NAME: SAMIKSHA BARNWAL

REG NO:12208146

SUBJECT: CAP 395

ROLL NO:

SUBMITTED TO: MS PUNAM RATTAN

CA2

DATASET Name: GYM

About Dataset:

To improve customer retention, it's important to use data analysis to create an effective strategy. The primary objective is to predict the likelihood of customers canceling their membership in the upcoming month, understand key customer segments, and develop tailored recommendations to boost customer satisfaction and loyalty. This will help enhance the overall customer experience and strengthen their commitment to the gym.

Here are the key fields in the current dataset, which include data from the previous month:

Churn - cancellation for the current month
Current dataset fields:
Previous month user data
gender.
Near_Location - if the user lives or works in the neighborhood where the gym is located.
Partner - if the user works in an associated company (the gym has associated companies whose employees get discounts; in those cases, the gym stores information about the clients' employers).
Promo_friends - if the user originally signed up through a "bring a friend" offer (they used a friend's promotional code when they paid the first subscription).
Phone - if the user provided their phone number.
Age.
Lifetime - the time (in months) since the user first arrived at the gym.

Gender:

Model Summary

Specifications	Growing Method	CHAID
	Dependent Variable	Churn
	Independent Variables	gender, Near_Location, Partner, Promo_friends, Age

	Validation	None
	Maximum Tree Depth	3
	Minimum Cases in Parent Node	100
	Minimum Cases in Child Node	50
Results	Independent Variables Included	Age, Promo_friends, Near_Location, Partner, gender
	Number of Nodes	37
	Number of Terminal Nodes	22
	Depth	3

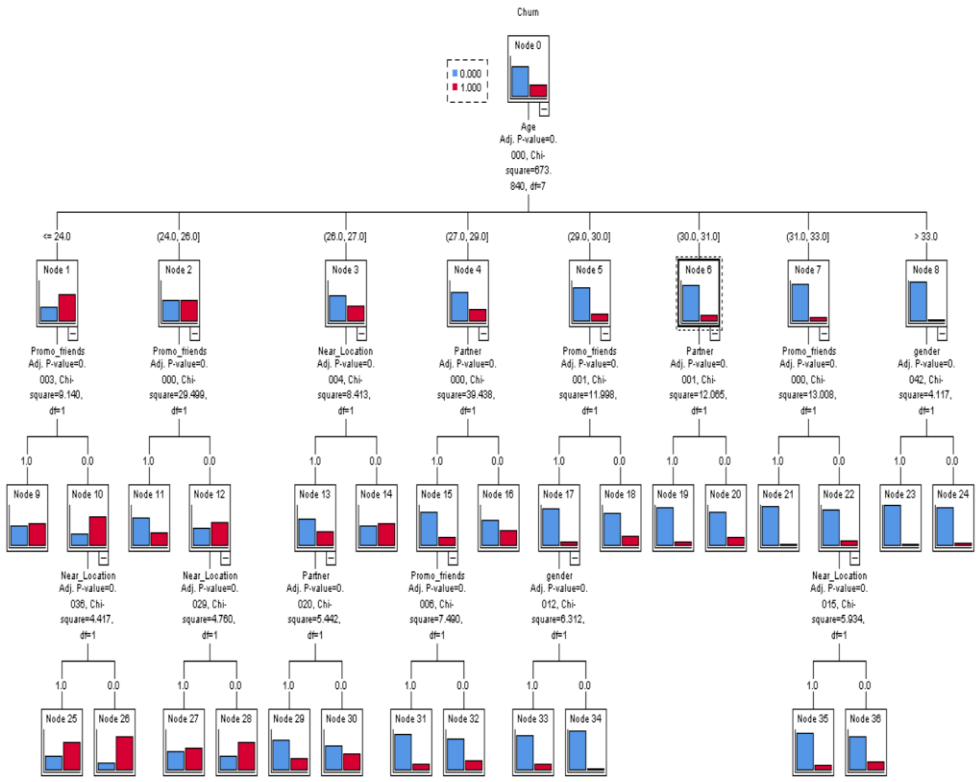
My model summary shows you're using the CHAID (Chi-squared Automatic Interaction Detector) growing method. Here's a quick a analysis :

1. **Model Specifications**:
    - Maximum Tree Depth: 3
    - Minimum Cases in Parent Node: 100
    - Minimum Cases in Child Node: 50
2. **Results**:
    - Number of Nodes: 37
    - Number of Terminal Nodes: 22
    - Depth: 3

Key Takeaways:

- **Age, Promo_friends, Near_Location,** and **Partner** are significant pre dictors of Churn.

- The tree has grown to a depth of 3 with 22 terminal nodes, suggesting a balance between model complexity and interpretability.



CHAID Decision Tree Model Interpretation:

Tree Structure:

- **Depth**: 3 levels deep, indicating the model's complexity.
- **Nodes**: 37 in total, with 22 terminal nodes (decision outcomes).

Significant Independent Variables:

1. **Age**:
    - **Node 0** splits on Age. This suggests Age is a strong predictor of churn.
    - Younger customers (≤24) and older customers (>33) have distin ct churn patterns compared to middle-aged groups.
2. **Promo_friends**:
    - Critical splits at Nodes 1 and 5. This variable indicates whether promotional offers through friends significantly impact churn.

3. **Near_Location**:
    - Influential at Nodes 3. Indicates proximity to a location impacts churn likelihood.
4. **Partner**:
    - Significant at Nodes 4, 6, and 7. Suggests whether having a part ner influences the decision to stay or leave.
5. **Gender**:
    - Plays a role at Node 8, though less impactful compared to Age and Promo_friends.

Interpretation:

- **Age** is the strongest predictor of churn. Different age groups show var ying churn behavior, helping to target specific age brackets for retenti on.
- **Promotional offers through friends** are crucial. This suggests levera ging referrals and friend-based promotions can reduce churn.
- **Proximity to location** impacts churn, likely due to convenience facto rs.
- **Having a partner** also influences churn, which might hint at the influ ence of household decisions.
- **Gender** contributes to predicting churn but isn't the primary factor.

