Take Your Analytics to the Next Level of Insight using Machine Learning in the Oracle Autonomous Data Warehouse

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

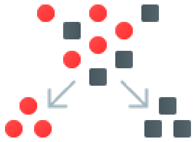
ORACLE

Safe Harbor Statement

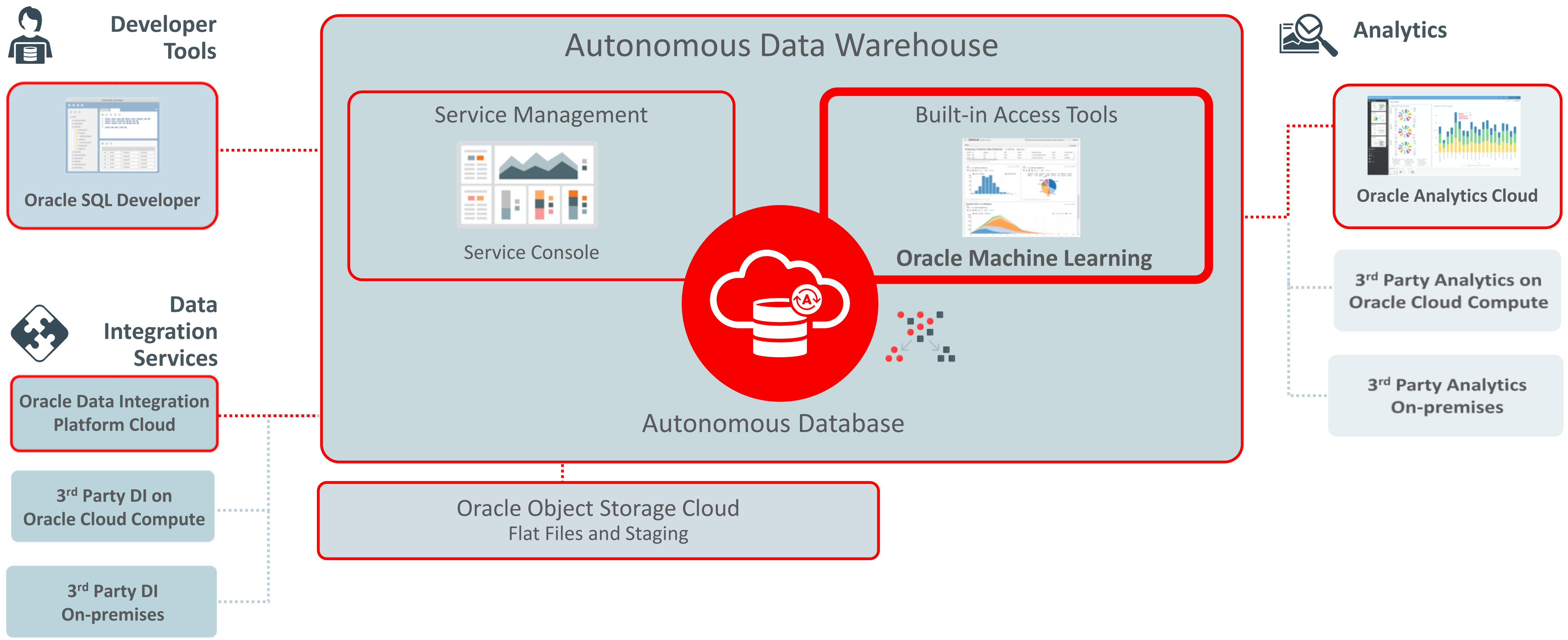
The following is intended to outline our general product direction. It is intended for information purposes only, and may not be incorporated into any contract. It is not a commitment to deliver any material, code, or functionality, and should not be relied upon in making purchasing decisions. The development,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

Autonomous Data Warehouse Cloud

- Easy
  - Fully-managed, pre-configured and optimized for DW workloads
  - Simply load data and run
    - No need to define indexes, create partitions, etc.
- Fast
  - Based on Exadata technology
- Elastic
  - Instant scaling of compute or storage with no downtime
- Powerful: Machine Learning included
  - Library of ML algorithms implemented as fully parallelized SQL functions



ADW + Oracle Machine Learning Notebooks



Powerful In-Database Machine Learning and Analytics

1 Extract value using SQL analytics

Hierarchical Analytics	Summary & Descriptive Statistics	SQL Windowing Functions	Tests for Statistical Correlations
Approximate Analytics	Pattern Matching	SQL Models	Advanced Aggregations
Ranking	Pivoting	Used-Defined PTFs	Text Analytics

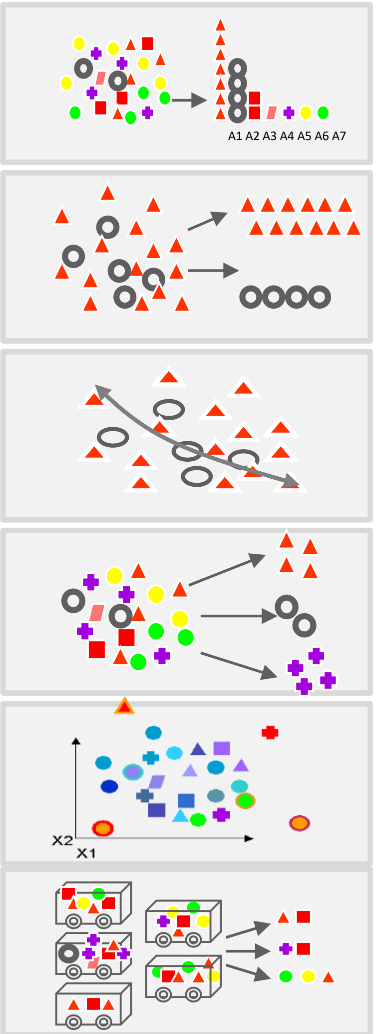
2 Gain insights, make predictions via ML

Classification / Prediction	Regression	Anomaly Detection	Attribute Importance
Association Rules / Market Basket Analysis	Clustering	Feature Extraction / Selection	Time Series / Forecasting
Cognitive Text Analytics	Ensemble Models	Predictive Queries	Text Mining

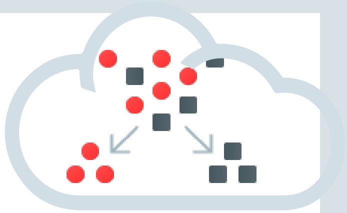
What is Machine Learning?

Automatically sift through large amounts of data to find hidden patterns, discover 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)*



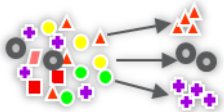
ADWC Machine Learning Algorithms



Classification

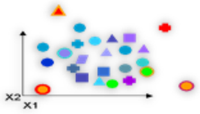
– Naïve Bayes
– Logistic Regression (GLM)
– Decision Tree
– Random Forest
– Neural Network
– Support Vector Machine
– Explicit Semantic Analysis

Clustering



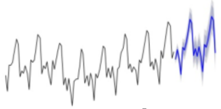
– Hierarchical K-Means
– Hierarchical O-Cluster
– Expectation Maximization (EM)

Anomaly Detection



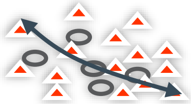
– One-Class SVM

Time Series

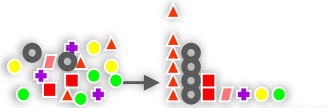


– Holt-Winters, Regular & Irregular, with and w/o trends & seasonal
– Single, Double Exp Smoothing

Regression



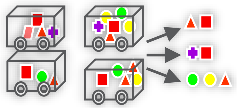
– Linear Model
– Generalized Linear Model
– Support Vector Machine (SVM)
– Stepwise Linear regression
– Neural Network



Attribute Importance

– Minimum Description Length
– Principal Comp Analysis (PCA)
– Unsupervised Pair-wise KL Div

Association Rules



– A priori/ market basket

Predictive Queries

– Predict, cluster, detect, features

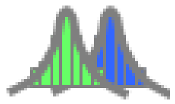
SQL Analytics

– SQL Windows, SQL Patterns, SQL Aggregates

Feature Extraction

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

Statistical Functions



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

ORACLE

•OAA includes support for Partitioned Models, Transactional, data. etc,

Oracle's Data Management and Machine Learning

Architectural Strategy

- In-Database proprietary implementations of machine learning algorithms
- Leverage strengths of the Database and adds new ML tech
  - Counting, conditional probabilities, sort, rank, partition, group-by, collections, etc.
  - Parallel execution, bitmap indexes, partitioning, aggregations, recursion w/in parallel infrastructure, IEEE float, frequent itemsets, Automatic Data Preparation (ADP), Text processing, etc.
- Focus on intelligent ML defaults, simplification & automation to enable applications
  - ADP, xforms, binning, missing values, Prediction_Details, Predictive_Queries, Model_views
- Machine learning models built via PL/SQL script; scored via SQL functions (1st class DB objects)

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

True power evident when scoring models using SQL functions, e.g.

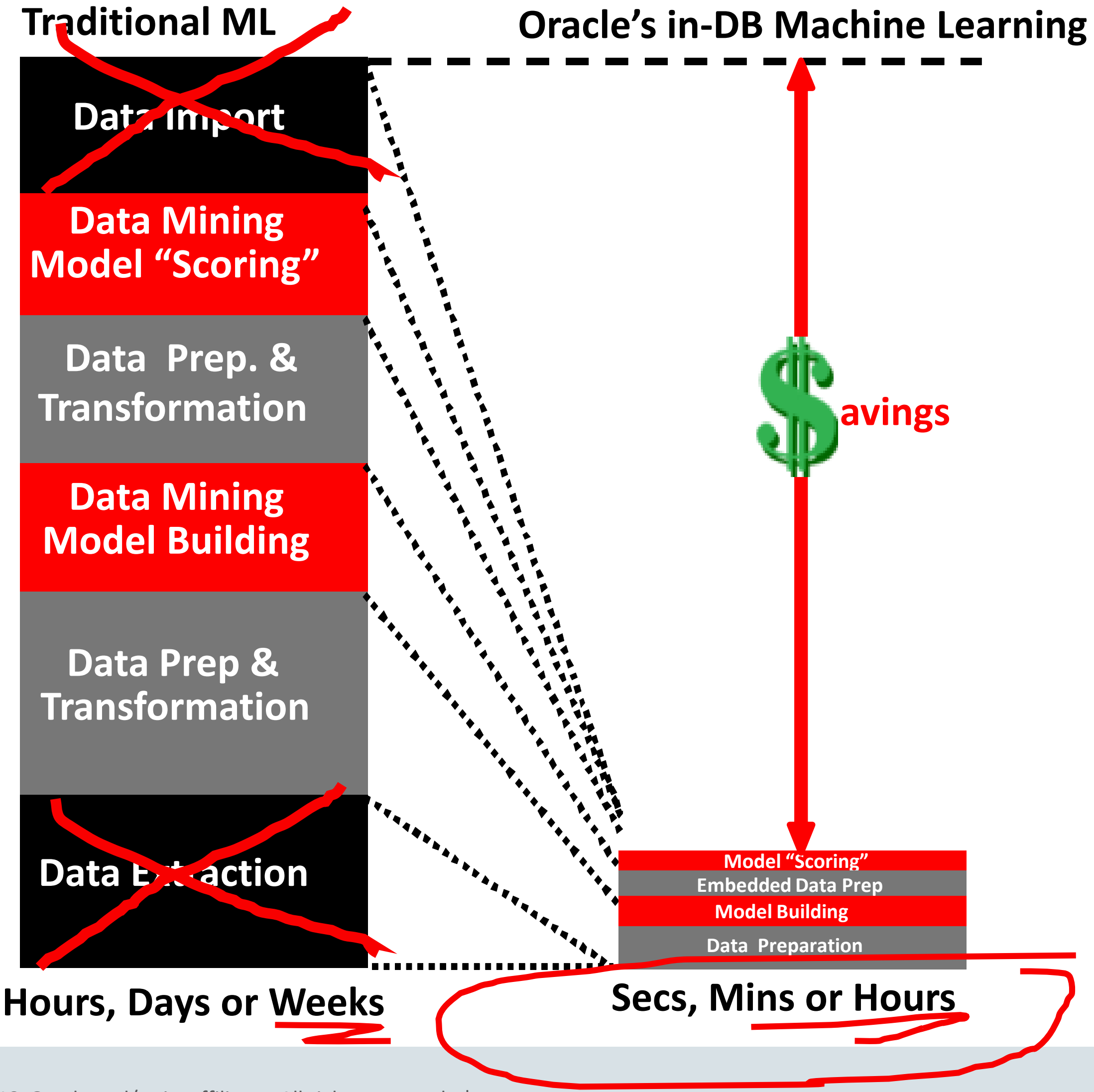
  - "Smart scan" ML model scoring "push down"; Supports OLTP and ATPC environments
- Machine Learning and Advanced Analytics are peer to rest of Oracle Data Mgmt features
  - Security, Back-up, Encryption, Scalability, "Big Data" eco-system, BDA, Big Data SQL, Cloud, Spark, etc.
  - Best ML & analytical development and deployment platform

