

Practical File

Name : Ganesh Agrhari

Subject : Predictive Analytics (BCADSN15301)

Class : BCA DS&AI; 33

Roll No. : 1230258126

Submitted To : Mr. Robin Tyagi

Practical: 2

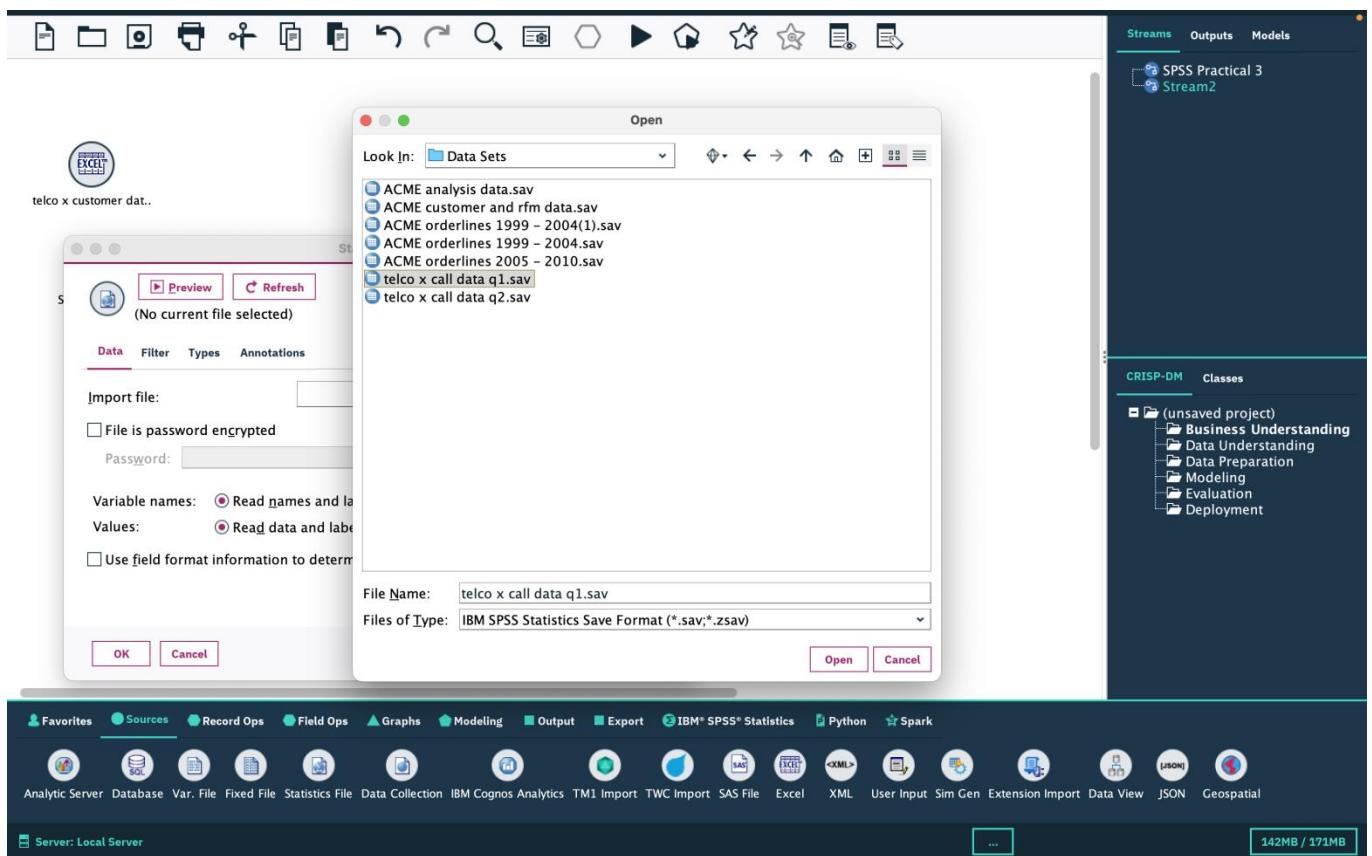
Definition: You work as a data miner for a telecommunications firm. You have to import data from various sources and examine the unit of analysis and the measurement levels of the fields.

Outcomes/Learning: We learned how to set up a basic data analysis stream in IBM SPSS Modeler by importing multiple raw data files, defining their field properties with the Type node, and verifying the contents using the Table node. The outcome is a prepared dataset ready for further analysis.

Required Tool: IBM SPSS Modeler.

Working: In this process shows how to import three separate datasets (customer, call, and products) and prepare them for analysis by using the Type node to correctly identify the data type of each variable. The Table node then provides a view of the cleaned data, confirming the preparation steps were successful.

Step 1: Importing Data: We start by importing multiple .sav files, including telco x customer dat.sav, telco x call data q1.sav, and telco x products.dat, into IBM SPSS Modeler.



