BABU BANARASI DAS UNIVERSITY

Session- 2025-26

Submitted To :

Mr. Vikas Kumar

Submitted By :

Amitesh Singh

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

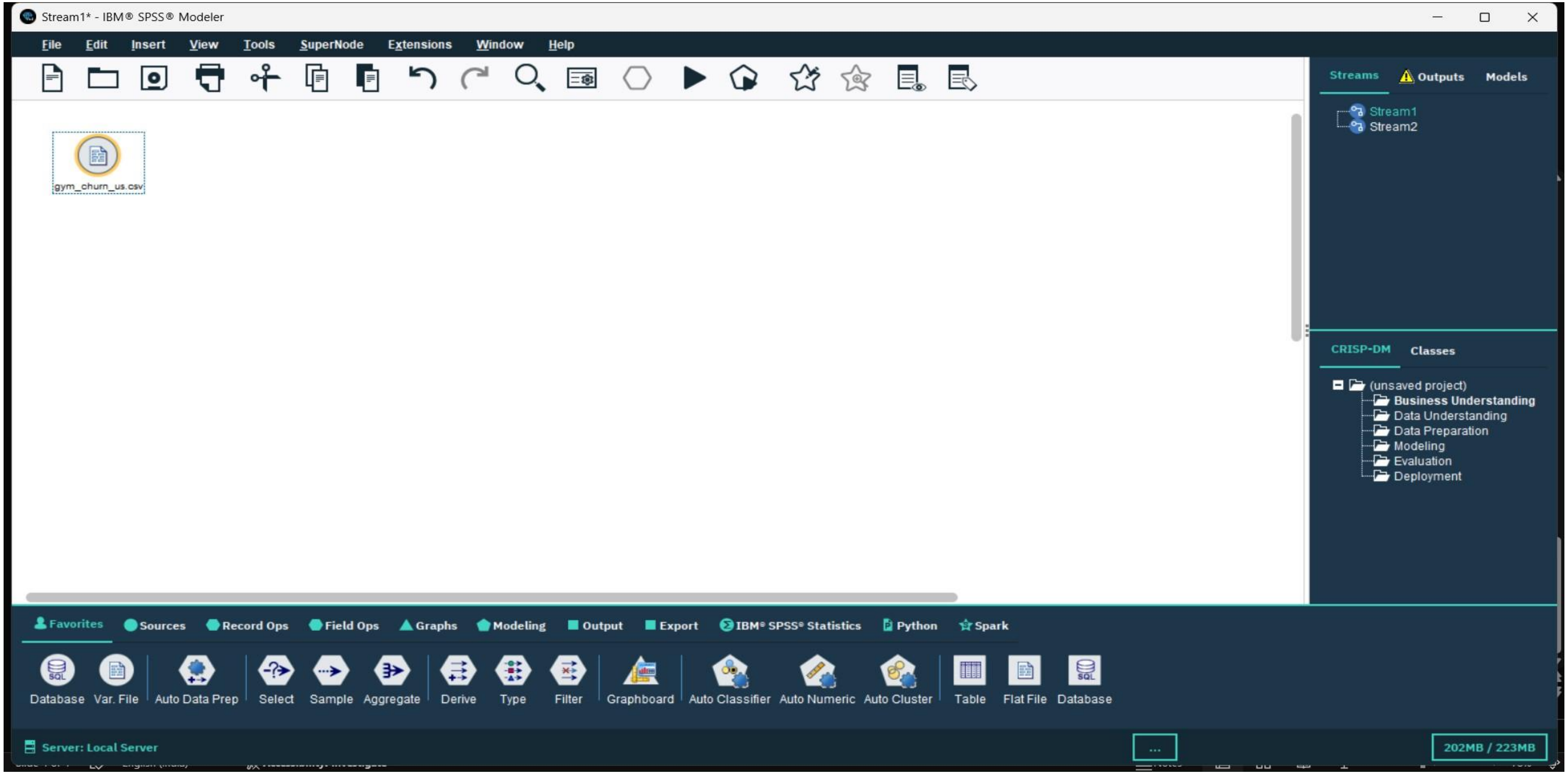
Outcomes/Learning: You will learn how to build a classification model to predict customer churn using CHAID in IBM SPSS Modeler. The project demonstrates the process of data preparation, model configuration, execution, and interpretation of results.

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.

