SPSS PRACTICAL

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

INDEX

SPSS Practical 11

This practical includes:

Practical: 11

Definition:

You will sample data using various techniques, and use partitioning to select the best predictive model.

Outcomes/Learning:

• Learned how to import and prepare data using **Var. File** and **Type Nodes**.
• Understood the use of **Simple** and **Complex Sampling** methods.
• Gained practical experience in **data partitioning** for model validation.
• Built an **Auto Numeric predictive model** using partitioned datasets.
• Visualized and balanced categorical data using **Graph** and **Balance Nodes**.

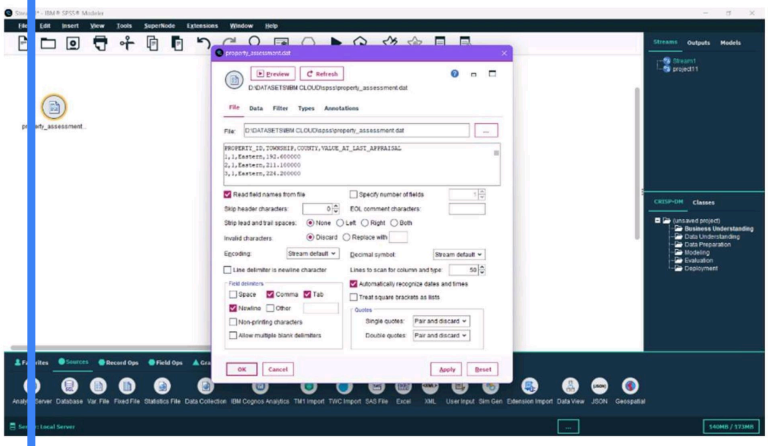
Required Tool:

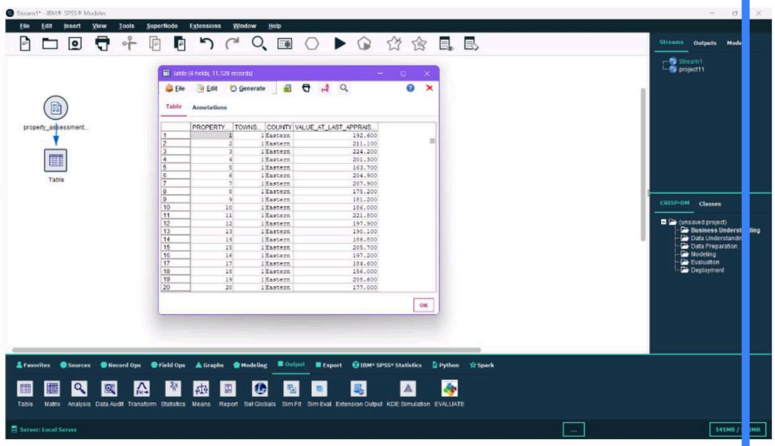
IBM SPSS MODELER

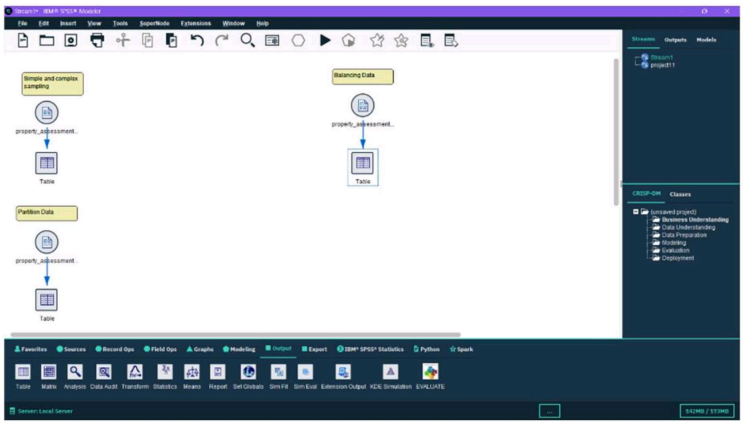
Working:

In this practical, the objective is to perform **data sampling, partitioning, and model building** for predictive analysis. The dataset (property_assessment.dat) is imported and explored, field roles are defined, and both **Simple** and **Complex Sampling** techniques are applied to obtain representative data subsets.
The dataset is then split into **Training (70%)** and **Testing (30%)** partitions using the Partition Node. An **Auto Numeric Node** is used to automatically build and evaluate predictive models based on the target field VALUE_AT_LAST_APPRAISAL. Finally, **Distribution Graph** and **Balance Nodes** are used to visualize and balance the COUNTRY field for fair model comparison.
Main Nodes Used:
• Var. File Node – to import and view data.
• Type Node – to define field roles and measurement levels.
• Sample Node – to perform simple and complex sampling.
• Partition Node – to split data into training and testing sets.
• Auto Numeric Node – to build predictive models.
• Balance & Distribution Graph Nodes – to visualize and balance categorical fields.

Step 1: Importing the Dataset
• Open **IBM SPSS Modeler** → create a **New Stream**.
• Add a **Var. File Node** → browse and select property_assessment.dat.
• Click **Apply** → **OK**, then attach a **Table Node** to view initial data.
• Duplicate the Var. File and Table Nodes twice for later use.

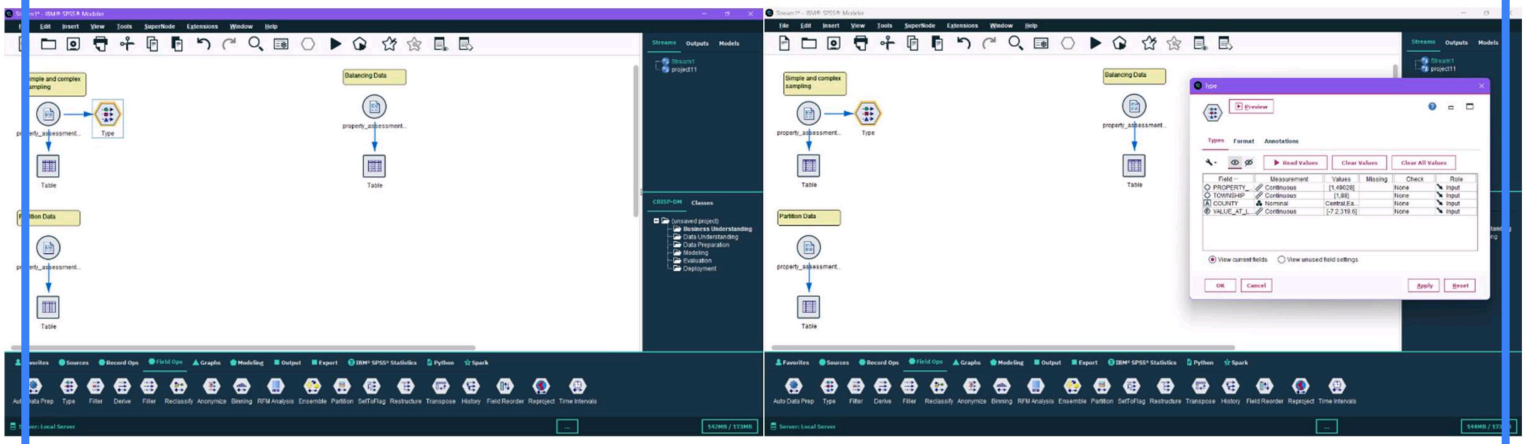






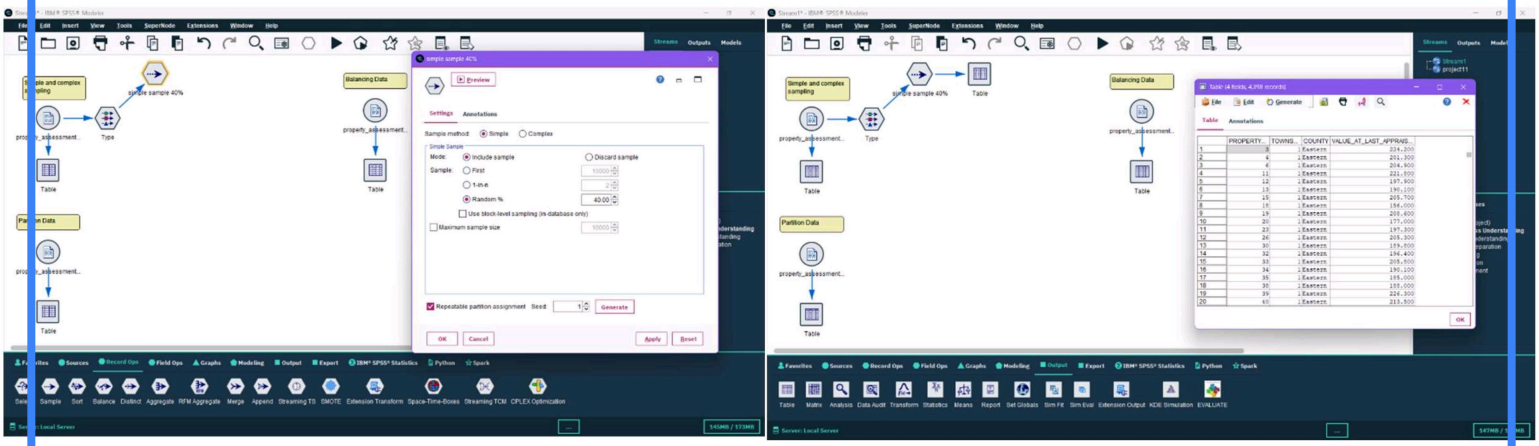
Step 2: Defining Field Roles

• Connect a **Type Node** to the Var. File Node.
• Click **Read Values** → review and adjust measurement levels and roles as required.
• Click **Apply**, then **OK**.



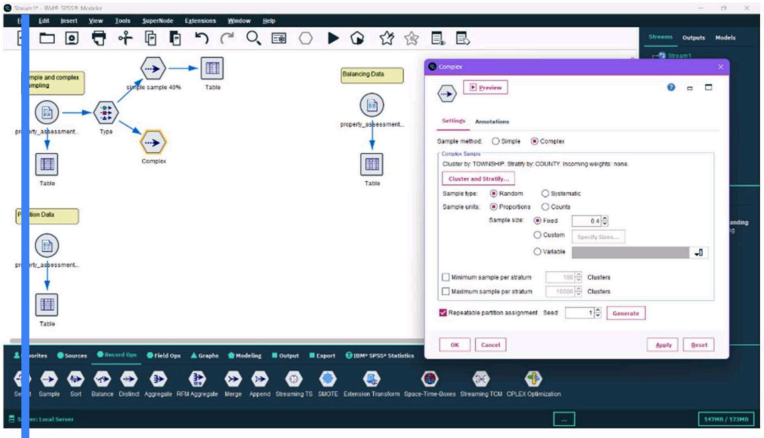
Step 3: Simple Sampling

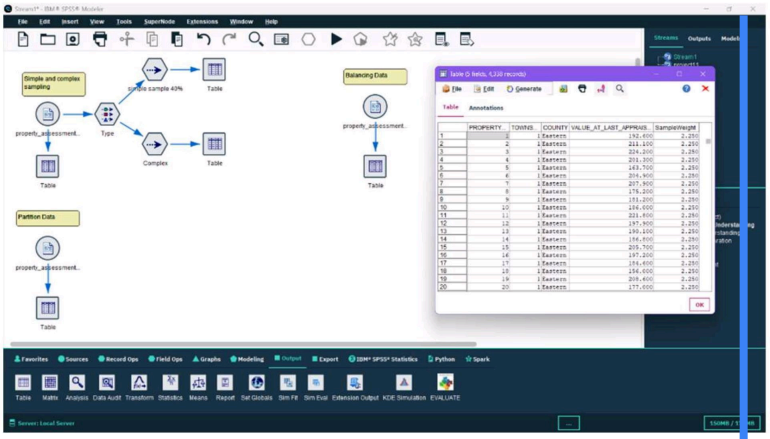
• Connect a **Sample Node** to the Type Node.
• Configure: **Method = Simple Sample**, **Type = Random %**, **Percentage = 40**, **Seed = 1**.
• Click **Apply**, **OK**, then view results with a **Table Node**.



