

# SPSS PRACTICAL

---

**Name:** Ganesh Agrahari

**Class:** BCA DS 33

**Roll No.:** 1230258176

**Submitted To:** Mr. Robin Tyagi

# INDEX

## SPSS Practical 7

This practical includes:

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

## Practical: 7

### Definition:

As a data miner in a telecommunications firm, you want to answer questions such as: Is churn related to handset? Is churn related to the number of dropped calls? And how strong is the relationship between number of gadgets purchased and revenues?

### Outcomes/Learning:

- Learned how to identify and analyze relationships between categorical and continuous fields.
- Understood the use of **Matrix, Means, and Statistics Nodes** for exploring correlations and group comparisons.
- Gained practical experience in interpreting data visually using **Distribution, Histogram, and Plot Nodes**.
- Enhanced understanding of data relationships essential for predictive modelling and business insights.

### Required Tool:

IBM SPSS MODELER

### Working:

In this practical, the goal is to explore and analyze relationships between categorical and continuous variables in the given dataset.

The dataset (telco x data.txt) is imported using a Var. File Node, and field properties are defined using a Type Node. Relationships between variables such as churn, handset, dropped calls, gadgets purchased, and revenues are examined using a combination of statistical and graphical nodes.

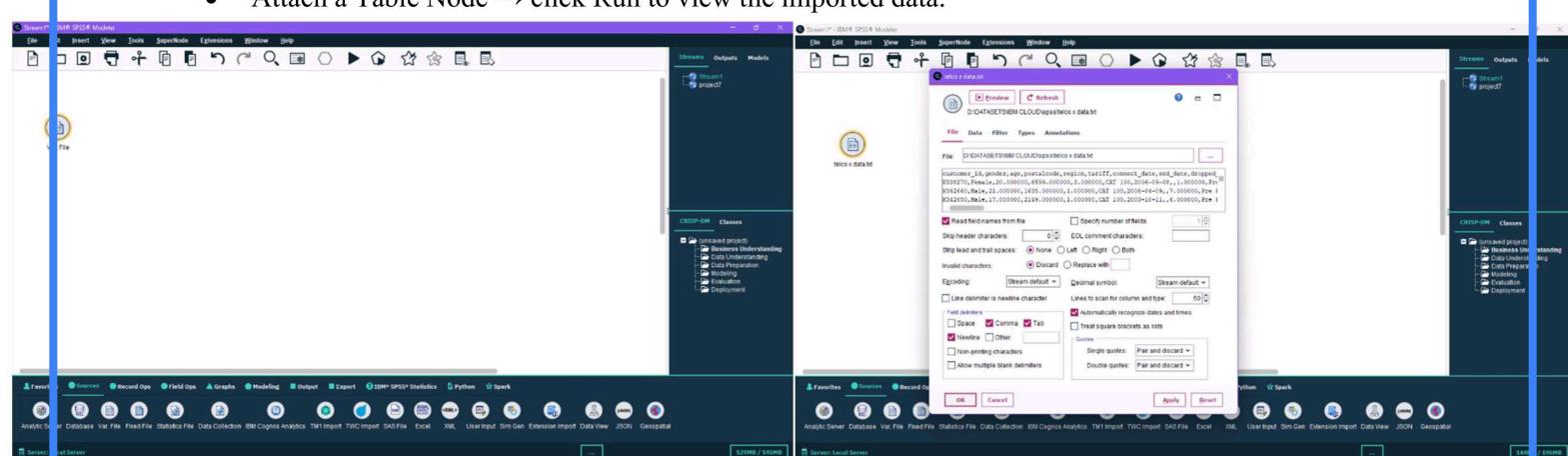
Different SPSS Modeler nodes are used to visualize, compare, and interpret data to identify patterns and correlations that can provide valuable business insights.

