Oracle Clustering Assignment

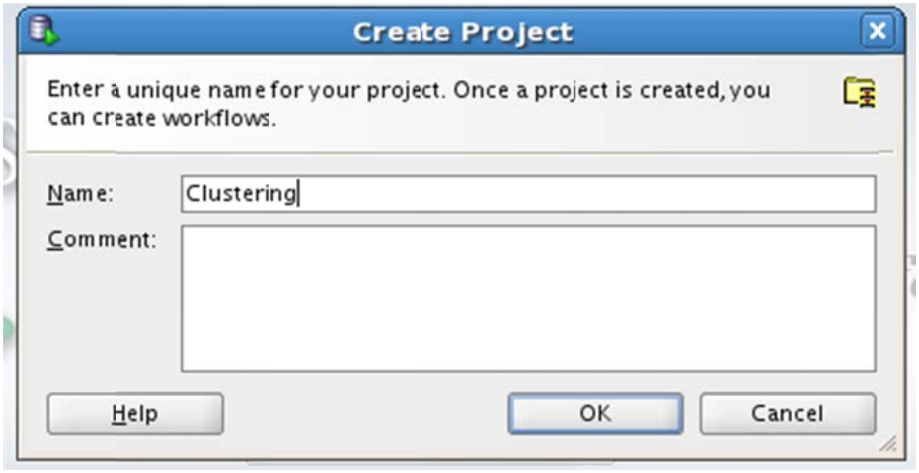
Note: There are three "**Required"** screens.

Start SQL Developer.

Create a project and a workflow

1. On the right side of the screen, there is a Data Miner tab. If it is not there, you can bring it in with View – Data Miner – Data Miner Connections.

2. Right click dmuser and select New Project



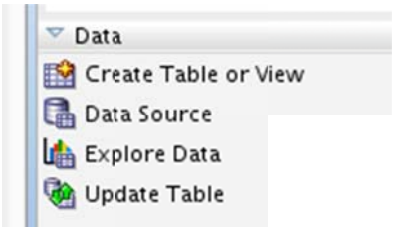
Type Clustering for Name, and click OK.

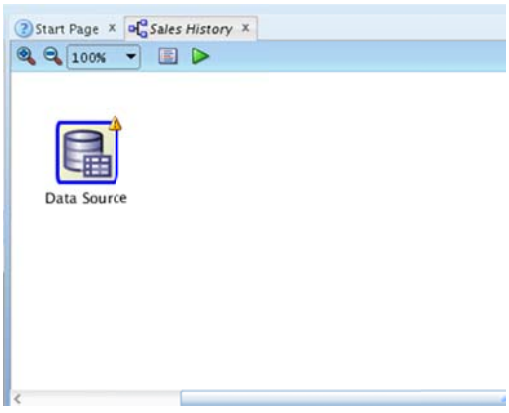
3. The project Clustering appears under dmuser. Right **c**lick the project name (Clustering) and select New Workflow.

Type Electronics Customers for Name, and click OK.

Setup a data source

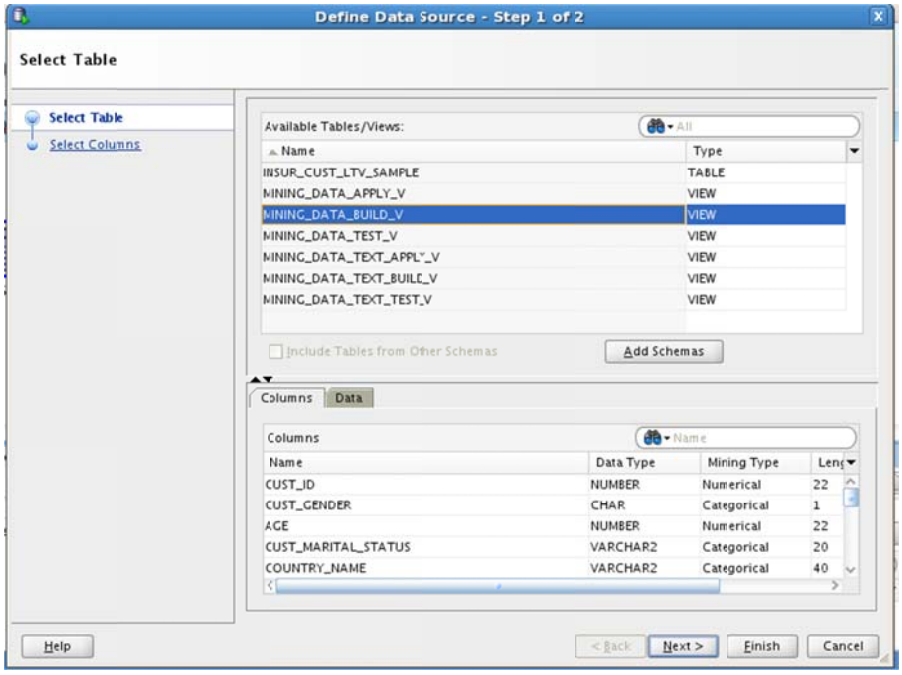
1. Under Component Palette, click Data. Then, click Data Source and drag it to the Electronics Customers workspace.



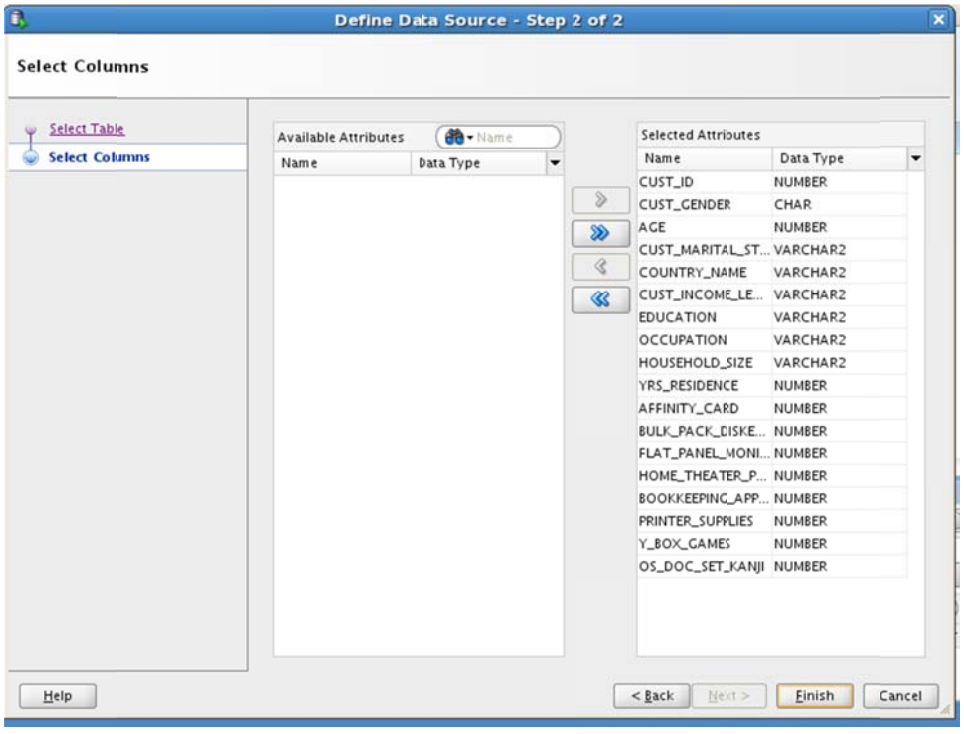


Data Source node is created in the workspace and Define Data Source window pops up.

