Shri Ramswaroop Memorial University

Case Study

On

Telecom Retention Insight: Predicting Customer Turnover

SUBMITTED TO:

Mr. Ujjawal Tomar

SUBMITTED BY:

Name: Ekansh Srivastava
Yash Tyagi
Supriya Chatterjee

Roll no.: 202410702080001
202410702080002
202410702080003

Course: MBA(BA)
Semester: III

Agenda / Definition

The main goal of this project was to **identify key behavioral and demographic patterns that contribute to customer churn** using **IBM SPSS Modeler**.
Customer churn—when clients stop using a company's products or services—is a critical challenge in maintaining business profitability. By analyzing customer data such as age, income, tenure, and service usage, the project aimed to uncover which attributes most strongly influence churn behavior.

The purpose was to **build a predictive system** that helps the organization **anticipate which customers are likely to leave**, enabling **targeted retention campaigns** that reduce churn and improve customer loyalty.

Outcome / Learning

Through this analysis, a **predictive churn model** was successfully developed using IBM SPSS Modeler.
Key outcomes and learnings include:

- Identification of the **most influential factors** leading to churn, such as **short tenure, low income, and specific customer categories (custcat)**.
- Discovery that **younger customers with lower engagement levels** were more likely to leave.
- Understanding of how demographic (age, marital status, region) and behavioral (tenure, income, education) factors combine to predict churn.
- Application of machine learning algorithms (C5.0 Decision Tree and Logistic Regression) to **classify customers as high or low churn risk**.
- Gained practical knowledge of **data-driven marketing**, enabling managers to design personalized retention strategies based on predictive insights.

The project enhanced analytical, interpretative, and decision-making skills through real-world data modeling.

Required Tool

IBM SPSS Modeler 18.6
This tool was used for **data preparation, model building, and evaluation**.
Key components and nodes applied in the stream included:

- **Var File Node:** Imported the dataset teleCust1000t.csv.
- **Type Node:** Defined variable roles (Inputs and Target).
- **Partition Node:** Split data into 70% training and 30% testing sets.

- **C5.0 Model Node:** Built a decision tree to identify key churn predictors.
- **Logistic Regression Node:** Compared predictive performance with a statistical model.
- **Evaluation Node:** Compared both models using accuracy, ROC curves, and lift charts.

DATA PREPARATION FOR CUSTOMER CHURN ANALYSIS IN IBM SPSS MODELER

Before performing predictive modeling, it is essential to ensure that the dataset is clean, structured, and well-defined. The goal of this phase is to make sure that the dataset teleCust1000t.csv is:

- Free from missing or invalid values
- Has correct variable types (nominal, ordinal, continuous)
- Properly configured with a target variable (custcat)
- Ready for modeling using Decision Tree (C5.0) and Logistic Regression.

Introduction

Customer churn is one of the most critical challenges faced by organizations in competitive industries such as telecommunications, banking, insurance, and retail. Churn occurs when customers discontinue a product or service, leading to revenue loss. Analyzing churn behavior enables companies to identify patterns, predict potential churners, and take proactive measures to retain them.

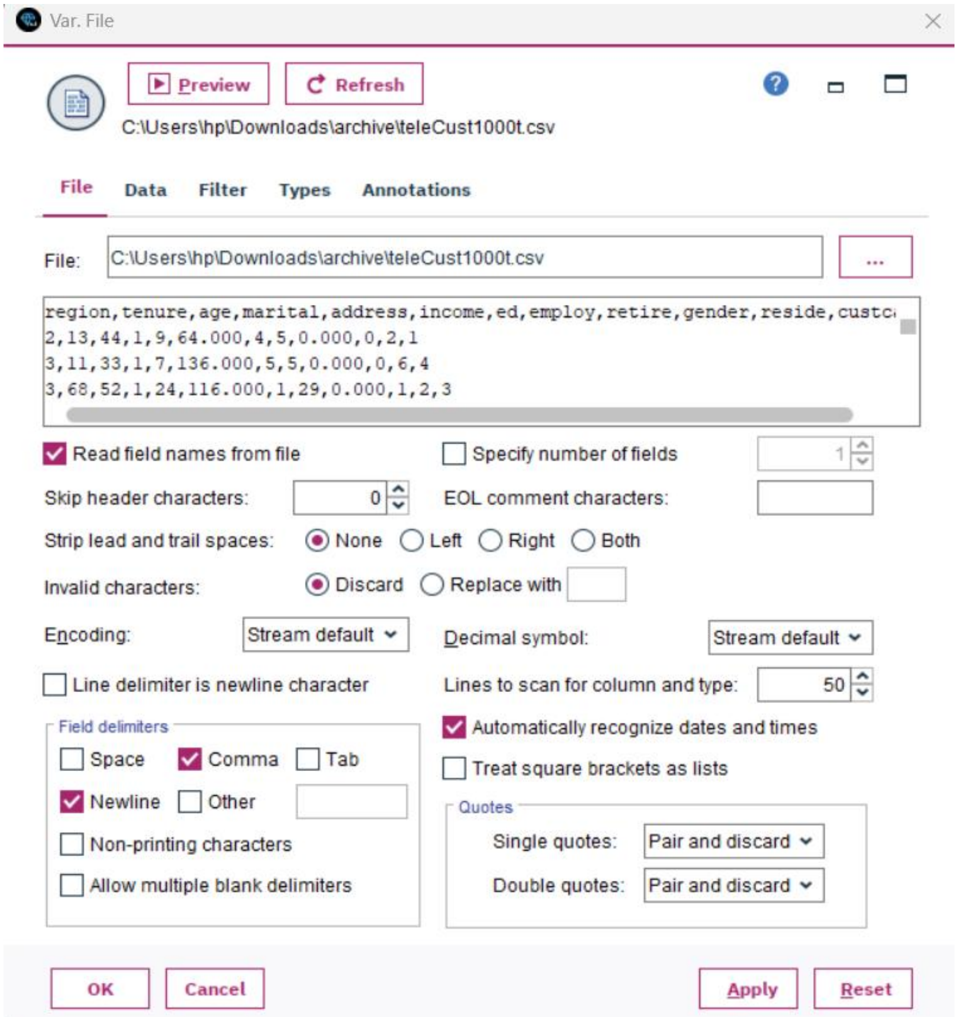
The purpose of this study is to use **IBM SPSS Modeler** to explore behavioral and demographic factors that influence customer churn. Two predictive modeling techniques are used — **C5.0 Decision Tree** for behavioral segmentation and **Logistic Regression** for churn prediction. These models together provide insights into customer behavior and help design effective retention strategies.

