

St. Joseph's University

Bangalore, Karnataka

PROJECT REPORT

ON

Foundation of Data Science

Submitted by:

Niveditha Ramesh-222BDA33

Anson J Maliackal-222BDA35

Mareena Polin -222BDA07

Submitted to:

JAYATI BHADRA

Head of Department

Department of Advanced Computing

St. Joseph's University

Recommendation engine based on graph theory

INTRODUCTION:

In today's world, there are countless options for dining out in Singapore, making it difficult for people to choose the perfect restaurant for their taste and budget. This is where recommendation systems come in. A recommendation system is a software program or algorithm that provides personalized suggestions to users based on their past behavior, preferences, and interests.

For our project, we are building a recommendation system for restaurants in Singapore based on their rating. We are using a combination of Networkx, seaborn, matplotlib, numpy, pandas, and ipywidgets to build a graph-based recommendation system. By representing the restaurant data as a graph and using graph-based algorithms to analyze the relationships between different entities, we can identify patterns and make personalized recommendations to users.

Our recommendation system is designed to help users find the perfect restaurant based on their preferences and budget. By analyzing the rating data for restaurants in Singapore, we can make accurate and personalized recommendations to users based on their past behavior and preferences. With the help of our recommendation system, users can enjoy dining out in Singapore like never before.

LITERATURE REVIEW:

In recent years, recommendation systems have become an increasingly popular research topic in various fields. With the rapid growth of the internet, e-commerce, and online platforms, recommendation systems have been widely used to provide personalized recommendations to users, which can improve their user experience and satisfaction.

In the field of restaurant recommendation systems, there have been various approaches proposed in the literature. Collaborative filtering and content-based filtering are two commonly used techniques. Collaborative filtering analyzes the user's past behavior and the behavior of other users to make recommendations, while content-based filtering analyzes the characteristics of items and matches them with users' preferences.

Graph-based models and algorithms have also been proposed in the literature for recommendation systems. By representing the data as a graph and analyzing the relationships between entities, graph-based models can capture the complex dependencies between users and items, leading to more accurate and personalized recommendations.

In terms of tools and libraries used for building recommendation systems, NetworkX, Seaborn, Matplotlib, NumPy, Pandas, and ipywidgets are commonly used in the literature. These libraries provide various functions and tools for data analysis, visualization, and modeling, which can be utilized to build effective and efficient recommendation systems.

For the specific context of your project, building a recommendation system for restaurants based on ratings in Singapore, the literature provides a strong foundation for using collaborative filtering, content-based filtering, and graph-based models to generate personalized recommendations for users. The use of the aforementioned tools and libraries can aid in the development of an effective recommendation system, which can benefit both users and restaurant businesses in Singapore.

AIM OF THE WORK:

The aim of the proposed work is to develop a recommendation engine based on graph theory that can provide personalized and accurate recommendations to users.

METHODS AND MATERIALS:

Data Collection:

The data for this project was obtained from a publicly available dataset containing information about restaurants in Singapore. The dataset includes various attributes such as restaurant name, location, cuisine type, price range, and user ratings.

Data Preprocessing:

The data was preprocessed using Python libraries such as Pandas and NumPy. The preprocessing steps included removing duplicates, handling missing values, and converting categorical data into numerical data.

Data Visualization:

The data was visualized using Python libraries such as Matplotlib and Seaborn. The visualization technique used is boxplot. Boxplot was used to identify outliers during the process of data cleaning.

Graph Creation:

The restaurant data was converted into a graph structure using the Networkx library in Python. The nodes in the graph represented the restaurants, and the edges represented the relationships between the restaurants based on their ratings.

Recommendation System:

The recommendation system was built using the graph created with the help of networkx library in python. The system uses collaborative filtering to recommend restaurants to users based on the rating.

User Interface:

The user interface for the recommendation system was built using the ipywidgets library in Python. The interface allows users to input their preferences and receive personalized restaurant recommendations based on those preferences.

EXPLORATORY DATA ANALYSIS:

Exploratory data analysis (EDA) is a critical step in any data science project, including the development of a recommendation system for restaurants based on Singapore ratings. EDA helps to understand the characteristics and distribution of the data, identify outliers, and inform the development of statistical models.

In this project, the following EDA techniques were used:

- 1.Data Cleaning: The first step in EDA is to clean the data. This involves removing duplicates, handling missing values, and ensuring that the data is in the correct format for analysis.

```

Jupyter 222BDA33,35,07 Last Checkpoint: an hour ago (autosaved)
Python 3 (pykernel)

File Edit View Insert Cell Kernel Widgets Help
Run

currency      10
delivery_cost  51
lat           0
lon           0
opening_hours  0
image_url     11
radius        0
rating        2132
reviews_nr    2534
delivery_options 0
promo        7359
loc_type      0
delivery_by   0
delivery_time 183
dtype: int64

In [5]: df.dropna(subset=['name'], inplace=True)

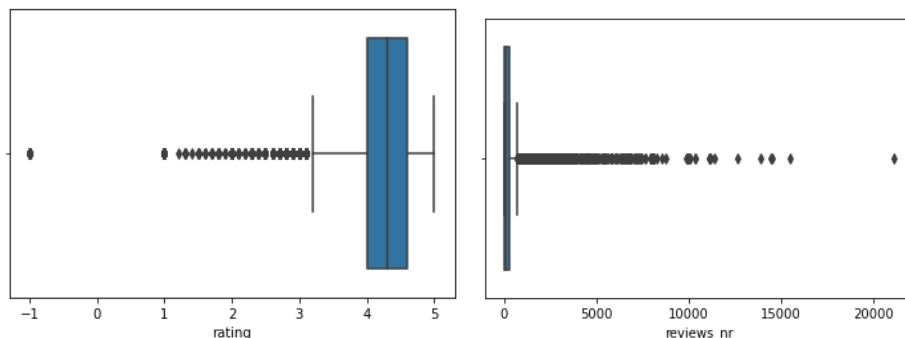
In [6]: df[['cuisine']].dropna(inplace=True)

In [7]: df[['cuisine']].unique()
Out[7]: array(['Western', 'Fast Food', 'Halal', 'Chicken', 'Beverages'],
              ['Healthy', 'Sandwiches', 'Fast Food', 'Halal', 'Western'],
              ['Same Prices In-Store', 'Beverages', 'Local', 'Bubble Tea'],
              ..., ['Chicken', 'Korean', 'Fried Chicken'],
              ['Asian', 'Bakery & Cake', 'Breakfast & Brunch', 'Singaporean'],
              ['Japanese', 'Noodles', 'Grill'], dtype=object)

In [8]: df[['cuisine']].value_counts()
Out[8]:
["Chinese"]      189
["Asian", "Chinese"] 175
["Local", "Chinese"] 135
["Western"]      134
["Local"]        120
...
["Chinese", "Seafood", "Fast Food"] 1
["Steak", "Western", "Ramen", "Islandwide Delivery"] 1
["Beverages", "Dessert", "Durian"] 1
["Asian", "BBQ", "Fried Chicken", "Malaysian", "Snack"] 1
["Japanese", "Noodles", "Grill"] 1
Name: cuisine, Length: 7488, dtype: int64

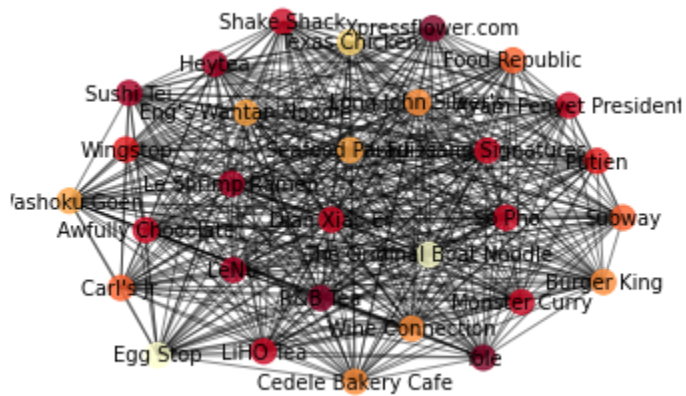
```

2.Data Visualization: The data was visualized using Python libraries such as Matplotlib and Seaborn. The visualization technique used is boxplot. Boxplot was used to identify outliers during the process of data cleaning.



Overall, EDA was an important step in this project as it helped to identify key patterns and relationships in the data that were later used to develop the recommendation system.

OUTPUT:



Mrs Pho

Restaurants similar to LeNu:

1. R&B Tea
2. Le Shrimp Ramen
3. Heytea
4. Ayam Penyet President
5. Eng's Wantan Noodle

Restaurants similar to Washoku Goen:

1. R&B Tea
2. Vole
3. Le Shrimp Ramen
4. Heytea
5. LeNu

Restaurants similar to Potato Corner:

1. R&B Tea
2. Vole
3. Xpressflower.com
4. Le Shrimp Ramen
5. Heytea

Restaurants similar to Pizzakaya:

1. Beyond Coffee
2. Tamago-En
3. innisfree
4. Lush
5. Da Paolo Gastronomia

Restaurants similar to Mrs Pho:

1. Beyond Coffee
2. Tamago-En
3. innisfree
4. Rive Gauche
5. Lush

GITHUB REPOSITORY LINK:

<https://github.com/Niveditha-RAMESH/RESTAURANT-RECCOMENDATION-SYSTEM-/blob/main/Reccomendation%20engine%20for%20restaurants%20based%20on%20rating.ipynb>

<https://github.com/MareenaTimothy/Recommendation-engine-based-on-graph-theory>

REFERENCES:

- Adomavicius, G., & Tuzhilin, A. (2005). Toward the next generation of recommender systems: a survey of the state-of-the-art and possible extensions. IEEE Transactions on knowledge and data engineering, 17(6), 734-749.
- Huang, M., & Sun, Y. (2013). A survey of collaborative filtering based social recommender systems. Computer Communications, 41, 1-10.
- Zhang, X., & Chen, M. (2016). A graph-based recommendation algorithm for cold-start problem in E-commerce. Journal of Computer and System Sciences, 82(7), 1140-1155.
- Chen, L., & Liu, N. (2018). A recommendation system for restaurant selection based on big data analytics. Journal of Big Data, 5(1), 1-13.