



Zwiększanie wartości analiz biznesowych dzięki uczeniu maszynowemu (ML)

Michał Grochowski
Konsultant BI/DWH





Alan Turing, 1950: “Can Machines Think?”

Today....

*Can A Machine Think
Like a Human Brain?*

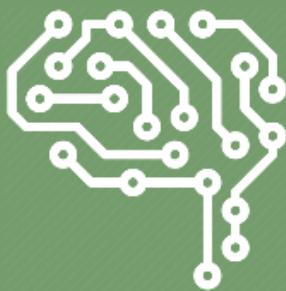
DATA SCIENCE

Teams, tools and techniques to use and deploy artificial intelligence

Machine learning uses algorithms to parse data, learn from that data, and make informed decisions based on what it has learned

ARTIFICIAL INTELLIGENCE

Any technique which enables computers to mimic human behavior



1950's

1960's

1970's

1980's

1990's

2000's

2010s

MACHINE LEARNING

AI techniques that give computers the ability to learn without being explicitly programmed to do so



DEEP LEARNING

A subset of ML which make the computation of multi-layer neural networks feasible



Oracle Machine Learning Algorithms and Analytics

CLASSIFICATION

Naïve Bayes
Logistic Regression (GLM)
Decision Tree
Random Forest
Neural Network
Support Vector Machine (SVM)
Explicit Semantic Analysis

CLUSTERING

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

ANOMALY DETECTION

One-Class SVM

TIME SERIES

Forecasting - Exponential Smoothing
Includes popular models
e.g. Holt-Winters with trends,
seasonality, irregularity, missing data

REGRESSION

Linear Model
Generalized Linear Model (GLM)
Support Vector Machine (SVM)
Stepwise Linear regression
Neural Network
LASSO

ATTRIBUTE IMPORTANCE

Minimum Description Length
Principal Component Analysis (PCA)
Unsupervised Pair-wise KL Div
CUR decomposition for row & AI

ASSOCIATION RULES

A priori/ market basket

PREDICTIVE QUERIES

Predict, cluster, detect, features

SQL ANALYTICS

SQL Windows
SQL Patterns
SQL Aggregates

FEATURE EXTRACTION

Principal Comp Analysis (PCA)
Non-negative Matrix Factorization
Singular Value Decomposition (SVD)
Explicit Semantic Analysis (ESA)

TEXT MINING SUPPORT

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

STATISTICAL FUNCTIONS

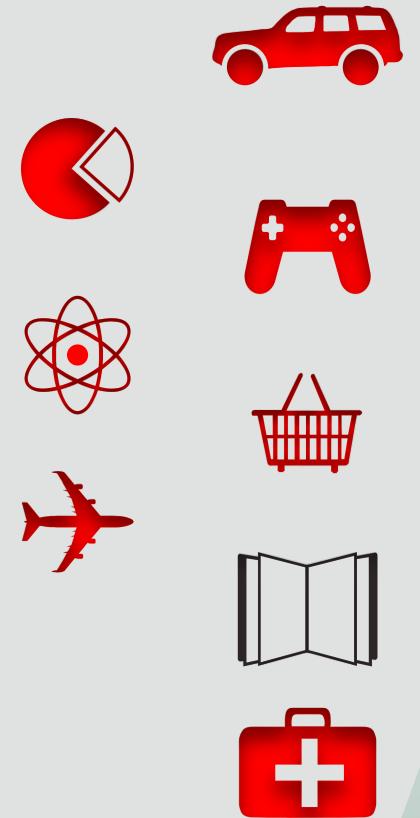
Basic statistics: min, max,
median, stdev, t-test, F-test,
Pearson's, Chi-Sq, ANOVA, etc.

R AND PYTHON PACKAGES

Third-party R and Python Packages
through Embedded Execution
Spark MLlib algorithm integration

Oracle ML Sample Use cases

- Detect fraud in customer transactions, insurance claims
- Identify which patients are at risk of developing certain conditions
- Target the right customer with the right offer
- Discover hidden customer segments
- Forecast customer demand for a product or service
- Find most profitable selling opportunities
- Anticipate and preventing customer churn
- Identify customers likely to churn and why
- Security and suspicious activity detection
- Understand sentiments in customer conversations
- Understand influencers in social networks
- Predict credit risk



Przykłady ML



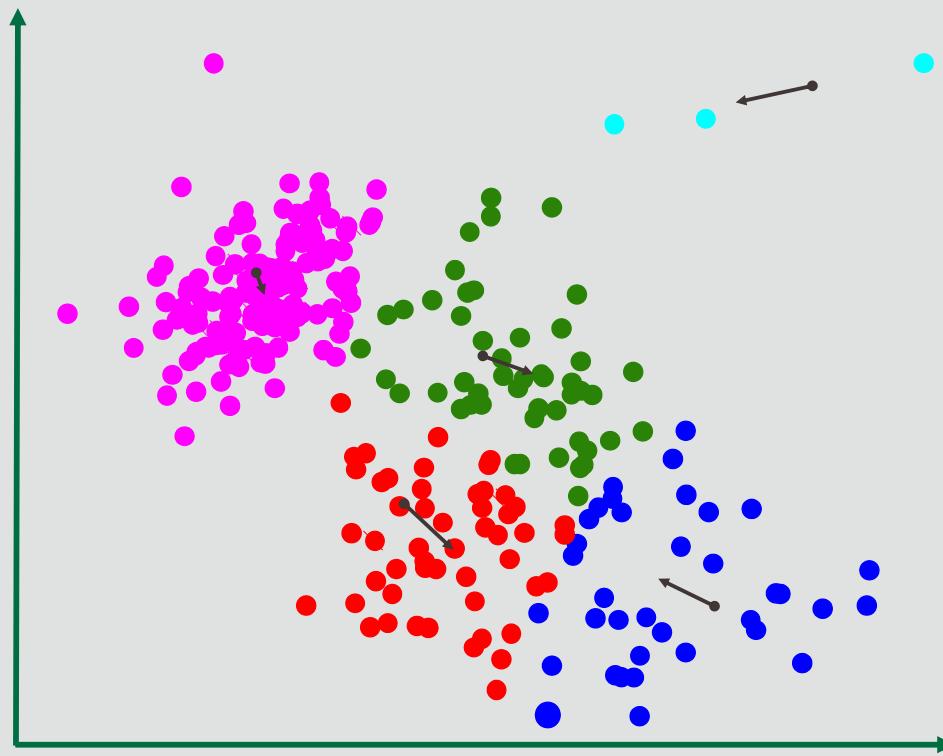
Customer Segmentation

Machine Learning - Clustering



Customer Segmentation

Machine Learning - Clustering



Predictive Maintenance

Machine Learning - Neural Networks

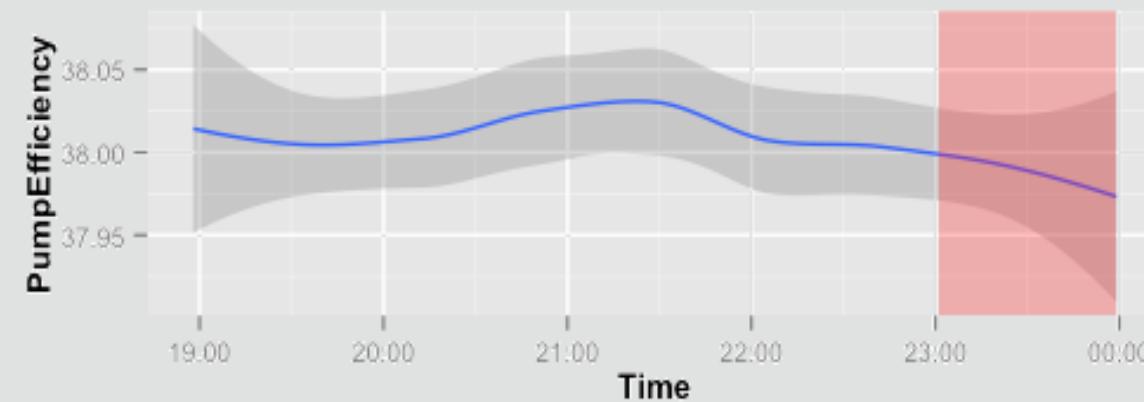
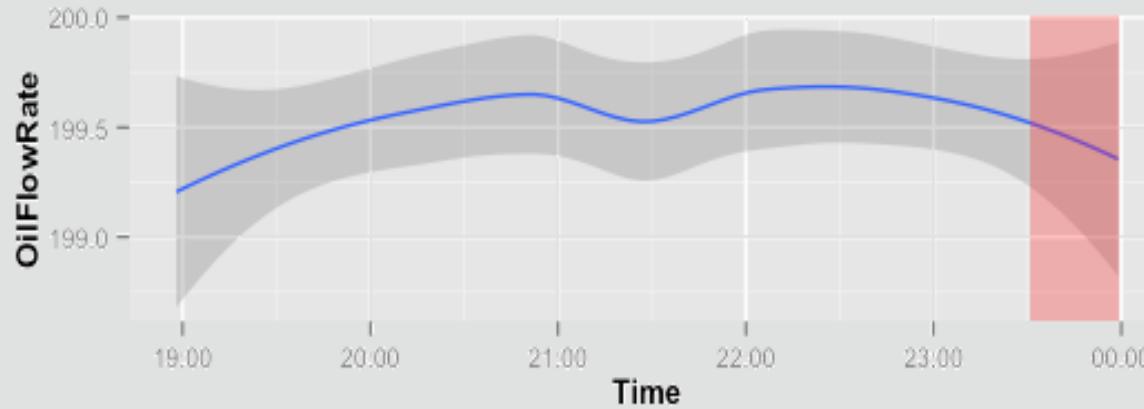
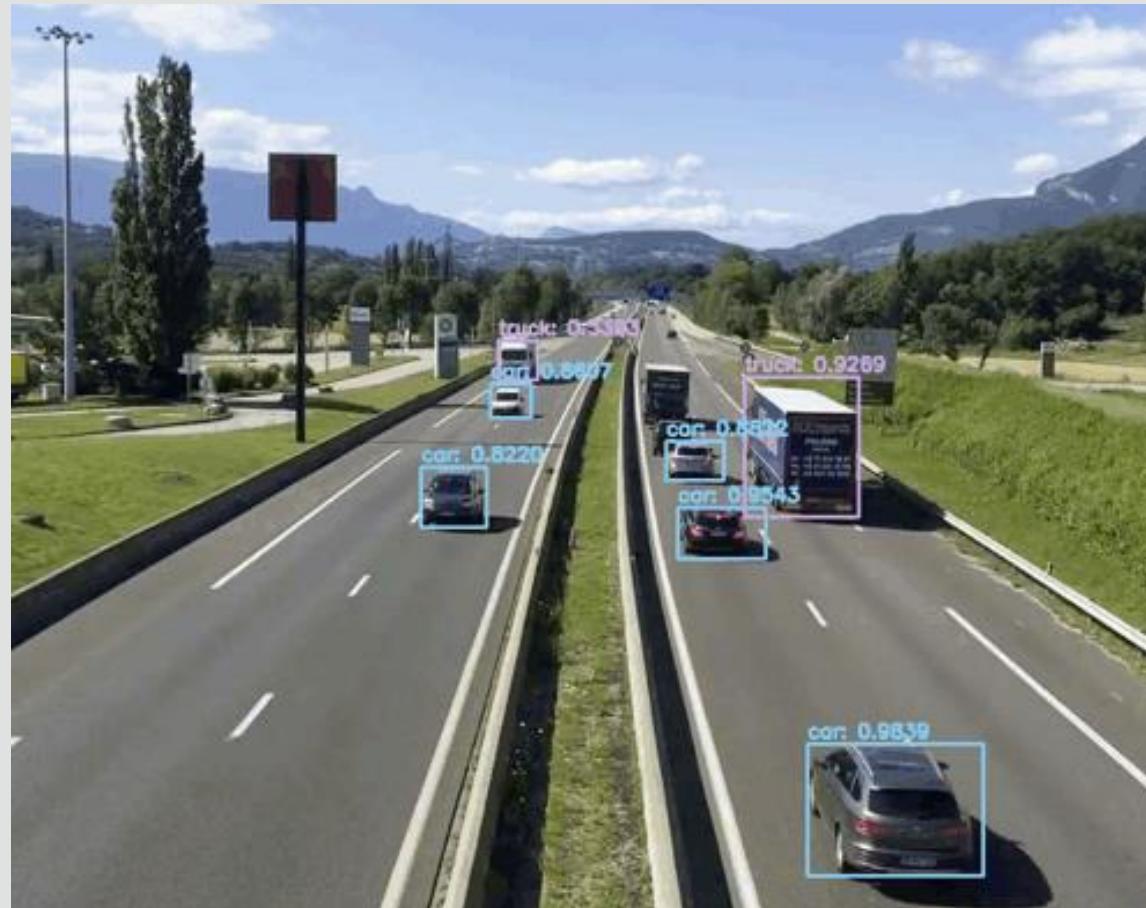
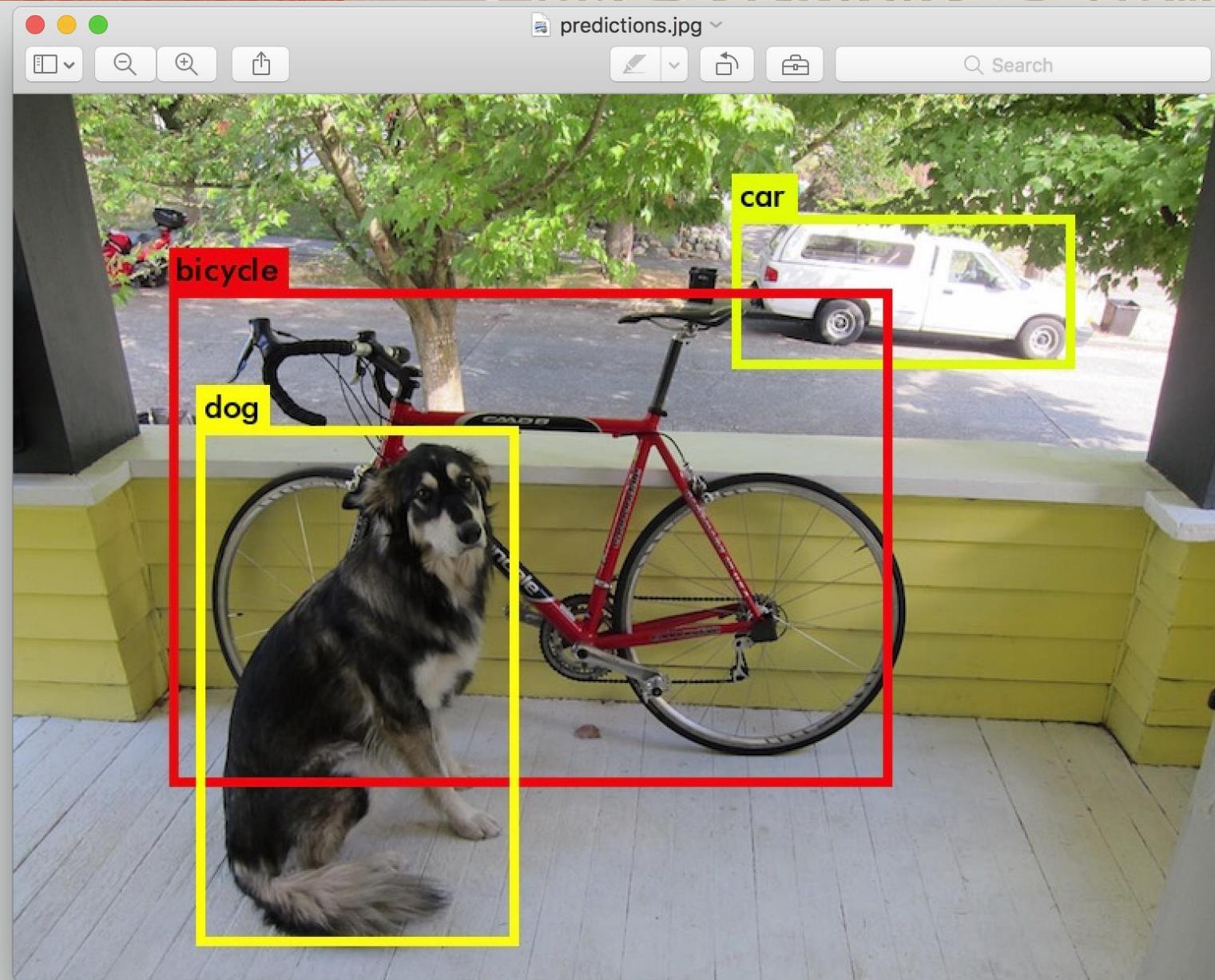
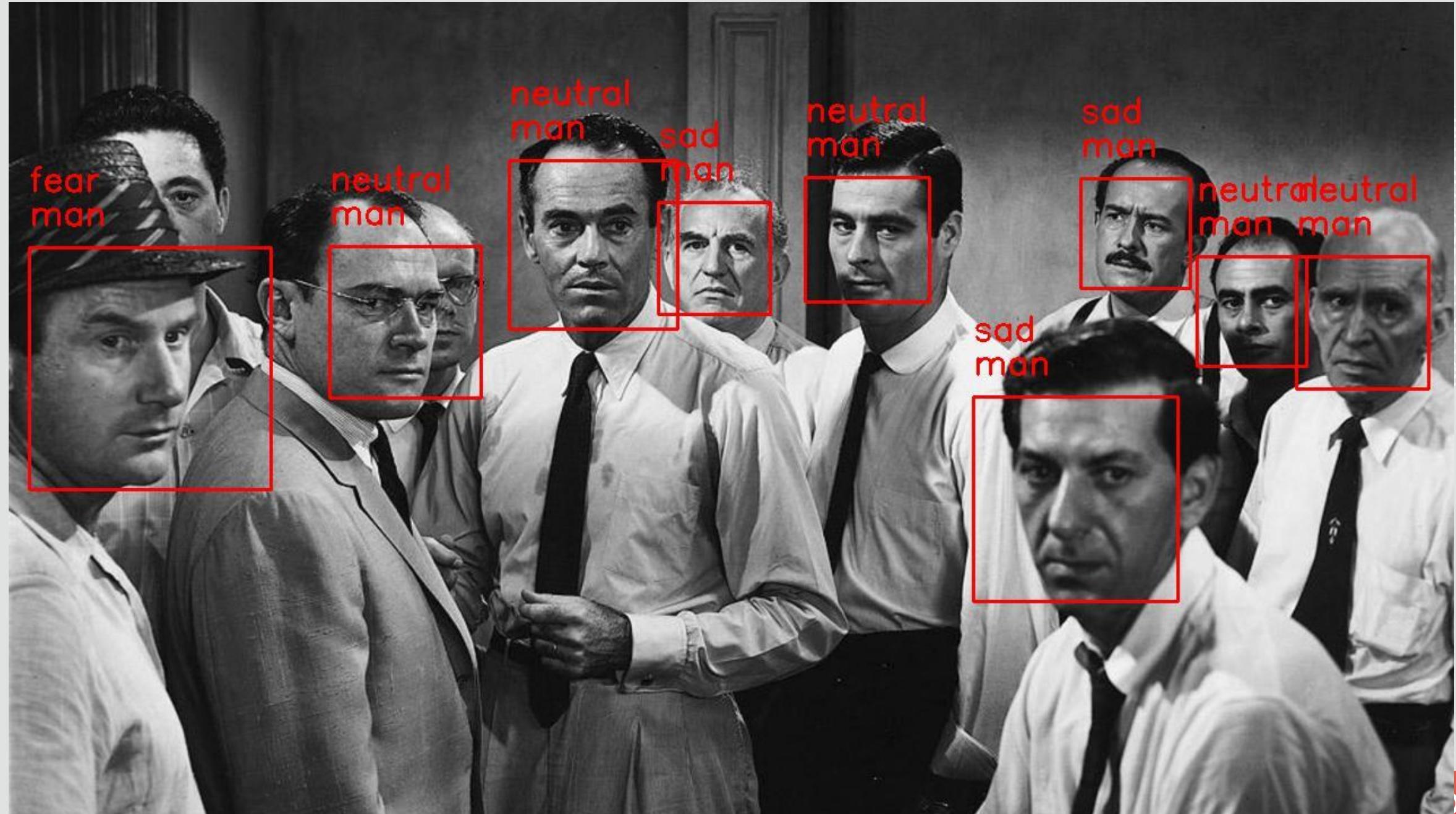


