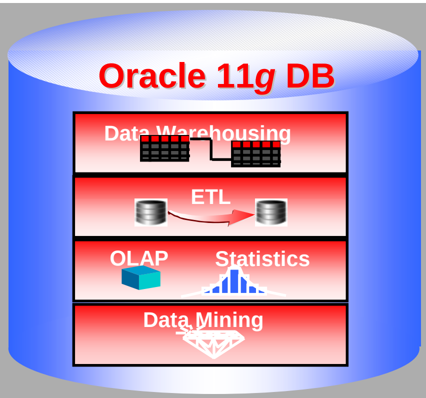


Oracle Data Mining 11g Release 2 Overview and Demo

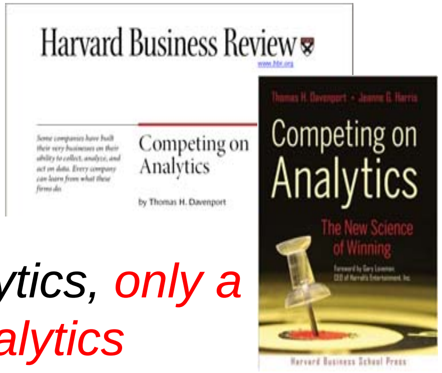
Charlie Berger
Sr. Director Product Management, Data Mining Technologies
Oracle Corporation
charlie.berger@oracle.com

The following 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.

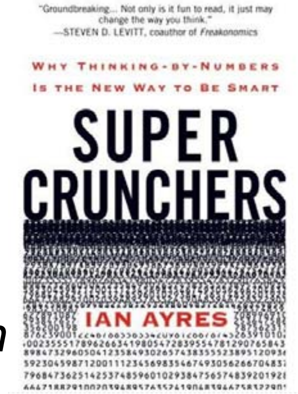
Outline

- Today's BI must go beyond simple reporting
- To succeed, companies must
  - Eliminate data movement
  - Collapse information latency
  - Deliver better BI through analytics
- ODM makes the Database an "Analytical Database"
  - Enables applications *"Powered by Oracle Data Mining"*
- Brief demonstrations
  1. Oracle Data Mining
  2. OBI EE Dashboards with ODM Results
  3. Oracle Sales Prospector with embedded ODM

Analytics: Strategic and Mission Critical

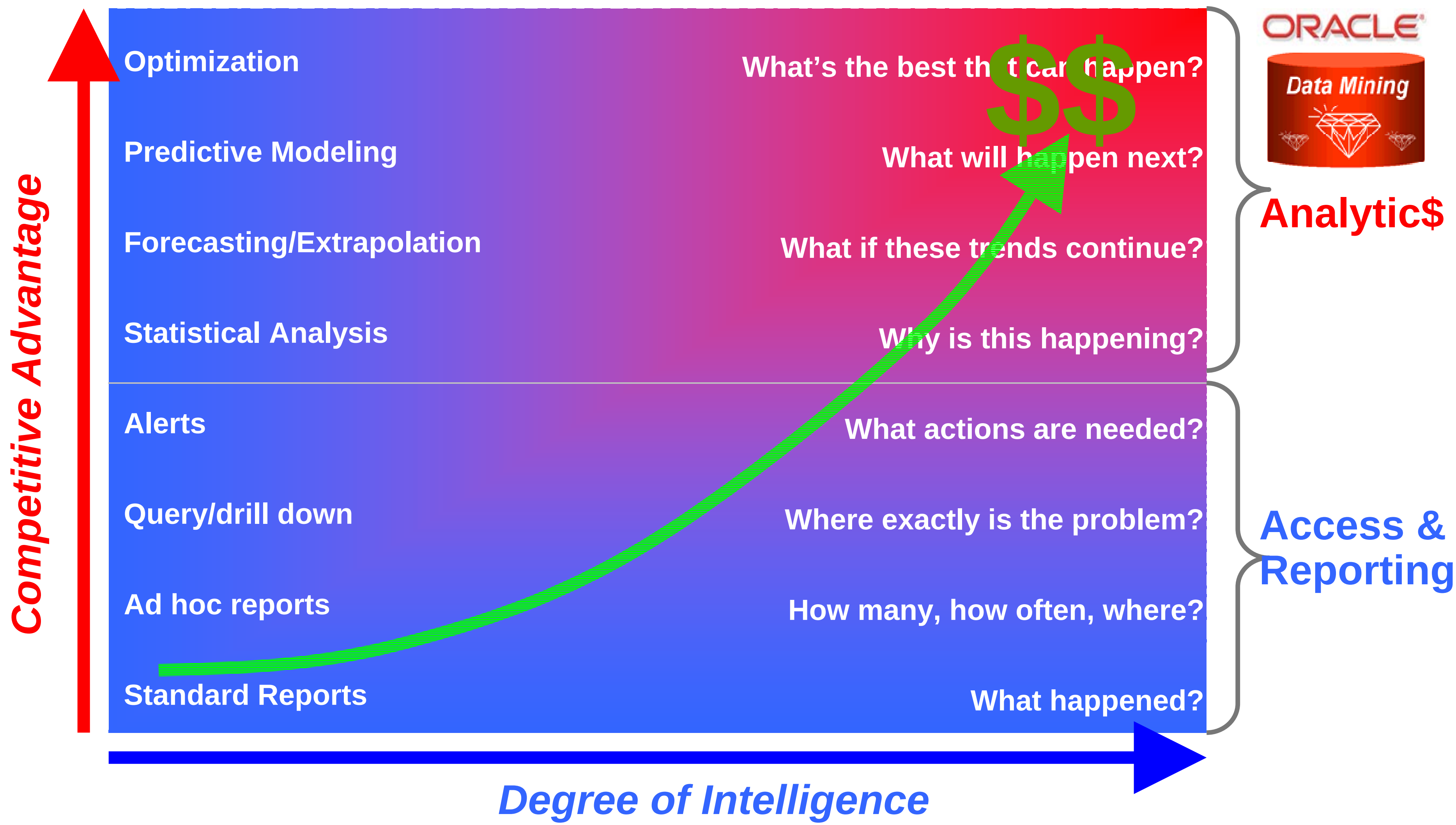


- *Competing on Analytics*, *by Tom Davenport*
  - *"Some companies have built their very businesses on their ability to collect, analyze, and act on data."*
  - *"Although numerous organizations are embracing analytics, only a handful have achieved this level of proficiency. But analytics competitors 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*



ORACLE

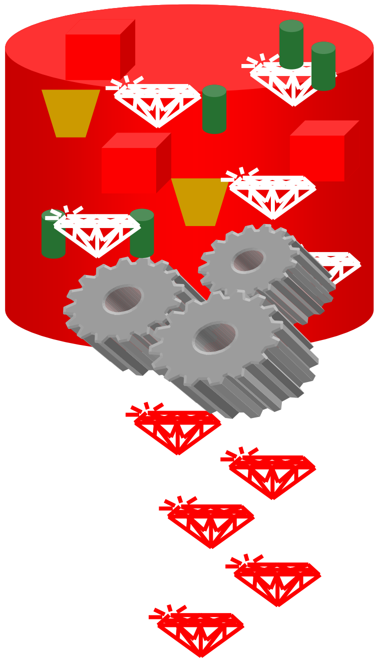
Competitive Advantage



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

Oracle Data Mining Option

What is Data Mining?



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

Oracle Data Mining Example Use Cases

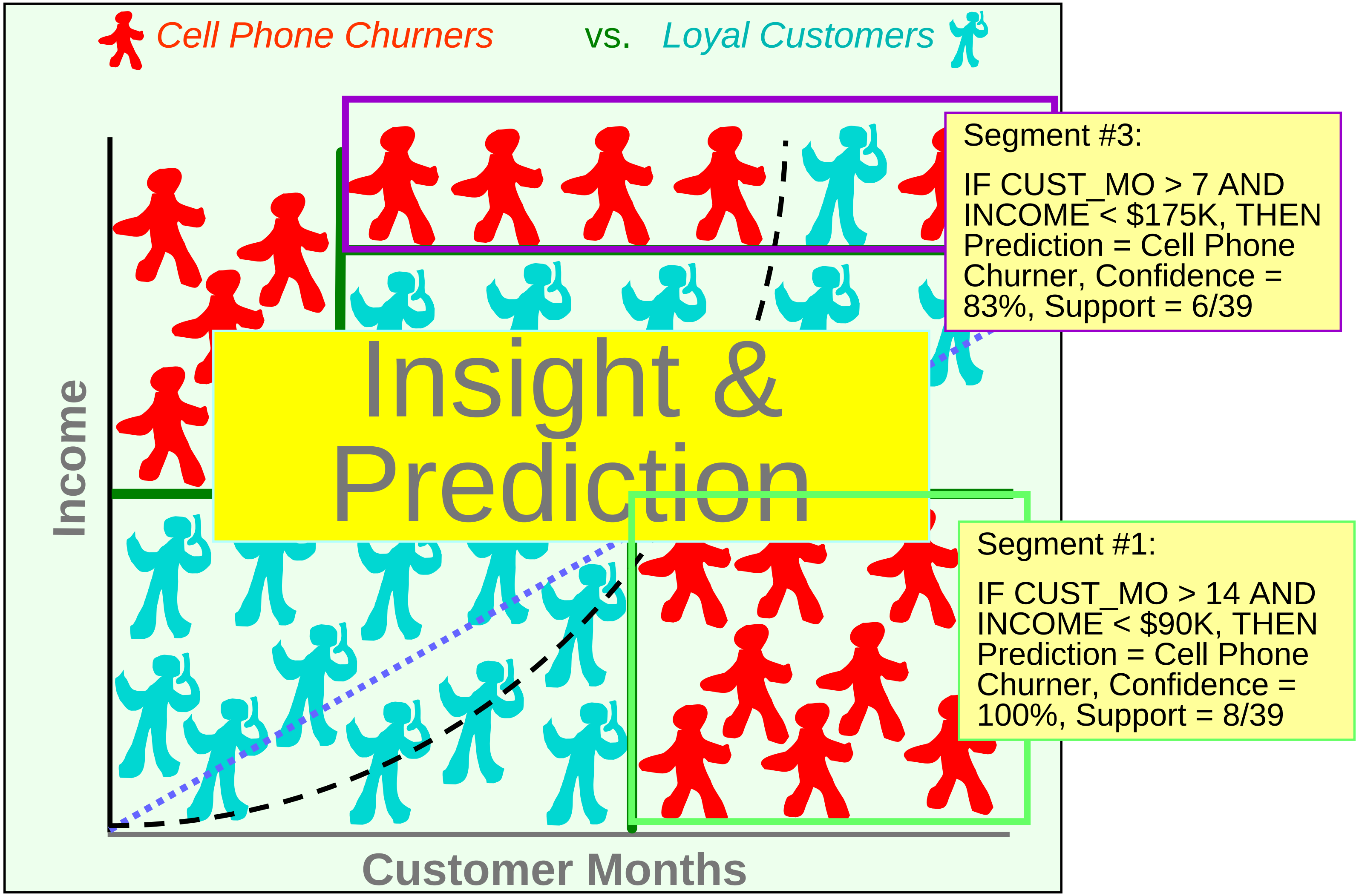
- **Retail**
  - Customer segmentation
  - Response modeling
  - Recommend next likely product
  - Profile high value customers
- **Banking**
  - Credit scoring
  - Probability of default
  - Customer profitability
  - Customer targeting
- **Insurance**
  - Risk factor identification
  - Claims fraud
  - Policy bundling
  - Employee retention
- **Higher Education**
  - Alumni donations
  - Student acquisition
  - Student retention
  - At-risk student identification

- **Healthcare**
  - Patient procedure recommendation
  - Patient outcome prediction
  - Fraud detection
  - Doctor & nurse note analysis
- **Life Sciences**
  - Drug discovery & interaction
  - Common factors in (un)healthy patients
  - Cancer cell classification
  - Drug safety surveillance
- **Telecommunications**
  - Customer churn
  - Identify cross-sell opportunities
  - Network intrusion detection
- **Public Sector**
  - Taxation fraud & anomalies
  - Crime analysis
  - Pattern recognition in military surveillance

