

Microsoft SQL Server 2008 Data Mining Addins for Microsoft Office 2010

October 2013

Data Mining Client for Excel 2010

The Data Mining Client addin enables you to go through the full data mining lifecycle within Excel by using your spreadsheet data. If you have the Data Mining Client addin installed, you should see the Data Mining ribbon when you launch MS Office Excel 2010 in your desktop. If you don't have Data Mining Client addin installed, you can download it for free and install it from Microsoft's website at: http://www.microsoft.com/downloads/details.aspx?FamilyId=896A493A2502479594AE-E00632BA6DE7&displaylang=en The addins installation wizard will walk you through the process.

The ribbon has Data Preparation, Data Modeling, Accuracy and Validation, Model Usage and Management tools, as shown below. In this document, we will go through all the tools in the Data Mining ribbon and see their functionality.



Now, before you can use these Data Mining tools, you need a connection to Analysis Services server. To connect to Analysis Services server, click on the No Connection button (shown above) in the ribbon under Data Mining. Analysis Services Connections screen comes up.

Analysis Services Connections	C -
Analysis Services Connections	
	Nø <u>w</u>
	Delete
	Make Corrent
Open a new connection by clicking the New button.	Test Connection
	Close

Open a new connection by clicking the New... button. Enter Server name as <u>http://ent-asrs.waltoncollege.uark.edu/olap/msmdpump.dll</u> (http connection to the Analysis Services server), select a specific user name and password as Log on credentials. Enter the credentials supplied to you by the University of Arkansas – (WALTON\ES##### and password. <u>NOTE: If you are using Remote Desktop, simply select "Use Windows Authentication</u>") and check the check box for Allow saving password. Then, select DMAddinsDB database from the dropdown list of Catalog names.

NOTE: The database named DMAddinsDB is created in the process of 'preparing' the server. DMAddinsDB database acts as a container for the mining models created by the Addins. It is a shared database that is supposed to be there to hold data **temporarily** while users connect to Analysis Services for the tools from Excel.

	https://pump.waitoncollege.uark.edu/olap/msmdpun
Log on credentials:	
O Use Windows Authen	tication
Use a specific user na	ame and password
User game:	walton/es90000
Eastword	
🔲 Blank passwor	d Allow gaving password
Catalog name:	DMAddinsDB
Eliandia name	(default) Analysis Services Project Dillards
Densy date.	OMAddinisDB DOLLGLAS
	ES90000AS UA_SAMSCLUB_STAR
Test Connection	OX Canc
Lest Connection	DK Canc
Lest Connection	UK Cano
Lest Connection	DK Cano
Lest Connection Inelysis Services Connections Relysis Services Connections Current Connection	DK Canc tions Neg torcolege.uark.edu)
Lest Connection Inalysis Services Connections Palysis Services Connections Current Connection	DK Cancellions

Then, click Test Connection to make sure a connection exists and click OK. Close the last screen that shows your Current Connection. Once the connection is made, you're all set to start using the Data Mining tools.

Current Cor	nnection	New
🖳 (detault)	(http:/pump.walkoncollege.uaik.edu)	Delete
		Make Current
	Success 🔀	Test Connection
	V Test Connection succeeded.	

As a side note, once you make the connection, it will remain in your excel 2010 settings until you Delete it. In other words, next time you open an excel workbook, you will see the connection already made in the Connection section of the Data Mining ribbon in Excel (shown on the left). You just need to click the Connection and click Test Connection to ensure it is still valid. If not valid, click New... to establish a connection again.



Data Preparation

Selecting the right attributes from the source data and getting them into the right format for mining typically takes up a large percentage of the time in a typical data mining project. This section provides tools that address common data preparation needs for data mining: explore data, clean data and partition data.

Explore Data Visually plots the distribution of discrete and continuous values and possibly add groupings back to the source data. It is designed to help users identify and resolve imperfections in a data set. The Explore Data tool allows view histograms of numeric and nonnumeric data in your worksheet and group numeric data into equal size buckets. First you will need to load the data to excel workbook. This tool works only on data present in Excel and not on external data. To demonstrate how this tool works, I have imported the Customers table in the MovieClick database in MSEnterprise server to Excel workbook (refer to page 28 of this document on how to import data from SQL server to Excel workbook).

Click Explore Data to get started. Click Next in the getting started page.

