SPSS PRACTICAL

Name:	Ganesh Agrahari
Class:	BCA DS 33
Roll No.:	1230258176
Submitted To:	Mr. Robin Tyagi

INDEX

SPSS Practical 6

This practical includes:

Practical: 6

Definition:

You work for a telecommunications firm and you need to cleanse and enrich a dataset in order to build models later.

Outcomes/Learning:

- Learned how to use the **Derive Node** to create new calculated fields such as billing details, segmentation, and discounts.
- Understood the use of the **Reclassify Node** to standardize and group categorical values for better analysis.
- Gained practical experience in cleansing and enriching raw datasets to prepare them for predictive modelling.

Required Tool:

IBM SPSS MODELER

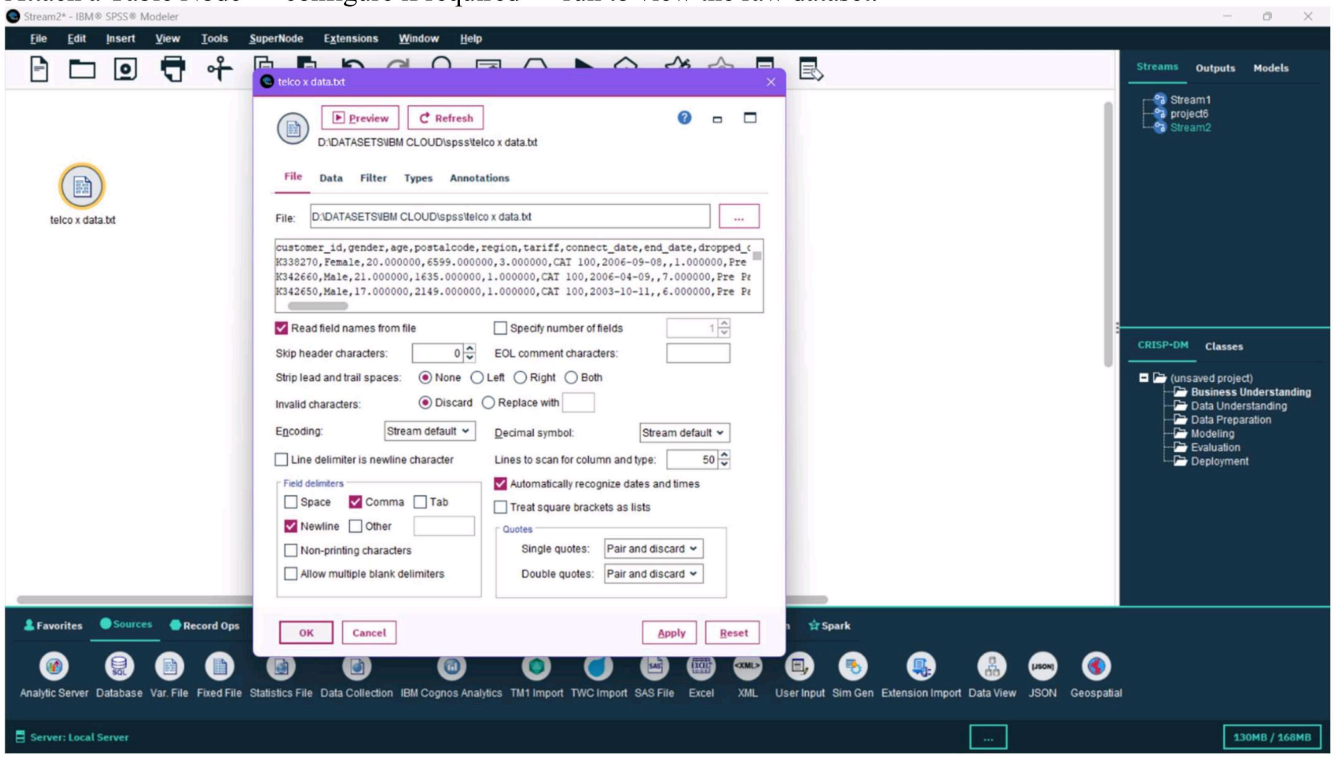
Working:

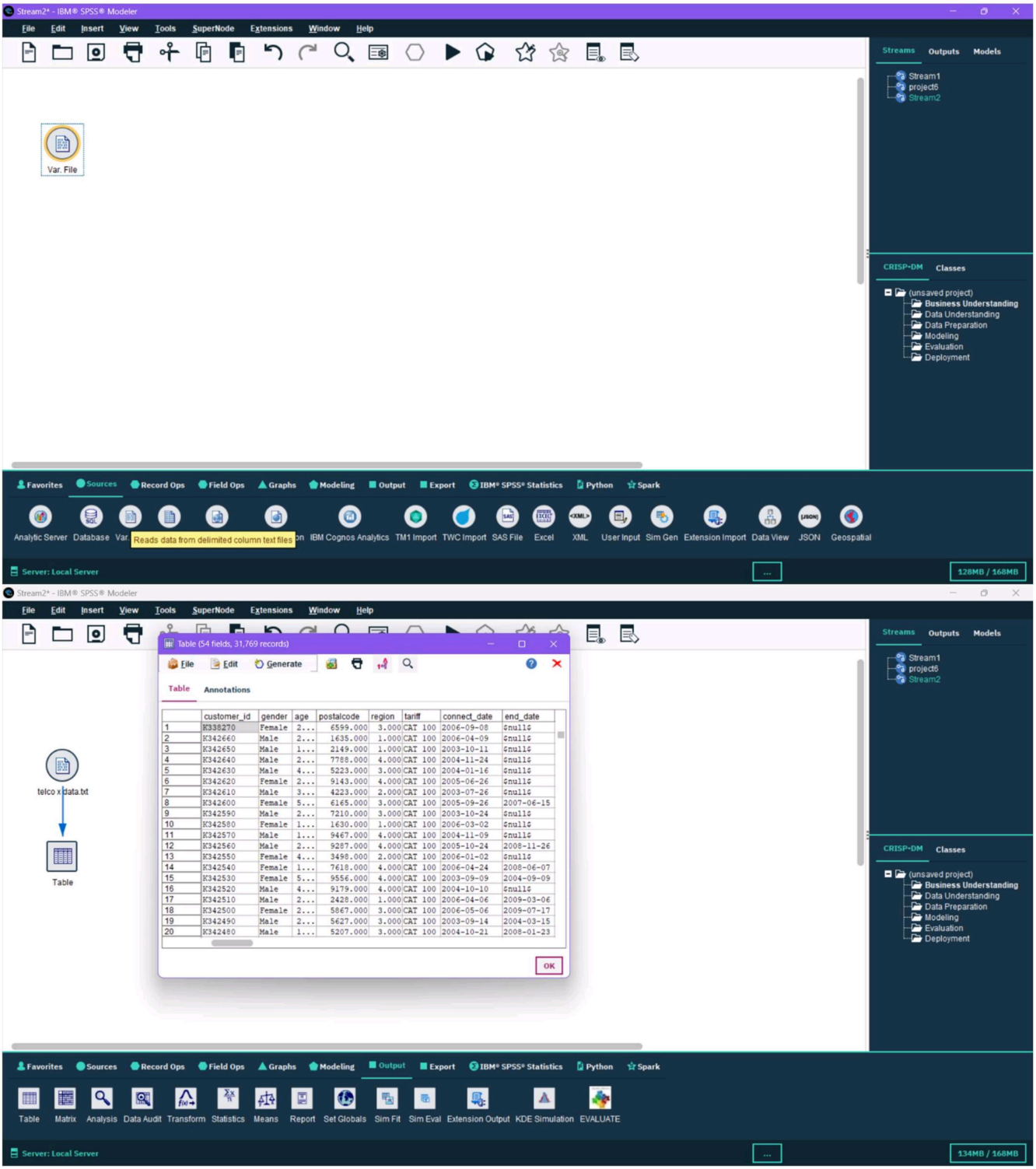
In this practical, the goal is to cleanse and enrich the dataset (*telco x data.txt*) for further modelling. Fields are standardized, billing details are calculated, customers are segmented, and discounts are derived. Finally, categorical values are reclassified into consistent groups.
The main nodes used are:

- **Type Node** – to define field roles and measurement levels.
- **Derive Nodes** – to create new fields such as billing, segmentation, and discounts.
- **Reclassify Node** – to restructure categorical values.
- **Table Nodes** – to validate results at each step.

Step 1: Importing the Dataset

Open IBM SPSS Modeler → create a new stream.
Add a Var. File Node and import *telco x data.txt*.
Attach a Table Node → configure if required → run to view the raw dataset.

