Executive Summary:

When a customer ceases to use the service from a business or ends their relationship with a company, the customer is known to have churned. Many businesses also treat a customer to have churned if they see some level of inactivity over a continued period of time. The loss due to churn is significant as it costs the business the direct revenue that was associated with the customer abandoning the business as well as the marketing cost to acquire the new ones. Hence the ability to identify the customers at churn risk, while there is still time to do something holds a huge potential for the company to curb the loss. Hence through this report, we are going to analyse the customer behaviour and predict churn in advance and derive insights for a targetted marketing approach that will help to retain the customers at churn risk.

Generally, the challenges of performing such analysis include relying on modelling techniques that are based on static data, that can help with limited information. Also, churn prediction techniques aim to understand precise customer behaviour and hence accuracy and precision matters, finding the right balance between the scores can help in proactively retaining the customers. Churn analysis for non-recurring business can be difficult as the churn definition is not proactively known. This report effectively covers ways of dealing with challenges like these such as, the customer information is broken down into temporal features using windowing function, which is comparatively more insightful as the customer’s behaviour over time is more dynamic. The scores like Accuracy, Precision and Recall are selected keeping the business requirement in mind. Furthermore, the churn definition for the business is defined and a logical parameter for the same is picked for the analysis, which makes it lesser challenging comparatively.

In order to leverage this analysis profitability, one needs to extract meaningful insights to implement as a part of their marketing strategy and hence this report also covers feature importance techniques, that illustrates which feature holds more importance over the other. This can help us understand the ultimate factors leading to customers retention/ churn over time.

The goal of this report is to

* Prevent revenue loss, by training the model with the best accuracy, sensitivity and precision
* Reduce marketing and sales costs, by precisely predicting the churners and targeting them better
* Extract meaningful insights for a good marketing campaign to reduce churning customers

Current levels of churn:

By referring to the individual report, the following can be concluded:

1. From figure 1, it can be concluded that beyond 55 days 75% of customers return for their consecutive visit and if the inactivity persists, the graph can be seen flattening. This clearly indicates that beyond this point, a consistent percentage of customers that return.
2. From figure 2, it is clear that from 75-80 days of inactivity, the number of customers that return generally don’t increase. The graphs flatten and reach a stagnation point

Hence from the above figures, it can be concluded that it is better to avoid targeting a huge percentage of people and to save up on marketing costs, to do so one must target lesser number of days for defining churn, and hence for this report, we’ll select 27 days as the churn definition. As this has a lesser magnitude of people to target when compared to 50 days. Also with 59.7% of customers returning for their consecutive visit makes ideal for studying the customer transactional data.

Technical Report:

The approach for this analysis is as follows:

1. Prepare a temporal dataset comprising of features that describe the buying nature of the customers for analysing the behaviour pattern over time.
2. Break the temporal set of features into input and output features using output window size and tumbling window size.
3. Evaluate the classification models for the best accuracy, precision and recall.
4. Make predictions using the best performing model
5. Use feature importance techniques to find out which features affect the output the most
6. Deriving insights

Building temporal dataset:

For this report, the methodology opted for churn analysis was done using Supervised learning technique. The features in the dataset were broken down into temporal features to understand the behaviour of the customers over time. The features that were chosen for the evaluation were revenue generated by customers using ‘value’, the amount of quantity purchased in every visit using ‘qty’, and the frequency of visit of every customer using ‘freq’. These features change with time and hence would give a clear understanding of the customer’s buying behaviour at a particular instance of time. To analyse these features over time, Python is used to create windows in order to frame the problem in a way such that our models get a precise understanding of the happenings during that instance of time. For this purpose, we use windowing parameters such as Tumbling Window Size and Output window size. Tumbling Window Size is kept at 7. Since we want the predict the data for a week’s duration, our supervised learning algorithm would study the behaviour pattern of the customers over a week’s duration and predict the outcome depending on this pattern. Output Window size is at 27, which will indicate whether a customer ultimately churns within the duration of churn definition or not. For our dataset, we pick receipt ID to check whether or not the customer has visited the store. These input and output features in our dataset will be useful in training, testing and prediction. ‘Num\_periods’ decides the number of temporal features in the dataset, this is at 10, if it is assigned number lesser than this, there is a decrease in the accuracy observed and it also reduces the number of observations significantly as the temporal window is smaller in size when compared to the output window size. If it is more than 10, the accuracy remains unaffected and hence for efficiency and to avoid redundancy, the value 10 is assigned to the variable num\_periods. ‘Reference\_day’ variable indicates the day with the most recent data, and for the testing, it equals to most recent date(max(purchased\_at))-output window size.

