



Qubix



Oracle Analytics & Machine Learning  
**Business Users and ML in Oracle Analytics**

Žiga Vaupot  
Country Manager & Business Analytics Consultant

# Introductions

---



**Žiga Vaupot**

**Qubix**

Country Manager, Slovenia/CEE  
Business Analytics Consultant

**SIOUG**

Board Member



[ziga.vaupot@qubix.com](mailto:ziga.vaupot@qubix.com)



[zigavaupot.blogspot.com](http://zigavaupot.blogspot.com)



[linkedin.com/in/zigavaupot](https://www.linkedin.com/in/zigavaupot)



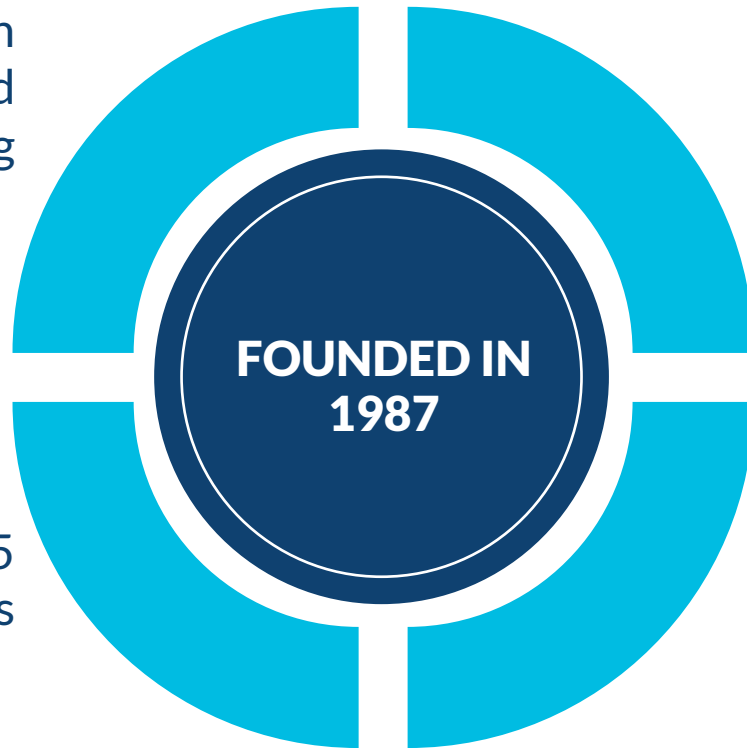
**ORACLE**  
ACE



**Qubix**

# Introducing Qubix

Capabilities focused on  
ERP, EPM, Analytics and  
Data Engineering



1,000+  
engagements

Offices in 5  
countries

Global Oracle multi  
award-winning  
partner



# Qubix

**Qubix**

# Business Users and ML in Oracle Analytics



# Typical users of Oracle Analytics

---



## Business users

Enterprise type of reporting  
Pixel-perfect  
Predefined dashboarding and reporting



## Business analysts

Effective data visualization and presentation  
Ad-hoc  
Self-service data preparation

# Machine Learning in Oracle Analytics

---

- (More or less) complete machine learning process cycle
- Effective data visualization
- Self-Service data preparation
- Self-Service model training and deployment
- Supports most popular supervised and unsupervised functions and algorithms (Classification, Regression, Clustering)
- Possibility to register, evaluate and use external data models (eg. from Autonomous DB)

# How typical user would most likely use ML in Oracle Analytics?



## Business users

No idea how ML works, possibly quite limited understanding.

In most cases they are able to define business problems and ask questions.

Will wait for results to be presented and interpreted by others.

## Business analysts



Possibly, have pretty good idea what ML is about and how it works.

Good candidates for self-service ML.

Will be able to understand the results and will use results in their projects.

# A tool of choice for Data Scientists?

---



Hell, no!

Well maybe, if I could use some serious ML tools and just feed the results back into Oracle Analytics.

# So, what options have we got then?

## a) Everything is done within Oracle Analytics



### Very likely scenario:

Business analyst has enough business knowledge to define the the business problem.

