Learn Predictive Analytics in 2 Hours!
Oracle Advanced Analytics Hands on Lab

Making Big Data + Analytics *Simple* —*On-Premise & Cloud*

Charlie Berger, MS Eng, MBA
Sr. Director Product Management, Data Mining and Advanced Analytics
charlie.berger@oracle.com www.twitter.com/CharlieDataMine

Brendan Tierney, President, Oralytics, Oracle ACE Director, Author

Tim Vlamis, VP & Analytics Strategist, Vlamis Software

Karl Rexer, Ph.D, President, Rexer Analytics

BIWA SUMMIT 2016

The Oracle Big Data + Analytics User Conference
January 26-28, 2016
Oracle Spatial Summit + YesSQL Summit

ORACLE

Learn Predictive Analytics in 2 Hours!

Oracle Advanced Analytics Hands on Lab

- **Lessons—Novices**
  1. **OAA Quick Overview**
  2. **Data Mining Concepts Briefly**
  3. **Quick Oracle Data Miner GUI Demo**
  4. **Take off! Do as many Tutorials as you can in the 2 hours HOL**
  5. **Ask questions! We're all here to help!**

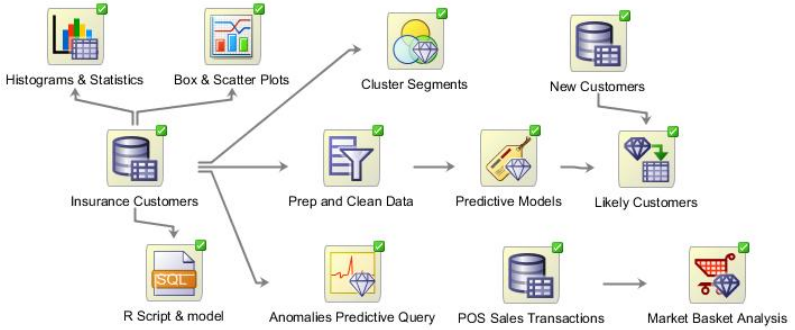
- **Take off!—Intermediate/Experts**
  1. **Environment**
     - Oracle 12c on the **Oracle Database Cloud**
     - Need SQL Developer 4.1.3
       - [See roaming instructors for memory stick]
       - Set up/configure Oracle Data Miner extension (already done, but read instructions)
  2. **Do 3-5 Tutorials**
     - Instructors will walk around helping

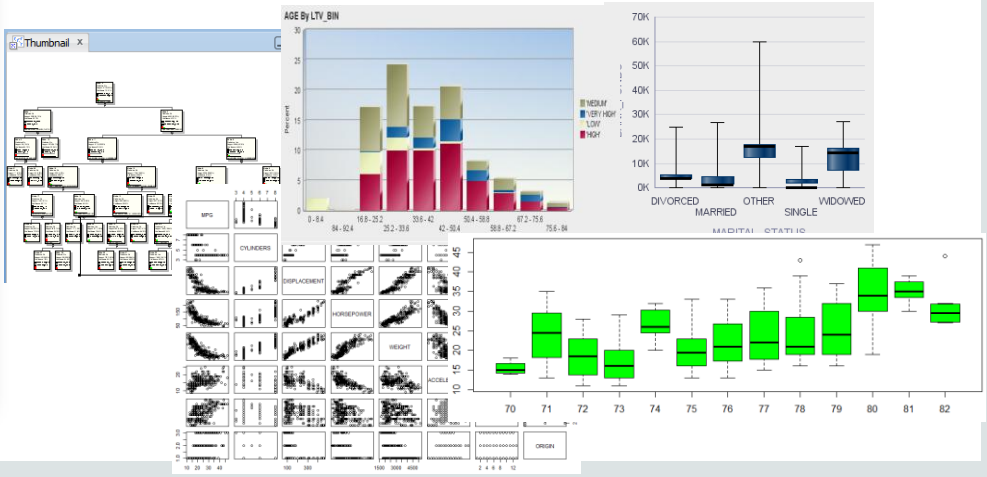
Oracle's Advanced Analytics

Fastest Way to Deliver Scalable Enterprise-wide Predictive Analytics

Key Features

- Scalable in-Database + Hadoop data mining algorithms and R integration
- Powerful predictive analytics and deployment platform
- Drag and drop workflow, R and SQL APIs
- Data analysts, data scientists & developers
- Enables enterprise predictive analytics applications



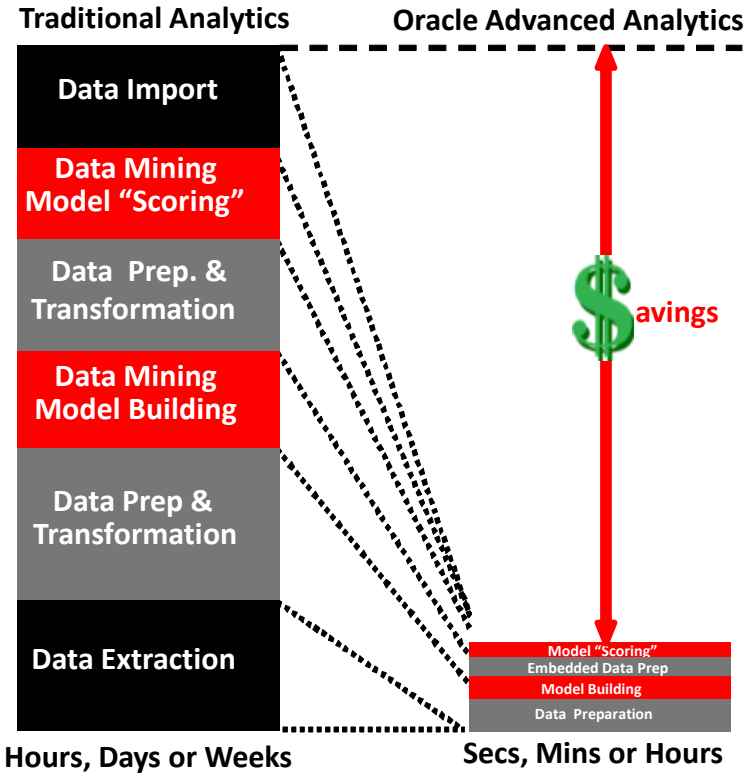


Oracle's Advanced Analytics

Fastest Way to Deliver Scalable Enterprise-wide Predictive Analytics

Major Benefits

- Data remains in Database & Hadoop
  - Model building and scoring occur in-database
  - Use R packages with data-parallel invocations
- Leverage investment in Oracle IT
  - Eliminate data duplication
  - Eliminate separate analytical servers
- Deliver enterprise-wide applications
  - GUI for predictive analytics & code generation
  - R interface leverages database as HPC engine



Oracle's Advanced Analytics

Multiple interfaces across platforms — SQL, R, GUI, Dashboards, Apps

Users

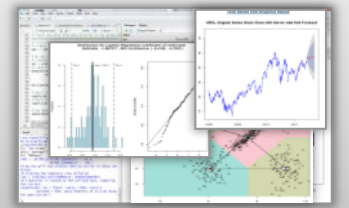
R programmers

Data & Business Analysts

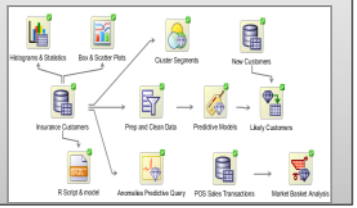
Business Analysts/Mgrs

Domain End Users

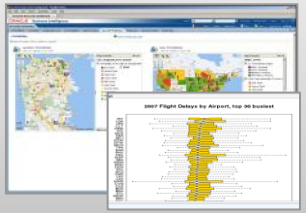
R Client



SQL Developer/ Oracle Data Miner

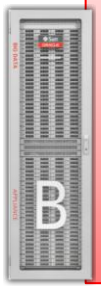


OBIEE



Applications

Platform



Hadoop

ORAAH

Parallel, distributed algorithms

Oracle Database Enterprise Edition

Oracle Advanced Analytics - Database Option

SQL Data Mining & Analytic Functions + R Integration for Scalable, Distributed, Parallel in-Database ML Execution

Oracle Cloud

Take off!—Intermediate/Experts

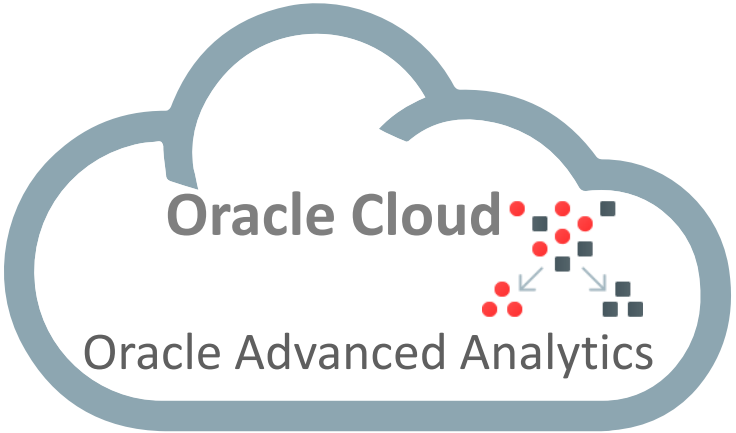
Quick Set up Overview

ORACLE

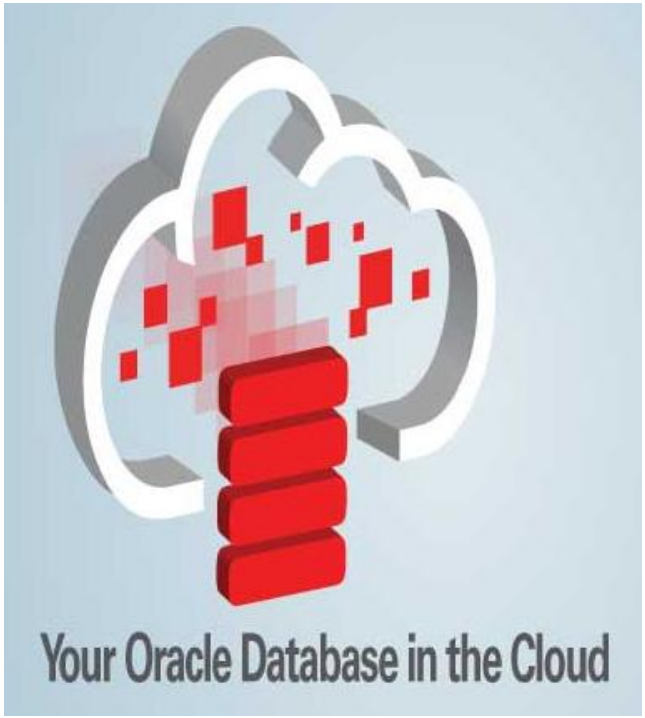
Learn Predictive Analytics in 2 Hours!

Oracle Advanced Analytics Hands on Lab

- We're using the Oracle Database Cloud!

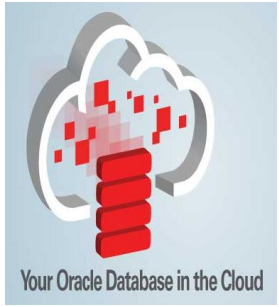


=

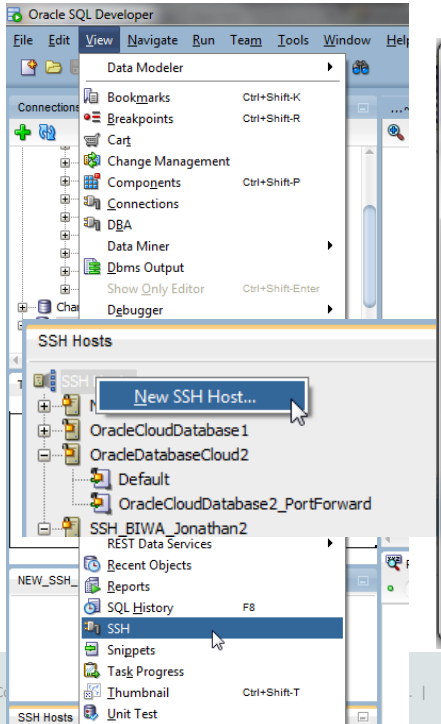


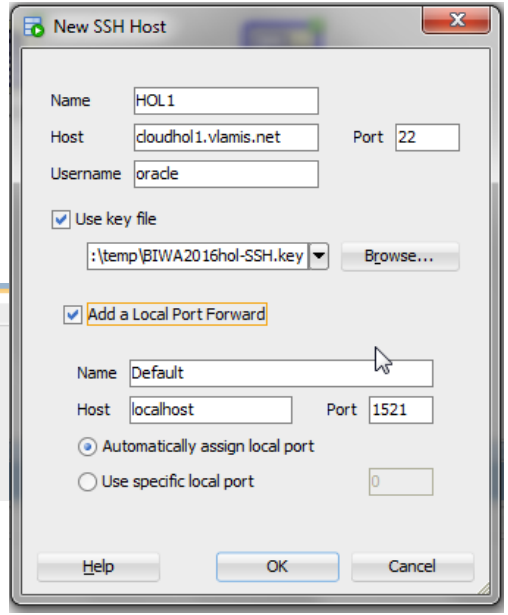
Learn Predictive Analytics in 2 Hours!

Oracle Advanced Analytics Hands on Lab



- Step 1—Install SQLDEV 4.1.3
- Step 2—Connect to Oracle Database Cloud
  - 1. Create SSH using “key”
    - www.vlamis.com/storage/biwa2016/BIWA2016hol-SSH.key
    - www.vlamis.com\storage\biwa.key
    - Cloudhol”N”.vlamis.net
    - Username is “oracle”
    - Add Port Forward and accept defaults

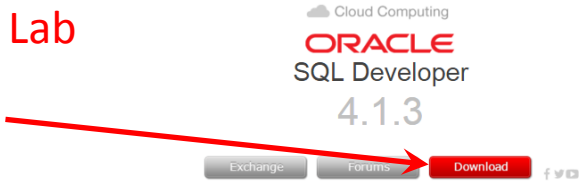


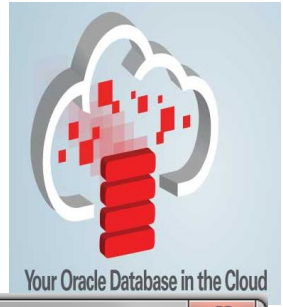


ORACLE

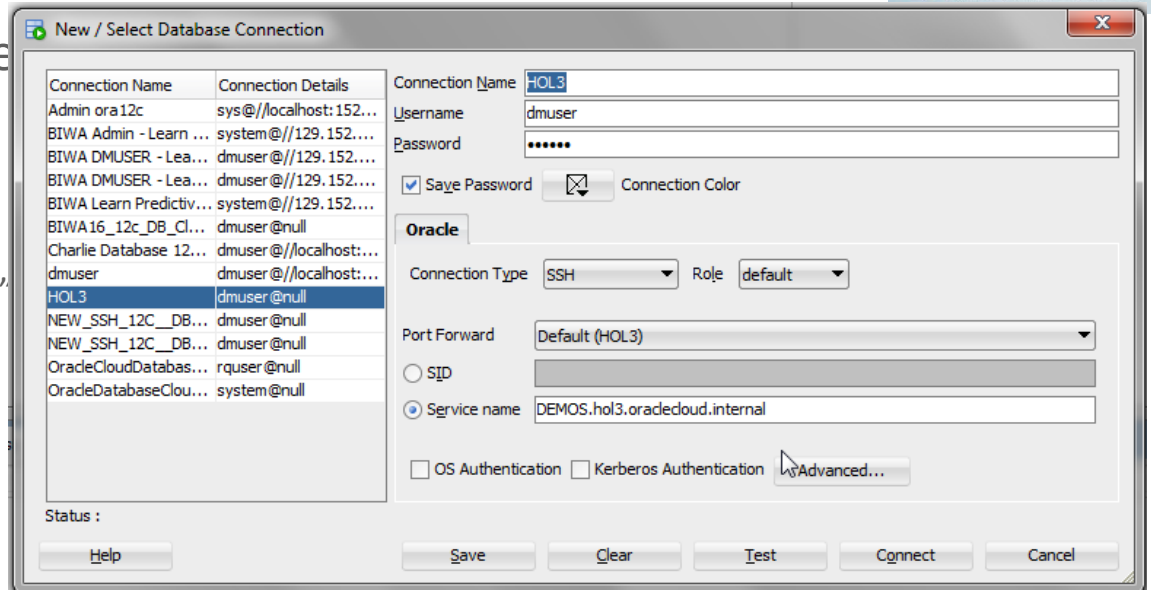
Learn Predictive Analytics in 2 Hours!

Oracle Advanced Analytics Hands on Lab





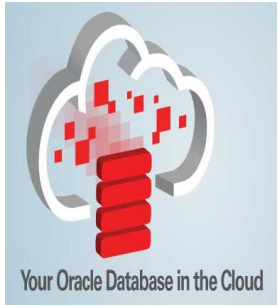
- Step 1—Install SQLDEV 4.1.3
- Step 2—Connect to Oracle Database Cloud
  - 1. Create SQL Database Connection
    - (Each student take unique HOL"N" assigned)
    - dmuser/dmuser (save password)
    - SSH
    - Port Forward (Default HOL"N")
    - Service ID (e.g. DEMOS.hol"N".oraclecloud.internal)

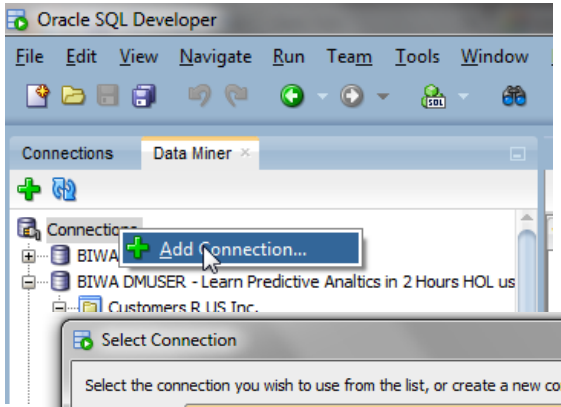


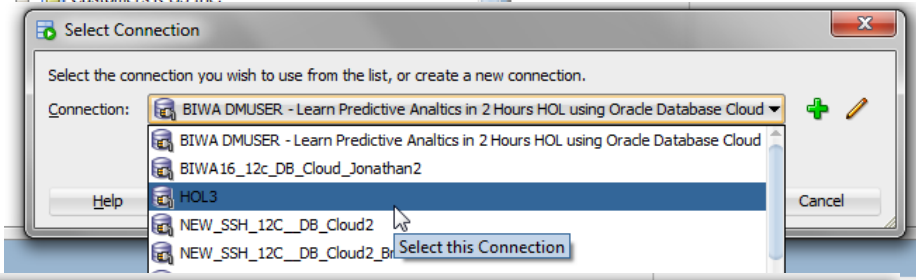
Learn Predictive Analytics in 2 Hours!

