



# Food Recommender

Zavier Soon  
DSIF 3

# Problem Statement

## What shall we eat? Anything.

- ❖ It is very common that we hang out with families, friends, and colleagues when comes to lunch or dinner time. There are an overwhelming number of restaurant choices in Singapore Central region.
- ❖ Dr. Brian Wansink from Cornell University claims that we make about 200 decisions about food each day
- ❖ We will develop webapps to recommend restaurants based on **similar restaurants** and **location-based**.

Project Focus : Singapore Central Region

Today 12:54

What do you want to have for dinner tonight?

I don't know, any cravings?

Ummm.... How about hotpot?

Any new place to recommend other than HDL?



Delivered



# Methodology

# Methodology





# Data Acquisition

Photo by [Carissa Gan](#) on [Unsplash](#)



# Data Acquisition



1

## Source

<https://www.yelp.com.sg/>

2

## Size

Restaurants Data : 1,230 x 14

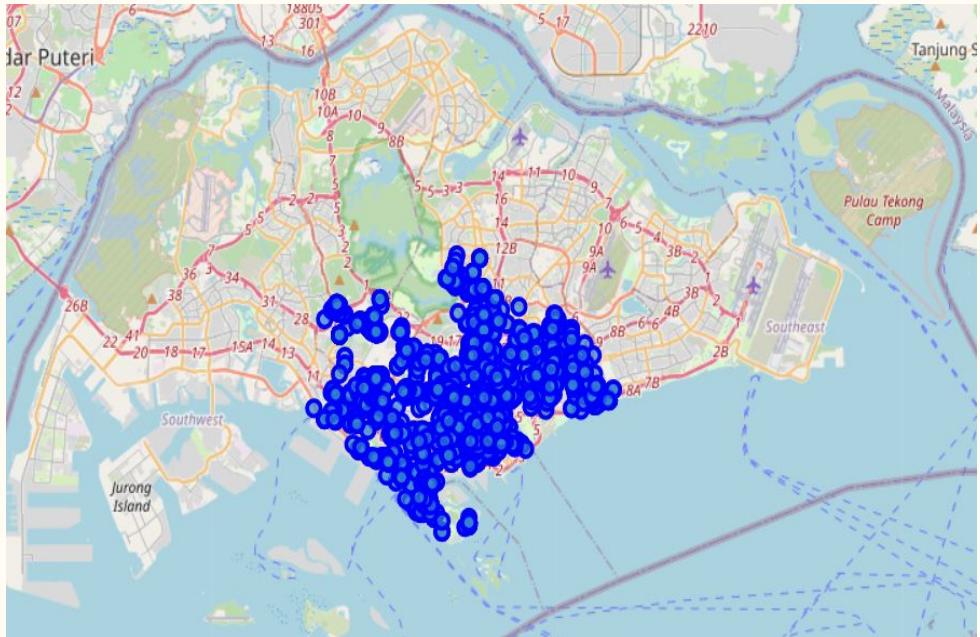
Restaurants Reviews: 18,710 x 7

3

## Features

Restaurants Details : id, name, categories, restaurant rating, price, address etc.

Restaurants Reviews: username, userid, businessid, comment\_text, rating





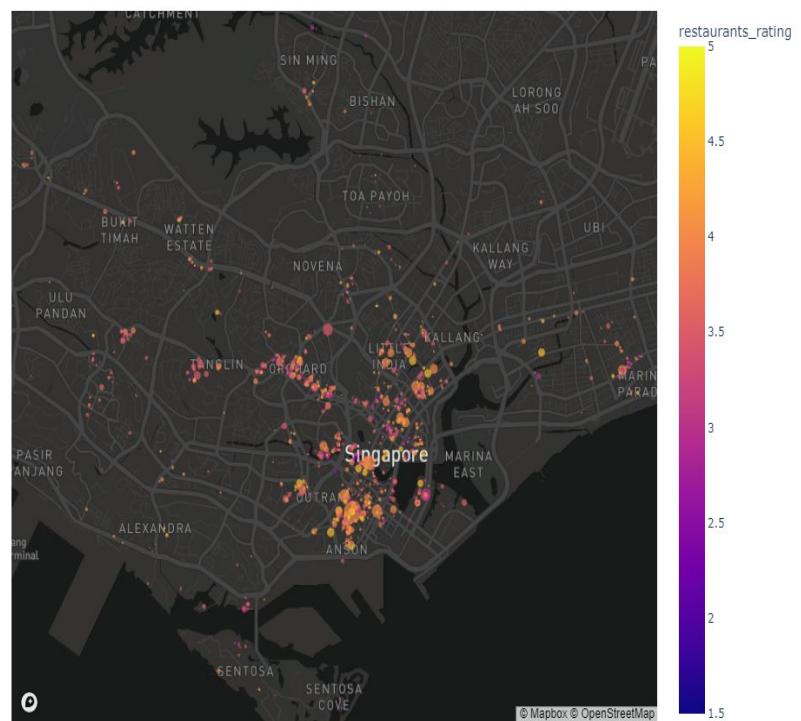
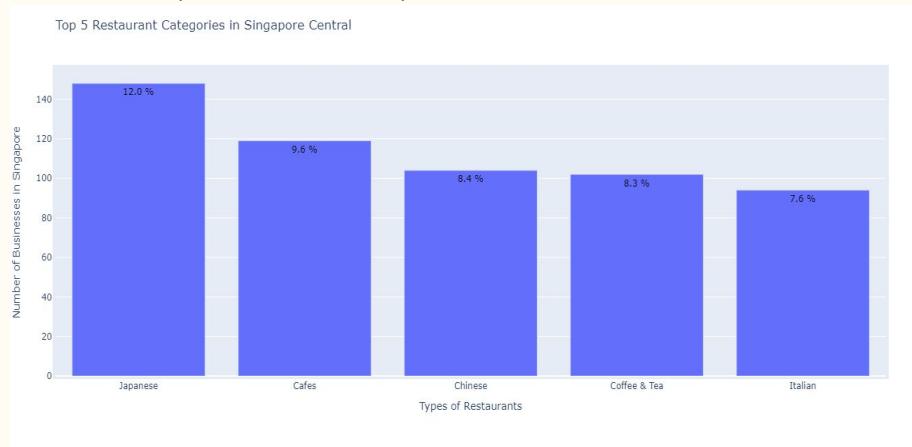
# Exploratory Data Analysis (EDA)



