Instagram likes analysis

Likes analysis based on time of uploading the post

1st Vishnu J G
Computer Science and Engineering
PES University
Bangalore, India
jgvishnu2001@gmail.com

2nd Praneeth A L
Computer Science and Engineering
PES University
Bangalore, India
praneeth.arasada@gmail.com

3rd Ajay Kumar K
Computer Science and Engineering
PES University
Bangalore, India
ajayk.sm2001@gmail.com

Abstract—**Instagram likes analysis is finding the influence of date and time of post on the number of likes that a post gets. This analysis gives a idea of the most suitable time to make a post, so that maximum likes can be obtained. This report has the literature survey for papers which address the changes in number of likes dude to various other factors such as age and followers count.**

Index Terms—**Instagram, post, likes, influence, time, date**

I. INTRODUCTION

Conventional methods of promoting a product such as TV advertisements, radios broadcasts, huge billboards and banners, pamphlets and events or fairs are nowadays being dominated by social media platforms. Influence of social media on product promotion and sales has increased tremendously in the past 2 years(2019,2020) and is expected to increase further this year as well.

Even if a company or business sector does not have a social media presence it can take the help of social media influencers to promote their products. These influencers are the key elements in improving the popularity of many budding companies and industries. Many small scale industries and home based businesses' only mode of marketing is social media. As we have seen many home-based businesses emerged during the period of corona lock-downs. These businesses were able to make a good sale of their products from their social media presence.

Even though social media influence can impact sales, it is important that some experience in the sales field is necessary. Being an influencer or a stand-alone home-based businesses, it is important to understand the type of customers they are dealing with and the product they are advertising. One of the key things to remember while advertising a product on a social media platform is the time when the post for advertising the product is made.

The number of likes on a post depends majorly on the number of people who view the post, so finding a right time to post in general means to find the right time when more people tend to come online or use Instagram. This way the post can be viewed by the maximum number of people.

Understanding this based on past data is important for the influencer to get more like. He/she needs to understand when his/her followers tend to view his contents and when people actively interact with them. Understanding many other factors about the followers such as mean age can help them decide when to make a post to get maximum likes.

The option to share the post is one more amazing feature of social media platforms(Tri, Utan and Euis, 2019). Facebook market helps build a online presence to many businesses. This way it is easier to connect both the seller and someone who is who is in need of a particular product, improving the supply chain. Sharing the post in turn gets more people to view the same post and possibly go further in purchasing the product.

Posts may also be categorised based on hashtags, here the more popular posts i.e., the ones with more likes are found at the top. People searing for a product may search based on hashtags and will see the one with the highest like(popularity).

It is clear that increasing the number of likes is the goal of the influencers. And to do this they need their post to be viewed by as many people as possible. This can be done by fixing on a time when most people come online. Not only the time but also the date, day and month also matters for their post to get more likes.

The main objective of this research is to understand the influence of data, time, day, month and number of hashtags on the likes and comments obtained on that post. Based on the results we prescribe a suitable date-time to make a particular post so that maximum people view and like the post which increases the probability of reaching the target customers.

II. LITERATURE SURVEY

Facebook is one of the major social media platform [1]. A post can receive many type of reactions likes, claps, cheers, angers, hearts, etc,. Considering all this as likes, a SUR model(generalized square OLS method)is built. But the

consideration of each day is not taken into account i.e., only weekday or weekend is considered. This is a more generalized approach as in SUR model heteroscedastic property is an assumption. Time is one other important variable the author has ignored stating a more generalized method of only considering year and month along with weekday or weekend. The author also concludes that the time of the post made has no effect on the number of likes the post gets, with a confidence of 64.17% and also stating that posting on a weekday i.e., during the busy hours gets more likes than posting on a weekend [1].

The above research [1] talks about brand building, social media accounts managed by companies to promote their products, but in our research we are not restricting our study to only companies, in fact we wish to address those people who wish to earn money from Instagram through likes on posts.

That people spend more time on Instagram as opposed to other social media applications.Despite such significant advantages, research on Instagram marketing remains scarce.... Content engagement is quantifiable interactions performed by users on contents on social media channels.The quantifications comprise likes, comments, shares , page views, clicks, and other aspects provided by the corresponding platforms . Specifically for Instagram, supplementary metrics include reach, impression, profile visit, website click, and bookmark.Day is another important variable such that no of likes they get depends on which day they post. Data of 2,958 Indonesian food brands posts were analysed using negative binomial regression to predict the number of likes and comments [2].

