

The Social Media Statistical software



West Bengal State University

Part III Examination 2019

Paper- VIII

Group B

PROJECT

Roll : 3191307 No. : 24632

Registration No. 1081511400222 of 2015

1. Summary

In the digital age we are living in, our social media presence is an identity proof itself. Not just an id proof but a means of character verification as well. From friends and relatives to interviewers at jobs, everyone gets up to date about the whereabouts of our life from our social media handles. But our social media presence is not limited to being our 'personal information bulletin'.

Even though the mental health community warns about the ill-effects of it, we just can't seem to stop scrolling down the screen of our mobile phones and personal computers. But given this situation, it doesn't mean there aren't real monetary benefits of being and maintaining **influencer-level status** on instagram, facebook, twitter, snapchat, youtube and other social media apps.

This software helps in achieving that purpose. The software uses data from the user's past posts to analyse and find out which genre of posts has the highest correlation to the number of followers the user has had simultaneously till now. This enables the user to gain better knowledge about what topics to post about to garner more followers and ultimately become social media influencers and become rich.

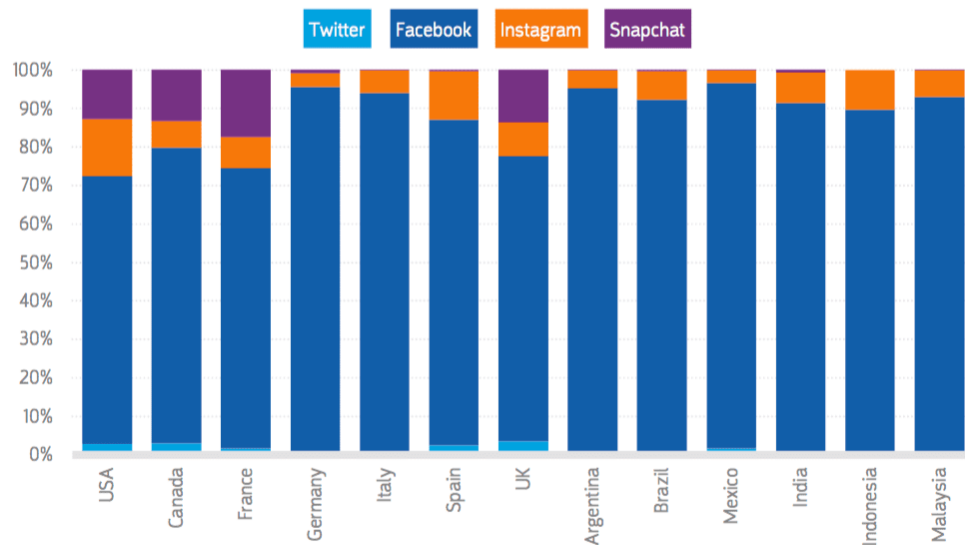
2. Introduction

It is the era when your talent need not have to be accompanied by contacts and opportunities and a great deal of luck for success. A camera and a good internet connection may just be enough. Some of the highest earning social media stars make upwards **6 million dollars** annually just by posting content online.

This is possible because of the paradigm shift of what defines '**leisure time**'. Earlier leisure time used to be occupied by the television or activities such as reading, playing games (indoor and outdoor), group discussions, etc. But after the advent of social media, the whole scenario has changed. It can be said with a certain level of confidence that now, the amount of time an average youth spends doing all of the aforementioned activities in totality is less than the time he/she spends being active in social media.