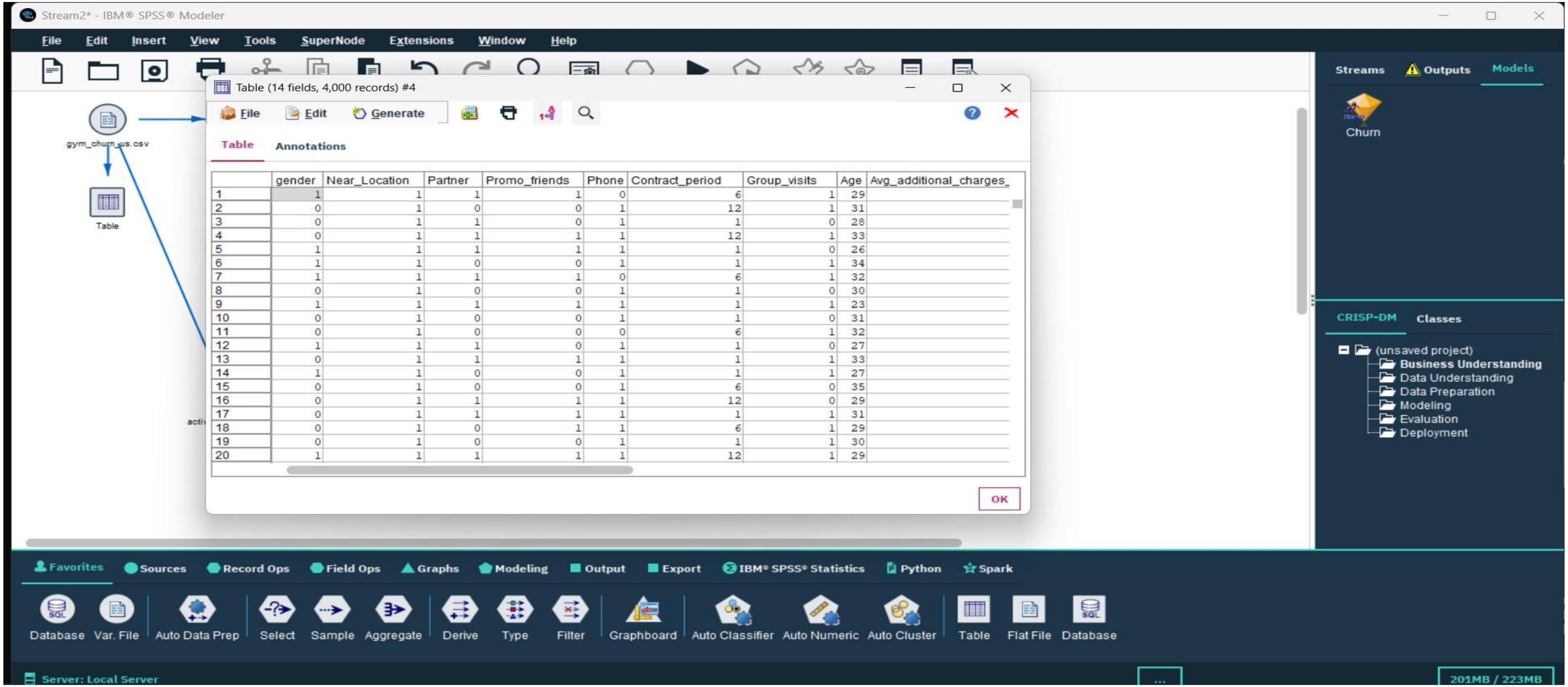
Step 1: Import Data

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



Step 2: Inspect and Prepare Data:

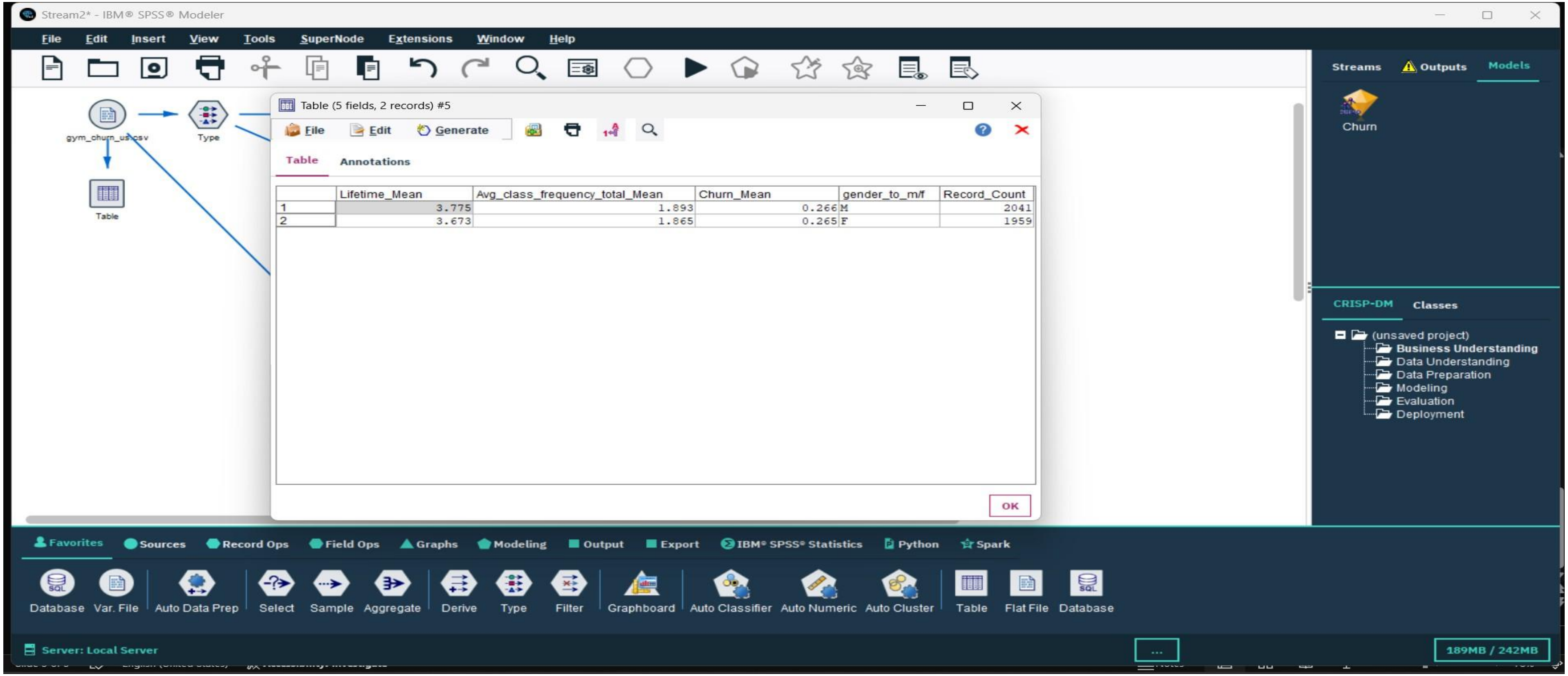
Checked for missing or invalid values and corrected any formatting or data type issues



