

# **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**

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

Platform: Kaggle

The image displays a 4x3 grid of 12 screenshots from a Microsoft Excel application running on a Windows operating system. The screenshots show various sheets of the 'Online Retail II' dataset, which is a large Excel file containing transactional data. The visible columns include Customer ID, Order ID, Description, and Quantity. The data spans multiple pages, indicating the large size of the dataset.

Customer ID	Order ID	Description	Quantity
1	1	DE 2014-01-05 001	1
2	2	DE 2014-01-05 002	1
3	3	DE 2014-01-05 003	1
4	4	DE 2014-01-05 004	1
5	5	DE 2014-01-05 005	1
6	6	DE 2014-01-05 006	1
7	7	DE 2014-01-05 007	1
8	8	DE 2014-01-05 008	1
9	9	DE 2014-01-05 009	1
10	10	DE 2014-01-05 010	1
11	11	DE 2014-01-05 011	1
12	12	DE 2014-01-05 012	1
13	13	DE 2014-01-05 013	1
14	14	DE 2014-01-05 014	1
15	15	DE 2014-01-05 015	1
16	16	DE 2014-01-05 016	1
17	17	DE 2014-01-05 017	1
18	18	DE 2014-01-05 018	1
19	19	DE 2014-01-05 019	1
20	20	DE 2014-01-05 020	1
21	21	DE 2014-01-05 021	1
22	22	DE 2014-01-05 022	1
23	23	DE 2014-01-05 023	1
24	24	DE 2014-01-05 024	1
25	25	DE 2014-01-05 025	1
26	26	DE 2014-01-05 026	1
27	27	DE 2014-01-05 027	1
28	28	DE 2014-01-05 028	1
29	29	DE 2014-01-05 029	1
30	30	DE 2014-01-05 030	1
31	31	DE 2014-01-05 031	1
32	32	DE 2014-01-05 032	1
33	33	DE 2014-01-05 033	1
34	34	DE 2014-01-05 034	1
35	35	DE 2014-01-05 035	1
36	36	DE 2014-01-05 036	1
37	37	DE 2014-01-05 037	1
38	38	DE 2014-01-05 038	1
39	39	DE 2014-01-05 039	1
40	40	DE 2014-01-05 040	1
41	41	DE 2014-01-05 041	1
42	42	DE 2014-01-05 042	1
43	43	DE 2014-01-05 043	1
44	44	DE 2014-01-05 044	1
45	45	DE 2014-01-05 045	1
46	46	DE 2014-01-05 046	1
47	47	DE 2014-01-05 047	1
48	48	DE 2014-01-05 048	1
49	49	DE 2014-01-05 049	1
50	50	DE 2014-01-05 050	1
51	51	DE 2014-01-05 051	1
52	52	DE 2014-01-05 052	1
53	53	DE 2014-01-05 053	1
54	54	DE 2014-01-05 054	1
55	55	DE 2014-01-05 055	1
56	56	DE 2014-01-05 056	1
57	57	DE 2014-01-05 057	1
58	58	DE 2014-01-05 058	1
59	59	DE 2014-01-05 059	1
60	60	DE 2014-01-05 060	1
61	61	DE 2014-01-05 061	1
62	62	DE 2014-01-05 062	1
63	63	DE 2014-01-05 063	1
64	64	DE 2014-01-05 064	1
65	65	DE 2014-01-05 065	1
66	66	DE 2014-01-05 066	1
67	67	DE 2014-01-05 067	1
68	68	DE 2014-01-05 068	1
69	69	DE 2014-01-05 069	1
70	70	DE 2014-01-05 070	1
71	71	DE 2014-01-05 071	1
72	72	DE 2014-01-05 072	1
73	73	DE 2014-01-05 073	1
74	74	DE 2014-01-05 074	1
75	75	DE 2014-01-05 075	1
76	76	DE 2014-01-05 076	1
77	77	DE 2014-01-05 077	1
78	78	DE 2014-01-05 078	1
79	79	DE 2014-01-05 079	1
80	80	DE 2014-01-05 080	1
81	81	DE 2014-01-05 081	1
82	82	DE 2014-01-05 082	1
83	83	DE 2014-01-05 083	1
84	84	DE 2014-01-05 084	1
85	85	DE 2014-01-05 085	1
86	86	DE 2014-01-05 086	1
87	87	DE 2014-01-05 087	1
88	88	DE 2014-01-05 088	1
89	89	DE 2014-01-05 089	1
90	90	DE 2014-01-05 090	1
91	91	DE 2014-01-05 091	1
92	92	DE 2014-01-05 092	1
93	93	DE 2014-01-05 093	1
94	94	DE 2014-01-05 094	1
95	95	DE 2014-01-05 095	1
96	96	DE 2014-01-05 096	1
97	97	DE 2014-01-05 097	1
98	98	DE 2014-01-05 098	1
99	99	DE 2014-01-05 099	1
100	100	DE 2014-01-05 100	1
101	101	DE 2014-01-05 101	1
102	102	DE 2014-01-05 102	1
103	103	DE 2014-01-05 103	1
104	104	DE 2014-01-05 104	1
105	105	DE 2014-01-05 105	1
106	106	DE 2014-01-05 106	1
107	107	DE 2014-01-05 107	1
108	108	DE 2014-01-05 108	1
109	109	DE 2014-01-05 109	1
110	110	DE 2014-01-05 110	1
111	111	DE 2014-01-05 111	1
112	112	DE 2014-01-05 112	1
