Same Mickey Different Cultures

Twitter Sentiment Analysis of Disney Parks

Justin Huang

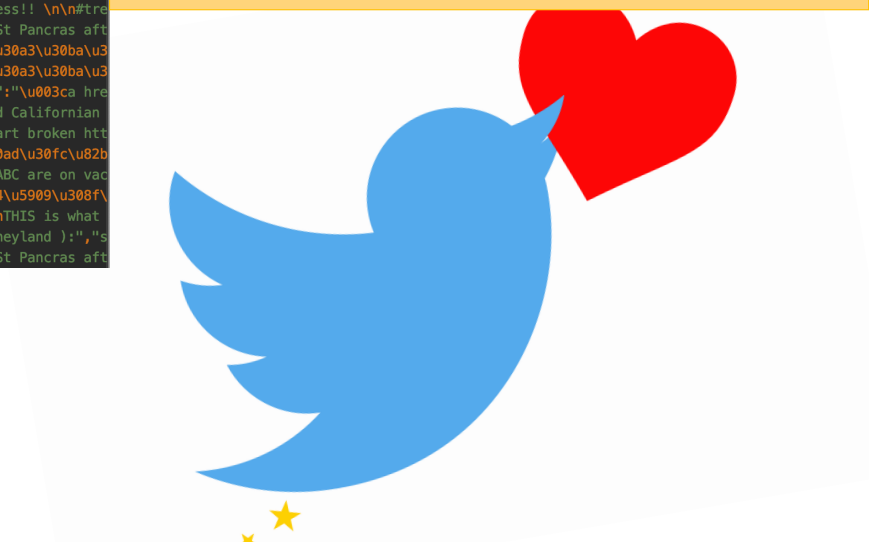
Intro

- Why Sentiment Analysis
  - By seeing trends of what people liked and disliked it can help relay information to Disney's business partners of what worked and didn't work.
- Imagineers
  - Imagineers are in charge of dreaming, designing and building Disney theme parks, attractions, cruise ships, resorts etc…
- Same name different Ownership
  - Oriental Lands
    - A Japanese leisure and tourism corporation headquartered in Urayasu, Chiba, Japan.
      - Tokyo Disneyland
      - Tokyo Disney Sea
  - Disney
    - Owners of several renown and loved character IPs
      - Disneyland
      - Disney California Adventure

Twitter API

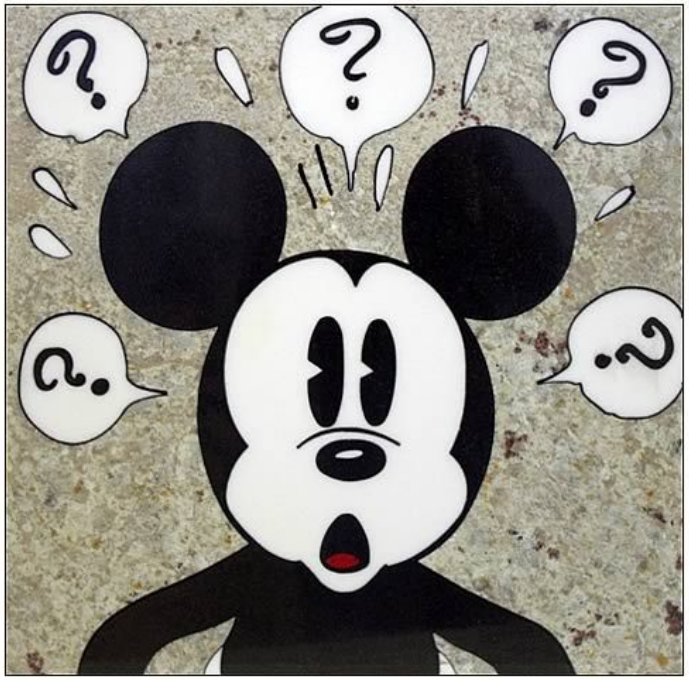

```
{"created_at":"Tue Sep 24 03:24:13 +0000 2019","id":1176336509575712768,"id_str":"1176336509575712768","text":"RT @tdrblognote: \u3053\u308c\u304c\u65b0\u3057\u
{"created_at":"Tue Sep 24 03:24:16 +0000 2019","id":1176336519969345536,"id_str":"1176336519969345536","text":"@OrangeGrove55 Galaxy\u2019s Edge at Walt Disney
{"created_at":"Tue Sep 24 03:24:19 +0000 2019","id":1176336531339919360,"id_str":"1176336531339919360","text":"RT @hp_yec: \u6628\u65e5\u3068\u306a\u3093\u304b\
{"created_at":"Tue Sep 24 03:24:25 +0000 2019","id":1176336557969567744,"id_str":"1176336557969567744","text":"RT @gourmetdyy: New #Halloween themed cups! #Shan
{"created_at":"Tue Sep 24 03:24:28 +0000 2019","id":1176336572305874944,"id_str":"1176336572305874944","text":"Why would Disney need to promote #StarWars: #Gala
{"created_at":"Tue Sep 24 03:24:44 +0000 2019","id":1176336637992681472,"id_str":"1176336637992681472","text":"RT @tdrblognote: \u3053\u308c\u304c\u65b0\u3057\u
{"created_at":"Tue Sep 24 03:24:47 +0000 2019","id":1176336652878303232,"id_str":"1176336652878303232","text":"RT @MezzoMikiD: \u6771\u4eac\u30c7\u30a3\u30ba\u3
{"created_at":"Tue Sep 24 03:24:49 +0000 2019","id":1176336658029002752,"id_str":"1176336658029002752","text":"You could win a vacation for four to a galaxy far
{"created_at":"Tue Sep 24 03:24:52 +0000 2019","id":1176336672872517633,"id_str":"1176336672872517633","text":"RT @_erikagirl00: a bitch just wants to go to Dis
{"created_at":"Tue Sep 24 03:24:54 +0000 2019","id":1176336681152208896,"id_str":"1176336681152208896","text":"RT @EricMMrozek: Why would Disney need to promote
{"created_at":"Tue Sep 24 03:24:58 +0000 2019","id":1176336698705186816,"id_str":"1176336698705186816","text":"RT @pinku_tai: \u0e01\u0e23\u0e35\u0e4a\u0e14\u0e
{"created_at":"Tue Sep 24 03:25:00 +0000 2019","id":1176336706124963840,"id_str":"1176336706124963840","text":"To Popcorn Bucket or not to Popcorn Bucket? That
