BABU BANARASI DAS UNIVERSITY

Predictive Analytics

ON

Super-store-ship-Churn

SUBMITTED TO:
Mr. VIKAS SIR
Roll No: 1230258325

SUBMITTED BY:
Name: Rahul Mishra

Project Overview

This project focuses on analyzing the shipment behavior of customers using the Sample Superstore dataset. The primary objective is to understand how different factors — such as customer segment, region, category, and order priority — influence the mode of shipment. Using IBM SPSS Modeler, a CHAID (Chi-squared Automatic Interaction Detection) model is built to classify and predict the most likely shipping method chosen by customers.

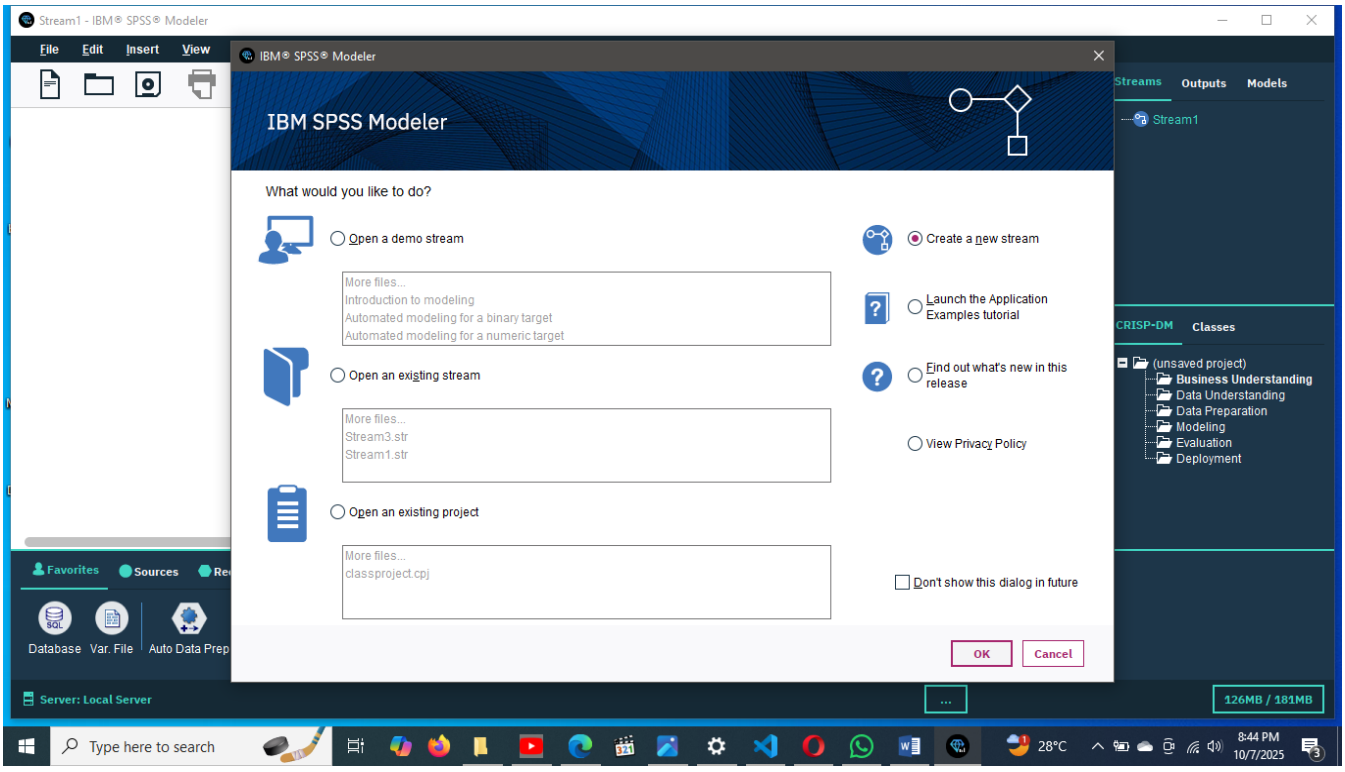
Problem Statement

The project aims to analyze product shipment patterns to predict the ship mode used by customers. Using IBM SPSS Modeler, a CHAID classification model is developed to identify key factors influencing shipping decisions. Here, "Second Class" shipments are considered potential churn cases due to delayed or unsatisfactory delivery experiences. The goal is to help the business improve logistics efficiency and reduce churn by optimizing shipping strategies.