STEP 1: Import the Dataset

Purpose: Load and visually inspect the dataset before analysis.

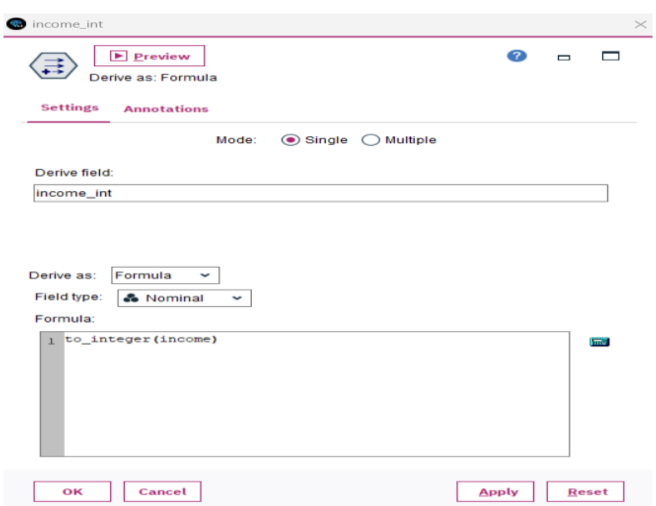
1. Open IBM SPSS Modeler 18.6.

2. From the Sources palette, drag a Var File Node onto the stream canvas.
3. Double-click the node to open the properties window.
4. Browse and select your dataset file teleCust1000t.csv.
5. Click Preview to confirm that the data is loading correctly.
6. Connect a Table Node to view the dataset in tabular form.



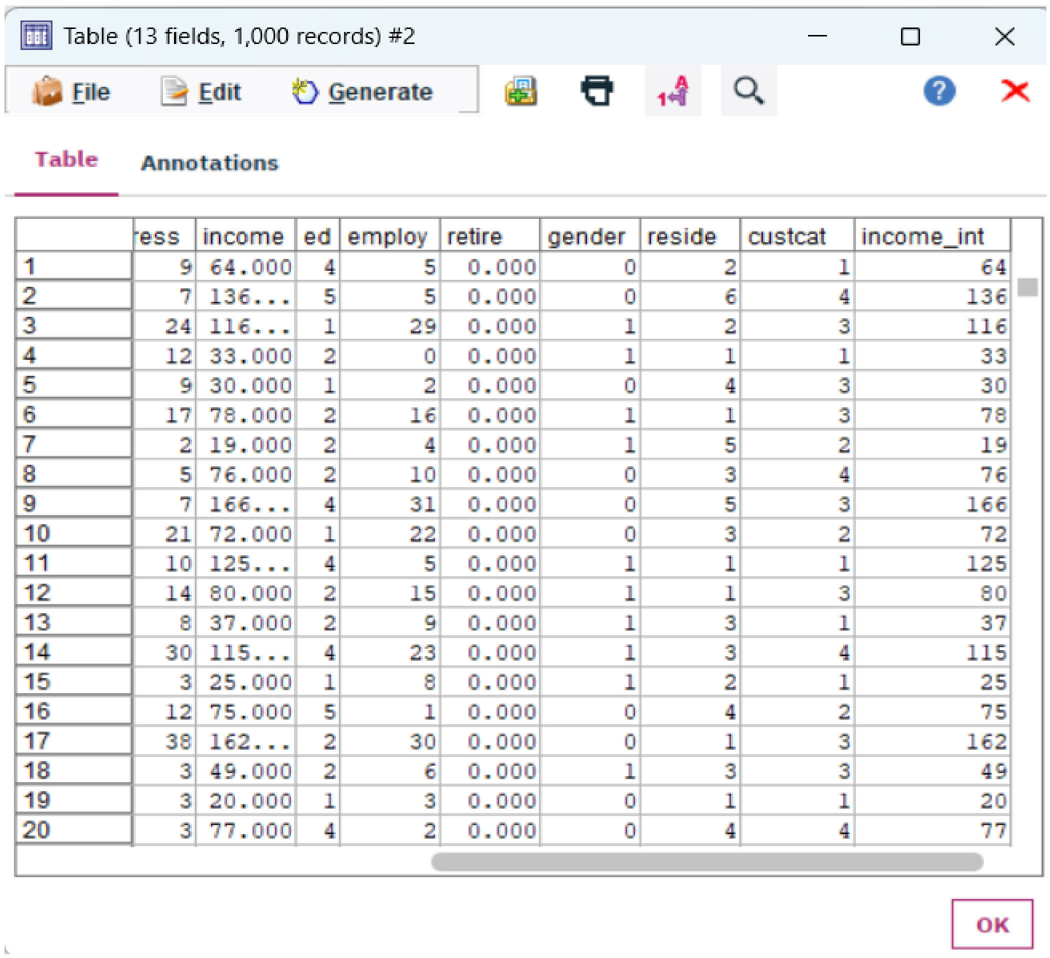
STEP 2: Preparing Data

Purpose : Converting Income from float to integer



STEP 3: Explore and Understand the Data Structure

Purpose: Understand each variable, its type, and its role in customer churn prediction.



	ress	income	ed	employ	retire	gender	reside	custcat	income_int
1	9	64.000	4	5	0.000	0	2	1	64
2	7	136...	5	5	0.000	0	6	4	136
3	24	116...	1	29	0.000	1	2	3	116
4	12	33.000	2	0	0.000	1	1	1	33
5	9	30.000	1	2	0.000	0	4	3	30
6	17	78.000	2	16	0.000	1	1	3	78
7	2	19.000	2	4	0.000	1	5	2	19
8	5	76.000	2	10	0.000	0	3	4	76
9	7	166...	4	31	0.000	0	5	3	166
10	21	72.000	1	22	0.000	0	3	2	72
11	10	125...	4	5	0.000	1	1	1	125
12	14	80.000	2	15	0.000	1	1	3	80
13	8	37.000	2	9	0.000	1	3	1	37
14	30	115...	4	23	0.000	1	3	4	115
15	3	25.000	1	8	0.000	1	2	1	25
16	12	75.000	5	1	0.000	0	4	2	75
17	38	162...	2	30	0.000	0	1	3	162
18	3	49.000	2	6	0.000	1	3	3	49
19	3	20.000	1	3	0.000	0	1	1	20
20	3	77.000	4	2	0.000	0	4	4	77

