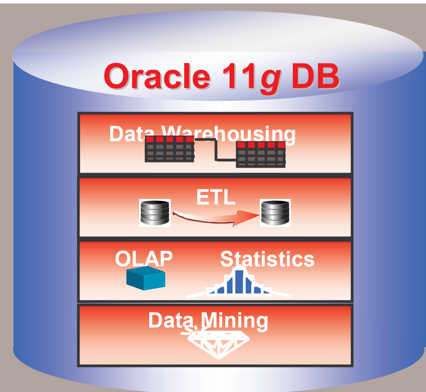
ORACLE OPEN WORLD

experience OPENWORLD

November 11–15, 2007

ORACLE



Predictive Analytics with Oracle Data Mining & Oracle BI EE

Charlie Berger
Sr. Director Product Management,
Data Mining Technologies & Life Sciences & Healthcare Industries
Oracle Corporation
charlie.berger@oracle.com

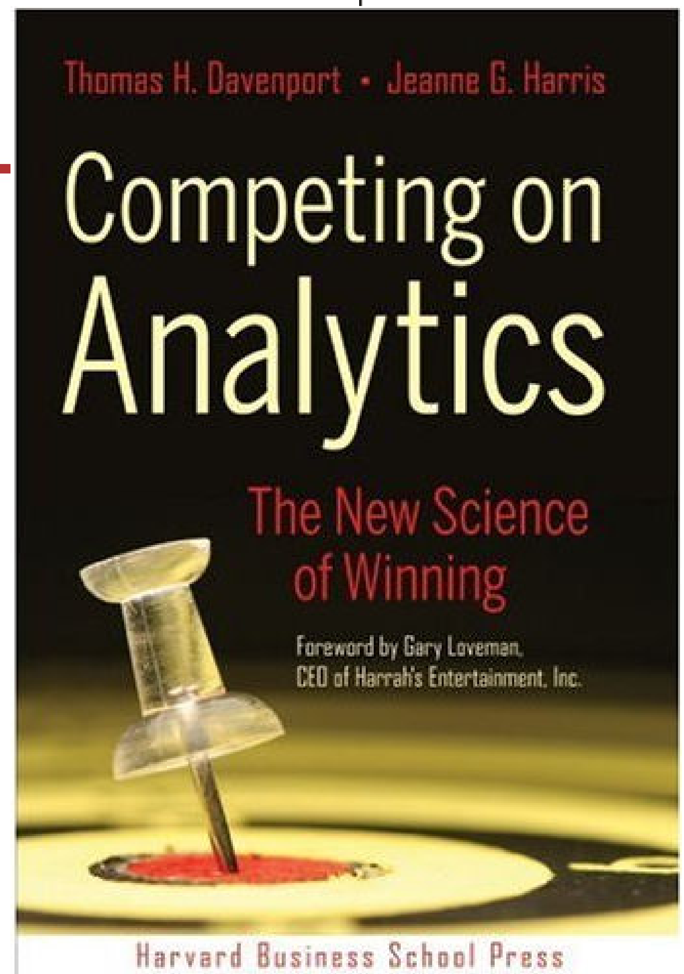
Harvard Business Review

www.hbr.org

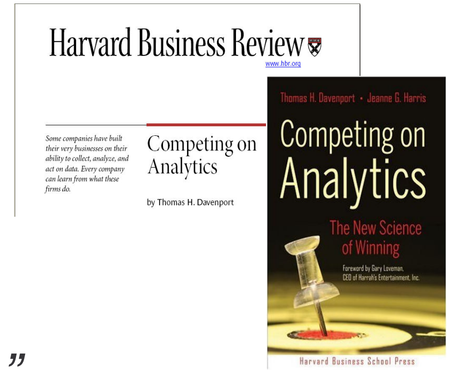
Some companies have built their very businesses on their ability to collect, analyze, and act on data. Every company can learn from what these firms do.

Competing on Analytics

by Thomas H. Davenport

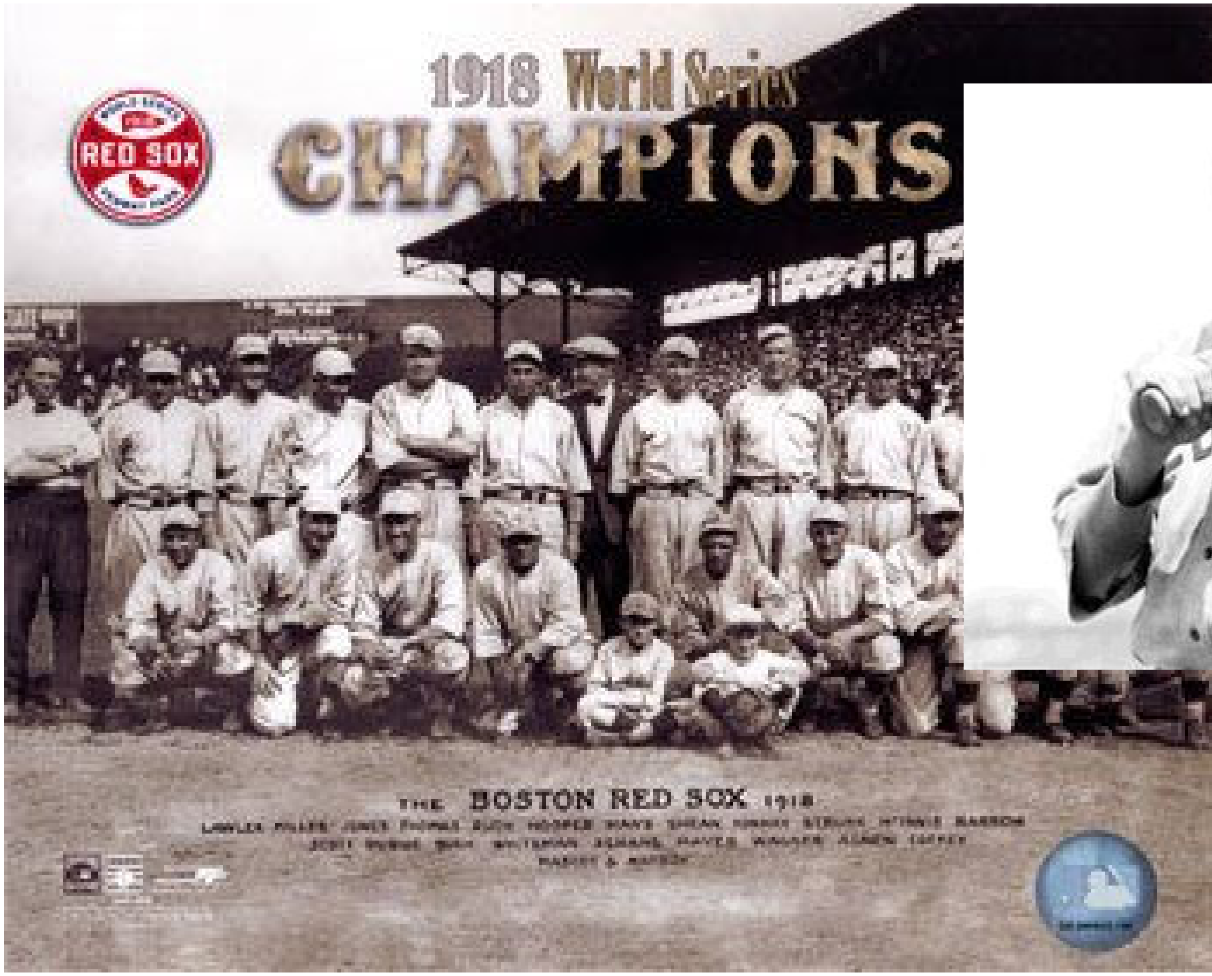


Competing on Analytics

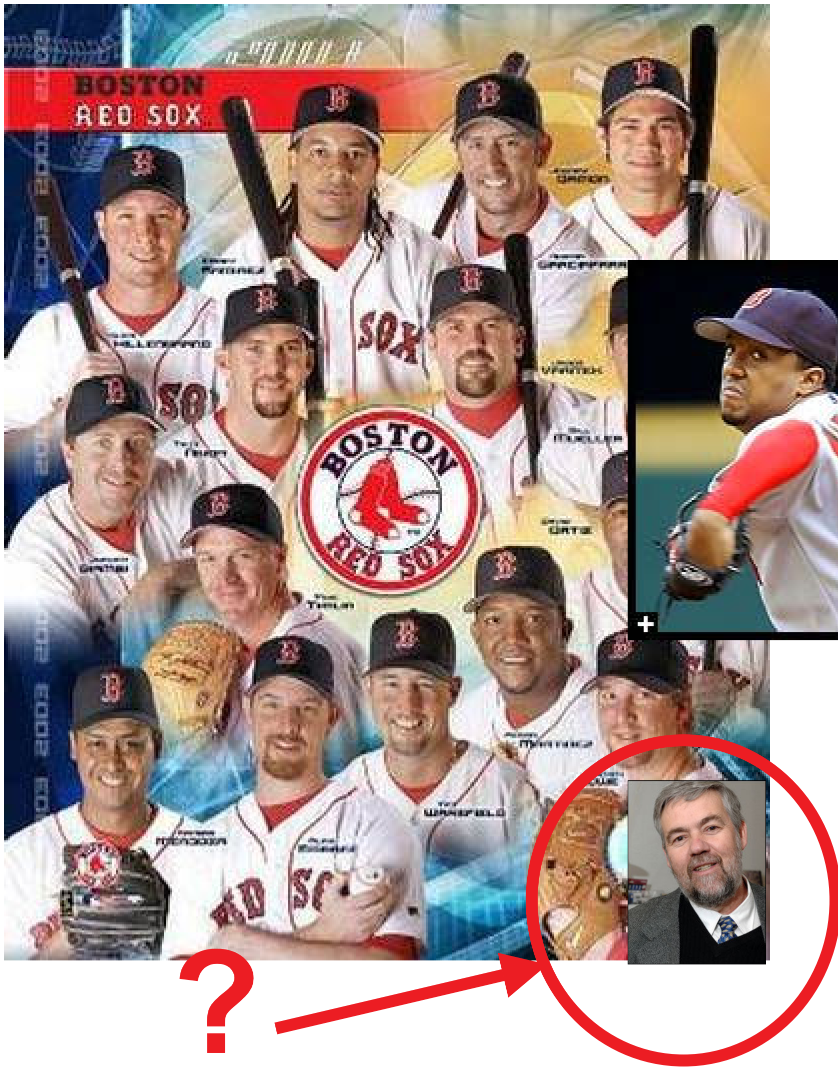


- *“Some companies have built their very businesses on their ability to collect, analyze, and act on data. Every company can learn from what these firms do.”*
  Competing on Analytics, by Thomas H. Davenport
- *“Although numerous organizations are embracing analytics, only a handful have achieved this level of proficiency. But analytics competitors are the leaders in their varied fields—consumer products, finance, retail, and travel and entertainment among them.” Competing on Analytics, by Thomas H. Davenport*
- Business intelligence cited as the top technology priority for CIOs in 2006 (surpassing security) - Gartner 2006
- Sizable market - 11.5% growth in 2005 for a market size of $5.7 billion in worldwide software revenue - IDC 2006
- “Organizations are moving beyond query and reporting” - IDC 2006

Boston Red Sox
1918 World Series Champions



2003 Red Sox AL Championship Series



October 16, 2003: Holding a 5-2 lead in the eighth inning of Game 7 at Yankee Stadium, Red Sox manager Grady Little elects to leave starter Pedro Martinez on the mound. Martinez proceeds to give up four hits and three runs in the inning, allowing the Yankees to tie the game. In the bottom of the eleventh inning, leadoff hitter Aaron Boone hits a solo home run off of Tim Wakefield to left field, ending the game and the series, giving the Yankees their 39th American League pennant.

Florida Marlins Win 2003 World Series

Josh Beckett pitches a complete game shut out

In the World Series, the Marlins put up their young roster with a \$54 million payroll up against the storied Yankees and their \$164 million payroll.

The Data Analyst

??

- George William “Bill” James is a baseball writer, historian, and statistician whose work has been widely influential. … His approach, … scientifically analyzes and studies baseball, often through the use of statistical data….
- In 2006, Time magazine named him as one of the 100 Most Influential People in the World
- In 2003, James was hired by former reader, John Henry, new owner of the Boston Red Sox
  - Innovations e.g.:
    - Runs Created. A statistic intended to quantify a player's contribution to runs scored, as well as a team's expected number of runs scored. Runs created is calculated from other offensive statistics. James' first version of it:
      Runs Created = ((Total Bases * (Hits + Walks))/(Plate Appearances).

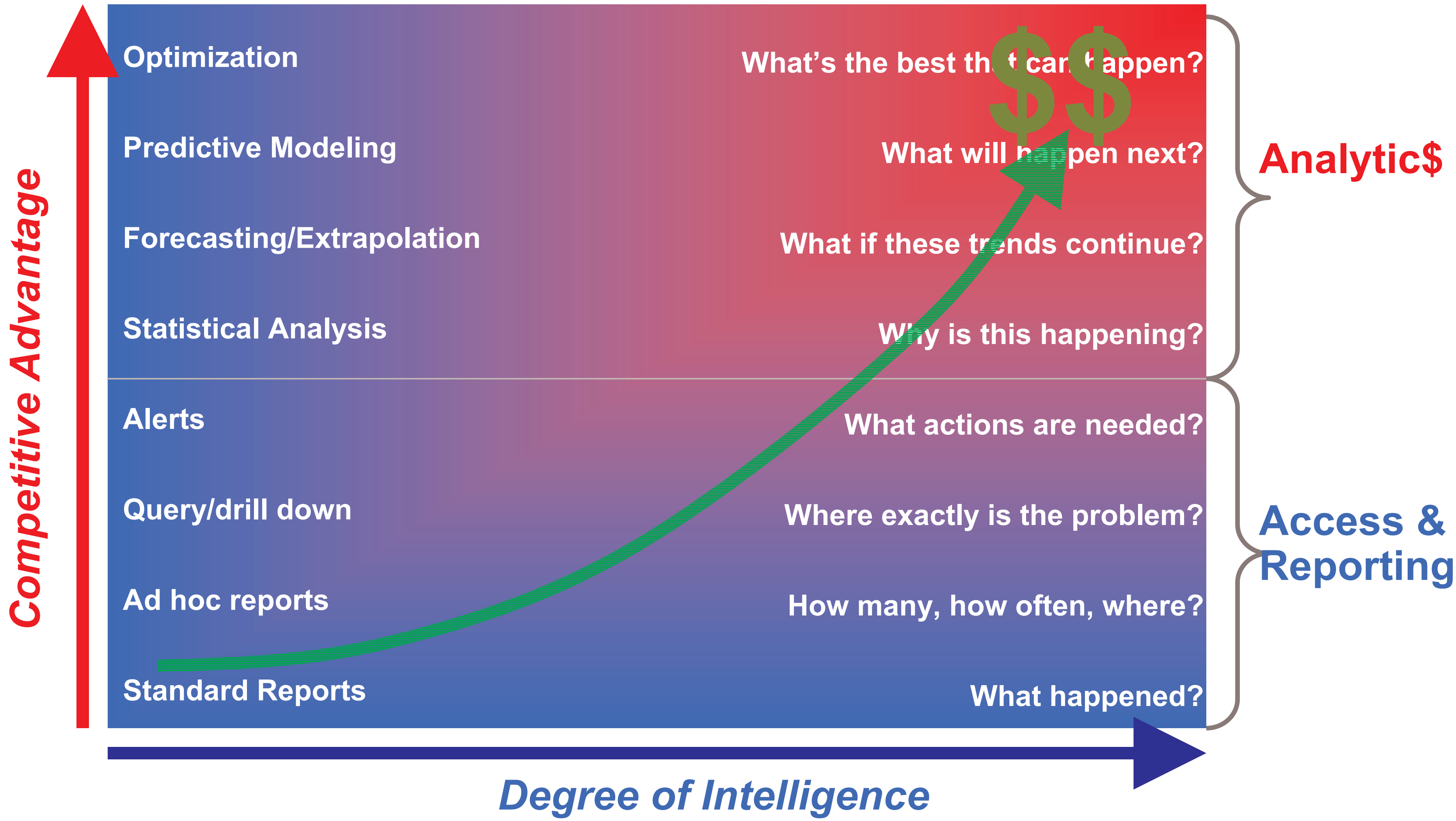
2004 World Series Champions!!!

Talent + Experience + Analytics = Red Sox are 2004 World Champions!

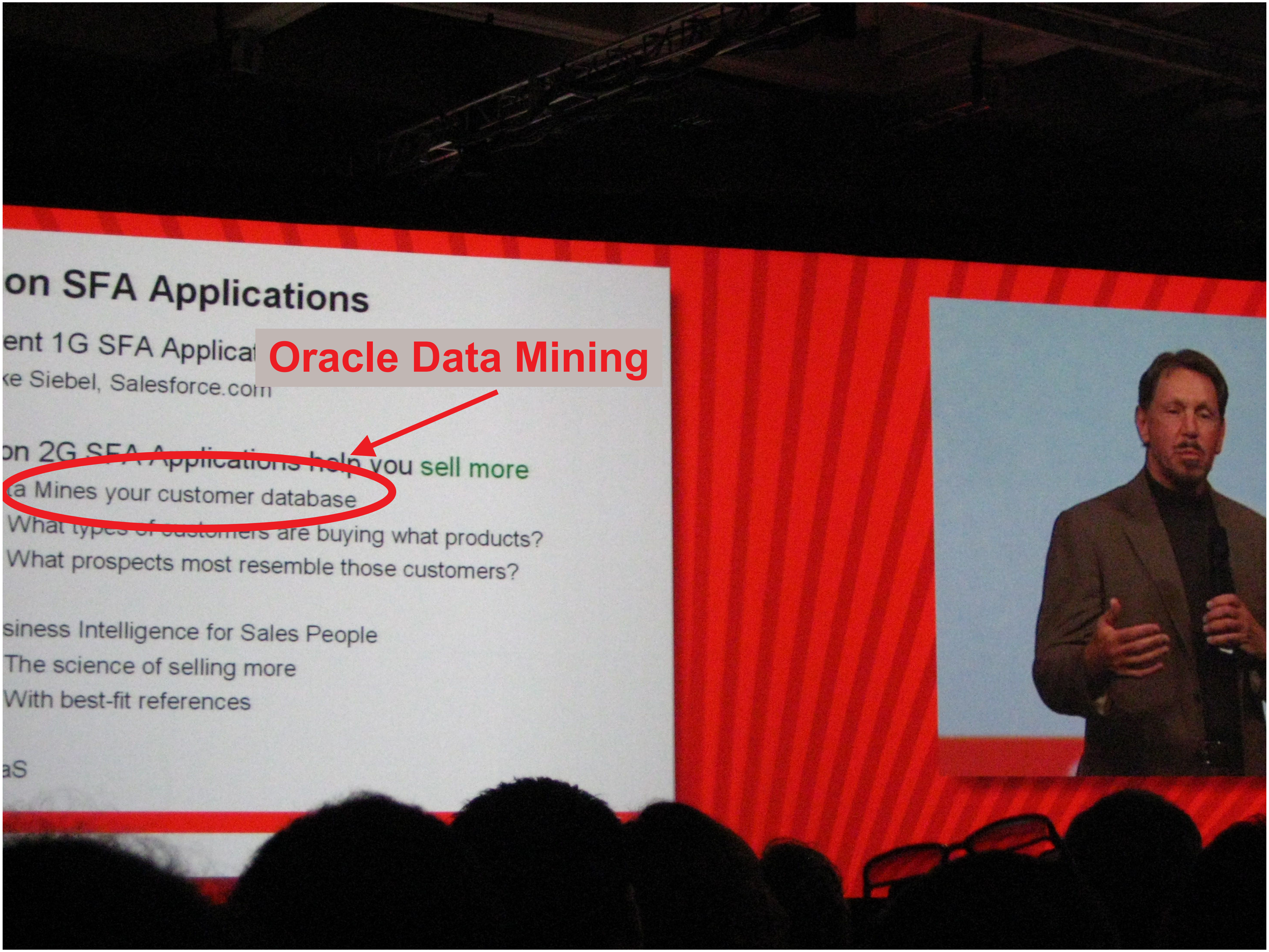
(1st time in 86 years!)

