

# ODTUG Kscope17



SAN ANTONIO, TEXAS ★ JUNE 25-29



PLEASE FILL OUT YOUR EVALUATIONS



# Starting with Oracle Data Science in the Cloud

Kscope 17

Tim VlamiS

Tuesday, June 27, 2017

[@VlamiSoftware](#)



# VlamiS Software Solutions

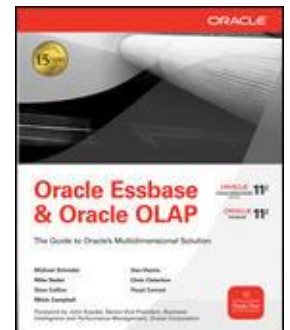
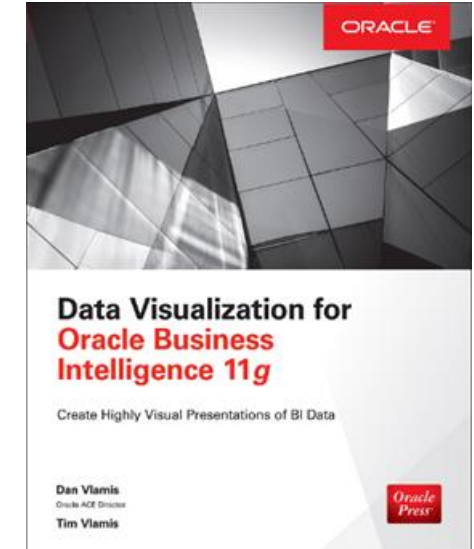
- VlamiS Software founded in 1992 in Kansas City, Missouri
- Developed 200+ Oracle BI and analytics systems
- Specializes in Oracle-based:
  - Enterprise Business Intelligence & Analytics
  - Analytic Warehousing
  - Data Mining and Predictive Analytics
  - Data Visualization
- Multiple Oracle ACEs, consultants average 15+ years
- [www.vlamiS.com](http://www.vlamiS.com) (blog, papers, newsletters, services)
- Co-authors of book “Data Visualization for OBI 11g”
- Co-author of book “Oracle Essbase & Oracle OLAP”
- Oracle University Partner
- Oracle Gold Partner

 EDUCATION RESELLER

 APPROVED  
EDUCATION CENTER

 Gold  
Partner


Specialized  
Oracle Business Intelligence  
Foundation Suite 11g





# Tim Vlamis

## Vice President & Analytics Strategist

- 30+ years in business modeling and valuation, forecasting, and scenario analyses
- Oracle ACE  ORACLE ACE
- Instructor for Oracle University's Predictive Analytics, Data Mining Techniques and Oracle R Enterprise Essentials Courses
- Professional Certified Marketer (PCM) from AMA
- Adjunct Professor of Business Benedictine College
- MBA Kellogg School of Management (Northwestern University)
- BA Economics Yale University



# Presentation Agenda

- Thinking about Advanced Analytics
- Background on Analytic Options to the Oracle DB
  - Data base cloud service high performance
  - Data base cloud service extreme performance
- Oracle Big Data Cloud Service
- Oracle Advanced Analytics
  - Oracle Data Mining
  - Oracle R Enterprise
- How to start with OAA – comparison of options



# Evidence-based analysis requires data

*“Data! Data! Data!” he cried impatiently. “I can’t make bricks without clay.”*

Sherlock Holmes in “The Adventure of the Copper Beeches”  
by Sir Arthur Conan Doyle





# But we have plenty of data...

- Est. Global IP traffic/month is **89 Exabytes** (89 Billion Gigabytes)
- Est. Global data stores total **4.4 Zetabytes** (44 Trillion Gigabytes)





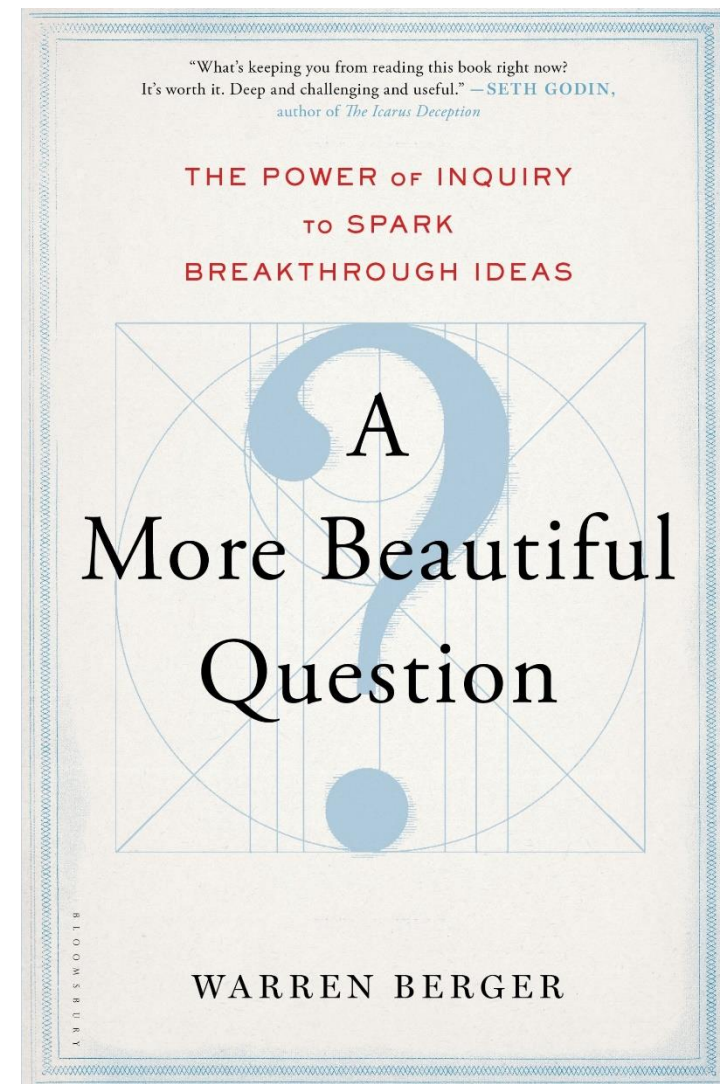


# Good Questions/Hypotheses are Needed

What behaviors in the past year are most significant in terms of segmenting our customers?

What is the Life Time Value of our customers?  
What's a potential new customer worth?

Which products are purchased together most often? Which products are purchased with our most profitable products?

