BABU BANARASI DAS UNIVERSITY

BBD
UNIVERSITY

Session- 2025-26

Submitted To :
Mr. Vikas Kumar

Submitted By :
Aman Yadav

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

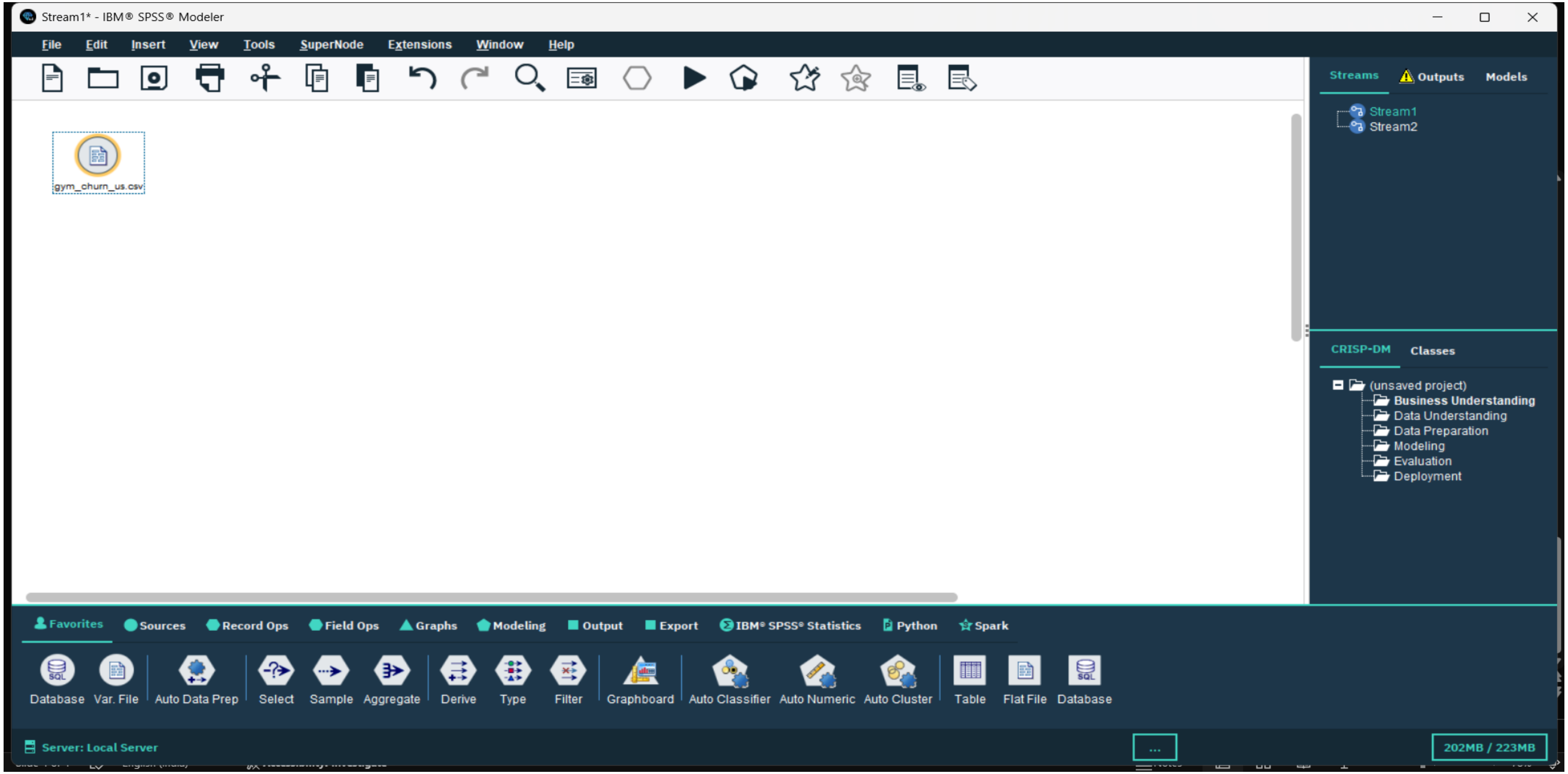
Outcomes/Learning: You will learn how to build a classification model to predict customer churn using CHAID in IBM SPSS Modeler. The project demonstrates the process of data preparation, model configuration, execution, and interpretation of results.

