



**SOFTWARE SOLUTIONS**



# **Oracle Data Mining Hands On Lab**

**Central States AUG 2012**

**Tim Vlamis**

**Vlamis Software Solutions**

**816-781-2880**

**<http://www.vlamis.com>**



# VlamiS Software Solutions

- Founded in 1992 by Dan VlamiS in Kansas City, MO
- Developed/implemented more than 200 Oracle BI systems
- Specializes in ORACLE-based:
  - Business Intelligence
  - Analytic Options to Oracle DB (OLAP, Data Mining, Spatial)
  - Data Warehousing
  - Training and mentoring
- Expert presenter at major Oracle conferences
- [www.vlamiS.com](http://www.vlamiS.com) (blog, papers, newsletters, services)
- Co-authored book “Oracle Essbase & Oracle OLAP”
- Beta tester for OBIEE 11g
- Reseller for Simba and NAVTEQ map data for OBIEE
- HOL Coordinator for 2012 Collaborate Conference





# Tim Vlami's Bio

- 20+ years experience in business modeling and valuation, forecasting, and scenario analyses
- Trainer for Oracle University Two-Day Data Mining Course
- Professional Certified Marketer (PCM) from AMA
- Active Member of NICO (Northwestern Institute on Complex Systems)
- Adjunct Professor of Business Benedictine College
- MBA Kellogg School of Management (Northwestern)
- BA Economics Yale University

