

**BABU BANARASI DAS UNIVERSITY**



## **CASE STUDY**

**ON**

**Predictive Analysis of Customer Purchase Behavior in Online Retail Using SPSS.**

**SUBMITTED TO:**

**Mr. Robin Tygai**

**SUBMITTED BY:**

**Name: Deeksha Singh**

**Roll No: 1230258150**

**Name: Janwi Kumari**

**Roll No: 1230258205**

**Name: Kirti Srivastava**

**Roll No: 1230258220**

**Section: BCADS33**

### Platform: Kaggle

[illegible][illegible][illegible]

# CASE STUDY

**Agenda/Definition:** We work as a data miner for an online retail firm, where we have to clean, aggregate and transform the data to identify high-value customers and prepare the dataset for further modeling.

**Learning:** We will learn how to:

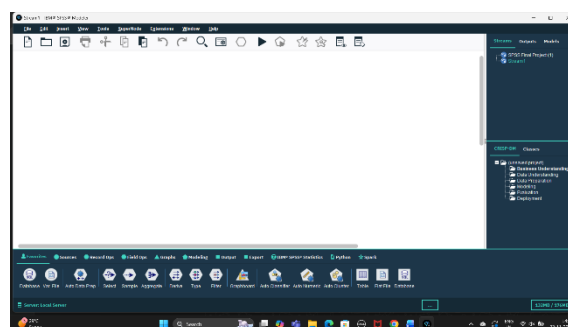
- Clean the dataset
- Aggregate the data
- Transform the data into relevant data

**Required Tool:** IBM SPSS Modeler 18.6

**Dataset Used:** Online Retail II Dataset (.xlsx)

**Working:** This project analyzes customer purchasing patterns in an online retail store, where the aim is to identify high value customers and prepare the dataset for further modeling. To achieve this, follow the given steps:

**STEP 1:** Open IBM SPSS Modeler 18.6 in your PC.



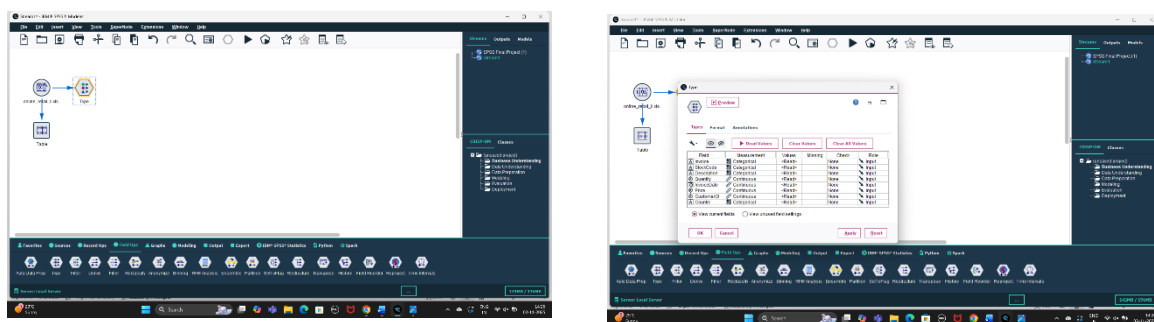
**STEP 2:** From the Sources Category, import the dataset named Online Retail II Dataset (.xlsx). This is an Excel dataset so, select the Excel Node.



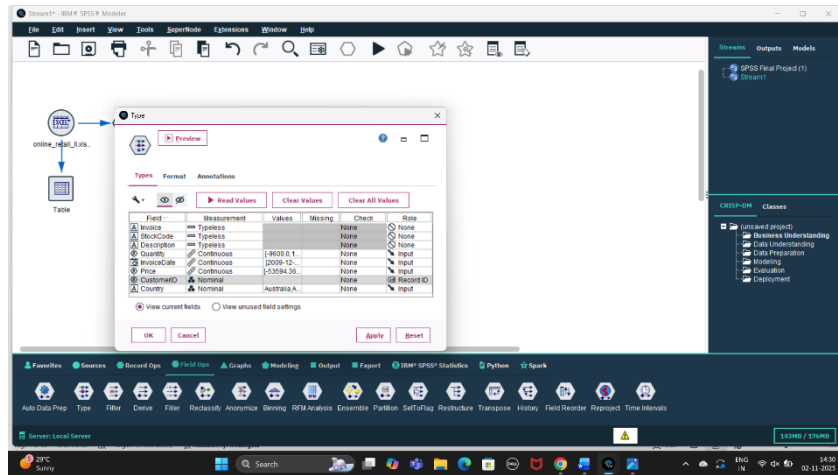
**STEP 3:** Connect a Table from Output Category to Preview the dataset.

