IBM SPSS Modeller

Name - Aradhya Dubey, Amogh Srivastava, Aditya Rai, Abhishek Giri

Class - BCADS - 31

Teacher - Mr. Vikas

Practical 1

Start a new stream.

1. Open SPSS Modeler and create a new Stream

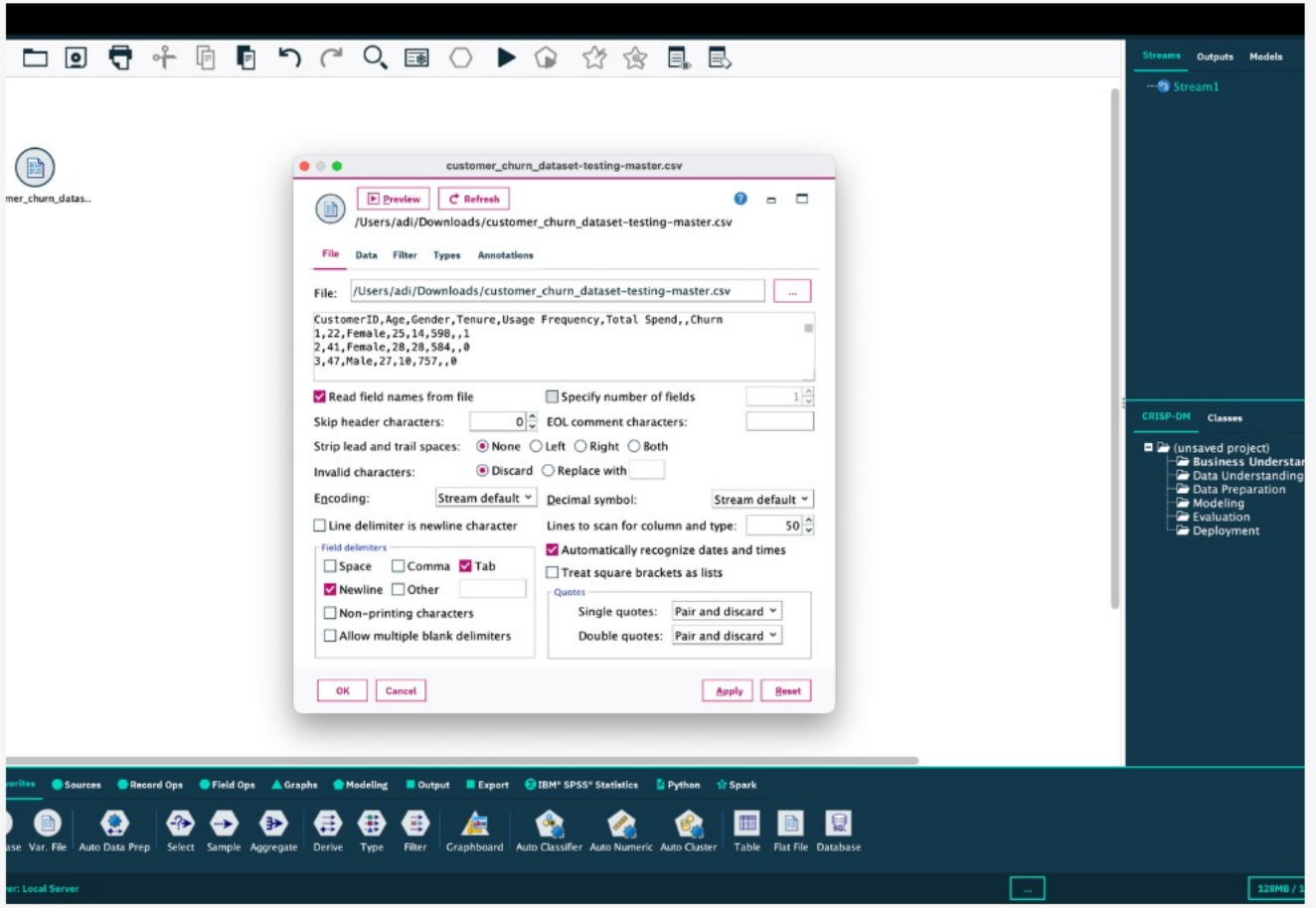
