**Module 7: Portfolio Project Paper**

**“Final Research Paper”**

Jose Vargas

Colorado State University Global

MIS581: Capstone – Business Intelligence & Data Analytics

Dr. Justin Bateh

February 2, 2024

**ABSTRACT**

The main topic discussed and addressed in this research paper is customer churn within the telecommunications industry. The research comprises literature that highlights an overview of the global telecommunications industry outlook, as well as some of its current challenges, such as customer churn and then dives into the issue, the proposed solution, and a discussion of the benefits of predictive models to counter the customer churn challenge in the industry. Some research questions and their corresponding null and alternate hypotheses revolve around the idea of measuring the influence, or lack thereof that some variables from the primary dataset might have on the target outcome, customer churn. A more detailed analysis is presented by building and running a logistic regression model that uses a combination of customer account and demographic predictor variables to predict the outcome. Methods used to conduct this study include the implementation of SAS Studio software to perform the required descriptive and statistical analysis on relevant variables to test the hypotheses. Findings will show which attributes are good indicators of customers who are at risk of churning and the benefits of leveraging predictive models to identify these. This research is important because it complements existing research and literature with concrete tests and results of predictive models to determine the severity of customer churn within the telecommunications industry and help companies in this sector plan effective customer retention strategies.

**INTRODUCTION**

For my capstone project, I decided to do conduct some research into the telecommunications industry because, regardless of what industry most people are in, telecommunications play a big role in the lives of everyone, from connecting us to our loved ones, to leveraging telecommunications for business and the sharing of all sorts of information. Formally, and according to AAI-Persona.org, telecommunications can be defined as “…the transmission of information through various media such as radio, television, telephone, and the internet. It is a critical component of modern society, enabling people to communicate, share information, and access entertainment from anywhere in the world.” (Persona, Pg. 1). Additionally, another reason why I selected to conduct my research in this industry is because despite of the recent advances in technology and AI, which may have negatively impacted other industries, the telecom industry has placed itself in a position where most business depend heavily on it to conduct their day-to-day business operations successfully and efficiently. In PwC’s most recent global telecom’s outlook, authors state that “The sector provides vital services on which billions of consumers and virtually all businesses rely.” (PwC, Pg. 1). It is evident that the telecom industry is an interesting example and use case for this capstone project as it seems to be one of those few industries that can use current and future business trends as either threats, or opportunities, or both depending on their use of analytics to adjust to these trends and advancements.

**OBJECTIVES**

The primary aim and most concrete product of this research is to generate a predictive model with customer churn in the telecom industry as the target outcome. Some objectives associated to the primary aim are the following:

* Produce business questions related to the research business problem along with their respective null and alternate hypotheses.
* Train and validate predictive model to predict customer churn.
* Pinpoint key elements leading to customer churn.
* Provide insights and recommendations to help telecom companies reduce customer churn.

**OVERVIEW OF STUDY**

Earlier deliverables of this project comprised building a foundation of the project by setting goals and objectives, identifying the required dataset(s), conducting preliminary research, drafting a literature review, and presenting the business questions and corresponding hypotheses. Research conducted for this study involves an outlook of the telecommunications industry, customer churn in the telecommunications industry, and the benefits of predictive modeling to minimize or mitigate customer churn in this industry. The primary software/tool used for this study is SAS Studio where descriptive and statistical analysis is conducted on relevant variables from a dataset that contains historical account and demographic data of customers who have churned or have stayed with their telecommunications provider. A logistic regression model is applied to the data to determine which variables seem to have a more significant impact on the target variable, churn. The results and findings will allow me to contribute to already existing research by providing the variables that could indicate a customer is at risk for churn, supporting or disproving current research, and ultimately determining what a good customer retention strategy could look like for a company in this industry.

**RESEARCH QUESTIONS AND HYPOTHESES**

My project aims to identify those common factors that place customers at risk for churn. In order to do this, a model must be built, and for a model to be built, I first have to select the relevant variables and ensure these variables have a significant effect on the target variable and not on any of the other independent variables. My research questions and hypotheses testing revolve around making sure the variables have some sort of influence on the target variable (churn). Findings from conducting the necessary steps to answer these questions can be used to directly inform telecommunications companies about the magnitude of influence the relevant variables have on customer churn, allowing these companies to identify ahead of time when a customer is at risk and helping them build effective customer retention strategies.

