

# **Babu Banarasi Das University**



## **Bank Customer Churn Prediction Using CHAID Decision Tree Model in IBM SPSS Modeler**

**SUBMITTED TO:**

**Mr. Vikas Kumar**

**SUBMITTED BY:**

**Akbar Naeem**

**Agenda/Definition:** The project aims to predict customer churn for a bank using the CHAID decision tree method. By analyzing customer data, the model identifies key factors influencing churn, helping the bank target retention efforts effectively

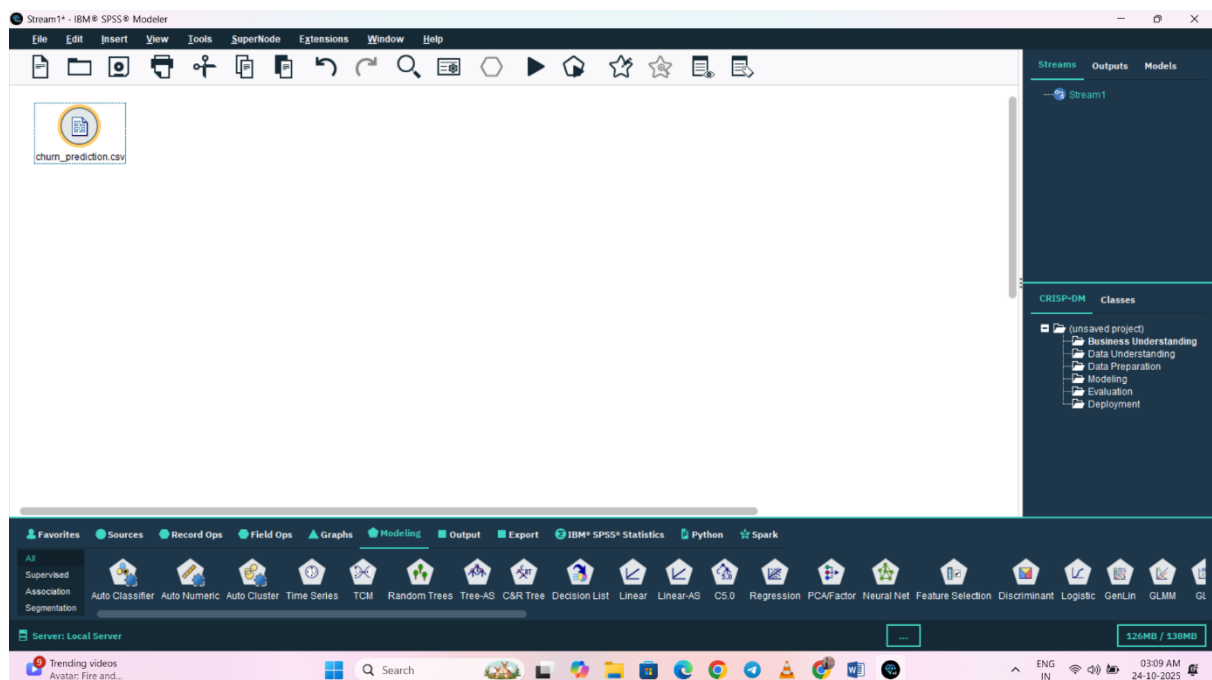
**Outcomes/Learning:** You will learn how to build a classification model to predict customer churn using CHAID in IBM SPSS Modeler. The project demonstrates the process of data preparation, model configuration, execution, and interpretation of results.

**Required Tool:** The tool used for this project is IBM SPSS Modeler.

**Working:** The project involves importing customer data, setting variable roles, configuring the CHAID model node, running the decision tree analysis, and interpreting the decision tree output. This workflow aids in understanding customer segments likely to churn.

### Step 1: Import Data

Loaded the dataset (churn\_prediction.csv) into SPSS Modeler and confirmed all fields were correctly recognized.



