



Oracle Data Mining 11g: Overview, Demos, Exadata and Road Map

Charlie Berger

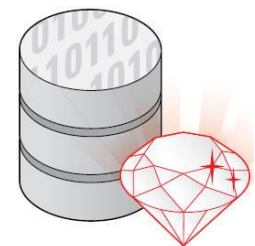
Sr. Director Product Management, Data Mining Technologies

Oracle Corporation

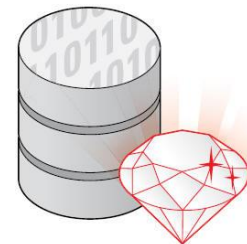
charlie.berger@oracle.com

www.twitter.com/CharlieDataMine

ORACLE®



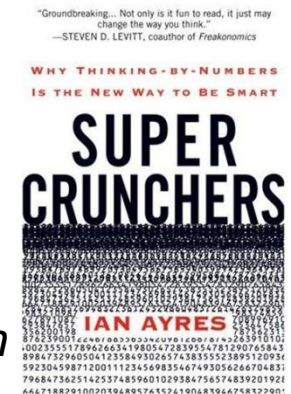
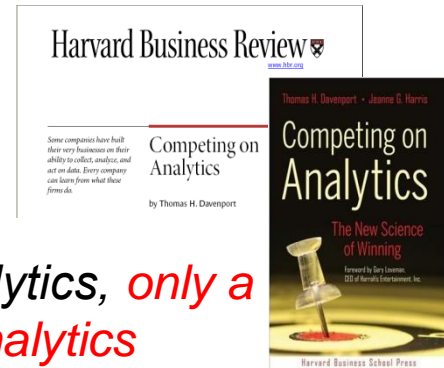
Copyright 2010 Oracle Corporation



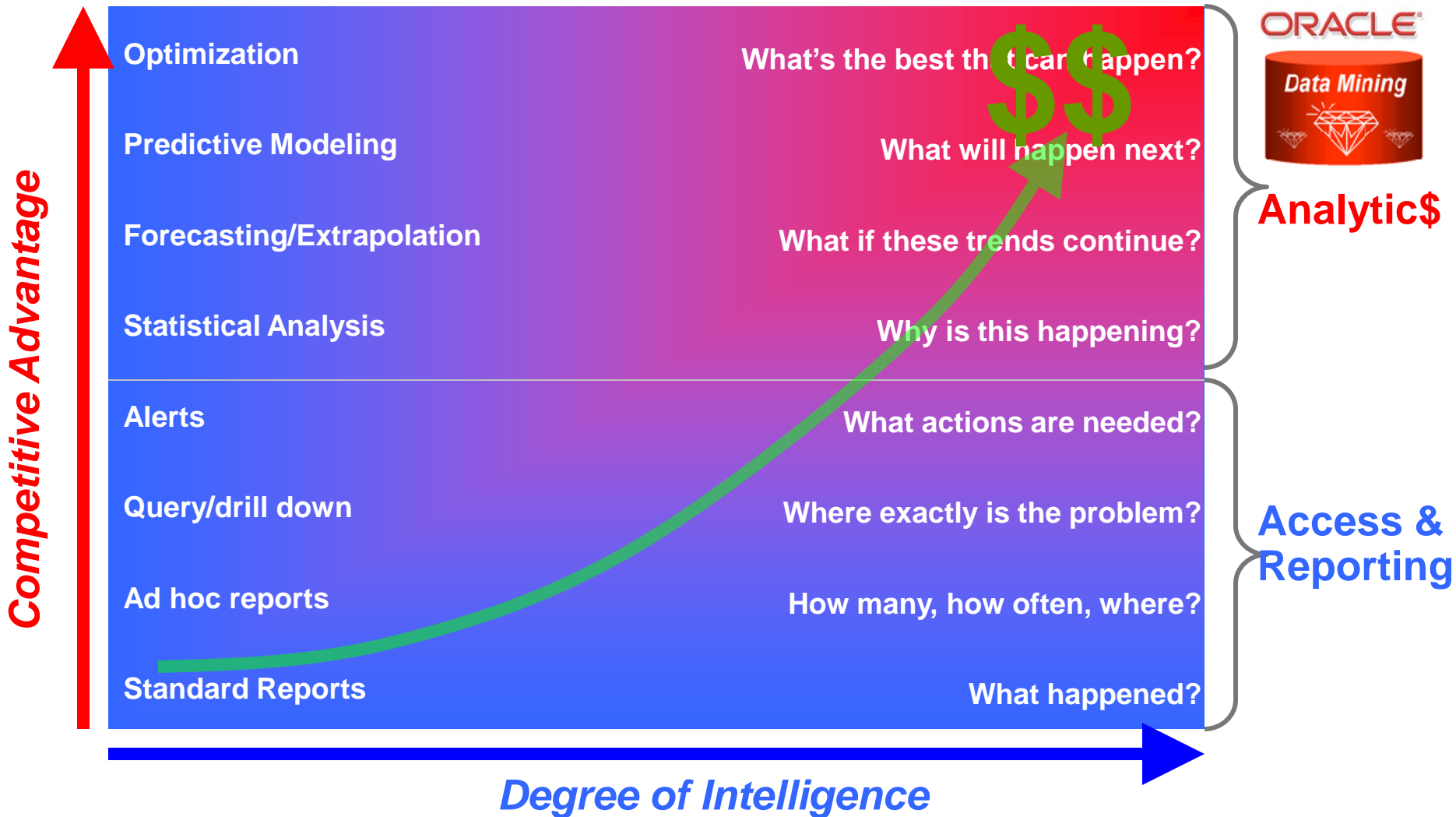
Market Drivers

Analytics: Strategic and Mission Critical

- *Competing on Analytics*, by Tom Davenport
 - “Some companies have built their very businesses on their ability to collect, analyze, and act on data.”
 - “Although numerous organizations are embracing analytics, *only a handful have achieved this level of proficiency. But analytics competitors are the leaders in their varied fields—consumer products finance, retail, and travel and entertainment among them.*”
 - “Organizations are moving beyond query and reporting” - IDC 2006
- *Super Crunchers*, by Ian Ayers
 - “In the past, one could get by on intuition and experience. Times have changed. *Today, the name of the game is data.*”
—Steven D. Levitt, author of *Freakonomics*
 - “*Data-mining and statistical analysis have suddenly become cool.... Dissecting marketing, politics, and even sports, stuff th complex and important shouldn't be this much fun to read.*” —Wired



Competitive Advantage



Analytics\$

Access & Reporting



- 11 years “stem celling analytics” into Oracle
 - Designed advanced analytics into database kernel to leverage relational database strengths
 - Naïve Bayes and Association Rules—1st algorithms added
 - Leverages counting, conditional probabilities, and much more
- Now, analytical database platform
 - 12 cutting edge machine learning algorithms and 50+ statistical functions
 - A data mining model is a schema object in the database, built via a PL/SQL API and scored via built-in SQL functions.
 - When building models, leverage existing scalable technology
 - (e.g., parallel execution, bitmap indexes, aggregation techniques) and add new core database technology (e.g., recursion within the parallel infrastructure, IEEE float, etc.)
 - True power of embedding within the database is evident when scoring models using built-in SQL functions (incl. Exadata)

```
select cust_id
from customers
where region = 'US'
and prediction probability(churnmod, 'Y' using *) > 0.8;
```

You Can Think of It Like This...

Traditional SQL

- “Human-driven” queries
- Domain expertise
- Any “rules” must be defined and managed

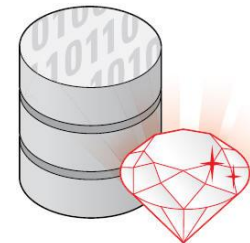
- SQL Queries
 - SELECT
 - DISTINCT
 - AGGREGATE
 - WHERE
 - AND OR
 - GROUP BY
 - ORDER BY
 - RANK



Oracle Data Mining

- Automated knowledge discovery, model building and deployment
- Domain expertise to assemble the “right” data to mine

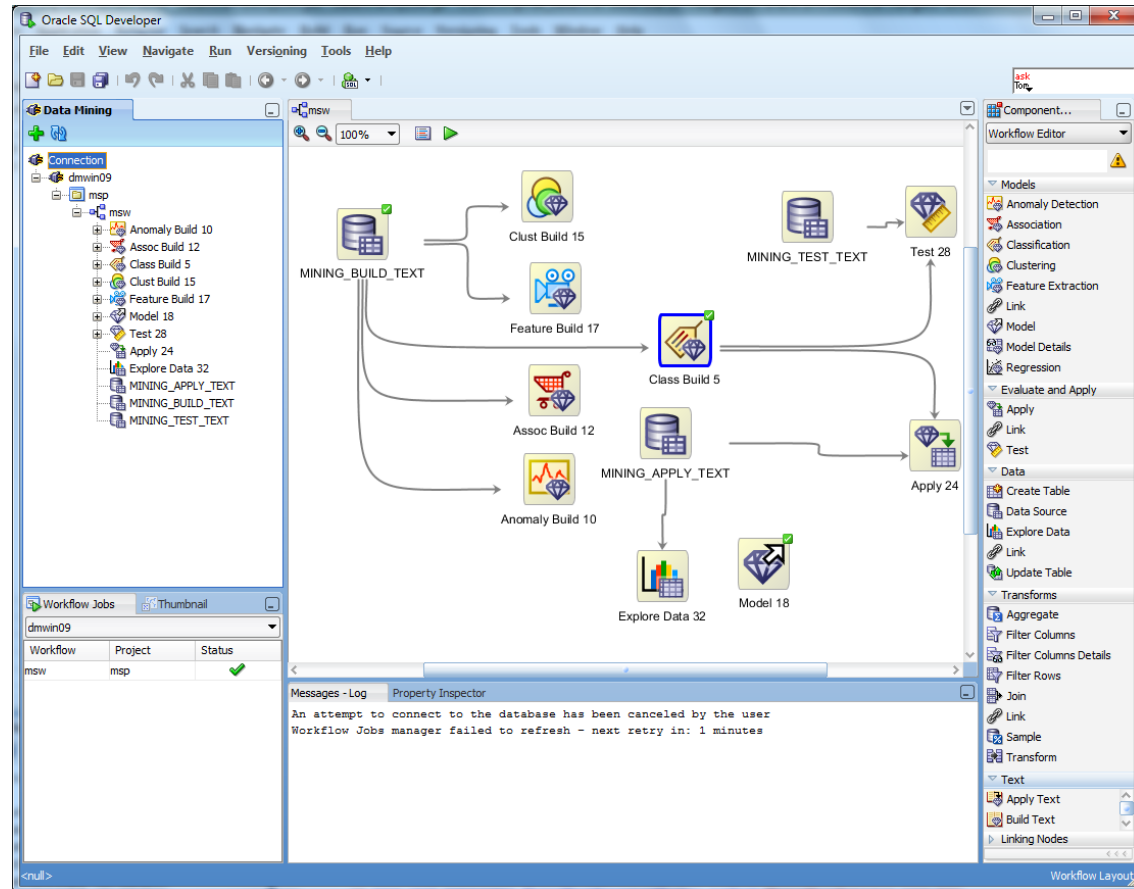
- ODM “Verbs”
 - PREDICT
 - DETECT
 - CLUSTER
 - CLASSIFY
 - REGRESS
 - PROFILE
 - IDENTIFY FACTORS
 - ASSOCIATE