Business analysis might posses enough (data flows/coding) skills to collect and prepare data. This one could be a tricky one (!).

Business analyst will be able to use Oracle Analytics toolset to train, evaluate and deploy ML models.

Business analyst might engage business user to obtain more business insights and/or data scientist for model improvements. With enough ML knowledge, business analyst might take option B.

## b) Part of the process happens outside of Oracle Analytics



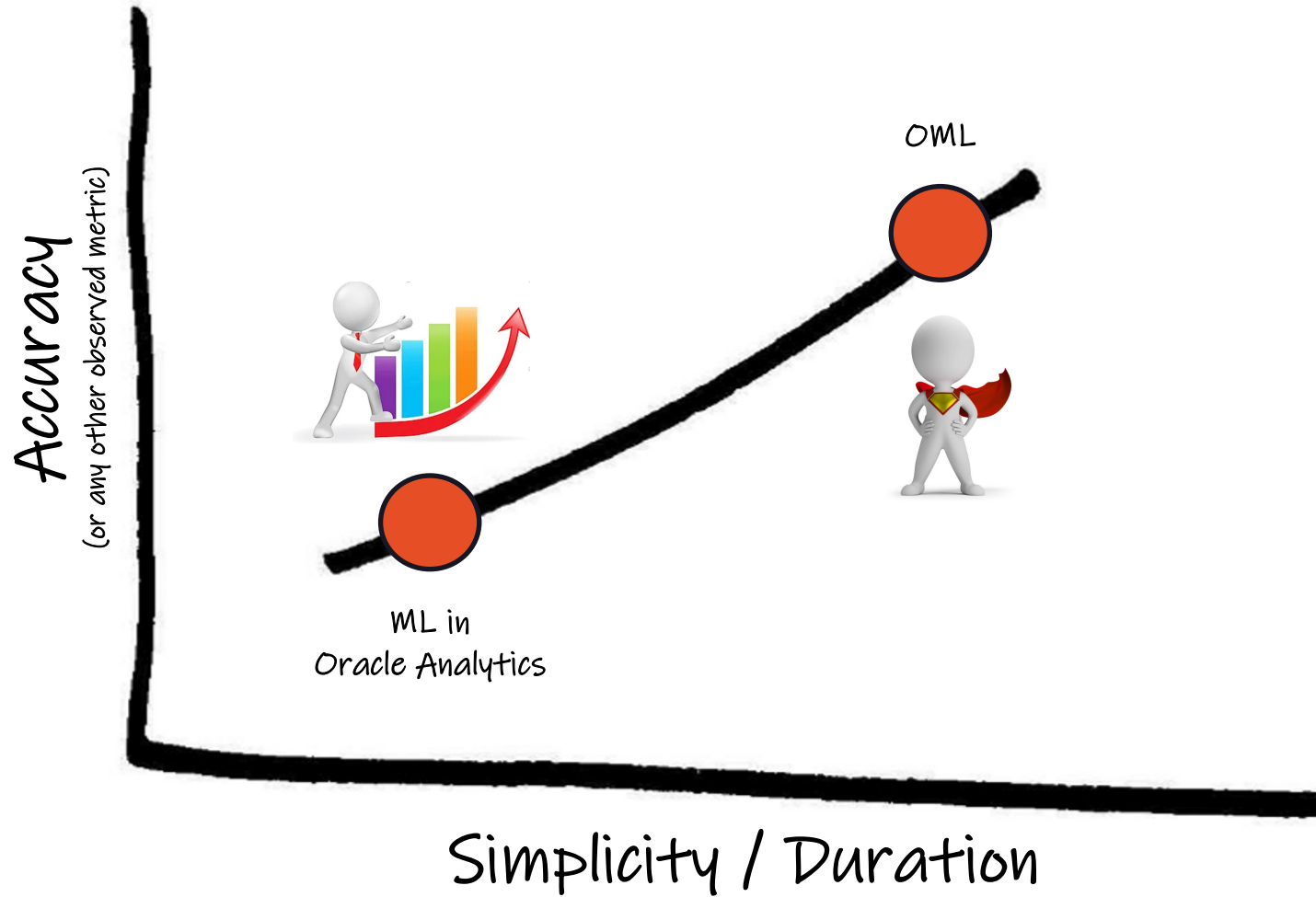
### Very likely scenario:

Business user will search for a data scientist (or will turn to Business analyst, but this is scenario A)

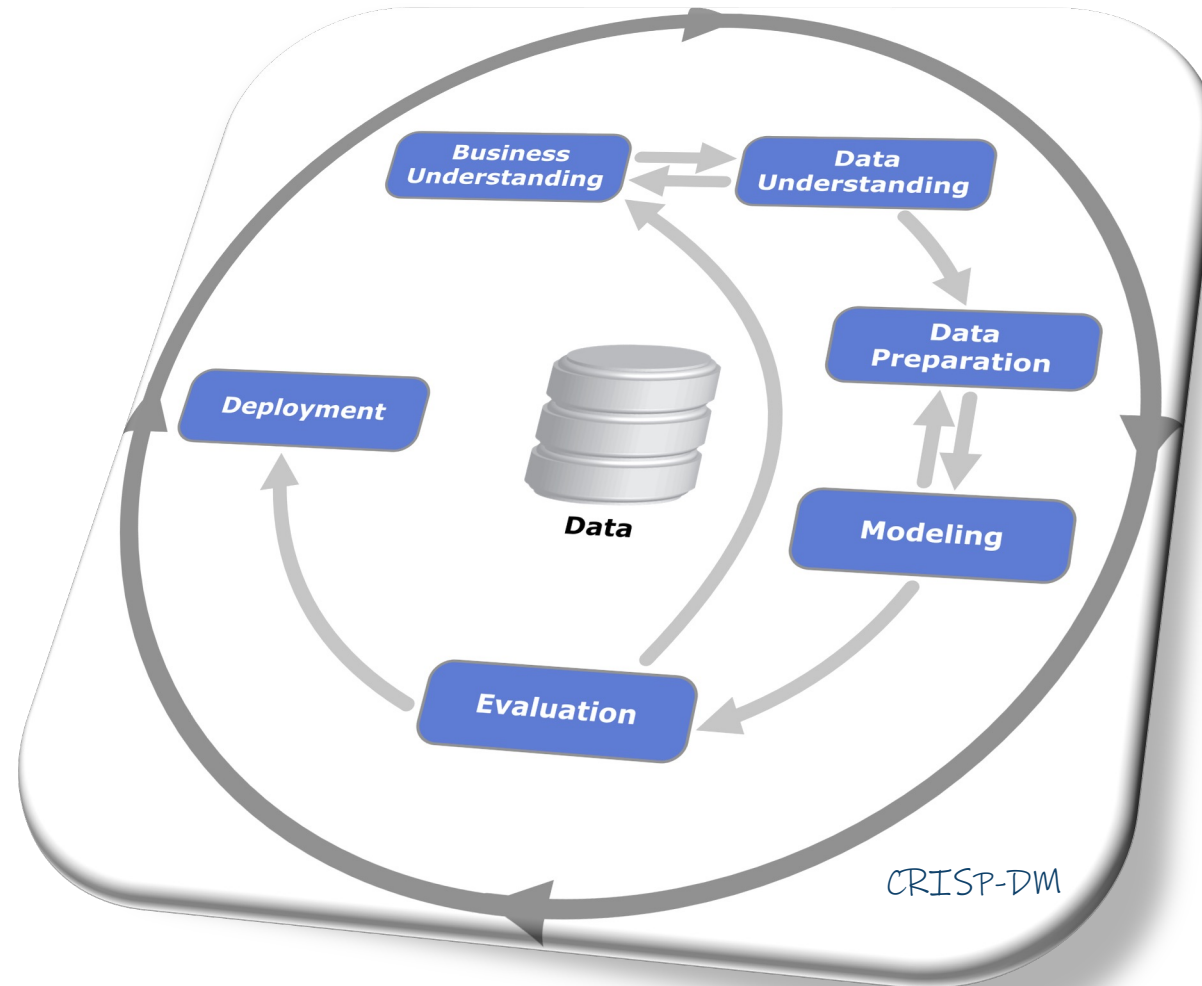
Data scientist will follow standard machine learning process using OML to come back with the best possible machine learning model. Model can already be deployed, in this case final results are presented to business user.