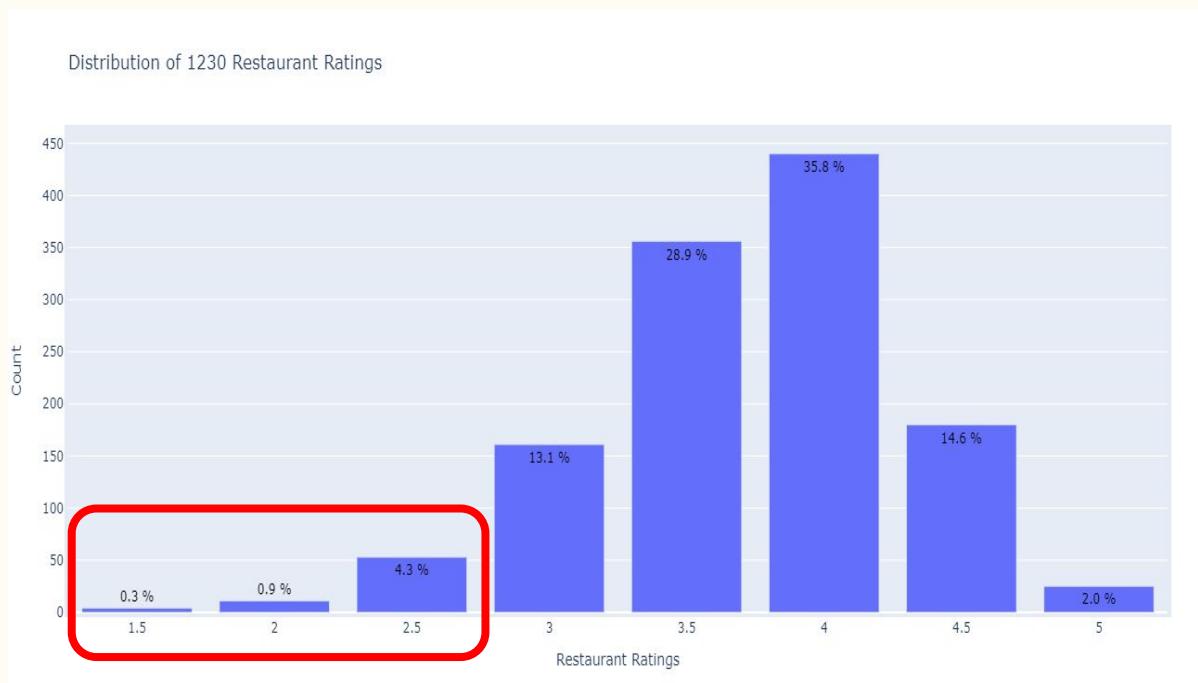
# EDA on Restaurant Data (1 / 2)

## Top 5 Most Popular Restaurant Categories:

- Japanese (148 Restaurants)
- Cafes (119 Restaurants)
- Chinese (104 Restaurants)
- Coffee & Tea (102 Restaurants)
- Italian (94 Restaurants)



