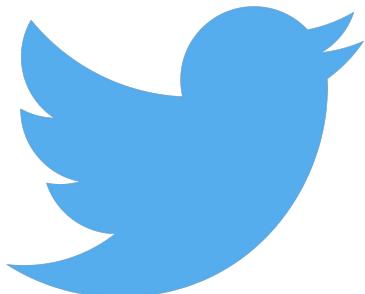


Tweet Learning



November 19, 2018



Context



OVERWATCH®

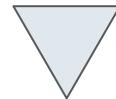


**Confirm expected outcomes
and discover other insights on
specific tweets**

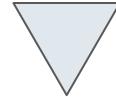


Approach

Collect Data



Pre-Process / Cluster / Model

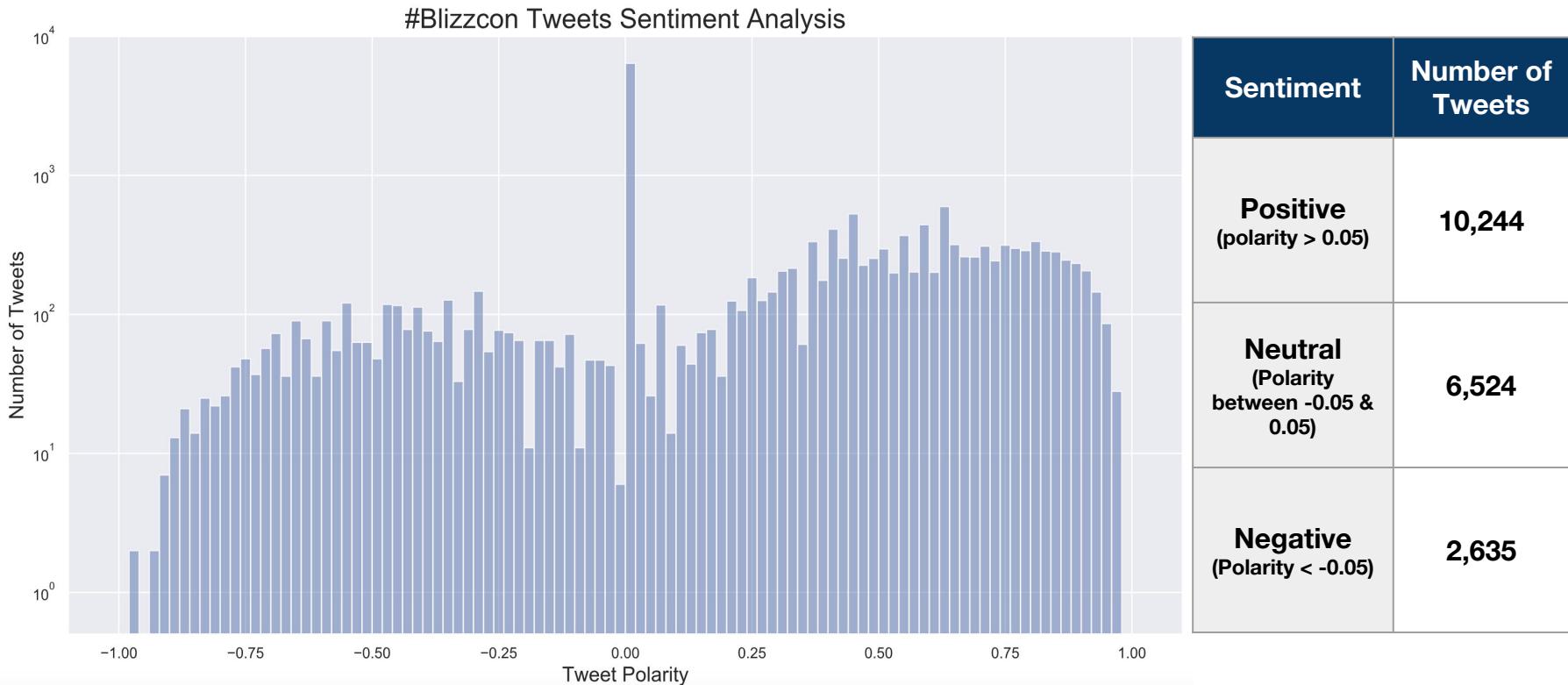


Visualize Outcomes





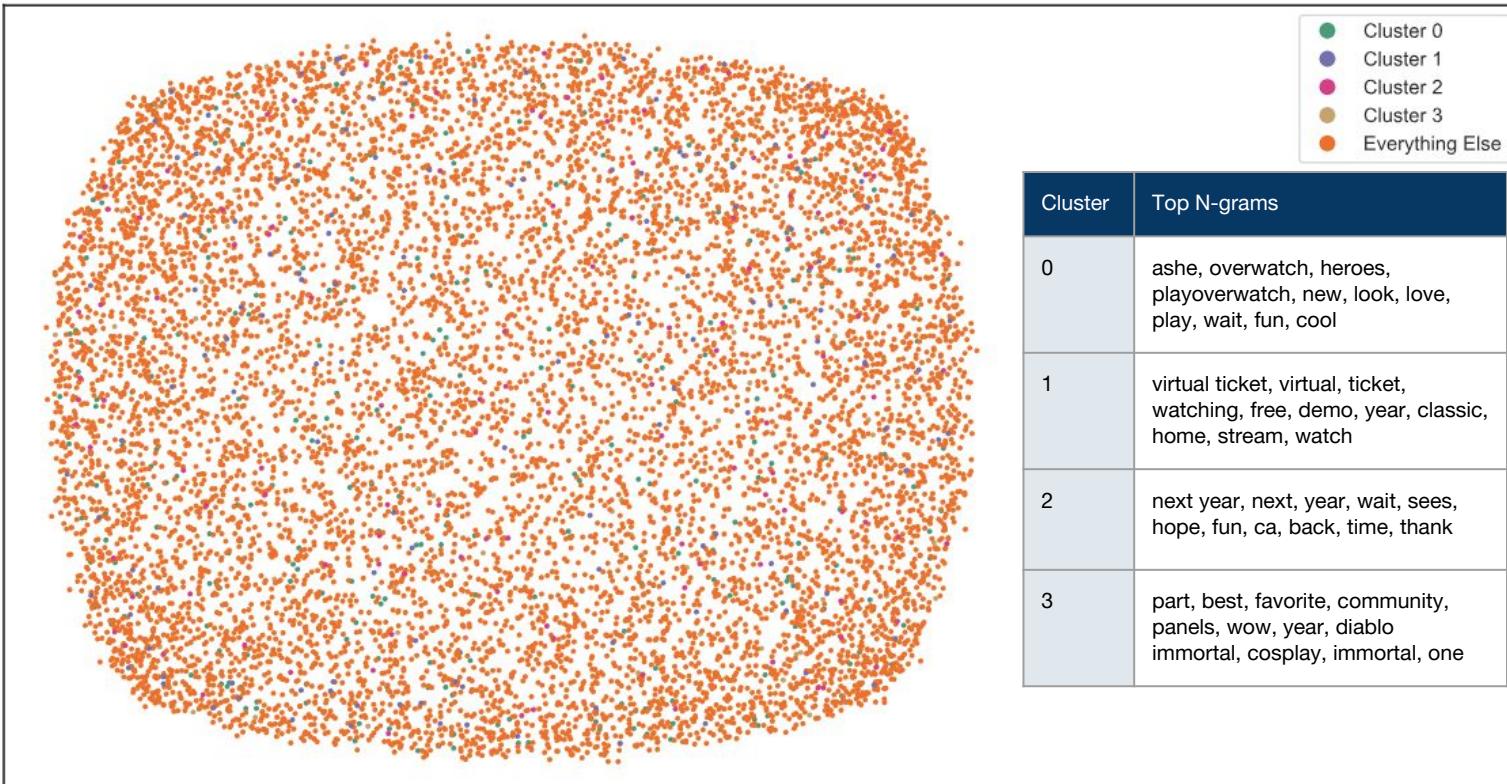
Sentiment Analysis





K-means Clustering (5 centroids)