Read the CSV

1. Add a **Var.File** node (or **Source → File →** CSV depending on your version).

2. Point it to your CSV file customer_churn_dataset-testing-master.csv.

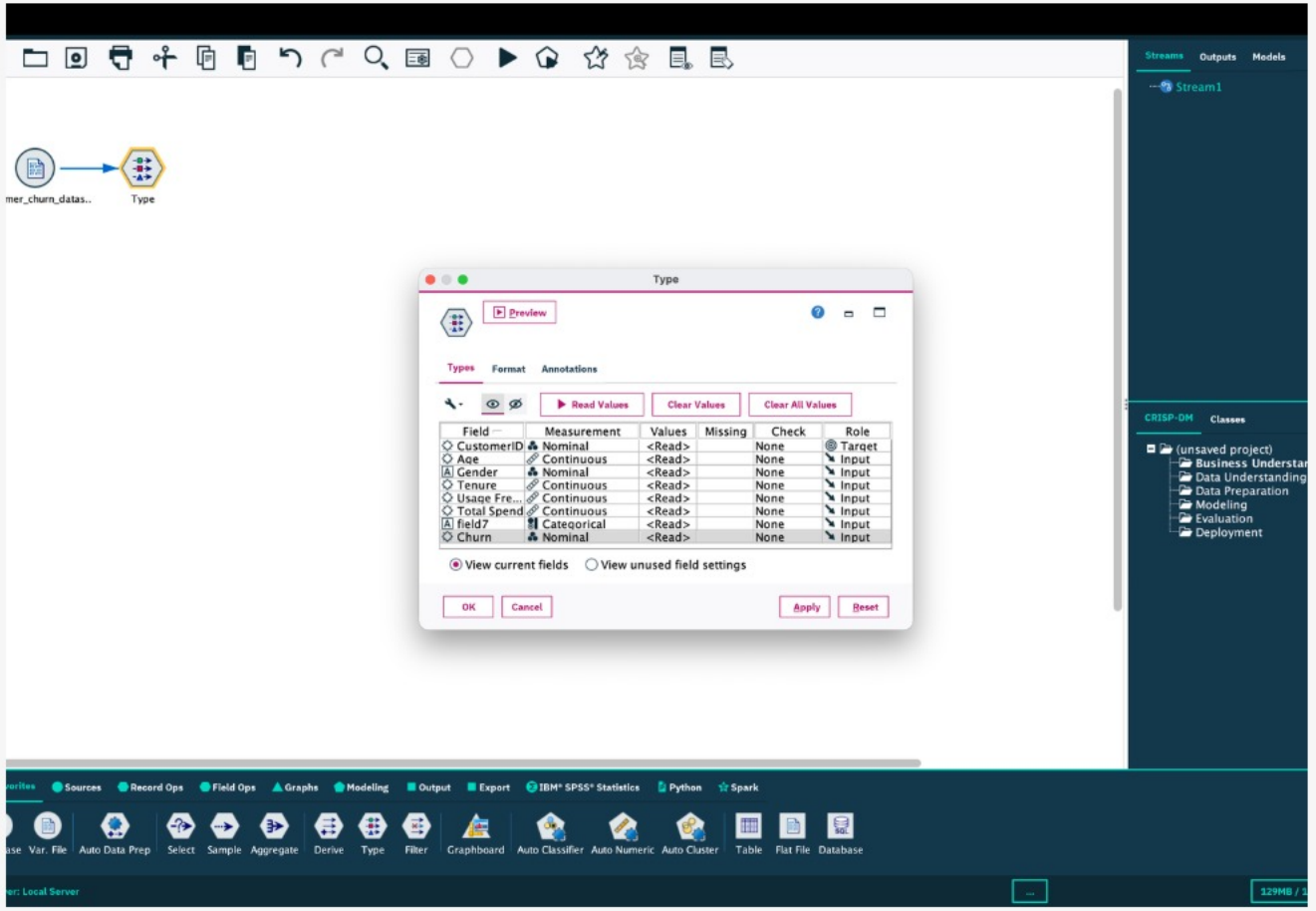
3. Check “First row contains field names”.

