BAU Journal - Science and Technology

Volume 1 | Issue 2
ISSN: 2706-784X

Article 7

June 2020

CUSTOMER CHURN ANALYSIS USING BINARY LOGISTIC REGRESSION MODEL

Mohammad Abiad
PhD Student, Faculty of Entrepreneurship, Business Engineering and Management, University Politehnica of Bucharest, Bucharest, Romania; College of Business Administration, American University of the Middle East, Kuwait, Mohammad.abiad@aum.edu.kw

Sorin Ionescu
Professor, Faculty of Entrepreneurship, Business Engineering and Management, University Politehnica of Bucharest, Bucharest, Romania, sc.ionescu@gmail.com

CUSTOMER CHURN ANALYSIS USING BINARY LOGISTIC REGRESSION MODEL

Abstract

These days, the fundamental target of all organizations is to accomplish a reasonable gainfulness. Organizations are looking to decide the elements, which has direct impact on their benefit and attempt to investigate them so as to pick up favorable circumstances on their rivals. With this solid aggressive market, holding customers turns into the fundamental objective of telecom mobile service providers opposed to drawing in new customers, since the expense of holding a customer is viewed as low contrasted with the expense of pulling in another customer as referenced before by numerous researchers. Thus, Customer Churn Analytics is one of the significant elements that organizations should concentrate on to help their goal of accomplishing a reasonable benefit. In Telecommunication area, the business aims to serve customers, and analyzing customer churn will give the executives a thought regarding the probability of a customer to leave the organization. The primary goal of this paper is to highlight the main factors that affect customer churn and then analyze them by using the Binary Logistic Regression Model. A case study in Kuwait telecommunication section was taken as an application for such algorithm. A non-probability sampling technique is used (Snowball) to select a sample of 120 customer and analyze them by using the statistical software SPSS and MINITAB, results shows that the Customers' Gender, Marital Status, Monthly Bill as well as the Service Quality have significant effects on Customer Churn.

Keywords

Telecommunication, Churn Rate, Binary Logistic Regression

1. INTRODUCTION

Customer Churn is the state of a customer leaving a company, regardless of whether to join a competitive one or not. "Customer churn refers to when a customer switches from one service provider to another" (Delvi et al, 2016). Improving company's profitability nowadays is based on the analysis of Customer churn especially in service based companies. "Building customer relationship is a backbone for all organizations in general, and companies in service industries in particular" (Angelova & Zekiri, 2011). At the point when fantastic relationship exist with customers, lower churn rate results. Henceforth the fundamental objective of any association is to check the factors that could influence this connection which influence the customer beat and examine it appropriately so as to evade such churn rate. At first, one would check the possible factors that can affect the customer choice to leave his company, this could be done by reading what scholars have significantly included in their research result, adding to it the input of experts from the same domain under study. Analyzing such factors will give an idea to companies about the significant factors they have in their market, and let them plan accordingly to avoid high number of churn rate. It is well known that looking for new customers is no more the main objective for companies, what is important nowadays is to plan satisfying your current customers and attract them more to stay with you. This is because of the high cost the company will face to attract new customers as stated by many scholars such as Sulaimon et al, 2016. Likewise, numerous scholars have referenced that held customers permits companies to improve their benefits, and companies may get higher profit by trying to lower the churn rate even for a small amount. Therefore, Customer Churn analysis provides advantages to those who analyze it properly. The difficult part in such analysis is usually the data collection and sample size. Higher sample size leads to more accurate results. Hence, in telecommunication sector the huge database omits this weakness and such data can be easily utilized for enhancing the company's performance which yields to decrease the customer churn rate and so increase their profit. "Customer Churn has a huge impact on companies and it is the prime focus area for the companies to remain competitive and profitable" (Mahajan et al, 2017).

Analyzing customer churn statistically has different methodologies that can be followed, and different stages should be implemented before reaching a result. In our research paper, we will use the Binary Logistic Regression Model to investigate the factors that has an effect with customer churn and try to highlight the important area for companies to focus on when planning to decrease the customer churn. Hence an algorithm for customer churn analysis using the binary logistic regression model is stated which has the following analysis workflow: A review of Customer churn background followed by listing the important factors that affect customer churn based on scholars' as well as experts input, Statistical Analysis used, and finally the algorithm implementation which provide an output that is represented by the regression equation which can be used for interpretation purposes which leads to suggestion plans for managers to avoid the high rate of customer churn. An application of this algorithm was taken in Kuwait Telecommunication Sector, a survey of 15 questions has been created, and data has been collected and then analyzed by using MINITAB and SPSS.

This paper is divided into 8 sections, Section 1: INTRODUCTION, Section 2: THE PROPOSED ANALYSIS WORKFLOW, Section 3: REVIEW ON CUSTOMER CHURN BARCKGROUND, Section 4: CUSTOMER CHURN IMPORTANT FACTORS, Section 5: STATISTICAL STAGES USED IN DATA ANALYSIS, Section 6: ALGORITHM IMPLEMENTATION, Section 7: APPLICATION IN KUWAIT TELECOMMUNICATION SECTOR, Section 8: CONCLUSION.