Invoice	StockCode	Description	Quantity	InvoiceDate	Price	CustomerID	Country
1	888286	100 CHOCOLATE GLAZED ROLL 20 SLICETS	12.00	2009-12-01 07:45:00	4.450	13005	United Kingdom
2	888284	1752109 FISH CHIPS	12.00	2009-12-01 07:45:00	6.750	13005	United Kingdom
3	888284	1752109 FISH CHIPS	40.00	2009-12-01 07:45:00	8.750	13005	United Kingdom
4	888284	100044 BAKED PASTRY - CHOCOLATE FILL	40.00	2009-12-01 07:45:00	2.100	13005	United Kingdom
5	888284	100044 BAKED PASTRY - CHOCOLATE FILL	24.00	2009-12-01 07:45:00	1.450	13005	United Kingdom
6	888284	100044 BAKED PASTRY - CHOCOLATE FILL	24.00	2009-12-01 07:45:00	1.450	13005	United Kingdom
7	888284	100044 BAKED PASTRY - CHOCOLATE FILL	24.00	2009-12-01 07:45:00	1.450	13005	United Kingdom
8	888284	100044 BAKED PASTRY - CHOCOLATE FILL	24.00	2009-12-01 07:45:00	1.450	13005	United Kingdom
9	888284	100044 BAKED PASTRY - CHOCOLATE FILL	24.00	2009-12-01 07:45:00	1.450	13005	United Kingdom
10	888284	100044 BAKED PASTRY - CHOCOLATE FILL	24.00	2009-12-01 07:45:00	1.450	13005	United Kingdom
11	888284	100044 BAKED PASTRY - CHOCOLATE FILL	24.00	2009-12-01 07:45:00	1.450	13005	United Kingdom
12	888284	100044 BAKED PASTRY - CHOCOLATE FILL	24.00	2009-12-01 07:45:00	1.450	13005	United Kingdom
13	888284	100044 BAKED PASTRY - CHOCOLATE FILL	24.00	2009-12-01 07:45:00	1.450	13005	United Kingdom
14	888284	100044 BAKED PASTRY - CHOCOLATE FILL	24.00	2009-12-01 07:45:00	1.450	13005	United Kingdom
15	888284	100044 BAKED PASTRY - CHOCOLATE FILL	24.00	2009-12-01 07:45:00	1.450	13005	United Kingdom
16	888284	100044 BAKED PASTRY - CHOCOLATE FILL	24.00	2009-12-01 07:45:00	1.450	13005	United Kingdom
17	888284	100044 BAKED PASTRY - CHOCOLATE FILL	24.00	2009-12-01 07:45:00	1.450	13005	United Kingdom
18	888284	100044 BAKED PASTRY - CHOCOLATE FILL	24.00	2009-12-01 07:45:00	1.450	13005	United Kingdom
19	888284	100044 BAKED PASTRY - CHOCOLATE FILL	24.00	2009-12-01 07:45:00	1.450	13005	United Kingdom
20	888284	100044 BAKED PASTRY - CHOCOLATE FILL	24.00	2009-12-01 07:45:00	1.450	13005	United Kingdom
21	888284	100044 BAKED PASTRY - CHOCOLATE FILL	24.00	2009-12-01 07:45:00	1.450	13005	United Kingdom
22	888284	100044 BAKED PASTRY - CHOCOLATE FILL	24.00	2009-12-01 07:45:00	1.450	13005	United Kingdom
23	888284	100044 BAKED PASTRY - CHOCOLATE FILL	24.00	2009-12-01 07:45:00	1.450	13005	United Kingdom
24	888284	100044 BAKED PASTRY - CHOCOLATE FILL	24.00	2009-12-01 07:45:00	1.450	13005	United Kingdom
25	888284	100044 BAKED PASTRY - CHOCOLATE FILL	24.00	2009-12-01 07:45:00	1.450	13005	United Kingdom
26	888284	100044 BAKED PASTRY - CHOCOLATE FILL	24.00	2009-12-01 07:45:00	1.450	13005	United Kingdom
27	888284	100044 BAKED PASTRY - CHOCOLATE FILL	24.00	2009-12-01 07:45:00	1.450	13005	United Kingdom
28	888284	100044 BAKED PASTRY - CHOCOLATE FILL	24.00	2009-12-01 07:45:00	1.450	13005	United Kingdom
29	888284	100044 BAKED PASTRY - CHOCOLATE FILL	24.00	2009-12-01 07:45:00	1.450	13005	United Kingdom
30	888284	100044 BAKED PASTRY - CHOCOLATE FILL	24.00	2009-12-01 07:45:00	1.450	13005	United Kingdom
31	888284	100044 BAKED PASTRY - CHOCOLATE FILL	24.00	2009-12-01 07:45:00	1.450	13005	United Kingdom
32	888284	100044 BAKED PASTRY - CHOCOLATE FILL	24.00	2009-12-01 07:45:00	1.450	13005	United Kingdom
33	888284	100044 BAKED PASTRY - CHOCOLATE FILL	24.00	2009-12-01 07:45:00	1.450	13005	United Kingdom
34	888284	100044 BAKED PASTRY - CHOCOLATE FILL	24.00	2009-12-01 07:45:00	1.450	13005	United Kingdom
35	888284	100044 BAKED PASTRY - CHOCOLATE FILL	24.00	2009-12-01 07:45:00	1.450	13005	United Kingdom
36	888284	100044 BAKED PASTRY - CHOCOLATE FILL	24.00	2009-12-01 07:45:00	1.450	13005	United Kingdom
37	888284	100044 BAKED PASTRY - CHOCOLATE FILL	24.00	2009-12-01 07:45:00	1.450	13005	United Kingdom
38	888284	100044 BAKED PASTRY - CHOCOLATE FILL	24.00	2009-12-01 07:45:00	1.450	13005	United Kingdom
39	888284	100044 BAKED PASTRY - CHOCOLATE FILL	24.00	2009-12-01 07:45:00	1.450	13005	United Kingdom
40	888284	100044 BAKED PASTRY - CHOCOLATE FILL	24.00	2009-12-01 07:45:00	1.450	13005	United Kingdom
41	888284	100044 BAKED PASTRY - CHOCOLATE FILL	24.00	2009-12-01 07:45:00	1.450	13005	United Kingdom
42	888284	100044 BAKED PASTRY - CHOCOLATE FILL	24.00	2009-12-01 07:45:00	1.450	13005	United Kingdom
43	888284	100044 BAKED PASTRY - CHOCOLATE FILL	24.00	2009-12-01 07:45:00	1.450	13005	United Kingdom
44	888284	100044 BAKED PASTRY - CHOCOLATE FILL	24.00	2009-12-01 07:45:00	1.450	13005	United Kingdom
45	888284	100044 BAKED PASTRY - CHOCOLATE FILL	24.00	2009-12-01 07:45:00	1.450	13005	United Kingdom
46	888284	100044 BAKED PASTRY - CHOCOLATE FILL	24.00	2009-12-01 07:45:00	1.450	13005	United Kingdom
47	888284	100044 BAKED PASTRY - CHOCOLATE FILL	24.00	2009-12-01 07:45:00	1.450	13005	United Kingdom
48	888284	100044 BAKED PASTRY - CHOCOLATE FILL	24.00	2009-12-01 07:45:00	1.450	13005	United Kingdom
49	888284	100044 BAKED PASTRY - CHOCOLATE FILL	24.00	2009-12-01 07:45:00	1.450	13005	United Kingdom
50	888284	100044 BAKED PASTRY - CHOCOLATE FILL	24.00	2009-12-01 07:45:00	1.450	13005	United Kingdom