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

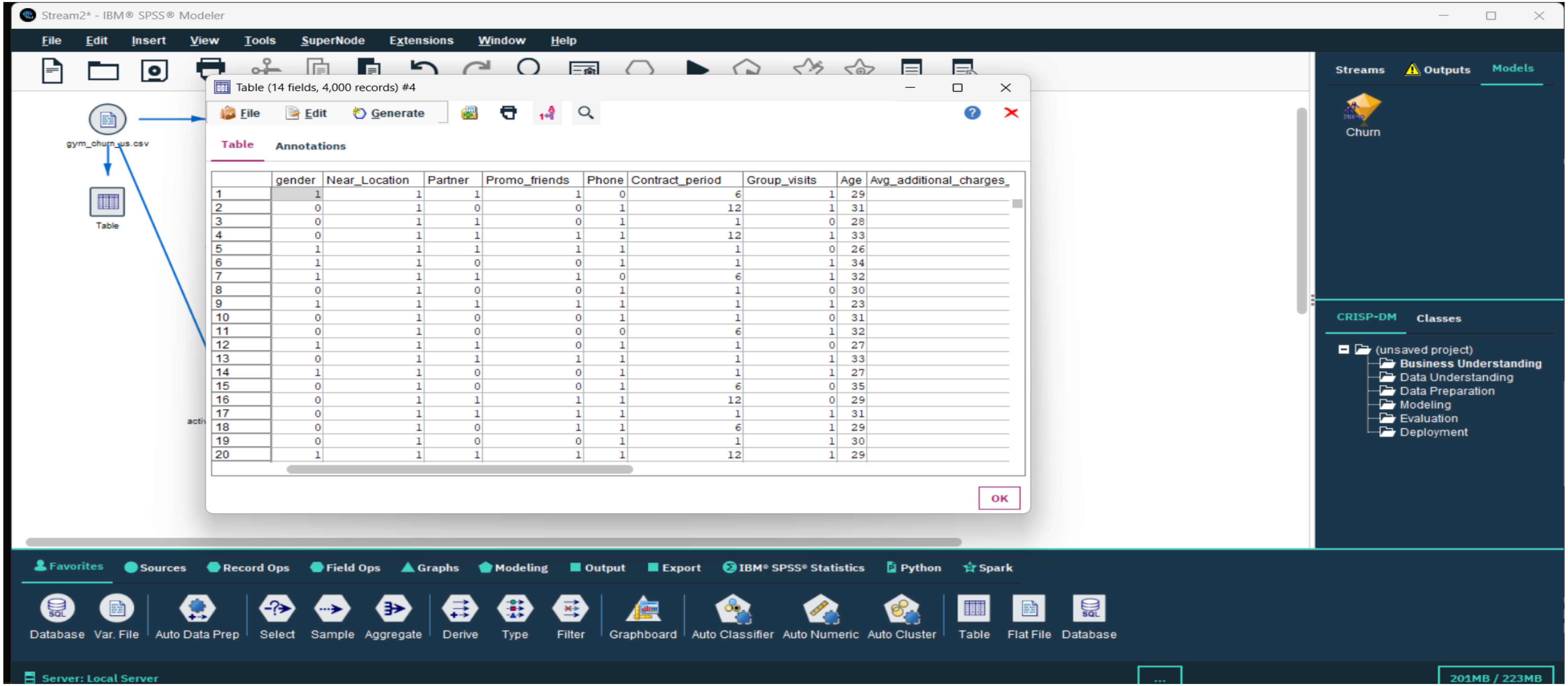
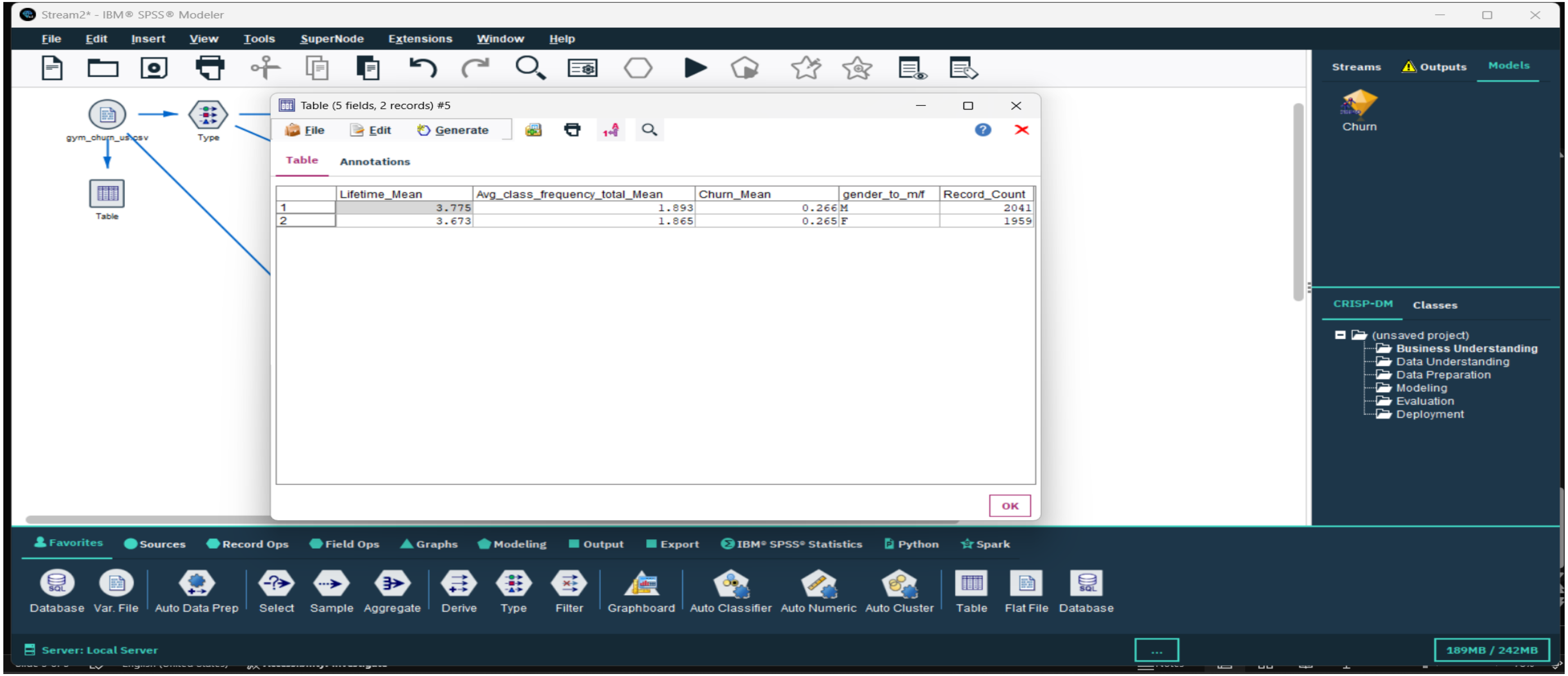