Step 2: Defining Data Types: A Type node is attached to each data source to automatically read and define the measurement level and role of each field (e.g., customer_id as Nominal, and peak_calls as Continuous).

The screenshot shows the SPSS Modeler interface with a process flow and a detailed configuration dialog.

Process Flow:

```

graph LR
    A[telco x customer dat..] --> B{Type}
    B --> C[Table]
    D[telco x call data q1..] --> E{Type}
    E --> F[Table]
    G[telco x products.dat] --> H{Type}
    H --> I[Table]
  
```

Type Dialog (Detailed View):

Fields Defined:

Field	Measurement	Values	Missing	Check	Role
customer_id	Nominal	<Read>	*	None	None
peak_calls	Continuous	<Read>	None	None	Input
peak_mins	Continuous	<Read>	None	None	Input
offpeak_calls	Continuous	<Read>	None	None	Input
offpeak_mins	Continuous	<Read>	None	None	Input
weekend_calls	Continuous	<Read>	None	None	Input
weekend_mins	Continuous	<Read>	None	None	Input
international_calls	Continuous	<Read>	None	None	Input

Dialog Buttons:

- OK
- Cancel
- Apply
- Reset

CRISP-DM Classes:

- (unsaved project)
 - Business Understanding
 - Data Understanding
 - Data Preparation
 - Modeling
 - Evaluation
 - Deployment

SPSS Modeler Main Interface:

- Favorites, Sources, Record Ops, Field Ops, Graphs, Modeling, Output, Export, IBM SPSS Statistics, Python, Spark
- Table, Matrix, Analysis, Data Audit, Transform, Statistics, Means, Report, Set Globals, Sim Fit, Sim Eval, Extension Output, KDE Simulation, EVALUATE
- Server: Local Server
- Memory: 144MB / 171MB

Step 3: Viewing Data: A Table node is then connected to each Type node to display the contents and structure of each individual dataset. The final outputs show the distinct data tables for customer information, call data, and product data.

The screenshot shows the SPSS Practical 3 Stream interface. On the left, a table titled "Telco x call data (10 fields, 95,307 records)" displays data for 20 rows. The columns include customer_id, peak_calls, peak_mins, offpeak_calls, offpeak_mins, weekend_calls, weekend_mins, and international_mins. The data shows various call patterns for different customers. An "OK" button is visible at the bottom right of the table view. On the right, the Stream interface shows a sidebar with "Streams", "Outputs", and "Models". Below that, a project named "SPSS Practical 3" is listed under "Stream2". Under "CRISP-DM Classes", there is an "(unsaved project)" entry with sub-items: Business Understanding, Data Understanding, Data Preparation, Modeling, Evaluation, and Deployment. At the bottom of the interface, there is a navigation bar with tabs like Favorites, Sources, Record Ops, Field Ops, Graphs, Modeling, Output, Export, IBM SPSS Statistics, Python, and Spark. Below the navigation bar are various tool icons for 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 bottom indicates "Server: Local Server", "...", and "145MB / 217MB".

	customer_id	peak_calls	peak_mins	offpeak_calls	offpeak_mins	weekend_calls	weekend_mins	international_mins
1	K100010	2.000	6.086	1.000	1.343	4.000	2.448	0.243
2	K100010	2.000	6.060	1.000	1.337	4.000	2.437	0.242
3	K100010	2.000	5.494	1.000	1.212	3.000	2.210	0.220
4	K100020	7.000	5.538	5.000	2.970	0.000	0.000	1.373
5	K100020	9.000	6.875	6.000	3.688	0.000	0.000	1.705
6	K100020	8.000	6.172	5.000	3.310	0.000	0.000	1.531
7	K100030	6.000	9.471	0.000	0.000	3.000	4.853	2.166
8	K100030	8.000	13.963	0.000	0.000	4.000	7.154	3.193
9	K100030	6.000	10.239	0.000	0.000	3.000	5.247	2.341
10	K100040	8.000	13.417	0.000	0.000	4.000	6.875	3.068
11	K100040	8.000	13.516	0.000	0.000	4.000	6.925	3.091
12	K100040	6.000	10.766	0.000	0.000	3.000	5.517	2.462
13	K100050	6.000	7.534	2.000	3.492	0.000	0.000	0.830
14	K100050	5.000	7.062	2.000	3.274	0.000	0.000	0.778
15	K100050	6.000	8.046	3.000	3.730	0.000	0.000	0.886
16	K100060	8.000	6.364	1.000	1.525	1.000	1.408	0.256
17	K100060	10.000	8.751	1.000	2.097	1.000	1.936	0.352
18	K100060	7.000	6.234	1.000	1.494	1.000	1.379	0.250
19	K100070	9.000	7.524	1.000	0.535	0.000	0.000	0.891
20	K100070	12.000	10.188	1.000	0.725	0.000	0.000	1.206