	gender	Near_Location	Partner	Promo_friends	Phone	Contract_period	Group_visits	Age	Avg_additional_charges_
1	1	1	1	1	0	6	1	29	
2	0	1	0	0	1	12	1	31	
3	0	1	1	0	1	1	0	28	
4	0	1	1	1	1	12	1	33	
5	1	1	1	1	1	1	0	26	
6	1	1	0	0	1	1	1	34	
7	1	1	1	1	0	6	1	32	
8	0	1	0	0	1	1	0	30	
9	1	1	1	1	1	1	1	23	
10	0	1	0	0	1	1	0	31	
11	0	1	0	0	0	6	1	32	
12	1	1	1	0	1	1	0	27	
13	0	1	1	1	1	1	1	33	
14	1	1	0	0	1	1	1	27	
15	0	1	0	0	1	6	0	35	
16	0	1	1	1	1	12	0	29	
17	0	1	1	1	1	1	1	31	
18	0	1	0	1	1	6	1	29	
19	0	1	0	0	1	1	1	30	
20	1	1	1	1	1	12	1	29	

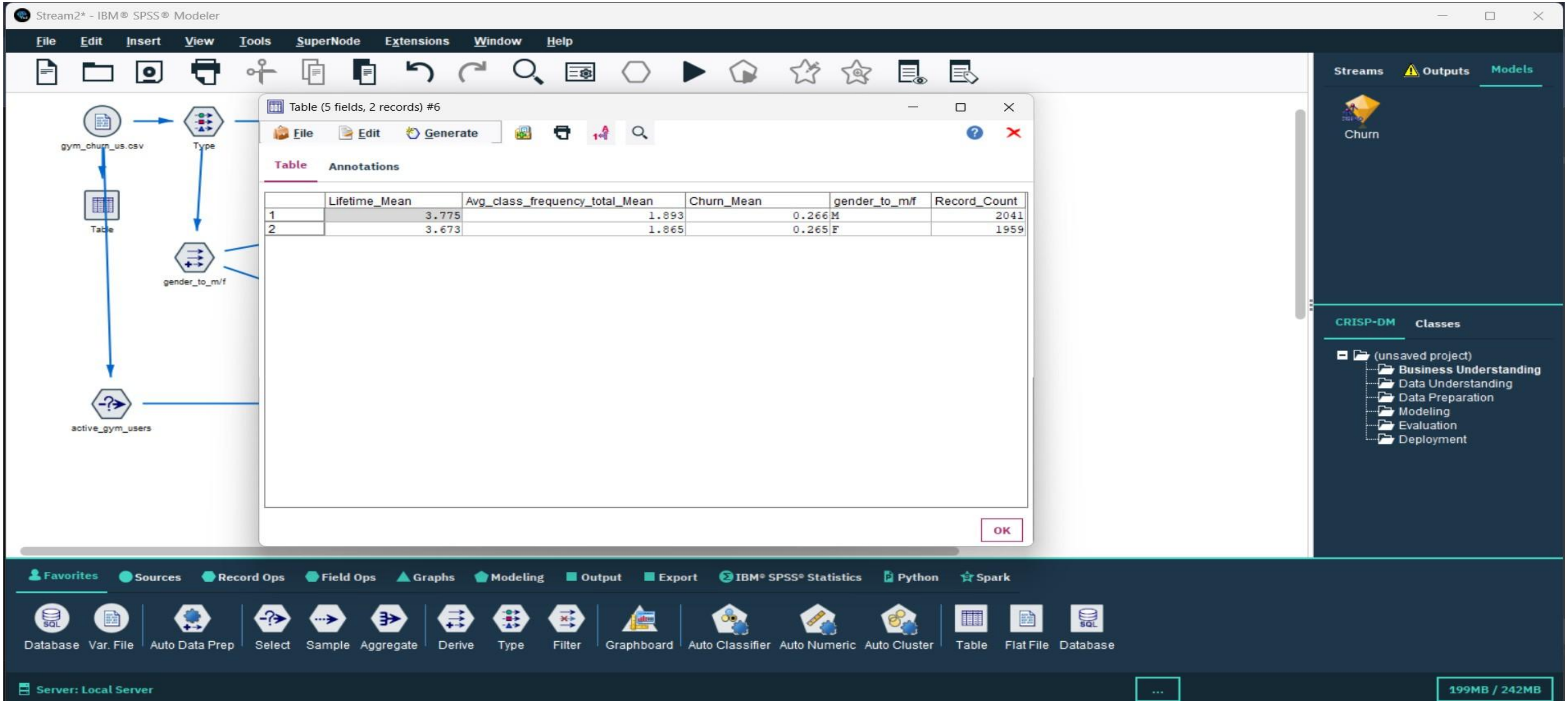
Step 3: Assign Variable Types/Roles :

