TikTok 2022 Influencers Analysis

Submitted to:

Lynn A. Agre, MPH, PhD

Statistics 390: Section M1

October 15, 2022

By:

Bingyi Long

193000028

Abstract:

This study will investigate the relationship between subscription amount and engagement rate, plus the association between influencers' niche field and how likely their content is to become viral. The purpose is to help sponsors identify appropriate TikTok influencers for their brands and optimize marketing. This analysis contains 31 TikTok influencers in Sep 2022 with average followers of around 20 Million and examines their watch rate, engagement rate, and virality. The result shows that Macro-influencers who have subscribers between 10M to 20M and Mid-tier Influencers with subscribers less than 10M both have an average higher engagement rate than top influencers who have subscribers over 20M, which indicates user conversion rate is more likely to be maximized by macro influencers.

Background:

As TikTok became the most downloaded app since 2018, businesses have recognized its power to reach audiences, spread the word, and engage with customers. When companies look for TikTok influencers to represent their brands and collaborate, several factors are considered: organic engagement and high virality of content, which indicates the user conversion rate, that is, how many users are attracted by the news, ads, or media actually become customers; influencers credits, that is their authoritative in the area and how much their subscribers believe in them; plus content's ability to resonate with audiences, in other words, how ads attract people. Therefore, this study will investigate what factors are related to virality and engagement rate and how they are influenced by these factors, such as the number of followers and the influencer's niche. The database source is "Social media influencers-TikTok Sep 2022" from Kaggle: https://www.kaggle.com/datasets/ramjasmaurya/top-1000-social-media-channels.

Methods:

The dataset contains 31 TikTok influencers with an average of 20M subscribers that corporations may have the intention to work with. This data set collects each influencer's subscribers amount plus average views, likes, comments, and shares in their posts. From there, the watch rate is calculated by average views/number of subscribers; the engagement rate is calculated by (average likes+comments+shares)/average views. Then, influencers are categorized into influential rank as variable "influencer" based on: "Top influencer"-subscribers greater than 20M, "Macro influencer"-subscribers between 10M and 20M; and "Mid-tier influencer"-subscribers less than 10M. Also, influencers are categorized into viral rank as variable "viral" based on: "viral"-average views between 1M and 5M, "viral sensation"-average views between 5M and 10M, and "viral creator"-average views greater than 10M. Besides, the data includes each influencer's niche: "DIY," "Dance," "Fashion," "Film," "Joke," and "Music."

The code is as following:

```
data Tiktok_influencers;
  Input tiktok_account $ subscribers_M views_avg_M likes_avg_K comments_avg_K shares_avg_K niche$;

  subscribers_M1=subscribers_M-0.2;
  comments_avg_K1=comments_avg_K-20;

  watch_rate=views_avg_M/subscribers_M;
  engagement_rate=(likes_avg_K+comments_avg_K+shares_avg_K)/(views_avg_M*1000);

  Datalines;
Stray_Kids 13.8 6.4 2300 50.2 34.2 dance
Khabane_lame 149.2 17.3 2300 15.2 8.7 joke
scarlett 2.1 17.9 845.8 53.9 6.3 music
Addison_Rae 88.7 22 906.6 7.6 26.2 fashion
enhypen 12.7 4.3 1100 24.8 14.3 dance
Kimberly_Loaiza 67.2 8.8 1200 12.1 4.3 fashion
```

```
ramonvitor 8.7 9.3 1900 7.3 2.4 DIY
TOMORROW_X_TOGETHER 18.1 3.8 1100 27.4 9.5 dance
emraynea 1.1 5.1 1400 6.6 13.5 joke
blackpinkofficial 31.2 4.5 839.6 13.9 22.2 dance

Lele_Pons 27.9 12.2 1000 1.5 1.7 joke
Matt 3.2 3.2 812 10.9 4.6 music
Rayssa_Buq 7.2 4.3 775.3 7.9 4.1 joke
charlidamelio 0.0646 5.4 804 6.9 1.7 fashion
Pk 19 4.8 893.3 3 5.6 joke
Selena_Gomez 43.2 5.2 804.9 4.2 4.5 fashion

Bella_Poarch 91.4 7.4 677.3 4.9 1.1 fashion
Max_Taylor 2.8 3.5 572.6 5.2 9.5 joke
Jordan_e_Mel 10.2 4.4 436.1 3 12.2 joke
Hannah_Stocking 26.5 7.8 840.8 1.6 2.5 joke
Xóchitl_Gómez 7.3 7 1300 26 2.9 fashion
Ruben_Tuesta 21.3 5.5 802.3 1.6 4.4 joke
Charlieee 5.1 12 470.2 1.7 2.9 joke

Calfreezy 0.96 6.5 869.1 1.7 2.4 joke
loveofhuns_x 0.144 3.5 382.4 8.3 8.9 film
Liza_Koshy 29.9 4.9 802.8 2.2 3.6 fashion
Victor_Melo 21.9 4.5 736.7 2.5  4.2 joke
Wisdom_Kaye 8.5 4.6 874.4 3.5 2 fashion
Q_Park 33.6 5 0.415 3.9 6.6 joke
LIL_G 1.7 4.4 0.521 5.5 4 music
Dylan_Mulvaney 7.3 3.8 730.8 6.6 1.2 fashion

;
RUN;
PROC PRINT data= Tiktok_influencers;
```

To answer the following questions:

1. Which niche is more likely to generate viral content (measured by the categorical variable "viral")?

2. Which niche has a higher engagement rate?

3. Is there an association between subscribers' amount (or categorical variable "influencer") and the watch rate or engagement rate?

First:

Basic analysis of the dataset to find mean, median, mode, and distribution of quantitative variables: the number of subscribers, watch rate, and engagement rate. And the frequency of categorical variable: niche.

```
Proc Freq data=tiktok_influencers;
table niche;
run;

Proc Means data=tiktok_influencers;
var subscribers_M;
RUN;

Proc Means data=tiktok_influencers;
var watch_rate;
RUN;

Proc Means data=tiktok_influencers;
var engagement_rate;
RUN;

PROC univariate data=tiktok_influencers;
var subscribers_M;
run;
```

```
PROC univariate data=tiktok_influencers;
var watch_rate;
run;

PROC univariate data=tiktok_influencers;
var engagement_rate;
run;
```

Second part:

using t test to examine weather the average engagement rate is greater than 10%, and weather average subscribers amount for top 1000 2020 Tiktok influencers is differ from 20 million.

```
/*  T-test hypothesis
  H0: average engagement rate =10%
  H1: average engagement rate is greater than 10%

  H0: average tiktok influencer's subscriber amount = 20 million
  H1: average tiktok influencer's subscriber amount != 20 million
   */

Proc TTest Sides=U;
  Var engagement_rate1;
RUN;

Proc MEANS N Mean STD T PRT;
  VAR subscribers_M1;
RUN;

/*
  T test result:
  t value =3.62, p value = 0.0005< 0.05, so reject null,
  there is enough evidenc to show that average engagement rate is greater than 10%.

  t value=0.77, p value=0.448>0.05, so do not reject null,
  there is no enough evidence to show that
  average subsribers amount for top 1000 2020 Tiktok influencers
  is differ from 20 million.
*/
```

then, rearrange dataset, remove the temporary variables used in t test and add new categorical variables "viral" and "influencer". And count the frequency of these two variables.

```
Data Tiktok_influencers15(drop=subscribers_M1 comments_avg_K1);
  set Tiktok_influencers;
RUN;

Data Tiktok_influencers2;
  set Tiktok_influencers15;
  if views_avg_M>10 then viral='viral creator';
  if 5<=views_avg_M<=10 then viral='viral sensation';
  if 1<views_avg_M<5 then viral='viral';


  if subscribers_M<=10 then influencer='Mid-tier influencer';
  if 10<subscribers_M<=20 then influencer='Macro influencer';
  if 20<subscribers_M then influencer='Top influencer';
RUN;

PROC FREQ data=Tiktok_influencers2;
  Tables influencer viral;
RUN;
```

Lastly, perform a correlation test to see if there is an association between subscription amount and engagement rate.

```
/*
Correlation Test:
H0: ro=0, there is no association between subscribtion amount and engagement rate
Ha: ro!=0, there is an association between subscribtion amount and engagement rate
*/

Proc Corr Data=tiktok_influencers2;
  Var subscribers_M;
  With engagement_rate;
RUN;

/*
Result:
p value= 0.2272>0.05, do not reject null.
There is no enough evidence to show that an association between subscribtion amount and engagement rate exists
*/
```

Third part:

Visualize categorical variables: viral, influencer by vertical bar

visualize quantitative variables: amount of subscribers, engagement rate, and watch rate by histogram.

```
proc gchart data=tiktok_influencers2;
  vbar viral influencer niche;
run;

proc univariate data=tiktok_influencers2;
  histogram;
  variables subscribers_M engagement_rate watch_rate;
run;
```

Visualize the intersection between viral and influencer.

Visualize the tabulate between niche and amount of subscribers

and the tabulate between niche and engagement rate

```
PROC Freq data=tiktok_influencers2;
  Table viral* influencer;
run;

proc tabulate data=tiktok_influencers2;
  class niche;
  var subscribers_M;
  table niche,
      subscribers_M*(N Mean Max);
run;

proc tabulate data=tiktok_influencers2;
  class niche;
  var engagement_rate;
  table niche,
      engagement_rate*(N Mean Max);
run;
```

Visualize engagement rate by box plot

visualize subscribers and engagement rate by scatterplot

visualize subscribers and watch rate by scatterplot

```
proc boxplot data=tiktok_influencers2;
  plot engagement_rate*group;
RUN;


PROC Plot data=tiktok_influencers2;
  Plot subscribers_M*engagement_rate='+';
RUN;


Proc sgscatter data=tiktok_influencers2;
  plot subscribers_M*watch_rate;
RUN;
```

Fourth part

Perform two chi-square test to examine whether the amount of subscribers (influential rank) is independent of niche, and whether virality is independent of niche.

```
/*
Chi-Square Test
H0:influencer is independent of niche
H1:influencer is not identipendent of niche

H0:virality is independent of niche
H1:virality is not identipendent of niche

*/
Proc Freq data=tiktok_influencers2;
  Table influencer*niche/CHISQ Expected Deviation Norow Nocol Nopercent;
RUN;


Proc Freq data=tiktok_influencers2;
  Table viral*niche/CHISQ Expected Deviation Norow Nocol Nopercent;
RUN;

/*
Chi-Square Conclusion:
With Chi-Square Value 19.1, and p-value is 0.039, which is less than 0.05, so we reject null,
and conclude that there is an association between influencer's subscription and niche.

With Chi-Square Value 8.0422, and p-value is 0.6247, which is greater than 0.05, so we do not reject null,
and conclude that there is not an association between virality and niche.


*/
```

and an Anova test to examine whether regardless of the subscription amount, the mean of engagement rate will be the same.

```
/* Anova Table Hypothesis:
  H0: regardless of the subscription, mean of engagement rate is the same.
  u1=u2=u3
  H1: At least one of the means is different from the others
  The significance level is a=0.05
  */

Proc GLM Data=tiktok_influencers2;
```

```
  Class influencer;
  Model engagement_rate=influencer;
  RUN;

  /*
  ANOVA Conclusion:
  with p-value =0.001<0.05, we do not reject null.
  There is sufficient evidence to rejection of the claim that all means are equal.
  */
```

Using paired t-test to survey if there is a difference between watch rate and engagement rate in each account.

Then, perform two-tailed significance test to survey if a non-zero correlation could exist between subscribers amount and engagement rate or watch rate.

```
  /*paired t test hypotheses:
  H0: u1=u2  the paired population means are equal
  H1: u1!=u2 the paired population means are not equal
  where u1 is the engagement rate of each account
  and u2 is the watch rate in each account
  */

  PROC TTEST DATA=tiktok_influencers2;
    PAIRED engagement_rate*watch_rate;
  RUN;

  /*
  given p=0.1174>0.05, do not reject null,
  and there is no enough evidence to
  conclude engagement_rate does not equal to watch_rate in each account
  */

  /*
  Two-Tailed significance test:
    H0: p=0, there is no association
    H1: p!=0, a none zero correlation could exist.
  */

  PROC CORR DATA=tiktok_influencers2 PLOTS=SCATTER(NVAR=all);
    VAR subscribers_M engagement_rate;
  RUN;

  PROC CORR DATA=tiktok_influencers2 PLOTS=SCATTER(NVAR=all);
    VAR subscribers_M watch_rate;
  RUN;

  /*
  p-value=0.2272>0.05, do not reject null,
  no enough evidence to prove that a none zero correlation could exist
  between subscription amount and engagement rate.

  p-value=0.245>0.05, do not reject null,
  no enough evidence to prove that a none zero correlation could exist
  between subscription amount and watch rate.
  */
```

All results are attached at the end of paper.

Results:

From analysis, we found there is enough evidence to show that the average engagement rate is greater than 10%, and not enough evidence to show that the average number of subscribers amount for the top 1000 2020 TikTok influencers is different from 20 million.

There is not enough evidence to show an association exists between subscription amount and engagement rate.

There is an association between subscription amount and niche.

There is no association between virality and niche.

There is sufficient evidence to reject that different subscription amount has the same engagement rate.

There is not enough evidence to reject that different niche has the same virality.

There is not enough evidence to conclude that the engagement rate does not equal watch_rate in each account.

No enough evidence to prove that a none zero correlation could exist between subscription amount and engagement rate.

No enough evidence to prove that a none zero correlation could exist between subscription amount and watch rate.

In summary, TikTok subscription amount or influential rank doesn't show a correlation with watch rate or engagement rate.