Oracle Data Miner 11gR2 New GUI

Optional GUI for Oracle Data Mining Option

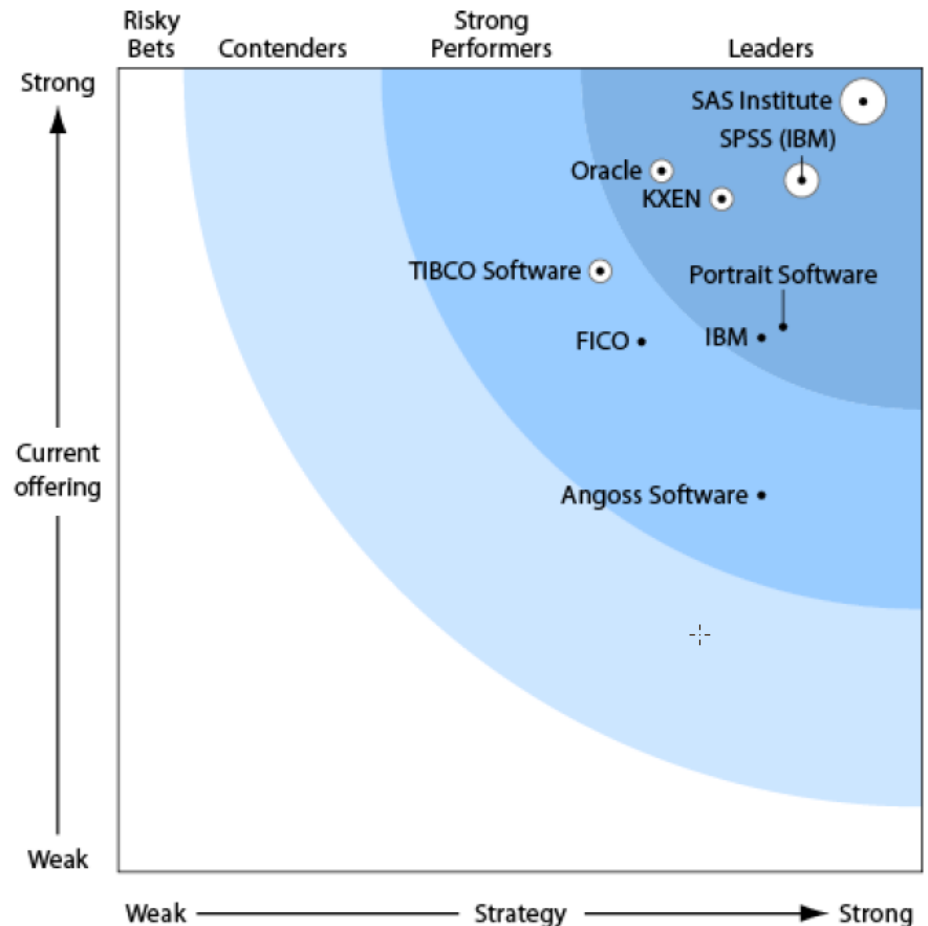
- Graphical User Interface for data analyst
- SQL Developer Extension (OTN download)
- Explore data—discover new insights
- Build and evaluate data mining models
- Apply predictive models
- Share analytical workflows
- Deploy SQL Apply code/scripts



The Forrester Wave™: Predictive Analytics And Data Mining Solutions, Q1 2010

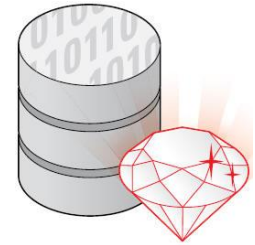
Oracle Data Mining Cited as a Leader; 2nd place in Current Offering

- Ranks 2nd place in Current Offering
- “Oracle focuses on in-database mining in the Oracle Database, on integration of Oracle Data Mining into the kernel of that database, and on leveraging that technology in Oracle’s branded applications.”

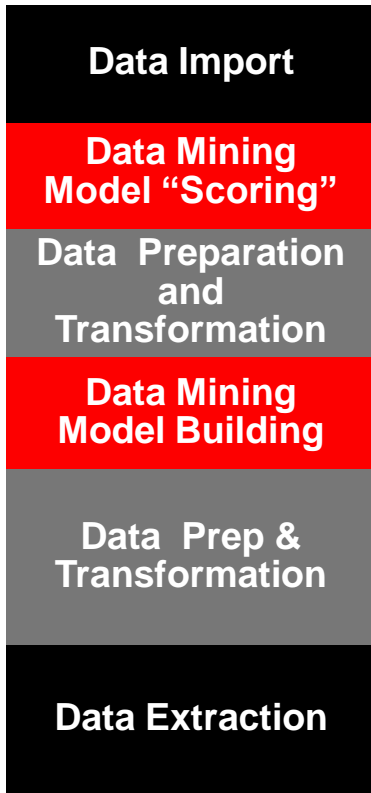


The Forrester Wave is copyrighted by Forrester Research, Inc. Forrester and Forrester Wave are trademarks of Forrester Research, Inc. The Forrester Wave is a graphical representation of Forrester's call on a market and is plotted using a detailed spreadsheet with exposed scores, weightings, and comments. Forrester does not endorse any vendor, product, or service depicted in the Forrester Wave. Information is based on best available resources. Opinions reflect judgment at the time and are subject to change.

In-Database Data Mining



Traditional Analytics



Oracle Data Mining



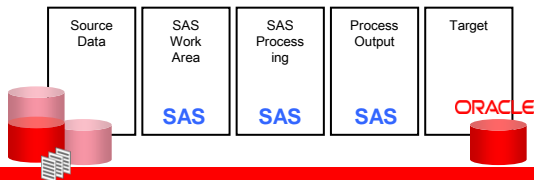
Savings

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

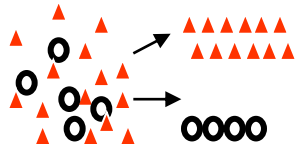
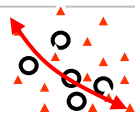
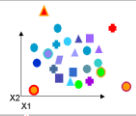
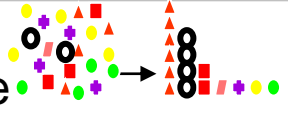
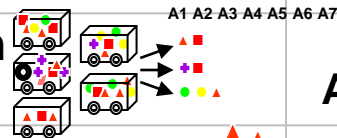
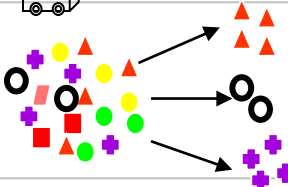
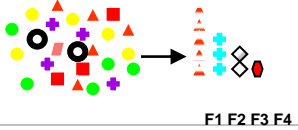
Hours, Days or Weeks



Secs. Mins or Hours



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 	Multiple Regression (GLM) Support Vector Machine	Classical statistical technique Wide / narrow data / text
Anomaly Detection 	One Class SVM	Lack examples
Attribute Importance 	Minimum Description Length (MDL)	Attribute reduction Identify useful data Reduce data noise
Association Rules 	Apriori	Market basket analysis Link analysis
Clustering 	Hierarchical K-Means Hierarchical O-Cluster	Product grouping Text mining Gene and protein analysis
Feature Extraction 	NMF	Text analysis Feature reduction

Oracle Data Mining + Exadata



- In 11gR2, SQL predicates and Oracle Data Mining models are pushed to storage level for execution

For example, find the US customers likely to churn:

```
select cust_id
from customers
where region = 'US'
and prediction_probability(churnmod, 'Y' using *) > 0.8;
```