113	113	DE 2014-01-05 113	1
114	114	DE 2014-01-05 114	1
115	115	DE 2014-01-05 115	1
116	116	DE 2014-01-05 116	1
117	117	DE 2014-01-05 117	1
118	118	DE 2014-01-05 118	1
119	119	DE 2014-01-05 119	1
120	120	DE 2014-01-05 120	1
121	121	DE 2014-01-05 121	1
122	122	DE 2014-01-05 122	1
123	123	DE 2014-01-05 123	1
124	124	DE 2014-01-05 124	1
125	125	DE 2014-01-05 125	1
126	126	DE 2014-01-05 126	1
127	127	DE 2014-01-05 127	1
128	128	DE 2014-01-05 128	1
129	129	DE 2014-01-05 129	1
130	130	DE 2014-01-05 130	1
131	131	DE 2014-01-05 131	1
132	132	DE 2014-01-05 132	1
133	133	DE 2014-01-05 133	1
134	134	DE 2014-01-05 134	1
135	135	DE 2014-01-05 135	1
136	136	DE 2014-01-05 136	1
137	137	DE 2014-01-05 137	1
138	138	DE 2014-01-05 138	1
139	139	DE 2014-01-05 139	1
140	140	DE 2014-01-05 140	1
141	141	DE 2014-01-05 141	1
142	142	DE 2014-01-05 142	1
143	143	DE 2014-01-05 143	1
144	144	DE 2014-01-05 144	1
145	145	DE 2014-01-05 145	1
146	146	DE 2014-01-05 146	1
147	147	DE 2014-01-05 147	1
148	148	DE 2014-01-05 148	1
149	149	DE 2014-01-05 149	1
150	150	DE 2014-01-05 150	1
151	151	DE 2014-01-05 151	1
152	152	DE 2014-01-05 152	1
153	153	DE 2014-01-05 153	1
154	154	DE 2014-01-05 154	1
155	155	DE 2014-01-05 155	1
156	156	DE 2014-01-05 156	1
157	157	DE 2014-01-05 157	1
158	158	DE 2014-01-05 158	1
159	159	DE 2014-01-05 159	1
160	160	DE 2014-01-05 160	1
161	161	DE 2014-01-05 161	1
162	162	DE 2014-01-05 162	1
163	163	DE 2014-01-05 163	1
164	164	DE 2014-01-05 164	1
165	165	DE 2014-01-05 165	1
166	166	DE 2014-01-05 166	1
167	167	DE 2014-01-05 167	1
168	168	DE 2014-01-05 168	1
169	169	DE 2014-01-05 169	1
170	170	DE 2014-01-05 170	1
171	171	DE 2014-01-05 171	1
172	172	DE 2014-01-05 172	1
173	173	DE 2014-01-05 173	1
174	174	DE 2014-01-05 174	1
175	175	DE 2014-01-05 175	1
176	176	DE 2014-01-05 176	1
177	177	DE 2014-01-05 177	1
178	178	DE 2014-01-05 178	1
179	179	DE 2014-01-05 179	1
180	180	DE 2014-01-05 180	1
181	181	DE 2014-01-05 181	1
182	182	DE 2014-01-05 182	1
183	183	DE 2014-01-05 183	1
184	184	DE 2014-01-05 184	1
185	185	DE 2014-01-05 185	1
186	186	DE 2014-01-05 186	1
187	187	DE 2014-01-05 187	1
188	188	DE 2014-01-05 188	1
189	189	DE 2014-01-05 189	1
190	190	DE 2014-01-05 190	1
191	191	DE 2014-01-05 191	1
192	192	DE 2014-01-05 192	1
193	193	DE 2014-01-05 193	1
194	194	DE 2014-01-05 194	1
195	195	DE 2014-01-05 195	1
196	196	DE 2014-01-05 196	1
197	197	DE 2014-01-05 197	1
198	198	DE 2014-01-05 198	1
199	199	DE 2014-01-05 199	1
200	200	DE 2014-01-05 200	1
201	201	DE 2014-01-05 201	1
202	202	DE 2014-01-05 202	1
203	203	DE 2014-01-05 203	1
204	204	DE 2014-01-05 204	1
205	205	DE 2014-01-05 205	1
206	206	DE 2014-01-05 206	1
207	207	DE 2014-01-05 207	1
208	208	DE 2014-01-05 208	1
209	209	DE 2014-01-05 209	1
210	210	DE 2014-01-05 210	1
211	211	DE 2014-01-05 211	1
212	212	DE 2014-01-05 212	1
213	213	DE 2014-01-05 213	1
214	214	DE 2014-01-05 214	1
215	215	DE 2014-01-05 215	1
216	216	DE 2014-01-05 216	1
217	217	DE 2014-01-05 217	1
218	218	DE 2014-01-05 218	1
219	219	DE 2014-01-05 219	1
220	220	DE 2014-01-05 220	1
221	221	DE 2014-01-05 221	1
222	222	DE 2014-01-05 222	1
223	223	DE 2014-01-05 223	1
224	224	DE 2014-01-05 224	1
225	225	DE 2014-01-05 225	1
226	226	DE 2014-01-05 226	1
227	227	DE 2014-01-05 227	1
228	228	DE 2014-01-05 228	1
229	229	DE 2014-01-05 229	1
230	230	DE 2014-01-05 230	1
231	231	DE 2014-01-05 231	1
232	232	DE 2014-01-05 232	1
233	233	DE 2014-01-05 233	1
234	234	DE 2014-01-05 234	1
235	235	DE 2014-01-05 235	1
236	236	DE 2014-01-05 236	1
237			

