NCOAUG
NORTH CENTRAL ORACLE APPS USER GROUP
TRAINING DAY
AUGUST 1, 2019

Leveraging Oracle Autonomous Data Warehouse for Advanced Analytics and Oracle Machine Learning

Kai Yu
ORACLE ACE Director
Distinguished Engineer, Database *Solutions Engineering, Dell EMC*
Jean Yu
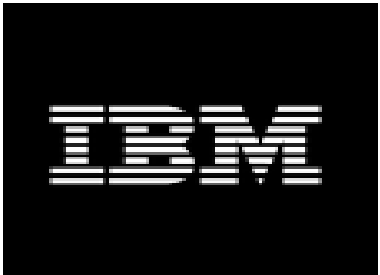Data Scientist, IBM

Kai Yu

DELLEMC

- ▪ Distinguished Engineer, Dell EMC Database Engineering
- ▪ 25+ years working in IT Industry
- ▪ Specializing in Oracle Database, Cloud, Virtualization
- ▪ Author and Speaker at IEEE and Oracle Conferences
- ▪ IOUG Cloud Computing SIG Co-founder and VP
- ▪ Oracle ACE Director
- ▪ Co-recipient of the 2011 OAUG Innovator of Year
- ▪ 2012 Oracle Excellence Award- Technologist of the
- • Year: Cloud Architect by Oracle Magazine
- ▪ My Blog: http://kyuoracleblog.wordpress.com/

Jean Yu

- **Data Scientist, IBM Multi-Cloud Management System**
- 21+ years of software development expertise
- Experience areas: Machine Learning, Multi-cloud management, storage system management, enterprise application management, IT enterprise deployment, workforce management, and banking arenas
- IBM Software Group Invention Review Board voting member since 2006
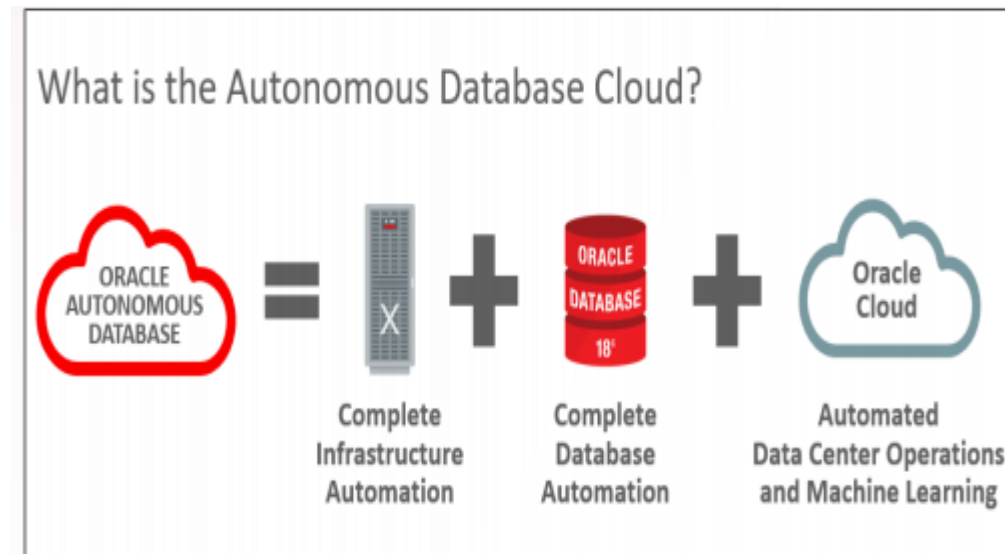- IBM Software Group Master Inventor with 12 patents granted and 4 invention plateaus.
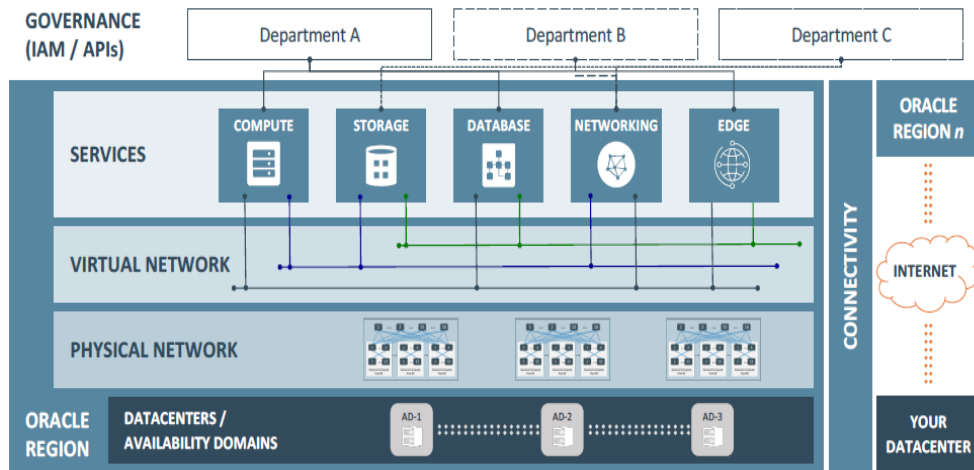
Agenda

- Oracle Autonomous Data Warehouse (ADW) Overview
- Working with ADW
- Running Oracle Machine Learning with ADW
- Machine Learning Model Building with Oracle Machine Learning

Oracle Autonomous Data Warehouse (ADW) Overview

- Oracle Autonomous Database : AI-based Automation for Data management and operation: **self-driving, self-securing, self-repairing**

- Oracle Autonomous Database built on Oracle Cloud Infrastructure (OCI)

- OCI (IaaS) provides a set of cloud infrastructure products Compute, storage, database, network Edge on the same flexibility virtual network.

- OCI supports Oracle cloud platform (PaaS) and Cloud Application(SaaS)

Oracle Autonomous Data Warehouse (ADW) Overview

- Oracle Autonomous Database : Optimized by workload:

One Autonomous Database - Optimized by Workload

ORACLE AUTONOMOUS DATABASE

Autonomous Data Warehouse (ADW)

Best for all Analytic Workloads:
- Data Warehouse, Data Mart
- Data Lake, Machine Learning

Autonomous Transaction Processing (ATP)

Best for TP and Mixed Workloads:
- Transactions, Batch, Reporting, IoT
- Application Dev, Machine Learning
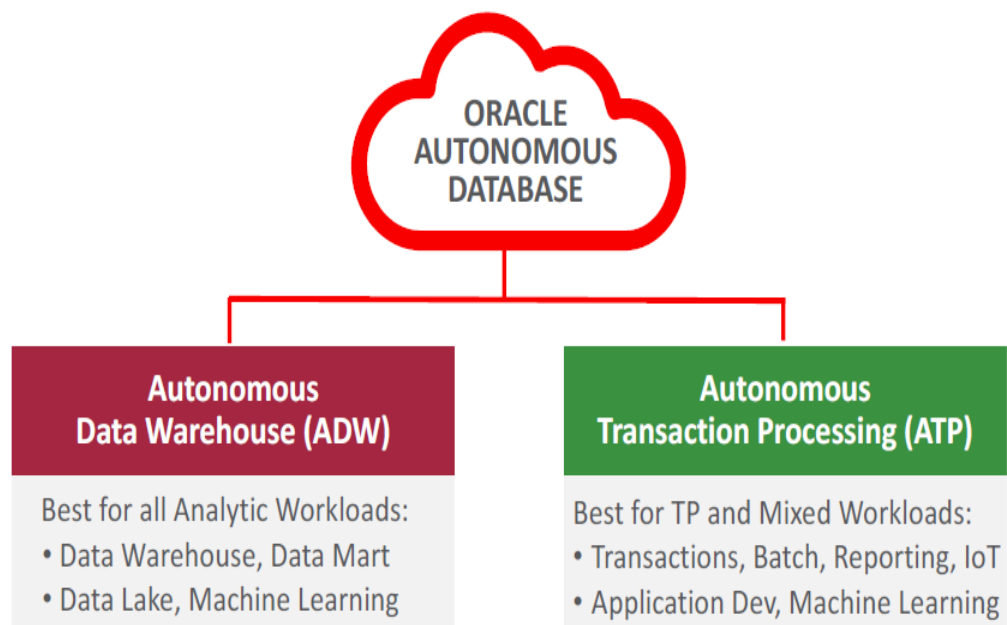
Autonomous Optimizations - Specialized by Workload

	ADW	ATP
Primary Goal	Fast Complex Analytics	Fast Simple Transactions
Data Formats	Columnar	Row
Data Access Acceleration	Creates Data Summaries	Creates Indexes Online*
Memory Usage	Parallel Joins and Aggregations	Data Caching to Avoid IO
Statistics	Updates optimizer statistics in real-time as data changes*	
SQL Plans	Manages SQL plans to prevent slowdowns and runaway SQL*	

- Autonomous Data Warehouse (ADW) provides an easy-to-use, fully autonomous data warehouse that scales elastically, delivers fast query performance and requires no database administration.