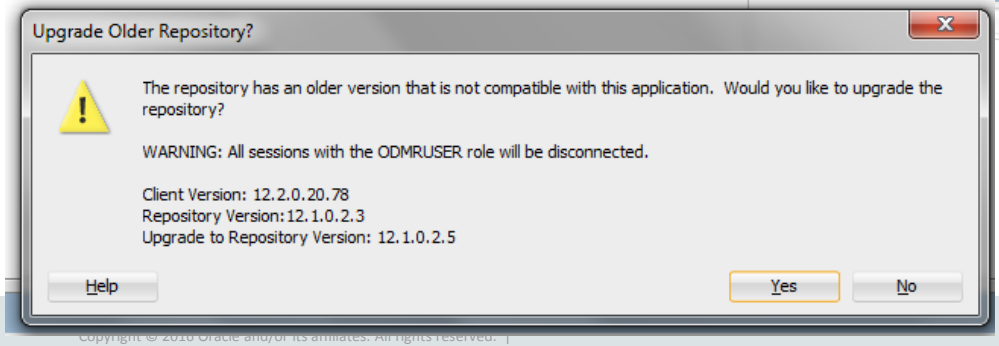
Oracle Advanced Analytics Hands on Lab

- Step 1—Install SQLDEV 4.1.3
- Step 2—Connect to Oracle Database Cloud
  - 1. Go to Oracle Data Miner & create a NEW Connection e.g. HOL”N”
  - Select HOL”N” from drop down menu
  - Optionally may need to upgrade older Data Mining repository (may take 3 mins)
  - You are done!



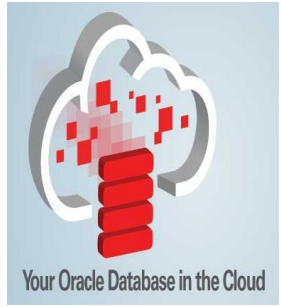




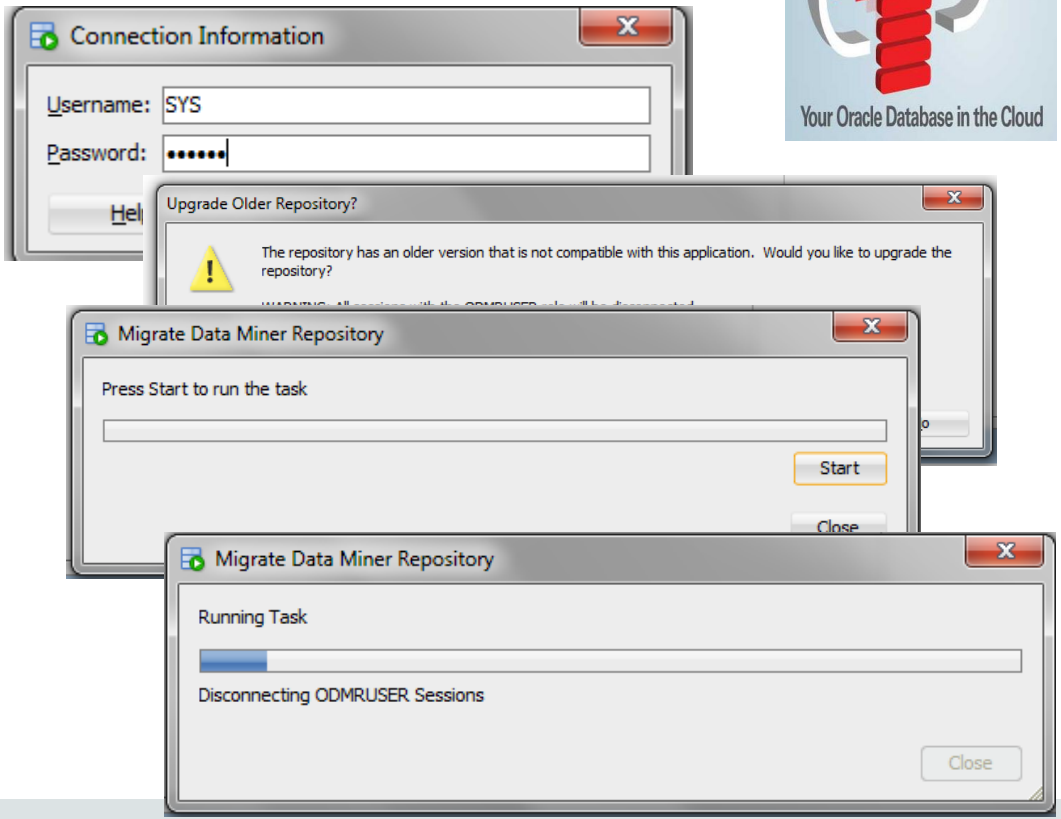


Learn Predictive Analytics in 2 Hours!

Oracle Advanced Analytics Hands on Lab

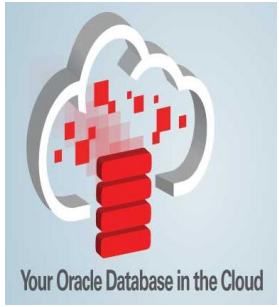


- Step 1—Install SQLDEV 4.1.3
- Step 2—Connect to Oracle Database Cloud
  - Connect as SYS/Welcome#1
  - Start to run task
  - Running task (may take 3 mins)

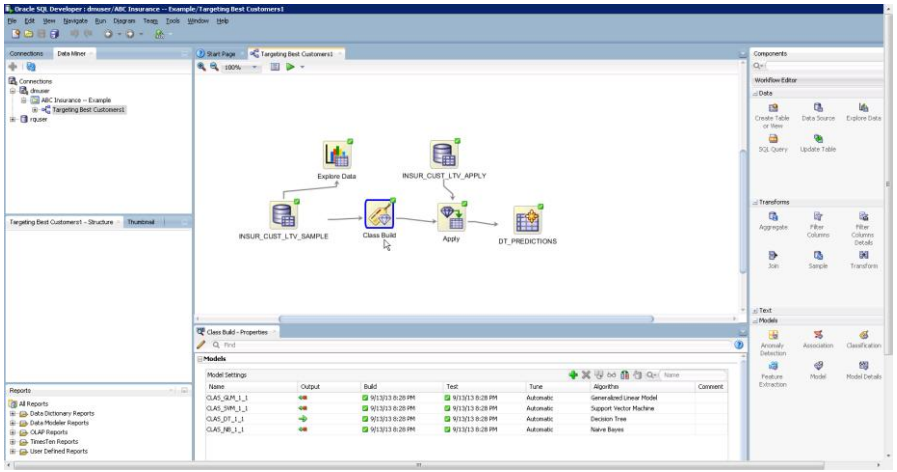


Learn Predictive Analytics in 2 Hours!

Oracle Advanced Analytics Hands on Lab



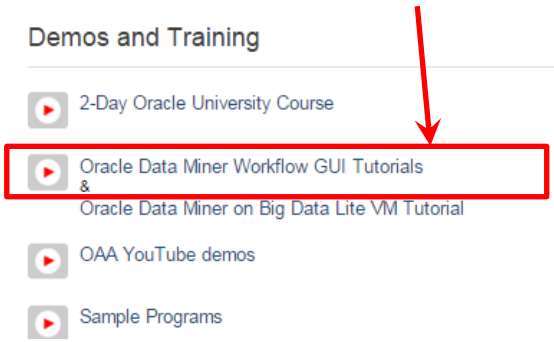
- Step 1—Install SQLDEV 4.1.3
- Step 2—Connect to Oracle Database Cloud
- Step 3—Start HOL!
  - dmuser/dmuser
  - Demo data for learning
  - Follow 3-5 OBE Online Tutorials
    1. Using Oracle Data Miner 4.1
    2. Star Schema Mining Using Oracle Data Miner 4.1
    3. Text Mining with an EM Clustering Model Using Data Miner 4.1
    4. **Anomaly Detection (CLAIMS)** *See Instructor for assistance*
    5. **Market Basket Analysis (SH.SALES)** *See Instructor for assistance*



OAA/Oracle Data Miner 4.1 HOL

We're Using the Oracle by Example Free Online Tutorials

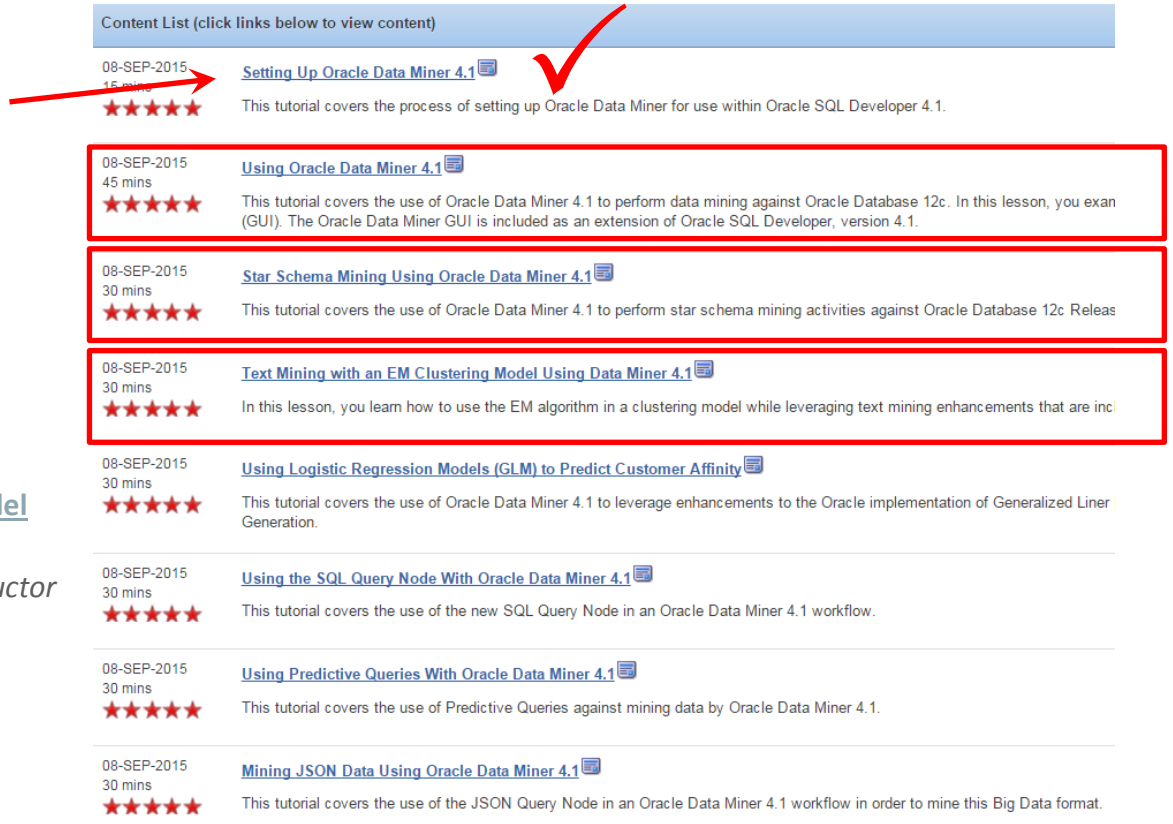
- Google "Oracle Data Miner"
- Scroll down to bottom of page & launch tutorials
  - https://apexapps.oracle.com/pls/apex/f?p=44785:24:::NO::P24_CONTENT_ID,P24_PREV_PAGE:11925,2



OAA/Oracle Data Miner 4.1 HOL

Uses Oracle by Example Free Online Tutorials

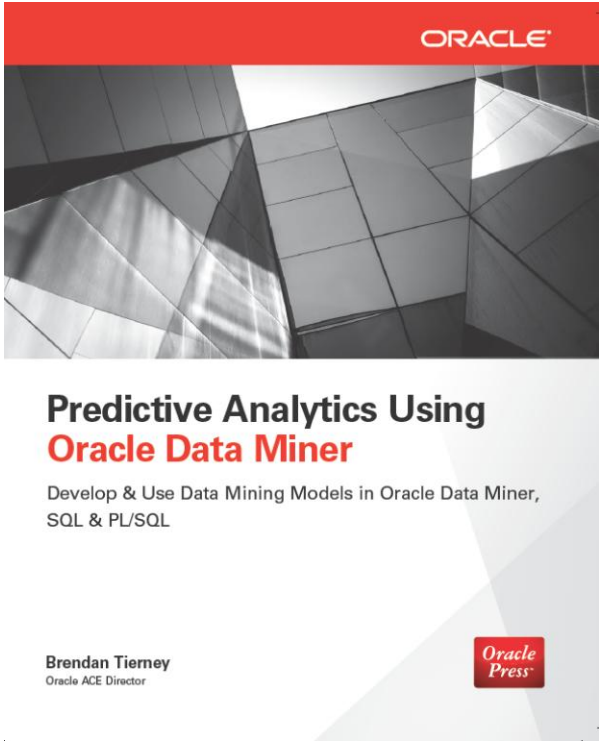
- There are 6 Tutorials
  - The first tutorial is already done for you
  - Recommend doing 3-5 Tutorials
    1. Using Oracle Data Miner 4.1
    2. Star Schema Mining Using Oracle Data Miner 4.1
    3. Text Mining with an EM Clustering Model Using Data Miner 4.1
    4. **Anomaly Detection (CLAIMS)** *See Instructor for assistance*
    5. **Market Basket Analysis (SH.SALES)** *See Instructor for assistance*



