

*Datawarehousing and Analytics*  
*WS 15/16*

*Assignment-2*

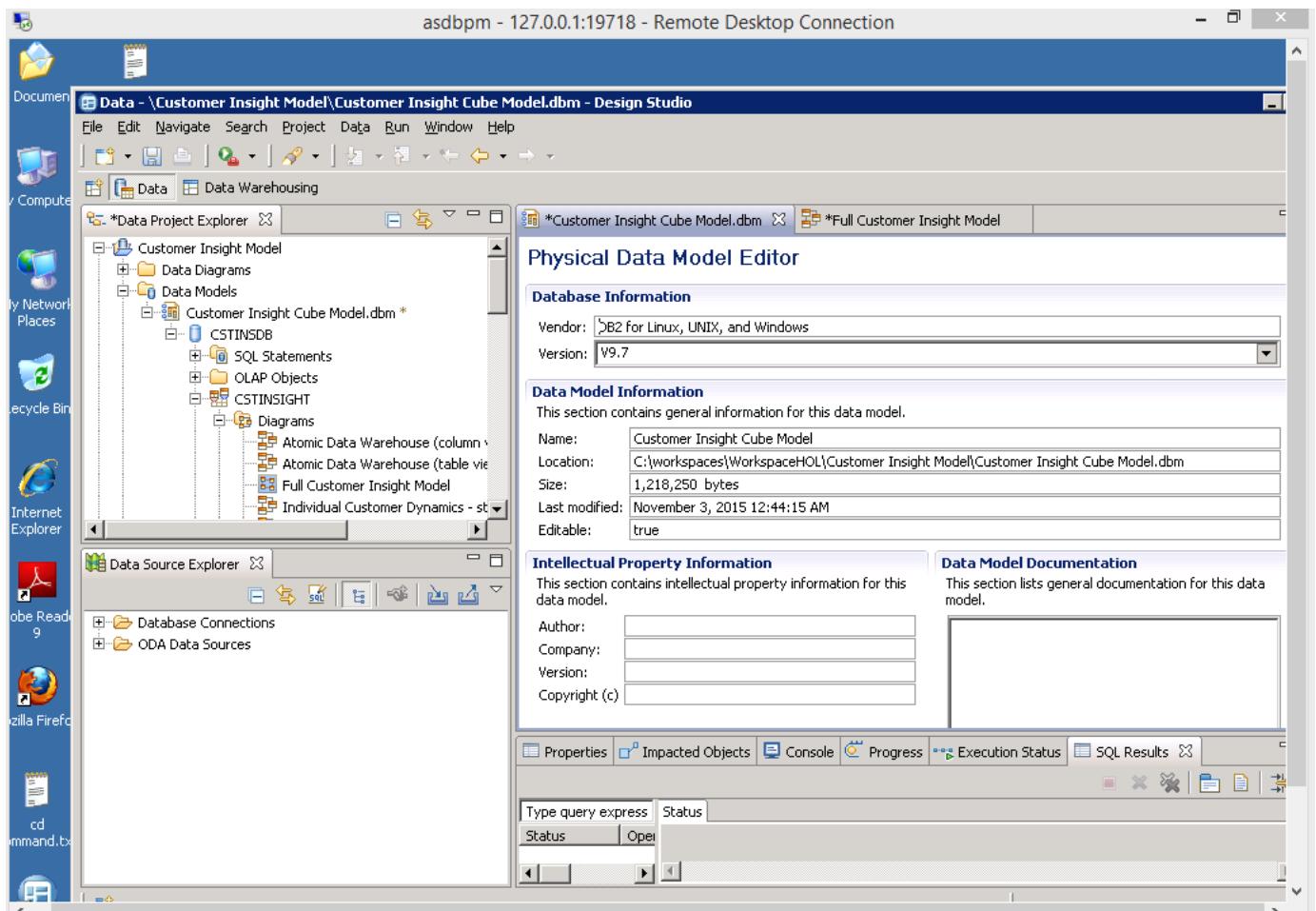
*Team 18*

*Krishna Damarla (2988256 )*  
*Mathumathi Anbalagan (2991285)*  
*Shalini Chellathurai( 2927550)*

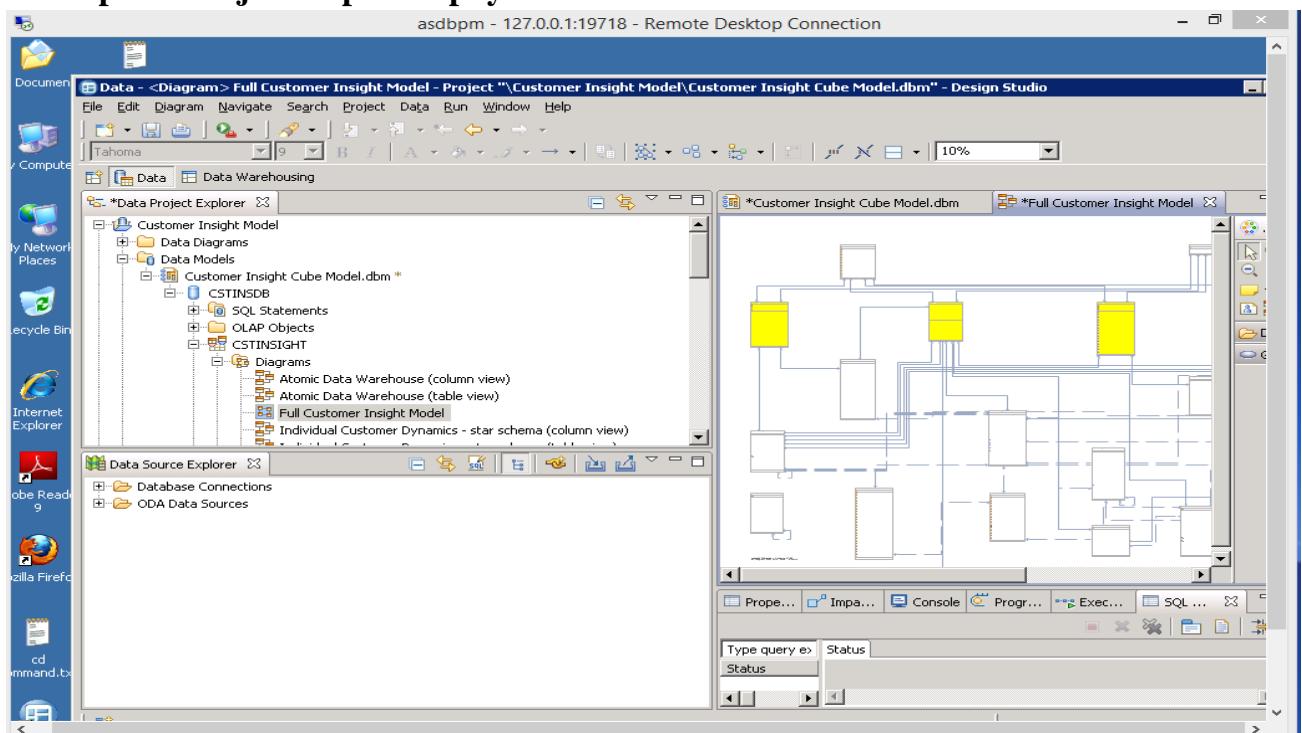
## **TASK 1**

**Using the InfoSphere Warehouse Model Pack for Customer Insight to deploy a basic analytical schema with appropriate Cognos reports**

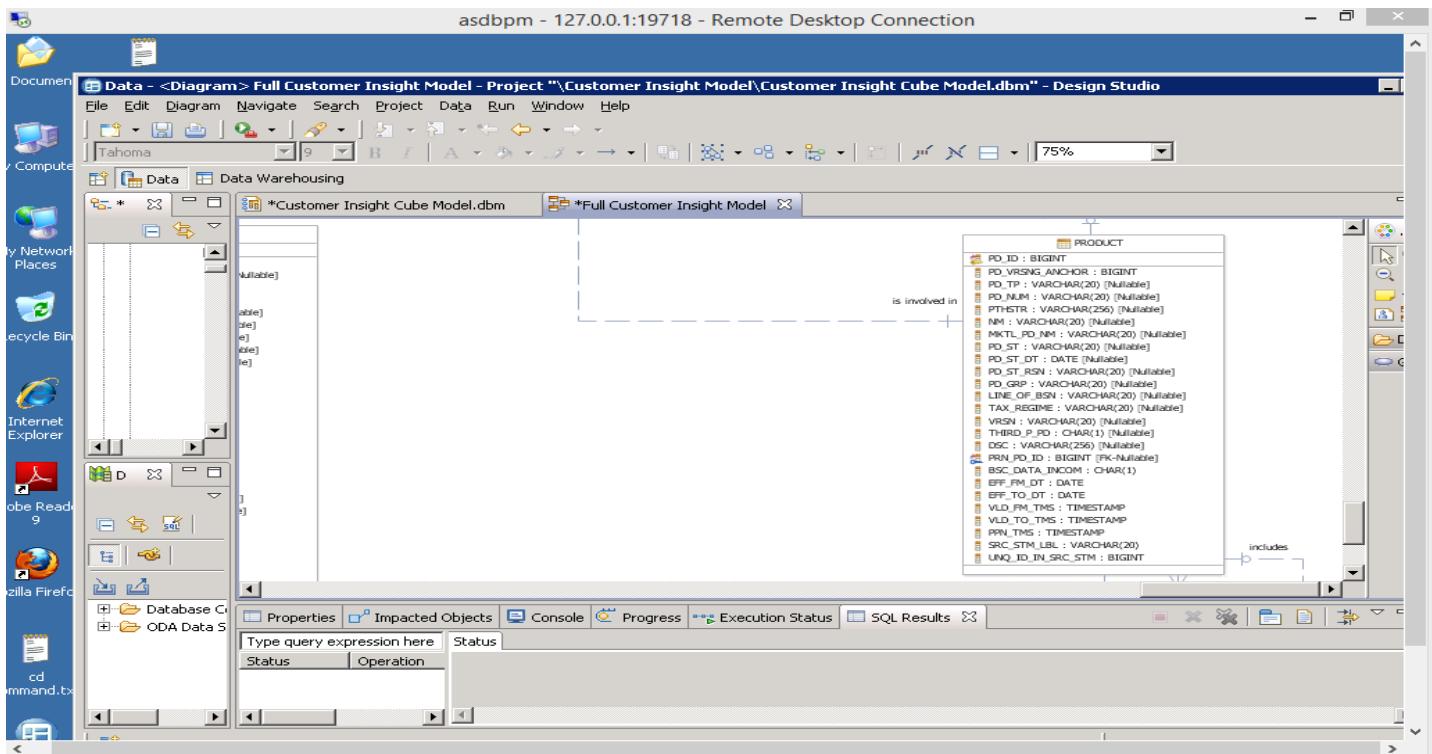
### **1.1 Import the physical data model**



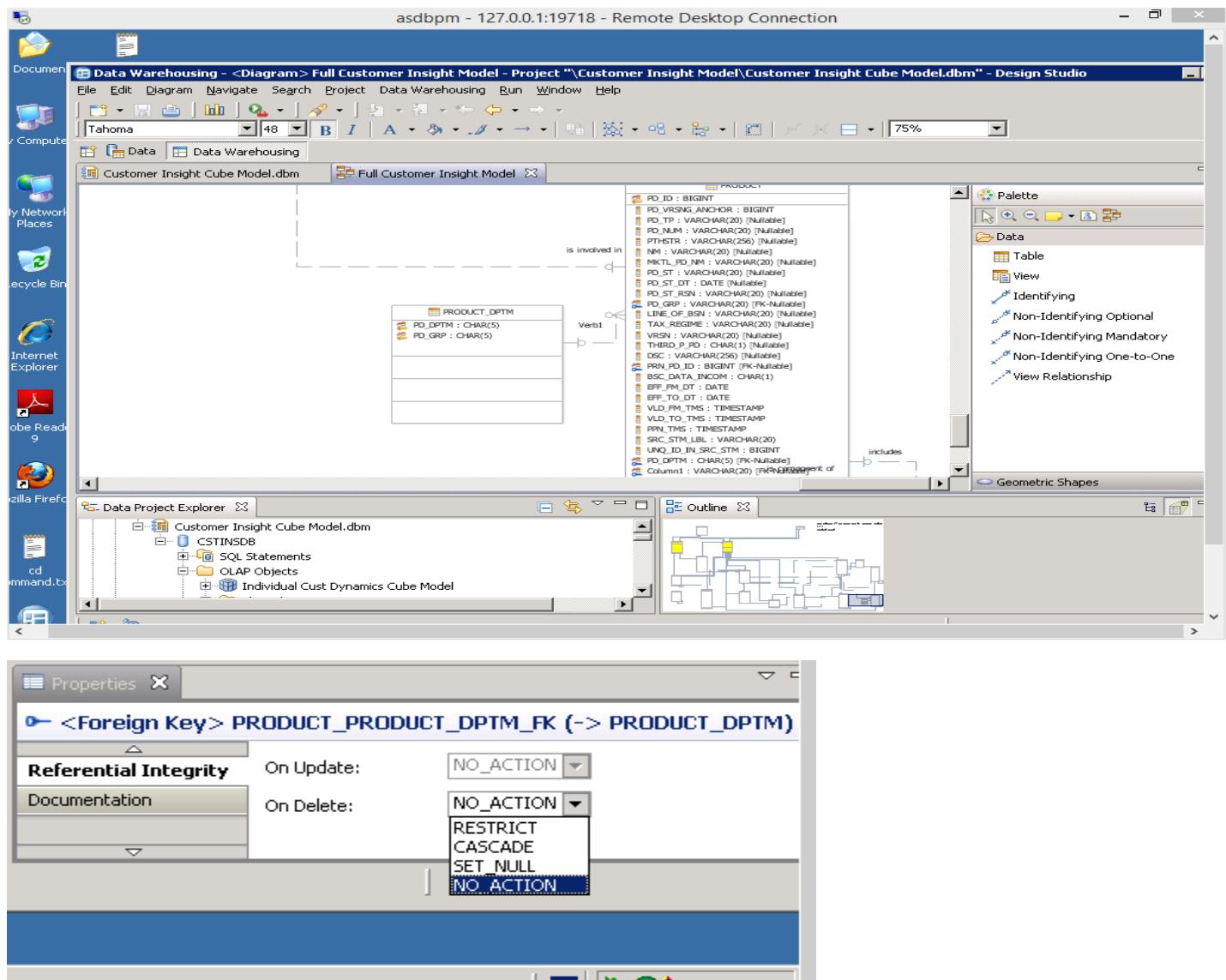
## 1.2 Explore the just imported physical database model



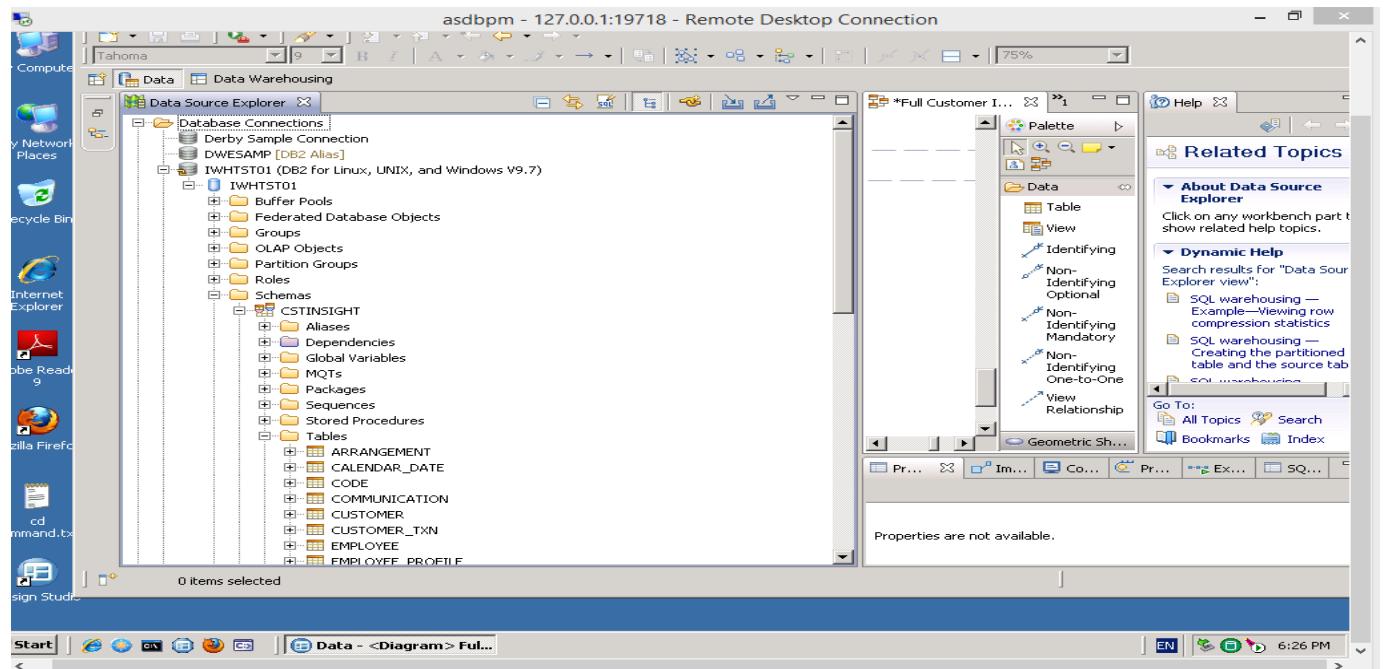
## 1.3 Customize the physical model



- Create a new table named 'PRODUCT\_DPTM' and make PD\_GRP as a primary key.
- After establishing Non-Identifying optional relationship between PRODUCT\_DPTM table and PRODUCT table, you can find the changes in PRODUCT table automatically reflecting PD\_GRP as primary key.

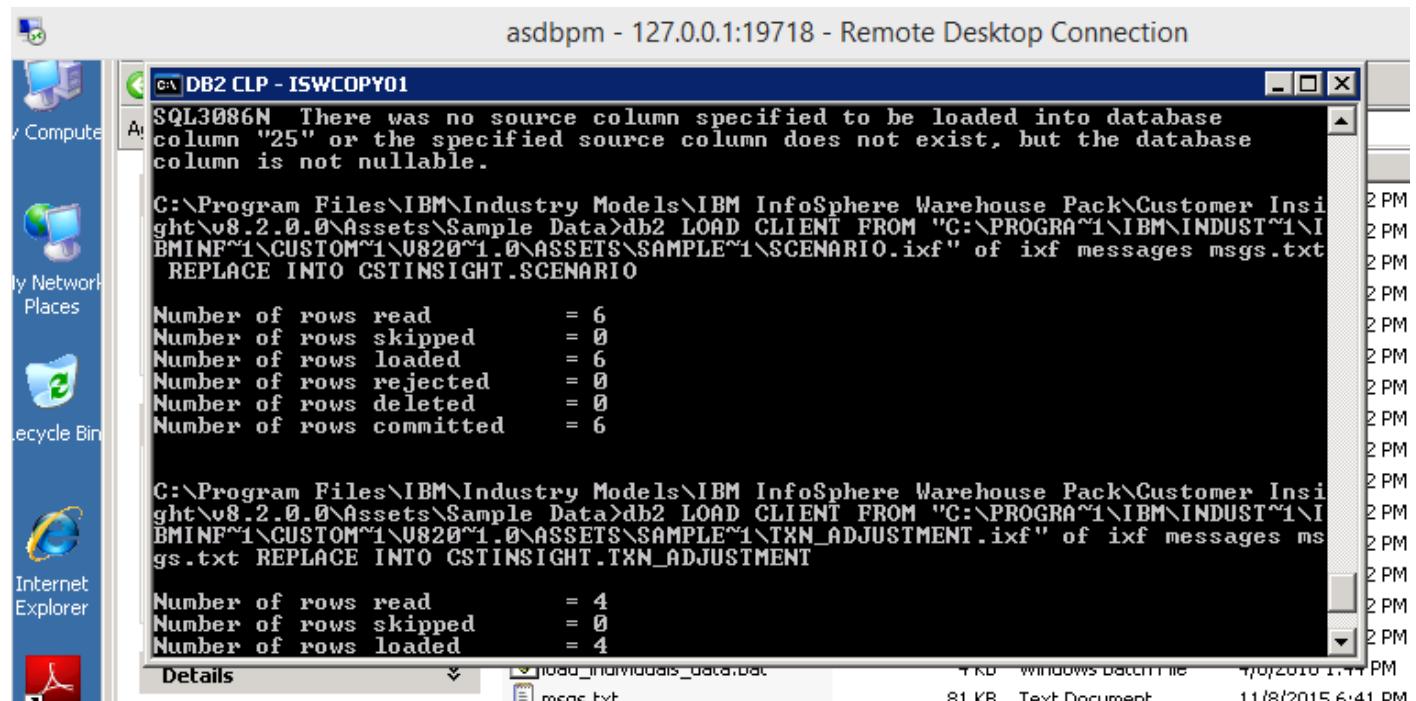


## 1.4 Deploy the physical data model to the database



## 1.5 Load sample data

Change the directory and give the command from windows command prompt.



