

Roll No. 33

Exam Seat No. _____

VIVEKANAND EDUCATION SOCIETY'S INSTITUTE

OF INFORMATION TECHNOLOGY

Hashu Advani Memorial Complex, Collector's Colony, R. C.
Marg, Chembur, Mumbai – 400074. Contact No. 02261532532



Since 1962

CERTIFICATE

Certified that Mr. /Miss **Ninad Avinash Patil** of MCA SECOND YEAR SECOND SHIFT has satisfactorily completed a course of the necessary experiments in DATA MINING AND BUSINESS INTELLIGENCE under my supervision in the Institute of Technology in the academic year 2019 - 2022.

Principal

Head of Department

Faculty In-charge

External Examiner



**V.E.S. Institute of Technology, Collector Colony,
Chembur, Mumbai
Department of M.C.A**

INDEX

Sr. No.	Contents	Date Of Performance	Date Of Submission	Marks	Signature
1	Open-Source BI Tools: Preparing Reports Preparing Dashboards Preparing Balanced Scorecards Analysis of Reports				
2	Data Warehousing using Oracle: Setting Up and Starting Warehouse Builder Introducing OWB Architecture and Configuration Defining Source Metadata Ensuring Data Quality Using Data Profiling Defining Staging Metadata and Mapping Tables Deriving Data Rules and Running Correction Mappings Defining a Relational Dimensional Model Handling Slowly Changing Dimensions				
3	OLAP with Oracle: Analytical Queries Grouping Functions Windowing Functions Rollup and Cube				
4	Data Mining Using Weka/R Miner: Introducing Weka/R Miner the Data Mining Process Using Classification Models Using Regression Models Using Clustering Models Performing Market Basket Analysis Performing Anomaly Detection Deploying Data Mining Results				
5	Mini Project: A Mini Projects based on Data Mining and Business Intelligence Techniques using advanced Web Technologies.				

Practical No 1

Assignment 1: Analysis operations

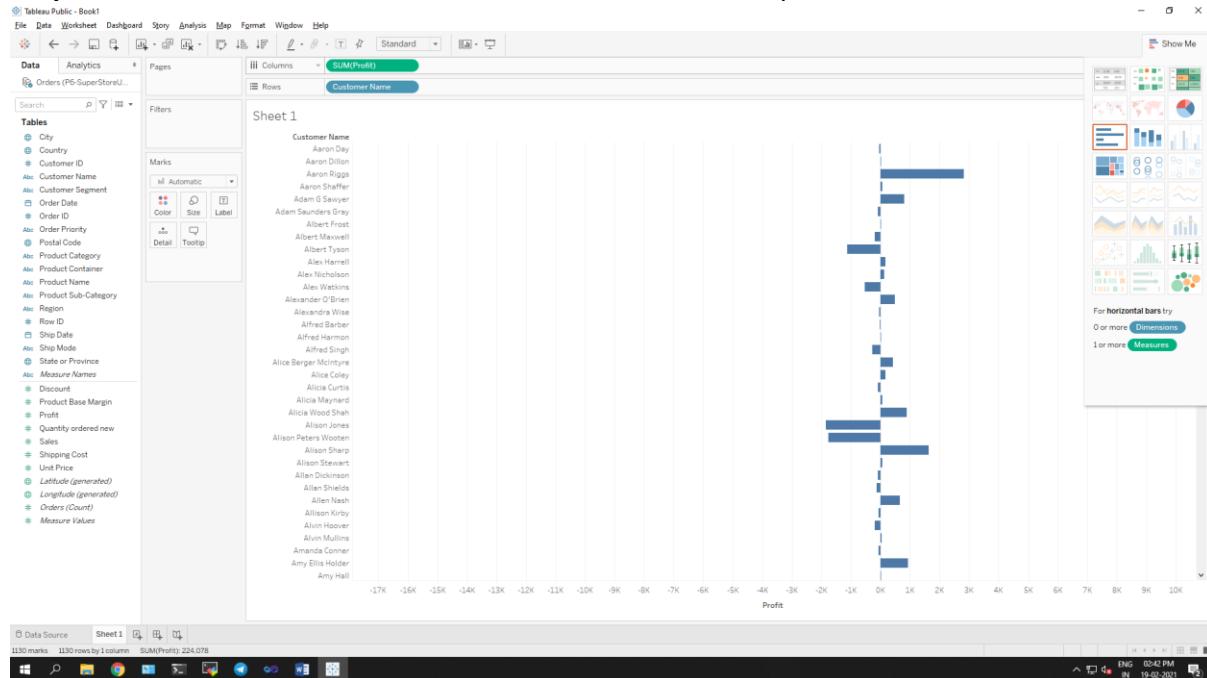
Q1: Find the customer with the highest overall profit. What is his/her profit ratio?

Ans:

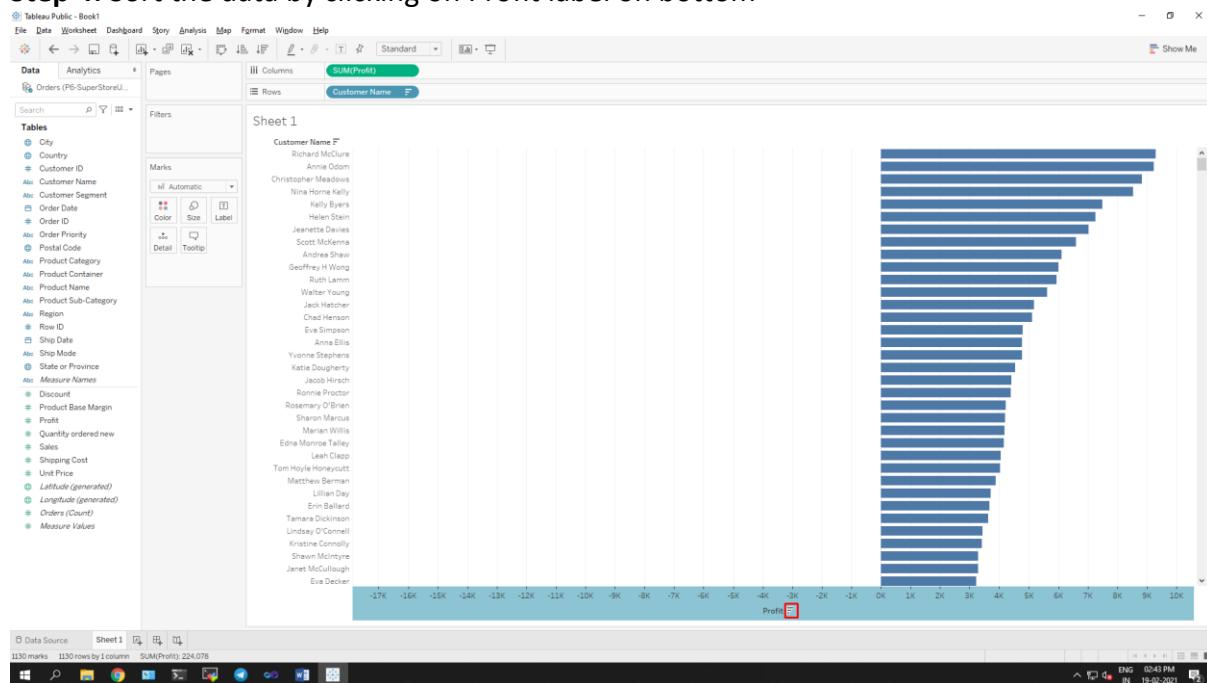
Step 1: Open the superstoreus2015 excel data set

Step 2: Drag Orders sheet to sheet area

Step 3: Go to sheet 1 and add Customer name as rows and profit as column



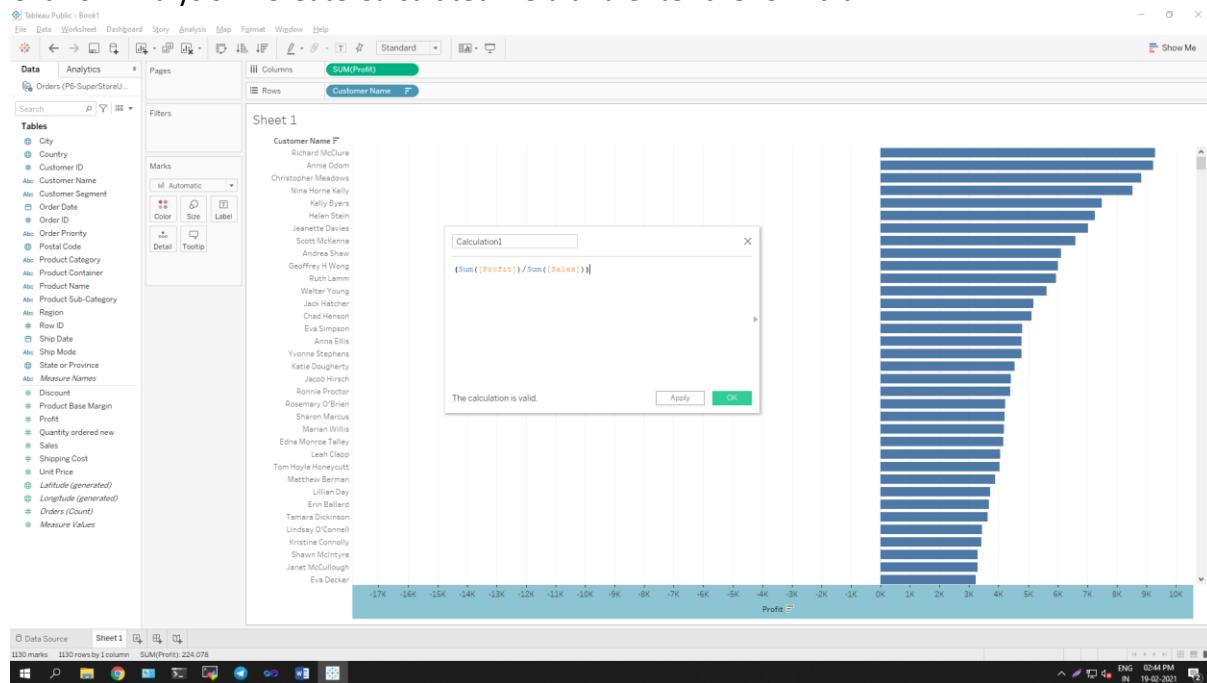
Step 4: Sort the data by clicking on Profit label on bottom



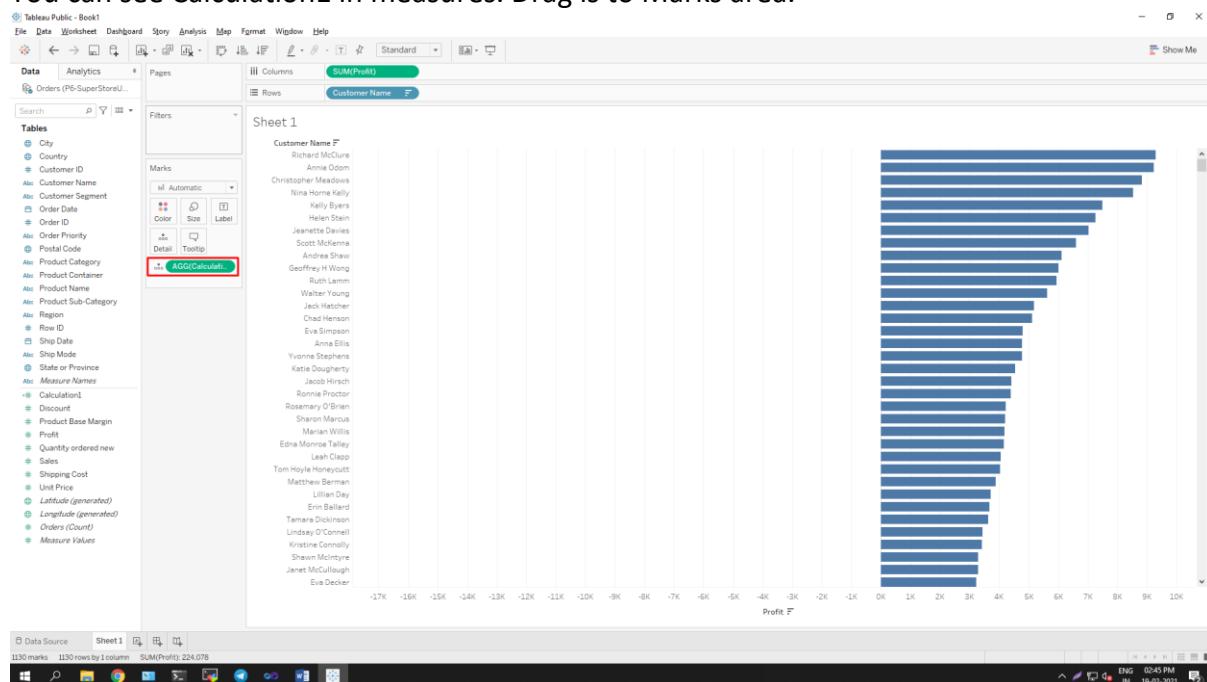
Step 5: To calculate profit Ratio

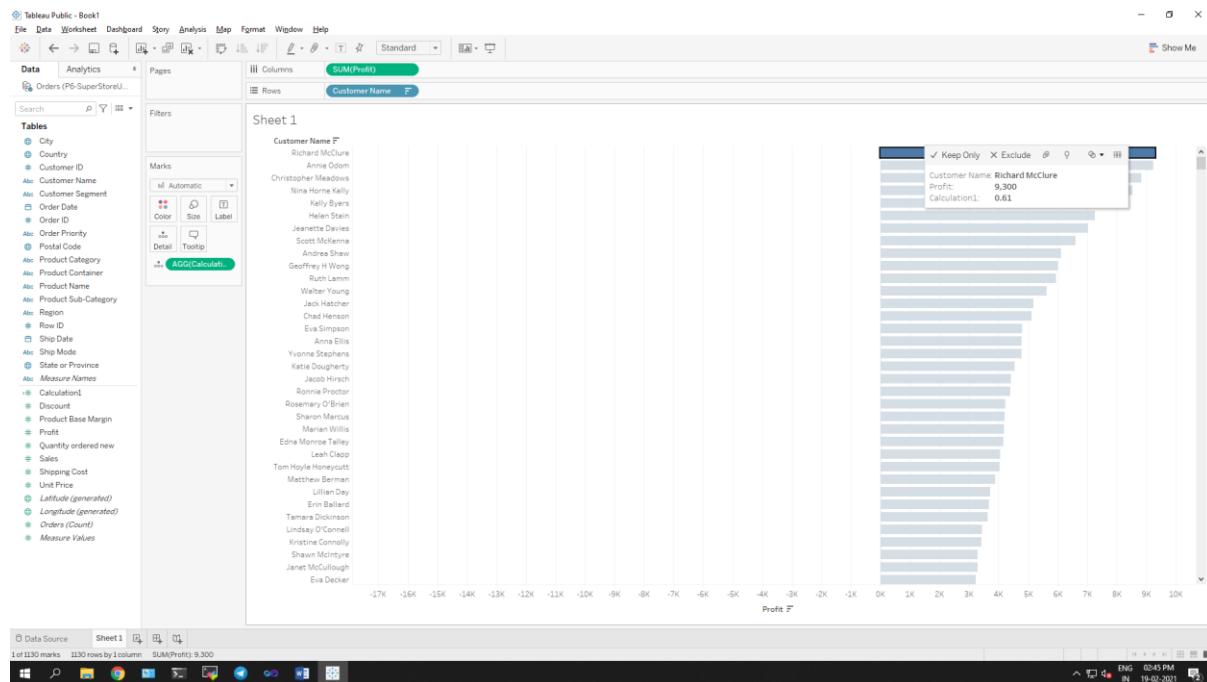
$$\text{Profit Ratio} = (\text{Sum}([\text{Profit}]) / \text{Sum}([\text{Sales}]))$$

This formula needs to be entered as tooltip or label
 Click on Analysis -> Create Calculated Field and enter the formula



You can see Calculation1 in measures. Drag is to Marks area.





Customer Name: Richard McClure

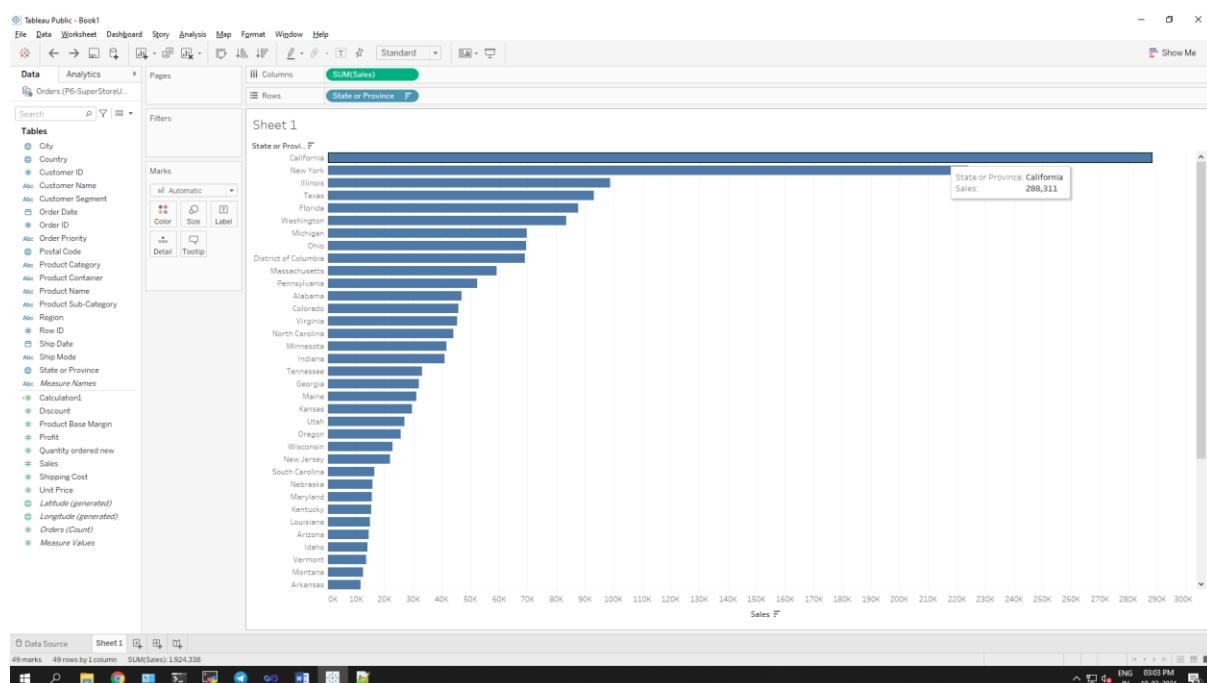
Profit: 9,300

Calculation1: 0.61

The customer with the highest overall profit is Richard McClure and his profit ratio are 0.61.

Q2: Which state has the highest Sales (Sum)? What are the total Sales for that state?

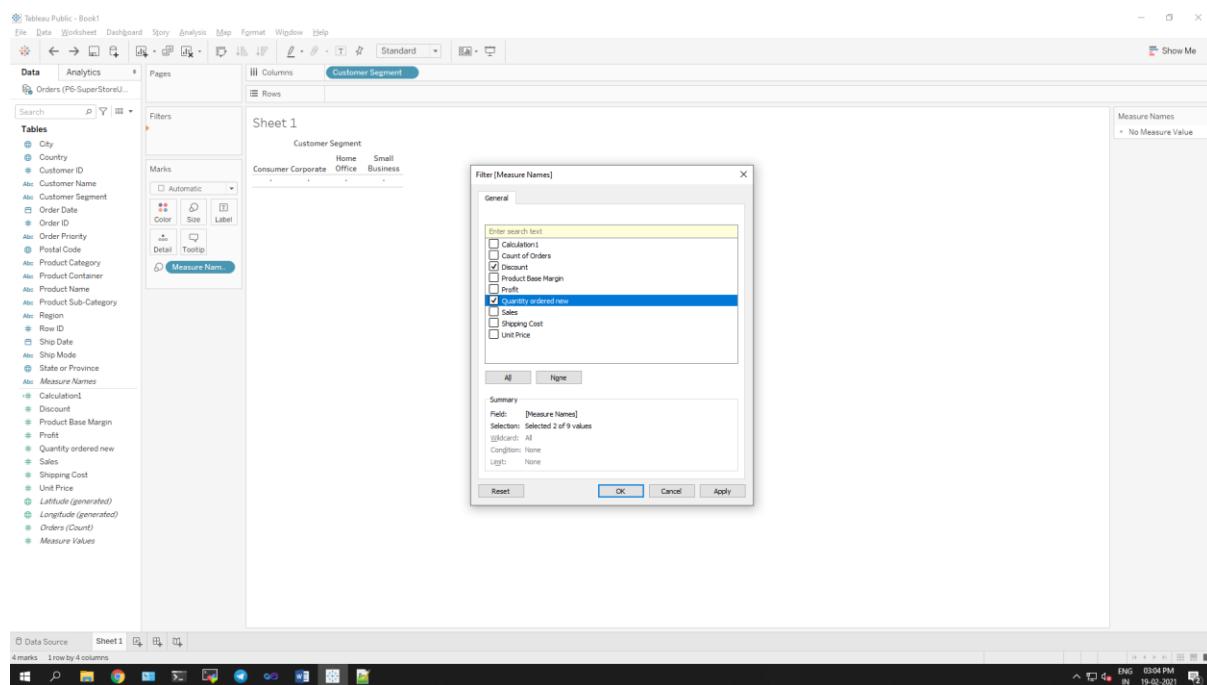
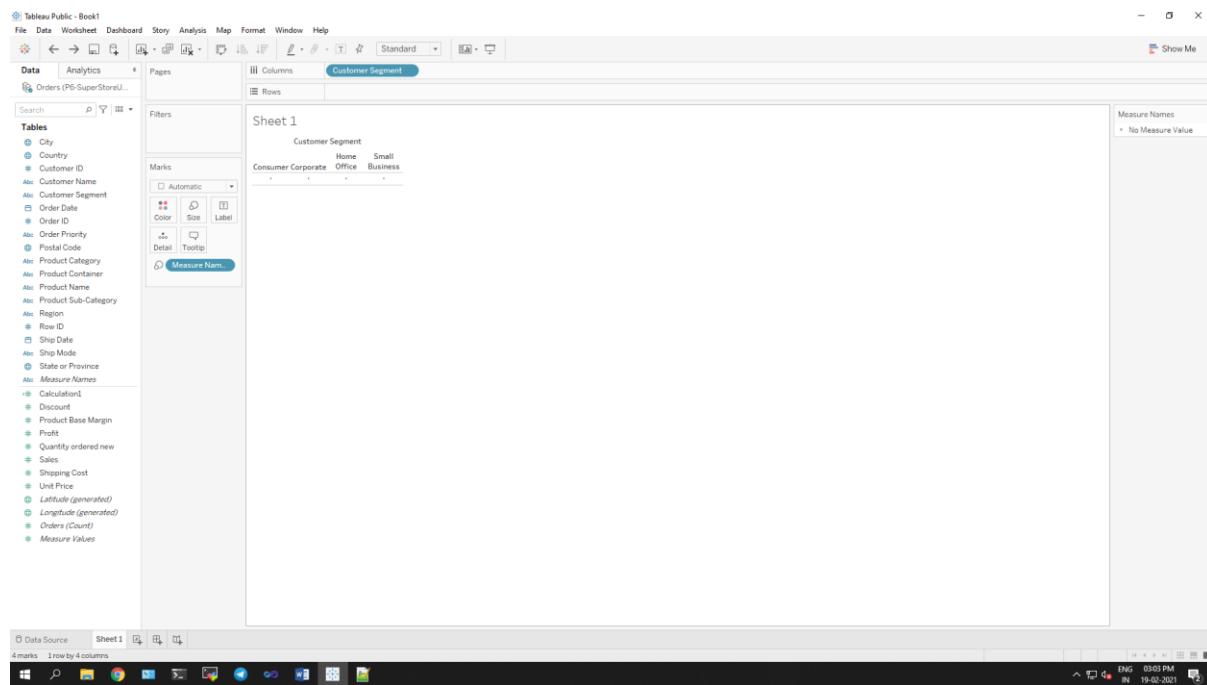
Ans:

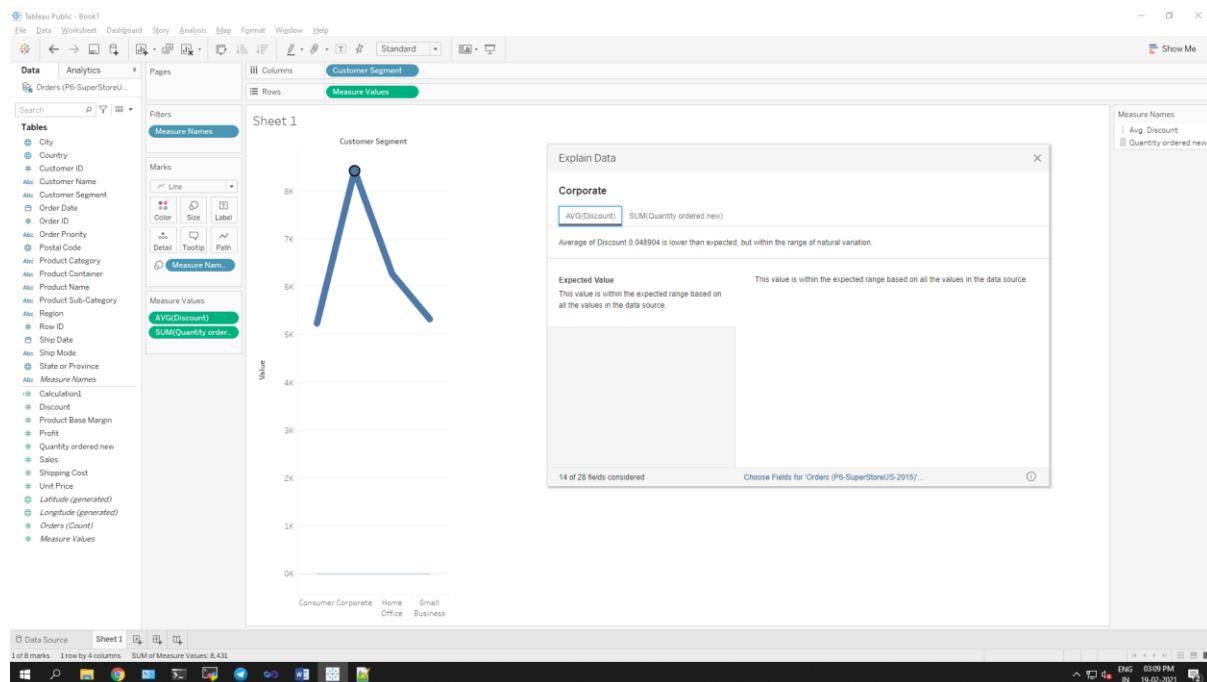
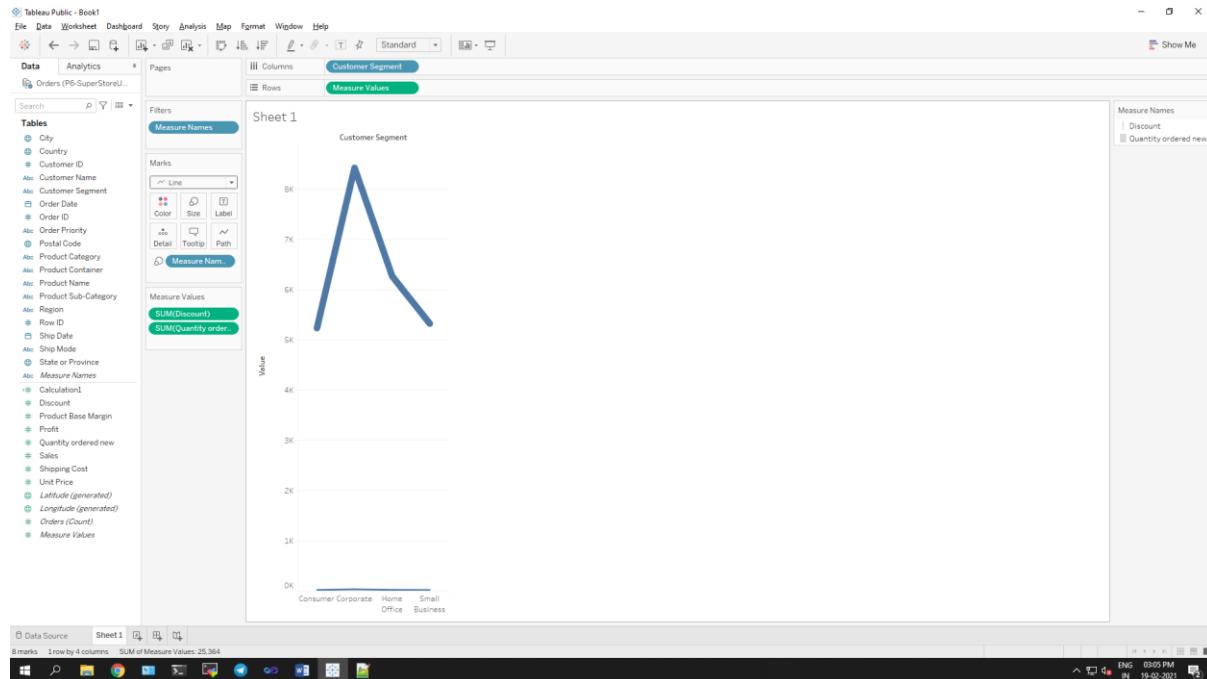


California has the highest Sales (Sum) and the total Sales is 288,311.

Q 3: Which customer segment has both the highest order quantity and average discount rate? What is the order quantity and average discount rate for that state?

Ans:



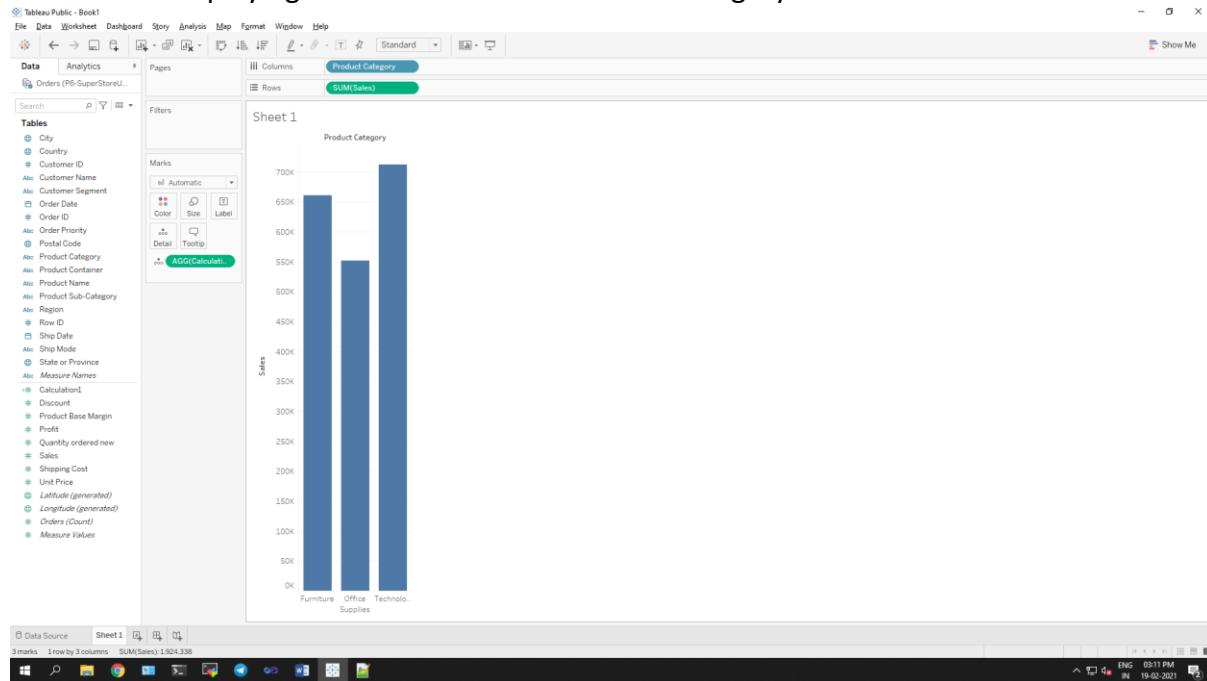


Corporate customer segment has both the highest order quantity and average discount rate. The order quantity and average discount rate for that segment is 8431 and 0.048904 respectively.

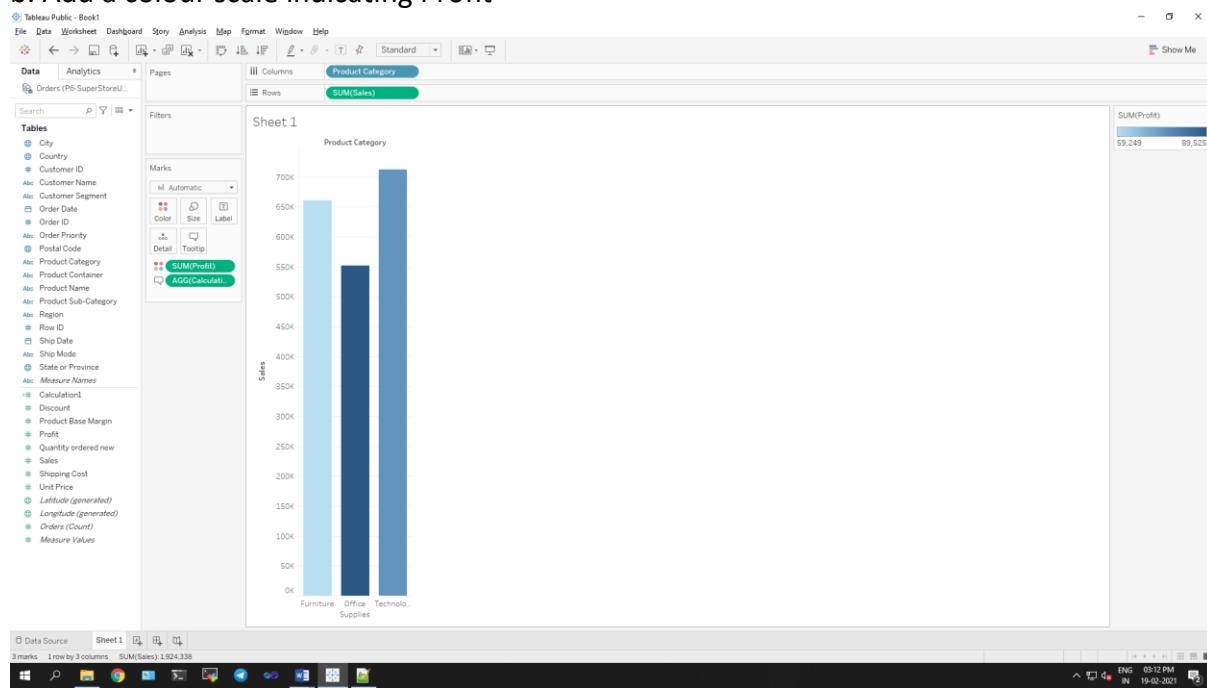
Q 4: Which Product Category has the highest total Sales? Which Product Category has the worst Profit? Name the Product Category and \$ amount for each.