Social Networks' Share of Time

Share of Combined Minutes for 4 Major Social Networking Apps



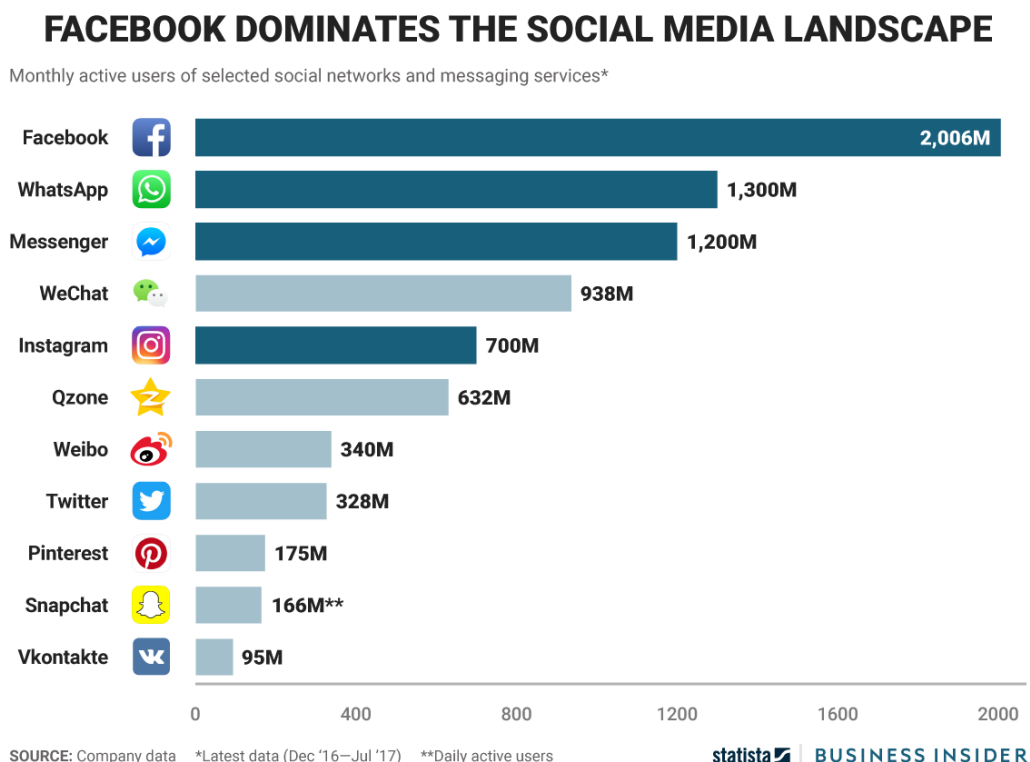
Hence, the main source of revenue in the entertainment industry which used to be advertisements through television has changed to advertisements through social media platforms. These outlets of the social media platforms are none other than users themselves, the main activities of whom, namely upvoting/liking, posting, sharing, commenting and following, result in the spreading of information among the masses. Users post a wide variety of posts such as humorous, emotional, informative, factual, scary, opinionated, entertaining, criticising, etc.

In such a time it is inevitable that some people will come out as more popular than others and their popularity will increase exponentially with time if they are able to manage their online content properly. Businesses and manufacturers target such persons with high social media influence, to advertise their products and in exchange pay them a lot to do so. Also social media platforms play or post advertisements alongside the posts of these social media influencers and pay them according to the number of followers they have.

The purpose of this **Social Media Statistical Software** is to let users know which genre of their posts, is helping them gain popularity in the social media world at any point in time, which in turn helps them establish themselves as influencer level personalities in the social media world.

3. Application

Consistency is a key to success and this is no easy get-rich-quick scheme that is sure to work. There are many struggle influencers trying to make a name for themselves online. It is obviously difficult to get attention from millions of users around the world and seldom does it happen overnight.



Thus in order to cater to millions of people around the globe one needs to understand the needs or wants of these people and also needs to know which scenarios are most rewarding in nature.

OBJECTIVE:

Some of the highest ranked social media influencers around the world are:

1. Kayla Itsines: Fitness mogul.
Estimated income- \$17 million from personal training app.
\$150,000 from each instagram post.
Followers- 8.6 million.

2. Evan Fong (VanossGaming): Gaming vlogger.
Estimated income- \$15.5 million.
Followers- 22.2 million.
3. Lilly Singh: Youtube content creator(entertainment)

Estimated income- \$10.5 million

Followers- 13.1 million (YouTube)

7.2 million (Instagram)
4. Ryan ToysReview: 7 year old youtube toy critic.

Estimated income- \$22 million

Followers- 17 million

(Source: www.cbsnews.com, www.complex.com, www.google.com)

As one can see from the above mentioned examples, this profession has a huge potential. There are no restrictions on type of posts, location, religion, caste, creed, gender or even age of the social media influencers. Ryan is a 7 year old boy who reviews toys in his own way, which his parents record and upload on youtube and has earned more than \$22 million through it.