Oracle’s Machine Learning & Advanced Analytics

Fastest Way to Deliver Enterprise-wide Predictive Analytics

Major Benefits

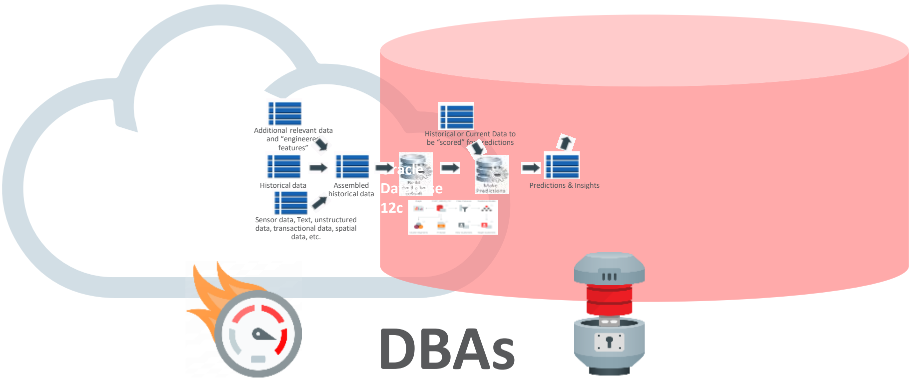
- Data remains in Database & Hadoop
  - Model building and scoring occur in-database
  - Use R packages with data-parallel invocations
- Leverage investment in Oracle IT
  - Eliminate data duplication
  - Eliminate separate analytical servers
- Deliver enterprise-wide applications
  - GUI for ML/Predictive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



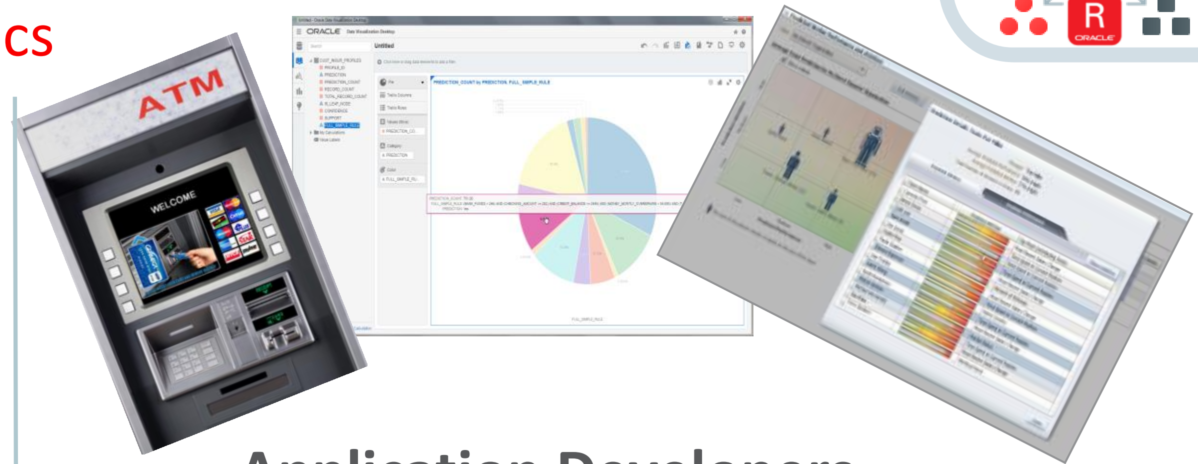
ORACLE

Multiple Data Scientist User Roles Supported

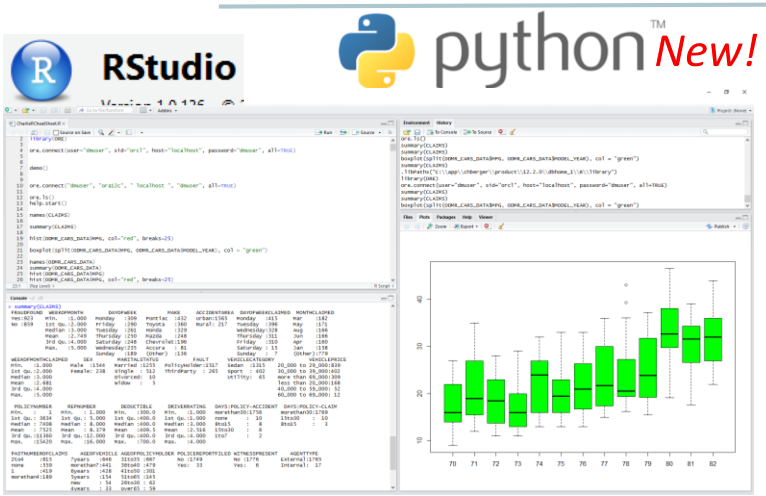
Oracle's Machine Learning/Advanced Analytics



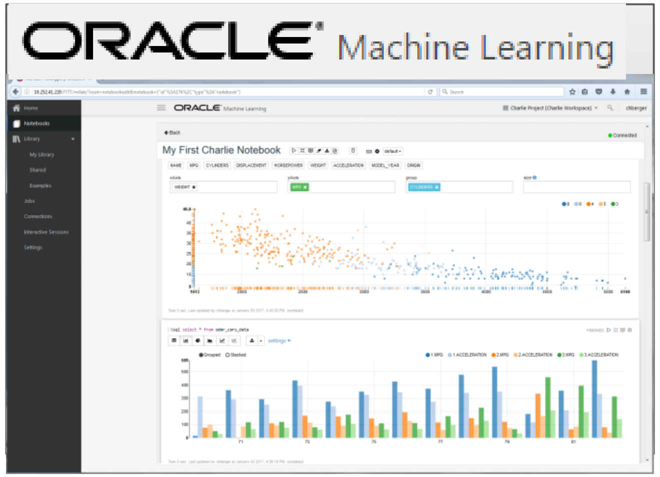
DBAs



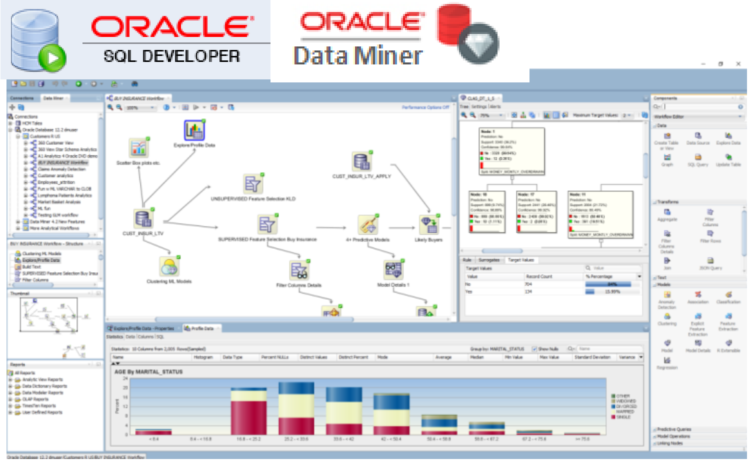
Application Developers



R, Python Users, Data Scientists



OML Notebook Users



Data Analysts, Citizen Data Scientists

OAA Model Build and Real-time SQL Apply Prediction

Simple SQL Syntax

ML Model Build (PL/SQL)

```
begin
dbms_data_mining.create_model('BUY_INSUR1', 'CLASSIFICATION',
  'CUST_INSUR_LTV', 'CUST_ID', 'BUY_INSURANCE', CUST_INSUR_LTV_SET');
end;
/
```