Image recognition & classification

Deep learning







13

New Open Select Stream

angry disgusted fearful happy sad neutral person: 99.63%

et est acce acca acca acca acca di di no

Analytics Cloud Details x Analiza emocji x +

https://mgoac-oraseemeaceeociworkshop.analytics.ocp.oraclecloud.com 80% ... Szukaj

Czesto odwiedzane Anuluj

ORACLE Analytics

Analiza emocji

Przygotowanie Wizualizacja Narracja Zapisz

EMOTION_TIMES Od 24.09.2019 00:00:00

ilosc wg EMOTION_NAME

EMOTION_NAME	ilosc
angry	20
disgusted	5
fearful	100
happy	80
neutral	85
sad	10

EMOTION_PROB wg EMOTION_TIMES (Godzina), EMOTION_TIMES (Minuta), EMOTION_NAME

EMOTION_NAME	09.24.2019 08 AM	09.24.2019 8:50 AM
angry	~1.5	~1.5
disgusted	~1.5	~1.5
fearful	~1.5	~1.5
happy	~1.5	~1.5
neutral	~1.5	~1.5
sad	~1.5	~1.5

7 Słupki Kanwa 1 Kanwa 2 +

ORACLE

Copyright © 2018, Oracle and/or its affiliates. All rights reserved.

Jaki interfejs wykorzystać?

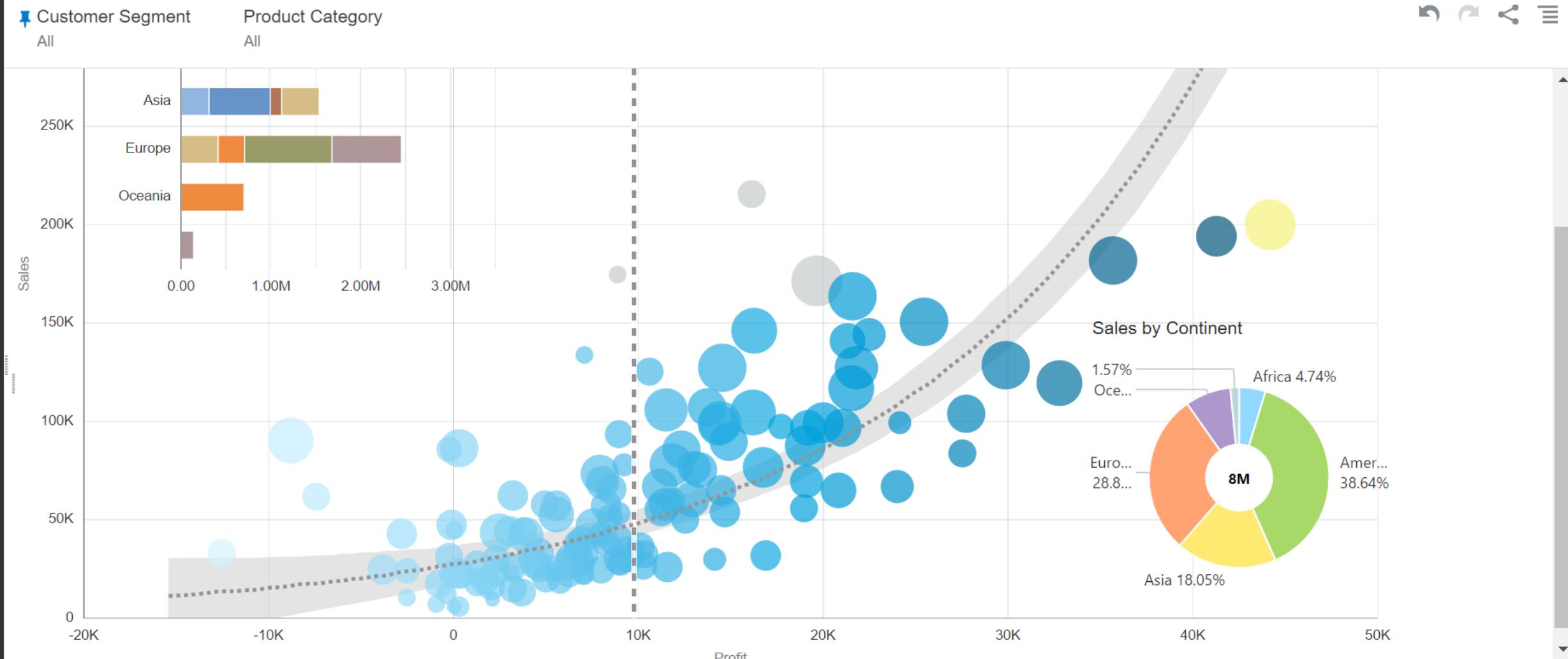
Sample Project - Project

Prepare

Visualize

Narrate

Auto Save



Untitled - Project

Prepare

Visualize

Narrate

Save

Data

- # of Orders
- A Order Priority
- A Customer ID
- # of Customers
- A Customer Name
- A Customer Segment
- A City**
- A Product Category
- A Product Sub Category
- A Grouped Sub Category
- A Product Container
- A Product Name

Discount, Sales by City

General

Title Auto

Legend Auto

Canvas 1

Click here or drag data to add a filter

Scatter

Trellis Columns

Trellis Rows

Values (Y-Axis)

Sales

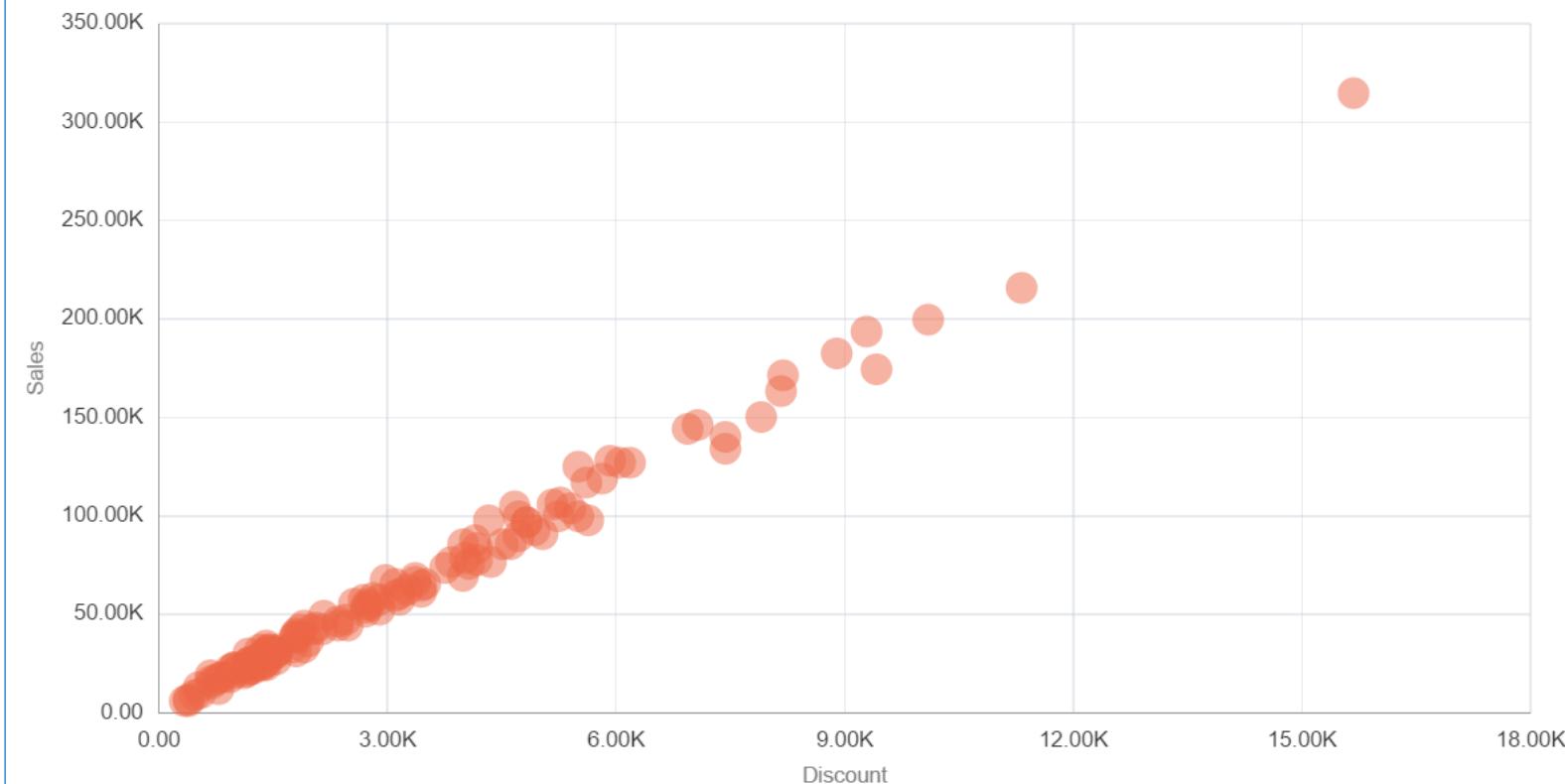
Values (X-Axis)

Discount

Category (Points)

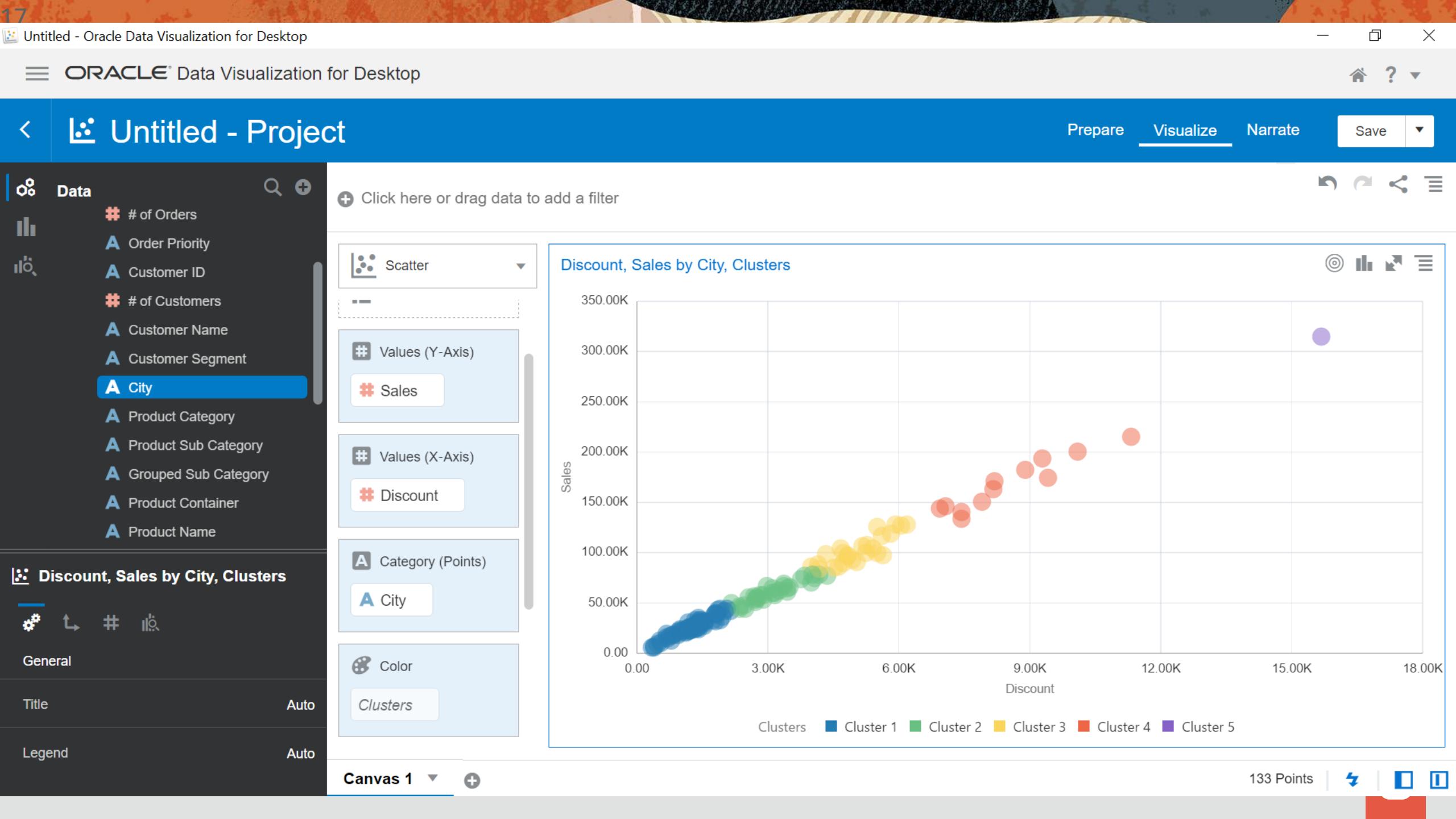
City

Discount, Sales by City



133 Points





Naive Bayes Apply Model - Attrition Prediction - Oracle Data Visualization for Desktop

ORACLE Data Visualization for Desktop

Create Connection

Select Connection Type

Search

Run Data Flow Save

100%

Data Flow Steps

- Add Data
- Join
- Union Rows
- Filter
- Aggregate
- Save Data Set
- Create Essbase Cube
- Add Columns
- Select Columns
- Rename Columns
- Transform Column
- Merge Columns

Oracle Applications

Oracle Autonomous Data Warehouse Cloud

Oracle Autonomous Transaction Processing

Oracle Big Data Cloud

Oracle Database

Oracle Content and Experience Cloud

Oracle Eloqua

Oracle Essbase

Oracle Service Cloud

Oracle Talent Acquisition Cloud

Oracle Storage Cloud Service

Actian Ingres

Actian Matrix

Actian Vector

Amazon Aurora

Cancel

Copyright © 2019 Oracle and/or its affiliates.

