GMM Post-level cluster analysis. Based on GMM\_EDA\_PostLevel.ipynb

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Cluster Number  Analysis Topic | **0** | **1** | **2** | **3** | **4** | **5** |
| Distribution of Social Platforms  Standard deviation is quite high (2x the means) |  | Lowest mix of Facebook, mostly Instagram/Twitter |  | Highest mix of Facebook, though not meaningfully higher | Highest mix of Twitter |  |
| Age/active years: average age is generally ~38 and standard deviation is very high across all clusters, so unlikely a big differentiating factor |  |  |  |  |  | Seems to have a little older profile than others. |
| Count of Followers by platform.  The standard deviation is quite high, so any differences are not very meaningful. | Has highest number of avg. Instagram followers. |  | Has highest number of avg. Instagram followers. | Has highest number of avg. Instagram followers. | Has highest number of twitter followers | Has highest number of facebook followers |
| Time of day posts  Quarter 1 = 12am-6am  Quarter 2 = 6am-12pm… | Posts most frequently quarter 4,3 | Posts more often in quarter 1 than others |  | Posts more often in quarter 1 than others | Posts most frequently quarter 3 | Posts most frequently quarter 4 |
| Genres  Differences likely not significant |  |  |  | R&B |  |  |
| Description Length/Hashtag Count  No meaningful differences |  |  |  |  |  |  |
| Post Type  Photo is by far the most popular, followed by link and video | Highest proportion of photo |  | Most video posts |  | Highest proportion of text, also high photo, low video. |  |
| Cluster Number  Analysis Topic | **0** | **1** | **2** | **3** | **4** | **5** |
| Number of days to track release.  Again, very high standard deviation (2x mean), so interpret w/ caution |  |  | Has highest number of days (96) | Has lowest number of days (71), so tends to post closer to track releases. |  | Has highest number of days (100) |
| Time since last post  High std dev and means are all fairly close to each other. |  |  | Tends to post more frequently. |  | Tends to post less frequently | Tends to post more frequently. |
| Correlations between features and social engagement scores  Higher post frequency led to lower score  Photo posts -> pos corr; others -> neg, especially links  Proximity to track release has almost no correlation with engagement | Posting in the evening slight positive corr  Shorter posts slight pos corr  R&B genre pos corr | Posting 12-6pm slight positive corr.  R&B/Pop/Electronic genre pos corr  More affected by post frequency  Benefits most from photo posts | Posting 12-6pm slight positive corr.  Shorter posts slight pos corr  More affected by post frequency | Posts in the morning slight positive corr.  Shorter posts slight pos corr  Rock/Electronic genre pos corr  Less affected by post frequency  Benefits most from photo posts | Posting 12-6pm slight positive corr.  Shorter posts slight pos corr  Rock genre pos corr  Benefits most from photo posts | Posting late morning slight positive corr.  Electronic/Rock/Pop genre pos corr |
| Best platform? | Better on Twitter than other clusters | Better on Instagram than other clusters |  |  | Better on Instagram than other clusters | Better on Facebook than other clusters |
|  |  |  |  |  |  |  |

Overall takeaways:

The clustering approach used resulted in six clusters with a relatively low silhouette score (.287). This means that the model has a difficult time creating clear delineations between the various clusters and that there is quite a bit of overlap. As such, many of our findings indicate that social media and top-level artist level features are insufficient to properly cluster different types of posts on social media. While analyzing the results, we found that the clusters saw relatively similar profiles across a number of attributes and that most differences would not be significant due to a high standard deviation. However, we were able to come away with the following high-level descriptions of the clusters:

Cluster 0 posts most frequently in the 2nd half of the day, tends to have a higher than average number of followers on Instagram, and posts photos more frequently than other groups. Posting in the evening has a slight benefit to their social engagement scores, as does being an R&B artist. They may be underutilizing Twitter as posts there tend to get higher engagement than other clusters.

Cluster 1 uses Facebook sparingly and tends to post in the early AM hours. Posting in the evening is probably underutilized. They should also post less frequently as that negatively impacts their social engagement scores more than other clusters. They benefit the most from posting photos, particularly when they use Instagram.

Cluster 2 tends to have a higher than average number of followers on Instagram, posts more videos, tends to post more frequently. However, shorter length posts have a slight positive correlation with social engagement. They should consider posting less frequently as they may be detrimental to average engagement.

Cluster 3 also tends to have higher than avg. number of followers on Instagram, posts frequently in the early morning and posts closer to track release dates. Early morning posts are effective, as are photo posts. They could do shorter posts more frequently.

Cluster 4 does better on Instagram than most other clusters and similar to Cluster 3 benefit from shorter/photo posts. They tend to post less frequently than other groups and benefit from posting in the afternoon. They use twitter a lot, but should consider Instagram.

Cluster 5 does better on Facebook than the other clusters and seems to have a slightly older profile. They post late in the evening, though posting later in the morning might be more effective.