

BABU BANARASI DAS UNIVERSITY



Predictive Analytics ON

Super-store-ship-Churn

**SUBMITTED TO:
Mr. VIKAS SIR
Roll No: 1230258342**

**SUBMITTED BY:
Name: Rishabh Singh**

Project Overview

This project focuses on analyzing the shipment behavior of customers using the Sample Superstore dataset. The primary objective is to understand how different factors — such as customer segment, region, category, and order priority — influence the mode of shipment. Using IBM SPSS Modeler, a CHAID (Chi-squared Automatic Interaction Detection) model is built to classify and predict the most likely shipping method chosen by customers.

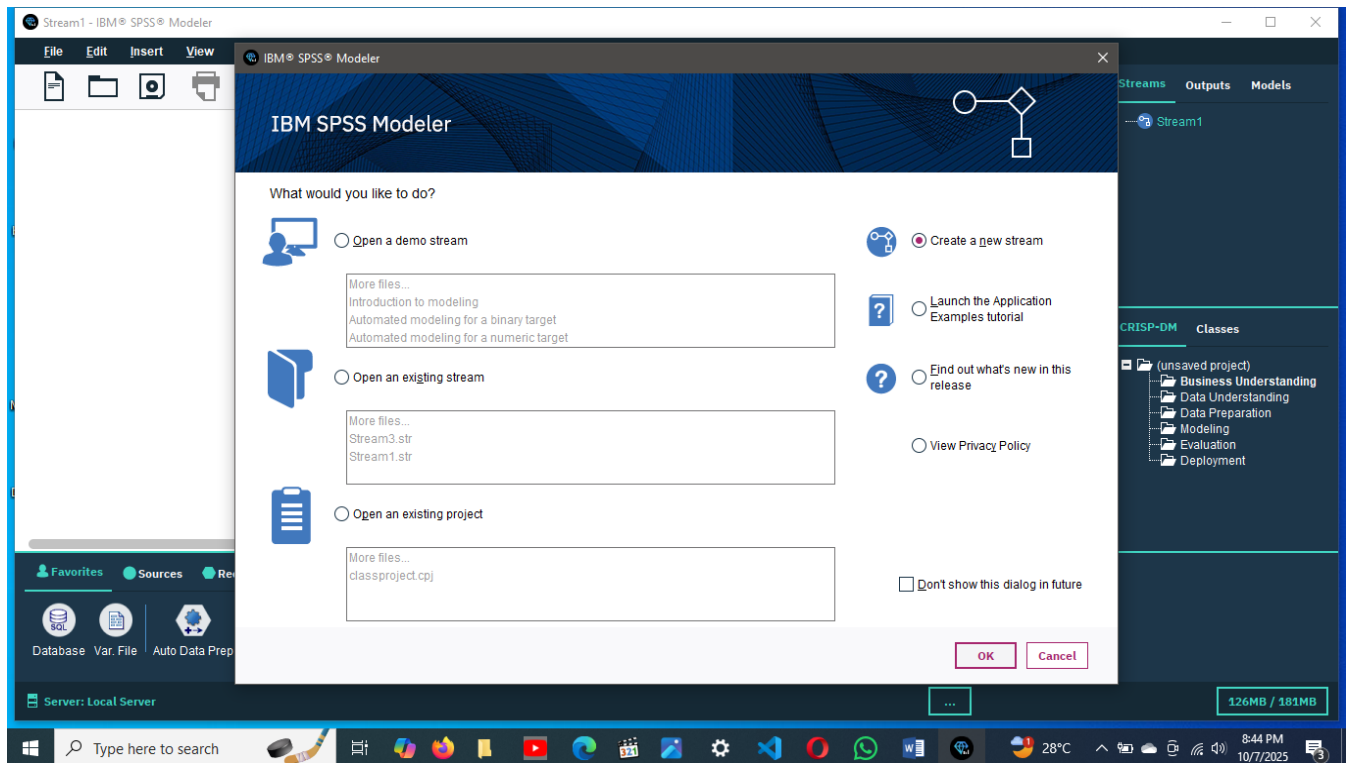
Problem Statement

The project aims to analyze product shipment patterns to predict the ship mode used by customers. Using IBM SPSS Modeler, a CHAID classification model is developed to identify key factors influencing shipping decisions. Here, “Second Class” shipments are considered potential churn cases due to delayed or unsatisfactory delivery experiences. The goal is to help the business improve logistics efficiency and reduce churn by optimizing shipping strategies.