+

Competitive Advantage of BI & Analytics

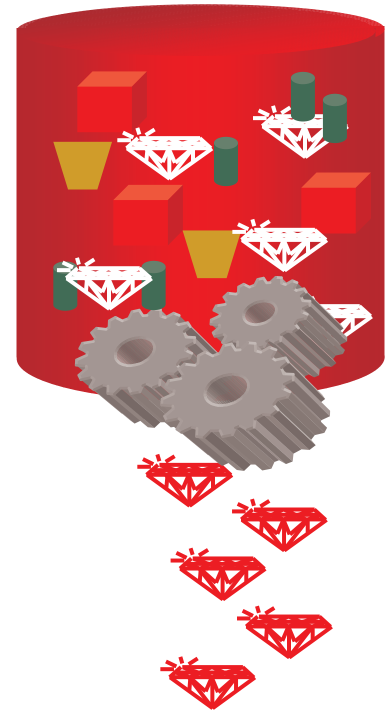


Source: Competing on Analytics, by T. Davenport & J. Harris



on SFA Applications
ent 1G SFA Applicat
ke Siebel, Salesforce.com
Oracle Data Mining
on 2G SFA Applications help you sell more
a Mines your customer database
What types of customers are buying what products?
What prospects most resemble those customers?
siness Intelligence for Sales People
The science of selling more
With best-fit references
aS

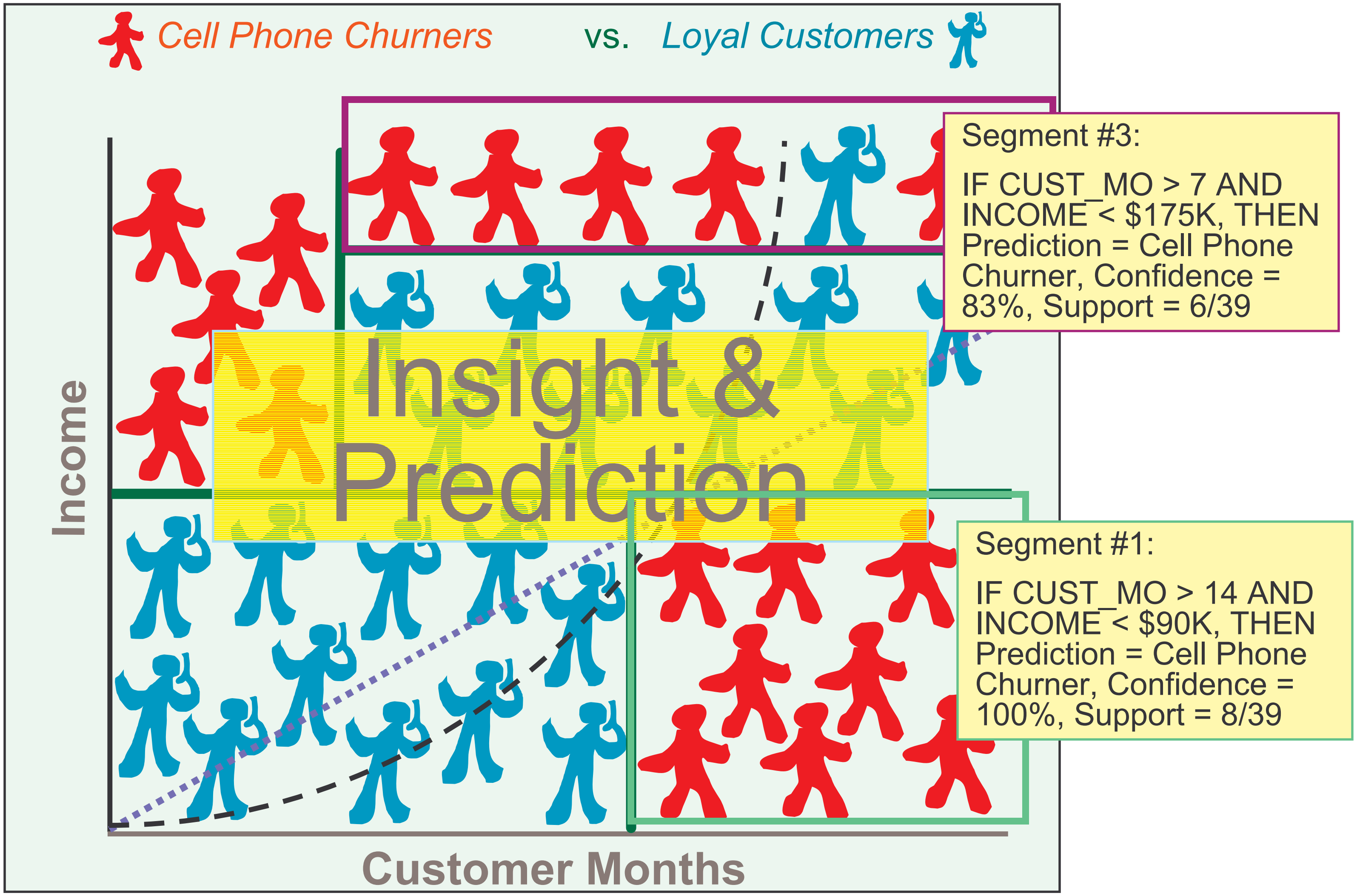
What is Data Mining?



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

Data Mining Provides

Better Information, Valuable Insights and Predictions



Oracle Data Mining

Overview (Classification)

Input Attributes

Target

Model

Cases

Name	Income	Age			Buy Product? 1 =Yes, 0 =No		
Historic Data							
Jones	30,000	30			1		
Smith	55,000	67			1		
Lee	25,000	23			0		
Rogers	50,000	44			0		
New Data							
Campos	40,500	52			?	1	.85
Horn	37,000	73			?	0	.74
Habers	57,200	32			?	0	.93
Berger	95,600	34			?	1	.65

Functional Relationship:

$Y = F(X_1, X_2, \ldots, X_m)$

ORACLE DATABASE 11g

Prediction Confidence

Business Intelligence & Analytics

ORACLE DATABASE 11g

Query and Reporting	OLAP	Data Mining
Extraction of detailed and roll up data	Summaries, trends and forecasts	Knowledge discovery of hidden patterns
"Information"	*"Analysis"*	*"Insight & Prediction"*
Who purchased mutual funds in the last 3 years?	**What is the** average **income** of mutual fund buyers, by region, by year?	**Who will buy** a mutual fund in the next 6 months and why?

Example Data Mining Applications

Financial Services
- Combat attrition (churn)
- Fraud detection
- Loan default (Basel II)
- Identify selling opportunities

Database Marketing
- Buy product x
- More targeted & successful campaigns
- Identify cross-sell & up-sell opportunities

Telecommunications
- Identify customers likely to leave Target highest lifetime value customers
- Identify cross-sell opportunities

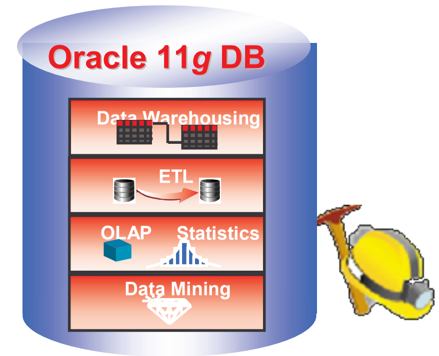
Insurance, Government
- Flag accounting anomalies (Sarbanes-Oxley)
- Reduce cost of investigating suspicious activity or false claims

Retail
- Loyalty programs
- Cross-sell
- Market-basket analysis
- Fraud detection

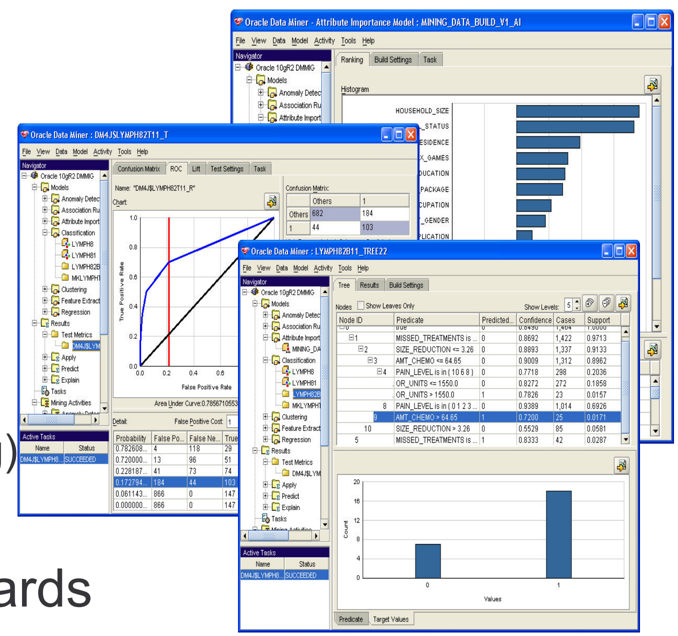
Life Sciences
- Find factors associated with healthy or unhealthy patients
- Discover gene and protein targets
- Identify leads for new drugs

Oracle Data Mining 11g

Oracle in-Database Mining Engine



- Data Mining Functions (Server)
  - PL/SQL & Java APIs
  - Develop & deploy predictive analytics applications
- Wide range of DM algorithms
  - Classification & regression
  - Clustering
  - Anomaly detection
  - Attribute importance
  - Feature extraction (NMF)
  - Association rules (Market Basket analysis)
  - Structured & unstructured data (text mining)
- Oracle Data Miner (GUI)
  - Simplified, guided data mining using wizards
- Predictive Analytics
  - *"1-click data mining"* from a spreadsheet



11g Statistics & SQL Analytics

FREE (Included in Oracle SE & EE)

- Ranking functions
  - rank, dense_rank, cum_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.

Note: Statistics and SQL Analytics are included in Oracle Database Standard Edition

- Descriptive Statistics
  - average, standard deviation, variance, min, max, median (via percentile_count), mode, group-by & roll-up
  - DBMS_STAT_FUNCS: summarizes numerical columns of a table and returns count, min, max, range, mean, stats_mode, variance, standard deviation, median, 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 Mining

Know More, Do More, Spend Less

Business Decision Makers

- Make Better Decisions
- Extract More Value from Your Data
- Lower Your Total Cost of Ownership

Data Analysts

- Get Results Faster
- Get More Results
- Easy to Use

Integrators and IT

- Create More Value for Your Organization
- Make Your Work Easier
- Transform IT from a Cost to a Profit Center

Predictive Analytics Use Case

- The cast:
  - Peter: a data mining analyst
  - Sally: a marketing manager
- Peter builds a decision tree classification model, tree_model
- Peter grants the ability to view/score the tree model to Sally

```
GRANT SELECT MODEL ON tree_model TO Sally;
```

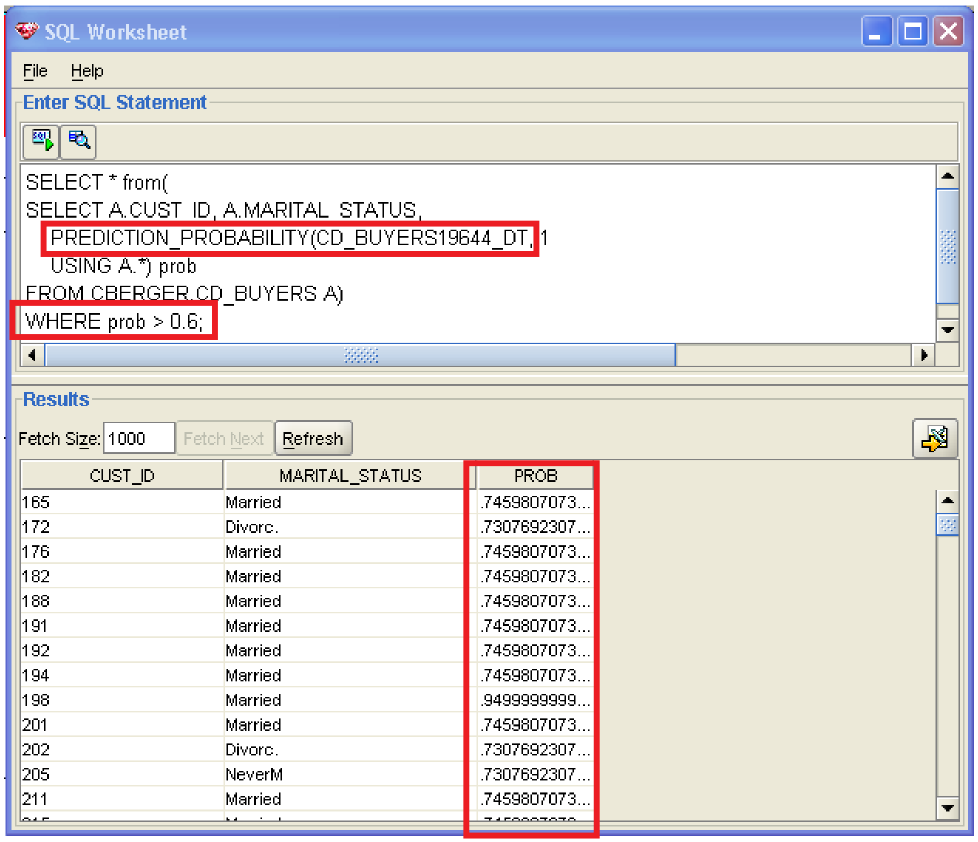
- Sally inspects the model, likes it, and wants it deployed
- Sally scores the customer database using the new model and his understanding of the cost of contacting a customer and sends the new contact list to the head of the sales department

```
CREATE TABLE AS SELECT cust_name, cust_phone FROM
customers
WHERE prediction(Peter.tree_model cost matrix (0,5,1,0) using *) =
‘responder’;
```

Example

Simple, Predictive SQL

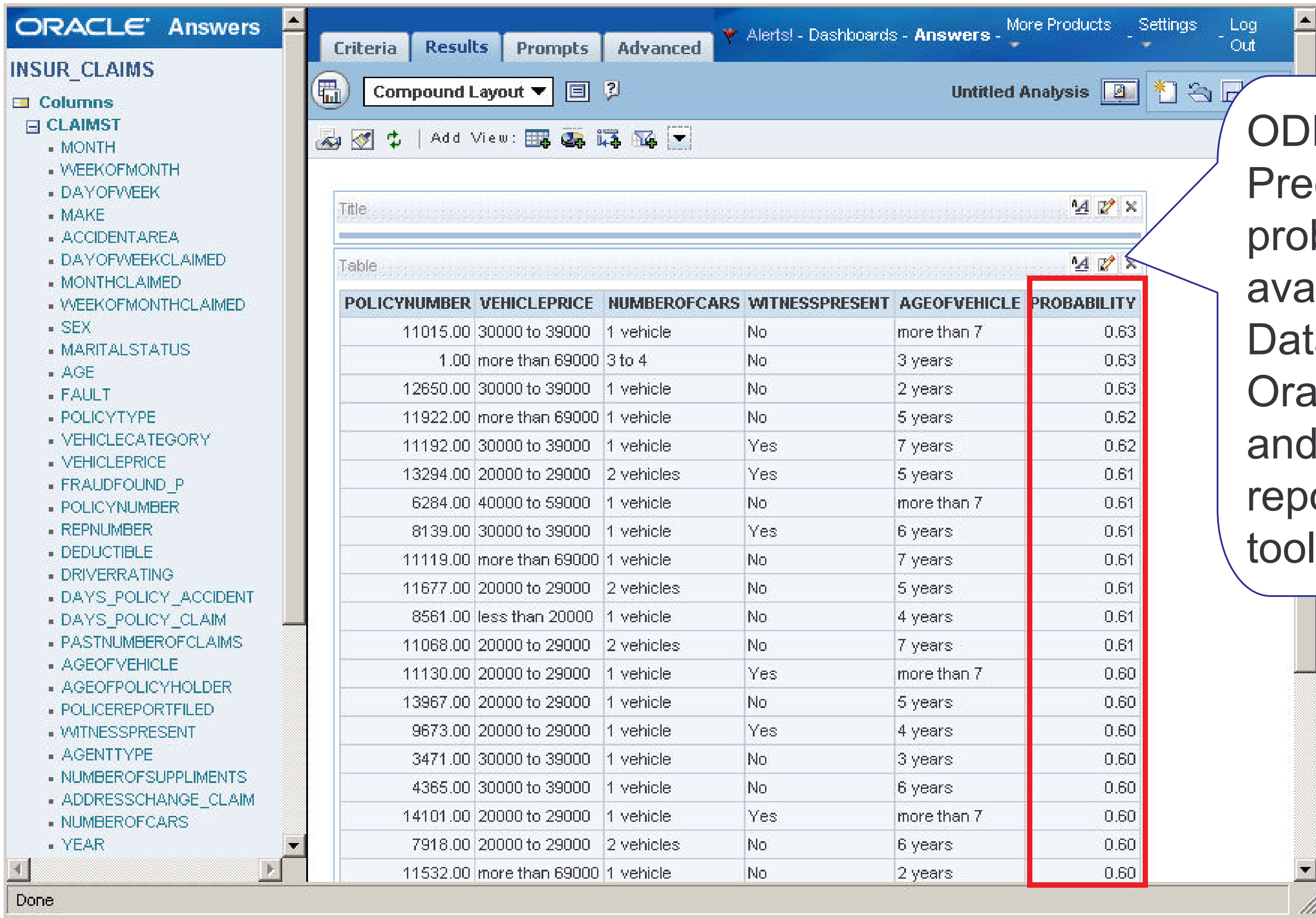
- Select customers who are more than 60% likely to purchase a product and display their marital status




```
SELECT * from(
SELECT A.CUST_ID, A.MARITAL_STATUS,
PREDICTION_PROBABILITY(CD_BUYERS77
117_DT, 1  USING A.*) prob
 FROM CBERGER.CD_BUYERS A)
 WHERE prob > 0.6;
```