Check the loaded sample data

The screenshot shows the IBM Data Studio interface. On the left, the Data Source Explorer displays the schema structure under 'CSTINSIGHT'. The 'Tables' section contains numerous tables such as ARRANGEMENT, CALENDAR\_DATE, CODE, COMMUNICATION, CUSTOMER, CUSTOMER\_TXN, EMPLOYEE, EMPLOYEE\_PROFILE, GEO\_AREA, HOUSEHOLD, INDIVIDUAL\_CST\_DYN\_FAC, INDIVIDUAL\_CST\_PROFILE, INDIVIDUAL\_CST\_TXN\_FAC, PERSON, PRODUCT, PRODUCT\_DPTM, and SCENARIO. On the right, the 'Field statistics' tab is active, showing a table with two rows: 'NUM\_OF\_PD\_S...' (Continuous, Value: 0) and 'PCS\_TM' (Continuous, Value: 0). Below the table, it says 'All 0 rows and 24 fields used' and 'Change ...'. The 'Charts for selected fields' tab is also visible, stating 'The field 'NUM\_OF\_PD SOLD' has no statistics information available.'

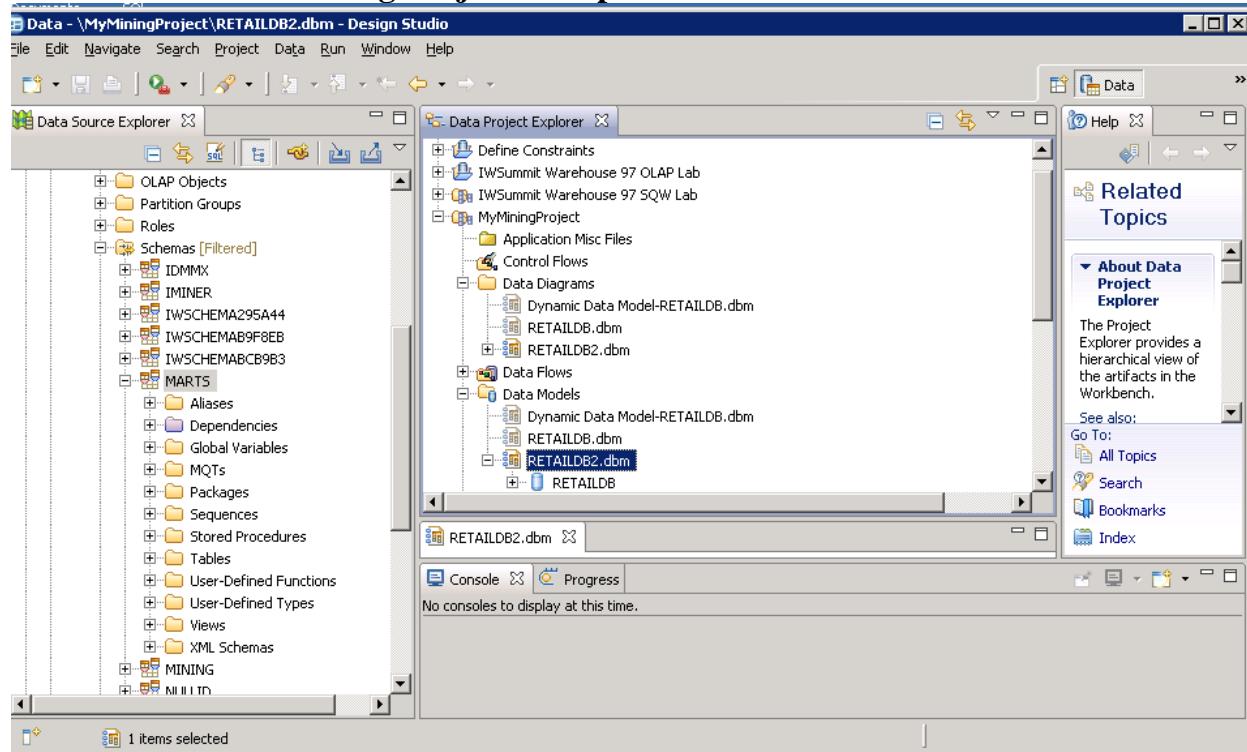
## 1.6 Import the Customer Insight Cognos (OLAP) reports

The screenshot shows the IBM Cognos Viewer interface. At the top, the title bar reads 'Year on Year Profitability Comparison (Individual) - Cognos Viewer - Windows Internet Explorer'. The main area displays a bar chart comparing four categories: Mid Manage & Prof, Skilled Worker, Student, and Unskilled Worker. The Y-axis ranges from 0% to 5%. Below the chart is a grid report with three sections: '2008', 'Change amount', and 'Change %'. The '2008' section includes columns for 'customers', 'income', and 'profit'. The 'Change amount' section includes columns for 'customers', 'income', and 'profit'. The 'Change %' section includes columns for 'customers', 'income', and 'profit'. The data for the '2008' section is as follows:

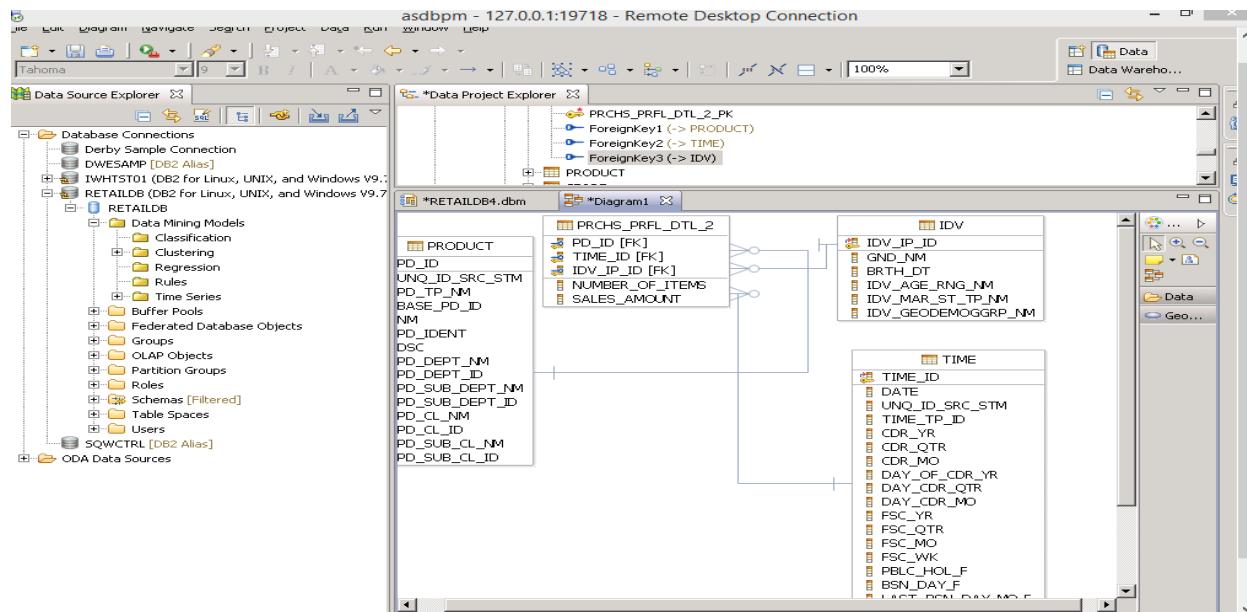
	2008	Change amount	Change %						
ers	income	profit	customers	income	profit	customers	income	profit	
203			2,071			132			6.37%
301			1,724			77			4.46%
350			328			22			6.60%
151			3,062			89			2.92%
<b>04</b>			<b>7,185</b>			<b>320</b>			<b>4.45%</b>

## **TASK 2**

### **2.1 Create a Warehousing Project & import data model**

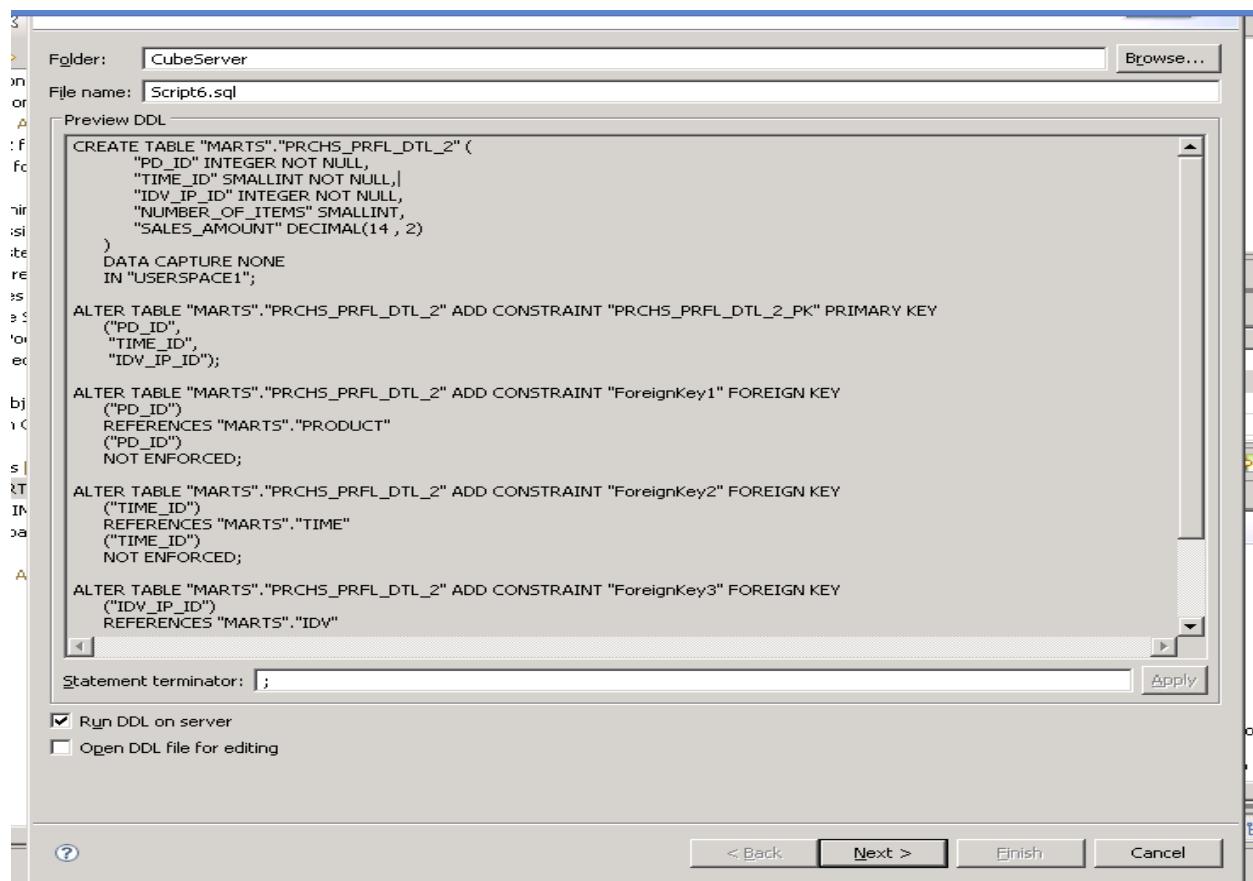


### **2.2 Transform the fact table to suit our needs**



## 2.3 Create the new fact table within the database

Generated DLL



```
CREATE TABLE "MARTS"."PRCHS_PRFL_DTL_2" (
    "PD_ID" INTEGER NOT NULL,
    "TIME_ID" SMALLINT NOT NULL,
    "IDV_IP_ID" INTEGER NOT NULL,
    "NUMBER_OF_ITEMS" SMALLINT,
    "SALES_AMOUNT" DECIMAL(14 , 2)
)
DATA CAPTURE NONE
IN "USERSPACE1";

ALTER TABLE "MARTS"."PRCHS_PRFL_DTL_2" ADD CONSTRAINT "PRCHS_PRFL_DTL_2_PK" PRIMARY KEY
("PD_ID",
"TIME_ID",
"IDV_IP_ID");

ALTER TABLE "MARTS"."PRCHS_PRFL_DTL_2" ADD CONSTRAINT "ForeignKey1" FOREIGN KEY
("PD_ID")
REFERENCES "MARTS"."PRODUCT"
("PD_ID")
NOT ENFORCED;

ALTER TABLE "MARTS"."PRCHS_PRFL_DTL_2" ADD CONSTRAINT "ForeignKey2" FOREIGN KEY
("TIME_ID")
REFERENCES "MARTS"."TIME"
("TIME_ID")
NOT ENFORCED;

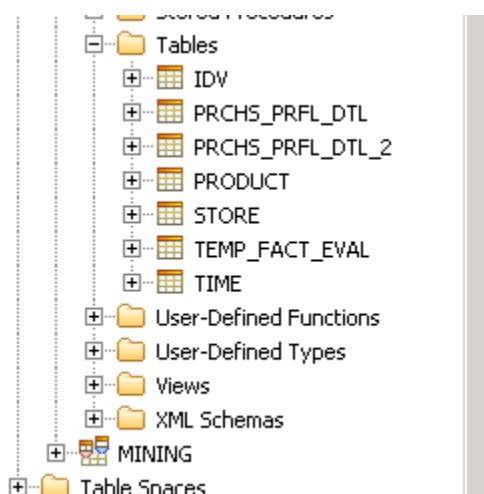
ALTER TABLE "MARTS"."PRCHS_PRFL_DTL_2" ADD CONSTRAINT "ForeignKey3" FOREIGN KEY
("IDV_IP_ID")
REFERENCES "MARTS"."IDV"

Statement terminator: ;

```

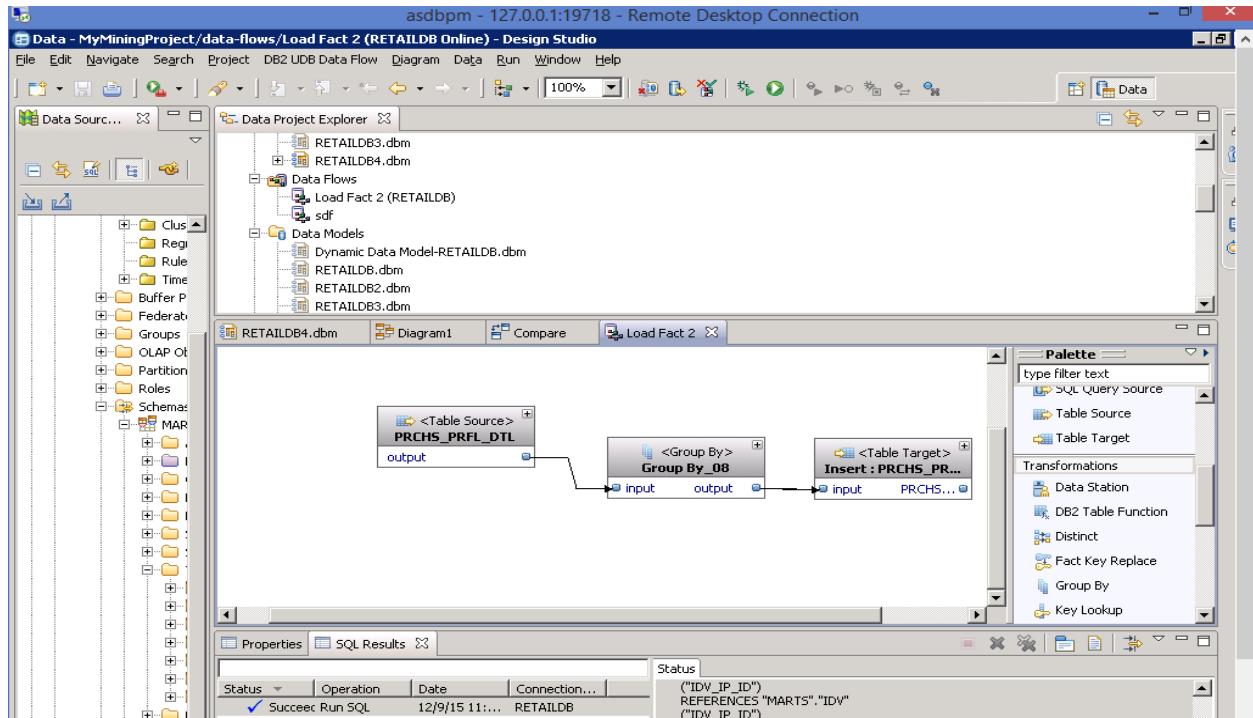
Run DDL on server  
 Open DDL file for editing

Newly assigned fact table in data source explorer

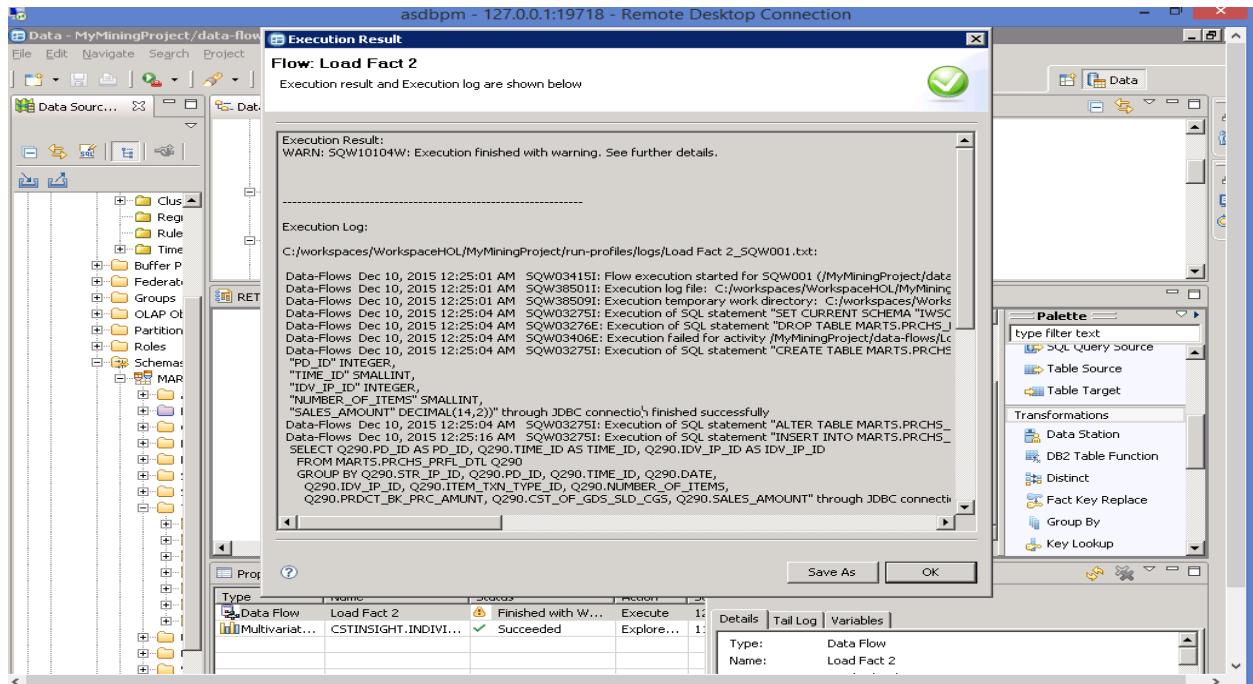


## 2.4 Data flow to load the fact data to the new fact table

Source table, target table, group by operators on canvas with connections



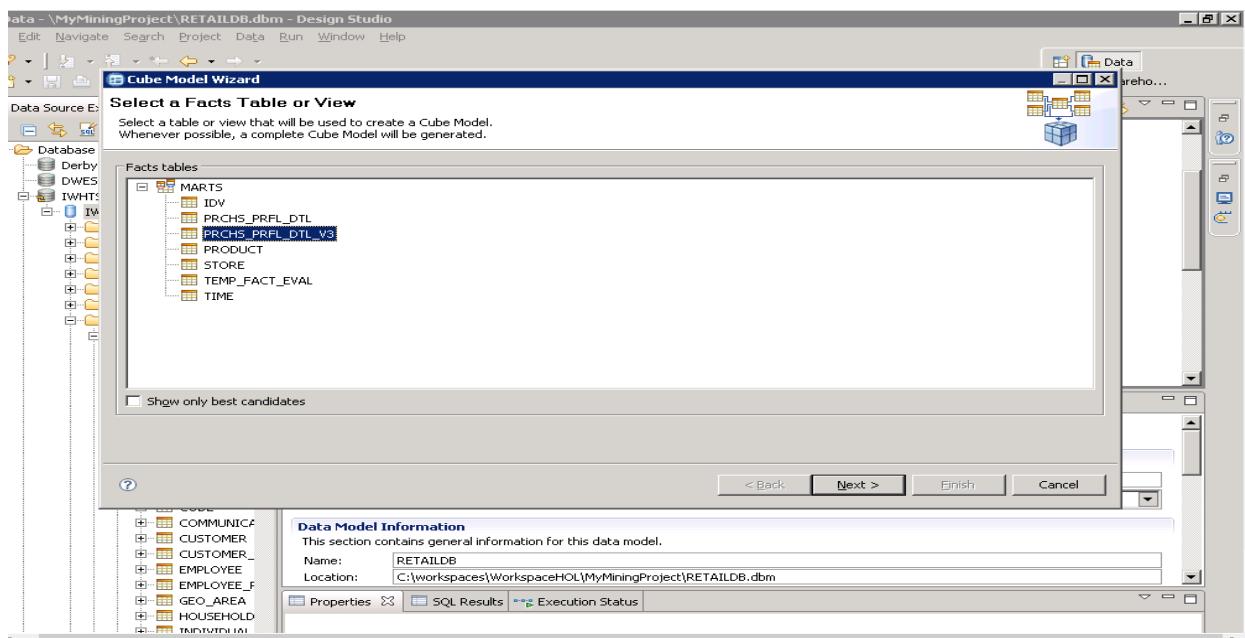
### Execution Summary:



## Sample results of the target table

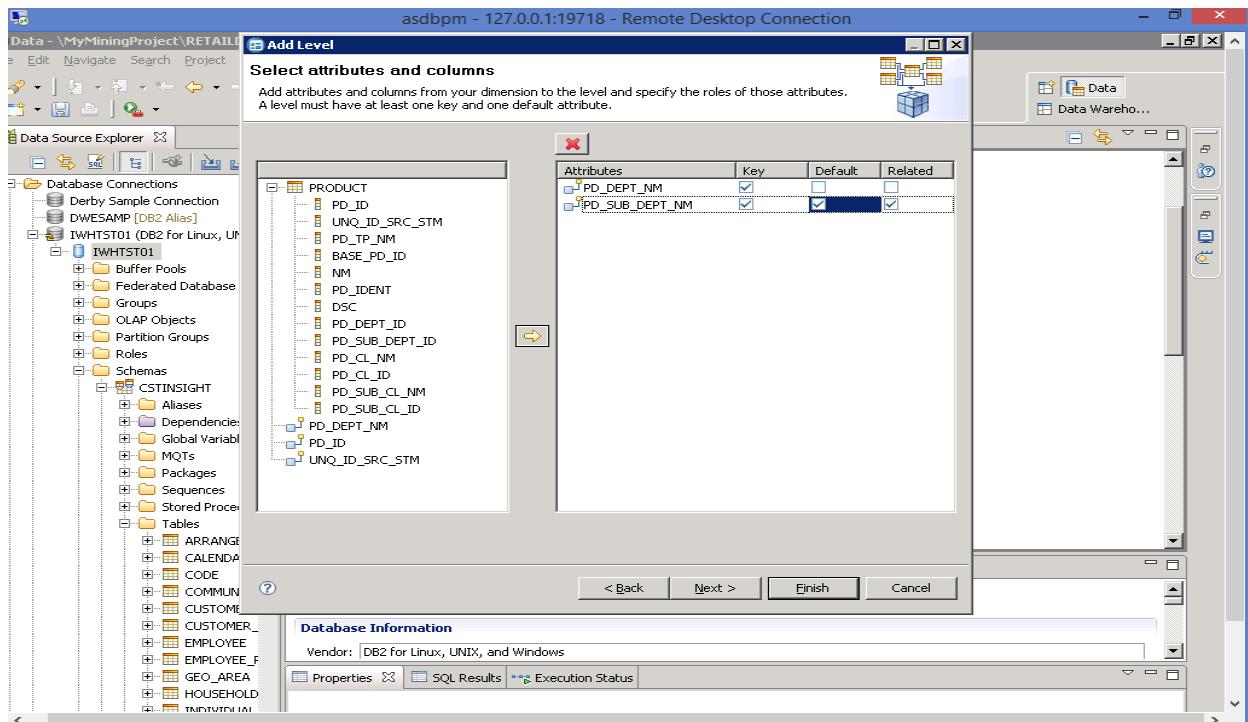
Properties				SQL Results		Execution Status					
Status	Operation	Date	Connection...	Status	Result1	PD_ID	TIME_ID	IDV_IP_ID	NUMBER_OF_ITEMS	SALES_AMOI	
✓	Suc CREATE TA...	11/8/15 6:1...	IWHTST01			1	37286	1041	2002	NULL	NULL
✓	Suc CREATE TA...	11/8/15 6:1...	IWHTST01			2	21603	1050	2007	NULL	NULL
✓	Suc CREATE TA...	11/8/15 6:1...	IWHTST01			3	1731	1243	2020	NULL	NULL
✓	Suc CREATE TA...	11/8/15 6:1...	IWHTST01			4	2668	1243	2020	NULL	NULL
✓	Suc CREATE TA...	11/8/15 6:1...	IWHTST01			5	12080	1243	2020	NULL	NULL
✓	Suc CREATE TA...	11/8/15 6:1...	IWHTST01			6	13350	1243	2020	NULL	NULL
✓	Suc CREATE TA...	11/8/15 6:1...	IWHTST01			7	16023	1243	2020	NULL	NULL
✓	Suc CREATE TA...	11/8/15 6:1...	IWHTST01			8	17626	1243	2020	NULL	NULL
✓	Suc CREATE TA...	11/8/15 6:1...	IWHTST01			9	26287	1243	2020	NULL	NULL
✓	Suc CREATE TA...	11/8/15 6:1...	IWHTST01			10	27247	1243	2020	NULL	NULL
✓	Suc CREATE TA...	11/8/15 6:1...	IWHTST01			11	28879	1243	2020	NULL	NULL
✓	Suc CREATE TA...	11/8/15 6:1...	IWHTST01			12	29389	1243	2020	NULL	NULL
✓	Suc CREATE TA...	11/8/15 6:1...	IWHTST01			13	32625	1243	2020	NULL	NULL
✓	Suc CREATE TA...	11/8/15 6:1...	IWHTST01			14	50654	1243	2020	NULL	NULL
✓	Suc CREATE IND...	11/8/15 6:1...	IWHTST01			15	716	780	2022	NULL	NULL
✓	Suc CREATE IND...	11/8/15 6:1...	IWHTST01			16	2998	780	2022	NULL	NULL
✓	Suc CREATE IND...	11/8/15 6:1...	IWHTST01			17	4799	1368	2022	NULL	NULL
✓	Suc CREATE IND...	11/8/15 6:1...	IWHTST01			18	6616	1344	2022	NULL	NULL
✓	Suc CREATE IND...	11/8/15 6:1...	IWHTST01			19	7850	1368	2022	NULL	NULL
✓	Suc CREATE IND...	11/8/15 6:1...	IWHTST01			20	9280	1368	2022	NULL	NULL
✓	Suc CREATE IND...	11/8/15 6:1...	IWHTST01			21	12160	1210	2022	NULL	NULL
✓	Suc CREATE IND...	11/8/15 6:1...	IWHTST01			22	15352	1368	2022	NULL	NULL
✓	Suc CREATE IND...	11/8/15 6:1...	IWHTST01			23	16660	1344	2022	NULL	NULL
✓	Suc CREATE IND...	11/8/15 6:1...	IWHTST01			24	18392	868	2022	NULL	NULL
✓	Suc CREATE IND...	11/8/15 6:1...	IWHTST01			25	18674	1368	2022	NULL	NULL
✓	Suc CREATE IND...	11/8/15 6:1...	IWHTST01			26	23538	1210	2022	NULL	NULL
✓	Suc CREATE IND...	11/8/15 6:1...	IWHTST01			27	23538	780	2022	NULL	NULL
✓	Suc CREATE IND...	11/8/15 6:1...	IWHTST01			28	26103	780	2022	NULL	NULL
✓	Suc CREATE IND...	11/8/15 6:1...	IWHTST01			29	30100	1210	2022	NULL	NULL

## 2.5 create cube model for PRCHS\_PRFL\_DTL\_V3 Table

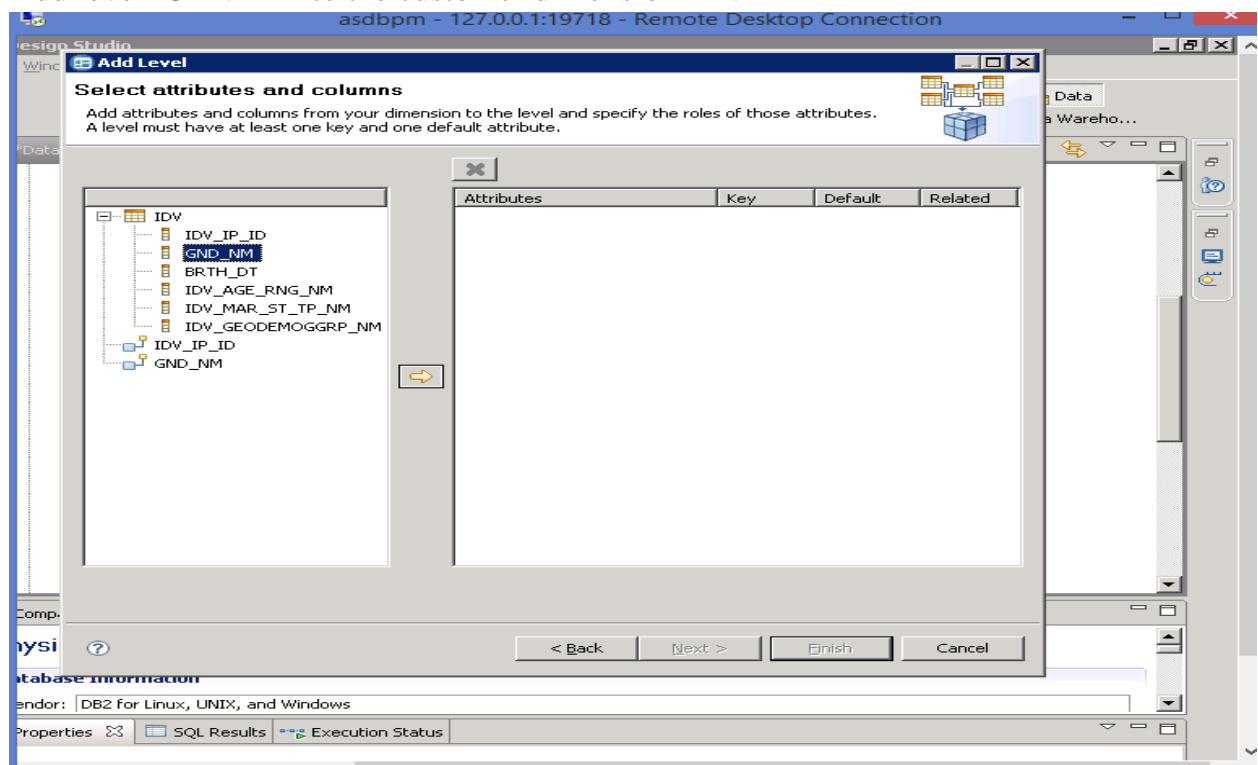


## 2.6 Define the levels and hierarchies for the dimensions

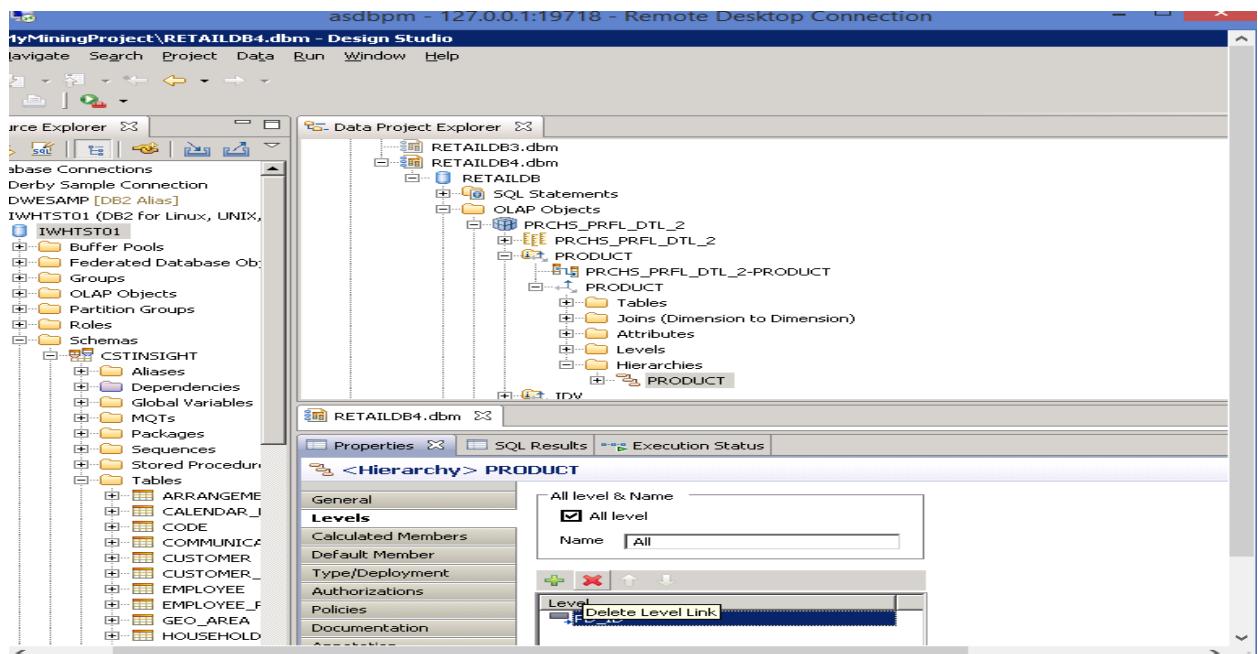
### a) PRODUCT, Time dimension levels



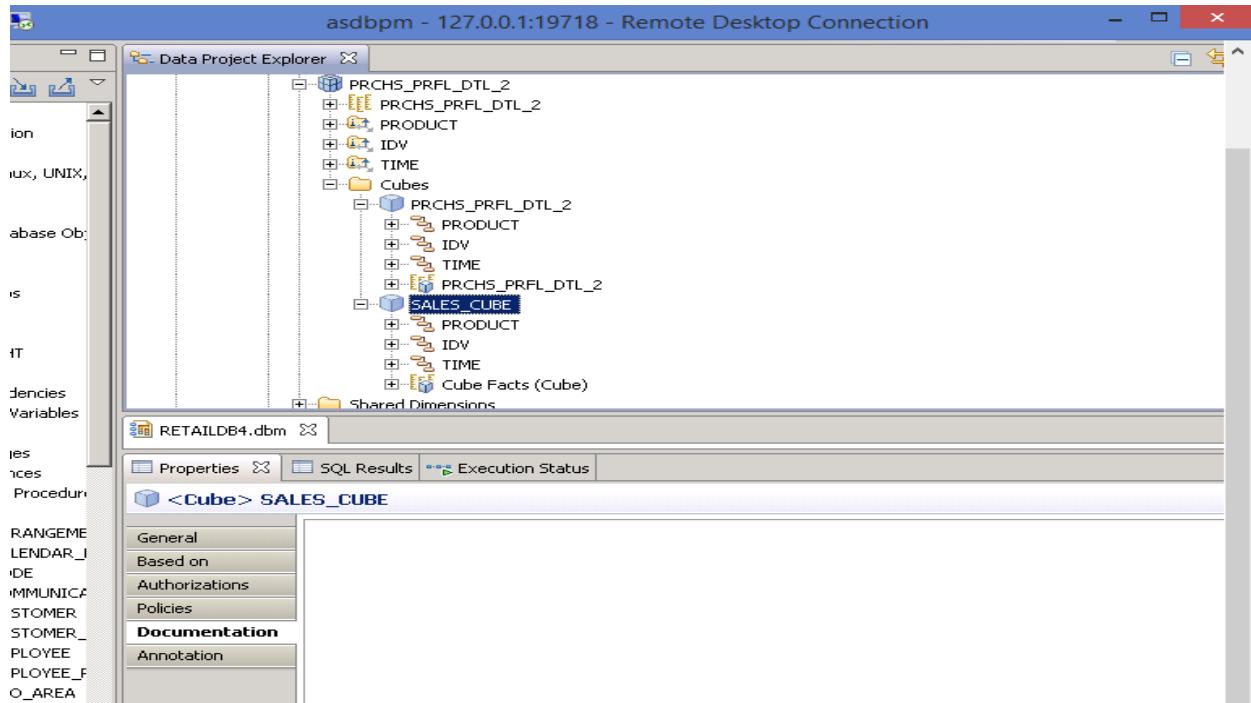
- b) Add level GENDER to the customer dimension IDV



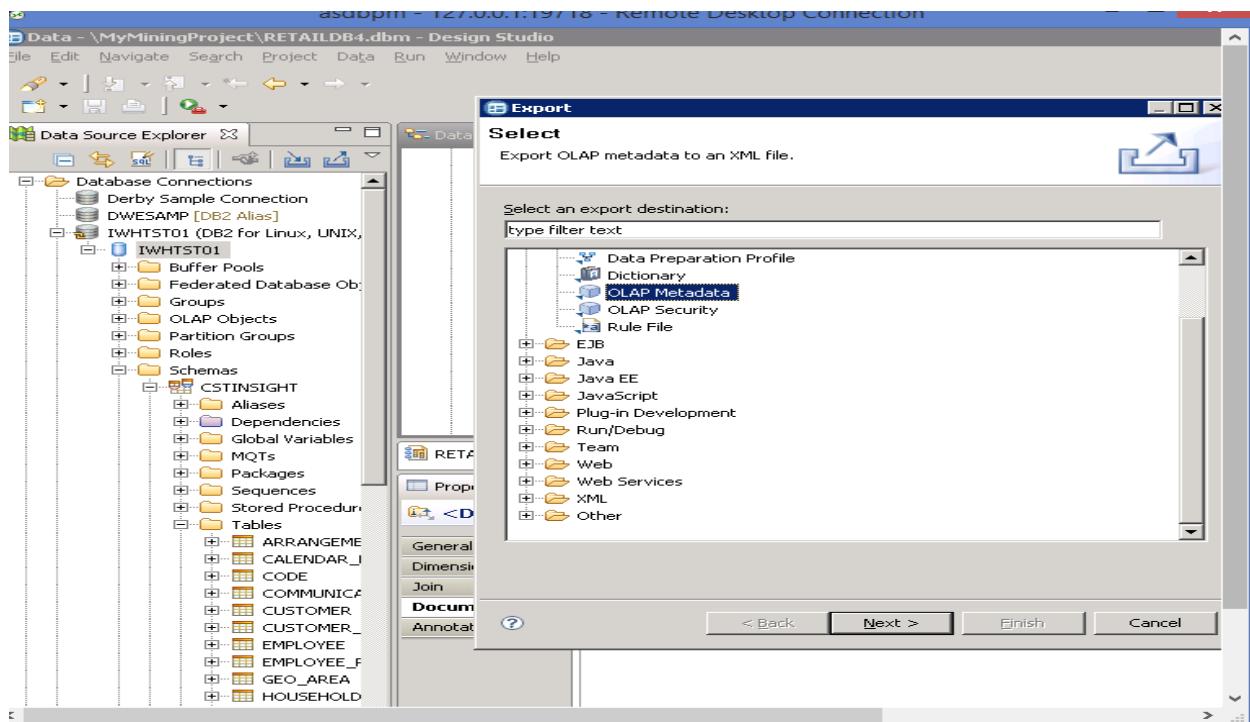
Now, create hierarchies for each dimension, Add your defined levels to Time, PRODUCT, idv dimension