» Naive Bayes Apply Model - Attrition Prediction - Data Flow

Run Data Flow

Save

Data Flow Steps



Add Data

Join

Merge Rows

Filter

Aggregate

Save Data Set

Create Essbase Cube

Add Columns

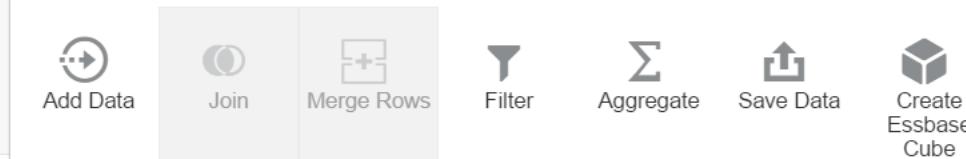
Select Columns

Rename Columns

Transform Column

Merge Columns

Attrition P.X



Add Data - Attrition



Data Set

Attrition

Description

Upload

When Run

Process



ab ID	ab First Name	ab Last Name	ab Position	ab Department	ab Distance
1001	Benjamin	Piper	52	Travel_Rarely	258
1002	Felicity	Russell	37	Travel_Rarely	1462

Show labels

100%



elections (37)



21 Naive Bayes Apply - Attrition Analysis - Oracle Data Visualization for Desktop

ORACLE Data Visualization for Desktop

Naive Bayes Apply - Attrition Analysis - Project

Prepare Visualize Narrate Auto Save

Predicted Attrition
Yes

EmployeeCount
41.00

EmployeeNumber	First Name	Last Name	Predicted Attrition	PredictionConfidence
1427	Brandon	Mackenzie	Yes	0.63
1433	Diana	Dickens	Yes	0.57
1439	Julia	Paterson	Yes	0.76
1458	Dan	Coleman	Yes	0.56
1487	Austin	Ross	Yes	0.85

EmployeeCount by JobRole, Department

Department	Count
Human Resources	2.44%
Sales	21.95%
Research & Development	17.07%
Other	21.95%

EmployeeCount by YearsInCurrentRole

EmployeeCount by JobSatisfaction

EmployeeCount by JobRole, OverTime

JobRole	OverTime	EmployeeCount
Healthcare Repre...	No	~2.5
Laboratory Techni...	Yes	~7.0
Research Scientist	Yes	~8.5
Sales Represent...	No	~2.5
Sales Represent...	Yes	~9.0

Canvas 1 Canvas 2 Canvas 3

Oracle SQL Developer

File Edit View Navigate Run Team Tools Window Help

Connections Data Miner

Start Page 360 View Star Schema Analytics BUY_INSURANCE_WORKFLOW BUY_INSURANCE

Parallel Query Off

Connections

- BITWA 16_12c_DB_Cloud_Jonathan2
- BITWA DMUSER - Learn Predictive Analytics
- CBERGER dbpm23 Cloud instance C
- Charlie Database 12c laptop 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
 - New Buy Insurance WF
 - New Customer Analytics

Scatter Box plots etc. 1 Explore Data

CUST_INSUR_LTV1

Filter Columns

Multiple Classification Models

Explore Data 1

CUST_INSUR_LTV_APPLY1

Explore Data 1

Most Likely Customers

Clustering Segmentation 1

Filter Columns Details 1

LIKELY_BUY_INSURANCE_CUSTMRS 1

Tree Settings

Parallel Query Off

Workflow Editor

Components

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

Transforms

- Aggregate
- Filter Columns
- Filter Columns Details
- Filter Rows
- Join
- 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

BUY INSURANCE - Structure

Clustering Segmentation 1

Explore Data

Explore Data 1

Filter Columns

Multiple Classification Models

Filter Columns Details 1

Scatter Box plots etc. 1

CUST_INSUR_LTV1

CUST_INSUR_LTV_APPLY1

Most Likely Customers

LIKELY_BUY_INSURANCE_CUSTMRS

Links

- From "CUST_INSUR_LTV1" to "Explore Data"
- From "CUST_INSUR_LTV1" to "Explore Data 1"
- From "Filter Columns" to "Multiple Classification Models"

Script Output

```
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,
```

Query Result

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

Rule Surrogates Target Values

Node Rule:

If BANK_FUNDS > 246
 And CHECKING_AMOUNT > 282
 And MONEY_MONTHLY_OVERDRAWN <= 54.215
 Then No

Confidence	0.8515671200473093
Support	0.1832863646217212

Multiple Classification Models - Properties

Models

Build Model Settings

OML4SQL: Model Build and Real-time Prediction

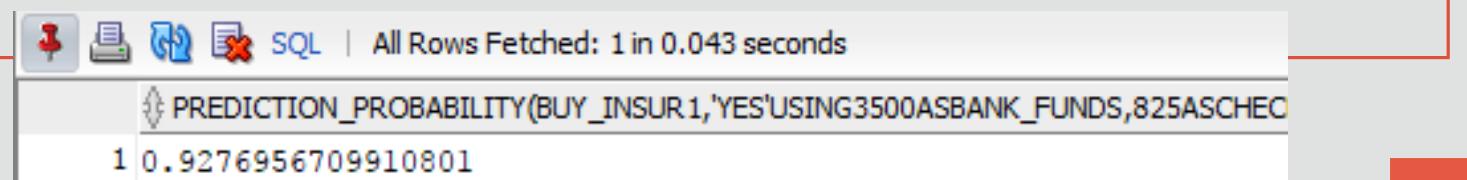
Simple SQL Syntax—Classification Model

Model build (PL/SQL)

```
BEGIN
  DBMS_DATA_MINING.CREATE_MODEL(
    model_name          => 'BUY_INSUR1',
    mining_function     => dbms_data_mining.classification,
    data_table_name     => 'CUST_INSUR_LTV',
    case_id_column_name => 'CUST_ID',
    target_column_name  => 'BUY_INSURANCE',
    settings_table_name => 'CUST_INSUR_LTV_SET');
END;
```

Real-time scoring (SQL query)

```
SELECT prediction_probability(BUY_INSUR1, 'Yes'
  USING 3500 as bank_funds, 825 as checking_amount, 400 as credit_balance, 22 as age,
  'Married' as marital_status, 93 as MONEY_MONTLY_OVERDRAWN, 1 as house_ownership)
FROM dual;
```



The screenshot shows the Oracle SQL developer interface with the following details:

- Toolbar icons: New, Open, Save, Undo, Redo, SQL.
- Status bar: All Rows Fetched: 1 in 0.043 seconds.
- Result grid:

	PREDICTION_PROBABILITY(BUY_INSUR1,'YES'USING3500ASBANK_FUNDS,825ASCHEC)
1	0.9276956709910801

```
#####
# create demo data set
#####

set.seed(123)      # enable repeatable results
options(digits=4)  # limit decimal output

-- generate a data set with three clusters

dat <- rbind(matrix(rnorm(1000, sd = 0.3), ncol = 2),          # cluster 1
              matrix(rnorm(1000, mean = 1, sd = 0.3), ncol = 2),  # cluster 2
              matrix(rnorm(1000, mean = 2.5, sd = 0.4), ncol = 2)) # cluster 3
colnames(dat) <- c("x", "y")
dat <- data.frame(dat)

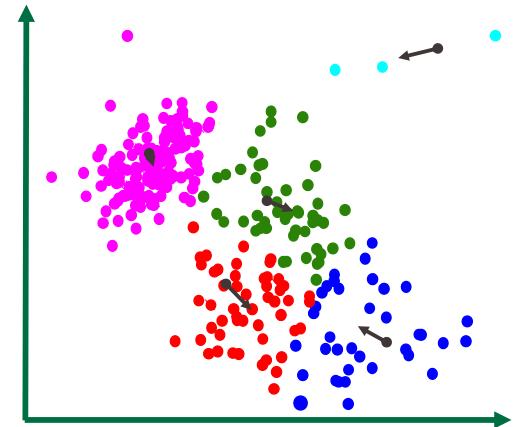
-- view the clusters

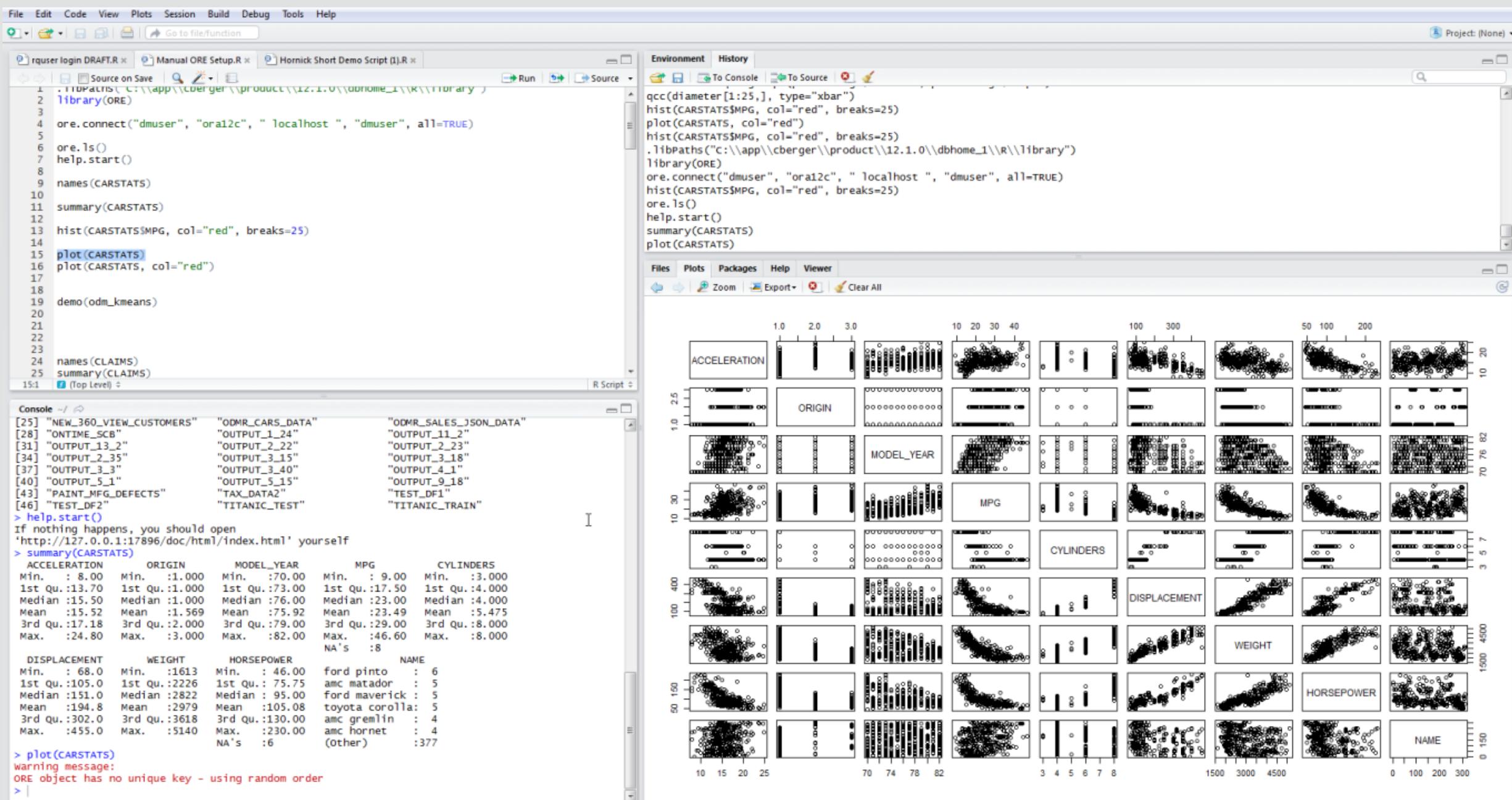
plot(dat$x, dat$y)

-- create a temporary database table using ore.push - object deleted when db connection ends

X <- ore.push (data.frame(dat))

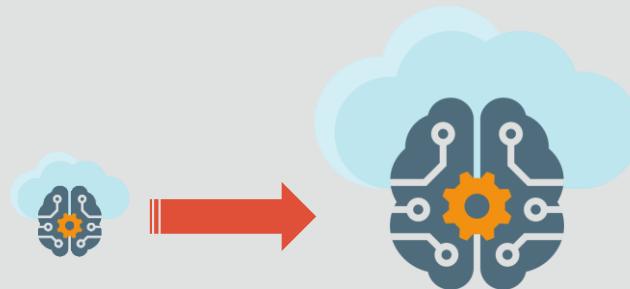
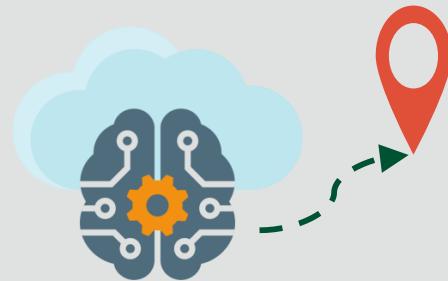
class(X)
```





Architektura Oracle AI/ML

Oracle Machine Learning Key Attributes



Automated

Get better results faster with less effort – even non-expert users

Scalable

Handle big data volumes using parallel, distributed algorithms – no data movement

Production-ready

Deploy and update data science solutions faster with integrated ML platform

Increase productivity, Achieve enterprise goals, Innovate More



Oracle Machine Learning

OML4SQL

Oracle Advanced Analytics
SQL API

OML4R

Oracle R Enterprise
R API

OML4Py*

Python API

OML Notebooks

with Apache Zeppelin on
Autonomous Database

Oracle Data Miner

Oracle SQL Developer extension

OML4Spark

Oracle R Advanced Analytics
for Hadoop

OML Microservices*

Supporting Oracle Applications
Image, Text, Scoring, Deployment,
Model Management