# EDA on Restaurant Data (2 / 2)

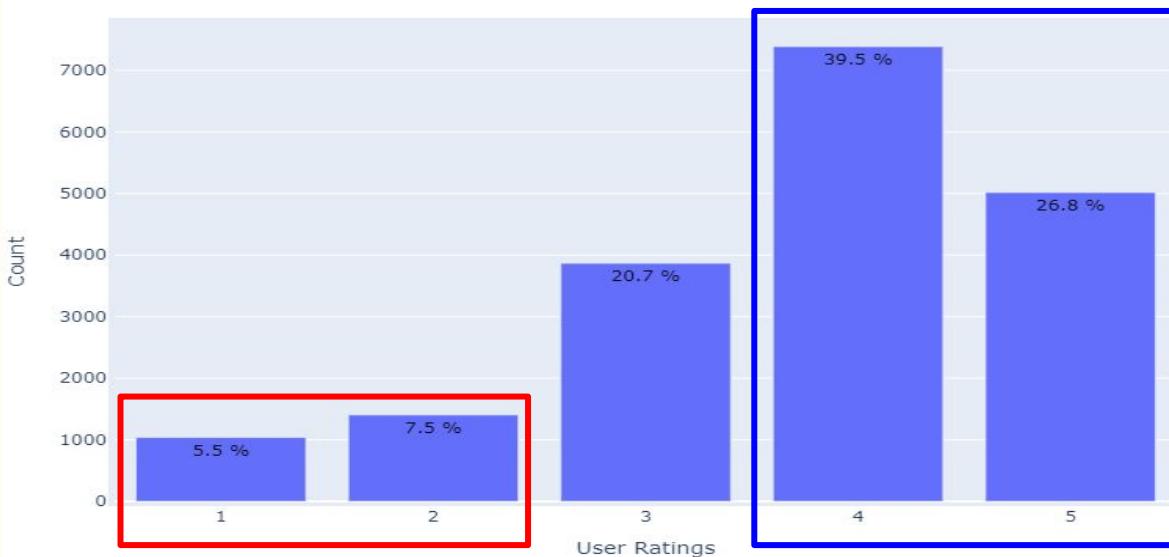


## Observations

- ❖ Distribution are slightly skewed to the left as most of the ratings are 3.5 & 4.
- ❖ User likely will not choose a restaurants with a rating below 3.

# EDA on Restaurant Reviews Data (1 / 2)

Distribution of 18710 Yelp User Ratings

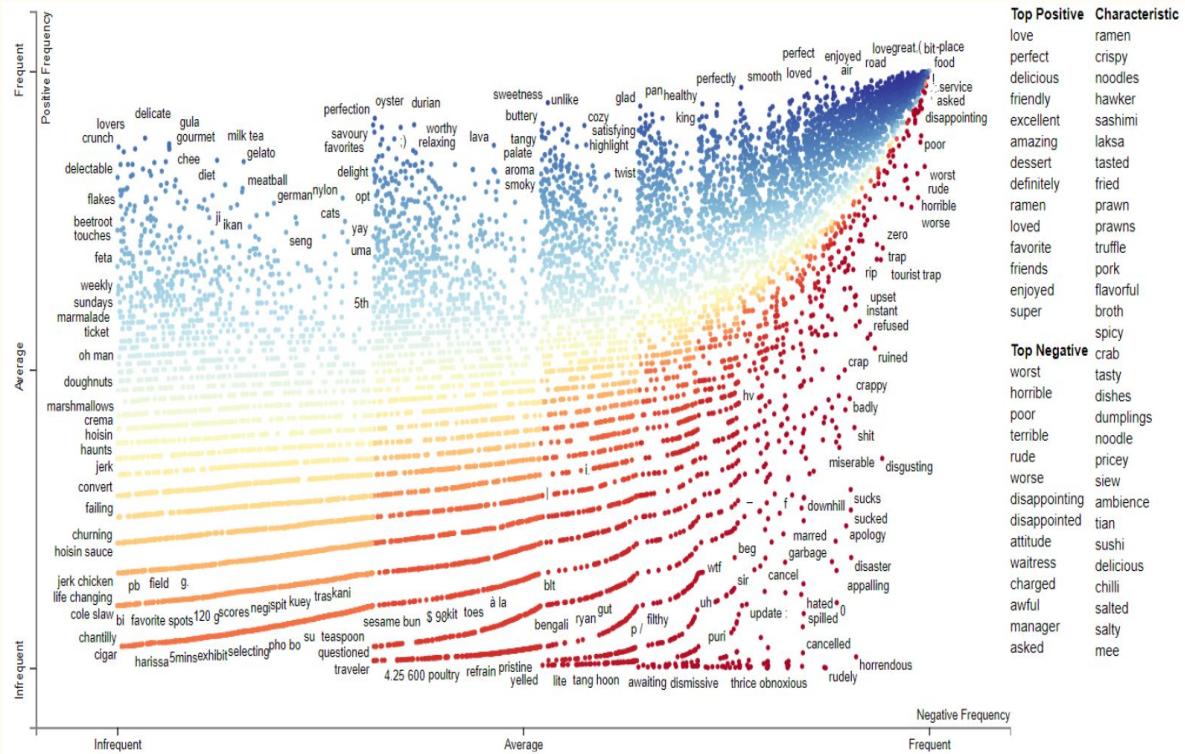


12,402 Positive Reviews  
2,441 Negative Reviews

## Observations

- The user rating distribution are skewed to the left as most of the ratings given by users are 4 & 5.