[tvlamis@vlamis.com](mailto:tvlamis@vlamis.com) 816-781-2880



# Spectrum of Oracle DB BI & Analytics

## OLAP

Summaries,  
hierarchies and  
dimensional data

*“Analysis”*

**What is the**  
average  
**income** of  
mutual fund  
buyers,  
**by** region,  
**by** year?

## Data Mining & R

Knowledge discovery  
of hidden patterns

*“Insight & Prediction”*

**Who is likely to**  
**purchase** a mutual  
fund in the next 6  
months and why?

## Spatial

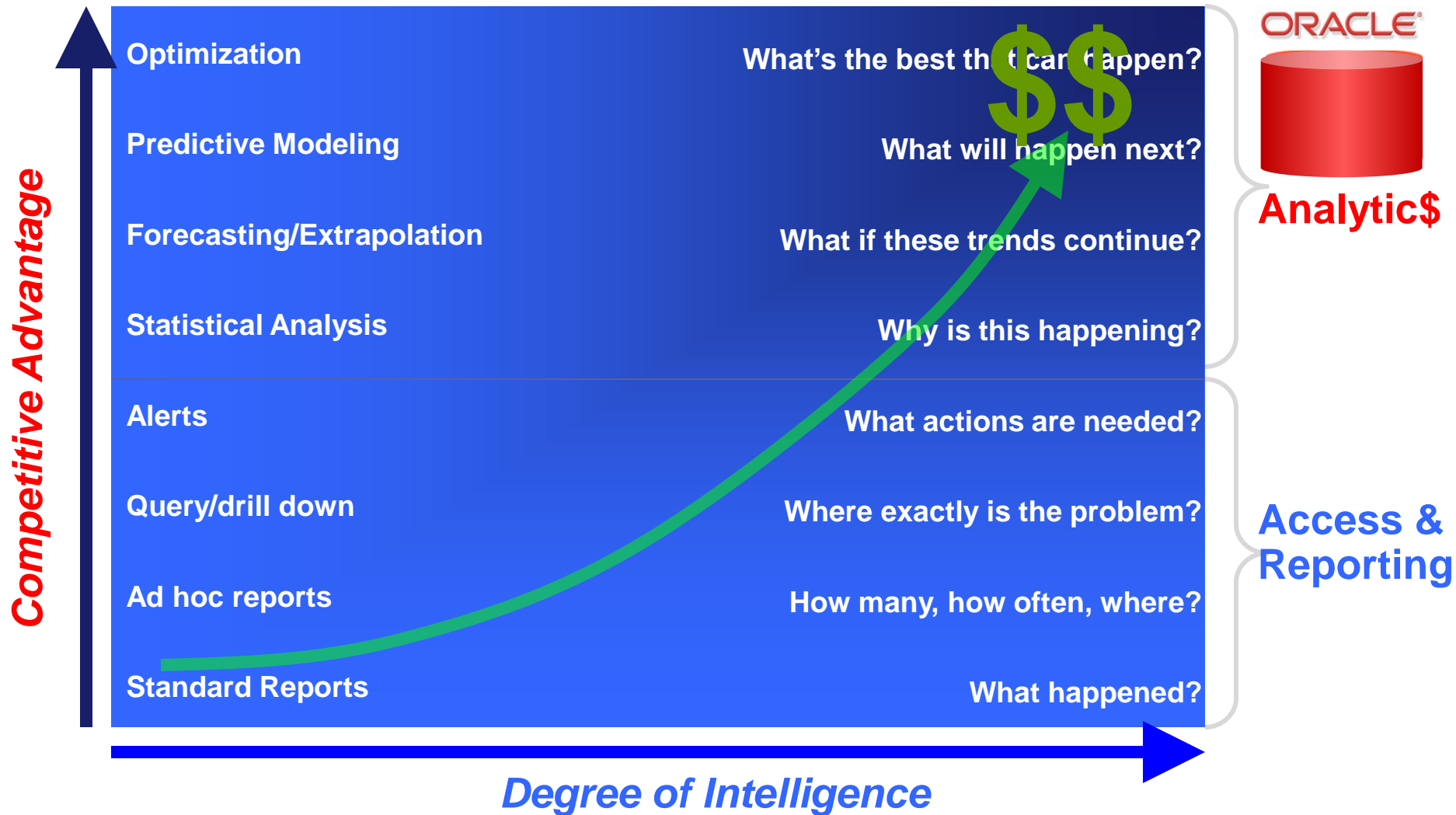
Spatial relationships  
between data

*“Location”*

**Where** were  
mutual funds  
purchased  
in the last 3  
years?



# Competitive Advantage of BI & Analytics



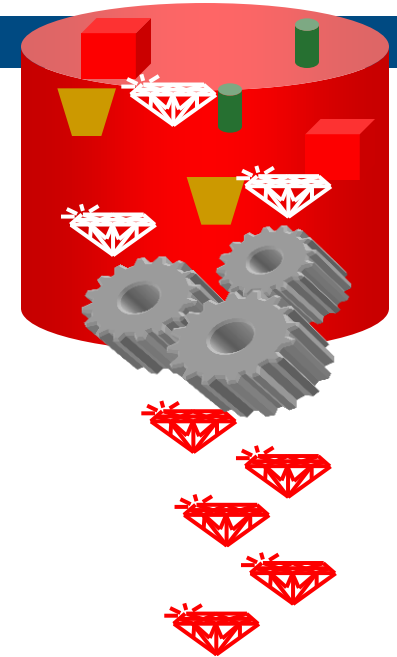
Source: Competing on Analytics, by T. Davenport & J. Harris



# What is Data Mining?



- Automatically sifts through data to find hidden patterns, discover new insights, and make predictions
- Data Mining can provide valuable results:
  - Predict customer behavior (*Classification*)
  - Predict or estimate a value (*Regression*)
  - Segment a population (*Clustering*)
  - Identify factors more associated with a business problem (*Attribute Importance*)
  - Find profiles of targeted people or items (*Decision Trees*)
  - Determine important relationships and “market baskets” within the population (*Associations*)
  - Find fraudulent or “rare events” (*Anomaly Detection*)

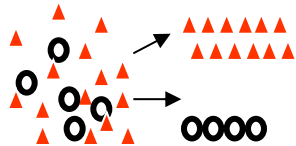
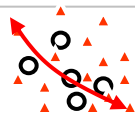
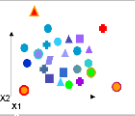
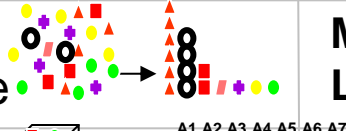
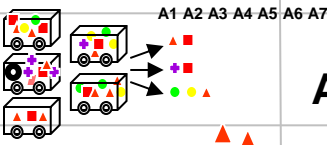
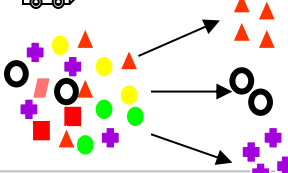
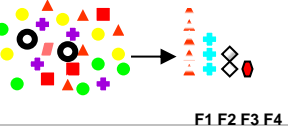




# Oracle Data Mining

- Oracle Data Mining is an option for the Enterprise Edition of the Oracle Database.
- A collection of APIs and specialized SQL functions.
- Includes a large number of specialized algorithms and built-in procedures.
- Makes use of many built-in capabilities of the Oracle Database
- ODM typically refers to “Oracle Data Mining”