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



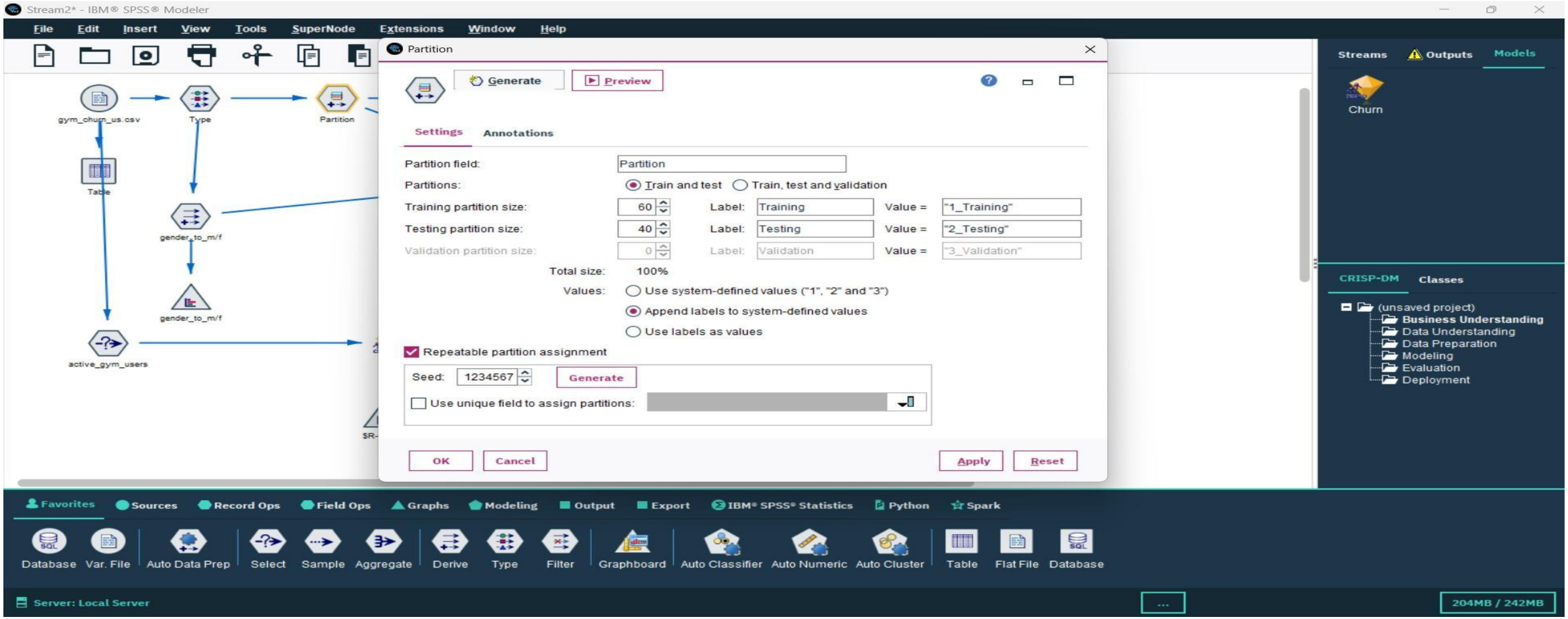
Step 4: Derive Node:

Derive Node converted the numeric gender codes (0 and 1) into categorical labels “F” and “M” for better readability.