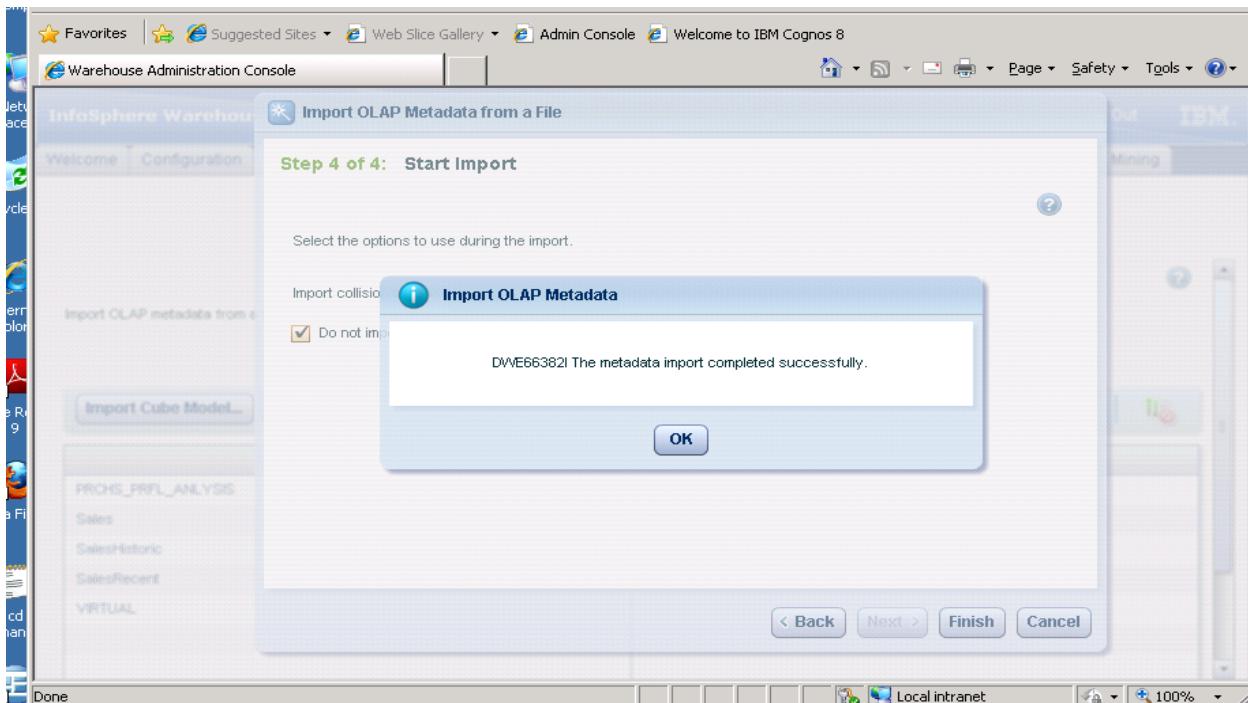
## 2.7 Add cube



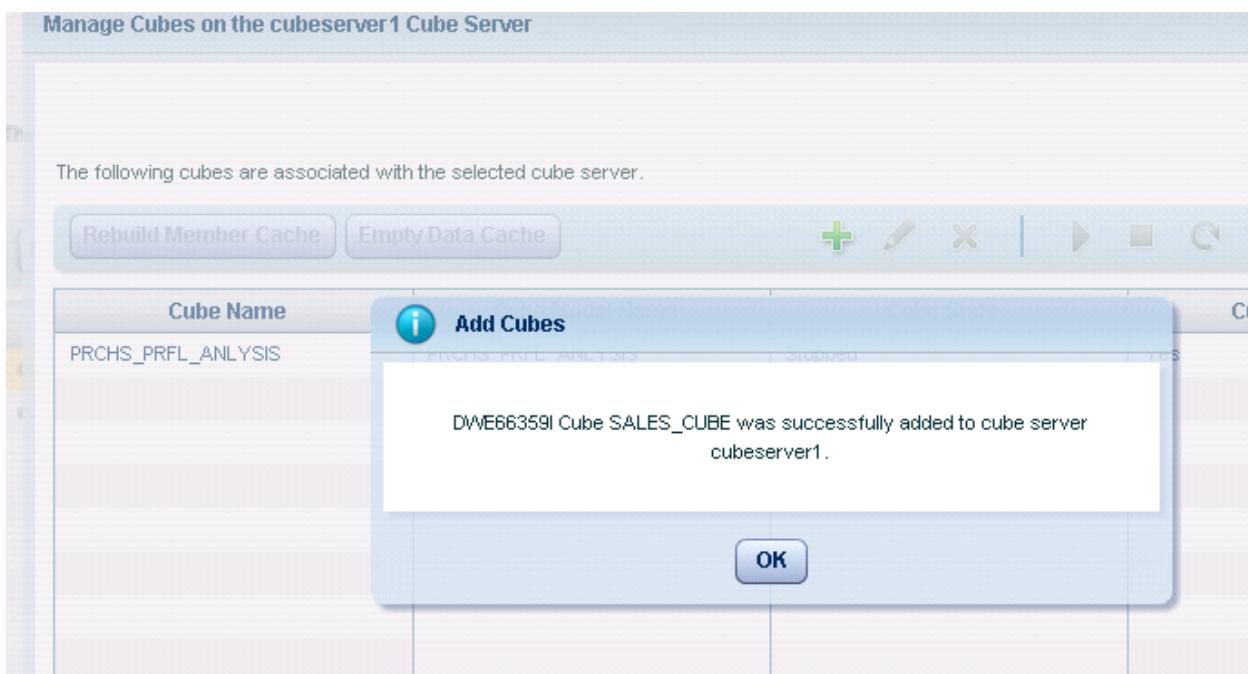
## 2.8 Export OLAP model

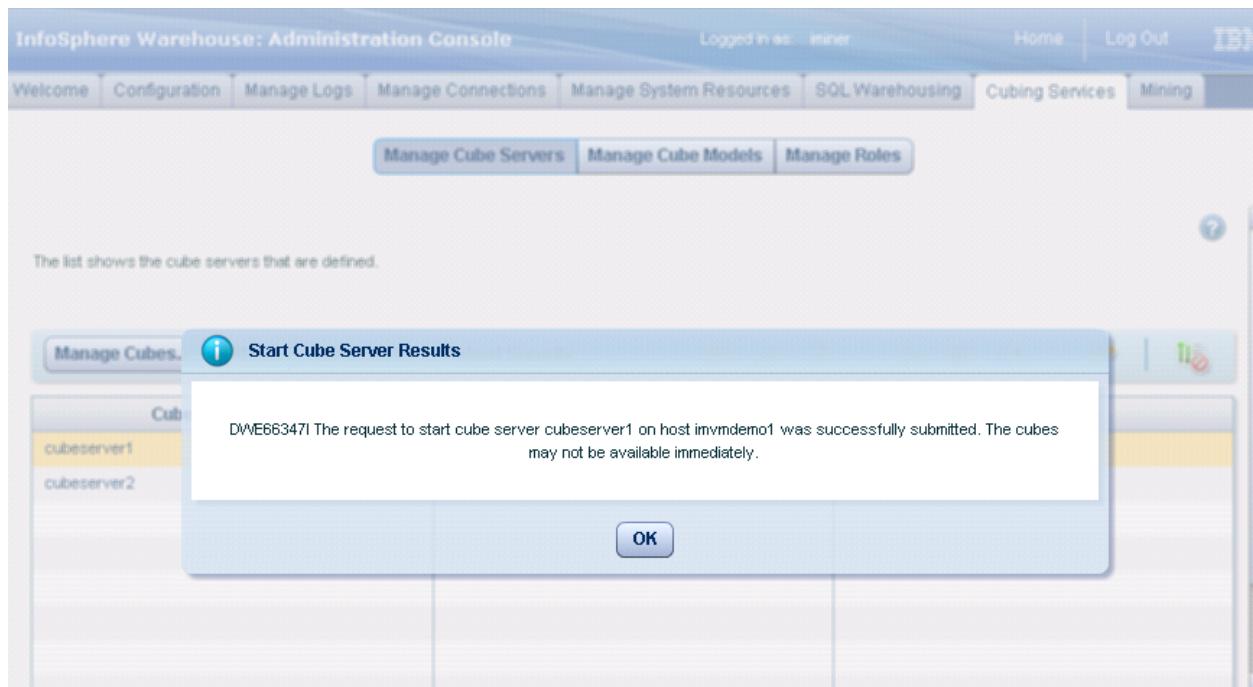


## 2.9 Deploy OLAP model to the Administration Console



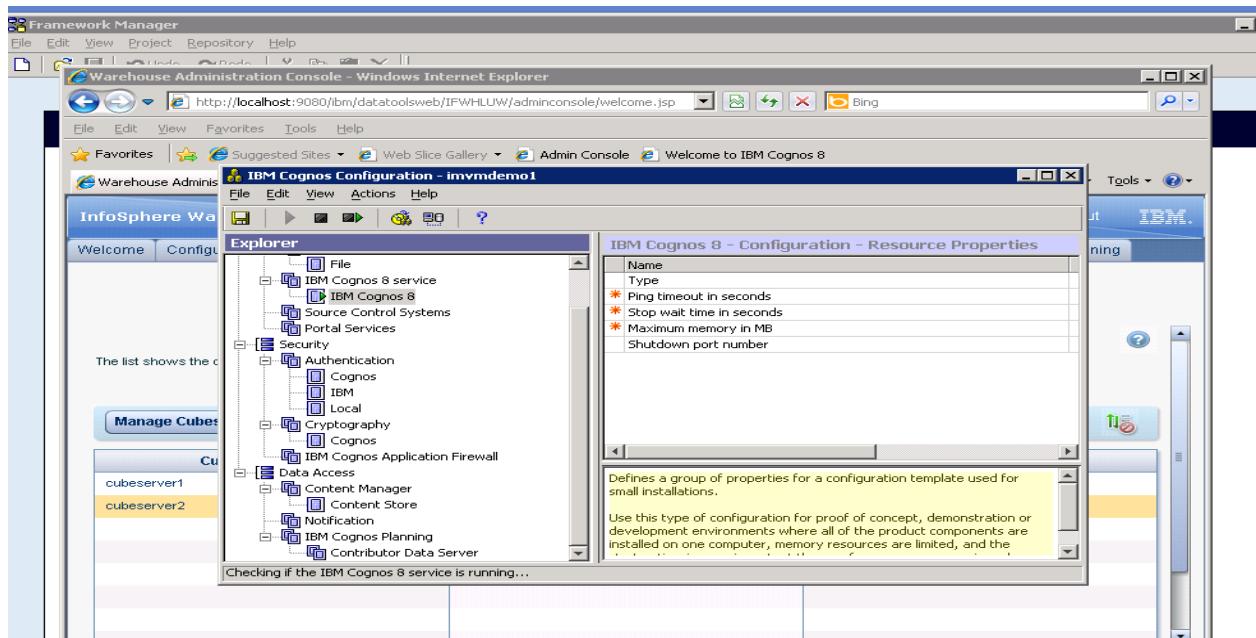
## 2.10 Run cube model within cube servers



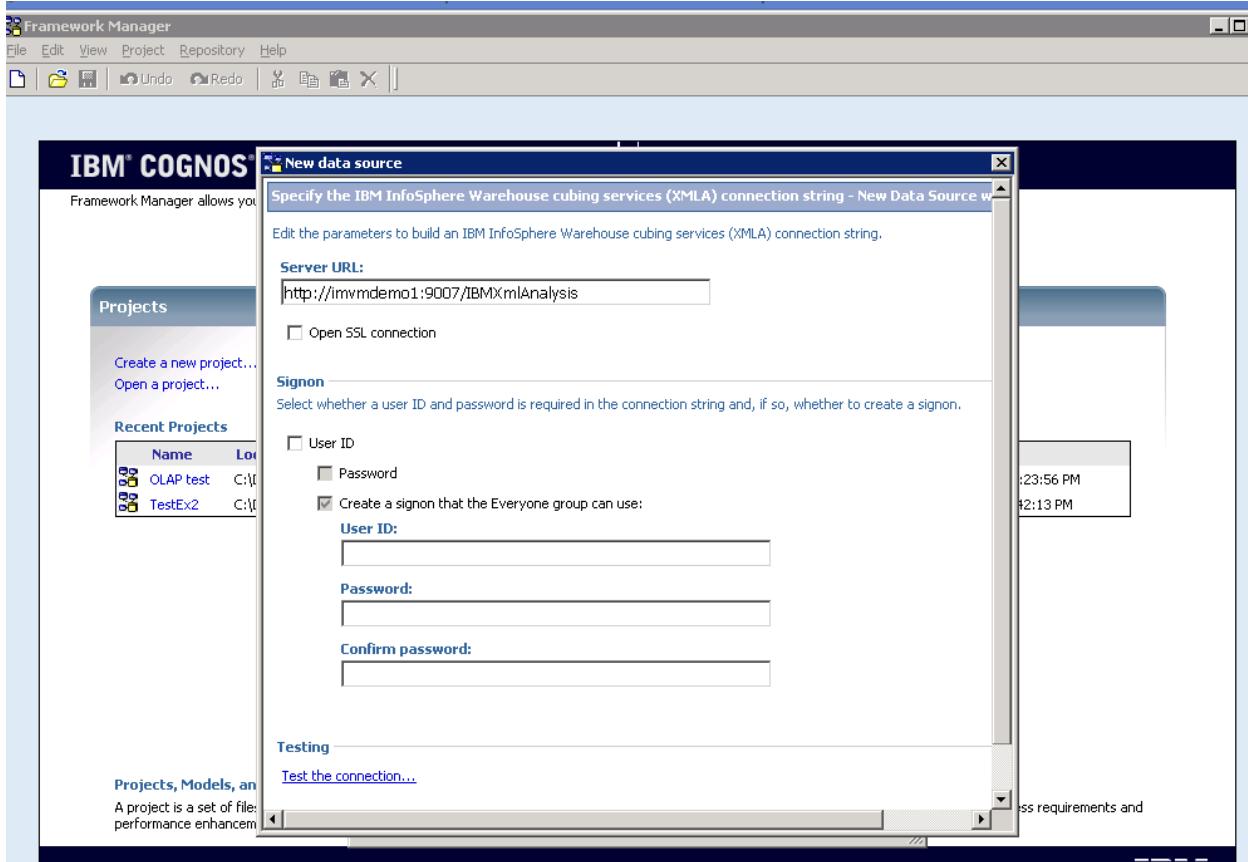


## 2.11 Cognos report using the OLAP cube

Start cognos server



## Cognos framework manager



## Publish wizard

asdbpm - 127.0.0.1:19718 - Remote Desktop Connection

New\* - Report Studio - Windows Internet Explorer

**Data**

- Filters...
- Suppress
- Sort
- Aggregate
- Calculate
- Insert Children
- Master Detail Relationships...
- Drill Behavior...**
- Default Data Formats...

Members here to create page layers

Context filter: Drop members

**Double click to edit text**

sure (y-axis):

Drop item here

Series:

- <#NUMBER\_OF\_ITEMS#>
- <#SALES\_AMOUNT#>
- Drop item here

(Default Legend Title)

Axis titles:

Categories (x-axis):

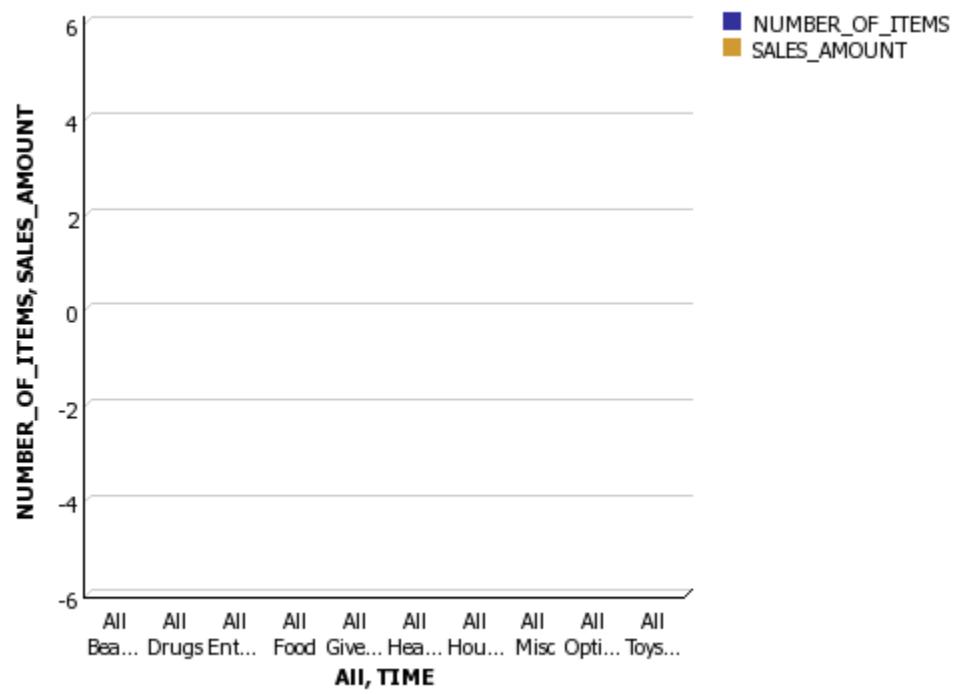
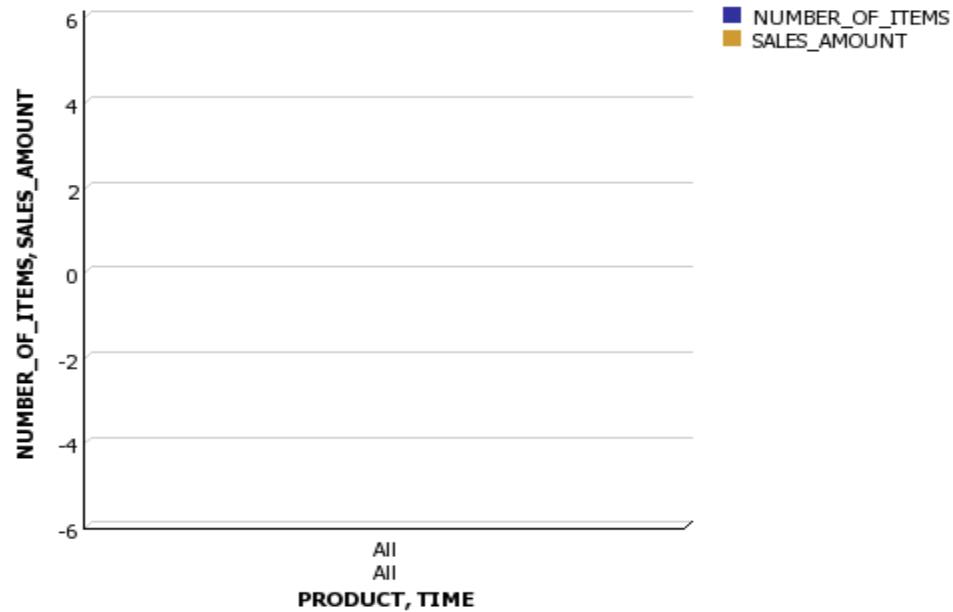
<#PRODUCT#> <#TIME#>

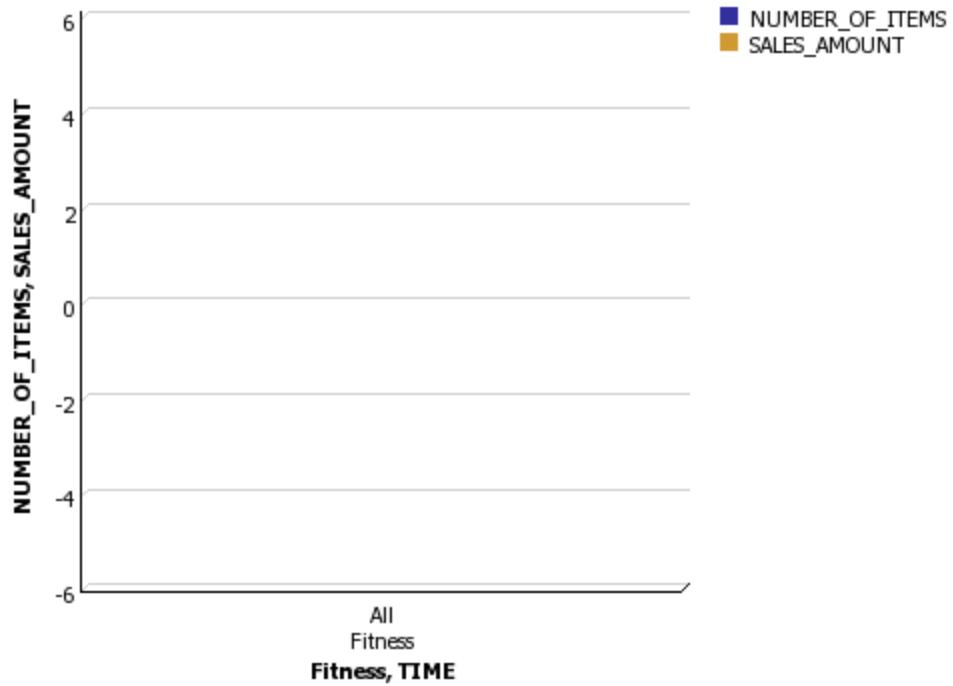
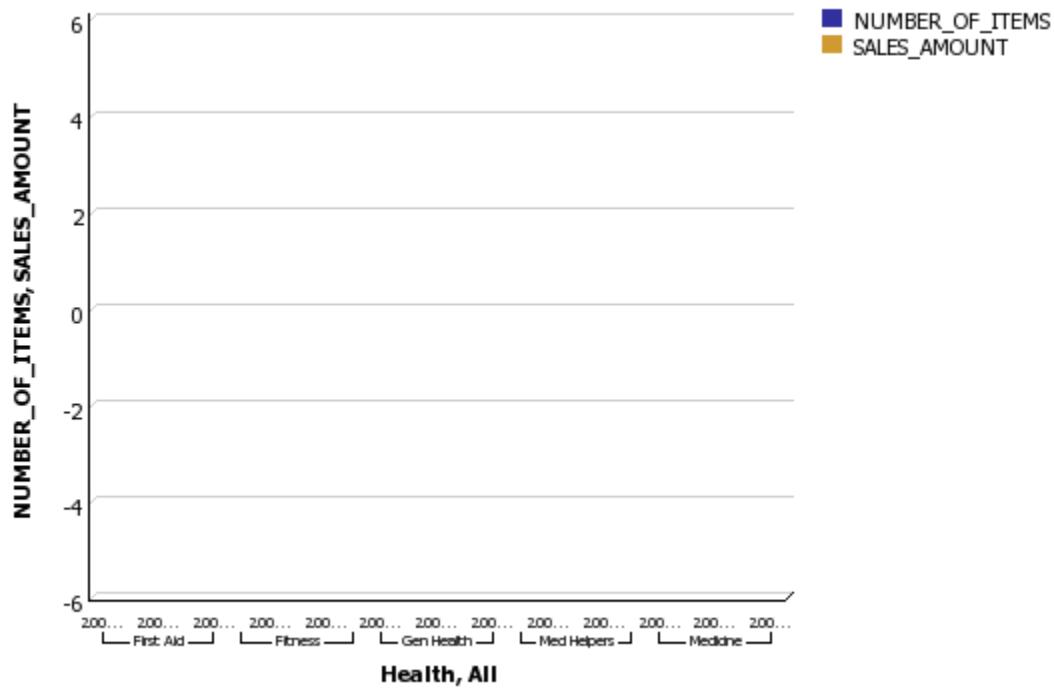
Properties - **Legend Title**

- Conditional**
- Conditional Styles
- Style Variable
- Data**
- Query: Query1
- Master Detail Relationships
- Properties
- General**
- Default Title: Yes
- Box**
- Box Type

The screenshot shows the Report Studio interface with a bar chart on the right. The chart has four bars per category, colored blue, orange, red, and yellow. The y-axis is labeled 'sure (y-axis)' with a scale from 0 to n. The x-axis categories are labeled 'abc' four times. The legend title is '(Default Legend Title)'. The properties panel on the left shows settings for conditional styles, data (query set to 'Query1'), general properties (default title set to 'Yes'), and a box type.