# Oracle Data Mining Algorithms

Problem	Algorithm	Applicability
Classification 	<b>Logistic Regression (GLM)</b> <b>Decision Trees</b> <b>Naïve Bayes</b> <b>Support Vector Machine</b>	<b>Classical statistical technique</b> <b>Popular / Rules / transparency</b> <b>Embedded app</b> <b>Wide / narrow data / text</b>
Regression 	<b>Multiple Regression (GLM)</b> <b>Support Vector Machine</b>	<b>Classical statistical technique</b> <b>Wide / narrow data / text</b>
Anomaly Detection 	<b>One Class SVM</b>	<b>Fraud Detection</b>
Attribute Importance 	<b>Minimum Description Length (MDL)</b>	<b>Attribute reduction</b> <b>Identify useful data</b> <b>Reduce data noise</b>
Association Rules 	<b>Apriori</b>	<b>Market basket analysis</b> <b>Link analysis</b>
Clustering 	<b>Hierarchical K-Means</b> <b>Hierarchical O-Cluster</b>	<b>Product grouping</b> <b>Text mining</b> <b>Gene and protein analysis</b>
Feature Extraction 	<b>NMF</b>	<b>Text analysis</b> <b>Feature reduction</b>





# 11g Statistics & SQL Analytics (Free)

- **Ranking functions**
  - rank, dense\_rank, cume\_dist, percent\_rank, ntile
- **Window Aggregate functions**  
(moving and cumulative)
  - Avg, sum, min, max, count, variance, stddev, first\_value, last\_value
- **LAG/LEAD functions**
  - Direct inter-row reference using offsets
- **Reporting Aggregate functions**
  - Sum, avg, min, max, variance, stddev, count, ratio\_to\_report
- **Statistical Aggregates**
  - Correlation, linear regression family, covariance
- **Linear regression**
  - Fitting of an ordinary-least-squares regression line to a set of number pairs.
  - Frequently combined with the COVAR\_POP, COVAR\_SAMP, and CORR functions

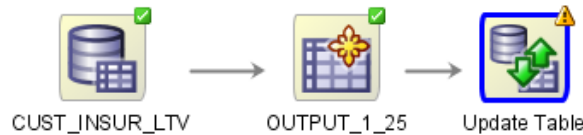
## Descriptive Statistics

- **DBMS\_STAT\_FUNCS:** summarizes numerical columns of a table and returns count, min, max, range, mean, median, stats\_mode, variance, standard deviation, quantile values, +/- n sigma values, top/bottom 5 values
- **Correlations**
  - Pearson's correlation coefficients, Spearman's and Kendall's (both nonparametric).
- **Cross Tabs**
  - Enhanced with % statistics: chi squared, phi coefficient, Cramer's V, contingency coefficient, Cohen's kappa
- **Hypothesis Testing**
  - Student t-test, F-test, Binomial test, Wilcoxon Signed Ranks test, Chi-square, Mann Whitney test, Kolmogorov-Smirnov test, One-way ANOVA
- **Distribution Fitting**
  - Kolmogorov-Smirnov Test, Anderson-Darling Test, Chi-Squared Test, Normal, Uniform, Weibull, Exponential

