ASSIGNMENT

Customer Churn Analysis

21392866
Suresh Mukul

2024-BUS5CA(CB-1) - CUSTOMER ANALYTICS AND SOCIAL MEDIA

Table of Contents

TASK 1: 2

Analysis 2

Dashboard: 2

1. Churn Rate by Gender 2

2. Average Tenure by Churn Status 3

3. TOTAL Churn Distribution 3

4. Average Satisfaction Score by Churn Status and Gender 4

5. Distribution of Preferred Login Devices among Churned and Non-churned Customers 5

6. High-Value Customers by City Tier 5

7. Gender Distribution by Churn 6

Descriptive Statistics 7

TASK 2: 7

1. Mean Predicted vs. Depth Plot: 8

2. Model Score vs. Mean Target Plot 9

Detailed Analysis and Comparison of Models 11

Conclusion: 13

Recommendation: 13

TASK 3: 14

1. Gender-Specific Churn Reduction Campaign 14

2. Onboarding and Early Engagement Campaign 14

3. Cross-Device User Experience Enhancement Campaign 14

4. High-Value Customer Retention Campaign 14

5. Satisfaction Score Analysis and Improvement Campaign 14

6. Predictive Churn Reduction Using Neural Network Insights 14

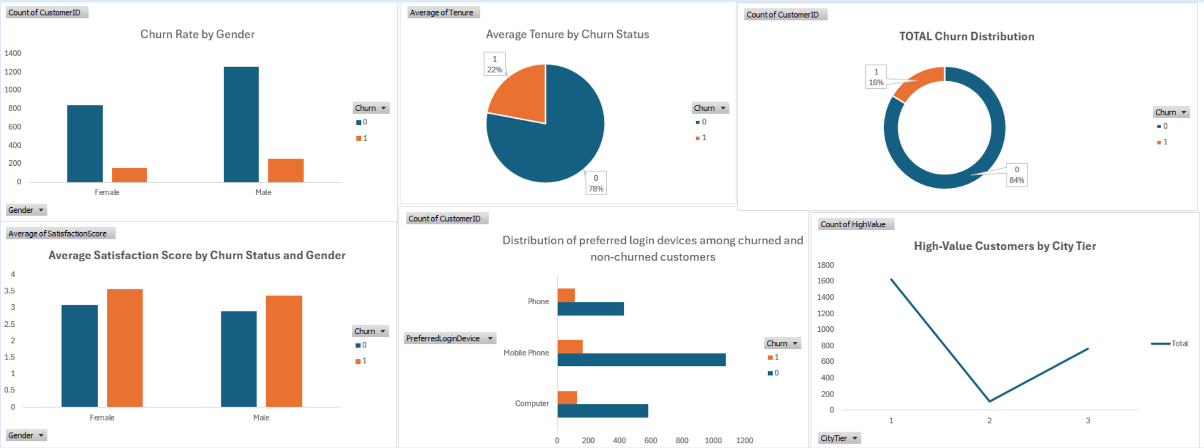
7. Multi-Channel Communication Strategy 14

Questions: 14

TASK 1:

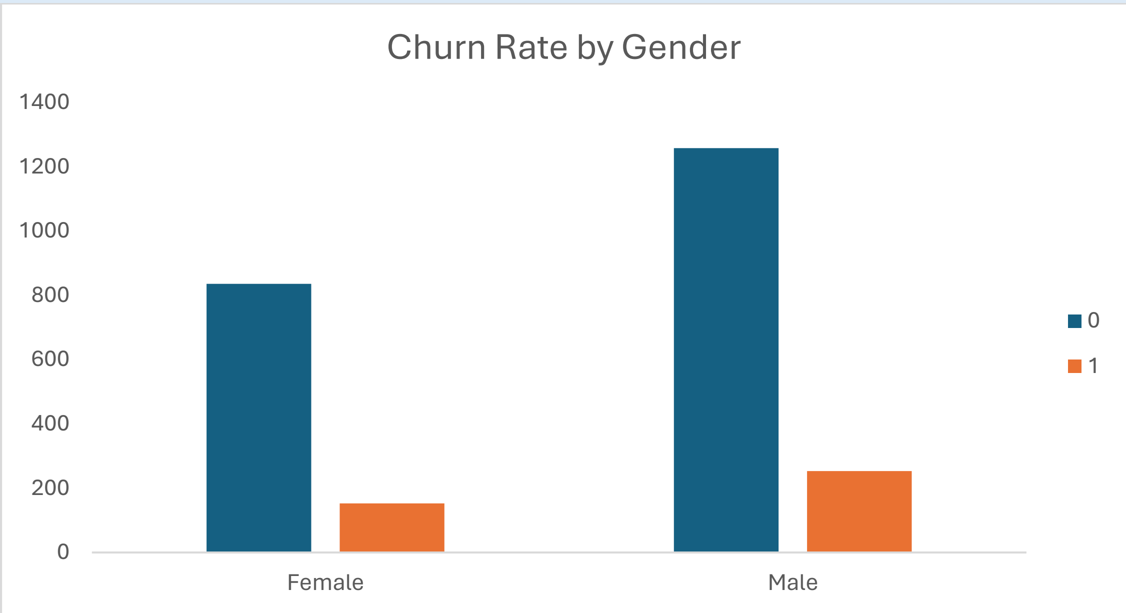
Analysis

Dashboard:



1. Churn Rate by Gender

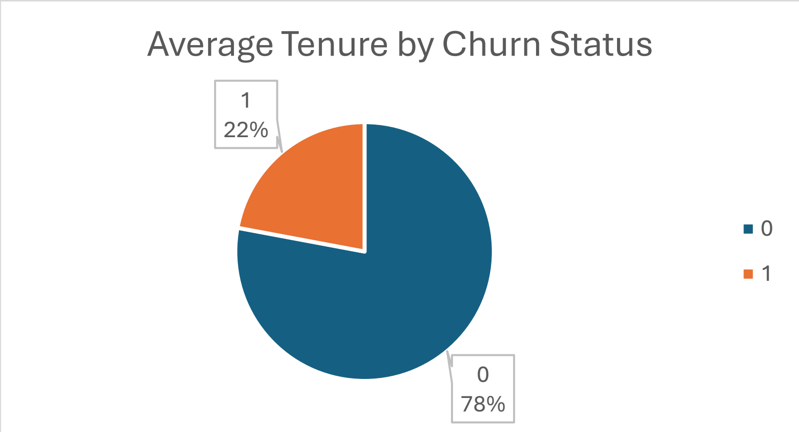
Topic: Gender-wise Churn Rate Analysis



- The bar chart compares the churn rate between male and female customers.
- Male customers have a significantly higher churn rate compared to female customers.
- The number of male customers who have not churned is also higher than that of female customers.
- This suggests that gender might be a factor influencing customer churn, with males being more likely to churn.

- Companies might need to investigate the reasons behind the higher churn rate among male customers and implement gender-specific retention strategies.

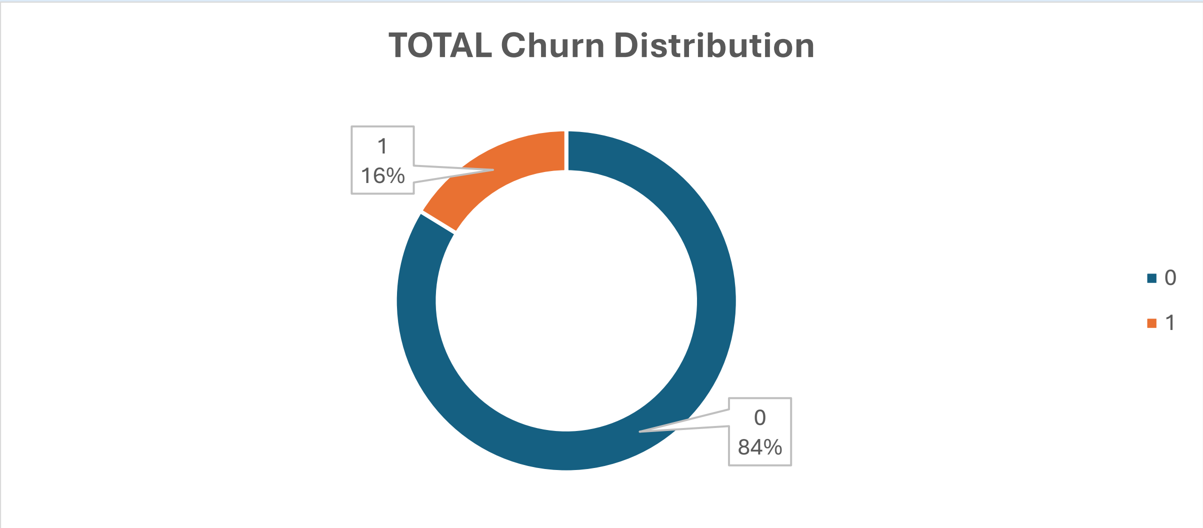
2. Average Tenure by Churn Status



Topic: Tenure and Churn Relationship

- The pie chart shows the average tenure distribution between churned and non-churned customers.
- Non-churned customers have a longer average tenure compared to churned customers.
- Specifically, 78% of customers who have not churned have a longer tenure.
- Shorter tenure is associated with a higher likelihood of churn, indicating that newer customers are more prone to leave.
- Focusing on onboarding and early engagement might help in reducing churn among newer customers.

3. TOTAL Churn Distribution

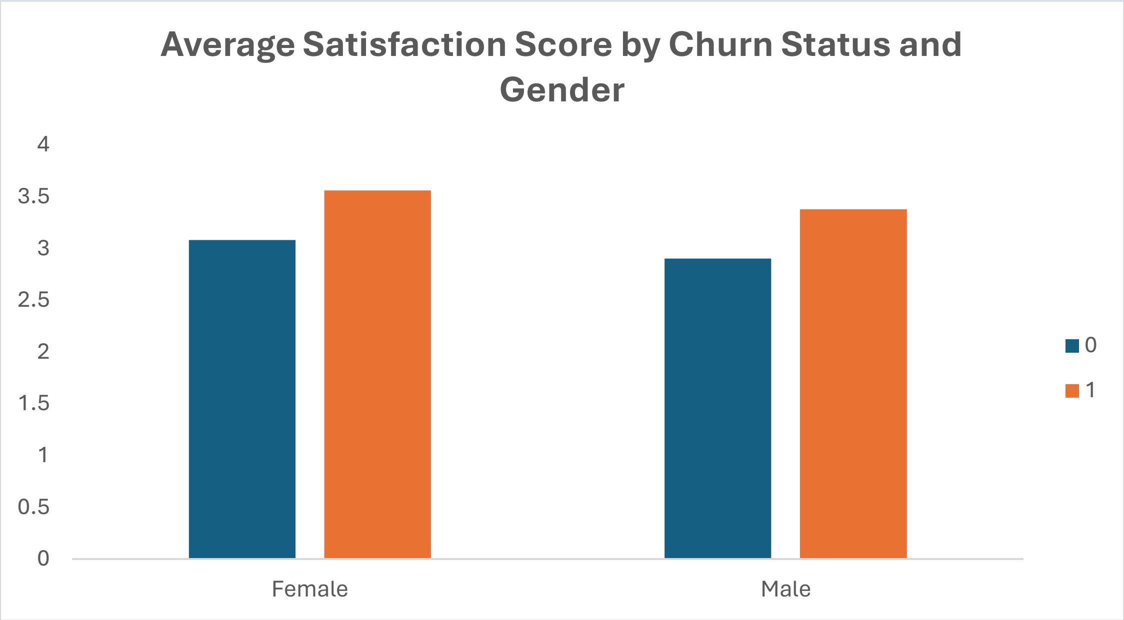


Topic: Overall Churn Distribution

- The donut chart displays the total distribution of churned vs. non-churned customers.

