Twitter Airline Sentiment Analysis

Aditya Chandel
Keval Shah
Anand Bora

Content

- Problem
- Dataset
- Exploratory Analysis
- Machine Learning Models
- Conclusion

Problem

Problem?

- Analyzing the tweet sentiments of travelers who flew on U.S. airlines in the month of February 2015.
- We want to answers various questions, like:
  - Text analysis of the user tweets to find out the reasons behind the user’s sentiments.
  - A language model for analyzing the sentiments using machine learning techniques.
  - Find out which airlines which provide best and worst customer satisfaction.
  - Get the most discussed topics among various airlines.
  - Other interesting stats and graphs.

Dataset

Twitter Data (Airline)

- 15000 users' tweets and metadata; about their reaction/opinion/review about a particular airline.
- CSV format.
- Tweets are labeled as Positive, Neutral, Negative by human labelers.
- Reason for negative sentiment ("Late Flight", "Rude Staff", etc.) is also in the dataset.
- Scraped in the month of February 2015.

Twitter Data (Airline)

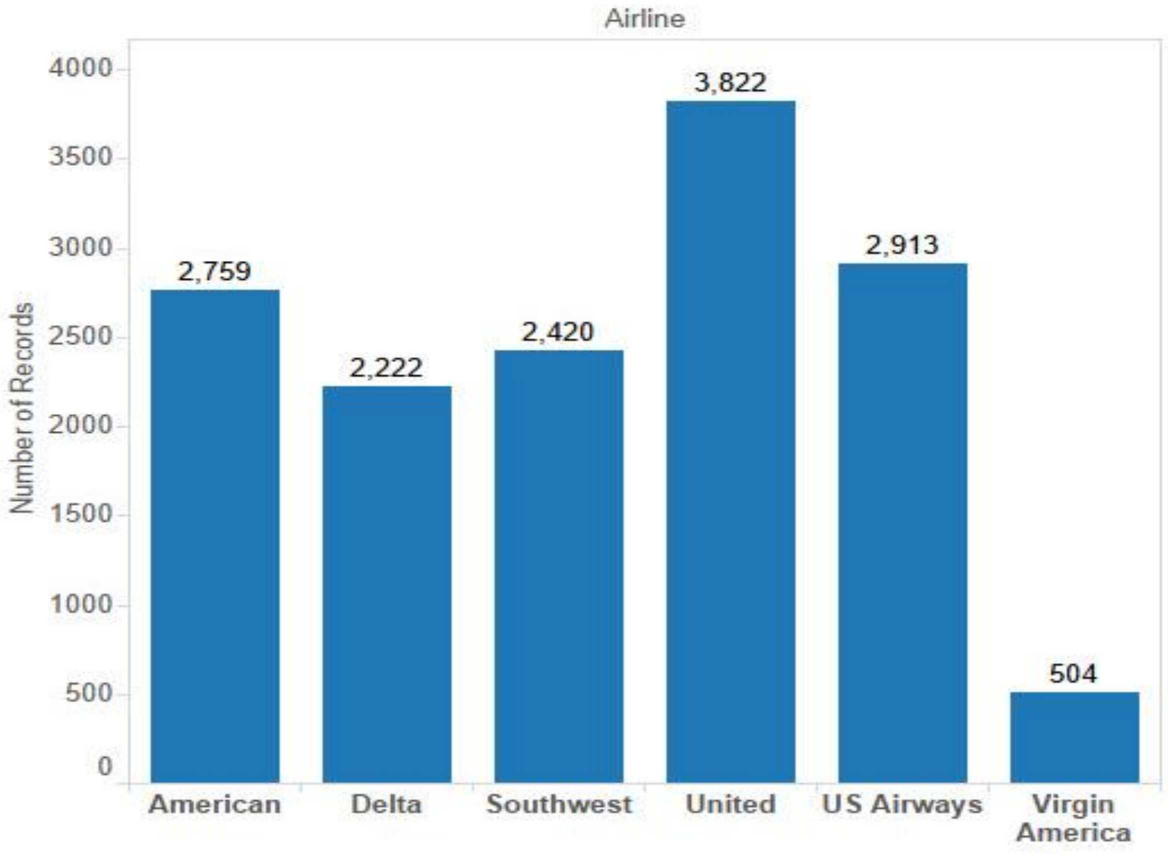
- The fields in the Tweets.csv file are:
  - tweet_id
  - airline_sentiment
  - negativereason
  - airline
  - name
  - text
  - tweet_coord
  - tweet_created
  - tweet_location
  - user_timezone

Twitter Data (Airline)

