

# SPSS PRACTICAL

---

**Name:** Ganesh Agrahari

**Class:** BCA DS 33

**Roll No.:** 1230258176

**Submitted To:** Mr. Robin Tyagi

# INDEX

## SPSS Practical 9

This practical includes:

- Objective and Introduction
- Theoretical Background
- Methodology and Procedure
- Data Analysis
- Results and Interpretation
- Screenshots and Output
- Conclusion
- References

## Practical: 9

### Definition:

In order to build better models later you will cleanse data, bin fields, and transform fields so that their distribution is similar to the normal distribution. You will use the Filler node, Binning node and Transform node to accomplish this.

### Outcomes/Learning:

- Learned how to manage missing and inconsistent data using Filler Nodes.
- Understood how to segment continuous variables with Equal Count and Optimal Binning.
- Gained practical experience in transforming and normalizing fields to reduce data bias.
- Applied CLEM expressions within Derive Nodes to create new analytical variables.
- Used the Data Audit Node to confirm statistical accuracy and data quality.

### Required Tool:

IBM SPSS MODELER

### Working:

In this practical, the focus is on preparing and improving data quality through cleansing, binning, and transformation operations. The dataset (telco x churn data.dat) is imported and processed using various nodes to handle missing values, normalize distributions, and create new analytical fields.

The process begins by importing the dataset, defining field roles, and then using multiple Filler Nodes to replace empty demographic, revenue, and text-based records. Binning Nodes are applied to categorize continuous variables like BILL\_TOTAL using Equal Count and Optimal Binning methods. Transformations are performed to reduce skewness, and a Derive Node is used to apply mathematical operations through CLEM expressions.

Finally, a Data Audit Node is used to validate the transformations with statistical and graphical summaries, ensuring data completeness and readiness for further analysis.

### Main Nodes Used:

- Type Node – to assign field roles and measurement levels.
- Filler Node – to replace missing or blank values.
- Binning Node – to group continuous variables.
- Transform Node – to normalize and balance data.
- Derive Node – to create new calculated fields using CLEM.
- Data Audit Node – to validate final data quality.

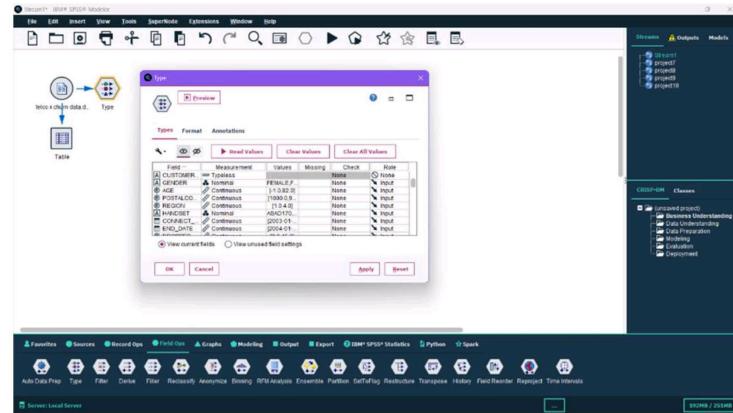
### **Step 1: Importing the Dataset and Viewing Initial Records**

- Open IBM SPSS Modeler and create a **New Stream**.
- Navigate to the **Sources** tab and double-click **Var. File**.
- Browse and select the dataset (telco x churn data.dat).
- Click **Apply**, then **OK** to load it.
- Add a **Table Node** to view initial records and click **Run** to display results.

The screenshot shows the IBM SPSS Modeler interface. On the left, the 'Sources' tab is active, displaying the 'Import' dialog for 'Var. File'. The file 'telco x churn data.dat' is selected, and the 'File' tab is chosen. The 'Content' tab shows the data structure: CUSTOMER\_ID, GENDER, AGE, INCOME, REGIONS, and CHURN\_DATE. Below this, 'Field roles' are defined: GENDER is 'Nominal', AGE is 'Continuous', INCOME is 'Continuous', REGIONS is 'Nominal', and CHURN\_DATE is 'Date'. The 'Transform' tab shows 'No transformations' applied. The 'OK' button is visible at the bottom. On the right, the 'Streams' tab is active, showing a stream named 'telco x churn data'. A 'Table' node is connected to the stream. A preview window shows the first 20 records of the dataset. The columns are CUSTOMER\_ID, GENDER, AGE, INCOME, REGIONS, and CHURN\_DATE. The data includes various gender and age categories, income levels, and churn dates. The 'OK' button is visible at the bottom of the preview window.