Ans:

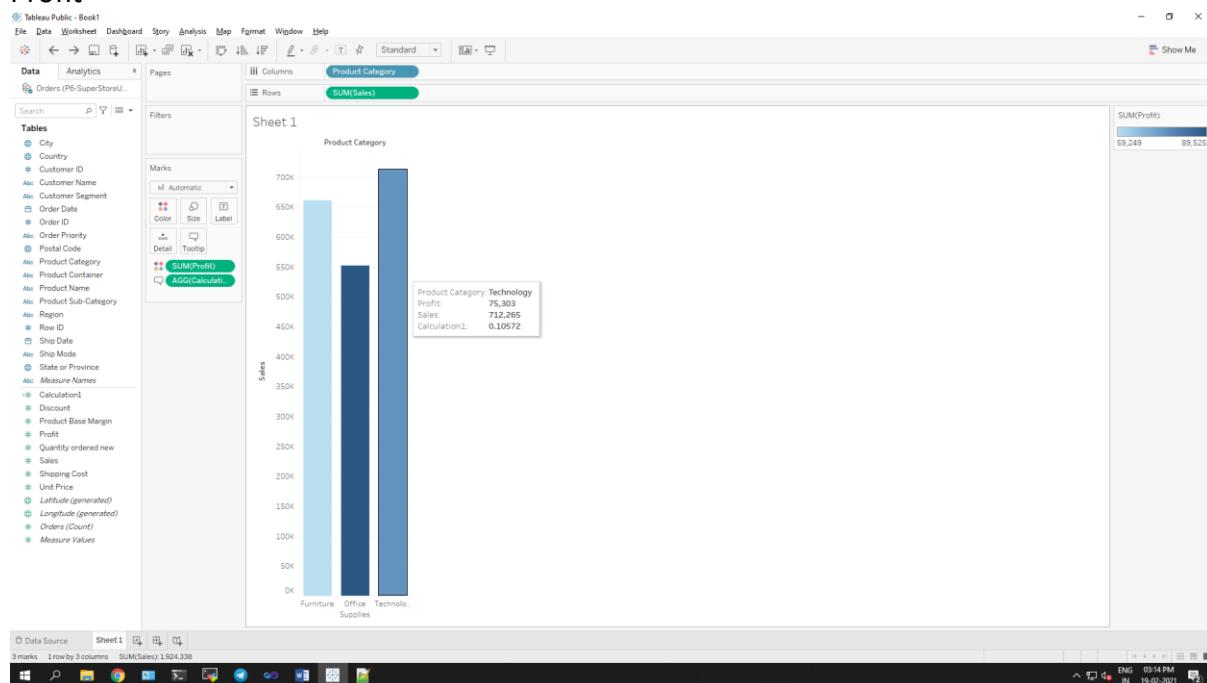
a. Bar Chart displaying total Sales for each Product Category



b. Add a colour scale indicating Profit



c. Each Product Category labelled with total Sales and Each Product Category labelled with Profit

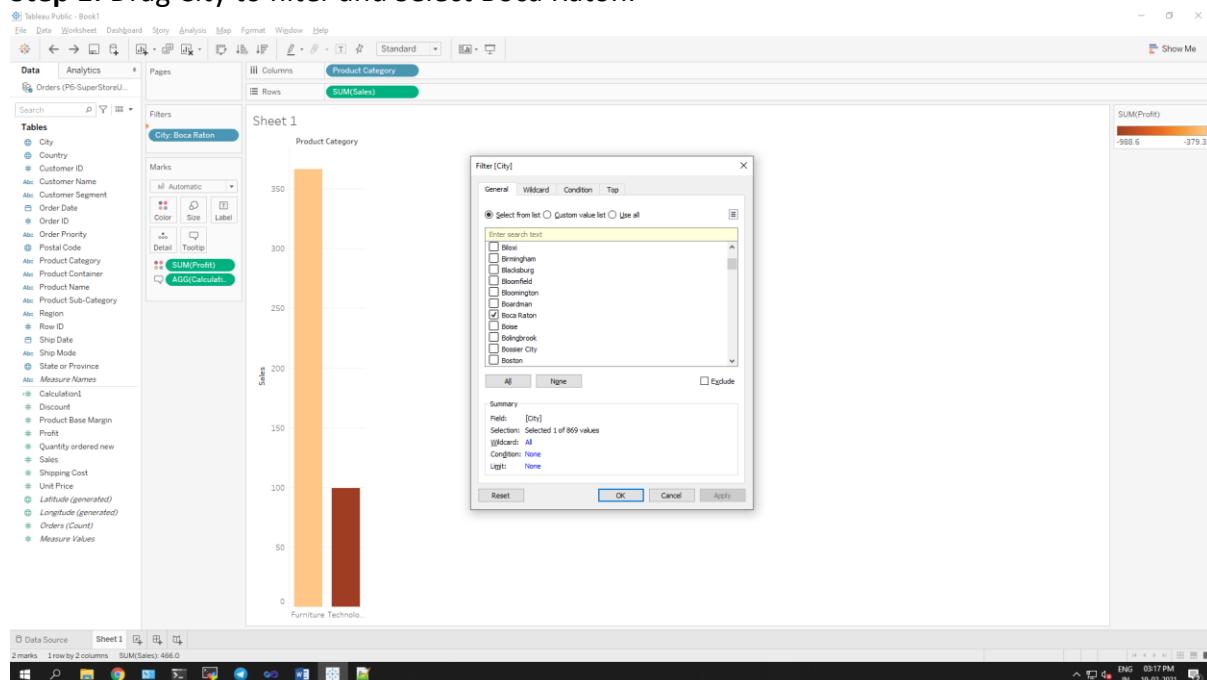


Product Category having the highest total Sales is Technology. Product Category having the worst Profit is Furniture. The Product Category and amount for each can be found from above image.

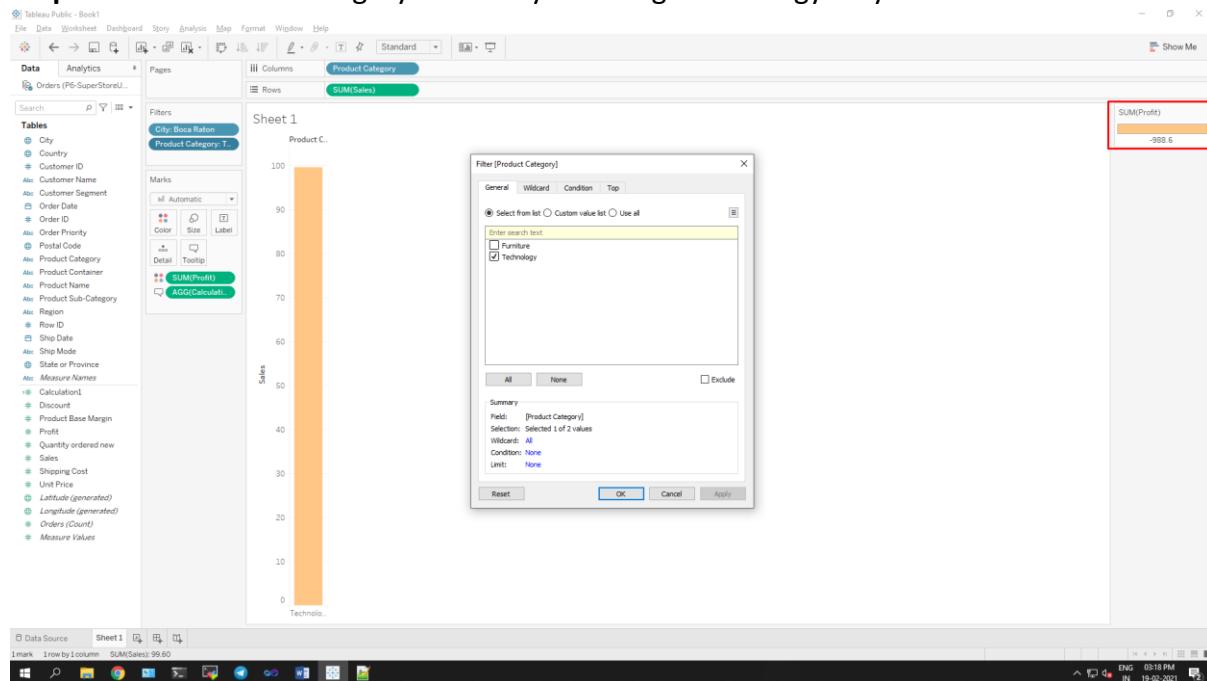
Q 5: Use the same visualization created for Question #4. What was the Profit on Technology (Product Category) in Boca Raton (City)?

Ans:

Step 1: Drag City to filter and Select Boca Raton.



Step 2: Filter Product Category as well by selecting Technology only.

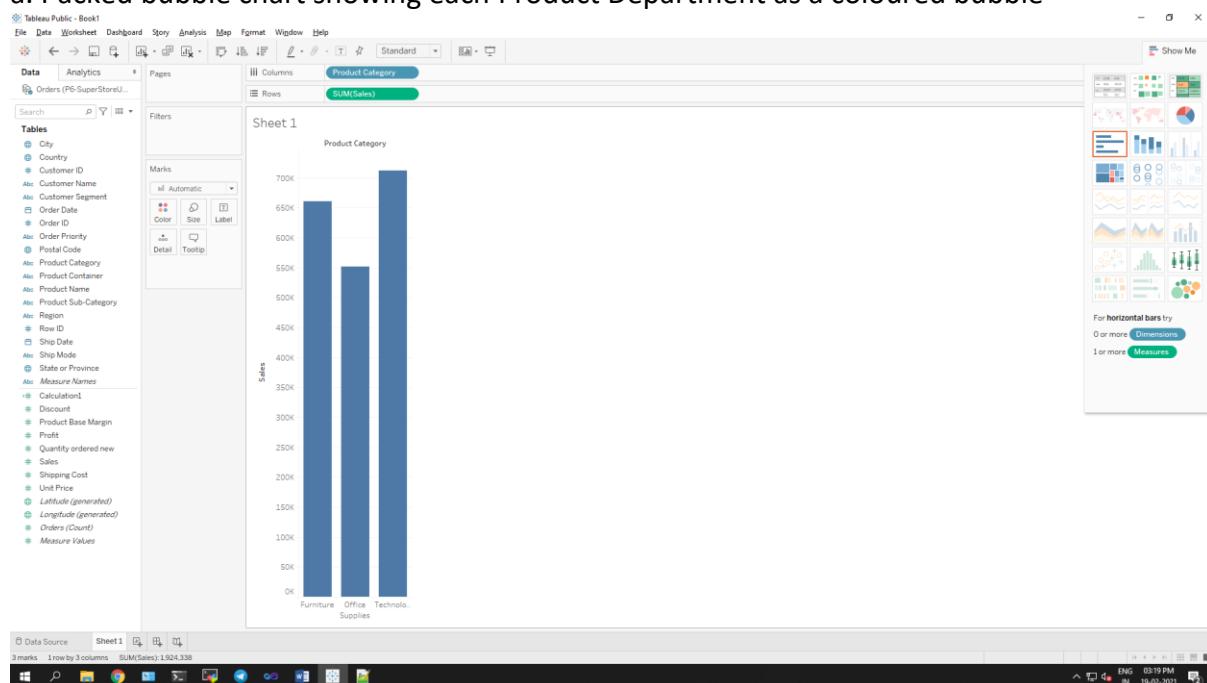


The Profit on Technology (Product Category) in Boca Raton (City) is -988.6.

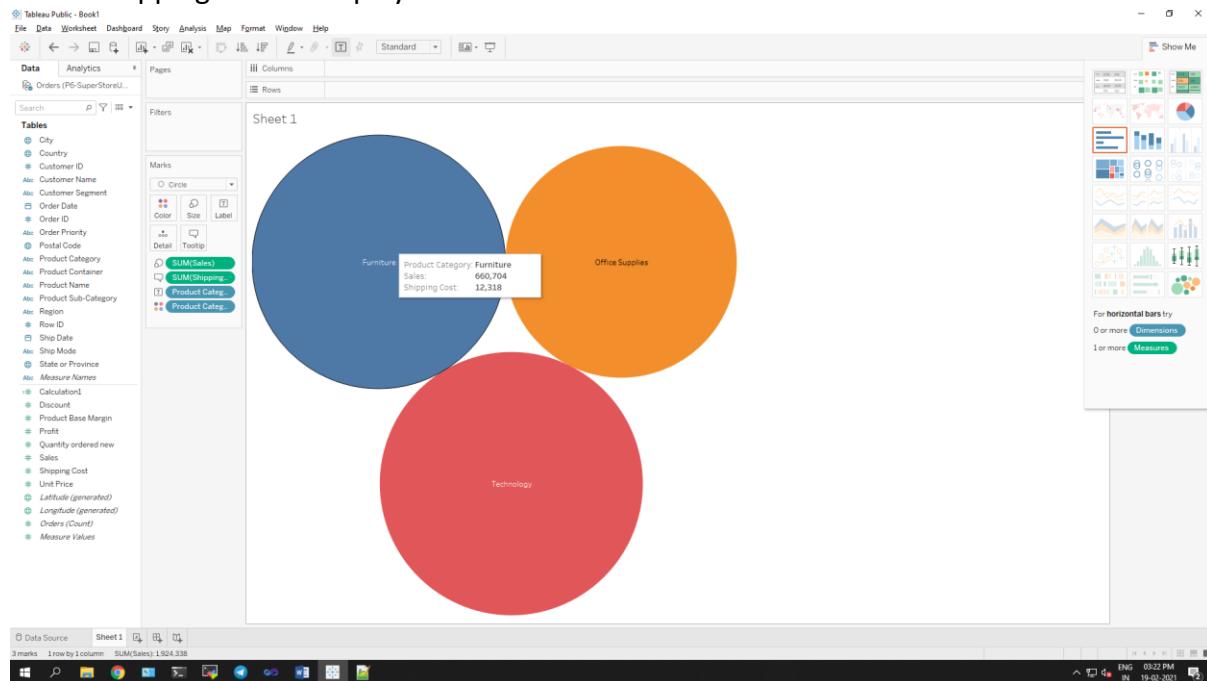
Q 6: Which Product Department has the highest Shipping Costs? Name the Department and cost.

Ans:

a. Packed bubble chart showing each Product Department as a coloured bubble



b. Use Shipping Cost to display the size of each bubble



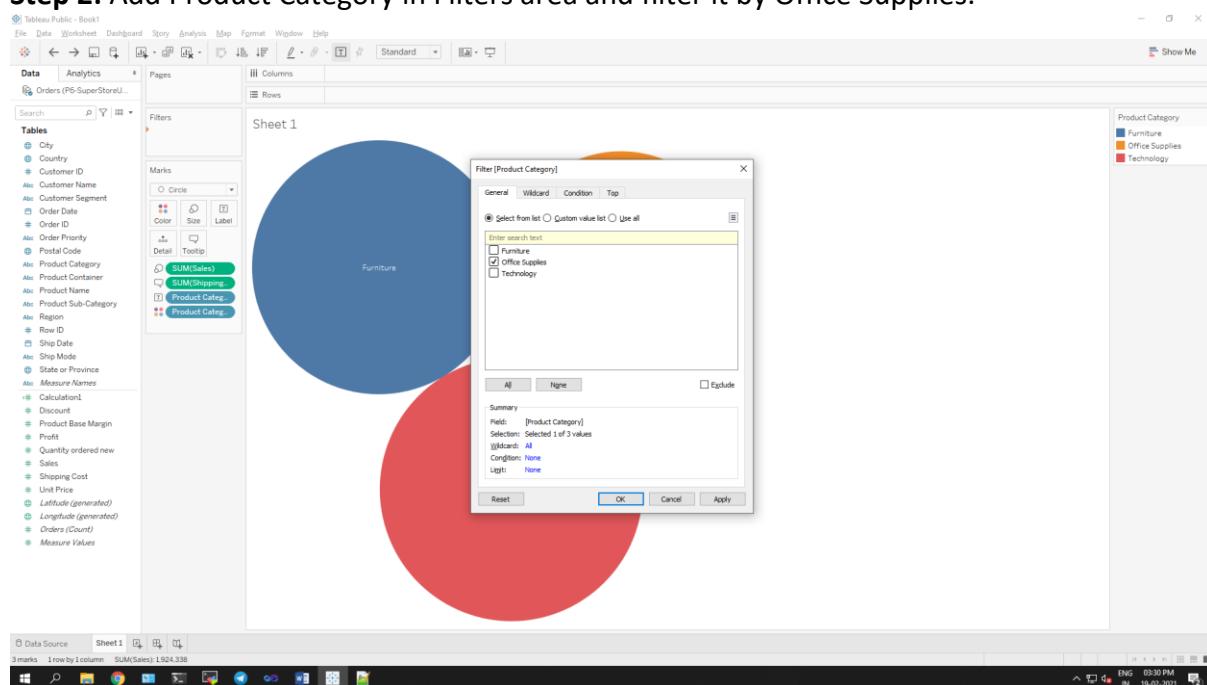
Product Department having the highest Shipping Costs is Furniture with Shipping cost 12.318.

Q 7: Use the same visualization created for Question #6. What was the shipping cost of Office Supplies for Xerox 1905 in the Home Customer Segment in Cambridge?

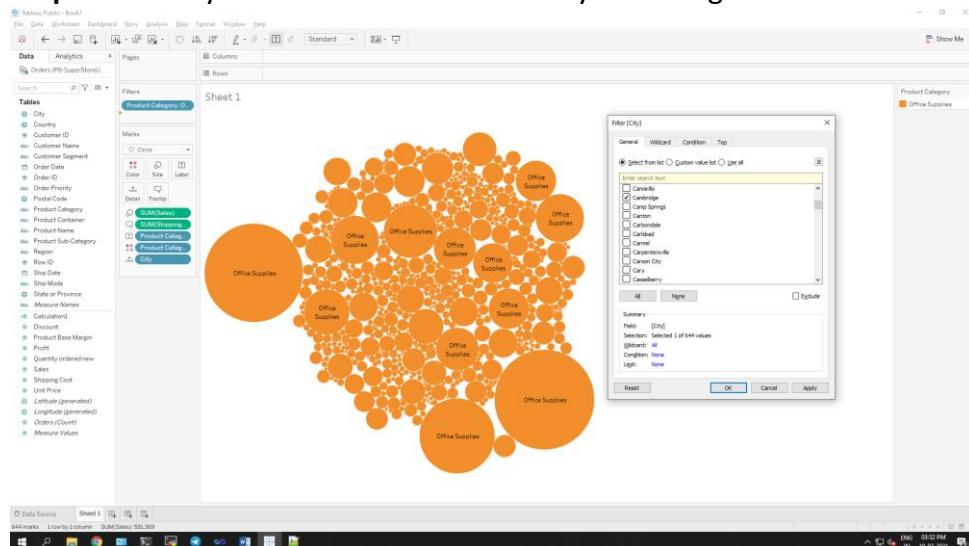
Ans:

Step 1: Apply filters as per requirement

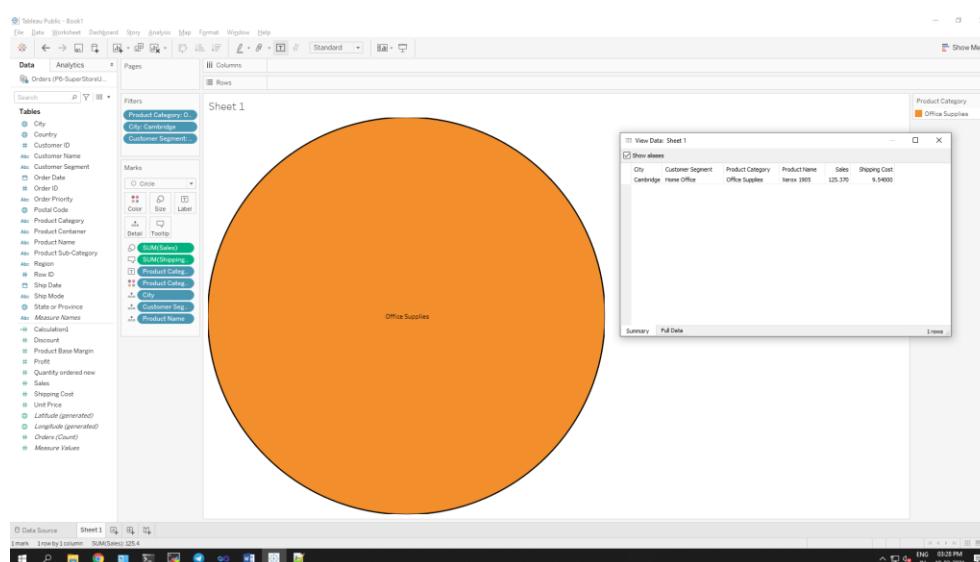
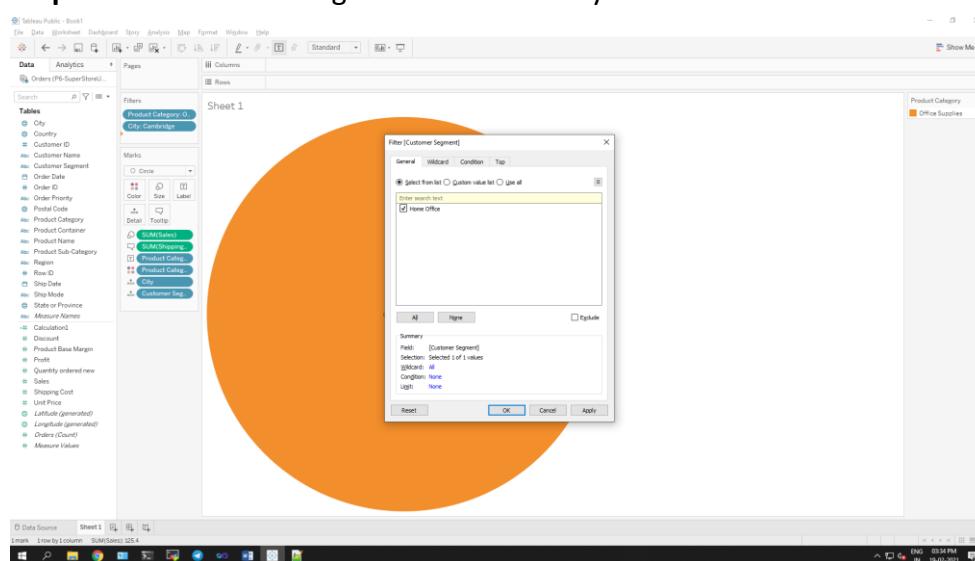
Step 2: Add Product Category in Filters area and filter it by Office Supplies.



Step 3: Add City in Filters area and filter it by Cambridge.



Step 4: Add Customer Segment and filter it by Home Office.



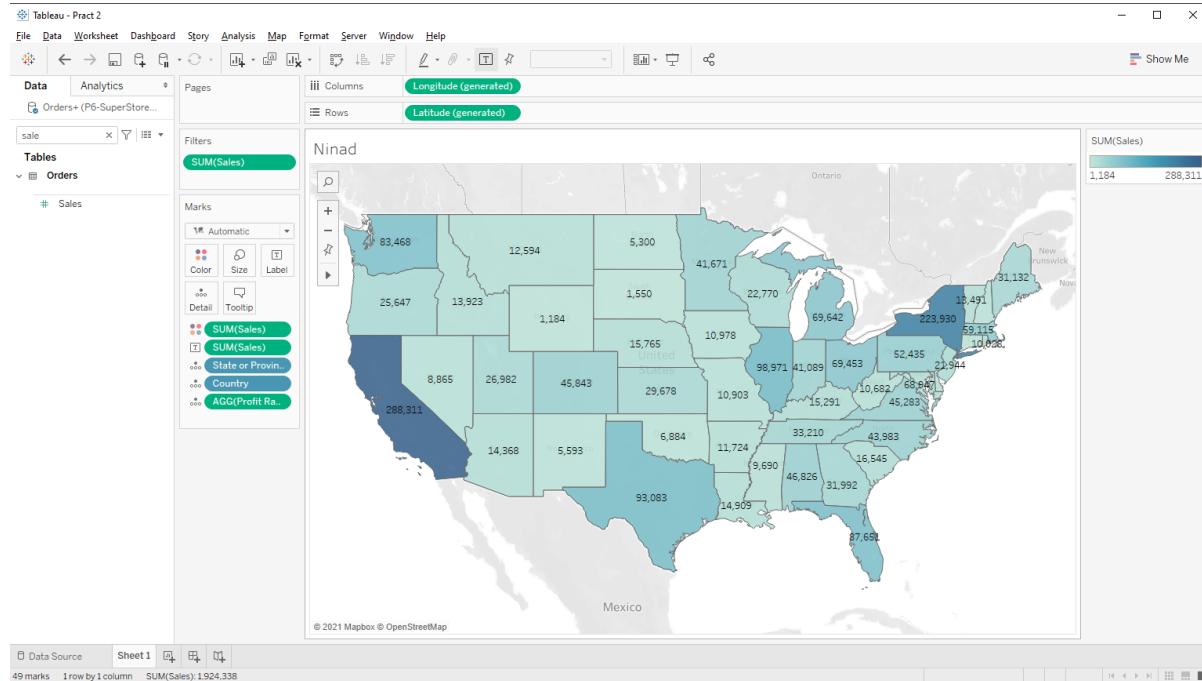
The shipping cost of Office Supplies for Xerox 1905 in the Home Customer Segment in Cambridge is 9.540.

Practical No 1

Assignment 2: Preparing Maps

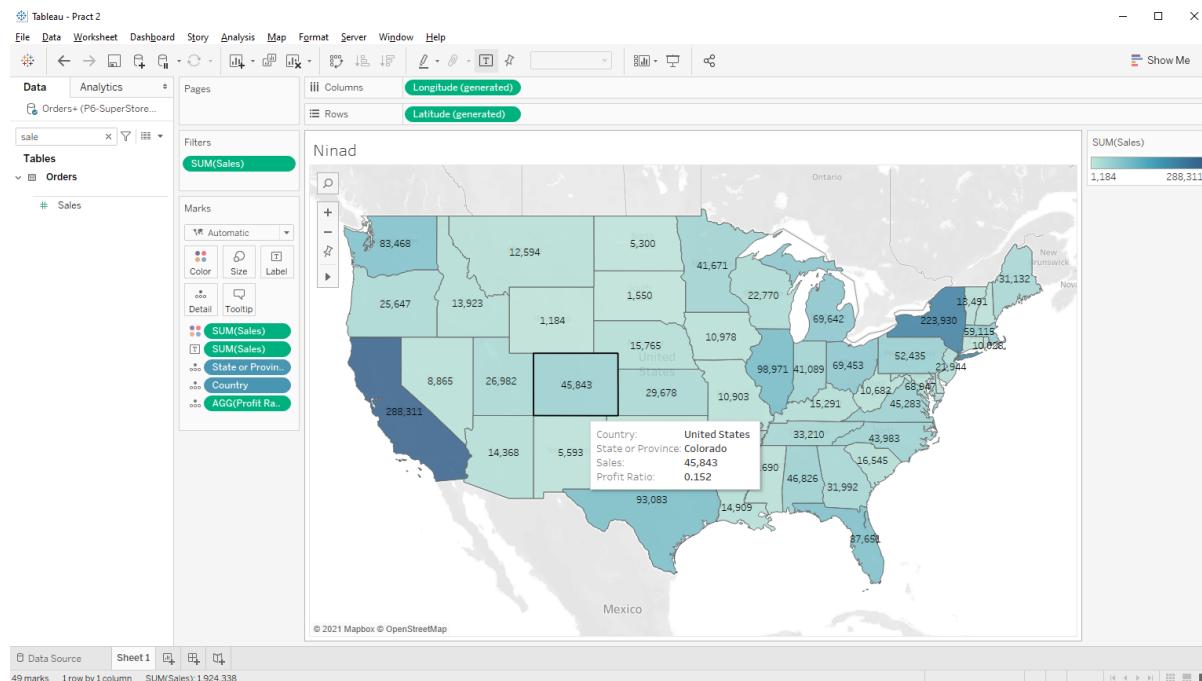
Q1: Prepare a Geographic map to show sales in each state.

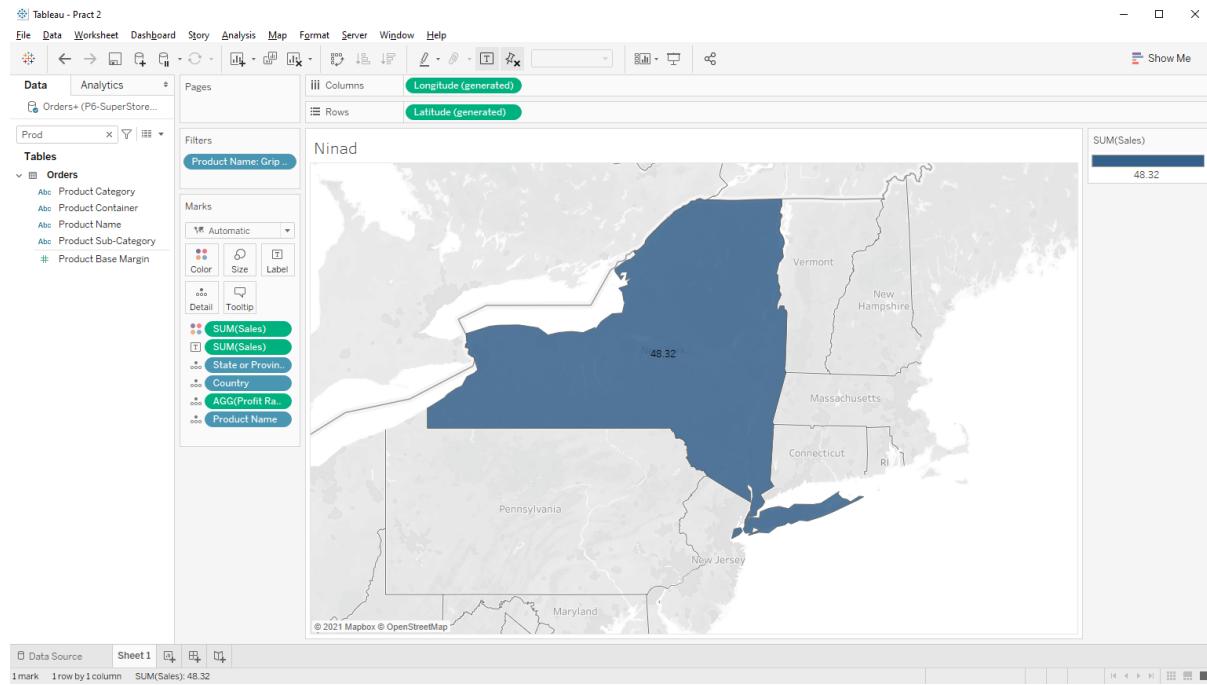
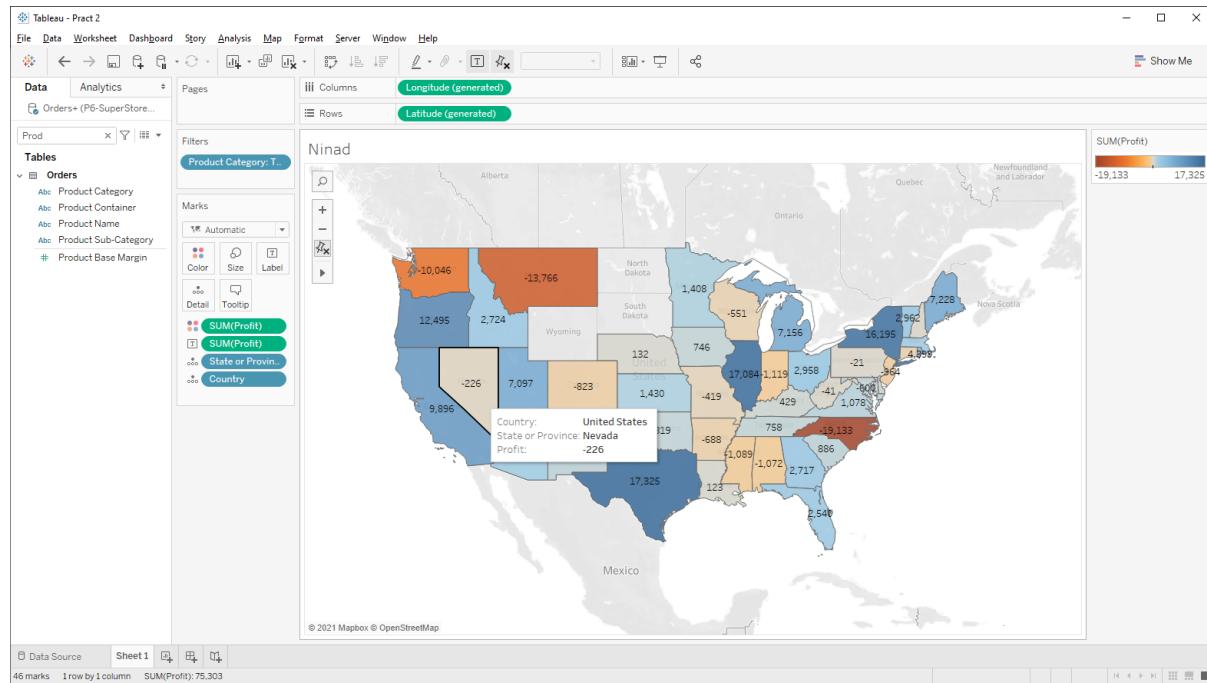
Ans:



Q2: Show Profit Ratio of each state as tooltip on map.