Oracle Autonomous Data Warehouse (ADW) Overview

- Autonomous Data Warehouse Characteristics
  - End-to-end management : provisioning, patching, backup, etc
  - Fully Tuned and provide good performance out of box , Ready to load and go
  - Scale compute and storage to fit workload without downtime
  - Support existing apps in cloud/ on–premise , connect with SQL*net, JDBC/ODBC
  - Use SQL Developer to connect to ADW for development and data loading.
  - Provide build-in web-based data analysis tools such as notebook for designing
     and sharing SQL based data-driven, interactive documents.
  - Business Intelligence tools: Oracle Analytic cloud, Oracle Data Virtualization Desktop, third party BI tool

Oracle Autonomous Data Warehouse (ADW) Overview

- Autonomous Data Warehouse Architecture:
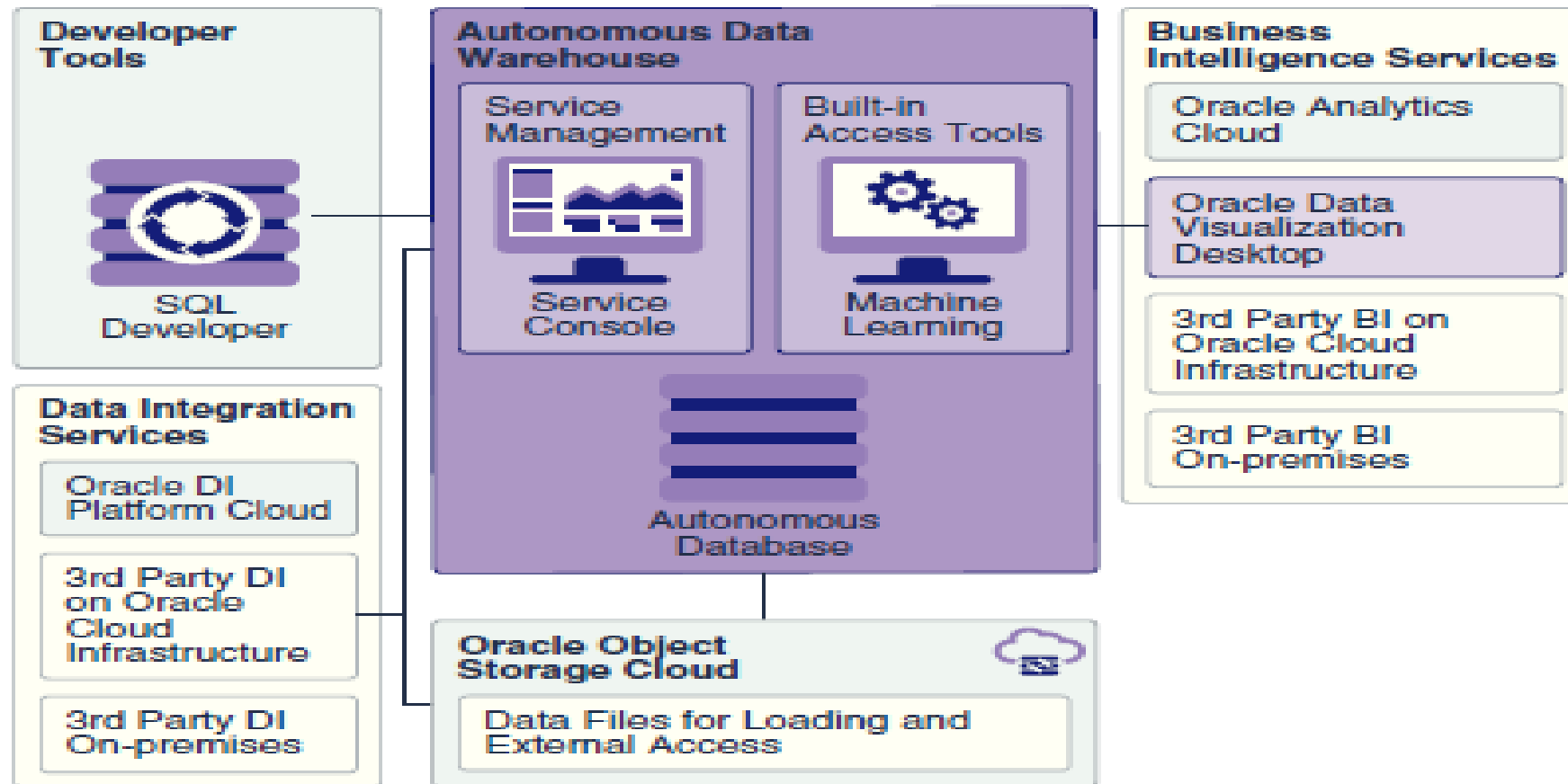
Working with Autonomous Data Warehouse

- Provisioning ADW Database Instance
  - Sign in to Oracle Cloud (you can leverage the free 30 days Try for Free)
  - Select Dashboard-> show Autonomous Data Warehouse->Click Autonomous Data Warehouse- to get Service: Autonomous Data Warehouse-> Open Service Console

Working with Autonomous Data Warehouse

- Provisioning ADW Database Instance
  - Click Create Autonomous Database-> Specify Workload Type, Database name, CPU core count, Storage (TB), Admin password -> take 10-15 minutes to create ADW Database service