Great book on Oracle Advanced Analytics

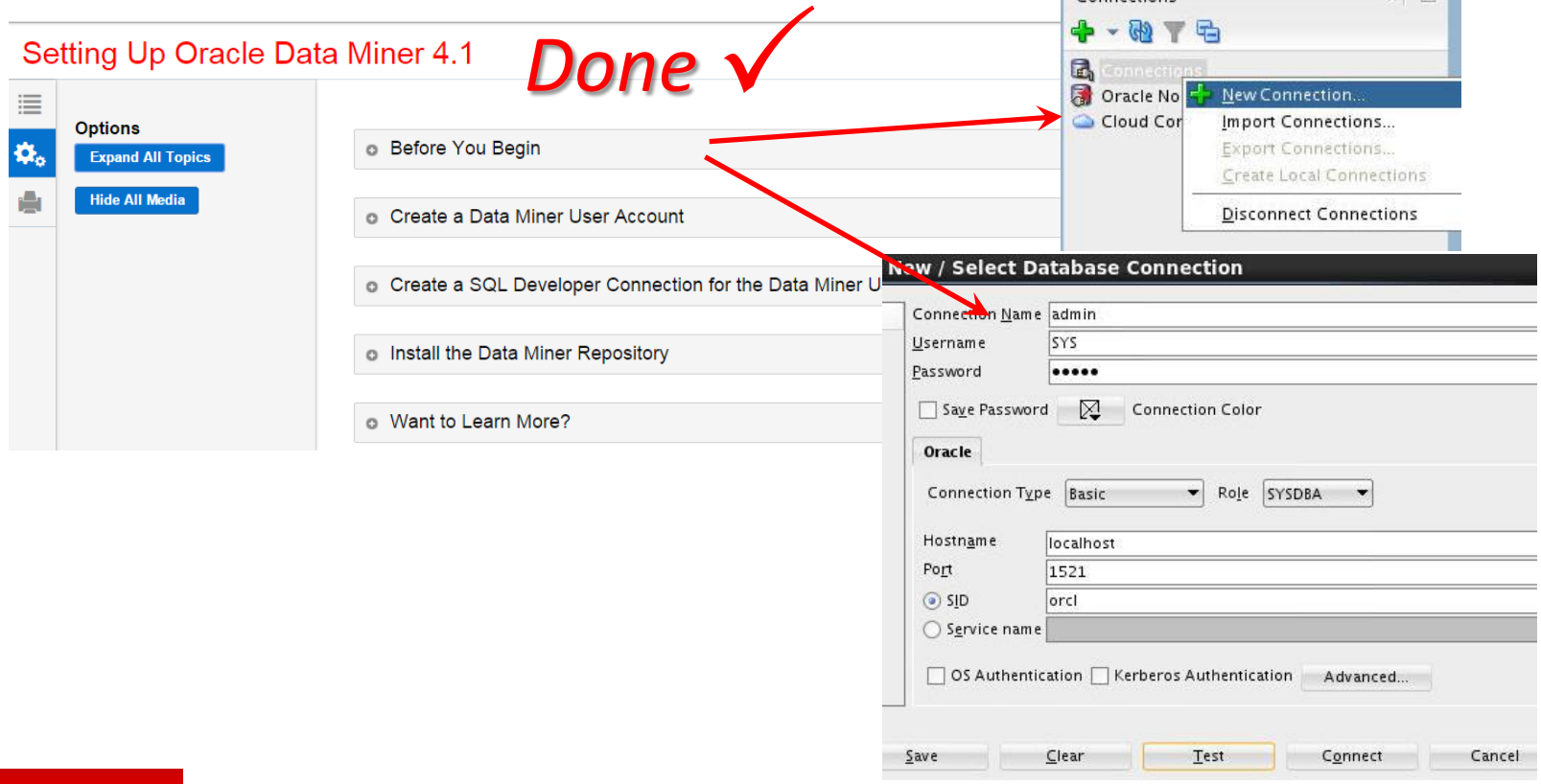
Available on Amazon or from Author

- Predictive Analytics Using Oracle Data Miner: Develop for ODM in SQL & PL/SQL



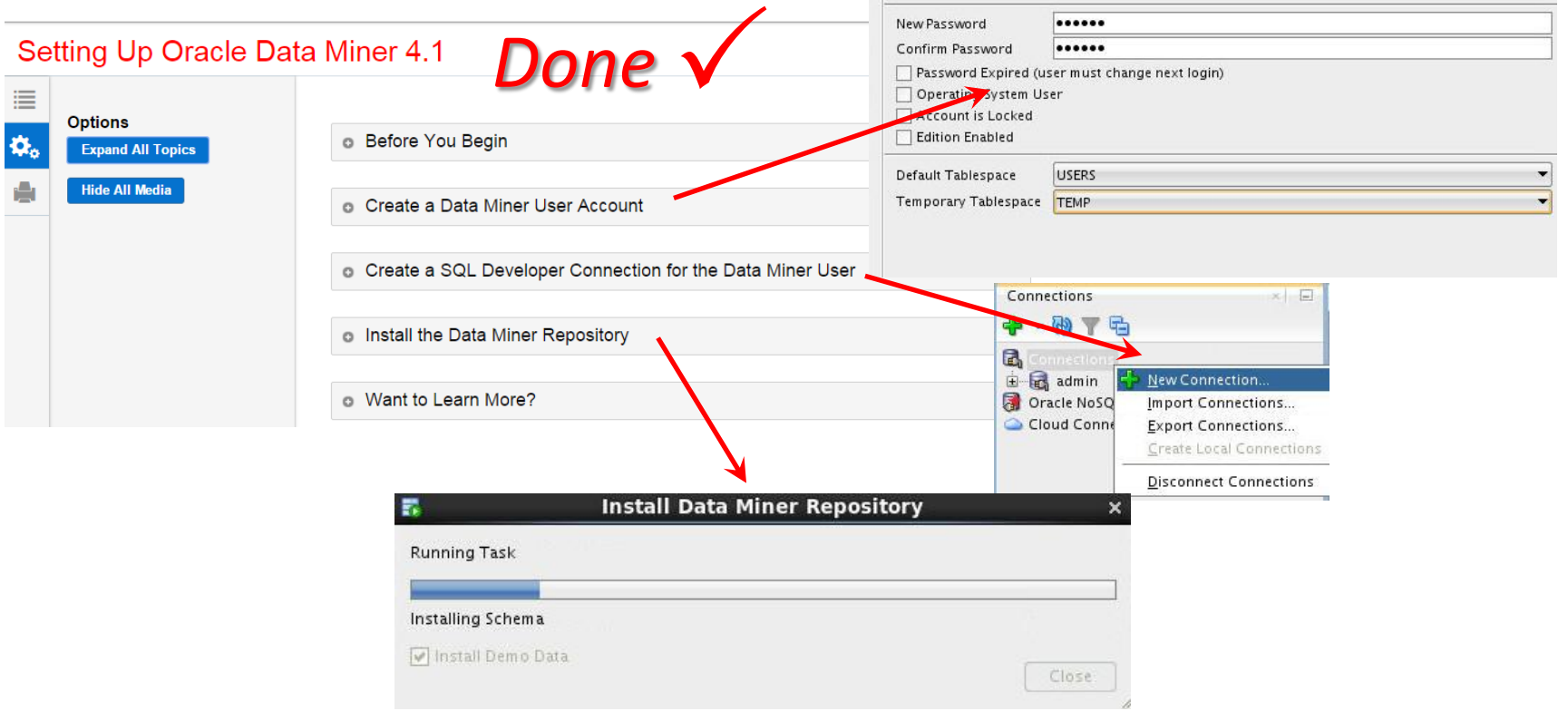
OAA/Oracle Data Miner 4.1 HOL

Setting Up Oracle Data Miner

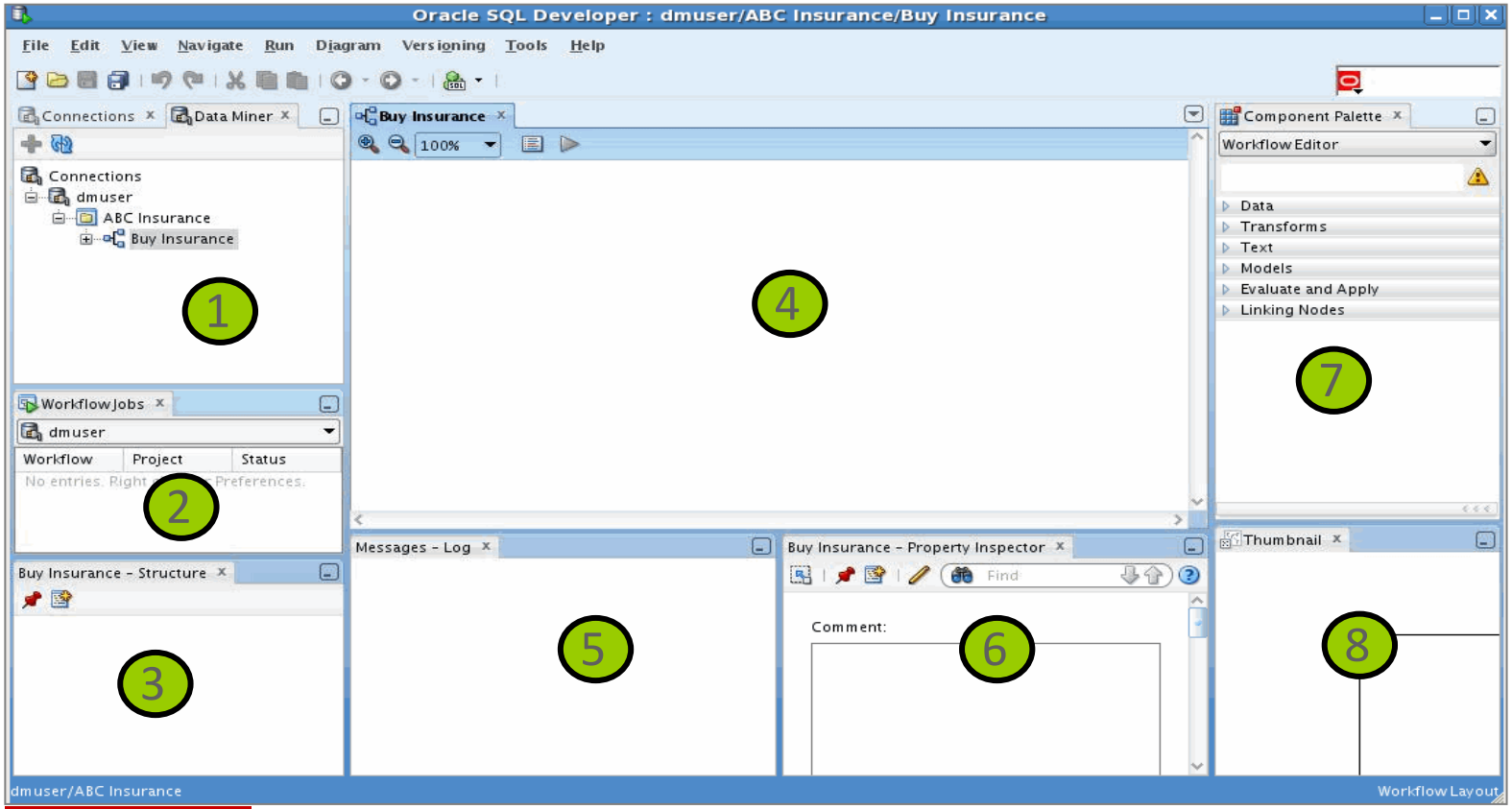


OAA/Oracle Data Miner 4.1 HOL

Setting Up Oracle Data Miner

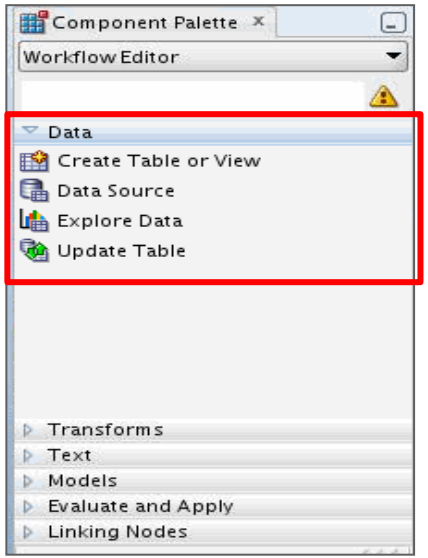


Introducing the Data Miner Interface

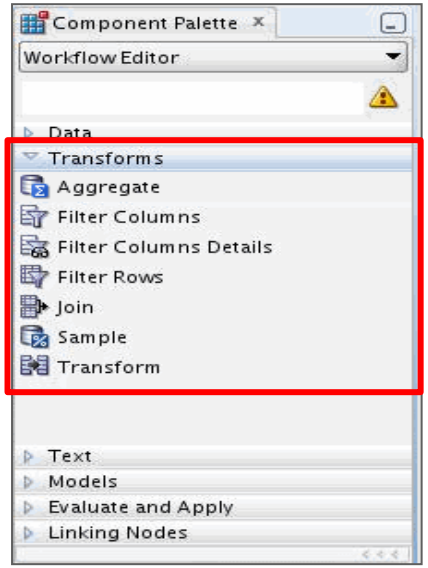


Examining Oracle Data Miner Nodes

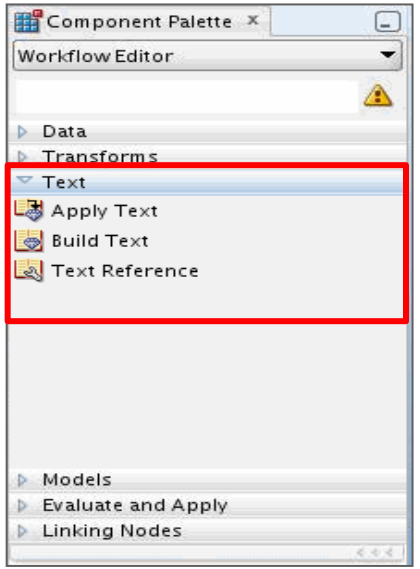
Data



Transforms

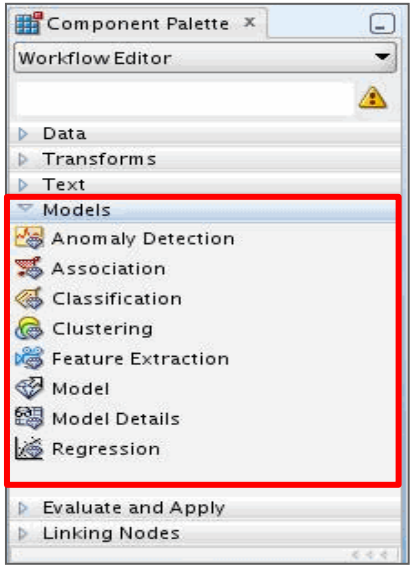


Text

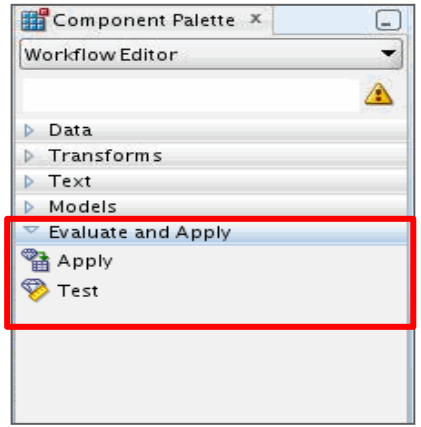


Examining Oracle Data Miner Nodes

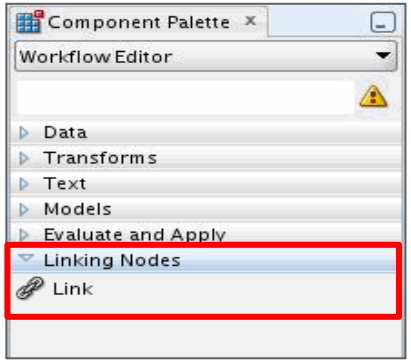
Models



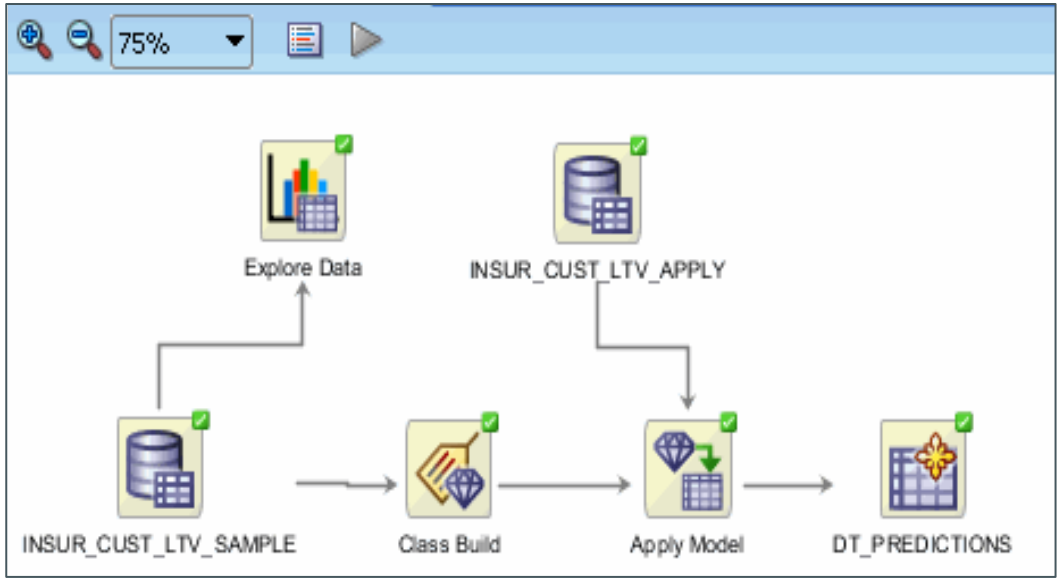
Evaluate and Apply



Linking



Previewing a Data Miner Workflow

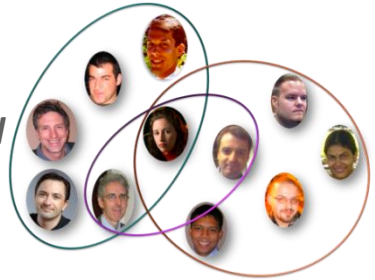


Multiple Data Sources/Types with Predictive Modeling

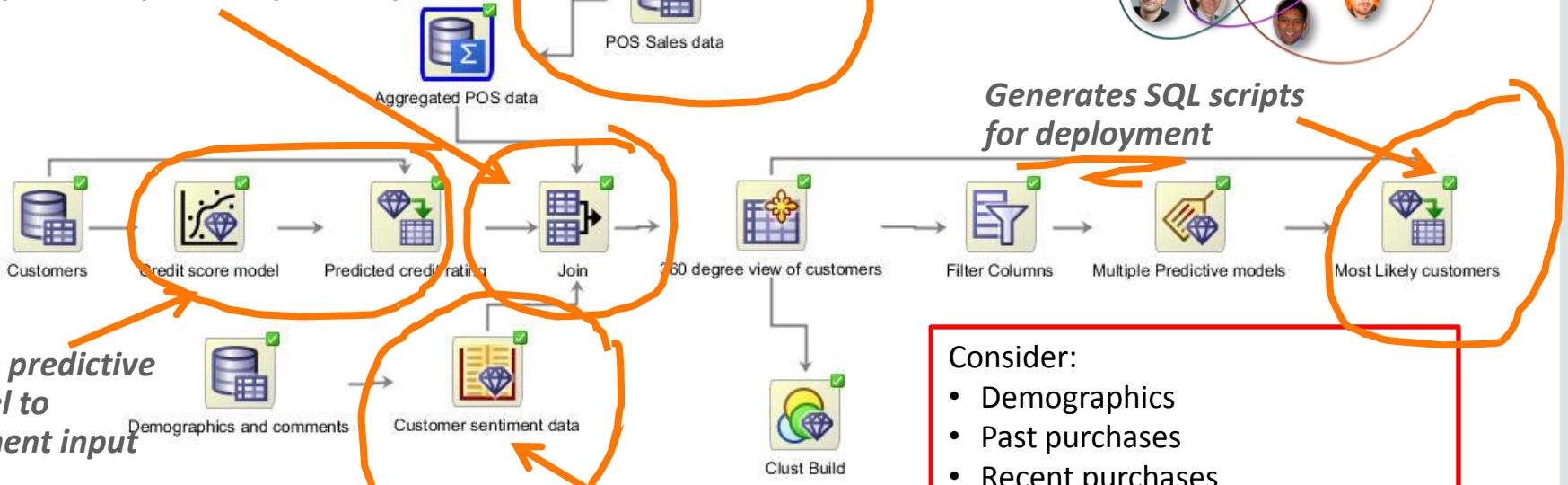
Ease of Deployment through SQL Script Generation

SQL Joins and arbitrary SQL transforms & queries – power of SQL

Transactional POS data



Generates SQL scripts for deployment



Inline predictive model to augment input data

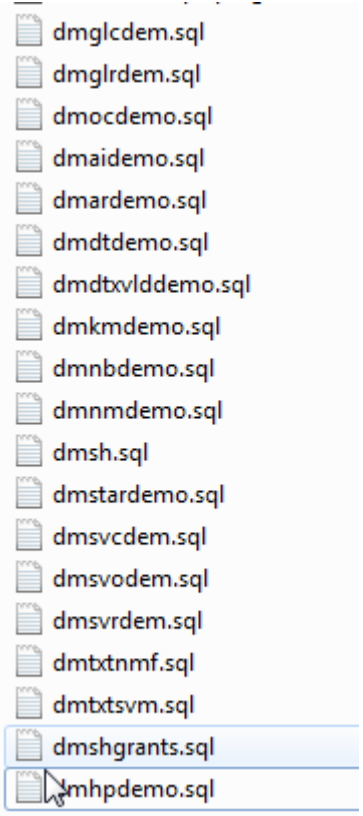
Unstructured data also mined by algorithms

Consider:

- Demographics
- Past purchases
- Recent purchases
- Customer comments & tweets

The Data Mining Sample Programs

- The Data Mining Sample Programs 12c Documentation
  - You can learn a great deal about the Oracle Data Mining API from the data mining sample programs. The programs illustrate typical approaches to data preparation, algorithm selection, algorithm tuning, testing, and scoring.
  - The programs are easy to use. They include extensive inline comments to help you understand the code. They delete all temporary objects on exit; you can run the programs repeatedly without setup or cleanup.
  - The data mining sample programs are installed with Oracle Database Examples in the demo directory under Oracle Home. The demo directory contains sample programs that illustrate many features of Oracle Database. You can locate the data mining files by doing a directory listing of dm*.sql.