Example

Better Information for OBI EE Reports and Dashboards

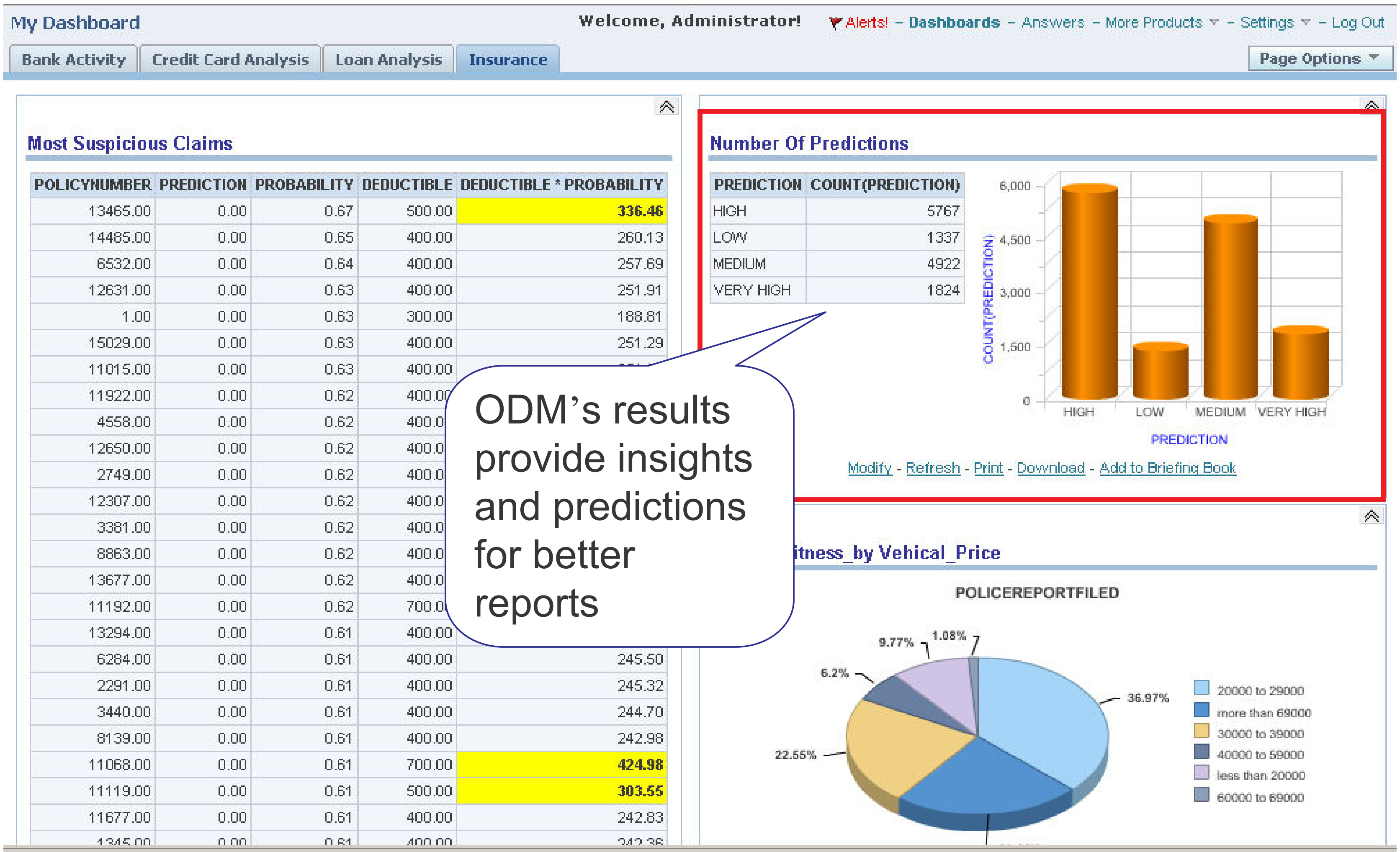


POLICYNUMBER	VEHICLEPRICE	NUMBEROFCARS	WITNESSPRESENT	AGEOFVEHICLE	PROBABILITY
11015.00	30000 to 39000	1 vehicle	No	more than 7	0.63
1.00	more than 69000	3 to 4	No	3 years	0.63
12650.00	30000 to 39000	1 vehicle	No	2 years	0.63
11922.00	more than 69000	1 vehicle	No	5 years	0.62
11192.00	30000 to 39000	1 vehicle	Yes	7 years	0.62
13294.00	20000 to 29000	2 vehicles	Yes	5 years	0.61
6284.00	40000 to 59000	1 vehicle	No	more than 7	0.61
8139.00	30000 to 39000	1 vehicle	Yes	6 years	0.61
11119.00	more than 69000	1 vehicle	No	7 years	0.61
11677.00	20000 to 29000	2 vehicles	No	5 years	0.61
8561.00	less than 20000	1 vehicle	No	4 years	0.61
11068.00	20000 to 29000	2 vehicles	No	7 years	0.61
11130.00	20000 to 29000	1 vehicle	Yes	more than 7	0.60
13967.00	20000 to 29000	1 vehicle	No	5 years	0.60
9673.00	20000 to 29000	1 vehicle	Yes	4 years	0.60
3471.00	30000 to 39000	1 vehicle	No	3 years	0.60
4365.00	30000 to 39000	1 vehicle	No	6 years	0.60
14101.00	20000 to 29000	1 vehicle	Yes	more than 7	0.60
7918.00	20000 to 29000	2 vehicles	No	6 years	0.60
11532.00	more than 69000	1 vehicle	No	2 years	0.60

ODM's Prediction probabilities available in Database for Oracle BI EE and other reporting tools

Example

Better Information for OBI EE Reports and Dashboards



Integration with Oracle BI EE

http://localhost/analytics/saw.dll?Dashboard

ORACLE BUSINESS INTELLIGENCE | *Intelligence Dashboards* | My Dashboard

My Dashboard

Welcome, John Smith! Dashboards - Answers - Advanced Report

Organization Analysis | Category Analysis

ODM provides likelihood of expense reporting fraud ….and other important questions.

Pick Any Time

Go

Potential Fraud by Organization

Org Level 2	Potential Fraud Cost
Midwest Region	2,896
Northeast Region	17,307
Southern Region	5,086
Western Region	11,252

Modify - Download

Most Suspicious Activities

Employee	Item	Day	Amount	Probabilit	Potential Fraud Cost
Louis Nagode	Misc. Employee Expenses	31-Dec-2003	15,740	59	9,265
Paul Laker	Misc. Employee Expenses	17-Dec-2003	4,996	56	2,792
Louis Nagode	Misc. Employee Expenses	30-Dec-2003	4,259	60	2,537
Dave Lindquist	Misc. Employee Expenses	01-Jan-2004	2,253	63	1,422
Steven Daniel	Hotel-Lodging	14-Dec-2004	2,304	52	1,205
Paul Laker	Hotel-Lodging	19-Dec-2004	2,219	54	1,201
Steven Daniel	Hotel-Lodging	22-Dec-2004	1,896	52	979
Christina Donohue	Hotel-Lodging	21-Dec-2004	1,744	53	919
Michael Cheng	Hotel-Lodging	21-Dec-2004	1,598	53	842
Dennis Haas	Hotel-Lodging	14-Dec-2004	1,539	52	805

Results

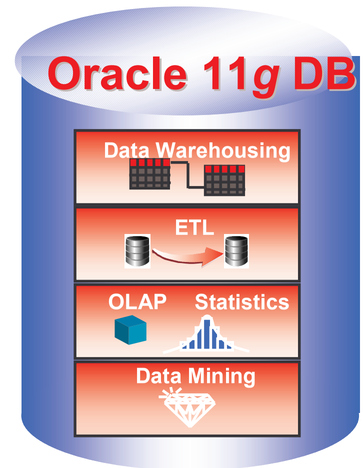
Modify - Download

Trends by Organization

Org Level 2	Quarter	Amount	Amt Parent Share	Amt % Chg Prior Per	Potential Fraud Cost	PFC Parent Share	PFC % Chg Prior Per
Midwest Region	Q1 2004	12,173	14	-9	1,056	17	37
	Q2 2004	6,528	14	-46	0		-100
	Q3 2004	7,427	13	14	0	0	
	Q4 2004	15,642	8	111	1,841	6	
Northeast Region	Q1 2004	28,182	32	-11	537	9	-75
	Q2 2004	14,986	33	-47	0		-100
	Q3 2004	21,287	38	42	742	100	
	Q4 2004	95,027	49	346	16,028	54	2,061
Southern Region	Q1 2004	15,773	18	13	2,003	33	155
	Q2 2004	7,458	16	-53	0		-100
	Q3 2004	8,674	15	16	0	0	
	Q4 2004	27,154	14	213	3,083	10	
Western Region	Q1 2004	30,909	36	-45	2,487	41	-84
	Q2 2004	16,883	37	-45	0		-100
	Q3 2004	18,612	33	10	0	0	

In-Database Data Mining

Advantages



- ODM architecture provides greater performance, scalability, and data security
- Data remains in the database at all times…with appropriate access security control mechanisms—fewer moving parts
- Straightforward inclusion within interesting and arbitrarily complex queries
- Real-world scalability—available for mission critical appls
- Enables pipelining of results without costly materialization
- Performant and scalable:
  - Fast scoring:
    - 2.5 million records scored in 6 seconds on a single CPU system
  - Real-time scoring:
    - 100 models on a single CPU: 0.085 seconds

Industry Analysts

PREDICTIVE ANALYTICS: Extending the Value of Your Data Warehousing Investment, By Wayne W. Eckerson

"…According to our survey, most organizations plan to significantly increase the analytic processing within a data warehouse database in the next three years, particularly for model building and scoring, which show 88% climbs. The amount of data preparation done in databases will only climb 36% in that time, but it will be done by almost two-thirds of all organizations (60%)—double the rate of companies planning to use the database to create or score analytical models."

"…it's surprising that about one-third of organizations plan to build analytical models in databases within three years."

"'We leverage the data warehouse database when possible,' says one analytics manager. He says most analysts download a data sample to their desktop and then upload it to the data warehouse once it's completed. 'Ultimately, however, everything will run in the data warehouse,' the manager says."

http://download.101com.com/pub/tdwi/Files/PA_Report_Q107_F.pdf

Industry Analysts

Oracle data mining: not only good but affordable too!

Published: 30th November, 2006, Bloor Research 2007

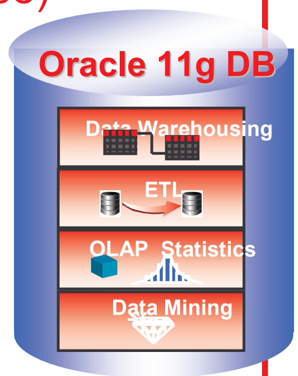
"…the Oracle data mining option is one of the great bargains available today because it is affordable and … is a real Rolls Royce of capability and features. …redesigned from scratch and put the algorithms into the database to ensure that, not only is the execution of the algorithms efficient, but the vast amounts of data handling that typifies traditional datamining is minimised. … Oracle are leaving the database in situ and mining it there, which saves a lot of effort and will greatly increase productivity.

This is a fully featured, highly sophisticated data mining capability to enable professionals to operate against Oracle data sets with productivity and precision. Oracle data mining has a broad range of available algorithms, which enable it to undertake virtually every kind of business and scientific analysis that one can think of. …

…Oracle are giving the data mining professional a real alternative to SAS and SPSS with an offering that is equally as well featured, but which promises to outperform any standalone offering.

ORACLE Analytics vs. SAS

1. **In-Database** Analytics Engine
   - Basic Statistics (Free)
   - Data Mining
   - Text Mining

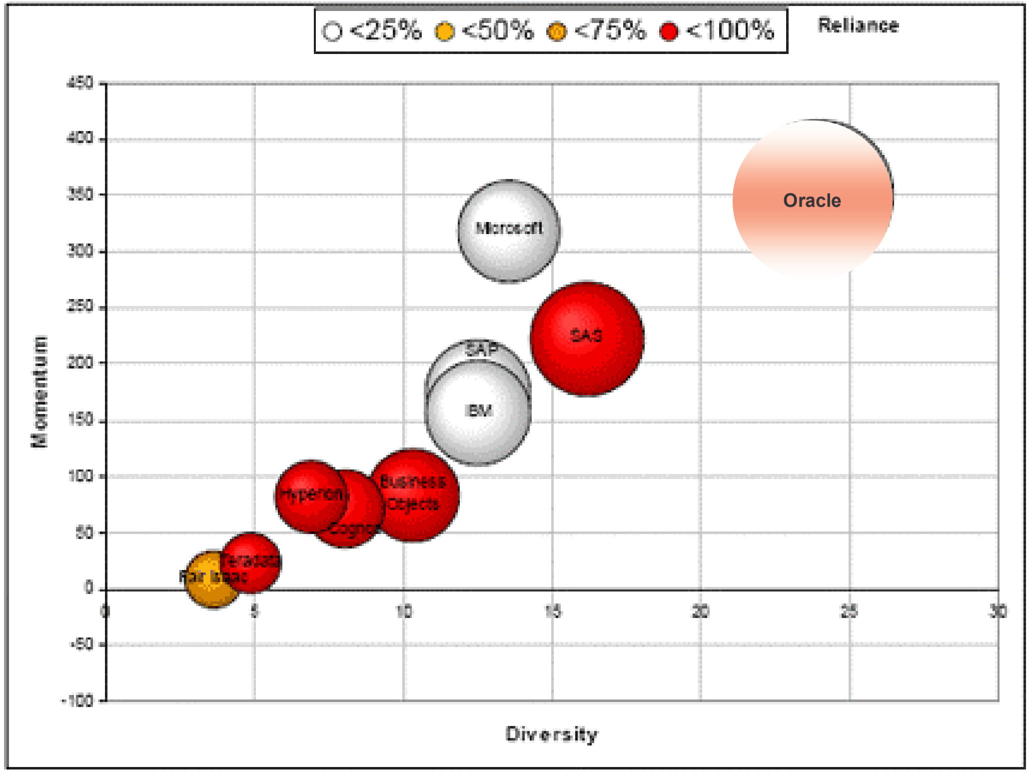


2. Standards Based Development Platform
   - Java (standard)
   - SQL (standard)
   - J2EE (standard)
3. **Costs** (ODM: $20K cpu)
   - Simplified environment
   - Single server
   - Security

1. **External** Analytical Engine
   - Basic Statistics
   - Data Mining
   - Text Mining (separate: SAS EM for Text)
   - Advanced Statistics
2. Proprietary Development Platform
   - SAS Code (proprietary)
3. **Costs** (SAS EM: $150K/5 users)
   - Annual Renewal Fee (~40% each year)

IDC Worldwide Business Analytics Software

Business Analytics Software Competitive Market Map, 2006



Notes

Size is the measure of a vendor's software revenue in the selected market.

Momentum is the size-adjusted annual software growth rate for the selected market.

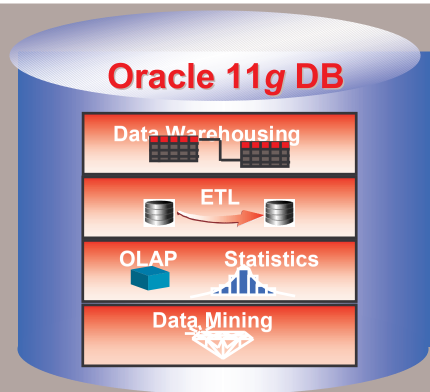
Reliance is a vendor's dependence on selected software revenue.

Diversity is the measure of the breadth and depth of product offerings within the selected software market.