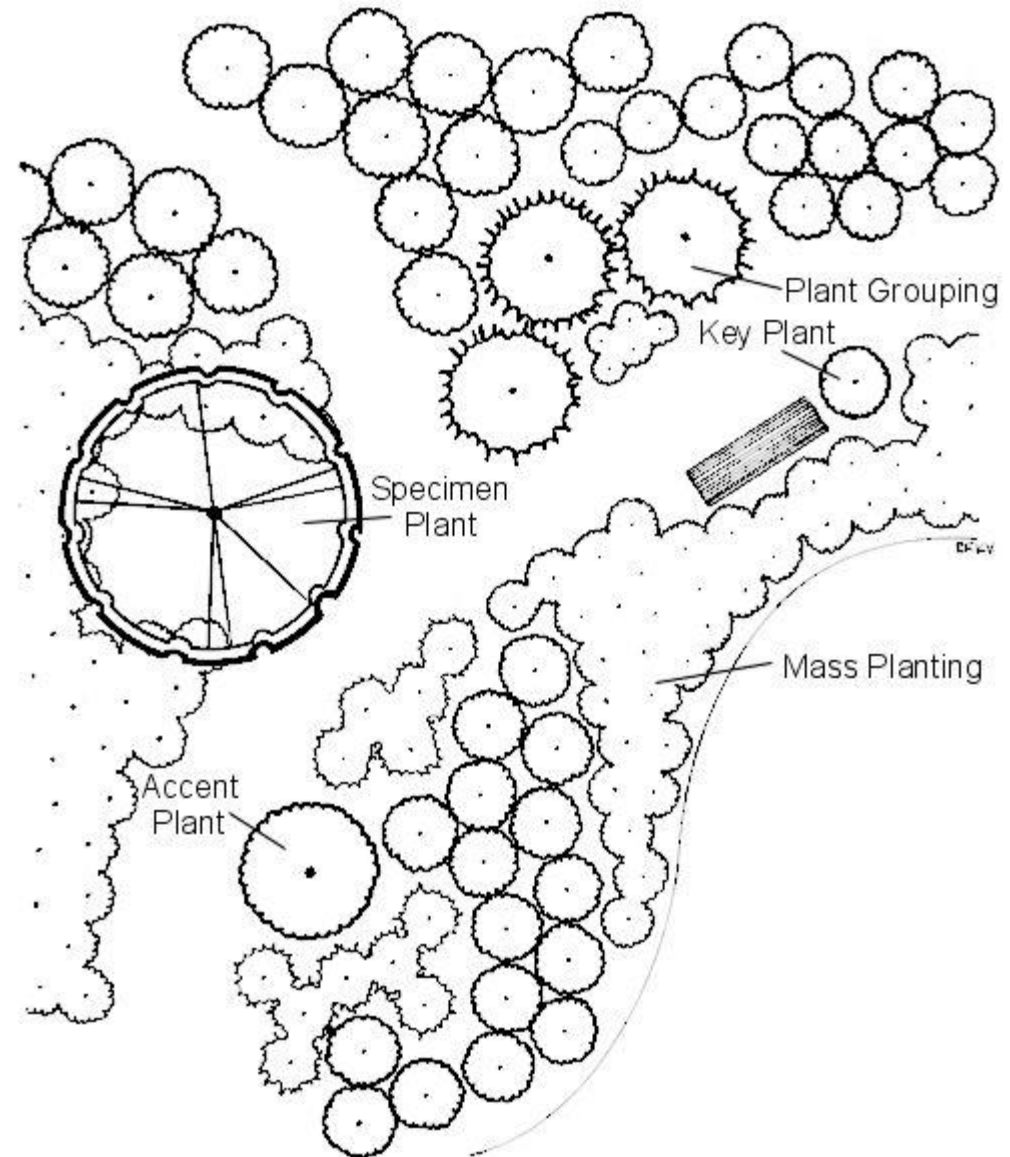






# Grow Your Analytics Naturally

- Start with a general plan





# Grow Your Analytics Naturally

- Start with a general plan
- Enhance existing features





# Grow Your Analytics Naturally

- Start with a general plan
- Enhance existing features
- Prepare one bed at a time







# Grow Your Analytics Naturally

- Start with a general plan
- Enhance existing features
- Prepare one bed at a time
- **Mulch and weed early on**





# Time is a Great Designer



Pictures Courtesy of National Trust Sissinghurst





# Time is a Great Designer



Pictures Courtesy of National Trust Sissinghurst





# Time is a Great Designer



Pictures Courtesy of National Trust Sissinghurst





# Time is a Great Designer



Pictures Courtesy of National Trust Sissinghurst





# Time is a Great Designer



Pictures Courtesy of National Trust Sissinghurst



# Time is a Great Designer



Pictures Courtesy of National Trust Sissinghurst



# Four Realms of Analytics

Probability Based

**Diagnostic  
Analytics**

**Predictive  
Analytics**

Rules Based

**Descriptive  
Analytics**

**Prescriptive  
Analytics**

Past

Future



# Oracle Database Cloud Service

- Has Oracle Advanced Analytics (Oracle Data Mining and ORE)
  - DBCS High Performance and Extreme Editions
  - Exadata Cloud Service
  - Exadata Bare Metal Cloud Service
  - Exadata Cloud Machine
- Does not have Oracle Advanced Analytics (ODM and ORE)
  - DBCS Standard and Enterprise Editions
  - Exadata Express Cloud Service
  - Database Schema Service
- If Oracle fully manages, you likely don't have OAA.





# Oracle Big Data Cloud Service

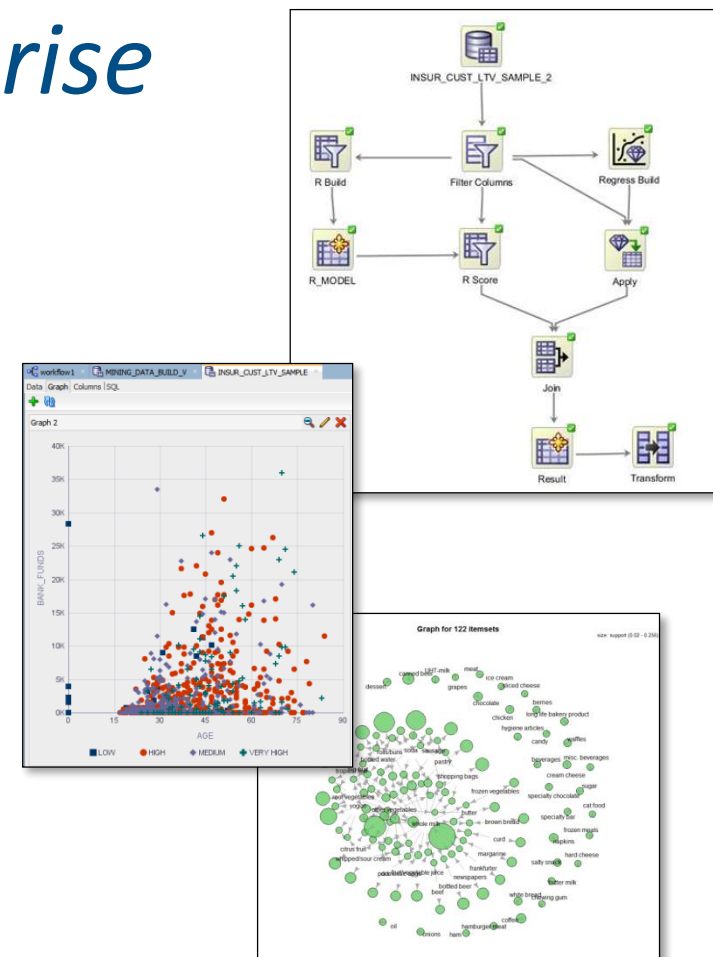
- Oracle Big Data Connectors include ORAAH (Oracle R Advanced Analytics for Hadoop)
- Oracle Big Data Spatial and Graph includes Oracle Property Graph algorithms
- Oracle Big Data Discovery Cloud Service
  - Fantastic for those who already have large sets of Hive tables