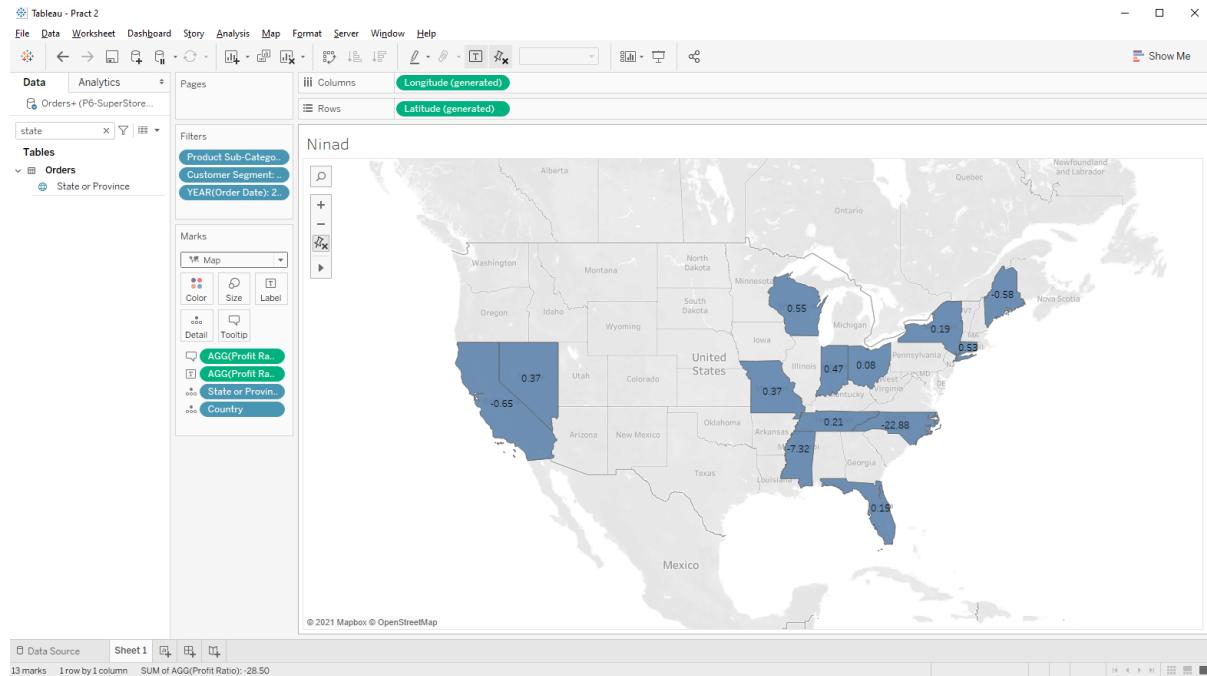
Ans:



Q 3: Show Profit ratio for Grip Envelop products.**Ans:****Q 4: In the technology product category which unprofitable state is surrounded by only profitable states.****Ans:**

Q 5: Which state has the worst Gross Profit Ratio on Envelopes in the Corporate Customer Segment that were Shipped in 2015?

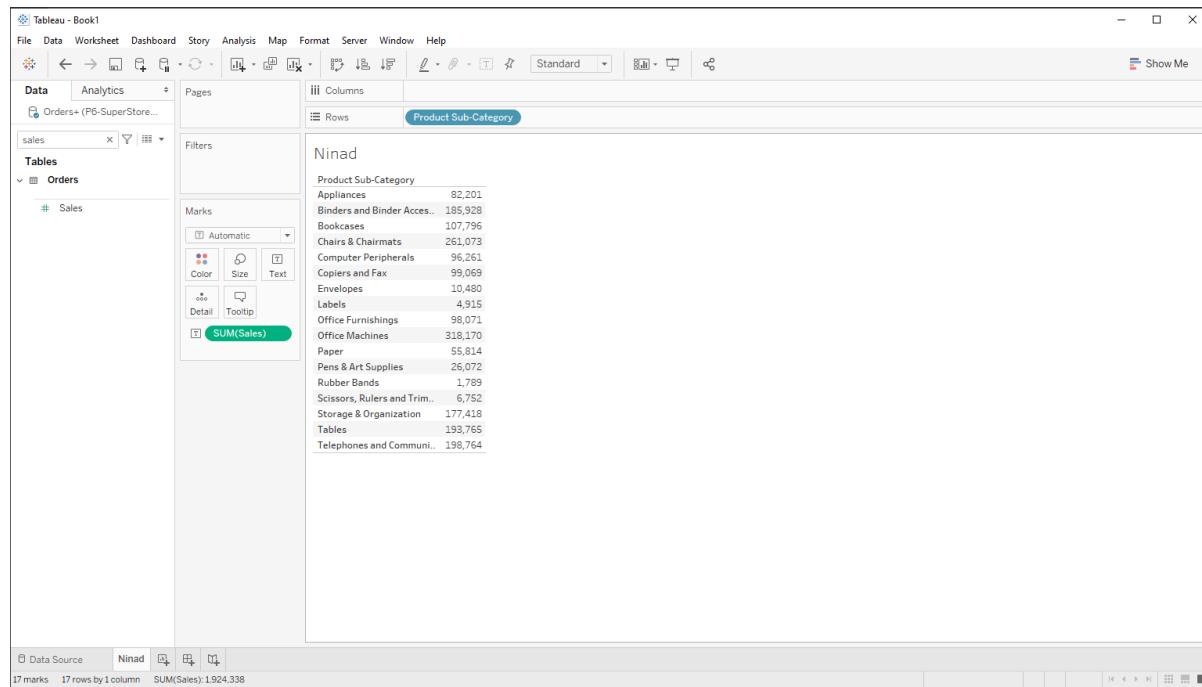
Ans:



Practical No 1
Assignment 3: Preparing Reports

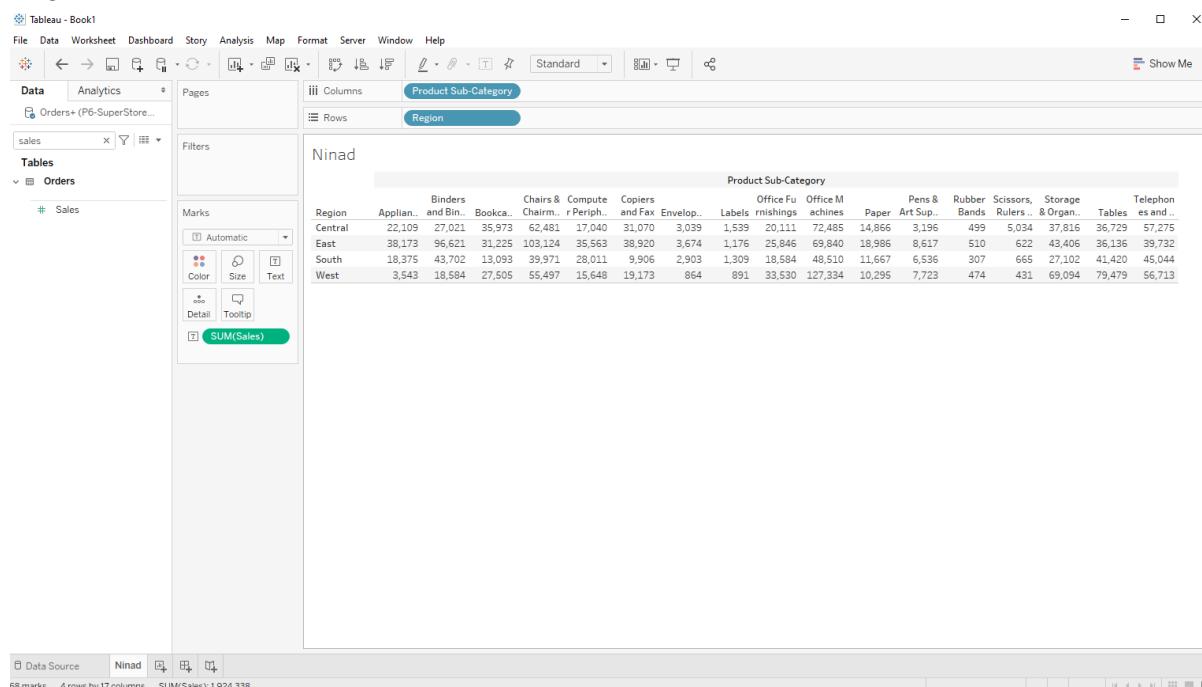
Q1: Prepare a report showing product category wise sales.

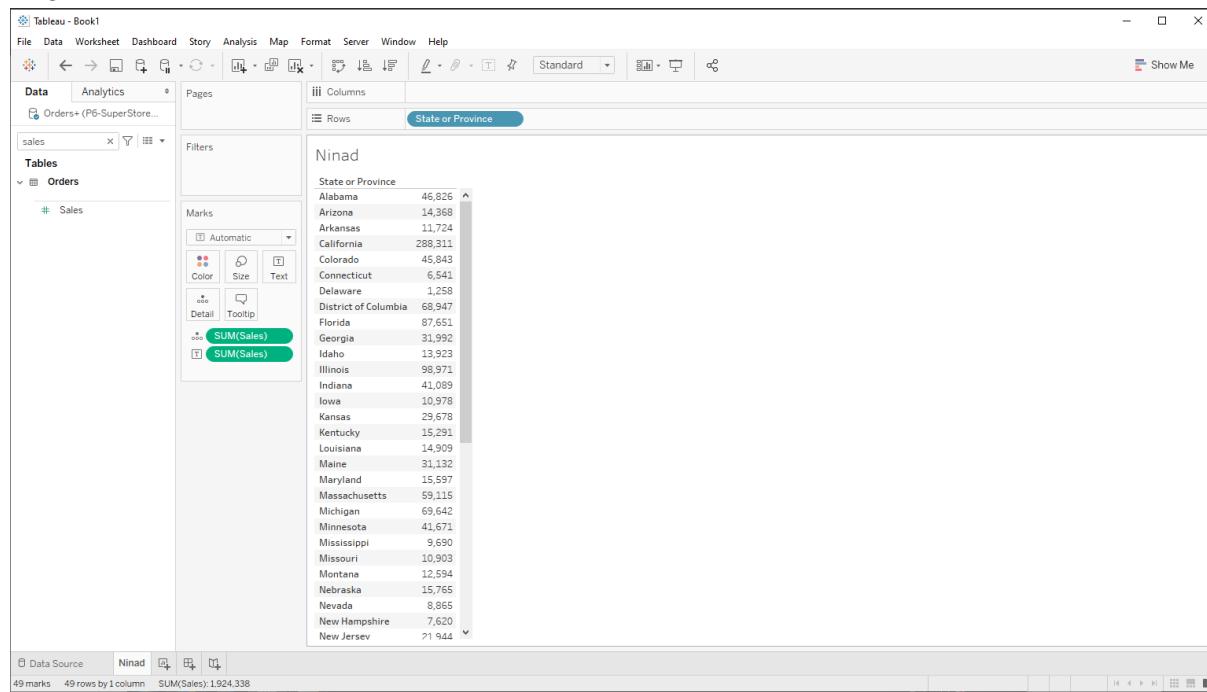
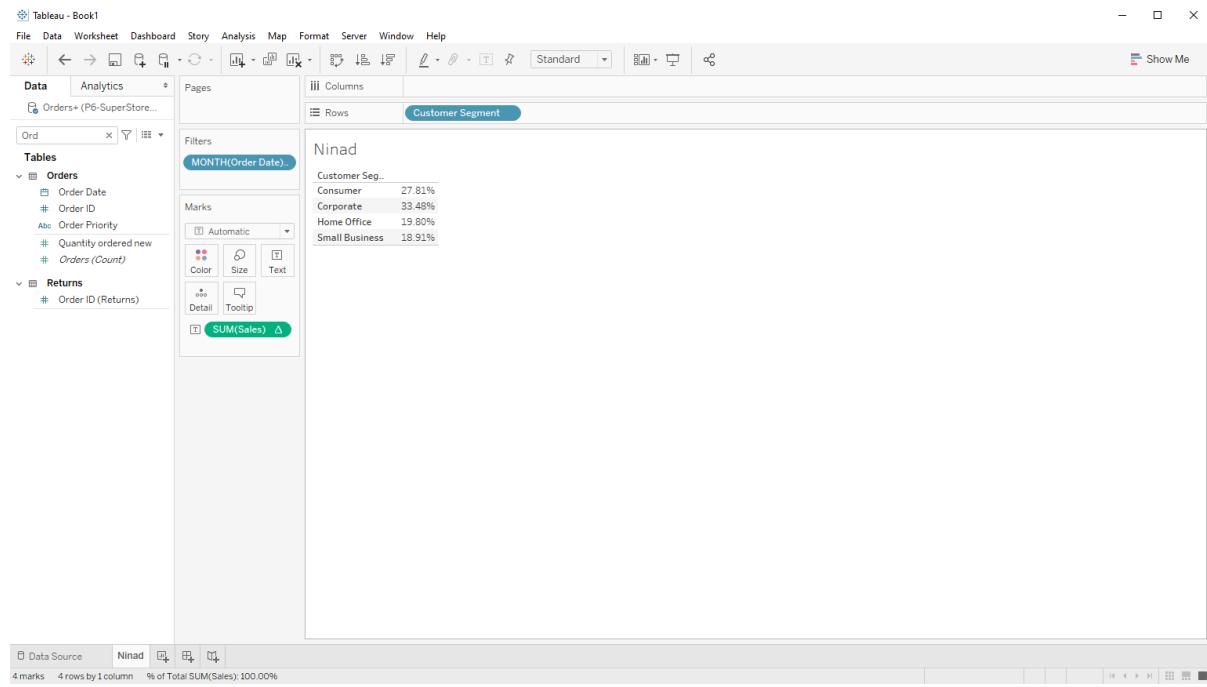
Ans:



Q2: Report showing region wise product wise sales.

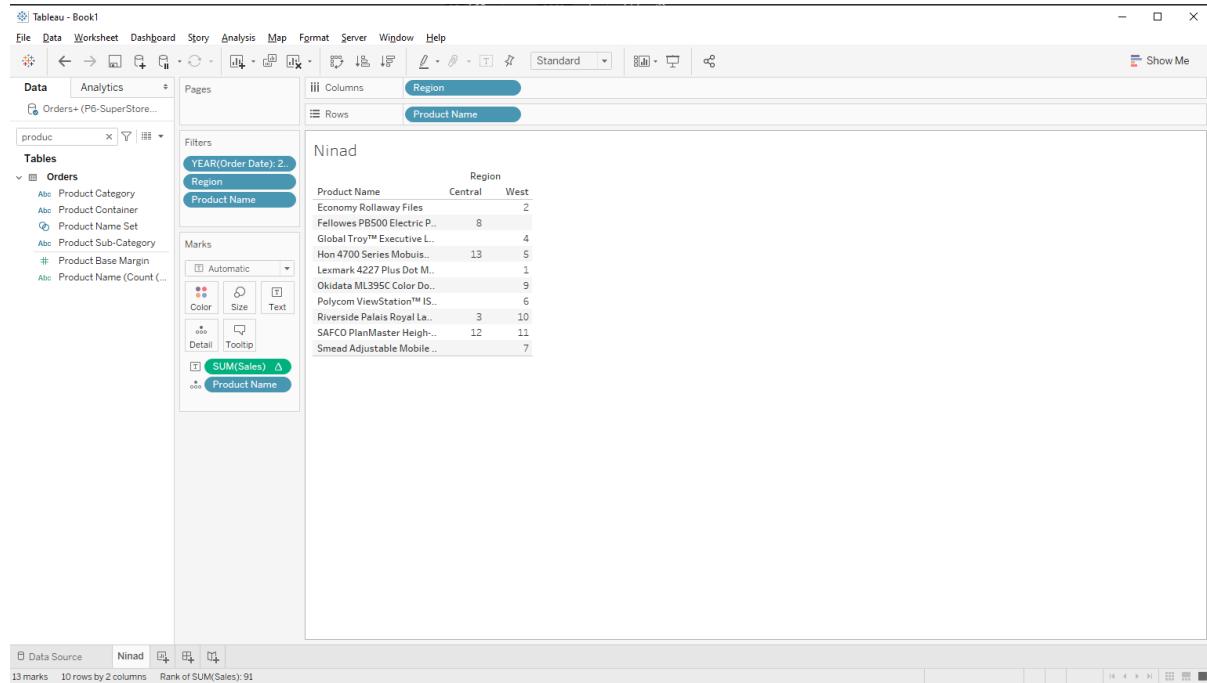
Ans:



Q 3: Report showing state wise sales.**Ans:****Q 4: What is the percent of total Sales for the 'Home Office' Customer Segment in July of 2014?****Ans:**

Q 5: Find the top 10 Product Names by Sales within each region. Which product is ranked #2 in both the Central & West regions in 2015?

Ans:



Practical No 2
DATA WAREHOUSING USING PENTAHO

Data Warehousing

Data warehousing is the process of constructing and using a data warehouse. A data warehouse is constructed by integrating data from multiple heterogeneous sources that support analytical reporting, structured and/or ad hoc queries, and decision making. Data warehousing involves data cleaning, data integration, and data consolidations.

The following are the functions of data warehouse tools and utilities -

- **Data Extraction** – Involves gathering data from multiple heterogeneous sources.
- **Data Cleaning** – Involves finding and correcting the errors in data.
- **Data Transformation** – Involves converting the data from legacy format to warehouse format.
- **Data Loading** – Involves sorting, summarizing, consolidating, checking integrity, and building indices and partitions.
- **Refreshing** – Involves updating from data sources to warehouse.

Who needs Data warehouse?

DWH (Data warehouse) is needed for all types of users like:

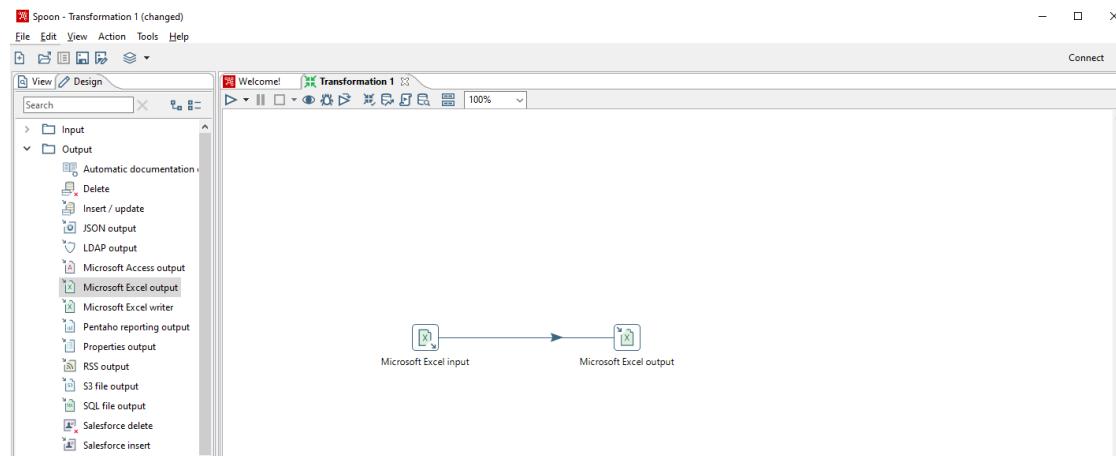
- Decision makers who rely on mass amount of data.
- Users who use customized, complex processes to obtain information from multiple data sources.
- It is also used by the people who want simple technology to access the data.
- It is also essential for those people who want a systematic approach for making decisions.
- If the user wants fast performance on a huge amount of data which is a necessity for reports, grids or charts, then Data warehouse proves useful.
- Data warehouse is a first step if you want to discover 'hidden patterns' of data-flows and groupings.

Advantages of Data Warehouse (DWH):

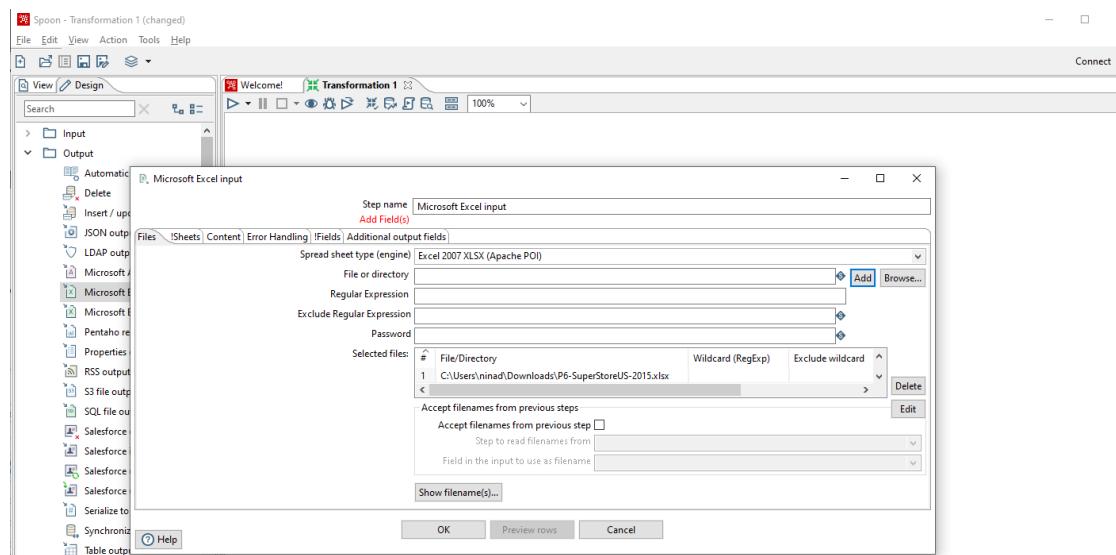
- Data warehouse allows business users to quickly access critical data from some sources all in one place.
- Data warehouse provides consistent information on various cross-functional activities. It is also supporting ad-hoc reporting and query.
- Data Warehouse helps to integrate many sources of data to reduce stress on the production system.
- Data warehouse helps to reduce total turnaround time for analysis and reporting.
- Restructuring and Integration make it easier for the user to use for reporting and analysis.
- Data warehouse allows users to access critical data from the number of sources in a single place. Therefore, it saves user's time of retrieving data from multiple sources.
- Data warehouse stores a large amount of historical data. This helps users to analyse different time periods and trends to make future predictions.

A) Microsoft Excel Input to Microsoft Excel Output

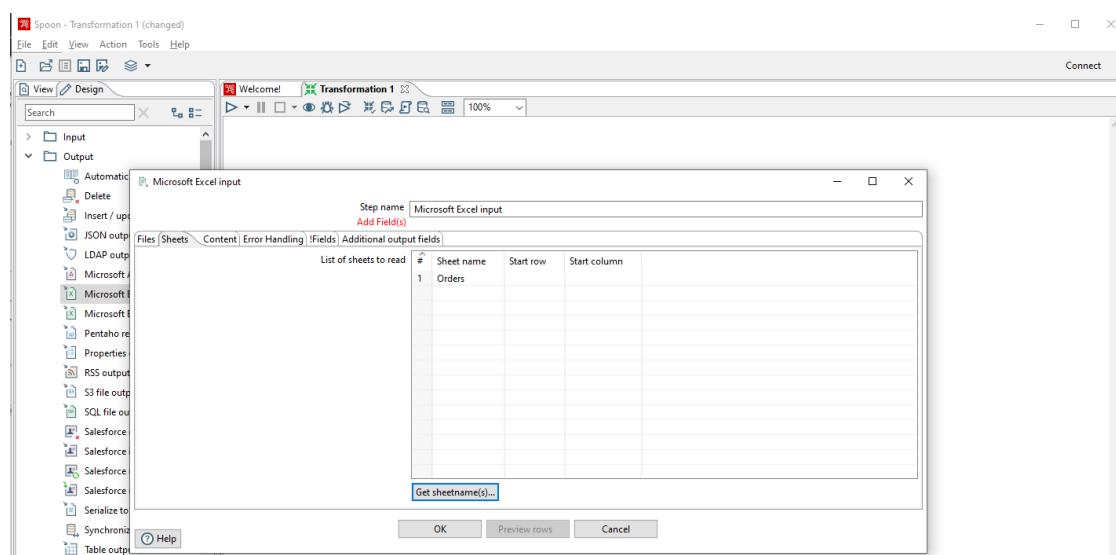
- 1) Drag and drop Microsoft Excel Input from input field and Microsoft Excel Output from output field. Create a hop between them.



- 2) Double click on Microsoft Excel Input and browse the excel file and click on add.



- 3) Click on Sheets -> Get sheet name. Select the Orders table.



4) Click on Fields -> Get Field Name. The field names should appear as shown below.

#	Name	Type	Length	Precision	Trim type	Repeat	Format	Currency	Decimal	Grouping
1	Row ID	Number			none	N				
2	Order Priority	String			none	N				
3	Discount	Number			none	N				
4	Unit Price	Number			none	N				
5	Shipping Cost	Number			none	N				
6	Customer ID	Number			none	N				
7	Customer Name	String			none	N				
8	Ship Mode	String			none	N				
9	Customer Segment	String			none	N				
10	Product Category	String			none	N				
11	Product Sub-Category	String			none	N				
12	Product Container	String			none	N				
13	Product Name	String			none	N				
14	Product Base Margin	Number			none	N				

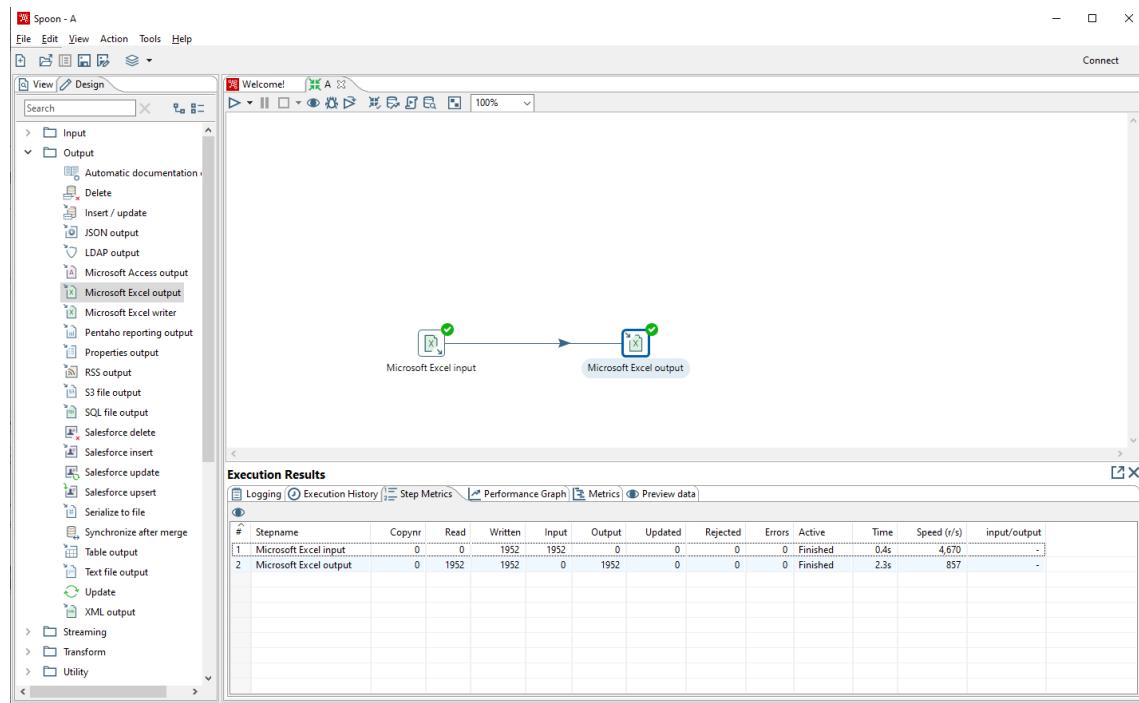
5) Preview the rows.

#	Row ID	Order Priority	Discount	Unit Price	Shipping Cost	Customer ID	Customer Name	Ship Mode	Customer Segment	Product Category	Product Sub-Category	Product Container	Product Name
1	208470	High	0.01	2.84	0.93	3.0	Bonne Potte	Express Air	Corporate	Office Supplies	Pens & Art Supplies	Wrap Bag	SARFORD Liquid Accent™ Tank-Style Highlighters
2	202048	Not Specified	0.02	500.00	502.93	5.0	Bonnie Dumb	Delivery Truck	Home Office	Furniture	Chair & Chairparts	Jumbo Box	Globe Executive Leather Low-Back Tilt
3	207610	Critical	0.06	9.48	7.29	11.0	Marion Dunlap	Regular Air	Small Business	Furniture	Office Furnishings	Small Pack	DAT Two-Tone Rosewood/Black Document Frame, Desktop, 5 x 7
4	248440	Medium	0.09	78.49	15.99	14.0	Gwendolyn F Tyson	Regular Air	Small Business	Furniture	Office Furnishings	Small Box	Howard Miller 12-3/4 Diameter Accurace 05™ Wall Clock
5	248460	Medium	0.08	2.03	2.31	14.0	Gwendolyn F Tyson	Regular Air	Small Business	Office Supplies	Pens & Art Supplies	Wrap Bag	Newell 321
6	248470	Medium	0.05	3.28	4.2	14.0	Gwendolyn F Tyson	Regular Air	Small Business	Office Supplies	Pens & Art Supplies	Wrap Bag	Newell 351
7	248480	Medium	0.05	3.38	1.63	14.0	Gwendolyn F Tyson	Regular Air	Small Business	Office Supplies	Rubber Bands	Wrap Bag	OIC Colored Binder Clips, Assorted Sizes
8	208110	Critical	0.02	4.42	4.97	15.0	Timothy Rees	Regular Air	Small Business	Office Supplies	Pens & Art Supplies	Small Box	Globe Executive Leather Low-Bac
9	208080	Medium	0.01	25.84	6.64	15.0	Timothy Rees	Regular Air	Small Business	Office Supplies	Envelopes	Small Box	Tyvek # Top-Opening Peel & Seal Envelopes, Gray
10	262670	High	0.04	2.86	1.58	16.0	Sarah Ramsey	Regular Air	Small Business	Furniture	Rubber Bands	Wrap Bag	Staples Gold Paper Clips
11	262680	High	0.05	115.99	2.5	16.0	Sarah Ramsey	Regular Air	Small Business	Technology	Telephones and Communication	Small Box	STATAC 7797
12	238900	High	0.05	26.48	6.93	18.0	Laura Hanna	Regular Air	Small Business	Furniture	Office Furnishings	Small Box	DAX Natural Wood-Tone Poster Frame
13	240630	Not Specified	0.07	12.99	9.44	19.0	Jim Rodgers	Regular Air	Small Business	Technology	Office Machines	Medium Box	Hewlett Packard 85 Scientific Calculator
14	240640	Low	0.05	26.48	6.93	19.0	Jim Rodgers	Regular Air	Small Business	Technology	Office Machines	Small Box	DAX Natural Wood-Tone Poster Frame
15	49820	Not Specified	0.08	5.00	3.39	21.0	Tony Wilkins Winters	Regular Air	Small Business	Office Supplies	Rubber Bands	Wrap Bag	Advantek Plastic Paper Clips
16	49830	Not Specified	0.07	12.99	9.44	21.0	Tony Wilkins Winters	Regular Air	Small Business	Technology	Office Machines	Medium Box	Hewlett Packard 85 Scientific Calculator
17	206310	High	0.06	55.48	14.3	24.0	Edna Thomas	Regular Air	Corporate	Office Supplies	Paper	Small Box	Xerox 100
18	206320	High	0.02	1.68	1.57	24.0	Edna Thomas	Regular Air	Corporate	Office Supplies	Pens & Art Supplies	Wrap Bag	Newell 323
19	239670	Not Specified	0.04	4.14	6.6	27.0	Guy Gallagher	Regular Air	Corporate	Furniture	Office Furnishings	Small Box	Elden Home Series Black Desk Accessories
20	239680	Low	0.08	34.49	7.73	32.0	Matthew Berman	Regular Air	Corporate	Office Supplies	Pens & Art Supplies	Small Box	Handi-Sure # Vacuum Mount K5 Pencil Sharpener
21	238120	High	0.01	17.00	8.31	32.0	Matthew Berman	Regular Air	Corporate	Technology	Office Machines	Medium Box	SkatAC 3000
22	232780	Medium	0.09	125.99	7.68	32.0	Matthew Berman	Express Air	Corporate	Technology	Telephones and Communication	Small Box	TimeportTP382
23	193550	Low	0.06	205.99	8.99	32.0	Matthew Berman	Regular Air	Corporate	Technology	Telephones and Communication	Small Box	StoneDunTech Recycled Plastic Frosted Binders
24	236540	Not Specified	0.03	4.24	5.41	33.0	Ricky Hersley	Regular Air	Corporate	Office Supplies	Binders and Binder Accessories	Small Box	Newell 338
25	236550	Not Specified	0.04	2.94	0.7	33.0	Ricky Hersley	Regular Air	Corporate	Office Supplies	Pens & Art Supplies	Wrap Bag	XTERRA 220 Dual Handset Phone With Caller ID/Call Waiting
26	236560	Low	0.03	98.49	19.99	43.0	Edna Thomas	Regular Air	Corporate	Office Supplies	Pens & Art Supplies	Small Box	6160
27	185310	Not Specified	0.0	115.99	2.5	52.0	Lorraine Kelly	Regular Air	Corporate	Technology	Telephones and Communication	Small Box	Okidata Pacemark 4410N Wide Format Dot Matrix Printer
28	221170	Critical	0.07	3502.14	52.0	53.0	Sidney Russell Austin	Delivery Truck	Corporate	Technology	Office Machines	Jumbo Box	Xerox 1003
29	185320	Not Specified	0.02	5.00	5.79	53.0	Sidney Russell Austin	Regular Air	Corporate	Office Supplies	Paper	Small Box	Durable Preboard Binders
30	206970	Medium	0.06	3.8	1.49	56.0	Randal Montgomery	Regular Air	Consumer	Office Supplies	Binders and Binder Accessories	Small Box	Newell 310
31	206980	Medium	0.06	1.76	0.7	56.0	Randal Montgomery	Regular Air	Consumer	Office Supplies	Pens & Art Supplies	Wrap Bag	Xerox 193
32	228900	High	0.02	5.98	3.15	62.0	Pam Gilbert	Regular Air	Corporate	Office Supplies	Paper	Small Box	

6) Double click on Microsoft Excel output. Click on Browse and save the output file.

