Project Report

on

Subscription-based Service Churn Prediction

by

Divya Venkatrao Pullivarthi

Presented To:

Prof. Nandhini Gulasingam

DSC 423: Data Analysis and Regression

9th June 2025

Table of Contents

Introduction **3**

Data Description **4**

Dummy Variables **5**

Data Exploration **7**

Boxplots 7

Descriptives 9

Data Analysis **11**

Full Model 11

Multicollinearity 12

Outliers 12

Influential Points 12

Standardized Estimate (STB) 12

Data Splitting **13**

Final Model **13**

Multicollinearity 14

Outliers 14

Influential Points 14

Standardized Estimate (STB) 14

Model Equation **15**

Equation Analysis 15

Classification Table **16**

Test Analysis **16**

Confusion Matrix 16

Predictions 17

Conclusion and Recommendation **18**

Appendix **19**

Subscription-based Service Churn Prediction

Introduction

This report represents insights from a logistic regression analysis conducted to examine the relationship between Customer Churn and various factors that can affect churn.

Customer Churn prediction is a crucial factor for an organization that provides services based on subscriptions. It helps the organization to proactively create and communicate retention plans to the customers who are predicted as someone who are likely to churn, i.e., they can leave the subscription or may switch to new service providers.

This way, the factors that are affecting churn such as Account Age, monthly charges for the type of subscription, followed by various kinds of usage of the subscription service like average duration of the watch, content types like TV Shows, Movies, or both, user rating, how many support tickets were submitted (how many times a customer faced issues and submitted a ticket), etc. can be understood.

By understanding the factors affecting churn, policymakers and organization staff can develop retention plans such as personalized offers, improved customer support, or tailored content recommendations to improve customer satisfaction so that customers continue to use their services. This way, long-term profitability could be maintained.

The research question to focus on is:

What is the likelihood of churn for a specific customer segment or profile? Which customer attributes or behaviors are most strongly associated with churn?

To address this question, a logistic regression analysis was conducted using a dataset with around 2000+ observations and around 20 predictors. These predictors involve customer demographic information and behavioral usage patterns like account age, monthly charges, subscription type, payment method, device registration, content preferences, support interactions, etc.

By observing the relationship between these variables and churn, the analysis aims to determine which factors strongly affect a customer's likelihood of discontinuing the subscription.

The hypotheses are as follows:

H_o (Null Hypothesis): $\beta_1 = \beta_2 = \beta_3 = \ldots = \beta_k = 0$

There is no significant predictor associated with customer churn.

H_a (Alternative Hypothesis): At least one $\beta_k \neq 0$

There is at least one significant predictor associated with customer churn.

We will either accept or reject the null hypothesis based on the insights gained from the analysis.

Data Description

Data Source Link: Kaggle - Subscription-based Service Churn Prediction

Number of Observations: 2226 observations after modifying the dataset.

Dependent Variable: Churn is binary (churn=1, no churn=0)

Independent Variables:

Variable	Description	Type
AccountAge	Age of the customer's subscription account (in months)	Number
MonthlyCharges	Monthly subscription charges	Number
TotalCharges	Total charges incurred by the customer	Number
SubscriptionType	Type of subscription plan chosen by the customer (e.g., Basic, Premium, Deluxe)	Text
PaymentMethod	Method used for payment (e.g., Credit Card, Electronic Check, PayPal)	Text
PaperlessBilling	Whether the customer uses paperless billing (Yes/No)	Text
ContentType	Type of content accessed by the customer (e.g., Movies, TV Shows, Documentaries)	Text
MultiDeviceAccess	Whether the customer has access on multiple devices (Yes/No)	Text
DeviceRegistered	Device registered by the customer (e.g., Smartphone, Smart TV, Laptop)	Text
ViewingHoursPerWeek	Average number of viewing hours per week	Number
AverageViewingDuration	Average duration of each viewing session	Number
ContentDownloadsPerMonth	Number of content downloads per month	Number

Variable	Description	Type
GenrePreference	Genre preference of the customer (e.g., Action, Drama, Comedy)	Text
UserRating	Customer satisfaction rating (1 to 5)	Number
SupportTicketsPerMonth	Number of customer support tickets raised per month	Number
Gender	Gender of the customer (Male/Female)	Text
WatchlistSize	Size of the customer's content watchlist	Number
ParentalControl	Whether parental control is enabled (Yes/No)	Text
SubtitlesEnabled	Whether subtitles are enabled (Yes/No)	Text
CustomerID	Unique identifier for each customer	Text

Dummy Variables

Dummy Variables were created by converting all categorical predictors.

1. **Subscription Type:**
   Reference: Basic
   Dummy Variables:
   - d_subtype2 = (SubscriptionType='Standard')
   - d_subtype3 = (SubscriptionType='Premium')

2. **Payment Method:**
   Reference: Mailed check
   Dummy Variables:
   - d_pay1 = (PaymentMethod='Bank transfer')
   - d_pay2 = (PaymentMethod='Credit card')
   - d_pay3 = (PaymentMethod='Electronic check')

3. **Content Type:**
   Reference: Both
   Dummy Variables:

- d_content1 = (ContentType='TV Shows')
- d_content2 = (ContentType='Movies')

4. **DeviceRegistered:**
   Reference: Tablet
   Dummy Variables:
   - d_device1 = (DeviceRegistered='Computer')
   - d_device2 = (DeviceRegistered='Mobile')
   - d_device3 = (DeviceRegistered='TV')

5. **GenrePreference:**
   Reference: Sci-fi
   Dummy Variables:
   - d_genre1 = (GenrePreference='Action')
   - d_genre2 = (GenrePreference='Comedy')
   - d_genre3 = (GenrePreference='Drama')
   - d_genre4 = (GenrePreference='Fantasy')

6. **Paperlessbilling:**
   d_paperlessbilling =Paperlessbilling 1 if Yes, 0 if No.

7. **MultiDeviceAccess:**
   d_MultiDeviceAccess = MultiDeviceAccess, 1 if Yes, 0 if No.

8. **Gender:**
   d_gender= gender, 1 if Female, 0 if Male.

9. **ParentalControl:**
   d_ParentalControl= ParentalControl is 1 if Yes, 0 if No.

10. **SubtitlesEnabled:**
    d_SubtitlesEnabled= SubtitlesEnabled is 1 if Yes, 0 if No.

Data Exploration

Boxplots:

Boxplot (B1):

By observing the boxplot for account age vs. churn from ***Appendix 1***, the diamond that represents the mean and the line that represents the median are quite close to each other in both churn groups, suggesting that the distribution of account age is approximately symmetric for both churn values.

There are no points present outside the whiskers in either the churn or non-churn groups, indicating that no outliers are present.

The minimum and maximum account age are the same for both groups. The account age ranges from 1 month to 120 months in the boxplot for the non-churn and churn groups.

For non-churn customers (churn = 0), the boxplot shows a wider range and a higher IQR. The IQR is larger for non-churn customers compared to churn customers, indicating greater age variability among customers who did not churn.

In contrast, for churned customers (churn = 1), the account age range and IQR are much narrower. The median and mean for non-churn customers are around 65, while for churn customers, they are around 45, suggesting that older customers with longer account histories tend to remain with the service, while customers with shorter account histories tend to churn.

Boxplot (B2):

By observing the boxplot for user rating vs. churn from ***Appendix 2***, the diamond that represents the mean and the line that represents the median are quite close to each other in both churn groups, suggesting that the distribution of user rating is approximately symmetric for both churn values.

There are no points present outside the whiskers in either the churn or non-churn groups, indicating that no outliers are present.

The minimum and maximum user ratings are the same for both groups. The user rating ranges from 1 to 5 in both groups.

For both groups, the overall spread and the IQR are similar, suggesting that the variability in user ratings is comparable for both churn groups.

The median and mean for user rating in the churn group are very slightly higher than the median and mean for the non-churn group, but the difference is minimal, and they can be considered almost similar (~3). This suggests that user rating alone is not a strong predictor of showing the difference between customers who churn and customers who do not.

Boxplot (B3):

By observing the boxplot for monthly charges vs. churn from ***Appendix 3***, the diamond that represents the mean and the line that represents the median are quite close to each other in both churn groups, suggesting that the distribution of monthly charges is approximately symmetric for both churn values.

There are no points present outside the whiskers in either the churn or non-churn groups, indicating that no outliers are present.

The minimum and maximum monthly charges are the same for both groups. The monthly charges range from $5 to $20 per month in both groups.

For both groups, the overall spread and the IQR are similar, suggesting that the variability in monthly charges is comparable for both churn groups.

The median and mean for monthly charges in the churn group are slightly higher than the median and mean for the non-churn group, suggesting that customers who are paying higher monthly charges may be on the margin of churning, but the difference is not significant

Boxplot (B4):

By observing the boxplot for support tickets per month vs. churn from ***Appendix 4***, the diamond that represents the mean and the line that represents the median are quite close to each other in both churn groups, suggesting that the distribution of support tickets per month is approximately symmetric for both churn values.

There are no points present outside the whiskers in either the churn or non-churn groups, indicating that no outliers are present.

The minimum and maximum support tickets per month are the same for both groups. The support tickets per month range from 0 to 9 tickets per month in both groups.

For both groups, the overall spread and the IQR are similar, suggesting that the variability in support tickets per month is comparable for both churn groups.

The median and mean for support tickets per month in the churn group are slightly higher than the median and mean for the non-churn group, suggesting that customers who submit higher numbers of support tickets may be on the margin of churning, but the difference is not significant.

The median and mean (~5) for support tickets per month in the churn group are slightly higher than the median and mean (~4) for the non-churn group, but the difference is minimal, suggesting that customers who raise support tickets per month may be on the margin of churning, but the difference is not significant.

Descriptives:

For Descriptive analysis from ***Appendix 5***, we computed a Five-number summary (i.e., Minimum, 25th Percentile (Q1), 50th Percentile (Median), 75th Percentile (Q3), and Maximum), along with the mean and mode. We also included the 99th percentile to compare values between the 99th percentile and maximum to check for any unusual values or patterns.

1. ***Churn:***
   The churn variable is binary, and hence we are not gaining any insights from it except that the mean is 56.29%, and the mode is 1, which means that 56.29% of customers have churned.

2. ***Account Age:***
   Customer account age ranges from 1 to 119 months, with a mean of 54.31 months and a median of 52.5 months. The difference between the 99th percentile (i.e., 117 months) and the maximum (i.e., 119 months) confirms that the data distribution is symmetric and that there are no unusual values, patterns, or outliers present.

3. ***Monthly Charges:***
   Monthly charges paid by customers on a monthly basis range from $4.99 to $19.82, with a mean of $13.06 and a median of $13.39. The difference between the 99th percentile (i.e., $19.29) and the maximum (i.e., $19.98) confirms that the data distribution is symmetric and that there are no unusual values, patterns, or outliers present.

4. ***Total Charges:***
   The total amount paid by customers ranges from $5.01 to $2534.76, with a mean of $709.82 and a median of $596.98. The difference between the 99th percentile (i.e., $2080.55) and the maximum (i.e., $2354.72) confirms that the data distribution is symmetric and that there are no unusual values, patterns, or outliers present due to long-term customers rather than anomalies.

5. ***Viewing Hours Per Week:***
   Weekly viewing hours range from 1.01 hours to 39.99 hours, with a mean of 18.96 hours and a median of 18.06 hours. The difference between the 99th percentile (i.e., 39.35 hours) and the maximum (i.e., 39.99 hours) confirms that the data distribution is symmetric and that there are no unusual values, patterns, or outliers present.

6. ***Average Viewing Duration:***

   Average viewing duration for each session ranges from 5 minutes to 179.97 minutes, with a mean of 84.85 minutes and a median of 80.22 minutes. The difference between the 99th percentile (i.e., 177.80 minutes) and the maximum (i.e., 179.97 minutes) confirms that the data distribution is symmetric and that there are no unusual values, patterns, or outliers present.

7. ***Content Downloads Per Month:***

   Content downloaded on a monthly basis range from 0 to 49, with a mean of 22.87 and a median of 22. The difference between the 99th percentile (i.e., 49) and the maximum (which is also 49) confirms that the data distribution is symmetric and that there are no unusual values, patterns, or outliers present.

8. ***User Rating:***

   Ratings given by users for the service range from 1 to 5, with a mean of 3.04 and a median of 3.07. The difference between the 99th percentile (i.e., 4.97) and the maximum (i.e., 4.99) confirms that the data distribution is symmetric and that there are no unusual values, patterns, or outliers present.

