STAT3613 Project Final Report

Analysis of factors that affect the views and number of likes of YouTube channel - The best league

Index

Introduction

Background

The best league is a YouTube channel which uploads the highlights of League of legends (LoL) e-sport matches. The views, number of like of different match highlight have a huge difference. The difference in views represent that the viewers have some preference for the match highlight. This analysis aims to find out the relationship between the factors of the match highlight and the views, number of like of YouTube channel - The best league. After finding out the result, the viewer's preference can be studied, it is useful for starting a YouTube channel about e-sport highlight and improving the existing YouTube channel. Based on our assumptions, conjoint analysis will be applied to the data.

Definition of key word

1. tournament - there are six big tournament in the League of legends (LoL) e-sport match, they are LEC, LCS, LCK, LPL, MSI and World

2. knock-out-stage - a competition in which only the winners of each stage play in the next stage, until one competitor or team is the final winner

3. Top player - In this study, it means a player who had gotten any champion title

4. Past Top team - In this study, it means a team who had gotten any champion title

5. Strong team - In this study, it means a team with higher ranking than its opponent before the match

Objective

1. Find out the factor(s) affect the views and like of video in Youtube channel - The best league

2. Find out the relationship between factors and views and likes of video in Youtube channel - The best league

3. Find the weighting of each factors that affect the views and likes of video in Youtube channel - The best league

Method

The analysis procedure

1. Sampling method design
2. Data collection
3. Data cleaning
4. Data analysis

Sampling method design

Multi-stage sampling As the factors is very different from each tournament, the six tournaments are divided into six stratums.

Based on the 10 randomly selected pilot sample from each match region, the sample size from each stratum are allocated as the following:

	A	B	C	D	E	F	G	H
1	Tournament	Worlds	LPL	LCK	LEC	LCS	MSI	Total
2	average view of ten random match day	21000	27000	65000	20000	9455	172166	
3		87000	49000	85000	6774	7103	70939	
4		94990	38000	31000	6945	93000	140998	
5		48000	50000	30000	10000	14000	74464	
6		95000	28000	77000	100000	24000	89744	
7		63000	44000	92000	15000	26000	61232	
8		52000	60000	25000	32000	15000	116642	
9		45000	32000	64000	11000	11000	82794	
10		75000	45000	73000	25000	83000	171167	
11		30000	36000	91000	10000	17000	53338	
12	sd	26361.21328	10598.21788	25746.84447	28058.53327	31246.9556	44342.74	
13	variance	694913565.6	112322222.2	662900000	787281289.4	976372233	1.97E+09	
14	total match day	17	74	51	18	18	8	186
15	match day times sd	448140.6257	784268.1231	1313089.068	505053.5989	562445.2	354741.9	3967738.551
16	Error bound(10 % of total average)							5371.251667
17	square of (Error bound / 2) times square of total match day							2.49527E+11
18	match day times variance	11813530614	8311844444	33807900000	14171063210	1.7575E+10	1.57E+10	1.01409E+11
19	sample size = square of H15 / (H18 + H17)							44.8598998
20	Neyman Allocation(sample size of each strata)	5.082574847	8.894755812	14.89236432	5.7280518	6.3789571	4.023296	45

