WWW Analyst Assessment

Brianna Lytle

- **Part 1: Who is the main audience? What demographic segment represents our core audience?**

```
1  www_fe['Age'].value_counts(normalize = True)
```

```
25-34    0.310182
35-44    0.210472
45-54    0.181530
55-64    0.160800
65+      0.083896
18-24    0.053122
Name: Age, dtype: float64
```

Figure 1 - Percentage spread of age groups in WWW Dataset

Age Group: Just over 50% of the audience is between ages 25-44. About 85% of the audience is between 25-64. The age groups 18-2 and 65+ has the least amount of audience members. *(Figure 1)*
Gender: About 89.5% of the visitors ere females and 10.5% were males.
Title Reference: The most popular Title reference of articles viewed across all demographics is from the original 'Who What Wear'. The least popular Title reference among all demographics is "The Thirty." However, this title reference is primarily read by people in the 25-34 age group.
Focus: For this assessment I will focus on the 4 most popular age groups. In Figure 2, we see these age groups consistently have the highest page views for each day in June 2019 and follow a similar pattern during the month. The youngest and oldest age groups do not follow this tend and are mostly consistent during the month.

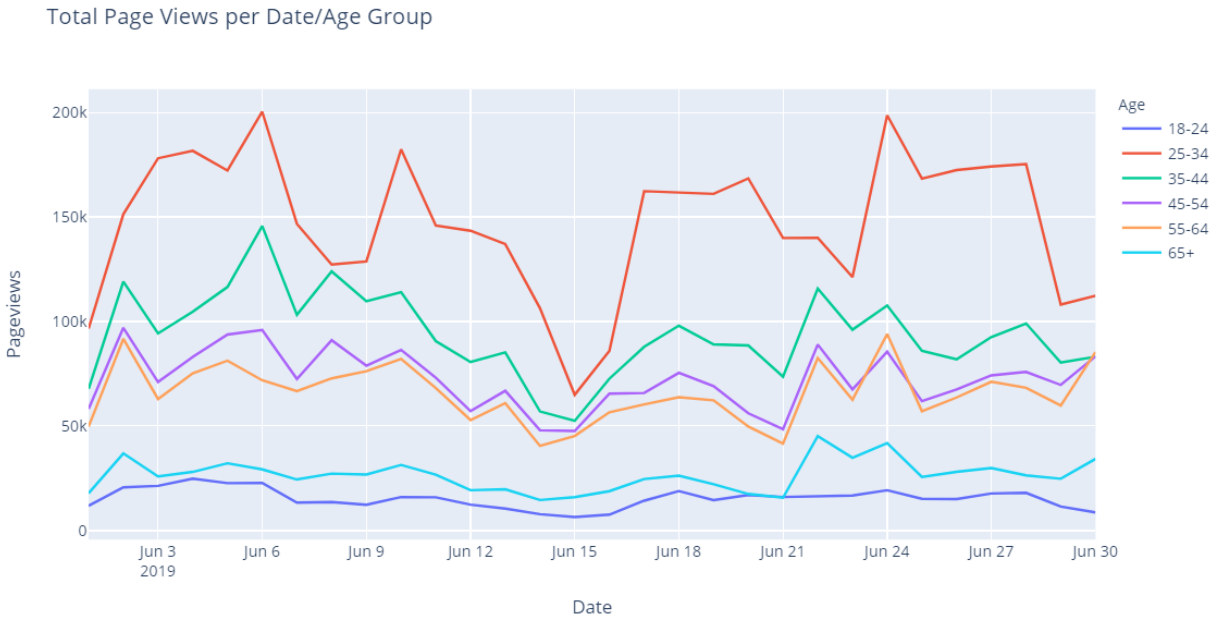

Figure 2 - Total Page Views in June 2019 based on Age Group

- **Part 2: What content resonates the most with this audience? What are the themes and frames that drive engagement?**

Note: At this part of my analysis I created a new dataframe that only contained the age groups between 25-64. This deceased the size of the dataframe from 32,755 rows to 28,267 rows.

NLP: I performed a quick analysis on the most common terms used in article titles. As expected, the most common titles are related to “fashion” and “summer”. Common titles include explaining the “x best” in fashion. In other words, lists are very common article titles. This analysis was also performed between the articles with the top 25% viewed articles, however, there wasn’t much difference. The main difference is that the top 25% of viewed articles in this age group shows more popularity of 'summer fashion' trends. This doesn't necessarily explain much about the audience but more about the time of year.

