

BABU BANARASI DAS UNIVERSITY

NAME : SHRASHTI KATIYAR

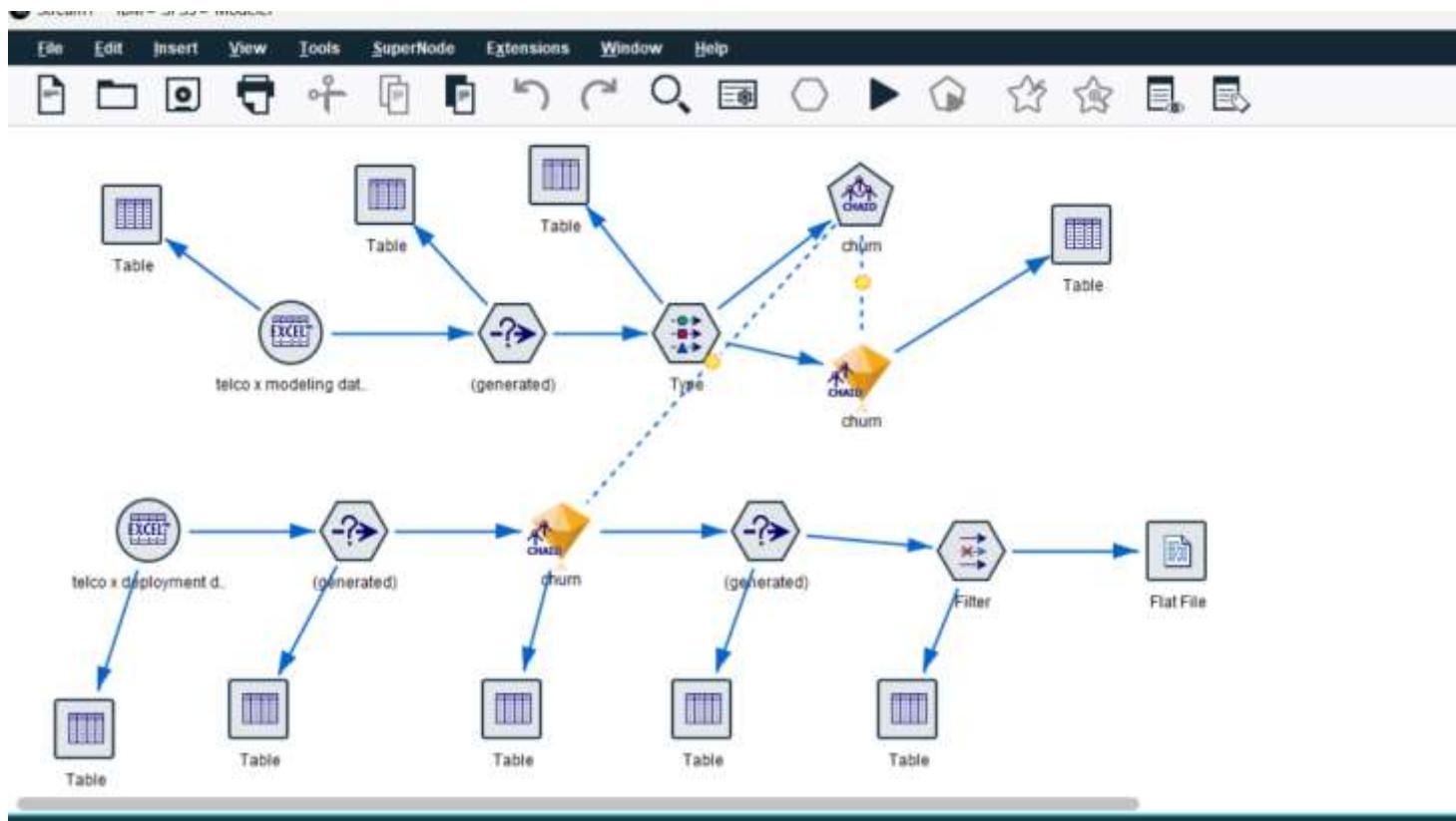
ROLL NO. : 1230258412

SUBMIT TO : MR. VIKASH SIR

ASSIGNMENT- PREDICTIVE ANALYTICS

- ▶ CREATING A DATA MINING PROJECT OF TELECOMMUNICATION DATASET FOR PREDICTING CHAID CHURN USING IBM SPSS MODELER.

STEP 1:



STEP 2:

The screenshot shows the IBM SPSS Modeler interface with a Stream titled "churn".

The Stream contains the following nodes:

- A "Table" node connected to the "churn" source.
- A "Filter" node connected to the "Table" node.
- A "Flat File" node connected to the "Filter" node.

The "churn" source is highlighted in yellow. The "Annotations" tab is selected in the top-left corner of the interface.

The bottom navigation bar includes icons for Favorites, Sources, Record Ops, Field Ops, Graphs, Modeling, Output, Export, IBM SPSS Statistics, Python, and Spark.

