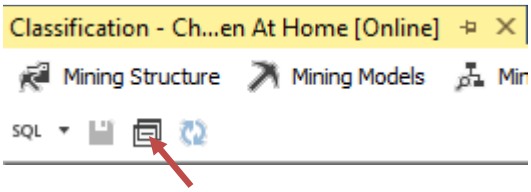
Hands on Exercise for Data Mining – Part 2

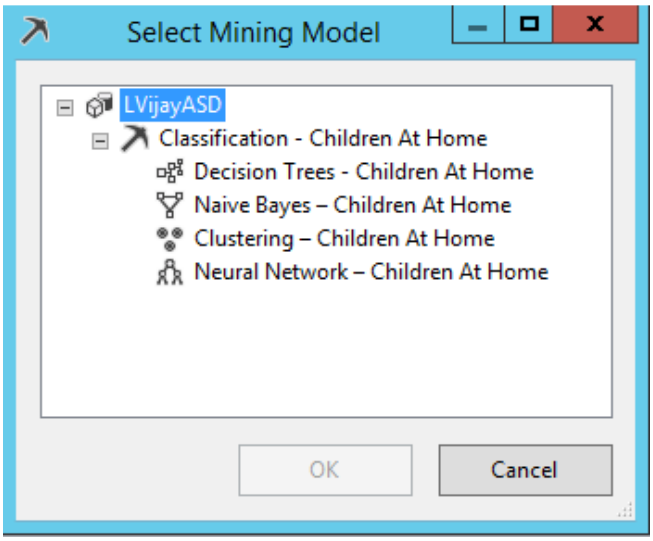
Log on to elab using <u>your eID credentials</u>

Learn by Doing – Creating a Singleton Query (From previous edition of Larson's Delivering BI book)

1. Open Visual Studio 2017
2. Click File→Open→Analysis Services Database. Select your analysis services database (it should be displayed in the box below the Database drop-down list) in the Connect to Database dialog box. Click OK. (**Note:** if your database is not displayed, then: In the Connect To Database dialog box, enter **buscissql\cisbi** in the Server textbox and select **your analysis services database (e.g., if your last name is Smith, your analysis services database will be called SmithASD)** from the Database drop-down list.
3. If the Data Mining Design tab for the Classification – Children At Home data mining structure is not displayed, double-click the entry for this data mining structure in the Solution Explorer window.
4. Select the Mining Model Prediction tab on the Data Mining Design tab.
5. Select the Singleton Query button on the Mining Model Prediction tab toolbar (shown below).