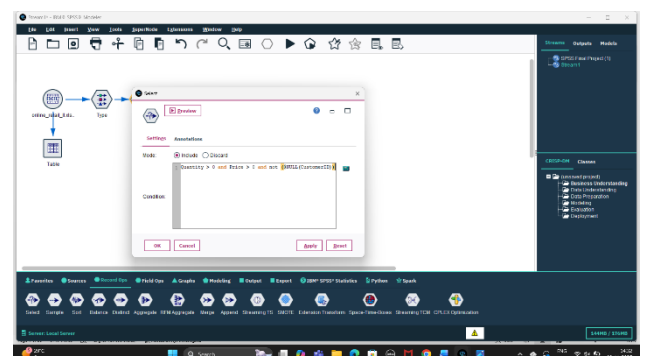
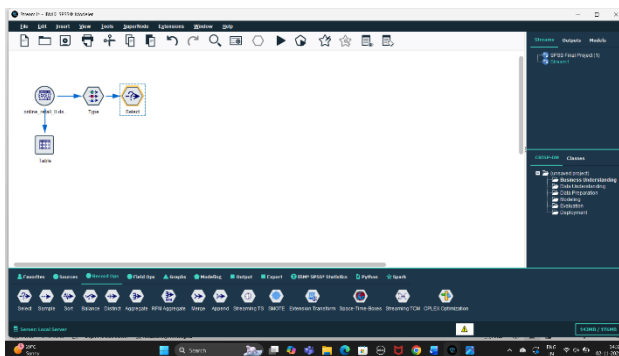
**STEP 4:** Connect the Type Node from Field Ops. This Node aids in understanding the different Fields Measurement Values.