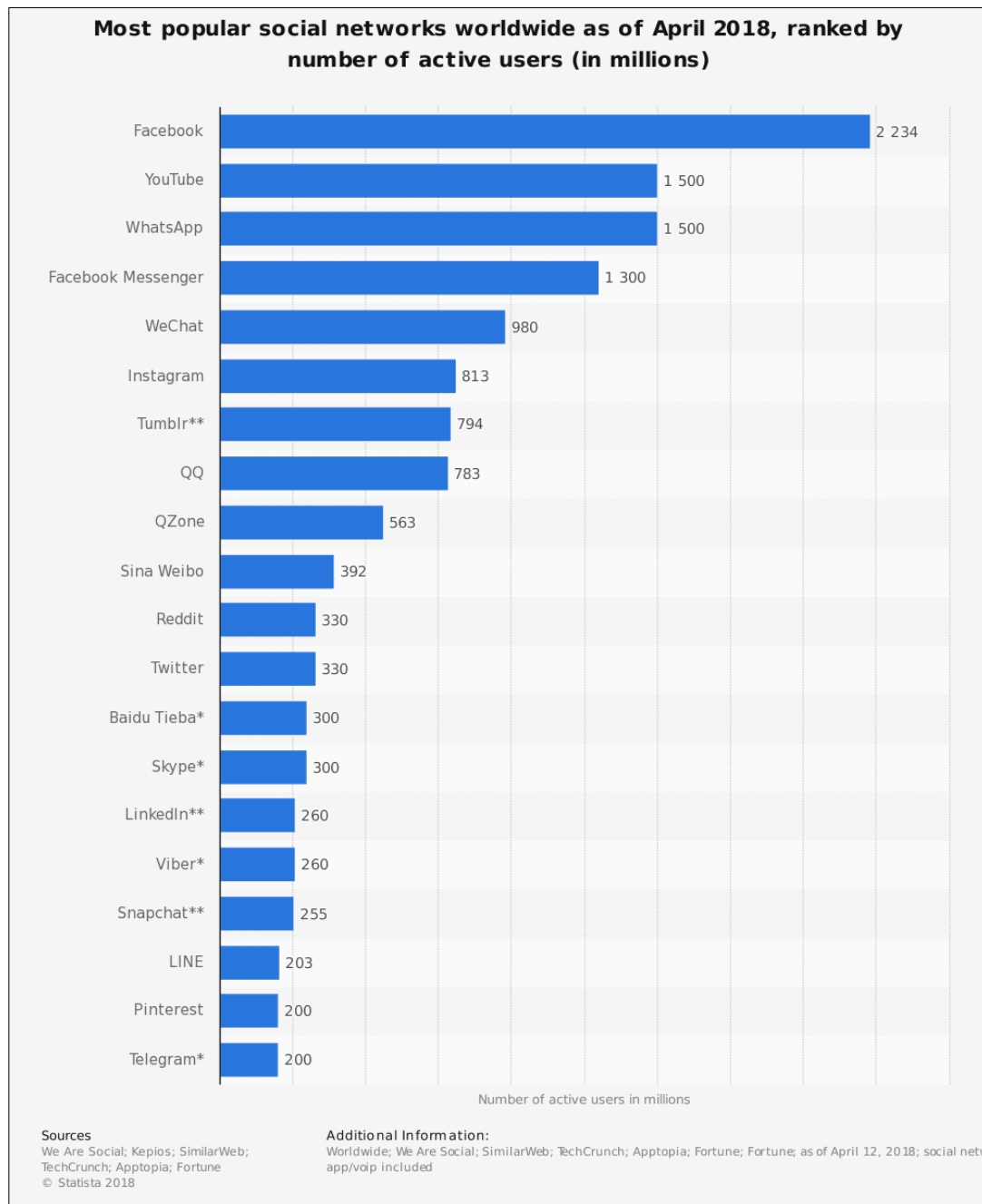
Thus one can see that with the right guidance and proper talent everyone gets an equal opportunity to achieve success.

The aim of this software is to let users know which genre of their post is the most influencing one and rank the genres in order of their capability to gain followers. Users input their data based on their previous social media posts and the software analyses the data to give in result the ranking of genres in order of their follower gaining capability. With proper and frequent use of this software, users can become popular social media influencers and highly paid content creators.

4. WORKING:

There are millions of active users around the world. The social media world is not restricted by geographical boundaries. In accumulation of all social media platforms there are thousands of millions of active users around the world.

Now, the main aim is to garner as many followers and subscribers as possible. Why?



This is because the main parts of online activity as mentioned before are- upvoting/liking, posting, sharing, commenting and following. Out of these, Businesses and Manufacturers look for the ones with the highest number of followers as it would imply that the more the number of followers one has, the more people, his/her posts would reach. Therefore more people would be able to see their products and services.