Figure 1: allocation base on variance and population size

Worlds: 5 match days will be sampled

LPL: 9 match days will be sampled

LCK: 15 match days will be sampled

LEC: 6 match days will be sampled

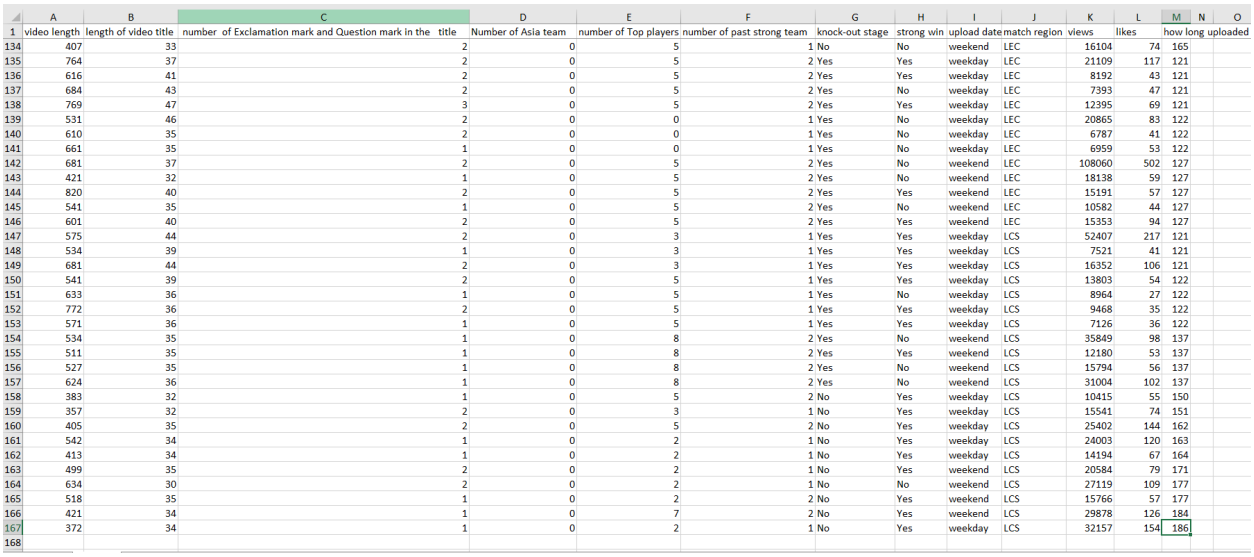
LCS: 5 match days will be sampled

MSI: 4 match days will be sampled

All match highlights video of the matches on the selected match days will be the sample.

Data Collection

The data is collected from the our studying YouTube channel - The best League, as the following:



	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
1	video length	length of video title	number of Exclamation mark and Question mark in the	title	Number of Asia team	number of Top players	number of past strong team	knock-out stage	strong win	upload date	match region	views	likes	how long	uploaded
134	407	33	2	0	5	1	No	No	weekend	LEC	16104	74	165		
135	764	37	2	0	5	2	Yes	Yes	weekday	LEC	21109	117	121		
136	616	41	2	0	5	2	Yes	Yes	weekday	LEC	8192	43	121		
137	684	43	2	0	5	2	Yes	No	weekday	LEC	7393	47	121		
138	769	47	3	0	5	2	Yes	Yes	weekday	LEC	12395	69	121		
139	531	46	2	0	0	1	Yes	No	weekday	LEC	20865	83	122		
140	610	35	2	0	0	1	Yes	No	weekday	LEC	6787	41	122		
141	661	35	1	0	0	1	Yes	No	weekday	LEC	6959	53	122		
142	681	37	2	0	5	2	Yes	No	weekend	LEC	108060	502	127		
143	421	32	1	0	5	2	Yes	No	weekend	LEC	18138	59	127		
144	820	40	2	0	5	2	Yes	Yes	weekend	LEC	15191	57	127		
145	541	35	1	0	5	2	Yes	No	weekend	LEC	10582	44	127		
146	601	40	2	0	5	2	Yes	Yes	weekend	LEC	15353	94	127		
147	575	44	2	0	3	1	Yes	Yes	weekday	LCS	52407	217	121		
148	534	39	1	0	3	1	Yes	Yes	weekday	LCS	7521	41	121		
149	681	44	2	0	3	1	Yes	Yes	weekday	LCS	16352	106	121		
150	541	39	2	0	5	1	Yes	Yes	weekday	LCS	13803	54	122		
151	633	36	1	0	5	1	Yes	No	weekday	LCS	8964	27	122		
152	772	36	2	0	5	1	Yes	Yes	weekday	LCS	9468	35	122		
153	571	36	1	0	5	1	Yes	Yes	weekday	LCS	7126	36	122		
154	534	35	1	0	8	2	Yes	No	weekend	LCS	35849	98	137		
155	511	35	1	0	8	2	Yes	Yes	weekend	LCS	12180	53	137		
156	527	35	1	0	8	2	Yes	No	weekend	LCS	15794	56	137		
157	624	36	1	0	8	2	Yes	No	weekend	LCS	31004	102	137		
158	383	32	1	0	5	2	No	Yes	weekday	LCS	10415	55	150		
159	357	32	2	0	3	1	No	Yes	weekday	LCS	15541	74	151		
160	405	35	2	0	5	2	No	Yes	weekday	LCS	25402	144	162		
161	542	34	1	0	2	1	No	Yes	weekday	LCS	24003	120	163		
162	413	34	1	0	2	1	No	Yes	weekday	LCS	14194	67	164		
163	499	35	2	0	2	1	No	Yes	weekend	LCS	20584	79	171		
164	634	30	2	0	2	1	No	No	weekend	LCS	27119	109	177		
165	518	35	1	0	2	2	No	Yes	weekend	LCS	15766	57	177		
166	421	34	1	0	7	2	No	Yes	weekend	LCS	29878	126	184		
167	372	34	1	0	2	1	No	Yes	weekday	LCS	32157	154	186		
168															

