

# Machine Learning - Final Project

Sentiment Analysis

Andrew Alvarez

Github: <https://github.com/Rewcifer/Machine-Learning-Final>

Spring 2023

# Contents



## **Business Use-case and Objective**



## **Industry Performance and Outlook**

- **US Domestic Airlines Industry Sentiment**
- **Major customer topics within US Domestic Airlines Industry**



## **Airline-specific sentiment evaluation**

- **Evaluate Bert topic modeling and sentiment analysis findings**
- **United Airlines positioning within broader industry**



## **Headwinds, Tailwinds, and Recommendations**

- **Review go-forward strategies and recommendations for client**
- **Opportunity area for United Airlines to increase loyalty**

# Business use-case and objective

---

1

**Loyalty programs** represent a significant source of profit for airlines in the US. Understanding what delights and disappoints customers is crucial to **customer strategy** and **business success**

2

In a post-pandemic environment where customer loyalty has been significantly disrupted, US Airlines are presented with a unique opportunity to increase their market share via **new customer acquisition**

3

Developing a robust analysis using **topic modeling** on tweets and a broader **sentiment analysis** will serve as an example of deep learning in action

# A customer-first approach to building a lasting strategy overlays customer needs and sentiment to drive value

1

## Total Addressable Market

What do customers care about?

How do we grow our customers?

2

## Brand Heath

What are our *surprises and delighters*?

What are our customers frustrated with and how do we address it?

Customer  
Scope



Total Market

Customer  
Strategy



**Acquire & Engage**  
Explore the ex-loyalty space post-pandemic  
Develop through meaningful connection

Customer  
Insights



**What are customer needs?**



United's potential loyalty customers



**Surprises/ Delights & Disappointers**  
what drives positive customer sentiment  
What drives negative brand engagement



**How** do we differentiate ourselves from the competition by maximizing delight?

# Data Assumptions and Hypotheses

## Tweet / Airline Representation

- This exercise assumes that the tweets used are a reasonable and actionable representation of sentiment toward US airlines
- Cultural nuances and other subtleties, such as sarcasm and irony, will be present and may affect model performance
- The distribution of tweets suggests that results for some airlines may suffer from underrepresentation
- As a result, **three airlines were chosen with the highest volume of tweets:** United, American, and Southwest

## Sentiment / Topic Correlation

- Sentiment toward airlines will differ among key categories, enabling actionable insights
- A key assumption is that certain topics will have a higher association with negative or positive sentiment (e.g. tweets about delays or cancelled flights will be more likely to be negative than positive)

## Model Hypotheses and Observations

- Leveraging a bidirectional encoder-decoder transformer model (BERT) will be well suited for sentiment analysis, as they excel in understanding sentence context
- Leveraging a BERT model fine-tuned for topic classification will enable stronger performance on classifying sentiment in my data
- Using a distilled version of BERT will increase computational efficiency without sacrificing performance

## Model Limitations and Bias

- While the fine-tuned DistilBERT-base-uncased model achieves 91.3% accuracy on the test set, additional validation will be required
- Known biases toward underrepresented populations and gender are documented within this model and therefore may affect overall results (Note: Add additional research into bias with datasets provided in hugging face to next steps)

# Data Overview

tweet_id	airline	name	text
0	1	United	rdowning76 @united thanks
1	2	United	CoreyASewart @united Thanks for taking care of that MR!! Happy customer.
2	3	United	CoralReefer420 @united still no refund or word via DM. Please resolve this issue as your Cancelled Flightled flight was useless to my assistant's trip.
3	4	United	Isalazarll @united Delayed due to lack of crew and now delayed again because there's a long line for deicing... Still need to improve service #united

## Dataset:

**8,807 Tweets** Sourced from Kaggle

**Composed of** tweets directed at American Airlines, Southwest Airlines, and United Airlines accounts

**Data size:** 803KB

## Data Variables:

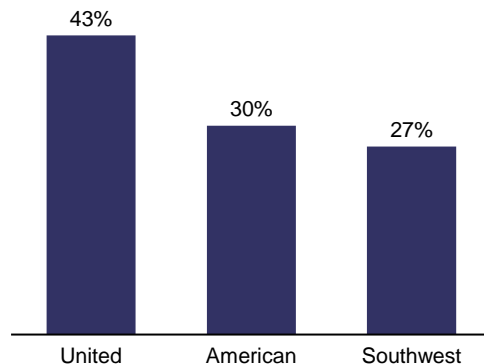
**1. tweet\_id**

**2. Airline**

**3. Name**

**4. Text**

## Share of Tweets:



# Data Cleaning and Feature Engineering/Preparation

1

## Data Cleaning

 [prasanthg3 / cleantext](#)

**Dropping any duplicate tweets from dataset**

2

**Cleaning the data for special characters and numbers**

3

**Checking for any nulls or potential datatype issues prior to modeling**

## Topic Modeling



**Removing stopwords within the tweets and stemming finding root of each word to enhance the interpretability of tweets (e.g., running → run)**

**Leveraging yake to help determine positive and negative associated words**

**Bigram and Trigram analysis to further understand drivers of sentiment via term frequency**

# Model Selection and Methodology

## Model Research

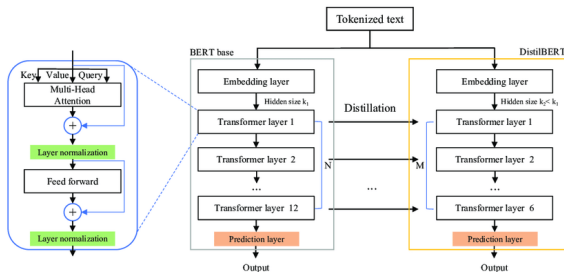
DistilBERT

### DistilBERT based uncased Model

A **self-supervised bi-directional encoder-decoder 6-layer transformer** model that is smaller and faster than BERT but trained to return the same probabilities as the BERT base model.