Some questions and their respective alternate and null hypotheses relevant to my research are the following:

1. **How does the type of contract influence customer churn in the telecom industry?**

The answer to this question will allow me to determine whether the type of contract has an influence on customer churn and whether it should be included in the final model as an independent variable.

* **Ho:** There is no significant difference in customer churn between customers with different types of contracts.
* **Ha:** There is a significant difference in customer churn in customers based on the type of contract.

1. **Is there a relationship between the type of internet service (DSL, Fiber Optic) and churn?**

The answer to this question will allow me to identify whether customers with DSL internet service are more likely to churn than customers with Fiber Optic internet service, or vice versa. If there is a significant difference, then it would be useful to include this variable as an independent variable in the model.

* **Ho:** There is no significant difference in churn rate between customers with Fiber Optic internet service, and those with DSL internet service.
* **Ha:** There is a significant difference in churn rate between customers with Fiber Optic internet service, and those with DSL internet service.

1. **Do accounts with multiple lines have lower churn rate than accounts with an individual line?**

The answer to this question will allow me to understand whether the number of lines under a single customer account has any effect on the churn of the customer has who owns that account. I chose this question because I assume switching providers is more time-consuming and troublesome when a customer own multiple lines, therefore I am expecting customers with multiple accounts to have a lower churn rate than customers with a single line.

* **Ho:** There is no significant difference in customer churn between account with one line, and accounts with multiple lines.
* **Ha**: There is a significant difference in customer churn between accounts with one line versus accounts with multiple lines.

**LITERATURE REVIEW**

**Introduction**

Customer churn is a prevalent and growing issue within the telecommunications industry, where companies must face intense competition and at the same time deal with an increasing number of customers discontinuing their services entirely or to switch to a competing provider. Understanding customer churn in this industry and its root causes in addition to being able to build a solution to predict future customer churn is critical and can help companies in this industry identify customers at risk for churn and develop strategies accordingly to retain these customers and reduce their rate of customer churn. The aim of this literature review is to explore and analyze existing research and literature pertaining to this industry, customer churn in the industry, and prediction of customer churn to present a view of what is already known and fill in any potential gaps in the research, if any.

**The Telecommunications Industry**

Telecommunications and the companies that operate within its industry are critical in global communication and in people’s ability to remain connected to their loved ones and to everyday entertainment. The telecommunications industry comprises various sectors, from streaming and internet connectivity, to satellite, wired, and wireless communication. According to an article by Adekunte et al. (2024), the industry has undergone tremendous growth and innovation over recent years, fueled by fast-paced technological advancements and the demand for state-of-the-art infrastructure. The intensely competitive nature of this industry has seen companies in the industry prioritize customer satisfaction and retention in addition to providing the highest quality service. Meena & Geng (2022) echo the innovation and improvements in quality of service that have resulted from this competitive environment, both critical components to retaining competitive advantage. My research as well as results from my predictive model can help support or disprove the theories discussed above. For instance, I can add some sort of customer satisfaction metric as an independent variable in my model and determine how significant this variable is against the outcome variable, customer churn.

**Customer Churn in the Telecommunications Industry**

Customer churn is a serious issue for companies within the telecommunications industry due to its immediate influence on a company’s revenue and profitability. Survey results produced by TechSee (2019), showed that customer churn in the telecommunications industry is most often due to high customer effort, this includes long wait times during calls, having to call the provider more than once, untrained customer service agents, etc. Additionally, survey data also shows that contract termination was a result of poor customer service as a primary reason in nearly 39% Americans reported.

To reduce customer churn in line with insights generated from surveys, companies are focusing on improving customer service and customer experience. TechSee (2019) highlights that reactive retention strategies, such as offering discounts or apologies to customers after canceling their service were ineffective in nearly 61% of occasions. Therefore, authors suggest conducting a more proactive approach as churned customers expressed that they would have continued their service with a provider had the provider offered a better service plan or an alternative to solve the issue, or even if the provider assured customers improvement in customer service. Additionally, CustomerGauge (Tessitore, 2024) emphasizes that understanding and benchmarking customer churn rates against industry standards can help companies plan targeted strategies to improve customer retention.

The research discussed above highlights the ineffectiveness of reactive retention strategies and the success of a proactive approach. However, the research defines the proactive approach as taking action once customers have expressed dissatisfaction. In my research and the goal of my project is to also use a proactive approach through the use of a predictive model, but instead of waiting for dissatisfaction, the model can be built to run based off other variables that can help identify customers at risk of churn before they even have to express dissatisfaction. I believe this alternate proactive approach strategy could yield better results related to customer retention.