- **Manufacturing**
  - Root cause analysis of defects
  - Warranty analysis
  - Reliability analysis
  - Yield analysis
- **Automotive**
  - Feature bundling for customer segments
  - Supplier quality analysis
  - Problem diagnosis
- **Chemical**
  - New compound discovery
  - Molecule clustering
  - Product yield analysis
- **Utilities**
  - Predict power line / equipment failure
  - Product bundling
  - Consumer fraud detection

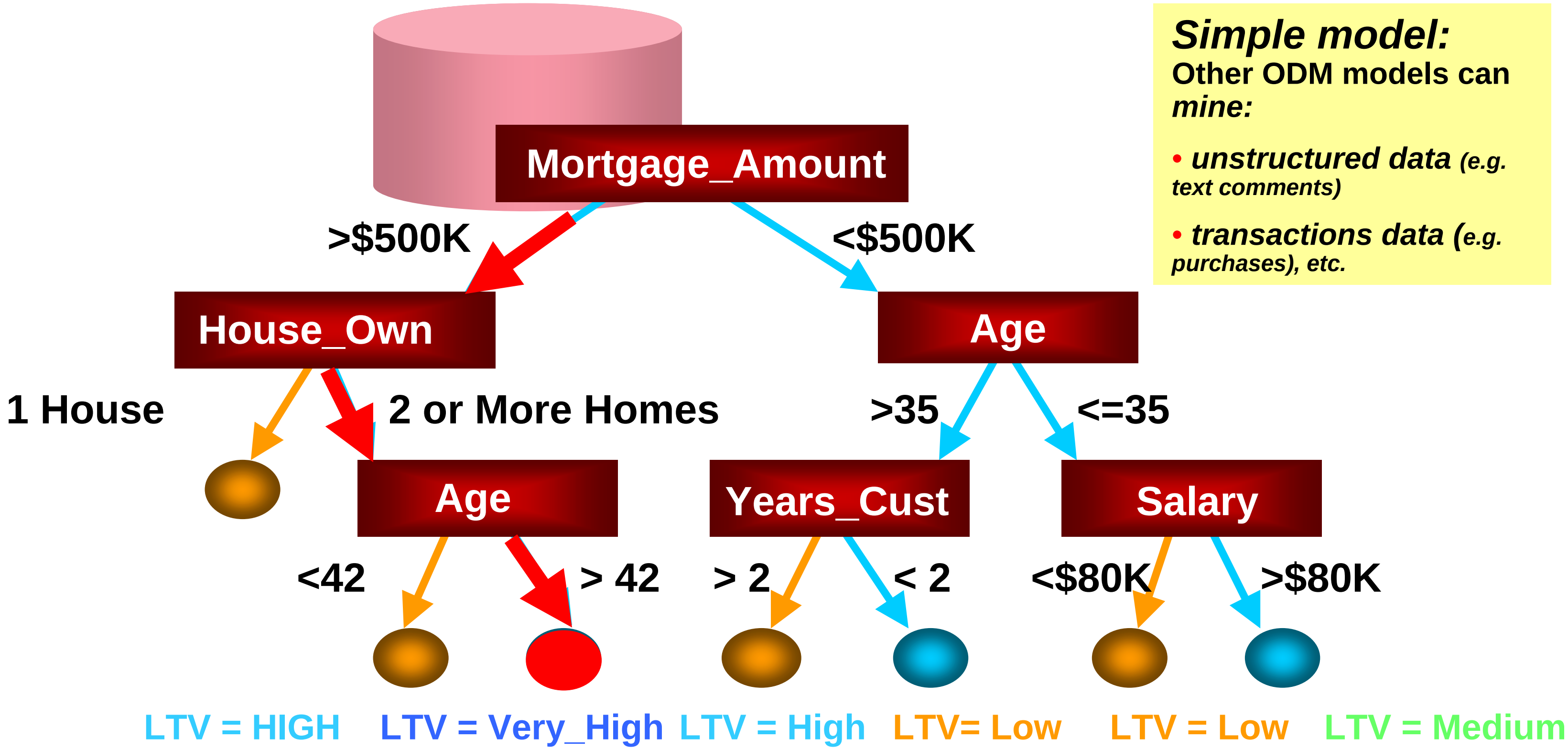
Data Mining Provides

Better Information, Valuable Insights and Predictions



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

Predicting High LTV Customers Using a Decision Tree Model



Simple model:
Other ODM models can *mine:*

- ***unstructured data** (e.g. text comments)*
- ***transactions data** (e.g. purchases), etc.*

IF (Mortgage_Amount > $500K AND House_Own = 2 or more AND Age = >42)
THEN Probability(Lifetime Customer Value is "VERY HIGH" = 77%, Support = 15%

“Essentially, all models are wrong, but some are useful.”

- George Box

(one of the most influential statisticians of the 20th century and a pioneer in the areas of quality control, time series analysis, design of experiments and Bayesian inference.)

Oracle Data Mining

Overview (Classification)

Model

Input Attributes

Target

Cases

Historic Data

Name	Income	Age	. . . .		Respond? 1 =Yes, 0 =No		
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:

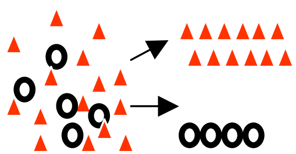
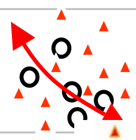
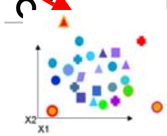
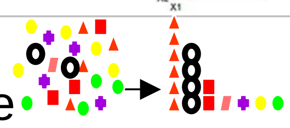
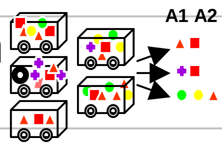
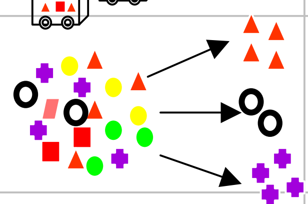
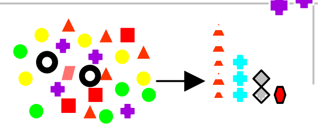
$$Y = F(X_1, X_2, \ldots, X_m)$$

Prediction Confidence

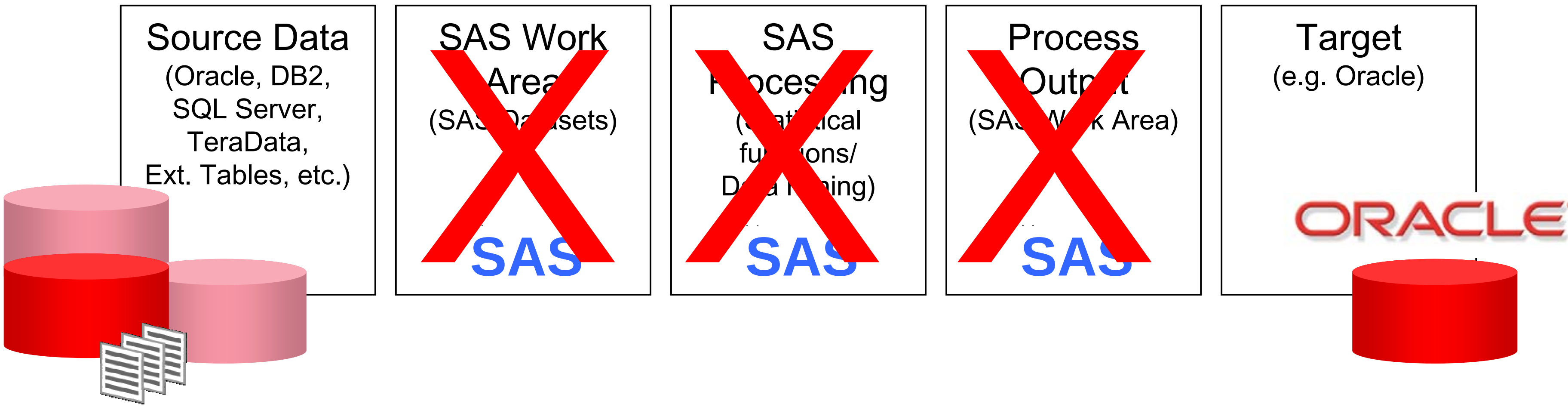
ORACLE DATABASE 11g

Oracle Data Mining

Algorithm Summary 11g

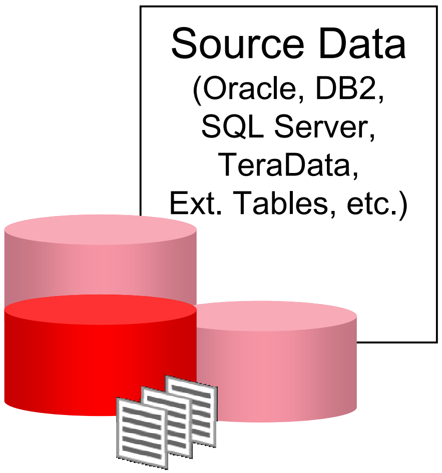
Problem	Algorithm	Applicability
Classification 	**Logistic Regression (GLM)** **Decision Trees** **Naïve Bayes** **Support Vector Machine**	**Classical statistical technique** **Popular / Rules / transparency** **Embedded app** **Wide / narrow data / text**
Regression 	**Multiple Regression (GLM)** **Support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

Traditional Analytics (SAS) Environment



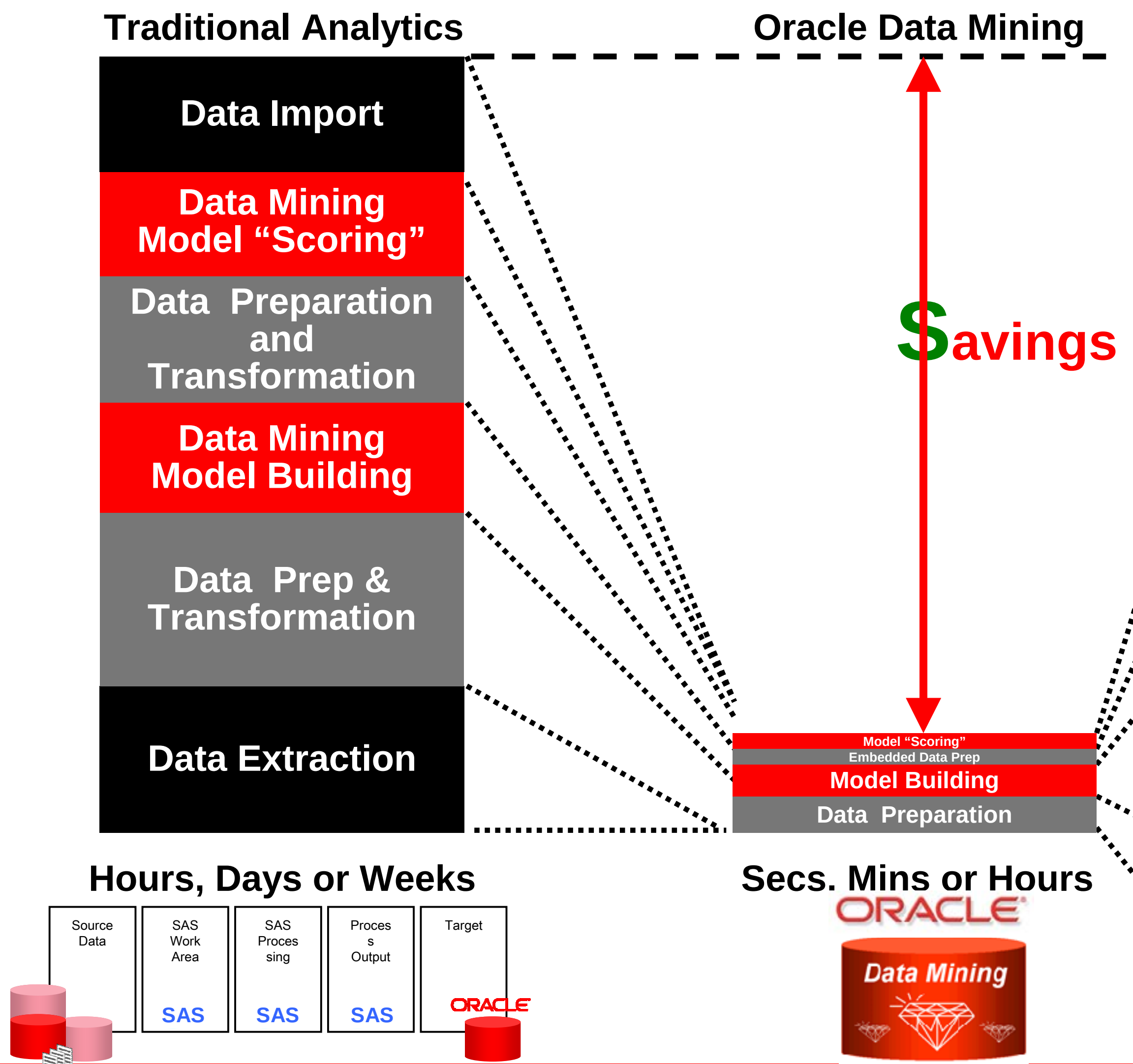
- SAS environment requires:
  - Data movement
  - Data duplication
  - Loss of security