{"created_at":"Tue Sep 24 03:25:04 +0000 2019","id":1176336721442566146,"id_str":"1176336721442566146","text":"RT @kelseyr1101: My diet consists of Starbucks an
{"created_at":"Tue Sep 24 03:25:06 +0000 2019","id":1176336732628762624,"id_str":"1176336732628762624","text":"RT @gourmetdyy: New #Halloween popcorn bucket~\n#
{"created_at":"Tue Sep 24 03:25:09 +0000 2019","id":1176336741344526336,"id_str":"1176336741344526336","text":"RT @MezzoMikiD: \u6771\u4eac\u30c7\u30a3\u30ba\u3
{"created_at":"Tue Sep 24 03:25:17 +0000 2019","id":1176336777348411392,"id_str":"1176336777348411392","text":"RT @MezzoMikiD: \u6771\u4eac\u30c7\u30a3\u30ba\u3
{"created_at":"Tue Sep 24 03:25:20 +0000 2019","id":1176336790166355968,"id_str":"1176336790166355968","text":"Late night sugar rush with my princess!! \n\n#tre
{"created_at":"Tue Sep 24 03:25:21 +0000 2019","id":1176336792372596737,"id_str":"1176336792372596737","text":"RT @KatyFBrand: Passport checks at St Pancras aft
{"created_at":"Tue Sep 24 03:25:22 +0000 2019","id":1176336797741117440,"id_str":"1176336797741117440","text":"RT @MezzoMikiD: \u6771\u4eac\u30c7\u30a3\u30ba\u3
{"created_at":"Tue Sep 24 03:25:24 +0000 2019","id":1176336805685161984,"id_str":"1176336805685161984","text":"RT @MezzoMikiD: \u6771\u4eac\u30c7\u30a3\u30ba\u3
{"created_at":"Tue Sep 24 03:25:32 +0000 2019","id":1176336838992089088,"id_str":"1176336838992089088","text":"i want to go to disneyland","source":"\u003ca hre
{"created_at":"Tue Sep 24 03:25:34 +0000 2019","id":1176336848899076098,"id_str":"1176336848899076098","text":"RT @DisneyParks: Disney\u2019s Grand Californian
{"created_at":"Tue Sep 24 03:25:39 +0000 2019","id":1176336867232538624,"id_str":"1176336867232538624","text":"RT @fairejk: im ready to have my heart broken htt
{"created_at":"Tue Sep 24 03:25:40 +0000 2019","id":1176336875209904128,"id_str":"1176336875209904128","text":"RT @aochann_disney: \u30df\u30c3\u30ad\u30fc\u82b
{"created_at":"Tue Sep 24 03:25:49 +0000 2019","id":1176336909762727937,"id_str":"1176336909762727937","text":"RT @DisneylandToday: .@TheGoldbergsABC are on vac
{"created_at":"Tue Sep 24 03:25:54 +0000 2019","id":1176336930838958081,"id_str":"1176336930838958081","text":"\u3042\u308c\u3048\uff01\uff01\u9854\u5909\u308f\
{"created_at":"Tue Sep 24 03:26:00 +0000 2019","id":1176336959075020800,"id_str":"1176336959075020800","text":"RT @Cartarsauce: Since you asked, \nTHIS is what
{"created_at":"Tue Sep 24 03:26:03 +0000 2019","id":1176336967794978817,"id_str":"1176336967794978817","text":"RT @lorrrrenaa: I NEED to go to Disneyland ):","s
{"created_at":"Tue Sep 24 03:26:04 +0000 2019","id":1176336972224311298,"id_str":"1176336972224311298","text":"RT @KatyFBrand: Passport checks at St Pancras aft
```