The Data Mining Sample Programs

Attribute Importance Sample Code

5.09600019 seconds

Worksheet | Query Builder

```
--------------------------------------------------------------------------------
--                              BUILD THE MODEL
--------------------------------------------------------------------------------


-------------------
-- Cleanup old output table for repeat runs
BEGIN EXECUTE IMMEDIATE 'DROP TABLE ai_explain_output';
EXCEPTION WHEN OTHERS THEN NULL; END;
/

-------------------
-- Run the EXPLAIN routine to get attribute importance results
BEGIN
  DBMS_PREDICTIVE_ANALYTICS.EXPLAIN(
    data_table_name     => 'mining_data_build_v',
    explain_column_name => 'affinity_card',
    result_table_name   => 'ai_explain_output');
END;
/

------------------------
-- DISPLAY RESULTS
```

Script Output ×

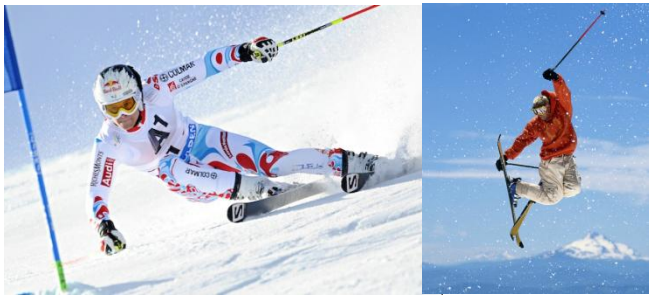
Task completed in 5.096 seconds

ATTRIBUTE_NAME	EXPLANATORY_VALUE	RANK
HOUSEHOLD_SIZE	.195	1
CUST_MARITAL_STATUS	.194	2
YRS_RESIDENCE	.115	3
EDUCATION	.106	4
AGE	.104	5
OCCUPATION	.092	6
Y_BOX_GAMES	.077	7
HOME_THEATER_PACKAGE	.069	8
CUST_GENDER	.043	9
BOOKKEEPING_APPLICATION	.024	10
BULK_PACK_DISKETTES	.000	11
COUNTRY_NAME	.000	11
CUST_ID	.000	11
CUST_INCOME_LEVEL	.000	11
FLAT_PANEL_MONITOR	.000	11
OS_DOC_SET_KANJI	.000	11
PRINTER_SUPPLIES	.000	11

17 rows selected

Learn Predictive Analytics in 2 Hours!

Oracle Advanced Analytics Hands on Lab



- Recommend doing 3-5 Tutorials
  1. Using Oracle Data Miner 4.1
  2. Star Schema Mining Using Oracle Data Miner 4.1
  3. Text Mining with an EM Clustering Model Using Data Miner 4.1
  4. **Anomaly Detection (CLAIMS)** *See Instructor for assistance*
  5. **Market Basket Analysis (SH.SALES)** *See Instructor for assistance*

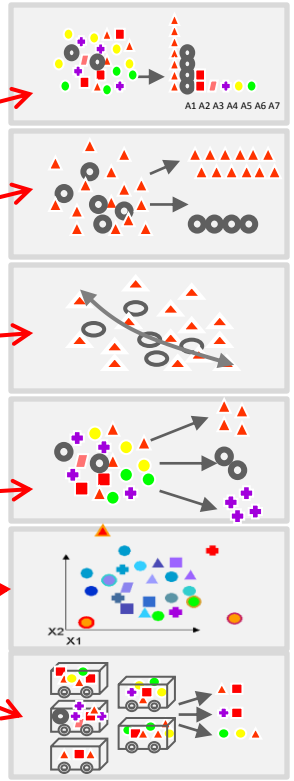
Novice/Introductory/Overviews

Quick Overview of Concepts, Process and Use Cases

What is Data Mining & Predictive Analytics?

Automatically sifting through large amounts of data to create models that find previously hidden patterns, discover valuable new insights and make predictions

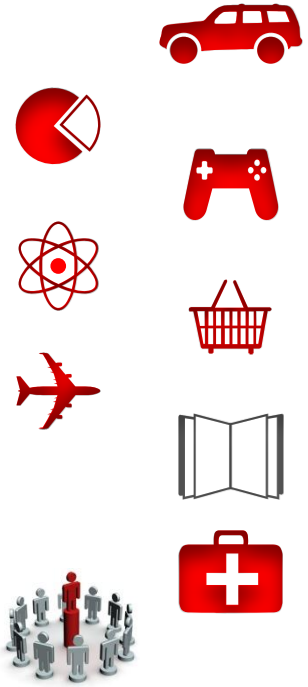
- Identify most important factor *(Attribute Importance)*
- Predict customer behavior *(Classification)*
- Predict or estimate a value *(Regression)*
- Find profiles of targeted people or items *(Decision Trees)*
- Segment a population *(Clustering)*
- Find fraudulent or “rare events” *(Anomaly Detection)*
- Determine co-occurring items in a “baskets” *(Associations)*



Predictive Analytics & Data Mining

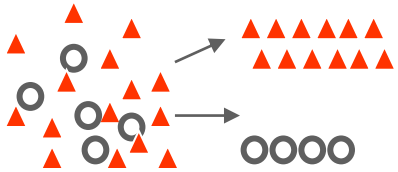
Typical Use Cases

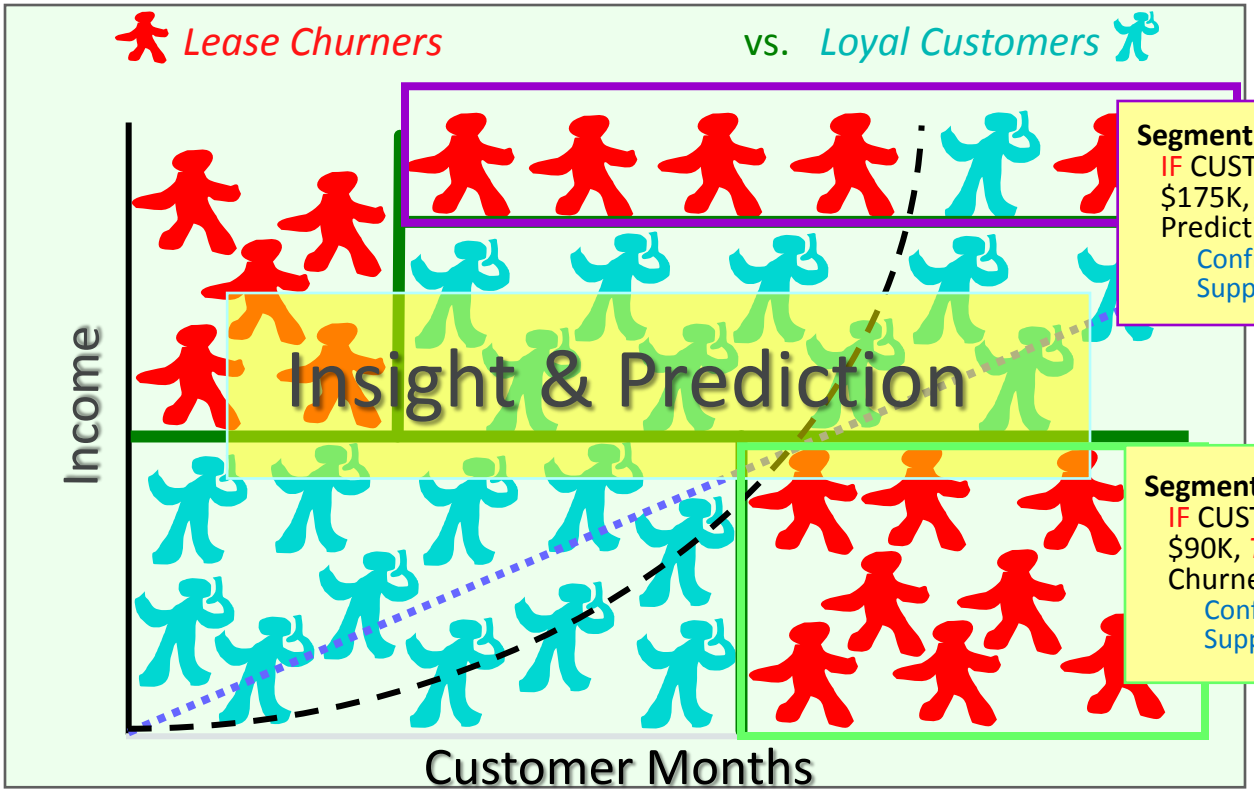
- Targeting the right customer with the right offer
- How is a customer likely to respond to an offer?
- Finding the most profitable growth opportunities
- Finding and preventing customer churn
- Maximizing cross-business impact
- Security and suspicious activity detection
- Understanding sentiments in customer conversations
- Reducing medical errors & improving quality of health
- Understanding influencers in social networks



Data Mining Provides

Better Information, Valuable Insights and Predictions





Segment #3
IF CUST_MO > 7 AND INCOME < $175K, *THEN* Prediction = Lease Churner,
Confidence = 83%
Support = 6/39

Segment #1
IF CUST_MO > 14 AND INCOME < $90K, *THEN* Prediction = Lease Churner
Confidence = 100%
Support = 8/39

Source: Inspired from *Data Mining Techniques: For Marketing, Sales, and Customer Relationship Management* by Michael J. A. Berry, Gordon S. Linoff

Oracle's Advanced Analytics

In-Database Data Mining Algorithms*—SQL & R & GUI Access

Classification

- Decision Tree
- Logistic Regression (GLM)
- Naïve Bayes
- Support Vector Machine (SVM)
- Random Forest

Regression

- Multiple Regression (GLM)
- Support Vector Machine (SVM)
- Linear Model
- Generalized Linear Model
- Multi-Layer Neural Networks
- Stepwise Linear Regression

Clustering

- Hierarchical k-Means
- Orthogonal Partitioning Clustering
- Expectation-Maximization

Attribute Importance

- Minimum Description Length
- Unsupervised pair-wise KL div.

Anomaly Detection

- 1 Class Support Vector Machine

Time Series

- Single & Double Exp. Smoothing

Predictive Queries

- Clustering
- Regression
- Anomaly Detection
- Feature Extraction

Feature Extraction & Creation

- Nonnegative Matrix Factorization
- Principal Component Analysis
- Singular Value Decomposition

Market Basket Analysis

- Apriori – Association Rules

Open Source R Algorithms

- Ability to run any R package via Embedded R mode

* supports partitioned models, text mining

ORACLE

Oracle University's Learn Predictive Analytics Using Oracle Data Mining Course Agenda

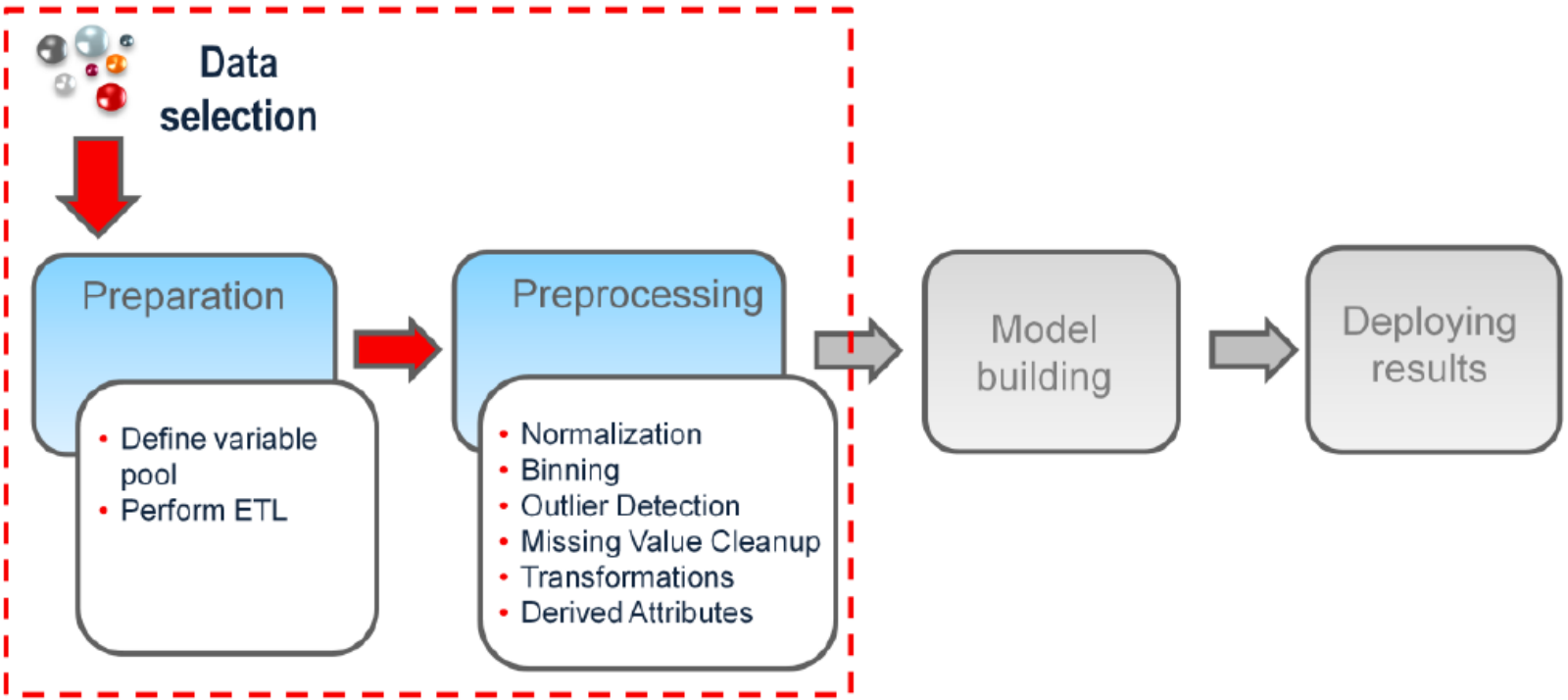
Day 1

1. Introduction
2. Predictive Analytics and Data Mining Concepts
3. The Data Mining Process
4. Introducing Oracle Data Miner 4.1
5. Using Classification Models
6. Using Regression Models

Day 2

7. Using Clustering Models
8. Performing Market Basket Analysis
9. Performing Anomaly Detection
10. Performing Star Schema Mining
11. Using Predictive Queries
12. Deploying Data Mining Results

The Data Mining Process



Data Mining Attributes

Data mining attributes can be:

- Categorical
- Numerical
- Text
- Target
- Active or Inactive

OCCUPATION
Prof.
Sales
Machine
Crafts

Categorical

AGE
40
47
32
63
40

Numerical

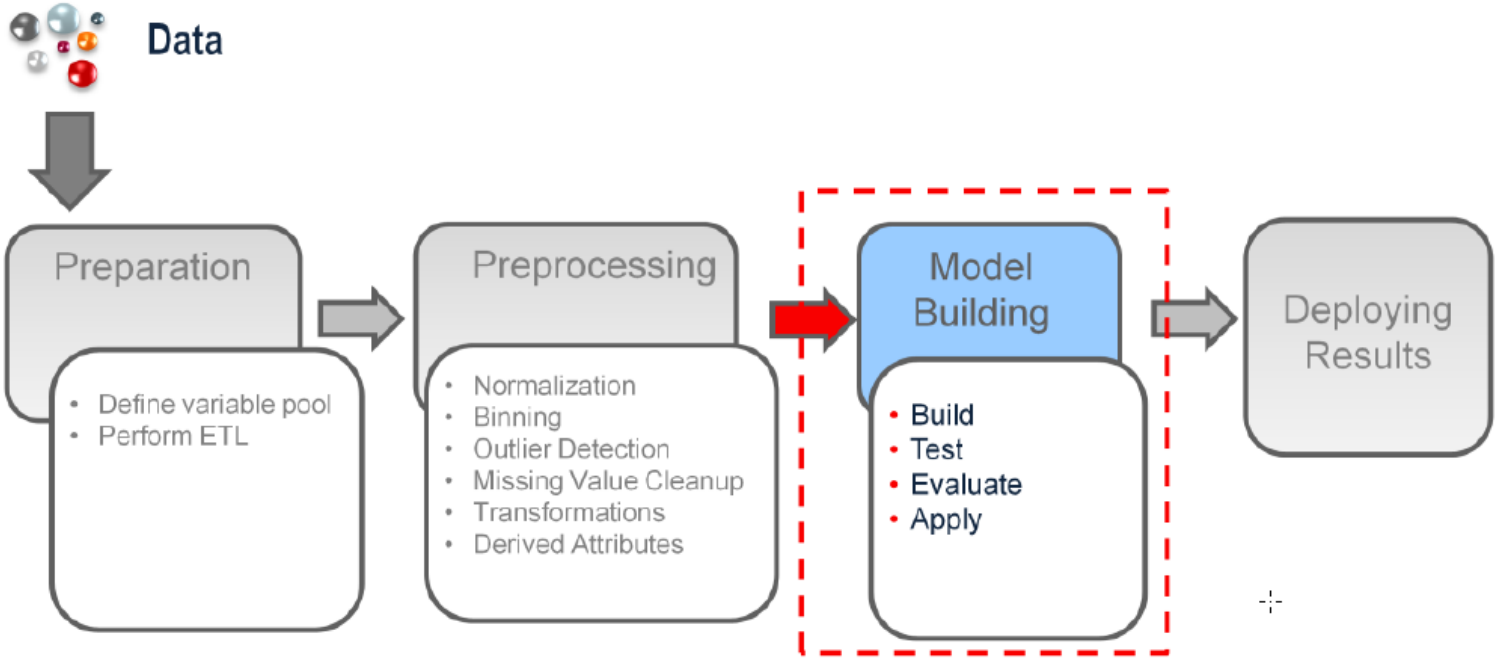
CLASS
1
0
0
0
0

Target

EDUCATION
Bach.
5th-6th
HS-grad
Bach.
HS-grad

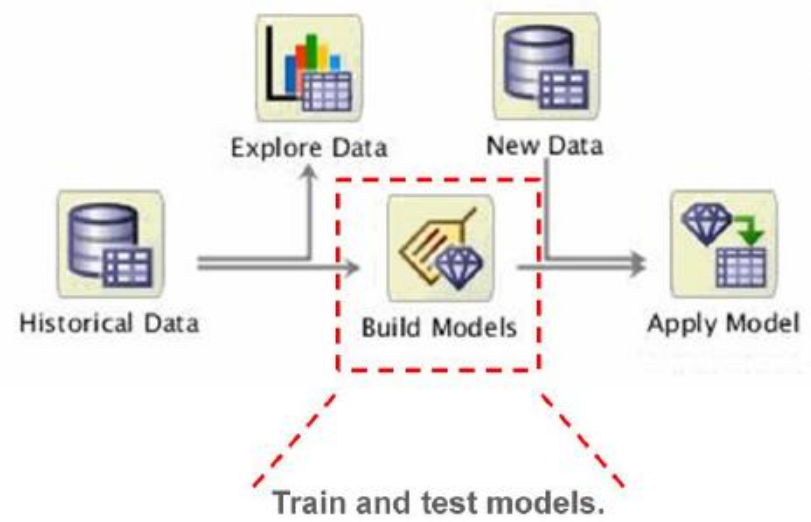
Active

Building and Evaluating Models

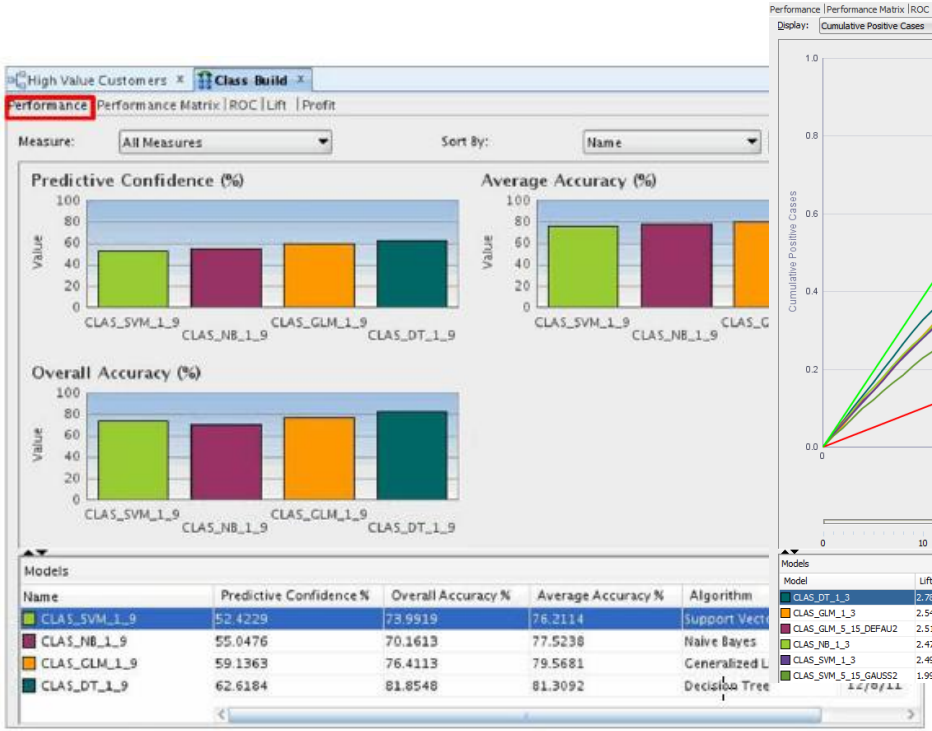


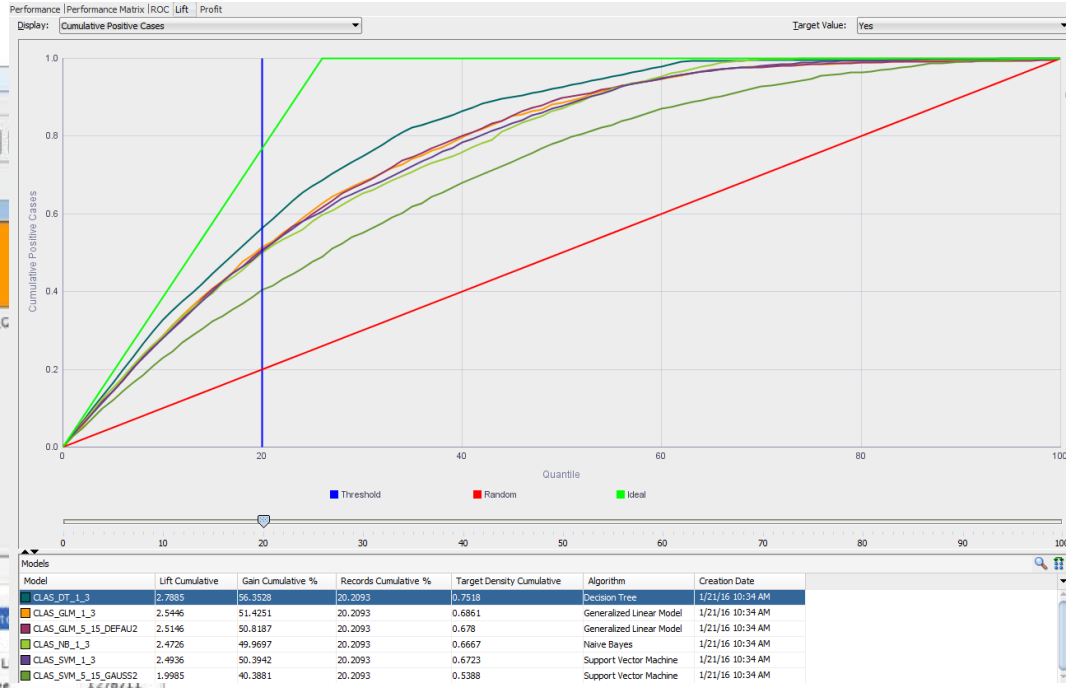
Model Building Tasks

Model Train & Test: Supervised Learning

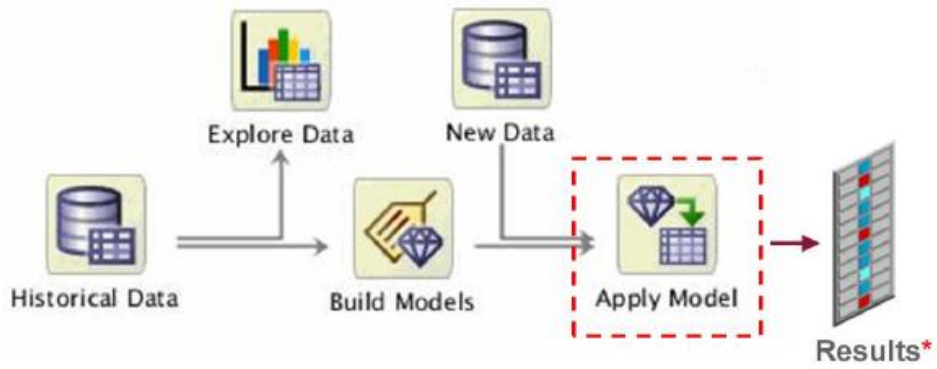


Model Evaluation: Supervised Learning





Applying the Selected Model(s)



(* Output settings may be specified by the user.)

Oracle Advanced Analytics

Brief Demos

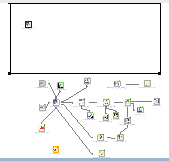
ORACLE

File Edit View Navigate Run Diagram Team Tools Window Help

Connections | Data Miner

Connections
- dmuser
  - ACME Mfg Paint Project
  - BERGERS R US
    - Fun with Gov
    - Predictive Analytics WF
  - Chicago Crime
  - Customers R Us Project
    - A+ Students OAA analytics
    - Big Data Analytcs w JSON
    - Big Data Analytics
    - *BUY_INSURANCE_WORKFLOW*
    - CARS_DATA_MINING
    - Churners01 work flow
    - Claims Fraud Clustering + SVM2
    - Customer Analytics incl RFM
    - Employees_attrition
    - Insurance Customer Analytics
    - Manufacturing Painting Ops
    - Market Basket Analysis
    - OOW14_CUSTOMER_ANALYTICS_360_I
    - Predictive Queries

Thumbnail



BUY_INSURANCE_WORKFLOW - ... | Reports

- Clustering Segmentation
- Explore/Profile Data
- Prediction Queries by Region
- Prediction Query
- Filter Cols_Attrib Importance
- Predictive Models for Student
- Class Build
- Predictive models incl TEXT
- Filter Columns Details
- Scatter Box plots etc.
- Graph
- Anomaly Detection Query
- CUST_INSUR_LTV
- CUST_INSUR_LTV_APPLY
- MINING_DATA_TEXT_BUILD_V

Start Page | *BUY_INSURANCE_WORKFLOW*

100% | Parallel Query Off

CUST_INSUR_LTV1

CUST_INSUR_LTV1 - Properties

Find

Data

Cache

Details

Source Table: DMUSER.CUST_INSUR_LTV

Data

Name

Name	Alias	Data Type
AGE		NUMBER
BANK_FUNDS		NUMBER
BUY_INSURANCE		VARCHAR2
CAR_OWNERSHIP		NUMBER
CHECKING_AMOUNT		NUMBER

Components

Workflow Editor

Data
- Create Table or View
- Data Source
- Explore Data
- Graph
- SQL Query
- Update Table

Transforms
- Aggregate
- Filter Columns
- Filter Columns Details
- Filter Rows
- Join
- JSON Query
- Sample
- Transform

Text

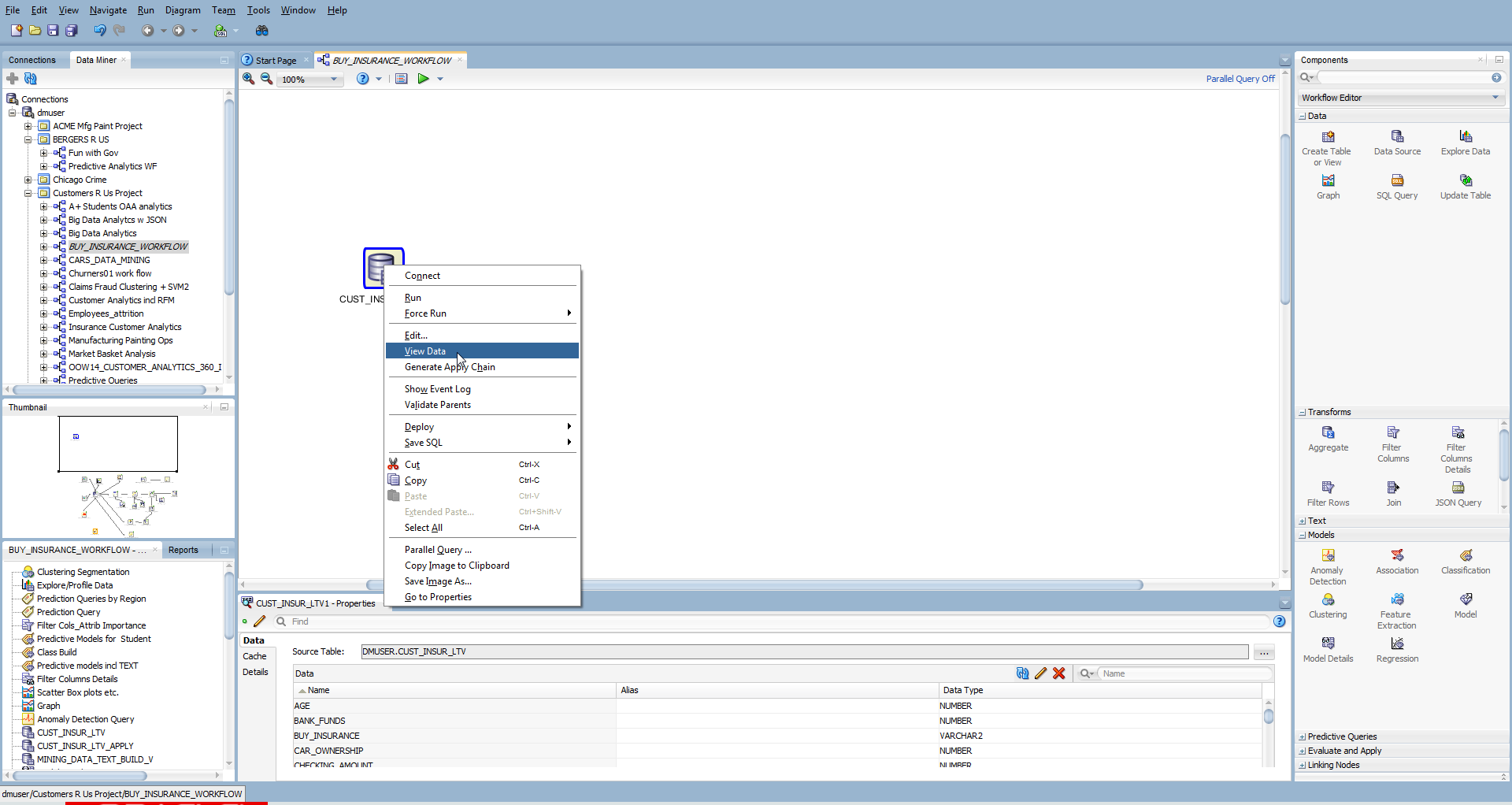
Models
- Anomaly Detection
- Association
- Classification
- Clustering
- Feature Extraction
- Model
- Model Details
- Regression