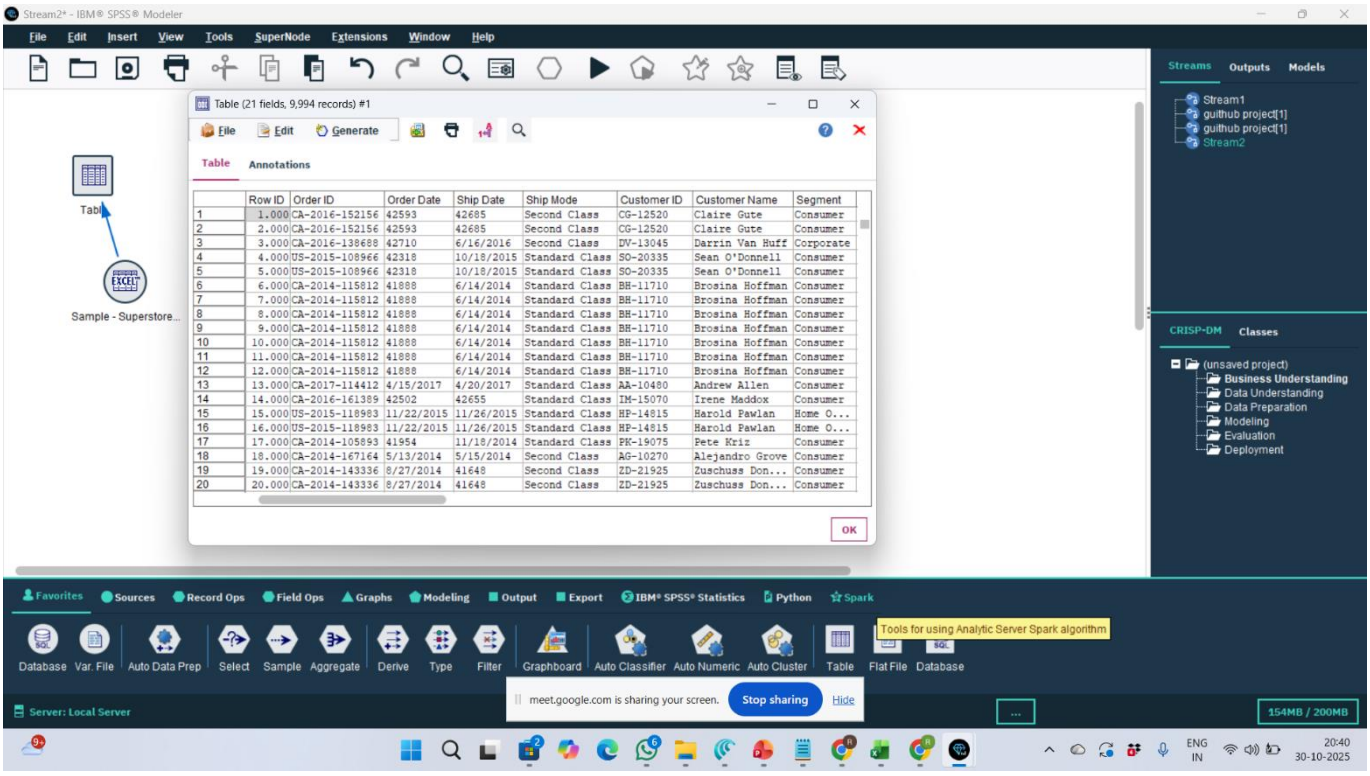
Tools & Technologies Used Category Tools / Technologies Purpose

| Category | Tools / Technologies | Purpose |
| --- | --- | --- |
| *Data Source* | Microsoft Excel (Sample Superstore Dataset) | Used for importing and managing raw shipment data |
| *Data Analysis & Modeling* | IBM SPSS Modeler | For building the CHAID classification model and performing data preprocessing |
| *Algorithm* | CHAID (Chi-squared Automatic Interaction Detection) | Used to classify and predict the shipping mode and identify key influencing factors |
| *Visualization* | IBM SPSS Charts & Decision Tree Output | For visual representation of model results and relationships between predictors |
| *Documentation* | GitHub & Markdown | Used to host the project and create structured documentation (README) |