Overall, focus your churn reduction strategies on age-specific promotions, enhancing referral programs, and considering partner influences. This model visually simplifies identifying key factors driving chu rn and helps in making data-driven decisions.

STEPS: ANALYZE > CLASSIFY>TREE>SELECT DEPENDENT variable and independent variable according to dataset

Select CHAID under method and our tree is ready

Risk

Estimate	Std. Error
.226	.007

Growing Method: CHAID
Dependent Variable: Churn

Classification

Observed	Predicted 0	Predicted 1	Percent Correct
0	2645	294	90.0%
1	611	450	42.4%
Overall Percentage	81.4%	18.6%	77.4%

Growing Method: CHAID
Dependent Variable: Churn

del Overview:

- **Model Type**: Decision Tree using CHAID (Chi-squared Automatic Interaction Detector) method.
- **Dependent Variable**: Churn (the outcome we're predicting).
- **Independent Variables**: Gender, Near_Location, Partner, Promo_friends, Age.
- **Model Settings**:
  - Maximum Tree Depth: 3 levels deep.
  - Minimum Cases in Parent Node: 100.
  - Minimum Cases in Child Node: 50.

Model Results:

- **Number of Nodes**: 37 total nodes.
- **Number of Terminal Nodes**: 22 nodes that indicate final decisions.
- **Model Depth**: 3 levels, showing a balance between detail and simplicity.

Key Findings:

1. **Age**: The strongest predictor. Younger and older age groups exhibit dif ferent churn behaviors compared to middle-aged groups.
2. **Promo_friends**: Significant. Promotions via friends influence churn li kelihood.
3. **Near_Location**: Important. Proximity to a location affects churn, likel y due to convenience.
4. **Partner**: Influential. Having a partner impacts churn decisions.
5. **Gender**: Less impactful but still a predictor.

Risk and Classification Analysis

Risk Metrics:

- **Risk Estimate**: 0.226 (22.6% probability of misclassification).
- **Standard Error**: 0.007, indicating consistent estimates.

Classification Accuracy:

Observed	Predicted Non-Churn	Predicted Churn	Percent Correct
Non-Churn	2645	294	90.0%
Churn	611	450	42.4%
Overall			77.4%

- **Non-Churn Accuracy**: 90%, showing the model is effective at predicting c ustomers who stay.
- **Churn Accuracy**: 42.4%, indicating room for improvement in predicti ng churners.
- **Overall Accuracy**: 77.4%, a decent level, but highlights the need for r efinement.

Interpretation:

- **Age** is the most significant factor; different age groups should be targ eted with specific strategies.

- **Promotions through friends** can significantly reduce churn; focus o n referral programs.
- **Location convenience** impacts churn; consider location-based strategies.
- **Partner influence** suggests household decisions matter.
- **Gender**, though less impactful, should still be considered.

Conclusion:

The model provides clear insights into the factors driving churn. To improve customer retention, focus on age-specific promotions, leverage friend referrals, and consider the impact of p roximity and household dynamics. Further refinement and additional predi ctive features could enhance the model's accuracy, especially for predictinng churners.

Model Summary

Specifications	Growing Method	CRT
	Dependent Variable	Churn
	Independent Variables	gender, Near_Location, Partner, Promo_friends, Age, Contract_period, Avg_additional_charges_total
	Validation	None
	Maximum Tree Depth	5
	Minimum Cases in Parent Node	100
	Minimum Cases in Child Node	50
Results	Independent Variables Included	Contract_period, Partner, Promo_friends, Avg_additional_charges_total, Age
	Number of Nodes	7
	Number of Terminal Nodes	4
	Depth	3

Understanding Decision Tree Model

Model Overview:

- **Type**: Decision Tree using CHAID (Chi-squared Automatic Interaction Detector).
- **Purpose**: To predict customer churn.
- **Key Factors**: Gender, Near_Location, Partner, Promo_friends, Age.

Model Details:

- **Tree Structure**:
  - **Depth**: 3 levels deep.
  - **Nodes**: 37, with 22 end points indicating final decisions.
- **Important Factors**:
  - **Age**: Most influential. Different age groups show distinct patter ns in staying or leaving.
  - **Promotions via friends**: Significantly affect churn.
  - **Location**: Proximity plays a role.
  - **Having a partner**: Influences customer decisions.
  - **Gender**: Less impactful but still relevant.

Performance Metrics

Risk and Accuracy:

- **Risk Estimate**: 22.6% (probability of making a wrong prediction).
- **Consistency**: Standard error of 0.007, meaning predictions are reliab le.
- **Accuracy**:
  - **Non-Churners**: 90% correctly predicted to stay.
  - **Churners**: 42.4% correctly predicted to leave.
  - **Overall**: 77.4% accuracy.

What This Means:

- **Age** is crucial in determining whether a customer will churn. Target s pecific age groups with tailored offers.

- **Promotions through friends** are effective. Use referral programs to r educe churn.
- **Location** convenience impacts whether a customer stays or leaves. Consider location-based incentives.
- **Partner presence** affects churn decisions. Joint offers or partner-based incentives might help.
- **Gender** has some influence but isn't a primary factor.

