

Project Title - Guest Segmentation using RFM Analysis and K-Means Clustering in a Hotel Booking System

Research Question - "How can RFM analysis combined with K-Means clustering be utilized to effectively segment hotel guests and improve targeted marketing strategies at Blue Water System Hotel (Exploratory research question)"

Project Objectives

- Collect and preprocess data to ensure it is clean and structured for analysis.
- Analyse the booking behaviour and spending patterns of each guest to determine their recency, frequency, and monetary values.
- Utilize a hybrid methodology that starts with rule-based segmentation and validates it through direct K-Means clustering to categorize guests according to their RFM scores.
- Utilize the knowledge of customer segmentation to develop targeted marketing strategies aimed at enhancing customer satisfaction and boosting hotel income.

Summary and Background

The customer is an asset for the organisation and should be given personalised treatment. The organisation may utilize its client database to do customer segmentation and profiling, which is highly beneficial for customer relationship management (CRM) and targeted marketing.

[2] RFM, which stands for Recency, Frequency, and Monetary value, is a crucial factor in customer segmentation. It considers the client's most recent visit, the frequency of their visits, and the revenue generated by the customer. This study employed sales data to compute RFC values and subsequently assigned a score ranging from 5 to 1 based on RFC values. Afterward, derive the RFM parameters for each individual client and organise the data into groups using the k-means clustering algorithm. Once the clustering process is complete, interpret the clusters using decision tree method to gain a comprehensive understanding of the distinct characteristics exhibited by each cluster.

[1] Customer segmentation based on the customer lifetime value (LTV) as input variable. This study focused on a new LTV model that considers the churn rate. The model not only considers the customer's previous profit contribution but also considers their future financial contribution, potential profit creation, and predicted service periods. This research evaluates customer value from three perspectives: current value, potential value, and customer loyalty. [The paper suggests marketing techniques based on the segmentation results obtained by dividing the customers based on these three perspectives.

[3] The study focuses on new customer segmentation model for banking industry based on two input variables, RFM, Demographic variables, followed by assessment of LTV. Perform initial segmentation based on RFM analysis and use the RFM score as input values to the K-means clustering. Performed further segmentation based on Demographic data. Education level, Occupation level and age are chosen as top three influencing variables. Finally, perform LTV analysis to assess the value of each customer cluster.

References

[1] Hwang, H., Jung, T. and Suh, E., 2004. An LTV model and customer segmentation based on customer value: a case study on the wireless telecommunication industry. *Expert systems with applications*, 26(2), pp.181-188.

[2] Kim, S.Y., Jung, T.S., Suh, E.H., and Hwang, H.S., 2006. Customer segmentation and strategy development based on customer lifetime value: A case study. *Expert systems with applications*, 31(1), pp.101-107.

[3] Namvar, M., Gholamian, M.R. and KhakAbi, S., 2010, January. A two-phase clustering method for intelligent customer segmentation. In *2010 International conference on intelligent systems, modelling, and simulation* (pp. 215-219). IEEE.