**STEP 5:** Rea the values to instantiate the different fields values. Now we will also change the role of some fields depending on our requirement. After the changes made, click on Apply and then on OK.



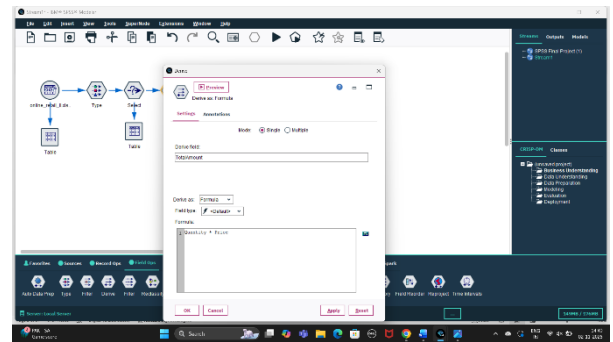
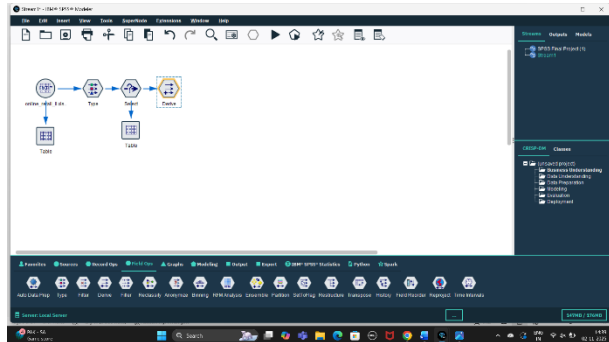
**STEP 6:** Connect A Select Node from Field Ops. This node allows us to include or discard the records. Type the following: Quantity > 0 and Price > 0 and not (@NULL(CustomerID)).



**STEP 7:** Connect a Table to see what changes your Select Node brings.

Field	Measurement	Values	Missing	Check	Role
Invoice	Typeless	None	None	<input checked="" type="checkbox"/>	None
StockCode	Typeless	None	None	<input checked="" type="checkbox"/>	None
Description	Typeless	None	None	<input checked="" type="checkbox"/>	None
Quantity	Continuous	1.90501	None	<input checked="" type="checkbox"/>	None
InvoiceDate	Continuous	2009-12	None	<input checked="" type="checkbox"/>	None
Price	Continuous	120094.30	None	<input checked="" type="checkbox"/>	None
CustomerID	Nominal	None	None	<input checked="" type="checkbox"/>	None
Country	Nominal	Australia	None	<input checked="" type="checkbox"/>	None

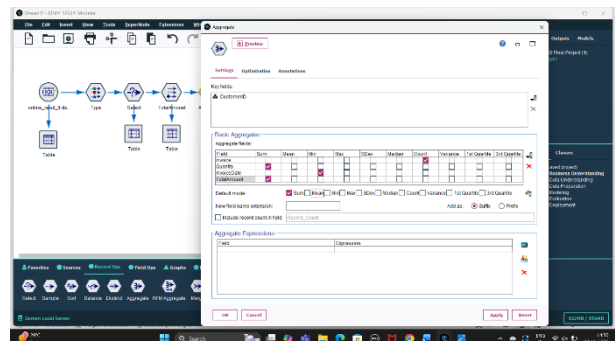
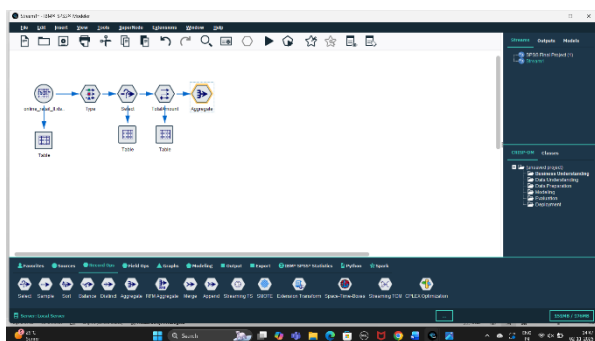
**STEP 8:** So now, we will create an additional field named TotalAmount. To create a new field, we use Derive Node from Field Ops. Double click on the node and from the expression builder choose the formula: price\*quantity. Click on Apply and OK.