While creating the dataset, a condition was given to select the customers only if they have visited the store more than once. This will eliminate one time visitors and would only analyse the behaviour of customers who have been more active. Another condition was given that would collect the data from all the stores from ‘2018-06-13’ onwards since store#3 starts collecting data from that period onwards. This would give us a consistent dataset across all the stores.

Model Evaluation:

We need to classify the customers into churners or non-churners and hence this qualifies to be a classification task. The Target class for the modelling purpose is Churners and the Non-Target class is non-churners, as we are focussing on targeting churners to provide them extra support and retain them. The Models that are chosen for the evaluation are Random Forest Classifier, Decision Tree and SVC.

Random forest is a good classification model, as there is no prescaling and standardization required, which is good for our dataset as it contains different variables across the timeline. This also provides Low Bias, Moderate Variance, which is apt for a good accuracy score. Its attribute n\_estimators is given the value 120 trees, as this provides the best accuracy and anything above this mark doesn’t improve the score.

SCV has a regularization metric-L2 by default, which ensures lesser overfitting. The kernel (RBF in this case) is very useful as this can take into consideration outliers and exceptions. The gamma for this is taken to be 1 because a higher value of gamma would tend to wiggle the decision boundary and may cause overfitting.

Decision Trees: The max\_depth chosen for decision tree is 20, for obtaining optimum results. The trees by default calculates Gini impurity of the clusters/ splits, which is important for the time-series data as they look at the time-series bisection with the highest purity clusters and pick the one that maximizes the separability of these clusters for better classification.

To be sure that we predict the churn of our active customers only, we put a condition in our model, wherein our test and training set only comprise of the customers who have been active in that particular time period by checking if the frequency of their visit is greater than 0 or not. This will ensure that we are predicting the potential churners from our active customer pool.

For our training models, we define variable ‘now’ to illustrate the most recent date, post which ideally the prediction of the churning customer is desired. For the model evaluation, we shift this date a little backwards, by subtracting output window size from ‘now’. The time after ‘now’ generates an output label and the time before ‘now’ generates an input label. This becomes our testing data for our machine learning model. Another variable ‘holdout\_set’ is defined to train the temporal data by the method known as repeated hold out. This variable is assigned the value 8, for obtaining the best accuracy, precision and recall.

To measure the performance of the training models accuracy, precision and recall are taken into consideration for each of the models. Accuracy is important as it describes the closeness of the prediction to the actual values. Precision illustrates positive instances among the retrieved instances and Recall illustrates relevant instances that were actually retrieved.

Since accuracy illustrates the fraction of predictions our model got right, hence it is important for our overall evaluation of the training models. Precision is a good measure to determine when the costs of False Positive is high, in FoodCorp’s business scenario false positive means that a non-churner (actual negative) has been identified as churner. The business might end up spending extra time and resources on the wrong customer and lose on a customer who is churning. The third score that is used to evaluate the models are Recall. It is an important matrix because it calculates how many of the Actual Positives our model captured as positives. For instance, if a churner (Actual Positive) is predicted as a non-churner, the consequence can be bad for the FoodCorp as it might target the wrong customer and end up losing on a customer who was about to churn. Post evaluation, Random Forest has the best accuracy (80.29%) and precision is (84.03%), and so we go ahead with the Random Forest model for our churn prediction.

Final Prediction:

For the purpose of prediction, we change the dates ‘now’ to the most recent date, in order to predict the churning customers for the next 27 days. After prediction, we get the values of churners and non-churners as 346 and 143 respectively.