## RESULT:





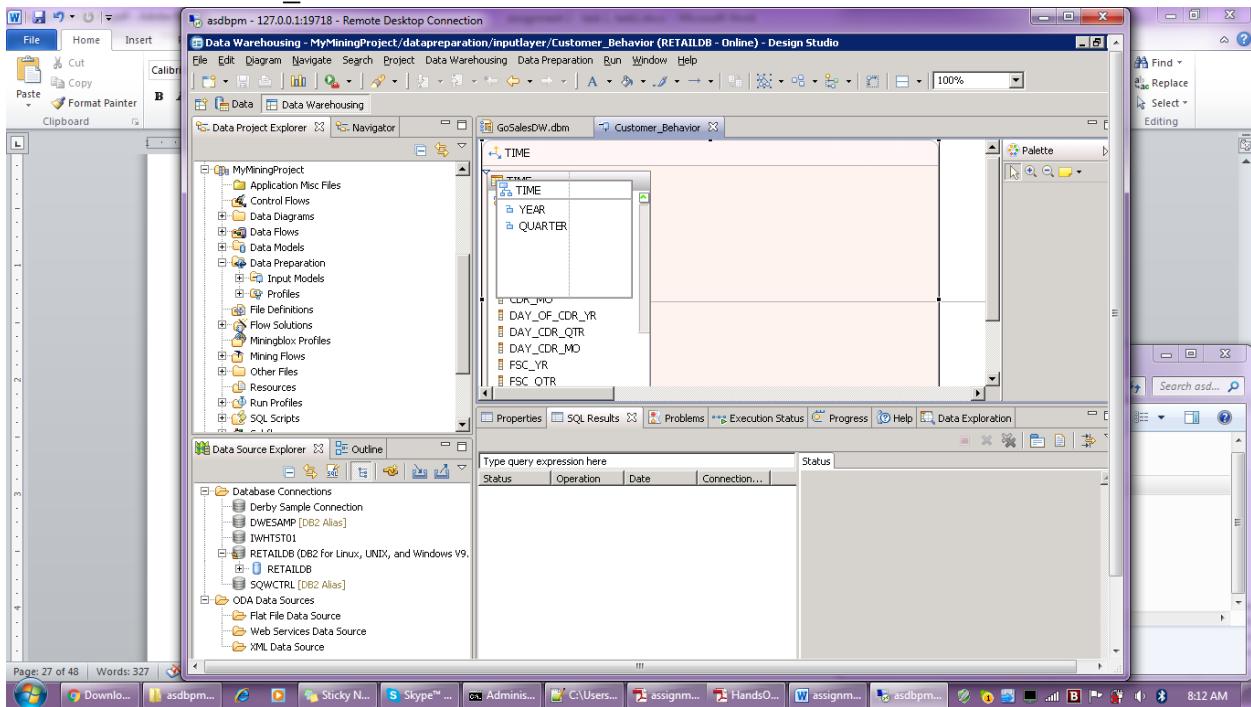
### **Task 3:**

**Descriptively prepare your data and perform customer segmentation using InfoSphere Warehouse:**

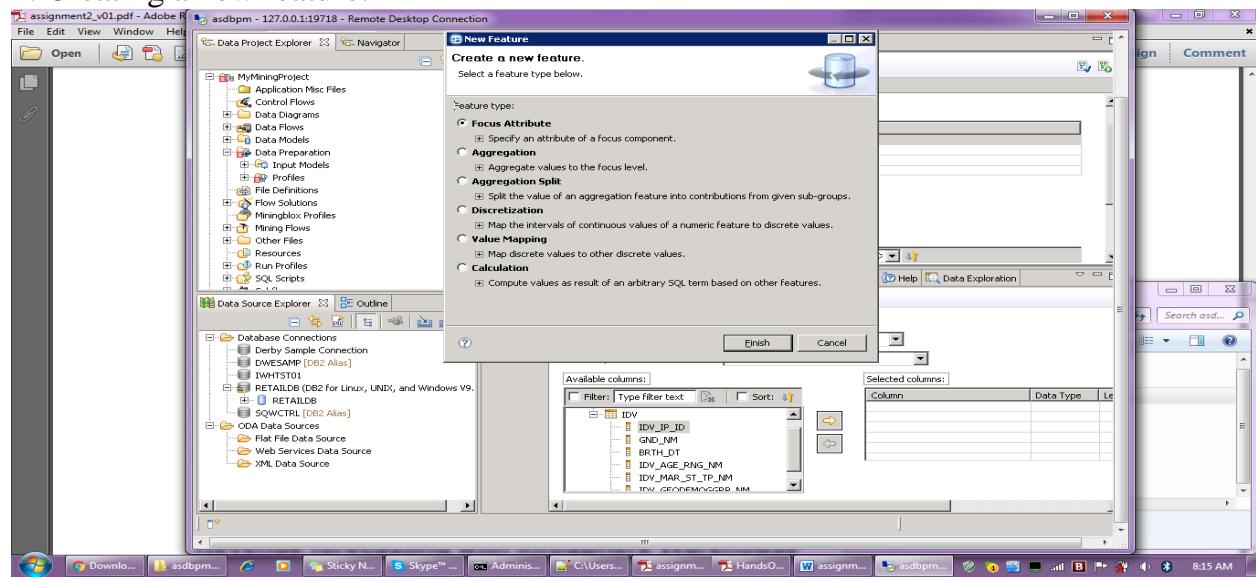
#### **3.1 Create and prepare data for customer segmentation**

1. Select the TEMP\_FACT\_EVAL cube from RETAILDB
2. Aggregate the Sales\_Amount and name it as Total Spent.
3. Do Aggregation split by Department and check Beauty, Food and Health.
4. Do Aggregation split by Sub Department.
5. Select Customer\_behavior and Mining Schema.

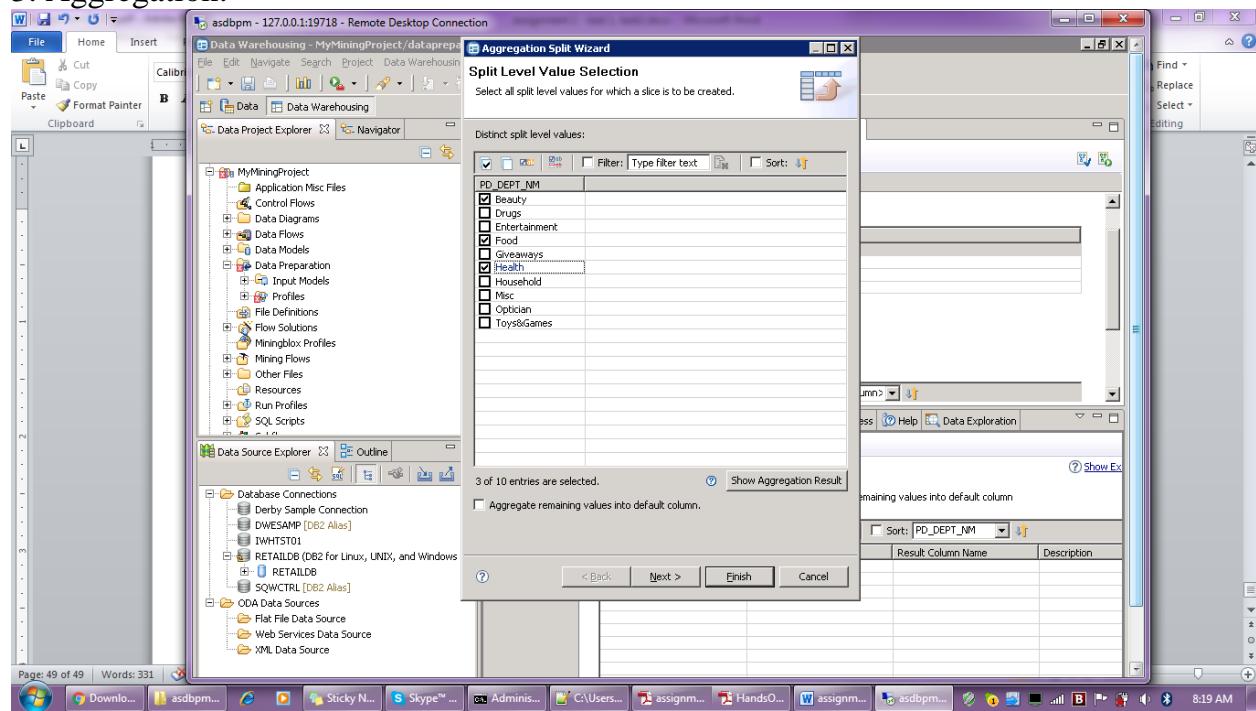
1. Created Customer\_Behaviour Data Model:



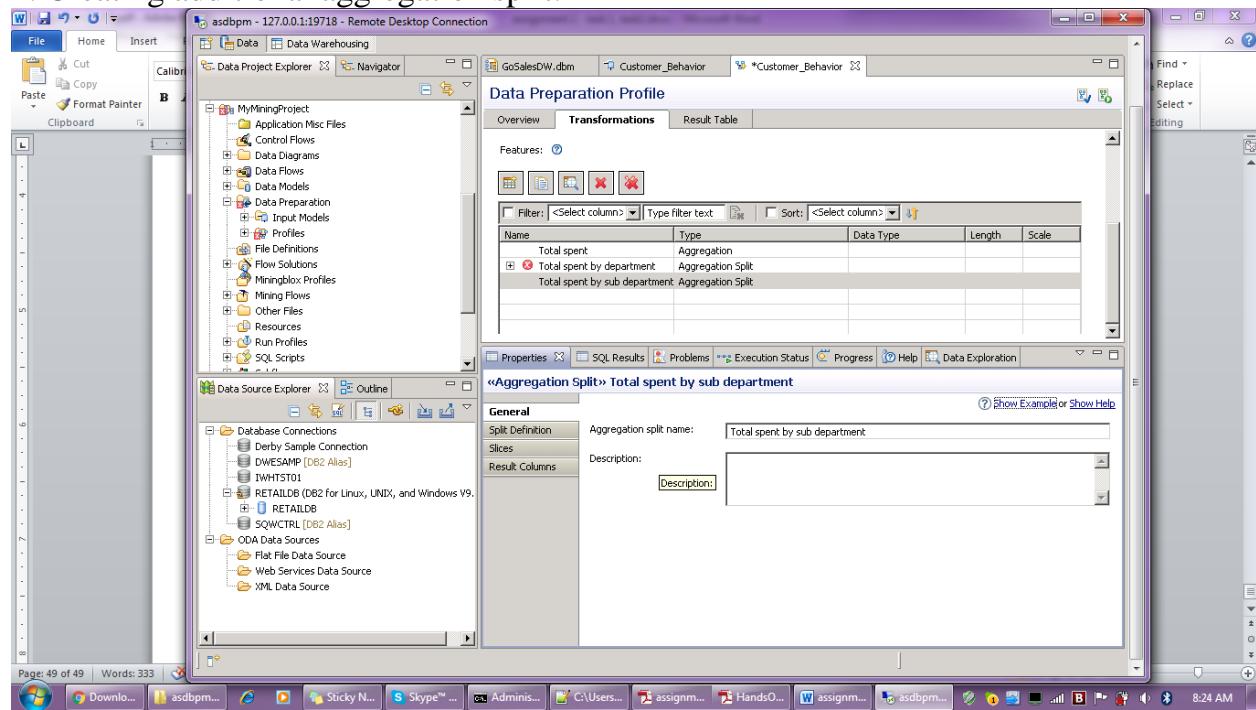
## 2. Creating a new feature:



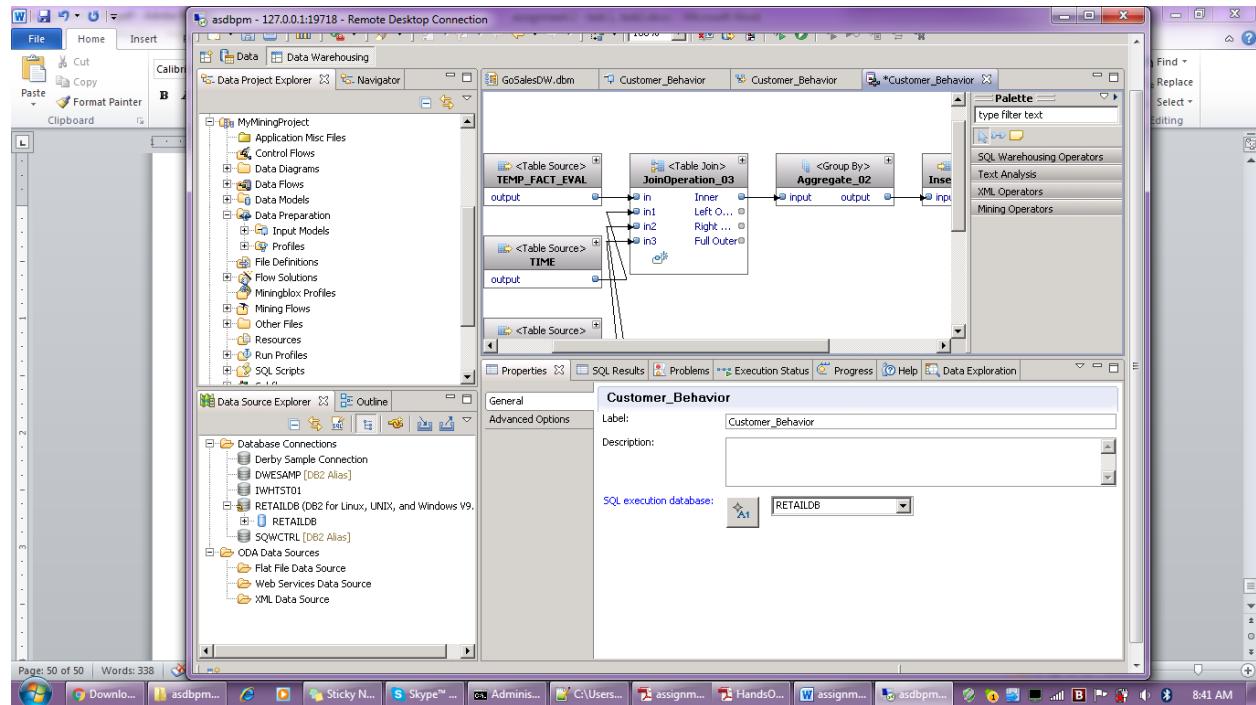
## 3. Aggregation:



#### 4. Creating additional aggregation split:



#### 5. New data flow:



## 6. Result:

Site	ID	IP_ID	TOTAL_SPENT	TOTAL_SPENT_ONBEAUTY	TOTAL_SPENT_ONFOOD	TOTAL_SPENT_ONHEALTH	TOTAL_SPENT_ONTRAVEL
1	219728	52.44	52.44	0.00	0.00	0.00	0.00
2	219732	304.74	289.20	5.69	9.85	0.00	0.00
3	219733	0.00	0.00	0.00	0.00	0.00	0.00
4	219736	3.46	0.00	3.46	0.00	0.00	0.00
5	219744	66.35	43.89	4.39	18.07	0.00	0.00
6	219741	13.07	13.07	0.00	0.00	0.00	0.00
7	219752	24.76	24.76	0.47	0.47	0.00	0.00
8	219770	19.73	21.73	0.37	0.00	0.00	0.00
9	219768	15.39	14.70	0.00	0.69	0.00	0.00
10	219790	17.30	14.53	2.77	0.00	0.00	0.00
11	219796	152.66	147.04	5.62	0.00	0.00	0.00
12	219806	26.43	13.94	0.00	5.91	0.00	0.00
13	219817	717.61	467.02	81.59	169.00	14.34	0.00
14	219827	424.16	304.04	26.31	71.44	7.77	0.00
15	219845	15.07	8.77	3.72	2.58	0.00	0.00
16	219849	29.37	26.89	0.00	2.48	0.00	0.00
17	219850	23.16	0.00	0.00	23.16	0.00	0.00
18	219852	250.52	250.05	0.00	17.02	0.00	0.00
19	219857	20.85	0.00	0.00	20.85	0.00	0.00
20	219862	124.09	87.73	2.10	34.26	0.00	0.00
21	219864	15.18	14.14	1.04	0.00	0.00	0.00
22	219868	74.78	73.82	0.96	0.00	0.00	0.00
23	219875	46.83	12.71	0.00	32.14	0.00	0.00
24	219876	19.06	10.44	2.71	5.91	0.00	0.00
25	219887	69.47	63.12	0.00	6.35	0.00	0.00
26	219904	6.68	5.10	1.58	0.00	0.00	0.00
27	219913	533.77	372.53	2.47	104.86	0.00	0.00
28	219914	23.86	23.22	0.00	0.64	0.00	0.00
29	219923	77.0k	77.0k	0.00	0.00	0.00	0.00

### 3.2 Create a customer Segmentation model:

1. Create Customer Segmentation as New Flow Solution with Marts.IDV as table.
2. Select IDV\_IP\_ID, GND\_NM and IDV\_AGE\_RNG\_NM to filter out the customers that contain an undefined gender value.
3. Select MINING.CUSTOMER\_BEHAVIOR and add IDV\_IP\_ID.
4. Create table CLUSTER\_INFO with MINING as schema.
5. Use CUST\_IN\_SEGMENT table and create the mining flow.

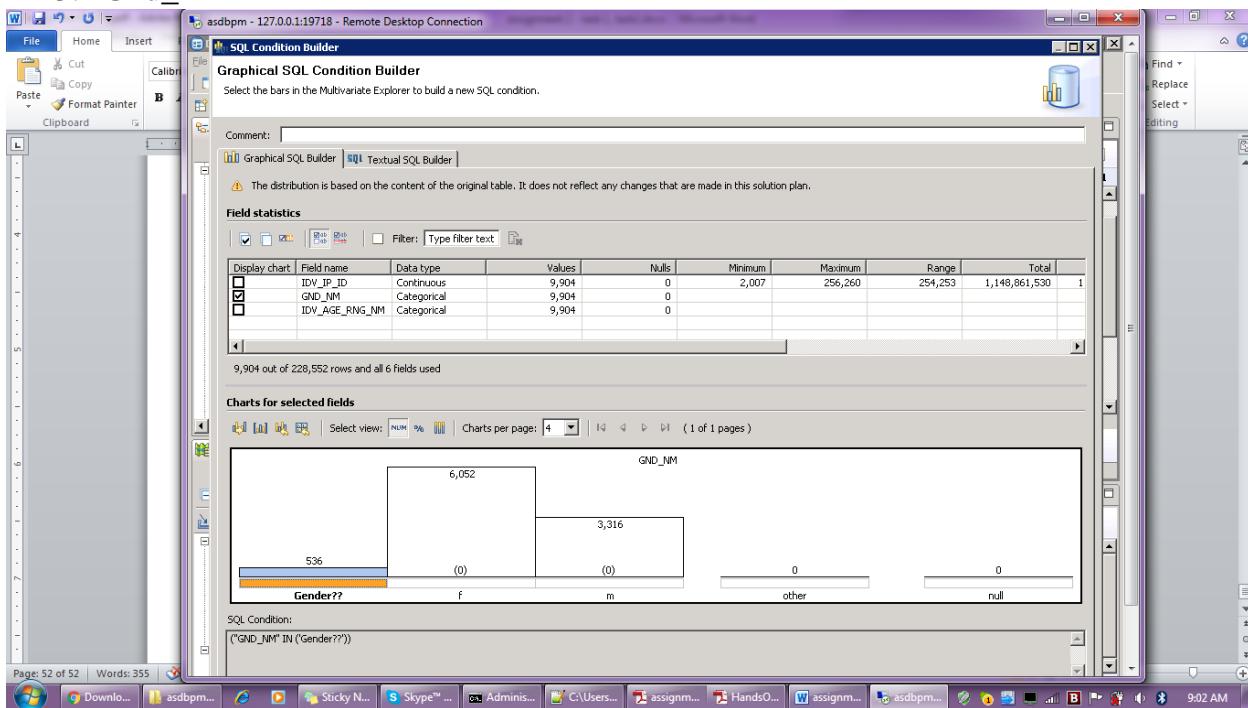
## 1. Customer segmentation:

The screenshot shows the Data Mining Project interface with the title bar "asdbpm - 127.0.0.1:19718 - Remote Desktop Connection". The main window displays the "Select demographic data" step, which is part of a flow named "2. Decision: Demographic". The left pane shows a tree view of the project structure under "MyMiningProject". The right pane shows a "Source table selection" dialog with a list of tables from the "MARTS" schema, including IDV, PRCHS\_PRFL\_DTL, PRODUCT, STORE, TEMP\_FACT\_EVAL, and TIME. A table preview window below shows sample data for the IDV table. The status bar at the bottom indicates "Page: 51 of 51" and "Words: 348".