# 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:

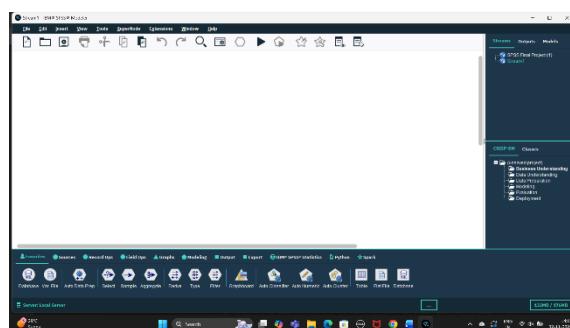
- Clean the dataset
- Aggregate the data
- Transform the data into relevant data

**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.

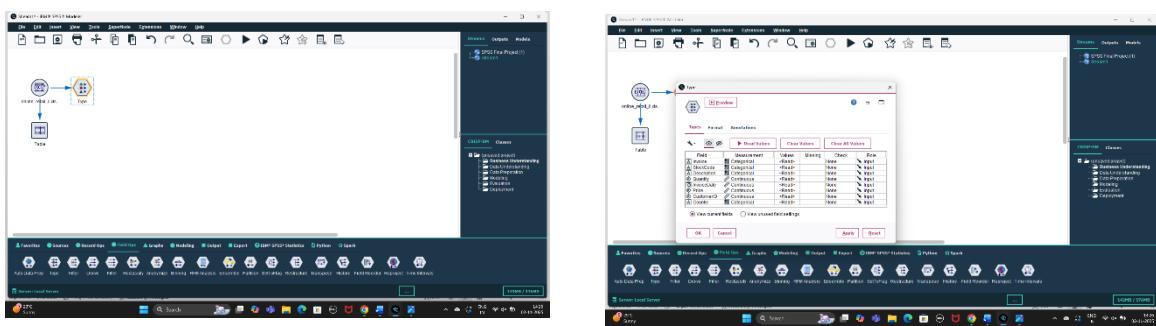


**STEP 2:** From the Sources Category, import the dataset named Online Retail II Dataset (.xlsx). This is an Excel dataset so, select the Excel Node.

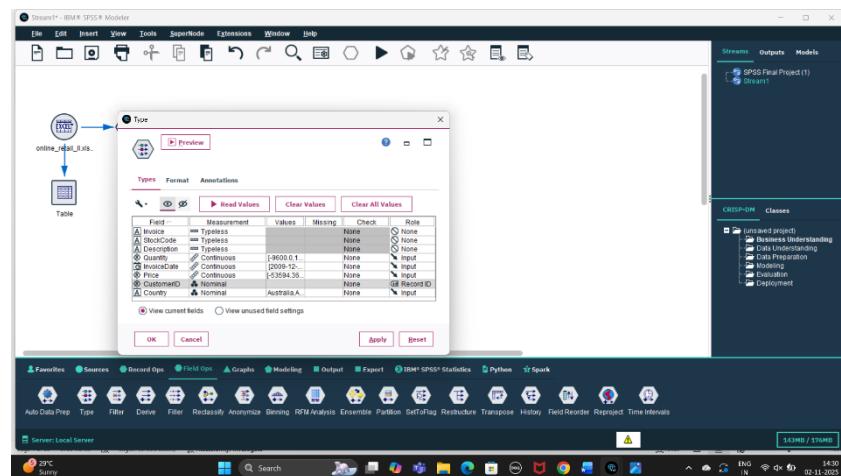


**STEP 3:** Connect a Table from Output Category to Preview the dataset.

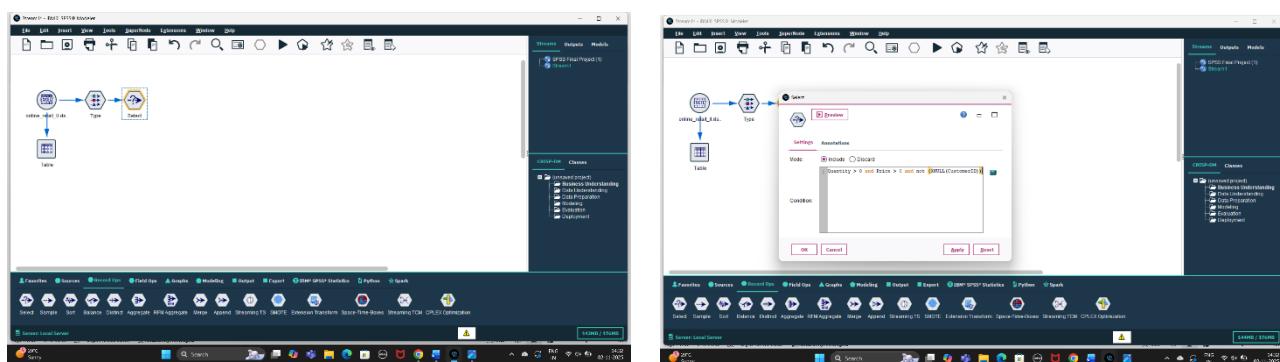
**STEP 4:** Connect the Type Node from Field Ops. This Node aids in understanding the different Fields Measurement Values.