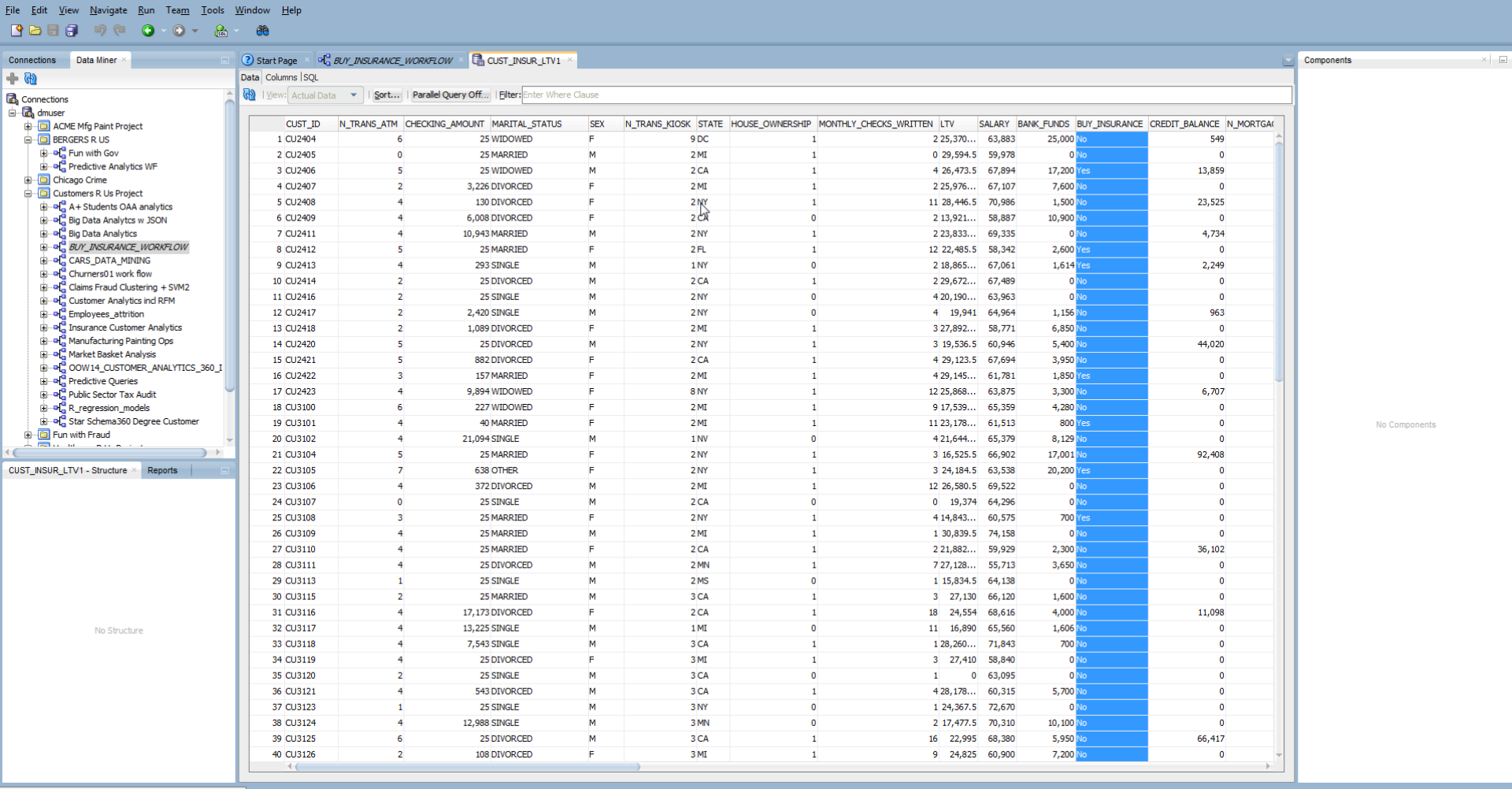
Predictive Queries

Evaluate and Apply

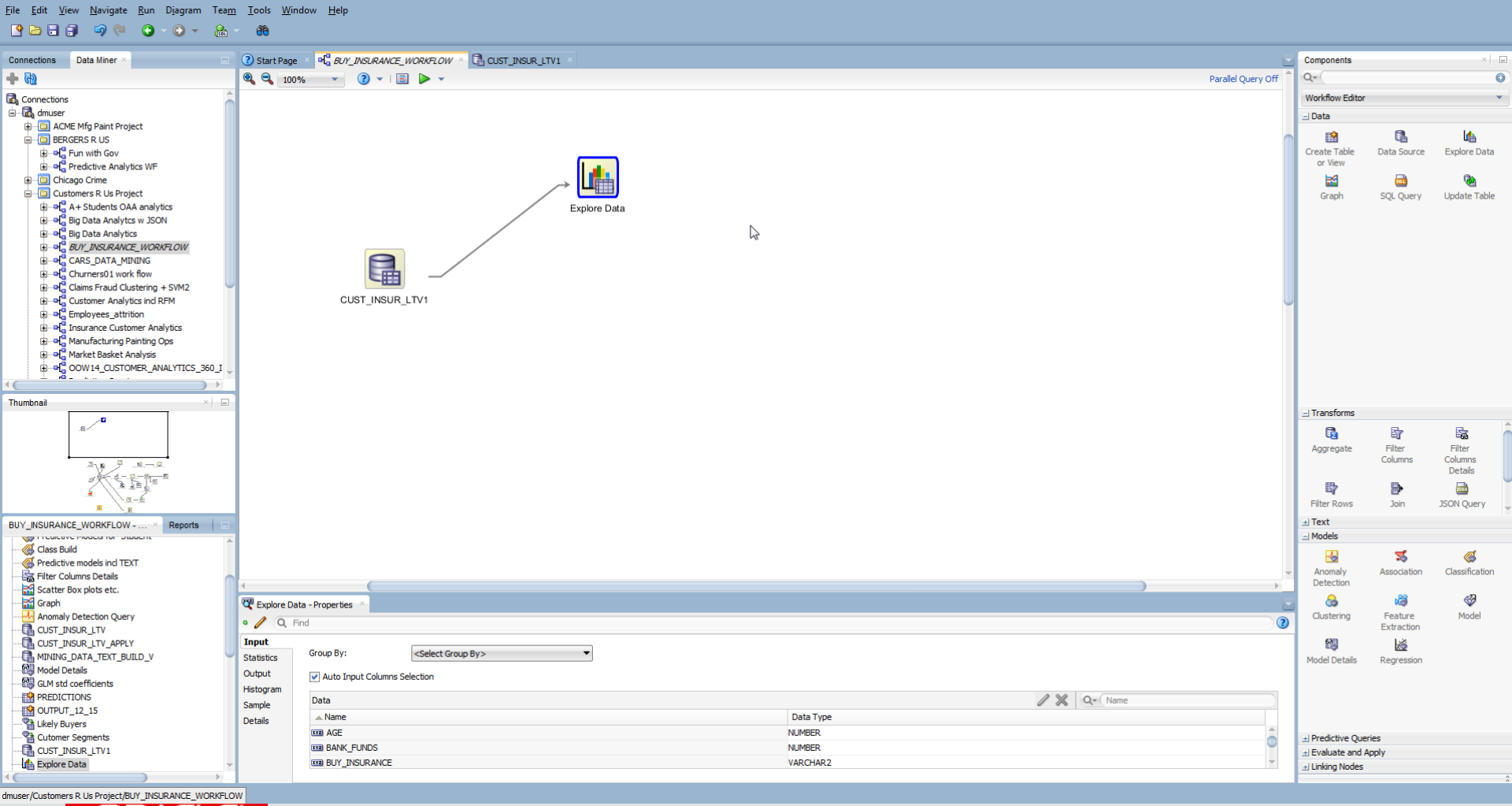
Linking Nodes

dmuser/Customers R Us Project/BUY_INSURANCE_WORKFLOW





	CUST_ID	N_TRANS_ATM	CHECKING_AMOUNT	MARITAL_STATUS	SEX	N_TRANS_KIOSK	STATE	HOUSE_OWNERSHIP	MONTHLY_CHECKS_WRITTEN	LTV	SALARY	BANK_FUNDS	BUY_INSURANCE	CREDIT_BALANCE	N_MORTGAG
1	CU2404	6	25	WIDOWED	F	9	DC	1	2	25,370...	63,883	25,000	No	549	
2	CU2405	0	25	MARRIED	M	2	MI	1	0	29,594.5	59,978	0	No	0	
3	CU2406	5	25	WIDOWED	M	2	CA	1	4	26,473.5	67,894	17,200	Yes	13,859	
4	CU2407	2	3,226	DIVORCED	F	2	MI	1	2	25,976...	67,107	7,600	No	0	
5	CU2408	4	130	DIVORCED	F	2	NY	1	11	28,446.5	70,986	1,500	No	23,525	
6	CU2409	4	6,008	DIVORCED	F	2	CA	0	2	13,921...	58,887	10,900	No	0	
7	CU2411	4	10,943	MARRIED	M	2	NY	1	2	23,833...	69,335	0	No	4,734	
8	CU2412	5	25	MARRIED	F	2	FL	1	12	22,485.5	58,342	2,600	Yes	0	
9	CU2413	4	293	SINGLE	M	1	NY	0	2	18,865...	67,061	1,614	Yes	2,249	
10	CU2414	2	25	DIVORCED	M	2	CA	1	2	29,672...	67,489	0	No	0	
11	CU2416	2	25	SINGLE	M	2	NY	0	4	20,190...	63,963	0	No	0	
12	CU2417	2	2,420	SINGLE	M	2	NY	0	4	19,941	64,964	1,156	No	963	
13	CU2418	2	1,089	DIVORCED	F	2	MI	1	3	27,892...	58,771	6,850	No	0	
14	CU2420	5	25	DIVORCED	M	2	NY	1	3	19,536.5	60,946	5,400	No	44,020	
15	CU2421	5	882	DIVORCED	F	2	CA	1	4	29,123.5	67,694	3,950	No	0	
16	CU2422	3	157	MARRIED	F	2	MI	1	4	29,145...	61,781	1,850	Yes	0	
17	CU2423	4	9,894	WIDOWED	F	8	NY	1	12	25,868...	63,875	3,300	No	6,707	
18	CU3100	6	227	WIDOWED	F	2	MI	1	9	17,539...	65,359	4,280	No	0	
19	CU3101	4	40	MARRIED	F	2	MI	1	11	23,178...	61,513	800	Yes	0	
20	CU3102	4	21,094	SINGLE	M	1	NV	0	4	21,644...	65,379	8,129	No	0	
21	CU3104	5	25	MARRIED	F	2	NY	1	3	16,525.5	66,902	17,001	No	92,408	
22	CU3105	7	638	OTHER	F	2	NY	1	3	24,184.5	63,538	20,200	Yes	0	
23	CU3106	4	372	DIVORCED	M	2	MI	1	12	26,580.5	69,522	0	No	0	
24	CU3107	0	25	SINGLE	M	2	CA	0	0	19,374	64,296	0	No	0	
25	CU3108	3	25	MARRIED	F	2	NY	1	4	14,843...	60,575	700	Yes	0	
26	CU3109	4	25	MARRIED	M	2	MI	1	1	30,839.5	74,158	0	No	0	
27	CU3110	4	25	MARRIED	F	2	CA	1	2	21,882...	59,929	2,300	No	36,102	
28	CU3111	4	25	DIVORCED	M	2	MN	1	7	27,128...	55,713	3,650	No	0	
29	CU3113	1	25	SINGLE	M	2	MS	0	1	15,834.5	64,138	0	No	0	
30	CU3115	2	25	MARRIED	M	3	CA	1	3	27,130	66,120	1,600	No	0	
31	CU3116	4	17,173	DIVORCED	F	2	CA	1	18	24,554	68,616	4,000	No	11,098	
32	CU3117	4	13,225	SINGLE	M	1	MI	0	11	16,890	65,560	1,606	No	0	
33	CU3118	4	7,543	SINGLE	M	3	CA	1	1	28,260...	71,843	700	No	0	
34	CU3119	4	25	DIVORCED	F	3	MI	1	3	27,410	58,840	0	No	0	
35	CU3120	2	25	SINGLE	M	3	CA	0	1	0	63,095	0	No	0	
36	CU3121	4	543	DIVORCED	M	3	CA	1	4	28,178...	60,315	5,700	No	0	
37	CU3123	1	25	SINGLE	M	3	NY	0	1	24,367.5	72,670	0	No	0	
38	CU3124	4	12,988	SINGLE	M	3	MN	0	2	17,477.5	70,310	10,100	No	0	
39	CU3125	6	25	DIVORCED	M	3	CA	1	16	22,995	68,380	5,950	No	66,417	
40	CU3126	2	108	DIVORCED	F	3	MI	1	9	24,825	60,900	7,200	No	0	



File Edit View Navigate Run Team Tools Window Help

Connections | Data Miner

- Connections
  - dmuser
    - ACME Mfg Paint Project
    - BERGERS R US
      - Fun with Gov
      - Predictive Analytics WF
    - Chicago Crime
    - Customers R Us Project
      - A+ Students OAA analytics
      - Big Data Analytcs w JSON
      - Big Data Analytics
      - BUY_INSURANCE_WORKFLOW
      - CARS_DATA_MINING
      - Churners01 work flow
      - Claims Fraud Clustering + SVM2
      - Customer Analytics incl RFM
      - Employees_attrition
      - Insurance Customer Analytics
      - Manufacturing Painting Ops
      - Market Basket Analysis
      - OOW14_CUSTOMER_ANALYTICS_360_I
      - Predictive Queries
      - Public Sector Tax Audit
      - R_regression_models
      - Star Schema360 Degree Customer

Explore Data - Structure | Reports

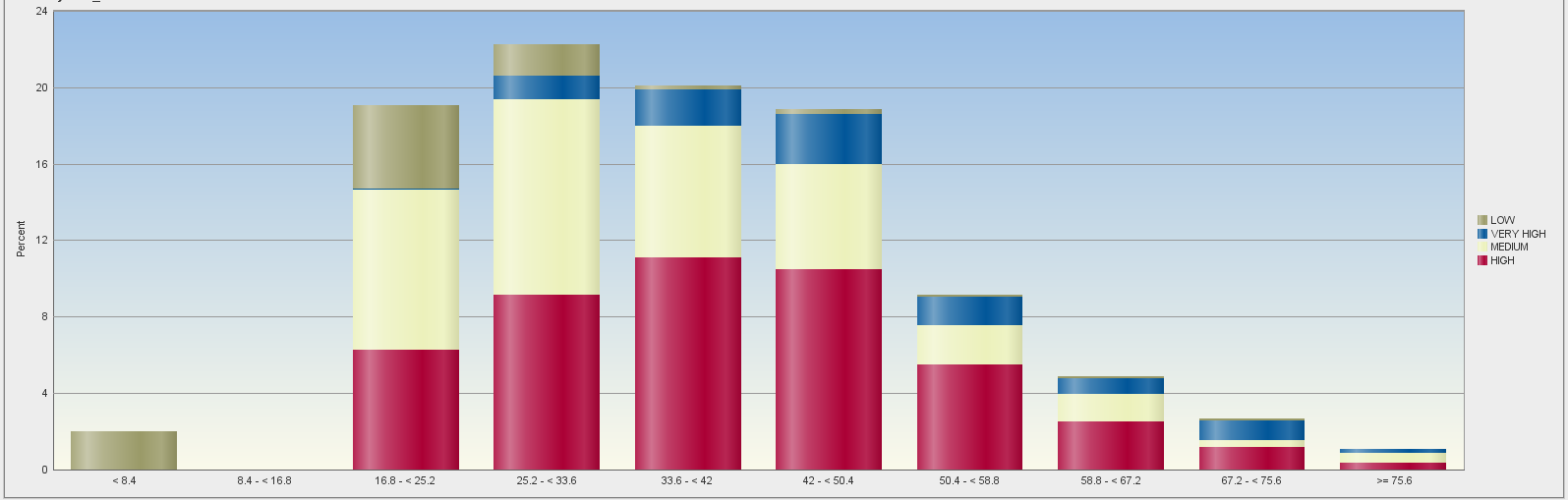
No Structure

Start Page | BUY_INSURANCE_WORKFLOW | CUST_INSUR_LTV1 | Explore Data

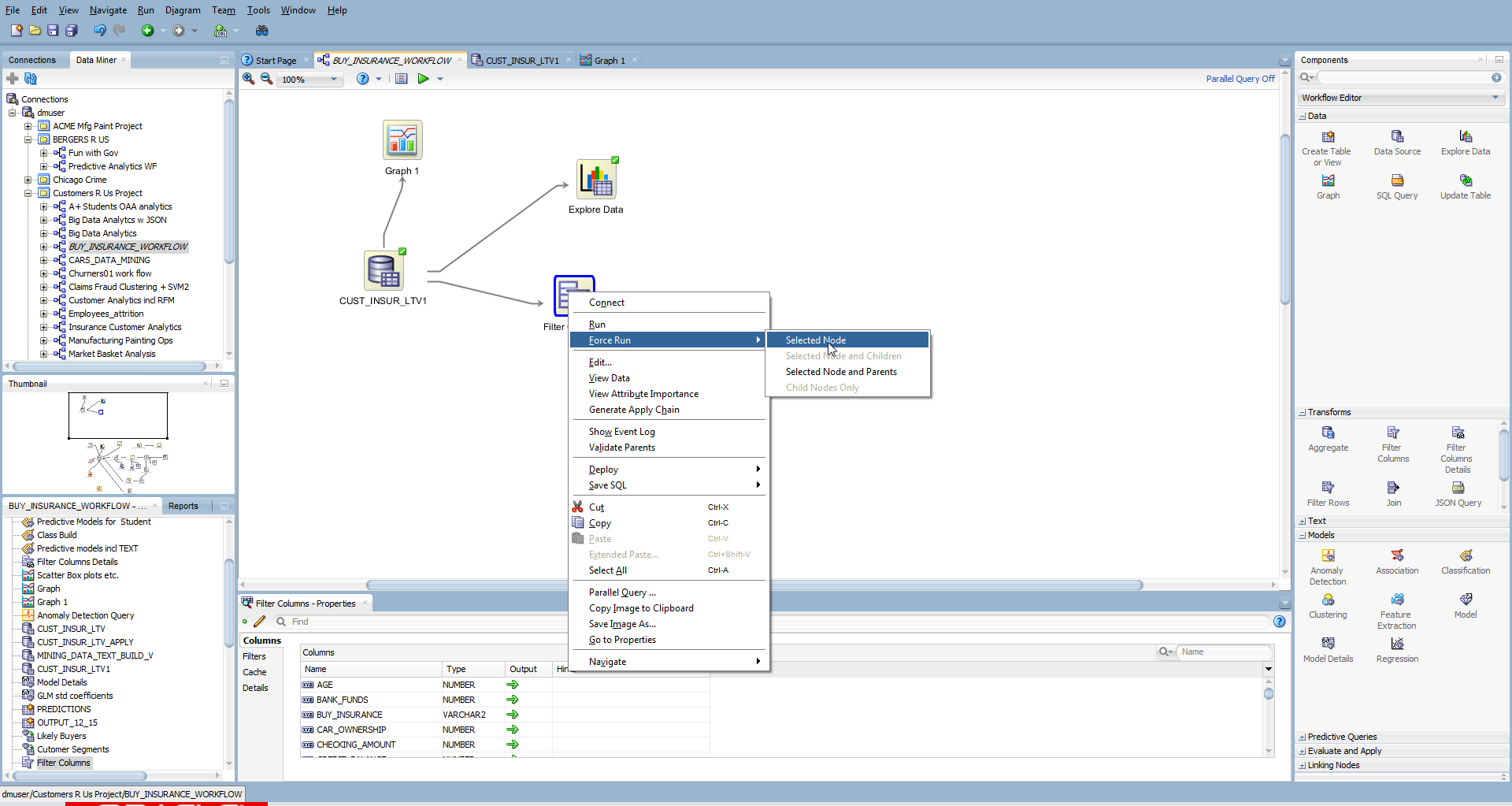
Statistics | Data | Columns | SQL

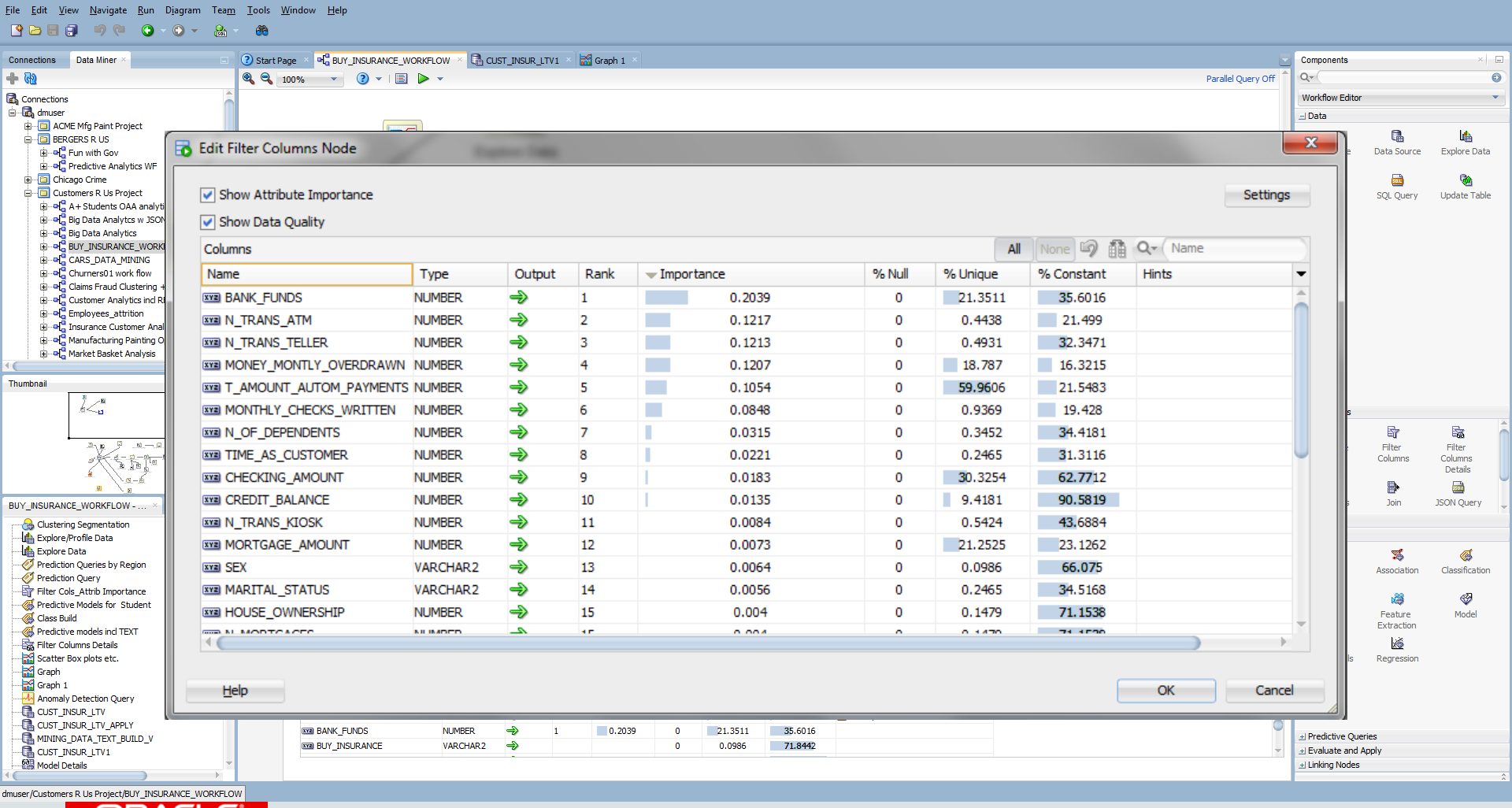
Statistics: 10 Columns from 2,005 Rows(Sampled) Group by: LTV_BIN Show Nulls Filter: Name

