PRODUCTIVE INSTAGRAM POSTING STRATEGY

TUYEN NGUYEN & NHI LAI | Mathematical Statistics | Oct 2019

Table of Contents

1. Introduction

1.1. Reasons for choosing the project – Case study

Online marketing campaigns are becoming more and more popular. A lot of companies have started to hire social media influencers to promote their products. Furthermore, they dedicate more time and effort to popularize, develop their own social media accounts like Instagram or Facebook. Therefore, they have to get to know the "trend" to find the most efficient way to advertise their products. As a result, the need to understand all the insights of social media like at which time of the day a post can reach most people is increasing rapidly.

On account of that, we came up with a case study:

Suppose that SP Jain School of Global Management is holding a contest. You have to make a video or take a photo featuring one of the most unique things about the school. The ones which have most likes or most comments will win the prizes. How can we find out a strategy to have the most popular post to be the champion?

Basing on this, we have decided to collect data from 511 posts from 48 popular accounts on Instagram to figure out the way to achieve most interaction for a post.

1.2. Objectives of the project

The main purposes of this project are:
- Figure out which day in a week a post can achieve most interaction.
- Figure out which type of post achieve more interaction.

2. Problem Approach

2.1. Loading the dataset

The dataset was originally prepared in a spreadsheet and exported as an open XML spreadsheet file named 'dataset.xlsx'. In our assignment, we only deal with sheet2 from the whole data collection.
As it required in the dataset, the type of each column is numberic, date, text, text, numeric, numeric (respectively)
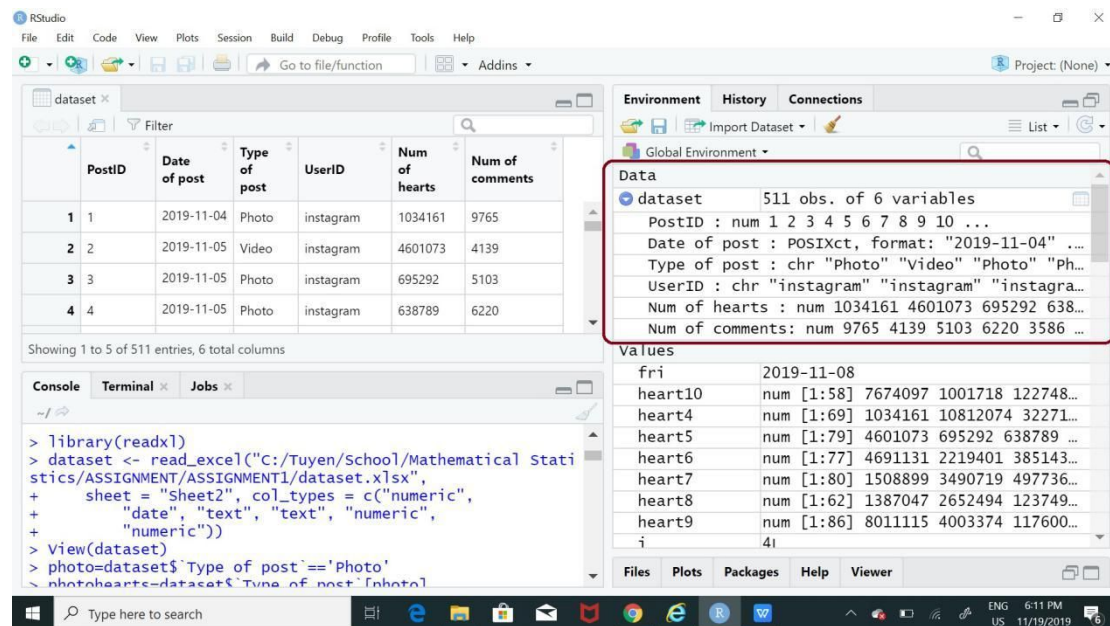
#importing data from excel
Input:
> library(readxl)

> dataset <- read_excel("C:/Tuyen/School/Mathematical Statistics/ASSIGNMENT/ASSIGNMENT1/dataset.xlsx", sheet = "Sheet2", col_types = c("numeric", "date", "text", "text", "numeric", "numeric"))

Output:
In the Global Environment field, we could see the summary of data frame.

2.2. Insights & Analysis

2.2.1. General Statistical Analysis

Task 1: taking summary of 'number of hearts' and 'number of comments'
Input:
> summary(dataset$`Num of hearts`)
> summary(dataset$`Num of comments`)
Output:
Num of hearts:

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
27323	173275	492100	1129544	1327724	108120774

Num of comments:

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
16	446	1548	8027	4422	1927881

Task 2: Computing the variation of values compared to mean:
*** Variance of number of hearts(Total)***
*** Variance of number of comments(Total)***

Task 3: Figuring the relationship between number of hearts and comments:
**Covariance:
**Correlation:

Task 4: Day with highest number of posts
Since, there is no built-in function in R Studio to find out the most common day for posting. We have to create it function with some logical mathematics.
#create Mode function
Input:

```
> Mode = function(x){
    ta = table(x)
    tam = max(ta)
    if (all(ta == tam))
        mod = NA
    else
        if(is.numeric(x))
    mod = as.numeric(names(ta)[ta == tam])
    else
        mod = names(ta)[ta == tam]
    return(mod)
}
```

Output:
now we have a Mode function to find the most common value in any set
#retrieve the day with highest number of post
Input:

```
> Mode(dataset$`Date of post`)
```

Output:

[1] "2019-11-09"

2.2.2. Detailed Statistical Analysis

2.2.2.A. Stratify Data:
Task 1: Classify 'number of hearts' and 'number of comments' according to 'Type of post'
Input:
 # Create vector with logical operation

```
> photo=dataset$`Type of post`=='Photo'
> video=dataset$`Type of post`=='Video'
```

 # Store data:
**Photo:

```
> photohearts=dataset$`Num of hearts`[photo]
> photocoms=dataset$`Num of comments`[photo]
```

