Social Media Analytics for Business (CIS 7029)

Assignment

* 1. As of an organization I’d like to select [Prime Group](https://www.primelands.lk/) real state giant,

issues they are facing are, Because of they are not handling properly, we can see limited number of people engage with their SM channels.

“*The value that SM creates for users is limited, and this lack of knowledge could limit the value accessible through SM engagement.*”(Cartwright, Davies and Archer-Brown, 2021, p. 130)

Due to SM channels context and trends changing time to time, we can see they are not accepting it, and adopting this behaviour. “*Brand image can be reinforced over time in various SM contexts, targeting B2B purchase decision-makers.*”(Tiwary *et al.*, 2021, p. 123)

By not identifying the correct age groups, SM context etc. it causes for incorrect measurements when it is branding through SM. because of they have limited resources and subject expertise. “*corporate branding through SM, firms can face barriers and measurement issues*,”(Tiwary *et al.*, 2021, p. 133)

challengers they are facing are, No matter big or small industry, it must build SM correctly “*Despite larger organisations possessing a greater amount of resources, they faced similar challenges to those experienced by smaller organisations when implementing SM.*”(Cartwright, Davies and Archer-Brown, 2021, p. 124)

Setting up relevant SM Channel and Content highly challenging its need reach correct audience “*Organisations often face challenges in settling upon an SM channel that is most appropriate for them and in developing the correct content that will reach the desired audience*”(Cartwright, Davies and Archer-Brown, 2021, p. 127)

In briefly other important points, we can realize from this.

“*No specific channel strategy, Low quality of content that is not engaging, Lack of dialogue amongst existing relationship partners, Difficulty in choosing the correct platform for very specific audiences, difficultly to continuously generate thought leadership content Engaging broad audiences increases risk in negative comments, long period of time to develop beneficial content, Relationships usually very complex that require consistent nurturing*” (Cartwright, Davies and Archer-Brown, 2021, p. 126)

Moreover, when try to implement SM as strategic tool, co-creators facing hesitation of managing it (Cartwright, Davies and Archer-Brown, 2021, p. 129)

Ethical consideration wise “*being honest, and disclosing relationships such as who you work for in personal posts when endorsing products and services, and not writing fake reviews.”(Joseph W. Barnes, 2015, p. 13)*

In order to marketing via SM it’s not a good approach to use free giveaways, (Muzumdar, Grant-Kels and Farshchian, 2021, p. 1) specially in real state industry its totally depend on trust.

1.2.

With the latest development with metaverse (Mystakidis, 2022), I can launch campaigns to design and develop to consumers to buy virtual land plots, virtual visits to fully completed condominium apartments and showcase the upcoming land and property development projects. “*The Metaverse is the post-reality universe, a perpetual and persistent multiuser environment merging physical reality with digital virtuality. It is based on the convergence of technologies that enable multisensory interactions with virtual environments, digital objects and people such as virtual reality (VR) and augmented reality (AR).*”(Mystakidis, 2022, p. 1)

With the emerge of CQA “*Community question answering (CQA) sites have emerged as platforms designed specifically for the exchange of questions and answers among communities of users.*” (Camacho, Luzón and Cambria, 2021, p. 319) its best to introduce platform specified bot assistance, which is not existing currently.

Letting consumers to decide what they want it another way of making it more engaging. “*One of the important ways in which social media differs from traditional media is that content is user-generated.”* (Vandenbosch, Fardouly and Tiggemann, 2022, p. 2) so, for that facilitating to consumers to design their own land plots, home decorations and interior designs using simple tools such as [AutoDraw](https://www.autodraw.com/)

Its oblivious more and more new SM platforms to popular in future (Vandenbosch, Fardouly and Tiggemann, 2022, p. 1), for an example like TikTok, so It will worthy to sharpening the SM team to work with new trends and new platforms to make this brand popular among youth.

With the introduction of SM platform’s integrated wallets approaches, like [facebook pay](https://pay.facebook.com/), [WhatsApp Pay](https://www.whatsapp.com/payments/in) and [TikTok shop](https://seller-sg.tiktok.com/) I’d like to introduce payment gateways to current brand from these platforms.

Facebook recently introduced Top Fan, Valued Commenter badges for Fan of Facebook pages. Such badges are already introduced for Facebook group fans, i.e Visual storyteller, Conversion starter. Likewise, I’ll introduce loyalty program and points-based system to make consumers more attach to the brand. When considering points, it will depend on their SM platform engagements too.