Source: IDC, 2007

http://www.oracle.com/corporate/analyst/reports/infrastructure/bi_dw/208699e.pdf

ORACLE DATABASE 11g



ODM Algorithms, Predictive Analytics & Statistical Functions

Oracle Data Mining

Algorithm Summary 11g

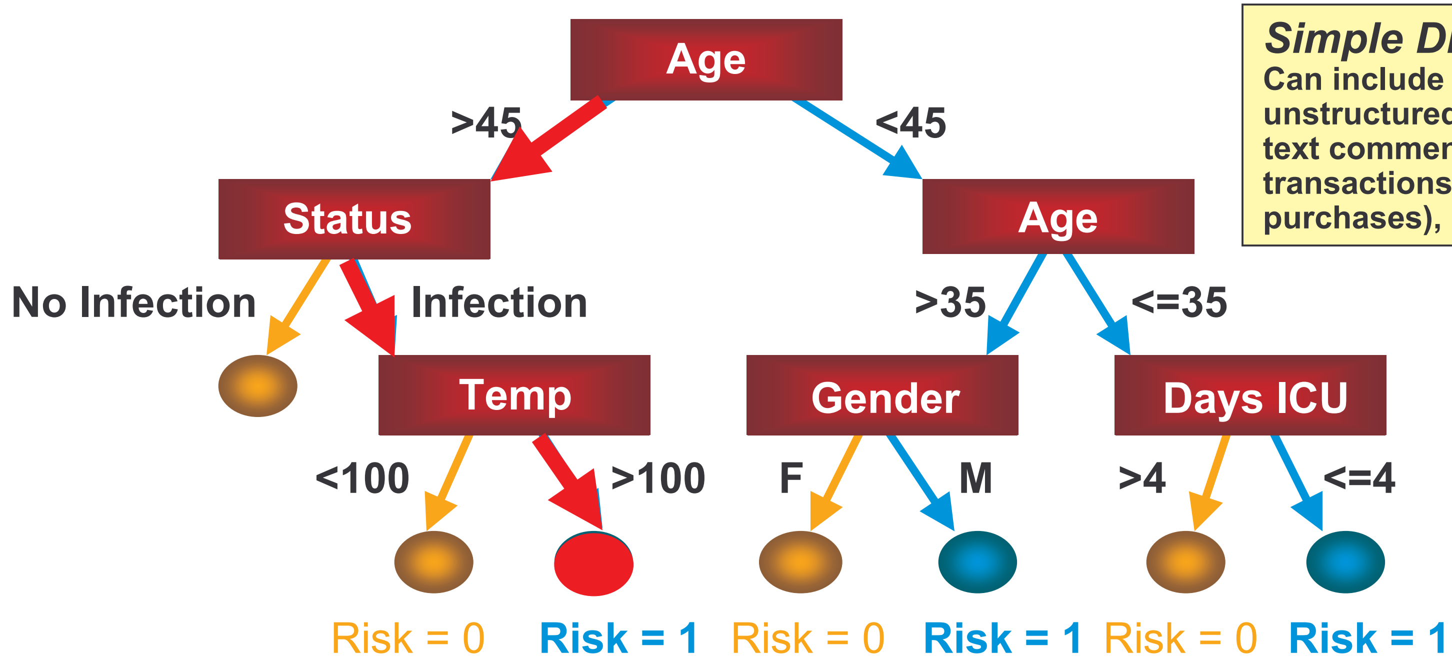
Problem	Algorithm	Applicability
Classification	**Logistic Regression (GLM)** **Decision Trees** **Naïve Bayes** **Support Vector Machine**	**Classical statistical technique** **Popular / Rules / transparency** **Embedded app** **Wide / narrow data / text**
Regression	**Multiple Regression (GLM)** **Support 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

Oracle Data Mining 10*g*R2

Decision Trees

- Classification, Prediction, Patient "profiling"



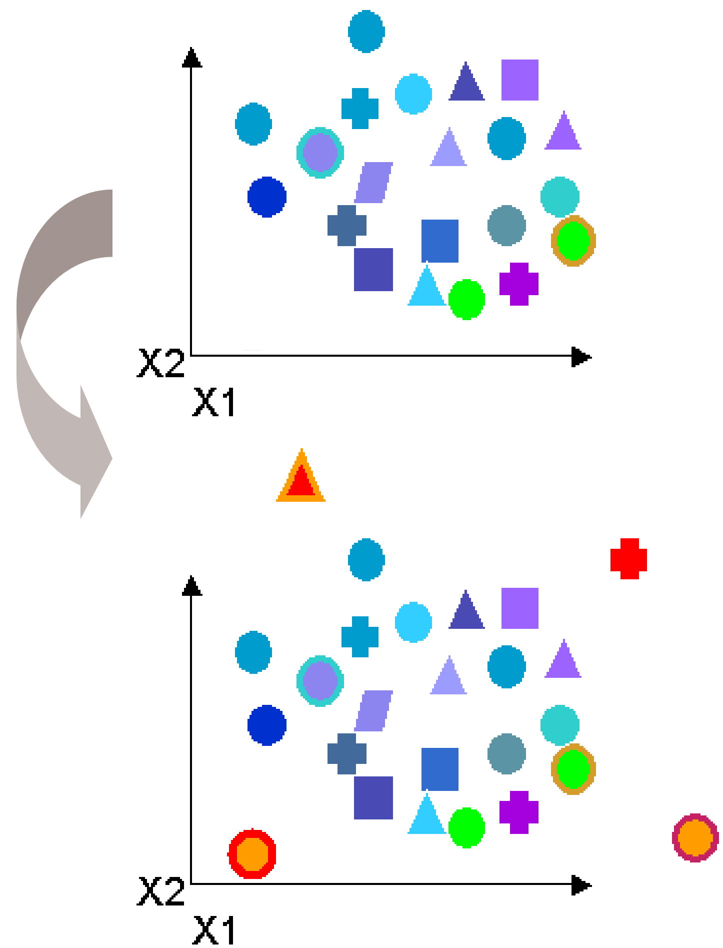
Simple DM model:
Can include unstructured data (e.g. text comments), transactions data (e.g. purchases), etc.

IF (Age > 45 AND Status = Infection AND Temp = >100)
THEN P(High Risk=1) = .77 Support = 250

Oracle Data Mining 11g
Anomaly Detection

- “One-Class” SVM Models
  - Fraud, noncompliance
  - Outlier detection
  - Network intrusion detection
  - Disease outbreaks
  - Rare events, true novelty

Problem: Detect rare cases

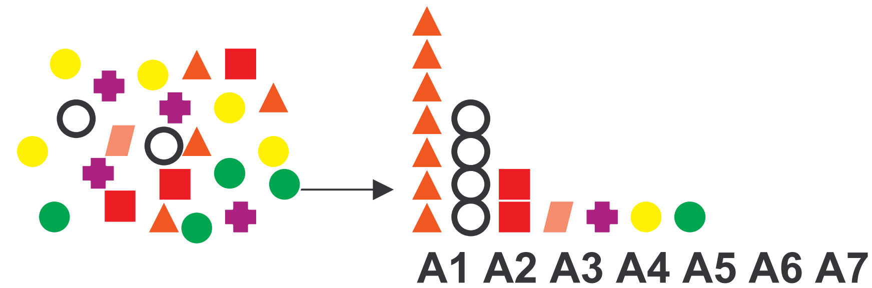


Oracle Data Mining

Algorithms & Example Applications

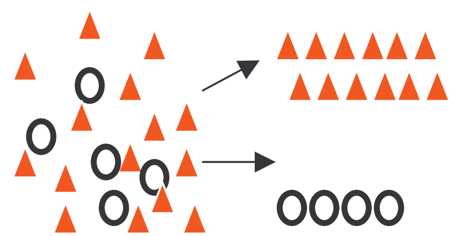
Attribute Importance

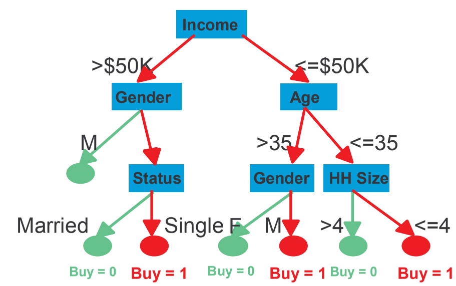
- Identify most influential attributes for a target attribute
  - Factors associated with high costs, responding to an offer, etc.



Classification and Prediction

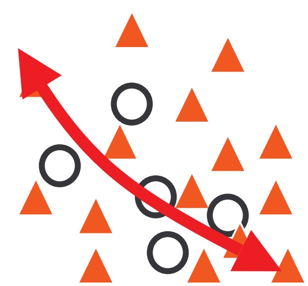
- Predict customers most likely to:
  - Respond to a campaign or offer
  - Incur the highest costs
- Target your best customers
- Develop customer profiles





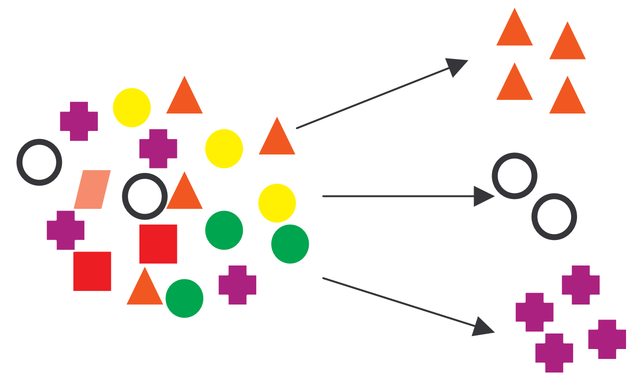
Regression

- Predict a numeric value
  - Predict a purchase amount or cost
  - Predict the value of a home



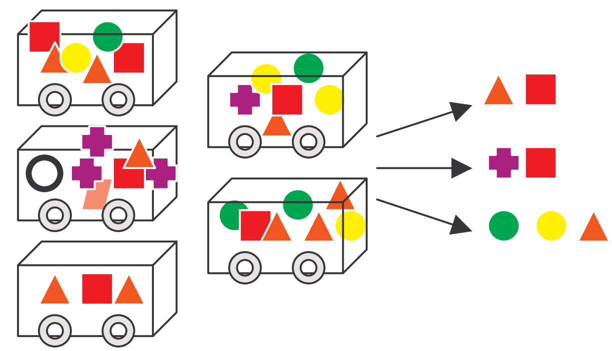
Oracle Data Mining

Algorithms & Example Applications



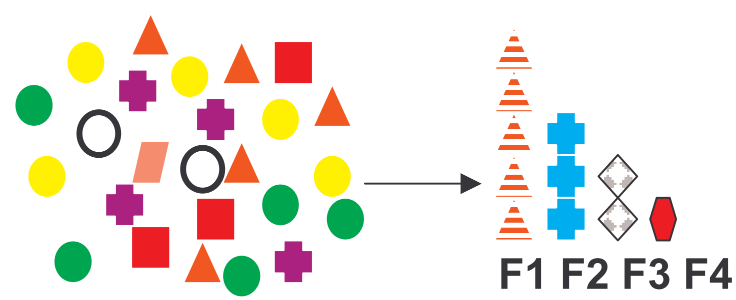
Clustering

- Find naturally occurring groups
  - Market segmentation
  - Find disease subgroups
  - Distinguish normal from non-normal behavior



Association Rules

- Find co-occurring items in a market basket
  - Suggest product combinations
  - Design better item placement on shelves



Feature Extraction

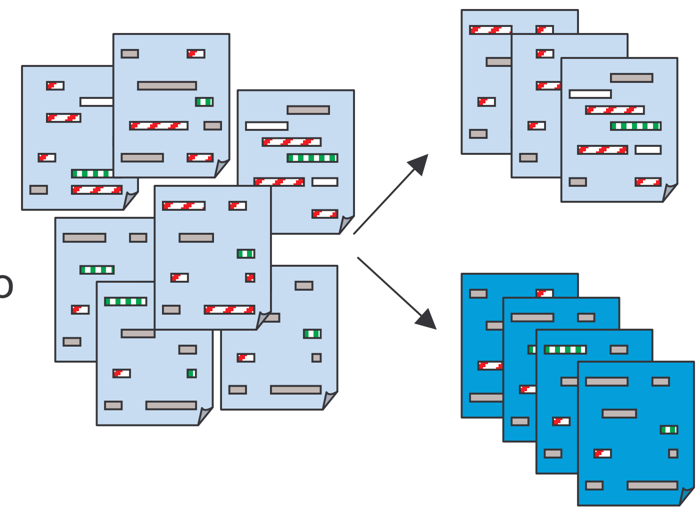
- Reduce a large dataset into representative new attributes
  - Useful for clustering and text mining

Oracle Data Mining

Algorithms & Example Applications

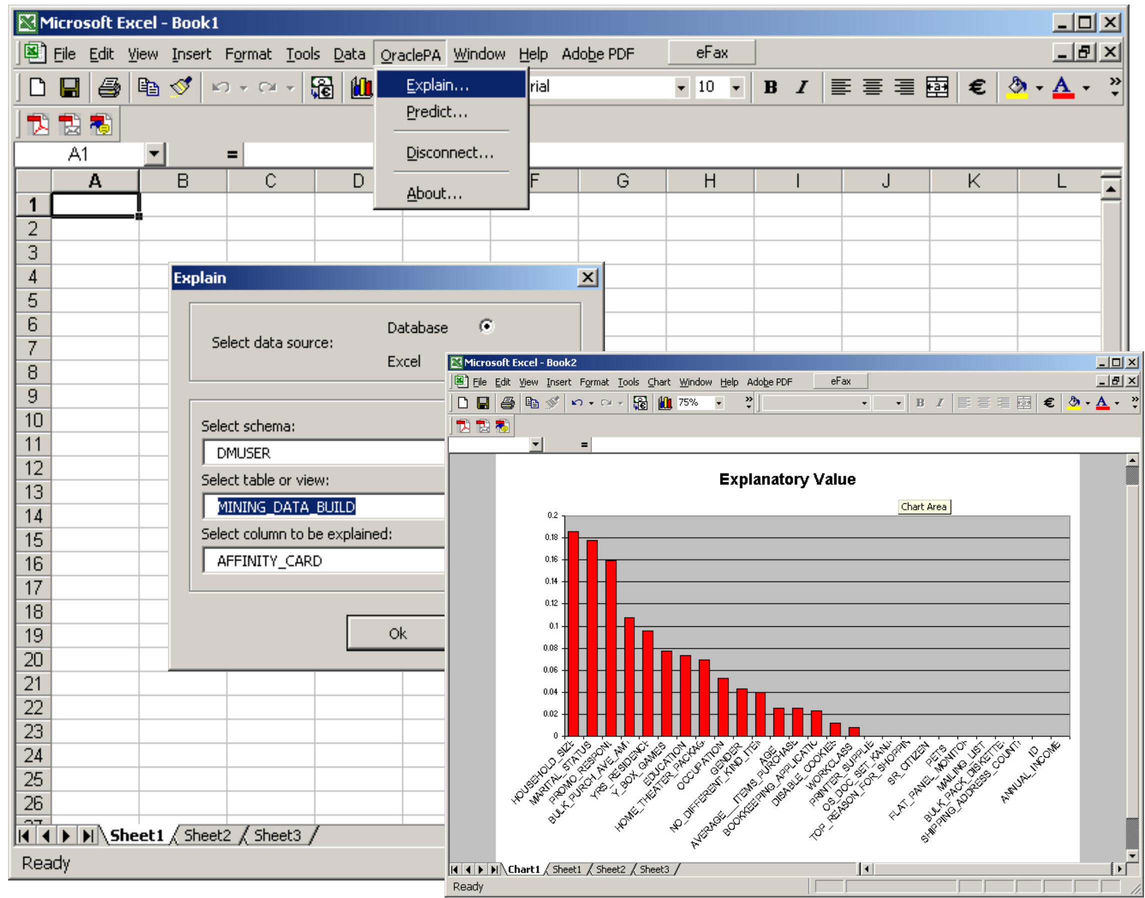
Text Mining

- Combine data and text for better models
  - Add unstructured text e.g. physician's notes to structured data e.g. age, weight, height, etc., to predict outcomes
- Classify and cluster documents
  - Combined with Oracle Text to develop advanced text mining applications e.g. Medline



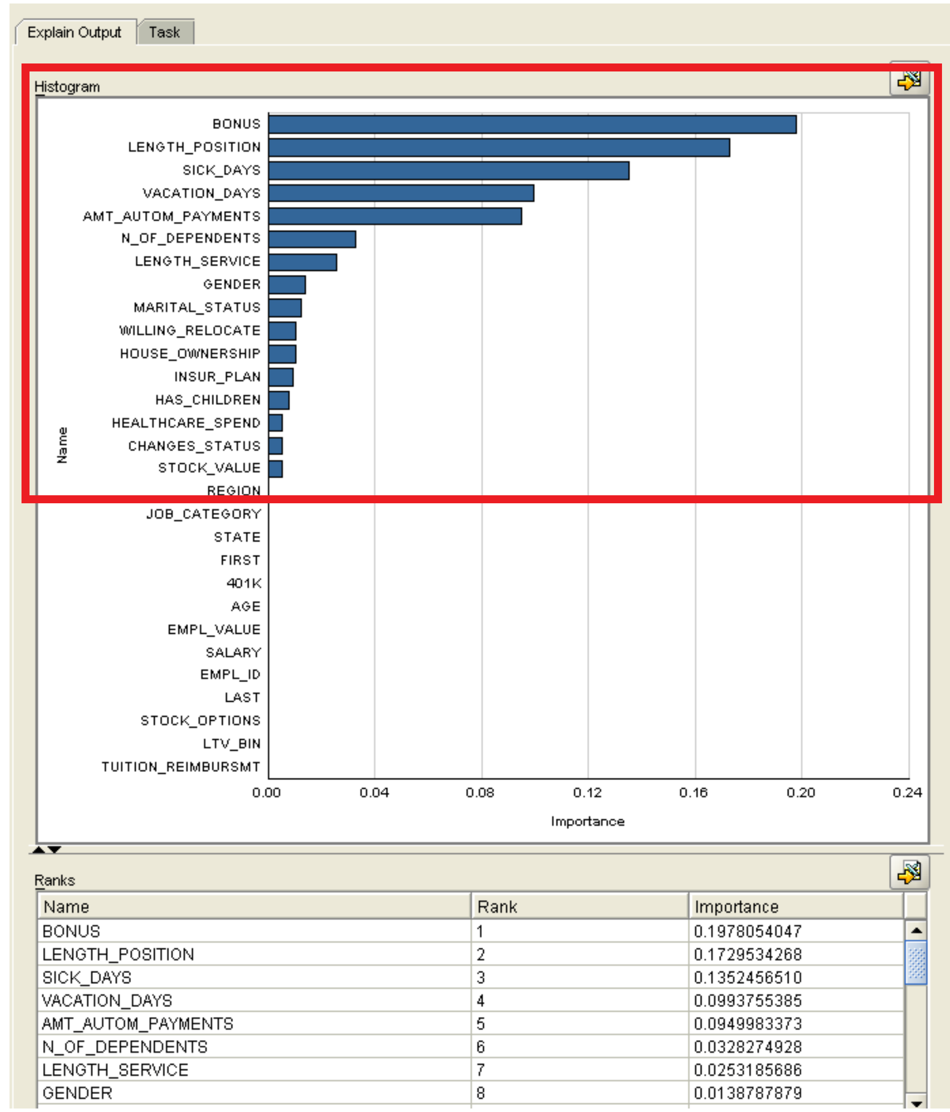
Predictive Analytics

- *"One-click data mining"*
  - Automates methodology used by expert data miners
- PL/SQL packages
  - PREDICT
  - EXPLAIN
  - PROFILE



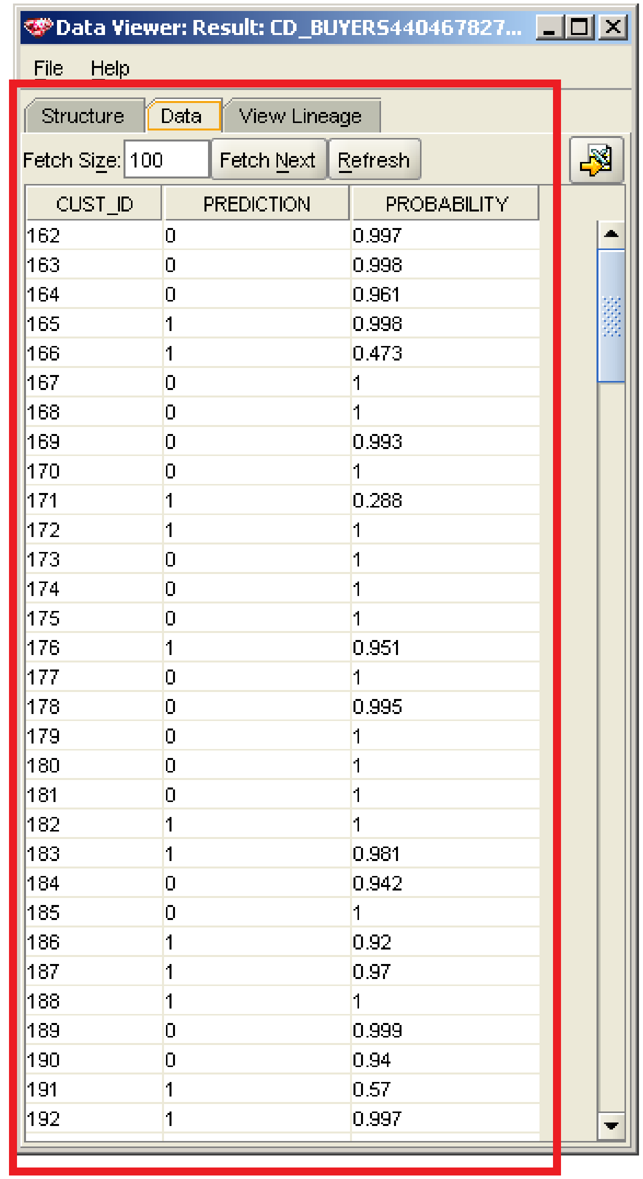
EXPLAIN PA

- Automated Attribute Importance
- Identifies key factors that have the strongest influence on the target variable
  - Satisfied customers
  - Hi-Value customers
  - Healthy patients
  - Responders
  - etc.