Scatter Plot
With 95% Prediction Ellipse

Observations	31
Correlation	-0.223
p-Value	0.2272

Scatter Plot
With 95% Prediction Ellipse

Observations	31
Correlation	-0.215
p-Value	0.2453
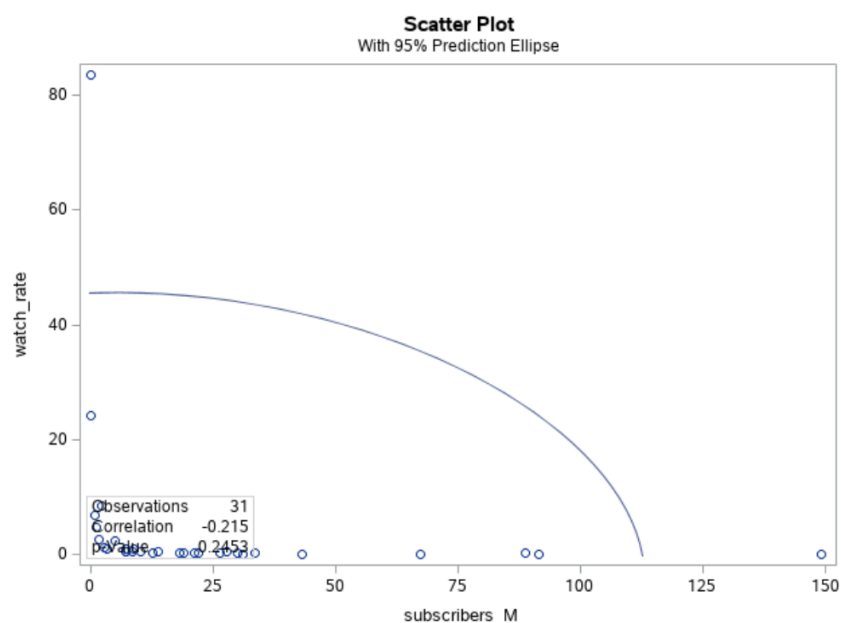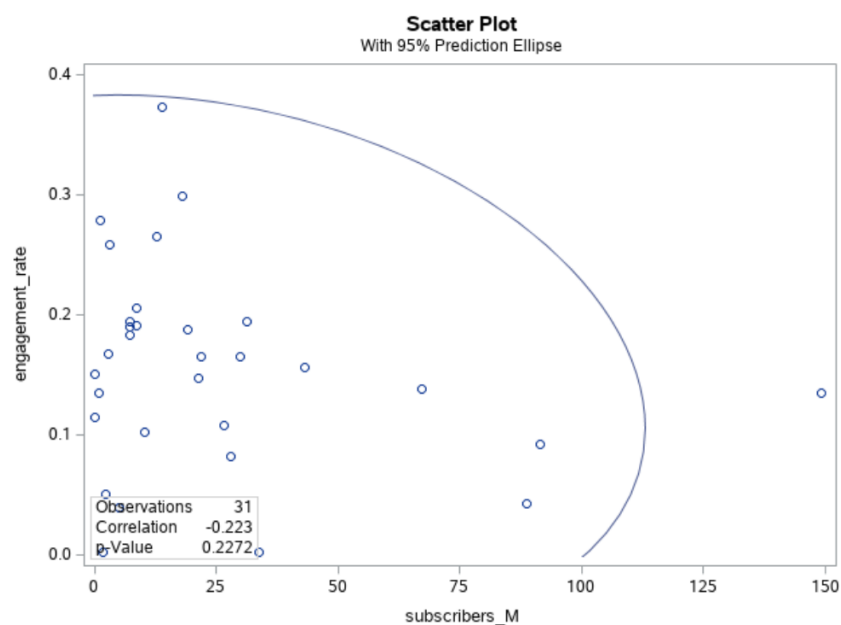
Niche doesn't differ the virality of influencers' content.
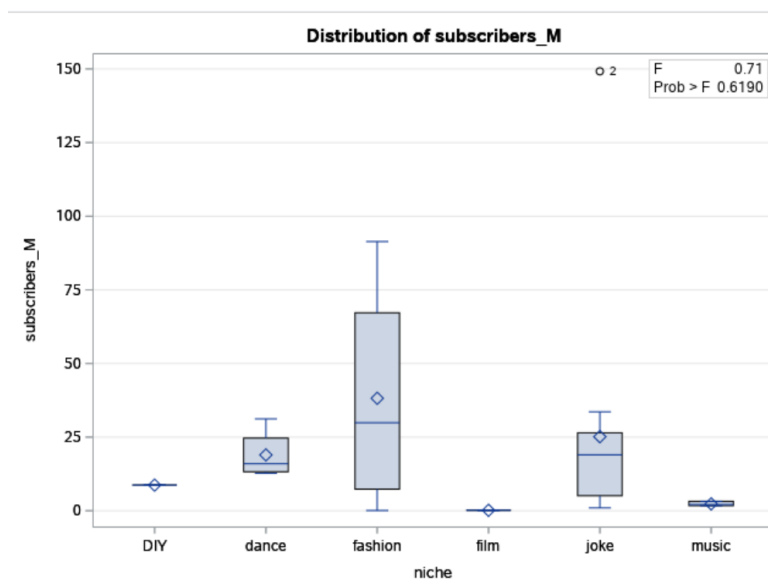
However, different subscription amount has different engagement rate.
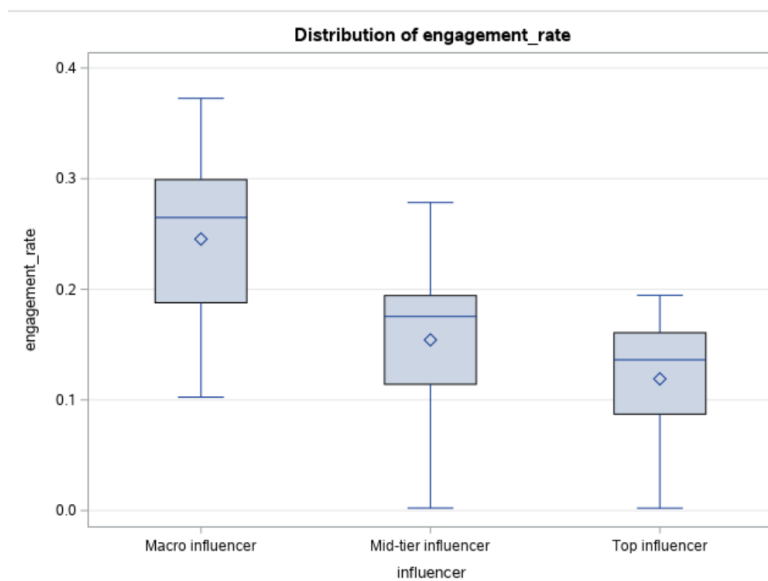
and niche does influence the subscription amount.

Distribution of subscribers_M
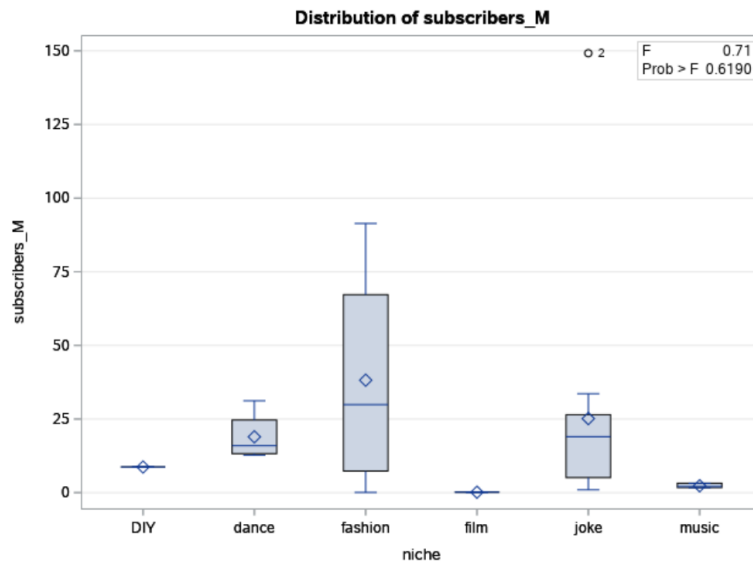
but niche doesn't influence engagement rate.

Distribution of engagement_rate
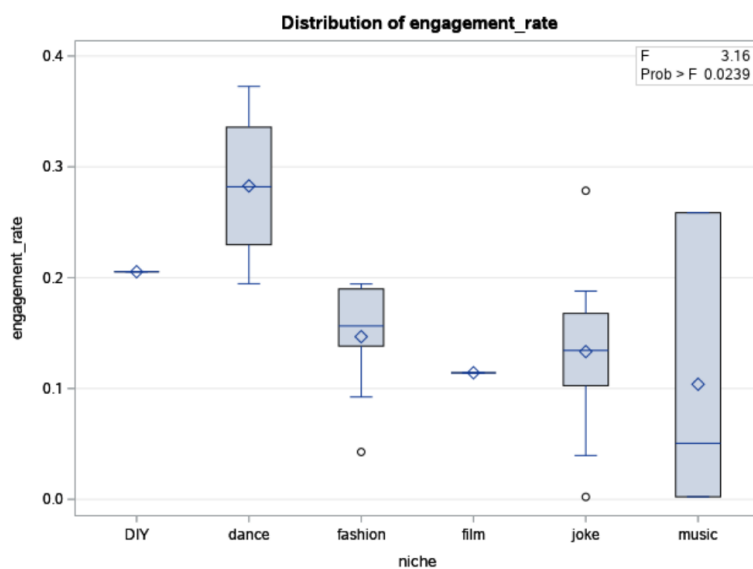
Interpretations:

The engagement rate or watch rate doesn't increase as the subscription amount increases. In reality, macro-influencers who have subscriptions between 5M and 10 M have the highest engagement rate, and mid-tier influencers who have subscriptions between 1 M and 5 have around the same engagement rate as Top influencers who have subscriptions of more than 20M. This finding is important because it shows that the highest followers on Tiktok don't make influencers' content engaging or interactive, which might lead to a lower user conversion rate because audiences tend to swipe away from this content sooner after watching rather than liking, leaving a comment, or sharing the post. In this case, audiences are less likely to respond to ads or external links of sponsored brands, so they are less likely to become customers. And the study shows that macro influencers have the highest engagement rate, so their ads can get more responses from audiences.

Second, although niche does influence the number of subscriptions, it doesn't differ in the potential of generating viral content. This result is significant because viral content best resonates with audiences and fosters secondary creation, which reaches audiences even outside of the influencer's followers. Thus, when corporations look for

influencers to work with, the niche doesn't prevent the content from going viral. For example, a joking type of video or a fashion type of video can both give rise to wide discussions and shared emotions with audiences.

For the initial questions:

1. Which niche is more likely to generate viral content (measured by the categorical variable "viral")?

2. Which niche has a higher engagement rate?

3. Is there an association between subscribers' amount (or categorical variable "influencer") and the watch rate or engagement rate?

The study shows that niche doesn't determine virality or engagement rate.

Subscribers' amount does influence engagement rate, but not in a correlated relationship, and subscribers' amount does not influence watch rate. This shows that influencers who have subscription amounts from low to high will have similar watch rates, which results in the number of people who watch the content ranging from low to high, but macro-influencers will have more engagement rate. This is a trade-off for corporations, whether the number of views of an ad or the number of people who respond to the ad is more important to the marketing campaign.

The limitations of analysis are it doesn't identify what factors result in the viral content. Similarly, it cannot predict what content will become viral. It only rules out that niche isn't a determining factor.

This research mainly helps business to build their marketing strategy with Tiktok influencers, but it can also provide insight to Tiktok creators who aims to grow their followers.

The future study can examine the effect of updated news and discussion on virality, such as the videos related to mental health issues during the pandemic or political jokes around voting season are more likely to become viral on Tiktok.

References

Alves, H., Fernandes, C., & Raposo, M. (2016). Social Media Marketing: A literature review and implications. *Psychology & Marketing*, *33*(12), 1029–1038. https://doi.org/10.1002/mar.20936

Montag, C., Yang, H., & Elhai, J. D. (2021). On the psychology of TikTok use: A first glimpse from empirical findings. *Frontiers in Public Health*, *9*. https://doi.org/10.3389/fpubh.2021.641673

Obs	tiktok_account	subscribers_M	views_avg_M	likes_avg_K	comments_avg_K	shares_avg_K	niche	watch_rate	engagement_rate	engagement_rate1	subscribers_M1	group
1	Stray_Ki	13.800	6.4	2300.00	50.2	34.2	dance	0.4638	0.37256	0.27256	-6.200	1
2	Khabane_	149.200	17.3	2300.00	15.2	8.7	joke	0.1160	0.13433	0.03433	129.200	1
3	scarlett	2.100	17.9	845.80	53.9	6.3	music	8.5238	0.05061	-0.04939	-17.900	1
4	Addison_	88.700	22.0	906.60	7.6	26.2	fashion	0.2480	0.04275	-0.05725	68.700	1
5	enhypen	12.700	4.3	1100.00	24.8	14.3	dance	0.3386	0.26491	0.16491	-7.300	1
6	Kimberly	67.200	8.8	1200.00	12.1	4.3	fashion	0.1310	0.13823	0.03823	47.200	1
7	ramonvit	8.700	9.3	1900.00	7.3	2.4	DIY	1.0690	0.20534	0.10534	-11.300	1
8	TOMORROW	18.100	3.8	1100.00	27.4	9.5	dance	0.2099	0.29918	0.19918	-1.900	1
9	emraynea	1.100	5.1	1400.00	6.6	13.5	joke	4.6364	0.27845	0.17845	-18.900	1
10	blackpin	31.200	4.5	839.60	13.9	22.2	dance	0.1442	0.19460	0.09460	11.200	1
11	Lele_Pon	27.900	12.2	1000.00	1.5	1.7	joke	0.4373	0.08223	-0.01777	7.900	1
12	Matt	3.200	3.2	812.00	10.9	4.6	music	1.0000	0.25859	0.15859	-16.800	1
13	Rayssa_B	7.200	4.3	775.30	7.9	4.1	joke	0.5972	0.18309	0.08309	-12.800	1
14	charlida	0.065	5.4	804.00	6.9	1.7	fashion	83.5913	0.15048	0.05048	-19.935	1
15	Pk	19.000	4.8	893.30	3.0	5.6	joke	0.2526	0.18790	0.08790	-1.000	1
16	Selena_G	43.200	5.2	804.90	4.2	4.5	fashion	0.1204	0.15646	0.05646	23.200	1
17	Bella_Po	91.400	7.4	677.30	4.9	1.1	fashion	0.0810	0.09234	-0.00766	71.400	1
18	Max_Tayl	2.800	3.5	572.60	5.2	9.5	joke	1.2500	0.16780	0.06780	-17.200	1
19	Jordan_e	10.200	4.4	436.10	3.0	12.2	joke	0.4314	0.10257	0.00257	-9.800	1
20	Hannah_S	26.500	7.8	840.80	1.6	2.5	joke	0.2943	0.10832	0.00832	6.500	1
21	Xóchitl	7.300	7.0	1300.00	26.0	2.9	fashion	0.9589	0.18984	0.08984	-12.700	1
22	Ruben_Tu	21.300	5.5	802.30	1.6	4.4	joke	0.2582	0.14696	0.04696	1.300	1
23	Charliee	5.100	12.0	470.20	1.7	2.9	joke	2.3529	0.03957	-0.06043	-14.900	1
24	Calfreez	0.960	6.5	869.10	1.7	2.4	joke	6.7708	0.13434	0.03434	-19.040	1
25	loveofhu	0.144	3.5	382.40	8.3	8.9	film	24.3056	0.11417	0.01417	-19.856	1
26	Liza_Kos	29.900	4.9	802.80	2.2	3.6	fashion	0.1639	0.16502	0.06502	9.900	1
27	Victor_M	21.900	4.5	736.70	2.5	4.2	joke	0.2055	0.16520	0.06520	1.900	1
28	Wisdom_K	8.500	4.6	874.40	3.5	2.0	fashion	0.5412	0.19128	0.09128	-11.500	1
29	Q_Park	33.600	5.0	0.42	3.9	6.6	joke	0.1488	0.00218	-0.09782	13.600	1
30	LIL_G	1.700	4.4	0.52	5.5	4.0	music	2.5882	0.00228	-0.09772	-18.300	1
31	Dylan_Mu	7.300	3.8	730.80	6.6	1.2	fashion	0.5205	0.19437	0.09437	-12.700	1

The FREQ Procedure

niche	Frequency	Percent	Cumulative Frequency	Cumulative Percent
DIY	1	3.23	1	3.23
dance	4	12.90	5	16.13
fashion	9	29.03	14	45.16
film	1	3.23	15	48.39
joke	13	41.94	28	90.32
music	3	9.68	31	100.00

The MEANS Procedure

Analysis Variable : subscribers_M				
N	Mean	Std Dev	Minimum	Maximum
31	24.5796323	33.1914239	0.0646000	149.2000000