# Oracle Advanced Analytics (OAA) DB Option

## *Oracle Data Mining + Oracle R Enterprise*

- Powerful **in-database** algorithms for **Data Mining** and **Statistical Analysis**
- **Easy** to add **predictive analytics** to enterprise applications and BI
- **Fastest** way to deliver **scalable, enterprise-wide** predictive analytics
- ORE eliminates R's limitations (memory and speed) for **Enterprise-scale analytics**

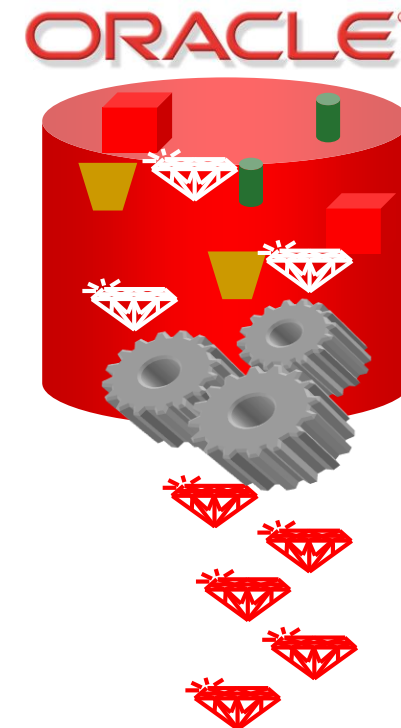






# What is Data Mining?

- 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 co-occurrences and “market baskets” within an event set (*Associations*)
  - Find fraudulent or “rare events” (*Anomaly Detection*)





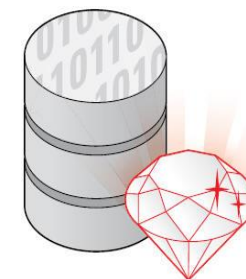
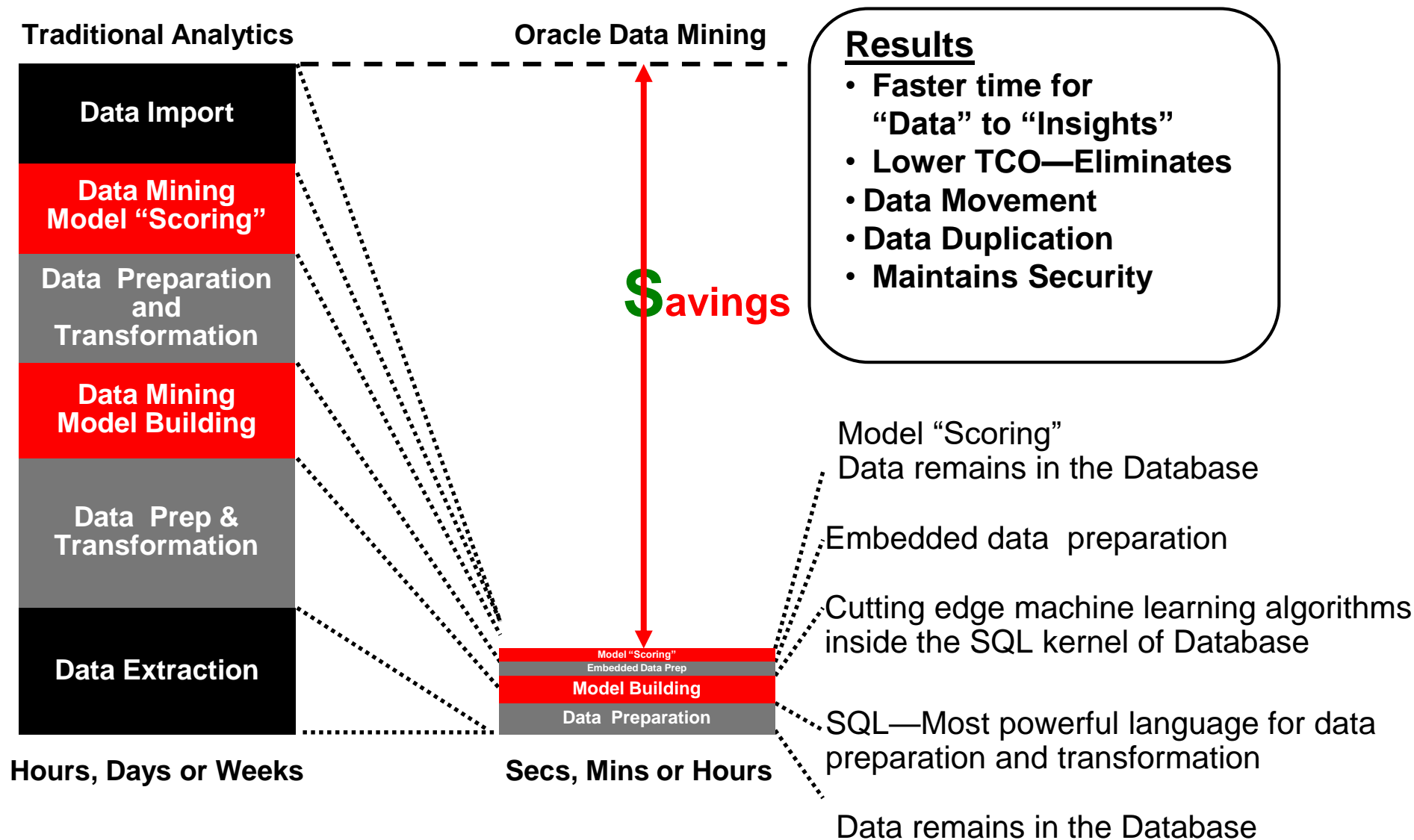
# Oracle Data Mining

- 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”



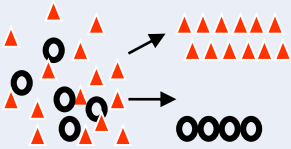

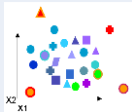
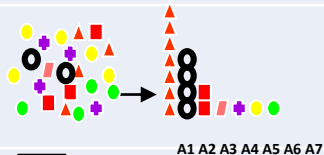
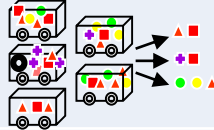
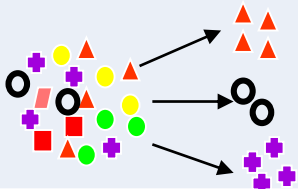