The status bar at the bottom indicates "542MB / 522MB" and "99%".

customer_id	data_usage	gender	age	tariff	dropped_calls	handset	year_mins
1	KL00019	yes	Male	4...	CAT 50	1,000 SCB10	36,121
2	KL00020	yes	Male	2...	CAT 50	0,000 SCB10	19,437
3	KL00025	yes	Male	3...	CAT 50	2,000 SCB20	72,650
4	KL00045	yes	Male	2...	CAT 50	2,000 SCB10	72,650
5	KL00050	yes	Male	4...	CAT 50	0,000 SCB10	40,698
6	KL00069	yes	Male	2...	CAT 50	1,000 SCB10	46,260
7	KL00070	yes	Male	3...	CAT 50	1,000 SCB20	54,270
8	KL00089	yes	Male	2...	CAT 50	0,000 SCB20	51,043
9	KL00098	yes	Male	3...	CAT 50	1,000 SCB20	99,090
10	KL00100	yes	Male	4...	CAT 50	0,000 SCB20	45,450
11	KL00110	yes	Male	3...	CAT 50	2,000 SCB20	45,450
12	KL00120	yes	Male	3...	CAT 50	0,000 SCB20	68,123
13	KL00130	yes	Male	4...	CAT 50	2,000 SCB20	63,352
14	KL00149	yes	Male	2...	CAT 50	1,000 SCB20	41,447
15	KL00150	yes	Male	3...	CAT 50	2,000 SCB20	94,090
16	KL00160	yes	Male	2...	CAT 50	0,000 SCB20	41,342
17	KL00170	yes	Male	3...	CAT 50	1,000 SCB10	100,200
18	KL00180	yes	Male	3...	CAT 50	1,000 SCB10	41,910
19	KL00190	yes	Male	3...	CAT 50	2,000 SCB10	58,792
20	KL00200	yes	Male	4...	CAT 50	0,000 SCB10	54,400

STEP 3:

The screenshot shows the SPSS Modeler interface. On the left, a data table window displays a dataset with 23 fields and 9,994 records. The columns include customer_id, data_known, gender, age, tariff, dropped_calls, handset, and peak_mins. A green arrow points from the 'Table' icon in the toolbar to the table icon in a process step. The process step itself has a list of six tables. The bottom of the screen shows the SPSS Modeler ribbon with various tabs like Favorites, Sources, Record Ops, Field Ops, etc., and a status bar indicating 'Server: Local Server... Records In: 25,544 Records Out: 43,456'.

STEP 4:

The screenshot displays the KNIME Analytics Platform interface, illustrating the fourth step in a data science process. The interface is divided into several panels:

- Left Panel:** A table viewer titled "Table (21 fields, 10,001 records)" showing a sample of data from a dataset. The columns include customer_id, data_known, gender, age, tariff, dropped_calls, handset, and peak_mins.
- Middle Panel:** A workflow canvas showing a sequence of nodes:
 - An input node (blue hexagon) connected to a "Tools" node (grid icon).
 - The "Tools" node connects to a "Filter" node (hexagon with a minus sign).
 - The "Filter" node connects to a "Flat File" output node (document icon).
- Right Panel:** A tree view titled "CRISP-DM Classes" showing the project structure:
 - (unSaved project)
 - Business Understanding
 - Data Understanding
 - Data Preparation
 - Modeling
 - Evaluation
 - Deployment
- Bottom Navigation:** A toolbar with various icons for different operations like Statistics, Sources, Record Ops, etc., and a status bar indicating "Server: Local Server" and memory usage "190MB / 173MB".

STEP 5:

The screenshot shows the IBM SPSS Modeler interface. At the top, there is a table viewer window titled "Table (2.3 Fields, 4,518 records)" displaying a dataset with columns: customer_id, data_known, gender, age, tariff, dropped_calls, handset, and peak_mins. The data includes various entries such as M376450, yes, Female, 4..., CAT 50, 6.000 CAS30, 136.200, etc. Below the table viewer is a toolbar with icons for File, Edit, Generate, and search.

The main workspace displays a data flow diagram. A "Table" node is connected to a "Filter" node, which then connects to a "Flat File" node. The "Flat File" node has an arrow pointing to a "Table" icon in the results pane at the bottom left. The results pane also shows five other "Table" icons.

The bottom navigation bar includes tabs for Favorites, Sources, Record Ops, Field Ops, Graphs, Modeling, Output, Export, IBM SPSS Statistics, Python, and Spark. It also features a toolbar with icons for Database, Var. File, Auto Data Prep, Select, Sample, Aggregate, Derive, Type, Filter, Graphboard, Auto Classifier, Auto Numeric, Auto Cluster, Table, Flat File, and Database. The status bar at the bottom indicates "Server: local Server" and "Uncoupling".

STEP 6:

The screenshot shows the IBM SPSS Modeler interface during the modeling phase of a data mining project.

Left Panel (Data View): A table view showing 23 fields and 31,789 records. The columns include customer_id, data_known, gender, age, tariff, dropped_calls, handset, and peak_mins. A specific row is highlighted with a yellow selection bar, showing values for customer_id K100010 through K100200.

Middle Panel (Flow View): A process flow diagram. It starts with two CHAID nodes (labeled "churn") connected to a "Table" output. The outputs from these nodes merge into a "Filter" node, which then connects to a "Flat File" output. Below the flow, there are five "Table" icons labeled "Table".

Right Panel (Project View): The "Streams" and "Outputs" sections show a list of tables and their details. The "CRISP-DM Classes" section indicates the project is at the "Business Understanding" stage.

Bottom Bar: A navigation bar with various icons and tabs: Favorites, Sources, Record Ops, Field Ops, Graphs, Modeling, Output, Export, IBM SPSS Statistics, Python, Spark, Database, Var File, Auto Data Prep, Select, Sample, Aggregate, Derive, Type, Filter, Graphboard, Auto Classifier, Auto Numeric, Auto Cluster, Table, Flat File, and Database.