## Step 2: Inspect and Prepare Data

Checked for missing or invalid values and corrected any formatting or data type issues.

	customer_id	vintage	age	gender	dependents	occupation	city	customer_nw_category
1	2101	44	Male	0.000	self_employed	197...		
2	2345	35	Male	0.000	self_employed	2null6		
3	2164	31	Male	0.000	salaried	146...		
4	2329	90		0.000	self_employed	102...		
5	1579	42	Male	2.000	self_employed	149...		
6	1923	42	Female	0.000	self_employed	109...		
7	2048	72	Male	0.000	retired	102...		
8	2009	46	Male	0.000	self_employed	623...		
9	2053	31	Male	0.000	salaried	109...		
10	2295	40	Male	3.000	self_employed	102...		
11	2359	68	Male	0.000	retired	405...		
12	2293	32	Male	0.000	salaried	109...		
13	2211	73	Male	0.000	retired	44.000		
14	1660	50	Male	0.000	salaried	409...		
15	1917	46	Female	0.000	self_employed	665...		
16	1516	51	Female	0.000	self_employed	123...		
17	2293	49	Male	0.000	self_employed	112...		
18	2131	52	Female	0.000	self_employed	109...		
19	2102	47	Female	1.000	self_employed	146...		
20	2181	41	Female	0.000	self_employed	102...		

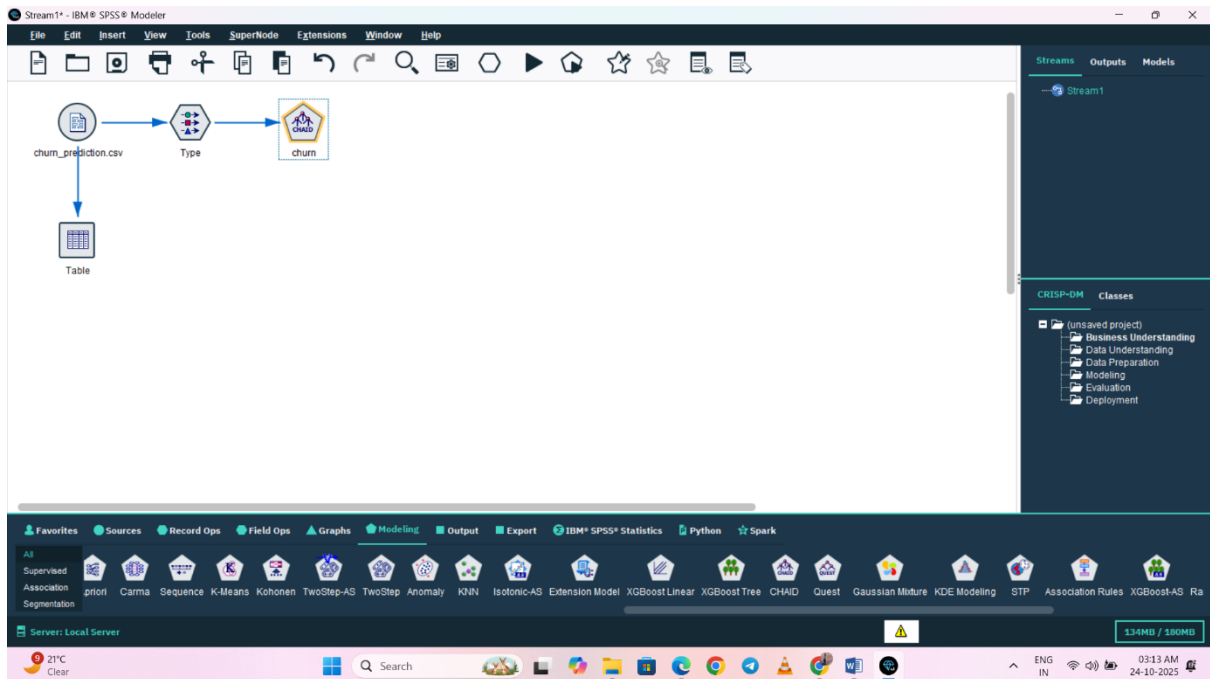
## Step 3: Assign Variable Types/Roles

Used the **Type** node to assign roles and measurement levels. The *churn* field was defined as the **target variable**