Oracle Architecture



- Oracle environment:
  - Eliminates data movement
  - Eliminates data duplication
  - Preserves security

In-Database Data Mining



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

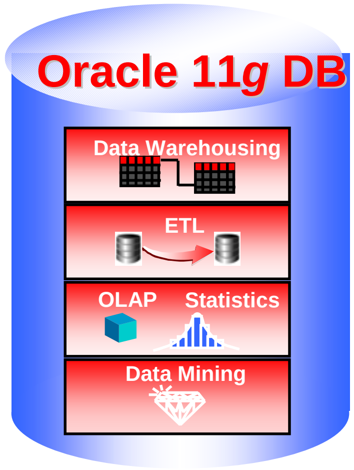
Cutting edge machine learning algorithms inside the SQL kernel of Database

SQL—Most powerful language for data preparation and transformation

Data remains in the Database

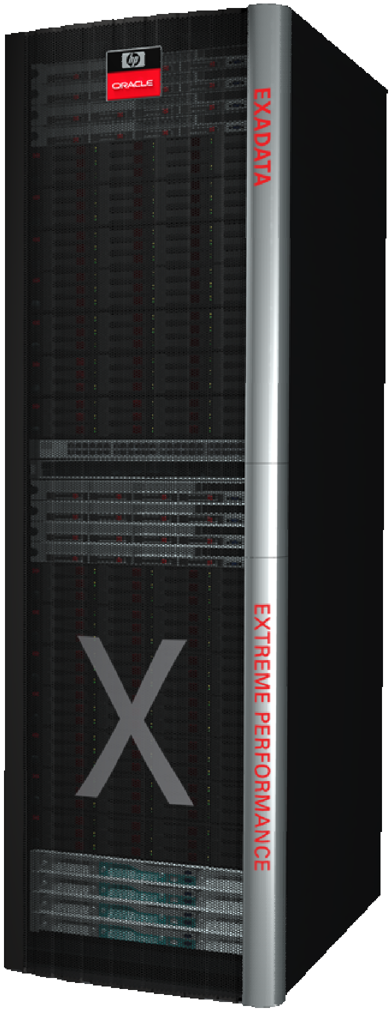
In-Database Data Mining

Advantages



- ODM architecture provides greater
  - Performance, scalability, and data security
- Data remains in the database
  - Fewer moving parts; shorter information latency
- Straightforward inclusion within interesting and arbitrarily complex queries
  - "SELECT Customers WHERE Income > 100K, AND Probability(Buy Product A) > .85;"
- 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

HP Oracle Database Machine & ODM



- Integrated data warehouse solution
- Extreme Performance
  - 10-100X faster than conventional DW systems
- Scalability to Petabytes
- Enterprise-Ready
  - Complete data warehouse functionality
  - Enterprise-level availability and security
- Scoring of Oracle Data Mining models
  - Blazingly fast performance
  - 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;
```

HP Oracle Database Machine & ODM

- 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;
```

ODM 11gR2 Scoring: Offloaded to Exadata

- Data mining scoring executed in Exadata:

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

Scoring function executed in Exadata

- All scoring functions offloaded to Exadata
- Benefits
  - Reduces data returned from Exadata to Database server
  - Reduces CPU utilization on Database Server
  - Up to 10x performance gains

“If I had one hour to save the world, I would spend fifty-five minutes defining the problem and only five minutes finding the solution”

- Albert Einstein

(see also http://www.wikihow.com/Define-a-Problem)

Where to Start?

"Wrong: Catalog everything you have, and decide what data is important.

Right: Work backward from the solution, define the problem explicitly, and map out the data needed to populate the investigation and models."

- James Taylor with Neil Raden, authors,
Smart (Enough) Systems

Oracle Data Mining and Unstructured Data

- Oracle Data Mining mines unstructured i.e. “text” data
- Include free text and comments in ODM models
- Cluster and Classify documents
- Oracle Text used to preprocess unstructured text

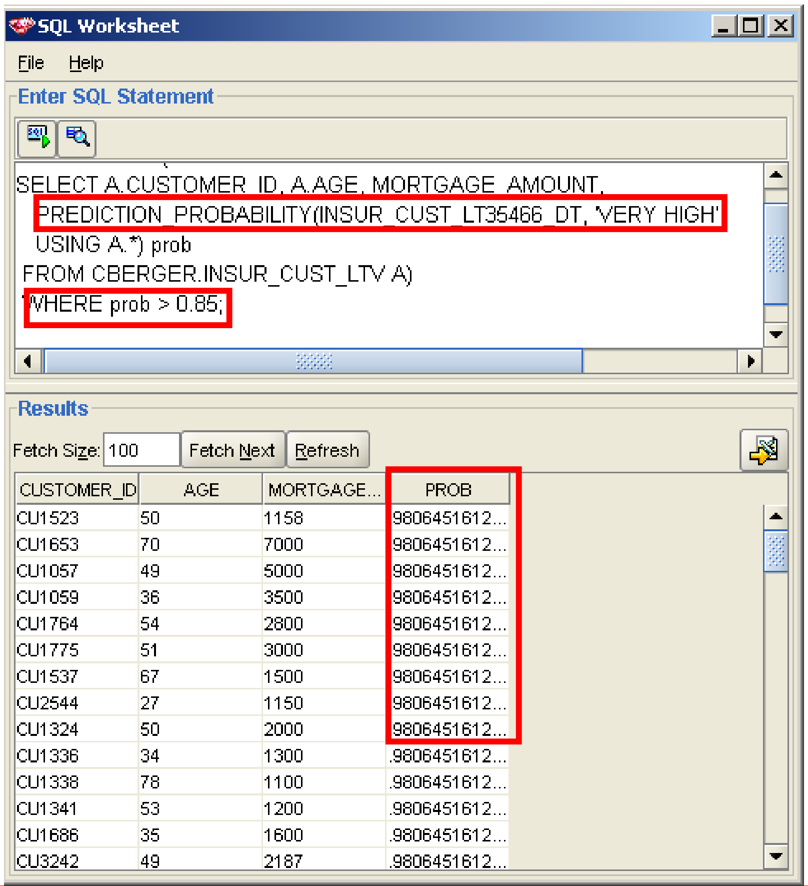
Structure | Data

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

Example: Simple, Predictive SQL

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




```
SELECT * from(
SELECT A.CUSTOMER_ID, A.AGE,
  MORTGAGE_AMOUNT,PREDICTION_PROBABILITY
  (INSUR_CUST_LT13126_DT, 'VERY HIGH'
   USING A.*) prob
 FROM CBERGER.INSUR_CUST_LTV A)
 WHERE prob > 0.85;
```

Fraud Prediction Demo

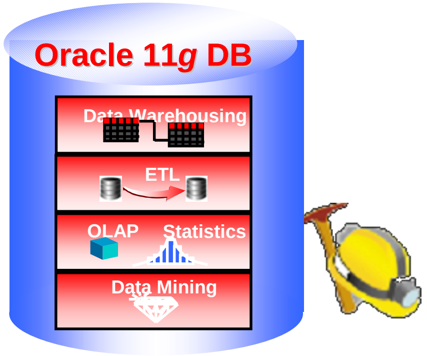
```
drop table CLAIMS_SET;
exec dbms_data_mining.drop_model('CLAIMSMODEL');
create table CLAIMS_SET (setting_name varchar2(30), setting_value varchar2(4000));
insert into CLAIMS_SET values
('ALGO_NAME','ALGO_SUPPORT_VECTOR_MACHINES');
insert into CLAIMS_SET values ('PREP_AUTO','ON');
commit;
```

```
begin
dbms_data_mining.create_model('CLAIMSMODEL', 'CLASSIFICATION',
  'CLAIMS', 'POLICYNUMBER', null, 'CLAIMS_SET');
end;
/
```