Model Apply (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;
```

SQL | All Rows Fetched: 1 in 0.043 seconds

	PREDICTION_PROBABILITY(BUY_INSUR1,'YES'USING3500ASBANK_FUNDS,825ASCHECKING_AMOUNT,400ASCREDIT_BALANCE
1	0.9276956709910801

ORACLE

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

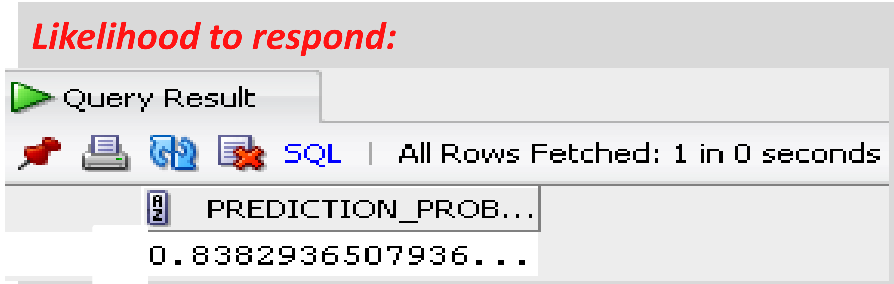
Time measure: set timing on;

ML Model Deployment for Real-Time Scoring

Real-Time Scoring, Predictions and Recommendations

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

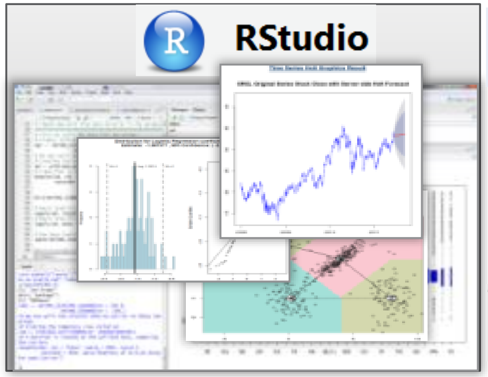
```
Select prediction_probability(CLAS_DT_1_15, 'Yes'
   USING 7800 as bank_funds, 125 as checking_amount, 20 as
   credit_balance, 55 as age, 'Married' as marital_status,
   250 as MONEY_MONTLY_OVERDRAWN, 1 as house_ownership)
from dual;
```

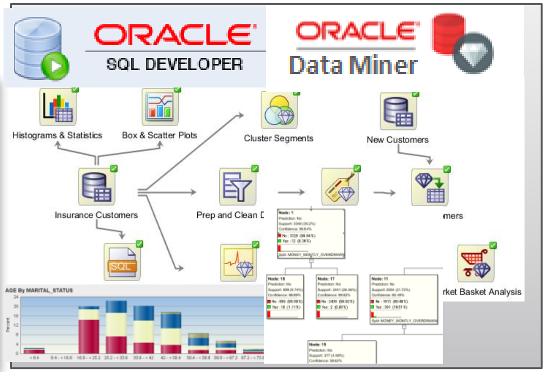


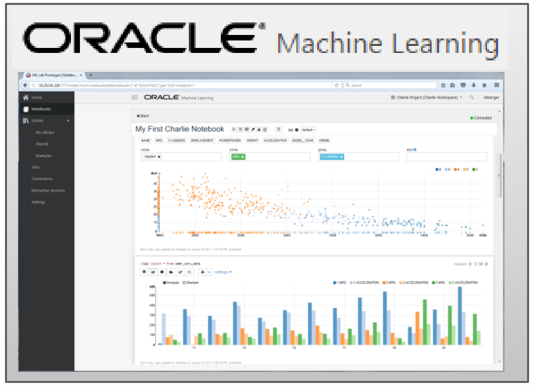
Manage and Analyze All Your Data

Data Scientists, R Users, Citizen Data Scientists

Architecturally, Many Options and Flexibility







SQL / R

Boil down the Data Lake

Big Data SQL / R

Object Store

NoSQL

ORACLE

kafka

"Engineered Features"

– Derived attributes that reflect domain knowledge—key to best models e.g.:

- Counts
- Totals
- Changes over time

ORACLE

Oracle's Data Management and Machine Learning

Market Observations:

- Machine learning, predictive analytics & "AI" now *must-have* requirements
- Separate islands for data management and data science just don't work
- Enterprises whose data science teams most rapidly extract insights and predictions win

ORACLE Conclusions:

- Must "operationalize" ML insights and predictions throughout enterprise
- Multilingual Machine Learning: SQL, R, Python, Workflow UI, Notebooks, Embed ML in Appls,
- Evolving towards <u>combined</u> Data Management + Machine Learning environment that can essentially to manage and "think" about data

Introducing:
Oracle Machine Learning SQL Notebooks

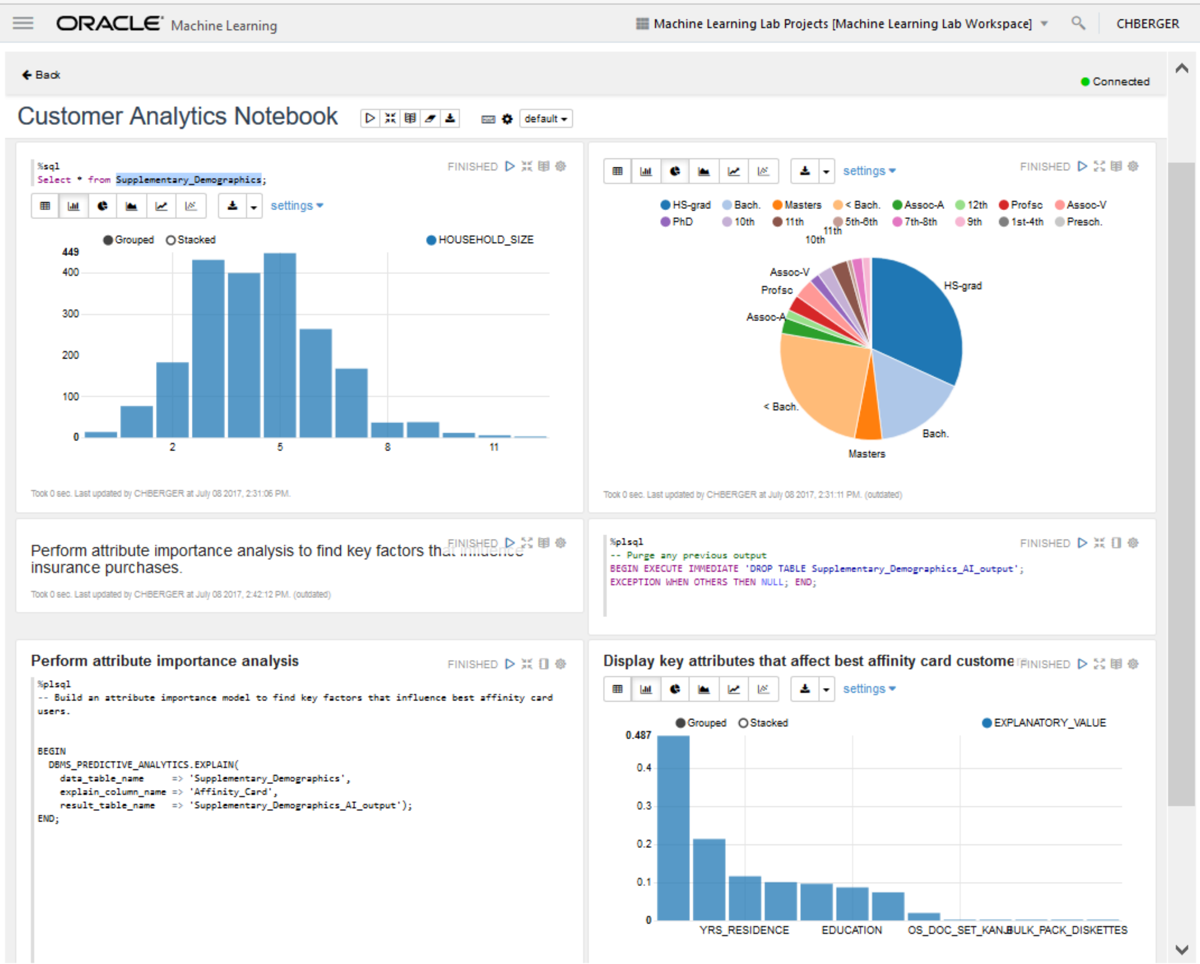
ORACLE

Oracle Machine Learning

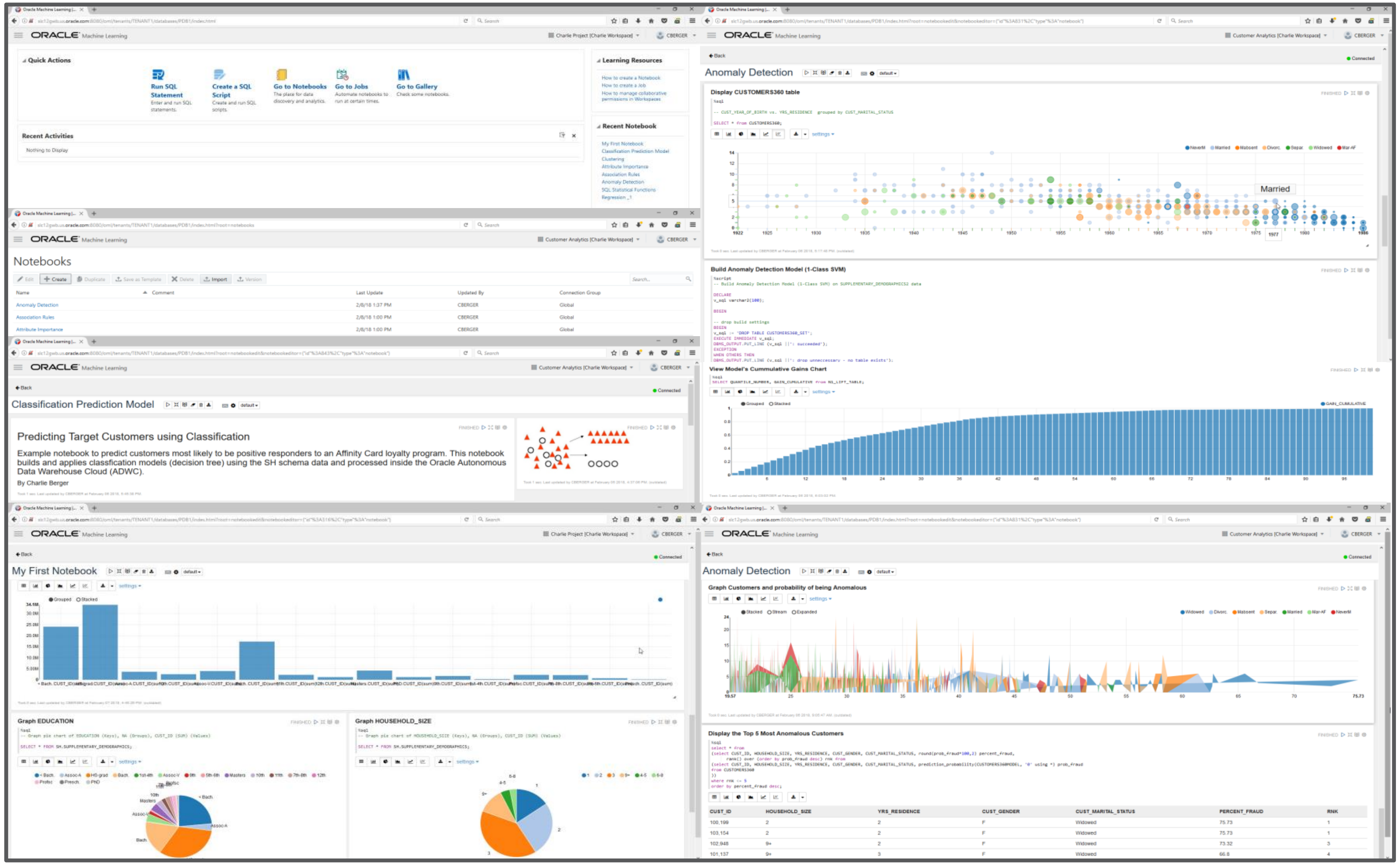
Machine Learning Notebook for Autonomous Data Warehouse Cloud

Key Features

- Collaborative UI for data scientists
  - Packaged with Autonomous Data Warehouse
  - Easy access to shared notebooks, templates, permissions, scheduler, etc.
- Supports development of ML methodologies in-ADW
  - SQL ML algorithms API
  - Predictions, churn, cross-sell, fraud, associations, statistics, correlations, forecasting, identify clusters, etc.



Oracle Machine Learning Quick DEMO



Oracle Machine Learning ...

slc12gwb.us.oracle.com:8080/omlusers/login.html?tenant=TENANT1&database=PDB1&redirect_uri=http://slc12gwb.us.oracle.com:8080/omlusers/api/oauth2/v1/login

Search

ORACLE Machine Learning

Sign In

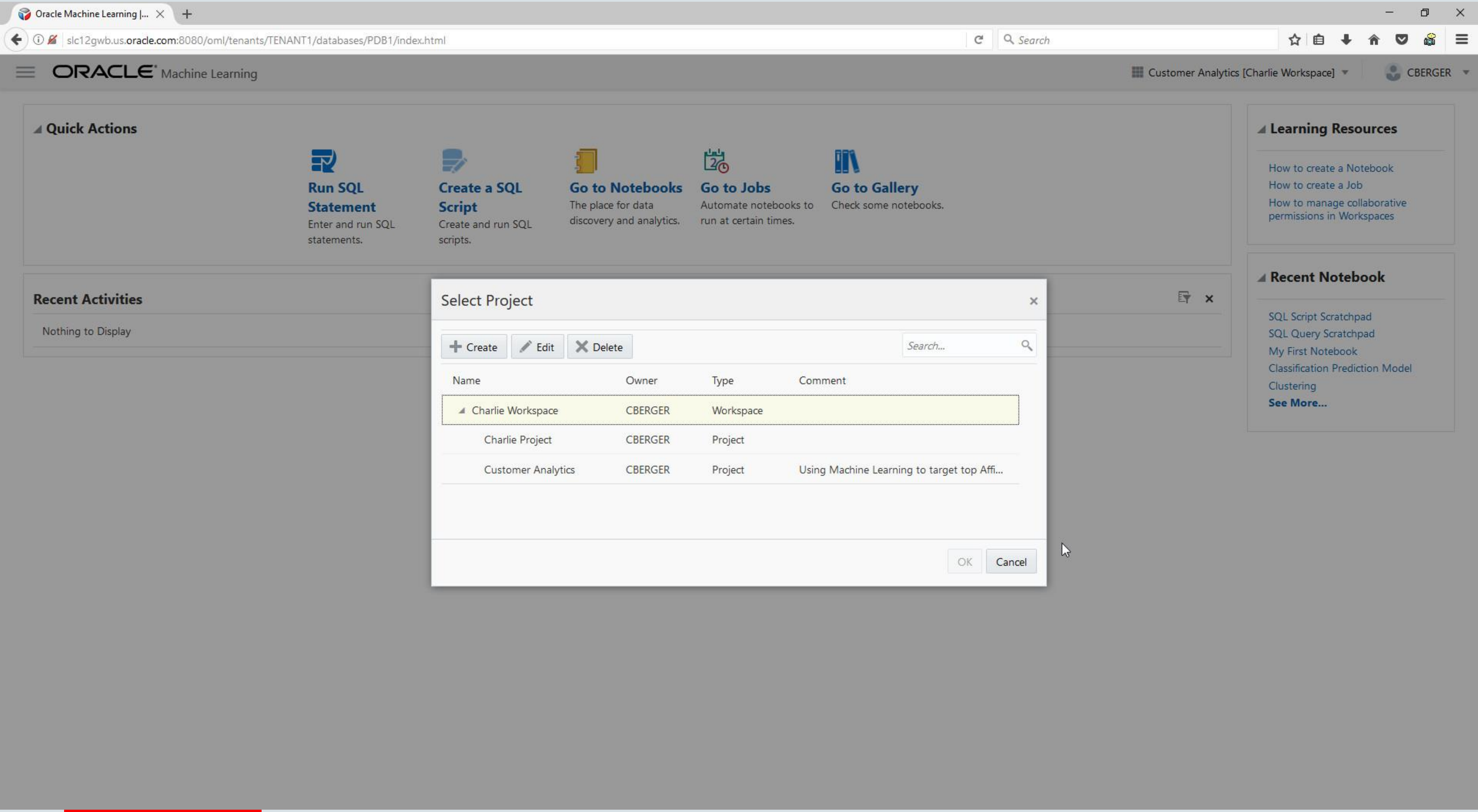
Tenant TENANT1

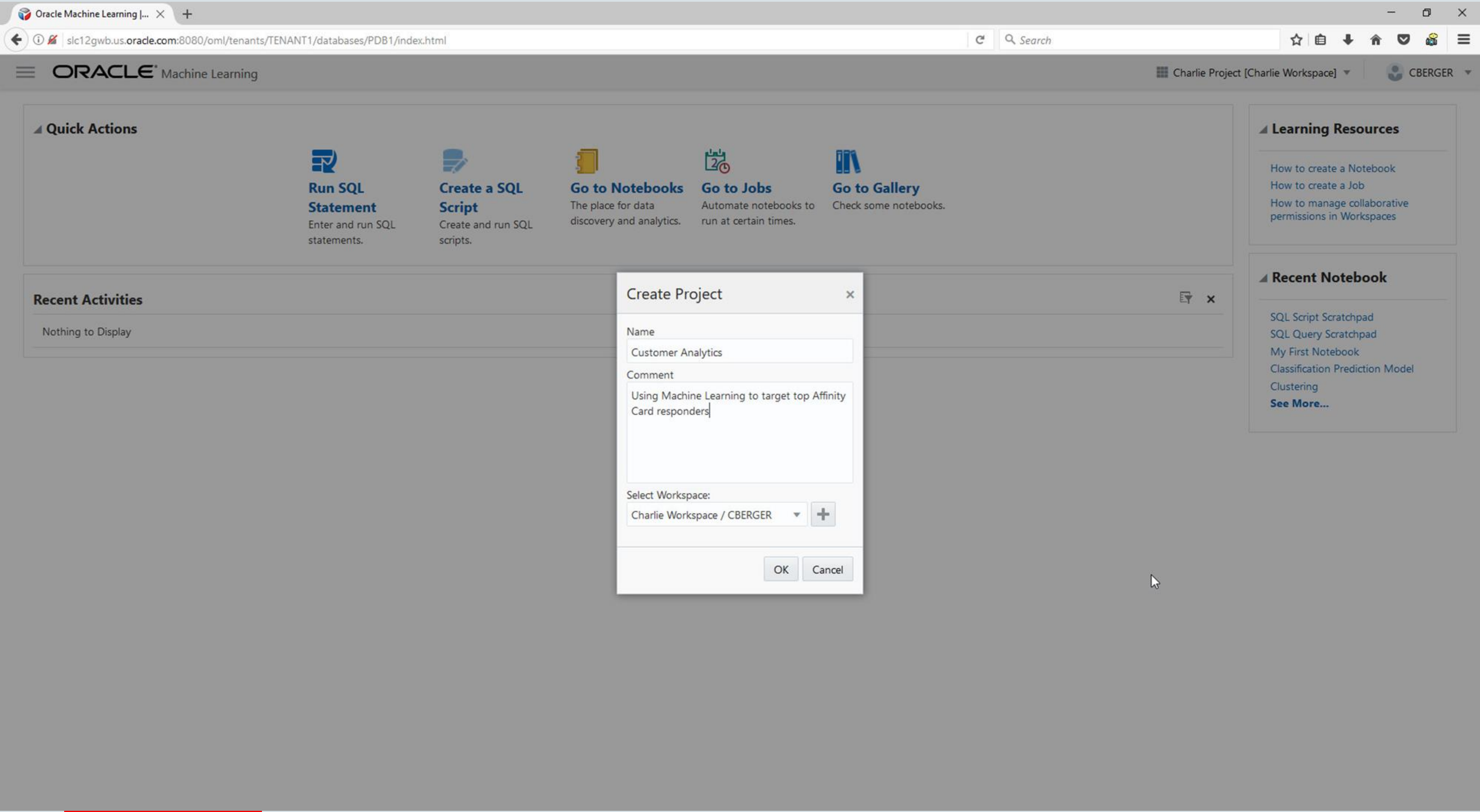
Database PDB1

* Username CBERGER

* Password •••••••••••••

Sign In


Oracle Machine Learning |...
slc12gwb.us.oracle.com:8080/oml/tenants/TENANT1/databases/PDB1/index.html
Search
ORACLE Machine Learning
Customer Analytics [Charlie Workspace]
CBERGER
Quick Actions
Run SQL Statement
Enter and run SQL statements.
Create a SQL Script
Create and run SQL scripts.
Go to Notebooks
The place for data discovery and analytics.
Go to Jobs
Automate notebooks to run at certain times.
Go to Gallery
Check some notebooks.
Learning Resources
How to create a Notebook
How to create a Job
How to manage collaborative permissions in Workspaces
Recent Activities
Nothing to Display
Recent Notebook
SQL Script Scratchpad
SQL Query Scratchpad
My First Notebook
Classification Prediction Model
Clustering
See More...
Select Project
Create
Edit
Delete
Search...
Name
Owner
Type
Comment
Charlie Workspace
CBERGER
Workspace
Charlie Project
CBERGER
Project
Customer Analytics
CBERGER
Project
Using Machine Learning to target top Affi...
OK
Cancel


Oracle Machine Learning |...
slc12gwb.us.oracle.com:8080/oml/tenants/TENANT1/databases/PDB1/index.html
Search
ORACLE Machine Learning
Charlie Project [Charlie Workspace]
CBERGER
Quick Actions
Run SQL Statement
Enter and run SQL statements.
Create a SQL Script
Create and run SQL scripts.
Go to Notebooks
The place for data discovery and analytics.
Go to Jobs
Automate notebooks to run at certain times.
Go to Gallery
Check some notebooks.
Learning Resources
How to create a Notebook
How to create a Job
How to manage collaborative permissions in Workspaces
Recent Activities
Nothing to Display
Recent Notebook
SQL Script Scratchpad
SQL Query Scratchpad
My First Notebook
Classification Prediction Model
Clustering
See More...
Create Project
Name
Customer Analytics
Comment
Using Machine Learning to target top Affinity Card responders
Select Workspace:
Charlie Workspace / CBERGER
OK
Cancel

Oracle Machine Learning |...

slc12gwb.us.oracle.com:8080/oml/tenants/TENANT1/databases/PDB1/index.html

ORACLE Machine Learning

Charlie Project [Charlie Workspace] CBERGER

Quick Actions

Run SQL Statement
Enter and run SQL statements.

Create a SQL Script
Create and run SQL scripts.

Go to Notebooks
The place for data discovery and analytics.

Go to Jobs
Automate notebooks to run at certain times.

Go to Gallery
Check some notebooks.

Recent Activities

Nothing to Display

Learning Resources

How to create a Notebook
How to create a Job
How to manage collaborative permissions in Workspaces

Recent Notebook

My First Notebook
Classification Prediction Model
Clustering
Attribute Importance
Association Rules
Anomaly Detection
SQL Statistical Functions
Regression _1

Oracle Machine Learning |...

slc12gwb.us.oracle.com:8080/oml/tenants/TENANT1/databases/PDB1/index.html

ORACLE Machine Learning

Charlie Project [Charlie Workspace]

CBERGER

Quick Actions

Run SQL Statement
Enter and run SQL statements.

Create a SQL Script
Create and run SQL scripts.

Go to Notebooks
The place for data discovery and analytics.

Go to Jobs
Automate notebooks to run at certain times.

Go to Gallery
Check some notebooks.

Learning Resources

Recent Notebook

SQL Query Scratchpad
My First Notebook
Classification Prediction Model
Clustering
Attribute Importance
See More...

Recent Activities

today

Charlie Berger **updated** SQL Query Scratchpad notebook in Charlie Project [Charlie Workspace]
2/8/18 12:14 PM

yesterday

Charlie Berger **updated** Classification Prediction Model notebook in Charlie Project [Charlie Workspace]
2/7/18 5:44 PM

last Tuesday

Charlie Berger **changed** notebook name from Statistical Functions to SQL Statistical Functions in Charlie Project [Charlie Workspace]
2/6/18 10:22 PM

Charlie Berger **changed** notebook name from SQL Statistical Functions to Statistical Functions in Charlie Project [Charlie Workspace]
2/6/18 10:17 PM

Charlie Berger **changed** notebook name from Statistical Functions to SQL Statistical Functions in Charlie Project [Charlie Workspace]
2/6/18 10:17 PM

Charlie Berger **changed** notebook name from SQL Statistical Functions to Statistical Functions in Charlie Project [Charlie Workspace]
2/6/18 10:16 PM

Oracle Machine Learning |...

slc12gwb.us.oracle.com:8080/oml/tenants/TENANT1/databases/PDB1/index.html?root=notebookedit&notebookeditor={"id"%3A817%2C"type"%3A"notebook"}

ORACLE Machine Learning

Charlie Project [Charlie Workspace] CBERGER

Back

Connected

SQL Query Scratchpad

default