P B ((((((((((((((((((T 20	N	TH 6	27 27 27	NK K	E. K	(2)		-	0 53.		
	AT O	-12			3 7				1.	1 88		
nor Calen handle Calent	a halanaha chada	er mintender	Fortari Adapter	that dates from the	CIGHE - JEERS	He Document Con	tio tibilage tibilage	Allow passes well	CORDACE. LICTURIDE (RELARD-REÉ	TEALE HOUSE		
Data Presentation	1748	le the second	ι.	Annaly and history		blacker trappe	Management		Detaitor	Him	A	
44 - 2	A.		¥			A real of the second se						
	0.00	Dis. D	Contract to 11	E	100 100 110	In Succession	CONTRACTOR	Thomas	1	With the last		- William
ALC . HUDDLATION	-	RANGER P	AN THE MENT	MN MARIANIAI	ih 🖃 Menmin	ALL PROPERTY AND	ALL THE PLANE	WINTE - Made	t sur 💌 benntse	LUMAN - Destr	VI DE PENNELO	-
47 Doctmate	Sister 1011	Own	Country Phil South Country	Married Com	Sciciarey	Partner	2	2	10012-000100	and the second second	2 Tenar	Mon
#T Doctarate	of produced	inter-		and the second second second	101	artnar	3	1	-1	1	1 Never	Non
47 Doctanale	and the state of	and the			1 in the second second	artras	7		1	-1	1 fearmer	Mon
at Doctorate	1		a Cation G	adad with the Evolution	Dista	petrier.	- 1	- 2	1.	1	1. Newsr	Mon
47.Doctanate	1 V.		+ Generg of	aneu wer die Espion	r Data	utnes		1	1		1 Same	More
67 Dectorate			anisting			stret	3	2	3	1	1 Never	Mon
47 Doctoreta		22				attar	1940	3	1		1 Second	Mart
all Doctorete						MIGHT	1	+	+		1. News	2000
AT Doctanate		1	What is not			etter		1	1		1 Second	Alexe
67 Doctorate		JA.	The Barrier O	an Weinel Weine gez Senne besie dat	aline	atter	3	1	1	1	1. tanuar	Mor
47 Doctorate			Education and	to the late in an antief lable or price	on anda	atteat	1	1	1	1	1 Never	Mer
at pochasing		9	What States I	a De?	Concert I	atteat	7	7	14		T Bellevelar	840
aT Doctatate			The stand etc	SCHI you to show a reason of a	ari barat -	attend .		1	-	1	1 Navar	Mor
ay Doctatate		100	- Z.			atzer	3	2	31		1. Same	Mor
AT Doctorete						atras	STO	1	3		2 Name	100
- AT Dochesida	4					atom	2	-	2		1. News	Alter
dif Doctorate	-					utrac.		1	1		1 barrent	allor.
all footbashe						articles.					1 Dave	Mor
all functionality						attains.	1	4			1 Danier	Aller
ST Toutsinging						steat		-			1 Name	100
all footballe			Philipped dates	For well-some proget science.		article.		-			1 March	100
all footsists						and the second				1	1 hanne	4100
all forstands	-					a codi		-	100		1 highlight	- 22
17 Decisional				And Sector	Caved	a to see			-	1		-99.0
al location	- Martin	Cause	.045	NAMES OF CASE	Sand State	arrige.		-		-	1.148.040	- 1990
AT BOOMER	Added	Control .	and a second sec	hard and the second sec	2224141	National Contractor		5		1	1 Distant	440.0
ATT Desetherable	- Adata	Church	MPN .	history and	Entering and	Contrast.	NT ALL	-			1	- 12
IST DOLDARD	and a second	Carr	100	00.01100.0	appender)	Production of the second se			-		A CONTRACT	1000
AT Decision	Male	Com	0.5	Married	2001/181	Suites:	-		-	-	1 Never	1000
and the second second	10100	Contract of Contract	100	Marked P	approxime taxage	Magencar.	-			4	1 10 10 10	4.4
all Incoherente	distan.	(internet)	04	Marrian	Service .	Martin as	100			-	1 March	-75
TT Destands	Adata	Com	05	hiterature .	East of the local day	Continues.	-	1		1	1 Manual	4.0
AT Doctoriana	South C	Course -	ine .	Married	200 LIN	Sattat		-	-	-	1 Manual	1.0-
art Portlander	didada.	Com.	110	himse	April 10	Margaren .		+	4	-	1.11000	1.000
all Destaults	and an	Control .	0.90	bewrine a	2004180	North and	324	-			1.148.000	- 640
al Doctorate	Adda a	Anno.	114	Number of	220.019	Constant.		1			3. 768 9.00	- 4900
art Decomana	Number of Street	ANNT:	100.	NUMTER'	SOCIETY:	Contrast.	-	-	-	-	1. Treasure	
et pocherate	And Add	Court.	05	Martine 2	appulley hadroid	the states	1.4	-	-	+	1 heart	1000
AT DATES A	- Topon	Consti	USL:	PARTIES .	2271110	Cartonia Contract		-			1. Transfer	
A MI Month Street	Winter .	Contra	ALM.	NUM TO BE	anioite/	Largonia			The later		1.110.00	7907

Select either the Table or Data range and click Next. Then, Select Column page comes up. You can select any numeric or nonnumeric column. In this example, select the nonnumeric Education Level column as shown in the screen below and Click Next to see its histogram.

efect Source Data		Q	Selecti	Column					(
n Table) Dete varge:	Seet11Table_painteptes_MeeeCick_Cultures*	-	Select Ortan	column woles	H	EDUCATION	-		1
	(21.0b) down rutings has frequence		TA.	HEDUCATION	SENDER	RENTOWN	INTERNETCONN	MARITALSTATUS	1
			47	0main	Male	Own	IDSL.	Married	1
			47	Distingle	Male	Own	OSL.	Married	
			47	Octobe	Male	Own	051.	Maried	
			47	Destants	Male	Quera	OSL.	Manuel	
			47	Gorrosale.	Mule	Own .	DSL.	Harred	
			47	Costruiste	Male	Own	OSL.	Married	
			47	Distruste	Male	Own	OSL.	Married	
			47	Coctoore.	Nale	Öven	061.	Married	
			47	Doctolate	Male	Own	OSL	Married	
			0		1				



At this point, you can click to copy a bitmap of the chart to a clipboard. You can also click Finish to close the wizard or click Back to explore other columns. Let's click Back to explore a numeric column.

ielect (Column						ç
Select	oline)		Num Ti	1			•
1	Wan C.	Ner Detr.	Sen L.	PPy Free	Prevec Buying Free	Press Format	+
1		0	2	Rarely	Blonttly	DVD	10
2		1	in -	Never	Monthly	DV0	1
2		2		Never	Revely	010	8
2		2	2	Harety	Barely	DVD	1
2		2	12.	Nevei	Randy	0/0	E.
2		0	m	Ranty	Ranniy	040	4
2		2	13	Nerver	Never	VHS	К.
2		0		Ramly	Rarely	0V0	E.
2		2	11	Neve	Monthly	DVD	£
4			-	18		1.0	61

Now, let's explore the Num of TVs column. Here, on the left, a numeric column (Num of TVs) is selected.



When viewing numeric data, the Explore Data wizard shows the data grouped to buckets of equal ranges.

Note that in the screen on the left, the number of Buckets selected is 8. You can increase or decrease that number by changing the number in the buckets control. You can also toggle between View as Discrete and View as Numeric to change how the data is viewed.





Notice that 'View as Numeric' is selected in the first screen, and 'View as Discrete' is selected in the second screen above. At this point, as explained above, you can click the copy button to copy a

bitmap of the chart to a clipboard, click Finish to close the wizard or click Back to explore other columns. In addition, if you want to use the bucketed data as a column in your worksheet for further analysis, you can click the Add New Column button which inserts a column of the appropriate ranges to your source data.

Clean Data Removes outliers and relabels discrete state values. For example, your source data may contain "Male" and "Female" for the Gender column and you prefer to use "M" and "F" when presenting the model results. For this example, let's use the MovieClickCustomers table from MSEnterprise we used for the above example. To start the relabel tool, click the Clean Data button and select Relabel.