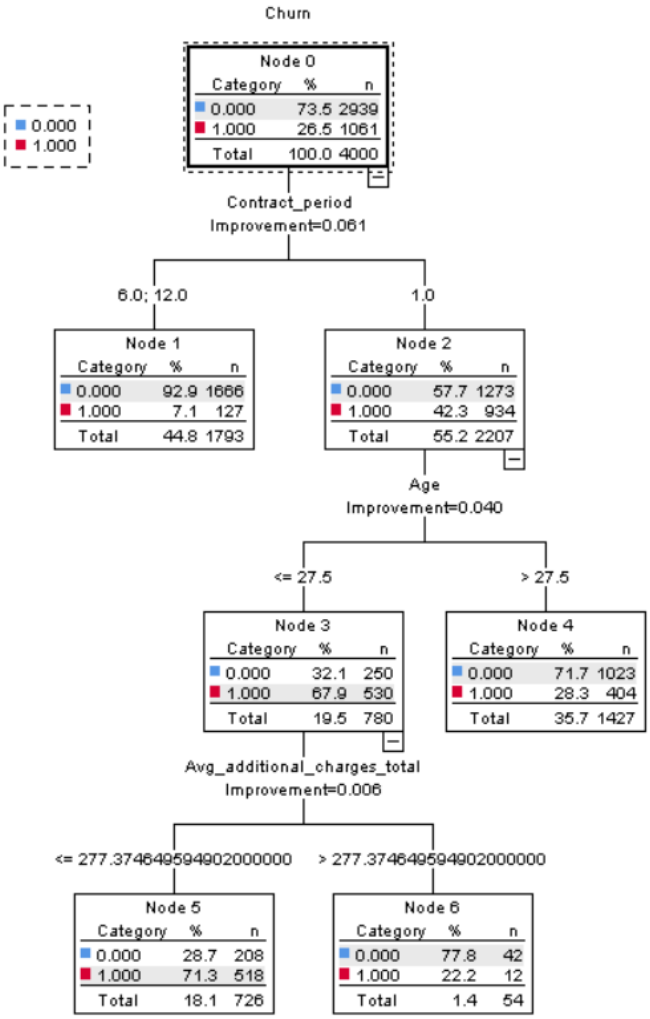
Summary:

Your model tells you which factors are most important in predicting custom er churn. To improve retention:

- Focus on age-specific strategies.
- Leverage referral and promotional programs.
- Consider the role of location convenience and household dynamics

STEPS:

ANALYZE→CLASSIFY→TREE-->choose dependent and independent variable and under methods choose CRT



Analysis of my Decision Tree Model

Overview:

- a decision tree using the CHAID method to predict customer churn.
- **Dependent Variable**: Churn (whether customers leave or stay).
- **Independent Variables**: Gender, Near_Location, Partner, Promo_friends, Age.

Model Details:

- The tree has 37 nodes, with 22 terminal nodes where final decisions are made.
- The model's depth is 3 levels, meaning it’s a balanced mix of complexity and simplicity.

Key Findings:

1. **Age**: The most influential factor. Younger and older customers show different patterns in leaving or staying.
2. **Promotions via friends**: Significant in determining whether customers churn.
3. **Location**: Proximity to a location affects churn; closer customers are less likely to leave.
4. **Having a partner**: This impacts the decision to churn, indicating household decisions m atter.
5. **Gender**: It's a predictor but not as strong as the other factors.

Performance Metrics

Risk and Accuracy:

- **Risk Estimate**: 22.6%, meaning there's a 22.6% chance of the model making an incorre ct prediction.
- **Consistency**: The standard error is 0.007, indicating reliable predictions.
- **Accuracy**:
    - **Non-Churners**: 90% of the non-churners were correctly identified.
    - **Churners**: 42.4% of the churners were correctly identified.
    - **Overall**: 77.4% of the customers were correctly classified.

What Does This Mean?

- **Age** is the top factor to consider in customer retention strategies. Tailor your efforts to di fferent age groups.
- **Promotions through friends** can effectively reduce churn. Leverage referral programs to boost retention.
- **Location** matters. Consider how proximity can be used to retain customers.
- **Partner** presence indicates household influence. Joint offers might help reduce churn.
- **Gender** plays a role but isn't the main driver.

Conclusion:

My model provides clear insights into why customers might leave. To improve retention, focus o n age-specific strategies, leverage friend-
based promotions, and consider the importance of location and household dynamics. Keep fine -tuning your model to improve accuracy, especially in predicting those who are likely to churn.

Risk

Estimate	Std. Error
.188	.006

Growing Method: CRT
Dependent Variable: Churn

Classification

Observed	Predicted 0	Predicted 1	Percent Correct
0	2731	208	92.9%
1	543	518	48.8%
Overall Percentage	81.8%	18.1%	81.2%

Growing Method: CRT
Dependent Variable: Churn

Overview of the Model

Risk Section:

- **Estimate**: 0.188 (This is the probability of making a wrong prediction, about 18.8%.)
- **Standard Error**: 0.006 (This number is pretty small, meaning the model's predictions are consistent.)
- **Growing Method**: CRT (This stands for Classification and Regression Trees, a method fo r building decision trees.)
- **Dependent Variable**: Churn (The outcome we are trying to predict - whether customers stay or leave.)