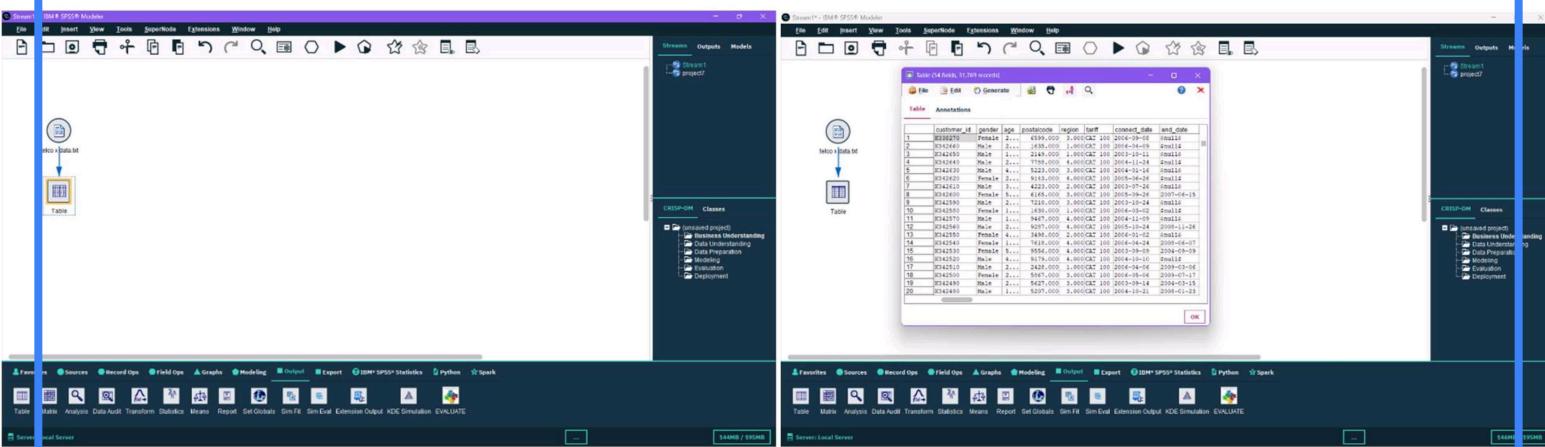
The main nodes used are:

- Type Node – to assign and define field roles and measurements.
- Matrix Node – to study the relationship between two categorical fields.
- Means & Histogram Nodes – to compare categorical and continuous fields.
- Statistics & Plot Nodes – to examine correlations between two continuous fields.
- Distribution Graph Node – to visualize categorical field distributions.
- Table Node – to display or validate data output at different stages.

### **Step 1: Importing the Dataset**

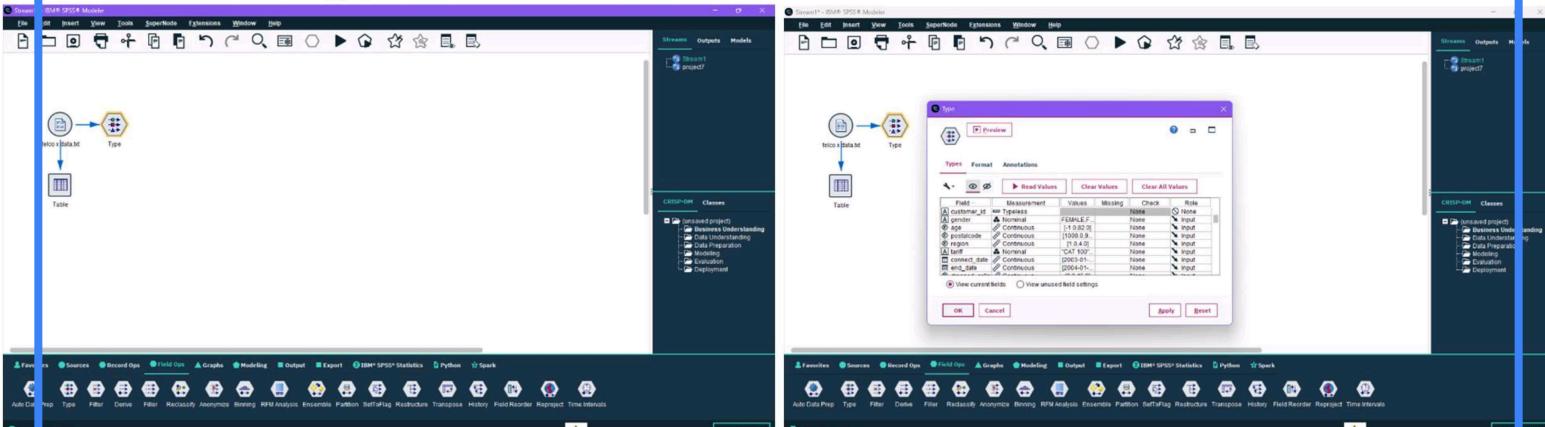
- Open IBM SPSS Modeler → create a new stream.
- Add a Var. File Node from the Sources tab → browse and select telco x data.txt.
- Click Apply → OK to load the dataset.
- Attach a Table Node → click Run to view the imported data.