**Masked Language Modeling** (a procedure to predict the missing words in a sentence based on the context of its surrounding words), was used in the training procedure.

**Chosen for its computational efficiency, performance in capturing the context and semantics of sentences, suitable for transfer learning, and can handle longer text**



## Model Selection

### DistilBERT based uncased finetuned SST-2

DistilBERT-based-uncased **fine-tuned** on the Stanford Sentiment Treebank (SST-2) dataset. This model was chosen given the downstream fine-tuning approach, **which trained the model to understand the compositional structure and effects of sentiment in language.**

**The model achieves 91.3% accuracy on the development set** vs the standard DistilBERT base model accuracy of 92.7%

### BERTopic

Python library **leveraging pre-trained transformer models** to create meaningful and interpretable topics for sentiment analysis via sentence embedding generation and dimensionality reduction to cluster and create topics based on language.

#### Hyperparameter Tuning

Hyperparameter tuning was used to find the optimal parameters for the UMAP and topic models. These optimal parameters were found via **random search.**

## Assessing Accuracy

### Classification Accuracy

In lieu of assessing and classifying all 8,807 tweets individually, **a random sample of 2,000 tweets will be fed into GPT 3.5-Turbo through the OpenAI API and prompted to classify each tweet.** This will be compared with the results from DistilBERT based uncased finetuned SST-2.

**prompt = "" "Categorize the sentiment of the writer as POSITIVE or NEGATIVE based on the text wrapped by triple backtick**

**return the result as a json value with the key name as label for value of POSITIVE or NEGATIVE**

**only you will return right after "You will return"**

**Question:**  
```{question}```  
**You will return**  
""

### BERTopic Topic Diversity Measure

A function to calculate topic diversity. It creates a list of topics (each comprising a list of words) and calculates the overall topic diversity using two variables: `unique_words` and `total_words`.

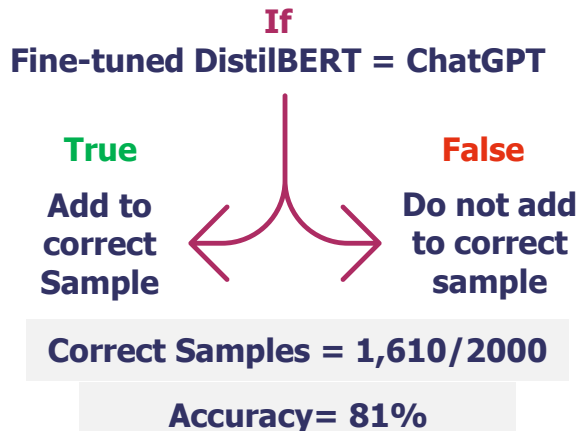
The primary goal here is to confirm that we have unique and discernable topics via the BERTopic modeling package to be able to understand the major components driving sentiment across the corpus of tweets.



# Model Results and Implications

## Fine-tuned DistilBERT

Sentiment Classification tested against OpenAI GPT-3 Turbo



**Mismatch Example:** americanair thx for showing me that your twitter appreciates me more than your employees. sure another airline would like my nohotel

DistilBERT Label: **POSITIVE**

GPT\_label: **NEGATIVE**

**Match Example:** united thanks so much my passport was recovered

DistilBERT Label: **POSITIVE**

GPT\_label: **POSITIVE**

## BERTopic Topic Modeling

Assessing Topic Diversity through Hyperparameter tuning and uniqueness

### 1 Parameter grid testing

```
"bertopic__nr_topics": [5, 10, 15, 20],  
"bertopic__min_topic_size": [5, 10, 15, 20],  
"umap__n_neighbors": np.arange(5, 50, 5),  
"umap__min_dist": np.linspace(0.0, 1.0, 11),  
"umap__n_components": [2, 3, 4, 5],
```

### Random Search Results

```
"bertopic__nr_topics": [5, 10, 15, 20],  
"bertopic__min_topic_size": [5, 10, 15, 20],  
"umap__n_neighbors": np.arange(5, 40, 5),  
"umap__min_dist": np.linspace(0.0, 0.9, 11),  
"umap__n_components": [2, 3, 4, 5],
```

### Final Hyperparameters

