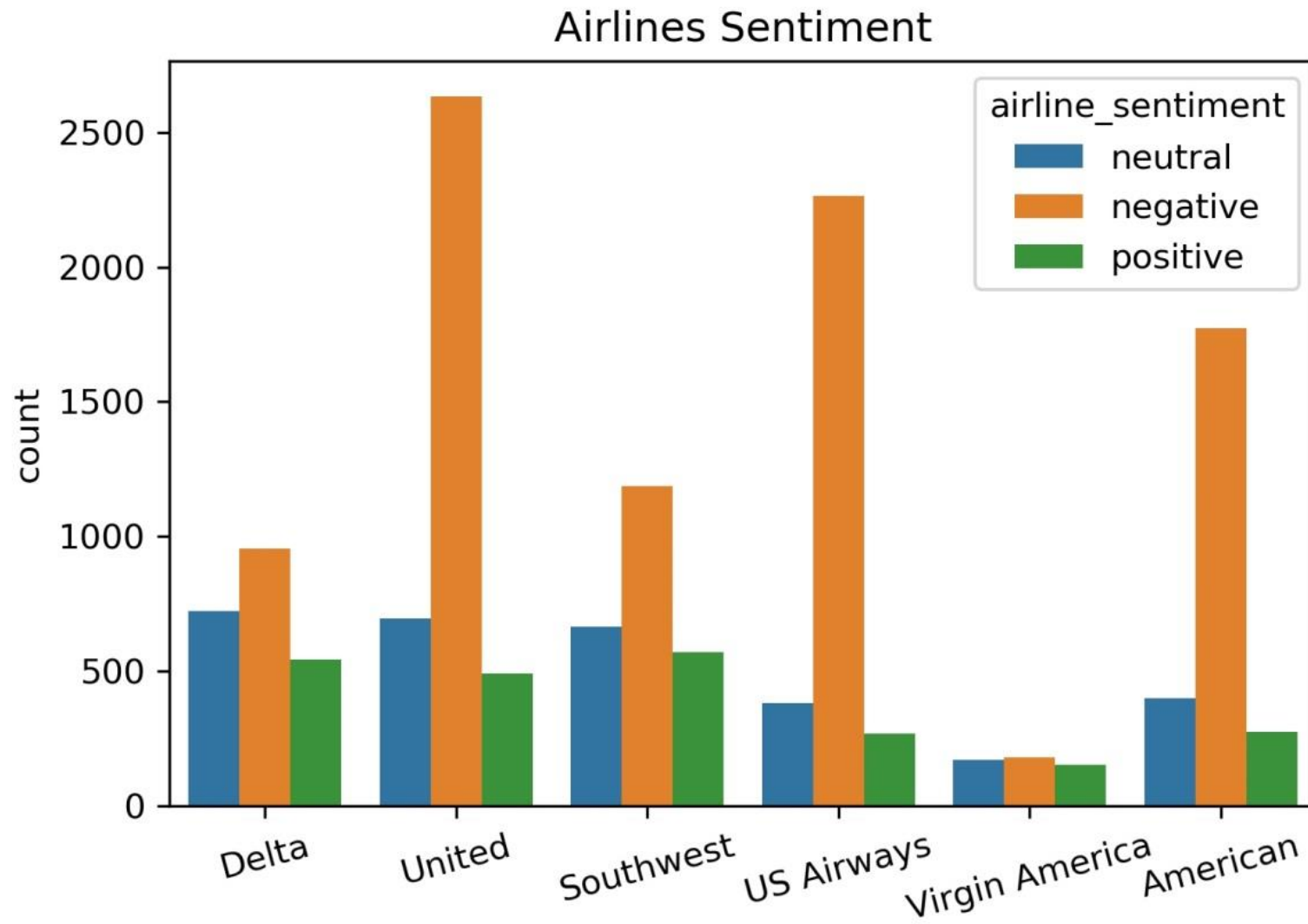


# **TWITTER US AIRLINE SENTIMENT ANALYSIS**


BY WILLIAM DEW

# DATA



# DATA WRANGLING

## Tweet cleaning steps

- 1.Convert text to lower-case
  - 2.Remove URLs
  - 3.Remove usernames
  - 4.Remove punctuation
  - 5.Remove stopwords
- 