With the rise of IOT (Internet of Things) and WOT (Web of things) it’s ideal convince higher management go for introduce smart homes and housing schemes, then based on these we can implement a SM-IOT collaborative systems, that data ans system can use for disaster managements situations like fire, flood, earthquakes, and other community bonding situation. “*With the rapid evolution of web technologies, Web 3.0 aims to expand on current and emerging social media platforms such as Facebook, Twitter, and TikTok, and integrate emerging computing paradigms, including the Internet of Things (IoT), named social media 3.0. The combinations of these platforms in Web 3.0 promises consumers greater integration, interaction, and more seamless movement between physical spaces.*” (Salim, Turnbull and Moustafa, 2022)

Diagram

Description automatically generated

Fig. 1. The functional architecture of the SM-IoT platform (Dridi, Sassi and Faiz, 2017, p. 1423)

2.1.1.

Real Estate industry is another essential area need to improve along with the population growth. As for living standards getting improve, this area needs to improve parallelly. I’d like to select **Prime Group** as for the company. When considering property market Prime Group is one of major player. But with the present market status, it’s been a challenge for this property developer to maintain their products such as lands, houses, apartments, housing schemes and commercial buildings. There are few large and small competitors for Prime Group presently, such as, **Home Lands**, **Blue Ocean**, **CBH Lands**, **Sky Line** and **HouseAndLandsLk** companies.

Based on following Real State Companies in Sri Lanka against Revenue and Employees table, it is showing other major players and their revenues in this business.

|  |  |  |  |
| --- | --- | --- | --- |
| **Rank** | **Company** | **Revenue(USD, Million)** | **Employee Count** |
| 1 | Trillium Property Management Services | 65.8 | 243 |
| 2 | Fairway Holdings Pvt | 62.3 | 334 |
| 3 | John Keells Properties | 59 | 892 |
| 4 | Upali Newspapers | 37.9 | 151 |
| 5 | Al Aman Group | 37.7 | 142 |
| 6 | Xavier Holdings Private | 36.7 | 335 |
| 7 | Prime Lands | 28.2 | 96 |
| 8 | Rush Lanka Group | 25.1 | 95 |
| 9 | JB Lands | 20.3 | 109 |
| 10 | The Home Lands Holdings | 18.5 | 62 |
| 11 | Overseas Realty (Ceylon | 18.2 | 271 |
| 12 | Coral Property Developers | 18.2 | 61 |
| 13 | Malwatte Valley Plantations | 17.2 | 14 |
| 14 | Port City Colombo | 16.9 | 49 |
| 15 | Altair | 16.6 | 47 |
| 16 | Nivasie Developers | 16.3 | 57 |
| 17 | Araliya Lands & Homes - Pvt | 16 | 51 |
| 18 | Dusit Thani | 14.4 | 54 |
| 19 | Gdc | 11.6 | 65 |
| 20 | Skyline Pvt | 11.5 | 46 |
|  |  |  |  |
|  | **Total** | **548.4** | **3174** |
|  | **Average/Mean** | **27.42** | **158.7** |
|  | **Median** | **18.35** | **80** |
|  | **Min** | **11.5** | **14** |
|  | **Max** | **65.8** | **892** |

(zoominfo, 2022)

We can see Prime Lands is on 7th position, making 28.2 million USD in the year of 2021. when this amount converts into Sri Lankan Rupees its around 980-1000 million, that amount tally in their following **Management Discussion & Analysis report for 2020/2021. (Prime Lands (Pvt) Ltd, 2021)**

Table

Description automatically generated

(Prime Group (Pvt) Ltd, 2021)

After calculating Total, Median and Average/Mean of above table we can see Prime Lands revenue is $27.25M is above the average revenue which is $27.25M, also it’s above the median revenue which is $18.35M

Also, they have around 96 direct employees comparing to Homelands and Skyline its higher amount, it’s above the median and average amount of employees which are respectively 78.5 and 36.99.

**Evolution of Condominium Market in Sri Lanka: A Review and Predict (Prathapasinghe, 2018)** and **Real Estate Market Analysis - First Quarter 2022(Statistics Department of the Central Bank of Sri Lanka, 2022)** showing its increasing **condominium product** market in future decade, so all these companies have this common product as one of their business scopes. So, they should prioritize that product, by playing their part properly and maintaining good ethical business practices can keep their space in this field.

|  |  |
| --- | --- |
| Chart  Description automatically generated with medium confidence  (Prathapasinghe, 2018) | Chart, line chart  Description automatically generated  (Statistics Department of the Central Bank of Sri Lanka, 2022) |

2.2.1.

