Contents

1. Introduction

As a popular sports team, Sydney Swans may need to know their fans' loyalty and how they can improve their business performance. This report analyses the result of a questionnaire to investigate the factors that influence their fans' love and the impact on the business performance. After processing the questionnaire data, Logit Model, Random Forest, Gradient Boosting and SVM are used to analyse which factors have a great effect or importance. Specifically, how fans' demographics and the love for the team influence their decision to stay and how Swans can improve fans' loyalty. Finally, recommendations are provided which may help Swans to improve its business performance and develop sustainability in the sports industry.

2. Data Pre-processing

Before data modelling, it's important to understand the data to identify if any pre-processing is needed. Firstly, the dummy values in q1.7 are summed up to get the total game attendance last year. Secondly, since q6 and q8 provide the importance of the criteria to assess Swans performance and organization and q7 and q9 provide fans' rating towards them, it is reasonable to combine them to get the weighted rating for data modelling. Besides, some demographical variables such as age, gender and income could be used to investigate what characteristics loyal fans have. Additionally, the average is used to fill in null values.

3. Analysis

3.1 Fans profile analysis

This section utilises Exploratory Data Analysis (EDA) to provide an overview of the characteristics of the fans that will stay compared to those that will not stay. The target variable is determined as "Membership renewal is automatic" (q14.6), the score >= 6 is categorised as "Stay", while the score < 6 is categorised as "Not Stay". According to the graphs in the Appendix, the members that will stay are more supportive and prefer to attend games with their friends or family. Besides, most of them are older aged with lower income, while members aged 25-34 tend not to stay.

3.2 Which hypothetical situations can influence fans' decision to stay?

This section aims to investigate how Swans can improve the probability of fans staying. q15 provides the ratings of ten hypothetical situations that would influence the likelihood that fans are willing to remain in the Swans membership. These ten hypotheses are determined to be independent variables and fans' "stay" (q14.6) is the dependent variable.

	Error rate	Sensitivity	Specificity	Precision	AUC	Cross-Entropy
Logit Model	0.242604	0.793814	0.708333	0.785714	0.833333	0.502818
Random Forests	0.242604	0.773196	0.736111	0.797872	0.839347	0.494608
Gradient Boosting	0.260355	0.721649	0.763889	0.804598	0.837772	0.501509
SVM	0.236686	0.804124	0.708333	0.787879	0.827320	0.000000

Figure 1 Model Evaluation

According to Figure 1, the SVM Model is selected because of the lowest error rate. Besides, the top three important hypotheses are "I would continue to be a member if the Swans came last on the ladder for one or more seasons" (q15.3), "I would continue to be a member if membership fees rose by 10%" (q15.5) and "I would continue to be a member if membership fees rose by 30%" (q15.2). Based on the result in the Logit Model, these three hypotheses are positively related to the target variable, and the result means that some of the fans are very loyal and they would stay in the club even if the membership fees increase or Swans come last on the ladder.

These three hypotheses show that the increased membership fees and backward ranking do not impact the loyal members' decision to stay. This indicates that fans' loyalty is important for the Swans to improve performance. Therefore, Swans should focus on retaining their loyal fans and improving fans' loyalty.

3.3 How does fans' love influence their decision to stay?

Fans are affected emotionally by a team's success but winning or losing does not represent the total picture of performance for loyal fans (Shuv-Ami, Vrontis & Thrassou, 2018). This section aims to examine the relationship between fans' loyalty and their decision to stay. 37 variables including team support (q15_team_sup), whether attend games last year (q18_game_last), game attendance (attendance_count), level of support (q3) and other questions about fans'

loyalty (q2, q5, q13 and q14) are selected as independent variables and fans' "stay" (q14.6) is the dependent variable.

	Error rate	Sensitivity	Specificity	Precision	AUC	Cross-Entropy
Logit Model	0.248521	0.793814	0.694444	0.777778	0.794387	0.592236
Random Forests	0.242604	0.793814	0.708333	0.785714	0.821163	0.508291
Gradient Boosting	0.207101	0.793814	0.791667	0.836957	0.821377	0.513455
SVM	0.284024	0.793814	0.611111	0.733333	0.793958	0.000000

Figure 2 Model Evaluation

Although the Logit Model does not have the lowest error rate, it is still chosen for analysis because its coefficients could help interpret the positive or negative effects of each variable. After fitting the Logit Model, 9 out of 37 variables are statistically significant with a p-value less than 0.05. "I tell others that I am a fan of the Swans" (q5.3) has the smallest negative coefficient, which means the more likely fans tell others they support Swans, the more likely they would not continue their membership. "Swans management behaves in a trustworthy manner towards me" (q13.4) has the largest positive coefficient, which means the trustworthiness of Swan's management has the greatest positive effect on fans' decision to stay among all the factors of fans' loyalty. This variable is used as a dependent variable in the section below to investigate how Swan's organization can achieve a trustworthy manner to influence fans to stay.