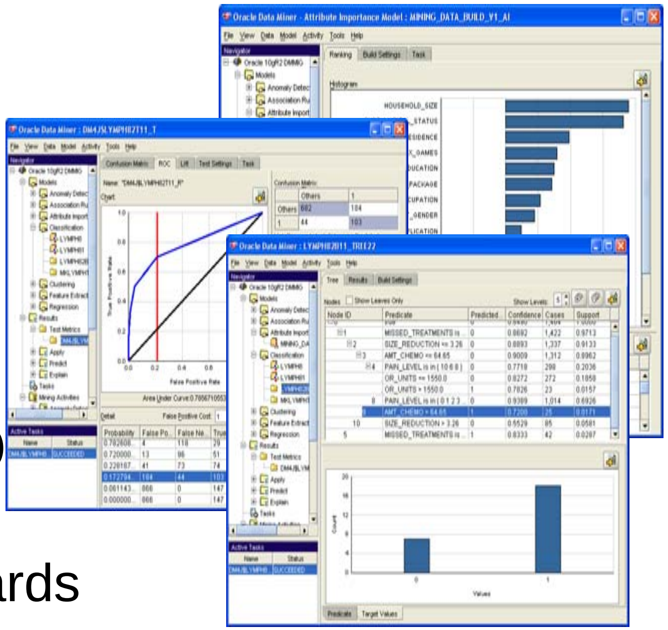
```
-- Top 5 most suspicious fraud policy holder claims
select * from
(select POLICYNUMBER, round(prob_fraud*100,2) percent_fraud,
     rank() over (order by prob_fraud desc) rnk from
(select POLICYNUMBER, prediction_probability(CLAIMSMODEL, '0' using *) prob_fraud
from CLAIMS
where PASTNUMBEROFCLAIMS 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

Oracle Data Mining 11g



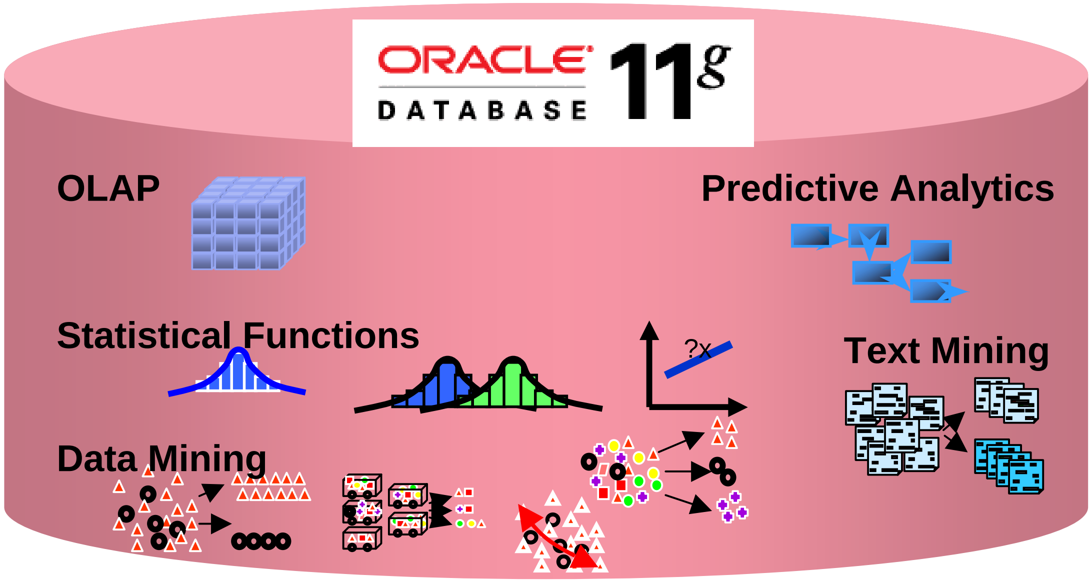
- Data Mining Functions (Server)
  - PL/SQL & Java APIs
  - Develop & deploy predictive analytics applications
- Wide range of DM algorithms (12)
  - Classification & regression
  - Clustering
  - Anomaly detection
  - Attribute importance
  - Feature extraction (NMF)
  - Association rules (Market Basket 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



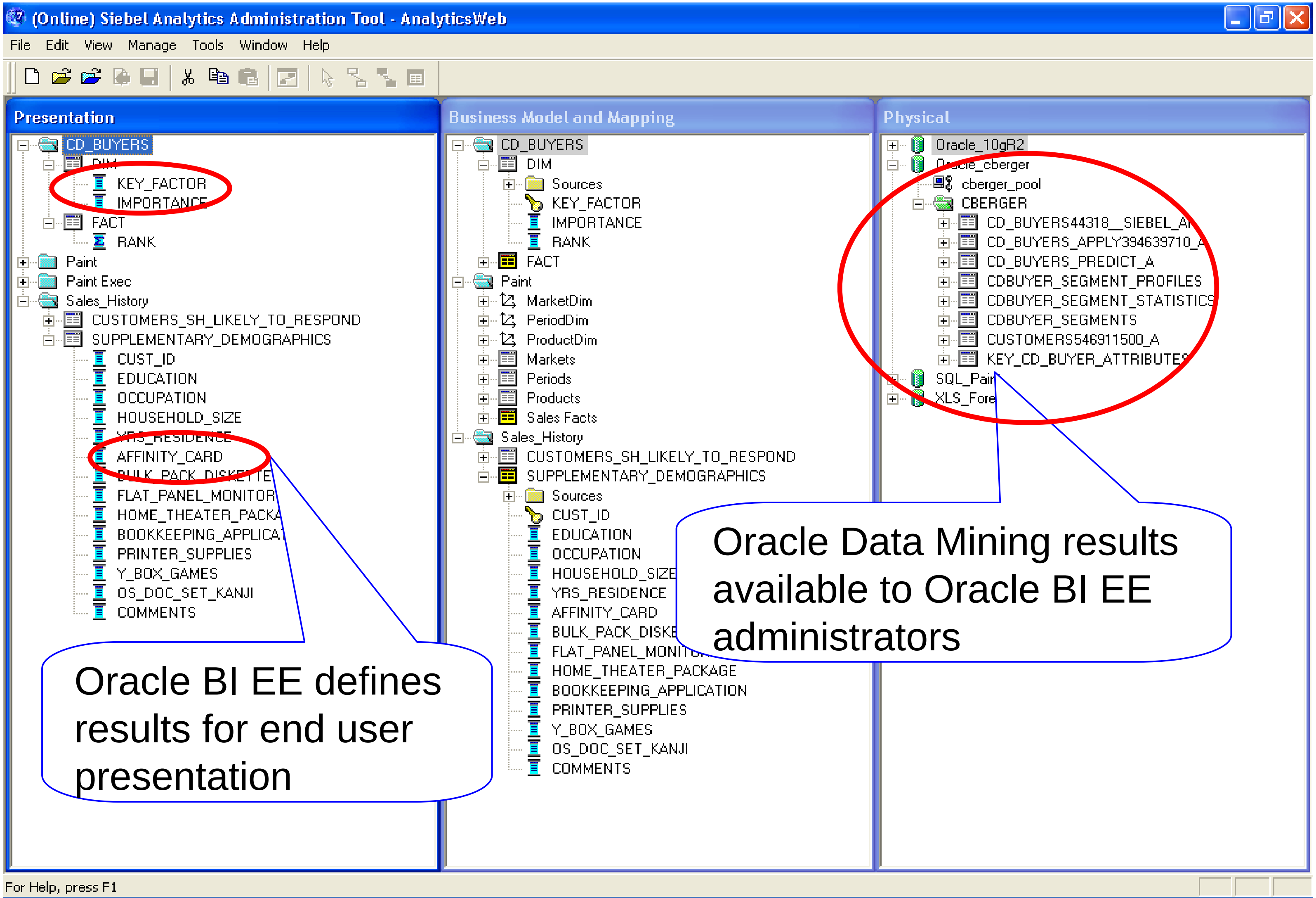
Analytical Database Changes**Everything**

It boils down to this:

Less data movement = **faster** analytics, and faster analytics = **better** BI throughout the enterprise

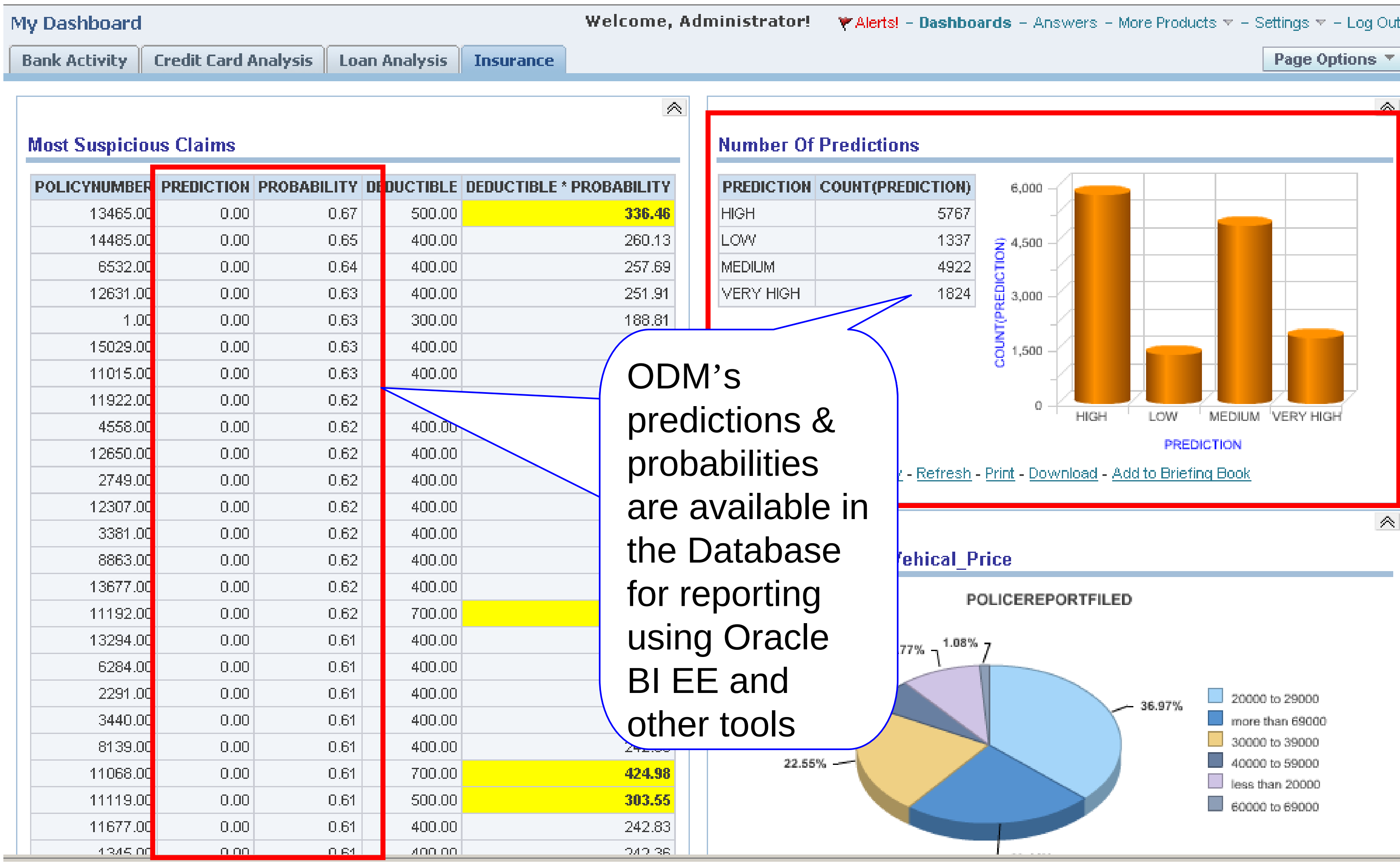


Integration with Oracle BI EE



Example

Better Information for OBI EE Reports and Dashboards

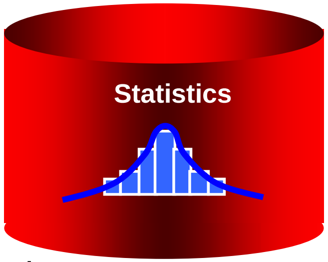


ODM's predictions & probabilities are available in the Database for reporting using Oracle BI EE and other tools

Oracle SQL Statistical Functions

(Free in Every Oracle Database)

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, 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

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

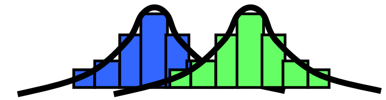
Descriptive Statistics

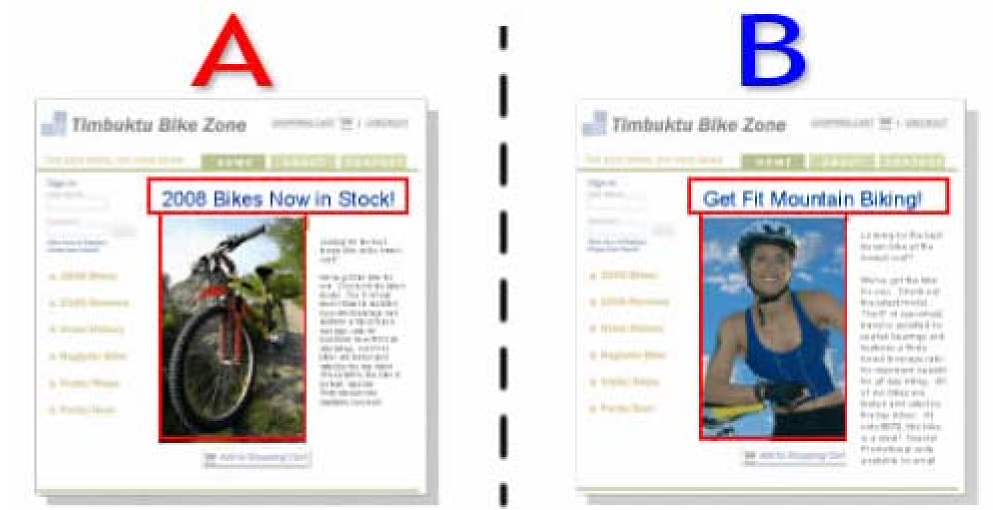
- MEDIAN & MODE
  - Median: takes numeric or datetype values and returns the middle value
  - Mode: returns the most common value

> SQL

```
A. SELECT STATS_MODE(AGE) from LYMPHOMA;

B. SELECT MEDIAN(AGE) from LYMPHOMA;

C. SELECT TREATMENT_PLAN, STATS_MODE(LYMPH_TYPE)
   from lymphoma GROUP BY TREATMENT_PLAN;

D. SELECT LYMPH_TYPE, MEDIAN(SIZE_REDUCTION) from
   LYMPHOMA GROUP BY LYMPH_TYPE ORDER BY
   MEDIAN(SIZE_REDUCTION) ASC;
```

Split Lot A/B Offer testing

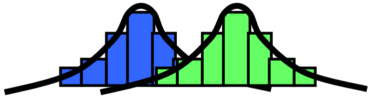




- Offer "A" to one population and "B" to another
- Over time period "t" calculate median purchase amounts of customers receiving offer A & B
- Perform t-test to compare
- If statistically significantly better results achieved from one offer over another, offer everyone higher performing offer

Independent Samples T-Test

(Pooled Variances)

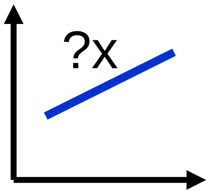


- Query compares the mean of AMOUNT_SOLD between MEN and WOMEN within CUST_INCOME_LEVEL ranges