Model will be then registered or deployed results dataset can be created with Oracle Analytic. This can be done by (if skilled) Business user or more likely Business Analyst.

# Trade-off



# Regardless of scenario chosen, ML process applies!

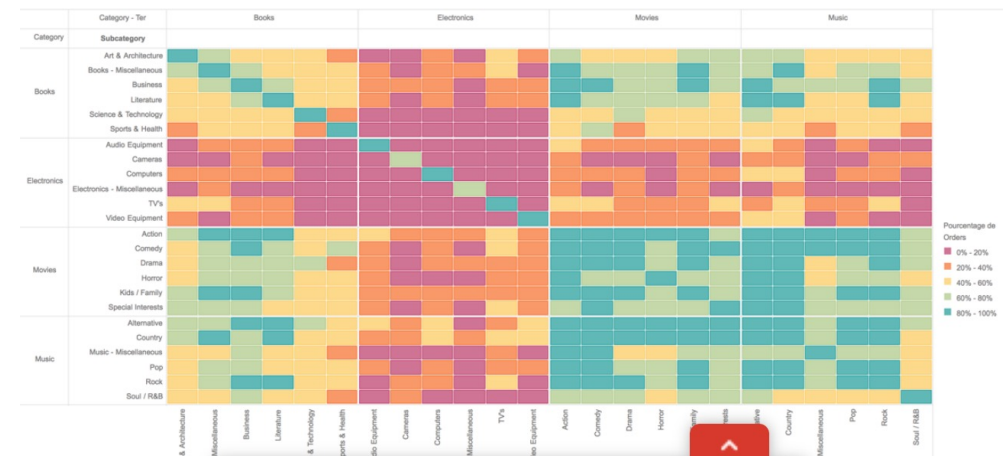


# Asking the Right Question

Which products/product subcategories are most often sold together?