9. ***Support Tickets Per Month:***

   Support tickets raised by customers per month range from 0 to 9, with a mean of 4.63 and a median of 5. The difference between the 99th percentile (i.e., 9) and the maximum (which is also 9) confirms that the data distribution is symmetric and that there are no unusual values, patterns, or outliers present.

10. ***Watchlist Size:***

    The watchlist size of customers ranges from 0 to 24 items, with a mean of 12.16 and a median of 12. The difference between the 99th percentile (i.e., 24) and the maximum (which is also 24) confirms that the data distribution is symmetric and that there are no unusual values, patterns, or outliers present.

11. ***Dummy Variables:***

    All dummy variables (d_subtype2, d_content1, d_pay2, etc.) are binary (0 or 1). Since they are binary, their descriptives do not reflect any insights that could be used, except for the mean, which suggests how each predictor divides the data. For example, d_ParentalControl has a mean of 0.49, suggesting that half of the users use the parental control feature while the other half do not.

Is the number of observations sufficient? Will the model be balanced?

A rule for a model to be stable and avoid overfitting is that each predictor should have at least 30 observations. We have a total of 28 predictors to work with now. So, we need 30 observations per predictor, which means we need at least 840 observations to ensure model stability and avoid overfitting. Since we have 2000+ observations, we can say the model would be balanced based on this.

Data Analysis

Full Model:

Upon running the full model, we can say from ***Appendix 6*** that the R^2 is 0.1809, i.e., 18.09%. So 18.09% of the outcome of churn could be explained by this full model, which includes predictors such as AccountAge, MonthlyCharges, TotalCharges, ViewingHoursPerWeek, AverageViewingDuration, ContentDownloadsPerMonth, UserRating, SupportTicketsPerMonth, WatchlistSize, d_subtype2, d_subtype3, d_pay1, d_pay2, d_pay3, d_content1, d_content2, d_device1, d_device2, d_device3, d_genre1, d_genre2, d_genre3, d_genre4, d_paperlessbilling, d_MultiDeviceAccess, d_gender, d_ParentalControl, and d_SubtitlesEnabled. It is below-average as a model, but still modest enough. Meanwhile, 81.91% of the outcome of churn remains unexplained by this model.

The AIC and SC values are low, i.e., 2664.466 and 2829.997, respectively, when compared to a model that includes only an intercept predictor. Since adding predictors reduces AIC and SC in comparison to a model with only an intercept predictor, this suggests improvement.

The Likelihood ratio is 444.1114, and the p-value of the likelihood ratio is <0.0001, which is less than 0.05. Thus, we reject the null hypothesis and state that at least one predictor in the model is significant. Since the likelihood ratio is 444.1114, we can say that the model is modest.

We can say that this model is balanced in another way as well. Specifically, if we take the ratio between the number of churn observations and the number of no-churn observations, and if the ratio is 1:1 or close to 1:1, we can conclude that the model is balanced. For this model, the number of churn observations is 1253, while the number of no-churn observations is 973. Calculating the ratio 1253/973, the result is 1:1.29, which is close to 1:1, so we can say this model would be stable.

From ***Appendix 7***, we can see that the full model has some predictors that are insignificant, meaning they have a p-value greater than 0.05. This issue will be further addressed in the final model selection.

Multicollinearity:

When we look at the Estimated Correlation Matrix from ***Appendix 8***, we can see that there is only one multicollinearity issue, i.e., between the AccountAge and TotalCharges predictors. They have a correlation value of 0.9479, so if the correlation value is greater than the absolute value of 0.9, i.e., |0.9|, then multicollinearity exists.

So, to solve this issue, the TotalCharges predictor was dropped because if we look at the Standardized Estimate from ***Appendix 7***, we can see that AccountAge is a stronger predictor than TotalCharges, and we don't need the TotalCharges predictor since it is the multiplication of AccountAge and MonthlyCharges.

Outliers:

We are looking at the Deviance Residuals from ***Appendix 9***. There are no data points that fall outside the range of +3 and -3, and so there are no outliers present in this full model.

Influential Points:

From ***Appendix 10 & 11***, the threshold for identifying influential points in this case is 2/√2226 = 0.0423. So, all the absolute DFBETA values greater than 0.0423 would be considered as influential points.

|DFBETA| > 0.0423 are influential points.

Can we ignore these influential points?

Looking at the descriptive output from ***Appendix 5***, we can see the data distribution is within a reasonable range for numerical predictors, as we can compare the 99th percentile with the maximum value.

Since the data points are distributed within a reasonable range and have no extreme values, patterns, or outliers, we can ignore these influential points. So, since the numerical data is reasonably well-distributed without any unusual patterns or values, we can choose to ignore the influential points as they are not distorting the model.

All the dummy variables are binary, having values 0 or 1. Their descriptives are of no particular use since nothing can be inferred from them. Since the values are fixed (0 or 1), we cannot determine if the data points are influential or not based on dummy predictors. Hence, these cannot be used to check for influential points. So overall, we would ignore these influential points.

Standardized Estimate (STB):

The absolute value of standardized estimates from ***Appendix 7*** is used to determine which predictor is strongest and which is weakest.

The strongest predictor in the full model is AccountAge (0.2907).

The weakest predictor is d_content1 (0.00090).

The predictors ranked from strongest to weakest in descending order are:

AccountAge, AverageViewingDuration, ContentDownloadsPerMonth, MonthlyCharges, ViewingHoursPerWeek, SupportTicketsPerMonth, d_pay2, d_subtype2, d_genre4, UserRating, TotalCharges, d_genre3, d_subtype3, d_pay1, d_genre2, WatchlistSize, d_ParentalControl, d_MultiDeviceAccess, d_content2, d_pay3, d_paperlessbilling, d_device1, d_SubtitlesEnabled, d_device2, d_genre1, d_device3, d_gender, d_content1.

Data Splitting

As shown in the Full Model Diagnostics output from ***Appendix 12***, we are looking at the sampling rate, and we can see that the dataset is split into a 70/30 ratio. 70% of observations from the dataset form our training set, while 30% of observations from the dataset form our testing set.

Final Model:

Upon running the final model selection using the forward method, we saw that 12 steps were performed to arrive at the final forward model, as inferred from ***Appendix 13***. The R^2 is 0.1654, i.e., 16.54%. Thus, 16.54% of the outcome of churn could be explained by this final model, which includes predictors such as AccountAge, MonthlyCharges, ViewingHoursPerWeek, AverageViewingDuration, ContentDownloadsPerMonth, UserRating, SupportTicketsPerMonth, d_subtype2, d_pay3, d_device3, d_genre2, and d_ParentalControl. While below average as a model, it is still modest enough. Meanwhile, 83.46% of the outcome of churn remains unexplained by this model.

The model contains a total of 12 predictors, so k = 12.

The AIC and SC values are low, i.e., 1883.316 and 1952.899, respectively, when compared to the full model, which had an AIC of 2664.466 and an SC of 2829.997.

The Likelihood ratio is 281.9119, and the p-value for the likelihood ratio is <0.0001, which is less than 0.05. Thus, we have significant predictors in this model, confirming that it is a modest model.

From ***Appendix 14***, we can see that the final model contains only significant predictors, meaning they all have a p-value less than 0.05, except for d_pay3, which has a p-value of 0.0501, just at the threshold. Since d_pay3 is a relevant predictor to have in the model, it was retained in this final model.

Multicollinearity:

When we look at the Estimated Correlation Matrix from ***Appendix 15***, we can see that there is no multicollinearity issue present in the final model. All of the predictors in the final model have a correlation value that is smaller than the absolute threshold of 0.9, i.e., |0.9|. Thus, no multicollinearity is present in the final model.

Outliers:

We are looking at the Deviance Residuals from ***Appendix 16***. There are no data points that fall outside the range of +3 and -3, and so there are no outliers present in this full model.

Influential Points:

From ***Appendix 17***, the threshold for identifying influential points in this case is $2/\sqrt{2226} = 0.0423$. So, all the absolute DFBETA values greater than 0.0423 would be considered as influential points.

|DFBETA| > 0.0423 are influential points.

Can we ignore these influential points?

Looking at the descriptive output from ***Appendix 5***, we can see the data distribution is within a reasonable range for numerical predictors, as we can compare the 99th percentile with the maximum value.

Since the data points are distributed within a reasonable range and have no extreme values, patterns, or outliers, we can ignore these influential points. So, since the numerical data is reasonably well-distributed without any unusual patterns or values, we can choose to ignore the influential points as they are not distorting the model.

All the dummy variables are binary, having values 0 or 1. Their descriptives are of no particular use since nothing can be inferred from them. Since the values are fixed (0 or 1), we cannot determine if the data points are influential or not based on dummy predictors. Hence, these cannot be used to check for influential points. So overall, we would ignore these influential points.

Standardized Estimate (STB):

The absolute value of standardized estimates from ***Appendix 18*** is used to determine which predictor is strongest and which is weakest.

The strongest predictor in the final model is AccountAge (0.3120).

The weakest predictor is d_pay3 (0.609).

The predictors ranked from strongest to weakest in descending order are:

AccountAge, AverageViewingDuration, ContentDownloadsPerMonth, ViewingHoursPerWeek, MonthlyCharges, SupportTicketsPerMonth, UserRating, d_subtype2, d_genre2, d_ParentalControl, d_device3, d_pay3.

Model Equation

From ***Appendix 18***, the model equation for the final model is:

log(churn=1/churn=0) = 1.424 – 0.016*accountage + 0.065*monthlycharges – 0.029*viewinghoursperweek – 0.008*averageviewingduration - 0.023* contentdownloadspermonth + 0.127*userrating + 0.085*supportticketspermonth – 0.253*d-subtype2 + 0.250*d_pay3 – 0.259*d_device3 + 0.280*d_genre2 – 0.229*d_parentalcontrol

where d_subtype2 = Standard, d-pay3 = Electronic check, d_device3 = TV, d_genre2=Comedy and d_parentalcontrol=Yes

Equation Analysis:

After retransformation of each variable from ***Appendix 18***, the effect on churn is:

- **AccountAge:** If the customer's account age increases by 1 month, then the odds (p/1-p) of churn decrease by 36.61%.

- **MonthlyCharges:** If the customer's monthly charges increase by $1, then the odds (p/1-p) of churn increase by 16.62%.

- **ViewingHoursPerWeek:** If the customer's viewing hours per week increase by 1 hour, then the odds (p/1-p) of churn decrease by 19.54%.

- **AverageViewingDuration:** If the customer's average viewing duration increases by 1 minute, then the odds (p/1-p) of churn decrease by 24.60%.

- **ContentDownloadsPerMonth:** If the customer's content downloads per month increase by 1 download, then the odds (p/1-p) of churn decrease by 19.85%.

- **UserRating:** If the customer's user rating increases by 1, then the odds (p/1-p) of churn increase by 8.48%.

- **SupportTicketsPerMonth:** If the number of support tickets submitted per month increases by 1 ticket, then the odds (p/1-p) of churn increase by 14.56%.

- **d_subtype2:** When the subscription type is Standard, then the odds (p/1-p) of churn decrease by 6.75%.

- **d_pay3:** When the payment method is Electronic Check, then the odds (p/1-p) of churn increase by 6.28%.

- **d_device3:** When the registered device is a TV, then the odds (p/1-p) of churn decrease by 6.35%.

- **d_genre2:** When the genre preference is Comedy, then the odds (p/1-p) of churn increase by 6.66%.

- **d_ParentalControl:** When Parental Control is enabled, i.e., Yes, then the odds (p/1-p) of churn decrease by 6.51%.

Classification Table

From the classification table in ***Appendix 19***, to determine the threshold, we add sensitivity and specificity. After adding 68.9 + 63.6, we get the highest sum value, i.e., 132.5, so its corresponding probability level, which is 0.55, is selected as the threshold for further test analysis.

Test Analysis

Confusion Matrix:

From the Confusion Matrix output in ***Appendix 20***, we get the following insights:

- **TP (True Positive) = 252** (Model correctly predicted “Will Churn.”)
- **TN (True Negative) = 192** (Model correctly predicted “Will Not Churn.”)
- **FP (False Positive) = 94** (Model incorrectly predicted “Will Churn.”)
- **FN (False Negative) = 129** (Model incorrectly predicted “Will Not Churn.”)

