GTA Restaurant Recommender

A subset of the Yelp Kaggle dataset

Capstone Project by Debra Goei

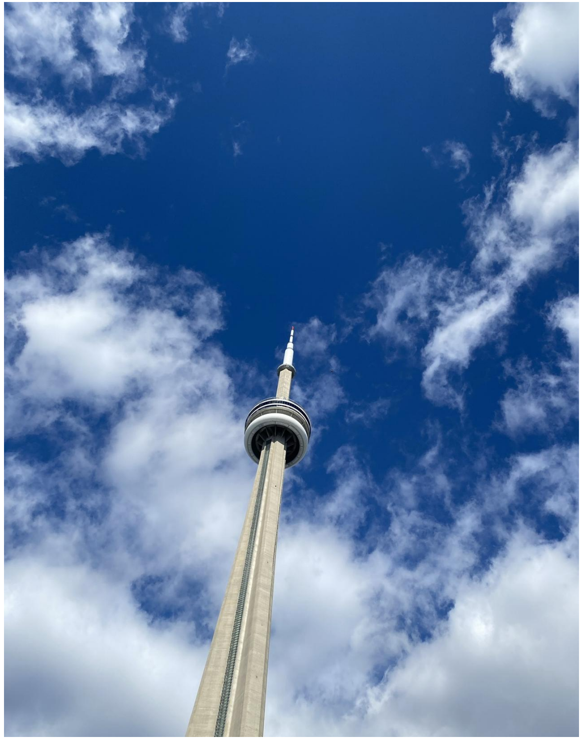
Problem Statement

Yelp is a business directory service:

- Users can leave ratings and create reviews
- Has a reservation system

Natural Language Processing on Yelp data:

- Different insight into user-driven reviews
- Provide feedback to businesses for improvement and upkeep
- Introduces small local businesses



Background

Recommender Systems

- Relevant and accurate information
- Learning user patterns & produce reliable outcomes

Yelp Dataset

- Originally over 6 million rows
- Cannot be scraped using an API
- Original features mostly kept

YELP Connects People with Great Businesses!.