PREDICT PA

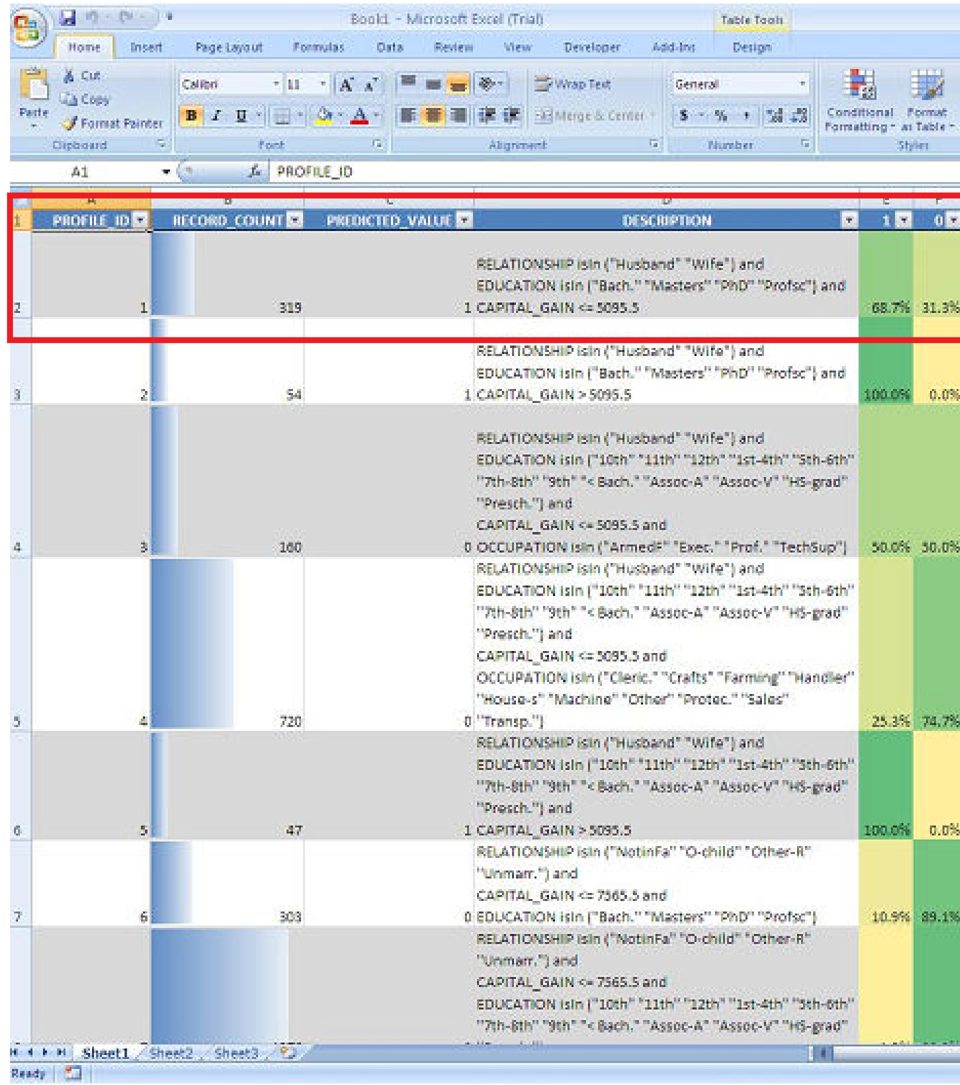
- Automated model build and model score
- Produces predictions:
  - People likely to respond
  - Customers likely to buy a product
  - Hi-Value customers
  - Likely Fraud
  - etc.



CUST_ID	PREDICTION	PROBABILITY
162	0	0.997
163	0	0.998
164	0	0.961
165	1	0.998
166	1	0.473
167	0	1
168	0	1
169	0	0.993
170	0	1
171	1	0.288
172	1	1
173	0	1
174	0	1
175	0	1
176	1	0.951
177	0	1
178	0	0.995
179	0	1
180	0	1
181	0	1
182	1	1
183	1	0.981
184	0	0.942
185	0	1
186	1	0.92
187	1	0.97
188	1	1
189	0	0.999
190	0	0.94
191	1	0.57
192	1	0.997

PROFILE PA

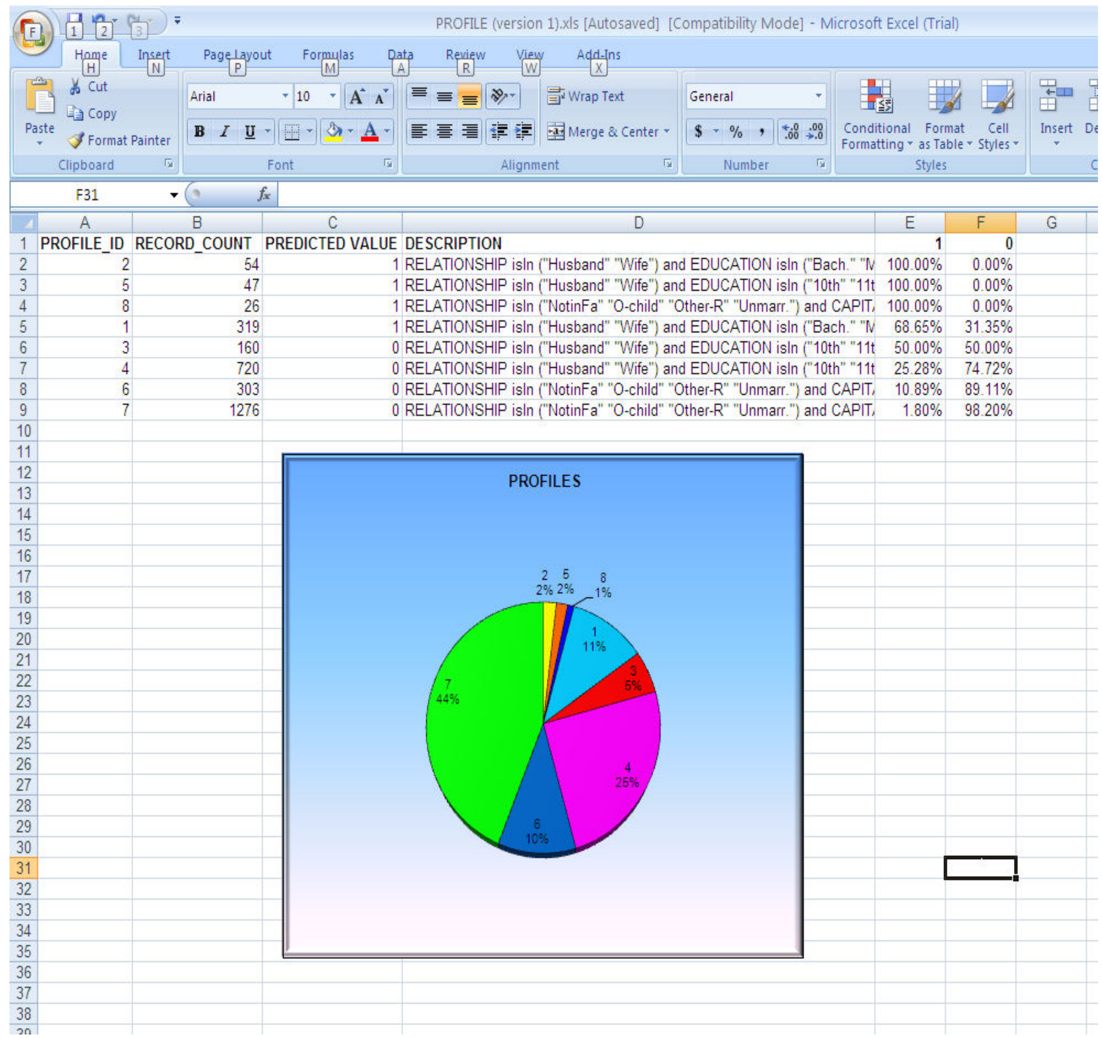
- PROFILE finds targeted customer segments, size and confidence and IF… THEN…. “rules” that describe them



PROFILE_ID	RECORD_COUNT	PREDICTED_VALUE	DESCRIPTION	1	0
1	319	1	RELATIONSHIP isIn ("Husband" "Wife") and EDUCATION isIn ("Bach." "Masters" "PhD" "Profsc") and CAPITAL_GAIN <= 5095.5	68.7%	31.3%
2	54	1	RELATIONSHIP isIn ("Husband" "Wife") and EDUCATION isIn ("Bach." "Masters" "PhD" "Profsc") and CAPITAL_GAIN > 5095.5	100.0%	0.0%
3	160	0	RELATIONSHIP isIn ("Husband" "Wife") and EDUCATION isIn ("10th" "11th" "12th" "1st-4th" "5th-6th" "7th-8th" "9th" "< Bach." "Assoc-A" "Assoc-V" "HS-grad" "Presch.") and CAPITAL_GAIN <= 5095.5 and OCCUPATION isIn ("ArmedF" "Exec." "Prof." "TechSup")	50.0%	50.0%
4	720	0	RELATIONSHIP isIn ("Husband" "Wife") and EDUCATION isIn ("10th" "11th" "12th" "1st-4th" "5th-6th" "7th-8th" "9th" "< Bach." "Assoc-A" "Assoc-V" "HS-grad" "Presch.") and CAPITAL_GAIN <= 5095.5 and OCCUPATION isIn ("Cleric." "Crafts" "Farming" "Handler" "House-s" "Machine" "Other" "Protec." "Sales" "Transp.")	25.3%	74.7%
5	47	1	RELATIONSHIP isIn ("Husband" "Wife") and EDUCATION isIn ("10th" "11th" "12th" "1st-4th" "5th-6th" "7th-8th" "9th" "< Bach." "Assoc-A" "Assoc-V" "HS-grad" "Presch.") and CAPITAL_GAIN > 5095.5	100.0%	0.0%
6	303	0	RELATIONSHIP isIn ("NotinFa" "O-child" "Other-R" "Unmarr.") and CAPITAL_GAIN <= 7565.5 and EDUCATION isIn ("Bach." "Masters" "PhD" "Profsc")	10.9%	89.1%
			RELATIONSHIP isIn ("NotinFa" "O-child" "Other-R" "Unmarr.") and CAPITAL_GAIN <= 7565.5 and EDUCATION isIn ("10th" "11th" "12th" "1st-4th" "5th-6th" "7th-8th" "9th" "< Bach." "Assoc-A" "Assoc-V" "HS-grad"		

PROFILE PA

- PROFILE finds targeted customer segments, size and confidence and IF… THEN…. “rules” that describe them

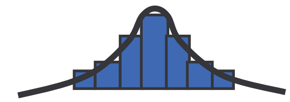


Links below point to Oracle 10g Release 2 SQL Reference documentation for database functions.

Numeric	Character	Aggregate	Aggregate (contd)	Analytic (contd)	DateTime	Conversion
ABS	CHR	AVG	STATS_MW_TEST	PERCENTILE_CONT	ADD_MONTHS	ASCIISTR
ACOS	CONCAT	COLLECT	STATS_T_TEST_*	PERCENTILE_DISC	CURRENT_DATE	BIN_TO_NUM
ASIN	INITCAP	CORR	STATS_WSR_TEST	RANK	CURRENT_TIMESTAMP	CAST
ATAN	LOWER	CORR_*	STDDEV	RATIO_TO_REPORT	DBTIMEZONE	CHARTOROWID
ATAN2	LPAD	CORR_S	STDDEV_POP	REGR_ (Linear Regr) *	EXTRACT (datetime)	COMPOSE
BITAND	LTRIM	CORR_K	STDDEV_SAMP	ROW_NUMBER	FROM_TZ	CONVERT
CEIL	NLS_INITCAP	COUNT	SUM	STDDEV *	LAST_DAY	DECOMPOSE
COS	NLS_LOWER	COVAR_POP	VAR_POP	STDDEV_POP *	LOCALTIMESTAMP	HEXTORAW
COSH	NLSSORT	COVAR_SAMP	VAR_SAMP	STDDEV_SAMP *	MONTHS_BETWEEN	NUMTODSINTERVAL
EXP	NLS_UPPER	CUME_DIST	VARIANCE	SUM *	NEW_TIME	NUMTOYMINTERVAL
FLOOR	REGEXP_REPLACE	DENSE_RANK	**Analytic**	VAR_POP *	NEXT_DAY	RAWTOHEX
LN	REGEXP_SUBSTR	FIRST	AVG *	VAR_SAMP *	NUMTODSINTERVAL	RAWTONHEX
LOG	REPLACE	GROUP_ID	CORR *	VARIANCE *	NUMTOYMINTERVAL	ROWIDTOCHAR
MOD	RPAD	GROUPING	COVAR_POP *	**Data Mining**	ROUND (date)	ROWIDTONCHAR
NANVL	RTRIM	GROUPING_ID	COVAR_SAMP *	CLUSTER_ID	SESSIONTIMEZONE	SCN_TO_TIMESTAMP
POWER	SOUNDEX	LAST	COUNT *	CLUSTER_PROBABILITY	SYS_EXTRACT_UTC	TIMESTAMP_TO_SCN
REMAINDER	SUBSTR	MAX	CUME_DIST	CLUSTER_SET	SYSDATE	TO_BINARY_DOUBLE
ROUND (number)	TRANSLATE	MEDIAN	DENSE_RANK	FEATURE_ID	SYSTIMESTAMP	TO_BINARY_FLOAT
SIGN	TREAT	MIN	FIRST	FEATURE_SET	TO_CHAR (datetime)	TO_CHAR (character)
SIN	TRIM	PERCENTILE_CONT	FIRST_VALUE *	FEATURE_VALUE	TO_TIMESTAMP	TO_CHAR (datetime)
SINH	UPPER	PERCENTILE_DISC	LAG	PREDICTION	TO_TIMESTAMP_TZ	TO_CHAR (number)
SQRT		PERCENT_RANK	LAST	PREDICTION_COST	TO_DSINTERVAL	TO_CLOB
TAN	ASCII	RANK	LAST_VALUE *	PREDICTION_DETAILS	TO_YMINTERVAL	TO_DATE
TANH	INSTR	REGR_ (Linear Regr)	LEAD	PREDICTION_PROBABILITY	TRUNC (date)	TO_DSINTERVAL
TRUNC (number)	LENGTH	STATS_BINOMIAL_TEST	MAX *	PREDICTION_SET	TZ_OFFSET	TO_LOB
WIDTH_BUCKET	REGEXP_INSTR	STATS_CROSSTAB	MIN *		**Comparison**	TO_MULTI_BYTE
		STATS_MODE	NTILE		GREATEST	TO_NCHAR (character)
		STATS_ONE_WAY_ANOVA	PERCENT_RANK		LEAST	TO_NCHAR (datetime)
		STATS_F_TEST				
		STATS_KS_TEST				

Note: * next to certain Analytic functions indicates that they support moving window specifications.

Descriptive Statistics



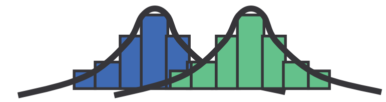
- MEDIAN & MODE

> SQL

  - Median: takes numeric or datetype values and returns the middle value
  - Mode: returns the most common value

```
A. SELECT STATS_MODE(EDUCATION) from CD_BUYERS;

B. SELECT MEDIAN(ANNUAL_INCOME) from CD_BUYERS;

C. SELECT EDUCATION, MEDIAN(ANNUAL_INCOME) from
   CD_BUYERS GROUP BY EDUCATION;

D. SELECT EDUCATION, MEDIAN(ANNUAL_INCOME) from
   CD_BUYERS GROUP BY EDUCATION ORDER BY
   MEDIAN(ANNUAL_INCOME) ASC;
```

One-Sample T-Test



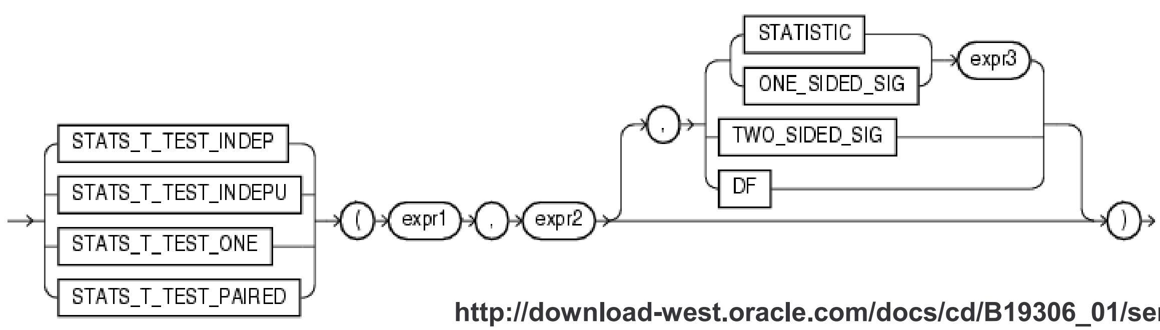
STATS_T_TEST_*

The t-test functions are:

- **STATS_T_TEST_ONE**: A one-sample t-test
- **STATS_T_TEST_PAIRED**: A two-sample, paired t-test (also known as a crossed t-test)
- **STATS_T_TEST_INDEP**: A t-test of two independent groups with the same variance (pooled variances)
- **STATS_T_TEST_INDEPU**: A t-test of two independent groups with unequal variance (unpooled variances)

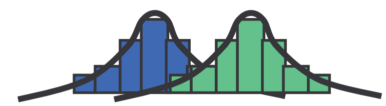
Syntax

stats_t_test::=



http://download-west.oracle.com/docs/cd/B19306_01/server.102/b14200/functions157.htm

Independent Samples T-Test



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

SQL Worksheet

File Help

Enter SQL Statement

```
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;
```