3.4 What can Swans organization do to achieve a trustworthy manner towards its fans?

This question aims to find out what factors behind the management of Swans may affect the trust of its fans. In the last section, "Swans management behaves in a trustworthy manner towards me" (q13.4) is the most important variable that could impact the fans' decision to stay, so further analysis of what can impact this variable is necessary to help the Swans retain more loyal fans. Ten sub-questions in q9 are selected as independent variables and the dependent variable is q13.4 with the score >= 6 as "Trustworthy" and the score < 6 as "Not Trustworthy".

	Error rate	Sensitivity	Specificity	Precision	AUC	Cross-Entropy
Logit Model	0.171598	0.595238	0.905512	0.675676	0.846457	0.419866
Random Forests	0.165680	0.452381	0.960630	0.791667	0.844957	0.419386
Gradient Boosting	0.183432	0.547619	0.905512	0.657143	0.826397	0.412651
SVM	0.177515	0.666667	0.874016	0.636364	0.847394	0.000000

Figure 3 Model Evaluation

According to Figure 3, the Random Forest Model is selected as the final model because its error rate is the lowest and the accuracy of all positive predictions is the highest compared with other models. Besides, the result of feature importance shows that the top three important features are "Employ courteous and professional staff" (q9.4.3), "Professionally market and manage the club" (q9.4.1) and "Keep me up to date on club matters" (q9.3.2).

Thus, three recommendations could be made to transform the club's trustworthy manner toward fans. Firstly, Swans could concentrate more on employment and staff training. In the recruiting process, Swans may observe if the candidates' manner is polite or considerate, and place new employees on short probation. Also, Swans could invest in personnel training through the employment of professional instructors and develop regular training programs for all employees.

Secondly, Swans could cooperate with a professional marketing organization to improve the marketing performance or invest in digital marketing platforms like Google Ads and Facebook to attract the targeted fans. Also, it could set strict rules and long-term development goals to effectively manage the club.

Thirdly, Swans could manage their social media accounts on several popular platforms to share the latest matters and information about the club. Besides, Swans may assign their trained personnel on these platforms to timely communicate and interact with followers. Also, ensuring Swans official website provide up-to-date and accurate information.

4. Conclusion and Limitation

To sum up, based on EDA and model analysis, Swans' fans that would stay are older aged with lower income and usually attend games with friends or families. It is recommended Swans

could launch a new CRM program for young fans since they have a low willingness to stay. Furthermore, model analysis shows that the Swans' fans are loyal and not price-sensitive, and "Trustworthy" is the most important factor for Swans management that can impact fans' loyalty. Therefore, Swans should focus on improving management and performance. To specify, Swans could provide professional training sessions to their employees to emphasise their professional image. Besides, to highlight their professional online presence and to better provide updates to the fans, Swans could invest in digital marketing platforms to engage with their fans on social media and regularly update their website with news and stories.

There are two potential limitations to this research. Firstly, this survey may be prone to untrue answers. Although the questionnaire has incorporated negative questions to avoid the halo effect, the survey may be long and time-consuming. This may undermine the willingness and trueness of the answer due to privacy issue or time limitation. Secondly, we are unaware of how the survey data is collected. Without verifying this, it is hard to say whether the sample is representative of the population. It could also affect the effectiveness of our model analysis and recommendations.

5. Reference

Shuv-Ami, A, Vrontis, D, & Thrassou, A 2018, 'Brand Lovemarks Scale of Sport Fans'
    Journal of Promotion Management, vol. 24, no. 2, pp. 215–232, doi:
    10.1080/10496491.2017.1360824.

6. Appendix

What are the characteristics of the fans that will stay?

This section utilises EDA to provide an overview of the characteristics of the fans that will stay (q14.6 >= 6) compared to those that will not stay (q14.6 < 6).