**Customer Churn Prediction in the Telecommunications Industry**

Customer churn prediction could drastically facilitate the process of developing effective customer retention strategies. Similar to my customer churn prediction model proposal for this capstone project, Rani et al. (2021) examine the use of logistic regression to predict customer churn in the telecommunications industry. The study shows that logistic regression can effectively identify customers at risk of churning through usage patterns and service quality data, and customer demographic information. This will be a good source to compare the results of my model with as I am currently using customer demographic information, and data related to customers’ accounts and usage patterns.

**Conclusion**

In conclusion, it is evident that one of the major challenges faced by companies within the telecommunications industry is customer churn. Previous research suggests this challenge is fueled by a combination of intense competition and the need to continuously innovate to adjust for new trends and technological advancements. Earlier in this review, I presented and analyzed existing research that highlights key causes of customer churn and the need for predictive models to address this issue. Results from these predictive models, if presented accurately and effectively, could help these organizations with their retention strategies, reducing customer churn. The literature sources discussed earlier in this paper complement my research by providing alternate or supporting views of the industry as a whole and of the issue of customer churn within the industry, such as considering customer dissatisfaction as predictor attribute and focusing on a proactive approach rather than a reactive approach.

**RESEARCH DESIGN**

**Methodology**

The main dataset used in this research is the telecommunications customer churn dataset, which is primarily composed of binary, categorical, numerical data. A quantitative methodology is employed as I am leveraging numeric and binary variables in statistical analysis to build a prediction of customer churn. Data collection, preprocessing, and initial analysis is conducted through SAS Studio, these steps include exploratory descriptive analysis to understand the distribution available attributes, and subsequent creation of the logistic regression model, as well as an interpretation of its results. Descriptive and statistical tests to be conducted are fueled by the research questions generated to either support or reject the null hypothesis.

**Methods**

A high overview of the descriptive and statistical tests I will conduct through SAS Studio to address the *Ho* and *Ha* for each respective hypothesis question is the following:

1. **How does the type of contract influence customer churn in the telecom industry?**

My research method for answering this question descriptively includes creating a bar chart showing, historically, the count of customers who discontinued their service by their type of contract. A descriptive analysis, however, is not sufficient to prove or disprove the *Ho*, which is why I will conduct a One-way ANOVA statistical test to determine whether this variable is a statistically significant predictor of churn, and to identify whether the means of customer churn across different types of contracts are different. If the means of the different types of contracts differ significantly, that will allow me to reject the *Ho*.

1. **Is there a relationship between the type of internet service (DSL, Fiber Optic) and churn?**

My research method for answering this question descriptively includes creating a bar chart showing the percentage of customers who discontinued their service by their type of internet service (if any). A descriptive analysis, however, is not sufficient to prove or disprove the *Ho*, which is why I will conduct a One-way ANOVA statistical test to determine whether this variable is a statistically significant predictor of churn, and to identify whether the means of customer churn across different types of internet service are different. If the means of the different levels of this variable differ significantly, that will allow me to reject the *Ho*.

1. **Do accounts with multiple lines have lower churn rate than accounts with an individual line?**

My research method for answering this question descriptively includes creating a bar chart showing the number of data points for every churned customer across different number of lines per customer account. A descriptive analysis, however, is not sufficient to prove or disprove the *Ho*, which is why I will conduct a One-way ANOVA statistical test to determine whether this variable is a statistically significant predictor of churn, and to identify whether the means of customer churn across number of lines are different for customers with a single line versus a customer with multiple lines. If the means of the different levels of this variable differ significantly, that will allow me to reject the *Ho*

**Limitations**

Some limitations regarding the data being used for this study include the following:

* Very few continuous variables will make the model more complex and more difficult to interpret.
* While the dataset has over 20 attributes, it may not capture all attributes that could influence customer churn including attributes discussed in existing literature that are not available as part of the dataset.
* The model is unable to adjust to market conditions which may be an issue in the future.

**Ethical Considerations**

Some ethical considerations for the type of data that I have chosen for my capstone project is the following:

* **Data Privacy**: Ensuring the dataset does not contain any personally identifiable information to comply with data privacy regulations such as GDPR and CCPA and protecting customer data privacy rights. (EU, 2016) (California State Legislature, 2018)
* **Data Bias:** The dataset was obtained from an online source. Although I am assuming bias will not be an issue since this is a public dataset, I am aware certain demographics may be disproportionately represented over others which could cause inaccurate predictions in the model.

