Practical File

Name : Ganesh Agrhari

Subject : Predictive Analytics (BCADSN15301)

Class : BCA DS&AI; 33

Roll No. : 1230258126

Submitted To : Mr. Robin Tyagi

Practical: 3

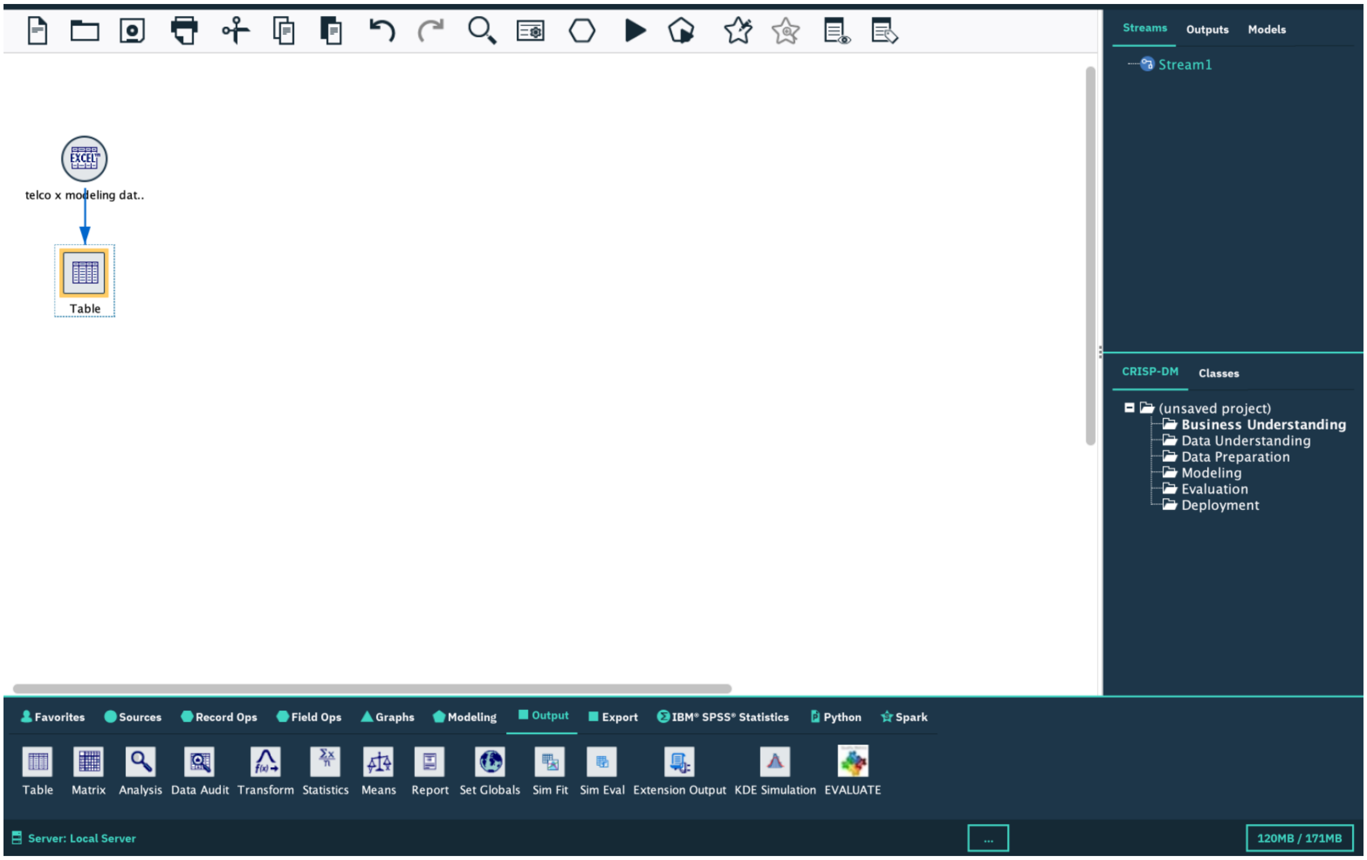
Definition: You work as a data miner for a telecommunications firm. You want to identify customers who are likely to cause churn by cancelling their subscription. Therefore you will build a model on historical data and apply this model to current customers.

Outcomes/Learning: From this process, we learned how to prepare data by filtering for a specific subset and then use a pre-built model to score the new data. The outcome is a final list of high-risk customers who are likely to churn, which can be used for targeted campaigns.

