MRKT 671 - Group 3 - Social Analytics Report

1. Why is engagement with social media content important for brands?

In recent years, social media has become the mainstream marketing channel of an increasing number of brands, invoking a shift in evaluation methods of performance of brands from traditional ways. Except for hard marketing metrics such as acquisition and purchases, soft metrics such as consumer attitude are also crucial in the evaluation of performance, and require separate evaluation methods when social media is the main marketing channel. On online platforms where brands invest great time and money to attract customers, user-generated content (UGC) on social media is regarded as a source that reflects "the wisdom of people" and enables analysis of customers' perceptions of a brand[1].

On consumers' side, social media content has a critical influence on their views of a campaign, a product, or a brand. News and commentary on social media postings are replacing traditional media as an information source, bringing about dramatic changes in the ways we interact with friends and family, learn about products, services, jobs; and think about news, politics, and religion[2].

The "paradigm shifts" on the brands' and consumers' sides both increase the importance and indicate the merits of engaging in social media content. It allows brands to understand customers' perceptions and delineate social media content-oriented marketing strategies flexibly and effectively. Additionally, an in-depth study on social media content and UGC can reveal patterns and relationships between online behaviors and preferred consumer experience. Because UGC and social media content demonstrate spontaneity—"your customers are your next ambassadors"—, a wise engagement strategy in online platform marketing would help to foster loyal customers and a self-sustainable online community that maintains a positive perception of the brand.

---

[1] Colicev, A., Kumar, A., & O'Connor, P. (2019). Modeling the relationship between firm and user generated content and the stages of the marketing funnel. International Journal of Research in Marketing, 36(1), 100–116. https://doi.org/10.1016/j.ijresmar.2018.09.005

[2] Luca, M. (2015). User-Generated Content and Social Media. Handbook of Media Economics, 563–592. https://doi.org/10.1016/b978-0-444-63685-0.00012-7

2. How accurately can engagement be predicted? What variables are most useful?

We first encoded categorical variables based on its natural, then applied dimension reduction due to multicollinearity using random forest. With 38 selected features, we explored multiple machine learning algorithms and decided not to take the linear regression approach as the actual data points are largely spread out from the line of best fit due to non-linearity. Below is the residual plot based on a linear regression model with standardized data. This nonlinearity might cause many regressors not working well with our data, thus, we should consider fitting nonlinear regressors, or group the target variables and perform classification instead.

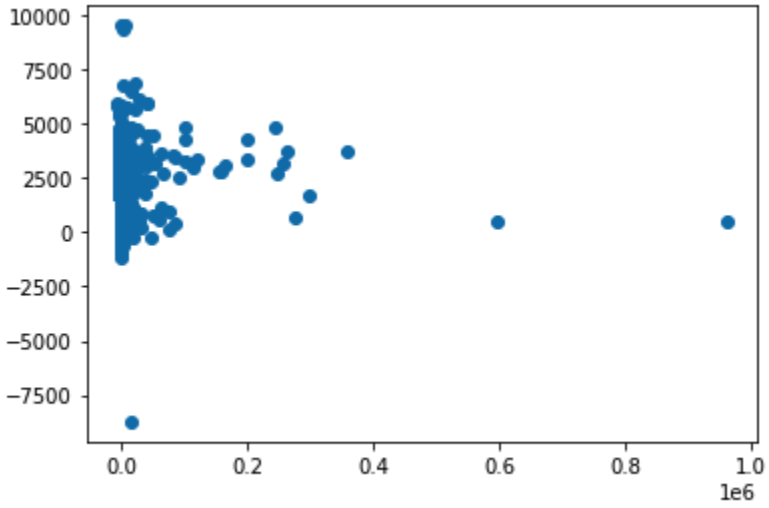

Exhibit 1 - Linear Regression Residual Plot

The best score mean absolute error we obtained was 455.21 with neural network via KerasRegressor, which added layers of weights multiplication to attributes with a non-linear activation function. The KerasRegressor we selected is based on the Tensorflow sequential model with a single fully connected hidden layer using the same number of neurons as input attributes. The following table is top 10 features observed from the feature importance retrieved using Neural Network's weights assigned to each attribute.