```
SELECT * FROM SH.CUSTOMERS;
```

FINISHED

CUST_ID	CUST_FIRST_NAME	CUST_LAST_NAME	CUST_GENDER	CUST_YEAR_OF_BIRTH	CUST_MARITAL_STATUS	CUST_STREET_ADDRESS	CUST_POSTAL_CODE	CUST_CITY	CUST_CITY_ID	CUST_STATE_PROVINCE	CUST_STATE_PROVINCE_
37,057	Bernard	Wright	M	1,941	married	107 East Catano Avenue	66,361	Velbert	52,436	Nordrhein-Westfalen	52,684
40,612	Bernard	Wright	M	1,947		107 South Prentiss Avenue	33,866	Bergen op Zoom	51,181	Noord-Brabant	52,682
44,167	Bernard	Wright	M	1,974	single	117 North Door Avenue	83,601	San Francisco	52,289	CA	52,567
34,880	Bernard	Wright	M	1,955	married	17 North Lehigh Court	59,862	Malaga	51,894	Malaga	52,661
47,946	Bernard	Wright	M	1,937		27 West Baraga Boulevard	46,864	Lauret	51,788	Languedoc-Roussillon	52,645
1,502	Bernard	Wright	M	1,945	single	37 Mountain View Street	80,841	Wolverhampton	52,514	England - West Midlands	52,593
5,057	Bernard	Wright	M	1,947	married	37 South Catano Street	34,216	Murnau	51,934	Bayern	52,561
8,612	Bernard	Wright	M	1,939		47 West Prentiss Road	72,059	Los Angeles	51,806	CA	52,567

Took 0 sec. Last updated by CBERGER at February 08 2018, 12:13:58 PM.

READY

Oracle Machine Learning |...

slc12gwb.us.oracle.com:8080/oml/tenants/TENANT1/databases/PDB1/index.html?root=notebooks

ORACLE Machine Learning

Customer Analytics [Charlie Workspace]

CBERGER

Notebooks

Edit | Create | Duplicate | Save as Template | Delete | Import | Version

Search...

Name	Comment	Last Update	Updated By	Connection Group
Anomaly Detection		2/8/18 1:37 PM	CBERGER	Global
Association Rules		2/8/18 1:00 PM	CBERGER	Global
Attribute Importance		2/8/18 1:00 PM	CBERGER	Global
Classification Prediction Model		2/8/18 1:00 PM	CBERGER	Global
Clustering		2/8/18 12:59 PM	CBERGER	Global
My First Notebook		2/8/18 1:00 PM	CBERGER	Global
Regression _1		2/8/18 1:00 PM	CBERGER	Global
SQL Query Scratchpad		2/8/18 1:00 PM	CBERGER	Global
SQL Script Scratchpad		2/8/18 1:00 PM	CBERGER	Global
SQL Statistical Functions		2/8/18 1:00 PM	CBERGER	Global

slc12gwb.us.oracle.com:8080/oml/tenants/TENANT1/databases/PDB1/index.html?root=notebooks#

My First Notebook

default

FINISHED

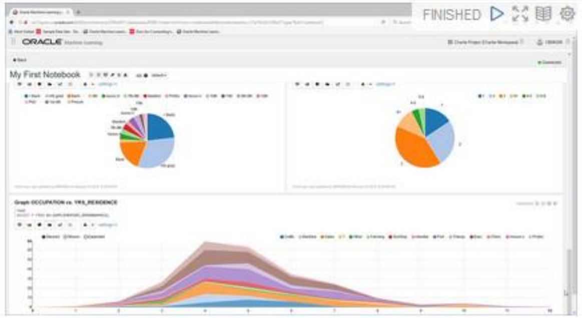
Simple Oracle Machine Learning notebook example

Oracle Machine Learning example notebook for learning basic functions using SH schema data and highlights basic data selection and data vewing using the Oracle Autonomous Data Warehouse Cloud (ADWC).

By Charlie Berger

Took 5 sec. Last updated by CBERGER at February 07 2018, 4:39:39 PM.

FINISHED



Took 0 sec. Last updated by CBERGER at February 07 2018, 4:39:39 PM.

FINISHED

For more information, check the Oracle ADWC Documentation https://docs.oracle.com/en/cloud/paas/autonomous-data-warehouse-cloud/index.html, Oracle Machine Learning folder on Oracle on Github https://github.com/oracle, Oracle Advanced Analytics http://www.oracle.com/technetwork/database/options/advanced-analytics/overview/index.html and Oracle Machine Learning on Oracle Technology Network and Introducing Oracle Machine Learning blog post https://blogs.oracle.com/datamining/introducing-oracle-machine-learning-sql-notebooks-for-the-oracle-autonomous-data-warehouse-cloud on Oracle Machine Learning blog.

Took 0 sec. Last updated by CBERGER at February 07 2018, 4:39:39 PM.

Show all tables

FINISHED

