

Oracle's Machine Learning and Advanced Analytics 12.2 New Features

Move the Algorithms; Not the Data!

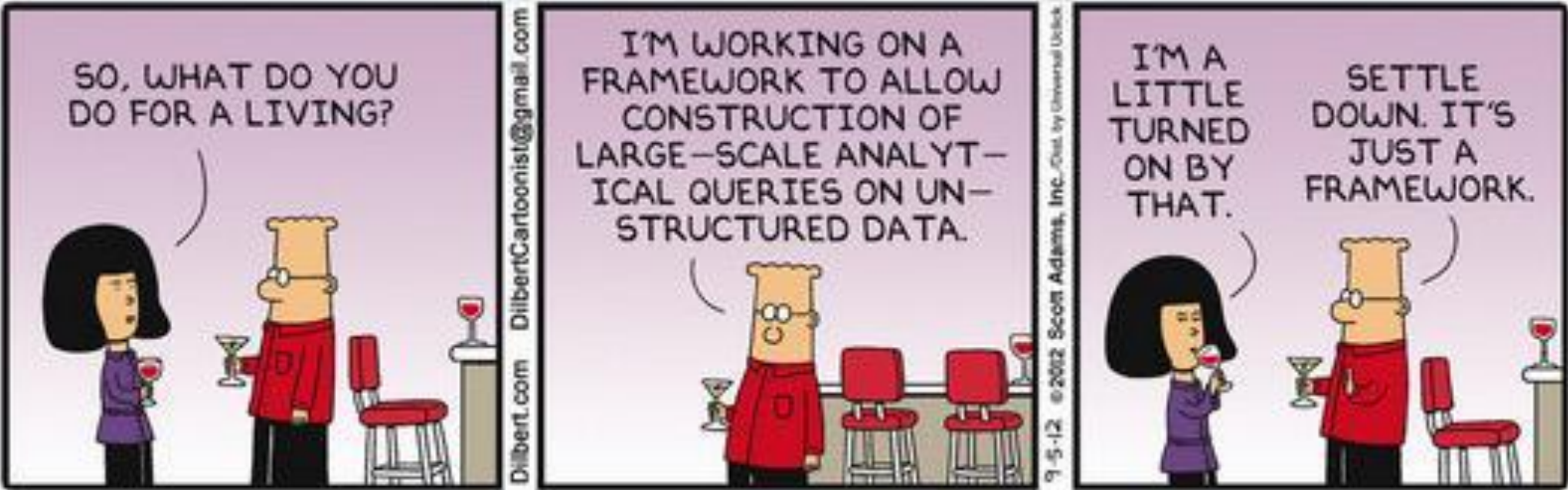
Charlie Berger, MS Engineering, MBA
Sr. Director Product Management,
Advanced Analytics and Machine Learning
charlie.berger@oracle.com
www.twitter.com/CharlieDataMine



Safe Harbor Statement

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.

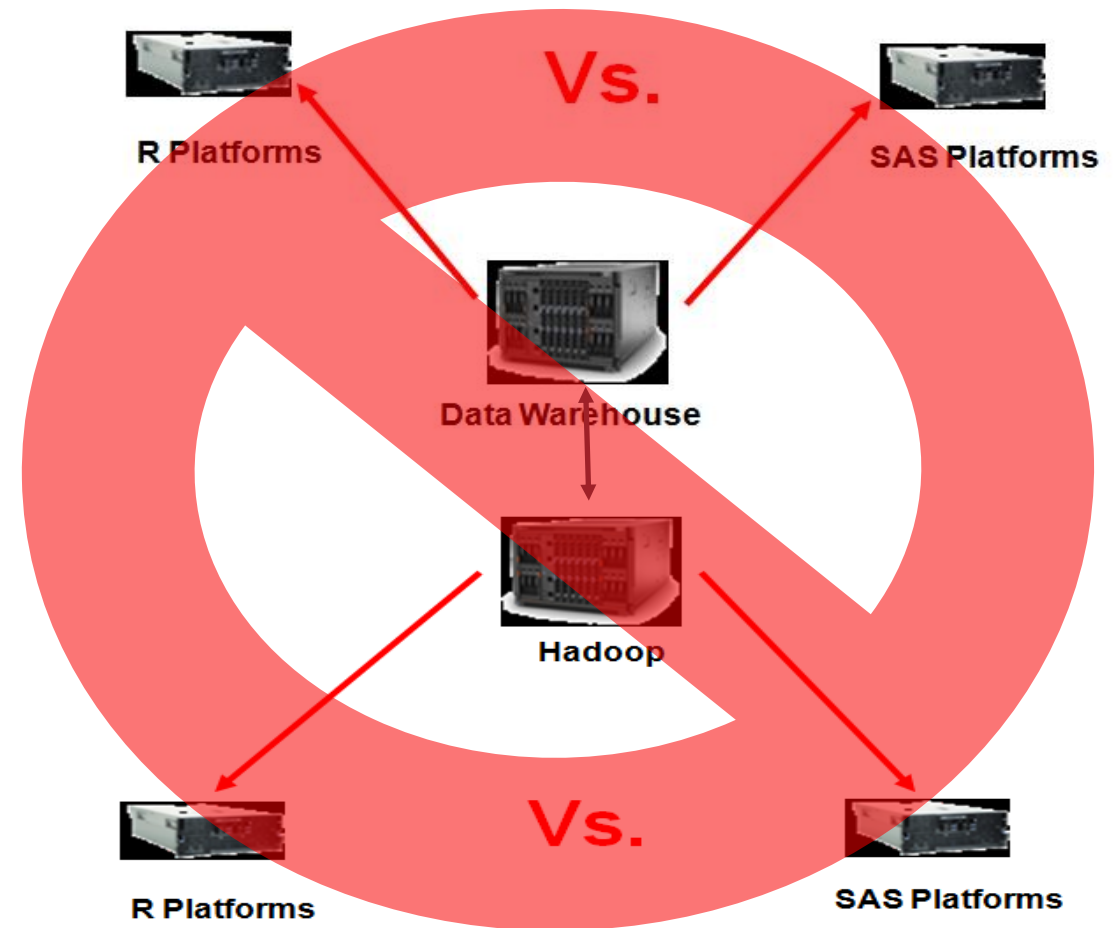
Dilbert on Big Data



Machine Learning/Analytics + Data Warehouse + Hadoop

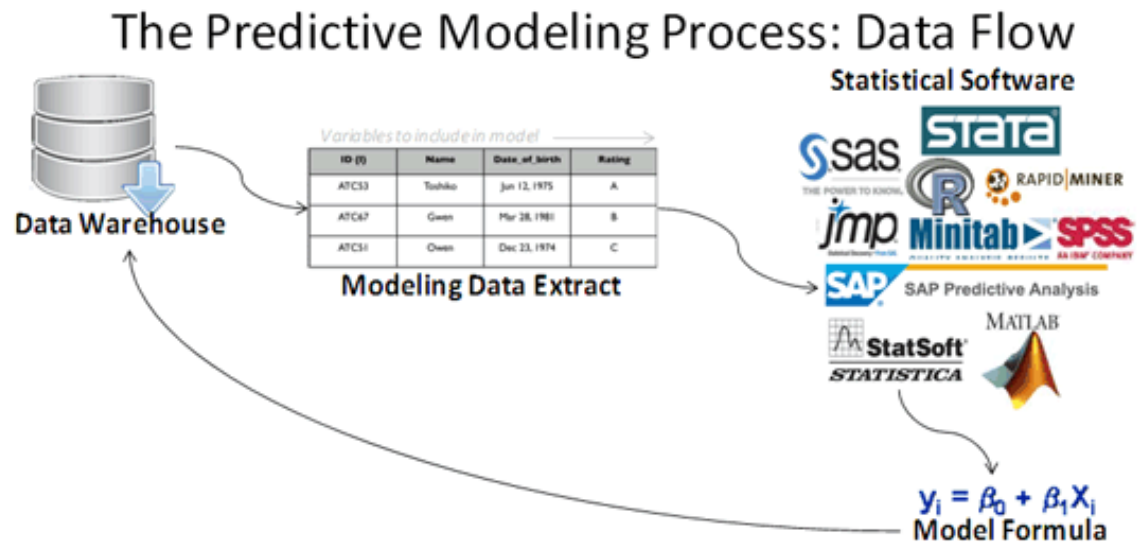
- Platform Sprawl

- More Duplicated Data
- More Data Movement Latency
- More Security challenges
- More Duplicated Storage
- More Duplicated Backups
- More Duplicated Systems
- More Space and Power



Traditional vs. Oracle Machine Learning/Predictive Analytics

- **Traditional**— *“Move the data”*

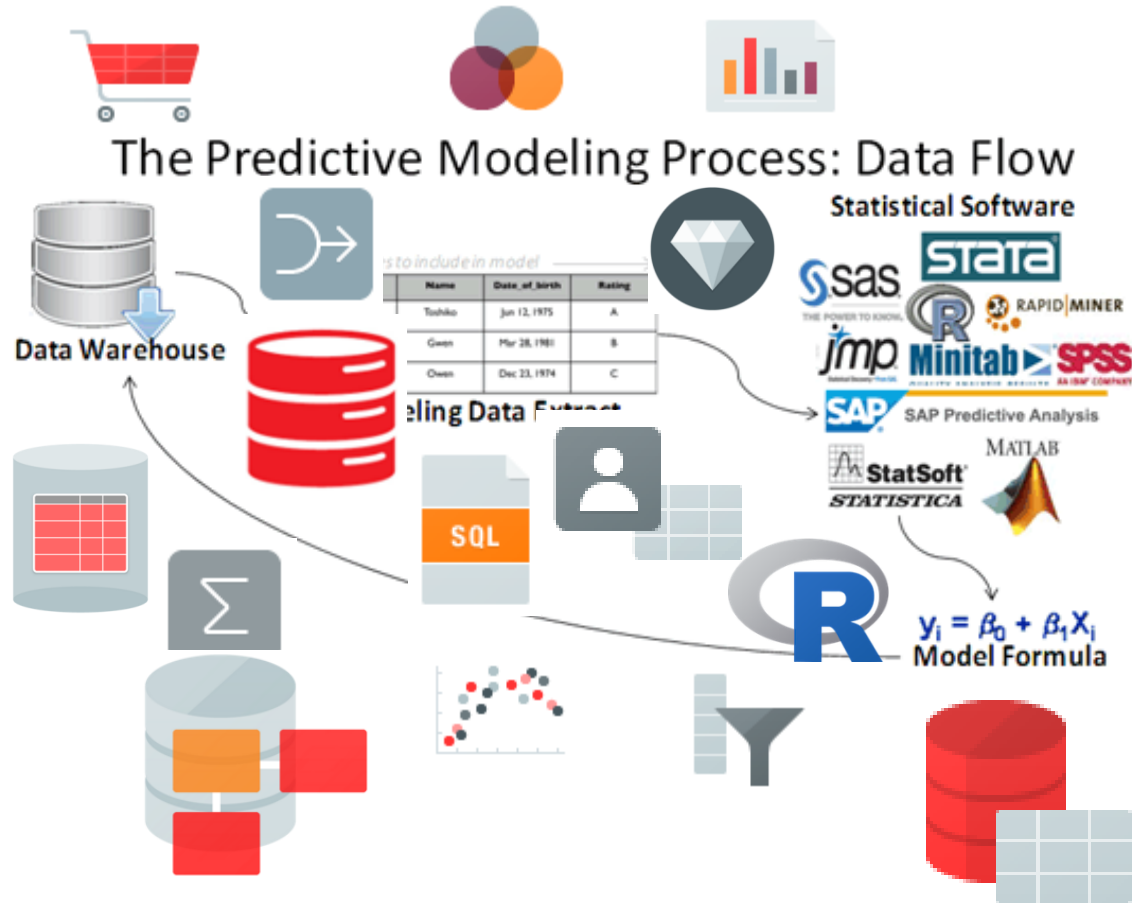


ORACLE® — *“Don’t move the data!”*



Traditional vs. Oracle Machine Learning/Predictive Analytics

- **Traditional**— *“Move the data”*



ORACLE® — *“Move the algorithms”*



Simpler, Smarter Data Management
+ Analytics / Machine Learning Architecture

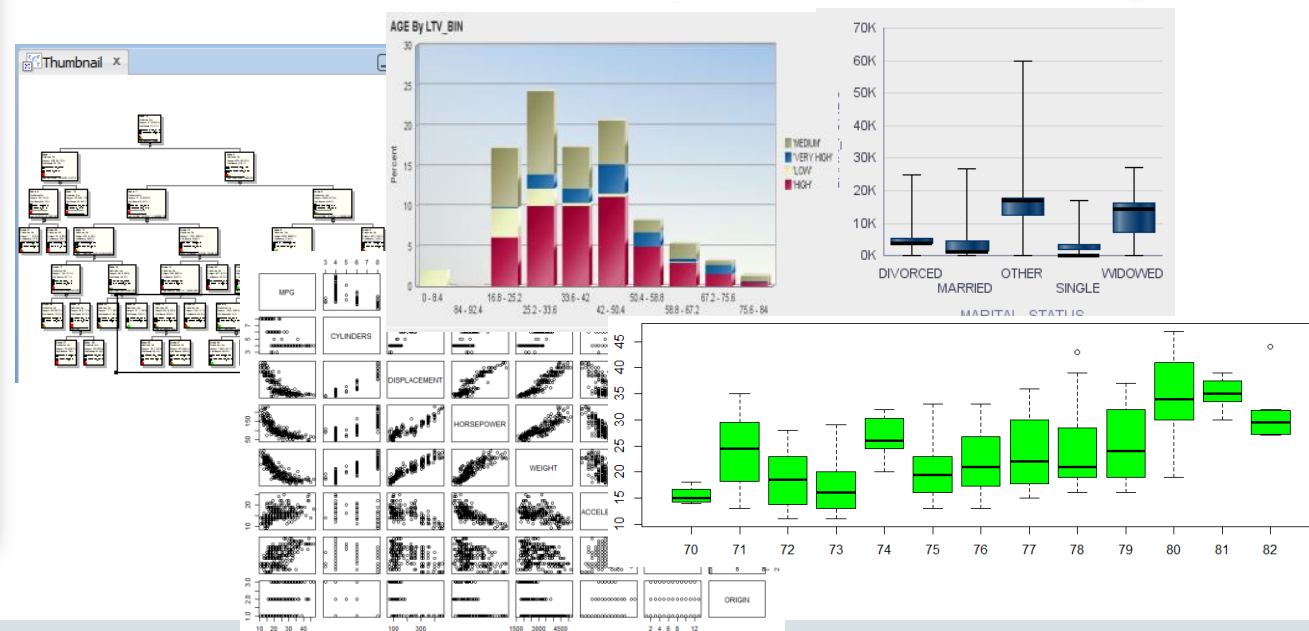
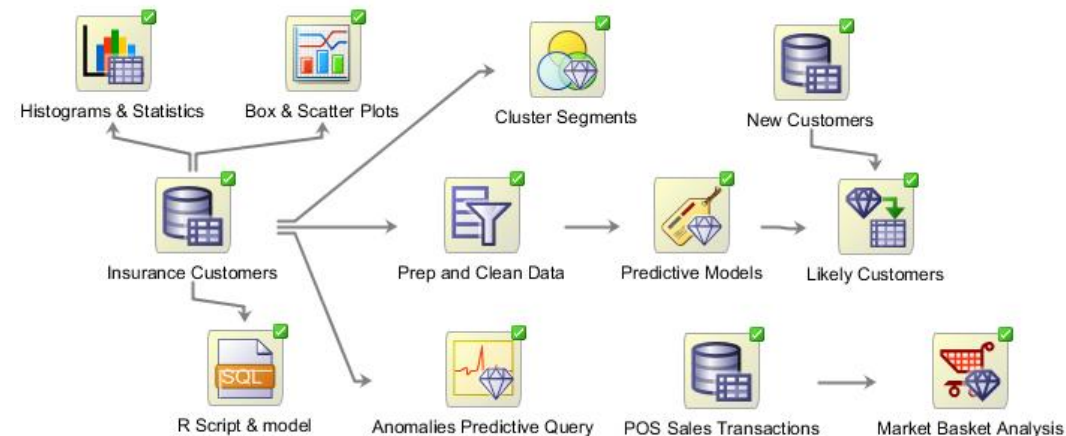
Oracle's Machine Learning/Advanced Analytics

Fastest Way to Deliver Scalable Enterprise-wide Predictive Analytics



Key Features

- Parallel, scalable data mining algorithms and R integration
- In-Database + Hadoop—Don't move the data
- Data analysts, data scientists & developers
- Drag and drop workflow, R and SQL APIs
- Extends data management into powerful advanced/predictive analytics platform
- Enables enterprise predictive analytics deployment + applications



Oracle's Machine Learning & Adv. Analytics Algorithms



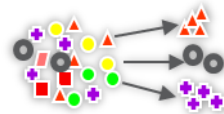
Classification

- Naïve Bayes
- Logistic Regression (GLM)
- Decision Tree
- Random Forest
- Neural Network
- Support Vector Machine
- Explicit Semantic Analysis
- Gaussian Mixture Models



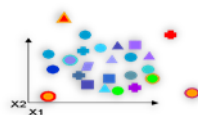
Clustering

- Hierarchical K-Means
- Hierarchical O-Cluster
- Expectation Maximization (EM)



Anomaly Detection

- One-Class Support Vector Machine (SVM)



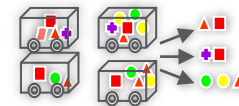
Regression

- Generalized Linear Model
- Support Vector Machine (SVM)
- Random Forest
- Linear Model
- Stepwise Linear regression
- LASSO



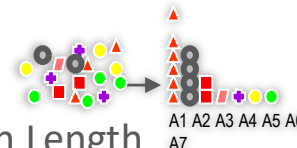
Association Rules

- A priori



Attribute Importance

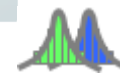
- Minimum Description Length
- Principal Component Analysis (PCA)
- Unsupervised Pair-wise KL Divergence



Predictive Queries



Statistical Functions



- Basic statistics: median, stdev, t-test, F-test, Pearson's, Chi-sq, Anova, etc.

Algorithm Support for Text

- Algorithms support text type
- Tokenization and theme extraction
- Explicit Semantic Analysis (ESA) for document similarity



Feature Extraction

- Principal Component Analysis (PCA)
- Non-negative Matrix Factorization
- Singular Value Decomposition (SVD)

Time Series

- Single Exponential Smoothing
- Double Exponential Smoothing

Open Source ML Algorithms



- CRAN R Algorithm Packages through Embedded R Execution
- Spark MLlib algorithm integration

Oracle's Machine Learning/Advanced Analytics Platforms

Machine Learning Algorithms Embedded in the Data Management Platforms

“Information Producers”

Data Scientists, R Users, Citizen Data Scientists

The 'Information Producers' section shows three main tools: RStudio, Oracle SQL Developer, and Oracle Data Miner. RStudio displays various data plots like histograms and line graphs. Oracle SQL Developer shows a workflow for data preparation and analysis, including steps like 'Prep and Clean C' and 'Cluster Segments'. Oracle Data Miner provides a visual interface for building predictive models, with a flowchart showing data from 'Insurance Customers' through 'Prep and Clean C' to 'Cluster Segments' and 'New Customers'.

“Information Consumers”

BI Analysts, Managers

Functional Users (HCM, CRM)

The 'Information Consumers' section shows two main tools: Oracle Data Visualization and Oracle Applications. Oracle Data Visualization displays a colorful pie chart and a scatter plot, representing data analysis results. Oracle Applications shows a user interface for managing resources, with a list of names and roles.

ORACLE® Data Management + Advanced Analytical Platform

Big Data SQL

Oracle BDA Hadoop

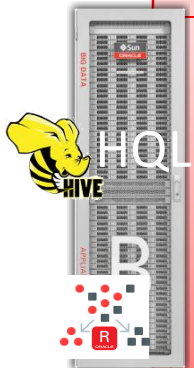
Oracle Database EE

“Oracle ML” Big Data Edition
Machine Learning Algorithms,
Statistical Functions + R Integration
for Scalable, Parallel, Distributed Execution

“Oracle Machine Learning” Database Edition
Machine Learning Algorithms,
Statistical Functions + R Integration
for Scalable, Parallel, Distributed, in-DB Execution

Oracle Cloud

Oracle Database EE

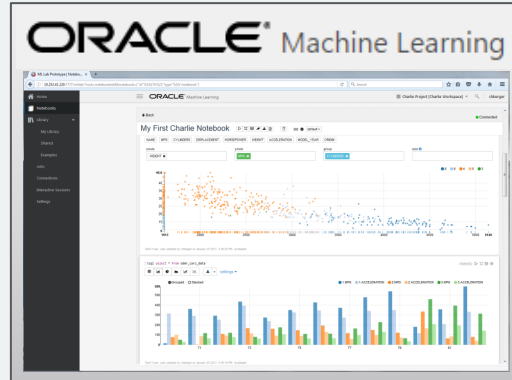
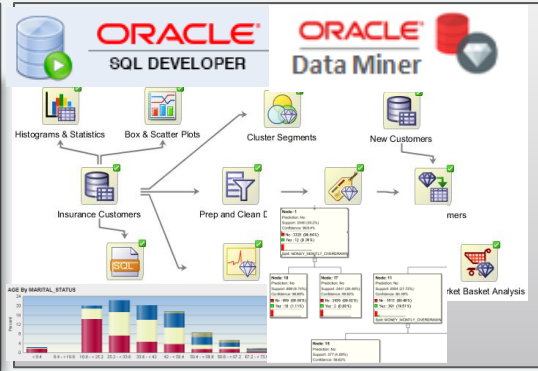
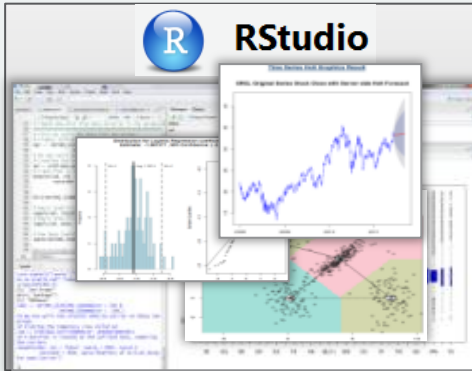


Oracle's Machine Learning/Advanced Analytics Platforms

Machine Learning Algorithms Embedded in the Data Management Platforms

“Information Producers”

Data Scientists, R Users, Citizen Data Scientists



New Zeppelin notebook based UI for data scientists collaborating and sharing ML analytical methodologies in Clouds

ORACLE® Data Management + Advanced Analytical Platform

Big Data SQL

Oracle BDA Hadoop

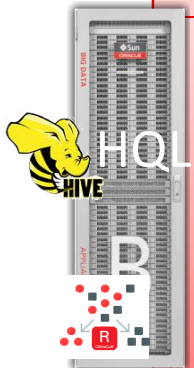
Oracle Database EE

“Oracle ML” Big Data Edition
Machine Learning Algorithms,
Statistical Functions + R Integration
for Scalable, Parallel, Distributed Execution

“Oracle Machine Learning” Database Edition
Machine Learning Algorithms,
Statistical Functions + R Integration
for Scalable, Parallel, Distributed, in-DB Execution

Oracle Cloud

Oracle Database EE



Oracle Advanced Analytics 12.2

Model Build Time Performance

Unofficial



NEW IN
12.2

OAA 12.2 Algorithms

Rows (Ms)