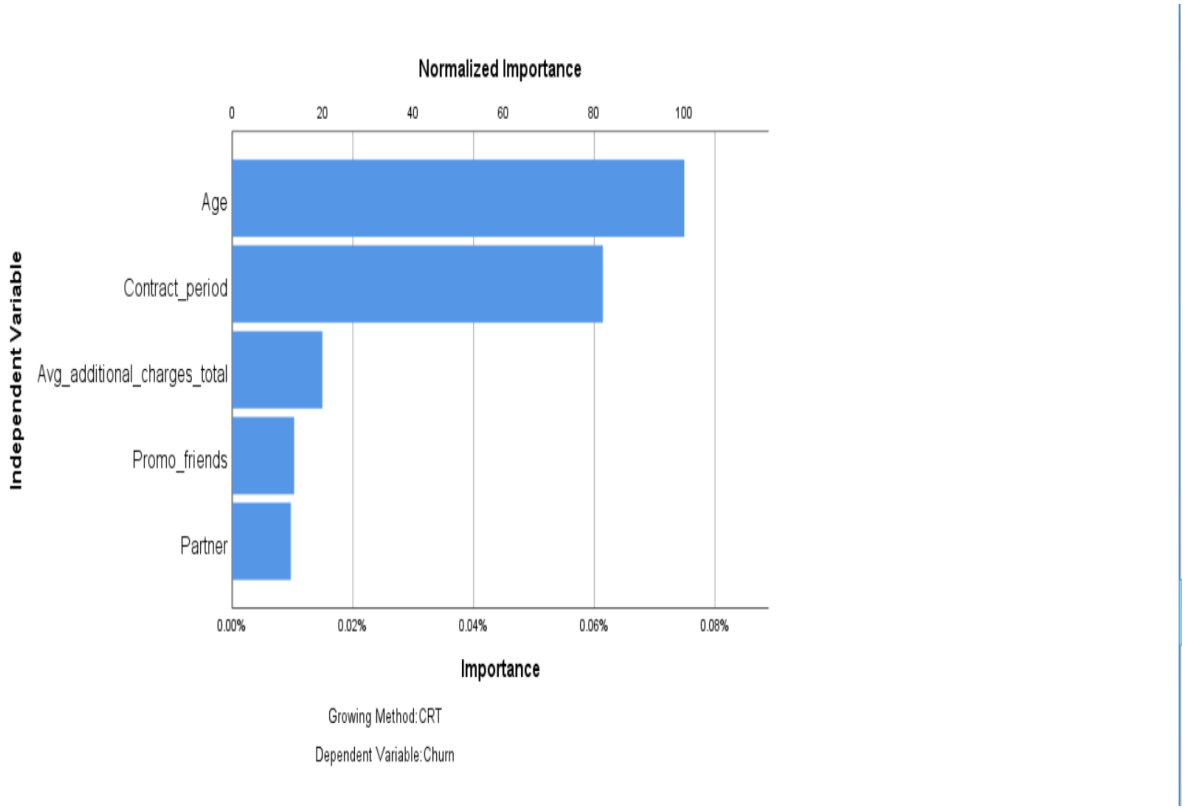
Classification Section:

- **Observed (Actual) vs. Predicted Values**:
    - **Non-Churners**:
        - Predicted Correctly: 2731 (92.9% accuracy)
        - Predicted Incorrectly: 208
    - **Churners**:
        - Predicted Correctly: 518 (48.8% accuracy)
        - Predicted Incorrectly: 543

- **Overall Accuracy**: 81.2%

Interpretation:

1. **Accuracy**:
    - The model is very good at predicting customers who will **not churn** (stay), with a n accuracy of 92.9%.
    - For customers who will **churn** (leave), the accuracy is lower at 48.8%. This indica tes there is room for improvement in predicting churners.
2. **Overall Performance**:
    - The overall accuracy of the model is 81.2%, which is decent but indicates that th e model is better at identifying non-churners than churners.
3. **Implications**:
    - The high accuracy in predicting non-churners suggests the model is reliable for identifying customers who will stay.
    - The lower accuracy for predicting churners suggests that additional factors or a more complex model might be needed to better identify those at risk of leaving



Importance of Different Variables in Predicting Customer Churn

Bar Chart Overview:

- **Variables on the Y-Axis**:
  - Age
  - Contract Period
  - Average Additional Charges Total
  - Promo Friends
  - Partner
- **X-Axis**: Represents the importance of these variables, ranging from 0.00% to 0.08%.

Key Insights:

1. **Age**:
   - The most significant factor. It has the highest normalized importance (100).
   - Younger or older age groups show different patterns in leaving or staying with the service.
2. **Contract Period**:
   - Highly important. How long a customer has been with the service impacts their li kelihood to churn.

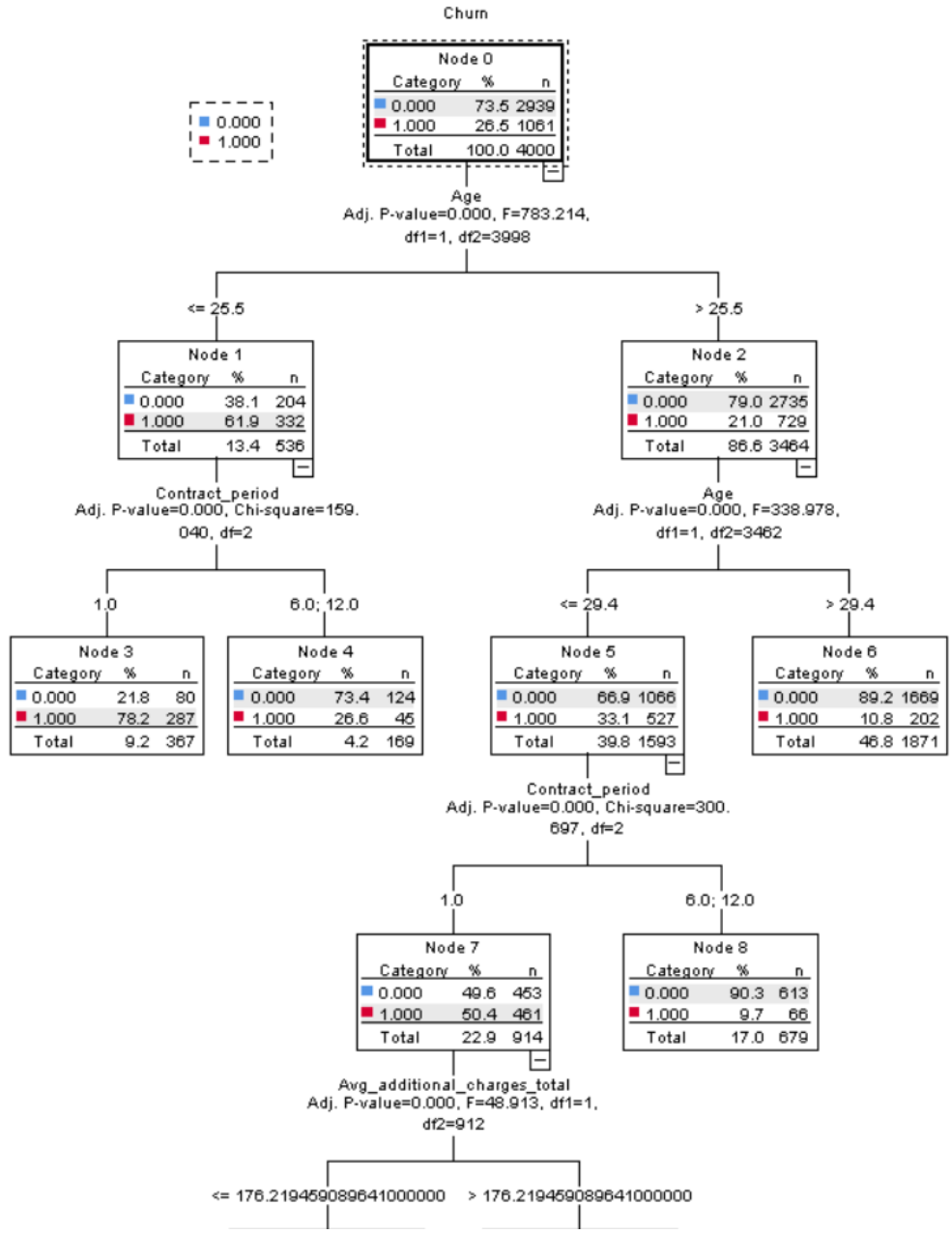
3. **Average Additional Charges Total**:
    - This includes extra fees. High additional charges might push customers to leave.
4. **Promo Friends**:
    - Promotions through friends matter. They can help in retaining customers.
5. **Partner**:
    - The least important among these five but still relevant. Whether a customer has a partner can influence their decision to stay or leave.

