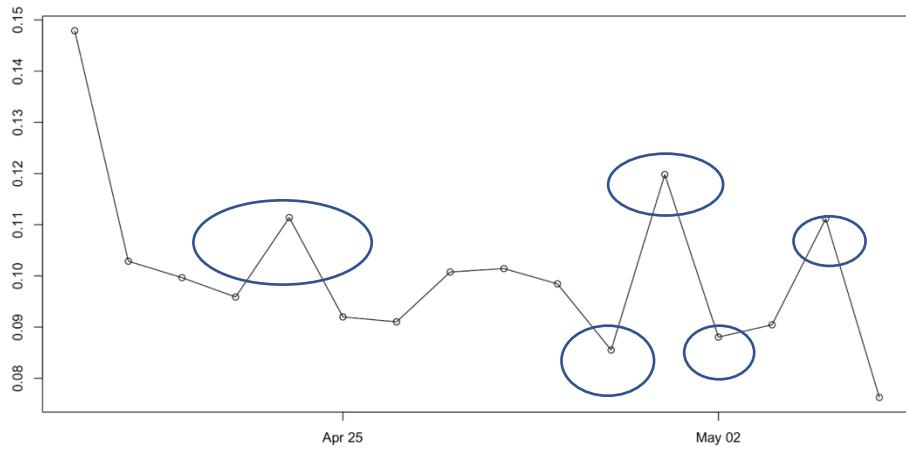


Topic Analysis and Word Cloud Visualizations



Based on graph above, we identified several dates where Twitter sentiment experienced radical changes. Then we dived into these dates to find out which topics have impacted the sentiment changes.





0423

The difference between these three wordclouds is “masterpiece” and “medium” and it positively increases the sentiment score of 0424.

```
> enn0424$masterpiece=grepl("masterpiece",enn0424$text)
> lm04241 = glm(text.sentiment0424 ~ masterpiece, data = enn0424)
> summary(lm04241)
```

Call:

```
glm(formula = text.sentiment0424 ~ masterpiece, data = enn0424)
```

Deviance Residuals:

Min 1Q Median 3Q Max
-1.26048 -0.10578 -0.01739 0.13238 1.38582

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	0.105775	0.003866	27.358	< 2e-16 ***
masterpieceTRUE	0.158872	0.023337	6.808	1.15e-11 ***

Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for gaussian family taken to be 0.05508718)

Null deviance: 211.17 on 3788 degrees of freedom
Residual deviance: 208.62 on 3787 degrees of freedom
AIC: -226.98

Number of Fisher Scoring iterations: 2

```

> enn0424$medium=grep("medium",enn0424$text)
> lm04242 = glm(text.sentiment0424 ~ medium, data = enn0424)
> summary(lm04242)

Call:
glm(formula = text.sentiment0424 ~ medium, data = enn0424)

Deviance Residuals:
    Min      1Q   Median      3Q      Max 
-1.2622 -0.1075 -0.0191  0.1421  0.9656 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept)  0.10748   0.00381  28.210 <2e-16 ***
mediumTRUE   0.41848   0.04787   8.741 <2e-16 ***  
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for gaussian family taken to be 0.05465847)

Null deviance: 211.17 on 3788 degrees of freedom
Residual deviance: 206.99 on 3787 degrees of freedom
AIC: -256.59

Number of Fisher Scoring iterations: 2

```



0429

0430 (low sentiment score) “fight”



0501(high sentiment score) enjoy/finally / performance/ please



0502(low sentiment score)

0430

Keyword "fight" may negatively affect the sentiment score of 4.30.

```
> enn0430$fight=grepl("fight",enn0430$text)
> lm04302 = glm(text.sentiment0430 ~ fight, data = enn0430)
> summary(lm04302)
```

Call:

```
glm(formula = text.sentiment0430 ~ fight, data = enn0430)
```

Deviance Residuals:

Min	1Q	Median	3Q	Max
-0.89372	-0.10071	0.00422	0.13343	1.14549

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	0.09335	0.00900	10.373	< 2e-16 ***
fightTRUE	-0.23787	0.05407	-4.399	1.24e-05 ***

Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for gaussian family taken to be 0.05969759)

Null deviance: 46.287 on 757 degrees of freedom
Residual deviance: 45.131 on 756 degrees of freedom
AIC: 18.713

Number of Fisher Scoring iterations: 2

0501

Keywords “enjoy” , “finally” “please” and “performance” may positively affect the sentiment score of 0501.

```
> summary(lm05012)
```

Call:

```
glm(formula = text.sentiment0501 ~ enjoy, data = enn0501)
```

Deviance Residuals:

Min	1Q	Median	3Q	Max
-0.99284	-0.11167	-0.01402	0.11637	1.04943

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	0.111673	0.009009	12.395	< 2e-16 ***
enjoyTRUE	0.227581	0.048307	4.711	2.9e-06 ***

Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for gaussian family taken to be 0.06306683)

Null deviance: 52.042 on 804 degrees of freedom
Residual deviance: 50.643 on 803 degrees of freedom
AIC: 63.823

```

> enn0501$performance=grep("performance",enn0501$text)
> lm05013 = glm(text.sentiment0501 ~ performance, data = enn0501)
> summary(lm05013)

Call:
glm(formula = text.sentiment0501 ~ performance, data = enn0501)

Deviance Residuals:
    Min      1Q   Median      3Q      Max 
-0.99235 -0.11118 -0.00724  0.11982  0.99299 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept)  0.111181  0.008959 12.411 < 2e-16 ***
performanceTRUE 0.270727  0.050835  5.326 1.31e-07 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for gaussian family taken to be 0.06259902)

Null deviance: 52.042 on 804 degrees of freedom
Residual deviance: 50.267 on 803 degrees of freedom
AIC: 57.829

Number of Fisher Scoring iterations: 2

> enn0501$finally=grep("finally",enn0501$text)
> lm05014 = glm(text.sentiment0501 ~ finally, data = enn0501)
> summary(lm05014)

Call:
glm(formula = text.sentiment0501 ~ finally, data = enn0501)

Deviance Residuals:
    Min      1Q   Median      3Q      Max 
-0.99551 -0.11434 -0.01739  0.11502  1.04676 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept) 0.114342  0.009109 12.552 < 2e-16 ***
finallyTRUE 0.131969  0.045689  2.888  0.00398 ** 
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for gaussian family taken to be 0.06414357)

Null deviance: 52.042 on 804 degrees of freedom
Residual deviance: 51.507 on 803 degrees of freedom
AIC: 77.45

> enn0501$please=grep("please",enn0501$text)
> lm05015 = glm(text.sentiment0501 ~ please, data = enn0501)
> summary(lm05015)

Call:
glm(formula = text.sentiment0501 ~ please, data = enn0501)

Deviance Residuals:
    Min      1Q   Median      3Q      Max 
-0.99100 -0.10983 -0.01288  0.12117  1.05126 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept) 0.109835  0.008889 12.357 < 2e-16 ***
pleaseTRUE  0.327148  0.051479  6.355 3.49e-10 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for gaussian family taken to be 0.06170659)

Null deviance: 52.042 on 804 degrees of freedom
Residual deviance: 49.550 on 803 degrees of freedom
AIC: 46.27

Number of Fisher Scoring iterations: 2