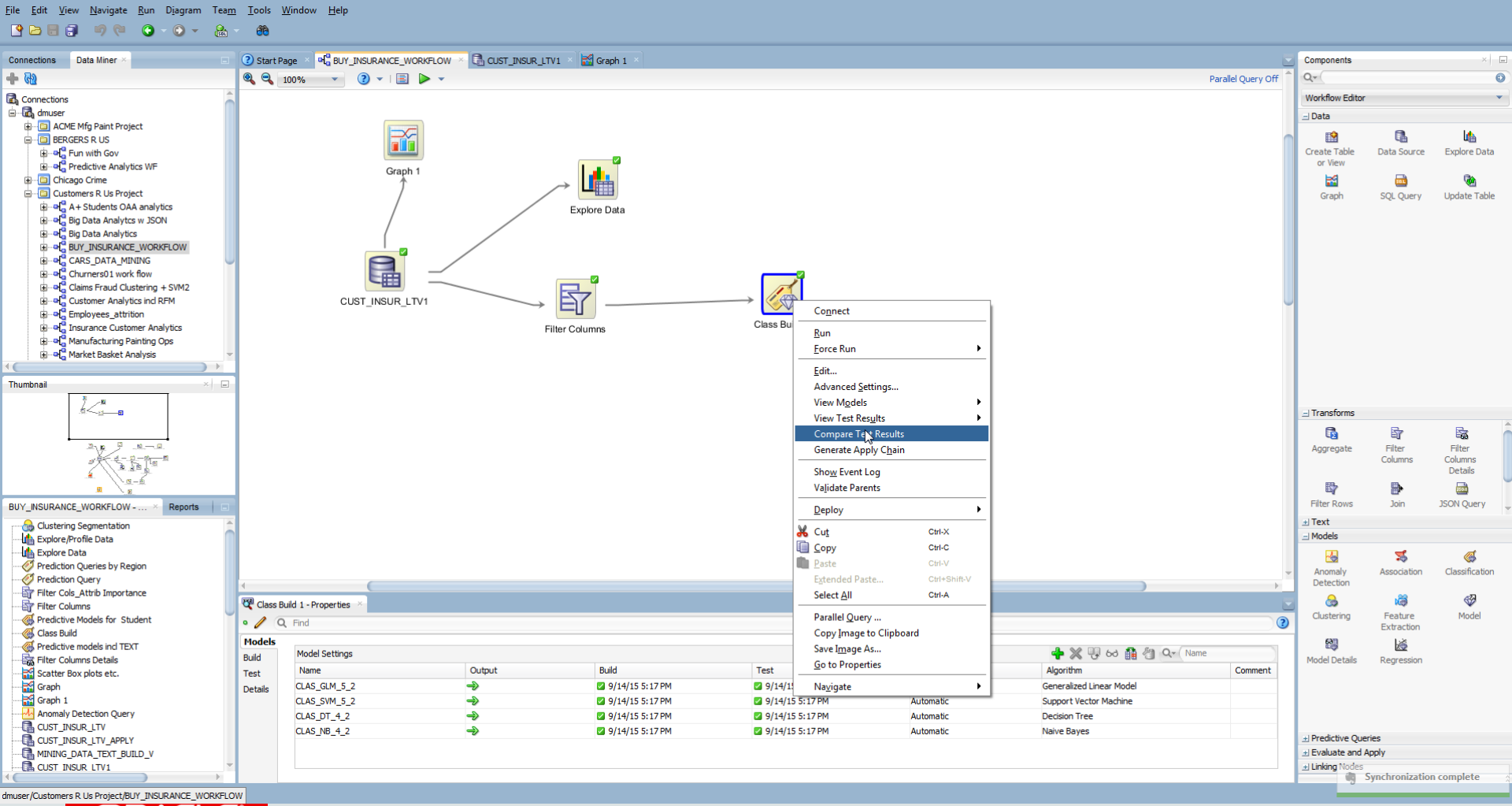
Name	Histogram	Data Type	Percent NULLs	Distinct Values	Distinct Percent	Mode	Average	Median	Min Value	Max Value	Standard Deviation	Variance
AGE		NUMBER	0	69	3.4414		37.6823	36	0	84	14.4737	209.4873
BANK_FUNDS		NUMBER	0	425	21.197		2,585.3736	551	0	30,200	4,739.2687	22,460,667.7132
BUY_INSURANCE		VARCHAR2	0	2	0.0998	No						
CAR_OWNERSHIP		NUMBER	0	2	0.0998		0.9247	1	0	1	0.264	0.0697
CHECKING_AMOUNT		NUMBER	0	626	31.2219		1,032.7312	25	25	24,471	3,103.6812	9,632,836.6997
CREDIT_BALANCE		NUMBER	0	187	9.3267		2,369.3825	0	0	185,775	12,753.4182	162,649,675.2952
CREDIT_CARD_LIMITS		NUMBER	0	26	1.2968		1,266.384	1,000	500	5,000	814.3591	663,180.7806
CUST_ID		VARCHAR2	0	2,005	100	<Other>						
FIRST		VARCHAR2	0	1,379	68.7781	<Other>						
HAS_CHILDREN		NUMBER	0	2	0.0998		0.5172	1	0	1	0.4998	0.2498
HOUSE_OWNERSHIP		NUMBER	0	3	0.1496		0.79	1	0	2	0.5116	0.2618



dmuser/Customers R Us Project/BUY_INSURANCE_WORKFLOW







File Edit View Navigate Run Team Tools Window Help

Connections | Data Miner

- Connections
  - dmuser
    - ACME Mfg Paint Project
    - BERGERS R US
      - Fun with Gov
      - Predictive Analytics WF
    - Chicago Crime
    - Customers R Us Project
      - A+ Students OAA analytics
      - Big Data Analytcs w JSON
      - Big Data Analytics
      - BUY_INSURANCE_WORKFLOW
      - CARS_DATA_MINING
      - Churners01 work flow
      - Claims Fraud Clustering + SVM2
      - Customer Analytics incl RFM
      - Employees_attrition
      - Insurance Customer Analytics
      - Manufacturing Painting Ops
      - Market Basket Analysis
      - OOW14_CUSTOMER_ANALYTICS_360_I
      - Predictive Queries
      - Public Sector Tax Audit
      - R_regression_models

Class Build 1 - Structure | Reports

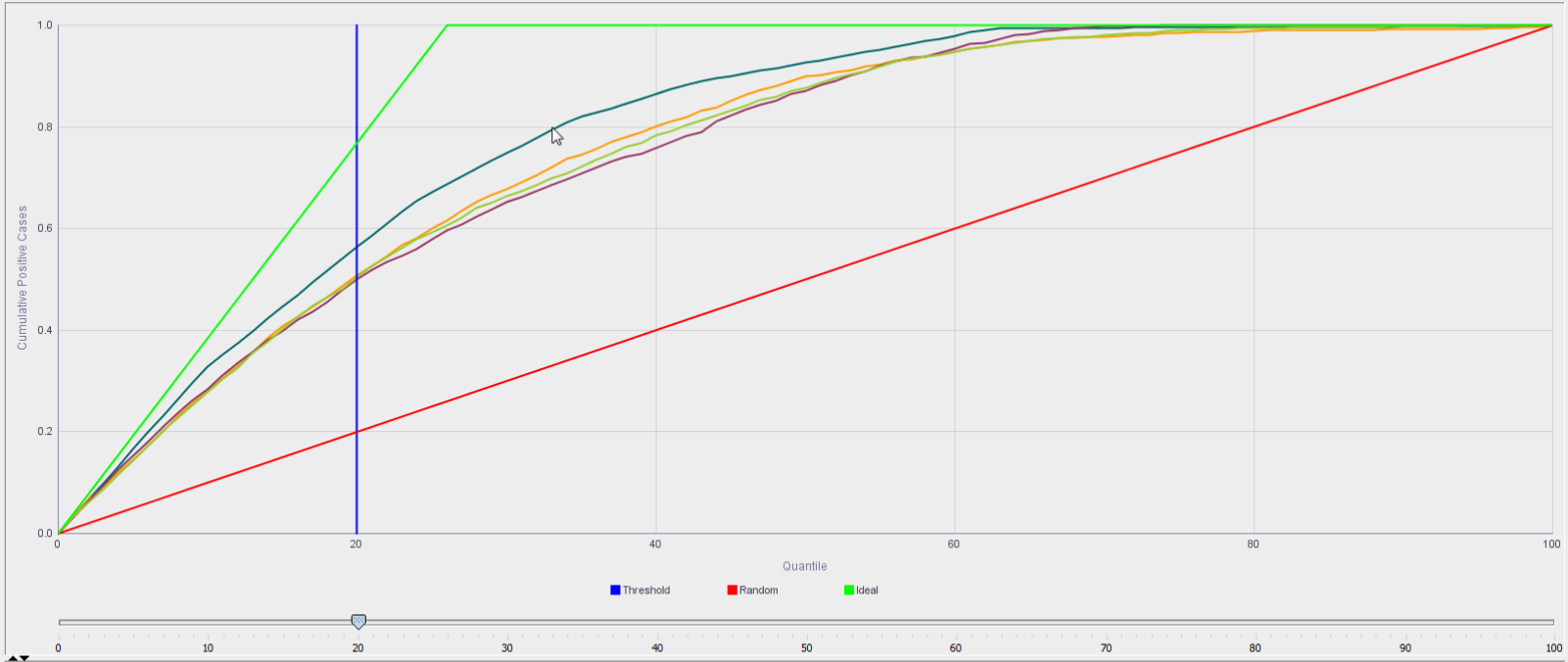
No Structure

Start Page | BUY_INSURANCE_WORKFLOW | CUST_INSUR_LTV1 | Graph 1 | Class Build 1

Performance | Performance Matrix | ROC | Lift | Profit

Display: Cumulative Positive Cases

Target Value: Yes



Models

Model	Lift Cumulative	Gain Cumulative %	Records Cumulative %	Target Density Cumulative	Algorithm	Creation Date
CLAS_DT_4_2	2.7885	56.3528	20.2093	0.7518	Decision Tree	9/14/15 5:17 PM
CLAS_GLM_5_2	2.5146	50.8187	20.2093	0.678	Generalized Linear Model	9/14/15 5:17 PM
CLAS_NB_4_2	2.4726	49.9697	20.2093	0.6667	Naive Bayes	9/14/15 5:17 PM
CLAS_SVM_5_2	2.4936	50.3942	20.2093	0.6723	Support Vector Machine	9/14/15 5:17 PM

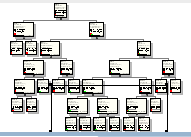
dmuser/Customers R Us Project/BUY_INSURANCE_WORKFLOW

File Edit View Navigate Run Diagram Team Tools Window Help

Connections | Data Miner

Connections
- dmuser
  - ACME Mfg Paint Project
  - BERGERS R US
    - Fun with Gov
    - Predictive Analytics WF
  - Chicago Crime
  - Customers R Us Project
    - A+ Students OAA analytics
    - Big Data Analytcs w JSON
    - Big Data Analytics
    - BUY_INSURANCE_WORKFLOW
    - CARS_DATA_MINING
    - Churners01 work flow
    - Claims Fraud Clustering + SVM2
    - Customer Analytics incl RFM
    - Employees_attrition
    - Insurance Customer Analytics
    - Manufacturing Painting Ops
    - Market Basket Analysis

Thumbnail



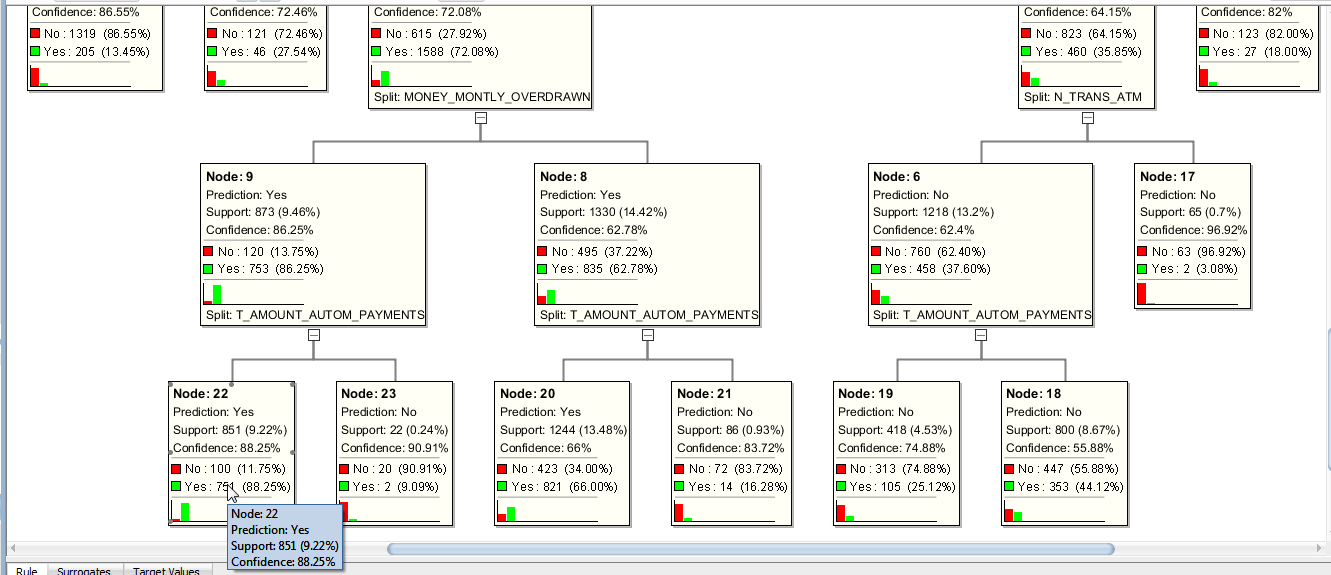
CLAS_DT_4_2 - Structure | Reports

- Node: 0
- Node: 1
- Node: 15
- Node: 16
- Node: 2
- Node: 3
- Node: 4
- Node: 5
- Node: 17
- Node: 6
- Node: 18
- Node: 19
- Node: 7
- Node: 8
- Node: 20
- Node: 21
- Node: 9
- Node: 22
- Node: 23
- Node: 10

Start Page | BUY_INSURANCE_WORKFLOW | CUST_INSUR_LTV1 | Graph 1 | CLAS_DT_4_2

Tree | Settings

100% | Maximum Target Values: 2 | Save Rules...