Table (14 fields, 4,000 records) #4

	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 :

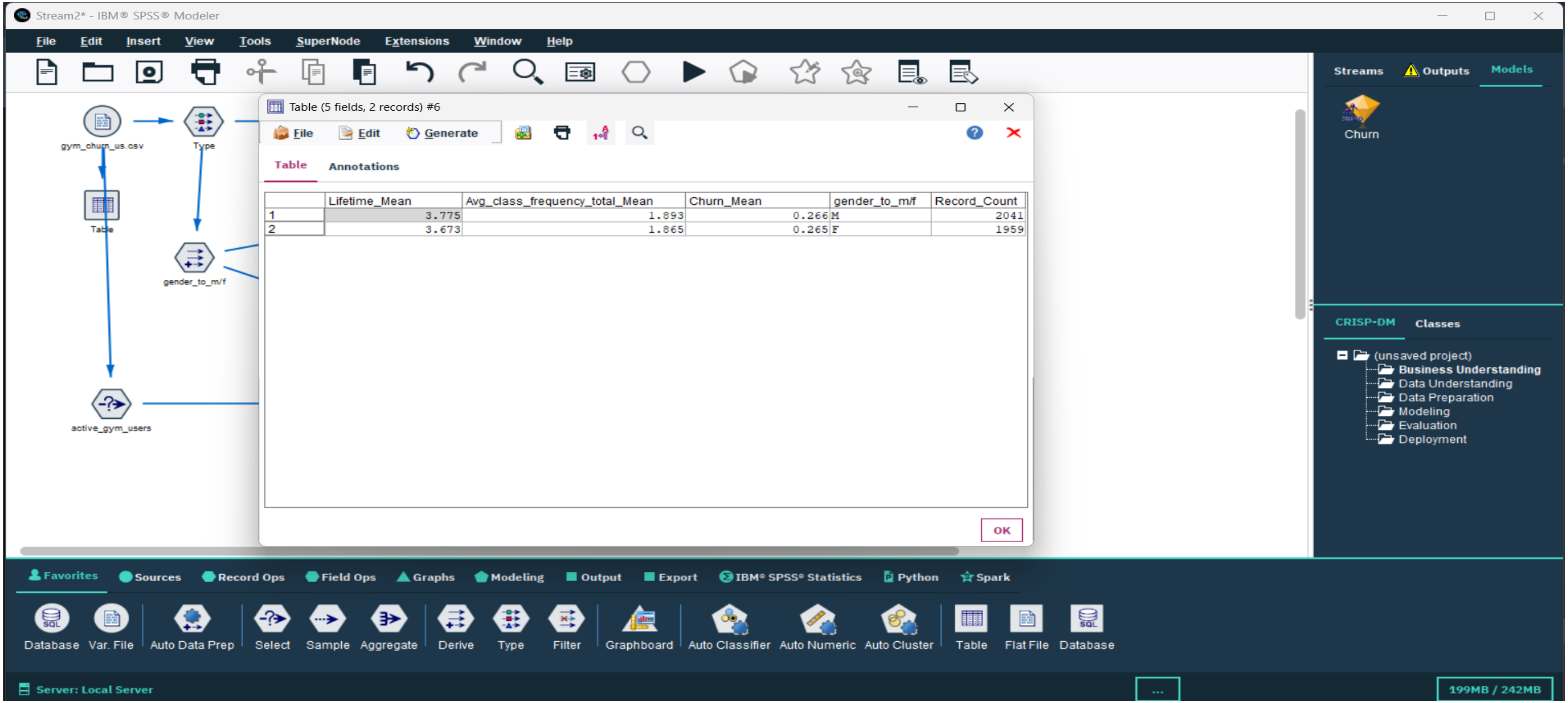
Used the **Type node** to assign roles and measurement levels. The churn field was defined as the **target variable.**