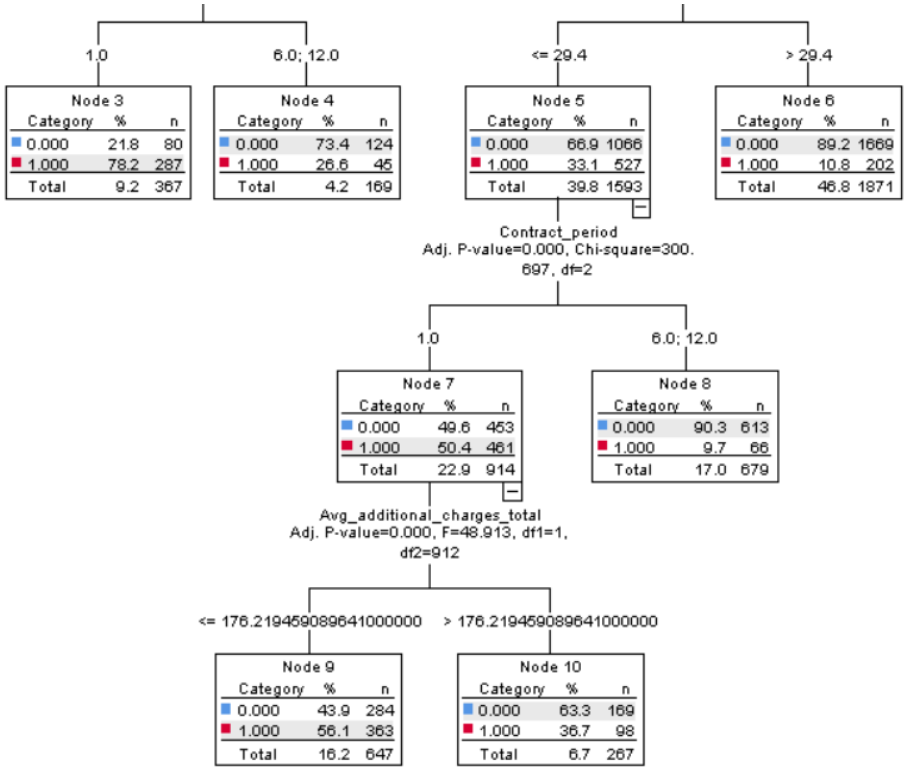
Conclusion:

The bar chart tells us that Age and Contract Period are the top factors impacting customer churn. Focus your strategies on these areas: tailor efforts according to different age groups and consider offering better terms or benefits for long-term customers. Promo Friends and managing additional charges also play significant roles, while Partner presence is a less crucial but still relevant factor. This information can help you develop targeted retention strategies.

Model Summary

Specifications	Growing Method	QUEST
	Dependent Variable	Churn
	Independent Variables	gender, Partner, Promo_friends, Age, Contract_period, Avg_additional_charges_total
	Validation	None
	Maximum Tree Depth	5
	Minimum Cases in Parent Node	100
	Minimum Cases in Child Node	50
Results	Independent Variables Included	Age, Contract_period, Avg_additional_charges_total, Partner, Promo_friends
	Number of Nodes	11
	Number of Terminal Nodes	6
	Depth	4