The MEANS Procedure

Analysis Variable : watch_rate				
N	Mean	Std Dev	Minimum	Maximum
31	4.6048607	15.3621754	0.0809628	83.5913313

The MEANS Procedure

Analysis Variable : engagement_rate				
N	Mean	Std Dev	Minimum	Maximum
31	0.1553536	0.0850748	0.0021830	0.3725625

The UNIVARIATE Procedure
Variable: subscribers_M

Moments			
N	31	Sum Weights	31
Mean	24.5796323	Sum Observations	761.9686
Std Deviation	33.1914239	Variance	1101.67062
Skewness	2.38357656	Kurtosis	6.21655036
Uncorrected SS	51779.0265	Corrected SS	33050.1185
Coeff Variation	135.036291	Std Error Mean	5.96135571

Basic Statistical Measures			
Location		Variability	
Mean	24.57963	Std Deviation	33.19142
Median	12.70000	Variance	1102
Mode	7.30000	Range	149.13540
		Interquartile Range	26.70000

Tests for Location: Mu0=0				
Test	Statistic		p Value	
Student's t	t	4.123161	Pr > |t|	0.0003
Sign	M	15.5	Pr >= |M|	<.0001
Signed Rank	S	248	Pr >= |S|	<.0001

Quantiles (Definition 5)	
Level	Quantile
100% Max	149.2000
99%	149.2000
95%	91.4000
90%	67.2000
75% Q3	29.9000
50% Median	12.7000
25% Q1	3.2000
10%	1.1000
5%	0.1440
1%	0.0646
0% Min	0.0646

Extreme Observations			
Lowest		Highest	
Value	Obs	Value	Obs
0.0646	14	43.2	16
0.1440	25	67.2	6
0.9600	24	88.7	4
1.1000	9	91.4	17
1.7000	30	149.2	2

The UNIVARIATE Procedure
Variable: watch_rate

Moments			
N	31	Sum Weights	31
Mean	4.60486068	Sum Observations	142.750681
Std Deviation	15.3621754	Variance	235.996432
Skewness	4.90131207	Kurtosis	25.250792
Uncorrected SS	7737.23997	Corrected SS	7079.89297
Coeff Variation	333.607821	Std Error Mean	2.75912815

Basic Statistical Measures			
Location		Variability	
Mean	4.604861	Std Deviation	15.36218
Median	0.437276	Variance	235.99643
Mode	.	Range	83.51037
		Interquartile Range	1.04452

Tests for Location: Mu0=0				
Test	Statistic		p Value	
Student's t	t	1.668955	Pr > |t|	0.1055
Sign	M	15.5	Pr >= |M|	<.0001
Signed Rank	S	248	Pr >= |S|	<.0001

Quantiles (Definition 5)

Level	Quantile
100% Max	83.5913313
99%	83.5913313
95%	24.3055556
90%	6.7708333
75% Q3	1.2500000
50% Median	0.4372760
25% Q1	0.2054795
10%	0.1309524
5%	0.1159517
1%	0.0809628
0% Min	0.0809628

Extreme Observations			
Lowest		Highest	
Value	Obs	Value	Obs
0.0809628	17	4.63636	9
0.1159517	2	6.77083	24
0.1203704	16	8.52381	3
0.1309524	6	24.30556	25
0.1442308	10	83.59133	14

The UNIVARIATE Procedure
Variable: engagement_rate

Moments			
N	31	Sum Weights	31
Mean	0.15535362	Sum Observations	4.81596214
Std Deviation	0.08507485	Variance	0.00723773
Skewness	0.3240225	Kurtosis	0.38891871
Uncorrected SS	0.96530903	Corrected SS	0.21713188
Coeff Variation	54.7620634	Std Error Mean	0.01527989

Basic Statistical Measures			
Location		Variability	
Mean	0.155354	Std Deviation	0.08507
Median	0.156462	Variance	0.00724
Mode	.	Range	0.37038
		Interquartile Range	0.09180

Tests for Location: Mu0=0				
Test		Statistic	p Value	
Student's t	t	10.16719	Pr > |t|	<.0001
Sign	M	15.5	Pr >= |M|	<.0001
Signed Rank	S	248	Pr >= |S|	<.0001

Quantiles (Definition 5)	
Level	Quantile
100% Max	0.3725625
99%	0.3725625
95%	0.2991842
90%	0.2649070
75% Q3	0.1943684
50% Median	0.1564615
25% Q1	0.1025682
10%	0.0427455
5%	0.0022775
1%	0.0021830
0% Min	0.0021830

Extreme Observations			
Lowest		Highest	
Value	Obs	Value	Obs
0.0021830	29	0.258594	12
0.0022775	30	0.264907	5
0.0395667	23	0.278451	9
0.0427455	4	0.299184	8
0.0506145	3	0.372563	1

The TTEST Procedure

Variable: engagement_rate1

N	Mean	Std Dev	Std Err	Minimum	Maximum
31	0.0554	0.0851	0.0153	-0.0978	0.2726

Mean	95% CL Mean		Std Dev	95% CL Std Dev	
0.0554	0.0294	Infty	0.0851	0.0680	0.1137

DF	t Value	Pr > t
30	3.62	0.0005

Distribution of engagement_rate1
With 95% Upper Confidence Interval for Mean

Q-Q Plot of engagement_rate1
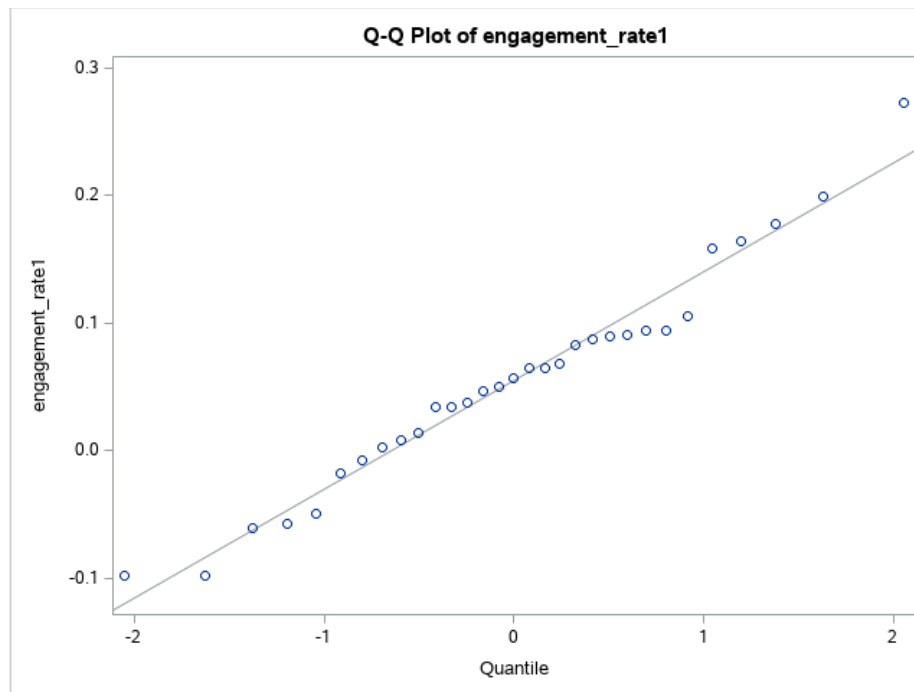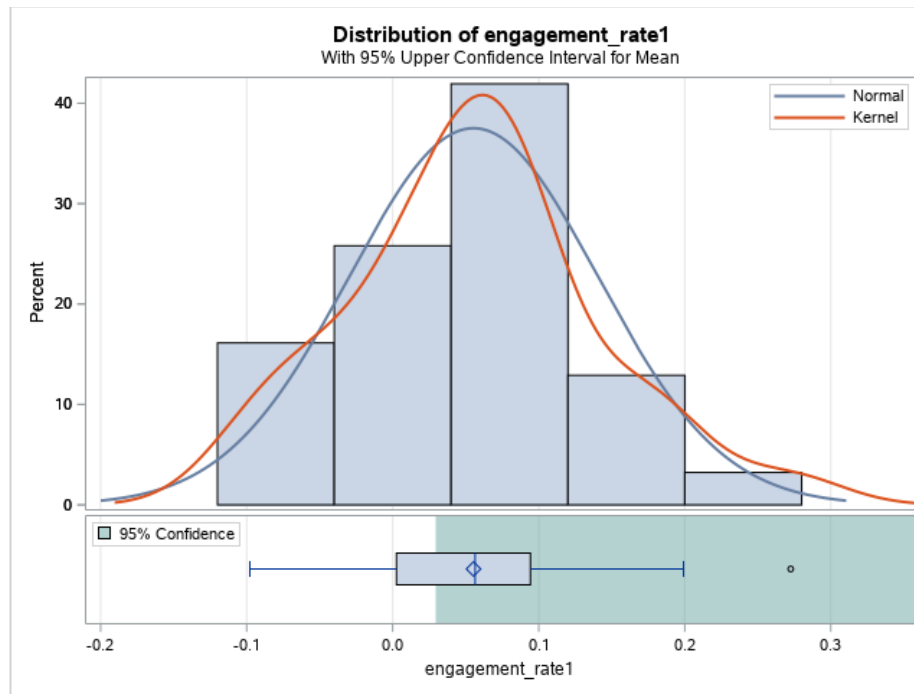
The MEANS Procedure

Analysis Variable : subscribers_M1				
N	Mean	Std Dev	t Value	Pr > |t|
31	4.5796323	33.1914239	0.77	0.4484

Obs	tiktok_account	subscribers_M	views_avg_M	likes_avg_K	comments_avg_K	shares_avg_K	niche	watch_rate	engagement_rate	engagement_rate1	group	viral	influencer
1	Stray_Ki	13.800	6.4	2300.00	50.2	34.2	dance	0.4638	0.37256	0.27256	1	viral sensati	Macro influencer
2	Khabane_	149.200	17.3	2300.00	15.2	8.7	joke	0.1160	0.13433	0.03433	1	viral creator	Top influencer
3	scarlett	2.100	17.9	845.80	53.9	6.3	music	8.5238	0.05061	-0.04939	1	viral creator	Mid-tier influencer
4	Addison_	88.700	22.0	906.60	7.6	26.2	fashion	0.2480	0.04275	-0.05725	1	viral creator	Top influencer
5	enhypen	12.700	4.3	1100.00	24.8	14.3	dance	0.3386	0.26491	0.16491	1	viral	Macro influencer

Obs	tiktok_account	subscribers_M	views_avg_M	likes_avg_K	comments_avg_K	shares_avg_K	niche	watch_rate	engagement_rate	engagement_rate1	group	viral	influencer
6	Kimberly	67.200	8.8	1200.00	12.1	4.3	fashion	0.1310	0.13823	0.03823	1	viral sensati	Top influencer
7	ramonvit	8.700	9.3	1900.00	7.3	2.4	DIY	1.0690	0.20534	0.10534	1	viral sensati	Mid-tier influencer
8	TOMORROW	18.100	3.8	1100.00	27.4	9.5	dance	0.2099	0.29918	0.19918	1	viral	Macro influencer
9	emraynea	1.100	5.1	1400.00	6.6	13.5	joke	4.6364	0.27845	0.17845	1	viral sensati	Mid-tier influencer
10	blackpin	31.200	4.5	839.60	13.9	22.2	dance	0.1442	0.19460	0.09460	1	viral	Top influencer
11	Lele_Pon	27.900	12.2	1000.00	1.5	1.7	joke	0.4373	0.08223	-0.01777	1	viral creator	Top influencer
12	Matt	3.200	3.2	812.00	10.9	4.6	music	1.0000	0.25859	0.15859	1	viral	Mid-tier influencer
13	Rayssa_B	7.200	4.3	775.30	7.9	4.1	joke	0.5972	0.18309	0.08309	1	viral	Mid-tier influencer
14	charlida	0.065	5.4	804.00	6.9	1.7	fashion	83.5913	0.15048	0.05048	1	viral sensati	Mid-tier influencer
15	Pk	19.000	4.8	893.30	3.0	5.6	joke	0.2526	0.18790	0.08790	1	viral	Macro influencer
16	Selena_G	43.200	5.2	804.90	4.2	4.5	fashion	0.1204	0.15646	0.05646	1	viral sensati	Top influencer
17	Bella_Po	91.400	7.4	677.30	4.9	1.1	fashion	0.0810	0.09234	-0.00766	1	viral sensati	Top influencer
18	Max_Tayl	2.800	3.5	572.60	5.2	9.5	joke	1.2500	0.16780	0.06780	1	viral	Mid-tier influencer
19	Jordan_e	10.200	4.4	436.10	3.0	12.2	joke	0.4314	0.10257	0.00257	1	viral	Macro influencer
20	Hannah_S	26.500	7.8	840.80	1.6	2.5	joke	0.2943	0.10832	0.00832	1	viral sensati	Top influencer
21	Xóchitl	7.300	7.0	1300.00	26.0	2.9	fashion	0.9589	0.18984	0.08984	1	viral sensati	Mid-tier influencer
22	Ruben_Tu	21.300	5.5	802.30	1.6	4.4	joke	0.2582	0.14696	0.04696	1	viral sensati	Top influencer
23	Charliee	5.100	12.0	470.20	1.7	2.9	joke	2.3529	0.03957	-0.06043	1	viral creator	Mid-tier influencer
24	Calfreez	0.960	6.5	869.10	1.7	2.4	joke	6.7708	0.13434	0.03434	1	viral sensati	Mid-tier influencer
25	loveofhu	0.144	3.5	382.40	8.3	8.9	film	24.3056	0.11417	0.01417	1	viral	Mid-tier influencer
26	Liza_Kos	29.900	4.9	802.80	2.2	3.6	fashion	0.1639	0.16502	0.06502	1	viral	Top influencer
27	Victor_M	21.900	4.5	736.70	2.5	4.2	joke	0.2055	0.16520	0.06520	1	viral	Top influencer
28	Wisdom_K	8.500	4.6	874.40	3.5	2.0	fashion	0.5412	0.19128	0.09128	1	viral	Mid-tier influencer
29	Q_Park	33.600	5.0	0.42	3.9	6.6	joke	0.1488	0.00218	-0.09782	1	viral sensati	Top influencer
30	LIL_G	1.700	4.4	0.52	5.5	4.0	music	2.5882	0.00228	-0.09772	1	viral	Mid-tier influencer
31	Dylan_Mu	7.300	3.8	730.80	6.6	1.2	fashion	0.5205	0.19437	0.09437	1	viral	Mid-tier influencer

The FREQ Procedure

influencer	Frequency	Percent	Cumulative Frequency	Cumulative Percent
Macro influencer	5	16.13	5	16.13
Mid-tier influencer	14	45.16	19	61.29
Top influencer	12	38.71	31	100.00

viral	Frequency	Percent	Cumulative Frequency	Cumulative Percent
viral	14	45.16	14	45.16
viral creator	5	16.13	19	61.29
viral sensati	12	38.71	31	100.00

The CORR Procedure

1 With Variables:	engagement_rate
1 Variables:	subscribers_M

Simple Statistics

Variable	N	Mean	Std Dev	Sum	Minimum	Maximum
engagement_rate	31	0.15535	0.08507	4.81596	0.00218	0.37256
subscribers_M	31	24.57963	33.19142	761.96860	0.06460	149.20000