File	Content	Custom	Fields
Filename	C:\Users\ninad\Downloads\file.xls <input type="button" value="Browse..."/>		
Create Parent folder	<input type="checkbox"/>		
Do not create file at start	<input type="checkbox"/>		
Extension	xls <input style="width: 20px; height: 20px; vertical-align: middle;" type="button" value="..."/>		
Include stepnr in filename?	<input type="checkbox"/>		
Include date in filename?	<input type="checkbox"/>		
Include time in filename?	<input type="checkbox"/>		
Specify Date/time format	<input style="width: 150px; height: 20px; vertical-align: middle;" type="text"/>		
Date/time format	<input style="width: 20px; height: 20px; vertical-align: middle;" type="button" value="..."/>		
Show filename(s)...	<input style="width: 100px; height: 20px; vertical-align: middle;" type="button" value="..."/>		
Add filenames to result	<input checked="" type="checkbox"/>		

7) Run the transformation.



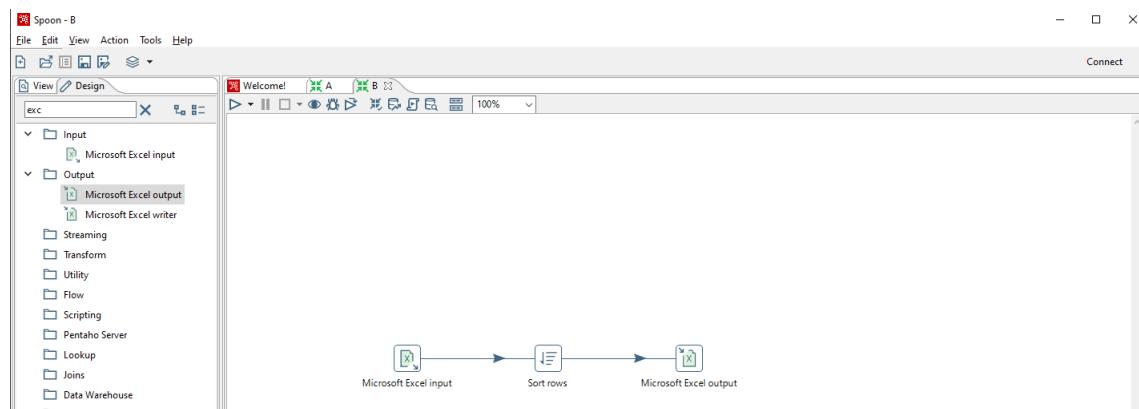
8) The output file should appear as shown below.

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V
1 Row ID	Order Pno/Discount	Customer	Customer	Ship Mode	Customer	Product C	Product S	Product Cr	Product N	Product Bi	Country	Region	State or City	Postal Co	Order Date	Ship Date	Profit				
2 20,847.00	High	01	2.84	.93	3.00	Bonnie Po Express	A Corporate	Office Sup	Pens & Ar Wrap Bag	SAIFORC	United Sta West	Washington	Anacortes	98,221.00	#####	#####				4.	
3 20,228.00	Not Specif	02	500.98	26.00	5.00	Ronnie Pro Delivery	Tr Home Office	Furniture	Chains & Jumbo Dru	Global Troy	United Sta West	California	San Gabri	91,776.60	#####	#####				4,390.	
4 21,776.00	Critical	06	9.48	7.29	11.00	Marcus Du Regular	Ai Home Office	Furniture	Office Furr Small Pac	DAX Two-T	United Sta East	New Jersey	Roselle	7,203.00	#####	#####				53.	
5 24,844.00	Medium	09	78.69	19.99	14.00	Gwendolyr	Regular	Ai Small Bus Furniture	Office Furr Small Box	Howard Mi	United Sta Central	Minnesota	Prior Lake	55,372.00	#####	#####				803.	
6 24,846.00	Medium	08	3.28	2.31	14.00	Gwendolyr	Regular	Ai Small Bus Furniture	Office Sup	Pens & Ar Wrap Bag	Newell 321	United Sta Central	Minnesota	Prior Lake	55,372.00	#####	#####			24.	
7 24,847.00	Medium	05	3.28	4.20	14.00	Gwendolyr	Regular	Ai Small Bus Furniture	Office Sup	Pens & Ar Wrap Bag	Newell 351	United Sta Central	Minnesota	Prior Lake	55,372.00	#####	#####			37.	
8 24,848.00	Medium	05	3.58	1.63	14.00	Gwendolyr	Regular	Ai Small Bus Office Sup	Rubber Ba Wrap Bag	OIC Colors	United Sta Central	Minnesota	Prior Lake	55,372.00	#####	#####					
9 18,181.00	Critical	00	4.42	4.99	15.00	Timothy R	Regular	Ai Small Bus Office Sup	Envelopes	Small Box Grip Seal f	United Sta East	New York	Smithtown	11,787.00	#####	#####				-59.	
10 20,925.00	Medium	01	35.94	6.66	15.00	Timothy R	Regular	Ai Small Bus Office Sup	Envelopes	Small Box Tyek	United Sta East	New York	Smithtown	11,787.00	#####	#####				261.	
11 26,267.00	High	04	2.98	1.58	16.00	Sarah Ran	Regular	Ai Small Bus Office Sup	Rubber Ba Wrap Bag	Staples G	United Sta East	New York	Syracuse	13,210.00	#####	#####				2.	
12 26,268.00	High	05	115.99	2.50	16.00	Sarah Ran	Regular	Ai Small Bus Technology	Telephone:Small Box StarTAC	7	United Sta East	New York	Syracuse	13,210.00	#####	#####				652.	
13 23,890.00	High	05	26.48	6.93	18.00	Laurie Han	Regular	Ai Small Bus Furniture	Office Furr	Small Box DAX Natur	United Sta West	Montana	Helena	59,601.00	#####	#####				314.	
14 24,053.00	Not Specif	07	12.99	9.44	19.00	Jim Rodge	Regular	Ai Small Bus Technology	Office Mac	Medium Bi-Hewlett Pa	United Sta West	Montana	Missoula	59,801.00	#####	#####				-114.	
15 5,890.00	High	05	26.48	6.93	21.00	Tony Wilki	Regular	Ai Small Bus Furniture	Office Furr	Small Box DAX Natur	United Sta East	New York	New York	10,012.00	#####	#####				384.	
16 6,052.00	Not Specif	08	5.00	3.39	21.00	Tony Wilki	Regular	Ai Small Bus Office Sup	Rubber Ba Wrap Bag	Advantus f	United Sta East	New York	New York	10,012.00	#####	#####				-17.	
17 6,063.00	Not Specif	07	12.99	9.44	21.00	Tony Wilki	Regular	Ai Small Bus Technology	Office Mac	Medium Bi-Hewlett Pa	United Sta East	New York	New York	10,012.00	#####	#####				-114.	
18 20,631.00	High	06	55.48	14.30	24.00	Edna Thon	Regular	Ai Corporate	Office Sup Paper	Small Box Xerox 194	United Sta West	California	Laguna Nt	92,677.00	#####	#####				28.	
19 20,632.00	High	02	1.68	1.57	24.00	Edna Thon	Regular	Ai Corporate	Office Sup	Pens & Ar Wrap Bag	Newell 323	United Sta West	California	Laguna Nt	92,677.00	#####	#####			-5.	
20 23,967.00	Not Specif	04	4.14	6.60	27.00	Guy Gallar	Regular	Ai Corporate	Furniture	Office Furr Small Box Eldon Imag	United Sta West	California	Lakewood	90,712.00	#####	#####				8.	
21 23,509.00	High	08	34.99	7.73	32.00	Matthew B	Regular	Ai Corporate	Office Sup Pens & Ar Small Box Hunt Bost	Technolog:Office Mac	Medium Bi-Cannon P1-	United Sta West	Oregon	Grants Pa:	97,525.00	#####	#####			144.	
22 23,612.00	High	01	17.98	8.51	32.00	Matthew B	Regular	Ai Corporate	Technolog:Office Mac	Medium Bi-Hewlett Pa	United Sta West	Oregon	Grants Pa:	97,525.00	#####	#####			-35.		
23 27,278.00	Medium	09	125.99	7.69	32.00	Matthew B	Express	A Corporate	Technolog:Telephone:Small Box StartAC 3	Technolog:Telephone:Small Box StartAC 3	United Sta West	Oregon	Grants Pa:	97,525.00	#####	#####			210.		
24 19,355.00	Low	06	205.99	8.99	32.00	Matthew B	Regular	Ai Corporate	Technolog:Telephone:Small Box TimeportP	Technolog:Telephone:Small Box TimeportP	United Sta West	Oregon	Grants Pa:	97,526.00	#####	#####			3,568.		
25 23,654.00	Not Specif	03	4.24	5.41	33.00	Ricky Hen	Regular	Ai Corporate	Office Sup Binders	an Small Box Storex Dur	United Sta West	Oregon	Gresham	97,030.00	#####	#####			-84.		
26 23,655.00	Not Specif	04	2.94	.70	33.00	Ricky Hen	Regular	Ai Corporate	Office Sup Pens & Ar Wrap Bag	Newell 338	United Sta West	Oregon	Gresham	97,030.00	#####	#####			24.		
27 25,933.00	High	00	99.99	19.99	43.00	Theodore I	Regular	Ai Consumer	Technolog:Office Mac	Small Box AT&T 230	United Sta West	Washington	Redmond	98,052.00	#####	#####			25.		

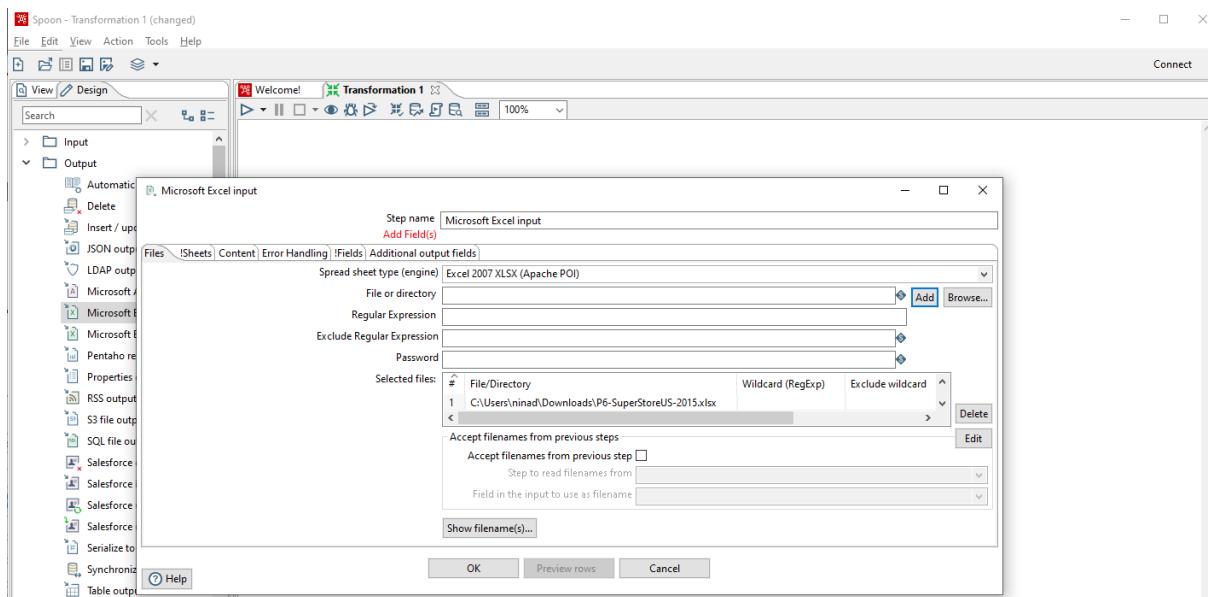
B) Applying transformations to rows.

1) Sort

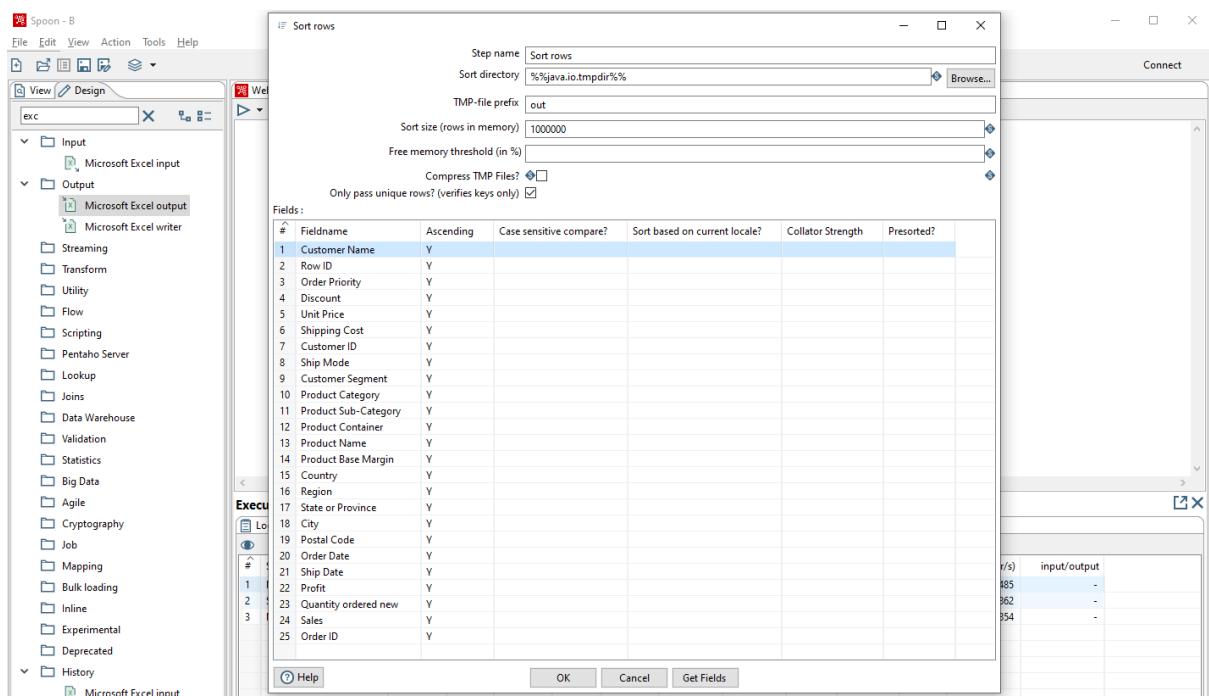
1) Drag and drop Microsoft Excel Input, Microsoft Excel Output and Sort from Transform field.



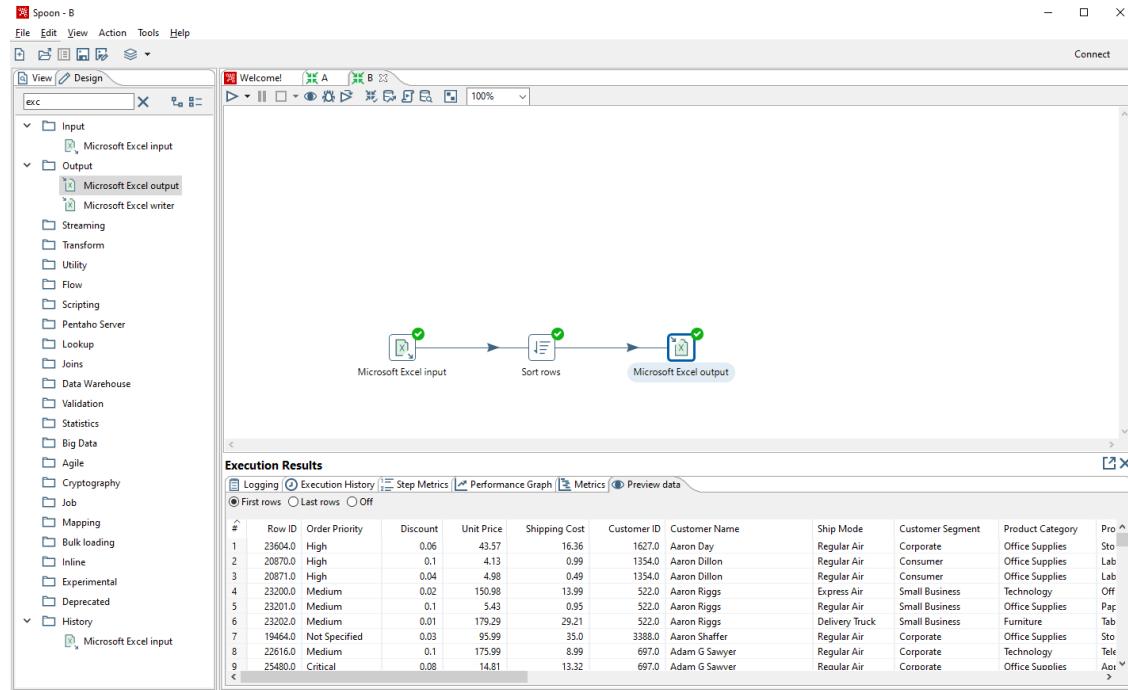
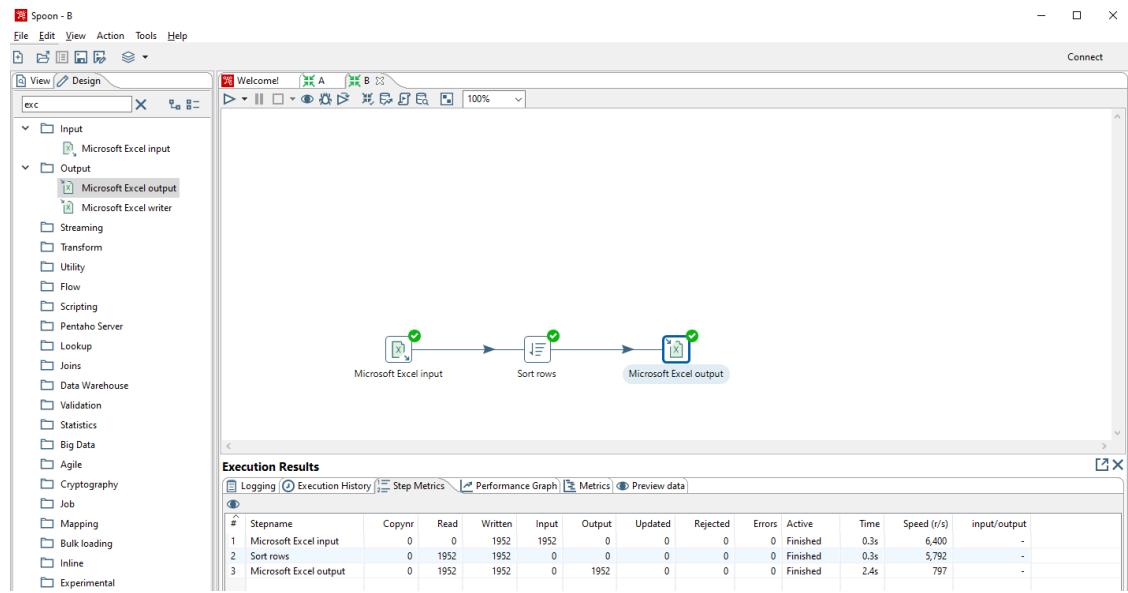
- 2) Set up the input file by choosing the appropriate Excel file. Choose Orders table and get field names. Preview rows.



- 3) Double click on Sort -> Get Fields. Move up the Customer Names field.



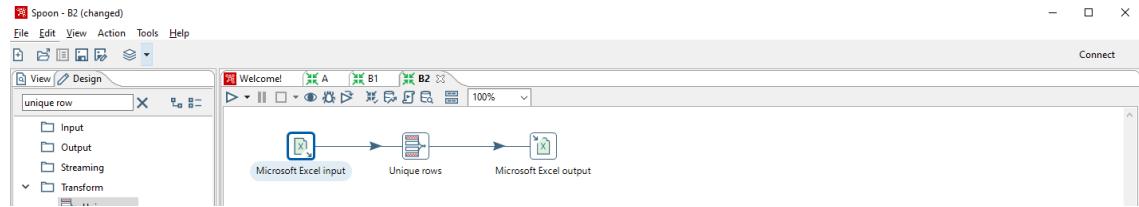
4) Double click on the output file and save it. Run the transformation.



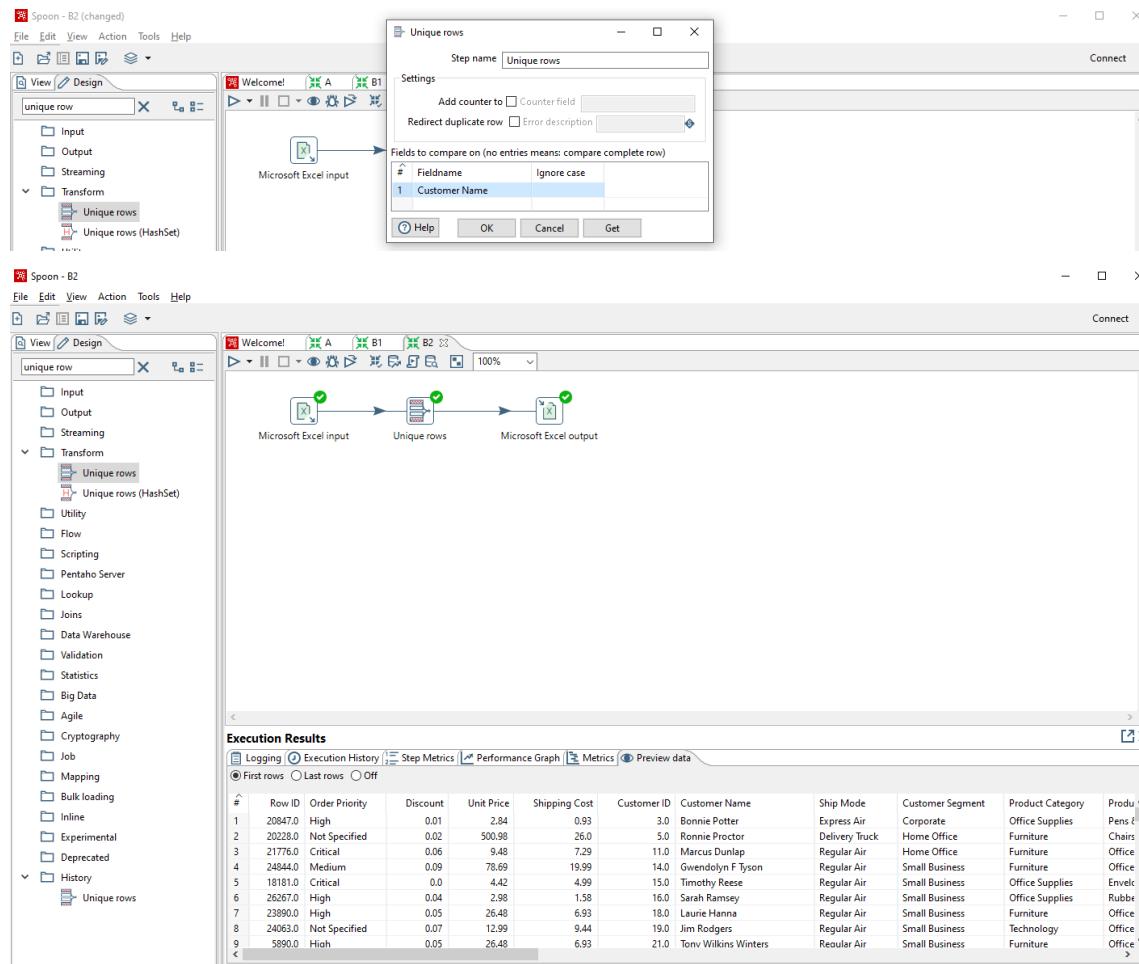
Row ID	Order Priority	Unit Price	Shipping Cost	Customer	Ship Mode	Customer Product C/Predict S/	Product I/Predict N/	Product R/Predict P/	Product U/Predict U/	Product Y/Predict Y/	Country	Region	State or P/City	Postal Code	Order Date	Ship Date	Phone	Quantity	Sales	Order ID
23,504.00	High	06	43.57	16.36	1,627	Aaron Day	Regular Air	Corporate	Office Supp Storage & Small Box Transf	File	55 United Sta	South	Tennessee Greeneville	37743	05/26/2010	05/27/2010	-	38.81	17.00	710 16 90,602.00
20,886.00	High	07	43.13	16.99	1,627	Aaron Day	Regular Air	Corporate	Office Supp Labels	Small Box Avery	55 United Sta	South	Tennessee Weatherford	37770	05/26/2010	05/27/2010	-	2.00	8.30	11,209.00
40,771.00	High	04	4.99	1.49	1,304.00	Aaron Dillon	Regular Air	Consumer	Office Supp Labels	Small Box Avery	55 United Sta	South	Tennessee Weatherford	37770	05/26/2010	05/27/2010	-	4.41	1.00	10,000.00
5,230.00	Medium	02	150.98	13.99	522.00	Aaron Rigg	Express Air	Mac Small Bus	Technology Bus Logos	Small Bus	38 United Sta	West	Oregon Redmond	97756	05/26/2010	05/27/2010	-	26.10	3.00	480 37 89,327.00
23,201.00	Medium	10	5.43	0.95	522.00	Aaron Rigg	Express Air	Small Business	Technology Bus Logos	Small Bus	38 United Sta	West	Oregon Redmond	97756	05/26/2010	05/27/2010	-	5.76	89,327.00	Technology
23,201.00	Medium	01	5.43	0.95	522.00	Aaron Rigg	Express Air	Small Business	Technology Bus Logos	Small Bus	38 United Sta	West	Oregon Redmond	97756	05/26/2010	05/27/2010	-	2.86	21.30	3,120.00
8,194.60	Not Specified	03	95.99	35.00	3,398.00	Aaron Sha	Regular Air	Corporate	Office Supp Storage & Large Box Safco Indu	Small Box	57 United Sta	Central	Illinois Chicago	60631	05/26/2010	05/27/2010	-	67.01	9.00	894 64 90,154.00
22,116.00	Medium	10	175.99	8.99	697.00	Adam G Sawyer	Regular Air	Corporate	Office Supp Storage & Large Box Safco Indu	Small Box	57 United Sta	Central	Indiana East Chic	46312	05/26/2010	05/27/2010	-	928.96	10.00	1,346.32
24,876.00	Critical	08	1,150.98	16.36	1,627	Adam G Sawyer	Regular Air	Corporate	Office Supp Appliances Small Box Holmes Re	Small Box	55 United Sta	South	Tennessee Memphis	38032	05/26/2010	05/27/2010	-	13.20	26.00	480 37 87,602.00
21,156.00	Medium	04	4.91	1.50	1,312.00	Adam O S Regular	Regular Air	Corporate	Office Supp Appliances Small Box Holmes Re	Small Box	36 United Sta	Central	Missouri Raytown	64133	05/26/2010	05/27/2010	-	95.05	2.00	480 37 87,602.00
26,959.00	Critical	06	161.55	19.99	674.00	Adam O S Regular	Regular Air	Corporate	Office Supp Storage & Small Box Fellowes 5	Small Box	66 United Sta	Central	Texas Weatherford	76102	05/26/2010	05/27/2010	-	7.58	3.00	485 01 188,174.00
22,638.00	Low	09	100.98	26.64	940.00	Albert Mac	Delivery Tr	Home Office Furniture	Postcard Ambro Bob Wht	Small Box	62 United Sta	East	Connectics New Milw	53777	05/26/2010	05/27/2010	-	193.58	4.00	396 19 90,644.00
21,747.00	Not Specified	09	100.98	26.64	940.00	Albert Mac	Delivery Tr	Home Office Furniture	Postcard Ambro Bob Wht	Small Box	76 United Sta	East	Connectics New Milw	53777	05/26/2010	05/27/2010	-	11.00	13,000.00	Postcard Ambro Bob Wht
15,217.00	Not Specified	00	2.88	79	1,814.00	Albert Rigg	Regular Air	Home Office	Office Supp Pens & Ar/Wrap Bag Newell 34C	Small Box	56 United Sta	South	Mississippi Olive Branch	38654	05/26/2010	05/27/2010	-	141.67	13.00	30 06 90,524.00
16,182.00	High	07	15.94	1.57	206.00	Albert Rigg	Regular Air	Corporate	Office Supp Pens & Ar/Wrap Bag Newell 34C	Small Box	59 United Sta	Central	Texas Gainesville	76240	05/26/2010	05/27/2010	-	152.67	23.00	221 24 88,361.00
18,322.00	High	07	15.94	1.57	206.00	Albert Rigg	Regular Air	Corporate	Office Supp Pens & Ar/Wrap Bag Newell 34C	Small Box	59 United Sta	Central	Texas Gainesville	76240	05/26/2010	05/27/2010	-	46.88	48.00	88,361.00
18,322.00	High	07	15.94	1.57	206.00	Albert Rigg	Regular Air	Corporate	Office Supp Pens & Ar/Wrap Bag Newell 34C	Small Box	59 United Sta	Central	Texas Gainesville	76240	05/26/2010	05/27/2010	-	29.15	21.30	3,120.00
18,322.00	High	07	15.94	1.57	206.00	Albert Rigg	Regular Air	Corporate	Office Supp Pens & Ar/Wrap Bag Newell 34C	Small Box	59 United Sta	Central	Texas Gainesville	76240	05/26/2010	05/27/2010	-	20.30	14.41	2,640.00
14,225.00	Critical	08	9.11	2.11	1,163.00	Albert Rigg	Regular Air	Corporate	Office Supp Pens & Ar/Wrap Bag Newell 34C	Small Box	52 United Sta	East	Connectics New Milw	53777	05/26/2010	05/27/2010	-	102.52	11.00	157.27 90,653.00
24,876.00	Critical	08	3.68	99	3,106.00	Alexander Regular	Regular Air	Corporate	Office Supp Storage & Large Box Safco Indu	Small Box	52 United Sta	East	Connectics New Milw	53777	05/26/2010	05/27/2010	-	102.52	11.00	157.27 90,653.00
24,876.00	Critical	08	3.68	99	3,106.00	Alexander Regular	Regular Air	Corporate	Office Supp Storage & Large Box Safco Indu	Small Box	52 United Sta	East	Connectics New Milw	53777	05/26/2010	05/27/2010	-	2.45	9.00	19 12 89,679.00
23,533.00	Critical	09	2.18	79	1,603.00	Alex Nichi	Regular Air	Consumer	Office Supp Storage & Large Box Safco Indu	Small Box	52 United Sta	East	Connectics New Milw	53777	05/26/2010	05/27/2010	-	29.15	10.00	157.27 90,653.00
23,533.00	Critical	09	2.18	79	1,603.00	Alex Nichi	Regular Air	Consumer	Office Supp Storage & Large Box Safco Indu	Small Box	52 United Sta	East	Connectics New Milw	53777	05/26/2010	05/27/2010	-	2.45	9.00	19 12 89,679.00
23,533.00	Critical	09	2.18	79	1,603.00	Alex Nichi	Regular Air	Consumer	Office Supp Storage & Large Box Safco Indu	Small Box	52 United Sta	East	Connectics New Milw	53777	05/26/2010	05/27/2010	-	29.15	10.00	157.27 90,653.00
23,533.00	Critical	09	2.18	79	1,603.00	Alex Nichi	Regular Air	Consumer	Office Supp Storage & Large Box Safco Indu	Small Box	52 United Sta	East	Connectics New Milw	53777	05/26/2010	05/27/2010	-	2.45	9.00	19 12 89,679.00
23,533.00	Critical	09	2.18	79	1,603.00	Alex Nichi	Regular Air	Consumer	Office Supp Storage & Large Box Safco Indu	Small Box	52 United Sta	East	Connectics New Milw	53777	05/26/2010	05/27/2010	-	29.15	10.00	157.27 90,653.00
23,533.00	Critical	09	2.18	79	1,603.00	Alex Nichi	Regular Air	Consumer	Office Supp Storage & Large Box Safco Indu	Small Box	52 United Sta	East	Connectics New Milw	53777	05/26/2010	05/27/2010	-	2.45	9.00	19 12 89,679.00
23,533.00	Critical	09	2.18	79	1,603.00	Alex Nichi	Regular Air	Consumer	Office Supp Storage & Large Box Safco Indu	Small Box	52 United Sta	East	Connectics New Milw	53777	05/26/2010	05/27/2010	-	29.15	10.00	157.27 90,653.00
23,533.00	Critical	09	2.18	79	1,603.00	Alex Nichi	Regular Air	Consumer	Office Supp Storage & Large Box Safco Indu	Small Box	52 United Sta	East	Connectics New Milw	53777	05/26/2010	05/27/2010	-	2.45	9.00	19 12 89,679.00
23,533.00	Critical	09	2.18	79	1,603.00	Alex Nichi	Regular Air	Consumer	Office Supp Storage & Large Box Safco Indu	Small Box	52 United Sta	East	Connectics New Milw	53777	05/26/2010	05/27/2010	-	29.15	10.00	157.27 90,653.00
23,533.00	Critical	09	2.18	79	1,603.00	Alex Nichi	Regular Air	Consumer	Office Supp Storage & Large Box Safco Indu	Small Box	52 United Sta	East	Connectics New Milw	53777	05/26/2010	05/27/2010	-	2.45	9.00	19 12 89,679.00
23,533.00	Critical	09	2.18	79	1,603.00	Alex Nichi	Regular Air	Consumer	Office Supp Storage & Large Box Safco Indu	Small Box	52 United Sta	East	Connectics New Milw	53777	05/26/2010	05/27/2010	-	29.15	10.00	157.27 90,653.00
23,533.00	Critical	09	2.18	79	1,603.00	Alex Nichi	Regular Air	Consumer	Office Supp Storage & Large Box Safco Indu	Small Box	52 United Sta	East	Connectics New Milw	53777	05/26/2010	05/27/2010	-	2.45	9.00	19 12 89,679.00
23,533.00	Critical	09	2.18	79	1,603.00	Alex Nichi	Regular Air	Consumer	Office Supp Storage & Large Box Safco Indu	Small Box	52 United Sta	East	Connectics New Milw	53777	05/26/2010	05/27/2010	-	29.15	10.00	157.27 90,653.00
23,533.00	Critical	09	2.18	79	1,603.00	Alex Nichi	Regular Air	Consumer	Office Supp Storage & Large Box Safco Indu	Small Box	52 United Sta	East	Connectics New Milw	53777	05/26/2010	05/27/2010	-	2.45	9.00	19 12 89,679.00
23,533.00	Critical	09	2.18	79	1,603.00	Alex Nichi	Regular Air	Consumer	Office Supp Storage & Large Box Safco Indu	Small Box	52 United Sta	East	Connectics New Milw	53777	05/26/2010	05/27/2010	-	29.15	10.00	157.27 90,653.00
23,533.00	Critical	09	2.18	79	1,603.00	Alex Nichi	Regular Air	Consumer	Office Supp Storage & Large Box Safco Indu	Small Box	52 United Sta	East	Connectics New Milw	53777	05/26/2010	05/27/2010	-	2.45	9.00	19 12 89,679.00
23,533.00	Critical	09	2.18	79	1,603.00	Alex Nichi	Regular Air	Consumer	Office Supp Storage & Large Box Safco Indu	Small Box	52 United Sta	East	Connectics New Milw	53777	05/26/2010	05/27/2010	-	29.15	10.00	157.27 90,653.00
23,533.00	Critical	09	2.18	79	1,603.00	Alex Nichi	Regular Air	Consumer	Office Supp Storage & Large Box Safco Indu	Small Box	52 United Sta	East	Connectics New Milw	53777	05/26/2010	05/27/2010	-	2.45	9.00	19 12 89,679.00
23,533.00	Critical	09	2.18	79	1,603.00	Alex Nichi	Regular Air	Consumer	Office Supp Storage & Large Box Safco Indu	Small Box	52 United Sta	East	Connectics New Milw	53777	05/26/2010	05/27/2010	-	29.15	10.00	157.27 90,653.00
23,533.00	Critical	09	2.18	79	1,603.00	Alex Nichi	Regular Air	Consumer	Office Supp Storage & Large Box Safco Indu	Small Box	52 United Sta	East	Connectics New Milw	53777	05/26/2010	05/27/2010	-	2.45	9.00	19 12 89,679.00
23,533.00	Critical	09	2.18	79	1,603.00	Alex Nichi	Regular Air	Consumer	Office Supp Storage & Large Box Safco Indu	Small Box	52 United Sta	East	Connectics New Milw	53777	05/26/2010	05/27/2010	-	29.15	10.00	157.27 90,653.00
23,533.00	Critical	09	2.18	79	1,603.00	Alex Nichi	Regular Air	Consumer	Office Supp Storage & Large Box Safco Indu	Small Box	52 United Sta	East	Connectics New Milw	53777	05/26/2010	05/27/2010	-	2.45	9.00	19 12 89,679.00
23,533.00	Critical	09	2.18	79	1,603.00	Alex Nichi	Regular Air	Consumer	Office Supp Storage & Large Box Safco Indu	Small Box	52 United Sta	East	Connectics New Milw	53777	05/26/2010	05/27/2010	-	29.15	10.00	157.27 90,653.00
23,533.00	Critical	09	2.18	79	1,603.00	Alex Nichi	Regular Air	Consumer	Office Supp Storage & Large Box Safco Indu	Small Box	52 United Sta	East	Connectics New Milw	53777	05/26/2010	05/27/2010	-	2.45	9.00	19 12 89,679.00
23,533.00	Critical	09	2.18	79	1,603.00	Alex Nichi	Regular Air	Consumer	Office Supp Storage & Large Box Safco Indu	Small Box	52 United Sta	East	Connectics New Milw	53777	05/26/2010	05/27/2010	-	29.15	10.00	157.27 90,653.0

2) Unique rows

1) Drag and drop Microsoft Excel Input, Microsoft Excel Output and Unique rows from transform field.



