

SPSS PRACTICAL

Name: Ganesh Agrahari

Class: BCA DS 33

Roll No.: 1230258176

Submitted To: Mr. Robin Tyagi

INDEX

SPSS Practical 10

This practical includes:

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

Practical: 10

Definition:

You will apply various transformations to sequence data, such as creating a record identifier and a moving average. You will restructure a transactional dataset into a dataset with one record per customer, which is the required data structure for correlational analyses.

Outcomes/Learning:

- Learned how to transform and restructure **sequence data** into a customer-level dataset.
- Understood the creation of **moving averages** and **count-based metrics** using Derive Node with CLEM expressions.
- Gained practical experience with the **Restructure** and **Aggregate Nodes** for summarizing and preparing data.
- Developed skills to prepare structured datasets suitable for **correlation and predictive analysis**.

Required Tool:

IBM SPSS MODELER

Working:

In this practical, the aim is to transform and restructure financial transaction data for analytical use. The dataset (year_balances.txt) is imported and enhanced with derived fields such as RECORD_ID, MA3 (three-period moving average), and NUMBER_OVERDRAWN (count of negative balances). The Type Node ensures proper measurement levels, maintaining consistency across fields.

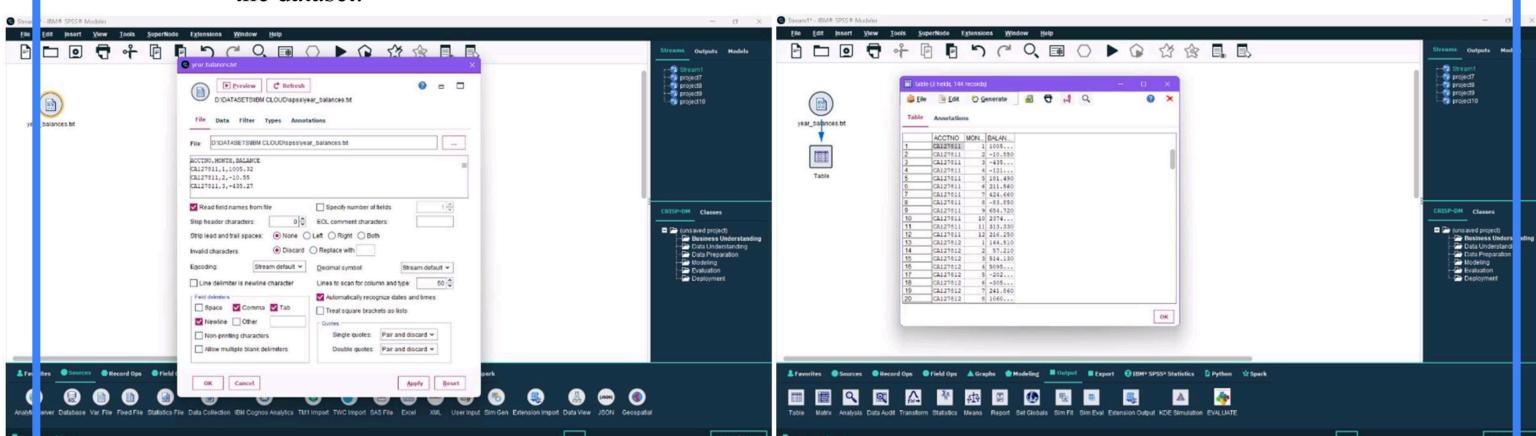
The Restructure Node converts time-series data into a single-record-per-customer format, facilitating customer-level analysis and modelling. Finally, the Aggregate Node summarizes the financial information at the account level, producing a dataset ready for advanced analysis.

Main Nodes Used:

- Var. File Node – to import the dataset.
- Derive Node – to create identifier, moving average, and calculated fields.
- Type Node – to define and adjust measurement levels.
- Restructure Node – to reshape transactional data.
- Aggregate Node – to summarize customer-level information.
- Table Node – to review and validate transformations.