As for the SMART goal, promoting condominiums among younger generation, age group of 25-35 will be more beneficial for the company by analysing below data sets. We can estimate that it can reach, by selling 10000 of apartments among these millennials period of Aug 2022 to Aug 2023.

Below is consideration of latest 10 facebook posts from each, Land Sale related posts and Condominium Apartment related posts from Prime Group.

Land Sale Related Reaction Measurements

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Date** | **Type** | **Comments** | **Shares** | **Likes** | **Angry** | **Love** | **Care** | **Haha** | **Wow** | **Sad** |
| 11/08/2022 | Land Sale | 2 | 4 | 12 | 0 | 0 | 0 | 0 | 0 | 0 |
| 09/08/2022 | Land Sale | 2 | 0 | 404 | 0 | 5 | 1 | 0 | 0 | 0 |
| 09/08/2022 | Land Sale | 1 | 2 | 335 | 0 | 3 | 1 | 1 | 0 | 1 |
| 09/08/2022 | Land Sale | 3 | 1 | 792 | 0 | 3 | 0 | 0 | 1 | 0 |
| 09/08/2022 | Land Sale | 4 | 2 | 533 | 0 | 10 | 2 | 2 | 0 | 0 |
| 09/08/2022 | Land Sale | 9 | 3 | 281 | 0 | 4 | 0 | 0 | 0 | 0 |
| 09/08/2022 | Land Sale | 0 | 11 | 143 | 0 | 0 | 0 | 0 | 0 | 0 |
| 09/08/2022 | Land Sale | 2 | 4 | 31 | 0 | 0 | 0 | 1 | 0 | 0 |
| 02/08/2022 | Land Sale | 10 | 7 | 1300 | 0 | 14 | 2 | 2 | 1 | 0 |
| 02/08/2022 | Land Sale | 23 | 3 | 1200 | 1 | 23 | 3 | 1 | 5 | 2 |
| **Total** |  | **56** | **37** | **5031** | **1** | **62** | **9** | **7** | **7** | **3** |
| **Median** |  | **2.5** | **3** | **369.5** | **0** | **3.5** | **0.5** | **0.5** | **0** | **0** |
| **Average** |  | **5.6** | **3.7** | **503.1** | **0.1** | **6.2** | **0.9** | **0.7** | **0.7** | **0.3** |
| **Min** |  | **0** | **0** | **12** | **0** | **0** | **0** | **0** | **0** | **0** |
| **Max** |  | **23** | **11** | **1300** | **1** | **23** | **3** | **2** | **5** | **2** |

Condominium Apartments Related Reaction Measurements

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Date** | **Type** | **Comments** | **Shares** | **Likes** | **Angry** | **Love** | **Care** | **Haha** | **Wow** | **Sad** |
| 20/04/2022 | Condo | 81 | 8 | 390 | 1 | 27 | 3 | 1 | 0 | 0 |
| 11/03/2021 | Condo | 77 | 26 | 193 | 0 | 8 | 3 | 0 | 0 | 1 |
| 09/03/2021 | Condo | 1100 | 262 | 12 | 3 | 312 | 23 | 19 | 22 | 19 |
| 16/02/2021 | Condo | 243 | 80 | 767 | 0 | 29 | 8 | 2 | 3 | 1 |
| 03/02/2021 | Condo | 58 | 13 | 808 | 0 | 22 | 0 | 0 | 0 | 0 |
| 03/02/2021 | Condo | 158 | 79 | 5200 | 0 | 85 | 3 | 6 | 5 | 2 |
| 06/07/2020 | Condo | 17 | 26 | 3300 | 1 | 26 | 2 | 4 | 2 | 1 |
| 24/06/2020 | Condo | 94 | 93 | 19000 | 2 | 158 | 19 | 32 | 18 | 3 |
| 19/06/2020 | Condo | 0 | 2 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 12/06/2020 | Condo | 38 | 34 | 1200 | 0 | 12 | 2 | 1 | 3 | 0 |
| **Total** |  | **1866** | **623** | **30870** | **7** | **679** | **63** | **65** | **53** | **27** |
| **Median** |  | **79** | **30** | **787.5** | **0** | **26.5** | **3** | **1.5** | **2.5** | **1** |
| **Average** |  | **186.6** | **62.3** | **3087** | **0.7** | **67.9** | **6.3** | **6.5** | **5.3** | **2.7** |
| **Min** |  | **0** | **2** | **0** | **0** | **0** | **0** | **0** | **0** | **0** |
| **Max** |  | **1100** | **262** | **19000** | **3** | **312** | **23** | **32** | **22** | **19** |