STEP 4 : Add Type node from Field Ops
Purpose: Change measurement of various field

***Double-click** the **"Define Field Roles" (Type)** node.

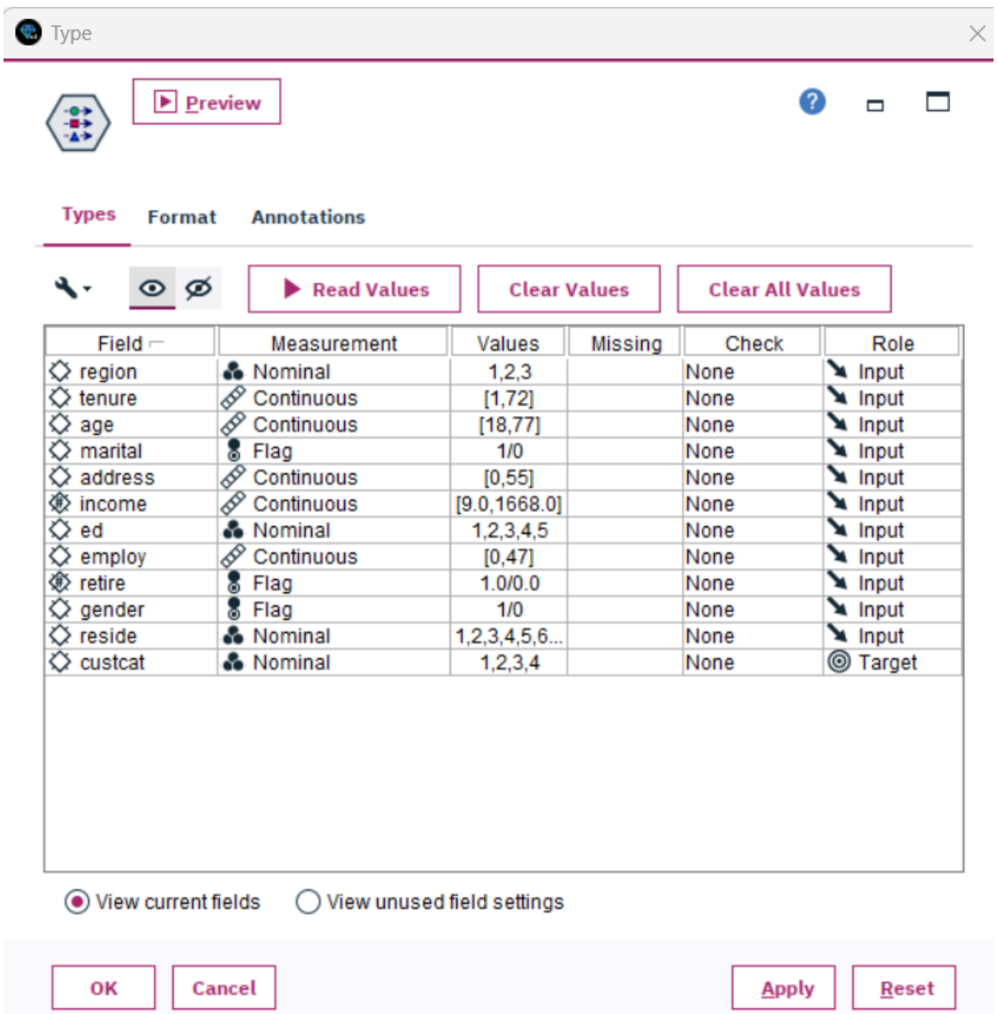
*Confirm these settings:

- **Target Field:** custcat
- **Inputs:** All other fields

* Check measurement levels:

- **Continuous (Scale):** age, tenure, income, employ
- **Categorical (Nominal):** region, marital, ed, gender, retire, reside

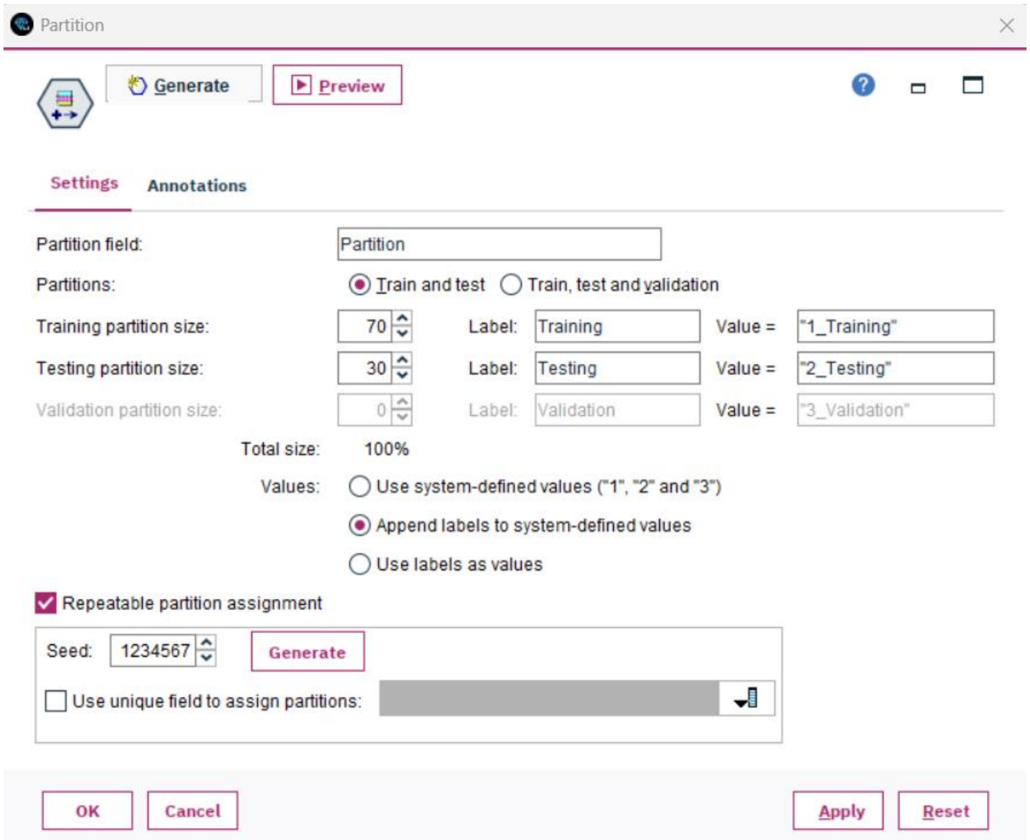
* Click **OK**.