Working with Autonomous Data Warehouse

- Connect to ADW database instance
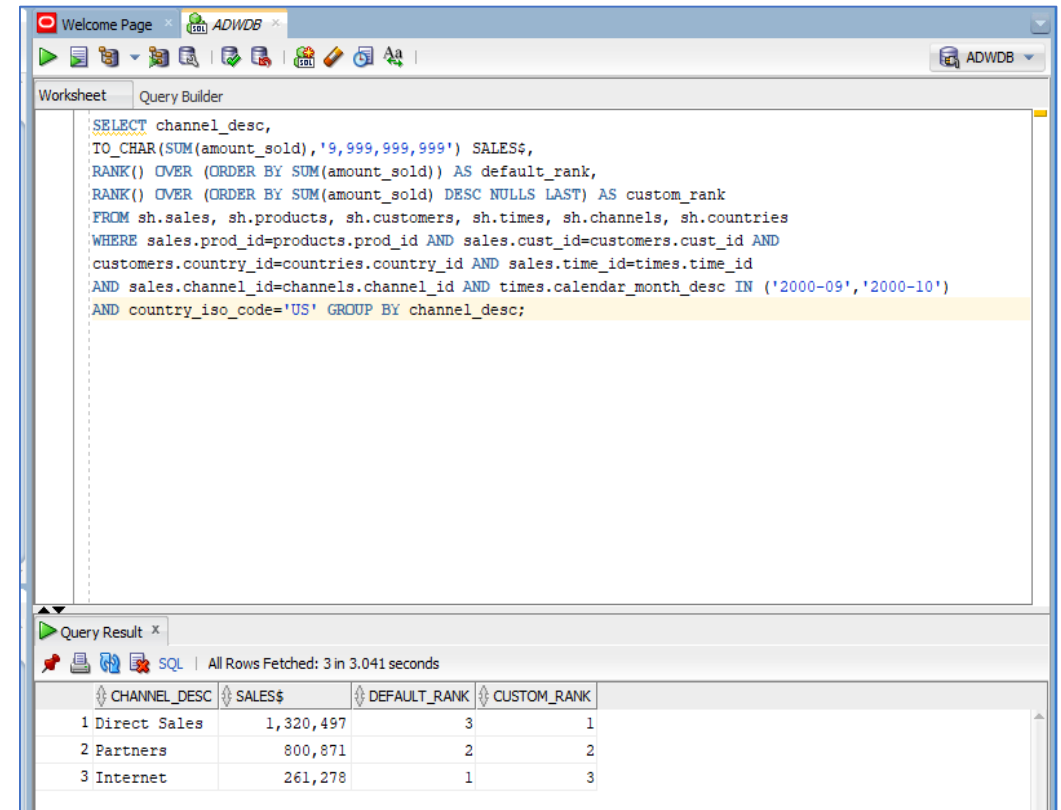 - Service Console->Administration-> Download a connection wallet
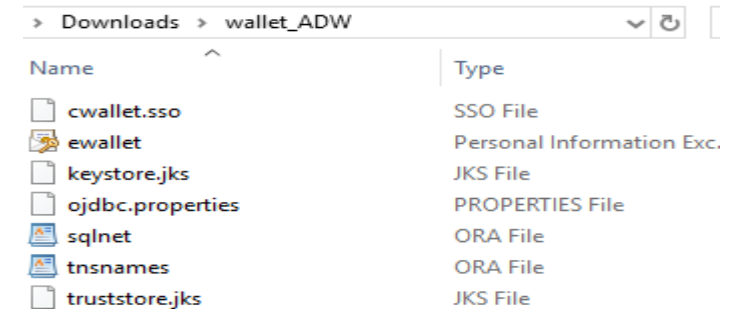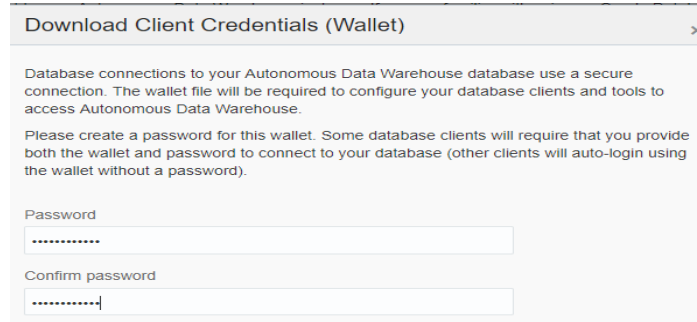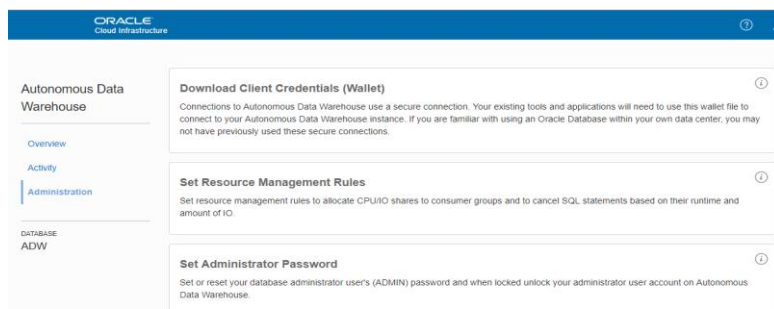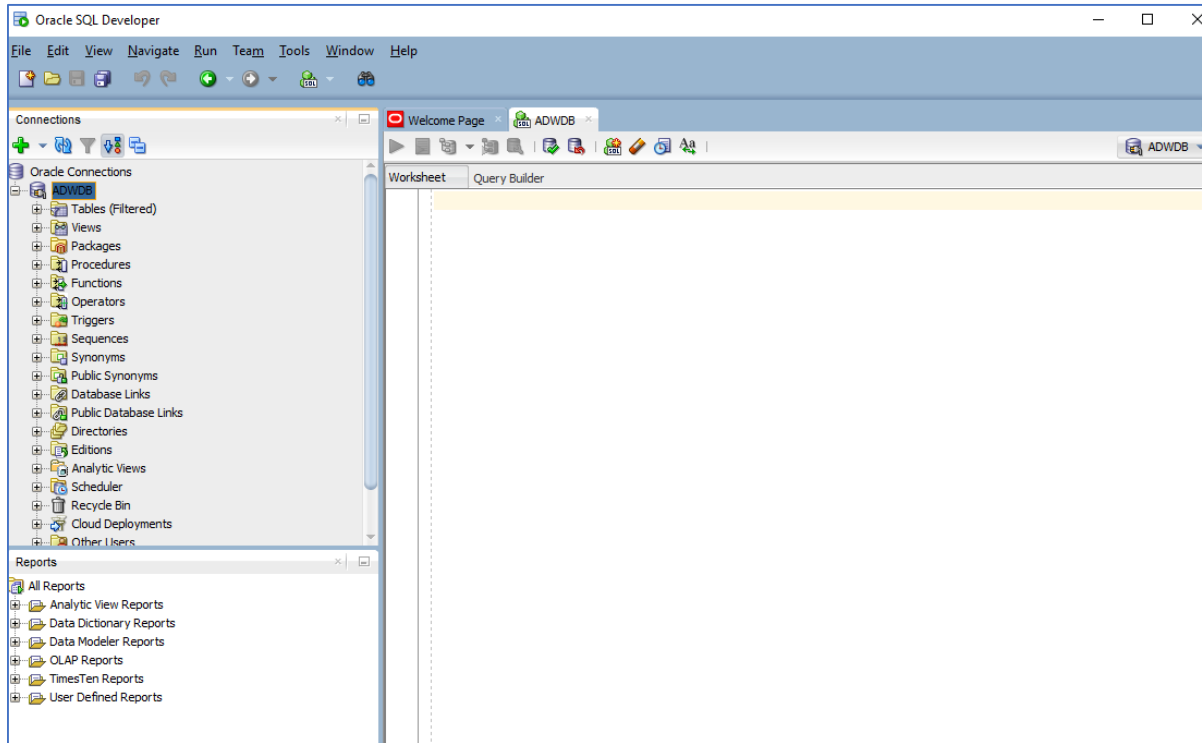