4. Click **Preview** to inspect rows.



Practical 1

Add a Type node (Data type & role correction.
Drag a **Type** node and connect the Var.File → Type.
In the Type node:
Set CustomerID → Role = **Id** (or **Key**),
Measurement = **Nominal** (but role = Id excludes from modeling).
Set Age → Role = **Input**, Measurement = **Ratio** or **Continuous**.
Gender → Role = **Input**, Measurement = **Nominal**.
Tenure → Role = **Input**, Measurement = **Ratio**.
Usage Frequency → Role = **Input**, Measurement = **Ratio**.
Total Spend → Role = **Input**, Measurement = **Ratio**.
Unnamed: 6 → Role = **None** (or remove column).
Churn → Role = **Target**, Measurement = **Nominal** (convert 0/1 to Nominal if it's numeric).



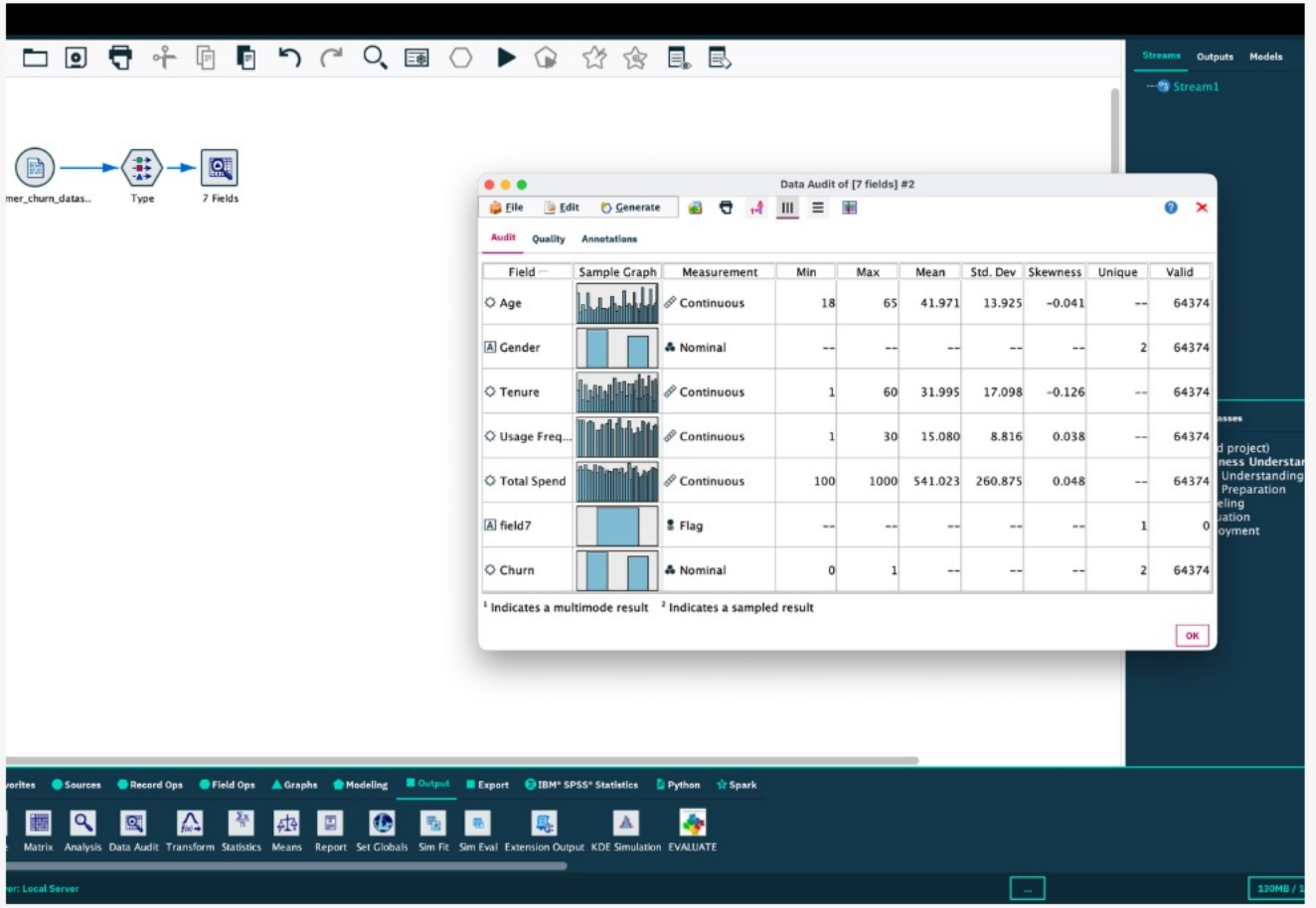
Practical 1

Add a Data Audit node

Connect **Type** → **Audit** (or use **Data Audit**).