	Lifetime_Mean	Avg_class_frequency_total_Mean	Churn_Mean	gender_to_m/f	Record_Count
1	3.775	1.893	0.266	M	2041
2	3.673	1.865	0.265	F	1959

Step 4: Derive Node:

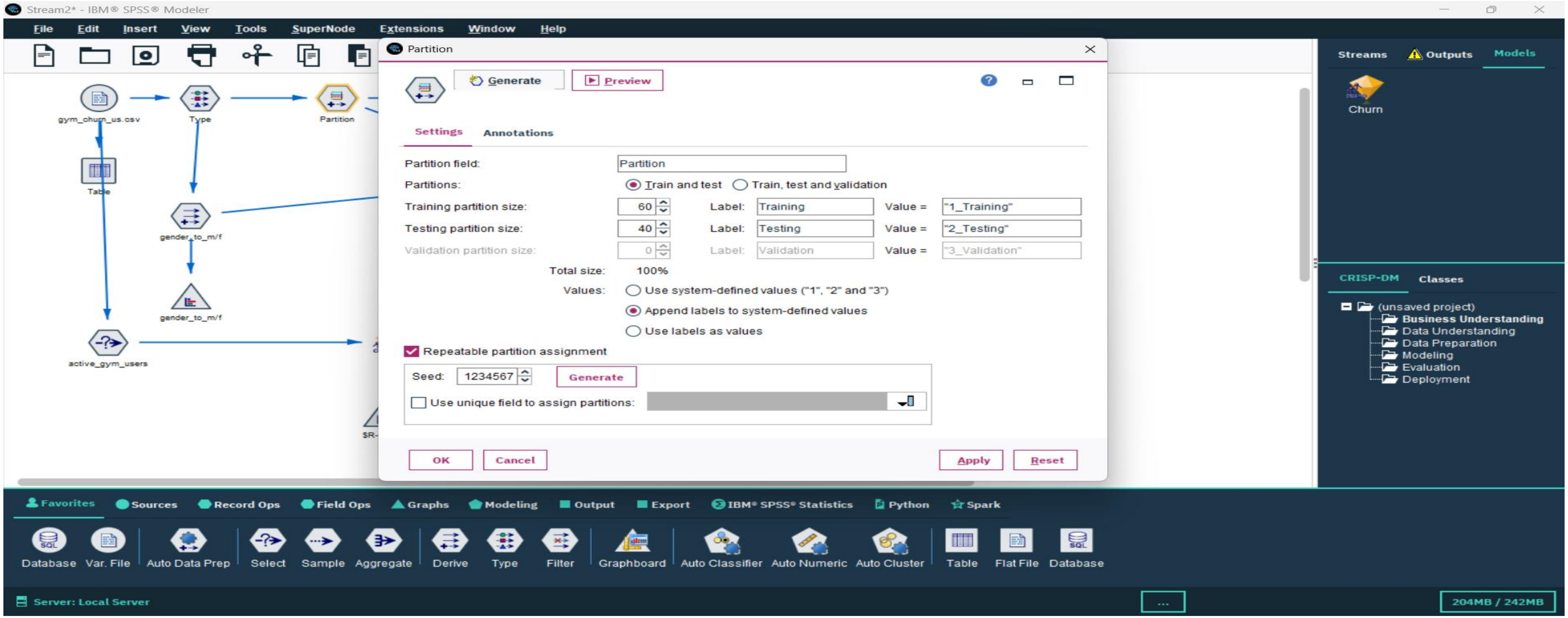
Derive Node converted the numeric gender codes (0 and 1) into categorical labels “F” and “M” for better readability.



	Lifetime_Mean	Avg_class_frequency_total_Mean	Churn_Mean	gender_to_m/f	Record_Count
1	3.775	1.893	0.266	M	2041
2	3.673	1.865	0.265	F	1959

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.



Partition

Partition field: Partition

Partitions: ● Train and test ○ Train, test and validation

Training partition size: 60 Label: Training Value = "1_Training"

Testing partition size: 40 Label: Testing Value = "2_Testing"

Validation partition size: 0 Label: Validation Value = "3_Validation"

Total size: 100%

Values: ○ Use system-defined values ("1", "2" and "3")