```
%sql
SELECT * FROM all_tables where owner = 'SH';
```

OWNER	TABLE_NAME	TABLESPACE_NAME	CLUSTER_NAME	IOT_NAME	STATUS	PCT_FREE	PCT_USED	INI_TRANS	MAX_TRANS	INITIAL_EXTENT	NEXT_EXTENT	MIN_EXTENTS	MAX_EXTENTS	PCT_INCREASE	FREELISTS	FREELIST_GROUPS	LOGGING	BACKED_UP
SH	SALES	SYSTEM			VALID	10	40	1	255	65,536	1,048,576	1	2,147,483,645		1	1	YES	N
SH	TIMES	SYSTEM			VALID	10	40	1	255	65,536	1,048,576	1	2,147,483,645		1	1	YES	N
SH	CHANNELS	SYSTEM			VALID	10	40	1	255	65,536	1,048,576	1	2,147,483,645		1	1	YES	N
SH	PROMOTIONS	SYSTEM			VALID	10	40	1	255	65,536	1,048,576	1	2,147,483,645		1	1	YES	N
SH	CUSTOMERS	SYSTEM			VALID	10	40	1	255	65,536	1,048,576	1	2,147,483,645		1	1	YES	N
SH	COUNTRIES	SYSTEM			VALID	10	40	1	255	65,536	1,048,576	1	2,147,483,645		1	1	YES	N
SH	SUPPLEMENTARY_DEMOGRAPHICS	SYSTEM			VALID	10	40	1	255	65,536	1,048,576	1	2,147,483,645		1	1	YES	N
SH	SALES_TRANSACTIONS_EXT	SYSTEM			VALID	10	40	1	255								YES	N

Took 9 sec. Last updated by CBERGER at February 07 2018, 4:39:43 PM. (outdated)

Display table

Oracle Machine Learning | Examp

https://adb.us-phoenix-1.oraclecloud.com/oml/tenants/OCID1.TENANCY.OC1..AAAAAAAAFCUE47PQMRF4VIGNEEBGBCMM...

ORACLE Machine Learning

Charlie's OML Projects [Charlie Workspace]

CHARLIE

Example Templates

New Notebook

Anomaly Detection

This notebook shows how to detect...

Author:

Date Added: 5/4/18 6:59 AM

Tags: 'Anomaly Detection' 'Machine...

1 Likes 45 2

Association Rules

Notebook to show the use of Assoc...

Author:

Date Added: 5/4/18 6:59 AM

Tags: 'SQL' 'Associations' 'Rules' 'M...

1 Likes 13 4

Attribute Importance

Notebook to identify key attributes...

Author:

Date Added: 5/4/18 6:59 AM

Tags: 'SQL' 'Attribute Importance' 'K...

0 Likes 16 3

Classification Prediction M...

Example notebook to predict custo...

Author:

Date Added: 5/4/18 6:59 AM

Tags: 'Classification' 'Prediction' 'De...

0 Likes 16 2

Clustering

This notebook shows how to identi...

Author:

Date Added: 5/4/18 6:59 AM

Tags: 'Clustering' 'K-Means' 'Expect...

0 Likes 16 2

My First Notebook

Oracle Machine Learning example ...

Author:

Date Added: 5/4/18 6:59 AM

Tags: 'SQL' 'Data' 'Graph'

0 Likes 36 1

Regression

This notebook shows how to predic...

Author:

Date Added: 5/4/18 6:59 AM

Tags: 'Regression' 'SVM' 'GLM' 'Logi...

0 Likes 13 1

Statistical Function

Oracle Machine Learning example ...

Author:

Date Added: 5/4/18 6:59 AM

Tags: 'Statistics' 'ANOVA' 'T-test' 'F-...

0 Likes 8 1

Oracle Machine Learning |...

slc12gwb.us.oracle.com:8080/oml/tenants/TENANT1/databases/PDB1/index.html?root=notebookedit&notebookeditor={"id"%3A831%2C"type"%3A"notebook"}

ORACLE Machine Learning   Customer Analytics [Charlie Workspace]   CBERGER

← Back   Connected

Anomaly Detection

default

FINISHED

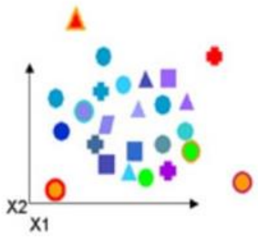
Anomaly Detection to Detect Suspicious or Rare Occurrences

This notebook shows how to detect rare records, customers or transactions using an unsupervised learning algorithm (1-Class Support Vector Machine). The notebook first builds a 1-Class SVM model and then applies the model to flag unusual or suspicious records. The anomaly detection model can also be applied to "score" new records. The entire machine learning methodology runs inside the Oracle Autonomous Data Warehouse Cloud (ADWC).

By Charlie Berger

Took 0 sec. Last updated by CBERGER at February 06 2018, 8:54:19 AM. (outdated)

FINISHED



Took 1 sec. Last updated by CBERGER at February 06 2018, 7:49:16 AM. (outdated)

FINISHED

For more information, check the Oracle ADWC Documentation https://docs.oracle.com/en/cloud/paas/autonomous-data-warehouse-cloud/index.html, Oracle Machine Learning folder on Oracle on Github https://github.com/oracle, Oracle Advanced Analytics http://www.oracle.com/technetwork/database/options/advanced-analytics/overview/index.html and Oracle Machine Learning on Oracle Technology Network and Introducing Oracle Machine Learning blog post

Took 0 sec. Last updated by CBERGER at February 06 2018, 9:25:06 AM. (outdated)

Clean up and drop any table if previously exists for notebook reproducibility

FINISHED

```
%script
BEGIN
     EXECUTE IMMEDIATE 'DROP Table SUPPLEMENTARY_DEMOGRAPHICS2';
  EXCEPTION
     WHEN OTHERS THEN NULL;
END;
```

```
PL/SQL procedure successfully completed.
```

Took 0 sec. Last updated by CBERGER at February 06 2018, 5:12:56 PM.

Create SUPPLEMENTARY_DEMOGRAPHICS2 table that remove COMMENTS unstructured data for simplicity.

FINISHED

```
%sql
CREATE Table SUPPLEMENTARY_DEMOGRAPHICS2
  AS (SELECT AFFINITY_CARD, BOOKKEEPING_APPLICATION, BULK_PACK_DISKETTES, CUST_ID, EDUCATION, FLAT_PANEL_MONITOR, HOME_THEATER_PACKAGE, HOUSEHOLD_SIZE, OCCUPATION, OS_DOC_SET_KANJI, PRINTER_SUPPLIES, YRS_RESIDENCE, Y_BOX_GAMES
     FROM SH.SUPPLEMENTARY_DEMOGRAPHICS);
```