2. THE PROPOSED ANALYSIS WORKFLOW

The best result of any statistical study comes from a well-organized and clear plan for the analysis. In this paper we will be proposing an algorithm for customer churn analysis using the Binary Logistic Regression Model. Fig. 1, shows the algorithm workflow in details and then in the next sections each step of the algorithm will be explained.

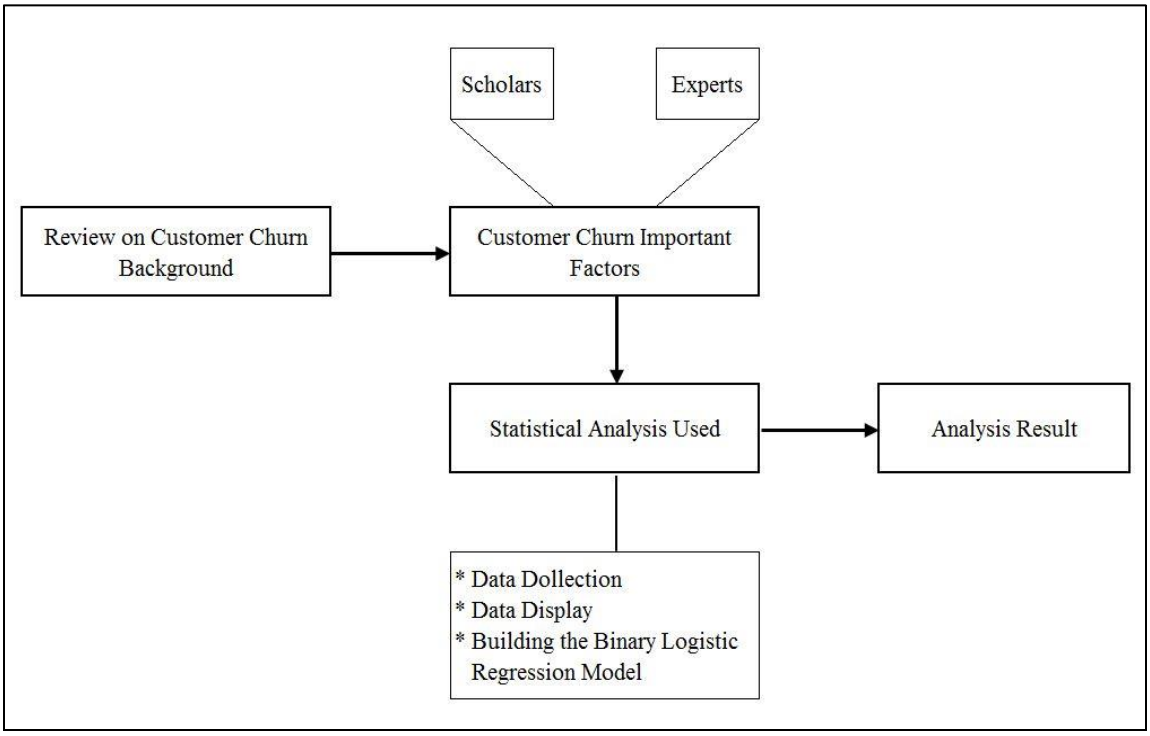

Fig.1: The Proposed Analysis Workflow
Reference: M. Abiad & S. Ionescu, November 2019

3. REVIEW ON CUSTOMER CHURN BACKGROUND

"Churn is a complicated phenomenon, which involves deciphering each and every nerve of customer behavior, analyzing technological advancements in the handset market (multiple SIM) and taking a comprehensive stock of competitor's product and service offerings" (Mahajan et al, 2017). "With proper management of customers, we can minimize the susceptibility to churn and maximize the profitability of the company" (Dahiya & Bhatia, 2015). The effect of Customer churn is very high on companies, since customer churn has a direct effect on company's profitability. Scholars discussed the result of customer churn in many research papers or conferences and all of which they highlight the importance of avoiding customer churn in service based companies. There are many factors that can affect directly or indirectly the customer decision of leaving the company, in another word that affect the customer churn. These factors are to be displayed, listed and then analyzed by appropriate tools in order to plan for decreasing such churn rate.

In telecommunication sectors, customers are of different types, some of them take immediate action and leave the company without prior warning or even prior discussion with customer service team, and another type of customers are those who give a chance, try to solve their problem and be satisfied before giving-up and taking the decision to leave.

"Customer Churn Prediction is one of the most important concerns of a telecom's companies" (Amin et al, 2018). There are many factors that make the customers stay even if they are not satisfied, but at the end, customers might have some limits and cannot hold on for a long time. For this reason, companies should play an important role with both types of customers, and seek for what types of action should be taken to keep their customers loyal to them and make them more satisfied and hence decrease the customer churn rate. Predicting which group having higher potential to churn can be analyzed statistically by studying many factors including the demographic factors, technical factors and cultural factors. "Churn Prediction model can help analyze the historical data available with the business to find the list of customers which are at high risk to churn" (Sulaimon et al, 2016).

Customer churn can be analyzed by using the historical data which includes the customer experience of their monthly bill, years of experience with the service provider and the level of satisfaction if monitored from time to time.