```
"bertopic__nr_topics": [5, 10, 15, 20],  
"bertopic__min_topic_size": [5, 10, 15, 20],  
"umap__n_neighbors": np.arange(5, 40, 5),  
"umap__min_dist": np.linspace(0.0, 0.9, 11),  
"umap__n_components": [2, 3, 4, 5],
```

*Limited to 10 topics for client interpretability. 15 Topics for this use case would be too many to action on.*

### 2 Topic Diversity

$$= \frac{\text{Length of Unique Words}}{\text{Total Words}}$$

**Diversity Score = 90%**

# Overall Airlines Sentiment

Total Addressable Market

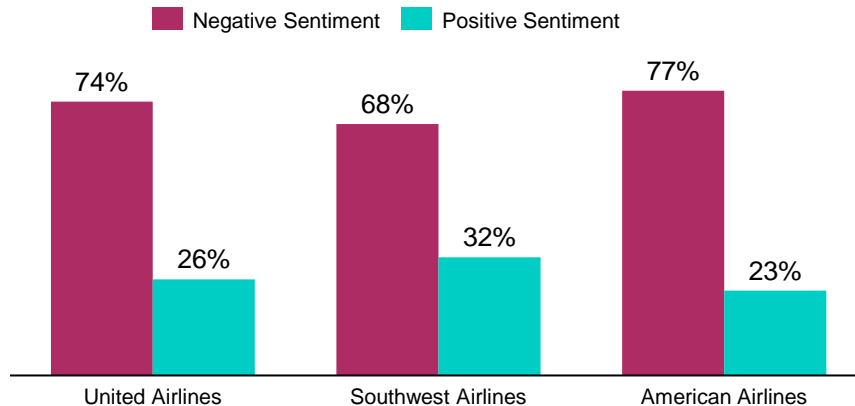
What do customers care about?

How do we grow our customers?

# In leveraging sentiment analysis, we were able to assign the highest count of sentiment to the overall tweet

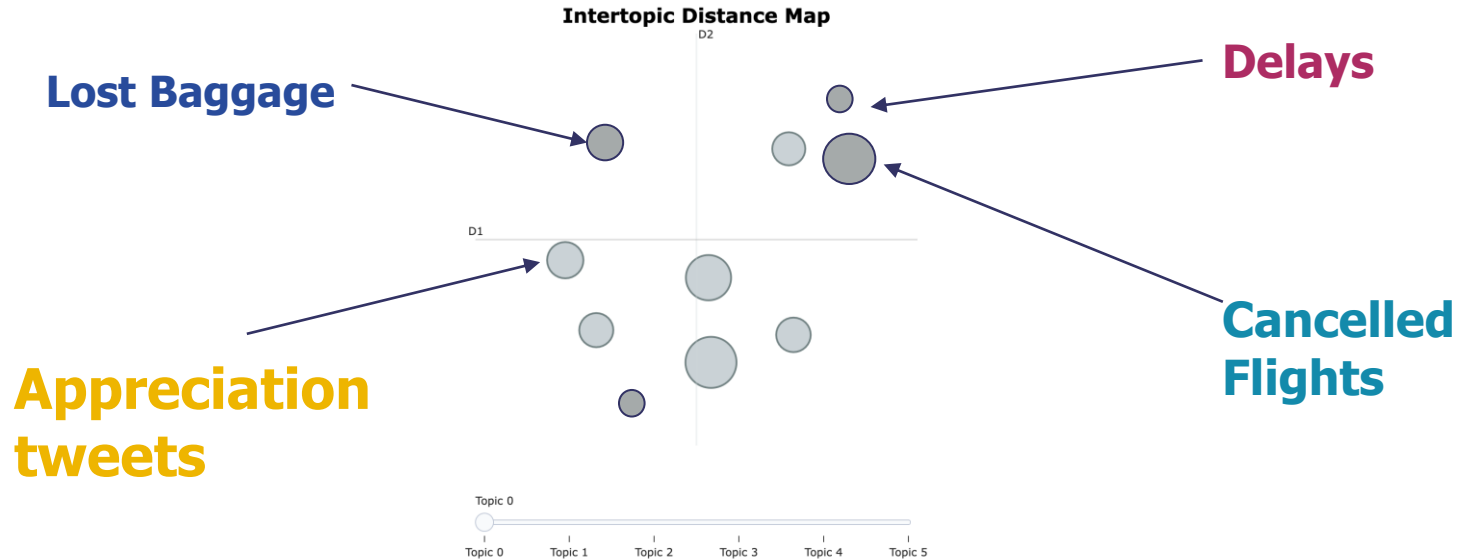
**Southwest Airlines** has the most positive sentiment percentage at **32%**

**United Airlines** positive tweet rate is **26%**, while **American Airlines** is **23%**



text	sentiment
united thanks	POSITIVE
united thanks for taking care of that mr happy customer.	POSITIVE
united still no refund or word via dm. please resolve this issue as your cancelled flightled flight was useless to my assistant s trip.	NEGATIVE
united delayed due to lack of crew and now delayed again because there s a long line for deicing... still need to improve service united	NEGATIVE
united thanks we filled it out. how s our luck with this is it common	POSITIVE

# Topic Modeling was leveraged to uncover what the customer bases care about across all airlines



**Within these topics we were able to detect trends and understand sticking points within the customer journey across all airlines**

Topic	Takeaways	Example Tweet
Baggage	Lost baggage issues and baggage check complaints overall	@AmericanAir Your customer service is deplorable. I am disgusted in your company and the ignorant people on the phones for lost baggage.
Change of plans	Cancellations, rebooking, and refund experiences are always top of mind	@united still no refund or word via DM. Please resolve this issue as your Cancelled Flightldt flight was useless to my assistant's trip.
Hold the Phone	Customer service wait times factored significantly	@united hence why I've been asking for the customer service phone number so I can speak to a real person to get this sorted out
Missed Connections	Delays and misconnects were a major topic across all customers	@AmericanAir many have missed connections already b/c of delayed flight which will finally board soon
Appreciate You	Good deeds do not go unnoticed. Appreciation and thank yous are returned	@united Thank you so much for your help with my birthday trip! Tickets are confirmed! :-)
Hold the Gate	The experience at the gate is as important as the flight experience too	@SouthwestAir First there were cockroaches crawling on the counter at the gate and visible mouse traps under the seats at the gate
Customer Service	Interactions with customer service also factored significantly	united not even mentioning how rude the customer service was to us. As a business owner, I'd be mortified if my employees acted as yours
Twitter Action	Taking actions (following, dm'ing, liking) to get ahold of customer service was common	@AmericanAir can u help rebook passenger via Twitter/DM. Been on hold for 1.5 hours. Thanks!
The Human Touch	Connecting with the flight attendants, crew, and gate agents also leave a strong impression	@AmericanAir Unbelievable that I cannot even wait on hold to speak to a human being to resolve my issue!!!!!! The system simply hangs up!
Duly Noted	Filing complaints also featured highly among tweets	@united worst flights I've ever had. ground crew ignored our plane, made me miss flight and now I have to cover the cost of a hotel. #DEN

# Competitor Specific

**Their** Brand Health

*What are **their** customers **talking about** and what is the **sentiment**?*

# Southwest surprise & delights customers with Twitter engagement & concert ticket giveaways, but cancellations are prominent

SWA Topic	% of tweets	Southwest Airlines Major Topics <i>(negative motivators)</i>	Sentiment Ratio <i>(Negative/Positive)</i>
0	26%	<b>Twitter specific actions and thank yous</b> <i>(driven by engagement and requests to follow to get critical updates)</i>	<div><div>30</div><div>70</div></div>
1	26%	<b>Cancelled flights, refunds, and rebooking</b> <i>(driven by lack of communication, sudden updates, and lack of empathy)</i>	<div><div>93</div><div>7%</div></div>
2	14%	<b>Customer service and wait times for call centers</b> <i>(driven by extremely long wait times, with customers suggesting they're "dying" at times)</i>	<div><div>90</div><div>10%</div></div>
3	13%	<b>In-flight Imagine Dragons Performance and giveaway</b> <i>(driven by a surprise event aimed to delight and surprise customers)</i>	<div><div>41</div><div>59</div></div>
4	11%	<b>Lost baggage</b> <i>(driven by policies that do not put the customer first, driving many to feel like they're being cheated)</i>	<div><div>86</div><div>14</div></div>
5	10%	<b>Negative customer service experiences</b> <i>(driven by lack of empathy and overall dissatisfaction with resolutions)</i>	<div><div>71</div><div>29</div></div>

# American has a polarizing effect on customers - struggles with call wait times & cancellations

AAL Topic	% of tweets	American Airlines Major Topics <i>(negative motivators)</i>	Sentiment Ratio <i>(Negative/Positive)</i>
0	31%	<b>Call wait times</b> <i>(driven by faulty automated systems and lack of true connectivity)</i>	<div><div>91</div><div>9</div></div>
1	21%	<b>Lost baggage</b> <i>(driven by logistical challenges, and inaccurate and confusing updates)</i>	<div><div>90</div><div>10%</div></div>
2	14%	<b>Appreciation and “thank yous”</b> <i>(driven by customers feeling a genuine sense of effort on the airline to resolve issues)</i>	<div><div>15</div><div>85</div></div>
3	13%	<b>Negative experiences with customer service</b> <i>(driven by significant call volume and extremely long wait times)</i>	<div><div>84</div><div>17</div></div>
4	12%	<b>Contacting AA via various mediums (email, call, app, twitter)</b> <i>(driven by immense frustration with a lack of a unified customer service experience)</i>	<div><div>73</div><div>27</div></div>
5	10%	<b>Cancelled flights</b> <i>(driven by a lack of communication, driving customers to feel left out and frustrated)</i>	<div><div>95</div><div>5%</div></div>



# United Focus

## Our Brand Health

*What are **our** customers talking about and how does the **sentiment** compare to our competitors?*

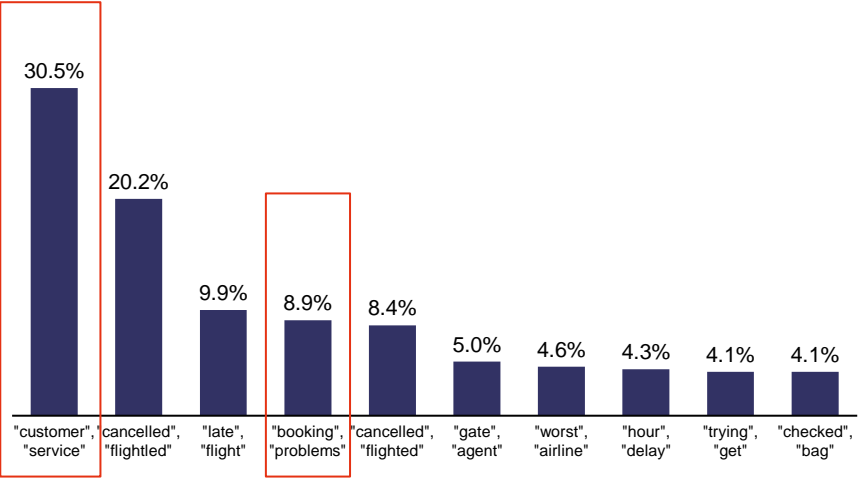
# While United’s strength is in appreciation the airline but struggles with lost baggage and cancellations

UAL Topic	% of tweets	United Airlines Major Topics <i>(negative motivators)</i>	Sentiment Ratio <i>(Negative/Positive)</i>
0	28%	<b>Lost baggage</b> <i>(driven by frequency, overweight capacity, flight delays, and rechecked bags)</i>	<div><div>83</div><div>17</div></div>
1	22%	<b>Appreciation and “thank yous”</b> <i>(driven by addressing travel-related stress and taking action)</i>	<div><div>35</div><div>64</div></div>
2	14%	<b>Cancellations and rebooking</b> <i>(driven by last minute cancellations, app glitches, and in-flight frustrations with crew)</i>	<div><div>93</div><div>7%</div></div>
3	13%	<b>Contacting customer service</b>	<div><div>79</div><div>21</div></div>
4	12%	<b>Delays and missed connections</b> <i>(driven by crew shortages, maintenance, connection challenges, and spurned plans)</i>	<div><div>91</div><div>9</div></div>
5	12%	<b>Gate, gate agents, and taxing delays</b> <i>(driven by crew shortages, maintenance, connection challenges, and spurned plans)</i>	<div><div>90</div><div>8</div></div>

# For United, flight cancellations rank highly, but rebooking and customer service frustrations represent even higher drivers of disappointment

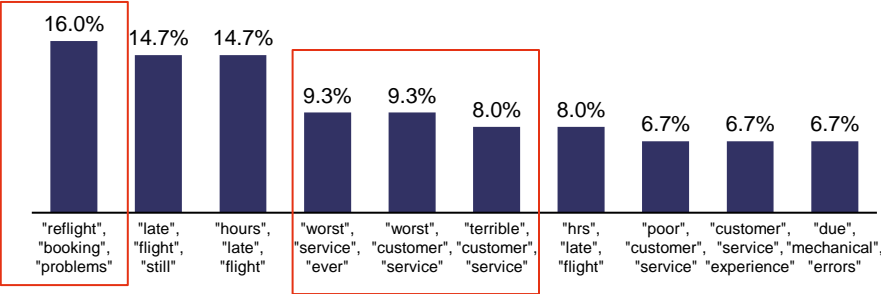
## United Airlines Negative Bigrams

% of Top 10 by Frequency



## United Airlines Negative Trigrams

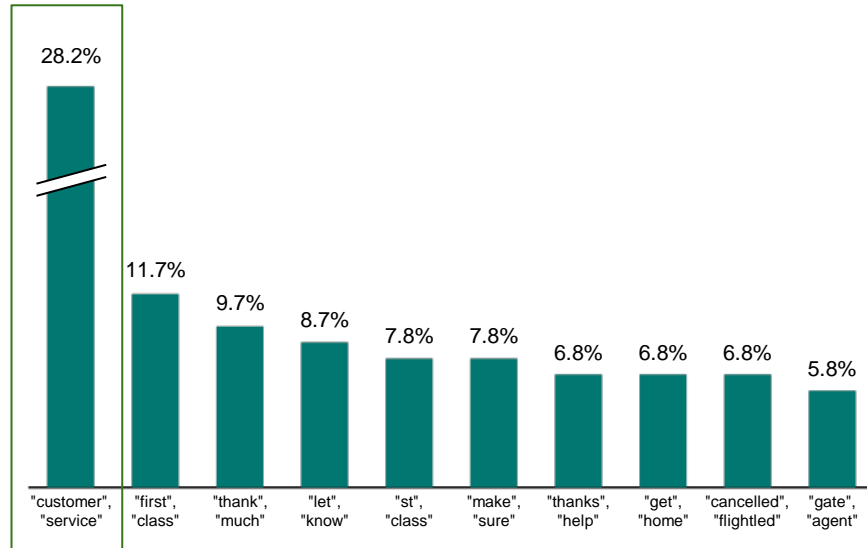
% of Top 10 by Frequency



# Positive sentiment points the same rebooking feedback - when people can easily rebook and are treated kindly, they are appreciative!

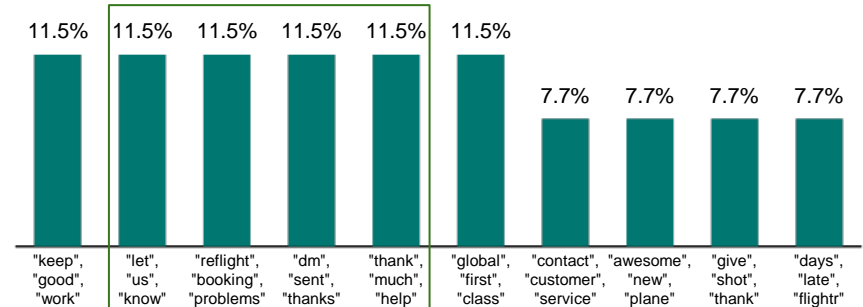
## United Airlines Positive Bigrams

% of Top 10 by Frequency



## United Airlines Positive Trigrams

% of Top 10 by Frequency



# Future work includes more analysis into the overall methodology used and other potential transformer models for classification purposes

---

## More analysis on model bias and overfitting

### Fine-tuning on SST-2 may affect performance on new data via overfitting

- Larger models, such as **DistillBERT** are **susceptible to overfitting**; smaller data sizes are drivers of this lapse in performance. It would be worth revisiting the regularization techniques used.
- **Fine-tuning on a larger dataset** than SST-2 could potentially improve performance.
- Additionally, experimenting on **adjusting the learning rate** may also improve performance. If the learning rate was too small, it may not learn enough from the data. If too large, it may have changed the model too much

## Optimize Methodology

### Improve model performance via indirect prompting and other improvements

- **Experiment with different tokenization techniques** and potentially padding/truncating the text to work better with DistillBERT.
- **Reframe the task to broaden sentiment**, moving beyond binary classification and attempting multiclassification (very positive, somewhat positive, neutral, somewhat negative, very negative).
- **Engage in a more thorough error analysis** to further understand where modifications can be made to the data or how to better frame the task at hand

## Consider different models/methodologies

### Consider leveraging state-of-the-art generative pre-trained transformers

- **While GPT-3 was used as a benchmark for this analysis**, additional research is necessary to understand if its unidirectional approach ultimately performs better than other bi-directional options.
- **GPT-3's size (175 billion parameters) enables it to capture a broad understanding of language**, which may ultimately improve performance. But what one gains in performance may not be worth the computational lift.
- **Experimentation with few-shot learning and instruction tuning** to further understand if GPTs can be adequately used for sentiment classification is also necessary.

## The bottom line - Control the Controllables | United Airlines

### Strategy & Execution

Transparency with a heavy dose of empathy

Gap to competition

Our competitors do a slightly better job with **transparency** and helping customers **access help** when issues arise from booking all the way to the gate

Where we can win

Create an **empathetic customer experience** to improve the overall experience

Empower customer service to become solutions managers when it comes to rebooking

Gap to competition

The competition does a better job helping customers regain **control** when travel plans change

Where we can win

Develop a better system to help customers **rebook their cancelled flights** seamlessly. Cancellations are inevitable, but frustration doesn't have to be!

Handle baggage issues with care

Gap to competition

Lost and mishandled baggage disproportionately impacted United customers

Where we can win

Develop a more **robust automated messaging system to track and update customers** on their bags journey from start to finish

Surprise and delight customers to drive loyalty

Gap to competition

Our competition is stronger at **delighting customers via special events and other partnerships**

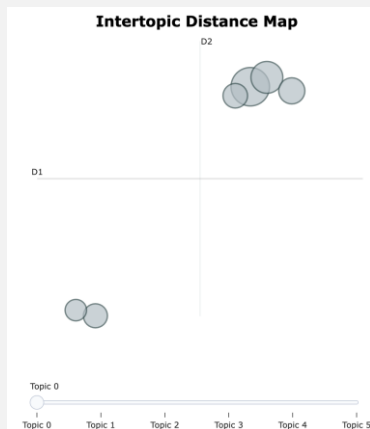
Where we can win

Enhance the typical flight experience **through stronger partnerships** to create events and positive buzz and drive new customer engagement

# Appendix

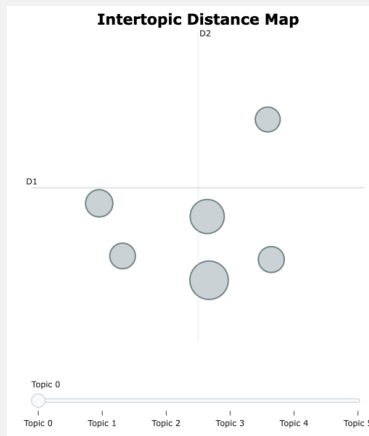
# Topic Modeling was leveraged to uncover core semantic structures in our twitter data set

## United Airlines



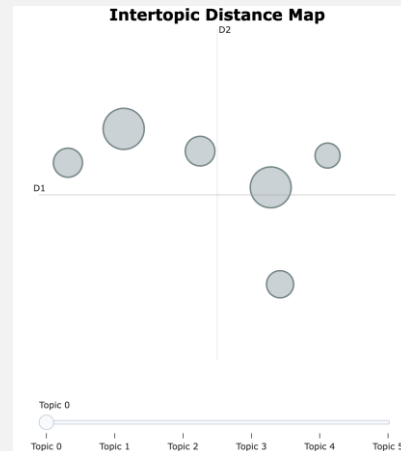
	Topic	Count	Name
0	-1	2727	-1_flight_get_co_http
1	0	331	0_bag_baggag_check_lost
2	1	227	1_thank_flight_plane_appreci
3	2	153	2_cancel_flight_flightl_get
4	3	135	3_custom_servic_care_thank
5	4	129	4_delay_flight_hour_time
6	5	103	5_gate_agent_wait_sit

## American Airlines



	Topic	Count	Name
0	-1	1922	-1_flight_get_cancel_aa
1	0	186	0_call_hour_hold_wait
2	1	146	1_bag_baggag_check_still
3	2	94	2_thank_appreci_much_done
4	3	85	3_servic_custom_line_terribl
5	4	83	4_email_respons_tweet_twitter
6	5	78	5_cancel_flight_flightl_dfw

## Southwest Airlines



	Topic	Count	Name
0	-1	1543	-1_flight_co_http_fli
1	0	226	0_thank_dm_follow_sent
2	1	224	1_cancel_flight_flightl_hold
3	2	118	2_hold_hour_call_phone
4	3	114	3_ticket_imaginedragon_show_would
5	4	99	4_bag_baggag_hour_lost
6	5	84	5_custom_servic_call_terribl



# Sample output of Sentiment Analysis and Topic Modeling

	tweet_id	airline	name	text	tweet_created	sentiment	topics
0	1	United	rdowning76	united thanks	2015-02-24 11:42:48	POSITIVE	4
2	3	United	CoralReefer420	united still no refund or word via dm. please resolve this issue as your cancelled flighted flight was useless to my assistant s trip.	2015-02-24 11:39:45	NEGATIVE	0
9	10	United	slandail	united gate agent hooked me up with alternate flights. if you have a way to prevent the constant issues that would rock.	2015-02-24 11:19:32	NEGATIVE	5
92	93	United	jakepoznak	united they helped me at the baggage service desk. said bc tsa screening was down in fil not all bags made it but bag will be in ewr	2015-02-24 08:31:20	NEGATIVE	1
96	97	United	ShawnaNewman	united thx for update	2015-02-24 08:27:37	NEGATIVE	4
97	98	United	Sri_Philips	united thank you	2015-02-24 08:26:30	POSITIVE	4
98	99	United	CATTastrophe	united dm these nuts	2015-02-24 08:22:52	NEGATIVE	8
103	104	United	jakepoznak	united despite my bag not making it to newark good informative email tracking updates help	2015-02-24 08:04:36	POSITIVE	1
104	105	United	RaquelSnyderDC	united. what s going on with ua outbound flight last thurs was delayed hrs how long will this delay be worst unitedflightsever	2015-02-24 08:03:25	NEGATIVE	3

# United Airlines Sentiment for each topic

topics	sentiment2	
-1	NEGATIVE	1928
	POSITIVE	825
0	NEGATIVE	272
	POSITIVE	52
1	NEGATIVE	145
	POSITIVE	15
2	NEGATIVE	123
	POSITIVE	33
3	NEGATIVE	132
	POSITIVE	12
4	NEGATIVE	100
	POSITIVE	34
5	NEGATIVE	117
	POSITIVE	17

