Analytics Accelerator – McDonald's

Complaint Data Analysis

Wharton Customer Analytics Student Team

Team 1 – Understanding Customer Complaint Data

Yung-Tang Chou
Engagement Lead

Joanna Yang
Senior Technical Analyst

Alyssa Nie
Senior Technical Analyst

Jeffrey Pan
Junior Technical Analyst

Qianqian Zhang
Junior Technical Analyst

Problem Statement: How can we quantify a complaint?

Customer Complaint
"I was so frustrated with not being able to use my App to order food that I decided not to order at all…Needless to say with the consistent issues that I have with this app and the inconsistency of this location honoring those who are attempting to use the app, I I am a few steps away from no longer being a consumer for this particular branded restaurant…Disappointed."

???

Modeling Techniques for Text Data

Sentiment Analysis

- Instead of calculating the "count" of complaints based on the source or their status, we need one go-to metric to evaluate these analysis. Sentiment trend is one of the primary quantification for text data, and is a good starting point to dive into issues in unstructured text data

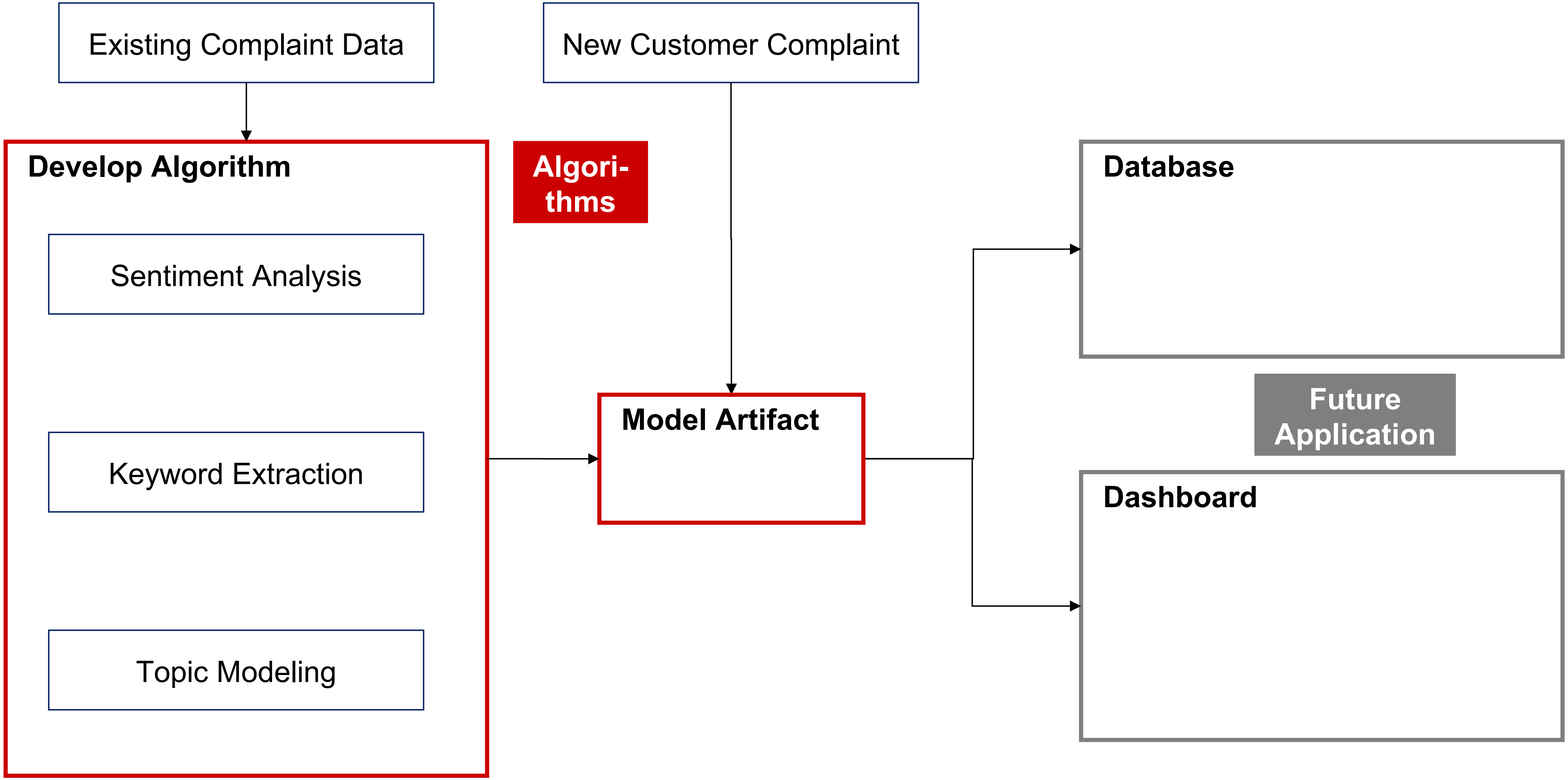
Keyword Extraction

- Word that appeared the most frequently doesn't indicate importance.
- Instead of looking at metrics for "frequency", we should look at statistics related to the "important" element, keywords, in the customer complaints

Topic Modeling

- Each corpus is constituted by some "topics", and each topic includes many "words"
- From the model we can create "latent topics" that helps us identify unidentified categorization of issues and monitor if there's any shift in complaint types.