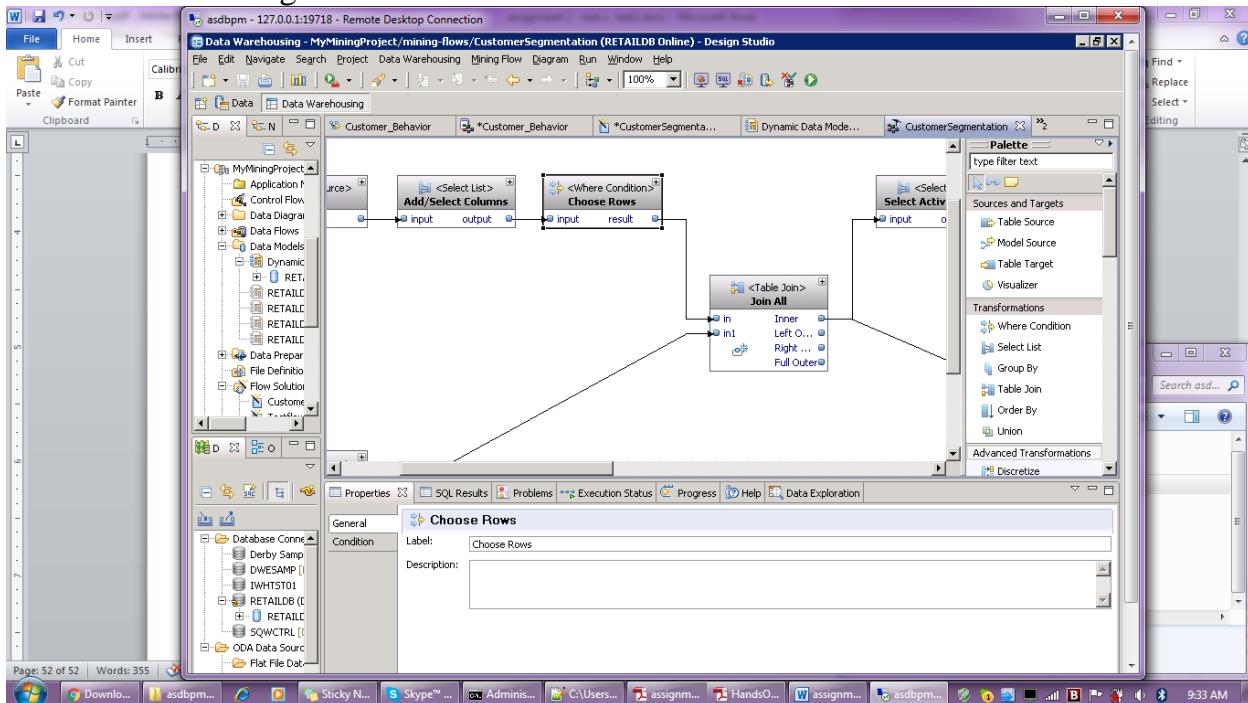
## 2. Field statistics:

The screenshot shows the Data Mining Project interface with the title bar "asdbpm - 127.0.0.1:19718 - Remote Desktop Connection". The main window displays the "Multivariate Distribution" step, which is part of a flow named "2. Decision: Demographic". The left pane shows a tree view of the project structure under "MyMiningProject". The right pane shows a "Field statistics" dialog with a table of field information and two charts below it. The charts are titled "BRTH\_DT" and "IDV\_AGE RNG\_NM". The status bar at the bottom indicates "Page: 51 of 51" and "Words: 351".

### 3. Gnd\_nm



### 4. Customer Segmentation Model:

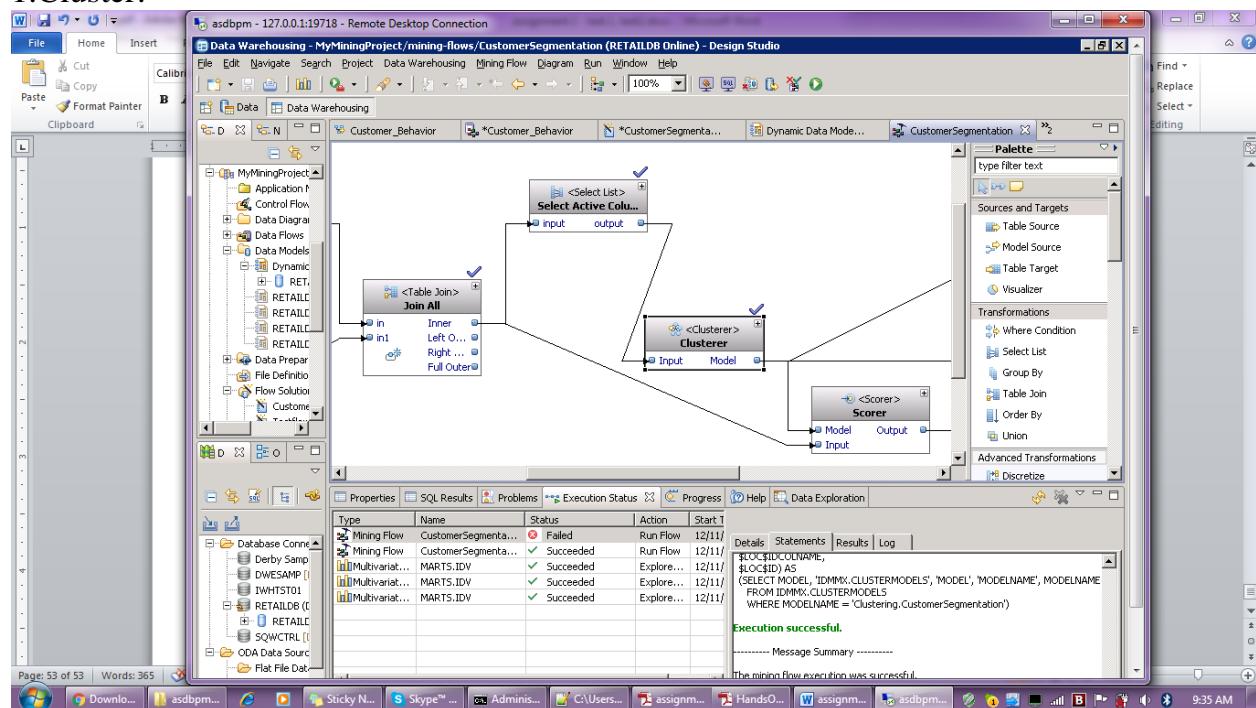


### 3.2 Analyse and enhance customer segments:

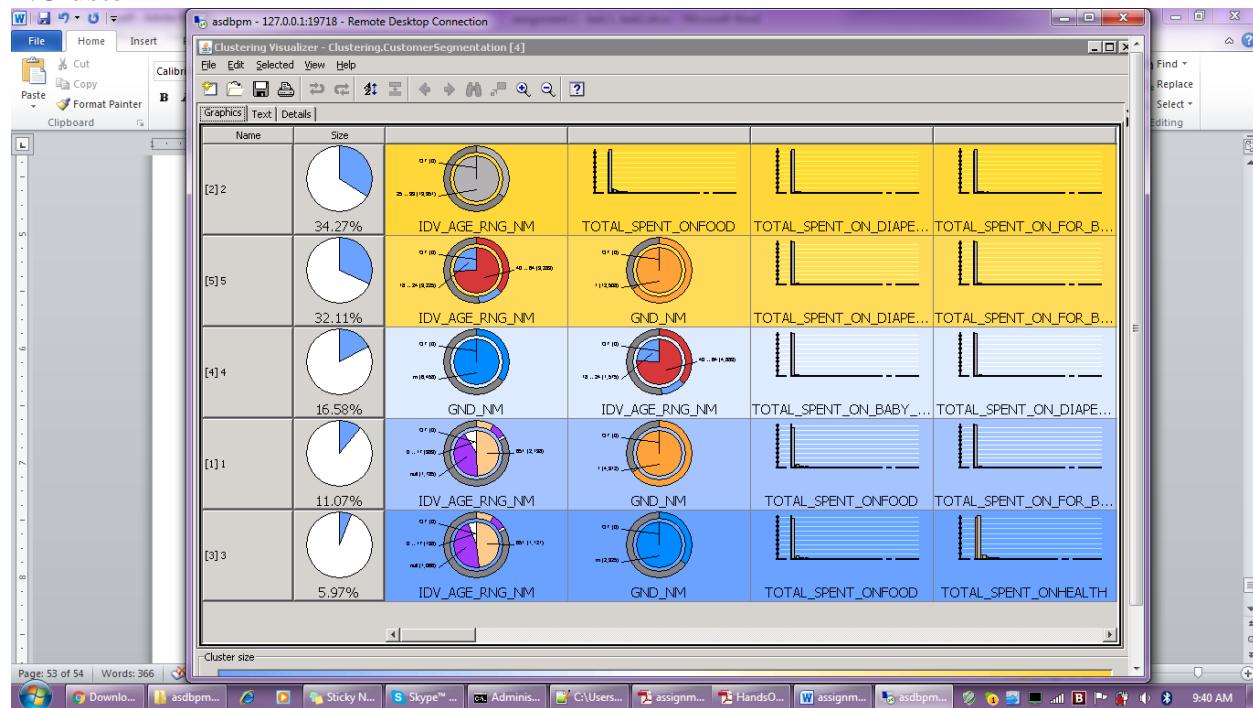
1. Right click the customer and click “Run to this step”.
2. Select Clusterer node and switch to properties tab.

3. double-click TOTAL\_SPENT\_ON\_BATHROOM\_BEAUTY, you see that overall, most customers spend between 0 and 2.50 Euros on bathroom beauty items.
4. To see the cluster assignment for each customer, go to the mining flow, right-click the CUST\_IN\_SEGMENT table target within the mining flow and select “Sample Contents.”
5. Value close to 1 means that the assigned cluster is the only cluster where this record fits in well. A value close to 0.5 means that there are other clusters this record might fit in well.

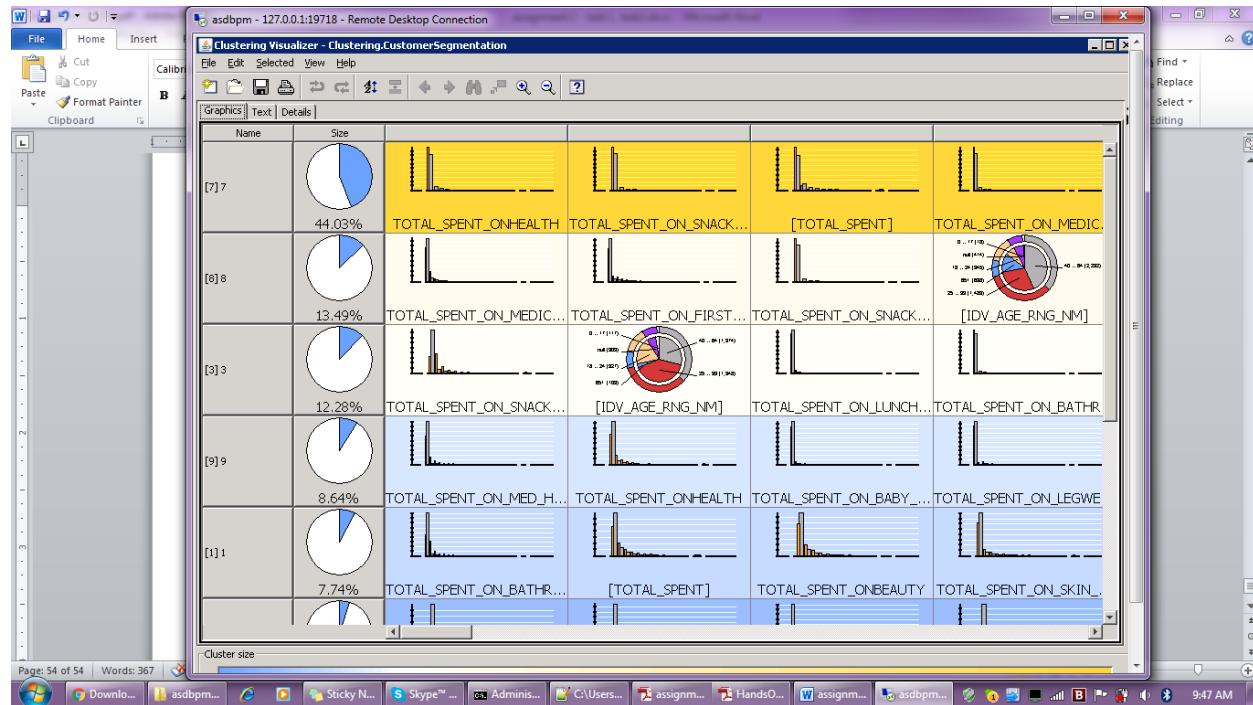
## 1. Cluster:



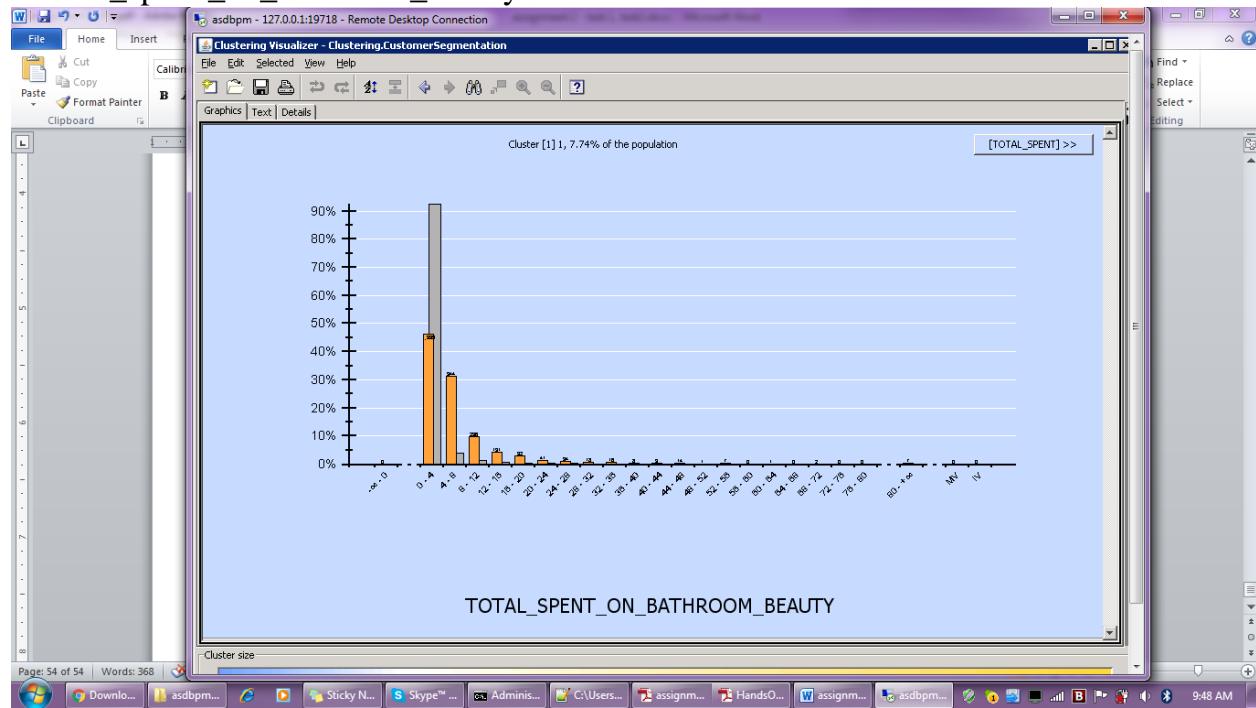
## 2.Cluster



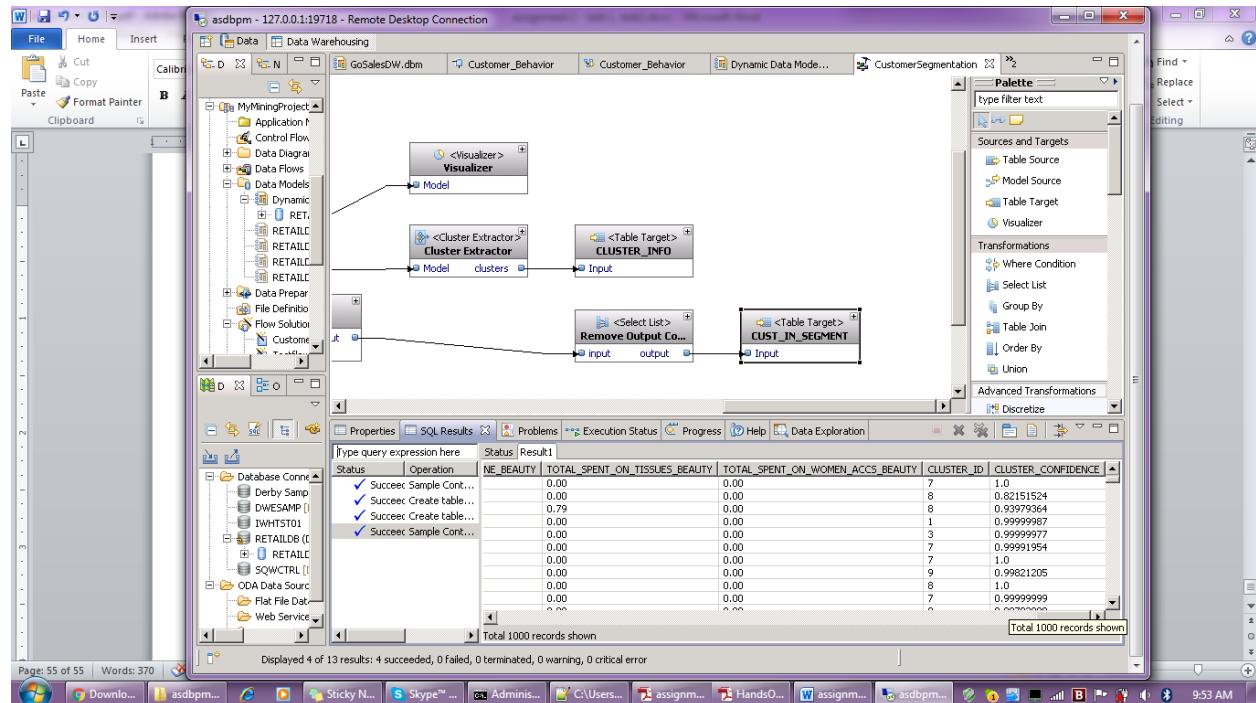
## 3.Cluster1



#### 4.Total\_spend\_on\_bathroom\_beauty



#### 4 Cluster Confidence:

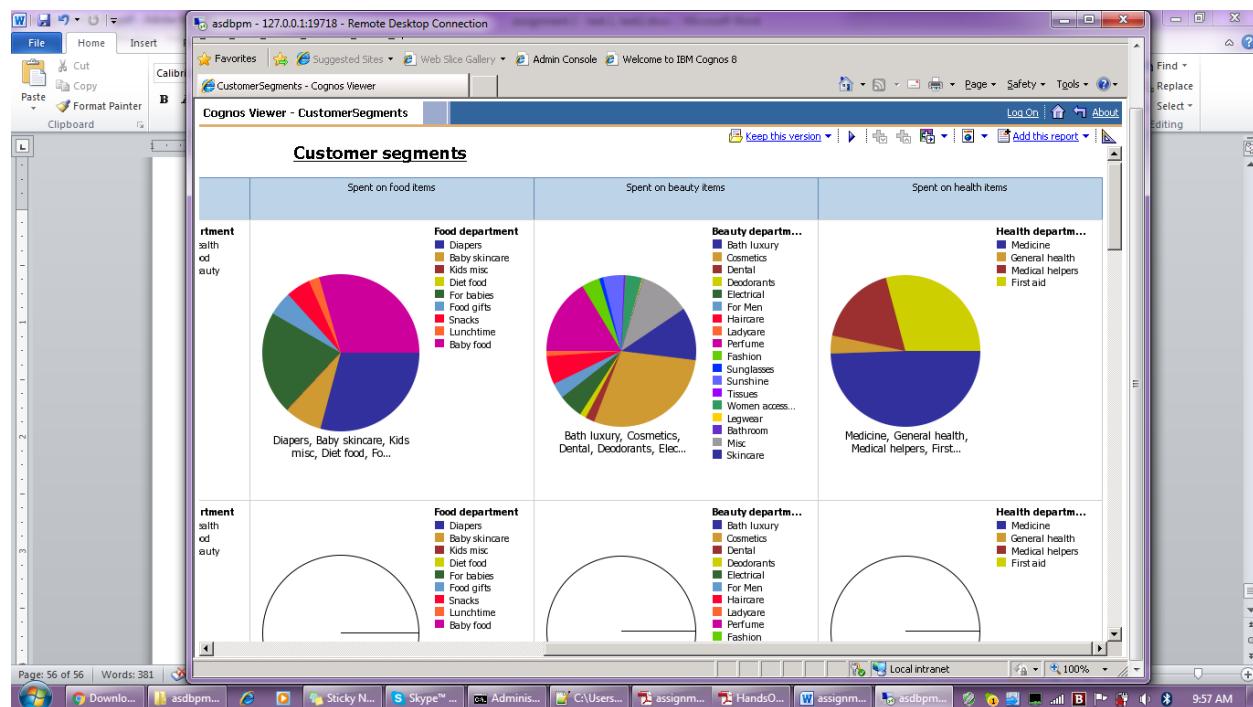
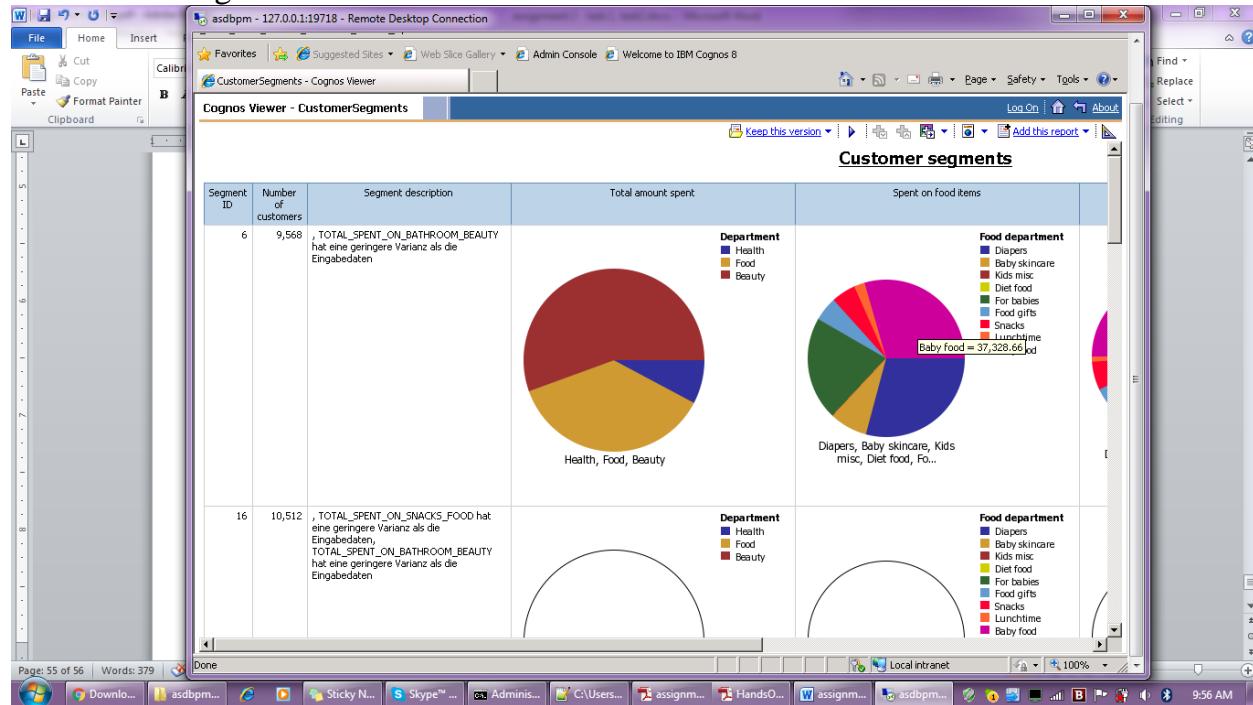