## Step 2: Defining Field Roles Using Type Node

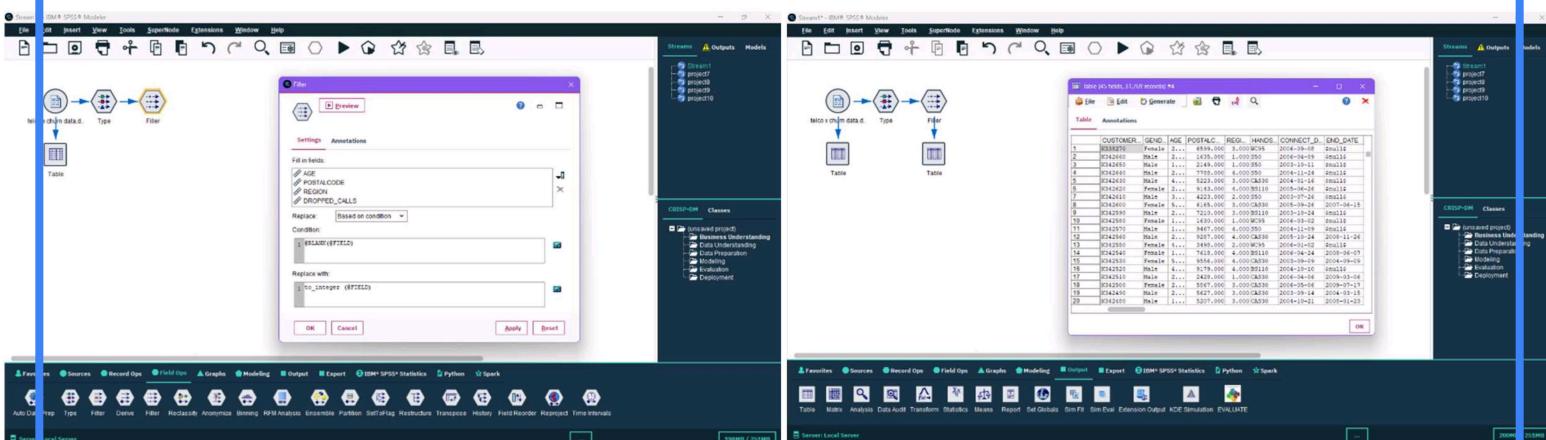
- Add a **Type Node** from the **Field Ops** tab and connect it to the dataset.
- Click **Read Values** to detect properties automatically.
- Adjust **Measurement Level** and **Field Roles** as needed.
- Click **Apply**, then **OK** to confirm.