At this stage of the project encompassing deliverables from week one to week five. I have fulfilled the requirements pertaining to providing an introduction which highlights the issue of customer churn within the telecommunications industry and my goal of building a logistic regression model to identify variables that have a greater influence on customer churn potentially helping companies in this industry determine when a customer is at risk of churn and implement effective customer retention strategies. I completed a literature review which leverages several scholastic articles that discuss the outlook of the telecommunications industry, the severity of customer churn in this industry, and the benefits of predictive modeling to address this issue. This is existing research that can either complement or disprove my research depending on the results to be presented in the following week. Lastly, research questions with their respective hypotheses have been presented and a research design plan has been developed and included in this first draft of my capstone project. Over the following weeks, I will add to this paper my findings, a conclusion from my research, and recommendations based on the results.

**FINDINGS**

In order to better understand the variables in my dataset, I conducted initial descriptive analysis specifically on the continuous variables of the dataset since these are the only variables for which computing basic summary statistics makes sense to understand their dispersion and central tendency. A high overview of these variables is displayed below:

**Figure 1**

*SAS Studio; Summary Statistics of Continuous Variables*

A screenshot of a computer

Description automatically generated

Figure 1 shows the results of running a summary statistics task in SAS Studio. The table shows useful metrics to understand the dispersion and central tendency of the continuous variables in my dataset such as mean, median, standard deviation, minimum and maximum. For the “Tenure” variable, for instance, a valid interpretation of these values is the following:

* Customers have been with this provider, on average, for 32.4 months.
* The standard deviation value is relatively close to the median value, indicating significant variability.
* A minimum value of 0 implies that some customers have been with this provider for 0 months which could indicate that these are new customers.
* At least one customer has been with this provider for a maximum of 72 months.
* The median value of 29 indicates that about half of the customers in this dataset have been with this provider for 29 months and the other half have been with this provider for more than 29 months.

A similar interpretation is true for the rest of the continuous variables. More detailed visuals of the distribution of the different continuous variables are shown below:

**Figure 2**

*SAS Studio; Distribution of Tenure*

A graph of a number of numbers

Description automatically generated

**Figure 3**

*SAS Studio; Distribution of MonthlyCharges*

A graph of a distribution of monthly charges

Description automatically generated

**Figure 4**

*SAS Studio; Distribution of TotalCharges*

A graph of a distribution of charge

Description automatically generated

The distributions of the variables “Tenure”, “MonthlyCharges”, and “TotalCharges” do not appear normal (see figures 2-4); they have either a positive skew, negative skew, or both. For example, "TotalCharges" shows that nearly 22% of customers have total charges less than $300, with a long tail up to nearly $9K, reflecting high variability.

**Statistical Analysis/Findings Related to Hypotheses**

The first research question in my study revolves around the influence of a customer’s type of contract on the target variable; churn.

**Figure 5**

*SAS Studio; Churn by Type of Contract*

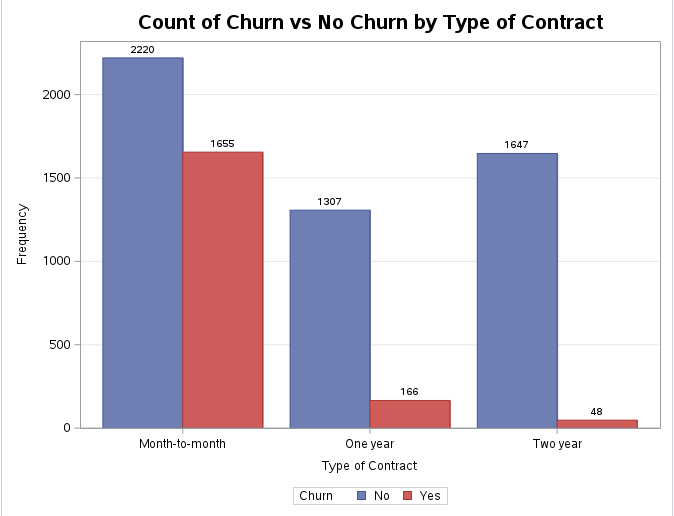
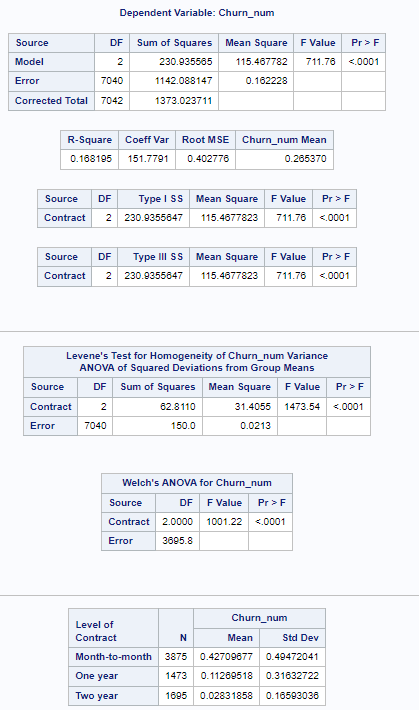
****