Run it to check missing values, distributions, outliers.

Action: You'll see Unnamed: 6 empty — confirm removal (if not already set in Type).



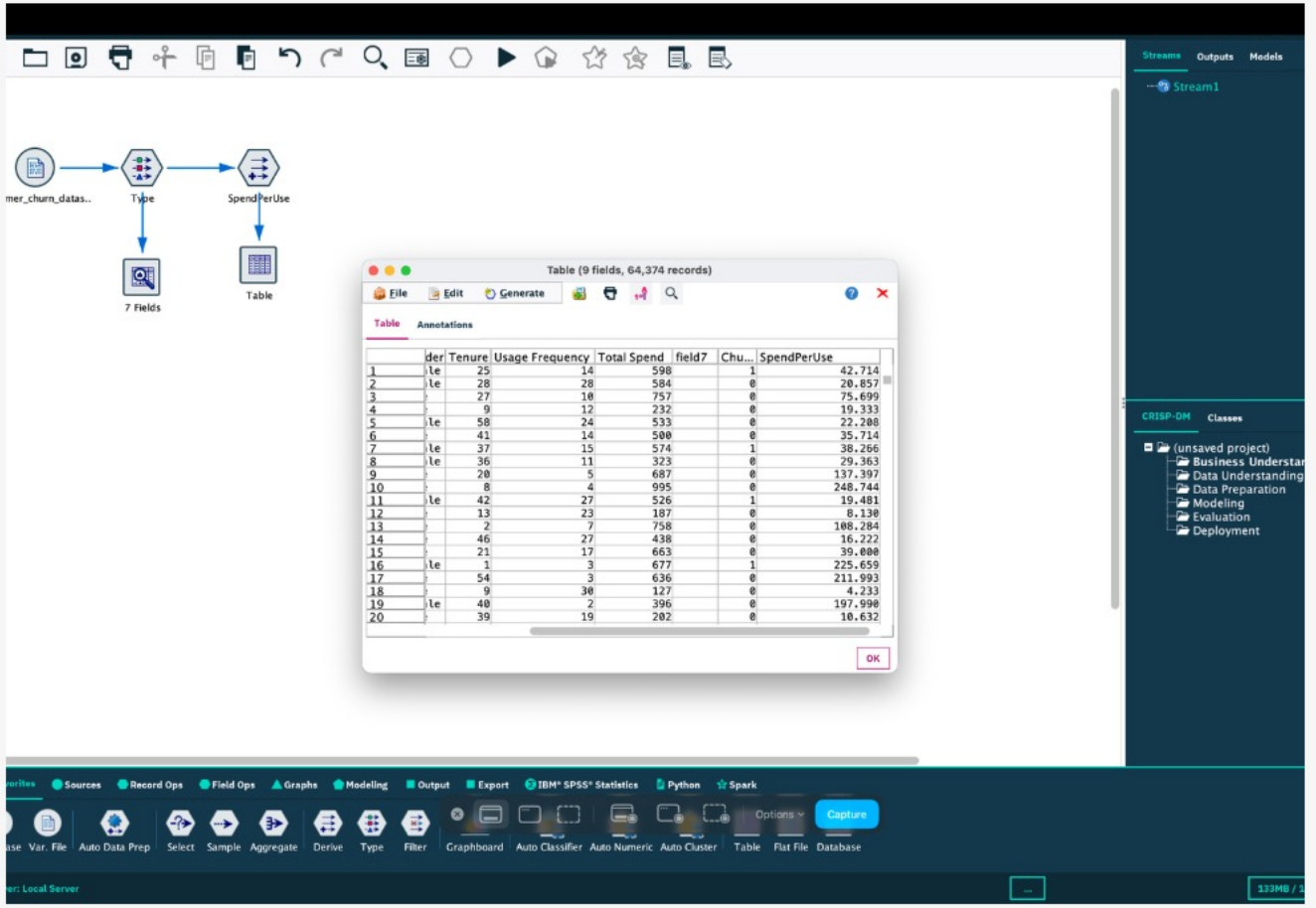
Practical 1

Feature engineering (Derive node)

Add a **Derive** node to create useful predictors.
Connect after Type / Filler.