## Step 3: Filling Empty Records (First Filler Node – Demographic Fields)

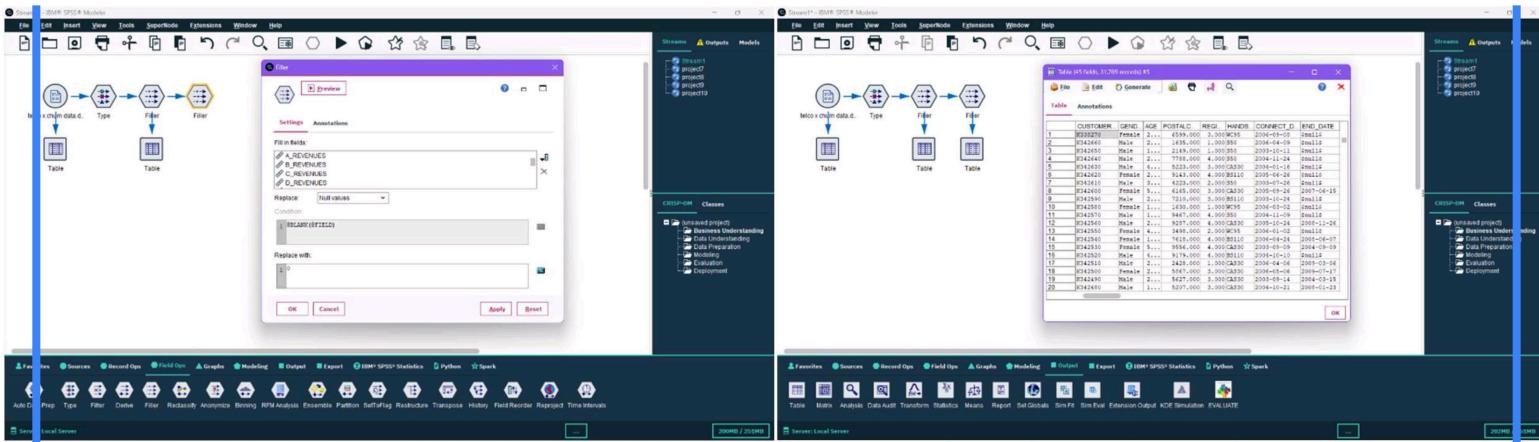
- Add a **Filler Node** connected to the Type Node.
- Configure:
  - **Target Fields:** AGE, POSTALCODE, REGION, DROPPED\_CALLS
  - **Replace Option:** Based on condition
  - **Condition Formula:** @BLANK(@FIELD) = 0
  - **Replace With Formula:** to\_integer(@FIELD)

Click **Apply**, then **OK**, and run using a Table Node.



## Step 4: Filling Empty Records (Second Filler Node – Revenue Fields)

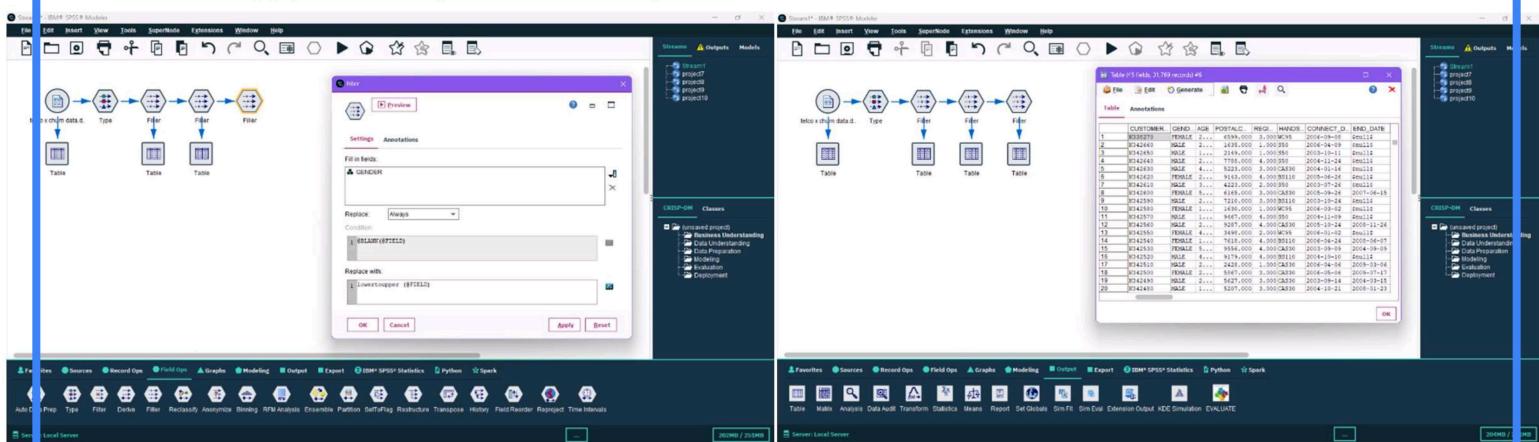
- Add another **Filler Node** connected to the previous one.
- Configure:
  - **Target Fields:** A\_REVENUES to L\_REVENUES
  - **Replace Option:** Based on condition
  - **Condition Formula:** @BLANK(@FIELD) = 0
  - **Replace With Formula:** 0
  - **Click Apply**, then **OK**, and run to verify the results.