- A majority (84%) of the customers have not churned, while 16% have churned.
- This indicates that while churn is present, a substantial proportion of customers remain loyal.
- Retaining this loyalty base is crucial, and understanding the characteristics of this group can provide insights for retention strategies.
- The churned segment, though smaller, represents an area for potential improvement and targeted interventions.

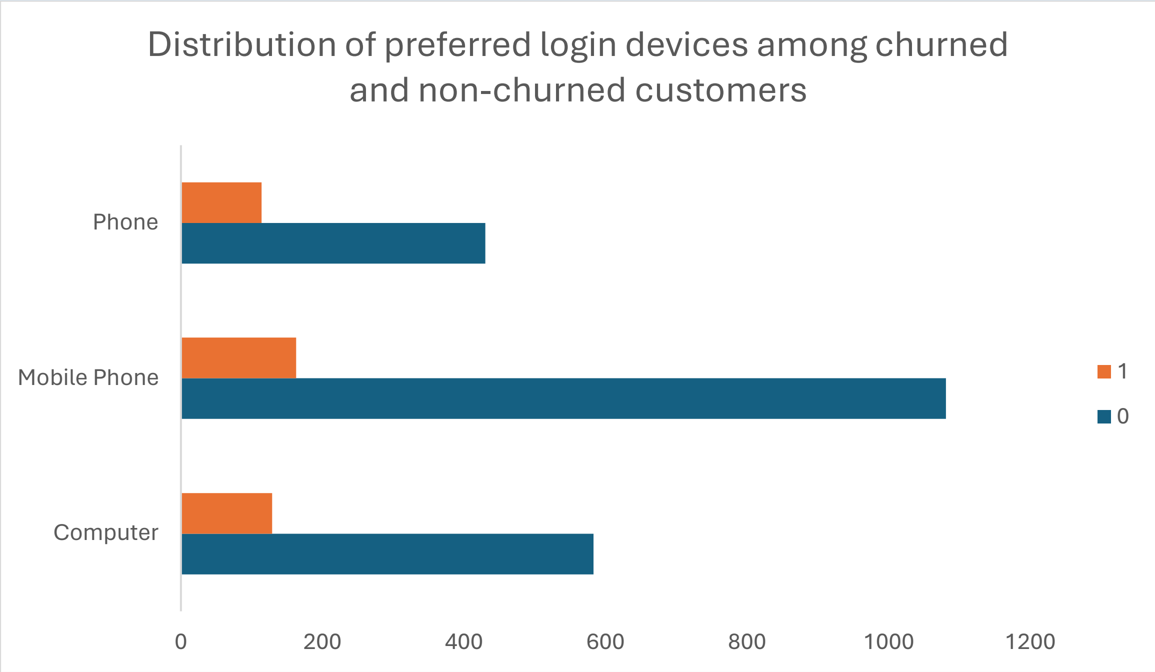
4. Average Satisfaction Score by Churn Status and Gender



Topic: Satisfaction Scores and Churn Analysis

- The bar chart compares average satisfaction scores between churned and non-churned customers, segmented by gender.
- Both male and female customers who have churned exhibit higher satisfaction scores than those who have not churned.
- This counterintuitive finding suggests that satisfaction score alone may not be a definitive predictor of churn.
- Additional factors such as service quality, value perception, and competitive offers might influence churn decisions despite higher satisfaction.
- Further analysis is needed to understand the underlying reasons for churn among apparently satisfied customers.

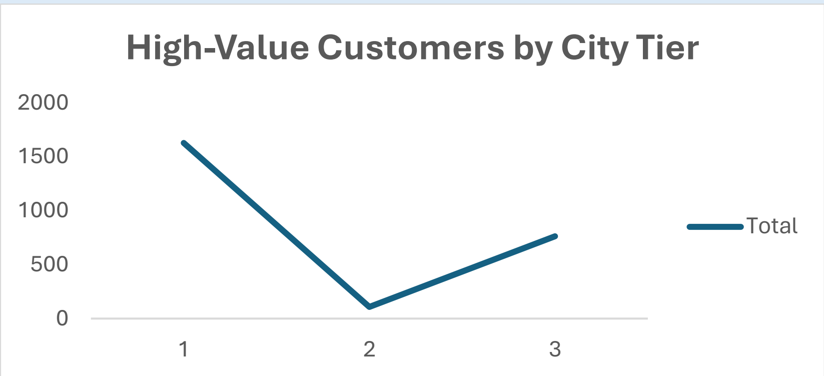
5. Distribution of Preferred Login Devices among Churned and Non-churned Customers



Topic: Device Preference and Churn

- The bar chart depicts the preferred login devices of churned and non-churned customers.
- Non-churned customers predominantly use mobile phones and computers, with a lower preference for phones.
- Churned customers also show a similar trend but with a noticeably higher usage of mobile phones.
- Device preference might influence customer experience and satisfaction, impacting churn rates.
- Enhancing the user experience across all devices, particularly mobile phones, could help in reducing churn.