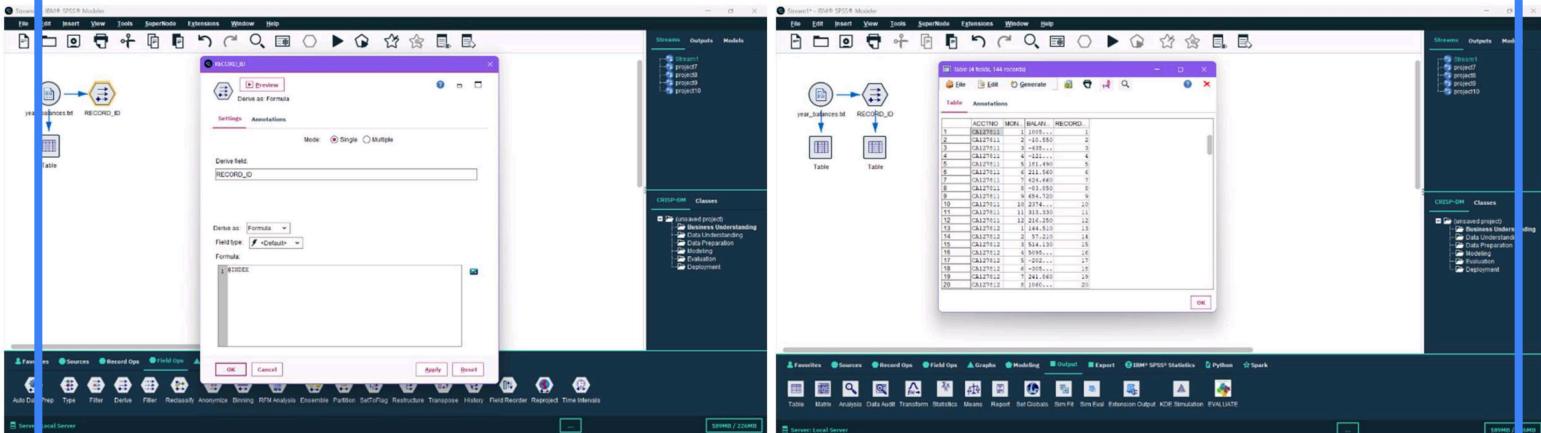
Step 1: Importing the Dataset and Viewing Initial Records

- Open IBM SPSS Modeler → create a New Stream.
- From the Sources tab, double-click Var. File to add a Var. File Source Node.
- Browse and select year_balances.txt using the three-dot button.
- Click Apply, then OK to load the dataset.
- Attach a Table Node → open it → configure display settings as required → click Apply then Run to view the dataset.



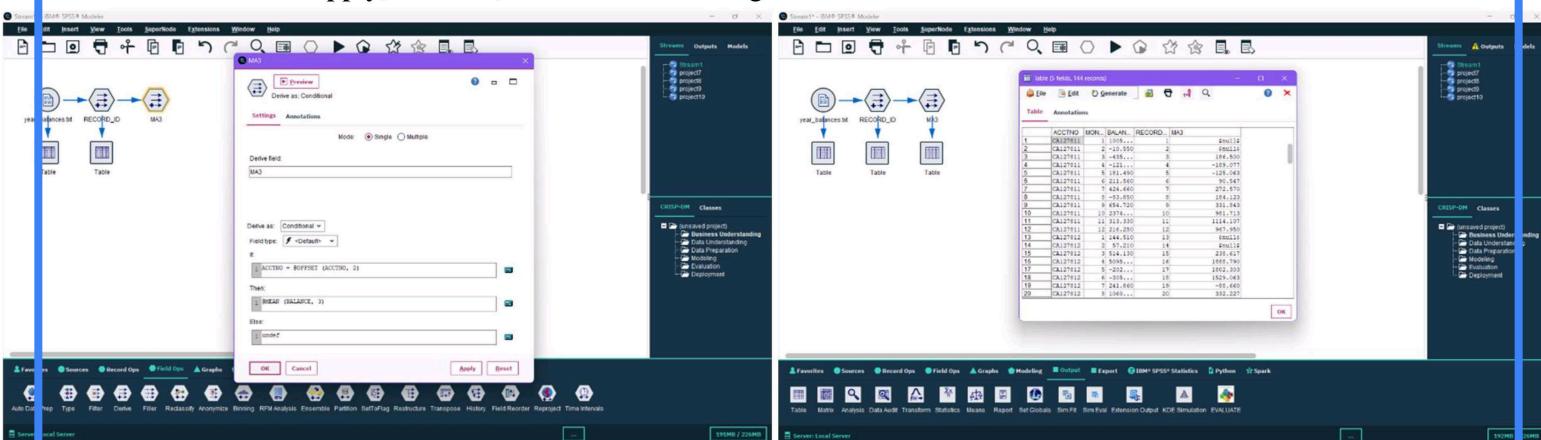
Step 2: Creating the RECORD_ID Field Using Derive Node

- Select the **Var. File Source Node**, go to **Field Ops** tab → double-click **Derive**.
- Configure as follows:
 - Rename field to **RECORD_ID**.
 - Mode: **Single**
 - Derive As: **Formula**
 - Field Type: **Default**
 - Formula: **@INDEX**
- Click **Apply**, then **OK**, and verify results using a **Table Node**.