Comparison results grouped by gender with statistical level of significance

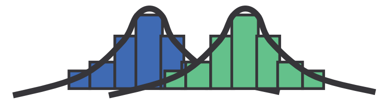
Results

Fetch Size: 100 Fetch Next Refresh

INCOME_LEVEL	SOLD_TO_MEN	SOLD_TO_WOMEN	T_OBSERVED	TWO_SIDED_P_VALUE
A: Below 30,000	1.05283489772952...	9.94281446665347...	-1.9880628862965...	.046811481623777469
B: 30,000 - 49,999	1.02596509506751...	1.09829641827200...	3.04330875305933...	.0023410534279906534
C: 50,000 - 69,999	1.05627588073092...	1.10127931012124...	2.36148671478380...	.018204221087140492
D: 70,000 - 89,999	1.06630299489770...	1.10472869932602...	2.28496442938685...	.022316997323373052
E: 90,000 - 109,999	1.03396741493733...	1.01610416258370...	-1.2544577321777...	.20967782263168744
F: 110,000 - 129,999	1.06764759620596...	1.05981311948214...	-6.0444998485409...	.54554530367779175
G: 130,000 - 149,999	1.08877532181064...	1.07313769857029...	-8.5298244925517...	.3936712177482169
H: 150,000 - 169,999	1.10987257925271...	1.07152191179957...	-1.9062363114356...	.056622982960629141
I: 170,000 - 189,999	1.02808237970970...	1.07435560141216...	2.18477851179480...	.028908565891614835
J: 190,000 - 249,999	1.08040563837250...	1.15343356029762...	2.58313424607376...	.0097945161125223348

ORACLE

Customer Example

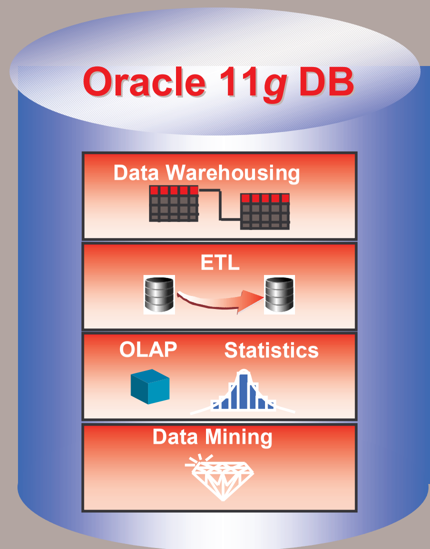


```
248 rows selected.

SQL> select peak_id peak,avg(decode(E.sample_group,'CNS',s.intensity,null)) avg_CNS, avg(decode(E.sa
mple_group,'ND',s.intensity,null)) avg_ND, stats_ks_test(E.sample_group,s.intensity,'STATISTIC') ks_
stat, stats_ks_test(E.sample_group,s.intensity) ks_p_value,  stats_t_test_indep(E.sample_group,s.int
ensity) t_test_p_value, avg(subs_mass) AVG_MASS from exp_descriptor E, celd_spectrum s where E.exp_i
d = s.exp_id and E.chip_id = s.chip_id and E.spot_number = s.spot_number and (sample_group = 'CNS' o
r sample_group='ND')  Group By peak_id order by  stats_t_test_indep(E.sample_group,s.intensity);

      PEAK    AVG_CNS     AVG_ND    KS_STAT KS_P_VALUE T_TEST_P_VALUE   AVG_MASS
---------- ---------- ---------- ---------- ---------- -------------- ----------
       178  1.3314339 2.17817187 .673333333 7.2556E-16     2.6544E-17 5952.91674
       181  5.0996028 7.89194275 .626666667 8.4480E-14     1.4848E-14  6075.9581
       180 2.27649538 3.47917519 .606666667 5.8453E-13     8.8539E-14 6055.14643
       182 1.82166302 2.70982458 .586666667 3.7986E-12     1.7684E-13 6093.52256
       112 1.43756807 .415726202         .6 1.0984E-12     4.5081E-13 4033.66603
       179 .470304995  .71366692 .546666667 1.3289E-10     6.0678E-13 5976.18528
       162  .32065549 .488111947 .606666667 5.8453E-13     6.3174E-13 5384.91078
       176 1.71447936 2.70554235 .553333333 7.4775E-11     1.7747E-12 5914.53224
       185 .336895407 .472142857        .55 9.9772E-11     1.9222E-12 6260.71013
       186 .401995708 .562915017 .506666667 3.6175E-09     2.1445E-12 6281.69466
       177  2.3623861 3.80160199 .586666667 3.7986E-12     4.1808E-12 5933.83033
```

"..Our experience suggests that Oracle 10g Statistics and Data Mining features can reduce development effort of analytical systems by an order of magnitude."

Sumeet Muju

Senior Member of Professional Staff, SRA International (SRA supports NIH bioinformatics development projects)

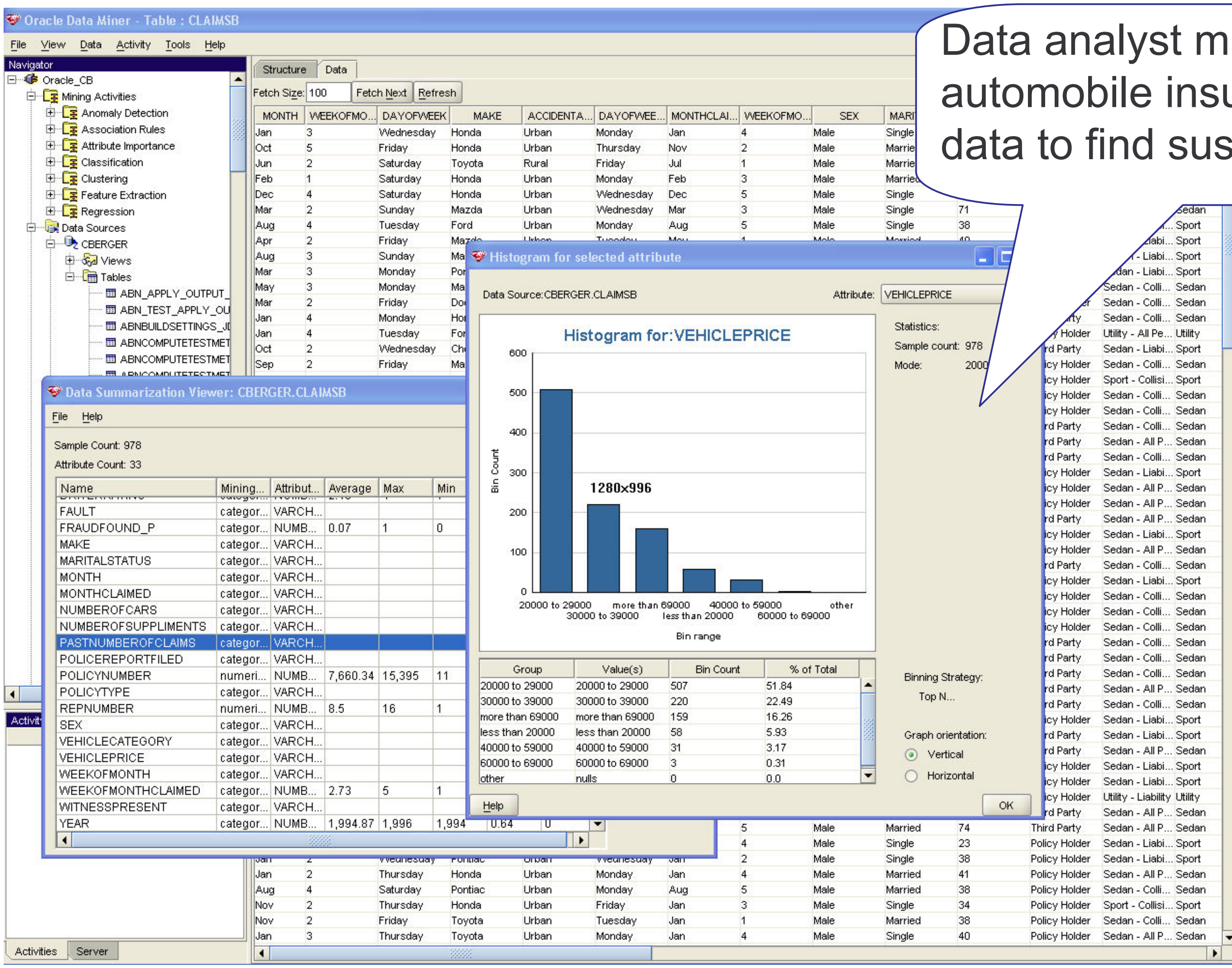


Oracle Data Mining + Oracle BI EE

D E M O N S T R A T I O N

Oracle Data Mining + OBI EE

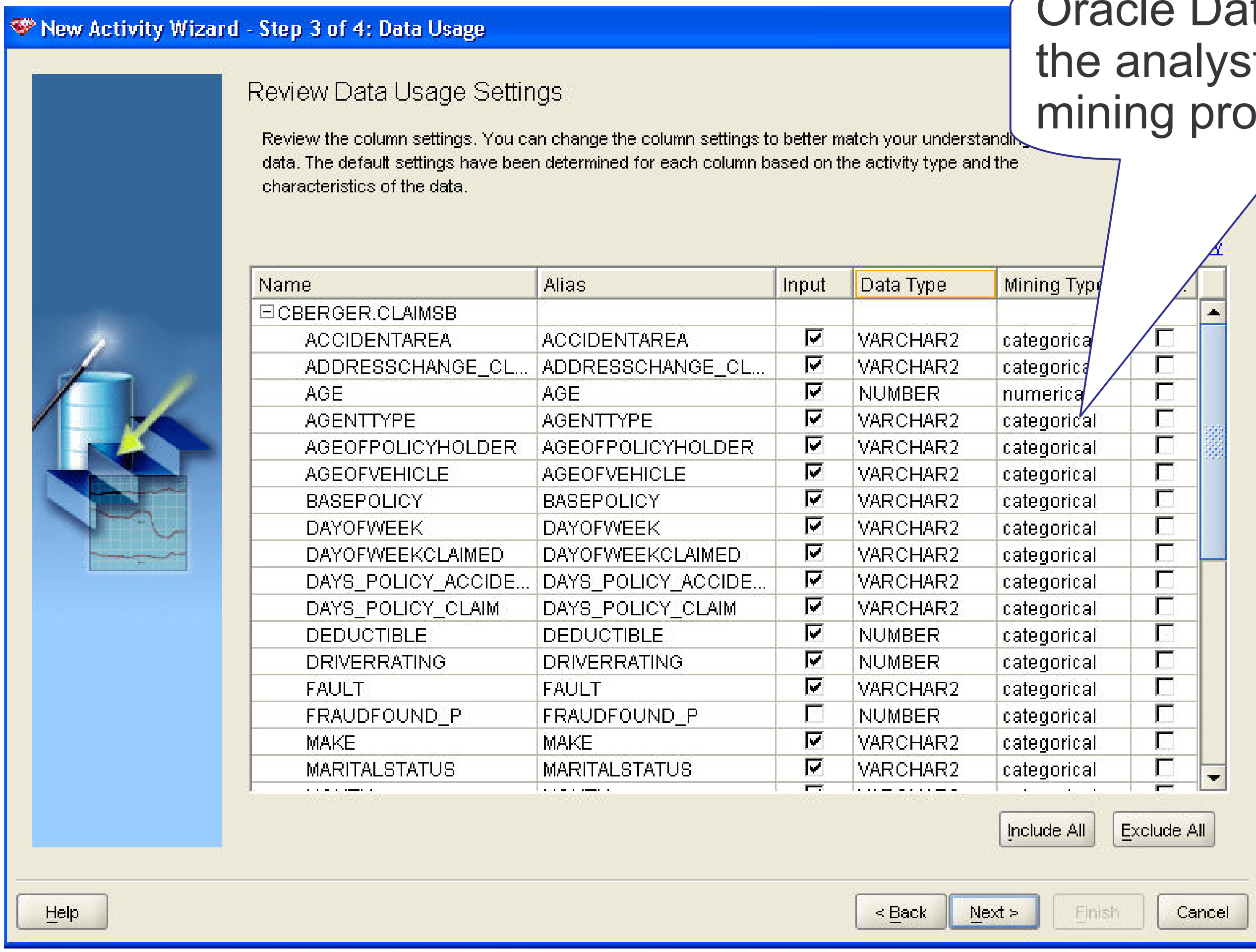
Auto Claims Fraud Demo



Data analyst mines automobile insurance claims data to find suspicious claims

Oracle Data Mining + OBI EE

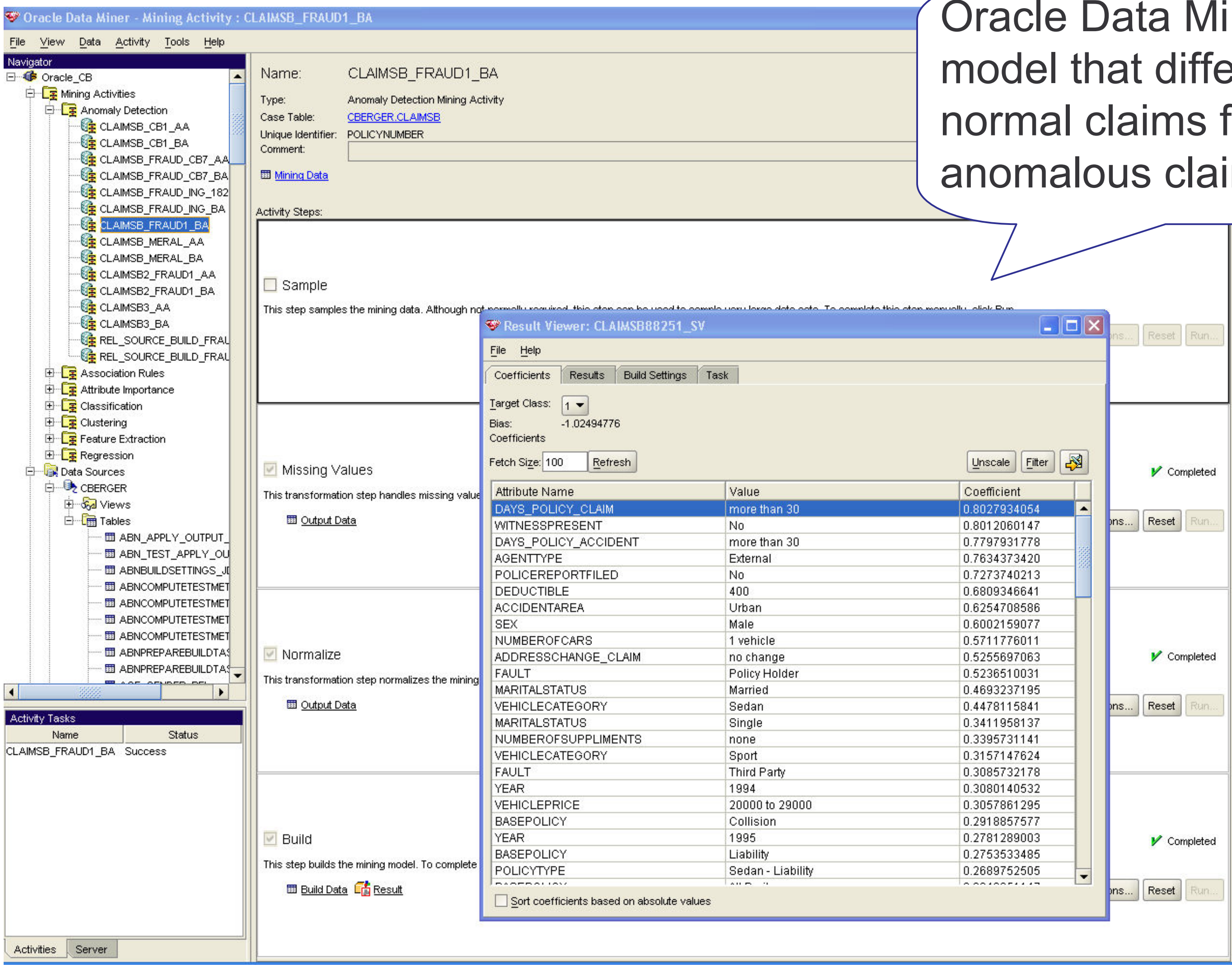
Auto Claims Fraud Demo



Oracle Data Miner guides the analyst through the data mining process

Oracle Data Mining + OBI EE

Auto Claims Fraud Demo



Oracle Data Mining builds a model that differentiates normal claims from anomalous claims

Oracle Data Mining + OBI EE

Auto Claims Fraud Demo

Oracle Data Mining creates a prioritized list of suspicious claims for investigation

Result Viewer: "CLAIMST768348894_A"

File Publish Help

Apply Output | Apply Settings | Task

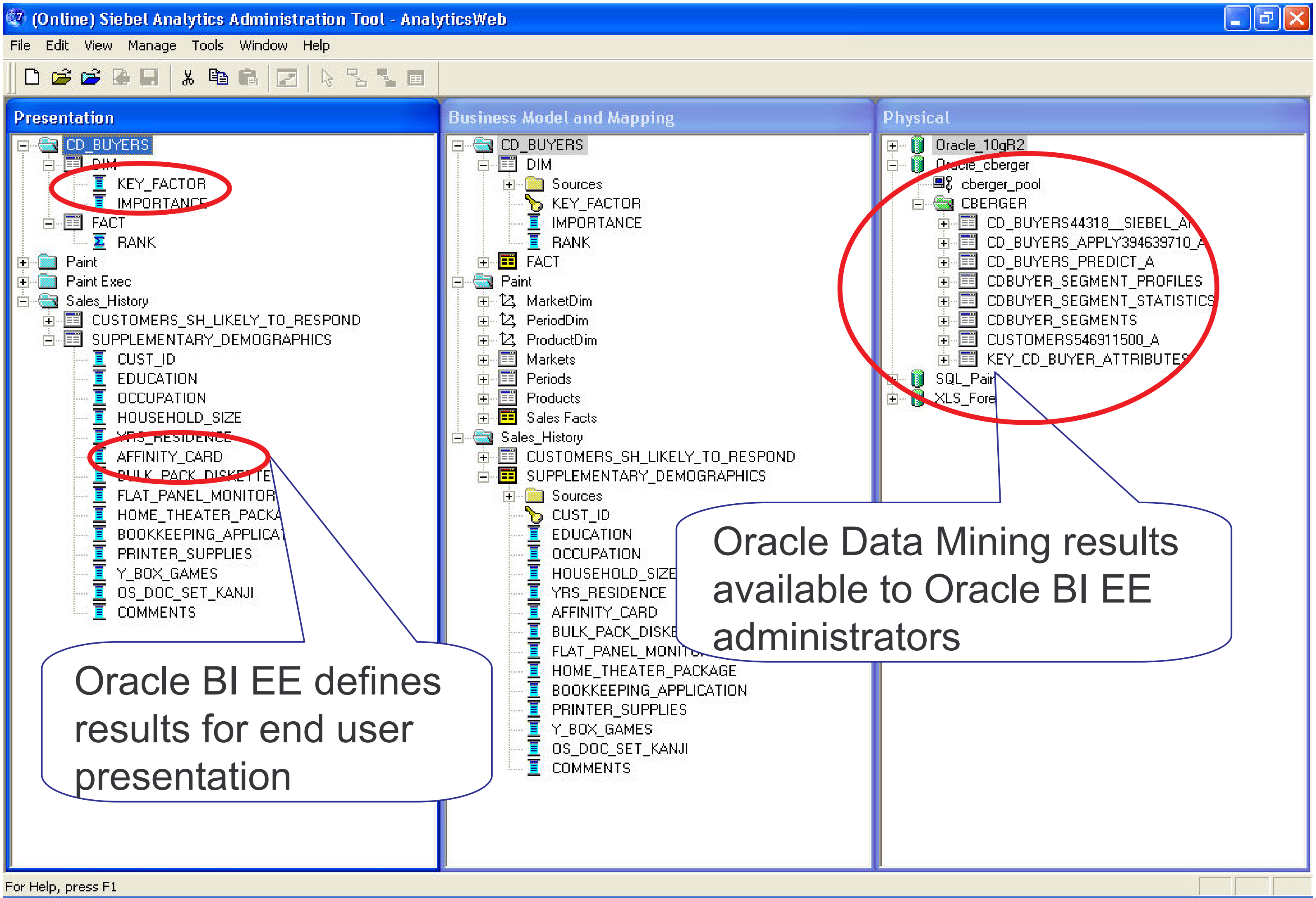
Apply Output Table:

Fetch Size: 10000 Refresh

DMR$CAS...	VEHICLEP...	POLICYTY...	NUMBER...	REPNUM...	AGE1	DRIVERR...	AGEOFVE...	POLICER...	WITNESS...	DEDUCTI...	ACCIDEN...	PREDICTI...	PROBAB...
11,015	30000 to 3...	Sedan - All ...	1 vehicle	7	45	1	more than 7	No	No	400	Urban	0	0.6311
1	more than ...	Sport - Lia...	3 to 4	12	21	1	3 years	No	No	300	Urban	0	0.6289
12,650	30000 to 3...	Sedan - Co...	1 vehicle	15	25	3	2 years	No	No	400	Urban	0	0.6278
11,922	more than ...	Utility - All ...	1 vehicle	15	46	1	5 years	No	No	400	Rural	0	0.6238
11,192	30000 to 3...	Sedan - Li...	1 vehicle	7	43	2	7 years	Yes	Yes	700	Rural	0	0.6162
13,294	20000 to 2...	Sedan - Co...	2 vehicles	7	28	2	5 years	No	Yes	400	Rural	0	0.6149
6,284	40000 to 5...	Sedan - Li...	1 vehicle	5	47	3	more than 7	No	No	400	Rural	0	0.6134
8,139	30000 to 3...	Sedan - Co...	1 vehicle	4	34	2	6 years	Yes	Yes	400	Rural	0	0.6115
11,119	more than ...	Sedan - All ...	1 vehicle	1	33	3	7 years	No	No	500	Rural	0	0.6078
11,677	20000 to 2...	Sedan - All ...	2 vehicles	15	49	1	5 years	No	No	400	Urban	0	0.6078
8,561	less than 2...	Sedan - All ...	1 vehicle	9	22	4	4 years	No	No	400	Urban	0	0.6056
11,068	20000 to 2...	Sedan - Co...	2 vehicles	3	54	2	7 years	Yes	No	700	Rural	0	0.6052
11,130	20000 to 2...	Sedan - All ...	1 vehicle	1	50	1	more than 7	Yes	Yes	700	Urban	0	0.6029
13,967	20000 to 2...	Sedan - Co...	1 vehicle	15	24	3	5 years	No	No	400	Rural	0	0.6025
9,673	20000 to 2...	Sedan - Co...	1 vehicle	7	22	2	4 years	Yes	Yes	400	Urban	0	0.6011
3,471	30000 to 3...	Sedan - Li...	1 vehicle	1	40	3	3 years	No	No	400	Urban	0	0.6005
4,365	30000 to 3...	Sedan - Co...	1 vehicle	6	60	3	6 years	No	No	400	Urban	0	0.5995
14,101	20000 to 2...	Sedan - Co...	1 vehicle	7	56	3	more than 7	Yes	Yes	400	Urban	0	0.5985
7,918	20000 to 2...	Sedan - All ...	2 vehicles	5	35	4	6 years	Yes	No	400	Urban	0	0.5965
11,532	more than ...	Sport - Coll...	1 vehicle	13	32	2	2 years	No	No	400	Urban	0	0.5964
11,229	30000 to 3...	Sedan - All ...	1 vehicle	7	54	3	5 years	No	No	400	Urban	0	0.5953
2,931	more than ...	Sport - Coll...	1 vehicle	10	42	4	7 years	No	No	400	Urban	0	0.5941
3,044	30000 to 3...	Sedan - Li...	1 vehicle	8	44	2	more than 7	Yes	No	400	Urban	0	0.5938
5,797	20000 to 2...	Sedan - Co...	2 vehicles	2	26	2	6 years	No	No	400	Urban	0	0.5924
9,398	20000 to 2...	Sedan - Co...	1 vehicle	1	59	4	7 years	Yes	Yes	400	Urban	0	0.5917
4,642	30000 to 3...	Sedan - Co...	2 vehicles	6	50	1	more than 7	No	No	500	Rural	0	0.5908
1,599	more than ...	Sport - Coll...	1 vehicle	7	22	4	5 years	No	No	400	Urban	0	0.5878
627	20000 to 2...	Sedan - Co...	1 vehicle	15	33	3	7 years	No	No	400	Urban	1	0.5873
12,041	20000 to 2...	Sedan - Co...	1 vehicle	11	33	1	7 years	No	No	400	Urban	1	0.5867
2,989	20000 to 2...	Sedan - Co...	1 vehicle	14	30	3	7 years	No	No	400	Urban	1	0.5865
2,092	20000 to 2...	Sedan - Li...	3 to 4	13	36	2	7 years	No	No	500	Urban	0	0.5864

Rule...

Integration with Oracle BI EE

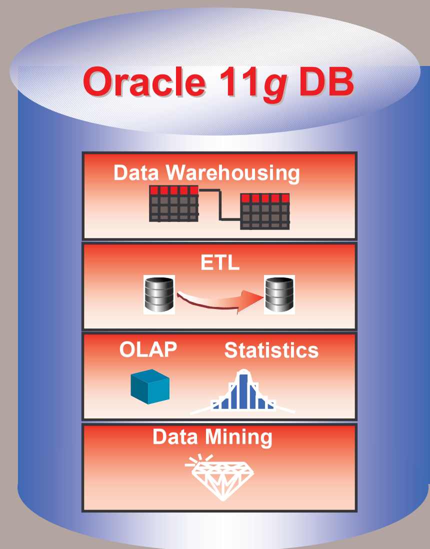


Integration with Oracle BI EE

POLICYNUMBER	PREDICTION	PROBABILITY	DEDUCTIBLE	DEDUCTIBLE * PROBABILITY
13465.0	0.00	0.67	500.00	**336.46**
14485.0	0.00	0.65	400.00	260.13
6532.0	0.00	0.64	400.00	257.69
12631.0	0.00	0.63	400.00	251.91
1.0	0.00	0.63	300.00	188.81
15029.0	0.00	0.63	400.00	251.29
11015.0	0.00	0.63	400.00	251.25
11922.0	0.00	0.62	400.00	249.84
4558.0	0.00	0.62	400.00	249.31
12650.0	0.00	0.62	400.00	249.17
2749.0	0.00	0.62	400.00	248.72
12307.0	0.00	0.62	400.00	248.56
3381.0	0.00	0.62	400.00	248.23
8863.0	0.00	0.62	400.00	247.51
13677.0	0.00	0.62	400.00	246.72
11192.0	0.00	0.62	700.00	**430.72**
13294.0	0.00	0.61	400.00	245.66
6284.0	0.00	0.61	400.00	245.50
2291.0	0.00	0.61	400.00	245.32
3440.0	0.00	0.61	400.00	244.70
8139.0	0.00	0.61	400.00	242.98
11068.0	0.00	0.61	700.00	**424.98**
11119.0	0.00	0.61	500.00	**303.55**
11677.0	0.00	0.61	400.00	242.83
1345.0	0.00	0.61	400.00	242.36

DRIVERRATING	SUM(DEDUCTIBLE * PROBABILITY)
1.00	352780.94
2.00	342473.34
3.00	340611.68
4.00	328747.06

Oracle Data Mining provides more information and better insight



Text Mining, Code Generation

Oracle Data Mining and Unstructured Data

- Oracle Data Mining includes unstructured i.e. “text” data
- Build classification and clustering models
- Oracle Text used to preprocess unstructured text

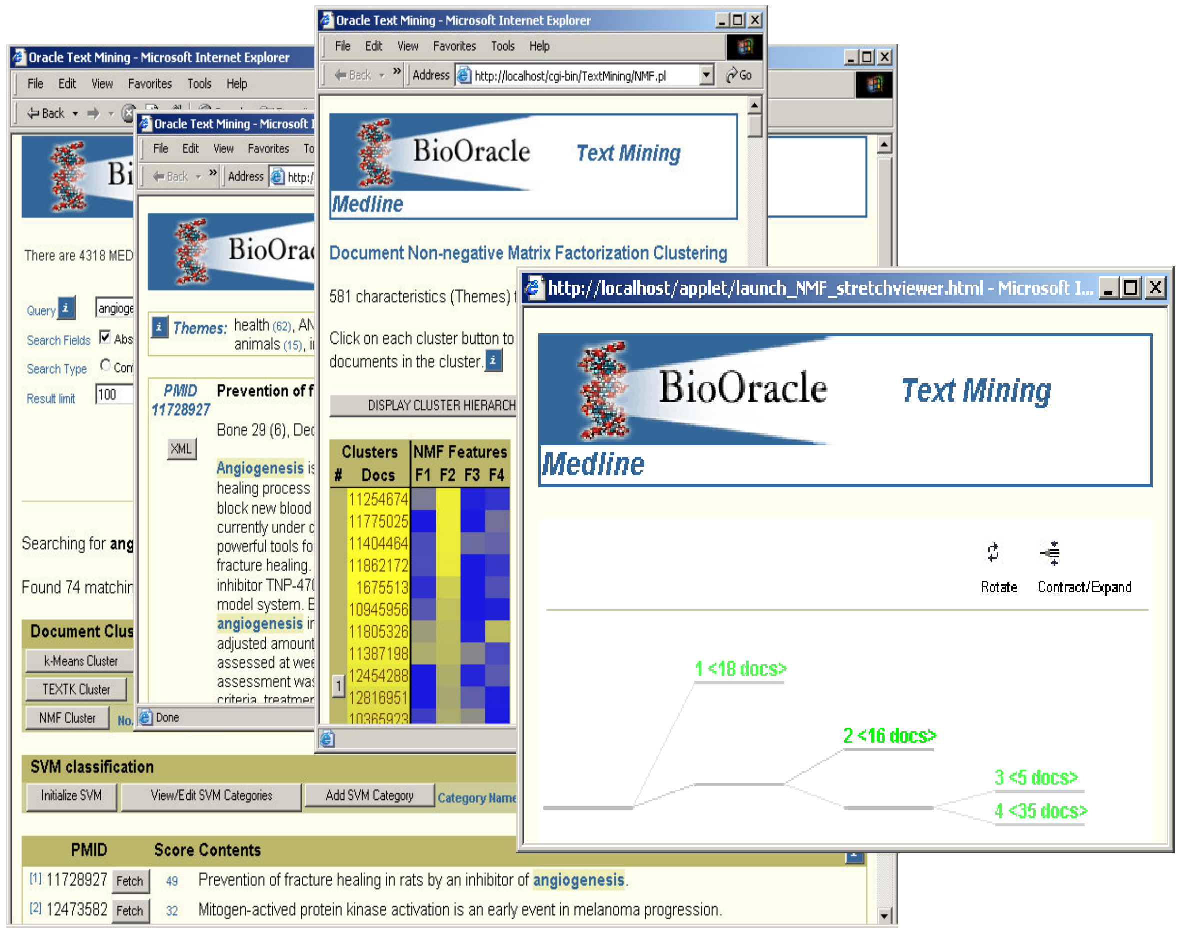
Structure | Data

Fetch Size: 100 | Fetch Next | Refresh

CUST_ID	AFFINITY_CARD	AGE	CU...	COMMENTS	CUST_MARI...
101501	0	41	F	Shopping at your store is a hassle. I rarely shop there and usually forget to bring your new loyalty c...	NeverM
101502	0	27	M	Affinity card is great. I think it is a hassle to have to remember to bring it in every time though.	NeverM
101503	0	20	F	I purchased a new computer recently, but the manuals weren't included. Could you ship them to me ...	NeverM
101504	1	45	M	Affinity card is great. I think it is a hassle to have to remember to bring it in every time though.	Married
101505	1	34	M	Why didn't you start a program like this before? Everyone else has been offering discounts like this f...	NeverM
101506	0	38	M	Forget it. I 'm not giving you all my personal information. I wish you'd give up and respect a customer...	Married
101507	0	28	M	It is a good way to attract new shoppers. After shopping at your store for more than a month, I am r...	Married
101508	0	19	M	I shop your store a lot. I love your weekly specials.	NeverM
101509	0	52	M	Affinity card makese sense only for bulk purchases. For all others, driving so far is not worth the di...	Married
101510	1	27	M	Could you send an Affinity Card to my mother in France? Let me know and I'll send you here address	NeverM
101511	0	30	M	Shopping at your store is a hassle. I rarely shop there and usually forget to bring your new loyalty c...	NeverM
101512	0	30	F	The new affinity card is great. Thank you. I do have to say that it is a hassle to remember to bring it ...	NeverM
101513	0	31	M	Thanks but even with your discounts, your products are too expensive. Sorry.	Married
101514	0	45	M	Affinity card is great. I think it is a hassle to have to remember to bring it in every time though.	NeverM
101515	0	36	F	I purchased the new mouse pads and love them. I also purchased one for my sister and one for my ...	NeverM
101516	0	33	M	Don't send me any more promotions. I get too much lousy junk mail already	Married
101517	0	38	F	Shopping at your store is a hassle. I rarely shop there and usually forget to bring your new loyalty c...	NeverM
101518	0	22	M	Don't send me any more promotions. I get too much lousy junk mail already	NeverM
101519	0	46	F	Shopping at your store is a hassle. I rarely shop there and usually forget to bring your new loyalty c...	Divorc.
101520	1	39	M	Affinity card is great. I think it is a hassle to have to remember to bring it in every time though.	Married
101521	0	61	M	I shop your store a lot. I love your weekly specials.	Married
101522	1	39	F	If I forget my affinity card, can I still shop here and get the discount?	NeverM
101523	0	22	M	A great program but I have to complain just a bit. Why do you need to know how many children I hav...	Mabsent
101524	0	38	M	Thank you, But please remove my name from your list.	Married
101525	0	18	F	My brother uses the affinity card a lot. I think the competitor has better prices without it.	NeverM

Custom Oracle Data & Text Mining Applications

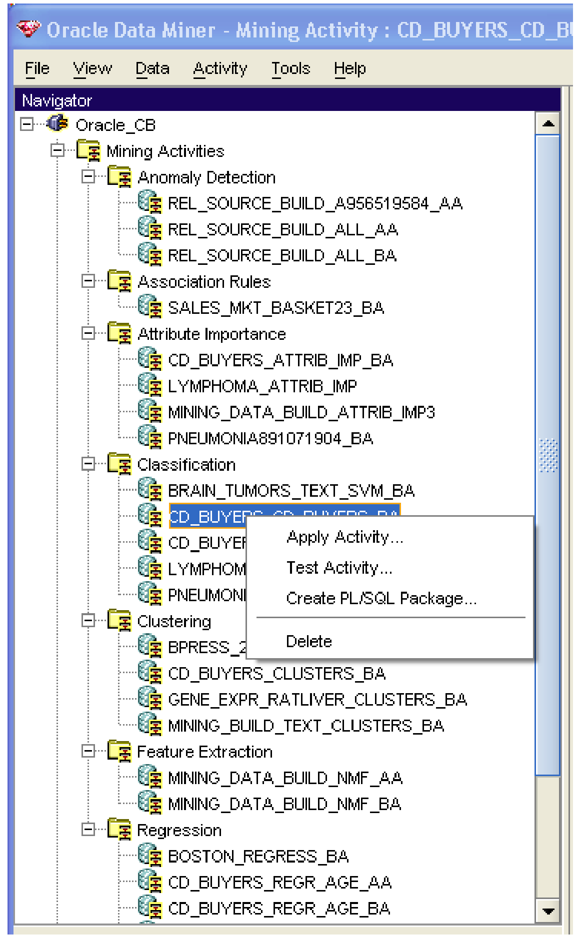
- Oracle Text (included in Oracle Database Standard Edition) preprocesses unstructured text
- Custom applications to mine:
  - Emails
  - Police reports
- Handles large volumes of “documents” or text

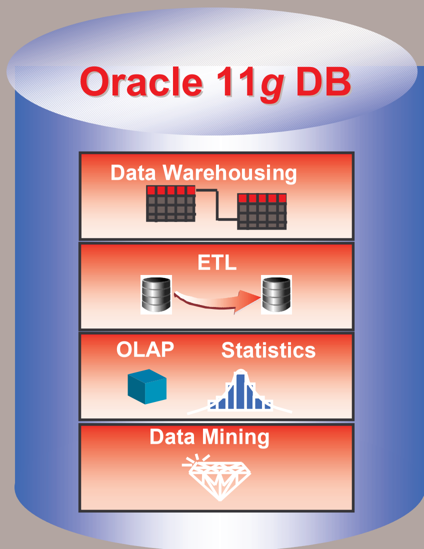


Oracle Data Miner (gui)

Code Generation

- PL/SQL code generation for Mining Activities





In-Database Analytics SQL Examples

Example #2

Better Insights & Information

- Select all customers who have a high propensity to attrite (> 80% chance) and have a customer value rating of more than 90 and have had a recent conversation with customer service regarding a Checking Plus account.

```
SELECT A.cust_name, A.contact_info
  FROM customers A
 WHERE PREDICTION_PROBABILITY(tree_model,
            'attrite' USING A.*) > 0.8
   AND A.cust_value > 90
   AND A.cust_id IN
       (SELECT B.cust_id
          FROM call_center@HQ_DB B
         WHERE B.call_date BETWEEN '01-Jan-2005'
                               AND '30-Jun-2005'
         AND CONTAINS(B.notes,'Checking Plus', 1)> 0);
```