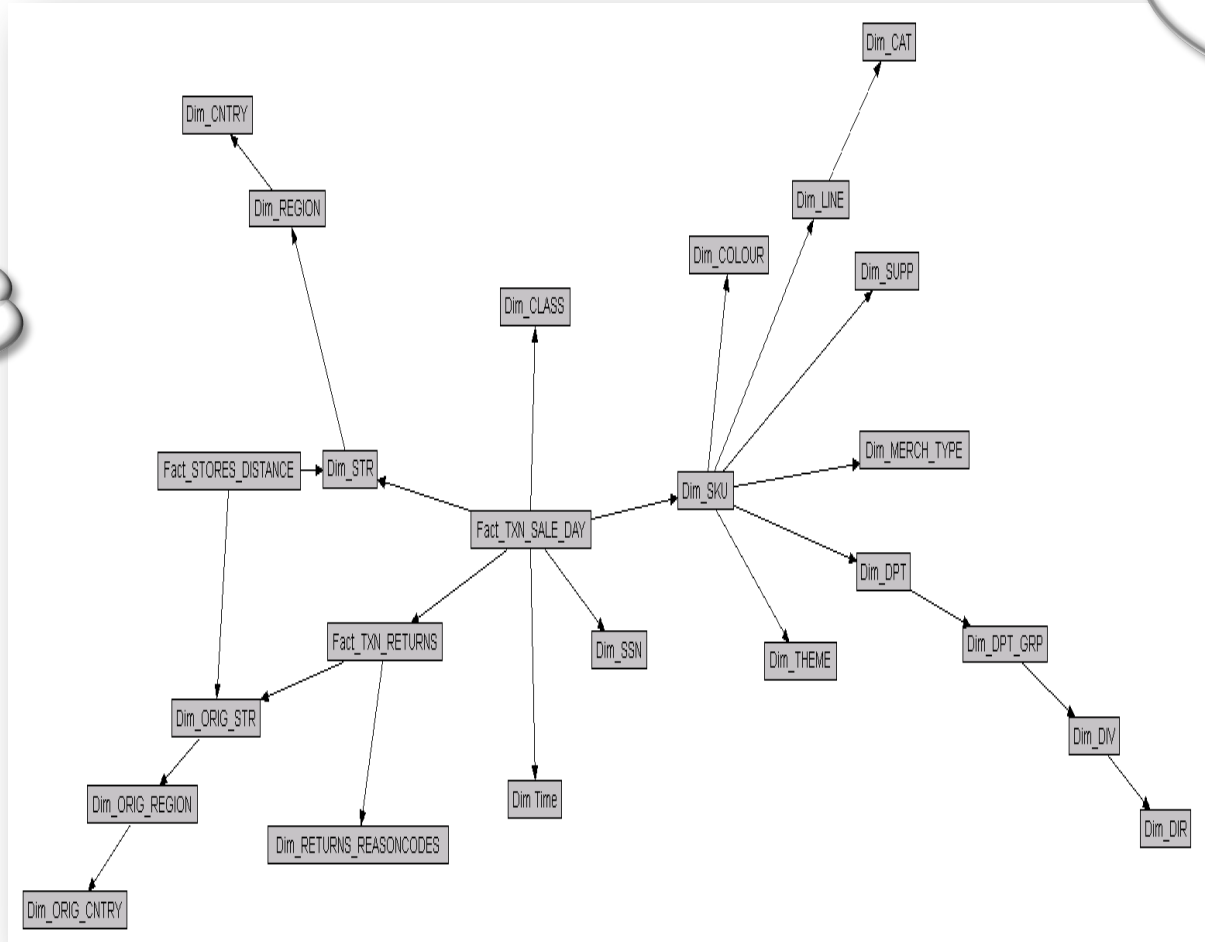
Which are the biggest opportunities for up/cross-sell?

Can we analyse this by each and every store by periods of the day?



# The Data

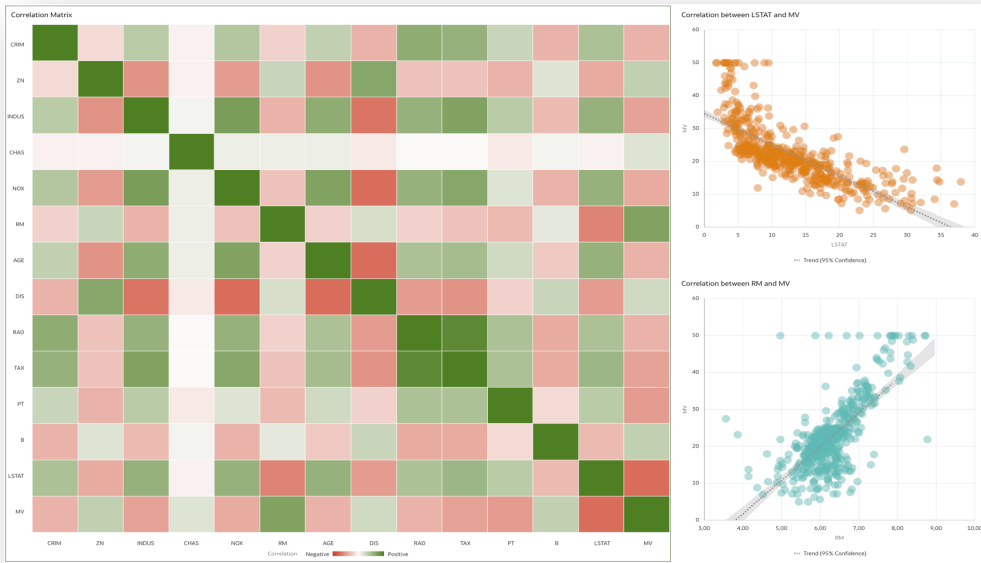
No, no, no, no ...



