Executive Summary:

The main aim of this report is to build on the past analysis which indicated that FoodCorp encounters the problem of customer churn. This report tends to use the data available by the company to classify the customers as churners or non-churners, so that preventive action could be taken to stop customers from switching elsewhere. In order to do this, supervised learning techniques were used. Building on the analysis provided by Consulting Corp, it was decided that a customer who does not visit the store(s) within 21 days will be classified as a potential churner. This was primarily done through analyzing the distribution of time between the visits of customers. The input features of the model were constructed based on the number of visits and the quantity bought by the active customers during specific time frames. After the essential steps of feature engineering and preprocessing, 4 different models were constructed. These were evaluated on the basis of several performance metrics and the results suggested that support vector machines classifier (SVM) tends to perform better than the rest of algorithms. Thus, SVM was chosen to be the final choice for training the model.

This report also talks about the distinct behavioral and spending patterns of the churners and the non-churners. The results reveal that churners are more likely to be infrequent, low spending customers, who just visit the stores once in a while to buy necessities. On the other hand, non-churners are consistent buyers who spend on a greater variety of items. They also spend more amount of money on average than the non-churners. The customer insights derived from this report can be used distinguish between churners and non-churners effectively. It can also serve as a basis for devising effective marketing strategy to keep the people engaged.

Churn Analysis:

In order to decide the churn definition, it is essential to understand that how frequently an average customer visits the store and what is the average time between the visits. To make the churn definition generalizable to the customer base, it must be applicable on at least 50% of the total customers. Furthermore, the analysis of proportion of people who actually churn within a specific time period will also be useful.

It was analyzed that if the inactivity period is chosen to be either 1 or 7 days, **only 2% - 26%** of the overall customer base is taken into account. This implies that more than half of the population will be left out and will be classified as churned if they don't visit every other day or within 7 days. This would waste considerable resources of the business in targeting all such people who do visit the stores more than once month but might not be fond of shopping very frequently. Thus, the option of choosing 1 or 7 days as churn period is eliminated.

The analysis also shows that on average, 75% of the customers visit the store within 36 days. **Fig 1** in Consulting Corp's report shows that the rate of change for customers arriving between 7 and 15 days is the steepest. There is almost **24.1% increase** in the customers with **additional 8 days**. On the other hand, the curve gets less steep when the number of days increase to 21 days. The customers increase by an **additional 9.95%** which equals to **60.42% of the total customer base**. If the window is chosen to be 36 days, the business will lose out more information rather than gaining insights. For example, if a customer initially visited 2 times a month and now visits only once a month, the change in buying pattern will not be captured by a churn window of 36 days. Moreover, **Fig 2** also shows that if the inactivity period is defined as 36 days, FoodCorp will expect to target only 9.79% of the people. This implies that many people who could have been targeted to improve their frequency of visits will not be captured. Thus, the option of choosing 36 days is also eliminated.

An inactivity period of 21 days could be an appropriate window as it covers more than half of the customer base and would effectively capture about 16% of the target population. It must be understood that the company will have limited resources and not all the customers could be targeted.

If a window of 15 days is chosen, although 21.5% of the customers might be targeted but it might not be worth the cost of spending on them as they might visit the store after 15 days themselves anyway. This is because it is more likely for a customer to visit once in 3 weeks than once in 2 weeks. A churn period of 21 days is a safe option to assume that if the customer does not visit in 3 weeks, there is a high chance that he can churn so the necessary policy could be adopted to prevent him from churning. By choosing this window, resources might be used more efficiently.

Feature Engineering and Selection:

1) **Lagged Features –** To construct the input variables for the classification problem, information related to the customers, their time of purchases, quantity bought, and the value of transaction was used. As the problem involves the element of time, lagged features were constructed based on the reference date of 29/10/2019. The tumbling window was chosen to be 7 days. One month's data used to build the model. The most recent data was used because given the context of the problem, in order to predict churn, the current customer base should be considered instead of people who had churned in the past and did not come back. In order to predict churn, an important aspect which needs to be considered could be the frequency of visits of customers in a given time period. If the customer has not visited the business recently, there is a potential chance that he might have churned. Furthermore, the quantity of items purchased might also reflect the buying behavior of the consumers and how often they visit. For example, a person who purchases more quantity might visit less often than a customer who buys less quantity but purchases more frequently. It might also be interesting to investigate the value of purchases. All these factors could be included while building the churn model.

