**RE-INNOVATING BUSINESS PRACTICES**

**Prepared for**

**ABC Academy of Music**

**Prepared by: Group 4**

Anjali Joshi

Krina Patel

Shuvam Jaggi

Srinivas Kyatam

Tanvi Ajmera

**December 3, 2024**

# **Executive Summary**

The Project aims to provide a detailed analysis and actionable recommendations to improve the profitability, customer retention and marketing efficacy of ABC Academy of Music situated in Toronto specializing in music lessons for all age categories.

The objective of the project is to address 3 prominent business questions using the dataset provided:

1. Recognize the most revenue generating music category.
2. Identify current customer churn rate and forecast future trends from 2024-2026 using predictive modelling.
3. Evaluate performance of marketing emails to establish the effective strategies for customer engagement and satisfaction.

For each business question, firstly the dataset was cleaned and prepared to ensure data’s quality and suitability for the analysis by checking missing values, duplicates and inconsistencies in the dataset. In the second step, Exploratory data analysis was performed using some statistical and visual analyses unveiling important patterns in sales, churn rate and email performance. Few features were engineered such as in the third question emails’ open rates were analysed using the feature day of the week that helped reveal customer acknowledgement patterns. To predict the churn rate of customers predictive modelling- ARIMA and ETS models were applied to incorporating seasonality, trends and residual error adjustments.

The key findings of the analysis were as follows:

1. The top 3 highest revenue generating categories are-
   1. Music Together- In School
   2. Music Together- Mixed Age
   3. M45- Music lessons
2. The top 3 most sold music lessons are-
   1. Music Together- Mixed Age
   2. M30- Music lessons
   3. M45- Music lessons
3. Churn rate of the academy in 2023 is 82%.
4. Forecasted churn rate from 2024 to 2026 is 73%.
5. The top 3 days to send marketing emails to customers are-
   1. Sunday
   2. Monday
   3. Friday
6. Multiple emails sent to a customer have a higher open rate than single emails.

According to the results derived from the analysis few of the measures to improve the sense of community and engagement, retention, revenue and optimize marketing can be recommended. The ABC Academy should organize jam sessions, workshops and showcase students’ performance on the social media handles to increase the reach and motivate students at the same time. The academy can reduce the churn rate by offering exclusive lessons to those continuing for multiple terms and celebrate student achievements through loyalty programs which can also result in increase in revenue. Lastly, as derived from the results, it is advised that marketing emails should be sent on Sundays ensuring higher open rate and better customer reach.

This analysis aims to empower ABC Academy of Music to formulate data-driven decisions to enhance the profitability and tailor the operations to strengthen its position in the market.

# **Table of Contents**

[**Executive Summary** 2](#_Toc183872917)

[**Table of Contents** 4](#_Toc183872918)

[**Table of Figures** 5](#_Toc183872919)

[**Introduction** 6](#_Toc183872920)

[**Situational Analysis** 6](#_Toc183872921)

[**Business Questions** 7](#_Toc183872922)

[**Methodology** 7](#_Toc183872923)

[**Results and Discussions** 14](#_Toc183872924)

[**Recommendations** 23](#_Toc183872925)

[**Conclusion** 24](#_Toc183872926)

[**References** 25](#_Toc183872927)

# **Table of Figures**

[Fig.1 Bar Plot of Status 10](#_Toc183860269)

[Fig. 2 Histogram of Membership Length 11](#_Toc183860270)

[Fig. 3 Box Plot of Membership Length 12](#_Toc183860271)

[Fig. 4 Total Quantity by Revenue Category 14](#_Toc183860272)

[Fig. 5 Total Sales by Revenue Category 15](#_Toc183860273)

[Fig. 6 Average Sales per Membership by Revenue Category 16](#_Toc183860274)

[Fig. 7 Summary of Q1 graphs 16](#_Toc183860275)

[Fig. 8 Snippet of new table 17](#_Toc183860276)

[Fig. 9 Total vs. Churned students over the years 18](#_Toc183860277)

[Fig. 10 Yearly Churn Rate (2017-2023) 19](#_Toc183860278)

[Fig. 11 ARIMA Forecasted and Actual Churn Rates (2017-2026) 19](#_Toc183860279)

[Fig. 12 Average open rate by Day of the Week 21](#_Toc183860280)

[Fig. 13 Open Rate: Single vs. Multiple Emails 22](#_Toc183860281)

# **Introduction**

In the dynamic business landscape, understanding customers’ needs and catering to that has become crucial for maintaining one-upmanship.