The rapid growth in technologies make it very competitive to telecommunication companies analyzing their customer behavior since historical data sometimes will not be useful for the analysis, therefore when running any prediction analysis you should be very careful to choosing the appropriate model.

4. CUSTOMER CHURN IMPORTANT FACTORS

Analyzing customer churn differs from country to another, from culture to another. Hence the factors that play an important role in this analysis might differ too. For that reason, in this paper, at first we will be listing the factors summarized by scholars which has an effect on customer relationship and then adding the factors by experts from the same country. Generally, scholars mentioned the importance of service quality in customer relationship with the company and how this factor affect the lifetime of the customer with the service provider. Service Quality has direct effect on customers and hence service providers are attentive this factor is one of the critical factors that affect the customer satisfaction as mentioned by Mannan et al, 2017. “Service quality was positively and significantly related to customer loyalty”, (Bhuian et al., 2018). “Customers considered call quality as the important antecedent in the retention decision”, (Adebiyi et al, 2015).

Another factor listed by scholars is the Promotion, “The use of this promotional mix has increased brand awareness, customer retention and profitability of the company”, (George et al, 2017). When customers are satisfied from the promotions received by their providers, there will be high chance of decreasing the churn rate of this customer, of which this should be tested in this study. In addition, the brand images was recently analyzed by scholars and found to be significant factors on customers, hence customers’ satisfaction from their service provider brand image is an additional factor that will be taken into consideration when analyzing the customer churn. “Brand image is the most important factor for customer satisfaction”, (Azam and Karim, 2017).

“Factors that significantly contribute the prediction of customer churning intention are; tariffs, transparency level, promotions, technical assistance, privacy and response to complaints” (Hamelin et al, 2009). Also, many scholars stated that the demographic factors such as Age, Gender and Location can affect the customer churn. Experts in the domain mentioned that the years of experience, marital status, and educational level could also be included as important factors in the analysis.

Accordingly, the factors that will be included in this study, to analyze the customer churn in Kuwait telecommunication section are as follow: Age, Nationality, Gender, Marital Status, Educational Level, Location, Customer’s Working Status, Name of Service Provider, Years of Experience with the service provider, Monthly bill, Customers level of satisfaction from the service quality, Brand Image, Promotion offered as well as the Level of Satisfaction from the customer service team.

5. STATISTICAL STAGES USED IN DATA ANALYSIS

Statistical stages in a study are aimed to reach a certain conclusion drawn from an input data which has a specific objective. In any statistical analysis, a cycle should be followed with the following stages: ‘Ask’, ‘Collect’ ‘Describe’, ‘Analyze’, and then ‘Conclude’. Hence before proceeding with data analysis, one should make the objective of the study to be clear in order to analyze data properly. Once the main objective of the study is clarified, the first stage is to collect data, this stage is directly followed by data cleaning and data summary, so that the data is described and ready to be easily analyzed. Appropriate models are applied in the next stage to reach significant conclusions.

In what follows, we will summarize the main stages used in data analysis starting with the details of data collection, followed by data summary and then discussing the Binary Logistic Regression Model that fits our case of study which is the customer churn analysis.

5.1 Data Collection

Data could be collected by using different techniques, such as Monitoring, Experiment, and Recording as well as by building a statistical survey. Data is called either primary or secondary. Primary data is the data collected and used for the first time whereas secondary data is an already existing data in the database, or it is the data collected and used earlier by someone else.

An important step in collecting and analyzing data is to know the types of variables included in your data, that is, whether the variable is considered to be quantitative which mean numerical, or qualitative which means categorical. The type of the variable affect the stage of data analysis when choosing the appropriate model to be used. After choosing the way of data collection, then an appropriate sample size should be selected to get significant results that can be generalized for the population.

In this paper, the main objective of the study is to analyze customer churn, hence data should be collected from customers and should include different types of variables to cover as much factors as possible that might affect the customer churn.

5.1.1 Statistical survey

A survey to be created targeting all the important factors that affect the dependent variable. In addition, the survey should include a question for the dependent variable of the study. Variables in the survey will be of different types, quantitative or qualitative all of which they are related to the case of study. Surveys should be tested on a small group of individuals before distributing it, so that questions and their answers might be modified in order to avoid weird cases in the collected data.

5.1.2 Choosing an Appropriate sample size

The sample size is the number of customers to be selected as a part of the total population in order to be used for data analysis. This sample should be selected in a way that makes it representative of the population. The more you increase your sample size the more accurate results you can reach, but due to a lot of constraints that should be taken into consideration, the sample size might not be large enough. Such constraints are like time limit, low budget for data analysis, security restrictions for some information and more. There are some measure used to determine the appropriate sample size to be selected based on a given target of significance level, such measures are like the Cochran measure. The use of MINITAB, SPSS and other statistical software are making it easier to calculate the sample size, since there are existing features that can do such calculation based on the input parameters.

After defining the survey and the appropriate sample size, sampling technique will be an important step in order to make our sample well representative for the population. Different sampling techniques are used such as Probability sampling such as stratified and random, as well as non-probability sampling such as snowball and convenience.