Figure 2: total 166 oberverion, 13 variables

Summary of Data:

```
##   video length    length of video title number of ! and ? Inf the title
##  Min.   : 121.0   Min.   :30.0          Min.   :1.000
##  1st Qu.: 524.2   1st Qu.:35.0          1st Qu.:1.000
##  Median : 616.5   Median :37.0          Median :2.000
##  Mean   : 648.7   Mean   :37.6          Mean   :1.789
##  3rd Qu.: 732.5   3rd Qu.:40.0          3rd Qu.:2.000
##  Max.   :2134.0   Max.   :47.0          Max.   :4.000
##  Number of Asia team Number of Top player Number of Past Strong Team
##  Min.   :0.000       Min.   : 0.000       Min.   :0.000
##  1st Qu.:0.000       1st Qu.: 2.000       1st Qu.:1.000
##  Median :2.000       Median : 5.000       Median :1.000
##  Mean   :1.343       Mean   : 4.422       Mean   :1.223
##  3rd Qu.:2.000       3rd Qu.: 6.000       3rd Qu.:2.000
##  Max.   :2.000       Max.   :10.000       Max.   :2.000
```

```
##      views            likes          how long uploaded
##  Min.   :  5992   Min.   :0.002734   Min.   : 23.0
##  1st Qu.: 18226   1st Qu.:0.004176   1st Qu.: 93.0
##  Median : 31251   Median :0.004829   Median :122.0
##  Mean   : 46014   Mean   :0.005077   Mean   :122.2
##  3rd Qu.: 66323   3rd Qu.:0.005600   3rd Qu.:151.0
##  Max.   :197193   Max.   :0.013055   Max.   :191.0
```

Defination to each variables

1. video length - the video length in seconds
2. length of video title - the word counts(including symbols) in the title
3. number of "!" and "?" in title - the counts of ! and ? symbols in the title
4. number of Asia team - number of Asia team in that match(max 2, min 0)
5. number of Top players - number of Top players(the players had got any champion title before) in that match(max 10, min 0)
6. number of past strong team - number of strong team(the team had got any champion title before) in that match(max 2, min 0)
7. knock-out stage - whether the match is in knock-out stage(Yes or No)
8. strong win - whether the stronger team(the team with higher rank) win the match(Yes or No)
9. upload date - the match video is uploaded on weekday or weekend
10. match region - the match is from which tournament
11. views - the views of the match video
12. likes - the number of likes of the match video
13. how long uploaded - (how long had the match video been uploaded)

Data analysis

Idea: in order to study how the factors affect and the weighting of the affect, conjoint analysis will be apply to the data.

Assumption 1: Since it is impossible to collect the profile of all the level of factors, some factors will be grouped.

video length, length of video title and number of "!" and "?" will be grouped as video factor since they are designed by the video maker.

number of Asia team, number of top player and number of past strong team will be grouped as match factor since they are fixed by the match result.

To determine the level of those group factor, cluster analysis will be applied. The data will be clustered by the grouped data. The charactertics of the cluster represent the level.