**Video:
```
> videohearts=dataset$`Num of hearts`[video]
> videocoms=dataset$`Num of comments`[video]
```

Output:

Task 2: Classify data according to 'Date of post'
#Use 'as.Date.POSIXct' function to have value with same format for logical operation
Input:
```
> mon=as.Date.POSIXct("2019-11-04",tz="UCT",Format="%Y-%m-%d")
> tue=as.Date.POSIXct("2019-11-05",tz="UCT",Format="%Y-%m-%d")
> wed=as.Date.POSIXct("2019-11-06",tz="UCT",Format="%Y-%m-%d")
> thu=as.Date.POSIXct("2019-11-07",tz="UCT",Format="%Y-%m-%d")
> fri=as.Date.POSIXct("2019-11-08",tz="UCT",Format="%Y-%m-%d")
> sat=as.Date.POSIXct("2019-11-09",tz="UCT",Format="%Y-%m-%d")
> sun=as.Date.POSIXct("2019-11-10",tz="UCT",Format="%Y-%m-%d")
```
Output:

 # Create vector with logical operation
Input:
**Post:
```
> post4=dataset$`Date of post`==mon
> post5=dataset$`Date of post`==tue
> post6=dataset$`Date of post`==wed
> post7=dataset$`Date of post`==thu
> post8=dataset$`Date of post`==fri
> post9=dataset$`Date of post`==sat
> post10=dataset$`Date of post`==sun
```

**Photo:
```
> photo4=dataset$`Type of post`=='Photo'&dataset$`Date of post`==mon
> photo5=dataset$`Type of post`=='Photo'&dataset$`Date of post`==tue
> photo6=dataset$`Type of post`=='Photo'&dataset$`Date of post`==wed
> photo7=dataset$`Type of post`=='Photo'&dataset$`Date of post`==thu
> photo8=dataset$`Type of post`=='Photo'&dataset$`Date of post`==fri
> photo9=dataset$`Type of post`=='Photo'&dataset$`Date of post`==sat
> photo10=dataset$`Type of post`=='Photo'&dataset$`Date of post`==sun
```

**Video:
```
> video4=dataset$`Type of post`=='Video'&dataset$`Date of post`==mon
> video5=dataset$`Type of post`=='Video'&dataset$`Date of post`==tue
> video6=dataset$`Type of post`=='Video'&dataset$`Date of post`==wed
> video7=dataset$`Type of post`=='Video'&dataset$`Date of post`==thu
> video8=dataset$`Type of post`=='Video'&dataset$`Date of post`==fri
> video9=dataset$`Type of post`=='Video'&dataset$`Date of post`==sat
> video10=dataset$`Type of post`=='Video'&dataset$`Date of post`==sun
```

Store Data According to 'Date of post':
Hearts(Total):
```
> heart4=dataset$`Num of hearts`[post4]
> heart5=dataset$`Num of hearts`[post5]
> heart6=dataset$`Num of hearts`[post6]
> heart7=dataset$`Num of hearts`[post7]
> heart8=dataset$`Num of hearts`[post8]
> heart9=dataset$`Num of hearts`[post9]
> heart10=dataset$`Num of hearts`[post10]
```

Hearts(Photo):
```
> photohearts4=dataset$`Num of hearts`[photo4]
> photohearts5=dataset$`Num of hearts`[photo5]
> photohearts6=dataset$`Num of hearts`[photo6]
> photohearts7=dataset$`Num of hearts`[photo7]
> photohearts8=dataset$`Num of hearts`[photo8]
> photohearts9=dataset$`Num of hearts`[photo9]
> photohearts10=dataset$`Num of hearts`[photo10]
```

Hearts(Video):
```
> videohearts4=dataset$`Num of hearts`[video4]
> videohearts5=dataset$`Num of hearts`[video5]
> videohearts6=dataset$`Num of hearts`[video6]
> videohearts7=dataset$`Num of hearts`[video7]
> videohearts8=dataset$`Num of hearts`[video8]
> videohearts9=dataset$`Num of hearts`[video9]
> videohearts10=dataset$`Num of hearts`[video10]
```

Comments(Total):
```
> comment4=dataset$`Num of comments`[post4]
> comment5=dataset$`Num of comments`[post5]
> comment6=dataset$`Num of comments`[post6]
> comment7=dataset$`Num of comments`[post7]
> comment8=dataset$`Num of comments`[post8]
> comment9=dataset$`Num of comments`[post9]
> comment10=dataset$`Num of comments`[post10]
```

Comments(Photo):
```
> photocomments4=dataset$`Num of comments`[photo4]
> photocomments5=dataset$`Num of comments`[photo5]
> photocomments6=dataset$`Num of comments`[photo6]
> photocomments7=dataset$`Num of comments`[photo7]
> photocomments8=dataset$`Num of comments`[photo8]
> photocomments9=dataset$`Num of comments`[photo9]
> photocomments10=dataset$`Num of comments`[photo10]
```

Comments(Video):

```
> videocomments4=dataset$`Num of comments`[video4]
> videocomments5=dataset$`Num of comments`[video5]
> videocomments6=dataset$`Num of comments`[video6]
> videocomments7=dataset$`Num of comments`[video7]
> videocomments8=dataset$`Num of comments`[video8]
> videocomments9=dataset$`Num of comments`[video9]
> videocomments10=dataset$`Num of comments`[video10]
```

2.2.2.B. Compute Data:

Task 1: Summation:
** Total number of hearts (photo)
```
> sum(dataset$`Num of hearts`[photo])
```

[1]   258112993

(a)   Total number of hearts (photo) - 04-11-2019 - Monday
```
> sum(dataset$`Num of hearts`[photo4])
```

[1]   45254199

(b)   Total number of hearts (photo) - 05-11-2019 - Tuesday
```
> sum(dataset$`Num of hearts`[photo5])
```

[1]   38492127

(c)   Total number of hearts (photo) - 06-11-2019 - Wednesday
```
> sum(dataset$`Num of hearts`[photo6])
```