2. Select MINING_DATA_BUILD_V and click Next.

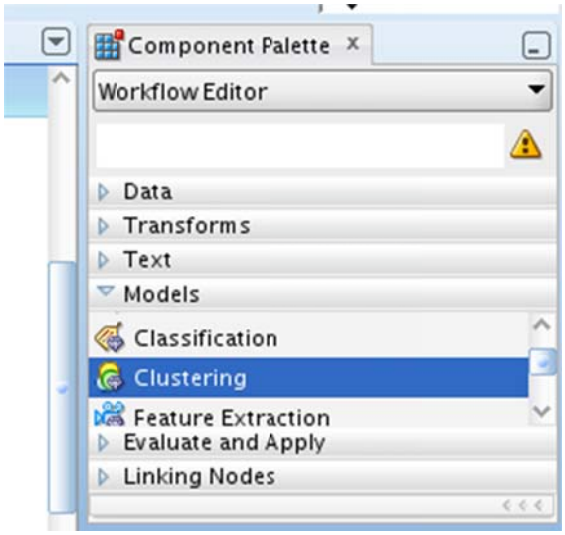


3. Make sure all attributes are select on the right side and click Finish.

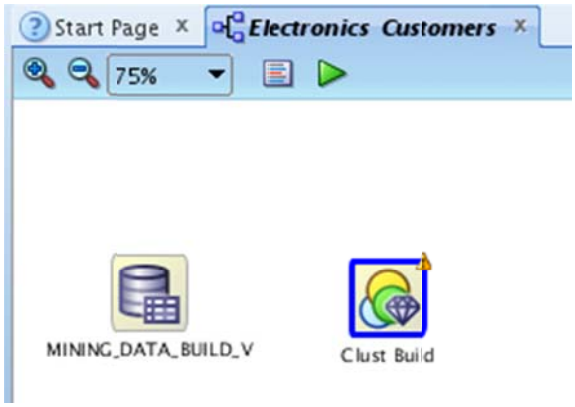


Perform Clustering and Build Model

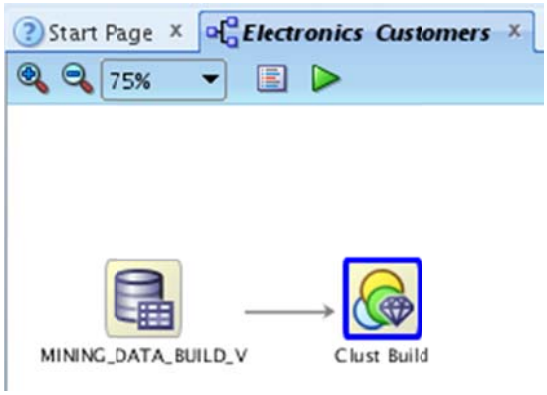
1. Under Component Palette, select and drag Clustering under Models to the work space.



Clust Build node is created in the workspace.



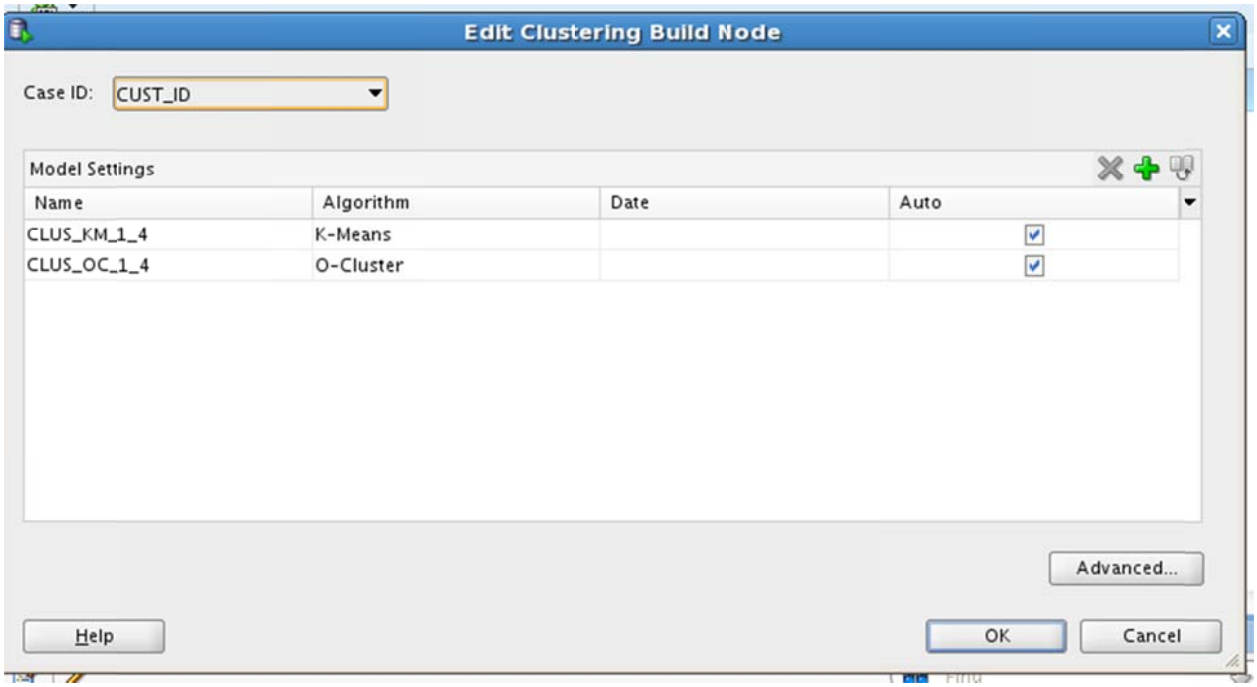
2. Right click MINING_DATA_BUILD_V node and connect it to Clust Build node.



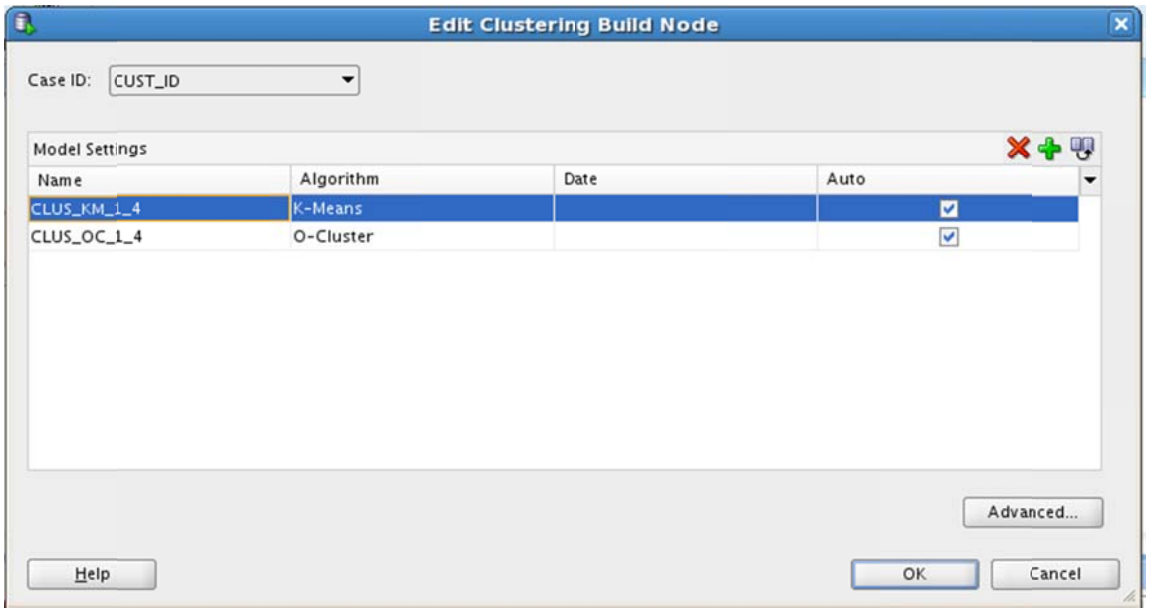
3. Right click Clust Build node and select Edit. Edit Clustering Build Node window appears.

Oracle has two clustering algorithms: enhanced version of K-means and O-cluster (Orthogonal Partitioning Clustering). The enhanced version of K-means creates clusters in a hierarchical manner (this is basically a variation of bisecting K-means). O-cluster is an Oracle proprietary algorithm. It is a grid-based hierarchical clustering algorithm. You can see both algorithms in the Edit Clustering Build Node.

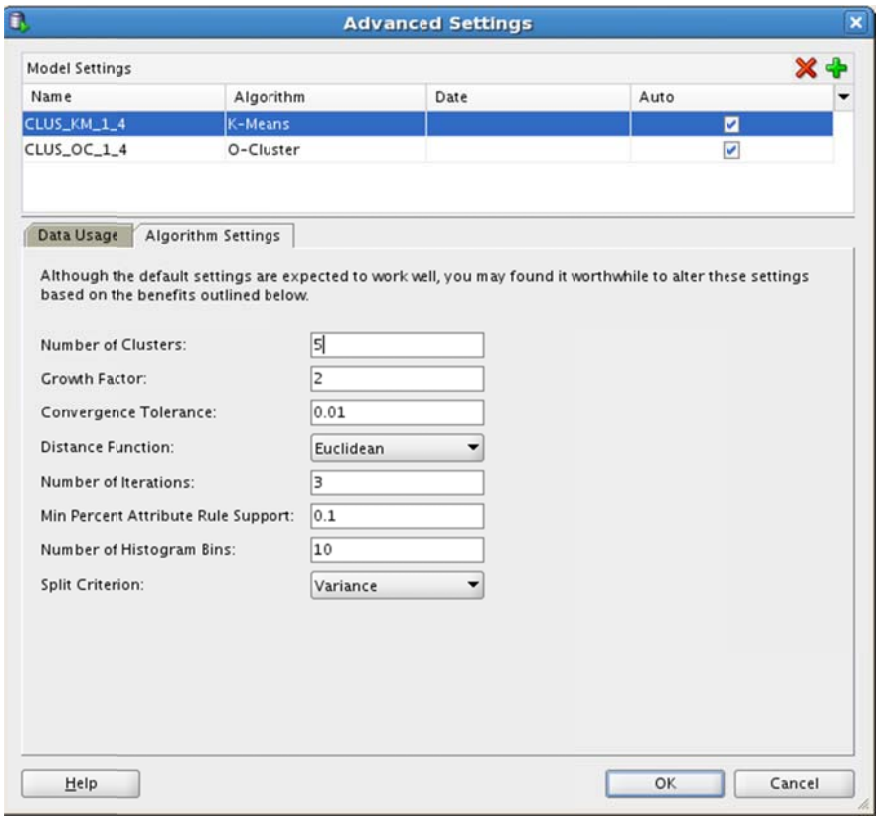
Select CUST_ID for Case ID.