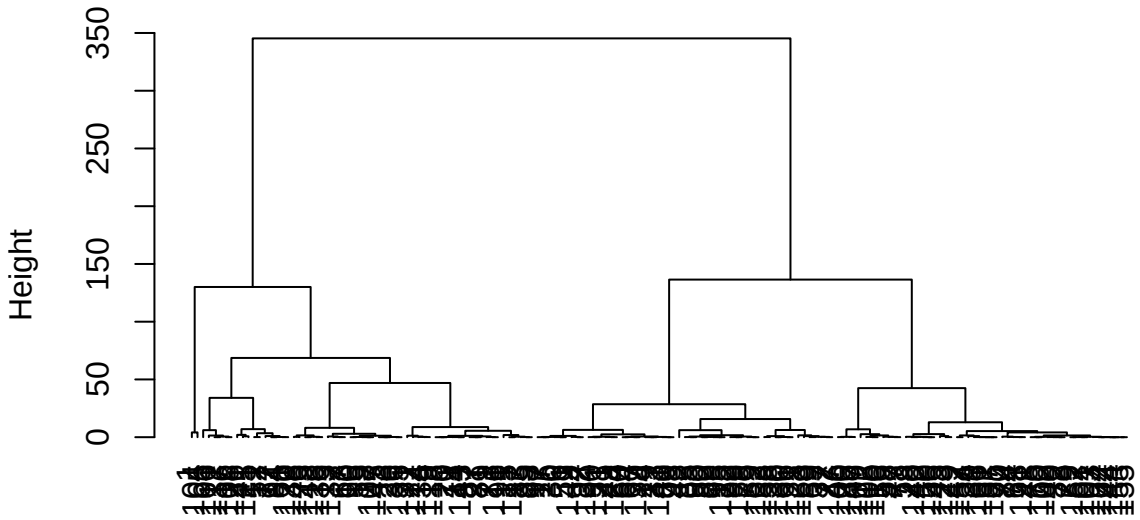
Assumption 2: The views and likes represent the rating of the viewers

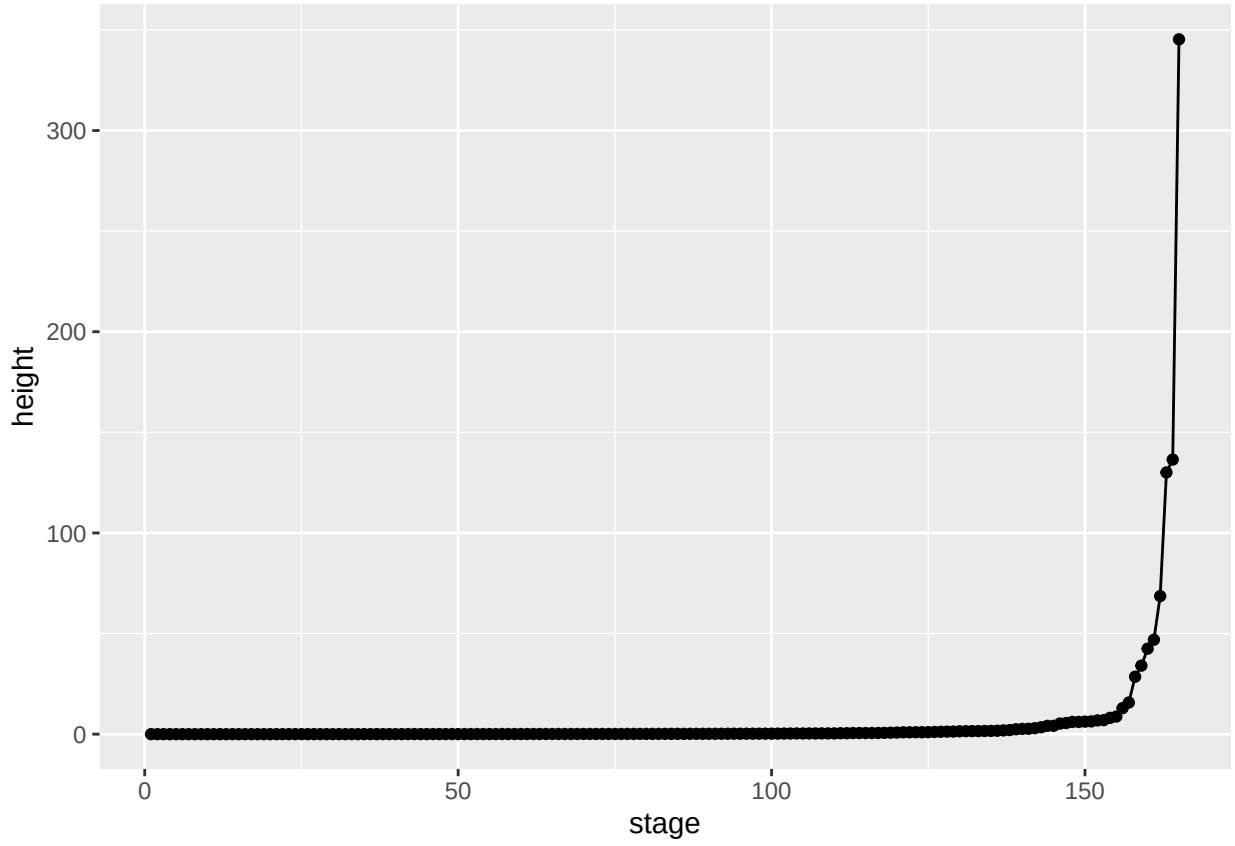
Assumption 3: If there is equal profile, the profile will be combined by there mean rating