```
{5: ['unit flight mke ord get delay gate attend commun ',
      'unit gate iad rdu ',
      'unit staralli gate far though ',
      'unit epic fail reagan jetway min tarmac ',
      'unit flight ua sit newark wait gate serious threehoursl flight ',
      'unit nicest gate attend newark airport gate right multipl delay still pleasant everyon '],
  4: ['unit thank program comment system assign seat behind even though pair avail chang back ',
      'unit mom left kindl flight today burgundi case light seat anyon find ',
      'unit seat avail pm ord rdu need seat pleas '],
  0: ['unit terribl illog rout cancel flightl connect made pay check bag lost bag ual ',
      'unit check carri watch handler throw bag miss convey belt sat min http co lyoox msh ',
      'unit fuck bag fuck fuck bag tell give number call human san '],
  1: ['unit realli need get act togeth realli disappoint chang flight delay flight etc ',
      'unit late flight everyon brave weather passeng wait flight crew realli ',
      'unit terribl flight experi happen everyon right flight ua delay hr total miss flight thnx ',
      'unit one suffer understand delay understand tri speed process delay extent ',
      'unit delay cs help notif ',
      'unit suck hour delay '],
  2: ['unit submit custom care form jan th still respons littl longer busi day ',
      'unit thank repli contact custom care get clarif issu ',
      'unit thank repli email custom care depart experi ',
      'unit custom custom servic http co qp aw nlip ',
      'unit better custom servic john wayn airport ',
      'unit custom servic ',
      'unit hung custom servic wait min hold guess vacat happen websit work ',
      'unit hour wait issu resolv contact custom servic comment wait time thank contact ',
      'unit thank rude custom servic hour delay realli help time wise afternoon meet hello delta '],
  3: ['unit chang dkyde gt instead gt gt origin reserv gt yesterday cancel flightl ',
      'unit offer flight cancel flight strand away home work ',
      'unit black white eu law cancel flightl flight want money ']]}
```

# Southwest Airlines Sentiment for each topic

topics	sentiment2	
-1	NEGATIVE	1111
	POSITIVE	532
0	NEGATIVE	232
	POSITIVE	19
1	NEGATIVE	58
	POSITIVE	59
2	NEGATIVE	15
	POSITIVE	102
3	NEGATIVE	83
	POSITIVE	18
4	NEGATIVE	86
	POSITIVE	6
5	NEGATIVE	56
	POSITIVE	31