**STEP 9:** Connect a Table to see your new Field.

Invoice	InvoiceDate	Customer	Quantity	Price	TotalAmount	Quantity
1000001	2010-01-01	Customer A	10	100	1000	10
1000002	2010-01-02	Customer B	20	200	4000	20
1000003	2010-01-03	Customer C	30	300	9000	30
1000004	2010-01-04	Customer D	40	400	16000	40
1000005	2010-01-05	Customer E	50	500	25000	50
1000006	2010-01-06	Customer F	60	600	36000	60
1000007	2010-01-07	Customer G	70	700	49000	70
1000008	2010-01-08	Customer H	80	800	64000	80
1000009	2010-01-09	Customer I	90	900	81000	90
1000010	2010-01-10	Customer J	100	1000	100000	100

**STEP 10:** From the Record Ops, select Aggregate Node. This node combines multiple records into one row. Select the Key field as CUSTOMER\_ID. In aggregate fields choose Invoice, InvoiceDate, TotalAmount, Quantity.



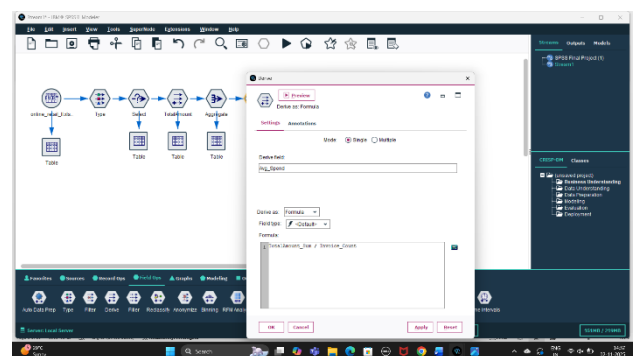
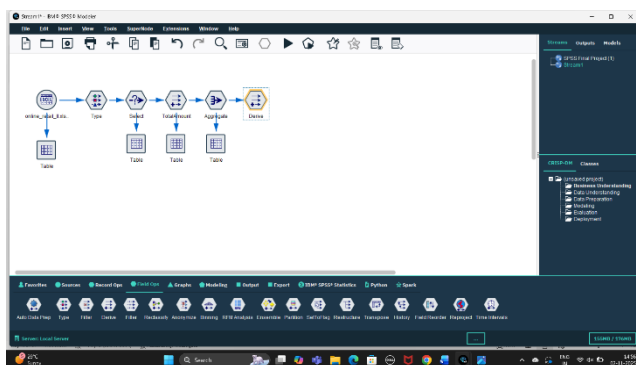


**STEP 11:** Connect a TABLE with this one as well.

[illegible]

**STEP 12:** Choose one more Derive node. This time we will use it to derive a field called AVG SPEND. From the Expression Builder choose the formula:

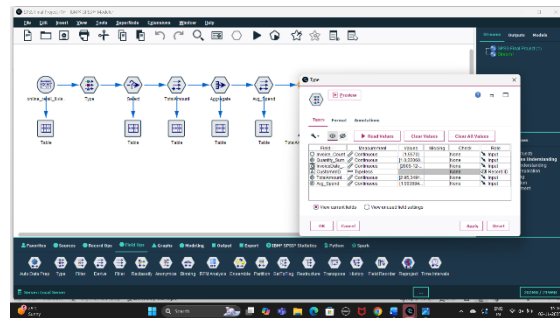
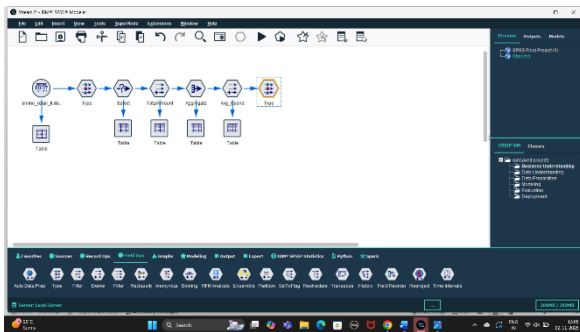
TotalAmount Sum / Invoice Count.