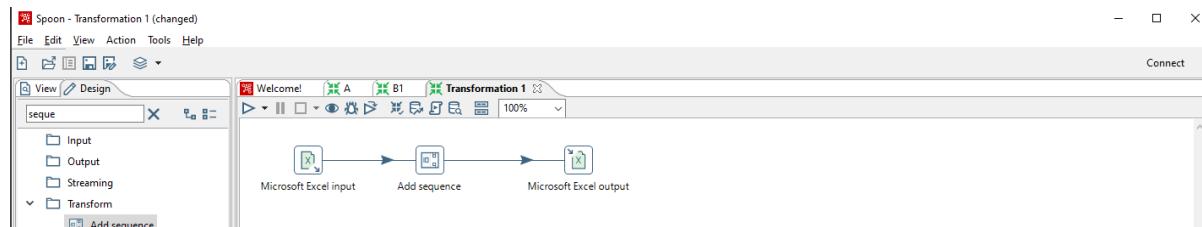
2) In Unique rows, select the column for which unique values need to be displayed. Run the transformation.



A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	AA	AB	AC											
Row ID	Order Priority	Discount	Unit Price	Shipping Cost	Customer	Customer Product C	Product D	Product E	Product F	Product G	Product H	Product I	Product J	Product K	Product L	Product M	Product N	Product O	Product P	Product Q	Product R	Product S	Product T	Product U	Product V	Product W	Product X	Product Y	Product Z	Product AA	Product AB	Product AC							
20,847.0	High	0.01	2.84	0.93	3.0	Bonnie Potter	Express A	Corporate	Office Supply	Pens & Writing	SAF/NCR	54 United States West	Washington/Anacortes	98,221.00	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****				
20,228.0	Not Specified	0.02	500.98	26.0	5.0	Ronnie Proctor	Delivery	Truck	Corporate	Office Supply	SAF/NCR	54 United States West	New Jersey/Garden City	72,203.00	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****				
21,776.0	Critical	0.06	9.48	7.29	11.0	Marcus Dunlap	Regular	Air	Home Office	Office Supply	SAF/NCR	45 United States East	New Jersey/Garden City	63,811.00	22.00	211.15	90,192.00	3.00																					
24,944.0	Medium	0.09	78.69	19.99	14.0	Gwendolyn F Tyson	Regular	Air	Small Business	Office Supply	SAF/NCR	43 United States Central	Minnesota/Prior Lake	55,372.00	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****				
26,267.0	High	0.04	2.98	1.58	16.0	Sarah Ramsey	Regular	Air	Small Business	Office Supply	SAF/NCR	43 United States Central	Minnesota/Prior Lake	803,471.00	16.00	1,164.45	86,838.00	4.00																					
23,990.0	High	0.05	26.48	6.93	18.0	Laurie Hanna	Regular	Air	Small Business	Office Supply	SAF/NCR	43 United States Central	Minnesota/Prior Lake	59,891.00	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****				
24,063.0	Not Specified	0.07	12.99	9.44	19.0	Jim Rodgers	Regular	Air	Small Business	Technology	SAF/NCR	43 United States Central	Minnesota/Prior Lake	114,641.00	18.00	231.79	90,032.00	13.00																					
5,890.0	High	0.05	26.48	6.93	21.0	Tony Winkles Winters	Regular	Air	Small Business	Technology	SAF/NCR	37 United States East	New York/New York	10,012.00	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****				
8,065.0	Not Specified	0.07	12.99	9.44	21.0	Tony Winkles Winters	Regular	Air	Small Business	Technology	SAF/NCR	37 United States East	New York/New York	10,012.00	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****				
23,605.0	Not Specified	0.07	12.99	9.44	21.0	Tony Winkles Winters	Regular	Air	Small Business	Technology	SAF/NCR	37 United States East	New York/New York	10,012.00	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****				
20,847.0	High	0.01	2.84	0.93	3.0	Bonnie Potter	Express	A	Corporate	Office Supply	SAF/NCR	54 United States West	Washington/Anacortes	13,075.00	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****				
20,228.0	Not Specified	0.02	500.98	26.0	5.0	Ronnie Proctor	Delivery	Truck	Corporate	Office Supply	SAF/NCR	54 United States West	New York/New York	13,075.00	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****				
21,776.0	Critical	0.06	9.48	7.29	11.0	Marcus Dunlap	Regular	Air	Home Office	Office Supply	SAF/NCR	45 United States East	New York/New York	13,075.00	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****				
24,944.0	Medium	0.09	78.69	19.99	14.0	Gwendolyn F Tyson	Regular	Air	Small Business	Office Supply	SAF/NCR	43 United States Central	Minnesota/Prior Lake	13,075.00	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****				
26,267.0	High	0.04	2.98	1.58	16.0	Sarah Ramsey	Regular	Air	Small Business	Office Supply	SAF/NCR	43 United States Central	Minnesota/Prior Lake	13,075.00	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****					
23,990.0	High	0.05	26.48	6.93	18.0	Laurie Hanna	Regular	Air	Small Business	Office Supply	SAF/NCR	43 United States Central	Minnesota/Prior Lake	13,075.00	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****					
24,063.0	Not Specified	0.07	12.99	9.44	19.0	Jim Rodgers	Regular	Air	Small Business	Technology	SAF/NCR	43 United States Central	Minnesota/Prior Lake	13,075.00	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****					
5,890.0	High	0.05	26.48	6.93	21.0	Tony Winkles Winters	Regular	Air	Small Business	Technology	SAF/NCR	37 United States East	New York/New York	10,012.00	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****				
8,065.0	Not Specified	0.07	12.99	9.44	21.0	Tony Winkles Winters	Regular	Air	Small Business	Technology	SAF/NCR	37 United States East	New York/New York	10,012.00	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****				
20,847.0	High	0.01	2.84	0.93	3.0	Bonnie Potter	Express	A	Corporate	Office Supply	SAF/NCR	54 United States West	Washington/Anacortes	13,075.00	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****				
20,228.0	Not Specified	0.02	500.98	26.0	5.0	Ronnie Proctor	Delivery	Truck	Corporate	Office Supply	SAF/NCR	54 United States West	New York/New York	13,075.00	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****				
21,776.0	Critical	0.06	9.48	7.29	11.0	Marcus Dunlap	Regular	Air	Home Office	Office Supply	SAF/NCR	45 United States East	New York/New York	13,075.00	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****				
24,944.0	Medium	0.09	78.69	19.99	14.0	Gwendolyn F Tyson	Regular	Air	Small Business	Office Supply	SAF/NCR	43 United States Central	Minnesota/Prior Lake	13,075.00	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****				
26,267.0	High	0.04	2.98	1.58	16.0	Sarah Ramsey	Regular	Air	Small Business	Office Supply	SAF/NCR	43 United States Central	Minnesota/Prior Lake	13,075.00	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****					
23,990.0	High	0.05	26.48	6.93	18.0	Laurie Hanna	Regular	Air	Small Business	Office Supply	SAF/NCR	43 United States Central	Minnesota/Prior Lake	13,075.00	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****					
24,063.0	Not Specified	0.07	12.99	9.44	19.0	Jim Rodgers	Regular	Air	Small Business	Technology	SAF/NCR	43 United States Central	Minnesota/Prior Lake	13,075.00	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****					
5,890.0	High	0.05	26.48	6.93	21.0	Tony Winkles Winters	Regular	Air	Small Business	Technology	SAF/NCR	37 United States East	New York/New York	10,012.00	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****				
8,065.0	Not Specified	0.07	12.99	9.44	21.0	Tony Winkles Winters	Regular	Air	Small Business	Technology	SAF/NCR	37 United States East	New York/New York	10,012.00	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****					
20,847.0	High	0.01	2.84	0.93	3.0	Bonnie Potter	Express	A	Corporate	Office Supply	SAF/NCR	54 United States West	Washington/Anacortes	13,075.00	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****					
20,228.0	Not Specified	0.02	500.98	26.0	5.0	Ronnie Proctor	Delivery	Truck	Corporate	Office Supply	SAF/NCR	54 United States West	New York/New York	13,075.00	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****					
21,776.0	Critical	0.06	9.48	7.29	11.0	Marcus Dunlap	Regular	Air	Home Office	Office Supply	SAF/NCR	45 United States East	New York/New York	13,075.00	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****					
24,944.0	Medium	0.09	78.69	19.99	14.0	Gwendolyn F Tyson	Regular	Air	Small Business	Office Supply	SAF/NCR	43 United States Central	Minnesota/Prior Lake	13,075.00	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****					
26,267.0	High	0.04	2.98	1.58	16.0	Sarah Ramsey	Regular	Air	Small Business	Office Supply	SAF/NCR	43 United States Central	Minnesota/Prior Lake	13,075.00	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****					
23,990.0	High	0.05	26.48	6.93	18.0	Laurie Hanna	Regular	Air	Small Business	Office Supply	SAF/NCR	43 United States Central	Minnesota/Prior Lake	13,075.00	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****					
24,063.0	Not Specified	0.07	12.99	9.44	19.0	Jim Rodgers	Regular	Air	Small Business	Technology	SAF/NCR	43 United States Central	Minnesota/Prior Lake	13,075.00	*****	*****																							

3) Add sequence to the file

- 1) Drag and drop Microsoft Excel Input, Microsoft Excel Output and Unique rows from transform field.



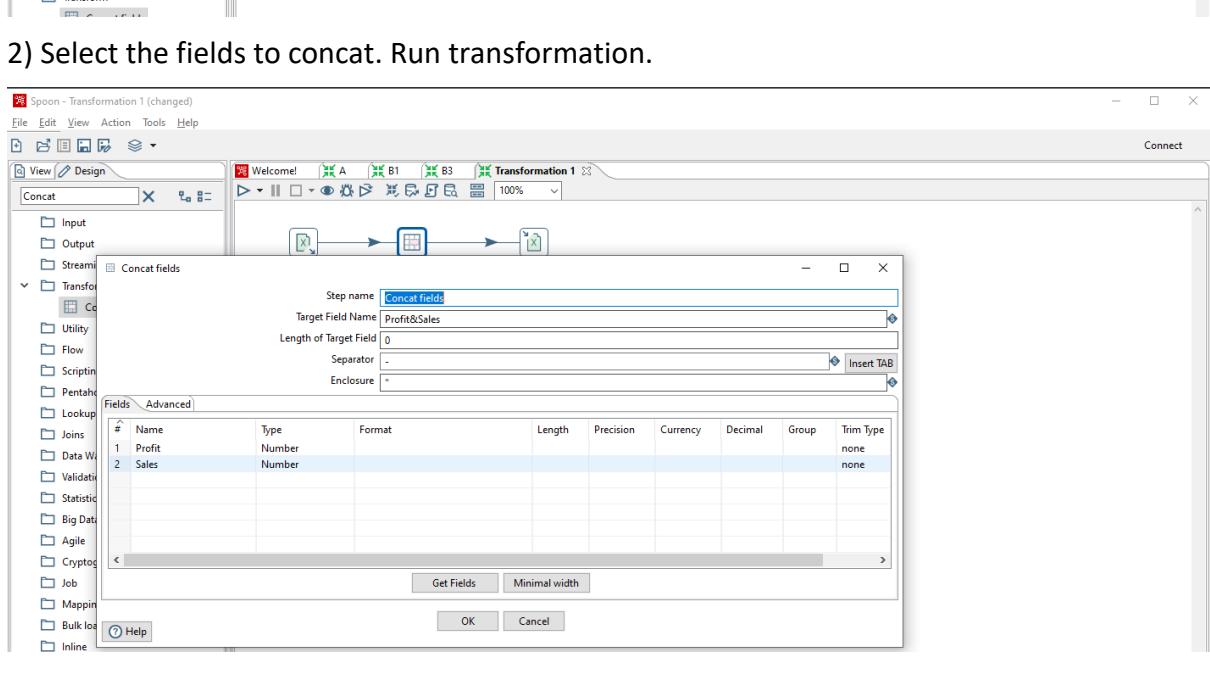
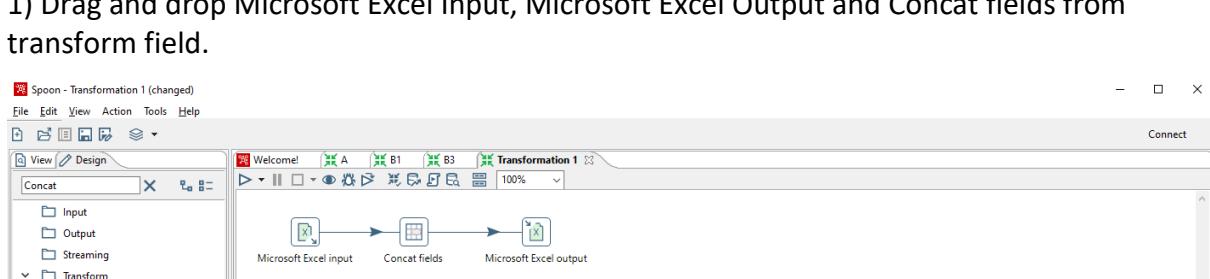
- 2) In Add Sequence, select the start value and increment value. Run the transformation.

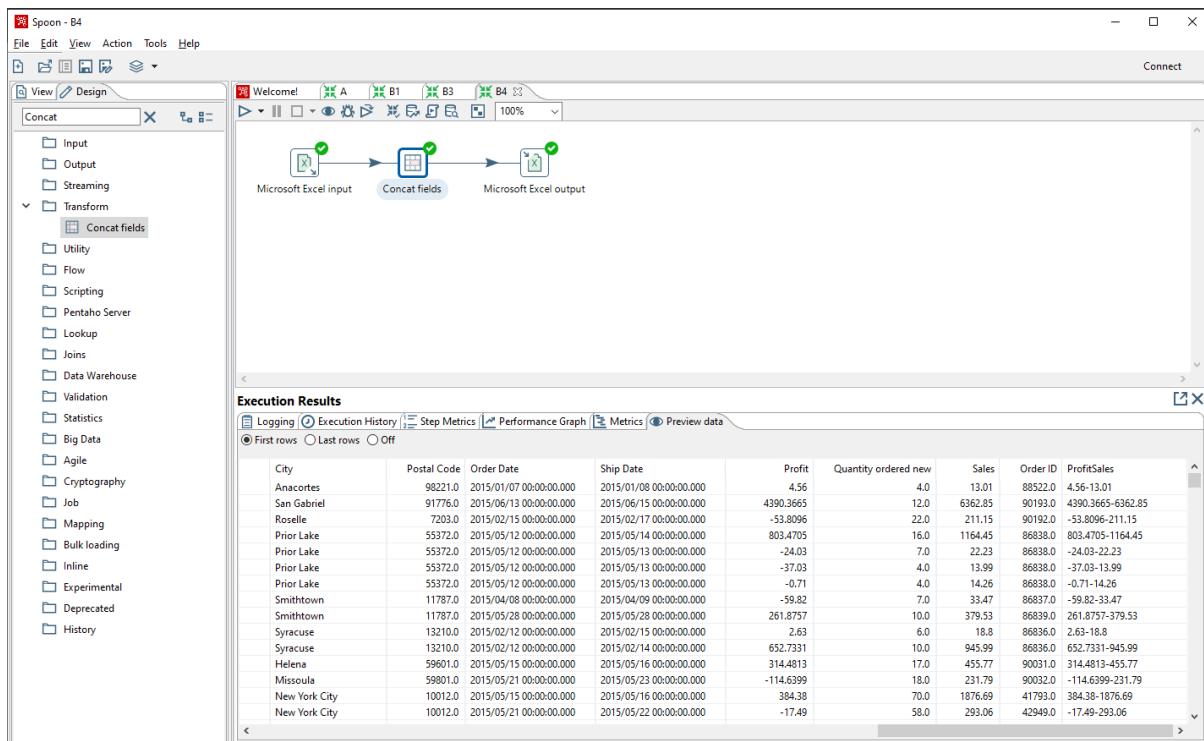
This screenshot shows the 'Add sequence' dialog box open in the center of the Spoon interface. The 'Step name' is set to 'Add sequence'. The 'Name of value' is 'Seq_val'. Under 'Use a database to generate the sequence', there is a checkbox 'Use DB to get sequence?' which is unchecked. Below it, 'Connection' and 'Schema name' dropdowns are shown, along with 'Sequences...' and 'Schemas...' buttons. Under 'Use a transformation counter to generate the sequence', there is a checkbox 'Use counter to calculate sequence?' which is checked. The 'Counter name (optional)' field is empty. The 'Start at value' is '1', 'Increment by' is '1', and 'Maximum value' is '999999999'. At the bottom are 'Help', 'OK', and 'Cancel' buttons.

Below the dialog, the main Spoon interface shows the transformation flow again: Microsoft Excel input -> Add sequence -> Microsoft Excel output. The 'Execution Results' window is open at the bottom, showing a table of data with columns: te or Province, City, Postal Code, Order Date, Ship Date, Profit, Quantity ordered new, Sales, Order ID, and Seq_val. The data includes rows for various cities and states, with the 'Seq_val' column showing values starting from 1 and increasing sequentially.

Z1	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
	Row ID	Order Ptn	Discount	Unit Price	Shipping Customer	Customer Ship Mode	Customer Product S	Product C	Product N	Product Bi	Country	Region	State or P/ City	Postal Co	Order Date	Ship Date	Profit	Quantity o	Sales	Order ID	Seq_val					
2	20.847.00	High	.01	2.84	93	3.00	Bonnie Po' Express	A Corporate	Office Sup	Pens & Ar Wrap Bag	SANFORC	54 United States West	WashingtonAnacates	98,221.00	#####	4,56	4.00	13,01	85,522.00	1.00						
3	20.228.00	Not Specif	.02	500.98	26.00	5.00	RON PrtDelivery Tr Home Offic	Furniture	Chairs & Jumbo DruGlobal	Two-T	60 United States West	California San Gabri	91,776.00	#####	4,390.37	12.00	6,362.85	90,193.00	2.00							
4	21.776.00	Critical	.06	9.48	7.29	11.00	Marcus Du Regular	A Home Offic	Furniture	Office Furr Small Pac DAX Two-T	45 United States East	New Jersey Roselle	7,203.00	#####	-53.81	22.00	211.15	90,192.00	3.00							
5	24.844.00	Medium	.09	78.69	19.99	14.00	Gwendolyn Regular	Ai Small Bus	Furniture	Office Furr Small Box Howard Mi	43 United States Central	Minnesota Prior Lake	55,372.00	#####	803.47	16.00	1,164.45	86,638.00	4.00							
6	24.846.00	Medium	.08	3.28	2.31	14.00	Gwendolyn Regular	Ai Small Bus	Office Sup	Pens & Ar Wrap Bag	Newell 321	55 United States Central	Minnesota Prior Lake	55,372.00	#####	-24.03	7.00	22.23	86,638.00	5.00						
7	24.846.00	Medium	.09	3.28	4.20	14.00	Gwendolyn Regular	Ai Small Bus	Office Sup	Pens & Ar Wrap Bag	Howard 351	45 United States Central	Minnesota Prior Lake	55,372.00	#####	-37.00	4.00	13.92	86,638.00	6.00						
8	24.846.00	Medium	.09	3.28	15.63	14.00	Gwendolyn Regular	Ai Small Bus	Office Sup	Small Box	Howard 351	45 United States Central	Minnesota Prior Lake	55,372.00	#####	-7.71	1.00	4.00	86,638.00	7.00						
9	18.191.00	Critical	.00	4.42	4.99	15.00	Timothy R Regular	Ai Small Bus	Office Sup	Envelope	Small Box Grip Seal F	38 United States East	New York Smithtown	11,787.00	#####	59.82	7.00	33.47	58,837.00	8.00						
10	20.925.00	Medium	.01	35.94	6.66	15.00	Timothy R Regular	Ai Small Bus	Office Sup	Envelope	Small Box Tyvek ® Tr	40 United States East	New York Smithtown	11,787.00	#####	261.88	10.00	379.53	58,837.00	9.00						
11	26.267.00	High	.04	2.98	1.58	16.00	Sarah Ran Regular	Ai Small Bus	Furniture	Office Furr Small Box DAX Natur	49 United States East	New York Syracuse	13,210.00	#####	2.63	6.00	18.80	86,836.00	10.00							
12	26.268.00	High	.05	115.99	2.50	16.00	Sarah Ran Regular	Ai Small Bus	Technology	Telephone:Small Box StarTAC 7	55 United States East	New York Syracuse	13,210.00	#####	652.73	10.00	945.99	86,836.00	11.00							
13	23.890.00	High	.05	26.48	6.93	18.00	Laura Hani Regular	Ai Small Bus	Furniture	Office Furr Small Box DAX Natur	49 United States West	Montana Helena	59,601.00	#####	314.48	17.00	455.77	90,031.00	12.00							
14	24.063.00	Not Specif	.07	12.99	9.44	19.00	Jim Rodge Regular	Ai Small Bus	Technology	Office Mac Medium Bi Hewlett Ps	39 United States East	Montana Missoula	98,001.00	#####	-114.64	18.00	231.79	90,032.00	13.00							
15	5.062.00	Medium	.05	26.48	15.35	21.00	Matthew B Regular	Ai Small Bus	Technology	Office Mac Medium Bi Hewlett Ps	40 United States East	Montana Missoula	98,001.00	#####	35.80	2.00	10.00	98,001.00	14.00							
16	5.062.00	Not Specif	.08	5.00	3.39	21.00	Tony Wilki Regular	Ai Small Bus	Office Sup	Rubber Ba Wrap Bag	Advantage F	37 United States East	New York New York	10,012.00	#####	-17.49	6.00	283.06	42,949.00	15.00						
17	6.063.00	Not Specif	.07	12.99	9.44	21.00	Tony Wilki Regular	Ai Small Bus	Technology	Office Mac Medium Bi Hewlett Ps	39 United States East	New York New York	10,012.00	#####	-114.64	7.00	914.22	42,949.00	16.00							
18	20.631.00	High	.06	55.48	14.30	24.00	Edu Thon Regular	Ai Corporate	Office Sup	Paper	Small Box Xerox 194	37 United States West	California Laguna Nig	92,677.00	#####	-28.30	1.00	67.43	87,651.00	17.00						
19	20.632.00	High	.02	1.68	1.57	24.00	Edu Thon Regular	Ai Corporate	Office Sup	Pens & Ar Wrap Bag	Newell 323	59 United States West	California Laguna Nig	92,677.00	#####	-5.31	1.00	2.25	87,651.00	18.00						
20	23.967.00	Not Specif	.04	4.14	6.60	27.00	Guy Gallar Regular	Ai Corporate	Furniture	Office Furr Small Box Eldon Imax	49 United States West	California Lakewood	90,712.00	#####	8.89	12.00	54.78	87,652.00	19.00							
21	23.509.00	High	.08	34.99	7.73	32.00	Matthew B Regular	Ai Corporate	Furniture	Office Sup Pens & Ar Small Box Hunt Bost	59 United States West	Oregon Grants Pt	97,526.00	#####	144.69	13.00	424.68	89,199.00	20.00							
22	23.509.00	High	.01	17.98	5.79	32.00	Matthew B Regular	Ai Corporate	Furniture	Office Sup Pens & Ar Small Box Hunt Bost	40 United States East	Oregon Grants Pt	97,526.00	#####	-35.80	2.00	10.00	97,526.00	21.00							
23	21.616.00	Medium	.05	125.69	6.69	32.00	Matthew B Regular	Ai Corporate	Technology	Telephone:Small Box StarTAC 3	59 United States West	Oregon Grants Pt	97,526.00	#####	21.00	6.00	783.55	89,202.00	22.00							
24	19.356.00	Low	.06	205.99	8.99	32.00	Matthew B Regular	Ai Corporate	Technology	Telephone:Small Box TransportP	56 United States West	Oregon Grants Pt	97,526.00	#####	3,569.10	22.00	3,838.14	89,203.00	23.00							
25	23.654.00	Not Specif	.03	4.24	5.41	33.00	Ricky Hen Regular	Ai Corporate	Office Sup Binders	Small Box StoreX Dur	35 United States West	Oregon Gresham	97,030.00	#####	-84.44	13.00	58.68	89,201.00	24.00							
26	23.655.00	Not Specif	.04	2.94	.70	33.00	Ricky Hen Regular	Ai Corporate	Office Sup	Pens & Ar Wrap Bag	Newell 338	58 United States West	Oregon Gresham	97,030.00	#####	24.31	18.00	53.10	89,201.00	25.00						
27	25.933.00	High	.00	99.99	19.99	43.00	Theodore K Regular	Ai Consumer	Technology	Office Mac Small Box AT&T 223L	52 United States West	WashingtonRedmond	98,052.00	#####	25.91	6.00	647.07	91,454.00	26.00							
28	18.551.00	Not Specif	.07	3,502.14	8.73	52.00	Lorraine Ki Regular	Ai Corporate	Technology	Telephone:Small Box 6160	57 United States West	WashingtonPuyallup	98,373.00	#####	162.67	6.00	627.04	88,426.00	27.00							
29	22.552.00	Medium	.05	1.68	5.89	53.00	Sidney Ru Regular	Ai Corporate	Office Sup	Binders	Small Box Boxbinders Pt	57 United States West	WashingtonRedmond	65,950.00	#####	3,267.53	19,825.25	28.00								
30	5.552.00	Not Specif	.07	3,502.14	8.73	53.00	Sidney Ru Regular	Ai Corporate	Office Sup	Binders	Small Box Boxbinders Pt	57 United States West	WashingtonRedmond	65,950.00	#####	-67.49	17.00	1,049.19	88,426.00	29.00						
31	20.697.00	Medium	.06	3.80	1.49	56.00	Randall M Regular	Ai Consumer	Office Sup	Binders	Small Box Durable Pr	38 United States East	New York Tonawand	14,150.00	#####	19.63	20.00	73.55	88,075.00	30.00						
32	20.698.00	Medium	.06	1.76	.70	56.00	Randall M Regular	Ai Consumer	Office Sup	Pens & Ar Wrap Bag	Newell 31C	56 United States East	New York Tonawand	14,150.00	#####	-1.65	17.00	29.57	88,075.00	31.00						
33	22.890.00	High	.02	5.98	5.15	62.00	Pam Gilbe Regular	Ai Corporate	Office Sup	Paper	Small Box Xerox 193	36 United States Central	Texas Round Roc	78,664.00	#####	2.14	3.00	22.85	87,407.00	32.00						
34	25.354.00	High	.09	29.14	4.88	62.00	Pam Gilbe Regular	Ai Corporate	Office Sup	Paper	Wrap Bag Snap-A-W	38 United States Central	Texas Round Roc	78,664.00	#####	349.41	17.00	506.39	87,408.00	33.00						
35	21.019.00	Not Specif	.02	175.99	4.99	64.00	Lynn Morri Regular	Ai Small Bus	Technology	Telephone:Small Box 5165	59 United States South	Virginia Salem	24,153.00	#####	-37.53	1.00	4.00	87,406.00	34.00							
36	21.019.00	Not Specif	.02	152.00	1.07	64.00	Lynn Morri Regular	Ai Small Bus	Technology	Telephone:Small Box 5165	59 United States South	Virginia Salem	24,153.00	#####	101.49	4.00	588.80	87,406.00	35.00							
37	19.217.00	Low	.00	291.73	48.80	68.00	Scott Burn Delivery Tr	Corporate	Furniture	Chairs & C.Jumbo DruHen 470 :	55 United States East	New York New York	10,177.00	#####	345.00	4.00	1,230.06	37,537.00	36.00							
38	19.217.00	Low	.07	100.98	45.00	68.00	Scott Burn Delivery Tr	Corporate	Furniture	Chairs & C.Jumbo DruHen Valut	69 United States East	New York New York	10,177.00	#####	-36.93	4.00	4,083.19	37,537.00	37.00							
39	5.273.00	Low	.07	155.06	7.07	68.00	Scott Burn Delivery Tr	Corporate	Furniture	Chairs & C.Jumbo DruGlobal Hig	59 United States East	New York New York	10,177.00	#####	-1,679.76	43.00	4,083.19	37,537.00	38.00							
40	5.274.00	Low	.07	122.99	70.20	68.00	Scott Burn Delivery Tr	Corporate	Furniture	Chairs & C.Jumbo DruGlobal Hig	74 United States East	New York New York	10,177.00	#####	575.40	32.00	4,902.30	37,537.00	39.00							
41	7.786.00	High	.09	122.99	70.20	68.00	Scott Burn Delivery Tr	Corporate	Furniture	Tables Jumbo DruHon 94000	74 United States East	Vermont Burlington	5,401.00	#####	-2,426.55	49.00	5,713.85	55,713.00	40.00							
42	25.786.00	High	.09	122.99	70.20	70.00	Annette BrDelivery Tr	Corporate	Furniture	Tables Jumbo DruHon 94000	76 United States East	Ohio Canton	44,708.00	#####	-2,426.55	12.00	1,400.53	87,947.00	41.00							
43	18.281.00	High	.04	296.18	54.12	83.00	Edgar StorDelivery Tr	Corporate	Furniture	Tables Jumbo DruHon 94000	76 United States East	Ohio Canton	44,708.00	#####	-2,426.55	6.00	1,821.89	87,365.00	42.00							

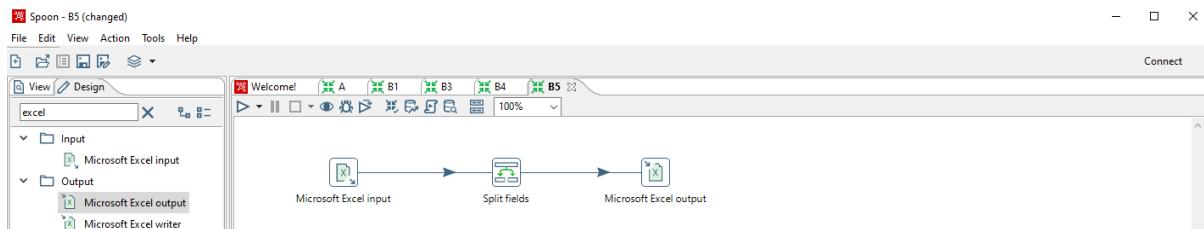
4) Concatenate



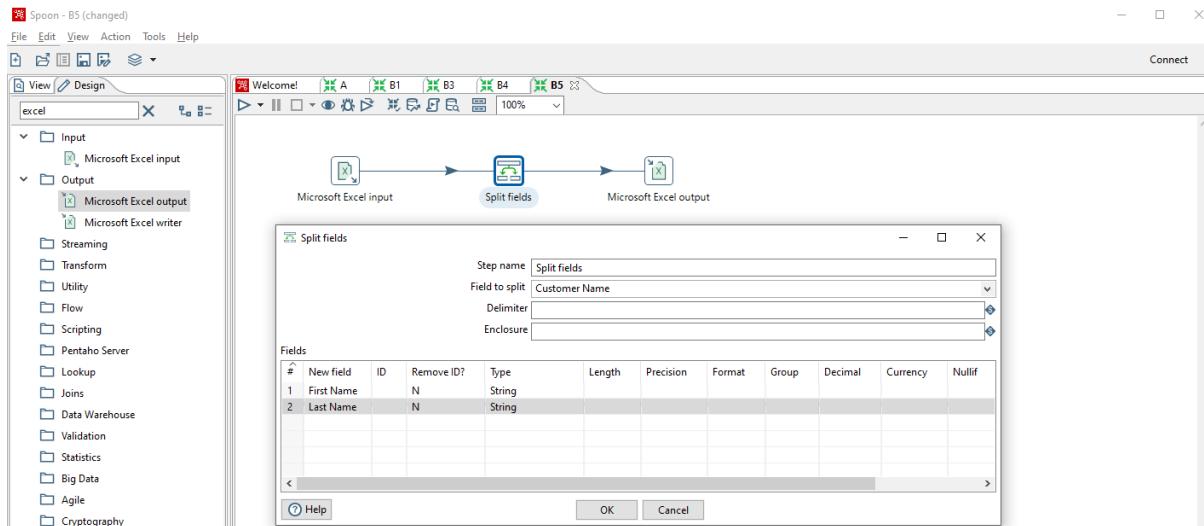


5) Split

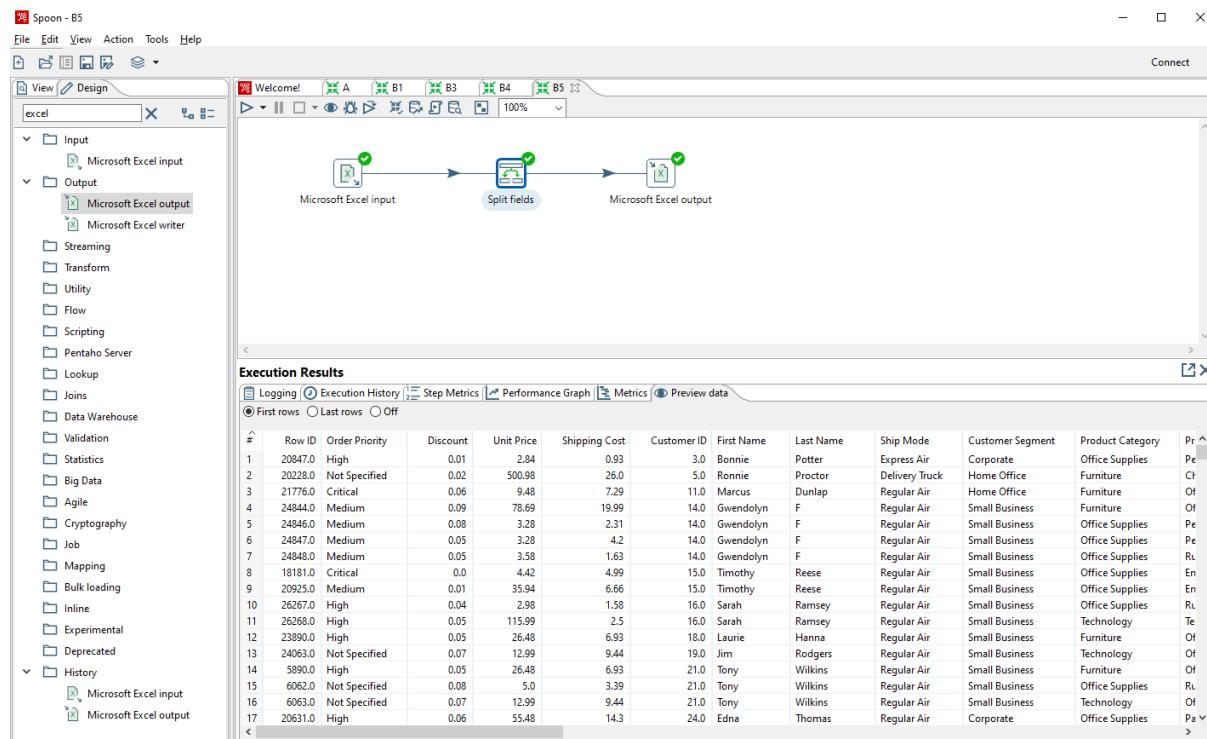
1) Drag and drop Microsoft Excel Input, Microsoft Excel Output and Split fields from transform field.