Total number of comments of condominium apartment related post counts of 1866, is 3332.14% higher than Lands related posts 56 of total comments. For comments mean/average this personage (3332.14%) is same since its same numbers. Also, we can see Maximum number of comments went to 1100 for condominium apartments, but 23 is the most number for the other category. Median is 3160% higher.

Likewise, by comparing these 2 data sets against Total, Median, Average, Min, and Max with criteria of Shares, Likes, Angry, Love, Care, Haha, Wow and Sad we can notice significant difference of user interactions.

With these rates its likely get 150-200 total comments, 50-80 of shares, 2000-3500 of likes, 0-1 of Angry reactions, 50-100 of Love reactions, 5-10 of Care reactions, 5-10 of Haha reactions, 5-10 of Wow reactions and 1-5 of Sad reactions for the next condominium related post.

2.2.2.

Since it has 198,058 of followers in Prime Lands Group facebook followers

For the condominium related post, we can find out following engagement rates as following

***“ER post = Total engagements on a post / Total followers \*100”***

**(**Katie Sehl and Shannon Tien, 2022)

|  |  |  |  |
| --- | --- | --- | --- |
| **Date** | **Type** | **Total Interactions** | **ER post** |
| 20/4/22 | Condo | 511 | 0.25800523% |
| 11/3/21 | Condo | 308 | 0.15551% |
| 9/3/21 | Condo | 1772 | 0.89468741% |
| 16/2/21 | Condo | 1133 | 0.57205465% |
| 3/2/21 | Condo | 901 | 0.45491725% |
| 3/2/21 | Condo | 5538 | 2.79615062% |
| 6/7/20 | Condo | 3511 | 1.77271304% |
| 24/6/20 | Condo | 19261 | 9.72492906% |
| 19/6/20 | Condo | 14 | 0.00706864% |
| 12/6/20 | Condo | 1957 | 0.9880944% |
| **Total ER** |  |  | **17.6241303%** |

*“****Average ER by post = Total ER by post / Total posts****”* (Katie Sehl and Shannon Tien, 2022)

Average ER = 17.62/10 is 1.7%

As we can see in above table engagements are consist around 0.5% to each post

After scraping overall facebook post of Prime Lands, date period of 01st of January 2021 to 01st of August 2022 (13 months), results are like these

Monthly reactions

|  |  |
| --- | --- |
| Chart, bar chart, histogram  Description automatically generated | A picture containing chart  Description automatically generated |

Most user feedbacks are positive and likes to engage (Reactions = like + love + haha + wow + sad + angry)

|  |  |
| --- | --- |
| Chart, histogram  Description automatically generated | Chart, histogram  Description automatically generated |

Share counts also steady when consider overall facebook posts. Around 2000 per month for total shares. reactions are overall above 8000 per month.

After analysing all twitter posts of Prime Lands, its comments outcome as follows

|  |  |
| --- | --- |
| Chart, shape, circle  Description automatically generated | Chart, pie chart  Description automatically generated |

With these healthy stats, it’s a good move for social media is a worthy investment for Prime Lands.

2.2.3.

I’m selecting as Facebook post and TikTok as communication platforms.

Facebook post with 1 image and English content would be ideal to share the information between younger generations. By looking at above overall data its user interaction getting higher for riddle related post, there for its will he arranges riddle-based post to boost communities. Since its sharing with English content, it will be easier to do all sentimental analysis and analyse data.

With below 3 surveys, its obvious, millennials are the ones going to become next work force which has more buying power and main users in social media in Sri Lanka.