#### 3.4 Show customer segments within Cognos:

1. Open IBM Cognos 8 and select IBM Cognos connection.
2. Select Customer Segmentation.

### 3. Verify the final report with the HandsOnLab.

#### 1. Customer Segments:



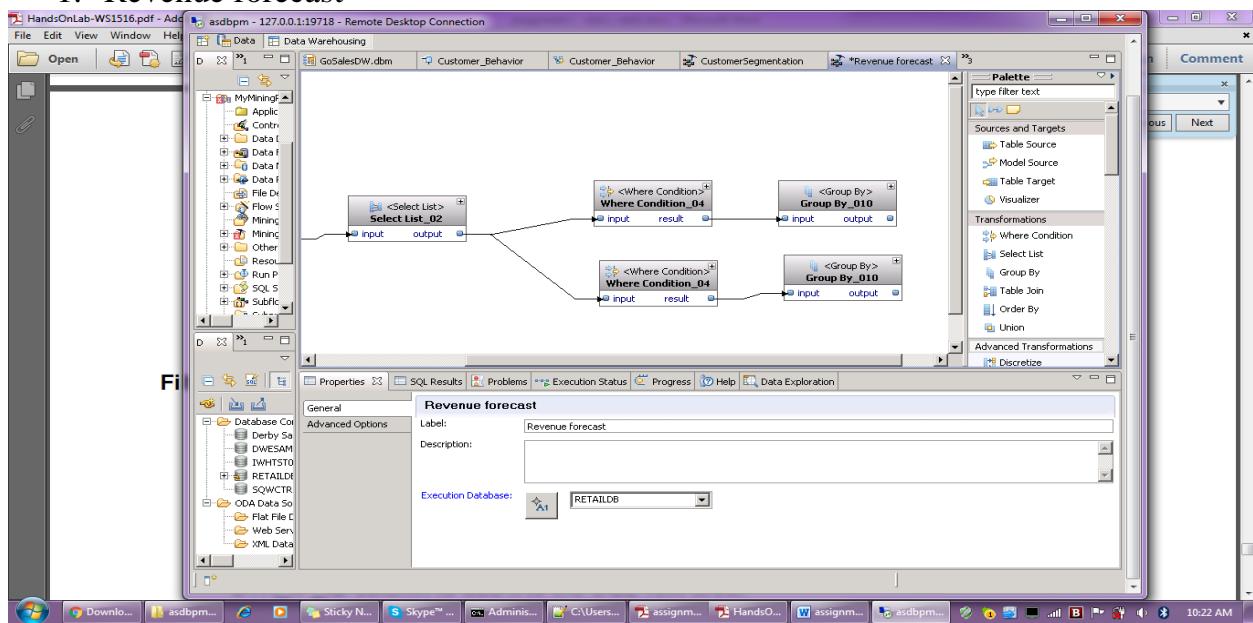
## Task 4:

### Revenue forecasting with InfoSphere Warehouse:

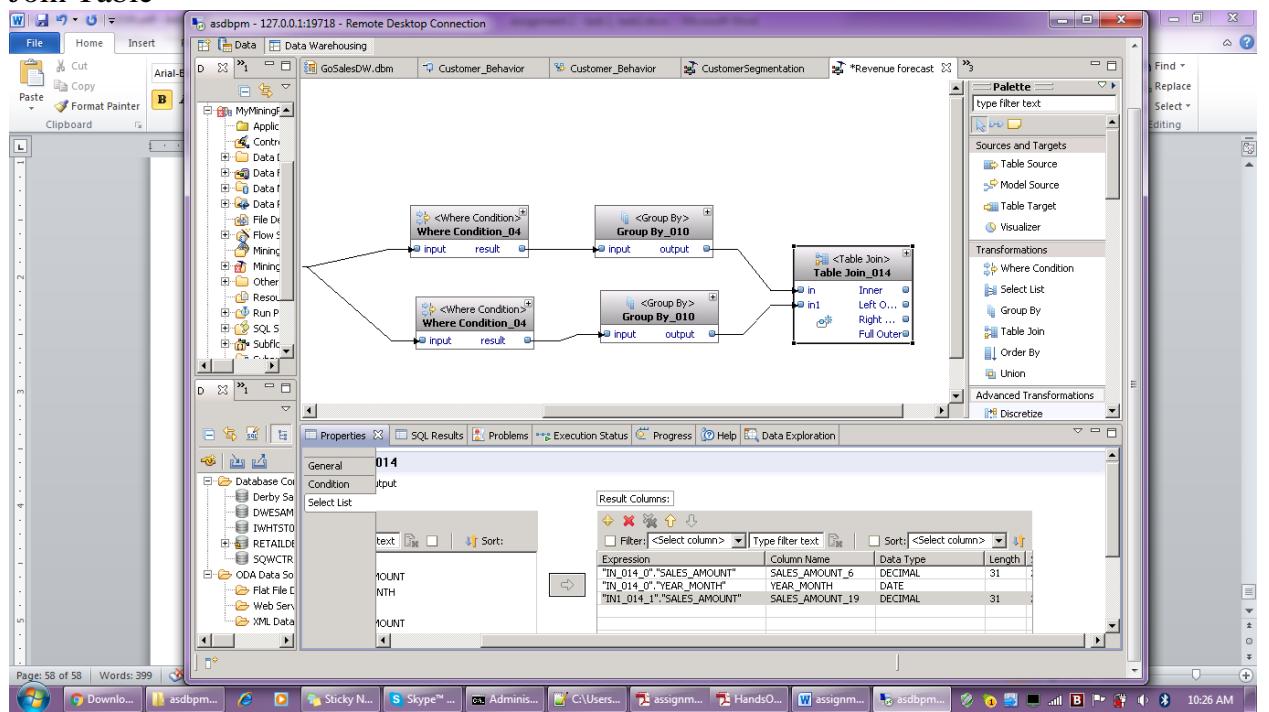
#### 4.1 Create a new mining flow in project “My mining project”

1. Create a new mining project with RETAILDB as source and name it as Revenue forecast.
2. Select the table PRCHS\_PRFL\_DTL.
3. Except Store id STR\_IP\_ID, the DATE and the revenue SALES\_AMOUNT remove the other columns.
4. Add a column YEAR\_MONTH with expression “date(year(date) || '-' || month(date) || '-01').”
5. Create two Where condition with STR\_IP\_ID =6 & 19.
6. Add a Groupby and aggregate by YEAR\_MONTH and sum the Sales\_amount for both the where clause.
7. Use a table join to Join the two tables.
8. Use a where condition YEAR\_MONTH < '2009-01-01'.
9. Add a time series operator
10. Add Year\_month to Time column and SALES\_AMOUNT\_6 and SALES\_AMOUNT\_19 to the value column.

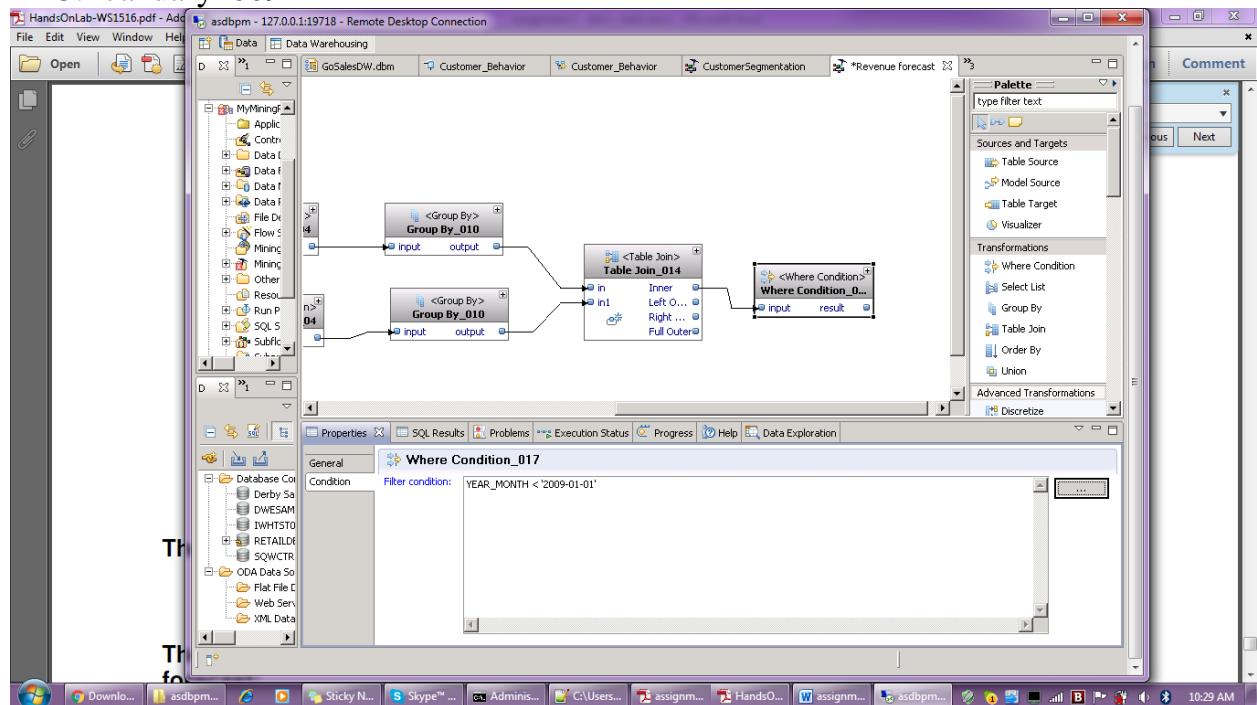
##### 1. Revenue forecast



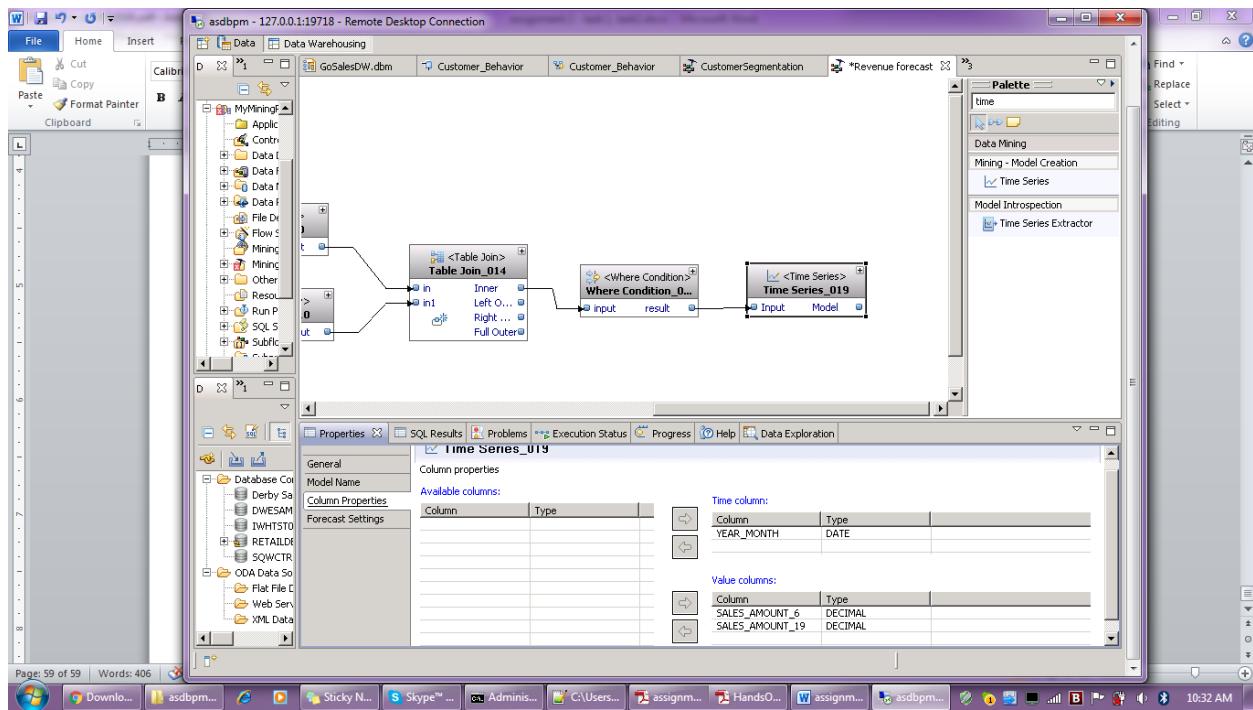
## 2. Join Table



## 3. January2009

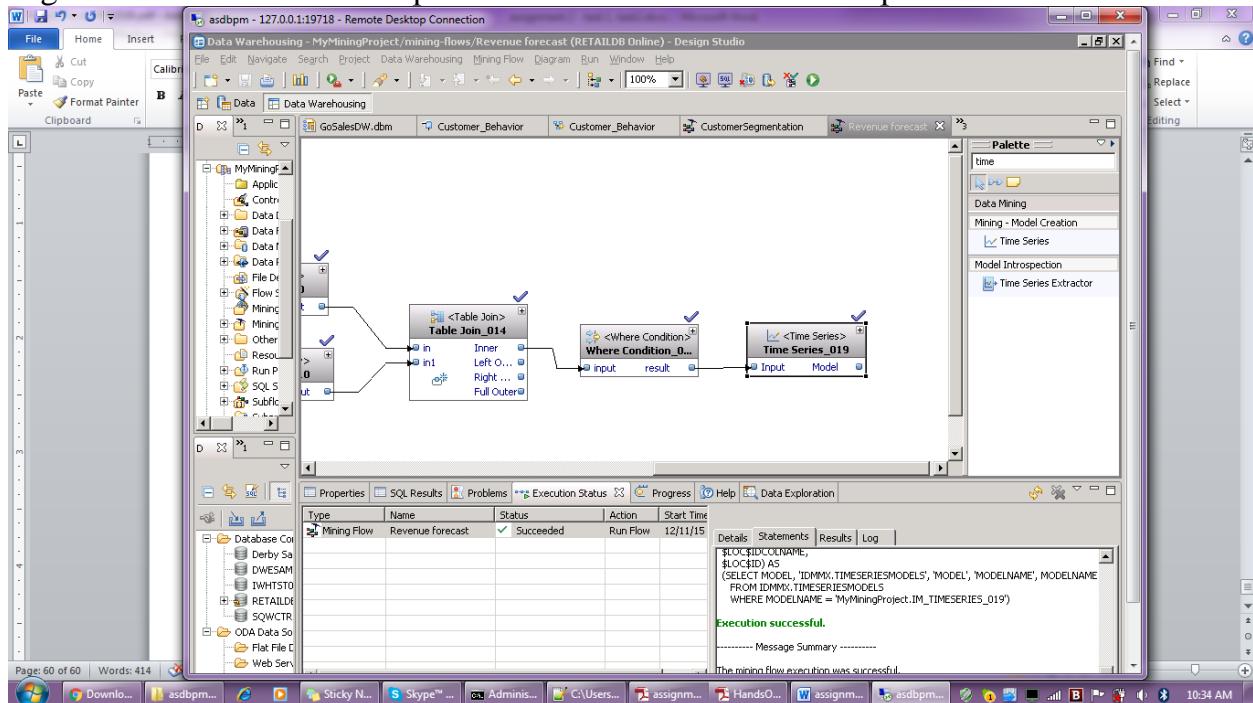


## 4. Time series operator:

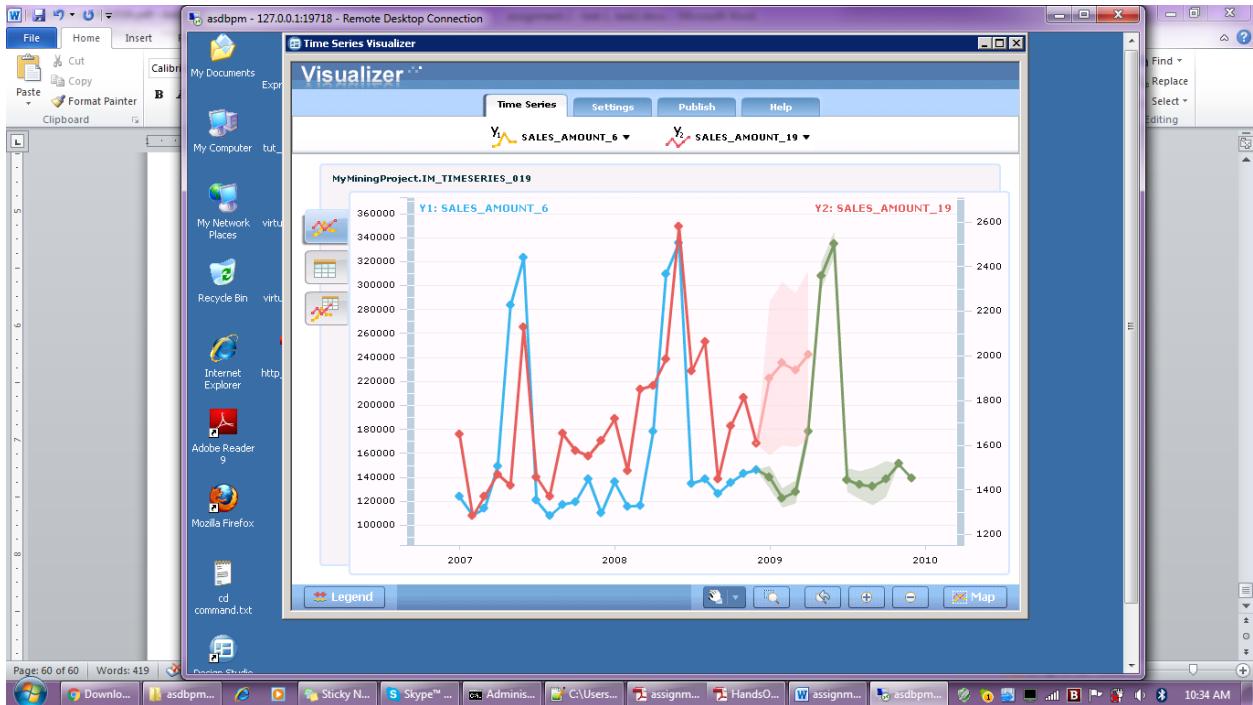


## 4.2 Execute the forecasting flow:

Right click the Time Series operator and select “Run to this step”.