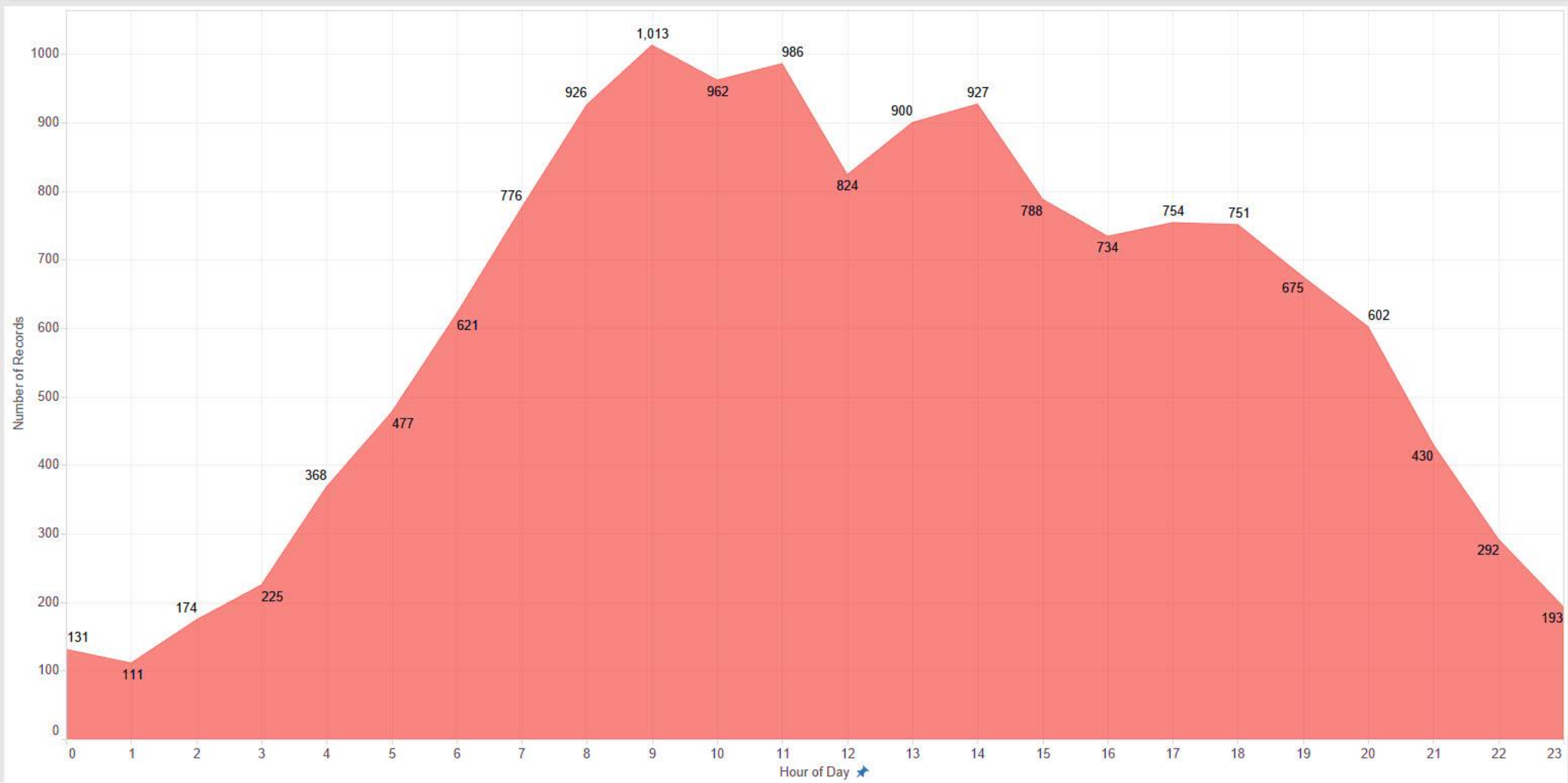
	tweet_id	airline_sentiment	negativereason	airline	name	text	tweet_coord	tweet_created	tweet_location	user_timezone
1	tweet_id	airline_sentiment	negativereason	airline	name	text	tweet_coord	tweet_created	tweet_location	user_timezone
2	5.70E+17	neutral		Virgin America	cairdin	@VirginAmerica What @dhepburn said.		2/24/2015 11:35		Eastern Time (US & Canada)
3	5.70E+17	positive		Virgin America	jnardino	@VirginAmerica plus you've added commercials to the experience.		2/24/2015 11:15		Pacific Time (US & Canada)
4	5.70E+17	neutral		Virgin America	yvonnalynn	@VirginAmerica I didn't today... Must mean I need to take another		2/24/2015 11:15	Lets Play	Central Time (US & Canada)
5	5.70E+17	negative	Bad Flight	Virgin America	jnardino	@VirginAmerica it's really aggressive to blast obnoxious "entertain		2/24/2015 11:15		Pacific Time (US & Canada)
6	5.70E+17	negative	Can't Tell	Virgin America	jnardino	@VirginAmerica and it's a really big bad thing about it		2/24/2015 11:14		Pacific Time (US & Canada)
7	5.70E+17	negative	Can't Tell	Virgin America	jnardino	@VirginAmerica seriously would pay $30 a		2/24/2015 11:14		Pacific Time (US & Canada)
8	5.70E+17	positive		Virgin America	cjmcginnis	@VirginAmerica yes, nearly every time I fly VX this â€œear wormâ€		2/24/2015 11:13	San Francisco C	Pacific Time (US & Canada)
9	5.70E+17	neutral		Virgin America	pilot	@VirginAmerica Really missed a prime opportunity for Men Withou		2/24/2015 11:12	Los Angeles	Pacific Time (US & Canada)
10	5.70E+17	positive		Virgin America	dhepburn	@virginamerica Well, I didn'tâ€¦but NOW I DO! :-D		2/24/2015 11:11	San Diego	Pacific Time (US & Canada)
11	5.70E+17	positive		Virgin America	YupitsTate	@VirginAmerica it was amazing, and arrived an hour early. You're t		2/24/2015 10:53	Los Angeles	Eastern Time (US & Canada)
12	5.70E+17	neutral		Virgin America	idk_but_youtube	@VirginAmerica did you know that suicide is the second leading ca		2/24/2015 10:48	1/1 loner squad	Eastern Time (US & Canada)
13	5.70E+17	positive		Virgin America	HyperCamiLax	@VirginAmerica I <3 pretty graphics. so much better than minim		2/24/2015 10:30	NYC	America/New_York
14	5.70E+17	positive		Virgin America	HyperCamiLax	@VirginAmerica This is such a great deal! Already thinking about m		2/24/2015 10:30	NYC	America/New_York
15	5.70E+17	positive		Virgin America	mollanderson	@VirginAmerica @virginmedia I'm flying your #fabulous #Seductive		2/24/2015 10:21		Eastern Time (US & Canada)
16	5.70E+17	positive		Virgin America	sjespers	@VirginAmerica Thanks!		2/24/2015 10:15	San Francisco, C	Pacific Time (US & Canada)
17	5.70E+17	negative	Late Flight	Virgin America	smartwatermelo	@VirginAmerica SFO-PDX schedule is still MIA.		2/24/2015 10:01	palo alto, ca	Pacific Time (US & Canada)
18	5.70E+17	positive		Virgin America	ItzBrianHunty	@VirginAmerica So excited for my first cross country flight LAX to M		2/24/2015 9:42	west covina	Pacific Time (US & Canada)
19	5.70E+17	negative	Bad Flight	Virgin America	heatherovieda	@VirginAmerica I flew from NYC to SFO last week and couldn't ful		2/24/2015 9:39	this place called	Eastern Time (US & Canada)