- Why Twitter
  - Twitter is popular in Japan and the US
- Twitter Data collected
  - Japanese and English Tweets
  - 270000 tweets for Halloween Event
  - 26000 for Chinese New Year event
  - Live Stream Tweets collected



Data Wrangling and cleaning

- Collected Disney tweets based off Anaheim and Tokyo Disney researched words
- Json file was structured using tweepy.
- Had to find Text through 4 deeply nested dictionaries
  - Extended Tweets
  - Quoted Status
  - Retweeted Status
  - Extended Entities
- Columns were formatted for Datetime and categorical
- Longitude and Latitude added back in using Geocoder
- Used office 365 translation to translate Japanese Tweets to English
- Then preprocessed the text
- Got the hashtags, mentions, links and emojis out



Exploratory Data Analysis

Most retweeted during Halloween collection period

retweetID	original_text	retweetID
1,187,535,711,000,000,000.000	RT @lobo240z: Fuck a fake ID, who makes fake Disneyland annual passes	14445
1,158,839,596,000,000,000.000	“Quick, act like this is your first tasting of the day” #🧐 #weekend #winetasting #jumpsuit #mrturk #californiaadventure #clothedcontent https://t.co/NbpoKzBuiU	7517
1,188,852,344,000,000,000.000	RT @TheSocietyDude: This guy is ready for whatever Disneyland will throw at him https://t.co/ZawaczjFax	6652
1,185,059,536,000,000,000.000	RT @babyhanners: I wish I was at Disneyland eating a churro right now	5140
1,190,101,152,000,000,000.000	✨✨✨✨✨✨✨\n　　5組10名様を\n#東京ディズニーランド®\n　新オープン施設の\n プレビューにご招待！\n✨✨✨✨✨✨✨\n\n新施設をいち早く体験するチャンス✧٩(ˊωˋ*)و✧\n\nこのツイートをリツイートして応募してね！\n\n#パークチケットもついてるよ🎁\n#JALプレビューチケットキャンペーン\n\n▼詳細▼	3135
1,184,630,277,000,000,000.000	People kept stopping me all day at Tokyo Disneyland with lovely compliments and asking for pictures! There's no face character for Tiana in the Japan parks so people really responded well to seeing one at TDL. More love for Princess Tiana in Japan yes 🤩💚 https://t.co/wvEC7hVwBg	3100
1,179,510,293,000,000,000.000	RT @iamAdrianRomero: It's always “wyd”, never, “lets go to Disneyland”.	3100
1,183,939,195,000,000,000.000	RT @Dream_GoesOn: ワンマン初回、今日から新演出…？ これは泣くよ…😭\n#TDR_now https://t.co/tHqRj5fhgs	1843
1,185,941,050,000,000,000.000	Yesterday I went into the Cantina at the new Star Wars land at Disneyland and the bartender leans over, sneakily gives me this, and whispers, “It's not often we get a Resistance General in here. Thank you for your service.” I almost cried. https://t.co/JM2YpRo2aF	1814
1,189,617,380,000,000,000.000	RT @christinekossa: Disneyland Paris with my hubby❤️💍🏰 https://t.co/QqoYb32mmi	1590
1,176,213,229,000,000,000.000	Passport checks at St Pancras after getting off the train from Paris tonight. First time I've seen this. A lot of ashen faces, including mine, seeing the reality kick in. Long queues. Young tired children wanting to get home after a trip to Disneyland. Welcome to Brexitland.	1396
1,189,573,144,000,000,000.000	The Chosen One himself, Hayden Christensen, had a most impressive visit to Star Wars: #GalaxysEdge at @Disneyland. The Force is strong with this one! https://t.co/P7BWi1vw33	1331
1,181,653,385,000,000,000.000	Taemin fansign highlights:\n\n- my name is Luna ➡ TM: are you a member of f (x)\n- i work in Disneyland ➡ TM: can i get in for free?\n- My name is Genesis ➡ TM: are you a hundai car? \n- Your English is so good ➡ TM: don't ..... lie https://t.co/rRsklsq5Vk	1319
1,189,616,281,000,000,000.000	RT @Genius: this guy made a beat out of sounds he recorded at disneyland and it's 🔥🔥\nhttps://t.co/bvs14ps9qM	1264
1,181,277,651,000,000,000.000	RT @intrnetdaughter: adults who are super into Disneyland are missing a part of their brain	1176
1,178,491,779,000,000,000.000	RT @deaddilf69: How is any of these pictures real https://t.co/fpCTO4AwBw	1044
1,178,801,532,000,000,000.000	RT @MezzoMikiD: 東京ディズニーランド\n本日10月1日よりシンデレラ城および城内は改修工事に伴い通行できなくなっています。\nhttps://t.co/6tEyanANLs https://t.co/qgp4JfaIxU	972
1,189,319,906,000,000,000.000	RT @dreamingup39: ドン引きなんだけど、年パス組にはこういう人もいるのよ\nダッフィー 毛布12500円で買う人も世の中いるから、転売はなくならない #TDR_now https://t.co/Mtr9iBNGgd	876
1,178,884,070,000,000,000.000	RT @markhoppus: I was trying to tell a story about the Disneyland ride “It's a Small World” and accidentally called it “All the Small World…	868
1,187,574,557,000,000,000.000	RT @Gg5w3f52ZhsV2aS: おじさん家\n#tdr_now https://t.co/eZaK9sOp2N	845

Exploratory Data Analysis

English Tweets

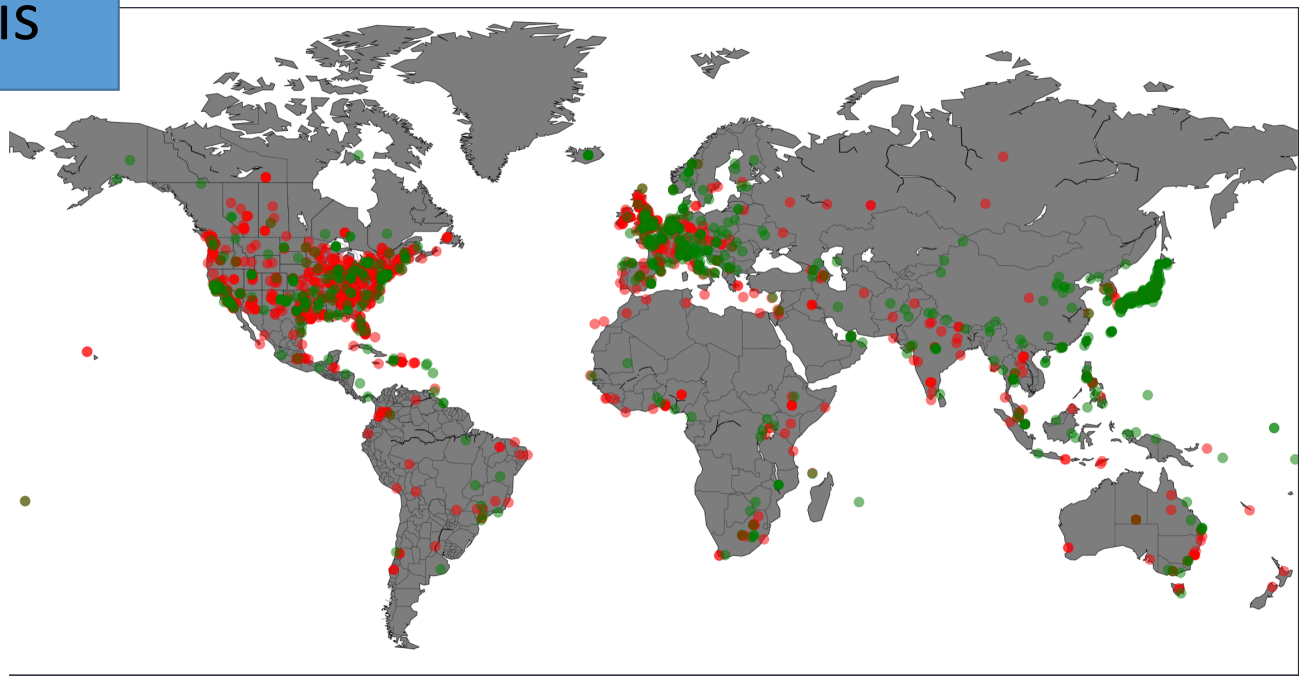
- Fake Annual Passes
- Quick
- First Tasting
- Stopping Lovely

Japanese Translated Tweets

- Experience
- Facility
- Talking about the last shows of One Mans Dream
- Cinderella's Castle
- Food

