



Starting Smart with Oracle Advanced Analytics

Heartland OUG Fall 2014

Dan Vlamis

Tim Vlamis

Vlamis Software Solutions

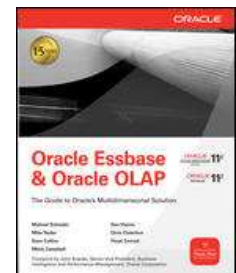
816-781-2880

<http://www.vlamis.com>



VlamiS Software Solutions

- VlamiS Software founded in 1992 in Kansas City, Missouri
- Developed more than 200 Oracle BI systems
- Specializes in ORACLE-based:
 - Data Warehousing
 - Business Intelligence
 - Data Mining and Predictive Analytics
 - Data Visualization
- Expert presenter at major Oracle conferences
- Authors of 2014 book “Data Visualization for Oracle BI 11g”
- Co-author of book “Oracle Essbase & Oracle OLAP”
- www.vlamiS.com (blog, papers, newsletters, services)
- Developer for IRI (former owners of Oracle OLAP)
- Beta tester for OBIEE 11g
- Conference chair for BIWA Summit 2014, 2015





Tim and Dan Vlamis

- Tim (business analyst and academic guy)
 - 25+ years in business modeling, valuation, and scenario analysis
 - 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 University)
 - BA Economics, Yale University
-
- Dan (OLAP expert and career IT guy)
 - 25+ Years in business intelligence/executive information systems
 - Led development team at IRI
 - Founded Vlamis Software Solutions 20 years ago in 1993
 - Author, speaker, Oracle ACE Director
 - BA Computer Science, Brown University



Presentation Agenda

- Background on Analytic Options to the Oracle DB
- Oracle Advanced Analytics
 - Oracle Data Mining
 - Oracle R Enterprise
- How to start with OAA – comparison of options
- Demo of OAA with Oracle BI Front end - SampleApp
- Bonus info on Oracle In Memory



Analytical Options to Oracle Database

- Oracle OLAP
 - Defines a multi-dimensional data structure that allows information for highly complex calculations to be done quickly.
 - Fast query performance and incremental update
 - Simplified access to analytic calculations
- Oracle Advanced Analytics (Data Mining & R)
 - Refers to the process of automatically sifting through data to find hidden patterns and make predictions.
 - Series of highly advanced algorithms and procedures.
 - Extends the “R” language to the Oracle Database
- Oracle Spatial & Graph
 - Provides the capability of relating data to geo positional coordinates, objects, and constructs.
 - Allows the construction and analysis of network topologies.



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

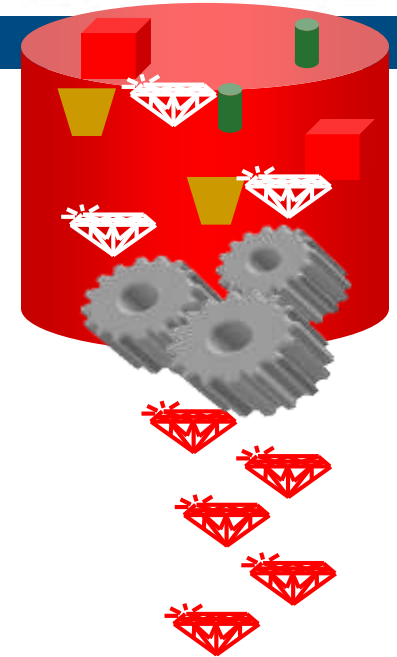


Source: Competing on Analytics, by T. Davenport & J. Harris

What is Data Mining?



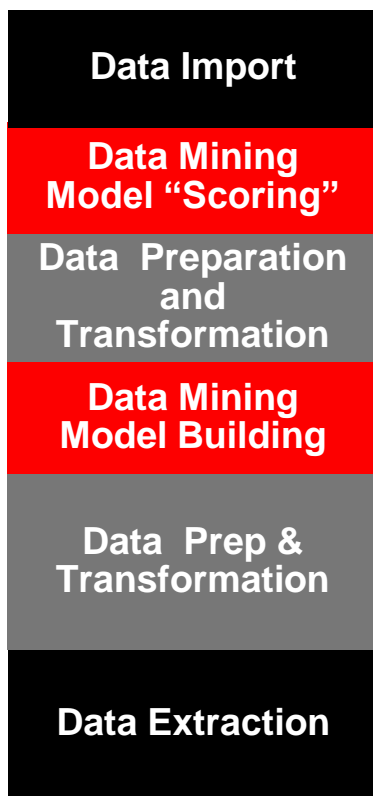
- Automatically sifts through data to find hidden patterns, discover new insights, and make predictions
- Data Mining can provide valuable results:
 - Predict customer behavior (*Classification*)
 - Predict or estimate a value (*Regression*)
 - Segment a population (*Clustering*)
 - Identify factors 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*)



In Database Data Mining



Traditional Analytics



Hours, Days or Weeks



Oracle Data Mining



Secs. Mins or Hours



\$avings

Results

- **Faster time for "Data" to "Insights"**
- **Lower TCO—Eliminates**
- **Data Movement**
- **Data Duplication**
- **Maintains Security**

- Model "Scoring"
- Data remains in the Database
- Embedded data preparation
- Cutting edge machine learning algorithms inside the SQL kernel of Database
- SQL—Most powerful language for data preparation and transformation
- Data remains in the Database