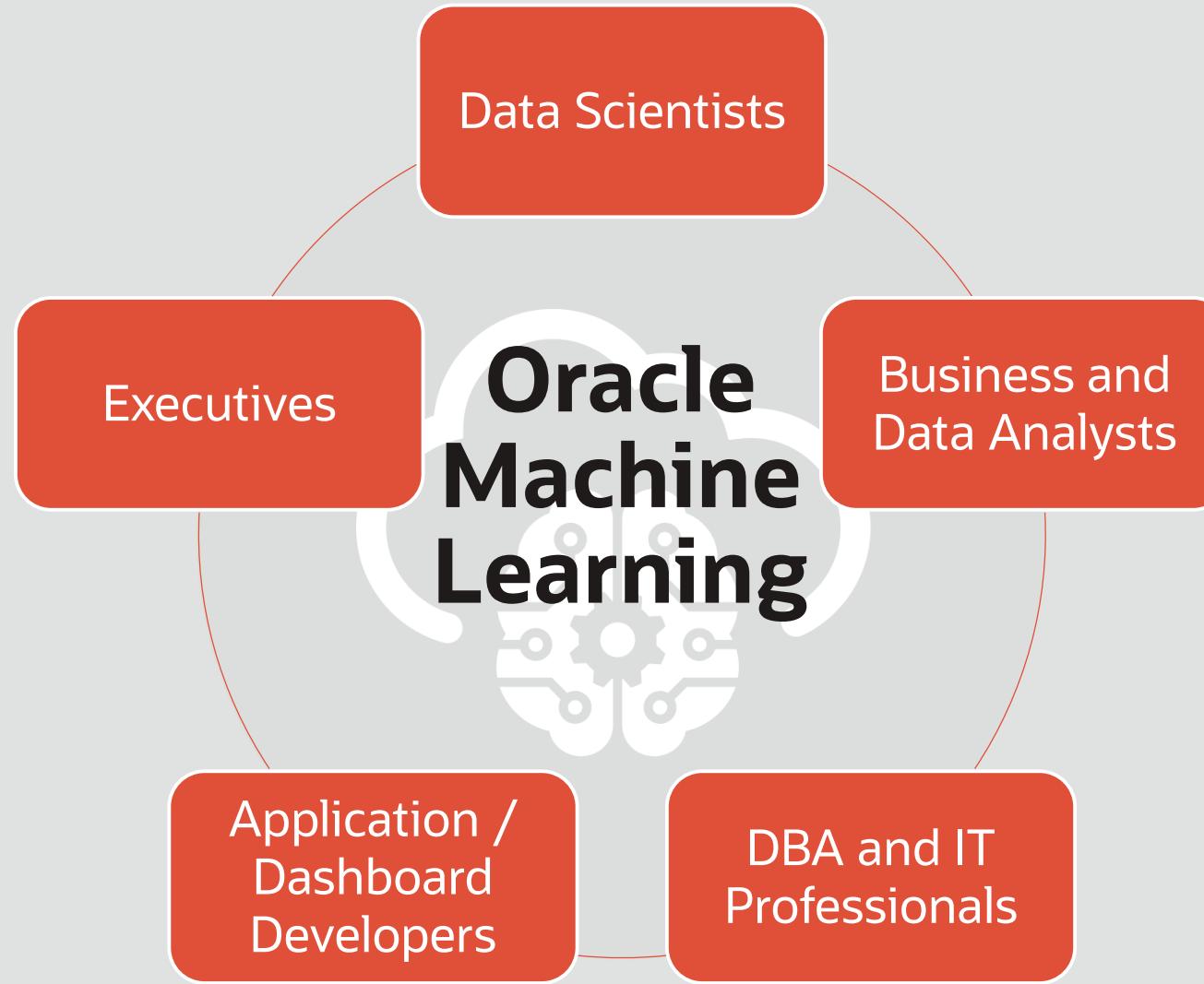
Oracle DV

ad-hoc ML

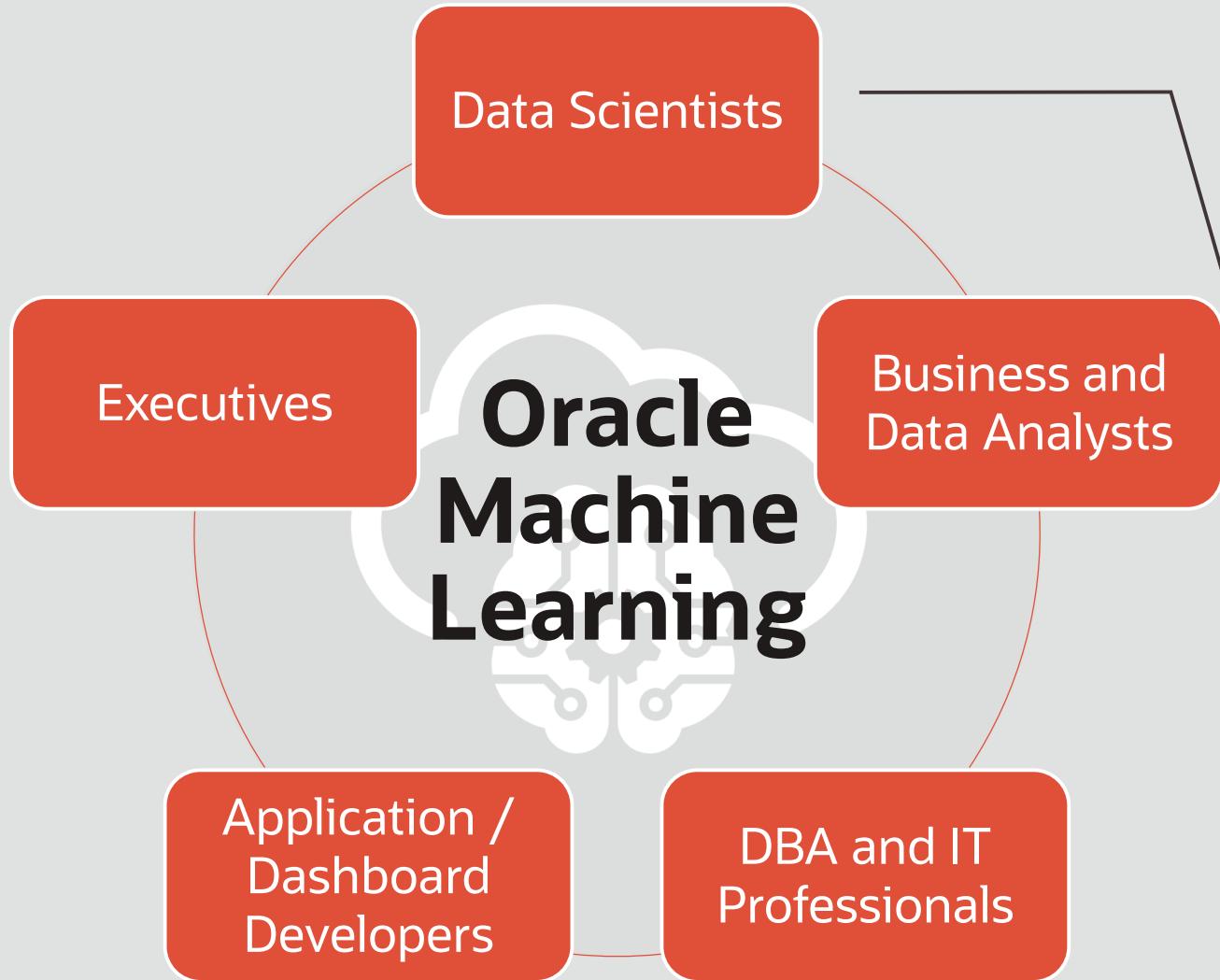


* Coming soon

OML empowers Enterprise Users

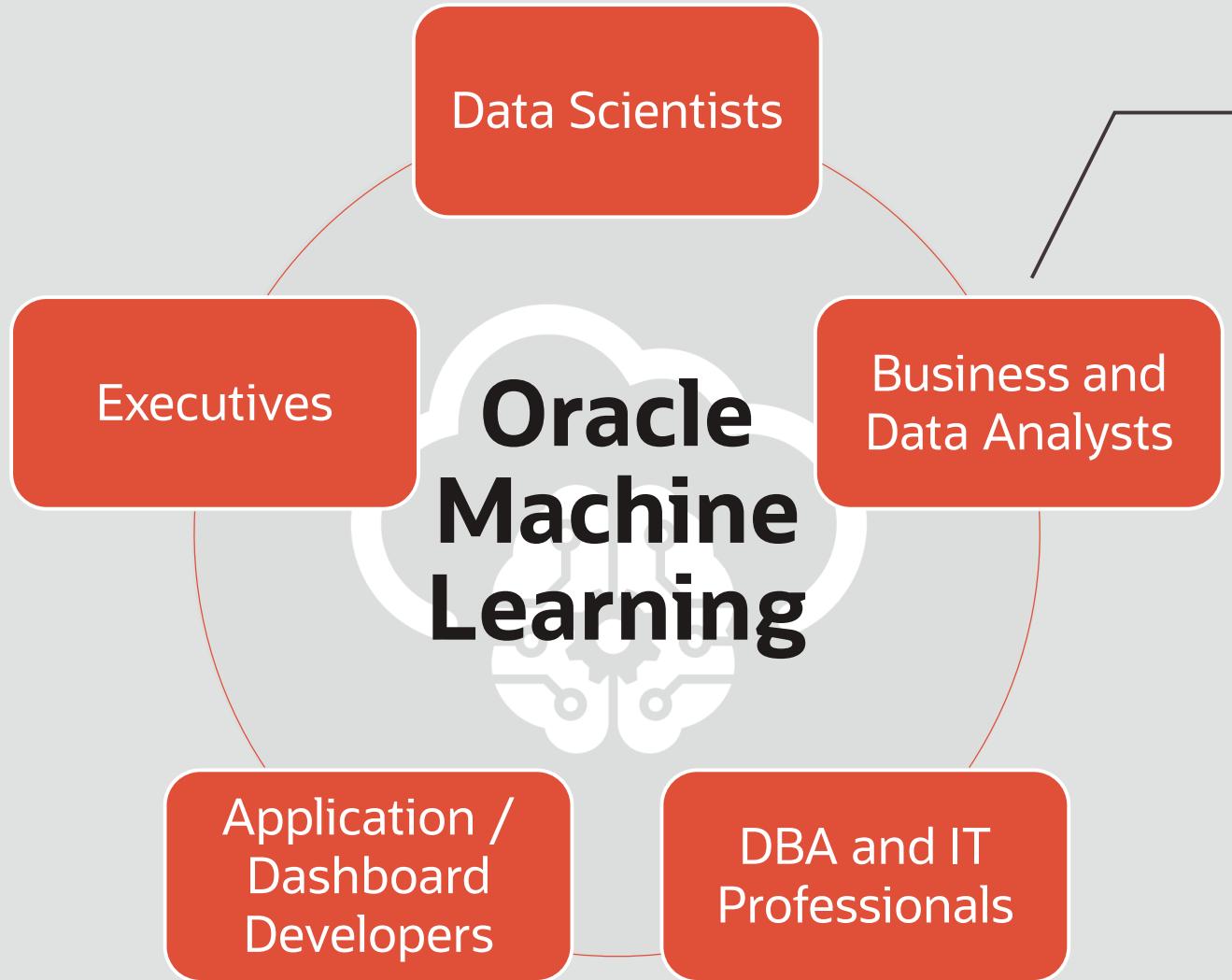


Data Scientists



- Popular data science languages: Python, R, SQL
- Augment with 3rd party packages
- Scalability and performance
- Automation-enhanced productivity
- Greater enterprise collaboration
- Integrate and analyze data across the enterprise

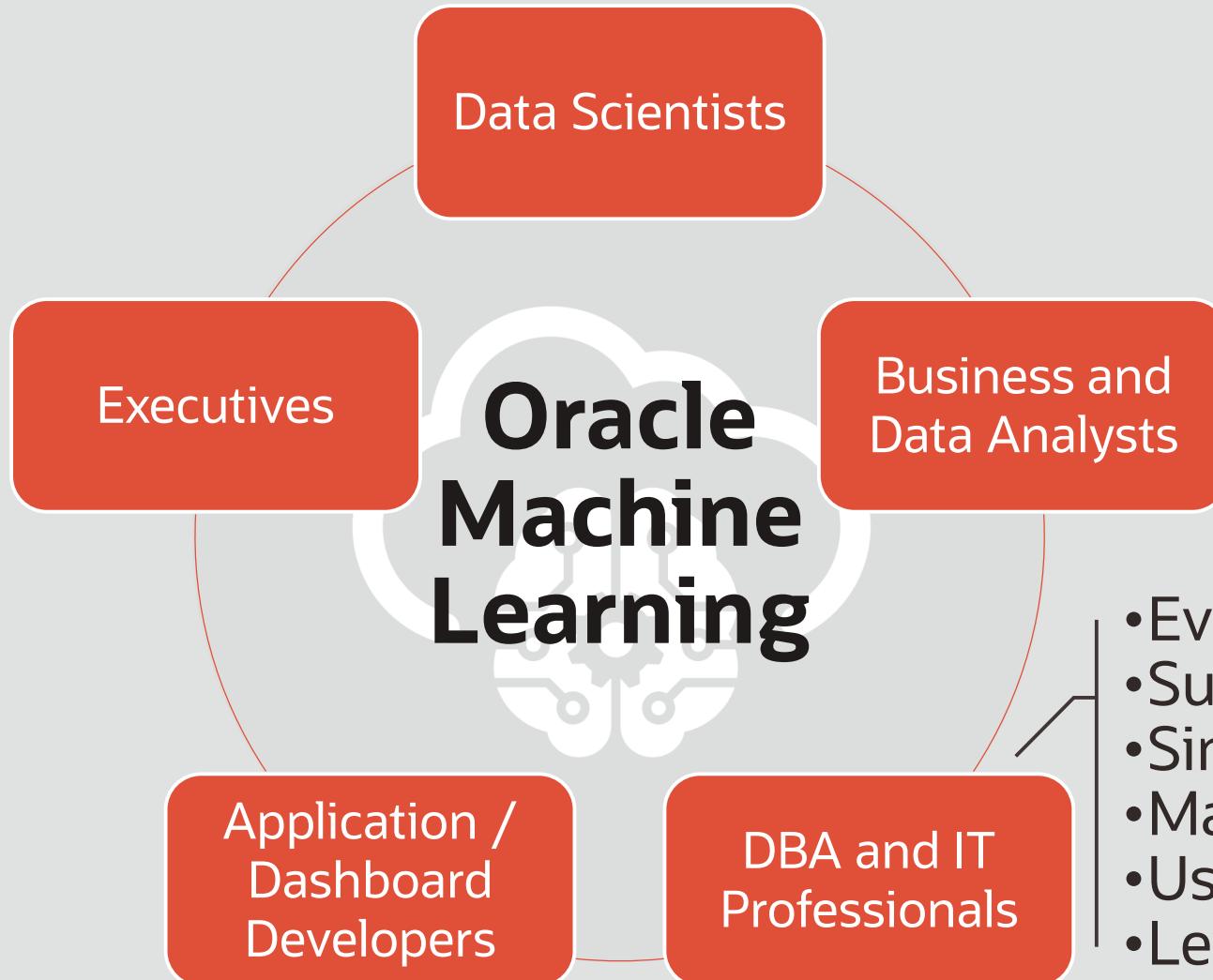
Business and Data Analysts



- Expand analytical tool set with ML
- Enable non-ML experts with AutoML
- Leverage domain knowledge for better results
- Collaborate with Data Scientists and IT



DBA and IT Professionals

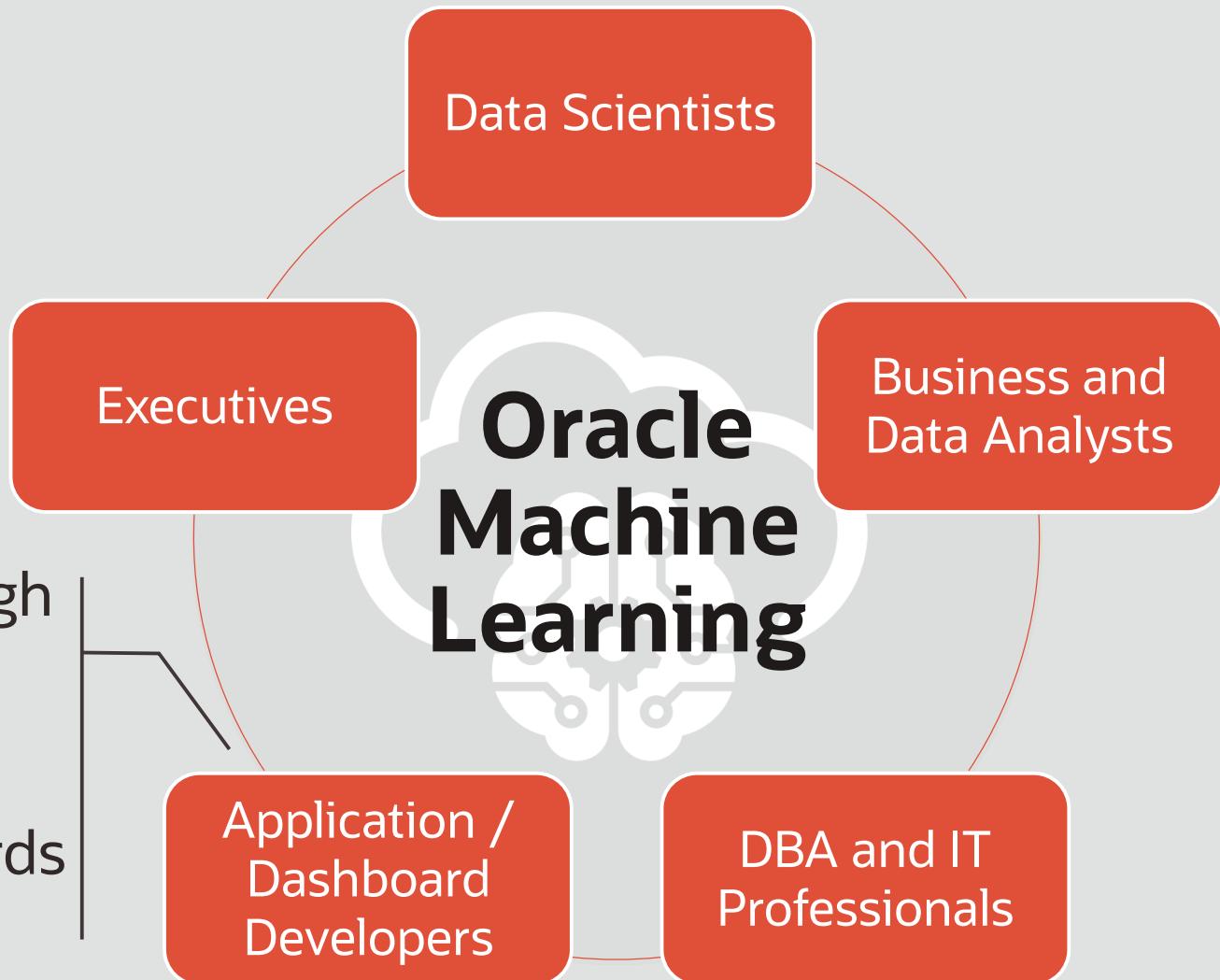


- Even greater value from Oracle investment
- Support scalability and performance
- Simpler, streamlined infrastructure
- Maintain data security, backup, recovery
- Use SQL, expand to Python and R
- Leverage Database and Big Data sources

Application and Dashboard Developers

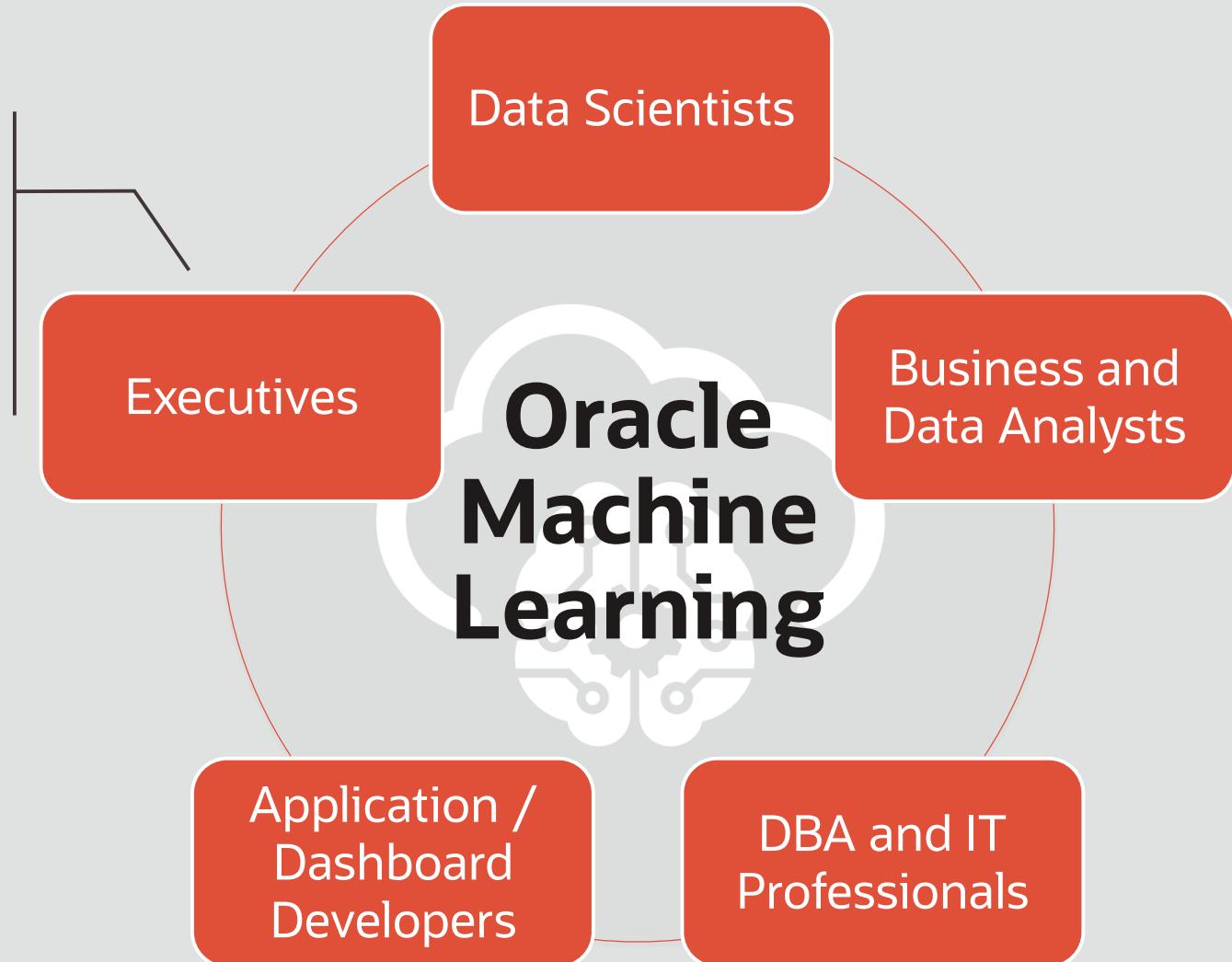


- Realize intelligent solutions faster through Oracle stack integration
- Easily uptake data scientists' R, Python, SQL scripts and rapidly deploy solutions
- Embed ML in applications and dashboards using SQL, REST, and SODA APIs