```
Updated 4500 row(s).
```

Took 1 sec. Last updated by CBERGER at February 06 2018, 5:13:01 PM.

Classification Prediction Model

default

Predicting Target Customers using Classification

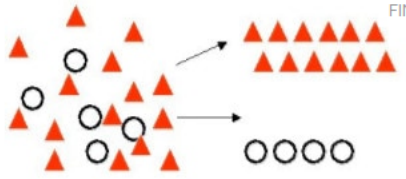
FINISHED

Example notebook to predict customers most likely to be positive responders to an Affinity Card loyalty program. This notebook builds and applies classfication models (decision tree) using the SH schema data and processed inside the Oracle Autonomous Data Warehouse Cloud (ADWC).

By Charlie Berger

Took 1 sec. Last updated by CBERGER at February 06 2018, 5:45:38 PM.

FINISHED



Took 1 sec. Last updated by CBERGER at February 05 2018, 4:37:06 PM. (outdated)

FINISHED

For more information, check the Oracle ADWC Documentation https://docs.oracle.com/en/cloud/paas/autonomous-data-warehouse-cloud/index.html, Oracle Machine Learning folder on Oracle on Github https://github.com/oracle, Oracle Advanced Analytics http://www.oracle.com/technetwork/database/options/advanced-analytics/overview/index.html and Oracle Machine Learning on Oracle Technology Network and Introducing Oracle Machine Learning blog post https://blogs.oracle.com/datamining/introducing-oracle-machine-learning-sql-notebooks-for-the-oracle-autonomous-data-warehouse-cloud on Oracle Machine Learning blog.

Took 1 sec. Last updated by CBERGER at February 03 2018, 9:48:09 PM. (outdated)

Display the SH.SUPPLEMENTARY_DEMOGRAPHICS data

FINISHED

```
%sql
Select * from SH.SUPPLEMENTARY_DEMOGRAPHICS;
```

CUST_ID	EDUCATION	OCCUPATION	HOUSEHOLD_SIZE	YRS_RESIDENCE	AFFINITY_CARD	BULK_PACK_DISKETTES	FLAT_PANEL_MONITOR	HOME_THEATER_PACKAGE	BOOKKEEPING_APPLICATION	PRINTER_SUPPLIES	Y_BOX_GAMES	OS_DOC_SET_KANJI	CC
100,931	< Bach.	Crafts	3	6	1	1	1	1	1	1	0	0	Sh
100,932	HS-grad	Machine	1	5	0	0	0	0	1	1	0	0	I ar
100,933	HS-grad	Sales	3	7	1	1	1	1	1	1	0	0	Aff
100,934	Bach.	?	2	4	0	0	0	0	1	1	1	0	Aff
100,935	< Bach.	Other	2	2	0	0	0	0	1	1	1	0	Gr
100,936	9th	Crafts	3	3	0	0	0	1	0	1	0	0	I ar
100,937	HS-grad	Crafts	2	4	0	1	1	0	1	1	1	0	Aff
100,938	HS-grad	Farming	1	1	0	1	1	1	1	1	0	0	I pı

Oracle Machine Learning × Oracle Machine Learning × Charlie

https://adwc.uscom-west-1.oraclecloud.com/oml/tenants/OCID1.TENANCY.OC1..AAAAAAAAFCUE47PQMRF4VIGNEEBGBCMMOY5R7XVOYPICJQQGE32...

ORACLE Machine Learning   Charlie Project [Charlie Workspace]   CHARLIE

← Back   Connected

Credit Score Predictions Simplified ...

default

FINISHED

Targeting Likely Good Credit Customers using Oracle Machine Learning's (OML) Classification Models

Heather has spent most of her time over the past couple of years extracting and preparing data for analysis. The large volumes of data need extracting and processing mean she spends most of her time waiting for jobs to finish and very little of her time analyzing the data. Demands from marketing are forcing a new approach whereby the data remains in the data warehouse and is processed there. The alternative cloud solution is more complex, and has no direct out of the box processes to analyze the data in place. She started taking a look at Oracle, and found the simple SQL commands in ADWC are familiar, and execute extremely fast, leveraging all the performance features of the platform. Further once she is done can can apply the learning models to incoming data on the fly, and allow end user analysts to immediately see mining results. This drastically reduces the cycle of data preparation, analysis, and publishing. It also means there is no change to analysis/reporting Data Visualization toolset that users are familiar with.

Scroll down this notebook and learn how to use OML to create predictive perspectives on data in ADWC, WITHOUT moving it. We will process a small 100k data set, but could use a 100M or billion row data set without worrying about processing time.

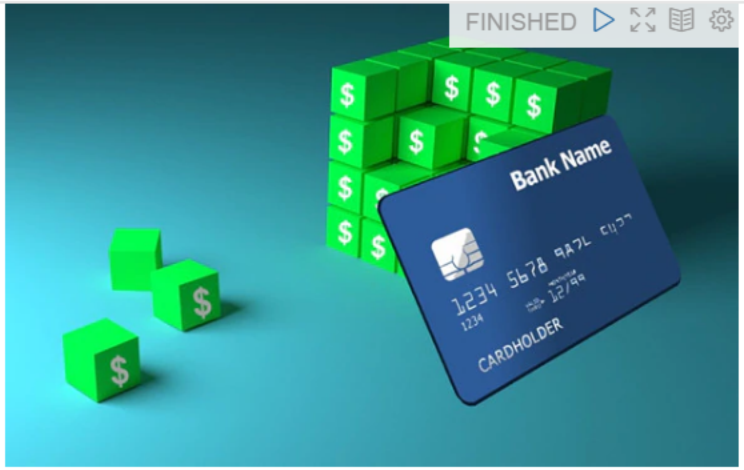
This is an extract of Alphaoffice customer information. We will first get acclimated to Apache Zepplin, the open source interface for interactive collaboration in a team environment.

The Business Problem:

Increase Sales by Targeting our Best Customers; Good Credit Customers!

Heather has a hunch that weakening sales may be due to the company selling to non-optimal customers; customers who

FINISHED



Took 0 sec. Last updated by CHARLIE at July 16 2018, 12:58:40 PM. (outdated)

Oracle Machine Learning | Oracle Machine Learning

https://adwc.uscom-west-1.oraclecloud.com/oml/tenants/OCID1.TENANCY.OC1..AAAAAAAAFCUE47PQMRF4VIGNEEBGBCMMOY5R7XVOYPICJQQGE32...

ORACLE Machine Learning   Charlie Project [Charlie Workspace]   CHARLIE

Back   Connected

Credit Score Predictions Simplified ...

default

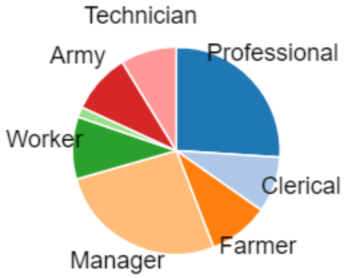
STEP 6: Review Data by Occupation

FINISHED

```sql
%sql
-- This shows an alternative presentation style - a pie chart.  Note that Zepplin visualizations are limited.  In lab 400 we will use Oracle Data Visualization to create more more interesting perspectives.

select customer_id, age, income, tenure, loan_type, loan_amount, occupation, marital_status
from credit_scoring_100k_v where rownum < 1000
```

settings

Professional Clerical Farmer Manager Worker NaN Army Technician



Took 0 sec. Last updated by CHARLIE at July 30 2018, 3:41:48 PM.

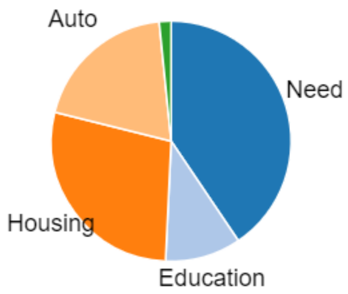
Another Pie Chart Example

FINISHED

```sql
%sql
-- This shows an alternative presentation style - a pie chart.  Note that Zepplin visualizations are limited.  In lab 400 we will use Oracle Data Visualization to create more more interesting perspectives.

select customer_id, age, income, tenure, loan_type, loan_amount, occupation, marital_status
from credit_scoring_100k_v where rownum < 1000
```

settings

Need Education Housing Auto NaN



Oracle Machine Learning × Oracle Machine Learning ×

https://adwc.uscom-west-1.oraclecloud.com/oml/tenants/OCID1.TENANCY.OC1..AAAAAAAAFCUE47PQMRF4VIGNEEBGBCMMOY5R7XVOYPICJQQGE32...

ORACLE Machine Learning Charlie Project [Charlie Workspace] CHARLIE

Back

Connected

Credit Score Predictions Simplified ...

default

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

FINISHED

Create Attribute Importance Machine Learning Model for Good Credit Customers

FINISHED

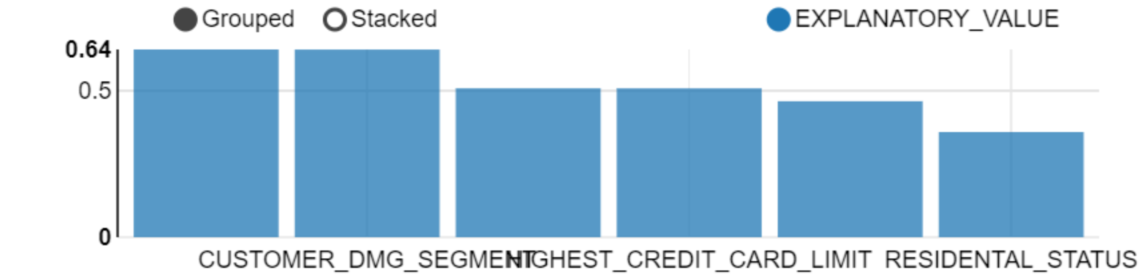
```
%script

-- 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(
    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