This study investigates the most effective posting timing on Instagram to enhance content engagements (i.e., likes and comments). This study reveals that the most effective posting days are on Monday, Tuesday, Wednesday, Thursday, and Friday as contents posted on Instagram on these days result in a higher number of likes and comments. In addition, this study also encourages food brands to post their contents during breakfast and lunch times as these times of the day prompt more likes and comments. The findings of this study will help brands and social media managers to include scheduling in their social media marketing strategy on Instagram. [2].

III. PROBLEM STATEMENT

This paper pertains to analysing the likes obtained on a post based on the time and date the post was made. It deals with the simple question of when to make a post to get the maximum number of likes.

The problem statement for this paper would be to analyse and predict the right time to make a post. This is a maximisation problem which can be looked at as a classification problem. Classification because we consider the post with the number of likes being more than the mean value for a particular user-id as a post made at the right time(high likes post) and one with less than or equal to the mean likes for an user-id as posts made at wrong time(low likes post).

We consider hour of the day when the post is made as we assume that people are more active on social media platforms during the later half of the day than the former half (busy hour of the day). This is a hypothesis, and proving this shows the correctness of our assumption.

Day of the week is one of the significant attribute that we will be dealing with in this paper. Tri, Utan and Euis [1] in their paper stated that posts get more likes on weekdays than on weekend, but this is a more generalized solution to what we are dealing with. Friday being considered as weekday gives an incorrect explanation to their model. Considering each day individually will help us get a deeper knowledge about the influence of the day of the week on the likes that a post can get (fig.1).
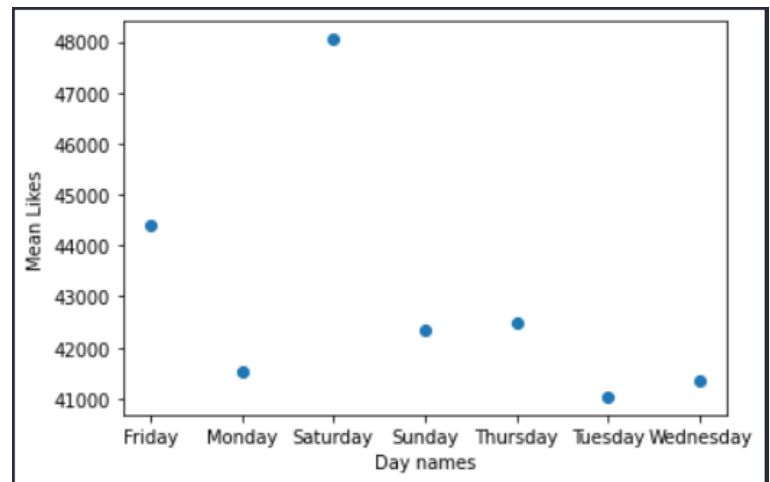

Fig. 1. Day vs mean Likes

Month of the year gives quite a few information about the likes on a post. Months with many festivals and holidays tend to have more likes than dryer months of the year. We can clearly see in fig.1 that more likes are seen at the end and beginning of the year than in the middle of the year (fig.2).

