Sample Case Study:
Wendy's and McDonald's Twitter Comments

William Kretz

Presentation Layout

- 1. Case Study Overview
- 2. Approach & Methods
- 3. Results
- 4. Recommendations
- 5. Limitations of the Data
- 6. Next Steps – Other Public Data
- 7. Conclusion

Case Study Overview

- Evaluating social media comments for 2 restaurant chains
- Sample Data:
  - 300 twitter mentions of Wendy's
  - 300 twitter mentions of McDonald's
- Goal: Find key insights that brand leadership might be interested in knowing

Approach & Methods

- 1. Clean data to prepare for analysis (tm package)
  - Extract tweets from excel file so they can be analyzed with R
  - Remove punctuation, capitalization, etc.
  - Combine words that go together (e.g. ice cream)
  - Example of Tweets:

Somebody bring me a two for three McDonald o got the �üí∞�üí∞�üí∞�üí∞�üí∞ the close at 10:30
Seeing everyone having mcd's and Im over here like eating healthy �üòÇ be worth it in the long run easy thenthough bye junk food notmissed @all

- 2. Visualize contents of data – Looking for things that stand out
  - Find popular topics for each brand by making word clouds

Approach & Methods - Continued

- 3. Conduct sentiment analysis
  - Measure how positive or negative each tweet is based on the contents
  - Determine whether a brand has a positive or negative social media image
  - Use a logistic model to try to predict sentiment based on words in a tweet
- 4. Cluster words together with word2vec Analysis
  - Use popular topics to find terms that are similar and associated with those topics
  - Use Kmeans clustering to compare the two brands based on their word vectors
  - Cluster popular topics that are shared by the brands to compare social media posts

Data Visualization – Word Clouds

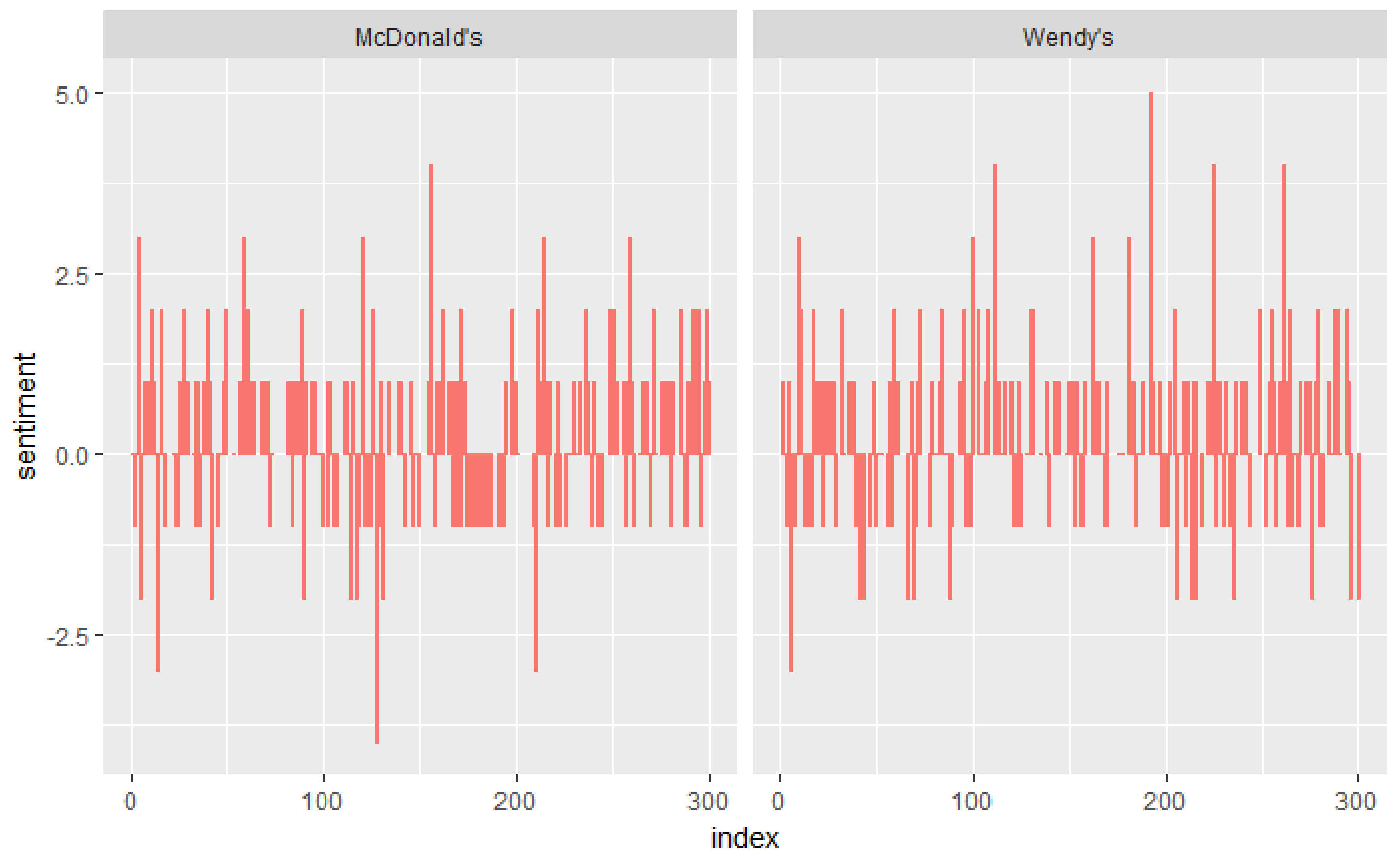
- Clean data with the tm package
- Generate visuals with the wordcloud package
- Size of word determined by frequency
- Result: Both restaurants have coffee and ice cream/frosty dominating the online conversation