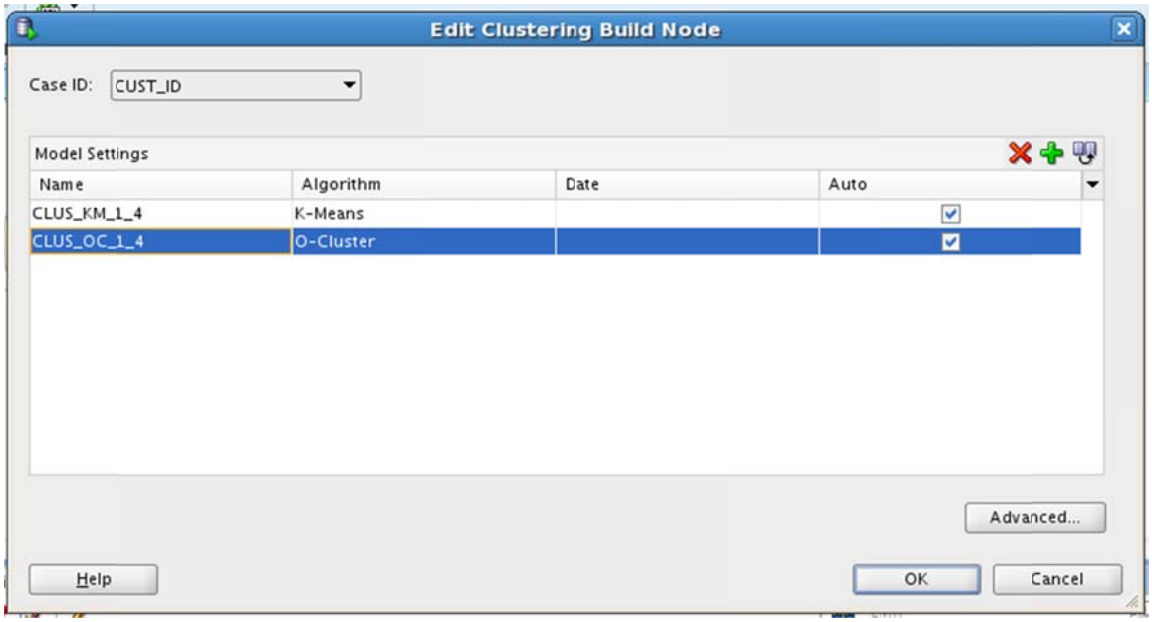
4. Select CLUS_KM_1_4 (you may have different numbers) and click Advanced.



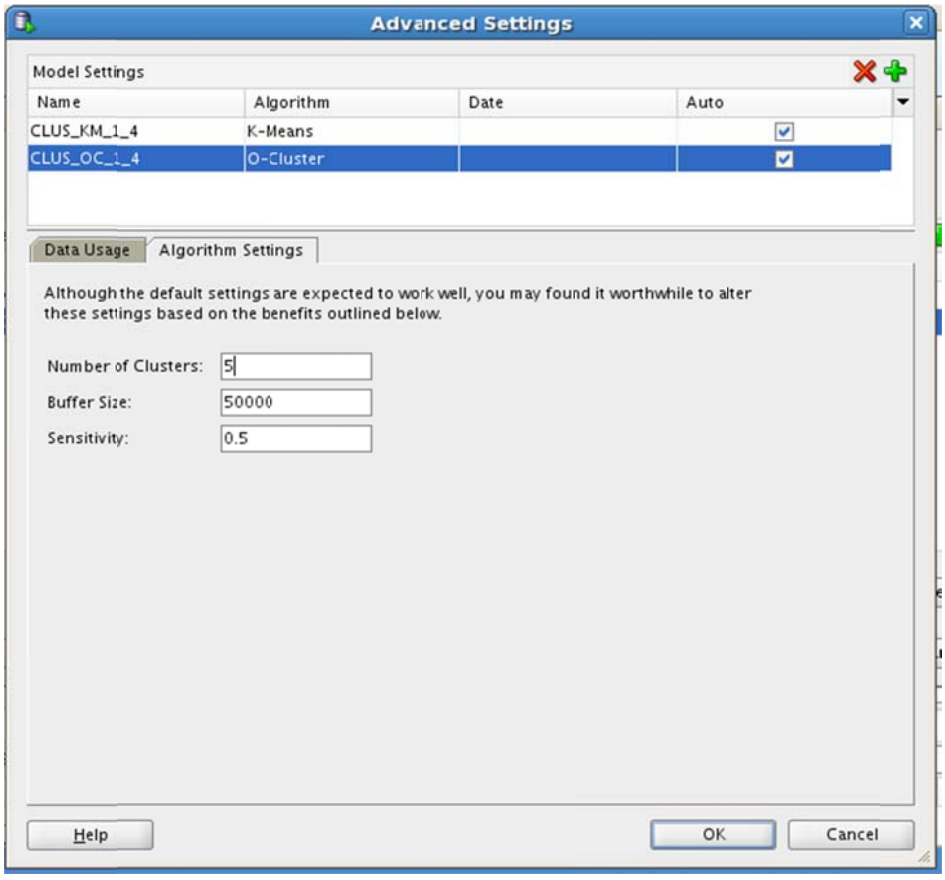
5. Click Algorithms Settings and change the Number of Clusters to 5, and click OK.



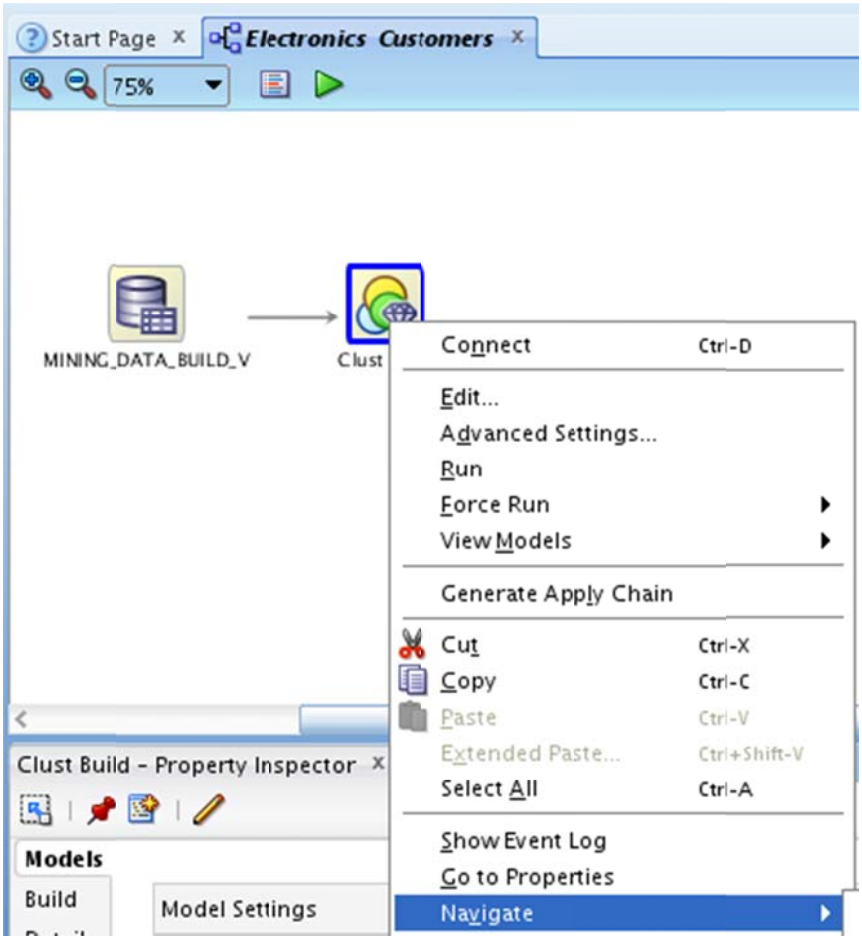
6. Select CLUS_OC_1_4 and click Advanced.