5.2 Describing Data

Data is considered the major element in any statistical analysis, without data there is no space for accurate analysis. Hence after we collect the needed information concerning what is needed to be analyzed, the step of data cleaning is reached. In this step, data should be cleaned from typos, missing information, outliers and so on. This could be done by using summary statistics which is displaying the data in either tabular form or graphs. Tables and graphs include frequencies and percentages that describe your data and from here you can spot extreme values, missing values so that you clean them before going to the analysis part. Describing data is not only used for data cleaning, hence even after cleaning your data, you should plan to summarize your data in order to get an initial idea of the study and get information about the current situation of the target.

The type of the variable under study, whether quantitative or qualitative, plays an important role in choosing the appropriate type of graphs and summary tables:

5.2.1 Qualitative variables

For categorical variables (Qualitative variables) we can use the summary table as well as one of the following graphs: Pie Chart, Bar Chart or Pareto Chart. Pie Chart is a circular representation used if the variable is categorical and has limited number of categories.

Bar Chart is a column representation of a variable, it could be drawn vertically or horizontally and the height of the bars represent the frequency/percentage of the category, this type of chart is easily used when we have large numbers of categories.

Pareto Chart is the Bar Chart in a descending order along with the cumulative line plot on the same graph, that is, you can read the cumulative percentage as well as the frequency of all categories on the same graph.

5.2.2 Quantitative variables

For numerical variables (Quantitative variables) we can use as one of the following graphs: Histogram, Boxplot, Scatterplot. In addition a summary table can be drawn which include some numerical measures that provides you information about your data location and dispersion, such measures are like the mean, median, standard deviation, coefficient of variation, skewness and so on.

The Histogram is a column graph, with classes for the variable, that is, the numerical classes are grouped into interval in a scale that fit the values, from this graph, and one can have an idea about the shape of the distribution or even study the skewness. Boxplot is a graphical representation that is usually used since it has more than one numerical measure on the graph, such as the minimum, maximum, quartiles and the mean. Boxplot is also used to detect outliers in your data. Scatterplot is used to study the relationship between any two numerical variables so that you can check whether you have a linear or non-linear relationship or even checking if there is no relationship. Sometimes, data should be transformed by using log transformation function in order to study the relationship, this is usually done when the scale of the two numerical variables are not close to each other. The scatter plot is the first step of studying the relationship as for linear relationship, the correlation coefficient should be used to determine how strong this relationship is.

5.3 Building Binary Logistic Regression Model

After collecting and describing your data, the stage of data analysis is reached. Different statistical analysis can be applied by using the statistical software such as SPSS and MINITAB and more. Regression Analysis have more than one type all of which is related to the number of independent variables as well as the type of the dependent variable. If the dependent variable is numerical, the we can either use the simple linear regression model in case of only one independent variable, or the multiple linear regression model in case of more than one independent variables. The independent variables should be numerical or dichotomous variables (has values 0 or 1), so in case of a categorical variable with two options, then these options should be recorded as 0 or 1. If the categorical variable has 3 options (possible answers) then this variable should be entered into the model as two separate variables, one variable represent the first option, and the second variable represent the second option, in this case, the interpretation will be reference to the third option of the variable.

In our paper, we are focusing on analyzing the customer churn by running the Binary Logistic Regression Model. "Logistic regression is used to predict a categorical (usually dichotomous) variable from a set of predictor variables" (Core.ecu.edu.). Binary Logistic Regression model is used in this analysis to determine the customer churn possibility. Assumptions of the binary logistic regression model are such as "dependent variable should be binary", "Observations are independent of each other", "little or no multicollinearity among the independent variables", "the independent variables needs to be linearly related ti the log odds" and finally "large sample size is required" (Assumptions of Logistic Regression - Statistics Solutions, 2020).

The output generated by running this model is summarized by the following general equation:

Eq. (1) $y = a + b_1x_1 + b_2x_2 + \cdots + b_nx_n$

Where a is the constant parameter, and $b_1, b_2, \ldots., b_n$ are the variables coefficients, noting that the variables should be either Measurable (continuous variables) or Dichotomous (variables that take values 0 or 1) and y is the log-odds ($log(\frac{p}{1-p})$) of the variable customer churn.

The first step in the Binary Logistic Regression Model is to prepare the variables in a way that they can fit the model, that is, if the variable is numerical then you may directly include it in the model. But if the variable is Categorical, then only dichotomous variables should be included directly, while the rest should be modified by transforming them into different variables and make them dichotomous with a reference value. As an example, we have three categories of the variable 'name of service provider": Zain, Ooredoo and Viva, this variable should be entered into the model as two variables, one for Ooredoo (has value 1 if the customer belongs to Ooredoo and 0 otherwise) and the second for Viva (has value 1 if the customer belongs to Viva and 0 otherwise) then in this case the reference value in the interpretation will be Zain service provider. As for the dependent variable of the model, it will be the customer churn question raised as dichotomous with value 0 if the there is no possibility for the customer to change the service provider and 1 otherwise.