Project Proposal Diagram

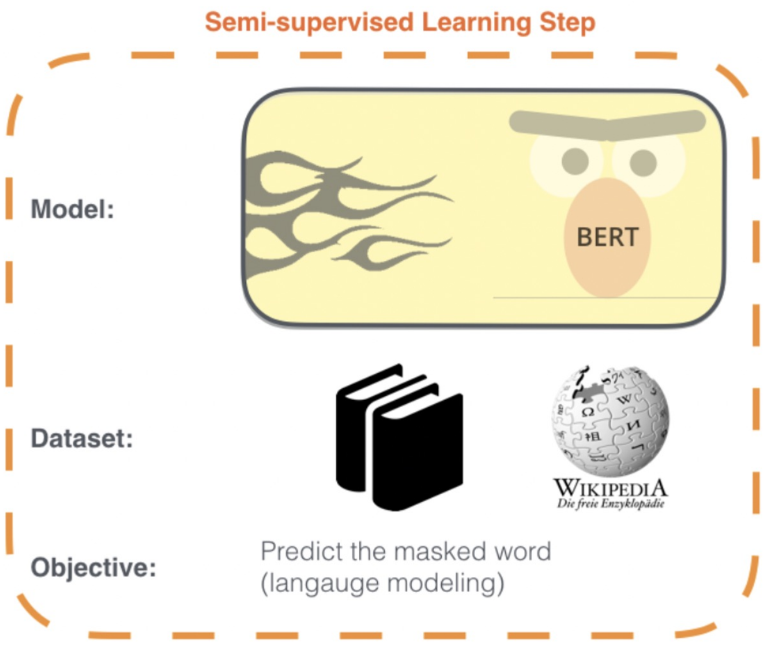


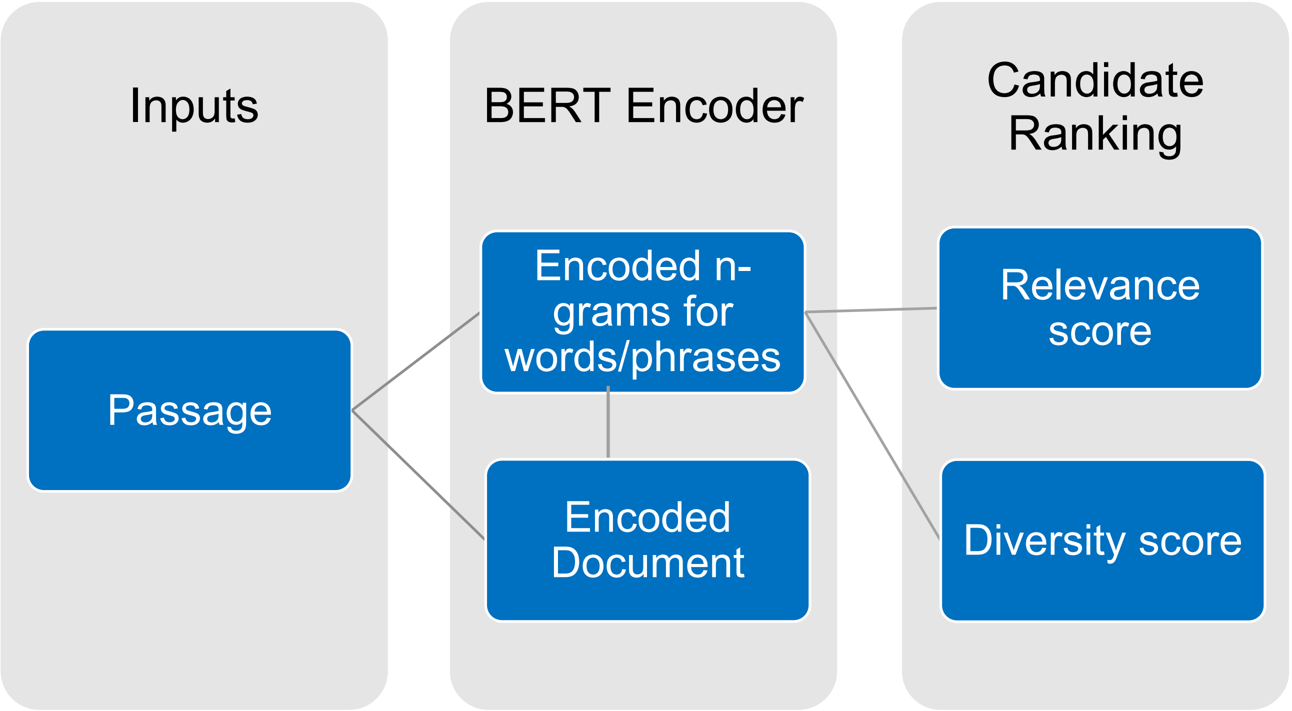
Agenda

- Keyword Extraction — 01
- Sentiment Analysis — 02
- Topic Modeling — 03
- K-Means Clustering — 04
- Conclusions and Recommendations — 05

High-level Explanation of Keyword Extraction

Keyword extraction is a technique used to automatically extract the most relevant words and phrases from text to help summarize the content.