Pearson Correlation Coefficients, N = 31
Prob > |r| under H0: Rho=0

	subscribers_M
engagement_rate	-0.22332 0.2272

FREQUENCY

influencer

FREQUENCY

niche
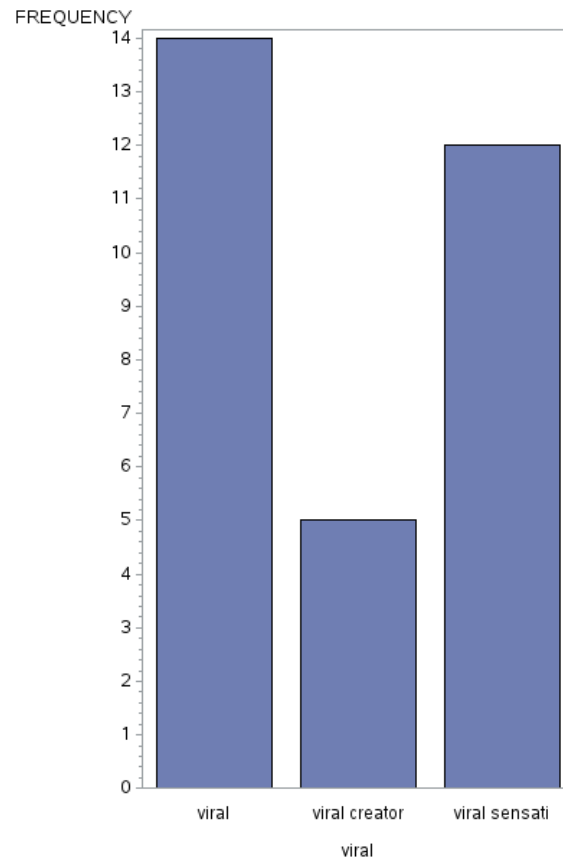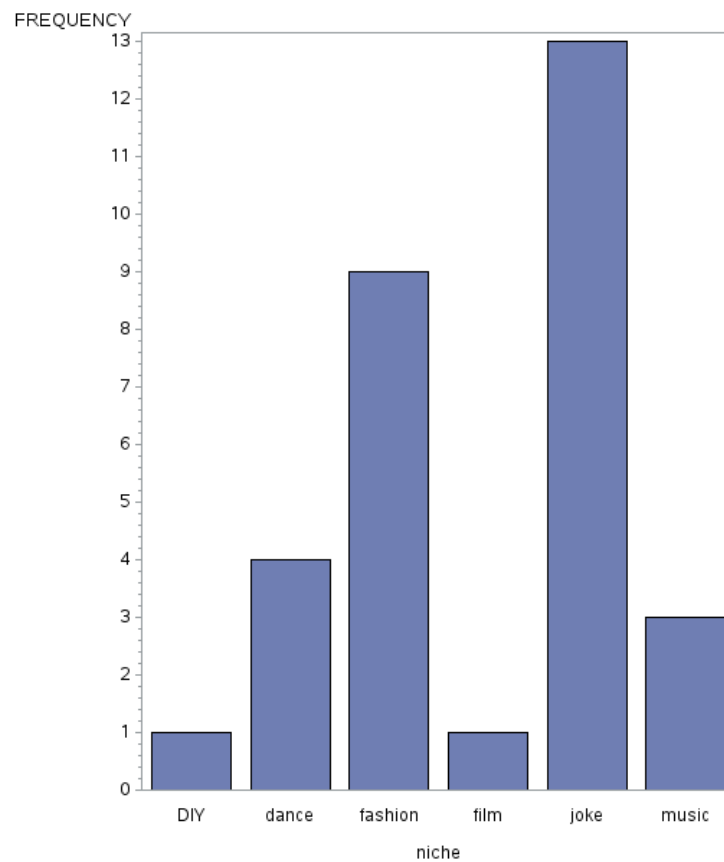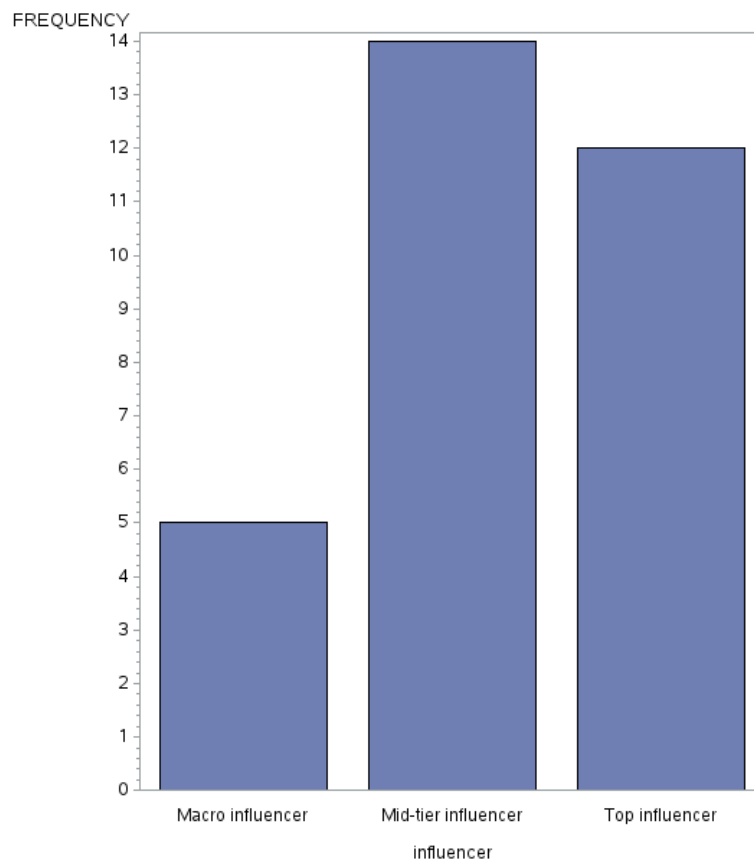
The UNIVARIATE Procedure
Variable: subscribers_M

Moments

Moments			
N	31	Sum Weights	31
Mean	24.5796323	Sum Observations	761.9686
Std Deviation	33.1914239	Variance	1101.67062
Skewness	2.38357656	Kurtosis	6.21655036
Uncorrected SS	51779.0265	Corrected SS	33050.1185
Coeff Variation	135.036291	Std Error Mean	5.96135571

Basic Statistical Measures			
Location		Variability	
Mean	24.57963	Std Deviation	33.19142
Median	12.70000	Variance	1102
Mode	7.30000	Range	149.13540
		Interquartile Range	26.70000

Tests for Location: Mu0=0				
Test		Statistic	p Value	
Student's t	t	4.123161	Pr > |t|	0.0003
Sign	M	15.5	Pr >= |M|	<.0001
Signed Rank	S	248	Pr >= |S|	<.0001

Quantiles (Definition 5)	
Level	Quantile
100% Max	149.2000
99%	149.2000
95%	91.4000
90%	67.2000
75% Q3	29.9000
50% Median	12.7000
25% Q1	3.2000
10%	1.1000
5%	0.1440
1%	0.0646
0% Min	0.0646

Extreme Observations			
Lowest		Highest	
Value	Obs	Value	Obs
0.0646	14	43.2	16
0.1440	25	67.2	6
0.9600	24	88.7	4
1.1000	9	91.4	17
1.7000	30	149.2	2

The UNIVARIATE Procedure

The UNIVARIATE Procedure
Variable: engagement_rate

Moments			
N	31	Sum Weights	31
Mean	0.15535362	Sum Observations	4.81596214
Std Deviation	0.08507485	Variance	0.00723773
Skewness	0.3240225	Kurtosis	0.38891871
Uncorrected SS	0.96530903	Corrected SS	0.21713188
Coeff Variation	54.7620634	Std Error Mean	0.01527989

Basic Statistical Measures			
Location		Variability	
Mean	0.155354	Std Deviation	0.08507
Median	0.156462	Variance	0.00724
Mode	.	Range	0.37038
		Interquartile Range	0.09180

Tests for Location: Mu0=0				
Test		Statistic	p Value	
Student's t	t	10.16719	Pr > |t|	<.0001
Sign	M	15.5	Pr >= |M|	<.0001
Signed Rank	S	248	Pr >= |S|	<.0001

Quantiles (Definition 5)	
Level	Quantile
100% Max	0.3725625
99%	0.3725625
95%	0.2991842
90%	0.2649070
75% Q3	0.1943684
50% Median	0.1564615
25% Q1	0.1025682
10%	0.0427455
5%	0.0022775
1%	0.0021830
0% Min	0.0021830

Extreme Observations			
Lowest		Highest	
Value	Obs	Value	Obs
0.0021830	29	0.258594	12
0.0022775	30	0.264907	5

Extreme Observations

Lowest		Highest	
Value	Obs	Value	Obs
0.0395667	23	0.278451	9
0.0427455	4	0.299184	8
0.0506145	3	0.372563	1

The UNIVARIATE Procedure

Distribution of engagement_rate

The UNIVARIATE Procedure
Variable: watch_rate

Moments

N	31	Sum Weights	31
Mean	4.60486068	Sum Observations	142.750681
Std Deviation	15.3621754	Variance	235.996432
Skewness	4.90131207	Kurtosis	25.250792
Uncorrected SS	7737.23997	Corrected SS	7079.89297
Coeff Variation	333.607821	Std Error Mean	2.75912815

Basic Statistical Measures

Location		Variability	
Mean	4.604861	Std Deviation	15.36218
Median	0.437276	Variance	235.99643
Mode	.	Range	83.51037
		Interquartile Range	1.04452

Tests for Location: Mu0=0

Test		Statistic	p Value	
Student's t	t	1.668955	Pr > |t|	0.1055
Sign	M	15.5	Pr >= |M|	<.0001
Signed Rank	S	248	Pr >= |S|	<.0001

Quantiles (Definition 5)

Level	Quantile
100% Max	83.5913313
99%	83.5913313
95%	24.3055556
90%	6.7708333
75% Q3	1.2500000
50% Median	0.4372760
25% Q1	0.2054795
10%	0.1309524

Quantiles (Definition 5)

Level	Quantile
5%	0.1159517
1%	0.0809628
0% Min	0.0809628

Extreme Observations

Lowest		Highest	
Value	Obs	Value	Obs
0.0809628	17	4.63636	9
0.1159517	2	6.77083	24
0.1203704	16	8.52381	3
0.1309524	6	24.30556	25
0.1442308	10	83.59133	14

The UNIVARIATE Procedure
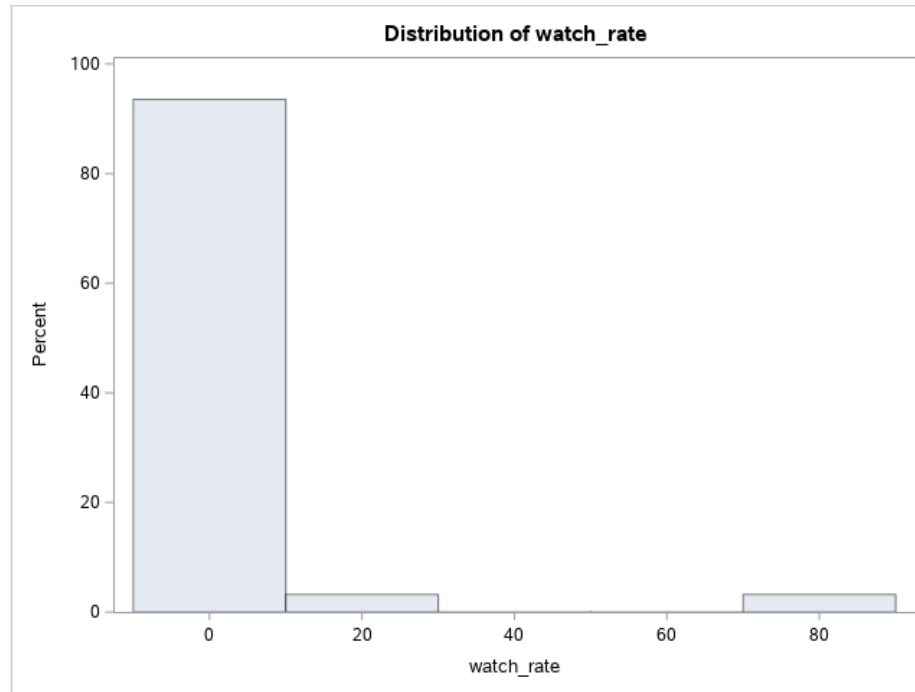
The FREQ Procedure

Frequency Percent Row Pct Col Pct	Table of viral by influencer			
		influencer		
viral	Macro influencer	Mid-tier influencer	Top influencer	Total
viral	4 12.90 28.57 80.00	7 22.58 50.00 50.00	3 9.68 21.43 25.00	14 45.16
viral creator	0 0.00 0.00 0.00	2 6.45 40.00 14.29	3 9.68 60.00 25.00	5 16.13
viral sensati	1 3.23 8.33 20.00	5 16.13 41.67 35.71	6 19.35 50.00 50.00	12 38.71
Total	5 16.13	14 45.16	12 38.71	31 100.00

	subscribers_M		
	N	Mean	Max
niche			
DIY	1	8.70	8.70
dance	4	18.95	31.20
fashion	9	38.17	91.40

	subscribers_M		
	N	Mean	Max
film	1	0.14	0.14
joke	13	25.14	149.20
music	3	2.33	3.20