ABC Academy of Music is in Toronto which offers a wide variety of music lessons for different ages, including babies, toddlers, children, teens and adults. Both private and group lessons in instruments such as piano, violin, guitar, drums, and many more are provided by the academy. The lessons can be in-person, online or in a hybrid format. University-trained instructors are appointed, which ensures high-quality learning experience for the customers. It has been in business since 2003 and modifies lessons according to students' interests, including specific styles and songs.

Businesses are always looking forward to opportunities and suggestions to grow and increase their productivity, which is directly proportional to profitability.

So, the objective of this project is to analyze the performance of different revenue categories, find out the customers’ churn rate and use the historical data to create a predictive model to forecast the future churn rate. Also, to derive the open rate of emails to formulate effective marketing strategies to promote sales, retain customers. These analyses would help us provide ABC Academy of Music with some recommendations to improve the profitability of their company and flourish among its competitors.

# **Situational Analysis**

**Mission:** “To provide students of all ages with the skills they need to enjoy music for life”.

**Vision:** “To Save Music by setting children and families on the path of a life of musical activity as a primary and necessary activity in life”.

**Industry Key success factors**

A music company works in a competitive industry which demands both artistic creativity and business strategies to be successful. Some of the numerous factors include Qualified and passionate Instructors, Diverse learning programs, Marketing and Brand building, Student retention, Financial Management and more.

* Qualified and passionate instructors- highly qualified and passionate instructors are required to give quality education to students to inspire and motivate students.
* Diverse Learning Program- it gives the opportunity to offer programs related to student’s interest and learning goals.
* Marketing and Brand building- promoting the brand through social media, online search engines and more to the target audience provides better business opportunities.
* Student Retention- student retention is one of the important parts of business success, because positive feedback will attract new students to join.
* Financial Management- Managing and spending the right amount of money on budgeting, salaries, facility costs, marketing and more is important for business profitability.

**SWOT Analysis**

* Strengths- Having highly skilled trainers and providing quality learning is one of the most important strengths companies must have to be successful. Other than that, providing a diverse range of courses.
* Weakness- The company may face difficulties in scaling the business and one of the reasons could be, students require a long-term commitment learning music and they might end up dropping out from the course which could result in inconsistent student retention.
* Opportunities- Collaborating with schools and colleges, organizing various workshops and events for new registrations and can open opportunities for revenue and could help in increasing scalability.
* Threats- shift in consumer preferences could be one of the threats that a company could face. Also, an increase in competition could result in a decrease in student retention. So, it is important to focus on the ways to provide quality education at a competitive price with diverse options.

**Stakeholders Preferences**

Stakeholders are likely to prefer a financially stable music school, with consistent revenue and high returns. They want great education from experienced instructors, cutting-edge facilities, and programs that attract a diverse student body. Other top considerations include maintaining a positive reputation and reaching out to local or global audiences. That interest may be fuelled by accessibility, inventive breakthroughs, and cultural influence, all of which increase their desire to support the school.

# **Business Questions**

In this Project, we will be answering three business questions and will try to help ABC Academy of Music in improving their profitability and marketing strategies by using the insights derived from our analysis.

1. Which is the most profitable revenue category generating the highest recurring sales?
2. What is the churn rate of the company?
3. A) On which day of the week, are the marketing emails most likely to be responded to?

B) Is sending multiple emails to a single customer generating a higher open rate?

# **Methodology**

**Q1) Which is the most profitable revenue category generating the highest recurring sales?**

**Data Preprocessing**

Data preprocessing steps are performed to ensure the data's quality and suitability. The advantage of this is that it ensures that the data required is suitable for the question, increases effectiveness, and reveals necessary insights into the data [[1]](#_References). This involved imputing suitable values for missing datapoints, differentiating pending payments from refunds, dealing with the multiple revenue categories, and identifying duplicate transaction IDs.  Also, data transformation was carried out where variables were transformed using standardization to ensure they all had one scale measurement.