The output of the Binary Logistic Regression Model includes the Model Significance Table, The Model Summary, The Classification Table and the variables Coefficients table.

5.3.1 Model significance table

In this model significance table, result will provide the significant level of the model, once the significance is low (less than 5%), then your model is considered to be a good model.

5.3.2 Model summary

This table includes the -2Likelehood which explain how poorly the model predicts your result, and the lowest the value the better the result will be. In addition, the table include the R square value which represent the percentage of the variation in the dependent variable that is explained by the independent variables.

5.3.3 Classification table

The classification is of two parts, the false positive as well as the false negative classification. False positive means that when you predict the occurrence of the event while in reality it will not occur. Whereas the False negative is the opposite case, that is, you do not predict the occurrence while in reality the event occur.

5.3.4 Variables coefficient table

This is the table where you can see the significance level of each of the independent variables as well as their coefficient in the regression equation. The level of significance is usually defined as 1%, 5% or 10% depending on the need of the study.

By running the model, we will get an idea about which group has higher potential for customer churn than others, as well as we will have significant factors that plays an important role in analyzing customer churn in Kuwait telecommunication sector.

6. ANALYSIS RESULTS

The last stage of the analysis is to run and interpret the Binary Logistic Regression Model by reading the result of the model significance, R-square, classification table and model coefficents.

Model significant means that at least one of the dependent variable has a significant effect on the dependent variable. The R-square represnts how much the variation of the dependent variable is explained by the significant independent variables, this can be usually enhanced by adding more factors to the study, that is, more independent variables that can affect the dependent variable. Finally getting the significant model coefficients which represent the Binary Logistic Regression Equation that can be analysed. Interpretation of the model is required and should be summarized for the managers in order to plan accordingly and see which group of customers having higher potential to leave and do something to avoid having the churn.

This is the final stage, where you read the results obtained from the analysis part and draw a conclusion of your study or even get information that you should take into consideration and then update your study from stage 1 and run the model again until your reach the goal of your analysis.

7. APPLICATION IN KUWAIT TELECOMMUNICATION SECTOR

After we checked the customer churn background and discovered the list of important factors listed by scholars or expert in the domain, we built a statistical survey of 15 questions targeting the analysis of the customer churn in Kuwait telecommunication sector. This survey contained qualitative and quantitative variables that targets demographic and the important factors that can affect the customer churn. The dependent variable of our study was the customer churn question which was raised by asking the customers to state by 'Yes' or 'No' whether there is any possibility for them to change their mobile service provider. The independent variables are classified as qualititative and quantitative variables, the quantitate variables that were included in the study are summarized by: Age, Years of Experience with the service provider, Monthly bill. In addition a scale from 0 to 100 was used to get data about the following variables: Customers level of satisfaction from the service Quality, Brand Image, Promotion offered as well as the level of satisfaction from the customer service team. As for the Qualitative variables, these variables include the Nationality, Gender, Marital Status, Educational Level, Location, Customer's Working Status and the Name of Service Provider.

A pilot survey was distributed on 10 scholars and experts in order to update the survey questions before collecting data from customers and then the survey questions were modified. The survey is now ready for data collection, and in our study, 120 customers have been selected by using a snowball sampling technique due to the limited time and budget constraints faced. The snowball sampling techniques is the process of collecting data from a person you know and then ask that person to collect data from another person that he knows and so on. The application was aimed to show the implementation of such algorithm. Results will be displayed in two parts, data display and data analysis.

7.1 Data Display

Factors that were taken into consideration when analyzing the Customer Churn in Kuwait telecommunication sector could be summarized and displayed using tables and graphs. Below some results obtained from the collected data with their interpretations.

Fig. 2 Displays at the percentage of customers that have possibility to churn when they were asked whether there is any possibility for them to leave the company or not. Results shows that 55.8% of Kuwait Telecommunication Customers have possibility to Churn, which make it interesting to analyze the factors that can affect them and let the mobile service providers plan accordingly.

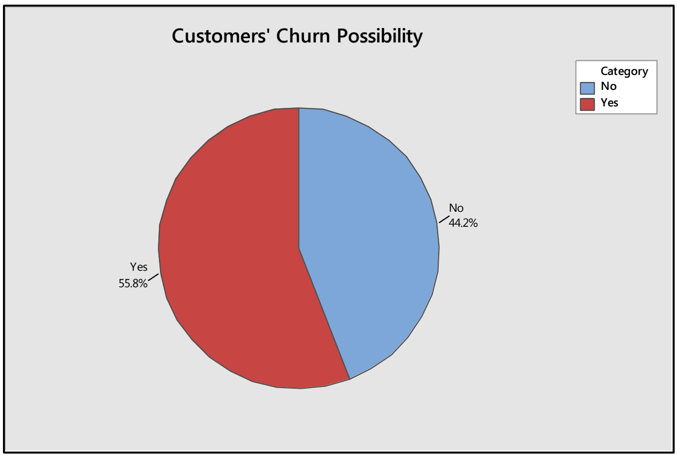

Fig.2: Customers' Churn Possibility
Reference: M. Abiad & S. Ionescu, November 2019

There are three Mobile Service Providers in Kuwait: 'Zain', 'Ooredoo' and 'Viva'. In our study, results, in Fig.3, shows that Zain and Ooredoo are leading the market with around 80% of telecommunication customers in Kuwait belong to one of these two companies.