6.1 Counts of games attendance

Figure 4 shows there's not much difference for fans that would "Stay" and "Not Stay" for those who attended less than nine games last year. However, fans are more likely to "Stay" if they attended more than nine games last year. The number of fans that would "Stay" is doubled comparing those who would "Not Stay" if they attended all eleven games last year.
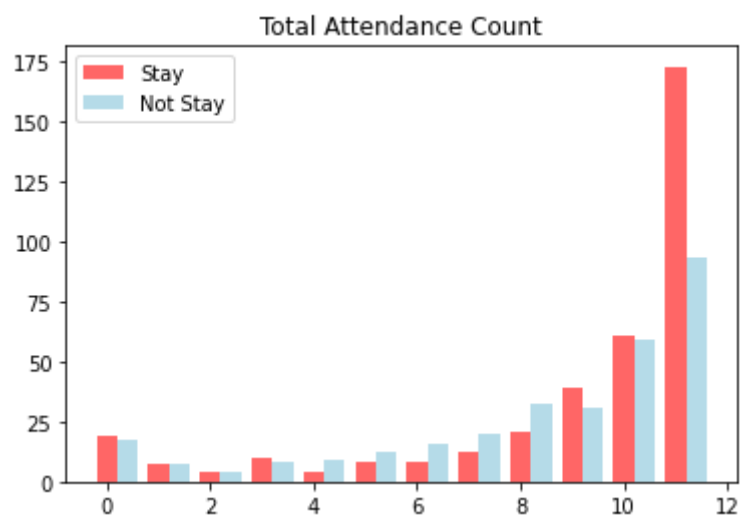

Figure 4 Total Attendance Count

6.2 Whom do they attend games with?

According to Figure 5, no matter whom fans attend games with, the number of fans that would "Stay" is greater than those would "Not Stay". And the most common way to attend games is with family or with friends.
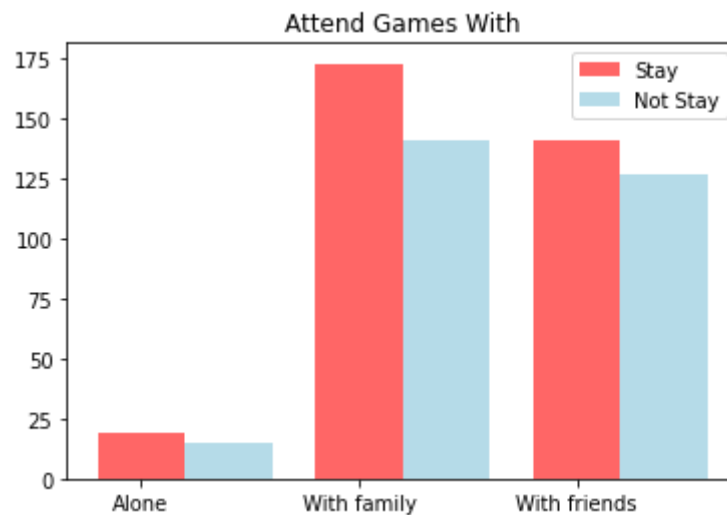
Figure 5 Attend Games With

6.3 What is their age?

According to Figure 6, across all age groups, more fans would "Stay" than "Not Stay", except for those aged 25-34. This indicates Swans may need to further identify the needs of this age group to ensure they would stay.
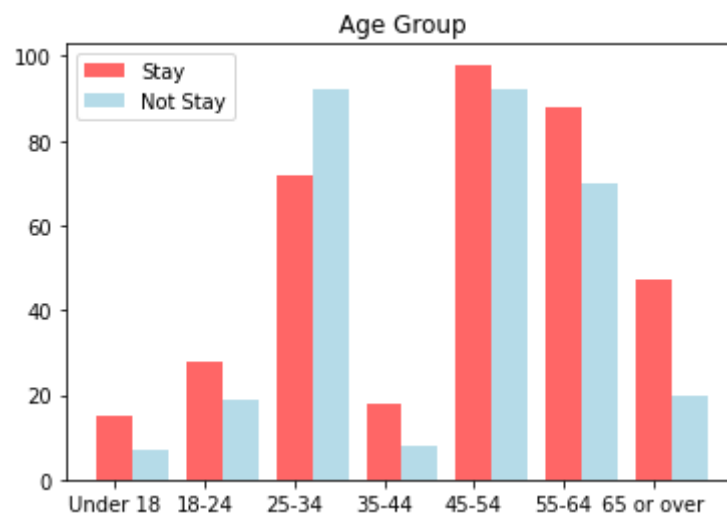

Figure 6 Age Group

6.4 What is their income?

According to Figure 7, interestingly, more fans would "Stay" in lower-income groups, while there are more fans that would "Not Stay" in higher-income groups. This is contrary to our expectation since we may assume that the higher income would be more willing to pay for the membership.

7

Figure 7 Income Group