```
{4: ['southwestair flight san diego delay per usual take home ',
'southwestair go stop fli delay everytim time due paperwork ',
'southwestair flight pilot spill tea radio yay delay almost hour '],
0: ['southwestair hold hour hung cancel flightl flight run daycar kid trap home ',
'southwestair bad weather multipl cancel flight flight us scare note next time ',
'southwestair need request refund flight cancel flightl morn due weather '],
2: ['southwestair thank touch ',
'southwestair thank keep check tri book first disney world vacat ',
'southwestair thank dming ',
'southwestair go airlin much say care see much right suck ',
'southwestair samoor thank kind respons acknowledg apolog go long way southwestrock ',
'southwestair thank follow appreci ',
'southwestair great la flight clarenc frank guy hoot thank ',
'southwestair hold music sound like super mario bro gameboy color ',
'southwestair thank day concert watch notic fist bump high five end rock version '],
3: ['southwestair bag make greenvil tonight forgiven highhop ',
'southwestair hr put tag bag sayin go greenvil instead raleigh kid ',
'southwestair thing free flight gonna spend board rental bc swa get baggag greenvil '],
1: ['southwestair gimm ',
'southwestair lacma fairfax ',
'southwestair ',
'southwestair love dragonss lol sorri see lauren oh god ahhhh ',
'southwestair love dragonss holi fuckinf shit ',
'southwestair love dragonss oh god lauren oh god oh god '],
5: ['southwestair fli southwest first time lga layov atlanta ',
'southwestair awesom first time fli ',
'southwestair ye much look forward use airlin futur flight never ']}
```

# American Airlines Sentiment for each topic

topics	sentiment2	
-1	NEGATIVE	1307
	POSITIVE	423
0	NEGATIVE	272
	POSITIVE	17
1	NEGATIVE	209
	POSITIVE	20
2	NEGATIVE	126
	POSITIVE	12
3	NEGATIVE	5
	POSITIVE	73
4	NEGATIVE	35
	POSITIVE	33
5	NEGATIVE	56
	POSITIVE	6