Rule | Surrogates | Target Values

Node Rule: Wrap

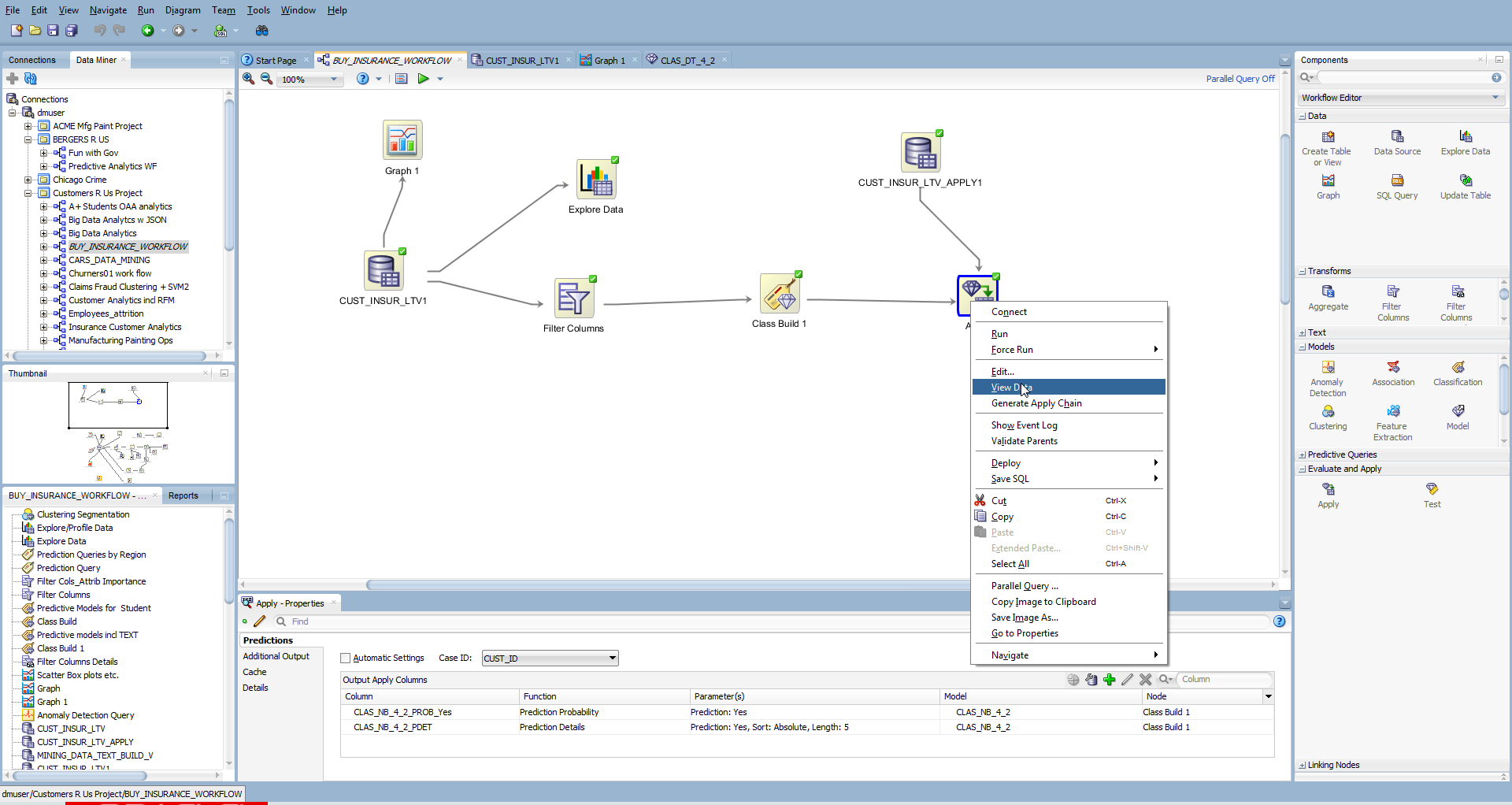
```
If   BANK_FUNDS > 246
And  CHECKING_AMOUNT <= 282
And  CREDIT_BALANCE <= 2445
And  MONEY_MONTLY_OVERDRAWN > 54.095
And  T_AMOUNT_AUTOM_PAYMENTS <= 14993
Then Yes
```

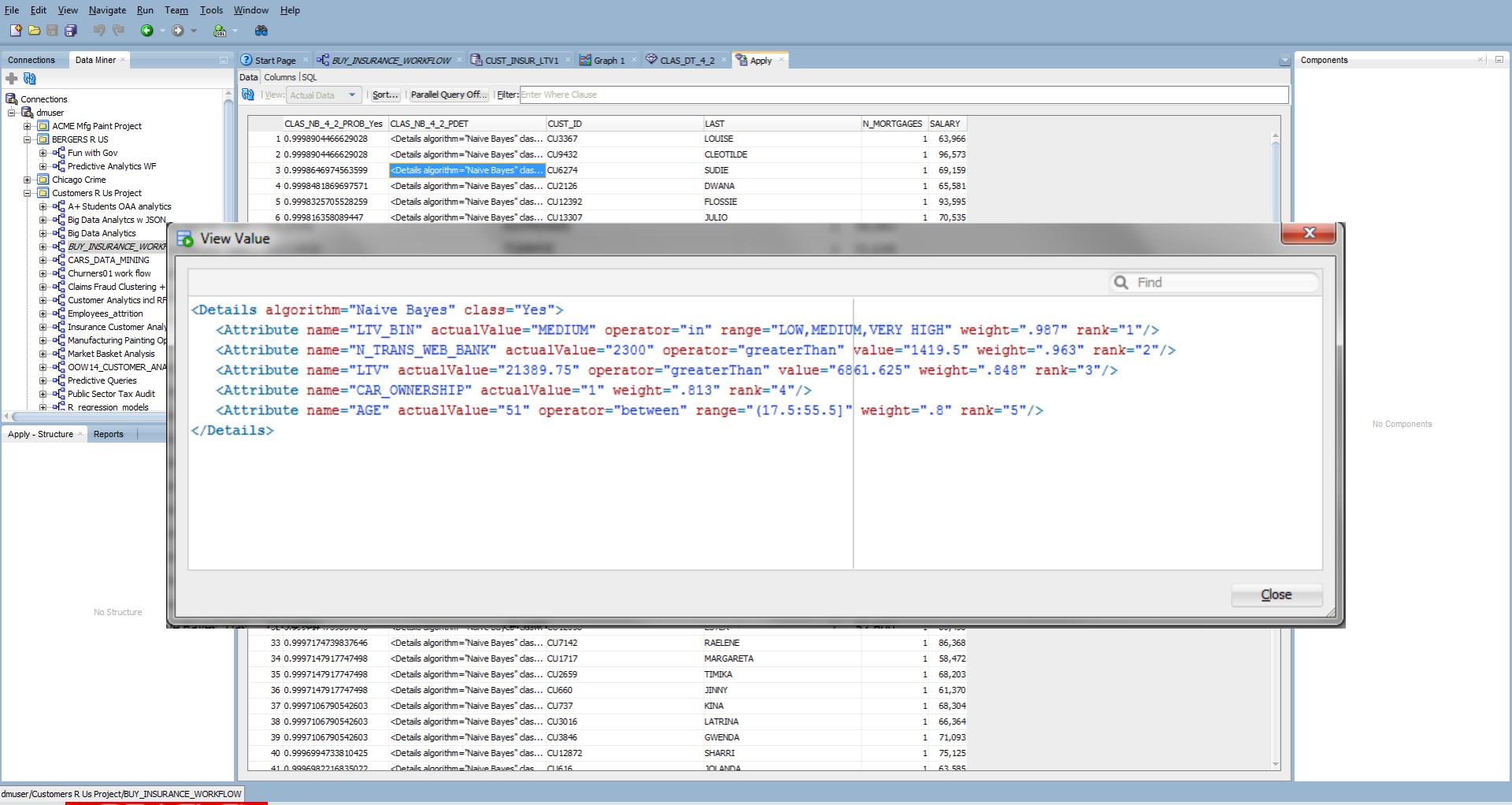
Confidence	0.882491186839013
Support	0.09223932365055279

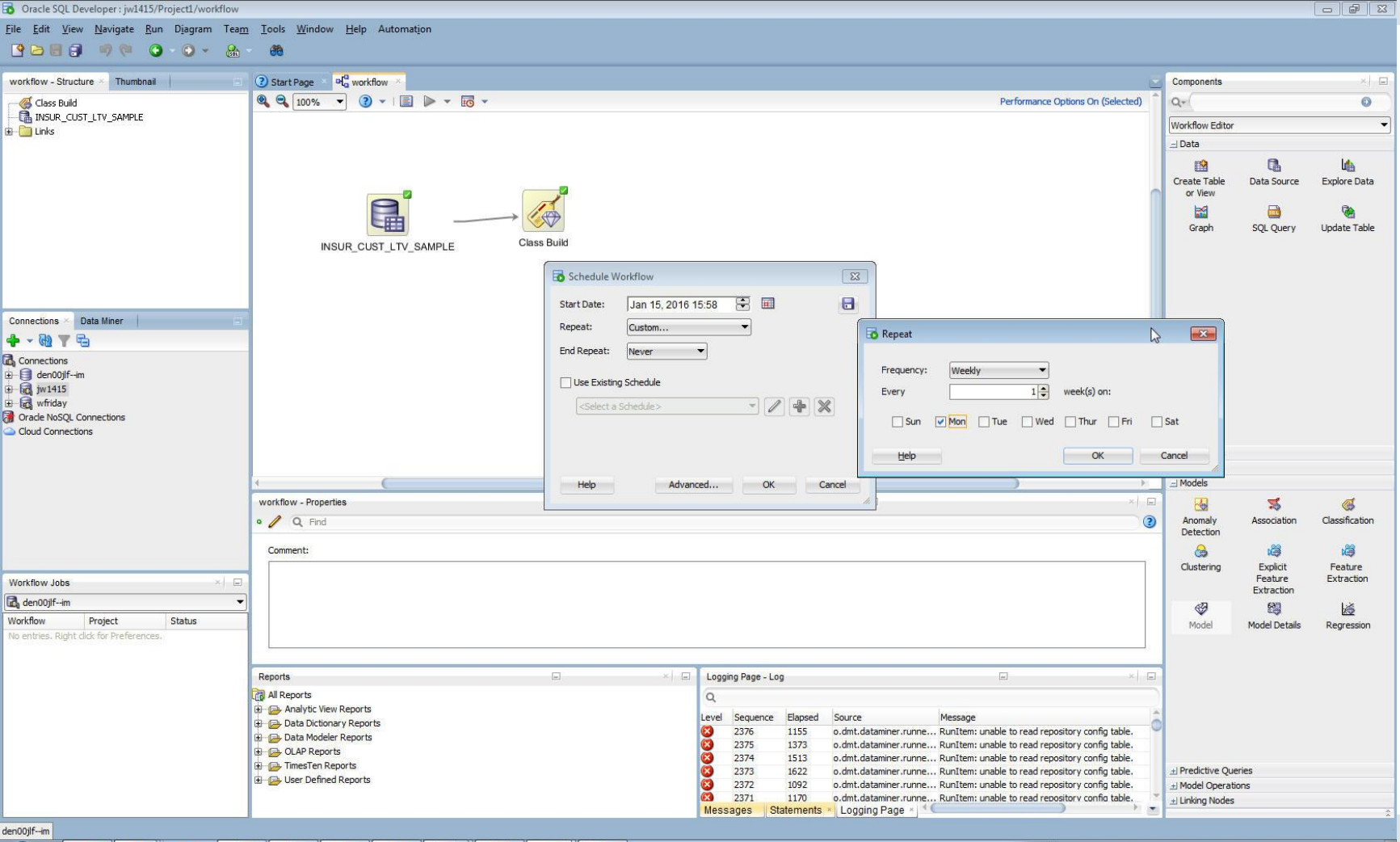
Components

No Components

dmuser/Customers R Us Project/BUY_INSURANCE_WORKFLOW






Oracle SQL Developer : jw1415/Project1/workflow
File Edit View Navigate Run Diagram Team Tools Window Help Automation
workflow - Structure
Thumbnail
Class Build
INSUR_CUST_LTV_SAMPLE
Links
Start Page
workflow
100%
Performance Options On (Selected)
INSUR_CUST_LTV_SAMPLE
Class Build
Schedule Workflow
Start Date:
Jan 15, 2016 15:58
Repeat:
Custom...
End Repeat:
Never
Use Existing Schedule
<Select a Schedule>
Help
Advanced...
OK
Cancel
Repeat
Frequency:
Weekly
Every
1
week(s) on:
Sun
Mon
Tue
Wed
Thur
Fri
Sat
Help
OK
Cancel
Components
Workflow Editor
Data
Create Table or View
Data Source
Explore Data
Graph
SQL Query
Update Table
Models
Anomaly Detection
Association
Classification
Clustering
Explicit Feature Extraction
Feature Extraction
Model
Model Details
Regression
Predictive Queries
Model Operations
Linking Nodes
Connections
Data Miner
Connections
den00jlf--im
jw1415
wfriday
Oracle NoSQL Connections
Cloud Connections
Workflow Jobs
den00jlf--im
Workflow
Project
Status
No entries. Right click for Preferences.
workflow - Properties
Find
Comment:
Reports
All Reports
Analytic View Reports
Data Dictionary Reports
Data Modeler Reports
OLAP Reports
TimesTen Reports
User Defined Reports
Logging Page - Log
Level
Sequence
Elapsed
Source
Message
2376 1155 o.dmt.dataminer.runne... RunItem: unable to read repository config table.
2375 1373 o.dmt.dataminer.runne... RunItem: unable to read repository config table.
2374 1513 o.dmt.dataminer.runne... RunItem: unable to read repository config table.
2373 1622 o.dmt.dataminer.runne... RunItem: unable to read repository config table.
2372 1092 o.dmt.dataminer.runne... RunItem: unable to read repository config table.
2371 1170 o.dmt.dataminer.runne... RunItem: unable to read repository config table.
Messages
Statements
Logging Page
den00jlf--im
EN
3:02 PM
1/15/2016

Fraud Prediction Demo

Automated In-DB Analytical Methodology

```
drop table CLAIMS_SET;
exec dbms_data_mining.drop_model('CLAIMSMODEL');
create table CLAIMS_SET (setting_name varchar2(30), setting_value varchar2(4000));
insert into CLAIMS_SET values ('ALGO_NAME','ALGO_SUPPORT_VECTOR_MACHINES');
insert into CLAIMS_SET values ('PREP_AUTO','ON');
commit;
```

```
begin
dbms_data_mining.create_model('CLAIMSMODEL', 'CLASSIFICATION',
  'CLAIMS', 'POLICYNUMBER', null, 'CLAIMS_SET');
end;
/
```

```
-- Top 5 most suspicious fraud policy holder claims
select * from
(select POLICYNUMBER, round(prob_fraud*100,2) percent_fraud,
      rank() over (order by prob_fraud desc) rnk from
(select POLICYNUMBER, prediction_probability(CLAIMSMODEL, '0' using *) prob_fraud
from CLAIMS
where PASTNUMBEROFCLAIMS in ('2to4', 'morethan4')))
where rnk <= 5
order by percent_fraud desc;
```

POLICYNUMBER	PERCENT_FRAUD	RNK
6532	64.78	1
2749	64.17	2
3440	63.22	3
654	63.1	4
12650	62.36	5

Automated Monthly "Application"! *Just add:*

Create
View CLAIMS2_30
As
Select * from CLAIMS2
Where mydate > SYSDATE – 30

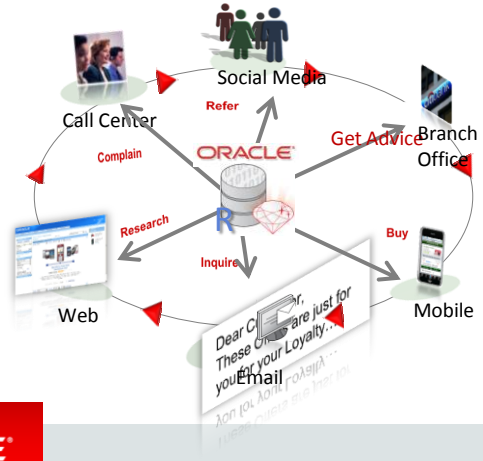
Time measure: set timing on;

Oracle Advanced Analytics

More Details

- On-the-fly, single record apply with new data (e.g. from call center)

```
Select prediction_probability(CLAS_DT_1_2, 'Yes'
  USING 7800 as bank_funds, 125 as checking_amount, 20 as
  credit_balance, 55 as age, 'Married' as marital_status,
  250 as MONEY_MONTLY_OVERDRAWN, 1 as house_ownership)
from dual;
```



Likelihood to respond:

Query Result

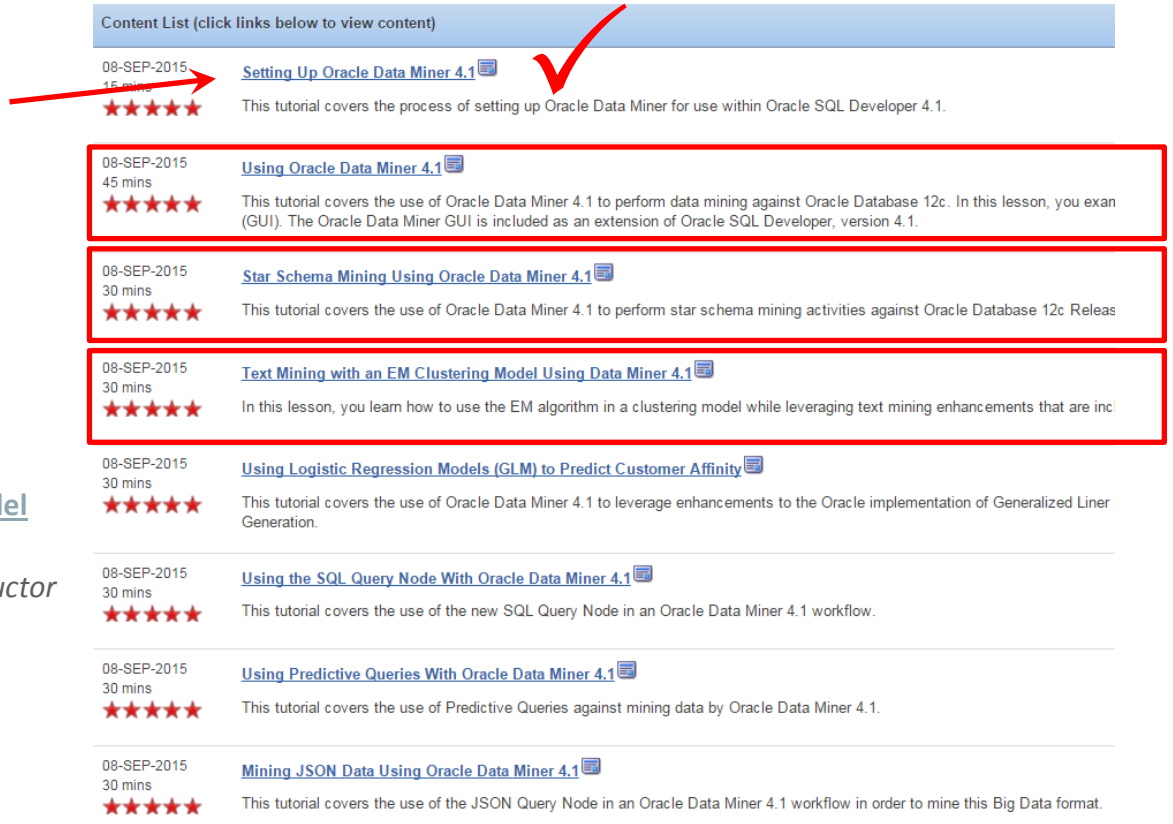
SQL | All Rows Fetched: 1 in 0 seconds

PREDICTION_PROB...
0.8382936507936...

OAA/Oracle Data Miner 4.1 HOL

Uses Oracle by Example Free Online Tutorials

- There are 6 Tutorials
  - The first tutorial is already done for you
  - Recommend doing 3-5 Tutorials
    1. Using Oracle Data Miner 4.1
    2. Star Schema Mining Using Oracle Data Miner 4.1
    3. Text Mining with an EM Clustering Model Using Data Miner 4.1
    4. **Anomaly Detection (CLAIMS)** *See Instructor for assistance*
    5. **Market Basket Analysis (SH.SALES)** *See Instructor for assistance*




ORACLE®