```

Keywords “man” , “lion” and “million” may negatively affect the sentiment score of 5.2.

```
> enn0502$man=grep("man",enn0502$text)
> lm05022 = glm(text.sentiment0502 ~ man, data = enn0502)
> summary(lm05022)

Call:
glm(formula = text.sentiment0502 ~ man, data = enn0502)

Deviance Residuals:
    Min      1Q   Median      3Q      Max 
-0.71677 -0.09867 -0.03721  0.10934  1.03741 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept) 0.09867   0.01294   7.628 1.64e-13 ***
manTRUE     -0.04365   0.02635  -1.657  0.0984 .  
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for gaussian family taken to be 0.05321132)

Null deviance: 22.335 on 418 degrees of freedom
Residual deviance: 22.189 on 417 degrees of freedom
AIC: -36.064

> enn0502$lion=grep("lion",enn0502$text)
> lm05023 = glm(text.sentiment0502 ~ lion, data = enn0502)
> summary(lm05023)

Call:
glm(formula = text.sentiment0502 ~ lion, data = enn0502)

Deviance Residuals:
    Min      1Q   Median      3Q      Max 
-0.71363 -0.09553 -0.02872  0.10401  1.04055 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept) 0.09553   0.01138   8.395 7.33e-16 ***
lionTRUE    -0.19330   0.05823  -3.320 0.000981 *** 
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for gaussian family taken to be 0.05218251)

Null deviance: 22.335 on 418 degrees of freedom
Residual deviance: 21.760 on 417 degrees of freedom
AIC: -44.245

Number of Fisher Scoring iterations: 2

> enn0502$million=grep("million",enn0502$text)
> lm05025 = glm(text.sentiment0502 ~ million, data = enn0502)
> summary(lm05025)

Call:
glm(formula = text.sentiment0502 ~ million, data = enn0502)

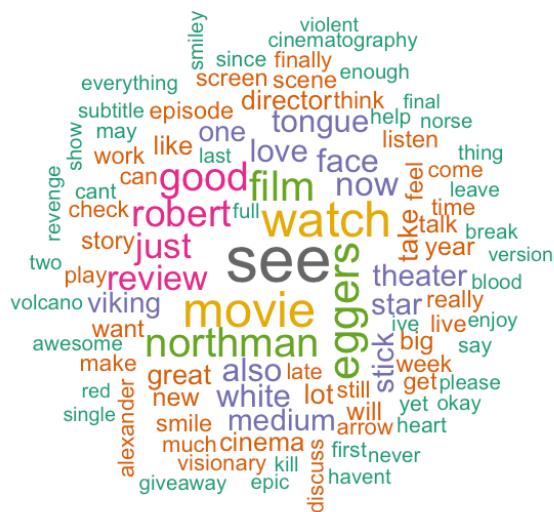
Deviance Residuals:
    Min      1Q   Median      3Q      Max 
-0.70945 -0.09135 -0.03316  0.10677  1.04472 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept) 0.09135   0.01131   8.075 7.3e-15 ***
millionTRUE -0.22368   0.09453  -2.366  0.0184 *  
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for gaussian family taken to be 0.05285191)

Null deviance: 22.335 on 418 degrees of freedom
Residual deviance: 22.039 on 417 degrees of freedom
AIC: -38.904

Number of Fisher Scoring iterations: 2
```



```
> enn0504$visionary=grepl("visionary",enn0504$text)
> lm05041 = glm(text.sentiment0504 ~ visionary, data = enn0504)
> summary(lm05041)
```

```
Call:  
glm(formula = text.sentiment0504 ~ visionary, data = enn0504)
```

Deviance Residuals:

Deviations Residuals:

Coefficients:

```

Estimate Std. Error t value Pr(>|t|)
(Intercept) 0.10871 0.01499 7.252 6.26e-12 ***
visionaryTRUE 0.18081 0.08054 2.245 0.0257 *
---
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0

```

(Dispersion parameter for gaussian family taken to be 0.05010116)

```
Null deviance: 11.726 on 230 degrees of freedom  
Residual deviance: 11.473 on 229 degrees of freedom  
AIC: -32.006
```