Exploratory Analysis

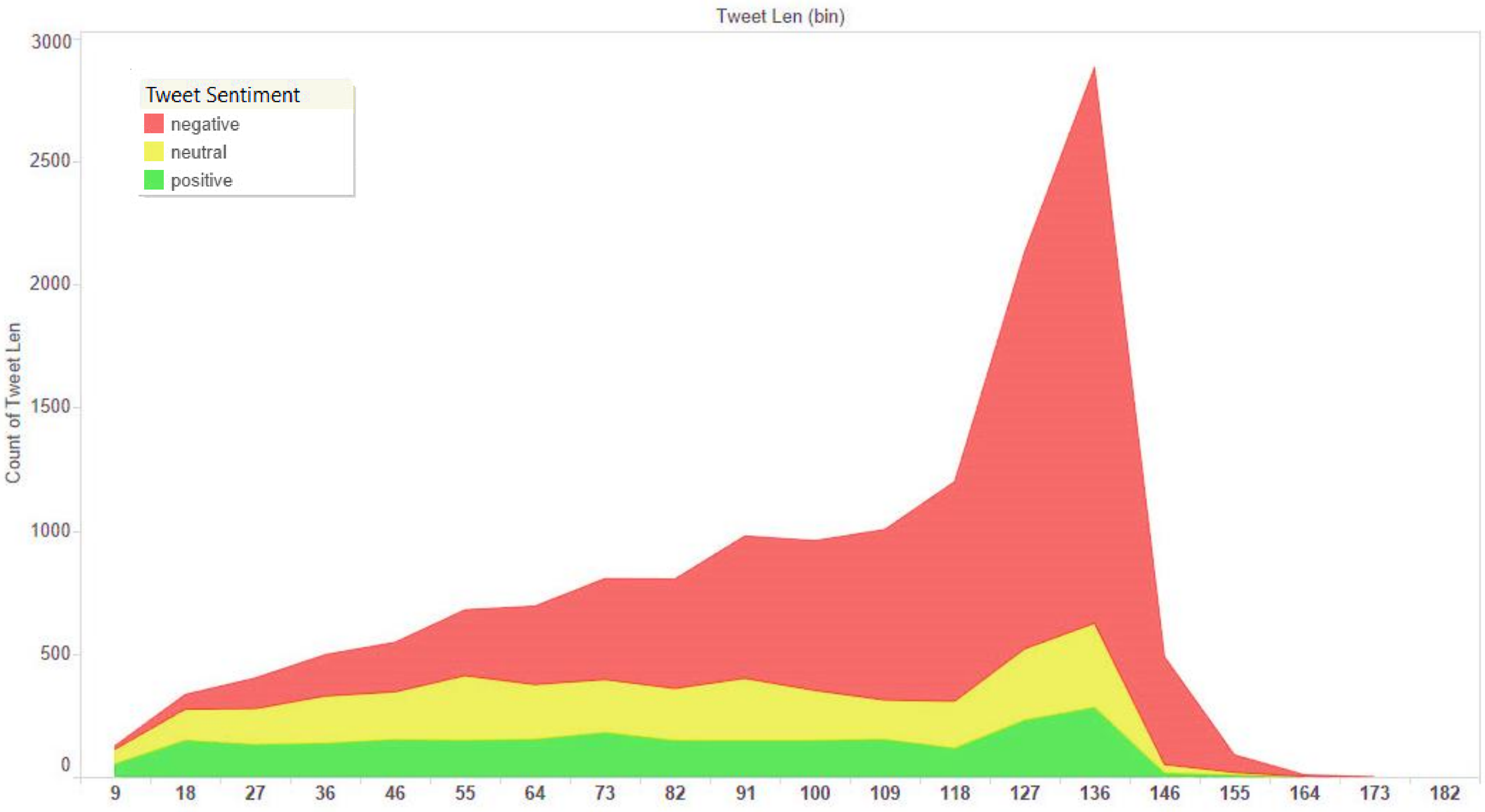
Tweets Per Airlines



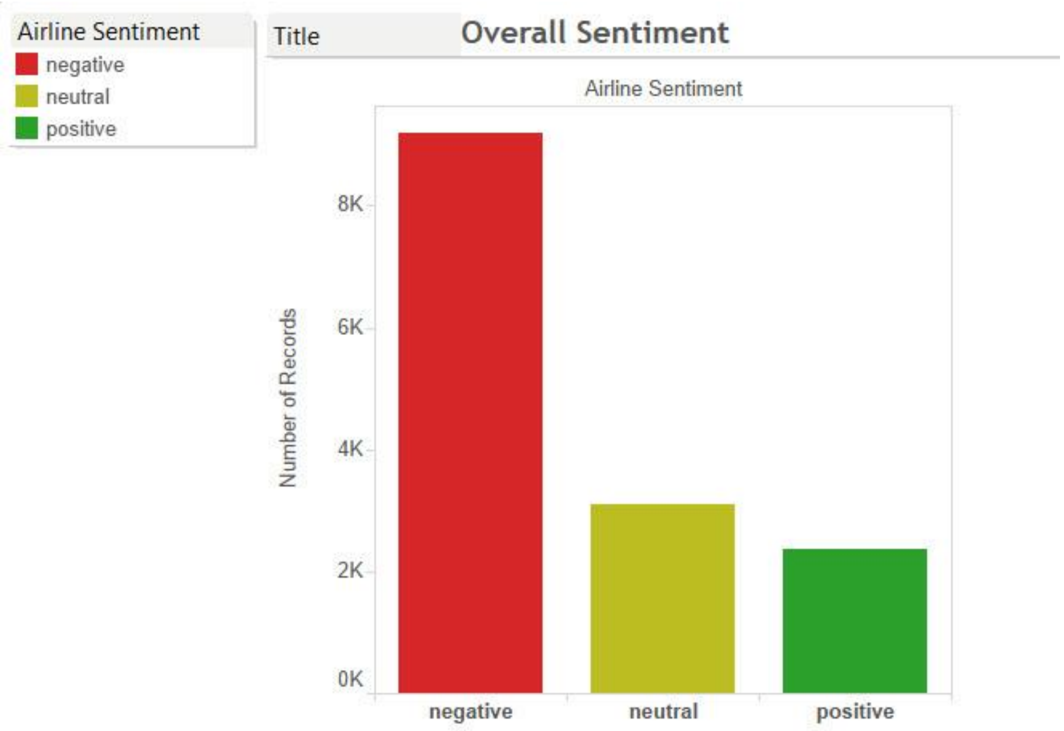
Time at which tweets were created

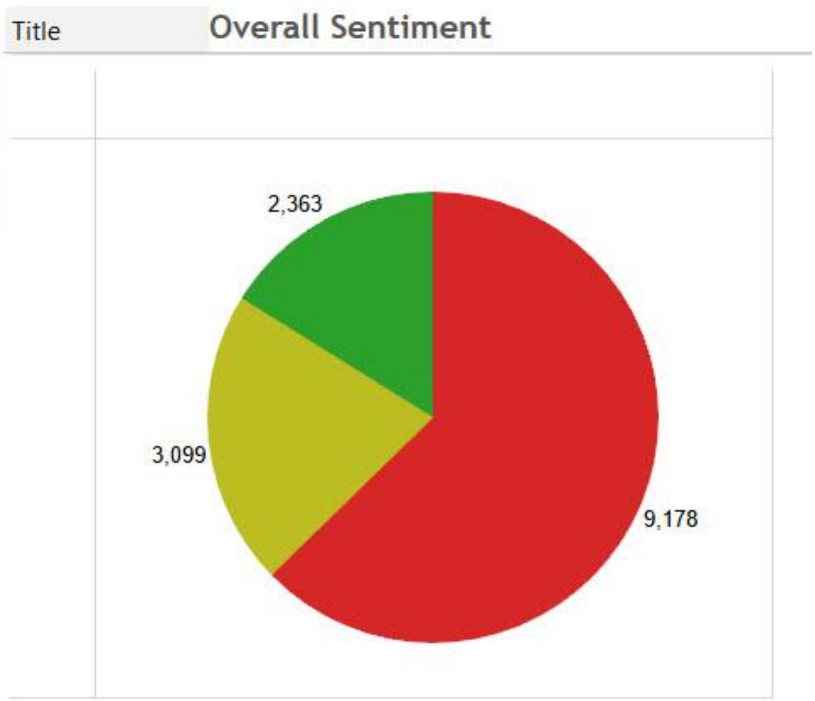