[1]   46893894

(d)   Total number of hearts (photo) - 07-11-2019 - Thursday
```
> sum(dataset$`Num of hearts`[photo7])
```

[1]   36567527

(e)   Total number of hearts (photo) - 08-11-2019 - Friday
```
> sum(dataset$`Num of hearts`[photo8])
```

[1]   16390365

(f)   Total number of hearts (photo) - 09-11-2019 - Saturday
```
> sum(dataset$`Num of hearts`[photo9])
```

[1]   41930450

(g)   Total number of hearts (photo) - 10-11-2019 - Sunday
```
> sum(dataset$`Num of hearts`[photo10])
```

[1] 32584431

** Total number of comments (photo)
> sum(dataset$`Num of comments`[photo])

[1] 3352232

(a)   Total number of comments (photo) - 04-11-2019 - Monday
> sum(dataset$`Num of comments`[photo4])

[1] 208024

(b)   Total number of comments (photo) - 05-11-2019 - Tuesday
> sum(dataset$`Num of comments`[photo5])

[1]   2138740

(c)   Total number of comments (photo) - 06-11-2019 - Wednesday
> sum(dataset$`Num of comments`[photo6])

[1]   287532

(d)   Total number of comments (photo) - 07-11-2019 - Thursday
> sum(dataset$`Num of comments`[photo7])

[1] 254580

(e)   Total number of comments (photo) - 08-11-2019 - Friday
> sum(dataset$`Num of comments`[photo8])

[1]   91856
(f)   Total number of comments (photo) - 09-11-2019 - Saturday
> sum(dataset$`Num of comments`[photo9])

[1]   208043

(g)   Total number of comments (photo) - 10-11-2019 - Sunday
> sum(dataset$`Num of comments`[photo10])

[1] 163457

** Total number of hearts (video)
> sum(dataset$`Num of hearts`[video])

[1] 319084217

(a)   Total number of hearts (video) - 04-11-2019 - Monday
> sum(dataset$`Num of hearts`[video4])

[1]   38512526

(b)   Total number of hearts (video) - 05-11-2019 - Tuesday
> sum(dataset$`Num of hearts`[video5])

[1]   50206227

(c)   Total number of hearts (video) - 06-11-2019 - Wednesday
> sum(dataset$`Num of hearts`[video6])

[1]   54677197

(d)   Total number of hearts (video) - 07-11-2019 - Thursday
> sum(dataset$`Num of hearts`[video7])

[1]   62780724

(e)   Total number of hearts (video) - 08-11-2019 - Friday
> sum(dataset$`Num of hearts`[video8])

[1]   40460361

(f)   Total number of hearts (video) - 09-11-2019 - Saturday
> sum(dataset$`Num of hearts`[video9])

[1]   52309974

(g)   Total number of hearts (video) - 10-11-2019 - Sunday
> sum(dataset$`Num of hearts`[video10])

[1] 20137208

*** Total number of comments (video)*
> sum(dataset$`Num of comments`[video])

[1] 749421

(a)   Total number of comments (video) - 04-11-2019 - Monday
> sum(dataset$`Num of comments`[video4])

[1]   77353

(b)   Total number of comments (video) - 05-11-2019 - Tuesday
> sum(dataset$`Num of comments`[video5])

[1]   72607

(c)   Total number of comments (video) - 06-11-2019 - Wednesday
> sum(dataset$`Num of comments`[video6])

[1]   164666

(d)   Total number of comments (video) - 07-11-2019 - Thursday
> sum(dataset$`Num of comments`[video7])

[1]   146050

(e)   Total number of comments (video) - 08-11-2019 - Friday
> sum(dataset$`Num of comments`[video8])

[1]   116741

(f)   Total number of comments (video) - 09-11-2019 - Saturday
> sum(dataset$`Num of comments`[video9])

[1]   102069

(g)   Total number of comments (video) - 10-11-2019 - Sunday
> sum(dataset$`Num of comments`[video10])

[1] 69935

Task 2: Expected value:
** Mean number of hearts (photo)
> mean(dataset$`Num of hearts`[photo])

[1] 739578.8

** Mean number of comments (photo)
> mean(dataset$`Num of comments`[photo])

[1] 9605.249

** Mean number of hearts (video)
> mean(dataset$`Num of hearts`[video])

[1] 1969656

** Mean number of comments(video)

> mean(dataset$`Num of comments`[video])

[1] 4626.056

Task 3: The variation compared to mean values:
** *Variance value of hearts (photo)*
> var(dataset$`Num of hearts`[photo])

[1]   1.328433e+12

** *Variance of comments (photo)*
> var(dataset$`Num of comments`[photo])

[1]   10682749140

** *Variance of hearts (video)*
> var(dataset$`Num of hearts`[video])

[1]   3.698916e+12

** *Variance of comments(video)*
> var(dataset$`Num of comments`[video])

[1] 53173680

Task 4: The relationship between number of hearts and comments:
**Correlation:*
> cor(dataset$`Num of hearts`[photo],dataset$`Num of comments`[photo])[1]
0.1615336
> cor(dataset$`Num of hearts`[video],dataset$`Num of comments`[video])[1]
0.5853755
> cor(dataset$`Num of hearts`,dataset$`Num of comments`)
[1] 0.1086016

2.2.2.C. Build Data Summary:

** Table 2.2.2.C (1) - Number of posts:

POST			
Day	No of Photos	No of Videos	Total
Mon	46	23	69
Tue	55	24	79
Wed	55	22	77
Thu	46	34	80
Fri	41	21	62
Sat	58	28	86

Sun	48	10	58
Total	349	162	511

** Table 2.2.2.C (2) - Number of Hearts and Comments:

Day	No of Hearts			No of Comments		
	Photo	Video	Total	Photo	Video	Total
Mon	45254199	38512526	83766725	208024	77353	285377
Tue	38492127	50206227	88698354	2138740	72607	2211347
Wed	46893894	54677197	101571091	287532	164666	452198
Thu	36567527	62780724	99348251	254580	146050	400630
Fri	16390365	40460361	56850726	91856	116741	208597
Sat	41930450	52309974	94240424	208043	102069	310112
Sun	32584431	20137208	52721639	163457	69935	233392
Total	258112993	319084217	577197210	3352232	749421	4101653