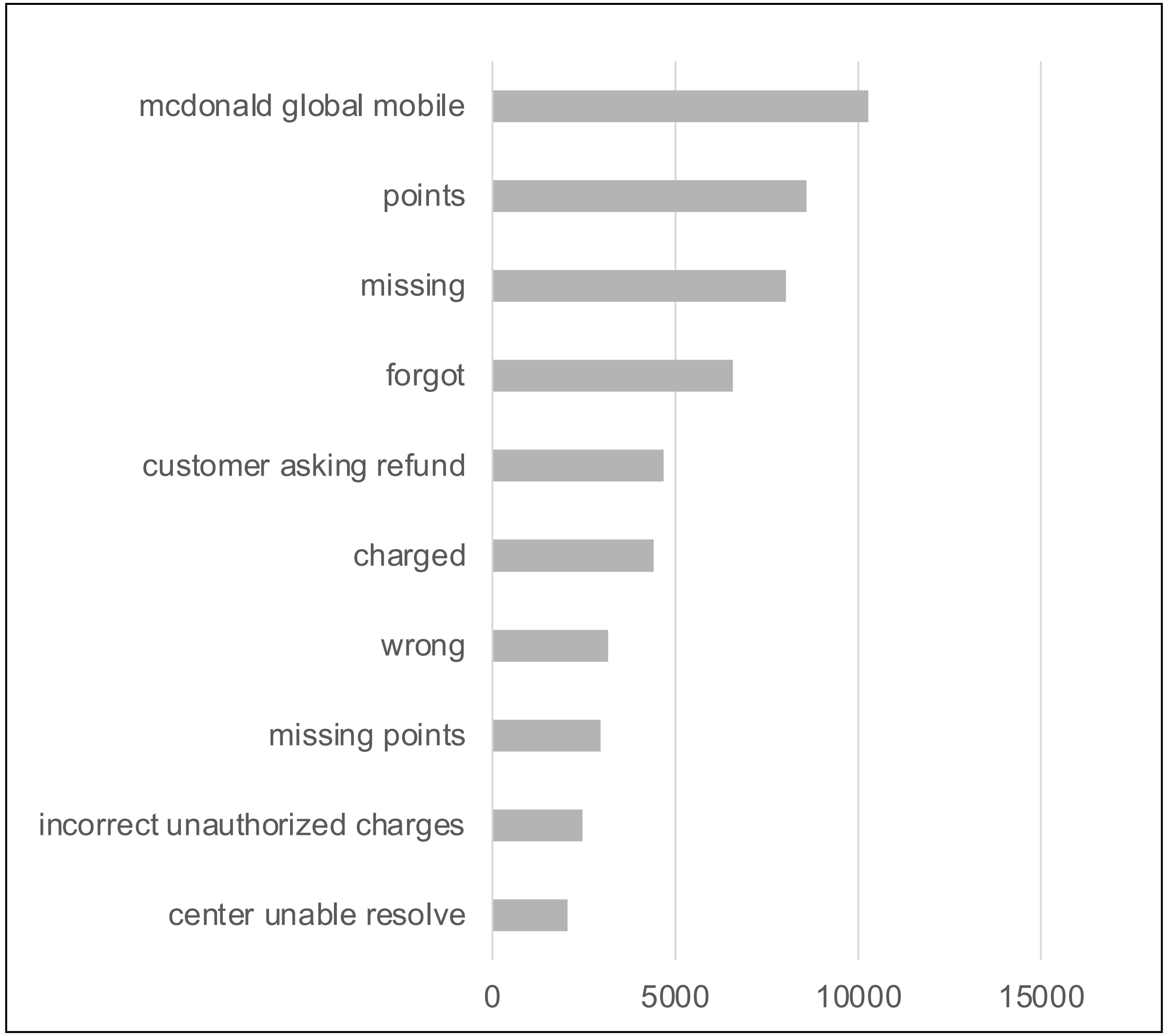


Explanation for the process

- **Inputs:** pass each customer comment into the keyword extraction pipeline
- **BERT Encoder**: BERT takes the comment and breaks it into the n-grams (phrases up to 3 words long) and words that could be keywords. Then BERT encodes both the entire comment and all candidates as vectors.
- **Candidate Ranking**: Candidates are selected based on how similar their vectors are to the comment vector. Candidate vectors are also compared to ensure that the selected keywords are different from each other.

Keyword Extraction result on the Entire Dataset

Top 10 Key Phrases across the Entire Dataset



Key Insights

- Customer complaints across the entire dataset primarily have to do with issues with the **mobile app, order completeness, and points/charges**, rather than food quality, service, or speed.
  - This doesn't imply that food quality is not a big issue; instead, it implies that customers are becoming unsatisfied **even before the ordering process is completed**.
  - We recommend McDonald's to look into its app design and mobile point issuing process to quickly improve its performance.
- "Missing/wrong" keywords relate to a few common gripes:
  - Missing items from order
  - Missing points from mobile orders
  - Wrong charges
  - Wrong order

Agenda

01	Keyword Extraction	
02	Sentiment Analysis	
03	Topic Modeling	
04	K-Means Clustering	
05	Conclusions and Recommendations	

Introduction to Sentiment Analysis

Motivation for Sentiment Analysis

- Sentiment Analysis is a Natural Language Processing technique used to determine whether the customer comment is positive, negative or neutral by assignment sentiment scores to the topics, categories or phrases.
- This can be used as the "first complaint metrics" to look at to determine the extent of negativity for each day, and can deep dive on specific dates to see what had happened

Methods for Sentiment Analysis

- VADER (Valence Aware Dictionary and Sentiment Reasoner)
  - Pros
    - > Contextual elements, like punctuation and degrees, are taken into account
    - > Able to customize the sentiment threshold
  - Cons
    - > Do not analyze the semantic structure, poor at complex sentence such as sarcasm
    - > Misspellings and grammatical mistakes may cause the analysis to overlook important words or usage