## Step 2: Defining Field Roles

- Connect a Type Node to the Var. File Node.
  - Open the Type Node → click Read Values to automatically detect field types.
  - Review and adjust Measurement (Nominal, Continuous, etc.) and Role as needed.
  - Click Apply → OK to confirm.



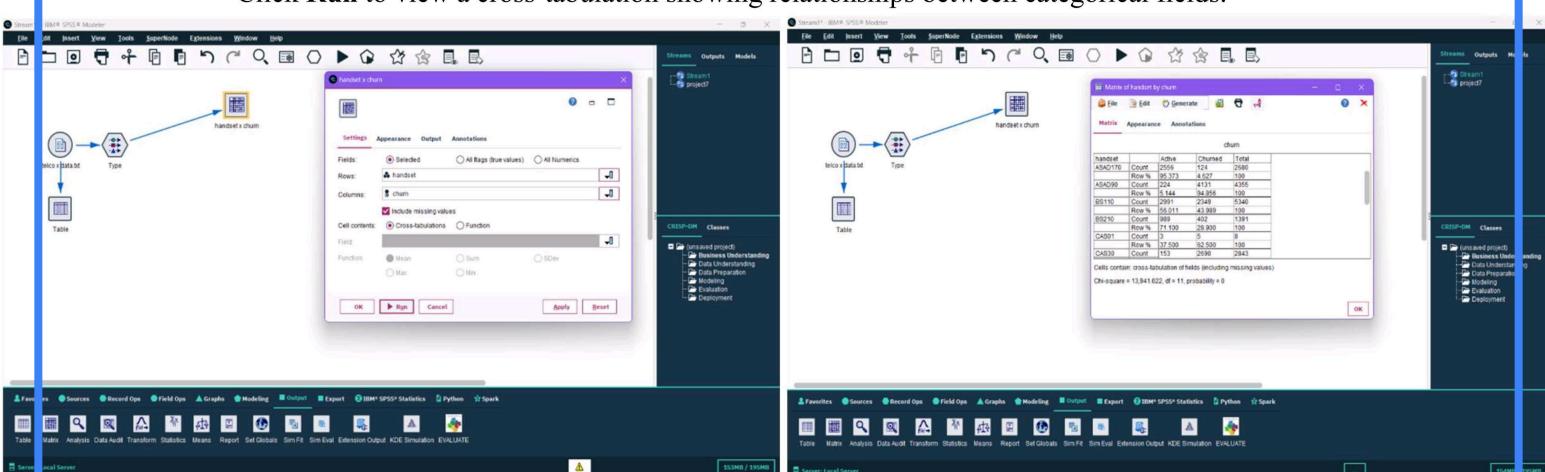
### Step 3: Analyzing Categorical Relationships

Attach a **Matrix Node** to the Type Node.