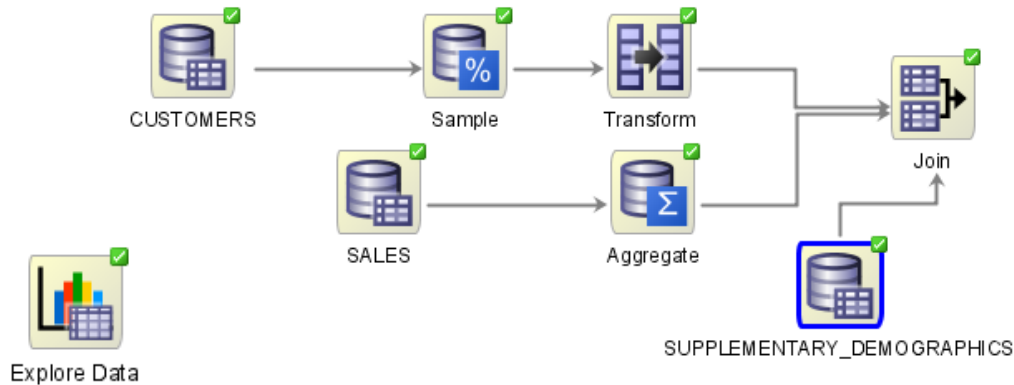


# Oracle Data Miner Nodes (Partial List)

## Tables and Views



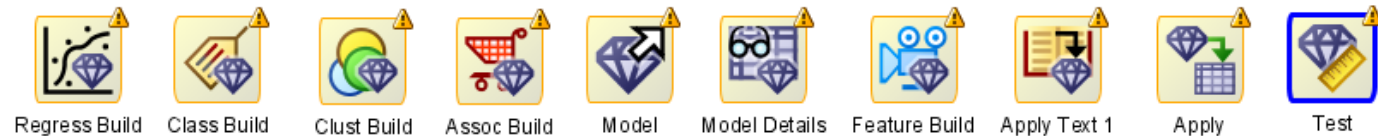
## Transformations



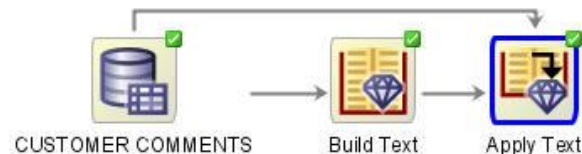
## Explore Data



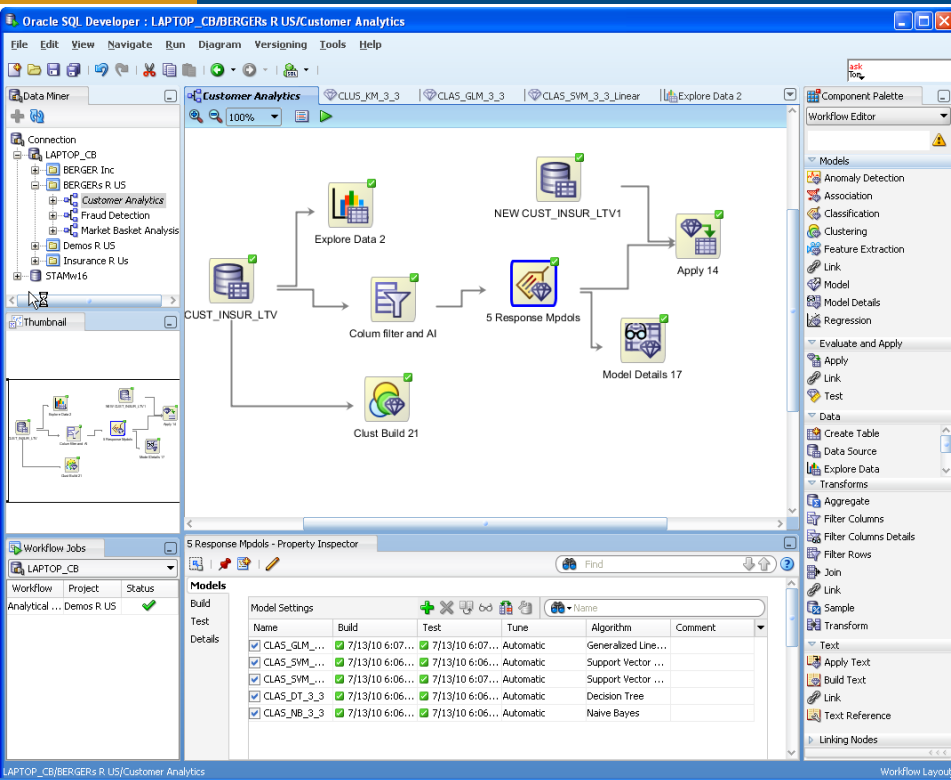
## Modeling



## Text



# Oracle Data Miner 11g Release 2



- Oracle Data Miner is a front end GUI for Oracle Data Mining.
- Extension for Oracle SQL Developer 3.x, a free utility program from Oracle that facilitates interaction with databases.
- Functions as an object oriented programming interface for designing data mining processes and procedures.