Weight	Feature
0.0163 ± 0.0013	Followers.at.Posting
0.0116 ± 0.0004	crosspost
0.0026 ± 0.0006	WC
0.0017 ± 0.0005	space
0.0015 ± 0.0007	see
0.0013 ± 0.0004	Message_len
0.0012 ± 0.0009	link_title_len
0.0011 ± 0.0001	ipron
0.0010 ± 0.0004	Live Video Complete
0.0009 ± 0.0004	focuspast

Exhibit 2 - Neural Network Feature Importance Based on Weights

Based on our best prediction model, extrinsic factors such as number of followers and likes demonstrated high importance; this represents a page's existing impact and potential amount of audience involved. If a page has a solid follower base, then it's more influential and possible to gain lots of post engagements. Most of our candidate models did not indicate high importance for when and who exactly made a post. A page's existing exposure and audience base are significant extrinsic factors to consider. Crosspost showed high importance as well, indicating the format of posting is significant. However, intrinsic factors took most place in the top 10 features. These attributes demonstrated the importance to look into intrinsic factors in detail to fully understand their impact both individually and together with extrinsic factors.

Regardless, the team used a random forest for better interpretability of feature importance due to its ease-of-use and performance on larger datasets.

3. What are the factors people engage with social posts? Extrinsic or intrinsic?

Our best model from the ensemble method with dimension-reduced data is a Gradient Boosting Regressor. This model produced a decent mean absolute value of 501. We noticed that both intrinsic and extrinsic attributes presented in top important features. Its output has many features overlapping with the Neural Network model, such as follower count and crossposting; but it also highlights content-related attributes, which lead us to further analyze the impact of intrinsic attributes.

Top 10 Feature Importance, Presented by Gradient Boosting	
attribute	importance
AllPunc	0.304631
Message_sentiment	0.171203
crosspost	0.132782
you	0.105025
see	0.052229
Likes.at.Posting	0.045861
Followers.at.Posting	0.030685
home	0.023611
social	0.020899
focusfuture	0.020775

Exhibit 3 - Feature Importance for Intrinsic/Extrinsic Factors

Following from this, two Random Forest models were run on the entire dataset for intrinsic & extrinsic factors respectively, in an attempt to identify the most important features within each that contribute to engagement rates; measured here by Total Interactions (i.e. the sum of likes, comments, shares and all other reactions).

Where extrinsic factors are concerned, it seems one of the most important factors (for Video content) is the number of views on the post. The length of a post's message, link text and the hour in which it was posted are the next most important factors, perhaps an indication that more attention should be paid to ensure these descriptive texts are succinct yet captivating enough to encourage engagement. Timing is also relevant here, likely pointing to content posted during 'peak' times (such as in the evening) when individuals would be more likely to spend time browsing Facebook. Numbers of likes/followers at time of posting then follow as next in

importance, which is not unreasonable, as already-popular pages are likely to have some sort of implied 'credibility' due to groupthink.

As for intrinsic factors, following an initial filter of LIWC attributes based on intuition, the most important factors where post content is concerned turns out to be the use of words that indicate analytical thinking, describe social processes (i.e. interpersonal relationships) and/or discuss drives (i.e. strong, impactful words often associated with achievement, power, risk and reward). This finding seems to indicate a preference among people for insightful content. It's no surprise that words describing social processes are highlighted as important as well, as it's a natural human instinct for one to seek connections[3]. Finally, posts that include strong, motivational content are likely to garner interactions as well, especially if these facilitate the process of self-actualization in individuals.