# In-Database Data Mining





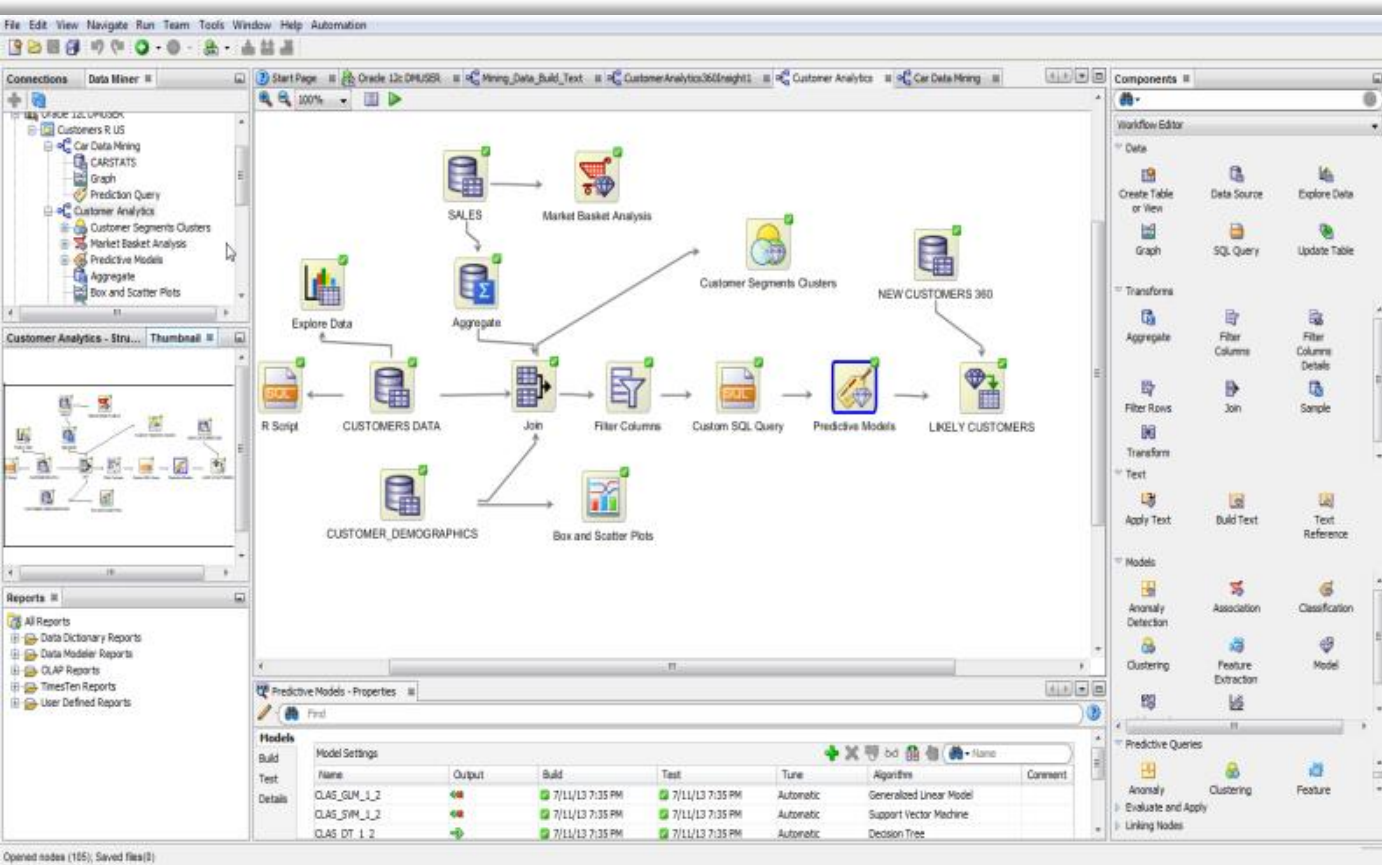
# Oracle Data Mining Algorithms

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 or Text
Regression 	Linear Regression (GLM) Support Vector Machine	Classical Statistical Technique Wide/Narrow Data or Text
Anomaly Detection 	One Class SVM	Unknown fraud cases or anomalies
Attribute Importance 	Minimum Description Length Principal Component Analysis	Attribute reduction Reduce data noise
Association Rules 	Apriori	Market Basket Analysis
Clustering 	Hierarchical K-Means Orthogonal Partitioning Expectation Maximization	Market Segmentation Product / Location Groupings Text analysis
Feature Extraction 	Non-negative Matrix Factorization Singular Value Decomposition	Feature Reduction Text Analysis





# 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

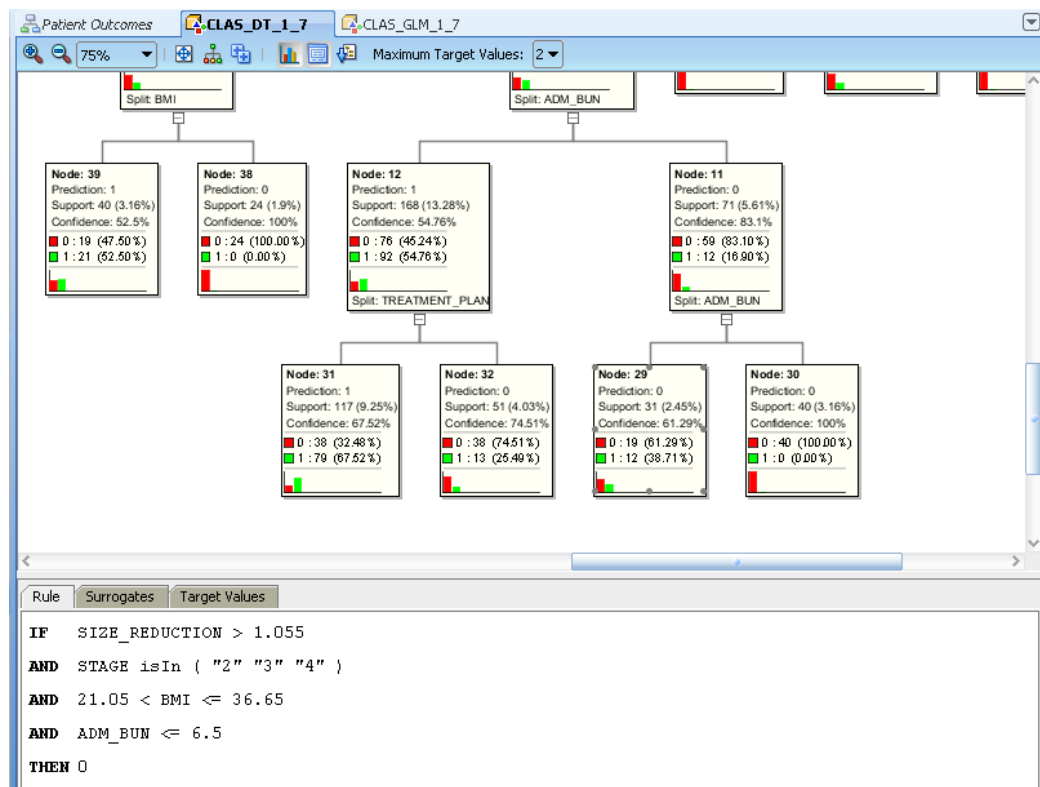


Table showing Coefficients: 297 out of 297. The table lists attributes, values, coefficients, standardized coefficients, and exponential coefficients.