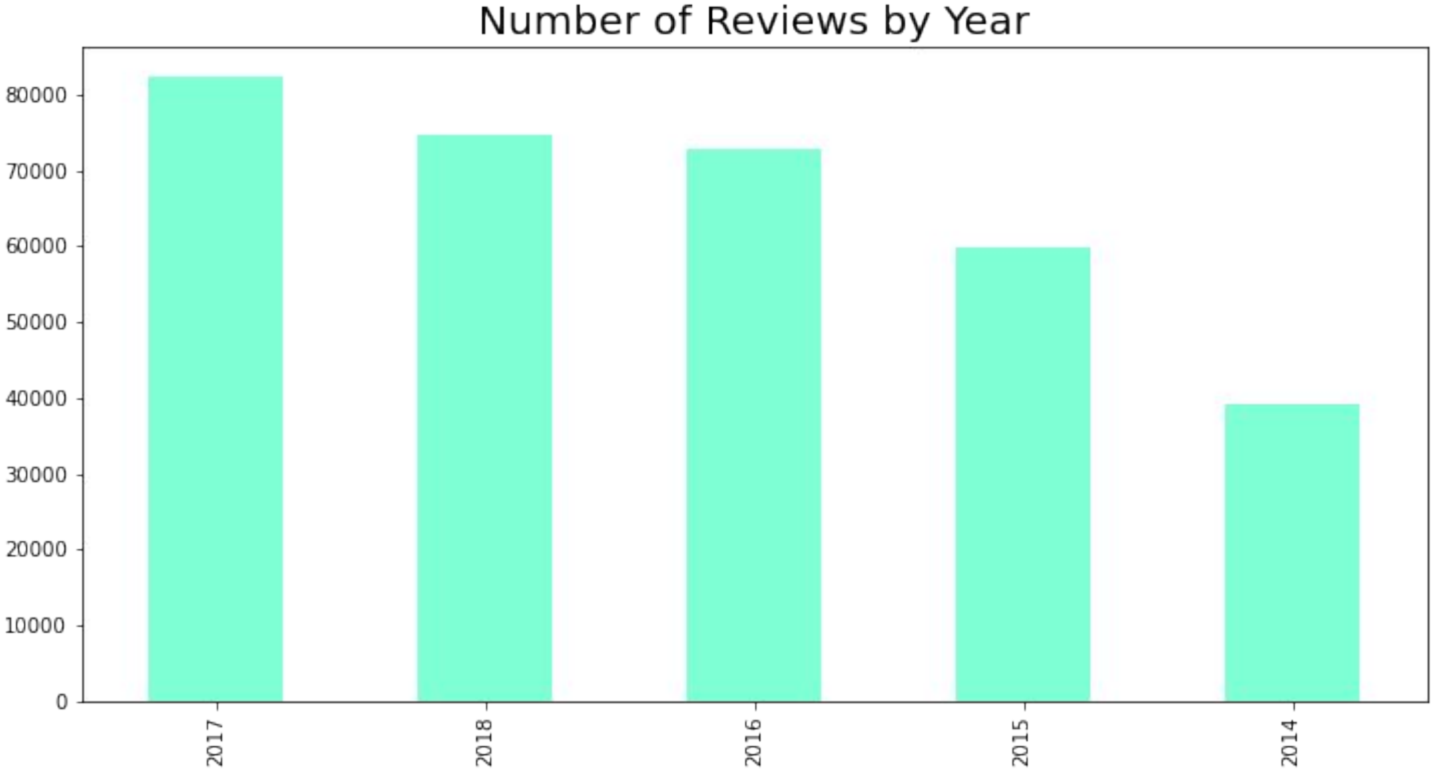
1. Outline

➔ **Data Cleaning**
Yelp JSON, transforming

➔ **EDA & Sentiment Analysis**
VADER

➔ **Recommender System**
Content-Based
Location-Based

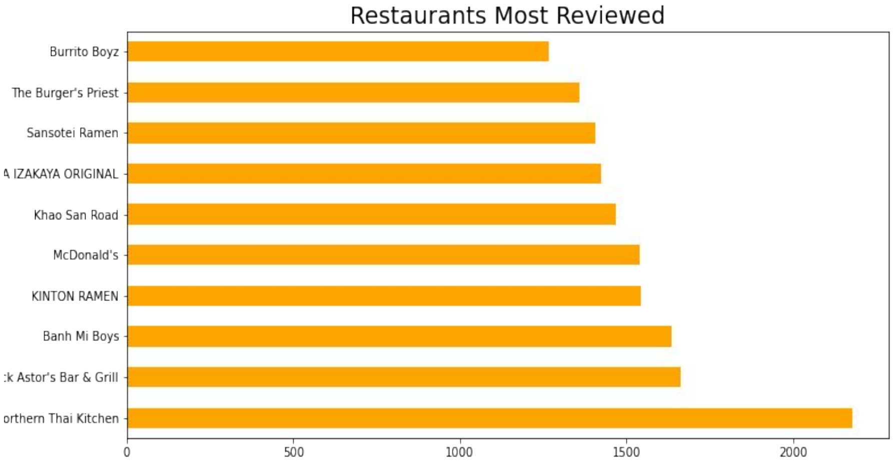
➔ **Conclusion & Takeaways**

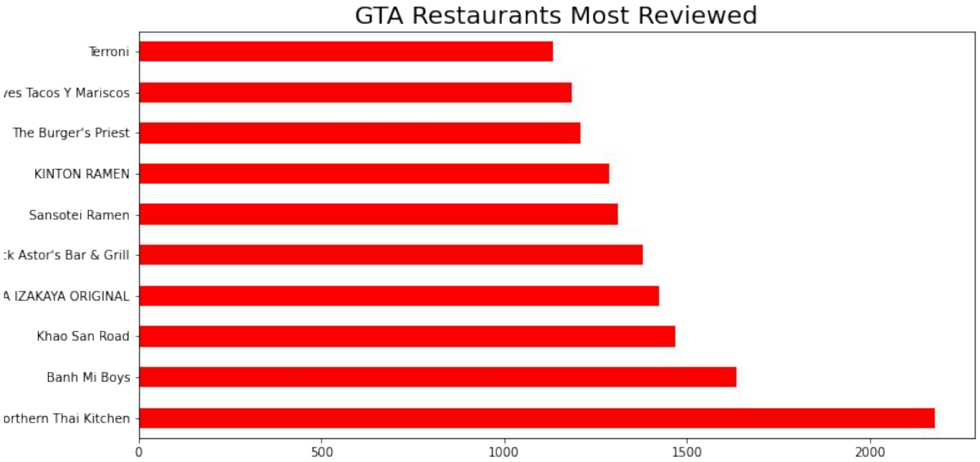

Number of Reviews by Year
80000
70000
60000
50000
40000
30000
20000
10000
0
2017
2018
2016
2015
2014