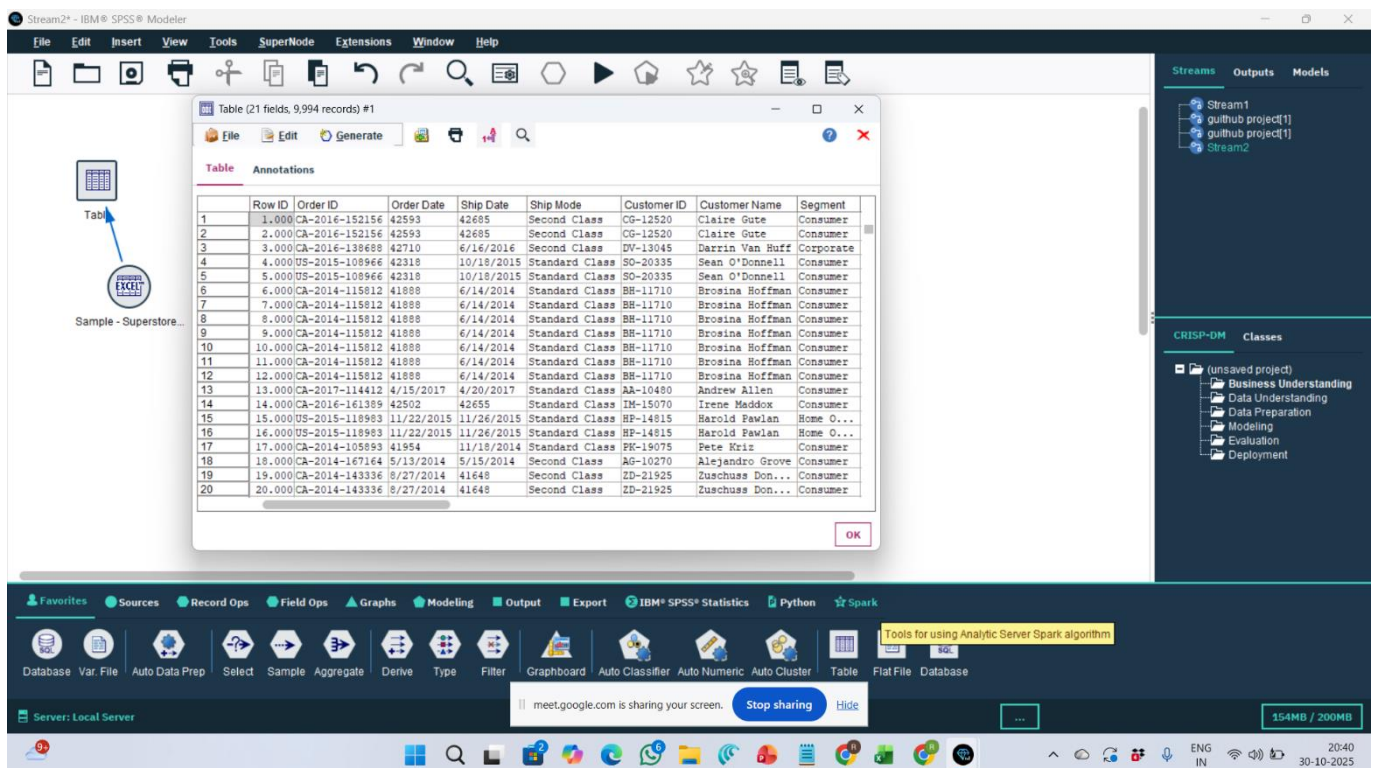
Tools & Technologies Used Category Tools / Technologies Purpose

Category	Tools / Technologies	Purpose
Data Source	Microsoft Excel (Sample Superstore Dataset)	Used for importing and managing raw shipment data
Data Analysis & Modeling	IBM SPSS Modeler	For building the CHAID classification model and performing data preprocessing
Algorithm	CHAID (Chi-squared Automatic Interaction Detection)	Used to classify and predict the shipping mode and identify key influencing factors
Visualization	IBM SPSS Charts & Decision Tree Output	For visual representation of model results and relationships between predictors
Documentation	GitHub & Markdown	Used to host the project and create structured documentation (README)

Step 1 LET's Start the Model



Step 2: From the palette, drag and drop the 'Excel. File' node under the 'Sources' tab to import a dataset (for example, a Excel file containing super-store data).