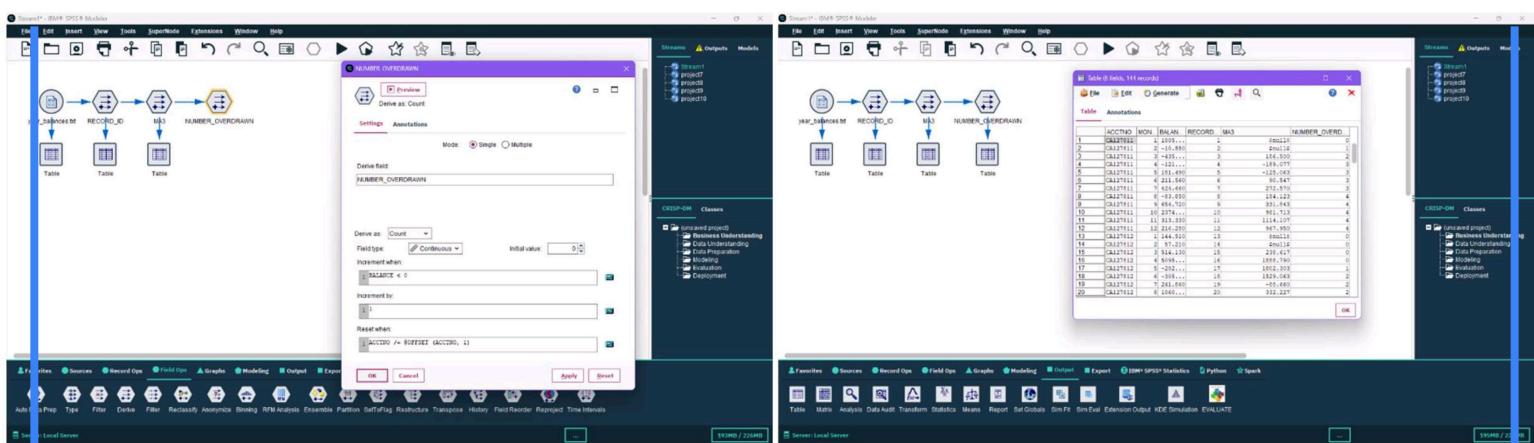
Step 3: Creating the MA3 Field Using Derive Node

- Select the **RECORD_ID Node**, go to **Field Ops** → double-click **Derive**.
- Configure as follows:
 - Rename field to **MA3**.
 - Mode: **Single**
 - Derive As: **Conditional**
 - Field Type: **Default**
- Formula:
If **ACCTNO = @OFFSET(ACCTNO, 2)** Then **@MEAN(BALANCE, 3)** Else **undef**
- Click **Apply**, then **OK**, and view results using a **Table Node**.



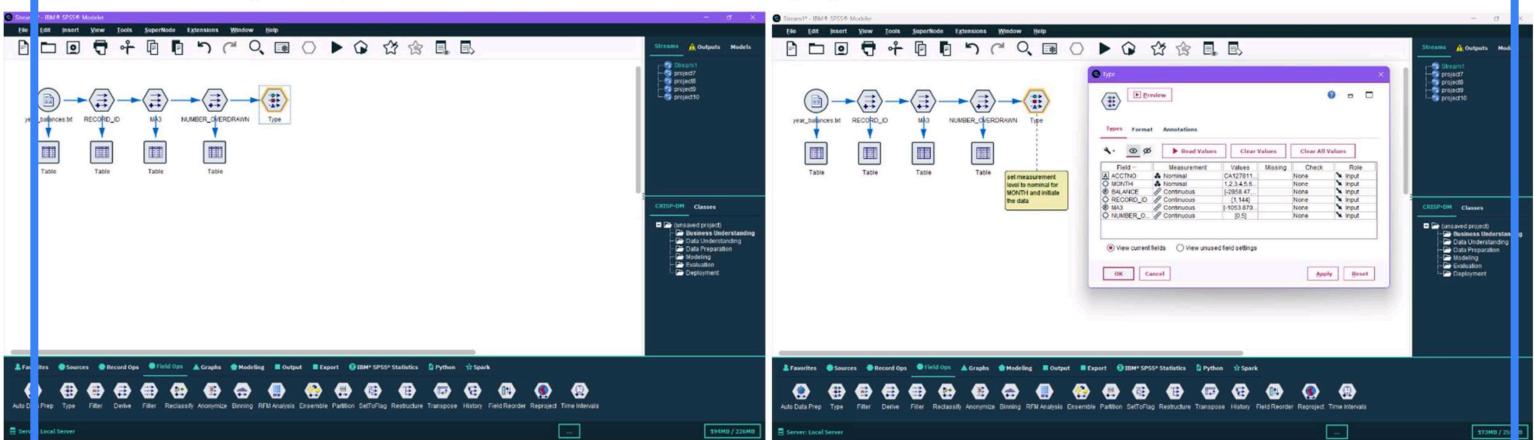
Step 4: Creating the NUMBER_OVERDRAWN Field Using Derive Node

- Select the **MA3 Node**, go to **Field Ops** → double-click **Derive**.
- Configure as follows:
 - Rename field to **NUMBER_OVERDRAWN**.
 - Mode: **Single**
 - Derive As: **Count**
 - Field Type: **Continuous**
- Formula:
Increment When **BALANCE < 0** Increment By 1 Reset When **ACCTNO /= @OFFSET(ACCTNO, 1)**
- Click **Apply**, then **OK**, and check results via a **Table Node**.



