**DATA 2206-01 – CAPSTONE PROJECT**

**LAKME Unisex Salon  
Customer Attrition Data Analysis**

**FINAL DATA ANALYSIS REPORT**

A picture containing text, clipart

Description automatically generated

**Professor** – Thomas Francescutti

Done by

Keerthana Kalaivanan – 100837032

Abraham Mathew – 100829875

Mathew Coe -100542534

**Contents**

[Introduction 3](#_Toc88872671)

[**Client Overview** 3](#_Toc88872672)

[**Capstone Objective** 3](#_Toc88872673)

[**Final Project Deliverable** 3](#_Toc88872674)

[**Recommendations** 3](#_Toc88872675)

[**Growth propellers** 4](#_Toc88872676)

[**Threat** 4](#_Toc88872677)

[Business Goals 4](#_Toc88872678)

[Data Analysis Goals 4](#_Toc88872679)

[Data description 5](#_Toc88872680)

[Data Exploration 6](#_Toc88872681)

[**PROCESS FLOW** 6](#_Toc88872682)

[Data Preparation & Transformation Process 7](#_Toc88872683)

[Data Exploration Insights: 9](#_Toc88872684)

[Data Analysis Approach: 11](#_Toc88872685)

[Conclusion & Recommendation to Client 12](#_Toc88872686)

## Introduction

**Client Overview**

Our client Lakme is a leading Unisex Hair & Style Salon offering a full range of grooming solutions (hair, body, and skin) for value conscious consumers They have currently 275 stores across India. While the stylists are employees, their branches are primarily franchises. They also have a luxury offshoot to cater to niche (rich) customers.

**Capstone Objective**

The CEO and CFO of the Group of Salons shared their concerns about customer attrition and the need for predicting attrition levels. The client wanted the capstone team to identify Customer Churn cycles and implement strategies that would improve customer retention and provide enhanced customer experience.

The following were agreed as Deliverables

**Final Project Deliverable**

* **Customer Segmentation:** Analyse and Identify the different customer segments of our client. Customer Segmentation will enable us to identify a customer’s value with the salon and market suitable beauty solution to them
* **Build Prediction Model:** Build a model to predict client churn. This churn model will give us a value which is the maximum number of days between two visits of a client, after which they will be considered as an attrited customer

**Recommendations**

1. We would analyse the raw data and report the irregularities in the way the data is being collected
2. We would suggest a suitable format for data collection in the future
3. We would also suggest new metrics and dimensions that will help the Salon Chain in conducting advance analysis and much more accurate results that can help their business grow

The Salon faces the following growth propellers and threats in current market scenario

**Growth propellers**

* Growing disposable income of middle class, peer pressure and ever-increasing exposure by media are some of the factors truly making India as the hottest market for high-end salon brands in India.
* Customers’ willingness to spend if they are getting something specifically different and much better than the regular services provided by mom-and-pop salons.

**Threat**

* International Brands coming into the domestic market can be a threat to the market share held by Client’s.
* Availability of cheap local brands and imitations

Due to these factors, the Salon is engaging us to achieve the following Business and Data Analysis goals

## Business Goals

Salon foresees challenges in sustaining the business against its competitors and wants to acquire new customers as well as avoid customers from leaving them and seek services from the competitors. Given the entry of international beauty brands in the market, Client realizes the importance of Customer retention for both sustenance and growth. Unlike other entities such as Banks where we can know a customer is an attrition when they close their account, the Salon does not know for sure if a customer is an attrition. To save the business, they would like to know the customer churn cycle to control attrition, attain retention, calculate customer lifetime value, segment them based on recency, frequency and monetary factors and finally improve marketing to the customers so that they will prefer us over the competitors.

## Data Analysis Goals

Based on the raw data that was provided for 2 salon chains which has information about transactions, memberships and products, the data analysis goals are as follows,

1. Cleanse the data to a machine-readable format for predictive modelling
2. Remove noise from data
3. Explore data for outliers and segment for further analysis
4. Derive additional metric for Recency, Frequency and Monetary analysis
5. Complete descriptive analysis on the data
6. Visualize data for insights
7. Complete predictive analysis to find the customer churn cycle
8. Provide prescription on data quality
9. Suggest additional metrics to be captured for future analytical activities

## Data description

* Client agreed to share the data from 2 branches for our study
* A branch where customer attrition is high
* A branch where customer attrition is low
* Customer and Transaction level Master data for 2 branches were shared as Excel data extracts
* Master Data covered the time of 12 months
* A total of 9999 unique clients were identified across two salons