● Append labels to system-defined values

○ Use labels as values

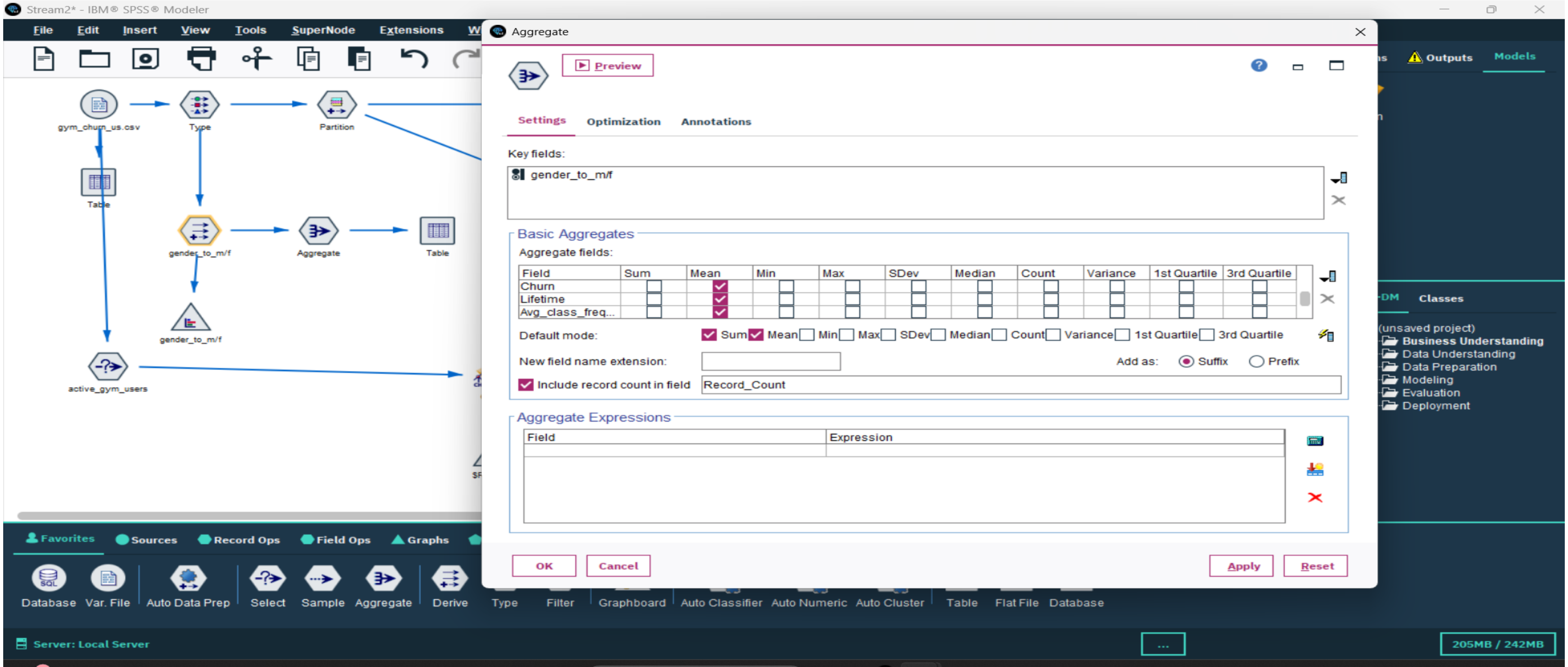
☑ Repeatable partition assignment

Seed: 1234567

□ Use unique field to assign partitions:

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.



Key fields:
gender_to_m/f

Basic Aggregates

Aggregate fields:

Field	Sum	Mean	Min	Max	SDev	Median	Count	Variance	1st Quartile	3rd Quartile
Churn	□	☑	□	□	□	□	□	□	□	□
Lifetime	□	☑	□	□	□	□	□	□	□	□
Avg_class_freq...	□	☑	□	□	□	□	□	□	□	□

Default mode: ☑ Sum ☑ Mean □ Min □ Max □ SDev □ Median □ Count □ Variance □ 1st Quartile □ 3rd Quartile

New field name extension:

Add as: ● Suffix ○ Prefix

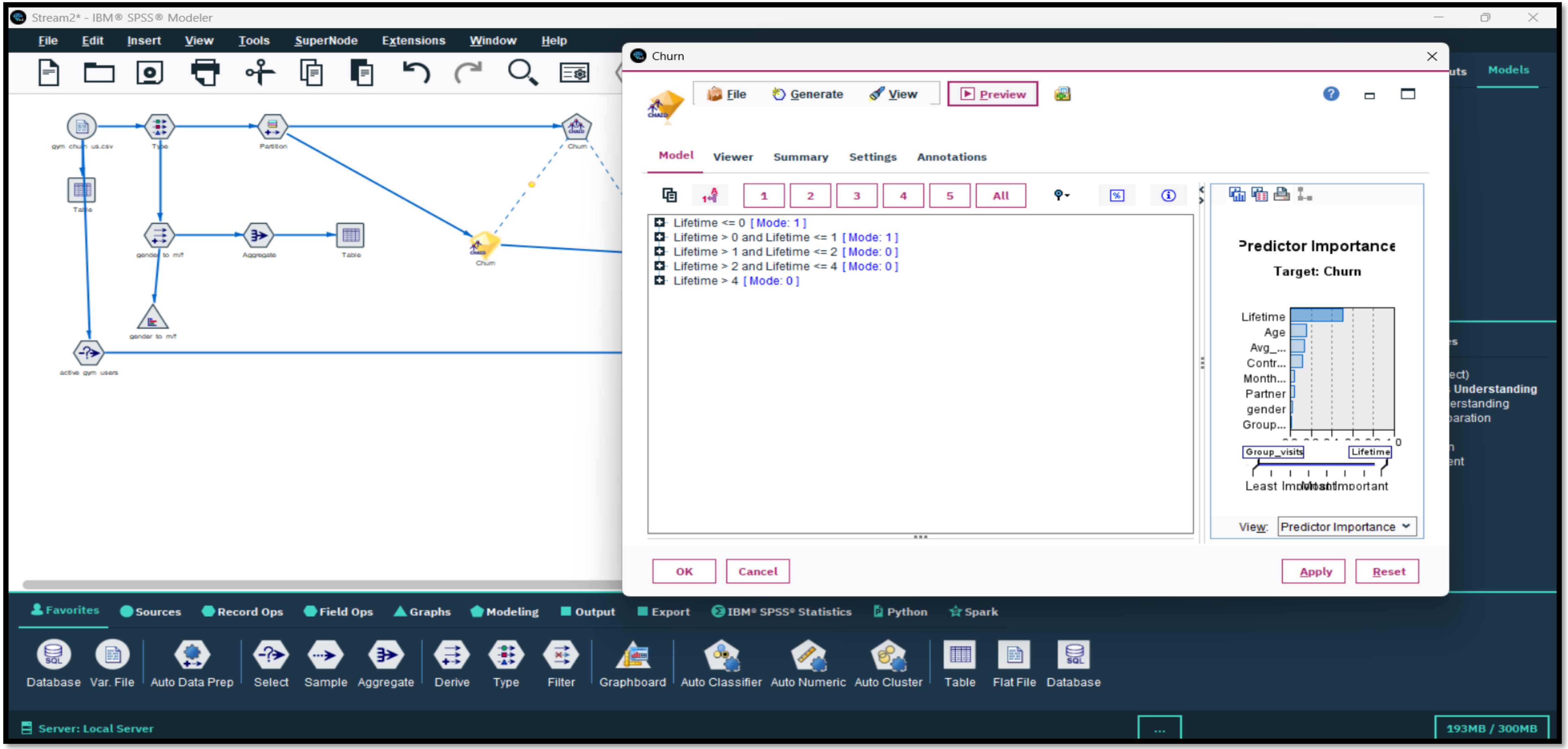
☑ Include record count in field Record_Count

Aggregate Expressions

Field	Expression

Step 7: Train the Model (Run CHAID)