- **Sensitivity =** TP / (TP+FN) = 0.661417 **(66.14%)**
- **Specificity =** TN / (TN+FP) = 0.671329 (**67.13%)**
- **Accuracy =** (TP+TN) / (TP+TN+FP+FN) = 0.665667 **(66.57%)**
- **Precision =** TP / (TP+FP) = 0.728324 **(72.83%)**

- **Sensitivity:** Out of the total actual churn cases, 66.14% of them were correctly identified as churn cases.
- **Specificity:** Out of the total actual non-churn cases, 67.13% of them were correctly identified as non-churn cases.
- **Accuracy:** The model accurately identifies 66.57% of the cases.
- **Precision:** 72.83% of the time, we are going to see similar results predicted by the model.

Predictions

Using the following sample data, the churn was predicted as shown in ***Appendix 21***:

Customer Profile: a 48-month subscriber who pays $18.48625748 monthly, has accumulated $886.90 in total charges, uses a basic subscription paid via bank transfer, does not use paperless billing, and watches TV shows on a registered TV device with multi-device access enabled. They watch 27.828635925 hours weekly, averaging 75.787314567 minutes per session, download 18 items per month, prefer drama, rate the service 4.162684247, submit 6 support tickets monthly, identify as male, have a watchlist of 22 items, use parental controls, but do not enable subtitles

Model Predicted: The predicted churn probability for the above customer profile is computed as p = 0.69841 (i.e., 69.84%), which means the customer with the given profile has a 69.84% chance of churning. Based on the model and the data, we are 95% confident that the true probability of churn for a customer with this profile falls between 88.84 % and 112.64%. suggesting targeted retention efforts for this customer.

Why is this customer marked as churn, even if the account age, content downloads, viewing, and average duration are high?

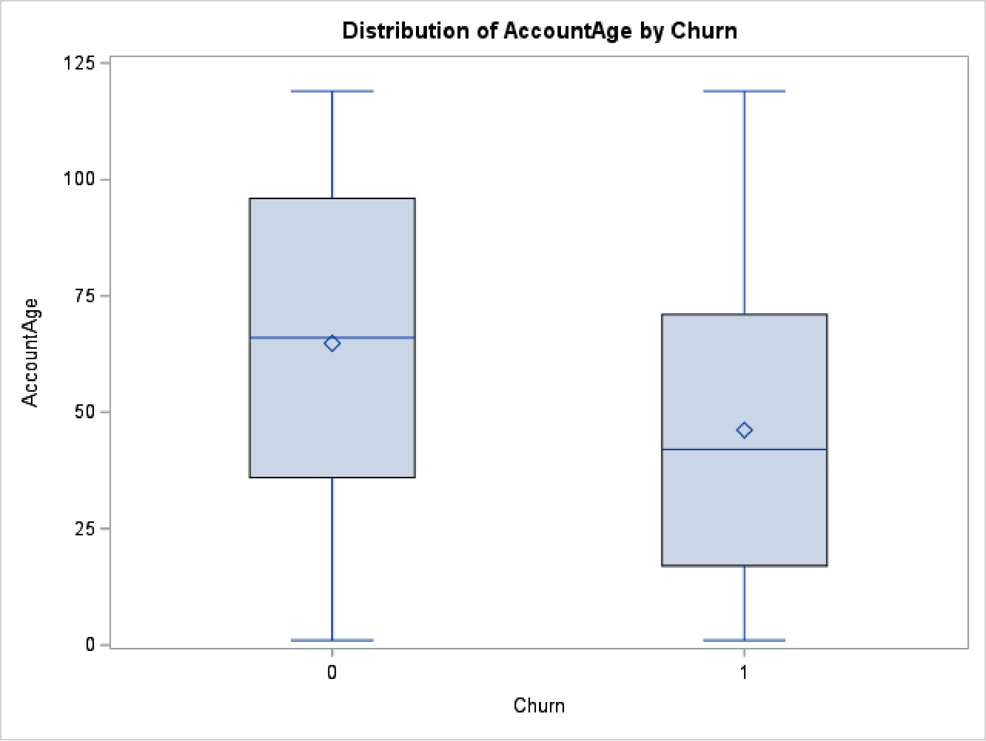
The customer tends to churn because they make electronic check payments, raise multiple support tickets, have a Standard subscription, watch mainly on a TV device, and prefer the Drama genre (all features are associated with higher churn likelihood in the model). Their account age, content downloads, viewing hours, and average session length are relatively high, but that wasn't enough to overcome the combination of other predictors that contributed to the churn risk.

Conclusion and Recommendation

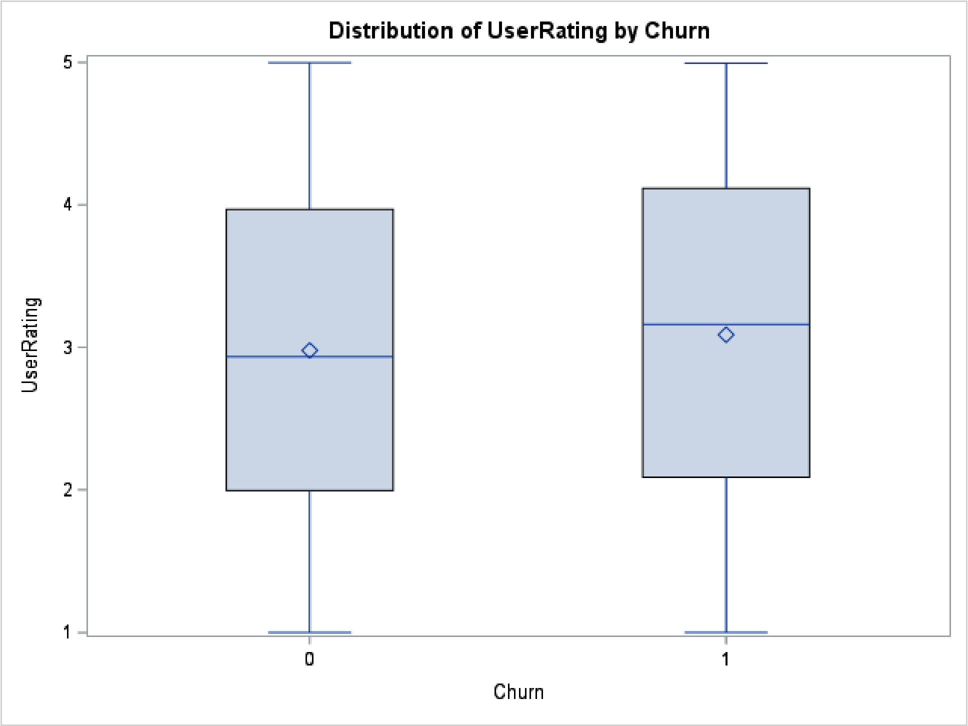
Because there is an important relationship between variables, such as payment method, amount of support tickets, subscription type, device type, and genre preference and customer churn, companies can improve customer experience in these particular places. This means communicating with customers to switch from electronic checks to less ambiguous payment methods, providing better support services that would reduce the amount of support tickets, and customizing content to the user's preferred genres. Also, customizing engagement strategies to those customers who are on Standard plans and TV devices can ultimately help retain the long-term users who show a lot of viewing and content usage, but churn based on other services.

Appendix

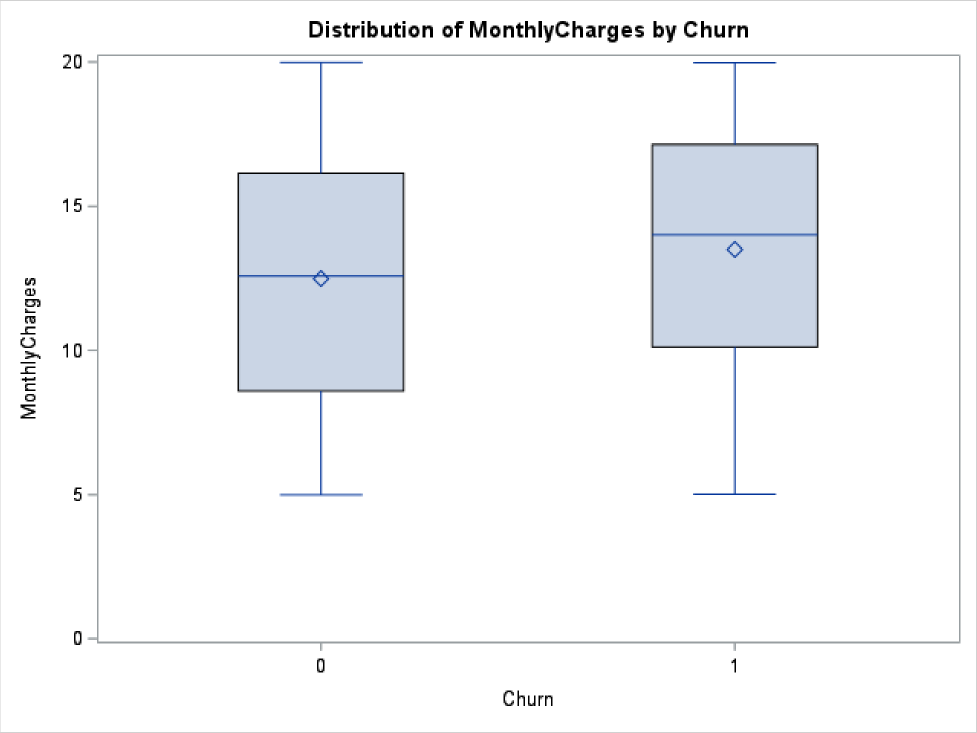
Appendix 1:



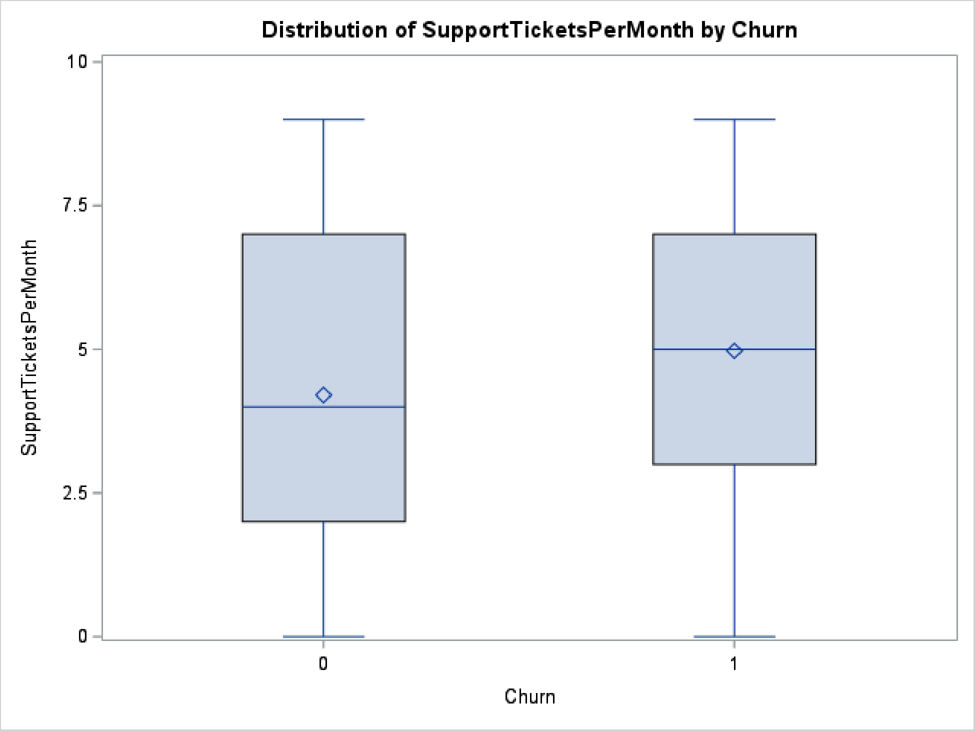
Appendix 2:



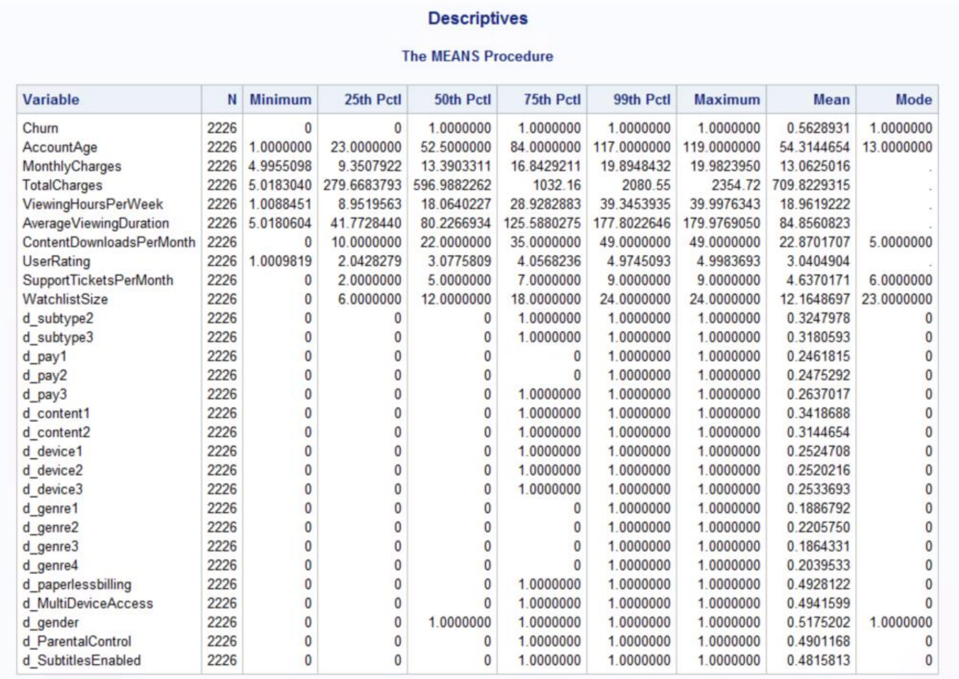
Appendix 3:



Appendix 4:



Appendix 5:

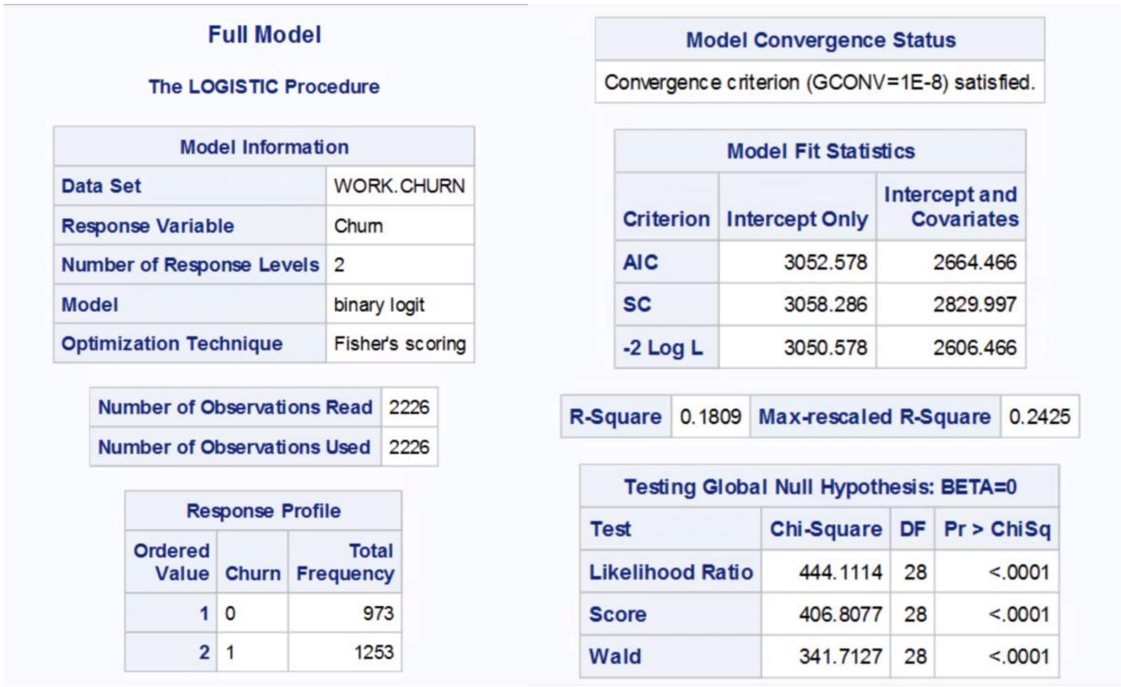


Descriptives

The MEANS Procedure

Variable	N	Minimum	25th Pctl	50th Pctl	75th Pctl	99th Pctl	Maximum	Mean	Mode
Churn	2226	0	0	1.0000000	1.0000000	1.0000000	1.0000000	0.5628931	1.0000000
AccountAge	2226	1.0000000	23.0000000	52.5000000	84.0000000	117.0000000	119.0000000	54.3144654	13.0000000
MonthlyCharges	2226	4.9955098	9.3507922	13.3903311	16.8429211	19.8948432	19.9823950	13.0625016	.
TotalCharges	2226	5.0183040	279.6683793	596.9882262	1032.16	2080.55	2354.72	709.8229315	.
ViewingHoursPerWeek	2226	1.0088451	8.9519563	18.0640227	28.9282883	39.3453935	39.9976343	18.9619222	.
AverageViewingDuration	2226	5.0180604	41.7728440	80.2266934	125.5880275	177.8022646	179.9769050	84.8560823	.
ContentDownloadsPerMonth	2226	0	10.0000000	22.0000000	35.0000000	49.0000000	49.0000000	22.8701707	5.0000000
UserRating	2226	1.0009819	2.0428279	3.0775809	4.0568236	4.9745093	4.9983693	3.0404904	.
SupportTicketsPerMonth	2226	0	2.0000000	5.0000000	7.0000000	9.0000000	9.0000000	4.6370171	6.0000000
WatchlistSize	2226	0	6.0000000	12.0000000	18.0000000	24.0000000	24.0000000	12.1648697	23.0000000
d_subtype2	2226	0	0	0	1.0000000	1.0000000	1.0000000	0.3247978	0
d_subtype3	2226	0	0	0	1.0000000	1.0000000	1.0000000	0.3180593	0
d_pay1	2226	0	0	0	0	1.0000000	1.0000000	0.2461815	0
d_pay2	2226	0	0	0	0	1.0000000	1.0000000	0.2475292	0
d_pay3	2226	0	0	0	1.0000000	1.0000000	1.0000000	0.2637017	0
d_content1	2226	0	0	0	1.0000000	1.0000000	1.0000000	0.3418688	0
d_content2	2226	0	0	0	1.0000000	1.0000000	1.0000000	0.3144654	0
d_device1	2226	0	0	0	1.0000000	1.0000000	1.0000000	0.2524708	0
d_device2	2226	0	0	0	1.0000000	1.0000000	1.0000000	0.2520216	0
d_device3	2226	0	0	0	1.0000000	1.0000000	1.0000000	0.2533693	0
d_genre1	2226	0	0	0	0	1.0000000	1.0000000	0.1886792	0
d_genre2	2226	0	0	0	0	1.0000000	1.0000000	0.2205750	0
d_genre3	2226	0	0	0	0	1.0000000	1.0000000	0.1864331	0
d_genre4	2226	0	0	0	0	1.0000000	1.0000000	0.2039533	0
d_paperlessbilling	2226	0	0	0	1.0000000	1.0000000	1.0000000	0.4928122	0
d_MultiDeviceAccess	2226	0	0	0	1.0000000	1.0000000	1.0000000	0.4941599	0
d_gender	2226	0	0	1.0000000	1.0000000	1.0000000	1.0000000	0.5175202	1.0000000
d_ParentalControl	2226	0	0	0	1.0000000	1.0000000	1.0000000	0.4901168	0
d_SubtitlesEnabled	2226	0	0	0	1.0000000	1.0000000	1.0000000	0.4815813	0

Appendix 6:



Full Model

The LOGISTIC Procedure

Model Information	
Data Set	WORK.CHURN
Response Variable	Churn
Number of Response Levels	2
Model	binary logit
Optimization Technique	Fisher's scoring

Number of Observations Read	2226
Number of Observations Used	2226

Response Profile

Ordered Value	Churn	Total Frequency
1	0	973
2	1	1253

Model Convergence Status
Convergence criterion (GCONV=1E-8) satisfied.

Model Fit Statistics

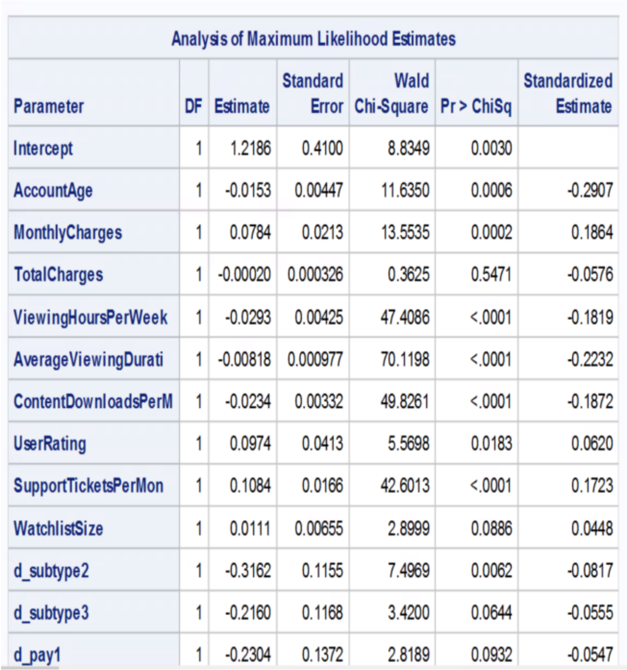
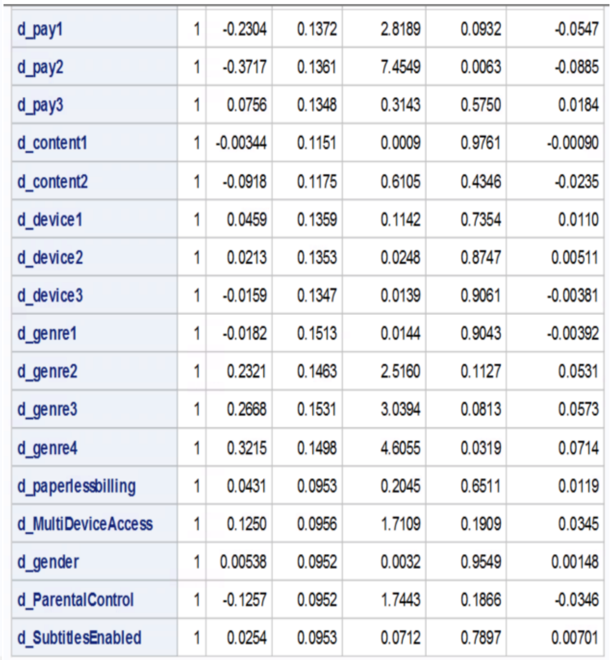
Criterion	Intercept Only	Intercept and Covariates
AIC	3052.578	2664.466
SC	3058.286	2829.997
-2 Log L	3050.578	2606.466