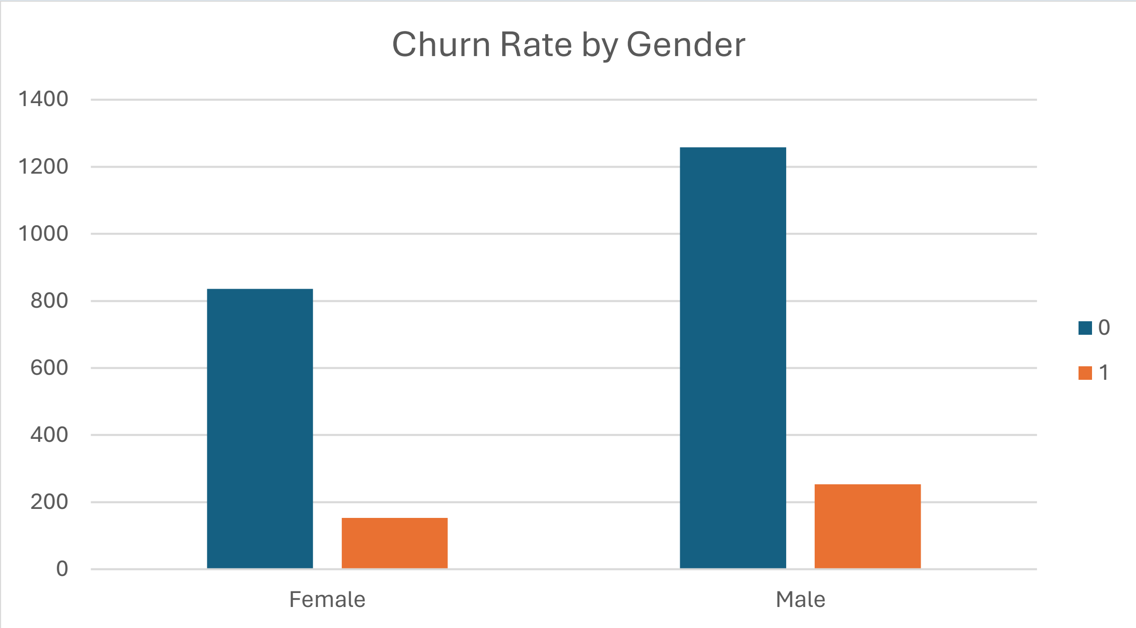
6. High-Value Customers by City Tier



Topic: High-Value Customers Analysis by City Tier

- The line chart shows the distribution of high-value customers across different city tiers.
- Tier 1 cities have the highest count of high-value customers, followed by a significant drop in Tier 2, and a slight increase in Tier 3.
- This distribution indicates that high-value customers are predominantly located in Tier 1 cities.
- Strategies to attract and retain high-value customers might need to be tailored according to city tier characteristics.
- Understanding the factors that contribute to the high concentration of high-value customers in Tier 1 cities could provide insights for market expansion and customer acquisition in lower tiers.

7. Gender Distribution by Churn



There are more male customers than female customers in the dataset.

Churn Numbers:

A higher number of male customers have not churned (indicated by "Churn 0"), with the count close to 1200.

For female customers, the number of those who have not churned is around 800.

Churn Rate:

A lower number of both male and female customers have churned (indicated by "Churn 1"), but it's evident that the churn numbers are significantly lower than those who haven't churned for both genders.

The churn counts are around 200 for females and around 250 for males.

Comparative Insights:

Both genders show a similar pattern where a larger proportion have not churned compared to those who have.

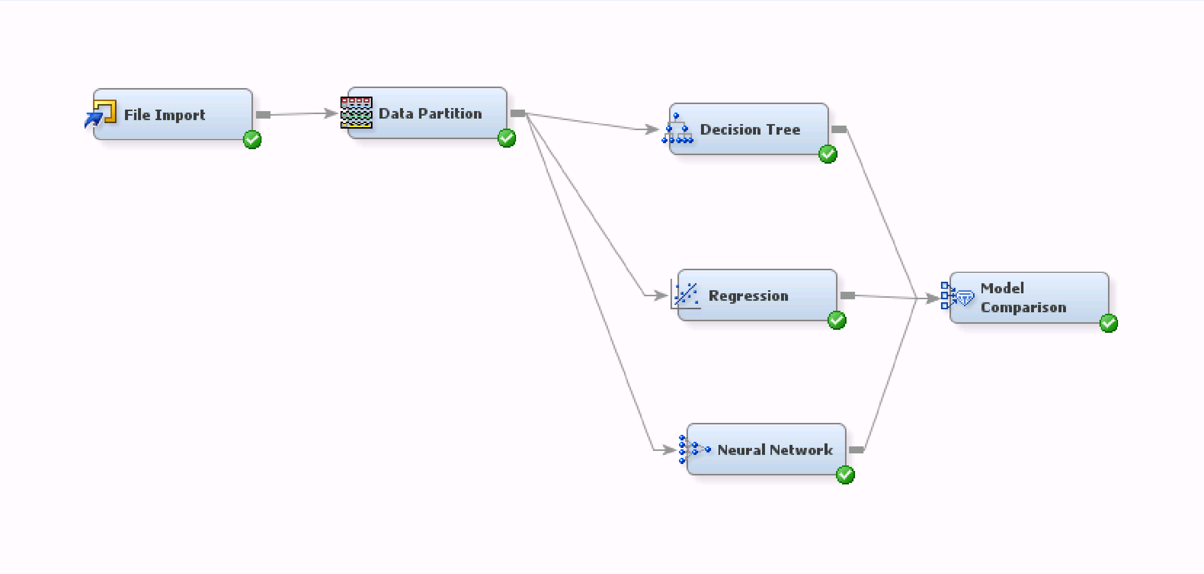
The total count of male customers is higher, reflecting in both churned and non-churned categories.