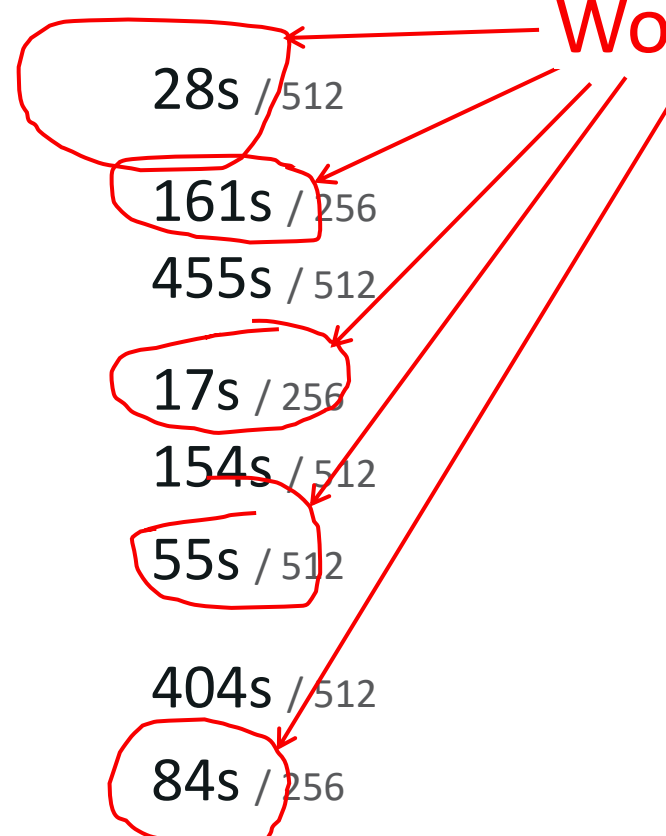
T7-4 (Sparc & Solaris)

X5-4 (Intel and Linux)

Model Build Time (Secs / Degree of Parallelism)

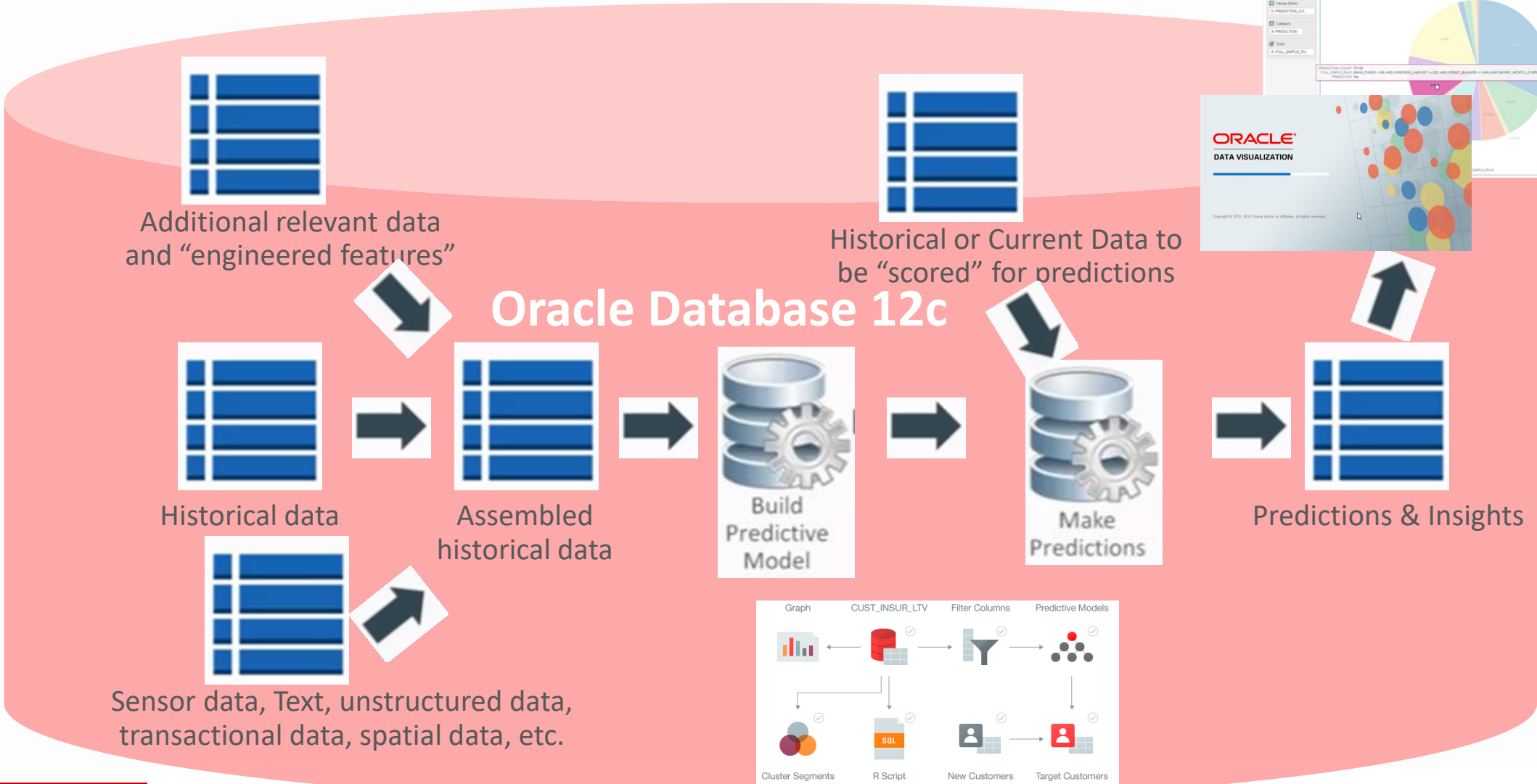
Attributes Importance	640	28s / 512	44s / 72
K Means Clustering	640	161s / 256	268s / 144
Expectation Maximization	159	455s / 512	588s / 144
Naive Bayes Classification	320	17s / 256	23s / 72
GLM Classification	640	154s / 512	363s / 144
GLM Regression	640	55s / 512	93s / 144
Support Vector Machine (IPM solver)	640	404s / 512	1411s / 144
Support Vector Machine (SGD solver)	640	84s / 256	188s / 72

Wow! That's Fast!



Machine Learning & Advanced Analytical Methodologies

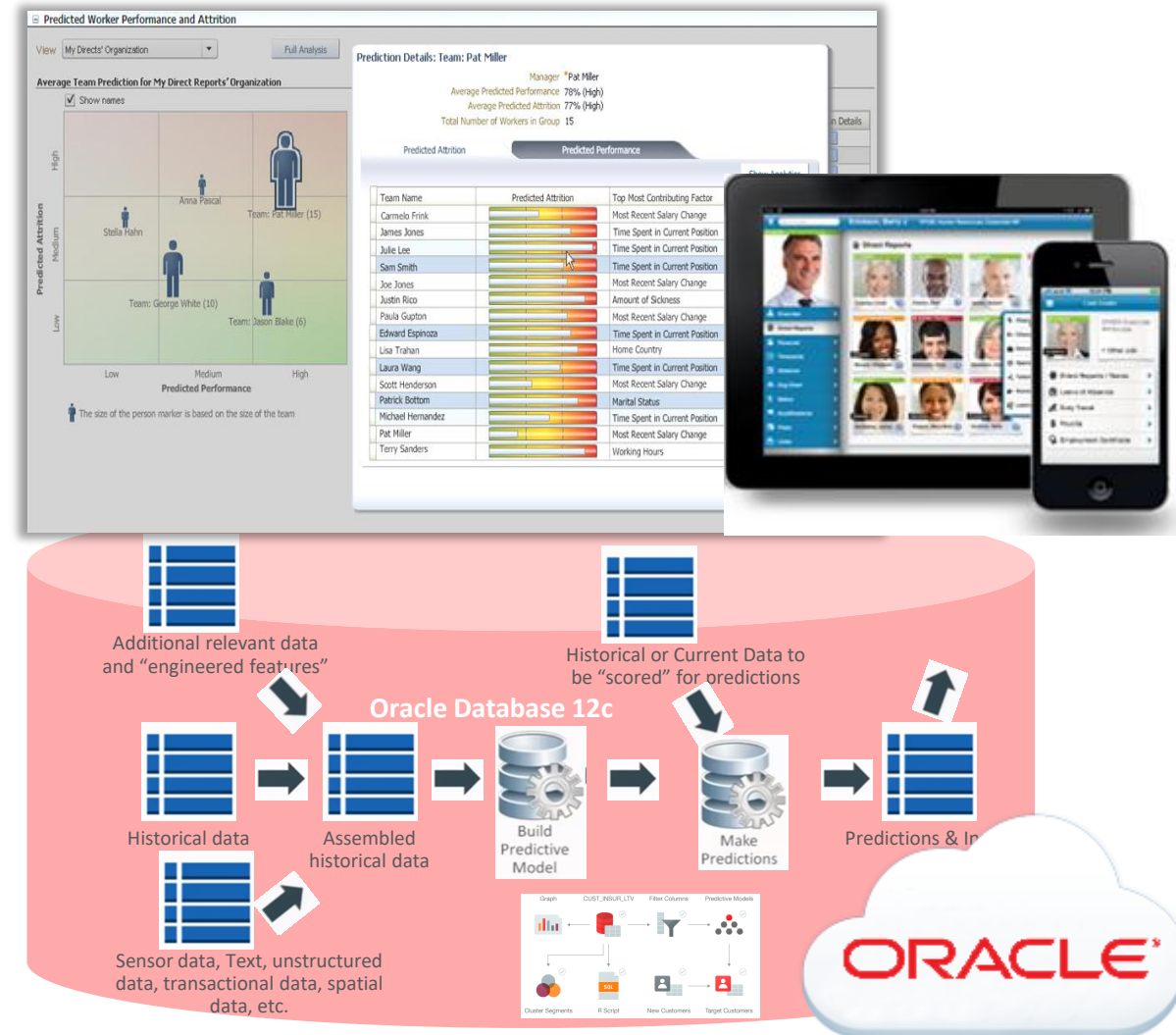
Data Preparation & Adv. Analytical Process Runs In-Database



Example Predictive Appl: HCM Cloud—Workforce Predictions

Complete, Integrated, Embedded, Automated and Interactive “Predictive HCM” Solution

- Integrated data management + embedded predictive analytics
- Full 360 degree employee view
- Single source of HCM data data
- Interactive dashboards and “What if” analysis
- Customizable if desired to add input variables to predictive models
- Mobile + Oracle Cloud solutions



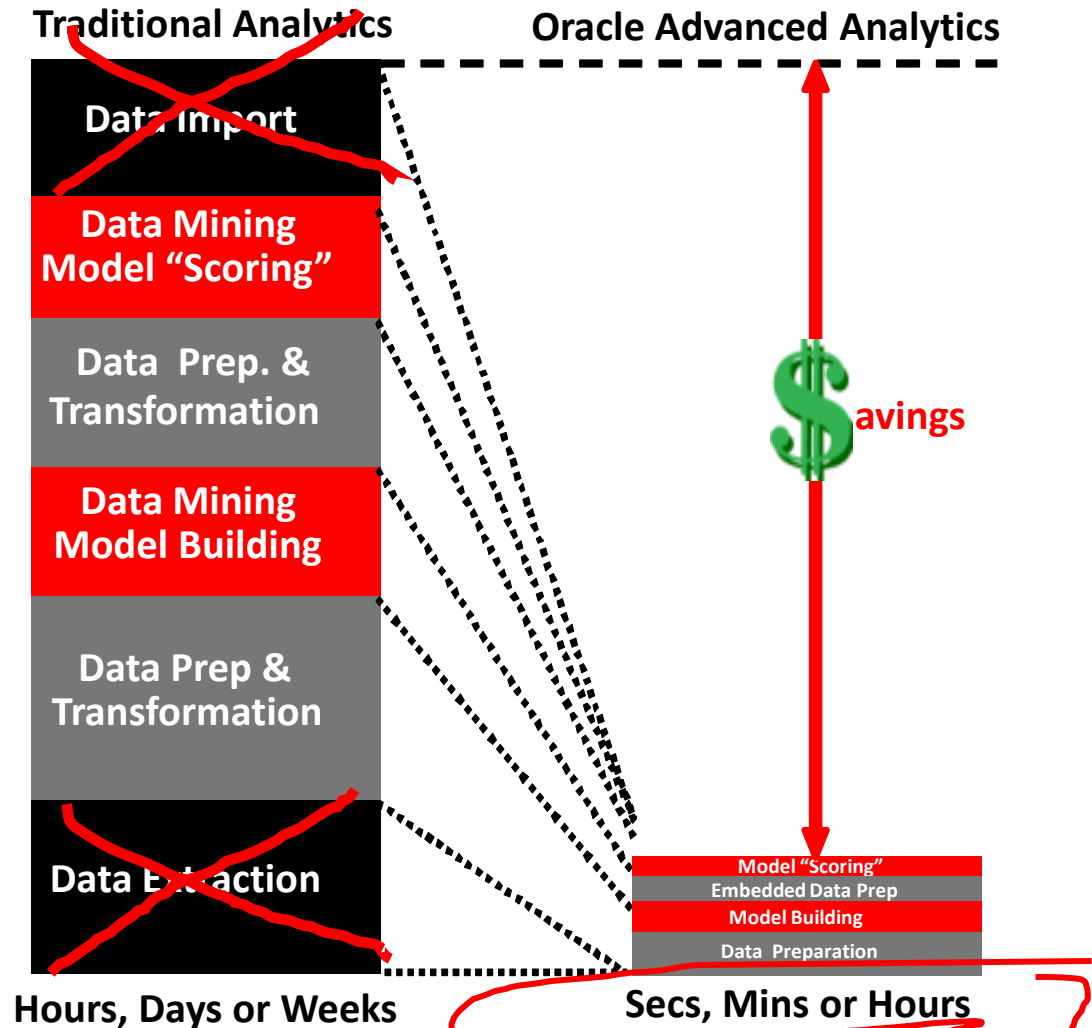
Oracle's Machine Learning/Advanced Analytics

Fastest Way to Deliver Scalable Enterprise-wide ML/Predictive Analytics



Major Benefits

- Data remains in Database & Hadoop
 - Model building and scoring occur in-database
 - Use R packages with data-parallel invocations
- Leverage investment in Oracle IT
 - Eliminate data duplication
 - Eliminate separate analytical servers
- Deliver enterprise-wide applications
 - GUI for ML/Predictive Analytics & code gen
 - R interface leverages database as HPC engine



Oracle's Machine Learning/Advanced Analytics

Fastest Way to Deliver Scalable Enterprise-wide Predictive Analytics



Key Features

- Parallel, scalable data mining algorithms and R integration
- In-Database + Hadoop—Don't move the data
- Data analysts, data scientists & developers
- Drag and drop workflow, R and SQL APIs
- Extends data management into powerful advanced/predictive analytics platform
- Enables enterprise predictive analytics deployment + applications

You Can Think of Oracle's Advanced Analytics Like This...

Traditional SQL

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

SQL Queries

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



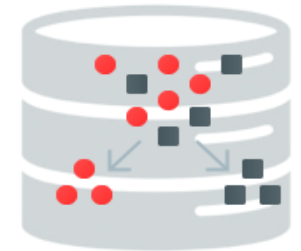
+

Oracle Advanced Analytics - SQL &

- Automated knowledge discovery, model building and deployment
- Domain expertise to assemble the "right" data to mine/analyze

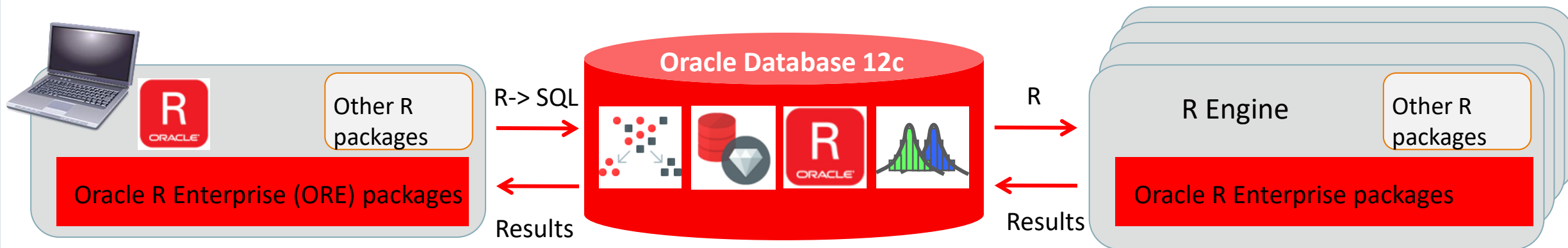
Analytical SQL "Verbs"

- PREDICT
- DETECT
- CLUSTER
- CLASSIFY
- REGRESS
- PROFILE
- IDENTIFY FACTORS
- ASSOCIATE



Oracle Advanced Analytics

How Oracle R Enterprise Compute Engines Work



1 R-> SQL Transparency “Push-Down”

- R language for interaction with the database
- R-SQL Transparency Framework overloads R functions for scalable in-database execution
- Function overload for data selection, manipulation and transforms
- Interactive display of graphical results and flow control as in standard R
- Submit user-defined R functions for execution at database server under control of Oracle Database

2 In-Database Adv Analytical SQL Functions

- 30+ Powerful data mining algorithms (regression, clustering, AR, DT, etc._)
- Run Oracle Data Mining SQL data mining functioning (ORE.odmSVM, ORE.odmDT, etc.)
- Speak “R” but executes as proprietary in-database SQL functions—machine learning algorithms and statistical functions
- Leverage database strengths: SQL parallelism, scale to large datasets, security
- Access big data in Database and Hadoop via SQL, R, and Big Data SQL

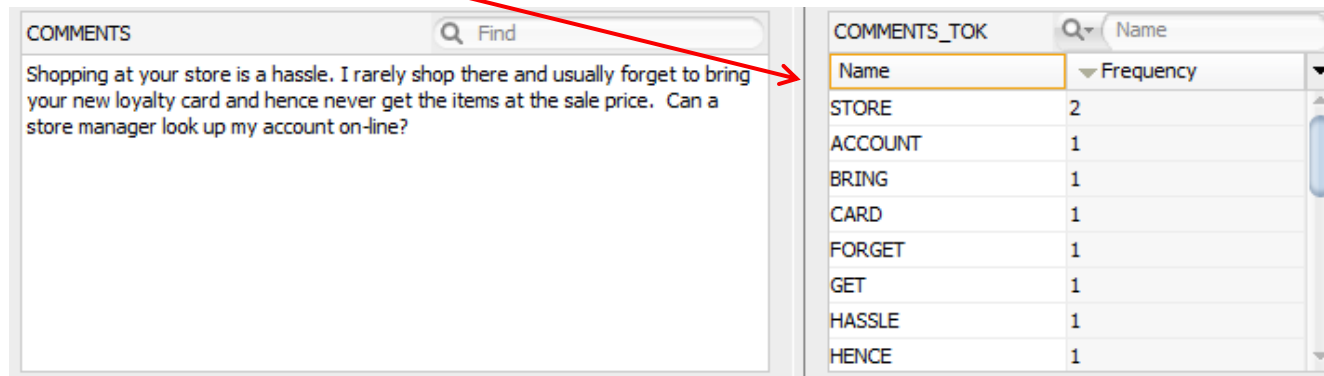
3 Embedded R Package Callouts

- R Engine(s) spawned by Oracle DB for database-managed parallelism
- ore.groupApply high performance scoring
- Efficient data transfer to spawned R engines
- Emulate map-reduce style algorithms and applications
- Enables production deployment and automated execution of R scripts

Oracle Text

Native Capability of every Oracle Database

- Oracle Text uses standard SQL to index, search, and analyze text and documents stored in the Oracle database, in files, and on the web.
- Oracle Text supports multiple languages and uses advanced relevance-ranking technology to improve search quality.
- Oracle Advanced Analytics leverages Oracle Text to pre-process (“**tokenize**”) unstructured data for the OAA **SQL ML/data mining functions**

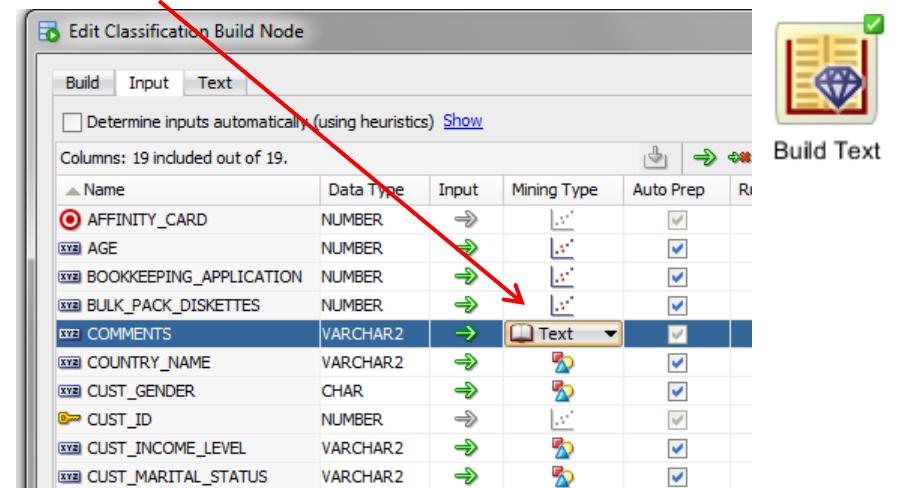


COMMENTS Find

Shopping at your store is a hassle. I rarely shop there and usually forget to bring your new loyalty card and hence never get the items at the sale price. Can a store manager look up my account on-line?

COMMENTS_TOK Name

Name	Frequency
STORE	2
ACCOUNT	1
BRING	1
CARD	1
FORGET	1
GET	1
HASSLE	1
HENCE	1




Edit Classification Build Node

Build Input Text

Determine inputs automatically (using heuristics) [Show](#)

Columns: 19 included out of 19.

Name	Data Type	Input	Mining Type	Auto Prep	Ri
AFFINITY_CARD	NUMBER	→		<input type="checkbox"/>	
AGE	NUMBER	→		<input checked="" type="checkbox"/>	
BOOKKEEPING_APPLICATION	NUMBER	→		<input checked="" type="checkbox"/>	
BULK_PACK_DISKETTES	NUMBER	→		<input checked="" type="checkbox"/>	
COMMENTS	VARCHAR2	→	Text	<input checked="" type="checkbox"/>	
COUNTRY_NAME	VARCHAR2	→		<input checked="" type="checkbox"/>	
CUST_GENDER	CHAR	→		<input checked="" type="checkbox"/>	
CUST_ID	NUMBER	→		<input checked="" type="checkbox"/>	
CUST_INCOME_LEVEL	VARCHAR2	→		<input checked="" type="checkbox"/>	
CUST_MARITAL_STATUS	VARCHAR2	→		<input checked="" type="checkbox"/>	

 Build Text

Objectives

- Prevent \$200M in losses every year using data to monitor, understand and anticipate fraud

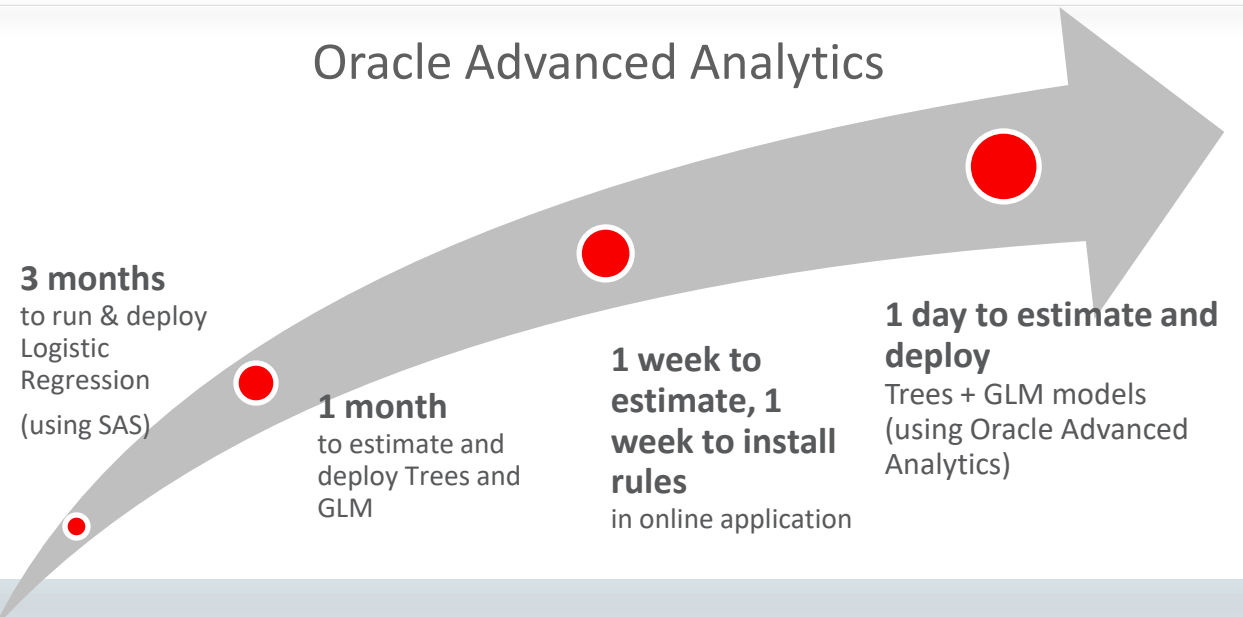
Solution

- We installed OAA analytics for model development during 2014
- When choosing the tools for fraud management, speed is a critical factor
- OAA provided a fast and flexible solution for model building, visualization and integration with production processes

“When choosing the tools for fraud management, speed is a critical factor. Oracle Advance Analytics provided a fast and flexible solution for model building, visualization and integration with production processes.”