MDS Representation of Positive Sentiment tweets

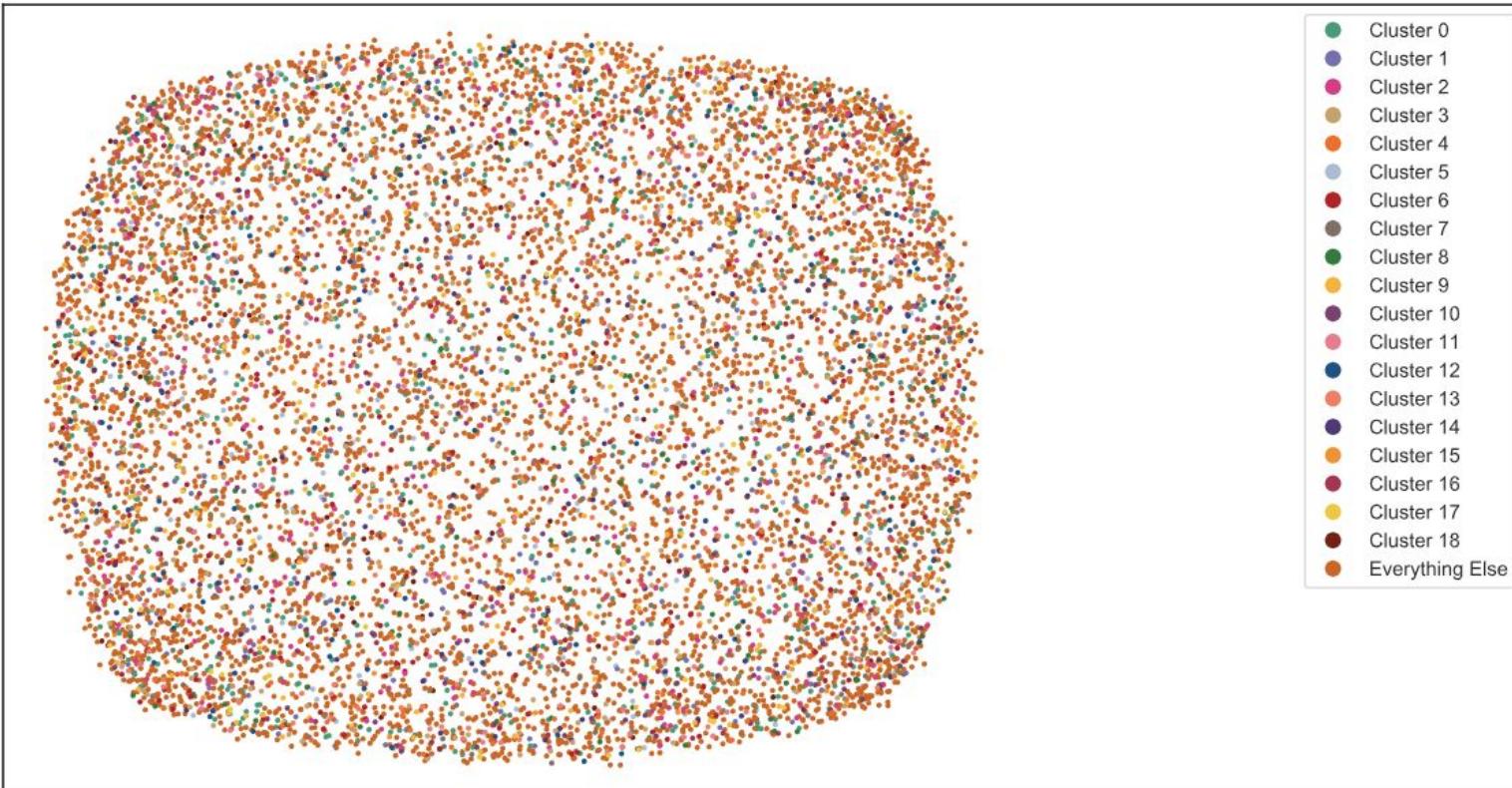


** MDS = Multidimensional Scaling



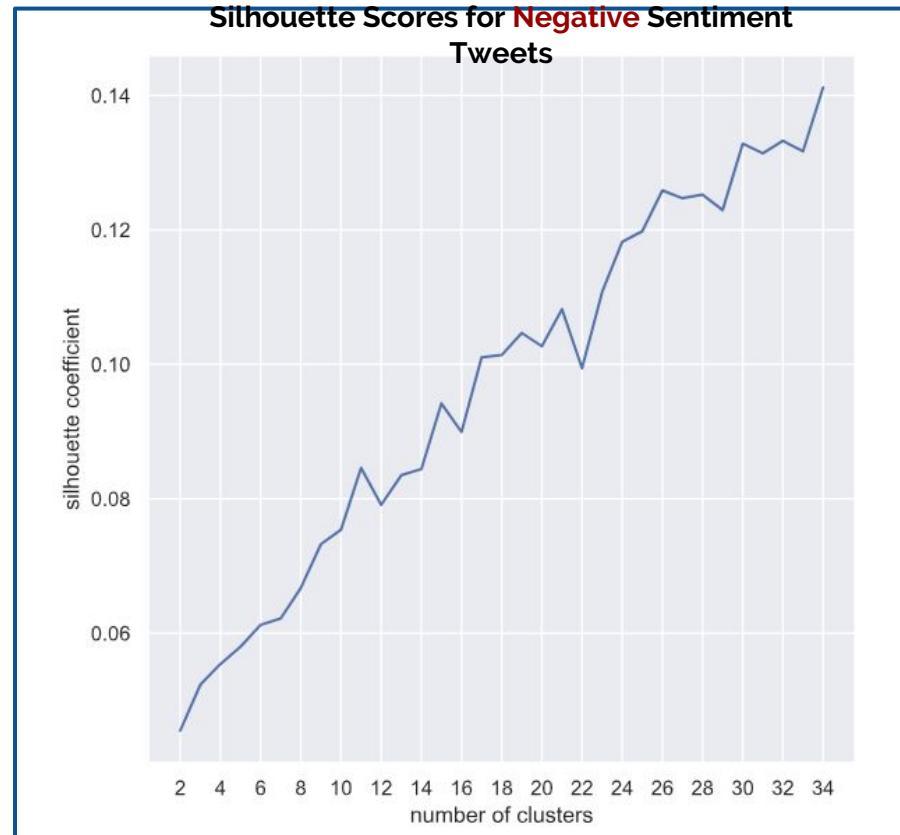