2) In Split fields, choose the row you want to split and specify the delimiter. In this, the Customer name is split into first and last name using the “space” delimiter.



3) Run the transformation.



Practical No 3
ORACLE ANALYTICAL QUERIES

Q) Implementation of Analytical queries like Roll_Up, Cube, First, Last, Lead, Lag, Rank, Dense Rank, etc.

Solution:

```
SELECT * FROM EMP;
```

SQL Worksheet

```
1 SELECT * FROM EMP;
```

EMPNO	DEPTNO	DOB	SALARY	COM	JOB
7931	2	21-JAN-98	26000	1500	SRMANAGER
7932	1	26-MAY-97	30000	1000	MANAGER
7933	3	03-AUG-98	36000	1100	CEO
7934	4	10-FEB-96	10000	1200	DEVELOPER
7935	1	14-JUL-99	16000	2000	MANAGER
7936	5	27-FEB-94	12000	500	CLERK

[Download CSV](#)
6 rows selected.

ROLL_UP:

```
SELECT DEPTNO,JOB,COUNT(*),SUM(SALARY)
FROM EMP
GROUP BY ROLLUP(DEPTNO,JOB);
```

SQL Worksheet

```
1 SELECT DEPTNO,JOB,COUNT(*),SUM(SALARY)
2 FROM EMP
3 GROUP BY ROLLUP(DEPTNO,JOB);
```

DEPTNO	JOB	COUNT(*)	SUM(SALARY)
1	MANAGER	2	46000
1	-	2	46000
2	SRMANAGER	1	26000
2	-	1	26000
3	CEO	1	36000
3	-	1	36000
4	DEVELOPER	1	10000
4	-	1	10000
5	CLERK	1	12000
5	-	1	12000
-	-	6	130000

[Download CSV](#)
11 rows selected.

CUBE:

```
SELECT DEPTNO,JOB,COUNT(*),SUM(SALARY)
FROM EMP
GROUP BY CUBE(DEPTNO,JOB);
```

SQL Worksheet

DEPTNO	JOB	COUNT(*)	SUM(SALARY)
-	-	6	130000
-	CEO	1	36000
-	CLERK	1	12000
-	MANAGER	2	46000
-	DEVELOPER	1	10000
-	SRMANAGER	1	26000
1	-	2	46000
1	MANAGER	2	46000
2	-	1	26000
2	SRMANAGER	1	26000
3	-	1	36000
3	CEO	1	36000
4	-	1	10000
4	DEVELOPER	1	10000
5	-	1	12000
5	CLERK	1	12000

[Download CSV](#)

16 rows selected.

Rank:

```
SELECT EMPNO,DEPTNO,SALARY,COM,
RANK() OVER(PARTITION BY DEPTNO ORDER BY SALARY)AS RANK FROM EMP;
```

SQL Worksheet

EMPNO	DEPTNO	SALARY	COM	RANK
101	1	2000	SA-Sales	1
102	1	3000	SA-Sales	1
103	1	3000	SA-Sales	1
104	1	3000	SA-Sales	1
105	1	3000	SA-Sales	1
106	1	3000	SA-Sales	1
107	1	3000	SA-Sales	1
108	1	3000	SA-Sales	1
109	1	3000	SA-Sales	1
110	1	3000	SA-Sales	1
111	1	3000	SA-Sales	1
112	1	3000	SA-Sales	1
113	1	3000	SA-Sales	1
114	1	3000	SA-Sales	1
115	1	3000	SA-Sales	1
116	1	3000	SA-Sales	1
117	1	3000	SA-Sales	1
118	1	3000	SA-Sales	1
119	1	3000	SA-Sales	1
120	1	3000	SA-Sales	1
121	1	3000	SA-Sales	1
122	1	3000	SA-Sales	1
123	1	3000	SA-Sales	1
124	1	3000	SA-Sales	1
125	1	3000	SA-Sales	1
126	1	3000	SA-Sales	1
127	1	3000	SA-Sales	1
128	1	3000	SA-Sales	1
129	1	3000	SA-Sales	1
130	1	3000	SA-Sales	1
131	1	3000	SA-Sales	1
132	1	3000	SA-Sales	1
133	1	3000	SA-Sales	1
134	1	3000	SA-Sales	1
135	1	3000	SA-Sales	1
136	1	3000	SA-Sales	1
137	1	3000	SA-Sales	1
138	1	3000	SA-Sales	1
139	1	3000	SA-Sales	1
140	1	3000	SA-Sales	1
141	1	3000	SA-Sales	1
142	1	3000	SA-Sales	1
143	1	3000	SA-Sales	1
144	1	3000	SA-Sales	1
145	1	3000	SA-Sales	1
146	1	3000	SA-Sales	1
147	1	3000	SA-Sales	1
148	1	3000	SA-Sales	1
149	1	3000	SA-Sales	1
150	1	3000	SA-Sales	1
151	1	3000	SA-Sales	1
152	1	3000	SA-Sales	1
153	1	3000	SA-Sales	1
154	1	3000	SA-Sales	1
155	1	3000	SA-Sales	1
156	1	3000	SA-Sales	1
157	1	3000	SA-Sales	1
158	1	3000	SA-Sales	1
159	1	3000	SA-Sales	1
160	1	3000	SA-Sales	1
161	1	3000	SA-Sales	1
162	1	3000	SA-Sales	1
163	1	3000	SA-Sales	1
164	1	3000	SA-Sales	1
165	1	3000	SA-Sales	1
166	1	3000	SA-Sales	1
167	1	3000	SA-Sales	1
168	1	3000	SA-Sales	1
169	1	3000	SA-Sales	1
170	1	3000	SA-Sales	1
171	1	3000	SA-Sales	1
172	1	3000	SA-Sales	1
173	1	3000	SA-Sales	1
174	1	3000	SA-Sales	1
175	1	3000	SA-Sales	1
176	1	3000	SA-Sales	1
177	1	3000	SA-Sales	1
178	1	3000	SA-Sales	1
179	1	3000	SA-Sales	1
180	1	3000	SA-Sales	1
181	1	3000	SA-Sales	1
182	1	3000	SA-Sales	1
183	1	3000	SA-Sales	1
184	1	3000	SA-Sales	1
185	1	3000	SA-Sales	1
186	1	3000	SA-Sales	1
187	1	3000	SA-Sales	1
188	1	3000	SA-Sales	1
189	1	3000	SA-Sales	1
190	1	3000	SA-Sales	1
191	1	3000	SA-Sales	1
192	1	3000	SA-Sales	1
193	1	3000	SA-Sales	1
194	1	3000	SA-Sales	1
195	1	3000	SA-Sales	1
196	1	3000	SA-Sales	1
197	1	3000	SA-Sales	1
198	1	3000	SA-Sales	1
199	1	3000	SA-Sales	1
200	1	3000	SA-Sales	1
201	1	3000	SA-Sales	1
202	1	3000	SA-Sales	1
203	1	3000	SA-Sales	1
204	1	3000	SA-Sales	1
205	1	3000	SA-Sales	1
206	1	3000	SA-Sales	1
207	1	3000	SA-Sales	1
208	1	3000	SA-Sales	1
209	1	3000	SA-Sales	1
210	1	3000	SA-Sales	1
211	1	3000	SA-Sales	1
212	1	3000	SA-Sales	1
213	1	3000	SA-Sales	1
214	1	3000	SA-Sales	1
215	1	3000	SA-Sales	1
216	1	3000	SA-Sales	1
217	1	3000	SA-Sales	1
218	1	3000	SA-Sales	1
219	1	3000	SA-Sales	1
220	1	3000	SA-Sales	1
221	1	3000	SA-Sales	1
222	1	3000	SA-Sales	1
223	1	3000	SA-Sales	1
224	1	3000	SA-Sales	1
225	1	3000	SA-Sales	1
226	1	3000	SA-Sales	1
227	1	3000	SA-Sales	1
228	1	3000	SA-Sales	1
229	1	3000	SA-Sales	1
230	1	3000	SA-Sales	1
231	1	3000	SA-Sales	1
232	1	3000	SA-Sales	1
233	1	3000	SA-Sales	1
234	1	3000	SA-Sales	1
235	1	3000	SA-Sales	1
236	1	3000	SA-Sales	1
237	1	3000	SA-Sales	1
238	1	3000	SA-Sales	1
239	1	3000	SA-Sales	1
240	1	3000	SA-Sales	1
241	1	3000	SA-Sales	1
242	1	3000	SA-Sales	1
243	1	3000	SA-Sales	1
244	1	3000	SA-Sales	1
245	1	3000	SA-Sales	1
246	1	3000	SA-Sales	1
247	1	3000	SA-Sales	1
248	1	3000	SA-Sales	1
249	1	3000	SA-Sales	1
250	1	3000	SA-Sales	1
251	1	3000	SA-Sales	1
252	1	3000	SA-Sales	1
253	1	3000	SA-Sales	1
254	1	3000	SA-Sales	1
255	1	3000	SA-Sales	1
256	1	3000	SA-Sales	1
257	1	3000	SA-Sales	1
258	1	3000	SA-Sales	1
259	1	3000	SA-Sales	1
260	1	3000	SA-Sales	1
261	1	3000	SA-Sales	1
262	1	3000	SA-Sales	1
263	1	3000	SA-Sales	1
264	1	3000	SA-Sales	1
265	1	3000	SA-Sales	1
266	1	3000	SA-Sales	1
267	1	3000	SA-Sales	1
268	1	3000	SA-Sales	1
269	1	3000	SA-Sales	1
270	1	3000	SA-Sales	1
271	1	3000	SA-Sales	1
272	1	3000	SA-Sales	1
273	1	3000	SA-Sales	1
274	1	3000	SA-Sales	1
275	1	3000	SA-Sales	1
276	1	3000	SA-Sales	1
277	1	3000	SA-Sales	1
278	1	3000	SA-Sales	1
279	1	3000	SA-Sales	1
280	1	3000	SA-Sales	1
281	1	3000	SA-Sales	1
282	1	3000	SA-Sales	1
283	1	3000	SA-Sales	1
284	1	3000	SA-Sales	1
285	1	3000	SA-Sales	1
286	1	3000	SA-Sales	1
287	1	3000	SA-Sales	1
288	1	3000	SA-Sales	1
289	1	3000	SA-Sales	1
290	1	3000	SA-Sales	1
291	1	3000	SA-Sales	1
292	1	3000	SA-Sales	1
293	1	3000	SA-Sales	1
294	1	3000	SA-Sales	1
295	1	3000	SA-Sales	1
296	1	3000	SA-Sales	1
297	1	3000	SA-Sales	1
298	1	3000	SA-Sales	1
299	1	3000	SA-Sales	1
300	1	3000	SA-Sales	1
301	1	3000	SA-Sales	1
302	1	3000	SA-Sales	1
303	1	3000	SA-Sales	1
304	1	3000	SA-Sales	1
305	1	3000	SA-Sales	1
306	1	3000	SA-Sales	1
307	1	3000	SA-Sales	1
308	1	3000	SA-Sales	1
309	1	3000	SA-Sales	1
310	1	3000	SA-Sales	1
311	1	3000	SA-Sales	1
312	1	3000	SA-Sales	1
313	1	3000	SA-Sales	1
314	1	3000	SA-Sales	1
315	1	3000	SA-Sales	1
316	1	3000	SA-Sales	1
317	1	3000	SA-Sales	1
318	1	3000	SA-Sales	1
319	1	3000	SA-Sales	1
320	1	3000	SA-Sales	1
321	1	3000	SA-Sales	1
322	1	3000	SA-Sales	1
323	1	3000	SA-Sales	1
324	1	3000	SA-Sales	1
325	1	3000	SA-Sales	1
326	1	3000	SA-Sales	1
327	1	3000	SA-Sales	1
328	1	3000	SA-Sales	1
329	1	3000	SA-Sales	1
330	1	3000	SA-Sales	1
331	1	3000	SA-Sales	1
332	1	3000	SA-Sales	1
333	1	3000	SA-Sales	1
334	1	3000	SA-Sales	1
335	1	3000	SA-Sales	1
336	1	3000	SA-Sales	1
337	1	3000	SA-Sales	1
338	1	3000	SA-Sales	1
339	1	3000	SA-Sales	1
340	1	3000	SA-Sales	1
341	1	3000	SA-Sales	1
342	1	3000	SA-Sales	1
343	1	3000	SA-Sales	1
344	1	3000	SA-Sales	1
345	1	3000	SA-Sales	1
346	1	3000	SA-Sales	1

EMPNO	DEPTNO	SALARY	COM	RANK
7935	1	16000	2000	1
7932	1	30000	1000	2
7931	2	26000	1500	1
7933	3	36000	1100	1
7934	4	10000	1200	1
7936	5	12000	500	1

[Download CSV](#)

6 rows selected.

UPDATE:

UPDATE EMP

SET SALARY=33000

WHERE EMPNO=7934;

SQL Worksheet

```

1 UPDATE EMP
2 SET SALARY=33000
3 WHERE EMPNO=7934;

```

1 row(s) updated.

SELECT:

SELECT * FROM EMP;

SQL Worksheet

```

1 SELECT * FROM EMP;

```

EMPNO	DEPTNO	DOB	SALARY	COM	JOB
7931	2	21-JAN-98	26000	1500	SRMANAGER
7932	1	26-MAY-97	30000	1000	MANAGER
7933	3	03-AUG-98	36000	1100	CEO
7934	4	10-FEB-96	33000	1200	DEVELOPER
7935	1	14-JUL-99	16000	2000	MANAGER
7936	5	27-FEB-94	12000	500	CLERK

[Download CSV](#)

6 rows selected.

Dense_rank:

```
SELECT EMPNO,DEPTNO,SALARY,COM,
DENSE_RANK() OVER(PARTITION BY DEPTNO ORDER BY SALARY)AS RANK FROM EMP;
```

SQL Worksheet

	EMPNO	DEPTNO	SALARY	COM	RANK
1	SELECT EMPNO,DEPTNO,SALARY,COM,				
2	DENSE_RANK() OVER(PARTITION BY DEPTNO ORDER BY SALARY)AS RANK FROM EMP;				

EMPNO	DEPTNO	SALARY	COM	RANK
7935	1	16000	2000	1
7932	1	30000	1000	2
7931	2	26000	1500	1
7933	3	36000	1100	1
7934	4	33000	1200	1
7936	5	12000	500	1

[Download CSV](#)

6 rows selected.

LEAD:

```
SELECT EMPNO,DOB,LEAD(DOB,1)
OVER(ORDER BY DOB) AS "NEXT" FROM EMP;
```

SQL Worksheet

	EMPNO	DOB	NEXT
1	SELECT EMPNO,DOB,LEAD(DOB,1)		
2	OVER(ORDER BY DOB) AS "NEXT" FROM EMP;		

EMPNO	DOB	NEXT
7936	27-FEB-94	10-FEB-96
7934	10-FEB-96	26-MAY-97
7932	26-MAY-97	21-JAN-98
7931	21-JAN-98	03-AUG-98
7933	03-AUG-98	14-JUL-99
7935	14-JUL-99	-

[Download CSV](#)

6 rows selected.

```
SELECT EMPNO,DOB,LEAD(DOB,1)
OVER(ORDER BY DOB) AS "NEXT" FROM EMP WHERE DEPTNO=1;
```

SQL Worksheet

```
1 SELECT EMPNO,DOB,LEAD(DOB,1)
2 OVER(ORDER BY DOB) AS "NEXT" FROM EMP WHERE DEPTNO=1;
```

EMPNO	DOB	NEXT
7932	26-MAY-97	14-JUL-99
7935	14-JUL-99	-

[Download CSV](#)

2 rows selected.

LAG:

```
SELECT EMPNO,DOB,LAG(DOB,1)
OVER(ORDER BY DOB) AS "PREVIOUS" FROM EMP;
```

SQL Worksheet

```
1 SELECT EMPNO,DOB,LAG(DOB,1)
2 OVER(ORDER BY DOB) AS "PREVIOUS" FROM EMP;
```

EMPNO	DOB	PREVIOUS
7936	27-FEB-94	-
7934	10-FEB-96	27-FEB-94
7932	26-MAY-97	10-FEB-96
7931	21-JAN-98	26-MAY-97
7933	03-AUG-98	21-JAN-98
7935	14-JUL-99	03-AUG-98

[Download CSV](#)

6 rows selected.

```
SELECT EMPNO,DOB,LAG(DOB,1)
OVER(ORDER BY DOB) AS "PREVIOUS" FROM EMP WHERE DEPTNO=2;
```

SQL Worksheet

```
1 SELECT EMPNO,DOB,LAG(DOB,1)
2 OVER(ORDER BY DOB) AS "PREVIOUS" FROM EMP WHERE DEPTNO=2;
```

EMPNO	DOB	PREVIOUS
7931	21-JAN-98	-

[Download CSV](#)

FIRST:

```
SELECT DEPTNO,SALARY,MAX(SALARY)KEEP(DENSE_RANK FIRST ORDER BY SALARY DESC)
OVER(PARTITION BY DEPTNO)"MAX" FROM EMP;
```

SQL Worksheet

1	SELECT DEPTNO,SALARY,MAX(SALARY)KEEP(DENSE_RANK FIRST ORDER BY SALARY DESC)
2	OVER(PARTITION BY DEPTNO)"MAX" FROM EMP;

DEPTNO	SALARY	MAX
1	30000	30000
1	16000	30000
2	26000	26000
3	36000	36000
4	33000	33000
5	12000	12000

[Download CSV](#)

6 rows selected.

LAST:

```
SELECT DEPTNO,SALARY,MIN(SALARY)KEEP(DENSE_RANK LAST ORDER BY SALARY DESC)
OVER(PARTITION BY DEPTNO)"MIN" FROM EMP;
```

SQL Worksheet

1	SELECT DEPTNO,SALARY,MIN(SALARY)KEEP(DENSE_RANK LAST ORDER BY SALARY DESC)
2	OVER(PARTITION BY DEPTNO)"MIN" FROM EMP;

DEPTNO	SALARY	MIN
1	30000	16000
1	16000	16000
2	26000	26000
3	36000	36000
4	33000	33000
5	12000	12000

[Download CSV](#)

6 rows selected.

Practical No 4

4.1- Introduction to weka

4.2 - Classification using WEKA

4.3 - Implementation of Apriori Algo using weka

4.1) INTRODUCTION TO WEKA.

Weka is a collection of machine learning algorithms for data mining tasks. Weka contains tools for data pre-processing, classification, regression, clustering, association rules, and visualization.

The buttons can be used to start the following applications:

Explorer: An environment for exploring data with WEKA.

Experimenter: An environment for performing experiments and conducting statistical tests between learning schemes.

Knowledge Flow: This environment supports essentially the same functions as the Explorer but with a drag-and-drop interface. One advantage is that it supports incremental learning.

Simple CLI: Provides a simple command-line interface that allows direct execution of WEKA commands for operating systems that do not provide their own command line interface.



1. Pre-process: It is used to choose and modify the data.
2. Classify: It is used to apply classification algorithms.
3. Cluster: Through this option we can learn different clustering algos for data.
4. Associate: It helps us to learn association rules for data.
5. Select Attributes: It helps us to select most relevant attributes in the data.
6. Visualize: It helps us to view the interactive 2D plot of the data.

Loading Data

The first three buttons at the top of the pre-process section enable us to load data into WEKA:

1. Open file.... Brings up a dialog box allowing us to browse for the data file on the local file system.
2. Open URL.... Asks for a Uniform Resource Locator address for where the data is stored.
3. Open DB.... Reads data from a database. (Note that to make this work you might have to edit the file in weka/experiment/DatabaseUtils.props.)

4.2) IMPLEMENTATION OF CLASSIFICATION TECHNIQUES.

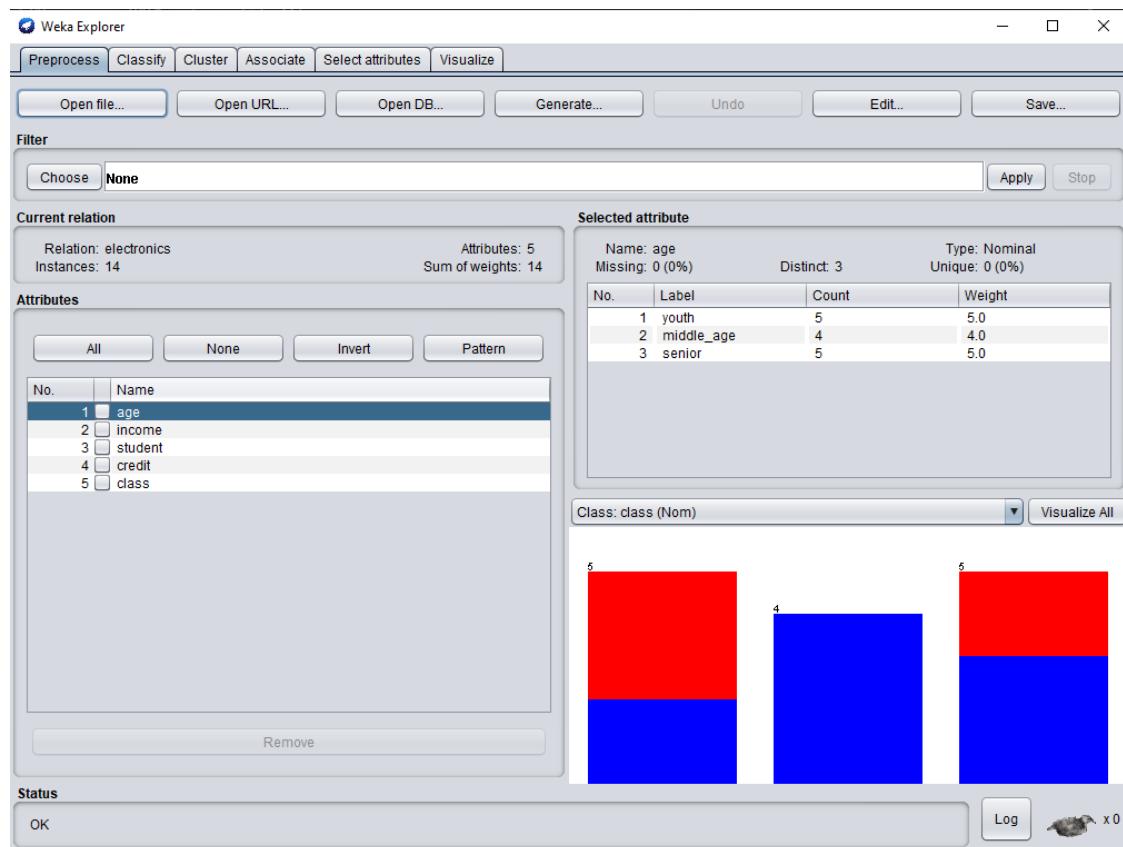
A) Create an ARFF (Attribute Relation File Format) file and construct decision tree for the following data.

RID	Age	Income	Student	Credit-rating	Class
1	Youth	High	No	Fair	No
2	Youth	High	No	Excellent	No
3	Middle-age	High	No	Fair	Yes
4	Senior	Medium	No	Fair	Yes
5	Senior	Low	Yes	Fair	Yes
6	Senior	Low	Yes	Excellent	No
7	Middle-age	Low	Yes	Excellent	Yes
8	Youth	Medium	No	Fair	No
9	Youth	Low	Yes	Fair	Yes
10	Senior	Medium	Yes	Fair	Yes
11	Youth	Medium	Yes	Excellent	Yes
12	Middle-age	Medium	No	Excellent	Yes
13	Middle-age	High	Yes	Fair	Yes
14	Senior	Medium	No	Excellent	No

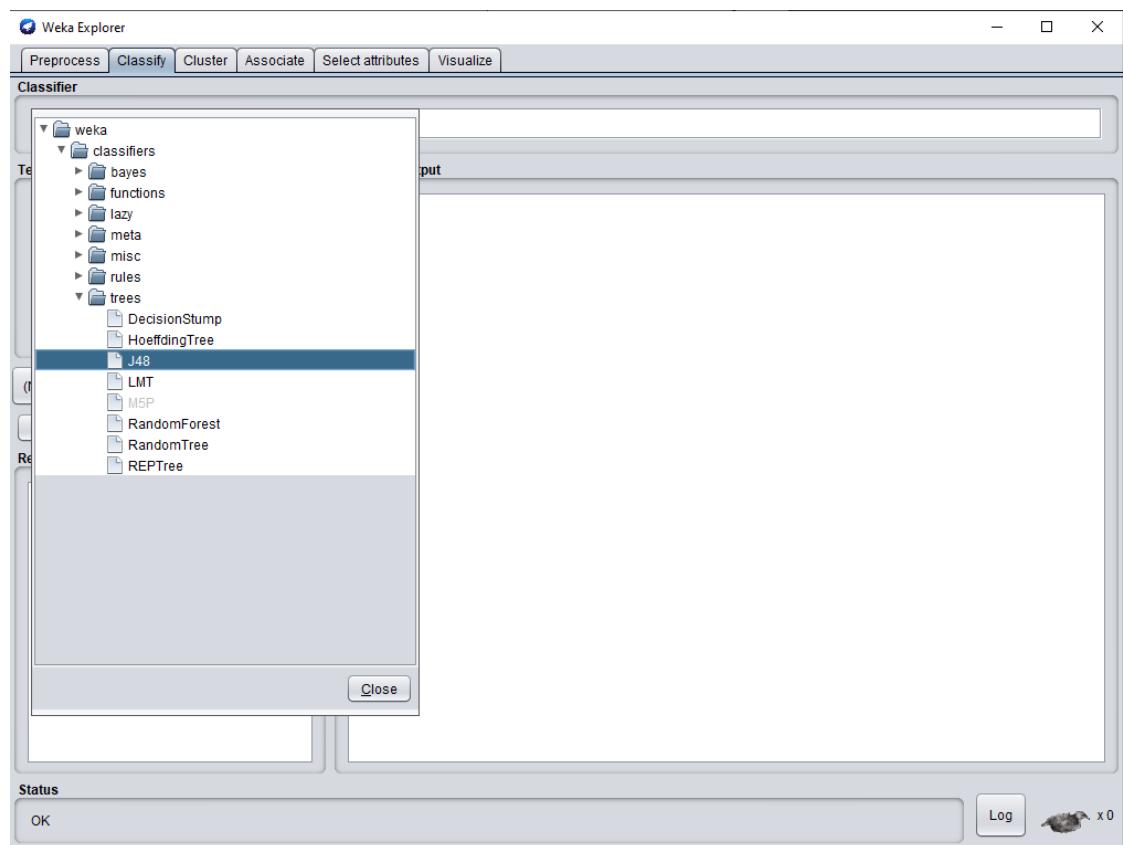
Step 1) Creating ARFF file: Open Notepad and type the following code and save it as electronics.arff.

