



SGAG'S SOCIAL MEDIA CONTENT ANALYSIS

ANLY482 ANALYTICS PRACTICUM PROPOSAL

ANITA | HOE XIU MING | SALLYANA

GROUP01 - EIPHANY

TABLE OF CONTENTS

Overview	3
Sponsor Background	3
Project Background & Motivation	3
Project Objectives	4
Data Collection and Description	5
SGAG Facebook Page Level Insights	5
SGAG Facebook Post Level Insights	6
SGAG Facebook Video Posts Insights	7
SGAG Twitter Activity Metrics	7
Challenges	8
Scope of Work	8
Proposed Methodology	9
Data Collection	9
Data Preparation	9
Exploratory Data Analysis	9
Data Cleaning	9
Data Normalization and Transformation	10
Cluster Analysis	10
Sentiment Analysis	10
Recommendations & Insights	10
Stakeholders	11
Deliverables	11
Project Timeline	12
References	13

Overview

In today's world, social media is no longer just a mean of connecting to peers. It has become a place where most of us get our entertainment and news update from, a channel to raise awareness and even a great platform for business start-ups to launch and manage their business. Locally, SGAG is the epitome of a successful social media business. In spite of their successful journey thus far, many other players have joined SGAG to generate humour contents on social media in recent years.

Under such competition, maintaining and diversifying their customer base have been a challenge to SGAG. Furthermore, with the ever growing post volume and traffic, evaluating the performance of its social media posts has become yet another concern to SGAG. Therefore, SGAG wishes to discover meaningful insights from the historical data of its various social media channels such as Facebook and Twitter.

Sponsor Background

Founded by Xiao Ming and Karl Mak in 2012, SGAG is one of Singapore's leading local humour content creators with the mission of "making every Singaporean's day a better one". SGAG generates creative contents across multiple content distribution platforms. These platforms include web application, mobile application and several popular social media sites such as Facebook, Twitter, YouTube, Instagram, and Snapchat.

Over the years, SGAG has built a loyal fan base on the various social media platforms. It has acquired a remarkable figure of nearly 495,000 likes on Facebook and 245,000 followers on Twitter. On top of that, it has also captured more than 250,000 followers on Instagram. Henceforth, SGAG looks forward to expanding and diversifying their customer base, as well as achieving greater expansion in the industry.

Project Background & Motivation

To achieve SGAG's motto of "to make readers laugh at least 5 times a day, 365 days a year", the company focus their attention on creating engaging and interesting content in their daily posts. SGAG creates two types of contents on the different social media platforms. These include branded content which is paid advertisement and organic content which features humors related to Singaporean's daily life. These contents are presented in the form of images memes, videos and listicles which can be liked, commented and shared by social media users.

Even though SGAG's journey has been a very successful one thus far, it still faces fierce competition from both local and international competitors due to the low barrier of entering the content creation market. In the local market, "SMRT Feedback" is one of SGAG's biggest competitors in generating humour content through Facebook and Twitter. Apart from "SMRT Feedback", there are also new players such as "TheSmartLocal" joining SGAG in the fray to generate sponsored video content. Besides that, SGAG also faces competition from bigger international players such as BuzzFeed and 9GAG.

Being in such a competitive market has resulted in SGAG's needs to constantly improve their content generation strategy to maintain their competitive advantage. To do so, SGAG would like to evaluate the performance of their contents on Facebook and Twitter. Particularly, SGAG would like to find out the following:

1. What are the different characteristics that constitute a "great" post? Does SGAG's audience still has the same content preferences as identified in last year? If not, what are the preferences of the audience now?
2. For paid content, what are the reasons affecting the performance of the posts? In particular, does the industry of the advertisers influence the performance of their posts?
3. How are the video posts performing on their social media platforms? Do video posts generate lesser views as compared to photos? If so, what are the factors that are affecting the performance of their video posts?

Project Objectives

Utilising data extracted from SGAG's social media channels such as Facebook and Twitter, we would like to discover meaningful and informative insights which will allow SGAG to better formulate their content and media strategy in order to reach out to more audience from different segments.

We will attempt to measure the effectiveness of paid and organic contents which SGAG has generated by examining the fans/followers' growth, audience reach and engagement on the different platforms. To assess the audience engagement, measurements such as number of "likes", "comments", "shares", "retweets", "views" and "link clicks" on the different posts will be used. For special case like video posts, factors such as completeness of the video viewing (percentage of video that has been watched) will also be taken into consideration.

To achieve the above mentioned, we will perform an in-depth analysis on the historical data gathered with the following aims:

1. To examine the popularity of paid versus organic content on Facebook and Twitter
2. To determine the performance and characteristics of different types of Facebook posts
3. To assess the popularity of posts according to the different types of format, namely: memes, videos and listicles
4. To understand the positive and negative perceptions of SGAG user's comments and tweets, and to examine the reason behind such sentiment.
5. To evaluate the performance of the branded paid advertisements by the industry of the advertisers

Data Collection and Description

To facilitate the data analysis, SGAG provided our team with the datasets of their two main social media channels which are Facebook and Twitter. Both of the datasets are obtained through the social media insights of the respective platform, ranging from September 2015 to August 2016.