Attribute	Value	Coefficient	Standardized Coeffi...	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

8.4 Oracle Datamining

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

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

Classification Tree Details  
Time run: 12/9/2011 1:03:03 PM

#	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',	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		

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



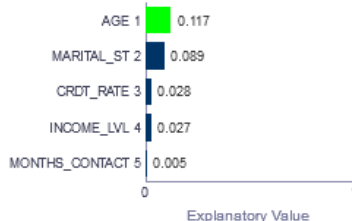
# Dynamically Using ODM From Oracle BI



## Model Attributes Significance

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

### Most Significant Attributes in the Model



Select Table Details By Credit Rate

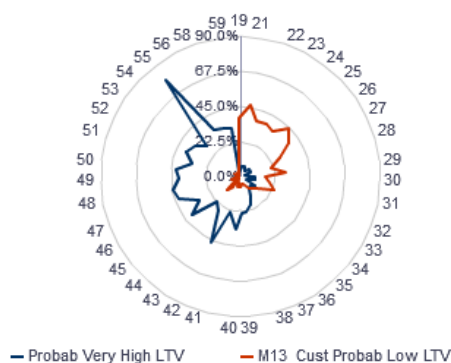
C6 Credit Rate	# of Custs	1- Revenue	Proba Low LTV	Probab Very High LTV
600	63	4,578,456	31.0%	11.0%
615	72	4,378,773	27.3%	8.0%
630	76	5,302,193	30.0%	7.6%
645	189	14,404,249	28.1%	9.7%
650	46	3,663,435	19.0%	12.9%
665	58	3,581,326	11.8%	15.3%
680	76	5,488,196	12.6%	14.3%
695	140	9,315,649	11.1%	22.7%
700	36	2,483,128	14.0%	23.9%
715	27	1,502,589	11.4%	33.1%
730	37	2,547,730	10.1%	28.3%
745	80	5,278,155	7.8%	21.7%
750	10	617,865	7.5%	11.5%
765	14	1,406,422	4.2%	39.6%



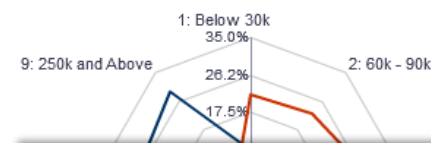
## Predicted LTV

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

### LTV Probability by Ages



### LTV Probability by Income Level



8.

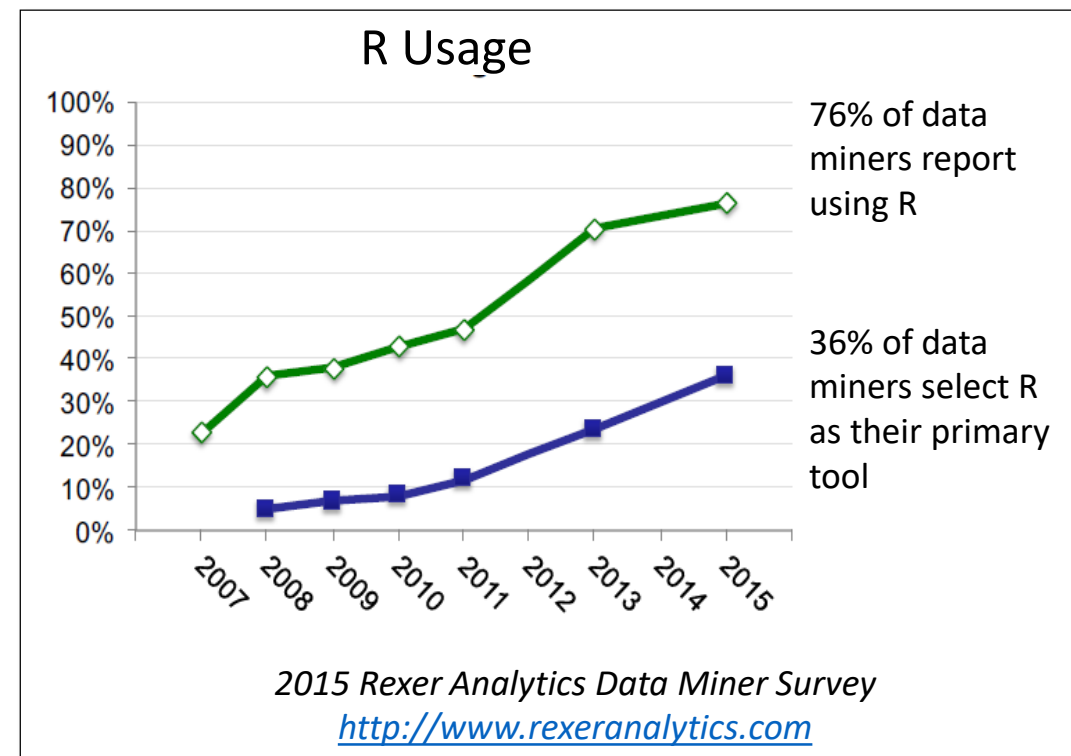
- SAMP\_USERS\_PROXIES
- SAMP\_USERS\_VIS
- SYS\_IOT\_OVER\_148455
- Views
  - ODM\_REGRDEMO\_DATA
  - ODM\_REGRDEMO\_DATA\_LOB
  - ODM\_REGRDEMO\_TRAIN
  - ODM\_SAMP\_CUSTOMERS\_LTV
  - ODM\_SAMP\_CUSTOMERS\_TRAI
  - ODM\_SAMP\_CUSTOMERS\_V
  - OLAP\_CUST\_SEGMENT\_HIER\_VI
  - OLAP\_CUST\_SEGMENT\_VIEW
  - OLAP\_EMPLOYEE\_ORG\_VIEW
- Reports
  - All Reports
  - Data Dictionary Reports

```
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?

- 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
- 2 million+ users worldwide and growing
- Thousands of R packages available
- Taught extensively in higher education







# *R is extensively used by Statisticians, Data Analysts, Students*

- Free (Open source)
- Graphical
- Powerful
- Extensible
- Ease to install and use
- Industry/subject specific packages
- Out-of-the-box functionality with many 'knobs', but smart defaults

The R Console shows the following code:

```
rgl.sr> ylen <- ylim[2] - ylim[1] + 1
rgl.sr> colorlut <- terrain.colors(ylen)
rgl.sr> col <- colorlut[y - ylim[1] + 1]
rgl.sr> rgl.clear()
rgl.sr> rgl.surface(x, z, y, color = col)
```

The R Data Editor shows a table with the following data:

height	weight
58	115
59	117
60	120
61	123
62	126
63	129
64	132
65	135
66	139
67	142
68	146
69	150
70	154
71	159
72	164

The Quartz window shows a 3D surface plot of the data. The R Workspace Browser shows a list of objects in the workspace. The R Package Manager shows a list of installed and available packages.

- Oracle R Distribution

- ROracle

*Open Source Software available to  
R Community for free*

- Oracle R Enterprise (ORE)

- Oracle R Advanced Analytics for Hadoop (ORAAH)

- Oracle R Connector for Hadoop (ORCH)



# Oracle R Enterprise

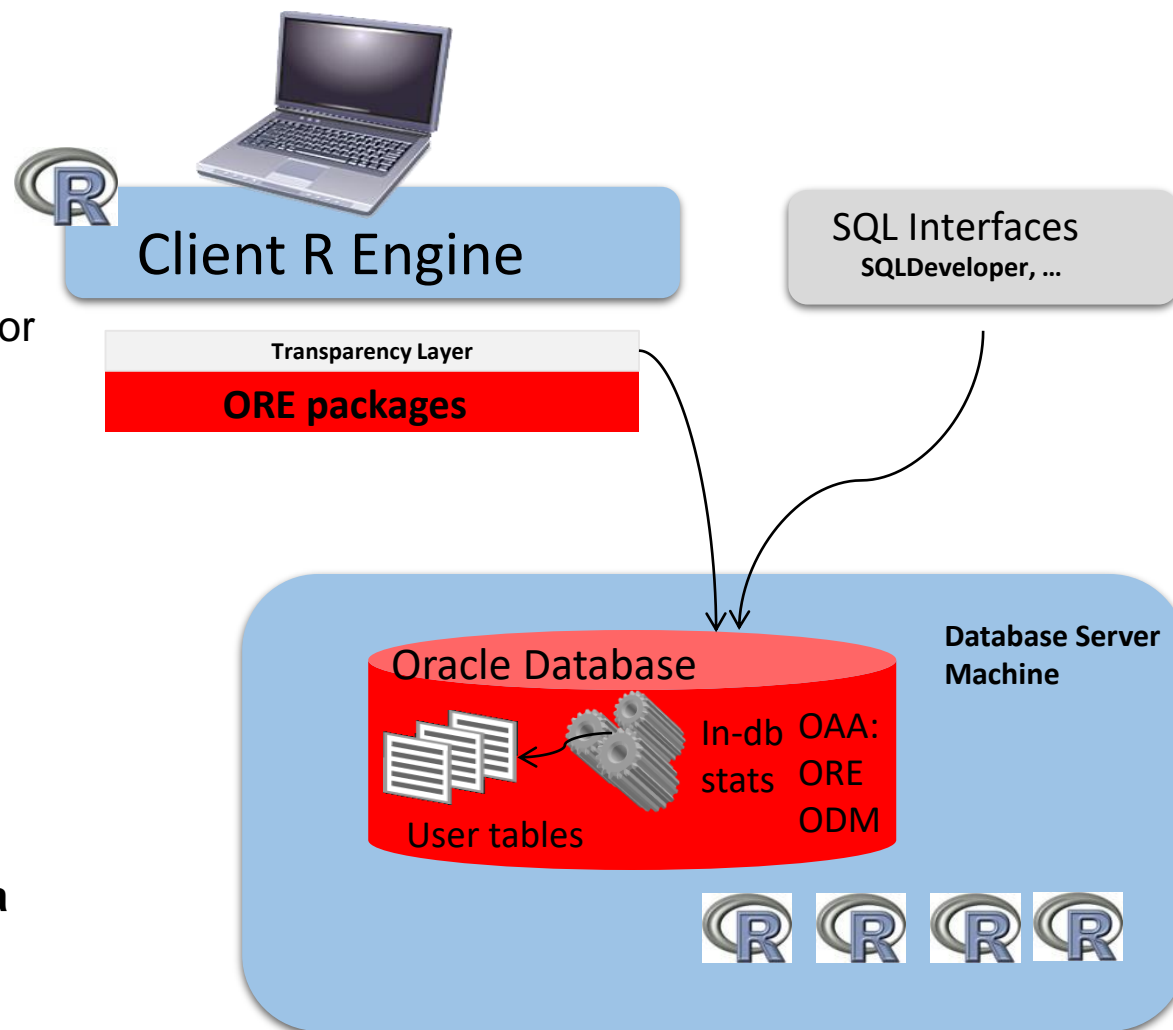
- Oracle R Enterprise (ORE) is a component of the Oracle Advanced Analytics (OAA) option to Oracle Database EE
- Provides transparent access to database-resident data from R
- Execute R scripts at the database machine managed by Oracle Database with data and task parallelism
- Execute R scripts from SQL
- Integrates R into the IT software stack
- Extends and enhances open source R





# Oracle R Enterprise

- A comprehensive, database-centric environment for end-to-end analytical processes in R, with immediate deployment to production environments
- Operationalize entire R scripts in production applications – eliminate porting R code
- Seamlessly leverage Oracle Database as an HPC environment for R scripts, providing data parallelism and resource management
- Avoid reinventing code to integrate R results into existing applications
- Transparently analyze and manipulate data in Oracle Database through R using versatile and customizable R functions
- Eliminate memory constraint of client R engine
- Score R models in Oracle Database
- Execute R scripts through Oracle Database server machine for scalability and performance
- **Get maximum value from your Oracle Database and Exadata**
- Enable integration and management through SQL
- Integrate R into the IT software stack, e.g. OBIEE





# R now integrated into OBIEE 11g and 12c

8.13 Oracle R Enterprise

Custom Link Home Catalog Favorites Dashboards New Open Signed In As Paulo Rodney

R Integration R End-User Interaction **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)

Clear All Filters

Script Name

- Balloon Plot
- BusiestAirports-1
- Conditional Histogram
- Correlation Matrix Ellipses
- Correlation matrix circles
- Enhanced Bar Plot
- Heatmap
- Multipanel Geo Lattice Plot
- Multivariate Star Plot
- PredictDelays-build
- PredictDelays-score
- QCC-1**
- RandomRedDots
- Times Series Forecasting
- Volcano Perspective Plot
- Volcano Perspective Plot2
- syncFunction

Param 1: cursor(select q'ALL' UNIQUE\_CARRIERS,q'both' GRA

Param 2: cursor(select 550 "ore.png.height", 550 "ore.i

Apply

**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, 1]  
      obj <- qcc(set1, type=gt, title=title)  
    } else {  
      title <- paste(gt, "Chart for", delayType, "delay -",  
                    years, startYear, "-", endYear)  
      set1 <- group[1: numYears, 1]  
      set2 <- group[numYears + 1: length(group), 1]  
      obj <- qcc(set1, set2, type=gt, title=title)  
    }  
  }  
}
```

