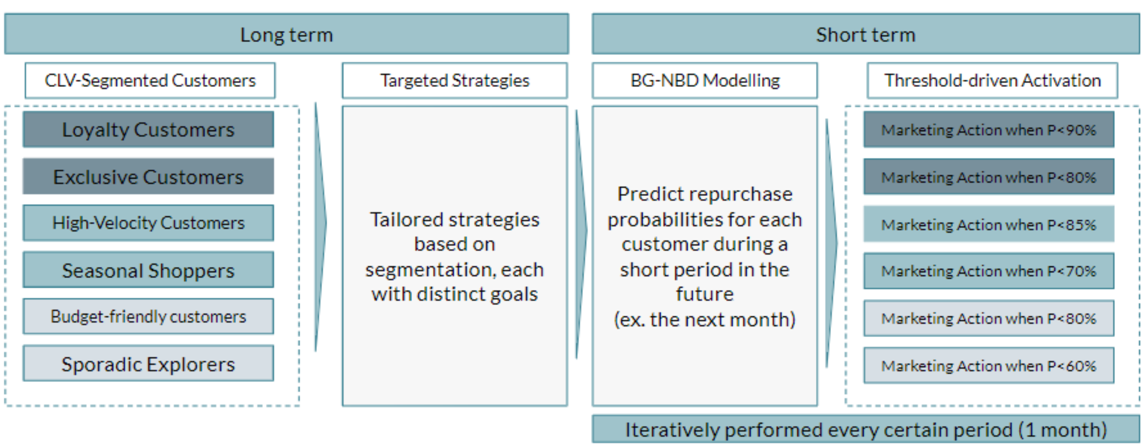
Proof of Concept

Group A

1. Introduction

In today's competitive hospitality landscape, understanding customer behavior and preferences is crucial for success. Marriott, a leader in the global hospitality industry, is keenly aware of the importance of data-driven insights in enhancing customer engagement. To this end, the "RFM-based Customer Lifetime Value Model" project was initiated. This innovative model segments Marriott's diverse customer base using Recency, Frequency, and Monetary (RFM) metrics. A key component of this project is the segmentation of customers through Customer Lifetime Value (CLV) models. By grouping customers based on distinct traits, the model enables the development of long-term strategies tailored to meet the specific needs of each group. Furthermore, the project employs BG-NBD (Beta Geometric/Negative Binomial Distribution) models to forecast the likelihood of customer repurchases. This predictive element allows for the creation of different engagement plans, particularly targeting customers who fall below a certain repurchase probability threshold. The stakeholders for this project include the data science team and the marketing team at Marriott. With the insights garnered from this project, the data science team can refine their algorithms and predictive models, while the marketing team can tailor their campaigns and outreach strategies to resonate more effectively with targeted customer segments. This Proof of Concept (POC) report serves as a testament to our commitment to refining our approach and methodologies, ensuring Marriott stands at the forefront of customer-centric innovation.



2. Success Criteria

In this section, we will present the business functional and technical outcomes integral to the success of our Marriott Customer Lifetime Value project. Our goal is to ensure alignment with the overarching business and technical objectives, fostering a clear understanding among stakeholders and obtaining the explicit sign-off from our sponsor.

a. Business Functional Outcomes

- **Long-term strategies** with RFM-based CLV prediction model and segmentation solution.
- **Short-term Threshold-Driven Marketing Activation strategies** based on iterative BG-NBD model dropout prediction.

The business functional outcomes are important success criteria that will reflect our project's impact on Marriott's business operations and customer relationships. Below are the detailed descriptions of our each outcome, aligning with the business-side success criteria:

i. Outcome 1: Enhanced Customer Value Realization

Using the RFM-based CLV prediction model, our goal is to significantly enhance the value realization from high-value customer segments within the Marriott Bonvoy program.

- Measurable & Verifiable: we can evaluate the success of this outcome by tracking the increase in annual spend from high-value customers in the next financial year. This assessment will be validated using advanced analytics on customer spending patterns and engagement levels, focusing on the high-value segments identified by our CLV model. This method ensures a data-driven approach to measure the effectiveness of our strategies.

ii. Outcome 2: Dynamic Marketing Engagement

Implementing Threshold-Driven Marketing Activation strategies based on our iterative BG-NBD model for dropout prediction, we aim to dynamically engage customers at critical junctures.