Number of Fisher Scoring iterations: 2

Total keyword analysis

```

> ennorthman=cbind(ennorthman,text_scoreall)
> ennorthman$director=grep("robert",ennorthman$text)
> lmall1 = glm(text.sentimentall ~ director, data = ennorthman)
> summary(lmall1)

Call:
glm(formula = text.sentimentall ~ director, data = ennorthman)

Deviance Residuals:
    Min      1Q   Median      3Q      Max 
-1.51250 -0.09980 -0.01975  0.12713  1.74156 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept) 0.094746  0.001704 55.600 <2e-16 ***
directorTRUE 0.046961  0.004916  9.552 <2e-16 ***  
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for gaussian family taken to be 0.05322776)

Null deviance: 1113.6 on 20832 degrees of freedom
Residual deviance: 1108.8 on 20831 degrees of freedom
AIC: -1981.3

Number of Fisher Scoring iterations: 2

> summary(lmall2)

Call:
glm(formula = text.sentimentall ~ director2, data = ennorthman)

Deviance Residuals:
    Min      1Q   Median      3Q      Max 
-1.51007 -0.10043 -0.02015  0.12633  1.74398 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept) 0.092319  0.001727 53.44 <2e-16 ***
director2TRUE 0.055210  0.004518 12.22 <2e-16 ***  
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for gaussian family taken to be 0.05308047)

Null deviance: 1113.6 on 20832 degrees of freedom
Residual deviance: 1105.7 on 20831 degrees of freedom
AIC: -2039.1

Number of Fisher Scoring iterations: 2

```

```

> ennorthman$story = grepl("story",ennorthman$text)
> lmall3 = glm(text.sentimentall ~ story, data = ennorthman)
> summary(lmall3)

Call:
glm(formula = text.sentimentall ~ story, data = ennorthman)

Deviance Residuals:
    Min      1Q   Median      3Q      Max 
-1.51714 -0.09939 -0.01774  0.12674  1.73691 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept) 0.099392  0.001647 60.358 < 2e-16 ***
storyTRUE   0.018470  0.007089  2.605  0.00919 **  
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for gaussian family taken to be 0.0534435)

Null deviance: 1113.6 on 20832 degrees of freedom
Residual deviance: 1113.3 on 20831 degrees of freedom
AIC: -1897.1

Number of Fisher Scoring iterations: 2

> ennorthman$northman = grepl("northman",ennorthman$text)
> lmall4 = glm(text.sentimentall ~ northman, data = ennorthman)
> summary(lmall4)

Call:
glm(formula = text.sentimentall ~ northman, data = ennorthman)

Deviance Residuals:
    Min      1Q   Median      3Q      Max 
-1.52045 -0.10270 -0.01856  0.12647  1.73361 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept) 0.102696  0.001720 59.703 < 2e-16 ***
northmanTRUE -0.017318  0.004712 -3.675 0.000238 *** 
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for gaussian family taken to be 0.05342627)

Null deviance: 1113.6 on 20832 degrees of freedom
Residual deviance: 1112.9 on 20831 degrees of freedom
AIC: -1903.8

Number of Fisher Scoring iterations: 2

```

```

> ennorthman$dollar = grepl("dollar",ennorthman$text)
> lmall5 = glm(text.sentimentall ~ dollar, data = ennorthman)
> summary(lmall5)

Call:
glm(formula = text.sentimentall ~ dollar, data = ennorthman)

Deviance Residuals:
    Min      1Q   Median      3Q      Max 
-1.51938 -0.10163 -0.01775  0.12515  1.73467 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept) 0.101629  0.001617 62.866 < 2e-16 ***
dollarTRUE -0.063656  0.011580 -5.497 3.91e-08 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for gaussian family taken to be 0.05338348)

Null deviance: 1113.6 on 20832 degrees of freedom
Residual deviance: 1112.0 on 20831 degrees of freedom
AIC: -1920.5

Number of Fisher Scoring iterations: 2

```

```
> ennorthman$bloody = grepl("bloody",ennorthman$text)
> lmall6 = glm(text.sentimentall ~ bloody, data = ennorthman)
> summary(lmall6)
```