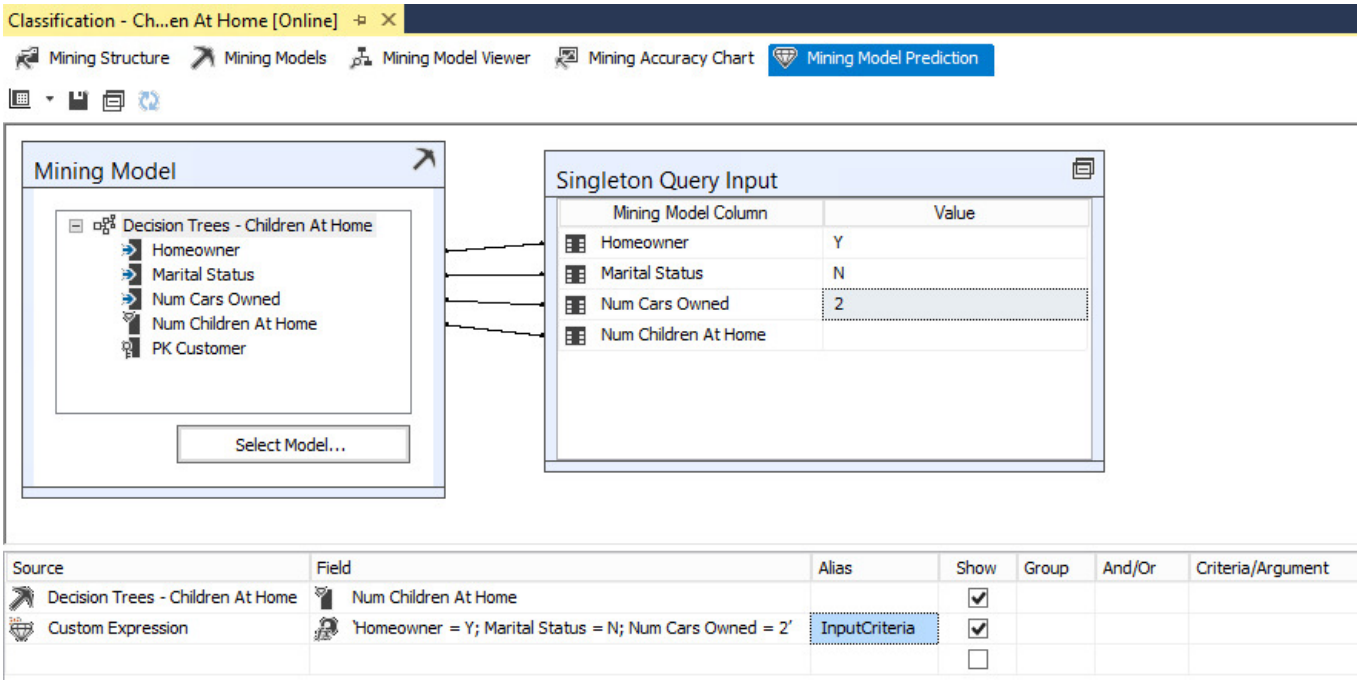


6. Click Select Model in the Mining Model window. The Select Mining Model dialog box appears, as shown below.

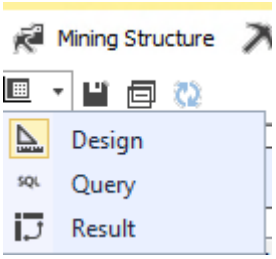


7. Select the Decision Trees – Children At Home mining model and click OK.
8. In the Singleton Query Input window, select Y from the Homeowner drop-down list in the Value column.
9. Select N from the Marital Status drop-down list.
10. Select 2 from the Num Cars Owned drop-down list.

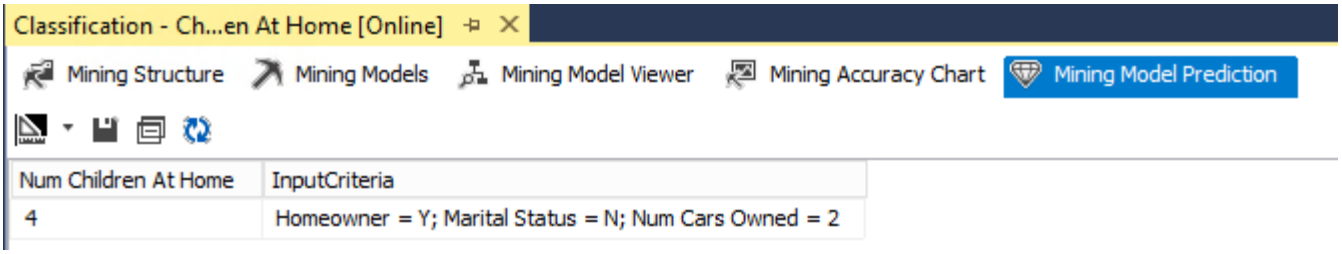
11. Now select the columns in the result set. In the first row under the Source column, select the Decision Trees – Children At Home mining model from the drop-down list. Num Children At Home is selected by default in the Field column because it is the only predictable field in the mining model.
12. In the second row under the Source column, select Custom Expression from the drop-down list.
13. In the second row under the Field column, enter the following:
    'Homeowner = Y; Marital Status = N; Num Cars Owned = 2'
14. In the second row under the Alias column, enter **InputCriteria**. The Mining Model Prediction tab appears as shown below.



15. Select Result from the View drop-down list on the Mining Model Prediction tab toolbar, as show below.



16. The result view is displayed, as shown below. The model predicts that someone who is a homeowner, is not married, and owns two cars will most likely have four children at home.



17. Select Design from the View drop-down list on the Mining Model Prediction tab toolbar to return to the design view.

A Prediction Join Query

A singleton query lets us use a single set of inputs to generate a single row result set. The prediction join query enables us to use a tableful of inputs to generate a multiple-row result set. We feed multiple records to the mining model, and it creates a prediction for each set of these records.