Length of Tweets

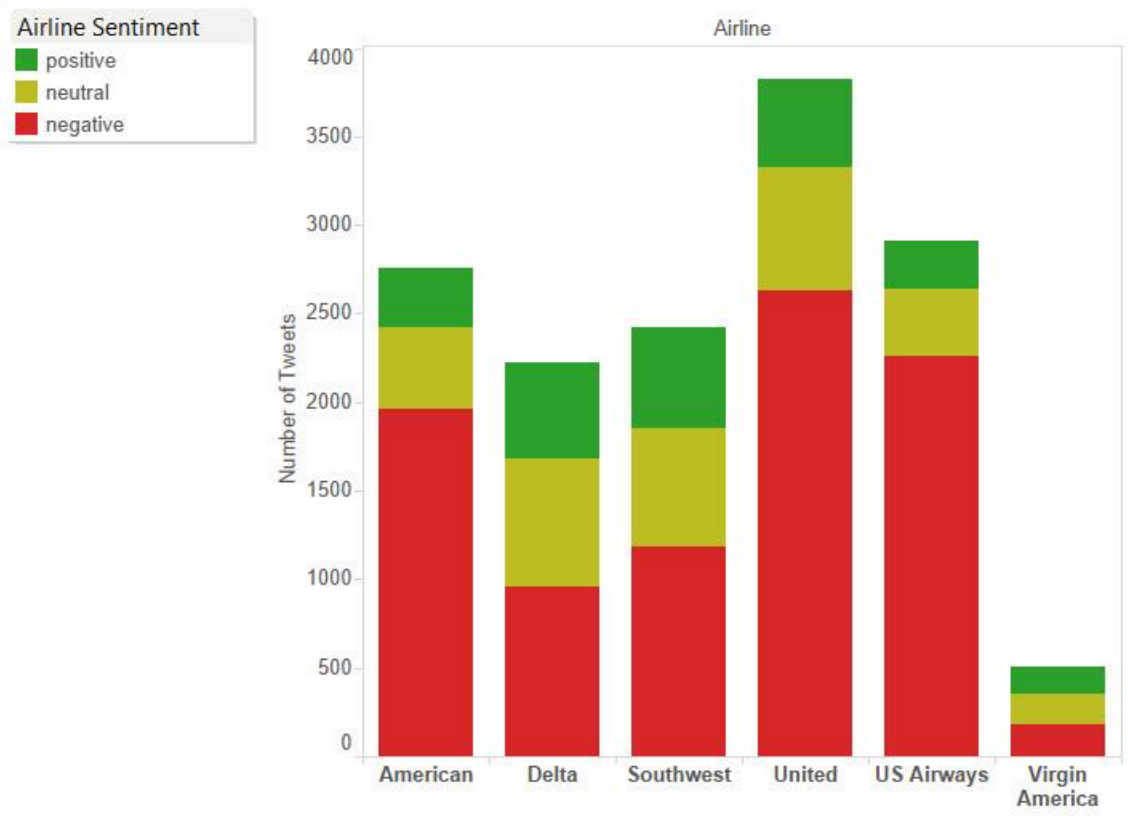


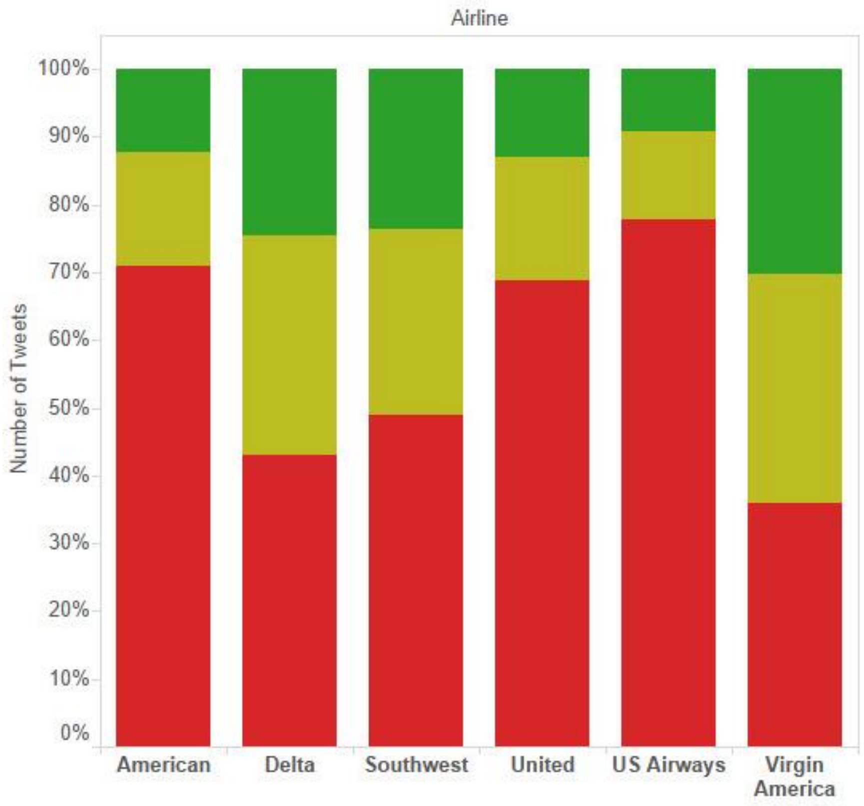
Overall Sentiments



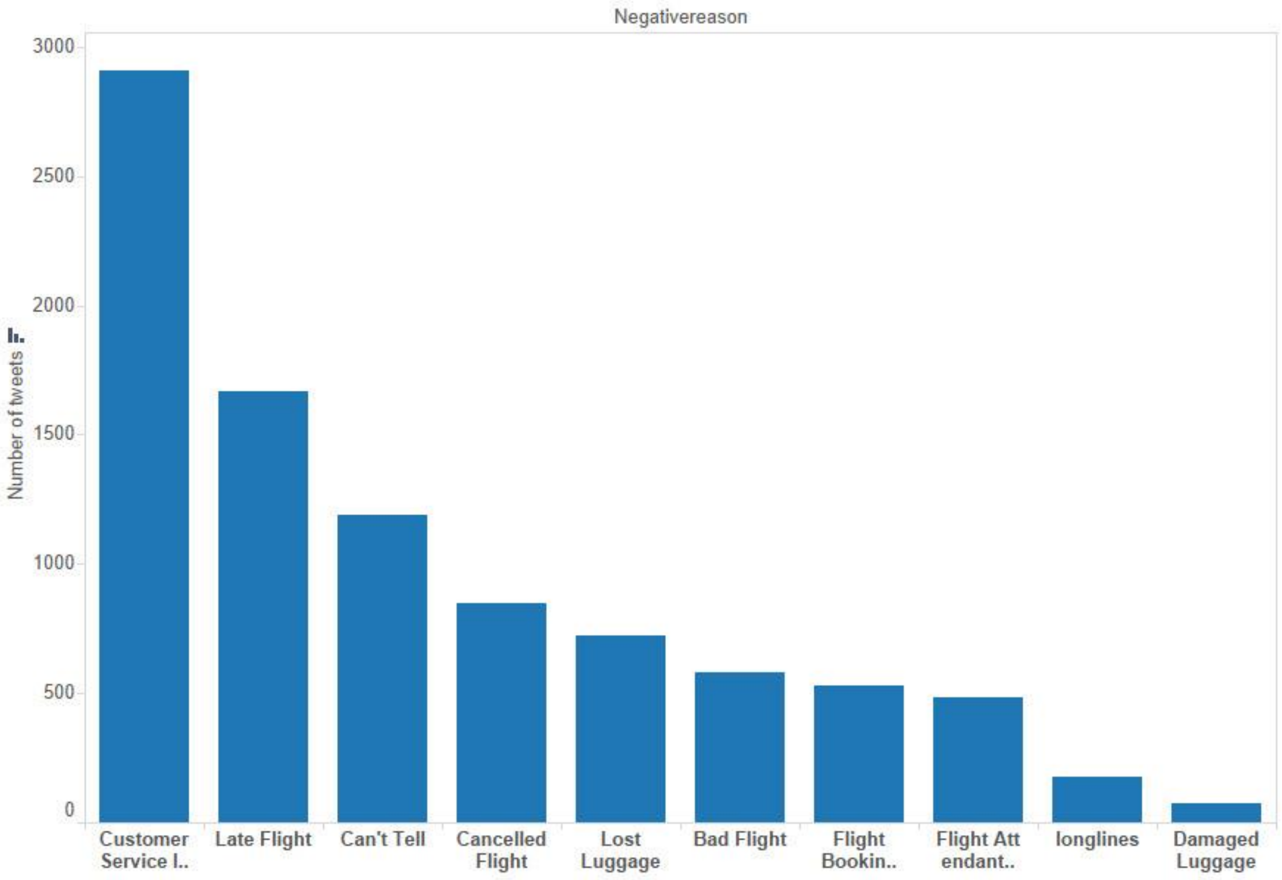


Sentiment for each Airline