**STEP 5:** Read 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.

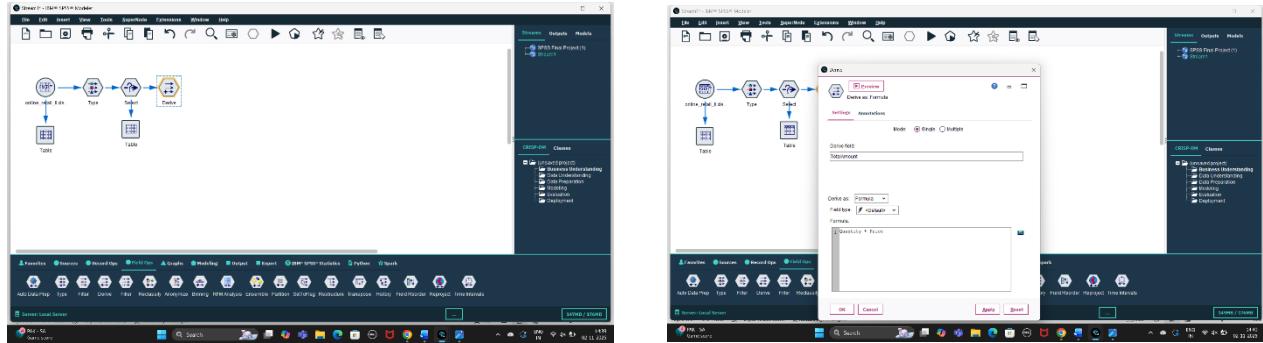


**STEP 6:** Connect A Select Node from Field Ops. This node allows us to include or discard the records. Type the following: Quantity > 0 and Price > 0 and not (@NULL(CustomerID)).



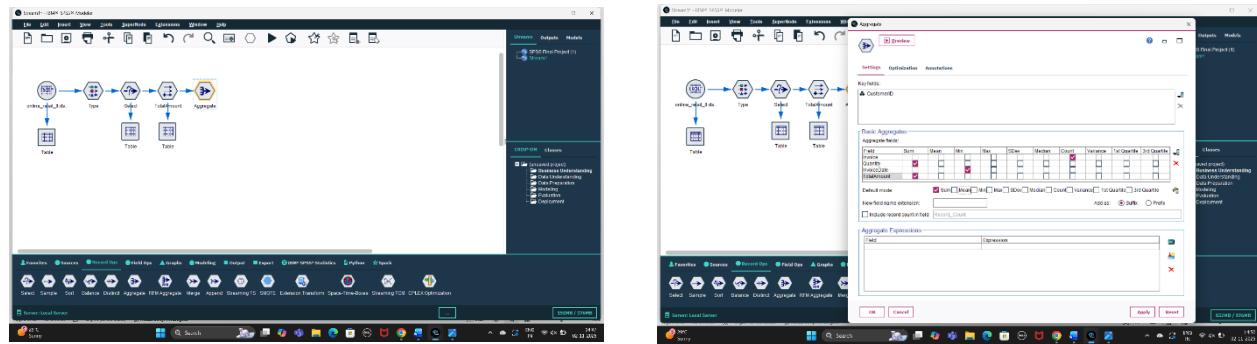
**STEP 7:** Connect a Table to see what changes your Select Node brings.

**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.

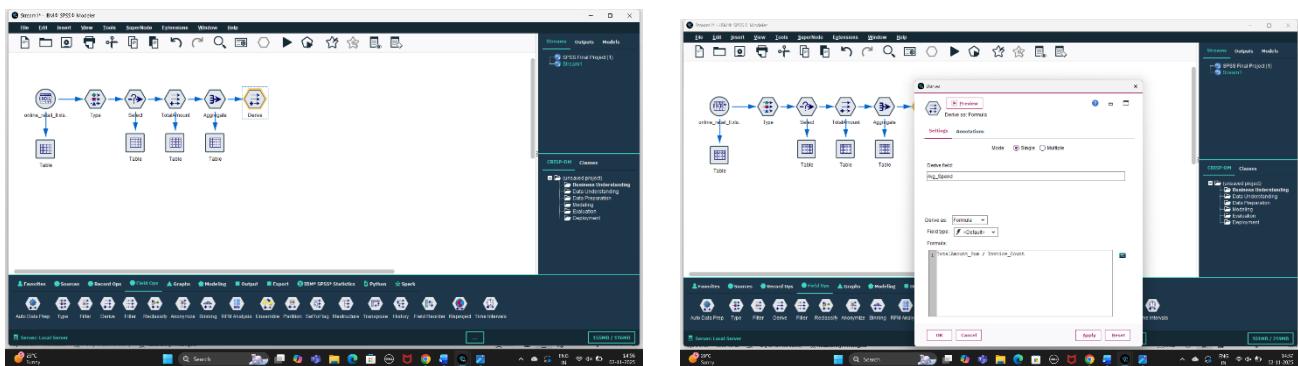
**STEP 10:** From the Record Ops, select Aggregate Node. This node combines multiple records into one row. Select the Key field as CUSTOMER\_ID. In aggregate fields choose Invoice, InvoiceDate, TotalAmount, Quantity.