```
{0: ['americanair flight cancel flightl call said call back hung say wait hr ',
     'americanair made flight still updat email land info ever sent second flight lose biz ',
     'americanair brother flight dfw cancel flightl yesterday fli ont dfw today '],
 4: ['americanair check dm pleas ',
     'americanair dm info ',
     'americanair dm anyth ',
     'americanair help ',
     'americanair understand busi still gotten answer need get home guy help ',
     'americanair expedia kill hour get human help '],
3: ['americanair thank ',
     'americanair thank show ',
     'americanair thank assist ',
     'americanair great thank ',
     'americanair thank chang anyth end good ',
     'americanair great thank ',
     'americanair thank gener comput gener respons accomod travel instead say sorri ',
     'americanair ok thank let know appreci respons ',
     'americanair thank respons got resolv counter '],
1: ['americanair absolut unaccept call day hung everi time need minut bit help ',
     'americanair messag delay respond ',
     'americanair get thru anyon phone '],
2: ['americanair ye manag lose two bag horrend airlin ',
     'americanair bag arent even load plane flight ',
     'americanair bag yet locat th time check track bag hard system '],
5: ['americanair misunderstood usairway would day flight chang gate agent said ',
     'americanair need chang flight cs tell take call chang onlin pleas help flight week ',
     'americanair flight tri chang hour tri reach agent unbeliev ']]}
```

# United Airlines - Negative sentiment explanations