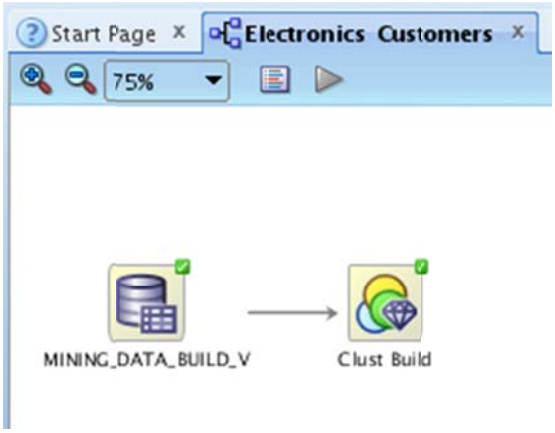
7. Under Algorithms Settings tab, set Number of Clusters to 5 and click OK, then OK.



8. Right click Clust Build node and click run.

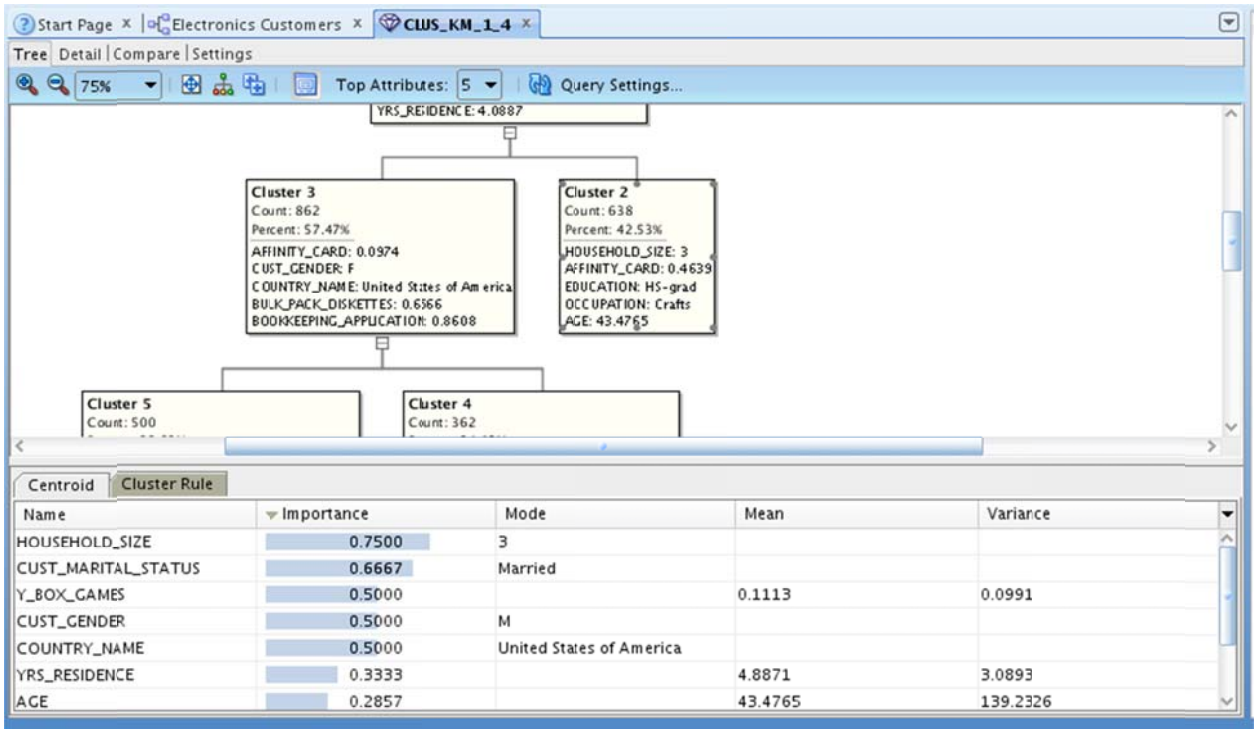


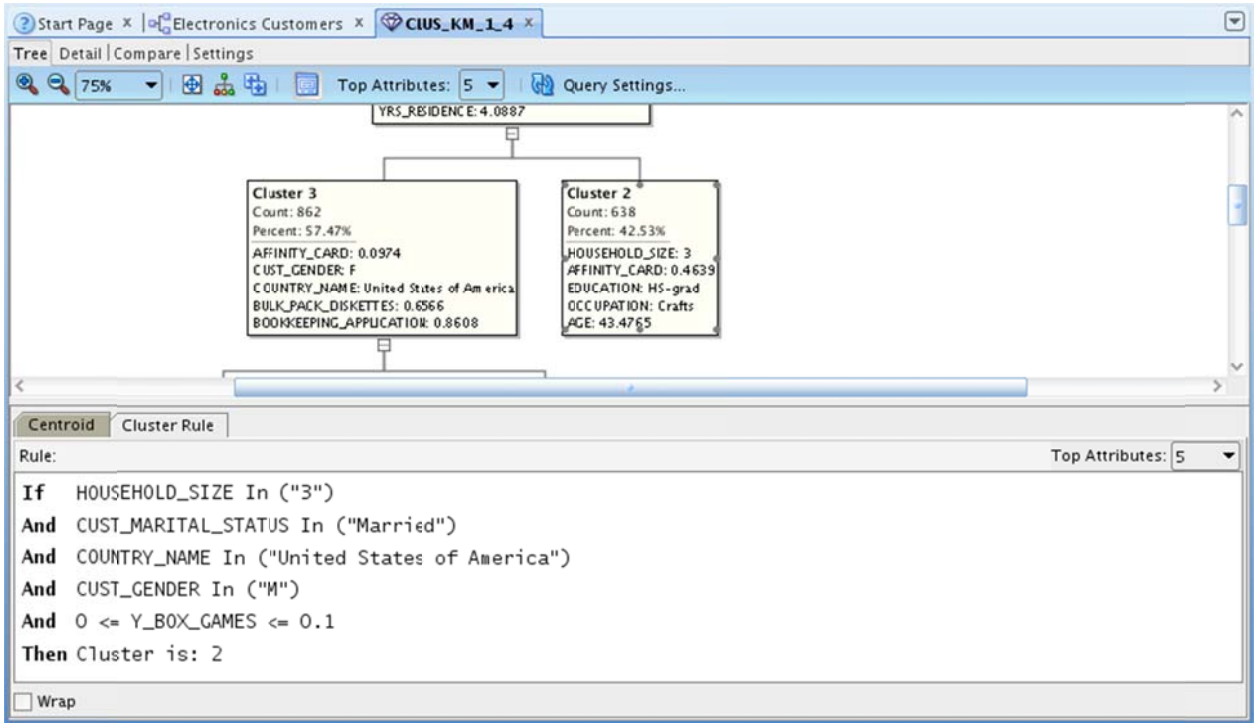
9. After the processing is finished, you will see a green check mark at the upper right corner of the nodes.



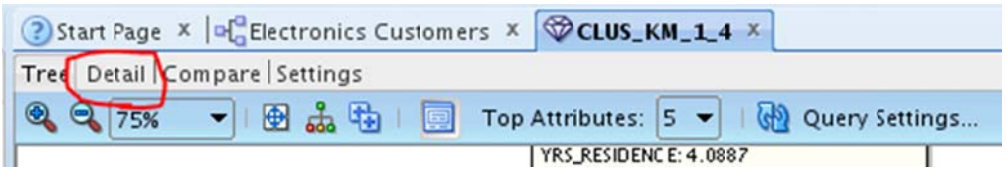
Required: Capture this screen and paste it onto your submission.

10. Right click Clust Build node, click View Models, and select the K-means model. If you select a cluster (in this illustration, Cluster 2 is selected), the centroid and cluster rule appear in a lower window.