Step 4: Complex Sampling

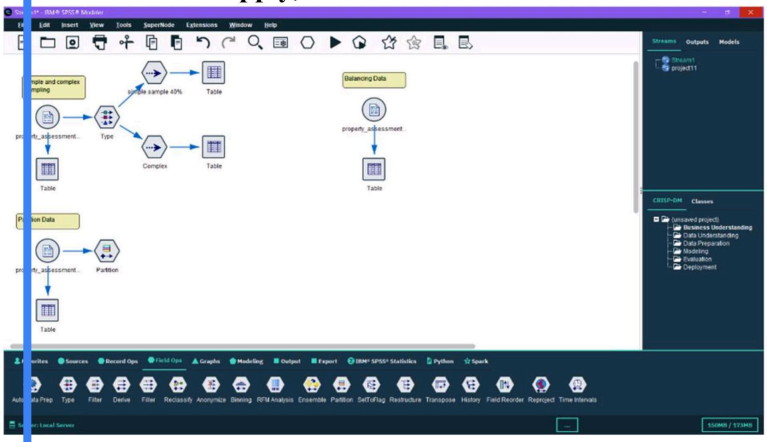
• Add another **Sample Node** to the same Type Node.
• Configure: **Method = Complex Sample**, **Cluster = TOWNSHIP**, **Stratify = COUNTRY**, **Sample Size = 0.4**, **Seed = 1**.
• Click **Apply**, **OK**, and check results via a **Table Node**.

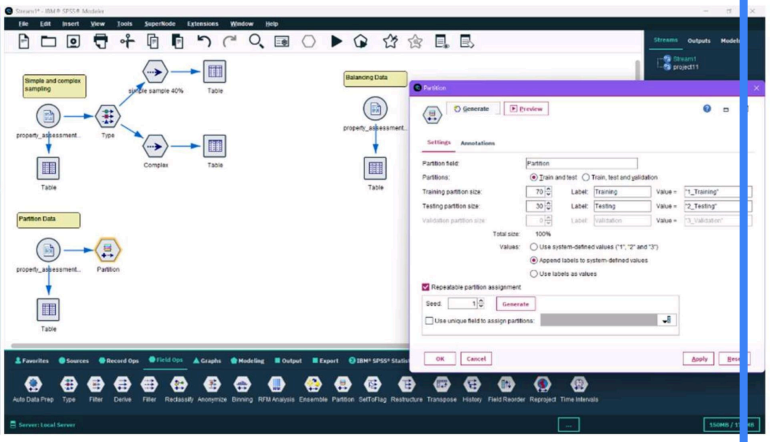




Step 5: Partitioning the Data

• Use the second **Var. File Node** → add a **Partition Node**.
• Configure: **Training = 70%**, **Testing = 30%**, **Seed = 1**, and enable repeatable assignment.
• Click **Apply**, then **OK**.