**Data Cleaning**

* **Negative Total Sales:** Rows with negative total sales were examined. Both, pending payments and refunds, are shown with negative values. Therefore, to differentiate between these two types of transactions, an assumption is taken into consideration:
  + If the value in ‘qty’ column, that signifies the quantity or number of items bought by the customer is zero (0), then those transactions were marked as ‘Refunds’.
  + But if the value of ‘qty’ is not zero and the ‘total sales’ value is negative, then it is categorized as pending payments.
* **Multiple Revenue Categories:** Dataset was converted into long form for better analysis of the categories. Hence, rows with multiple revenue categories were split into separate rows, and the corresponding quantity and total sales were adjusted accordingly.

**Data Analysis - Categorical Analysis**

* The dataset was grouped by the revenue\_category variable to calculate the total quantity and total sales for each category.
* The grouped data was visualized using appropriate charts to understand the distribution of quantity and sales across different revenue categories.

**Q2) What is the churn rate of the company?**

The purpose of this question is to find out the churn rate of the business and develop a predictive model to forecast the future churn rate from the given data.

To solve this question, the membership data was used. The data consists of 2091 instances and 16 variables.

**Data preprocessing**

Data preprocessing is an integral part of the data analysis process. [[1]](#_References)

* First, unnecessary columns that were not relevant for the analysis were removed from the dataset. The columns that were excluded to reduce the dimensionality of the data were as follows- last renewal date, cancellation reason, cancellation note, termination date, conversion type, converted to and hold period.
* The columns’ names were converted to snake case naming convention to ensure easier readability and uniformity in the code.
* The dataset was checked for duplicates, and it was found that 19 rows are duplicates so to maintain quality and accuracy of the result, those rows were removed.
* Missing values were checked, there were in total 1584 missing values. On column wise inspection- start date had 7 missing values and cancellation date had 1577 missing values.

To address these missing values some assumptions were considered:

1. For missing start date- It is assumed that the start date is same as the purchase date.
2. For missing cancellation date- The values of missing cancellation date were assigned based on their status.
   * If the status of the observation of the missing cancellation date is ‘expired’ then the end date of the membership is considered its cancellation date.
   * In all other cases, i.e. if the status is active, active/on hold, not activated, the cancellation date is assumed to be not available (NA).

**Feature Engineering**

It is the process of creating or modifying the features to enhance the performance of the model. [[2]](#_References)

* **Membership length:** By using the dyplr’s mutate function, membership tenure was captured in this new variable. The instances in which membership is not available (NA), i.e. observations with the status as active, active/on hold and not activated, the difference between end date and start date in days is taken as its membership length.

While for the observations with ‘expired’ status, the difference between cancellation date and start date in days is taken as its membership length.

After inspection, it was noticed that a few of the instances of membership length were negative. And after viewing these instances, it was seen that they are negative because the cancellation date was before the start date which means that the customer cancelled the membership before the start date that is even before he/she started taking the classes. So, the membership length for such customers is 0. Therefore, we changed all the negative instances of membership length to 0.

**Machine learning method**

For the analysis and forecasting of the time series data related to churn rate, AutoRegressive Integrated Moving Average (ARIMA) and ETS models were applied.

1. **AutoRegressive Integrated Moving Average (ARIMA) Model**

ARIMA is a popular statistical modeling technique used for forecasting of time series data based on past and current data. It can be subdivided into 3 components [[3]](#_References):

* AR (Auto Regressive) – It examines the past and current values to forecast the next upcoming values. So, after the calculation of the churn rate, it was used to further determine the future churn rate.
* I (Integrated) – In this step helps to make the data more balanced by removing the any trends or patterns that might exist and make the data stationary as the predictions might be impacted by these patterns.
* MA (Moving Average) – this part of the ARIMA focuses on balancing the residual errors or noise in the data and it also models the dependency between the variable and the noise from a moving average.

A general structure to represent an ARIMA model is ARIMA (p, d, q), where [[3]](#_References):

* p is the number of autoregressive terms
* d is the number of differences
* q is the number of moving averages

To automate the model selection process, auto.arima () function is used from ‘forecast’ package in R. It utilizes Akaike Information Criterion (AIC) value of ARIMA model to compare various ARIMA models and select the best one out of them.