```
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

Correlation Functions



- The CORR_S and CORR_K functions support nonparametric or rank correlation (finding correlations between expressions that are ordinal scaled).
- Correlation coefficients take on a value ranging from –1 to 1, where:
  - 1 indicates a perfect relationship
  - –1 indicates a perfect inverse relationship
  - 0 indicates no relationship
- The following query determines whether there is a correlation between the AGE and WEIGHT of people, using Spearman's correlation:

```
select CORR_S(AGE, WEIGHT)
    coefficient,
    CORR_S(AGE, WEIGHT,
    'TWO_SIDED_SIG')
    p_value,
substr(TREATMENT_PLAN, 1,15)
as TREATMENT_PLAN
from CBERGER.LYMPHOMA
GROUP BY TREATMENT_PLAN;
```

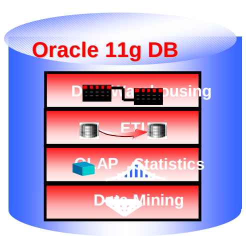
COEFFICIENT	P_VALUE	TREATMENT_PLAN
.1862586290028...	.019908367365...	Chemo&Radiation
-.0575579915035...	.072279268481...	Chemo_only
-.0746488538574...	.288631463930...	Experimental
-.1254971583227...	.000018140526...	Radiation

ORACLE Analytics vs. SAS

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

2. Costs (ODM: **$23K cpu**)
   - Simplified environment
   - Single server
   - Security

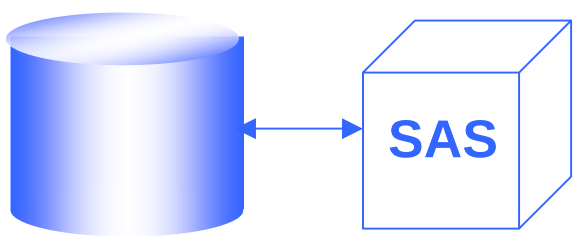
3. IT Platform
   - SQL (standard)
   - Java (standard)



1. ***External*** Analytical Engine
   - Basic Statistics
   - Data Mining
   - Text Mining (separate: SAS EM for Text)
   - *Advanced* Statistics

2. Costs (SAS EM: **$150K/5 users**)
   - Duplicates data
   - Annual Renewal Fee (AUF)
     (~45% each year)

3. IT Platform
   - SAS Code (proprietary)

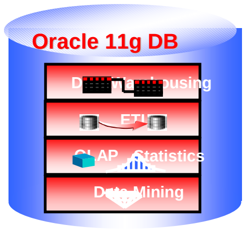


ORACLE Analytics vs. SAS

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

2. Costs (ODM: **$23K cpu**)
   - Simplified environment
   - Single server
   - Security

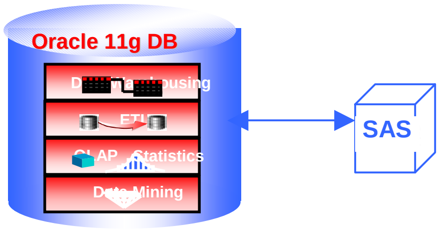
3. IT Platform
   - SQL (standard)
   - Java (standard)



1. ***External*** Analytical Engine
   - Basic Statistics
   - Data Mining
   - Text Mining (separate: SAS EM for Text)
   - *Advanced* Statistics

2. Costs (SAS EM: **$150K/5 users**)
   - Duplicates data
   - Annual Renewal Fee (AUF)
     (~45% each year)

3. IT Platform
   - SAS Code (proprietary)



SAS In-Database Processing

3-Year Road Map

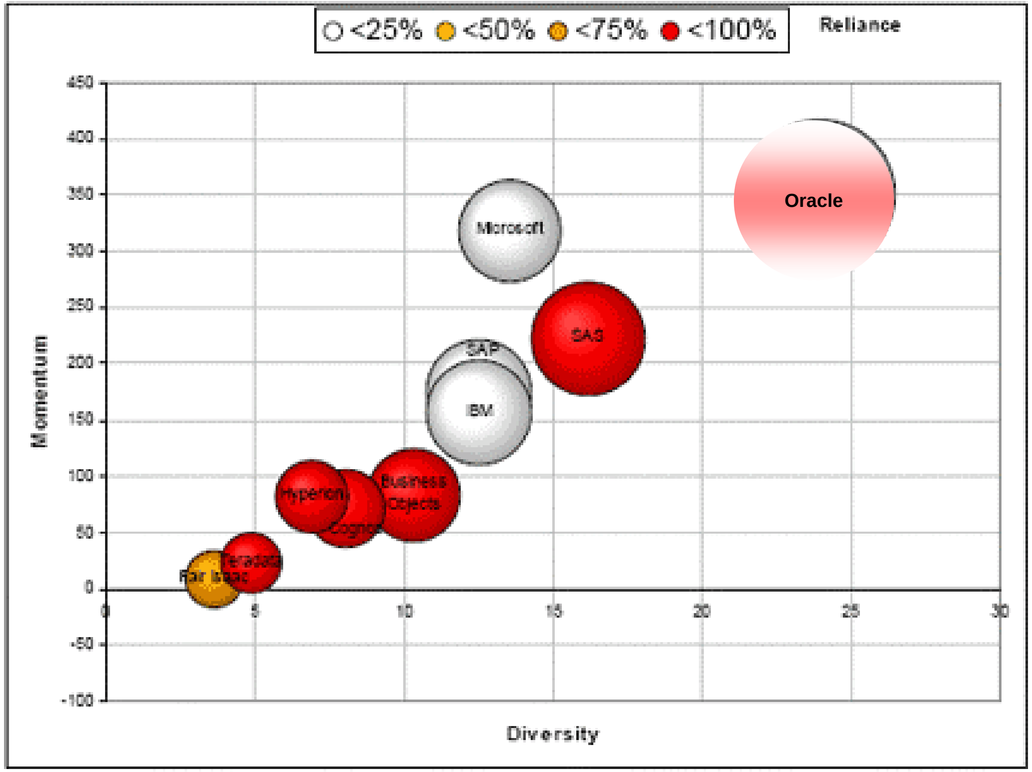

sas
Technical Paper
SAS® In-Database Processing
A Roadmap for Deeper Technical Integration with Database Management Systems

- “The goal of the SAS In-Database initiative is … to achieve deeper technical integration with database providers..
- …, the SAS engine often must load and extract data over a network to and from the DBMS. This presents a series of challenges:
- …Network bottlenecks between SAS and the DBMS constrain access to large volumes of data
- … the results of the SAS processing must be transferred back to the DBMS for final storage, which further increases the cost.

Source: **SAS In-Database Processing** ***White Paper—October 2007***

ORACLE

IDC Worldwide Business Analytics Software



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

Brief Demonstrations

1. Oracle Data Mining
2. Oracle Business Intelligence EE
3. CRM Sales Prospector

Oracle Data Mining + OBI EE

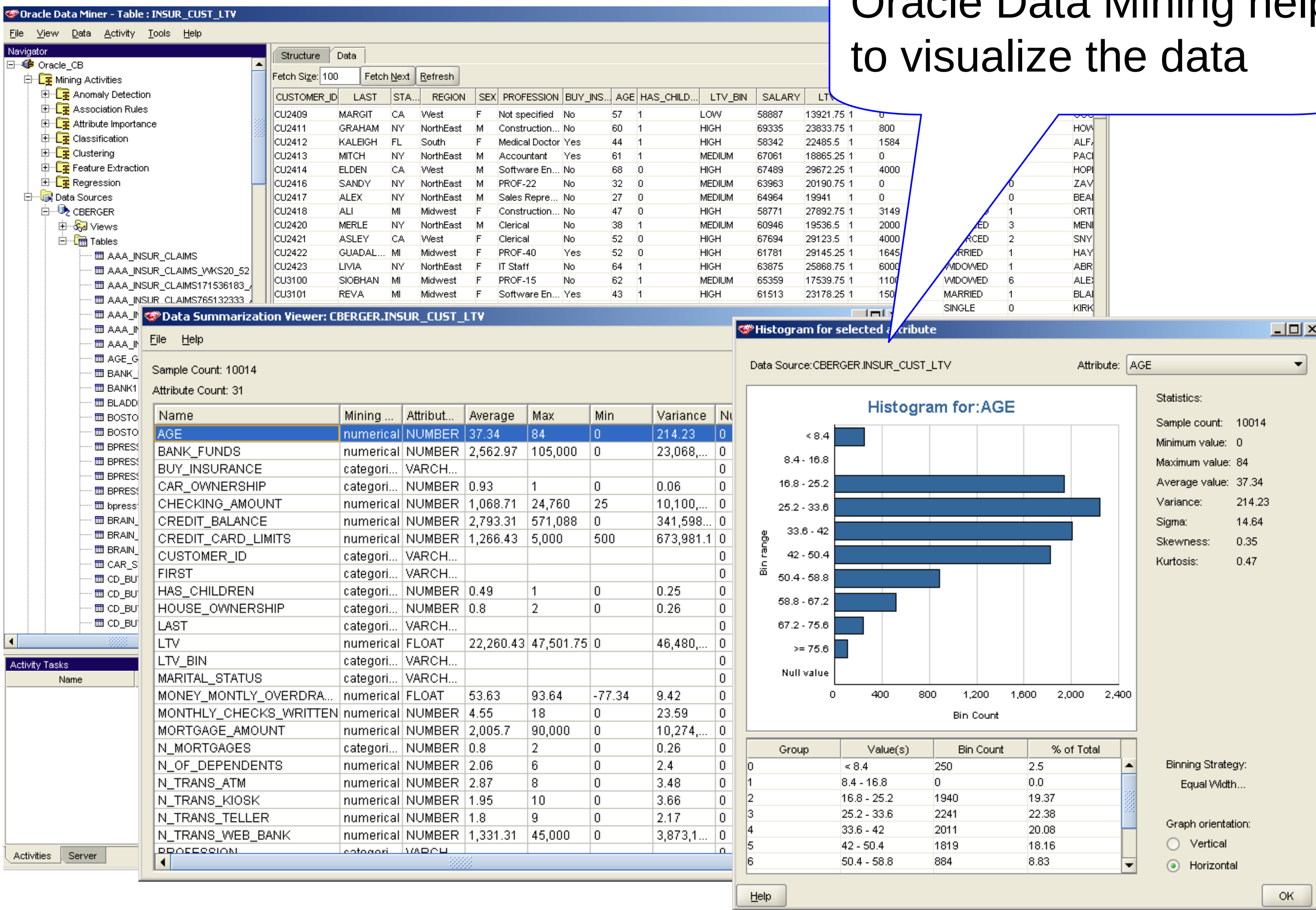
Quick Demo: Oracle Data Mining

- Scenario: Insurance Company
- Business problem(s):
  1. Better understand the business by looking at graphs of the data
  2. Identify the factors (attributes) most associated with Customer who BUY_INSURANCE
  3. Target Best Customers
     a. Build a predictive model to understand who will be a VERY_HIGH VALUE Customer …. And WHY (IF… THEN.. Rules that can describe them)
     b. Predict who is likely to be a VERY_HIGH VALUE Customer in the future
     c. View results in an OBI EE Dashboard
        - Including other business problems e.g. Fraud, Cross-Sell, etc.
        - (Entire process **can be automated** w/ PL/SQL and/or Java APIs)