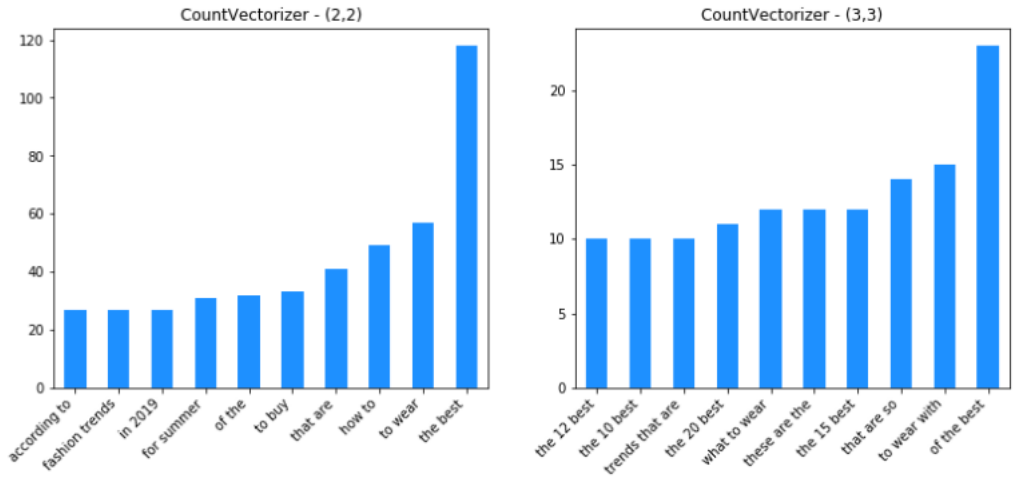

Figure 3- Most common bi- and tri-grams for audience focus

	date_count	Pageviews	Unique Pageviews	upv_average
date_count	1.000000	0.302855	0.294224	-0.118101
Pageviews	0.302855	1.000000	0.998311	0.132740
Unique Pageviews	0.294224	0.998311	1.000000	0.144413
upv_average	-0.118101	0.132740	0.144413	1.000000

Figure 4- Correlation between values in article_analysis.csv

Date count: I created a pivot table which counted the amount of days and article was viewed. More than half of the articles were not viewed more than 3 days of the month. 65 out of the 1088 articles (5.9%) in the article_analysis.csv were viewed more than 15 times in the month. 13 of those articles were viewed for the entire 30 days of the month. In figure 4, we see that the amount of days an article is read doesn’t necessarily impact the total pageviews the article will receive.

Pageviews vs. Date count content: There is a slight difference between the top 20 articles with the highest pageviews compared the articles that were consistently viewed during June 2019. The 6 of the top 20 pageviewed articles mentioned a celebrity name and 4 of the articles contained some sort of list (ex. '7 Outdated Sandal Trends..'). On the other hand, out of the top 20 daily viewed articles, only one contained a celebrity name and 9 contained lists. In addition, the articles with the highest pageiews have a higher upv_average while the most common daily viewed articles have a low upv_average. This shows us readers are most likely searching for these list-related

articles more than once during the month. The articles with a high viewership are articles that a person would only read once.

- Part 3.1: What are the channels that are driving engagement?

Most popular Medium/Source by age Group:

- Age 25-34:
  - Among the highest page viewership, the medium with the highest unique page viewership percent is a "paid" source - 308 views, 96.7% unique.
- Age 35-44:
  - Among the highest page viewership, the medium with the highest unique page viewership percent is a "facebook" source – 259 views, 91.9% unique.
  - Suggestion: To target this demographic, affiliate marketing with other online fashion/beauty magazines. Continue strong SEO. Most overall page views come from organic/referral sites as well as other online beauty magazines.
- Age 45 – 54:
  - Among the highest page viewership, the medium with the highest unique page viewership percent is a "facebook" source – 224.8 views, 90.4% unique.
  - Note: If it's not in their email they probably won't bother to look for an article with WWW
- Age 55-65:
  - Among the highest page viewership, the medium with the highest unique page viewership percent is a "facebook" source – 198 views, 86.4% unique.
  - All mediums with a unique page viewership above 90% had an overall average total page view less than 100.

Where are most of the articles being source from?

- Google - 29.7%
- m.facebook.com – 16%
- newsletter – 15.9%
- from.flipoard.com – 11.5%
- www-who-what-wear.cdn.ampproject.org – 9.3%
- pinterest.com – 9.3%
- yahoo.com – 4.1%

What is the most popular medium for articles?

- Referral – 52.37%
- Organic – 29.72%
- Email – 15.93%
- (none) – 1.3%

What are the major sources overall?