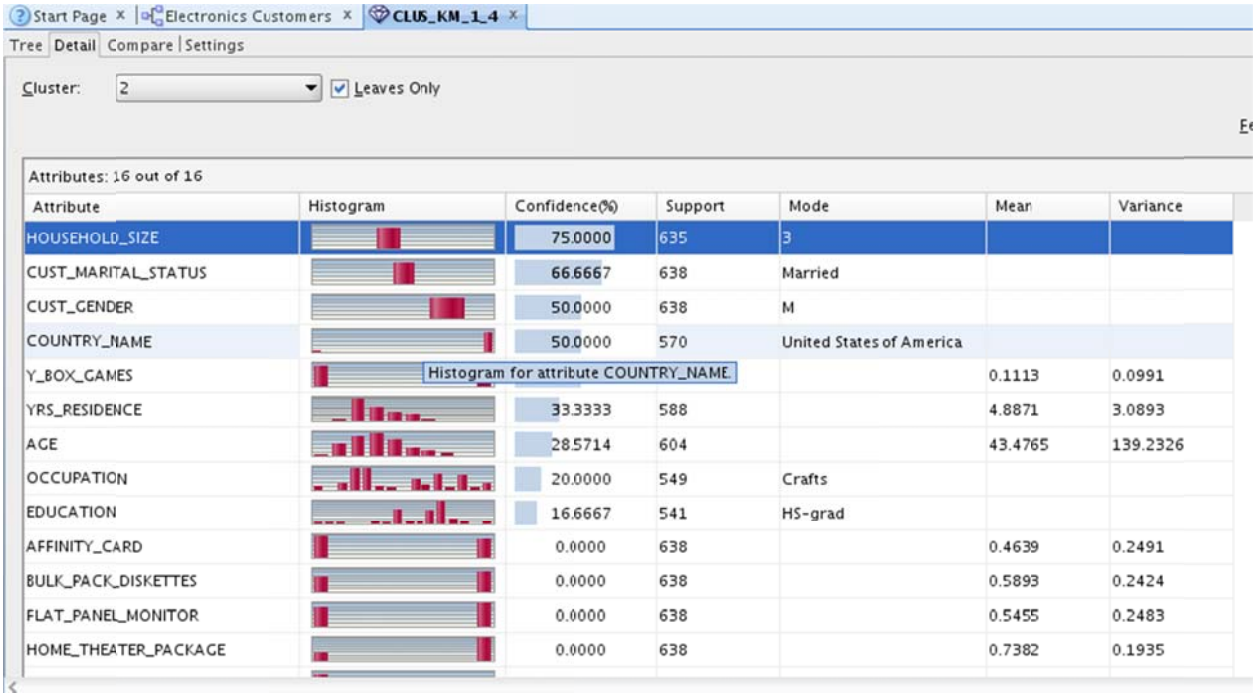




11. Click on Detail tab

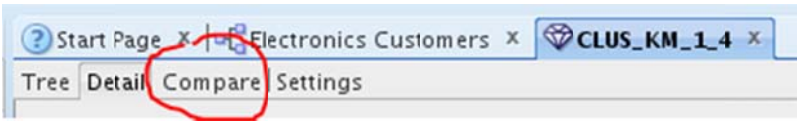


Details of each cluster are displayed (in this example, details of Cluster 2 are shown).

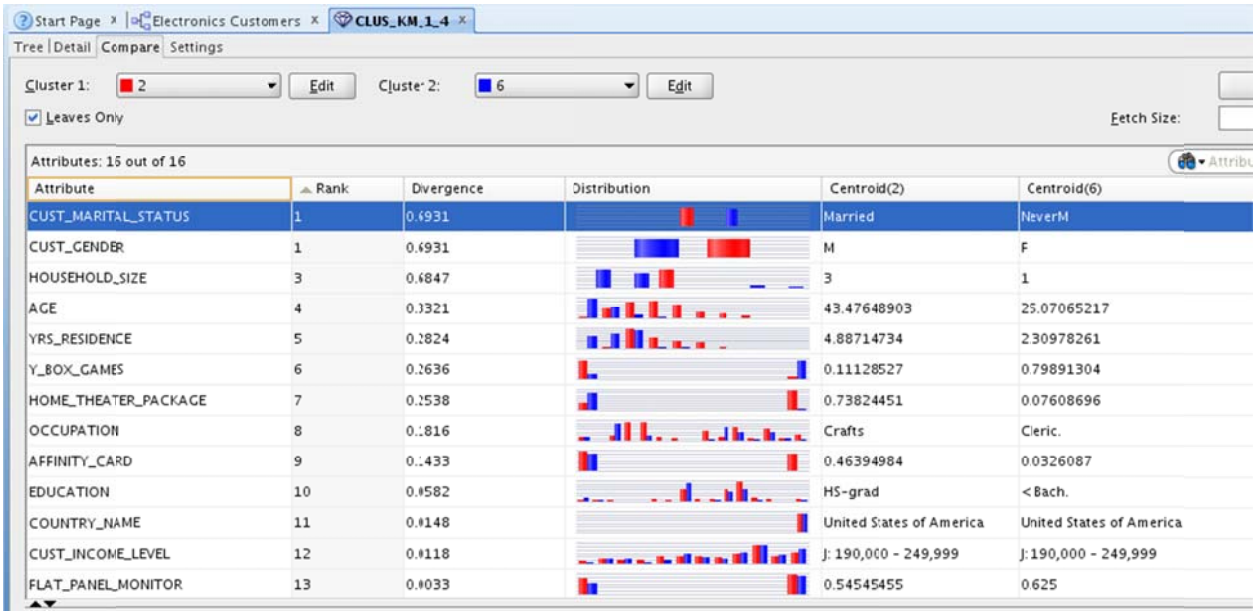


If you scroll down, you can see the histogram of a chosen attribute.

12. You can compare properties of two clusters by clicking Compare tab



Cluster 2 and Cluster 6 are compared:



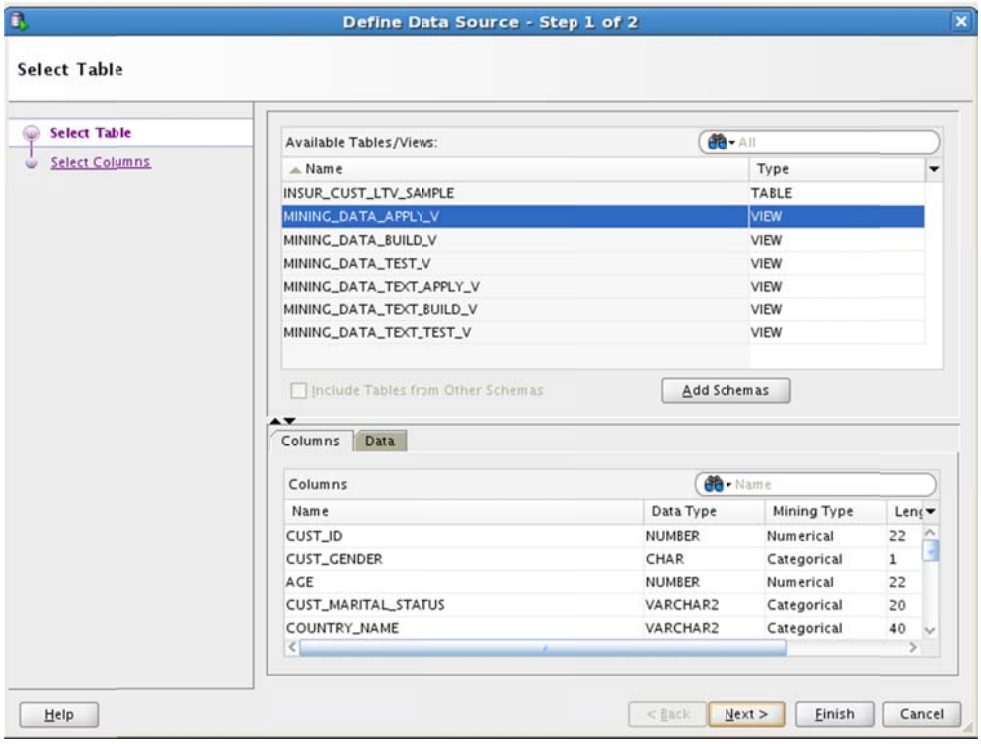
Attribute	Rank	Divergence	Distribution	Centroid(2)	Centroid(6)
CUST_MARITAL_STATUS	1	0.6931		Married	NeverM
CUST_GENDER	1	0.6931		M	F
HOUSEHOLD_SIZE	3	0.6847		3	1
AGE	4	0.3321		43.47648903	25.07065217
YRS_RESIDENCE	5	0.2824		4.88714734	230978261
Y_BOX_GAMES	6	0.2636		0.11128527	0.79891304
HOME_THEATER_PACKAGE	7	0.2538		0.73824451	0.07608696
OCCUPATION	8	0.1816		Crafts	Cleric.
AFFINITY_CARD	9	0.1433		0.46394984	0.0326087
EDUCATION	10	0.0582		HS-grad	<Bach.
COUNTRY_NAME	11	0.0148		United States of America	United States of America
CUST_INCOME_LEVEL	12	0.0118		J: 190,000 - 249,999	J: 190,000 - 249,999
FLAT_PANEL_MONITOR	13	0.0033		0.54545455	0.625

Required: Capture this screen and paste it onto your submission.

Apply the Model

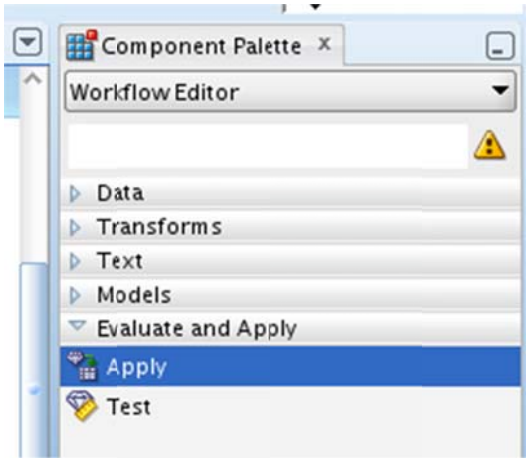
When Oracle performs clustering on a dataset, it also builds a probabilistic model which can be used to assign clusters to objects in an unknown dataset **(**this is similar to using a classifier model to assign class labels to objects in an unknown dataset). You can do this by applying the model to an unknown dataset using Apply Node. You will apply the model to the dataset MINING_DATA_APPLY_V.