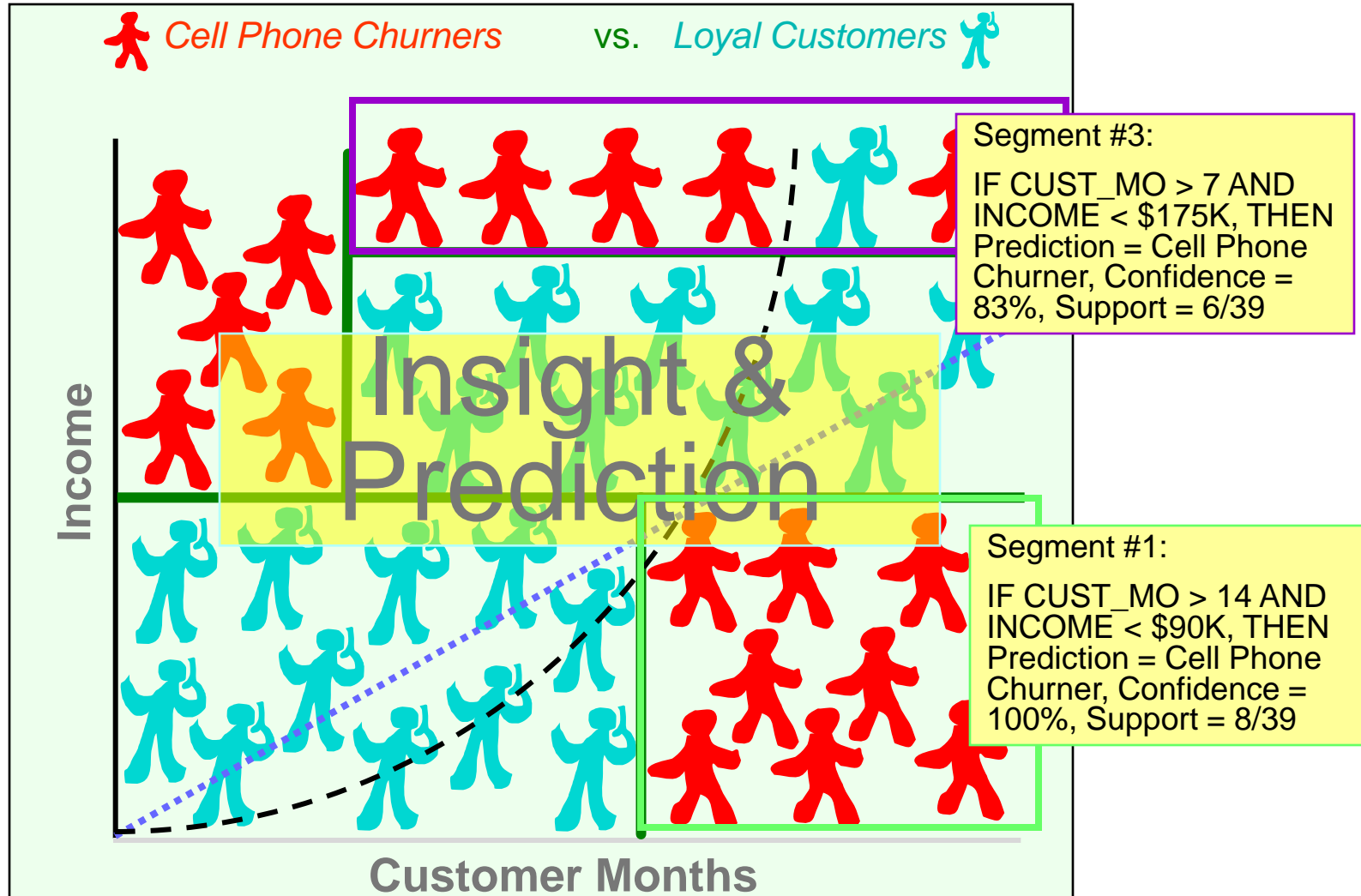
# Classification

- Prediction model for non-continuous information
  - Binary such as yes/no
  - Limited set (low/medium/high)
- Involves “supervised learning”
  - Prediction directed by a previously known dependent variable or “target” variable.
  - Commonly includes three phases:
    - Training
    - Testing
    - Scoring
- Results in predictive models that are applied to new data sets.
- In our example, we predict which prospects are likely to buy insurance.