(Che.)	10	0	Ŧ	Main	ecute M	bostatt Inc	-		1	able Roots					
9	Hura	Juant	Page Layout	Parmates	Data	Restore	View	Cala Mirson	ų	Amalyzz	Design				
	1	R.	子子	😵 💥		5	Ø			瓜	6	R	×	P	EX,
Data	Dota *	Data	Charley Estimate	Curret Associat	e morecase	ADMINIST	Ourt	Matrix	Orart	DIDWIE	Query	Models	(Mpenterprine)	Trace	THEP
Ret		Outlins		Data Mudeles	ł	-	A(5)	ind and Merry	thes	Hitettel	Chage .	Management	Contemin		Party
🥛 See	300		onnettions hav	e been duabled	Optia	й)									
-	1	Relates	5	2											

The getting started with the Relabel data wizard screen (not shown here) comes up; click Next. Select the data source and the column in the next screens.

~ Related		Rela	et.				= 3	-
Select Source Data	I	Select	Colores				0]
· Jatie O Den range	Steet 11 Table_resenterprise_MonierGick_Cultureer •	Selec Deter	colums: langles	6	ender			•
	12/18 data spractice backless	EA.	Education Level	Gender	Home Ownership	Internet Connection	Raital Sub	4.4
	Added to The second	13	Doctorate	Vale	Öwn	Dial-Up	Married	
		47	Declarate	Mater	Own	DSL.	Marred	499
		100	Doctorate	Net	Qum	DSL.	Hannel	
		35	Bachelar's Degree	144	Own	Cable Modern	Maned	10
		32	Bechelor's Degree	Name	Own	Cable Modern	Maned	
		12	Bachelor's Degree	Max.	Own	No Internet Connection	Harred	
		22	Master's Degree	Mar	Own	031	Married	
		12	Bachelor's Degree	Note:	Revit	DSI,	Maried .	
		33	Master's Degree	-	Own	Cable Modern	Harried	4
		125	1.141				1	
	s Back Next > Carcol				•B	eck Fied >	Cano	ei

You will see the original labels of your data for the column selected and you enter the new labels.



Select your destination and you will see your column relabeled as shown below.

Lete Pissonitum	Oral Maine Dett Oral Maine Dett Oral Maine Dett	any bianagement ge bianagement	۲	Security Warning Data c	annections
Security Warring Buts Coronal	- Tables (•
- 10 (5)	Referet Destination	*		A B C	D
AN ANDRESS	Desired Constrained			Education Gend	e liome
A B C		J	1	Agentevel t	• Daniel
Education careful in	Sharib share to not the motilized date	the second li	2	33 Doctorate M	Own
32 Contorate Male O	dant was the second of the second of the	Contraction of the second	3	47 Doctorate M	Own
d7 Dectorate Male D	C dat as a new column to the current worksheet	Marian	4	Doctorate M	Own
Doctorate Male D	a management of the second	I Never	5	35 Bachelor's M	Own
15 Bathelior's Male O	E Copy sheet data with changes to a new worksheet	D Bareh	6	12 Bachelor's M	Own
S2 Bachelor's Make		a Never	1	12 Bachalor's M	Own
12 Bachelor's Male O	Orange gate in place	Barely	-	12 Master's DM	Channe
32 Master's O Male O		Never	1	an master s this	- Charte
12 Bathelor's Male Br		 Banely 	A	32 sachelor 1 M	Rem
33 Master's O Male O		Never	10	33 Master's DM	Que:
-84 Master's D Male D		Pérner	11	44 Master's D M	Clem
38 Matter's D.Male D		B Haroby	12	39 Master's D M	Qwn
29 Maiber's D'Male fe		Never	13	29 Master's DM	Rent
41 Doctorate Male D		Neve	14	41 Doctorate M	Own
44 Doctorate Female O		Never	15	44 Doctorate F	Own
AN MAIDER'S CIMARE C		A Never	16	59 Master's D M	Own
AS Decrement's Make O	e Sack Brief Cercait	D Never	17	45 Bachelor's M	Own
22 Decision S mane L		- manual	100	the later of the second	the state of

The Outliers tool detects and cleans up outliers from your data. To start the remove Outliers tool, click the Clean Data button, select Outliers and follow the instructions on the screen.

Partition Data Splits the source data into training and test sets, takes a random sample of the input data or perform oversampling to adjust for skewed distributions.

Data Modeling

This section covers the actual model definition and processing. It provides wizards that help you easily build common types of mining models without worrying about the actual mining algorithms and associated parameters supported on the server. Also included in this section are advanced options that allow the user to pick the exact mining algorithm and tweak additional parameters. To create a mining model, click Create Mining Structure under Advanced in the Data Modeling section of the Data Mining ribbon.

section of the Data Mining ribbon.

E a vier	Fige Learner	Terrater 1	tata ile	new Viter Data (Kong	Ultrace 3: mign Deign	ala - Manaafi E	itel		
Datas Data Datas	Ethner Oute	e filologija		Accused Designations Brefs (Sati - Martin	Nerse Marsage Marsage Marsage	E2000045 Piter punto existencinge a Connectinge	at ata	X
B A	а Э	D	-	Create names structure with no must	es d	M	1 1 2		E
1 ACC - ELEBORATION	M	Own	DE:	Stationer.DMCbardtLAddb- Press Pt Dis add 20 Ordu	Spouse/Partner	anneanta <u>s</u> inna 2		RECORDINEN	E NUMBER
3 47 Doctorala	M	-Own	05.	Manied	1pouse/Partner	2	2 2		1

1

Click the Next button in the welcome screen (not shown here). We will use the churn table for this example. The churn table is imported to excel from SQL Server 2008. Refer to steps on Import Data on Page 5. Select PUBLIC_DATASETS_DM database and Churn table when asked for the database and table.