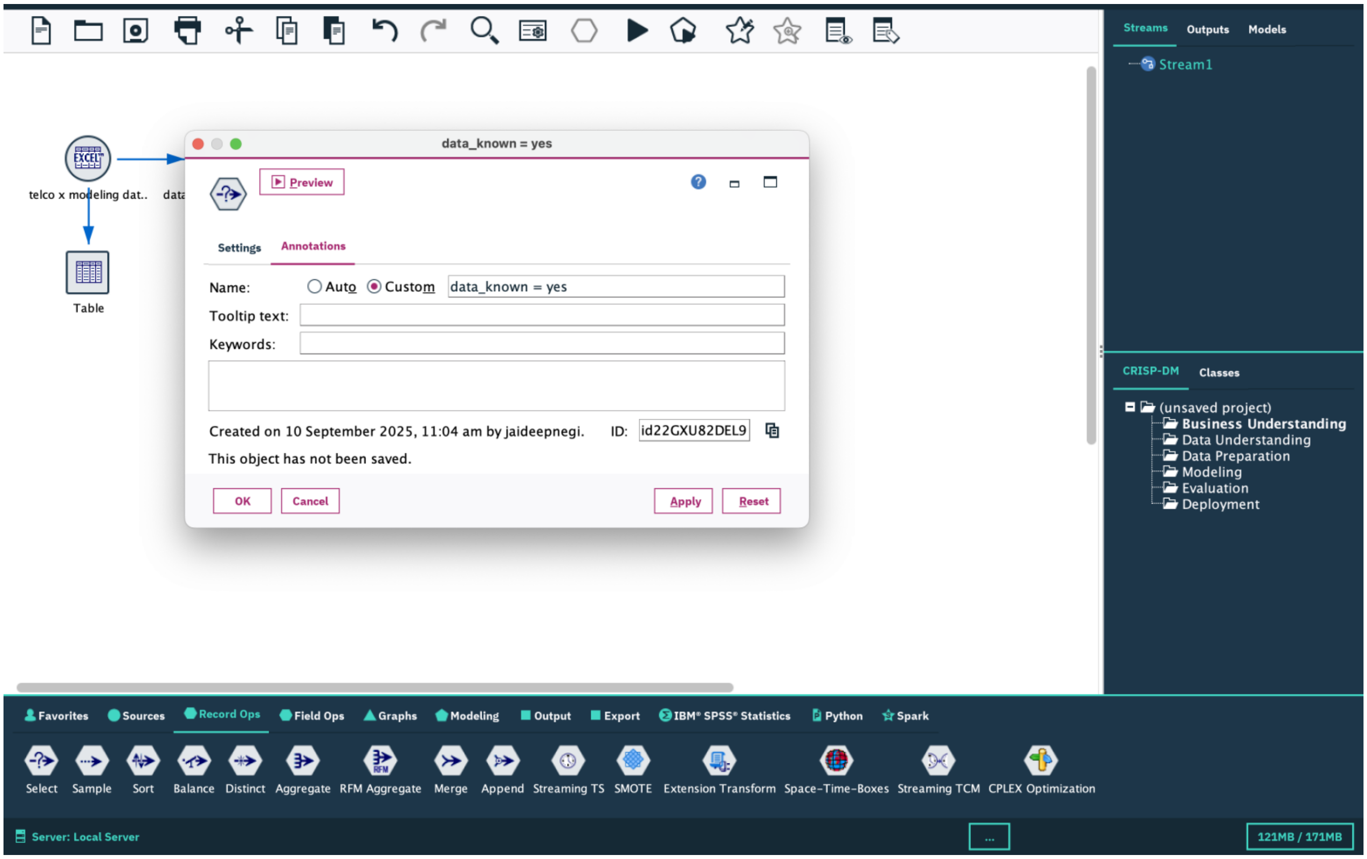
Required Tool: IBM SPSS Modeler.

Working: This workflow demonstrates the use of a pre-trained C5.0 model to score new data for customer churn prediction. It specifically identifies a target group of "customers at risk" by filtering for those with a high-confidence prediction of churning, and then outputs a clean list for targeted actions.

Step 1: Data Import: The process starts by importing an Excel file, telco x modeling dat.xlsx, into the workspace. A Table node is connected directly to the data source to provide a first look at the raw data.