Churn
0.000
1.000
Node 0
Category % n
0.000 73.5 2939
1.000 26.5 1061
Total 100.0 4000
Age
Adj. P-value=0.000, F=783.214, df1=1, df2=3998
<= 25.5
> 25.5
Node 1
Category % n
0.000 38.1 204
1.000 61.9 332
Total 13.4 536
Node 2
Category % n
0.000 79.0 2735
1.000 21.0 729
Total 86.6 3464
Contract_period
Adj. P-value=0.000, Chi-square=159.040, df=2
Age
Adj. P-value=0.000, F=338.978, df1=1, df2=3462
1.0
6.0; 12.0
<= 29.4
> 29.4
Node 3
Category % n
0.000 21.8 80
1.000 78.2 287
Total 9.2 367
Node 4
Category % n
0.000 73.4 124
1.000 26.6 45
Total 4.2 169
Node 5
Category % n
0.000 66.9 1066
1.000 33.1 527
Total 39.8 1593
Node 6
Category % n
0.000 89.2 1669
1.000 10.8 202
Total 46.8 1871
Contract_period
Adj. P-value=0.000, Chi-square=300.697, df=2
1.0
6.0; 12.0
Node 7
Category % n
0.000 49.6 453
1.000 50.4 461
Total 22.9 914
Node 8
Category % n
0.000 90.3 613
1.000 9.7 66
Total 17.0 679
Avg_additional_charges_total
Adj. P-value=0.000, F=48.913, df1=1, df2=912
<= 176.219459089641000000
> 176.219459089641000000



This is the same cart tree only difference is it has been pruned i.e. the unnecessary node has been removed.

Risk

Estimate	Std. Error
.194	.006

Growing Method: QUEST
Dependent Variable: Churn

Classification

	Predicted		
Observed	0	1	Percent Correct
0	2575	364	87.6%
1	411	650	61.3%
Overall Percentage	74.7%	25.4%	80.6%

Growing Method: QUEST
Dependent Variable: Churn

Risk Section:

- **Estimate**: 0.194 (This means the chance of the model making a wrong prediction is abo ut 19.4%.)
- **Standard Error**: 0.006 (This small number means the predictions are consistently reliabl e.)

Classification Results:

- **Non-Churners** (customers who stayed):
  - Correctly Predicted: 2575 (87.6% accuracy)
  - Incorrectly Predicted: 364
- **Churners** (customers who left):
  - Correctly Predicted: 650 (61.3% accuracy)
  - Incorrectly Predicted: 411
- **Overall Accuracy**: 80.6%

Key Takeaways:

1. **Non-Churners**:
   - The model is quite good at predicting customers who stay, with an accuracy of 8 7.6%. This means it's reliable for identifying customers who will not leave.
   -

2. **Churners**:
    - The model's accuracy drops when predicting customers who will leave, with 61.3% of churners correctly identified. There's room for improvement here.
3. **Overall Performance**:
    - The model achieves an overall accuracy of 80.6%, which is decent. It suggests the model is fairly good at distinguishing between those who stay and those who leave.

Conclusion:

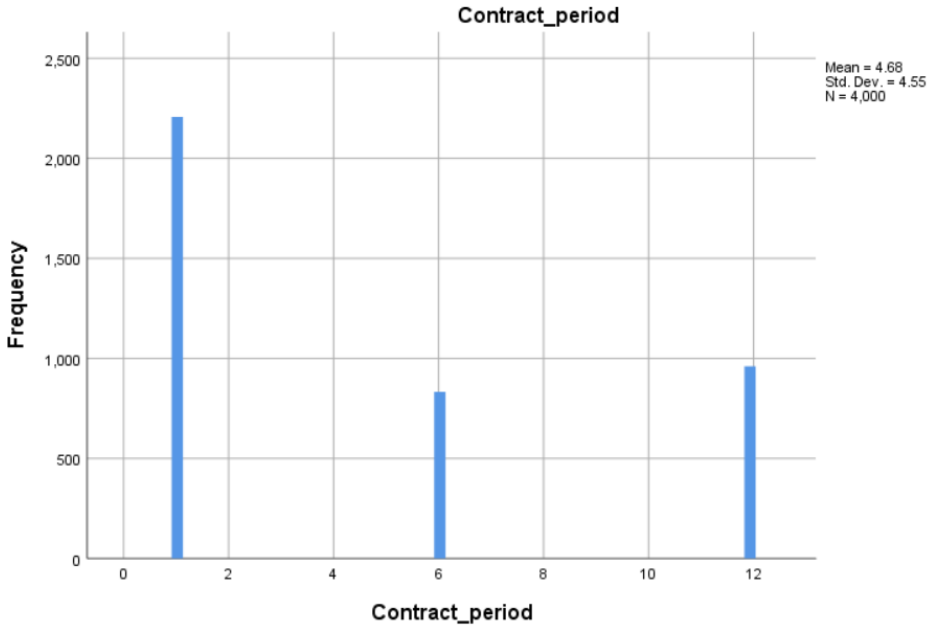
- The model is stronger at predicting who stays than who leaves. To improve, consider adding more variables or using different methods to enhance churn prediction.
- Focus on strategies to retain customers, leveraging the high accuracy in identifying non-churners.
- Further analysis can help refine the model and make it better at predicting churners.