```
['because', 'manually', 'enter', 'baggage', 'tags', 'really', 'worst', 'cust']  
['because']  
['because', 'flights', 'overbooked']  
['because', 'glitch', 'notfair']  
['because', 'apparently', 'ua', 'overweight', 'amp', 'returned', 'gate', 'wtf']  
['because', 'would', 'rather', 'complain', 'public', 'forum']  
['because', 'flight', 'staff', 'police', 'overhead', 'bins']  
['because', 'print', 'piece', 'paper']  
['because', 'reach', 'anyone', 'deliver']  
['because', 'missed', 'connection']  
['because', 'every', 'opportunity', 'job', 'managed', 'fail', 'time', 'lazy']  
['because', 'one', 'knew', 'simple', 'jobs']  
['because', 'plane', 'overweight', 'great', 'knew', 'soon', 'landed', 'instead']  
['because', 'lowered', 'award', 'tickets', 'want', 'get', 'hard', 'earn']  
['because', 'talking', 'someone', 'thailand']  
['because', 'delay']  
['because', 'delay', 'stuff', 'happens', 'get', 'wish', 'handled', 'better']  
['because', 'united', 'booked', 'us', 'grumpykim']  
['because', 'snow', 'storm', 'denver', 'dc', 'waiting', 'pilot', 'get']  
['because', 'copilot', 'hours', 'flying', 'raleigh', 'disappointing']  
['because', 'catering', 'issues', 'poorservice']  
['because', 'staff', 'took', 'break', 'happy']  
['because', 'salt', 'floor', 'previous', 'passengers', 'want', 'fly', 'care']  
['because', 'site', 'errors', 'agents', 'ask', 'pin', 'see', 'platinum']  
['because', 'ua']  
['because', 'sent', 'dca', 'first', 'pvd', 'may', 'rent', 'car']  
['because', 'never', 'get', 'use', 'tsa', 'prescreen', 'paid']  
['because', 'head', 'pounding']  
['because', 'pressurization', 'failure', 'worst', 'pain', 'ever', 'felt', 'thought']  
['because', 'wait', 'bag', 'baggage', 'claim']
```

# American Airlines - Negative sentiment explanations

