Celebrity Hype Dashboard: An R Shiny Web Application

Tigran Melkonian
melkonian.t@northeastern.edu
Northeastern University
Boston, MA
December 6, 2021

***Abstract* — The Celebrity Hype Dashboard is an near real-time awareness trend search tool that provides a graphical representation of the relative scale of public interest (hype) for designated celebrities of interest. The tool allows for the tracking of various celebrities and search terms that have a verified Wikipedia page over time. The primary goal of this feature is to help users understand, track, and compare how much public awareness is behind a specific set of celebrity life events.**

1 Objectives and Background

Humans are social beings who have evolved to dedicate a disproportionate amount of intimate attention to those individuals with the greatest social status as a means of maintaining and guaranteeing their own social success and status. When the general population is constantly exposed to media personas (celebrities), an opportunity arises for fans to develop a sense of identification and friendship. As a result, many celebrities can influence social trends, drive engagement, and generate conversation among their ever-growing audience.

This extreme level of online connectedness feeds a social trend life cycle that is almost entirely controlled by celebrities. Therefore, as a means of staying up-to-date with current trends, fans will reference and interact with multiple online resources such as Instagram, Meta, Twitter, and Wikipedia.

I believe these online interactions across various media platforms can be quantified in-order to better understand the level of awareness behind any given celebrity at any given point in time throughout their careers. In this paper I present a web-application I developed using R, Shiny, and Wikipedia pageview data to generate a Celebrity Hype score that more intuitively allows fans to understand the relative level of awareness behind their favorite celebrities.

2 Proposed Approach

2.1 Data

Wikipedia page views refers to the number of times a particular Wikipedia page has been requested. Using the R package 'pageviews', it is possible to see statistics on how often any Wikipedia page has been viewed during various times, including data as far back as July 1, 2015.

Pageview statistics do not consider how long people have stayed on the article. So, whether they load the article and read from beginning to end, or if they leave it in seconds, it will count as one view.

It is important to note that in some rare cases pageviews might have been purposely manipulated by the use of web-scraping scripts or malware. As a result the pageview counts do not accurately reflect the number of unique visitors a given Wikipedia page has received.

The R package 'pageviews' empowers users with additional view count granularity through providing categorical information on the technology source of a pageview. Therefore, users can query for pageview counts over time with respect to technology source type: desk-top, mobile-app, mobile-web, and all three combined.

2.2 Method and Implementation

Raw daily wikipedia pageview data is transformed to mimic the Google Trends score format. In this section, I outline my unique approach to generating a Celebrity Hype Score that is based on Wikipedia pageview data.

Hype Score Generation and Standardization Process

Search Window

The R Shiny server script initially takes into account a user's designated search window to query relevant pageview data. Generally, due to the variability of pageview counts on a day-to-day basis, for comparative analyses, a large observation window size may help flush out insignificant spikes in the generated weekly hype scores and highlight true variations in hype for a given set of designated celebrities over time.

Score Calculation

In creating the Hype score, the daily pageview data is first grouped by total weekly pageview volume and divided by the total pageview volume of the trailing search window designated by the user. These ratios are finally rescaled on a range of 0-100 based on a week's respective pageview volume proportion to the total pageview volume in the designated trailing window.

Tech

I developed the Celebrity Hype Dashboard web app using R, Shiny (R package for web-development), pageviews (R package for accessing Wikipedia pageview data), rsconnect (R package for remotely deploying and updating Shiny web-applications to shinyapps.io), and shinyapps.io to host and serve my Shiny application for free.

3 Sample Use-case

Step 1

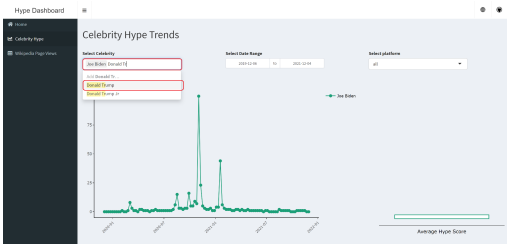
Navigate to the Celebrity Hype Dashboard following this link: https://tigranmelkonian.shinyapps.io/celebrity_hype_dashboard/. You will initially be directed to the home page of the dashboard containing relevant information about the web-application, Wikipedia pageviews data source, and methodology used in generating the Hype score metric.

Step 2

Navigate to the 'Celebrity Hype' tab. Here you will be able to intuitively query for your celebrities of interest with respect to data platform source and observation window. The trend of Hype score and average hype score over time are plotted below as a trendline and bar plot respectively.