1. **Exponential Smoothing State Space (ETS) Model**

In addition to the ARIMA model, the ETS model is also applied to forecast the churn rates. ETS is another popular time series forecasting model widely used in data analysis. It is built on the principles of exponential smoothing, in which the forecast is adjusted by placing more weight on the recent observations. ETS has 3 components, namely Error, Trend, Seasonality which are described below [[3]](#_References).

* Error – It means the random fluctuations that don’t follow a particular trend or pattern in the data. Overall, ETS helps in handling such unpredictable changes.
* Trend – It overlooks any upward or downward movement in the data over a period.
* Seasonality – This part of ETS captures any specific repetition in pattern over time such as daily, monthly or yearly changes.

For the analysis, ets () function from the ‘forecast’ package in R is utilized to automate the process of optimal ETS model selection. It help in evaluating the best configurations of error, trend and seasonality component of the model based on the data characteristics.

**Exploratory Data Analysis (EDA)**

EDA is performed on the dataset to understand its overall structure and distribution of the data [[4]](#_References). Here, R is used to perform the exploratory data analysis part.

A graph with a bar chart

Description automatically generated with medium confidence

Fig.1 Bar Plot of Status

The above bar chart shows the distribution of the membership data categorized in terms of its status. It can be noted that most of the members have ‘expired’ membership, with 1467 instances. In addition to that, approx. 500 members have actively ‘cancelled’ their membership. It indicated that a large proportion of the memberships are not renewed resulting in higher churn rates. Against these negative numbers, the academy has 78 ‘active’ members that are and 5 members that are ‘on hold’ status. While there are around 30 customers that have bought memberships for future music classes.

Histogram and Boxplot were used to define the distributions and to check the outliers in the dataset. According to Investopedia, “histogram is a graphical representation of data points organized into user- specified ranges”. The range of classes were represented on the x-axis and number of count or percentage of occurrences in data for each variable is represented on the y-axis [[5]](#_References).

Whereas a boxplot is a type of graph which presents a visual of how dataset is spread into 25th percentile, 50th percentile and 75th percentile and the comparison between them and how maximum, minimum and outlier values are spread. It is drawn as a box with a line inside it which is further extended from each side known as whiskers. The box represents the interquartile range (IQR) that is the difference between the third and the first quartile or the amount spread in the middle 50% of a dataset. The whiskers indicate the variation in maximum, minimum and any outlier data points compared to the interquartile range. The longer whisker means more variability of the data points to the IQR and vice versa [[6]](#_References).

A graph of a person with a bar graph

Description automatically generated with medium confidence

Fig. 2 Histogram of Membership Length

The above histogram portrays the distribution of membership length (in days) by plotting it as number of occurrences in various ranges of 30 days. It reveals a positively skewed distribution, with a large concentration of members having shorter memberships. The most common membership length falls between 60 - 90 days. Apart from this, another peak can be noticed in the histogram around 20 to 300 days mark which suggests that there is a small subset of members with extended membership.

A graph with a blue line and red dots

Description automatically generated

Fig. 3 Box Plot of Membership Length

The above boxplot also illustrates the distribution of membership lengths in days with other statistical insights. The median of the membership length is 91 days, which represents the central tendency of the data. The box plot spans over Q1 (83 days) to Q3 (180 days), with the interquartile range (IQR) of 97 days. This suggests that 50% of the data lies between these two quartiles, having a moderate spread of around 3 months. Additionally, several points beyond the upper whisker can be observed, indicating potential outliers. It points out that there are several members that extend their membership beyond the typical range, reaching over 500 days.

**Q3)**

1. **On which day of the week, are the marketing emails most likely to be responded to?**

The purpose of this question is to find the most appropriate day of the week to send marketing emails to customers based on the open rate. To solve this problem, the anonymized marketing emails dataset was used.

**Data Cleaning**

Before using the dataset, we used SQL for data cleaning and

* All the column names were converted to snake case.
* Column ‘client\_groups’ was removed from the dataset as it was not necessary and had null values.

**Data Preparation**

We loaded the dataset in R and checked the string type of each feature (column). Since the ‘send\_date’ column was in character type format, we converted it into date type.

Feature Engineering

The current dataset does not have a column showcasing the ‘day’ an email was sent. To solve the question, we created a new column called “day\_of\_week’. This column extracts days from the date mentioned in the data. We created a column of ‘open\_rate’ using the formula (opened / (sent – failed)) \* 100. Using this new column along with the open rate column, we calculated the average open rate for each day.