1. Return to Electronics Customers workspace, select and drag Data Source node to the workplace. In the Define Data Source window, select MINING_DATA_APPLY_V and click Finish.

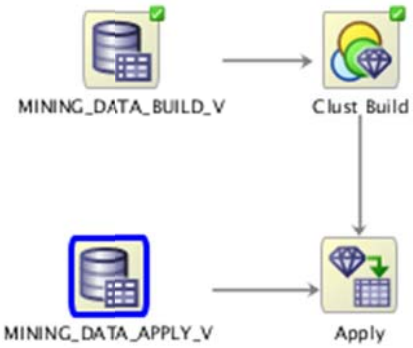


MINING_DATA_APPLY_V node is created in the workspace.

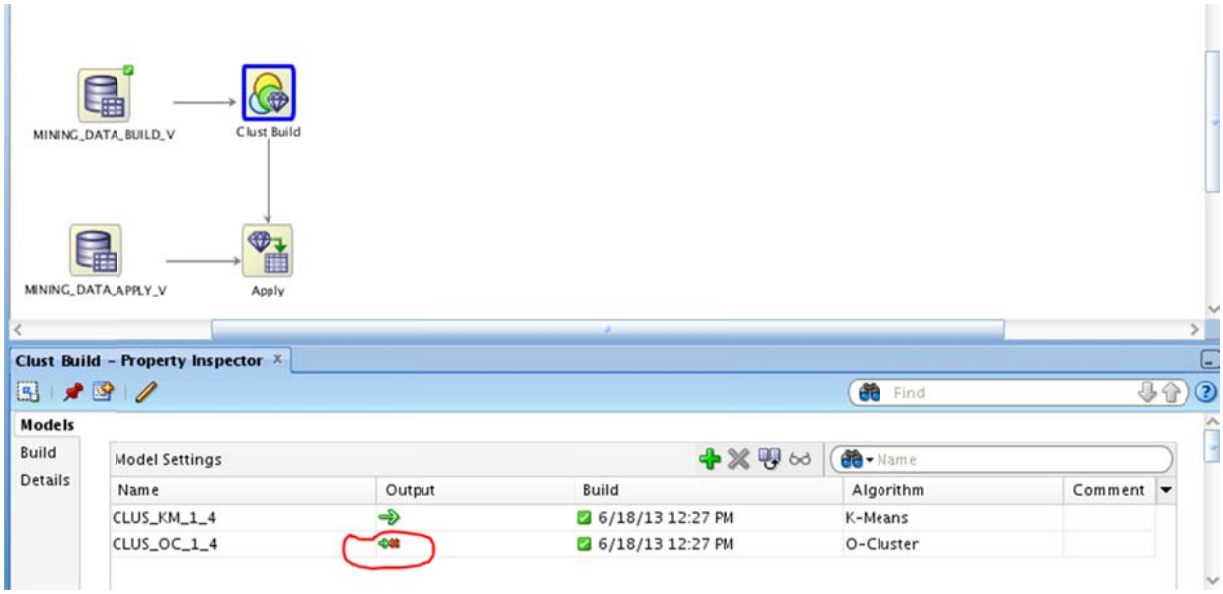
2. Select and drag Apply node under Evaluate and Apply in Component Palette to the workspace.



3. Connect Clust Build node and MINING_DATA_APPLY_V node to Apply node.

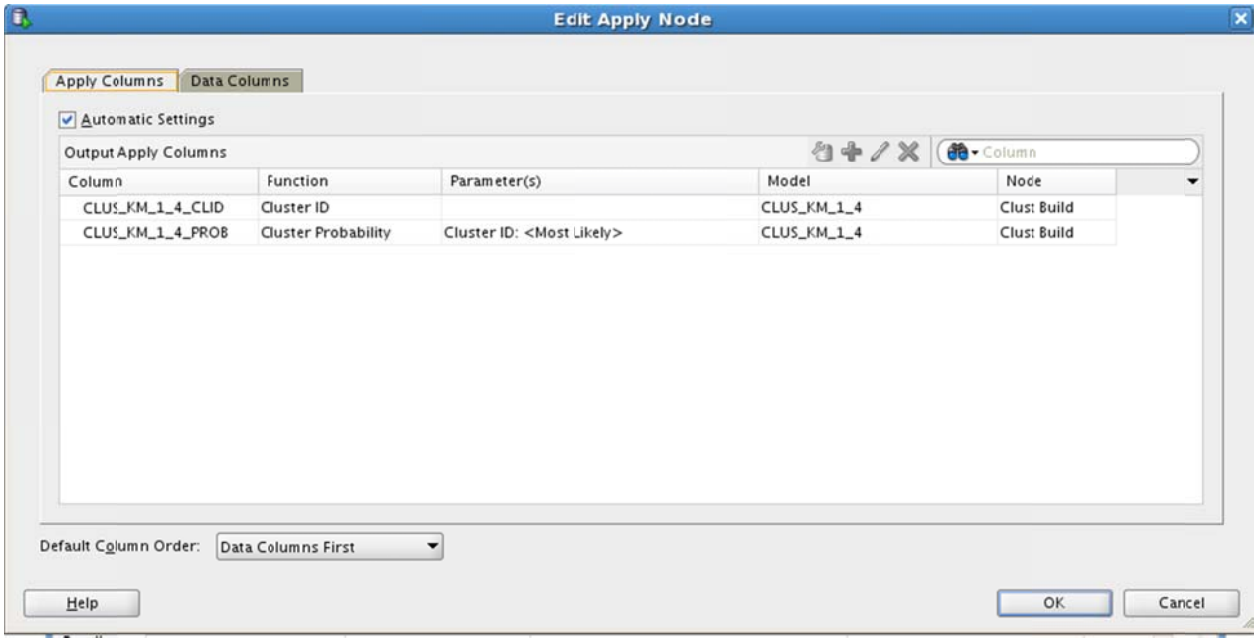


4. You will apply the model of K-means to MINING_DATA_APPLY_V node. To do that, click Clust Build node. In the Clust Build – Property Inspector window, click Output column of CLUS_OC_1_4 (O-cluster). A small, red “x” will appear in the box.

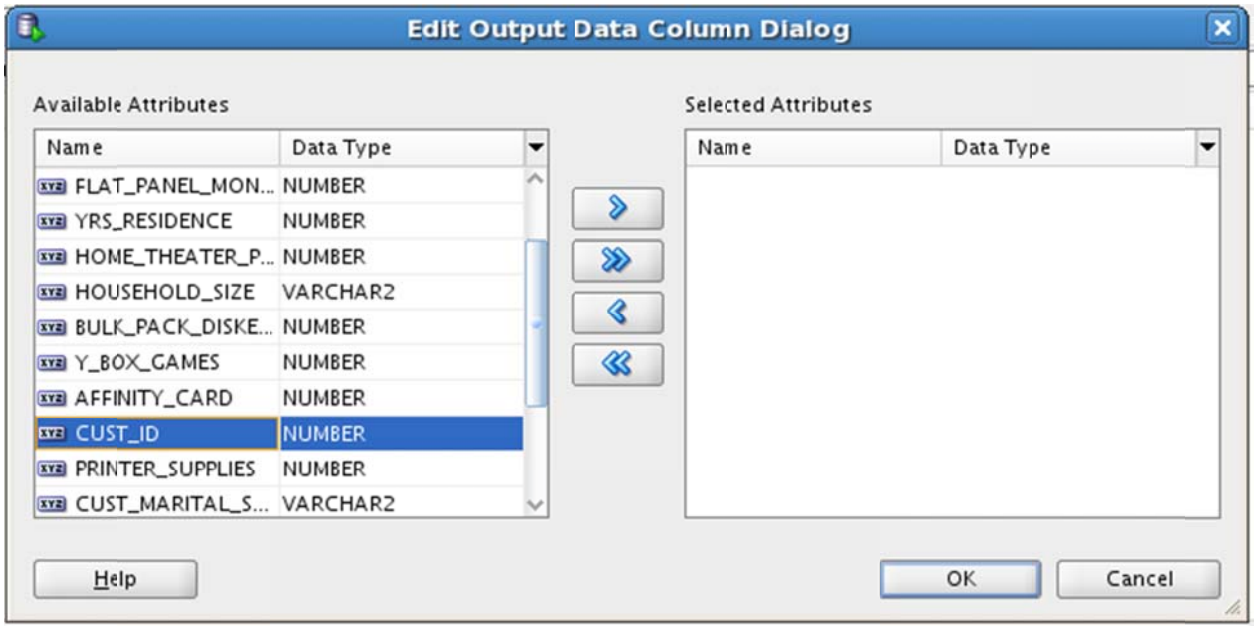


5. The result of applying a model will include two columns by default: cluster id and probability. Since you also want to see cluster assignment of each customer, you will include the customer id in the result (you can include other attributes if you want).