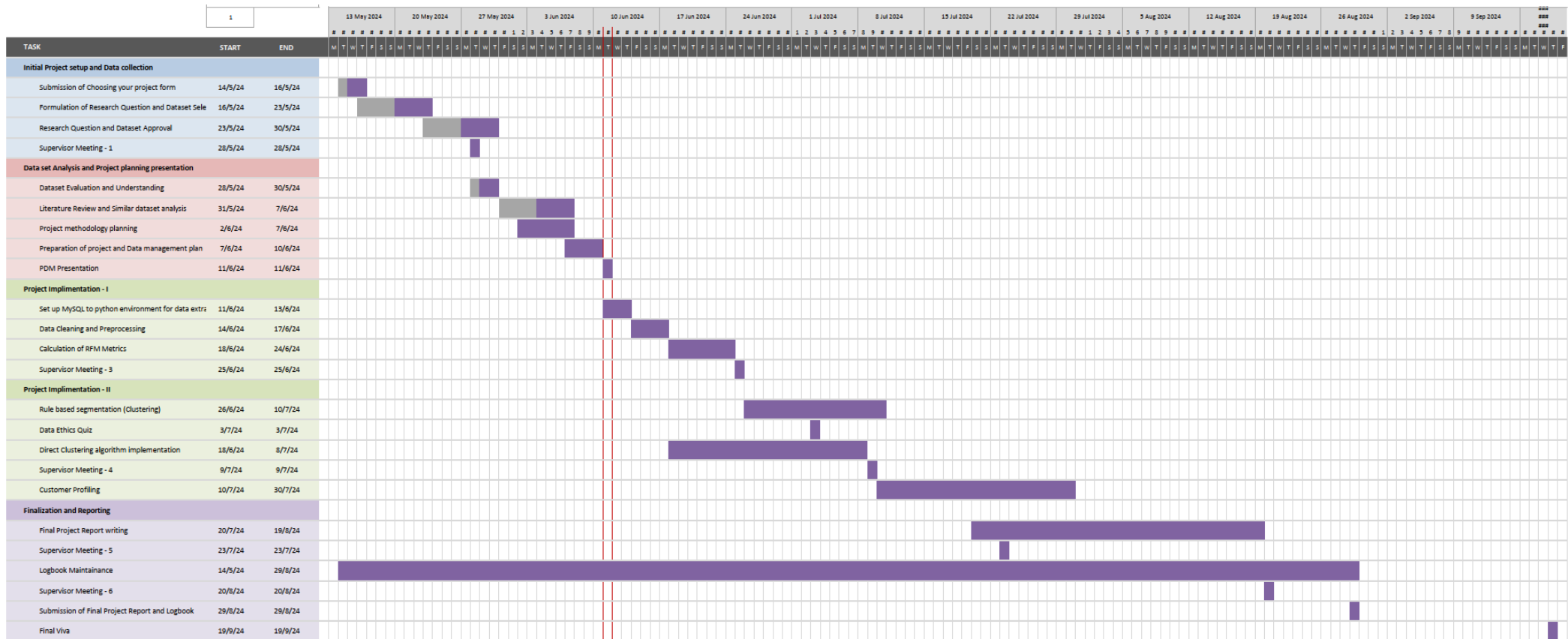
Task List

Initial Project setup and data collection – Develop research question, select data set, and obtain approval for the research question and dataset.

Data analysis and Project planning Presentation – Analyse data, review similar works, plan project methodology and develop a detailed PDM.

Project Implementation – Extract MySQL data to python environment, data cleaning and preprocessing, RFM analysis, Perform Rule based and Direct k-means clustering for guest segmentation and guest profiling.

Finalization and Reporting – Maintaining project logbook, project final report writing, submission of reports and final viva.



Overview of the dataset

The dataset originated from the Blue-Water Lake resort by Serendib, a hospitality establishment situated in Malawi. It was provided by Seeduwa Software Development Solutions (SSD) a Sri Lankan company which is the main ICT provider for the companies like Globe Internet Limited - Malawi, Serendib Hotels & Resorts – Malawi as well as Globe Computer Systems Ltd – Malawi. MySQL database contains approximately 340 tables covering all aspects of Blue-water hotel's operations. For my research, I consider the guest reservation details that includes guest details, reservations data, and expenses.

- Guest Data – contains guest demographics and contact details.
- Reservation Data – contains guest reservation details (check in check out dates, room types)
- Expenses – contains the room chargers, additional service chargers.

The data enables the hotel management to comprehend guest demographic data, behaviour, preferences, and spending patterns, thereby enhancing marketing strategies.

Summary of Data

The data is stored in a MySQL relational database. Reservations table contains 82466 records, guest table consist of 6323 records whereas expenses table containing 113,443 records. Expected size of the data files – expenses table (26.4 MB), Guest table (1.50MB), Reservations table (15.7MB)

Document Control

GitHub repository will be utilized as the central hub for the project development, backup and tracking which ensures the project components are organized, backup and versioned effectively. Regular commits will be published to the repository to ensure that all modifications are recorded and recoverable.

- GitHub Address - https://github.com/samindika/Hotel_Guest_Segmentaion_RFM_Kmeans
- Data Files – data/reservations_data.csv, data/guest_data.csv, data/expense_data.csv
- Code Files – src/k-meansclustering.py.
- Documentation – store in the root directory (docs/data_management_plan.pdf)

A comprehensive Readme file is provided with the repository includes instructions for users and code descriptions for anyone interest in using or comprehend the data and code related to this project.

Security and storage

- Local backup - computer's local copy of the repository is up to date with the GitHub repository.
- Cloud backup – use one drive to periodically store the entire repository backup for extra safety.

Version control

- Commit changes – Use descriptive commit message for each change.
- Branching – implement branches for various aspects of the project. (Branches for RFM analysis, k-means clustering)
- Merging - Use pull requests to merge changes from feature branches into the main branch.
- Tags and Release – Use tags to mark significant versions.

Ethical Requirement

This is a proprietary dataset from the Blue Water Lake Resort and accessed through the established channels between SSD and the Malawian software business. This is done while adhering to data privacy and security rules. SSD has given the email confirmation to use the reservation data, expenses data, and guest data for the project.