Oracle's Advanced Analytics 12.2c

New Features & Road Map

Bigger, Better, Faster, More!

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

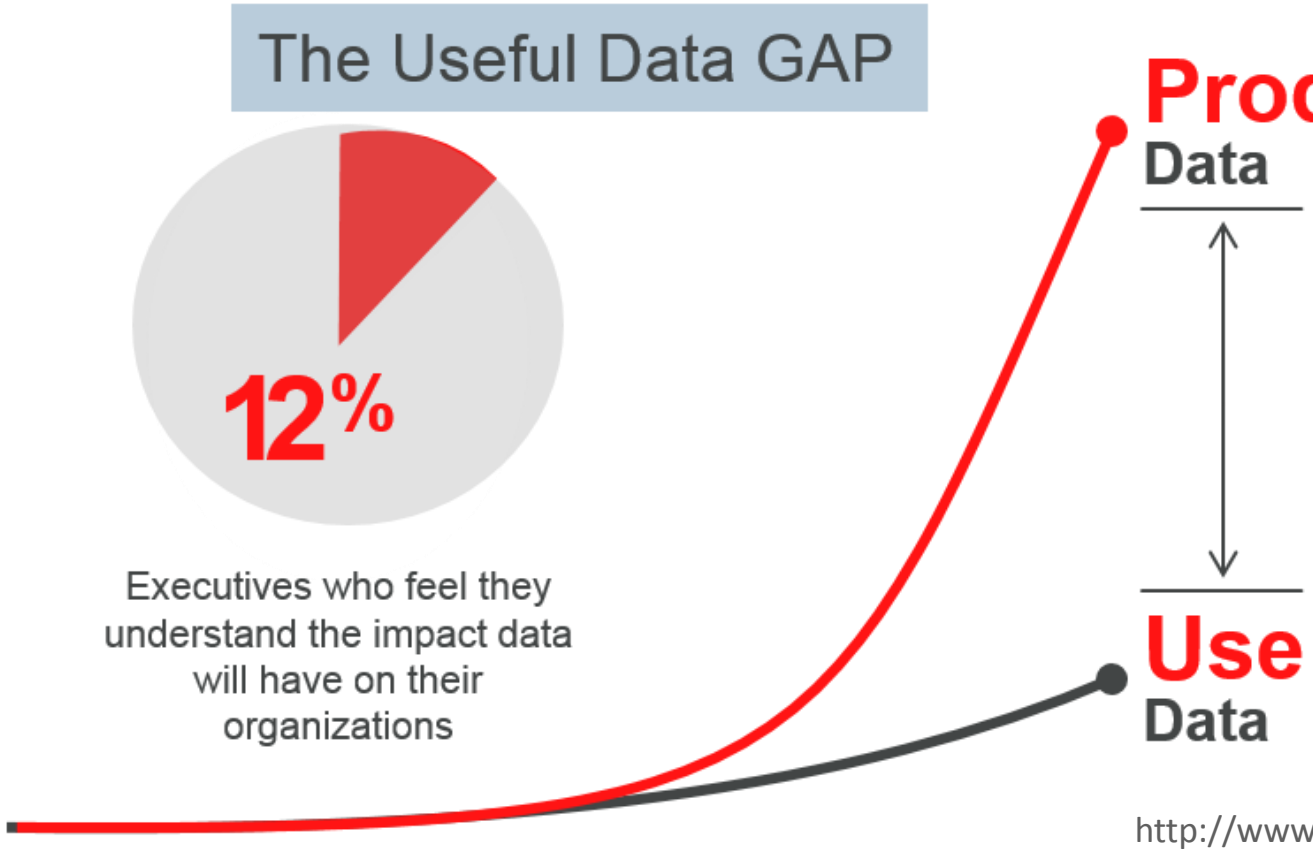
ORACLE

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.

Predictive Analytics 101

- Data, data everywhere – explosive growth
- Growth of data exponentially greater than growth of data analysts!



Machine Learning/Data Analysis platforms requirements:

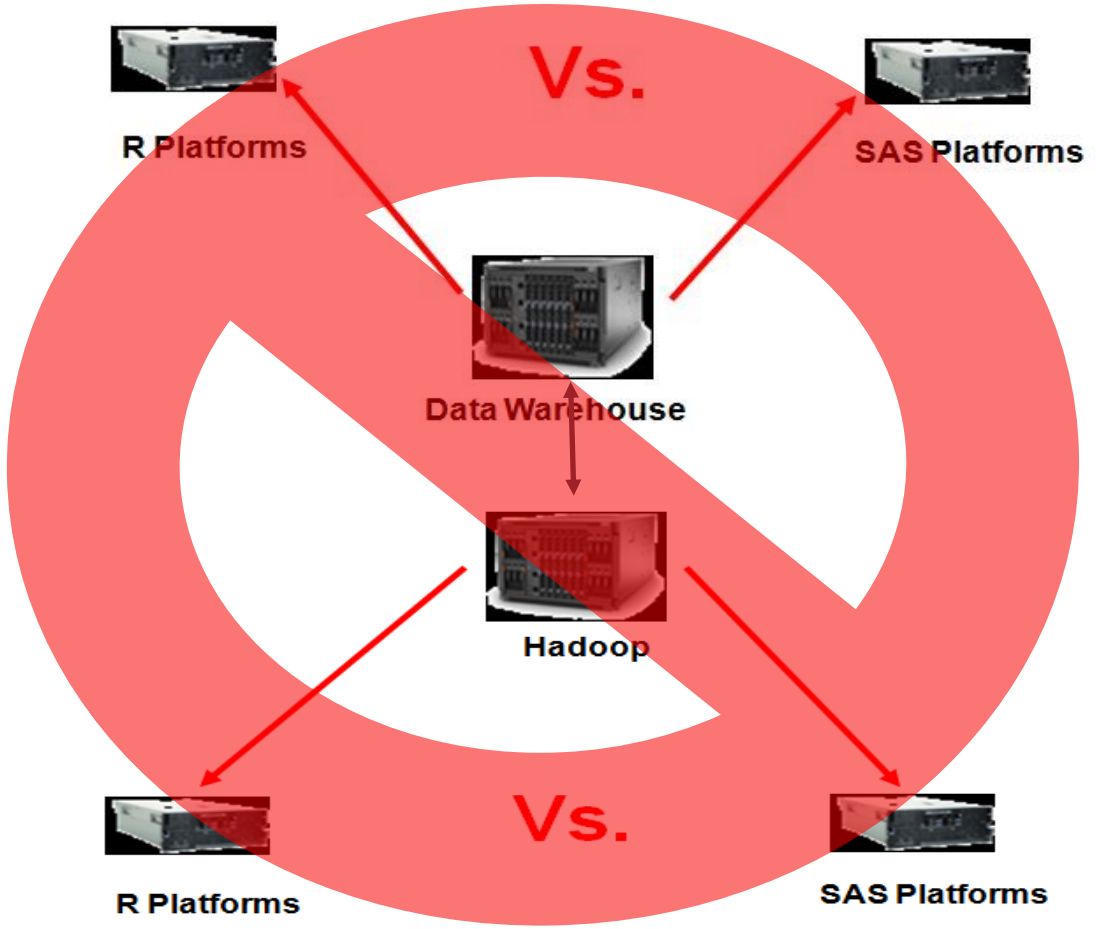
- Be extremely powerful and handle large data volumes
- Be easy to learn
- Be highly automated & enable deployment

http://www.delphianalytics.net/more-data-than-analysts-the-real-big-data-problem/
http://uk.emc.com/collateral/analyst-reports/ar-the-economist-data-data-everywhere.pdf

ORACLE

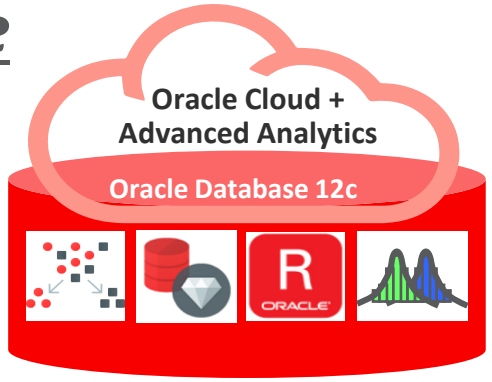
Machine Learning/Analytics + Data Warehouse + Hadoop

- 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



Vision

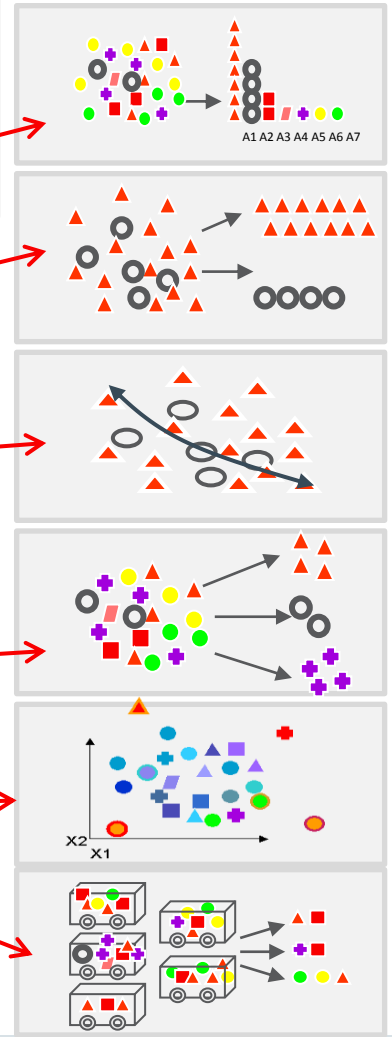
- Big Data + Machine Learning/Analytics Platform for the Era of Big Data and Cloud
  - **Make Big Data + ML/Analytics Model Discovery *Simple***
    - Any data size, on any computer infrastructure—on-premise and/or cloud
    - Any variety of data (structured, unstructured, transactional, geospatial), in any combination
  - **Make Big Data + ML/Analytics Model Deployment *Simple***
    - As a service, as a platform, as an application
    - On-premise and/or cloud



What is Machine Learning, Data Mining & Predictive Analytics?

Automatically sifting through large amounts of data to create models that find previously hidden patterns, discover valuable new insights and make predictions

- Identify most important factor *(Attribute Importance)*
- Predict customer behavior *(Classification)*
- Predict or estimate a value *(Regression)*
- Find profiles of targeted people or items *(Decision Trees)*
- Segment a population *(Clustering)*
- Find fraudulent or “rare events” *(Anomaly Detection)*
- Determine co-occurring items in a “baskets” *(Associations)*



Oracle Advanced Analytics DB Option

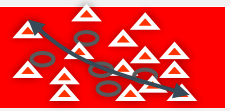
Advanced Analytics

In-Database Machine Learning Algorithms*—SQL & R & GUI Access

Classification

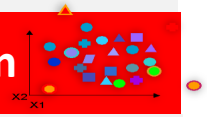
- Decision Tree
- Logistic Regression (GLM)
- Naïve Bayes
- Support Vector Machine (SVM)
- Random Forest

Regression



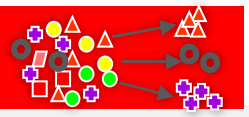
- Multiple Regression (GLM)
- Support Vector Machine (SVM)
- Stepwise Linear Regression
- Linear Model
- Generalized Linear Model
- Multi-Layer Neural Networks

Anomaly Detection



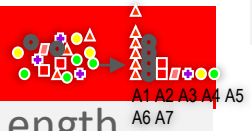
- 1-Class Support Vector Machine

Clustering



- Hierarchical k-Means
- Orthogonal Partitioning Clustering
- Expectation-Maximization

Attribute Importance



- Minimum Description Length
- Unsupervised pair-wise KL div.

Market Basket Analysis

- Apriori – Association Rules

Text Mining

- All OAA/ODM SQL ML support
- Explicit Semantic Analysis

Predictive Queries

- Clustering
- Regression
- Anomaly Detection
- Feature Extraction

Feature Extraction & Creation

- Nonnegative Matrix Factorization
- Principal Component Analysis
- Singular Value Decomposition

Time Series

- Single & Double Exp. Smoothing

Open Source R Algorithms

- Ability to run any R package (9,000+)via Embedded R mode

+ Ability to Mine Unstructured, Structured & Transactional data
+ Partitioned Models

ORACLE

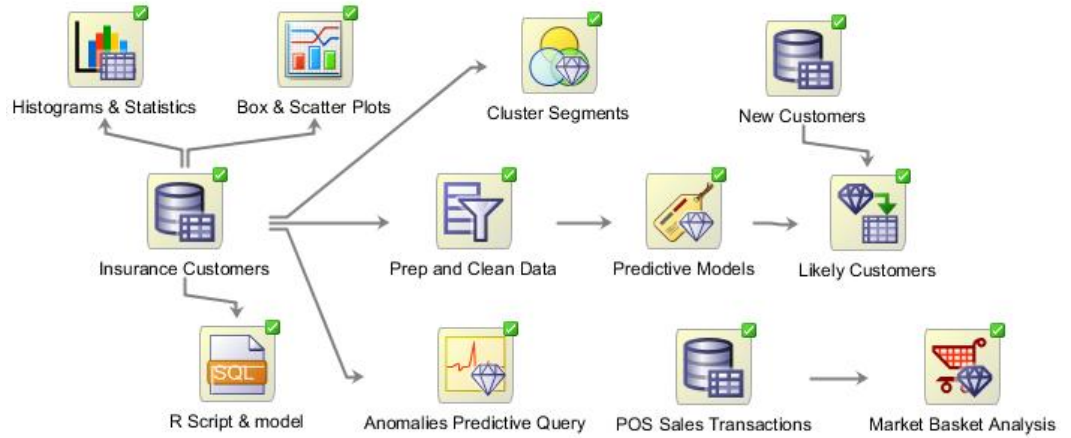
Oracle's Advanced Analytics

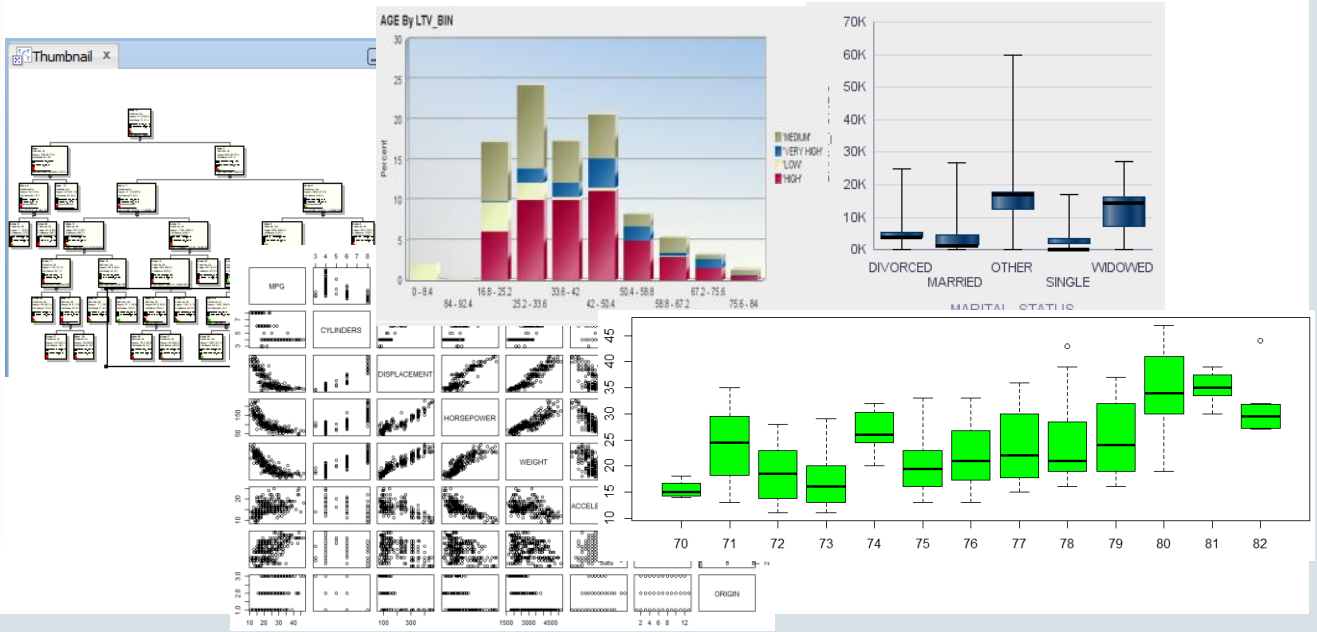
Fastest Way to Deliver Scalable Enterprise-wide Predictive Analytics

Advanced Analytics

Key Features

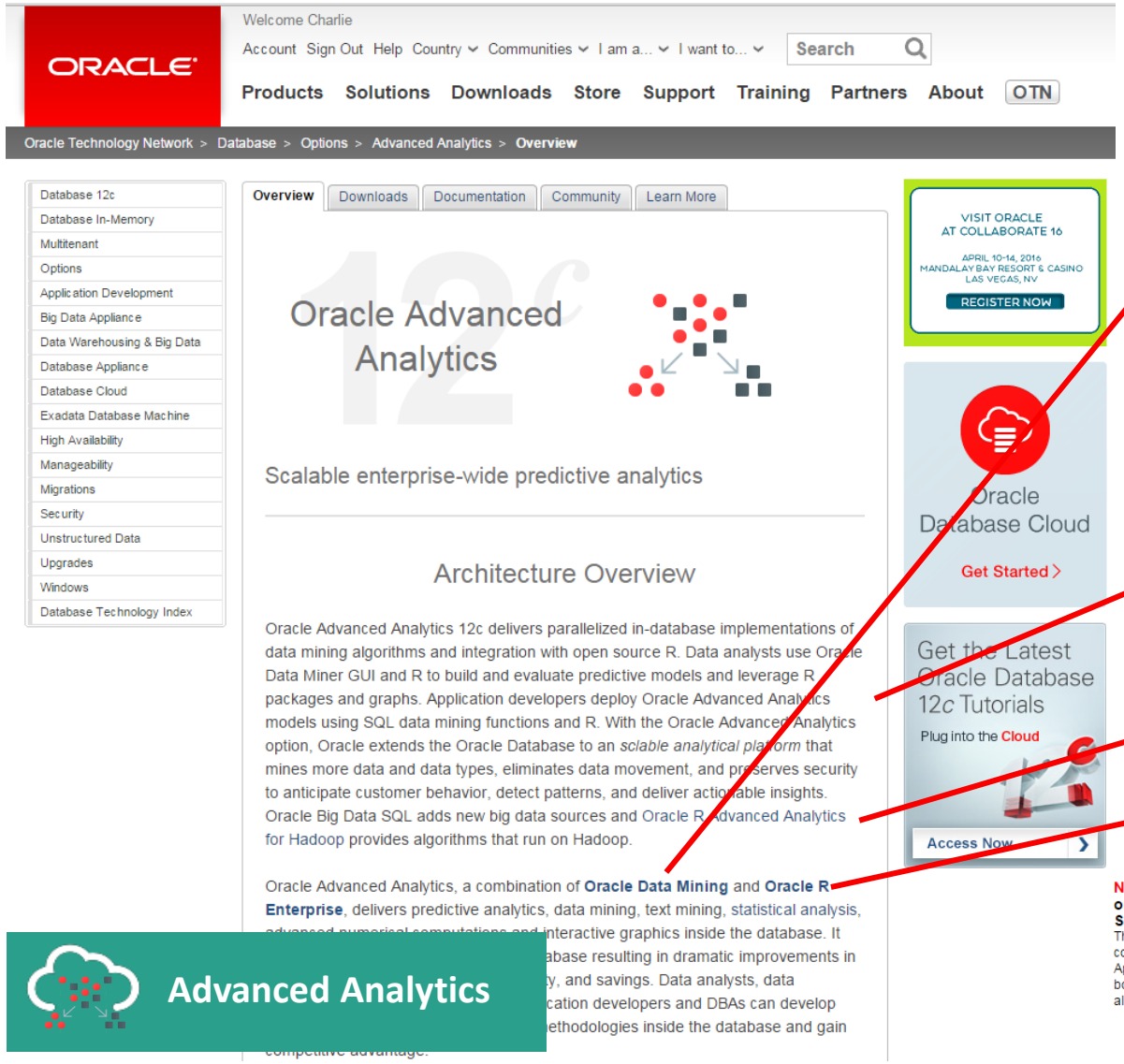
- Parallel, scalable data mining algorithms and R integration
- In-Database + Hadoop—Don't move the data
- Data analysts, data scientists & developers
- Drag and drop workflow, R and SQL APIs
- Extends data management into powerful advanced/predictive analytics platform
- Enables enterprise predictive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

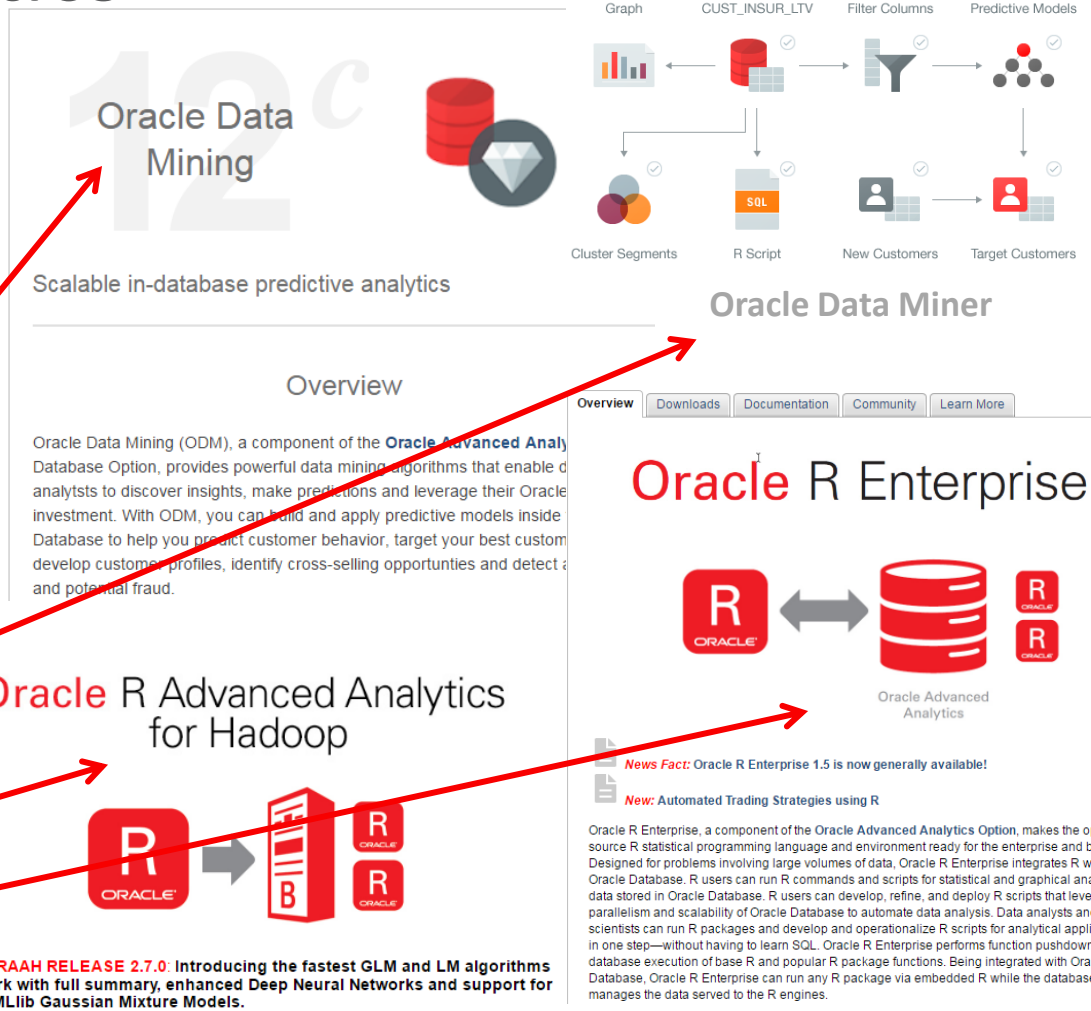




ORACLE

Google “Oracle Advanced Analytics”