   4 lagged features (f1-f4) were created for the number of visits of active customer within a particular week for an entire month. Similarly, an additional 4 features (f5-f8) were created for the quantity of items purchased by a consumer in the given time frame. Another 4 features (f9-f12) were created to analyze the value of products purchased per customer.

2) **Output Variable -** The output variable was constructed as a binary variable i.e. if a customer visited the store once or more times in the 21-day window, he will be given a value of 1. On the other hand, if the customer did not show up, a value of 0 was given. Here it is important to remember that churned customers were given the value 0 and not churned customers were given the value 1. So, the target class was 0.

3) **Correlation -** Feature selection is an important step in model development to make sure that only relevant variables are selected which can predict output accurately. To see the relevance of the variables to the output, a correlation matrix was used. **Fig 1.1** shows that the features f1-f4 show the

	f1	f2	f3	f4	f5	f6	f7	f8	f9	f10	f11	f12	churn
f1	1.000000	0.385306	0.458550	0.497579	0.674713	0.182275	0.254392	0.310622	0.639454	0.132800	0.208169	0.299329	0.254070
f2	0.385306	1.000000	0.412212	0.511062	0.246631	0.595545	0.227485	0.309416	0.192491	0.528400	0.169754	0.286537	0.230641
f3	0.458550	0.412212	1.000000	0.533278	0.288646	0.174155	0.648657	0.297511	0.243806	0.110931	0.608137	0.276058	0.256803
f4	0.497579	0.511062	0.533278	1.000000	0.313698	0.255514	0.298045	0.638303	0.274668	0.206514	0.258450	0.642849	0.319971
f5	0.674713	0.246631	0.288646	0.313698	1.000000	0.293931	0.289329	0.388891	0.893041	0.209728	0.223559	0.345664	0.189953
f6	0.182275	0.595545	0.174155	0.255514	0.293931	1.000000	0.231933	0.341061	0.228314	0.795523	0.157018	0.289329	0.149823
f7	0.254392	0.227485	0.648657	0.298045	0.289329	0.231933	1.000000	0.322193	0.233458	0.140717	0.860022	0.288756	0.193903
f8	0.310622	0.309416	0.297511	0.638303	0.388891	0.341061	0.322193	1.000000	0.324752	0.238859	0.258454	0.909690	0.244389
f9	0.639454	0.192491	0.243806	0.274668	0.893041	0.228314	0.233458	0.324752	1.000000	0.207139	0.233899	0.335206	0.186641
f10	0.132800	0.528400	0.110931	0.206514	0.209728	0.795523	0.140717	0.238859	0.207139	1.000000	0.126232	0.250039	0.133032
f11	0.208169	0.169754	0.608137	0.258450	0.223559	0.157018	0.860022	0.258454	0.233899	0.126232	1.000000	0.282694	0.170740
f12	0.299329	0.286537	0.276058	0.642849	0.345664	0.28[illegible]	[illegible]	0.909690	0.335206	0.250039	0.282694	1.000000	0.258121
churn	0.254070	0.230641	0.256803	0.319971	0.189953	0.14[illegible]	[illegible]	0.244389	0.186641	0.133032	0.170740	0.258121	1.000000

Fig 1.1

highest correlation with the churn feature. Thus, it might be useful to include these features. On the other hand, it is observed that there is a very high correlation between quantity features (f5-f9) and value features (f9-f12). This high correlation can lead to the problem of multicollinearity and reduce the effectiveness of the model. Thus, it was decided to drop f9-f12 as they had comparatively lower correlation with the output variable than features f5-f8.

4) **Wrapper Method -** In order to remove features with no/low variance Boruta package was used. This method helped in understanding how specific features affect prediction accuracy individually and in combination with other features. The method threw away f4. However, comparison of the accuracy scores as a result of predicting with boruta selected features and all features showed that it is better to include all features from f1 to f8. The accuracy score with all features was 73.34% as compared to 73% of Boruta selected features. Thus, it was decided to include features from f1-f8 in the model.