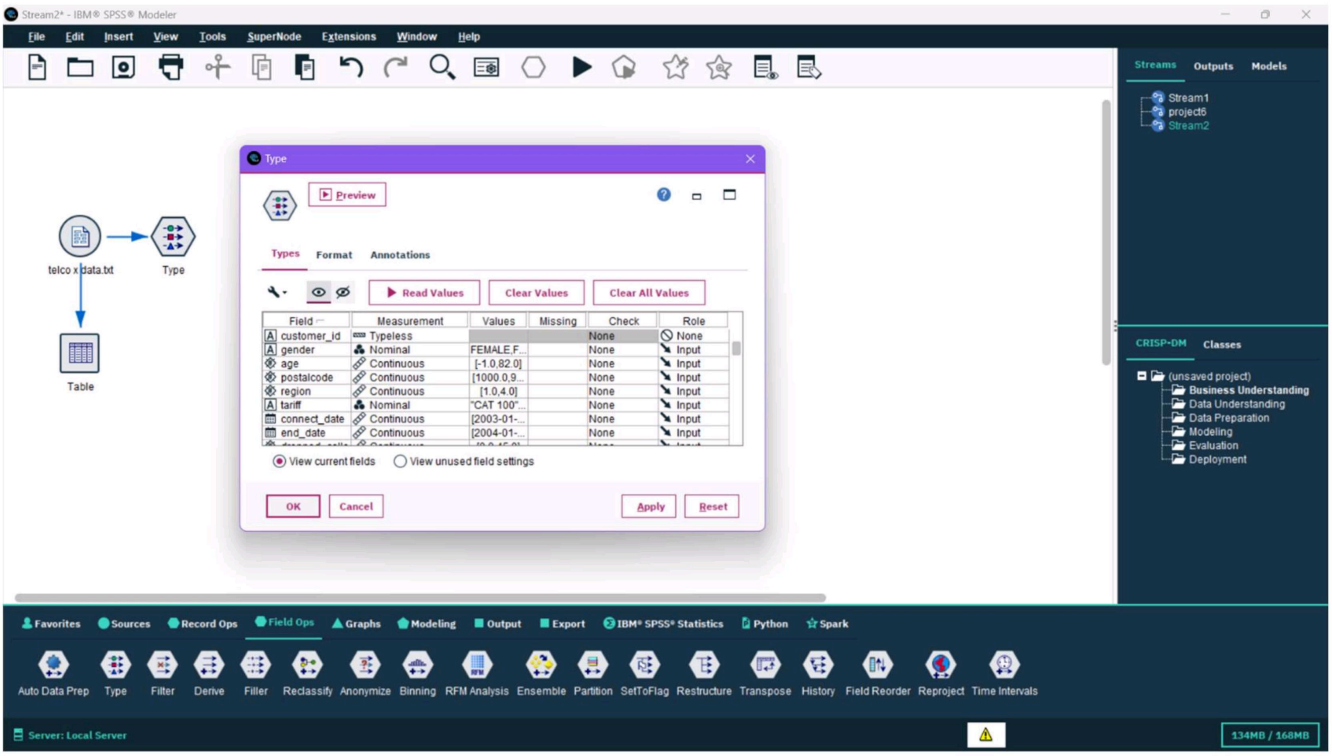




Step 2: Defining Field Properties

Connect a Type Node to the Var. File Node.

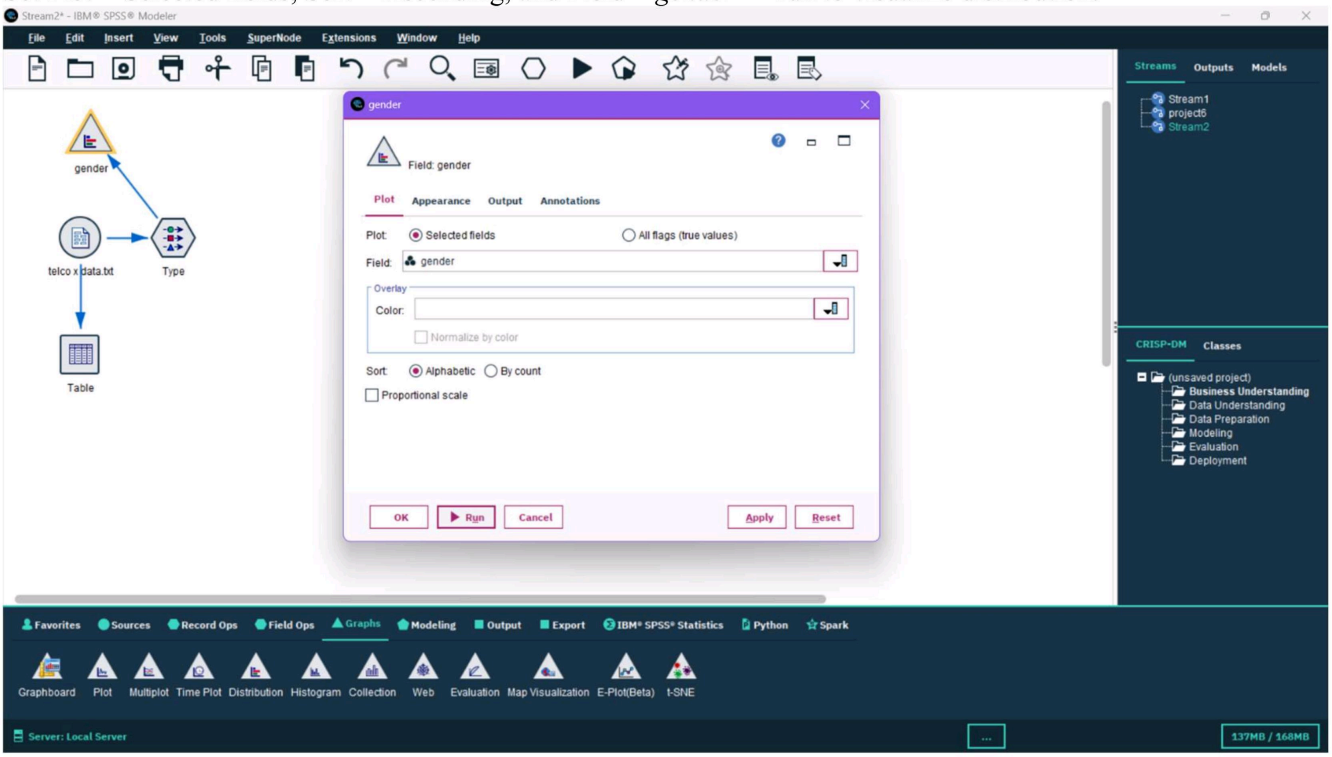
Open it → click Read Values → observe updated measurement types (Nominal, Ordinal, etc.).