Step 2: Filtering Data with the Select Node: A Select node is then added to the stream to filter the data. The condition data_known = 'yes' is applied, which keeps only the records where the value of the data_known field is 'yes'. This step is crucial for preparing a clean dataset for a specific analysis.



data_known = yes

Settings | Annotations

Name: ○ Auto ● Custom data_known = yes

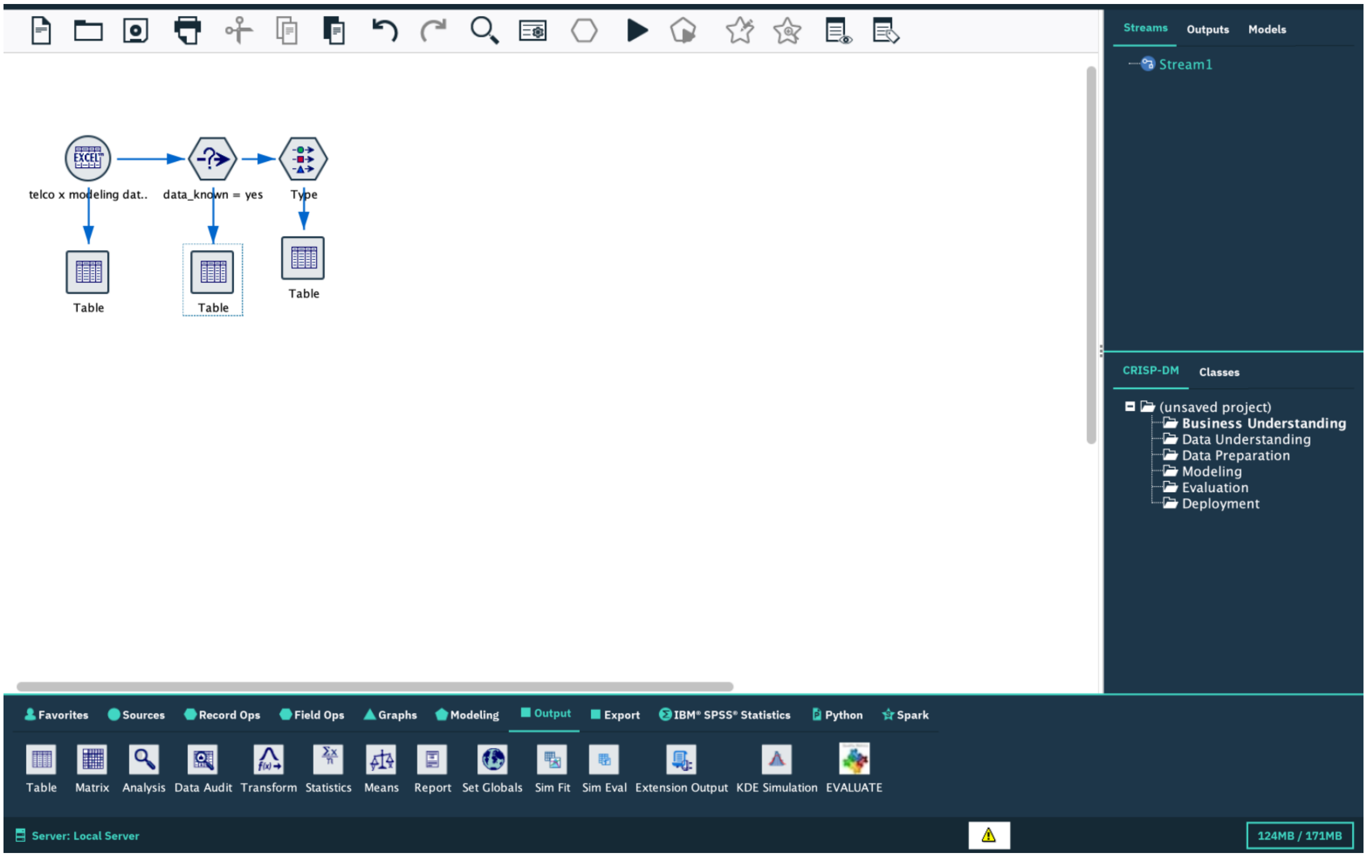
Tooltip text:

Keywords:

Created on 10 September 2025, 11:04 am by jaideepnegi. ID: id22GXU82DEL9

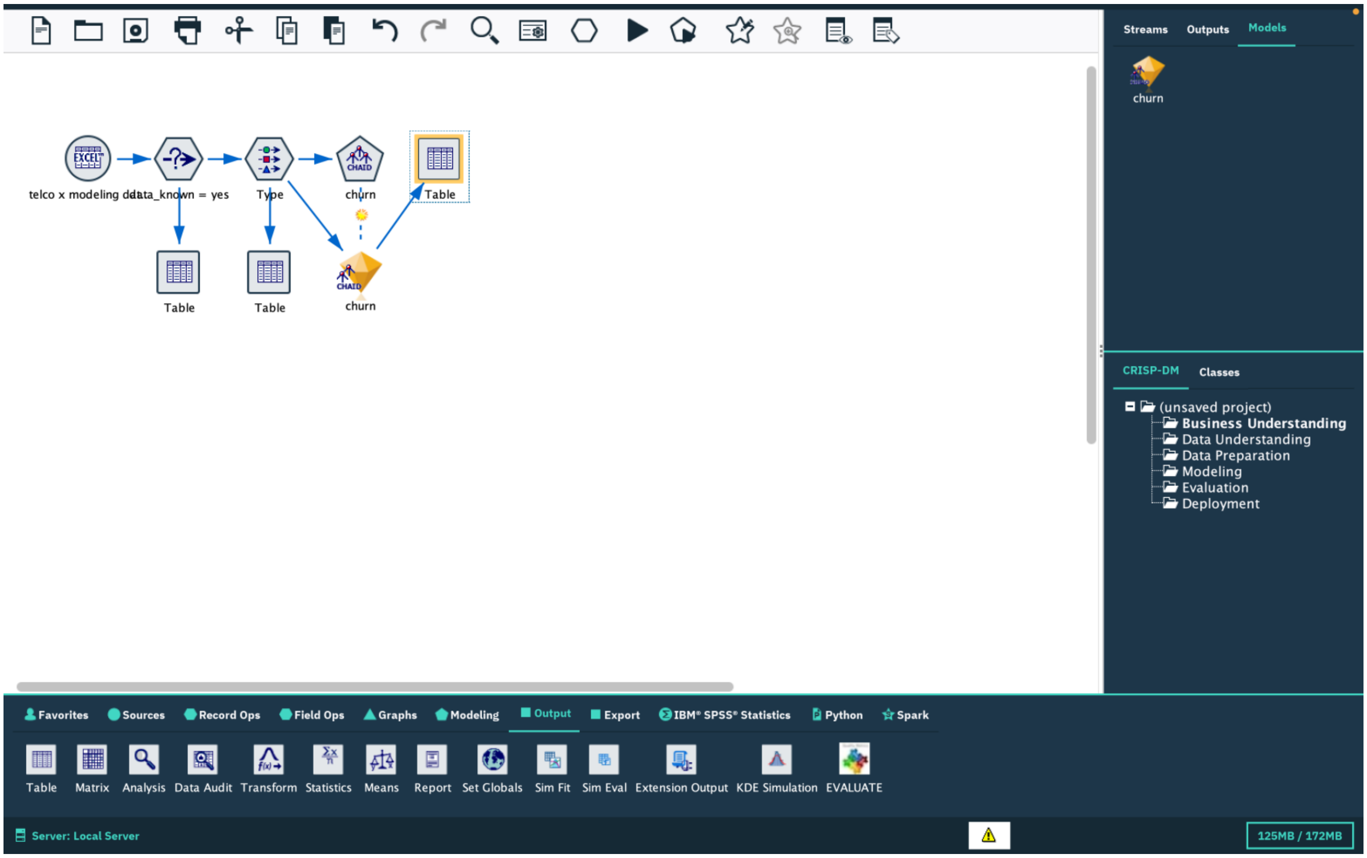
This object has not been saved.

Step 3: Defining Field Properties with the Type Node : Following the filter, a Type node is used to define the properties of each field in the filtered dataset. It automatically reads the data and sets the measurement level (e.g., Flag for categorical yes/no data, Continuous for numerical values) and the role (e.g., Input for predictor variables and Target for the variable to be predicted, in this case, churn).