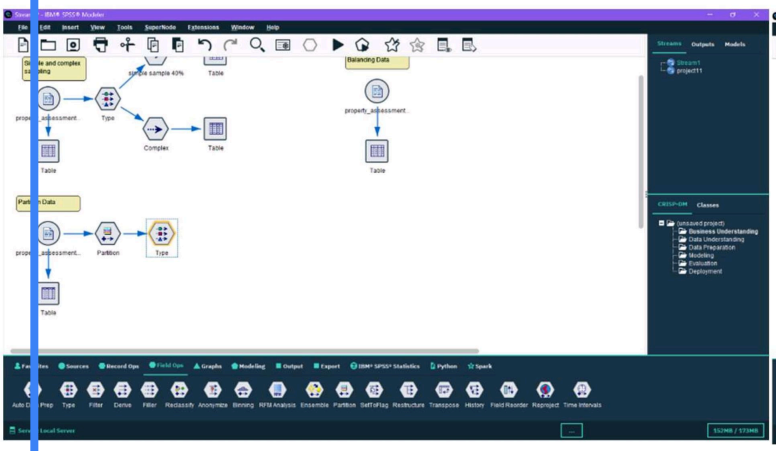


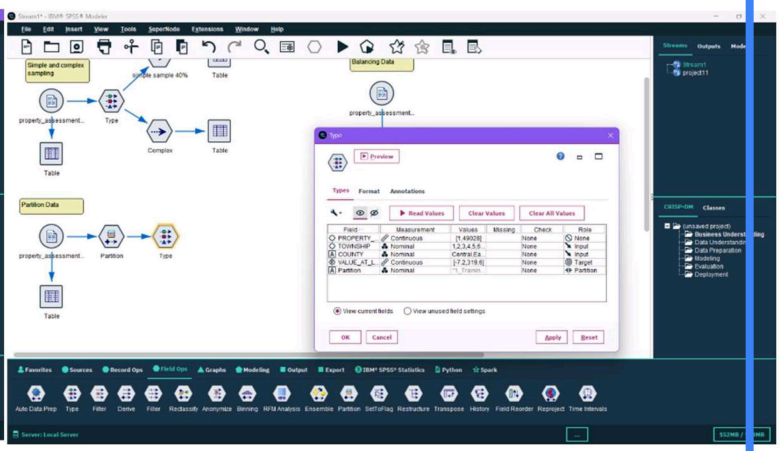


Step 6: Assigning Field Roles

• Connect a **Type Node** to the Partition Node.
• Set:

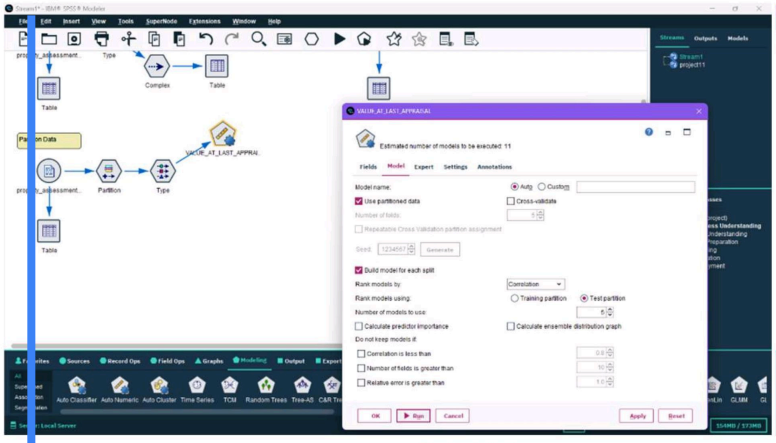
- **TOWNSHIP, COUNTRY:** Input
- **VALUE_AT_LAST_APPRAISAL:** Target
- **partition:** Partition
  • Click **Apply**, then **OK**.

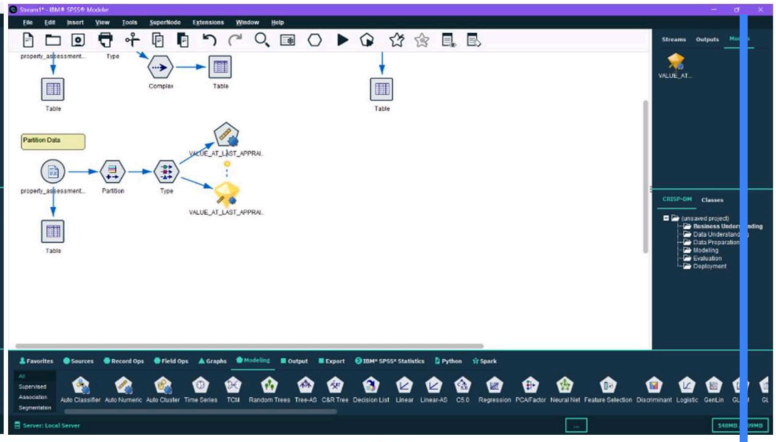


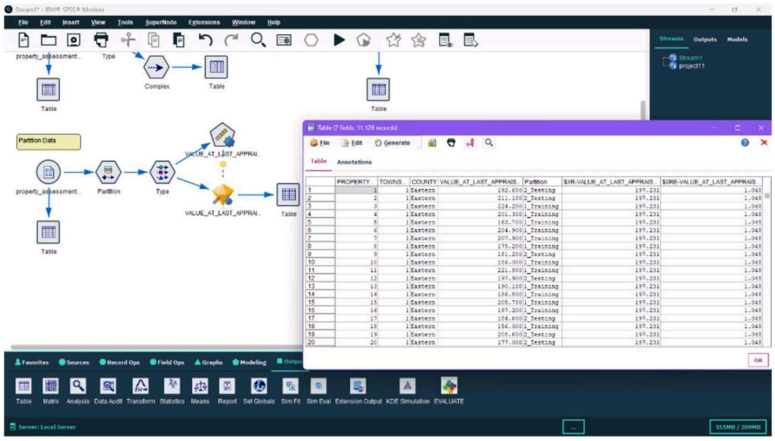


Step 7: Building the Auto Numeric Model

• Add an **Auto Numeric Node** to the Type Node.
• Use predefined roles, enable partitioned data, and build top 5 models ranked by correlation.
• Click **Apply**, **Run**, and review the results with a **Table Node**.