- SQL Developer connects to an ADW instance as admin user using a connection wallet
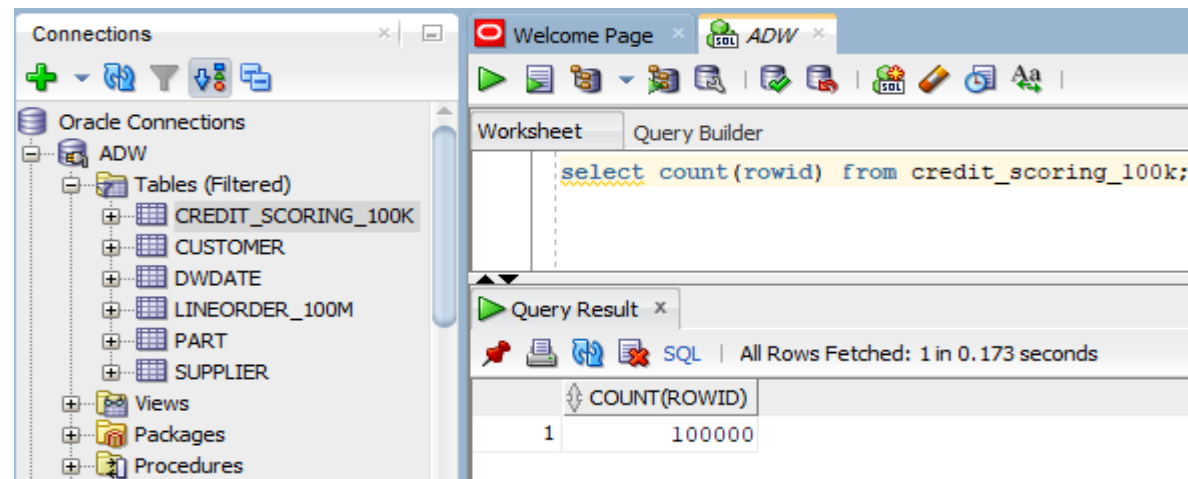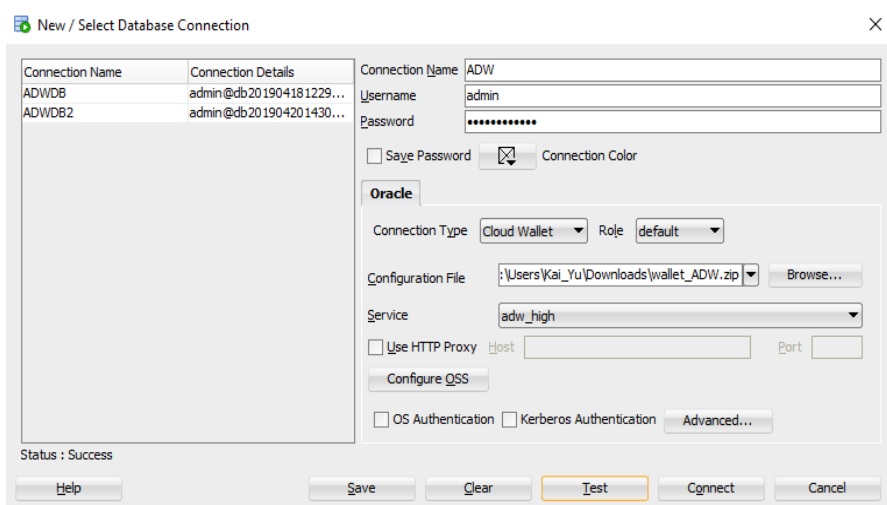
Working with Autonomous Data Warehouse
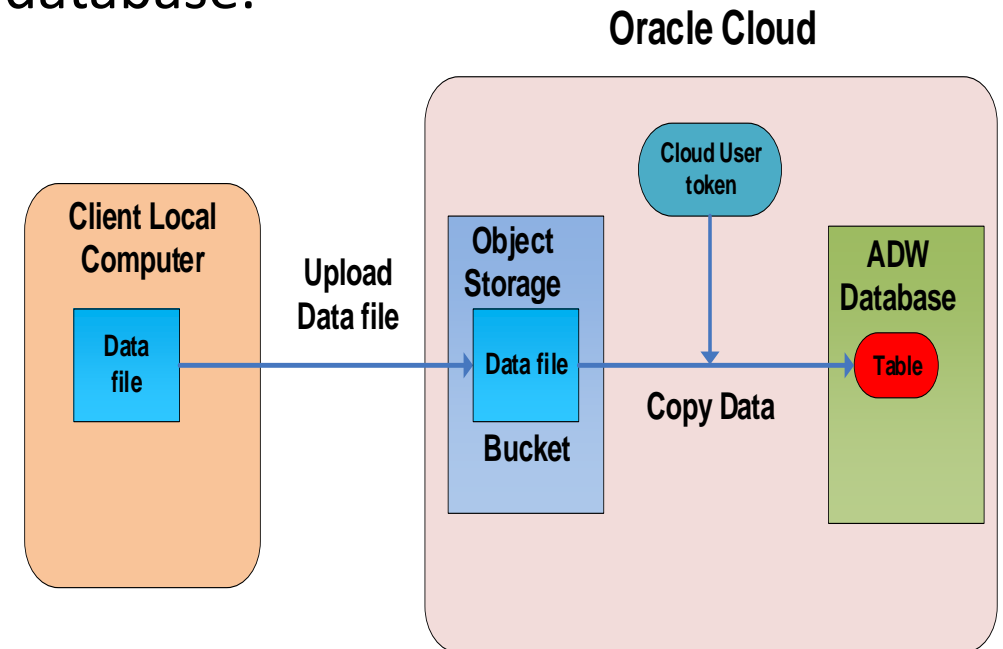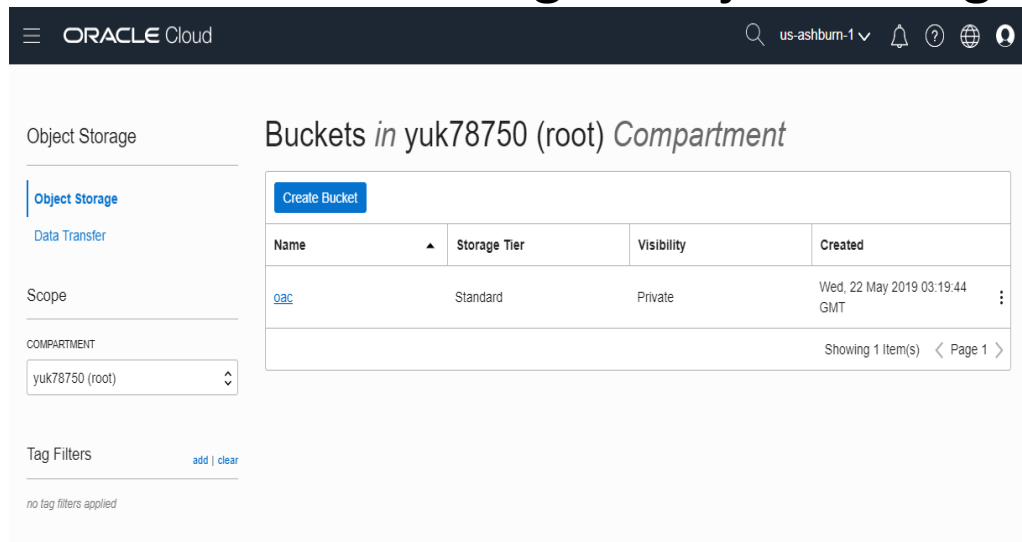
- Run queries against ADW

Working with Autonomous Data Warehouse

- Upload data from local computer to ADW table
  - Load data from local computer to Object storage in Oracle cloud
  - Copy data from Object storage to table of ADW database.
- Detailed steps:
  1. Create Bucket in object storage

Menu->block storage->Object storage

Working with Autonomous Data Warehouse

- Upload object from local computer to object storage

Working with Autonomous Data Warehouse

- Get the URL of the object

Working with Autonomous Data Warehouse

- Create auth token and create credential

Working with Autonomous Data Warehouse

Copy data to ADW database table

```
begin
  dbms_cloud.copy_data(
      table_name =>'credit_scoring_100k',
      credential_name =>'ADWC_TOKEN',
      file_uri_list => 'https://objectstorage.us-ashburn-1.oraclecloud.com/n/idodgb3rtcqf/b/adwc/o/credit_scoring_100
      format => json_object('ignoremissingcolumns' value 'true', 'removequotes' value 'true', 'dateformat' value 'YYY
  );
end;
```

Script Output ✕

Task completed in 16.237 seconds

PL/SQL procedure successfully completed.

Working with Autonomous Data Warehouse

- Trouble shooting tips: object not found

Check object URL link:

select object_name, bytes from dbms_cloud.list_objects('adwc_token3','https://objectstorage.us-ashburn1.oraclecloud.com/v1/yuk78750/adwc/credit_scoring_100k.csv');

ORA-20404: Object not found - https://objectstorage.us-ashburn-1.oraclecloud.com/v1/yuk78750/adwc/credit_scoring_100k.csv

ORA-06512: at "C##CLOUD$SERVICE.DBMS_CLOUD", line 558

ORA-06512: at "C##CLOUD$SERVICE.DBMS_CLOUD", line 1333

Verify object URL:

Working with Autonomous Data Warehouse

- Data loading Trouble shooting tips: Copy data failure:

Review error message in ADMIN"."COPY$33_LOG

```
begin
  dbms_cloud.copy_data(
    table_name =>'credit_scoring_100k',
    credential_name =>'ADWC_TOKEN3',
    file_uri_list => 'https://objectstorage.us-ashburn-1.oraclecloud.com/n/idodgb3rtcqf/b/adwc/o/credit_sc
    format => json_object('ignoremissingcolumns' value 'true', 'removequotes' value 'true', 'dateformat' v
  );
end;
```

Query Result ✕ | Script Output ✕

📌 ✏️ 💾 🖨 📋 | Task completed in 3.835 seconds

```
Error starting at line : 58 in command -
begin
  dbms_cloud.copy_data(
    table_name =>'credit_scoring_100k',
    credential_name =>'ADWC_TOKEN3',
    file_uri_list => 'https://objectstorage.us-ashburn-1.oraclecloud.com/n/idodgb3rtcqf/b/adwc/o/credit_scoring_1
    format => json_object('ignoremissingcolumns' value 'true', 'removequotes' value 'true', 'dateformat' value 'Y
  );
end;
Error report -
ORA-20003: Reject limit reached, query table "ADMIN"."COPY$33_LOG" for error details
ORA-06512: at "C##CLOUD$SERVICE.DBMS_CLOUD", line 540
ORA-06512: at "C##CLOUD$SERVICE.DBMS_CLOUD", line 1008
ORA-06512: at "C##CLOUD$SERVICE.DBMS_CLOUD", line 1031
ORA-06512: at line 2
```

Query Result ✕

📌 🖨 🔍 📋 SQL | All Rows Fetched: 197 in 2.561 seconds

RECORD		
181	Terminated by ","	
182	CHURN_RATE_OF_ACCOUNT_NO1	CHAR (255)
183	Terminated by ","	
184	CHURN_RATE__OF_ACCOUNT_NO2	CHAR (255)
185	Terminated by ","	
186	CHURN_RATE_OF_ACCOUNT_NON	CHAR (255)
187	Terminated by ","	
188	HEALTH_SCORE	CHAR (255)
189	Terminated by ","	
190	CUSTOMER_DEPTH	CHAR (255)
191	Terminated by ","	
192	LIFECYCLE_STAGE	CHAR (255)
193	Terminated by ","	
194	CREDIT_SCORE_BIN	CHAR (255)
195	Terminated by ","	
196	error processing column CUSTOMER_ID in row 1 for datafile https://objectstorage.us-ashburn-1.oraclecloud.c...	
197	ORA-01722: invalid number	

ORA-01722 invalid number. Cause: The attempted conversion of a character string to a **number** failed because the character string was not a valid numeric literal. Only numeric fields or character fields containing numeric data may be used in arithmetic functions or expressions.