Exploratory Data Analysis

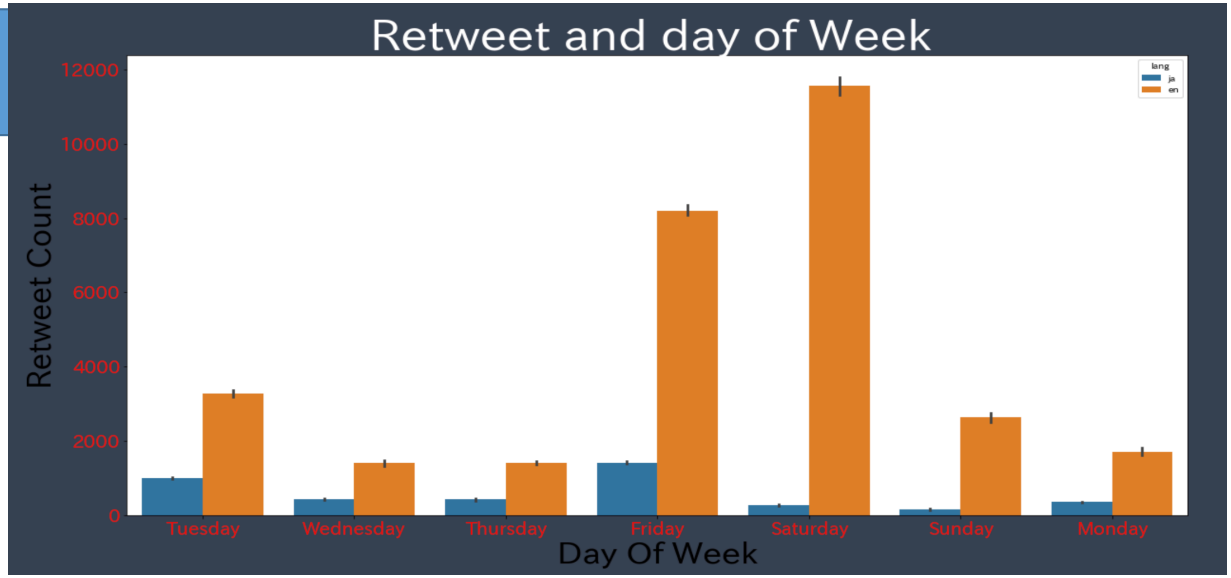
```
# Geocoder api
from opencage.geocoder import OpenCageGeocode
from pprint import pprint
key = # get api key from:  https://opencagedata.com
geocoder = OpenCageGeocode(key)


en_lat = []   # create empty lists
en_long = []

for x in en_loc3000.user_location: # iterate over rows in dataframe
    try:
        query = x
        results = geocoder.geocode(query)
        lat = results[0]['geometry']['lat']
        long = results[0]['geometry']['lng']
        en_lat.append(lat)
        en_long.append(long)
    except IndexError:
        en_lat.append(np.nan)
        en_long.append(np.nan)
```

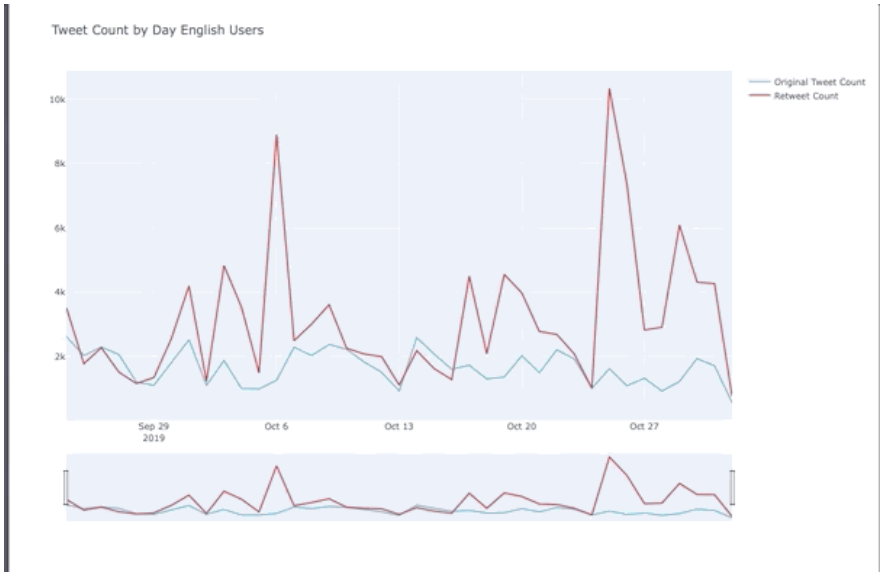


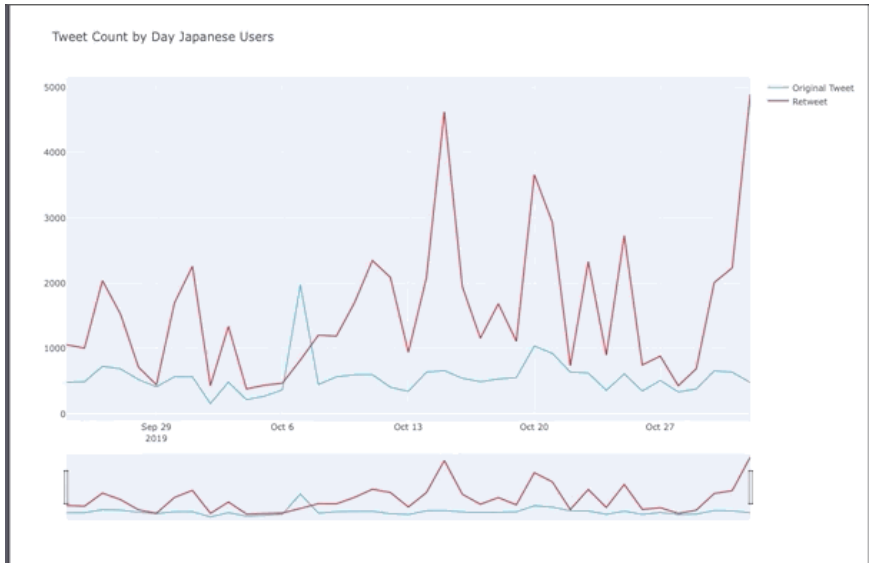
- Geocoder to find longitude and latitude for user location
- Most US influencers were in the US marked in red
- Most Japanese influencers marked in green were in Japan and also in Europe

Exploratory Data Analysis

- For both English and Japanese users retweets and tweets happened mainly on Friday and Saturday.
- For peak periods the biggest peak was the weekend before Halloween for English users
- Japanese users was during the JAL event and Halloween day.