The screenshot shows the IBM SPSS Modeler interface. In the main workspace, a 'churn\_prediction.csv' data source is connected to a 'Table' node. A 'Type' node is also connected to the 'Table' node. The 'Type' node is highlighted with a red box. The 'Type' node is also connected to the 'Table' node in the main workspace. The 'Type' node is highlighted with a red box. The 'Type' node is also connected to the 'Table' node in the main workspace. The 'Type' node is highlighted with a red box.

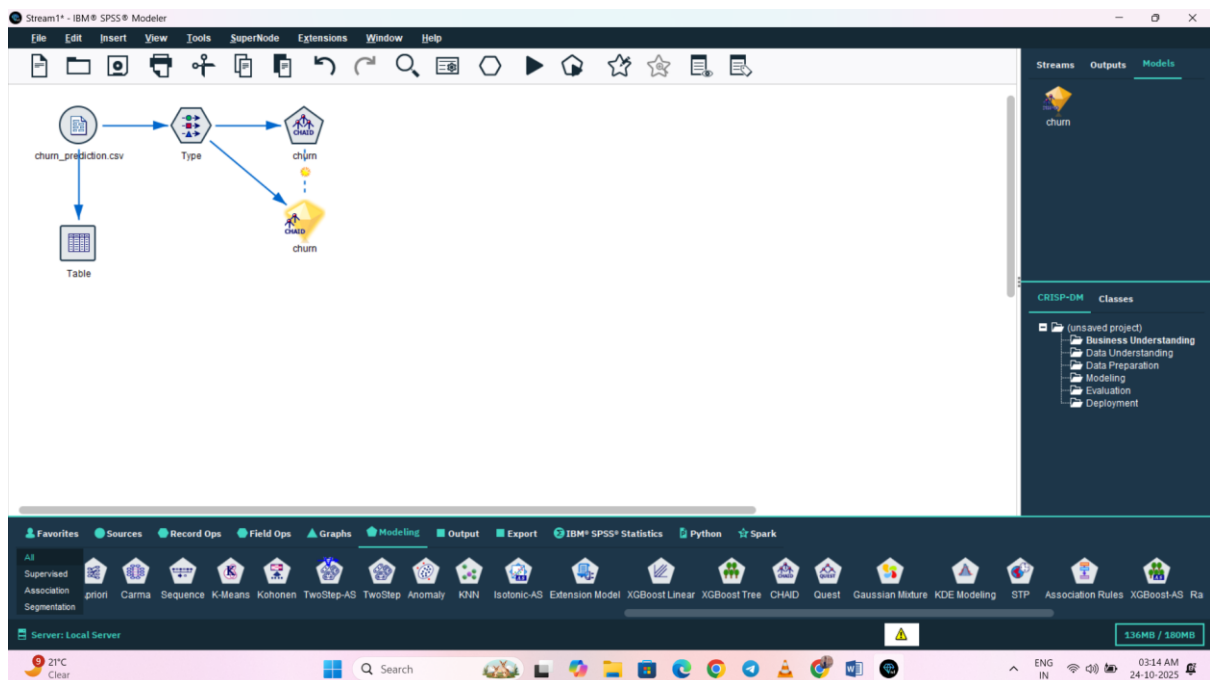
## Step 4: Add CHAID Node

Inserted the **CHAID** modeling node and connected it to the Type node output.