- Miguel Barrera, Director of Risk Analytics, Fiserv Inc.
- Julia Minkowski, Risk Analytics Manager, Fiserv Inc.

Oracle Advanced Analytics

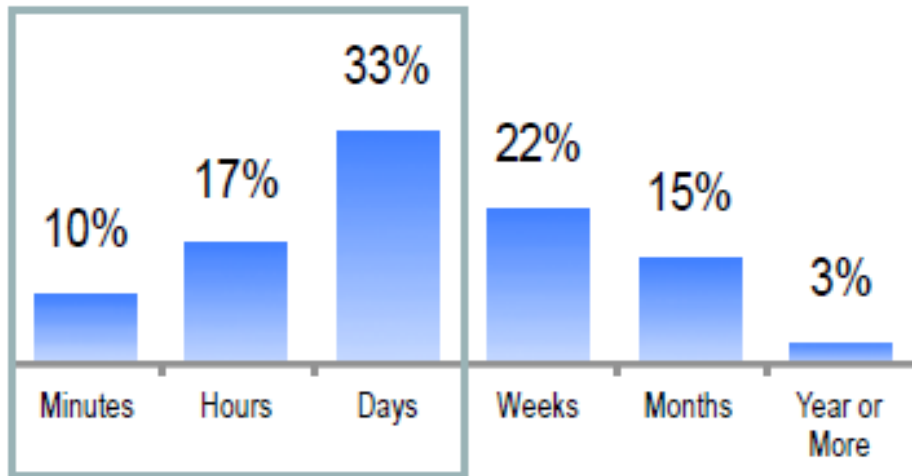


Ease of Deployment

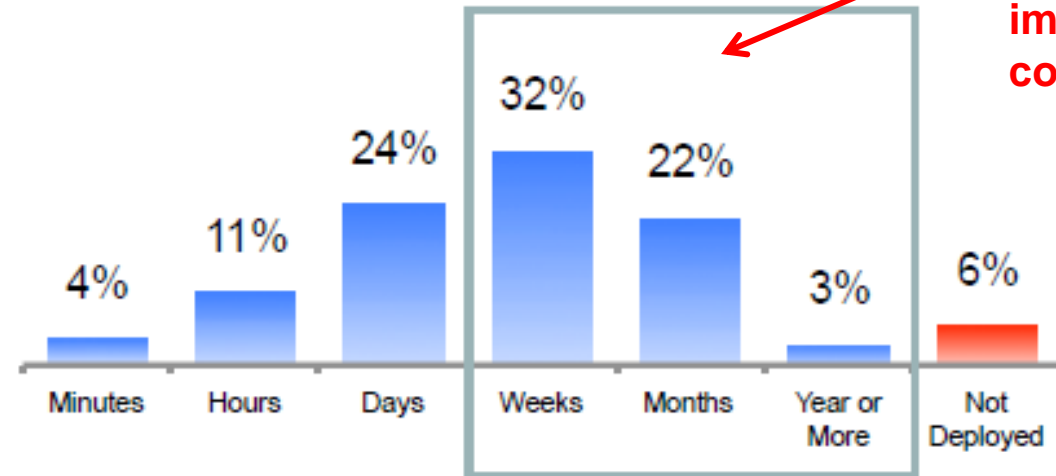
Data Miner Survey 2016 by Rexer Analytics

While 6 out of 10 data miners report the data is available for analysis within days of capture, the time to deploy the models takes substantially longer. For 60% of the respondents the deployment time will range between 3 weeks and 1 year.

Time to Data Analysis



Time to Deployment



Everyone forgets about deployment – but is most important component!





Business Services Authority

UK National Health Service

Combating Healthcare Fraud

Objectives

- Use new insight to help identify cost savings and meet goals
- Identify and prevent healthcare fraud and benefit eligibility errors to save costs
- Leverage existing data to transform business and productivity

Solution

- Identified up to GBP100 million (~~US\$156 million~~) **potentially** saved through benefit fraud and error reduction
- Used anomaly detection to uncover fraudulent activity where some dentists split a single course of treatment into multiple parts and presented claims for multiple treatments
- Analyzed billions of records at one time to measure longer-term patient journeys and to analyze drug prescribing patterns to improve patient care

Update: £300M confirmed fraud
£400+M additional potential identified

Now moving to Cloud

Oracle Exadata Database Machine

Oracle Advanced Analytics



Oracle Exalytics In-Memory Machine

Oracle Endeca Information Discovery

Oracle Business Intelligence EE

Rapidly Build, Evaluate & Deploy Analytical Methodologies

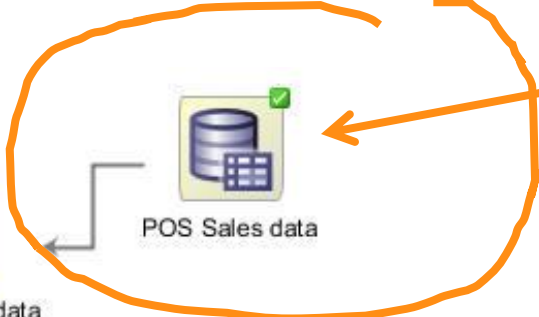
Leveraging a Variety of Data Sources and Types



SQL Joins and arbitrary SQL transforms & queries – power of SQL

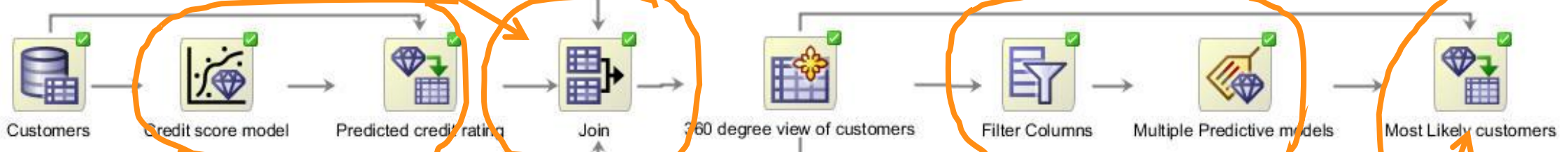


Aggregated POS data



Transactional POS data

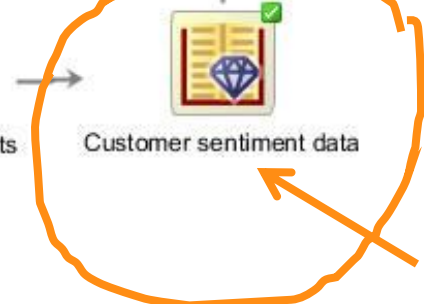
Modeling Approaches



Inline predictive model to augment input data



Demographics and comments



Unstructured data also mined by algorithms

Consider:

- Demographics
- Past purchases
- Recent purchases
- Comments & tweets

Generates SQL scripts and workflow API for deployment



More Data Variety—Better Predictive Models

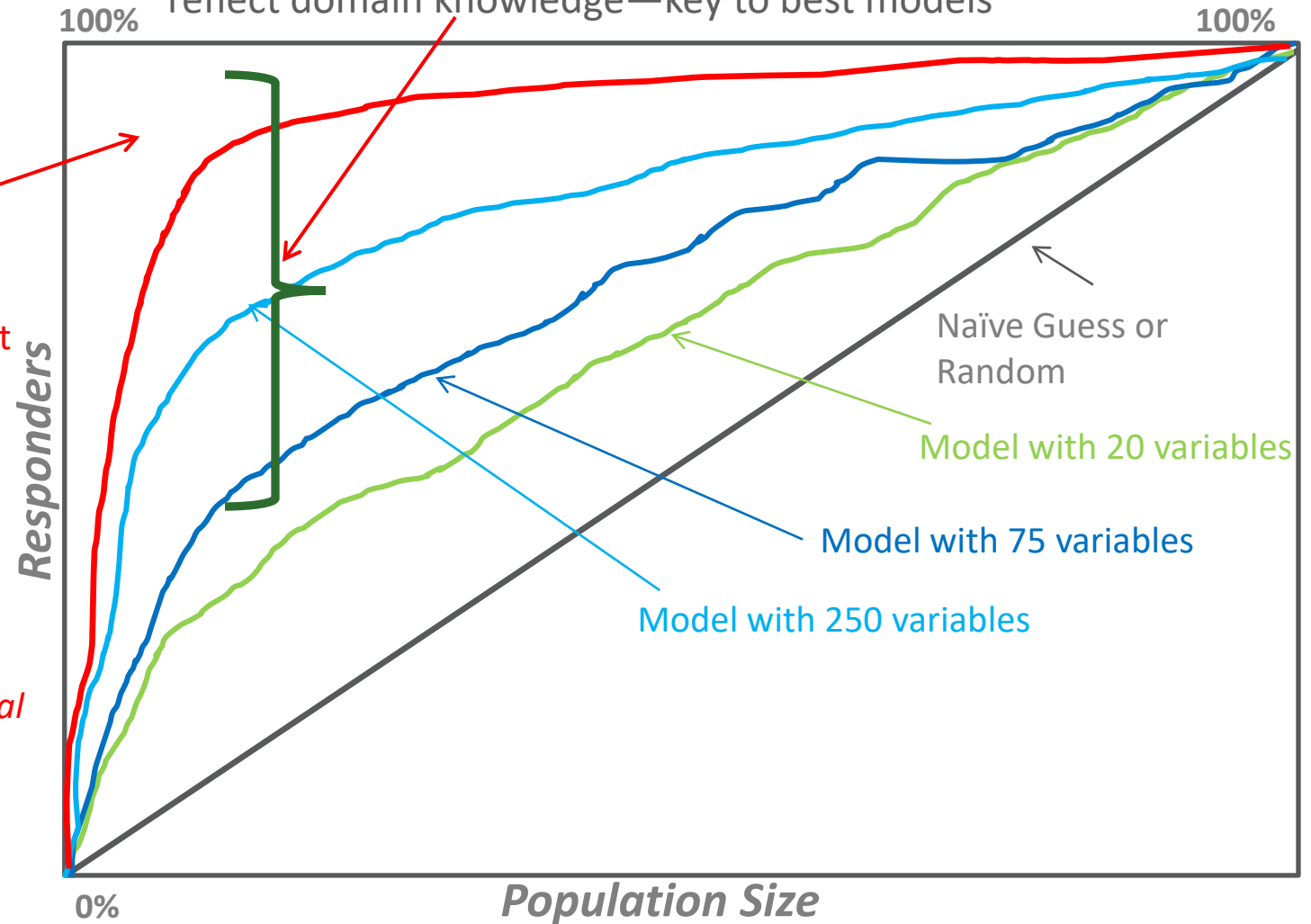
- Increasing sources of relevant data can boost model accuracy



Model with “Big Data” and hundreds -- thousands of input variables including:

- Demographic data
- Purchase POS transactional data
- “Unstructured data”, text & comments
- Spatial location data
- Long term vs. recent historical behavior
- Web visits
- Sensor data
- etc.

Engineered Features – Derived attributes/variable that reflect domain knowledge—key to best models



Big Data Analytics using w Graph

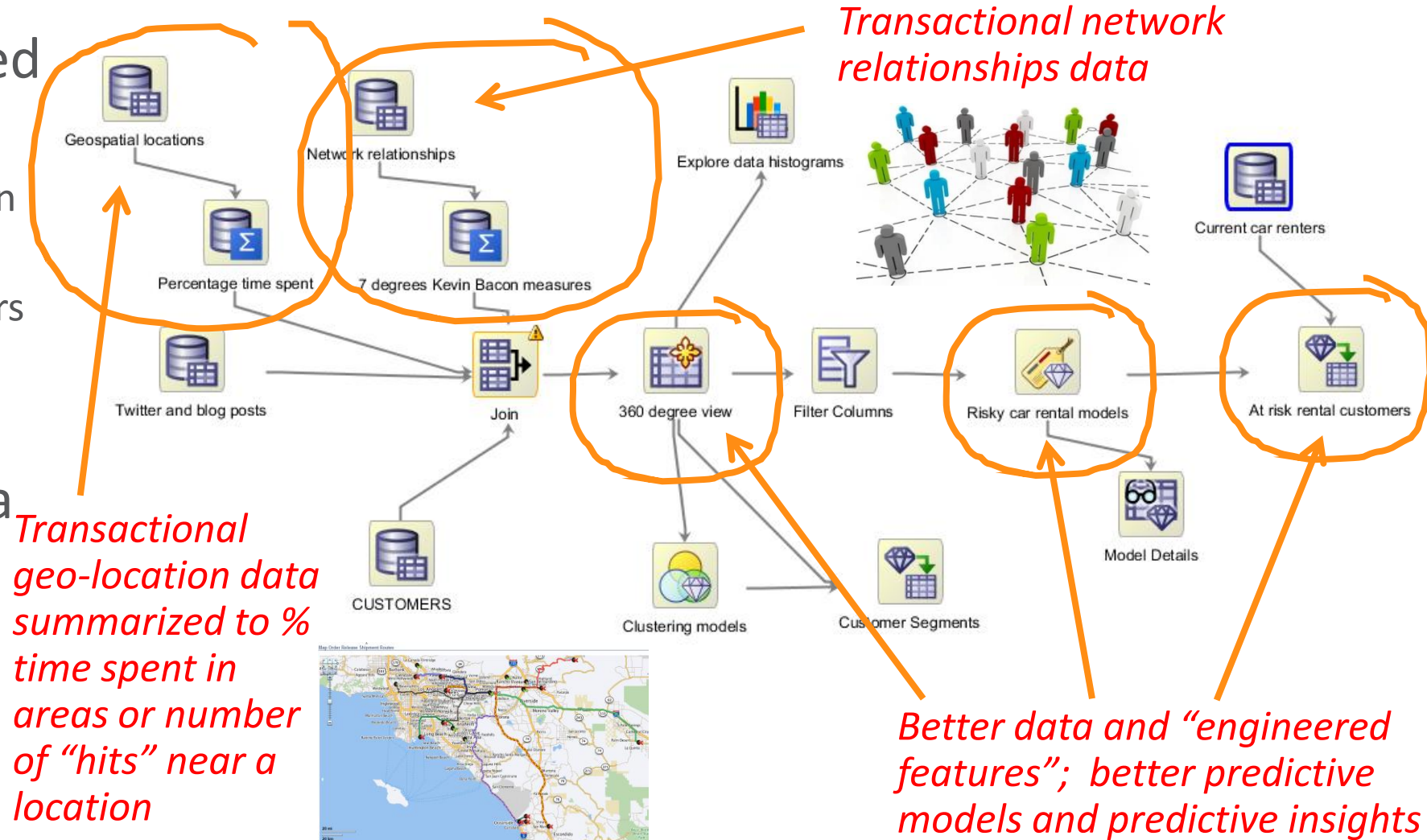
Oracle Advanced Analytics/Machine Learning with Enhanced Graph & Spatial Data Sources

- Add new engineered features

- Percentage time spent in zones
- Amount time/encounters with persons of interest

- Better predictions using available data

- At risk customers
- Government approval processes
- Medical claims
- IoT predictive analytics



Transactional geo-location data summarized to % time spent in areas or number of "hits" near a location

Transactional network relationships data

Better data and "engineered features"; better predictive models and predictive insights



DX Marketing

Cloud Based Predictive Analytics/Database Marketing

DX Marketing

Objectives

- Cloud-based solution
- Increase revenue
- Reduce time-to-market

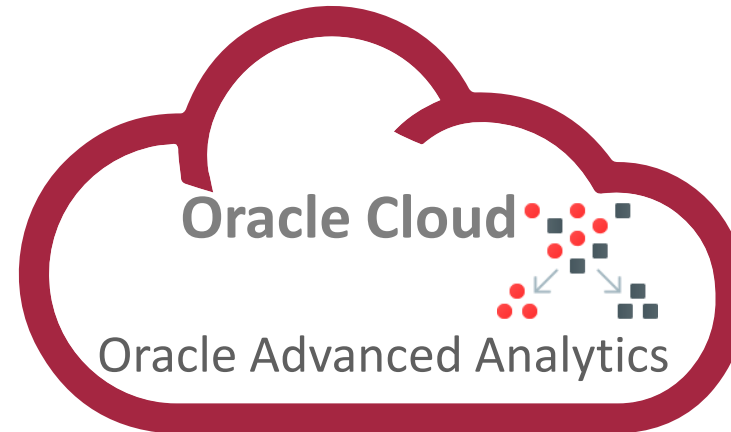
Solution

The company considered only two solution vendors --SAS and Oracle to host its consumer data. SAS offered to help build the IT infrastructure from scratch and helped develop a one-year plan. But when they looked at the number of personnel needed to manage the infrastructure including administrators, security specialists and analysts as well as Security & HIPPA compliance needed, Oracle's DBCS solution looked far more attractive. Hence, they decided to go with Oracle. Oracle's solution offered:

- Scalability
- Built in analytical tools including data mining.
- Built in HIPPA compliance and security features.
- Required fewer resources --only two analysts --Data Engineer and an expert in Predictive Analytics who now manage the entire eco system.

“Time to market has significantly improved from 4-6 weeks to less than a week with the result the company can bring new clients on board faster. This has helped boost revenues by 25% in the six months since using Oracle's DBCS..”

– DX Marketing



[DX Marketing Expands Customer Acquisition with Oracle Cloud](#) – YouTube video

Zagrebačka Bank (biggest bank in Croatia)



Increases Cash Loans by 15% Within 18 Months of Deployment

Objectives

- Needed to speed up entire advanced analytics process; data prep was taking 3 days; model building 24 hours
- Faster time to “actionable analytics” for Credit Risk Modeling and Targeted Customer Campaigns

Solution

- Zaba migrated from SAS to the Oracle Advanced Analytics platform for statistical modeling and predictive analytics
- Increased prediction performance by leveraging the security, reliability, performance, and scalability of Oracle Database and Oracle Advanced Analytics for predictive analytics—running data preparation, transformation, model building, and model scoring within the database

“With Oracle Advanced Analytics we execute computations on thousands of attributes in parallel—impossible with open-source R. Analyzing in Oracle Database without moving data increases our agility. Oracle Advanced Analytics enables us to make quality decisions on time, increasing our cash loans business 15%.”

– Jadranka Novoselovic, Head of BI Dev., Zagrebačka Bank

“We chose Oracle because our entire data modeling process runs on the same machine with the highest performance and level of integration. With Oracle Database we simply switched on the Oracle Advanced Analytics option and needed no new tools,”

– Sinisa Behin, ICT coordinator at BI Dev. Zagrebačka Bank



[ZabaBank Oracle Customer Snapshot on OTN](#)



Oracle Data Miner GUI

Easy to Use for “Citizen Data Scientist”

- Easy to use to define analytical methodologies that can be shared
- SQL Developer Extension
- Workflow API and generates SQL code for immediate deployment

The screenshot displays the Oracle SQL Developer interface with the Data Miner extension. The main workspace shows a workflow diagram for 'BUY INSURANCE' with nodes: 'CUST_INSUR_LTV1', 'Filter Columns', 'Multiple Classification Models', 'Most Likely Customers', and 'LIKELY_BUY_INSURANCE_CUSTMRS1'. A 'Worksheet' window shows the following SQL code:

```
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,
```

The 'Query Result' window shows the following data:

POLICYNUMBER	PERCENT_FRAUD	RNK	
1	654	61.87	1
2	11068	57.37	2
3	7435	55.47	3

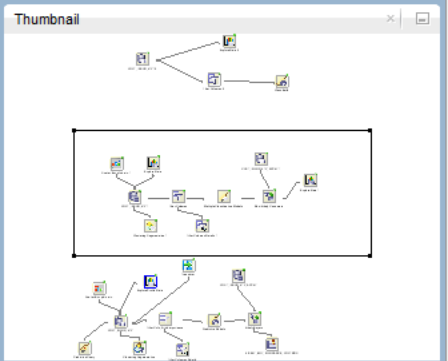
The 'Rule' window shows a classification rule:

```
Rule:
If BANK_FUNDS > 246
And CHECKING_AMOUNT > 282
And MONEY_MONTHLY_OVERDRAWN <= 54.215
Then No
Confidence 0.8515671200473093
Support 0.1832863646217212
```



Connections Data Miner

- Connections
- BIWA16_12c_DB_Cloud_Jonathan2
- BIWA DMUSER - Learn Predictive Analytics in 2 Hours HO
- CBERGER dbpm23 Cloud instance Connection
- Charlie Database 12c laptop dmuser_dmuser
 - AAA Customer Analytics
 - 360 View Star Schema Analytics
 - BUY INSURANCE
 - Claims Anomaly Detection
 - Customer 360 Clustering
 - Customer analytics NEW
 - Fun Analytics WF
 - Market Basket Analysis
 - My First Workflow
 - New Buy Insurance WF
 - New Customer Analytics
 - New WF
 - Predictive Queries
 - RFM input4 Customer Analytics
 - Structured Unstructured Data