Step 1 LET's Start the Model

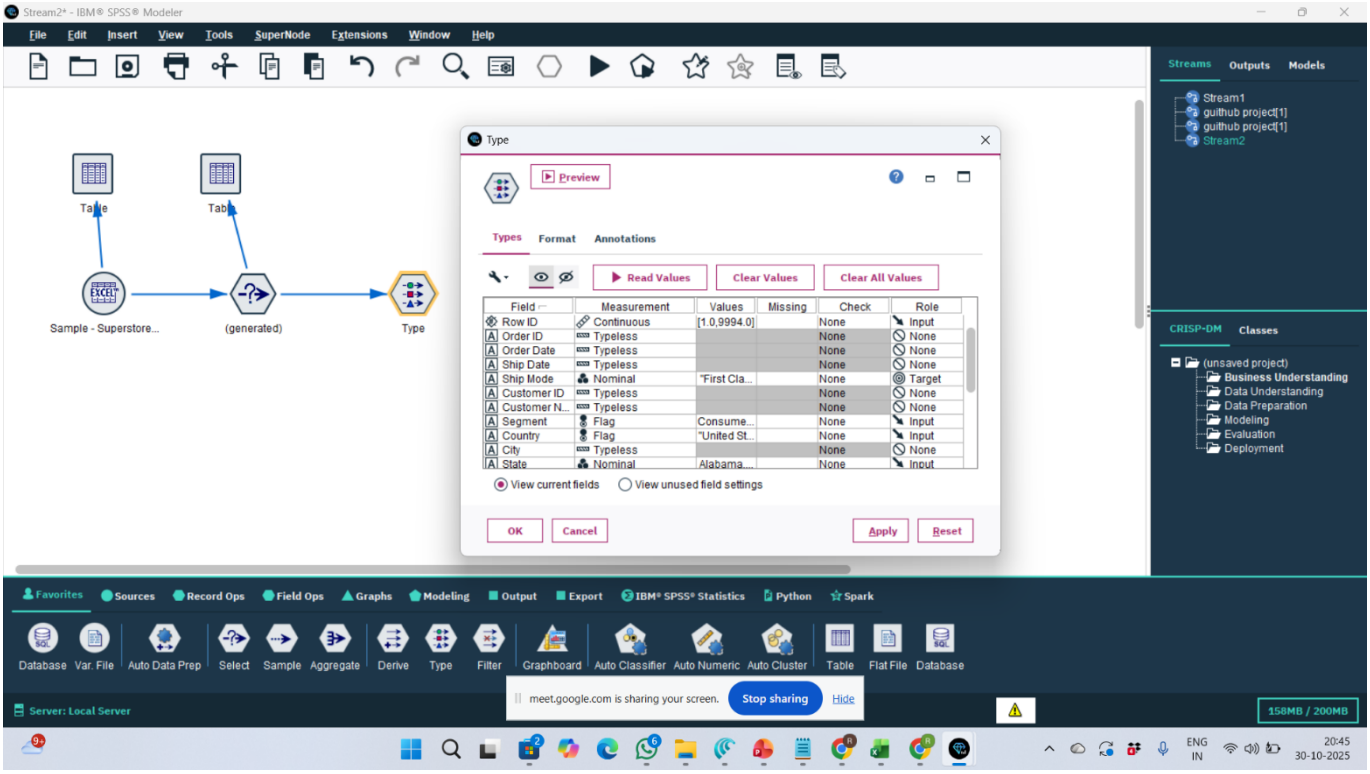


Step 2: From the palette, drag and drop the 'Excel. File' node under the 'Sources' tab to import a dataset (for example, a Excel file containing super-store data).

