure 3(a): Polarity of Articles (Positive)**



**Figure 3(b): Polarity of Articles (Negative)**

**1.4 Analysis of Plots**

1. Best Day of Release

Figure 1 shows how the average number of shares of all types of released articles varies within a week. Each colored line represents one type of article. We can see that different types of articles have different curves. However, in general, they are all shared more times on weekend than weekdays.

1. Optimal Number of Images / Words

Figure 2 shows that there is a very slight relationship between the number of images in an article with the number of shares. When we tested the relationship between the number of words in an article with the number of shares it showed that there was no significant relationship. The plot shows that there is a peak in the range of 7-25 images, making that the optimal number of images in an article for maximum shares.

1. Polarity of News

Figure 3 shows the month wise average number of shares for negative and positive news respectively. From the above plots we can clearly see that the maximum negative news is shared in January, followed closely by June and October. Sharing of negative news is on the lower-end in November and December. The number of shares for positive news is maximum in December, followed closely by April.

**1.5 Conclusion**

Based on our analysis we encourage Digital Media Firms for the following:

* To post articles in the weekend for maximizing their shares. To overcome the stigma of publishing articles of only a particular genre as our analysis shows no significant relationship between the genres with the number of shares.
* To include an average number of pictures (7-25) in an article and not in the extremes of too less or too many.
* To publish more positive news in the months of December and April as compared to the other months and to focus more on negative news in the months of January and June. Overall, more resources should be put on promoting negative news as negative news receives higher number of shares on average.

*Word Count - 553*

**1.6 References**

* UCI Machine Learning Repository: Data Set. (2015, May 31). *Online News Popularity Dataset.* Retrieved from https://archive.ics.uci.edu/ml/datasets/Online News Popularity# on Sep 26, 2016 11:15:23 AM.
* K. Fernandes, P. Vinagre and P. Cortez. (2015) A Proactive Intelligent Decision Support System for Predicting the Popularity of Online News. In *Proceedings of the 17th EPIA 2015 - Portuguese Conference on Artificial Intelligence*, September, Coimbra, Portugal.