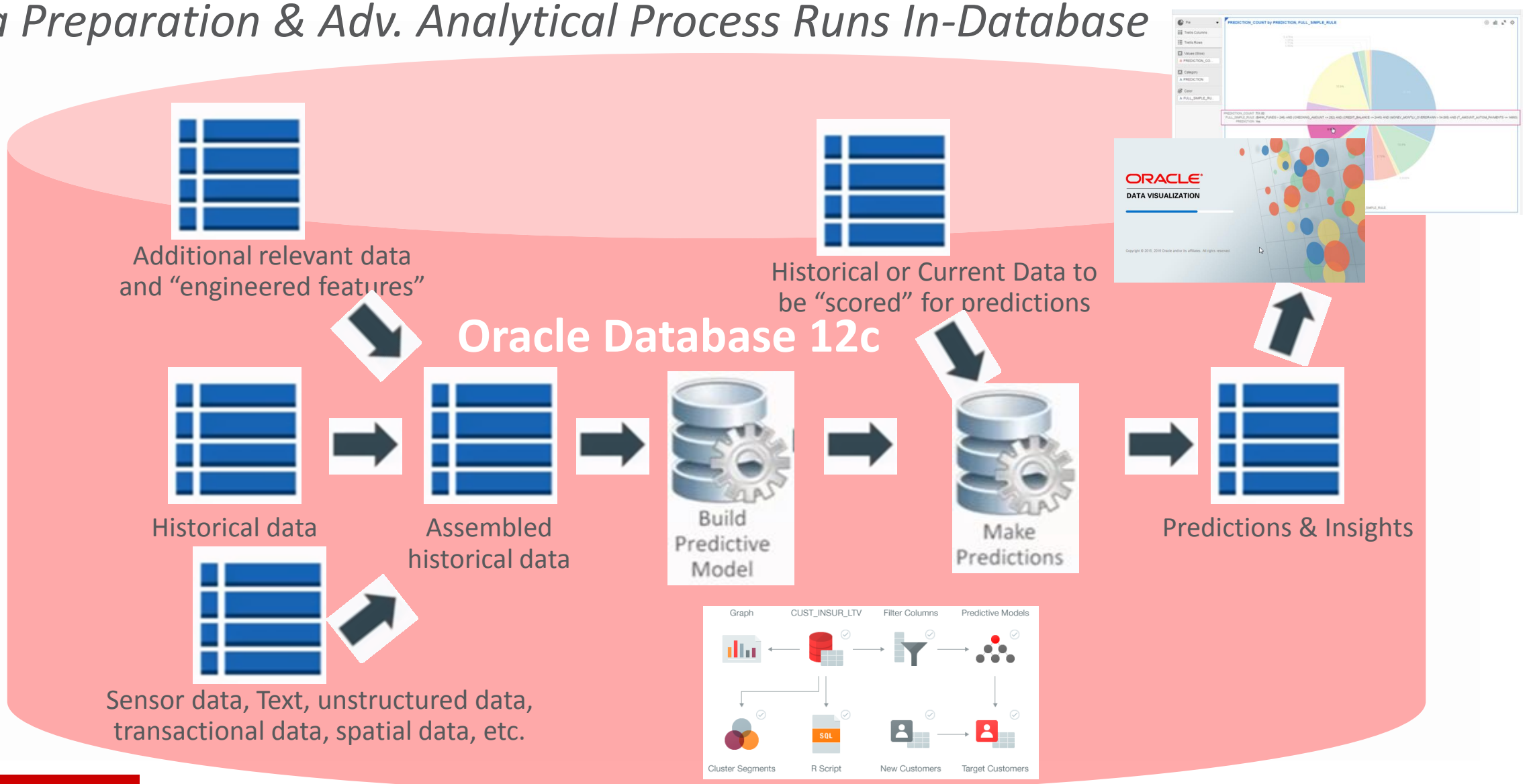




Advanced Analytics

Oracle Data Mining/ Machine Learning/Predictive Analytics

Data Preparation & Adv. Analytical Process Runs In-Database



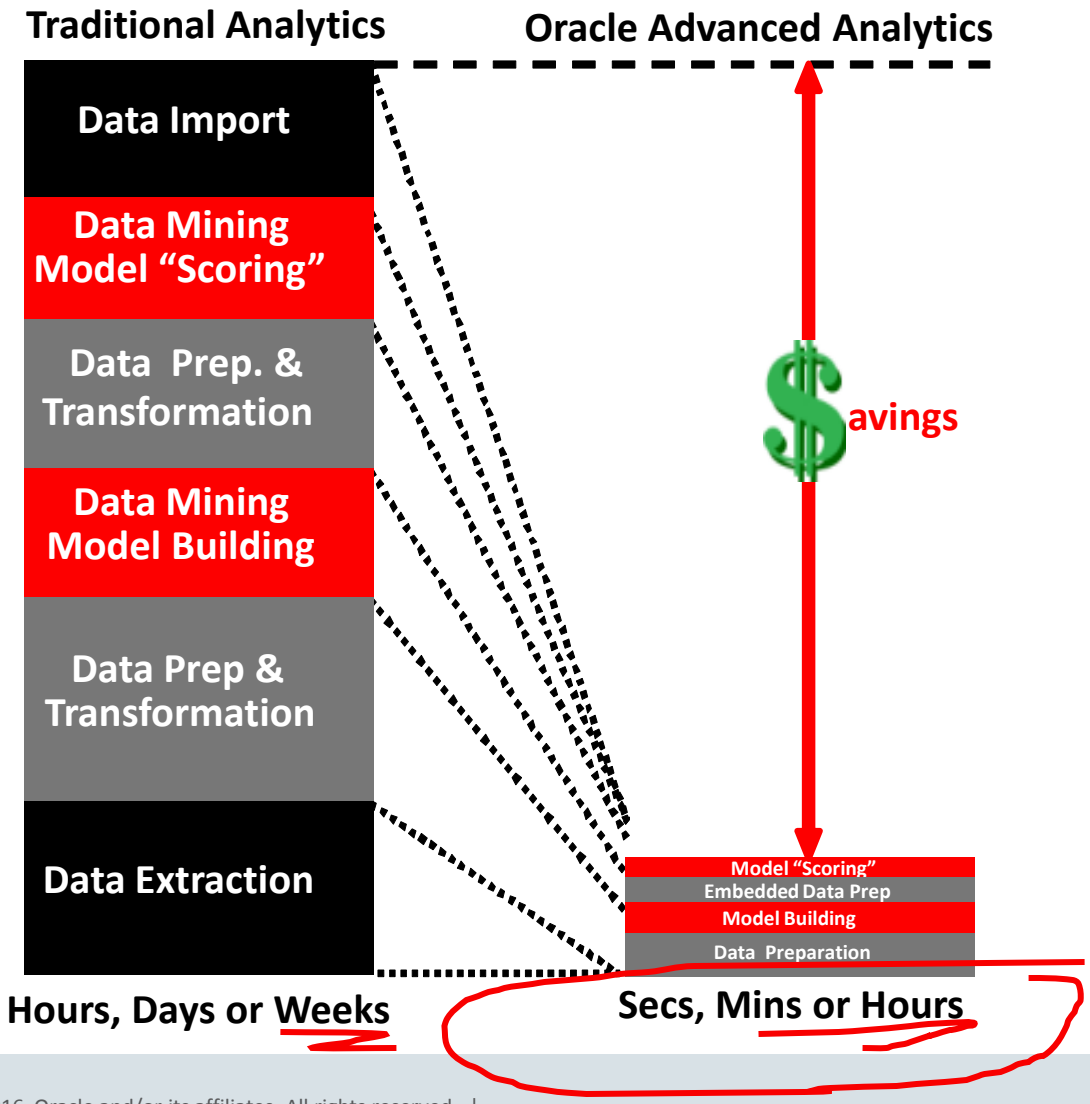
Oracle's Advanced Analytics

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

Oracle Advanced Analytics 12.2

Model Build Time Performance

Prelim/Unofficial

NEW IN 12.2

OAA 12.2 Algorithms	Rows (Ms)	T7-4 (Sparc & Solaris) Model Build Time (Secs / Degree of Parallelism)	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!*

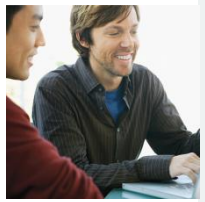
ORACLE

The way to read their results is that they compare 2 chips: X5 (Intel and Linux) and T7 (Sparc and Solaris). They are measuring scalability (time in seconds) with increase degree of parallelism (dop). The data also has high cardinality categorical columns which translates in 9K mining attributes (when algorithms require explosion). There are no comparisons to 12.1 and it is fair to say that the 12.1 algorithms could not run on data of this size.

Oracle's Advanced Analytics and Machine Learning Platform

Multiple interfaces across platforms — SQL, R, GUI, Dashboards, Apps

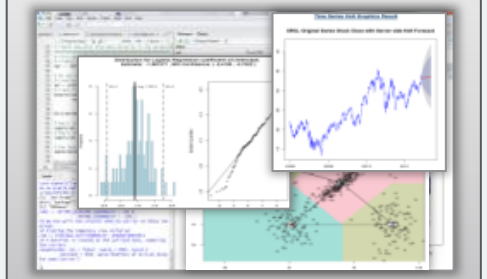
Users



Information Producers

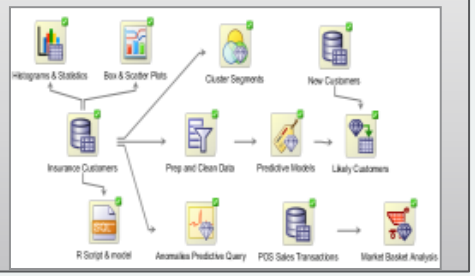
R programmers

R Client



Data & Business Analysts

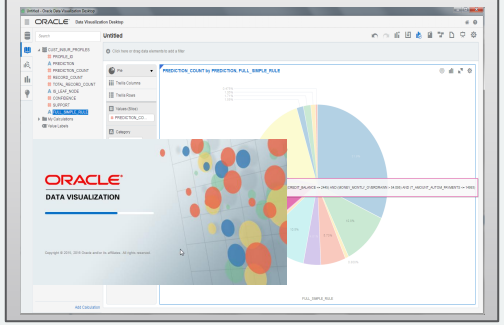
SQLDEV/ Oracle Data Miner



Information Consumers

Business Analysts/Mgrs

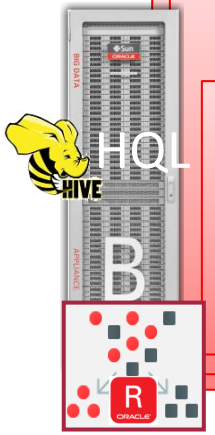
OBIEE/DV



Domain End Users (HCM, CRM)

Applications

Platform



Hadoop

ORAAH

Parallel, distributed algorithms

Oracle Database Enterprise Edition

Oracle Advanced Analytics - Database Option

SQL Data Mining, ML & Analytic Functions + R Integration for Scalable, Distributed, Parallel in-DB ML Execution

Oracle Database 12c

Oracle Cloud

ORACLE

You Can Think of Oracle 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`

+

SQL Statistical Functions - SQL & R

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

Statistical SQL "Verbs"

- `MEAN, STDEV`
- `MEDIAN`
- `SUMMARY`
- `CORRELATE`
- `FIT`
- `COMPARE`
- `ANOVA`

FREE!

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 & R

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

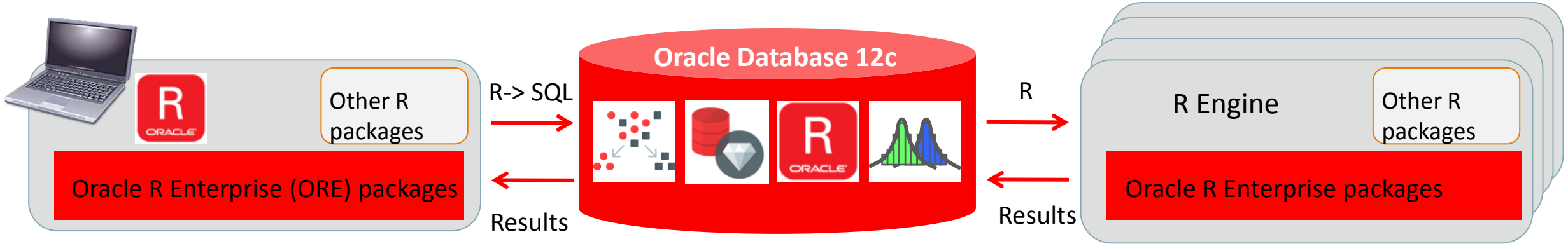
Analytical SQL "Verbs"

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

Oracle Advanced Analytics

How Oracle R Enterprise Compute Engines Work

Advanced Analytics



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

- 15+ 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

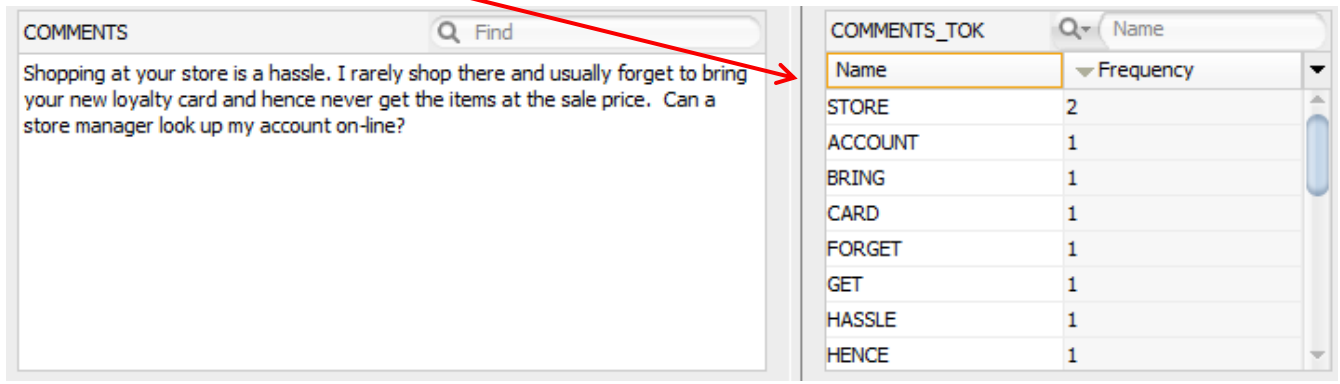
ORACLE

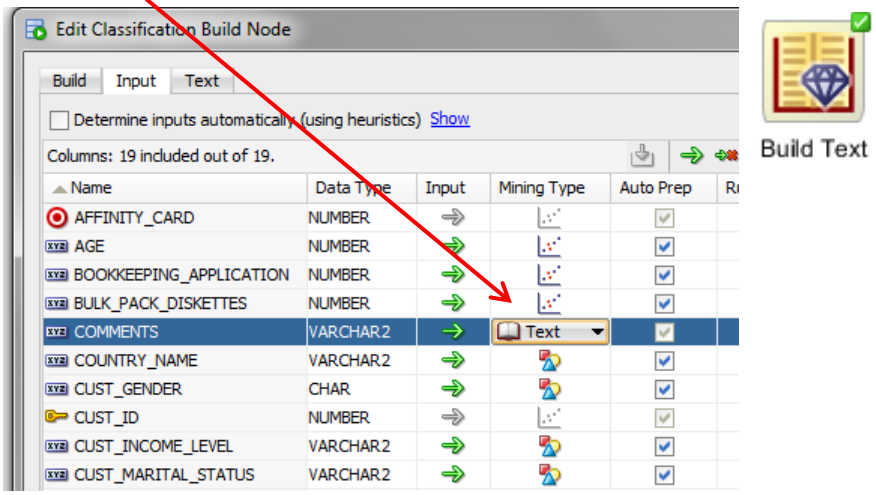
Oracle Text

ORACLE DATABASE 12c

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

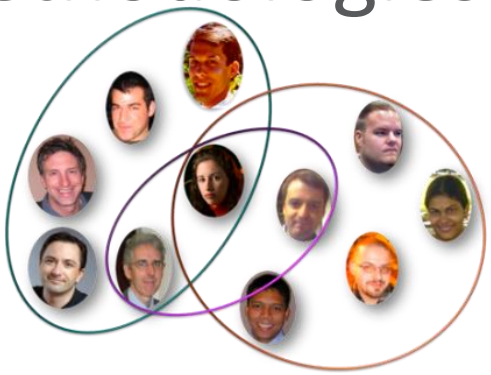




ORACLE

Rapidly Build, Evaluate & Deploy Analytical Methodologies

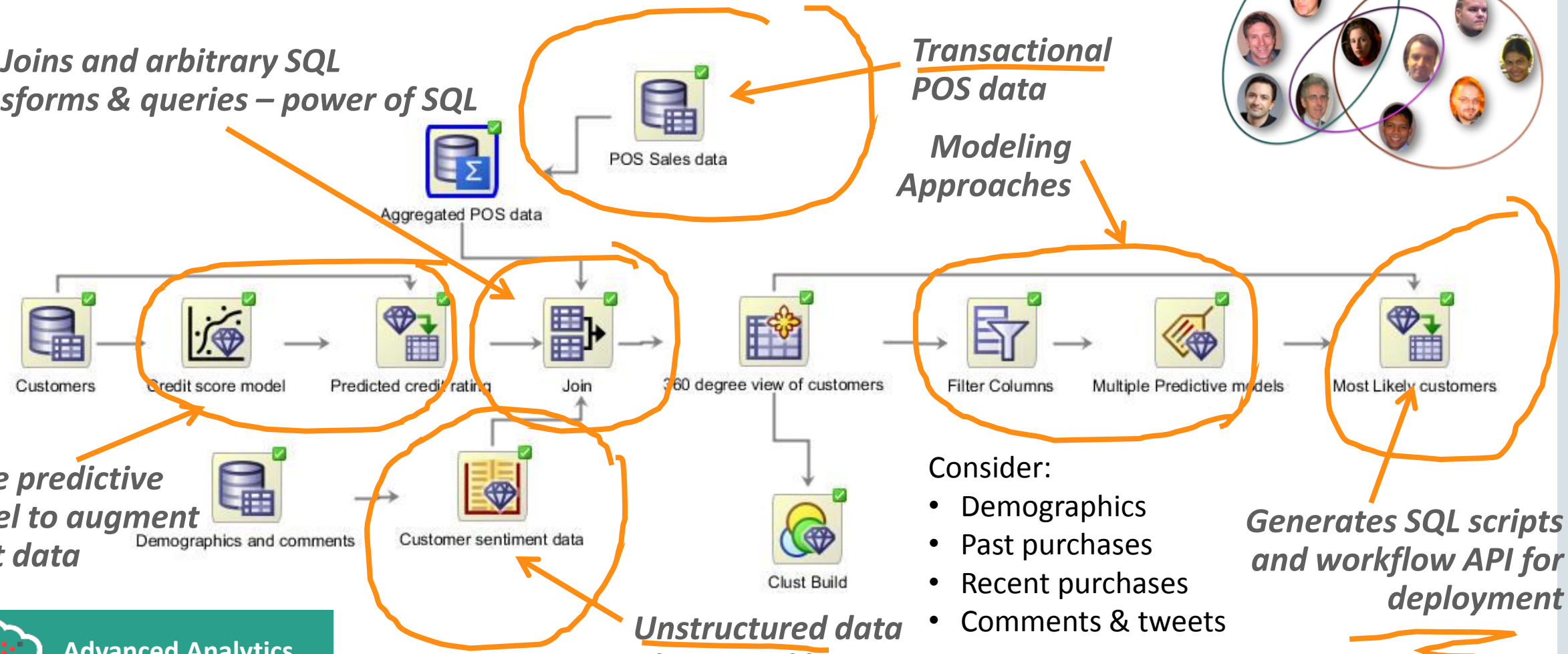
Leveraging a Variety of Data Sources and Types



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

Transactional POS data

Modeling Approaches



Inline predictive model to augment input data

Unstructured data also mined by algorithms

Consider:

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

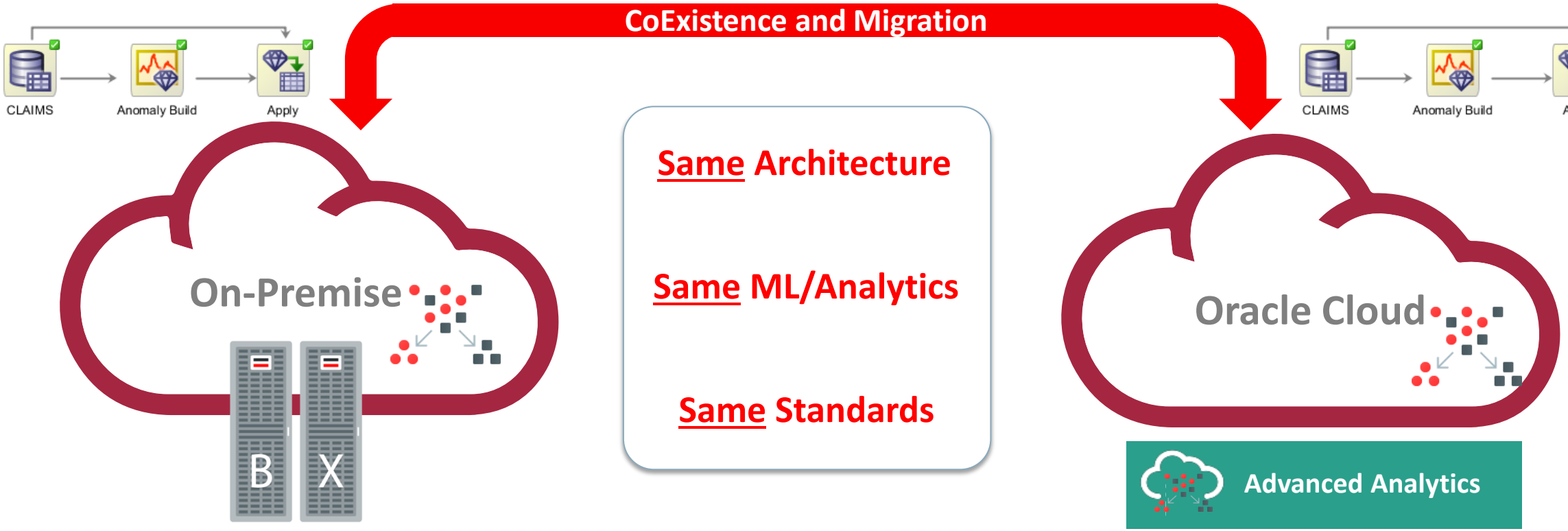
Generates SQL scripts and workflow API for deployment

Advanced Analytics

ORACLE

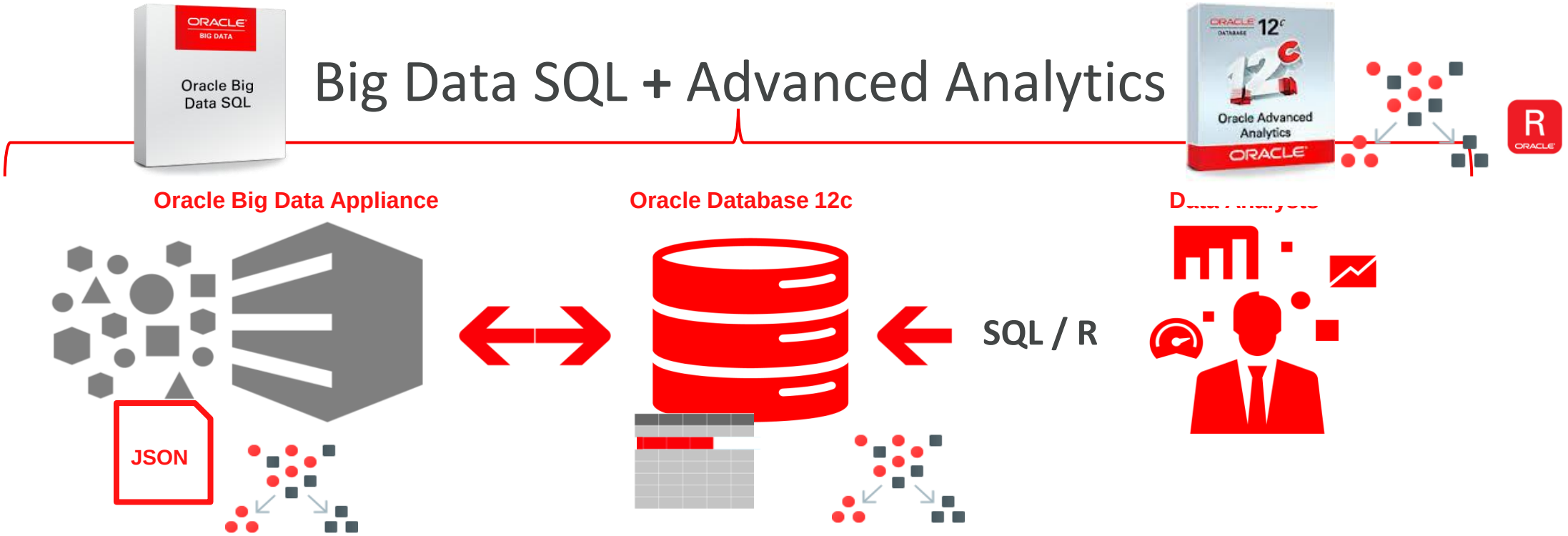
Oracle Advanced Analytics—On Premise or Cloud