Creating a Prediction Join Query

The prediction join query functions similarly to the singleton query. The difference is that the Singleton Query Input window is replaced by the Select Input Table(s) window. This window lets us select the table or joined set of tables that will serve as the record source for the prediction join.

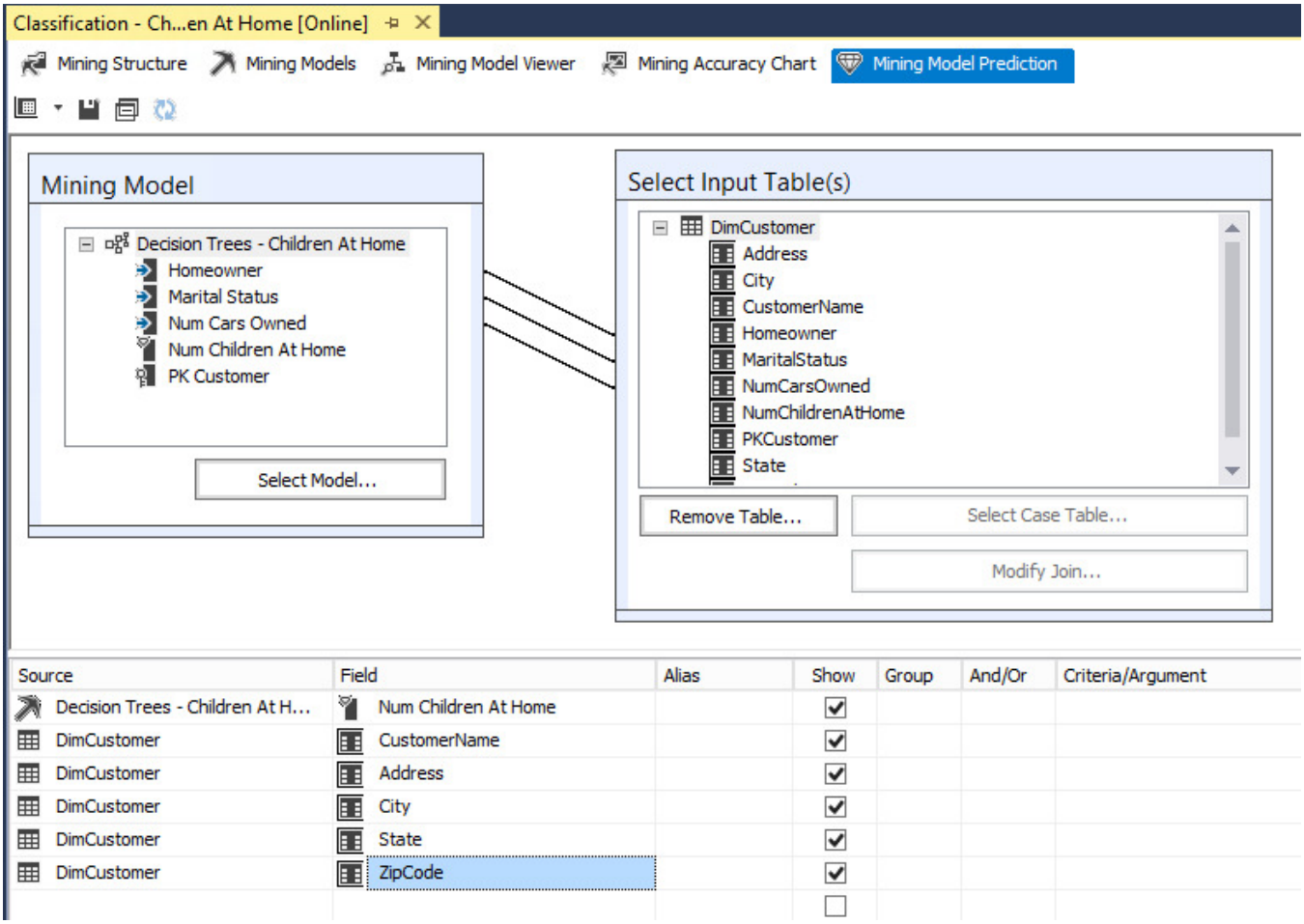
Learn by Doing – Creating a Prediction Join Query

Business Need: In a previous Learn By Doing exercise, we discussed a scenario where the Maximum Miniatures Marketing Department purchased a mailing list that they wanted classified into those households likely to have no children at home and those that have one or more children at home. We now have a trained and tested mining model ready to perform that classification.

