**Babu Banarasi Das University**

A red and blue logo

AI-generated content may be incorrect.

**Case Study**

**On**

**CUSTOMER SEGMENTATION AND CHRUN PREDICTION**

**SUBMITTED TO: SUBMITTED BY:**

**Mr. Vikas Kumar Rishi Singh**

**Riya Pandey**

**Roshni K. Verma**

**Riya Saxena**

**Agenda/Definition:**

The objective of this project is to perform customer segmentation using K-Means clustering and predict customer churn using the CHAID decision tree model in IBM SPSS Modeler 18.6.

**Outcomes/Learning:**

How to perform customer segmentation, derive meaningful insights, and apply CHAID for churn prediction.

**Required Tool:**

IBM SPSS Modeler 18.6

**Working:**

Customer data is segmented using K-Means, followed by derivation of customer segments and churn prediction using CHAID decision tree analysis.

**Step 1: Data Loading and Type Node**

The dataset 'mall\_cust\_training.xlsx' is loaded and connected to the Type node to assign measurement levels and data types.

A screenshot of a computer

AI-generated content may be incorrect.

Screenshot shows data fields being defined in the Type node.

**Step 2: K-Means Customer Segmentation**

K-Means clustering is applied to segment customers based on their Spending Score and Income levels.

A screenshot of a computer

AI-generated content may be incorrect.

Screenshot shows the K-Means node setup with spending score input.

**Step 3: K-Means Model Summary**

The Model Summary shows three clusters formed based on customer spending patterns.

A screenshot of a computer

AI-generated content may be incorrect.

Screenshot displays cluster sizes and silhouette scores.

**Step 4: Derive Customer Segment**

,I used a Derive node to create a new field named 'Customer\_Segment'. This converted the numeric cluster labels into meaningful names:  
- Cluster 1 → Loyal Regulars  
- Cluster 2 → VIP Customers  
- Cluster 3 → Budget Shoppers  
This made the clusters more understandable for business interpretation.

.

A screenshot of a computer

AI-generated content may be incorrect.

Screenshot shows the Derive node rules mapping clusters to segment names.

**Step 5: Derive Churn Field (chrun)**

I created a new derived field named 'chrun' using the Derive node. This field labels each customer as 'Churn' or 'Retained' based on spending score and customer segment.

A screenshot of a computer

AI-generated content may be incorrect.

Screenshot displays rules used for deriving churn or retained customers.

**Step 6: CHAID Model for Churn Prediction**

I used the CHAID model to predict churn using 'chrun' as the target variable and predictors such as Gender, Age, Annual Income, Spending Score, and Customer\_Segment. The CHAID tree provided clear decision rules and helped identify which segments are more likely to churn.

A screenshot of a computer

AI-generated content may be incorrect.

Screenshot shows CHAID tree structure with decision nodes.

**Step 7: Evaluation of Results**

The model results are visualized and evaluated to identify churn-prone segments for business insights.

A screenshot of a computer

AI-generated content may be incorrect.

Screenshot presents churn distribution across segments.

**Step 8: Final Model Flow**

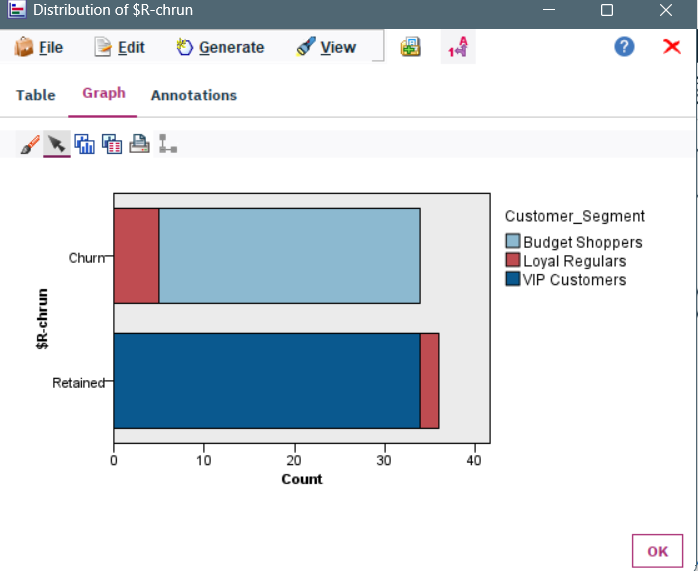
The final SPSS Modeler flow demonstrates the complete pipeline from data input to churn prediction output.A diagram of a diagram

AI-generated content may be incorrect.

Screenshot shows the complete workflow in IBM SPSS Modeler.

**Step 9: Output Visualization**

The CHAID model output nodes highlight churn and retention patterns visually.



Screenshot shows the CHAID decision tree and churn visualization results.

**Step 10: Conclusion**

VIP Customers now show 0% churn, while Budget Shoppers are the new high-risk segment with 29 churned customers.