Ontario is not just where **JUSTIN BIEBER, THE WEEKND & DRAKE** call home…

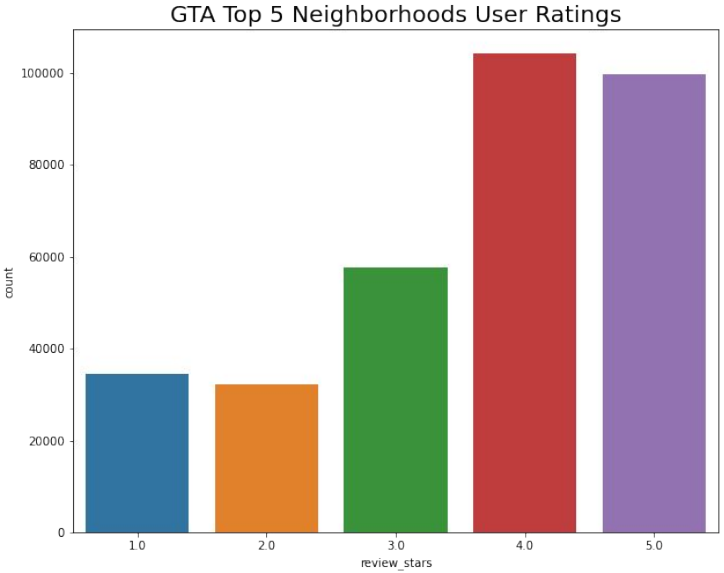
Image Source: Forbes

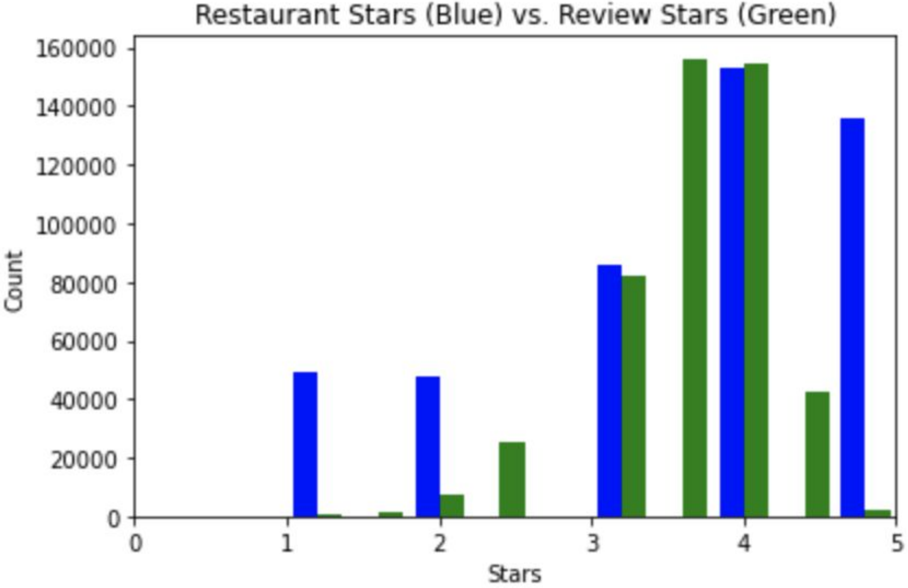
Home to Many Thai Restaurants
& Fast Food Chains Too!