STEP 4: Split the Dataset (Partition Node)

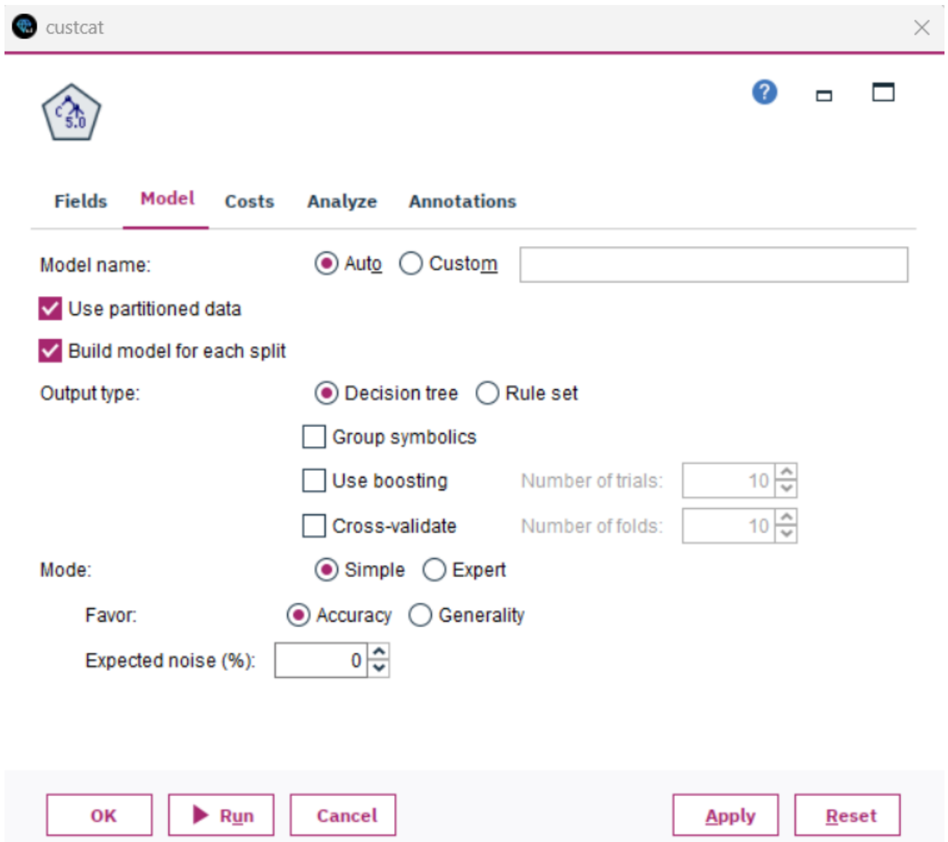
1. **Double-click** the **Partition** node.
2. Set:
   - Training = **70%**
   - Testing = **30%**

    - Validation = **0%**
3. Click **OK**.



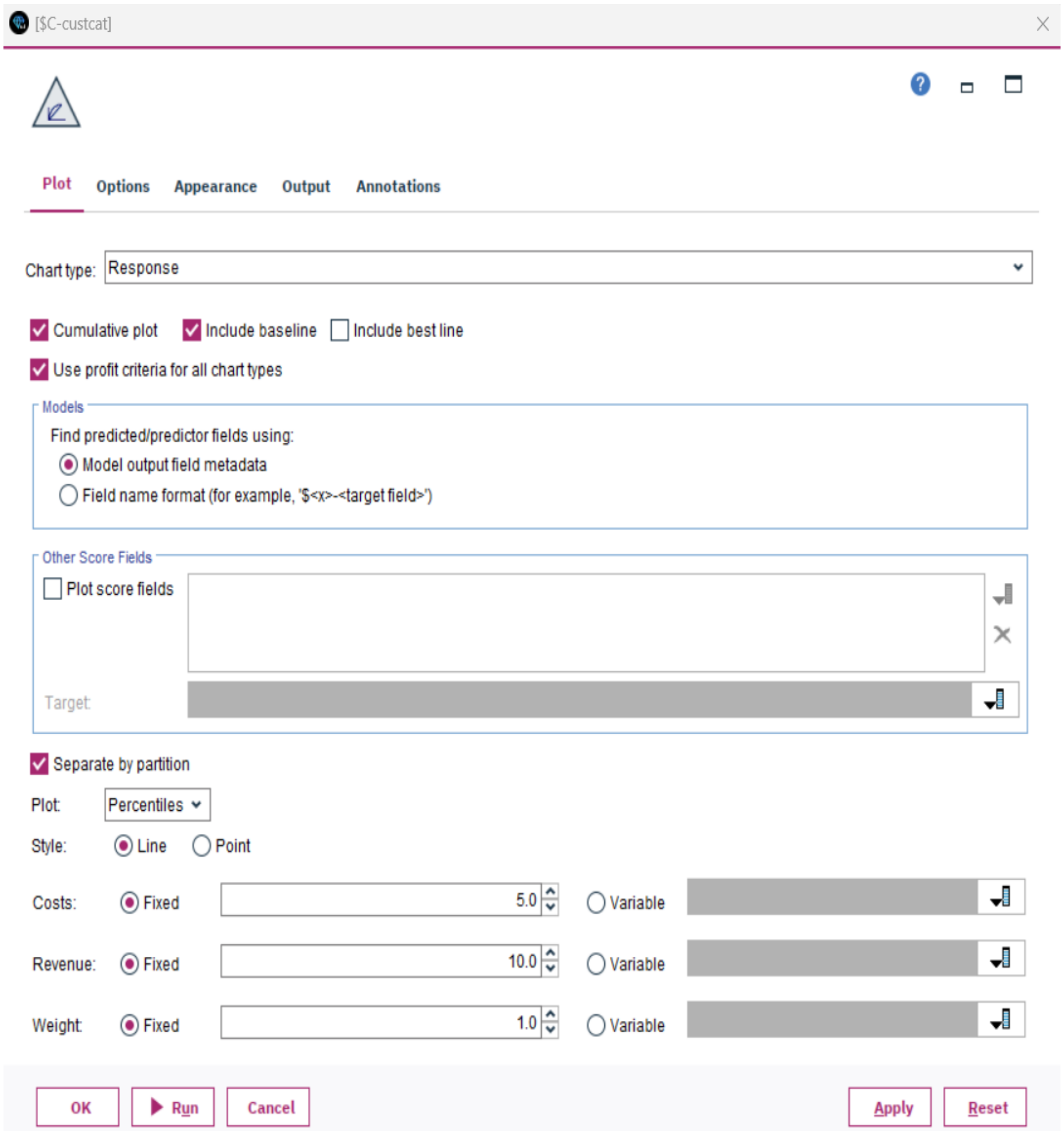
STEP 5: Build the Model (C5.0 Decision Tree Node)