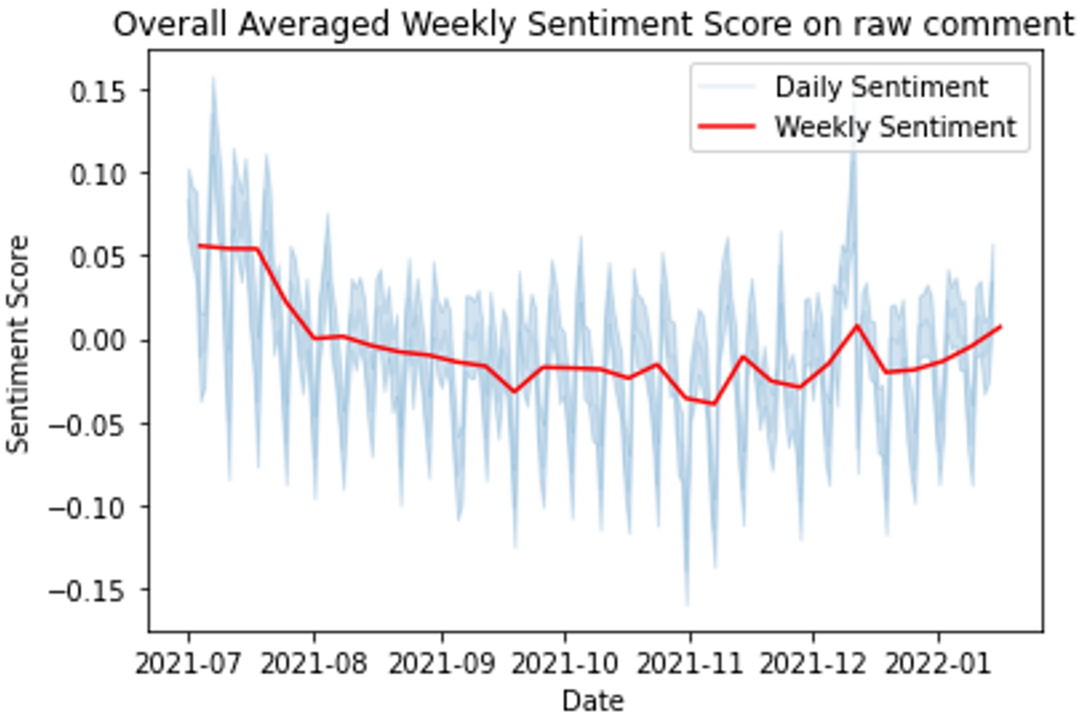
Text Data

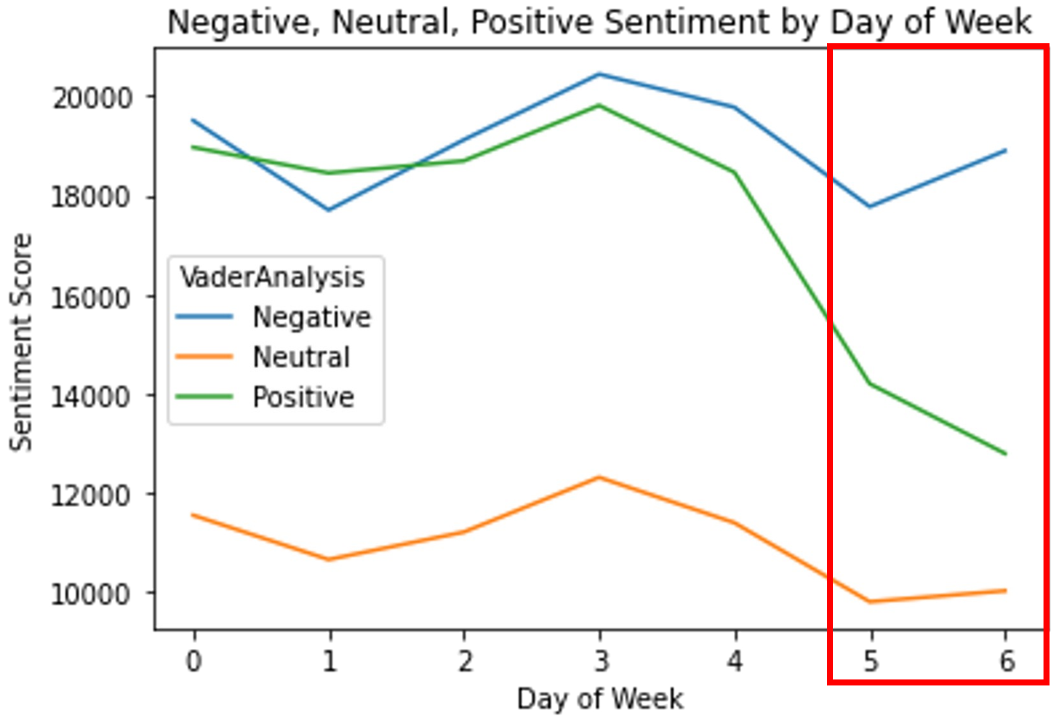
Lexicon-based Model

Based on lexicon dictionary

Categorical		Numerical	
word	sentiment	word	sentiment
nice	pos	nice	2
beautiful	pos	beautiful	3
amazing	pos	amazing	4
ugly	neg	ugly	-3
stupid	neg	stupid	-2

Sentiment Analysis on Day of Week





Key Insights

- At the onset of the program weekly aggregated sentiment score is the highest in July across all months. After then, aggregated sentiment fluctuates around 0 with the lowest level in November.
- Daily sentiment fluctuates strongly between weekdays and weekends; generally, there is a higher sentiment on Tuesday/Wednesday, and lower sentiment appears on weekends.
  - The plummet in weekend results in the significant decrease in positive reviews and increased negative reviews.

Agenda

	Keyword Extraction	01
	Sentiment Analysis	02
	Topic Modeling	03
	K-Means Clustering	04
	Conclusions and Recommendations	05

Topic Modeling result on the Entire Dataset

Motivation for Topic Modeling

- Topic Modeling is a technique that scans through documents, examines how words and phrases co-occur, and **learns clusters of words that best characterize these documents**

- **Latent Dirichlet Allocation (LDA):** assume probability distribution and using an iterative process to update best topic
  - Pros
    - > Simple to implement & fast
    - > Takes into account bigrams (words that appear close together)
  - Cons
    - > Requires lots of data due to high variance
    - > Lack interpretability
    - > More inaccurate to shorter texts