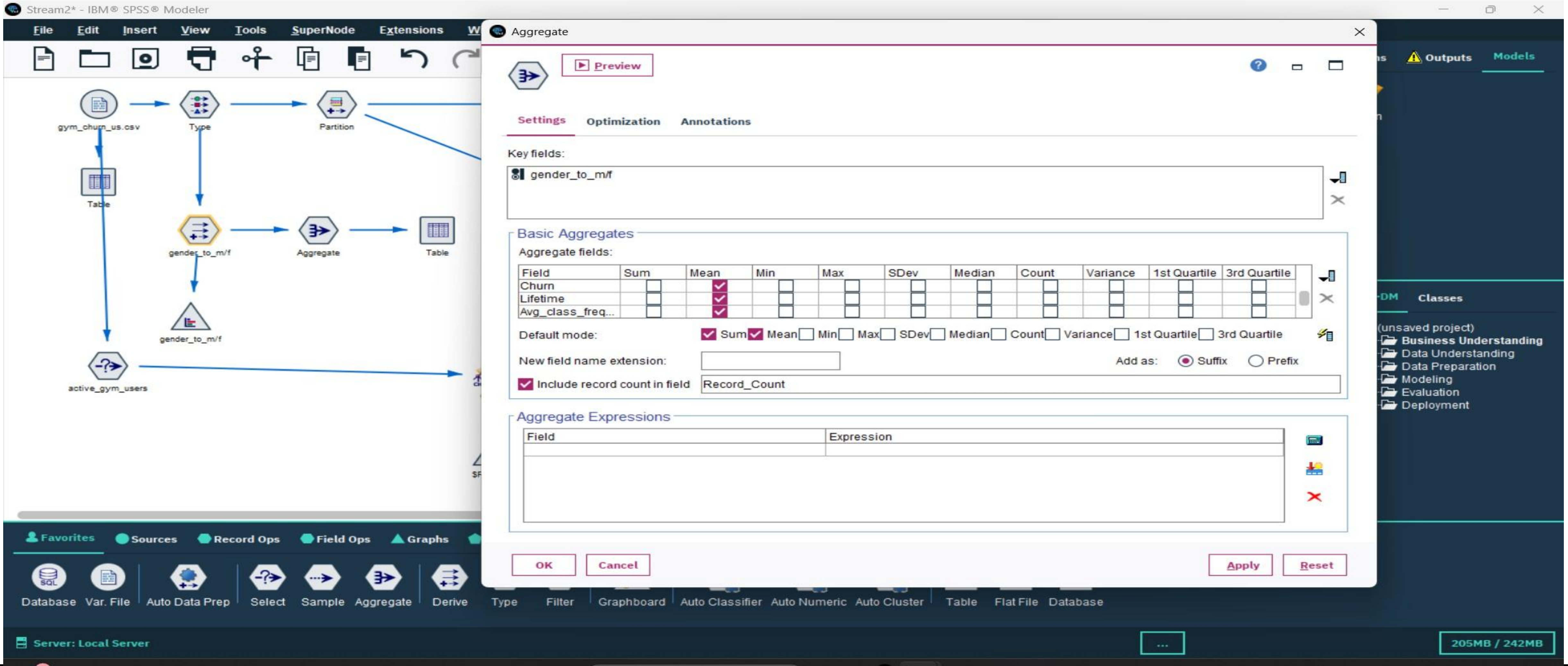
Step 5: Partitoin node:

A **Partition Node** in IBM SPSS Modeler is used to split the dataset into separate subsets, such as **training** and **testing** samples.
It helps in **model validation** by allowing you to test the model’s accuracy on unseen data.