Disadvantages / Drawbacks:

1. Human kind is unpredictable and trends change very frequently. Hence the genre of post gathering many followers today might turn uninteresting tomorrow. Therefore the software needs to be used frequently.
2. The software is just an indicator of which genre is attracting most followers but it doesn't give a magnitude of how much. Hence it does not ensure a high number of followers upon usage.

5. DATA AND ANALYSIS:

Data:

Our data for analysis will be the total number of likes or upvotes the user received in each of the posts he/she is considering for analysis from every genre and the real time number of followers he/she had corresponding to them. This is because, the number of shares a user's post gets inadvertently results in greater number of likes for that post and comments are subjective in nature and are not a reliable source of statistical information/data.

The software will accept three broad classifications of the genres from the user. It will then accept the number of followers corresponding to the genre posts. The user may use information from as many posts as he/she wishes to use for the analysis and they need not be in chronological order. But the posts from various genres corresponding to the number of followers must be from the same point of time, i.e. a user will choose a point of time and then observe the number of likes he/she received from the various posts of each of the three genres and the total number of followers he/she had corresponding to that time period. These will be the data used in the software.

Analysis:

For our analysis, we shall first look into our aim. Our aim was to find the genre of posts which enables the user to garner the most number of followers. Thus we are looking for the genre of posts which has the highest correlation of the number of likes/upvotes to the number of followers one has.

Since, our aim is to find that particular genre whose increase in post will bring us more followers, we want to find the correlation between a genre and the number of followers, eliminating the effect of the other genres from it.

Thus we are interested in the Partial Correlation Coefficient of the genres to the number of followers.

We have used C Programming to create a software, that will take in as data the number of likes the user received from the various posts of each of three genres (chosen by the user) and the total number of followers the user had all of which are corresponding to a particular time period. The software will give in return as result the rank of genres in correspondence their partial correlation coefficients, i.e. it will tell us which genre of posts helps in attracting most number of followers.



6. Working example of the software

Let us take a real life scenario and use data to analyse and provide us with a result. Suppose we have the following data:

Genre classifications: Emotional, Entertainment, Informative.

Each column under the genres contain the number of likes/upvotes one received from the corresponding number of post and the column followers contains the number of followers the user had corresponding to that number of post (at the time of those posts).

Post	Emotional	Entertainment	Informative	Followers
1	50	40	33	15
2	45	43	37	15
3	57	48	49	18
4	63	52	45	19
5	79	65	57	25
6	82	63	69	26
7	90	73	65	26
8	95	78	60	28
9	90	80	85	30
10	100	85	70	35

We used the Social Media Statistical Software to analyse the data and we got the following result:

The name of your genres are:				
Emotional,	Entertainment,	Informative	and final column is the Number of followers	
Your data set is:				
50	40	33	15	
45	43	37	15	
57	48	49	18	
63	52	45	19	
79	65	57	25	
82	63	69	26	
90	73	65	26	
95	78	60	28	
90	80	85	30	
100	85	70	35	
Genre Entertainment brings you most followers, followed by genre Informative and then genre Emotional.				

Now in order to check the validity of the software, let us calculate the partial correlation coefficients by hand and see if we obtain the same result.

Let us assume our given data set as:

Post	S2	S3	S4	S1
1	50	40	33	15
2	45	43	37	15
3	57	48	49	18
4	63	52	45	19
5	79	65	57	25
6	82	63	69	26
7	90	73	65	26
8	95	78	60	28
9	90	80	85	30
10	100	85	70	35

Hence,

Mean of $S1 = s1 = 23.7$; Mean of $S2 = s2 = 75.1$; Mean of $S3 = s3 = 62.7$;
Mean of $S4 = s4 = 57$.

Now, for $j = 1, 2, \dots, 10$

$S2j-s2$	$S3j-s3$	$S4j-s4$	$S1j-s1$
-25.1	-22.7	-24	-8.7
-30.1	-19.7	-20	-8.7
-18.1	-14.7	-8	-5.7
-12.1	-10.7	-12	-4.7
3.9	2.3	0	1.3
6.9	0.3	12	2.3
14.9	10.3	8	2.3
19.9	15.3	3	4.3
14.9	17.3	28	6.3
24.9	22.3	13	11.3

$\text{Cov}(S1,S1) = 44.31$; $\text{Cov}(S2,S2) = 392.54$; $\text{Cov}(S3,S3) = 264.011$;
 $\text{Cov}(S4,S4) = 261.55$; $\text{Cov}(S1,S2) = 128.478$; $\text{Cov}(S1,S3) = 106.344$;
 $\text{Cov}(S1,S4) = 96.33$; $\text{Cov}(S2,S3) = 315.589$; $\text{Cov}(S2,S4) = 277.444$;
 $\text{Cov}(S3,S4) = 232.333$.