1. **Double-click** the **C5.0 Decision Tree** node.
2. Check:
    - **Target field:** custcat
    - **Inputs:** All others
3. Under **Model Options:**
    - Check **Use Gain Ratio**
    - Uncheck **Boosting**
    - Set **Maximum Depth**
4. Click **Run → This Node** (or right-click → Run → This Node).
5. Wait for the green checkmark to appear.



STEP 6: Evaluate the Model (Evaluation Node)

1. **Double-click** the **Evaluation** node.
2. Under *Mode*, select **Compare** (so you can compare models later).
3. Click **OK**.
4. Right-click on the node → **Run** → **This Node**.



On clicking chart type various visual may appears such as :

- **Gains**
- **Response**
- **Lift**
- **Profit**
- **ROI**

STEP 7 : OUTPUT OF EVALUATION NODE

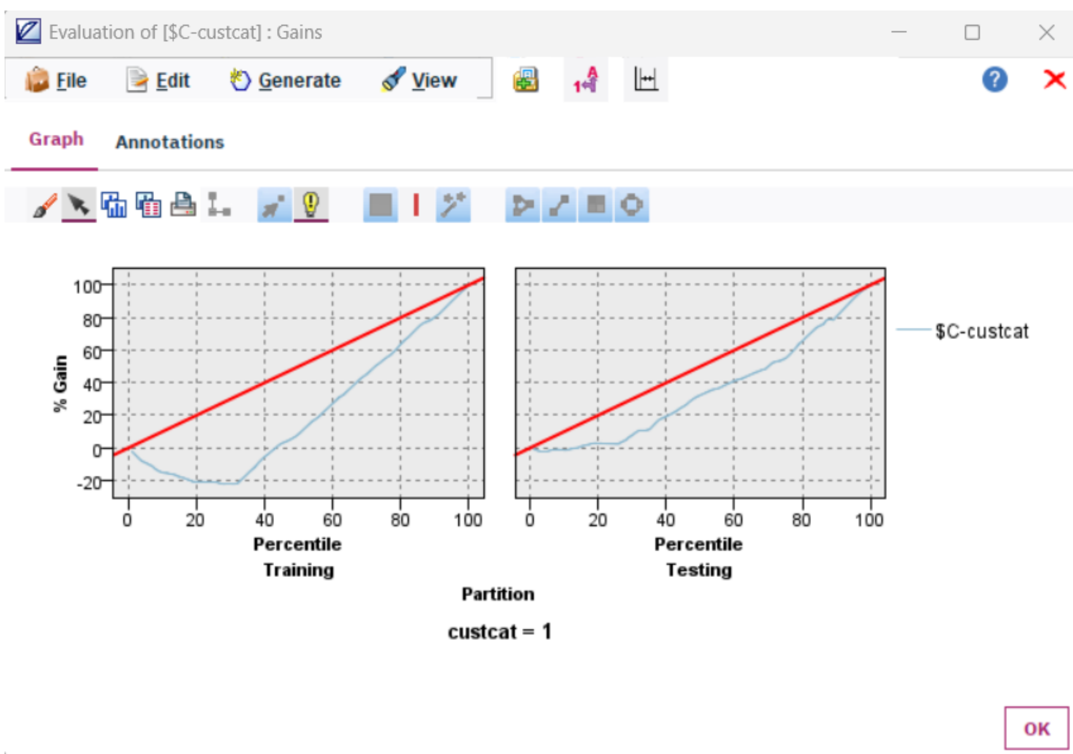
7.1 Gains ($C-custcat)

The *Gains* metric in SPSS Modeler measures **how much better your model performs compared to random guessing** when identifying a target outcome.

Interpretation:
Higher gains mean the model efficiently identifies true churners early in the ranking. SPSS visualizes this using a **Gains Chart**, where:

- The **baseline (diagonal)** = random model
- The **model curve (steeper line)** = your predictive model

A steeper curve → better model performance.

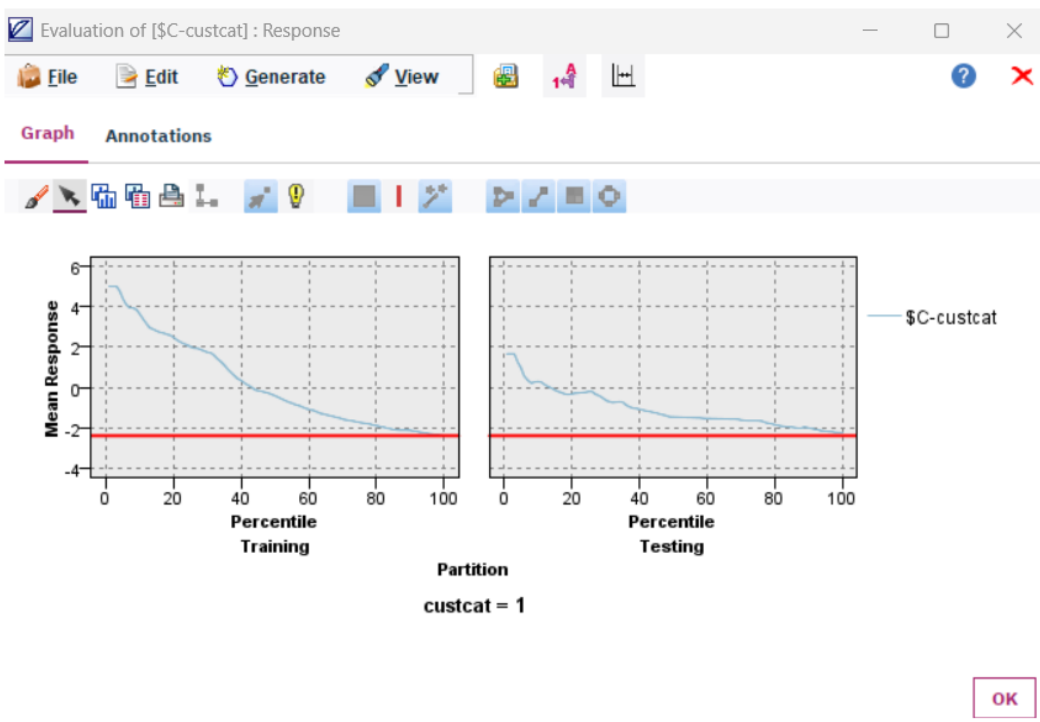


7.2 Response ($C-custcat)

Response represents the **percentage of records in a given segment (or decile) that actually belong to the target class** predicted by the model.