Result

First Step:

use cluster analysis to divide group the videos on cluster base on the factors 1 - 3(the video factors), which are video length, length of video title and number of "!" and "?". Base on the number of cluster suggested, we can find out how many level of video factor is suitable.



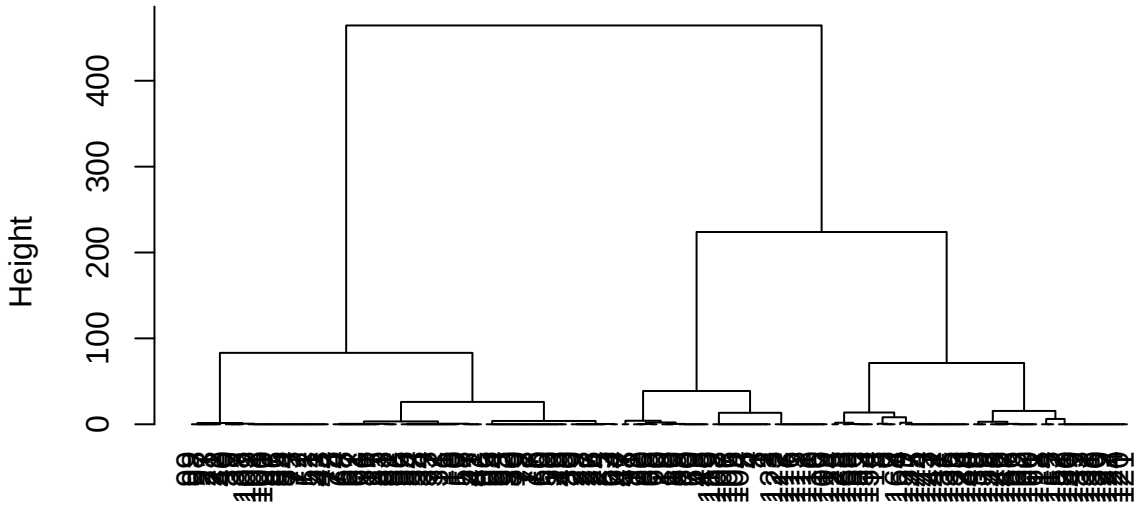


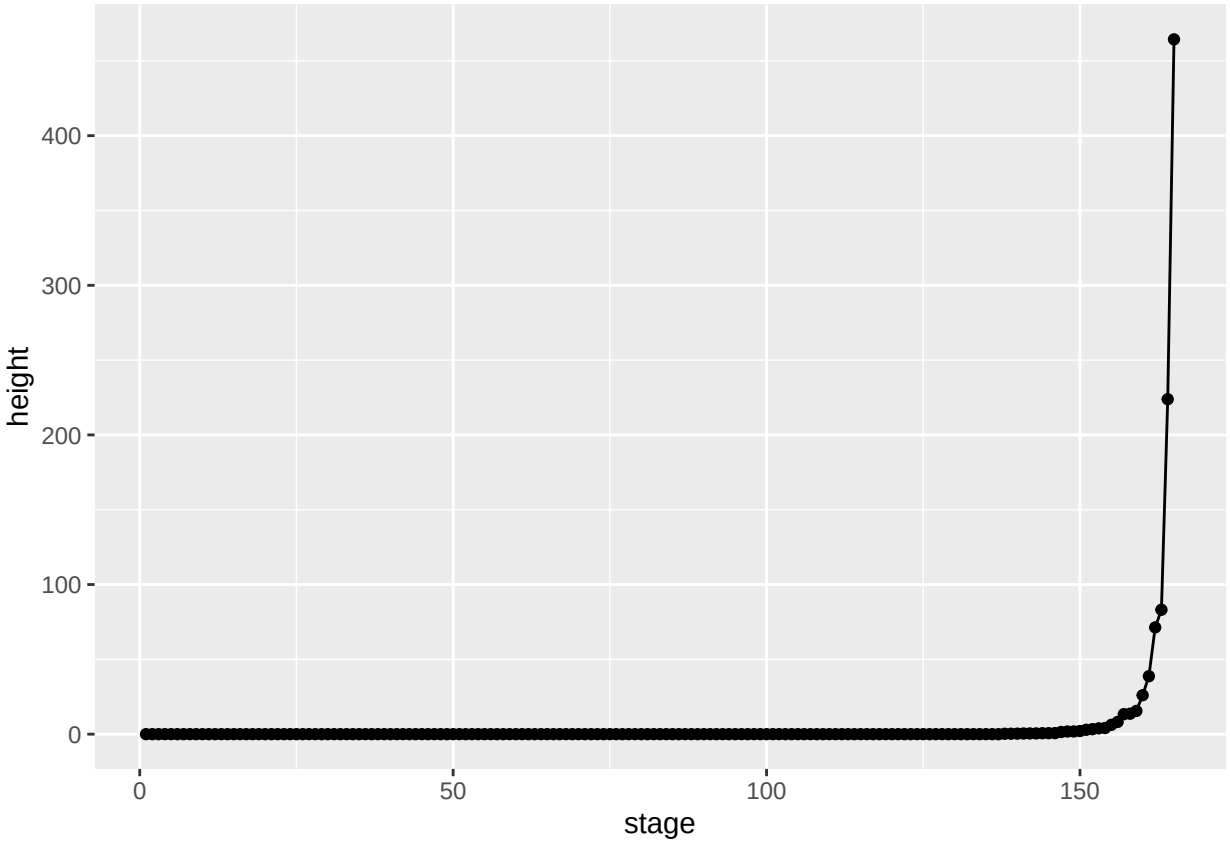

```
##   Group.1 video length length of video title number of ! and ? Inf the title
## 1       1     757.6557              41.16393                        2.278689
## 2       2     585.4381              35.53333                        1.504762
```

Based on the result, two cluster solution is suggested. The video factor is going to be divided in two level.
Level 1: Long video length, long length of video title and more number of ! and ? Inf the title
Level 2: Short video length, Short length of video title and less number of ! and ? Inf the title

Second Step:

use cluster analysis to divide group the videos on cluster base on the factors 4 - 6(the match factors), which are the number of Asia team, number of top player and number of past strong team. Base on the number of cluster suggested, we can find out how many level of match factor is suitable






```
##   Group.1 Number of Asia team Number of Top player Number of Past Strong Team
## 1       1           1.7297297             8.216216                  2.0000000
## 2       2           0.1320755             4.754717                  1.4716981
## 3       3           2.0000000             2.342105                  0.6710526
```

Based on the result, three cluster solution is suggested. The video factor is going to be divided in three level.