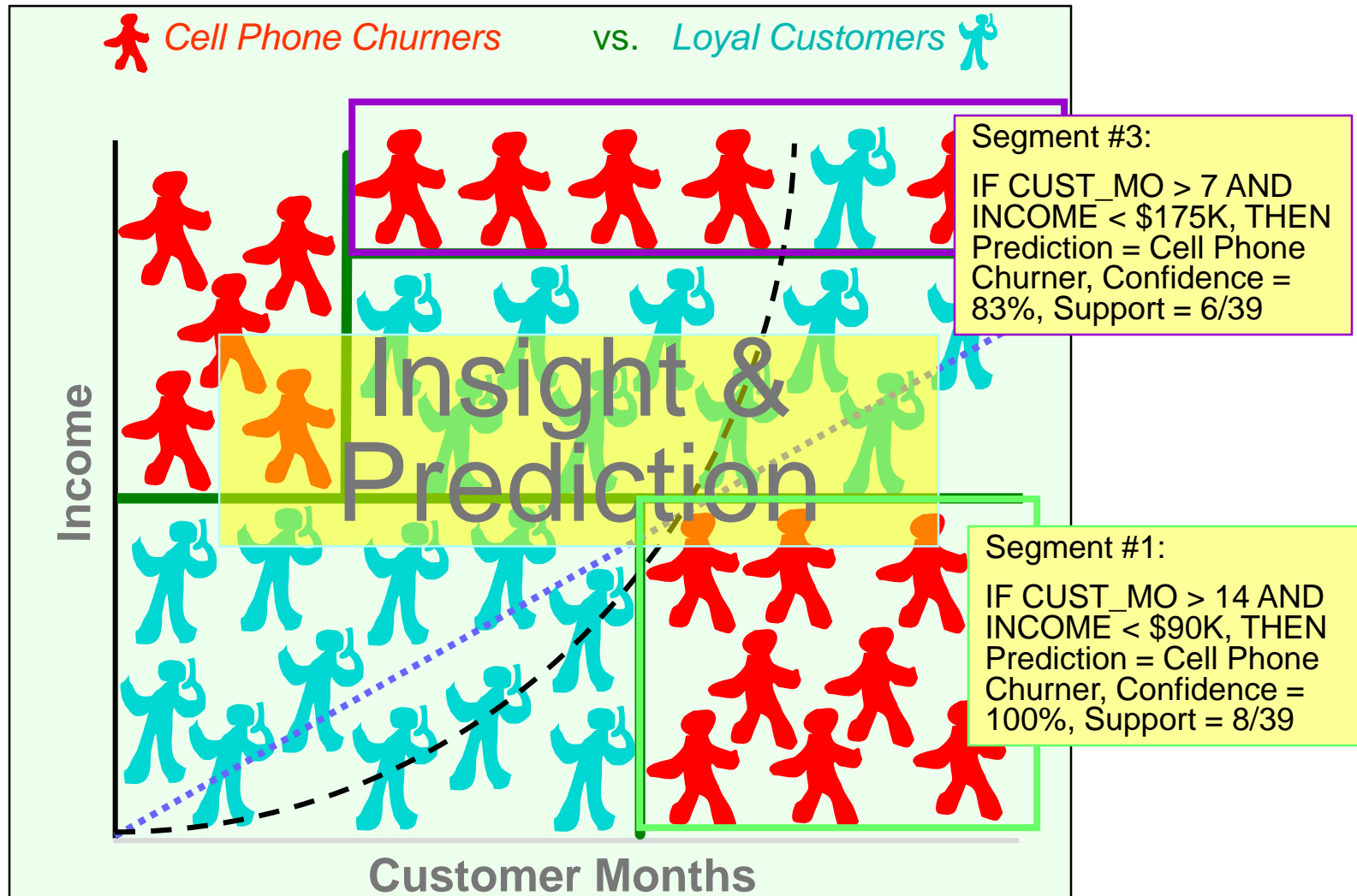
Oracle Data Mining

- 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”



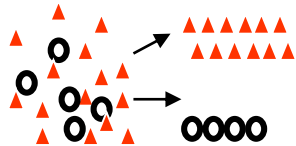
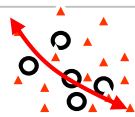
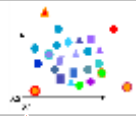
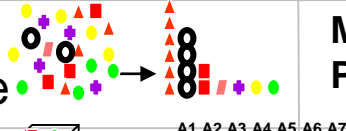
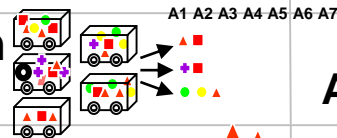
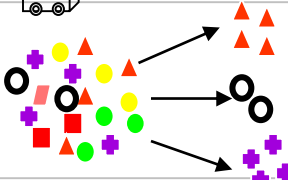
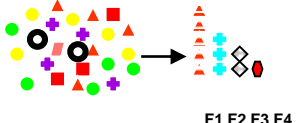
Data Mining Provides

Better Information, Valuable Insights and Predictions



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

Oracle Data Mining Algorithms

Problem	Algorithm	Applicability
Classification 	Logistic Regression (GLM) Decision Trees Naïve Bayes Support Vector Machine	Classical statistical technique Popular / Rules / transparency Embedded app Wide / narrow data / text
Regression 	Linear Regression (GLM) Support Vector Machine	Classical statistical technique Wide / narrow data / text
Anomaly Detection 	One Class SVM	Unknown fraud cases or anomalies
Attribute Importance 	Minimum Description Length Principal Component Analysis	Attribute reduction Identify useful data Reduce data noise
Association Rules 	Apriori	Market basket analysis Next Best Offer
Clustering 	Hierarchical K-Means Hierarchical O-Cluster	Product grouping Text mining Gene and protein analysis
Feature Extraction 	Nonnegative Matrix Factorization (NMF) Singular Value Decomposition (SVD)	Text analysis Feature reduction