Figure 5 shows the comparison of the number of customers that churned based on the type of contract. For instance, I can see that for customers with a month-to-month contract, approximately 1655 of them discontinued service with their provider versus 2220 who did not churn resulting in a 42.7% churn rate (1655/3875), compared to churn rates of 11.3% and 2.8% for one-year contracts and two-year contracts, respectively. This would make sense as short-term contracts such as month-to-month contracts make it easier for customers to end their service as opposed to having to wait a year or two. This analysis, however, does not establish correlation between reasonable doubt which is why I also conducted a one-way ANOVA test to measure the statistical significance of the relationship between these two variables. The results are displayed below:

**Figure 6**

*SAS Studio; One-way ANOVA Results*

****

Generally, when running a One-way ANOVA test, the larger the F-value, the greater the variance in the means of the different variables. Based on the results from figure 6, a high f-value of 711.76 suggests strong evidence of variance between at least one of the means of type of contract and churn. Coupled with a p-value (<.0001) below the specified significance level (.05), this indicates this metric is statistically significant. Therefore, the results suggest that the differences in the group means are highly significant, and it is unlikely that the churn rates as they relate to the type of contract are coincidental. Based on this analysis, I can conclude that the type of contract does in fact have a significant influence in customer churn and provides strong evidence to reject the null hypothesis.

The second research question in my study attempts to determine whether there is a relationship between the type of internet service and customer churn. Descriptive and statistical analyses are displayed below:

**Figure 7**

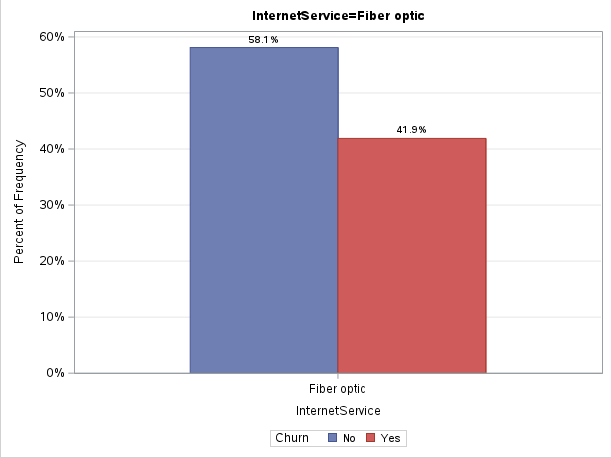
*SAS Studio; Customer Churn by Type of Internet Service - DSL*