Level 1: most of Asia team, most of top player and all past strong team
Level 2: most of West team, about half of top player and most of past strong team
Level 3: all the Asia team, less of top player and less past strong team

Conjoint analysis of how the factors affect and the weight of affect of the factors on the video's view

We cheat the views as a customers giving rating.

The views are the customer's rating, more views means more prefer.

Combining the video factor and match factor level from the previous section, we have the following results

```
##                  [,1]
## ko.No       0.3750000
## ko.Yes     -0.3750000
## sw.No       0.6666667
## sw.Yes     -0.6666667
## ud.weekday -2.1250000
## ud.weekend  2.1250000
## vl.1       -5.5000000
## vl.2        5.5000000
## ml.1       -8.7500000
## ml.2       10.6875000
## ml.3       -1.9375000
```

```
##          [,1]
## ko 0.02039660
## sw 0.03626062
## ud 0.11558074
## vl 0.29915014
## ml 0.52861190
```

Based on the result we find that
1. Knock-out stage have little impact on the views. Not knock-out stage matches are more preferable
2. Strong team win or not have little impact on the views. weak team matches are more preferable
3. Upload Date impact the views on more than 10%. Weekend uploaded video are more preferable
4. video factor impact the views on 30%. Level 2(Short video length, Short length of video title and less number of ! and ? Inf the title) are more preferable
5. match factor impact the views the most on more than 50%. Level 2: most of West team, about half of top player and most of past strong team are the most preferable

Conjoint analysis of how the factors affect and the weight of affect of the factors on the video's likes