1. **Is sending multiple emails to a single customer generating a higher open rate?**

By answering this question, we will understand whether sending multiple emails to a single customer generates a higher open rate or not. We have used the anonymized emails dataset for this question.

**Data cleaning**

Using SQL, we converted all the column names of the dataset to snake case.

**Data Preparation**

After loading the dataset in R, we performed a few steps on the dataset.

* On checking the string type of each column, it was discovered that the ‘creation\_day’ column was a character type rather than date. Therefore, we converted it to date.
* There were certain records as ‘failed’ in the ‘status’ column. With the assumption that these are the emails that were not delivered to the customer, we have chosen to remove those records.
* We have converted the status column into binary format for the convenience of our analysis.
* The emails were categorized into single and multiple emails based on the number of emails sent to a single customer.

**Feature Engineering**

Like the previous question, this dataset also didn’t have a column dedicated to showing the day of the email been sent to a customer. So, we created a new column ‘day’.

**Exploratory Data Analysis**

Bar graph and pie chart were created to visualize the average open rate and open rate of multiple emails.

# **Results and Discussions**

**Q1) Which is the most profitable revenue category generating the highest recurring sales?**

A graph of a bar

Description automatically generated

Fig. 4 Total Quantity by Revenue Category

The above bar plot portrays the number of registrations each music lesson category got within the timeline of the sales dataset. For the interpretation of the above and other plots for question 1, we will be ignoring the ‘Uncategorized’ revenue category as it doesn’t distinguish the reason of all the transactions recorded under it. Also, it contains the records of all the payments received by the ABC Music Academy, or the refunds sanctioned to its customers when the academy’s systems were down, and manual entry of these transactions were made without mentioning the category it belongs to.

The plot depicts a great popularity of ‘Music Together – Mixed School’ among the customers compared to other categories, with ~ 550 memberships. ‘M30 – Music Lessons’ and ‘M45 - Music Lessons’ follows 2nd and 3rd most popular categories, with approximately 240 and 130 memberships sold respectively.

A graph of sales

Description automatically generated

Fig. 5 Total Sales by Revenue Category

In the above bar graph, the revenue categories are organized in terms of decreasing total sales value. It can be noted that although ‘Music Together – In-School’ is not a very popular category, with only 50 instances in the dataset, it generates the largest revenue for the academy. It leads the list with around 45,000 CAD in total sales.

Due to its high popularity, ‘Music Together – Mixed Age’ comes as the runner up category to generate large amounts of sales, with total sales value reaching ~ 31,000 CAD. Music lessons: M45, M30 and M60 are 3rd, 4th and 5th most revenue generating categories respectively.

A graph of sales per membership

Description automatically generated

Fig. 6 Average Sales per Membership by Revenue Category

The above bar graph depicts the average revenue the academy generates with each membership for each type of revenue category. Although ‘Music Together – In-School’ has only 50 memberships in the sales data timeline, it generates most revenue per membership, with around $900 per membership. ‘Group Classes’ follows next with approximately $600 per membership. While ‘Summer – M60’ generates around $150 per membership, securing 3rd position in this list.

A screenshot of a data sheet

Description automatically generated

Fig. 7 Summary of Q1 graphs

The above visual records the values of the 3 graphs for question 1 in a tabular form. It is arranged according to the descending total sales value. It summarizes the dataset by categorizing it by different revenue categories to calculate total memberships sold (in total\_qty column), total revenue generated (in total\_sales column) and average revenue generated per membership (in avg\_sales\_per\_membership column).

**Q2) What is the churn rate of the company?**

Using R, a new function is defined to categorize data based on different years, i.e, 2016-2024.

A screenshot of a data

Description automatically generated

Fig. 8 Snippet of new table

Through this categorized data, a new table is created containing columns:

* Total\_Students: The total number of students in a given year is recorded under this column. The basis to calculate total students in a given year is the number of students that registered in the academy in that given year and the number of students that continued from the preceding years.
* No\_of\_Churned\_Students: It is the number of students that churned from the academy i.e. left the academy in the given year.
* Churn\_rate: The churn rate of a given year is calculated by the number of churned students by the total number of students in that given year.