Data Preprocessing:

1) **Data Scaling –** As the features included values with different units, it was important to scale them before training the model. Standard Scaler was used to standardize all the numerical variables.
2) **Imbalance Classes –** It was analyzed that the churned customers were around 27% as compared to 73% non-churned customers in the training set. SMOTE technique was used to oversample the minority class i.e. churned customers to improve model performance. It was also analyzed that if oversampling was performed before the temporal holdout evaluation, it led to overfitting which resulted in over-optimistic results for all the models. Thus, it was done within the loop while conducting evaluation of models.

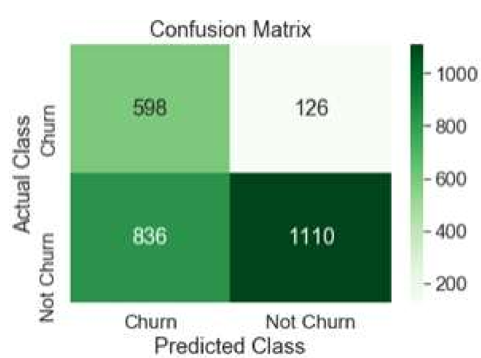
Choice of Classifiers:

Four different models were tested for the purpose of choosing the best classifier for the given problem. The models tried included Random Forest, SVM, Logistic Regression and KNN.

Hyperparameter Tuning

In order to improve the performance of the models, hyperparameter tuning was undertaken. Random Search was used for this purpose. It evaluated *n* uniform random points in the hyperparameter space, and selected the parameters producing the best performance. This was used instead of GridSearch because it took less of the computational time and yielded the same results.

Random Forest Classifier:

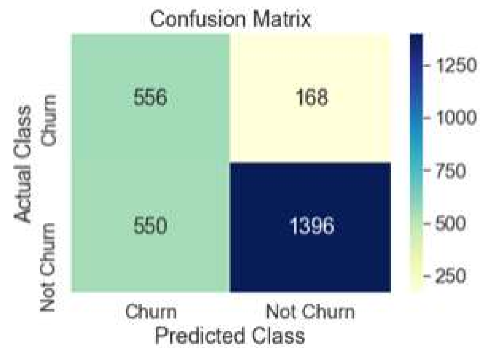


In random forest, the max_features were set to "auto". This allowed the algorithm to include all the features which made sense to every tree. There were no restrictions on the individual tree. Secondly, n_estimators were chosen to be 40 as it made the model predictions better and faster. A higher number was also tested however, it failed to significantly improve the performance of the model and just made the code run more slowly. After training the model, it was able to generate an overall accuracy of 64%. The sensitivity was 0.83 which was higher than the specificity of 0.57.

	precision	recall	f1-score	support
0	0.42	0.83	0.55	724
1	0.90	0.57	0.70	1946
accuracy			0.64	2670
macro avg	0.66	0.70	0.63	2670
weighted avg	0.77	0.64	0.66	2670

This implies that this model predicted the churned customers correctly 83% of the times, however, it did not predict the non-target class i.e. the non-churners well. Furthermore, this model has an AUC score of 0.7 which shows that the ratio of predicting true positives over false positives. This is a good performance indicator as we would preferably want to predict more actual churners that falsely predicting the non-churners as churners. The F1 score is 0.55. F1 score combines the precision and sensitivity. In this case, it is lower as compared to other models because of the lower precision. This implies that when it comes to evaluating the proportion of actual churned consumers which were correctly classified as churned, the model gives the right result only 42% of the times. In this dataset, as we have class imbalance, precision would be a better measure of performance as it does not take into account the true negatives i.e. non churners which were accurately classified as non-churners.

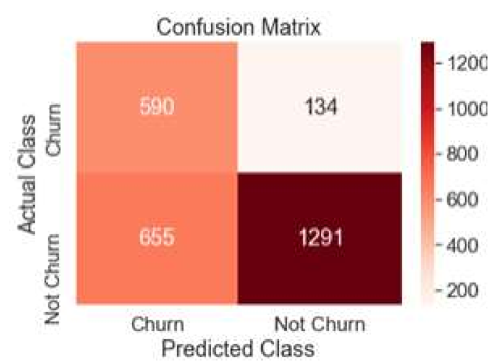
Support Vector Machine:



While setting the hyperparameters in SVM, the c-value was set to 1. After testing for lower as well as high values of c, it was decided that a relatively lower value tends to give results which are more generalizable. Secondly, rbf kernel was chosen and the gamma value was set to 0.1 as a result of hyperparameter tuning. The results after training suggested that the model has an overall accuracy of 73% which is better than random forest. This model performed much better in classifying the proportion of churned customers correctly. 78% of the churned customers were classified correctly whereas 71% of the non-target class was also predicted accurately. The F1 score is 0.61 which ranks as one of the best amongst rest of the models. The precision of the model is 50% which is better than random forest however, it must be noted that this model also incorrectly classifies many non-churn customers as churners which could potentially lead to wastage of company's resources. However, it is interesting to note that the false positives are greater than the false negatives. The AUC score of 0.74 reflects that if the business has more tolerance of true positives over false positives, then this model is the best as compared to rest of the models.

	precision	recall	f1-score	support
0	0.50	0.78	0.61	724
1	0.90	0.71	0.80	1946
accuracy			0.73	2670
macro avg	0.70	0.75	0.70	2670
weighted avg	0.79	0.73	0.75	2670

K- Nearest Neighbor:

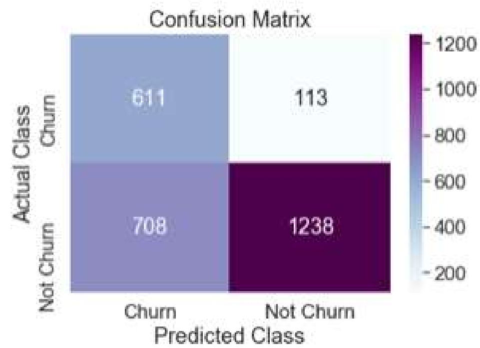


The model was tested for several parameters. After comparing a lower value of k such as 5 and a value as high as 27, a higher value was chosen. This was because it led the model to become less sensitive to noise and was smoother between the boundaries of classes. Euclidean distance was chosen. This model performs really well in terms of sensitivity. The churn customers were accurately predicted 81% of the times. The overall accuracy of 70% was nearly same as rest of the models. The precision of the model is slightly worse than SVM, however, the F1 score is the almost the same at 0.6. If the AUC score is analyzed it is 0.74 which is equivalent to the SVM model. This implies that this model misclassifies a greater proportion of actual non-churn customers as churned than correctly classifying churned customers. In the business context, if the cost of targeting

	precision	recall	f1-score	support
0	0.47	0.81	0.60	724
1	0.91	0.66	0.77	1946
accuracy			0.70	2670
macro avg	0.69	0.74	0.68	2670
weighted avg	0.79	0.70	0.72	2670

customers is greater than the cost of losing out customers which haven't been targeted then this model might not be the best option.

Logistic Regression:



In this model, a higher level of c was chosen after hyperparameter tuning. It corresponds to lower level of regularization. This implies that the model will be less sensitive to the noise in the data. After training, the results showed that this model outperforms all of the models in terms of sensitivity. It predicts the churned customers correctly 83% of the times. However, it must be noted that the false positives are the highest among all the models. If the cost of targeting customers is very high, this model might not be a good choice as the precision of 0.46 is second lowest among all models. On the other hand, it will misclassify a low number of customers as non-churn which means that less customers might be potentially lost because of the inability of the business to identify customers with high probability of churning. The overall accuracy of this model is also second lowest after random forest at 69%.

```
              precision    recall  f1-score   support

           0       0.46      0.84      0.60       724
           1       0.92      0.64      0.75      1946

    accuracy                           0.69      2670
   macro avg       0.69      0.74      0.67      2670
weighted avg       0.79      0.69      0.71      2670
```

Evaluation Strategy

Before choosing the right model, it is important to understand whether the cost to target wrong consumers is higher than the cost of revenue lost by not re-engaging the potential churners. In the former case, specificity of the model will be the key to assess its success as the cost of predicting the non-target class incorrectly will be too high. This is because resources will be wasted targeting incorrect consumers. It might also annoy the consumers if they are bombarded with irrelevant emails or marketing schemes. In the latter case, sensitivity will be important because as the churners will not be targeted accurately, the cost of acquiring new customers over long term will be significant.