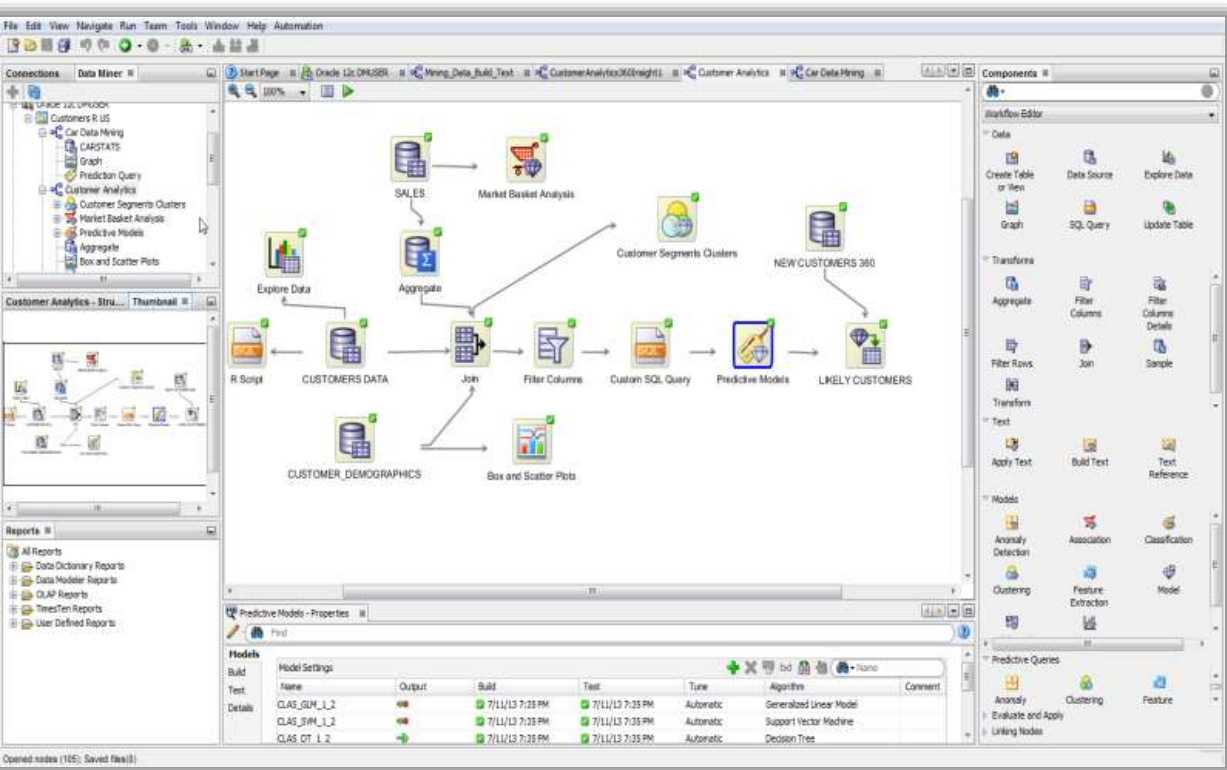
11g Statistics & SQL Analytics (Free)

- **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

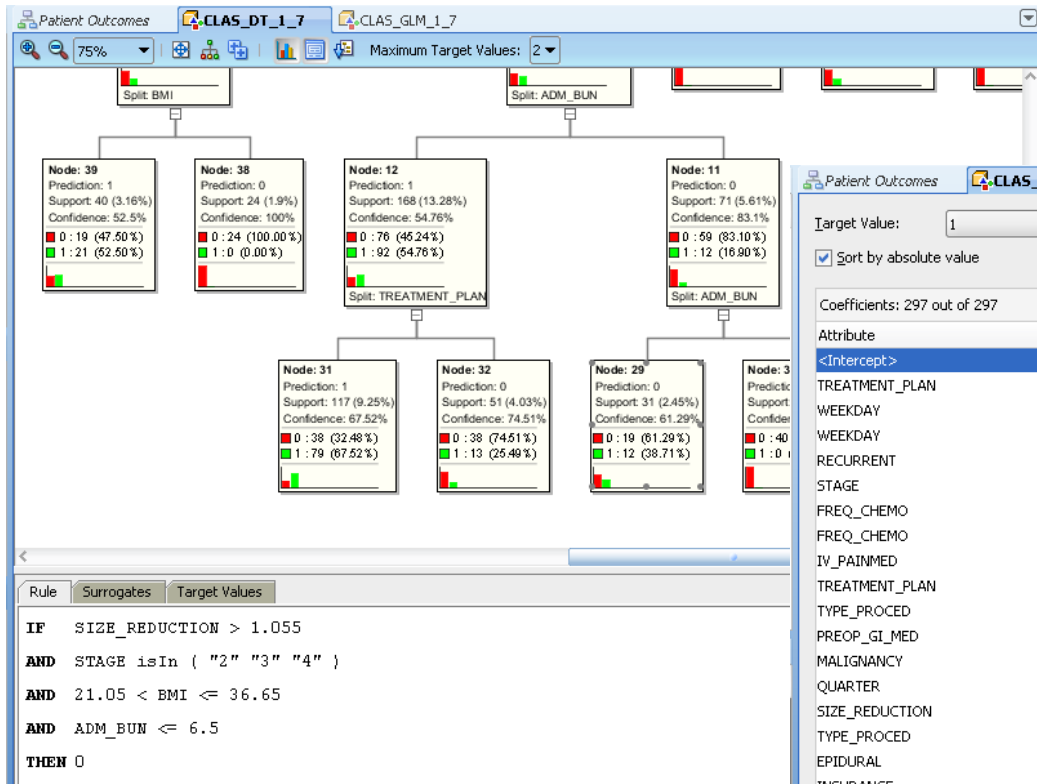
Oracle Data Miner



- Easy to Use
 - Oracle Data Miner GUI for data analysts
 - “Work flow” paradigm
- Powerful
 - Multiple algorithms & data transformations
 - Runs 100% in-DB
 - Build, evaluate and apply models
- Automate and Deploy
 - Save and share analytical workflows
 - Generate SQL scripts for deployment

Understand Model Details

- Interactive model viewers



Outcome predictive models table for the CLAS_GLM_1_7 model. The table displays coefficients for 297 attributes, sorted by absolute value. The columns are 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	-0.40697858	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

LTV Prediction LTV Details **Classification Tree** LTV Probabilities What If Scoring Geo LTV P

Classification Tree

Page Information (click to collapse or expand)

Classification Tree

Time run: 12/9/2011 1:03:03 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		

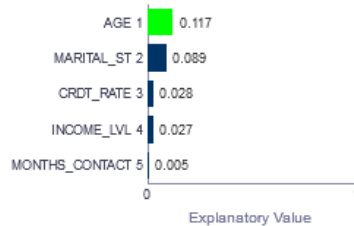
Dynamically Using ODM From Oracle BI



Model Attributes Significance

Time run: 5/15/2014 7:37:48 AM

Most Significant Attributes in the Model



Select Table Details By Credit Rate

C6 Credit Rate	# of Custs	1- Revenue	Proba Low LTV	Probab Very High LTV
600	63	4,578,456	31.0%	11.0%
615	72	4,378,773	27.3%	8.0%
630	76	5,302,193	30.0%	7.6%
645	189	14,404,249	28.1%	9.7%
650	46	3,663,435	19.0%	12.9%