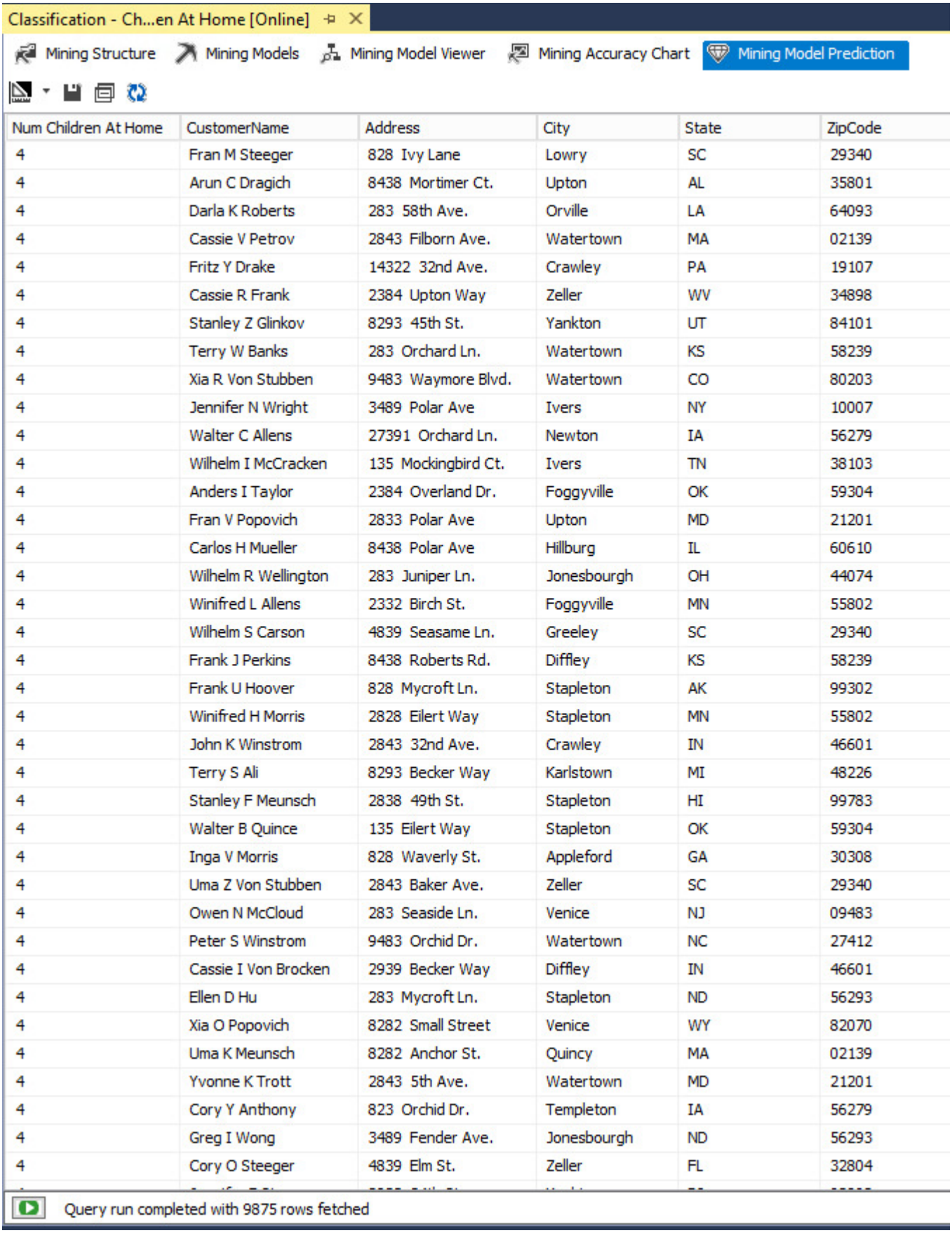
To keep things simple, we can use the same customer table and a little imagination to pretend that this is the newly purchased list. Remember, the reason we had to do this classification was because the newly purchased list did not include the number of children at home. Therefore, if we have our existing customer table play the role of the newly purchased list, we must ignore the fact that it already has a number of children at home field.