*“Majority of users (77.2%) are belong to age 18-27 which means that young generation in Sri Lanka is much engaged in Facebook.”(Rathnayake and Rathnayake, 2017, p. 340)*

|  |  |
| --- | --- |
| Chart, bar chart  Description automatically generated(Seven Media Group, 2022) | Table  Description automatically generated(Seven Media Group, 2022) |

Timeline

Description automatically generated with low confidence

(Simon Kemp, 2022)

As we see here TikTok is considered as most engaging social media in now and upcoming years. There for it will worthy to share short clips with following topics about condominium, contents like its helping to minimise carbon footprint, easy to waste management, compact environment, secure neighbourhood and read made product.

Then when we talk about this **condominium purchasing** information flows between this group we can focus information diffusion of networks, so let’s divide 25-35 age group into following groups.

Group A - people who like buy land plot and build their dream home as they want  
Group B - people who like buy ready made apartments

If Prime lands give nice condominium purchasing offer to a Group B person, and then he’ll adopt this offer that information will share among close peers, due to this age group are mostly office colleagues, schoolmates, classmates, and university mates.

office

Group B Guy

school

class

office

office

office

office

office

office

office

office

office

office

office

office

office

office

university

office

office

office

office

office

There are about 10-12 million population of 25-35 age group in Sri Lanka, if few Group B people adopt this condominium purchasing behaviour, it will create network phenonium. If its model this complete Group B as network like below every person in this Group B, we can represent as Node, also we can put edge between two nodes if those nodes represented people, are talking to each other, following each other, listen to each other. so this product start somewhere in this network Group B. Group B guys classmates adopting it, then their friends adopting it, likewise in other sub sets. So this condominium purchase movement will diffuse through this social network.

According to this there can be few possibilities with diffusion

1. Entire population in this Group B adopting this
2. It can die away soon, Nobody adopting this
3. Mixture of decision, Very few people adopting

So when above adoption behaviour applying to model, we can analyse to get adopt above behaviour to someone call **X** who is 30 years old, what requirements to fulfil.

Let’s apply the model for this now

**Group A** represent as **Action A**

**Group B** represent as **Action B**

To do action A, need **10** Millions amount of LKR

To do action B, need **20** Millions amount of LKR

Let’s say, **p fraction** of friends of mine adopted A and remaining **1-p fraction** of adopted B

To get adopt A to me it will be **px10**

To get adopt B to me it will be **(1-p)x20**

so if that **X** to buy condominium apartment(Action B), it should full this

**(1-p)x20 >= px10**

Example :If **X** has 100 friends (classmates, colleagues, schoolmates etc), p is 40 of them adopted Action A

Then with above 60x20 >= 40x10 is true, **X** to adopt Action B

2.2.4.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Platform | Audience | Influencers | Content Type | Response from |
| Facebook | 25-35  Age group | Actors and Actresses, Sportsmen, Comedians | Images with English Content and Hashtags | 25-35 Age group, general public |
| TikTok | 20-30  Age group | TikTok influencers, environmentalist | Clips about Environmental Effects,  Short clips with higher Living Standards | Millennials |

Prime Lands has, 198091 of Facebook followers and 10400 of TikTok followers

|  |  |
| --- | --- |
| Facebook Graph for Prime Lands | A picture containing sky, accessory  Description automatically generated  TikTok Graph for Prime Lands |

Degree of centrality as follows

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Platform | nx.degree\_centrality(g) | nx.out\_degree\_centrality(g) | nx.in\_degree\_centrality(g) | nx.closeness\_centrality(g) | nx.betweenness\_centrality(g) | nx.eigenvector\_centrality(g) |
| Facebook | Prime\_Lands Node : 1.0  For Other Nodes 0.00000505 | Prime\_Lands Node : 1.0  For Other Nodes 0.00000505 | Prime\_Lands Node : 1.0  For Other Nodes 0.0 | Prime\_Lands Node : 1.0  For Other Nodes 0.0 | Prime\_Lands Node : 1.0  For Other Nodes 0.0 | 0.00 |
| TikTok | Prime\_Lands Node : 1.0  For Other Nodes 0.00009615 | Prime\_Lands Node : 1.0  For Other Nodes 0.00009615 | Prime\_Lands Node : 1.0  For Other Nodes 0.0 | Prime\_Lands Node : 1.0  For Other Nodes 0.0 | Prime\_Lands Node : 1.0  For Other Nodes 0.0 | 0.01 |

Clustering coefficient is 0.1-0.9 for both Facebook and TikTok due to there are connection between of their followers, but those followers not connected with each other’s.

2.3.1.