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

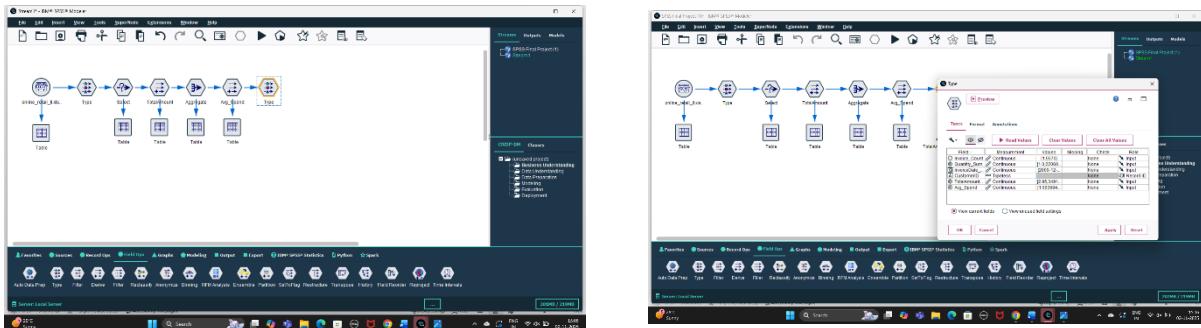
**STEP 12:** Choose one more Derive node. This time we will use it to derive a field called AVG\_SPEND. From the Expression Builder choose the formula:

TotalAmount Sum / Invoice Count.

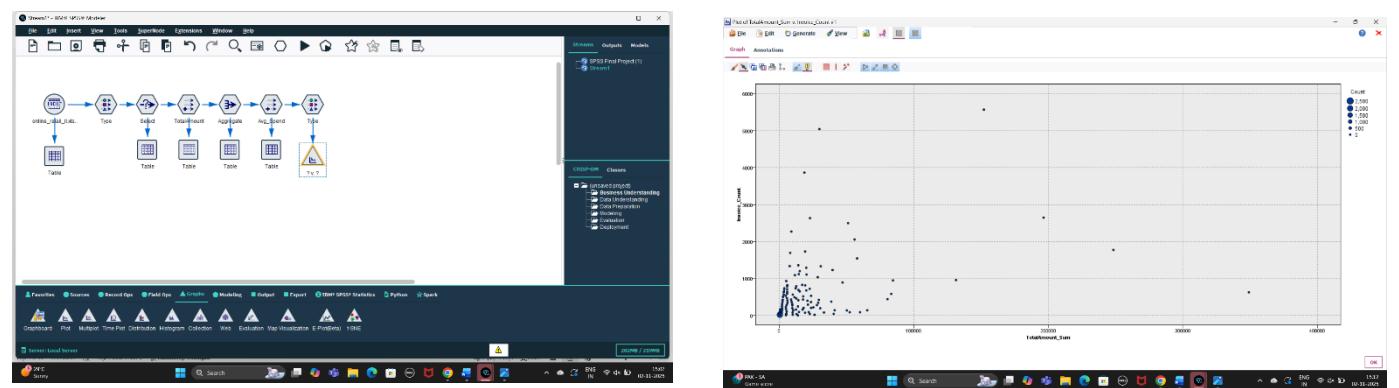


**STEP 13:** Connect a TABLE with this one also.

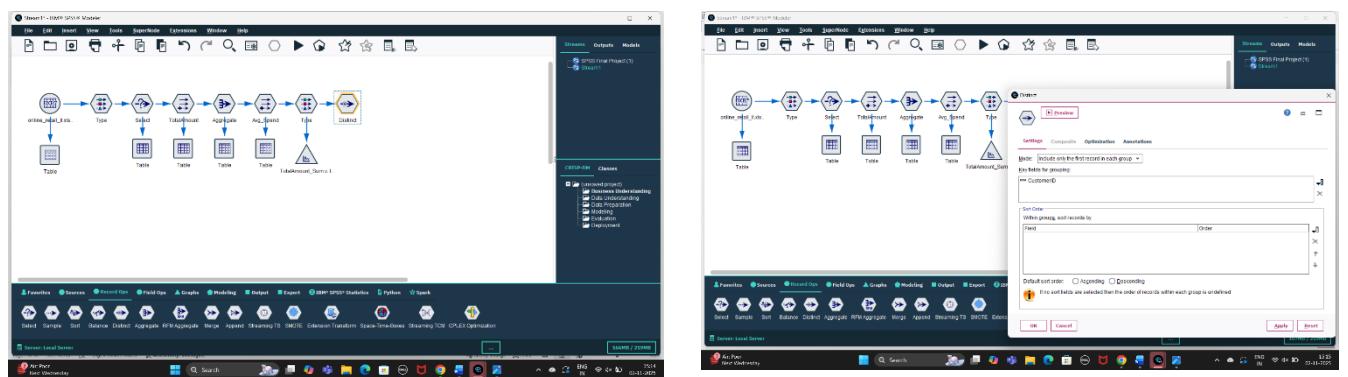
**STEP 14:** Connect the Type node. Read the values.