# WORD BALLOONS


The figure displays three word clouds, each representing a different sentiment. The first cloud (Negative) features words like 'flight', 'plane', 'help', 'hour', 'customer service', and 'bag'. The second cloud (Neutral) includes 'please', 'ticket', 'im', 'light', 'time', 'need', 'change', and 'reservation'. The third cloud (Positive) shows 'thank', 'awesome', 'great', 'love', 'amazing', 'service', and 'crew'.

Negative

Neutral


Positive

# WORD BALLOONS




A word cloud representing negative sentiment. The most prominent words are 'flight', 'plane', 'time', 'bag', 'help', 'hour', 'get', 'customer', and 'service'. Other visible words include 'delayed', 'cancelled', 'flight', 'luggage', 'waiting', 'gate', 'call', 'agent', 'told', 'minute', 'im', 'cancelled', 'flight', 'luggage', 'waiting', 'gate', 'call', 'agent', 'told', 'minute', 'im', 'cancelled', 'flight', 'luggage', 'waiting', 'gate', 'call', 'agent', 'told', 'minute', 'im'.

Negative



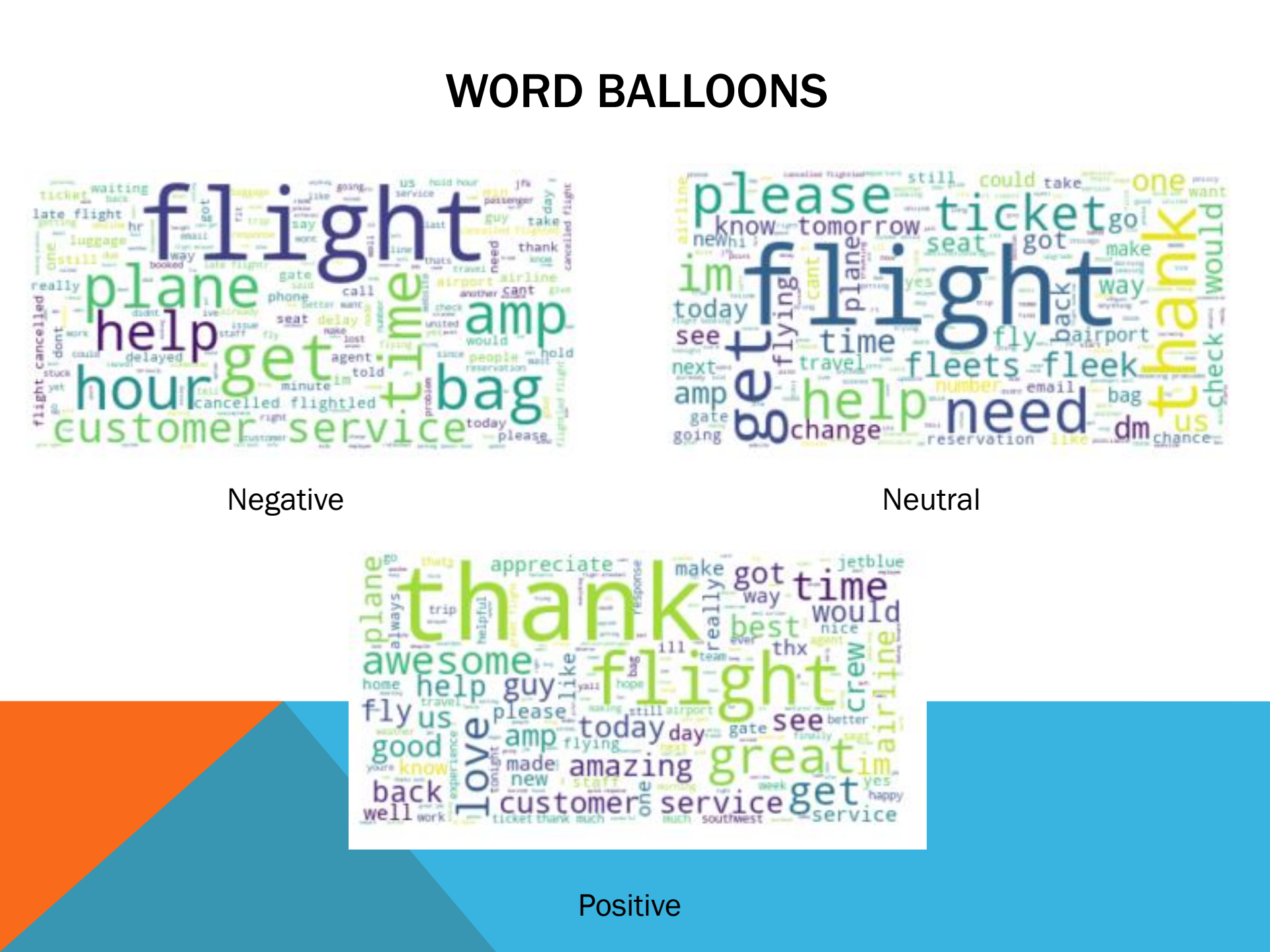
A word cloud representing neutral sentiment. The most prominent words are 'flight', 'plane', 'time', 'bag', 'help', 'hour', 'get', 'customer', and 'service'. Other visible words include 'delayed', 'cancelled', 'flight', 'luggage', 'waiting', 'gate', 'call', 'agent', 'told', 'minute', 'im', 'cancelled', 'flight', 'luggage', 'waiting', 'gate', 'call', 'agent', 'told', 'minute', 'im', 'cancelled', 'flight', 'luggage', 'waiting', 'gate', 'call', 'agent', 'told', 'minute', 'im'.