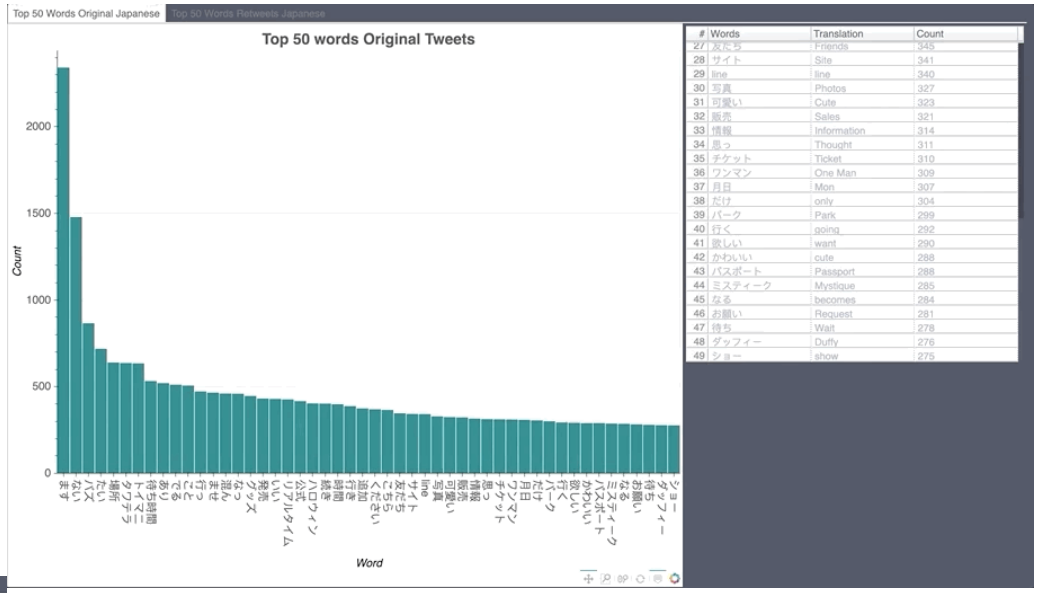


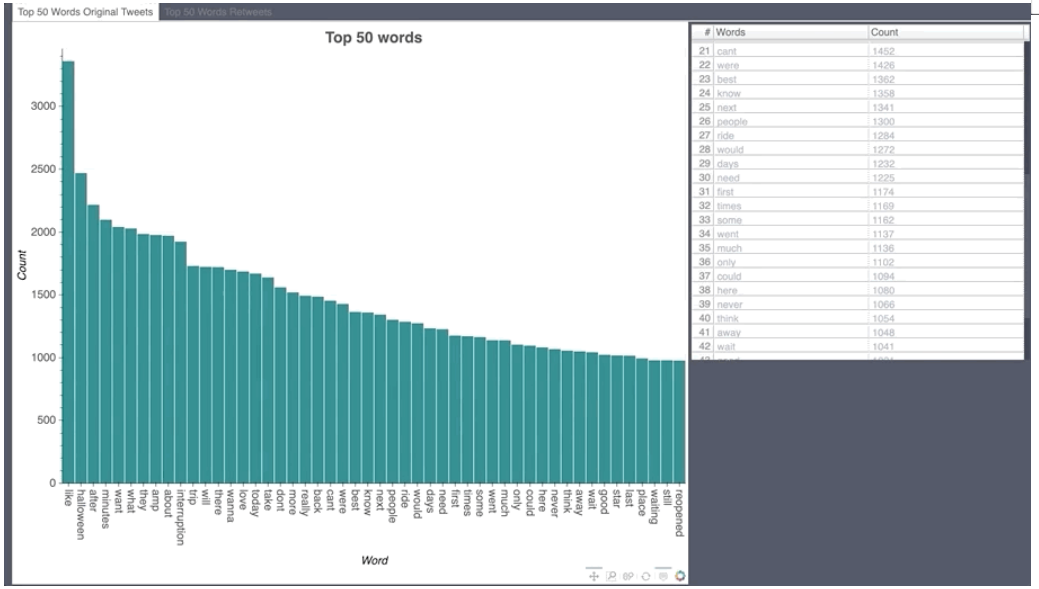




Exploratory Data Analysis

- Top 2 words original tweet English Users
  - Like and Halloween
- Top 2 words retweets
  - Fuck and annual