Year is another attribute which is considered. Year doesn't not give much significant information about the likes on the post as our data has post up till mid of 2021 and social media gained popularity only after 2018 end or 2019 beginning. But year helps us easily visualize the influence of other formats on the likes of a post (fig.3).
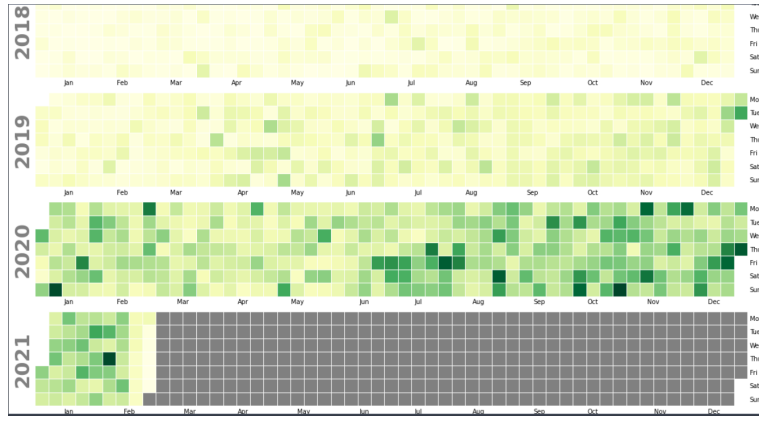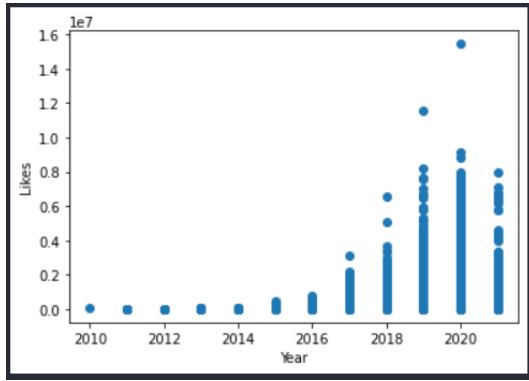
Fig. 2. Calendar graph

Fig. 3. Year vs Likes

IV. How is it different

A. Dataset

The data [3] has 13 columns in all, with 178922 rows or data items (fig.4). The mean and some other basic summary statistics are given in fig.5.
There is a small +ve correlation between Likes and number of comments on a post but this is not of any importance to us.

The dataset consists of Date posted which is in DateTime format. We extract the individual date, month, hour and year. We compute the day of the week based on the extracted values using the calendar module in python. We also have like score which is to generalize likes and followers. The type variable specifies the type of post made i.e., image, video or a collection of images. Minutes is something we will assume to be non-significant in our analysis.

B. Approach

We divide the dataset into train, test and validation as a way of reducing numerosity. We build a simple regression model with likes as the explained variable, considering hour, day, date, month and year as explanatory variables. These are the factors that we mentioned that we are going to study and understand how these factors affect the explained variable

i.e., number of Likes.

$$y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 \tag{1}$$

Here we have $X_1 - hour, X_2 - day, X_3 - date$ $and X_4 - year. \beta_i - are respective coefficients (regression).$

Similarly we wish to build similar other models such as random forest, decision trees and AdaBoost classifier to try and get a model which can predict values as accurate as possible. Accuracy being the measure the grade these models, we can have many more but we will restrict to these as we do not have many ratio values to deal with.

C. Differences

Our approach differs from [1] as we are considering 2 interval values and multiple models, to check accuracy and compare them. The main intention is to get the predicted number of likes to as close as possible to observed value. We compute error using OLS which is not the case in Tri,Utan and Euis [1], who used GLS(Generalized least square)models to estimate errors.

Tri,Utan and Euis [1] use number of tags, manager comments and other factors as well when building the model, which can cause some spurious correlations to the model when predicting likes based soley on time and date. In our research we are using only time and date based explanatory variable to study their effect on the target variable i.e., number of Likes.

Most of the above mentions papers use like score which is a generalization of like for all the users, but we propose to use Likes and restrict it to each users thereby overcoming the problem of normalizing the dataset or depending on some other derived variables such as likes score.

Number of comments is one other variable we can look into, but we are not considering in this research of ours as one user can make only one like on a post but one user can

Fig. 4. Dataset

	User uuid	Likes	Days passed from post	Likes Score	Type	Numer of Tags	Numer of Comments	Date Posted	Year	Month	Day	Hour	Minute	Day_of_week
0	1	506	657	0.74	Image	3	11	2019-05-12 21:18:39	2019	5	12	21	18	Sunday
1	1	385	657	0.57	Image	0	1	2019-05-12 14:25:55	2019	5	12	14	25	Sunday
2	1	373	666	0.55	Images	4	15	2019-05-03 15:18:37	2019	5	3	15	18	Friday
3	1	312	667	0.46	Images	4	1	2019-05-02 16:40:38	2019	5	2	16	40	Thursday
4	1	513	674	0.75	Images	5	0	2019-04-25 16:51:58	2019	4	25	16	51	Thursday
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
178917	1089	9	821	0.07	Image	0	1	2018-11-29 14:40:11	2018	11	29	14	40	Thursday
178918	1089	3	822	0.02	Image	3	0	2018-11-28 20:59:19	2018	11	28	20	59	Wednesday
178919	1089	2	822	0.01	Image	0	0	2018-11-28 16:27:32	2018	11	28	16	27	Wednesday
178920	1089	14	822	0.10	Image	0	2	2018-11-28 14:37:13	2018	11	28	14	37	Wednesday
178921	1089	9	823	0.07	Image	0	0	2018-11-27 14:27:42	2018	11	27	14	27	Tuesday

