

Practical File

Name : Ganesh Agrhari

Subject : Predictive Analytics (BCADSN15301)

Class : BCA DS&AI; 33

Roll No. : 1230258126

Submitted To : Mr. Robin Tyagi

10				
----	--	--	--	--

Practical: 1

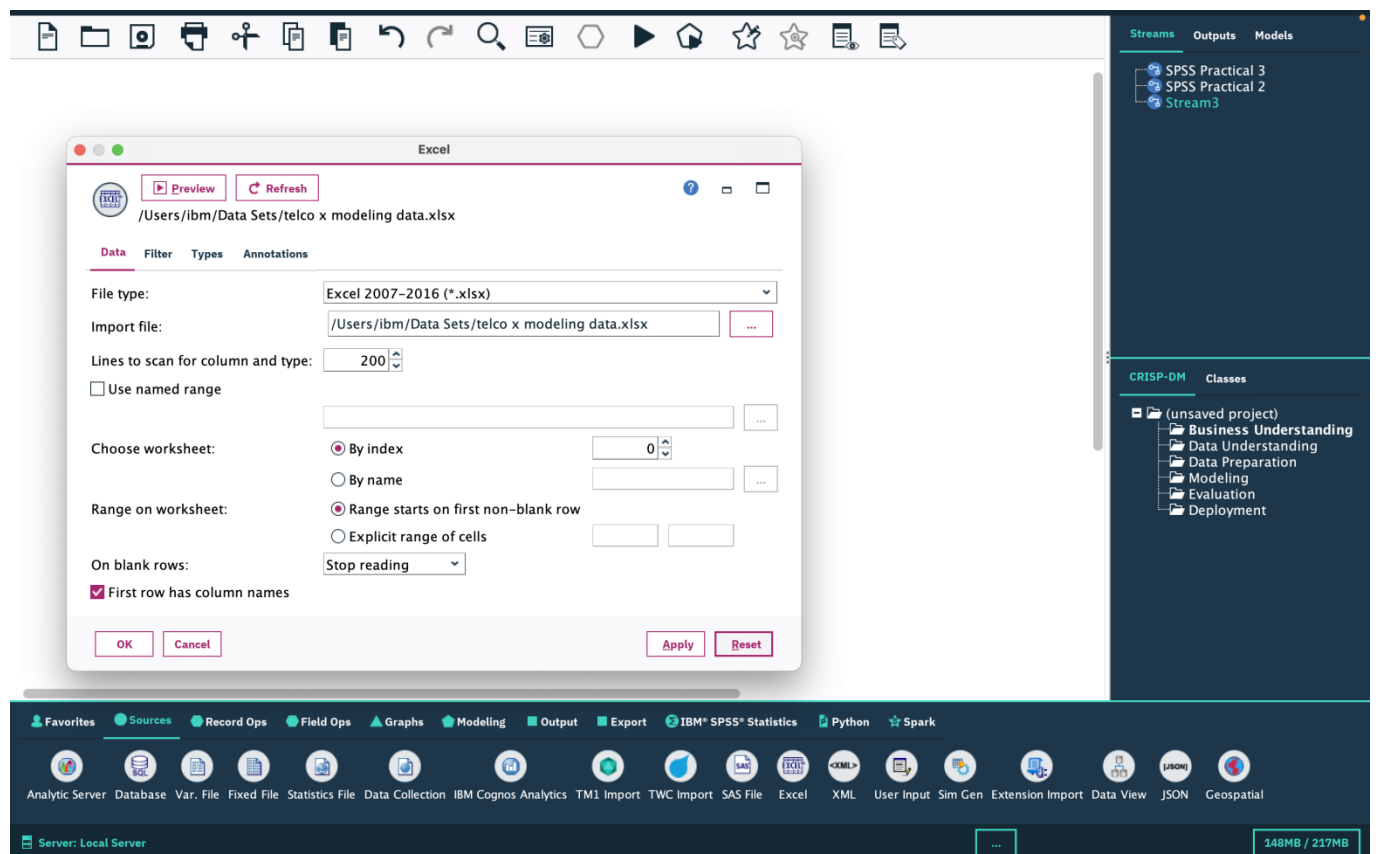
Definition: You want to become familiar with the IBM SPSS Modeler user interface that you will use to create and edit streams.

Outcomes/Learning: Through this process, we learned how to perform a basic data preparation workflow in IBM SPSS Modeler by importing a raw dataset, filtering out unnecessary fields, and viewing the cleaned data in a table format. The key outcome is a streamlined dataset containing only the relevant variables for further analysis.

Required Tool: IBM SPSS Modeler.