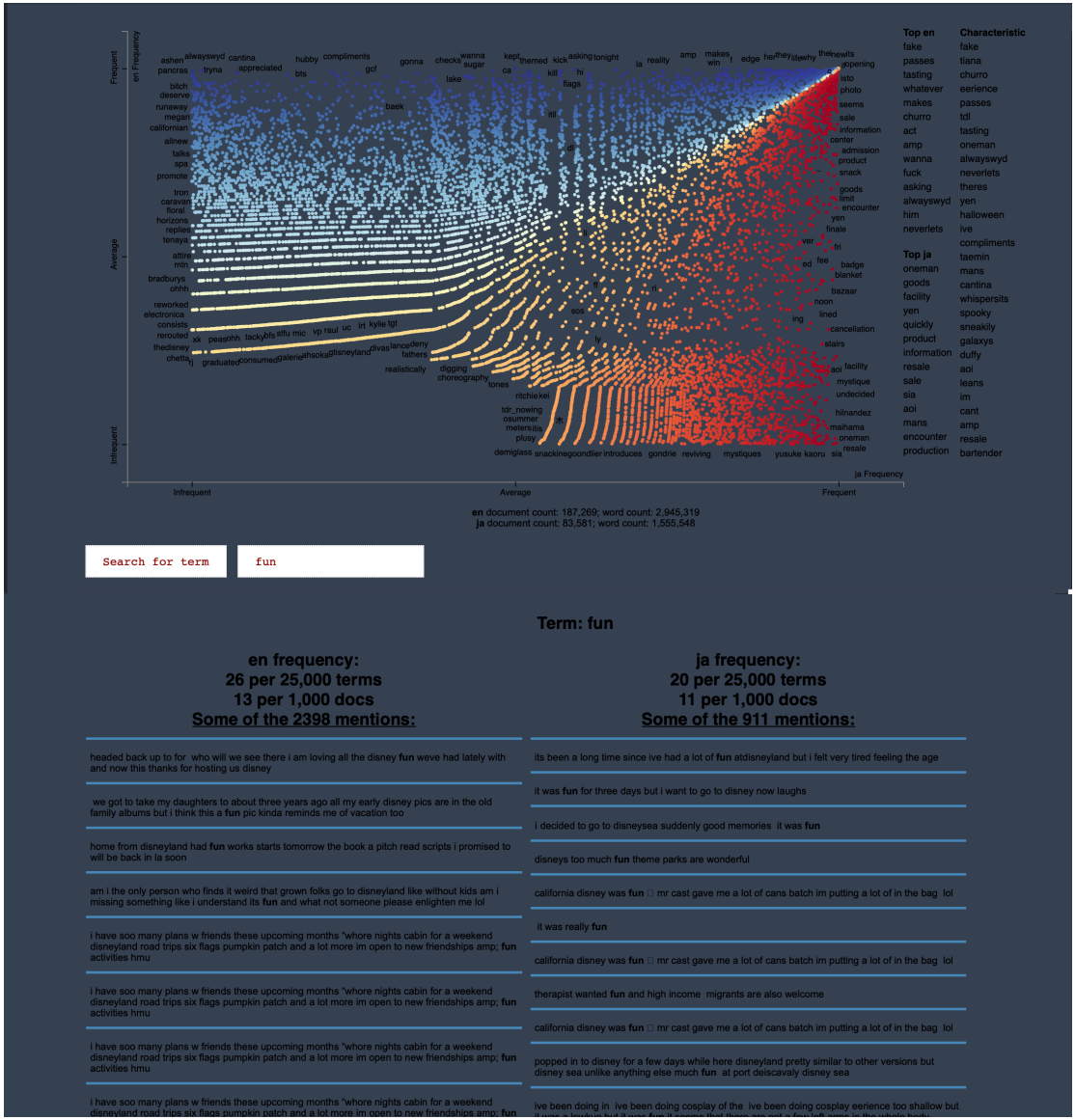


- Top 2 words original tweet Japanese Users
  - Monday and Sale
- Top 2 words retweets
  - Friend and Site

Exploratory Data Analysis

Scatter Text part of the Spacy Library

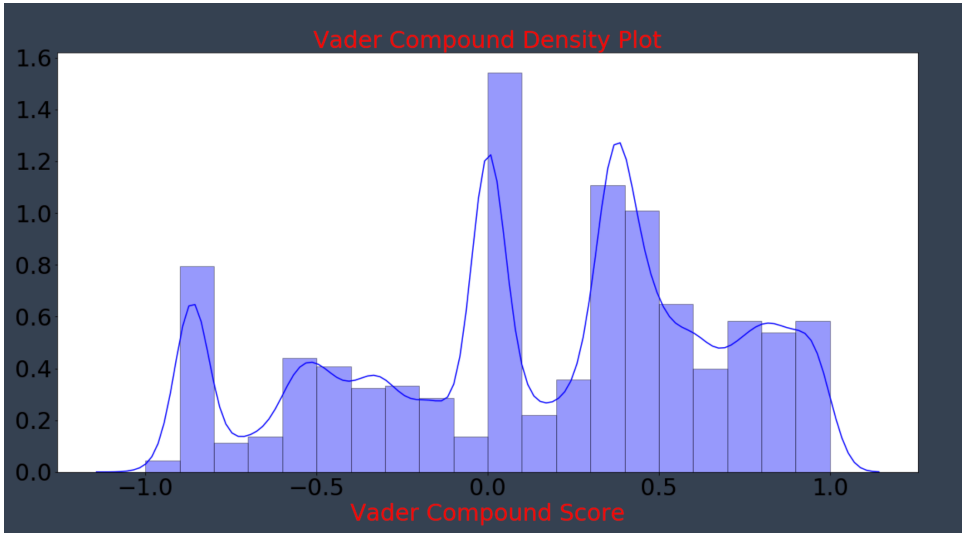
- Able to see word frequency across documents
- Able to do a search to see how many times it shows up per document or how many mentions
- Did Japanese documents that were translated compared to English documents

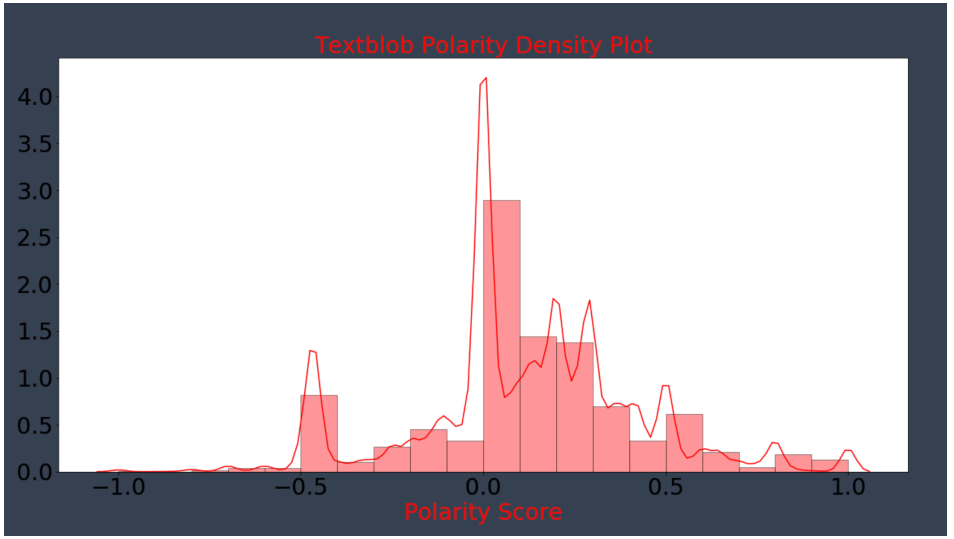


Stats

TextBlob and Vader sentiment compound score were used to compare:

- Japanese vs English user Sentiment towards the Disney Parks
- Tokyo Disney Resort vs Anaheim Disney Resort to see which one was more favorably received.





First had to test normality. Both were not normal. However should have large enough population for CLM to hold.