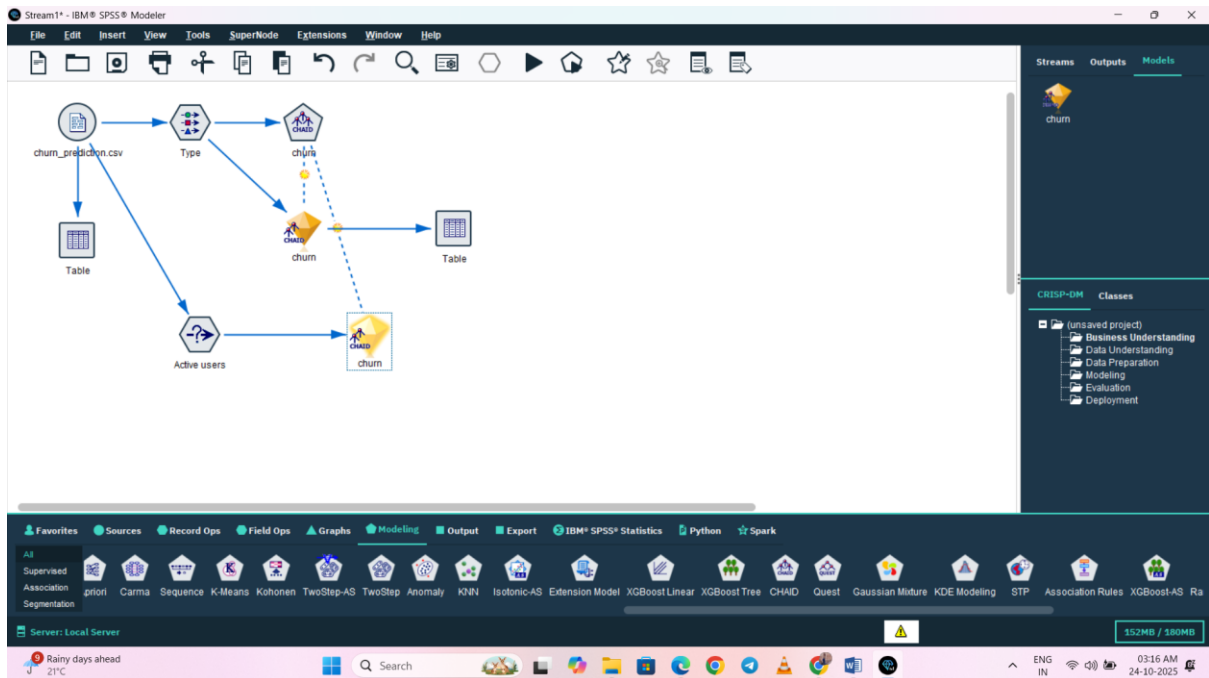
## Step 5: Configure CHAID Model

Defined model settings such as significance level, target variable, and maximum depth before saving the configuration.



### Step 6: Train the Model (Run CHAID)

Executed the model stream and generated the CHAID decision tree output.