**4.3 Analyse the forecasting results:**  
Right click Time Series operator and select “Open Model”



### **Task 5:**

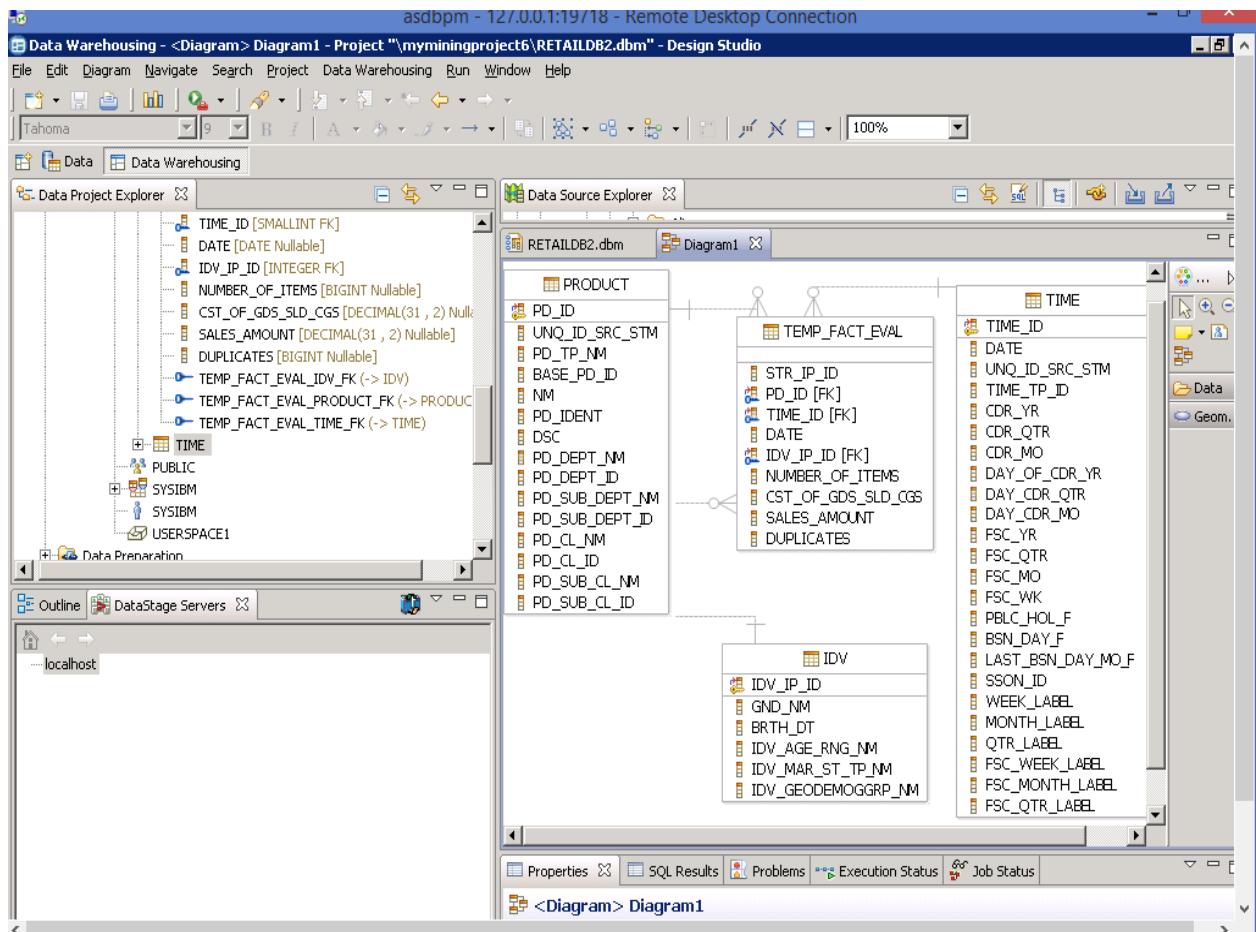
(Add.task1)

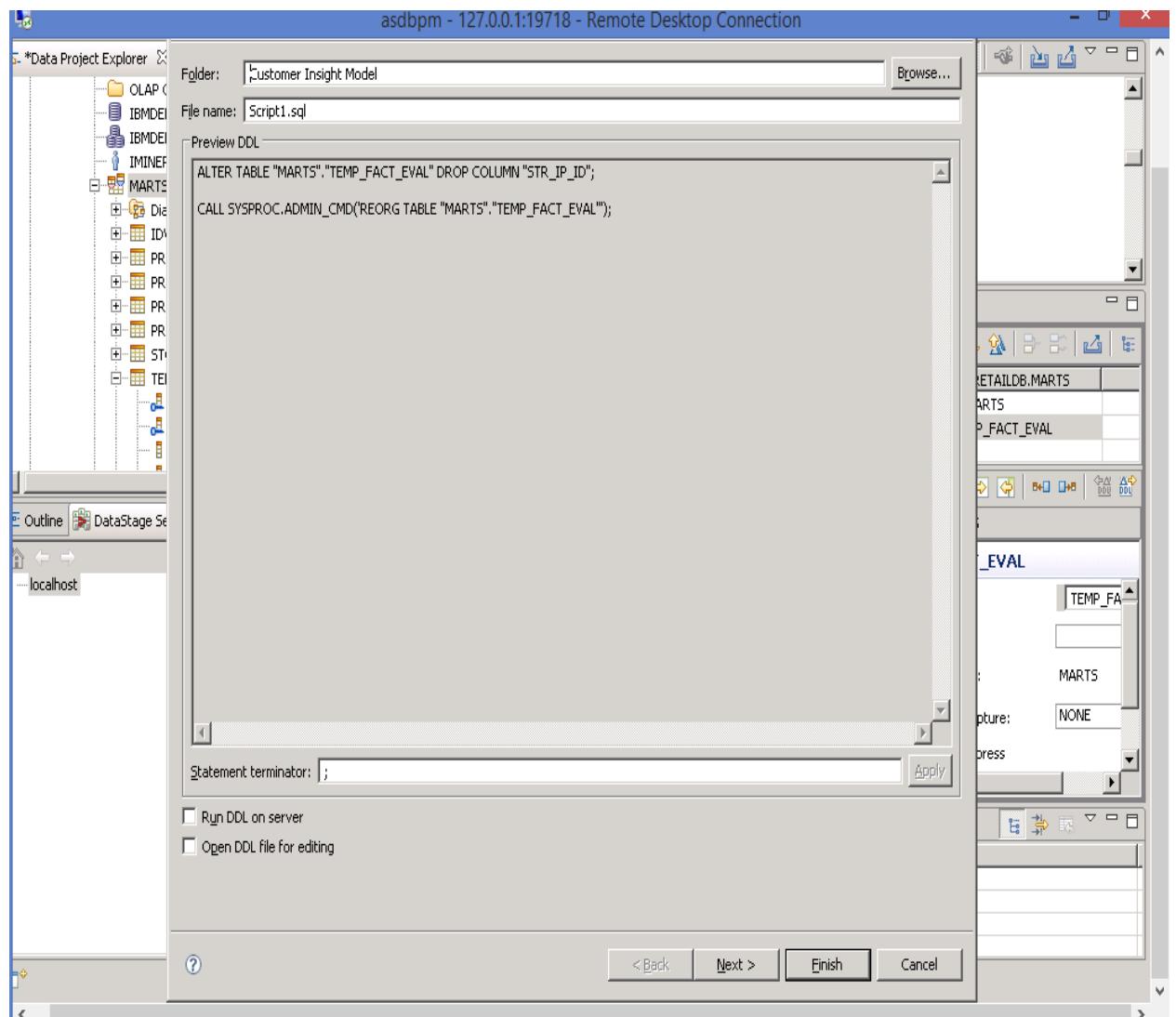
Steps:

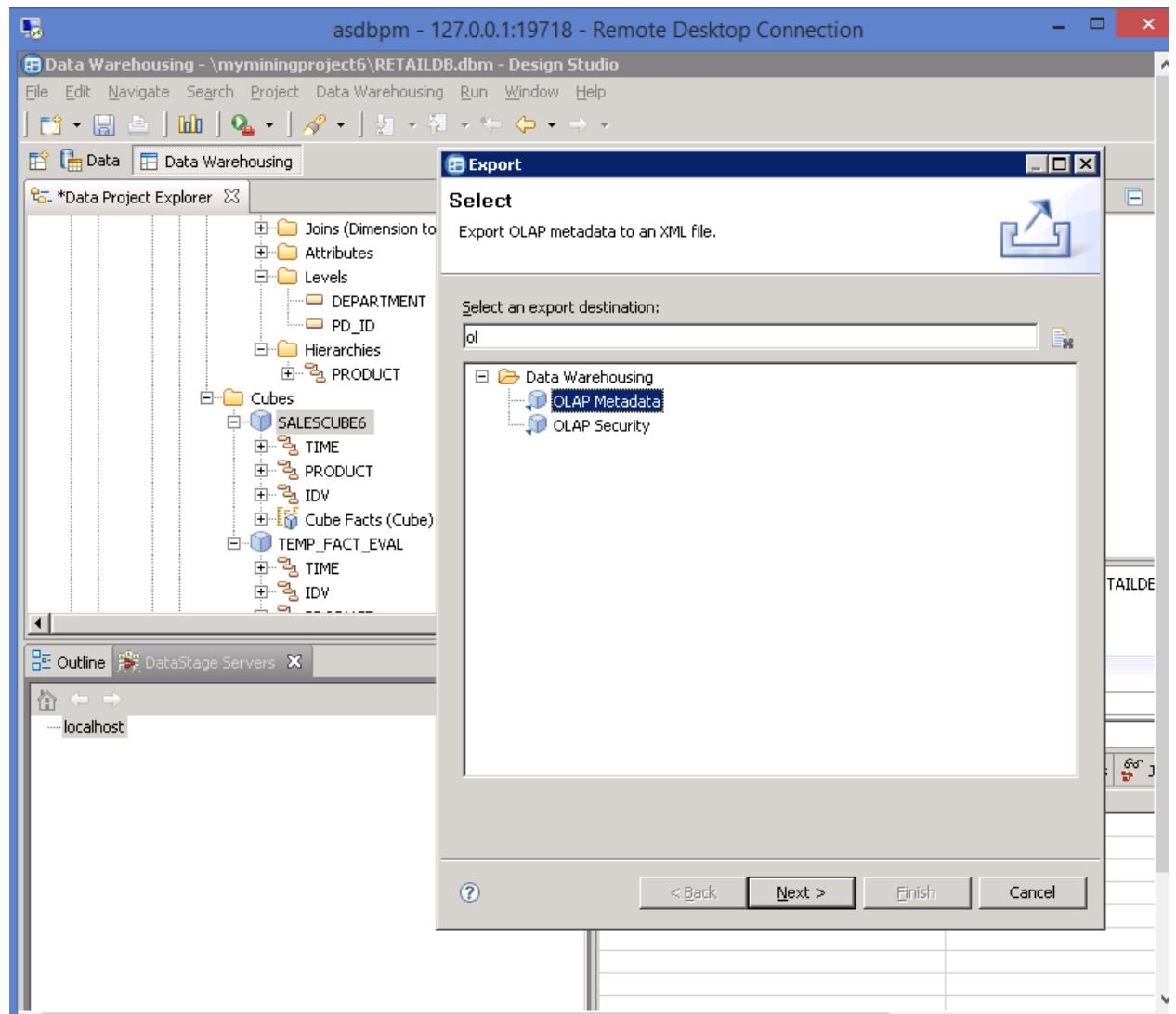
1. Copy the cognos report (Year on Year Profit Comparison Individual) generated in task
2. Edit the report (which is copied) in report studio
3. Change the label "Individual-Year on Year Profit Comparison" to "Individual-Year on Year Income Comparison".
4. Change the label "Profit previous and current year" to "Income previous and current year".
5. Change the x-axis measure in the graph "Income previous and current year" from Profit amount to Income amount.
6. Add the x-axis measure as Profit amount in the graph "% Increase from previous year"

## TASK 6

Same steps as task2 but cube is define based on TEMP\_FACT\_EVAL table







File Edit View Favorites Tools Help

Favorites Suggested Sites Web Slice Gallery Admin Console Welcome to IBM Cognos 8

Warehouse Administration Console Log Out Mining

InfoSphere Warehouse Import OLAP Metadata from a File

Welcome Configuration Step 2 of 4: Specify Metadata XML File to Import

Select which metadata XML file to use for the import operation.

Client miningproject6exportcubemodel.xml

Server

Import OLAP metadata from a file

Import Cube Model

PRCHS\_PRFL\_ANLYSIS  
Sales  
SalesHistoric  
SalesRecent  
PRCHS\_PRFL\_DTL\_2  
VIRTUAL

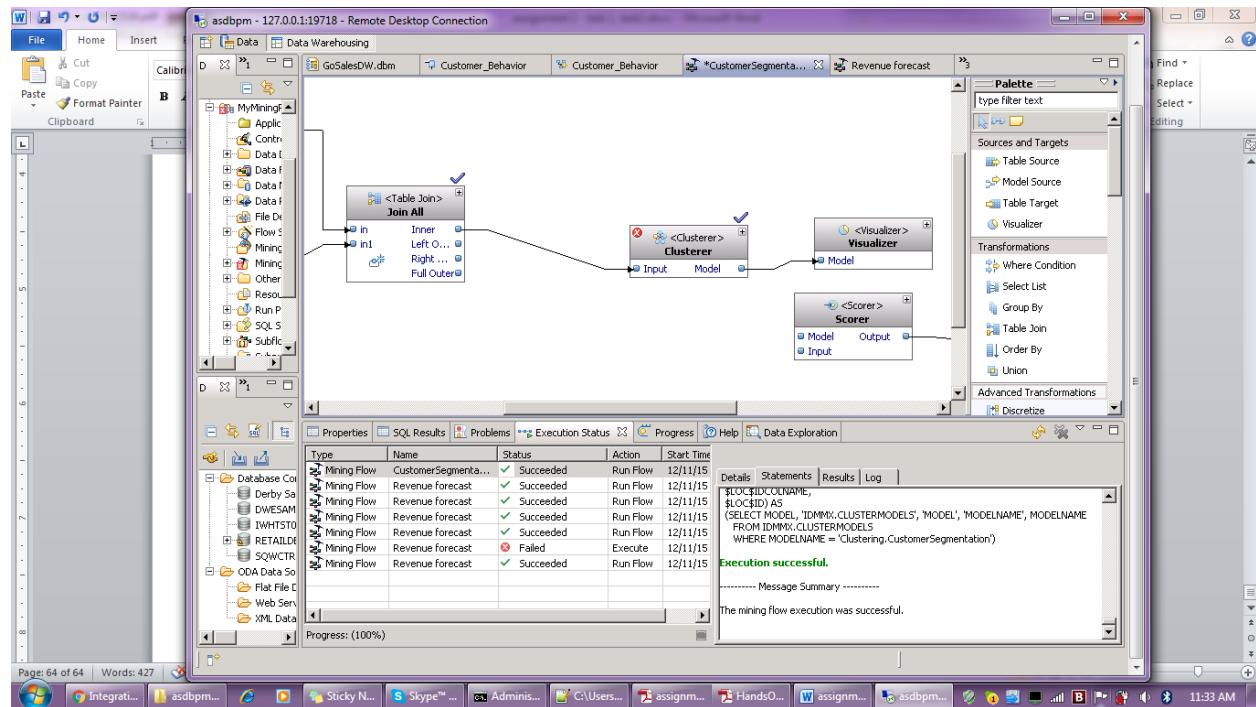
RETAILDB Back Next Finish Cancel

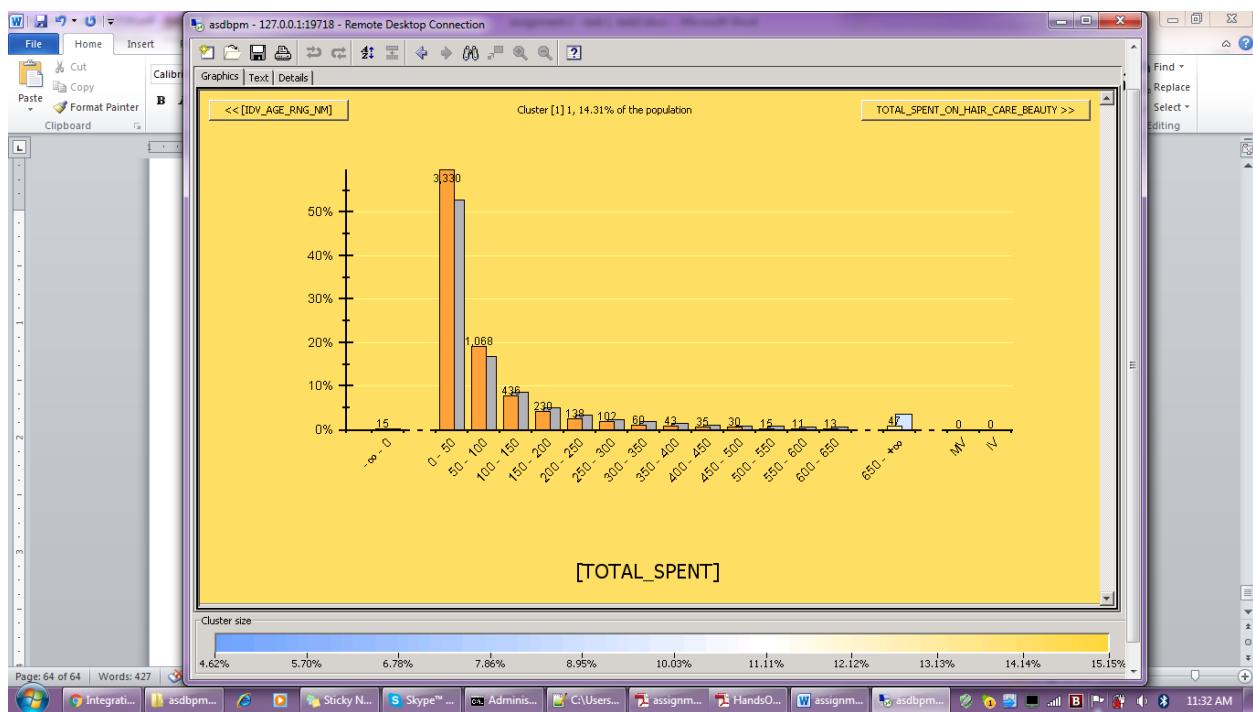
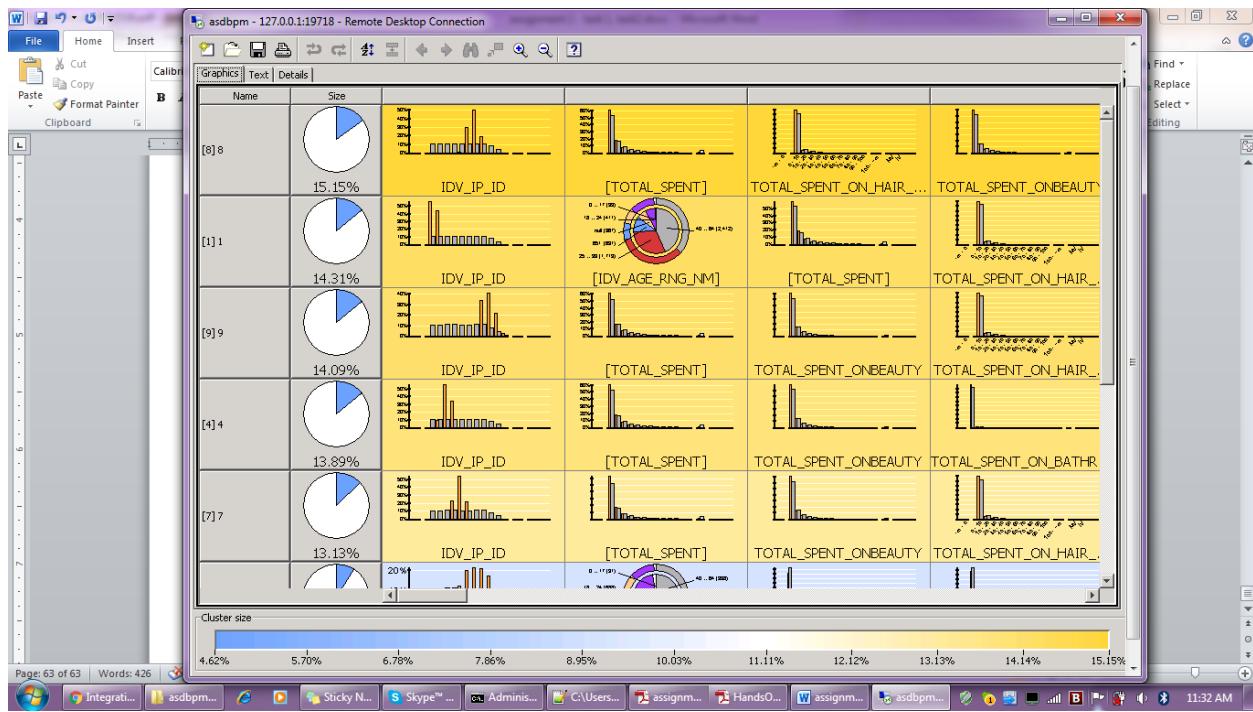
This screenshot shows the 'Import OLAP Metadata from a File' dialog box in the InfoSphere Warehouse Administration Console. The dialog is titled 'Step 2 of 4: Specify Metadata XML File to Import'. It asks the user to select a metadata XML file for the import operation. There are two radio buttons: 'Client' (selected) and 'Server'. The 'Client' radio button has the file path 'miningproject6exportcubemodel.xml' in its input field, with a 'Browse...' button next to it. The 'Server' radio button has an empty input field. Below the radio buttons is a note: 'Select which metadata XML file to use for the import operation.' To the left of the dialog, there's a sidebar with a tree view of cube models: PRCHS\_PRFL\_ANLYSIS (expanded, showing Sales, SalesHistoric, SalesRecent), PRCHS\_PRFL\_DTL\_2, and VIRTUAL. At the bottom right of the dialog are buttons for Back, Next, Finish, and Cancel. The background shows the main administration console interface with various tabs and links.

## Task 7: Additional Data Mining Model

### Task3:

1. In the Data Model obtained from Task3 in the Join table select Inner Join and connect it to the Clusterer.
2. Right Click the clustere and Run it.
3. Connect a Visualizer and Run it.
4. In the Graph obtained duble click it to view the individual values.





## Add.task4:

1. In the data model obtained from Task 4 in Select List select the output as Sales\_amount and TimeID.
2. Use a Clusterer to forecast the Sales Amount for the time period.
3. Right Click the Clusterer and Run it.

4. In the chart obtained double click to view of individual graph.