Hence, Dispersion matrix $\Sigma =$

44.31	128.478	106.344	96.333
128.478	392.54	315.589	277.444
106.344	315.589	264.011	232.333
96.333	277.444	232.333	261.55

Thus,

$$|\Sigma_{11}| = 230642.01664504 ; |\Sigma_{22}| = 20232.52675833 ; |\Sigma_{33}| = 46062.644014092$$

$$|\Sigma_{44}| = 5469.165562812 ; |\Sigma_{12}| = -20676.383826495 ; |\Sigma_{13}| = -58293.44191614$$

$$|\Sigma_{14}| = -11201.294633685$$

Therefore,

$$\rho_{12.34} = (-\Sigma_{12})/(\sqrt{\Sigma_{11}} * \sqrt{\Sigma_{22}}) = 0.302677641$$

$$\rho_{13.24} = (-\Sigma_{13})/(\sqrt{\Sigma_{11}} * \sqrt{\Sigma_{33}}) = 0.565556613$$

$$\rho_{14.23} = (-\Sigma_{14})/(\sqrt{\Sigma_{11}} * \sqrt{\Sigma_{44}}) = 0.315383162$$

Hence, partial correlation between S1 and S3 is highest followed by partial correlation between S1 and S4 and then S1 and S2.

7. Conclusion:

The result provided by the Social Media Statistical software is in accordance with our calculated result. S1 referred to the number of followers, S2 referred to the number of likes in the genre Emotional, S3 referred to the number of likes in the genre Entertainment and S4 referred to the number of likes in the genre Informative. Therefore even our theoretically calculated result shows that the genre Entertainment will bring the most number of followers, followed by the genre Informative and then, the genre Emotional.

Therefore the Software is authentic and provides us with correct results.

The software can be used by users to find out which genre of their posts brings them the highest number of followers. The software can be used at any point of time and it can analyse any amount of data input by the user. The more the data, better is the analysis which is true for any statistical analysis.

The users can then post more in that genre to garner more followers and possibly become social media influencers and earn a great deal of money while being very popular at the same time.

8. Appendix

The entire code of our Social Media Statistical Software is given as follows:

```
#include<stdio.h>
#include<stdlib.h>
#include<conio.h>
#include<math.h>
#include<unistd.h>
double covariance(int ds1[], int ds2[], int size)
{
    int x;
    double cov, xbar, ybar, temp=0, sx=0, sy=0, n;
```

```

    for(x=0;x<size;x++)
    {
        sx= sx+ ds1[x];
        sy= sy+ ds2[x];
    }
    n= size-1;
    xbar=(double)sx/size;
    ybar=(double)sy/size;
    for(x=0;x<size;x++)
    {
        temp= temp+((ds1[x]-xbar)*(ds2[x]-ybar));
    }
    cov=(double)temp/n;
    return(cov);
}
double detof(double dispmat[4][4],int row, int col)
{
    double detsigma[3][3], determinant;
    int i,j,m=0, n=0;
    for(i=0; i<4; i++)
    {
        for(j=0;j<4; j++)
        {
            if(i!=row && j!=col)
            {
                detsigma[m][n]=dispmat[i][j];
                n++;
            }
        }
        if(i!=row)
            m++; n=0;
    }
    determinant=(pow(-1,row+col))*
(detsigma[0][0]*(((detsigma[1][1])*(detsigma[2][2]))-
((detsigma[1][2])*(detsigma[2][1])))-
detsigma[0][1]*(((detsigma[1][0])*(detsigma[2][2]))-
((detsigma[2][0])*(detsigma[1][2])))+detsigma[0][2]*(((detsigma[1][0])*(detsigma[2][1]
)))-((detsigma[2][0])*(detsigma[1][1]))));

    return(determinant);
}

int main()