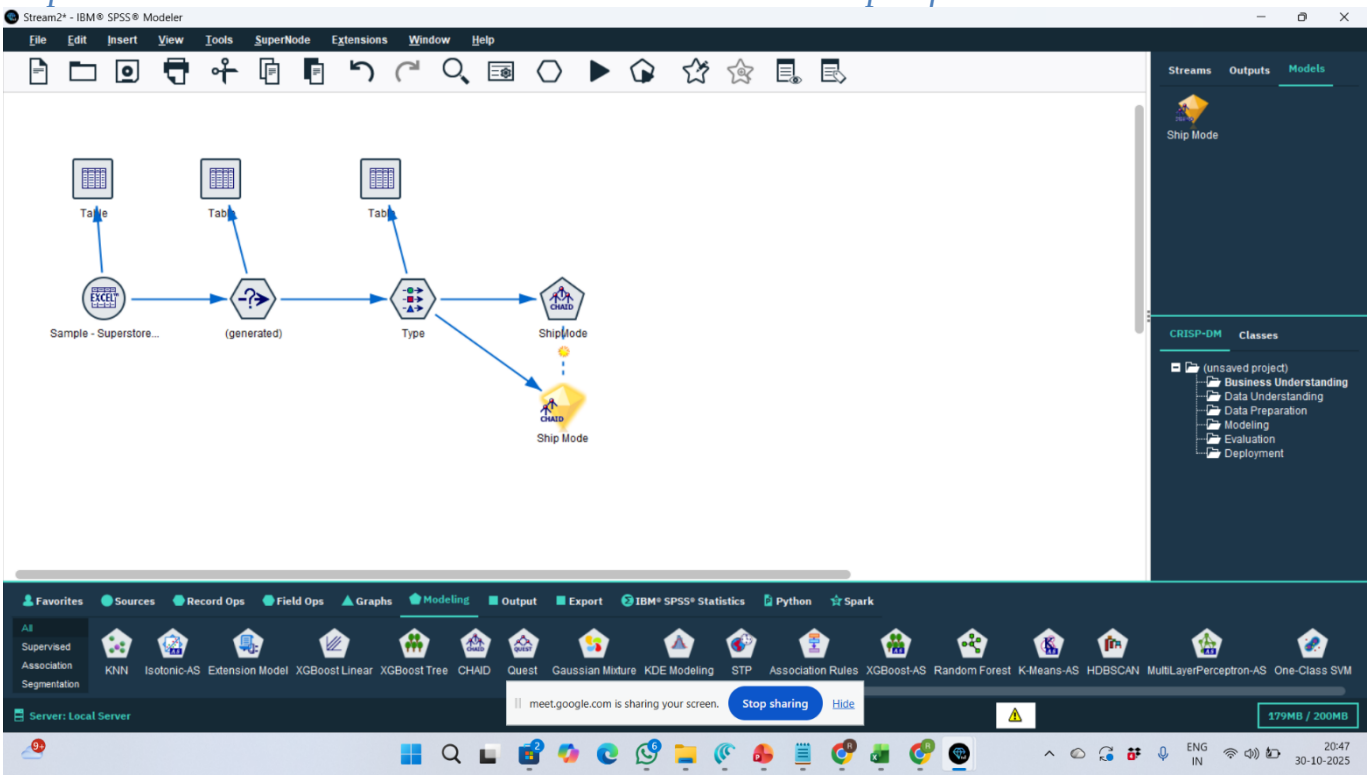


Double-click the node and browse to select your dataset. Click OK to load it.