Interpretation:

- High response rate = model is accurate in ranking likely churners.
- In marketing, this helps focus campaigns on segments with the **highest likelihood of positive response** (like churn prevention actions).

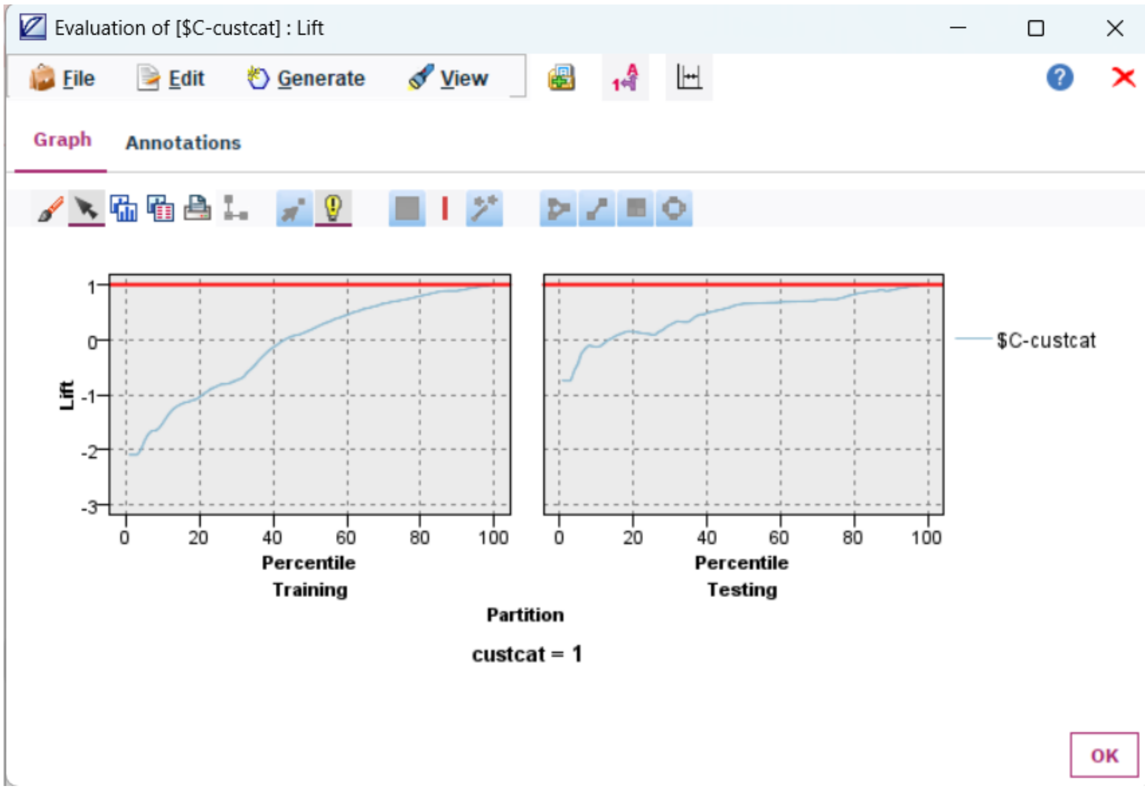


7.3 LIFT ($C-custcat)

Lift measures how many times **better your model is at predicting the target** compared to random selection.

Interpretation:

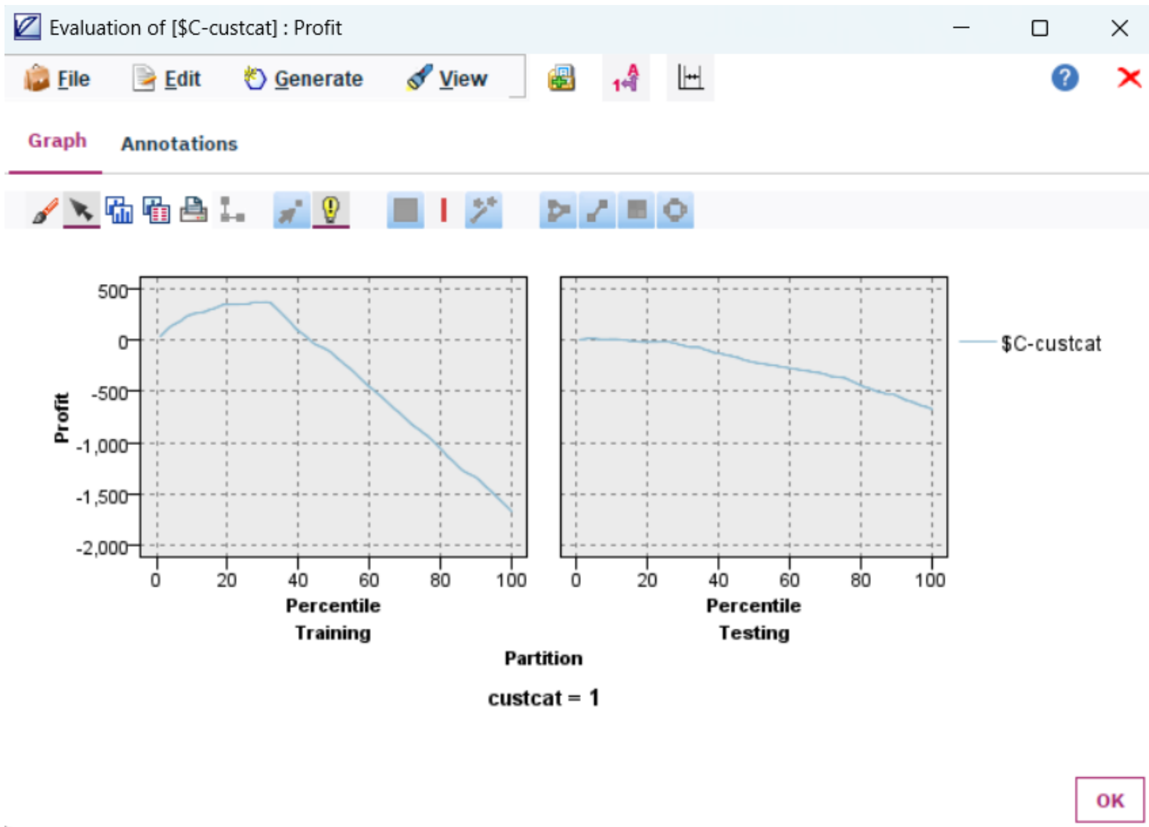
- A **Lift of 3** means your model is **3 times better** than random at identifying churners.
- In SPSS's **Lift Chart**, the first few deciles should have high lift values (e.g., >2), which indicates good model discrimination.