Examples to create:

SpendPerUse = Total Spend / (Usage Frequency + 0.001) — helps capture intensity.



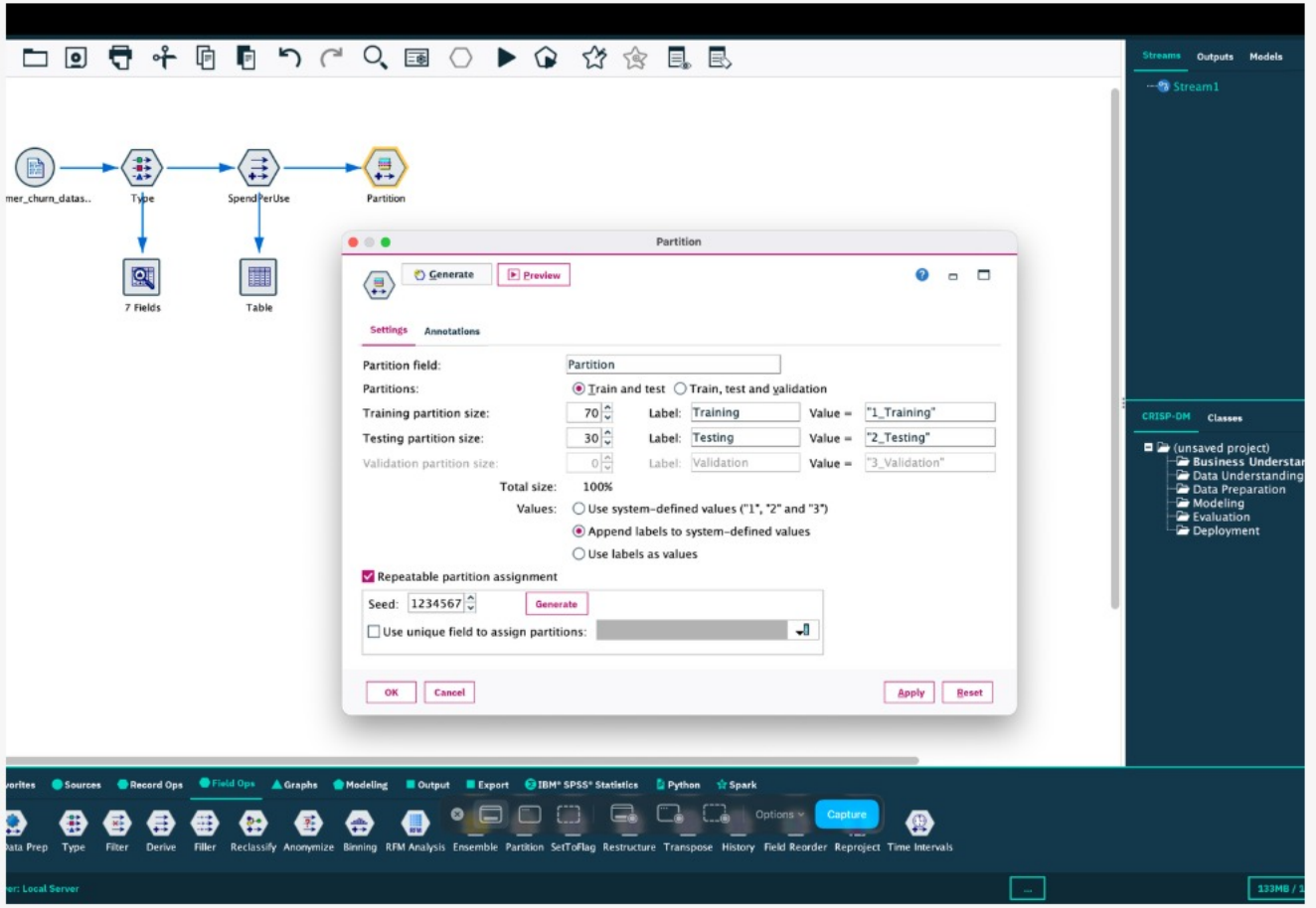
Practical 2

Partition the data

Add a **Partition** node (important for honest evaluation).

Set proportions: **Training = 70%, Test = 30%** (or 60/40). Use random seed for reproducibility (e.g., seed = 12345).