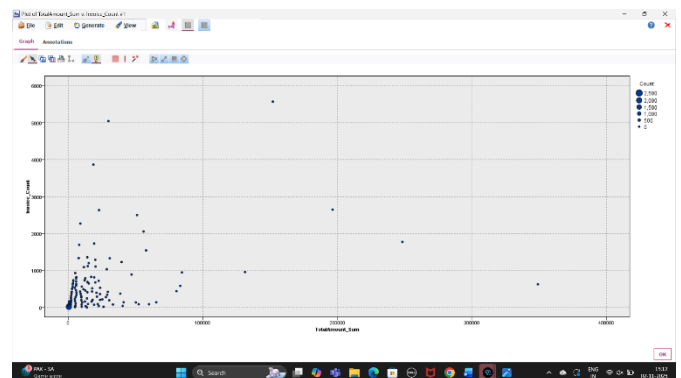
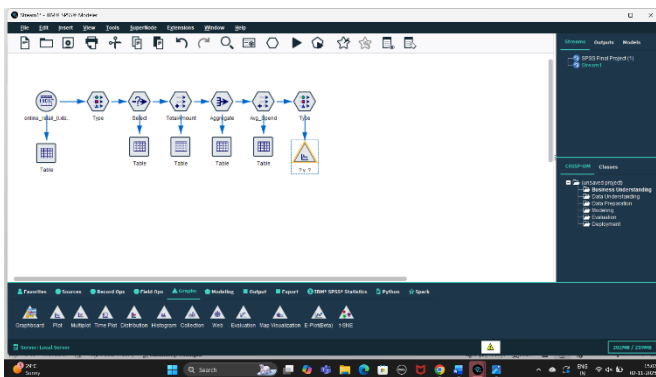
**STEP 13:** Connect a TABLE with this one also.

[illegible]

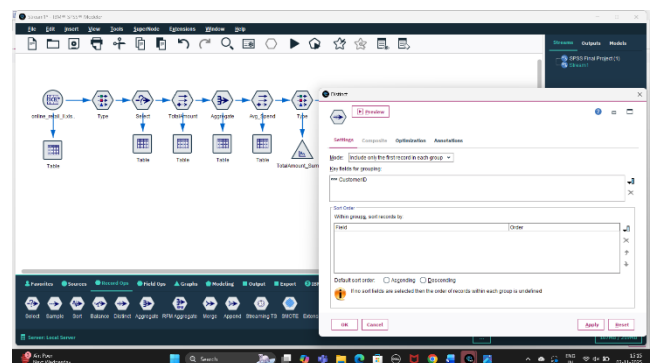
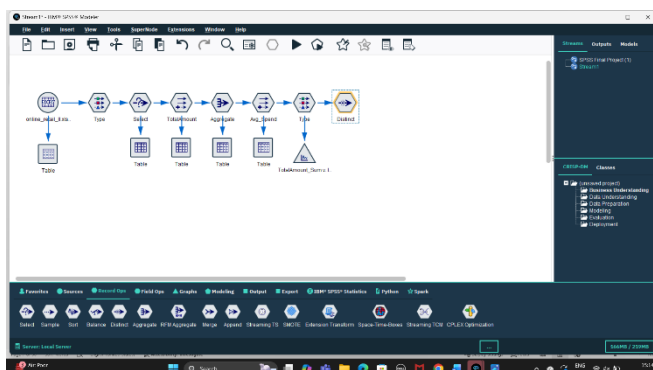
**STEP 14:** Connect the Type node. Read the values.



**STEP 15:** Connect a Plot graph from Graphs. Graphs aid in visualization and better understanding. Double click on it and give the X-axis and Y-axis TotalAmount\_Sum and Invoice\_Count respectively. Click on Apply and Run.



**STEP 16:** Select Distinct Node from Record Ops. Include only first record in each group should be chosen in Mode. In key fields for grouping choose CUSTOMER\_ID. Click on Apply and OK.



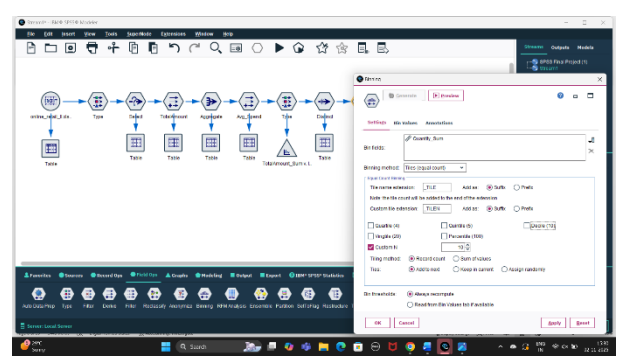
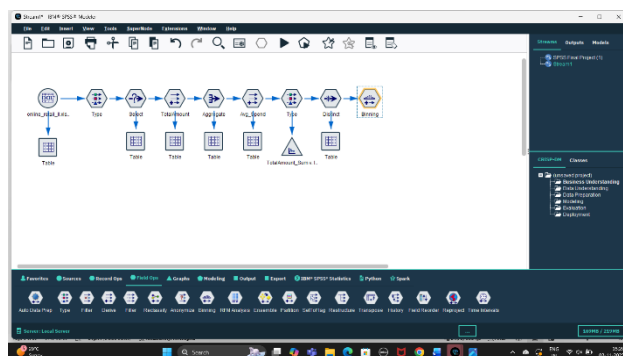