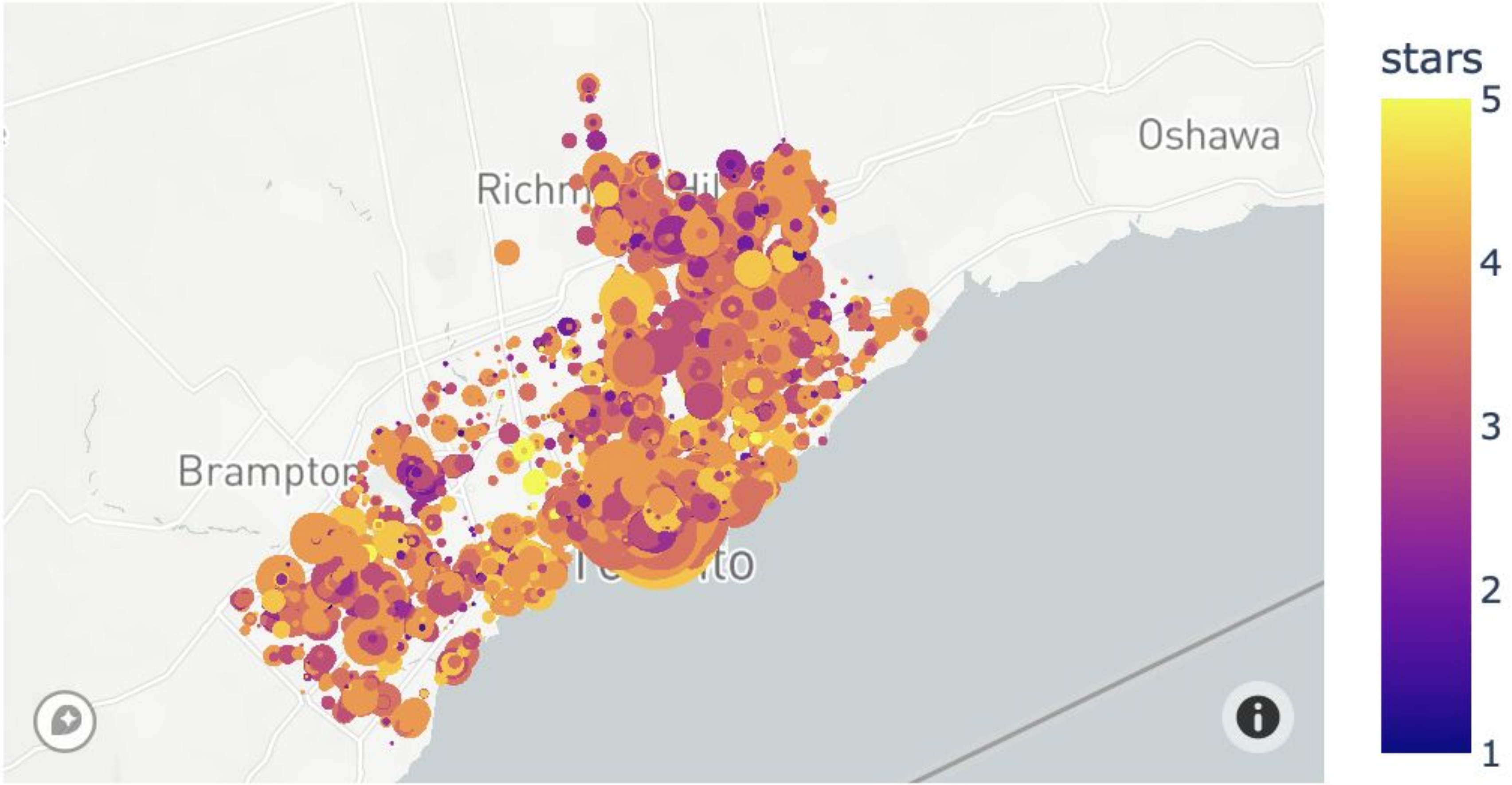


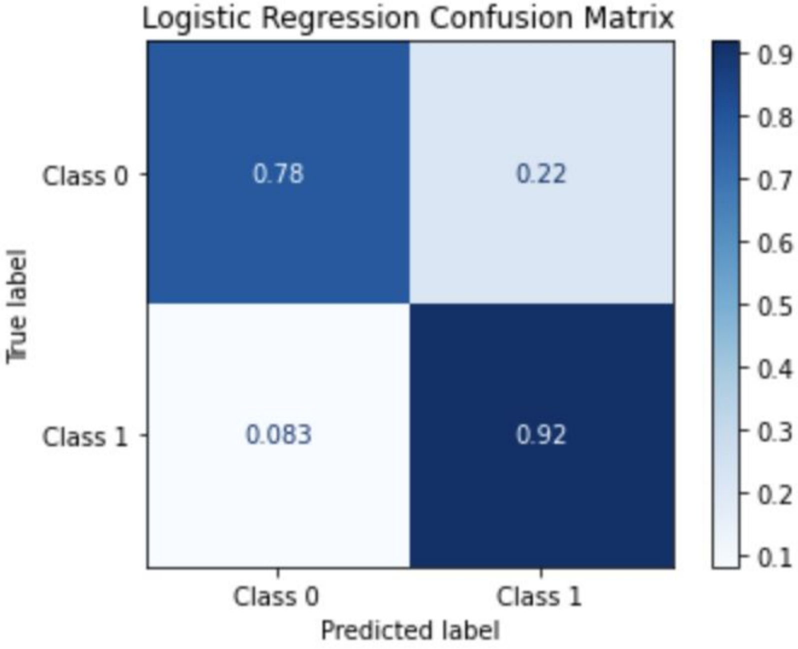