Some instances of the datasets that are provided to us are SGAG Facebook Page Level Insights, Post Level Insights, Video Posts Insights and SGAG Twitter Activity Metrics. To further understand the datasets given, we will attempt to look at the details of each datasets in the following section:

SGAG Facebook Page Level Insights

Facebook Page Level Insights Data shows us a general view of the overall page statistics and performance which is linked to the growth of SGAG's Facebook Fan Page. Key metrics sheet that can be found in this dataset includes Lifetime Total Likes, Daily Total Reach, Daily Organic Reach, New Likes and Weekly Total Impressions.

The detailed description of the metrics is as follow:

Column Name	Description
Lifetime Total Likes	The total number of unique users who have liked SGAG's Page.

Daily count of fans online	The number of unique users who liked SGAG's Page and who were online on the specified day.
Weekly Page Engaged Users	The number of unique users who engaged with SGAG's Page. Engagement includes any click or story created.
28 Days Organic Reach	The number of unique users who visited SGAG's Page, or saw your Page or one of its posts in their news feed or ticker. These unique users include fans who have liked SGAG's Page as well as those who haven't.

SGAG Facebook Post Level Insights

Facebook Page Level Data only allows us to view the top-content analytics which is inadequate in analysing the performance of SGAG's posts. Hence, SGAG also provided us with their Facebook Post Level Data which gives us a deeper understanding by breaking the metrics down into post level.

Following are some of the Metrics on the Post Level:

Column Name	Description
Lifetime Post Organic Reach	The number of unique users who saw SGAG post in their news feed or ticker, or on SGAG page's timeline.
Lifetime Post Total Impressions	The total count in number of impressions of SGAG's post.
Lifetime Post reach by people who like your Page	The total number of users who liked SGAG's page and saw SGAG's post in their news feed. In other words, this metric refers to SGAG Fan Reach for both organic and promoted post reach.
Lifetime Post Consumers	The number of unique users who clicks anywhere on SGAG's post, regardless of whether it results in a story. The clicks can come from comments, video play, expand comments, click profile of commentator.

SGAG Facebook Video Posts Insights

SGAG does not only create picture or text content, they also produce video content on their social media platforms. The performance of its video posts is also a crucial information that SGAG would like to find out. Hence, insights on video posts will be able to help us in analysing unique metrics pertaining to their video posts.

Following are some of the critical Video Metrics:

Column Name	Description
Lifetime Unique Video Views	The number of unique users who viewed SGAG's video for more than 3 seconds.
Lifetime Total 30-Second Views	The total number of times SGAG's video was viewed for 30 seconds or viewed to the end, whichever came first.
Lifetime Unique Views to 95%	The number of unique users who viewed SGAG's video to 95% of the video length.
Lifetime Auto-Played Video Views	The total number of times SGAG's video started playing automatically and users viewed it for more than 3 seconds.
Lifetime Clicked-to-Play Video Views	The total number of times users clicked to play SGAG's video and viewed it more than 3 seconds.

SGAG Twitter Activity Metrics

The various activity metrics of Twitter enable us to gain a better understanding of SGAG tweets performance and show how SGAG's tweets resonate with their audience.

The description of Twitter Metrics is as follows:

Column Name	Description
Impressions	The total number of times a user is served a Tweet in their timeline or search results.
Engagement	The total number of times a user interacted with a Tweet by clicking anywhere on the Tweet, including retweets, replies, follows, likes, links, cards, hashtags, embedded media, username, profile

	photo, or Tweet expansion.
Likes	The total number of times a user liked SGAG's Tweet.
User profile clicks	The total number of clicks on the SGAG's profile, @handle, or profile photo SGAG.
Follows	The total number of times a user follows SGAG's Twitter account directly from the Tweet.

Challenges

A potential difficulty that we might be facing while dealing with the data is the different formats of the datasets given. Facebook post's and page's data comes in Microsoft Excel spreadsheets, while other information such as the advertisers' data is stored in PDF format. As a result, manual reconciliation of advertiser's data to the post's data (advertisement) might need to be performed.

Scope of Work

With all the aforementioned objectives in mind, we will aim to perform analysis using data gathered from the two distributions channels: Facebook and Twitter, ranging from September 2015 to August 2016.

Our proposed scope of the project includes the following:

1. Data Collection
2. Data Preparation
3. Exploratory Data Analysis
4. Data Cleaning
5. Data Normalization and Transformation
6. Cluster Analysis
7. Sentiment Analysis
8. Recommendations and Insights

Proposed Methodology

Data Collection

Apart from gathering data from SGAG's Facebook and Twitter, we will attempt to obtain additional data such as Facebook's comments and Twitter's tweets to find out the different sentiment that the user might have upon writing the comments and tweets. In addition, we would also like to look at other data such as the dates of public holiday in Singapore, in order to examine the effects of such occasions on the traffic of SGAG's social media platforms.