Connect output to modeling nodes (the partition node has ports for training/test — most model nodes will use training, and Evaluation node will use both).



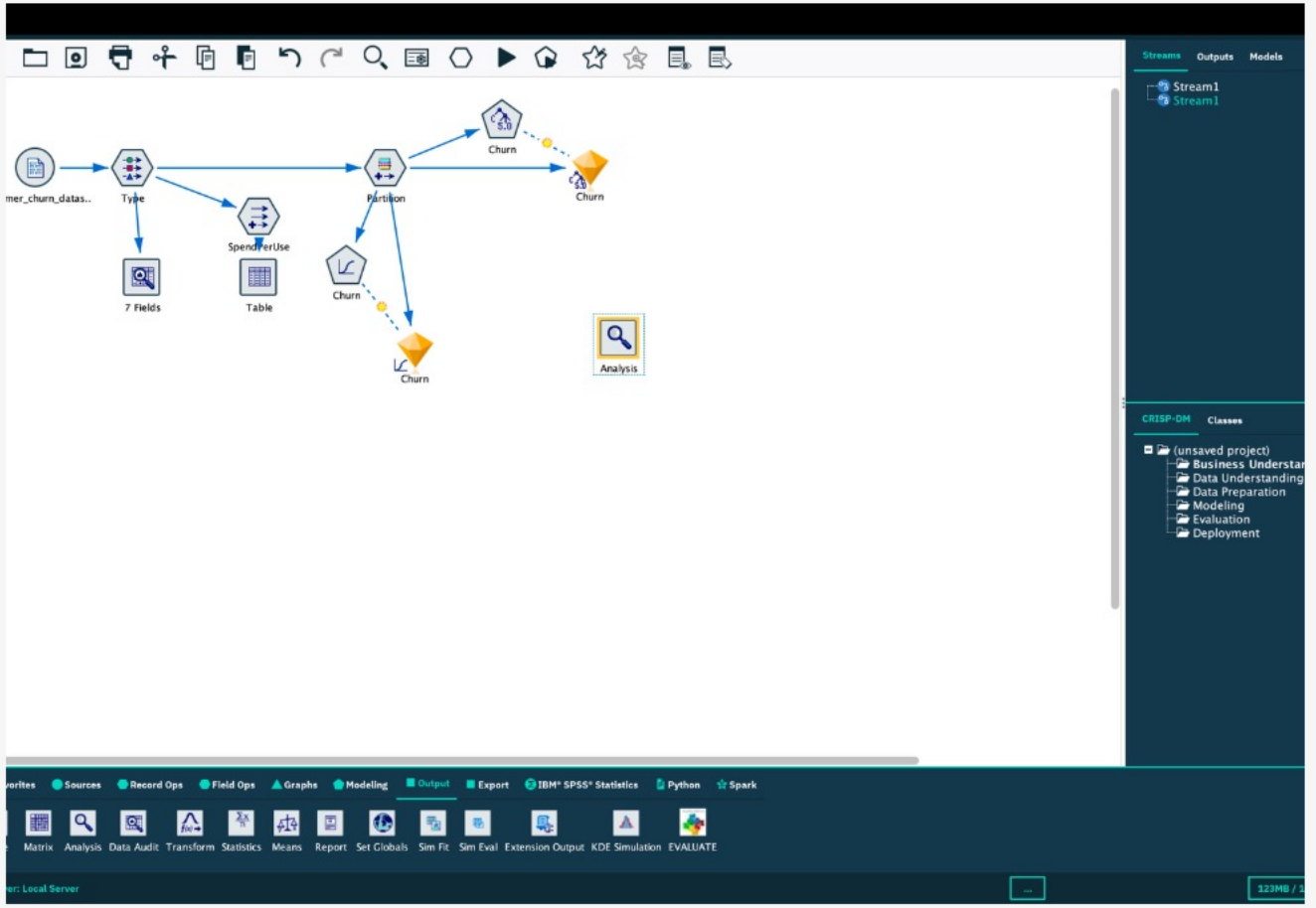
Practical 2

Choose & add models

Add at least two different model nodes so you can compare:

Auto Classifier or **C5.0** (decision tree) — good interpretability.

Logistic Regression (for baseline & coefficients).



