Samuel Wolfe

September 19, 2023

MSBA 212-70

Homework 1

1. I searched for ‘tim gunn’ using the example code provided during class and with Octoparse 8. I managed to get both to return similar enough data to make a comparison applicable. Neither the supplied code nor the automated “non-code” method provides reaction data. After looking into Pinterest, it appears that it has some type of analytics data, but they are data Pinterest supplies to businesses already. Check the stats info here [Pinterest Stats](https://help.pinterest.com/en/business/article/pin-stats).
2. Table 1 below.

|  |  |  |
| --- | --- | --- |
| Table 1 | Title | Description |
| Minimum | 7 | 13 |
| Mean | 45.85 | 139.12 |
| Maximum | 91 | 415 |
| Standard Deviation | 23.86 | 106.87 |

1. See file example8results.csv.
2. \*grams
   1. Index 35 (high)
      1. Most Common Description Unigrams: [(('.',), 4), (('project',), 3), (('runway',), 3)]
      2. Most Common Title Unigrams: [(('wedding',), 1), (('gown',), 1)]
      3. Most Common Description Bigrams: [(('project', 'runway'), 3), (('runway', 'bridal'), 1), (('bridal', 'gown'), 1)]
      4. Most Common Title Biigrams: [(('wedding', 'gown'), 1)]
      5. Most Common Description Trigrams: [(('project', 'runway', 'bridal'), 1), (('runway', 'bridal', 'gown'), 1), (('bridal', 'gown', 'inspiration'), 1)]
      6. Most Common Title Trigrams: []
   2. Index 4 (low)
      1. Most Common Description Unigrams: [(("n't",), 1), (('call',), 1), ((',',), 1)]
      2. Most Common Title Unigrams: [((':',), 1), (('project',), 1), (('runway',), 1)]
      3. Most Common Description Bigrams: [(("n't", 'call'), 1), (('call', ','), 1), ((',', 'long'), 1)]
      4. Most Common Title Biigrams: [((':', 'project'), 1), (('project', 'runway'), 1), (('runway', "'s"), 1)]
      5. Most Common Description Trigrams: [(("n't", 'call', ','), 1), (('call', ',', 'long'), 1), ((',', 'long', 'fits'), 1)]
      6. Most Common Title Trigrams: [((':', 'project', 'runway'), 1), (('project', 'runway', "'s"), 1), (('runway', "'s", 'ashley'), 1)]
   3. When I search for ‘tim’, none of these do. Once I make it to ‘tim g’ the search suggests ‘tim gunn’. After that the only two suggestions are ‘wardrobe essentials’ and ‘quotas’.
3. Start with Table 2. This is using the Gunnings-Fog Index, as it was the simplest for me to understand.

|  |  |  |
| --- | --- | --- |
| Table 2 | Values | Notes |
| Minimum | 0.4 | Index 41 |
| Mean | 8.42 |  |
| Maximum | 18.56 | Index 14 |
| Standard Deviation | 4.21 |  |

Looking at the min and max scores, we can see quickly that the min is mostly micro links to other websites, which do not translate very well into comprehensible data. The maximum on the other hand is proven to be descriptions taken directly from published material. It makes sense that the average would fall around 8.5 and the std would be around 4. That puts the first ranges for our mean at below 6th grade and around high school, which is around the average US reading level. This is obvious because the created descriptions at the higher end are by the companies involved in the shows Tim Gunn participates in and they are seeking an audience desiring to seem fashionable and smart. While the average person on Pinterest, a website about visuals, may not have the same level of vocabulary.

1. Regarding Reactions, I tried Apify and Octoparse to collect reactions via scraping. I could not for the life of me figure out how to get either of these programs to scape the data. I got farthest with Octoparse in that I manually crated a step by step “click through” instructions for the scraper that would only ever grab the reactions to the first icon. I recognize that you may be able to or there may be scrapers on Apify that can do this, but every single one is locked behind a paywall at this point for me.
2. Please review file “swolfe Homework 1 video”.