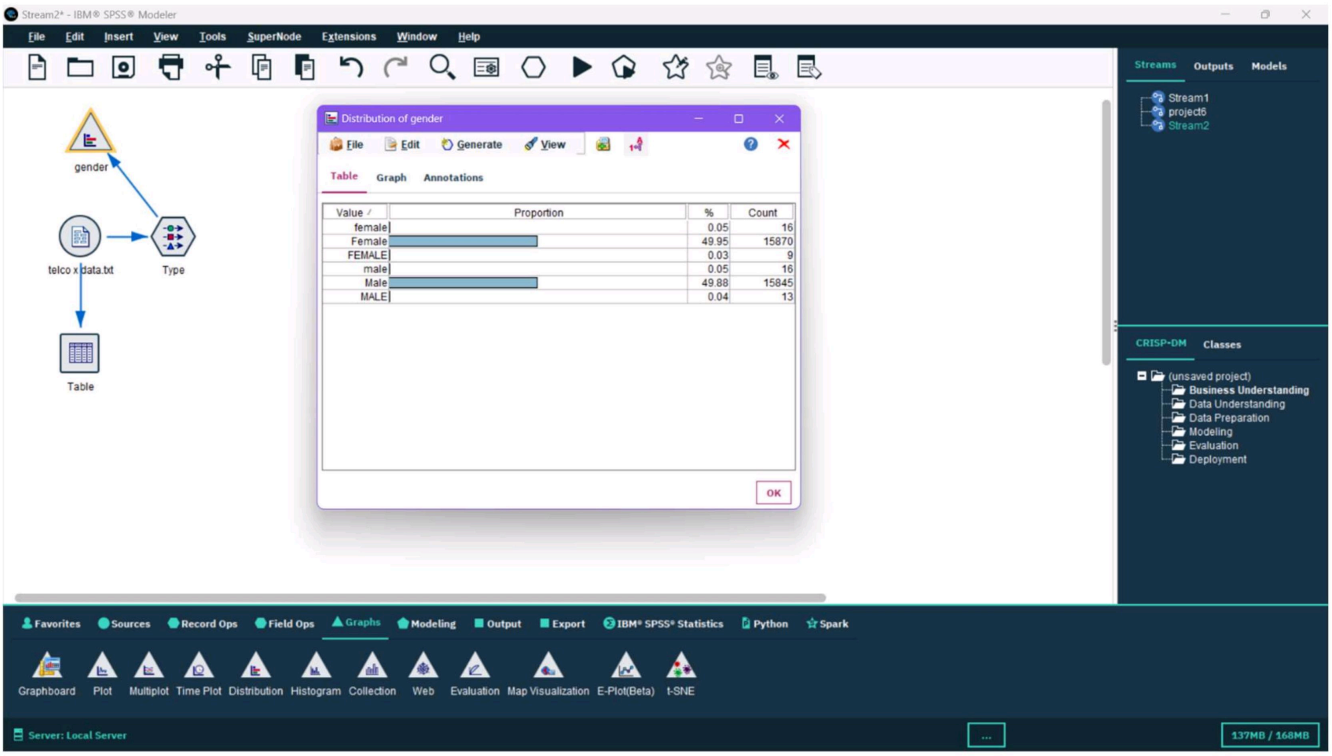


Step 3: Exploring with Distribution Graph

Attach a Distribution Node to the Type Node.

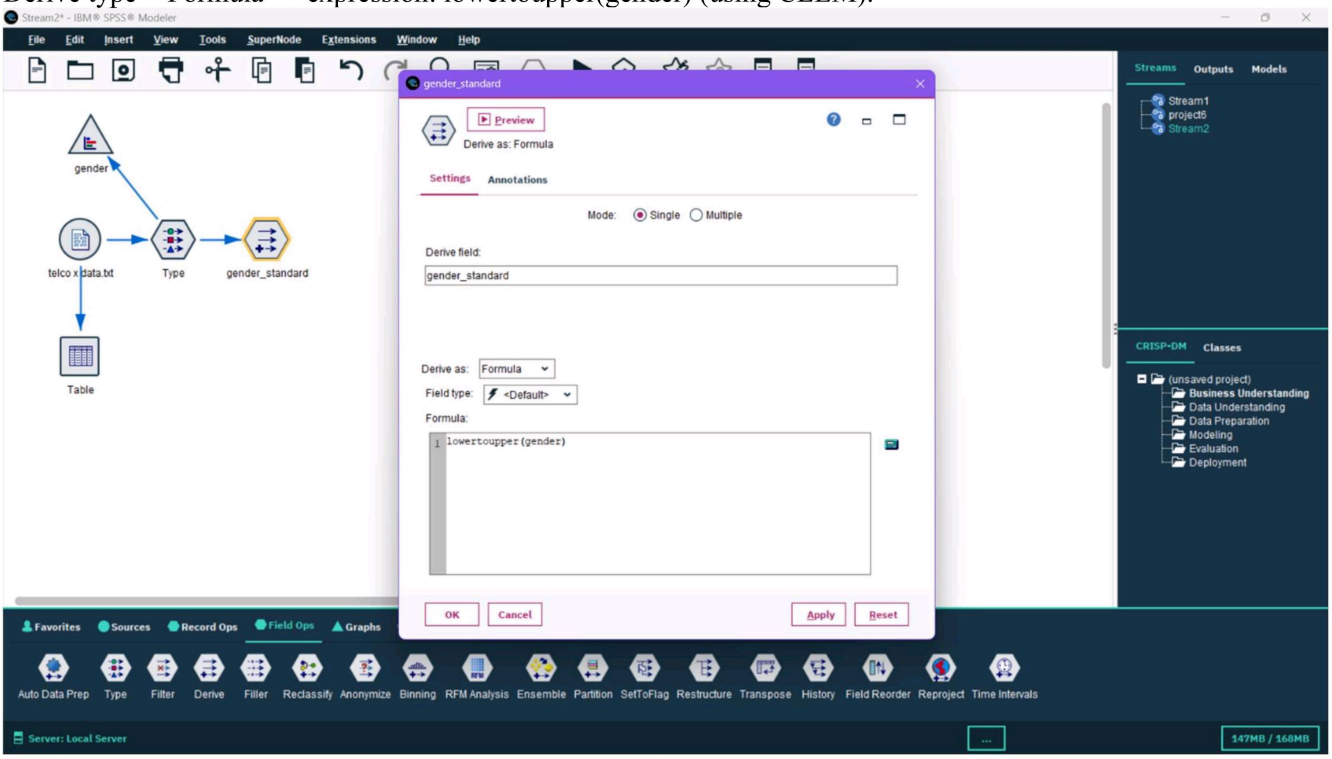
Set Plot = Selected fields, Sort = Ascending, and Field = *gender* → run to visualize distribution.