Running Oracle Machine Learning with ADW

- Why Machine Learning: Analytics value and Maturity:
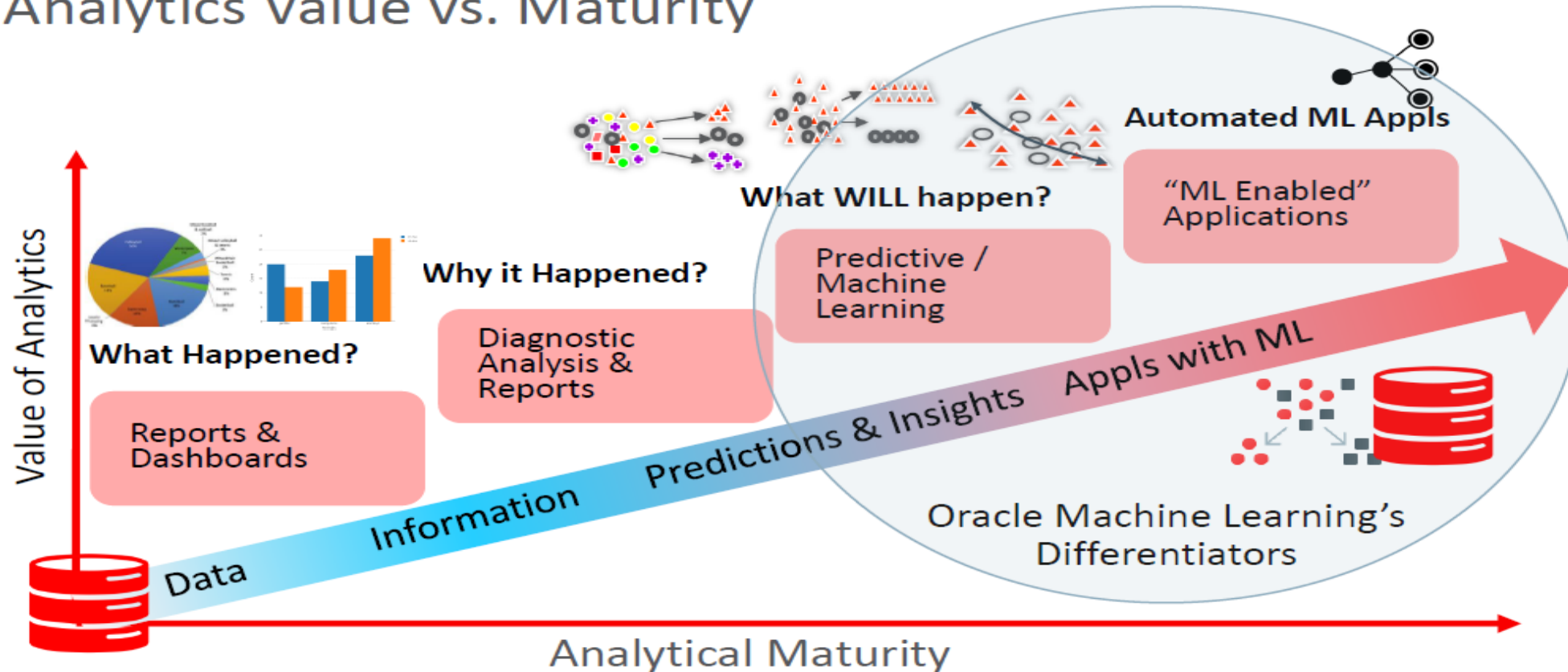
Running Oracle Machine Learning with ADW

- Machine Learning: algorithms automatically examine large amount of data to identify patterns, , discover new insight am make predications
- Machine learning algorithms implemented as SQL functions inside Oracle DB

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

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

ATTRIBUTE IMPORTANCE
- Minimum Description Length
- Principal Comp 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
- 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.
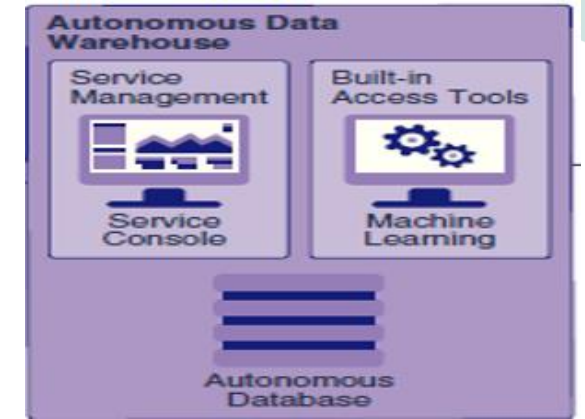
R AND PYTHON PACKAGES
- Third-party R and Python Packages through Embedded Execution
- Spark MLlib algorithm integration

Running Oracle Machine Learning with ADW

- Oracle Machine Learning: a part of ADWC collaborate environment
  - A web-based development to create data mining notebook
  - Used by data scientists, developer and business to users to perform data analytics, data discover and data virtualizations.
  - Leverage the ADWC scalability and performance in Oracle cloud .

- Use Oracle Machine Learning: start with Oracle ML users from ADW Admin

Running Oracle Machine Learning with ADW

- Use Oracle Machine Learning: Login to Oracle ML, taken into OML home page

- **Select Run SQL Statement to** SQL Query Scratchpad

Running Oracle Machine Learning with ADW

- The report menu bar lets you change the result to a graph and/or export result

- Set the report type to one of the graphs, a Settings link appears to the right of the menu. Settings allows you to control the layout of columns within the graph.
  :histogram, pie chart, line chart, cumulative gain chart, scatter plot.

Running Oracle Machine Learning with ADWC

- Save the Scratchpad as a new notebook. Click Edit button, give names and details of notebook, then save.

Running Oracle Machine Learning with ADW

- Use Oracle Machine Learning: Login to Oracle ML And look at Notebook Examples

Machine Learning Model building with OML

- Import a notebook of Data Exploration Analysis and OML model building
  (https://github.com/oracle/learning-library/blob/master/workshops/adwc4dev/L300.md)

Machine Learning Model Building with OML

- ## Data Gathering and Preparation
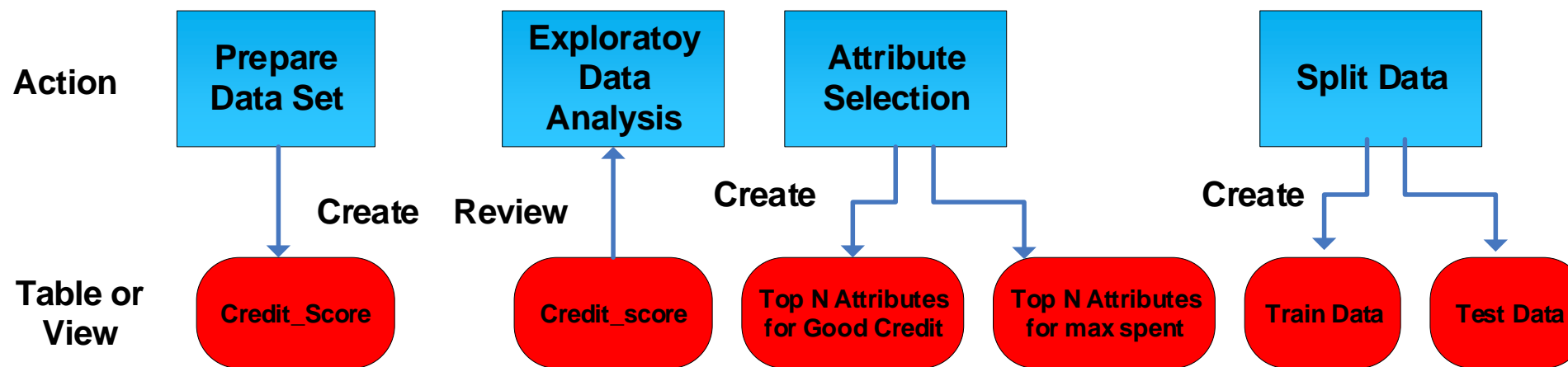
- ## ML Model Building, Evaluation and Application
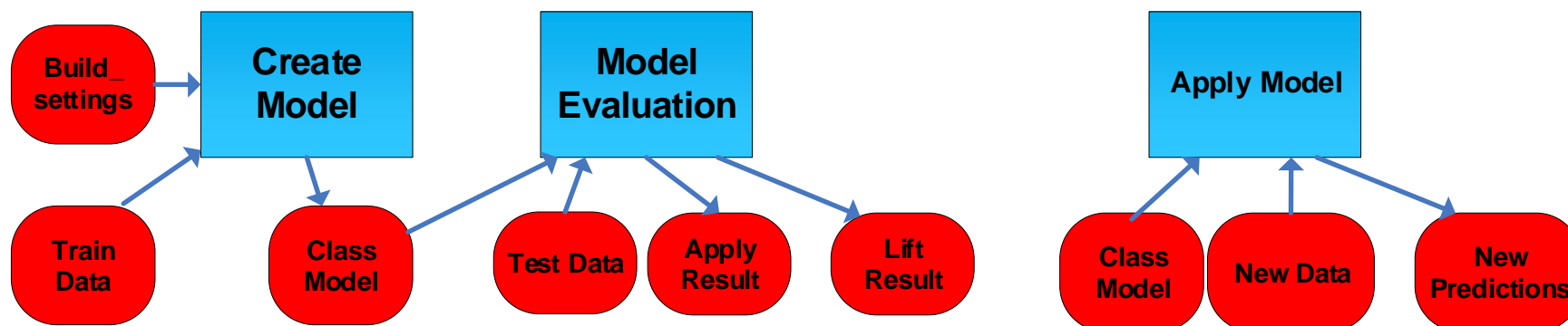
Data Gathering and Preparation with OML