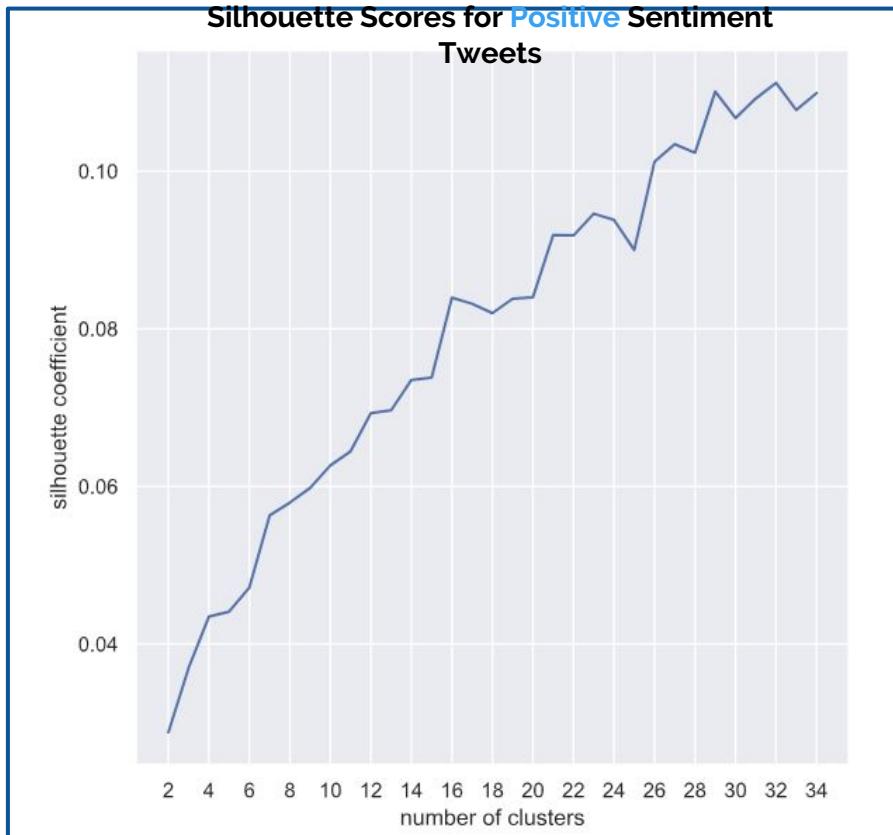
K-means Clustering (20 centroids)

MDS Representation of Positive Sentiment tweets





K-means Clustering Optimization





Latent Dirichlet Allocation (LDA)