To evaluate the models, several metrics were used. A high sensitivity score is usually preferable because accurately predicting the churners will help the business devise an effective strategy to reengage them. Furthermore, specificity will also be considered as the company might want to keep the cost of targeting the incorrect customers to the minimum. F1 Score which is the weighted average of Precision and Sensitivity will also be used instead of accuracy because accuracy is useful if false positives and false negatives have similar cost. In the business scenarios, it is safe to assume that cost of lost revenue differs from cost of targeting wrong customers. The models will also be evaluated on the basis of Receiver Operating Characteristic Curve. AUC will measure the quality of models' prediction against the parameters of True Positive and False Positive Rate.

Classifiers	**Sensitivity**	**Specificity**	**Accuracy**	**AUC**	**F1 Score**
Random Forest	0.83	0.57	0.64	0.70	0.55
SVM	0.78	0.71	0.73	0.74	0.61
KNN	0.81	0.66	0.70	0.74	0.60
Logistic Regression	0.84	0.64	0.69	0.74	0.60

Model Selection:

In order to evaluate the performance of 4 chosen classifiers, repeated temporal holdout method was used. This is because the traditional cross validation randomly splits the data and as our observations have a time element attached to them, there is a possibility that the model would be trained based on the knowledge of future which we will not exist in real life scenarios.

The features formed by the data of active customers in one-time window were used to train a model. The model made the prediction for the next time step. The prediction was then stored and evaluated against the known values. Tumbling window was used to go back one step at a time and hold out the test set to repeat the process. 10 hold out sets were used, and the average balanced accuracy and F1 Score was used to evaluate the performance of the model. The reason behind choosing balanced accuracy and F1 score was that in this case, both false positives and false negatives are important to determine the effectiveness of the model. Furthermore, the ability of the model to precisely predict the proportion of churned customers is also valuable.

The results suggest that SVM and Logistic Regression perform considerably well. However, the final model selected could be SVM. This is because in this case we are assuming that the cost of false positives is high. Logistic Regression misclassifies a lot of non-churners as churners which might add burden on the company's resources and divert the attention from churners that are actually important. SVM on the other hand performs better in terms of precision and F1 score. It also performs better on the dimension of specificity i.e. it is better in predicting the non-churners as well. The AUC score of SVM is also high which means that the ratio of predicting true positives over false positives is high which makes it a suitable choice for classification.

	Random Forest	**SVM**	**KNN**	**Logistic Regression**
F1 Score	0.71	0.837	0.838	0.833
Balanced Accuracy	0.693	0.693	0.658	0.695

Marketing Insights:

The analysis suggests that in the month of November, out of a total of 2670 customers, 724 customers churned. This is 27% of the active customers. This implies that these customers have probably switched to competitors. There are also consumers who have reduced the frequency of their purchases. They visit the stores less often and buy less quantity of items. As the cost of attracting new customers is very high compared to retaining existing customers, it is important to identify the particular customers who have stopped purchasing from FoodCorp. An effective marketing strategy needs to be implemented to re-engage these customers and encourage them to continue buying. After analyzing the characteristics of these customers, pen portraits of two groups of consumers were made which are discussed in the later section.