Double-click the node and browse to select your dataset. Click OK to load it.

Step 3: Next, from the 'Record Ops' tab, drag and drop the 'Select' node. Connect it to the data import node.

The screenshot shows the IBM SPSS Modeler interface. A workflow diagram on the left shows a 'Sample - Superstore...' node connected to a '(generated)' node, which is then connected to a 'Type' node. A 'Type' dialog box is open in the center, displaying a table of field properties. The 'Types' tab is selected, showing a list of fields with their measurement types, values, missing values, check boxes, and roles. The 'Record Ops' tab is active in the bottom toolbar.

Field	Measurement	Values	Missing	Check	Role
Row ID	Continuous	11 0.9994 0		None	Input
Order ID	Typeless			None	None
Order Date	Typeless			None	None
Ship Date	Typeless			None	None
Ship Mode	Nominal	*First Cla...		None	Target
Customer ID	Typeless			None	None
Customer N...	Typeless			None	None
Segment	Flag	Consume...		None	Input
Country	Flag	*United St...		None	Input
City	Typeless			None	None
State	Nominal	Alabama		None	Input

Step 4: Double-click and set a churn condition to select specific records

The screenshot shows the IBM SPSS Modeler interface. A workflow diagram on the left shows a 'Sample - Superstore...' node connected to a '(generated)' node, which is then connected to a 'Type' node. The 'Type' node is now connected to a 'Ship Mode' node. The 'Modeling' tab is active in the bottom toolbar.

Step 5 and find the avg profit in sales

The screenshot shows the IBM SPSS Modeler interface. On the left, a workflow diagram includes nodes for 'Sample - Superstore...', '(generated)', 'Type', 'Ship Mode', 'Average_Profit', 'Profit & Loss', and 'Reclassify'. A central window displays a table with 22 fields and 5,191 records. The table has columns: Sales, Quantity, Discount, Profit, and Average_Profit. The right sidebar shows 'Streams' (Stream1, Stream2, Stream3) and 'CRISP-DM Classes' (Business Understanding, Data Understanding, Data Preparation, Modeling, Evaluation, Deployment). The bottom status bar indicates 'Server: Local Server' and '177MB / 224MB'.

		Sales	Quantity	Discount	Profit	Average_Profit
1	Bookcase	261.960	2.000	0.000	41.914	109.797
2	erced Stacking Chairs, Rounded Back	731.940	3.000	0.000	219.582	1607.208
3	lim Rectangular Table	957.577	5.000	0.450	-383.031	-3667.819
4	System	22.368	2.000	0.200	2.516	0.563
5	nd Plastic Desk Accessories, Cherry Wood	48.860	7.000	0.000	14.169	6.923
6		7.280	4.000	0.000	1.966	0.143
7	phone	907.152	6.000	0.200	90.715	822.925
8	ath Locking Rings by Sameill	18.504	3.000	0.200	5.782	1.070
9	et Surge	114.900	5.000	0.000	34.470	39.606
10	onference Tables	1706.184	9.000	0.200	85.309	1455.532
11	hone - Charcoal black	911.424	4.000	0.200	68.357	623.020
12		15.552	3.000	0.200	5.443	0.847
13	omb Binding Machine	407.976	3.000	0.200	132.592	540.944
14	rtical 5-Shelf: 72"H x 36"W x 18 1/2"D	665.880	6.000	0.000	13.318	88.679
15	er	55.500	2.000	0.000	9.990	5.544
16		8.560	2.000	0.000	2.482	0.212
17		213.480	3.000	0.200	16.011	34.180
18	w Binder, White, 1"	22.720	4.000	0.200	7.384	1.678
19	hair, Gray	71.372	2.000	0.300	-1.020	-0.728
20	lim Rectangular Table	1044.630	3.000	0.000	240.265	2509.879

Step 6: Find the profit and loss statement for the data set. Define that a negative value (–) in the avg_profit column indicates a loss statement.

The screenshot shows the IBM SPSS Modeler interface. The workflow diagram is similar to the previous one, but the 'Reclassify' node is highlighted. A central window displays a table with 23 fields and 5,191 records. The table has columns: Sales, Quantity, Discount, Profit, Average_Profit, and Profit & Loss. The right sidebar shows 'Streams' and 'CRISP-DM Classes'. The bottom status bar indicates 'Server: Local Server' and '180MB / 224MB'.