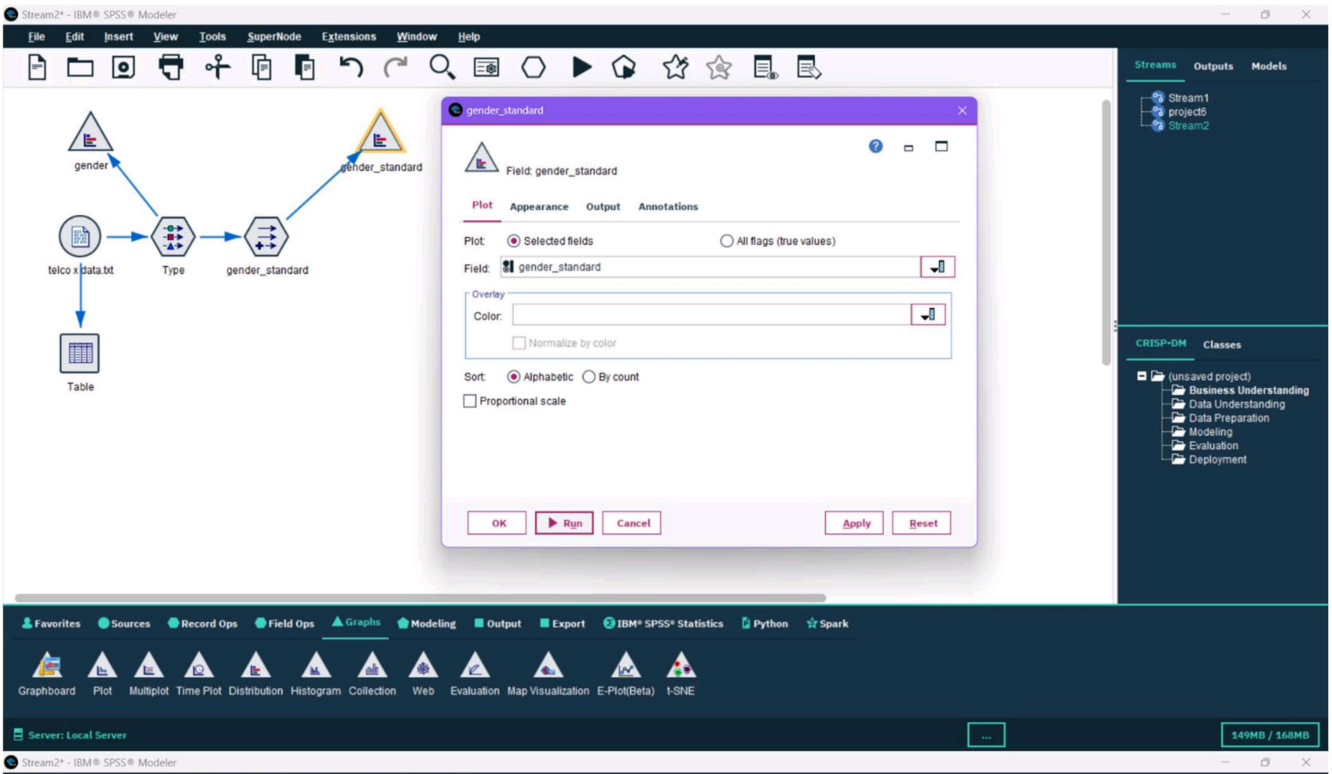
Step 4: Standardizing Gender Field

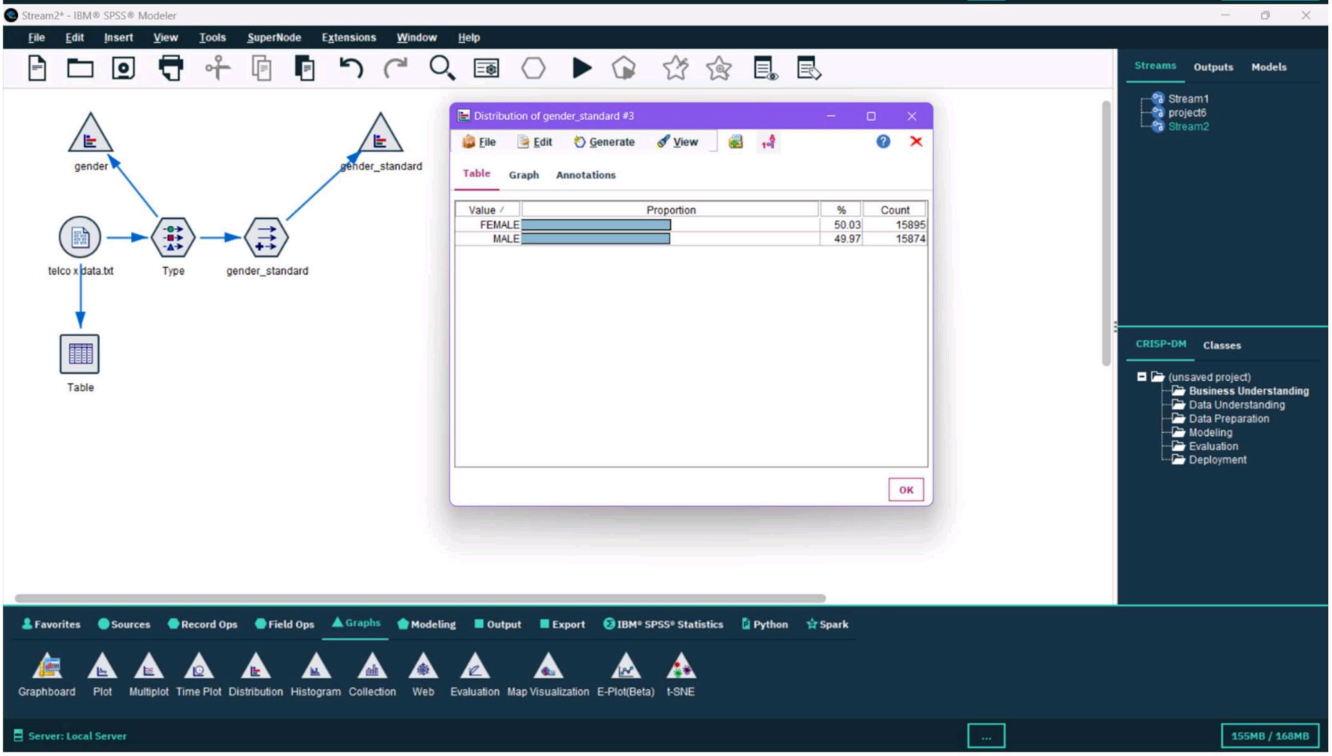
Connect a Derive Node to the Type Node.
Rename field = *gender_standard*.
Derive type = Formula → expression: lowertoupper(gender) (using CLEM).



Step 5: Visualizing Standardized Gender

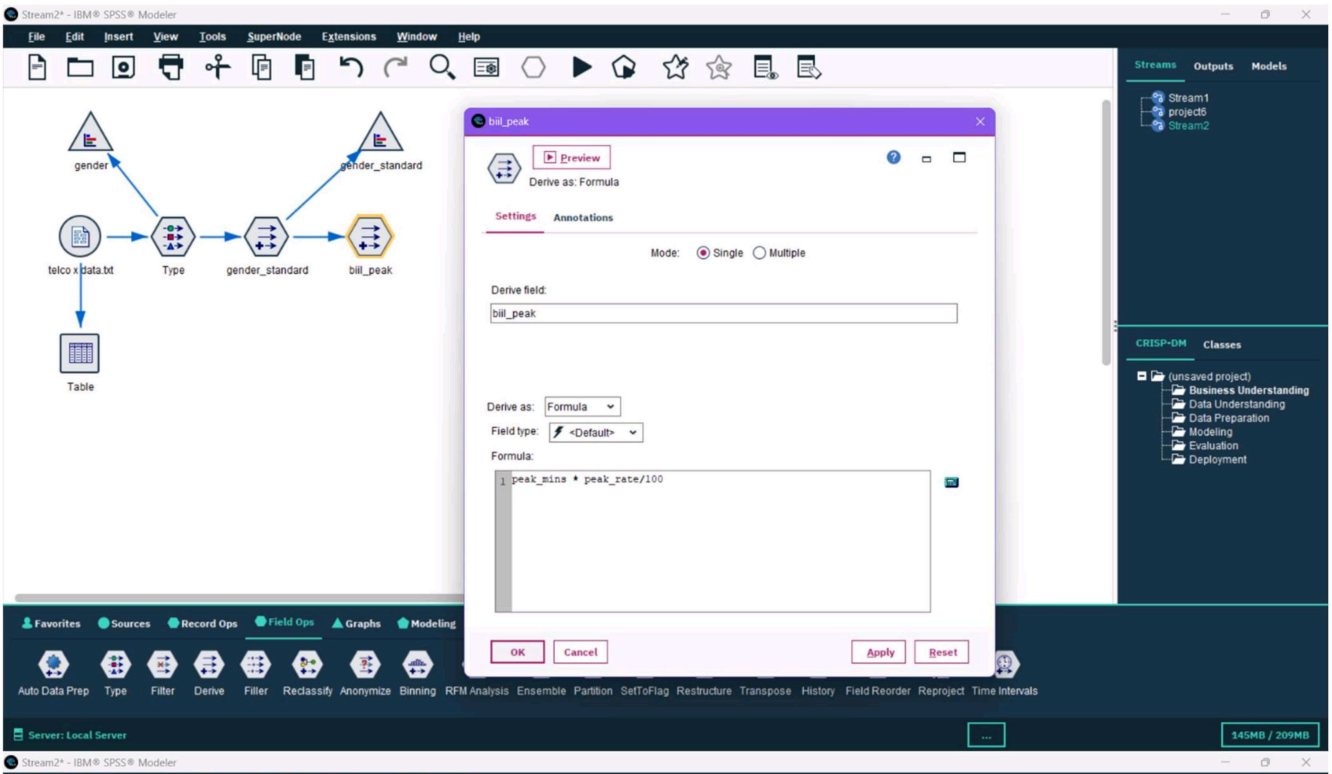
Attach a Distribution Node to the *gender_standard* Derive Node.
Set field = *gender* → run to confirm corrected values.