Average Visits ***Fig 2.1*** **Comparison:**

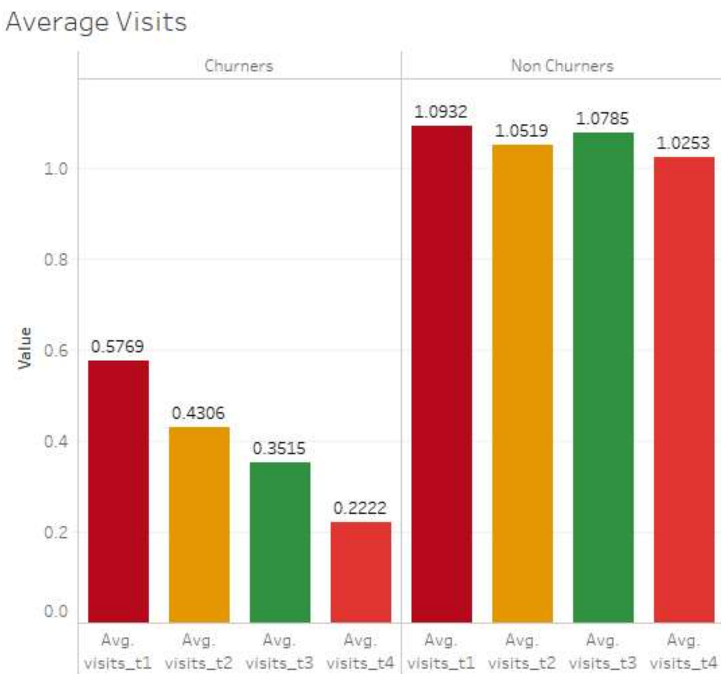


Fig 2.1 shows that 4 periods of 21 days each (t1-t4) were seleced starting from the reference date and moving backwards by subtracting 21 days from each period. The graph shows that there is a difference between the average visits of a customer who is likely to churn and the non-churners. This reaffirms the findings that an average churner is less likely to visit the store within 21 days. On the other hand, the non-churners usually visit the store at least once or more within 21 days. Their visiting patterns are consistent. The declining average visits of the churners show that these customers can visit once in a while and then disappear for a period of time which is why the standard deviation of their visits is higher than the non-churners.

Department Wise Spend

department_name	Churners	Non Churners
BAKERY	546	20,213
BEVERAGES	62	8,553
CIGARETTES & TOBACC	111	7,761
CONCESSION 1		10
CONCESSION 5		48
CONCESSION BAKERY	8	1,648
CUISINE DE FRANCE/I	95	8,483
DELICATESSEN	206	6,291
ELECTRICAL	28	275
EMB		119
FLOORCOVERINGS		0
NEWSPAPERS & MAGAZI	399	12,744
NON FOOD		16
OTHER GROCERY		294
PAYPOINT	37	3,369
PETROL		591
POSTAGE STAMPS		303
POWER CARD/KEY CARD		65
PROMOTIONS		19
PROVISIONS DAIRY	752	21,855
PROVISIONS PREPARED	364	14,340
SEASONAL	3	3,069
SOFT BEVERAGES	434	14,451
SOFT FURNISHINGS		285
STORE EXPENSES	12	564
SUSPENSE ACCOUNT		206

Departmental Spend Analysis:

The basket analysis can also help in identification of customers. If the baskets of the churners and non churners is analyzed **in Fig 2.1**, it is interesting to note that across all departments, the spending of churners is considerably lower than the non churners. This could be attributed to the fact that the churners visit the store less frequently and buy less items than the non-churners.

Another important observation is that the churners spend only in food related departments. The highlighted departments show that there are no transactions related to churners in these areas. It is evident that the churners do not buy clothing and housng products from the store. This implies that these mightbe one off customers who just visit the store once in a while for buying necessities. They might not belong to the regular customer base and might shop somewhere else. On the other hand, it can be seen that non churners spend in departments such as petrol, lottery, promotions and homeware products. This could indicate that these consumers might be frequent shoppers and visit FoodCorp for their weekly and monthly shopping**.**

Pen Portraits:

Low Spending Infrequent Customers – These are the customers who usually visit once after more than 21 days. They can also disappear for more than 42 days. They generally spend on low value products such as grocery, tobacco etc. A major part of their spending is food related items. Their spending patterns are inconsistent and there is a lot of variation in the amount of money that they spend on each visit.

Potential Loyal Customers – These customers visit the business at least once within 21 days. They generally spend on a wide variety of products ranging from grocery, clothing and household products. They are more likely to be loyal to the business and might be part of the loyalty scheme. Their spending patterns are consistent, and they usually spend a greater amount of money on each visit. heir basket size is relatively larger than the low spending infrequent customers.