- McDonald's Word Cloud

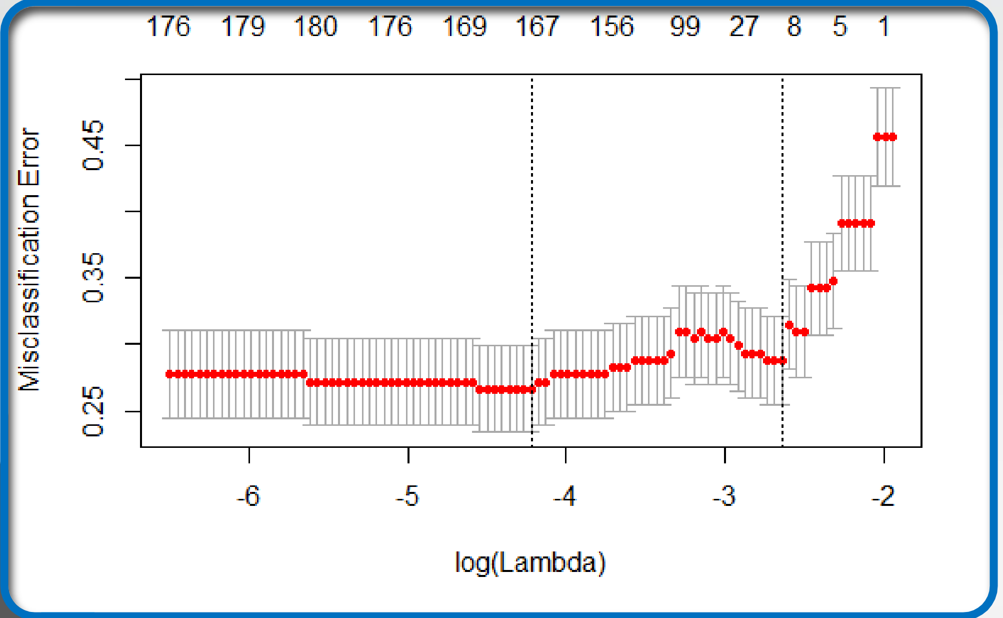
- Wendy's Word Cloud

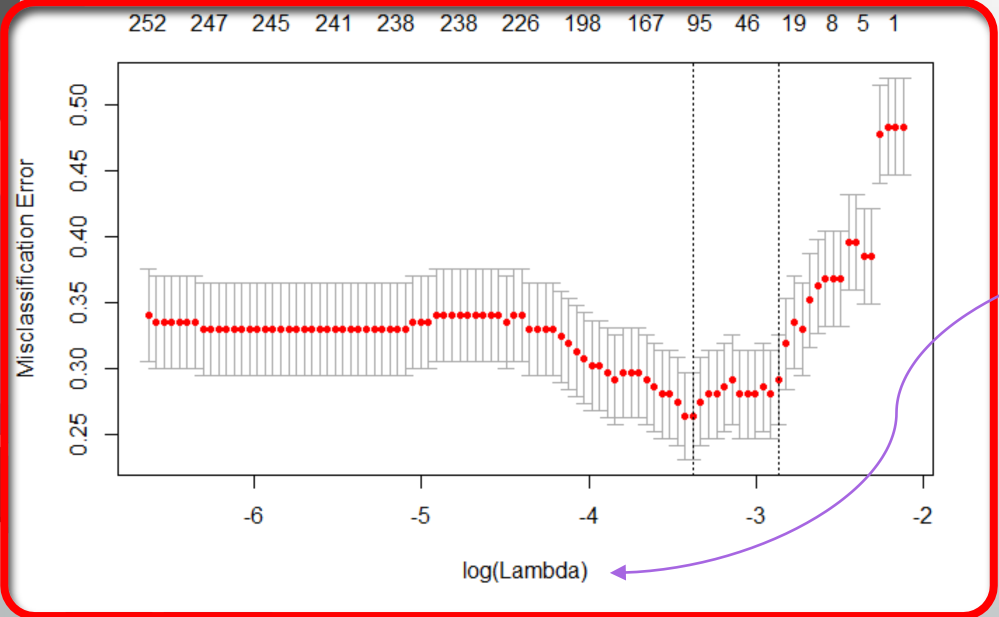
Results – Sentiment Analysis

- Utilize the tidytext package
- Observe 2 emotions – positive & negative using the ‘bing’ dictionary
- Result: Inconclusive trends without more data, but Wendy’s and McDonald’s had similar net sentiments of +63 and +61, respectively



Predicting Sentiment– Lasso Method





- Turn the calculated sentiment into a binary response:
  - 0 if sentiment <1
  - 1 if sentiment >1
- Use the count of each term as a predictor
- Utilize Leave-One-Out Cross Validation with glmnet package to fit logit model
- Penalty term determined at misclassification minimum
- Predict if a tweet is positive (1) or negative (0)

$$\sum_{i=1}^{n}\left(y_i-\beta_0-\sum_{j=1}^{p}\beta_j x_{ij}\right)^2+\lambda\sum_{j=1}^{p}|\beta_j| = \mathrm{RSS}+\lambda\sum_{j=1}^{p}|\beta_j|$$

Results - Lasso Logit Model

Wendy's	
Term	**Model Coefficient**
s***	-1.888512e+00
tired	-1.875498e+00
broke	-1.483457e+00
icecream	-1.275532e+00
used	1.433030e+00
thank	1.229372e+00
like	1.177548e+00
offer	1.071537e+00
free	9.586808e-01
used	1.433030e+00

McDonald's	
Term	**Model Coefficient**
broken	-4.337562e+00
expensive	-4.170277e+00
f***	-3.804703e+00
costs	-3.683289e+00
crack	-3.484951e+00
good	1.165448e+00
work	9.840688e-01
worth	7.242181e-01
right	6.348992e-01
best	5.999352e-01

$$logit(\pi(\boldsymbol{X})) = ln\left(\frac{\pi(\boldsymbol{X})}{1-\pi(\boldsymbol{X})}\right) = \beta_0 + \boldsymbol{\beta}^T\boldsymbol{X}$$

- Look at which words contribute to a negative and positive tweets
- McDonald's most negative contribution comes from 'broken'
- The presence of 'broken' in a tweet is associated with an decrease of -4.3 in the log odds of the tweet being positive for McDonalds or a multiplicative decrease in odds of 0.014
- S*** has the most negative effect on Wendy's tweets

Word2Vec Analysis

- Use word2vec to find which words are associated with each other
- Matrix of word vectors
- Each column and row of the matrix corresponds to a term in the dictionary
- At the intersection of a column and row, the element is the weight
- The weights are a measurement of how well the row term describes the column term after being trained on the data set