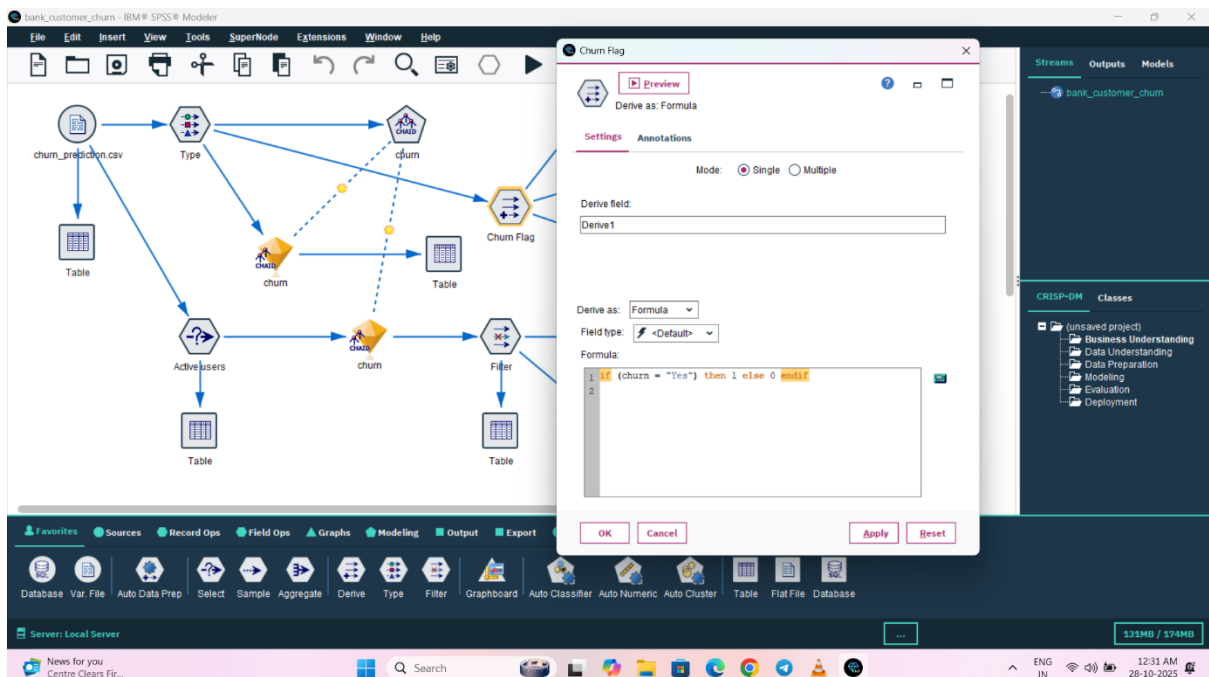


## Step 7: Derive Node Creation

Added a **Derive** node to create a new field `Churn_Flag` using the formula:

```
if (churn = "Yes") then 1 else 0 endif
```

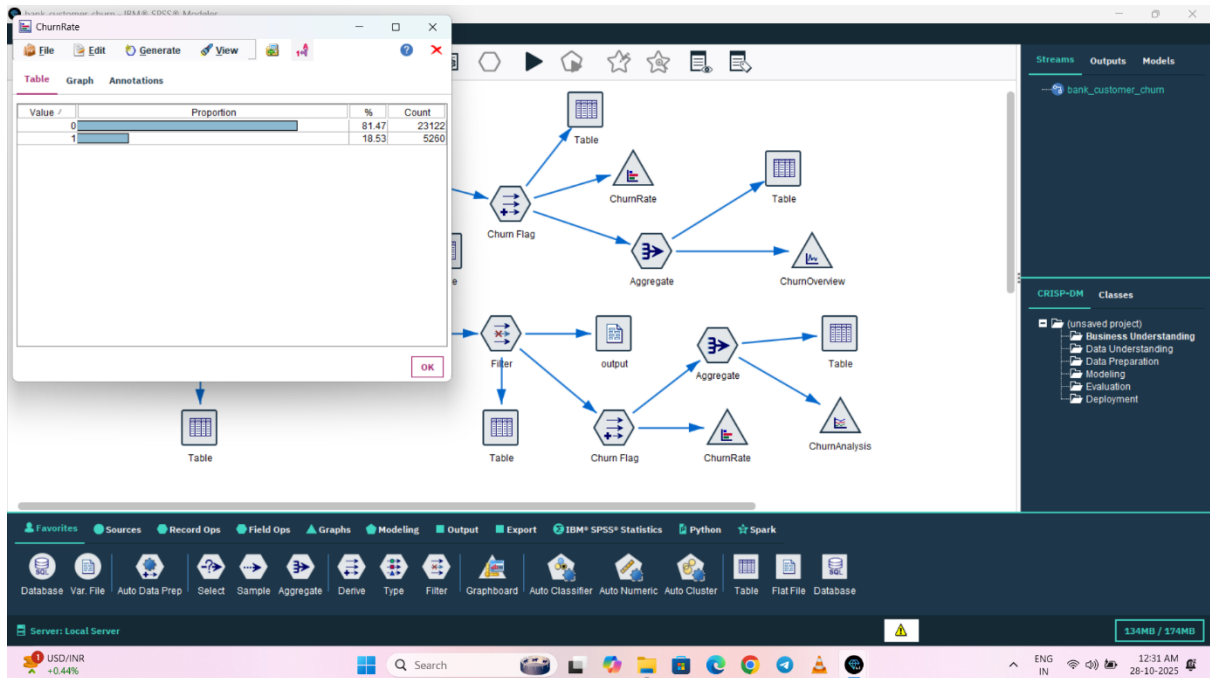
This numeric flag enabled better aggregation and analysis in subsequent nodes.