```
['because', 'flight', 'cancelled', 'flighttled', 'even', 'give', 'food', 'comp']
['because', 'jetway', 'broken', 'steps', 'planb', 'waiting', 'nearly', 'hour']
['because', 'aa', 'incompetence', 'poor', 'equip', 'checks']
['because', 'aa', 'delay', 'pending', 'cancelled', 'flighttlation']
['because', 'mechanical', 'issues', 'perhaps', 'consider', 'maintenance']
['because', 'industry', 'embarrassed']
['because', 'companies', 'job', 'guys', 'failed', 'first', 'place', 'upgrade']
['because', 'wait', 'another', 'hours', 'receive', 'automated', 'phone', 'call']
['because', 'mins', 'next', 'flight', 'sat', 'plane', 'hour']
['because', 'nightmare', 'knew', 'would']
['because', 'last', 'name', 'right', 'thing', 'reinstate', 'tickets']
['because', 'deaf', 'hold', 'long']
['because', 'airline', 'sucks']
['because', 'pilots', 'spend', 'hours', 'phone', 'get', 'ep', 'agent']
['because']
['because', 'gate', 'open', 'hope', 'miss', 'connection', 'logistics', 'americanview']
['because', 'family', 'severe', 'weather', 'excuse', 'call', 'make', 'change']
['because', 'evil', 'wants', 'soul']
['because']
['because', 'one', 'number', 'rubbed']
['because', 'thanks', 'fam']
['because', 'concerned', 'flight', 'make', 'able', 'leave', 'airport']
['because', 'aa', 'took', 'hrs', 'call', 'back', 'ended', 'paying']
['because', 'plane', 'toilet', 'working', 'needed', 'gas', 'flight', 'leaving']
['because', 'lines', 'busy']
['because', 'plane', 'old']
['because', 'weather', 'dallas', 'without', 'charged']
['because', 'add', 'ktn', 'asked', 'numerous', 'times', 'one', 'answers']
['because', 'accept', 'us', 'card', 'way', 'nyc']
['because', 'status', 'disgusted']
['because', 'buy', 'first', 'class', 'delta', 'learn']
['because', 'worked']
['because', 'automated', 'system', 'keeps', 'hanging', 'two', 'days', 'help']
```

## Southwest Airlines - Negative sentiment explanations

```
['because', 'expected', 'snow']
['because', 'first', 'time', 'forgot', 'look', 'ids', 'brutal']
['because', 'fat', 'whatever', 'reason', 'thanks']
['because', 'meeting', 'today']
['because', 'badpolicy', 'allows', 'cheatcustomers']
['because', 'cousin', 'amp', 'probably', 'sit', 'together', 'flight', 'tomorrow']
['because', 'follow', 'notanymore', 'wrong', 'badpolicy']
['because', 'playing', 'detective', 'trying', 'figure', 'one', 'atl']
['because', 'abused', 'every', 'airport']
['because', 'roads', 'snow']
['because', 'everyone', 'deserves', 'redcarpet', 'treatment', 'even', 'mom', 'co']
['because', 'flight', 'cancelled', 'flightled', 'apparent', 'reason', 'told', 'sorry']
['because', 'equipment', 'working', 'job', 'get']
['because', 'pink', 'paper', 'needed', 'white']
['because', 'opposed', 'flying', 'las', 'future']
['because', 'equipment', 'failure', 'compensation', 'offered']
['because', 'yall', 'cancelled', 'flightled', 'return', 'flight', 'answer', 'phone']
['because', 'volume', 'high']
['because', 'cavalli', 'calves', 'interfering', 'onboard', 'equipment']
['because', 'lied', 'told', 'ice', 'runway', 'every', 'carrier', 'able']
['because', 'according', 'flight', 'attendant', 'husband', 'talk', 'english', 'fact']
['because', 'follow']
['because', 'flight', 'progress']
['because', 'gate', 'attendant', 'let', 'someone', 'board', 'whose', 'boarding']
['because', 'awful', 'service', 'summer']
['because', 'need', 'follow', 'back']
['because', 'bags', 'loaded', 'correctly', 'gives']
['because', 'told', 'flythroughs', 'good', 'flight', 'otherwise']
['because', 'responsible', 'adult', 'got', 'number', 'call', 'ty']
['because', 'everyone', 'late', 'flight', 'co', 'boh', 'mh', 'cb']
['because', 'accommodate', 'rebook', 'timely', 'manner', 'horrible']
['because', 'routed', 'total', 'passengers', 'impossible', 'triple', 'connection', 'well']
['because', 'phone', 'dying', 'customerservice']
```