Real-time Prediction

```
with
 records as (select
     178255 ANNUAL_INCOME,
      0          CAPITAL_GAIN,
      83        SAVINGS_BALANCE,
      246      AVE_CHECKING_BALANCE,
      30        AGE,
      'Bach.'  EDUCATION,
      'SelfENI' WORKCLASS,
      'Married' MARITAL_STATUS,
      'Sales'    OCCUPATION,
      'Husband' RELATIONSHIP,
      'White'   RACE,
      'Male' SEX,
      70 HOURS_PER_WEEK,
      '?' NATIVE_COUNTRY,
      98 PAYROLL_DEDUCTION from dual)
 select s.prediction prediction, s.probability probability
 from (
   select PREDICTION_SET(CD_BUYERS76485_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)

Real-time Prediction Multiple Models

```
with records as (select
    178255 ANNUAL_INCOME,
        0 CAPITAL_GAIN,
        83 SAVINGS_BALANCE,
        246 AVE_CHECKING_BALANCE,
    30 AGE,
        'Bach.' EDUCATION,
        'SelfENI' WORKCLASS,
        'Married' MARITAL_STATUS,
        'Sales' OCCUPATION,
        'Husband' RELATIONSHIP,
        'White' RACE,
        'Male' SEX,
        70 HOURS_PER_WEEK,
        '?' NATIVE_COUNTRY,
        98 PAYROLL_DEDUCTION from dual)
    select t.*
    from (
      select 'CAR_MODEL' MODEL, s1.prediction prediction, s1.probability probability,
s1.probability*25000 as expected_revenue from (
        select PREDICTION_SET(NBMODEL_JDM, 1  USING *) pset
        from records ) t1, TABLE(t1.pset) s1
      UNION
      select 'MOTOCYCLE_MODEL' MODEL, s2.prediction prediction, s2.probability probability,
s1.probability*2000 as expected_revenue from (
        select PREDICTION_SET(ABNMODEL_JDM, 1  USING *) pset
        from records ) t2, TABLE(t2.pset) s2
      UNION
      select 'TRICYCLE_MODEL' MODEL, s3.prediction prediction, s3.probability probability,
s1.probability*50 as expected_revenue from (
        select PREDICTION_SET(TREEMODEL_JDM, 1  USING *) pset
        from records ) t3, TABLE(t3.pset) s3
      UNION
      select 'BICYCLE_MODEL' MODEL, s4.prediction prediction, s4.probability probability,
s1.probability*200 as expected_revenue from (
        select PREDICTION_SET(SVMCMODEL_JDM, 1  USING *) pset
        from records ) t4, TABLE(t4.pset) s4
      )  t
      order by t.expected_revenue desc;
```

On-the-fly, single record apply with multiple models; sort by expected revenues

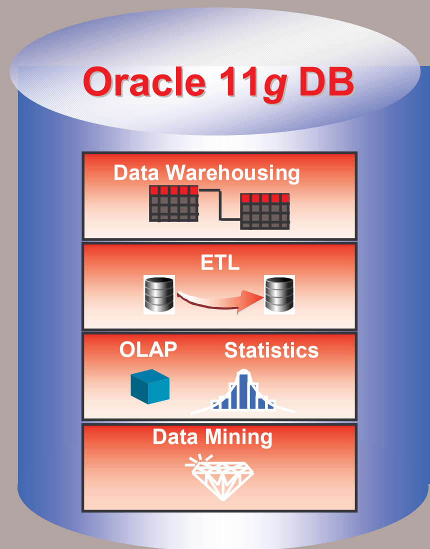
ORACLE

Example #3

Launch & Evaluate a Marketing Campaign

1. Given a previously built response model,…predict who will respond to a campaign, …and why
2. …find out how much each customer spent 3 months before and after the campaign
3. …how much for just DVDs?
4. Is the success statistically significant?

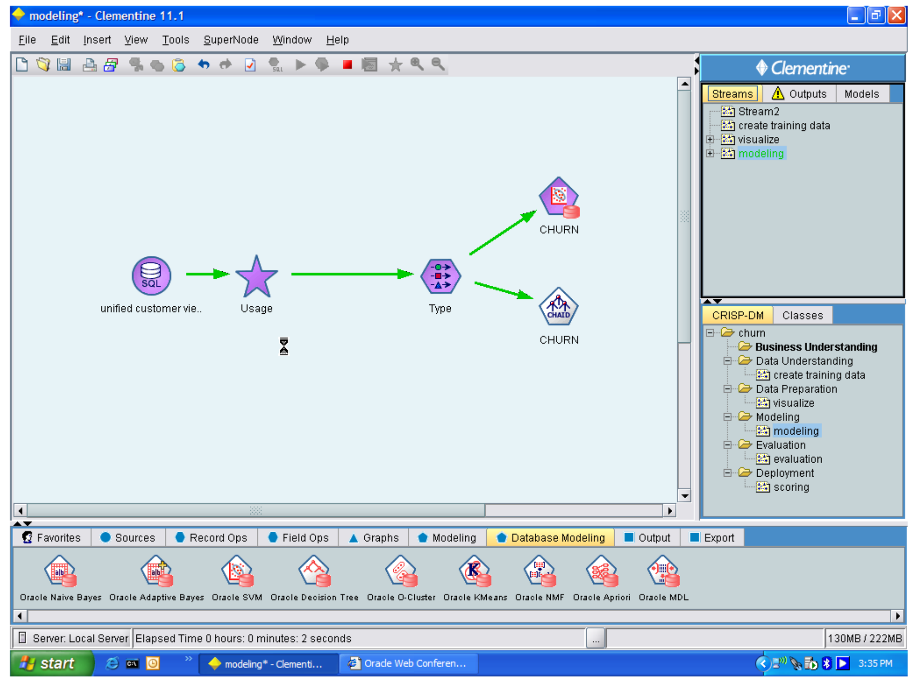
```
select responder, cust_region, count(*) as cnt,
       sum(post_purch - pre_purch) as tot_increase,
       avg(post_purch - pre_purch) as avg_increase,
       stats_t_test_paired(pre_purch, post_purch) as
  significance
from (
 select cust_name,
        prediction(campaign_model using *) as responder,
        sum(case when purchase_date < 15-Apr-2005 then
           purchase_amt else 0 end) as pre_purch,
        sum(case when purchase_date >= 15-Apr-2005 then
           purchase_amt else 0 end) as post_purch
 from customers, sales, products@PRODDB
 where sales.cust_id = customers.cust_id
   and purchase_date between 15-Jan-2005 and 14-Jul-2005
   and sales.prod_id = products.prod_id
   and contains(prod_description, 'DVD') > 0
 group by cust_id, prediction(campaign_model using *) )
group by rollup responder, cust_region order by 4 desc;
```


Oracle 11g DB
Data Warehousing
ETL
OLAP
Statistics
Data Mining

Partners

SPSS Clementine

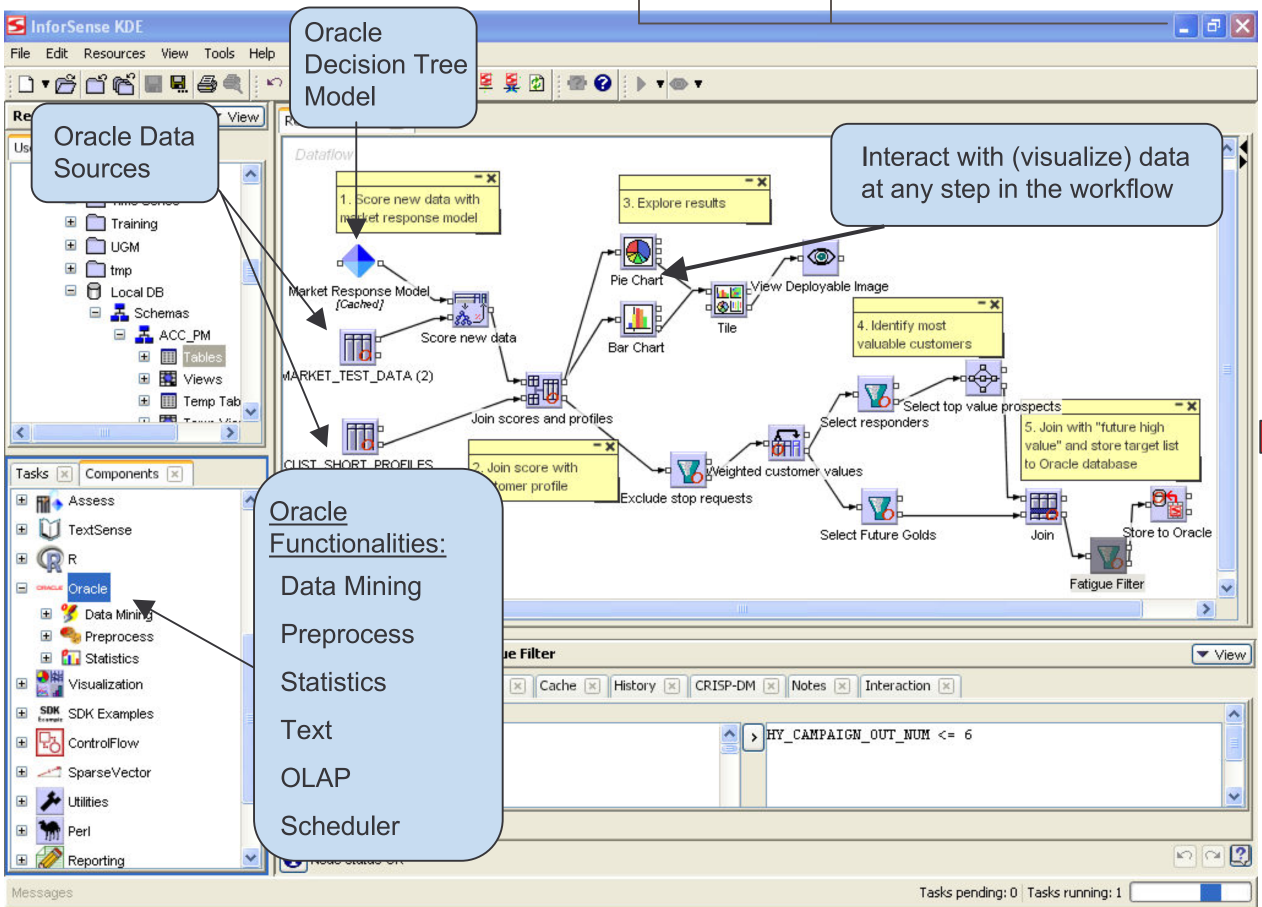
- 35+ year heritage in analytic technologies with operations in over 60 countries
- Industry Dominance in DM
  - Banking
  - Insurance
  - Telco
  - Retail
  - Public Sector
- Combine SPSS Clementine ease of use with ODM in-Database functionality & scalability
  - Build, store, browse and score models in the Database for optimal performance



- For more information :
  - SPSS – Mike Bittner, Strategic Alliance Manager, 770.329.3870 or mbittner@spss.com
  - Oracle – Alan Manewitz, TBU, (925) 984-9910 or alan.manewitz@oracle.com
  - Oracle – Charlie Berger, Product Management (781) 744-0324 or charlie.berger@oracle.com

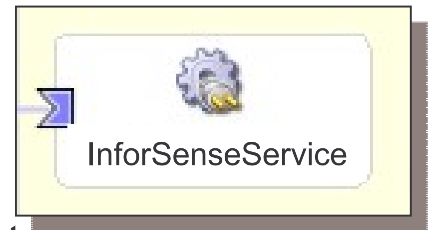
InforSense -- A Single Optimized Environment for Real Time Business Analytics within the Database

InforSense
Embedding Intelligence Throughout the Enterprise

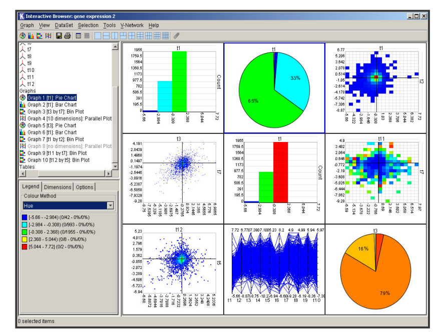


Deployment

Deploy the analytic workflow as a service embedding to BPEL, SFA, CRM



Deploy the analytic workflow as an Oracle Portal



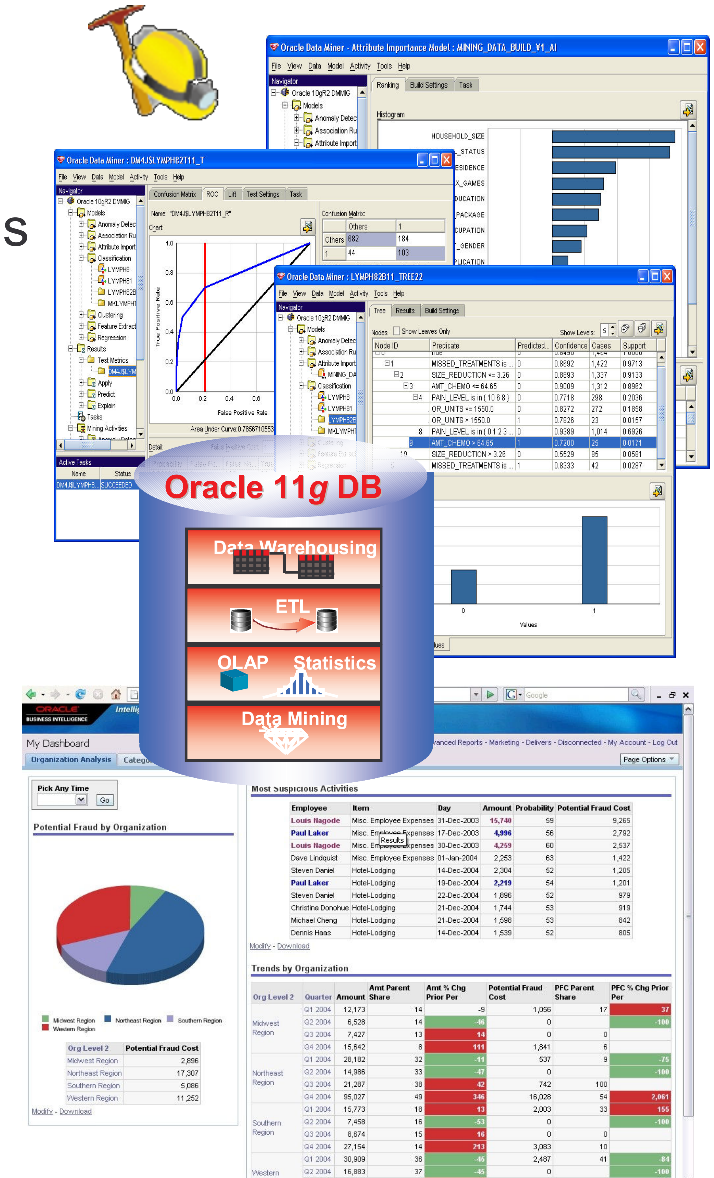
SAS free analytics: leverage Oracle analytics
SQL free analytics: drag-drop application build
Visual analytics: interactive visualisation

Integrative analytics: unified analytical environment
Automated analytics: deploy to Oracle Portal and BPEL

ORACLE

Predictive Analytics with Oracle Data Mining & Oracle BI EE

- **Know More!**
  - Wide range of data mining algorithms & (free) statistical functions
  - Mine data, discover insights, make predictions
- **Do More!**
  - Mine for new information —deploy using Oracle BI EE
  - Build applications that help you *“Compete on Analytics”*
- **Spend Less!**
  - Eliminate redundant data management and data analysis systems
  - Leverage investment in Oracle Database Technology



BIWA Summit 2008
Better Information—Better Results

- **Dec. 2-3, 2008** in San Francisco at Oracle HQ Conference Center
- 2nd BIWA Summit—presentations available at www.olsug.org)
- Currently FREE to join BIWA
- IOUG SIG for users who share expertise and vision dedicated to mutual success in leveraging:
  - Oracle Database-centric BI, Warehousing and Analytics
- BIWA's goals include the sharing of *best practices* and *novel and interesting use cases*

CONFERENCE ANNOUNCEMENT AND CALL FOR SPEAKER PROPOSALS

OLSUG-2008 INTERNATIONAL ORACLE LIFE SCIENCES AND HEALTHCARE USER GROUP MEETING

April 27-28, 2008 | World Trade Center | Boston, MA

Venue: World Trade Center, Boston, MA, USA
in conjunction with Bio-IT 2008

Submit your proposal today!

Submit Speaker Proposal

Oracle Life Science User Group

The Oracle Life Sciences User Group (OLSUG) represents a worldwide association of people who share a vested interest in the successful deployment and use of Oracle products within the life sciences and healthcare industry. The OLSUG-2008 International OLSUG meeting is designed to accelerate the success of users of Oracle products by providing a forum for sharing useful information and best practices, discovering novel and innovative practices, and exposing issues and limitations concerning Oracle products.

Important Dates

Last date for submission of abstracts: November 16, 2007

Presentation notification: December 7, 2007

International OLSUG2008 Meeting: April 27-28, 2008

Topics of Interest as Related to Life Science and Healtcare IT

Distributed Data Access: Gateways, External Tables & Generic Connectivity.
Data Integration: Relational and XML.
Semantic Web: Use of Oracle's RDF enabling data integration and knowledge representation
Healthcare IT: Translational research (Benchside to Bedside)
Healthcare IT: Molecular profiling (Bedside to Benchside)
Healthcare IT: Evidence-based practice
Supporting Many Data Types: Text, Images, XML and RDF.
Managing Lots of Data: Partitioning, RAC, ASM, Streams, Data Guard and Grid.
Enabling Secure Collaboration: Collaboration Suite, Portal and Security.
Finding Patterns and Insights: Data Mining, Statistics and Text Mining.
Application Development: Fusion Middleware, JDeveloper and Application Express.
Database Performance: Tuning tips

ORACLE

More Information:

Oracle Data Mining 11g

- oracle.com/technology/products/bi/odm/index.html

Oracle Statistical Functions

- http://www.oracle.com/technology/products/bi/stats_fns/index.html

Oracle Business Intelligence Solutions

- oracle.com/bi

http://search.oracle.com

oracle data mining

Contact Information: Email: Charlie.berger@oracle.com

ORACLE IS THE INFORMATION COMPANY

Q&A

QUESTIONS
ANSWERS

ORACLE

The preceding is intended to outline our general product direction. It is intended for information purposes only, and may not be incorporated into any contract. It is not a commitment to deliver any material, code, or functionality, and should not be relied upon in making purchasing decisions.
The development, release, and timing of any features or functionality described for Oracle's products remains at the sole discretion of Oracle.