** Table 2.2.2.C (3) - Photo Summary

	Number of hearts						Number of comments					
	Min	Max	Mean	Median	Q1	Q3	Min	Max	Mean	Median	Q1	Q3
04-11-2019	28747	10812074	983787	218338	77152	1028594	16	59298	4522.3	728	157.2	4768.5
05-11-2019	27323	3293649	699857	480285	179827	903921	18	1927881	38886.2	2282	612.5	4995.5
06-11-2019	31050	4691131	852616	492100	162634	1295096	33	66722	5228	2019	544	4874
07-11-2019	40821	6692005	794946	386784	160592	788479	82	82283	5534	1738	525	4308
08-11-2019	45898	1387047	399765	225801	107587	494572	75	8447	2240	1072	304	4049
09-11-2019	57929	6712609	722939	403114	144270	842355	43	62205	3586.9	1294.5	407.2	3376.2
10-11-2019	46677	7674097	678842	287446	163829	842891	153	44696	3405.3	1463	470.5	4008.5
Total (whole week)	27323	10812074	739579	353353	135298	894732	16	1927881	9605	1516	402	4356

** Table 2.2.2.C (4) - Video Summary

	Number of hearts						Number of comments					
	Min	Max	Mean	Median	Q1	Q3	Min	Max	Mean	Median	Q1	Q3
04-11-2019	81548	724431	1674458	585810	340904	1850958	60	30015	3383.2	1894	422.5	3601.5
05-11-2019	82255	7250801	2091926	1214511	282574	3524078	42	17392	3025.3	940	255.8	4011.5
06-11-2019	196709	8781018	2485327	2199310	661732	3439911	117	37677	7485	4816	1099	8311
07-11-2019	132867	6104185	1846492	1571164	764582	2466408	71	21198	4295.6	1577.5	766.8	3734
08-11-2019	89068	4849749	1926684	1530618	698622	2697768	72	40207	559	2129	515	6182
09-11-2019	95800	8011115	1868213	1155444	24460	2493800	94	22414	3645.3	913	428,5	3277
10-11-2019	278442	7260083	2013721	1126151	412701	2946640	135	33962	6993.5	1305	427.8	6777.8
Total (whole week)	81548	8781018	1969656	1294526	396520	3137977	42	40207	4626.1	1565.5	510.2	4627.8

** Table 2.2.2.C (5) - Post Summary

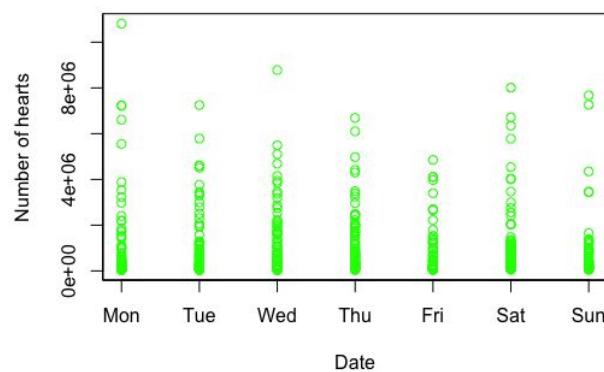|	Number of hearts						Number of comments					
	Min	Max	Mean	Median	Q1	Q3	Min	Max	Mean	Median	Q1	Q3
04-11-2019	28747	10812074	1214011	358092	107661	28747	16	59298	4136	1228	188	4095
05-11-2019	27323	7250801	1122764	568520	210878	1222034	18	1927881	27991.7	1921	577.5	4560
06-11-2019	31050	8781018	1319105	775019	235919	1751710	33	66722	5873	2820	545	6387
07-11-2019	40821	6692005	1241853	676933	254805	1807192	71	82283	5007.9	1638.5	605.8	82283
08-11-2019	45898	4849749	916947	416116	130844	1234600	72	40207	3364	1236	350	4435
09-11-2019	57929	8011115	1095819	449600	159920	1158096	43	62205	3605.9	1128	407.2	3376.2
10-11-2019	46677	7674097	908994	355610	169638	932291	135	44696	4024	1463	463.5	4069.5
Total (whole week)	27323	10812074	1129544	492100	173275	1327724	16	1927881	8027	1548	446	4422

2.3. Visualizing Data:

2.3.1. Processing Raw Data:

Firstly, we use plot() function to visualize the number of hearts and comments grouped by date of post.
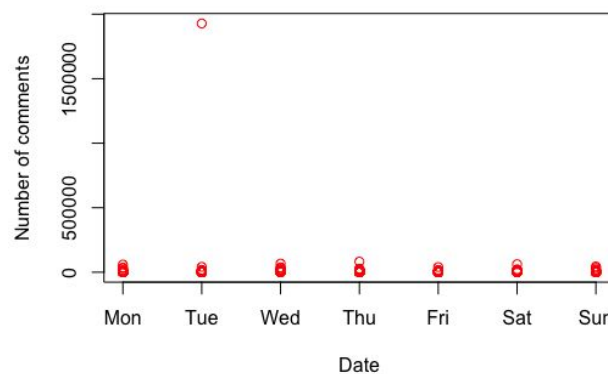
Number of hearts

plot(dataset_copy$`Date of post`,dataset_copy$`Num of hearts`, xlab ='Date', ylab ='Number of hearts',col ='green')

Dot plot 2.3.1.1

Number of comments

plot(dataset_copy$`Date of post`,dataset_copy$`Num of comments`,xlab ='Date', ylab ='Number of comments',col ='red')

Dot plot 2.3.1.2

From this graph, we can tell that there is a significant outlier in the number of comments. We detect this outlier by using the outlier() function.

> install.packages("outliers")
> library(outliers)

> outlier(dataset$`Num of comments`)

[1] 1927881

We removed the data of that post. Then we graph the box plot of the number of hearts and the number of comments grouped by date of post to see if we need to remove any more outliers for smaller range and a more objective result.