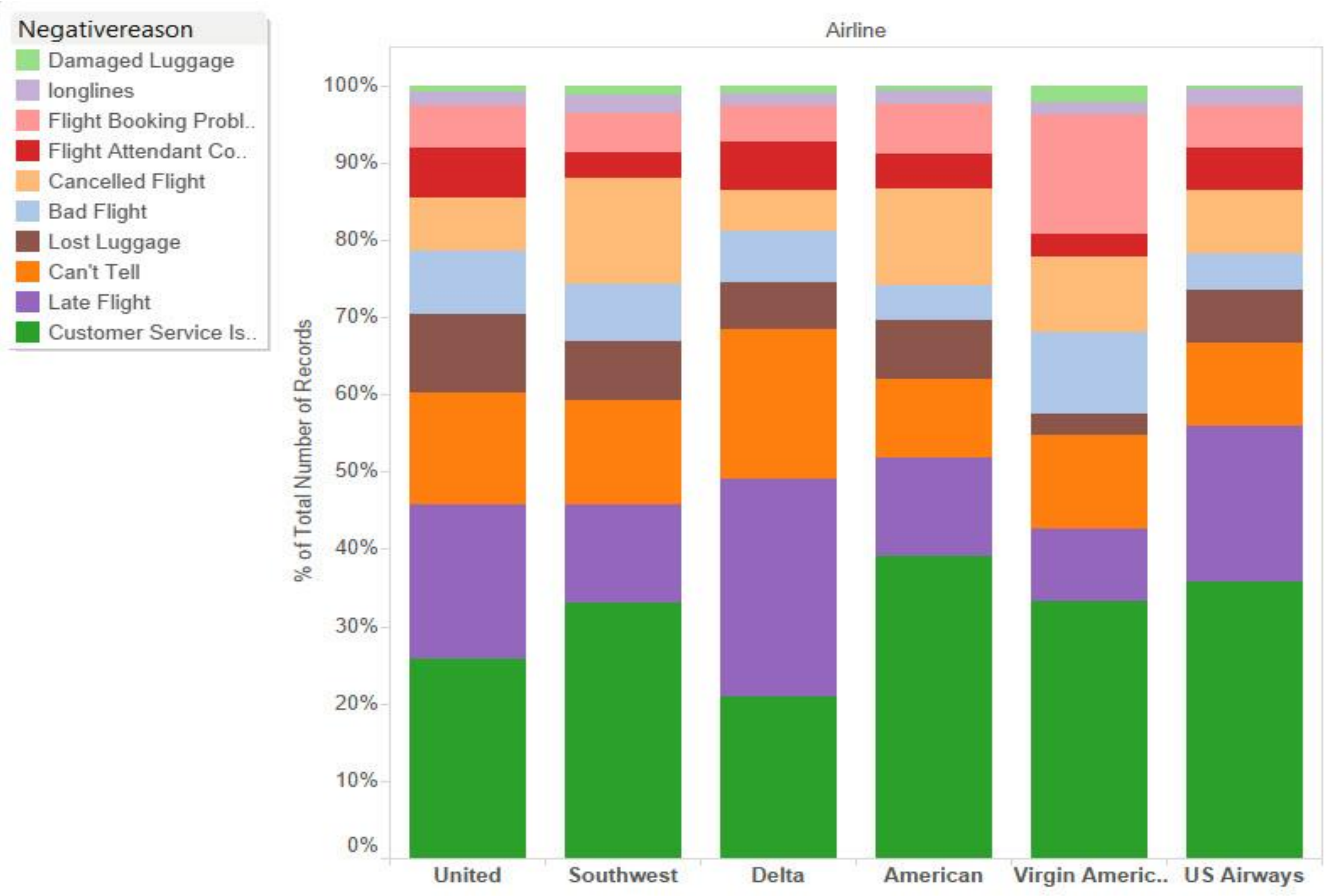


Reason for Negative Sentiment



Negativereason	
Bad Flight	580
Can't Tell	1,190
Cancelled Flight	847
Customer Service Issue	2,910
Damaged Luggage	74