[3] https://www.scientificamerican.com/article/why-we-are-wired-to-connect/

4. Where are the implications for social media marketing managers?

To identify one interesting brand with potential, we ran a K-Means clustering model on all brands and used a random forest to select the most important features in determining the clusters. Through clustering, brands with similar posts and corresponding engagement were highlighted and this allowed us to analyze how consumer perception varies with different marketing strategies. We generated 6 clusters each with varying size and characteristics. Initially we intended to pick the outliers such as the fourth cluster which consisted of only one brand, 'Tasty' to assess how it differentiates in its marketing strategy. However, through further analysis, it was determined that despite promising customer interactions, it was a new brand and so had less depth to offer. Instead, we decided to shed light on the general advertising trends in the market by focusing on the characteristics of the largest cluster: 0; which consisted of 94 brands. From here, we selected Tech Insider, the brand with the highest value for number of Likes.to.Posting which was the most important feature identified earlier.

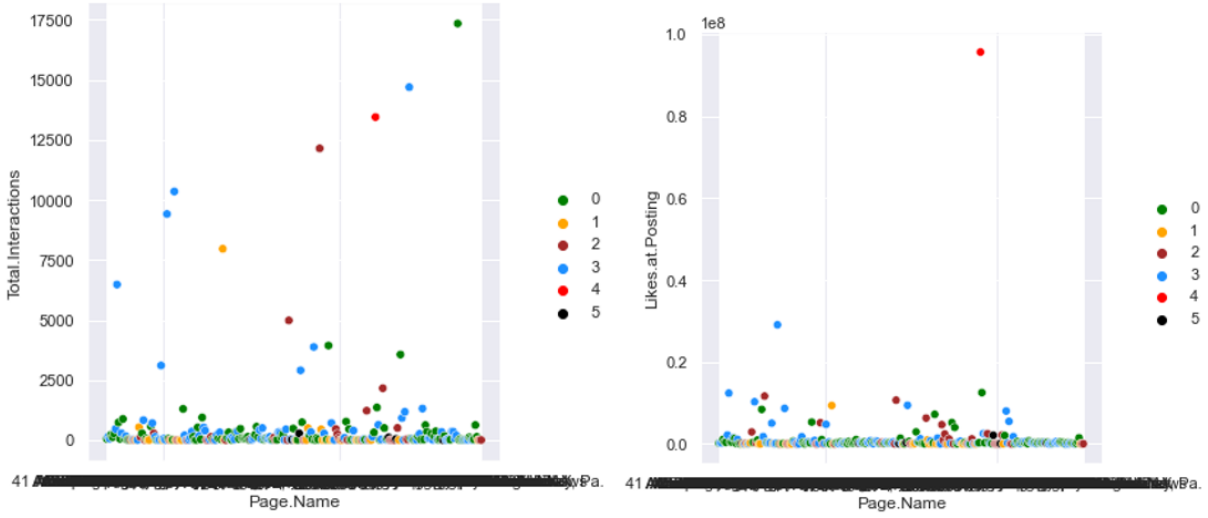

Exhibit 4 - Brands Clusters with Total Interactions and Likes at Posting features

Tech Insider is a section of Business Insider established in 2015, which contents around technology are published on the same website[4] yet different Facebook pages (@techinsider) with Business Insider (@businessinsider). Even though the two pages appear in the same cluster and top the likes and followers at post, that of Tech Insider is 1.5 times higher and both have

[4] https://www.businessinsider.com/sai

percentage interactions of likes and followers below 0.01%, 10 times lower than the 0.10% cluster average. However, there is room for improvement.

Our team would recommend a three-phase strategy for Tech Insider to boost engagement - (i) video crossposting quick-wins, (ii) content creation dos-and-don'ts and (iii) strategic repositioning with Business Insider. From the extrinsic analysis, we have learnt that cross-posting from a central video library to several pages is the trick to promote interactions. This feature is available to Tech Insider but not currently utilized (used once only as data revealed). Instead of sharing videos directly on Tech Insider page, the social media marketing manager can upload videos to Business Insider then crosspost to Tech Insider.