Box plot of the number of hearts:

ggplot(dataset,aes(x =as.factor(dataset$`Date of post`),y =dataset$`Num of hearts`)) +geom_boxplot() + xlab("Date of post") + ylab("Number of hearts")+ ggtitle("Number of hearts before removing outliers") + theme(plot.title = element_text(hjust = 0.5))
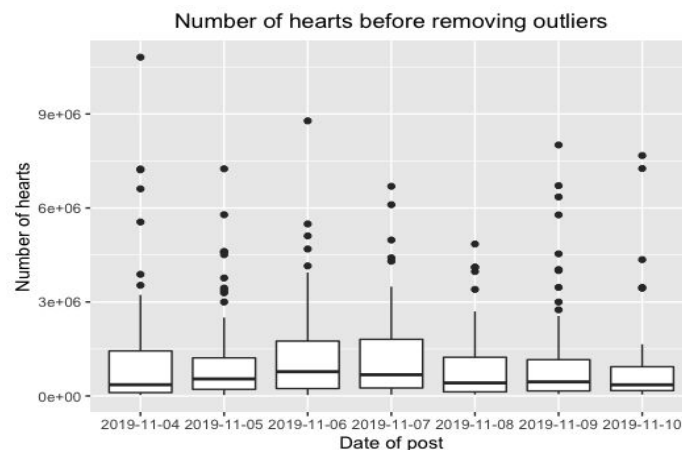
Box plot 2.3.1.3

We kept the data within the 3-standard deviation range and removed the others. This is the box plot after removing the significant outliers:
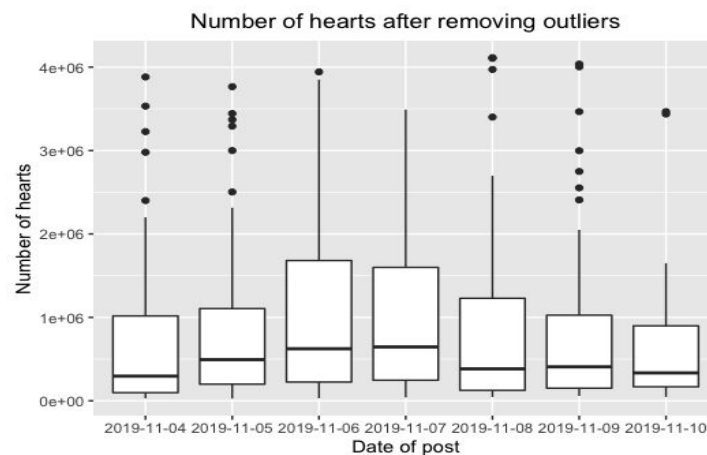
Box plot 2.3.1.4

It is obviously showed that the majority of numbers of hearts per post are less than 2,000,000. Furthermore, there is no outlier in the Thursday data.

Box plot of the number of comments:

ggplot(dataset,aes(x =as.factor(dataset$`Date of post`),y =dataset$`Num of comments`)) +geom_boxplot() + xlab("Date of post") + ylab("Number of comments")+ggtitle("Number of comments before removing outliers") + theme(plot.title = element_text(hjust = 0.5))
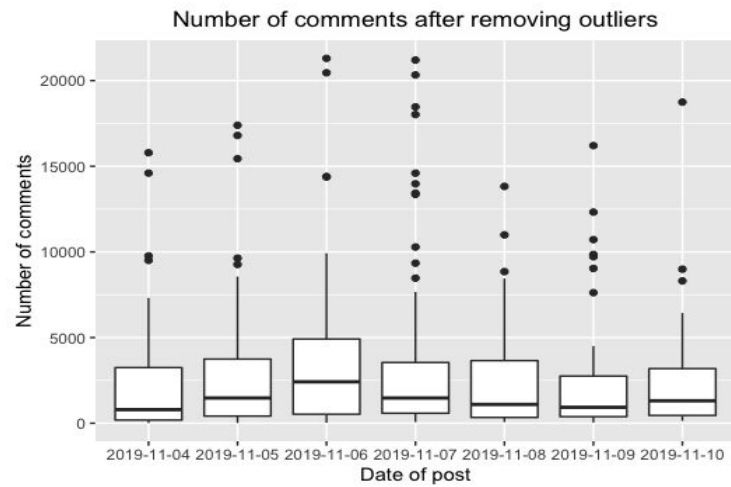
Box plot 2.3.1.5

We kept the data within the 3-standard deviation range and remove the others. This is the boxplot after removing the outliers which shows that most of the number of comments are less than 10,000.
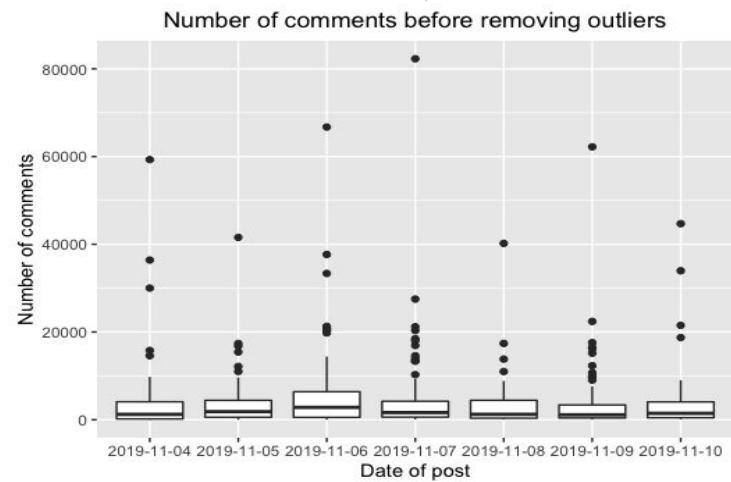
Box plot 2.3.1.6

2.3.2. Analyzing Processed Data:

2.3.2.A. Processed Data Summary:

Number of posts:

table(dataset_copy$`Date of post`) #total post on each day
table(dataset_copy$`Date of post`[photo]) #total photo posts on each day
table(dataset_copy$`Date of post`[video]) #total video posts on each day

POST			
Day	No of Photos	No of Videos	Total
Mon	44	20	64
Tue	53	19	72
Wed	52	19	71
Thu	44	30	74
Fri	41	19	60
Sat	57	24	81
Sun	47	8	55
Total	338	139	477

Table 2.3.2.1

Number of Hearts and Comments:

Day	No of Hearts			No of Comments		
	Photo	Video	Total	Photo	Video	Total
Mon	27224039	19103478	46327517	112320	38510	150830
Tue	35031316	23448312	58479628	169289	34277	203566
Wed	36886135	35295850	72181985	166706	95084	261790
Thu	27653648	42987601	70641249	144789	117038	261827
Fri	16390365	34911990	51302355	91856	59125	150981
Sat	35217841	27630491	62848332	145838	30403	176241
Sun	24910334	8525380	33435714	118761	14457	133218
Total	203313678	191903102	395216780	3352232	749421	4101653

Table 2.3.2.2

Day	% of Hearts			% of Comments		
	Photo	Video	Total	Photo	Video	Total
Mon	6.9	4.8	11.7	2.7	0.9	3.7
Tue	8.9	5.9	14.8	4.1	0.8	5
Wed	9.3	8.9	18.3	4.1	2.3	6.4
Thu	7	10.9	17.9	3.5	2.9	6.4
Fri	4.1	8.8	13	2.2	1.4	3.7
Sat	8.9	7	15.9	3.6	0.7	4.3
Sun	6.3	2.2	8.5	2.9	0.4	3.2
Total	51.4	48.6	100	81.7	18.3	100

Table 2.3.2.3

Speaking of number of hearts, there is no obvious distinction in preference for photo and video, 51.4% an 48.6%, respectively. While users tend to be more interested in photo posts when it comes to comments, 81.7% compared to 18.3%.

Descriptive Statistics of all the Posts:

summary(dataset_copy$`Num of comments`[post4]) #each day
summary(dataset_copy$`Num of comments`) #total

	Number of hearts						Number of comments					
	Min	Max	Mean	Median	Q1	Q3	Min	Max	Mean	Median	Q1	Q3
04-11-2019	28747	3882052	723867	295714	97444	1017458	16	15790	2357	793	185	3249
05-11-2019	27323	3765651	812217	493286	199578	1105474	18	17392	2827.3	1473	418.8	3744.2
06-11-2019	31050	3942798	1016648	623969	224718	1681574	33	21297	3687	2418	526	4916
07-11-2019	40821	3490719	954611	645836	246774	1599390	71	21198	3538	1472	584	3546
08-11-2019	45898	4113148	855039	382962	125471	1228806	72	13824	2516.3	1097	388.8	3655.2
09-11-2019	57929	4036508	775905	407598	151466	1026583	43	16204	2176	931	384	2753
10-11-2019	46677	3466179	607922	334611	168530	898446	135	18742	2422.1	1310	456.5	3188.5
Total (whole week)	27323	4113148	828547	433073	165220	1160748	16	21297	2806	1344	384	3634

Table 2.3.2.4

Descriptive Statistics of Photo posts:

summary(dataset_copy$`Num of hearts`[post4photo]) #each day
summary(dataset_copy$`Num of hearts`[photo]) #total

	Number of hearts						Number of comments					
	Min	Max	Mean	Median	Q1	Q3	Min	Max	Mean	Median	Q1	Q3
04-11-2019	28747	3227172	618728	213956	75416	966934	16	15790	2552.7	576.5	156.8	4211.2
05-11-2019	27323	3292649	660968	415519	171347	889416	18	16796	3194	2142	589	4671
06-11-2019	31050	2543355	709349	459434	145279	1135968	33	20457	3205.9	1902	525.8	4270.2
07-11-2019	40821	3490719	628492	361664	157210	716542	82	20330	3290.7	1553	522.8	3918
08-11-2019	45898	1387047	399765	225801	107587	494572	75	8447	2240	1072	304	4049
09-11-2019	57929	4003374	617857	399504	144239	793828	43	16204	2559	1199	402	3326
10-11-2019	46677	3438023	530007	273433	163154	798225	153	18742	2527	1435	467	3800
Total (whole week)	27323	4003374	601520	325507	126460	850530	16	20457	2809.3	1408	392.2	3942.5

Table 2.3.2.5

Descriptive Statistics of Video posts:

summary(dataset_copy$`Num of comments`[post4video]) #each day
summary(dataset_copy$`Num of comments`[video]) #total

	Number of hearts						Number of comments					
	Min	Max	Mean	Median	Q1	Q3	Min	Max	Mean	Median	Q1	Q3
04-11-2019	81548	3882052	955174	484499	308320	1219617	60	7318	1925.5	1465	364.5	3007.5
05-11-2019	82255	7250801	1234122	774987	210878	1771336	42	17392	1804.1	672	193.5	1411.5
06-11-2019	196709	3942798	1857676	1751710	565232	3026722	117	21297	5004	4510	6786	1411.5
07-11-2019	132867	3385412	1432920	1417626	662864	2011549	71	21198	3901	1360	649	2607
08-11-2019	89068	4113148	1837473	1530618	515537	2693422	72	13824	3112	1548	481	3218
09-11-2019	95800	4036508	1151270	1067468	229507	1497267	94	4507	1266.8	752	375.5	1536.5
10-11-2019	278442	3466179	1065672	686116	374574	1341262	135	8307	1807.1	830	294.8	1789.5
Total (whole week)	81548	4113148	1380598	1119340	315022	2176860	42	21297	2797.8	1062	379.5	2947

Table 2.3.2.6

2.3.2.B. Perfect Visualization:

2.3.2.B.1. Number of posts

Basing on the data from **Table 2.3.2.1**, we draw a pie graph illustrating the proportion of total posts on each day in a week.
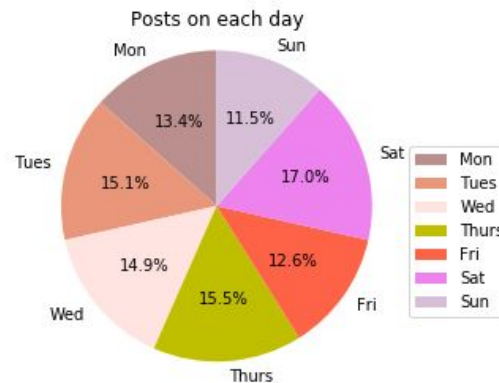

Pie graph 2.3.2.7

From this pie chart, we can say that people tend to post mostly on Saturday (17%). Besides, Thursday, Tuesday and Wednesday are also favourable days for people to post because they take up 15.5%, 15.1% and 14.9% respectively the number of posts in a week.
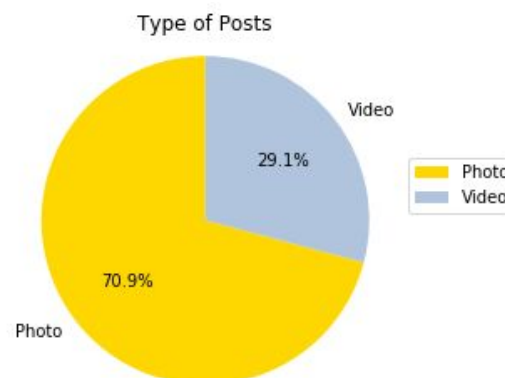
Pie graph 2.3.2.8

This 'Type of post' pie chart illustrates the number of photo posts compared to the number of video posts in a week. The portion of photos is larger than videos (70.9% > 29.1%). Hence, we can say that people have a tendency to post a photo more than a video.
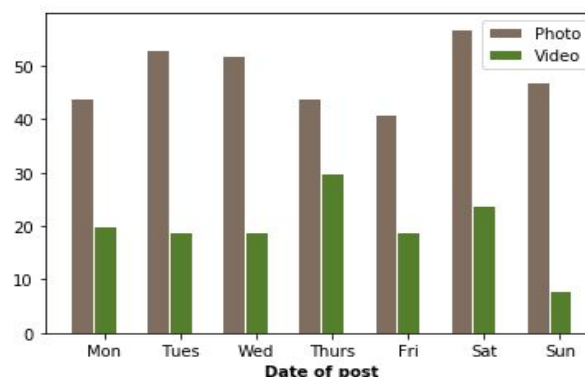
Grouped bars chart 2.3.2.9

From this grouped bar chart, we can see that in each day, the number of posts in 'Photo' type mostly at least double those in 'Video' type and even 4 times on Sunday. People tends to post more photos on Saturday and more videos on Thursday.

2.3.2.B.2. The correlation between hearts and comments
>cor(dataset_copy$`Num of hearts`,dataset_copy$`Num of comments`)
[1] 0.5687156
>plot(dataset_copy$`Num of hearts`,dataset_copy$`Num of comments`, xlab = "Number of hearts", ylab="Number of comments")
>abline(lm(dataset_copy$`Num of comments`~ dataset_copy$`Num of hearts`))
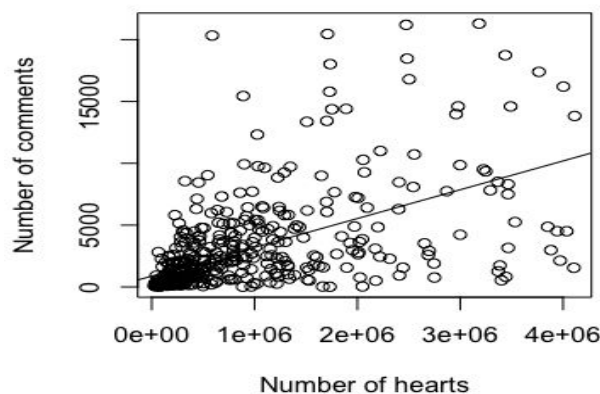
Scatter plot 2.3.2.10

We draw a scatter plot with a regression line to visualize the positive correlation between the number of hearts and comments. In general, it appears that a post with higher number of hearts tends to have higher number of comments.

2.3.2.B.3. Number of Hearts
We chose pie charts to visualize the number of hearts grouped by type of post on each day of the week so that we could easily see the proportion. We draw these charts basing on the **Table 2.3.2.2**.
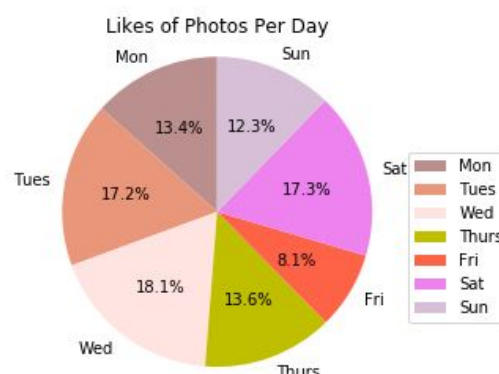
Pie chart 2.3.2.11
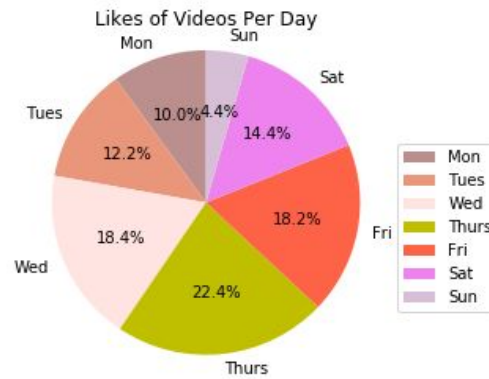
Pie chart 2.3.2.12

Regarding the photo posts, they tend to gain most of hearts on Wednesday (18.1%) or Saturday (17.3%).

Regarding the video posts, they tend to gain most of hearts on Thursday (22.4%) or Wednesday (18.4%).