100% Compatibility Enables Easy Coexistence and Migration



Transparently move workloads and ML/analytical methodologies between On-premise and public cloud

Manage and Analyze All Data—SQL & Oracle Big Data SQL



Structured and Unstructured Data Reservoir
- JSON data
- HDFS / Hive
- NoSQL
- Spatial and Graph data
- Image and Video data
- Social Media

Store business-critical data in Oracle
- Customer data
- Transactional data
- Unstructured documents, comments
- Spatial and Graph data
- Image and Video data
- Social Media

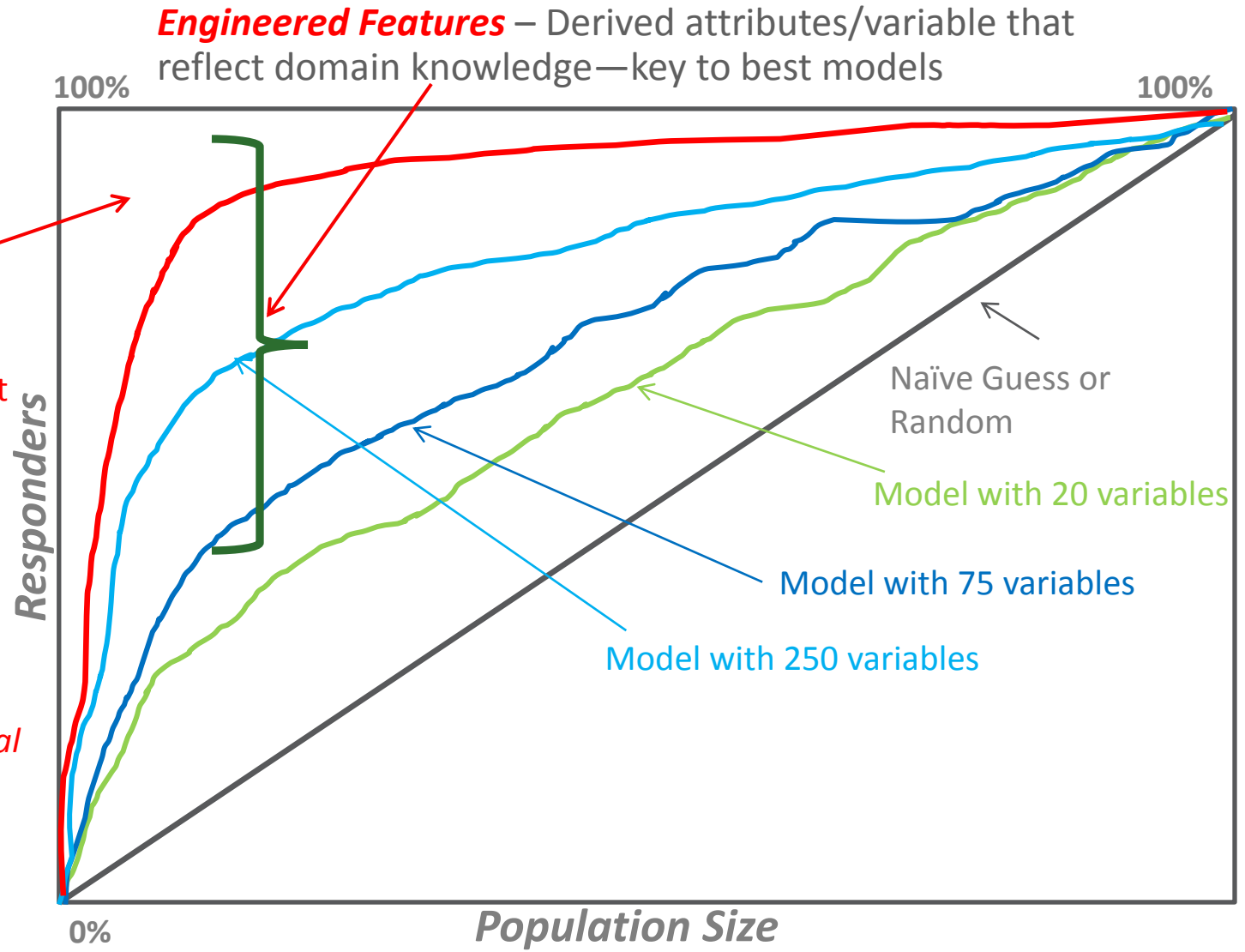
Data analyzed via SQL / R / GUI
- R Clients
- SQL Clients
- Oracle Data Miner

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.*



Oracle Advanced Analytics

Brief Demos

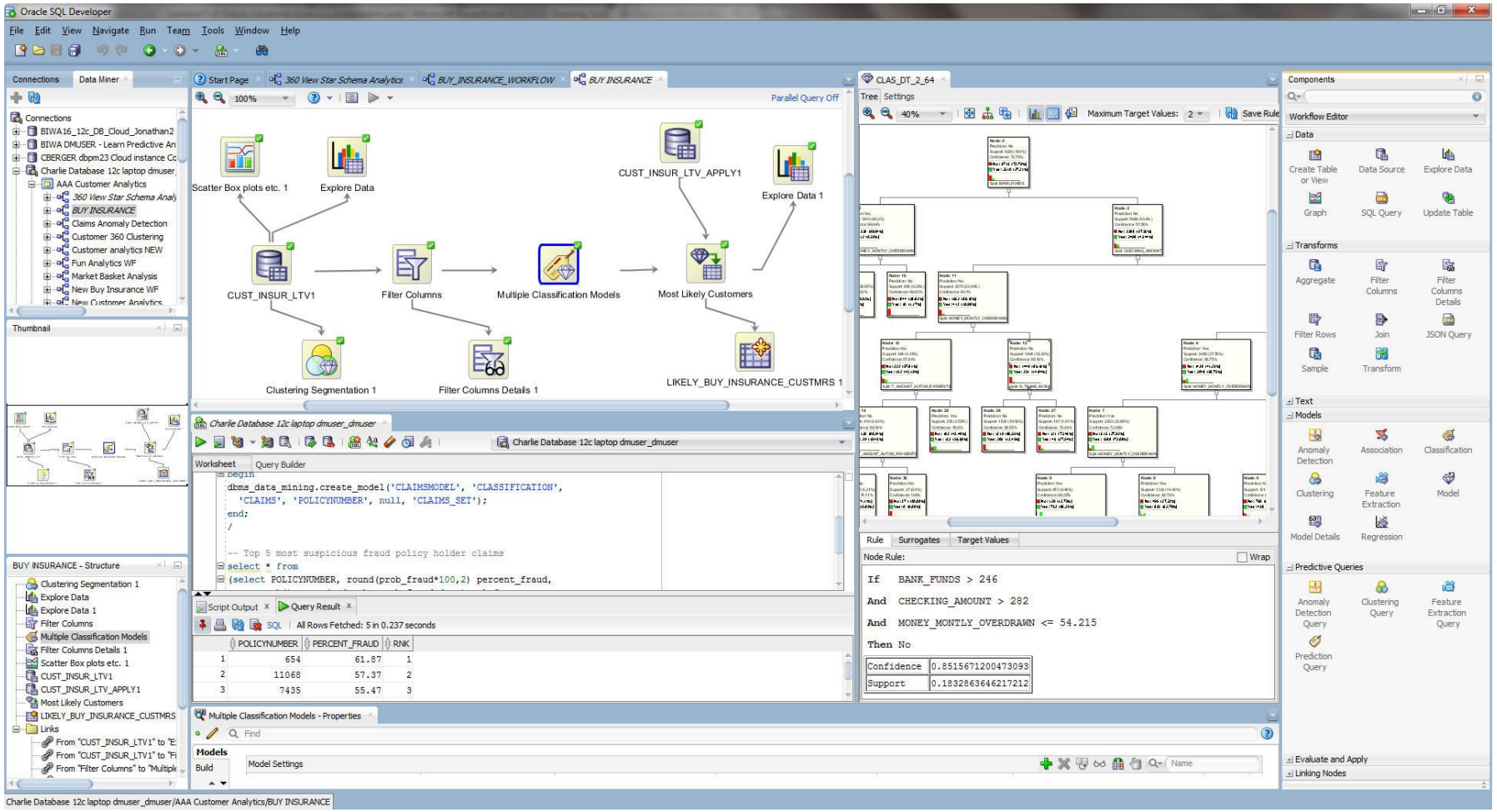
ORACLE

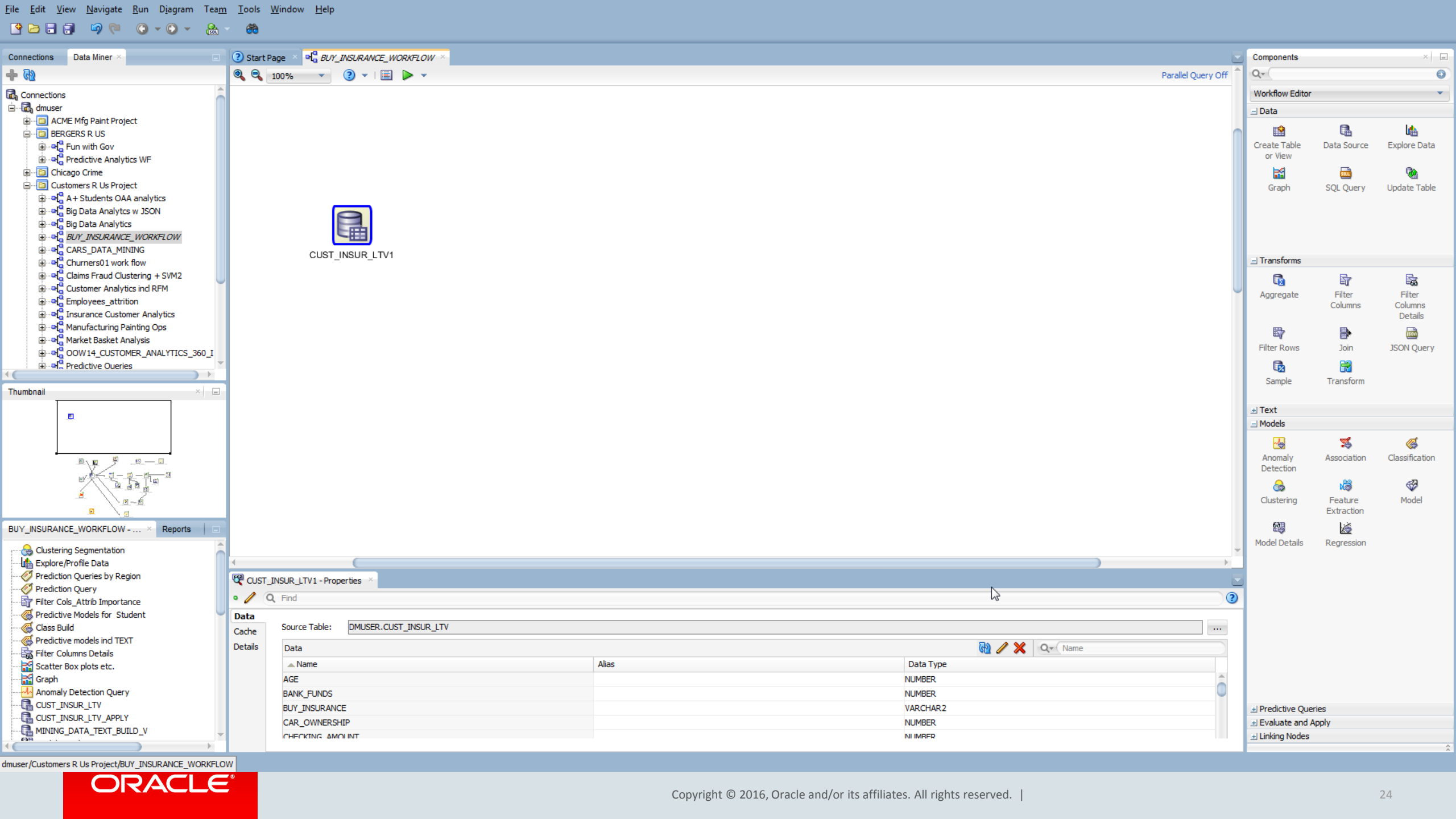
Oracle Data Miner GUI

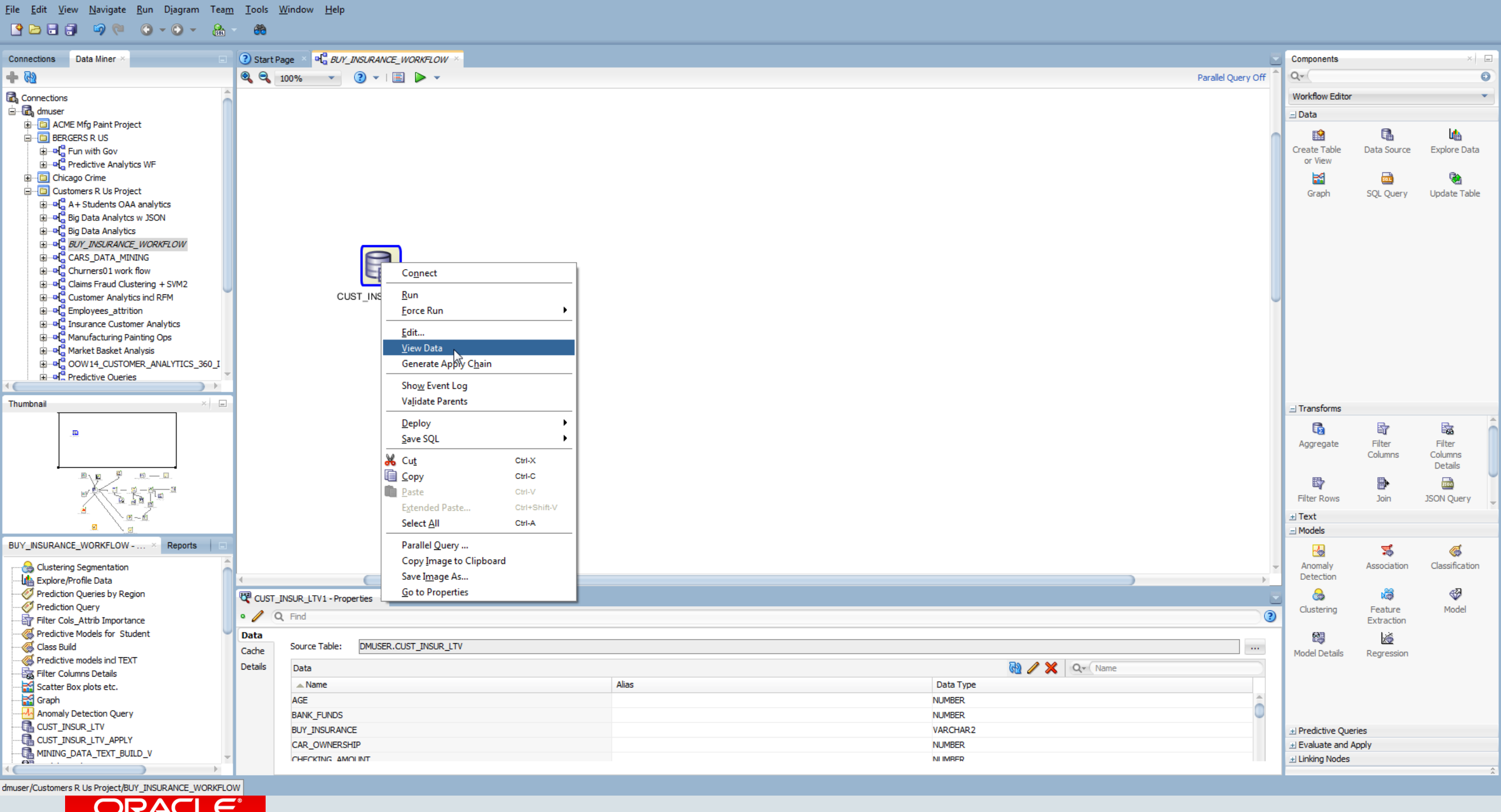
Easy to Use for "Citizen Data Scientist"

Advanced Analytics

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







File Edit View Navigate Run Team Tools Window Help

Connections | Data Miner

- 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 incl RFM
      - Employees_attrition
      - Insurance Customer Analytics
      - Manufacturing Painting Ops
      - Market Basket Analysis
      - OOW14_CUSTOMER_ANALYTICS_360_I
      - Predictive Queries
      - Public Sector Tax Audit
      - R_regression_models
      - Star Schema360 Degree Customer
    - Fun with Fraud

CUST_INSUR_LTV1 - Structure | Reports

No Structure

Start Page | *BUY_INSURANCE_WORKFLOW* | CUST_INSUR_LTV1

Data | Columns | SQL

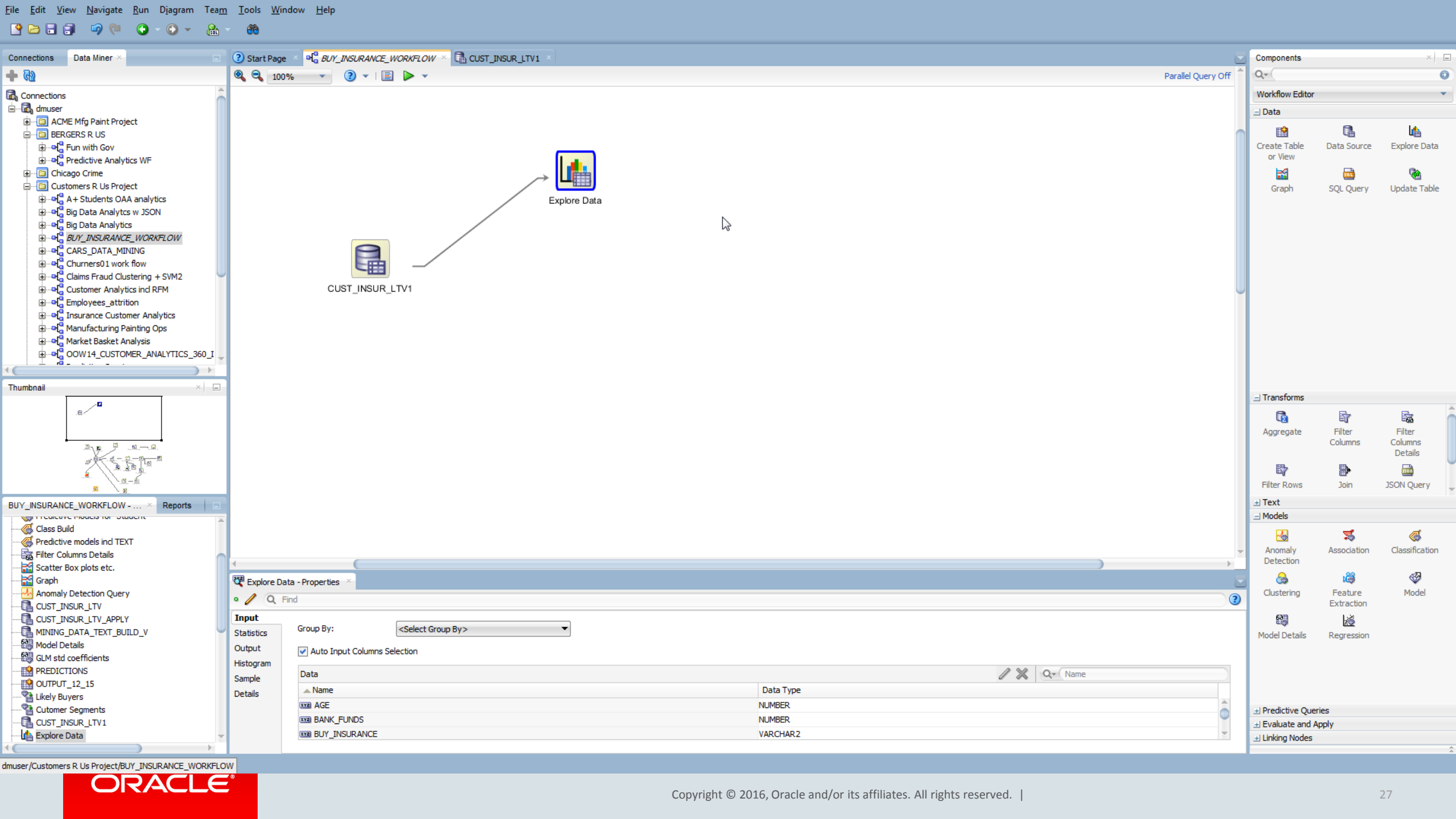
View: Actual Data | Sort... | Parallel Query Off... | Filter: Enter Where Clause

	CUST_ID	N_TRANS_ATM	CHECKING_AMOUNT	MARITAL_STATUS	SEX	N_TRANS_KIOSK	STATE	HOUSE_OWNERSHIP	MONTHLY_CHECKS_WRITTEN	LTV	SALARY	BANK_FUNDS	BUY_INSURANCE	CREDIT_BALANCE	N_MORTGA
1	CU2404	6	25	WIDOWED	F	9	DC	1	2	25,370...	63,883	25,000	No	549	
2	CU2405	0	25	MARRIED	M	2	MI	1	0	29,594.5	59,978	0	No	0	
3	CU2406	5	25	WIDOWED	M	2	CA	1	4	26,473.5	67,894	17,200	Yes	13,859	
4	CU2407	2	3,226	DIVORCED	F	2	MI	1	2	25,976...	67,107	7,600	No	0	
5	CU2408	4	130	DIVORCED	F	2	NY	1	11	28,446.5	70,986	1,500	No	23,525	
6	CU2409	4	6,008	DIVORCED	F	2	CA	0	2	13,921...	58,887	10,900	No	0	
7	CU2411	4	10,943	MARRIED	M	2	NY	1	2	23,833...	69,335	0	No	4,734	
8	CU2412	5	25	MARRIED	F	2	FL	1	12	22,485.5	58,342	2,600	Yes	0	
9	CU2413	4	293	SINGLE	M	1	NY	0	2	18,865...	67,061	1,614	Yes	2,249	
10	CU2414	2	25	DIVORCED	M	2	CA	1	2	29,672...	67,489	0	No	0	
11	CU2416	2	25	SINGLE	M	2	NY	0	4	20,190...	63,963	0	No	0	
12	CU2417	2	2,420	SINGLE	M	2	NY	0	4	19,941	64,964	1,156	No	963	
13	CU2418	2	1,089	DIVORCED	F	2	MI	1	3	27,892...	58,771	6,850	No	0	
14	CU2420	5	25	DIVORCED	M	2	NY	1	3	19,536.5	60,946	5,400	No	44,020	
15	CU2421	5	882	DIVORCED	F	2	CA	1	4	29,123.5	67,694	3,950	No	0	
16	CU2422	3	157	MARRIED	F	2	MI	1	4	29,145...	61,781	1,850	Yes	0	
17	CU2423	4	9,894	WIDOWED	F	8	NY	1	12	25,868...	63,875	3,300	No	6,707	
18	CU3100	6	227	WIDOWED	F	2	MI	1	9	17,539...	65,359	4,280	No	0	
19	CU3101	4	40	MARRIED	F	2	MI	1	11	23,178...	61,513	800	Yes	0	
20	CU3102	4	21,094	SINGLE	M	1	NV	0	4	21,644...	65,379	8,129	No	0	
21	CU3104	5	25	MARRIED	F	2	NY	1	3	16,525.5	66,902	17,001	No	92,408	
22	CU3105	7	638	OTHER	F	2	NY	1	3	24,184.5	63,538	20,200	Yes	0	
23	CU3106	4	372	DIVORCED	M	2	MI	1	12	26,580.5	69,522	0	No	0	
24	CU3107	0	25	SINGLE	M	2	CA	0	0	19,374	64,296	0	No	0	
25	CU3108	3	25	MARRIED	F	2	NY	1	4	14,843...	60,575	700	Yes	0	
26	CU3109	4	25	MARRIED	M	2	MI	1	1	30,839.5	74,158	0	No	0	
27	CU3110	4	25	MARRIED	F	2	CA	1	2	21,882...	59,929	2,300	No	36,102	
28	CU3111	4	25	DIVORCED	M	2	MN	1	7	27,128...	55,713	3,650	No	0	
29	CU3113	1	25	SINGLE	M	2	MS	0	1	15,834.5	64,138	0	No	0	
30	CU3115	2	25	MARRIED	M	3	CA	1	3	27,130	66,120	1,600	No	0	
31	CU3116	4	17,173	DIVORCED	F	2	CA	1	18	24,554	68,616	4,000	No	11,098	
32	CU3117	4	13,225	SINGLE	M	1	MI	0	11	16,890	65,560	1,606	No	0	
33	CU3118	4	7,543	SINGLE	M	3	CA	1	1	28,260...	71,843	700	No	0	
34	CU3119	4	25	DIVORCED	F	3	MI	1	3	27,410	58,840	0	No	0	
35	CU3120	2	25	SINGLE	M	3	CA	0	1	0	63,095	0	No	0	
36	CU3121	4	543	DIVORCED	M	3	CA	1	4	28,178...	60,315	5,700	No	0	
37	CU3123	1	25	SINGLE	M	3	NY	0	1	24,367.5	72,670	0	No	0	
38	CU3124	4	12,988	SINGLE	M	3	MN	0	2	17,477.5	70,310	10,100	No	0	
39	CU3125	6	25	DIVORCED	M	3	CA	1	16	22,995	68,380	5,950	No	66,417	
40	CU3126	2	108	DIVORCED	F	3	MI	1	9	24,825	60,900	7,200	No	0	

Components

No Components

dmuser/Customers R Us Project/BUY_INSURANCE_WORKFLOW



Oracle SQL Developer : Charlie Database 12c laptop dmuser_dmuser/AAA Customer Analytics/BUY INSURANCE/Scatter Box plots etc.