A graph of a bar

Description automatically generated with medium confidence

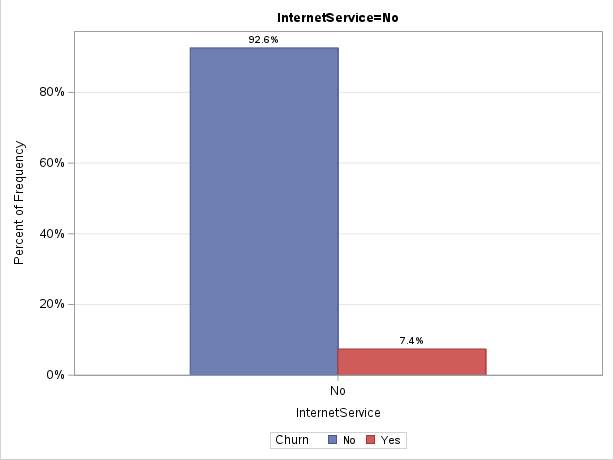
**Figure 8**

*SAS Studio; Customer Churn by Type of Internet Service – Fiber Optic*



**Figure 9**

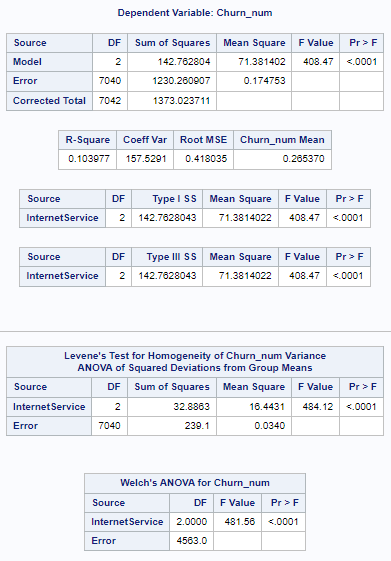
*SAS Studio; Customer Churn by Type of Internet Service – No Internet*



The visuals displaying customer churn (see figures 7-9) show customer churn rates of 19%, 41.9%, and 7.4% for customers with DSL internet service, fiber optic internet service, and no internet service, respectively. Assuming a true relationship between the variables of churn and internet service, it is evident that customers with Fiber Optic internet service are significantly more likely to churn than those with no internet service or DSL internet service. To confirm or disprove this relationship beyond random chance, I conducted a One-way ANOVA test to measure the significance of the relationship between these two variables.

**Figure 10**

*SAS Studio; One-way ANOVA Results*

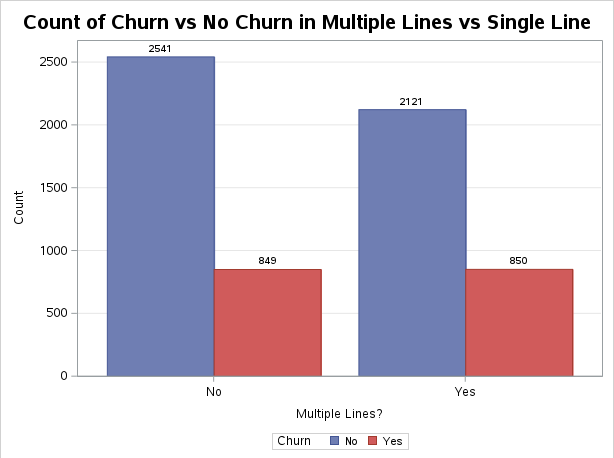
**

Similar to the analysis of the first research question, the high f-value of 408.47 between internet service and customer churn indicates strong variance between the group means. A p-value of <.0001, which is below the specified significance level of 0.05 confirms the statistical significance of this result. These results suggest that the churn rates as they relate to the type of internet service in a customer’s account are not due to random chance and provide strong evidence against the null hypothesis.

The last research question attempts to determine whether the number of lines in a customer account has an impact on customer churn. Descriptive analysis is displayed below, followed by its statistical analysis.

**Figure 11**

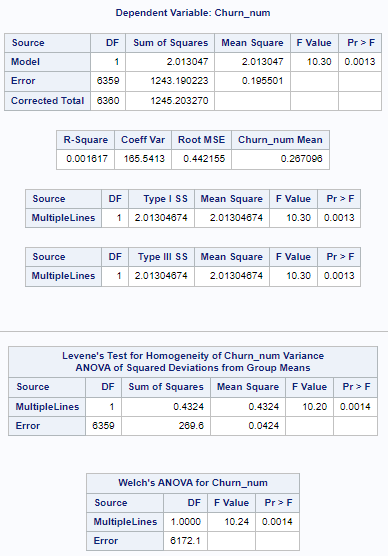
*SAS Studio; Customer Churn by Number of Lines*



The bar chart in figure 11 shows a customer churn rate of 25% (849/3390) for customers with a single line against a churn rate of approximately 28.6% for customers with multiple lines. The visual suggests that customers with multiple lines in their account are slightly more likely to churn. Statistical analysis to complement or counter this assumption is displayed below.

**Figure 12**

*SAS Studio; One-way ANOVA Results*

**

Unlike the results from the previous research questions, the f-value (10.30) seems relatively low, indicating that there is not significant variability between the group means. However, its corresponding p-value (0.0013) still falls below the significance level (0.05) which suggests that even though the variability is not significant, the relationship between these two variables is still true and the churn rate is not due to random chance. Therefore, I still have strong enough evidence to reject the null hypothesis and conclude that customers with multiple lines are in fact slightly more likely to churn than customers with a single line in their account.