Scoring function executed in Exadata

Oracle Data Miner 11gR2 GUI

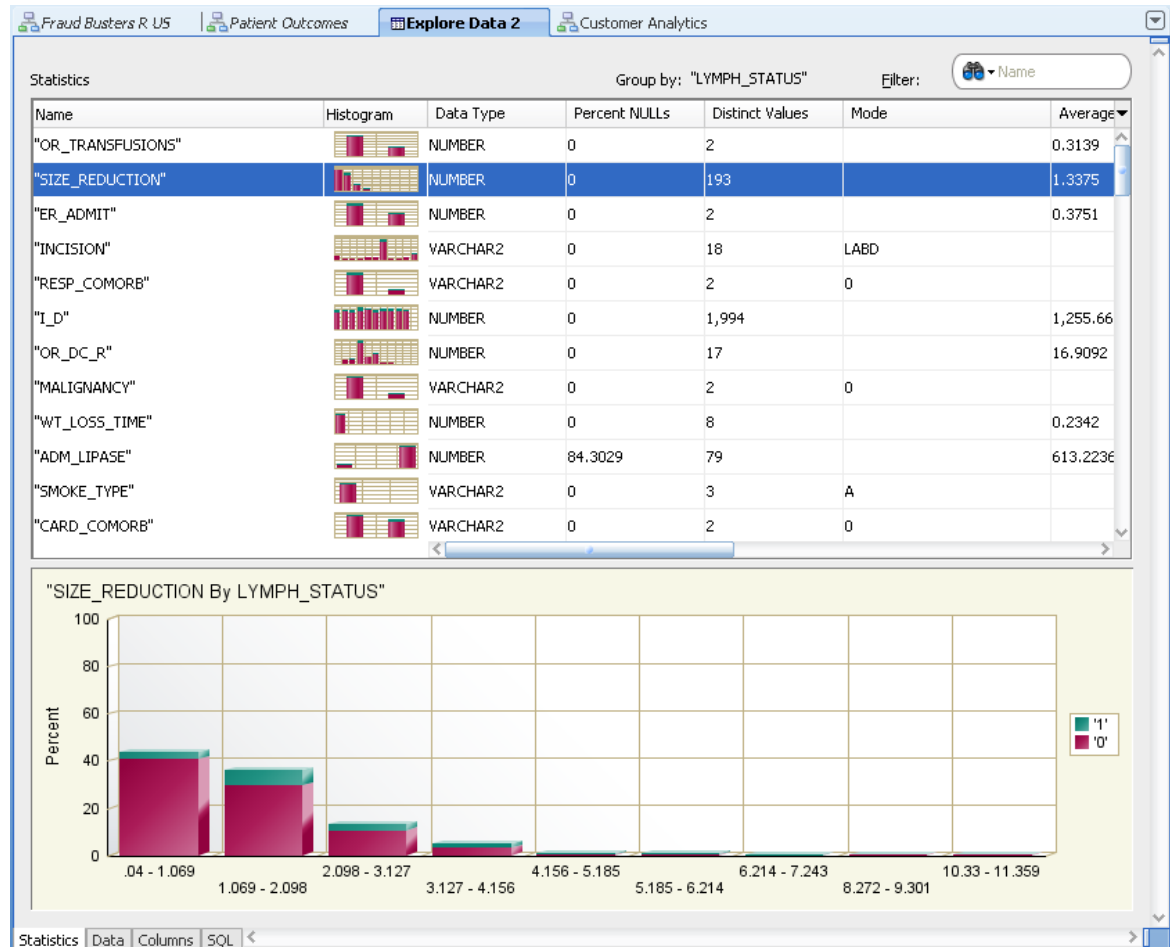
- Predict customer behavior
- Identify key factors
- Predict next-likely product
- Customer profiling
- Detect fraud & anomalies
- Mine “text” and unstructured data

The screenshot displays the Oracle Data Miner 11gR2 GUI. The main window shows a workflow editor with a process flow: Lymphoma -> Explore Data 2 -> Filter Columns 4 -> Outcome predictive models -> At Risk patients. A 'Discover patient clusters' process is also connected to the 'Lymphoma' data source. The 'Outcome predictive models - Property Inspector' window is open, showing a table of models.

Name	Build	Test	Tune	Algorithm	Comment
CLAS_GLM_1_7	6/15/10 11:00...	6/15/10 11:00...	Automatic	Generalized Linea...	
CLAS_SVM_1_7	6/15/10 10:59...	6/15/10 11:00...	Automatic	Support Vector M...	
CLAS_DT_1_7	6/15/10 11:08...	6/15/10 11:08...	Automatic	Decision Tree	
CLAS_NB_1_7	6/15/10 10:59...	6/15/10 10:59...	Automatic	Naive Bayes	

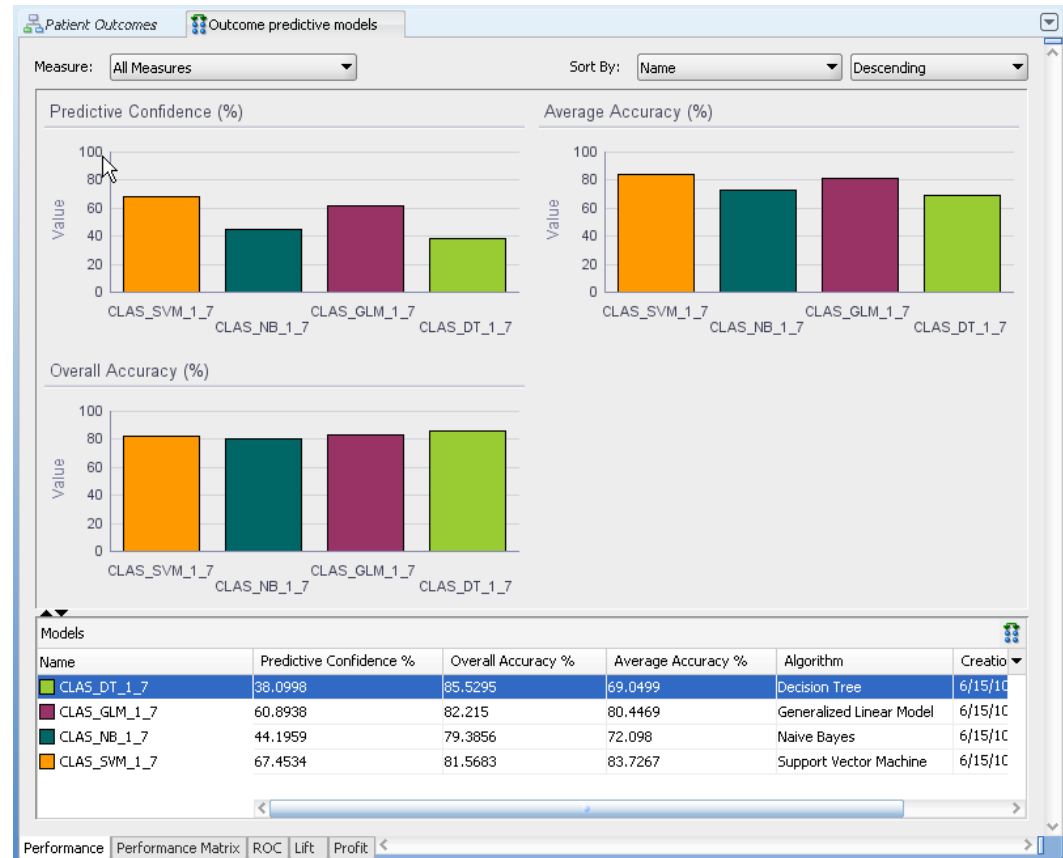
Explore Data

- Thumbnail distributions of every attribute
 - Grouped by another attribute
- Summary statistics for all attributes
 - Min, max, stdev, variance
 - median, mean, skewness, kurtosis, etc.



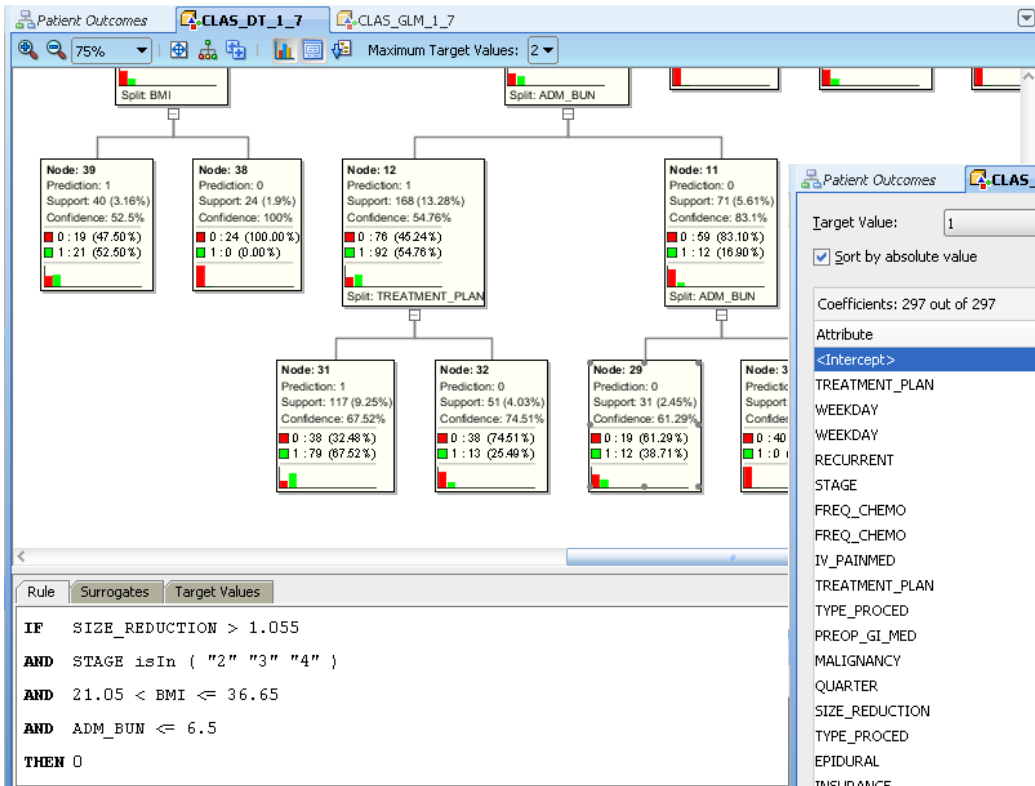
Build and Evaluate Models

- Comparative model performance results
- Adjust and tune predictive models



Understand Model Details

- Interactive model viewers



```

Rule Surrogates Target Values
IF SIZE_REDUCTION > 1.055
AND STAGE isIn ( "2" "3" "4" )
AND 21.05 < BMI <= 36.65
AND ADM_BUN <= 6.5
THEN 0
    
```