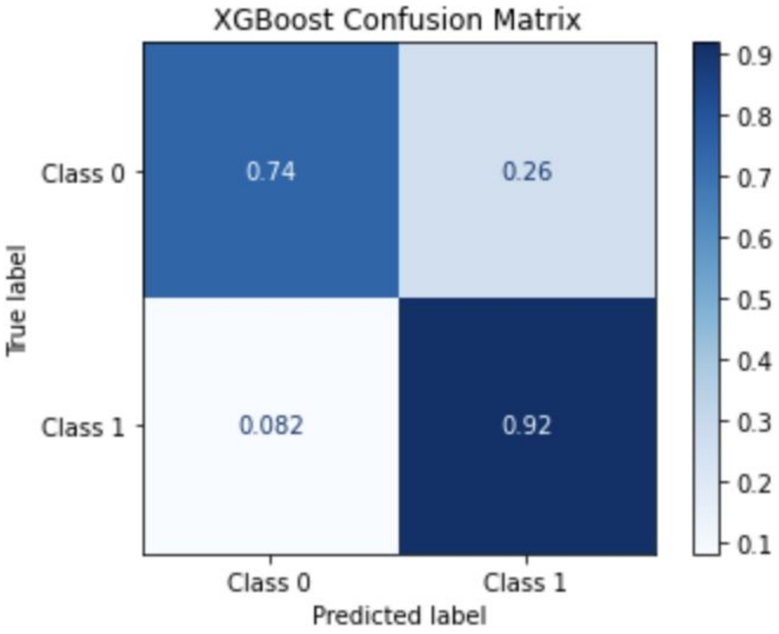
What Can a Numeric Rating Really Tell Us?