# EDA on Restaurant Reviews Data (2 / 2)



## Observations

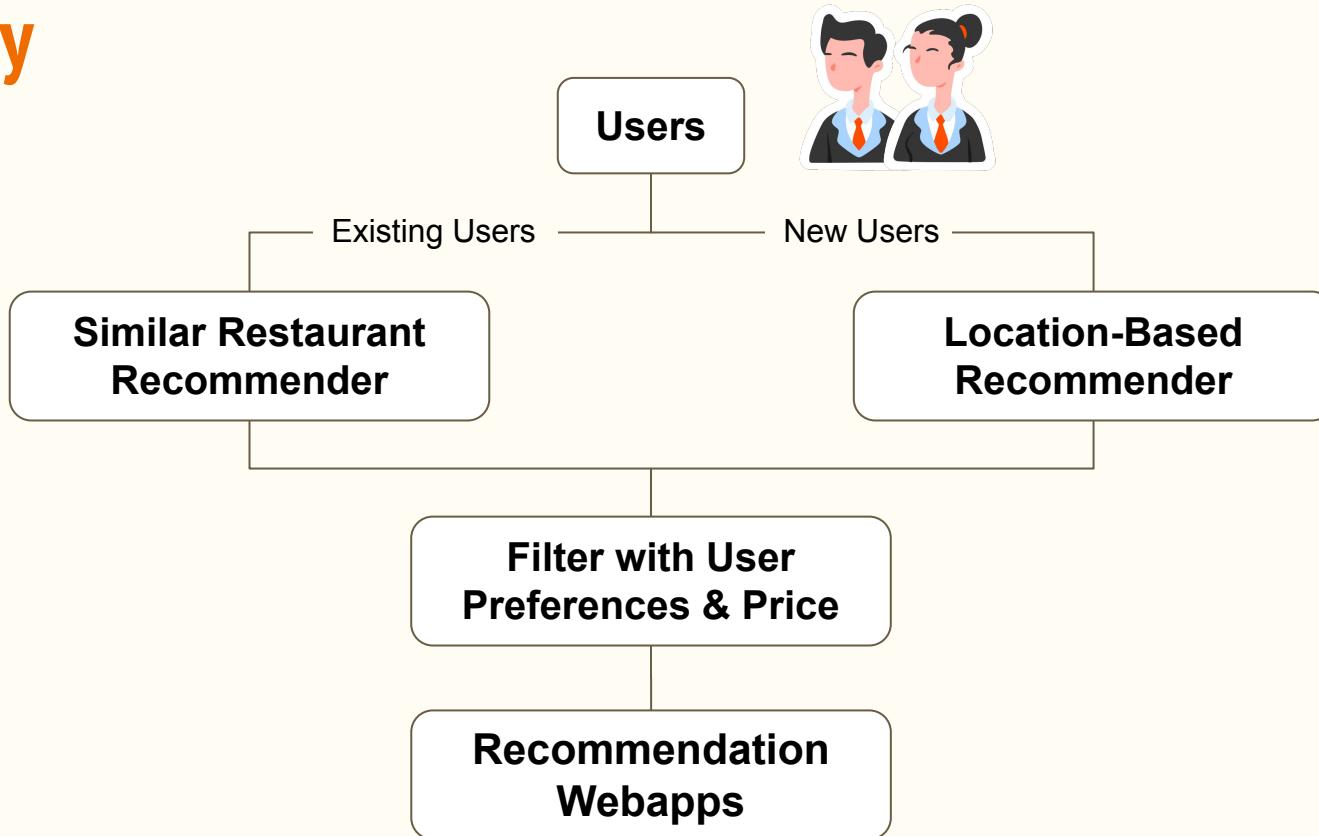
- ❖ Words most often associated with bad reviews are on the bottom right. They include rude, bad, crap, tourist trap & etc.
- ❖ Words most frequently associated with good reviews are on the top left. Among them are delicate, perfection, buttery & etc.