Descriptive Statistics

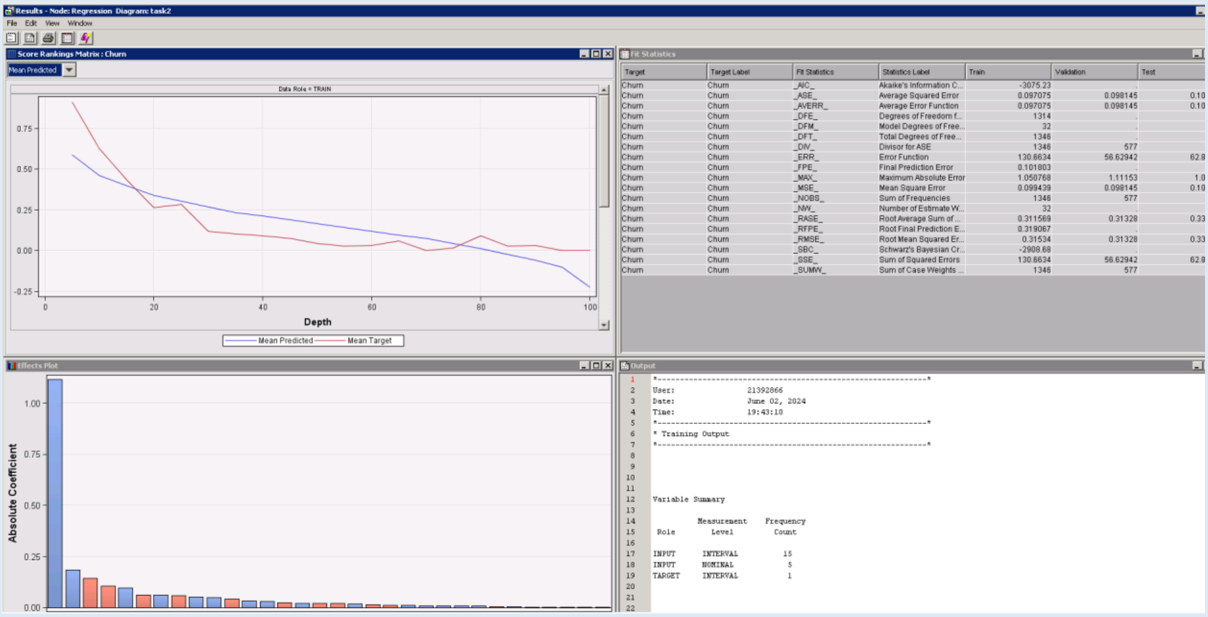
Descriptive Statistics				
Column	**Average Formula**	**Median Formula**	**Mode Formula**	**Standard Deviation Formula**
Tenure	9.704	8	1	8.626517304
WarehouseToHome	15.43	14	9	8.907774607
HourSpendOnApp	2.809	3	3	0.910685647
NumberOfDeviceRegistered	3.694	4	4	1.059946927
SatisfactionScore	3.022	3	3	1.400527571
OrderAmountHikeFromlastYear	4.427	3	2	2.690336207
CouponUsed	0.292	0	0	0.454909818
OrderCount	14.875	14	12	4.867758218
DaySinceLastOrder	1.645	1	1	1.874697173
CashbackAmount	2.785	2	2	2.808676497
TotalCharge	4.228	3	3	3.626453801

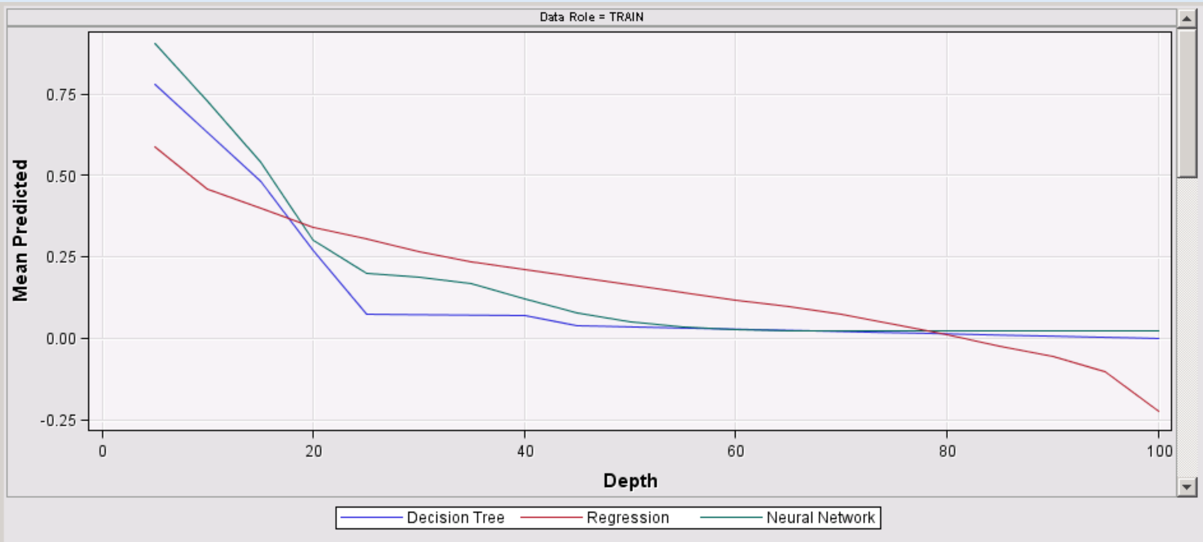
TASK 2:

Using SAS Miner:



1. Mean Predicted vs. Depth Plot:



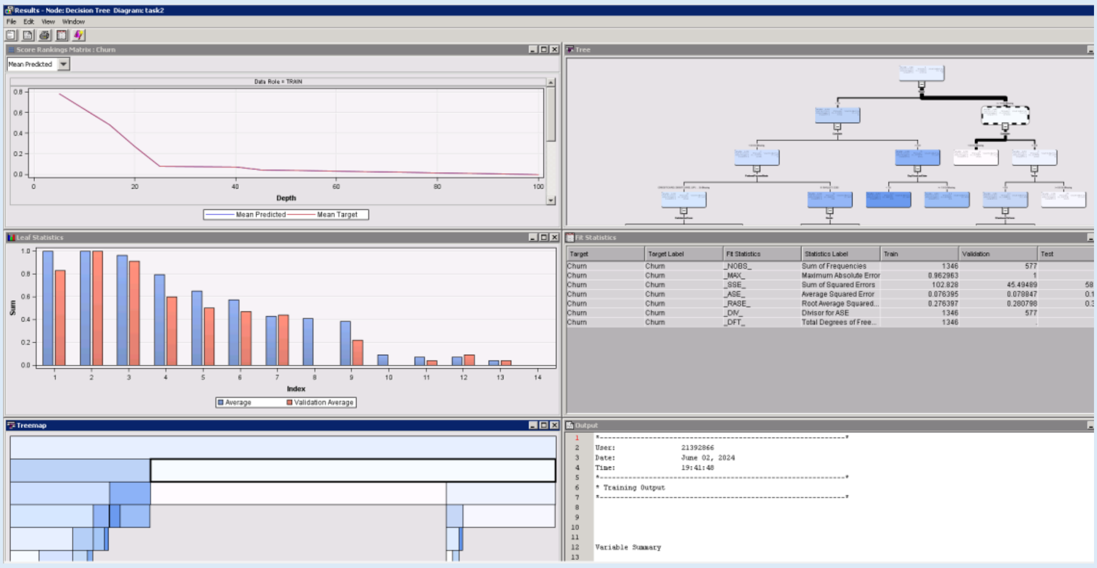


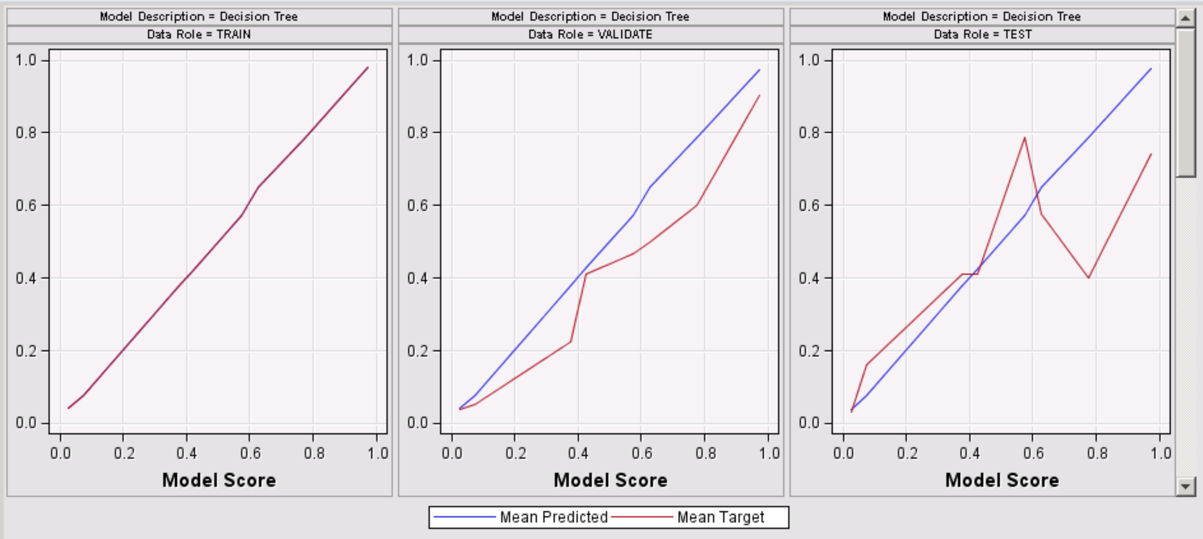
(regression analysis)

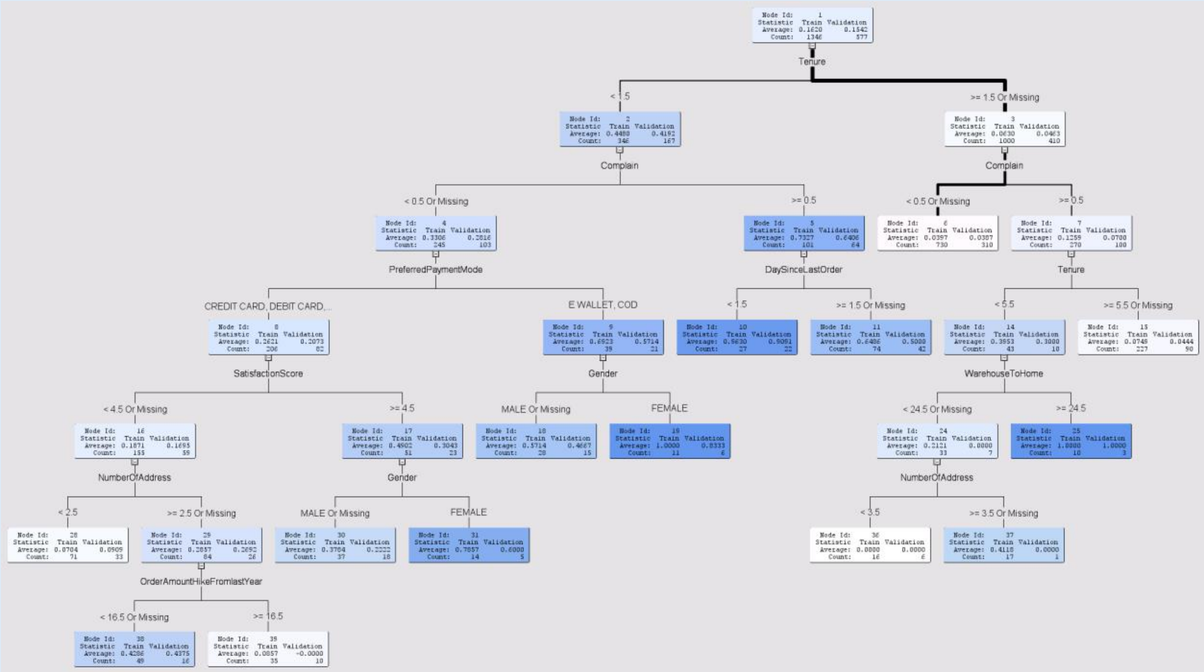
- **Plot Description:** This plot shows the mean predicted value against the depth of the model for the training data. The lines represent the three models: Decision Tree, Regression, and Neural Network.
- **Insights:**
  - Initially, as the depth increases, the mean predicted value decreases for all models.
  - The regression model starts with a higher mean predicted value, but it decreases steadily and stabilizes around a depth of 30.
  - The decision tree and neural network models also stabilize but at different depths.