Data Preparation

As the datasets we receive are in different formats and files, we will attempt to consolidate the data into suitable and consistent format to be used for analysis.

Different spreadsheets that are broken down by the date range will be combined into one CSV file.

Exploratory Data Analysis

In exploratory data analysis, we will firstly identify missing values and potential outliers in the datasets. After which, we will look at the trends across all the posts in terms of average number of likes, comments, shares, video views and average duration of video views.

Data Cleaning

Missing values and outliers identified in the data exploration may cause inaccuracy and skewness in the data. Hence, we will perform the necessary techniques in our attempt of cleaning the data. To handle the missing value, we will look at the amount of missing values identified to determine whether we should eliminate the data objects or to fill the value in automatically. As for outliers, we will look at the reason of the existence of such outliers to decide whether we should ignore the data objects or to include it as a different group in our analysis.

Data Normalization and Transformation

As the distribution of the certain attributes may be skewed, we will attempt to normalize such attributes prior to performing our analysis, such that these attributes will not dominate other attributes during our analysis. Furthermore, data transformation techniques such as discretization or binarization will also be performed should our analysis or model building requires the data to be in categorical or binary form.

Cluster Analysis

Cluster analysis will be carried out to determine the existence of clusters in SGAG's Facebook posts. Thereafter, we will attempt to identify the profiles of each cluster of posts and examine the reasons affecting the performance of each cluster. This will help SGAG in reviewing the performance of the different type of posts and to better improve the quality of the posts.

Sentiment Analysis

In order for SGAG to better understand its' users, text mining and sentiment analysis will be conducted on SGAG Facebook's comments or Twitter tweets. We will attempt to discover the various reasons causing happiness and unhappiness of comments and tweets related to SGAG's posts. By doing so, SGAG will also be able to realize the effect of its users' sentiments on its posts' performance.

Recommendations & Insights

Based on the insights gathered from the different analysis, we will provide recommendations for SGAG to improve on its content generation strategy. If there is any outstanding or possible future works, we will also hand over for SGAG to look into and work on.

Stakeholders

The main stakeholders of this project includes:

- Project Supervisor: Prof Kam Tin Seong, Associate Professor of Information Systems; Senior Advisor, SIS (Programme in Analytics)
- Sponsor: Karl Mak, Co-Founder of SGAG & MGAG

Deliverables

The final deliverables of this project are as following:

- Project Proposal
- Mid-term Report
- Mid-term Presentation
- Final Report
- Final Presentation
- Project Wiki Page
- Project Poster
- Recommendations to Sponsor

Project Timeline

Tasks		Week 1	Week 2	Week 3	Week 4	Week 5	Week 6	Week 7	Week 8	Week 9	Week 10	Week 11	Week 12	Week 13
Requirement Gathering	Confirm Project Sponsor	Planned	Actual											
	Gather Requirements	Planned												
Data Gathering	Industry and Competitors		Planned	Actual										
	Gather Data from Client		Planned	Actual										
Project Proposal	Refining Project Scope		Planned											
	Proposal Preparation			Planned	Actual									
	Wiki Page Preparation				Planned	Actual								
Proposal Deadline 4 September, 2359														
Data Exploratory	Data Preparation					Planned								
	Data Exploration				Planned	Actual								
	Data Cleaning					Planned								
	Data Normalization & Transformation						Planned							
	Cluster Analysis							Planned						
Midterm	Midterm Report Preparation						Planned							
	Midterm Wiki Update							Planned						
	Midterm Presentation Preparation								Planned					
Midterm Presentation Week 7														
Insights & Recommendation	Sentiment Analysis							Planned						
	Create Visualization from Analysis Results								Planned					
	Generating Insights									Planned				
	Formulate Recommendations										Planned			
Final	Final Report Preparation										Planned			
	Final Wiki Update											Planned		
	Final Presentation Prepatation												Planned	
	Final Submission													Planned
	Final Presentation week 13													

Agenda	
Planned	
Actual	
Milestone	

References

TechInAsia. (2016, April 11). We're Karl Mak & Xiao Ming, co-founders of SGAG. Ask us anything! Retrieved from <https://www.techinasia.com/talk/sgag-ama>

Digital Insights. (2013, December 30). How to measure your Facebook Page with Facebook Insights. Retrieved from <http://blog.digitalinsights.in/the-ultimate-guide-to-facebook-insights/05118987.html>

Andrea Vahl. (2014, October 16). How to Do a Deep Dive into Your Facebook Insights in 6 Steps. Retrieved from <http://www.andreavahl.com/facebook-insights/deep-dive-facebook-insights.php>

VerticalResponse. (2013, July 18). A Guide to Facebook Insights. Retrieved from <http://www.verticalresponse.com/blog/a-guide-to-facebook-insights/>

Facebook. (n.d.). Understand How Your Videos Perform on Facebook. Retrieved from <https://www.facebook.com/facebookmedia/best-practices/video-metrics>

Twitter. (n.d.). Twitter activity dashboard. Retrieved from <https://support.twitter.com/articles/20171990>