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

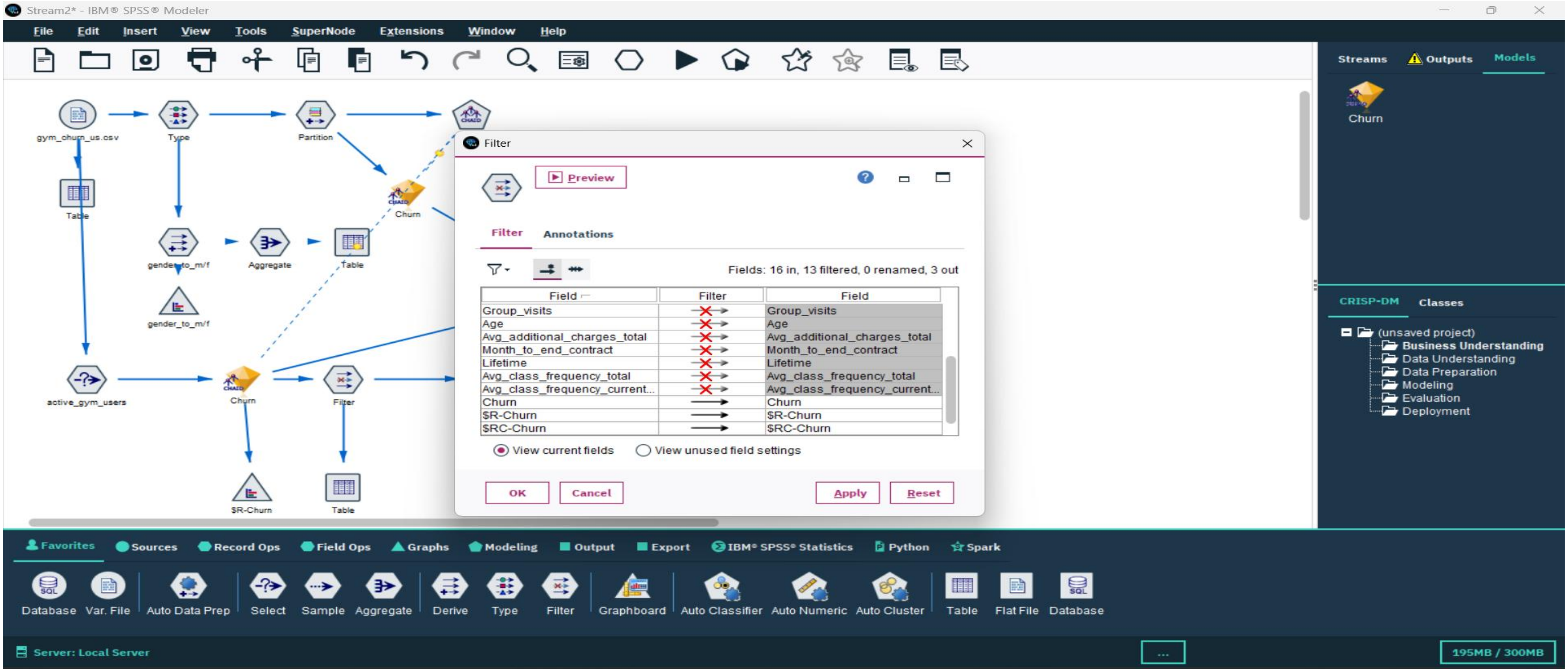


Stream2* - IBM® SPSS® Modeler
File Edit Insert View Tools SuperNode Extensions Window Help
gym churn us.csv
Type
Partition
Churn
Table
gender to m/f
Aggregate
Table
Churn
gender to m/f
active gym users
Churn
File Generate View Preview
Model Viewer Summary Settings Annotations
1 2 3 4 5 All
Predictor Importance
Target: Churn
Lifetime
Age
Avg_...
Contr...
Month...
Partner
gender
Group...
Group_visits
Lifetime
Least Important
Most Important
View: Predictor Importance
OK Cancel Apply Reset
Models
Favorites Sources Record Ops Field Ops Graphs Modeling Output Export IBM® SPSS® Statistics Python Spark
Database Var. File Auto Data Prep Select Sample Aggregate Derive Type Filter Graphboard Auto Classifier Auto Numeric Auto Cluster Table Flat File Database
Server: Local Server
193MB / 300MB

```
Lifetime <= 0 [ Mode: 1 ]
Lifetime > 0 and Lifetime <= 1 [ Mode: 1 ]
Lifetime > 1 and Lifetime <= 2 [ Mode: 0 ]
Lifetime > 2 and Lifetime <= 4 [ Mode: 0 ]
Lifetime > 4 [ Mode: 0 ]
```

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.



Fields: 16 in, 13 filtered, 0 renamed, 3 out

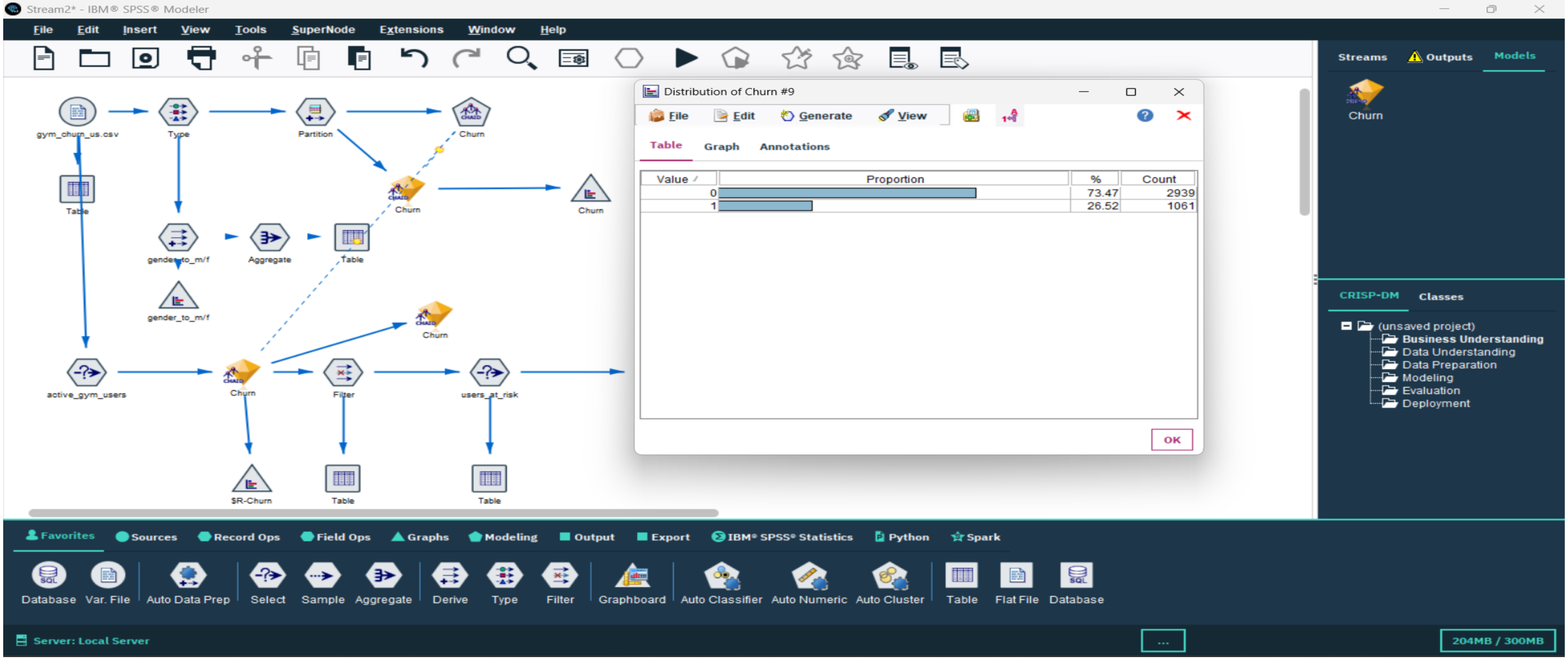
Field	Filter	Field
Group_visits	✗→	Group_visits
Age	✗→	Age
Avg_additional_charges_total	✗→	Avg_additional_charges_total
Month_to_end_contract	✗→	Month_to_end_contract
Lifetime	✗→	Lifetime
Avg_class_frequency_total	✗→	Avg_class_frequency_total
Avg_class_frequency_current...	✗→	Avg_class_frequency_current...
Churn	→	Churn
$R-Churn	→	$R-Churn
$RC-Churn	→	$RC-Churn