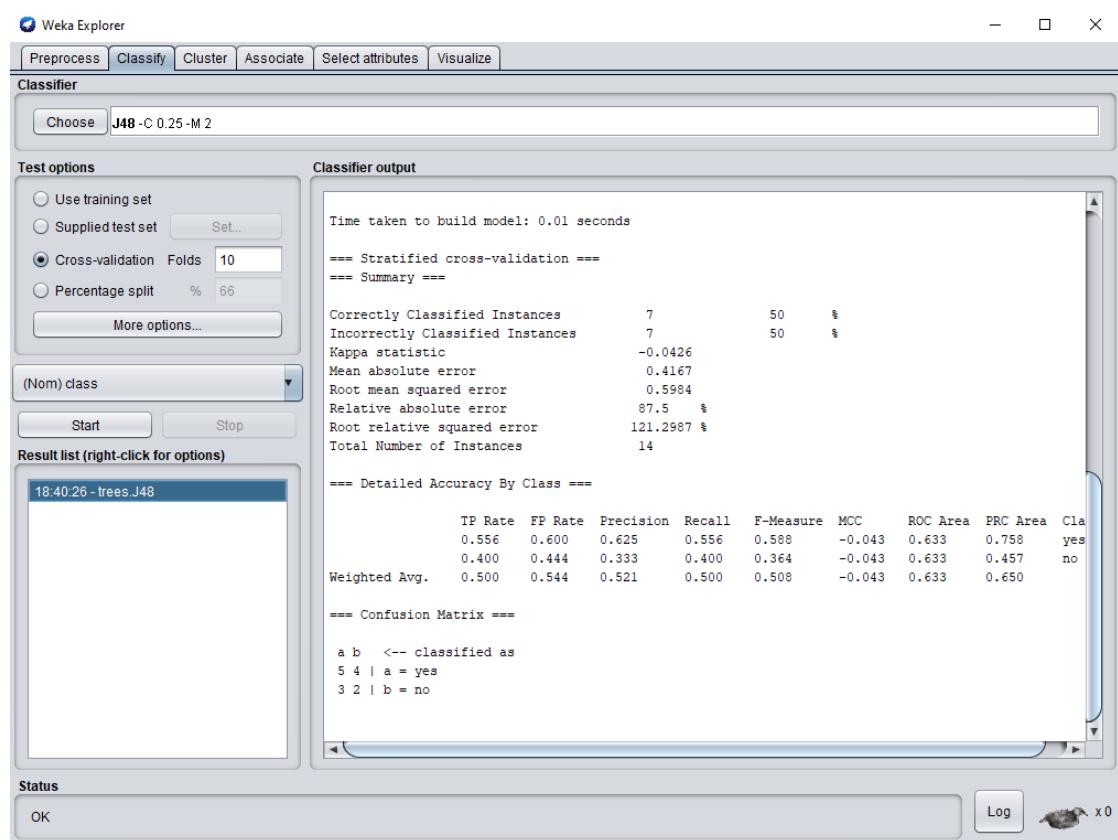
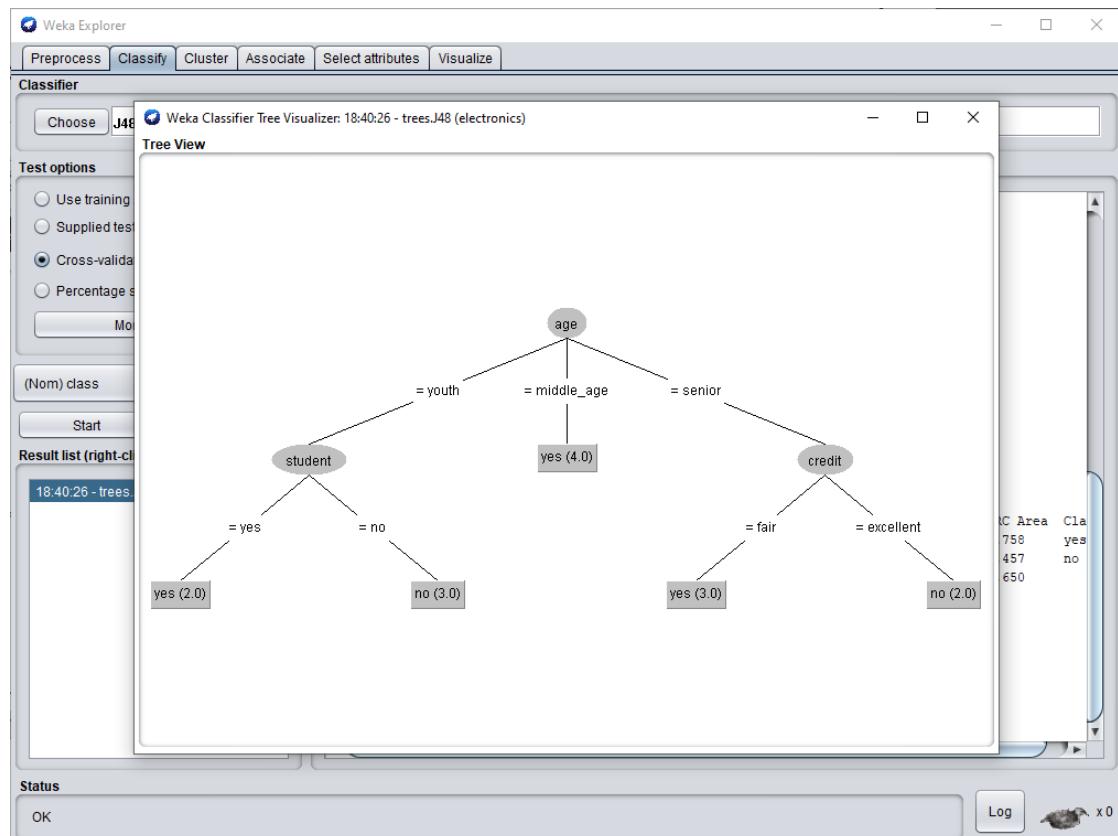
```
@relation electronics
@attribute age{youth,middle_age,senior}
@attribute income{high,medium,low}
@attribute student{yes,no}
@attribute credit{fair,excellent}
@attribute class{yes,no}
@data
youth,high,no,fair,no
youth,high,no,excellent,no
middle_age,high,no,fair,yes
senior,medium,no,fair,yes
senior,low,yes,fair,yes
senior,low,yes,excellent,no
middle_age,low,yes,excellent,yes
youth,medium,no,fair,no
youth,low,yes,fair,yes
senior,medium,yes,fair,yes
youth,medium,yes,excellent,yes
middle_age,medium,no,excellent,yes
middle_age,high,yes,fair,yes
senior,medium,no,excellent,no
```

Step 2) Open this file in WEKA by clicking on open file button.



Step 3) click on classify tab and then click on choose file and under tree select J48.



Step 4) Click on Start.**Step 5) Right Click on J48 tree and select visualize tree.**

4.3) Implementation of Apriori Algo using weka**Q1**

5.3 A database has five transactions. Let $\text{min_sup} = 60\%$ and $\text{min_conf} = 80\%$.

TID	items_bought
T100	{M, O, N, K, E, Y}
T200	{D, O, N, K, E, Y }
T300	{M, A, K, E}
T400	{M, U, C, K, Y}
T500	{C, O, O, K, I,E}

- (a) Find all frequent itemsets using Apriori and FP-growth, respectively. Compare the efficiency of the two mining processes.
- (b) List all of the *strong* association rules (with support s and confidence c) matching the following metarule, where X is a variable representing customers, and item_i denotes variables representing items (e.g., "A", "B", etc.):

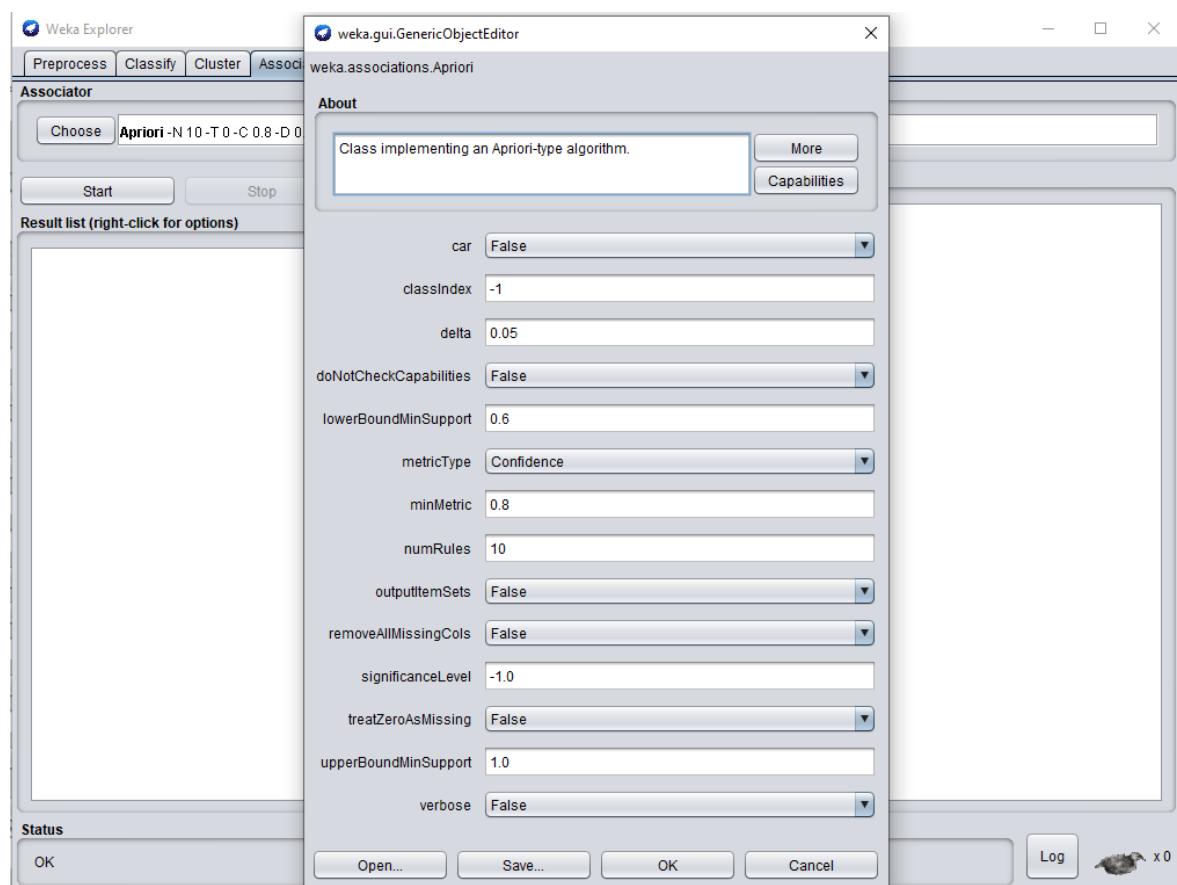
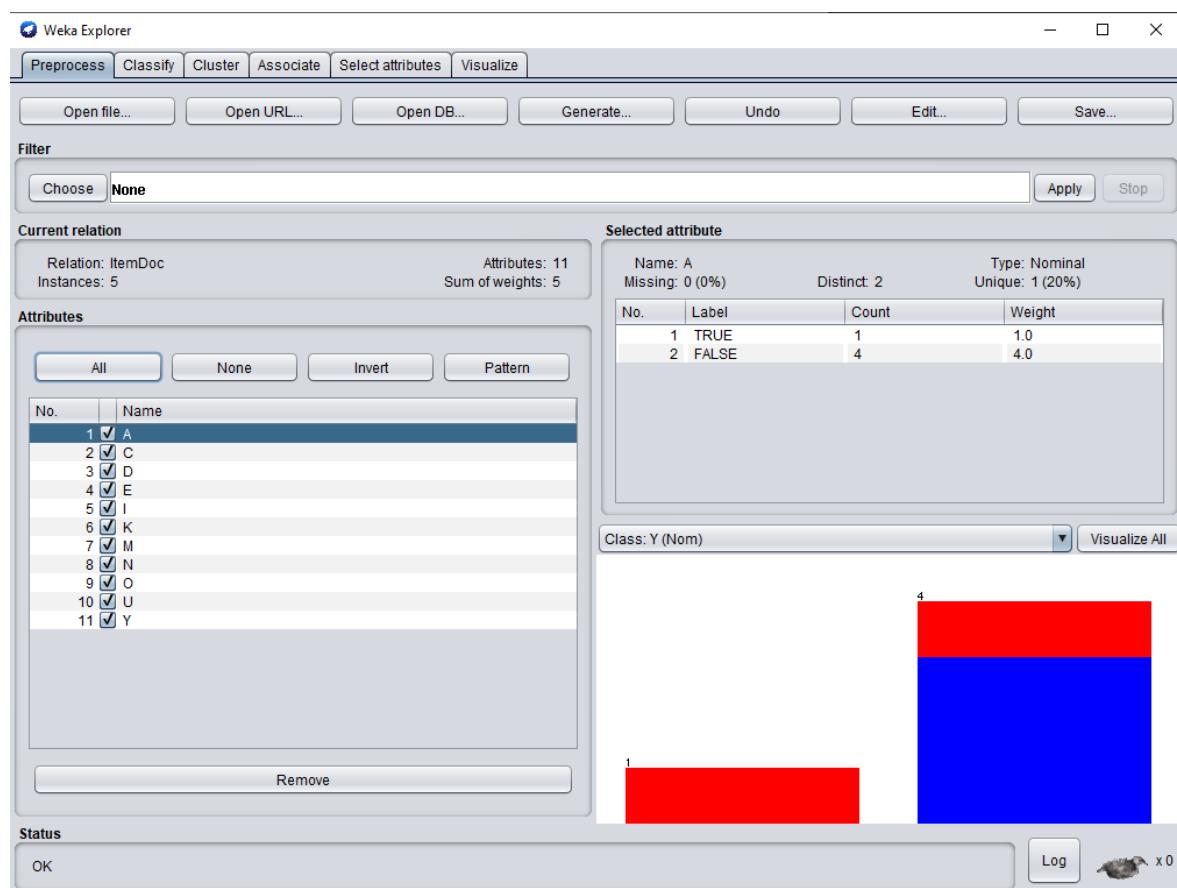
$$\forall x \in \text{transaction}, \text{buys}(X, \text{item}_1) \wedge \text{buys}(X, \text{item}_2) \Rightarrow \text{buys}(X, \text{item}_3) \quad [s, c]$$

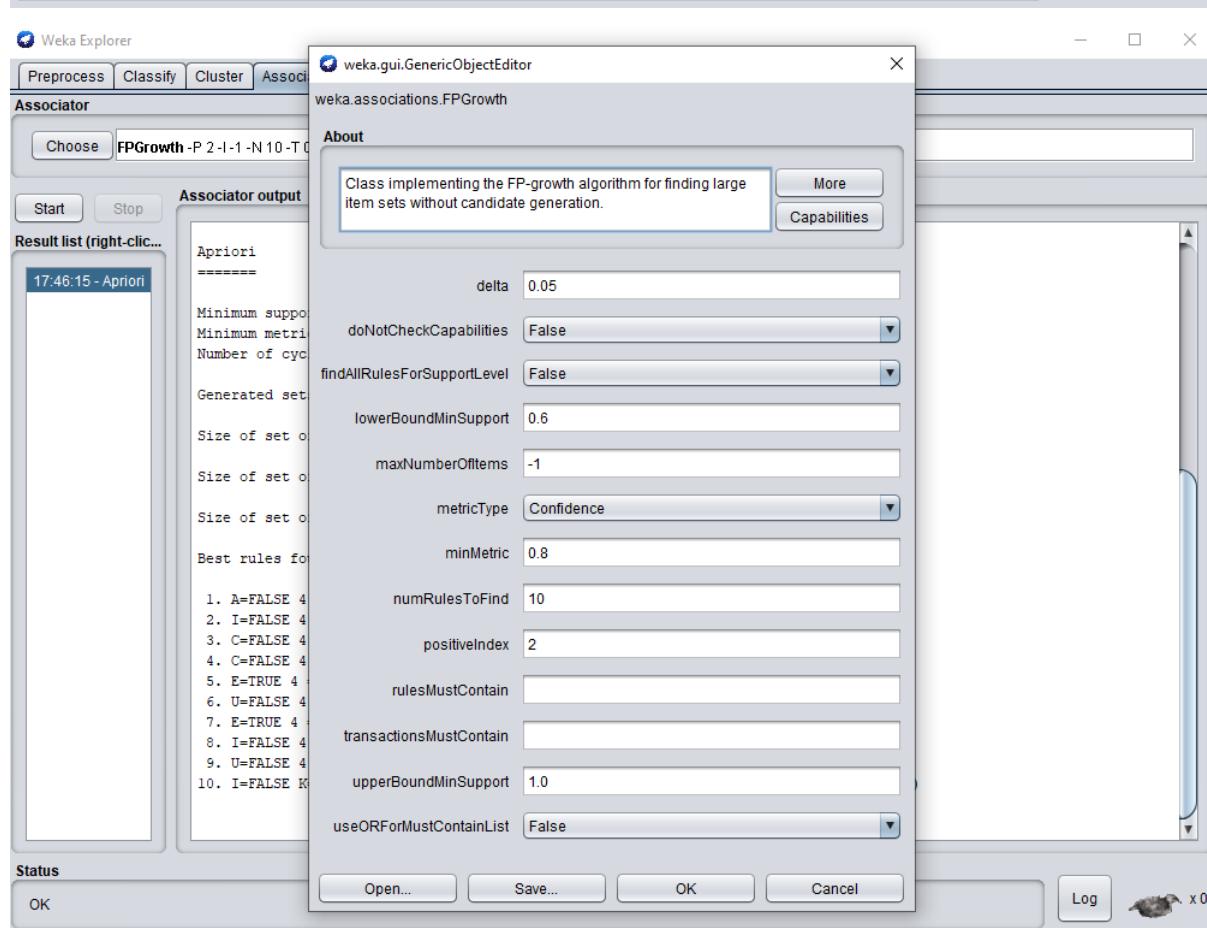
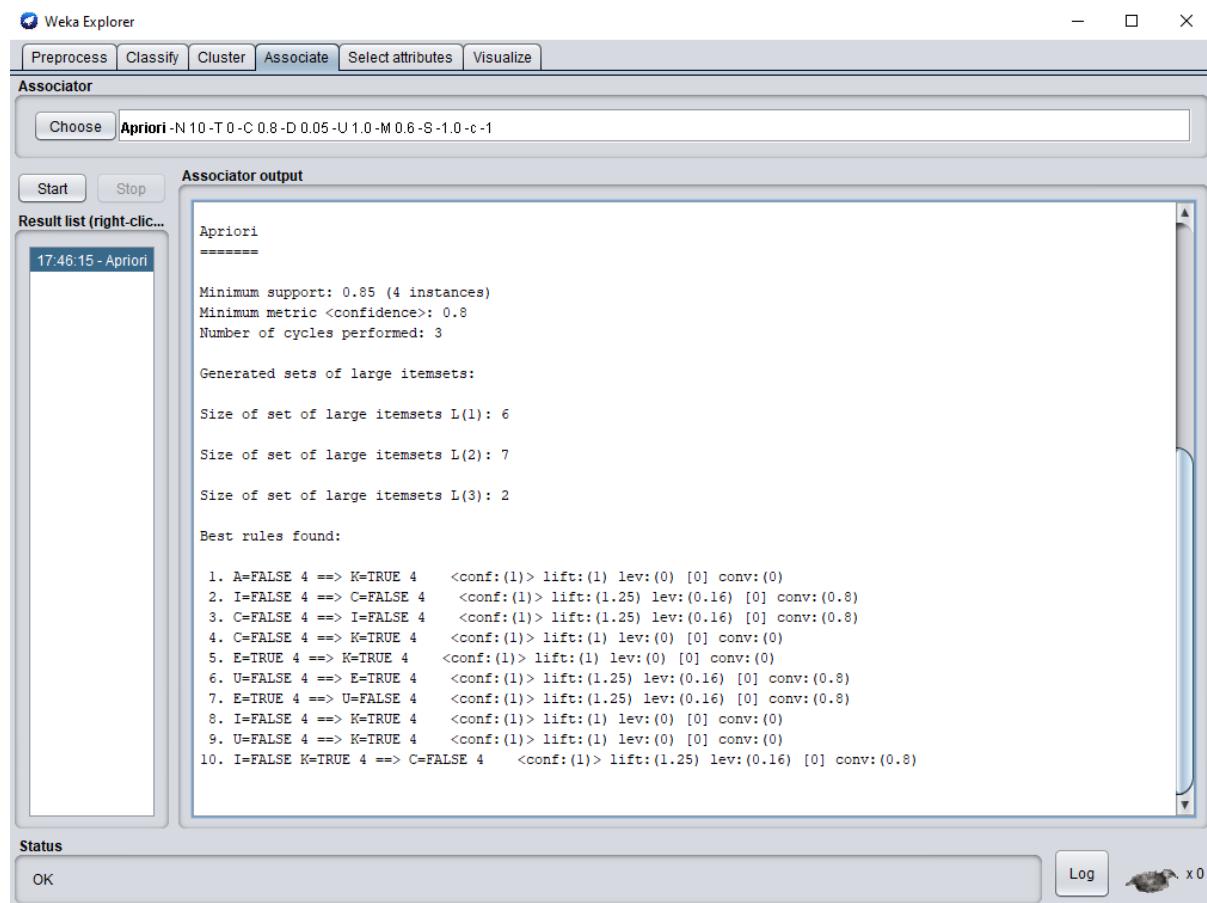
Data file:

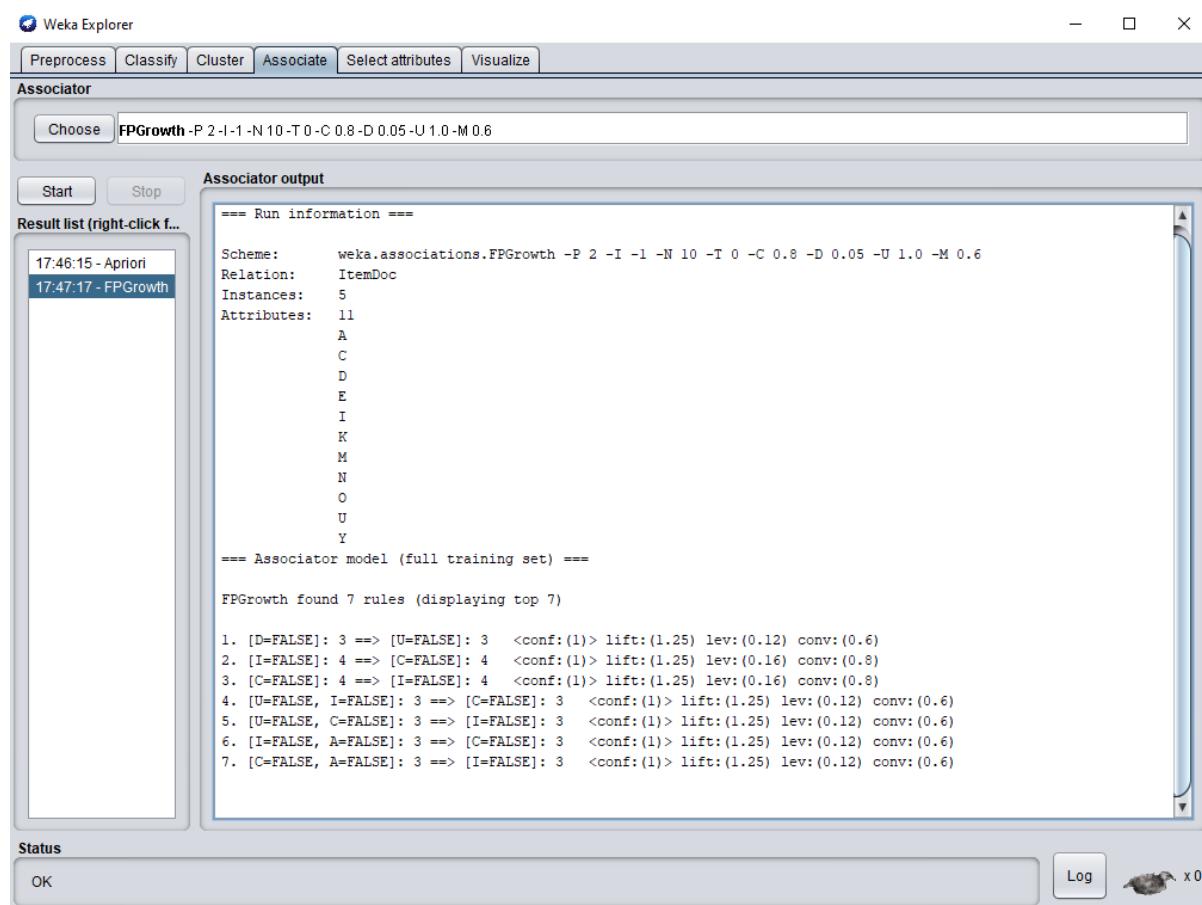
```

@relation ItemDoc
@attribute A {TRUE, FALSE}
@attribute C {TRUE, FALSE}
@attribute D {TRUE, FALSE}
@attribute E {TRUE, FALSE}
@attribute I {TRUE, FALSE}
@attribute K {TRUE, FALSE}
@attribute M {TRUE, FALSE}
@attribute N {TRUE, FALSE}
@attribute O {TRUE, FALSE}
@attribute U {TRUE, FALSE}
@attribute Y {TRUE, FALSE}
@data
FALSE,FALSE,FALSE,TRUE,FALSE,TRUE,TRUE,TRUE,TRUE,FALSE,TRUE
FALSE,FALSE,TRUE,TRUE,FALSE,TRUE,FALSE,TRUE,TRUE,FALSE,TRUE
TRUE,FALSE,FALSE,TRUE,FALSE,TRUE,TRUE,FALSE,FALSE,FALSE,FALSE
FALSE,FALSE,TRUE,FALSE,FALSE,TRUE,TRUE,FALSE,FALSE,TRUE,TRUE
FALSE,TRUE,FALSE,TRUE,TRUE,FALSE,FALSE,TRUE,FALSE,FALSE,TRUE

```

Output:





Q2.

5.5 A database has four transactions. Let $\text{min_sup} = 60\%$ and $\text{min_conf} = 80\%$.

<i>cust_ID</i>	<i>TID</i>	<i>items_bought</i> (in the form of brand-item_category)
01	T100	{King's-Crab, Sunset-Milk, Dairyland-Cheese, Best-Bread}
02	T200	{Best-Cheese, Dairyland-Milk, Goldenfarm-Apple, Tasty-Pie, Wonder-Bread}
01	T300	{Westcoast-Apple, Dairyland-Milk, Wonder-Bread, Tasty-Pie}
03	T400	{Wonder-Bread, Sunset-Milk, Dairyland-Cheese}

- (a) At the granularity of *item_category* (e.g., $item_i$ could be "Milk"), for the following rule template,

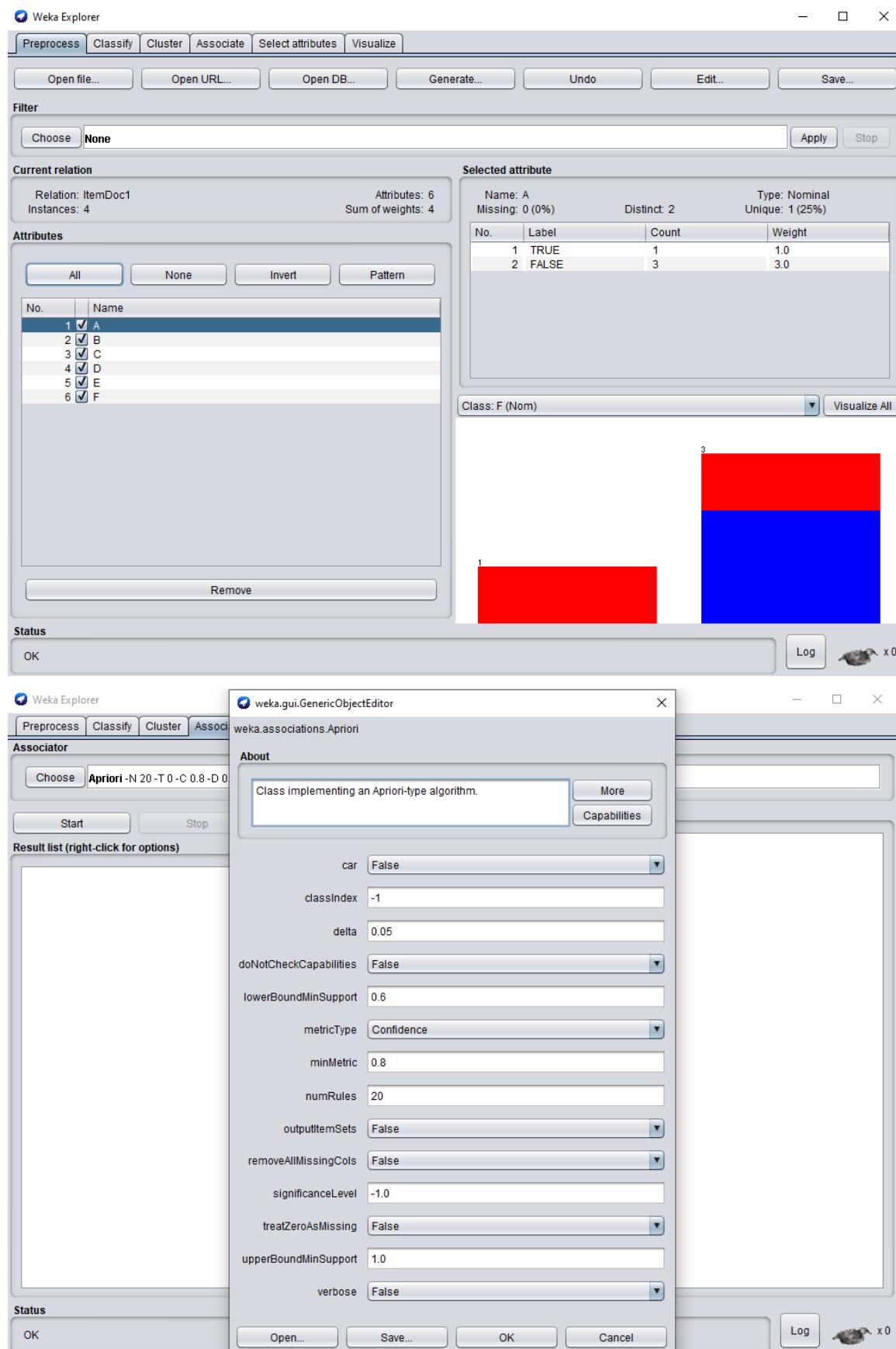
$$\forall X \in \text{transaction}, \text{buys}(X, item_1) \wedge \text{buys}(X, item_2) \Rightarrow \text{buys}(X, item_3) \quad [s, c]$$

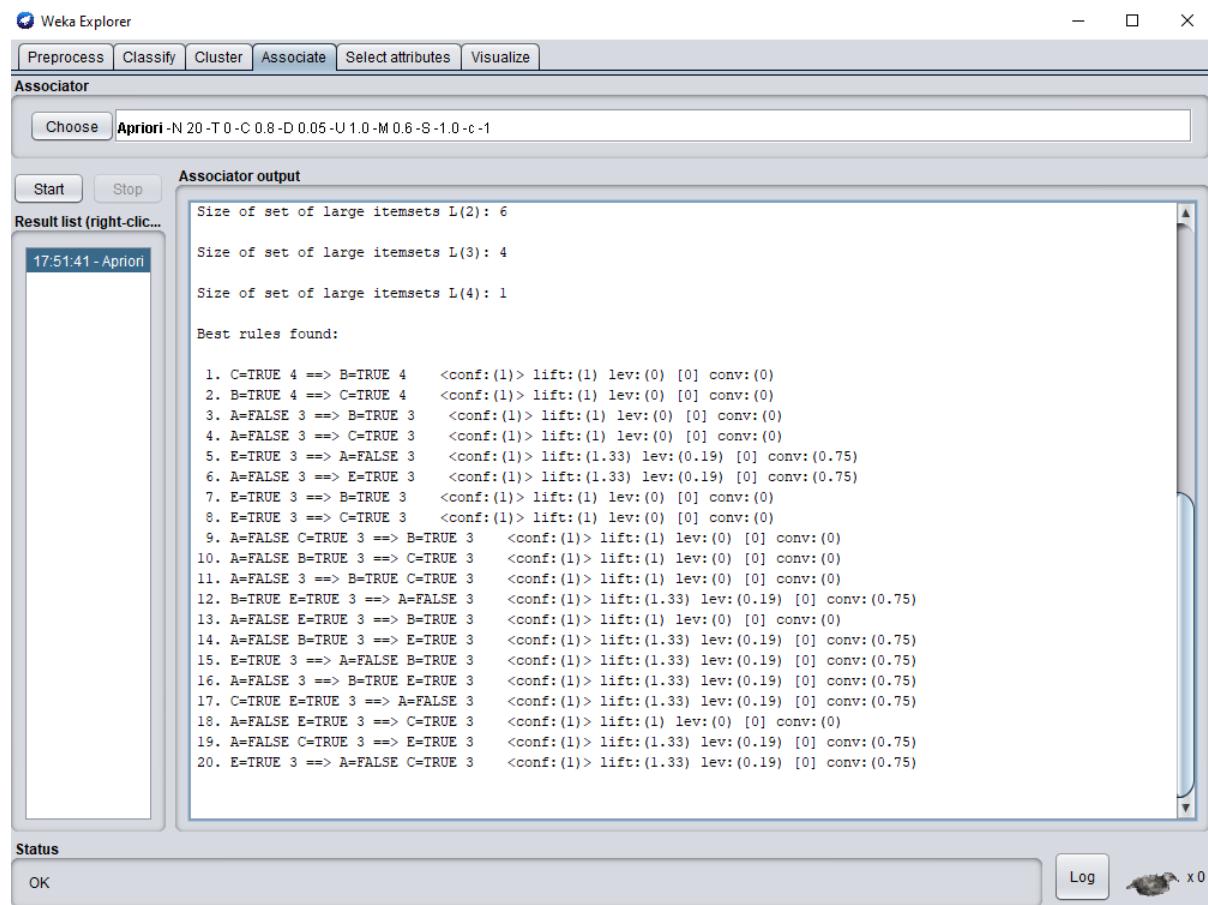
Data file:

```

@relation ItemDoc1
@attribute A {TRUE, FALSE}
@attribute B {TRUE, FALSE}
@attribute C {TRUE, FALSE}
@attribute D {TRUE, FALSE}
@attribute E {TRUE, FALSE}
@attribute F {TRUE, FALSE}
@data
100,TRUE,TRUE,TRUE,TRUE,FALSE,FALSE
200,FALSE,TRUE,TRUE,TRUE,TRUE,TRUE
300,FALSE,TRUE,TRUE,FALSE,TRUE,TRUE
400,FALSE,TRUE,TRUE,FALSE,TRUE,FALSE

```

Output:



(Second Year Afternoon Shift)

(Advanced Web Technology & Data Mining and Business Intelligence Lab)

MINI PROJECT**CERTIFICATE**

This Mini Project duly signed in this Documentation represents the bonafide works by:

NAME: NINAD A PATIL

ROLL NO: 33

NAME: SARATH SHANKARANARAYANAN

ROLL NO: 41

For SEMESTER - IV of Second Year of the Master in Computer Application (MCA) [Afternoon Shift] in the Computer Laboratory of this College during the academic year 2020- 2021

Lecturer In-Charge

Head, Department of MCA

Lab Incharge

External Examiner

ACKNOWLEDGEMENT

It is in good fortune that we find the opportunity to express our deep sense of gratitude to all those people who helped us with their guidance and assistance without which this project would not be possible.

The successful completion of any task would be incomplete without the mention of those people whose cooperation made it possible, whose constant guidance and encouragement crown all our efforts with success.

We would like to give our heartiest thanks to the Project guide, **Prof. MANJU AHUJA AND Prof. DHANAMMA JAGLI** giving us their precious time, incessant encouragement, and for their valuable advice and guidance without which this project would not have seen the light of day.

We also thank our group members and colleagues who have helped in the successful completion of this project. And last but not least, we would like to thank all those who contributed to this project either morally or materially. Thank you all.

INDEX

Sr. No	Topic	Remark
1	Introduction	
2	Abstract	
3	Algorithm Used	
4	Hardware / Software Requirements	
5	Code	
6	Screenshot	
7	Limitation & Future Enhancement	
8	Future Scope Of The Project	
9	Conclusion	
10	Reference	

INTRODUCTION

A decision tree is a structure that contains nodes (rectangular boxes) and edges(arrows) and is built from a dataset (table of columns representing features/attributes and rows corresponds to records). Each node is either used to make a decision (known as decision node) or represent an outcome (known as leaf node).

A decision tree is a decision support tool that uses a tree-like model of decisions and their possible consequences, including chance event outcomes, resource costs, and utility. It is one way to display an algorithm that only contains conditional control statements.