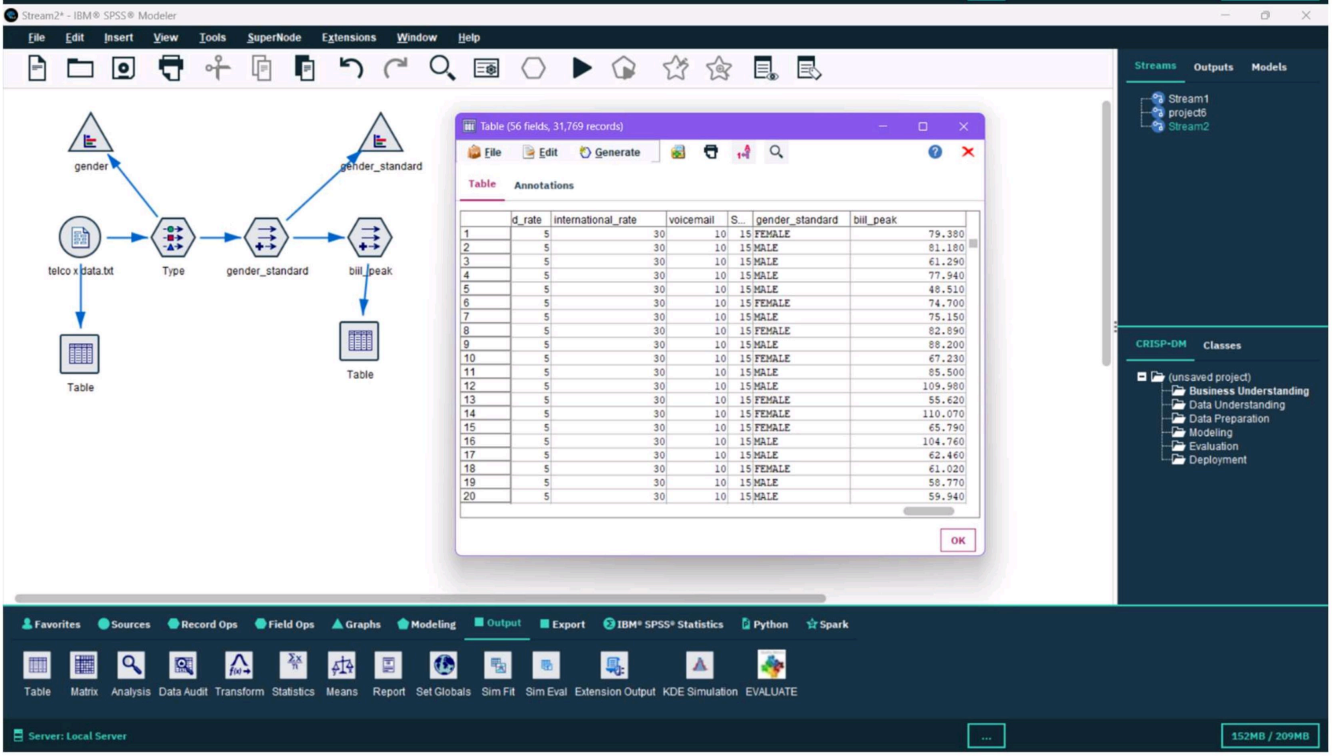




Step 6: Calculating Peak Bill

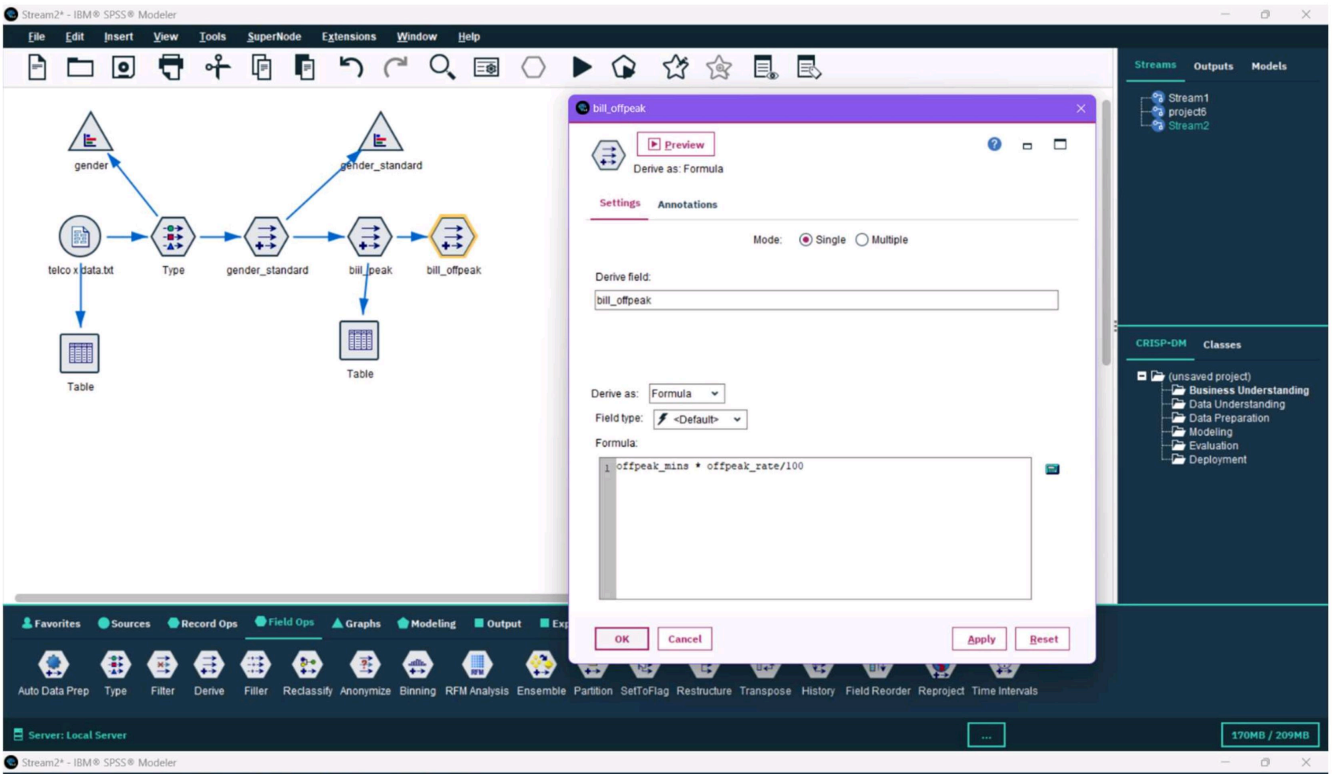
Add a new Derive Node after *gender_standard*.
Rename field = *bill_peak*.
Formula = peak_mins * peak_rate / 100.
Attach a Table Node → run to view results.