Neutral

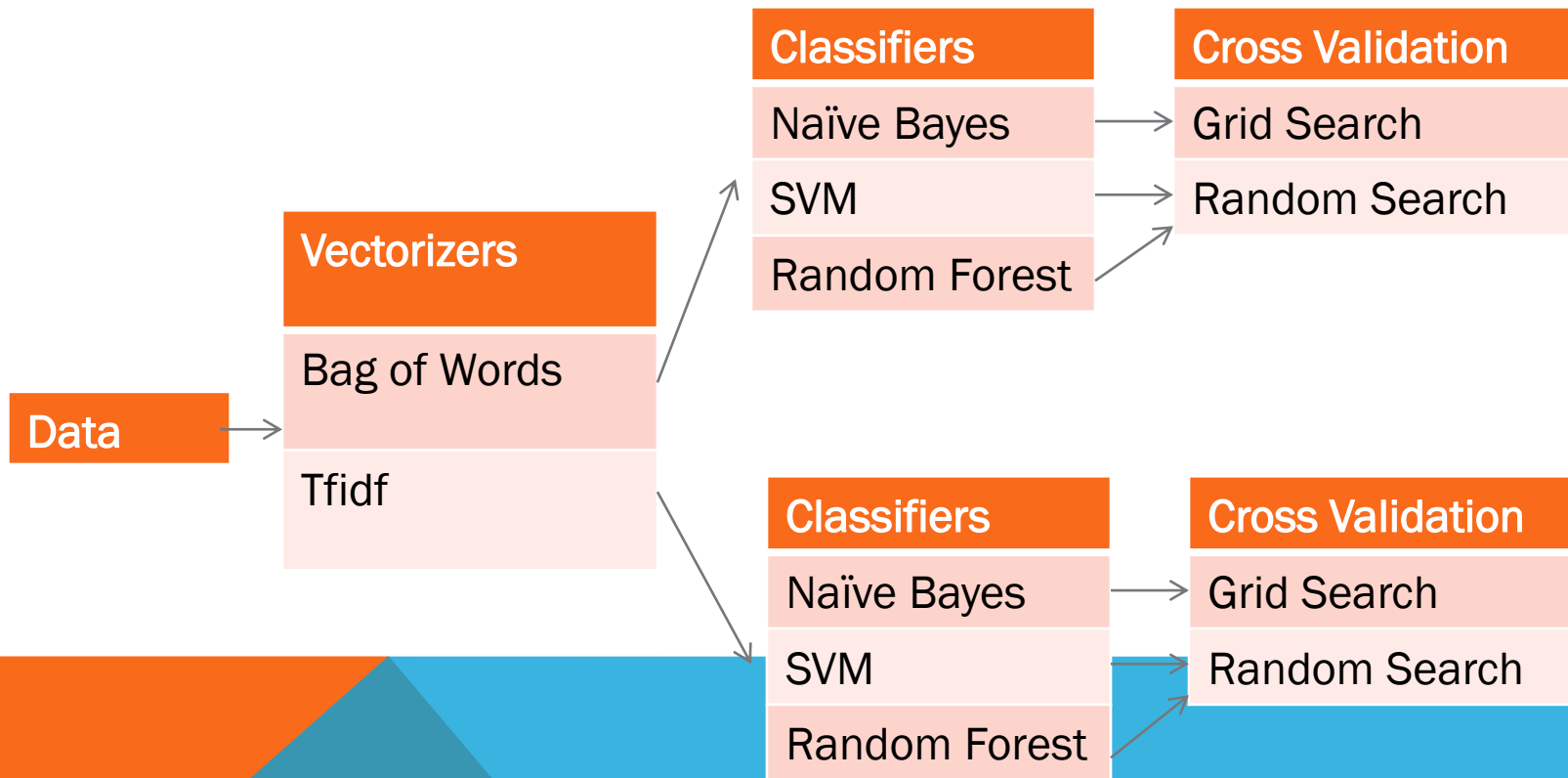


A word cloud representing positive sentiment. The most prominent words are 'thank', 'flight', 'time', 'great', 'service', 'customer', 'love', 'awesome', 'help', 'guy', 'like', 'please', 'today', 'day', 'gate', 'see', 'better', 'finally', 'great', 'im', 'happy', 'well', 'work', 'ticket', 'thank', 'much', 'much', 'southwest', 'service', 'service'.

Positive

[illegible][illegible]

# MODELS



# CROSS VALIDATION


Vectorizer	Classifier	Training Set F1	Validation Set F1
Bag of Words	Naïve Bayes	0.969 (0.002)	0.634 (0.009)
<i>Bag of Words</i>	<i>SVM</i>	<i>0.968 (0.001)</i>	<i>0.707 (0.010)</i>
Bag of Words	Random Forest	0.991 (0.001)	0.701 (0.012)
Tfidf	Naïve Bayes	0.985 (0.001)	0.639 (0.009)
Tfidf	SVM	0.880 (0.001)	0.695 (0.005)
Tfidf	Random Forest	0.990 (0.001)	0.698 (0.011)

# TEST SET RESULTS

Vectorizer	Classifier	Precision	Recall	F1
Bag of Words	SVM	0.731	0.699	0.713

- Test Set F1 score (0.713) is higher than Train (0.707) and Validation (0.707) F1 scores
- Precision and Recall given to show model performance

## FURTHER INQUIRIES

1. Explore URLs and usernames
  2. Validate more hyper parameters for Bag of Words and Tfidf
  3. Look into oversampling and downsampling data
  4. Build model with non-text features
- 
- A decorative graphic at the bottom of the slide consisting of three overlapping triangles. The leftmost triangle is orange, the middle one is a darker blue, and the rightmost one is a lighter blue. They are arranged in a way that creates a sense of depth and modern design.