Positive Sentiment Tweets Results

Cluster	Associated Text	Topic
1	ready, sees, time, opening, come, excited, ceremony, say	Opening Ceremony
2	overwatch, good, ashe, new, heroes, excited, going, ready	Overwatch - New Content
3	world, wcs, finals, cup, get, starcraft, serral, win	Starcraft Competition

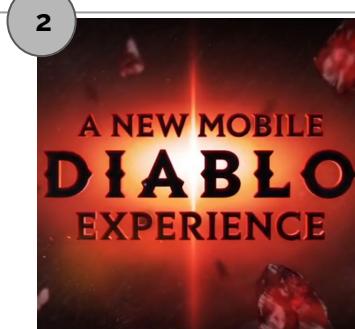




Latent Dirichlet Allocation (LDA)

Negative Sentiment Tweets Results

Cluster	Associated Text	Topic
1	game, diablo, mobile, blizzard, ass, dead, was, mobile game	Diablo mobile game
2	crying, diablo, damn, was, sad, live, blizzconreveals, war	Diablo mobile game
3	diablo, mobile, tonight, stream, diablo mobile, ass, blizzard, live	Diablo mobile game





Results Do Not Lie...

0: [0.0062*'diablo' + 0.0051*'game' + 0.0047*'mobile' + 0.0042*'ashe' + 0.0039*'blizzard' + 0.0038*'overwatch' + 0.0030*'mobile game' + 0.0029*'ass' + 0.0028*'want' + 0.0026*'flight' + 0.0023*'morning' + 0.0021*'line' + 0.0019*'fuck' + 0.0019*'day' + 0.0019*'was' + 0.0018*'dead' + 0.0018*'red' + 0.0018*'hours' + 0.0018*'bob' + 0.0016*'cut'

(Probability sum of displayed n-grams)/(total probability) = 0.058031/1.000000
Number of documents in plot = 312

1: 0.0037*[crying] + 0.0030*fuck' + 0.0025*'diablo' + 0.0023*'damn' + 0.0022*'was]' + 0.0022*'sad' + 0.0019*'live' + 0.0019*'blizzard' + 0.0018*'like' + 0.0018*'ways' + 0.0018*'get' + 0.0018*'crying' + 0.0017*'game' + 0.0016*'around' + 0.0015*'overwatch' + 0.0014*'bed' + 0.0014*'mobile' + 0.0014*'blizzconreveals' + 0.0014*'pc' + 0.0013*'war'

(Probability sum of displayed n-grams)/(total probability) = 0.038672/1.000000
Number of documents in plot = 301

2: 0.0077*'diablo' + 0.0041*'hearthstone' + 0.0032*'mobile' + 0.0030*'tonight' + 0.0029*'stream' + 0.0027*'blizzard' + 0.0023*'diablo mobile' + 0.0022*'two' + 0.0020*'ass]' + 0.0020*'remaster' + 0.0019*'blizzard ent' + 0.0019*'ent' + 0.0018*'idea' + 0.0017*'right' + 0.0017*'everything' + 0.0017*'fucking' + 0.0017*'night' + 0.0016*'annoucement' + 0.0016*'ass' + 0.0016*'live'

(Probability sum of displayed n-grams)/(total probability) = 0.049437/1.000000
Number of documents in plot = 289

3: 0.0034*'blizzard' + 0.0033*'overwatch' + 0.0029*'world' + 0.0023*'started' + 0.0022*'wait' + 0.0021*'see' + 0.0019*'weird' + 0.0019*'people' + 0.0019*'jealous' + 0.0018*'anaheim' + 0.0018*'day' + 0.0017*'uk' + 0.0017*'get' + 0.0016*'cup' + 0.0016*'though' + 0.0016*'battle' + 0.0015*'game' + 0.0015*'owwc' + 0.0015*'blizzard ent' + 0.0015*'ent'

(Probability sum of displayed n-grams)/(total probability) = 0.039925/1.000000
Number of documents in plot = 285

4: 0.0041*'blizzard' + 0.0037*'come' - 0.0029*'diablo' + 0.0027*'warcraft' + 0.0024*'one' + 0.0021*'day' + 0.0020*'reforged' + 0.0019*'destiny' + 0.0018*'overwatch' + 0.0018*'was]' + 0.0018*'convention' + 0.0016*'still' + 0.0016*'amp' + 0.0016*'ugh' + 0.0016*'oh' + 0.0015*'coming' + 0.0015*'convention center' + 0.0015*'center' + 0.0015*'jeff' + 0.0014*'dead'