- Prepare Data set: create view credit_scoring_100k as select * from credit_score
- Exploratory Data Analysis: by credit score; by Job and Income

Data Gathering and Preparation with OML

- Prepare Data set: create view credit_scoring_100k as select * from credit_score
- Exploratory Data Analysis: bin the variable

PL/SQL code:

```
dbms_data_mining_transform.create_bin_num(
                bin_table_name    => 'bin_num_tbl');
dbms_data_mining_transform.insert_autobin_num_eqwidth(
        bin_table_name   => 'bin_num_tbl',
        data_table_name  => 'CREDIT_SCORING_100K_V',
        bin_num          => 5,
        max_bin_num      => 10,
        exclude_list     =>
dbms_data_mining_transform.COLUMN_LIST('CUSTOMER_ID')
    );
dbms_data_mining_transform.xform_bin_num(
        bin_table_name     => 'bin_num_tbl',
        data_table_name    => 'CREDIT_SCORING_100K_V',
        xform_view_name    => 'mining_data_bin_view');
```


Show Histogram of Age (All Numeric Variables are Binned, This is an Example)   FINISHED

%sql

-- Note: Typically data scientists shold explore their data to make sure it makes sense and matches the data they believe they are working with.

select customer_id, age, income, tenure, loan_type, loan_amount, occupation, education_level from mining_data_bin_view;

●Grouped ○Stacked   ●CUSTOMER_ID
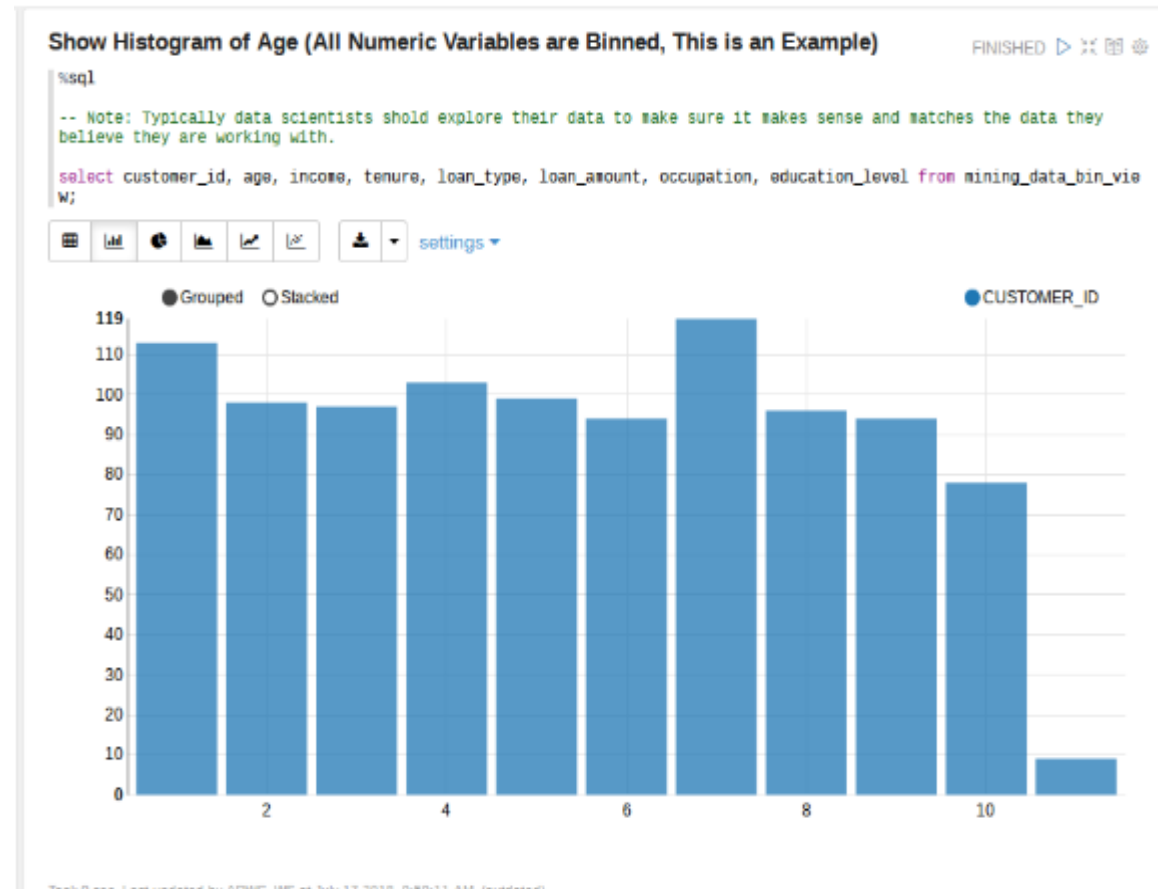
Data Gathering and Preparation with OML

- Attribute Selection aka Feature Engineering/Feature Selection
  Create attribute importance Machine Learning model for Good Credit
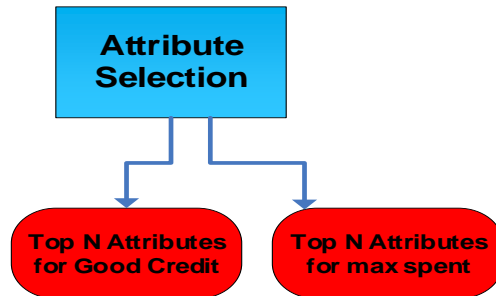

```
%script
DECLARE
v_sql varchar2(100);

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;
```

PL/SQL procedure successfully completed.