File Edit View Navigate Run Team Tools Window Help

Connections | Data Miner

Connections
- BIWA16_12c_DB_Cloud_Jonathan2
- BIWA DMUSER - Learn Predictive Analtics 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

Thumbnail

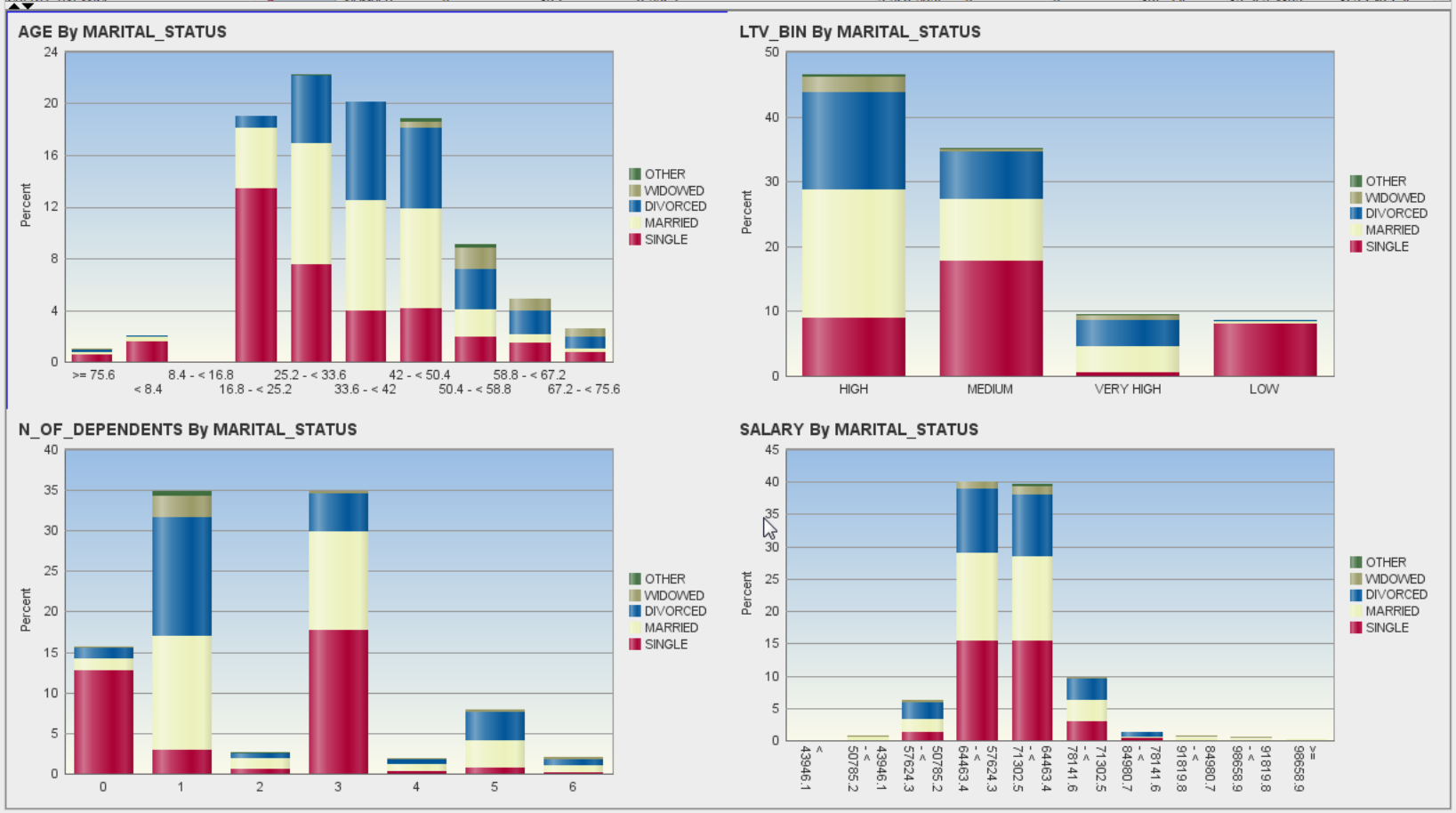
SSH Hosts
- SSH Hosts
  - CBERGER dbpm23 Cloud instance
  - Express Cloud PDB May 25
  - HOL3
  - NEW_SSH_BIWA_Brian2
  - OracleCloudDatabase1
  - OracleDatabaseCloud2
  - SSH_BIWA_Jonathan2

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

Statistics | Data | Columns | SQL

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

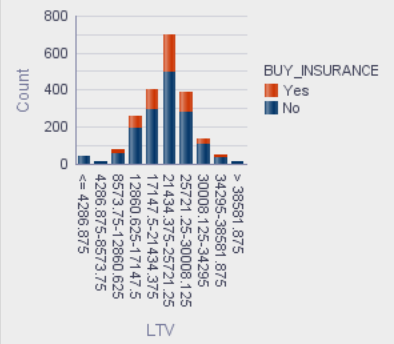


Scatter Box plots etc. | Components

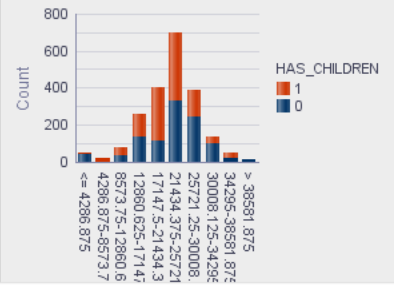
View: Sample Data | Parallel Query Off...

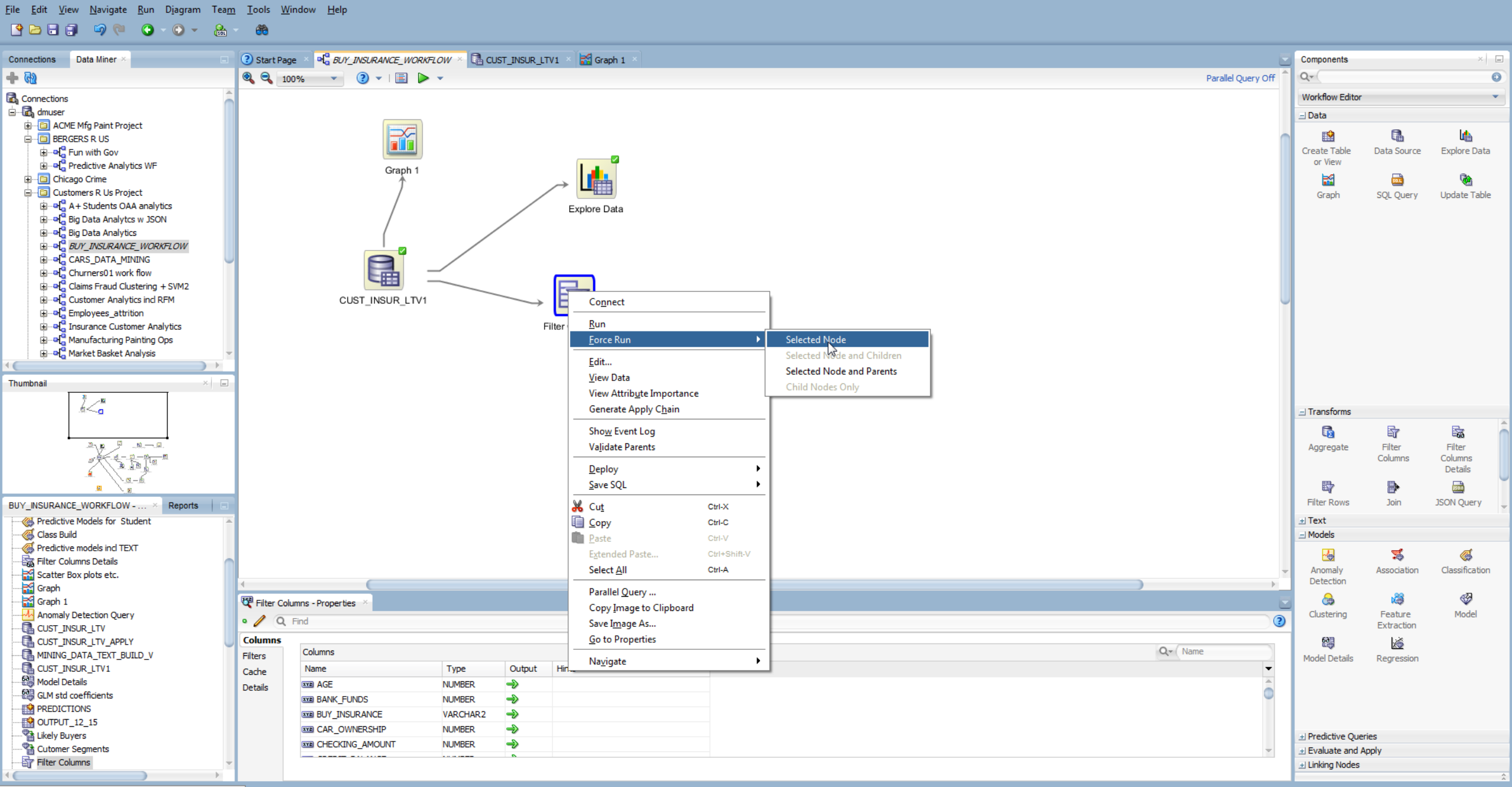
Graph1

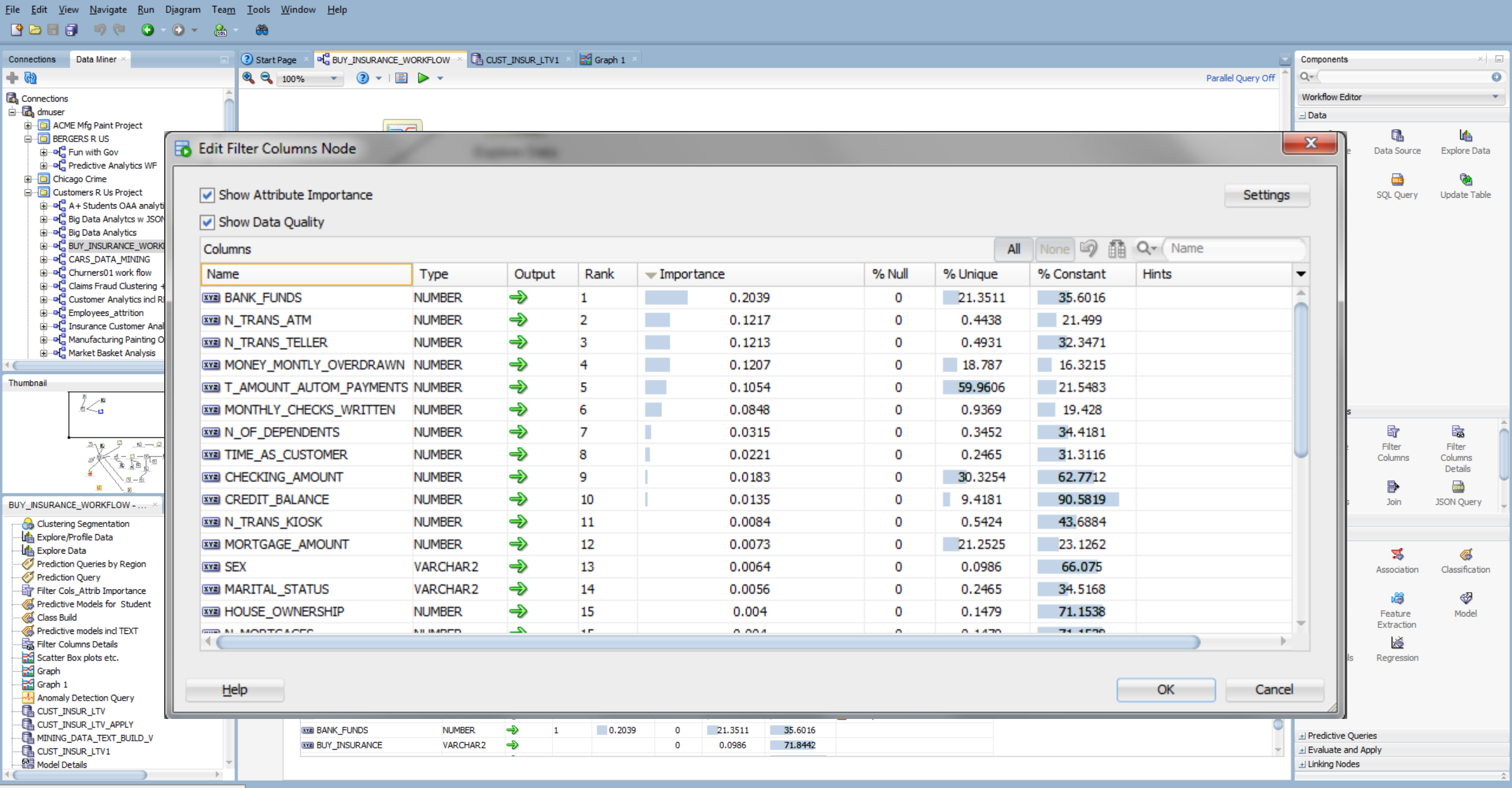
Graph2

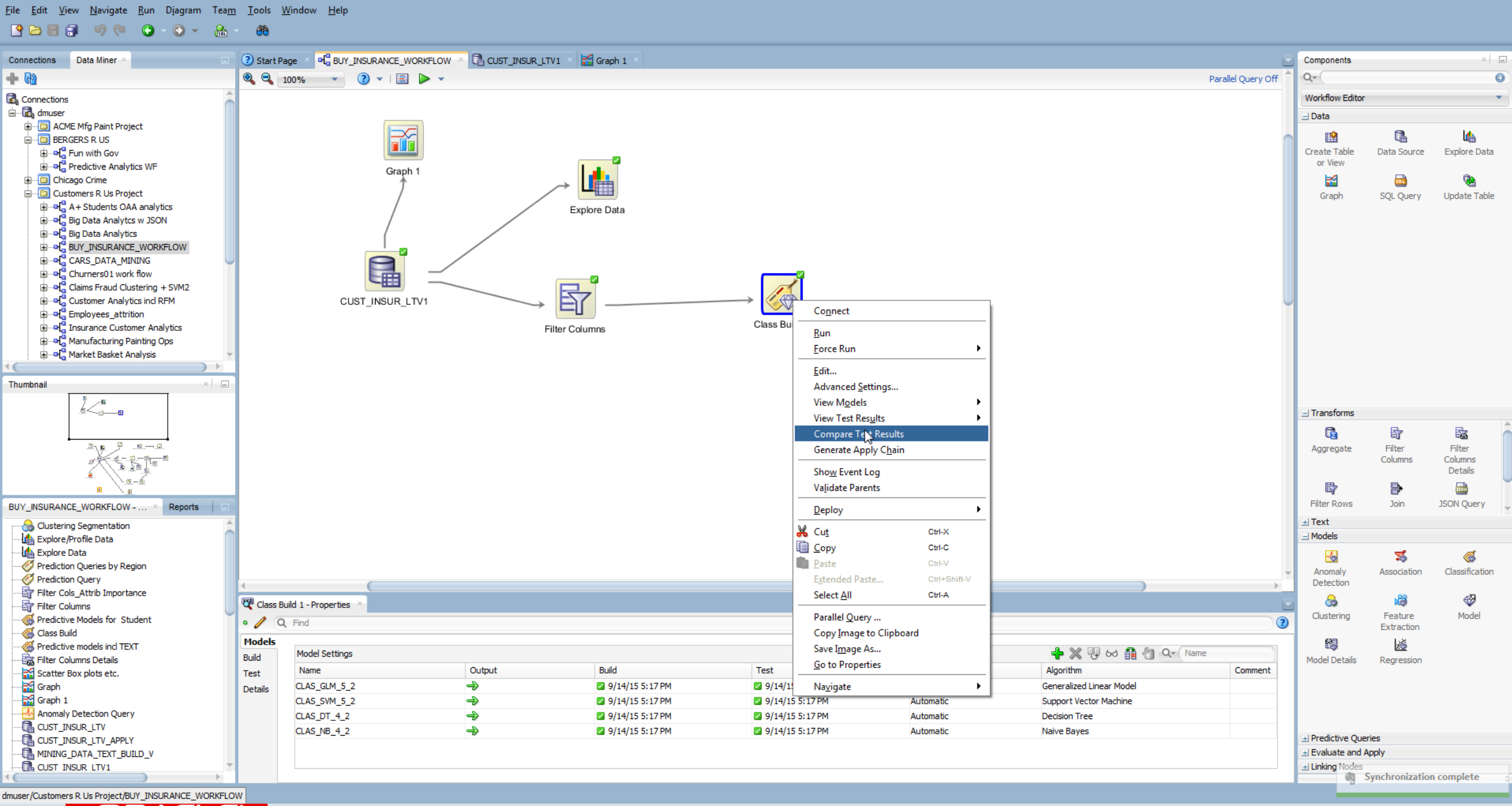


Graph3









File Edit View Navigate Run Team Tools Window Help

Connections | Data Miner

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 incl RFM
    - Employees_attrition
    - Insurance Customer Analytics
    - Manufacturing Painting Ops
    - Market Basket Analysis
    - OOW14_CUSTOMER_ANALYTICS_360_I
    - Predictive Queries
    - Public Sector Tax Audit
    - R_regression_models

Class Build 1 - Structure | Reports

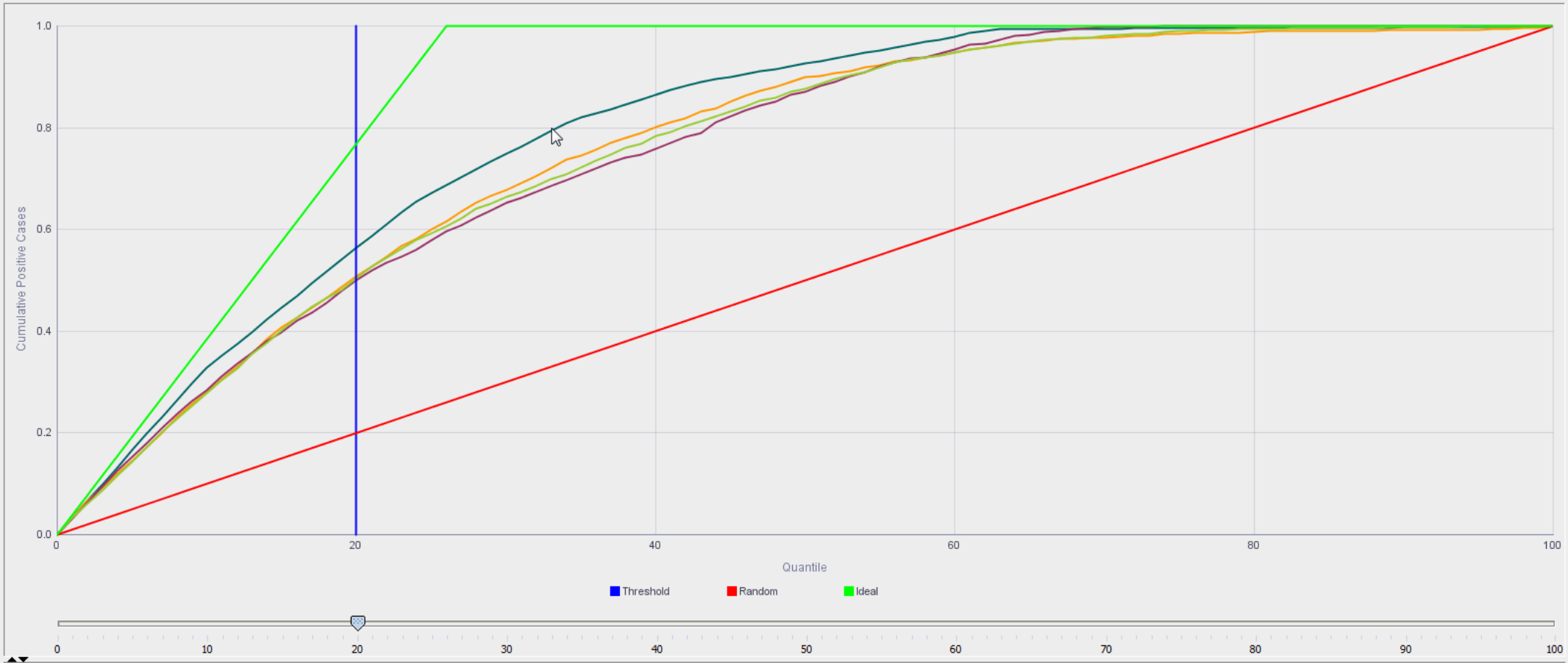
No Structure

Start Page | BUY_INSURANCE_WORKFLOW | CUST_INSUR_LTV1 | Graph 1 | Class Build 1

Performance | Performance Matrix | ROC | Lift | Profit

Display: Cumulative Positive Cases

Target Value: Yes



Models

Model	Lift Cumulative	Gain Cumulative %	Records Cumulative %	Target Density Cumulative	Algorithm	Creation Date
CLAS_DT_4_2	2.7885	56.3528	20.2093	0.7518	Decision Tree	9/14/15 5:17 PM
CLAS_GLM_5_2	2.5146	50.8187	20.2093	0.678	Generalized Linear Model	9/14/15 5:17 PM
CLAS_NB_4_2	2.4726	49.9697	20.2093	0.6667	Naive Bayes	9/14/15 5:17 PM
CLAS_SVM_5_2	2.4936	50.3942	20.2093	0.6723	Support Vector Machine	9/14/15 5:17 PM

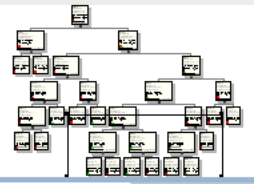
dmuser/Customers R Us Project/BUY_INSURANCE_WORKFLOW

File Edit View Navigate Run Diagram Team Tools Window Help

Connections | Data Miner

- 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 incl RFM
      - Employees_attrition
      - Insurance Customer Analytics
      - Manufacturing Painting Ops
      - Market Basket Analysis