R-Square	0.1809	Max-rescaled R-Square	0.2425

Testing Global Null Hypothesis: BETA=0

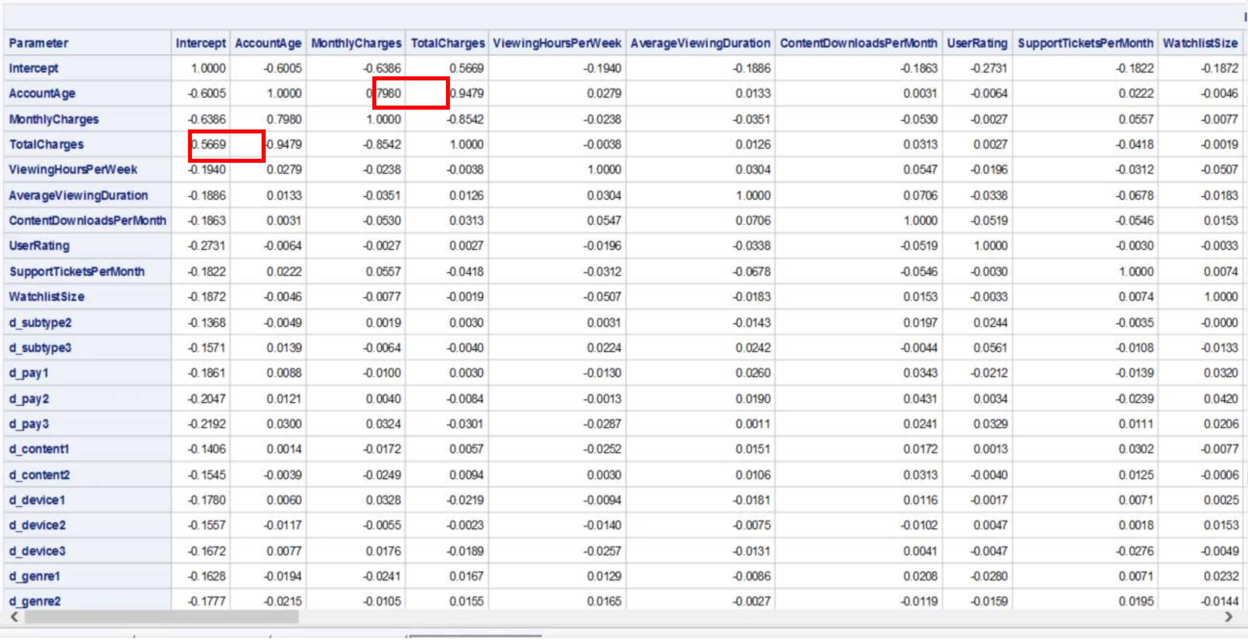
Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	444.1114	28	<.0001
Score	406.8077	28	<.0001
Wald	341.7127	28	<.0001

Appendix 7:

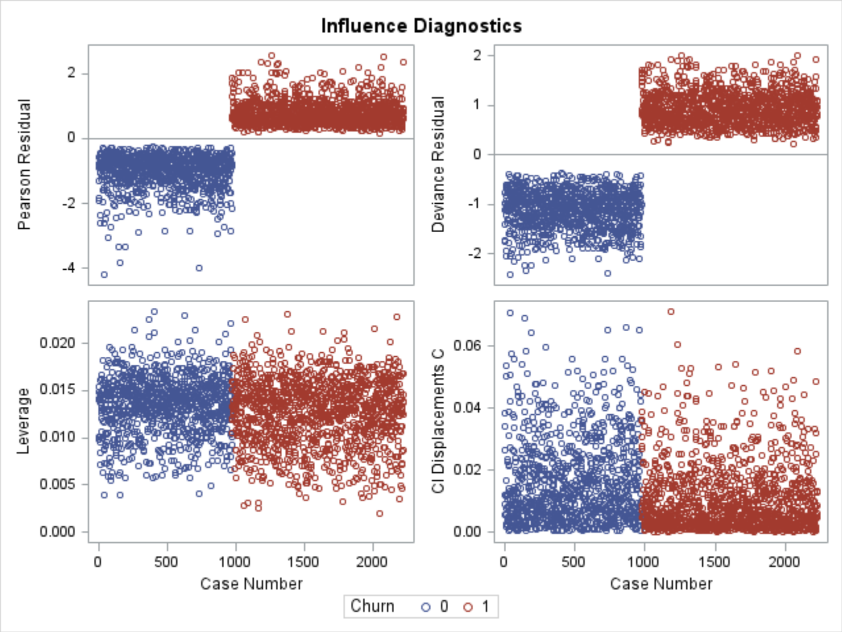
Analysis of Maximum Likelihood Estimates						
Parameter	DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq	Standardized Estimate
Intercept	1	1.2186	0.4100	8.8349	0.0030	
AccountAge	1	-0.0153	0.00447	11.6350	0.0006	-0.2907
MonthlyCharges	1	0.0784	0.0213	13.5535	0.0002	0.1864
TotalCharges	1	-0.00020	0.000326	0.3625	0.5471	-0.0576
ViewingHoursPerWeek	1	-0.0293	0.00425	47.4086	<.0001	-0.1819
AverageViewingDurati	1	-0.00818	0.000977	70.1198	<.0001	-0.2232
ContentDownloadsPerM	1	-0.0234	0.00332	49.8261	<.0001	-0.1872
UserRating	1	0.0974	0.0413	5.5698	0.0183	0.0620
SupportTicketsPerMon	1	0.1084	0.0166	42.6013	<.0001	0.1723
WatchlistSize	1	0.0111	0.00655	2.8999	0.0886	0.0448
d_subtype2	1	-0.3162	0.1155	7.4969	0.0062	-0.0817
d_subtype3	1	-0.2160	0.1168	3.4200	0.0644	-0.0555
d_pay1	1	-0.2304	0.1372	2.8189	0.0932	-0.0547
d_pay2	1	-0.3717	0.1361	7.4549	0.0063	-0.0885
d_pay3	1	0.0756	0.1348	0.3143	0.5750	0.0184
d_content1	1	-0.00344	0.1151	0.0009	0.9761	-0.00090
d_content2	1	-0.0918	0.1175	0.6105	0.4346	-0.0235
d_device1	1	0.0459	0.1359	0.1142	0.7354	0.0110
d_device2	1	0.0213	0.1353	0.0248	0.8747	0.00511
d_device3	1	-0.0159	0.1347	0.0139	0.9061	-0.00381
d_genre1	1	-0.0182	0.1513	0.0144	0.9043	-0.00392
d_genre2	1	0.2321	0.1463	2.5160	0.1127	0.0531
d_genre3	1	0.2668	0.1531	3.0394	0.0813	0.0573
d_genre4	1	0.3215	0.1498	4.6055	0.0319	0.0714
d_paperlessbilling	1	0.0431	0.0953	0.2045	0.6511	0.0119
d_MultiDeviceAccess	1	0.1250	0.0956	1.7109	0.1909	0.0345
d_gender	1	0.00538	0.0952	0.0032	0.9549	0.00148
d_ParentalControl	1	-0.1257	0.0952	1.7443	0.1866	-0.0346
d_SubtitlesEnabled	1	0.0254	0.0953	0.0712	0.7897	0.00701

Appendix 8:

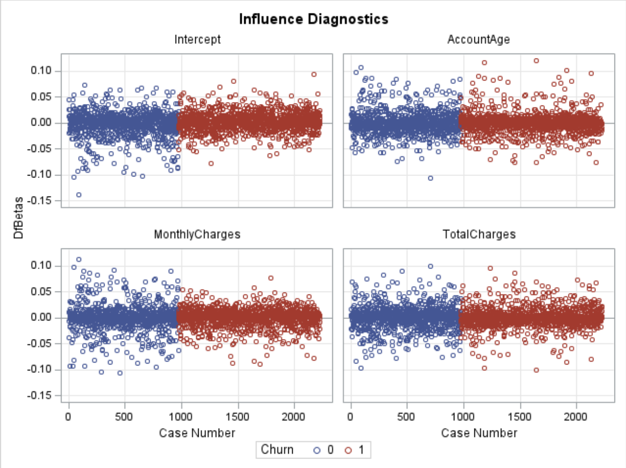


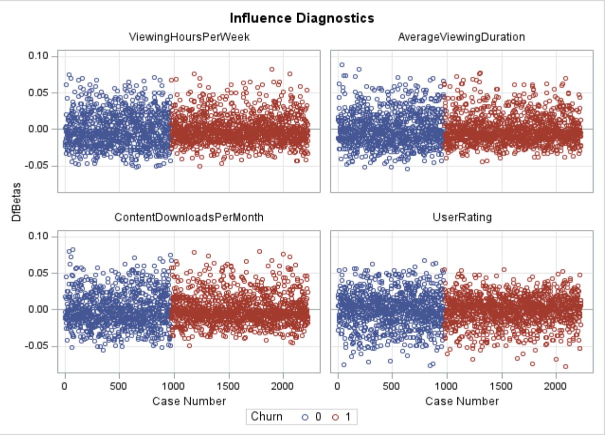
Parameter	Intercept	AccountAge	MonthlyCharges	TotalCharges	ViewingHoursPerWeek	AverageViewingDuration	ContentDownloadsPerMonth	UserRating	SupportTicketsPerMonth	WatchlistSize
Intercept	1.0000	-0.6005	-0.6386	0.5669	-0.1940	-0.1886	-0.1863	-0.2731	-0.1822	-0.1872
AccountAge	-0.6005	1.0000	0.7980	0.9479	0.0279	0.0133	0.0031	-0.0064	0.0222	-0.0046
MonthlyCharges	-0.6386	0.7980	1.0000	-0.8542	-0.0238	-0.0351	-0.0530	-0.0027	0.0557	-0.0077
TotalCharges	0.5669	0.9479	-0.8542	1.0000	-0.0038	0.0126	0.0313	0.0027	-0.0418	-0.0019
ViewingHoursPerWeek	-0.1940	0.0279	-0.0238	-0.0038	1.0000	0.0304	0.0547	-0.0196	-0.0312	-0.0507
AverageViewingDuration	-0.1886	0.0133	-0.0351	0.0126	0.0304	1.0000	0.0706	-0.0338	-0.0678	-0.0183
ContentDownloadsPerMonth	-0.1863	0.0031	-0.0530	0.0313	0.0547	0.0706	1.0000	-0.0519	-0.0546	0.0153
UserRating	-0.2731	-0.0064	-0.0027	0.0027	-0.0196	-0.0338	-0.0519	1.0000	-0.0030	-0.0033
SupportTicketsPerMonth	-0.1822	0.0222	0.0557	-0.0418	-0.0312	-0.0678	-0.0546	-0.0030	1.0000	0.0074
WatchlistSize	-0.1872	-0.0046	-0.0077	-0.0019	-0.0507	-0.0183	0.0153	-0.0033	0.0074	1.0000
d_subtype2	-0.1368	-0.0049	0.0019	0.0030	0.0031	-0.0143	0.0197	0.0244	-0.0035	-0.0000
d_subtype3	-0.1571	0.0139	-0.0064	-0.0040	0.0224	0.0242	-0.0044	0.0561	-0.0108	-0.0133
d_pay1	-0.1861	0.0088	-0.0100	0.0030	-0.0130	0.0260	0.0343	-0.0212	-0.0139	0.0320
d_pay2	-0.2047	0.0121	0.0040	-0.0084	-0.0013	0.0190	0.0431	0.0034	-0.0239	0.0420
d_pay3	-0.2192	0.0300	0.0324	-0.0301	-0.0287	0.0011	0.0241	0.0329	0.0111	0.0206
d_content1	-0.1406	0.0014	-0.0172	0.0057	-0.0252	0.0151	0.0172	0.0013	0.0302	-0.0077
d_content2	-0.1545	-0.0039	-0.0249	0.0094	0.0030	0.0106	0.0313	-0.0040	0.0125	-0.0006
d_device1	-0.1780	0.0060	0.0328	-0.0219	-0.0094	-0.0181	0.0116	-0.0017	0.0071	0.0025
d_device2	-0.1557	-0.0117	-0.0055	-0.0023	-0.0140	-0.0075	-0.0102	0.0047	0.0018	0.0153
d_device3	-0.1672	0.0077	0.0176	-0.0189	-0.0257	-0.0131	0.0041	-0.0047	-0.0276	-0.0049
d_genre1	-0.1628	-0.0194	-0.0241	0.0167	0.0129	-0.0086	0.0208	-0.0280	0.0071	0.0232
d_genre2	-0.1777	-0.0215	-0.0105	0.0155	0.0165	-0.0027	-0.0119	-0.0159	0.0195	-0.0144

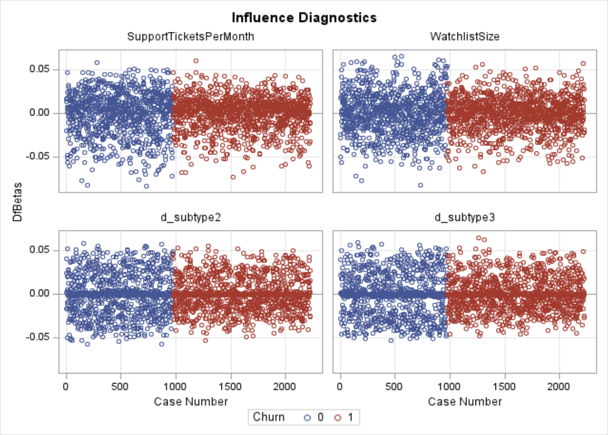
Appendix 9:

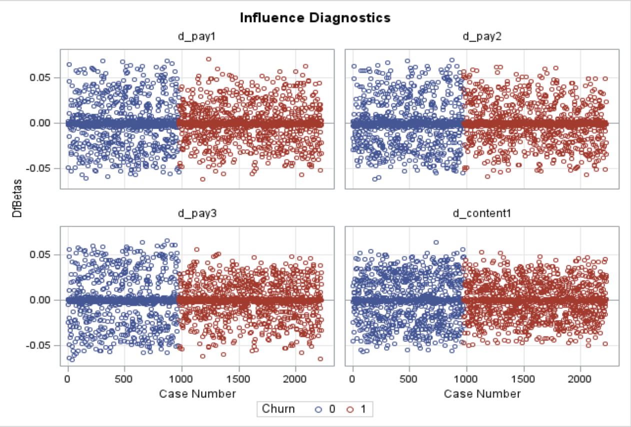


Appendix 10:

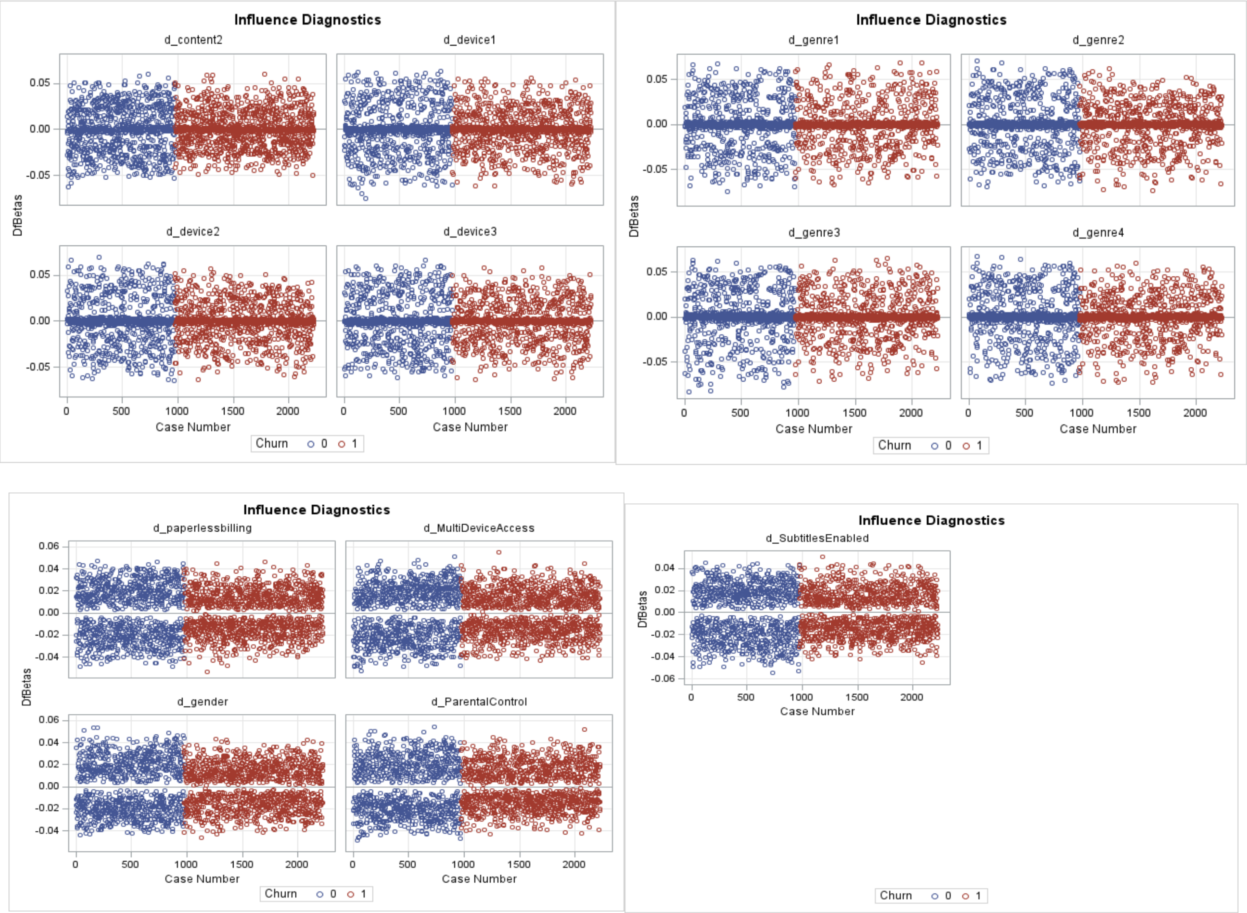




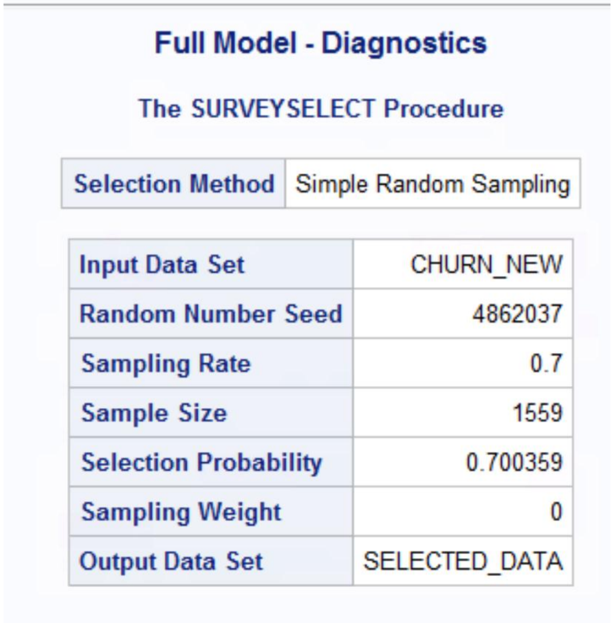




Appendix 11:



Appendix 12:



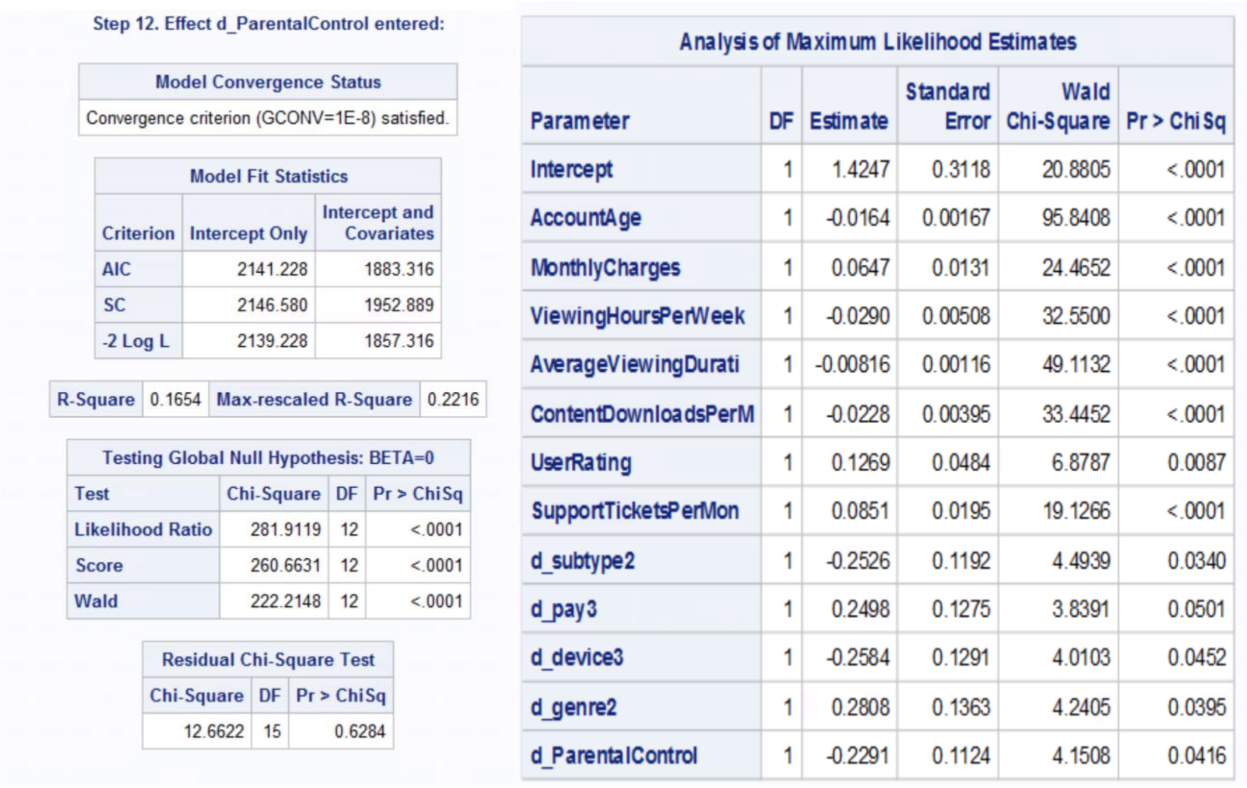
Full Model - Diagnostics

The SURVEYSELECT Procedure

Selection Method	Simple Random Sampling

Input Data Set	CHURN_NEW
Random Number Seed	4862037
Sampling Rate	0.7
Sample Size	1559
Selection Probability	0.700359
Sampling Weight	0
Output Data Set	SELECTED_DATA

Appendix 13:



Step 12. Effect d_ParentalControl entered:

Model Convergence Status
Convergence criterion (GCONV=1E-8) satisfied.

Model Fit Statistics

Criterion	Intercept Only	Intercept and Covariates
AIC	2141.228	1883.316
SC	2146.580	1952.889
-2 Log L	2139.228	1857.316

R-Square	0.1654	Max-rescaled R-Square	0.2216

Testing Global Null Hypothesis: BETA=0

Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	281.9119	12	<.0001
Score	260.6631	12	<.0001
Wald	222.2148	12	<.0001

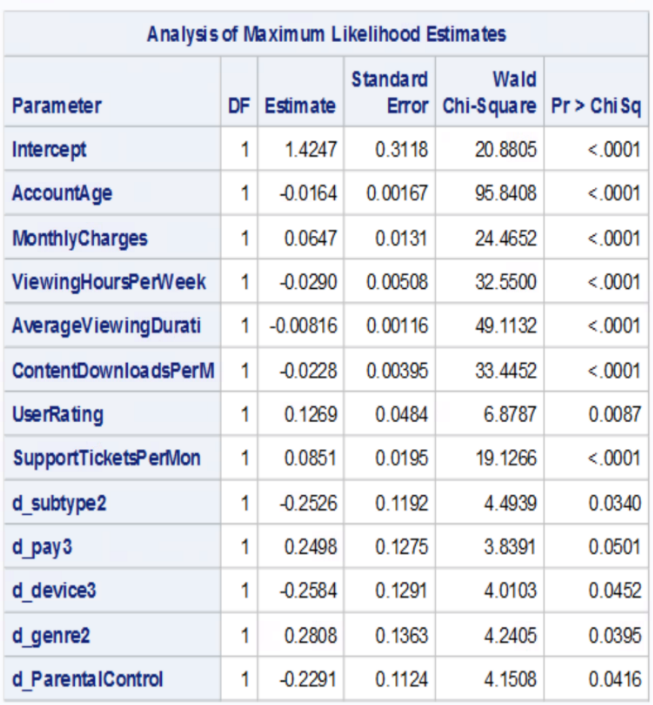
Residual Chi-Square Test

Chi-Square	DF	Pr > ChiSq
12.6622	15	0.6284

Analysis of Maximum Likelihood Estimates

Parameter	DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept	1	1.4247	0.3118	20.8805	<.0001
AccountAge	1	-0.0164	0.00167	95.8408	<.0001
MonthlyCharges	1	0.0647	0.0131	24.4652	<.0001
ViewingHoursPerWeek	1	-0.0290	0.00508	32.5500	<.0001
AverageViewingDurati	1	-0.00816	0.00116	49.1132	<.0001
ContentDownloadsPerM	1	-0.0228	0.00395	33.4452	<.0001
UserRating	1	0.1269	0.0484	6.8787	0.0087
SupportTicketsPerMon	1	0.0851	0.0195	19.1266	<.0001
d_subtype2	1	-0.2526	0.1192	4.4939	0.0340
d_pay3	1	0.2498	0.1275	3.8391	0.0501
d_device3	1	-0.2584	0.1291	4.0103	0.0452
d_genre2	1	0.2808	0.1363	4.2405	0.0395
d_ParentalControl	1	-0.2291	0.1124	4.1508	0.0416