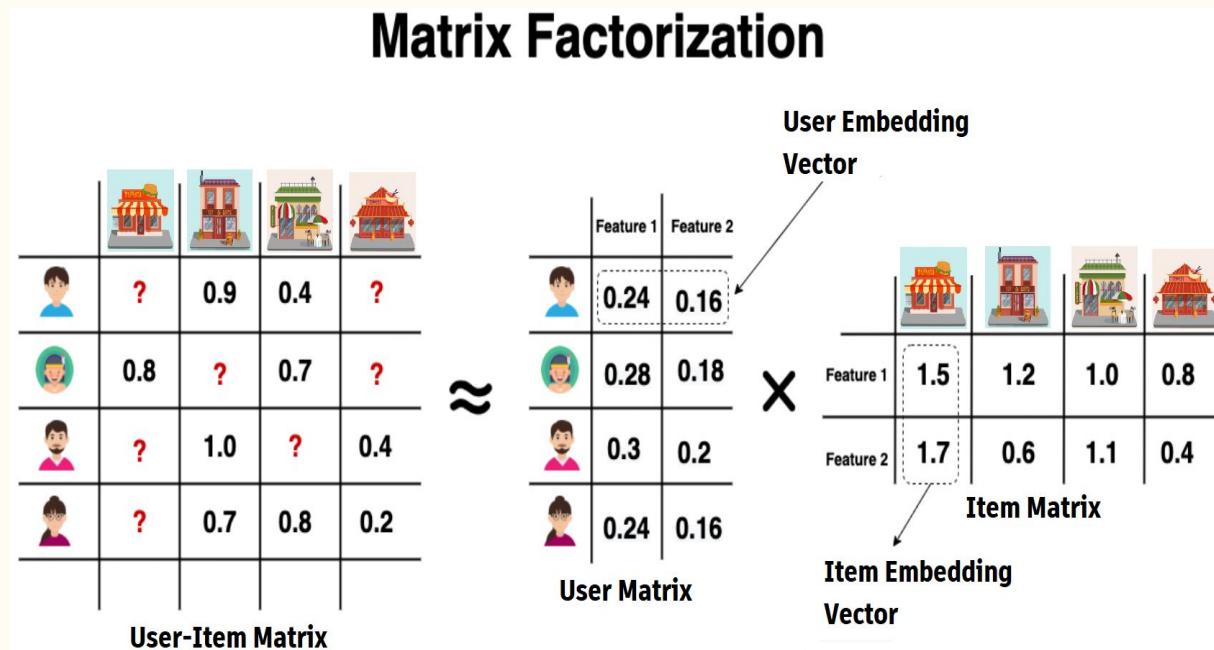
# Recommender System



# Strategy

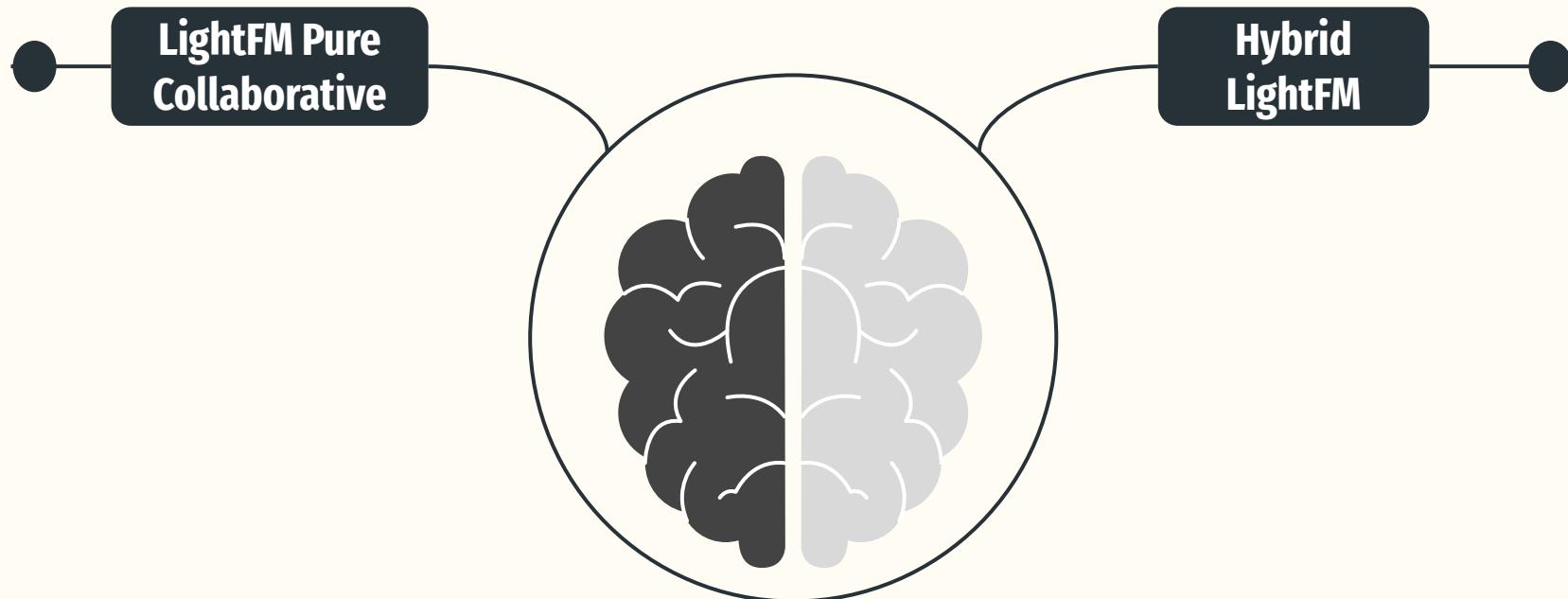


# Similar Restaurants Recommender - LightFM



# Similar Restaurants Recommender - LightFM

## Model



# Similar Restaurants Recommender - LightFM

## Model Evaluation

Model	Train AUC Score	Test AUC Score	Precision (k = 10)	Recall (k = 10)
LightFM (Pure Collaborative Filtering)				
Hybrid LightFM				

### Remarks:

- ❖ AUC Score is the probability that a randomly chosen positive example has a higher score than a randomly chosen negative example
- ❖ Precision at k is the proportion of recommended items in the top-k set that are relevant
- ❖ Recall at k is the proportion of relevant items found in the top-k recommendations