Target Value: 1

Sort by absolute value

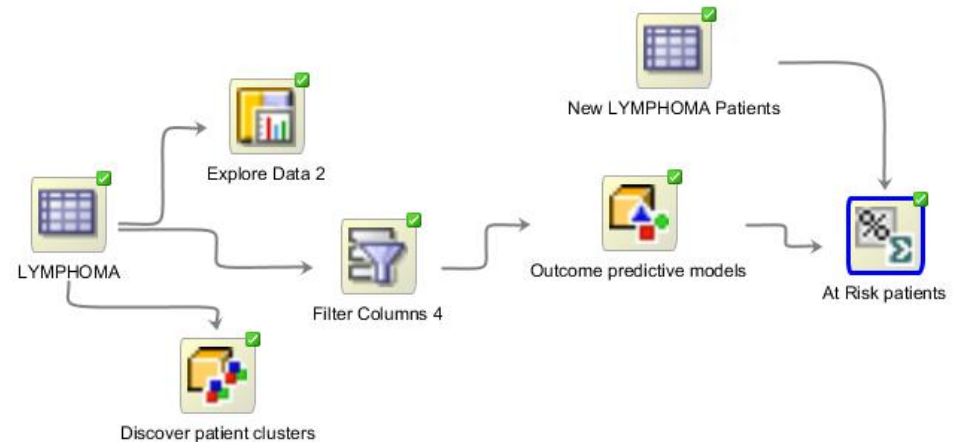
Fetch Size: 10,000

Coefficients: 297 out of 297

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_PROCED	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_PROCED	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

Analytical “Work Flow” Methodologies

- *Build, share and automate predictive analytics methodologies*

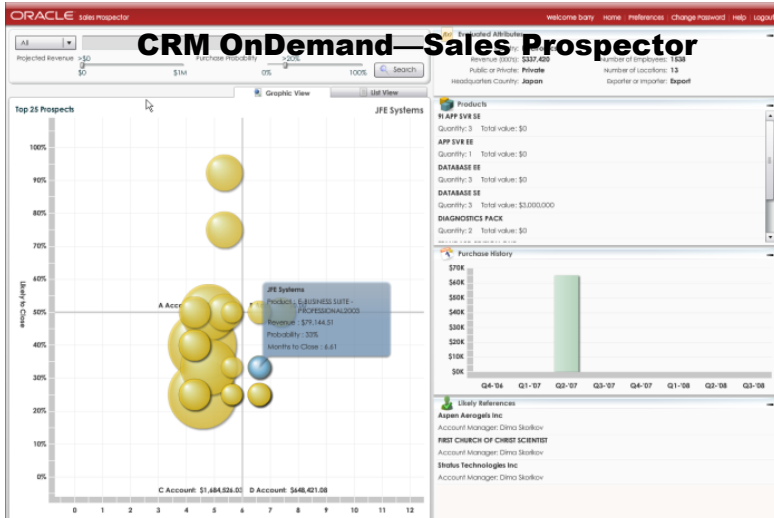


At Risk patients

	CLAS_SVM_1_7_PRED	CLAS_SVM_1_7_PROB	LYMPH_TYPE	SIZE_TUMOR_MM	MARITAL	ADM_ALBUMIN	AMT_CHEMO	FREQ_CHEMO	CLAS_DT_
1	1	0.99865991	Agressive	7,100	M	1.6	42.17	1	
2	1	0.99865991	Agressive	7,100	M	1.6	42.17	1	
3	1	0.99865991	Agressive	7,100	M	1.6	42.17	1	
4	1	0.99351446	Agressive	5,200	M	2.4	52	1	
5	1	0.99351446	Agressive	5,200	M	2.4	52	1	
6	1	0.99351446	Agressive	5,200	M	2.4	52	1	
7	1	0.99149541	Agressive	1,350	S	2.4	37.01	2	
8	1	0.99149541	Agressive	1,350	S	2.4	37.01	2	
9	1	0.99149541	Agressive	1,350	S	2.4	37.01	2	
10	1	0.99149541	Agressive	1,350	S	2.4	37.01	2	
11	1	0.9912111	Indolent	3,400	W		3.25	2	
12	1	0.9912111	Indolent	3,400	W		3.25	2	
13	1	0.9912111	Indolent	3,400	W		3.25	2	
14	1	0.9912111	Indolent	3,400	W		3.25	2	
15	1	0.98217842	Indolent	1,000	M	2.4	52.25	1	
16	1	0.98217842	Indolent	1,000	M	2.4	52.25	1	

Predictive Analytics Applications

Powered by Oracle Data Mining (Partial List as of March 2010)

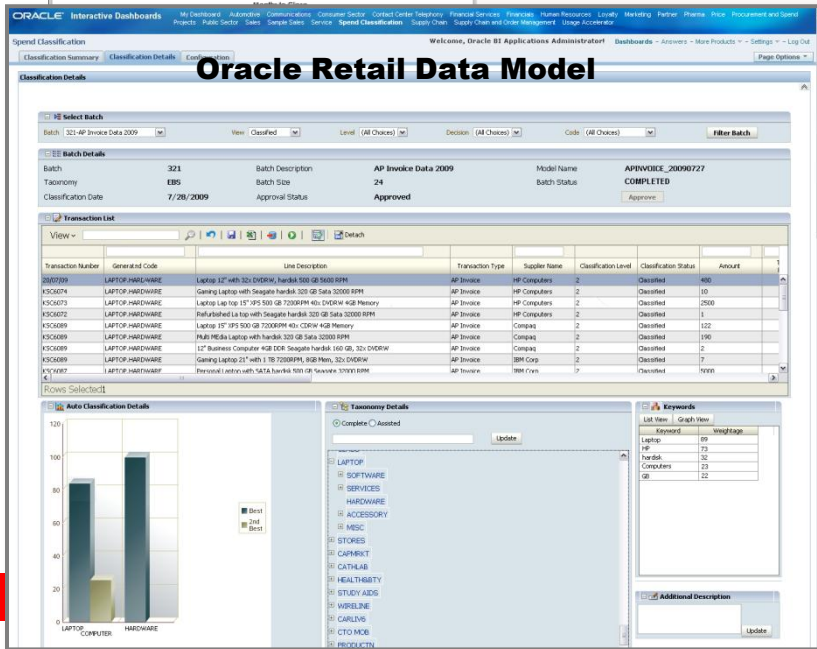


Oracle Communications Data Model

Churn Prediction by Customer Segment

Customer Segment	Customer Name	Cell Phone No	Contract Value	Month Revenue	Debt Value	LTV Band	LTV Value	LTV Months	ARPU Band	Churn Indicator	Sentiment	Churn Probability	Customer Segment Key
	Chloe White	9985005370	\$0.00	\$222.00			\$65,000.00	10		▲+		56	101
	Delora Walker	9985009300	\$0.00	\$130.00			\$85,000.00	18		▲+		30	101
	Max Gerber	9985006181	\$3,000.00	\$2,500.00	\$222.00		\$79,000.00	17		▲+		39	101
	Olen Christian	9985008393	\$0.00	\$130.00			\$59,000.00	13		▼-		82	101
	Mason Murray	9985007979	\$9,000.00	\$7,500.00	\$70.00		\$56,000.00	21		▼-		94	101
	Deb Coe	9985007379	\$3,000.00	\$2,500.00	\$130.00		\$89,000.00	10		▲+			101

Oracle Open World - Schedule Builder



Spend Classification

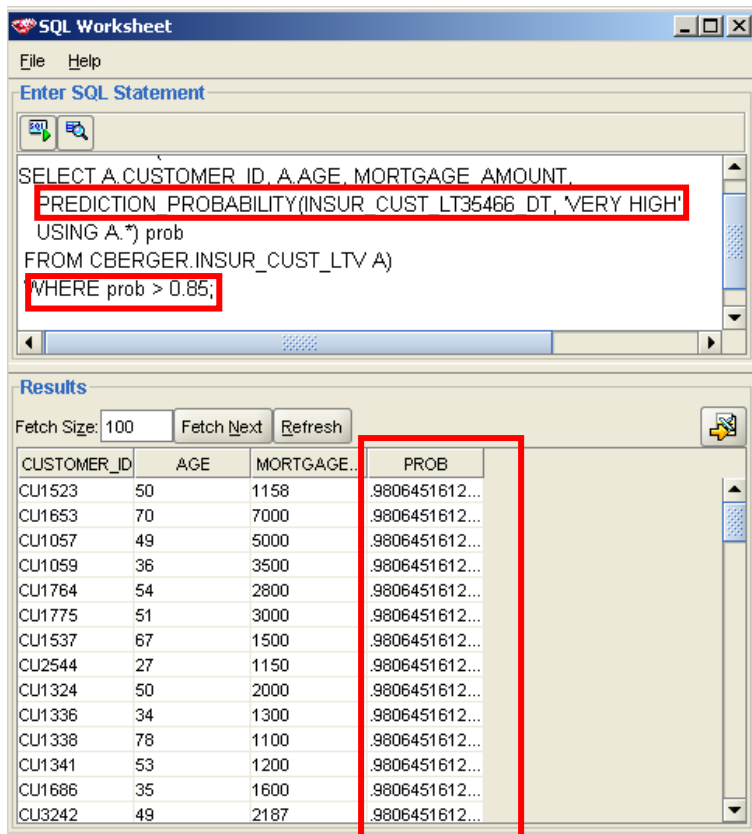
Year: 2007, Month: BY 2007 M2, Division: Pittsburgh Division, District: , Store Name: PITTSBURGH 109