	% of All	Characteristics for the Topic	Top Words Associated with the Topic
Topic 1	22.7%	App related: points and rewards system	”points”, “app”, “time”, “code”, “rewards”, “asked”, “able”, ”phone”
Topic 2	5.1%	Extra coupons & deals for food	“deal”, “fries”, “free”, “large”, “coffee”, “one”, “two”, “coupon”
Topic 3	12.6%	Waiting time and pick-up at drive throughs	“food”, “minutes”, “order”, “drive”, “thru”, ”paid”, “went”, ”manager”
Topic 4	12.3%	Problems with app: logging in and installing	“app”, “email”, “mcdonalds”, “mobile”, “reinstall”, “uninstall”
Topic 5	27.5%	Refunds and mobile-processed orders	”order”, “mobile”, “app”, “would”, “placed”, “mcdonalds”, “refund”
Topic 6	15.3%	Finance-related: charges related complaints	“charged”, “contact”, “back”, “information”, “incorrect”, “bank”
Topic 7	4.5%	Issues with contacting representative.	“case”, ”issue”, “information”, “call”, “per”, “steps”, “center”, “guide”

Agenda

	Keyword Extraction	01
	Sentiment Analysis	02
	Topic Modeling	03
	K-Means Clustering	04
	Conclusions and Recommendations	05

Clustering on Complaint Dataset

Features that we used

We used a total of 109 features in the final aggregate clustering analysis, which used all NLP techniques mentioned earlier to classify users into 3 distinct groups.

Frequency: number of times the user has logged a complaint or follow-up.

Topic Scores (7): the average scores for each topic found for each user. This is from Topic Modeling.

Keywords (100): the top keyword for every complaint and its associated score, for all keywords in the top 100 most common keywords. This is from Keyword Extraction.

Recency: a variable named "days_from_start" that counts the number of days since the first recorded complaint that user has logged a complaint. Based on most recent complaint.

Description of each cluster

Cluster 0: Digital
This cluster is associated heavily with the keywords 'sign' and 'log', likely referencing when consumers had trouble signing in or logging in. This is further confirmed by the keywords 'Facebook', 'login', 'activate', 'reinstall', 'refresh', and etc. In essence, this is the cluster that deals most heavily with digital problems with consumers.

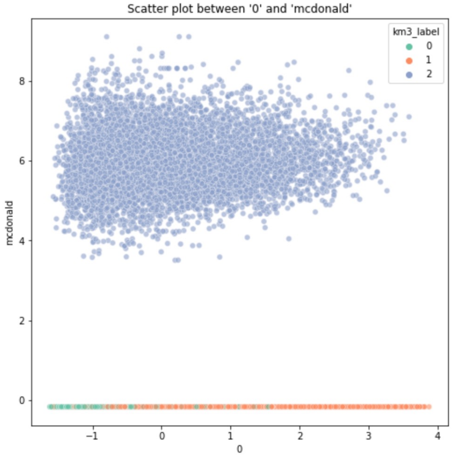
Cluster 1: Product Purchasing
This cluster is associated heavily with all food products (fry, pie, pies, egg, frappe, McMuffin, coffee, etc.), and financial issues (subtotal, reimbursement, charge, charges, refund). Hence, this is likely the product-focused category, where most complaints center around errors with charge.

Cluster 3: McDonalds
This cluster is thus associated solely with the keyword 'mcdonald' (100% of complaints with keyword 'mcdonald' are in this cluster), which likely references when consumers complain about the franchise name itself, rather than any quality of service.

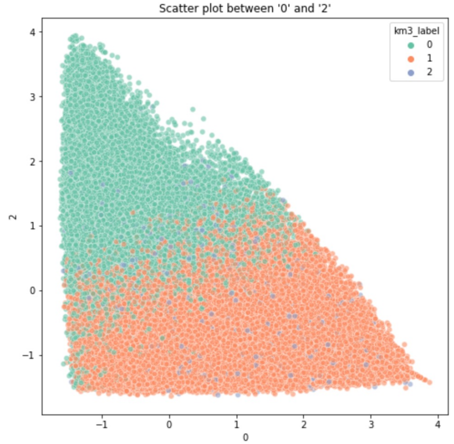
Descriptions of Clusters

Pairwise Scatterplots

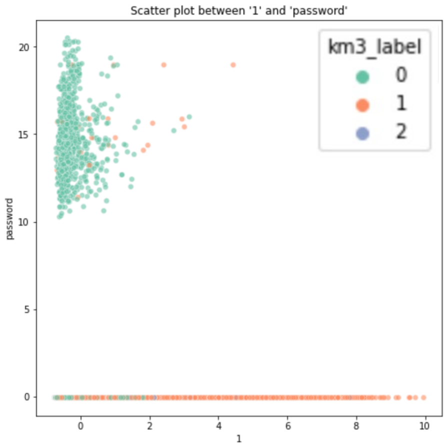
Topic Points Reward x Keyword ‘mcdonald’