(Probability sum of displayed n-grams)/(total probability) = 0.040871/1.000000
Number of documents in plot = 279

5: 0.0034*'diablo' + 0.0030*'tomorrow' + 0.0030*'time' + 0.0028*'thing' + 0.0024*'year' + 0.0024*'every' + 0.0023*'see' + 0.0022*'today' + 0.0021*'kill' + 0.0021*'people' + 0.0021*'get' + 0.0019*'ass' + 0.0019*'day' + 0.0018*'mythic' + 0.0018*'new' + 0.0018*'going' + 0.0018*'was]' + 0.0018*'bad' + 0.0017*'wait' + 0.0017*'us'

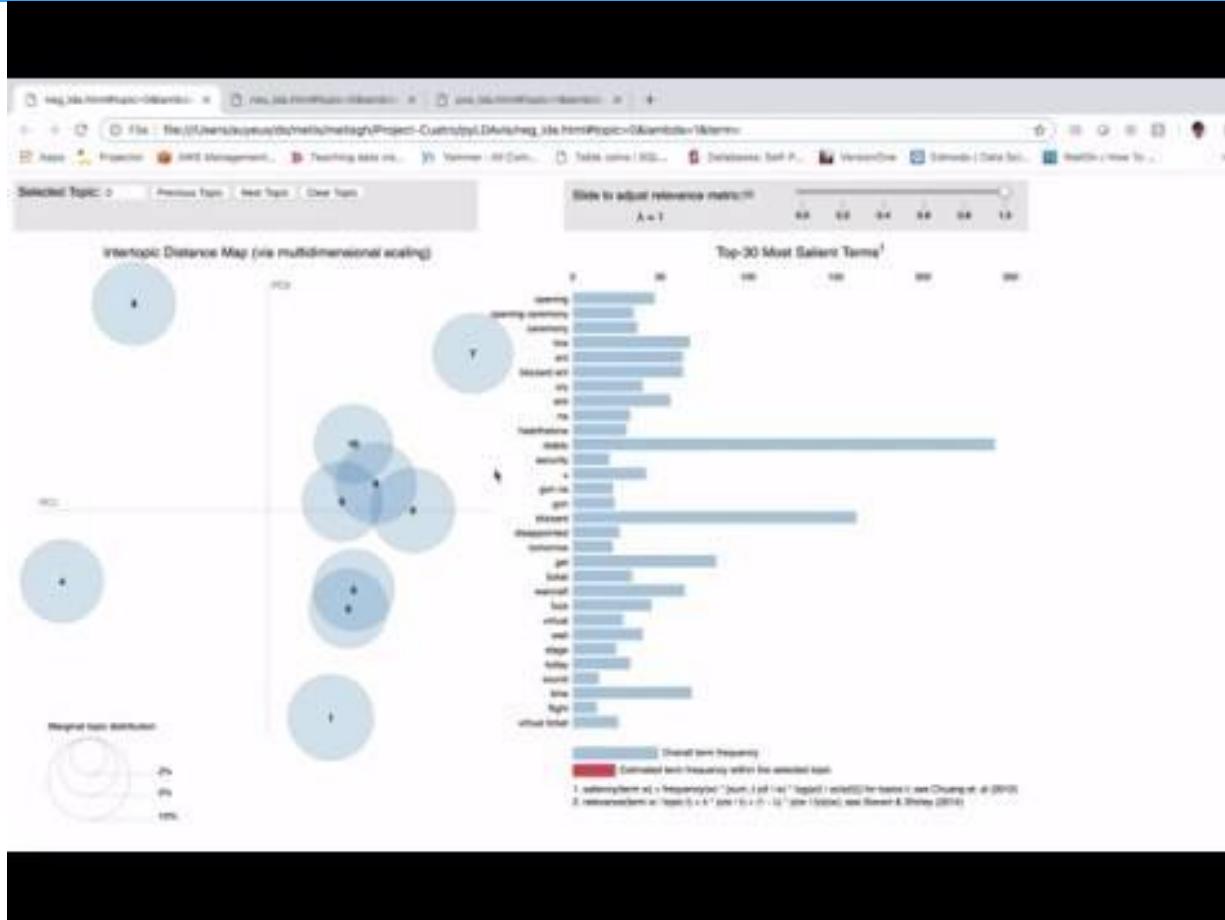
(Probability sum of displayed n-grams)/(total probability) = 0.044072/1.000000
Number of documents in plot = 271