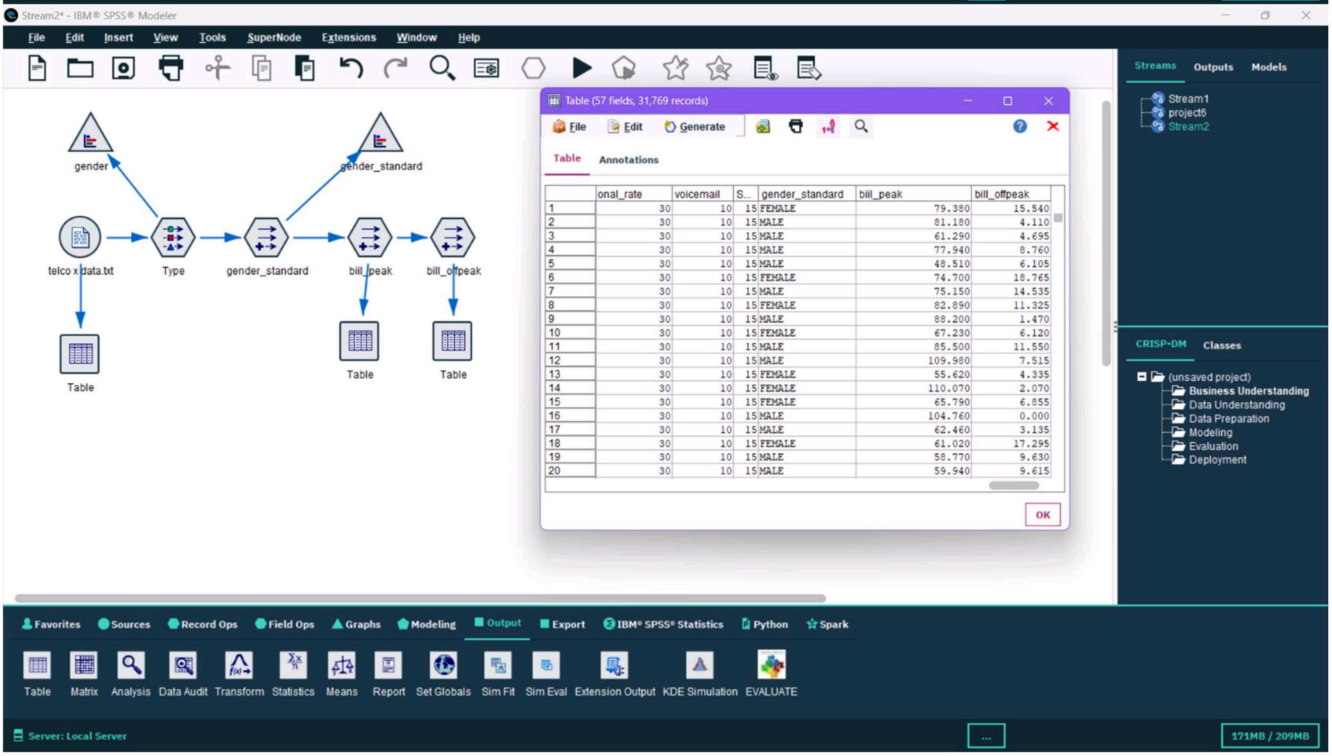




Step 7: Calculating Off-Peak Bill

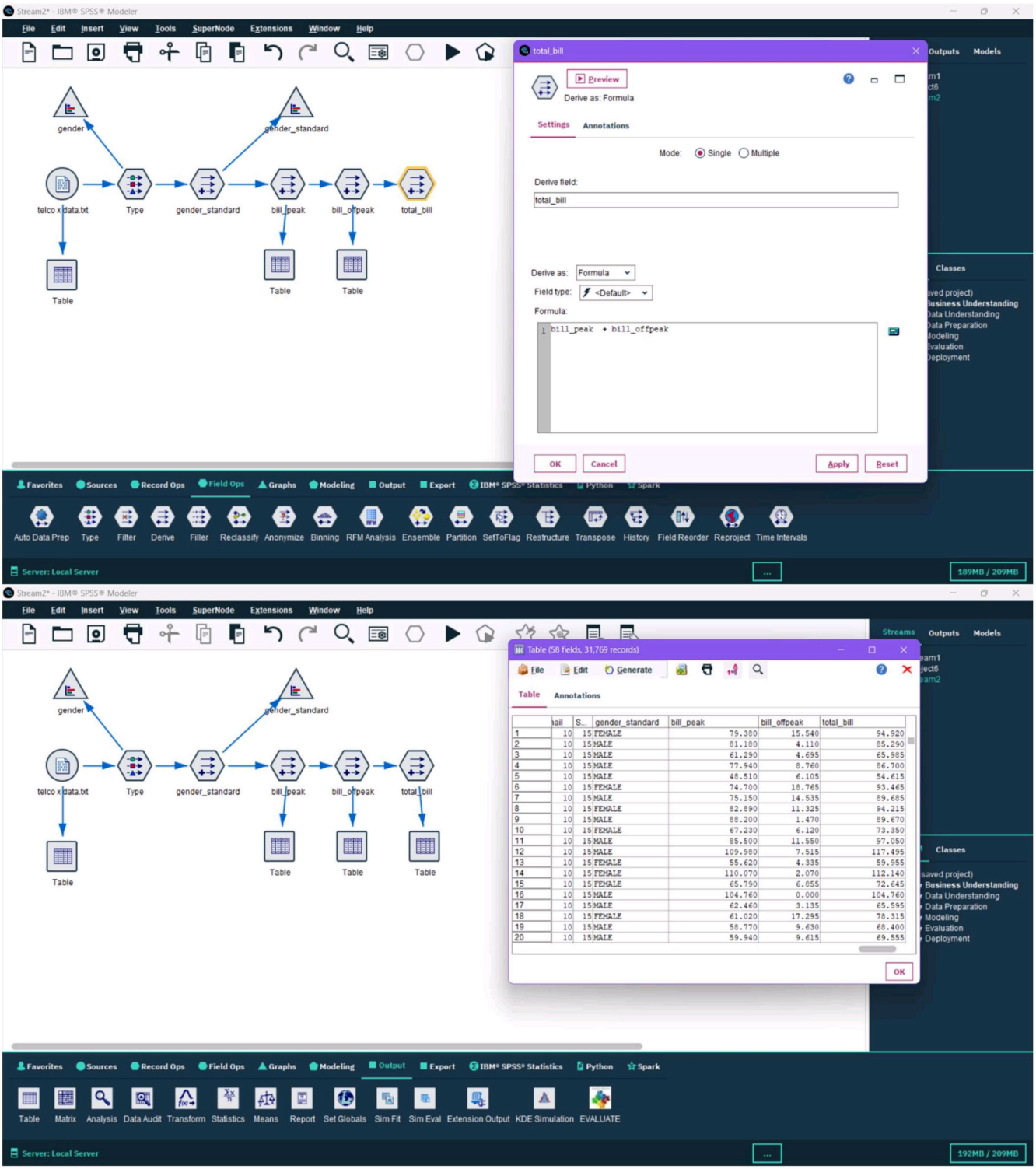
Add another Derive Node after *bill_peak*.
Rename field = *offbill_peak*.
Formula = offpeak_mins * offpeak_rate / 100.
Attach a Table Node → run to confirm results.





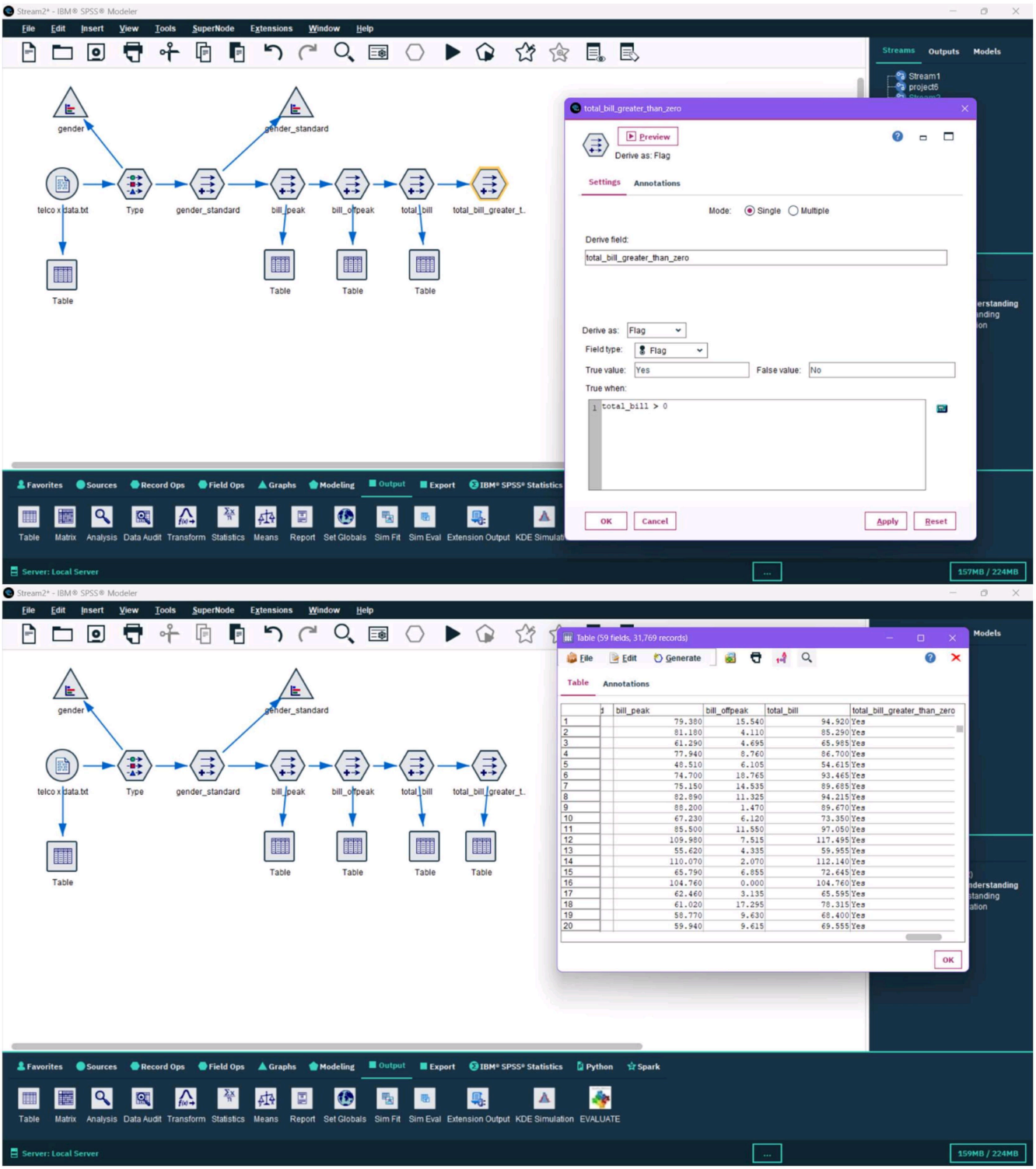
Step 8: Calculating Total Bill

Connect a Derive Node to *offbill_peak*.
Rename field = *total_bill*.
Formula = bill_peak + offbill_peak.
Attach a Table Node → run to display total bill values.



Step 9: Flagging Valid Bills

Insert another Derive Node after *total_bill*.
Rename field = *total_bill_greater_than_zero*.
Type = Flag → condition: bill_total > 0.
Attach a Table Node → run to validate flags.