Attribute Selection

Top N Attributes for Good Credit

Top N Attributes for max spent

Display the Top N Attributes for Good Credit Customers

FINISHED

```
%sql

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

settings ▾

○ Grouped ○ Stacked

● EXPLANATORY_VALUE

0.646 / 0.6 / 0.5 / 0.4 / 0.3 / 0.2 / 0.1 / 0

WEALTH CUSTOMER_DMG_SEGMENT INCOME HIGHEST_CREDIT_CARD_LIMIT CUSTOMER_VALUE_SEGMENT RESIDENTAL_STATUS
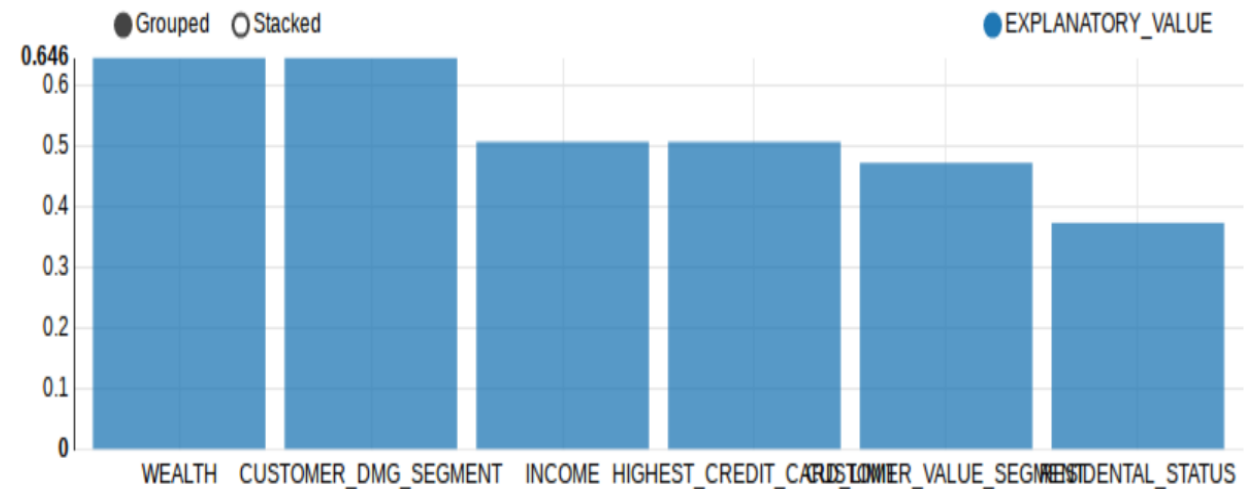
Data Gathering and Preparation with OML

- Attribute importance model for MAX_CC_SPENT_AMOUNT:

```
BEGIN
  DBMS_PREDICTIVE_ANALYTICS.EXPLAIN(
    data_table_name     => 'Credit_Scoring_100k_v',
    explain_column_name => 'MAX_CC_SPENT_AMOUNT',
    result_table_name   => 'ai_explain_output_MAX_CC_SPENT_AMOUNT');
END;
```

Display Top N Attributes for MAX_CC_SPENT_AMOUNT   FINISHED ▷ ✕ 📖 ⚙

%sql

Select * from ai_explain_output_MAX_CC_SPENT_AMOUNT where rownum < 7;

ATTRIBUTE_NAME	ATTRIBUTE_SUBNAME	EXPLANATORY_VALUE	RANK
INCOME		0.08842	1
HIGHEST_CREDIT_CARD_LIMIT		0.08841	2
CUSTOMER_DMG_SEGMENT		0.07173	3
WEALTH		0.07165	4
MAX_CC_SPENT_AMOUNT_PREV		0.06453	5
CHURN_RATE_OF_CC1		0.05036	6

- Split Data into Train data (60% data): N1_TRAIN_DATA
               Test Data   (40% data) : N1_TEST_DATA

Split Data

Create

Train Data   **Test Data**
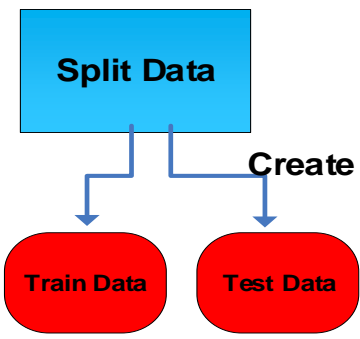
Create or replace view N1_TRAIN_DATA as SELECT * from CREDIT_scoring_100k_V SAMPLE (60) SEED (1)

Create or replace view N2_TRAIN_DATA as SELECT * from CREDIT_scoring_100k MINUS
               SELECT * FROM N1_ TRAIN_DATA

ML Model Building, Evaluation and Application with OML

- ## Create Predictive Model to Target Good Credit Customers
  - Add settings to n1_build_settings table
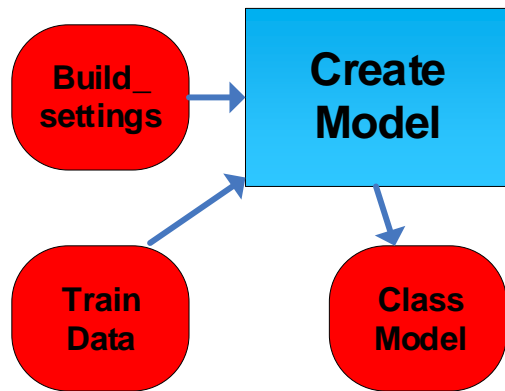

```
/* Create a Build Setting (DT) for Model Build */

EXECUTE IMMEDIATE 'CREATE TABLE n1_build_settings (setting_name VARCHAR2(30),setting_value VARCHAR2(4000))';
EXECUTE IMMEDIATE 'INSERT INTO n1_build_settings (setting_name, setting_value) VALUES (''ALGO_NAME'', ''ALGO_DECISION_TREE'')';
EXECUTE IMMEDIATE 'INSERT INTO n1_build_settings (setting_name, setting_value) VALUES (''PREP_AUTO'', ''ON'')';


DBMS_OUTPUT.PUT_LINE ('Created model build settings table: n1_build_settings ');

/*
-- Populate and Adjust Model Setting (DT) for Model Build
EXECUTE IMMEDIATE 'INSERT INTO n1_build_settings VALUES (''TREE_TERM_MAX_DEPTH'', 7)';
EXECUTE IMMEDIATE 'INSERT INTO n1_build_settings VALUES (''TREE_TERM_MINREC_SPLIT'', 20)';
EXECUTE IMMEDIATE 'INSERT INTO n1_build_settings VALUES (''TREE_TERM_MINPCT_SPLIT'', .1)';
EXECUTE IMMEDIATE 'INSERT INTO n1_build_settings VALUES (''TREE_TERM_MINREC_NODE'', 10)';
EXECUTE IMMEDIATE 'INSERT INTO n1_build_settings VALUES (''TREE_TERM_MINPCT_NODE'', 0.05)';
*/
```

  - Build a Classification model

Execute DBMS_DATA_MINING.CREATE_MODEL Procedure:
DBMS_DATA_MINING.APPLY('N1_CLASS_MODEL','CLASSFICATION','N1_TRAIN_DATA',
'CUSTOMER_ID','CREDIT_SCORE_BIN','n1_build_settings')

```
EXECUTE IMMEDIATE 'CALL DBMS_DATA_MINING.CREATE_MODEL(''N1_CLASS_MODEL'', ''CLASSIFICATION'', ''N1_TRAIN_DATA'', ''CUSTOMER_ID'',''
    CREDIT_SCORE_BIN'', ''n1_build_settings'')';
```

ML Model Building, Evaluation and Application with OML

- ## Model Evaluation
  - Test the Model by generating an apply result
    Execute DBMS_DATA_MINING.APPLY Procedure:

DBMS_DATA_MINING.APPLY('N1_CLASS_MODEL','N1_TEST_DATA','CUSTOMER_ID','N1_APPLY_RESULT')
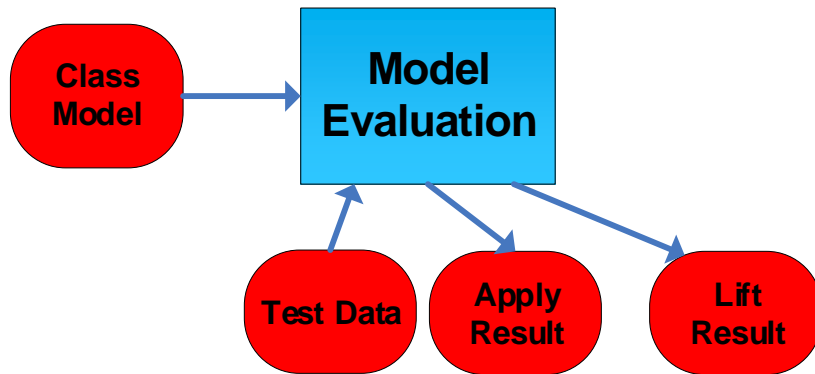
  - Create a lift result
    Execute DBMS_DATA_MINING.COMPUTE_LIFT Procedure:

DBMS_DATA_MINING. COMPUTE_LIFT ('N1_APPLY_RESULT', 'N1_TEST_DATA', 'CUSTOMER_ID',
'CREDIT_SCORE_BIN', 'N1_LEFT_TABLE', 'GOOD CREDIT', 'PREDICATION', 'PROBABILITY', 100)