The word '**diablo**' represents **4 out of the top 5** represented topics for negative sentiment tweets





pyLDAvis - Data Visualization for LDA





Conclusion

- LDA is very capable with capturing social trends from Twitter data
- Texts mapped to each K-means cluster or LDA topic can provide valuable insights on trends



Future Considerations

- DBSCAN
- Word2Vec
- Flask App



***“Imagination is the
key to discovery”***

Winston



THANKS!

APPENDIX



Text Pre-processing

Original Tweet:

“Can we all go back to #Blizzcon with more time, more food, more drinks and less waiting for drinks? Yea, I'd like that.”

After tokenization:

```
[ 'can', 'we', 'all', 'go', 'back', 'to', 'blizzcon', 'with', 'more', 'time',
'more', 'food', 'more', 'drinks', 'and', 'less', 'waiting', 'for', 'drinks',
'yea', 'i', "'d", 'like', 'that' ]
```

After lemmatization:

```
[ 'can', 'we', 'all', 'go', 'back', 'to', 'blizzcon', 'with', 'more', 'time',
'more', 'food', 'more', 'drink', 'and', 'le', 'waiting', 'for', 'drink',
'yea', 'i', "'d", 'like', 'that' ]
```

After removing stop words:

```
[ 'go', 'back', 'time', 'food', 'drink', 'le', 'waiting', 'drink', 'yea',
'like' ]
```

Vectorization / TF-IDF (Positive Tweets)

Number of documents = 10244, Number of TF-IDF features = 139

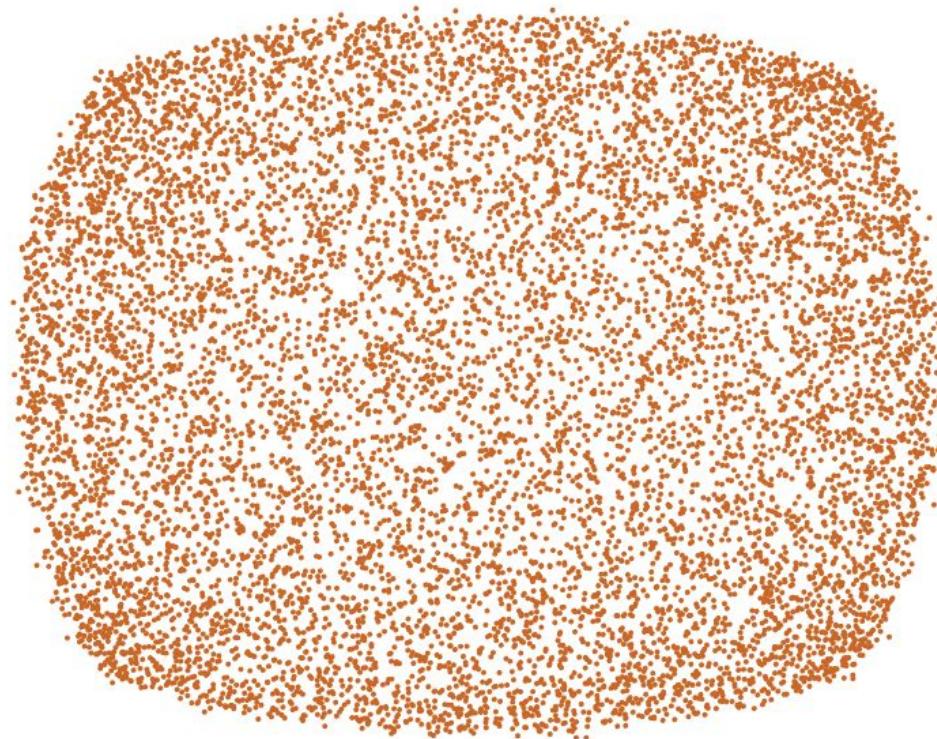
TF-IDF features, with scores:

0. 'll:	5.48	1. also:	5.49	2. always:	5.29	3. amazing:	3.91
4. amp:	4.73	5. anaheim:	5.15	6. announcement:	5.18	7. another:	5.32
8. anyone:	5.57	9. arena:	5.57	10. ashe:	4.45	11. awesome:	4.34
12. back:	4.75	13. best:	4.38	14. better:	5.55	15. blizzard:	3.42
16. blizzard ent:	4.28	17. blizzcon...:	5.30	18. ca:	5.20	19. ceremony:	5.06
20. classic:	4.90	21. come:	4.72	22. community:	5.31	23. cool:	4.91
24. cosplay:	4.79	25. could:	5.48	26. day:	4.09	27. demo:	5.37
28. diablo:	3.65	29. diablo immortal:	4.71	30. diabloimmortal:	5.54	31. ent:	4.28
32. even:	5.37	33. ever:	5.34	34. everyone:	4.48	35. excited:	4.35
36. fan:	5.32	37. favorite:	5.05	38. feel:	5.36	39. final:	4.99
40. finally:	5.34	41. first:	4.31	42. free:	5.12	43. friend:	4.47
44. fun:	4.36	45. game:	4.10	46. get:	4.14	47. getting:	5.12
48. go:	4.73	49. going:	4.61	50. good:	4.15	51. got:	4.18
52. great:	4.12	53. guy:	4.85	54. happy:	4.68	55. hearthstone:	5.24
56. hero:	4.27	57. home:	4.98	58. hope:	4.68	59. immortal:	4.66
60. know:	4.98	61. last:	5.06	62. let:	4.97	63. like:	4.03
64. line:	5.37	65. live:	4.81	66. look:	4.45	67. looking:	5.45
68. love:	4.00	69. made:	4.95	70. make:	4.77	71. many:	5.17
72. meet:	4.82	73. met:	5.39	74. mobile:	4.94	75. much:	4.59
76. na:	5.24	77. need:	5.22	78. new:	4.00	79. next:	4.64
80. next year:	5.45	81. night:	5.29	82. oh:	5.37	83. one:	4.22
84. opening:	5.06	85. opening ceremony:	5.24	86. overwatch:	3.70	87. owc2018:	5.23
88. panel:	4.92	89. part:	5.58	90. party:	5.08	91. people:	4.62
92. play:	4.48	93. playing:	5.08	94. playoverwatch:	4.93	95. please:	5.09
96. pretty:	5.12	97. ready:	4.52	98. really:	4.74	99. right:	5.26
100. say:	5.05	101. see:	3.97	102. seeing:	5.48	103. serral:	4.94
104. stage:	5.48	105. starcraft:	5.20	106. still:	5.08	107. stream:	5.41
108. super:	5.18	109. sure:	5.42	110. take:	5.38	111. team:	4.87
112. thank:	4.36	113. thanks:	4.50	114. thing:	5.03	115. think:	5.15
116. ticket:	5.18	117. time:	3.79	118. today:	4.34	119. tomorrow:	5.35
120. tonight:	5.53	121. u:	4.85	122. virtual:	5.26	123. virtual ticket:	5.35
124. wait:	4.93	125. want:	4.54	126. warcraft:	4.09	127. watch:	4.98
128. watching:	4.82	129. way:	5.11	130. wcs:	4.92	131. weekend:	4.86
132. well:	4.78	133. win:	5.15	134. work:	5.31	135. world:	4.70
136. would:	5.34	137. wow:	4.04	138. year:	3.80		



Multidimensional Scaling (MDS)

MDS Representation of Positive Sentiment tweets



TF-IDF Features

139
↓
2

GRAVEYARD

Latent Dirichlet Allocation (LDA)

Positive Sentiment Tweets Results

Cluster	Associated Text	Topic
1	ready, sees, time, opening, come, excited, ceremony, say	Opening Ceremony
2	overwatch, good, Ashe, new, heroes, excited, going, ready	Overwatch - New Content
3	world, wcs, finals, cup, get, starcraft, serral, win	Starcraft Competition

Negative Sentiment Tweets Results

Cluster	Associated Text	Topic
1	game, diablo, mobile, blizzard, ass, dead, was, mobile game	Diablo mobile game
2	crying, diablo, damn, was, sad, live, blizzconreveals, war	Diablo mobile game
3	diablo, mobile, tonight, stream, diablo mobile, ass, blizzard, live	Diablo mobile game