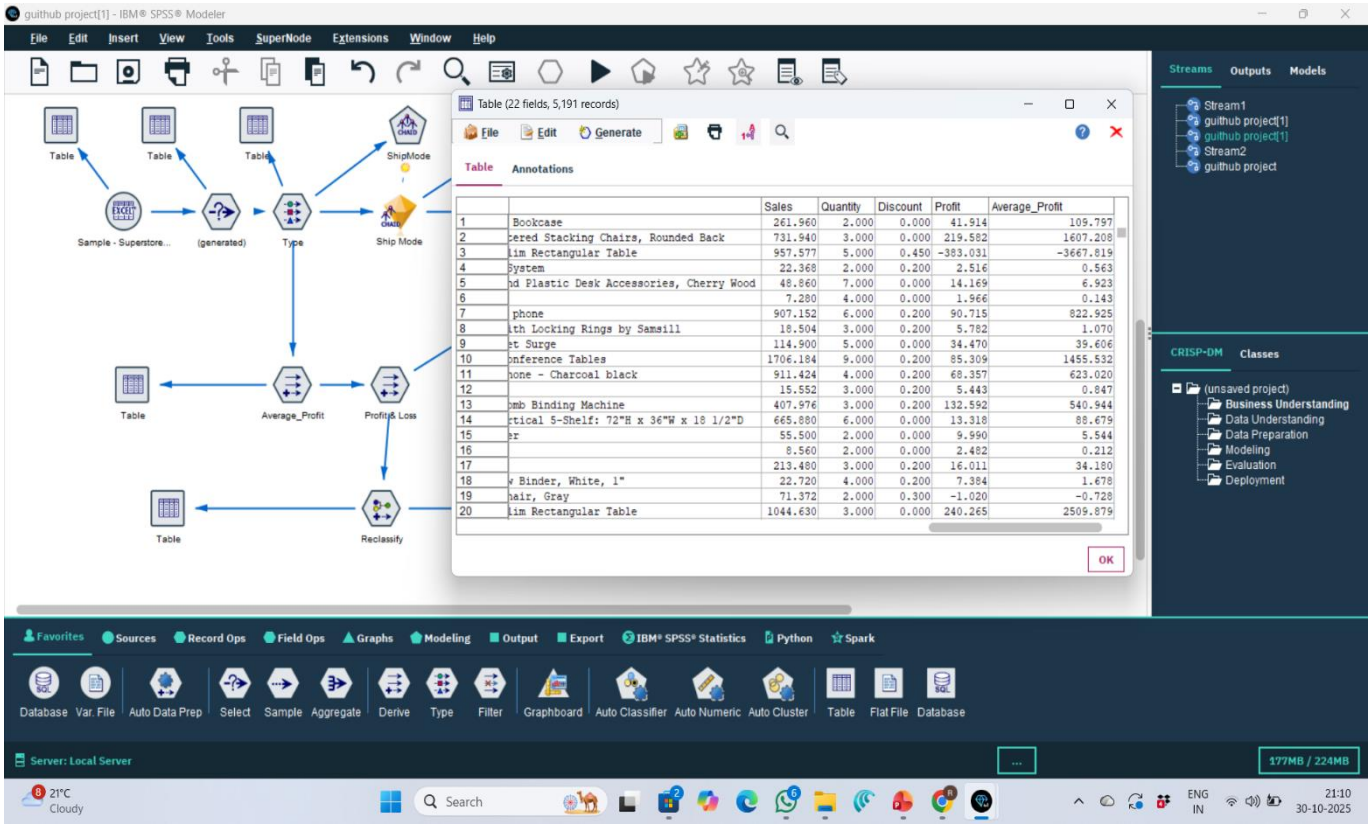
Step 3: Next, from the 'Record Ops' tab, drag and drop the 'Select' node. Connect it to the data import node.