1. Click the Singleton Query button on the Mining Model Prediction tab toolbar to toggle back to the Prediction Join Query mode. Click Yes to continue when warned about losing your query.
2. Click Select Case Table in the Select Input Table(s) window. The Select Table dialog box appears.
3. Select Max Min Mining DM data source view (the second of the two listed) from the Data Source drop-down list. DimCustomer should be the only Table/View Name and it should be selected. Click OK. The fields in the Mining Model and the Input Table should be automatically mapped based on field names.
4. We are not supposed to have a Num Children At Home field in this table, so we need to delete this mapping and predict this value instead. Click the mapping line going from Num Children At Home to NumChildrenAtHome so it is highlighted.
5. Right-click the mapping line and select Delete from the context menu.
6. In the first row under the Source column, select Decision Trees-Children At Home mining model from the drop-down list. Again, Num Children At Home is selected by default in the Field column because it is the only predictable field in the mining model.
7. In the second row under the Source column, select DimCustomer table from the drop-down list.
8. In the second row under the Field column, select CustomerName from the drop-down list.
9. In the third row under the Source column, select DimCustomer table from the drop-down list.
10. In the third row under the Field column, select Address from the drop-down list.

11. Continue this process to include City, State, and ZipCode in the result set. The Mining Model Prediction tab appears as shown below.



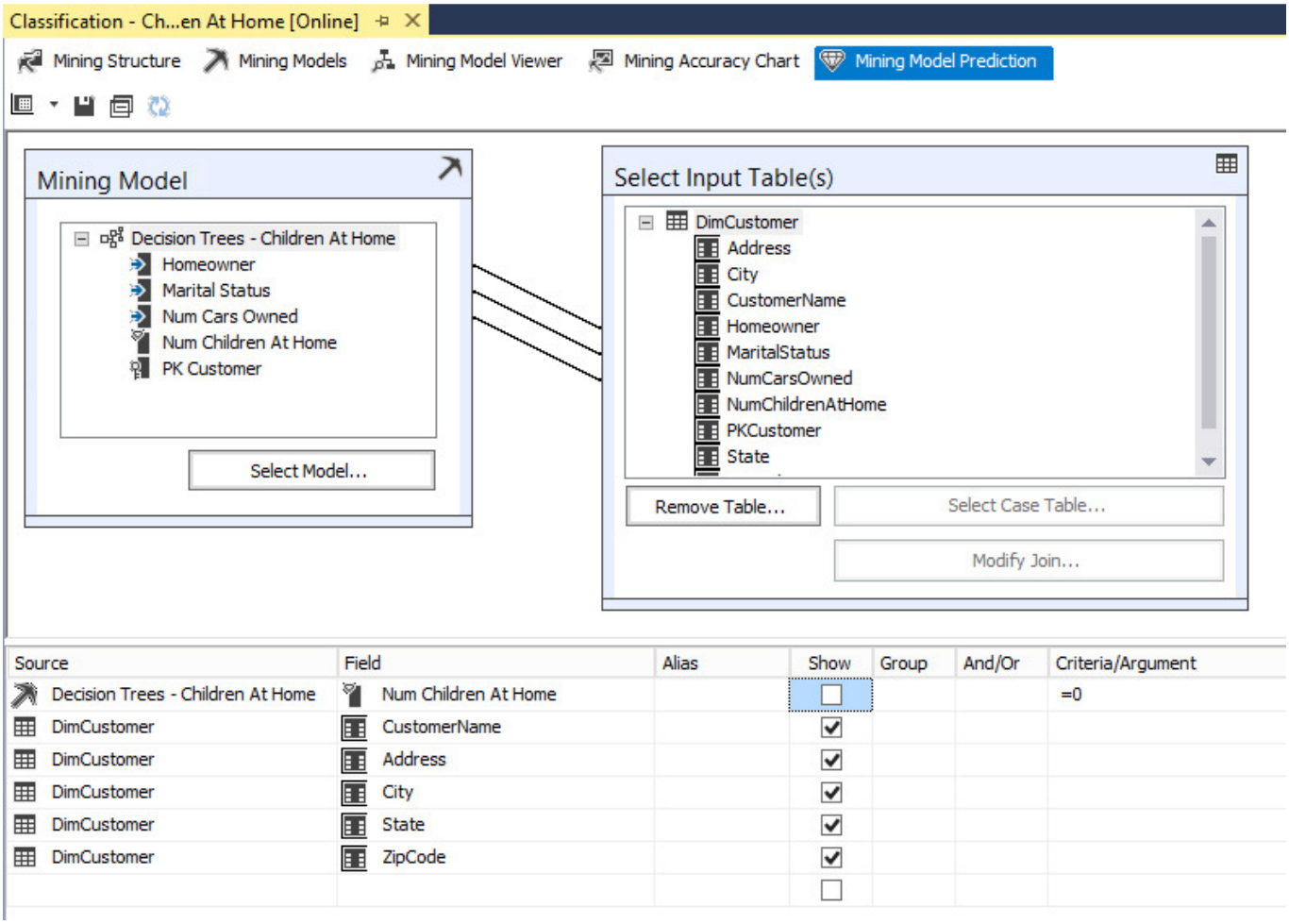
12. Select Result from the View drop-down list on the Mining Model Prediction tab toolbar. The result view is displayed as shown below.