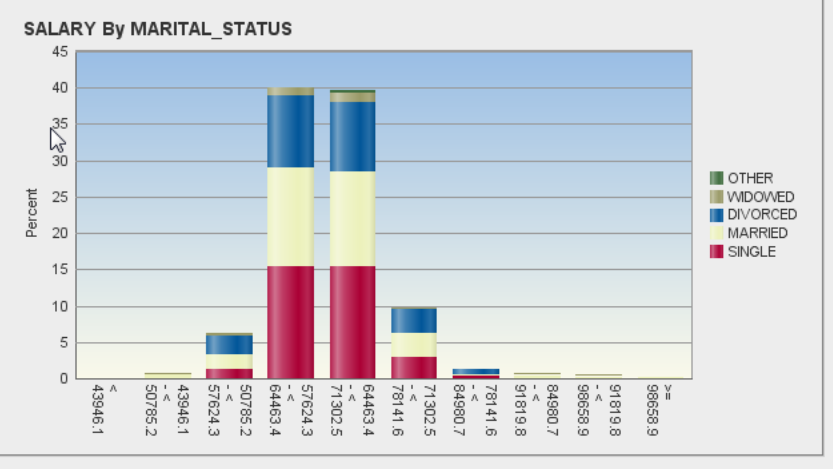
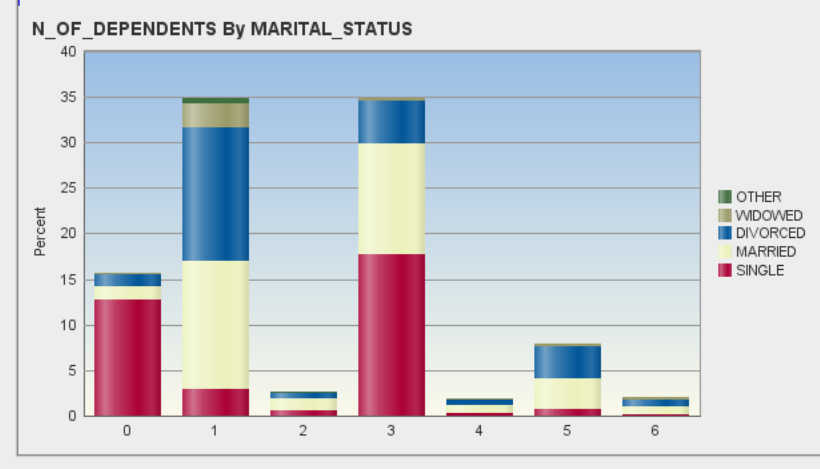
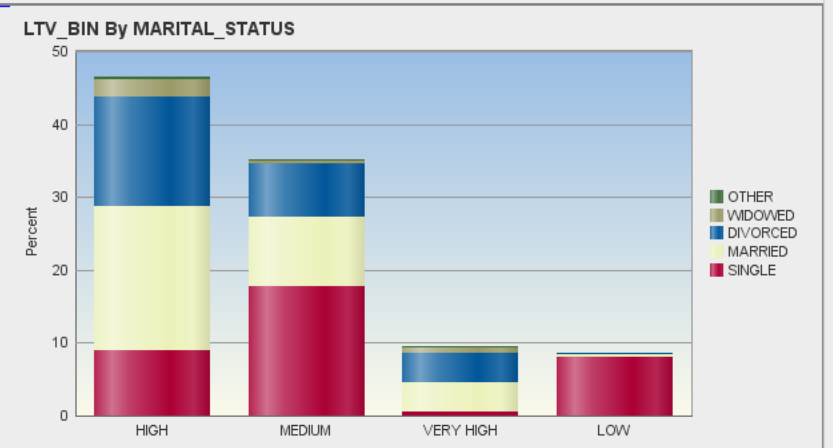
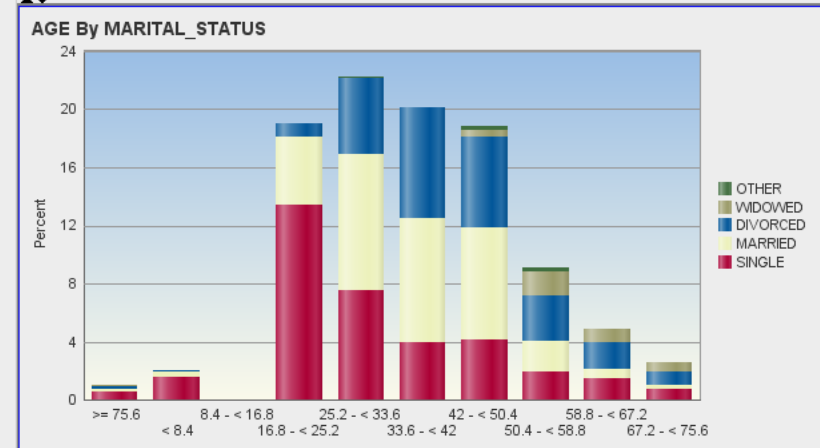
SSH Hosts

- SSH Hosts
- CBERGER dbpm23 Cloud instance
- Express Cloud PDB May 25
- HOL3
- NEW_SSH_BIWA_Brian2
- OracleCloudDatabase 1
- OracleDatabaseCloud2
- SSH_BIWA_Jonathan2

Start Page BUY INSURANCE Charlie Database 12c laptop dmuser_dmuser2.sql Profile Data

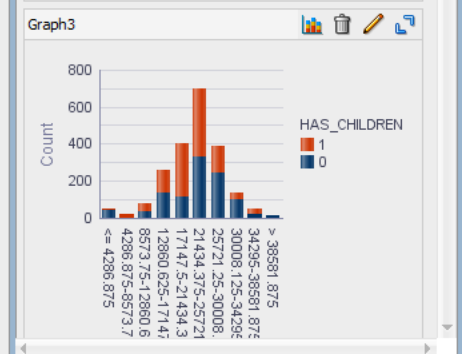
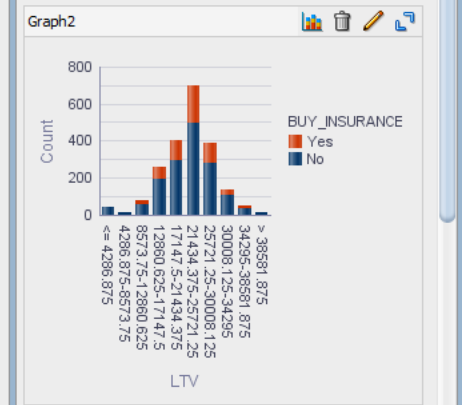
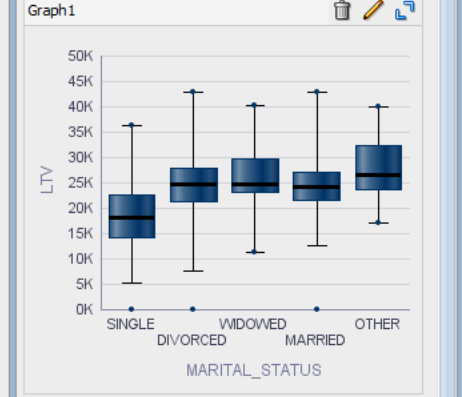
Statistics: 10 Columns from 2,005 Rows(Sampled) Group by: MARITAL_STATUS Show Nulls Filter: Name

Name	Histogram	Data Type	Percent NULLs	Distinct Values	Distinct Per...	Mode	Average	Median	Min Value	Max Value	Standard Devi...	Variance
AGE		NUMBER	0	69	3.4414		37.6823	36	0	84	14.4737	209.4873
BANK_FUNDS		NUMBER	0	425	21.197		2,585.3736	551	0	30,200	4,739.2687	22,460,667.7...
BUY_INSURANCE		VARCHAR2	0	2	0.0998	No						
CAR_OWNERSHIP		NUMBER	0	2	0.0998		0.9247	1	0	1	0.264	0.0697
CHECKING_AMOUNT		NUMBER	0	626	31.2219		1,032.7312	25	25	24,471	3,103.6812	9,632,836.6997



Scatter Box plots etc. Components

View: Sample Data Parallel Query Off...



Connections Data Miner Start Page BUY_INSURANCE_WORKFLOW CUST_INSUR_LTV1 Graph 1

100% Parallel Query Off

Components Workflow Editor Data

Data Source Explore Data SQL Query Update Table

Edit Filter Columns Node

Show Attribute Importance

Show Data Quality

Settings

Columns All None Name

Name	Type	Output	Rank	Importance	% Null	% Unique	% Constant	Hints
BANK_FUNDS	NUMBER	→	1	0.2039	0	21.3511	35.6016	
N_TRANS_ATM	NUMBER	→	2	0.1217	0	0.4438	21.499	
N_TRANS_TELLER	NUMBER	→	3	0.1213	0	0.4931	32.3471	
MONEY_MONTHLY_OVERDRAWN	NUMBER	→	4	0.1207	0	18.787	16.3215	
T_AMOUNT_AUTOM_PAYMENTS	NUMBER	→	5	0.1054	0	59.9606	21.5483	
MONTHLY_CHECKS_WRITTEN	NUMBER	→	6	0.0848	0	0.9369	19.428	
N_OF_DEPENDENTS	NUMBER	→	7	0.0315	0	0.3452	34.4181	
TIME_AS_CUSTOMER	NUMBER	→	8	0.0221	0	0.2465	31.3116	
CHECKING_AMOUNT	NUMBER	→	9	0.0183	0	30.3254	62.7712	
CREDIT_BALANCE	NUMBER	→	10	0.0135	0	9.4181	90.5819	
N_TRANS_KIOSK	NUMBER	→	11	0.0084	0	0.5424	43.6884	
MORTGAGE_AMOUNT	NUMBER	→	12	0.0073	0	21.2525	23.1262	
SEX	VARCHAR2	→	13	0.0064	0	0.0986	66.075	
MARITAL_STATUS	VARCHAR2	→	14	0.0056	0	0.2465	34.5168	
HOUSE_OWNERSHIP	NUMBER	→	15	0.004	0	0.1479	71.1538	

Help OK Cancel

Thumbnail

BUY_INSURANCE_WORKFLOW

- Clustering Segmentation
- Explore/Profile Data
- Explore Data
- Prediction Queries by Region
- Prediction Query
- Filter Cols_Attrib Importance
- Predictive Models for Student
- Class Build
- Predictive models ind TEXT
- Filter Columns Details
- Scatter Box plots etc.
- Graph
- Graph 1
- Anomaly Detection Query
- CUST_INSUR_LTV
- CUST_INSUR_LTV_APPLY
- MINING_DATA_TEXT_BUILD_V
- CUST_INSUR_LTV1
- Model Details

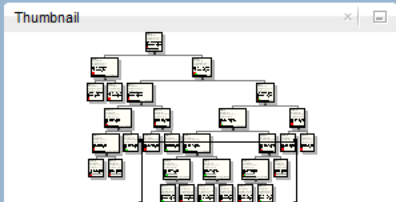
Filter Columns Filter Columns Details Join JSON Query Association Classification Feature Extraction Model Regression

Predictive Queries Evaluate and Apply Linking Nodes



Connections

- dmuser
 - ACME Mfg Paint Project
 - BERGERS R US
 - Fun with Gov
 - Predictive Analytics WF
 - Chicago Crime
 - Customers R Us Project
 - A + Students OAA analytics
 - Big Data Analytcs w JSON
 - Big Data Analytics
 - BUY_INSURANCE_WORKFLOW
 - CARS_DATA_MINING
 - Churners01 work flow
 - Claims Fraud Clustering + SVM2
 - Customer Analytics ind RFM
 - Employees_attrition
 - Insurance Customer Analytics
 - Manufacturing Painting Ops
 - Market Basket Analysis



- CLAS_DT_4_2 - Structure Reports
- Node: 0
 - Node: 1
 - Node: 15
 - Node: 16
 - Node: 2
 - Node: 3
 - Node: 4
 - Node: 5
 - Node: 17
 - Node: 6
 - Node: 18
 - Node: 19
 - Node: 7
 - Node: 8
 - Node: 20
 - Node: 21
 - Node: 9
 - Node: 22
 - Node: 23
 - Node: 10

Tree Settings

100% Maximum Target Values: 2 Save Rules...

Confidence: 86.55%
 No: 1319 (86.55%)
 Yes: 205 (13.45%)

Confidence: 72.46%
 No: 121 (72.46%)
 Yes: 46 (27.54%)

Confidence: 72.08%
 No: 615 (27.92%)
 Yes: 1588 (72.08%)
 Split: MONEY_MONTHLY_OVERDRAWN

Confidence: 64.15%
 No: 823 (64.15%)
 Yes: 460 (35.85%)
 Split: N_TRANS_ATM

Confidence: 82%
 No: 123 (82.00%)
 Yes: 27 (18.00%)

Node: 9
 Prediction: Yes
 Support: 873 (9.46%)
 Confidence: 86.25%
 No: 120 (13.75%)
 Yes: 753 (86.25%)
 Split: T_AMOUNT_AUTOM_PAYMENTS

Node: 8
 Prediction: Yes
 Support: 1330 (14.42%)
 Confidence: 62.78%
 No: 495 (37.22%)
 Yes: 835 (62.78%)
 Split: T_AMOUNT_AUTOM_PAYMENTS

Node: 6
 Prediction: No
 Support: 1218 (13.2%)
 Confidence: 62.4%
 No: 760 (62.40%)
 Yes: 458 (37.60%)
 Split: T_AMOUNT_AUTOM_PAYMENTS

Node: 17
 Prediction: No
 Support: 65 (0.7%)
 Confidence: 96.92%
 No: 63 (96.92%)
 Yes: 2 (3.08%)

Node: 22
 Prediction: Yes
 Support: 851 (9.22%)
 Confidence: 88.25%
 No: 100 (11.75%)
 Yes: 751 (88.25%)

Node: 23
 Prediction: No
 Support: 22 (0.24%)
 Confidence: 90.91%
 No: 20 (90.91%)
 Yes: 2 (9.09%)

Node: 20
 Prediction: Yes
 Support: 1244 (13.48%)
 Confidence: 66%
 No: 423 (34.00%)
 Yes: 821 (66.00%)

Node: 21
 Prediction: No
 Support: 86 (0.93%)
 Confidence: 83.72%
 No: 72 (83.72%)
 Yes: 14 (16.28%)

Node: 19
 Prediction: No
 Support: 418 (4.53%)
 Confidence: 74.88%
 No: 313 (74.88%)
 Yes: 105 (25.12%)

Node: 18
 Prediction: No
 Support: 800 (8.67%)
 Confidence: 55.88%
 No: 447 (55.88%)
 Yes: 353 (44.12%)

Rule Surrogates Target Values

Node Rule: Wrap

If BANK_FUNDS > 246
 And CHECKING_AMOUNT <= 282
 And CREDIT_BALANCE <= 2445
 And MONEY_MONTHLY_OVERDRAWN > 54.095
 And T_AMOUNT_AUTOM_PAYMENTS <= 14993
 Then Yes

Confidence	0.882491186839013
Support	0.09223932365055279

Components

No Components



Connections Data Miner

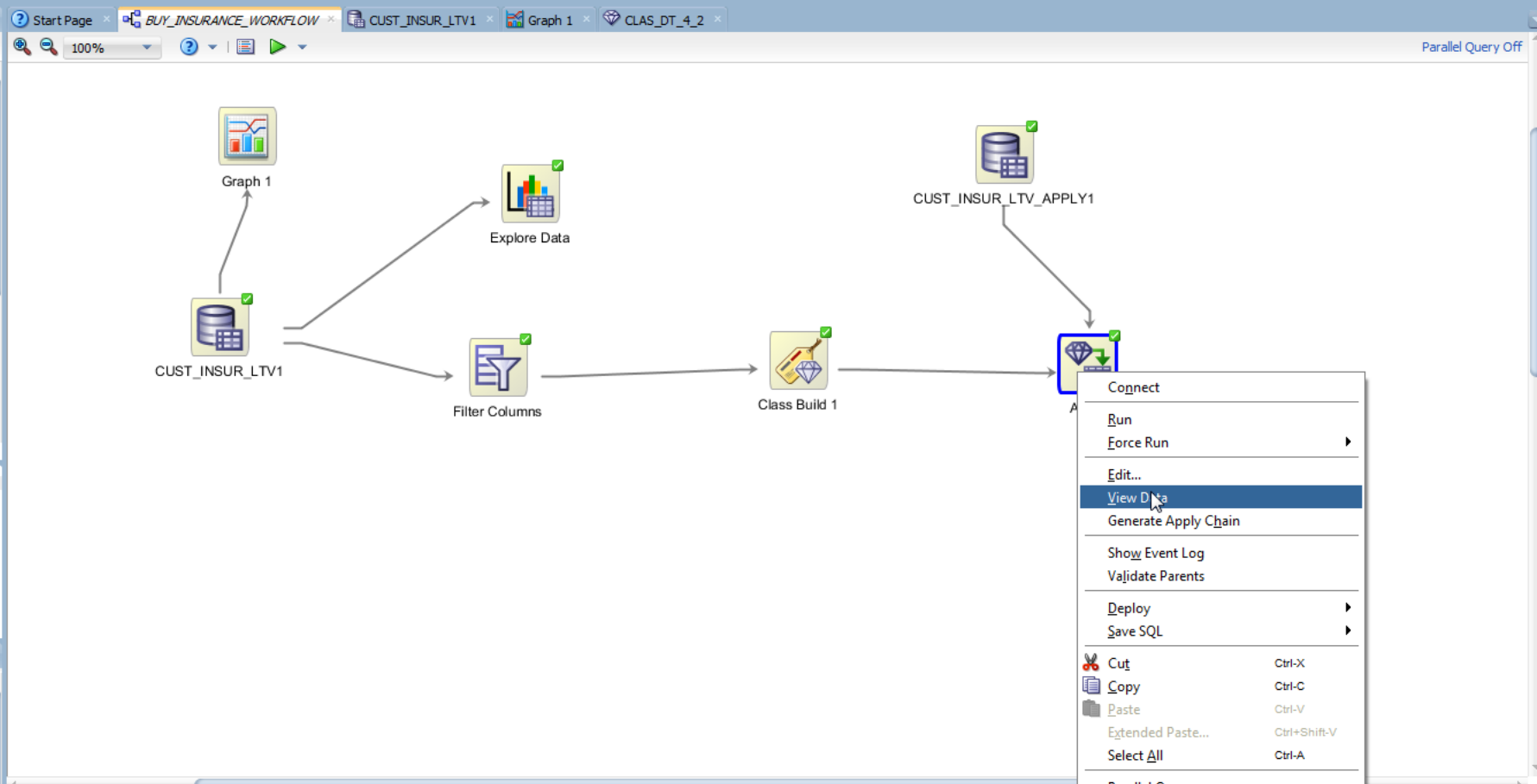
dmuser

- ACME Mfg Paint Project
- BERGERS R US
 - Fun with Gov
 - Predictive Analytics WF
- Chicago Crime
- Customers R Us Project
 - A + Students OAA analytics
 - Big Data Analytics w JSON
 - Big Data Analytics
 - BUY_INSURANCE_WORKFLOW**
 - CARS_DATA_MINING
 - Churners01 work flow
 - Claims Fraud Clustering + SVM2
 - Customer Analytics ind RFM
 - Employees_attrition
 - Insurance Customer Analytics
 - Manufacturing Painting Ops

Thumbnail

BUY_INSURANCE_WORKFLOW - ... Reports

- Clustering Segmentation
- Explore/Profile Data
- Explore Data
- Prediction Queries by Region
- Prediction Query
- Filter Cols_Attrib Importance
- Filter Columns
- Predictive Models for Student
- Class Build
- Predictive models ind TEXT
- Class Build 1
- Filter Columns Details
- Scatter Box plots etc.
- Graph
- Graph 1
- Anomaly Detection Query
- CUST_INSUR_LTV
- CUST_INSUR_LTV_APPLY
- MINING_DATA_TEXT_BUILD_V
- CUST_INSUR_LTV1



- Connect
- Run
- Force Run
- Edit...
- View Data**
- Generate Apply Chain
- Show Event Log
- Validate Parents
- Deploy
- Save SQL
- Cut Ctrl-X
- Copy Ctrl-C
- Paste Ctrl-V
- Extended Paste... Ctrl+Shift-V
- Select All Ctrl-A
- Parallel Query ...
- Copy Image to Clipboard
- Save Image As...
- Go to Properties
- Navigate

Apply - Properties

Additional Output Automatic Settings Case ID: CUST_ID

Cache

Details

Column	Function	Parameter(s)	Model	Node
CLAS_NB_4_2_PROB_Yes	Prediction Probability	Prediction: Yes	CLAS_NB_4_2	Class Build 1
CLAS_NB_4_2_PDET	Prediction Details	Prediction: Yes, Sort: Absolute, Length: 5	CLAS_NB_4_2	Class Build 1

Components

Workflow Editor

Data

- Create Table or View
- Data Source
- Explore Data
- Graph
- SQL Query
- Update Table

Transforms

- Aggregate
- Filter Columns
- Filter Columns

Text

Models

- Anomaly Detection
- Association
- Classification
- Clustering
- Feature Extraction
- Model

Predictive Queries

Evaluate and Apply

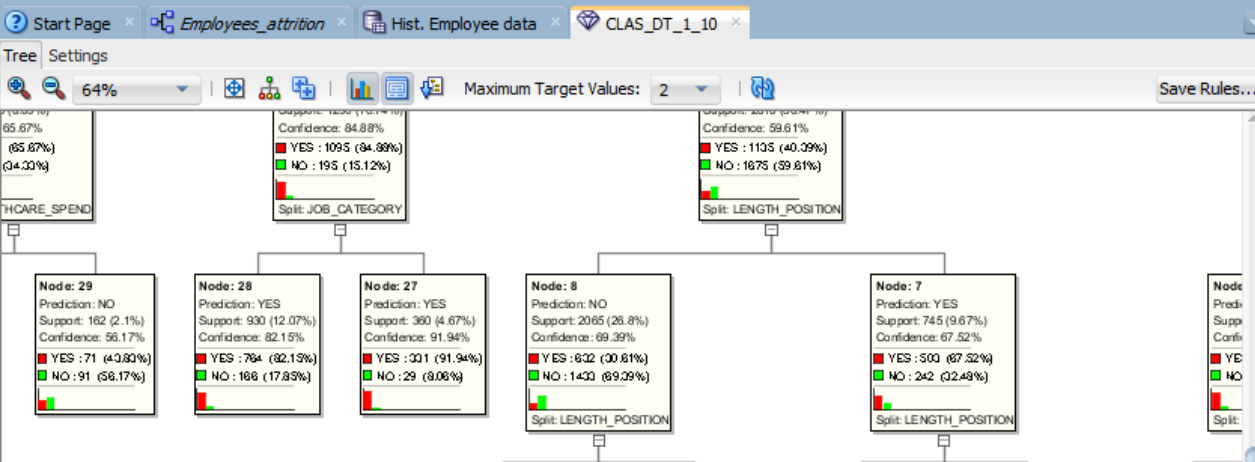
- Apply
- Test

Linking Nodes



Connections Data Miner x

- Customers R Us Project
 - A+ Students OAA analytics
 - Big Data Analytics w JSON
 - Big Data Analytics
 - BUY_INSURANCE_WORKFLOW
 - CARS_DATA_MINING
 - Churners01 work flow
 - Claims Fraud Clustering + SVM2
 - Customer Analytics ind RFM
 - Employees_attrition
 - Insurance Customer Analytics
 - Manufacturing Painting Ops
 - Market Basket Analysis
 - OOW14_CUSTOMER_ANALYTICS_360_
 - PAW Cust Analytics
 - Predictive Queries
 - Public Sector Tax Audit
 - R_regression_models
 - RFM Customer Analytics
 - RFM CUSTOMER ANALYTICS2
 - Star Schema360 Degree Customer workflow for

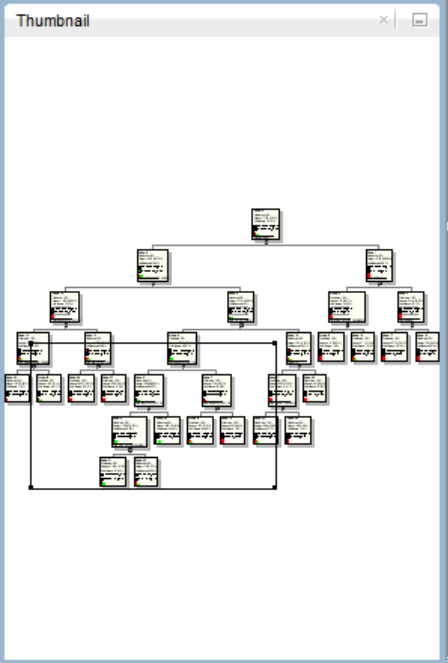


At Risk Employees x

Data Columns | SQL

View: Cache Data Sort... Parallel Query Off... Filter: Enter Where Clause

CLAS_GLM_1_10_PROB_NO	CLAS_GLM_1_10_PDET
1 0.4790120104773802	<Details algorithm="Generalized Linea...
2 0.03718578019711771	<Details algorithm="Generalized Linea...
3 0.1457710019211804	<Details algorithm="Generalized Linea...
4 0.12768928733510213	<Details algorithm="Generalized Linea...
5 0.5698507390589098	<Details algorithm="Generalized Linea...
6 0.8124155905506835	<Details algorithm="Generalized Linea...
7 0.2038408741589256	<Details algorithm="Generalized Linea...
8 0.8968650734058157	<Details algorithm="Generalized Linea...
9 0.6750795593148387	<Details algorithm="Generalized Linea...
10 0.7915741097722724	<Details algorithm="Generalized Linea...