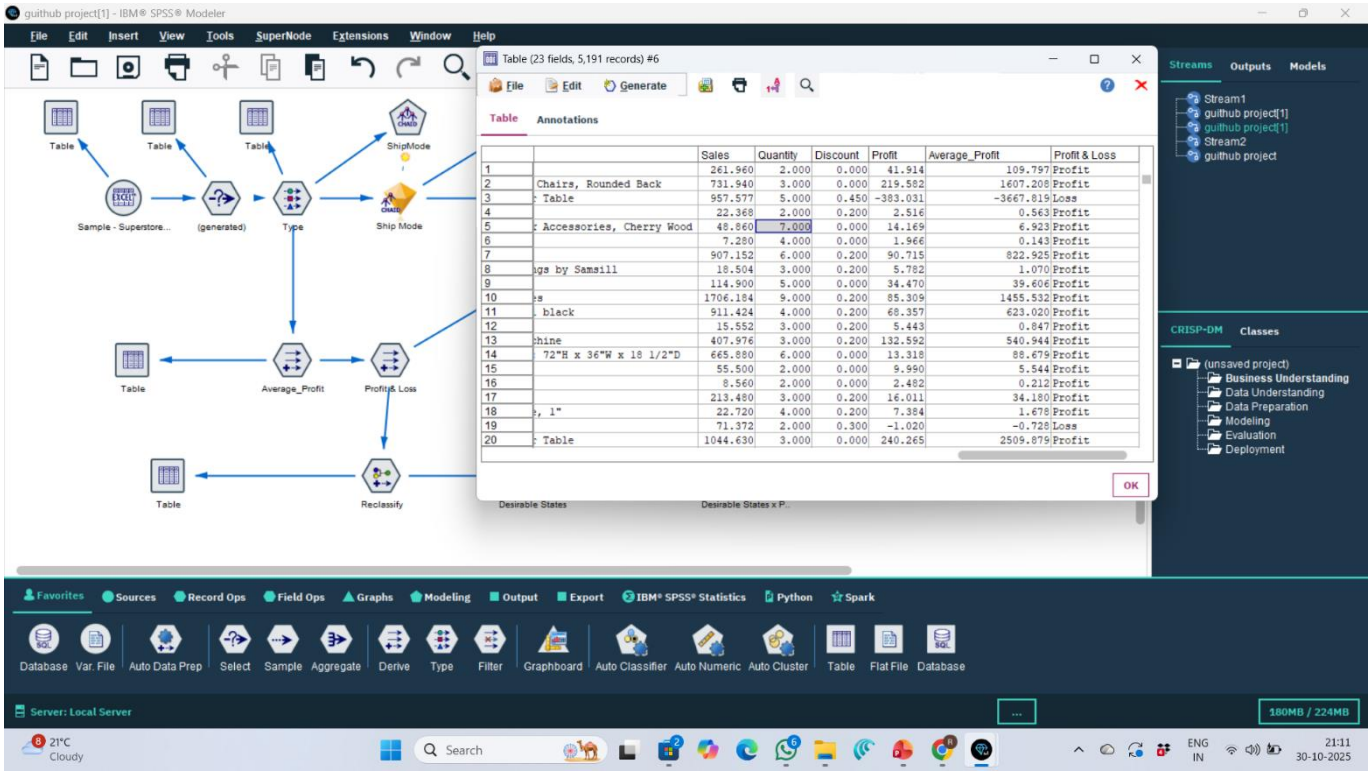
Step 4: Double-click and set a churn condition to select specific records