- Mean difference test showed that there was a difference between Japanese and English twitter users as well as the parks.
- Confirmed with non-parametric test, Mann- Whitney U Test. Both extremely low p-value at alpha 0.05

Stats

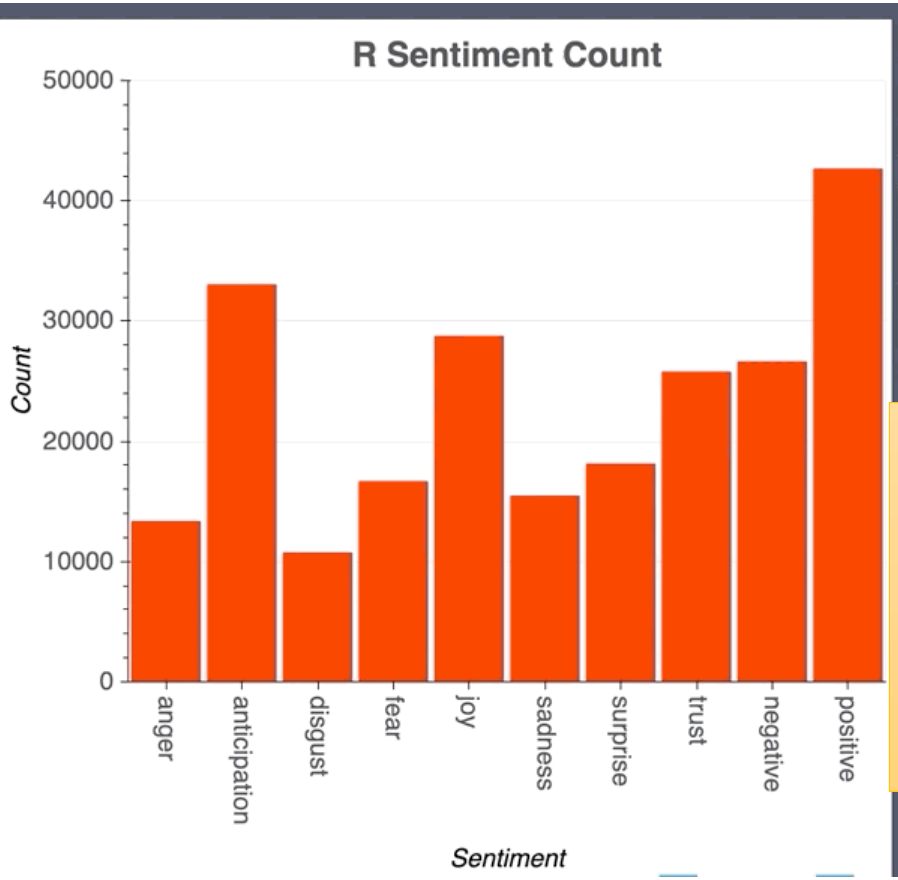
	user_favorite	user_follower	retweet	retweet_count	retweet_fav_count	text_length	polarity_tb	subjectivity_tb	vader_pos	vader_score
user_favorite	1	-0.0019	0.14	-0.023	-0.0093	0.06	0.0098	0.009	-0.0021	0.014
user_follower	-0.0019	1	-0.02	-0.0056	-0.0054	0.0062	0.0017	0.00089	-0.0005	0.0016
retweet	0.14	-0.02	1	0.22	0.22	0.18	-0.013	0.12	-0.0027	-0.019
retweet_count	-0.023	-0.0056	0.22	1	0.92	-0.16	-0.34	0.32	-0.092	-0.34
retweet_fav_count	-0.0093	-0.0054	0.22	0.92	1	-0.13	-0.26	0.28	-0.071	-0.27
text_length	0.06	0.0062	0.18	-0.16	-0.13	1	0.24	0.15	0.088	0.26
polarity_tb	0.0098	0.0017	-0.013	-0.34	-0.26	0.24	1	0.16	0.49	0.62
subjectivity_tb	0.009	0.00089	0.12	0.32	0.28	0.15	0.16	1	0.25	0.023
vader_pos	-0.0021	-0.0005	-0.0027	-0.092	-0.071	0.088	0.49	0.25	1	0.72
vader_score	0.014	0.0016	-0.019	-0.34	-0.27	0.26	0.62	0.023	0.72	1

- Not too correlated.
- Only thing that is really correlated is retweet count and favorite count with a correlation of 0.92
- Vader pos and compound score of course would have a high correlation at 0.72

Machine Learning

- Text preprocess steps
  - Remove amp,&, /n, lxml
  - Remove href
  - Remove @ mentions
  - Remove RT
  - Lemmatize
  - Remove punctuation
  - Remove short words less than 2 characters
  - Remove special Japanese Punctuation
  - Tokenize

Machine Learning



#	Sentiment	Count
9	positive	42646
1	anticipation	33016
4	joy	28734
8	negative	26603
7	trust	25758
6	surprise	18106
3	fear	16645
5	sadness	15442
0	anger	13331
2	disgust	10719

Feature engineering and selection

- R Sentiment Emotion
- Text Length
- Number of Capital letters
- Number of Punctuations
- Number of Emojis
- Number of Hashtags

Machine Learning

- Document term matrix selection
  - CountVectorizer best results overall
  - TFIDF Naïve Bayes TFIDF produced better results
  - Word2Vec worst results however an show word similarity
- Hyper Tuning
  - Random Forest best
  - Naïve Bayes second best
  - Logistic Regression third best
  - KNN worst