**Predictive Analysis: Logistic Regression**

The end goal of this research paper is to build a linear regression model that uses relevant variables including those discussed earlier in this paper as predictor variables to predict the target outcome, churn. A fast backward selection was used in this model in order to run a model that uses all of the specified variables and eliminates insignificant variables within the first iteration instead of removing a single variable during every step. Results of the fast backward elimination logistic regression model are displayed below:

**Figure 13**

*SAS Studio; Initial Step of Fast Backward Elimination*

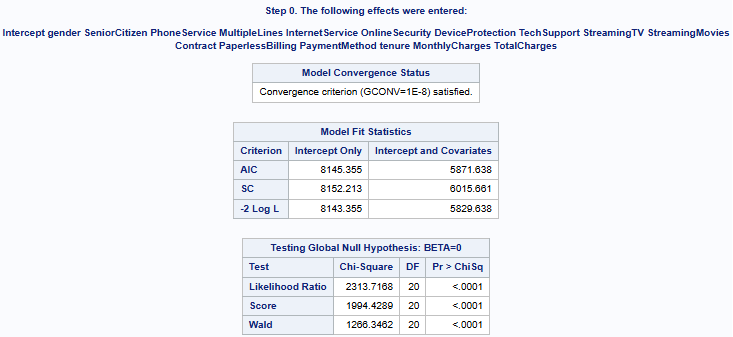


Figure 13 shows the results of the first step after running the model in which a full model including the intercept and selected variables is executed. The model fit statistics table provides several metrics that measure the fit of the data for a model containing only the intercept versus a model containing the intercept and the selected predictor variables. Akaike’s Information Criterion (AIC), Schwarz Criterion (BIC or SC) and the Likelihood Ratio (-2 Log L) are all metrics that measure goodness of fit. Generally, the lower the value for all of these metrics indicates a better fitting model. Looking at the table, a model that consists of the intercept and all of the predictor variables generates lower AIC, BIC, and -2 Log L values at 5871.638, 6015.661, 5829.638, respectively against their corresponding intercept-only model values, suggesting that a model that contains intercept and covariates provides a better fit for the data.

The bottom table of the visual tests whether the regression coefficients of the predictor variables are equal to 0, this would indicate that the predictors are not statistically significant. The chi-square values for all of the tests are significantly high and have a corresponding p-value of <.0001 which suggests that all, if not most, of the predictor variables generate a model with better predictive performance than a model that only contains the intercept.

Overall, the results from the initial step indicate that the predictor variables used in the model can explain the variability in churn.

**Figure 14**

*SAS Studio; Fast Backward Elimination – Step 1*

*A screenshot of a computer

Description automatically generated*

Figure 14 shows a summary of effects (variables) removed from the model after determining that they are not statistically significant, and their removal will not harm the performance of the model. Given a significance of 0.05, pr > Chi Sq values of .7330, .4893, .4216, and .0664 for the variables gender, device protection, phone service, and dependents, respectively suggest that removing these variables will not have a negative impact on the model. As a result, these variables are removed in the backward elimination process. It is important to note that the residual chi-square value increases post removal of each variable, indicating an improvement in model fit*.*

Like the analysis in the initial step, the model fit statistics table shows lower values for AIC, BIC and -2 Log L in a model that includes intercept and covariates, indicating better fit. Furthermore, the high chi-square values with corresponding <.0001 p-value suggest that the remaining predictors contribute significantly to the model.

**Figure 15**

*SAS Studio; Fast Backward Elimination – Step 1 (continued)*

*A screenshot of a computer

Description automatically generated*

The metrics shown in the tables in figure 15 focus on the model’s predictive performance. In the first table, for instance, the results show that the model produced correct predictions 84.8% of the time and incorrect predictions 15.2% of the time. The “c” score, which is equal to the percent concordant, is a value of .848 out of 1 measuring the model’s ability to predict the correct outcome, the higher the “c” score, the better.

In the residual Chi-square test table, the p-value of .1773 is greater than the significance level of .05, this confirms the model fits the data well. Lastly, the last table provides a summary of the variables that can be removed from the model because they don’t necessarily improve the predictive performance of the model.

**CONCLUSION**

I started this research project by giving a high overview of the telecommunications industry and what seems to be its most critical problem in today’s world due to competition and technological advancements, customer churn. The study aims to investigate the industry, customer churn within the industry, and the potential benefits of implementing predictive solutions for a company in this industry to be able to proactively identify customers at risk of churn based on customer account and demographic information.