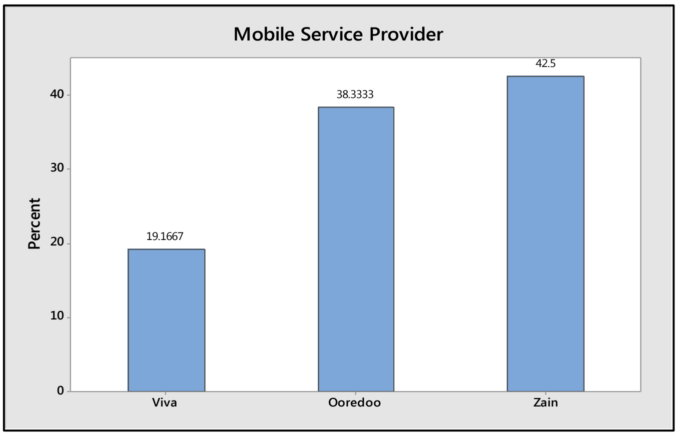

Fig.3: Customers' Churn Possibility
Reference: M. Abiad & S. Ionescu, November 2019

As a part of the data display for demographic variables, Fig. 4 shows that the gender selection in our sample was almost the same with 50.8% for Female and 49.2% for Male. In addition we can see that the majority of female customers in Kuwait are single almost double the result of non-single, where is for male it is approximately the same.

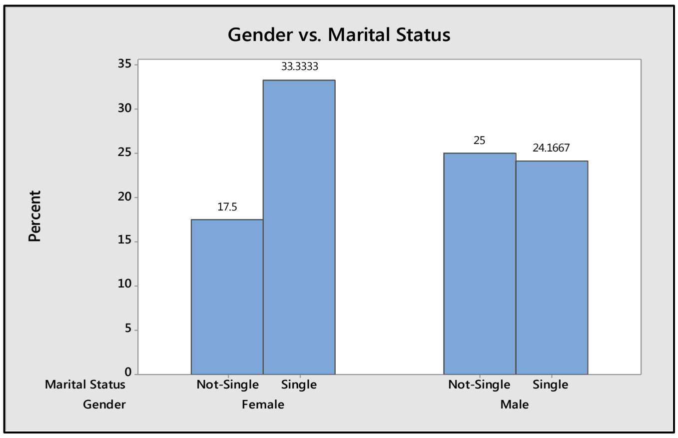

Fig.4: Gender vs. Marital Status
Reference: M. Abiad & S. Ionescu, November 2019

Data can be also displayed by using tabular forms, Table 1 represents the numerical measure calculated for the variables Age, Years of Experience, Monthly Bills, Level of Satisfaction of customers from the Service Quality, Customer Service, Brand image as well as the Promotion.

Table 1: Numerical Measure for the quantitative variables
Reference: M. Abiad & S. Ionescu

		Mean	Variance	Minimum	Q1	Median	Q3	Maximum
Age		30.5	87.58	12	22	31	36	71
Years of Experience		6.548	21.18	0.5	3	5	9.75	20
Monthly Bill		20.8	248.22	3	10	15	30	80
Level of Satisfaction from	Service Quality	72.74	259.53	10	60	78.5	80	100
	Customer Service	66.79	416.47	0	60	70	80	100
	Brand Image	65.95	655.64	0	50	70	80	100
	Promotion=	70.78	593.69	0	60	77.5	90	100

This table shows that the average age of customer in our sample was 30.5 years and that 50% of customers' age is less than 31 years old. As an average, customers have around 6.5 years of experience with the same service provider with 25% of them having more than 9.75 years. The average monthly bill of customers is 20.8 KD (Kuwaiti Dinar) and ranges from 3 to 80 KD with 75% of customer are paying less than or equal to 30 KD. In addition, results shows that customers gave a range from 65 to 73 for the level of satisfaction of the service quality, customer service, brand image and promotion. Some of the customers were extremely satisfied and some of them were not.

Another Summary table is displayed to show the distribution for customers' nationality along with their working status, this is shown in Table 2.

Table 2: Customer Nationality vs. Working Status
Reference: M. Abiad & S. Ionescu

	Not Working	Working	Total
Arab Expats	8.33%	41.67%	50.00%
Kuwaiti	30.00%	7.50%	37.50%
Non-Arab Expats	0.00%	12.50%	12.50%
Total	38.33%	61.67%	100.00%

We can deduce from table 2 that the majority of our sample goes to Arab Expat customers with 50% followed by Kuwaitis with 37.5%. An interesting result obtained from our sample that all of the selected Non-Arab expats are working, this might not reflect the correct fact in Kuwait and this is due to the method of sampling technique used. Again, the aim of this application, is to show how data can be summarized and show an example of the algorithm.