Rule Surrog

Node Rule:

If BONU:
And HEAL:
And STOC:
And 2.5
And AMT:
Then NO

Confidence	0.6183533447684391
Support	0.15131066701271736



Find

```
Model" class="NO">
actualValue="5" weight=".246" rank="1"/>
value="1" weight="-.046" rank="2"/>
="F" weight=".036" rank="3"/>
actualValue="25" weight=".022" rank="4"/>
27" weight=".021" rank="5"/>
```

Close

0587337939665	<Details algorithm="Generalized Linea...
148971355681	<Details algorithm="Generalized Linea...
5366586052554	<Details algorithm="Generalized Linea...
07470752471	<Details algorithm="Generalized Linea...
695885974645	<Details algorithm="Generalized Linea...
31135950403	<Details algorithm="Generalized Linea...
6143770649	<Details algorithm="Generalized Linea...
46131803526	<Details algorithm="Generalized Linea...
71284917635	<Details algorithm="Generalized Linea...
58226130772	<Details algorithm="Generalized Linea...
36 0.43108362457662586	<Details algorithm="Generalized Linea...
37 0.2988898709403637	<Details algorithm="Generalized Linea...
38 0.31307063556114767	<Details algorithm="Generalized Linea...



Search

Untitled



CUST_INSUR_PROFILES

- PROFILE_ID
- PREDICTION
- PREDICTION_COUNT
- RECORD_COUNT
- TOTAL_RECORD_COUNT
- IS_LEAF_NODE
- CONFIDENCE
- SUPPORT
- FULL_SIMPLE_RULE**
- My Calculations
- Value Labels



+ Click here or drag data elements to add a filter

Pie

Trellis Columns

Trellis Rows

Values (Slice)

PREDICTION_CO...

Category

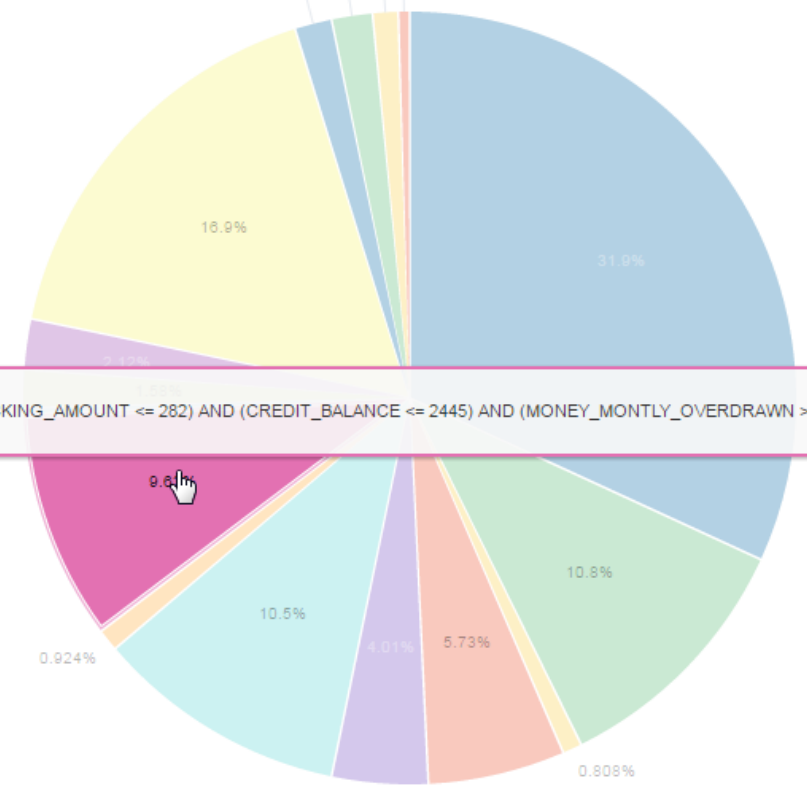
PREDICTION

Color

FULL_SIMPLE_RU..

PREDICTION_COUNT by PREDICTION, FULL_SIMPLE_RULE

0.475%
1.08%
1.71%
1.55%



PREDICTION_COUNT 751.00
FULL_SIMPLE_RULE (BANK_FUNDS > 246) AND (CHECKING_AMOUNT <= 282) AND (CREDIT_BALANCE <= 2445) AND (MONEY_MONTHLY_OVERDRAWN > 54.095) AND (T_AMOUNT_AUTOM_PAYMENTS <= 14993)
PREDICTION Yes

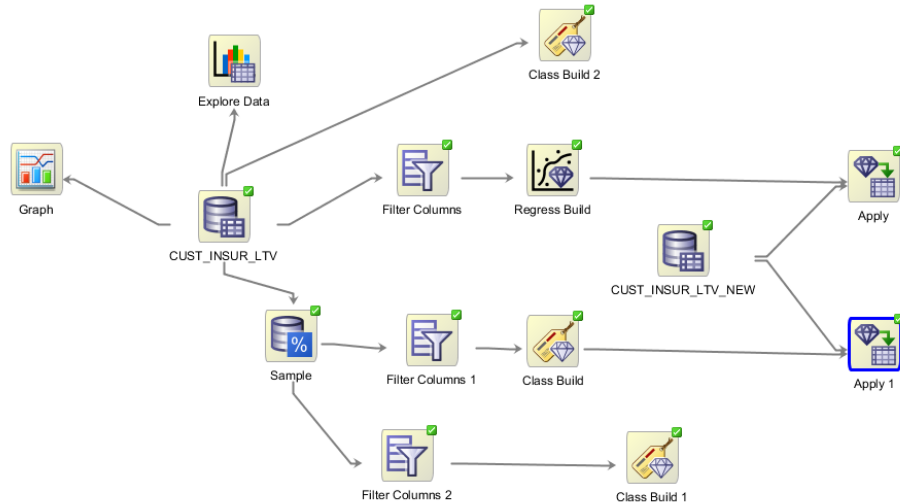
FULL_SIMPLE_RULE

Sharing, Automation and Deployment

Immediately Go to “Productionization” of Analytical Methodologies



- Share ODMr workflows
- Workflow API for 100% automation
 - Immediate deployment of data analyst’s methodologies
- SQL Script Generation
 - Deploy methodology as SQL scripts



Generate SQL Script - Step 2 of 2

Script Directory

Target Database:

Script Directory:

Base Directory: Browse...

Directory Path: C:\SQLDEV Oracle Data Miner Feb 17 2013\sqldeveloper\sqldeveloper\bin\ODM work flow

Name	Date modified	Type	Size
Apply 1.sql	7/24/2013 4:12 PM	SQL File	3 KB
Class Build.sql	7/24/2013 4:12 PM	SQL File	56 KB
CUST_INSUR_LTV.sql	7/24/2013 4:12 PM	SQL File	4 KB
CUST_INSUR_LTV_NEW.sql	7/24/2013 4:12 PM	SQL File	4 KB
Filter Columns 1.sql	7/24/2013 4:12 PM	SQL File	8 KB
Predicting LTV_BEST.png	7/24/2013 4:12 PM	PNG image	64 KB
Predicting LTV_BEST_Drop.sql	7/24/2013 4:12 PM	SQL File	3 KB
Predicting LTV_BEST_Run.sql	7/24/2013 4:12 PM	SQL File	6 KB
Sample.sql	7/24/2013 4:12 PM	SQL File	4 KB

Help < Back Next > Finish Cancel

Fraud Prediction Demo

Automated In-DB Analytical Methodology



```
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 ('2to4', 'morethan4')))
where rnk <= 5
order by percent_fraud desc;
```

	POLICYNUMBER	PERCENT_FRAUD	RNK
1	654	61.87	1
2	11068	57.37	2
3	7435	55.47	3
4	3599	55.4	4
5	14877	55.37	5

Automated Monthly “Application”! *Just add:*

```
Create
View CLAIMS2_30
As
Select * from CLAIMS2
Where mydate > SYSDATE – 30
```

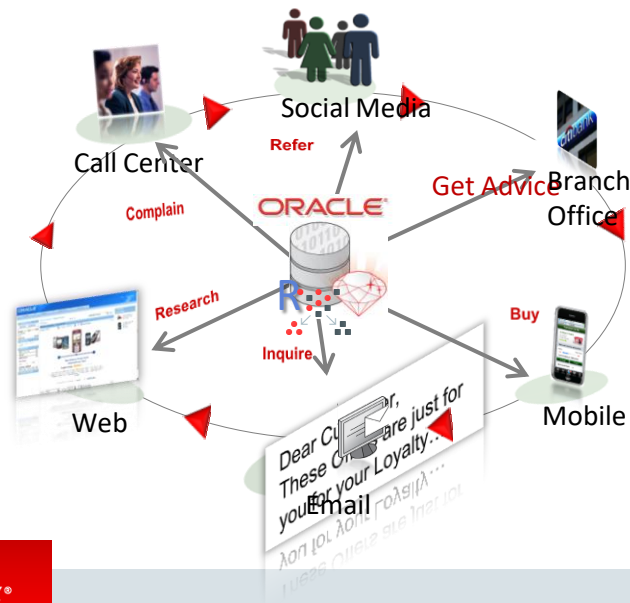
Time measure: set timing on;

Oracle Advanced Analytics

Real-Time Scoring, Predictions and Recommendations

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

```
Select prediction_probability(CLAS_DT_1_5, 'Yes'
  USING 7800 as bank_funds, 125 as checking_amount, 20 as
  credit_balance, 55 as age, 'Married' as marital_status,
  250 as MONEY_MONTHLY_OVERDRAWN, 1 as house_ownership)
from dual;
```



Likelihood to respond:

Query Result	
All Rows Fetched: 1 in 0 seconds	
PREDICTION_PROB...	0.8382936507936...

R: Transparency via function overloading

Invoke in-database aggregation function

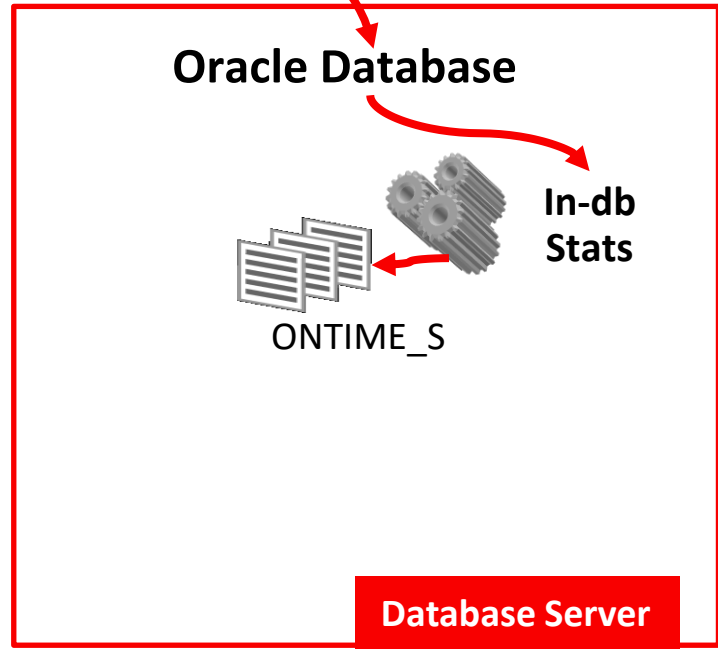
```
R Console
Oracle Distribution of R version 3.0.1 (2012-06-22) -- "Good Sport"

> aggdata <- aggregate(ONTIME_S$DEST,
+                      by = list(ONTIME_S$DEST),
+                      FUN = length)

> class(aggdata)
[1] "ore.frame"
attr(,"package")
[1] "OREbase"
> head(aggdata)
  Group.1    x
1 ABE      237
2 ABI       34
3 ABQ     1357
4 ABY       10
5 ACK        3
6 ACT       33
```



```
Oracle SQL
select DEST, count(*)
from ONTIME_S
group by DEST
```



R: Transparency via function overloading

Invoke in-database Data Mining model (Support Vector Machine)

```

R Console
Oracle Distribution of R version 3.0.1 (2012-06-22) -- "Good Sport"
> svm_mod <- ore.odmSVM(BUY~INCOME+YRS_CUST+MARITAL_STATUS, data=CUST,
                        "classification", kernel="linear")
> summary(svm_mod)
Call:
ore.odmSVM(formula = BUY ~ INCOME + YRS_CUST + MARITAL_STATUS, data = CUST,
            type = "classification", kernel.function = "linear")

Settings:
      prep.auto      on
active.learning al.enable
complexity.factor 46.044899
conv.tolerance    1e-04
kernel.function   linear

Coefficients:
  class  variable value estimate
1      0      INCOME  5.204561e-05
2      0 MARITAL_STATUS M -4.531359e-05
3      0 MARITAL_STATUS S  4.531359e-05
4      0      YRS_CUST  1.264948e-04
5      0 (Intercept)  9.999269e-01
6      1      INCOME  2.032340e-05
7      1 MARITAL_STATUS M  2.636552e-06
8      1 MARITAL_STATUS S -2.636555e-06
9      1      YRS_CUST -1.588211e-04
10     1 (Intercept) -9.999324e-01
    
```

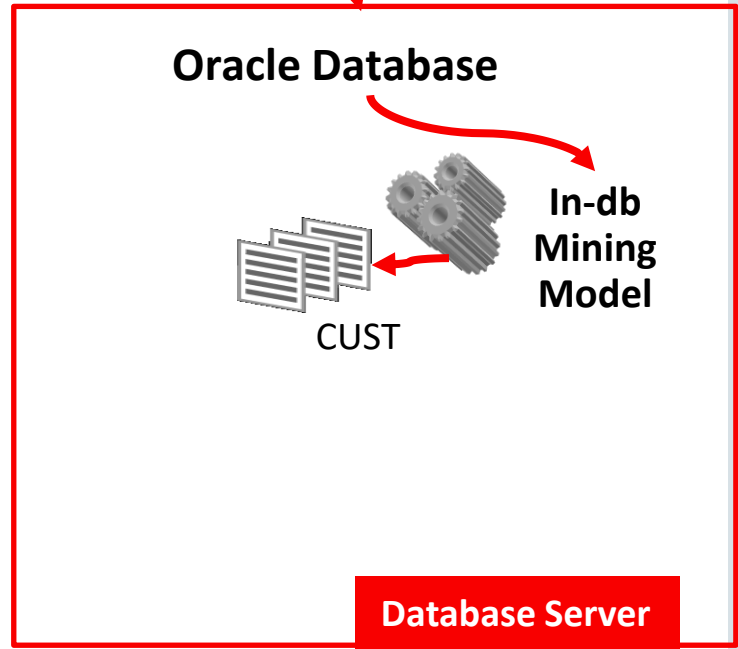
Oracle Advanced Analytics
ORE Client Packages

Transparency Layer

Oracle PL/SQL

```

BEGIN
DBMS_DATA_MINING.CREATE_MODEL (
 model_name => ' SVM_MOD' ,
 mining_function =>
 dbms_data_mining.classification
 ...
    
```

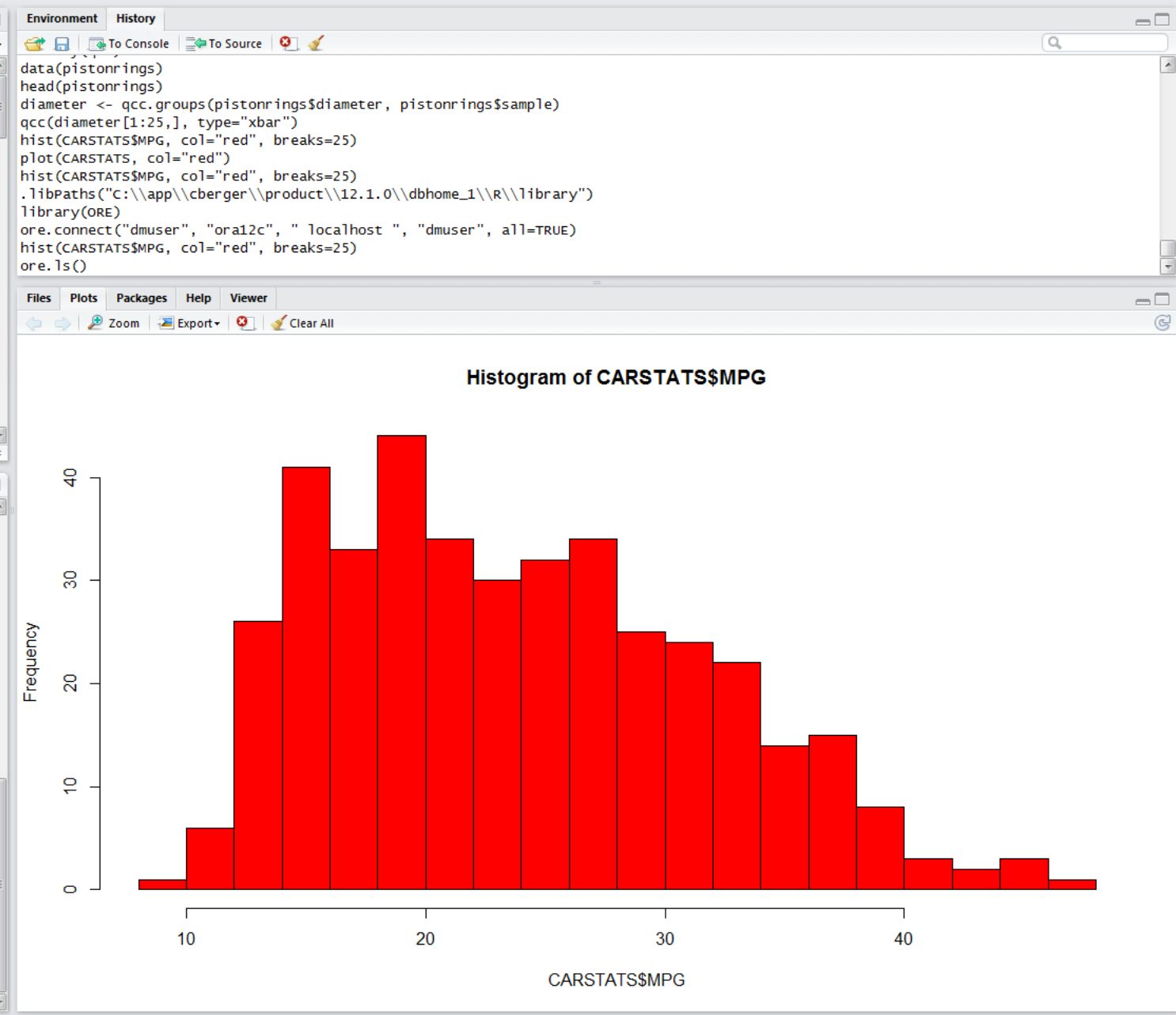


```

1 .libPaths("C:\\app\\cberger\\product\\12.1.0\\dbhome_1\\R\\library")
2 library(ORE)
3
4 ore.connect("dmuser", "ora12c", " localhost ", "dmuser", all=TRUE)
5
6 ore.ls()
7 help.start()
8
9 names(CARSTATS)
10
11
12 hist(CARSTATS$MPG, col="red", breaks=25)
13
14
15 plot(CARSTATS)
16 plot(CARSTATS, col="red")
17
18 demo(odm_kmeans)
19
20
21
22
23
24 names(CLAIMS)
25 summary(CLAIMS)
6:9 (Top Level)
    
```

```

Console ~/
Loading required package: OREmodels
Loading required package: OREdm
Loading required package: lattice
Loading required package: OREpredict
Loading required package: ORExml
>
> ore.connect("dmuser", "ora12c", " localhost ", "dmuser", all=TRUE)
Loading required package: ROracle
Loading required package: DBI
The following object is masked _by_ .GlobalEnv:
  IRIS
There were 39 warnings (use warnings() to see them)
> hist(CARSTATS$MPG, col="red", breaks=25)
> ore.ls()
 [1] "CARSTATS"           "CENSUS"           "CHICAGO_CRIMES"
 [4] "CHURNERS01"         "CLAIMS"           "CLAIMS_FRAUD"
 [7] "CLAIMS_SET"         "CRIME_SCENES_DATA" "CUST_INSUR_LTV"
[10] "CUST_INSUR_LTV_APPLY" "EMPL_DATA"        "EMPL_DATA_JUNE07"
[13] "EXPENSE_VIEW_T"     "INSUR_CUST_LTV_SAMPLE" "IRIS"
[16] "LYMPHOMA"           "MINING_DATA_APPLY_V" "MINING_DATA_BUILD_V"
[19] "MINING_DATA_TEST_V" "MINING_DATA_TEXT_APPLY_V" "MINING_DATA_TEXT_BUILD_V"
[22] "MINING_DATA_TEXT_TEST_V" "NCI60_DATA"       "NCI60_DATA_MAP"
[25] "NEW_360_VIEW_CUSTOMERS" "ODMR_CARS_DATA"   "ODMR_SALES_JSON_DATA"
[28] "ONTIME_SCB"         "OUTPUT_1_24"      "OUTPUT_11_2"
[31] "OUTPUT_13_2"       "OUTPUT_2_22"      "OUTPUT_2_23"
[34] "OUTPUT_2_35"       "OUTPUT_3_15"      "OUTPUT_3_18"
[37] "OUTPUT_3_3"        "OUTPUT_3_40"      "OUTPUT_4_1"
[40] "OUTPUT_5_1"        "OUTPUT_5_15"      "OUTPUT_9_18"
[43] "PAINT_MFG_DEFECTS" "TAX_DATA2"        "TEST_DF1"
[46] "TEST_DF2"          "TITANIC_TEST"     "TITANIC_TRAIN"
    
```




```

1 .libPaths("C:\\app\\cberger\\product\\12.1.0\\dbhome_1\\R\\library")
2 library(ORE)
3
4 ore.connect("dmuser", "ora12c", "localhost", "dmuser", all=TRUE)
5
6 ore.ls()
7 help.start()
8
9 names(CARSTATS)
10
11 summary(CARSTATS)
12
13 hist(CARSTATS$MPG, col="red", breaks=25)
14
15 plot(CARSTATS)
16 plot(CARSTATS, col="red")
17
18 demo(odm_kmeans)
19
20
21
22
23 names(CLAIMS)
24 summary(CLAIMS)
25
15:1 (Top Level)