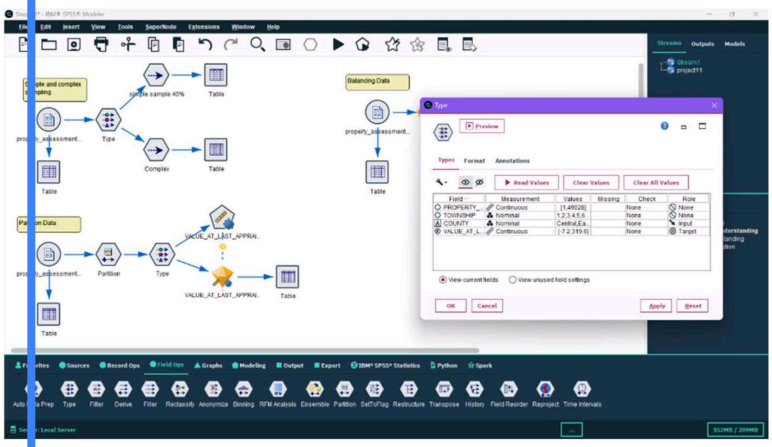


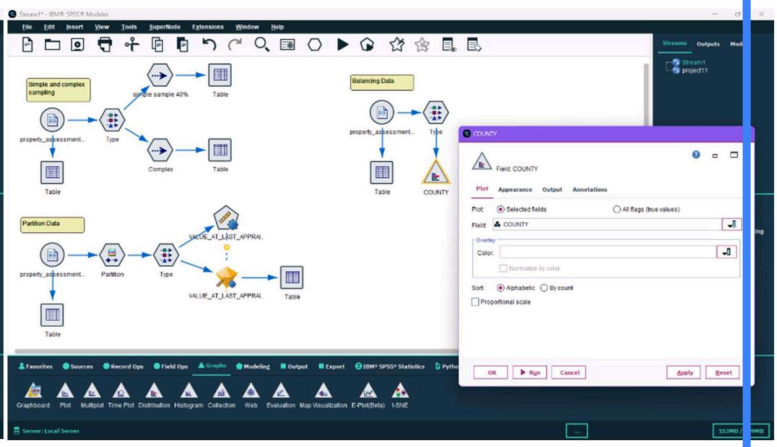


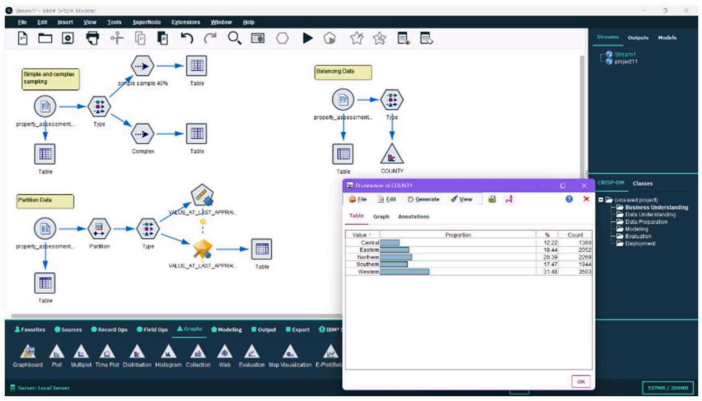


Step 8: Visualizing Country Distribution

- From the third **Var. File Node**, attach a **Type Node**, then a **Distribution Graph Node**.
- Set **Field = COUNTRY**, **Plot Type = Selected Fields**, **Sort = Alphabetic**.
- Click **Apply**, then **Run** to view the distribution.







Step 9: Balancing the Country Field

- Connect a **Balance Node** to the same Type Node.
- Configure balancing options as required → **Apply**, **OK**.
- Attach another **Distribution Graph Node** → view balanced results by **COUNTRY**.