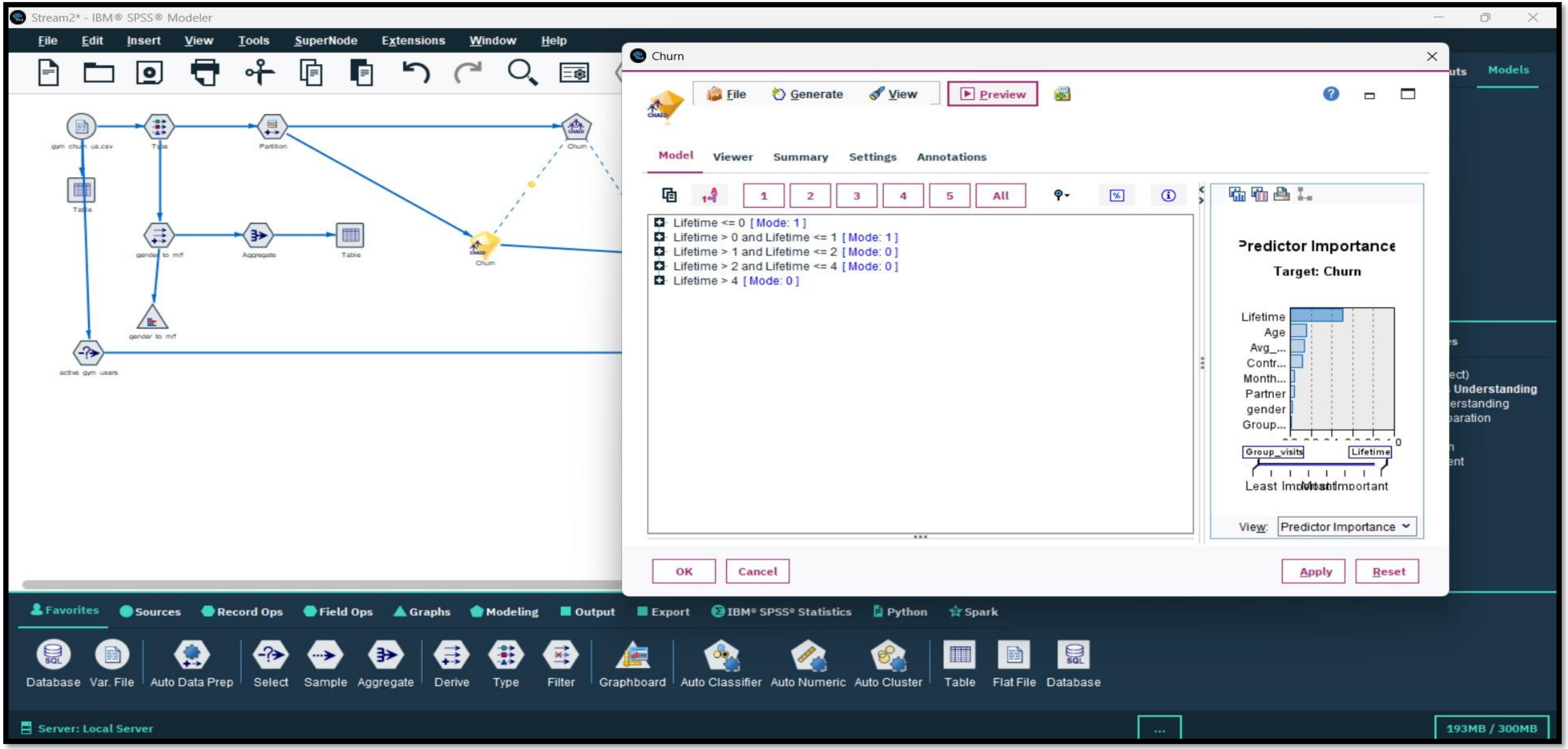
Step 6: Aggregate Node:

The **Aggregate Node** in IBM SPSS Modeler is used to **summarize data by grouping records** based on key fields. It helps compute statistics like **mean, sum, count, or maximum** for each group to identify overall trends and patterns.



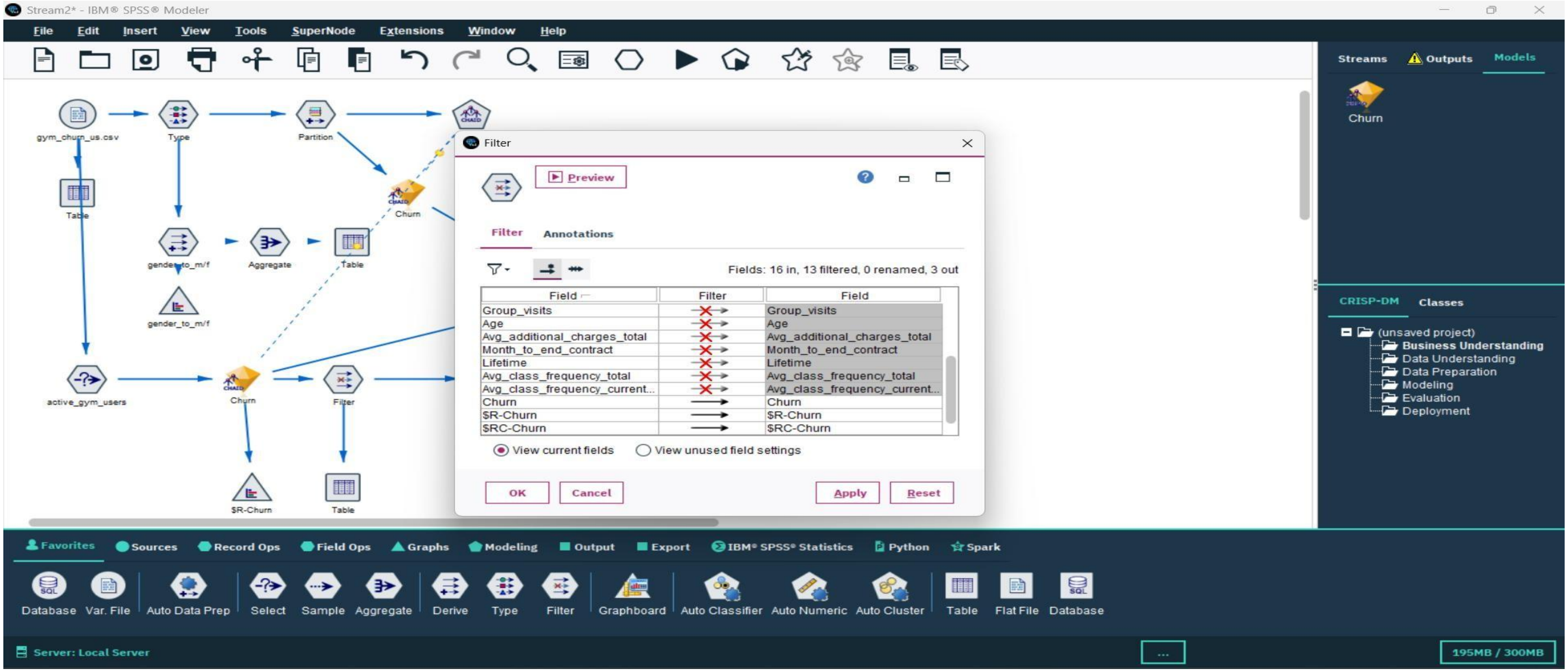
Step 7: Train the Model (Run CHAID)