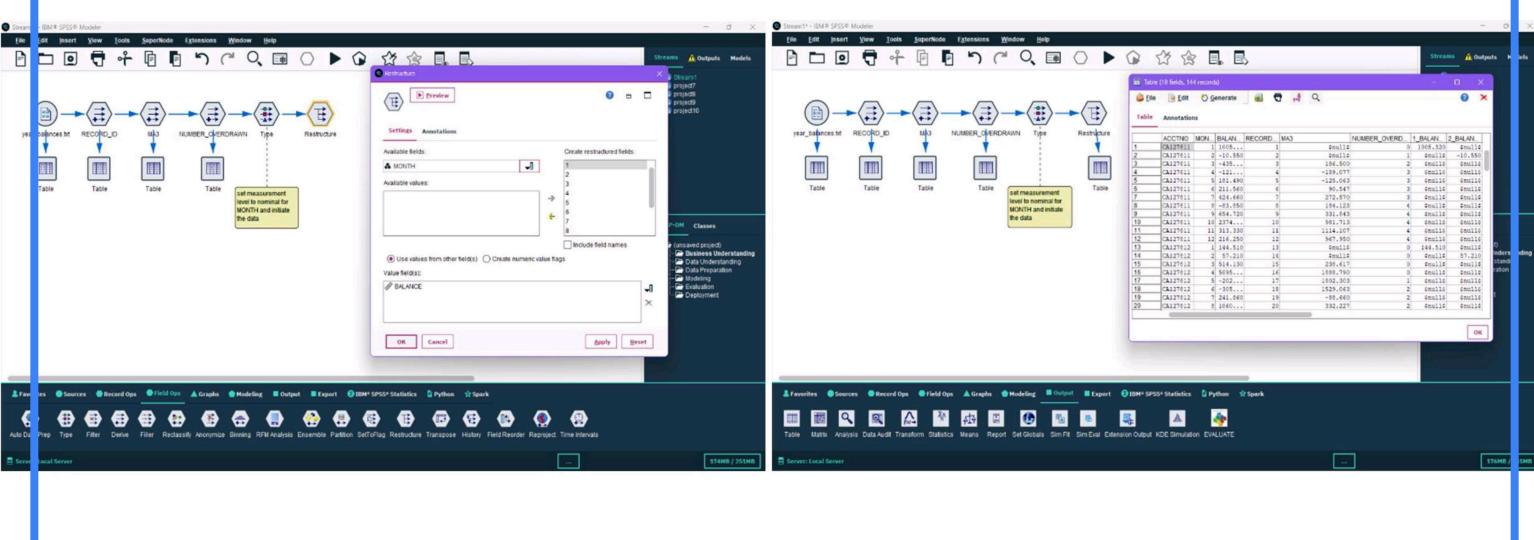
Step 5: Defining Field Roles with the Type Node

- Select the **NUMBER_OVERDRAWN** Node, go to **Field Ops** → double-click **Type**.
- Click **Read Values** to automatically detect field properties.
- Review and adjust the **Measurement** column (Nominal, Ordinal, Continuous, etc.) and set:
 - **MONTH** → Measurement Level = **Nominal**.
- Click **Apply**, then **OK** to confirm.
- Optionally, add a **Comment Box** to describe the purpose of this node.



Step 6: Restructuring Data Using the Restructure Node

- Select the **Type Node**, go to **Field Ops** → double-click **Restructure**.
 - Configure as follows:
 - In **Available Fields**, select **MONTH**.
 - Move all available values to create restructured fields.
 - Choose **Use values from other field(s)**.
 - In **Value Fields**, select **BALANCE**.
- Click **Apply**, then **OK**, and verify transformation with a Table Node.



Step 7: Aggregating Data Using the Aggregate Node

- Select the **Restructure Node**, go to **Record Ops** → double-click **Aggregate**.
 - Configure as follows:
 - **Key Field(s):** ACCTNO
 - **Default Mode:** Sum
- Click Apply, then OK, and view the summarized results using a Table Node.