A graph with red and blue lines

Description automatically generated

Fig. 9 Total vs. Churned students over the years

The line plot provides a visualization of total students and churned students from 2016-2024. With the number of students on the y axis and year on the x axis. It presents the trend of students registering in and leaving the academy. It is observed that the academy experienced a rise in the number of students joining the academy till the year 2019, after which a gradual dip in the total registered students can be seen till the year 2021. The academy saw a spike in the number of students joining the academy in the subsequent years, where the total number of students peaked in 2023 with around 700 students. The number of students leaving the academy follows a similar pattern as the number of registered students. The rate of sudden decline in the years 2020 to 2021 can be assumed to be the result of the COVID pandemic.

A graph with a line going up

Description automatically generated

Fig. 10 Yearly Churn Rate (2017-2023)

The visualization demonstrates the churn rate from 2017 to 2023. A sharp uptrend in the churn rate can be seen from 2017 to 2020. It dipped in the following year, 2021, which can be the effect of the improving situation of the pandemic. Later, the churn rate continued to move in upward direction and peaked in 2023 i.e. ~ 82%.

A graph with a line and a line

Description automatically generated with medium confidence

Fig. 11 ARIMA Forecasted and Actual Churn Rates (2017-2026)

* ARIMA and ETS models are used to forecast the values of churn rate for 3 years, 2024 to 2026. Based on these models the graphs were plotted using the autoplot () function. The above graph shows the forecast and the actual churn rate from 2017-2026 based on the ARIMA model. The black line represents the actual churn rates across various years. While the red line represents the fitted values according to the ARIMA model. On the other hand, the blue line shows the forecasted churn rate for the upcoming years, i.e. 2024 to 2026. Based on the model prediction, the churn rate for the following 3 years will be ~ 73%.
* The dark blue band depicts the 80% confidence interval while the light blue band portrays the 95% confidence interval.
* According to Hayes, A. (n.d.), a confidence interval refers to the probability of a population parameter (such as mean) lies within a range of values. Overall confidence intervals reflect the certainty of an estimate.

A graph with a line and a line

Description automatically generated with medium confidence

Figure 9 ETS Forecasted and Actual Churn Rate (2017-2026)

The graph portrays the forecasted values based on the ETS model. It provides almost similar result to that of ARIMA model, where the forecasted values of churn rate by ETS is also around 73%.

**Q3)**

1. **On which day of the week, are the marketing emails most likely to be responded to?**

After data cleaning, feature engineering in which a new column called day\_of\_week and open\_rate were created using the (opened / (sent – failed)) \* 100 formula.

A graph of blue bars

Description automatically generated

Fig. 12 Average open rate by Day of the Week

The processed data was visualized through a bar plot that shows average open rate by day of the week. It was created using ggplot2 library, day of the week on the X axis while average open rate on Y axis. According to the visualization, Sunday is the best day during the week to send out marketing emails to the customers because the opening rate is the highest on that day, pointing to the fact that people are relatively free during the weekend. The second most favorable day is Monday.

The table below specifies the day of the week in the order of the most favorable day to least favorable day based on the open rate.

|  |
| --- |
| **Day of the week** |
| Sunday |
| Monday |
| Friday |
| Thursday |
| Saturday |
| Wednesday |
| Tuesday |

1. **Is sending multiple emails to a single customer generating a higher open rate?**

A blue pie chart with a number of emails

Description automatically generated

Fig. 13 Open Rate: Single vs. Multiple Emails

Based on our analysis, we reviewed whether sending a single email or multiple emails to customers had an impact on the likelihood of the email being opened. This was done to interpret the effectiveness of follow-up emails in improving customer engagement.

Once data cleaning was done, we used cleaned data to find out what type of mails are more likely to open and visualized through pie chart to provide a clear view. As seen in the pie chart, customers who received many emails had a slightly higher open rate (53.4%) than those who received only one email (46.6%). Frequent email communication with customers increases the likelihood of email opening, as repeated messages from the same sender can prompt higher engagement.