**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.



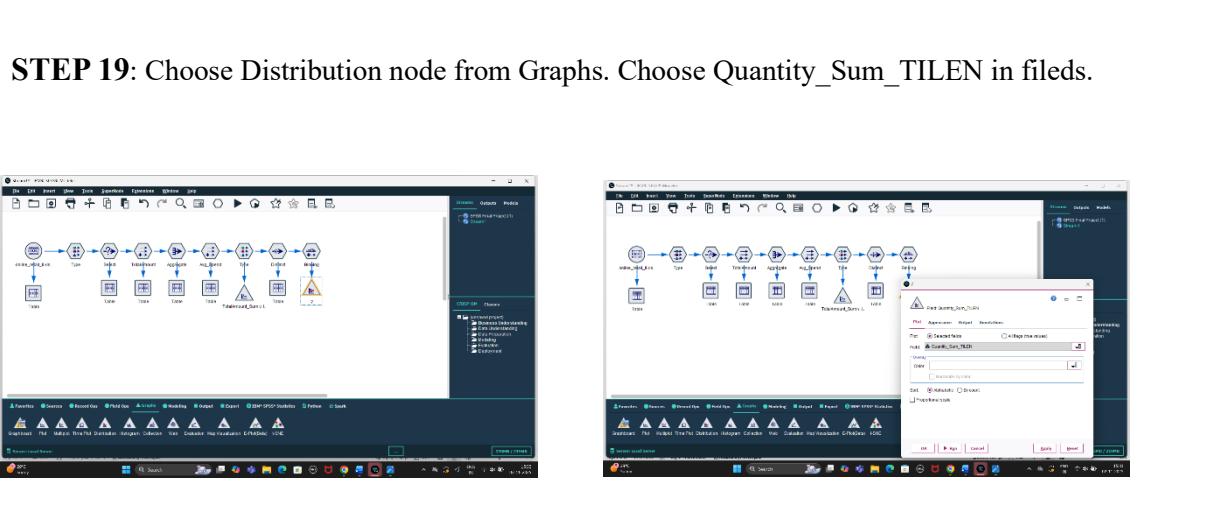
**STEP 17:** Connect a Table again to see how it works.