As for a Native tool for Facebook going to use is Facebook Insights as for Non-native tool Hootsuite Analytics($49-$739 per month)

As for a Native tool for TikTok going to use is creatorportal and as for Non-native tool Iconosquare($49-$79 per month)

2.3.2.

For the facebook It will consider following data

Page Views based on Gender, Page Views Based on age groups, Interaction based on Gender, Interaction based on Age groups, Demographic Data, Clicks, Comments, Reach, Shares and Trends

For TikTok It will consider following data

Number of views, likes count, comments count, shares count, view time average, full video watch time, traffic source, type and audience territories. Moreover performance, engagement, standalone video detailed analytics, additive videos, ideal time to post, history & density and Media lifespan.

2.3.3.

If we consider we maintain 12 of team member for each platform

2 – Content Creators for Facebook, 2 – Content Creators for TikTok, 2 – Graphic, Designer, 2 – Videographers , 2 – Video editors, 2 – Photo Editors

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Platform | Staff Cost(LKR) | Tools Cost(LKR) | Miscellanies | Total Cost(LKR) |
| Facebook | 900000 | 500000 | 200000 | 1600000 |
| TikTok | 900000 | 500000 | 100000 | 1500000 |

Based on Annual Report of Prime Lands its mentioning they are selling 2000-3000 condominium apartments per year, but according to my goal withing next 12 months I plan to bring this number to 10000. Their current profit is 1000M in LKR

But after execute this it will be 3500M LKR since (1000/2000)(7000).

Apart from this ROI, they will able to make experienced SM Team and in SM channels user base increase too. After 12 months it can find out who are the people bought theses 10000 apartments by giving them survey, like how they decide to purchase this product, any of their friends has buy this before likewise, we can analyse this sets of data and come to conclusion.

2.4.1.

Facebook data 01st of January 2021 to 01st of August 2022 extracted using *facepager* with special settings, since *facebook\_scraper* not giving actual datasets. Twitter data extracted via *tweepy* and twitter API. Details description about how extracted data I have mentioned in data\_extraction\_guide.docx

2.4.2.1. **Hosted URL -** [**https://service-qpjothvhra-as.a.run.app/app**](https://service-qpjothvhra-as.a.run.app/app)

Graphical user interface, application

Description automatically generated

2.4.2.2.

Chart

Description automatically generated with medium confidence

2.4.2.3.

Chart

Description automatically generated with medium confidence

2.4.2.4.

Graphical user interface, application

Description automatically generated

Graphical user interface, chart, application, pie chart

Description automatically generated

2.5.1.

*Collaborative filtering for social media influencer analysis*

Suppose we have general user call **U**, whom we want make product recommendation, also there are social media influencers call **I1**, **I2**, **I3**, **I4**, **I5**. Mapping U with specific influencer(s) we can focus on metrics such as Gender, Age, Languages, Interests, Locations of influencers.

There may some product that both influencers and **U** likes, some products both influencers and **U** don’t like. We can call it that influencer group as neighbours of user **U** By analysing influencers choices we can to find group of influencers that who are like and dislike similar to **U**.

Once we found set **N** of influencers or the neighbourhood of users similar to users **U**

We can find other products that are likes by set **N**, and recommend those items to the user **U**.

Key trick is to find the users that are similar to user **U**, to do that we need define notion of similarity between influencers.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Products  Influencers | P1 | P2 | P3 | P4 | P5 | P6 |
| I1 | 8 |  |  | 10 | 2 |  |
| I2 | 10 | 10 | 4 |  |  |  |
| I3 |  |  |  | 2 | 8 |  |
| I4 |  | 4 |  |  |  | 6 |
| I5 | 2 | 6 |  | 4 |  |  |

I1 – 8 out of 10 P1 product, 10 out of 10 P3 and 2 out of 10 of P4

I2 – 10 out of 10 P1 and P2 , 4 out 10 P2 and so on.

There are some products with empty rating by some influencers.

We can consider influencer **I** and **J** with rating vectors **r**I and **r**J, we can find similarity matrix **Sim(I,** **J)**

There are lots of products both I1 and I2 not rated, Both I1 and I2 rated P1 as high, I1 and I3 rated P4, P4 common but there rating are very dissimilar. I1 likes P4 dislike P4 , vice versa for I3, so we can group I1 and I2 are similar while I1 and I3 dissimilar which is

Sim(I1, I2) **>** Sim(I1, I3)**.** Users with similar taste have higher similarity than users with dissimilarities.