Configure as follows:

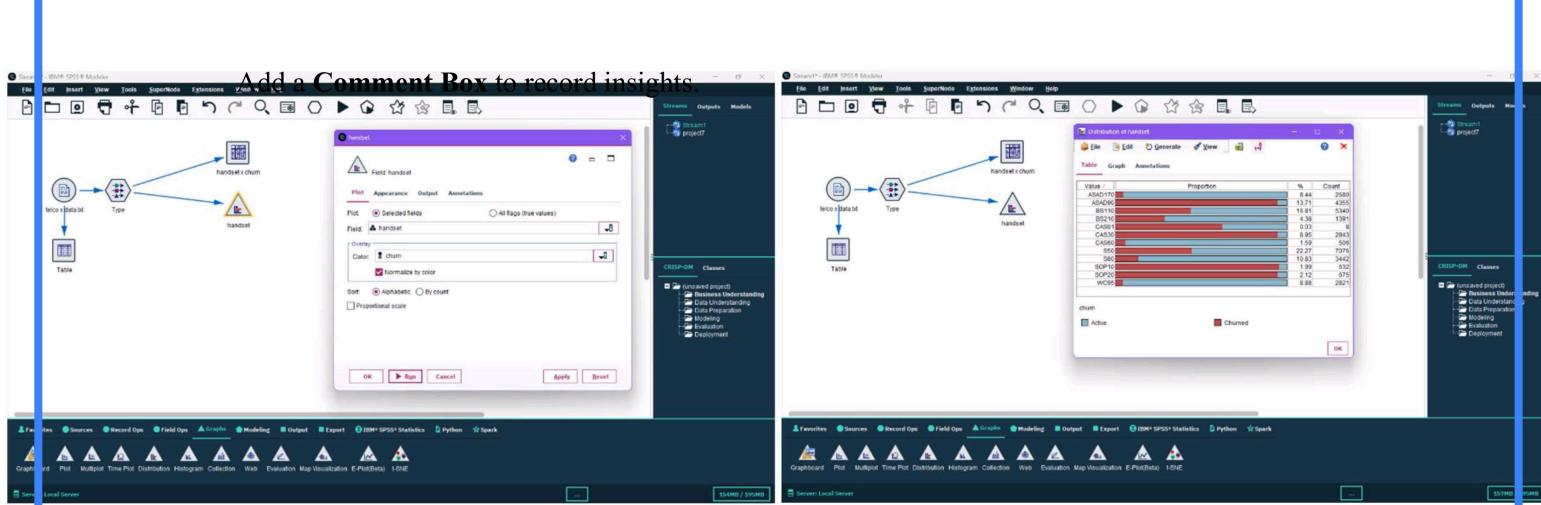
- **Rows:** handset
  - **Columns:** churn
  - Enable **Counts**, **Percentage of Rows**, and **Include Totals**

Enable Counts, Percentage of Rows, and Include Totals.  
Click **Run** to view a cross-tabulation showing relationships between categorical fields.