Decision trees are commonly used in operations research, specifically in decision analysis, to help identify a strategy most likely to reach a goal, but are also a popular tool in machine learning.

A decision tree is a flowchart-like structure in which each internal node represents a "test" on an attribute (e.g., whether a coin flip comes up heads or tails), each branch represents the outcome of the test, and each leaf node represents a class label (decision taken after computing all attributes). The paths from root to leaf represent classification rules.

A decision tree consists of three types of nodes:

- **Decision nodes** – typically represented by squares
- **Chance nodes** – typically represented by circles
- **End nodes** – typically represented by triangles

Decision trees are commonly used in operations research and operations management. If, in practice, decisions have to be taken online with no recall under incomplete knowledge, a decision tree should be paralleled by a probability model as a best choice model or online selection model algorithm. Another use of decision tree is as a descriptive means for calculating conditional probabilities.

Among decision support tools, decision trees (and influence diagrams) have several advantages. Decision trees:

- Are simple to understand and interpret. People are able to understand decision tree models
- after a brief explanation.
- Help determine worst, best and expected values for different scenarios.

Disadvantages of decision trees:

- They are unstable, meaning that a small change in the data can lead to a large change in the structure of the optimal decision tree.
- They are often relatively inaccurate. Many other predictors perform better with similar data. This can be remedied by replacing a single decision tree with a random forest of decision trees, but a random forest is not as easy to interpret as a single decision tree.

The project has been developed using: C#

ABSTRACT

“Decision Tree Simulation” is a project which uses ID3 Algorithm for building decision tree and find best attribute. The algorithm builds a decision tree from a sample of data. The training phase results in a decision tree, which will be used to classify future input data. Every leaf represents a value of the result class and a non-leaf node represents a decision node. The main idea behind the implementation is to take the dataset and then break it into smaller pieces until the problem is small enough to be solved. This section will describe how this was achieved and will give an overview about the flow in the code. At the start of the program, the user can either enter data by hand or import a .csv file or text file with the training data. Importing the data is pretty straight forward. The first row holds the column title and the last column contains the outcome. Entering the data by hand works accordingly. First the user has to declare the number of columns. This number must be an integer and greater than 1. Next the user is prompted to enter the title of every column. Then the user can enter the values for the columns. After the program has received the data for training, either by hand or by importing a text file, the program creates the decision tree. This process is called training or learning.

Once the learning process is complete the user has different options on how to interact with the program. The user can either print the created tree, export entered data or let the program predict the result. Printing the tree and exporting the test data are pretty straight forward. The print function will be explained in more detail in the next section.

If the user wants to get a prediction, the program prompts the user to enter a value for every previously defined column. After that the program will predict the outcome and print which route it took to iterate through the tree. If no valid route was found, an error message will be printed.

ALGORITHM USED

- ID3 algorithm is a simple decision tree learning algorithm developed by Ross Quinlan (1983) This algorithm uses the greedy search technique on a given dataset, to test each attribute at every tree node.
- We used this metric to minimize the depth of the tree (minimize asked questions).
- It is a function that measure which questions provide the most balanced splitting the information gain measures how well a given attribute separates training examples into class labels.
- Information gain is expected reduction of entropy related to specific attribute when we split a decision tree node. The information gain Gain (S, A) of example set S, on attribute A is Gain (S, A) = Entropy(S) - $\sum ((|S_v| / |S|) * \text{Entropy}(S_v))$

ALGORITHM

Decision tree builds classification or regression models in the form of a tree structure. It breaks down a dataset into smaller subsets with increase in depth of tree. The final result is a tree with decision nodes and leaf nodes. A decision node (e.g., Outlook) has two or more branches (e.g., Sunny, Overcast and Rainy). Leaf node (e.g., Play) represents a classification or decision. The topmost decision node in a tree which corresponds to the best predictor is called root node. Decision trees can handle both categorical and numerical data.

Types of decision trees

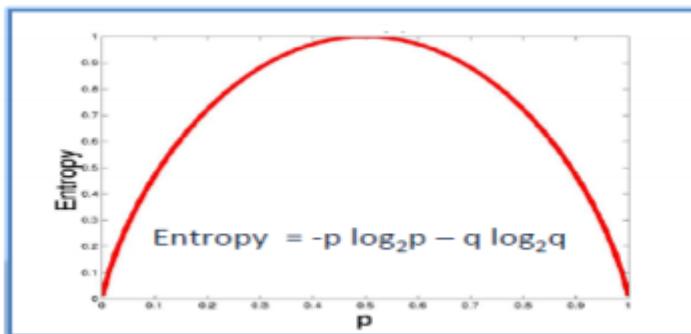
- Categorical Variable Decision Tree: Decision Tree which has categorical target variable then it called as categorical variable decision tree.
- Continuous Variable Decision Tree: Decision Tree which has continuous target variable then it is called as Continuous Variable Decision Tree.

ID3

The core algorithm for building decision trees is called ID3. Developed by J. R. Quinlan, this algorithm employs a top-down, greedy search through the space of possible branches with no backtracking. ID3 uses Entropy and Information Gain to construct a decision tree.

Entropy

A decision tree is built top-down from a root node and involves partitioning the data into subsets that contain instances with similar values (homogeneous). ID3 algorithm uses entropy to calculate the homogeneity of a sample. If the sample is completely homogeneous the entropy is zero and if the sample is equally divided then it has entropy of one.



a) Calculate Entropy (Amount of uncertainty in dataset):

$$\text{Entropy} = \frac{-p}{p+n} \log_2 \left(\frac{p}{p+n} \right) - \frac{n}{p+n} \log_2 \left(\frac{n}{p+n} \right)$$

b) Calculate Average Information:

$$I(Attribute) = \sum \frac{p_i + n_i}{p + n} Entropy(A)$$

Information Gain: The information gain is based on the decrease in entropy after a data-set is split on an attribute. Constructing a decision tree is all about finding attribute that returns the highest information gain (i.e., the most homogeneous branches).

Step 1: Calculate entropy of the target.

Step 2: The dataset is then split on the different attributes. The entropy for each branch is calculated. Then it is added proportionally, to get total entropy for the split. The resulting entropy is subtracted from the entropy before the split. The result is the Information Gain, or decrease in entropy

$$\text{Gain}(T, X) = \text{Entropy}(T) - \text{Entropy}(T, X)$$

Step 3: Choose attribute with the largest information gain as the decision node, divide the dataset by its branches and repeat the same process on every branch.

Step 4a: A branch with entropy of 0 is a leaf node.

Step 4b: A branch with entropy more than 0 needs further splitting.

Step 5: The ID3 algorithm is run recursively on the non-leaf branches, until all data is classified.

HARDWARE/SOFTWARE REQUIREMENTS

HARDWARE REQUIREMENT

- Intel(R) Core (TM) i5-4440 CPU @ 3.10GHz
- 8 GB DDR3 RAM.
- 10 GB Hard Disk

SOFTWARE REQUIREMENT:

- **Operating System:** Windows 7 or higher
- **Software Used:** Visual Studio 2019
- **Programming Language:** C#

CODE

Main.cs

```
using System;
using System.Collections.Generic;
using System.Drawing;
using System.IO;
using System.Linq;
using System.Windows.Forms;

namespace ID3
{
    public partial class Main : Form
    {
        List<Attribute> Attributes = new List<Attribute>();
        DecisionTreeID3 DTID3;
        List<List<string>> Examples = new List<List<string>>();
        int Height, Width = 0;

        public Main()
        {
            InitializeComponent();
            this.StartPosition = FormStartPosition.CenterScreen;
        }

        private void btnnew_Click(object sender, EventArgs e)
        {
            rtxresult.Clear();
            dgvmain.Columns.Clear();
            btnRun.Enabled = false;
            Attributes.Clear();
            Width = 0;
            Bitmap imageDelete = new Bitmap(pbxpaint.Width, pbxpaint.Height);
            pbxpaint.Image = imageDelete;
        }

        private void btnlearn_Click(object sender, EventArgs e)
        {
            Examples.Clear();
            for (int i = 0; i < dgvmain.Rows.Count - 1; i++)
            {
                List<string> example = new List<string>();
                for (int j = 0; j < dgvmain.Columns.Count; j++)
                {
                    example.Add(dgvmain.Rows[i].Cells[j].Value.ToString().ToLower());
                }
                Examples.Add(example);
            }
            List<Attribute> at = new List<Attribute>();
            for (int i = 0; i < Attributes.Count; i++)
            {

```

```

        at.Add(Attributes[i]);
    }
    DTID3 = new DecisionTreeID3(Examples, at);
    DTID3.GetTree();
    Height = DTID3.Depth * 200;
    Width = DTID3.Tree.NumberLabel * 100;
    pbxpaint.Invalidate();
    rtxresult.Text = DTID3.Solution;
}
private void cbxc_CheckedChanged(object sender, EventArgs e)
{
}
private void pbxpaint_Paint(object sender, PaintEventArgs e)
{
    if (Width > 0)
    {
        pbxpaint.Width = Width;
        pbxpaint.Height = Height;
        PaintTree(DTID3.Tree, e, 0, 50);
    }
}
private void PaintTree(TreeNode tree, PaintEventArgs e, int X, int Y)
{
    int XStart = X;
    X = (tree.NumberLabel * 100 + 2 * X) / 2 - 50;
    if (tree.Attributes.Name.ToString() != "")
    {
        e.Graphics.FillRectangle(Brushes.Blue, X, Y, tree.Attributes.Name.Length * 15,
30);
        e.Graphics.DrawString(tree.Attributes.Name.ToString(), new Font("Arial", 20),
Brushes.Red, new PointF(X, Y));
    }
    else
    {
        e.Graphics.FillRectangle(Brushes.Green, X + 25, Y, tree.Attributes.Label.Length *
20, 30);
        e.Graphics.DrawString(tree.Attributes.Label, new Font("Arial", 20),
Brushes.Yellow, new PointF(X + 25, Y));
    }
    int XEndA;
    for (int i = 0; i < tree.Attributes.Value.Count; i++)
    {
        XEndA = tree.Childs[i].NumberLabel * 100 + XStart;
        int XA = (XStart + XEndA) / 2 - 50;
        e.Graphics.DrawLine(new Pen(Brushes.Black, 2), X + 50, Y + 30, XA + 50, Y +
100);
        e.Graphics.DrawString(tree.Attributes.Value[i].ToString(), new Font("Arial", 20),
Brushes.Blue, new PointF(XA, Y + 100));
    }
}

```

```

e.Graphics.DrawLine(new Pen(Brushes.Black, 2), XA + 50, YA + 130, XA + 50, YA
+ 200);
    PaintTree(tree.Childs[i], e, XStart, YA + 200);
    XStart = XEndA;
}
}

private void btnloaddata_Click(object sender, EventArgs e)
{
    dgvmain.Columns.Clear();
    rtxresult.Clear();
    Attributes.Clear();
    OpenFileDialog OpenDiag = new OpenFileDialog();
    OpenDiag.InitialDirectory = Application.StartupPath;
    OpenDiag.DefaultExt = ".txt";
    OpenDiag.Filter = "Text documents (.txt)|*.txt";
    if (OpenDiag.ShowDialog() == DialogResult.OK)
    {
        StreamReader sr = new StreamReader(OpenDiag.FileName);
        string line = "";
        if ((line = sr.ReadLine()) != null)
        {
            string[] attributeName = line.Trim().ToLower().Split('\t').ToArray();
            for (int i = 0; i < attributeName.Length - 1; i++)
            {
                Attribute temp = new Attribute();
                temp.Name = attributeName[i];
                Attributes.Add(temp);
                DataGridViewTextBoxColumn column = new
                DataGridViewTextBoxColumn();
                column.HeaderText = attributeName[i].ToUpper();
                column.Name = attributeName[i].ToUpper();
                dgvmain.Columns.Add(column);
            }
            DataGridViewTextBoxColumn columnend = new
            DataGridViewTextBoxColumn();
            columnend.HeaderText = attributeName[attributeName.Length - 1].ToUpper();
            columnend.Name = attributeName[attributeName.Length - 1].ToUpper();
            dgvmain.Columns.Add(columnend);
            btnRun.Enabled = true;
        }
        while ((line = sr.ReadLine()) != null)
        {
            line = line.Trim();
            string[] value = line.Trim().ToLower().Split('\t').ToArray();
            DataGridViewRow dgvr = new DataGridViewRow();
            for (int i = 0; i < value.Length - 1; i++)
            {
                Attributes[i].AddValue(value[i]);
            }
            string[] value2 = line.Trim().ToUpper().Split('\t').ToArray();

```

```
        dgvmain.Rows.Add(value);
    }
    sr.Close();
}
}

private void ID3_Load(object sender, EventArgs e)
{
    btnRun.Enabled = false;
}

private void button1_Click(object sender, EventArgs e)
{
    About frm = new About();
    frm.Show();
    frm.StartPosition = FormStartPosition.CenterScreen;
}
}
```

ID3.cs

```
using System;
using System.Collections.Generic;
namespace ID3
{
    class DecisionTreeID3
    {
        List<List<string>> Examples;
        List<Attribute> Attributes;
        public List<string> RuleID3 = new List<string>();
        TreeNode _tree;
        int _depth;
        string _solution;
        public int ruleCount;
        string Rule;
        internal TreeNode Tree
        {
            get { return _tree; }
            set { _tree = value; }
        }
        public int Depth
        {
            get { return _depth; }
            set { _depth = value; }
        }
        public string Solution
        {
            get { return _solution; }
            set { _solution = value; }
        }
        public DecisionTreeID3(List<List<string>> Examples, List<Attribute> Attributes)
```

```

{
    this.Examples = Examples;
    this.Attributes = Attributes;
    this.Tree = null;
    Depth = 0;
}
// entropy calculation
private double GetEntropy(int Positives, int Negatives)
{
    if (Positives == 0)
        return 0;
    if (Negatives == 0)
        return 0;
    double Entropy;
    int total = Negatives + Positives;
    double RatePositves = (double)Positives / total;
    double RateNegatives = (double)Negatives / total;
    Entropy = -RatePositves * Math.Log(RatePositves, 2) - RateNegatives *
    Math.Log(RateNegatives, 2);
    return Entropy;
}
// count Gain(bestat,A);
private double Gain(List<List<string>> Examples, Attribute A, string bestat)
{
    double result;
    int CountPositives = 0;
    int[] CountPositivesA = new int[A.Value.Count];
    int[] CountNegativeA = new int[A.Value.Count];
    int Col = Attributes.IndexOf(A);
    for (int i = 0; i < A.Value.Count; i++)
    {
        CountPositivesA[i] = 0;
        CountNegativeA[i] = 0;
    }
    for (int i = 0; i < Examples.Count; i++)
    {
        int j = A.Value.IndexOf(Examples[i][Col].ToString());
        if (Examples[i][Examples[0].Count - 1] == "1")
        {
            CountPositives++;
            CountPositivesA[j]++;
        }
        else
        {
            CountNegativeA[j]++;
        }
    }
    result = GetEntropy(CountPositives, Examples.Count - CountPositives);
    for (int i = 0; i < A.Value.Count; i++)
    {
}

```

```

        double RateValue = (double)(CountPositivesA[i] + CountNegativeA[i]) /
Examples.Count;
        result = result - RateValue * GetEntropy(CountPositivesA[i], CountNegativeA[i]);
    }
    Solution = Solution + "\n * Gain(" + bestat + "," + A.Name + ") = " +
result.ToString();
    return result;
}
// ID3 algorithm
private TreeNode ID3(List<List<string>> Examples, List<Attribute> Attribute, string
bestat)
{
    Solution = Solution + "----- Review " + bestat + " -----";
    if (CheckAllPositive(Examples))
    {
        Solution += "\n All samples assert => Returns the root node with the label 1";
        return new TreeNode(new Attribute("1"));
    }
    if (CheckAllNegative(Examples))
    {
        Solution += "\n All samples are negative => Returns the root node with the label
0";
        return new TreeNode(new Attribute("0"));
    }
    if (Attribute.Count == 0)
    {
        Solution += "\n Empty properties => Returns the root node with the most common
value ";
        return new TreeNode(new Attribute(GetMostCommonValue(Examples)));
    }
    Attribute BestAttribute = GetBestAttribute(Examples, Attribute, bestat);
    int LocationBA = Attributes.IndexOf(BestAttribute);
    TreeNode Root = new TreeNode(BestAttribute);
    for (int i = 0; i < BestAttribute.Value.Count; i++)
    {
        List<List<string>> Examplesvi = new List<List<string>>();
        for (int j = 0; j < Examples.Count; j++)
        {
            if (Examples[j][LocationBA].ToString() == BestAttribute.Value[i].ToString())
                Examplesvi.Add(Examples[j]);
        }
        if (Examplesvi.Count == 0)
        {
            Solution += "\n Empty properties => Returns the root node with the most
common value ";
            return new TreeNode(new Attribute(GetMostCommonValue(Examplesvi)));
        }
        else
        {

```

```

        Solution += "\n";
        Attribute.Remove(BestAttribute);
        Root.AddNode(ID3(Examplesvi, Attribute, BestAttribute.Value[i]));
    }
}
return Root;
}
// Get the property with the highest Gain
private Attribute GetBestAttribute(List<List<string>> Examples, List<Attribute>
Attributes, string bestat)
{
    double MaxGain = Gain(Examples, Attributes[0], bestat);
    int Max = 0;
    for (int i = 1; i < Attributes.Count; i++)
    {
        double GainCurrent = Gain(Examples, Attributes[i], bestat);
        if (MaxGain < GainCurrent)
        {
            MaxGain = GainCurrent;
            Max = i;
        }
    }
    Solution = Solution + "\n\t=> We choose the biggest Gain : " +
Attributes[Max].Name;
    return Attributes[Max];
}
// take the most common value of the target set
private string GetMostCommonValue(List<List<string>> Examples)
{
    int CountPositive = 0;
    for (int i = 0; i < Examples.Count; i++)
    {
        if (Examples[i][Examples[0].Count - 1] == "1")
            CountPositive++;
    }
    int CountNegative = Examples.Count - CountPositive;
    string Label;
    if (CountPositive > CountNegative)
        Label = "1";
    else
        Label = "0";
    Solution = Solution + " là " + Label;
    return Label;
}
// checks if all sets are positive
private bool CheckAllPositive(List<List<string>> Examples)
{
    for (int i = 0; i < Examples.Count; i++)
    {
        if (Examples[i][Examples[0].Count - 1].ToString() == "0")

```

```

        return false;
    }
    return true;
}
// Check if all episodes are Negative
private bool CheckAllNegative(List<List<string>> Examples)
{
    for (int i = 0; i < Examples.Count; i++)
    {
        if (Examples[i][Examples[0].Count - 1] == "1")
            return false;
    }
    return true;
}
// tree construction
public void GetTree()
{
    Solution = "";
    List<Attribute> at = new List<Attribute>();
    for (int i = 0; i < Attributes.Count; i++)
    {
        at.Add(Attributes[i]);
    }
    Tree = ID3(Examples, at, "S");
    Depth = GetDepth(Tree);
}
// Find the law
public void SearchRule(TreeNode Rule)
{
    if (Rule.Attributes.Value.Count == 0)
    {
    }
    else
    {
        string temp1 = "";
        _solution += Rule.Attributes.Name + " = ";
        temp1 += _solution + " ";
        for (int i = 0; i < Rule.Attributes.Value.Count; i++)
        {
            string temp2 = "";
            temp2 = temp1 + Rule.Attributes.Value[i] + ", ";
            if (Rule.Childs[i].Attributes.Value.Count == 0)
            {
                ruleCount++;
                _solution = temp2 + "} THEN { " + Rule.Childs[i].Attributes.Label + " } ";
                RuleID3.Add(_solution);
            }
            else
            {
                if (Rule.Attributes.Value.Count == 0)
                {

```

```

        SearchRule(Rule.Childs[i]);
    }
    else
    {
        _solution = temp2;
        SearchRule(Rule.Childs[i]);
    }
}
}

public void GetRule(TreeNode tree)
{
    _solution = "";
    Rule += " WITHDRAWAL FROM DECISION PLANT ID3 \n\n";
    SearchRule(tree);
    for (int i = 0; i < ruleCount; i++)
        Rule += " Rule [" + i + "]: IF {" + RuleID3[i] + "\n";
    Rule += "\n Total Law : " + ruleCount;
    ruleCount = 0;
}
// Decision Tree Optimization.
public bool CheckAllLabelNegative(TreeNode tree)
{
    int test = 0;
    string temp;
    temp = "No";
    for (int i = 0; i < tree.Attributes.Value.Count; i++)
    {
        if (tree.Childs[i].Attributes.Label == temp)
            test++;
    }
    if ((test > 1) && (test == tree.Attributes.Value.Count))
        return true;
    else
        return false;
}
// get the depth of the tree
private int GetDepth(TreeNode tree)
{
    int depth;
    if (tree.Childs.Length == 0)
        return 1;
    else
    {
        depth = GetDepth(tree.Childs[0]);
        for (int i = 1; i < tree.Childs.Length; i++)
        {
            int depthchild = GetDepth(tree.Childs[i]);
            if (depth < depthchild)

```

```

    depth = depthchild;
}
depth++;
}
return depth;
}
}
}
```

Attributes.cs

```
using System.Collections.Generic;
```

```
namespace ID3
{
    class Attribute
    {
        List<string> _value;
        string _name;
        string _label;
        public List<string> Value
        {
            get { return _value; }
            set { _value = value; }
        }
        public string Name
        {
            get { return _name; }
            set { _name = value; }
        }
        public string Label
        {
            get { return _label; }
            set { _label = value; }
        }
        public Attribute()
        {
            this.Name = "";
            this.Label = "";
            this.Value = new List<string>();
        }
        public Attribute(List<string> Value, string Name)
        {
            this.Value = Value;
            this.Name = Name;
            this.Label = "";
        }
        public Attribute(string Label)
        {
            this.Label = Label;
        }
    }
}
```

```

        this.Name = string.Empty;
        Value = new List<string>();
    }
    public void AddValue(string Value)
    {
        if (!_value.Contains(Value))
            _value.Add(Value);
    }
}

```

Treenode.cs

```

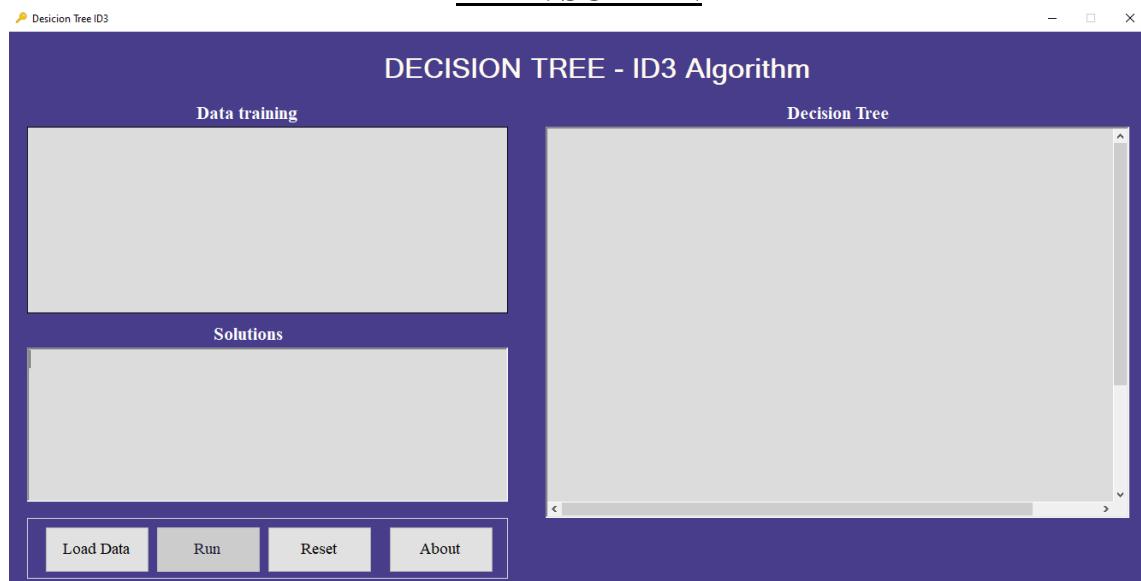
namespace ID3
{
    class TreeNode
    {
        Attribute attributes;
        TreeNode[] child;
        int n;
        int _numberLabel;
        internal Attribute Attributes
        {
            get { return attributes; }
            set { attributes = value; }
        }
        internal TreeNode[] Childs
        {
            get { return child; }
            set { child = value; }
        }
        public int NumberLabel
        {
            get { return _numberLabel; }
            set { _numberLabel = value; }
        }
        public TreeNode(Attribute Attributes)
        {
            this.Attributes = Attributes;
            this.Childs = new TreeNode[Attributes.Value.Count];
            n = 0;
            for (int i = 0; i < Attributes.Value.Count; i++)
            {
                Childs[i] = null;
            }
            if (Attributes.Value.Count == 0)
                NumberLabel = 1;
            else
                NumberLabel = 0;
        }
    }
}

```

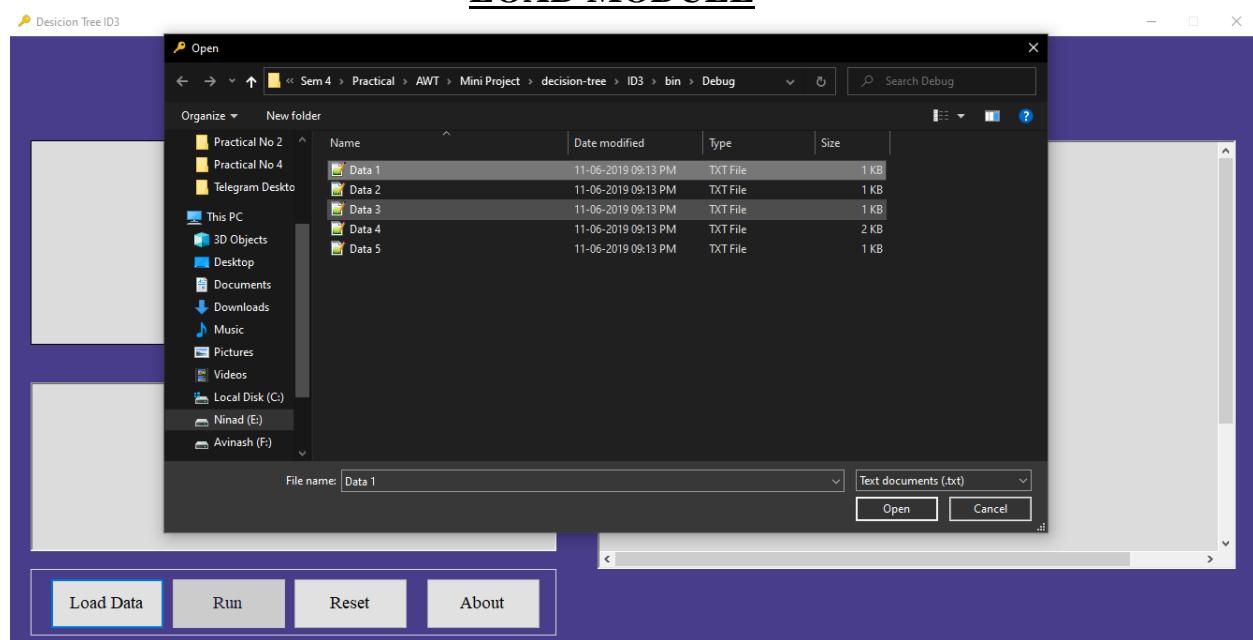
```
public void AddNode(TreeNode Child)
{
    if (n < Childs.Length)
    {
        Childs[n] = Child;
        NumberLabel = NumberLabel + Child.NumberLabel;
    }
    n++;
}
}
```

(SCREENSHOT)

MAIN SCREEN



LOAD MODULE



 Desicion Tree ID3

DECISION TREE - ID3 Algorithm

Data training

	OUTLOOK	TEMPERATUF	HUMIDITY	WIND	PLAYTENNIS
▶	sunny	hot	high	weak	0
	sunny	hot	high	strong	0
	overcast	hot	high	weak	1
	rain	mild	high	weak	1
	rain	cool	normal	weak	1
	rain	cool	normal	strong	0
	overcast	cool	normal	strong	1
	sunny	mild	high	weak	0
	sunny	cool	normal	weak	1

Decision Tree



Solutions

Load Data
Run
Reset
About

RUN MODULE

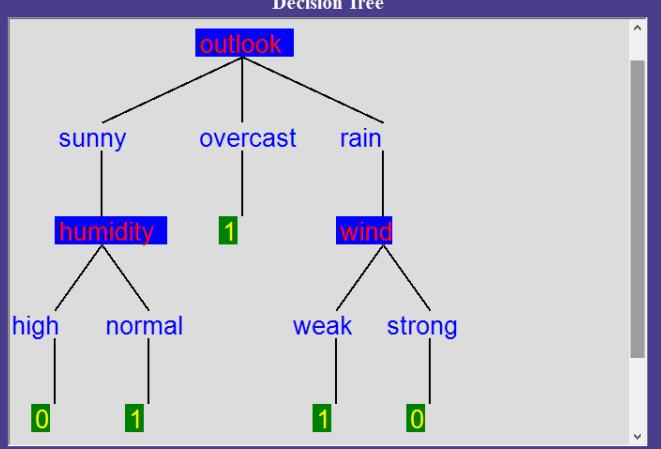
 Desicion Tree ID3

DECISION TREE - ID3 Algorithm

Data training

	OUTLOOK	TEMPERATUF	HUMIDITY	WIND	PLAYTENNIS
▶	sunny	hot	high	weak	0
	sunny	hot	high	strong	0
	overcast	hot	high	weak	1
	rain	mild	high	weak	1
	rain	cool	normal	weak	1
	rain	cool	normal	strong	0
	overcast	cool	normal	strong	1
	sunny	mild	high	weak	0
	sunny	cool	normal	weak	1

Decision Tree



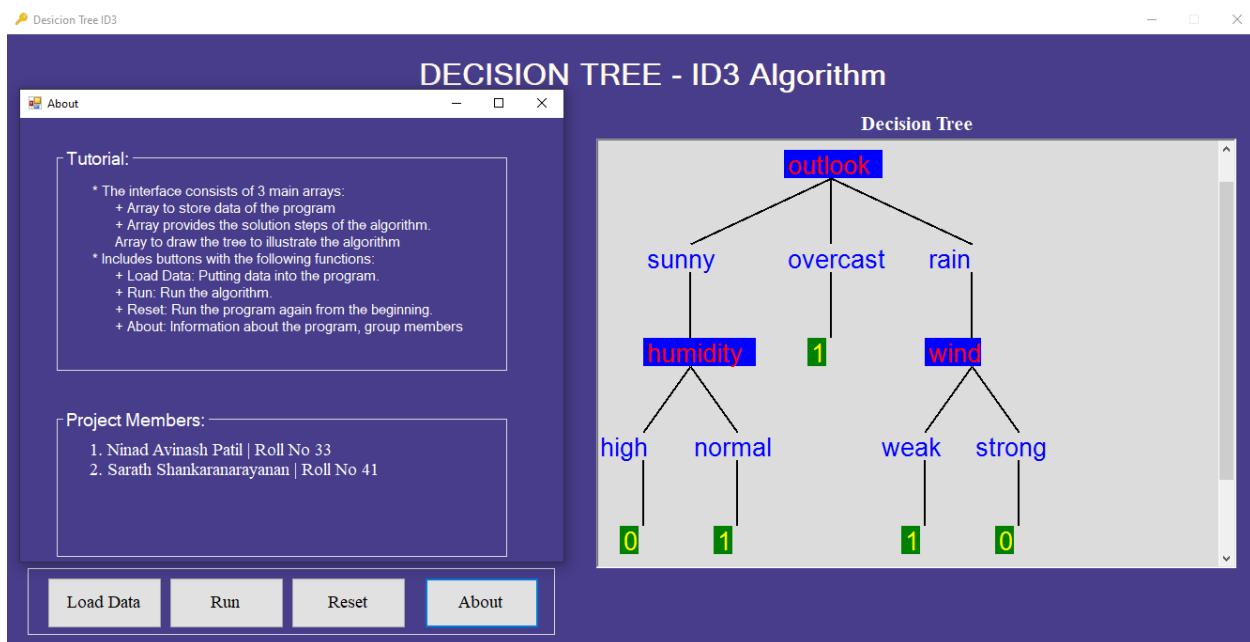
Solutions

Review S
Review sunny

Load Data
Run
Reset
About

```
* Gain(S,outlook) = 0.246749819774439
* Gain(S,temperature) = 0.029225656589546
* Gain(S,humidity) = 0.151835501362341
* Gain(S,wind) = 0.0481270304082693
=> We choose the biggest Gain : outlook
----- Review sunny -----
* Gain(sunny,temperature) = 0.570950594454660
```

ABOUT MODULE



LIMITATIONS AND FUTURE ENHANCEMENT

- The first limitation is that the last column of the training data, the result column, can only contain two different values. These values should be 1 and 0. If the result values are two different values, but not 1 and 0, the program will work but the leaves are not identified as such and therefore printed in a wrong colour.
- The second limitation affects the ability to print the tree. Due to the nature of the console, it is not possible to create a print function which can print all sizes and variations of a tree in a user-friendly way. Therefore, the approach with different colours, separators and casings were chosen.

FUTURE SCOPE OF THE PROJECT

- To add other algorithms
- Future Scope is to add different values to the end column other than 1 and 0

CONCLUSION

- So, we have created a project based on algorithm ID3.
- We have tested this project with different dataset and the result we are getting is in good accuracy.
- So, this project using ID3 algorithm can predict the possibility using the dataset.
- According to the data it will generate a decision tree and we can predict the possibility according to the decision tree.

REFERENCE

Web Reference:

- <https://www.javatpoint.com/machine-learning-decision-tree-classification-algorithm>
- <https://www.javatpoint.com/decision-tree-induction>
- <https://www.guru99.com/c-sharp-windows-forms-application.html>

Textbook Reference:

- Programming in C# by E Balagurusamy.
- Data Mining and Business Intelligence by Abhishek Swaroop, Devesh Agarwal.
- Data Mining Concepts and Techniques 3rd Edition by Jinesh Melvin.