To apply this, we can use **Jaccard** Similarity (Rana and Deeba, 2019, p. 3)

Sim(I1, I2) = | r I1 r I2 | / | rI1 rI2 |

Jaccard is intersection of rI1, rI2 and divided by union of rI1, rI2,  it’s just take intersection of vectors and divided by union, in the other hand its 1-(Jaccard distance of rI1, rI2)

So for above example Sim(I1, I2) = 1/5 because both together rated P1 , others 4 separately.

And Sim(I1, I3) = 2/4, Sim(I1, I2) < Sim(I1, I3), its measures the quantity of items rated not the qualitative values of it, due to its ignoring rating value we have to abandon this methodology.

Other option we can consider is **Cosine** Similarity (Najafabadi, Mohamed and Onn, 2019, p. 532)

Sim(I1, I2) = Cos(r I1, r I2) =

In order to calculate this we have to fill all empty rated rows as 0 for I1, I2, I3

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | P1 | P2 | P2 | P3 | P4 | P5 |
| I1 | 8 | 0 | 0 | 10 | 2 | 0 |
| I2 | 10 | 10 | 4 | 0 | 0 | 0 |
| I3 | 0 | 0 | 0 | 2 | 8 | 0 |
| I4 |  | 4 |  |  |  | 6 |
| I5 | 2 | 6 |  | 4 |  |  |

Sim(I1, I2) = 0.41996052556580804 , Sim(I1, I2), α = 65.16790467078617°

Sim(I1, I3) = 0.3368165348543039, Sim(I1, I3), α = 70.31696159072192°

Even though its meet our expectation Sim(I1, I2) > Sim(I1, I3). We can see here it’s not a significant difference here, since we made all non-rated as 0 there is a problem here, its treats empty ratings as negative. 0 is the worst possible rating. So we sort of assumption here, if I1 didn’t give rating to P2 , we imposed rating to 0. Which is a bad assumption. Even though it may get some rating for P2 .

To fix above problem there is third option **Centered Cosine/Pearson Correlation** Similarity (Li *et al.*, 2017, p. 204)

To achieve this we have to normalize for a given influencer by subtracting row mean or the average rating of the influencer

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  | P1 | P2 | P3 | P4 | P5 | P6 | Average Rating |
| I1 | 8 |  |  | 10 | 2 |  | 20/3 |
| I2 | 10 | 10 | 4 |  |  |  | 24/3 |
| I3 |  |  |  | 2 | 8 |  | 10/2 |
| I4 |  | 4 |  |  |  | 6 | 10/2 |
| I5 | 2 | 6 |  | 4 |  |  | 12/2 |

We have to subtract each rating from that row average rating, except blank rating.

After subtract it our matrix is like this, we consider blank values as 0

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  | P1 | P2 | P3 | P4 | P5 | P6 | Average Rating |
| I1 | 4/3 |  |  | 10/3 | -14/3 |  | 20/3 |
| I2 | 6/3 | 6/3 | -12/3 |  |  |  | 24/3 |
| I3 |  |  |  | -6/2 | 6/2 |  | 10/2 |
| I4 |  | -2/2 |  |  |  | 2/2 | 10/2 |
| I5 | -8/2 | 0 |  | -4/2 |  |  | 12/2 |

If its sum all the values in each row it become 0 here, What we done here is, we centred influencer’s rating of each influencer around 0, 0 become average rating for every influencer. Positive rating means influencer like that product more than average, negative means vice versa.

Sim(I1, I2) = 0.09245003270420486, Sim(I1, I2), α = 84.69542856090239°

Sim(I1, I3) = -0.9607689228305228, Sim(I1, I3), α = 163.897886248014°

With these values its with bigger margins, showing Sim(I1, I2) > Sim(I1, I3), reason for success of this is missing rating treated as average, also some rates mean to be tough rates and some people not, this handles it well.

With that we can move to predictions or recommendations to general user **U**

Let’s consider **rI** be the vector of influencer **I**’s rating

set **N** of influencers which we call the neighbourhood. who are **K** consist of influencers, most similar to influencers **I**.