Practical 2

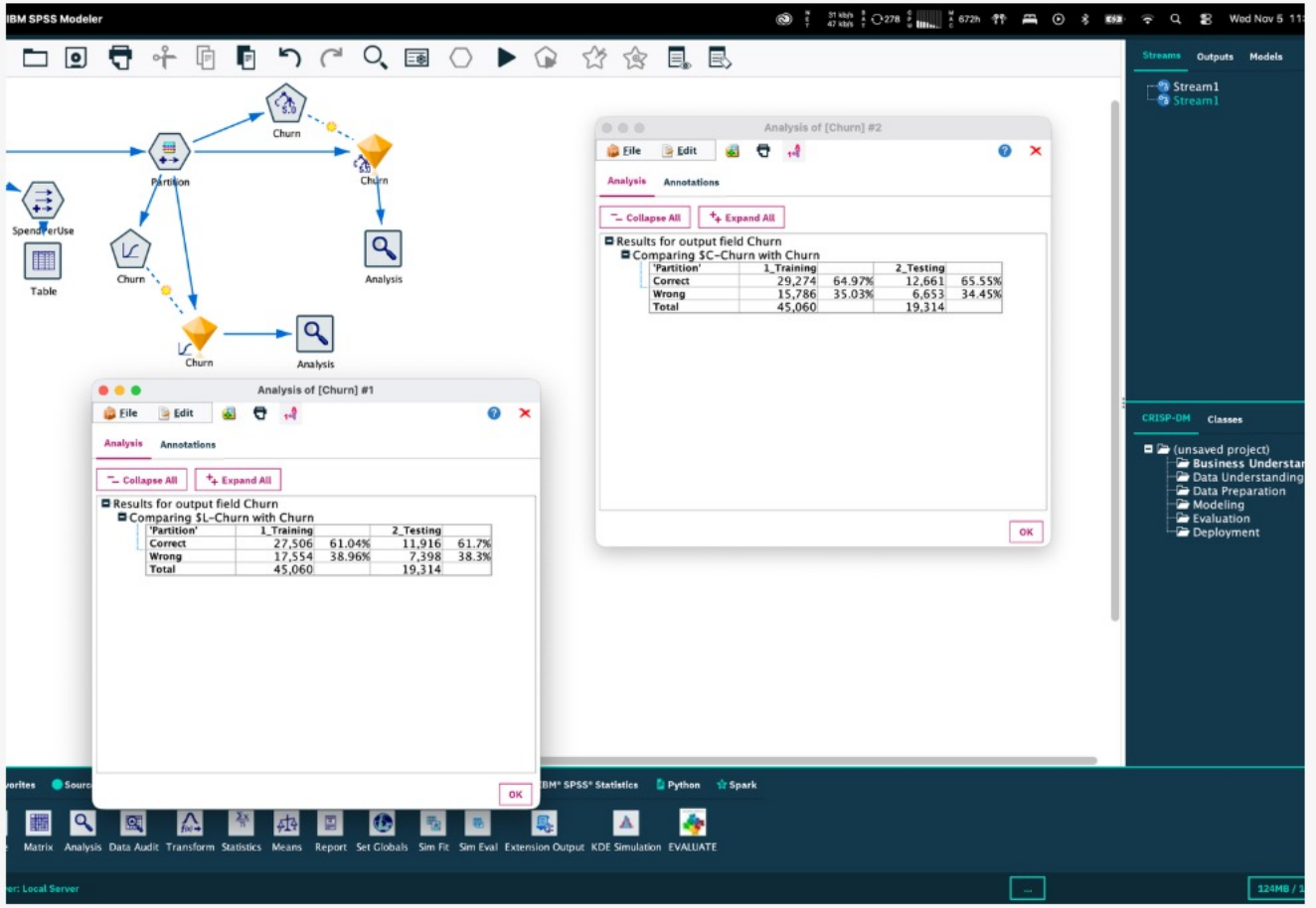
Evaluation node

Add an **Analysis → Evaluation → Analysis / Evaluation** node (called **Analysis**, **Analysis node**, or **Evaluation** depending on version).

Connect all model outputs into the Evaluation node (drag connecting lines from model nodes to Evaluation).

Run the Evaluation node; it will compare models on Test partition and give you:

Confusion matrix, accuracy, precision, recall, ROC curve, AUC, lift chart.



Thank You!!!