**Meta Data Information**

|  |  |  |  |
| --- | --- | --- | --- |
| **Data Set Name** | **Business Description** | **Key Business Elements** | **Challenges** |
| Client Master | Client Information Repository of all Active Clients | * Salon Name * Client Id * Name * Phone * DOB * Client Create Date. | * Bad client id values (starts with “.”, combination of name & number * DOB values incorrect * Emails missing |
| Transaction Ticket Master | Customer Transaction Repository. | * Salon Name * Client Id * Name * Ticket Id * Transaction Date | * Has client ids that are not present in client master * Data Entry incorrect (high individual transactions) |
| Membership Master | Membership Information of Customer | * Salon Name * Member card no Client Name * Phone * Membership date | * No proper link with client master * Data inconsistent |
| Transaction Ticket Service Detail Master | Services (Hair/Skin/Body) taken by Customer | * Salon Name * Ticket ID * Service Description Discount Amt | * Client level data are not correct. * Client ids missing in client master |
| Service Master | Meta Data about Services | * Service Codes * Service Names * Rate Card Type | * Meta data table of the Hair, Skin and Body services offered |

## Data Exploration

**PROCESS FLOW**

Diagram

Description automatically generated

**Assumptions & Constraints:**

* Data set is provided for only 2 branches. Insights will be restricted only to the 2 branches
* Invalid Client Ids were identified in Client Master dataset and were not used for prediction
* Client Master has the list of all active clients only
* Client Id in Client master should be unique
* Churn Prediction was done based on the data in Transaction master as it had reasonable quality data

**Steps Involved:**

* 1. Exploratory Data Analysis was performed on
* Datasets given by client
* Client Master
* Transaction Service Master
* Membership Master
* Service Master
  1. Data Exploration revealed the following
* Identified a client in the client master and transaction ticket master datasets based on “Client ID “
* Performed Data Classification to find Qualitative and Quantitative Business Elements
* Attrition or churn cycle of client was calculated based on
* Data from transaction ticket master
* Recency of visit – found via transaction date from “Transaction Ticket Master”
* Frequency of visit – found via no of ticket ids for a client from “Transaction Ticket Master”
* Number of days between each client visit (days difference between each visit)
* High level client segmentation was done based on client Gender as data was good
* Next level segmentation was done on arriving at a score based on recency and frequency of visits
  1. Created Derived values to calculate
* Frequency of visits
* Recency of visits
* No of days between visits
* Segment Customers based on
  + - Gender
    - Services Offered
    - Membership
    - Recency, Frequency of Visits
* Identified Outlier data
* Data Exclusion - Exclude Customers that have more than 50% missing values

**Techniques & Tools**

* RFM Analysis
* Power BI
* MS Excel

## Data Preparation & Transformation Process

* The Data Exploration revealed the following information
* Attrition data needs to find at customer level using transaction level information
* Data largely present in Transaction Ticket Master dataset (transactional data)
* Need for creating a new master dataset which has both customer and transaction level information
* Created a new Dataset called “**Customer Activity**” to contain the Key Business Elements below
* **Transaction Level Data**

Transaction level information for every client between 2014 and 2015 from “Transaction Ticket Master”.  Key Business Elements include (but not limited to)

* Ticket Id
* Transaction Date
* Client Name /Phone/Email
* Transaction Amount
* **Customer Level Data**

Mapped each client from Transaction Ticket Master with “Client Master” dataset and added client information. Key Business Elements include (but not limited to)

* Client ID
* Ticket ID (1 ticket = 1 transaction)
* Dob, Phone, Email etc

* **Membership Data**

For every client in “Transaction Ticket Master” identified membership information from “Membership Master”. Key Business Elements include (but not limited to)

* Membership Card No
* Name
* Location
* Frequent service
* **Transaction Service Data**

For every client in “Transaction Ticket Master” identified Client Service-related details from “Transaction Ticket Service Detail Master”. Key Business Elements include (but not limited to)

* Ticket ID
* Name
* Service Desc
* Retail price
* The data in “**Customer Activity**” gave us the following derived values
* ClientID
* Sex
* Frequency
* Recency
* Days Difference between Visits
* Repeat Cycle
* Total Spend
* Average Spend

**Assumptions & Constraints**

* Data sets is given for only 2 branches. Insights will be restricted only to the 2 branches
* Invalid Client Ids were identified in Client Master dataset and were excluded for exploration
* Data Imputation will not be performed on customer level information due to a lot of dependencies
* New dataset was created ONLY with clean data from Raw datasets
* Relationships between master datasets were identified ONLY based on 1 year data and assumed to be consistent across all branches (outside the scope of branches) and all years (outside the scope of time)

## Data Exploration Insights:

Tools used: Power BI

Chart, pie chart

Description automatically generated**Analysis by Gender – By Client Count and Membership**

Chart, bar chart

Description automatically generated

-There are 66% non-members compared to 34% of membership holders

-By count, the number of Male customers is considerably high than female customers

-Also, there are much more male than female membership holder. Male numbers are almost high by 50%

**Analysis of Spend by Sex and Membership**

Chart, bar chart

Description automatically generated

-Overall, we could see that Male clients spend more with the Salon than female clients

-Overall, we could see that membership holders spend more with the Salon than Non-Members

Analysis of Average Spend by Sex and age groupChart, bar chart

Description automatically generated

-By sex and age group combination we could see that Females between the ages of 31 and 40 are the best group while female between 41-50 age group spends the lowest with the Salon.