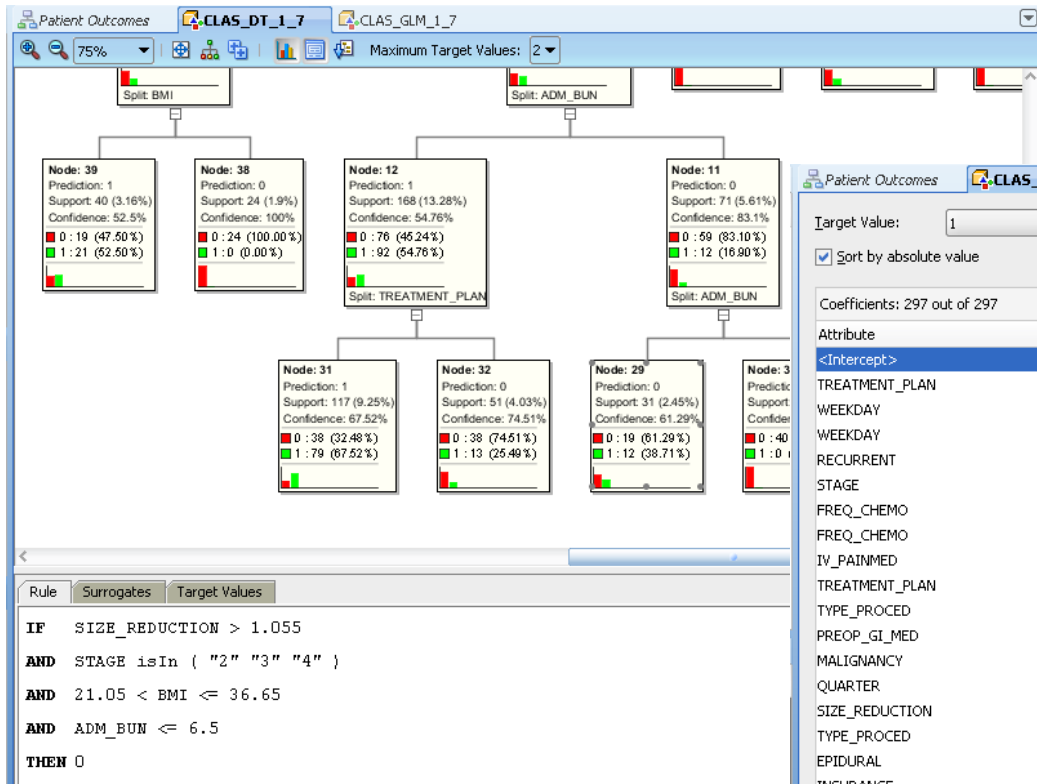
# Data Mining Provides

*Better Information, Valuable Insights and Predictions*



# Understand Model Details

- Interactive model viewers



Outcome predictive models viewer for CLAS\_GLM\_1\_7. It displays a table of coefficients for 297 attributes. The table includes columns for Attribute, Value, Coefficient, Standardized Coefficient, and Exp(Coefficient).

Attribute	Value	Coefficient	Standardized Coefficient	Exp(Coefficient)
<Intercept>	NULL	-1.83481346	0	6.26396556
TREATMENT_PLAN	Chemo_only	-0.46513283	0.11735002	1.59222567
WEEKDAY	W	-0.0697858	0.0869471	1.50227193
WEEKDAY	Th	-0.34941526	0.05883753	1.418238
RECURRENT	1	-0.33993936	0.07348783	1.4048624
STAGE	3	0.29916993	-0.06150948	0.74143341
FREQ_CHEMO	1	0.29378459	-0.06262496	0.74543705
FREQ_CHEMO	0	-0.26376819	0.05597178	1.30182638
IV_PAINMED	DEM	-0.26085980	0.036163	1.29804567
TREATMENT_PLAN	Chemo&Radiation	-0.25534174	0.03324906	1.2909027
TYPE_PROCD	closed	0.25466832	-0.01992872	0.77517356
PREOP_GI_MED	1	0.25194913	-0.06873117	0.77728428
MALIGNANCY	1	0.24061736	-0.05486614	0.78614238
QUARTER	A	0.23306129	-0.05746447	0.79210502
SIZE_REDUCTION	NULL	0.22915110	-0.15356344	0.79520837
TYPE_PROCD	1	-0.22759025	0.03846051	1.25557075
EPIDURAL	1	-0.22715954	0.05119796	1.25503009
INSURANCE	B	0.21168257	-0.05517357	0.80922152
OR_TRANSFUSIONS	1	0.20613024	-0.0550411	0.81372709
TYPE_ABX	Cipro	0.20248206	-0.02044382	0.81670114
EKG	SB	0.19228831	-0.02216336	0.82506896
IV_PAINMED	TORD	-0.19105185	0.01912802	1.21052222
INCISION	KNEE	-0.18882816	0.01878139	1.20783338
INSURANCE	C	0.18859100	-0.02710814	0.82812514
WT_LOSS_TIME	NULL	-0.17535293	0.11368976	1.19166672
WEEKDAY	Sa	0.17096336	-0.02674837	0.84285246



# Oracle Data Mining & OBI 11g

ODM's predictions & probabilities are available in the Database for reporting using Oracle BI EE and other tools