- SAMP_USERS_PROXIES
- SAMP_USERS_VIS
- SYS_IOT_OVER_148455
- Views
 - ODM_REGRDEMO_DATA
 - ODM_REGRDEMO_DATA_LOB
 - ODM_REGRDEMO_TRAIN
 - ODM_SAMP_CUSTOMERS_LTV**
 - ODM_SAMP_CUSTOMERS_TRAI
 - ODM_SAMP_CUSTOMERS_V
 - OLAP_CUST_SEGMENT_HIER_VI
 - OLAP_CUST_SEGMENT_VIEW
 - OLAP_EMPLOYEE_ORG_VIEW

Reports

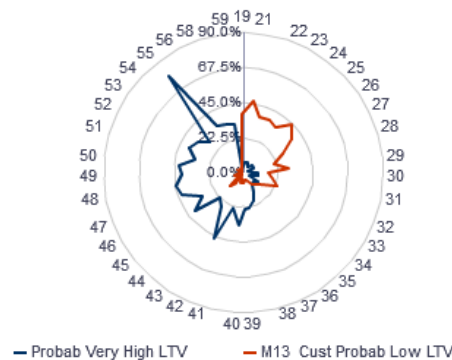
- All Reports
- Data Dictionary Reports



Predicted LTV

Time run: 5/15/2014 7:37:48 AM

LTV Probability by Ages



LTV Probability by Income Level

1: Below 30k

```
CREATE OR REPLACE FORCE VIEW "BISAMPLE"."ODM_SAMP_CUSTOMERS_LTV" ("CUST_KEY", "M_INCOME_LVL", "M_MARITAL_ST", "M_CRDT_RATE", "M_LTV", "M_AGE", "M_MONTHS_CONTACT",
to_number(extractValue(PREDICTION_DETAILS(ODM_LTV_BIN USING *), 'Details/@node')) M_PRED_LTV_NODE,
PREDICTION_PROBABILITY(ODM_LTV_BIN, 'VERY HIGH' USING *) M_PRED_PROB_VH,
PREDICTION_PROBABILITY(ODM_LTV_BIN, 'HIGH' USING *) M_PRED_PROB_H,
PREDICTION_PROBABILITY(ODM_LTV_BIN, 'MEDIUM' USING *) M_PRED_PROB_M,
PREDICTION_PROBABILITY(ODM_LTV_BIN, 'LOW' USING *) M_PRED_PROB_L,
PREDICTION(ODM_LTV_BIN USING *) M_PRED_BIN
FROM ODM_SAMP_CUSTOMERS_V C;
```



What is R?

CRAN Task Views

- R is an Open Source scripting language and environment for statistical computing and graphics
<http://www.R-project.org/>
- Popular alternative to SAS, SPSS & other proprietary statistical environments
- Around 2 million R users worldwide
- Thousands of R packages available



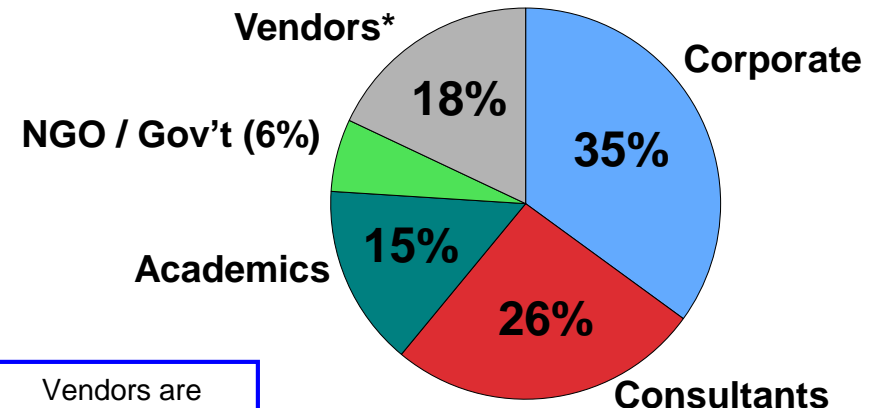
Bayesian	Bayesian Inference
ChemPhys	Chemometrics and Computational Physics
ClinicalTrials	Clinical Trial Design, Monitoring, and Analysis
Cluster	Cluster Analysis & Finite Mixture Models
DifferentialEquations	Differential Equations
Distributions	Probability Distributions
Econometrics	Computational Econometrics
Environmetrics	Analysis of Ecological and Environmental Data
ExperimentalDesign	Design of Experiments (DoE) & Analysis of Experimental Data
Finance	Empirical Finance
Genetics	Statistical Genetics
Graphics	Graphic Displays & Dynamic Graphics & Graphic Devices & Visualization
HighPerformanceComputing	High-Performance and Parallel Computing with R
MachineLearning	Machine Learning & Statistical Learning
MedicalImaging	Medical Image Analysis
MetaAnalysis	Meta-Analysis
Multivariate	Multivariate Statistics
NaturalLanguageProcessing	Natural Language Processing
OfficialStatistics	Official Statistics & Survey Methodology
Optimization	Optimization and Mathematical Programming
Pharmacokinetics	Analysis of Pharmacokinetic Data
Phylogenetics	Phylogenetics, Especially Comparative Methods
Psychometrics	Psychometric Models and Methods
ReproducibleResearch	Reproducible Research
Robust	Robust Statistical Methods
SocialSciences	Statistics for the Social Sciences
Spatial	Analysis of Spatial Data
SpatioTemporal	Handling and Analyzing Spatio-Temporal Data
Survival	Survival Analysis
TimeSeries	Time Series Analysis
gR	gRaphical Models in R