# Similar Restaurants Recommender - LightFM

## Model Evaluation

Model	Train AUC Score	Test AUC Score	Precision (k = 10)	Recall (k = 10)
LightFM (Pure Collaborative Filtering)	0.9554	0.6004	0.0075	0.0579
Hybrid LightFM	0.9887	0.6352	0.0093	0.0595

### Remarks:

- ❖ AUC Score is the probability that a randomly chosen positive example has a higher score than a randomly chosen negative example
- ❖ Precision at k is the proportion of recommended items in the top-k set that are relevant
- ❖ Recall at k is the proportion of relevant items found in the top-k recommendations

# Similar Restaurants Recommender - LightFM

## Model Hyperparameter Tuning

Model	Train AUC Score	Test AUC Score	Precision (k = 10)	Recall (k = 10)
Hybrid LightFM	0.9887	0.6352	0.0093	0.0595
Hybrid LightFM (Hyperparameter Tuning)	0.9590	0.6727	0.0165	0.1042

### Remarks:

- ❖ AUC Score is the probability that a randomly chosen positive example has a higher score than a randomly chosen negative example
- ❖ Precision at k is the proportion of recommended items in the top-k set that are relevant
- ❖ Recall at k is the proportion of relevant items found in the top-k recommendations

# Similar Restaurants Recommender Case Study

Case Study : Tomi Sushi



Ki-sho 葵匠

Correlation: 0.9998



Aoki

Correlation: 0.9993

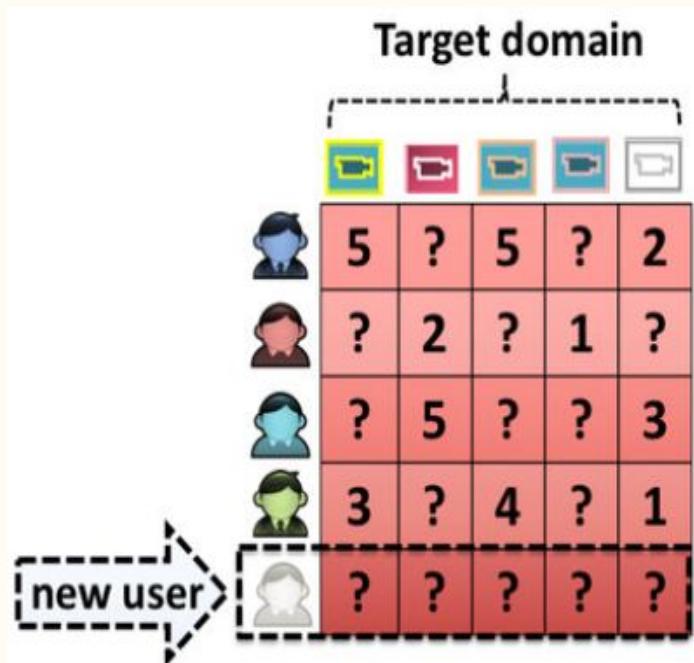


Kyoaji Dining

Correlation: 0.9991



# Location-Based Recommender



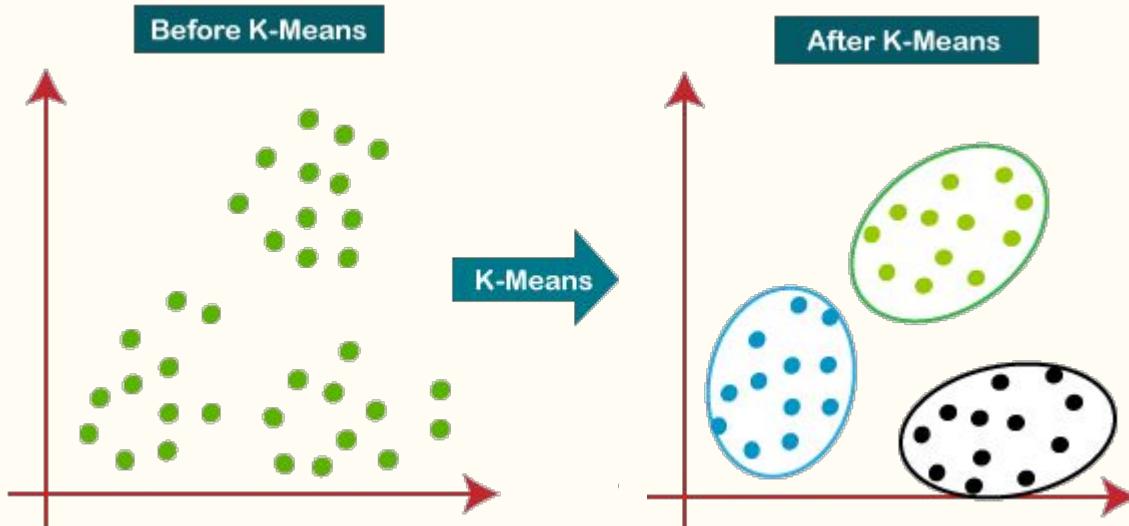
## New Users Cold Start

- ❖ Insufficient information due to unable to draw any inference for new users.
- ❖ It is difficult to provide personalized recommendations.