Some key objectives of this project include producing business questions related to the business problem along with their respective null and alternate hypotheses, training a predictive model to predict customer churn, pinpointing key elements leading to customer churn, and providing insights and recommendations to help telecom companies reduce customer churn.

The study starts by first identifying important questions related to the relationship between selected predictors and their influence on customer churn through hypothesis questions. Research and hypotheses are followed by a literature review in which I discussed and analyzed existing literature pertaining to the industry, the implications of customer churn in the industry, and the benefits of predictive modeling to mitigate this issue. Next, I established a research design in which I outlined the methodology, methods, limitations, and ethical considerations of my research and the analysis conducted to address the research topic.

Lastly, I presented my findings derived from the results of conducting descriptive, statistical, and predictive analysis on the telecommunications customer churn dataset through SAS Studio. Descriptive analysis comprised visuals to show the distribution of continuous variables and diagrams to picture the relationship of predictor variables against the target variable. Statistical analysis such as one-way ANOVA helped address the null hypothesis presented by the research questions by measuring the statistical significance of the relationship, or lack thereof, between key predictor variables and the target variable. Predictive analysis consisted of building a logistic regression model with a fast backward selection to identify which variables are statistically relevant in explaining a variability in customer churn versus those that are not and can be removed from the model to improve predictive performance.

The results provide strong evidence that my model is significantly better at predicting customer churn with the variables I selected than a model with no predictors aside from suggesting the removal of four of the variables which show minimal contribution to the predictive capability of the model.

**RECOMMENDATIONS**

In terms of business processes and practices and in support of the positive results of my model, I suggest companies in this industry leverage predictive models to determine which customer attributes have an influence in customer churn. Identifying these attributes with anticipation can help companies identify customers who are more likely to churn and design more effective customer retention strategies and incentives tailored to these specific customers. For instance, based on the results, a company can offer discounts to customers with Fiber Optic internet service as these have shown to be more likely to churn. Another example could be to implement a cancellation fee to the month-to-month contract to keep those customers from churning even though that sounds like a more aggressive solution.

Regarding the data collected, my research specifically showed that some of the variables are not relevant to build an effective model, such variables can be removed from data collection to save time and space. Contrastingly, to build a more robust model, other key attributes can be added to data collection to include more detailed information about customers, such as education level, income, economic class, geospatial data, age, marital status, number of previous providers, etc.

**REFERENCES**

Adekunte, P. A., Sadiku, J. O., & Sadiku, M. N. O. (2024, November 20). *Telecommunications*

*industry: An overview*. International Journal of Trend in Scientific Research and Development. https://www.ijtsrd.com/engineering/telecommunications/71624/

telecommunications-industry-an-overview/matthew-n-o-sadiku

BlastChar. (2018, February 23). *Telco customer churn*. Kaggle.

https://www.kaggle.com/datasets/blastchar/telco-customer-churn

Kumar, S. (n.d.). *Perspectives from the Global Telecom Outlook 2023–2027*. PwC.

https://www.pwc.com/gx/en/industries/tmt/assets/pwc-gto-2023.pdf

Meena, M., & Geng, J. (2022, April 26). *Dynamic Competition in Telecommunications: A*

*systematic Literature Review*. Sage Journals.

https://journals.sagepub.com/doi/10.1177/21582440221094609

Rani, K. S., Thalisma, S., Prasanna, N. G. L., Vindhya, R., & Srilakshmi, P. (2021, September

8). *Analysis of Customer Churn Prediction in Telecom Industry Using Logistic Regression*. SSRN. https://papers.ssrn.com/sol3/papers.cfm?abstract\_id=3902033

*Reasons for customer churn in Telecoms [survey results]*. TechSee. (2019).

<https://techsee.com/resources/reports/2019-telecom-churn-survey/>

*SAS ondemand for academics*. SAS OnDemand for Academics. (n.d.).

https://welcome.oda.sas.com/

Tessitore, S. (2024, April 18). *What’s the average churn rate by industry?* CustomerGauge.

https://customergauge.com/blog/average-churn-rate-by-industry

Understanding the role of telecommunications in the modern world. Advancing Communication

Technology for a Connected World. (2024, December 2). <https://www.aal>

persona.org/understanding-the-role-of-telecommunications-in-the-modern-world/

Wagh, S., Andhale, A., Wagh, K., Pansare, J., & Ambadekar, S. (2023, November 13). *Customer*

*churn prediction in telecom sector using Machine Learning Techniques*. Results in

Control and Optimization.https://www.sciencedirect.com/science/article/

pii/S2666720723001443