Working: This project file demonstrates a streamlined data preparation process using IBM SPSS Modeler. It imports raw data, filters for relevant fields, and outputs a clean, ready-to-use table for subsequent analysis.

Step 1: Import Data: The process begins by importing an Excel file named telco x modeling data.xlsx into the IBM SPSS Modeler.



Step 2: Filter Data: A **Filter** node is used to select specific fields from the imported data. In this case, fields like customer_id, data_known, gender, age, and churn were kept, while others like tariff and dropped_calls were discarded.

The screenshot displays the IBM SPSS Modeler interface. A 'Filter' node dialog box is open, showing the configuration for filtering data. The dialog has a 'Filter' tab and an 'Annotations' tab. The 'Filter' tab is active, showing a list of fields with checkboxes indicating whether they are to be kept or discarded. The fields are: customer_id, data_known, gender, age, tariff, dropped_calls, handset, peak_mins, peak_rate, and bill_peak. The 'Filter' column shows that customer_id, data_known, gender, and age are kept (indicated by a right arrow), while tariff, dropped_calls, handset, peak_mins, peak_rate, and bill_peak are discarded (indicated by a red X). The 'Annotations' column is empty. The dialog also shows 'Fields: 23 in, 18 filtered, 0 renamed, 5 out'. At the bottom of the dialog are buttons for 'OK', 'Cancel', 'Apply', and 'Reset'. The background shows the main workspace with a toolbar at the top and a sidebar on the right. The sidebar has tabs for 'Streams', 'Outputs', and 'Models'. Under 'Streams', there are three items: 'SPSS Practical 3', 'SPSS Practical 2', and 'Stream3'. Under 'Outputs', there is a folder named '(unsaved project)' containing several sub-folders: 'Business Understanding', 'Data Understanding', 'Data Preparation', 'Modeling', 'Evaluation', and 'Deployment'. The bottom status bar shows 'Server: Local Server' and '146MB / 217MB'.

Field	Filter	Field
customer_id	→	customer_id
data_known	→	data_known
gender	→	gender
age	→	age
tariff	✗	tariff
dropped_calls	✗	dropped_calls
handset	✗	handset
peak_mins	✗	peak_mins
peak_rate	✗	peak_rate
bill_peak	✗	bill_peak

Step 3: View Results: A Table node is connected to the Filter node to display the filtered data.

The screenshot displays the IBM SPSS Modeler software interface. At the top, a toolbar contains various icons for file operations, navigation, and execution. Below the toolbar, a workflow diagram shows three nodes: 'telco x modeling dat..' (a source node), 'Filter' (a transformation node), and 'Table' (a target node), connected by green arrows. In the center, an 'Execution Feedback' dialog box is open, displaying a progress bar and a table with the following content:

Node	State
Table	Running

Below the table, the dialog shows 'Time elapsed 00:00:07' and a 'Stop Execution' button. A checkbox labeled 'Close dialog once execution is complete' is checked. The right sidebar shows a project tree under 'CRISP-DM' with folders for 'Business Understanding', 'Data Understanding', 'Data Preparation', 'Modeling', 'Evaluation', and 'Deployment'. The bottom status bar indicates 'Server: Local Server', 'Records In : 15,000', 'Records Out : 15,000', and a memory usage of '146MB / 217MB'.

Step 4: Execute and Display: The execution of the stream is initiated, and the final output is a table showing a preview of the data with the selected fields.

The screenshot displays the IBM SPSS Modeler software interface. At the top, a toolbar contains icons for file operations, editing, and execution. The main workspace shows a workflow stream with three components: a source node labeled 'telco x modeling dat..', a transformation node labeled 'Filter', and a target node labeled 'Table'. On the right side, a panel titled 'Streams' lists the components of the stream: 'SPSS Practical 3', 'SPSS Practical 2', and 'Stream3'. Below this, a 'CRISP-DM' section shows a project structure with folders for 'Business Understanding', 'Data Understanding', 'Data Preparation', 'Modeling', 'Evaluation', and 'Deployment'. At the bottom, a navigation bar includes tabs for 'Favorites', 'Sources', 'Record Ops', 'Field Ops', 'Graphs', 'Modeling', 'Output', 'Export', 'IBM® SPSS® Statistics', 'Python', and 'Spark'. The 'Output' tab is currently selected, showing a list of available output types: 'Table', 'Matrix', 'Analysis', 'Data Audit', 'Transform', 'Statistics', 'Means', 'Report', 'Set Globals', 'Sim Fit', 'Sim Eval', 'Extension Output', 'KDE Simulation', and 'EVALUATE'. The status bar at the very bottom indicates 'Server: Local Server' and shows a memory usage indicator of '147MB / 217MB'.