## Solution

- ❖ Location-based recommendation system using K-Means Clustering.

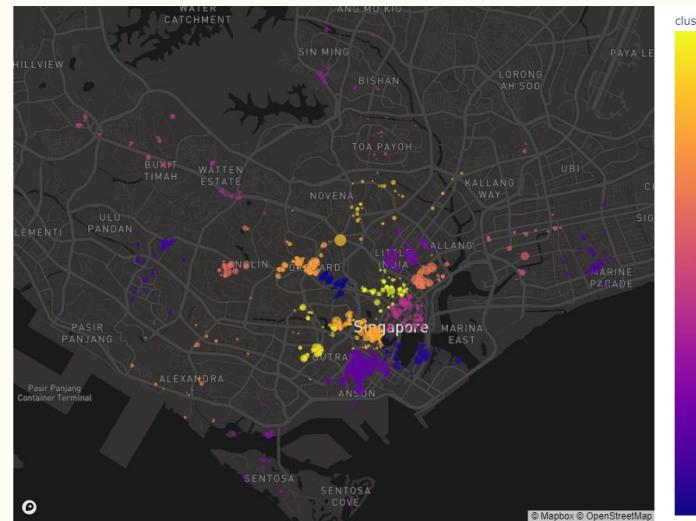
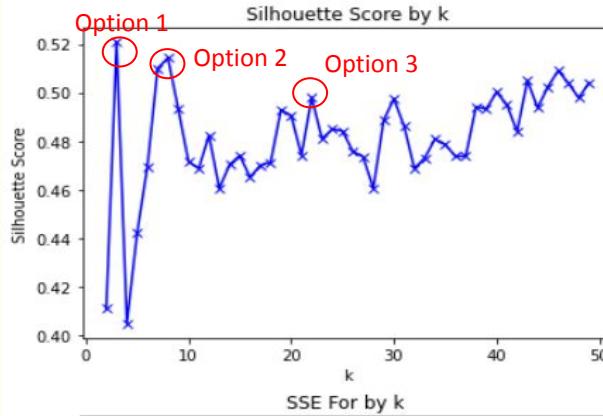
# Location-Based Recommender



## K - Means Clustering

- ❖ A method of vector quantization.
- ❖ The goal is to partition the data into sets of points, such that the total sum of squared distances from each point to the mean point of the cluster is minimized.
- ❖ One method of evaluating results is Silhouette Score. A metric used to calculate the goodness of a clustering technique.

# Location-Based Recommender

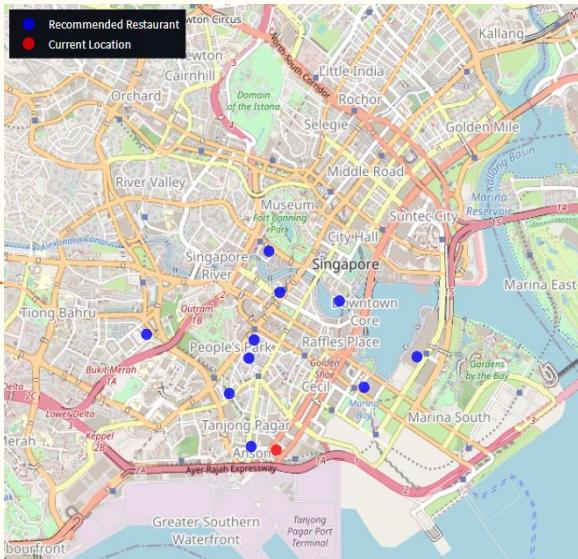


## Observations

- ❖ Optimal cluster is  $k = 3$ . It would be better to have more clusters that are smaller and more spread out.
- ❖ As less cluster might results in easily repeated recommendation and also discourage users from visiting the restaurants that we recommend if they are too far from their location.

# Location-Based Recommendation Model Case Study

Case Study :  
79 Anson Road, Singapore



Name	Address
Hokkaido Ramen Santouka	6 Eu Tong Sen St, #02-76, Singapore 059817, Singapore
NUDE Seafood	Street Level, 12 Marina Boulevard, Marina Bay Financial Centre Tower 3, Singapore 018982, Singapore
Old Chengdu Sichuan Cuisine	80/82 Pagoda St, Singapore 059239, Singapore
Dragon Phoenix	177A River Valley Rd, #06, Singapore 179031, Singapore
Man Man Japanese Unagi	1 Keong Saik Rd, #01-01, Singapore 089109, Singapore
Lian He Ben Ji Claypot Rice	335 Smith St, 02-197 - 199, Singapore 050335, Singapore
Empress	1 Empress Pl, #01-03, Asian Civilisations Museum, Singapore 179555, Singapore
YAYOI	100 Tras St, #03-12, 100 AM, Singapore 569922, Singapore
Imperial Treasure Fine Chinese Cuisine	10 Bayfront Ave, #02-04, Singapore 018956, Singapore
Tiong Bahru Bakery	56 Eng Hoon St, #01-70, Singapore 160056, Singapore



# Conclusion

# Conclusion & Looking Forward

## Conclusion

- ❖ Evaluated LightFM code package and compared the performance of its pure and hybrid models for similar restaurant recommender.
- ❖ Hybrid LightFM's outperformed pure collaborative filtering model not only in terms of AUC score but also precision & recall score.
- ❖ However, for new users there is an existing problem such as cold start.
- ❖ Thus, location based recommender using K-mean clustering is developed.
- ❖ The location-based recommender aim to provide a quick and dirty service for passing users.