Step 10: Customer Segmentation

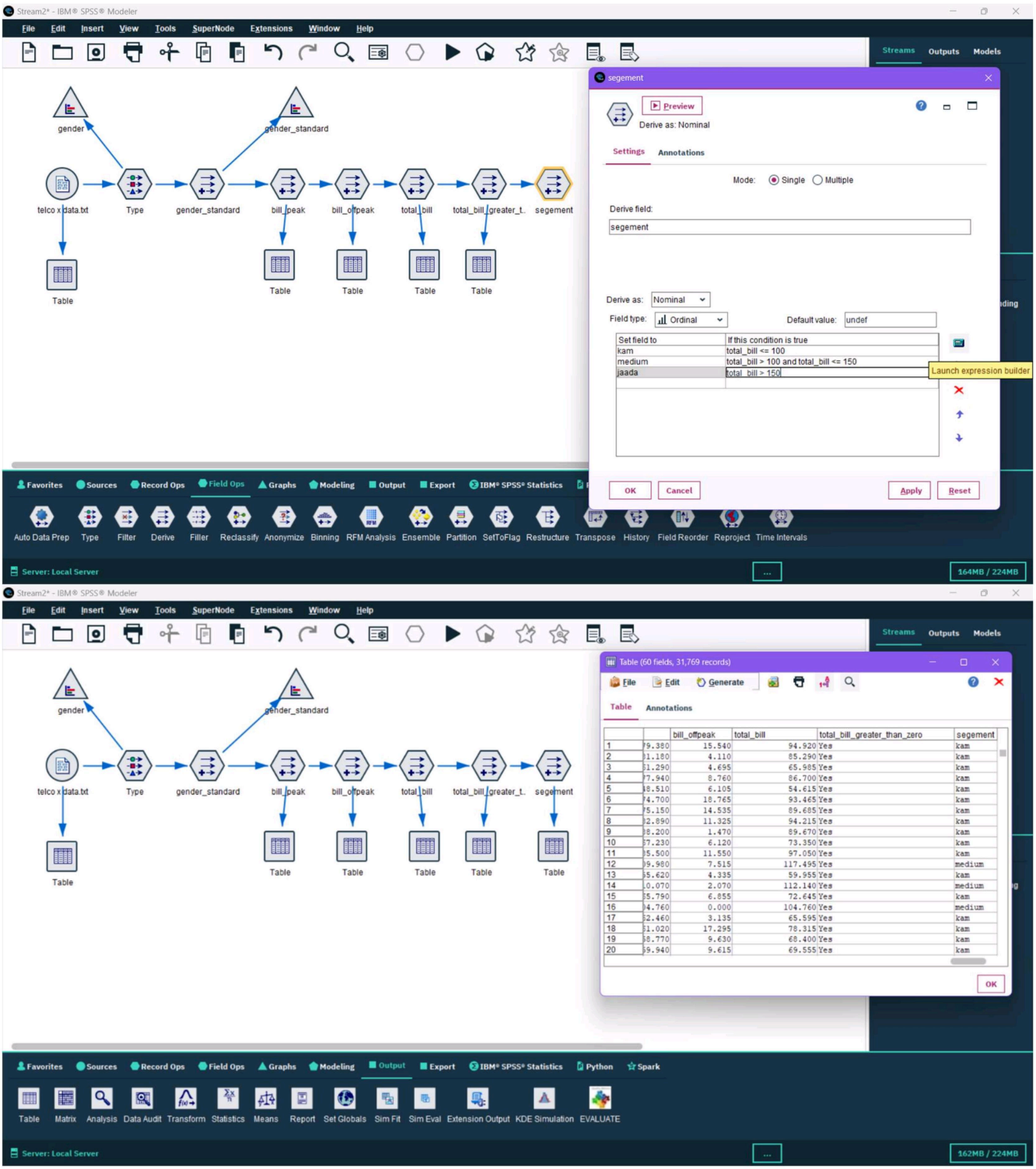
Add a Derive Node after the flag.

Rename field = *segment*.

Type = Nominal and Ordinal → expression:

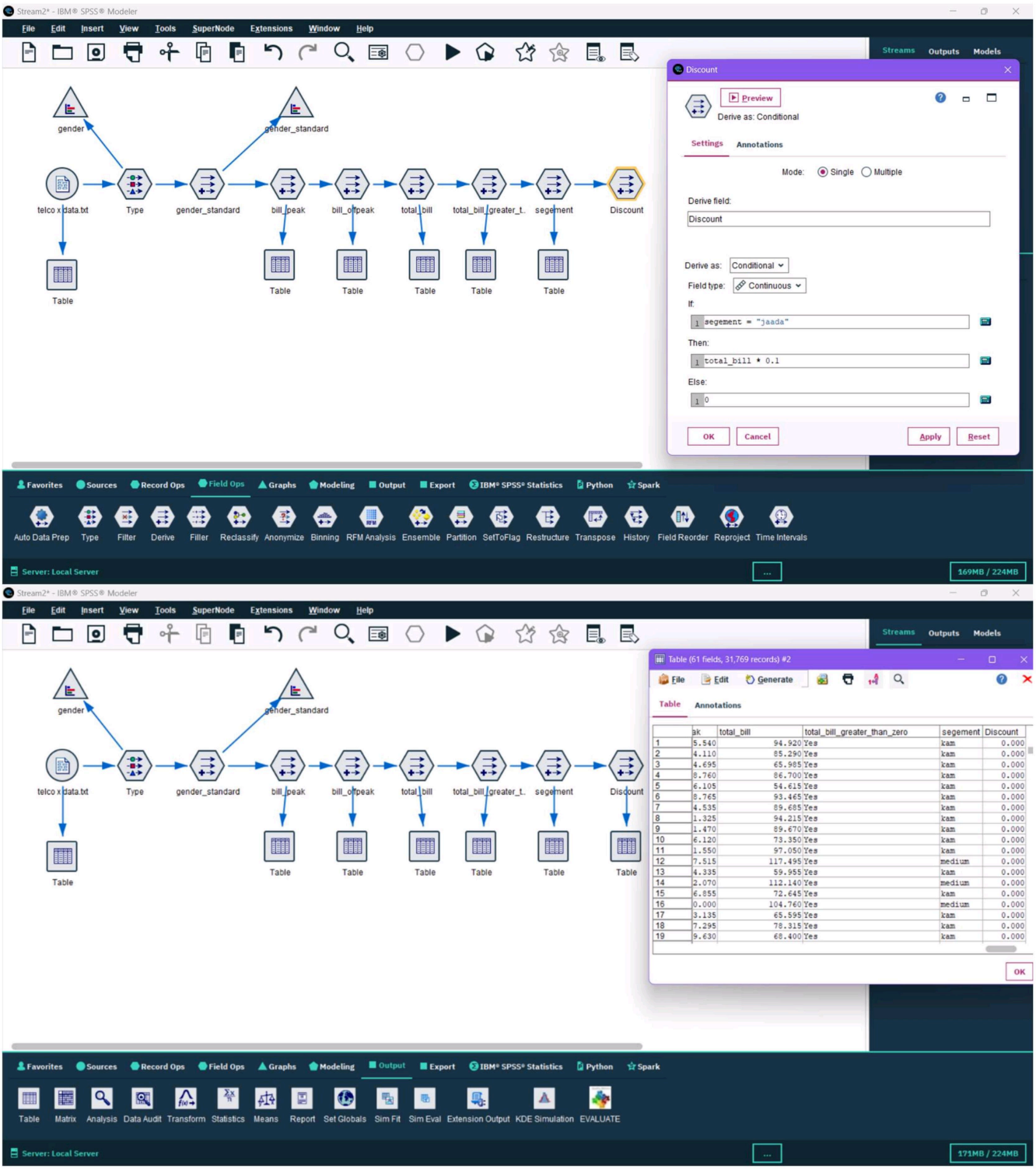
- *kam* if bill_total <= 100
- *medium* if bill_total > 100 and <= 150
- *jaada* if bill_total > 150.

Attach a Table Node → run to confirm segment allocation.



Step 11: Applying Discount Rules

Connect another Derive Node to *segment*.
Rename field = *discount*.
Type = Conditional Continuous → expression:
If segment = "jaada" Then bill_total * 0.1 Else 0.
Attach a Table Node → run to calculate discounts.