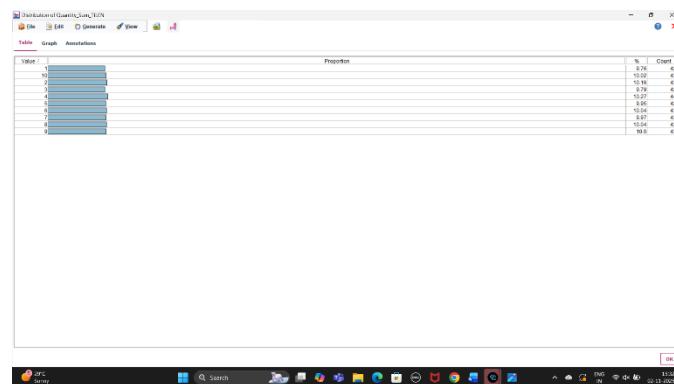
Product_Cat	Order_Day	InvDate	InvTime	Qty	Quantity_Sum	Avg_Open
1	1	2011-01-01 00:00:00	00:00:00	12345	12345	24.432
2	2	2011-01-01 00:00:00	00:00:00	12345	12345	24.432
3	3	2011-01-01 00:00:00	00:00:00	12345	12345	24.432
4	4	2011-01-01 00:00:00	00:00:00	12345	12345	24.432
5	5	2011-01-01 00:00:00	00:00:00	12345	12345	24.432
6	6	2011-01-01 00:00:00	00:00:00	12345	12345	24.432
7	7	2011-01-01 00:00:00	00:00:00	12345	12345	24.432
8	8	2011-01-01 00:00:00	00:00:00	12345	12345	24.432
9	9	2011-01-01 00:00:00	00:00:00	12345	12345	24.432
10	10	2011-01-01 00:00:00	00:00:00	12345	12345	24.432
11	11	2011-01-01 00:00:00	00:00:00	12345	12345	24.432
12	12	2011-01-01 00:00:00	00:00:00	12345	12345	24.432
13	13	2011-01-01 00:00:00	00:00:00	12345	12345	24.432
14	14	2011-01-01 00:00:00	00:00:00	12345	12345	24.432
15	15	2011-01-01 00:00:00	00:00:00	12345	12345	24.432
16	16	2011-01-01 00:00:00	00:00:00	12345	12345	24.432
17	17	2011-01-01 00:00:00	00:00:00	12345	12345	24.432
18	18	2011-01-01 00:00:00	00:00:00	12345	12345	24.432
19	19	2011-01-01 00:00:00	00:00:00	12345	12345	24.432
20	20	2011-01-01 00:00:00	00:00:00	12345	12345	24.432
21	21	2011-01-01 00:00:00	00:00:00	12345	12345	24.432
22	22	2011-01-01 00:00:00	00:00:00	12345	12345	24.432
23	23	2011-01-01 00:00:00	00:00:00	12345	12345	24.432
24	24	2011-01-01 00:00:00	00:00:00	12345	12345	24.432
25	25	2011-01-01 00:00:00	00:00:00	12345	12345	24.432
26	26	2011-01-01 00:00:00	00:00:00	12345	12345	24.432
27	27	2011-01-01 00:00:00	00:00:00	12345	12345	24.432
28	28	2011-01-01 00:00:00	00:00:00	12345	12345	24.432
29	29	2011-01-01 00:00:00	00:00:00	12345	12345	24.432
30	30	2011-01-01 00:00:00	00:00:00	12345	12345	24.432
31	31	2011-01-01 00:00:00	00:00:00	12345	12345	24.432
32	32	2011-01-01 00:00:00	00:00:00	12345	12345	24.432
33	33	2011-01-01 00:00:00	00:00:00	12345	12345	24.432
34	34	2011-01-01 00:00:00	00:00:00	12345	12345	24.432
35	35	2011-01-01 00:00:00	00:00:00	12345	12345	24.432
36	36	2011-01-01 00:00:00	00:00:00	12345	12345	24.432
37	37	2011-01-01 00:00:00	00:00:00	12345	12345	24.432
38	38	2011-01-01 00:00:00	00:00:00	12345	12345	24.432
39	39	2011-01-01 00:00:00	00:00:00	12345	12345	24.432
40	40	2011-01-01 00:00:00	00:00:00	12345	12345	24.432
41	41	2011-01-01 00:00:00	00:00:00	12345	12345	24.432
42	42	2011-01-01 00:00:00	00:00:00	12345	12345	24.432
43	43	2011-01-01 00:00:00	00:00:00	12345	12345	24.432
44	44	2011-01-01 00:00:00	00:00:00	12345	12345	24.432
45	45	2011-01-01 00:00:00	00:00:00	12345	12345	24.432
46	46	2011-01-01 00:00:00	00:00:00	12345	12345	24.432
47	47	2011-01-01 00:00:00	00:00:00	12345	12345	24.432
48	48	2011-01-01 00:00:00	00:00:00	12345	12345	24.432
49	49	2011-01-01 00:00:00	00:00:00	12345	12345	24.432
50	50	2011-01-01 00:00:00	00:00:00	12345	12345	24.432
51	51	2011-01-01 00:00:00	00:00:00	12345	12345	24.432
52	52	2011-01-01 00:00:00	00:00:00	12345	12345	24.432
53	53	2011-01-01 00:00:00	00:00:00	12345	12345	24.432
54	54	2011-01-01 00:00:00	00:00:00	12345	12345	24.432
55	55	2011-01-01 00:00:00	00:00:00	12345	12345	24.432
56	56	2011-01-01 00:00:00	00:00:00	12345	12345	24.432
57	57	2011-01-01 00:00:00	00:00:00	12345	12345	24.432
58	58	2011-01-01 00:00:00	00:00:00	12345	12345	24.432
59	59	2011-01-01 00:00:00	00:00:00	12345	12345	24.432
60	60	2011-01-01 00:00:00	00:00:00	12345	12345	24.432
61	61	2011-01-01 00:00:00	00:00:00	12345	12345	24.432
62	62	2011-01-01 00:00:00	00:00:00	12345	12345	24.432
63	63	2011-01-01 00:00:00	00:00:00	12345	12345	24.432
64	64	2011-01-01 00:00:00	00:00:00	12345	12345	24.432
65	65	2011-01-01 00:00:00	00:00:00	12345	12345	24.432
66	66	2011-01-01 00:00:00	00:00:00	12345	12345	24.432
67	67	2011-01-01 00:00:00	00:00:00	12345	12345	24.432
68	68	2011-01-01 00:00:00	00:00:00	12345	12345	24.432
69	69	2011-01-01 00:00:00	00:00:00	12345	12345	24.432
70	70	2011-01-01 00:00:00	00:00:00	12345	12345	24.432
71	71	2011-01-01 00:00:00	00:00:00	12345	12345	24.432
72	72	2011-01-01 00:00:00	00:00:00	12345	12345	24.432
73	73	2011-01-01 00:00:00	00:00:00	12345	12345	24.432
74	74	2011-01-01 00:00:00	00:00:00	12345	12345	24.432
75	75	2011-01-01 00:00:00	00:00:00	12345	12345	24.432
76	76	2011-01-01 00:00:00	00:00:00	12345	12345	24.432
77	77	2011-01-01 00:00:00	00:00:00	12345	12345	24.432
78	78	2011-01-01 00:00:00	00:00:00	12345	12345	24.432
79	79	2011-01-01 00:00:00	00:00:00	12345	12345	24.432
80	80	2011-01-01 00:00:00	00:00:00	12345	12345	24.432
81	81	2011-01-01 00:00:00	00:00:00	12345	12345	24.432
82	82	2011-01-01 00:00:00	00:00:00	12345	12345	24.432
83	83	2011-01-01 00:00:00	00:00:00	12345	12345	24.432
84	84	2011-01-01 00:00:00	00:00:00	12345	12345	24.432
85	85	2011-01-01 00:00:00	00:00:00	12345	12345	24.432
86	86	2011-01-01 00:00:00	00:00:00	12345	12345	24.432
87	87	2011-01-01 00:00:00	00:00:00	12345	12345	24.432
88	88	2011-01-01 00:00:00	00:00:00	12345	12345	24.432
89	89	2011-01-01 00:00:00	00:00:00	12345	12345	24.432
90	90	2011-01-01 00:00:00	00:00:00	12345	12345	24.432
91	91	2011-01-01 00:00:00	00:00:00	12345	12345	24.432
92	92	2011-01-01 00:00:00	00:00:00	12345	12345	24.432
93	93	2011-01-01 00:00:00	00:00:00	12345	12345	24.432
94	94	2011-01-01 00:00:00	00:00:00	12345	12345	24.432
95	95	2011-01-01 00:00:00	00:00:00	12345	12345	24.432
96	96	2011-01-01 00:00:00	00:00:00	12345	12345	24.432
97	97	2011-01-01 00:00:00	00:00:00	12345	12345	24.432
98	98	2011-01-01 00:00:00	00:00:00	12345	12345	24.432
99	99	2011-01-01 00:00:00	00:00:00	12345	12345	24.432
100	100	2011-01-01 00:00:00	00:00:00	12345	12345	24.432