## Step 8: Calculate Churn Rate

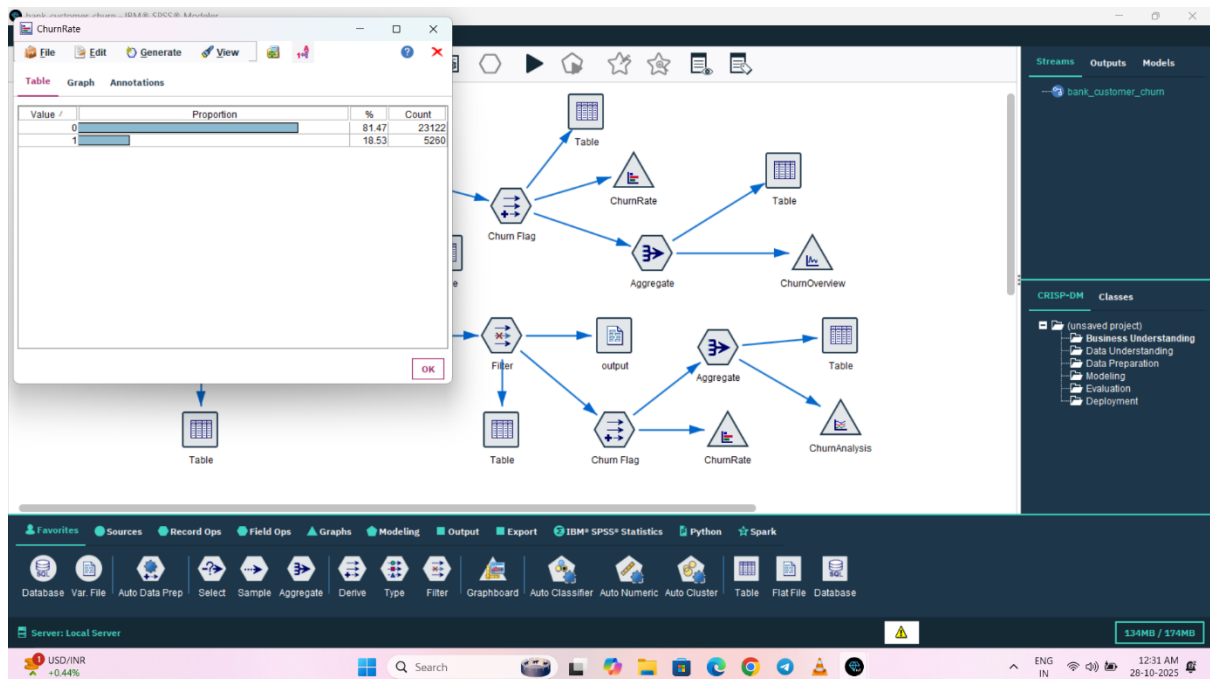
Used **Aggregate** and **Table** nodes to compute churn proportions —

- 0 → 81.47% (Non-churned)
- 1 → 18.53% (Churned)