8.4 Oracle Datamining					Home Catalog Dashboards New Open Signed In As				
LTV Prediction LTV Details Classification Tree LTV Probabilities What If Scoring Geo LTV P					Map				
Classification Tree									
Page Information (click to collapse or expand)									
Classification Tree					ails				
Time run: 12/9/2011 1:03:03 PM					3:03 PM				
20 Actual Unit Price									
		2008	2009	2010	Grand Total				
0 - All Individuals	MEDIUM	9,302	9,302	9,382	9,331				
1 - M_MARITAL_ST in 'DIVORCED', 'SINGLE'	MEDIUM	9,207	9,329	9,421	9,322				
2 - M_CRDT_RATE <= 657.5	LOW	9,225	9,164	9,377	9,261				
12 - M_INCOME_LVL in 'LEVEL 5', 'LEVEL 6', 'LEVEL 7', 'LEVEL 8', 'LEVEL 9'	MEDIUM	8,904	9,131	9,670	9,261				
13 - M_INCOME_LVL in 'LEVEL 1', 'LEVEL 2', 'LEVEL 3', 'LEVEL 4'	LOW	9,345	9,176	9,259	9,261				
3 - M_CRDT_RATE > 657.5	MEDIUM	9,193	9,462	9,454	9,370				
14 - M_MONTHS_CONTACT <= 12.5	VERY HIGH	8,815	9,418	8,690	8,951				
4 - M_MONTHS_CONTACT > 12.5	MEDIUM	9,242	9,468	9,543	9,421				
7 - M_MARITAL_ST in 'MARRIED', 'WIDOW'	HIGH	9,397	9,276	9,343	9,341				
1- Revenue									
		2008	2009	2010	Grand Total				
0 - All Individuals	MEDIUM	16,500,000	15,000,000	18,500,000	50,000,000				
1 - M_MARITAL_ST in 'DIVORCED', 'SINGLE'	MEDIUM	8,155,247	7,589,505	9,289,014	25,033,766				
2 - M_CRDT_RATE <= 657.5	LOW	3,560,875	3,340,550	4,015,646	10,917,071				
12 - M_INCOME_LVL in 'LEVEL 5', 'LEVEL 6', 'LEVEL 7', 'LEVEL 8', 'LEVEL 9'	MEDIUM	938,983	889,059	1,189,016	3,017,058				
13 - M_INCOME_LVL in 'LEVEL 1', 'LEVEL 2', 'LEVEL 3', 'LEVEL 4'	LOW	2,621,892	2,451,491	2,826,630	7,900,013				

#	M23 Full Rule	Predicted LTV	# of Cust	1- Revenue	Trend
12	M_MARITAL_ST in 'DIVORCED', 'SINGLE' ; AND M_CRDT_RATE <= 657.5; AND M_INCOME_LVL in 'LEVEL 5', 'LEVEL 6';	MEDIUM	0		
13	M_MARITAL_ST in 'DIVORCED', 'SINGLE' ; AND M_CRDT_RATE <= 657.5; AND M_INCOME_LVL in 'LEVEL 1', 'LEVEL 2';	LOW	0		
14	M_MARITAL_ST in 'DIVORCED', 'SINGLE' ; AND M_CRDT_RATE > 657.5; AND M_MONTHS_CONTACT <= 12.5	VERY HIGH	0		
15	M_MARITAL_ST in 'DIVORCED', 'SINGLE' ; AND M_CRDT_RATE > 657.5; AND M_MONTHS_CONTACT > 12.5;	MEDIUM	0		
16	M_MARITAL_ST in 'DIVORCED', 'SINGLE' ; AND M_CRDT_RATE > 657.5; AND M_MONTHS_CONTACT > 12.5;	LOW	0		
17	M_MARITAL_ST in 'DIVORCED', 'SINGLE' ; AND M_CRDT_RATE > 657.5; AND M_MONTHS_CONTACT > 12.5;	MEDIUM	0		
18	M_MARITAL_ST in 'DIVORCED', 'SINGLE' ; AND M_MONTHS_CONTACT > 12.5; AND M_INCOME_LVL in 'LEVEL 1', 'LEVEL 2', 'LEVEL 3', 'LEVEL 4';	HIGH	18	48,866	
19	M_MARITAL_ST in 'MARRIED', 'WIDOW' ; AND M_INCOME_LVL in 'LEVEL 1', 'LEVEL 2', 'LEVEL 3', 'LEVEL 4';	MEDIUM	0		
20	M_MARITAL_ST in 'MARRIED', 'WIDOW' ; AND M_INCOME_LVL in 'LEVEL 1', 'LEVEL 2';	HIGH	0		

# Oracle Data Mining & Spatial

ORACLE Business Intelligence

8.4 Oracle Datamining

LTV Prediction LTV Details Classification Tree LTV Probabilities What If Scoring **Geo LTV Prediction** Regression Regression Model Results Map