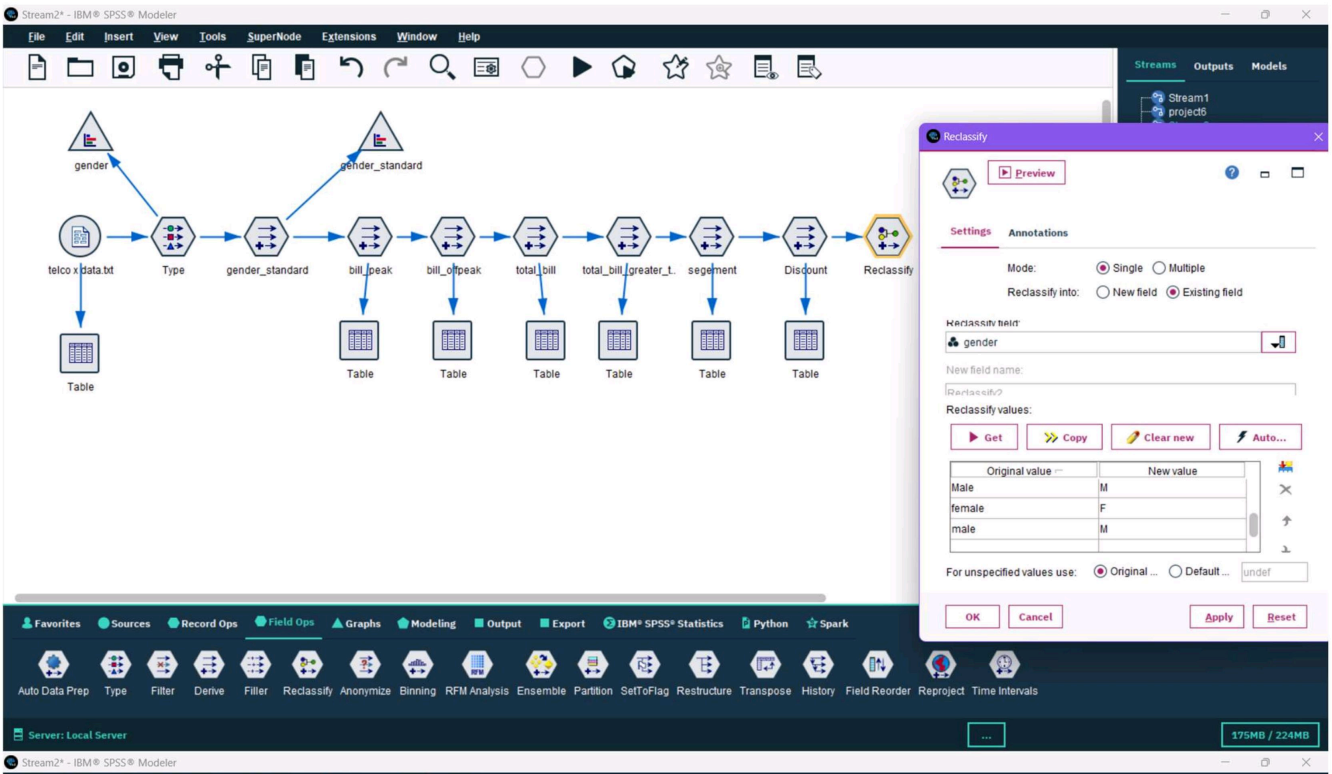


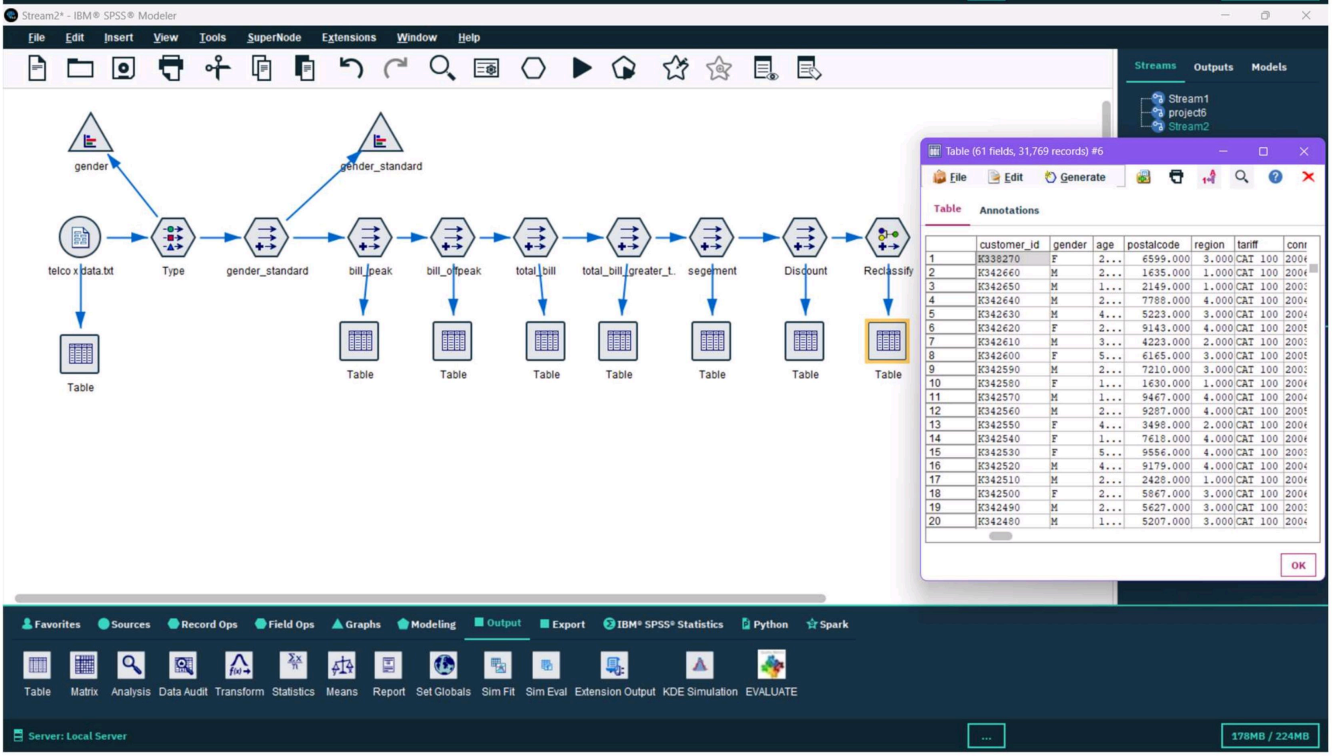
Step 12: Reclassifying Gender Field

Finally, add a Reclassify Node after *segment*.

Reclassify field = *gender* → map values into standardized categories (M, F).

Attach a Table Node → run to generate the final cleaned dataset.


Stream2* - IBM® SPSS® Modeler
File Edit Insert View Tools SuperNode Extensions Window Help
Streams Outputs Models
Stream1
project6
gender
gender_standard
telco x data.txt
Type
gender_standard
bill_peak
bill_offpeak
total_bill
total_bill_greater_t..
segement
Discount
Reclassify
Table
Reclassify
Preview
Settings Annotations
Mode: Single Multiple
Reclassify into: New field Existing field
Reclassify field:
gender
New field name:
Reclassify values:
Get Copy Clear new Auto...
Original value
New value
Male M
female F
male M
For unspecified values use: Original ... Default ... undef
OK Cancel Apply Reset
Favorites Sources Record Ops Field Ops Graphs Modeling Output Export IBM® SPSS® Statistics Python Spark
Auto Data Prep Type Filter Derive Filler Reclassify Anonymize Binning RFM Analysis Ensemble Partition SetToFlag Restructure Transpose History Field Reorder Reproject Time Intervals
Server: Local Server
175MB / 224MB


Stream2* - IBM® SPSS® Modeler
File Edit Insert View Tools SuperNode Extensions Window Help
Streams Outputs Models
Stream1
project6
Stream2
gender
gender_standard
telco x data.txt
Type
gender_standard
bill_peak
bill_offpeak
total_bill
total_bill_greater_t..
segement
Discount
Reclassify
Table
Table (61 fields, 31,769 records) #6
File Edit Generate
Table Annotations
customer_id gender age postalcode region tariff
1 K338270 F 2... 6599.000 3.000 CAT 100 2006
2 K342660 M 2... 1635.000 1.000 CAT 100 2006
3 K342650 M 1... 2149.000 1.000 CAT 100 2003
4 K342640 M 2... 7788.000 4.000 CAT 100 2004
5 K342630 M 4... 5223.000 3.000 CAT 100 2004
6 K342620 F 2... 9143.000 4.000 CAT 100 2005
7 K342610 M 3... 4223.000 2.000 CAT 100 2003
8 K342600 F 5... 6165.000 3.000 CAT 100 2005
9 K342590 M 2... 7210.000 3.000 CAT 100 2003
10 K342580 F 1... 1630.000 1.000 CAT 100 2004
11 K342570 M 1... 9467.000 4.000 CAT 100 2004
12 K342560 M 2... 9287.000 4.000 CAT 100 2005
13 K342550 F 4... 3498.000 2.000 CAT 100 2004
14 K342540 F 1... 7618.000 4.000 CAT 100 2004
15 K342530 F 5... 9556.000 4.000 CAT 100 2003
16 K342520 M 4... 9179.000 4.000 CAT 100 2004
17 K342510 M 2... 2428.000 1.000 CAT 100 2004
18 K342500 F 2... 5867.000 3.000 CAT 100 2004
19 K342490 M 2... 5627.000 3.000 CAT 100 2003
20 K342480 M 1... 5207.000 3.000 CAT 100 2004
OK
Favorites Sources Record Ops Field Ops Graphs Modeling Output Export IBM® SPSS® Statistics Python Spark
Table Matrix Analysis Data Audit Transform Statistics Means Report Set Globals Sim Fit Sim Eval Extension Output KDE Simulation EVALUATE
Server: Local Server
178MB / 224MB