	engagement_rate		
	N	Mean	Max
niche			
DIY	1	0.21	0.21
dance	4	0.28	0.37
fashion	9	0.15	0.19
film	1	0.11	0.11
joke	13	0.13	0.28
music	3	0.10	0.26
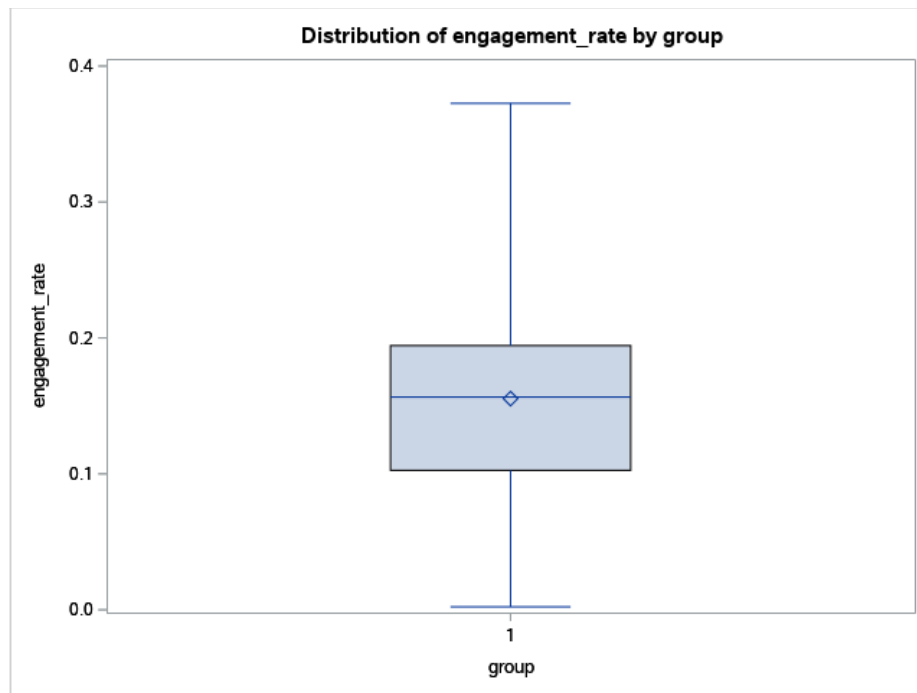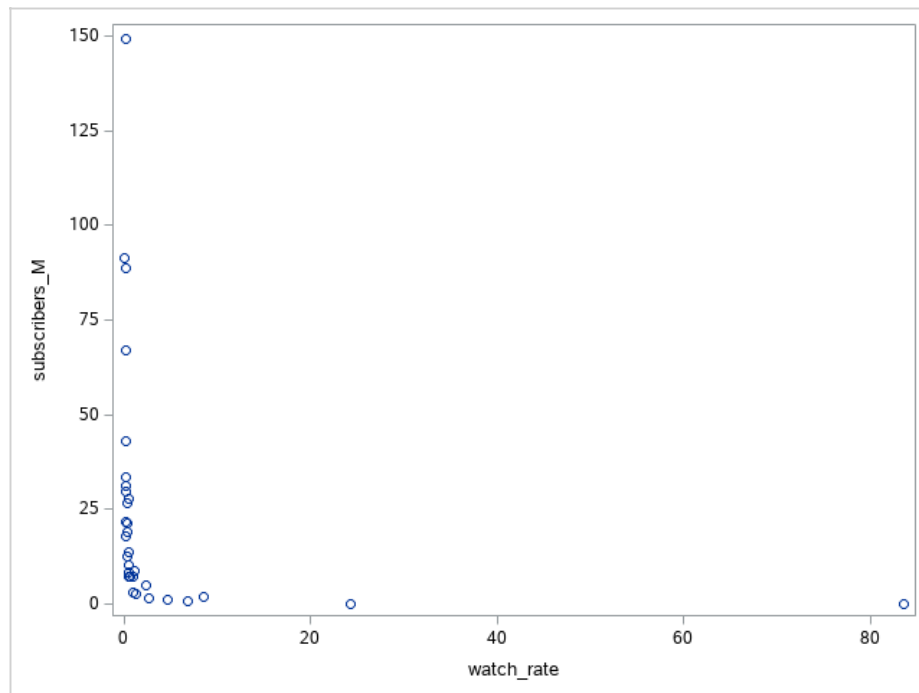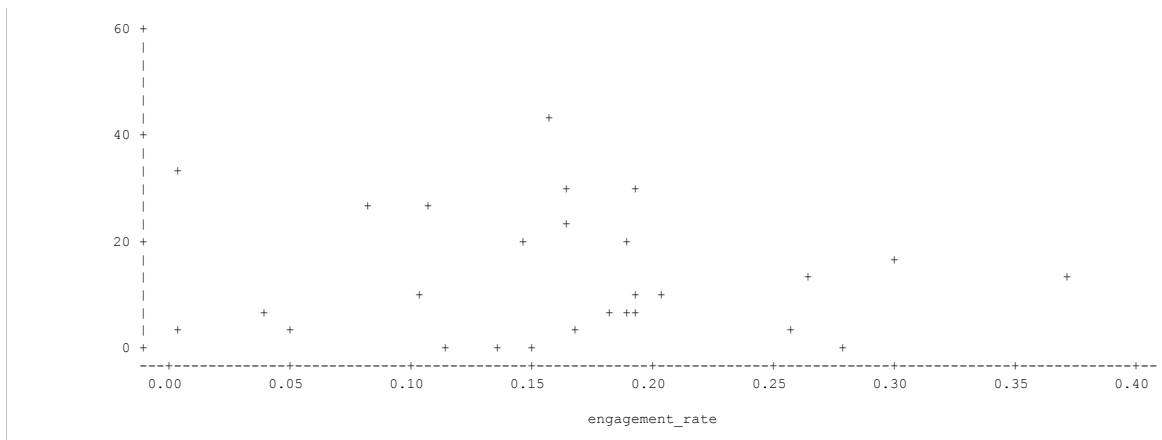
Distribution of engagement_rate by group

```
Plot of subscribers_M*engagement_rate.  Symbol used is '+'.

              160 +
                  |
                  |
                  |                                   +
                  |
              140 +
                  |
                  |
                  |
                  |
              120 +
                  |
subscribers_M     |
                  |
              100 +
                  |
                  |              +             +
                  |
               80 +
                  |
                  |
                  |                     +
                  |
```

The FREQ Procedure

viral	Frequency	Percent	Cumulative Frequency	Cumulative Percent
viral	14	45.16	14	45.16
viral creator	5	16.13	19	61.29
viral sensati	12	38.71	31	100.00

influencer	Frequency	Percent	Cumulative Frequency	Cumulative Percent
Macro influencer	5	16.13	5	16.13
Mid-tier influencer	14	45.16	19	61.29
Top influencer	12	38.71	31	100.00

niche	Frequency	Percent	Cumulative Frequency	Cumulative Percent
DIY	1	3.23	1	3.23
dance	4	12.90	5	16.13
fashion	9	29.03	14	45.16
film	1	3.23	15	48.39
joke	13	41.94	28	90.32
music	3	9.68	31	100.00

The FREQ Procedure

Frequency Expected Deviation	Table of influencer by niche						
		niche					
influencer	DIY	dance	fashion	film	joke	music	Total
Macro influencer	0 0.1613 -0.161	3 0.6452 2.3548	0 1.4516 -1.452	0 0.1613 -0.161	2 2.0968 -0.097	0 0.4839 -0.484	5
Mid-tier influencer	1 0.4516 0.5484	0 1.8065 -1.806	4 4.0645 -0.065	1 0.4516 0.5484	5 5.871 -0.871	3 1.3548 1.6452	14
Top influencer	0 0.3871 -0.387	1 1.5484 -0.548	5 3.4839 1.5161	0 0.3871 -0.387	6 5.0323 0.9677	0 1.1613 -1.161	12
Total	1	4	9	1	13	3	31

Statistics for Table of influencer by niche

Statistic	DF	Value	Prob
Chi-Square	10	19.0995	0.0390
Likelihood Ratio Chi-Square	10	20.0969	0.0283
Mantel-Haenszel Chi-Square	1	0.3433	0.5579
Phi Coefficient		0.7849	
Contingency Coefficient		0.6174	
Cramer's V		0.5550	
WARNING: 89% of the cells have expected counts less than 5. Chi-Square may not be a valid test.			

Sample Size = 31

The FREQ Procedure

Frequency Expected Deviation	Table of viral by niche						
		niche					
viral	DIY	dance	fashion	film	joke	music	Total
viral	0 0.4516 -0.452	3 1.8065 1.1935	3 4.0645 -1.065	1 0.4516 0.5484	5 5.871 -0.871	2 1.3548 0.6452	14
viral creator	0 0.1613 -0.161	0 0.6452 -0.645	1 1.4516 -0.452	0 0.1613 -0.161	3 2.0968 0.9032	1 0.4839 0.5161	5
viral sensati	1 0.3871 0.6129	1 1.5484 -0.548	5 3.4839 1.5161	0 0.3871 -0.387	5 5.0323 -0.032	0 1.1613 -1.161	12
Total	1	4	9	1	13	3	31

Statistics for Table of viral by niche

Statistic	DF	Value	Prob
Chi-Square	10	8.0422	0.6247
Likelihood Ratio Chi-Square	10	10.1915	0.4239
Mantel-Haenszel Chi-Square	1	0.5044	0.4776
Phi Coefficient		0.5093	
Contingency Coefficient		0.4539	
Cramer's V		0.3602	
WARNING: 89% of the cells have expected counts less than 5. Chi-Square may not be a valid test.			

Sample Size = 31

The GLM Procedure

Class Level Information		
Class	Levels	Values
influencer	3	Macro influencer Mid-tier influencer Top influencer

Number of Observations Read	31
Number of Observations Used	31

The GLM Procedure

Dependent Variable: engagement_rate

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	2	0.05639252	0.02819626	4.91	0.0148
Error	28	0.16073936	0.00574069		
Corrected Total	30	0.21713188			

R-Square	Coeff Var	Root MSE	engagement_rate Mean
0.259716	48.77090	0.075767	0.155354

Source	DF	Type I SS	Mean Square	F Value	Pr > F
influencer	2	0.05639252	0.02819626	4.91	0.0148

Source	DF	Type III SS	Mean Square	F Value	Pr > F
influencer	2	0.05639252	0.02819626	4.91	0.0148

The ANOVA Procedure

Class Level Information		
Class	Levels	Values
influencer	3	Macro influencer Mid-tier influencer Top influencer

Number of Observations Read	31
Number of Observations Used	31

The ANOVA Procedure

Dependent Variable: engagement_rate

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	2	0.05639252	0.02819626	4.91	0.0148
Error	28	0.16073936	0.00574069		
Corrected Total	30	0.21713188			

R-Square	Coeff Var	Root MSE	engagement_rate Mean
0.259716	48.77090	0.075767	0.155354

Source	DF	Anova SS	Mean Square	F Value	Pr > F
influencer	2	0.05639252	0.02819626	4.91	0.0148

The ANOVA Procedure

The ANOVA Procedure

Tukey's Studentized Range (HSD) Test for engagement_rate

Note: This test controls the Type I experimentwise error rate.

Alpha	0.05
Error Degrees of Freedom	28
Error Mean Square	0.005741
Critical Value of Studentized Range	3.49918

Comparisons significant at the 0.05 level are indicated by ***.				
influencer Comparison	Difference Between Means	Simultaneous 95% Confidence Limits		
Macro influencer - Mid-tier influencer	0.09112	-0.00655	0.18879	

influencer Comparison	Difference Between Means	Simultaneous 95% Confidence Limits	
Macro influencer - Top influencer	0.12637	0.02658	0.22616	***
Mid-tier influencer - Macro influencer	-0.09112	-0.18879	0.00655	
Mid-tier influencer - Top influencer	0.03525	-0.03850	0.10900	
Top influencer - Macro influencer	-0.12637	-0.22616	-0.02658	***
Top influencer - Mid-tier influencer	-0.03525	-0.10900	0.03850	

Comparisons significant at the 0.05 level are indicated by ***.

The ANOVA Procedure

Class Level Information

Class	Levels	Values
niche	6	DIY dance fashion film joke music

Number of Observations Read	31
Number of Observations Used	31

The ANOVA Procedure

Dependent Variable: views_avg_M

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	5	48.7814916	9.7562983	0.41	0.8366
Error	25	593.6178632	23.7447145		
Corrected Total	30	642.3993548			

R-Square	Coeff Var	Root MSE	views_avg_M Mean
0.075936	68.88213	4.872855	7.074194

Source	DF	Anova SS	Mean Square	F Value	Pr > F
niche	5	48.78149159	9.75629832	0.41	0.8366

Distribution of views_avg_M
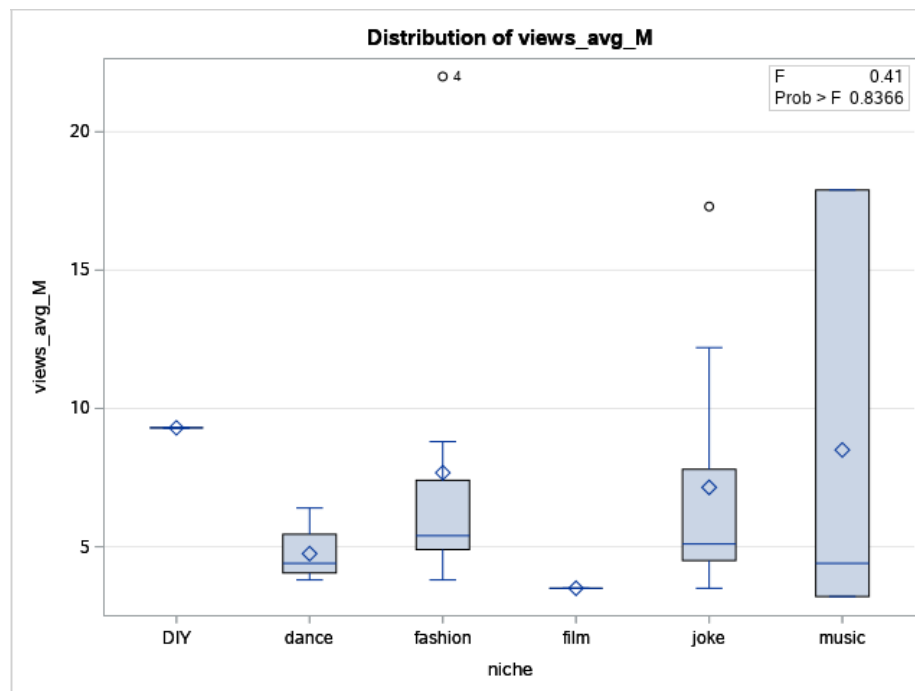
The ANOVA Procedure
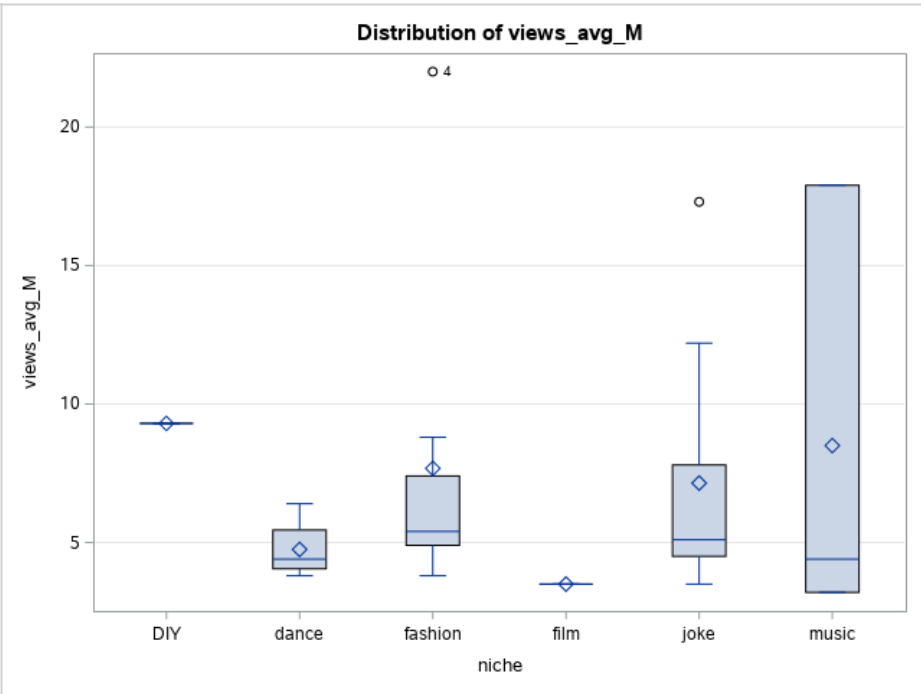
The ANOVA Procedure

Tukey's Studentized Range (HSD) Test for views_avg_M

Note: This test controls the Type I experimentwise error rate.