The content creation dos-and-don'ts and strategic repositioning are based on intrinsic analysis specific to the brand. We computed CrowdTangle's overperforming score for each post of 256 unique brands for a more apple-to-apple comparison of interactions and re-ran the random forest model for overperforming and underperforming posts. Comparing relative feature importance, we can formulate the recipe of Tech Insider's success in formulating their content around gender, health, affiliation, percept, risk and social as part of their editorial guidelines. On the contrary, contents related to money, reward, sad, work, power and home shall be avoided as these are not the audiences' concerns. People might fancy breakthroughs of technology for the common good of mankind and find the association of technology to money and power less appealing. Or simply they are not looking for content for wealth building from a technology-focused page.

In the longer-term, Tech Insider and Business Insider marketing managers should sit together and align the strategic positioning of both sites as well as their social media platforms. Readers have shown their different interests to the two entities from interactions, for example 'male' (gender) ranks as the 3rd top feature in over-performing Tech Insider posts but 19th in Business Insider, and 'home' is the 1st in Business Insider but 19th in Tech. Should the brand fulfill its audience to their taste or shape them into Insider's overall branding and positioning? Perhaps more business considerations shall be taken into account but marketing/text analytics is definitely playing a big role in answering those.

Tech Insider Over-performing Posts			
#	Predictor	Importance	Relative pos diff
1	social	0.292	4
2	Analytic	0.138	2
3	male	0.098	20
4	affiliation	0.097	6
5	drives	0.064	-2
6	percept	0.063	5
7	cogproc	0.057	0
8	work	0.025	-7
9	power	0.019	-7
10	Message_Subjective	0.018	6
11	health	0.016	9
12	achieve	0.015	-4
13	Message _Positive	0.014	6
14	hear	0.014	-2
15	Type_Native Video	0.010	3
16	risk	0.009	5
17	see	0.009	0
18	money	0.008	-12
19	home	0.006	-6
20	leisure	0.006	-5

Tech Insider Under-performing Posts			
#	Predictor	Importance	Relative pos diff
1	work	0.091	7
2	power	0.077	7
3	drives	0.074	2
4	Analytic	0.070	-2
5	social	0.069	-4
6	money	0.065	12
7	cogproc	0.060	0
8	achieve	0.056	4
9	reward	0.054	12
10	affiliation	0.047	-6
11	percept	0.046	-5
12	hear	0.043	2
13	home	0.043	6
14	sad	0.033	8
15	leisure	0.031	5
16	Message_Subjective	0.028	-6
17	see	0.025	0
18	Type_Native Video	0.019	-3
19	Message _Positive	0.015	-6
20	health	0.011	-9

Business Insider Over-performing Posts			
#	Predictor	Importance	Relative pos diff
1	home	0.263	12
2	cogproc	0.122	5
3	social	0.093	2
4	Analytic	0.090	4
5	work	0.085	6
6	drives	0.050	-4
7	power	0.037	3
8	achieve	0.033	1
9	risk	0.025	8
10	percept	0.024	2
11	leisure	0.022	-7
12	Type_Native Video	0.022	4
13	money	0.021	-7
14	hear	0.014	10
15	reward	0.014	-14
16	Message _Positive	0.014	3
17	see	0.013	-3
18	affiliation	0.010	-15
19	male	0.008	4
20	Message_Subjective	0.008	2

Business Insider Under-performing Posts			
#	Predictor	Importance	Relative pos diff
1	reward	0.112	14
2	drives	0.089	4
3	affiliation	0.075	15
4	leisure	0.070	7
5	social	0.064	-2
6	money	0.064	7
7	cogproc	0.061	-5
8	Analytic	0.059	-4
9	achieve	0.058	-1
10	power	0.057	-3
11	work	0.056	-6
12	percept	0.038	-2
13	home	0.025	-12
14	see	0.024	3
15	family	0.024	13
16	Type_Native Video	0.020	-4
17	risk	0.019	-8
18	health	0.018	5
19	Message _Positive	0.012	-3
20	feel	0.011	6

Exhibit 5 - Feature Importance on Over/under-performing Posts