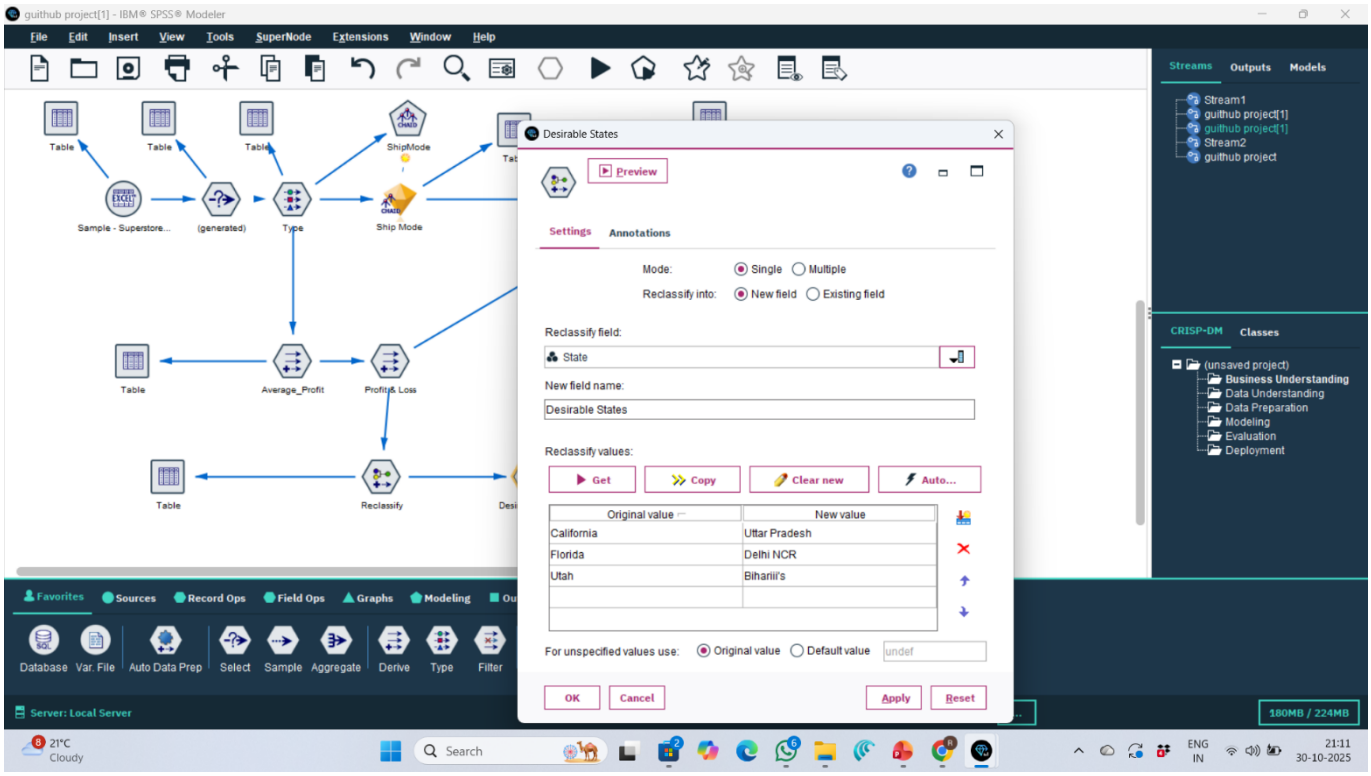
Step 5 and find the avg profit in sales



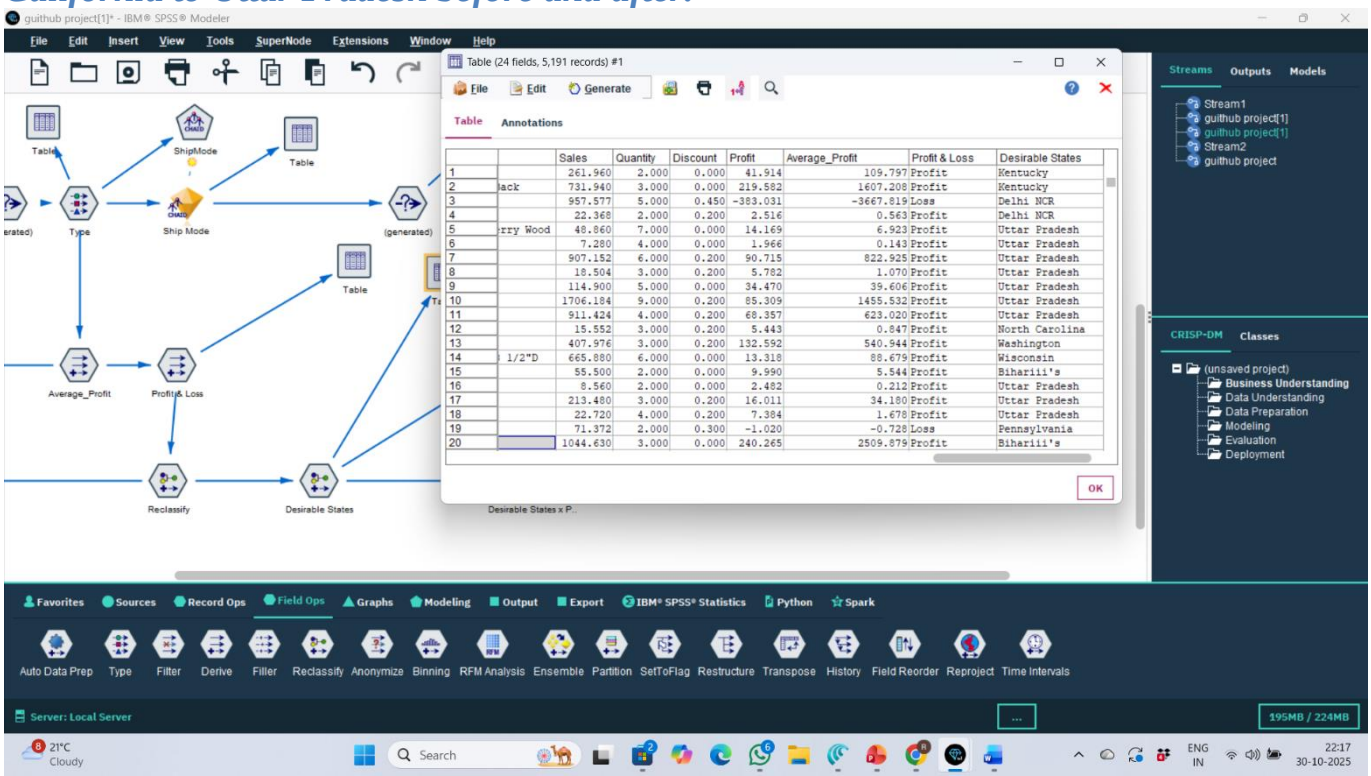
Step 6: Find the profit and loss statement for the data set. Define that a negative value (–) in the avg_profit column indicates a loss statement.