Flight Attendant Complaints	481
Flight Booking Problems	529
Late Flight	1,665
longlines	178
Lost Luggage	724

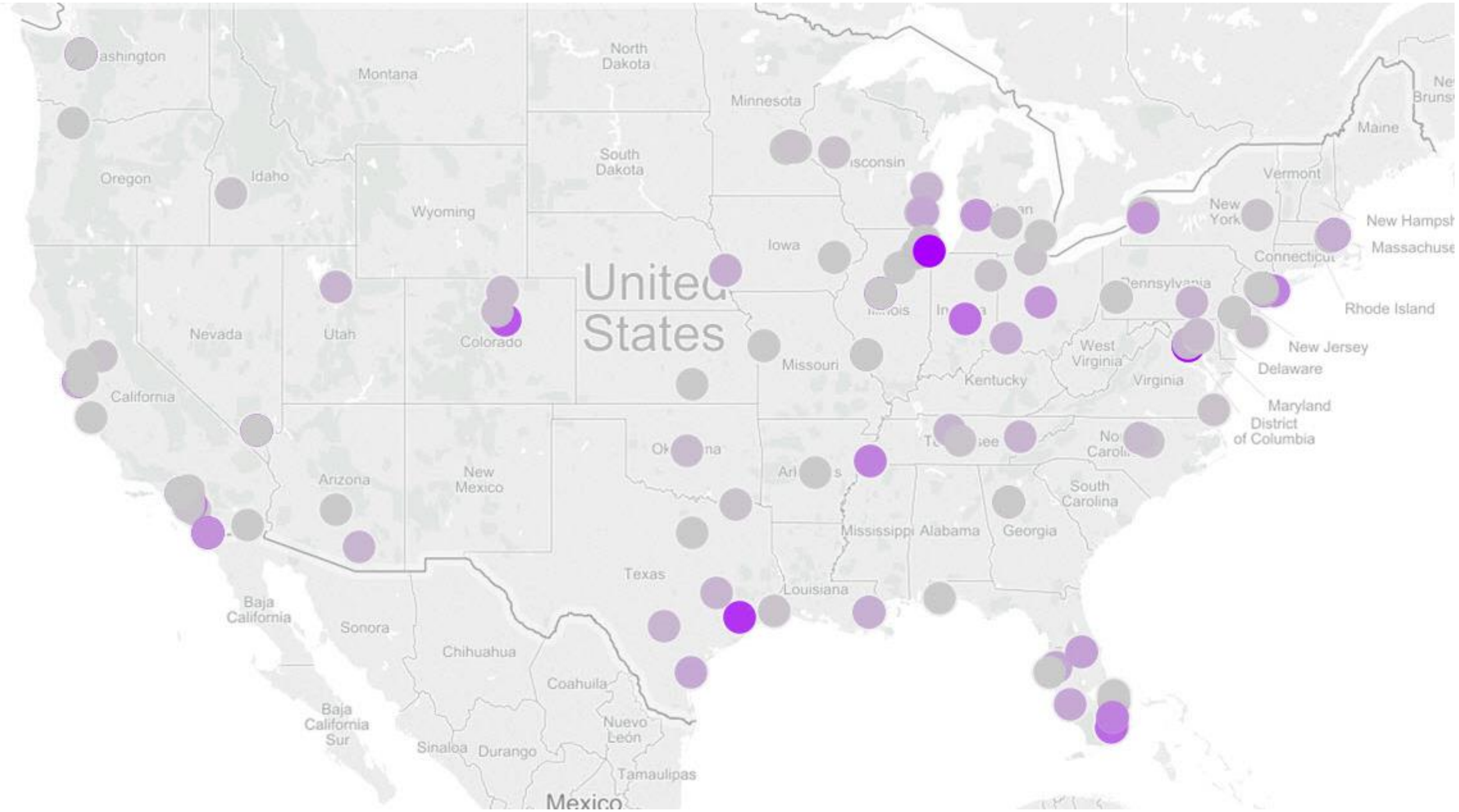
Reason for Negative Sentiment by Airline