Step 3

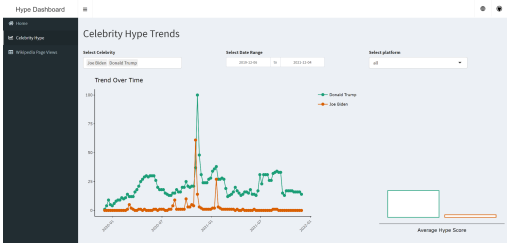
By default the application will load data from Kim Kardashian and Kanye West. Please clear these inputs and search for your celebrity/ celebrities of interest. You will have the option of choosing from over 20000 preloaded celebrities. If you are not able to query for your desired celebrity, then please make sure that they have a valid wikipedia page, if they do then you can simply add them to the list when prompted to do so through the search bar.



Step 4

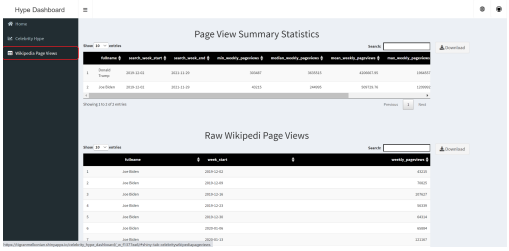
Please allow some time after selecting celebrities of interest for the data to load and for the visualizations to manifest on the dashboard. This process should not take

more than a few seconds. Note the new Celebrity Hype trend line over time and average hype score bar plot adjacent. In this example we can see that Joe Biden had a Hype score of 61 and Donald Trump had a Hype Score of 37 during the week of the presidential elections. However, Donald Trump persistently experiences a greater level of Hype over time. This is probably due to the fact that he is an actual social celebrity and has been president for the past four years (among other reasons).



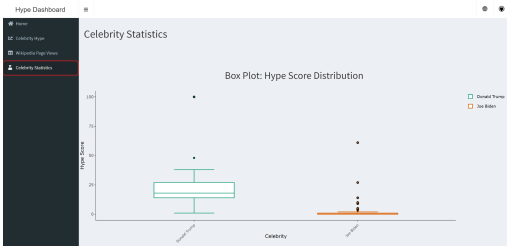
Step 5

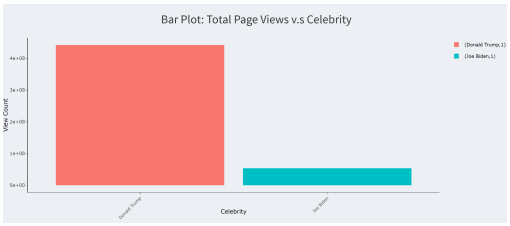
Next navigate to the Wikipedia Page Views tab for access to both descriptive summary statistics and full data output for the celebrities you queried for on the previous tab.

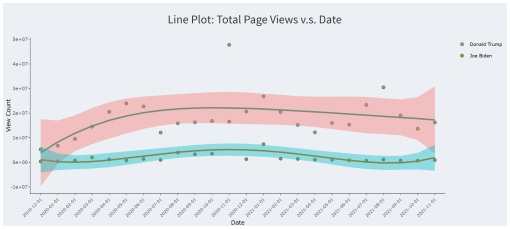


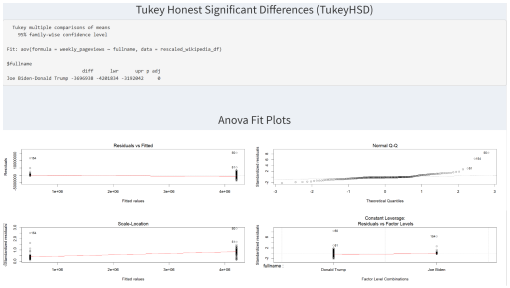
Step 6

Finally navigate to the Celebrity Statistics tab to access multiple data visualizations that summarise both Hype score and pageview counts with respect to user designated celebrities including: a box plot for Hype Score distribution, a bar plot for total pageview counts, a linear plot fit to quarterly pageview totals, Tukey Honest Significant Differences table, and summary plots of the N group anova fit (where N is the number of celebrities designated by the user on the Celebrity Hype tab.









We can note that for the given example comparing Joe Biden and Donald Trump, the adjusted p-value for the difference between their weekly pageviews with respect is zero. Meaning there exists sufficient evidence that the difference between weekly pageview counts is significant, i.e. not zero. In this case Donald Trump has a significantly greater level of awareness over time given the designated time window of 2019-12-06 - 2021-12-04.

5 References

[1] Celebrity Hype Dashboard

[2] Celebrity Hype Dashboard GitHub Repository

[3] FAQ about Google Trends data - Trends Help

[4] Wikipedia:Pageview statistics - Wikipedia

[5] Wikipedia