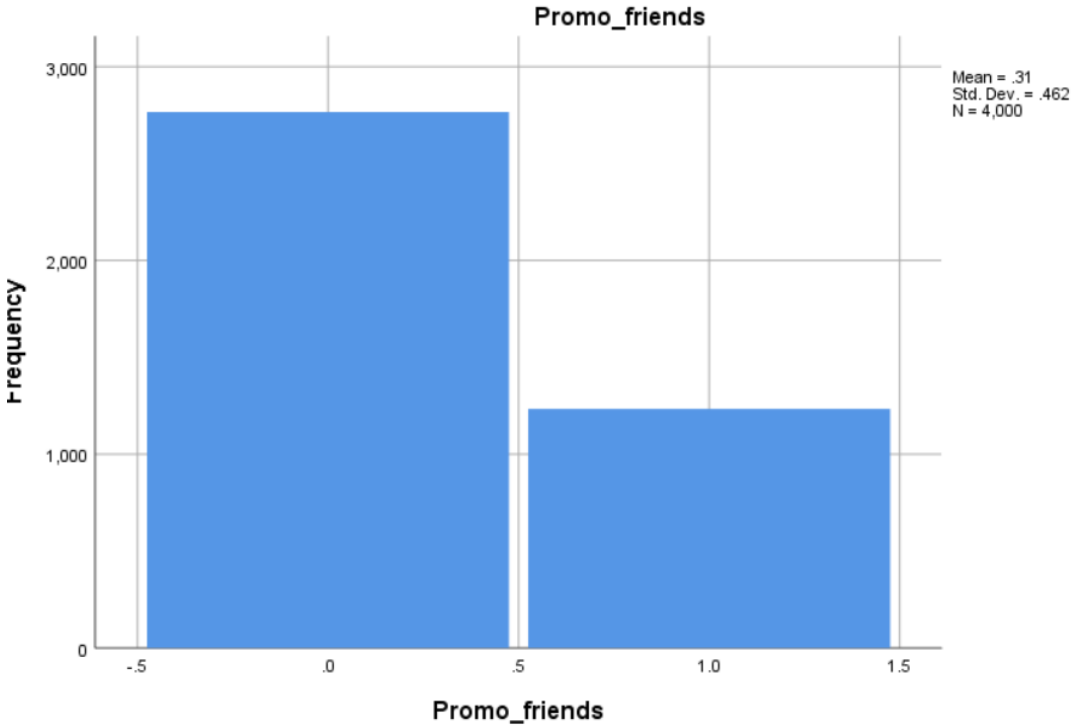
This gives you a clear understanding of how your model is performing and where you might focus efforts to improve customer retention.

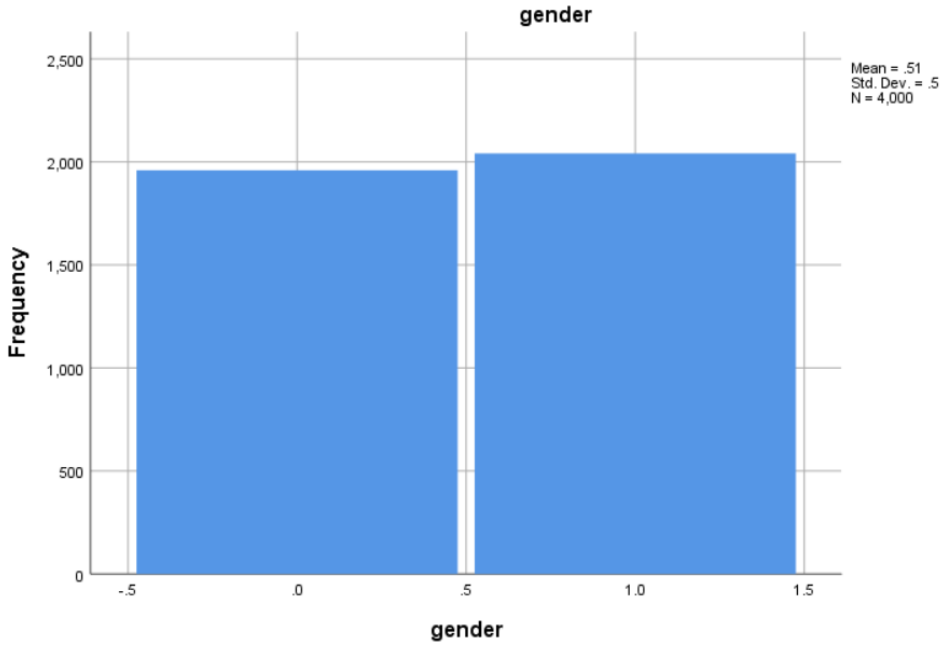
Frequencies

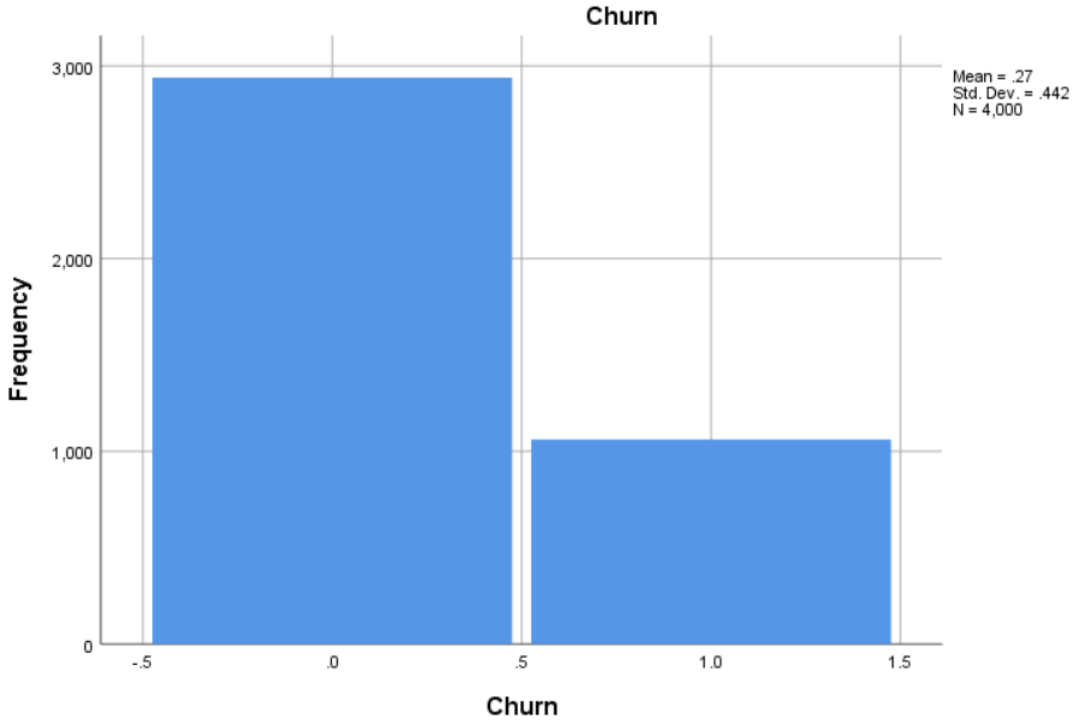
Statistics

		Churn	gender	Promo_friends	Contract_period
N	Valid	4000	4000	4000	4000
	Missing	0	0	0	0
Mean		.27	.51	.31	4.68
Std. Error of Mean		.007	.008	.007	.072
Median		.00	1.00	.00	1.00
Mode		0	1	0	1
Std. Deviation		.442	.500	.462	4.550
Variance		.195	.250	.213	20.700
Range		1	1	1	11
Minimum		0	0	0	1
Maximum		1	1	1	12