### Step 5: Filling Empty Records (Third Filler Node – Text Fields)

- Add a **Filler Node** and configure:

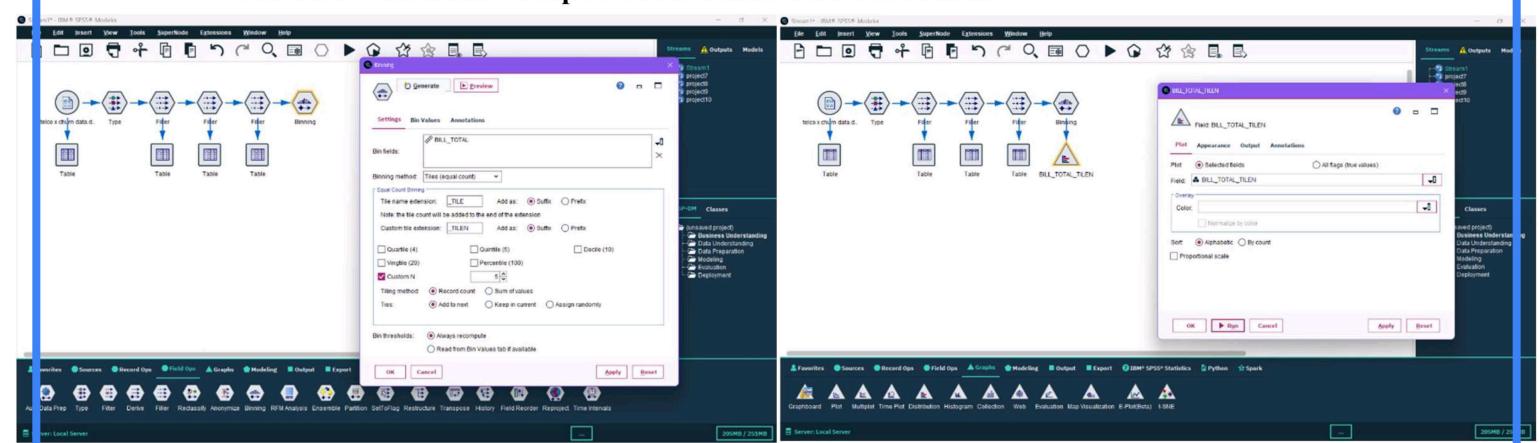
- **Target Field:** GENDER
- **Replace Option:** Always
- **Condition Formula:** @BLANK(@FIELD) = 0
- **Replace With Formula:** lowertoupper(@FIELD)
- **Apply and view updated results using a Table Node.**

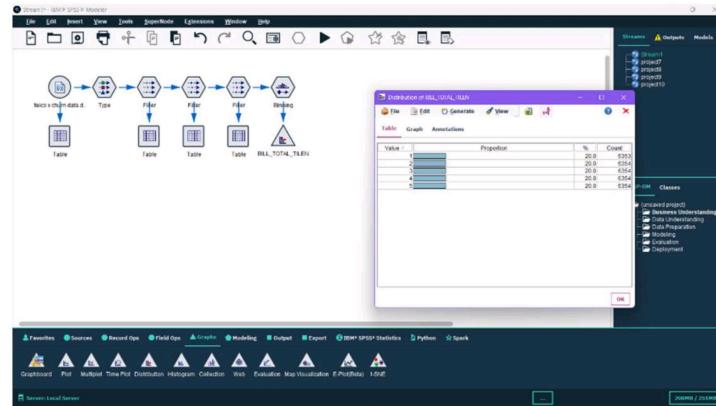


### Step 6: Applying First Binning Node (Equal Count)

- Add a **Binning Node** connected to the last Filler Node.
- Configure:

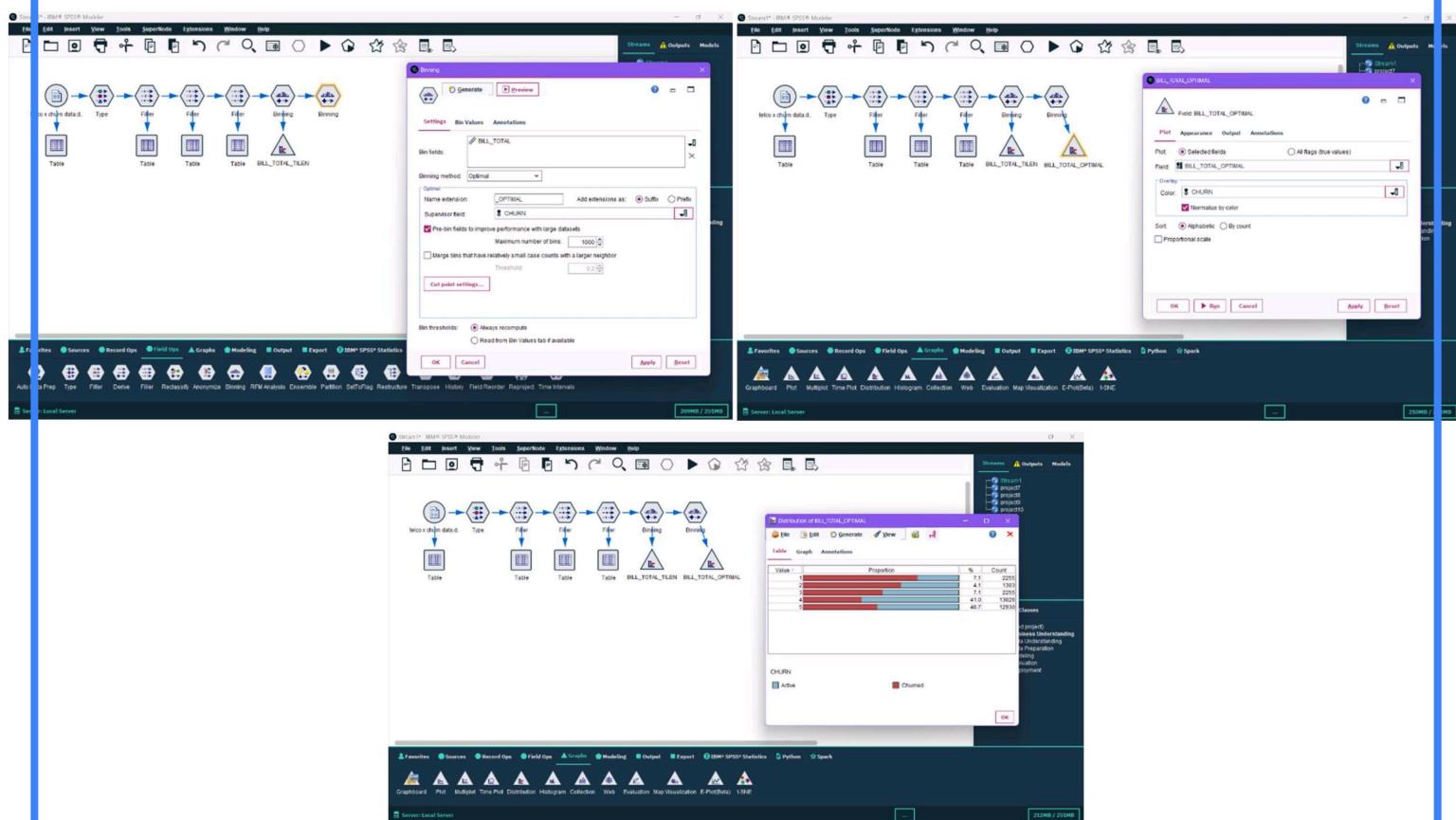
- **Bin Field:** BILL\_TOTAL
- **Binning Method:** Tiles (Equal Count)
- **Add Suffix:** \_TILEN
- **Number of Tiles:** as required (Quartile, Quintile, etc.)
- **Connect a Distribution Graph Node** to visualize the binned data.