file	Performance Measure Value	% of Supporting Transaction	Probability
since = '1', '2', '3', '4', '5' and Year's_OF_RESIDENCE = '1', '2', '3' and Household Size = '3+' and IDENCE = '1'	LEAST LOYAL	2.45%	94.76%
since = '1', '2', '3', '4', '5' and Year's_OF_RESIDENCE = '1', '2', '3' and Household Size = '3+' and IDENCE = '1' and Marital Status = 'UNMARRIED', 'SEPARATED'	LEAST LOYAL	0.13%	100.00%
since = '1', '2', '3', '4', '5' and Year's_OF_RESIDENCE = '1', '2', '3' and Household Size = '3+' and IDENCE = '1' and Marital Status = 'MARRIED', 'SINGLE'	LEAST LOYAL	1.60%	87.85%
since = '1', '2', '3', '4', '5' and Year's_OF_RESIDENCE = '4', '5' and Household Size = '3+'	PRETTY LOYAL	10.93%	83.17%
since = '1', '2', '3', '4', '5' and Year's_OF_RESIDENCE = '4', '5' and Household Size = '3+' and Marital Status = 'SINGLE'	PRETTY LOYAL	7.38%	81.58%
since = '1', '2', '3', '4', '5' and Year's_OF_RESIDENCE = '4', '5' and Household Size = 'LESS THAN 3'	MARGINALLY LOYAL	9.60%	86.29%
since = '1', '2', '3', '4', '5' and Year's_OF_RESIDENCE = '4', '5' and Household Size = 'LESS THAN 3' and 'MARRIED', 'SINGLE'	MARGINALLY LOYAL	6.52%	84.17%
since = '10', '8', '7', '8', '9' and Household Size = '3+'	MOST LOYAL	25.01%	88.28%
since = '10', '8', '7', '8', '9' and Household Size = '3+' and Marital Status = 'MARRIED', 'SINGLE'	MOST LOYAL	17.11%	86.10%
since = '10', '8', '7', '8', '9' and Household Size = 'LESS THAN 3'	PRETTY LOYAL	26.20%	80.88%



Example: Simple, Predictive SQL

Select customers who are **more than 85% likely to be HIGH VALUE customers** & display their AGE & MORTGAGE_AMOUNT



The screenshot shows an Oracle SQL Worksheet window. The top section is titled "Enter SQL Statement" and contains the following SQL query:

```
SELECT A.CUSTOMER_ID, A.AGE, MORTGAGE_AMOUNT,  
PREDICTION_PROBABILITY(INSUR_CUST_LT35466_DT, 'VERY HIGH'  
USING A.*) prob  
FROM CBERGER.INSUR_CUST_LTV A)  
WHERE prob > 0.85;
```

The bottom section is titled "Results" and displays a table with the following data:

CUSTOMER_ID	AGE	MORTGAGE..	PROB
CU1523	50	1158	.9806451612...
CU1653	70	7000	.9806451612...
CU1057	49	5000	.9806451612...
CU1059	36	3500	.9806451612...
CU1764	54	2800	.9806451612...
CU1775	51	3000	.9806451612...
CU1537	67	1500	.9806451612...
CU2544	27	1150	.9806451612...
CU1324	50	2000	.9806451612...
CU1336	34	1300	.9806451612...
CU1338	78	1100	.9806451612...
CU1341	53	1200	.9806451612...
CU1686	35	1600	.9806451612...
CU3242	49	2187	.9806451612...

```
SELECT * from(  
SELECT A.CUST_ID, A.AGE,  
MORTGAGE_AMOUNT, PREDICTION_PROBABILITY  
(CUST_INSUR_LT46939_DT, 'VERY HIGH'  
USING A.*) prob  
FROM CBERGER.CUST_INSUR_LTV A)  
WHERE prob > 0.85;
```

Fraud Prediction Demo

```
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',
'CLAIMS2', '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 CLAIMS2
where PASTNUMBEROFCLAIMS in ('2 to 4', 'more than 4')))
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



More Interesting SQL

(Missing Value Imputation Example)

Select the 10 customers who are **most likely to attrite** based solely on: age, gender, annual_income, and zipcode. In addition, since annual_income is often missing, perform **null/missing value imputation** for the annual_income attribute using all of the customer demographics.

```
SELECT * FROM (
  SELECT cust_name, cust_contact_info,
         rank() over (ORDER BY
           PREDICTION_PROBABILITY(attrition_model, 'attrite'
            USING age, gender, zipcode,
              NVL(annual_income,
                 PREDICTION(estim_income USING *)))
           as annual_income) DESC) as cust_rank
  FROM customers)
WHERE cust_rank < 11;
```

Real-time Prediction

with

```
records as (select
  78000 SALARY,
  250000 MORTGAGE_AMOUNT,
  6 TIME_AS_CUSTOMER,
  12 MONTHLY_CHECKS_WRITTEN,
  55 AGE,
  423 BANK_FUNDS,
  'Married' MARITAL_STATUS,
  'Nurse' PROFESSION,
  'M' SEX,
  4000 CREDIT_CARD_LIMITS,
  2 N_OF_DEPENDENTS,
  1 HOUSE_OWNERSHIP from dual)
```

```
select s.prediction prediction, s.probability probability
```

```
from (
```

```
  select PREDICTION_SET(CUST_INSUR_LT46939_DT, 1 USING *) pset
  from records) t, TABLE(t.pset) s;
```

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

PREDICTION	PROBABILITY
HIGH	.65123504738232096

Example of Embedded Predictive SQL

Powers Next Generation Predictive Marketing Tools

Adobe Reader - [1166547846129out.pdf]

File Edit View Document Tools Window Help

Pages

UNIVERSITY.edu

Promoting educational excellence

University.edu
Management Administration
1234 State Street
Bloomington, IN 88888

December 19, 2006

Dear 101504,

Over the past year and a half the University.edu Alumni Association has been busy lending its expertise to the U.edu staff, orchestrating equipment donations, coordinating and participating in a major station clean-up, and establishing an alumni-student mentoring program.

E.EAA wants to raise funds to provide a stipend for the U.Edu general manager and to pay an engineer to repair vital equipment.

Because you have been so generous in the past we are first asking you, our alumni, for financial assistance. We suggest a donation of \$25, \$50, or whatever feels right. (If you or your company would like information on underwriting a program, please contact us at 303.575.3548.)

Your donation is fully tax deductible. We urge you to take your donation this tax year by sending in your donation by December 31, 2006.

In the coming months look for fundraising effort updates in our newsletter. And, be sure to mark your calendar now for our annual meeting and awards dinner -- Saturday evening, April 27, at U.ECP. You'll soon receive more information on this special fund-raiser.

Sincerely,

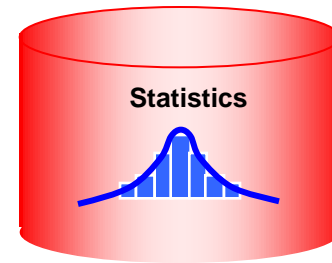
University.edu Fund Raising Staff

Attachments

Notice that the id of the likely responder is included.

Letter personalized with embedded predictive analytics

11g Statistics & SQL Analytics

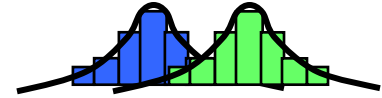


- Ranking functions
 - rank, dense_rank, cume_dist, percent_rank, ntile
- Window Aggregate functions (moving and cumulative)
 - Avg, sum, min, max, count, variance, stddev, first_value, last_value
- LAG/LEAD functions
 - Direct inter-row reference using offsets
- Reporting Aggregate functions
 - Sum, 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

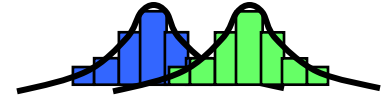
Split Lot A/B Offer testing



- Offer “A” to one population and “B” to another
- Over time period “t” calculate **median** purchase amounts of customers receiving offer A & B
- Perform **t-test** to compare
- If statistically significantly better results achieved from one offer over another, offer everyone higher performing offer



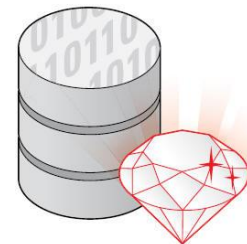
Independent Samples T-Test (Pooled Variances)



- Query compares the mean of AMOUNT_SOLD between MEN and WOMEN within CUST_INCOME_LEVEL ranges