● View current fields ○ View unused field settings

Step 8: Calculate Churn Rate:

Used Aggregate and Table nodes to compute churn proportions.

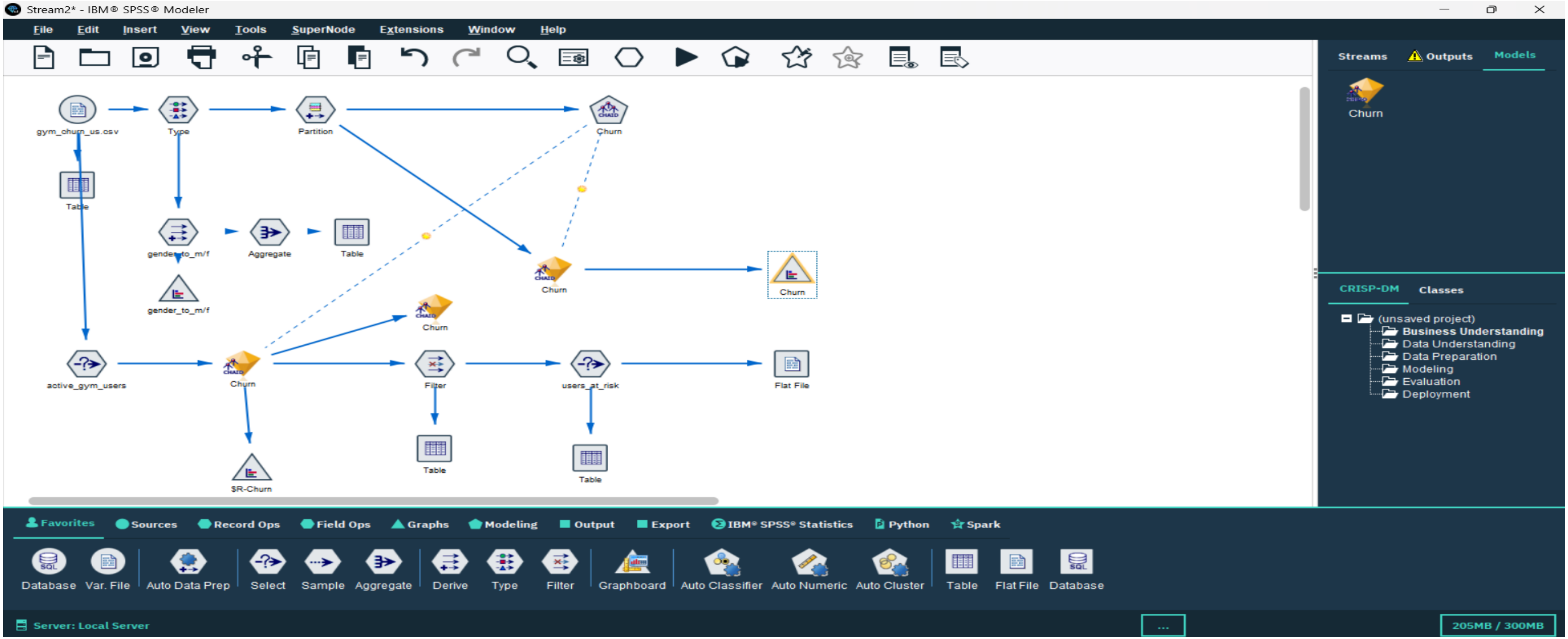
- 0 → 81.47% (Non-churned
- 1 → 18.53% (Churned)



Value	Proportion	%	Count
0		73.47	2939
1		26.52	1061

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