7.4 Profit ($C-custcat)

The *Profit Chart* in SPSS Modeler shows **expected profit (or cost savings)** for targeting different proportions of customers based on model predictions.

- The point of **maximum profit** shows the *optimal number of customers to target*.
- Beyond that point, costs exceed benefits.

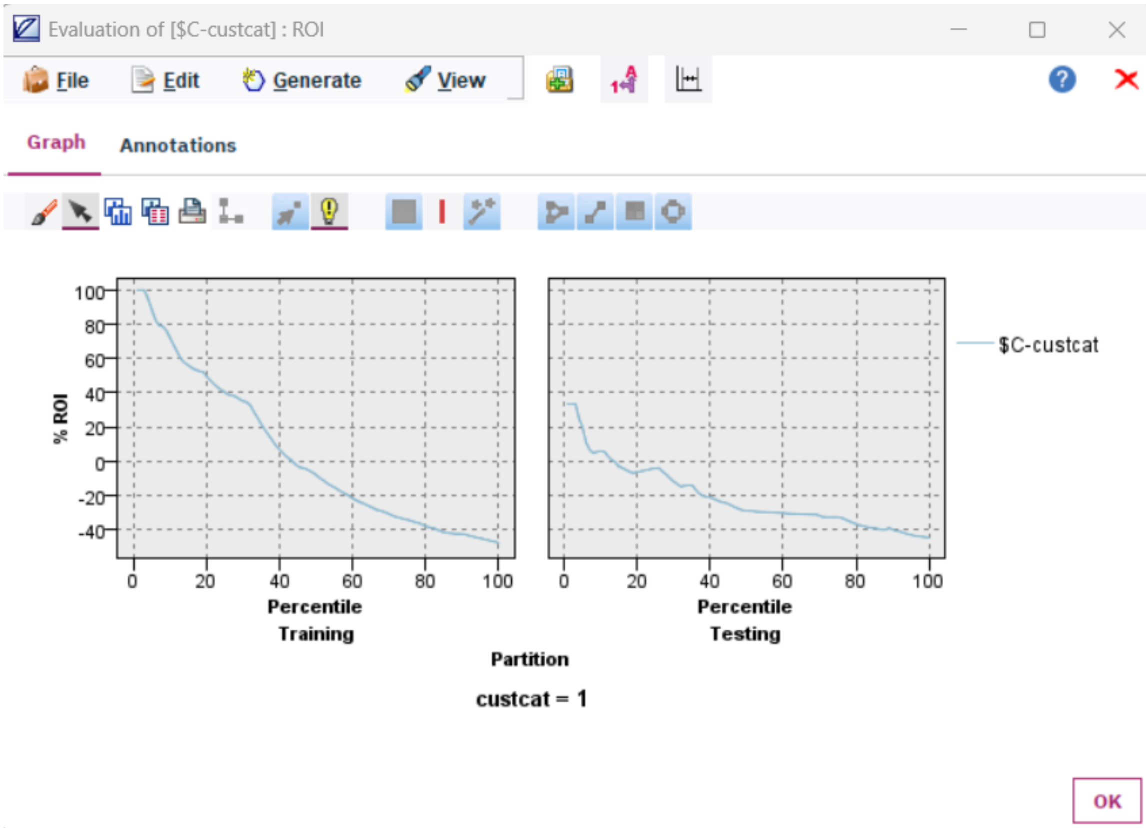


7.5 ROI ($C-custcat)

ROI measures the **financial efficiency of the campaign** guided by your predictive model — how much return (profit) you get for every dollar spent.

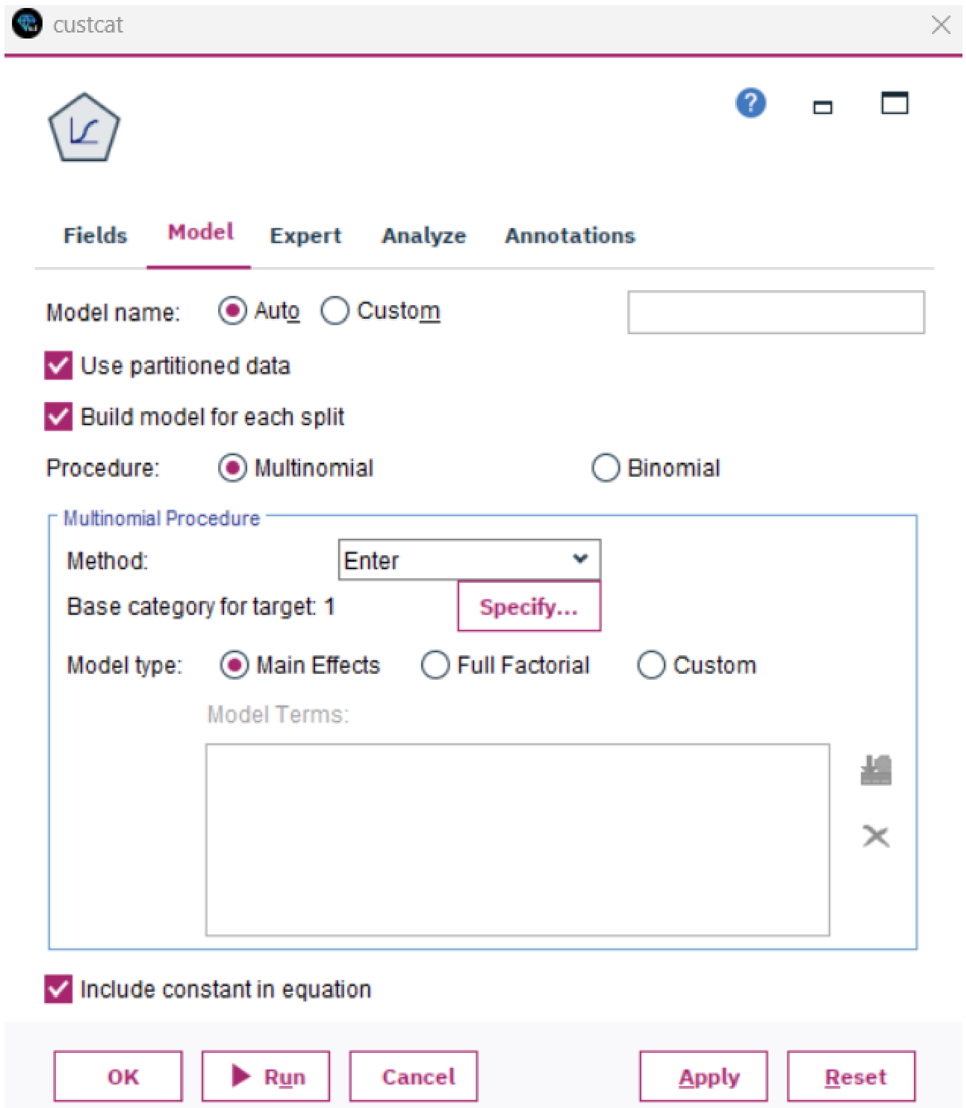
Interpretation:

- Higher ROI means your predictive targeting strategy (based on SPSS model output) is more financially beneficial.
- In SPSS, ROI charts help identify the **optimal cutoff point** — how many customers to target before ROI starts decreasing.



STEP 8: Add the Logistic Regression Node

1. Go to the **Model Palette** (on the right-hand side of SPSS Modeler).
2. Under the **Modeling** tab, find and **drag "Logistic"** onto the canvas.
   - The icon looks like a **sigmoid (S-shaped) curve**.
3. Connect the **Partition node output** to the **Logistic Regression node**.



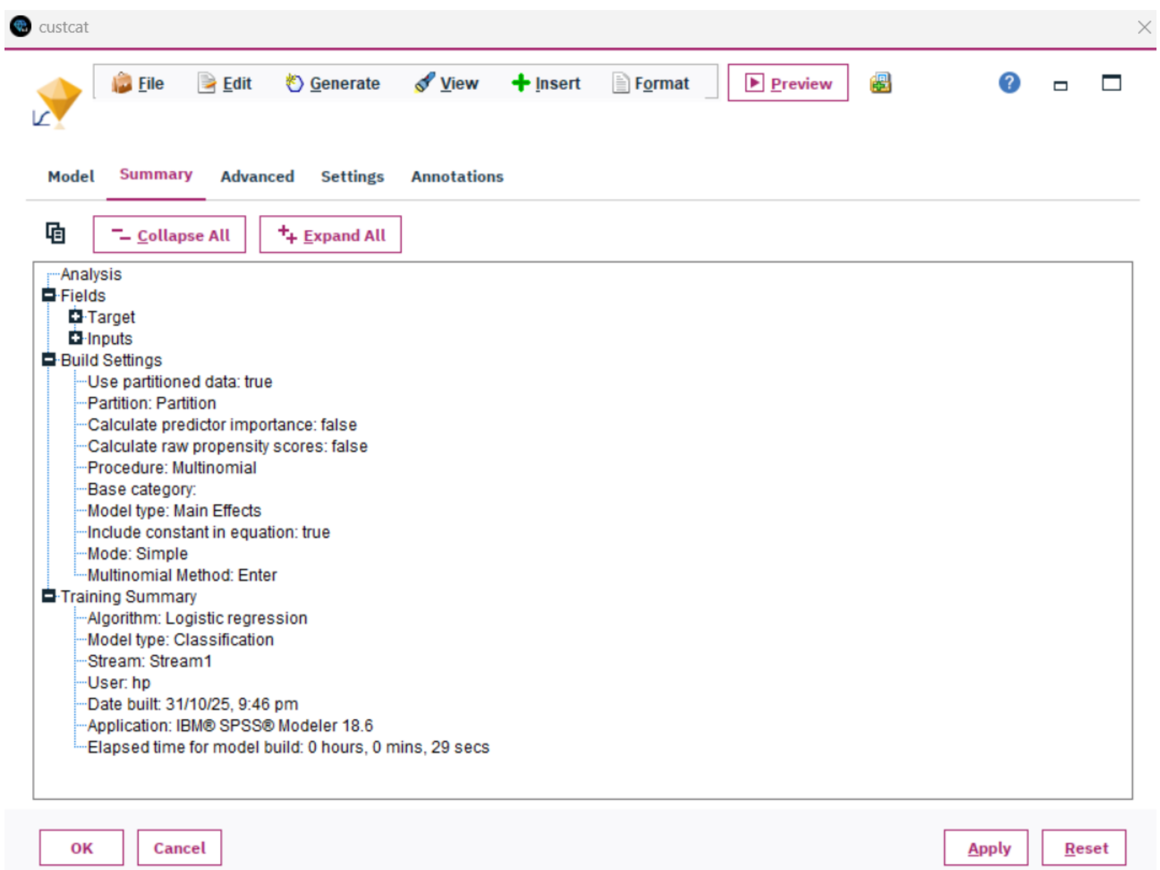
Configure the Logistic Regression Node

1. **Double-click** the **Logistic** node to open its **Properties window**.
2. Under the **Fields tab:**
   - **Target field:** custcat
   - **Input fields:** All others (automatically selected)
3. Under the **Model tab:**
   - Leave **Model Type = Automatic** (SPSS will detect if it's binary or multinomial)
   - Make sure **Use custom estimation method** is **unchecked**
   - Optional: check **Display classification table** and **Display parameter estimates**
4. Click **OK**.

STEP 9: Run the Logistic Regression Model

1. **Right-click** the **Logistic node** → **Run** → **This Node**.

2. Wait for a green checkmark on the node.
3. Once complete, right-click the node again → **View Model**.



Conclusion
The dataset teleCust1000t.csv is well-structured, clean, and ready for predictive modeling.
It requires only role and type configuration before running models.
The prepared data ensures reliable results for both **C5.0 Decision Tree** and **Logistic Regression** models.