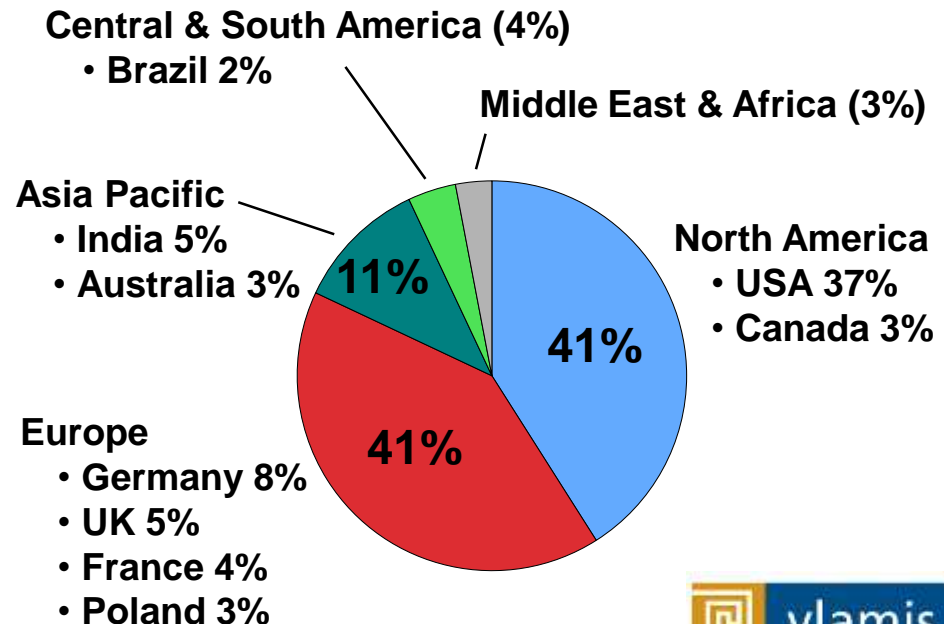
2013 Rexer Analytics Data Miner Survey

- 6th survey since 2007
- 68 questions
- 10,000+ invitations emailed, plus promoted by newsgroups, vendors, and bloggers
- Respondents: 1,259 data miners from 75 countries
- Data collected in first half of 2013

*Data from software vendors is excluded from analyses in this presentation unless otherwise noted.



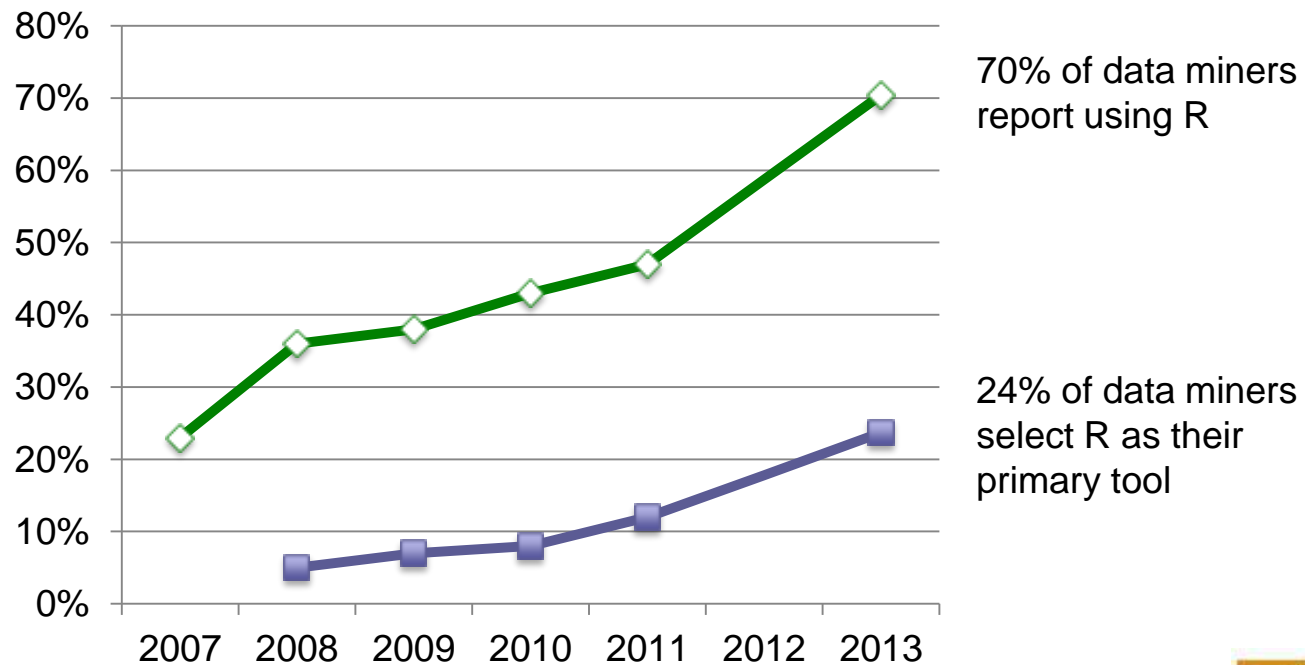
Vendors are included in this analysis.



2013 Rexer Analytics Data Miner Survey (2)

The proportion of data miners using R is rapidly growing, and since 2010, R has been the most-used data mining tool. While R is frequently used along with other tools, an increasing number of data miners also select R as their primary tool. Among data miners who say they are likely to switch their primary package in the coming year, R is frequently identified as the tool they are plan to switch to – more than 2.5 times more often that any other tool.