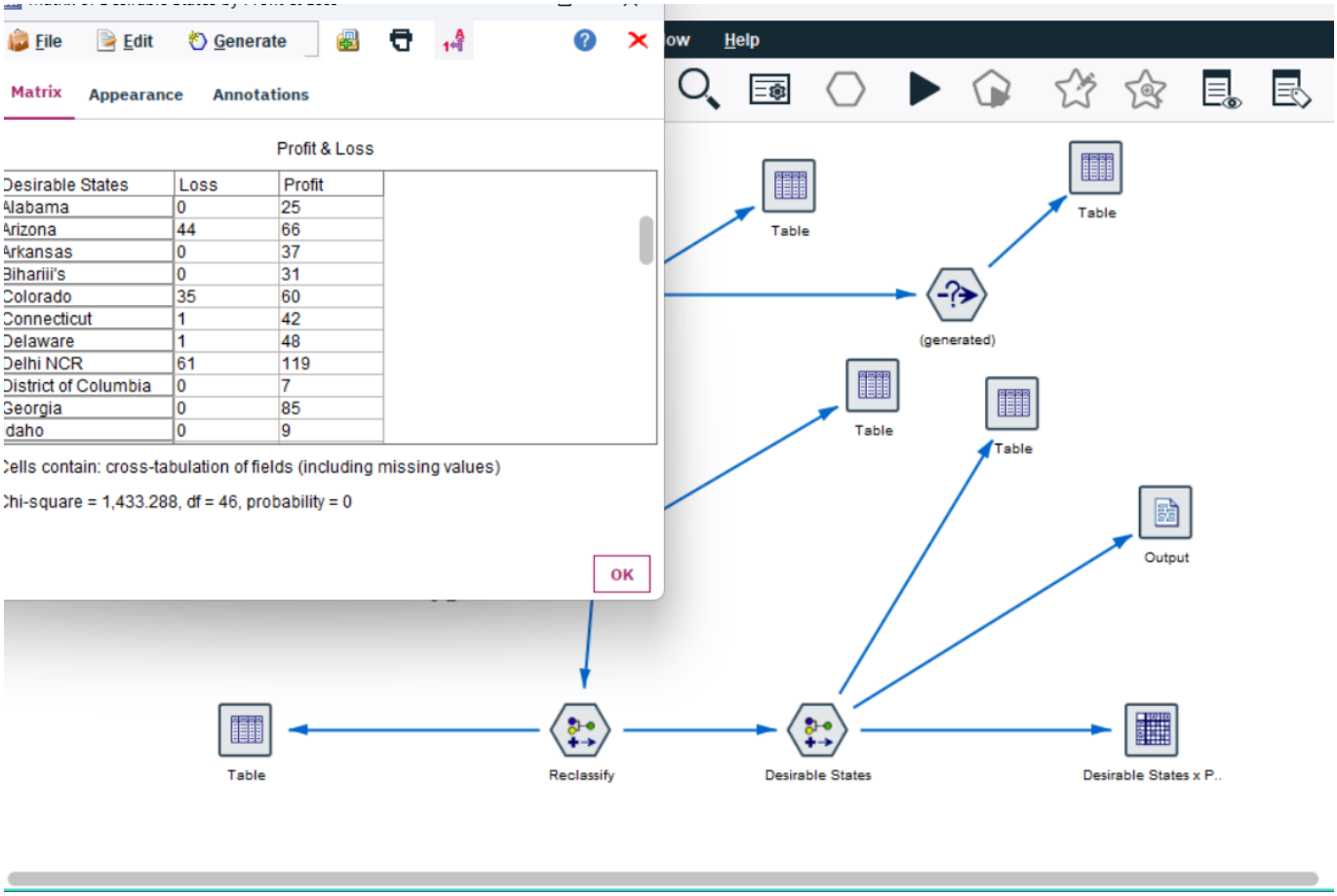
Step 7: The report shows desirable sales using the Re-Classify method.



Step 8: The report shows desirable sales. You can see the changes in city names from California to Uttar Pradesh before and after.



Step 9: The report shows metrics of profit and loss.



Step 10: Final output of the model