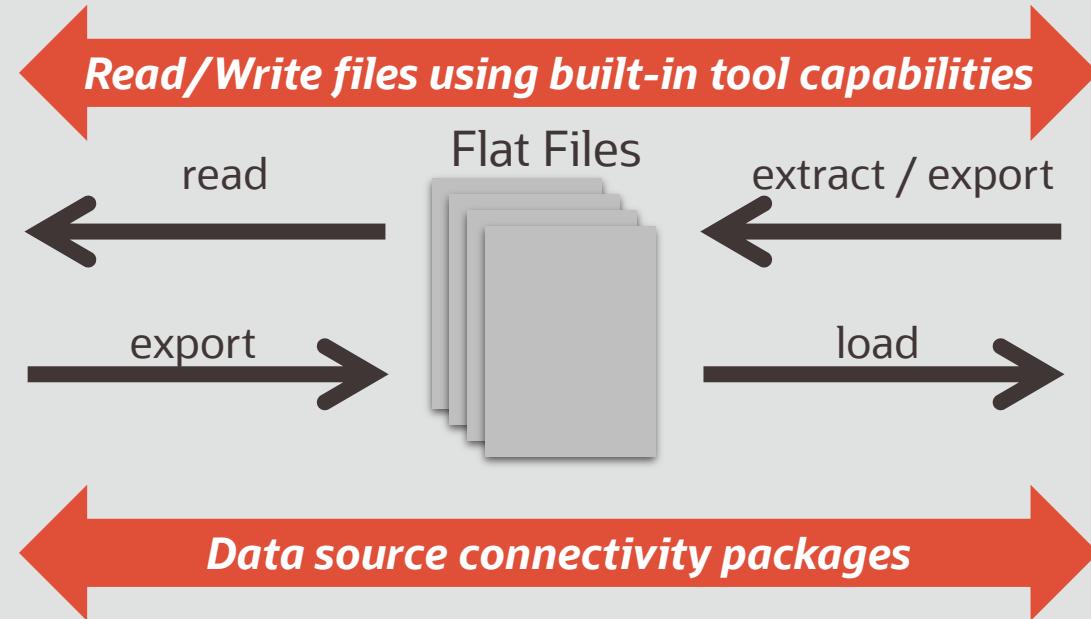


Executives

- Benefit from world-class data management technology and support
- Democratize ML across the enterprise to enable better data-driven decisions
- Deploy solutions faster to realize ROI



Traditional Analytics and Data Source Interaction



Deployment
Ad hoc
cron job

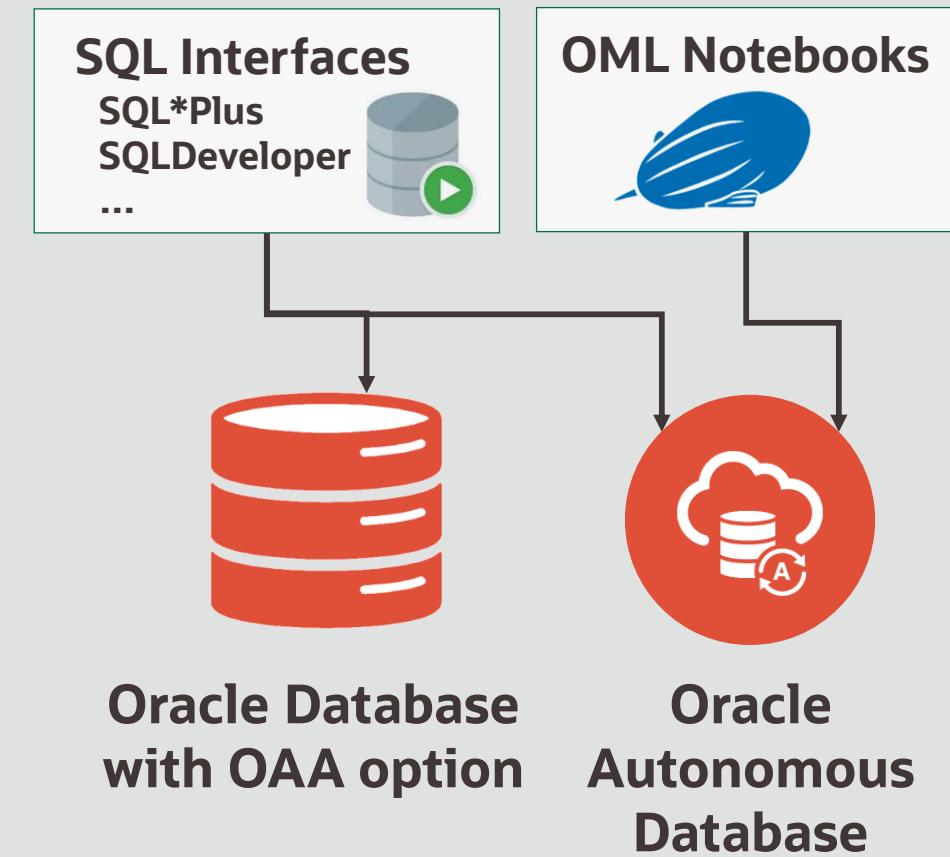
Access latency
Paradigm shift: R/Python → *Data Access Language* → R/Python
Memory limitation – data size, in-memory processing
Single threaded
Issues for backup, recovery, security
Ad hoc production deployment

Oracle Machine Learning for SQL (OML4SQL)

Oracle Machine Learning for SQL

Component of Oracle Autonomous Database and
Oracle Advanced Analytics option to Oracle Database

In-database, parallel, distributed algorithms
ML models as first class database objects
Export / import models across databases
Batch and real-time scoring
Explanatory predictive details
Leverage ML across Oracle stack



OML4SQL: Model Build and Real-time Prediction

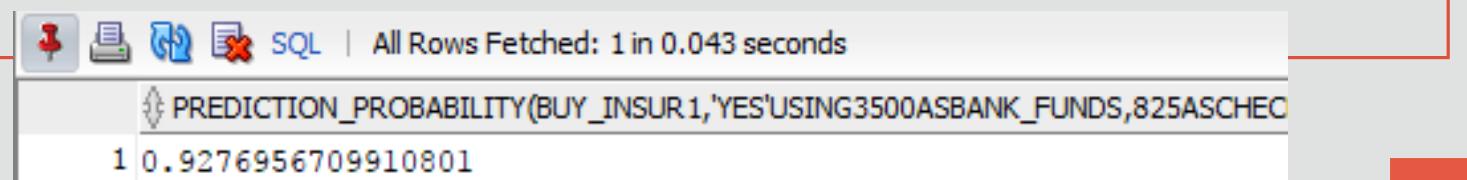
Simple SQL Syntax—Classification Model

Model build (PL/SQL)

```
BEGIN
  DBMS_DATA_MINING.CREATE_MODEL(
    model_name          => 'BUY_INSUR1',
    mining_function     => dbms_data_mining.classification,
    data_table_name     => 'CUST_INSUR_LTV',
    case_id_column_name => 'CUST_ID',
    target_column_name  => 'BUY_INSURANCE',
    settings_table_name => 'CUST_INSUR_LTV_SET');
END;
```

Real-time scoring (SQL query)

```
SELECT prediction_probability(BUY_INSUR1, 'Yes'
  USING 3500 as bank_funds, 825 as checking_amount, 400 as credit_balance, 22 as age,
  'Married' as marital_status, 93 as MONEY_MONTLY_OVERDRAWN, 1 as house_ownership)
FROM dual;
```



The screenshot shows the Oracle SQL developer interface. At the top, there are several icons: a red pushpin, a blue document, a blue gear, and a red X. To the right of these is the text "SQL | All Rows Fetched: 1 in 0.043 seconds". Below this is a table with one row of data. The table has two columns: the first is a small icon of a document with a red X, and the second is the text "PREDICTION_PROBABILITY(BUY_INSUR1,'YES'USING3500ASBANK_FUNDS,825ASCHEC)". The row itself has two columns: the first is the number "1" and the second is the probability value "0.9276956709910801".

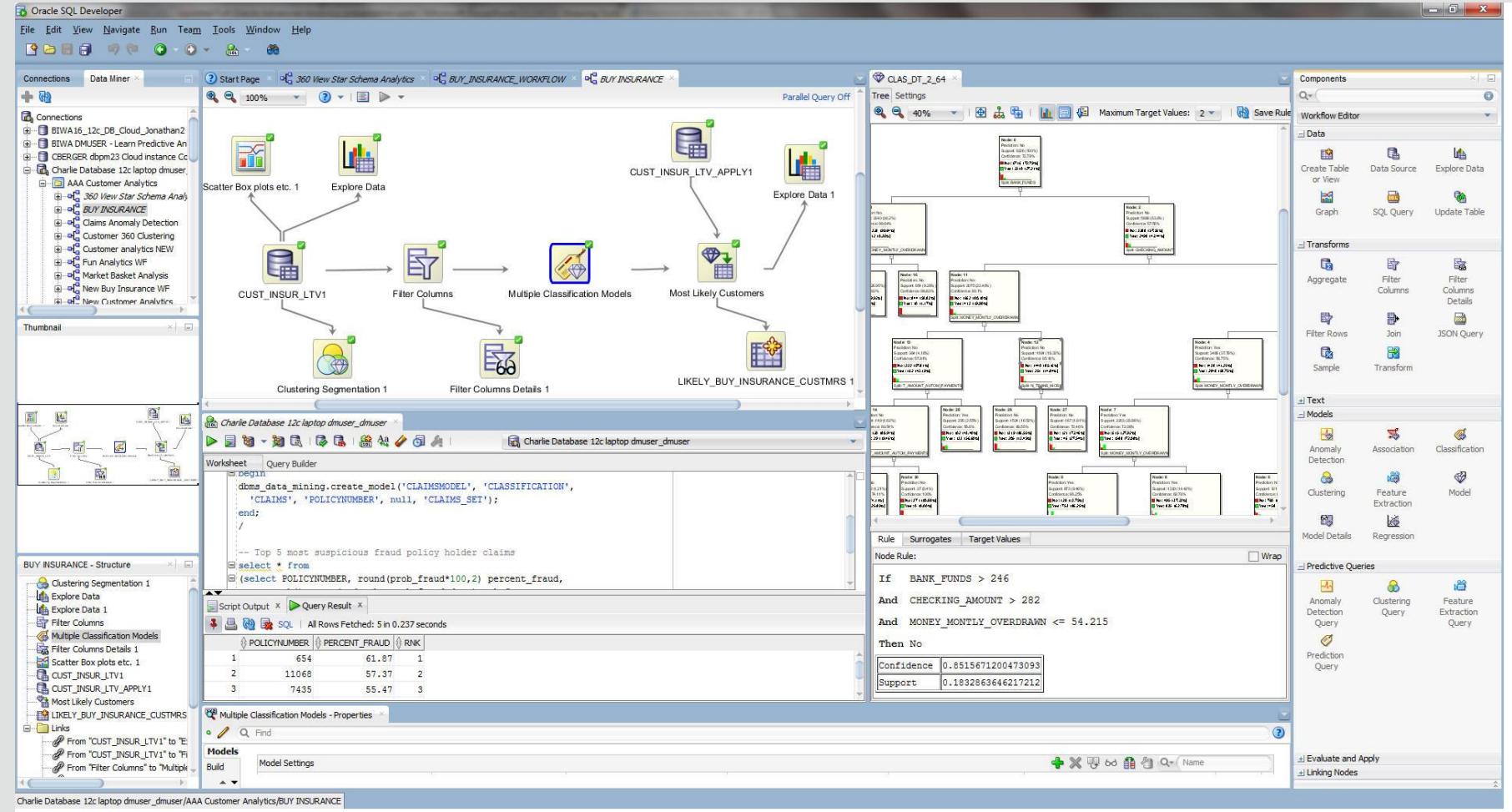
1	0.9276956709910801
---	--------------------

Oracle Data Miner User Interface

Create analytical workflows – supports “Citizen Data Scientists”



SQL Developer Extension
Automates typical data science steps
Easy to use drag-and-drop interface
Analytical workflows quickly defined and shared
Wide range of algorithms and data transformations
Generate SQL code for immediate deployment





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
DEP_BALANCE		NUMBER	0	167	2.8667		1,032.7312	25	25	24,471	3,103.6812	9,632,836.6997

AGE By MARITAL_STATUS

LTV_BIN By MARITAL_STATUS

N_OF_DEPENDENTS By MARITAL_STATUS

SALARY By MARITAL_STATUS

Scatter Box plots etc. Components

Graph1

Graph2

Graph3

Connections Data Miner

Start Page BUY_INSURANCE_WORKFLOW CUST_INSUR_LTV1 Graph 1 Parallel Query Off

Connections dmuser

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

Thumbnail

BUY_INSURANCE_WORKFLOW - ...

- Clustering Segmentation
- Explore/Profile Data
- Explore Data
- Prediction Queries by Region
- Prediction Query
- Filter Cols_Attr 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

dmuser/Customers R Us Project/BUY_INSURANCE_WORKFLOW

Components

Workflow Editor

Data

Data Source Explore Data

SQL Query Update Table

Filter Columns

Filter Columns Details

Join

JSON Query

Association Classification

Feature Extraction Model

Regression

Help

OK Cancel

100% 100%

Parallel Query Off

Edit Filter Columns Node

Show Attribute Importance

Show Data Quality

Settings

Columns

Name	Type	Output	Rank	Importance	% Null	% Unique	% Constant	Hints
XYZ BANK_FUNDS	NUMBER	→	1	0.2039	0	21.3511	35.6016	
XYZ N_TRANS_ATM	NUMBER	→	2	0.1217	0	0.4438	21.499	
XYZ N_TRANS_TELLER	NUMBER	→	3	0.1213	0	0.4931	32.3471	
XYZ MONEY_MONTLY_OVERDRAWN	NUMBER	→	4	0.1207	0	18.787	16.3215	
XYZ T_AMOUNT_AUTOM_PAYMENTS	NUMBER	→	5	0.1054	0	59.9606	21.5483	
XYZ MONTHLY_CHECKS_WRITTEN	NUMBER	→	6	0.0848	0	0.9369	19.428	
XYZ N_OF_DEPENDENTS	NUMBER	→	7	0.0315	0	0.3452	34.4181	
XYZ TIME_AS_CUSTOMER	NUMBER	→	8	0.0221	0	0.2465	31.3116	
XYZ CHECKING_AMOUNT	NUMBER	→	9	0.0183	0	30.3254	62.7712	
XYZ CREDIT_BALANCE	NUMBER	→	10	0.0135	0	9.4181	90.5819	
XYZ N_TRANS_KIOSK	NUMBER	→	11	0.0084	0	0.5424	43.6884	
XYZ MORTGAGE_AMOUNT	NUMBER	→	12	0.0073	0	21.2525	23.1262	
XYZ SEX	VARCHAR2	→	13	0.0064	0	0.0986	66.075	
XYZ MARITAL_STATUS	VARCHAR2	→	14	0.0056	0	0.2465	34.5168	
XYZ HOUSE_OWNERSHIP	NUMBER	→	15	0.004	0	0.1479	71.1538	
XYZ N_MORTGAGES	NUMBER	→	16	0.004	0	0.1479	71.1538	

Predictive Queries Evaluate and Apply Linking Nodes



Connections Data Miner X | Welcome Page X | Predicting LTV X | Targeting Best Customers X | test X | Components X

Connections

BI APPS dmuser

- OOA Testdrive
 - Clustering
 - Market Ba
 - Predicting
 - Targeting
 - test
- SampleApp
 - Customer
 - LTV Bin Cl
 - Market Ba
 - Predicting
 - Predictive
 - Revenue

Ravello DMUSER

Ravello RQUSER

Columns

Name	Type
AGE	NUMBER
BANK_FUNDS	NUMBER
BUY_INSURANCE	VARCHAR2
CAR_OWNERSHIP	NUMBER
CHECKING_AMOUNT	NUMBER
CREDIT_BALANCE	NUMBER
CREDIT_CARD_LIMITS	NUMBER
CUSTOMER_ID	VARCHAR2
FIRST	VARCHAR2
HAS_CHILDREN	NUMBER
HOUSE_OWNERSHIP	NUMBER
LAST	VARCHAR2
LTV	NUMBER
LTV_BIN	VARCHAR2
MARITAL_STATUS	VARCHAR2
MONEY_MONTHLY_OVERDRAWN	NUMBER