R Usage





Oracle R Enterprise

- Part of the Advanced Analytics Option to the Oracle Database Enterprise Edition
- Provides transparent access to database-resident data from R
- Embedded R script execution through database managed R engines with SQL language integration
- Provides data and task parallelism and full power of Oracle database for R
- Enables advanced statistics for in-database execution
- Integrates R into the IT software stack
- Extends and enhances open source R



Oracle R Distribution



ability to dynamically load:

Intel Math Kernel Library (MKL)

AMD Core Math Library

Solaris Sun Performance Library



Oracle
Support

- Improved scalability at client and database for embedded R execution
- Enhanced linear algebra performance using Intel's MKL, AMD's ACML, and Solaris Sun Performance Library
- Enterprise support for customers of Oracle Advanced Analytics option, Big Data Appliance, and Oracle Linux
- Available as a free download from Oracle
- Oracle to contribute bug fixes and enhancements to open source R



Other R Offerings

- ROracle
 - Open source Oracle database interface driver for R based on OCI
 - Maintained by Oracle, optimizations and bug fixes released to open source community
- Oracle R Connector for Hadoop
 - R interface to Oracle Hadoop Cluster on Big Data Appliance
 - Access and manipulate data in HDFS, database, and file system
 - Write MapReduce functions using R and execute through R
- Rstudio
 - Popular open source user interface for R
 - Integrated Development Environment
- Rcommander
 - Extended GUI for R
 - Used for plugins (e.g. text mining)
 - Often used with Rstudio

R now integrated into OBIEE 11.1.1.7

8.13 Oracle R Enterprise

R Integration R End-User Interaction

R Workbench

R Sourcing from BI CEIM

R Sourcing from BI CEIM (end-user)

Custom Link

Home

Catalog

Favorites

Dashboards

New

Open

Signed In As Paulo Ro

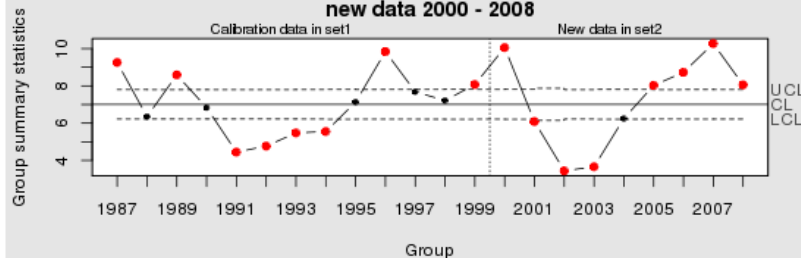
QCC-1

Edit Script

```
function(dat) {
  uniquecarrier <- dat$UNIQUE_CARRIERS
  delayType <- dat$DELAY_TYPE
  startYear <- dat$START_YEAR
  endYear <- dat$END_YEAR
  newDataStartYear <- dat$NEW_DATA_START_YEAR
  newDataEndYear <- dat$NEW_DATA_END_YEAR
  graphTypes <- dat$GRAPH_TYPES
  ore.sync(table="ONTIME_S")
  ore.attach()
  library("qcc", lib.loc="/home/oracle/R/library")
  if(graphTypes == "both") graphTypes <- c("xbar","S")
  dat <- ONTIME_S
  if(newDataStartYear != -1) {
    dat <- dat[(dat$YEAR>=startYear & dat$YEAR<=endYear) |
      (dat$YEAR>=newDataStartYear & dat$YEAR<=newDataEndYear),]
  } else {
    dat <- dat[dat$YEAR>=startYear & dat$YEAR<=endYear,]
  }
  if(uniquecarrier != "ALL") {
    dat <- dat[dat$UNIQUECARRIER == uniquecarrier, ]
  }
  if (delayType=="arrival") {
    dat <- dat[,c("ARRDELAY", "YEAR")]
  } else {
    dat <- dat[,c("DEPDELAY", "YEAR")]
  }
  names(dat) <- c("DELAY", "YEAR")
  delayData <- ore.pull(dat)
  library(qcc)
  group <- qcc.groups(delayData$DELAY, delayData$YEAR)
  numYears <- endYear - startYear + 1
  par(mfrow=c(length(graphTypes),1))
  for(gt in graphTypes) {
    if (newDataStartYear == -1) { # not using new data comparison
      title <- paste(gt, "Chart for", delayType, "delay -",
        years, startYear, "-", endYear)
      set1 <- group[1: numYears,]
      obj <- qcc(set1, type=gt, title=title)
    } else {
      title <- paste(gt, "Chart for", delayType, "delay -",
        years, startYear, "-", endYear, "New data",
        newDataStartYear, "-", newDataEndYear)
      set1 <- group[1: numYears,]
      set2 <- group[numYears+1: numYears+numYears,]
      obj <- qcc(set1, type=gt, title=title)
    }
  }
}
```