```

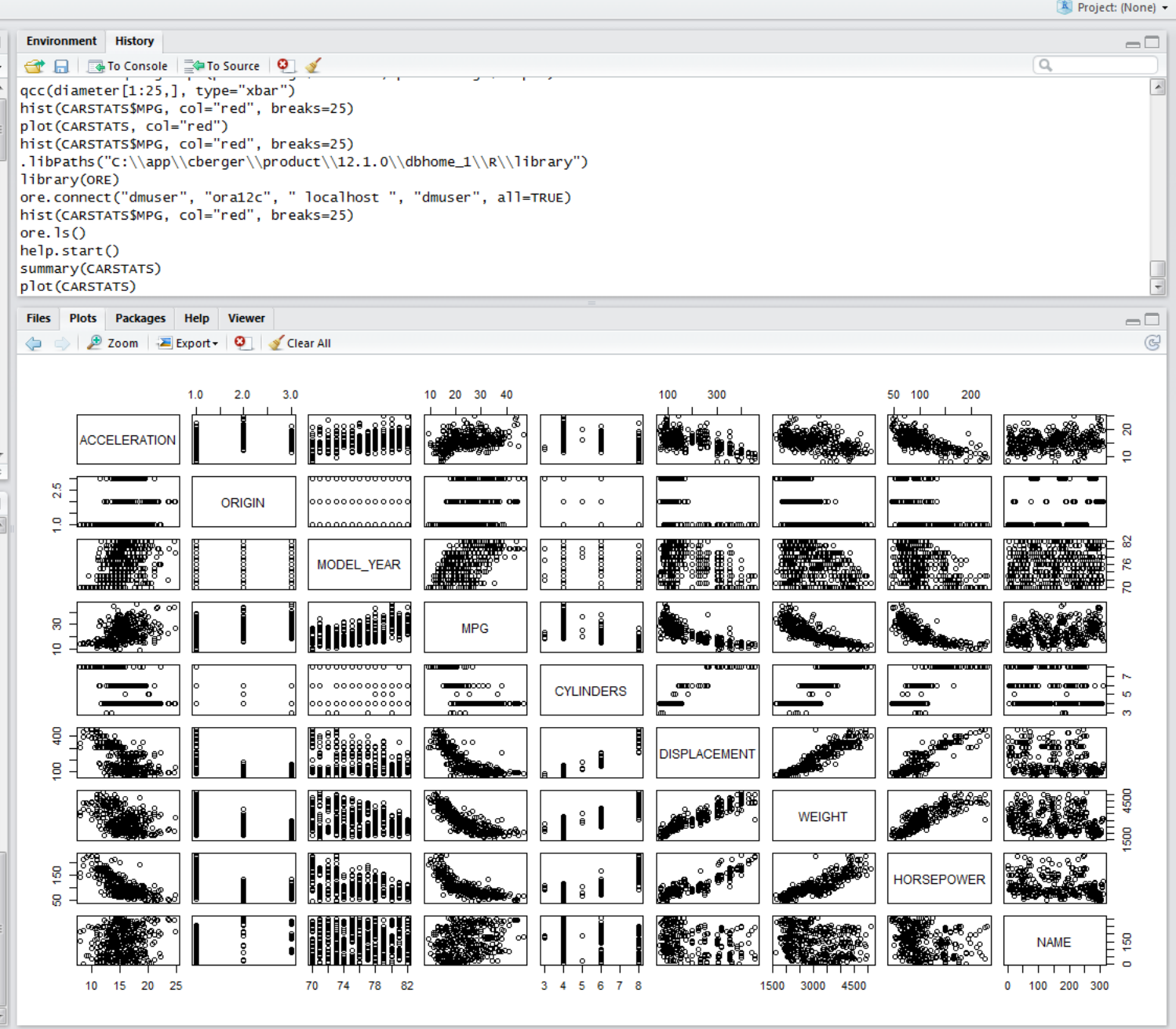
```

Console ~/
[25] "NEW_360_VIEW_CUSTOMERS" "ODMR_CARS_DATA" "ODMR_SALES_JSON_DATA"
[28] "ONTIME_SCB" "OUTPUT_1_24" "OUTPUT_11_2"
[31] "OUTPUT_13_2" "OUTPUT_2_22" "OUTPUT_2_23"
[34] "OUTPUT_2_35" "OUTPUT_3_15" "OUTPUT_3_18"
[37] "OUTPUT_3_3" "OUTPUT_3_40" "OUTPUT_4_1"
[40] "OUTPUT_5_1" "OUTPUT_5_15" "OUTPUT_9_18"
[43] "PAINT_MFG_DEFECTS" "TAX_DATA2" "TEST_DF1"
[46] "TEST_DF2" "TITANIC_TEST" "TITANIC_TRAIN"
> help.start()
If nothing happens, you should open
'http://127.0.0.1:17896/doc/html/index.html' yourself
> summary(CARSTATS)
  ACCELERATION  ORIGIN  MODEL_YEAR  MPG  CYLINDERS
Min. : 8.00  Min. :1.000  Min. :70.00  Min. : 9.00  Min. :3.000
1st Qu.:13.70 1st Qu.:1.000 1st Qu.:73.00 1st Qu.:17.50 1st Qu.:4.000
Median :15.50 Median :1.000 Median :76.00 Median :23.00 Median :4.000
Mean :15.52 Mean :1.569 Mean :75.92 Mean :23.49 Mean :5.475
3rd Qu.:17.18 3rd Qu.:2.000 3rd Qu.:79.00 3rd Qu.:29.00 3rd Qu.:8.000
Max. :24.80 Max. :3.000 Max. :82.00 Max. :46.60 Max. :8.000
NA's :8

  DISPLACEMENT  WEIGHT  HORSEPOWER  NAME
Min. : 68.0  Min. :1613  Min. : 46.00  ford pinto : 6
1st Qu.:105.0 1st Qu.:2226 1st Qu.: 75.75  amc matador : 5
Median :151.0 Median :2822 Median : 95.00  ford maverick : 5
Mean :194.8 Mean :2979 Mean :105.08  toyota corolla : 5
3rd Qu.:302.0 3rd Qu.:3618 3rd Qu.:130.00  amc gremlin : 4
Max. :455.0 Max. :5140 Max. :230.00  amc hornet : 4
NA's :6 (other) :377

> plot(CARSTATS)
warning message:
ORE object has no unique key - using random order
>

```



```

rqruser login DRAFT.R x Manual ORE Setup.R x Hornick Short Demo Script (1).R x
Source on Save Run Source
1 .libPaths("C:\\app\\cberger\\product\\12.1.0\\dbhome_1\\R\\library")
2 library(ORE)
3
4 ore.connect("dmuser", "ora12c", "localhost", "dmuser", all=TRUE)
5
6 ore.ls()
7 help.start()
8
9 names(CARSTATS)
10
11 summary(CARSTATS)
12
13 hist(CARSTATS$MPG, col="red", breaks=25)
14
15 plot(CARSTATS)
16 plot(CARSTATS, col="red")
17
18 demo(odm_svm)
19
20 demo(odm_kmeans)
21
22
23
24
25 names(CLAIMS)
18:1 (Top Level) R Script

```

```

Console -/
> svm.mod2 <- NULL
> svm.mod2 <- ore.odmSVM(Y~X,dat, "regression")
> summary(svm.mod2)

Call:
ore.odmSVM(formula = Y ~ X, data = dat, type = "regression")

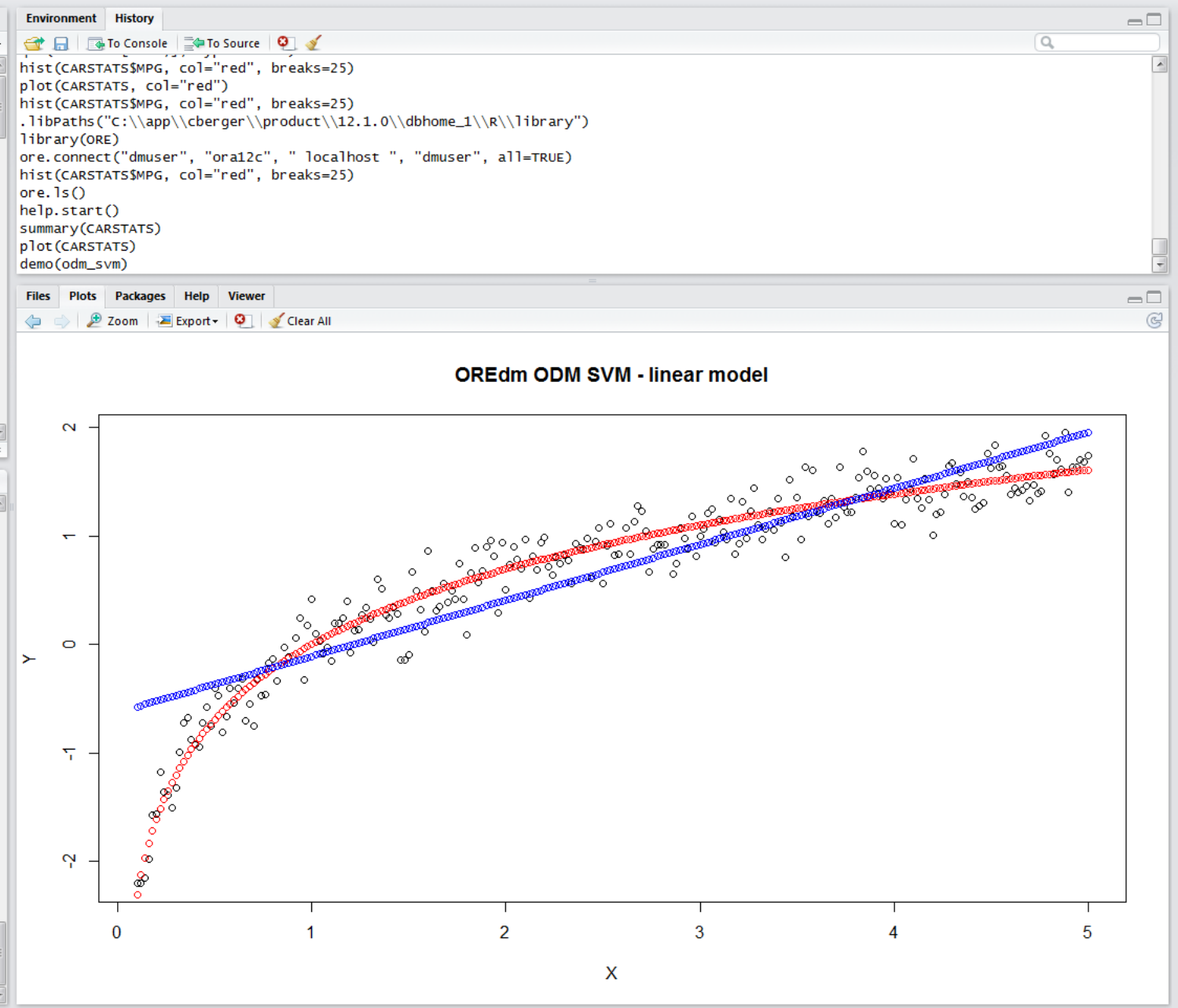
Settings:
      value
prep.auto      on
active.learning a1.enable
complexity.factor 0.614366
conv.tolerance  1e-04
epsilon        0.042761
kernel.cache.size 50000000
kernel.function gaussian
std.dev        0.048980

Residuals:
      Min.      1st Qu.      Median      Mean      3rd Qu.      Max.
-0.530900 -0.147100 -0.012980 -0.008536  0.117600  0.532100

Coefficients:
[1] No coefficients with gaussian kernel

> svm.res2 <- predict(svm.mod2,dat,supplemental.cols="x")
> plot(dat,main="Comparing SVM Model Results")
Hit <Return> to see next plot:

```



ORAAH: Machine Learning in Spark against HDFS data

Invoke ORAAH custom parallel distributed GLM Model using Spark Caching

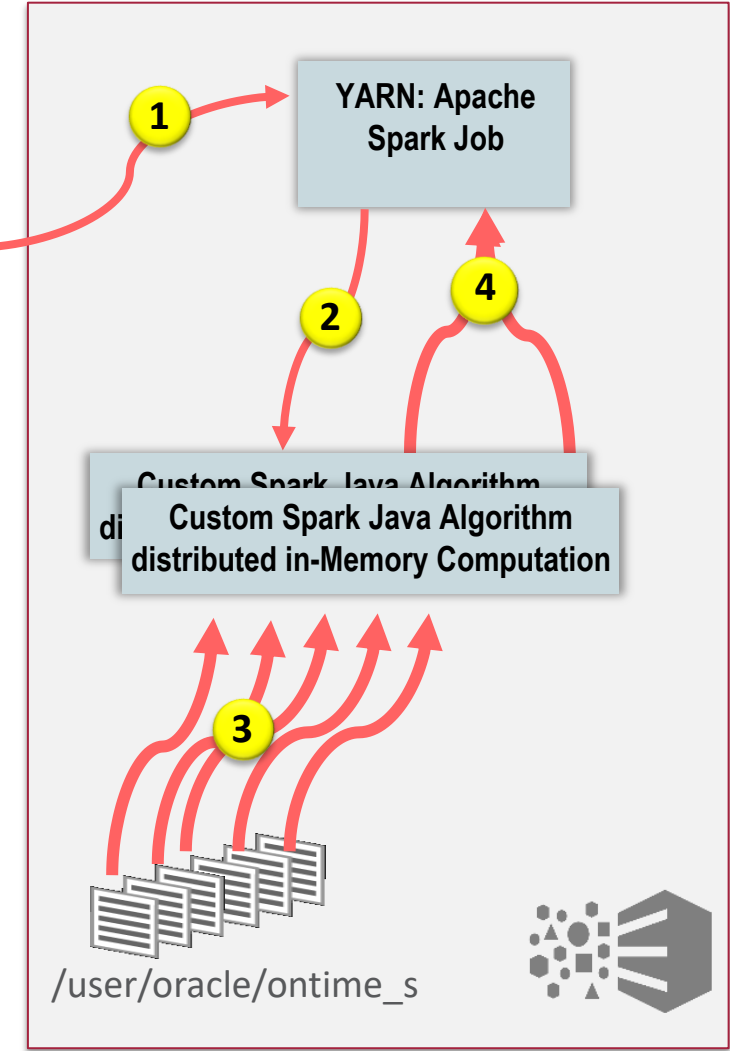
```
R Console
Oracle Distribution of R version 3.1.1 (-- -- "Sock it to Me")
> Connects to Spark
> spark.connect("yarn-client",memory="24g")

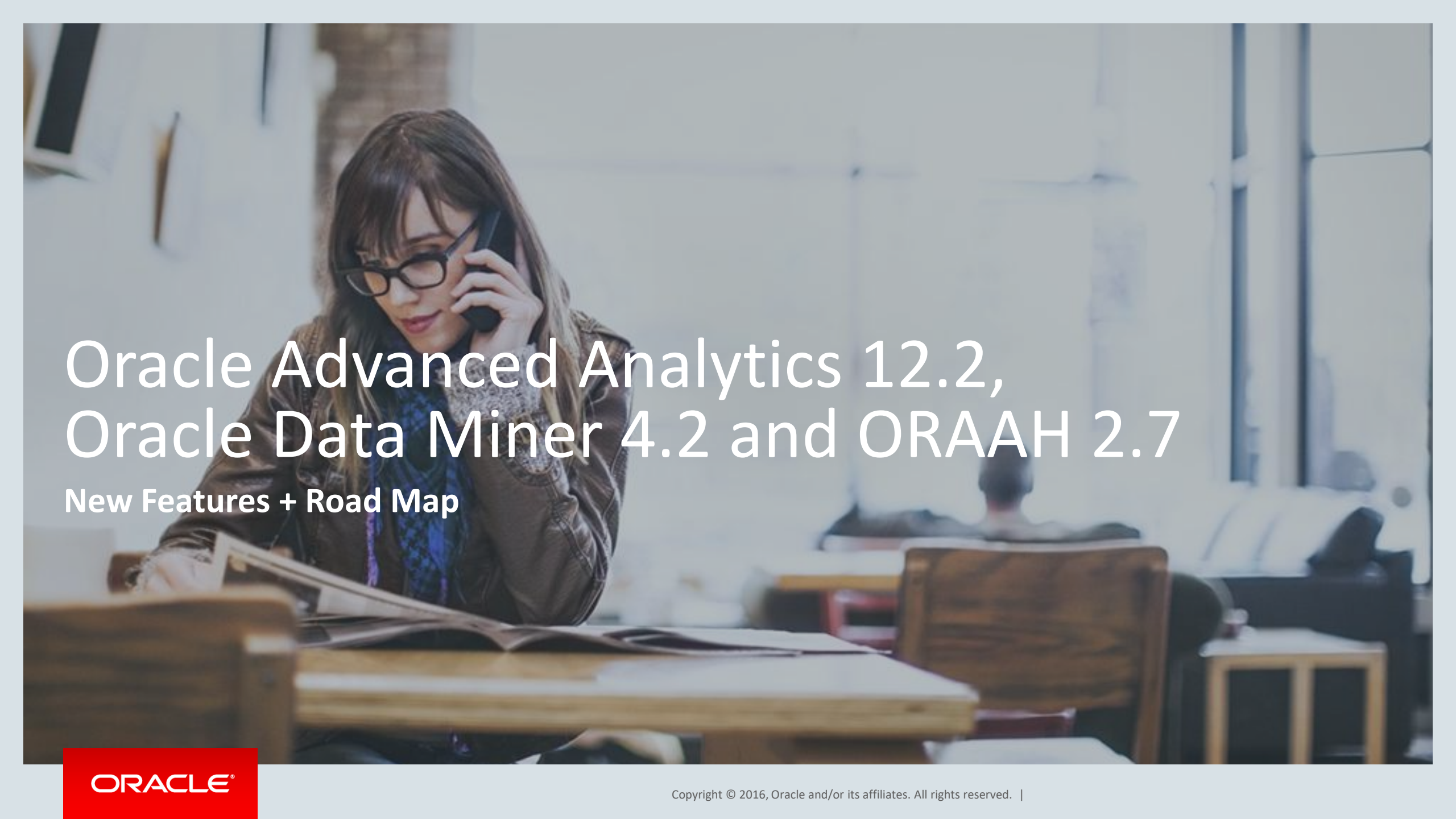
> # Attaches the HDFS file for use within R
> ont1bi <- hdfs.attach("/user/oracle/ontime_1bi")

> # Formula definition: Cancelled flights (0 or 1) based on other attributes
> form_oraa_h_glm2 <- CANCELLED ~ DISTANCE + ORIGIN + DEST + F(YEAR) + F(MONTH) +
+ F(DAYOFMONTH) + F(DAYOFWEEK)
> system.time(m_spark_glm <- orch.glm2(formula=form_oraa_h_glm2, ont1bi))
ORCH GLM: processed 6 factor variables, 25.806 sec
ORCH GLM: created model matrix, 100128 partitions, 32.871 sec
ORCH GLM: iter 1, deviance 1.38433414089348300E+09, elapsed time 9.582 sec
ORCH GLM: iter 2, deviance 3.39315388583931150E+08, elapsed time 9.213 sec
ORCH GLM: iter 3, deviance 2.06855738812683250E+08, elapsed time 9.218 sec
ORCH GLM: iter 4, deviance 1.75868100359263200E+08, elapsed time 9.104 sec
ORCH GLM: iter 5, deviance 1.70023181759611580E+08, elapsed time 9.132 sec
ORCH GLM: iter 6, deviance 1.69476890425481350E+08, elapsed time 9.124 sec
ORCH GLM: iter 7, deviance 1.69467586045954760E+08, elapsed time 9.077 sec
ORCH GLM: iter 8, deviance 1.69467574351380850E+08, elapsed time 9.164 sec
user system elapsed
84.107 5.606 143.591
```

Oracle R Advanced Analytics
for Hadoop Client Packages

Spark-Based Machine
Learning algorithms
module



A woman with long brown hair and glasses is sitting at a wooden table in a cafe. She is wearing a brown leather jacket and a blue patterned scarf. She is holding a black smartphone to her ear with her right hand and looking down at a newspaper or magazine on the table with her left hand. The background is a blurred cafe interior with other tables and chairs.

Oracle Advanced Analytics 12.2, Oracle Data Miner 4.2 and ORAAH 2.7

New Features + Road Map

Oracle Advanced Analytics 12.2

New Oracle Database Features



- **Significant Performance Improvements for all Algorithms**

- New parallel model build / apply redesigned infrastructure to enable faster new algorithm introduction
- Scale to larger data volumes found in big data and cloud use cases

- **Unsupervised Feature Selection**

- Unsupervised algorithm for pair-wise correlations (Kullback-Leibler Divergence (KLD)) for numeric & categorical attributes to find highest “information containing” attributes

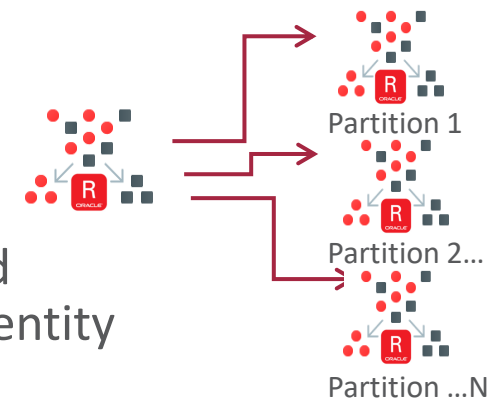
Name	Type	Output	Rank	Importance
HOUSE_OWNERSHIP	NUMBER	→	1	0.3331
N_MORTGAGES	NUMBER	→	2	0.3330
MORTGAGE_AMOUNT	NUMBER	→	3	0.3301
TIME_AS_CUSTOMER	NUMBER	→	4	0.2918
LTV	NUMBER	→	5	0.2900
LTV_BIN	VARCHAR2	→	6	0.2880
N_OF_DEPENDENTS	NUMBER	→	7	0.2870
MARITAL_STATUS	VARCHAR2	→	8	0.2684
N_TRANS_WEB_BANK	NUMBER	→	9	0.2454
N_TRANS_ATM	NUMBER	→	10	0.2394
STATE	VARCHAR2	→	11	0.2335
MONEY_MONTHLY_OVERDRAWN	NUMBER	→	12	0.2293
REGION	VARCHAR2	→	13	0.2286
SEX	VARCHAR2	→	14	0.2057
BANK_FUNDS	NUMBER	→	15	0.1790
N_TRANS_TELLER	NUMBER	→	16	0.1475
T_AMOUNT_AUTOM_PAYMENTS	NUMBER	→	17	0.1216
CAR_OWNERSHIP	NUMBER	→	18	0.1157
BUY_INSURANCE	VARCHAR2	→	19	0.1039
HAS_CHILDREN	NUMBER	→	20	0.0976

- **Association Rules Enhancements**

- Adds calculation of values associated with AR rules such as sales amount to indicate the value of co-occurring items in baskets
- Can filter input items prior to market basket analysis

- **Partitioned Models**

- Instead of building, naming and referencing 10s or 1000s of models, partitioned models organize and represent multiple models as partitions in a single model entity



Oracle Advanced Analytics 12.2

New Oracle Database Features



- **Explicit Semantic Analysis (ESA) algorithm**

- Useful technique for extracting meaningful, interpretable features; better than LDA
- English Wikipedia is Text corpus default to equate tokens with human identifiable features and concepts
- ESA improves text processing, classification, document similarity and topic identification
- Compare documents that may not even mention same topics e.g. al-Qa ida or Osama bin Laden:

Document 1

- *'Senior members of the Saudi royal family paid at least \$560 million to Osama bin Laden terror group and the Taliban for an agreement his forces would not attack targets in Saudi Arabia, according to court documents. The papers, filed in a \$US3000 billion (\$5500 billion) lawsuit in the US, allege the deal was made after two secret meetings between Saudi royals and leaders of al-Qa ida, including bin Laden. The money enabled al-Qa ida to fund training camps in Afghanistan later attended by the September 11 hijackers. The disclosures will increase tensions between the US and Saudi Arabia.'*

Document 2

- *'The Saudi Interior Ministry on Sunday confirmed it is holding a 21-year-old Saudi man the FBI is seeking for alleged links to the Sept. 11 hijackers. Authorities are interrogating Saud Abdulaziz Saud al-Rasheed "and if it is proven that he was connected to terrorism, he will be referred to the sharia (Islamic) court," the official Saudi Press Agency quoted an unidentified ministry official as saying.'*

ESA Similarity Score = 0.62

Oracle Advanced Analytics 12.2

New Oracle Database Features



- **Explicit Semantic Analysis (ESA) algorithm**
- *"The more things change... Yes, I'm inclined to agree, especially with regards to the historical relationship between stock prices and bond yields. The two have generally traded together, rising during periods of economic growth and falling during periods of contraction. Consider the period from 1998 through 2010, during which the U.S. economy experienced two expansions as well as two recessions: Then central banks came to the rescue. Fed Chairman Ben Bernanke led from Washington with the help of the bank's current \$3.6T balance sheet. He's accompanied by Mario Draghi at the European Central Bank and an equally forthright Shinzo Abe in Japan. Their coordinated monetary expansion has provided all the sugar needed for an equities moonshot, while they vowed to hold global borrowing costs at record lows"*
- Top topics (**concepts, people, organizations, events**) discovered by ESA using Wikipedia as model source data
 - *Recession, Ben Bernanke, Lost Decade Japan, Mario Draghi, Quantitative easing, Long Depression, Great Recession, Federal Open Market Committee, Bank of Canada, Monetary policy, Japanese asset price bubble, Money supply, Great Depression, Central bank, Federal Reserve System*
- If instead of using the entire Wikipedia, we limit ourselves to the source dataset comprised of concepts only, this result would translate to:
 - *Recession, Quantitative easing, Monetary policy, Money supply, Central bank, Federal Reserve System*

Oracle Advanced Analytics 12.2

New Oracle Database Features



- **Explicit Semantic Analysis (ESA) algorithm**
- *"The more things change... Yes, I'm inclined to agree, especially with regards to the historical relationship between stock prices and bond yields. The two have generally traded together, rising during periods of economic growth and falling during periods of contraction. Consider the period from 1998 through 2010, during which the U.S. economy experienced two expansions as well as two recessions: Then central banks came to the rescue. Fed Chairman Ben Bernanke led from Washington with the help of the bank's current \$3.6T balance sheet. He's accompanied by Mario Draghi at the European Central Bank and an equally forthright Shinzo Abe in Japan. Their coordinated monetary expansion has provided all the sugar needed for an equities moonshot, while they vowed to hold global borrowing costs at record lows"*
- Top topics (**concepts, people, organizations, events**) discovered by ESA using Wikipedia as model source data
 - *Recession, Ben Bernanke, Lost Decade Japan, Mario Draghi, Quantitative easing, Long Depression, Great Recession, Federal Open Market Committee, Bank of Canada, Monetary policy, Japanese asset price bubble, Money supply, Great Depression, Central bank, Federal Reserve System*
- If instead of using the entire Wikipedia, we limit ourselves to the source dataset comprised of concepts only, this result would translate to:
 - *Recession, Quantitative easing, Monetary policy, Money supply, Central bank, Federal Reserve System*

Oracle Advanced Analytics 12.2

ESA vs. LDA (Latent Dirichlet Allocation); ESA is more interpretable than LDA

Topics discovered by LDA are *latent*, meaning difficult to interpret

- Topics are defined by their keywords, i.e., they have no names, no abstract descriptions
- To give meaning to topics, keywords can be extracted by LDA
- Definitions solely based on keywords are fuzzy, and keywords for different topics usually overlap
- Extracted keywords can be just generic words
- Set of automatically extracted keywords for a topic does not map to a convenient English topic name

Biggest LDA problem; set of topics is fluid

- Topic set changes with any changes to the training data
- Modification of training data changes topic boundaries
- Training data is almost never static

ESA discovers topics from a given set of topics in a knowledge base

- Topics are defined by humans → topics are well understood.
- Topic set of interest can be selected and augmented if necessary → full control of the selection of topics
- Set of topics can be geared toward a specific task, .e.g., knowledge base for topic modeling of online messages possibly related to terrorist activities, which is different than one for topic modeling of technical reports from academia
- Can combine multiple knowledge bases, each with its own topic set, which may or may not overlap
- Topic overlapping does not affect ESA's capability to detect relevant topics

Oracle Advanced Analytics 12.2

New Oracle Database Features



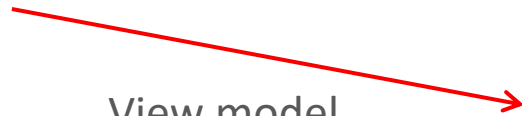
- **Explicit Semantic Analysis (ESA) algorithm**



WIKISAMPLE



Explicit Feature Extraction Build



View model



Apply model

The screenshot shows the Oracle Advanced Analytics interface. At the top, there are tabs for 'Start Page', 'ESA', 'Oracle Database 12.2 dmuser.sql', 'dmardemo.sql', and 'FEAT_ESA_1_11'. Below the tabs, there are controls for 'Sorting Order' (Ascending/Descending), 'Fetch Size' (2,000), and 'Items Per Feature' (50). A 'Filter' section is set to 'No Filter'. The main area displays a table with 'Feature ID' and 'Items' columns. The 'Abraham Lincoln' feature is highlighted, showing a list of related terms. Below the table, there are options for 'Tag Cloud', 'Coefficients', and 'Order By' (Coefficients, Ascending, Descending). A large text box displays the top terms for the selected feature.