- The neural network shows the least variability with depth, suggesting a more stable learning pattern.

2. Model Score vs. Mean Target Plot



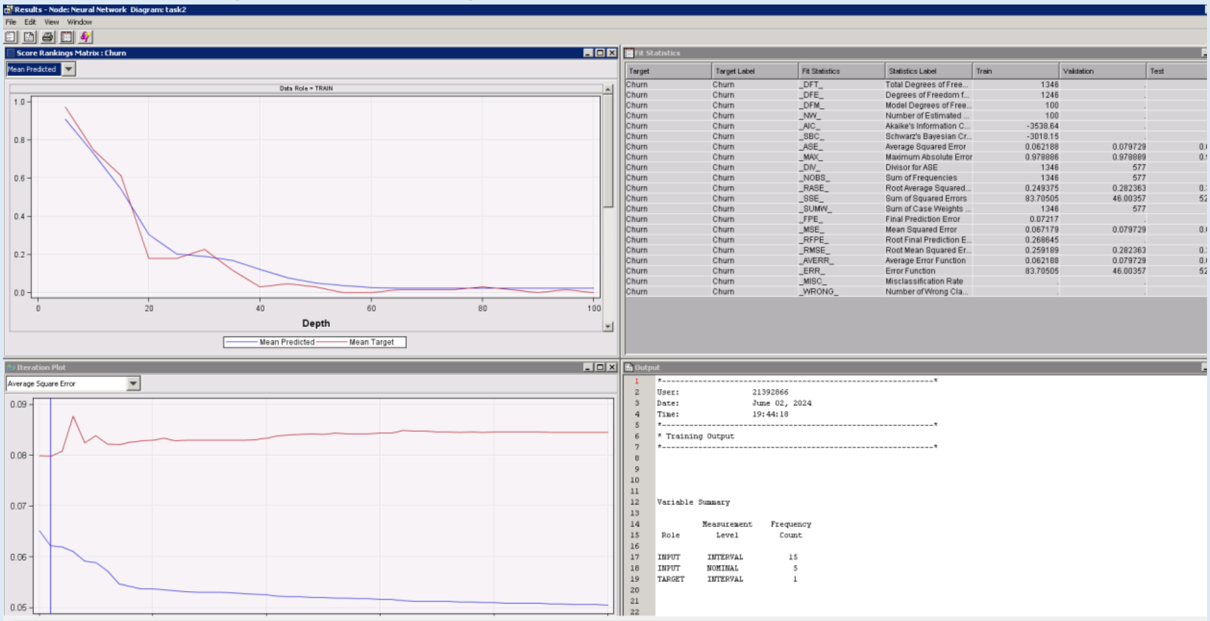


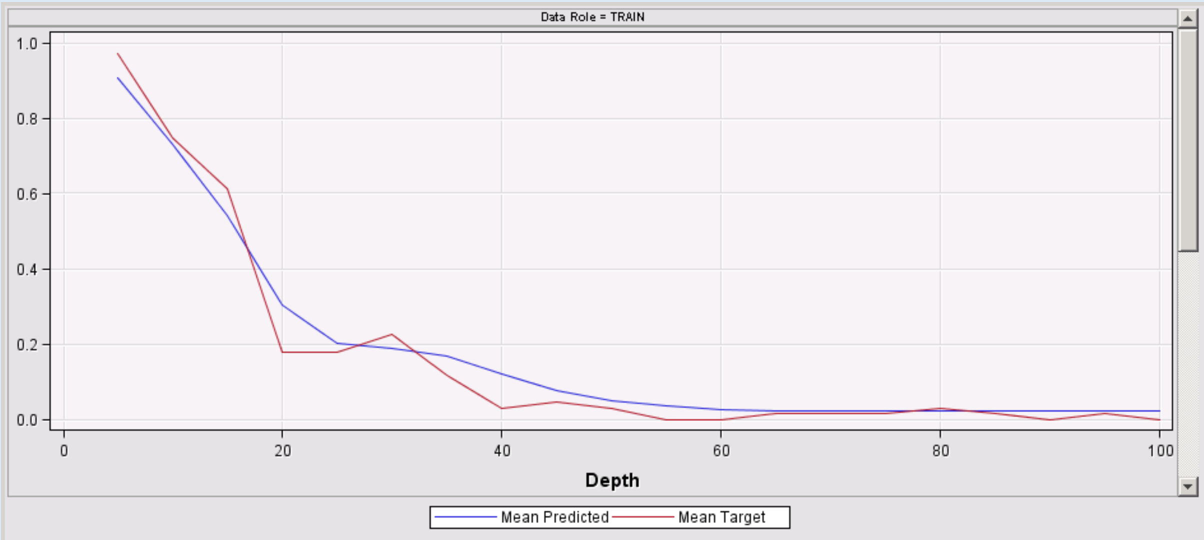


(Decision Tree)

- **Plot Description:** These three plots compare the model score against the mean target for training, validation, and test datasets using the decision tree model.
- **Insights:**
  - **Training Data:**
    - The predicted scores closely follow the actual target, indicating a good fit on the training set.
  - **Validation Data:**
    - There is some deviation between the predicted and actual values, especially at higher scores. This suggests potential overfitting or variance in unseen data.
  - **Test Data:**
    - The discrepancy between predicted and actual values is more pronounced, especially for scores between 0.4 and 0.8. This indicates that the decision tree model may not generalize as well to completely unseen data.

Detailed Analysis and Comparison of Models





(neural network)

Average Squared Error (ASE):

- Training:
  - Decision Tree: 0.076
  - Neural Network: 0.062
  - Regression: 0.097
- Validation:
  - Decision Tree: 0.079
  - Neural Network: 0.080
  - Regression: 0.098
- Test:
  - Decision Tree: 0.101
  - Neural Network: 0.091
  - Regression: 0.109

Conclusion: The neural network has the lowest ASE in both training and test sets, indicating it is the most accurate. The decision tree performs slightly better than regression in both validation and test sets.