FINISHED

```
%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



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

Credit Score Predictions Simplified ...

default

STEP 11: Create Predictive Model to Target Good Credit Customers

FINISHED

Now that Heather has found the key attributes that most influence findig more Good Credit customers and also making better Maximum Credit Card Amount decisions, she wants to leverage Oracle Machine Learning's powerful in-Database, parallelized algorithms to build predictive models that help her company to better target "the right customers" with the "right offers".

Lets' quickly review the machine learning process:

Problem Definition: Target Good Credit Customers

Data Gathering and Preparation: We've assembled 100K records with 100+ variable about each customer and have created a target field (Good Customer/Other Customer) so we can use OML's Supervised Algorithms, specifically let's start by using a decision tree algorithm.

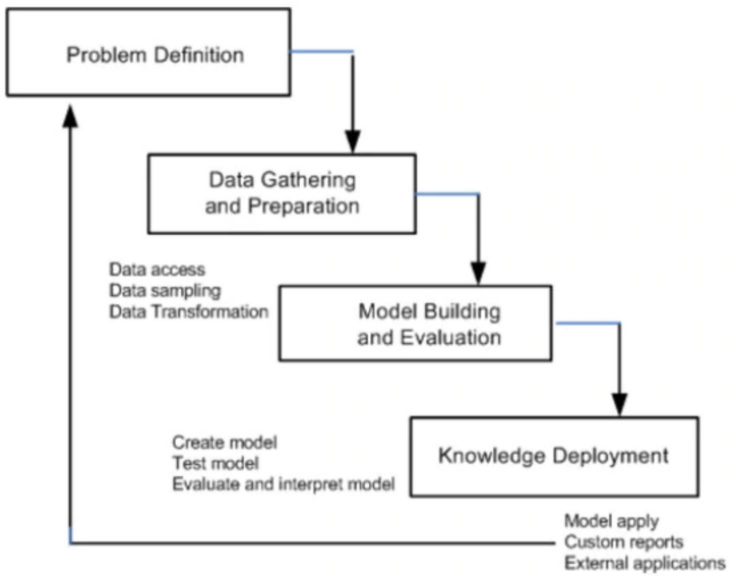
Model Building and Evaluation: We'll create a randomly selected sample from our Credit_Scoring_100k historical data and use 60% as training data for the machine learning model building phase. Then, we'll use the remaining 40% as a holdout sampple to test our model's accuracy using various model evaluation tools such as a "lift chart".

Knowledge Deployment: Once we're satisfied that we have a useful ML model that can predict with some accuracy which customers we should target (Good Credit customers), we want to apply our OML model to new customer data inside ADWC and then take a deeper look a them. Lastly, we'll jump over to Oracle Analytics Cloud for a more interactive, exploratory data analysis experience but now focusing on our customers of interest (Good Credit customers).

Took 0 sec. Last updated by CHARLIE at July 19 2018, 3:06:35 PM. (outdated)

Data Mining and Machine Learning Process

FINISHED



Took 0 sec. Last updated by CHARLIE at July 16 2018, 5:31:12 PM. (outdated)

Oracle Machine Learning × Oracle Machine Learning ×

https://adwc.uscom-west-1.oraclecloud.com/oml/tenants/OCID1.TENANCY.OC1..AAAAAAAAFCUE47PQMRF4VIGNEEBGBCMMOY5R7XVOYPICJQQGE32...

← Back

Connected

Credit Score Predictions Simplified ...

default

STEP 17: Review the CREDIT_SCORING_NEW_PREDICTIONS Table and rank Good Customers Based on Prediction Probablity, and Other Factors

FINISHED

CUSTOMER_ID	PROB_GOOD_CREDIT	AGE	INCOME	TENURE	LOAN_TYPE	LOAN_AMOUNT	OCCUPATION	EDUCATION_LEVEL	MARITAL_STATUS
41,434	98.94	54	4,250	9	Need	30,000	Manager	Bachelor's Degree	Married
49,459	98.94	27	5,250	18	Auto	40,000	Manager	Bachelor's Degree	Married

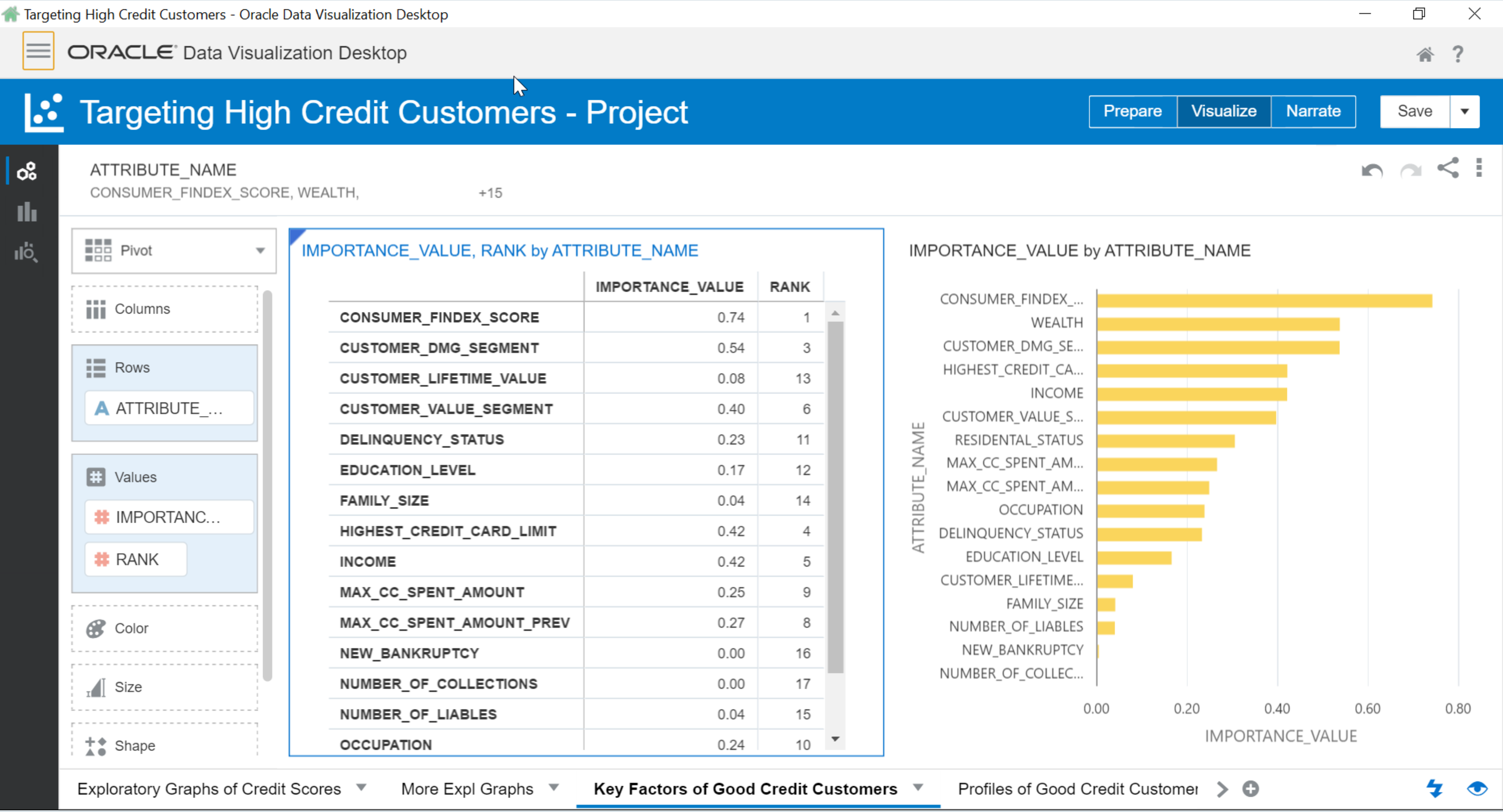
Took 1 sec. Last updated by CHARLIE at August 03 2018, 4:54:24 PM. (outdated)

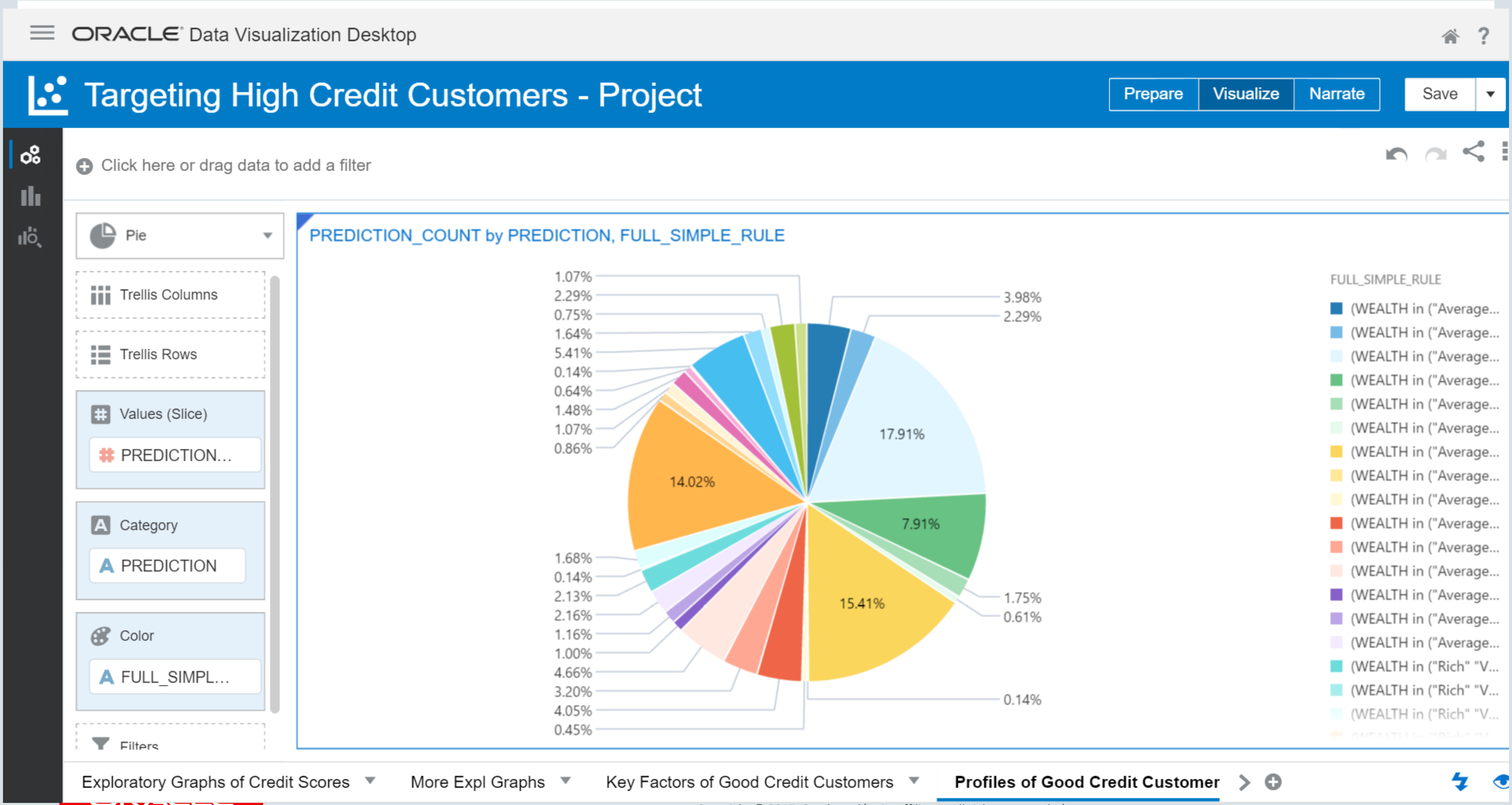
STEP 18: Apply a ML Model to a Single Record in a Transactional Application

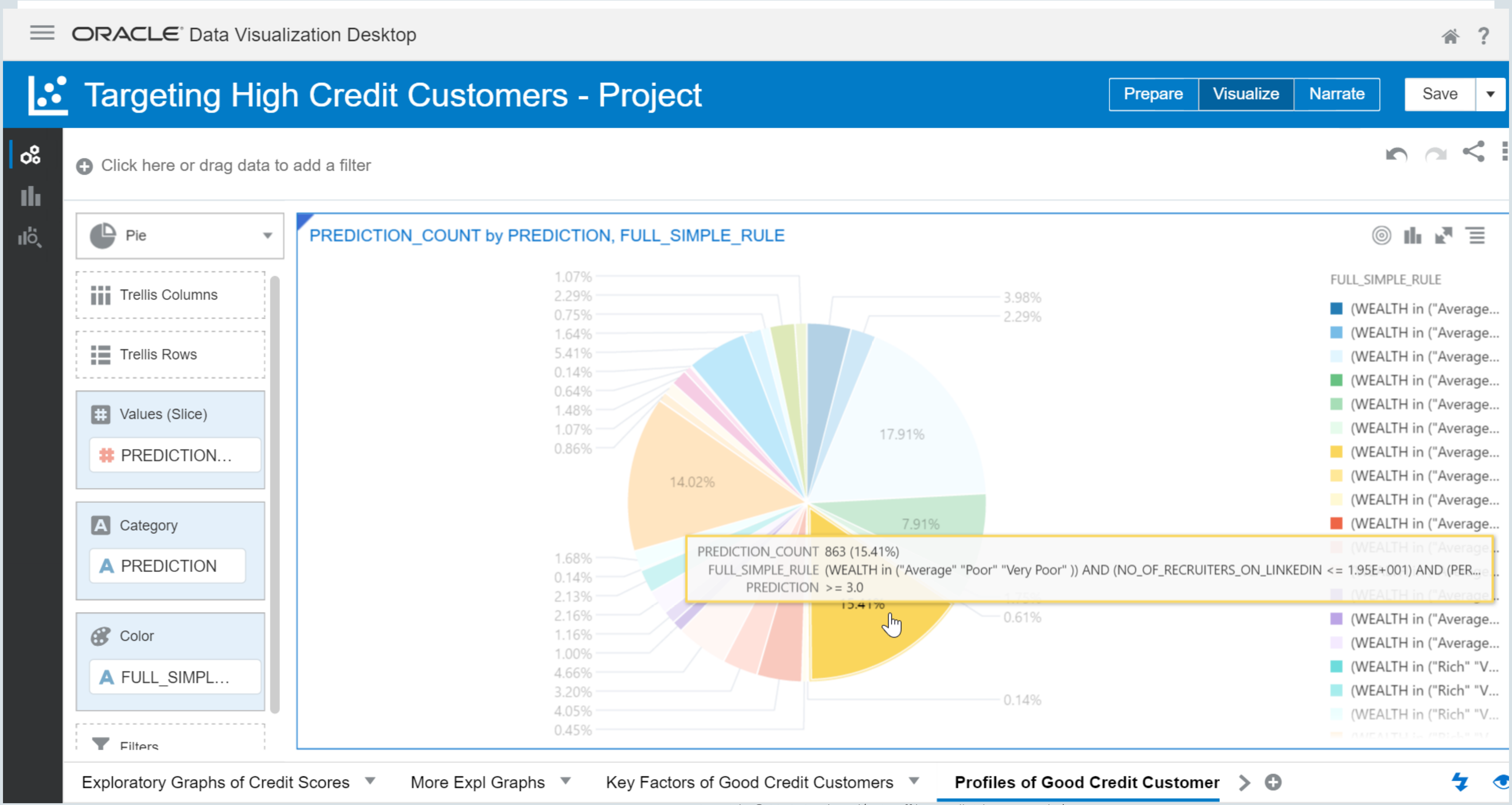
FINISHED