Call:

```
glm(formula = text.sentimentall ~ bloody, data = ennorthman)
```

Deviance Residuals:

Min	1Q	Median	3Q	Max
-1.51943	-0.10168	-0.01816	0.12510	1.73462

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	0.101683	0.001613	63.030	< 2e-16 ***
bloodyTRUE	-0.081470	0.012799	-6.366	1.99e-10 ***

Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for gaussian family taken to be 0.05335712)

Null deviance: 1113.6 on 20832 degrees of freedom
Residual deviance: 1111.5 on 20831 degrees of freedom
AIC: -1930.8

Number of Fisher Scoring iterations: 2

```
> ennorthman$cinematography = grepl("cinematography ",ennorthman$text)
> lmall7 = glm(text.sentimentall ~ cinematography , data = ennorthman)
> summary(lmall7)
```

Call:

```
glm(formula = text.sentimentall ~ cinematography, data = ennorthman)
```

Deviance Residuals:

Min	1Q	Median	3Q	Max
-1.51704	-0.09929	-0.01713	0.12594	1.65265

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	0.099291	0.001611	61.631	< 2e-16 ***
cinematographyTRUE	0.084357	0.014126	5.972	2.38e-09 ***

Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for gaussian family taken to be 0.05336954)

Null deviance: 1113.6 on 20832 degrees of freedom
Residual deviance: 1111.7 on 20831 degrees of freedom
AIC: -1925.9

Number of Fisher Scoring iterations: 2

```
> ennorthman$brutal = grepl("brutal",ennorthman$text)
> lmall8 = glm(text.sentimentall ~ brutal , data = ennorthman)
> summary(lmall8)
```

Call:

```
glm(formula = text.sentimentall ~ brutal, data = ennorthman)
```

Deviance Residuals:

Min	1Q	Median	3Q	Max
-1.52103	-0.10328	-0.01778	0.12476	1.73302

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	0.103282	0.001643	62.847	< 2e-16 ***
brutalTRUE	-0.055103	0.007171	-7.684	1.61e-14 ***

Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for gaussian family taken to be 0.05330983)

Null deviance: 1113.6 on 20832 degrees of freedom
Residual deviance: 1110.5 on 20831 degrees of freedom
AIC: -1949.3

Number of Fisher Scoring iterations: 2

```
> ennorthman$scene = grepl("scene",ennorthman$text)
> lmall9 = glm(text.sentimentall ~ scene , data = ennorthman)
> summary(lmall9)
```

Call:

```
glm(formula = text.sentimentall ~ scene, data = ennorthman)
```

Deviance Residuals:

Min	1Q	Median	3Q	Max
-1.49900	-0.10100	-0.01767	0.12578	1.73530

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	0.101003	0.001627	62.069	<2e-16 ***
sceneTRUE	-0.019751	0.009227	-2.141	0.0323 *

Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for gaussian family taken to be 0.05344916)

Null deviance: 1113.6 on 20832 degrees of freedom
Residual deviance: 1113.4 on 20831 degrees of freedom
AIC: -1894.9

Number of Fisher Scoring iterations: 2

```
> ennorthman$revenge = grepl("revenge",ennorthman$text)
> lmall10 = glm(text.sentimentall ~ revenge , data = ennorthman)
> summary(lmall10)
```

Call:

```
glm(formula = text.sentimentall ~ revenge, data = ennorthman)
```

Deviance Residuals:

Min	1Q	Median	3Q	Max
-1.52290	-0.10515	-0.01765	0.12481	1.73116

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	0.105148	0.001633	64.40	<2e-16 ***
revengeTRUE	-0.103832	0.007626	-13.62	<2e-16 ***

Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for gaussian family taken to be 0.05298931)

Null deviance: 1113.6 on 20832 degrees of freedom
Residual deviance: 1103.8 on 20831 degrees of freedom
AIC: -2074.9

Number of Fisher Scoring iterations: 2