Define Filter Column Settings

Data Quality

% Nulls less than or equal 95

% Unique less than or equal 95

% Constant less than or equal 95

Attribute Importance

Target: <Select a Target>

Importance Cutoff 0

Top N 100

Sample Technique: System Determined

Stratified Cutoff 10

Sampling (Data Quality and Attribute Importance)

Sample Size 10 000

Pomoc OK Anuluj

Target: <Select a Target> Settings

Name

OK Anuluj

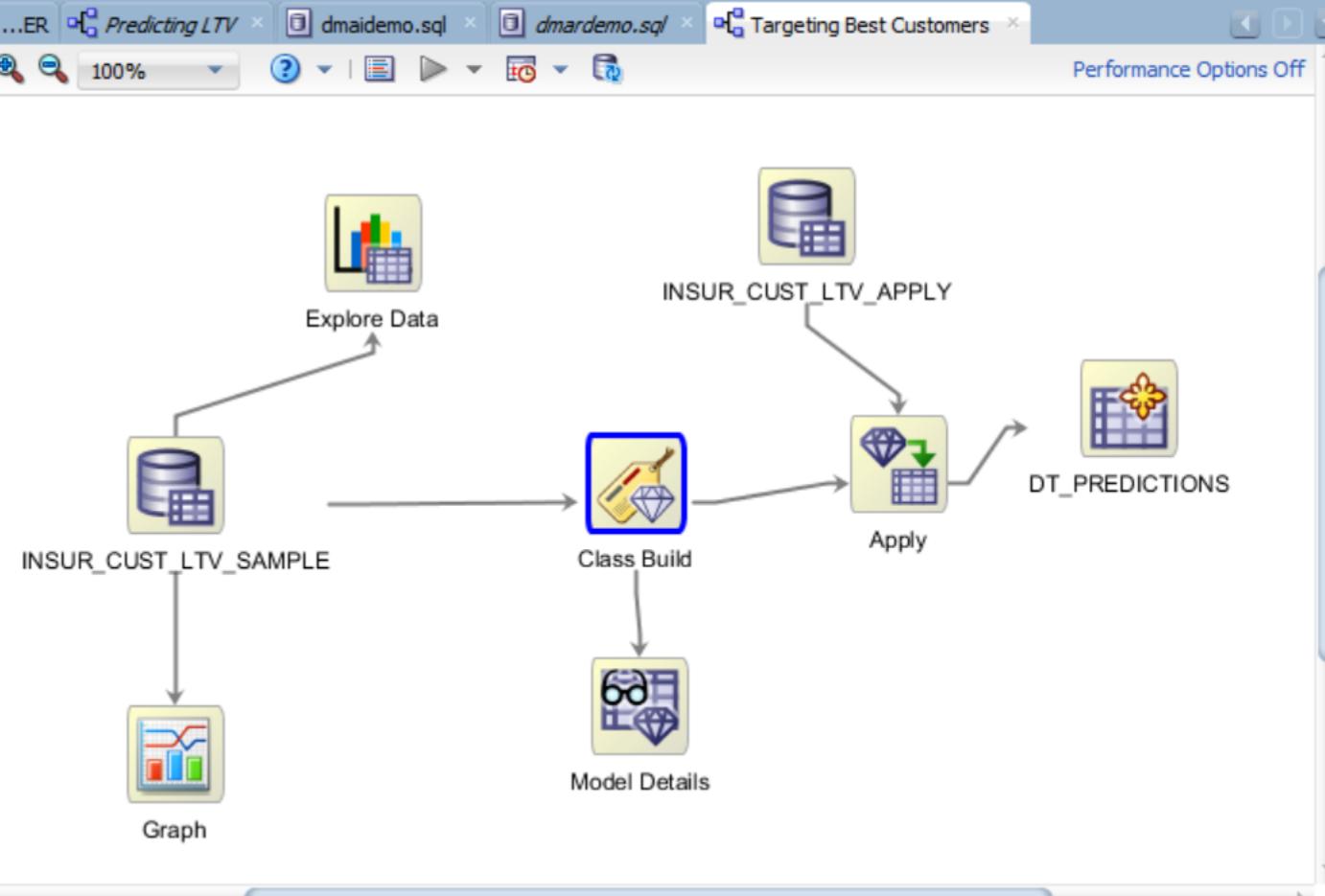
Attribute Importance

Target:



Connections

- BI APPS dmuser
 - OAA Testdrive Solutions
 - Clustering EM
 - Market Basket Analysis
 - Predicting LTV
 - Targeting Best Customers
- SampleApp
 - Customer Segmentation
 - LTV Bin Classification
 - Market Basket Analysis
 - Predicting Sales
 - Predictive Queries
 - Revenue Regression
- Ravello DMUSER
- Ravello RQUSER



Messages - Log

```

ODMR_ENGINE_SEC Compiled (with errors)
ODMR_ENGINE_TRANSFORMS_SEC Compiled (with errors)
ODMR RNGTNR MTNTNG SEC Compiled (with errors)
  
```

Messages

Logging Page

Statements

Compiler

Components

Workflow Editor

Data

or View



Transforms

Text

Models

Predictive Queries

Model Operations

Linking Nodes



Class Build - Properties

Find

 Generate Selected Test Results for Model T

Test Data

 Use All Mining Build Data for Testing

 Use Split Build Data for Testing

Split for Test (%): 50

Create Split as:

Table

 Use Test Data Source for Testing

Details



Oracle Machine Learning Notebooks

Autonomous Database as a Data Science Platform

Collaborative UI

Based on Apache Zeppelin

Supports data scientists, data analysts,
application developers, DBAs

Easy sharing of notebooks and templates

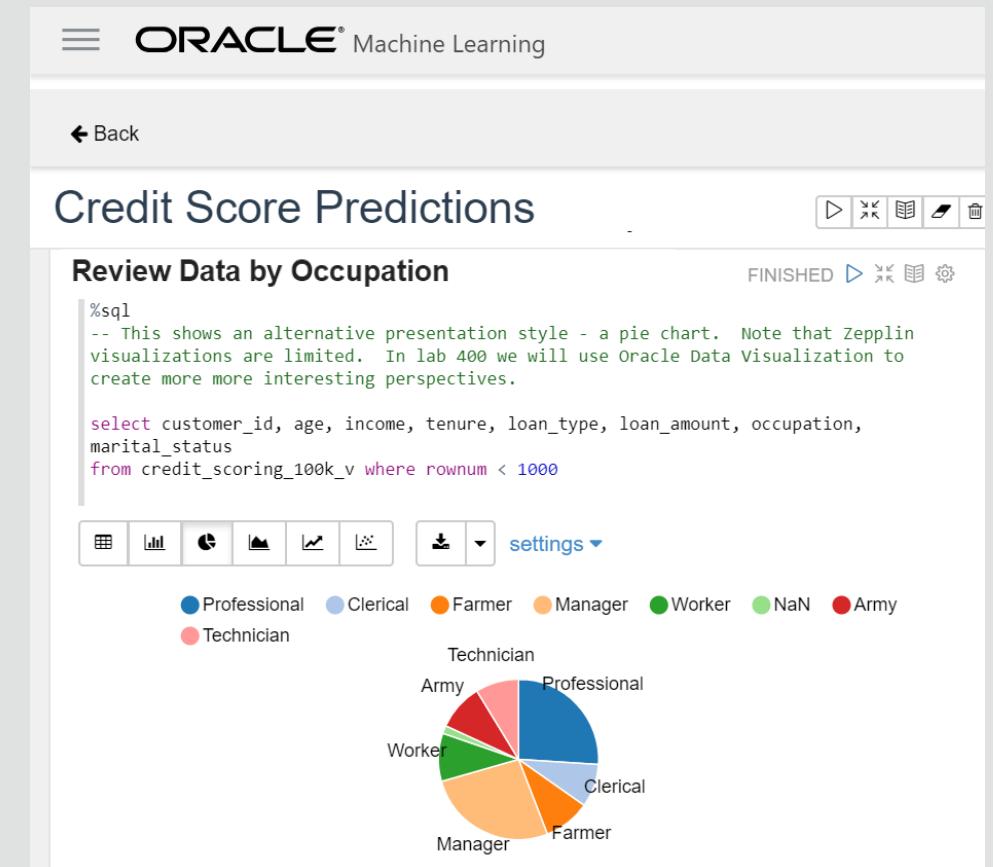
Permissions, versioning, and execution scheduling

Included with Autonomous Database

Automatically provisioned, managed, backed up

In-database SQL algorithms and analytics functions

Soon to be augmented with Python and R



https://adwc.uscom-west-1.oraclecloud.com/oml/tenants/OCID1.TENANCY.OC1..AAAAAAAFCUE47PQMRF4VIGNEEBGBCMMOY5R7XVOYPIJCQQGE32...

ORACLE® Machine Learning Charlie Project [Charlie Workspace] CHARLIE Connected

Back

Credit Score Predictions Simplified ...

STEP 9: Enough with Simple Charting; Let's Run Some OML Machine Learning Algorithms!

FINISHED

Create Attribute Importance Machine Learning Model

Good Credit Customers

```
%script
-- Find the importance of attributes that independently impact the target attribute:
CREDIT_SCORE_BIN

DECLARE
v_sql varchar2(100);

BEGIN
BEGIN EXECUTE IMMEDIATE 'DROP TABLE ai_explain_output_credit_score_bin';
EXCEPTION WHEN OTHERS THEN NULL;
END;

BEGIN
DBMS_PREDICTIVE_ANALYTICS.EXPLAIN( Attribute Importance Model
  data_table_name      -> 'CREDIT_SCORING_100K_V',
  explain_column_name -> 'CREDIT_SCORE_BIN',
  result_table_name   -> 'AI_EXPLAIN_OUTPUT_CREDIT_SCORE_BIN');
END;

End;
```

Display the Top N Attributes for Good Credit Customers

```
%sql
-- Display those attributes that most influence the target field (Good Credit
customers)

Select * from ai_explain_output_CREDIT_SCORE_BIN where rownum < 7;
```

settings

EXPLANATORY_VALUE

0.64
0.5
0

CUSTOMER_DMG_SEGMENT
GUEST_CREDIT_CARD_LIMIT
RESIDENTIAL_STATUS

Took 1 sec. Last updated by CHARLIE at July 30 2018, 3:42:43 PM. (outdated)

Shared Templates

[+ Create Notebook](#)[Edit](#)[Delete](#) [Anomaly Detection](#)

Author: USER07

Date Added: 8/13/19 11:31 PM

0 Likes 0 0

[Association Rules](#)

Author: USER07

Date Added: 8/13/19 11:32 PM

0 Likes 1 1

[Attribute Importance](#)

Author: USER07

Date Added: 8/14/19 6:00 PM

1 Likes 3 4

[Classification Prediction M...](#)

Author: USER07

Date Added: 8/13/19 11:32 PM

1 Likes 0 0

[Clustering](#)

Author: USER07

Date Added: 8/13/19 11:33 PM

1 Likes 0 0

[Credit Score Predictions W...](#)

Author: USER07

Date Added: 8/14/19 7:31 PM

0 Likes 0 0

[Credit Score Predictions W...](#)

10k version

Author: USER07

Date Added: 8/16/19 8:03 PM

0 Likes 0 1

[Credit Score Predictions W...](#)

10k version

Author: USER07

Date Added: 8/16/19 7:55 PM

0 Likes 0 2

[My First Notebook](#)

Author: USER07

Date Added: 8/13/19 11:37 PM

0 Likes 2 3

[Regression](#)

Author: USER07

Date Added: 8/13/19 11:34 PM

0 Likes 1 0

[SQL Query Scratchpad](#)

Author: USER07

Date Added: 8/13/19 11:51 PM

0 Likes 0 0

[SQL Script Scratchpad](#)

Author: USER07

Date Added: 8/13/19 11:35 PM

0 Likes 0 0

[SQL Statistical Functions](#)

Author: USER07

Date Added: 8/13/19 11:36 PM

0 Likes 0 0

[Targeting Top Customers 1...](#)

100K version

Author: USER07

Date Added: 8/15/19 9:46 PM

0 Likes 0 0

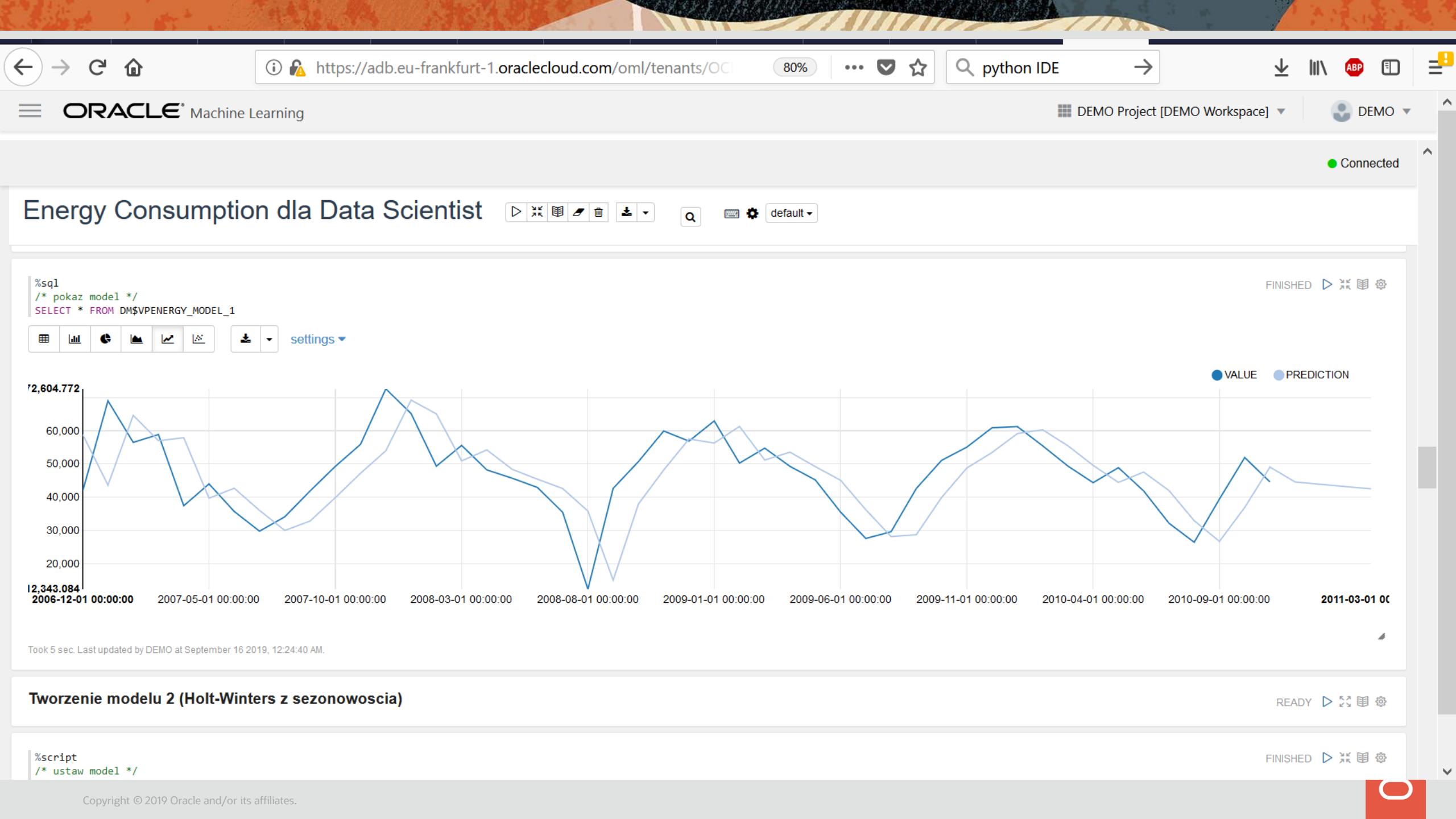
[Time Series Forecasting](#)

Author: USER07

Date Added: 8/13/19 11:35 PM

Tags: 'Time Series Forecasting'