Wizard					?
se and 1 ise and Ta	able ble/Cube whic	h contains t	he data you want.		K
that conta	ins the data y	ou want:			
DM		*			
cific table:	e!				
Owner	Description	Modified	Created	Туре	^
dbo			10/5/2008 9:45:57 PM	VIEW	
dbo			10/5/2008 10:30:18 Pf	WIEW	
dbo			10/24/2007 3:23:00 Pf	A TABLE	
dbo			3/30/2008 7:36:24 PM	TABLE	
doo			10/22/2007 10:06:15/	AM TABLE	
dbo			10/22/2007 10:06:15/	AM TABLE	
dbo			10/24/2007 3:23:00 Pf	M TABLE	
2200			10/04/0007 0:00:00	A TADLE	1000
	Wizard e and Ta se and Ta that conta DM offic table: Owner dbo dbo dbo dbo dbo dbo dbo dbo dbo	Wizard e and Table se and Table/Cube which that contains the data y DM offic table: Owner Description dbo dbo dbo dbo dbo dbo dbo dbo dbo	Wizard e and Table se and Table/Cube which contains to that contains the data you want: DM v offic table: Owner Description Modified dbo dbo dbo dbo dbo dbo dbo dbo	Wizard e and Table se and Table/Cube which contains the data you want. that contains the data you want: DM Control Description Modified Created dbo 10/5/2008 9:45:57 PM dbo 10/5/2008 10:30:18 Pf dbo 10/24/2007 3:23:00 Pf dbo 10/22/2007 10:06:15 / db0	Wizard e and Table se and Table/Cube which contains the data you want. Image: Cube which contains the data you want. Image: Cube which contains the data you want. Chat contains the data you want: Image: Cube which contains the data you want. Owner Description Modified Created Type dbo 10/5/2008 9:45:57 PM VIEW dbo 10/5/2008 10:30:18 PM VIEW dbo 10/24/2007 3:23:00 PM TABLE dbo 10/22/2007 10:06:15 AM TABLE

Select the table with the data to be create a Mining Structure in the Select Source Data page and click the Next button. Use the churn table.

		Hann brint Pi	ugn Layout, Prinniana Dieta	Renter View Data Mining Analyze Da	tige				
On the next	and the	the Dear September	abort Quay Icani Illinga	Marage Montpetter	DHAOSING DSDDO Corres	9-544 1 0-11			
	۲	Security Warning Data men	uemore frame parts disabled	tern					
		617 - 3	A N						
		A	C Property and the second		-	Carrows	2 martine	K.	Stowald
	1	Recentile 📲 State 🖬 Acco	munt i	e Witcard		tins of Day (alls 🕶 Day (harge 🖪 Ev	e Mins 💌 Fi
	-2	141 M0	Select Source Date		12	161.5	52	27.46	173.5
	3	642 \04				165.1	120	28.5	198.5
	4	643 K		The second se		195	92	33.15	210.5
	1	644 OK	1 ² Table	Sheet117able_papps_PUBLIC_DATASETS_DM_Chain'	*	213.8	79	36.35	265.1
	6	545 00		Communication of the second second second	and the	205.5	114	34.34	215.7
	17	846 MT	T Data congr			352.9	112	42.99	13728
		647 WV		🖉 Malata projektos halatare		225.6	74	40.01	227.9
	*	640 ME				192	91	32.64	127.6
	49	4.00 101	External data and be			69.4	122	11.78	101.2
	11	650 MO	Dide mare sent		124	261.7	97	44.49	210.6
	12	651 NY			-	295.5	£1	40.04	257.2
	11	#52 CT	Guarge:		10	211.4	100	36.28	204.9
	14	653.AR				206.9	543	35.17	127.6
	15	654 WY				363.0	-65	44.05	185,4
	15	655 CA				181.4	126	21.14	195.5
	17	656 CA				\$57.6	15	26.78	194,1
	10	457 MPE				175.6	147	29.05	141.6
	19	1658 AK				242.5	101	41.23	192.8
	20	\$50 WA				151	102	25.67	131.4
	71	660 00		i i i i i i i i i i i i i i i i i i i	1.11	138.1	103	23,48	1645
	22	561 MT		c Back Next > Can	CH	264.7	69	41	305
	Contract of	1000 000	and the second s		1.000	1000 10	1944	100.000	1.0.0

screen, select which columns to use and click Next. Mark RecordID as Key and do not use columns related to Charge, State, Area Code and Phone. From exploratory data analysis (not shown here), it was determined that the variables (columns) of **State**, **Area Code** and **Phone** contained bad data. Also, all the columns related to Charge were perfectly correlated to the corresponding Mins (Minutes) column. So, none of the Charge columns will be used in the analysis. Mark all these columns as Do not use. Mark the remaining columns as Include. Click Next after selections are made.

Table Column	Usage		18	-
Eve Mins	Include	-	-	
Eve Cals	Include	-1	-	
Eve Dhage	Dia not use	*		
Night Mins	Include	*	-	
Night Calls	Include	+	-	
Night Charge	Do not use	-		
nt Mins	Include	-	-	
nt Cals	Include	+	_	
nil Charge	Do not use	+	-	
DustServ Calls	Include	-		
Dhum	Include	*	1	

On the next screen you are given the option of selecting what percentage of the data should be used for testing. Keep the default settings and click Next.

of the entry in the interview and the time set	21		1
in one internet and the stary set			12
Sercentage of data to testing		20	-
Laximum number of rows:		1	-1
Oexception			-
rounder or round for best man the Preci-	emage of data for letting.	roess' to 0, this limit will not b	RI I
used The partitions canaled by the viewed and testing sets. Because the partition is not affected.	n die Ingenitation namen of nie and cardina and die your name	versus representative loss ng structure, your source de	ng ta
used The partitions cansised by the viscent , and being sett. Receive the partitio is not affected.	we conden, which happe to	erone representative have	nd ta

Click Finish on the next screen to create the Mining structure.

Create Mining Structure W	Azard 📃 🖬 🖬
inish	*
Structure name:	Lable repor FUBLIC DATASETS DM. Chum Stucture
Structure description:	Structure for the Table_zeppo_PUBLIC_DATASETS_DM_Churn table
Options	
C Use temporary structure	0(1
	(Park Dirich Caucal
	C gach Lawren Cancel

To add a mining model to the structure that was previously created, select the Add Model to Structure option under Advanced in the Data Modeling section of the Data Mining ribbon.



Select the structure which was created above as shown below and click Next.