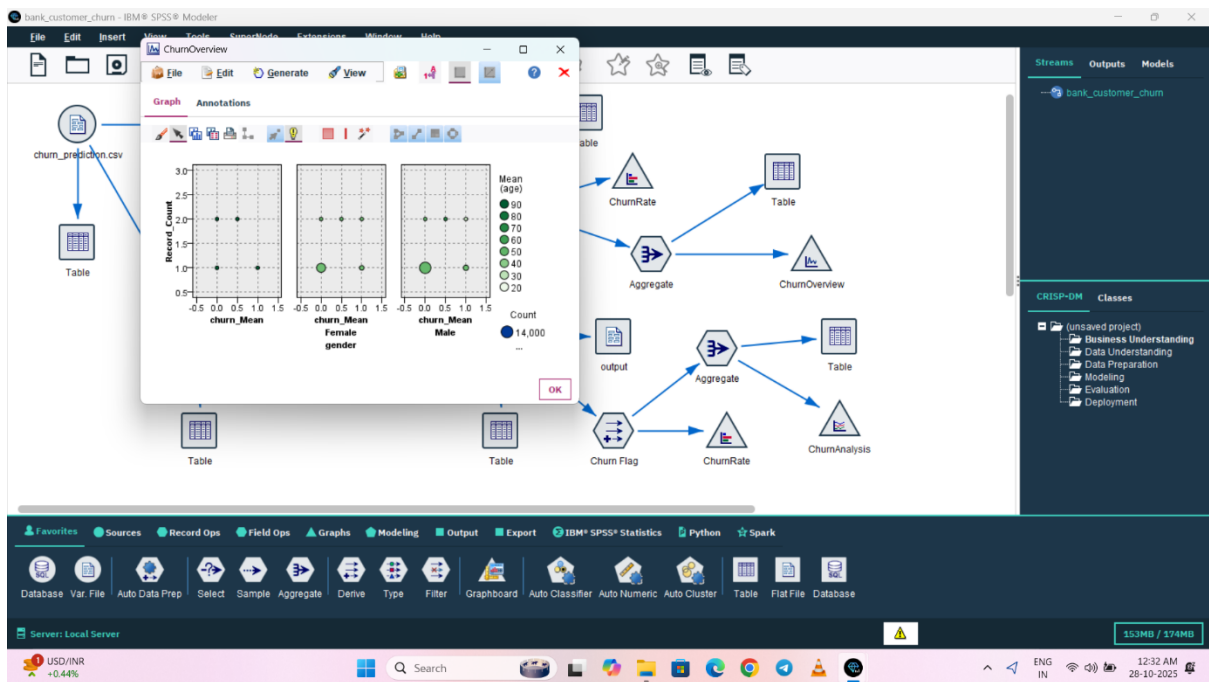
## Step 9: Analyze Churn by Demographics

Applied **Aggregate** with **vintage**, **age**, and **gender** as key fields. Calculated mean churn and record count for each group to observe segment-wise trends.