Feature ID	Items
ASCII	H CHARACTERS CHARACTER CONTROL ENTERED SYSTEMS REPRESENT TEXT SPACE US ESCAPE CODE D
Abraham Lincoln	VICTORY FEDERAL DEPARTMENT THOMAS REPUBLICAN FEBRUARY U EXECUTIVE PRESIDENTIAL LARGEL
Abu Dhabi	OIL AL ARAB LARGEST ISLAND AREA CITY SETTLED COAST LOCATED MILES CENTURY BROTHER OFF MA
Academy Award	ACADEMY FILM PROMINENT RULES ANGELES OFFICIAL LOS ITSELF MONTHS WIN PUBLIC FILMS NIGHT F
Achilles	SON HERO POEM ASKED MENTION GODS CULTURE LINES ANCIENT GOD KING GREEK EUROPEAN FATHER
Africa	AFRICA MILES CONTINENT KM ASIA EGYPT NORTHERN C MAIN MEANING ORIGIN APPROXIMATELY NATI
Afro-Asiatic languages	LANGUAGES M BERBER AFRICA LANGUAGE N FEATURES SUGGESTS FAMILY BRANCHES G RED WIDESPRE
Agricultural science	AGRICULTURAL SCIENCE AGRICULTURE PRODUCTION RESEARCH MANAGEMENT SOIL PRODUCTS ENVIR
Agriculture	AGRICULTURE AGRICULTURAL FARMING PRODUCTS FOOD FARM PLANTS ANIMALS SCIENCE ANIMAL PR
Al Gore	SCHOOL TENNESSEE AL CARTHAGE ARMY WASHINGTON SERVED ALBERT OCTOBER ELECTION BORN MIL

Tag Cloud: Coefficients

Order By: Coefficients (Ascending/Descending)

VICTORY FEDERAL THOMAS REPUBLICAN FEBRUARY
DEPARTMENT U PRESIDENTIAL LARGELY EXECUTIVE STATES
PRESIDENT WAR LED EFFORTS AMERICA ROLE CIVIL UNION POWERS GIVEN
GOVERNMENT WRITINGS SETTING QUALITIES PROVED PROPERTIES ORDERS NICKNAMED
NEVADA NATION HELPED FIGHTING EXPAND ELECTION DECLARED CONSIDERATIONS
CONFLICT CO CITIZEN BOUNDARIES BIRTH BEGINNING ATTEMPT AMERICANS ENDING
SOMETIMES KM OFTEN WIN ULTIMATELY SUPPORT ROOM RELATIVELY MILLIONS MILES GOAL
GETTING DEFFAT COUNTY CLAIM AGRICULTURE SOON OPPOSED LEADER FORMER EFFORT DIRECTED

```
SELECT FEATURE_COMPARE(feats_esa_1_1
    USING 'Oracle Database is the best available for managing your data' text
    AND USING 'The SQL language is the one language that all databases have in common'
    text) similarity
FROM DUAL;
```

The screenshot shows the Oracle SQL Developer interface with the query results. The 'Query Result' tab is active, showing a single row with the similarity score.

SIMILARITY
1 0.7629282835628297

The result we get is 0.7629.

Oracle Advanced Analytics 12.2

New Oracle Database Features



- **Extensibility for R Models**

- Register R models as in-database models for build, apply, settings, and viewing
- Supports data with “nested” attributes, handling text and aggregated transactional data for open source R packages
- Extends ease of advanced analytics development from R to Oracle Database
- Enables R users to roll out new analytics and more rapidly take advantage of existing R packages

Oracle Data Miner 4.2

New Features for OAA



NEW IN
4.2

- Add/Expose all 12.2 features in Oracle Data Miner UI

Oracle SQL Developer

Connections: BIWA 16_12c_DB_Cloud_Jonathan2, BIWA DMUSER - Learn Predictive An..., CBERGER dbpm23 Cloud instance Co..., Charlie Database 12c laptop dmuser...

Start Page: 360 View Star Schema Analytics, BUY_INSURANCE_WORKFLOW, BUY INSURANCE

Parallel Query Off

Scatter Box plots etc. 1, Explore Data, CUST_INSUR_LTV1, Filter Columns, Multiple Classification Models, Most Likely Customers, Explore Data 1, Clustering Segmentation 1, Filter Columns Details 1, LIKELY_BUY_INSURANCE_CUSTMRS 1

Worksheet: Query Builder

```
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,
```

Script Output: Query Result

POLICYNUMBER	PERCENT_FRAUD	RNK	
1	654	61.87	1
2	11068	57.37	2
3	7435	55.47	3

Multiple Classification Models - Properties

Models: Build, Model Settings

Rule Surrogates Target Values

Node Rule: Wrap

```
If BANK_FUNDS > 246
And CHECKING_AMOUNT > 282
And MONEY_MONTHLY_OVERDRAWN <= 54.215

Then No
```

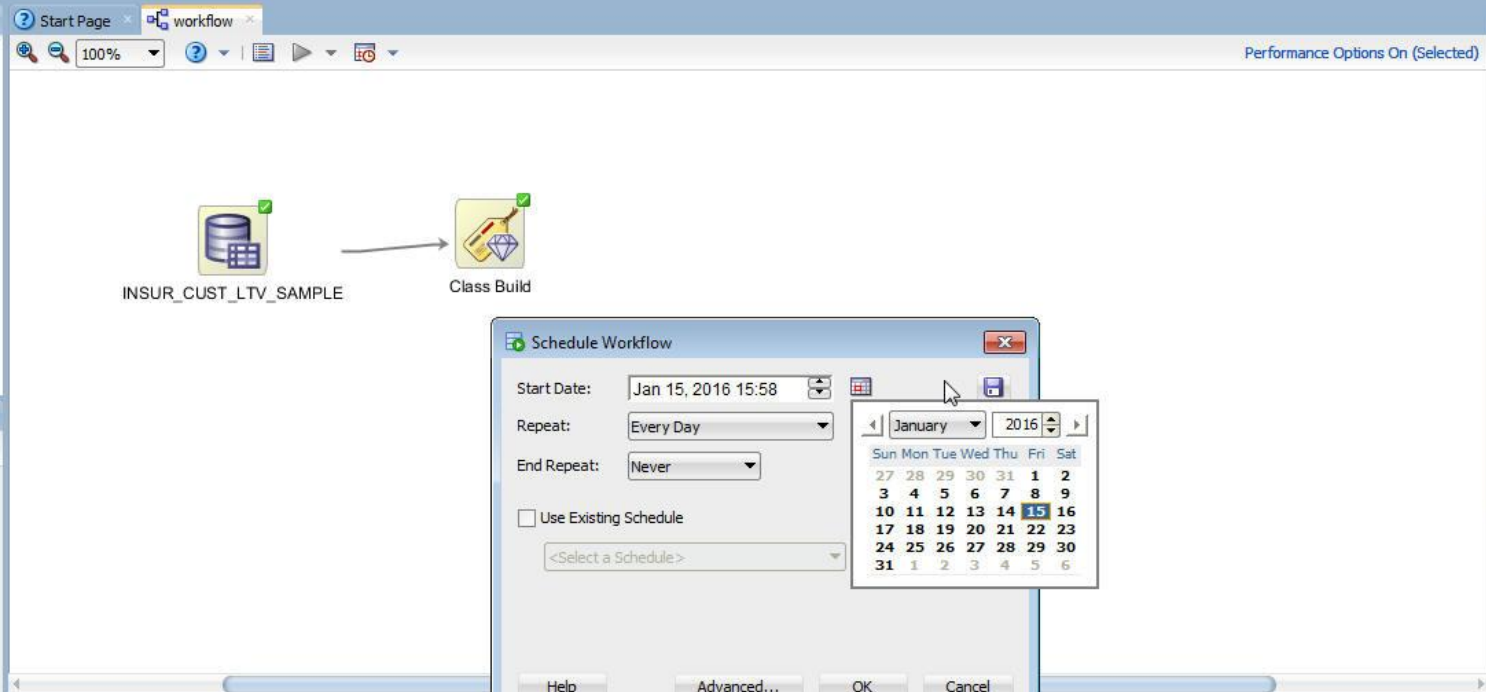
Confidence	Support
0.8515671200473093	0.1832863646217212

Components: Workflow Editor, Data (Create Table or View, Data Source, Explore Data, Graph, SQL Query, Update Table), Transforms (Aggregate, Filter Columns, Filter Columns Details, Filter Rows, Join, JSON Query, Sample, Transform), Text, Models (Anomaly Detection, Association, Classification, Clustering, Feature Extraction, Model, Model Details, Regression), Predictive Queries (Anomaly Detection Query, Clustering Query, Feature Extraction Query, Prediction Query), Evaluate and Apply, Linking Nodes



workflow - Structure Thumbnail

- Class Build
- INSUR_CUST_LTV_SAMPLE
- Links



Schedule Workflow

Start Date: Jan 15, 2016 15:58

Repeat: Every Day

End Repeat: Never

Use Existing Schedule

<Select a Schedule>

Help Advanced... OK Cancel

January 2016

Sun	Mon	Tue	Wed	Thu	Fri	Sat
27	28	29	30	31	1	2
3	4	5	6	7	8	9
10	11	12	13	14	15	16
17	18	19	20	21	22	23
24	25	26	27	28	29	30
31	1	2	3	4	5	6

Connections Data Miner

- Connections
- den00jlf-im
- jw1415
- wfriday
- Oracle NoSQL Connections
- Cloud Connections

workflow - Properties

Find

Comment:

Workflow Jobs

den00jlf-im

Workflow	Project	Status
No entries. Right click for Preferences.		

Reports

- All Reports
- Analytic View Reports
- Data Dictionary Reports
- Data Modeler Reports
- OLAP Reports
- TimesTen Reports
- User Defined Reports

Logging Page - Log

Level	Sequence	Elapsed	Source	Message
✖	2376	1155	o.dmt.dataminer.runne...	RunItem: unable to read repository config table.
✖	2375	1373	o.dmt.dataminer.runne...	RunItem: unable to read repository config table.
✖	2374	1513	o.dmt.dataminer.runne...	RunItem: unable to read repository config table.
✖	2373	1622	o.dmt.dataminer.runne...	RunItem: unable to read repository config table.
✖	2372	1092	o.dmt.dataminer.runne...	RunItem: unable to read repository config table.
✖	2371	1170	o.dmt.dataminer.runne...	RunItem: unable to read repository config table.

Messages Statements Logging Page

Components

Workflow Editor

Data

- Create Table or View
- Data Source
- Explore Data
- Graph
- SQL Query
- Update Table

Transforms

Text

Models

- Anomaly Detection
- Association
- Classification
- Clustering
- Explicit Feature Extraction
- Feature Extraction
- Model
- Model Details
- Regression

Predictive Queries

Model Operations

Linking Nodes

NEW IN
4.2

Oracle R Advanced Analytics for Hadoop

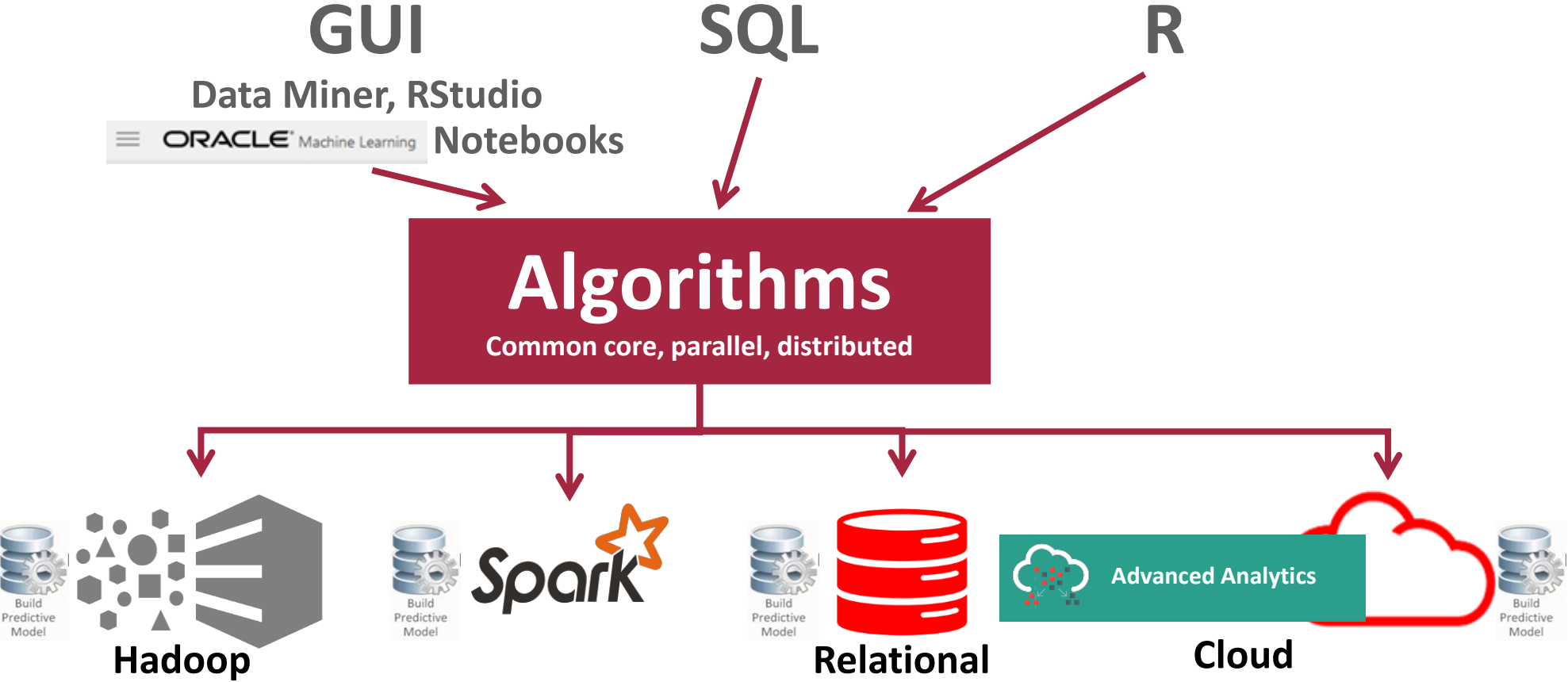
New Features in ORAAH 2.7



- Updated ORAAH GLM and LM algorithms which are much faster, stable and light on memory than comparable GLM and LM methods from Spark Mllib
- Both methods also bring a new summary feature that makes them comparable to solutions from open-source R glm and lm, but capable of handling Big Data at enterprise scale
- The Neural Networks algorithm has been enhanced to support the full formula processing and a full build and scoring in Spark
- The new Gaussian Mixture Models is an addition to the set of algorithms supported in Spark Mllib
- ORAAH's Spark-based LM with full formula support and summary - `orch.lm2()`
- ORAAH's Spark-based GLM with full formula support and summary - `orch.glm2()`

Oracle Advanced Analytics Strategy & Road Map

- One server side product, with a single analytic library, supporting multiple data platforms, analytical engines, UIs and deployment strategies



What is a “notebook”



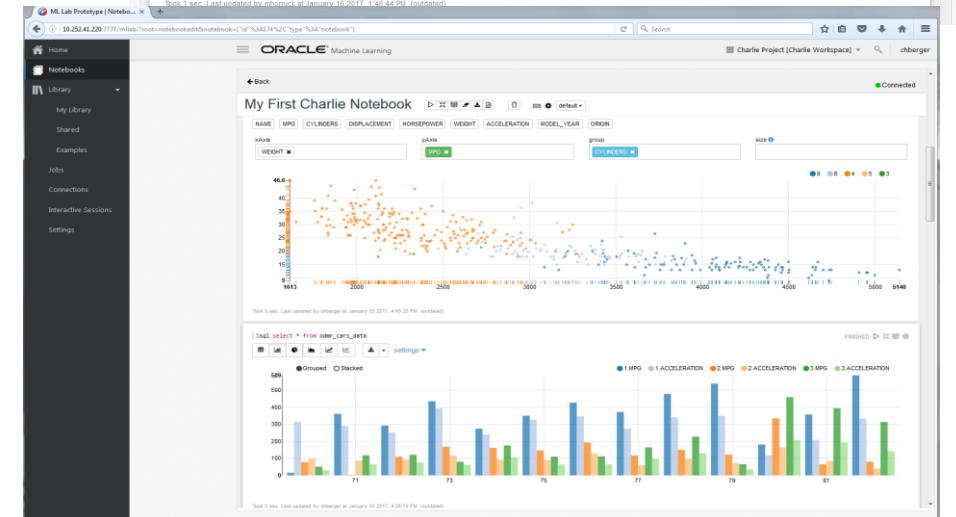
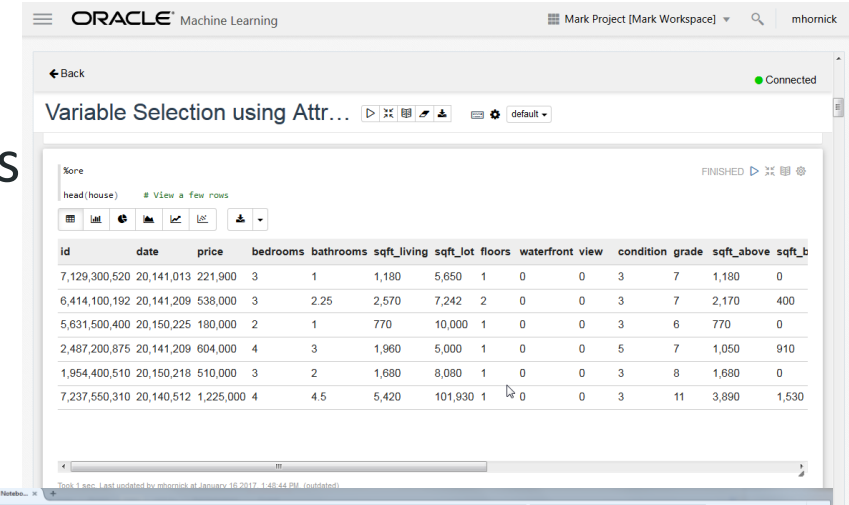
Oracle Machine Learning

Multi-Language, Multi-Server Engine Oracle Machine Learning for Clouds



Key Features

- Collaborative ML environment for data scientists
 - Shared Zeppelin notebooks, templates, and permissions
- Language—SQL ML algorithms API (ODM)
- DWCS server—Oracle Database
- Supports deployment of ML analytics solutions
 - Enables publishing libraries, templates, use cases
- Road map
 - Multi-Language support
 - R language
 - Multi-Server Engines
 - R, ORE, ORAAH, Spark





Oracle's Advanced Analytics

Predictive Applications + OBIEE Integration

Enabling “Predictive” Enterprise Applications

Oracle Applications Using Oracle Advanced Analytics—Partial List

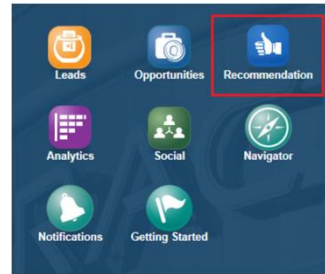
• Oracle HCM Cloud

- Employee turnover and performance prediction and “What if?” analysis



• Oracle Sales Cloud

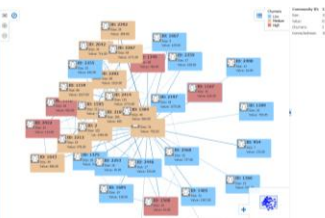
- Prediction of sales opportunities, what to sell, amount, timing, etc.



• Oracle Industry Data Models

- **Communications Data Model** churn prediction, segmentation, profiling, etc.
- **Retail Data Model** loyalty and market basket analysis
- **Airline Data Model** analysis frequent flyers, loyalty, etc.
- **Utilities Data Model** customer churn, cross-sell, loyalty, etc.