Microsoft Trees from dropdown Mining and click

	structure mizaru		
elect MiningAl	goiltm		
Select the algorit parameters.	hm you want to use to build the m	ning model, and optionally set th	he algorithm
Algorithm:	Microsoft Decision Trees		*
Description			

On the next screen, select the type of usage for the columns that were included. Mark RecordID as Key and Churn? column as Predict Only. Notice that the columns Charge, State, Area Code and Phone which were marked as "Do not use" no longer appear on the select columns screen. Leave the remaining columns as Input and click Next.

elect Column					ł
90 B					
Table Column		Usage			
Account Length		input			1
Chum_	1	Predict only		-	
CustServ Calls		Input			
Day Calls		Input		-	1
Day Mins		Input	-		
Eve Calls		Input			
Eve Mina		Input	*		
Int_I Plan		Input	-		1
Init Calls		Input	+	-+++	2
Ind Mina		Input		-	
Night Calls		Inpút.	-	-+1 -	
Night Mins		Input	<u>.</u>		
			_		

Click Finish on the next screen to add the model to the mining structure. Optionally, you can give the model any name or leave the default name for the model.

inish	
Shuchure name:	Table_mannepris_FUBLIC_DATASET6_DM_Chum Times
Structure gescription	Structure for the Table_menterprise_PUBLIC_DATASETS_DM_Cham
Model name:	Table membronie_PUBLIC_DATASETS_DM_Chum - Tree
Mgdel description:	Microsoft_Decision_Trees model built in the Table_msenterprise_PUBLIC_DATASETS_DM_Chum
Options	
Browse model	💽 Enable drillthrough
Use temporary model	Process model
	Rack Finish Cancel

The Decision Tree Page comes up. You can copy the diagram to Excel by clicking the Copy to Excel button in the left bottom section of the screen.



You can also add another model to existing mining structure by clicking Add Model to Structure, under under the Advanced toolbar.

2	1	ite al	•		Harry - Mry	ou ion to	ad			taske forola		1			
9	PEDAG	Sviert	Page Layon	t. Pormulais	Deta	Enter	e Vee	Data Mini	10	Anabas	Denge.				
Durore Usta	Ores Data *	Canple Cats	A Æ	Curter Asiatish	Ferrecart	(Salara	Activity Obst	Cisuationical Materia	Prom Prom Chart	Croil - Validation	S. Demain	Dooment Que	v Maringe Modeu		E.
Data	ority Wa	then la	ta consection) h	California and an	Ogts	3	Create Minin	g Structure	rinder	len	1	da dal Ukaga	Monagement	Cornection	((the)
	617		.6 1	N	(Included)	-	Add Model b	e Structure	C						

the Advanced toolbar.

Accuracy and Validation

This section includes charts for validating and verifying the accuracy of the mining models. The three charts provided are: Accuracy Chart, Classification Matrix and Profit Charge.

Accuracy Chart – Evaluates the performance of the model against test data by drawing a lift chart for classification models and a scatter plot for estimation models. To see the Accuracy Chart, click the Accuracy Chart in the Accuracy and Validation Section of the Data Mining ribbon.

9	Have	lisert	Page Lapsur	tamitis	Bits Bit	iew Mew	Data Mini	10	Arialgok	Design			
torine	Clean 1	Cample Sample	A A	😵 💥	Porcast Advin	and Acores	Charletter	Pott	Churs-	Downe Document Query	Manage		Rep.
Data .	Data +	Ditta			141161174	Chert	INMATRIX .	Chart.	Validation	Model	Modell	(04(p(p)))	7.40
Date	S Preparat	966		Bala Modeling			Accurbing land	ralitati	181	ModelUisge	Management	Connection	itets .

The getting Started page comes up. Click Next to select the existing model. Select the Trees model created above and click Next.



elect Structure or Model		
Structures and models	Description:	
Churn Churn	Model Name	Table_msenterprise_PUBLIC_DAT
Table_msenterprise_PUBLIC_DATASETS_C	Algorithm	Microsoft_Decision_Trees
Table_mserverprise_PUBLIC_DATASETS_DM	Description	Microsoft_Decision_Trees model t Table_msenterprise_PUBLIC_DAT Structure mining structure
	Output	
	Churn	-
	input	
	Accou CuatSe Day Ca Day M Eve Ca Eve M	nt Length erv Calls els ns its ns
¢ >	e	2

Specify Column to Predict (in this example, churn_) and Value to Predict (False) and click Next.

K Accuracy Chart	E C 🔛
Specify Column to Predict and Value to Predict	7
Mining column to predict Value to predict Description	False.
This task, analyzes the performance of the selected models in predicting the 'Churn,' column for the text data that you will choose in the nest page of this wicard. The task generates a chart report describing prediction accuracy for selected models on 'Churn,' This chart shows how the number of correct predictions grows with the number of cases analyzed by the models.	80% 60% 40% 20% 0% 0% 20% 40% 60% 80% 100%
	<back next=""> Cancel</back>

Select Source Data and click Next.

c necessary control		
Select Source Data		1
O Test data from godel		
⊙ Lable:	Sheet19Table_moenterprise_PUBLIC_DATASETS_DM_MovieClick	Y
O Data range:	[-
	🕑 My data range too beaders	
C External data source:		
Data gource name:		÷
Query.	1	-
		2
		el.

Accept the defaults in the Specify Relationship page and click Finish.

pecily Relationship		
specify the relationship between m	odel columns and input columns	
Mining Column	Table Column	
Account Length	Account Length	*
Chun,	Chum,	2
CurtServ Cells	CustServ Calt	-
Day Calls	Day Calls	<u>.</u>
Day Mina	Day Mine	*
Eve Cali	Eve Cala	
Eve Minu	Eve Mint	*
Int_1Plan	int_I Plan	*
Indi Calla	Intl Calls	
inil Mina	Intl Miran	*
Night Calls	Night Calls	*
Night Mins	Night Minz	1
VMail Mestage	VMail Message	*1
4.4.4.79	1+A+ + P+	- 14

The chart will be in a new tab called 'Accuracy Chart' as shown below.