178922 rows × 14 columns

	User uuid	Likes	Days passed from post	Likes Score	Numer of Tags	Numer of Comments	Year	Month	Day	Hour	Minute
count	178922.000000	1.789220e+05	178922.000000	178922.000000	178922.000000	1.789220e+05	178922.000000	178922.000000	178922.000000	178922.000000	178922.000000
mean	546.316948	4.298806e+04	447.477510	0.270961	3.578045	5.622285e+02	2019.412107	6.764814	15.694219	13.396435	28.741737
std	317.200142	2.604648e+05	492.620418	0.210007	6.877008	1.201380e+04	1.374731	3.753390	8.824645	7.385343	17.678904
min	1.000000	0.000000e+00	0.000000	0.000000	0.000000	0.000000e+00	2010.000000	1.000000	1.000000	0.000000	0.000000
25%	273.000000	1.340000e+02	107.000000	0.100000	0.000000	4.000000e+00	2019.000000	3.000000	8.000000	7.000000	13.000000
50%	540.000000	1.064000e+03	260.000000	0.220000	0.000000	2.300000e+01	2020.000000	7.000000	16.000000	16.000000	29.000000
75%	829.000000	1.074300e+04	612.000000	0.390000	3.000000	1.357500e+02	2020.000000	10.000000	23.000000	20.000000	44.000000
max	1089.000000	1.544569e+07	3775.000000	1.000000	31.000000	2.907644e+06	2021.000000	12.000000	31.000000	23.000000	59.000000

Fig. 5. Summary Statistics

or influencers who wish to interact with the maximum number of followers at a time.

hyperref

```
Data columns (total 14 columns):
 #   Column                 Non-Null Count    Dtype
---  ------                 --------------    -----
 0   User uuid              178922 non-null   int64
 1   Likes                  178922 non-null   int64
 2   Days passed from post  178922 non-null   int64
 3   Likes Score            178922 non-null   float64
 4   Type                   178922 non-null   object
 5   Numer of Tags          178922 non-null   int64
 6   Numer of Comments      178922 non-null   int64
 7   Date Posted            178922 non-null   object
 8   Year                   178922 non-null   int64
 9   Month                  178922 non-null   int64
 10  Day                    178922 non-null   int64
 11  Hour                   178922 non-null   int64
 12  Minute                 178922 non-null   int64
 13  Day_of_week            178922 non-null   object
```

Fig. 6. Basic info about columns

make many comments on a single post. This gives rise to fake comments(bot comments) or spamming which may lead to incorrect estimations.

V. CONCLUSION

Number of likes depend on number of people who view it or the reach. We consider date and time as a parameter to predict when the number of Instagram users will be maximum. This way we can get a clear idea of when to make an Instagram post to get maximum number of likes. This will help people who want to promote their products through social media platforms

REFERENCES

[1] Tri Hanifawati, Utan Sahiro Ritonga, and Euis Evi Puspitasari, "MANAGING BRANDS' POPULARITY ON FACEBOOK: POST TIME, CONTENT AND BRAND COMMUNICATION STRATEGIES" Journal of INDONESIAN ECONOMY and BUSINESS. Volume 34, Number 2, 2019.
[2] Risqo M.Wahid ,Mumahhad Wadud "SOCIAL MEDIA MARKETING ON INSTAGRAM: WHEN IS THE MOST EFFECTIVE POSTING TIMING?", 2020.
[3] https://www.kaggle.com/vasileiosmpletsos/1100-instagram-users-datetime-posts-data has the dataset that is used.
[4] https://www.kaggle.com/datasets
[5] https://en.wikipedia.org/wiki/List_of_social_platforms_with_at_least_100_million_activ
[6] https://towardsdatascience.com/predict-the-number-of-likes-on-instagram-a7ec5c020203
[7] https://github.com/gvsi/instagram-like-predictor