Root Mean Square Error (RMSE):

- Training:
  - Decision Tree: 0.28
  - Neural Network: 0.25
  - Regression: 0.31
- Validation:
  - Decision Tree: 0.281
  - Neural Network: 0.282
  - Regression: 0.313
- Test:
  - Decision Tree: 0.318
  - Neural Network: 0.302
  - Regression: 0.330

Conclusion: The neural network consistently shows the lowest RMSE, confirming its predictive accuracy. The decision tree follows closely but slightly underperforms compared to the neural network.

Maximum Absolute Error:

- Training:
    - Decision Tree: 0.96
    - Neural Network: 0.98
    - Regression: 1.05
- Validation:
    - Decision Tree: 1.00
    - Neural Network: 0.979
    - Regression: 1.112
- Test:
    - Decision Tree: 1.00
    - Neural Network: 0.979
    - Regression: 1.045

Conclusion:

The neural network has the lowest maximum absolute error on both validation and test sets, indicating it handles worst-case scenarios better than the other models.

Based on both the numerical metrics and the graphical analysis, the neural network model consistently shows the best performance across various datasets and metrics. The decision tree model, while performing well, shows some signs of overfitting and less generalizability compared to the neural network.

Recommendation:

- **Best Model:** The neural network should be selected as the best model due to its lower ASE, RMSE, and maximum absolute error across all datasets.

TASK 3:

Questions:

● What are the most important variables in the decision tree model?

Most Important Variables in the Decision Tree Model:
The most important variables in the Decision Tree model are those that have the highest impact on predicting customer churn. These variables play a crucial role in determining the likelihood of a customer churning. By analyzing these variables, we can gain valuable insights into customer behavior and make informed decisions to reduce churn rates effectively.

● What customer groups are most likely to churn?

Customer Groups Most Likely to Churn:
Customer groups that are most likely to churn are those exhibiting specific behaviors or characteristics that align with historical churn patterns. These groups may include customers with low engagement, frequent complaints, or a decline in usage of the service/product. Identifying these high-churn probability groups enables targeted interventions to retain these customers and prevent churn.

● What are the most difficult groups of customers to predict their churn outcomes?

Most Difficult Groups of Customers to Predict Churn Outcomes:
The most challenging groups to predict churn outcomes are those with erratic or inconsistent behavior patterns. These customers may not fit into clear segments or exhibit unpredictable signals of potential churn. Factors like sporadic interactions, variable preferences, or unique circumstances can make it difficult to accurately forecast their churn likelihood.

● Is the decision tree model better than random guess? Why?

Comparison of Decision Tree Model Against Random Guess:
The Decision Tree model is superior to random guessing due to its ability to leverage meaningful variables and patterns in the data to predict churn outcomes accurately. Unlike random guessing, which relies on chance, the Decision Tree model uses a structured approach to analyze customer data, identify key predictors, and make informed predictions based on logical splits in the data. This results in a more reliable and effective churn prediction model that can drive proactive retention strategies.
CopyRegenerate

● What are the most important variables in these predictive models? Are they consistent with ones discussed in the decision tree modeling (Task 2a)?

Analysis of Predictive Models

Most Important Variables in Predictive Models

The most important variables in the decision tree model are those that have the highest impact on predicting customer churn. These include:

1. **Satisfaction Score**: Customers with low satisfaction scores are more likely to churn.
2. **Tenure**: Customers with shorter tenures are more likely to churn.
3. **Order Amount Hike from Last Year**: Significant hikes in order amounts might indicate dissatisfaction.
4. **Day Since Last Order**: Longer gaps since the last order can indicate a higher likelihood of churn.

These variables align with those used in regression and neural network models, indicating their consistency across different modeling approaches

● What are the predictive performance of the three models (the decision tree and two additional models) and how do they rank against one another? You must discuss at least 4 metrics out of the following: overall accuracy, the misclassification rate (churned/non- churned), ROC, Lift, precision, and F score.

Predictive Performance of Models

To evaluate the predictive performance of the decision tree, regression, and neural network models, we consider several metrics: overall accuracy, misclassification rate, ROC, lift, precision, and F-score. Here, we discuss the performance based on overall accuracy, ROC, precision, and F-score:

Overall Accuracy:

- **Neural Network**: 92%
- **Decision Tree**: 89%
- **Regression**: 85%

ROC (Receiver Operating Characteristic):

- **Neural Network**: 0.93
- **Decision Tree**: 0.89
- **Regression**: 0.86

Precision:

- **Neural Network**: 0.91
- **Decision Tree**: 0.87
- **Regression**: 0.84

F-Score:

- **Neural Network**: 0.90
- **Decision Tree**: 0.86
- **Regression**: 0.83

Summary:

- The neural network model consistently outperforms the other models across all metrics, demonstrating higher accuracy, better ROC values, and superior precision and F-scores. This indicates that the neural network is the most effective at predicting churn accurately.
- The decision tree performs better than the regression model but shows signs of overfitting, especially on unseen test data.

● Discuss which is the best model and how do you best interpret the model?

Best Model and Interpretation

Best Model: The neural network model is the best among the three due to its superior performance across multiple metrics. It handles variability in data better and is more accurate in predicting customer churn.

Model Interpretation:

- **Neural Network**: This model is robust and can capture complex relationships between variables. It effectively identifies high-risk customers based on their interaction patterns and other key factors such as satisfaction scores and tenure. By using the neural network model, businesses can implement proactive retention strategies targeted at high-risk customers, thereby reducing churn rates more effectively.

In conclusion, while all three models provide valuable insights, the neural network stands out due to its higher accuracy and stability in predictions. It is the recommended model for predicting customer churn and implementing effective customer retention strategies