Gaue Sandis Case Data - Data Propagation A1 +	n, Etherada Caster Annorado Fore Unite Mananting & Accuracy Chart 1	or Model Table moan	Caustonson Porta C Mates Chart Va Universited Voldstein Herpinae_Public_DAT	ASETS_DW_Churn - Trees	An Marage Mashi Manigtows	essoi interarmente	OLAS Accelerge wark, ecke Generalizen	Tierz erz erz	
	Accuracy Cha	rt for Model 'Tab	le_msenterpris	PUBLIC_DATASETS	DM Churn	- Trees'		-	1
		1	Predicted Culume 10	nm_"+Talié.		11 11	_		-
ible_menterprise_P	WBLIC_DAT 111.24%								
33 %.									
12 %		//	(Post Area)	Via Mazel 	PLOCE DATASETS,	,DM_OW			

Classification Matrix Displays a matrix of correct and incorrect classifications by evaluating your model against test data. To create Classification Matrix, click the Classification Matrix button in the Accuracy and Validation Section of the Data Mining ribbon.



Click Next in the getting Started page (not shown) and select the ...Trees model (the model created in the process before) from the Select Model screen and click Next.

Classification Matrix				
Specify Column to Predict				
Mining column to predict.	2	n_		2
Show moulds as gourt				
Description This task analyzes the performance of the selected models		À	a	e I
will choose in the next page of the wizard. The task	1	Counts of i	correct/inc	orrect clar
generates a matrix report, and displays the conect classifications service the mechanication and conect	8			
each model on 'Chun,'	9		0(Actual)	1[Artuəl]
The matrix shows, for each distinct state of the "Dham."	12	1	2499	6158
column, the number of occurrences in the test data as well	12			
each nodel.	13	Correct	6853	6158
Correct predictions appear on the main diagonal of the dissilication matrix. A non-zero value which is not on the main diagonal of the matrix represents one or more mechanism.	14	Misclassifi	2499	2974

Specify Column to predict (churn_) in the next page and click Next.

Choose the churn table in the Select Source Data wizard page and click Next.

Classification Matrix		- 🗆 🛛
Select Source Data		
O Test data from model		
⊙ Iable:	Sheet21'Table_monterpline_PUBLIC_DATASETS_DM_Dtur/	•
O Data range:		
C External data source Dele gource name		_ •
Query:		1
		9
	(Back Next>	Cancel

Accept the default values in the Specify Relationship page and click Finish.

pecily Relationship		
Specify the relationship between m	odel columns and input columns:	
Mining Column	Table Column	1.1
Account Length	Account Length	
Chum_	Chum_	*
CustServ Calls	CurtServ Calls	
Day Calls	Day Cale	1
Day Mins	Day Mins	~
Eve Cals	Eve Cala	~
Eve Mins	Eve Mins	8
Int_1 Plan	Int_I Plan	×
Inti Calls	Inti Celt:	*
Intl Mins	Inti Mina	3
Night Calls	Night Calls	*
Night Mins	Night Mins	
VMail Message	VMail Message	8

The Classification Matrix output will open a new tab in the same spreadsheet and will look like the screenshot below.

Z.	₩ 10 × 01 × 14		Back2 + Microsoft Estat	
	teren brunt Papeta	and Parmala Data Brenn Stan Datablioreg		
Distance in the second	The Desir Lander Charley Erweiter	Contra Assessing Frances Advances Data Assessing Frances Advances Data Moderny	Sarkan Manar Common Common Manager Man	olit Tuke Hep- tolege Lanit Atta
	41	Counts of correct/incoment classification for model Table_r	neerderprise_PUBLIC_DATASETS_DM_Churn-Trees'	
-	A	1	4	0 2 *
+	Counts of correct/incorr	ect classification for model 'Table msenterpris	e PUBLIC DATASETS DM Churn - Trees'	
1	Pendicted Column 'Cham."			
1	Columns correspond to actual vel	(Jan)		
	Raws correspond to predicted va	fver -		
\$	Para di constituti di Canada di	Wall	which which is preside to switch the set of the set of the set	
쓰	model same:	Table_Interterprise_PORDL_DATASETS_DW_Churt-Trees	table insemergence Polatic particults poli cham- trees	
	Total composition and the all	92.24.%	507	
	Total miscialisment:	1.00.76	<i>B</i> I	
-	Results as Percentages for Mode	Table menterprise PUBLIC DATASETS DM Chum Trees		
11		* False (Actual)	True (Actual) +	
J.	False.	JF \$4,29	22.77%	
13	True,	5.33%	77.23%	
10	Famela	22.47 M		
10	Minianified	94.87 %	77.25% 83.77%	
17				
28	Results as Counts for Model 'Tab	le musiefergrine PUBLIC DATAGETS DM Churn - Trees"		
11		* False (Actual) *	True (Actual) (*	
	False,	2698	110	
문	True,	153	- m	
44	Fanset	1404	171	
34	Misclassified	152	110	
25				
26				
27				
28				
47				
31				
32				
33				
34				
36				
17				
38				
18. 19. 1	r n. Smith Share? Amer	en Olar Classification Habra (194412) 12	1	w(
10	dille and the second second	and the second se		35

Profit Chart – Graphically models profit for targeted campaigns based on usersupplied parameters for cost of existing mining models.

Model Usage

This section covers the two standard tasks that you would perform with trained mining models: Browse and Query.

Browse – Explores the patterns and rules learned by the mining algorithm from the training data. This "mining model content" is visualized in different ways depending on the type of model you're

browsing. To see the functionality of this tool, click browse in the Model Usage section of the Data Mining ribbon and select the model you would like to browse.

9	Home	Insert	Page	Layout	Fau	mulas	Data	Review	View	Data Mining	a Al	byze	Design	
			80	Ł	00	3%		-		8	m	A	6	
Explore Data	Clean Data =	Data	Classify E	Ectimate	Outer	Associate	Forecast	Advanced	Chart.	Classification Matrix	Profit Chart	BEGANER	Query	Models
Dat	a Prepar	ation			Data	Modeling			Accus	racy and Valida	thitm	Model	Usage	Management

vlodel.	Properties:	
Table_zeppo_PUBLIC_DATASETS_DM_	Model Name	Table_peppo_PUBLIC_DATASET *
These stations and the rest of the state	Algorithm	Mcrosoff_Decision_Trees
	Description	Microsoft_Decision_Trees model Table_zeppo_PUBLIC_DATASET Structure mining structure
	Output	
	Churn	
	input	
	Accou CustSi Day Ci	nt Length erv Calls alts
	Eve Cr Eve M Ini J Pi Ini Cal	afo fot an fo
	Indi Min	a



Below is an example of the visualization of a decision tree model for the Trees model created before.

predictions on new data. The Query task supports a wizard for building simple queries as well as an advanced editor where you can use DMX templates to build queries or manually type in the DMX statement.