Segment Avg Debt value



• Oracle Retail GBU Cloud Services

- Market Basket Analysis Insights
- Customer Insights & Clustering

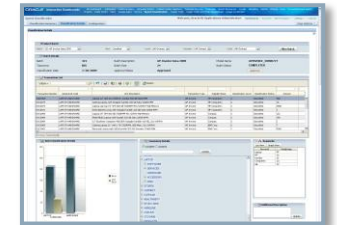


• Oracle Customer Support

- Predictive Incident Monitoring (PIM)

• Oracle Spend Classification

- Real-time and batch flagging of noncompliance and anomalies in expense submissions



• Oracle FinServ Analytic Applications

- Customer Insight, Enterprise Risk Management, Enterprise Performance, Financial Crime and Compliance

• Oracle Adaptive Access Manager

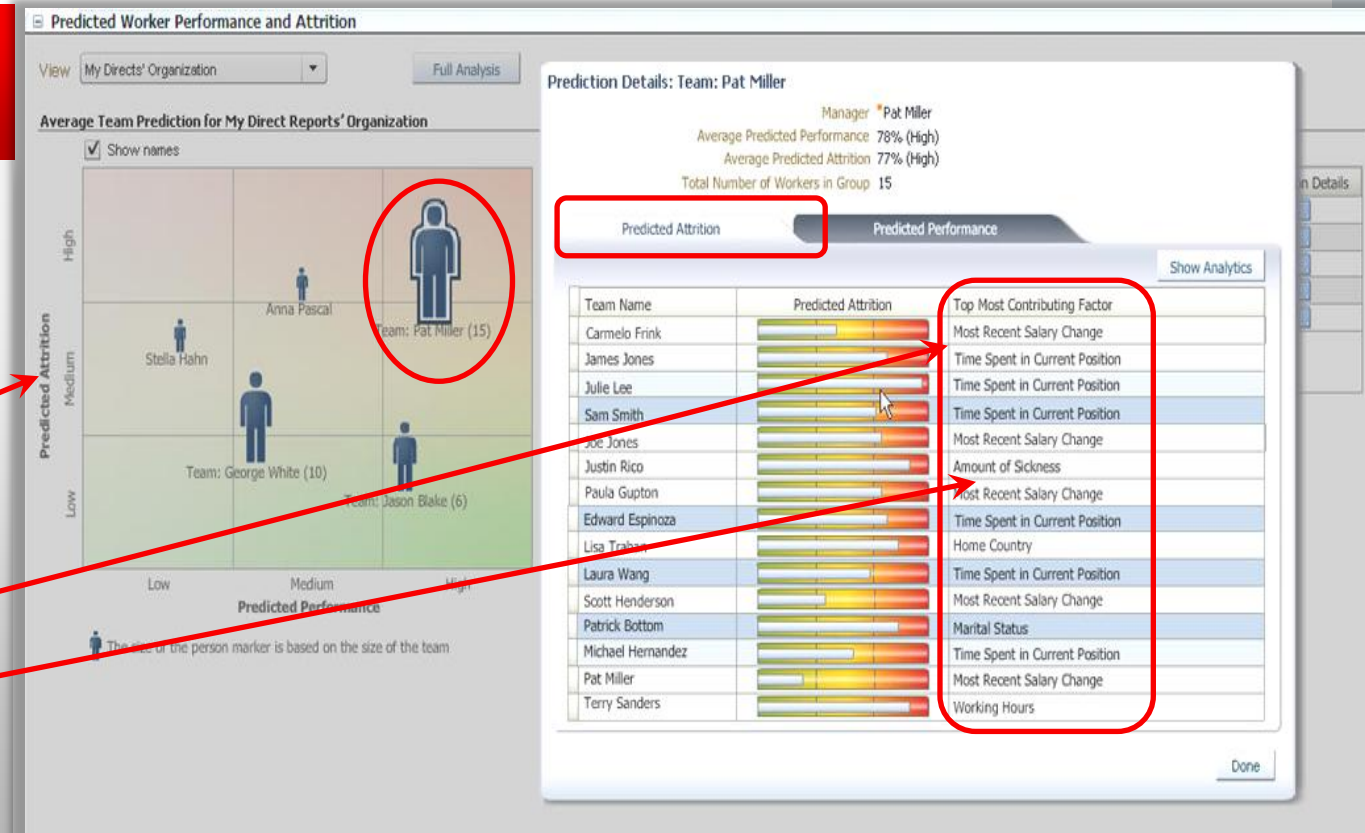
- Real-time security and fraud analytics

HCM Predictive Workforce

Predictive Analytics Applications

Human Capital Management Powered by OAA

- Oracle Advanced Analytics factory-installed predictive analytics
- Employees likely to leave and predicted performance
- Top reasons, expected behavior
- Real-time "What if?" analysis



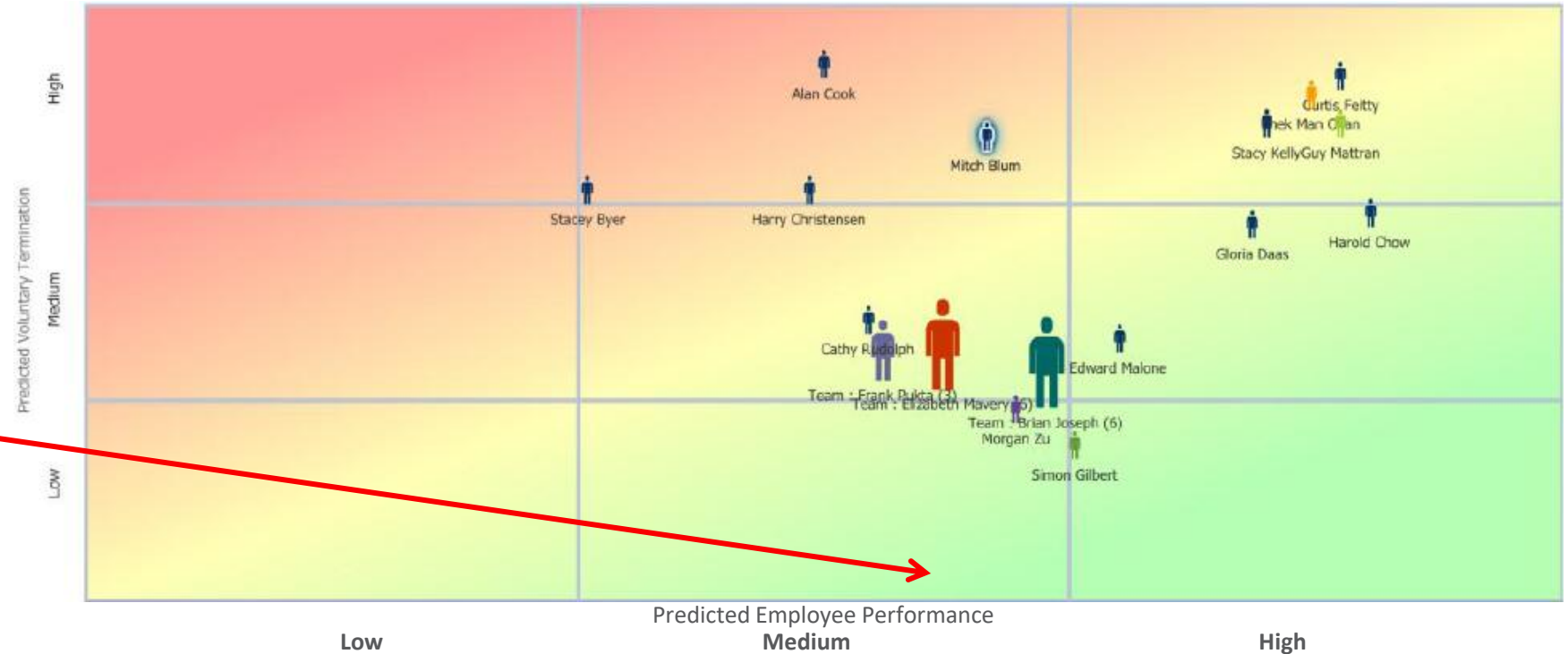
[Link to Oracle HCM on O.com](#)
[HCM Predictive Workforce demo](#)

Performance and Voluntary Termination Predictions

Let's Walk Through Again But Go More Slowly...

- Predicted Voluntary Termination for Worker

- Predicted Performance for Individual Worker



Predicted Performance for Individual Worker

Prediction Details: Gloria Daas



Name **Gloria Daas** Current Performance Rating 2-Inconsistent
 Manager **Blum, Mitch** Predicted Voluntary Termination 63%
 Predicted Performance 79%

Predicted Voluntary Termination **Predicted Performance**

Contribution	Predicted Performance Reason	Current Value
	Current performance rating	2
	Latest salary change	2.9 %
	Time with current manager	146.2 Months
	Current grade	Admin03
	Time since last sickness	1.42 Months
	Worker's performance compared to pe...	58.06
	Number of manager changes in the las..	1
	Current or most recent manager	Mitch Blum
	Time in current grade	56.52 Months

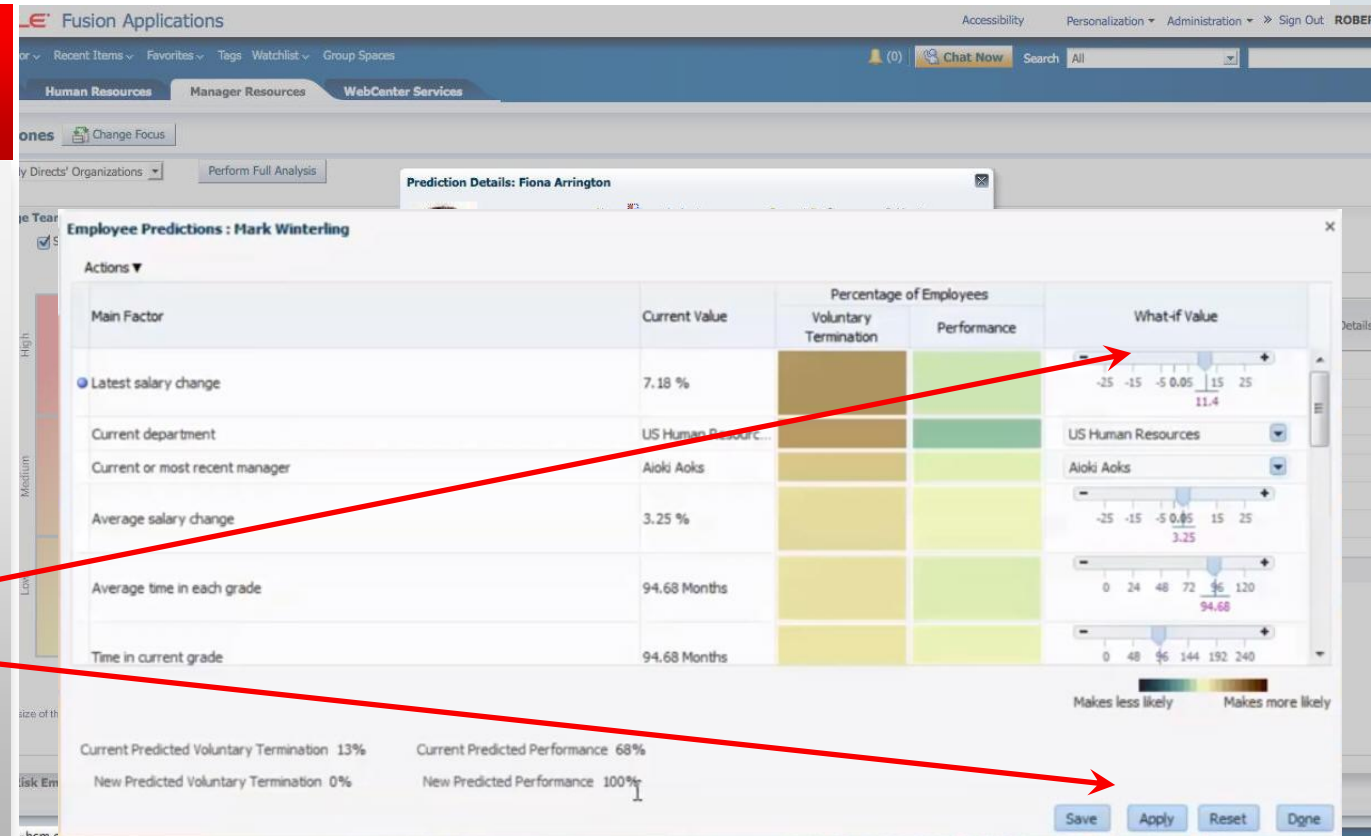


HCM Predictive Workforce

Predictive Analytics Applications

Fusion Human Capital Management Powered by OAA

- Oracle Advanced Analytics factory-installed predictive analytics
- Employees likely to leave and predicted performance
- Top reasons, expected behavior
- Real-time "What if?" analysis



[Link to Oracle HCM on O.com](#)
[HCM Predictive Workforce demo](#)

Oracle Sales Cloud Sales Predictor

Predictive Analytics Applications

CRM Sales Predictions Powered by OAA

- Sales Predictor helps sales reps answer critical sales questions:
 - Which products should be offered to a customer?
 - Who are the customers buying products?
 - What are the reasons a product is being bought?
- Sales Predictor offers product recommendations that have a higher likelihood of being converted to a win.

White Space Analysis

	Business Tablets	Software Manual Set	Vario 7000 Tablet	Vario 7500 Tablet	Vario 8000 Tablet	Vario 8500 Tablet
Imaging Innovatio... Total Potential: \$12,442,354	\$11,345,000	\$55,378	Rejected	\$88,896	\$885,534	\$67,546
SynergyTek Inc Total Potential: \$9,762,340	No potential \$0	Assets	\$35,224	\$9,000,000	\$681,861	\$45,255
Times Technology Total Potential: \$3,088,816	\$2,335,731	\$36,000	\$35,224	No potential \$0	\$681,861	Rejected
Star Coffee HQ Total Potential: \$2,136,323	\$566,764	\$1,500,000	\$24,304	No potential \$0	Rejected	\$45,255
4M Technologies Total Potential: \$1,472,080	No potential \$0	No potential \$0	\$1,159,800	\$312,280	No potential \$0	No potential \$0
Pacific General H... Total Potential: \$773,711	Assets	No potential \$0	\$24,304	No potential \$0	\$681,861	\$67,546

[Link to Oracle CRM SPE on O.com](#)

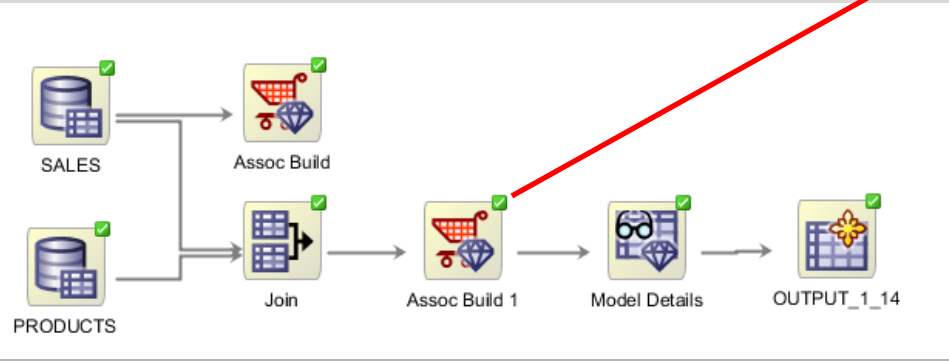
Oracle Retail Market Basket Insights Cloud Service

Market Basket Analysis



Pre-Built Market Basket Analysis

- Gain actionable insight into your shoppers' behavior.
- Pre-built market-basket analysis identifies product affinities

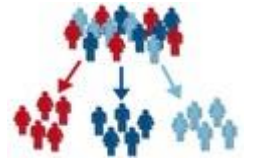


[Link to Oracle Retail MBA on O.com](#)



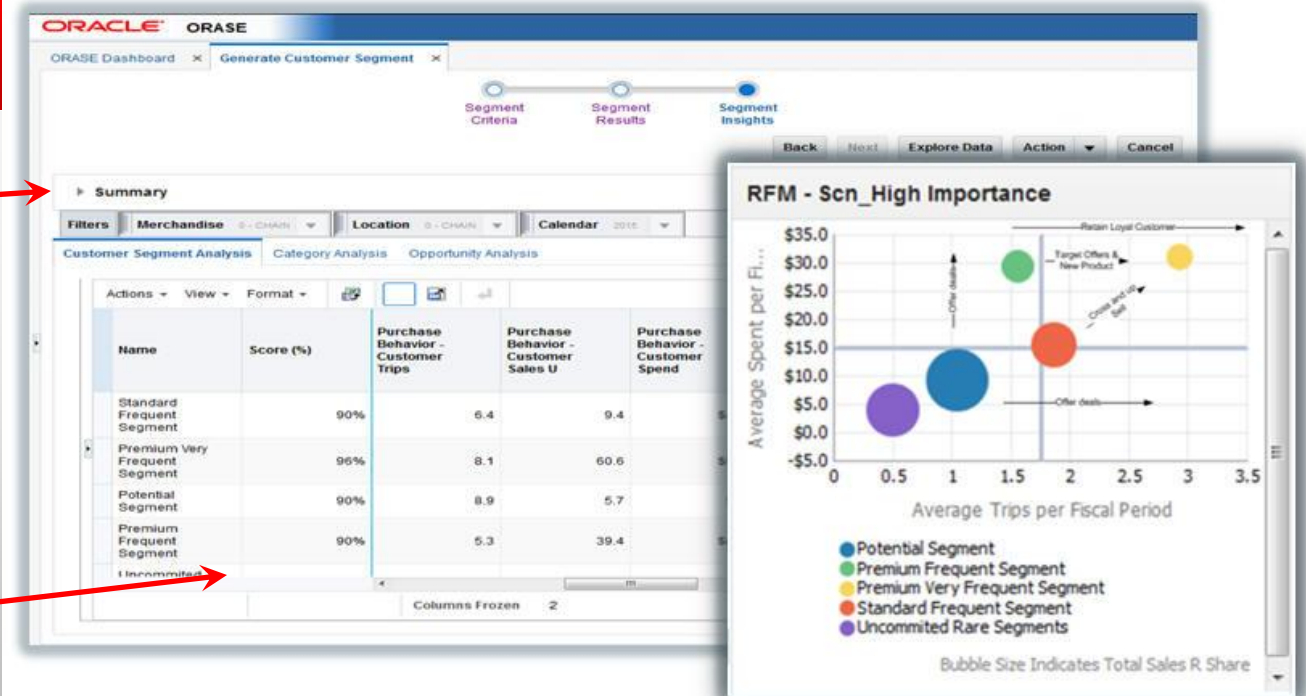
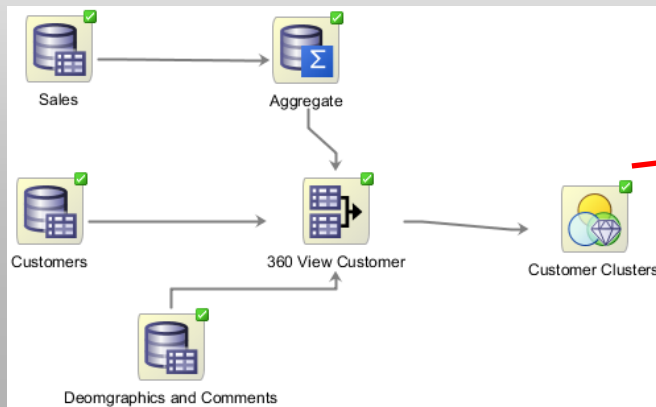
Oracle Retail Customer Insights Cloud Service

Customer Segmentation/Clustering Analysis



Pre-Built Customer Clustering Models

- Gain actionable insight into your customer's behavior.
- Pre-built clustering models identify hidden customer segments



[Link to Oracle Retail CI Cloud on O.com](#)

Oracle Communications Industry Data Model

Example Predictive Analytics Application

Pre-Built Predictive Models

- Fastest Way to Deliver Scalable Enterprise-wide Predictive Analytics
- OAA's clustering and predictions available in-DB for OBIEE
- Automatic Customer Segmentation, Churn Predictions, and Sentiment Analysis

Business Intelligence

Customer Segmentation Details

Cell Phone No	Contract Value	Month Revenue	Debt Value	LTV Band	LTV Value	Customer Segment	Churn Probability	Customer Segment Key	Community Role	Community Size	Churn Ratio
9985007046	\$18,000.00	\$15,600.00	\$140.00	LTV_1	\$41,000.00	Age Young and PAY TV user	59	104	LOCAL		3
9985007589	\$18,000.00	\$16,200.00	\$444.00		\$49,000.00	Bad phone number and Low usage	45	104	PASSIVE		3
9985006289	\$18,000.00	\$16,800.00	\$140.00		\$34,000.00	Family User, High Revenue	71	104	LOCAL		4
9985003794	\$18,000.00	\$14,000.00	\$140.00		\$82,000.00	High end insensitive to Loyalty Program	16	104	PASSIVE		7
9985005144	\$6,000.00	\$5,478.26	\$260.00		\$85,000.00	High value Organizational Customer	19	104	LOCAL		4
9985002105	\$6,000.00	\$5,555.56	\$444.00		\$56,000.00	High value and use loyalty program	76	104			
9985000594	\$6,000.00	\$5,538.46	\$180.00		\$76,000.00	Low Revenue	16	104			

[Link to OCDM on OTN](#)

A woman with long brown hair and glasses is sitting at a wooden table in a cafe. She is wearing a brown leather jacket over a blue patterned scarf. She is holding a black mobile phone to her ear with her right hand and looking down at a newspaper or magazine on the table with her left hand. The background is a bright, slightly blurred cafe interior with other tables and chairs. The text "Getting started" is overlaid in white on the left side of the image.

Getting started

Getting started: Oracle's AA/ML Links and Resources

ORACLE Oracle Advanced Analytics Overview Information

- [Oracle's Machine Learning and Advanced Analytics 12.2c and Oracle Data Miner 4.2 New Features presentation](#)
- [Oracle Advanced Analytics Public Customer References](#)
- [Big Data Analytics with Oracle Advanced Analytics: Making Big Data and Analytics Simple white paper](#) on OTN
- [Oracle INTERNAL ONLY OAA Product Management Wiki and Beehive Workspace](#)

YouTube recorded Oracle Advanced Analytics Presentations and Demos, White Papers

- [Oracle's Machine Learning & Advanced Analytics 12.2 & Oracle Data Miner 4.2 New Features YouTube video](#)
- [Library of YouTube Movies](#) on Oracle Advanced Analytics, Data Mining, Machine Learning (7+ "live" Demos e.g. Oracle Data Miner 4.0 New Features, Retail, Fraud, Loyalty, Overview, etc.)
- [Overview YouTube video](#) of Oracle's Advanced Analytics and Machine Learning

ORACLE UNIVERSITY Getting Started/Training/Tutorials

- Link to [OAA/Oracle Data Miner Workflow GUI Online \(free\) Tutorial Series](#) on OTN
- Link to [OAA/Oracle R Enterprise \(free\) Tutorial Series](#) on OTN
- Link to [Try the Oracle Cloud Now!](#)
- Link to [Getting Started w/ ODM blog entry](#)
- Link to [New OAA/Oracle Data Mining 2-Day Instructor Led Oracle University course.](#)
- [Oracle Data Mining Sample Code Examples](#)

ORACLE Help Center Additional Resources, Documentation & OTN Discussion Forums

- [Oracle Advanced Analytics Option on OTN page](#)
- [OAA/Oracle Data Mining on OTN page](#), [ODM Documentation](#) & [ODM Blog](#)
- [OAA/Oracle R Enterprise page on OTN page](#), [ORE Documentation](#) & [ORE Blog](#)
- [Oracle SQL based Basic Statistical functions](#) on OTN
- [Oracle R Advanced Analytics for Hadoop \(ORAAH\)](#) on OTN

BIWA SIG User Community www.biwasummit.org

- Business Intelligence, Warehousing & Analytics—[BIWA SUMMIT 2018 WITH SPATIAL SUMMIT](#) *THE* Big Data + Adv. Analytics/ML + Spatial + Cloud + IoT + "Everything Cool" User Conference *January 30 - February 1, 2018, Redwood Shores, CA* (contains links to past presentations)



[HOME](#)

[BECOME A SPONSOR](#)

[BIWA 2017](#)

[REGISTER](#)

[SESSIONS](#)

[SPEAKERS](#)

[NEWS](#)

BIWA SUMMIT 2018 WITH SPATIAL SUMMIT

January 30 - February 1, 2018

THE Big Data + Analytics + Spatial + Cloud + IoT + Everything Cool User Conference

www.biwasummit.org



ORACLE®