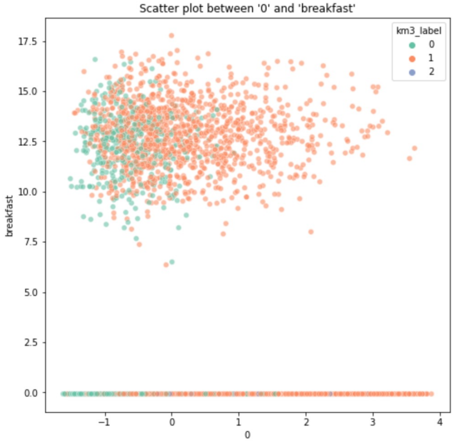
Topic Points Reward x Topic Drive Thru/Speed



Topic Coupon/Deal x Keyword ’password’



Topic Points Reward x Keyword ’breakfast’



Keywords of Each Cluster

Cluster 0: Digital	
sign	0.997354
log	0.996564
facebook	0.994061
login	0.987144
reinstalled	0.984733
activate	0.980806
refresh	0.979508
reinstall	0.977650
password	0.975845
email	0.973347
gmail	0.968804
account	0.961211
unable	0.950375

Cluster 2: McDonald	
mcdonald	1.0000

Cluster 1: Product Purchasing	
fry	0.967033
subtotal	0.965932
scan	0.965357
charged	0.961585
reimbursement	0.957045
pie	0.954972
unauthorized	0.953353
pies	0.952055
charge	0.949932
egg	0.938338

Agenda

- Keyword Extraction — 01
- Sentiment Analysis — 02
- Topic Modeling — 03
- K-Means Clustering — 04
- **Conclusions and Recommendations — 05**

Conclusion and Recommendtation

Conclusion and Recommendation

- With **keyword extraction**, we can **extract key phrases** that are the most similar to the entire document, which serves as a simple summarization cross comparison method across different complaints.
  - From the analysis we realize that customer talks more about issues with the **mobile app, order completeness, and points/charges**, rather than food quality, service, or speed.
  - For keywords related to “missing/wrong”, there are some common gripes such as **missing items, missing reward points, wrong charges, wrong orders**, etc.
  - We recommend McDonald’s start tacking the most frequent complaints to mitigate probability of customer churn
- With **sentiment analysis**, we can **quantify customer complaints** and monitor on a daily basis thanks to the instant availability (no need to re-train the model on the entire text corpus)
  - We realize that transcription from agents might have neutralized the sentiment from calls, leading to the seemingly higher sentiment score compared to emails.
  - We recommend
    - > using sentiment as the starting point for daily monitor of customer opinion
    - > (if needed) alter the transcribing process to mitigate the bias introduced through agent’s transcription
- With **topic modeling**, we can **conduct opinion mining** to understand what are the latent topics customer cares about.
  - Compared to keyword extraction, topic modeling puts more emphasis on the context of the corpus, which makes this modeling technique data-hungry
  - App-related issues and refund issues are the top 2 latent topics customer cares about -> should be the McDonald’s priority to mitigate customer complaints
- With **clustering analysis**, we **group customer through text behaviors** into three clusters
  - 1) Digital Non-Payment Issues, 2) Customer Interactions, 3) General Purchasing Experience
  - We realize that that regular buyers are the most common and are stable across all three clusters, which corroborates the fact that no specific complaint type affect churning behavior more
  - We recommend McDonald’s start with either the most frequent complaint (most frequent keywords) to mitigate customer dissatisfaction