Alpha	0.05
Error Degrees of Freedom	25
Error Mean Square	23.74471
Critical Value of Studentized Range	4.35830

niche Comparison	Difference Between Means	Simultaneous 95% Confidence Limits	
DIY - music	0.800	-16.540	18.140
DIY - fashion	1.622	-14.207	17.452
DIY - joke	2.154	-13.430	17.738
DIY - dance	4.550	-12.240	21.340
DIY - film	5.800	-15.437	27.037
music - DIY	-0.800	-18.140	16.540
music - fashion	0.822	-9.189	10.834
music - joke	1.354	-8.265	10.972
music - dance	3.750	-7.719	15.219
music - film	5.000	-12.340	22.340
fashion - DIY	-1.622	-17.452	14.207
fashion - music	-0.822	-10.834	9.189
fashion - joke	0.532	-5.980	7.043
fashion - dance	2.928	-6.096	11.952
fashion - film	4.178	-11.652	20.007
joke - DIY	-2.154	-17.738	13.430
joke - music	-1.354	-10.972	8.265
joke - fashion	-0.532	-7.043	5.980
joke - dance	2.396	-6.190	10.983
joke - film	3.646	-11.938	19.230
dance - DIY	-4.550	-21.340	12.240
dance - music	-3.750	-15.219	7.719
dance - fashion	-2.928	-11.952	6.096
dance - joke	-2.396	-10.983	6.190
dance - film	1.250	-15.540	18.040
film - DIY	-5.800	-27.037	15.437
film - music	-5.000	-22.340	12.340
film - fashion	-4.178	-20.007	11.652
film - joke	-3.646	-19.230	11.938
film - dance	-1.250	-18.040	15.540

Comparisons significant at the 0.05 level are indicated by *.**

The ANOVA Procedure

Class Level Information

Class	Levels	Values
niche	6	DIY dance fashion film joke music

Number of Observations Read	31
Number of Observations Used	31

The ANOVA Procedure

Dependent Variable: subscribers_M

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	5	4127.96596	825.59319	0.71	0.6190
Error	25	28922.15257	1156.88610		
Corrected Total	30	33050.11853			

R-Square	Coeff Var	Root MSE	subscribers_M Mean
0.124900	138.3789	34.01303	24.57963

Source	DF	Anova SS	Mean Square	F Value	Pr > F
niche	5	4127.965963	825.593193	0.71	0.6190
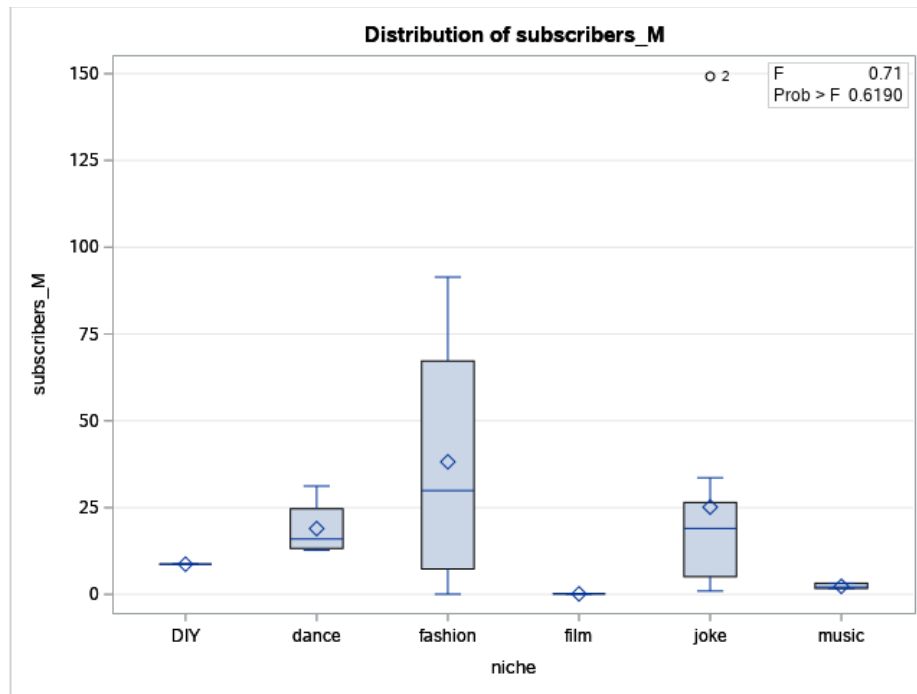
The ANOVA Procedure
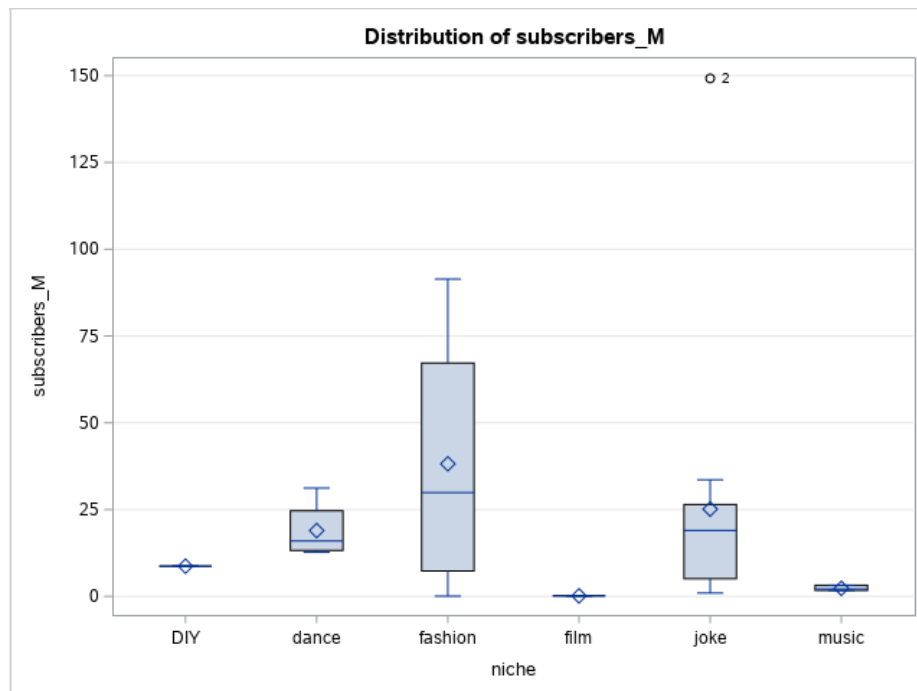
The ANOVA Procedure

Tukey's Studentized Range (HSD) Test for subscribers_M

Note: This test controls the Type I experimentwise error rate.

Alpha	0.05
Error Degrees of Freedom	25
Error Mean Square	1156.886
Critical Value of Studentized Range	4.35830

Comparisons significant at the 0.05 level are indicated by ***.				
niche Comparison	Difference Between Means	Simultaneous 95% Confidence Limits		
fashion - joke	13.04	-32.41	58.49	

niche Comparison	Difference Between Means	Simultaneous 95% Confidence Limits	
Comparisons significant at the 0.05 level are indicated by *.**			
fashion - dance	19.22	-43.77	82.21
fashion - DIY	29.47	-81.02	139.96
fashion - music	35.84	-34.04	105.72
fashion - film	38.03	-72.46	148.52
joke - fashion	-13.04	-58.49	32.41
joke - dance	6.19	-53.75	66.12
joke - DIY	16.44	-92.34	125.21
joke - music	22.80	-44.34	89.94
joke - film	24.99	-83.79	133.77
dance - fashion	-19.22	-82.21	43.77
dance - joke	-6.19	-66.12	53.75
dance - DIY	10.25	-106.94	127.44
dance - music	16.62	-63.44	96.67
dance - film	18.81	-98.39	136.00
DIY - fashion	-29.47	-139.96	81.02
DIY - joke	-16.44	-125.21	92.34
DIY - dance	-10.25	-127.44	106.94
DIY - music	6.37	-114.67	127.40
DIY - film	8.56	-139.68	156.80
music - fashion	-35.84	-105.72	34.04
music - joke	-22.80	-89.94	44.34
music - dance	-16.62	-96.67	63.44
music - DIY	-6.37	-127.40	114.67
music - film	2.19	-118.85	123.23
film - fashion	-38.03	-148.52	72.46
film - joke	-24.99	-133.77	83.79
film - dance	-18.81	-136.00	98.39
film - DIY	-8.56	-156.80	139.68
film - music	-2.19	-123.23	118.85

The ANOVA Procedure

Class Level Information

Class	Levels	Values
niche	6	DIY dance fashion film joke music

Number of Observations Read	31
Number of Observations Used	31

The ANOVA Procedure

Dependent Variable: engagement_rate

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	5	0.08413038	0.01682608	3.16	0.0239
Error	25	0.13300151	0.00532006		
Corrected Total	30	0.21713188			

R-Square	Coeff Var	Root MSE	engagement_rate Mean
0.387462	46.95014	0.072939	0.155354

Source	DF	Anova SS	Mean Square	F Value	Pr > F
niche	5	0.08413038	0.01682608	3.16	0.0239
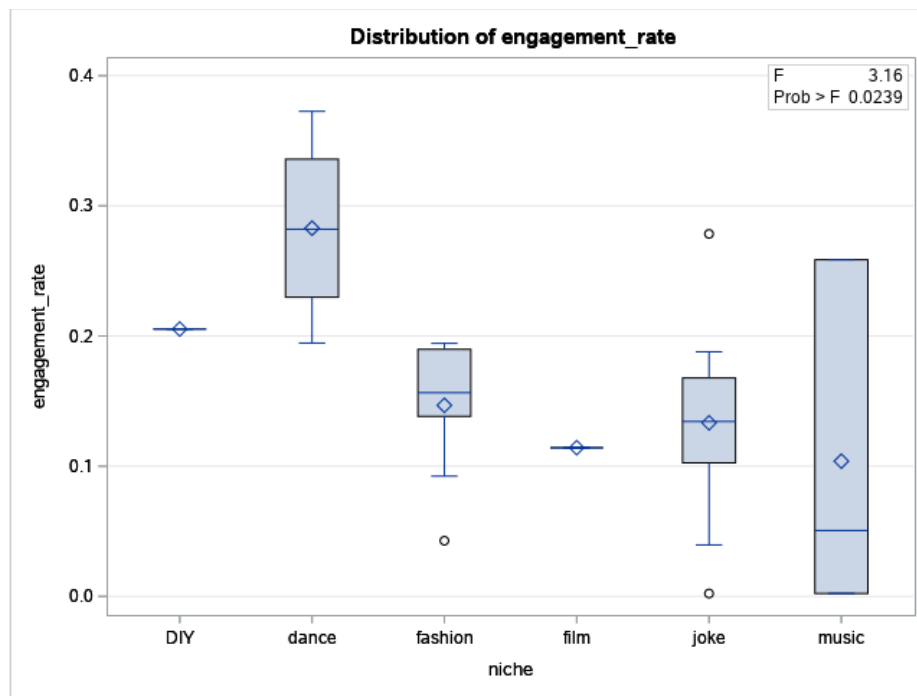
The ANOVA Procedure
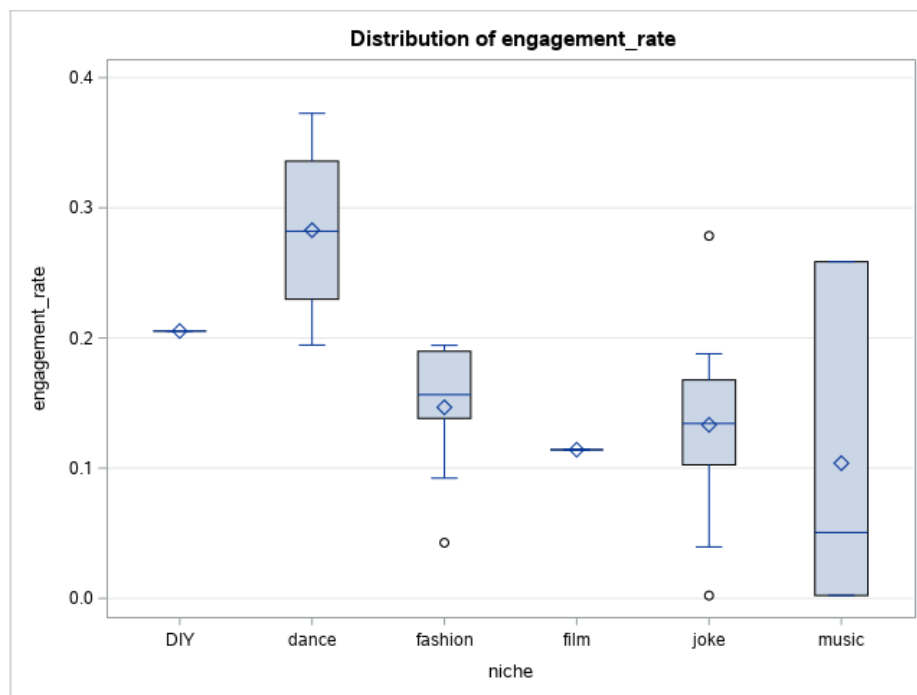
Tukey's Studentized Range (HSD) Test for engagement_rate

Note: This test controls the Type I experimentwise error rate.

Alpha	0.05
Error Degrees of Freedom	25
Error Mean Square	0.00532
Critical Value of Studentized Range	4.35830

Comparisons significant at the 0.05 level are indicated by ***.				
niche Comparison	Difference Between Means	Simultaneous 95% Confidence Limits		
dance - DIY	0.07747	-0.17384	0.32878	

niche Comparison	Difference Between Means	Simultaneous 95% Confidence Limits		
dance - fashion	0.13606	0.00098	0.27114	***
dance - joke	0.14951	0.02099	0.27803	***
dance - film	0.16864	-0.08267	0.41996	
dance - music	0.17898	0.00731	0.35066	***
DIY - dance	-0.07747	-0.32878	0.17384	
DIY - fashion	0.05859	-0.17835	0.29553	
DIY - joke	0.07204	-0.16123	0.30531	
DIY - film	0.09117	-0.22672	0.40906	
DIY - music	0.10152	-0.15804	0.36107	
fashion - dance	-0.13606	-0.27114	-0.00098	***
fashion - DIY	-0.05859	-0.29553	0.17835	
fashion - joke	0.01345	-0.08402	0.11092	
fashion - film	0.03258	-0.20436	0.26952	
fashion - music	0.04292	-0.10693	0.19278	
joke - dance	-0.14951	-0.27803	-0.02099	***
joke - DIY	-0.07204	-0.30531	0.16123	
joke - fashion	-0.01345	-0.11092	0.08402	
joke - film	0.01913	-0.21414	0.25240	
joke - music	0.02947	-0.11450	0.17345	
film - dance	-0.16864	-0.41996	0.08267	
film - DIY	-0.09117	-0.40906	0.22672	
film - fashion	-0.03258	-0.26952	0.20436	
film - joke	-0.01913	-0.25240	0.21414	
film - music	0.01034	-0.24921	0.26990	
music - dance	-0.17898	-0.35066	-0.00731	***
music - DIY	-0.10152	-0.36107	0.15804	
music - fashion	-0.04292	-0.19278	0.10693	
music - joke	-0.02947	-0.17345	0.11450	
music - film	-0.01034	-0.26990	0.24921	

Comparisons significant at the 0.05 level are indicated by ***.