Management

This gives the ability to manage existing models in the Analysis Services database you are connected to. You can rename, delete, clear, reprocess, export, or import mining structures and models as shown below.



Shuctures and models	Lasks:	
Churn Churn Table_msenterprise_PUBLIC_DA Table_msenterprise_PUBLIC_DATA	 Rename this mining structure Delete this mining structure Clear this mining structure Process this mining structure with original data Process this mining structure with new data Export this mining structure Import 	
	Description: Structure: Churn Description: Models: Churn Columns: Account Length Churn Cust Serv Cata	

As we saw at the beginning of this document, the Connection button allows to create and manage connections to Analysis Services databases – a connection to an Analysis Services database is required before you can run any of the analysis tools described above.

Trace

The Trace button allows you to trace the commands that are sent by the data mining addin(s) to the Analysis Services instance you're currently connected to.

Help

The Help button provides access to the documentation included with the addins as well as the Getting Started wizard and online tutorials.

Sources include:

Microsoft's website sqlserverdatamining.com

Importing data to Excel from SQL Server 2008

To demonstrate how to import data to excel workbook from SQL Server 2008, we will import the churn table from Public_Datasets_DM database in SQL Server 2008. You will need to log in to ts-mec.waltoncollege.uark.edu with your WALTON credentials to be able to do this. On your computer use Remote Desktop Connection to connect to **ts-mec.waltoncollege.uark.edu** and click Options.



Click the Local Resources tab and click the More... button

Remote D	esktop Connection		
4	Remote Des Connecti	sktop i on	
General D	splay Local Resource	es Programs Experien	ce Advanced
Remote a.	idio Configure remote aud Settings	sio settings.	
Keyboard	Apply Windows igey	combinations:	
	Only when using the Example: ALT+TAB	tuli screen	*
Local devi	ces and resources		
-	Choose the devices your remote session.	and resources that you w	ant to use in
	Prințers	Cipboard	
0.000		Connard	Hab

In the Local devices and resources, expand Drives and check the check box for C: drive as shown in the screen below. Click OK.



Then, click Connect.

eneral D	splay Local Resource	es Programs	Experience	Advanced
evernote al	Configure remote au	udio settings.]		
Keyboard	Apply Windows key	combinations:		
0	Only when using th	e full screen		•
	Example: ALT+TAB			
local devi	ces and resources			
-	Choose the devices your remote session	and resources	that you want	to use in
	V Prințers	V q	pboard	
	More	1		

Click Connect.

•	Do you trust tras tem	OLG COMBRIDATION ?
his rem emote c	ote connection could ha omputer before you con	im your local or remote computer. Make sure that you trust the nect.
	Type:	Remote Desktop Connection
CO S	Remote computer:	ts-mec wattoncollege uark.edu
	Gateway server:	ent-sysgw.waitoncollege.uark.edu
Dont	ask me again for conne	ctions to this computer

Enter your credentials (Walton\Es##### and password provided by the University of Arkansas) and click OK.

ncollege.uark.edu (remote computer)
GACL\jzaman
walton\es90000
Domain: walton
nber my credentials

Click Yes.

	The identity of the remote computer cannot be verified. Do you want to connect anyway?
The rem security	note computer could not be authenticated due to problems with its certificate. It may be unsafe to proceed.
Certifi	cate name
-	Name in the certificate from the remote computer: TS-MEC1.uark.edu
Certifi	cate errors
The f	following errors were encountered while validating the remote suter's certificate:
A	The certificate is not from a trusted certifying authority.
Do you	want to connect despite these certificate errors?
Don	t ask me again for connections to this computer

Once you're logged in, open My Computer as shown in the screen below.

Command Prompt	Documents Computer Network Control Panel Devices and Printers Administrative Tools Help and Support Run
All Programs	Windows Security
Search programs and Ries	Log off

You will see YOUR C: drive connected to the ts-mec.waltoncollege.uark.edu server under Other drives. You can double click the drive to see the contents of YOUR C drive from REMOTE server.



Then, let's start the process of importing data from SQL server to Excel workbook. On the Remote server, click Start and

Open Excel.



Import Data

To import data from SQL server to Excel workbook, follow these steps. Go to the Data tab \rightarrow From Other Sources \rightarrow From SQL Server.

🗶 🖬 🕫 - 1	(*) (*								
File Ho	mr 1	Insert 1	Page Layout	Formulas	Data Re	view View	, Dat	a Mining	
From From Web	From Text	From Other Sources*	Existing Connections	Refresh All *	Connections Properties Edit Links	호나 <u>호류</u> 친나 Sort	Filter	To Const To Reausty Sy Advanced	Tex Colu
A1	Get Ed		rom SQL Server reate a connection vio Excel as a Tab	on to a SQE lie or PivotT	Server table. Imp able report.	port data	Hort & Fil	101	
A 2 3 4 5 6 7 8	8		rom Analysis Ser reate a connection oport data into E rom XML Data In open or map a Xh rom Data Conne nport data for an onnection Wizar onn Microsoft Q oport data for an usery Wizard and	vices on to a SQL ixtel as a Tain nport it. file into f ction Wizar i unlisted fo d and QLED upy t unlisted fo 1 ODBC.	Server Analysis 5 ble or PriotTable Escel. d ormat by using 19 DB. ormat by using 19	ienvices cube. report. ne Data ne Microsoft	H	1	1

Type in msenterprise.waltoncollege.uark.edu for Server Name. Enter your ES account credentials (**without** the Walton domain). Then click Next.

Data Connection Wi	zard	? ×
Connect to Data Enter the information	base Server tion required to connect to the database server.	
1. <u>S</u> erver name:	msenterprise.waltoncollege.uark.edu	
2. Log on creden © Use <u>W</u> ind © Use <u>t</u> he fo	tials ows Authentication Ilowing User Name and Password	
<u>U</u> ser Namo <u>P</u> assword	e: ES90000	
	Cancel < <u>B</u> ack <u>N</u> ext >	Finish

Select PUBLIC_DATASETS_DM from the drop down list and select MovieClick database. Click Next.