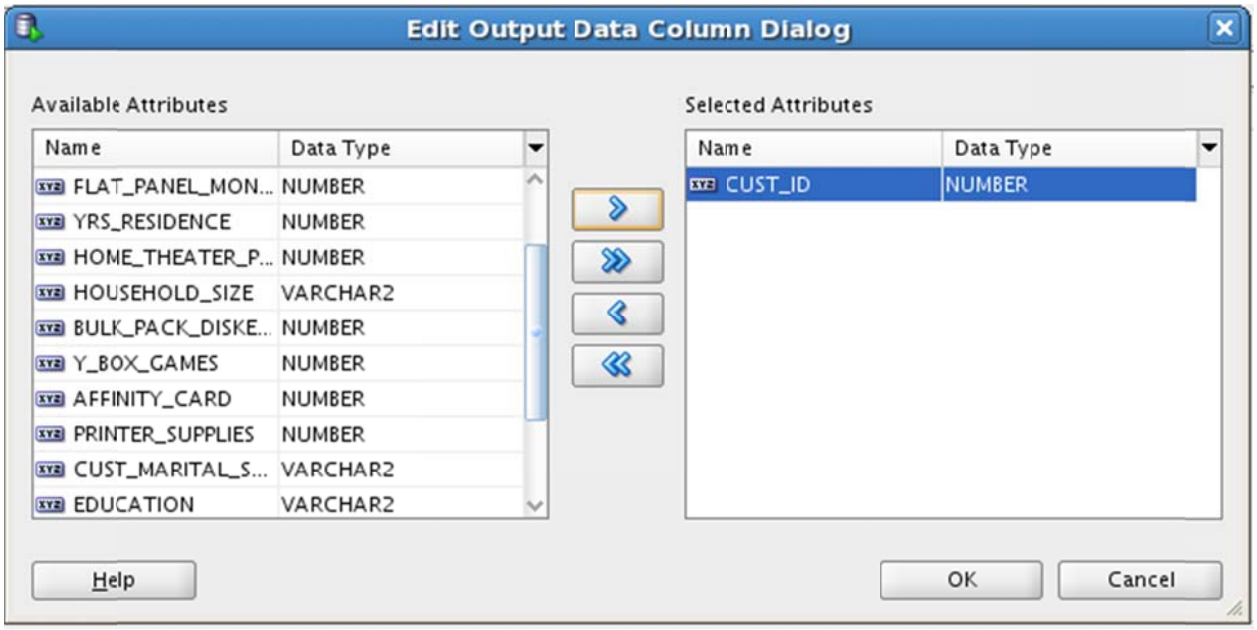
Right click Apply node and select Edit. Edit Apply Node window appears.



6. Click Data Columns tab and click green “+” symbol.

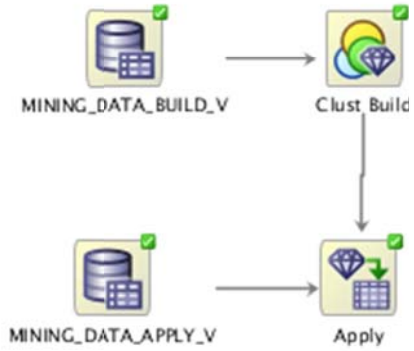


7. In the Edit Output Data Column Dialog window, select CUST_ID and move it to the right.

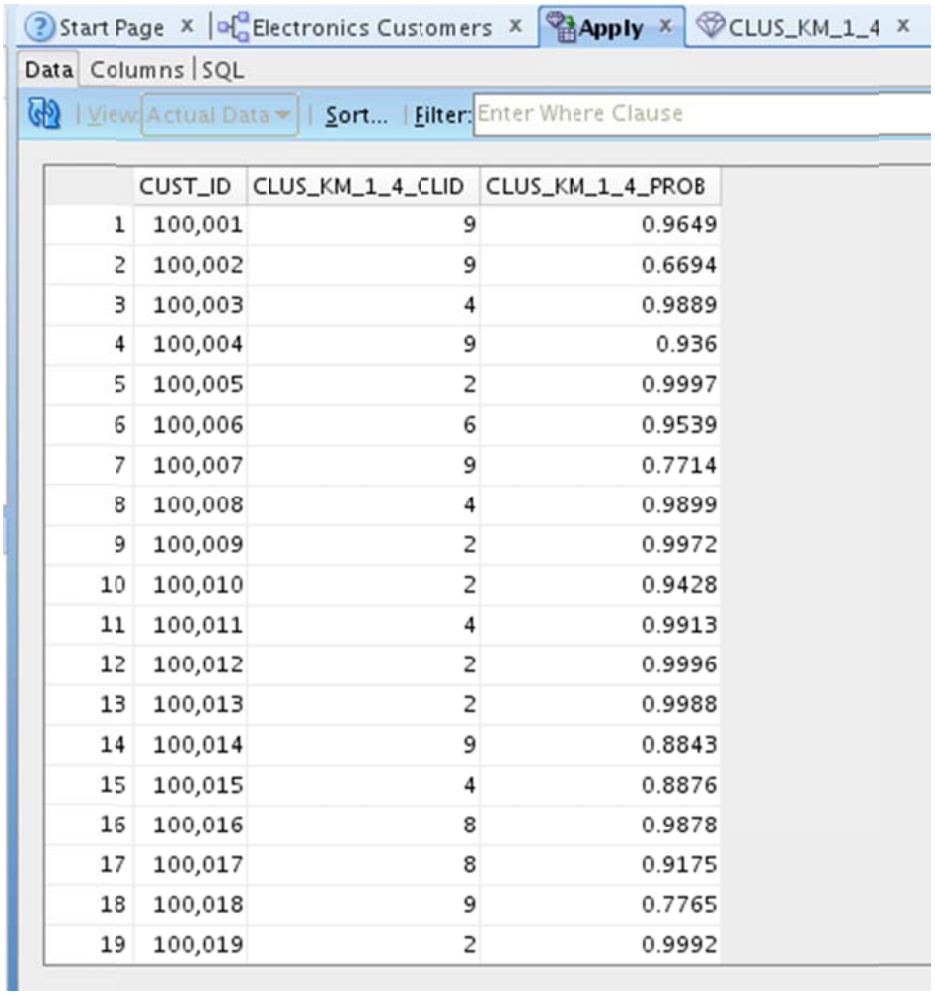


Click OK, and OK.

8. Right click Apply node and click Run.

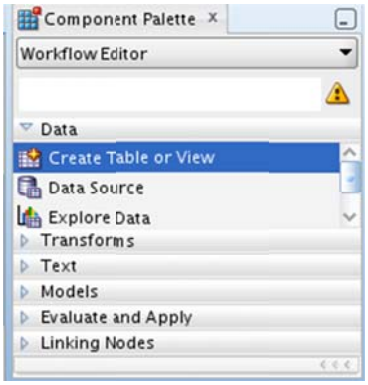


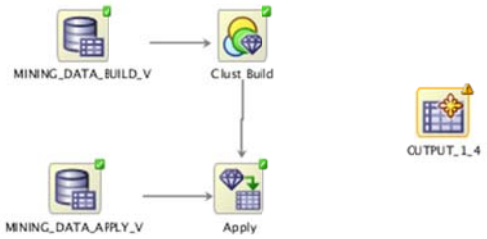
9. To see the result, right click Apply node and click View Data. You can see cluster assignments of all customers in the unknown dataset.



	CUST_ID	CLUS_KM_1_4_CLID	CLUS_KM_1_4_PROB
1	100,001	9	0.9649
2	100,002	9	0.6694
3	100,003	4	0.9889
4	100,004	9	0.936
5	100,005	2	0.9997
6	100,006	6	0.9539
7	100,007	9	0.7714
8	100,008	4	0.9899
9	100,009	2	0.9972
10	100,010	2	0.9428
11	100,011	4	0.9913
12	100,012	2	0.9996
13	100,013	2	0.9988
14	100,014	9	0.8843
15	100,015	4	0.8876
16	100,016	8	0.9878
17	100,017	8	0.9175
18	100,018	9	0.7765
19	100,019	2	0.9992

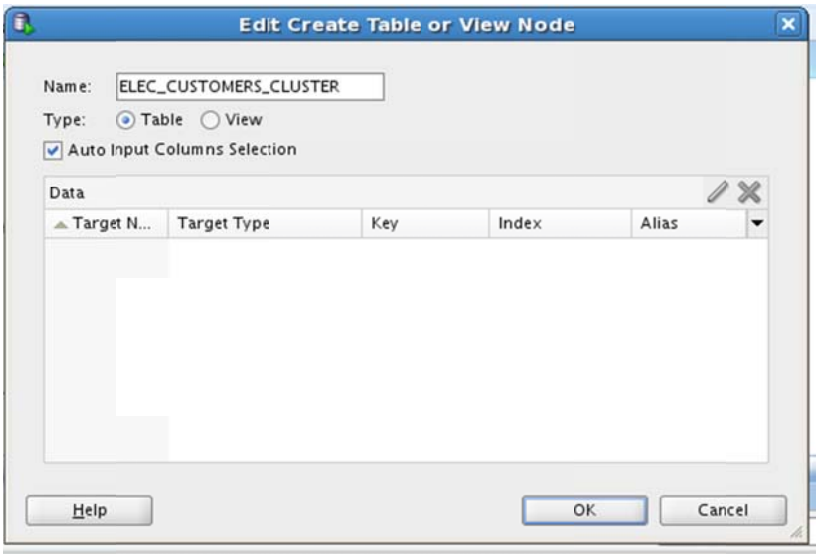
10. You can export the cluster assignments to a table. To do that, return to Electronics Customers workspace, drag Create Table or View from Data under Component Palette into the workspace.