```
SELECT substr(cust_income_level,1,22) income_level,  
       avg(decode(cust_gender, 'M', amount_sold, null)) sold_to_men,  
       avg(decode(cust_gender, 'F', amount_sold, null)) sold_to_women,  
       stats_t_test_indep(cust_gender, amount_sold, 'STATISTIC', 'F')  
       t_observed,  
       stats_t_test_indep(cust_gender, amount_sold) two_sided_p_value  
FROM sh.customers c, sh.sales s  
WHERE c.cust_id=s.cust_id  
GROUP BY rollup(cust_income_level)  
ORDER BY 1;
```

SQL Worksheet



Applications *Powered by* Oracle Data Mining

CRM OnDemand—Sales Prospector

Predictions

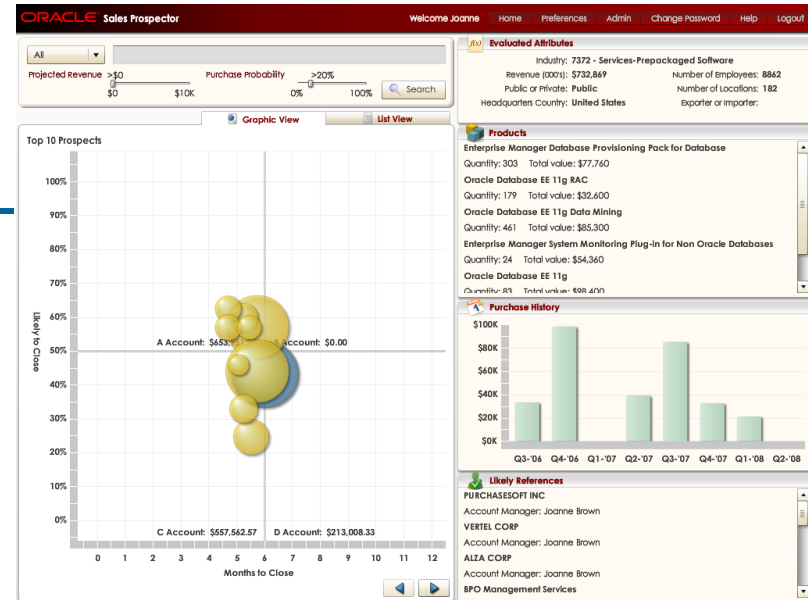
- Revenue
- Probability
- Time to close

Analysis

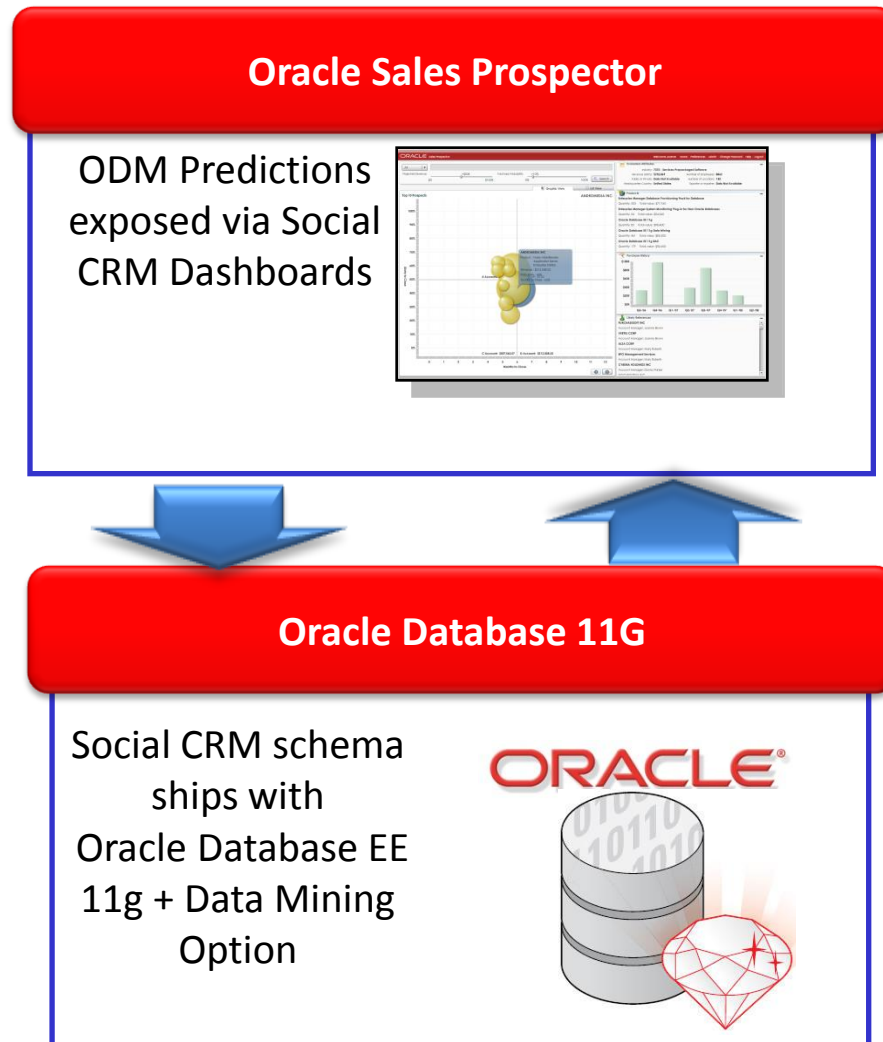
- Customer attributes
- Products owned
- Purchase history

References

- Similar customers
- Similar products



CRM OnDemand—Sales Prospector

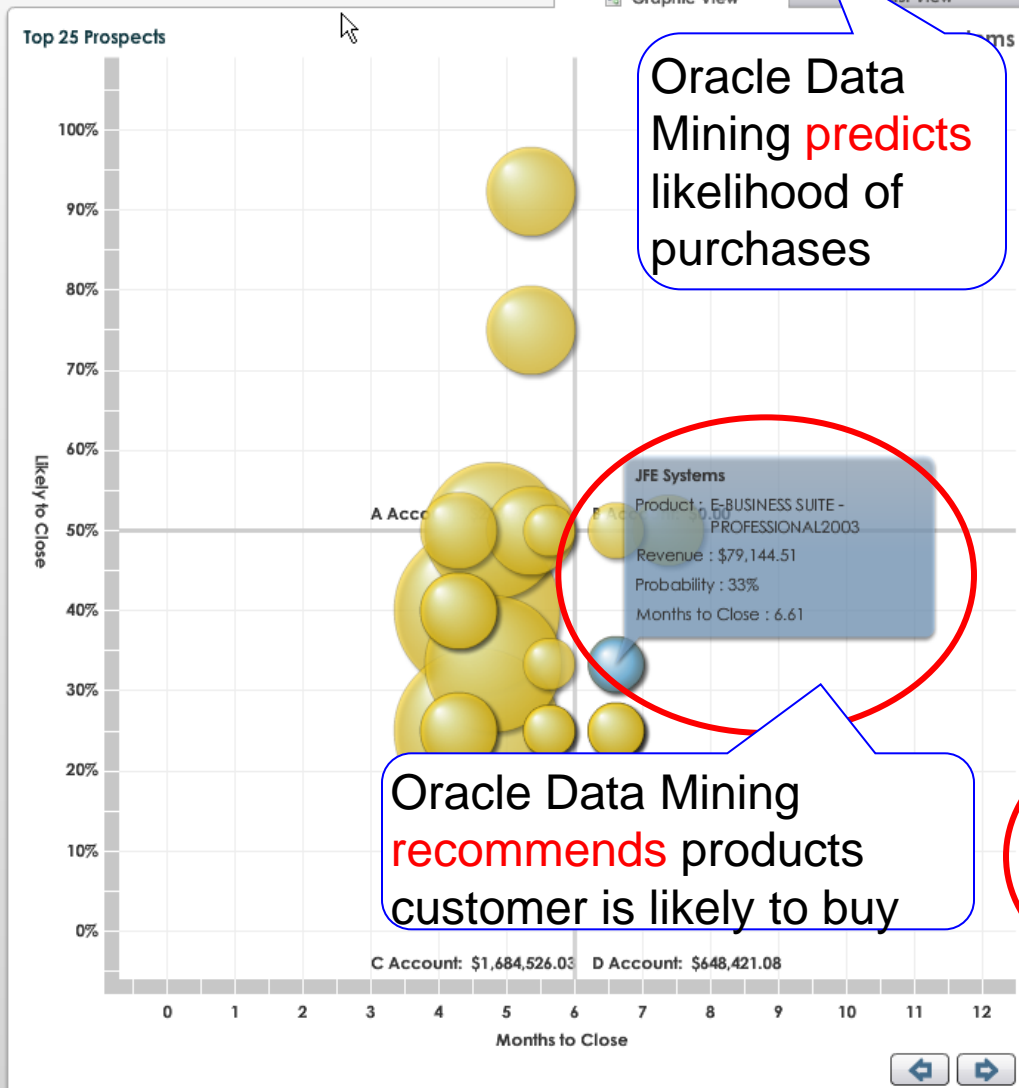


All

Projected Revenue >\$0 Purchase Probability >20%

\$0 \$1M 0% 100%

Search



Oracle Data Mining **predicts** likelihood of purchases

Oracle Data Mining **recommends** products customer is likely to buy

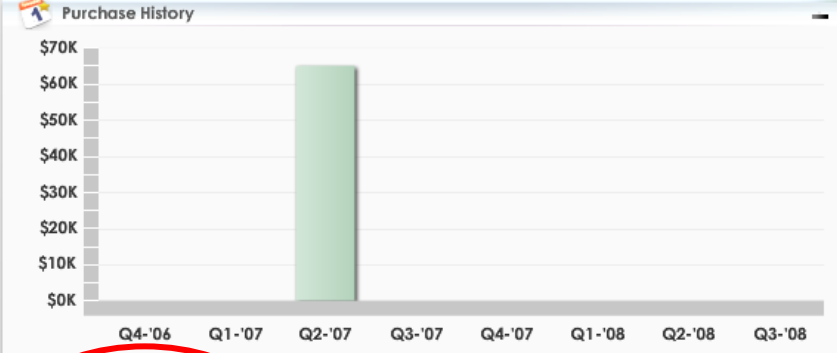
Evaluated Attributes

Industry: **Electronics**
 Revenue (000's): **\$337,420**
 Public or Private: **Private**
 Headquarters Country: **Japan**

Number of Employees: **1538**
 Number of Locations: **13**
 Exporter or Importer: **Export**

Products

91 APP SVR SE	Quantity: 3	Total value: \$0
APP SVR EE	Quantity: 1	Total value: \$0
DATABASE EE	Quantity: 3	Total value: \$0
DATABASE SE	Quantity: 3	Total value: \$3,000,000
DIAGNOSTICS PACK	Quantity: 2	Total value: \$0



Likely References

Aspen Aerogels Inc	Account Manager: Dima Skorikov
FIRST CHURCH OF CHRIST SCIENTIST	Account Manager: Dima Skorikov
Stratus Technologies Inc	Account Manager: Dima Skorikov

Oracle Data Mining **suggests** likely references

Oracle Open World (OOW) Schedule Builder

Session Recommendation Engine

- Build Personal OOW Agendas
 - Recommends sessions, exhibitors and demos based on profile
 - Identify related sessions to selected session
- Get Recommendations
- Status
 - Production use at OOW'08 and OOW'09
 - 40,000+ attendees
- Tech details
 - Solution includes in-database transformations, ODM clustering (text mining) and classification algorithms with code generation from Oracle Data Miner

The screenshot displays the Oracle Open World Schedule Builder interface. At the top, it shows the event details: "OCTOBER 11-15, 2009 MOSCONE CENTER SAN FRANCISCO" and the slogan "Come with questions. Leave with answers." Below this is a navigation bar with tabs for "Home", "Build Schedule", "Saved Schedule & Interests", "Session Changes", and "Logout".

The main content area includes a search section with "Basic Search" and "Advanced Search" options, and a "Recommended Exhibitors" list featuring companies like SAP AG, Microstrategy, and IT Convergence. A "Recommended Oracle Demos" section is also visible.

The central part of the interface is a calendar grid showing sessions from Sunday, Oct 11 to Thursday, Oct 15. The grid is organized by time slots (8:00 to 15:00). Key sessions include:

- Monday, Oct 12, 9:00:** Keynote: Oracle Develop--What Are We Still Doing Wrong?
- Monday, Oct 12, 10:00:** Keynote: Capitalizing on
- Monday, Oct 12, 13:00:** (*R) Oracle Data Mining 11g: Overview, Demos, Oracle Exadata, and
- Tuesday, Oct 13, 9:00:** Keynote: The Art of the Possible--Charles Phillips and Safra Catz, Oracle
- Tuesday, Oct 13, 10:00:** Keynote: The Future of Enterprise
- Wednesday, Oct 14, 10:00:** Keynote: Innovation Across the Stack--Thomas Kurian, Oracle
- Wednesday, Oct 14, 11:00:** Keynote: A Thorough Program Modern Customer
- Thursday, Oct 15, 9:00:** Keynote: Primavera Project Portfolio Management Road
- Thursday, Oct 15, 14:30:** Keynote: Seven

Oracle Retail Data Model

The screenshot displays the Oracle Interactive Dashboards interface. At the top, there are navigation links for 'My Dashboard', 'OLAP', 'ORDM Demo Analytics', 'ORDM Demo KPIs', and 'ORDM MetaData'. The main header includes 'Mining Analysis' and a welcome message 'Welcome, NRFdemo!'. Below this is a menu of analysis tools: Associate Basket Analysis, Associate Loss Analysis, Associate Sales Analysis, Product Category Mix, Customer Loyalty Analysis, Frequent Shopper Category Mix, Item Basket Analysis, Item POS Loss, and Store Loss. A red oval highlights this menu.

Below the menu is a filter section with dropdowns for Year (2007), Month (BY 2007 M2), Division (Pittsburgh Division), District, and Store Name (PITTSBURGH 109). A 'Performance Measure' dropdown is set to 'Customer Loyalty Type' with a 'Go' button.

A blue callout bubble on the left contains the text: "Out-of-the box, Oracle Data Mining generates profiles of customers".

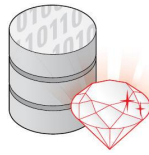
A blue callout bubble on the right contains the text: "Oracle Data Mining automatically mines data for analysis reports".

The main content area shows a table titled 'Customer Profile' with the following data:

Customer Profile	Performance Measure Value	% of Supporting Transaction	Probability
Years Of Residence = '1', '2', '3', '4', '5' and YearS_OF_RESIDENCE = '1', '2', '3' and Household Size = '3+' and YearS_OF_RESIDENCE = '1'	LEAST LOYAL	2.45%	84.76%
Years Of Residence = '1', '2', '3', '4', '5' and YearS_OF_RESIDENCE = '1', '2', '3' and Household Size = '3+' and YearS_OF_RESIDENCE = '1' and Marital Status = 'DIVORCED', 'SEPARATED'	LEAST LOYAL	0.13%	100.00%
Years Of Residence = '1', '2', '3', '4', '5' and YearS_OF_RESIDENCE = '1', '2', '3' and Household Size = '3+' and YearS_OF_RESIDENCE = '1' and Marital Status = 'MARRIED', 'SINGLE'	LEAST LOYAL	1.60%	87.85%