Richmond Hill
Oshawa
Brampton
Toronto
stars
5
4
3
2
1

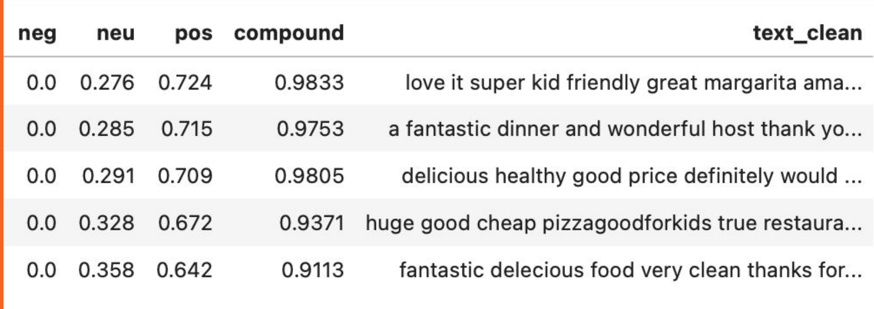




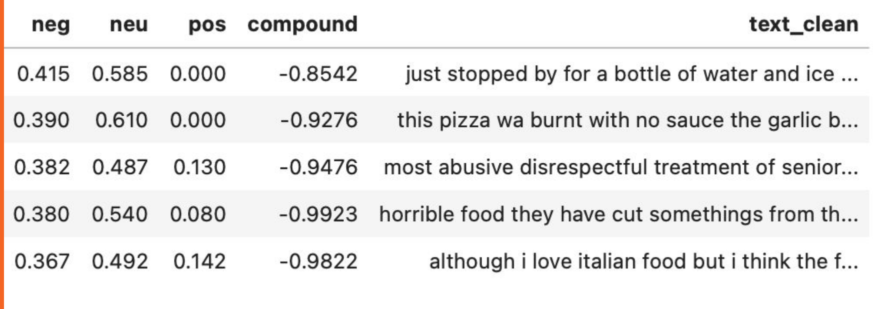
	Train	Test
LogReg	0.8810	0.8666
XGBoost	0.8667	0.8153

Sentiment Analysis

VADER works very well on social media type text and generalizes to multiple domains without really suffering from a speed-performance tradeoff



neg	neu	pos	compound	text_clean
0.0	0.276	0.724	0.9833	love it super kid friendly great margarita ama...
0.0	0.285	0.715	0.9753	a fantastic dinner and wonderful host thank yo...
0.0	0.291	0.709	0.9805	delicious healthy good price definitely would ...
0.0	0.328	0.672	0.9371	huge good cheap pizzagoodforkids true restaura...
0.0	0.358	0.642	0.9113	fantastic delecious food very clean thanks for...



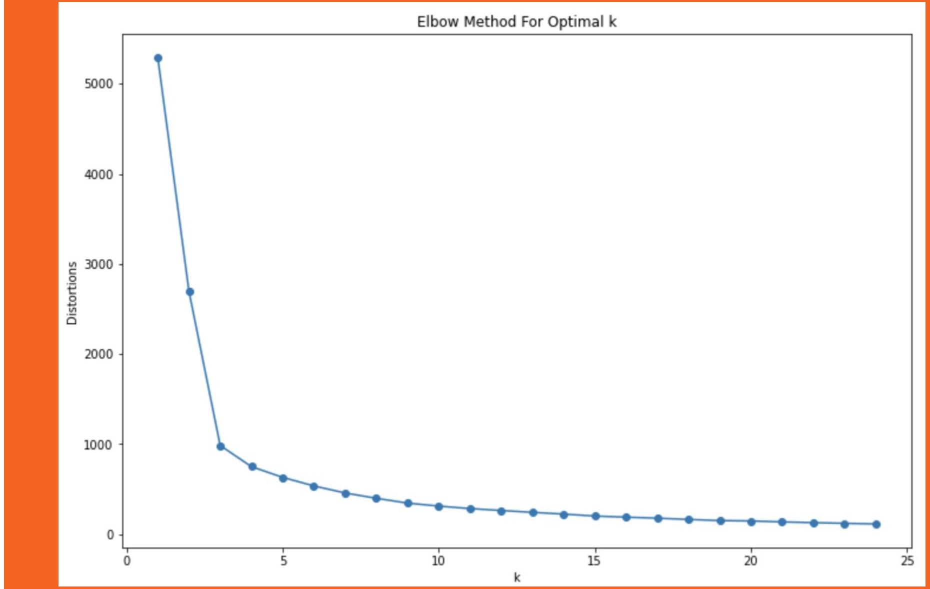
neg	neu	pos	compound	text_clean
0.415	0.585	0.000	-0.8542	just stopped by for a bottle of water and ice ...
0.390	0.610	0.000	-0.9276	this pizza wa burnt with no sauce the garlic b...
0.382	0.487	0.130	-0.9476	most abusive disrespectful treatment of senior...
0.380	0.540	0.080	-0.9923	horrible food they have cut somethings from th...
0.367	0.492	0.142	-0.9822	although i love italian food but i think the f...

Content-Based Recommender

- ➔ CountVectorizer
- ➔ Cosine Similarity
- ➔ Count Matrix

Location-Based Recommender

- KMeans Clustering
- Elbow Method



Source: travel.trade.gov

Conclusion & Takeaways

Incorporate Neural Network and Deep Learning Concepts

More data will improve on accuracy scores

Stakeholders can identify negative words to assist businesses with improvement

Thank You!

Time to test out the

Streamlit App (still being developed)!

Sources have been included in the main project repo: github.com/debragoei