- Other- 39.6%
- Google related source – 32.1%

- Facebook related source – 16.4%
- Pinterest related source – 6.4%
- Yahoo related source -5.1%
- Instagram related source - <1%

Figure 5 shows us most of our page views come from a referral medium (52.37%). This means that visitors clicked from another site to reach the WWW article. The second most popular medium is organic (29.72%) which is mostly displayed in the Google source. This makes sense since Google is the most popular search engine. The third most popular medium is email (15.93%). The only Source that has a medium source is 'newsletter'. We can assume this audience comes from those who clicked articles from email newsletters

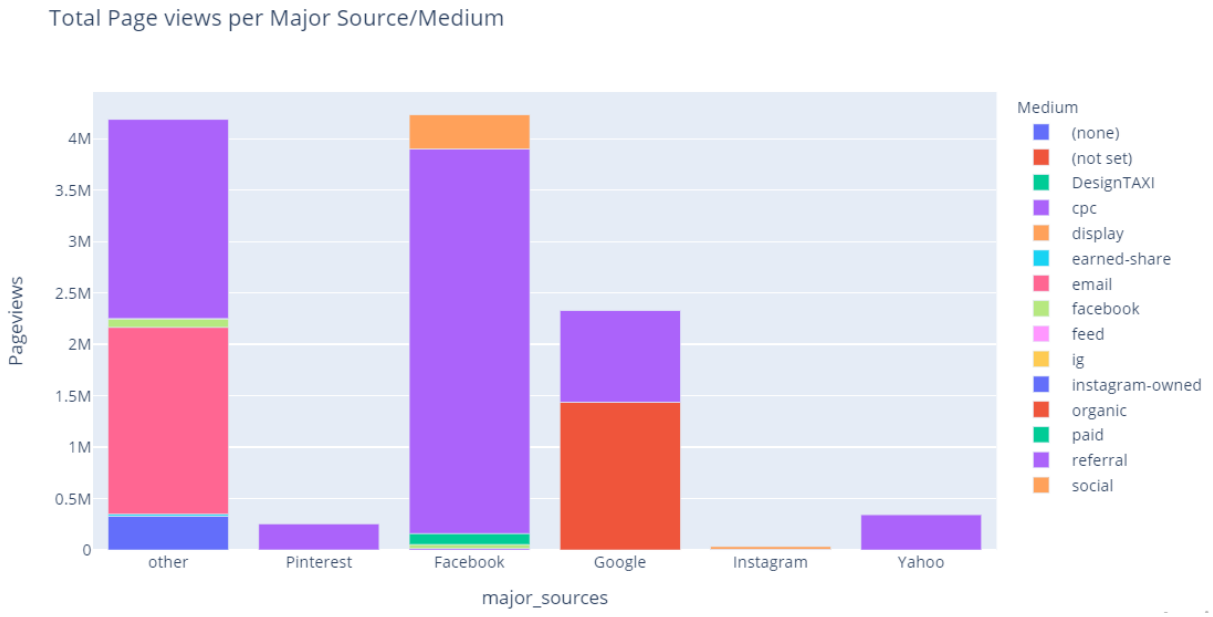

Figure 5

- **Part 3.2: Which traffic sources should we lean into, which traffic sources should we deprioritize?**

“Lean Into”:

- Celebrity articles - These articles do have a significantly high page viewership. Though they are not consistently viewed during the month, it does produce a high amount of unique page viewers. This can be a strategy to produce name exposure.
- SEO - Search engines are powerful. Article content for search engines can be related to the time of year. For example, the articles containing “Summer trends” and “Summer beauty” were among the top viewed articles of the month and appeared in the top 20 articles with a high date count.
- Flipboard – Though this was not a highly significant source for ages 25-34, it is a popular source for ages 35-64. For ages 35-64, there were more page views from a Flipboard source than the combined google sources.

Deprioritize:

- Instagram – This is not a successful source in producing pageviews. In general, Instagram is not a place that users go to for articles. Based on Figure 5, we can see our audience is lead to our website through Facebook, emails, or search engine results. Facebook and email platform(s) are article friendly. An audience would use a search engine if they are seeking a more describe read. It is difficult to gain users from Instagram because you cannot have a hyperlink in a caption. It must be in a bio or through an ad button.
- CPC/Facebook (medium/source) – The page viewership is only 13,927 (less than 1% of total audience pageviews). Especially if the CPC is high, this is not a useful channel and efforts would be useful elsewhere.
- Earned-share mediums – Page viewership is 16,491 (less than 1% of total audience pageviews) and a 78.3% unique page viewership – if WWW is spending a significant amount on media sharing, then efforts would be useful elsewhere.
- Design Taxi – Total page viewership is 1,261.

Next Steps:

- Perform clustering analysis
- More analysis on themes of articles
- Investigate themes of articles from specific sources