Feature Importance:

For better analysis, we check the variable importance of the temporal features that we have used. By doing so, we can get valuable insights as to which features impacts the visit of the customers the most and potentially find out the reason why FoodCorp’s active customers are churning. In order to do so, we create a function to compare and print top 10 variables that are importances. Top 10 features would illustrate which feature amongst Value, Frequency of visits and quantity purchased is the most significant to determine whether a customer would come back or not.

After this, we proceed to remove the variable with low variance and to do so we set the variance threshold at 0.0, by doing so all the values with negative ranking is removed from the variable importance list. This provides a clear and competitive list of features that can be selected for deriving further actionable insights.

We use the same data frame that we made for the model evaluation. We break the temporal features into training and test set and use Random Forest to train and test. The trained model is then wrapped in a permutation importance object. Post this a .fit object is called to generate feature importance. We assign cv=prefit, to ensure that the model we want to check the importance for is already fitted.

Summary:

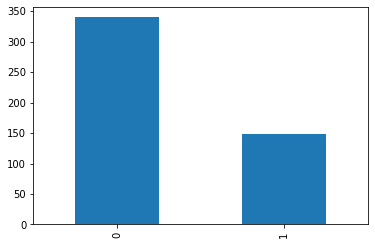
The final model chosen for prediction is Random forest classifier as the accuracy, precision and recall are favouring the business need of targetting the churners and non-churners effectively.

Insight Report:

It is observed that 346 active customers churn out of 489 active customers, and based on our variable importance analysis, it is established that the variables that impact the output the most are ‘value’ and ‘freq’. Through this insight when we look at the description of the final churners’ dataset, we get to know that the mean value spent ranges from £25 to £16, while the average spending of an active customer is in between £110 to £76. It is clear from this insight that people who tend to churn don’t prefer spending more money. It is further made clear by referring to the max spend of the customers. The active customers who tend to churn spend up to a max limit of £300-£200 whereas the customers to remain active spend up to a limit from £900-£700. Hence, when the spending behaviour of a customer has been observed to be shrinking, that can be the first indication of this customer being at a churn risk. To encourage such customers to spend more, FoodCorp must roll out offers that incentivise the customers and make them feel valued. This can effectively increase the value spend by the customer in the store per visit.

It is also observed that the mean frequency of visit of an active customer who is at churn risk is reduced to 1 time per week, whereas the frequency that of a customer who is retained visits approximately 7-5 times in a week. It is further made clear with the maximum and the minimum number of visits of both customers and churn risk and otherwise. The customer at churn risk visits the store 1 or sometimes doesn’t even show up in a week time, whereas the customer who is retained visits the store almost a maximum every day. Keeping in mind that FoodCorp is a grocery store that makes it an essential business, and hence it can be assumed that the frequency for something as essential and regularly consumed as groceries should be visited quite frequently. Hence, if the reduction in the frequency of visit is observed, a targeted marketing action must be taken to curb any chances of that customer becoming a churn risk to the company. To make this happen, FoodCorp can introduce weekly offers to attract customers on a regular basis. Regularly visiting the store might also indicate that the person’s average value contribution would also increase, which is an obvious benefit.

Finally, we observe the quantity that our customers shop, every time they visit the store. The active customer, who is at a churn risk buys 2-3 items on an average in a week, wheres a customer who is not at a churn risk buys 3-4 items on an average in a week. This is not a very significant number, which is similarly indicated in the variable importance analysis. FoodCorp can introduce a marketing campaign such as buy one get another free, to encourage the customers to shop in bulk. This would also lead to better customer relationship and retention. Bulk buying also leads to an increase in the revenue from the customers and hence bulk buying can lead a customer to spend more every time he/she visits the store.



From the above graph, it is clear that the customers who are at churn risk(0) are more in number than the customers who are ultimately retained(1). And hence focusing on retaining them would help FoodCorp save up on acquisition costs.

The following pen portraits is derived from the dataset above:

Average amount spent by churners till date= 234.1074025974027

Avgerage amount spent by non churners till date= 254.07635820895504

Avgerage number of visits by churners till date= 16.396103896103895

Avgerage number of visits by non churners till date= 15.605970149253732