Years Of Residence = '1', '2', '3', '4', '5' and YearS_OF_RESIDENCE = '4', '5' and Household Size = '3+'	PRETTY LOYAL	10.93%	83.17%
Years Of Residence = '1', '2', '3', '4', '5' and YearS_OF_RESIDENCE = '4', '5' and Household Size = '3+' and Marital Status = 'MARRIED', 'SINGLE'	PRETTY LOYAL	7.31%	81.58%
Years Of Residence = '1', '2', '3', '4', '5' and YearS_OF_RESIDENCE = '4', '5' and Household Size = 'LESS THAN 3'	MARGINALLY LOYAL	9.00%	86.29%
Years Of Residence = '1', '2', '3', '4', '5' and YearS_OF_RESIDENCE = '4', '5' and Household Size = 'LESS THAN 3' and Marital Status = 'MARRIED', 'SINGLE'	MARGINALLY LOYAL	6.52%	84.17%
Years Of Residence = '10', '6', '7', '8', '9' and Household Size = '3+'	MOST LOYAL	25.01%	88.28%
Years Of Residence = '10', '6', '7', '8', '9' and Household Size = '3+' and Marital Status = 'MARRIED', 'SINGLE'	MOST LOYAL	17.11%	86.10%
Years Of Residence = '10', '6', '7', '8', '9' and Household Size = 'LESS THAN 3'	PRETTY LOYAL	26.20%	80.88%

A red oval highlights the table content.

OCDM—Pre-Built Data Mining Models



- Churn Prediction
- Customer Profiling/
Segmentation
- Customer churn
factors
- Cross-Sell
Opportunity
- Sentiment Analysis
- Life Time Value
Prediction

ORACLE Interactive Dashboards My Dashboard OCDM MetaData Welcome to OCDM

Churn By Customer Segment Welcome, ocdm! Dashboards - Answers - More Products - Settings - Log Out

Churn Prediction by Customer Segment

Customer Segment Name is equal to Segment_3

Customer Segment	Customer Name	Cell Phone No	Contract Value	Month Revenue	Debt Value	LTV Band	LTV Value	LTV Months	ARPU Band	Churn Indicator	Sentiment	Churn Probability	Customer Segment Key
	Chloe Waite	9985005370	\$0.00		\$222.00		\$65,000.00	10			▲+	56	101
	Delora Walker	9985009300	\$0.00		\$130.00		\$85,000.00	18			▲+	30	101
	Max Gerber	9985006161	\$3,000.00	\$2,500.00	\$222.00		\$79,000.00	17			▲+	39	101
	Glen Christian	9985008393	\$0.00		\$130.00		\$59,000.00	13		● Probability of Churning is very high	▼-	82	101
	Mason Murray	9985007979	\$9,000.00	\$7,500.00	\$70.00		\$56,000.00	21		● Probability of Churning is very high	▼-	94	101
	Deb Coe	9985007379	\$3,000.00	\$2,500.00	\$130.00		\$89,000.00	10			▲+	69	101
	Murray Walker	9985007504	\$0.00		\$70.00		\$80,000.00	44			▲+	36	101
	Phil Hurst	9985006286	\$0.00		\$130.00		\$54,000.00	21			▲+	32	101
	Candide Rodrick	9985009904	\$0.00		\$222.00		\$75,000.00	35			▲+	46	101
	Tiffany Hatcher	9985002670	\$0.00		\$70.00		\$35,000.00	11		● Probability of Churning is very high	▼-	74	101



Spend Classification

Classify Spend into Purchasing Categories

FEATURES

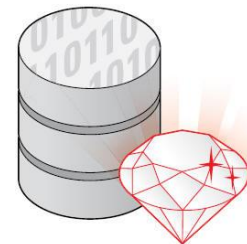
- Hierarchical classification and scoring (using Oracle Data Mining)
- Auto Spend Classification – Inline and Batch
- Assisted or manual updates online or excel interface offline
- Native integration to OBIA Procurement & Spend Analytics 7.9.6

BENEFITS

- Classifies spend data from various sources into procurement category hierarchies
- Category normalization aids strategic sourcing and contract negotiation
- Business user usability
- In-line mode integrated with EBS iProcurement

The screenshot displays the Oracle Spend Classification web interface. The top navigation bar includes 'ORACLE Interactive Dashboards' and various application areas like 'My Dashboard', 'Autonomous Communications', 'Consumer Spend', 'Control Center', 'Telephony', 'Financial Services', 'Employee', 'Human Resources', 'Loyalty', 'Marketing', 'Partner', 'Pricing', 'Procurement and Sourcing', 'Projects', 'Public Sector', 'Sales', 'Supply Sales', 'Service', 'Spend Classification', 'Supply Chain', 'Supply Chain and Order Management', and 'Usage Accelerator'. The main content area is titled 'Spend Classification' and includes a 'Classification Summary' and 'Classification Details' tab. The 'Classification Details' section shows a table with columns for 'Transaction Number', 'General Code', 'Line Description', 'Transaction Type', 'Supplier Name', 'Classification Level', 'Classification Status', and 'Amount'. Below this is an 'Auto Classification Details' section with a bar chart comparing 'LAPTOP COMPUTER' and 'HARDWARE' categories. The 'Taxonomy Details' section shows a hierarchical tree of procurement categories, and the 'Keywords' section shows a list of keywords with their weights.

Transaction Number	General Code	Line Description	Transaction Type	Supplier Name	Classification Level	Classification Status	Amount
2007099	LAPTOP-HARDWARE	Laptop 12" with 32k DVD-RW, hardisk 500 GB 5680 RPM	AP Invoice	HP Computers	2	Classified	450
K5C6074	LAPTOP-HARDWARE	General Laptop with Seagate hardisk 320 GB SATA 30000 RPM	AP Invoice	HP Computers	2	Classified	10
K5C6073	LAPTOP-HARDWARE	Laptop Lap top 15" SP5 500 GB 7200RPM 4kb DVD-RW 4GB Memory	AP Invoice	HP Computers	2	Classified	2500
K5C6072	LAPTOP-HARDWARE	Fullboxed Laptop with Seagate hardisk 320 GB SATA 30000 RPM	AP Invoice	HP Computers	2	Classified	1
K5C6089	LAPTOP-HARDWARE	Laptop 15" SP5 500 GB 7200RPM 4kb DVD-RW 4GB Memory	AP Invoice	Compaq	2	Classified	122
K5C6089	LAPTOP-HARDWARE	Full 15.5" Laptop with hardisk 320 GB SATA 30000 RPM	AP Invoice	Compaq	2	Classified	190
K5C6089	LAPTOP-HARDWARE	17" Business Computer 4GB DDR Seagate hardisk 160 GB, 32k DVD-RW	AP Invoice	Compaq	2	Classified	2
K5C6089	LAPTOP-HARDWARE	Gateway Laptop 21" with 1 TB 7200RPM, 8GB Mem, 32k DVD-RW	AP Invoice	IBM Corp	2	Classified	7
K5C6082	LAPTOP-HARDWARE	Personal laptop with SATA hardisk 500 GB Seagate 30000 RPM	AP Invoice	IBM Corp	2	Classified	5000



Exadata V2 + Oracle Data Mining 11gR2

Exadata V2 + Oracle Data Mining 11gR2

“DM Scoring” Pushed to Storage!



- In 11gR2, SQL predicates and Oracle Data Mining models are pushed to storage level for execution

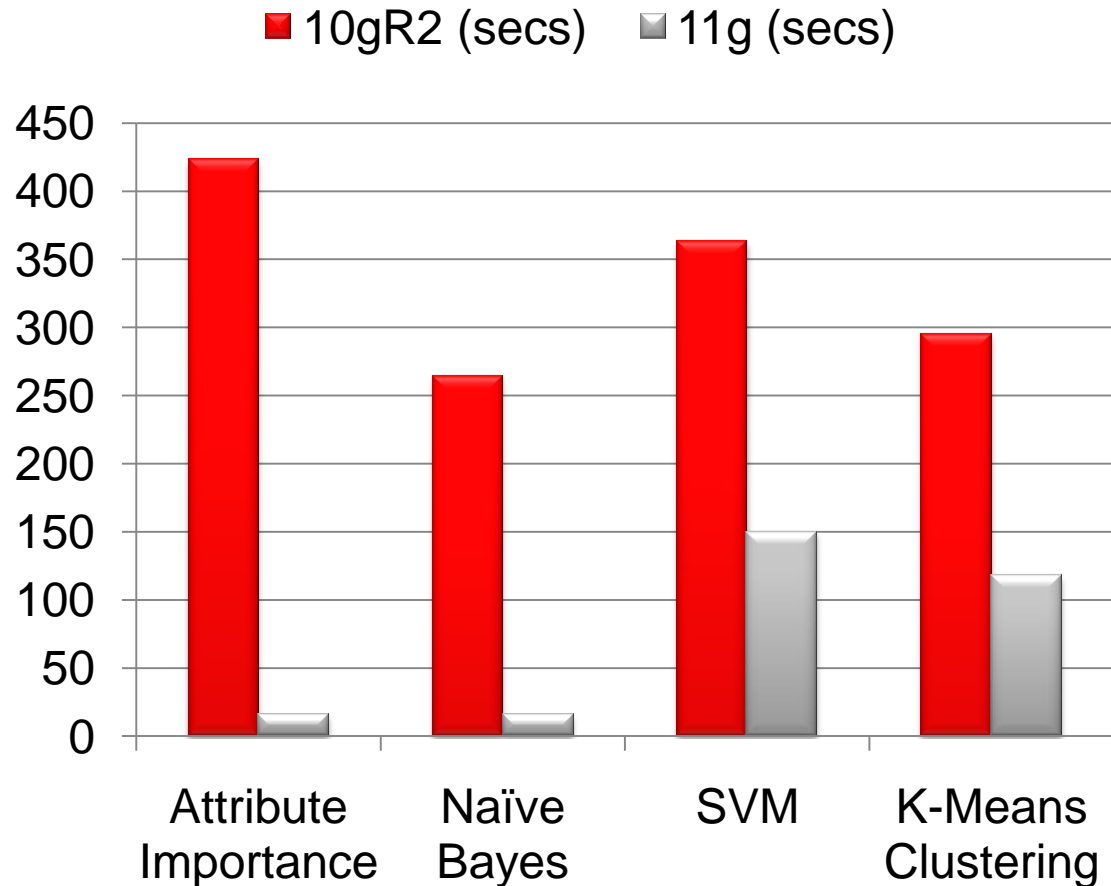
For example, find the US customers likely to churn:

```
select cust_id
from customers
where region = 'US'
and prediction_probability(churnmod, 'Y' using *) > 0.8;
```

Scoring function executed in Exadata

Model Build Performance Improvement

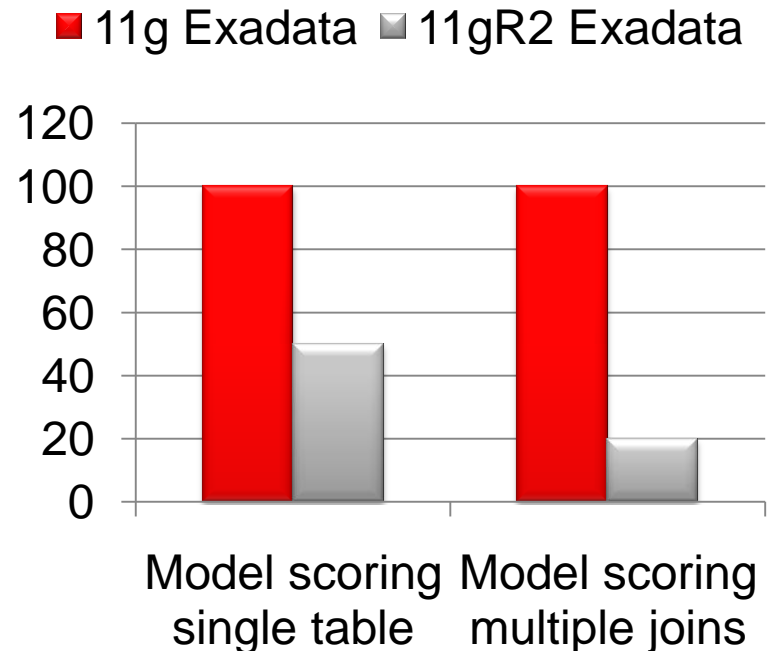
- ODM *model building* phase reduced by factors of 2-26X times faster in 11g



Source: Performance Improvement of Model Building in Oracle 11.1 Data Mining, An Oracle White Paper, May 2008

Exadata Smart Scan Model Scoring

- ODM model scoring 2-5X+ times faster on Exadata
 - Results achieved depend on the number of joins performed to assemble the data that will be “scored” with the ODM prediction mining function

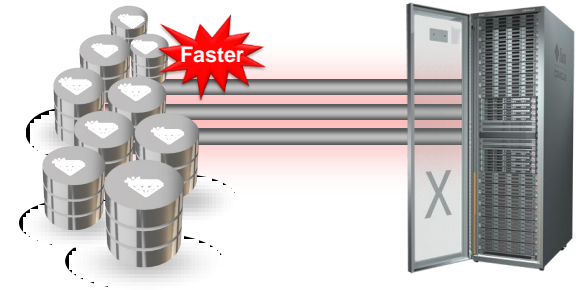


Conceptual example: depends on size of data, algorithm and number of joins

Exadata V2 + Oracle Data Mining 11gR2

Benefits

- Eliminates data movement
 - 2X-5X+ faster scoring on Exadata
 - Depends on number of joins involved with data for scoring
- Preserves security
- Significant architecture and performance advantages over SAS Institute
 - Years ahead of SAS's road map to move SAS analytics towards RDBMSs (<http://support.sas.com/resources/papers/InDatabase07.pdf>)
- Netezza performance but using industry standard RDBMS + SQL-based in-database advanced analytics
- Best platform for building enterprise predictive analytics applications e.g. Fusion Applications →
“Analytical iPod for the Enterprise”





Getting Started

Data Mining Projects

- “The vast majority of BI professionals are excited about the prospects of data mining, but are fully mystified about where to begin or even how to prepare”
- “Of those who did initiate a modeling initiative, ...51% of data mining projects either never left the ground, did not realize value or the ultimate results were not measurable”
- “In most cases, those who attempted an implementation ended up building excellent predictive models that answer the wrong questions”
- “For any organization with annual revenues more than \$50 million, employing data mining technology is not a matter of whether, but when”



<http://www.the-modeling-agency.com>



ORACLE

Getting Started with Oracle Data Mining

- You can download a **free evaluation copy** of Oracle Data Mining and try it out on your own computer. See the [Oracle Data Mining Administrators Guide](#), which tells how to install a database and set up a user account. Download the Oracle Database Enterprise Edition (10gR2 or 11g) from the [Oracle Technology Network](#). The Oracle Data Mining Option is installed by default with Oracle Database EE. For data analysts or those new to data mining, you will also want to download and install [Oracle Data Miner](#), the free, optional graphical user interface. A summary of algorithms supported by ODM with links to the documentation is posted [here](#).
- To get started quickly, Part I of [ODM Concepts](#) introduces you to the features and terminology of Oracle Data Mining. Then, use the [Oracle Data Mining Tutorial](#) to provide step-by-step guidance for using the Oracle Data Miner graphical interface. ... You can use the Oracle Data Miner (*Data --> Import...*) to import your own data in .csv text files and begin mining.
- For application developers, the [ODM Application Developer's Guide](#) along with the Oracle Data Mining sample programs gets you started writing SQL- or Java-based data mining applications.
- Some additional datasets for learning Oracle Data Mining include:
CUST_INSUR_LTV (dmp file), CD_BUYERS (dmp file), EMPL_DATA (dmp file), LYMPHOMA (dmp file)
- Application developers can integrate predictive analytics into any report or enterprise application using ODM's server-based PL/SQL or Java APIs. See [ODM Sample Programs](#) for demo sample code.
- **Oracle Data Mining Education through Oracle University**
 - [Installing Data Miner](#) (Oracle By Example)
 - [Solving Business Problems with Data Mining](#) (Oracle By Example)





ORACLE®