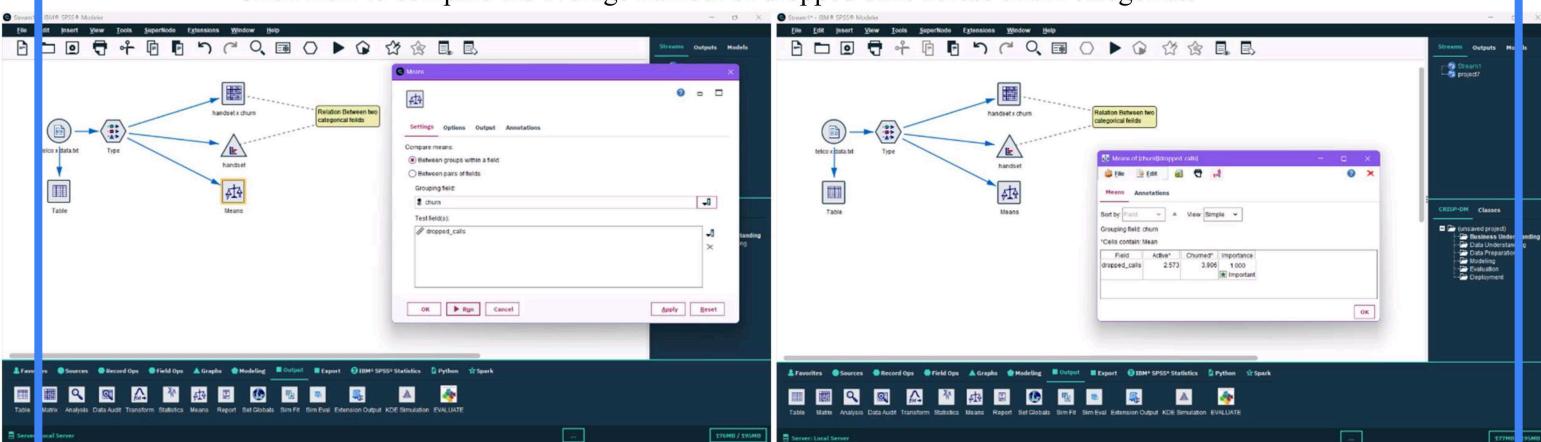
## Step 4: Visualizing Handset Distribution

- Attach a **Distribution Graph Node** to the Type Node.
  - Set **Field = handset**, **Overlay = churn**, and enable **Normalize by Colour**.
  - Click **Run** to visualize handset distribution by churn status.



## Step 5: Comparing Group Means

- Connect a **Means Node** to the Type Node.
- Set **Grouping Field = churn**, **Test Field = dropped\_calls**.
- Click **Run** to compare the average number of dropped calls across churn categories.



## Step 6: Visualizing Dropped Calls Distribution

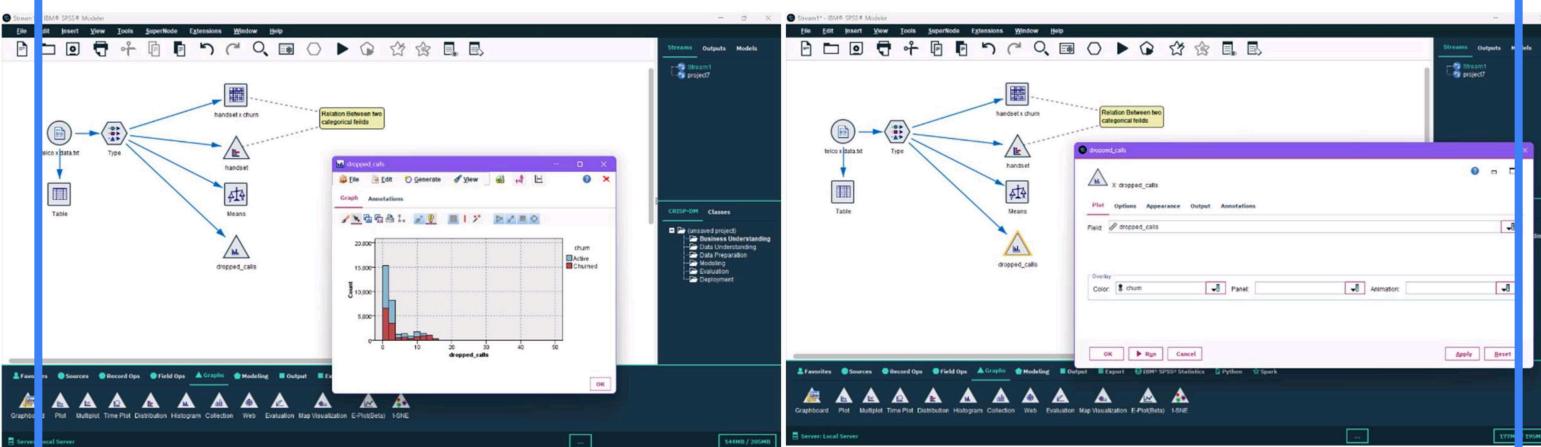
Attach a **Histogram Node** to the Type Node.