Other Insights from Data:

-Lots of missing data better record keeping in the future needs to be completed by the salons for better insights

-We have a lot more male customers than female customers

-Female customers on average spend more than male customers

-The demographics that spend most per visit are males and females aged 31-40

-On average a members spend 3K more per year than non-members

-30 to 35% of our customers are members

-Our target audience should be females between the ages of 31 to 40 to get them memberships

## Data Analysis Approach:

For our problem statement, we are required to analyse the data through 3 major methodologies namely

1. RFM Analysis

This analysis is done for the purpose of Customer Segmentation and better marketing of the Salons products

* RFM was done to calculate Recency Frequency and Monetary Parameters of Customers
* The “**Customer Activity”** data from Data Preparation was used to build RFM score for each customer with the help of following Key Business Fields
  + Frequency of Customer Visits
  + Recency (Last or recently visited date)
  + Difference In Days between Visits
  + Money Spent during Visits
* An RFM score was given to each of the observations ranging from 1 to 5 which will be grouped as R-1, F-1, M-1 for a customer who is not worth marketing for and R-5, F-5, M-5 for the best customer in the system

1. Customer Lifetime Value Analysis

Using the scores from the above RFM Analysis, the Customers Life Time Value can be calculated to find out which customers are most valuable to the Salon Chain

Using AHP (Analytic Hierarchy Process) to derive Weightage for Recency, Frequency and Monetary Business Fields. A standard formula for CLV Calculation at Customer Level

|  |
| --- |
| **(Weightage for Recency) X** **(Recency Score) +**  **(Weightage for Frequency)** X **(Frequency Score) +**  **(Weightage for Monetary) X (Monetary Score)**  ***RFM Score range from 1 thru 5 respectively*** |

1. Predictive Model Analysis

This step of analysis is to identify a perfect predictive model for our analysis.

Our initial consideration was to evaluate Logistic regression and Survival analysis to predict the Attrition Cycles and Churn rate of our client’s customers.

**Our client’s dataset has time to event data which depicts**

* Till what point in time the customer has been visiting our client
* What was the last visited date/month/year

Logistic Regression:

Is used when the research objective is focused on whether or not an event occurred, rather than when it occurred i.e., time course information is not used. Instead of building a predictive model for “Y (Response)” directly, the approach models Log Odds (Y)

Survival Analysis:

Survival Analysis is used to analyse the data in which the time until the event is of interest. The response is often referred to as survival time, failure time or event time.

On further discussion we are going forward with the **Survival Analysis** model since is Best Suited for our Client’s Customer Attrition Prediction since we need a period within which a customer doesn’t make a transaction with the client, he/she will be considered an attrition.

## Conclusion & Recommendation to Client

Thus, we can suggest the client with the following Process Recommendations and few other Operational Recommendations as well so that client can keep up the current market competition and make it a pleasure for the customers as well by well catered marketing campaigns. The following are the recommendations for the client

**Process to Improve Data Quality**

* Build a standard system for data management
* Implement data profiling, to discover the structure of the data and understand it
* Define data validity rules to evaluate compliance and quality of the data
* Specify a cleansing strategy to be used with data not compliant to data rules
* Create data quality scorecard
* Remove dependency on technology staff by implementing straight through processing
* Improve training for order entry staff to enhance data entry quality
* Improve technology/application stack to help with prefilling of customer data during order entry/updates

**Client Retention Recommendations**

* 1. Focus on arresting Customer Churn
     1. Customer Outreach programs
     2. Repeat service offers
     3. Family package offers
  2. Focus on customers who have higher intervals between their visits
     1. Bi-monthly calls from branch to understand Customer pulse
     2. Implement Structured feedback process after every visit
  3. Improve awareness, advertisement of Gender based services to arrest churn
     1. Implement women services to improve repeat women clients
     2. Offer Packages to attract women and children

**Enhancing Customer Experience Recommendations**

* 1. Discuss feasibility of Online Appointments through Android Apps and services
  2. Create channel to discuss feedback on services and products online
  3. Provide channel to recognize good stylists via emails/Facebook blogs, tweets

**Employee Franchisee Model Recommendations**

* 1. Improve communication channel between client and franchisees
  2. Implement quarterly audits to oversee Franchisee Performance
  3. Roll out surveys to Franchisee customers to improve Franchisee processes
  4. Create alignment between Client Employees and Franchisee employees by clearly laying out engagement models.