Thumbnail



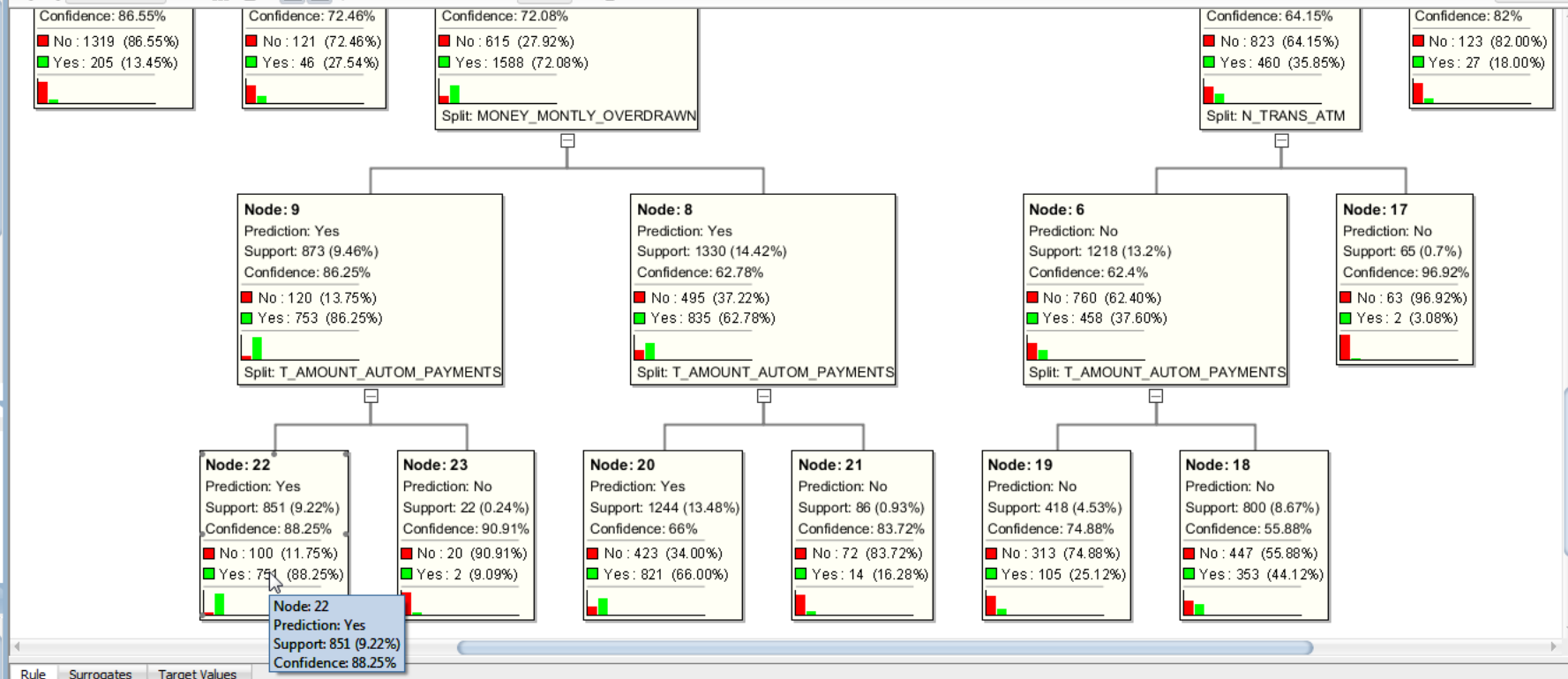
CLAS_DT_4_2 - Structure | Reports

- 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

Start Page | BUY_INSURANCE_WORKFLOW | CUST_INSUR_LTV1 | Graph 1 | CLAS_DT_4_2

Tree | Settings

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



Rule | Surrogates | Target Values

Node Rule: Wrap

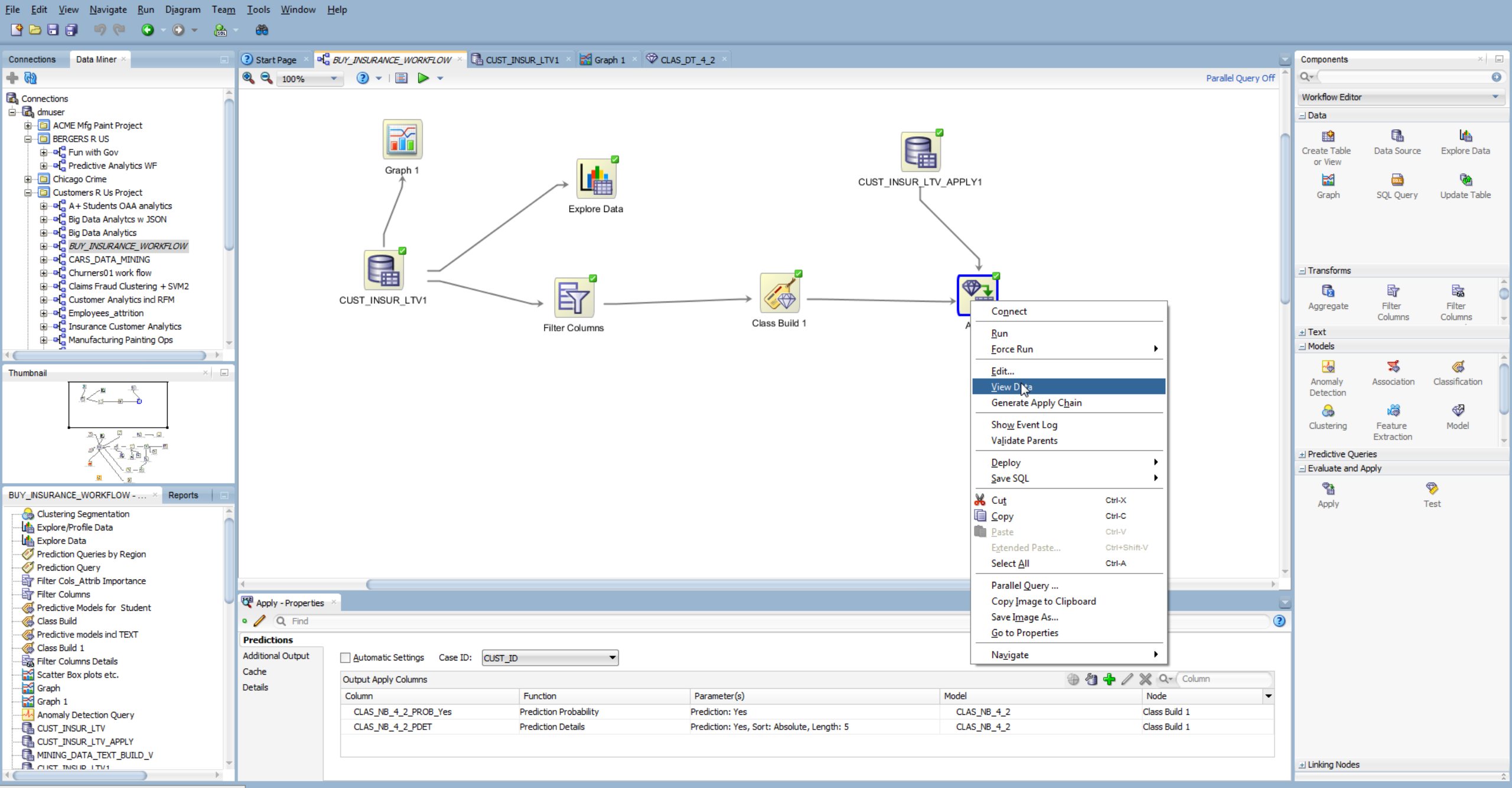
```
If   BANK_FUNDS > 246
And  CHECKING_AMOUNT <= 282
And  CREDIT_BALANCE <= 2445
And  MONEY_MONTLY_OVERDRAWN > 54.095
And  T_AMOUNT_AUTOM_PAYMENTS <= 14993
Then Yes
```

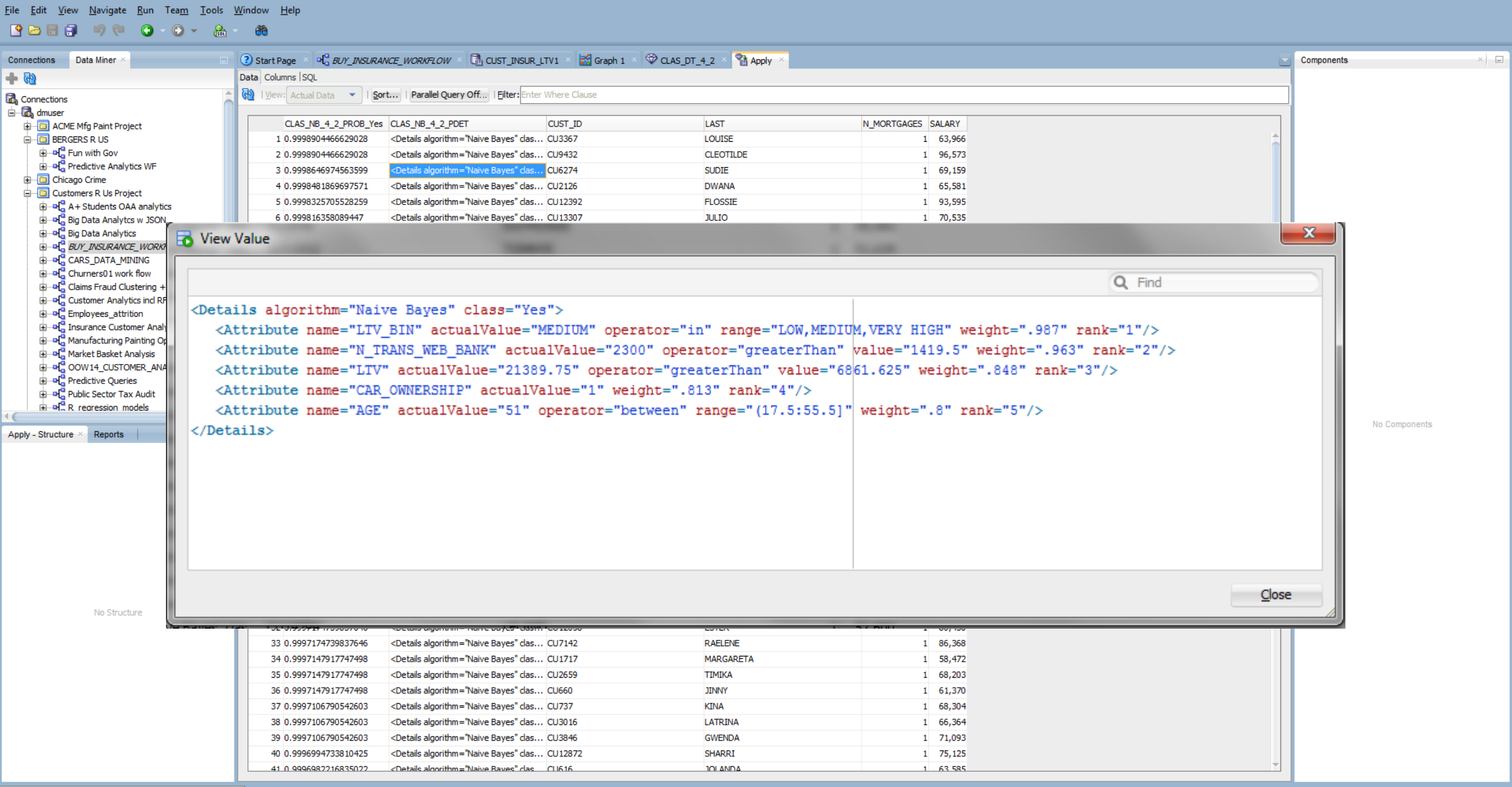
Confidence	0.882491186839013
Support	0.09223932365055279

Components

No Components

dmuser/Customers R Us Project/BUY_INSURANCE_WORKFLOW



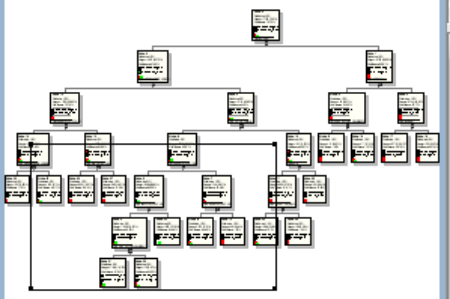


File Edit View Navigate Run Diagram Team Tools Window Help

Connections | Data Miner

- 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 incl 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

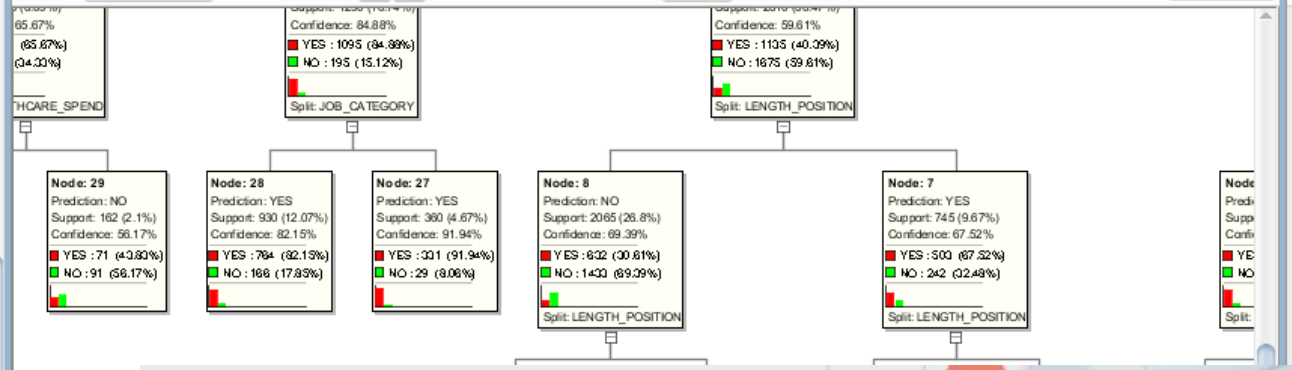
Thumbnail



Start Page | *Employees_attrition* | Hist. Employee data | CLAS_DT_1_10

Tree | Settings

64% | Maximum Target Values: 2 | Save Rules...

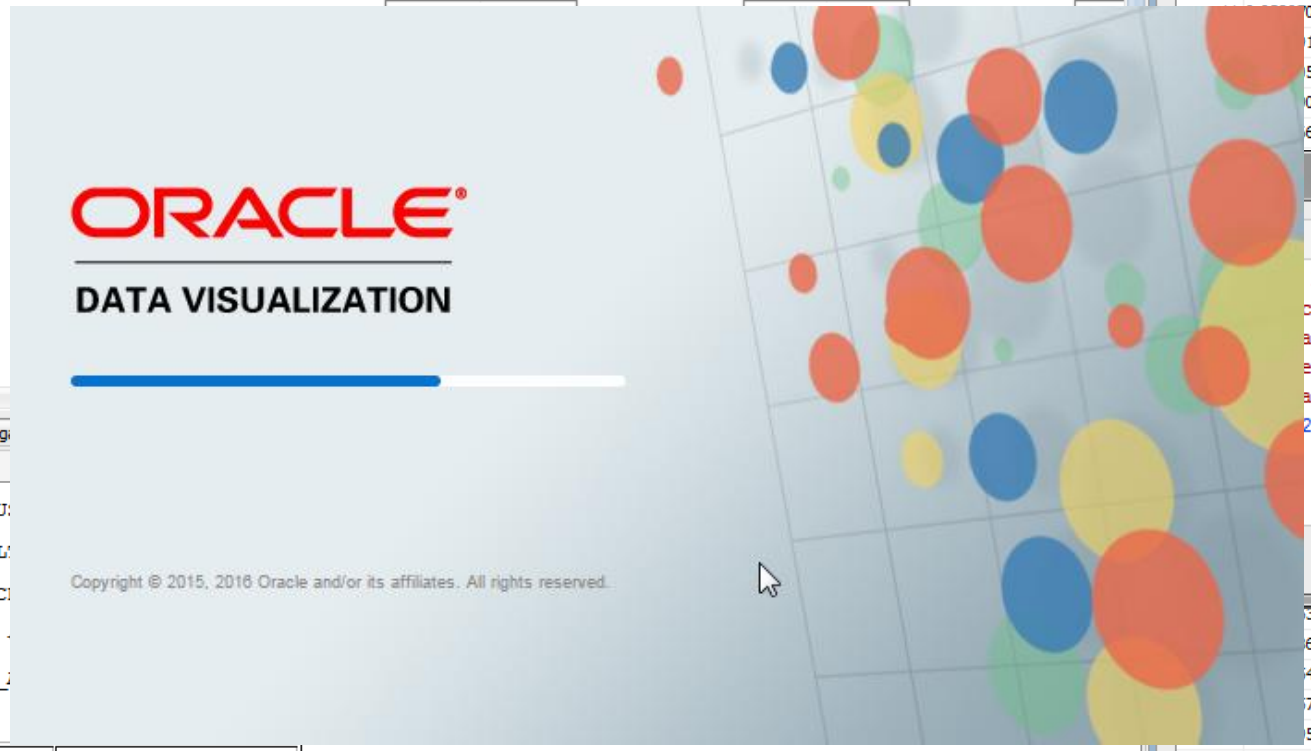


Rule | Surrog

Node Rule:

```
If   BONU
And  HEAL
And  STOC
And  2.5
And  AMT_
Then NO
```

Confidence	0.6183533447684391
Support	0.15131066701271736

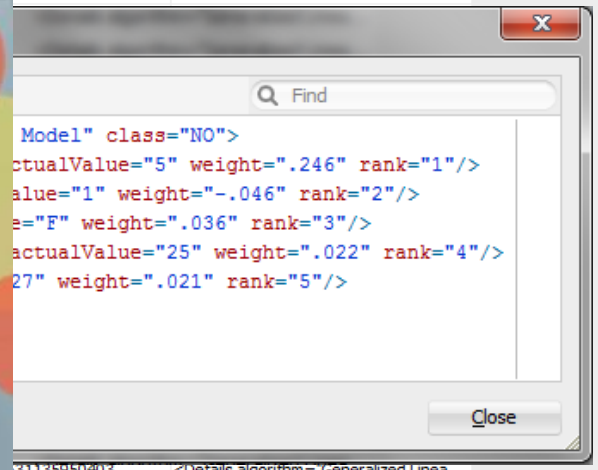


At Risk Employees

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...
	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...



Charlie Database 12c laptop dmuser _dmuser/Customers R Us Project/Employees_attrition

Untitled - Oracle Data Visualization Desktop

ORACLE Data Visualization Desktop

Search

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

Untitled

Click here or drag data elements to add a filter

Pie

Trellis Columns

Trellis Rows

Values (Slice)

PREDICTION_CO...

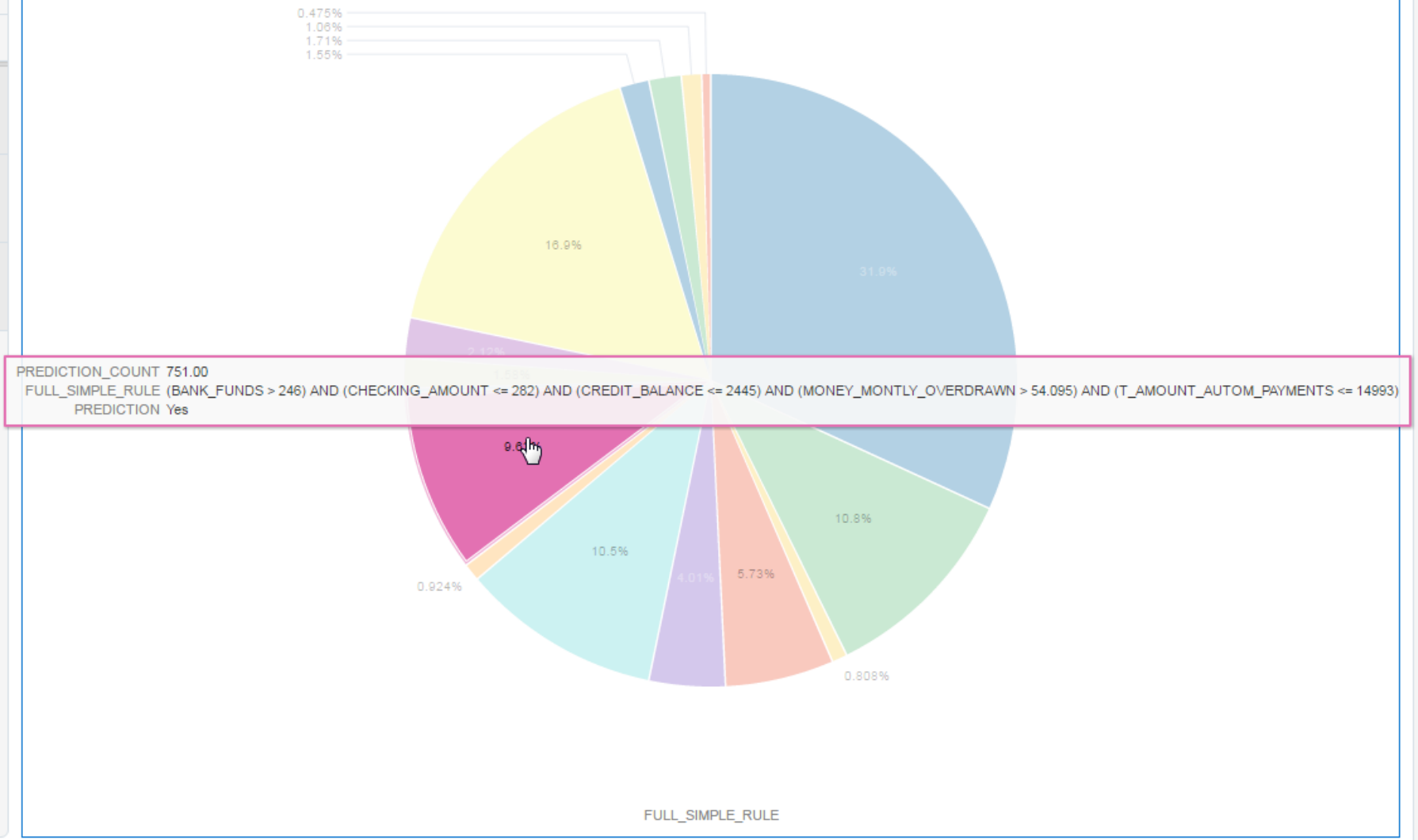
Category

PREDICTION

Color

FULL_SIMPLE_RU..

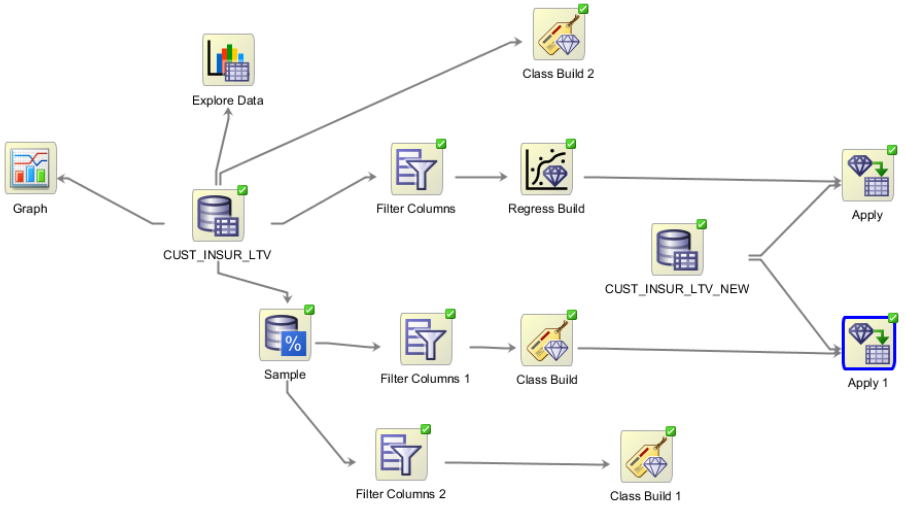
PREDICTION_COUNT by PREDICTION, FULL_SIMPLE_RULE

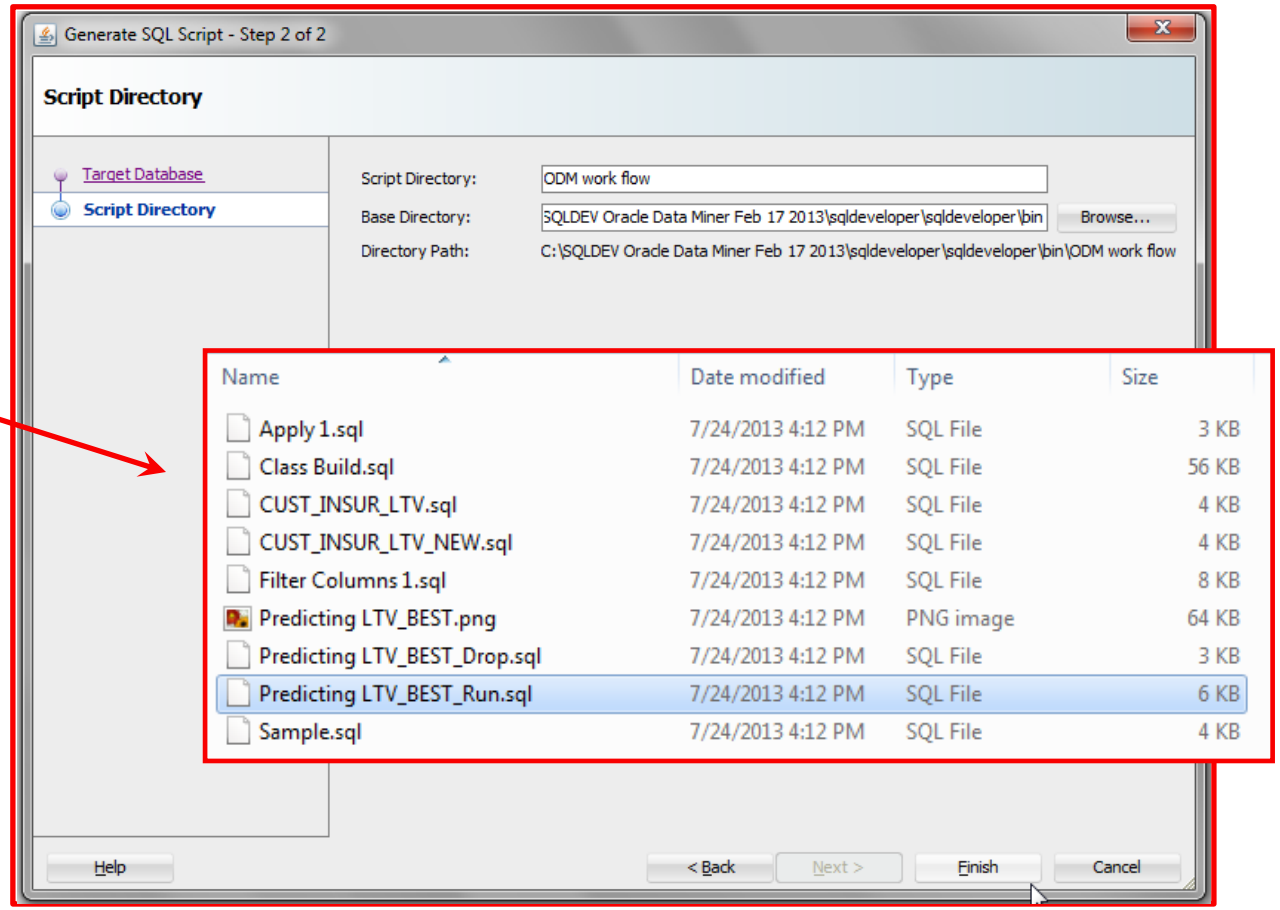


Sharing, Automation and Deployment

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





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;
```

Script Output | Query Result

SQL | All Rows Fetched: 5 in 0.064 seconds

	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_64, 'Yes'
  USING 7800 as bank_funds, 125 as checking_amount, 20 as
   credit_balance, 55 as age, 'Married' as marital_status,
   250 as MONEY_MONTLY_OVERDRAWN, 1 as house_ownership)
from dual;
```

Likelihood to respond:

Query Result

SQL | All Rows Fetched: 1 in 0 seconds

PREDICTION_PROB...
0.8382936507936...