## Looking Forward

- ❖ Scale the recommender to include all subzone area.
- ❖ Use deep learning algorithms to predict more accurate recommendations.
- ❖ Incorporate Graph Theory for location-based recommender to optimize travelling routes.

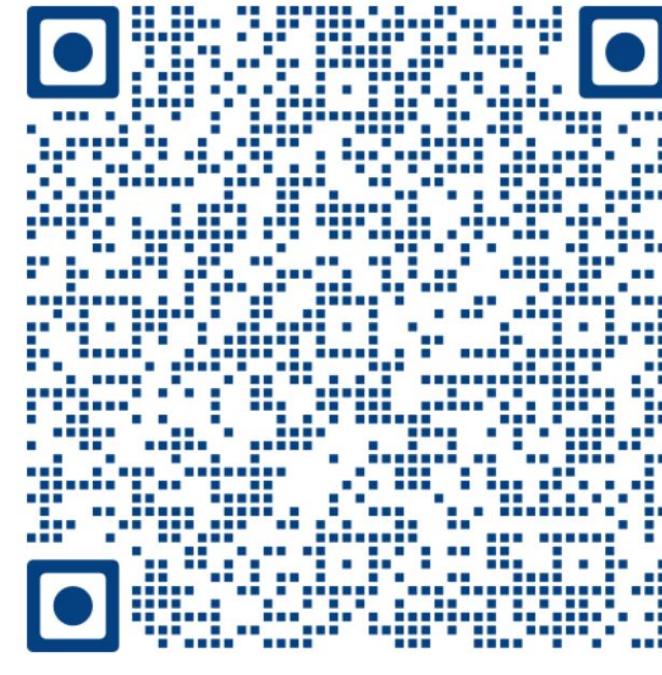


# Web Application

Photo by [Madie Hamilton](#) on [Unsplash](#)



# Deployment



<https://share.streamlit.io/zaviersoon/food-recommender-system/app.py>

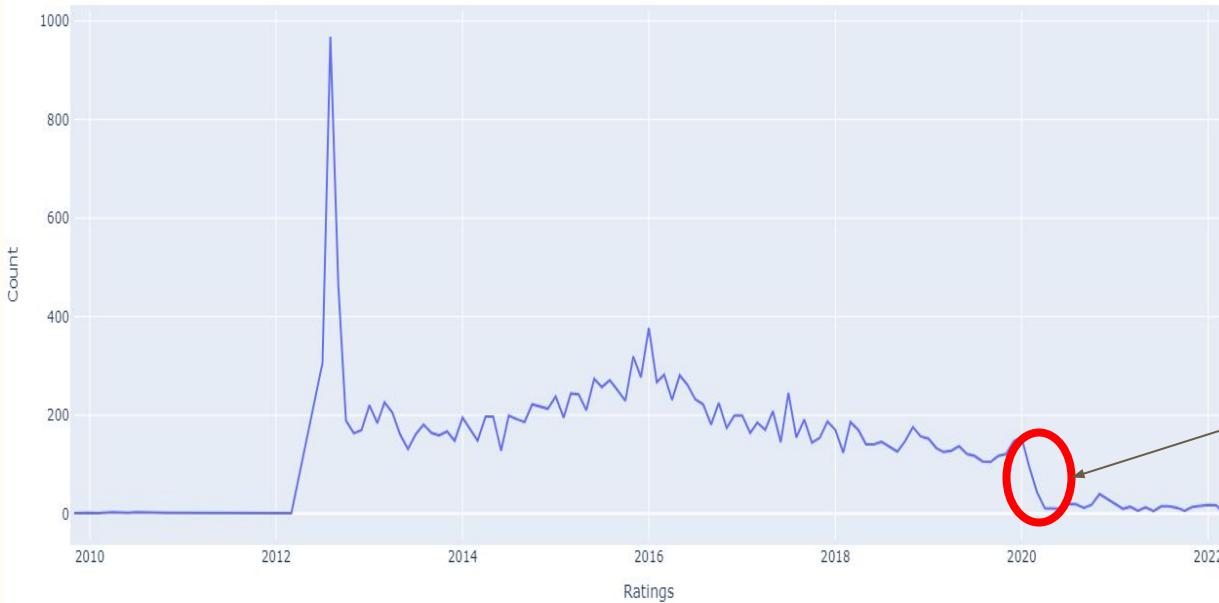


# THANK YOU

*CREDITS: This presentation template was designed by [Freeslides](#). Some of the icons are from [Flaticon](#) & images are from [Unsplash](#), [500px](#) & [sortedmedia](#). Please keep this slide for attribution.*

# EDA on Restaurant Reviews Data

Review trend from 2009 to 2022



Covid-19