- Measurable & Verifiable: we can assess the effectiveness of our strategies by examining customer dropout rates and engagement increase rates within targeted marketing campaigns over the next six months. The measure of success will include monitoring campaign response rates, customer retention statistics, and a decrease in predicted dropout rates as forecasted by our BG-NBD model. This multi-faceted approach will provide a comprehensive view of the impact of our marketing initiatives.

iii. Outcome 3: Enhanced Personalization for Cross-Sell and Upsell Opportunities

Use insights from our RFM-based CLV model and segmentation solution to drive targeted cross-selling and upselling.

- Measurable & Verifiable: we can measure the success of our strategies by observing the increase in cross-sell and upsell rates among targeted segments, within the next financial year. Additionally, the uptake rate for new packages or offers introduced to these segments will serve as an additional metric of success. Our strategy involves closely tracking changes in sales data, with a particular focus on additional purchases and customer responses to targeted offers. This monitoring will be conducted in combination with the insights provided by our modeling, ensuring that the right customers are being targeted at the most appropriate times.

b. Technical Functional Outcomes

The technical functional outcomes are a series of technical milestones we would achieve during the course of the project. Each would lay the foundation for our long-term (CLV-based customer segmentation) and short-term (threshold-driven marketing activation) strategic recommendations.

i. Outcome 1: Engineered Features Capturing RFM Patterns

This outcome is the first technical milestone we achieve, and would be foundational to the proceeding success of both technical and business outcomes. Since the dataset provides only limited features, many of which aggregated, to offer us insights on the RFM patterns of customers, we would utilize feature engineering to extract more useful information.

- Measurable & Verifiable: we would investigate the engineered features using statistical and analytic techniques, and perform continuing version control of our dataset. The successful continuing incorporation of new features can be indirectly measured by the improvements in our predictive model, using metrics like RMSE.

ii. Outcome 2: Tailored RFM-Based CLV Model

Building on the first outcome, the features and the enhanced dataset, we would use machine learning methods to predict the CLVs of each customer, paving the way for further analysis. We define CLV as Customer Value (CV) * Customer Lifespan (CLS): CV is measured by the revenue contributed by each customer in a typical year, and predicted with machine learning; CLS is the time of engagement each customer has with Marriott, and we calculate this by FirstRecency*Frequency, with the rationale that a high first recency (time interval between first transaction batch and last) and a high frequency indicates that a customer is active for a long time, therefore with a potential higher lifespan.

- Measurable & Verifiable: the success of this model can be measured by a variety of tests and metrics. Furthermore, since we may use different machine learning models for different customer segments, the models can also be cross-compared. As in outcome 1, we will have strict version control on the dataset to ensure verifiability.

iii. Outcome 3: Segmentation Solution Based on CLV Model

This outcome is the key technical objective that builds the foundation for actionable long-term and short-term strategies. We segment the customers based on their value, lifespan, and frequency (all of which are obtained by our RFM analysis), dividing them into six segments each with distinct characteristics.

- Measurable & Verifiable: We will cross-examine the customer profiles of each segment with public sources to ensure their reliability.

iii. Outcome 4: BG-NBD Model For Dropout Prediction

This outcome is the key technical objective for our short-term threshold-driven marketing activation strategy. We will build a BG-NBD model for the customer base to examine their probabilities of dropping out, therefore determining the appropriate time for marketing actions by setting thresholds for each segment.

- Measurable & Verifiable: We will perform extensive research on segmentation characteristics to set appropriate and actionable thresholds.

3. Schedule

a. Timeline

In Week 1, our sponsor will start digging into what comes next after implementing the model. They'll have a thorough discussion to understand how the model works and what the results and our recommended strategies imply. This week, key members in the team like project managers and data scientists will work together to address any questions, clarifications, or insights that arise from the initial review.

In Week 2, our sponsor will shift to the data preparation phase. They will dedicate time and resources to make sure that their internal data aligns with the model requirements. This process involves data cleaning, preprocessing, and re-structuring. The goal of this week is to set the stage for a smooth integration in the weeks that follow.

In Week 3, the sponsor will start running the model on their own data. The sponsor's tech team will be in charge of this phase, closely monitoring the model's performance and making any necessary adjustments to enhance its accuracy. At the same time, the team will start preparing for communication with the marketing team.

In Week 4, the tech team will hold meetings with the marketing team. Collaborative efforts will be made to translate the model's outputs into actionable marketing strategies. Our sponsor will be able to identify opportunities and refine target parameters based on our recommended strategies in order to reach the overall marketing goal.