```
%sql

--  Try this out by running query and then change customer_value_segment from 'Gold' to 'Silver' to see lower probability of customer having Good Credit

select prediction_probability(N1_CLASS_MODEL, 'Good Credit'
  USING 'Very Rich' as WEALTH, 2000 as income, 'Silver' as customer_value_segment, 'Owner' as residential_status) Prediction_Probability
from dual;
```

PREDICTION_PROBABILITY
0.98945


Targeting High Credit Customers - Oracle Data Visualization Desktop
ORACLE Data Visualization Desktop
Targeting High Credit Customers - Project
Prepare
Visualize
Narrate
Save
ATTRIBUTE_NAME
CONSUMER_FINDEX_SCORE, WEALTH, +15
Pivot
Columns
Rows
ATTRIBUTE_...
Values
IMPORTANC...
RANK
Color
Size
Shape
IMPORTANCE_VALUE, RANK by ATTRIBUTE_NAME
IMPORTANCE_VALUE RANK
CONSUMER_FINDEX_SCORE 0.74 1
CUSTOMER_DMG_SEGMENT 0.54 3
CUSTOMER_LIFETIME_VALUE 0.08 13
CUSTOMER_VALUE_SEGMENT 0.40 6
DELINQUENCY_STATUS 0.23 11
EDUCATION_LEVEL 0.17 12
FAMILY_SIZE 0.04 14
HIGHEST_CREDIT_CARD_LIMIT 0.42 4
INCOME 0.42 5
MAX_CC_SPENT_AMOUNT 0.25 9
MAX_CC_SPENT_AMOUNT_PREV 0.27 8
NEW_BANKRUPTCY 0.00 16
NUMBER_OF_COLLECTIONS 0.00 17
NUMBER_OF_LIABLES 0.04 15
OCCUPATION 0.24 10
IMPORTANCE_VALUE by ATTRIBUTE_NAME
CONSUMER_FINDEX_...
WEALTH
CUSTOMER_DMG_SE...
HIGHEST_CREDIT_CA...
INCOME
CUSTOMER_VALUE_S...
RESIDENTAL_STATUS
MAX_CC_SPENT_AM...
MAX_CC_SPENT_AM...
OCCUPATION
DELINQUENCY_STATUS
EDUCATION_LEVEL
CUSTOMER_LIFETIME...
FAMILY_SIZE
NUMBER_OF_LIABLES
NEW_BANKRUPTCY
NUMBER_OF_COLLEC...
ATTRIBUTE_NAME
0.00
0.20
0.40
0.60
0.80
IMPORTANCE_VALUE
Exploratory Graphs of Credit Scores
More Expl Graphs
Key Factors of Good Credit Customers
Profiles of Good Credit Customer


ORACLE Data Visualization Desktop
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
PREDICTION
Color
FULL_SIMPL...
Filters
PREDICTION_COUNT by PREDICTION, FULL_SIMPLE_RULE
1.07%
2.29%
0.75%
1.64%
5.41%
0.14%
0.64%
1.48%
1.07%
0.86%
3.98%
2.29%
17.91%
14.02%
7.91%
15.41%
1.68%
0.14%
2.13%
2.16%
1.16%
1.00%
4.66%
3.20%
4.05%
0.45%
1.75%
0.61%
0.14%
FULL_SIMPLE_RULE
(WEALTH in ("Average...
(WEALTH in ("Average...
(WEALTH in ("Average...
(WEALTH in ("Average...
(WEALTH in ("Average...
(WEALTH in ("Average...
(WEALTH in ("Average...
(WEALTH in ("Average...
(WEALTH in ("Average...
(WEALTH in ("Average...
(WEALTH in ("Average...
(WEALTH in ("Average...
(WEALTH in ("Average...
(WEALTH in ("Average...
(WEALTH in ("Average...
(WEALTH in ("Rich" "V...
(WEALTH in ("Rich" "V...
(WEALTH in ("Rich" "V...
Exploratory Graphs of Credit Scores
More Expl Graphs
Key Factors of Good Credit Customers
Profiles of Good Credit Customer


ORACLE Data Visualization Desktop
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
PREDICTION
Color
FULL_SIMPL...
Filters
PREDICTION_COUNT by PREDICTION, FULL_SIMPLE_RULE
FULL_SIMPLE_RULE
(WEALTH in ("Average...
(WEALTH in ("Rich" "V...
17.91%
14.02%
7.91%
15.41%
PREDICTION_COUNT 863 (15.41%)
FULL_SIMPLE_RULE (WEALTH in ("Average" "Poor" "Very Poor")) AND (NO_OF_RECRUITERS_ON_LINKEDIN <= 1.95E+001) AND (PER...
PREDICTION >= 3.0
Exploratory Graphs of Credit Scores
More Expl Graphs
Key Factors of Good Credit Customers
Profiles of Good Credit Customer

Example Oracle ML Customer References

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

Update: £300M confirmed fraud
£700M additional potential identified

£1 Billion in savings
…Moving to Cloud

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

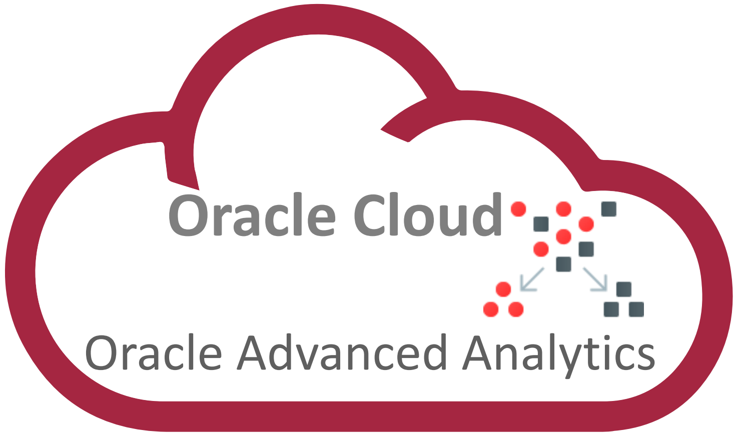
Solution

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%.”

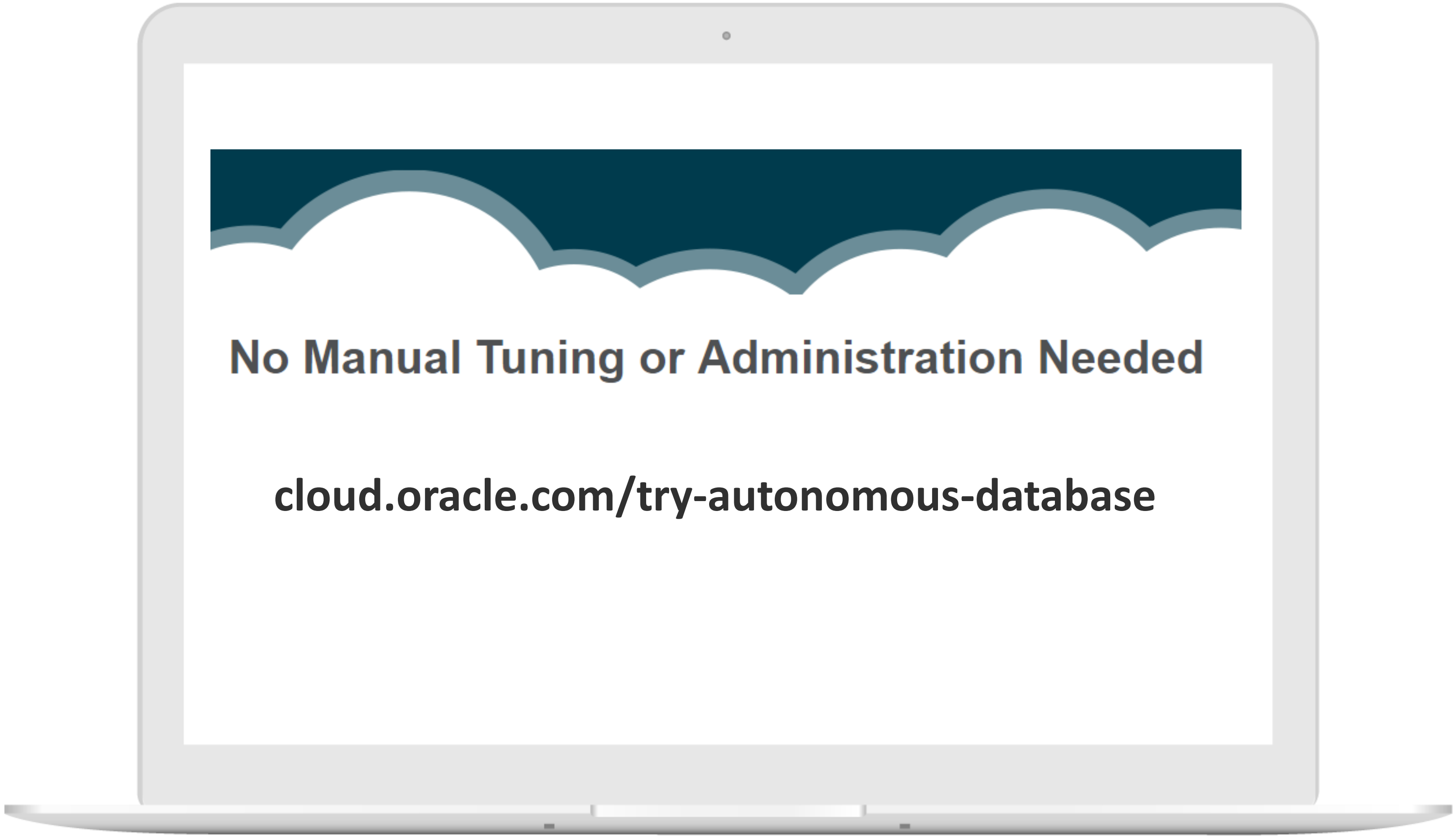
– Jadranka Novoselovic, Head of BI Dev., Zagrebačka Bank

“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

Get 3300 hours, 2 TB of Exadata Storage on Oracle Cloud for free*



* Trial expires upon usage of 3300 CPU hours or trial has reached 30 days, whichever comes first.

ORACLE

Analytics and Data Summit

All Analytics. All Data. No Nonsense.

March 12 – 14, 2019

BIWA

Formerly called the BIWA Summit with the Spatial and Graph Summit
Same great technical content…new name!
www.AnalyticsandDataSummit.org

Oracle Spatial & Graph SIG