The ANOVA Procedure

Class Level Information

Class	Levels	Values
influencer	3	Macro influencer Mid-tier influencer Top influencer

Number of Observations Read	31
Number of Observations Used	31

The ANOVA Procedure

Dependent Variable: watch_rate

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	2	717.925453	358.962726	1.58	0.2238
Error	28	6361.967522	227.213126		
Corrected Total	30	7079.892975			

R-Square	Coeff Var	Root MSE	watch_rate Mean
0.101403	327.3409	15.07359	4.604861

Source	DF	Anova SS	Mean Square	F Value	Pr > F
influencer	2	717.9254528	358.9627264	1.58	0.2238
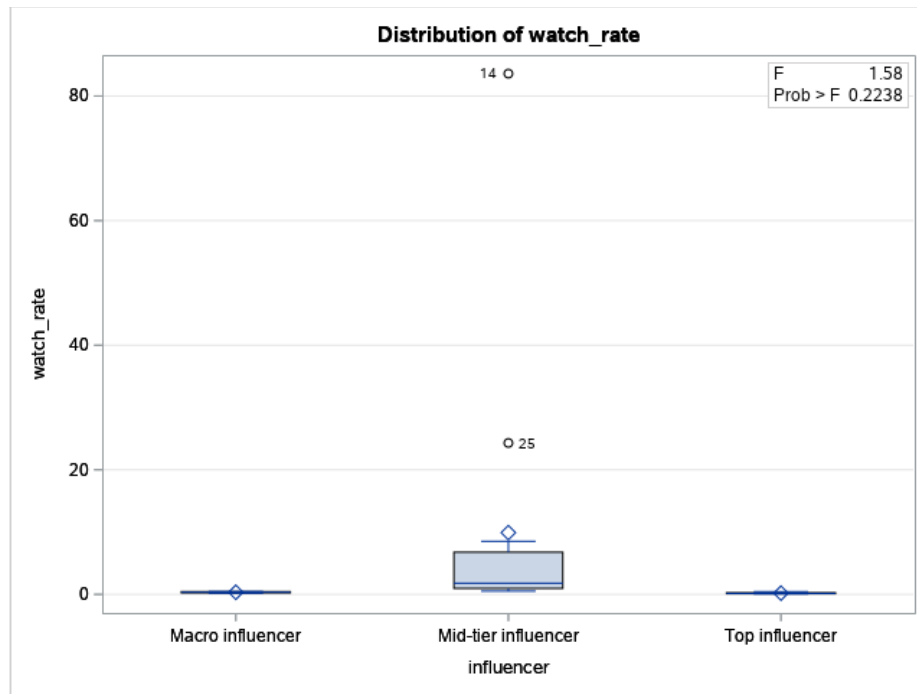
The ANOVA Procedure
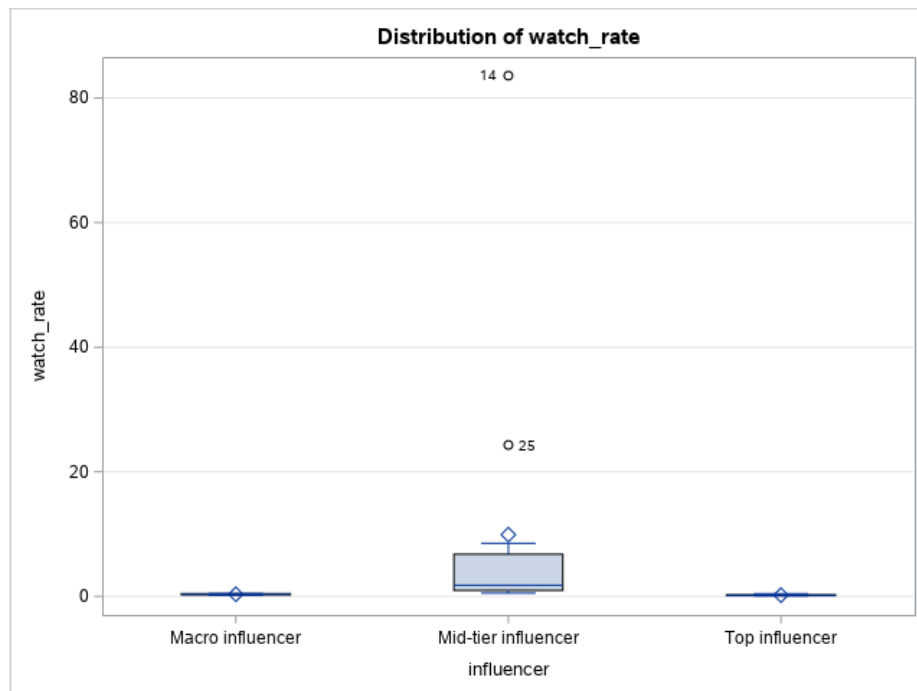
The ANOVA Procedure

Tukey's Studentized Range (HSD) Test for watch_rate

Note: This test controls the Type I experimentwise error rate.

Alpha	0.05
Error Degrees of Freedom	28
Error Mean Square	227.2131
Critical Value of Studentized Range	3.49918

Comparisons significant at the 0.05 level are indicated by ***.				
influencer Comparison	Difference Between Means	Simultaneous 95% Confidence Limits		
Mid-tier influencer - Macro influencer	9.568	-9.863	28.999	

Comparisons significant at the 0.05 level are indicated by ***.

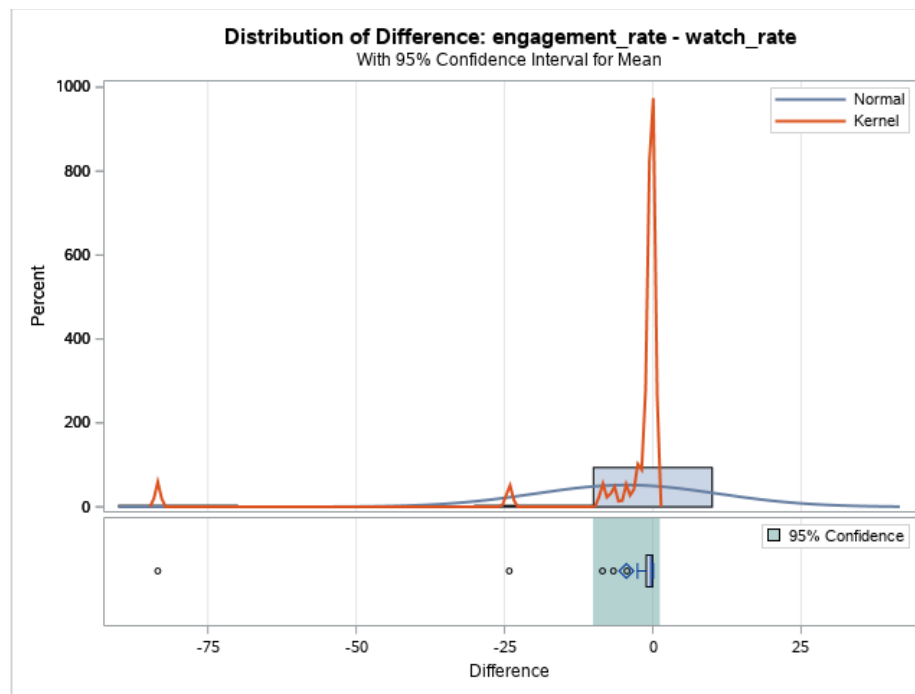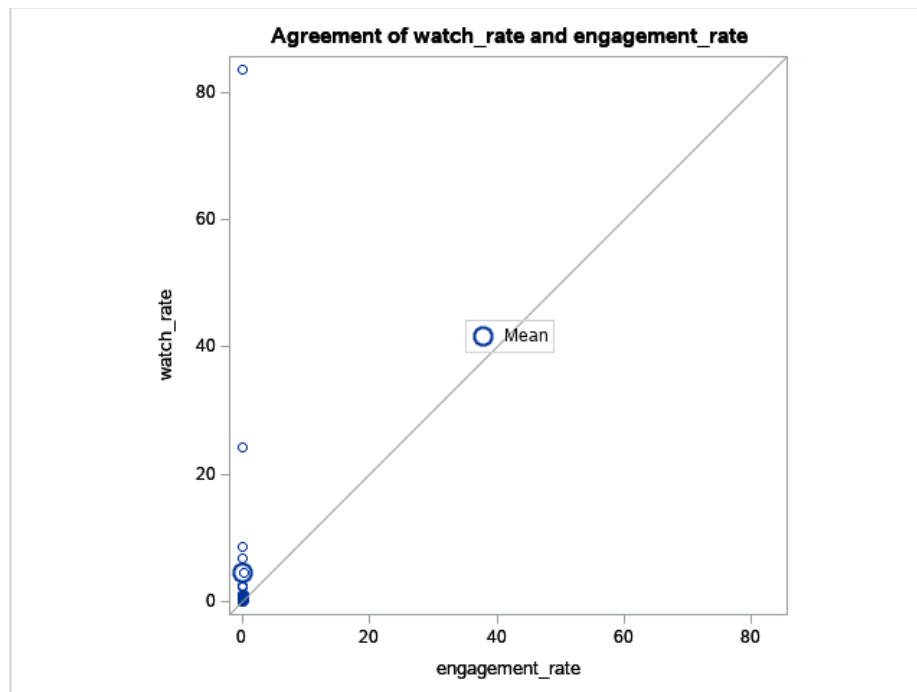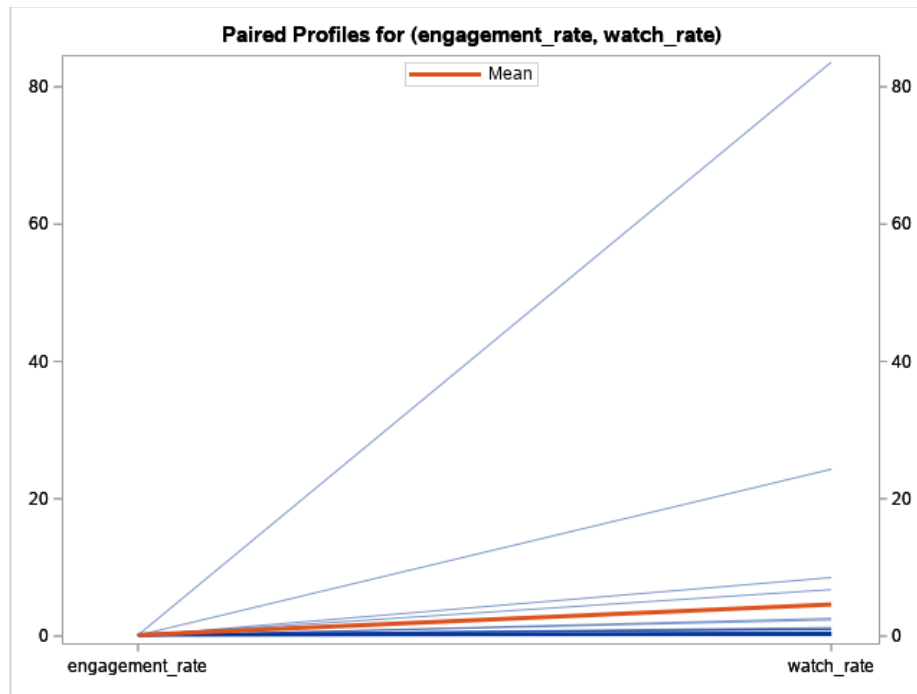| influencer Comparison	Difference Between Means	Simultaneous 95% Confidence Limits	
Mid-tier influencer - Top influencer	9.712	-4.961	24.384
Macro influencer - Mid-tier influencer	-9.568	-28.999	9.863
Macro influencer - Top influencer	0.144	-19.709	19.996
Top influencer - Mid-tier influencer	-9.712	-24.384	4.961
Top influencer - Macro influencer	-0.144	-19.996	19.709

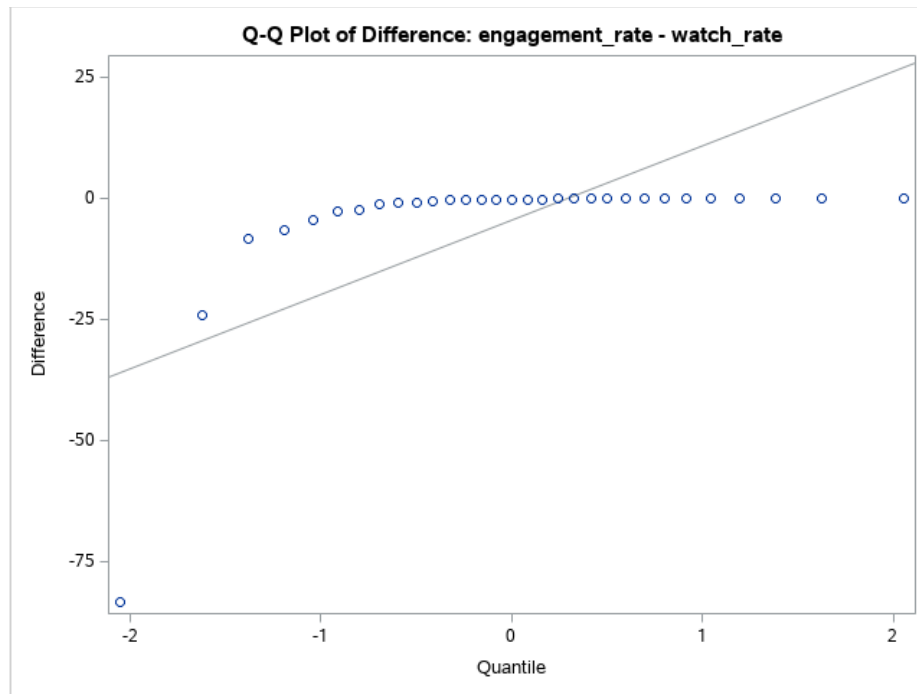**The TTEST Procedure**

Difference: engagement_rate - watch_rate

N	Mean	Std Dev	Std Err	Minimum	Maximum
31	-4.4495	15.3672	2.7600	-83.4408	0.0892

Mean	95% CL Mean		Std Dev	95% CL Std Dev	
-4.4495	-10.0863	1.1872	15.3672	12.2801	20.5410

DF	t Value	Pr > |t|
30	-1.61	0.1174

Distribution of Difference: engagement_rate - watch_rate
With 95% Confidence Interval for Mean

Paired Profiles for (engagement_rate, watch_rate)

Agreement of watch_rate and engagement_rate

Q-Q Plot of Difference: engagement_rate - watch_rate

The CORR Procedure

2 Variables:	subscribers_M engagement_rate

Simple Statistics

Variable	N	Mean	Std Dev	Sum	Minimum	Maximum
subscribers_M	31	24.57963	33.19142	761.96860	0.06460	149.20000
engagement_rate	31	0.15535	0.08507	4.81596	0.00218	0.37256

Pearson Correlation Coefficients, N = 31
Prob > |r| under H0: Rho=0

	subscribers_M	engagement_rate
subscribers_M	1.00000	-0.22332 0.2272
engagement_rate	-0.22332 0.2272	1.00000