## Step 10: Visualize Churn Overview

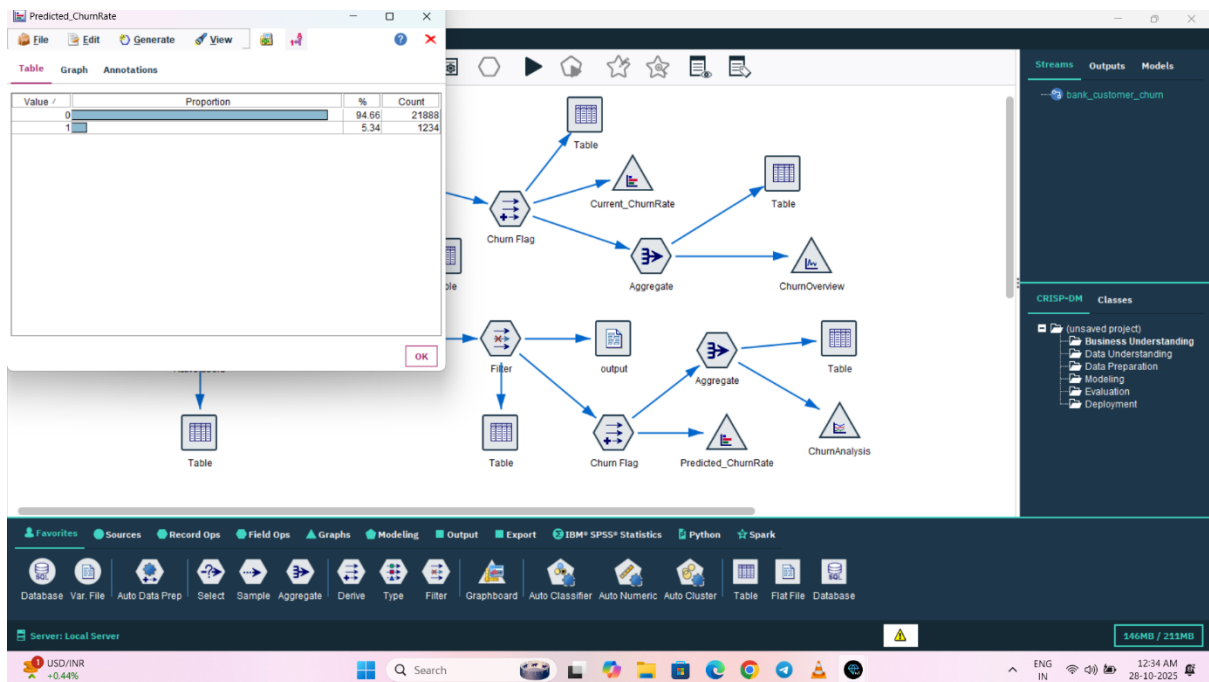
Displayed a **Bubble Plot** showing churn means by gender and average age, revealing visible patterns in churn tendency.



## Step 11: Predictive Churn Rate Analysis

Derived **Predicted\_ChurnRate** from model output —

- Predicted Non-Churn (0): 94.66%
- Predicted Churn (1): 5.34%





Generated a **Line Graph** displaying churn mean across record counts, reflecting gradual churn variation across customer segments.

[illegible]



# Conclusion

The project “**Bank Customer Churn Prediction Using CHAID Decision Tree Model in IBM SPSS Modeler**” effectively demonstrates how predictive analytics can help identify customers likely to leave a bank. The objective was to analyze customer data, discover key churn factors, and use the CHAID decision tree algorithm to predict churn behavior accurately.

The workflow began with importing and inspecting the dataset, ensuring proper formatting and resolving missing values. Variable roles were assigned through the **Type node**, with the *churn* field defined as the target. A **Derive node** was created to generate a numeric *Churn\_Flag*, enabling more precise aggregation and visualization of churn rates. Using **Aggregate** and **Table** nodes, the overall churn rate was calculated — about **18.53% churned** and **81.47% retained** — offering a clear understanding of the data distribution.

The **CHAID model** was then configured with parameters such as maximum tree depth and significance level, producing a decision tree that divided customers into segments based on churn likelihood. The model identified important predictors such as customer age, vintage, and account balance. These insights revealed that newer and younger customers were more prone to churn, while long-tenured clients showed stronger retention.

Further analysis and visualizations — including bubble and line plots — provided a deeper view of churn patterns across demographics and segments. The **final model evaluation** compared actual and predicted churn values, confirming the CHAID model’s reliability. The **stream diagram** summarizes this entire workflow, from data preparation to model interpretation, visually representing how each analytical stage connects within IBM SPSS Modeler.

In summary, this project successfully applied the CHAID decision tree to uncover actionable insights for customer retention. It highlights how data-driven approaches can help banks anticipate churn, improve engagement, and make informed strategic decisions. The knowledge gained from this workflow strengthens analytical proficiency in SPSS Modeler and lays a foundation for future enhancements using advanced machine learning models or automated churn monitoring systems.

---