0 Likes 0 0



Targeting High Credit Customers - Project

Prepare

Visualize

Narrate

Save



+ Click here or drag data to add a filter

Pie

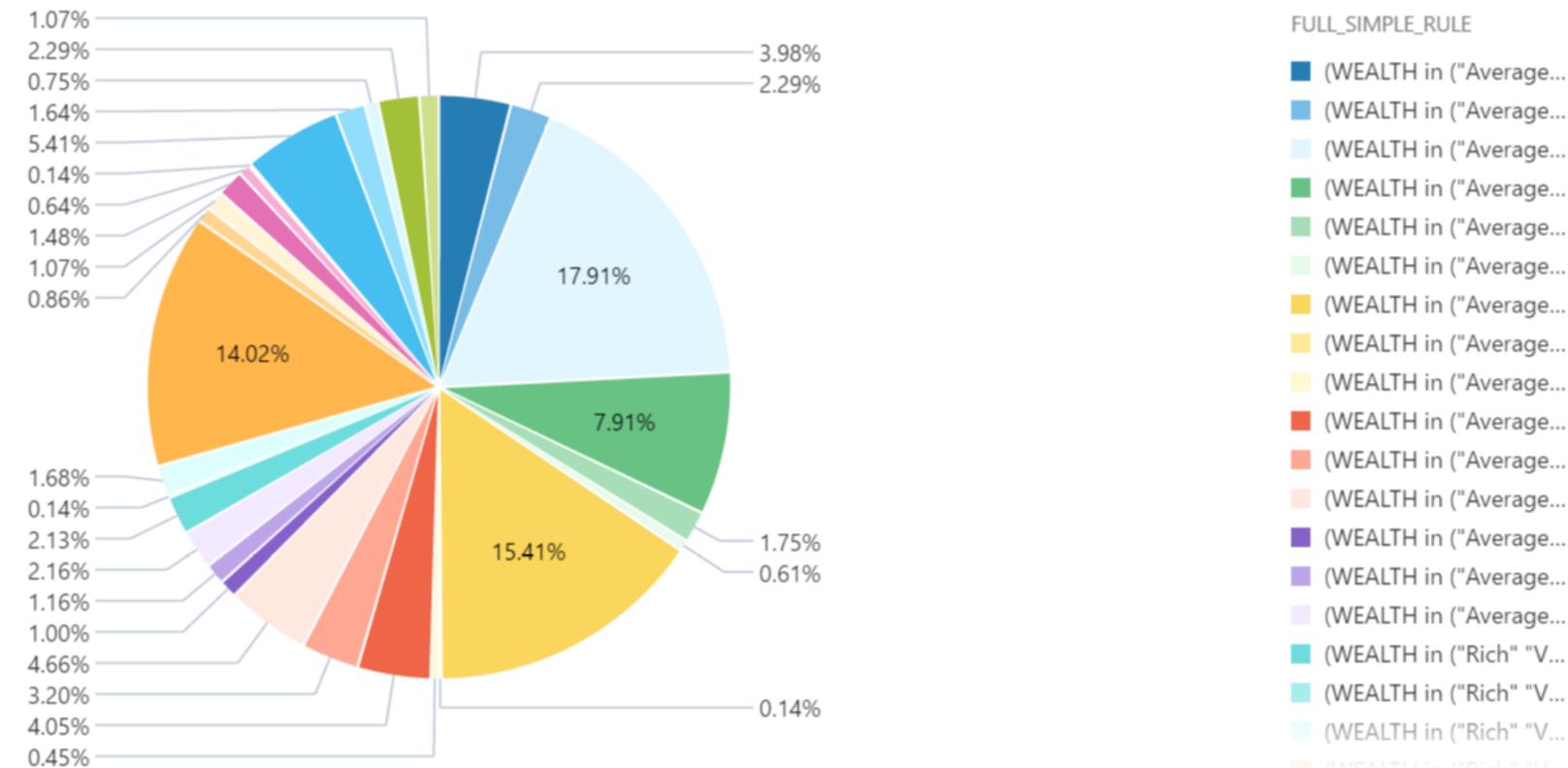
Trellis Columns

Trellis Rows

Values (Slice)
PREDICTION...Category
A PREDICTIONColor
A FULL_SIMPL...

Filters

PREDICTION_COUNT by PREDICTION, FULL_SIMPLE_RULE



Oracle Machine Learning for R (OML4R) Python (OML4Py)

Oracle is a founding member of **R consortium**

R Consortium mission

Work with and provide support to key organizations that develop, maintain, distribute and use R software through the identification, development and implementation of infrastructure projects

Promotes the growth and development of R as a leading platform for data science and statistical computing

Supports and collaborates with the [R Foundation](#), the governing body of the R Project

Funding projects to enhance R and support its users

Fostering the continued growth of R community and the data science ecosystem

Enables the use of R in commercial environments, and fostering collaboration between companies investing in R

See <https://www.r-consortium.org>

```
1 .TOPDIRS("C:\app\coherer\product\12.1.0\odmone_1\RK\TOPDIRS")
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
19 demo(odm_kmeans)
20
21
22
23
24 names(CLAIMS)
25 summary(CLAIMS)
15:1 [ (Top Level) ]
```

```
Console ~/ 
[25] "NEW_360_VIEW_CUSTOMERS"      "ODMR_CARS_DATA"          "ODMR_SALES_JSON_DA
[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.   :3.0
1st Qu.:13.70  1st Qu.:1.000   1st Qu.:73.00   1st Qu.:17.50  1st Qu.:4.0
Median :15.50  Median :1.000   Median :76.00   Median :23.00  Median :4.0
Mean   :15.52  Mean   :1.569   Mean   :75.92   Mean   :23.49  Mean   :5.4
3rd Qu.:17.18  3rd Qu.:2.000   3rd Qu.:79.00   3rd Qu.:29.00  3rd Qu.:8.0
Max.   :24.80  Max.   :3.000   Max.   :82.00   Max.   :46.60  Max.   :8.0
```

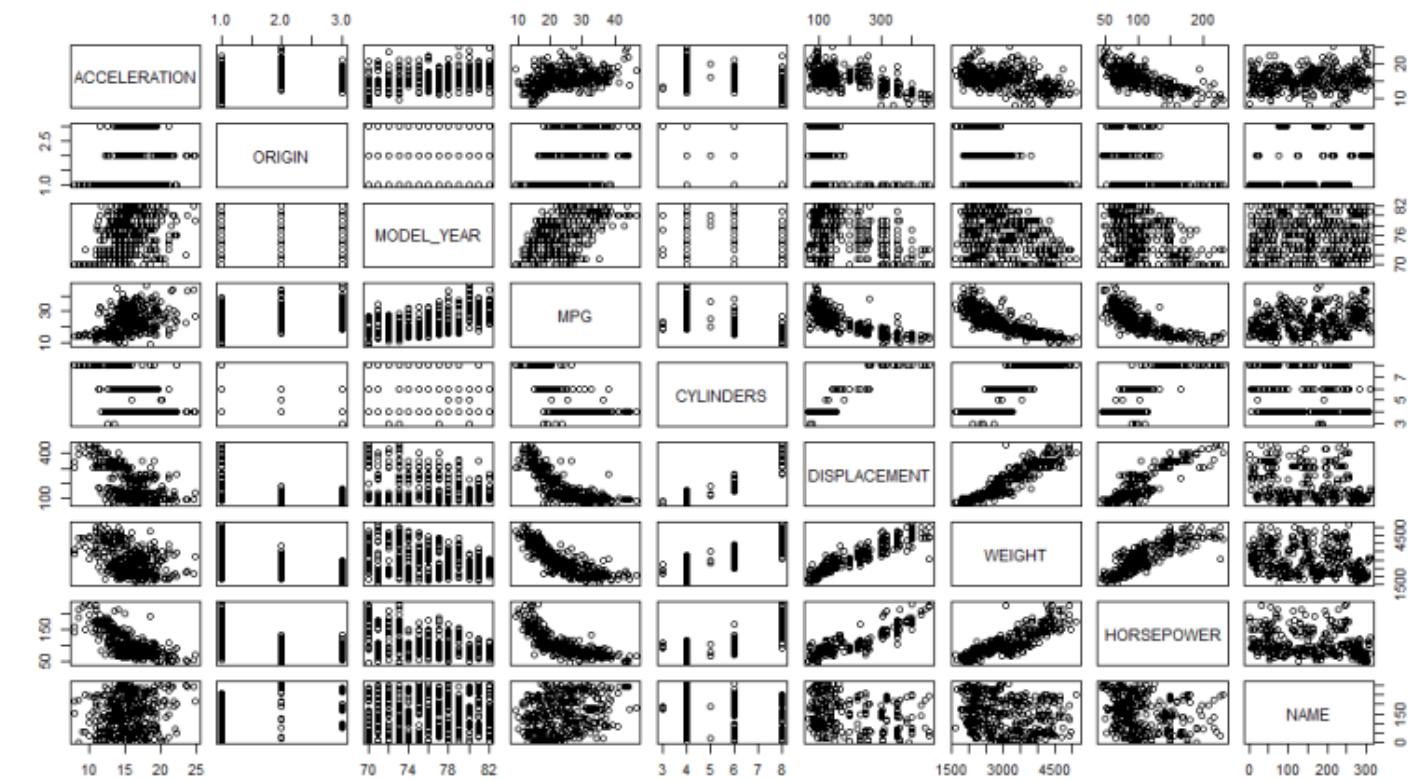
DISPLACEMENT	WEIGHT	HORSEPOWER	NA'S	:	8
Min. : 68.0	Min. : 1613	Min. : 46.00	ford pinto	:	
1st Qu.: 105.0	1st Qu.: 2226	1st Qu.: 75.75	amc matador	:	
Median : 151.0	Median : 2822	Median : 95.00	ford maverick	:	
Mean : 194.8	Mean : 2979	Mean : 105.08	toyota corolla	:	
3rd Qu.: 302.0	3rd Qu.: 3618	3rd Qu.: 130.00	amc gremlin	:	
Max. : 455.0	Max. : 5140	Max. : 230.00	amc hornet	:	
		NA's : 6	(other)	:	

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

Environment History

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

Files Plots Packages Help Viewer



Oracle Machine Learning for R and Python

Components of Oracle Advanced Analytics option to Oracle Database

Transparency layer

- Leverage proxy objects so data remain in database
- Overload native functions translating functionality to SQL
- Use familiar R / Python syntax on database data

Parallel, distributed algorithms

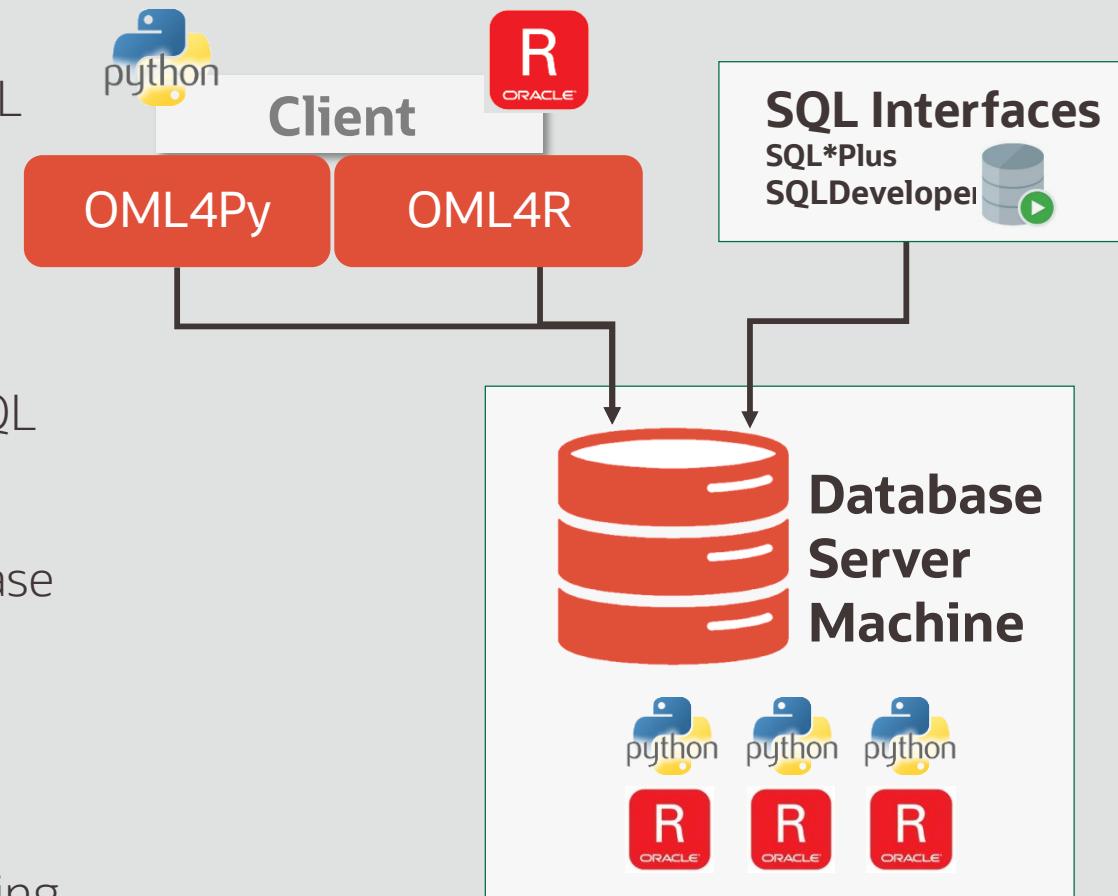
- Scalability and performance
- Exposes in-database algorithms available from OML4SQL

Embedded execution

- Manage and invoke R or Python scripts in Oracle Database
- Data-parallel, task-parallel, and non-parallel execution
- Use open source packages to augment functionality

OML4Py AutoML

- Model selection, feature selection, hyper-parameter tuning



Proxy objects

Example using OML4R interface

	Sepal.Length	Sepal.Width	Petal.Length	Petal.Width	Species
1	5.1	3.5	1.4	0.2	setosa
2	4.9	3.0	1.4	0.2	setosa
3	4.7	3.2	1.3	0.2	setosa
4	4.6	3.1	1.5	0.2	setosa
5	5.0	3.6	1.4	0.2	setosa
6	5.4	3.9	1.7	0.4	setosa

data.frame

Inherits from

Proxy
data.frame



```
> str(iris)
'data.frame': 150 obs. of 5 variables:
 $ Sepal.Length: num 5.1 4.9 4.7 4.6 5 5.4 4.6 5 4.4 4.9 ...
 $ Sepal.Width : num 3.5 3 3.2 3.1 3.6 3.9 3.4 3.4 2.9 3.1 ...
 $ Petal.Length: num 1.4 1.4 1.3 1.5 1.4 1.7 1.4 1.5 1.4 1.5 ...
 $ Petal.Width : num 0.2 0.2 0.2 0.2 0.2 0.4 0.3 0.2 0.2 0.1 ...
 $ Species     : Factor w/ 3 levels "setosa", "versicolor", ...: 1 1 1 1 1 1 1 1 1 1 ...

```



```
> str(iris)
'data.frame': 150 obs. of 5 variables:
Formal class 'ore.frame' [package "OREbase"] with 12 slots
 ..@ .Data : list()
 ..@ dataQry : Named chr "( select /*+ no_merge(t) */ \"Sepal.Length\" VAL001, \"Sepal.Width\" VAL002, \"Petal.Length\" VAL003, \"Petal.Width\" VAL004, \"Species\" VAL005 from \"RQUSER\".\"IRIS\" t )"
 ..@ sqlName : chr
 ..@ sqlValue : chr  "\"Sepal.Length\"\" \"Sepal.Width\"\" \"Petal.Length\"\" \"Petal.Width\"\" ...
 ..@ sqlTable : chr "\"RQUSER\".\"IRIS\""
 ..@ sqlPred : chr ""
 ..@ extRef : list()
 ..@ names : chr
 ..@ row.names: int
 ..@ s3class : chr "data.frame"
```

Manipulating Data – SQL equivalent

R

SQL

Column selection

```
df <- ONTIME_S[,c("YEAR", "DEST", "ARRDELAY")]
head(df)
head(ONTIME_S[,c(1,4,23)])
head(ONTIME_S[,-(1:22)])
```

Row selection

```
df1 <- df[df$DEST=="SFO",]
df2 <- df[df$DEST=="SFO",c(1,3)]
df3 <- df[df$DEST=="SFO" | df$DEST=="BOS",1:3]
```

Benefits of ORE transparency:

In-database execution

Deferred execution

Leverage indexes, partitioning, query optimization, parallelism

Column selection

```
create view df as
select YEAR, DEST, ARRDELAY
from ONTIME_S;
```

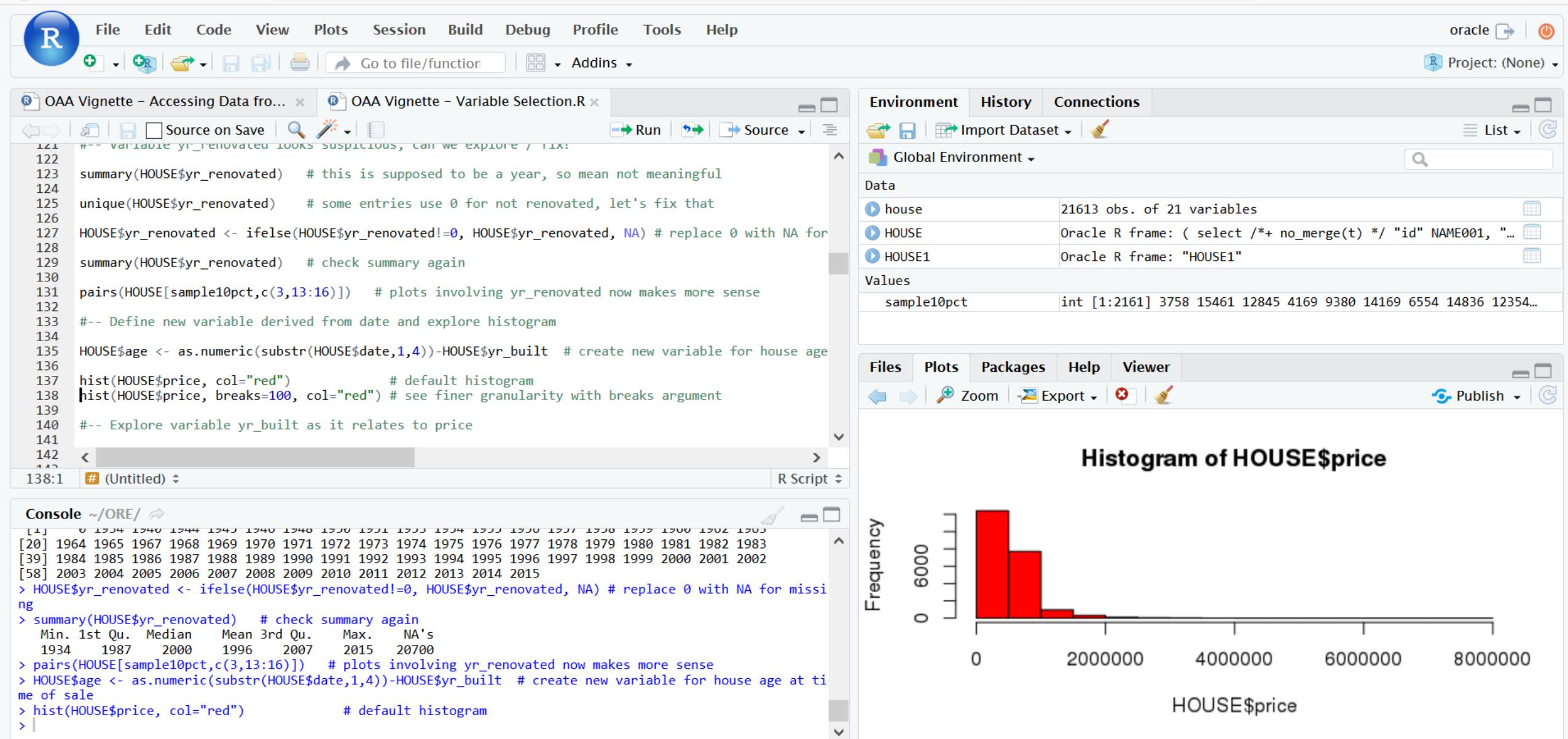
-- cannot do next two in SQL,
e.g., column selection by number & exclusion