QCC-1 Graphics Result

xbar Chart for arrival delay - years 1987 - 1999 new data 2000 - 2008

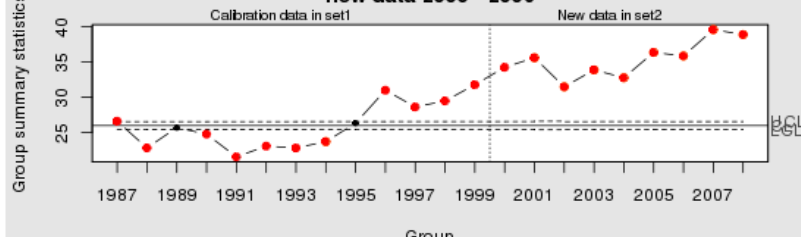


Number of groups = 22
Center = 7.008499
StdDev = 26.1959

LCL is variable
UCL is variable

Number beyond limits = 16
Number violating runs = 0

S Chart for arrival delay - years 1987 - 1999 new data 2000 - 2008



R now integrated into OBIEE 11.1.1.7

R Workbench

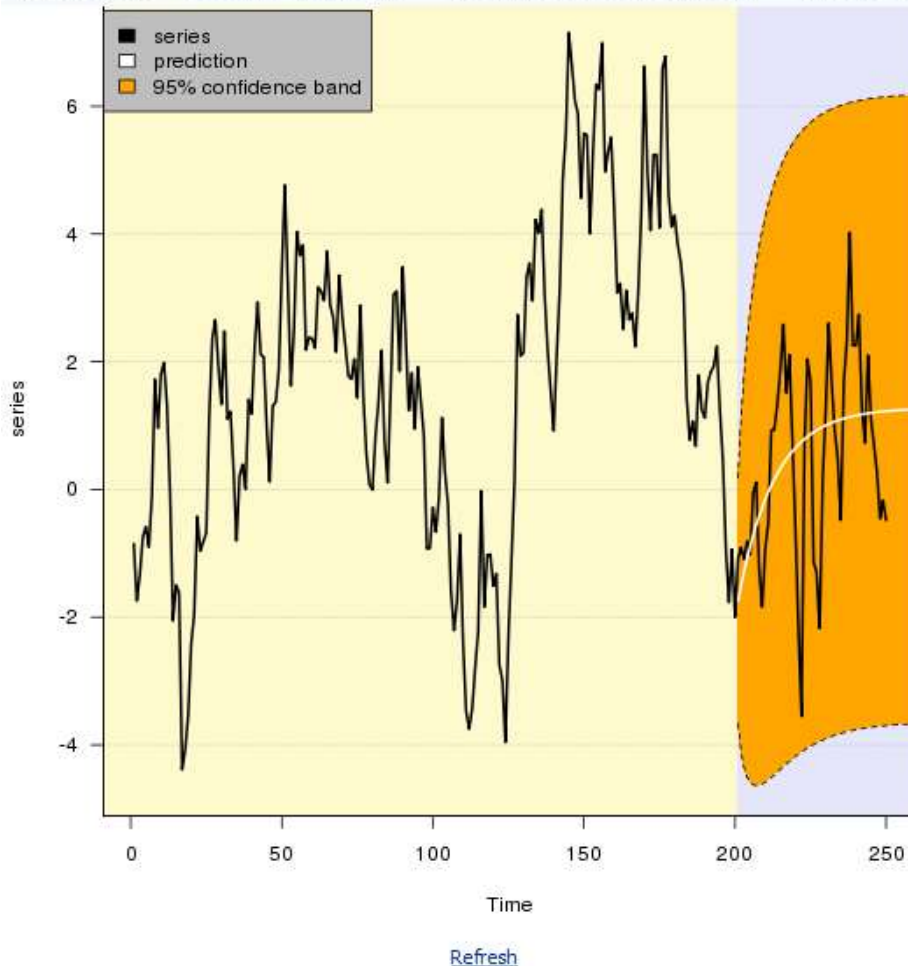
R Sourcing from BI CEIM

R Sourcing from BI CEIM (end-user)

R Results Object in RPD

BIP Sourcing from R

Quality Control Chart (BIP)



```
require(gplots)

set.seed(120)
# simulate an AR(1) process
coefs <- 0.95
series <- arima.sim(list(ar=coefs),n=250)

# fit AR(1) with the 200 first data
model <- arima(series[1:200],c(1, # AR part
                                0, # I order
                                0)) # MA part

# make forecast from the model
forecast <- predict(model,80)

# compute the limits of the graph
ylim <- c( min(series[1:200],forecast$pred - 1.96 * forecast$se),
           max(series[1:200],forecast$pred + 1.96 * forecast$se))

# prepare the space where to plot
opar <- par(mar=c(4,4,2,2),las=1)
plot(series,ylim=ylim,type="n",xlim=c(1,250))
usr <- par("usr")

# split the figure in two parts
# - the part used to fit the model
rect(usr[1],usr[3],201,usr[4],border=NA,col="lemonchiffon")

# - the part used to make the forecast
rect(201,usr[3],usr[2],usr[4],border=NA,col="lavender")

abline(h= (-3:3)*2, col="gray", lty=3)

# draw a 95% confidence band
polygon( c(201:280,280:201),
         c(forecast$pred - 1.96*forecast$se,rev(forecast$pred + 1.96*f
         col="orange",
         lty=2,border=NA))

lines( 201:280, forecast$pred - 1.96*forecast$se, lty=2)
lines( 201:280, forecast$pred + 1.96*forecast$se, lty=2)
```

Oracle Advanced Analytics & 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

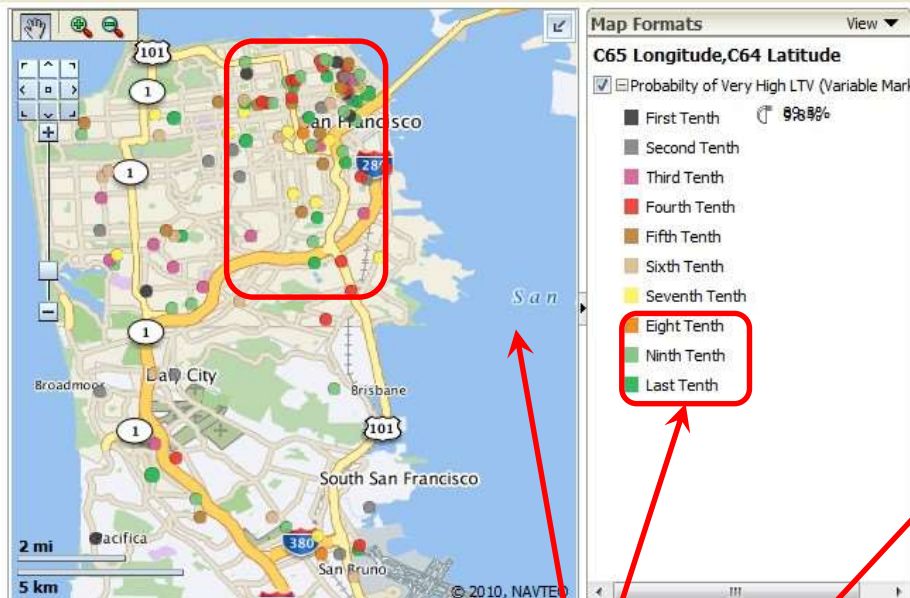
Geo LTV Prediction

[Return to Main Index page](#)

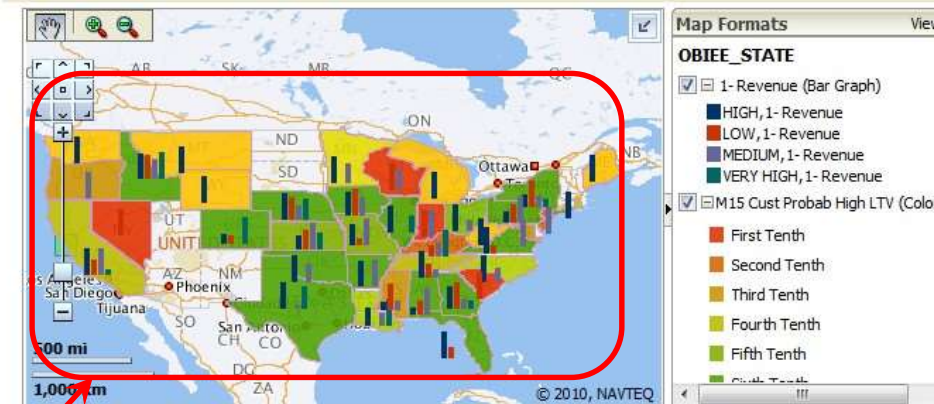
Page Information (click to collapse or expand)

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



What is Spatial Data?