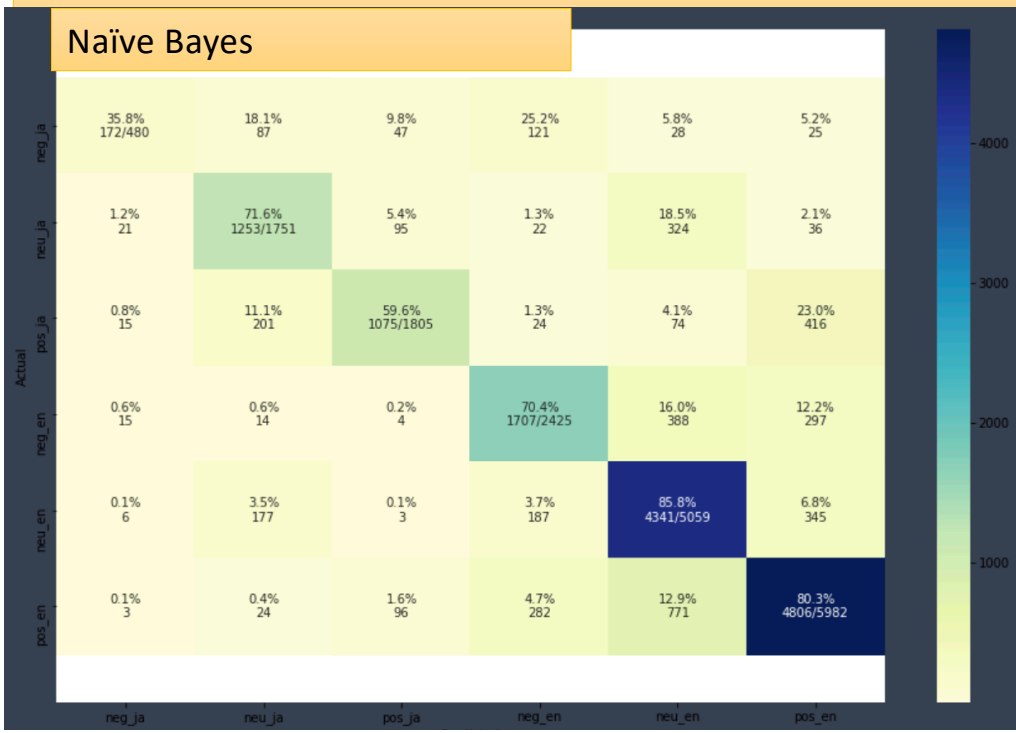
Important Features from Random Forest

```
[(0.07264998189604532, 'tweet_len'),
 (0.04641853216744782, 'cap_count'),
 (0.04272377734677494, 'punc_count'),
 (0.038893919132389765, 'hash_count'),
 (0.02146655519317955, 'sadness'),
 (0.018894953953547726, 'joy'),
 (0.017807906473101077, 'fear'),
 (0.016860613435313247, 'emoji_count'),
 (0.014306119628743563, 'anger'),
 (0.014305397515462942, 'anticipation'),
 (0.012611158225566781, 'trust'),
 (0.012546505297928749, 'disgust'),
 (0.009523934533368706, 172),
 (0.008760837448836236, 'surprise'),
 (0.008609675090072008, 470),
 (0.0050045463033912265, 547),
 (0.004917858903756468, 937),
 (0.004756201692522895, 496),
 (0.004426082061033796, 318),
 (0.004410321722824947, 182)]
```

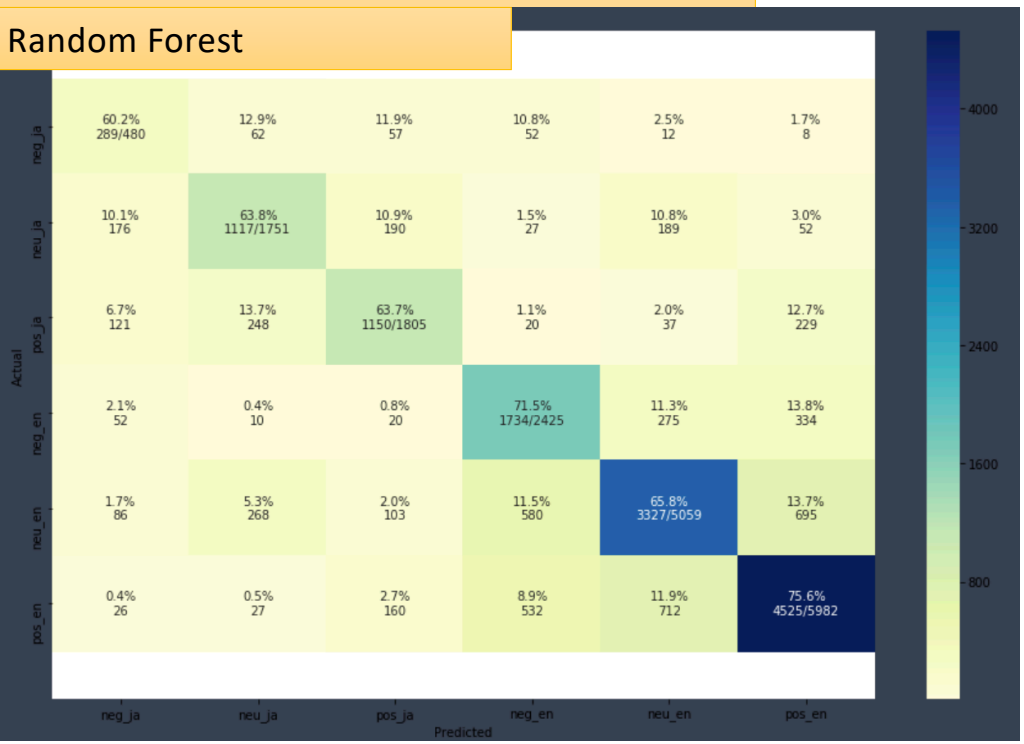
Machine Learning

- Naïve Bayes vs Random Forest
  - Naïve Bayes worked with 15000 max features and (1,3) grams best
  - Random Forest worked with unigrams at 5000 features best

Naïve Bayes



Random Forest



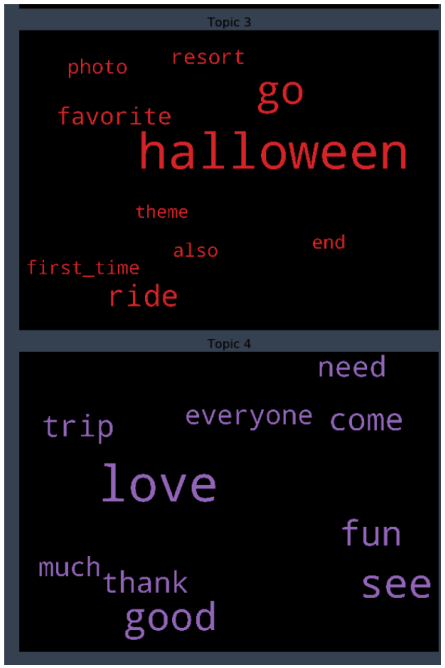
Machine Learning LDA

Japanese Positive

- One mans dream with Donald and friends
- Food and show with Mickey
- Good Happy and Halloween

English Positive

- Halloween and resort first time for the ride
- Cute Amazing friends
- Love good thank and fun



Chinese New Year Model Validation

- Model performed worse
- Only 46 percent accuracy using RF single n-gram model.
- 33 percent 1 to 3-gram Naïve Bayes model