7.2 Data Analysis

After displaying the data, the analysis part is ready to be done. In our paper we are mainly focusing on customer churn analysis using the binary logistic regression model. The assumptions of the model were checked and the model was generated to provide an example of how such analysis can be implemented. The use of the SPSS or any other Statistical software can be helpful in running this model.

The output result should first summarize the significance of the model, followed by the R-Square value, then by the classification table before the model coefficients. At first, we recall the dependent and independent variables that will be included in the study. These variables were at the beginning of this section.

Starting with Model significant results, Table 3. Shows that the model is significant, therefore at least one of the independent variable explain the variation of the dependent variable.

Table 3: Omnibus Tests of Model Coefficients
Reference: M. Abiad & S. Ionescu

		Chi-square	df	Sig.
Step 1	Step	61.823	16	.000
	Block	61.823	16	.000
	Model	61.823	16	.000

Model significant is 0.000, hence our model is good and we proceed with the next result. Table 4. represent the model summary and shows the R2 value for Cox & Snell as well as for Nagelkerke. The results shows that R square is 0.539 which means that 53.9% of the variation in customer churn possibility is explained by the independent variables.

Table 4: Model Summary
Reference: M. Abiad & S. Ionescu

Step	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
1	102.895[a]	.403	.539

Binary Logistic Regression Model output provides a classification table which represent the percentage of error for classifying the groups. The classification is of two parts, false positive and false negative. As an overall percentage of correct prediction, we have it 77.5% as shown in Table 5. Which means that we are 77.5% sure that our prediction is correct.

Table 5: Classification Table
Reference: M. Abiad & S. Ionescu

Observed			Predicted		
			Churn Possibility		Percentage Correct
			No	Yes	
Step 1	Churn Possibility	No	39	14	73.6
		Yes	13	54	80.6
	Overall Percentage				77.5

Finally, the last result of the SPSS output for running the Binary Logistic Regression Model is represented by the Varialbes Coefficient table as shown in Table 6.

Table 6: Variables Coefficient Table
Reference: M. Abiad & S. Ionescu

	B	S.E.	Wald	df	Sig.	Exp(B)
Age	-.032	.055	.351	1	.554	.968
Gender	-1.130	.571	3.917	1	.048	.323
MaritalStatus	1.379	.697	3.917	1	.048	3.971
WorkStatus	.054	.919	.003	1	.953	1.056
ExperiencewithMSP	-.090	.077	1.358	1	.244	.914
MonthlyBill	-.030	.018	2.755	1	.097	.970
ServiceQuality	-.111	.029	14.740	1	.000	.895
CustomerService	.011	.016	.415	1	.520	1.011
BrandImage	-.007	.013	.294	1	.588	.993
Promotion	-.016	.012	1.948	1	.163	.984
Arab	-1.374	.943	2.122	1	.145	.253
NonArab	-.267	1.514	.031	1	.860	.766
Secondary	.187	.861	.047	1	.828	1.206
Masterormore	.828	.870	.906	1	.341	2.289
Ooredoo	.811	.695	1.361	1	.243	2.251
Viva	.656	.698	.883	1	.347	1.927
Constant	11.563	2.849	16.474	1	.000	105135.179

Table 6 shows the result of significance of each variable in the model as well as its coefficient value. As a summary of this result, we can conclude the following:

7.2.1 Significant variables:

Gender, Marital Status, Monthly Bill and Service Quality are considered to be significat at 1%, 5% or 10% level of significance. The binary logistic equation that can be generated from this model will be summarized by the following equation:

Eq. (2) $$y = 11.563 - 1.13 * Gender + 1.379 * Marital\ Status - 0.03 * Monthly\ Bill - 0.111 * Service\ Quality$$

As an interpretation, we can say that Gender has a negative impact on customer churn, that is, Male customers are less likey to leave the service provider than female customers, therefore mobile service providers may need to put more effort to retain female customers. similar results obtained for service quality and hence the more the customer is satisfied from the service quality the less the churn possibility will be for the customer. Marital status has positive significant of Customer churn, which means that Single customers are more likely to leave the company than non-single customers. As for the Monthly bill, this factor is significant at 10% level of significance and has negative impact, that means the higher the monthly bill it is the lower the churn rate will be, this could be for the reason that when customers are paying large amount of monthly bill they could be connected with their mobile service provider with some packages and so there will be less change for them to leave the company and loose such packages.

7.2.2 Non-Significant variables:

The remaining variables in the model, such as the Age, Nationality, Work Status, Brand Image satisfaction, promotion satisfaction as well as the satisfaction from customer service team are not considered to be significant based on our sample of study. Which means that customers are not affected by such variables and hence mobile service providers should target other than these variables when they want to decrease the churn rate they have.

8. CONCLUSION

Customer Relation with the Service Provider becomes the main focus of managers in service companies. When the relationship is in a good condition, customers feel more loyal to the company, this will directly affect the company's profitability. Therefore, companies need to analyze the reason behind customer churn and try to decrease this rate. In this paper, an algorithm is defined for customer churn analysis by using one of the known statistical methodologies, the Binary Logistic Regression Model. The idea of such algorithm is to make it clear for companies, how they can analyze their date and determine the group of customers with higher churn probability and determine what factors plays an important role in such analysis, so that they can focus on these factors and suggest plans to decrease such probability.