```
EXECUTE IMMEDIATE 'CALL DBMS_DATA_MINING.COMPUTE_LIFT(''N1_APPLY_RESULT'',''N1_TEST_DATA'',''CUSTOMER_ID'',''CREDIT_SCORE_BIN''
    ,''N1_LIFT_TABLE'',''Good Credit'',''PREDICTION'',''PROBABILITY'',100);
```
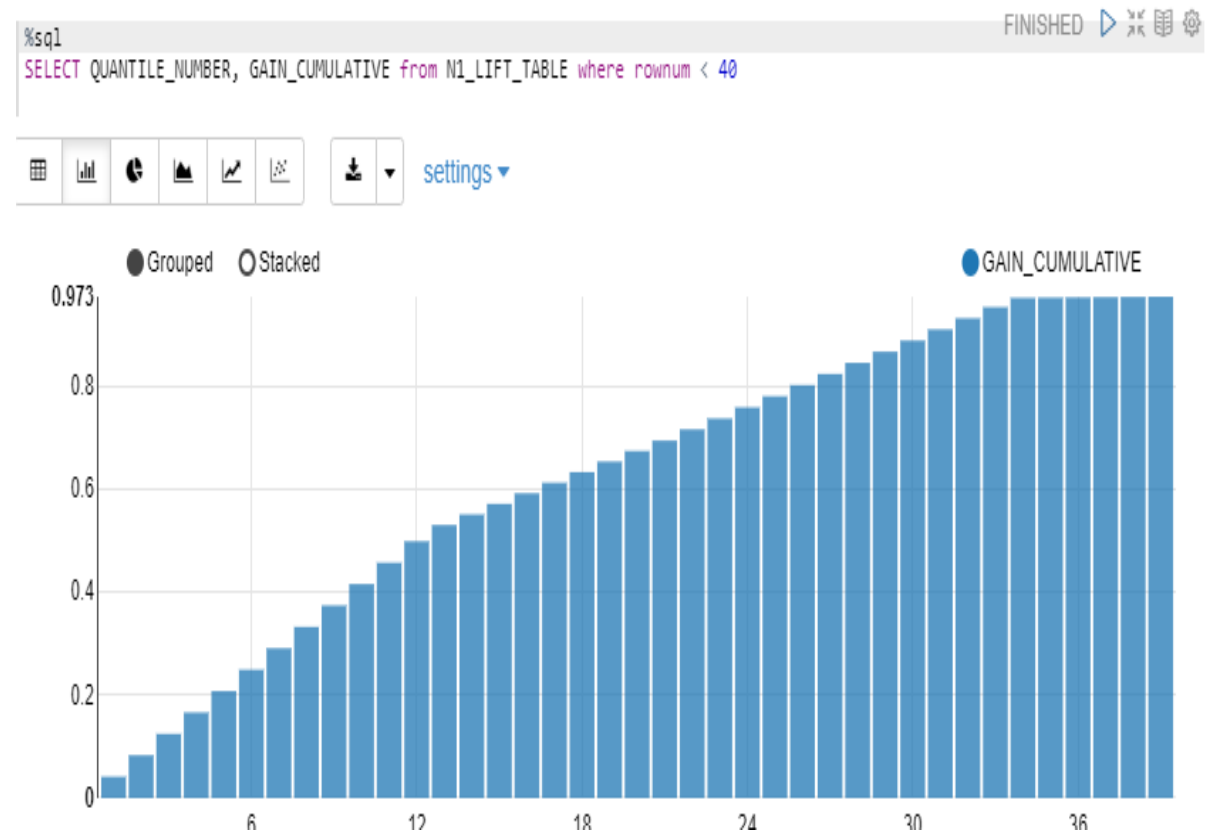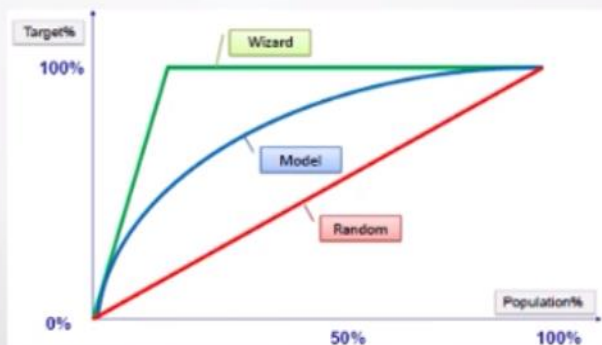
ML Model Building, Evaluation and Application with OML

- ## Model Evaluation
  - Review Lift result :View Model's Cummulative **Gains Chart and decide if its a good model**

ML Model Building, Evaluation and Application with OML

- ## Apply Model to New data:
  Apply the Oracle Machine Learning Model to New Customers to Show Customers Most Likely to Have Good Credit

%sql

select a.customer_id , a.prob_Credit_Score_Bin , b.age, b.income, b.tenure, b.loan_type,

      b.loan_amount, b.occupation, b.education_level, b.marital_status

from (select * from (select Customer_id, round(prob_Credit_Score_Bin *100,2) prob_Credit_Score_Bin

      from (select Customer_ID, prediction_probability(N1_CLASS_MODEL, NULL using *)

      prob_Credit_Score_Bin from credit_scoring_new_cust_v))) a, credit_scoring_100k_v b

where a.customer_id = b.customer_id

order by a.prob_Credit_Score_Bin desc

Put the predication results into credit_score_new_predictions table

```
%
create table credit_score_new_predictions as
select a.customer_id
     , a.prob_good_credit
     , b.age, b.income, b.tenure, b.loan_type, b.loan_amount, b.occupation, b.education_level, b.marital_status
from (select * from (select Customer_id, round(prob_good_credit *100,2) prob_good_credit from (select Customer_ID,
     prediction_probability(N1_CLASS_MODEL, 'Good Credit' using *) prob_good_credit from credit_scoring_new_cust_v))) a
, credit_scoring_100k_v b
where a.customer_id = b.customer_id
```

FINISHED

Updated 9920 row(s).
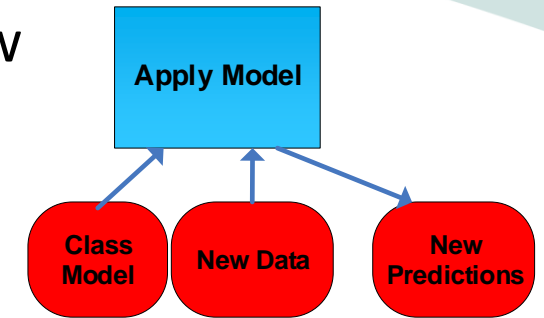
ML Model Building, Evaluation and Application with OML

- ## Apply Model to New data:
  Review new predications table
   select * from credit_score_new_predictions
  order by rank() over (order by prob_good_credit desc)

Apply a ML Model to a Single Record in a Transactional Application:
select prediction_probability(N1_CLASS_MODEL, 'Good Credit' USING 'Rich' as WEALTH, 2000 as income, 'Silver' as customer_value_segment) Prediction_Probability
from dual;


```
%
/* Score and rank new customers that are a professional, married, and borrowing for education (in this case).  You can
    substitute or add other filters. */

select * from credit_score_new_predictions
order by rank() over (order by prob_good_credit desc)
```

CUSTOMER_ID	PROB_GOOD_CREDIT	AGE	INCOME	TENURE	LOAN_TYPE	LOAN_AMOUNT	OCCUPAT
46459	100	20	5250	17	Housing	95000	Profession
89734	100	22	4250	14	Housing	95000	Profession
93597	100	24	5250	26	Housing	95000	Profession
18317	100	25	5250	25	Housing	95000	Profession
9193	100	31	5250	36	Housing	95000	Profession
7172	100	32	4250	10	Housing	95000	Profession
75821	100	36	4250	29	Housing	95000	Profession
17989	100	45	6250	3	Housing	95000	Profession

Took 1 min 21 sec. Last updated by ADWC_WS at May 24 2019, 11:18:13 PM.

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

PREDICTION_PROBABILITY
0.5098856966326846

References:

1. Oracle Cloud Using Oracle Autonomous Data Warehouse E85417-33 April 2019
2. Oracle Cloud Using Oracle Machine Learning E78535-12 March 2019
3. Oracle Cloud Oracle Autonomous Data Warehouse , Hans on Workshop Lab 100-400
(https://github.com/oracle/learning-library/tree/master/workshops/adwc4dev
4. Take your Analytics to the Next Level of Insight using Machine Learning in the Oracle
Autonomous Database, IOUG Collaborate 19, Charlie Berger, Oracle

Thank You and QA

Contact me at kai_yu@dell.com or visit my Oracle Blog at
http://kyuoracleblog.wordpress.com/

NCOAUG
NORTH CENTRAL ORACLE APPS USER GROUP
TRAINING DAY
AUGUST 1, 2019