In Week 5, the sponsor will finalize their strategies based on the model results and previous discussions. Additionally, they will initiate the process of communicating the findings across relevant departments, ensuring a mutual understanding of the model and strategies' implications for the company.

b. Risk & Dependencies

Risk 1: Overfitting Models

Since we plan to build our prediction model based on the data and features, we may encounter overfitting, which may lead to poor generalization and inaccurate customer segmentation. To address this risk, we will fine-tune the hyper-parameters, adjust our models, and evaluate our models using metrics such as RMSE and ROC-AUC.

Risk 2: Unsatisfactory Feedback

If we receive unsatisfactory feedback from our professor, we may need to make significant revisions to our model or strategies, potentially resulting in unexpected delays in our preparation for the client presentation. To address this risk, we need to allocate spare time for the modifications after the rehearsal.

Risk 3: Miscommunication

We are divided into tech and business teams. There is a risk of miscommunication between the tech and business teams, which could lead to a lack of coherence in the final combination. To mitigate this risk, we need to develop efficient communication channels between the two teams and establish the meeting on a regular basis.

Risk 4: Incomplete Data Source

The data source used for the analysis may be incomplete, lacking essential demographic and other relevant information, which could result in inaccurate customer segmentation and decision-making process. To address this risk, we may consider ensuring the completeness of data sources to improve the accuracy of future RFM-based strategies.

Risk 5: Decreased Effectiveness Over Time

The effectiveness of the strategies may decrease over time as customer preferences evolve, resulting in less accurate initial segmentation. To address this risk, the company should regularly review and update its RFM models and customer segmentation to ensure they remain aligned with changing customer behavior.

Risk 6: Customer Dissatisfaction

There is a risk of a customer backlash if they perceive the strategies as overly targeted, which could harm the company's reputation. To address this risk, the company should balance between personalized marketing and respecting customer privacy.

Risk 7: Regulatory Changes

RFM model-based strategies involve collecting and using customer data, which may be impacted by changes in data privacy and marketing regulations. To address this risk, the company should stay up-to-date with relevant regulations and adapt its strategies accordingly to avoid potential legal issues.

Dependency 1: Feature Engineering and Model Construction

The first dependency of our project is on the data processing phase. Since our subsequent analysis and recommendations hinge on the results of our data analysis, any delay in this phase may disrupt the whole schedule. To address this, we need to follow a strict schedule and ensure timely results.

Dependency 2: Strategies and Recommendations

The second dependency of our project lies in the strategic planning phase. In order to prepare for the final presentation, we need to develop sound and practical recommendations. If we encounter challenges or obstacles in this process, we may need to work more efficiently in the later stages.

Dependency 3: Feedback

The third dependency of our project is based on the feedback received from the professor and the sponsor during each week’s presentation and rehearsal. Our team’s ability to quickly implement changes based on the feedback we received is also important for our project.

4. Deliverables

a. Interim Deliverables

i. As an initial deconstruction of the project, we deliver the understanding of the market in which Marriott operates, encompassing a SWOT analysis, a Porter's Five Forces, identification of the key terms and vocabularies used in the hotel industry, issues and challenges faced by the companies, and core competitors.
ii. With the initial understanding, we deliver questions to the sponsor about their market, business, and data.
iii. After the initial understanding, we deliver our team organization and communication method, our project objectives, and the relevance of our project outcome to the client.
iv. With the data preparation and exploratory data analysis, we deliver key points and insights from the EDA and the relevance of each insight to the goal of the entire project.
v. With the data preparation and initial insights at hand, we move on to suggest the initial analysis strategy and approach, as well as deliver the initial modeling approach and results.
vi. We define and deliver the success criteria for the entire project.
vii. Based on the updated model, we update the insights and findings on the estimate of customer lifetime value and refine our segmentation of customers to best identify the top value group.
viii. With the identification of the top customers, we design the marketing strategies for increased customer retention, reactivation, product cross-sell and up-sell.
ix. After developing and fine-tuning the initial CLV model, we provide a preliminary assessment of its performance and accuracy.

b. Final Deliverable

1. **Project Summary and Data Analysis**
   A detailed overview of the project goals, scope, and methodologies used.
   Insights and findings derived from our data analysis, supplemented with visualizations that highlight key trends and patterns in customer behavior.