Oracle's Advanced Analytics

Example Customer References

Fiserv

Risk Analytics in Electronic Payments

Objectives

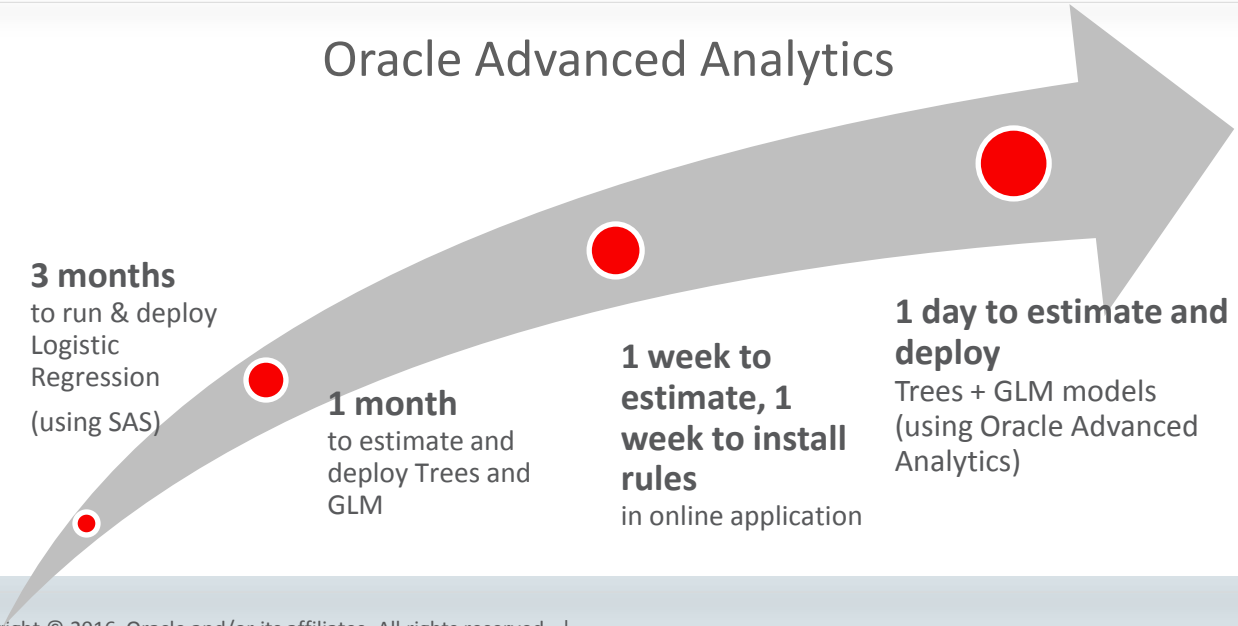
- Prevent $200M in losses every year using data to monitor, understand and anticipate fraud

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.



UK National Health Service

Combating Healthcare Fraud

NHS Business Services Authority

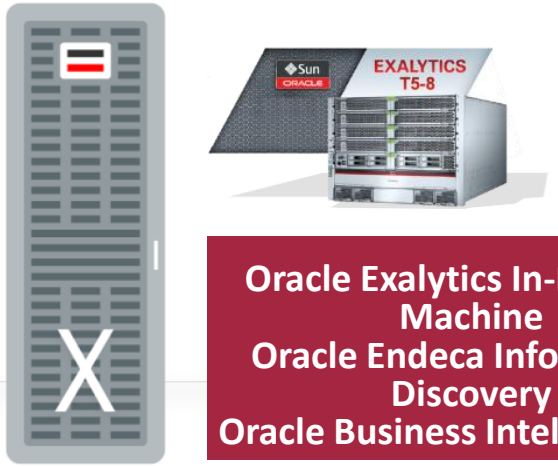
Objectives

- 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

- "Oracle Advanced Analytics' data mining capabilities and Oracle Exalytics' performance really impressed us. The overall solution is very fast, and our investment very quickly provided value. We can now do so much more with our data, resulting in significant savings for the NHS as a whole"
  – Nina Monckton, Head of Information Services, NHS Business Services Authority



Oracle Exadata Database Machine

Oracle Advanced Analytics

Oracle Exalytics In-Memory Machine
Oracle Endeca Information Discovery
Oracle Business Intelligence EE

ORACLE

DX Marketing

Cloud Based Predictive Analytics/Database Marketing

DX | Marketing

Objectives

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

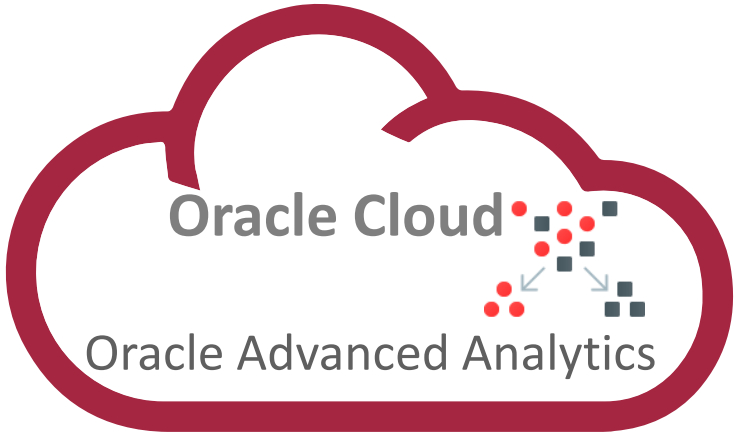
Solution

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

ORACLE

Zagrebačka Bank (biggest bank in Croatia)

Increases Cash Loans by 15% Within 18 Months of Deployment

Zagrebačka banka
UniCredit Group

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

An Post

Boosts Retail & Postal Services with Big Data & Analytics Platform

Objectives

- Provide a scalable big data and analytics platform to manage millions of daily transactions, facilitate new postal services, and align with market needs and the growth in post office retail services

Solution

- Deployed a big data and analytics platform using Oracle Exadata Database Machine,Oracle Advanced Analytics, … to …, combat fraud, and enable readily-available enterprise wide business …within core areas such as mails, parcels, and retail
- Enabled An Post to rapidly analyze and respond …using Oracle Advanced Analytics…
- Vertice (partner) provided expert services including …in-depth product and data mining expertise services

- "With our Oracle solution for Big Data and Analytics we now have an analytics platform that has completely transformed our service delivery model. It provides An Post with a single source of truth while automatically consolidating data from all post offices nationwide."

– John Cronin, Group Chief Information Officer, An Post Limited

"Big data is a vital part of our future," says John Cronin, CIO of Ireland's postal network.

http://www.oracle.com/us/corporate/customers/customersearch/an-post-1-big-data-2759887.html

https://www.siliconrepublic.com/enterprise/2016/02/05/an-post-oracle-john-cronin-five-minute-cio

Oracle Advanced Analytics

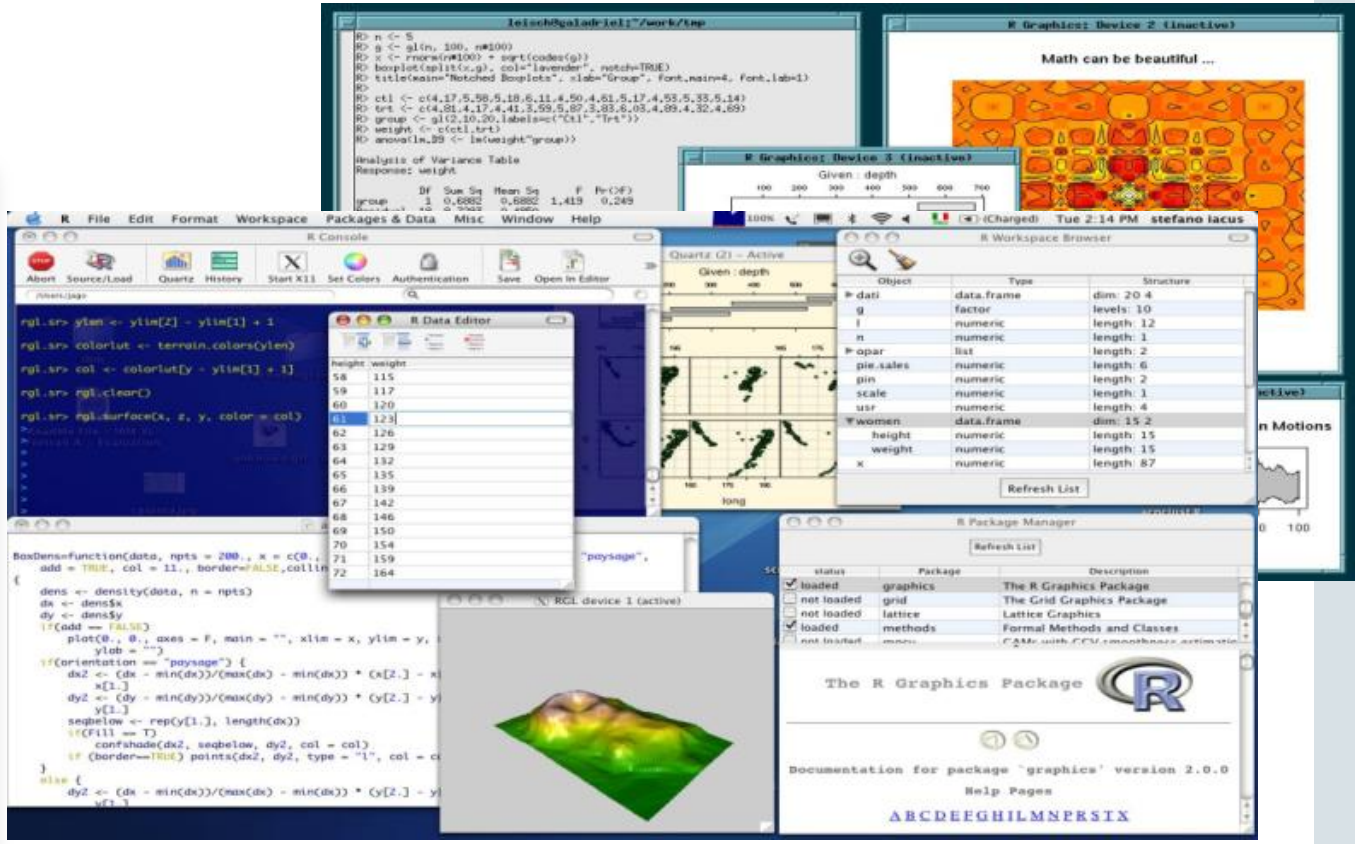
OAA/Oracle R Enterprise (R integration)

R—Widely Popular

R is a statistics language similar to Base SAS or SPSS statistics

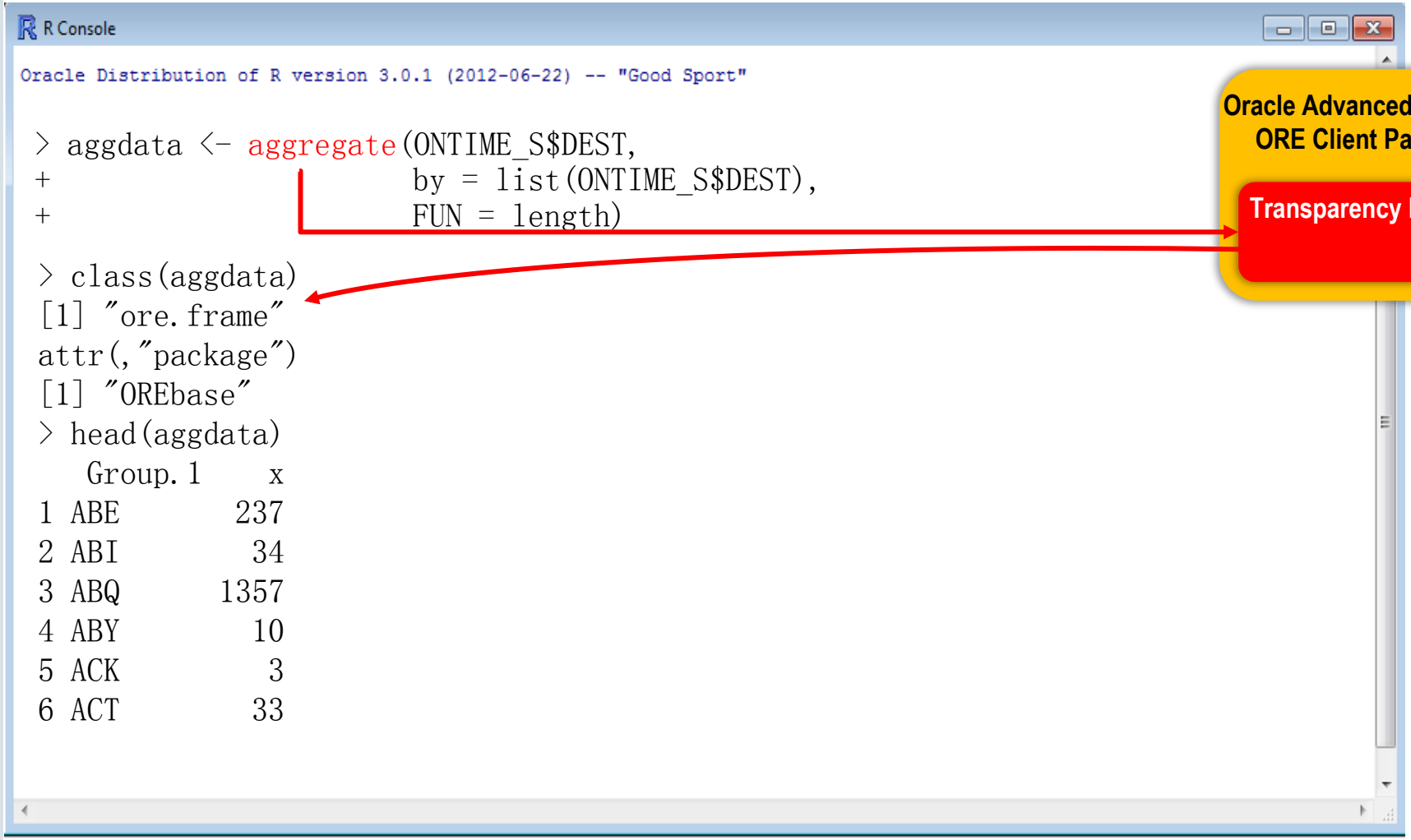
R environment

- Strengths
  - Powerful & Extensible
  - Graphical & Extensive statistics
  - Free—open source
- Challenges
  - Memory constrained
  - Single threaded
  - Outer loop—slows down process
  - Not industrial strength



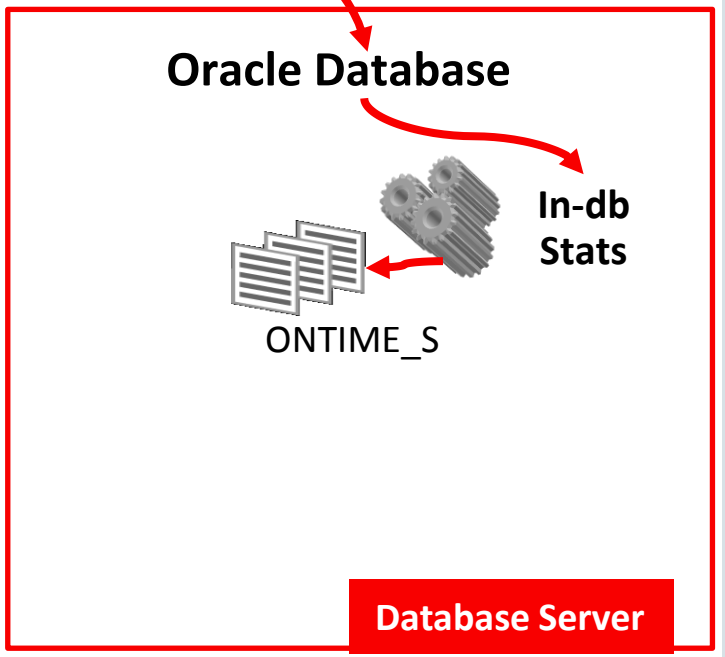
R: Transparency via function overloading

Invoke in-database aggregation function



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

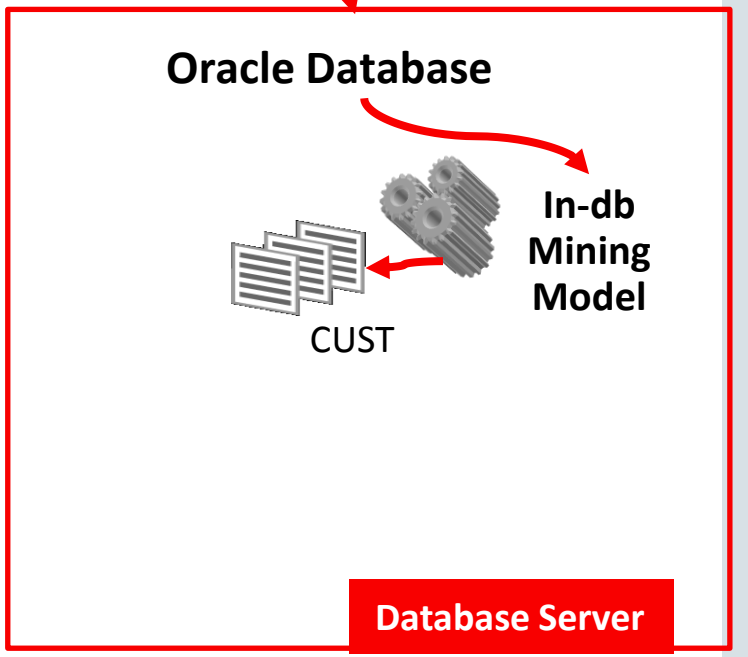
```
Coefficients:
   class       variable value      estimate
1      0          INCOME        5.204561e-05
2      0 MARITAL_STATUS      M -4.531359e-05
3      0 MARITAL_STATUS      S  4.531359e-05
4      0       YRS_CUST        1.264948e-04
5      0    (Intercept)        9.999269e-01
6      1          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
...
```



File Edit Code View Plots Session Build Debug Tools Help

Go to file/function

Project: (None)

rquser login DRAFT.R | Manual ORE Setup.R | Hornick Short Demo Script (1).R

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
12
13 hist(CARSTATS$MPG, col="red", breaks=25)
14
15 plot(CARSTATS)
16 plot(CARSTATS, col="red")
17
18
19 demo(odm_kmeans)
20
21
22
23
24 names(CLAIMS)
25 summary(CLAIMS)
```

6:9 (Top Level) R Script

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"
>
```

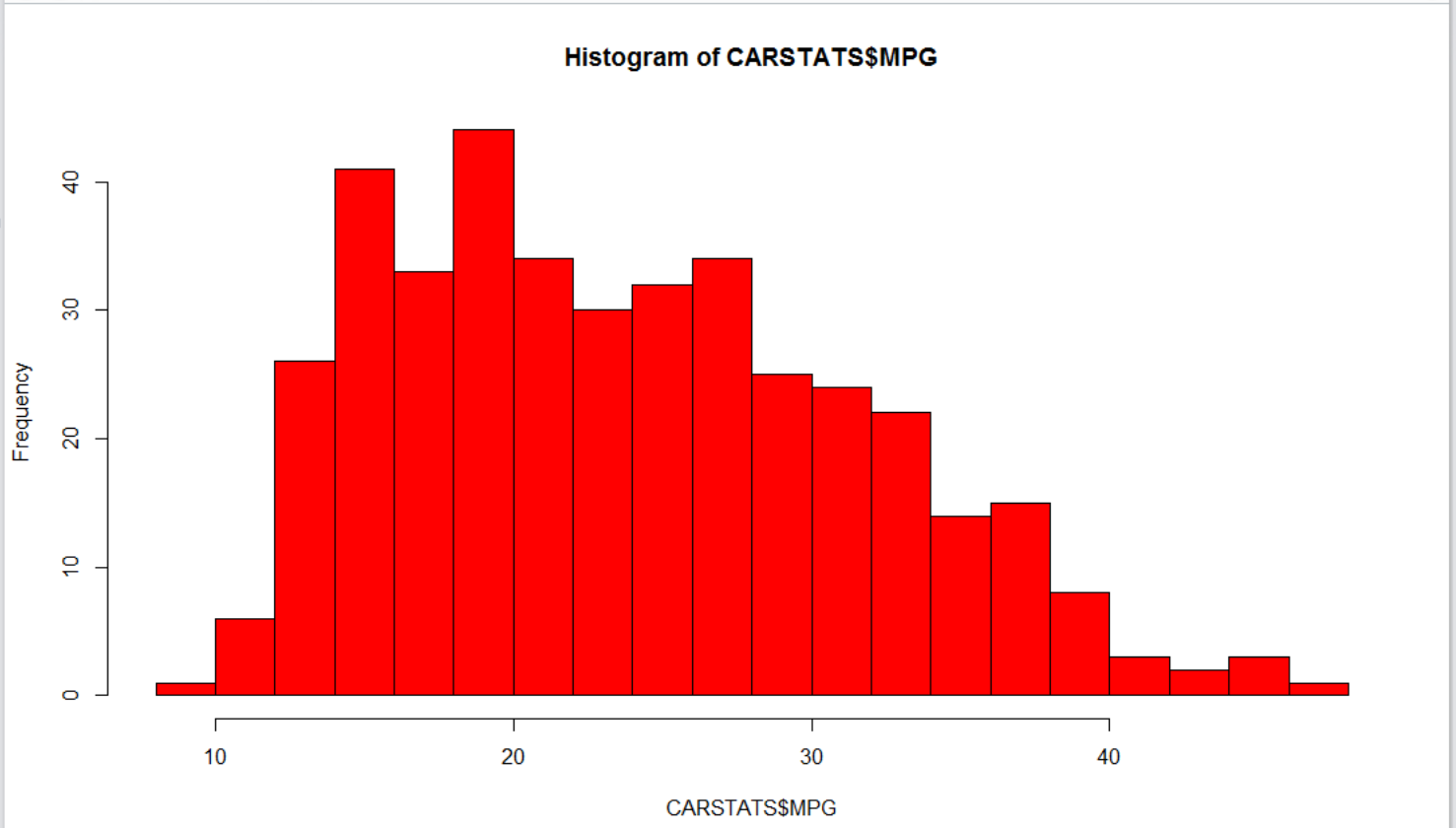
Environment | History

To Console | To Source

```
data(pistonrings)
head(pistonrings)
diameter <- qcc.groups(pistonrings$diameter, pistonrings$sample)
qcc(diameter[1:25,], type="xbar")
hist(CARSTATS$MPG, col="red", breaks=25)
plot(CARSTATS, col="red")
hist(CARSTATS$MPG, col="red", breaks=25)
.libPaths("C:\\app\\cberger\\product\\12.1.0\\dbhome_1\\R\\library")
library(ORE)
ore.connect("dmuser", "ora12c", " localhost ", "dmuser", all=TRUE)
hist(CARSTATS$MPG, col="red", breaks=25)
ore.ls()
```