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

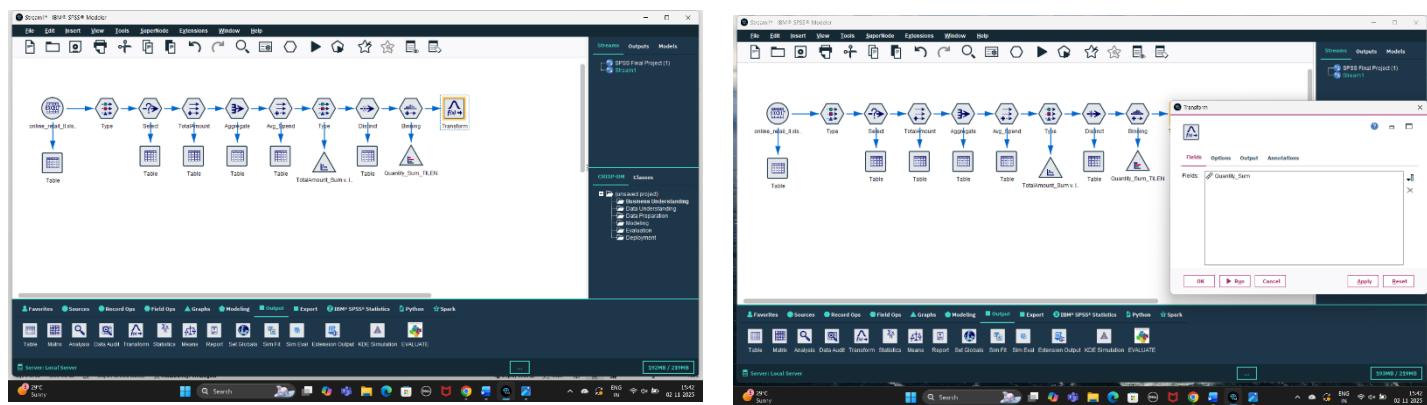
Double click on this node, select Binning method as Tiles and in binning fields choose Quantity\_Sum.

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

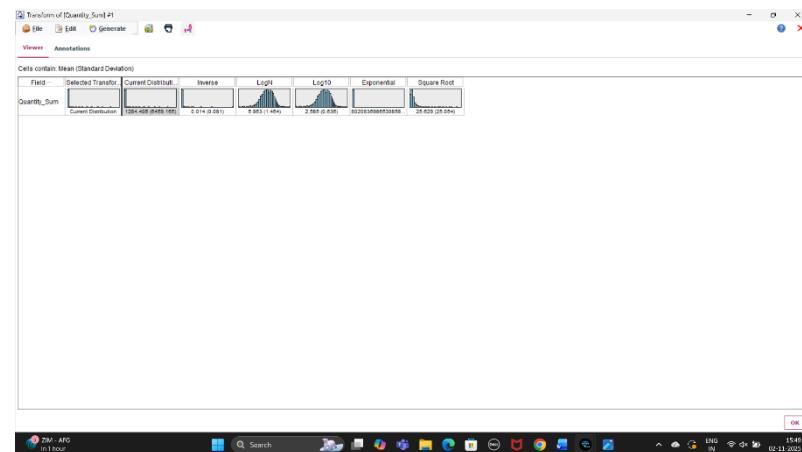




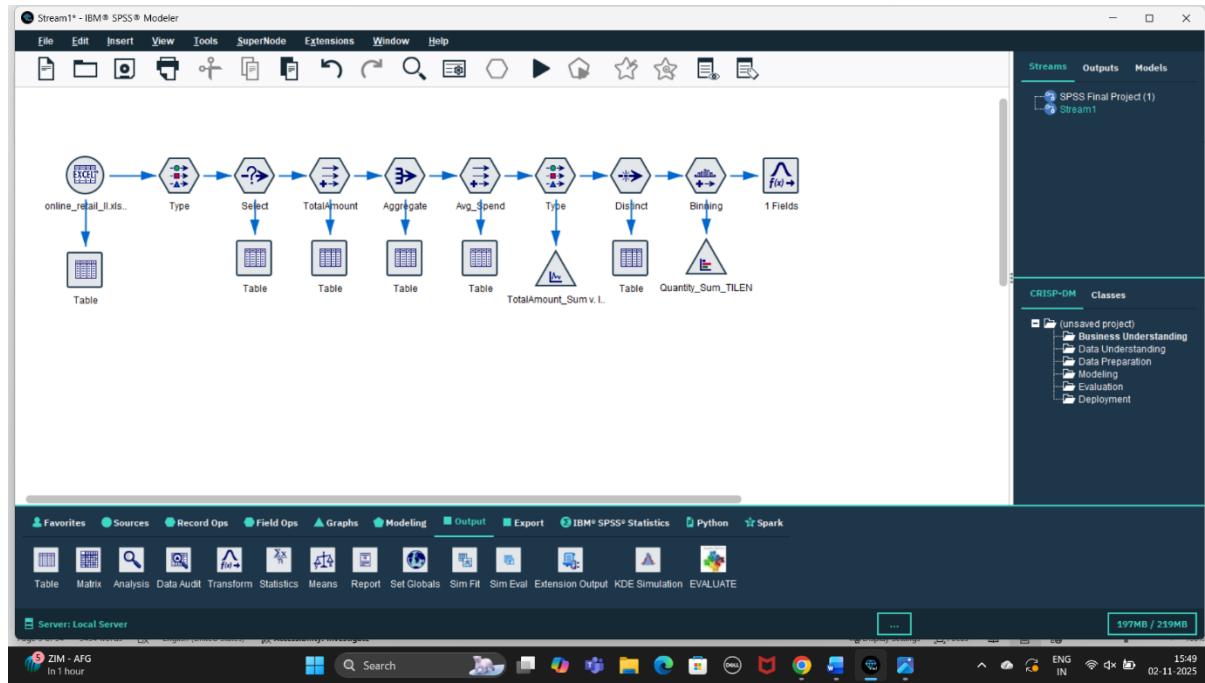
**STEP 21:** From Output Category choose Transform Node, this node will transform the data in a standard format. Double click on the node and choose Quantity\_Sum in the fields.



**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-----