Fig 2.2

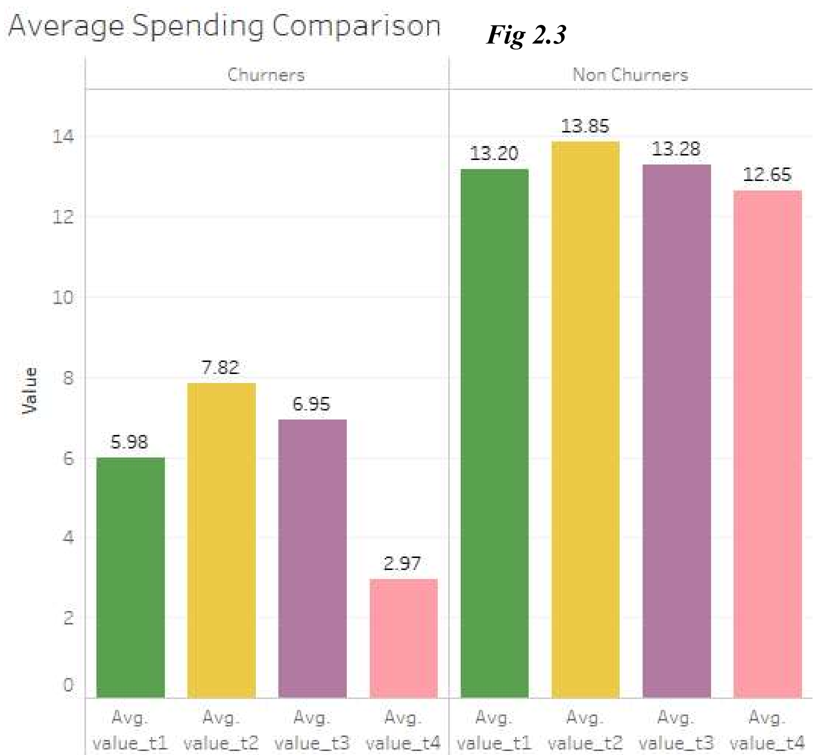


Fig 2.3

The data regarding the average spending of active customers during these periods was analyzed. It was interesting to note that the average spend of the non-churners was relatively consistent than the churners. When the standard deviation of their spent is analyzed, it ranges between £18 and £21 across periods. On the other hand, if churners are analyzed, it is seen that their average spending is inconsistent across different periods. If standard deviation is considered, the spent ranges between £8 to £15. This implies that churners are generally not habitual buyers. The spending patterns of these consumers are hard to predict as they vary a lot. On the other hand, the non-churners might be regular buyers who spend a greater proportion of money on their visits and plan their purchases.

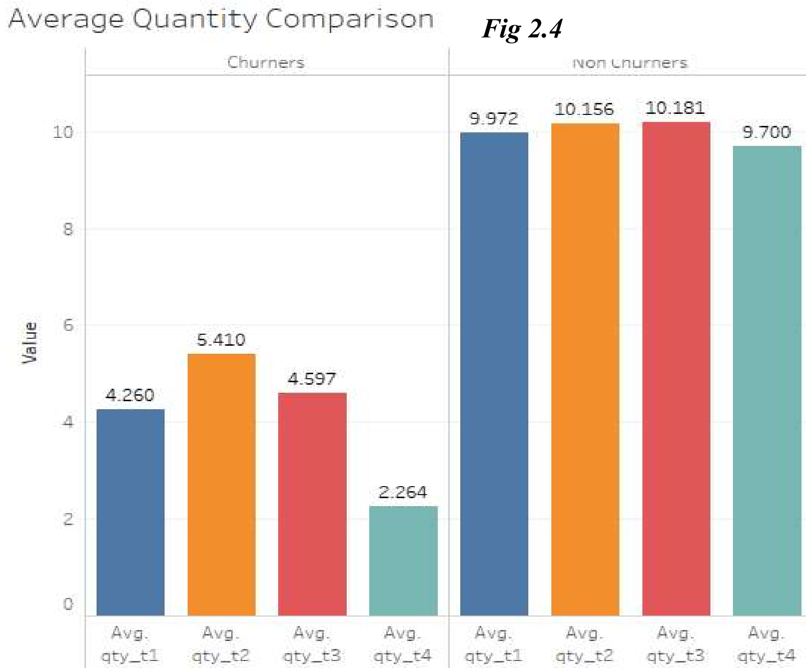


Fig 2.4

If the average quantities are analyzed, the same pattern appears across churners and non-churners overtime. The non-churners on average buy larger quantities of products than the churners. This again can be attributed to the fact that these people might be habitual shoppers.

Recommendations:

The above analysis suggests that if the business wants to identify the churners, it is important to keep track of their purchase patterns. If a customer buys a quantity of less than 6 items, and on average spends less than £8 over a time frame, his probability of churning is more than a customer who buys an average more than 9 units per visit and spends at least £12 pounds. The first step in improving the churn rates is identifying the people who are likely to churn. Secondly, their baskets should be analyzed, and they could be given promotional offers based on their interests. They could also be incentivized to be a part of loyalty programs to keep them engaged.