Files | Plots | Packages | Help | Viewer

Zoom | Export | Clear All



File Edit Code View Plots Session Build Debug Tools Help

Project: (None)

rquser login DRAFT.R | Manual ORE Setup.R | Hornick Short Demo Script (1).R

Source on Save | Run | Source

```
.libPaths("C:\\app\\cberger\\product\\12.1.0\\dbhome_1\\R\\library")
library(ORE)

ore.connect("dmuser", "ora12c", " localhost ", "dmuser", all=TRUE)

ore.ls()
help.start()

names(CARSTATS)

summary(CARSTATS)

hist(CARSTATS$MPG, col="red", breaks=25)

plot(CARSTATS)
plot(CARSTATS, col="red")


demo(odm_kmeans)




names(CLAIMS)
summary(CLAIMS)
```

15:1 (Top Level) R Script

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
>
```

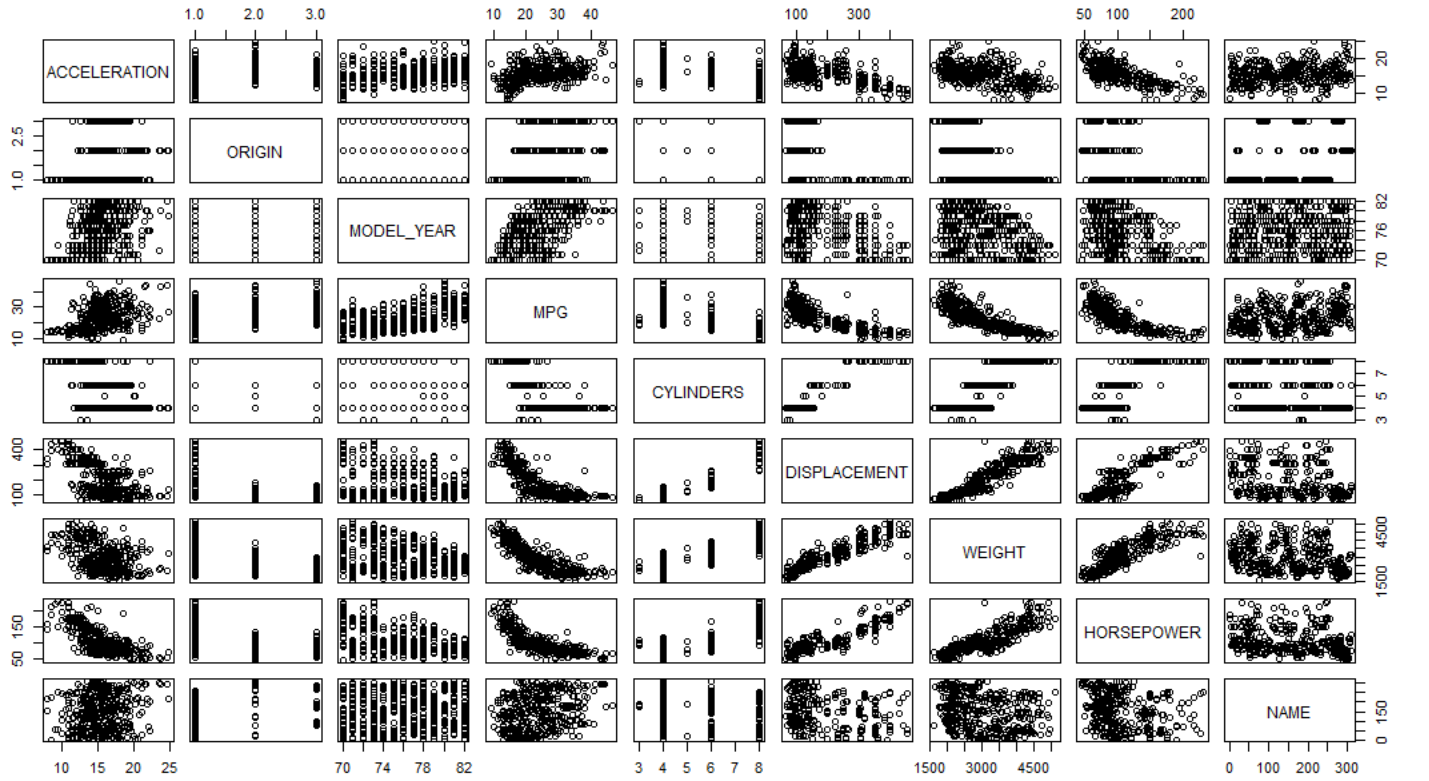
Environment | History

To Console | To Source