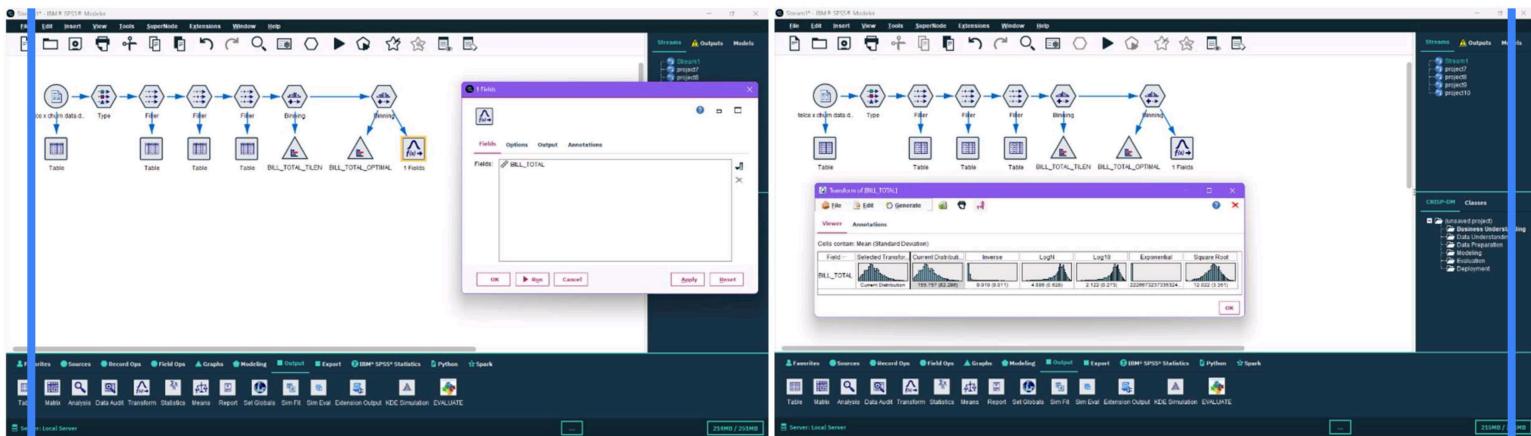
### Step 7: Applying Second Binning Node (Optimal Binning)

- Add another **Binning Node** connected to the previous one.
- Configure:
  - **Bin Field:** BILL\_TOTAL
  - **Binning Method:** Optimal
  - **Add Suffix:** \_OPTIMAL
  - **Supervisor Field:** CHURN
  - Connect a **Distribution Graph Node** and run to view results.



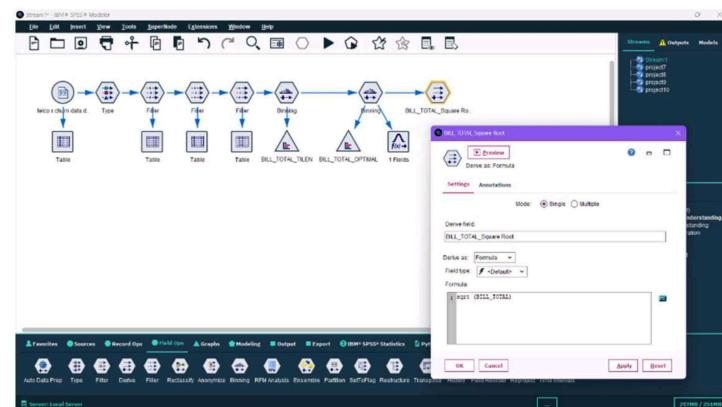
### Step 8: Applying Transform Node to Remove Bias

- Add a **Transform Node** connected to the second Binning Node.
- Select **BILL\_TOTAL** as the field to adjust.
- Click **Apply**, then **Run** to normalize the data distribution.



### Step 9: Deriving Square Root Transformation Using Derive Node

- Add a **Derive Node** connected to the Transform Node.
- Configure:
  - **New Field Name:** BILL\_TOTAL\_Square Root
  - **Derive Type:** Formula
  - **Field Type:** Default
  - **Formula:**  $\text{sqrt}(\text{BILL\_TOTAL})$
  - Apply and run to generate the derived field.



### Step 10: Performing Final Data Audit

- Add a **Data Audit Node** connected to the Derive Node.
- Configure:
  - **Option:** Use Custom Fields
  - **Display Options:** Graphs, Basic Statistics, Advanced Statistics
  - **Fields:** BILL\_TOTAL and BILL\_TOTAL\_Square Root
  - Apply and run to generate the audit report for validation.