```

```

{
    double dispmat[4][4],sigma11,sigma22,sigma33,sigma44, sigma12, sigma13,
    sigma14;
    int np, l;
    char g1[30], g2[30], g3[30];
    double parcorr12, parcorr13, parcorr14;
    printf("\n\tEnter the name of the first genre of your posts \n");
    scanf("%s",g1);
    printf("\n\tEnter the name of the second genre of your posts \n");
    scanf("%s",g2);
    printf("\n\tEnter the name of the third genre of your posts \n");
    scanf("%s",g3);
    printf("\n\tEnter the number of posts in each genre \n");
    scanf("%d",&np);
    int dg1[np],dg2[np],dg3[np],df[np];
    for(l=0;l<np;l++)
    {
        printf("\n");
        printf("\tEnter the number of likes you got in post %d of genre %s
\n",l+1,g1);
        scanf("%d",&dg1[l]);
        printf("\tEnter the number of likes you got in post %d of genre %s
\n",l+1,g2);
        scanf("%d",&dg2[l]);
        printf("\tEnter the number of likes you got in post %d of genre %s
\n",l+1,g3);
        scanf("%d",&dg3[l]);
        printf("\tEnter the number of followers you had corresponding to
these posts\n");
        scanf("%d",&df[l]);
    }
    printf("\n\tThe name of your genres are: \n\t%s,\t%s,\t%s\t and final
column is the Number of followers\n",g1, g2, g3);
    printf("\t Your data set is:\n");
    for(l=0;l<np;l++)
    {
        printf("\t%d\t%d\t%d\t%d\n",dg1[l], dg2[l], dg3[l], df[l]);
    }
    sleep(3);
    dispmat[0][0]= covariance(df,df,np);
    dispmat[0][1]= covariance(df,dg1,np);
    dispmat[0][2]= covariance(df,dg2,np);
    dispmat[0][3]= covariance(df,dg3,np);
}

```

```

dispmat[1][0]= covariance(dg1,df,np);
dispmat[1][1]= covariance(dg1,dg1,np);
dispmat[1][2]= covariance(dg1,dg2,np);
dispmat[1][3]= covariance(dg1,dg3,np);
dispmat[2][0]= covariance(dg2,df,np);
dispmat[2][1]= covariance(dg2,dg1,np);
dispmat[2][2]= covariance(dg2,dg2,np);
dispmat[2][3]= covariance(dg2,dg3,np);
dispmat[3][0]= covariance(dg3,df,np);
dispmat[3][1]= covariance(dg3,dg1,np);
dispmat[3][2]= covariance(dg3,dg2,np);
dispmat[3][3]= covariance(dg3,dg3,np);

```

```

sigma11= detof(dispmat,0,0);
sigma22= detof(dispmat,1,1);
sigma33= detof(dispmat,2,2);
sigma44= detof(dispmat,3,3);
sigma12= detof(dispmat,0,1);
sigma13= detof(dispmat,0,2);
sigma14= detof(dispmat,0,3);

```

```

parcorr12= ((-1)*sigma12)/((sqrt(abs(sigma11)))*(sqrt(abs(sigma22))));
parcorr13= ((-1)*sigma13)/((sqrt(abs(sigma11)))*(sqrt(abs(sigma33))));
parcorr14= ((-1)*sigma14)/((sqrt(abs(sigma11)))*(sqrt(abs(sigma44))));

```

```

if(parcorr12> parcorr13 && parcorr13> parcorr14)
    printf("\n\tGenre %s brings you most followers, followed by genre %s and
then genre %s.\n",g1, g2, g3);
    else if(parcorr12> parcorr14 && parcorr14> parcorr13)
        printf("\n\tGenre %s brings you most followers, followed by genre %s and
then genre %s.\n",g1, g3, g2);
        else if(parcorr13> parcorr12 && parcorr12> parcorr14)
            printf("\n\tGenre %s brings you most followers, followed by genre %s and
then genre %s.\n",g2, g1, g3);
            else if(parcorr13> parcorr14 && parcorr14> parcorr12)
                printf("\n\tGenre %s brings you most followers, followed by genre %s and
then genre %s.\n",g2, g3, g1);
                else if(parcorr14> parcorr12 && parcorr12> parcorr13)
                    printf("\n\tGenre %s brings you most followers, followed by genre %s and
then genre %s.\n",g3, g1, g2);
                    else if(parcorr14> parcorr13 && parcorr13> parcorr12)
                        printf("\n\tGenre %s brings you most followers, followed by genre %s and
then genre %s.\n",g3, g2, g1);

```