**STEP 17:** Connect a Table again to see how it works.

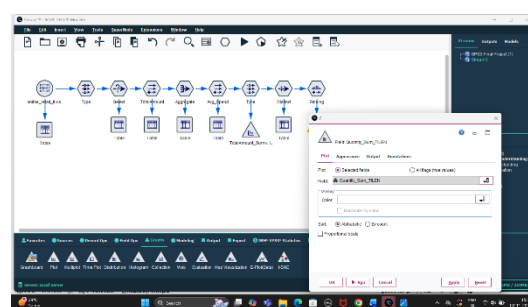
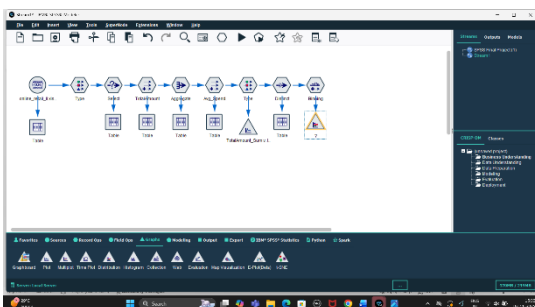
Index_Count	Quantity_Sum	TimeStamp	TimeStamp_Sum
1	100	2019-12-12 12:12:12	121212
2	200	2019-12-12 12:22:22	222222
3	300	2019-12-12 12:32:32	323232
4	400	2019-12-12 12:42:42	424242
5	500	2019-12-12 12:52:52	525252
6	600	2019-12-12 13:02:02	626262
7	700	2019-12-12 13:12:12	727272
8	800	2019-12-12 13:22:22	828282
9	900	2019-12-12 13:32:32	929292
10	1000	2019-12-12 13:42:42	103424
11	1100	2019-12-12 13:52:52	114535
12	1200	2019-12-12 14:02:02	125646
13	1300	2019-12-12 14:12:12	136757
14	1400	2019-12-12 14:22:22	147868
15	1500	2019-12-12 14:32:32	158979
16	1600	2019-12-12 14:42:42	170090
17	1700	2019-12-12 14:52:52	181201
18	1800	2019-12-12 15:02:02	192312
19	1900	2019-12-12 15:12:12	203423
20	2000	2019-12-12 15:22:22	214534

**STEP 18:** From Field Ops, choose Binning Node. This node bins the data into equal distributions.

Double click on this node, select Binning method as Tiles and in binning fields choose Quantity\_Sum.



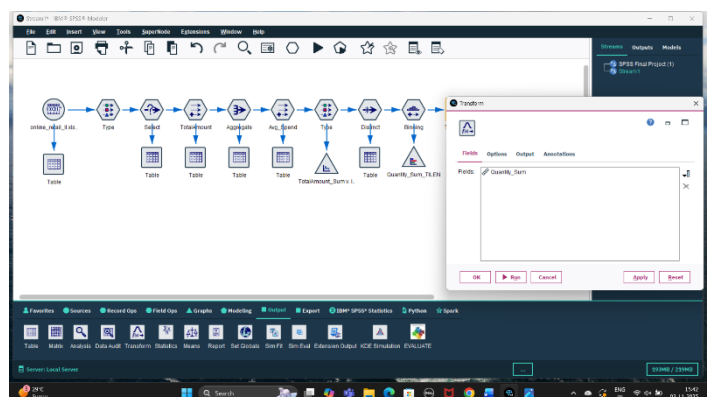
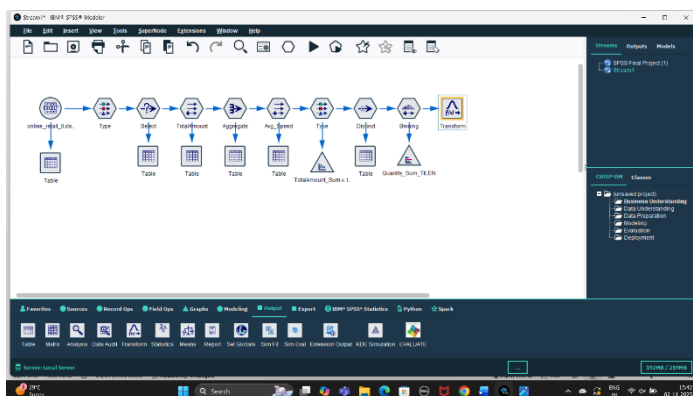
**STEP 19:** Choose Distribution node from Graphs. Choose Quantity\_Sum\_TILEN in files.