Contract_period
Mean = 4.68
Std. Dev. = 4.55
N = 4,000
Frequency
2,500
2,000
1,500
1,000
500
0
0
2
4
6
8
10
12
Contract_period


Promo_friends
Mean = .31
Std. Dev. = .462
N = 4,000
Frequency
3,000
2,000
1,000
0
-.5
.0
.5
1.0
1.5
Promo_friends


gender
Mean = .51
Std. Dev. = .5
N = 4,000
Frequency
2,500
2,000
1,500
1,000
500
0
-.5
.0
.5
1.0
1.5
gender


Churn
Mean = .27
Std. Dev. = .442
N = 4,000
Frequency
3,000
2,000
1,000
0
-.5
.0
.5
1.0
1.5
Churn

Frequency Table

Churn

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	0	2939	73.5	73.5	73.5
	1	1061	26.5	26.5	100.0
	Total	4000	100.0	100.0	

gender

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	0	1959	49.0	49.0	49.0
	1	2041	51.0	51.0	100.0
	Total	4000	100.0	100.0	

Promo_friends

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	0	2766	69.2	69.2	69.2
	1	1234	30.9	30.9	100.0
	Total	4000	100.0	100.0	

Contract_period

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1	2207	55.2	55.2	55.2
	6	833	20.8	20.8	76.0
	12	960	24.0	24.0	100.0
	Total	4000	100.0	100.0	

STEPS: ANALYZE-→DESCRIPTIVE →FREQUENCY → Under STATISTICS Choose MEAN, MEDIAN , MODE ,etc

Overview:

The table lists frequencies and percentages for four categories: **Churn**, **Gender**, **Promo Friends** , and **Contract Period**.

Churn:

- **0 (Did not churn)**:
    - **Frequency**: 2939
    - **Percentage**: 73.5% of the total customers did not leave.
- **1 (Churned)**:
    - **Frequency**: 1061
    - **Percentage**: 26.5% of the total customers left.

Gender:

- **0 (Probably Male)**:
    - **Frequency**: 1959
    - **Percentage**: 49% of the customers.
- **1 (Probably Female)**:
    - **Frequency**: 2041
    - **Percentage**: 51% of the customers.

Promo Friends:

- **0 (No Promo from Friends)**:
    - **Frequency**: 2766
    - **Percentage**: 69.2% did not receive promotions from friends.
- **1 (Promo from Friends)**:
    - **Frequency**: 1234
    - **Percentage**: 30.9% received promotions from friends.

Contract Period:

- **1 Month**:
    - **Frequency**: 2207
    - **Percentage**: 55.2% of the customers had a 1-month contract.
- **6 Months**:
    - **Frequency**: 833
    - **Percentage**: 20.8% of the customers had a 6-month contract.
- **12 Months**:
    - **Frequency**: 960
    - **Percentage**: 24% of the customers had a 12-month contract.

Key Points:

- A majority of customers did not churn, with 73.5% staying.
- Gender distribution is almost equal, with slightly more females (51%) than males (49%).
- Most customers did not receive promotional offers from friends (69.2%).
- Over half of the customers had a 1-month contract (55.2%), while the rest had either 6-month (20.8%) or 12-month contracts (24%).

This table helps to understand the distribution of customers based on different attributes and th eir likelihood to churn. It indicates that contract duration and promotional offers might play role s in customer retention.

FINAL CONCLUSION :

1. Customer Retention Strategies:

- **Age-Specific Offers**:
    - **Insight**: Age is the most significant factor in predicting churn.
    - **Action**: Tailor membership offers and classes to different age groups. For younger customers, offer high-energy classes like HIIT or dance. For older customers,consider low-impact activities such as yoga or pilates.
- **Contract Periods**:
    - **Insight**: Longer contract periods are linked to lower churn.
    - **Action**: Introduce incentives for long-term memberships. Offer discounts or perks for customers who sign up for 6 or 12-month contracts.

2. Promotional Strategies:

- **Friends' Referrals**:
    - **Insight**: Promotions through friends significantly impact retention.
    - **Action**: Implement a referral program where current members get benefit s for bringing in friends. This could be free classes, merchandise, or disco unts on membership.

3. Operational Improvements:

- **Location Convenience**:
    - **Insight**: Proximity to the gym affects churn rates.
    - **Action**: If possible, consider opening smaller satellite locations in areas w ith a high concentration of members. Alternatively, provide easy transpor tation options or parking benefits to make it convenient for members to vi sit.

4. Service Enhancements:

- **Customer Segmentation**:

- **Insight**: Different factors like Age, Partner presence, and Gender affect ch urn differently.
- **Action**: Segment your members based on these factors and personalize t heir experiences. Use targeted communication, like emails or app notifica tions, to engage them with relevant content and offers.

5. Feedback and Adaptation:

- **Regular Feedback**:
  - **Insight**: Understanding why customers churn or stay can refine your strat egies.
  - **Action**: Regularly gather feedback from your members through surveys an d use this data to continuously adapt your offerings and services.

Conclusion:

To sum up, focus on tailored offers for different age groups, promote longer membershi ps with incentives, leverage referrals, enhance convenience, personalize experiences based on member segmentation, and keep adapting based on feedback. These steps, in formed by your data analysis, can significantly boost gym's retention and growth.