Questions

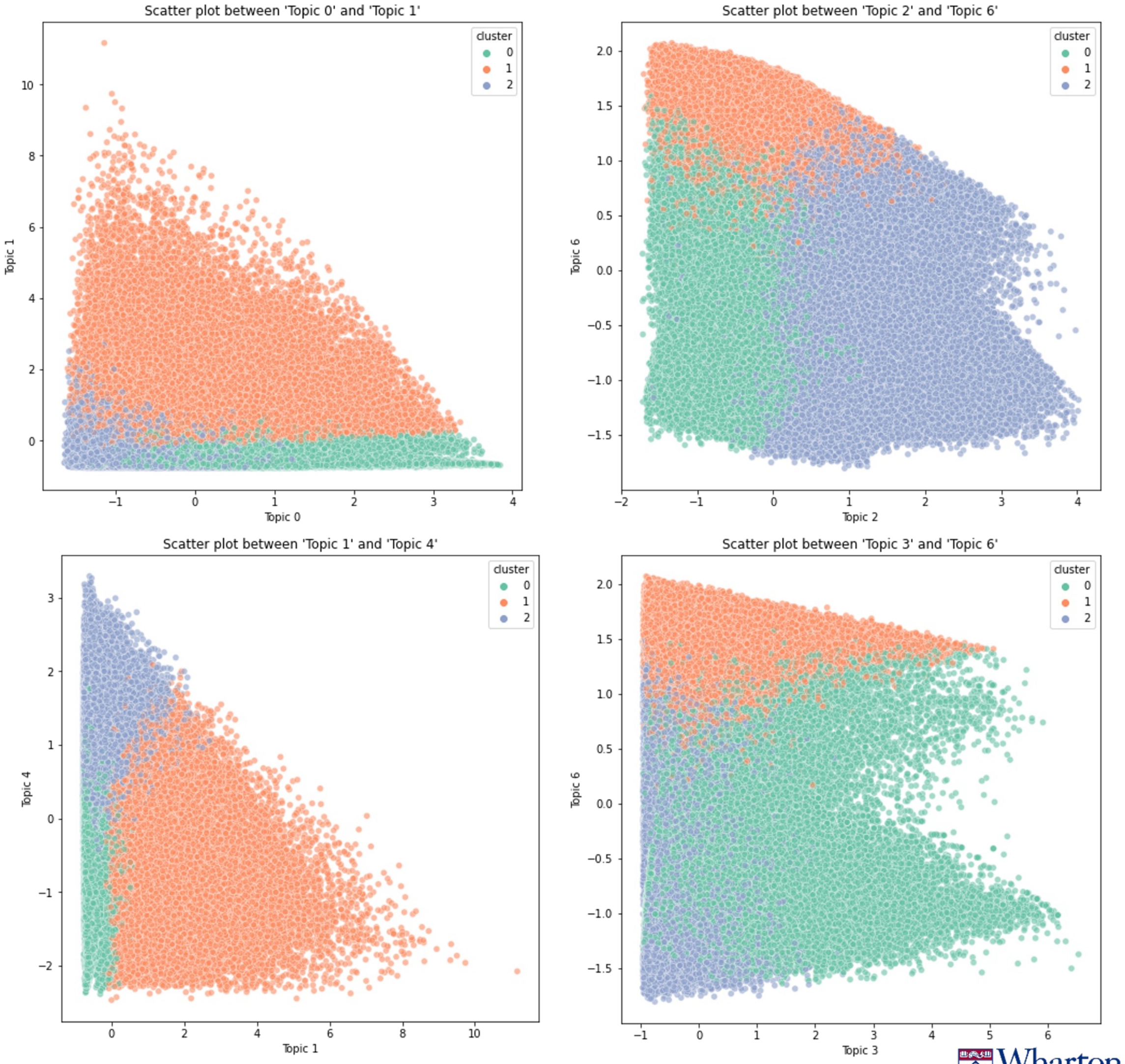
?

Clustering on Complaint Dataset

Features that we used

- **Frequency:** number of times the user has logged a complaint or follow-up.
- **Topic Composition** (7): percentage of each topic extracted from the result of topic modeling.
  - **Topic 1**: App related complaints such as the points and rewards system
  - **Topic 2**: Extra coupons & deals for food
  - **Topic 3**: Waiting time and pick-up at drive throughs
  - **Topic 4**: Problems with the account or app: logging in, installing, etc.
  - **Topic 5**: Refunds or problems with mobile-processed orders
  - **Topic 6**: Finance-related: charges related complaints
  - **Topic 7**: Issues with calling or contacting representative.
- **Sentiment Score**: sentiment score associated with each comment from the VADER algorithm

Pairwise Scatterplots



Descriptions of Clusters

	Topic 1	Topic 2	Topic 3	Topic 4	Topic 5	Topic 6	Topic 7	compound	count
cluster									
0	0.603727	-0.505178	-0.656397	0.456447	-0.529650	-0.006611	-0.430672	-0.005415	0.032432
1	-0.042217	1.332605	-0.326324	-0.168311	-0.491027	0.037917	1.398317	0.160628	-0.036734
2	-0.683209	-0.354809	1.008649	-0.419130	0.976514	-0.019268	-0.489775	-0.108319	-0.012086

Why Clustering on Text Data?

- **Cluster 0**: **Digital Non-Payment Issues**: incorporates points, logins, etc.
  - Topic 1: App related complaints such as the points and rewards system
  - Topic 4: Problems with the account or app: logging in, installing, etc.
- **Cluster 1**: **Customer Interactions**: deals, customer services, etc.
  - Topic 2: Extra coupons & deals for food
  - Topic 7: Issues with calling or contacting representative.
- **Cluster 2**: **General Purchasing Experience:** friction associated with day to day purchasing
  - Topic 3: Waiting time and pick-up at drive throughs
  - Topic 5: Refunds or problems with mobile-processed orders