Conclusions:

- American Airline has worst customer service.
- Delta has most late flights.
- Southwest and American has most cancelled flights.
- United is top in lost luggage.

Tweet Location

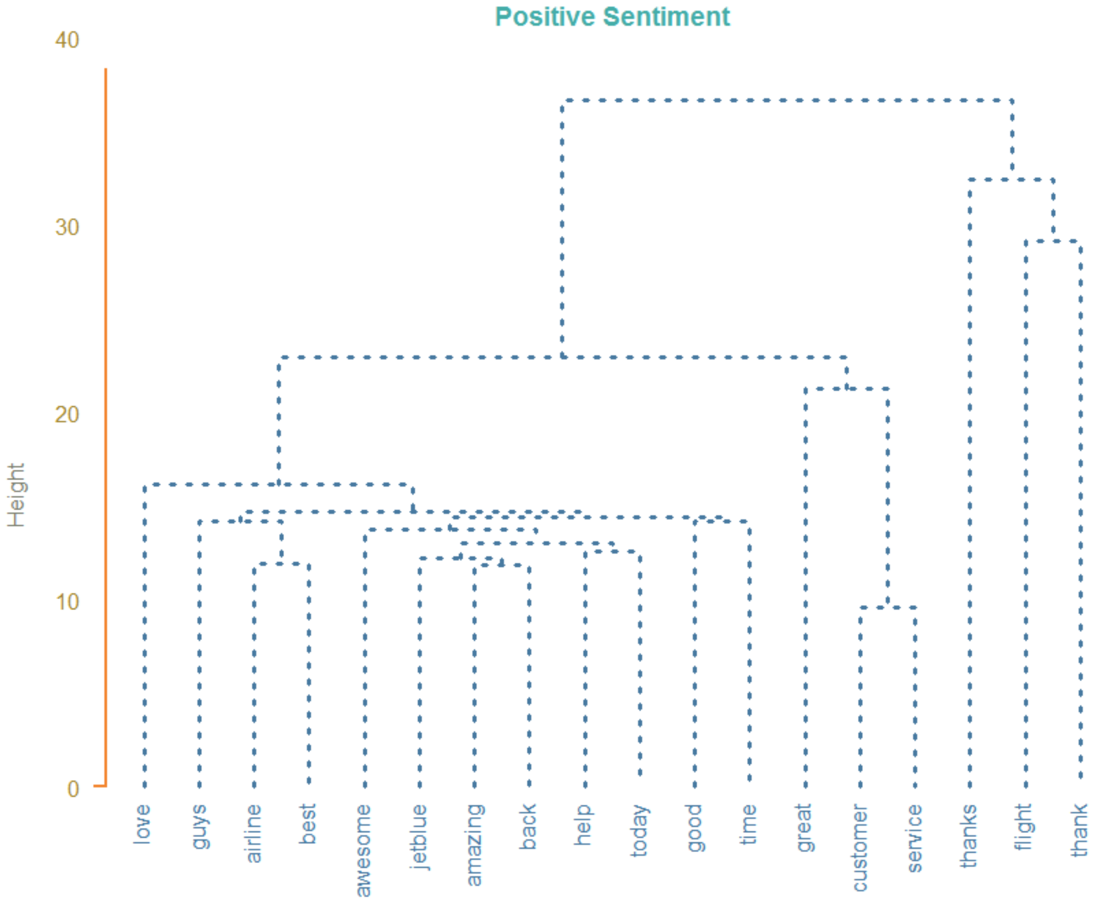


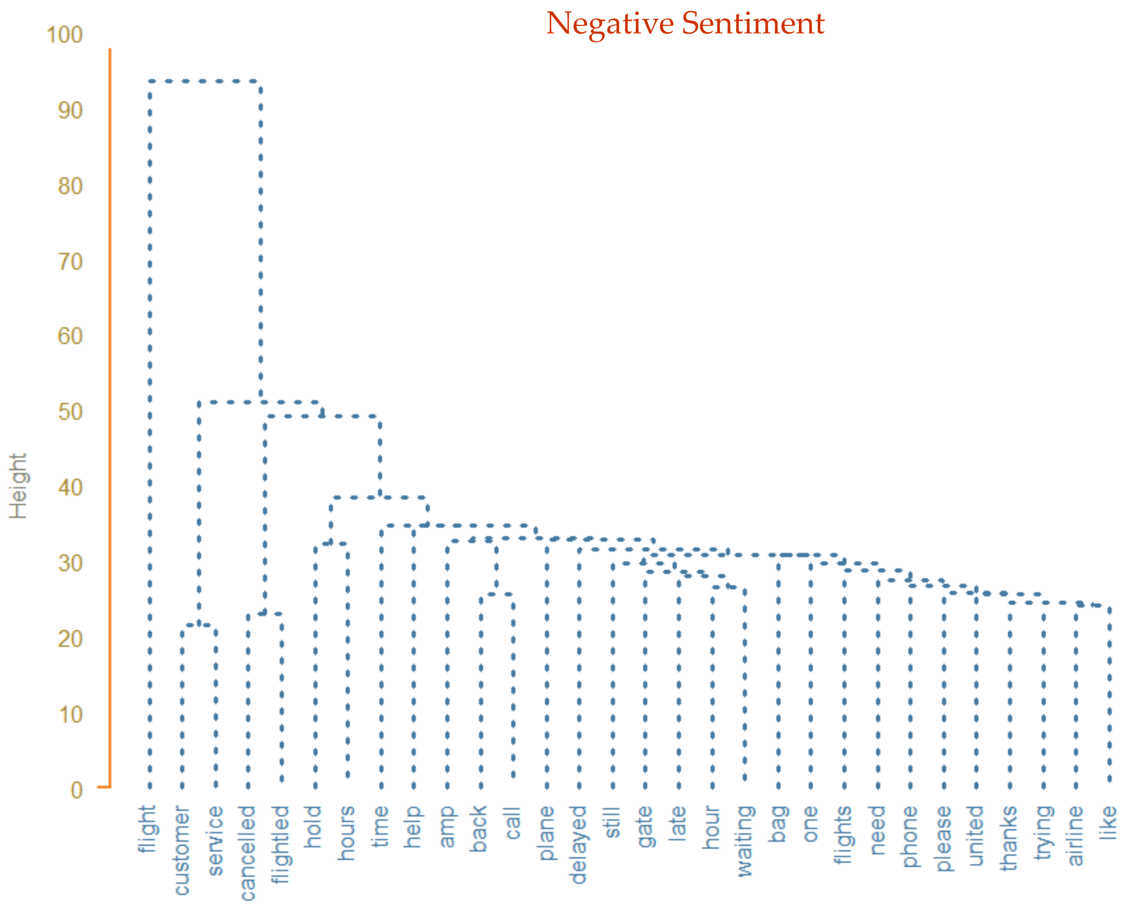
Top words: Positive & Negative Sentiments

Top 10 words in +ve sentiment

Top 10 words in -ve sentiment

Word Associations (Dendrogram)





Machine Learning Models

Build Machine Learning Classifiers

- Goal: To build a model for classifying flight tweets into destination class (i.e. positive, negative and neutral) using the features extracted from tweet text.
- Approach:
  1. Load tweets into Sframe.
  2. Preprocess Tweets:
     - Remove punctuations.
     - Convert to lower case.
     - Convert www.* or https://* to URL
     - Convert @username to AT_USER
     - Remove additional white spaces
     - Replace #word with word

Build Machine Learning Classifiers

- Approach (cont.):

3. Building Feature Vector from Tweets (Feature Extraction):
   - Replace words with two or more occurrences.
   - Strip punctuations.
   - Remove words not starting with alphabets.
   - Remove Stop Words.

4. Train classifiers using the processed tweets and extracted features.
   - Multi-class classification problem.
   - Classifiers used: ***Logistic Regression, Random Forest, Boosted Tree, Decision Tree.***
   - 10 Iterations.
   - 10-fold Cross Validation.
   - Calculate Accuracy, AUC, F1 Score, Precision, Recall.

Results / Evaluation

Evaluation (1-Gram Model)

Classifier	Accuracy	AUC	F1 Score	Precision	Recall
Logistic Regression	0.692	0.804	0.587	0.589	0.588
Random Forest	0.669	0.706	0.41	0.714	0.417
Decision Tree	0.665	0.67	0.404	0.679	0.409