**QCC-1 Graphics Result**

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

Calibration data in set1 New data in set2

Group summary statistics

UCL CL LCL

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

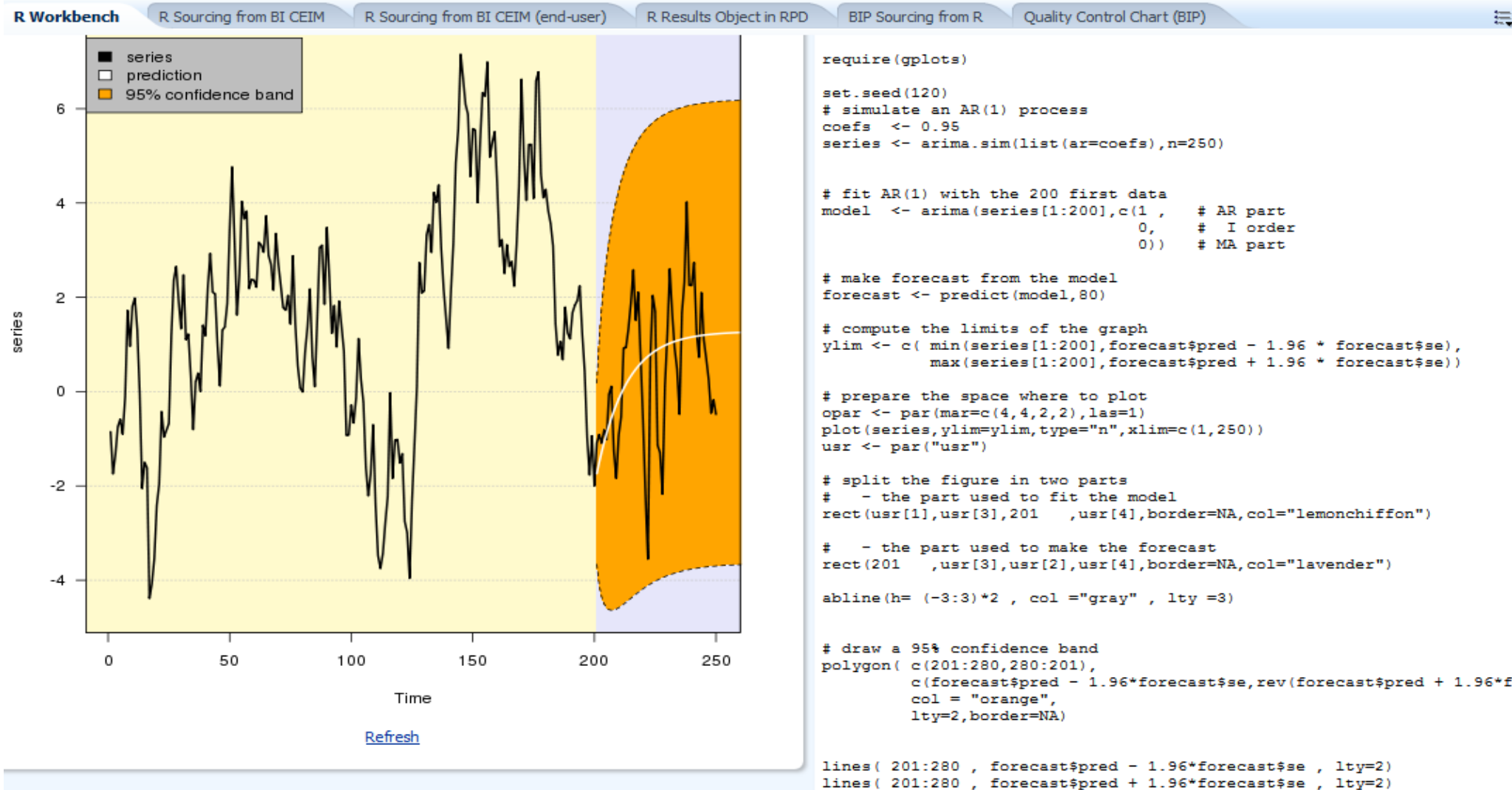
Calibration data in set1 New data in set2

Group summary statistics

UCL CL LCL



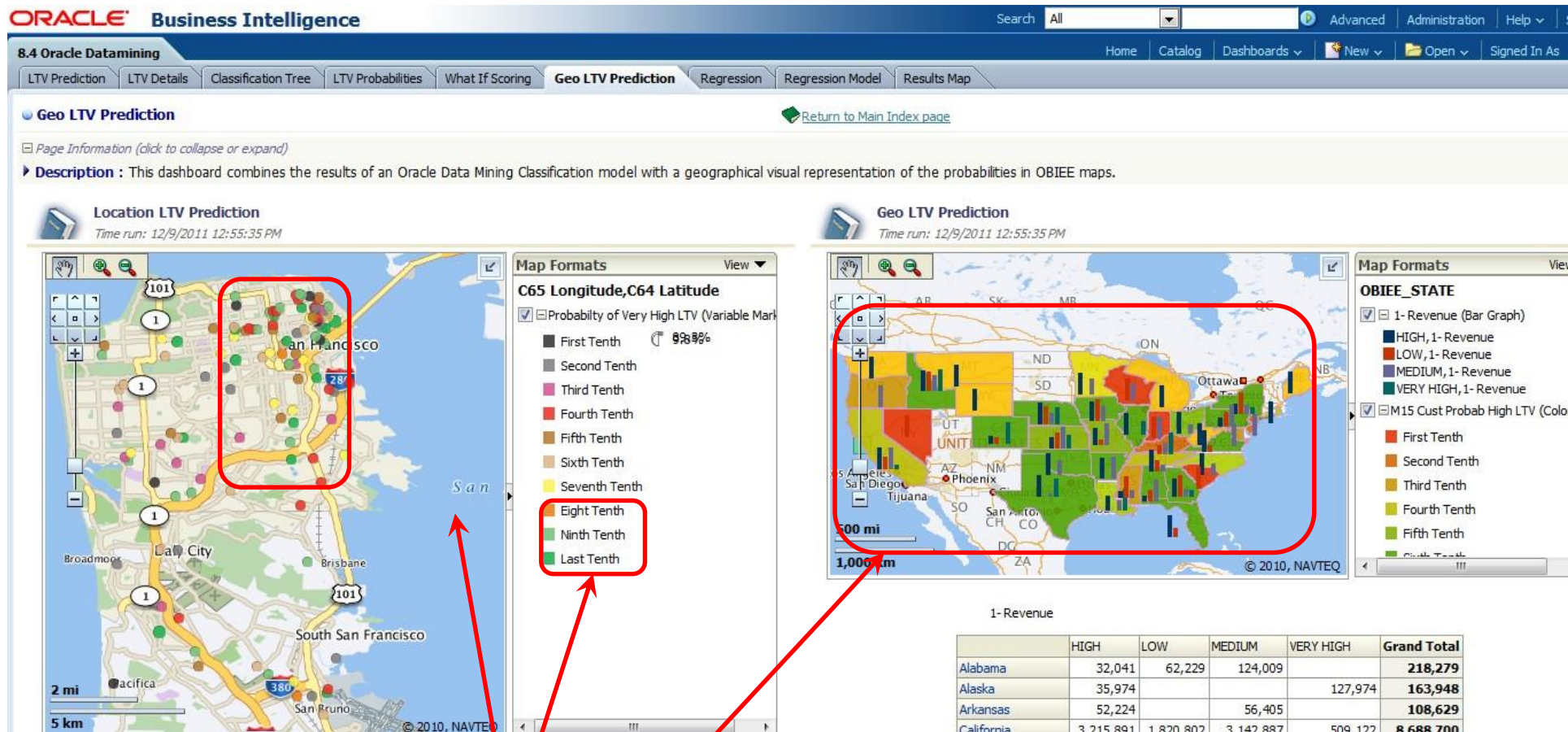
# R now integrated into OBIEE 11g and 12c







