Building a recommendation system based on the reviews on attractions across Europe

Abstract

In recent days the recommender systems play an essential role in the entertainment industry particularly in the travel and tourism section. Social media platforms like Facebook, Instagram, Twitter, Travel Advisor and Airbnb provide huge volume of data in the form of reviews, forums, blogs, feedbacks, etc. This data can be an essential input for leisure activity recommendations. The ultimate goal of this experimental analysis is to use clustering algorithms and create a recommendation system using the social media datasets that are related to travel and tourism.

Additional Key Words and Phrases: Travel Recommendation Systems, Clustering Algorithms, Europe Tourism Analysis

Background

The goal of this project is to help travellers plan their destinations and associated activities, with minimal information. To achieve this, the abundant social media data available on the internet will support. In this enormous quantity there might be some redundant or false data. In such cases recommender systems help in filtering relevant information from this huge volume of data that is rapidly increasing. This data analysis project will also help the hotel, restaurant and leisure activity providers to work on real, tangible improvements in a company's business performance.

Literature Review

There are several experiments conducted to research the travelling patterns and destinations. Dr. Shini Renjith worked on the evaluation of partitioning clustering algorithms to process the social media data in tourism domain.

Methodology

This dataset is contributed by Dr. Shini Renjith and contains the travel reviews on attractions from 24 categories across Europe. Google user rating ranges from 1 to 5 and average user rating per category is calculated. Each traveller rating is mapped as Excellent (4), Very Good (3), Average (2), Poor (1), and Terrible (0). The attributes included in the dataset are Churches, Resorts, Beaches, Parks, Theatres, Museums, Malls, Zoo, Restaurants, Pubs/Bars, Art Galleries, Dance Clubs, Swimming Pools, Gyms, Bakeries, Beauty & Spas, Cafes, View Points, Monuments and Gardens. With this data we propose to apply K- means clustering algorithm to build a model that can predict the most visited attraction attribute based on the review records on the Trip Advisor platform.

Observations

Some attributes such as **pubs, bars and restaurants** have wide-range of rating distribution, as they are the common activities which most of tourist enjoy. Few attributes such as **gyms, bakeries and swimming pools** are marked with relatively low rating. It might be interesting to find out why most of users gave low rating to these types of attractions. Since we don't have enough information about attraction itself or any descriptive user reviews, we'll focus on segmenting users into different cluster based on their preferences. Malls have the highest average rating and gyms have the lowest average rating implying that travellers prefer malls than gym. It is a common phenomenon that gyms are not usually visited by all the tourists. Entertainment has the highest average rating and Services have the lowest rating implying that people are more interested in entertainment.