We are ready for some great predictions. We've got all of our data in a data warehouse!



# Data Collection & Exploratory Data Analysis



## Oracle Analytics

- Data Visualizations: Correlation Matrix, Trendline (regression), Boxplot ...
- Explain, Autoinsights



## Zeppelin Notebooks

- SQL, PL/SQL and Python for data wrangling
- SQL/Python Statistical Functions
- Visualizations



# Facts

---

Data preparation can take

60-90% of total ML project time.

Data quality contributes 75% - 90%  
to overall ML project success.



# Yes, this is true ...

Model Training Script **Linear Regression for model training**

\* Target **MV**  
target, the target(label) to learn/predict

Regression Method **Ordinary Least Squares**  
Method for linear regression training.

Regularization Weight **0**  
Regularization Weight(L1 Ratio or L2 Ratio). Please enter 0

**Categorical Column Imputation** **Most Frequent**  
The mode method for categorical features to fill NA. Two o

**Numerical Column Imputation** **Mean**  
The mode method for numeric features to fill NA. Four opt

**Categorical Encoding Method** **Indexer**  
Encoding method.

Maximum Null Value Percent **80**  
Maximum Null Value Percent

Train Partition Percent **80**  
The percentage of original data used for training, default is

**Standardization** **False**  
Standarize data before training.

Resolving missing values issue

Handling Categorical Values

Standardization/ Normalisation

## ... but pay attention ...

Which formula is used for standardization and what are the consequences?

$$v'_i = \frac{v_i}{10^k}$$

$$v'_i = \frac{v_i - \min_v}{\max_v - \min_v} * (\max_{v'} - \min_{v'}) + \min_{v'}$$

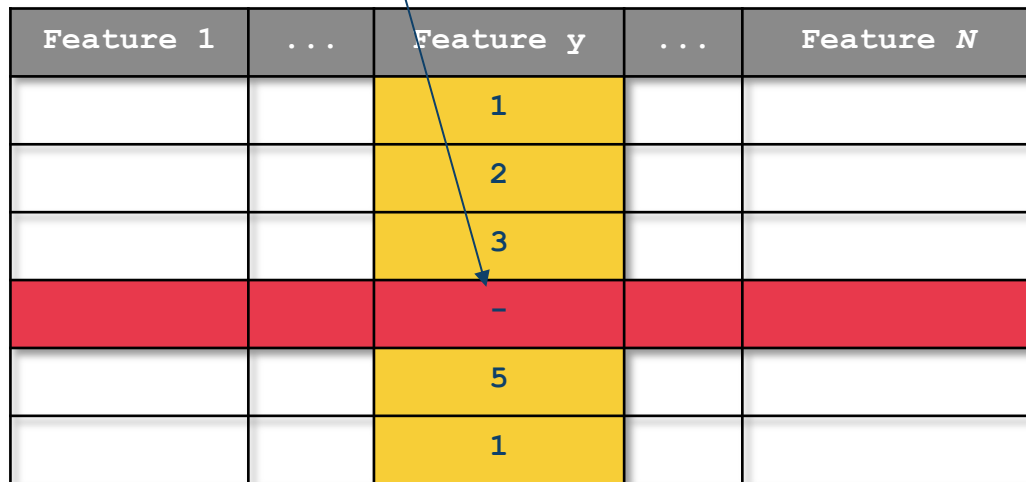
$$v'_i = \frac{v_i - \bar{V}}{\sigma_v}$$

$$v_n = \frac{1}{1 + e^{-v_i}}$$

(replace  $v_i$  in formula above with  $v_t = \frac{(v_i - V)}{\lambda(\sigma_v/2\pi)}$ )