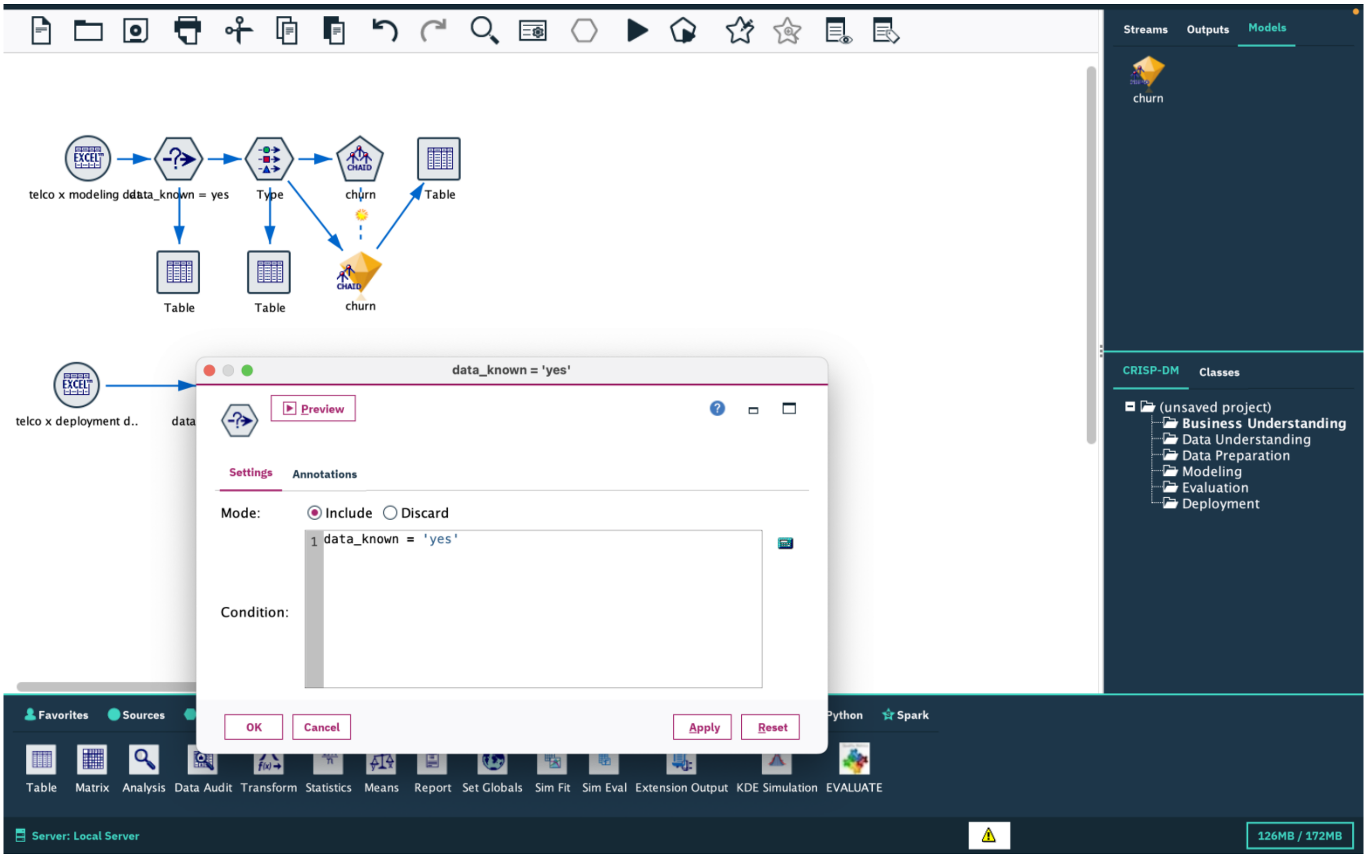


Step 4: Modeling with the C5.0 Node : A C5.0 node is added to the stream. This node is a classification model that builds decision trees to predict the `churn` variable, which has been designated as the target. The C5.0 algorithm is specifically designed for this type of supervised learning task.

A Table node is connected to the output of the C5.0 node to view the results of the model.



Step 6: Data Import and Preparation : The process begins by importing the telco x deployment d... dataset. A Select node is then used to filter the records, keeping only those where the data_known field is 'yes', preparing a clean dataset for scoring.



data_known = 'yes'