Select Datab	ase and base and 1	Table (able/Cube wh	ich contain	s the data you w	int.		×
elect the databa	se that con	tains the data	you want:				
PUBLIC_DATASET	IS_DM						
✓ Connect to a s	pecific tabl	e:					
Name	Owner	Description	Modified	Created		Туре	
HMEQ	dbo			7/25/2008 1:16:	11 PM	TABLE	
HMEQID\$	dbo			7/25/2008 1:23:	25 PM	TABLE	
HomeEQ	dbo			7/17/2008 4:35:	19 PM	TABLE	
🛄 Loan	dbo			10/24/2007 3:23	:00 PM	TABLE	
MovieClick	dbo			12/15/2009 3:12	::37 PM	TABLE	
Order 🔤	dbo			10/24/2007 3:23	:00 PM	TABLE	
Prospect	dbo			7/17/2008 4:40:	56 PM	TABLE	
Shopping	dbo			7/17/2008 4:27:	03 PM	TABLE	
			Carcel	C Back	Next	<u> </u>	Finish

Now click Finish.

Data Connection Wizard	? ×
Save Data Connection File and Finish Enter a name and description for your new Data Connection file, and press Finis save.	sh to
File Name:	
msenterprise PUBLIC_DATASETS_DM MovieClick.odc	Browse
Save password in file	
Friendly Name:	
msenterprise PUBLIC_DATASETS_DM MovieClick	_
Search Keywords	
2earch hoymonds.	
Always attempt to use this file to refresh data Excel Services: Authentication Settings	

Click OK to insert the data in existing worksheet.



Enter your password again, if prompted.

SQL Server Log	jin		
Server:	msenterprise	•	OK
Use Truster	d Connection		Cancel
Login ID:	E\$90000		Help
Password	moones		Options >>

Then, you will see the churn table data opened in your worksheet, as shown in the next screen.

		-										
	-A.	#	210-21C-	P	10 E 20	de 1 de	1 00	11	in the star	-425(0):	N.	LM
1	ADE -	HERRICATION	• GENINER •	RENT OWNER	MITTHEFEOWN	* MARTALITALIS *	MOVIESHEETON	Manaazing 🖛 n	HEADENS 💌 HI	186.621 2 1	ais(hillian 💌 🛛	THEAT BEACHTER
-	4	7 Dectorate	Adula;	Pan	05,	Married	Eposie/Partner	2	1	21	+	1 former
3	- 4	7 Doctorate	Mate	Own	C192.	Married	Spouse/Partner	. 2	2	2	1	1 Never
4	- 4	7 Doctorate	Mille	Own	DSL.	Married	lipouse/Partner	4	1	2	1	1 Newlet
5	- 4	7 Doctorate	Mala	Own	0%	Married	fipouse/Partner	2	1	2	ï	1 Never
18		7 Doctorete	Male	Own.	EXEL	Marriett	Spitule/Partner	3	102	2	1	1 Never
1	- 4	/ Doctorate	Male	Own	05L	Married	spouse/Partner	4		3.	1.	1 Never
.8	- 4	Doctorate	Male.	Own	050	Matried	Epouse/Parmer.	2	2	2	1	1 Never.
D	- 40	7 Doctorate	Main	Own	011.	Married	Spoule/Pariner	1	. 2	30	1	1 Narrer
30	4	2 Dioctorata	Mile	ŵwn.	this.	Married	toouse/Partner	2	22	10	1	1 Never
11	4	7 Doctorate	btelo .	Own	09.	Married	tipouse/Partner	2	12	- <u>i</u>)	1	1 Never
12	4	7 Doctorate	Male	Own.	0.51	Married	Sprinke/Partner	-2	02	1	1	1 Never
13	- 4	7 Doctorate	Male	Own	DSC.	Married	Spouse/Partner	2	23	2	1	1 forver
14	4	7 Doctorate	their .	Own	0%	Married	Spouse/Parmer	1	24	1	1	1 Never
15	4	7 Doctorate	state	Own	03.	Married.	spoure/Fartner	1	124	- 4 2	3.	1 Never
36	4	7 Doctorate	Malp	Ówn.	051	Marnad	spouse/Partner	2	1	27	1	1 Never
17	- 4	Doctorate	Male	Own	050	Mamed	Spouse/Partner	1	1	1	1	1 Never
58		2 Doctorate	Male	Own	019	Idamied	Spoce@Pattner	0	3	33	I	1 Novene
12	4	/ Doctorate	Male	Own	0%	Married	Spouse/Partner	2		1	1.	1 Never
-20	4	Doctoreta	Male	Own	092	Married	spolice/Partner.	2	.2	21	1	1 Never
15	- 14	7 Doctorate	Atale	Own	0%	Married	Spouse/Partner	1	1	1	1	1 Never

Now, Save this worksheet to YOUR desktop. (note that this worksheet is open in the Remote server). Click File > Save

As...



And make sure to select YOUR C: drive by double clicking it.



computer.

An American				3
S Houseft Office Excel	Name -	Date modified	Туре	See
	dell	5/25/2010 2:12 PM	File folder	
Favorites	🗼 drvrtnp	5/25/2010 2:14 PM	File folder	
Desktop	🗼 Intel	5/25/2010 2:12 PM	File folder	
Downloads	PerfLogs	7/13/2009 9:37 PM	File Folder	
All POLETIC Fraces	L Program Files	10/1/2010 12-08 PM	File folder	
📷 Libraries	Jusers	10/4/2010 2:49 PM	file folder	
Documents Music Pictures Videos	Undows 🚽	7/9/2010 2:29 PM	File folder	
👫 Computer				
Network.				
File genne: Baur				
Save as type: Excel	Workbook			
de étanos porose		Tage: Arbit a tag	Title: Ad	id a HRIe

workbook in Remote server and log off ts-mec.waltoncollege.uark.edu

Now, on YOUR computer, open YOUR C: drive, and then open the Churn Excel file and you can proceed with Data Mining from Excel.