Num Children At Home	CustomerName	Address	City	State	ZipCode
4	Fran M Steeger	828 Ivy Lane	Lowry	SC	29340
4	Arun C Dragich	8438 Mortimer Ct.	Upton	AL	35801
4	Darla K Roberts	283 58th Ave.	Orville	LA	64093
4	Cassie V Petrov	2843 Filborn Ave.	Watertown	MA	02139
4	Fritz Y Drake	14322 32nd Ave.	Crawley	PA	19107
4	Cassie R Frank	2384 Upton Way	Zeller	WV	34898
4	Stanley Z Glinkov	8293 45th St.	Yankton	UT	84101
4	Terry W Banks	283 Orchard Ln.	Watertown	KS	58239
4	Xia R Von Stubben	9483 Waymore Blvd.	Watertown	CO	80203
4	Jennifer N Wright	3489 Polar Ave	Ivers	NY	10007
4	Walter C Allens	27391 Orchard Ln.	Newton	IA	56279
4	Wilhelm I McCracken	135 Mockingbird Ct.	Ivers	TN	38103
4	Anders I Taylor	2384 Overland Dr.	Foggyville	OK	59304
4	Fran V Popovich	2833 Polar Ave	Upton	MD	21201
4	Carlos H Mueller	8438 Polar Ave	Hillburg	IL	60610
4	Wilhelm R Wellington	283 Juniper Ln.	Jonesbourgh	OH	44074
4	Winifred L Allens	2332 Birch St.	Foggyville	MN	55802
4	Wilhelm S Carson	4839 Seasame Ln.	Greeley	SC	29340
4	Frank J Perkins	8438 Roberts Rd.	Diffley	KS	58239
4	Frank U Hoover	828 Mycroft Ln.	Stapleton	AK	99302
4	Winifred H Morris	2828 Eilert Way	Stapleton	MN	55802
4	John K Winstrom	2843 32nd Ave.	Crawley	IN	46601
4	Terry S Ali	8293 Becker Way	Karlstown	MI	48226
4	Stanley F Meunsch	2838 49th St.	Stapleton	HI	99783
4	Walter B Quince	135 Eilert Way	Stapleton	OK	59304
4	Inga V Morris	828 Waverly St.	Appleford	GA	30308
4	Uma Z Von Stubben	2843 Baker Ave.	Zeller	SC	29340
4	Owen N McCloud	283 Seaside Ln.	Venice	NJ	09483
4	Peter S Winstrom	9483 Orchid Dr.	Watertown	NC	27412
4	Cassie I Von Brocken	2939 Becker Way	Diffley	IN	46601
4	Ellen D Hu	283 Mycroft Ln.	Stapleton	ND	56293
4	Xia O Popovich	8282 Small Street	Venice	WY	82070
4	Uma K Meunsch	8282 Anchor St.	Quincy	MA	02139
4	Yvonne K Trott	2843 5th Ave.	Watertown	MD	21201
4	Cory Y Anthony	823 Orchid Dr.	Templeton	IA	56279
4	Greg I Wong	3489 Fender Ave.	Jonesbourgh	ND	56293
4	Cory O Steeger	4839 Elm St.	Zeller	FL	32804

13. One issue remains. The Marketing Department only wants to mail to households that are predicted to have no children at home. All of the others should be eliminated from the list. Select Design from the View drop-down list on the Mining Model Prediction tab toolbar to return to the design view.