		Sales	Quantity	Discount	Profit	Average_Profit	Profit & Loss
1		261.960	2.000	0.000	41.914	109.797	Profit
2	Chairs, Rounded Back	731.940	3.000	0.000	219.582	1607.208	Profit
3	Table	957.577	5.000	0.450	-383.031	-3667.819	Loss
4		22.368	2.000	0.200	2.516	0.563	Profit
5	Accessories, Cherry Wood	48.860	7.000	0.000	14.169	6.923	Profit
6		7.280	4.000	0.000	1.966	0.143	Profit
7		907.152	6.000	0.200	90.715	822.925	Profit
8	ge by Sameill	18.504	3.000	0.200	5.782	1.070	Profit
9		114.900	5.000	0.000	34.470	39.606	Profit
10	s	1706.184	9.000	0.200	85.309	1455.532	Profit
11	s black	911.424	4.000	0.200	68.357	623.020	Profit
12		15.552	3.000	0.200	5.443	0.847	Profit
13	hine	407.976	3.000	0.200	132.592	540.944	Profit
14	72"H x 36"W x 18 1/2"D	665.880	6.000	0.000	13.318	88.679	Profit
15		55.500	2.000	0.000	9.990	5.544	Profit
16		8.560	2.000	0.000	2.482	0.212	Profit
17		213.480	3.000	0.200	16.011	34.180	Profit
18	, 1"	22.720	4.000	0.200	7.384	1.678	Profit
19		71.372	2.000	0.300	-1.020	-0.728	Loss
20	Table	1044.630	3.000	0.000	240.265	2509.879	Profit

Step 7: The report shows desirable sales using the Re-Classify method.

Desirable States

Settings Annotations

Mode: ☒ Single ☐ Multiple

Reclassify into: ☒ New field ☐ Existing field

Reclassify field: State

New field name: Desirable States

Reclassify values:

Get Copy Clear new Auto...

Original value	New value
California	Uttar Pradesh
Florida	Delhi NCR
Utah	Biharlii's

For unspecified values use: ☒ Original value ☐ Default value undef

OK Cancel Apply Reset

Step 8: The report shows desirable sales. You can see the changes in city names from California to Uttar Pradesh before and after.

Table (24 fields, 5,191 records) #1

	Sales	Quantity	Discount	Profit	Average Profit	Profit & Loss	Desirable States
1	261.960	2.000	0.000	41.914	109.797	Profit	Kentucky
2	731.940	3.000	0.000	219.582	1607.208	Profit	Kentucky
3	957.577	5.000	0.450	-383.031	-3667.819	Loss	Delhi NCR
4	22.368	2.000	0.200	2.516	0.563	Profit	Delhi NCR
5	40.860	7.000	0.000	14.169	6.923	Profit	Uttar Pradesh
6	7.280	4.000	0.000	1.966	0.143	Profit	Uttar Pradesh
7	907.152	6.000	0.200	90.715	822.925	Profit	Uttar Pradesh
8	18.504	3.000	0.200	5.762	1.070	Profit	Uttar Pradesh
9	114.900	5.000	0.000	34.470	39.606	Profit	Uttar Pradesh
10	1706.194	9.000	0.200	85.309	1455.932	Profit	Uttar Pradesh
11	911.424	4.000	0.200	68.357	623.020	Profit	Uttar Pradesh
12	15.552	3.000	0.200	5.443	0.847	Profit	North Carolina
13	407.976	3.000	0.200	132.592	540.944	Profit	Washington
14	665.880	6.000	0.000	13.318	88.679	Profit	Wisconsin
15	55.500	2.000	0.000	9.990	5.544	Profit	Biharlii's
16	8.560	2.000	0.000	2.482	0.212	Profit	Uttar Pradesh
17	213.480	3.000	0.200	16.011	34.180	Profit	Uttar Pradesh
18	22.720	4.000	0.200	7.384	1.678	Profit	Uttar Pradesh
19	71.372	2.000	0.300	-1.020	-0.728	Loss	Pennsylvania
20	1044.630	3.000	0.000	240.265	2509.879	Profit	Biharlii's

Step 9: The report shows metrics of profit and loss.