Oracle Data Mining + OBI EE

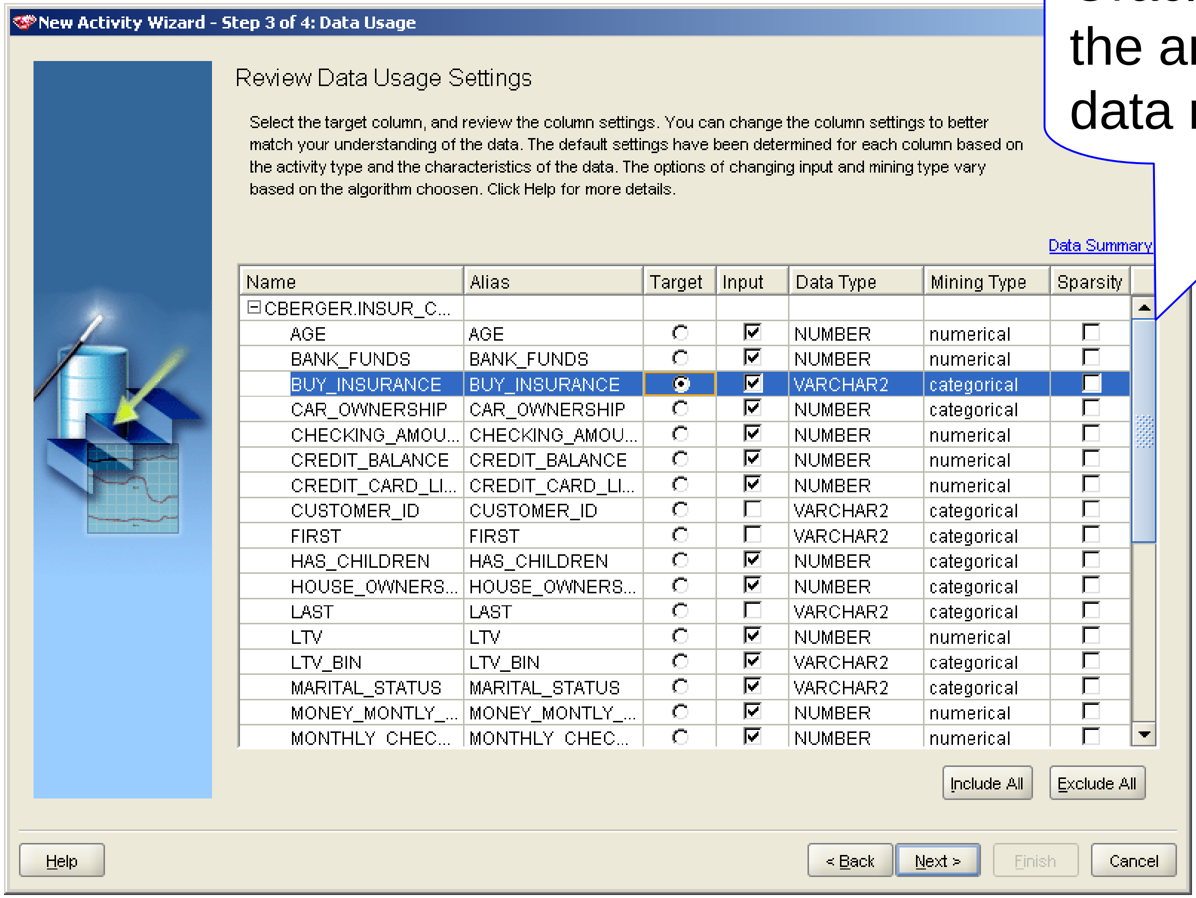
Understand the Data

Oracle Data Mining helps to visualize the data



ORACLE

Oracle Data Mining + OBI EE
Target the Right Customers

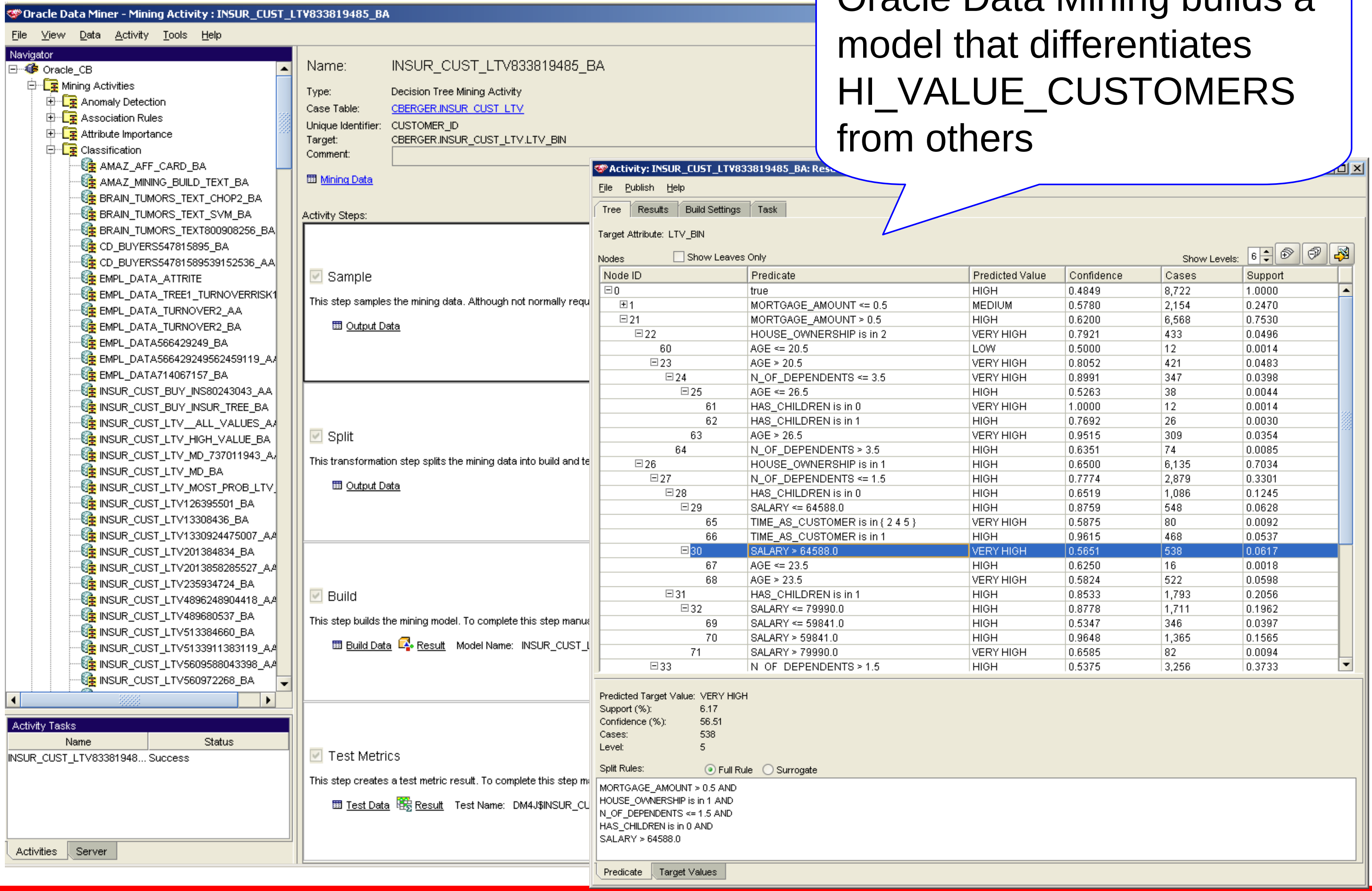


Oracle Data Miner guides the analyst through the data mining process

Oracle Data Mining + OBI EE

Targeting High Value Customers

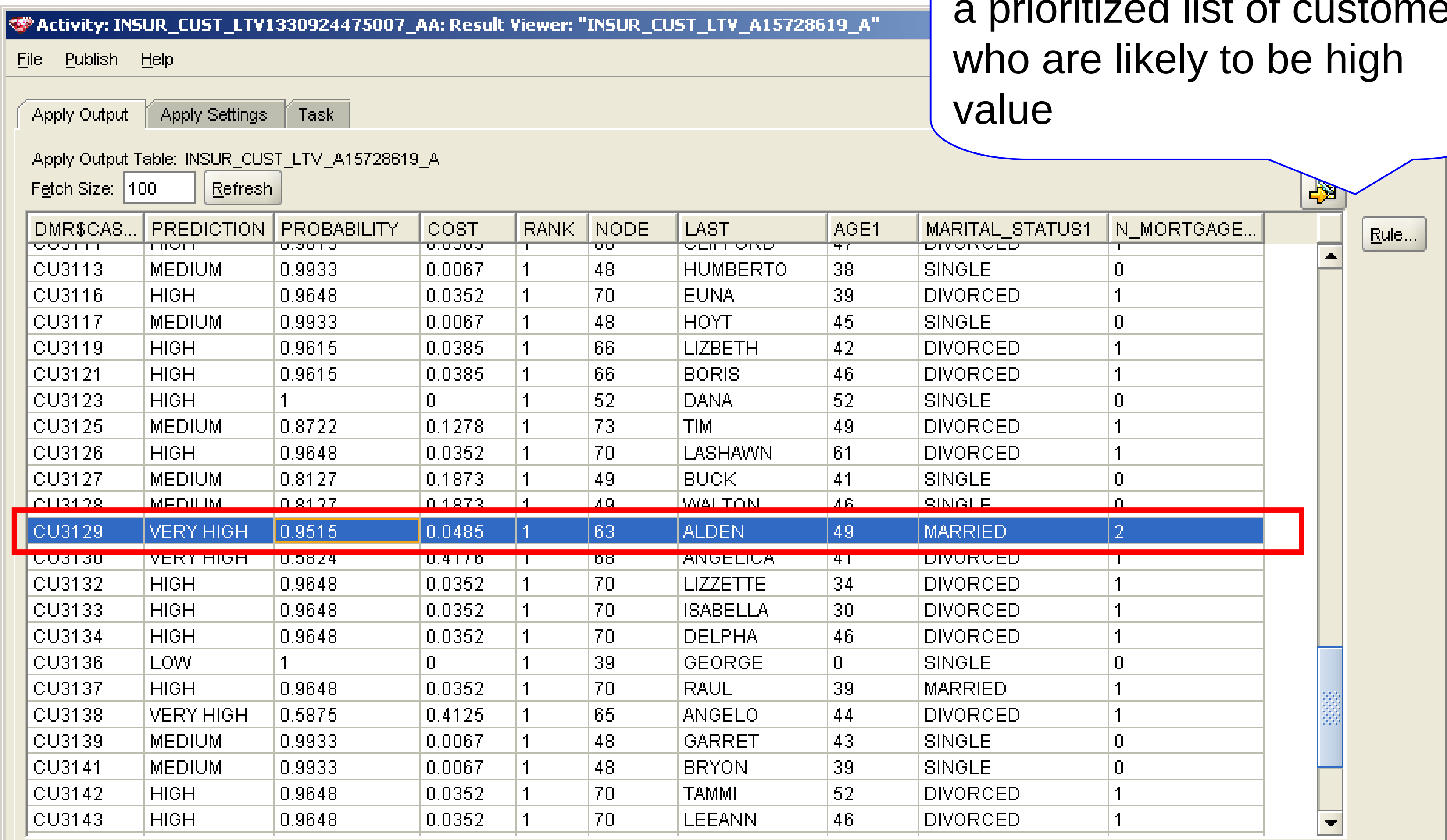
Oracle Data Mining builds a model that differentiates HI_VALUE_CUSTOMERS from others