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



Step 8: Filter Node :

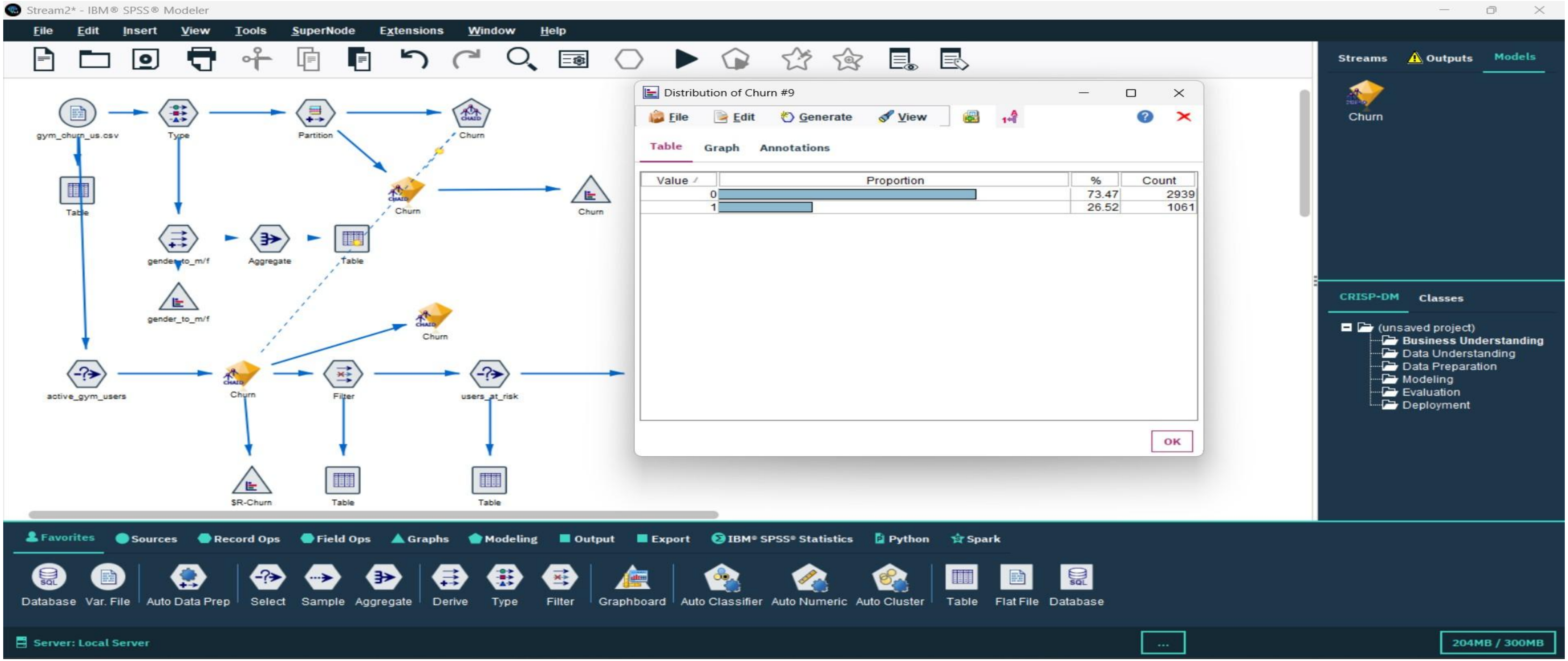
A **Filter Node** in IBM SPSS Modeler is used to **include or exclude specific fields** from the dataset.
It helps in **removing irrelevant or unwanted variables** before analysis or modeling.