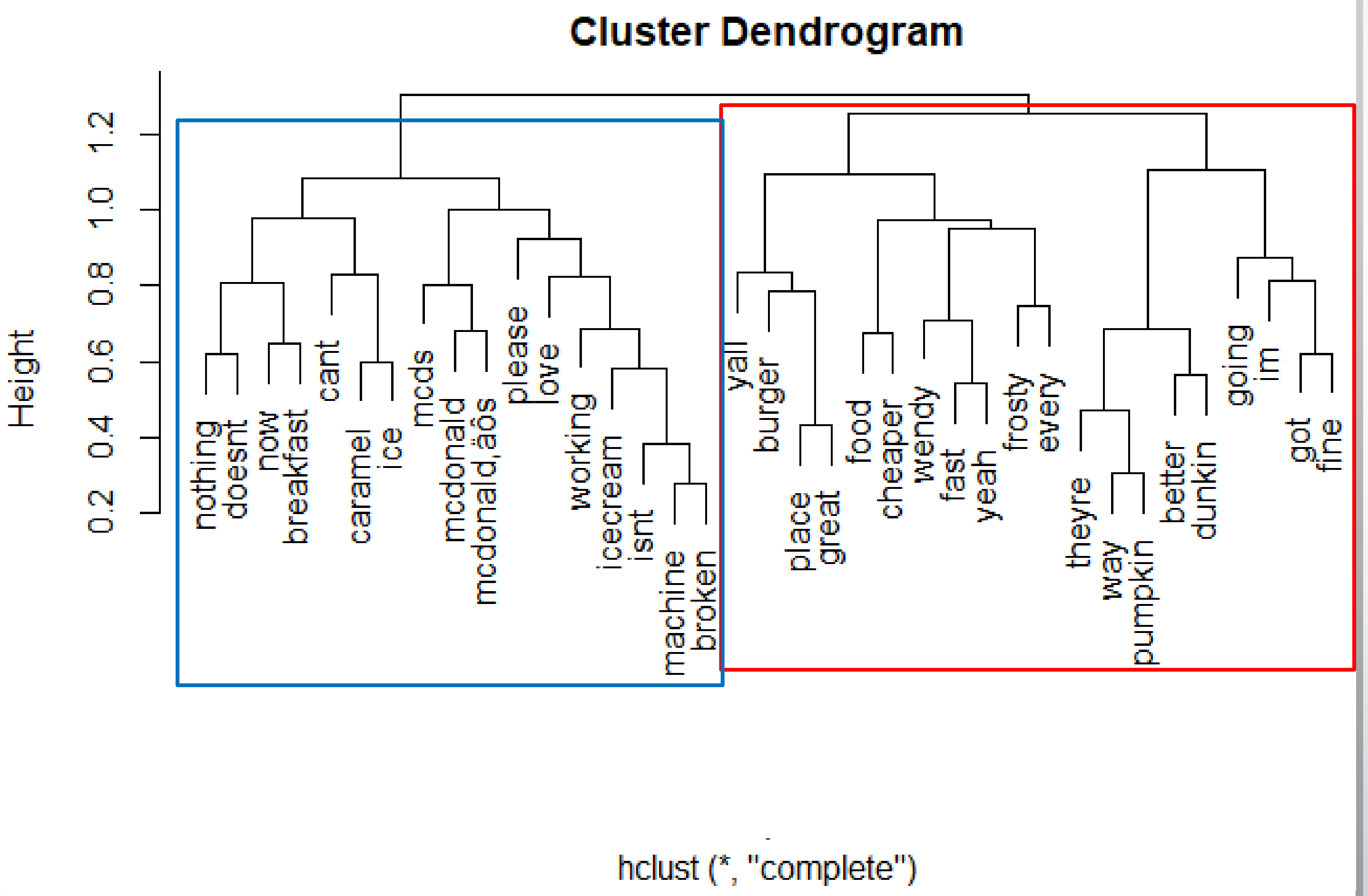
	Dog	Cat
Dog	1.00	0.01
Cat	0.01	1.00
Bark	0.98	0.02
Purr	0.02	0.99
Run	0.63	0.54

Results - Word2Vec Analysis

- Use word2vec to determine which words are similar
- Words that appear together often will have higher similarity