Prediction/recommendation for influencer **I** and product **P**

Simple prediction is to take average rating like this from neighbourhood

Average rating of all the users for product **P** in the neighbourhood (Quella, 2021, p. 40)

Option 1 : **rIP = 1/K**

Its ignores actual similarity values who are similarity value **P**, there may be range of similarity values within the neighbourhood, it might contain influencers who are highly similar to influencers **I** and few other influencers who’re not that similar to influencer **I** (Quella, 2021, p. 40)

Weight the average rating by the similarity values, its weighted average,  **= Sim(I**,**J)**

Option 2 : **rIP =**

We look at neighbourhood **N**, and for each influencer **J** neighbourhood **N**,weight **J**’s rating for product **P** , by. The similarity of **I** and **J**,then normalize it by taking sum of the similarities.

2.5.2.

*Community detection in social networks*

Let’s consider small network example like following

Chart, line chart

Description automatically generated

For a plain eyesight we can say there are 3 communities in this network

Diagram

Description automatically generated

Since this easy to see visually, it’s a small example, but as for nodes and edges adding to this networks in thousands, it’s become complex to tell how many communities are there.

So for option we can use **Girvan-Newman** (Girvan and Newman, 2002, p. 7822) Algorithm to make it automatically find these communities.

In order to do, have to calculate matrix of every single Edge of this network, edge is the lines that map these nodes. We have to calculate **edge betweenness** which is one of **network centrality** measures.

Number of Nodes above network = 12

Total Pairs of Nodes above Network = (12 \* 11) / 2 = 64

Edge betweenness explain like this, every pairs of node in the network, how many of those pairs have shortest path between those nodes, that have to pass through given edge.

Diagram

Description automatically generated with medium confidence

To apply above edge betweenness, If we consider (B2,G7) edge from above network, we can see there are 5 nodes for B2 side community and 7 nodes for G7 side community.

EB of (B2, G7) = ( B2 Nodes \* G7 Nodes ) / Total Pairs of Nodes = (5 \* 7) / 64 = 0.54687

Other example,

EB of (L12, I9) = ( L12 Nodes \* I9 Nodes ) / Total Pairs of Nodes = (3 \* 9) / 64 = 0.42187

Using following function we can do same calculation above.

**nx.edge\_betweenness\_centrality(G)**

('A1', 'B2'): 0.3181818181818182

('A1', 'D4'): 0.16666666666666669

('A1', 'E5'): 0.10606060606060606

('B2', 'C3'): 0.19696969696969696

('B2', 'G7'): 0.5303030303030303

('C3', 'E5'): 0.09090909090909091

('F6', 'G7'): 0.09090909090909091

('F6', 'I9'): 0.06060606060606061

('F6', 'H8'): 0.015151515151515152

('G7', 'H8'): 0.09090909090909091

('G7', 'I9'): 0.36363636363636365

('H8', 'I9'): 0.06060606060606061

('I9', 'L12'): 0.4090909090909091

('J10', 'K11'): 0.015151515151515152

('J10', 'L12'): 0.15151515151515152

('K11', 'L12'): 0.15151515151515152

To identify communities we have to notice most important edges of the network, which are have the highest edge betweenness. If an edge has very high edge betweenness, compared others that means its very often on the shortest path, between some cluster of nodes and to rest of the network, means its holding these two communities to the rest of the network.

For the edges that have very low edge betweenness consider lower importance for the shortest path, not contribute hold communities.

Girvan-Newman saying in order to separate communities, identify highest edge betweenness edges, delete them to separate communities from rest of network. Re-evaluate same process up to some context. By doing so we can do the community detection.

Tricky part here is to know how many of deletion need to do, if its doing too many, it will separate all the nodes, which isn’t correct. Too few means it’s not separating communities which are actually need to separate.

Let’s try to simulate this.

Original Network

A picture containing transport, wire

Description automatically generated

After Step 1

A picture containing accessory, necklet

Description automatically generated

After Step 2

A picture containing graphical user interface

Description automatically generated

After Step 3

Graphical user interface, application, Teams

Description automatically generated

Summary

In first section I selected a company discussed discuss the Challenges Issues, and Ethical Consideration in Social Media usage for the specific area. And I’m the one become Lead of SM analytics team how to drive this brand cope with new trends.

Second part of the section discussed about SMART goal and social media strategy. Compared 2 products of the company, analysed it. Implemented ROI and dashboard for those.

Third Section I’ve discussed important concepts in social media analytics.

Abbreviations

SM - social media

Prime Lands - Prime Group PLC

EB - Edge Betweenness