The CORR Procedure
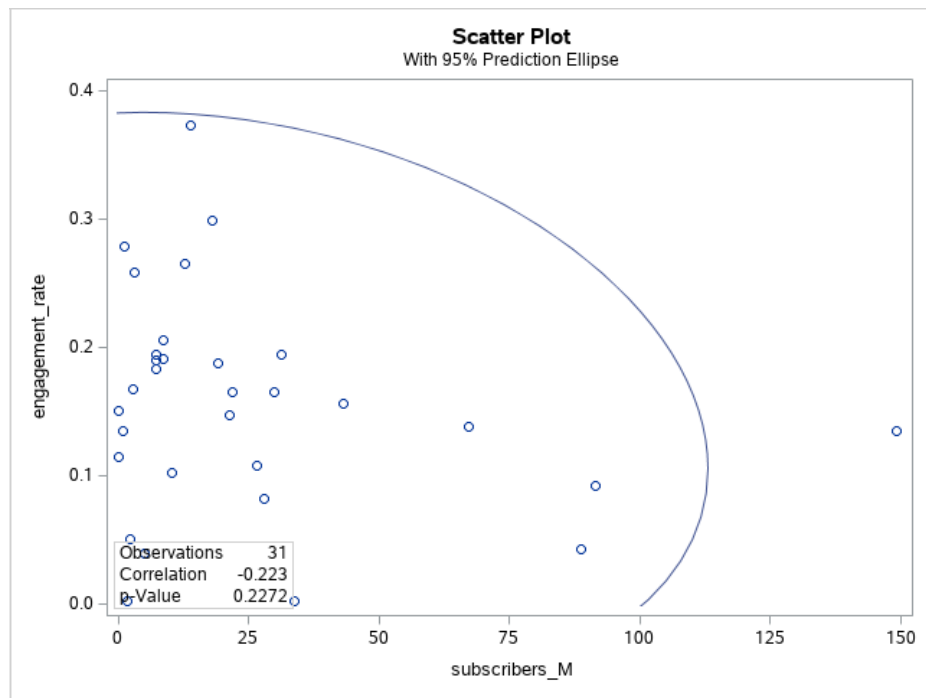
The CORR Procedure

2 Variables:	subscribers_M watch_rate

Simple Statistics

Variable	N	Mean	Std Dev	Sum	Minimum	Maximum
subscribers_M	31	24.57963	33.19142	761.96860	0.06460	149.20000
watch_rate	31	4.60486	15.36218	142.75068	0.08096	83.59133

Pearson Correlation Coefficients, N = 31
Prob > |r| under H0: Rho=0

	subscribers_M	watch_rate
subscribers_M	1.00000	-0.21504 0.2453
watch_rate	-0.21504 0.2453	1.00000

The CORR Procedure

Scatter Plot
With 95% Prediction Ellipse

Observations	31
Correlation	-0.215
P-value	0.2453
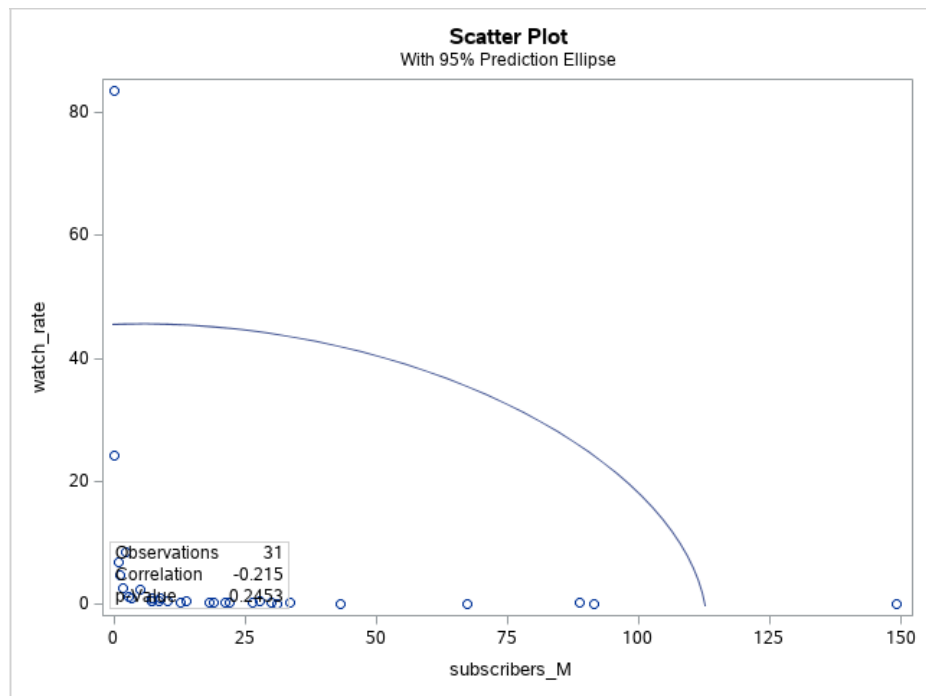
The CORR Procedure

2 Variables:	subscribers_M views_avg_M

Simple Statistics						
Variable	N	Mean	Std Dev	Sum	Minimum	Maximum
subscribers_M	31	24.57963	33.19142	761.96860	0.06460	149.20000
views_avg_M	31	7.07419	4.62745	219.30000	3.20000	22.00000

Pearson Correlation Coefficients, N = 31 Prob > |r| under H0: Rho=0		
	subscribers_M	views_avg_M
subscribers_M	1.00000	0.55752 0.0011
views_avg_M	0.55752 0.0011	1.00000

The CORR Procedure

Scatter Plot
With 95% Prediction Ellipse

Observations	31
Correlation	0.5575
p-Value	0.0011
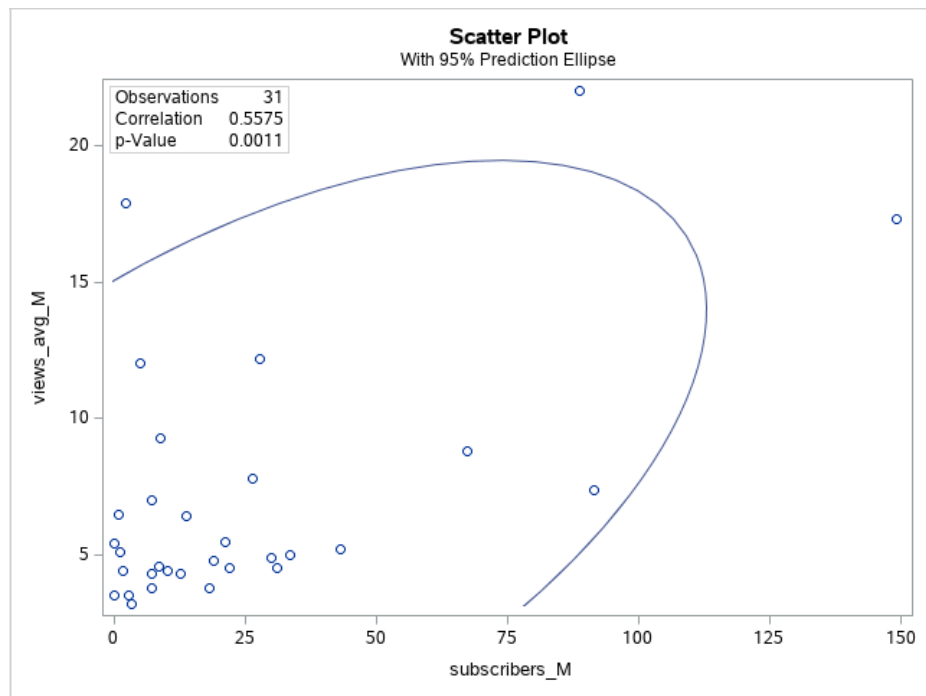
The REG Procedure
Model: MODEL1
Dependent Variable: subscribers_M

Number of Observations Read	31
Number of Observations Used	31

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	1	1648.31180	1648.31180	1.52	0.2272
Error	29	31402	1082.82092		
Corrected Total	30	33050			

Root MSE	32.90624	R-Square	0.0499
Dependent Mean	24.57963	Adj R-Sq	0.0171
Coeff Var	133.87606		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	38.11529	12.46146	3.06	0.0047
engagement_rate	1	-87.12803	70.61818	-1.23	0.2272

The REG Procedure
Model: MODEL1
Dependent Variable: subscribers_M

Fit Diagnostics for subscribers_M

Observations	31
Parameters	2
Error DF	29
MSE	1082.8
R-Square	0.0499
Adj R-Square	0.0171

Residuals for subscribers_M
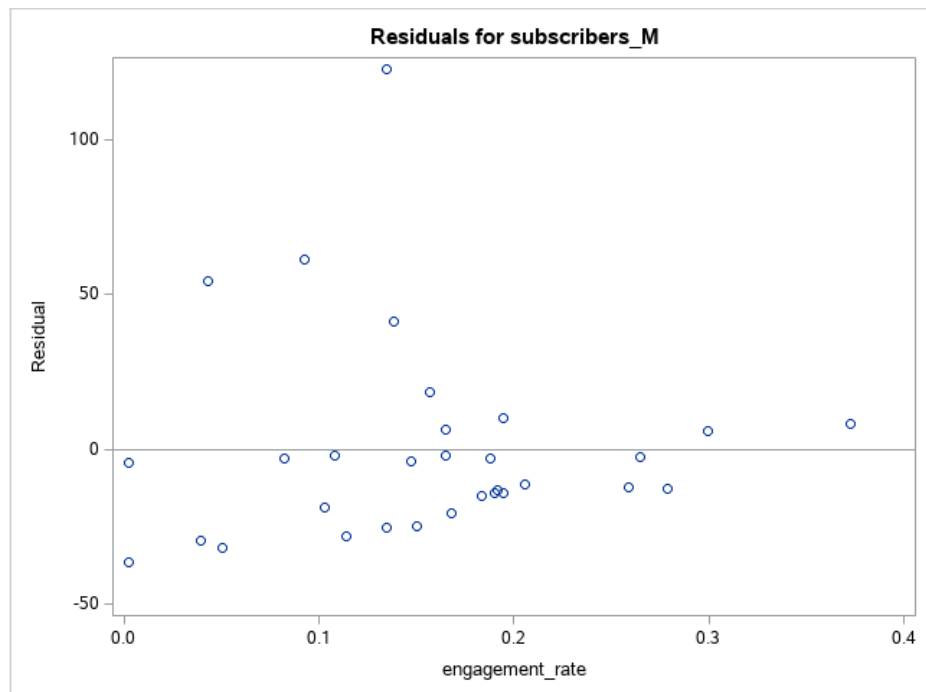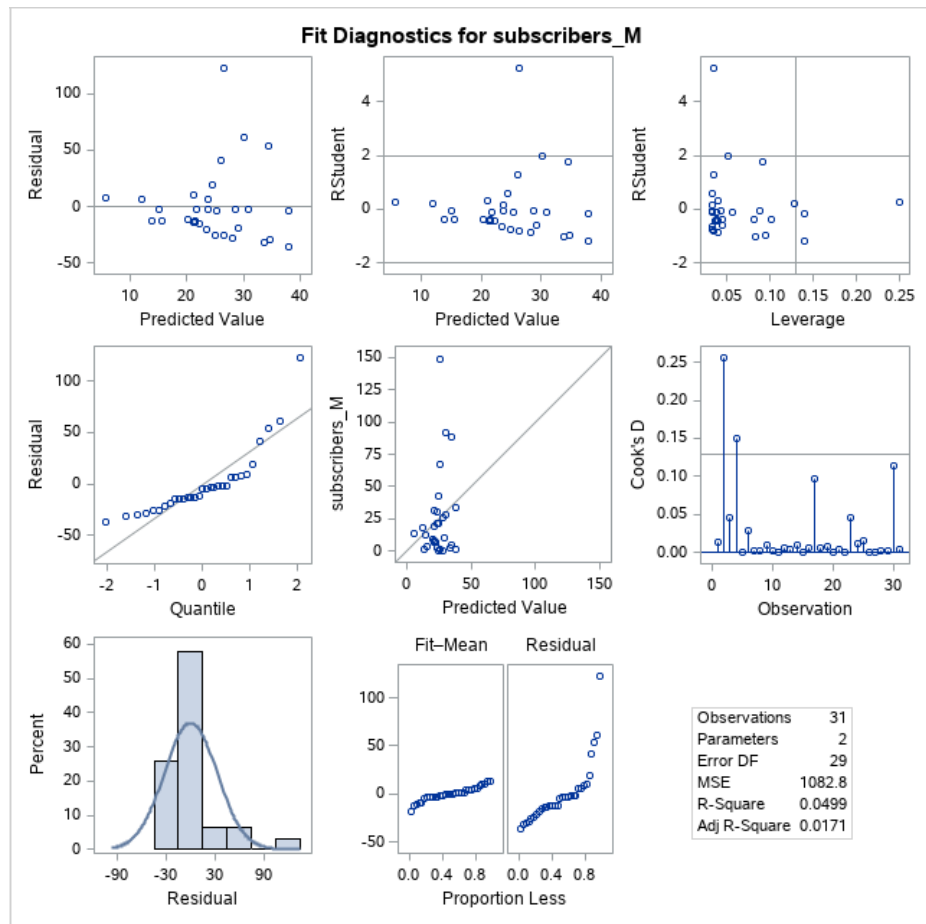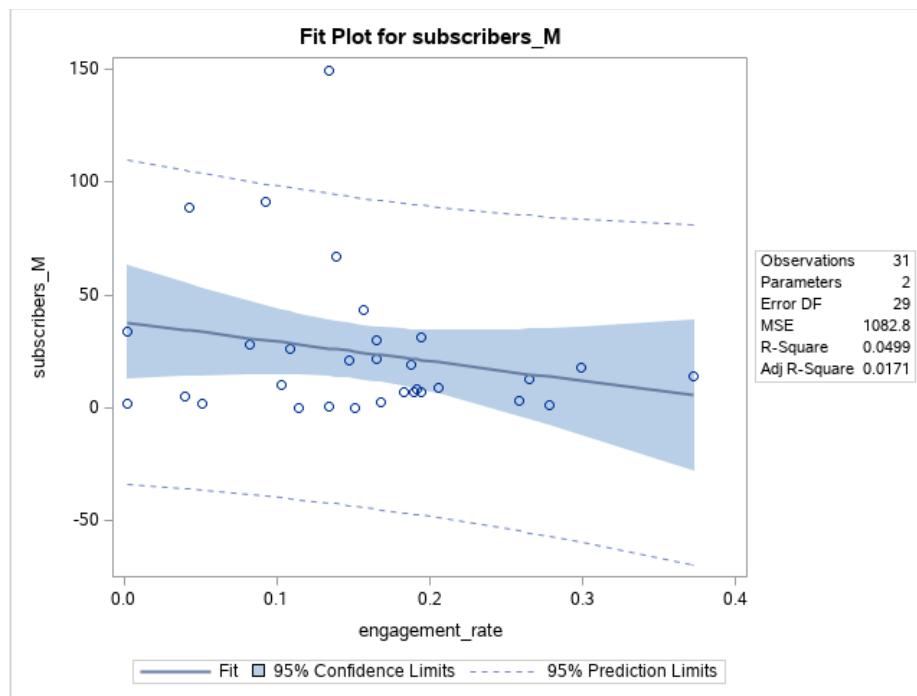
The GLM Procedure

Number of Observations Read	31
Number of Observations Used	31

The GLM Procedure

Dependent Variable: subscribers_M

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	1	121.70176	121.70176	0.11	0.7457
Error	29	32928.41677	1135.46265		
Corrected Total	30	33050.11853			

R-Square	Coeff Var	Root MSE	subscribers_M Mean
0.003682	137.0917	33.69663	24.57963

Source	DF	Type I SS	Mean Square	F Value	Pr > F
comments_avg_K	1	121.7017614	121.7017614	0.11	0.7457

Source	DF	Type III SS	Mean Square	F Value	Pr > F
comments_avg_K	1	121.7017614	121.7017614	0.11	0.7457

Parameter	Estimate	Standard Error	t Value	Pr > |t|
Intercept	26.22316718	7.86319024	3.33	0.0023
comments_avg_K	-0.15364772	0.46931464	-0.33	0.7457