Coffee – Wendy's		Coffee – McDonald's		Frosty – Wendy's		Ice Cream – McDonald's	
Term	Similarity	Term	Similarity	Term	Similarity	Term	Similarity
coffee	1.0000000	coffee	1.0000000	frosty	1.0000000	icecream	1.0000000
guys	0.4763786	cream	0.5141529	coupons	0.5064501	broken	0.5107056
s***	0.4014006	iced	0.4954903	oh	0.4625504	machine	0.4990655
value	0.3815113	id	0.4872462	thank	0.3796040	isnt	0.4579619
can	0.3769041	asked	0.4663842	wait	0.3774013	working	0.4374439
offer	0.3763297	cash	0.4029563	eat	0.3583451	3	0.4234742
starbucks	0.3688373	discount	0.3812199	value	0.3538442	said	0.3706398
menu	0.3434407	work	0.3786789	fries	0.3505563	get	0.3640188
roast	0.3379123	best	0.3595603	wendy	0.3500079	broke	0.3514559
cash	0.3260823	starbucks	0.3587209	day	0.3491884	wendy	0.3333257

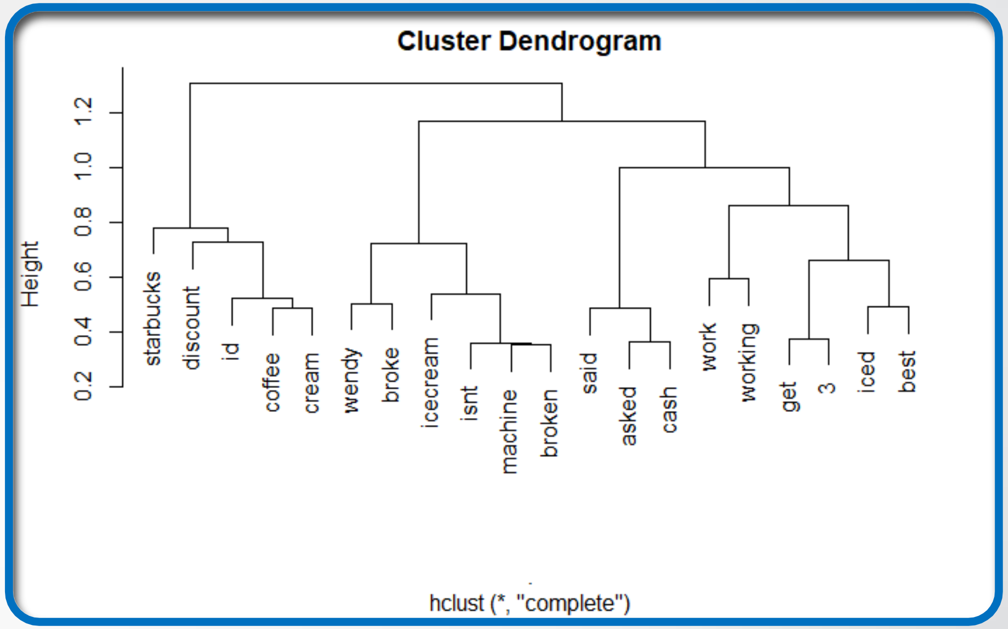
Results -Cluster Analysis

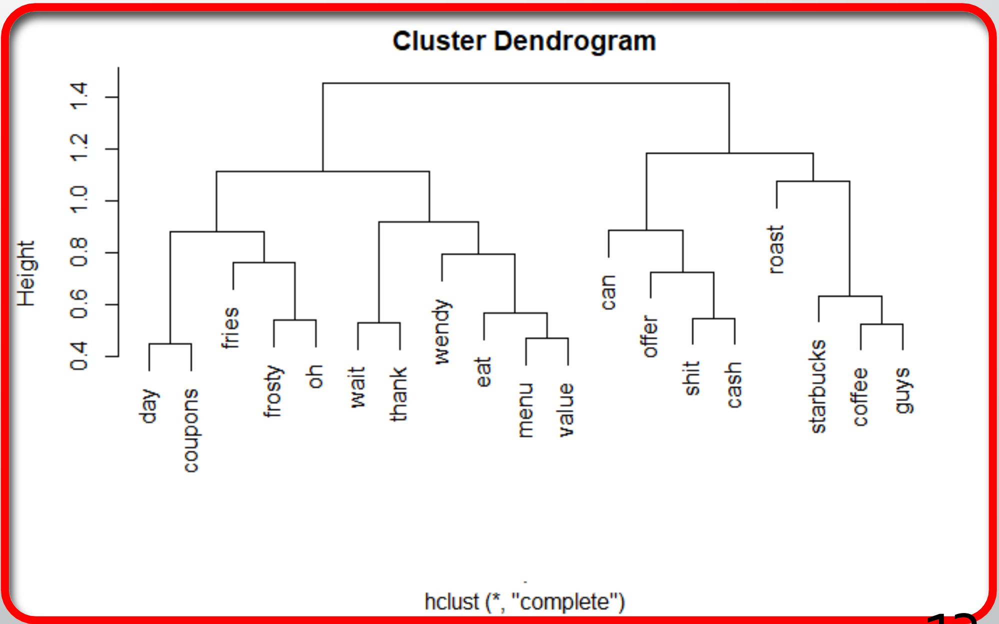


- Use Kmeans clustering with the word2vec results to find differences between the conversations associated with each brand
- The **left branch**, encapsulated in blue, represents the McDonald's conversation; the **right branch**, encapsulated in red, represents the Wendy's conversation

Results - Cluster Analysis Ice Cream vs Coffee

- McDonald's - Top
  - Coffee associated with 'starbucks' and 'cream'
  - Ice cream again associated with 'broken' 'machine'
  - When the ice cream machines are broken, customers also reference Wendy's - a competitor of McDonald's
- Wendy's - Bottom
  - Coffee is associated with 's***', 'guys', and 'starbucks'
  - Their frosty has a better image than the coffee, and is associated with 'fries' and 'coupons'
- Both brands have Starbucks in their coffee cluster





Summary of Results

- Frequency Word Clouds
  - both brands have coffee and ice cream/frosty driving their twitter conversation
- Sentiment – Qualitative Analysis
  - Without more data, there is no clear trend with sentiment
  - Both brands have a mix of positive and negative tweets
- Sentiment – Logit Model
  - For Wendy's, 's***' has the largest negative influence on tweets
  - For McDonald's, 'broken' has the largest negative influence on tweets
- Word2Vec – Similarity Clustering
  - Wendy's coffee associated with 's***' and 'Starbucks', a competitor
  - McDonald's coffee associated with neutral descriptors like 'cream' and 'Starbucks'
  - Wendy's frosty was associated with 'coupons'
  - McDonald's ice cream associated with 'broken', 'machine', 'isn't', and 'working'