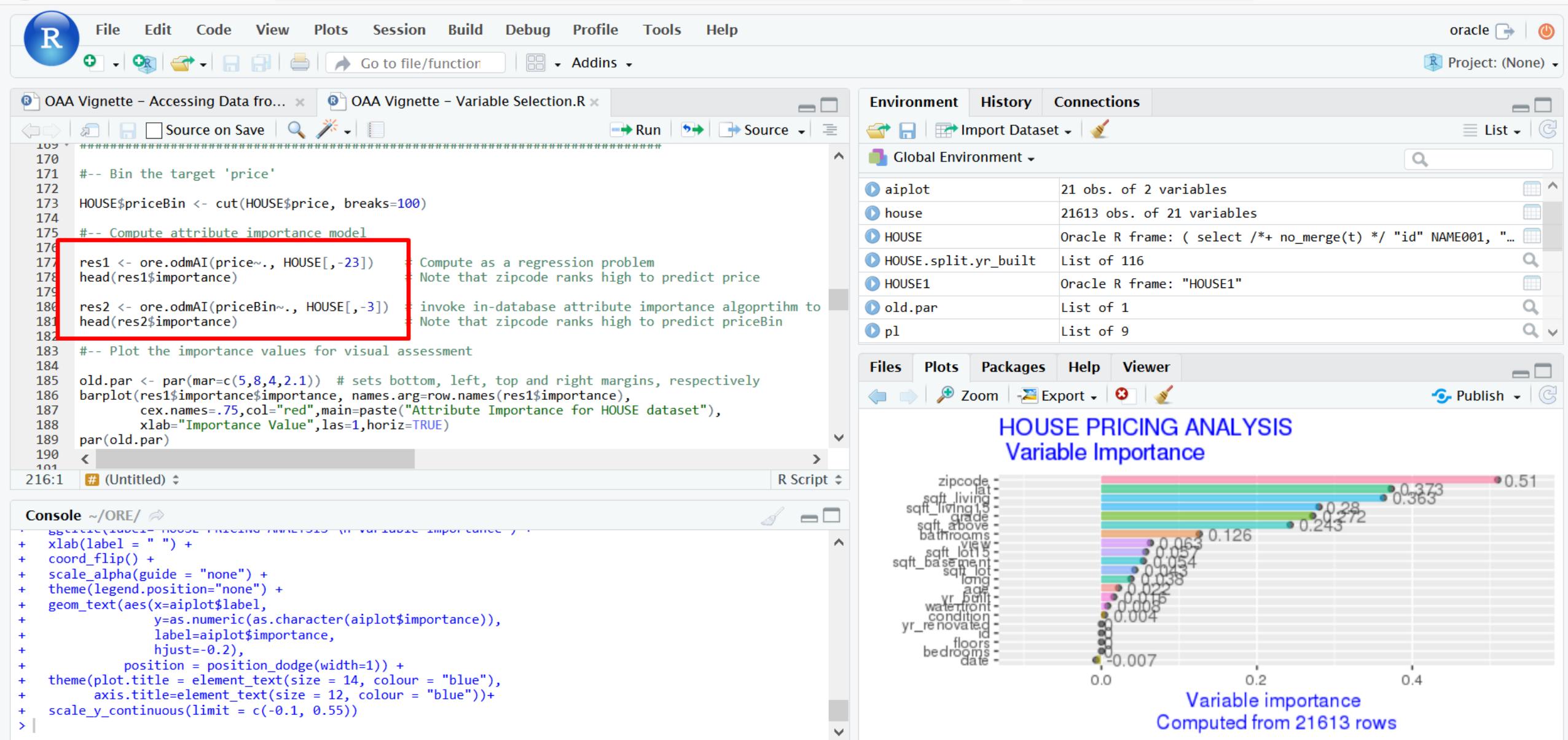
Row selection

```
create view df1 as
select * from df where DEST='SFO' ;
```

```
create view df2 as
select YEAR, ARRDELAY from df where DEST='SFO'
```

```
create view df3 as
select YEAR, DEST, ARRDELAY from df
where DEST='SFO' or DEST='BOS'
```





Automatic Data Preparation

Automatically performs transformations required by each algorithm

Supported transformations

- **Binning**: applies a supervised transformation to numeric data to generate categorical bins
- **Normalization**: normalizes numeric data to fit required range, e.g., 0..1
- **Missing value treatment**: algorithm-dependent, replaces *missing at random* numeric values with the mean and categorical values with the mode, or handles missing at random data natively
- **Outlier treatment**: removes values that deviate significantly from most other values in the column, which can affect normalization and binning

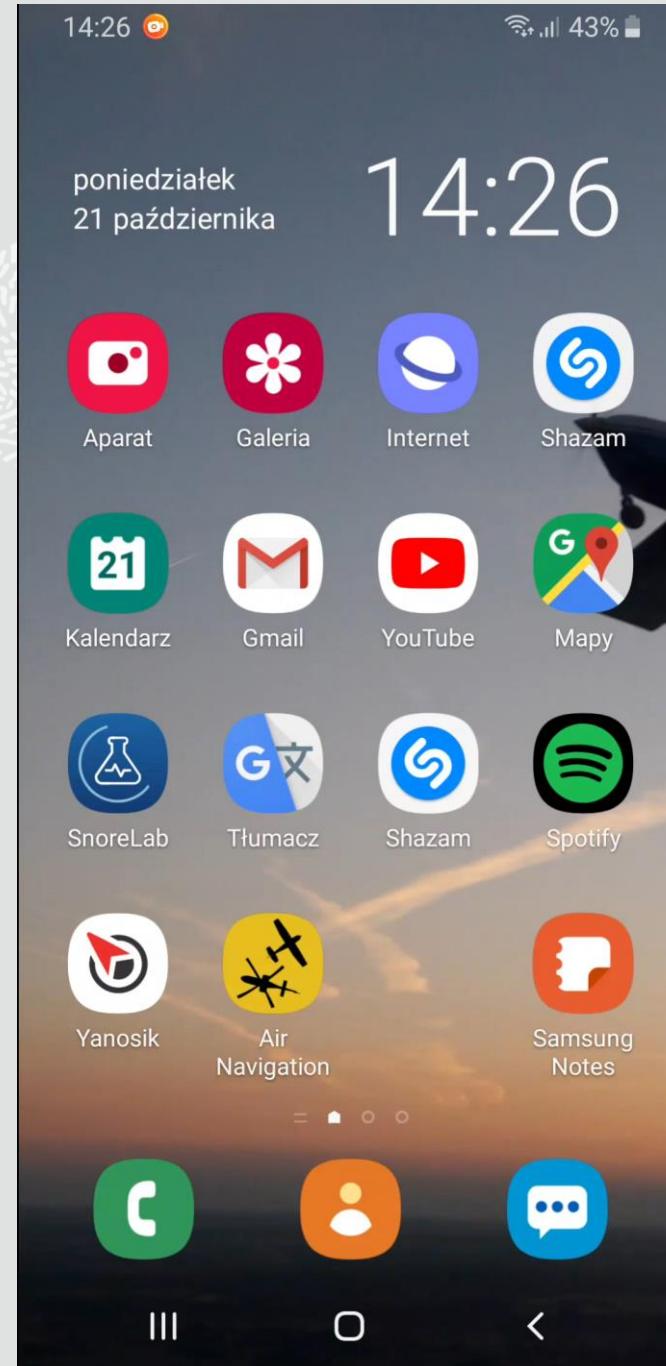
Transformation “instructions” embedded in model for automatic application during scoring

Can turn off automatic data preparation if user needs more control over preparation stages

Oracle Data Visualization and ML

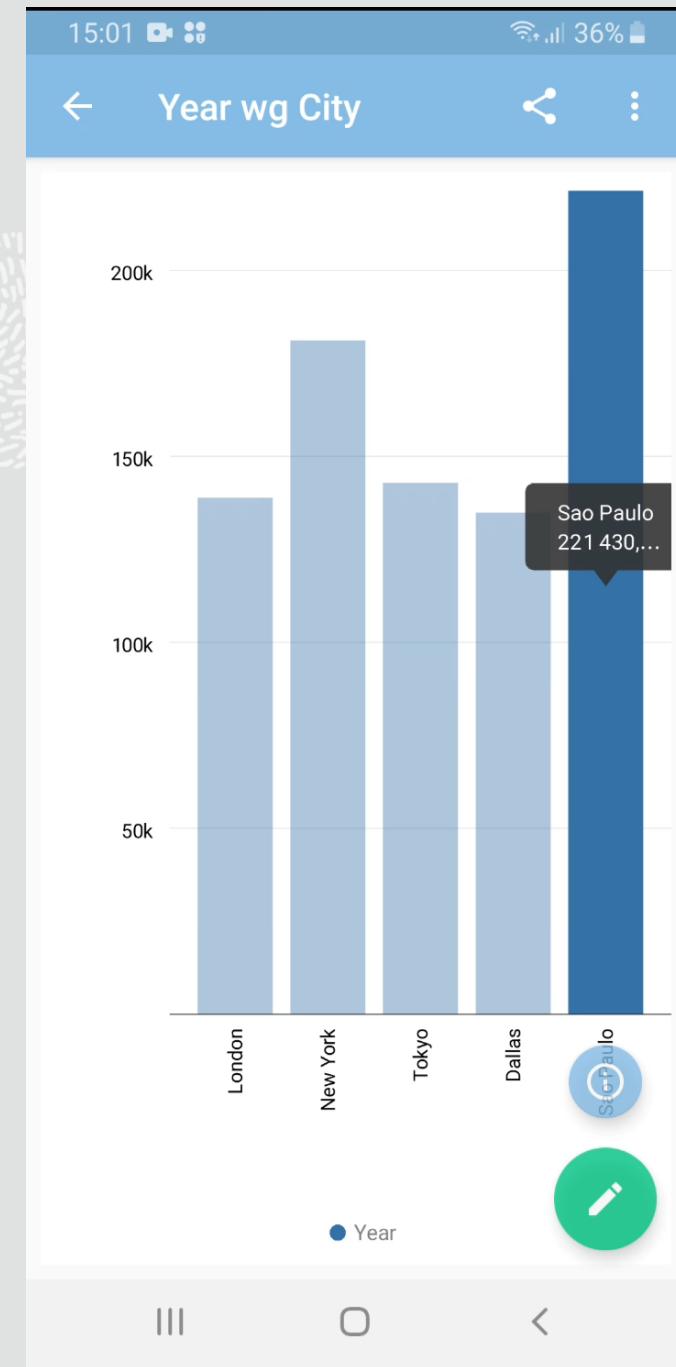
Oracle Synopsis

- Smart summaries
- Automatic chart generation
- Ability to filter, sort, and zoom in on data
- Fingerprint lock
- 100% on-device analytics
- Works with iOS and Android devices
- Works with Excel and CSV files



Oracle Synopsis

- Smart summaries
- Automatic chart generation
- Ability to filter, sort, and zoom in on data
- Fingerprint lock
- 100% on-device analytics
- Works with iOS and Android devices
- Works with Excel and CSV files



6

The screenshot shows the Oracle Analytics Cloud interface. At the top, there is a navigation bar with icons for back, forward, search, and user profile. The URL in the address bar is oac910virtualappli-portalv910c22aemea-rhcnhetb.srv.ravcloud.com. The page title is "Bez tytułu". The top menu bar includes "Przygotowanie", "Wizualizacja" (which is underlined, indicating it is the active tab), "Narracja", and "Zapisz".

The main content area is titled "Dane" (Data) and shows a hierarchical tree structure under "T&E2". The "Month" node is expanded, showing sub-options: "Rok", "Kwartał", "Kwartał roku", "Miesiąc", "Miesiąc roku" (which is highlighted with a yellow circle and a cursor), "Tydzień", "Tydzień roku", "Dzień tygodnia", "Dzień", and "Dzień roku".

Below the tree, there is a text box: "Aby dodać filtr, proszę tutaj kliknąć albo przeciągnąć dane".

The main workspace is divided into three sections:

- Left Section:** Text: "Aby wyświetlić szczegóły, proszę wybrać wizualizację".
- Center Section:** Text: "Proszę tutaj przeciągnąć wizualizacje lub dane".
- Right Section:** Text: "Aby dodać filtr, proszę tutaj kliknąć albo przeciągnąć dane".

At the bottom of the workspace, there is a footer with the text "Wyjaśnienie: Out-of-Policy Expense" and "Kanwa 2".

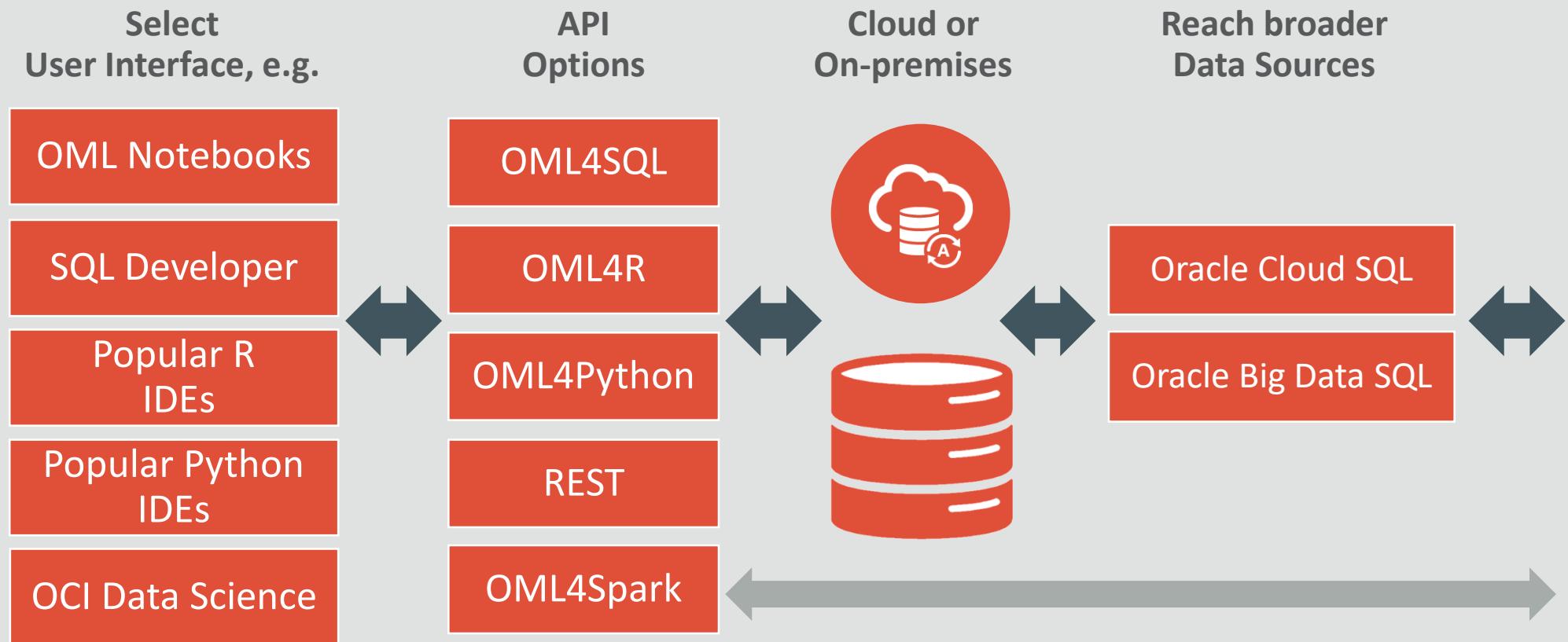
On the far right, there are several small icons for navigation and sharing.

Podsumowanie



Cross-Platform Machine Learning

Multiple user interfaces and APIs
Deployed in cloud and on-premises
From database to entire data management ecosystem



Data Lake

Oracle Database

Oracle Object Storage

Amazon S3

Azure Blob Storage

NoSQL Databases

Kafka Streams

Big Data Service (HDFS)

Dziękuję

—