Boosted Tree	0.671	0.764	0.421	0.707	0.424

- **Logistic Regression**: Best ***Accuracy, AUC, Recall and F1 score***.
- **Boosted Tree**: Best ***Precision***.

Evaluation (2-Gram Model)

Classifier	Accuracy	AUC	F1 Score	Precision	Recall
Logistic Regression	0.766	0.863	0.689	0.707	0.675
Random Forest	0.703	0.803	0.499	0.731	0.512
Decision Tree	0.7	0.766	0.511	0.662	0.518
Boosted Tree	0.722	0.85	0.552	0.744	0.544

- 10% increase in performance over 1-Gram Model.
- **Logistic Regression**: Best ***Accuracy, AUC, Recall and F1 score***.
- **Boosted Tree**: Best ***Precision***.

Evaluation (TF-IDF Model)

Classifier	Accuracy	AUC	F1 Score	Precision	Recall
Logistic Regression	0.732	0.834	0.66	0.66	0.661
Random Forest	0.653	0.69	0.378	0.696	0.401
Decision Tree	0.654	0.659	0.384	0.706	0.403
Boosted Tree	0.659	0.761	0.398	0.717	0.411

- All TF-IDF classifiers performed worse than non TF-IDF models.
- 10-15% drop in measures (i.e. Accuracy, AUC, etc.).
- Reason: Less characters in tweets (<140 words). So all the words used have high significance.

Summary & Conclusion

Summary

- First we performed thorough exploratory analysis on the data to find interesting information about airlines, like:
  - Flights with best and worst customer satisfaction.
  - Reasons for negative sentiment.
  - Reasons for negative sentiment by airlines.
  - Top negative and positive words.
  - Word associations (Dendrogram)
- Then we built various machine learning models to classify tweets to positive, neutral and negative categories, and to check if they confirm with human labeling.
- 2-Gram model performed the best, with classification accuracy of 76% and AUC of 84%.
- TF-IDF model performed worst.
- We have carried out what we had proposed in our proposal.