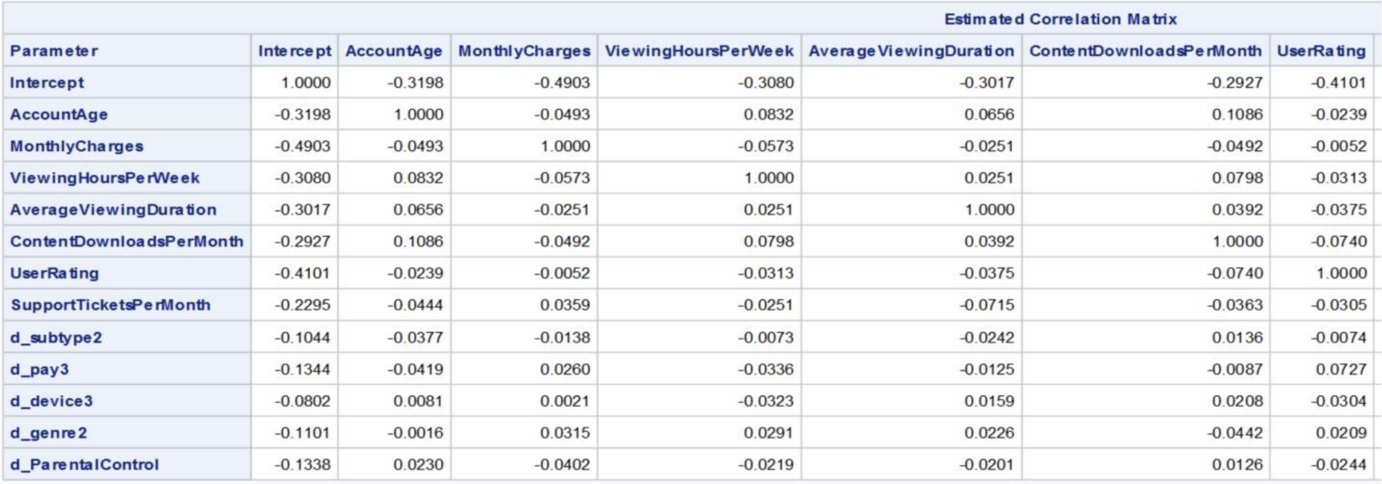
Appendix 14:



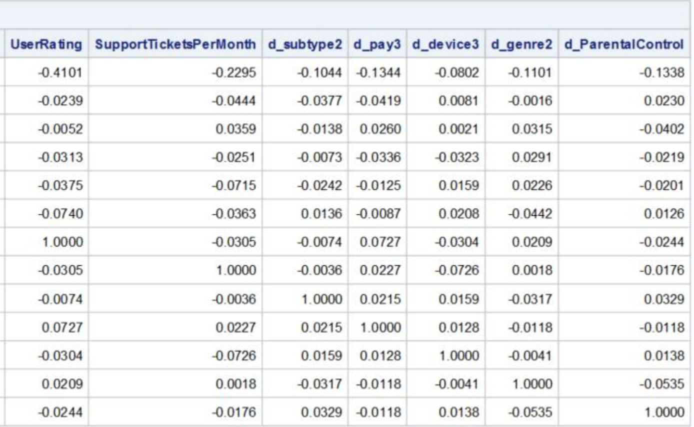
Analysis of Maximum Likelihood Estimates

Parameter	DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept	1	1.4247	0.3118	20.8805	<.0001
AccountAge	1	-0.0164	0.00167	95.8408	<.0001
MonthlyCharges	1	0.0647	0.0131	24.4652	<.0001
ViewingHoursPerWeek	1	-0.0290	0.00508	32.5500	<.0001
AverageViewingDurati	1	-0.00816	0.00116	49.1132	<.0001
ContentDownloadsPerM	1	-0.0228	0.00395	33.4452	<.0001
UserRating	1	0.1269	0.0484	6.8787	0.0087
SupportTicketsPerMon	1	0.0851	0.0195	19.1266	<.0001
d_subtype2	1	-0.2526	0.1192	4.4939	0.0340
d_pay3	1	0.2498	0.1275	3.8391	0.0501
d_device3	1	-0.2584	0.1291	4.0103	0.0452
d_genre2	1	0.2808	0.1363	4.2405	0.0395
d_ParentalControl	1	-0.2291	0.1124	4.1508	0.0416

Appendix 15:

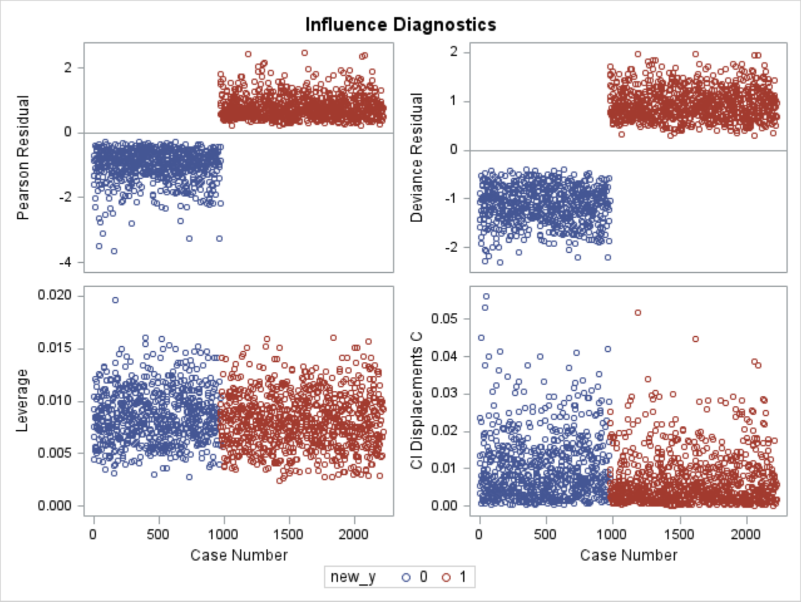


					Estimated Correlation Matrix		
Parameter	Intercept	AccountAge	MonthlyCharges	ViewingHoursPerWeek	AverageViewingDuration	ContentDownloadsPerMonth	UserRating
Intercept	1.0000	-0.3198	-0.4903	-0.3080	-0.3017	-0.2927	-0.4101
AccountAge	-0.3198	1.0000	-0.0493	0.0832	0.0656	0.1086	-0.0239
MonthlyCharges	-0.4903	-0.0493	1.0000	-0.0573	-0.0251	-0.0492	-0.0052
ViewingHoursPerWeek	-0.3080	0.0832	-0.0573	1.0000	0.0251	0.0798	-0.0313
AverageViewingDuration	-0.3017	0.0656	-0.0251	0.0251	1.0000	0.0392	-0.0375
ContentDownloadsPerMonth	-0.2927	0.1086	-0.0492	0.0798	0.0392	1.0000	-0.0740
UserRating	-0.4101	-0.0239	-0.0052	-0.0313	-0.0375	-0.0740	1.0000
SupportTicketsPerMonth	-0.2295	-0.0444	0.0359	-0.0251	-0.0715	-0.0363	-0.0305
d_subtype2	-0.1044	-0.0377	-0.0138	-0.0073	-0.0242	0.0136	-0.0074
d_pay3	-0.1344	-0.0419	0.0260	-0.0336	-0.0125	-0.0087	0.0727
d_device3	-0.0802	0.0081	0.0021	-0.0323	0.0159	0.0208	-0.0304
d_genre2	-0.1101	-0.0016	0.0315	0.0291	0.0226	-0.0442	0.0209
d_ParentalControl	-0.1338	0.0230	-0.0402	-0.0219	-0.0201	0.0126	-0.0244

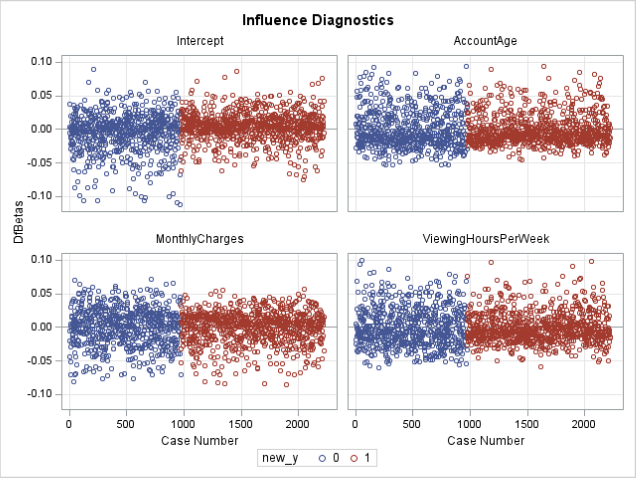


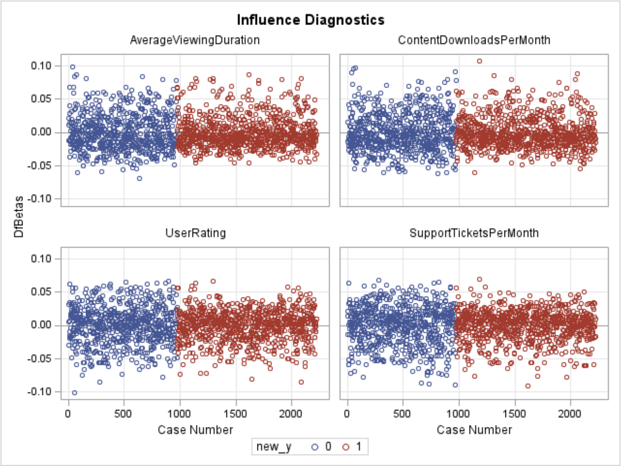
UserRating	SupportTicketsPerMonth	d_subtype2	d_pay3	d_device3	d_genre2	d_ParentalControl
-0.4101	-0.2295	-0.1044	-0.1344	-0.0802	-0.1101	-0.1338
-0.0239	-0.0444	-0.0377	-0.0419	0.0081	-0.0016	0.0230
-0.0052	0.0359	-0.0138	0.0260	0.0021	0.0315	-0.0402
-0.0313	-0.0251	-0.0073	-0.0336	-0.0323	0.0291	-0.0219
-0.0375	-0.0715	-0.0242	-0.0125	0.0159	0.0226	-0.0201
-0.0740	-0.0363	0.0136	-0.0087	0.0208	-0.0442	0.0126
1.0000	-0.0305	-0.0074	0.0727	-0.0304	0.0209	-0.0244
-0.0305	1.0000	-0.0036	0.0227	-0.0726	0.0018	-0.0176
-0.0074	-0.0036	1.0000	0.0215	0.0159	-0.0317	0.0329
0.0727	0.0227	0.0215	1.0000	0.0128	-0.0118	-0.0118
-0.0304	-0.0726	0.0159	0.0128	1.0000	-0.0041	0.0138
0.0209	0.0018	-0.0317	-0.0118	-0.0041	1.0000	-0.0535
-0.0244	-0.0176	0.0329	-0.0118	0.0138	-0.0535	1.0000

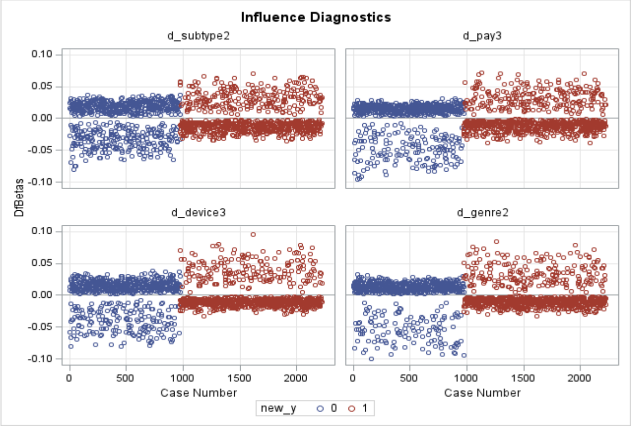
Appendix 16:

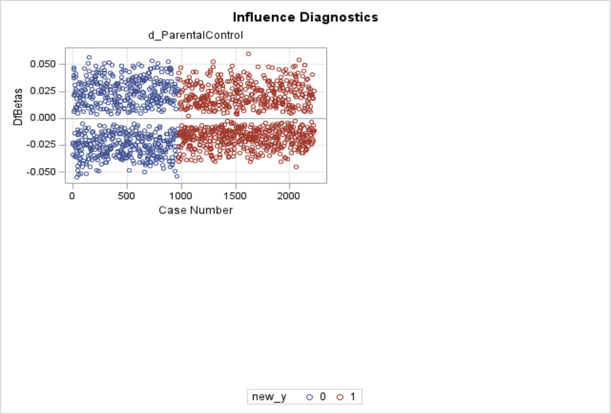


Appendix 17:

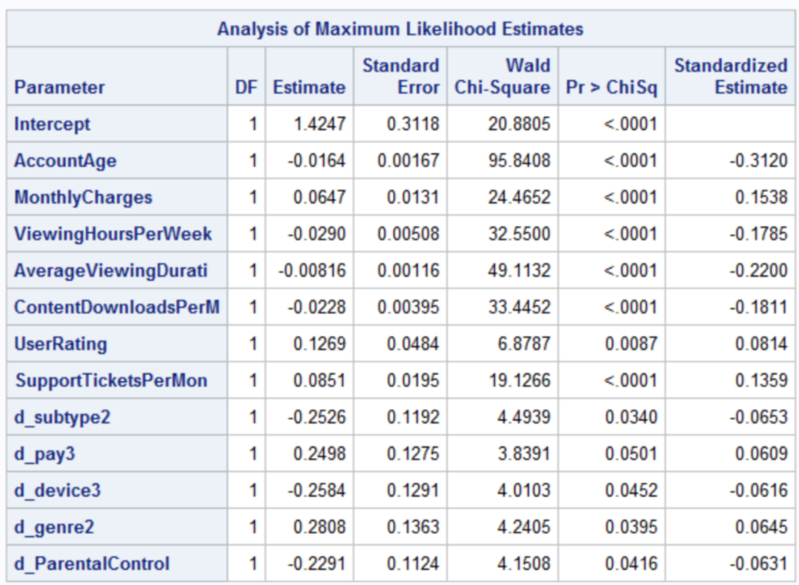








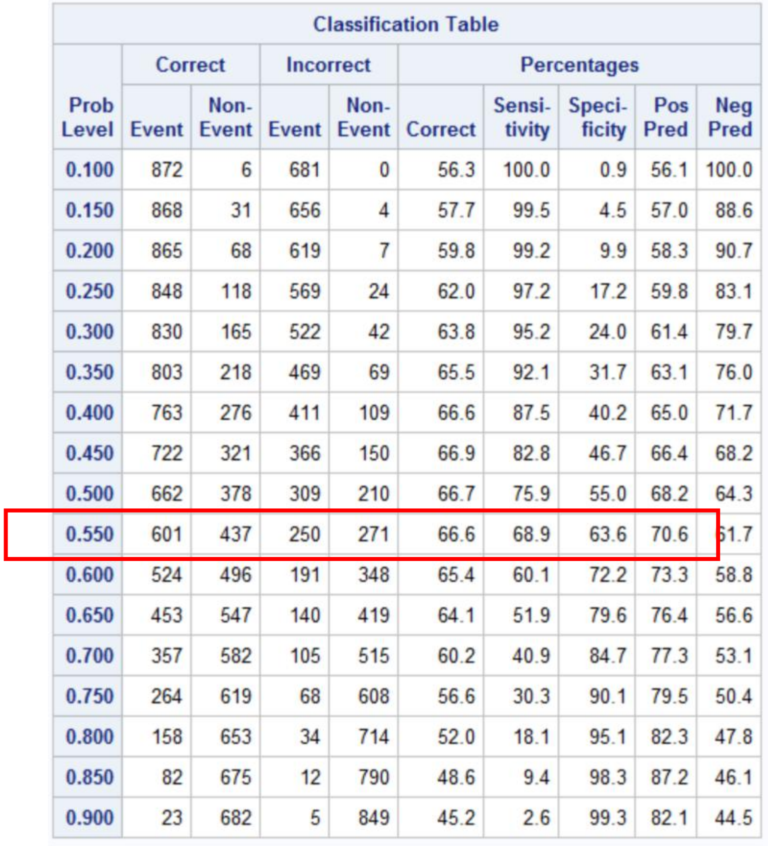
Appendix 18:



Analysis of Maximum Likelihood Estimates

Parameter	DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq	Standardized Estimate
Intercept	1	1.4247	0.3118	20.8805	<.0001	
AccountAge	1	-0.0164	0.00167	95.8408	<.0001	-0.3120
MonthlyCharges	1	0.0647	0.0131	24.4652	<.0001	0.1538
ViewingHoursPerWeek	1	-0.0290	0.00508	32.5500	<.0001	-0.1785
AverageViewingDurati	1	-0.00816	0.00116	49.1132	<.0001	-0.2200
ContentDownloadsPerM	1	-0.0228	0.00395	33.4452	<.0001	-0.1811
UserRating	1	0.1269	0.0484	6.8787	0.0087	0.0814
SupportTicketsPerMon	1	0.0851	0.0195	19.1266	<.0001	0.1359
d_subtype2	1	-0.2526	0.1192	4.4939	0.0340	-0.0653
d_pay3	1	0.2498	0.1275	3.8391	0.0501	0.0609
d_device3	1	-0.2584	0.1291	4.0103	0.0452	-0.0616
d_genre2	1	0.2808	0.1363	4.2405	0.0395	0.0645
d_ParentalControl	1	-0.2291	0.1124	4.1508	0.0416	-0.0631

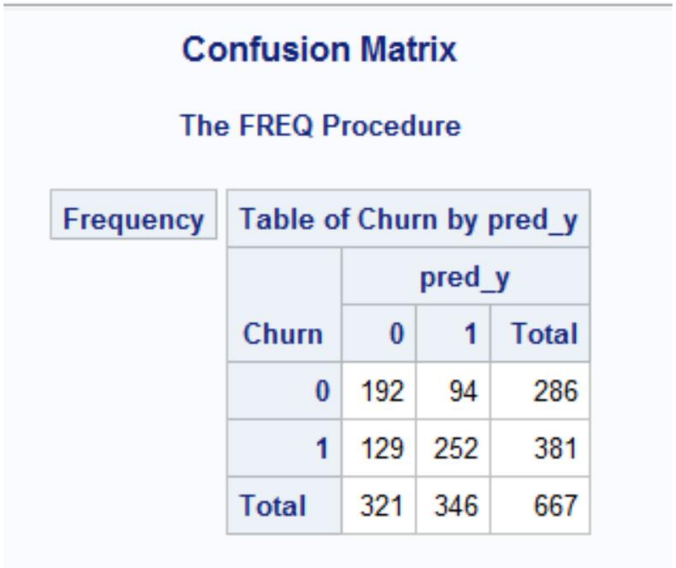
Appendix 19:



Classification Table

Prob Level	Correct		Incorrect		Percentages				
	Event	Non-Event	Event	Non-Event	Correct	Sensi-tivity	Speci-ficity	Pos Pred	Neg Pred
0.100	872	6	681	0	56.3	100.0	0.9	56.1	100.0
0.150	868	31	656	4	57.7	99.5	4.5	57.0	88.6
0.200	865	68	619	7	59.8	99.2	9.9	58.3	90.7
0.250	848	118	569	24	62.0	97.2	17.2	59.8	83.1
0.300	830	165	522	42	63.8	95.2	24.0	61.4	79.7
0.350	803	218	469	69	65.5	92.1	31.7	63.1	76.0
0.400	763	276	411	109	66.6	87.5	40.2	65.0	71.7
0.450	722	321	366	150	66.9	82.8	46.7	66.4	68.2
0.500	662	378	309	210	66.7	75.9	55.0	68.2	64.3
0.550	601	437	250	271	66.6	68.9	63.6	70.6	61.7
0.600	524	496	191	348	65.4	60.1	72.2	73.3	58.8
0.650	453	547	140	419	64.1	51.9	79.6	76.4	56.6
0.700	357	582	105	515	60.2	40.9	84.7	77.3	53.1
0.750	264	619	68	608	56.6	30.3	90.1	79.5	50.4
0.800	158	653	34	714	52.0	18.1	95.1	82.3	47.8
0.850	82	675	12	790	48.6	9.4	98.3	87.2	46.1
0.900	23	682	5	849	45.2	2.6	99.3	82.1	44.5

Appendix 20:

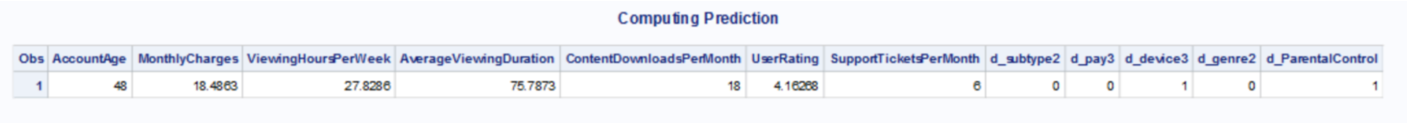


Confusion Matrix

The FREQ Procedure

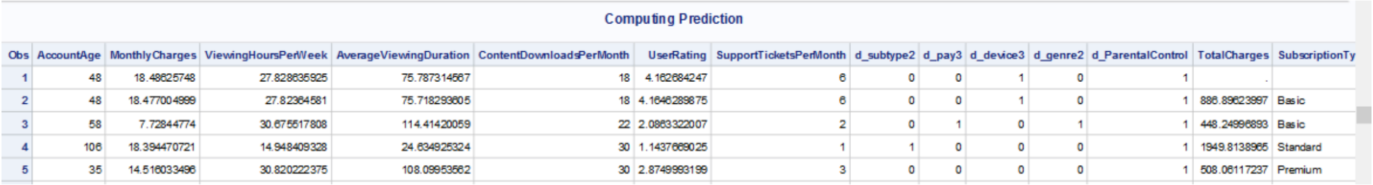
Frequency	Table of Churn by pred_y			
		pred_y		
	Churn	0	1	Total
	0	192	94	286
	1	129	252	381
	Total	321	346	667

Appendix 21:



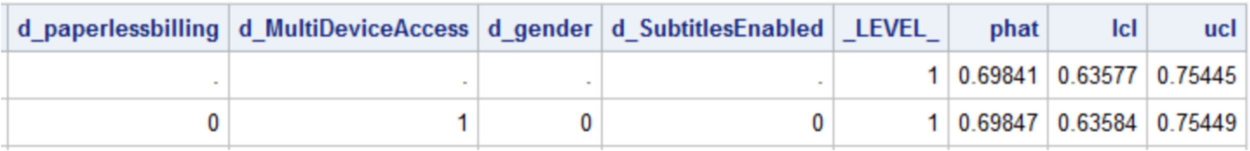
Computing Prediction

Obs	AccountAge	MonthlyCharges	ViewingHoursPerWeek	AverageViewingDuration	ContentDownloadsPerMonth	UserRating	SupportTicketsPerMonth	d_subtype2	d_pay3	d_device3	d_genre2	d_ParentalControl
1	48	18.4863	27.8286	75.7873	18	4.16268	6	0	0	1	0	1



Computing Prediction

Obs	AccountAge	MonthlyCharges	ViewingHoursPerWeek	AverageViewingDuration	ContentDownloadsPerMonth	UserRating	SupportTicketsPerMonth	d_subtype2	d_pay3	d_device3	d_genre2	d_ParentalControl	TotalCharges	SubscriptionTy
1	48	18.48625748	27.828635925	75.787314567	18	4.162684247	6	0	0	1	0	1	.	
2	48	18.477004999	27.82364581	75.718293605	18	4.1646289875	6	0	0	1	0	1	886.89623997	Basic
3	58	7.72844774	30.675517808	114.41420059	22	2.0863322007	2	0	1	0	1	1	448.24996893	Basic
4	106	18.394470721	14.948409328	24.634925324	30	1.1437669025	1	1	0	0	0	1	1949.8138965	Standard
5	35	14.516033496	30.820222375	108.09953562	30	2.8749993199	3	0	0	0	0	1	508.06117237	Premium



d_paperlessbilling	d_MultiDeviceAccess	d_gender	d_SubtitlesEnabled	_LEVEL_	phat	lcl	ucl
.	.	.	.	1	0.69841	0.63577	0.75445
0	1	0	0	1	0.69847	0.63584	0.75449