

Final Paper

Grace DesJardins

2024-12-10

Patterns in Popularity: A Cross-Platform Analysis of Wikipedia and Google Trends

Introduction

For my final project, I web scraped information from the “Most Popular Wikipedia Articles of the Week” webpage from four consecutive weeks (October 6 to 12, 2024; October 13 to 19, 2024; October 20 to 26, 2024; and October 27 to November 2, 2024). In gathering this information, I was curious if, over this four week period, there were any words/themes that appeared that were in common, and how they compared when looking at Google Trends data. With this question of interest, I performed a text analysis on the information that I scraped to find common words/themes, and then performed a Google Trends analysis using those common words/themes.

Wikipedia Web Scraping

During my Wikipedia web scraping, I essentially repeated the same process four times because I had four consecutive weeks to scrape. First, I would read in the given Wikipedia html page as an R object. Then, I would extract the table from the web page. This table had 25 rows because of the fact that it was the 25 most popular Wikipedia articles from a given week, and then 6 columns: Rank, Article, Class, Views, Image, Notes/About. Once I extracted the table, I printed the result and noticed that there were NA’s in the Class and Image columns, which made sense because there were images in both of those columns, which cannot be scraped and were not important to my analyses. Thus, I cleaned out those columns with NA, leaving me with the Rank, Article, Views, and Notes/About columns. As previously mentioned, I did this four times for each article: October 6 to 12, 2024; October 13 to 19, 2024; October 20 to 26, 2024; and October 27 to November 2, 2024. Below are previews of the four tables I obtained from these web scrapings.

October 6 to 12, 2024

```
## # A tibble: 6 x 4
##   Rank Article           Views   'Notes/about'
##   <int> <chr>           <chr>   <chr>
## 1     1 Ratan Tata       5,034,616 "Ratan Tata, a Parsi Indian industrial-
## 2     2 Lyle and Erik Menendez 2,686,453 "Down from the top spot after three we-
## 3     3 Tata family         1,656,736 "The prestigious family of #1, includi-
## 4     4 Joker: Folie à Deux    1,517,588 "Does the description \"jukebox musica-
## 5     5 Noel Tata            1,198,976 "The half-brother of #1, and his succe-
## 6     6 Hurricane Milton     1,005,478 "Less than two weeks after Hurricane H-
```

October 13 to 19, 2024

```
## # A tibble: 6 x 4
##   Rank Article           Views   'Notes/about'
##   <int> <chr>           <chr>   <chr>
## 1     1 Liam Payne       7,904,304 "\"It seems to me that when I dieThese~
## 2     2 Cheryl (singer)  2,011,202 "From 2016 to 2018, #1 dated this sing~
## 3     3 Lyle and Erik Menendez 1,800,441 "People who watch the Netflix show Mon~
## 4     4 One Direction      1,757,832 "In 2010, #1 had just had his second u~
## 5     5 Baba Siddique       1,747,594 "Indian politician Baba Siddique was a~
## 6     6 Lawrence Bishnoi    1,560,755 "Indian politician Baba Siddique was a~
```

October 20 to 26, 2024

```
## # A tibble: 6 x 4
##   Rank Article           Views   'Notes/about'
##   <int> <chr>           <chr>   <chr>
## 1     1 Rodney Alcala      2,490,646 "In 1979, this reprehensible man who a~
## 2     2 Lyle and Erik Menendez 1,468,336 "More murderers on Netflix, albeit wit~
## 3     3 Liam Payne             1,269,238 "The shocking death of the former One ~
## 4     4 Venom: The Last Dance  1,266,811 "Venom wasn't a good case for a franch~
## 5     5 Woman of the Hour      1,242,489 "This American crime thriller film, re~
## 6     6 Deaths in 2024        1,028,425 "Mumbling good morning & raising my he~
```

October 27 to November 2, 2024

```
## # A tibble: 6 x 4
##   Rank Article           Views   'Notes/about'
##   <int> <chr>           <chr>   <chr>
## 1     1 Teri Garr           1,355,055 "This American actre~
## 2     2 2024 Ballon d'Or       1,273,764 "European champion R~
## 3     3 Rodney Alcala          1,258,084 "Netflix brought att~
## 4     4 2024 United States presidential election 1,234,532 "At least it's over?~
## 5     5 Tony Hinchcliffe       1,121,021 "The 2024 Trump rall~
## 6     6 Rúben Amorim           1,110,284 "Manchester United h~
```