# Oracle Advanced Analytics & Spatial



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



# 5 Common use cases for predictive analytics

- 1) Customer Segmentation using Clustering algorithms
  - Discovered patterns can be extremely meaningful
  - Able to include hundreds of dimensions
  - Great first project
- 2) Predict Lifetime Customer Value
  - Measure impact of different product purchases on LCV
  - Promote and incentive profitable purchases



# 5 Common use cases for predictive analytics

- 3) Market Basket Analysis for retailers and warehouses
  - Understand purchasing and picking patterns
- 4) Employee Retention analysis
  - Classify employees into basic categories
  - Understand impact of different incentives and rewards
- 5) Optimize Customer Service and Next Best Offer
  - Use decision trees to determine rules for customers
  - Dramatically increase effectiveness of offers





# Basic Ways to Get Started

- Do a POC project on your own
- Conduct a workshop for key stakeholders to build support
  - One hour to one day
- Conduct a “predictive analytics summit” BYOD workshop
- Conduct ODM and ORE training classes with 1-day workshop
- Use a defined Quick Start program (2 weeks)



# ODM Quick Start Overview

- Oracle Database Cloud Service
  - High Performance or Extreme Performance
- 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)



# Quick Start 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





# Bring Your Own Data BYOD Workshop

- Requires preparation and pre-workshop activities by participants and organizers
- Develop business use cases and data sets in advance
- Prepare environments and verify data sets in advance
- Determine depth of predictive/advanced analytics to be used in workshop
- Include participants from across the organization
- Have experts/proctors with expertise in interface and analytics
- Spread risk across many business cases (VC model)



# Oracle Data Mining Training (2 days)

- Introduction
- Data Mining Concepts and Terminology
- The Data Mining Process
- Introducing Oracle Data Miner
- 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



# Important Factors in Getting Started

- Lots of internal experts and 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
  - Bring people along on the journey, establish culture
  - Everyone shares a minimum common knowledge base
- Use workshops (JAD style session) for 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



# Oracle Test Drive

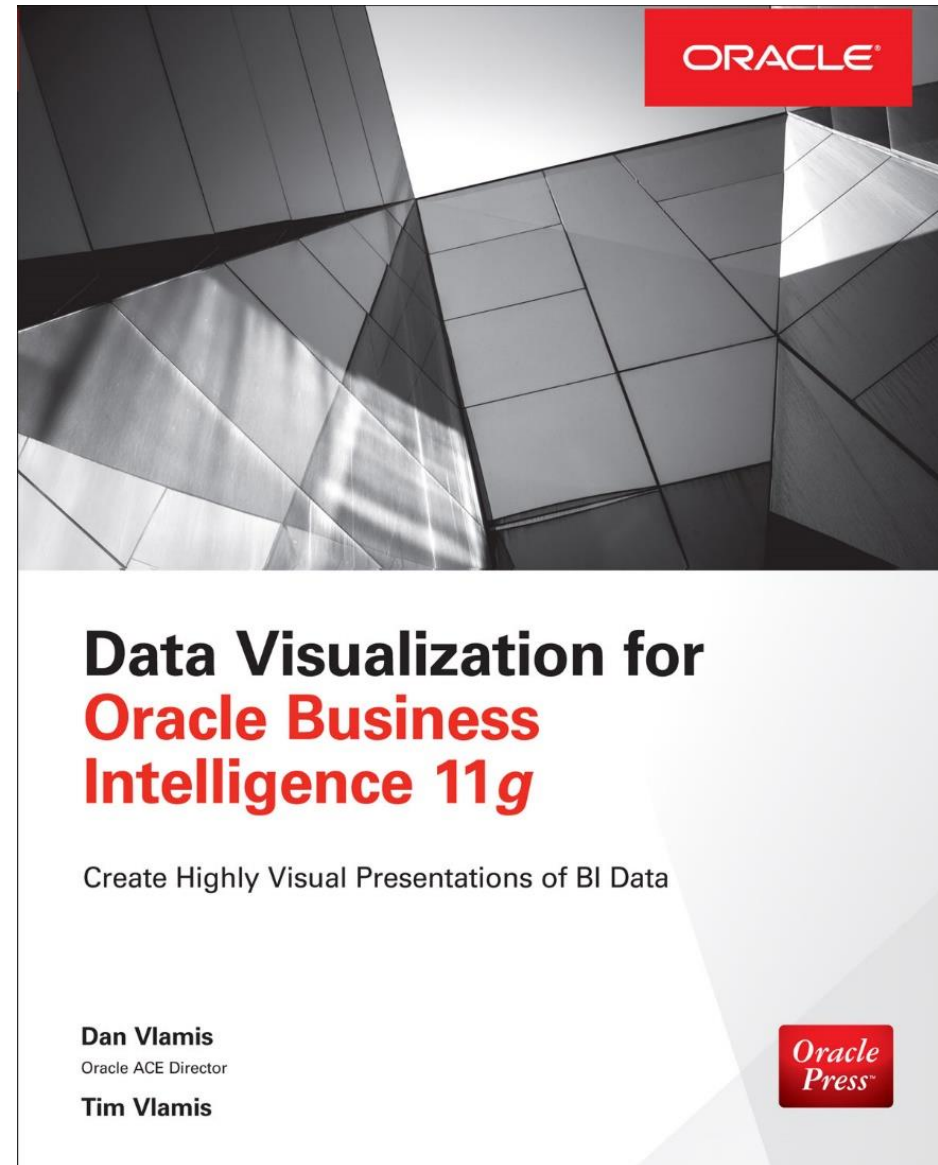
- Free to try Oracle BI, Advanced Analytics Go to [www.vlamis.com/td](http://www.vlamis.com/td)
- Runs off of Oracle Cloud
- Test Drives for:
  - Oracle BI
  - Oracle Advanced Analytics
- Once sign up, you have private instance for one day
- Available now





# Drawing for Free Book

Add business card to basket  
or fill out card





# Thank You!

## Starting Smart with Oracle Advanced Analytics

Tim Vlami

[tvlamis@vlamis.com](mailto:tvlamis@vlamis.com)

[@TimVlami](#)

[@VlamiSoftware](#)

[www.vlamis.com](http://www.vlamis.com)

# ODTUG Kscope17



SAN ANTONIO, TEXAS ★ JUNE 25-29



PLEASE FILL OUT YOUR EVALUATIONS