2. **RFM Segmentation Analysis**
   In-depth analysis of customer segments based on Recency, Frequency, and Monetary (RFM) values.
   Comparative analysis of different customer segments, emphasizing their unique characteristics and potential value to Marriott.

3. **Customer Lifetime Value Prediction Model**
   A fully developed machine learning model designed to estimate the lifetime value of Marriott's customers.

Documentation on model development, including data preprocessing steps, feature engineering, model selection, training, and validation processes.

4. **Recommendation and Strategies**
   Strategic recommendation tailored to various customer segments identified through the RFM analysis.
   Actionable strategies for customer engagement, retention, and reactivation, aligned with Marriott's business objectives. Strategies are grouped into short-term strategies and long-term strategies.

5. **Supporting Documentation and Code**
   All Python Notebooks (*.ipynb) with code for data analysis, model building, and validation.
   Additional documentation providing context and explanation for the code and methodologies used.

5. Research & Critical Analysis

Understanding customer behavior in the hotel industry is both beneficial and necessary for Marriott in this highly competitive sector. We believe that a refined RFM (Recency, Frequency, Monetary) model with multiple features might be an efficient way to segment its diversified customer base and calculate accurate Customer Lifetime Value (CLV) based on thorough study and data analysis.

a. Benefit of the Solution to Marriott

Gaining Knowledge about Customer Preferences: Marriott may obtain vital insight into client preferences by utilizing the RFM approach, allowing for more targeted and cost-effective marketing initiatives.

Expand Business Opportunities: Marriott employs this methodology to identify their most profitable consumers and design marketing campaigns to maximize income.

Marketing Intervention for At-Risk Customers: Our approach segments customers by purchasing patterns and applies BG-NBD model thresholds within each category. This enables Marriott to deploy targeted strategies, promptly re-engaging customers who are nearing the risk threshold."

Upselling and cross-selling opportunities: Gaining knowledge into customers categories enables businesses to more accurately personalize product or service suggestions and improve sales prospects.

b. Solution to Business Problems

While Marriott's existing Bonvoy program is admirable, it emphasizes frequency as the main measure of customer value. Our RFM-based CLV model addresses this shortcoming by considering not only frequency but also when their last transaction occurred and how much money they typically spend; our multifaceted approach provides a much clearer picture of their true customer value, allowing marketing efforts to be targeted more effectively.

Issues or concerns that have yet to be addressed: Despite the fact that our model delivers an extensive examination, some flaws or concerns remain:

Scalability: As Marriott's client base grows and diversifies, our approach will need to be refined on a regular basis to remain effective.

Real-Time Data Integration: While our present model works well with batch processing, we intend to add real-time data integration - an important feature in a dynamic market like hospitality - in future iterations of our model.

While our model takes into account a variety of factors, more research into how cultural and geographic differences influence customer behavior may be required.

In sum, our RFM-based CLV model provides many benefits to Marriott in solving the business problem; nevertheless, we recognize it is an evolving tool and commit to further enhancing it to put us at the forefront of customer-centric strategies in hospitality.

6. Conclusion & Sign-Off

In conclusion, our team has made significant progress in the development of the long-term RFM-based Customer Lifetime Value Model and short-term BG-NBD Model for Dropout Prediction for Marriott. Through comprehensive data analysis and iterative model improvements, we discover crucial aspects for improving client engagement, the Marriott Bonvoy program, and revenue forecasts. Our multifaceted strategy, centered on personalized tactics and data-driven insights, has the potential to transform Marriott's customer-centric activities and secure the company's position as a leader in the competitive hospitality industry. While we recognize that continued adaptation and refinement are required, we are optimistic that our current efforts constitute a significant step forward in accomplishing Marriott's main business and customer relationship goals.

We, the undersigned, hereby certify that the contents of this Proof of Concept report accurately reflect the status and direction of Marriott's "RFM-based Customer Lifetime Value Model" project. We agree to the defined success criteria, timeframe, identified risks, dependencies, suggested deliverables, and analysis of the business problem solution. With our support, we demonstrate our belief in the team's strengths and the potential influence of this project on Marriott's future success.

Sponsor's Name: Marriott International

Sponsor's Signature:

Date:

Project Member’s Name: Yujie Lu, Ya-Chien Yang, Yvette Tong, Ruijie Shi, Kunjia Shi, Xinyi Ma, Yuqi Liu, Yi-An Ko, Xiaofan Ji

Date: 10/26/2023