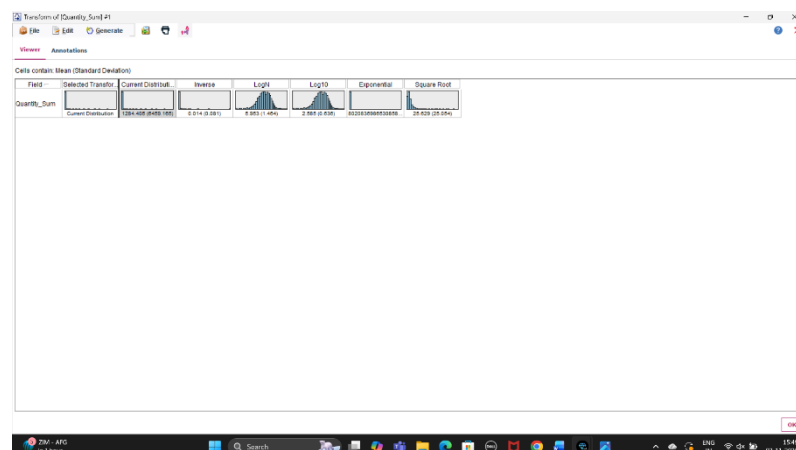
**STEP 20:** Click on Apply and Run.

Table	Proportion	Count
1	0.78	431
2	0.80	430
3	0.79	430
4	0.79	430
5	0.79	430
6	0.79	430
7	0.79	430
8	0.79	430
9	0.79	430
10	0.79	430

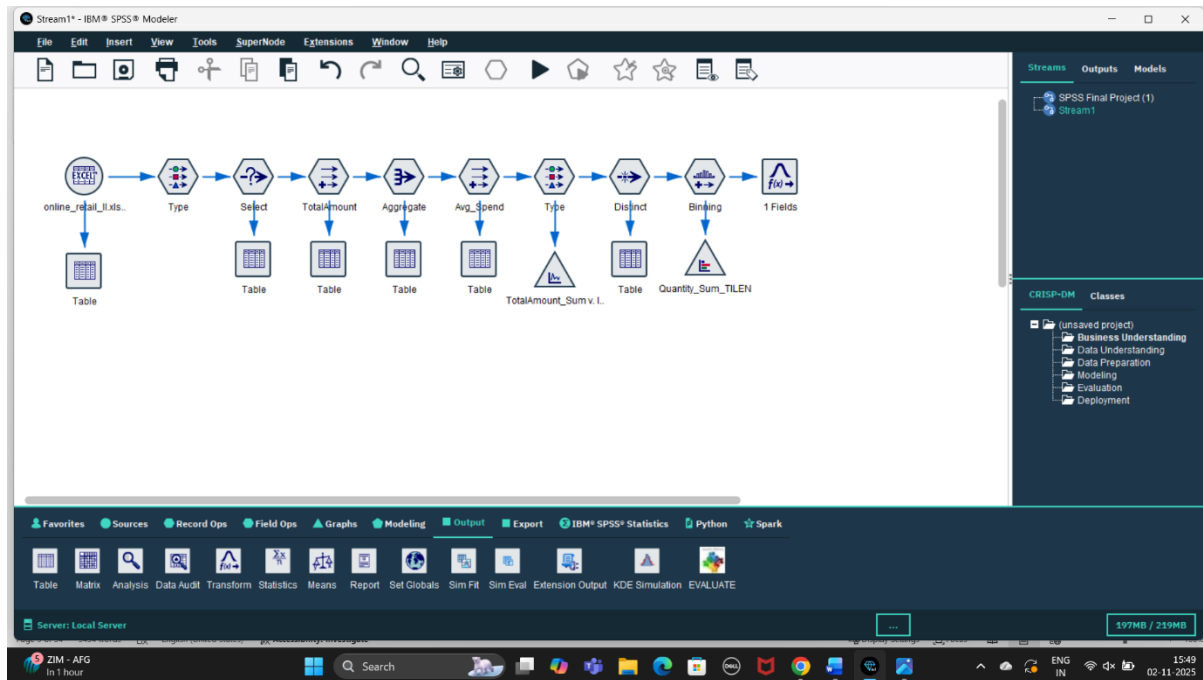
**STEP 21:** From Output Category choose Transform Node, this node will transform the data in a standard format. Double click on the node and choose Quantity\_Sum in the fields.



**STEP 22:** Click on Apply and Run.



**STEP 23:** The final flow of nodes looks like this:



This marks the end of our SPSS Case Study.

-----THANK YOU-----