14. In the first row under the Criteria/Argument column, enter **= 0**.
15. We do not have to see the predictable value in our result set, especially now that it will be 0 for every row. Uncheck the check box in the first row of the Show column. The Mining Model Prediction tab appears as shown below.

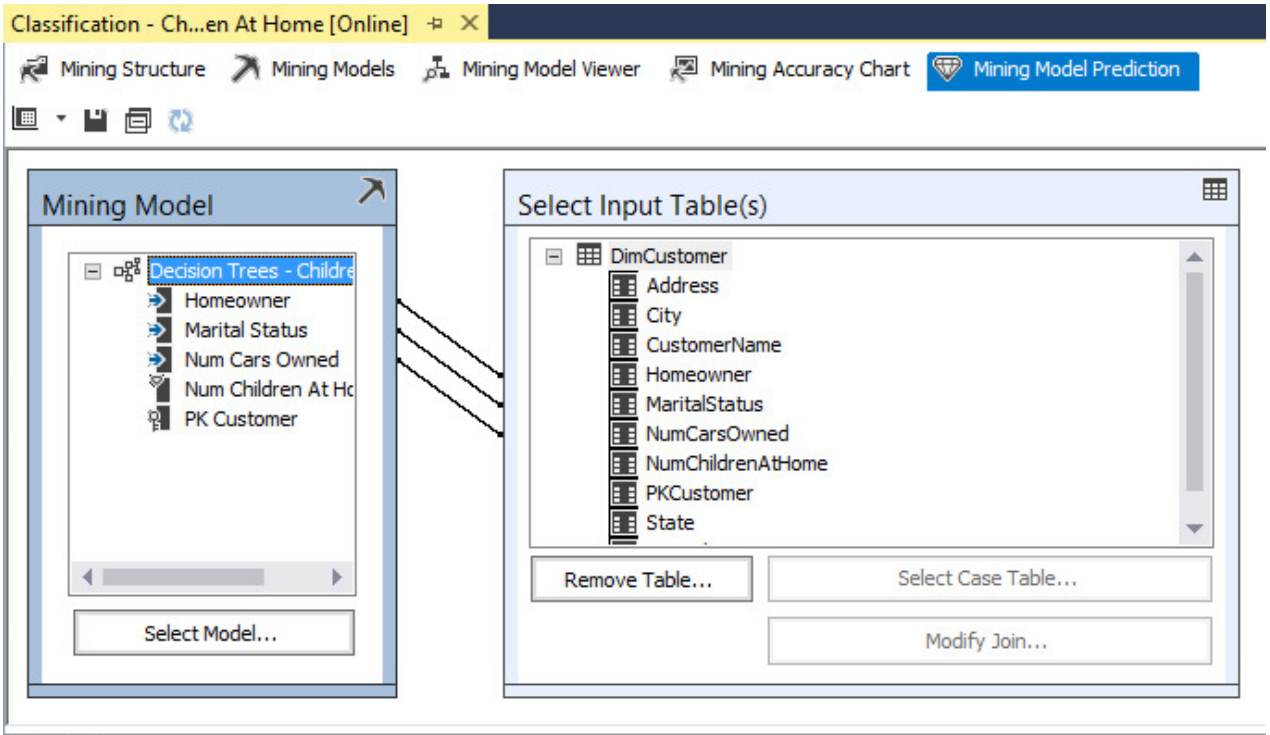


16. Select Result from the View drop-down list on the Mining Model Prediction tab toolbar. Now only those households predicted to have no children at home are included in our result set as shown below.