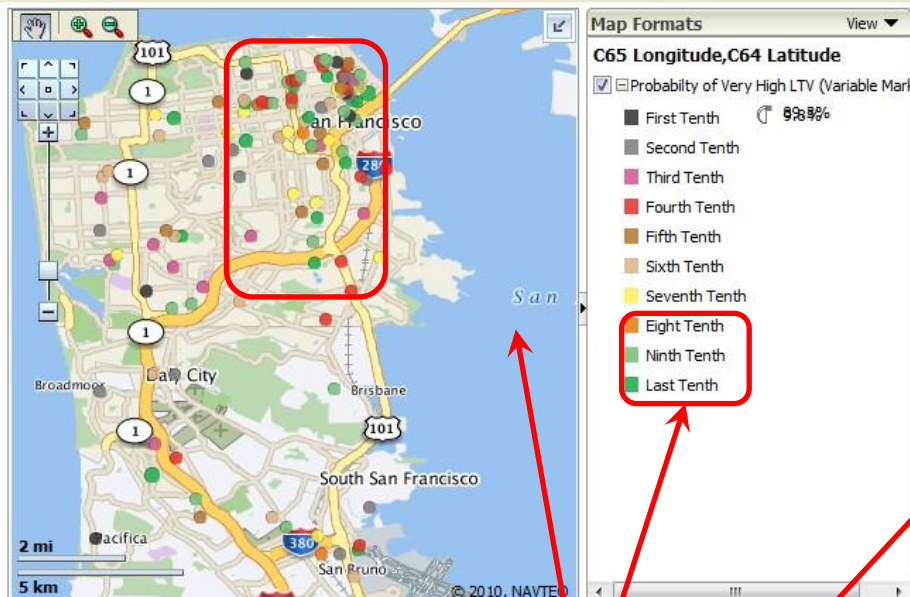
Geo LTV Prediction

[Return to Main Index page](#)

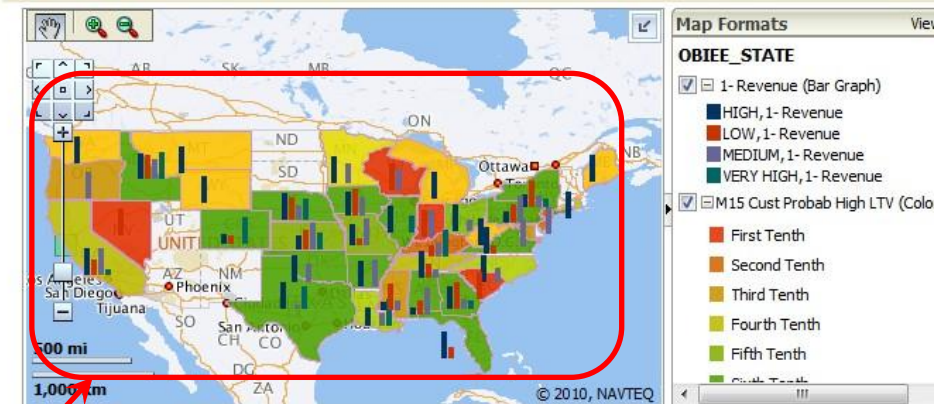
Page Information (click to collapse or expand)

**Description** : This dashboard combines the results of an Oracle Data Mining Classification model with a geographical visual representation of the probabilities in OBIEE maps.

**Location LTV Prediction**  
Time run: 12/9/2011 12:55:35 PM



**Geo LTV Prediction**  
Time run: 12/9/2011 12:55:35 PM



1- Revenue

	HIGH	LOW	MEDIUM	VERY HIGH	Grand Total
Alabama	32,041	62,229	124,009		218,279
Alaska	35,974			127,974	163,948
Arkansas	52,224		56,405		108,629
California	3,215,891	1,820,802	3,142,887	509,122	8,688,700

Customer "most likely" be  
be HIGH and VERY HIGH  
value customer in the future







# Let's get started with the hands-on

- Use wireless connection “HoteLumiere Conference”
- Password: Cardinals
- Use Remote Desktop Connection Utility (Windows) to connect to the IP address written on your handout.
- .compute-1.amazonaws.com
- Username: administrator
- Password: csoaug



# Oracle Test Drive

- Free to try out Oracle BI
- Go to [www.vlamis.com/testdrive-registration/](http://www.vlamis.com/testdrive-registration/)
- Runs off of Amazon AWS
- Hands-on Labs based on Collaborate 2012 HOLs
- Test Drives for:
  - Oracle BI
  - BI Publisher
  - Microsoft Excel against Oracle OLAP
  - Oracle Data Mining
  - Map Views in OBIEE
- Once sign up, you have private instance for 5 hours
- Available now