We cheat the likes as a customers giving rating.
The views are the customer's rating, more likes per views means more prefer.
Combining the video factor and match factor level from the previous section, we have the following results.

```
##                   [,1]
## ko.No      -2.4583333
## ko.Yes      2.4583333
## sw.No      -0.1666667
## sw.Yes      0.1666667
## ud.weekday -0.6666667
## ud.weekend  0.6666667
## vl.1       -6.2916667
## vl.2        6.2916667
## ml.1        0.5000000
## ml.2       -3.5000000
## ml.3        3.0000000
```

```
##          [,1]
## ko 0.19155844
## sw 0.01298701
## ud 0.05194805
## vl 0.49025974
## ml 0.25324675
```

Based on the result we find that
1. Knock-out stage have 20% impact on the likes. knock-out stage matches gain more likes
2. Strong team win or not have little impact on the likes. strong team matches gain more likes
3. Upload Date have little impact on the likes. Weekend uploaded video gain more likes
4. video factor impact the likes the most on around 50%. Level 2(Short video length, Short length of video title and less number of ! and ? Inf the title) gain more likes
5. match factor impact the likes the most on more than 25%. Level 3(all the Asia team, less of top player and less past strong team) gain more likes

Limitations

The fitness of conjoint analysis on the factors impact on likes is not quite well. The R-squared value is only 0.2779. It implies that the model assumption may not fit the study on likes or the data collection is not enough.

The results are based on the following assumption(s):

1. The views and likes of the video will not have signified growth after it is uploaded 3 weeks
2. The number of likes and views grow in same growth rate
3. As the channel - the best league is a Chinese channel, the number Asia team is a factor affecting views and likes
4. The number of Top player data are collected by the rules: a player who had gotten any champion title is a top player However, a player may not be considered as a top player by others view by the following possible reasons:

    a) The player had the champion title a lot time ago
    b) The player is the substitute player

5. The number of Past Top team data are collected by the rules: a team who had gotten any champion title is a past top team However, a team may not be considered as a top team by others view by the following possible reasons:

    a) The team had the champion title a lot time ago
    b) The team have different players after it wins a tournament

6. The views and likes represent the rating of the viewers
7. The combined factors can represented the factors well

Conclusion

This study aims to study how the factors and what weights the factors affect the video's views and likes of a e-sport highlight YouTube channels by conjoint analysis. Based on initial guess and assumptions, some factors are considered as potential target. The sample observations data are collected from the targeted study YouTube channel, the best league based on Multi-stage sampling. After assumptions and cluster analysis, some of the factors are divided into suitable factor level for the conjoint analysis. The results of the conjoint analysis are the following:

For the improtant factors affect the likes:

Video length, length of video title and number of ! and ? marks in the title are the main factors(50%) affect the likes The viewers give more likes to short video length, short length of video title and less number of ! and ? in the title

Following by the number of Asia team, number of top player and number of past strong team(25%)
The viewers give more likes to all the Asia team, less of top player and less past strong team

Following by whether the match is knock-on stage(19%)
The viewers give more likes to the knock-on stage match video

For the improtant factors affect the views:

number of Asia team, number of top player and number of past strong team are the main factors(53%) affect the views
The views are more for most of West team, about half of top player and most of past strong team matches

Video length, length of video title and number of ! and ? marks in the title are the main factors(30%)
The views are more for Short video length, Short length of video title and less number of ! and ? in the title video

Following by whether the match is upload date(12%)
The views are more for the video uploaded on weekend

	A	B	C	D	E	F	G	H	I	J	K	L
1		ko - stage	strong win	upload date	video	match						
2	Weight on likes	19%	1%	5%	50%	25%						
3	Weight on views	2%	3%	12%	30%	53%						
4	prefer by viewers(more likes)	Yes	Yes	weekend	Short video length, Short length of video title and less number of ! and ? Inf the title	all the Asia team, less of top player and less past strong team						
5	prefer by viewers(more views)	No	No	weekend	Short video length, Short length of video title and less number of ! and ? Inf the title	most of West team, about half of top player and most of past strong team						

Figure 3: summary table of the study result

Sugguestion

In order to have a high views and likes video, upload the video on weekend, having Short video length, Short length of video title and less number of ! and ? in the title video is preferred.

By Member - Wu Tsz Lok (3035694824) - 100%

End of the Project