Oracle Data Mining + OBI EE

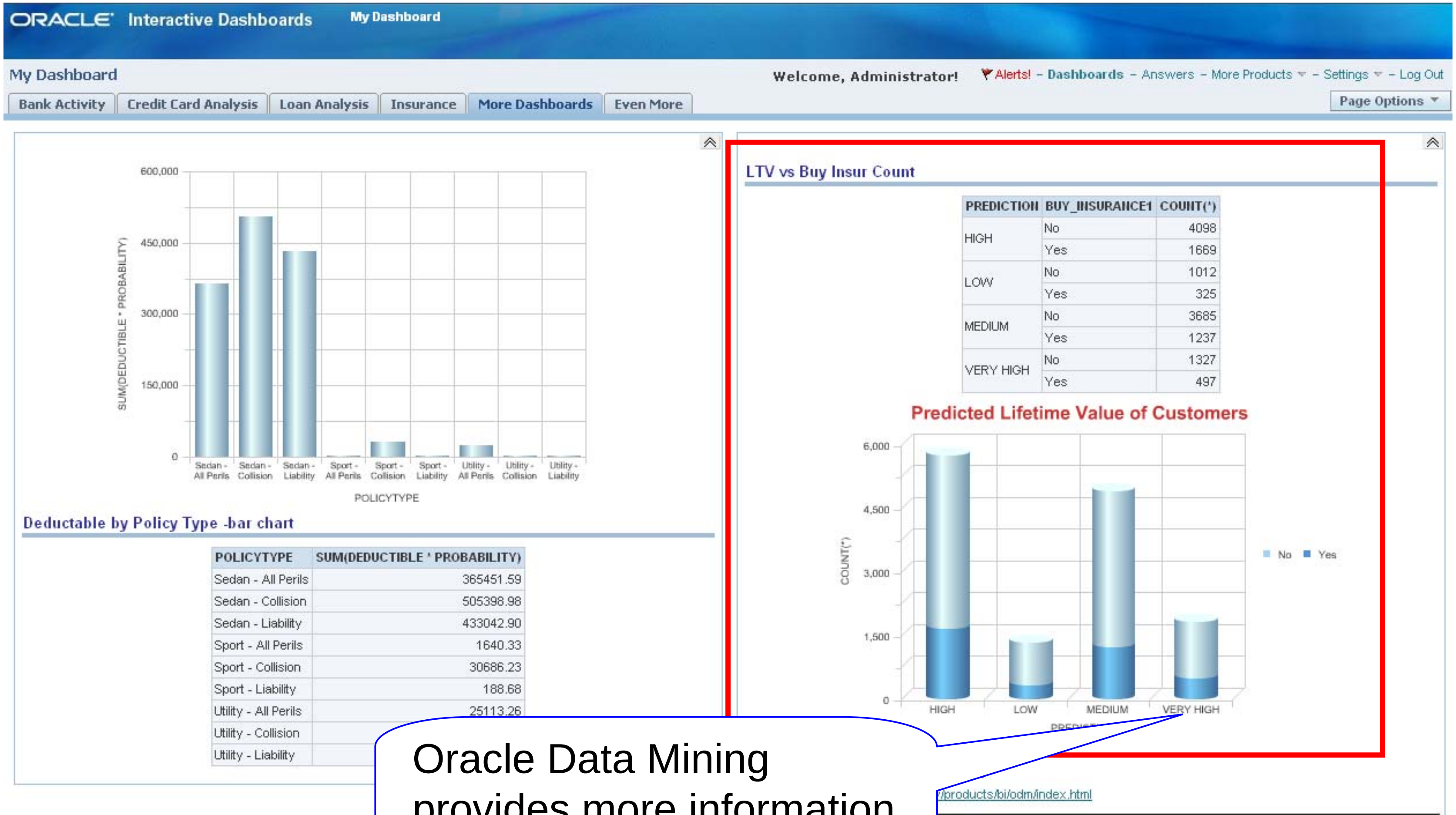
Targeting High Value Customers

Oracle Data Mining creates a prioritized list of customer who are likely to be high value



DMR$CAS...	PREDICTION	PROBABILITY	COST	RANK	NODE	LAST	AGE1	MARITAL_STATUS1	N_MORTGAGE...
CU3113	MEDIUM	0.9933	0.0067	1	48	HUMBERTO	38	SINGLE	0
CU3116	HIGH	0.9648	0.0352	1	70	EUNA	39	DIVORCED	1
CU3117	MEDIUM	0.9933	0.0067	1	48	HOYT	45	SINGLE	0
CU3119	HIGH	0.9615	0.0385	1	66	LIZBETH	42	DIVORCED	1
CU3121	HIGH	0.9615	0.0385	1	66	BORIS	46	DIVORCED	1
CU3123	HIGH	1	0	1	52	DANA	52	SINGLE	0
CU3125	MEDIUM	0.8722	0.1278	1	73	TIM	49	DIVORCED	1
CU3126	HIGH	0.9648	0.0352	1	70	LASHAWN	61	DIVORCED	1
CU3127	MEDIUM	0.8127	0.1873	1	49	BUCK	41	SINGLE	0
CU3128	MEDIUM	0.8127	0.1873	1	49	WALTON	46	SINGLE	0
CU3129	VERY HIGH	0.9515	0.0485	1	63	ALDEN	49	MARRIED	2
CU3130	VERY HIGH	0.5824	0.4176	1	68	ANGELICA	41	DIVORCED	1
CU3132	HIGH	0.9648	0.0352	1	70	LIZZETTE	34	DIVORCED	1
CU3133	HIGH	0.9648	0.0352	1	70	ISABELLA	30	DIVORCED	1
CU3134	HIGH	0.9648	0.0352	1	70	DELPHA	46	DIVORCED	1
CU3136	LOW	1	0	1	39	GEORGE	0	SINGLE	0
CU3137	HIGH	0.9648	0.0352	1	70	RAUL	39	MARRIED	1
CU3138	VERY HIGH	0.5875	0.4125	1	65	ANGELO	44	DIVORCED	1
CU3139	MEDIUM	0.9933	0.0067	1	48	GARRET	43	SINGLE	0
CU3141	MEDIUM	0.9933	0.0067	1	48	BRYON	39	SINGLE	0
CU3142	HIGH	0.9648	0.0352	1	70	TAMMI	52	DIVORCED	1
CU3143	HIGH	0.9648	0.0352	1	70	LEEANN	46	DIVORCED	1

Integration with Oracle BI EE



Oracle Data Mining provides more information and better insight

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**

Oracle Data Mining (SQL & Java) APIs

HCM Prediction Demo

```
drop table HCM_SET;
exec dbms_data_mining.drop_model('HCMMODEL');

create table HCM_SET (setting_name varchar2(30), setting_value varchar2(4000));
insert into HCM_SET values ('ALGO_NAME','ALGO_SUPPORT_VECTOR_MACHINES');
insert into HCM_SET values ('PREP_AUTO','ON');
commit;

begin
dbms_data_mining.create_model('HCMMODEL', 'CLASSIFICATION',
  'EMPL_DATA', 'EMPL_ID', 'CURR_EMPL', 'HCM_SET');
end;
/

-- accuracy (per-class and overall)
col actual format a6
select actual, round(corr*100/total,2) percent, corr, total-corr incorr, total from
(select actual, sum(decode(actual,predicted,1,0)) corr, count(*) total from
(select CURR_EMPL actual, prediction(HCMMODEL using *) predicted
from EMPL_DATA_JUNE07)
group by rollup(actual));

-- top 5 very high value, current employees most likely to leave
select * from
(select empl_id, round(prob_leave*100,2) percent_leave,
    rank() over (order by prob_leave desc) rnk from
(select empl_id, prediction_probability(HCMMODEL, 'NO' using *) prob_leave
from EMPL_DATA_JUNE07
where CURR_EMPL = 'YES' and LTV_BIN = 'VERY HIGH'))
where rnk <= 5
order by percent_leave desc;
```

```
ACTUAL  PERCENT  CORR    INCORR  TOTAL
------- -------- ------- ------- -------
NO      84.04    3133    595     3728
YES     80.61    8159    1963    10122
        81.53    11292   2558    13850
Elapsed: 00:00:01.51
SQL>

EMPL_ID    PERCENT_LEAVE   RNK
---------- -------------   ----------
772858     96.84           1
775441     95.65           2
777992     92.1            3
773473     91.51           4
771813     90.21           5
Elapsed: 00:00:00.29
SQL>
```

ORACLE

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';
```

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(INSUR_CUST_LT68054_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

Prediction Multiple Models/Optimization

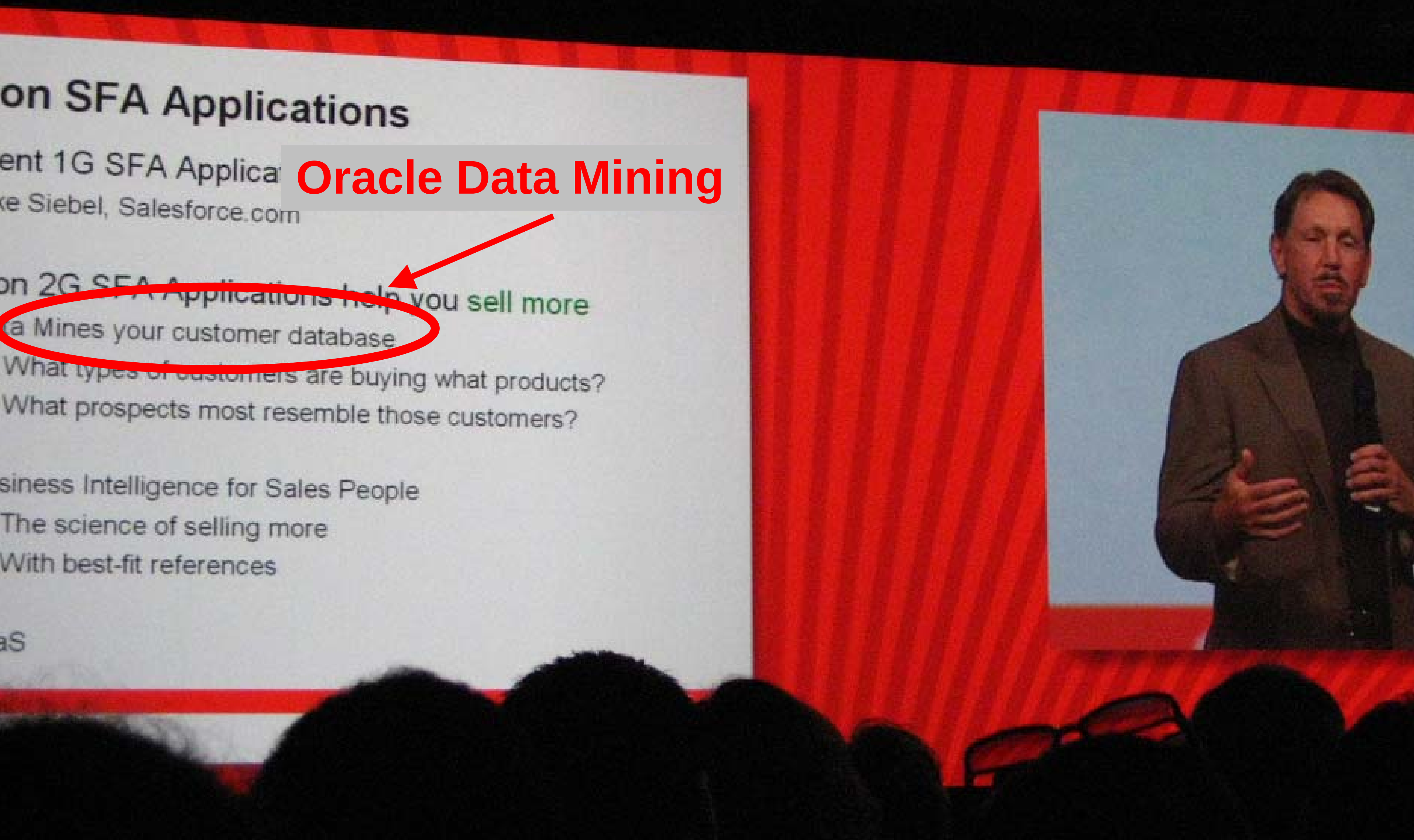
On-the-fly, multiple models; then sort by expected revenues