Configure:

- **Field = dropped\_calls**
- **Overlay = churn**
- Enable **Automatic X-Range** and specify the number of bins.

Click **Run** to generate a histogram comparing churn behavior with dropped calls.

Add a **Comment Box** to summarize the visualization.



## Step 7: Computing Field Statistics

Attach a **Statistics Node** to the Type Node.

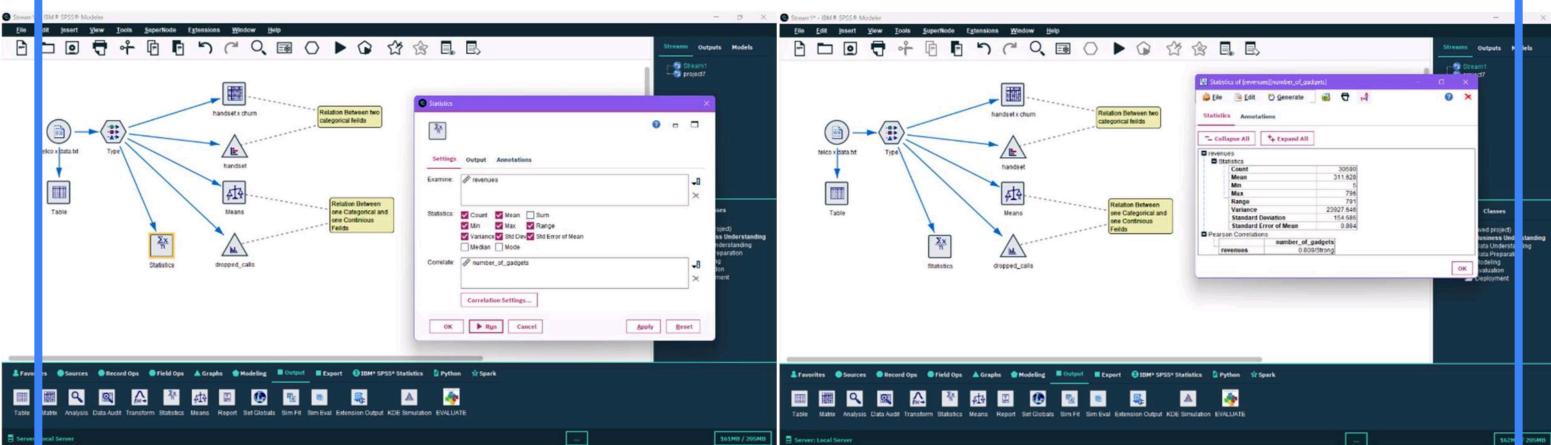
Configure:

- **Examine Field = revenues**

- **Correlate Field = number\_of\_gadgets**

Choose the correlation method (e.g., Pearson or Spearman).

Click **Run** to compute summary statistics and correlations for the selected fields.



### Step 8: Visualizing Relationships Between Continuous Fields

- Connect a **Plot Graph Node** to the **Type Node**.
- Set **X Field = number\_of\_gadgets**, **Y Field = revenues**.
- Click **Run** to create a scatter plot showing how gadgets purchased relate to revenue.
- Add a **Comment Box** to document findings.

### Step 9: Reviewing Outputs

- Attach **Table Nodes** after key analysis nodes (Matrix, Means, Statistics) to view outputs.
- Run each Table Node to validate and interpret results.
- Summarize all observations using **Comment Boxes** for proper documentation.