Preview

Settings | Annotations

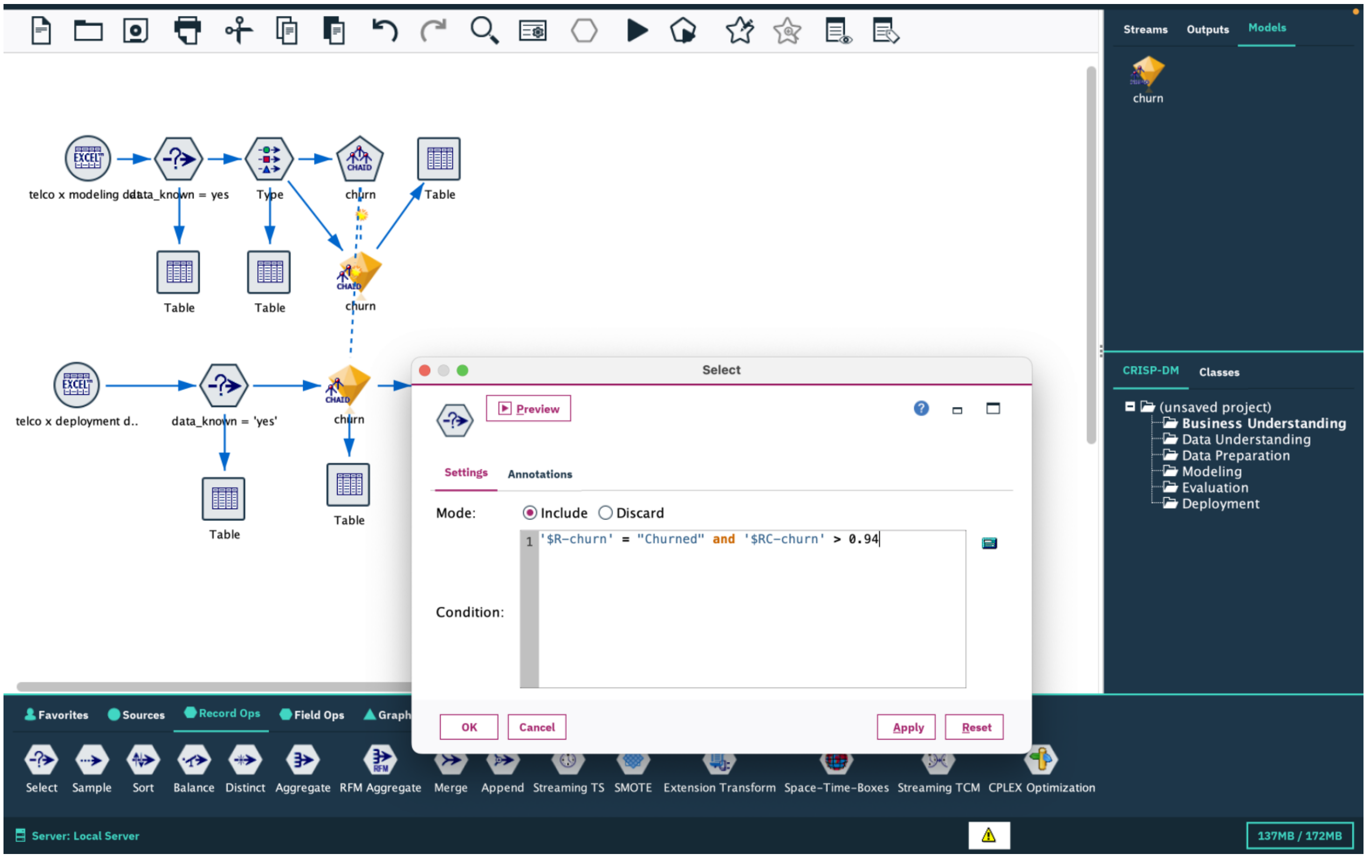
Mode: ● Include ○ Discard

Condition:

```
data_known = 'yes'
```

OK | Cancel | Apply | Reset

Step 7: Scoring Data with a Pre-built Model : The prepared data is fed into a C5.0 model, which was trained in a previous workflow. This model scores the new data to predict the likelihood of "churn" for each customer. The output of this node is a table showing the scored results, with a new field, `$SC-churn`, indicating the prediction.