```
qcc(diameter[1:25,], type="xbar")
hist(CARSTATS$MPG, col="red", breaks=25)
plot(CARSTATS, col="red")
hist(CARSTATS$MPG, col="red", breaks=25)
.libPaths("C:\\app\\cberger\\product\\12.1.0\\dbhome_1\\R\\library")
library(ORE)
ore.connect("dmuser", "ora12c", " localhost ", "dmuser", all=TRUE)
hist(CARSTATS$MPG, col="red", breaks=25)
ore.ls()
help.start()
summary(CARSTATS)
plot(CARSTATS)
```

Files | Plots | Packages | Help | Viewer

Zoom | Export | Clear All



Oracle Advanced Analytics for Hadoop

Predictive algorithms that execute in a parallel/distributed manner on Hadoop with data in HDFS

ORACLE

Oracle R Advanced Analytics for Hadoop

Advanced Analytics

Using Hadoop and HIVE Integration, plus R Engine and Open-Source R Packages

Hadoop Cluster

with **Oracle R Advanced Analytics for Hadoop (ORAAH)**

R interface to HQL Basic Statistics, Data Prep, Joins and View creation — HQL

Parallel, distributed algorithms:

MLP Neural Nets*, GLM*, LM, PCA, k-Means, NMF, LMF

** Spark-Caching enabled*

Use of Open-source R packages via custom R Mappers / Reducers

R Client

R Analytics
Oracle R Advanced Analytics for Hadoop

Oracle Database
with Advanced Analytics option

SQL

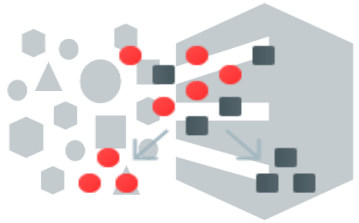
SQL Client

SQL Developer
Other SQL Apps

ORACLE

Oracle R Advanced Analytics for Hadoop

AA Algorithms in a Hadoop Cluster: Map-Reduce and Spark (2.7)



Classification

- GLM ORAAH (hadoop MapReduce)
- Logistic Regression ORAAH (Spark)
- Random Forests (Spark MLlib)
- Decision Trees (Spark MLlib)
- Support Vector Machines (Spark MLlib)
- Gaussian Mixture Models

Clustering

- K-Means (hadoop MapReduce)
- K-Means (Spark MLlib)

Regression

- MLP Neural Networks ORAAH (Spark)
- LASSO (Spark MLlib)
- Ridge Regression (Spark MLlib)
- Support Vector Machines (Spark MLlib)
- Random Forest (Spark MLlib)
- Linear Regression (Spark MLlib)

Basic Statistics

- Correlation/Covariance (hadoop MapReduce)

Feature Extraction

- Non-negative Matrix Factorization (hadoop MapReduce)
- Collaborative Filtering (LMF) (hadoop MapReduce)

Attribute Importance

- Principal Components Analysis (hadoop MapReduce)
- Principal Components Analysis (Spark MLlib)
- LASSO (Spark MLlib)

ORACLE

Oracle’s Advanced Analytics

Predictive Applications + OBIEE Integration

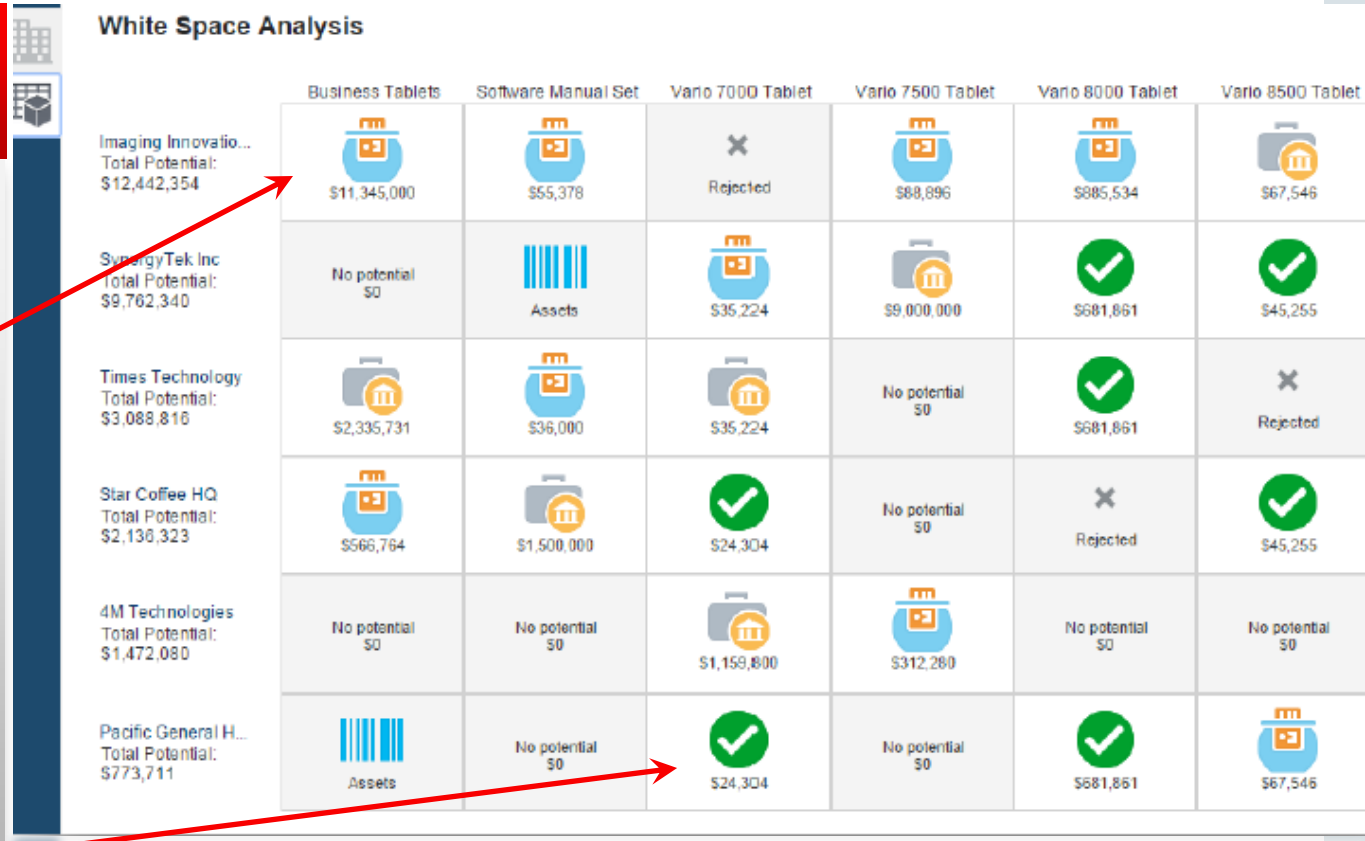
ORACLE

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.



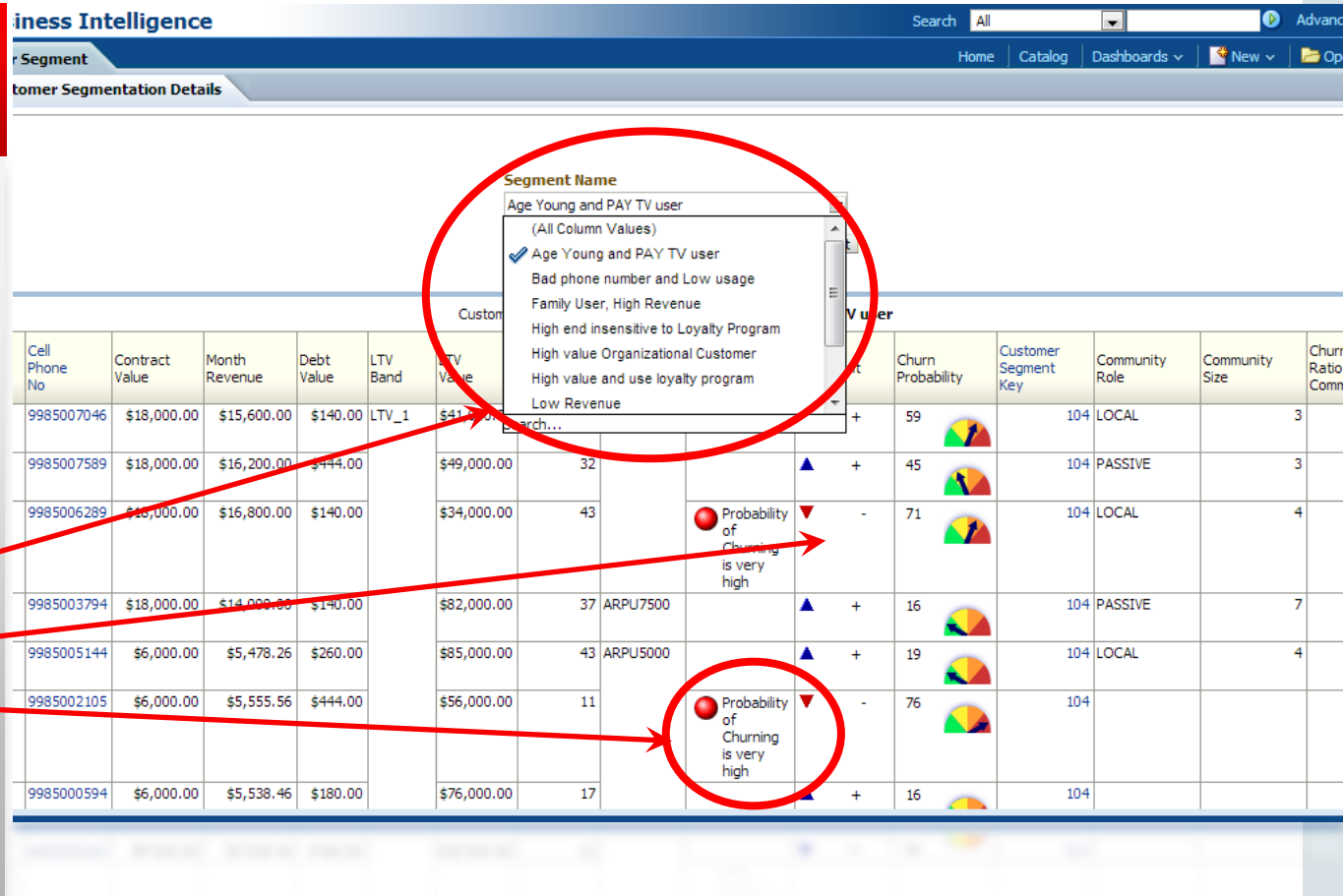
[Link to Oracle CRM SPE on O.com]

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

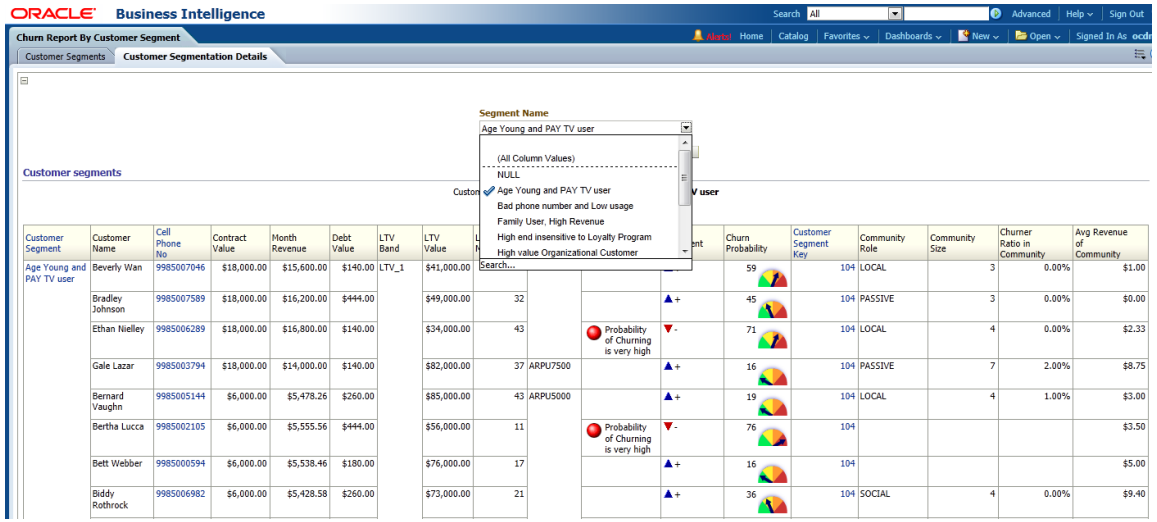


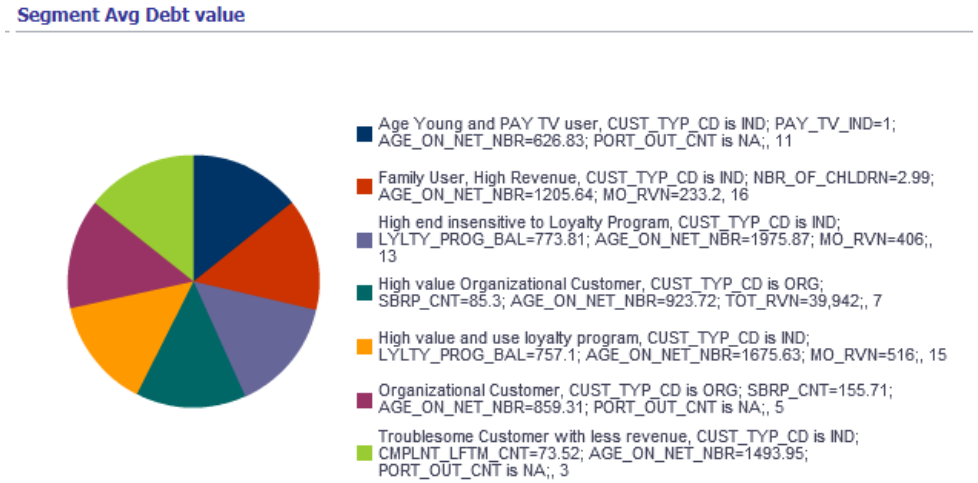
[Link to OCDM on OTN]

Oracle Communications Data Model

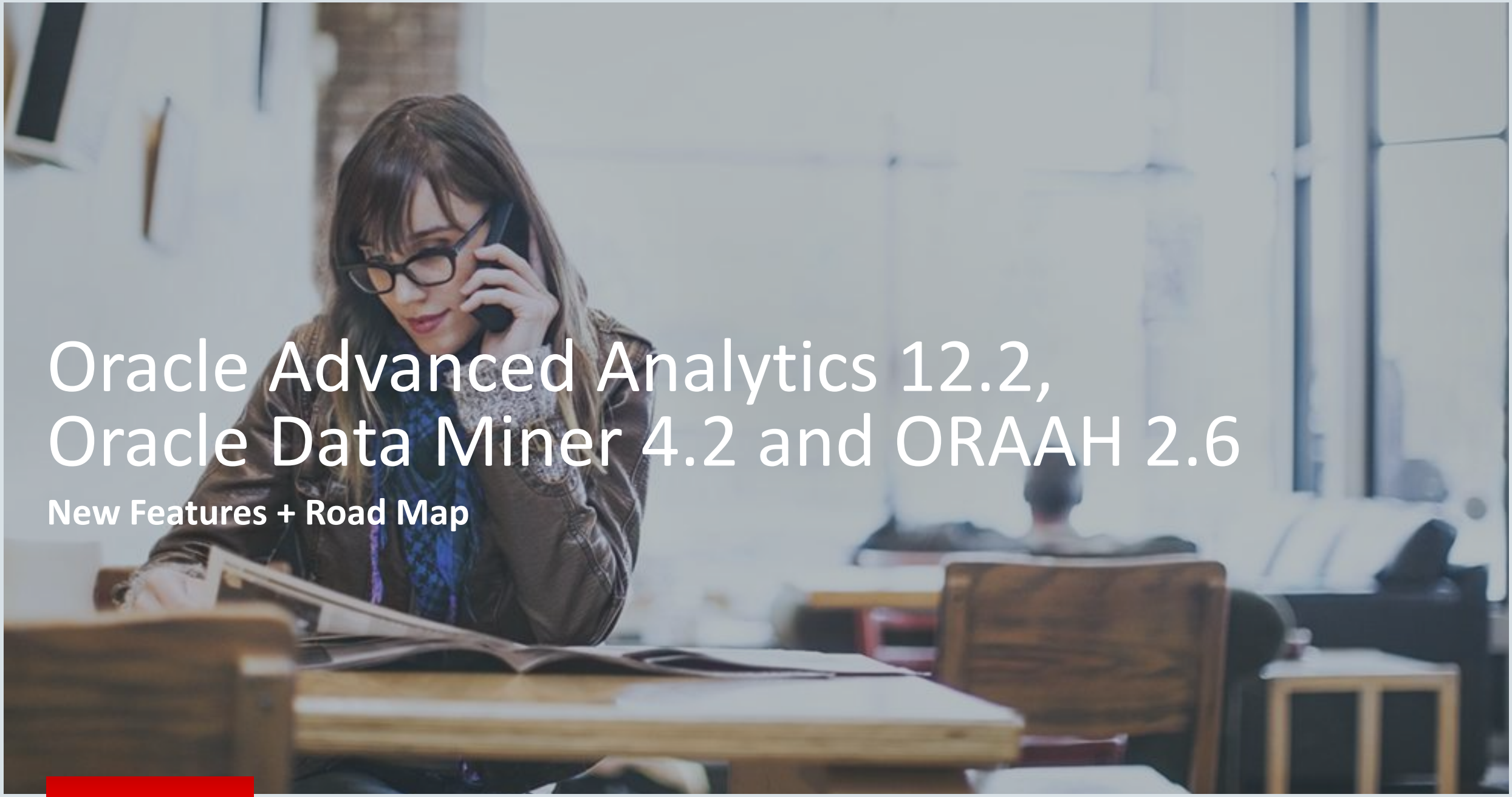
Pre-Built Data Mining Models

1. Churn Prediction
2. Customer Profiling
3. Customer Churn Factor
4. Cross-Sell Opportunity
5. Customer Life Time Value
6. Customer Sentiment
7. Customer Life Time Value





Link to OCDM on OTN



Oracle Advanced Analytics 12.2, Oracle Data Miner 4.2 and ORAAH 2.6

New Features + Road Map

ORACLE

Oracle Advanced Analytics 12.2

New Oracle Database Features

- **Major Performance Improvements for Algorithms**
  - New parallel model build / apply infrastructure to enable faster algorithm introduction
  - Scale to larger data volumes found in big data and cloud use cases
- **Unsupervised Feature Selection**
  - Uses unsupervised pair-wise Kullback-Leibler Divergence (KLD - correlations analysis - numeric and categorical attributes) to find highest “information containing” attributes
- **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
- **Partitioned Models**
  - Instead of building, naming and referencing 10s or 1000s of models, a partitioned model will organize and represent these multiple models as partitions in a single model entity

ORACLE

Oracle Advanced Analytics 12.2

New Oracle Database Features

NEW IN 12.2

- **Explicit Semantic Analysis (ESA) algorithm**
  - Useful technique for extracting meaningful, interpretable features, better than LDA
  - Use ESA for document similarity and topic identification
  - Wikipedia provides a large corpus of existing documents to provide sensible features and topics
  - *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 0.62

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

NEW IN 4.2

Oracle Data Miner

4.2 New Features

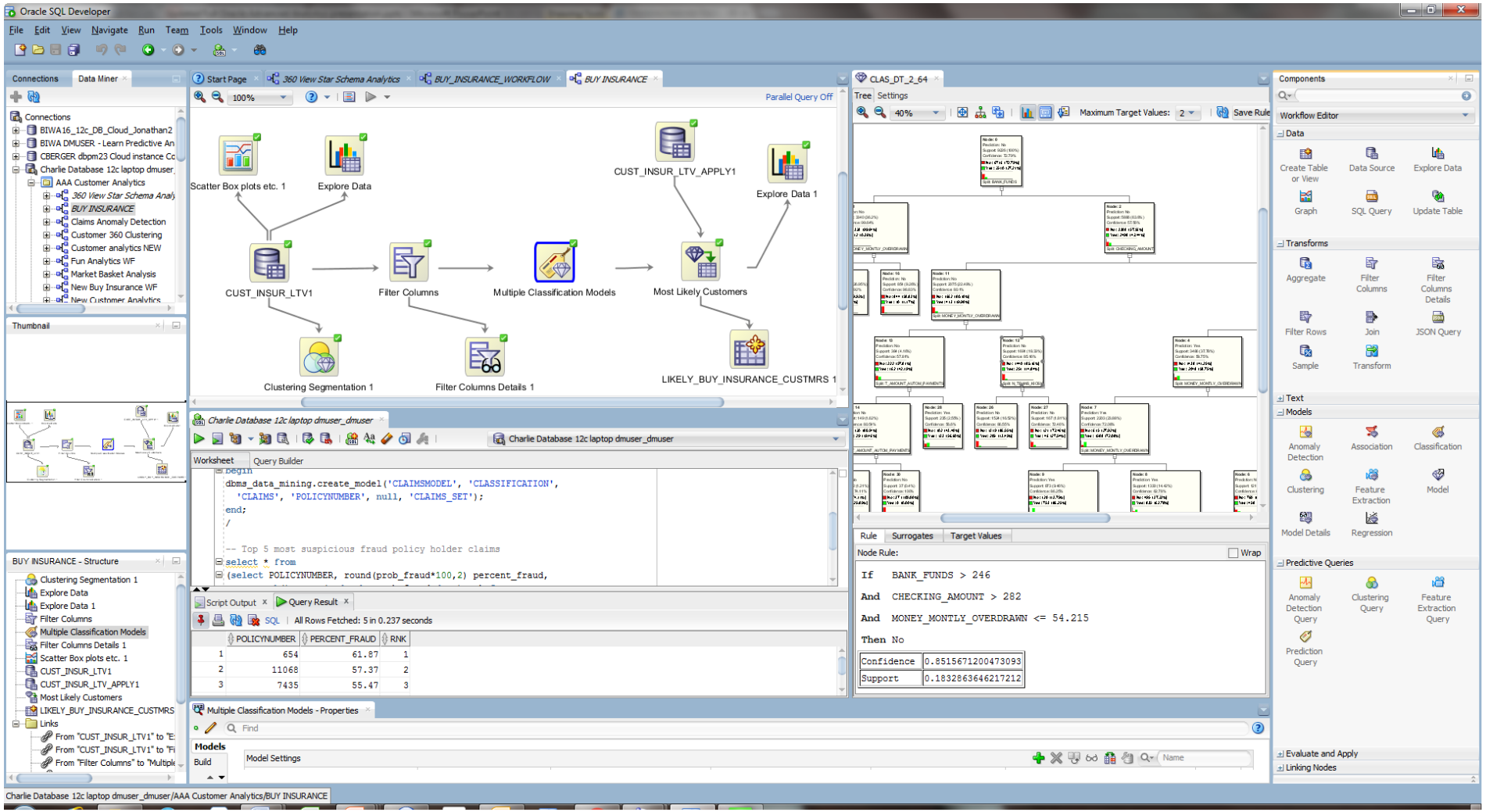
ORACLE

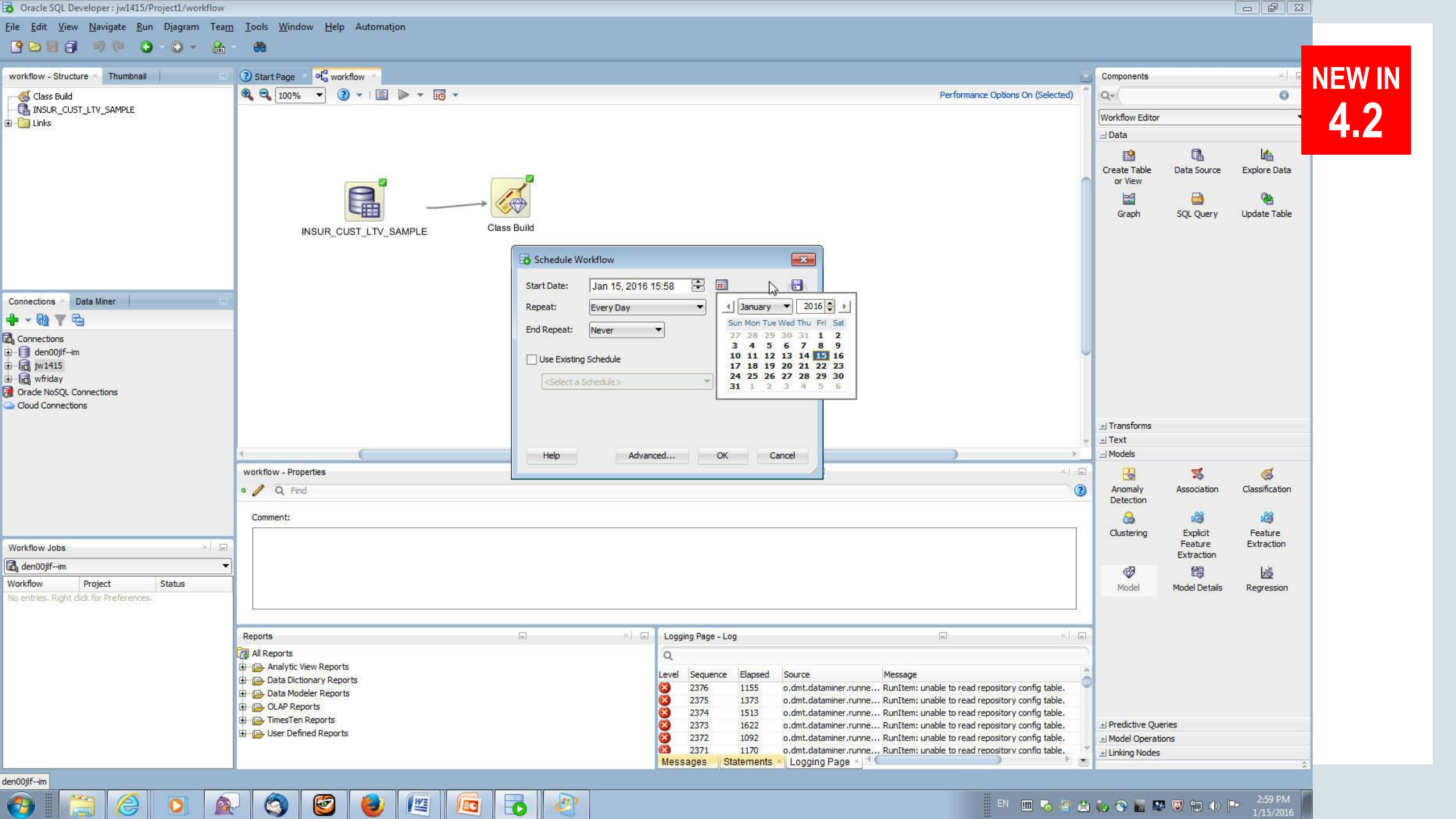
Oracle Data Miner 4.2

New Features for OAA

NEW IN 4.2

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




Oracle SQL Developer : jw1415/Project1/workflow
File Edit View Navigate Run Diagram Team Tools Window Help Automation
workflow - Structure
Thumbnail
Class Build
INSUR_CUST_LTV_SAMPLE
Links
Start Page
workflow
100%
Performance Options On (Selected)
INSUR_CUST_LTV_SAMPLE
Class Build
Schedule Workflow
Start Date: Jan 15, 2016 15:58
Repeat: Every Day
End Repeat: Never
Use Existing Schedule
<Select a Schedule>
January 2016
Sun Mon Tue Wed Thu Fri Sat
Help
Advanced...
OK
Cancel
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
Connections
Data Miner
Connections
den00jlf--im
jw1415
wfriday
Oracle NoSQL Connections
Cloud Connections
Workflow Jobs
den00jlf--im
Workflow
Project
Status
No entries. Right click for Preferences.
workflow - Properties
Find
Comment:
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
den00jlf--im
EN
2:59 PM
1/15/2016
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

Getting started

Getting started: OAA Links and Resources

Oracle Advanced Analytics Overview:

- OAA presentation— Big Data Analytics with Oracle Advanced Analytics …or just watch Watch YouTube video presentation and demo(s)
- Big Data Analytics with Oracle Advanced Analytics: Making Big Data and Analytics Simple white paper on OTN
- Oracle Internal OAA Product Management Wiki and Workspace
- Oracle Advanced Analytics Customer Successes

YouTube recorded OAA Presentations and Demos:

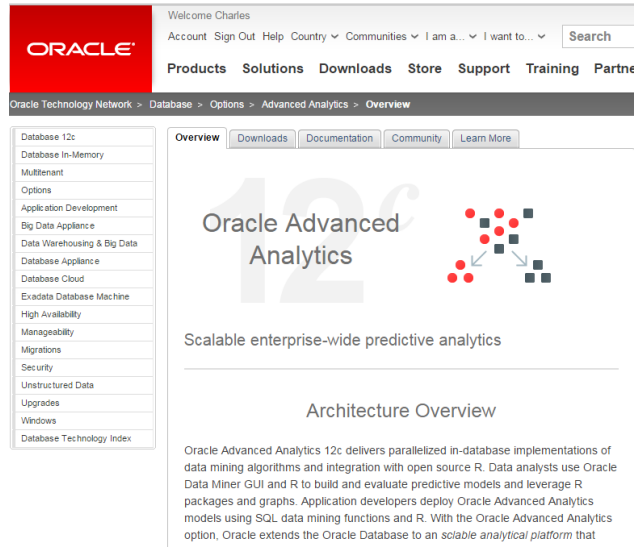
- Oracle Advanced Analytics and Data Mining at the YouTube Movies (6 + OAA "live" Demos on ODM'r 4.0 New Features, Retail, Fraud, Loyalty, Overview, etc.)

Getting Started:

- 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 Free Oracle Advanced Analytics **"Test Drives" on Oracle Cloud** via Vlamis Partner
- 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

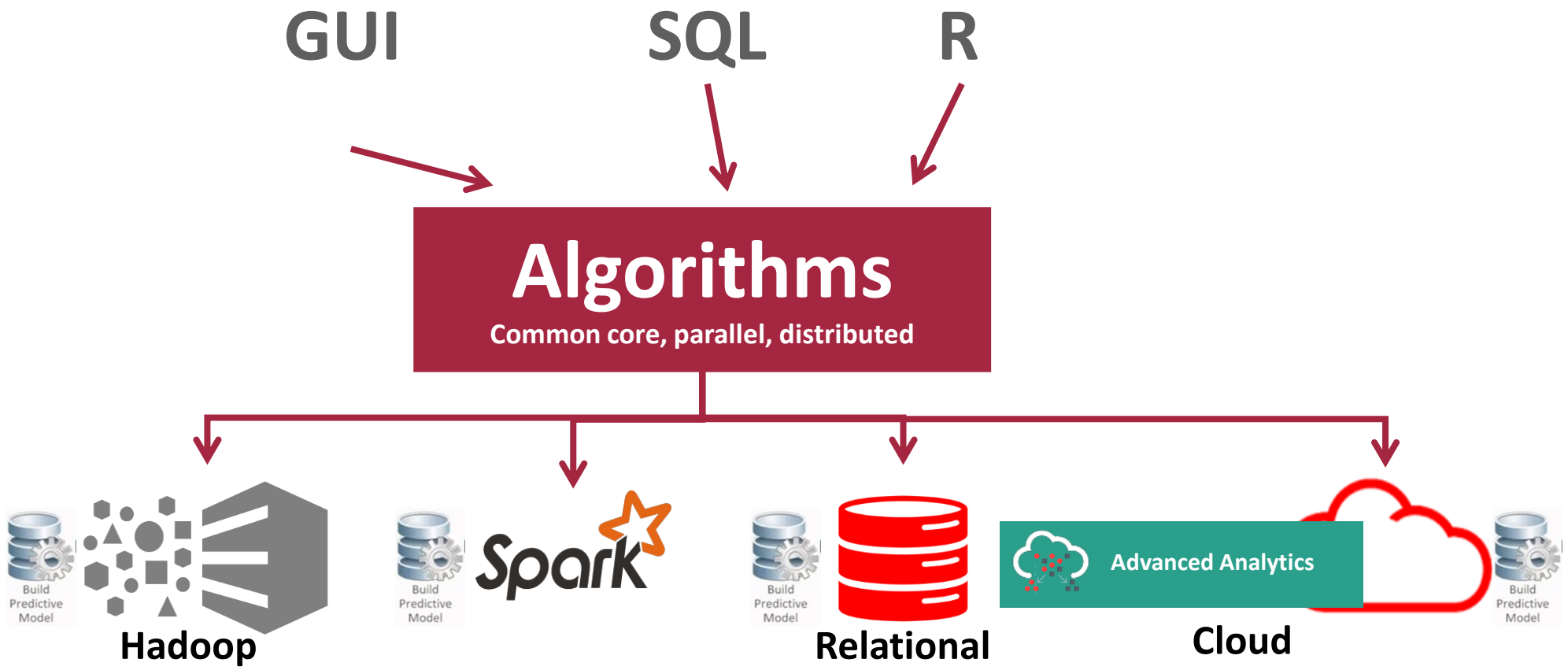
Additional Resources:

- 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
- **Business Intelligence, Warehousing & Analytics—BIWA Summit'17**, Jan 31, Feb 1 & 2, 2017 at Oracle HQ Conference Center (w/ links to customer presentations)



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

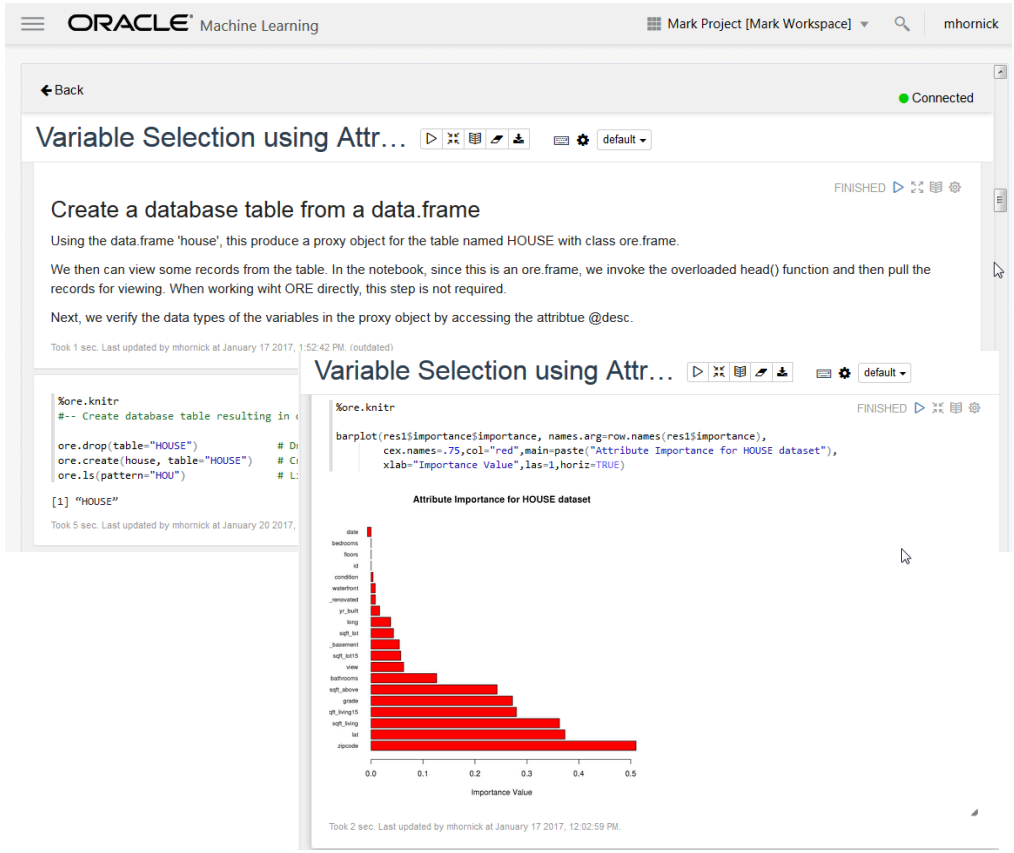


Oracle Machine Learning

Multi-Platform, Multi-Engine Oracle Machine Learning for the Cloud

Introduction

- A collaborative unified notebook user interface and cloud platform for Oracle's machine learning capabilities
- Enables teams to…
  - Leverage the languages, platforms and engines of their choice
  - Explore and prepare data
  - Build, access, and use machine learning models
  - Deploy machine learning solutions

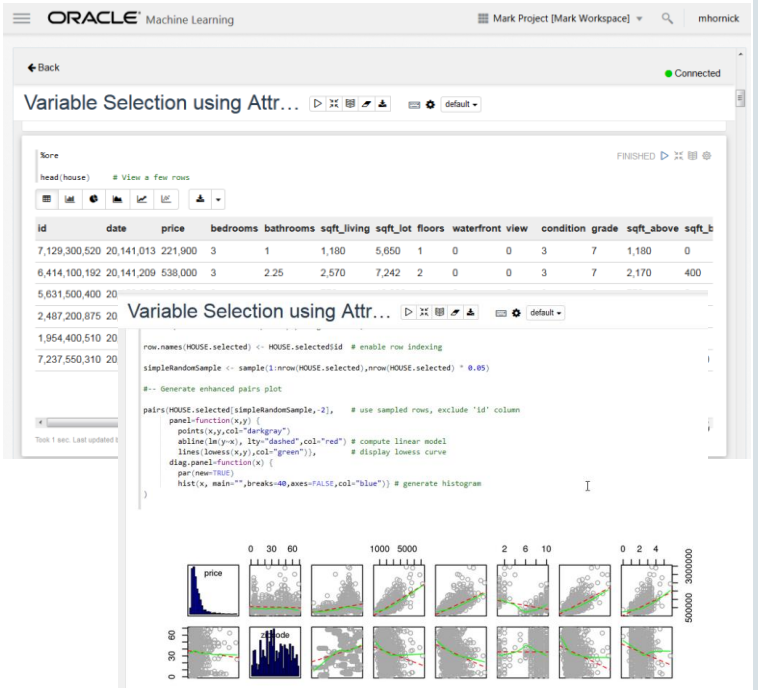


Oracle Machine Learning

Multi-Platform, Multi-Engine Oracle Machine Learning for the Cloud

Features

- Zeppelin-based Unifying Notebook UI
  - Multiple backend servers, analytical engines and languages
  - Oracle and open source algorithms, platforms, tools, and data sources (SQL, R, Spark, Python)
  - Provides a flexible toolbox for the data scientist
- Supports collaborative development
  - Shared notebooks and templates, with access permissions
- Supports deployment of predictive analytics solutions
  - Enables publishing libraries, templates, examples of common use cases
- Dynamic and expandable platform



Oracle's Advanced Analytics and Machine Learning Platform

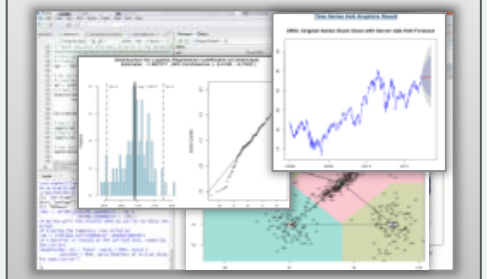
Multiple interfaces across platforms — SQL, R, GUI, Dashboards, Apps

Information Producers

Users

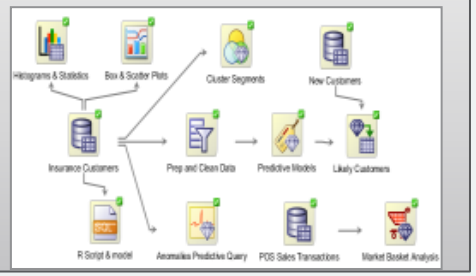
R programmers

R Client



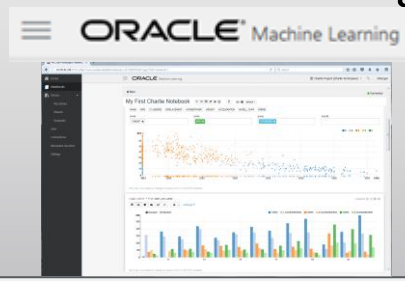
Data & Business Analysts

SQLDEV/
Oracle Data Miner



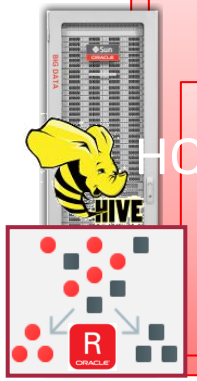
Data Scientists (R, SQL, Python, etc.)

Oracle
Machine Learning



Open Source MLLib, RServ, Spark,, etc.

Platform



HQL

Hadoop

ORAAH
Parallel, distributed algorithms

Oracle Database Enterprise Edition

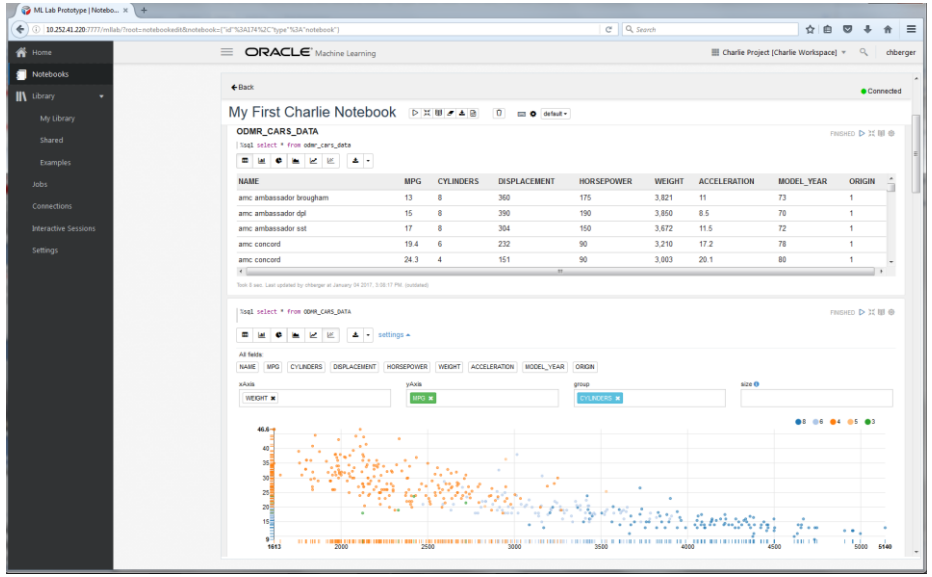
Oracle Advanced Analytics - Database Option
SQL Data Mining, ML & Analytic Functions + R Integration for Scalable, Distributed, Parallel in-DB ML Execution

Oracle Database 12c

Oracle Cloud

ORACLE

Oracle Machine Learning



Beta Program

- Seeking knowledgeable, articulate data scientists/data analysts users for early product access, beta testing and providing valuable customer feedback
- Oracle Machine Learning will be hosting as “i-Betas” in the Oracle Cloud for easier customer access and testing
- OML Beta Program 2017

OML Contacts in Product Management

- Charlie Berger – charlie.berger@oracle.com
- Mark Hornick – mark.hornick@oracle.com
- Marcos Arancibia –marcos.arancibia@oracle.com


ORACLE®