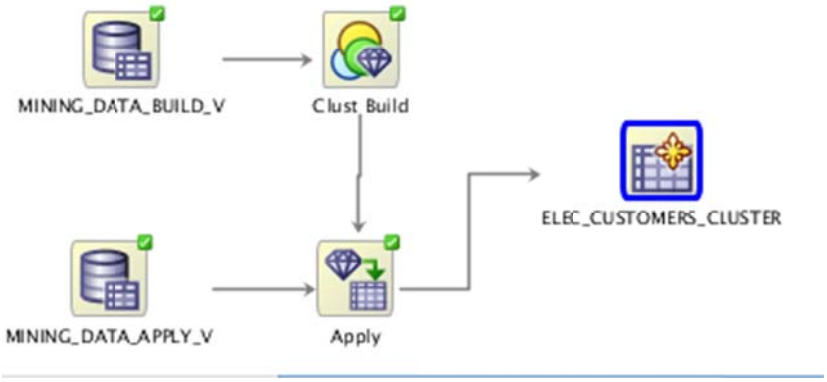


11. Rename the node (OUTPUT_1_4 in this example) to ELEC_CUSTOMER_CLUSTER. To do that, right click the new node (OUTPUT_1_4 in this example) and click Edit.

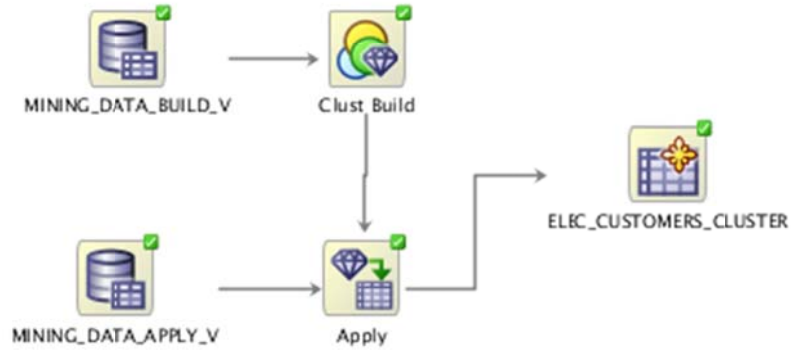
In Edit Create Table or View Node window, type ELEC_CUSTOMER_CLUSTER for name. Make sure that Table is selected and click OK.



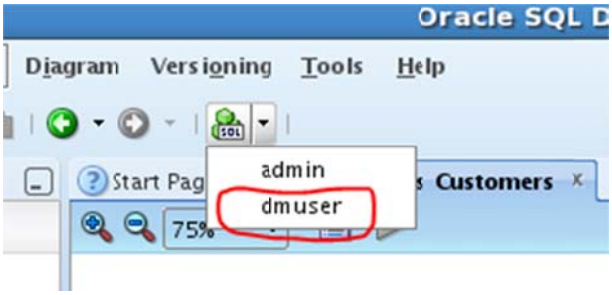
12. Connect Apply node to ELEC_CUSTOMERS_CLUSTER node.



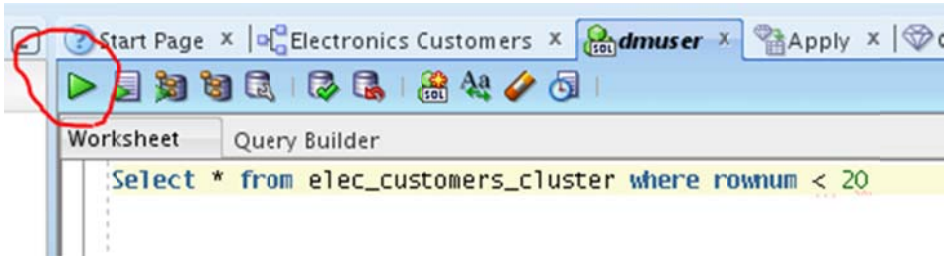
13. Right click ELEC_CUSTOMERS_CLUSTER node and click run. Green check mark appears at the upper right corner of ELEC_CUSTOMERS_CLUSTER



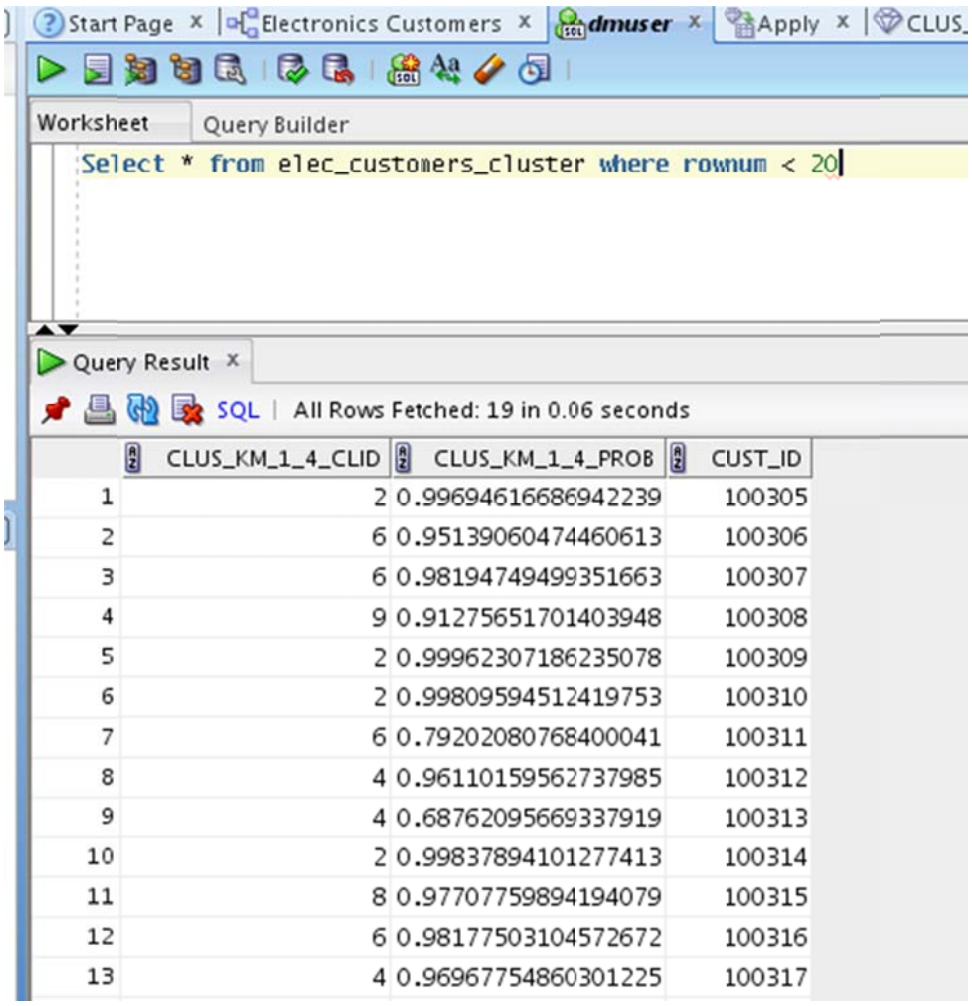
14. To verify the creation of the table, open SQL Worksheet by clicking dmuser from the SQL Worksheet drop down menu.



15. Type the query select * from elec_customers_cluster where rownum < 20 in the worksheet window and click the green triangle as shown below:



The query result will show up in the Query Result window.



	CLUS_KM_1_4_CLID	CLUS_KM_1_4_PROB	CUST_ID
1	2	0.99694616686942239	100305
2	6	0.95139060474460613	100306
3	6	0.98194749499351663	100307
4	9	0.91275651701403948	100308
5	2	0.99962307186235078	100309
6	2	0.99809594512419753	100310
7	6	0.79202080768400041	100311
8	4	0.96110159562737985	100312
9	4	0.68762095669337919	100313
10	2	0.99837894101277413	100314
11	8	0.97707759894194079	100315
12	6	0.98177503104572672	100316
13	4	0.96967754860301225	100317

Required: Capture this screen and paste it onto your submission.