Select

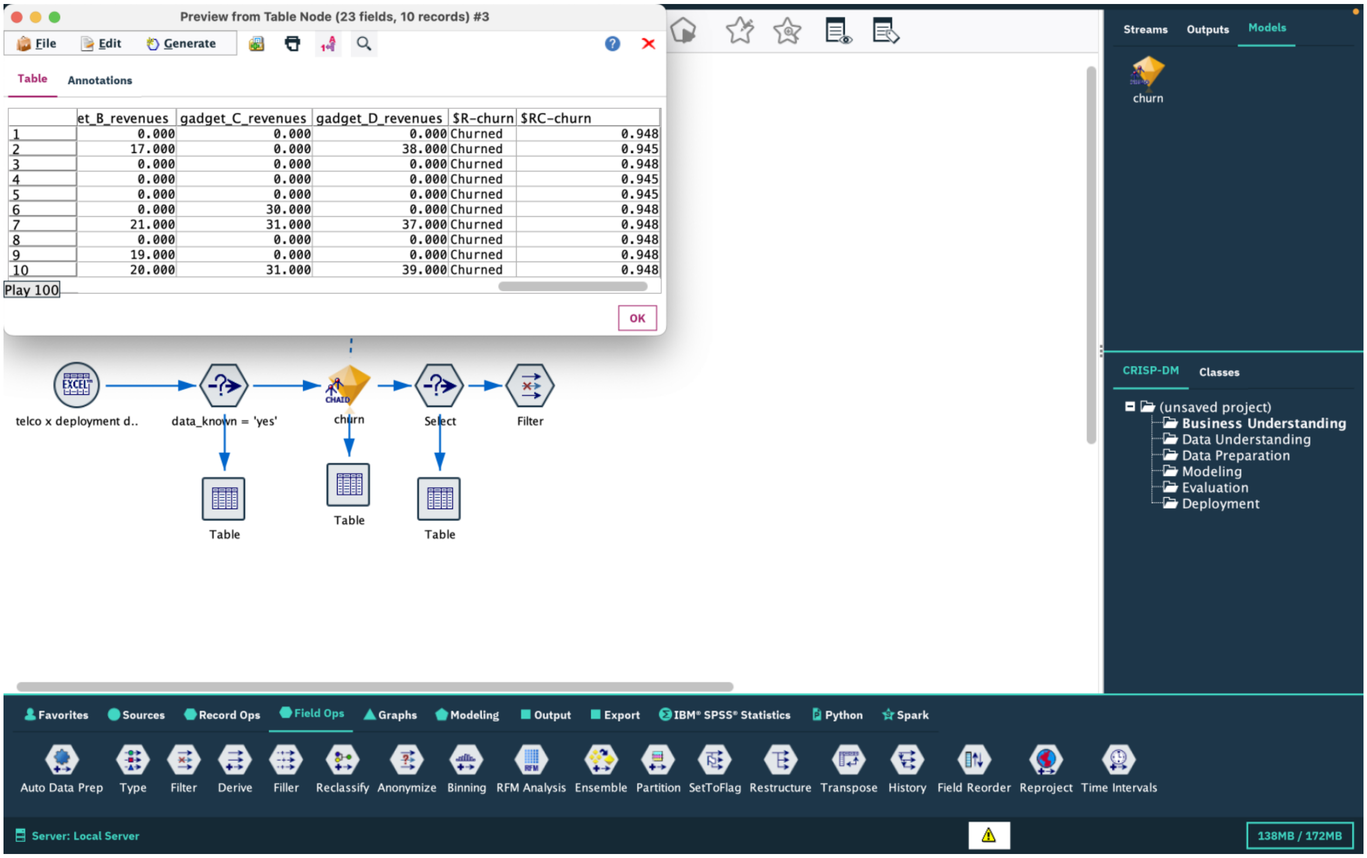
Settings | Annotations

Mode: ● Include ○ Discard

Condition:

```
'$R-churn' = "Churned" and '$RC-churn' > 0.94
```

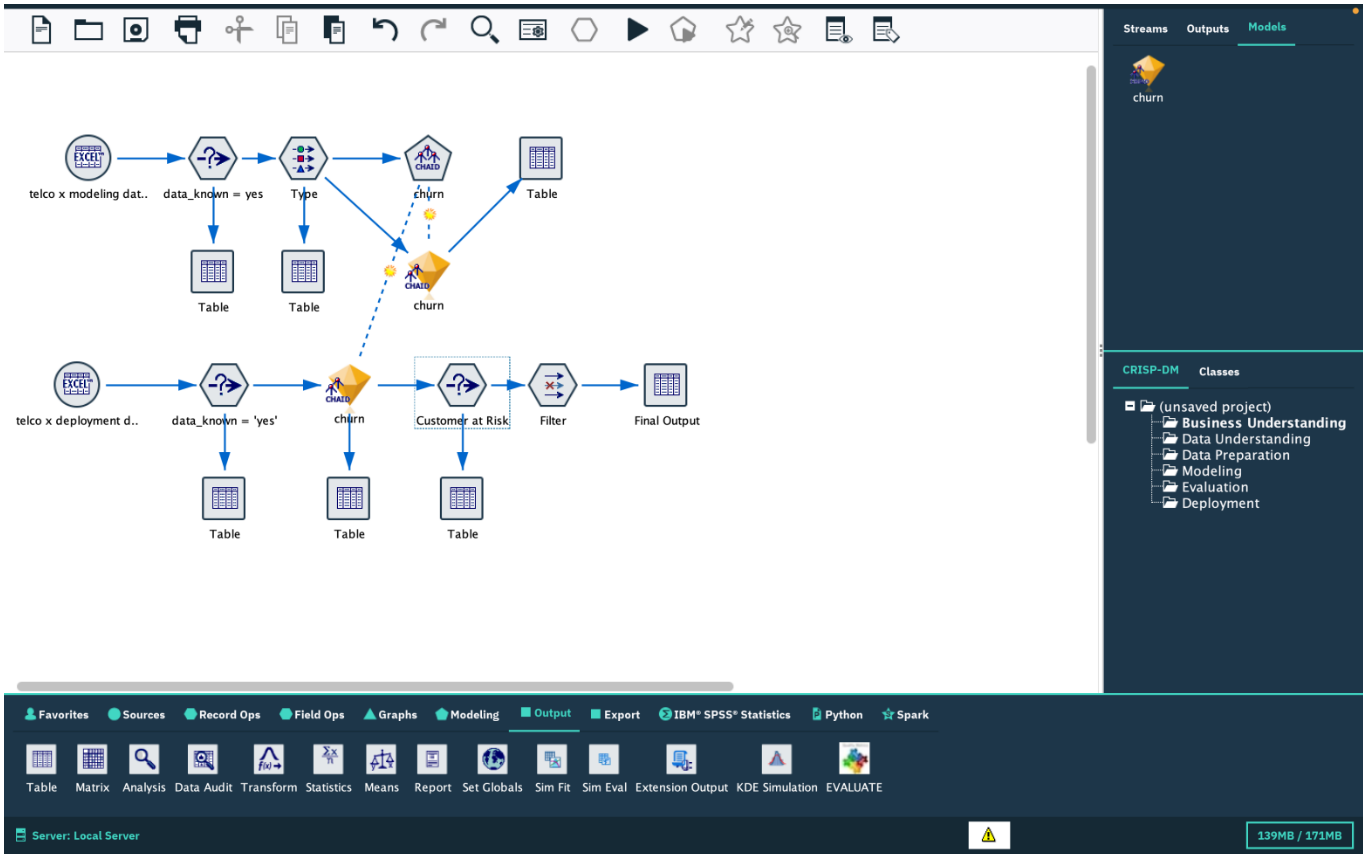
Step 8: Identifying High-Risk Customers : Another Select node is applied to the scored data. This node filters the results to identify "customers at risk," specifically those predicted to churn with a high degree of confidence. The condition for this is `$SC-churn` (the predicted churn) being equal to 'Churned' and the confidence score (`$SC-churn`) being greater than 0.94.



Preview from Table Node (23 fields, 10 records) #3

	et_B_revenues	gadget_C_revenues	gadget_D_revenues	$R-churn	$RC-churn
1	0.000	0.000	0.000	Churned	0.948
2	17.000	0.000	38.000	Churned	0.945
3	0.000	0.000	0.000	Churned	0.948
4	0.000	0.000	0.000	Churned	0.945
5	0.000	0.000	0.000	Churned	0.945
6	0.000	30.000	0.000	Churned	0.948
7	21.000	31.000	37.000	Churned	0.948
8	0.000	0.000	0.000	Churned	0.948
9	19.000	0.000	0.000	Churned	0.948
10	20.000	31.000	39.000	Churned	0.948

Step 9: Final Output : Finally, a Table and a Filter node are used to generate the final output. The Filter node removes unnecessary fields, such as the confidence score, leaving a clean table of high-risk customers, which is the final product of this deployment stream.



Practical: 4

Definition: You work as a data miner for a telecommunications firm. It is your job, in order to merge the datasets later, to remove duplicate records in the customer dataset and to transform a transactional dataset into a dataset that has one record per customer.

Outcomes/Learning: We learned how to set up a basic data analysis stream in IBM SPSS Modeler by importing multiple raw data files, defining their field properties with the Type node, and verifying the contents using the Table node. The outcome is a prepared dataset ready for further analysis.

Required Tool: IBM SPSS Modeler.

Working: In this process shows how to import three separate datasets (customer, call, and products) and prepare them for analysis by using the Type node to correctly identify the data type of each variable. The Table node then provides a view of the cleaned data, confirming the preparation steps were successful.

Step 1: Importing Data: We start by importing multiple .sav files, including telco x customer dat.sav, telco x call data q1.sav, and telco x products.dat, into IBM SPSS Modeler.

•   •   •

Continue……………………………………….

Note: heading size should be times new roman, 12 , bold
Normal text size should be times new roman, 11 , normal