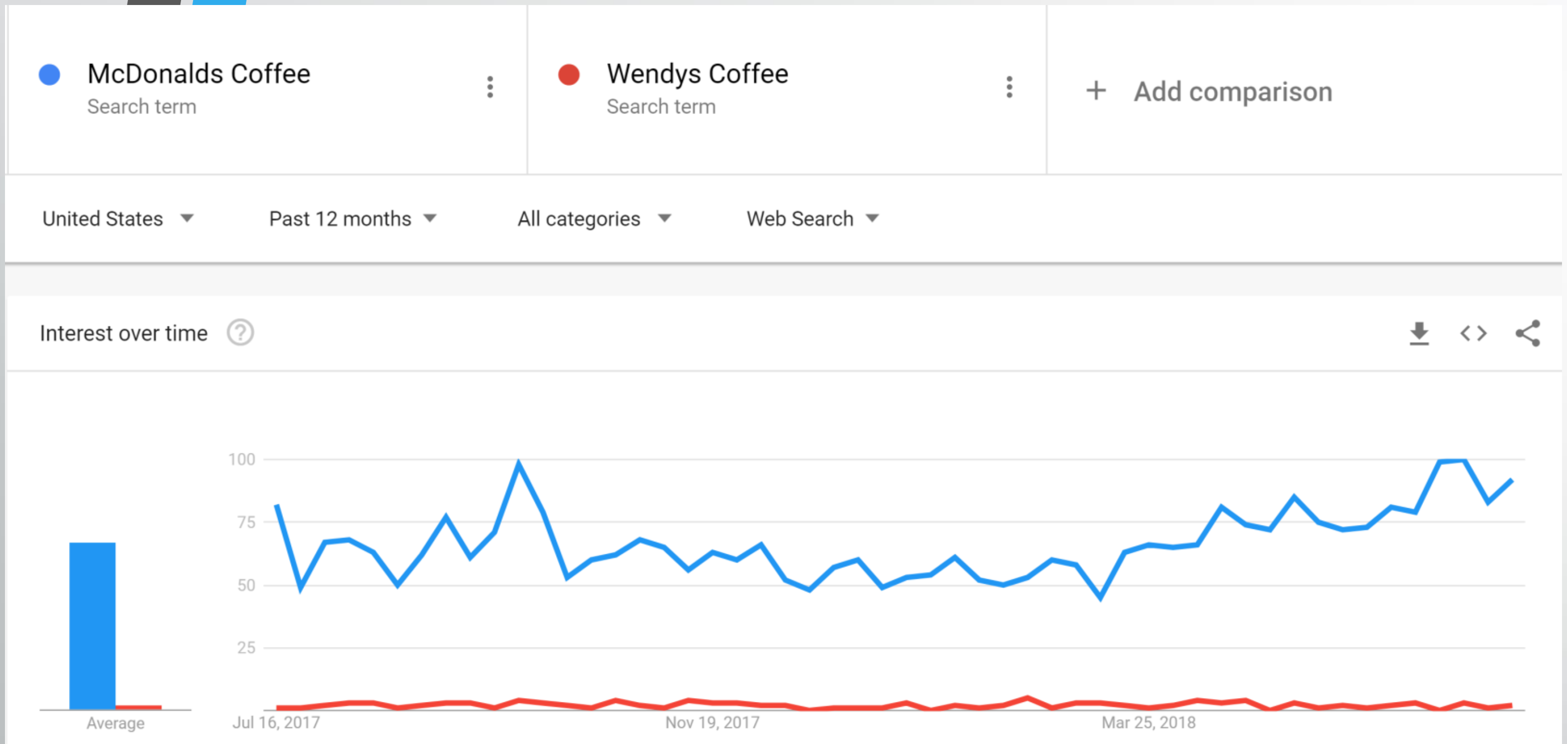
Recommendations

- McDonald's
  - McDonald's should look at improving ice cream machines because a large part of the online conversation is that they are broken
  - McDonald's coffee is popular on social media, but they should look at what Starbucks offers since customers tend to compare coffee with them
- Wendy's
  - Wendy's should look at improving their coffee because it is often associated with complaints on social media and it is their 2nd most popular topic
  - Wendy's should look at what Starbucks offers since customers tend to compare coffee with them
  - Customers like to tweet about the frosties and often associate them with coupons, so Wendy's should look at ways to bring customers in with coupons and keep them coming back to pay full price

Limitations of the Data

- Problems with Dataset
  - No time data -> would allow to track changes in sentiment or topic relevance over time
    - Example) Are ice cream machines broken all year long or just during the summer?
  - No location data -> would allow for testing if sentiment or topic relevance between regions differs
    - Example) Do customers complain about broken ice cream machines at the same rates in NJ and FL?
  - No Username/Profile -> would allow tracking of how often a specific customer complains on twitter
    - Example) Are all negative tweets posted by the same person or are many people complaining?
  - No Likes/Retweets Info -> Track the influence of a tweet
    - Example) Did a celebrity tweet a complaint that was then retweeted and seen by 1000s of people, or did the complaint go unnoticed?

Next Steps - Other Public Data



- Facebook data
  - Public commentary
  - Franchise pages
  - Follower demographics
- Google trends
- HealthData.gov
- National Climate Data Center
- Another source for the brands (though nonpublic) is their app users

Conclusion

- Both McDondald's and Wendy's social media conversations are heavily focused on their ice cream options and their coffee
- McDonald's should invest in improving their ice cream machines and try to promote their coffee as a better option to consumers
- Wendy's should improve their coffee while finding ways to attract people to their frostys without having to rely on coupons
- The data was just the text of tweets, having access to the meta data would allow for further analysis like trending the topics for a time analysis or comparing the sentiment between different regions

References

- Word2vec
  - https://blog.acolyer.org/2016/04/21/the-amazing-power-of-word-vectors/
  - https://github.com/bmschmidt/wordVectors/blob/master/vignettes/introduction.Rmd
- Sentiment Analysis
  - http://uc-r.github.io/tidy_text
- Lasso & glmnet
  - https://web.stanford.edu/~hastie/glmnet/glmnet_alpha.html
- James, G. (2014). *An introduction to statistical learning: With applications in R.* New York: Springer

Thank you!

Questions?