To get more insights, we illustrate the largest number of hearts according to type with a grouped bar chart below:

>ggplot(dataset_copy, aes(fill=dataset_copy$`Type of post`, y=sum(dataset_copy$`Num of hearts`), x=dataset_copy$`Date of post`)) + geom_bar(position="dodge", stat="identity") + xlab("Date of post") + ylab("Number of hearts")+ggtitle("Hearts on each day") + labs(fill = 'Type of post') + theme(plot.title = element_text(hjust = 0.5))
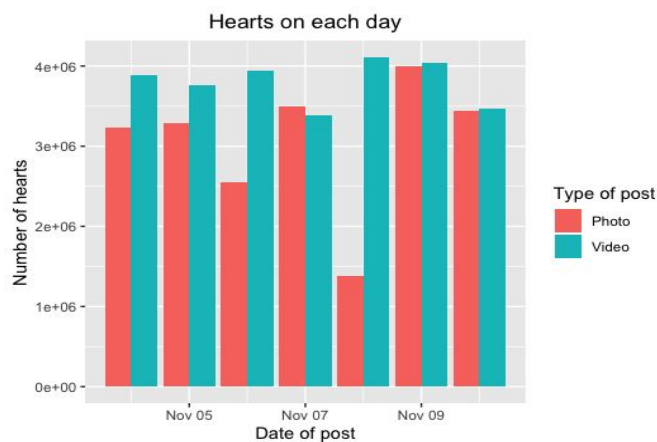
Grouped bars chart 2.3.2.13

Generally, video posts gain more hearts than photos for almost all the week except for Thursday. We can see a huge difference in hearts between these two types on Friday.

2.3.2.B.4. Number of Comments

The pie charts below show the number of comments grouped by type of post on each day of the week. We draw these pie charts basing on the **Table 2.3.2.2**.

Regarding the photos, they gain most comments on Tuesday (17.8%) or Wednesday (17.6%).

Regarding the videos, they gain most comments on Thursday (30.1%) or Wednesday (24.4%).
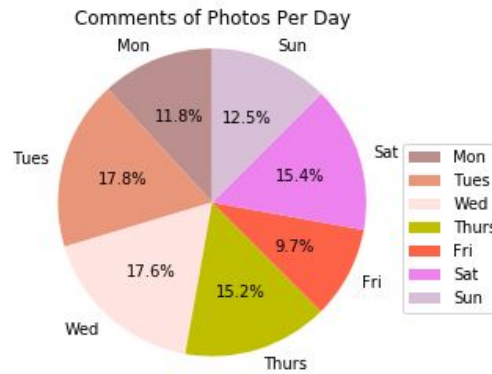
Pie chart 2.3.2.14
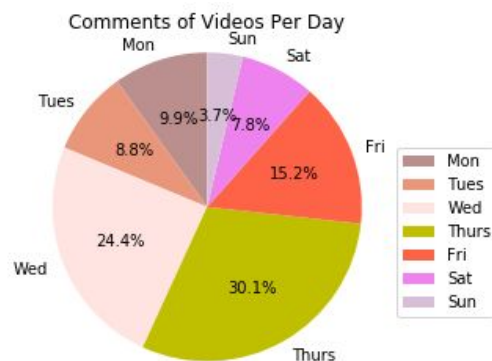
Pie chart 2.3.2.15

Below is the grouped bar chart of the largest number of comments according to type of post:

```
> ggplot(dataset_copy, aes(fill=dataset_copy$`Type of
post`,x=dataset_copy$`Date of post`,y=dataset_copy$`Num of comments`)) +
geom_bar(position="dodge", stat="identity") + xlab("Date of post") +
ylab("Number of comments") + ggtitle("Comments on each day") + labs(fill =
'Type of post') + theme(plot.title = element_text(hjust = 0.5))
```
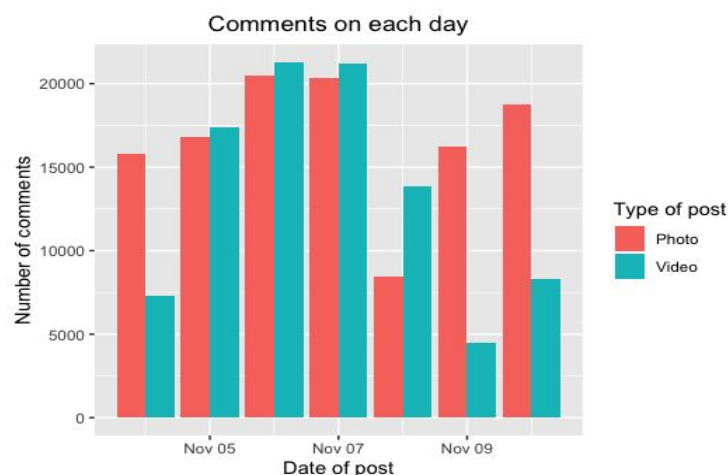

Grouped bars chart 2.3.2.16

From the chart, we can say that the largest number of comments on Monday, Saturday and Sunday belongs to photo posts. However, on Tuesday, Wednesday, Thursday and Friday, video posts gain the largest number of comments.

3. Conclusion

Throughout the study of more than 500 randomly picked posts on Instagram from most 48 popular Instagram users, as we can obviously see the culture of posting to achieve the huge success in having attention and support from other accounts. More precisely, the intensity of interactions always peaks at the end of the week, mostly on Saturday. Furthermore, wise marketers seem to prior photo-typed posts over video ones due to its convinience in making and attraction to normal users. All the statistics and visualizations we get illustrate the strongly dependent relationship between number of hearts and comments of each post. In other words, posts with larger quantity of comments are more likely to achieve more hearts and interests. Generally, to answer the question in the case study, the best strategy to win the prize is that that student should post a photo at weekend, especially Saturday. Trivially, he or she could use a comment trick (as the number of comments are not restricted for any accounts on Instagram): ask for support from friends, family to comment a lot in order to create a popular 'vibe' for his or her post, which would attracts more hearts in the future.

4. Index of tables and graphs