```
with records as (select
   178255 ANNUAL_INCOME,
      30 AGE,
      'Bach.' EDUCATION,
      'Married' MARITAL_STATUS,
      'Male' SEX,
      70 HOURS_PER_WEEK,
      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;
```

Oracle Sales Prospector

Larry Ellison Oracle Open World Keynote
November 2007

- **Announces Fusion Edge CRM On-Demand Hosted Application with integrated data mining to mine customer database**

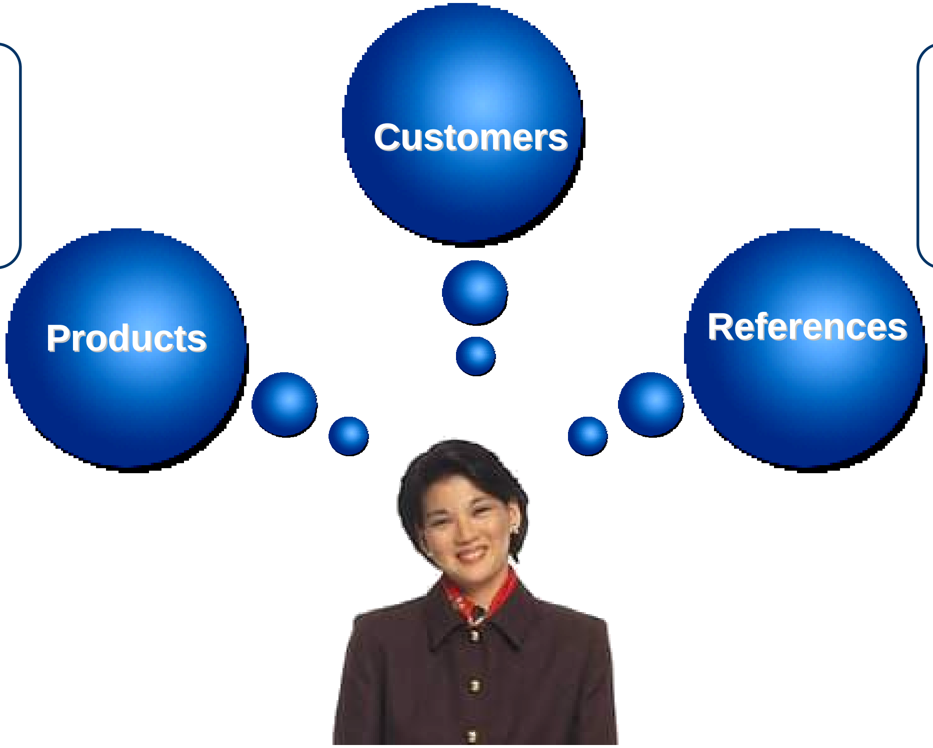


How Can I Sell More?

- Which prospects most resemble those customers?

- Which types of customers are buying which products?

- Which references can I use to help me close my deals?

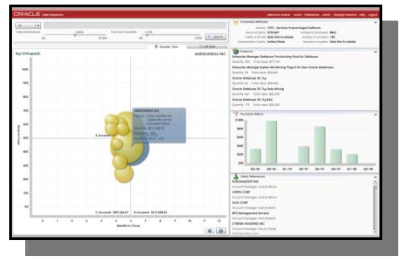


Sales Rep

Oracle Data Mining = the Science of Selling

Oracle Sales Prospector

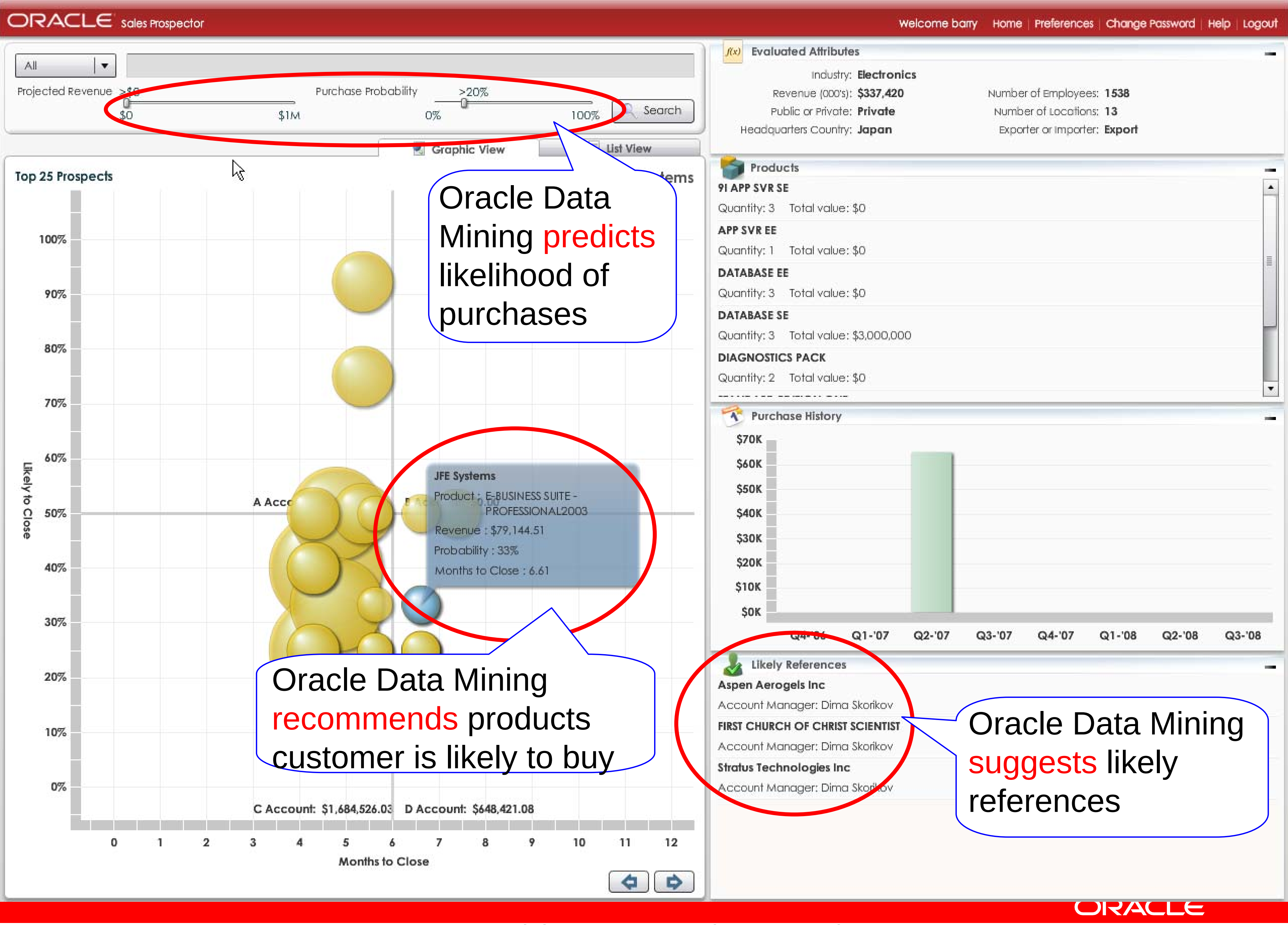
ODM Predictions exposed via Social CRM Dashboards



Oracle Database 11G

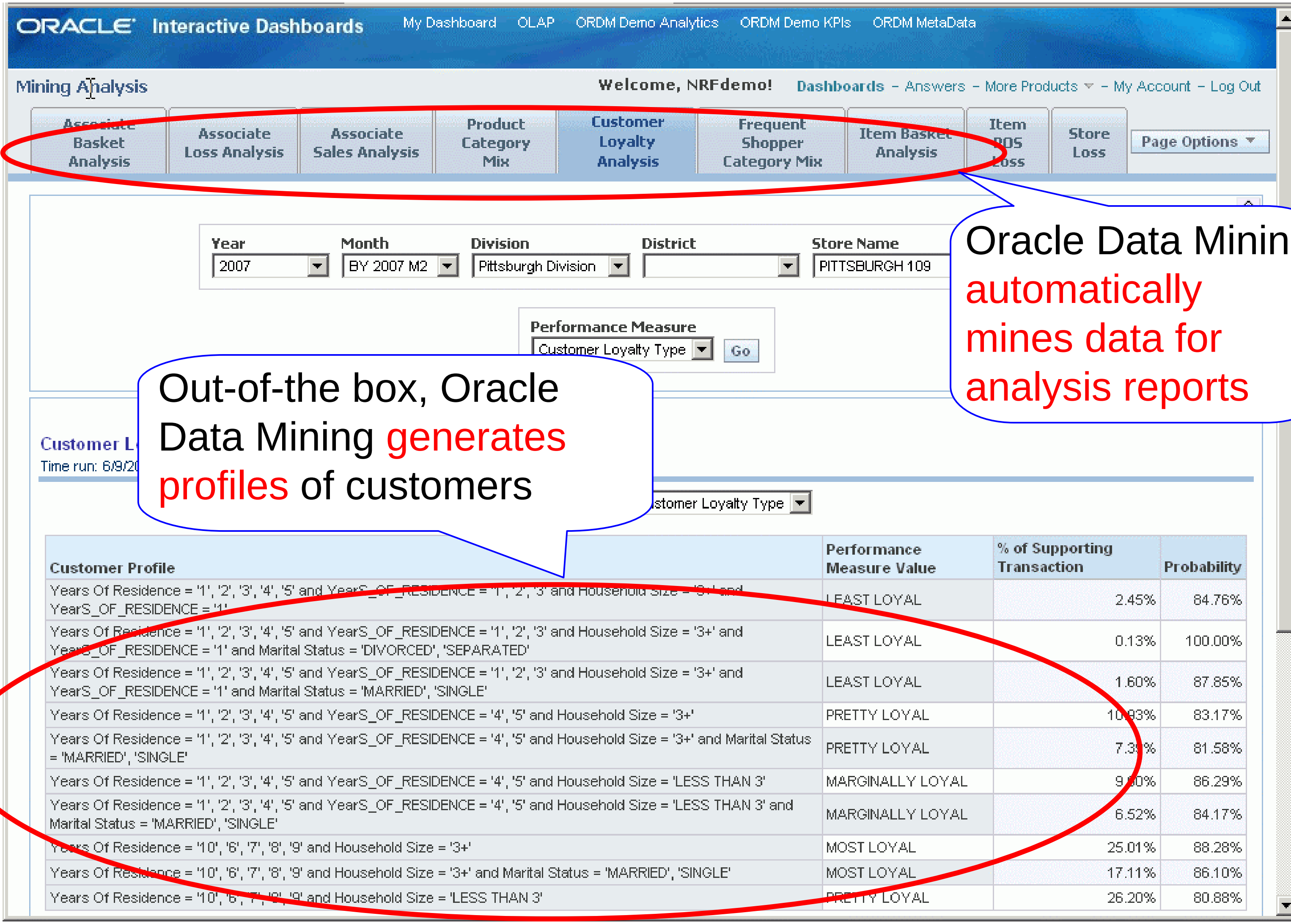
Social CRM schema ships with Oracle Database EE 11g + Data Mining Option

ORACLE


ORACLE Sales Prospector
Welcome barry | Home | Preferences | Change Password | Help | Logout
Projected Revenue >$0
Purchase Probability >20%
Search
Graphic View
List View
Top 25 Prospects
Oracle Data Mining predicts likelihood of purchases
JFE Systems
Product : E-BUSINESS SUITE - PROFESSIONAL2003
Revenue : $79,144.51
Probability : 33%
Months to Close : 6.61
Oracle Data Mining recommends products customer is likely to buy
C Account: $1,684,526.03
D Account: $648,421.08
Likely to Close
Months to Close
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
9I 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
Purchase History
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

Oracle Retail Data Model

Oracle Retail Data Model



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[illegible]3%	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	7.3[illegible]%	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[illegible]0%	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%

Oracle Data Mining automatically mines data for analysis reports

Out-of-the box, Oracle Data Mining generates profiles of customers

Summary

Oracle Data Mining Summary

- Powers Next-Generation Predictive Applications
  - Rapidly Build Applications that Automatically Mine Data
  - Code Once, Run Anywhere
  - Parallel and Distributed Processing
  - Industry Standard SQL and Java APIs

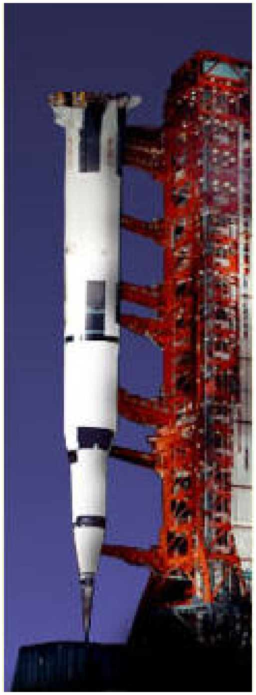
- Industry Leader in In-Database Data Mining
  - Option to the Industry Leading RDBMS—Oracle Database
  - Classification, Regression, Attribute Importance
  - Clustering, Market Basket Analysis, Anomaly Detection, Feature Extraction
  - Cutting Edge Algorithms: SVM, One-Class SVM, NMF, Scalable GLM

Oracle Data Mining Summary

- More Information from More Data
  - Easy to use Oracle Data Miner Graphical User Interface
  - Wide Range of In-Database Data Mining Algorithms and Statistics
  - Mine Text, Transactional, and Star Schema Data
  - Mine XML, Semantic RDF, Spatial, and OLAP Data
- Eliminate Barriers Between Analysts and IT
  - Quickly Disseminate Analytical Results and Models Throughout the Organization
  - Include Real-Time Predictive Models and New Insights in SQL queries
  - Eliminate Data Movement, Maximize Security

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

the modeling agency

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 (10*g*R2 or 11*g*) 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 TECHNOLOGY NETWORK

ORACLE Data Mining

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

QUESTIONS

ANSWERS