CustomerName	Address	City	State	ZipCode
Thomas R Ali	2839 Park Ct.	Rowley	OR	97205
Fritz N McCloud	283 58th Ave.	Newton	CT	06032
Wilhelm V Trott	3827 23rd St.	Karlstown	CO	80203
Valerie N Winstrom	8438 23rd St.	Yankton	NE	57203
Valerie F Winstrom	2838 23rd St.	Appleford	GA	30308
Xia W Petrov	2823 28th St.	Jonesbourgh	CO	80203
Darla D Fulston	2843 Elm St.	Foggyville	MO	63301
Zeb X Carson	823 Angle St.	Zeller	WV	34898
Gary J Drake	8293 Small Street	Edgeview	MA	02139
Greg Z Popovich	2838 Filborn Ave.	Xavier	AR	62312
Ralph S Meunsch	29838 32nd Ave.	Big Fork	CA	94109
Fritz Z Wong	2823 Overland Dr.	Crawley	NM	88044
Lia Z Petrov	3948 Filborn Ave.	Jonesbourgh	TX	75201
Melissa C Unger	14322 32nd Ave.	Jonesbourgh	IN	46601
Petra Y McCloud	8293 49th St.	Hillburg	CT	06032
Wilhelm D Unger	828 50th St.	Ivers	OR	97205
Frank W Anthony	283 Mayfair Ln.	Foggyville	IL	60610
Winifred M Perkins	2772 Mortimer Ct.	Newton	WY	82070
Uma C Taylor	283 Becker Way	Orville	ND	56293
Carlos P Carson	283 Mycroft Ln.	Rowley	MT	59715
Inga H Elert	3288 Quay St.	Quincy	KS	58239
Vern C Zipperer	9293 Drummin Lane	Yankton	GA	30308
Quincy D Hoover	2833 Hwy 45	Edgeview	NV	89701
Gary V Frank	823 Upton Way	Martinsville	VT	04394
Stanley S Smith	2838 Polar Ave	Diffley	NY	10007
Paul R Winstrom	2823 58th Ave.	Greeley	PA	19107
Melissa H Olsen	2839 Mycroft Ln.	Karlstown	IA	56279
Stanley R Kline	2839 Polar Ave	Edgeview	NE	57203
Edward O Petrov	3827 Birch St.	Ivers	TN	38103
Anders W Fulton	823 129th St.	Watertown	MD	21201
Edward R Roosevelt	8232 Poplar St.	Foggyville	VT	04394
Gary D Unger	2983 73rd Ave.	Zeller	MN	55802
Owen K Frank	283 Highview Blvd.	Jonesbourgh	CT	06032
Edward V Taylor	27391 Waymore Blvd.	Ivers	ME	03904
Vern E Frank	283 Mycroft Ln.	Venice	TN	38103
Fran N Trott	2833 34th St.	Edgeview	MT	59715
Winifred N Wellington	2843 Seaside Ln.	Stapleton	NM	88044

17. Right-click anywhere on the result set.
18. Select Copy All from the context menu.
19. Paste the results to an Excel workbook.

20. Save the Excel workbook to your class folder for CIS 570 (i.e., “S: drive”). Name the file, **YourLastnameDM2Output** (e.g., SmithDM2Output).
21. Back in VS 2017, click Query from the View drop-down list on the Mining Model Prediction tab toolbar. The query appears as shown below. This is the Data Mining Extensions (DMX) query you created using the graphical tools.




```
SELECT
  t.[CustomerName],
  t.[Address],
  t.[City],
  t.[State],
  t.[ZipCode]
From
  [Decision Trees - Children At Home]
PREDICTION JOIN
  OPENQUERY([Max Min Mining DM],
    'SELECT
      [CustomerName],
      [Address],
      [City],
      [State],
      [ZipCode],
      [Homeowner],
      [MaritalStatus],
      [NumCarsOwned]
    FROM
      [dbo].[DimCustomer]
    ') AS t
ON
  [Decision Trees - Children At Home].[Homeowner] = t.[Homeowner] AND
  [Decision Trees - Children At Home].[Marital Status] = t.[MaritalStatus] AND
  [Decision Trees - Children At Home].[Num Cars Owned] = t.[NumCarsOwned]
WHERE
  [Decision Trees - Children At Home].[Num Children At Home] =0
```

22. Close Visual Studio 2017.