Step 8: Calculate Churn Rate:

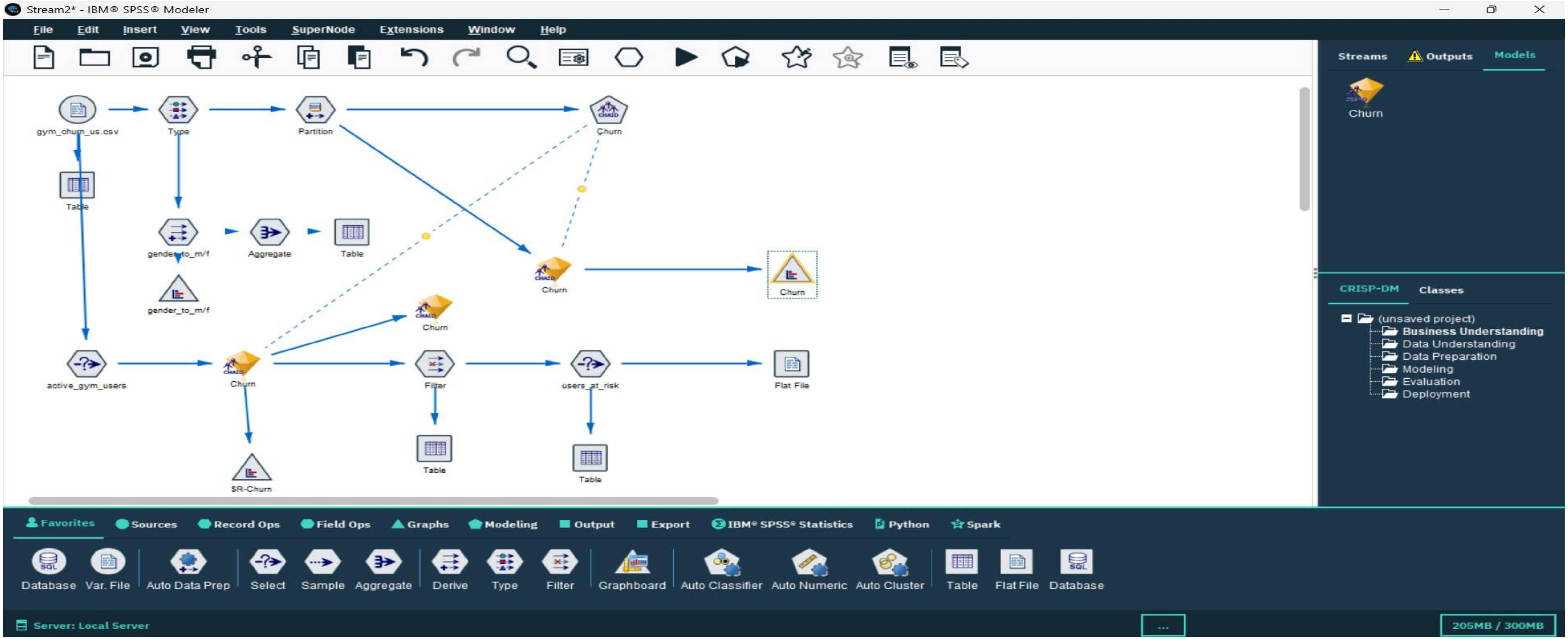
Used Aggregate and Table nodes to compute churn proportions.

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



Step 13: Model Evaluation & Summary

Compared actual vs. predicted churn rates to evaluate model performance and interpret findings for actionable retention planning. The complete SPSS Modeler stream (shown below) illustrates the workflow from data import to churn prediction and analysis:



🧾 Conclusion

The churn analysis conducted using **IBM SPSS Modeler** provided valuable insights into customer behavior and retention at the gym. Through systematic data preparation and transformation, key variables such as **gender**, **lifetime**, and **average class frequency** were analyzed to understand their relationship with churn. The **Derive Node** was effectively used to convert numeric gender codes into readable labels (“M” and “F”), improving the interpretability of the results.

Further, by using the **Aggregate Node**, important statistical summaries like mean lifetime, average class frequency, and churn rate were computed for each gender group. The analysis revealed that both male and female customers have similar churn rates, but slight variations in engagement and lifetime values. These findings highlight the importance of personalized engagement strategies to reduce member dropout and improve retention.

Overall, the project demonstrates how **IBM SPSS Modeler** can be leveraged to perform data preparation, transformation, and statistical analysis in a structured way. It also emphasizes the role of data-driven decision-making in understanding customer patterns and supporting effective business strategies.

Summary

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