The analysis starts with reviewing the customer churn background, then by listing the most important factors that has been listed by scholars, as well as getting the input from experts in the domain, which has a significant effect on customer churn decision. In addition, the algorithm includes the statiscal stages used in data analysis such as Data collection, Data Display and Data analysis are the statistical cycle used in any analysis. The final stage is the algorithm implementation of which the result of the Binary Logistic regression Model interpreted in the correct way.As an application of this algorithm, a case study was taken from Kuwait telecommunication sector and data was analyzed using the statistical software MINITAB and SPSS. A sample of 120 customers were used with a survey of 15 questions that target the factors that affects the customer churn decision. At first from the data display output, around 55% of customers have the possibility to leave the company according to their input, and the result of the Binary Logistic Regression model was significant with around 54% of the variation in customer churn are explained by the factors included in the model. Our results, based on our sample, concluded that Gender, Marital Status, Monthly bill and Service Quality affecting the customer churn in Kuwait while the Age, Educational Level and the rest of the variables did not show a significant effect on customer churn. An equation was generated from this model that computes the log odds of the variable customer churn for the significant independent variable.

The benefit of this paper, is to provide a clear plan of analyzing customer churn using the binary logistic regression model, since the steps needed for such analysis are explained in the algorithm. The study can be enhanced by incluing more factors that might has an important role on customer churn to increase the value of the R-square, i.e to increase the percentage of variation that can be explained by the independent factor. Also, the study has difficulty to make reliable results due to the fact of the small sample size, hence in future study, this paper can be used as the basis of running a customer churn analysis by the telecommunication companies, as they can include more customers in the sample and hence having a representative sample of the market place.

ACKNOWLEDGEMENT

Many thanks to the American University of the Middle East for the academic support they are offering to help us proceed in our research, as well as supporting me in my PHD thesis

REFERENCES

- Adebiyi, S.O., Oyatoye, E.O. and Kuye, O.L.. (2015). An Analytic Hierarchy Process Analysis: Application to Subscriber Retention Decisions in the Nigerian Mobile Telecommunications. *International Journal of Management and Economics*, *48*(1), 63-83.
- Amin, A., Shah, B., Khattak, A. M., Baker, T., & Anwar, S. (2018, July). Just-in-time customer churn prediction: With and without data transformation. In *2018 IEEE Congress on Evolutionary Computation (CEC)* (pp. 1-6). IEEE.
- Angelova, B., & Zekiri, J. (2011). Measuring customer satisfaction with service quality using American Customer Satisfaction Model (ACSI Model). *International Journal of Academic Research İn Business and Social Sciences*, *1*(3), 232.
- Azam, S. F., & Karim, K. S. (2017). Factors Influencing Customers'satisfaction on Bangladeshi Telecommunication Service Providers. *European Journal of Management and Marketing Studies-Volume*, *2*(3), 116.
- Bhuian, S.N., Al-Balushi, M. and Butt, I., 2018. Antecedents and consequences of customer loyalty in Qatar. *Journal for Global Business Advancement*, *11*(1), pp.41-63.

- Core.ecu.edu. (2019). *Binary Logistic Regression with SPSS©* [online] Retrieved from www.core.ecu.edu
- Dalvi, P. K., Khandge, S. K., Deomore, A., Bankar, A., & Kanade, V. A. (2016, March). Analysis of customer churn prediction in telecom industry using decision trees and logistic regression. In *2016 Symposium on Colossal Data Analysis and Networking (CDAN)* (pp. 1-4). IEEE.
- Dahiya, K., & Bhatia, S. (2015, September). Customer churn analysis in telecom industry. In *2015 4th International Conference on Reliability, Infocom Technologies and Optimization (ICRITO)(Trends and Future Directions)* (pp. 1-6). IEEE.
- George Onyango, A., Caleb, K. and Patrick, G., .(2017). Effect of Brand Awareness on Customer Retention in the Mobile Telecommunication Industry in Kenya: A Case of Airtel Kenya.
- Hamelin, N., Nassali, A., & Harcar, T. (2009). Determining churn drivers in moroccan telecom sector. *Research Yearbook*, 573.
- Mahajan, V., Misra, R., & Mahajan, R. (2017). Review on factors affecting customer churn in telecom sector. *International Journal of Data Analysis Techniques and Strategies*, *9*(2), 122-144.
- Mannan, M., Mohiuddin, M.F., Chowdhury, N. and Sarker, P., .(2017). Customer satisfaction, switching intentions, perceived switching costs, and perceived alternative attractiveness in Bangladesh mobile telecommunications market. *South Asian Journal of Business Studies*, *6*(2), 142-160.
- Statistics Solutions. (2020). *Assumptions Of Logistic Regression - Statistics Solutions*. Retrieved from www.statisticssolutions.com
- Sulaimon, O. S., Emmanuel, O. E., & Bilqis, B. B. (2016). Relevant Drivers for Customers Churn and Retention Decision in the Nigerian Mobile Telecommunication Industry. *Journal of Competitiveness*, *8*(3).