# **Recommendations**

1. **Community-building methods:** Organizing various events, workshops, jam sessions to create a sense of belonginess among students. Also, adding success stories, event clips, and performances of students on social media to improve reach and impact.
2. **Performance Opportunities:** Organizing competitions and general events where students can participate and showcase their talents to motivate other students and give them a milestone to look up to.
3. **Reward achievements:** Offering discounts or exclusive lessons for students who continue for multiple terms, which can help reduce the retention rates.
4. **Loyalty programs:** Celebrating student achievements, such as completing levels or participating in performance could help build healthy relationships and can motivate others.
5. **Targeted discounts to boost sales and registration:** A limited time discount on certain lower revenue-contributing services could be given to new customers to increase sales of that music lesson and improve overall business profits. Other than that, a discount on the 1st month could be provided to new people to increase student registration.
6. **Optimized email timing for higher engagement:**As per the analysis, it is recommended to send email on Sunday as on that day people tend to open their emails more as compared to the rest of the week.

# **Conclusion**

This report provides a key highlight to enhance the performance and revenue of ABC Academy of music and provides the scope of improvement in customer retention of the company. Through this analysis we cater to 3 business questions to get the result. Firstly, **Music Together- In School, Music Together – Mixed Age, and M45 – Music Lessons** are the 3 most categories contributing highest to revenue. Similarly, M45 – Music Lessons and M30 – Music Lessons are the top categories being sold. However, by using the data provided we found that the company’s churn rate is **82%** in 2023 and the predicted churn rate from 2024 to 2026 could be **73%**. By tackling the churn rate and focusing on the most demanded products of the company could help achieve sustainable growth and enhance profitability of the company and could maintain a competitive position in the market.

Other than that, by analyzing the emails data we found that sending mails on **Sunday, Monday and Friday** provides the highest chances of mails being opened. Also, we concluded that sending multiple emails has a higher engagement than sending single mail to the customer. With this help, the company could improve their customer engagement and could maintain a good relationship with the customer. Overall, by tackling the churn rate, focusing on higher contributing products and constructing enhanced marketing strategies could improve overall profitability and revenue which will ultimately improve the company’s growth.

# **References**

1. *Alasadi, S. A., & Bhaya, W. S. (2017).* [*Review of data preprocessing techniques in data mining. Journal of Engineering and Applied Sciences*](https://d1wqtxts1xzle7.cloudfront.net/54509277/4102-4107-libre.pdf?1506113528=&response-content-disposition=inline%3B+filename%3DReview_of_Data_Preprocessing_Techniques.pdf&Expires=1732997865&Signature=gpNMBRTPLPgRsuHJwpqldrIkQvDy~3-bG0iFq8TInqdB30ot5ZAYkPCxOD2VYO60G6Crj7MnZHoNpI31c1Klu7EPfmC1K5d-3SsU7tGl5ZvBz1tAjft9QS4xqXqTj8cC7MFaTo-wNKP5eR3k~OhUlZRrZfR-hU8C-yM0jTOR3B0-dMX1JtAt-Y7TdEl0qYlnDMqFH3ikCJde~NS3w0fwPczMpPPq1pa5~ydkwIFyF-VW8SZXwai-48NUU-RVxzJDxeEXvizsLnW0I8kMCo~EFXRHrO4dV5DDFWeYl-4ZkRaqVy9f2TSHgqUlI1es-581crgt9RO8WUn6~KVd99OsvA__&Key-Pair-Id=APKAJLOHF5GGSLRBV4ZA)*, 12(16), 4102-4107.*
2. *What is feature engineering?. Domino Data Lab. (n.d.).* [*https://domino.ai/data-science-dictionary/feature-engineering*](https://domino.ai/data-science-dictionary/feature-engineering)
3. Hyndman, R. J., & Athanasopoulos, G. (2021). [*Forecasting: Principles and practice*.](https://otexts.com/fpp2/) OTexts.
4. Ibm. (2024, November 7). *What is exploratory data analysis?*. IBM. [https://www.ibm.com/topics/exploratory-data-analysis](https://www.ibm.com/topics/exploratory-data-analysis%20)
5. Chen, J. (n.d.). *How a histogram works to display data*. Investopedia. <https://www.investopedia.com/terms/h/histogram.asp>
6. *Atlassian. (n.d.). A complete guide to box plots.* [*https://www.atlassian.com/data/charts/box-plot-complete-guide*](https://www.atlassian.com/data/charts/box-plot-complete-guide%20)