Keiichiro Watanabe

Data Analysis on Article Popularity

Project Purpose:
Writing a popular news article is a difficult task. Often, a KPI (Key Performance Indicator) for a writer involves getting the most page views and shares in a single article. In the past, I had done an internship as a writer for an online news media company, and my performance was also measured by the popularity of an article. I can speak from my own experience that it is often troublesome for a writer to identify which articles become popular and why they do so. Thus, I thought it would be compelling if I could identify unique combinations of factors that influence an article's popularity. Do certain topics, such as technology, get more attention than others, such as lifestyle or business? What days of the week should we publish articles to get the most shares? How long should a news article really be if we want people to read it? The purpose of this project is to analyze the factors that influence the popularity of online news articles. I aim to identify the key variables and patterns that contribute to an article's success in terms of social media shares.

Data Description and Collection:
I utilized a dataset of Mashable articles written before 2015 obtained from the UCI Machine Learning Repository. This dataset includes 60 explanatory variables and a responsive variable, which is the number of shares an article has on social networks. These variables encompass various aspects such as article content, category, author, and publication details. For more information on data, view the dataDescription.pdf , located in the same folder.

Target Users:
This Shiny App was developed primarily for online news writers and publishers. It can also be valuable to content creators and editors seeking to optimize their articles for increased readership and engagement.

Research Questions:
- What are the primary factors influencing the popularity of online news articles?
- Do specific article topics, such as technology or lifestyle, have a significant impact on popularity?
- Are there optimal days and times for publishing articles to maximize shares?

Insights from Data:
My analysis revealed several significant insights:
- Publishing articles on weekdays (M,T,W,TH) can boost shares compared to weekends.
- The number of sources an article cites does not play a role in predicting its popularity.
- Article topics do not tend to be a useful predictor for how much shares an article gets.
- Articles with shorter headlines (between 7 and 14 words) tend to be more popular.

Keiichiro Watanabe

Description of Design Decisions

What	Why	How
Visual elements, such as colors, styles, and layouts, have been chosen to create an aesthetically pleasing and user-friendly interface. A navy blue background and white text ensure a clean, simple, and readable design. Labels for data visualizations are concise and convey the purpose of each visual.	The primary purpose is to facilitate data analysis on factors affecting online article popularity. The app aims to help users understand which variables influence the popularity of news articles, aiding content creators and writers in optimizing their work.	R packages, including Shiny, ggplot2, and corrplot, are used to create a functional and interactive data analysis tool. Reactive expressions, data tables, and plots are utilized to dynamically present and visualize data. Download links are provided for users to access data and documentation.

Areas for Improvement:

While my analysis yielded valuable insights, there are areas for improvement in the project:

- Combining key variables to create a regression model.
- Integration of real-time data for more accurate predictions.
- Analyzing multiple datasets from other publishers

Sources or References:

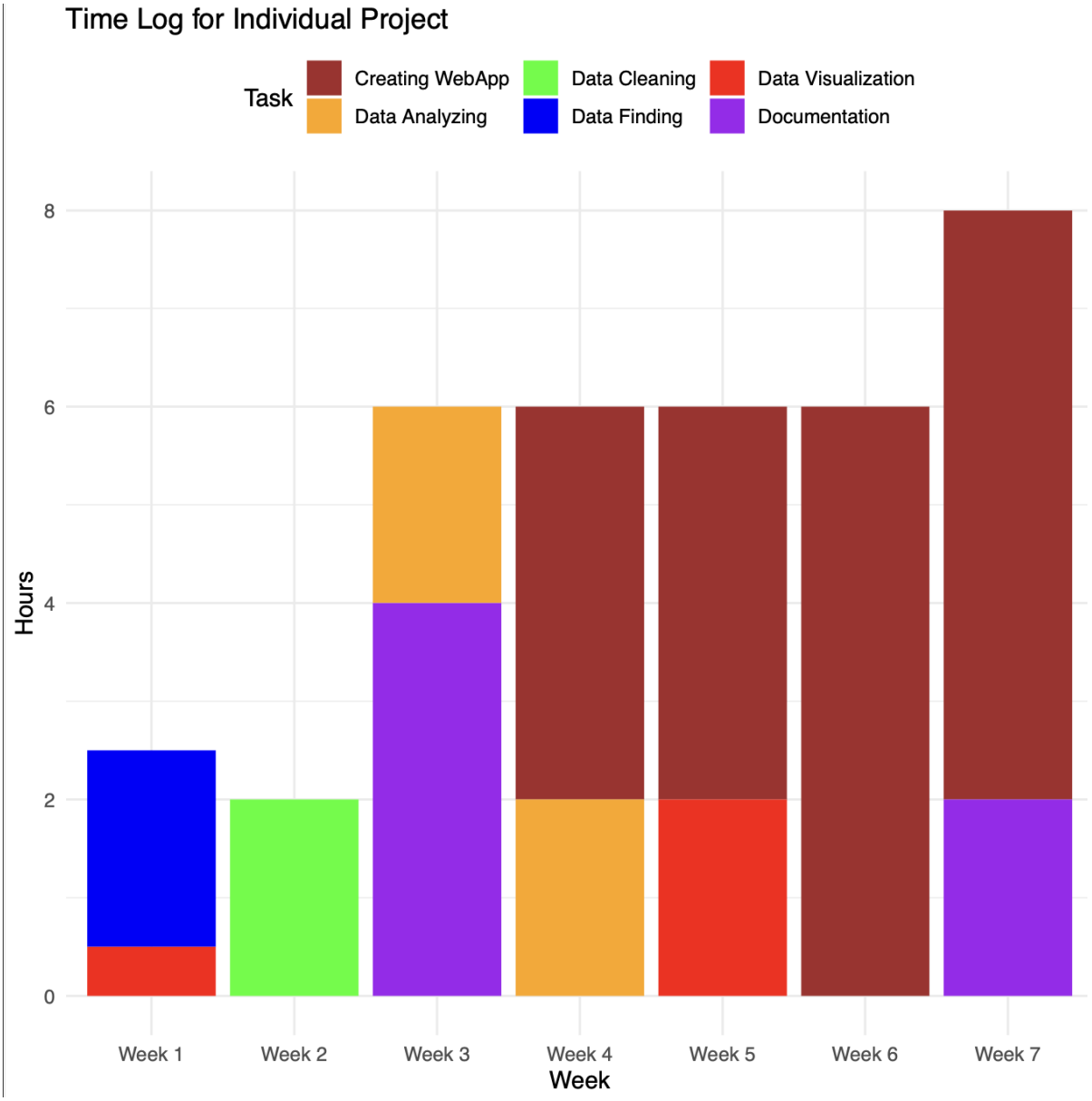
Data: https://archive.ics.uci.edu/ml/datasets/Online+News+Popularity

Berger、J., & Milkman, K. L. (2012). What Makes Online Content Viral? Journal of Marketing Research, 49(2), 192-205.

Dorne, B. (2021). Factors Predicting the Popularity of Online News Articles. Unpublished Manuscript.