- Business data that contains or describes location
 - Street and postal address (customers, stores, factory, etc.)
 - Sales data (sales territory, customer registration, etc.)
 - Assets (cell towers, pipe lines, electrical transformers, etc.)
 - Geographic features (roads, rivers, parks, etc.)
- Anything connected to a physical location
- Any data sets that contain “link and node” relationships between data objects. Can be directional or non-directional.



How to Get Started

- Lots of internal experts and lots of people who would like to be involved and learn
- Lots of people intimidated by what they don't know
- Start by “level setting” and establishing a strong foundation with basic training (2 days)
 - Bring people along on the journey, establish culture
- Immediately conduct a workshop (JAD style session) investigation of possibilities
 - Evaluation of data sources and data sets
 - Recognition of major business issues
 - Review of basic algorithms
 - Identification of potential PoC projects (plusses and minuses)
- Decide on pilot projects and who works on it
- Start simple and return value quickly



Need Representatives from Across Organization

- Get different perspectives
- Data Mining requires synergistic insights
- Builds broad support
- Non-siloed
- Takes Data mining out of IT and into Line of Business
- Participants learn from each other
- Formal statistics background not required
- Only prerequisite: analytical mind set / love data



Methodology

- Implement a highly scalable infrastructure
- Establish a common foundational understanding of data mining
- Demonstrate the Value of Analytics by Completing a Market Basket Project Immediately



ODM Machine Bundle Overview

- Hardware
 - Oracle Database Appliance
- Software
 - Oracle Database 12c (with options)
 - Oracle Advanced Analytics Option including Oracle Data Mining
 - Oracle SQL Developer: Data Miner Add-in (free download)
- Services
 - Implementation and configuration from VlamiS Software Solutions (Oracle Gold Partner)
 - Oracle University Oracle Data Mining Techniques course (taught by VlamiS Software Solutions)
 - Market Basket Analysis Project performed on company data
- Time frame: 9 business days (less than 2 weeks)



Compressed Schedule

- Day 1:
 - Two consultants meet with client team to review project plan, review data sources, identification of best data to start with, set technical objectives for project (basic market basket analysis deliverable)
- Day 2:
 - Consultant One: Install ODA and configure to network (need support from client tech staff)
 - Consultant Two: Conduct first day of ODM class with client team
- Day 3:
 - Consultant One: Install new pluggable Database, SQL Developer
 - Consultant Two: Conduct second day of ODM class with client team
- Day 4:
 - Two consultants establish data plan for project with client and import data
- Day 5:
 - Consultant One: Prepare tables for mining (add keys, new tables, transforms, etc.)
 - Consultant Two: Document data plan
- Day 6:
 - Consultant Two: Build market basket workflow
- Day 7:
 - Consultant Two: Conduct market basket analyses
- Day 8:
 - Consultant Two: Prepare presentation of findings from market basket analyses
- Day 9:
 - Consultant Two: Deliver presentation with client



Oracle Data Mining Training (2 days)

- Introduction
- Data Mining Concepts and Terminology
- The Data Mining Process
- Introducing Oracle Data Miner 11g Release 2
- Using Classification Models
- Using Regression Models
- Using Clustering Models
- Performing Market Basket Analysis
- Performing Anomaly Detection
- Deploying Data Mining Results



Oracle R Enterprise Training (2 Days)

- Oracle R Enterprise technologies introduction
- Introduction to R hands-on
- ORE transparency layer with hands-on exercises
- ORE embedded R execution with hands-on exercises
- ORE predictive analytics with hands-on exercises
- Using ROracle
- Overview of ORE with OBIEE



Comparison of Training Courses

Oracle Data Mining

- Organized by algorithm
- Intro to data mining
- MBAs, BI Admin, DBAs
- Focused on business issues
- Uses GUI
- Approachable for new users

Oracle R Enterprise

- Organized by process
- Intro to Oracle R Enterprise
- Data Scientists, BI Admin, DBAs
- Focused on executing R in Oracle Database
- Uses R scripts
- Technical



Oracle In Memory

- Announced at Oracle OpenWorld 2013
- Launched by Larry Ellison June 2014
- Not released yet, so packaging not final – Option? (likely) Exadata only? (likely not)
- Allows Oracle tables to be loaded into memory
- Super-fast scans of data using SIMD and vector instruction sets
- Works with OLTP
- First release likely limited in scope
- Allows for more flexible aggregation



Oracle In Memory Predictions

- In Memory capabilities will start overlapping with Oracle OLAP capabilities
- Allows for more flexible deployments
- No need to pre-define hierarchies
- Does not have metadata for pre-defined hierarchies yet
- Does not have capabilities for OLAP calculations yet
- Does not involve misunderstood cube technology – simple for DBAs to understand



Oracle Test Drive

- Free to try Oracle BI, Advanced Analytics and Big Data
- Go to www.vlamiS.com/td
- Runs off of Amazon AWS
- Test Drives for:
 - Oracle BI
 - Oracle Advanced Analytics
 - Big Data
- Once sign up, you have private instance for 5 hours
- Available now



Mark Your Calendars Now!

BIWA Summit 2015, Jan 27-29 Oracle HQ Conference Center

Accepting Abstracts NOW!

Business Intelligence, Warehousing and Analytics
IOUG Special Interest Group

www.biwasummit.com



Thank You!

Thank You for Attending Session **Capitalizing on Analytics in the Oracle Database in BI Applications**

Presenter Information

Dan Vlami, President

Vlami Software Solutions, Inc.

816-781-2880

dvlami@vlami.com

For more information go to www.vlami.com