Overall Rating Distribution

No of reviews for each category

Average rating per Category
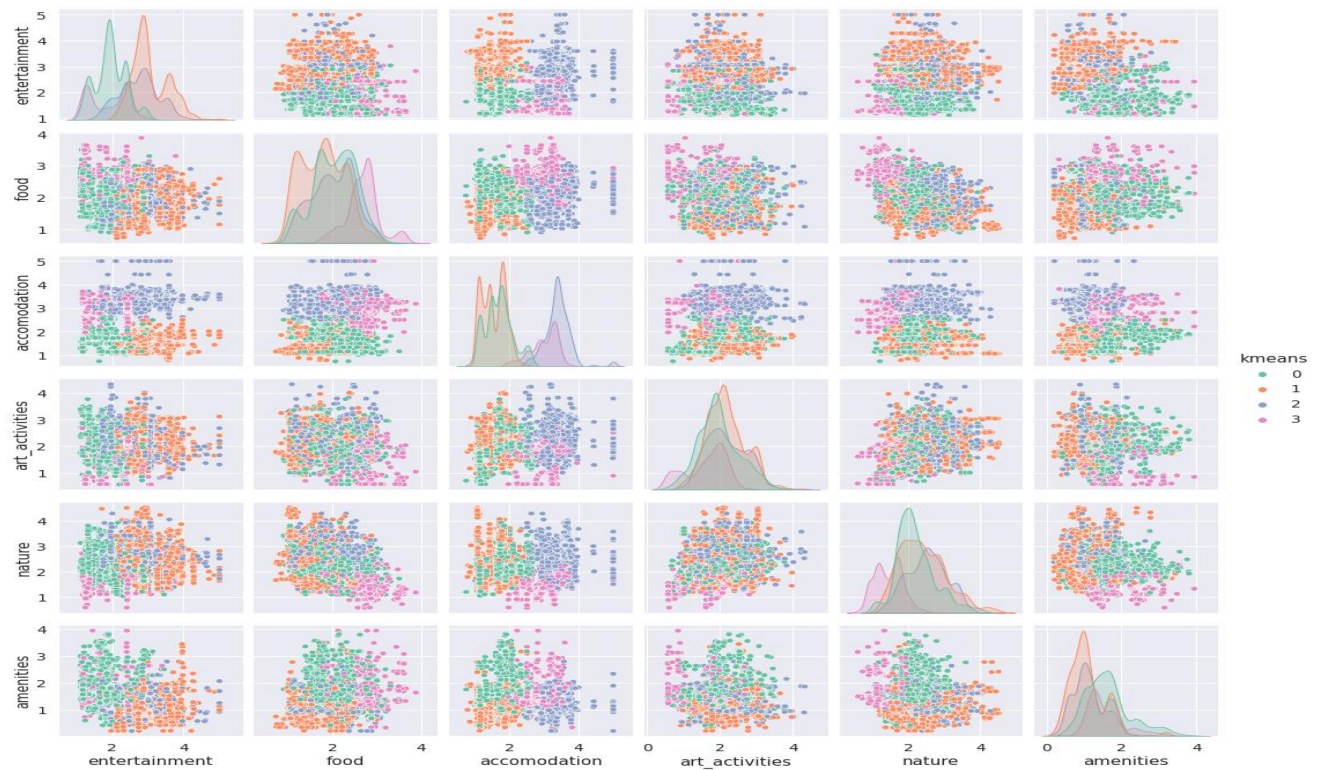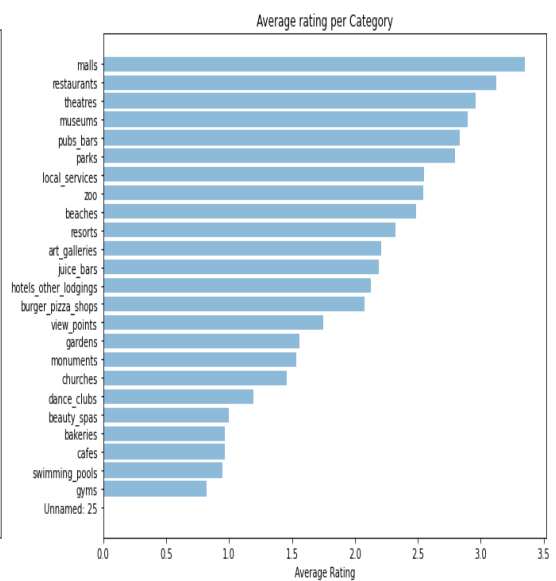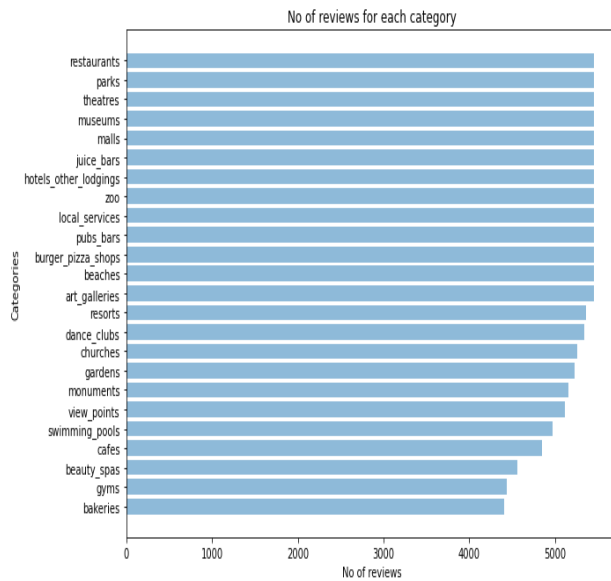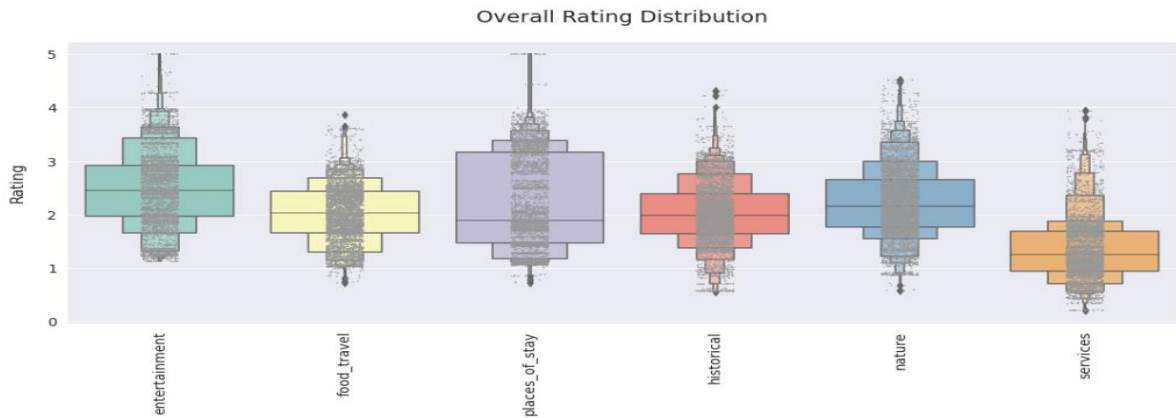
Conclusion

The output of this project denotes the mostly visited category which is the **entertainment** attributes that include **theatres, dance clubs and malls** among the entire set of attributes mentioned above. This work depicts that behavioral data of the customer can always be used as an input to clustering process. In this work, we considered user reviews, feedbacks and rating information captured from google reviews for attractions. However, travel service providers may have multiple options to record user traits and interests by tracking the types of queries coming to them, taking direct feedback with the help of questionnaires or surveys, tracking the user transactions and monitoring the reviews on travel forums and portals. Depending on the data volume and its distribution pattern in consideration, they can adopt optimistic clustering algorithms to segment their customer base so that they can meet the needs of their target customers and appropriate travel solutions can be offered.

.
References
S. Renjith and C. Anjali, "A personalized mobile travel recommender system using hybrid algorithm," 2014 First International Conference on Computational Systems and Communications (ICCSC), 2014, pp. 12-17, doi: 10.1109/COMPSC.2014.7032612.
S. Renjith, A. Sreekumar and M. Jathavedan, "Evaluation of Partitioning Clustering Algorithms for Processing Social Media Data in Tourism Domain," 2018 IEEE Recent Advances in Intelligent Computational Systems (RAICS), 2018, pp. 127-131, doi: 10.1109/RAICS.2018.8635080.+

https://github.com/titov-vladislav/Travel-Review-Rating-Clustering