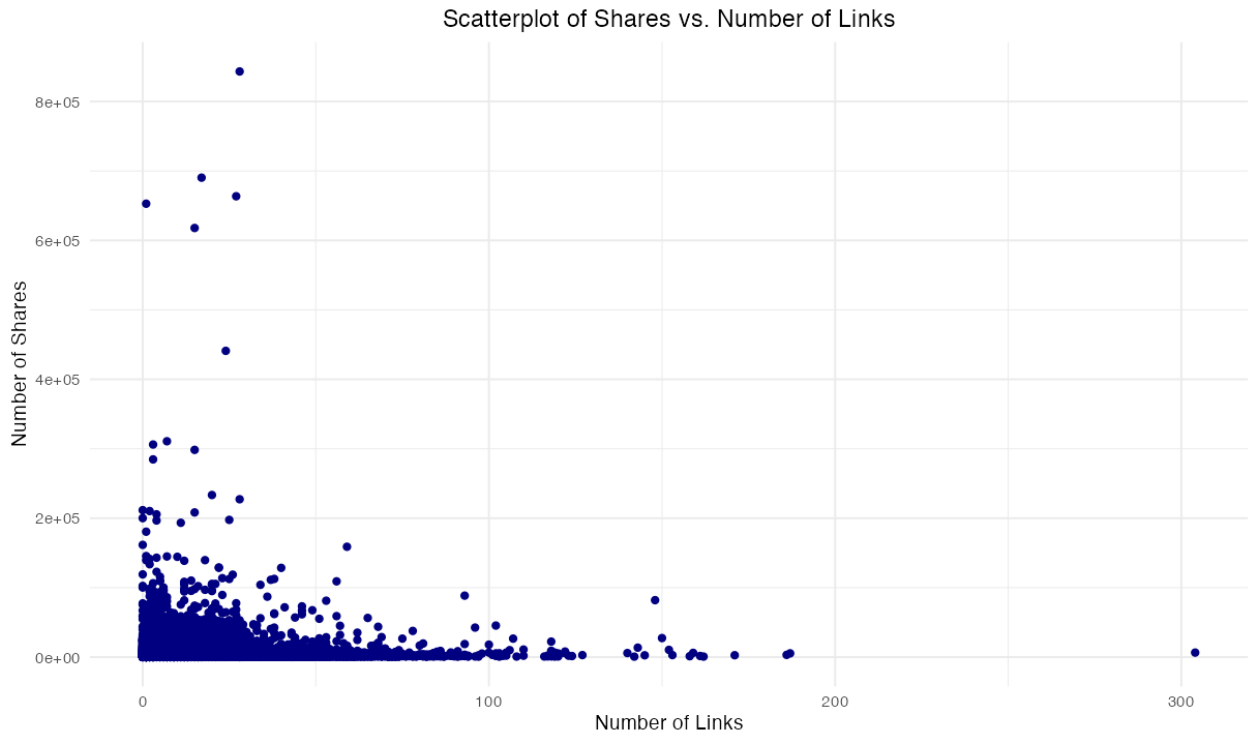
Appendix

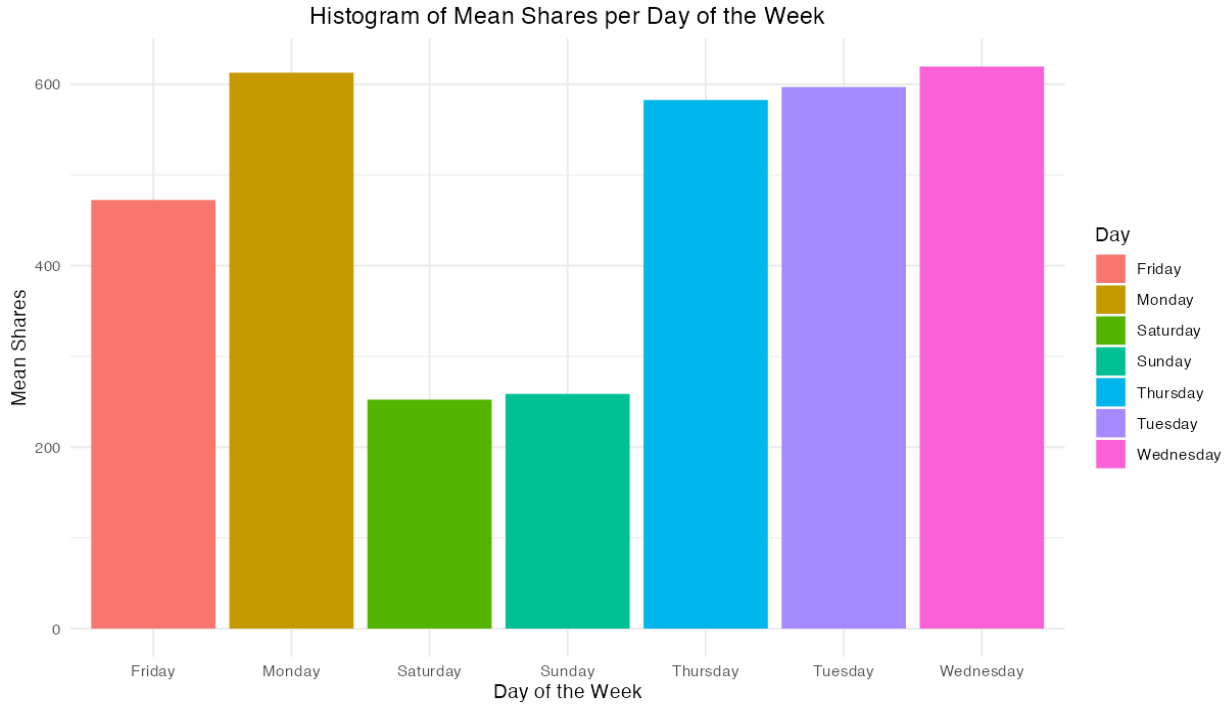
Time Log + Chart in R



Keiichiro Watanabe

DataViz





Keiichiro Watanabe