## ... or consider the following:

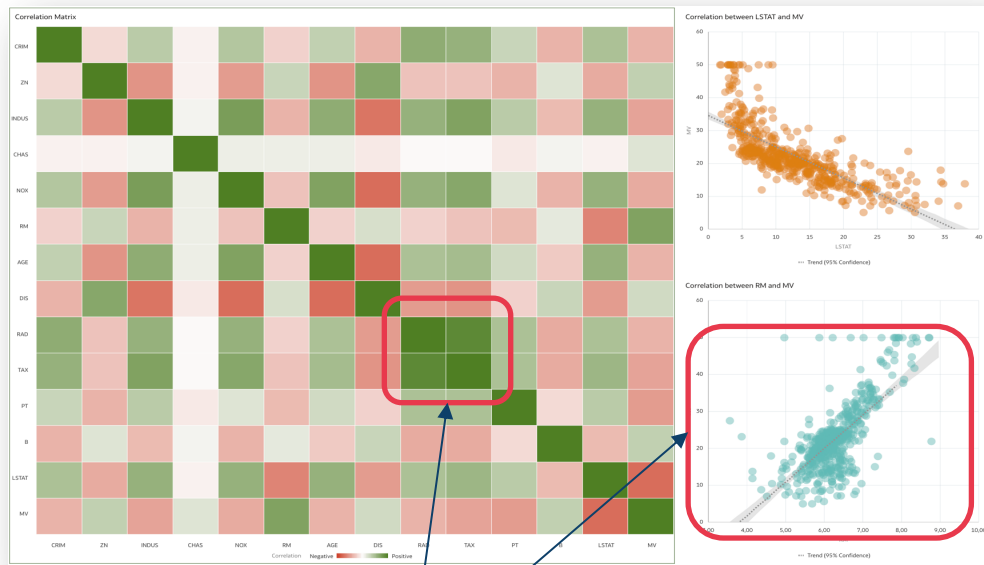
Which value should replace missing values?



Feature 1	...	Feature y	...	Feature N
		1		
		2		
		3		
		-		
		5		
		1		

- Remove row with missing value or
- Replace missing value with
  - some generic constant
  - some attribute statistics like average or median
  - some stitistic value that belongs to the same specific group. For example, customer segment average.
  - the most probable value.
  - a value that retains characteristics of the sample. For example, calculate a value that retains variability.

# ... and don't forget, correlations won't be removed automatically.



High correlation identified between two columns



Data flows is a perfect self-service tool to perform data preparation!

Select Columns - REMOVE TAX    TAX and RAD are highly correlated, hence removing one of these two attributes.

Search     Add all    Add selected    Selected (14/15)    Remove all    Remove selected

TAX    IDX

# Data Scientist will follow the best practices

- ... using SQL and/or python for required data preparation transformations.
- There might be other data preparation challenges to deal with which would require:
  - data sampling
  - dimensionality reduction
  - aggregations
  - etc.
- When running OML algorithms, in most cases, PREP\_AUTO parameter will be set to ON. This will automatically perform data transformations required by specific algorithm.

```
%python
try:
    oml.drop(model="ANOMALY_DETECTION_MODEL_CLAIMS0")
except:
    print("Model does not exist")

odm_settings = {'prep_auto' : 'ON'}

svm_mod = oml.svm("anomaly_detection", **odm_settings)
svm_mod.fit(CLAIMS0, None, model_name = 'ANOMALY_DETECTION_MODEL_CLAIMS0')
```

Model does not exist

Model Name: ANOMALY\_DETECTION\_MODEL\_CLAIMS0

Model Owner: ML\_USER

Algorithm Name: Support Vector Machine

Mining Function: ANOMALY\_DETECTION

Settings:

	setting name	setting value
0	ALGO_NAME	ALGO_SUPPORT_VECTOR_MACHINES
1	ODMS_DETAILS	ODMS_ENABLE
2	ODMS_MISSING_VALUE_TREATMENT	ODMS_MISSING_VALUE_AUTO
3	ODMS_SAMPLING	ODMS_SAMPLING_DISABLE
4	PREP_AUTO	ON
5	SVMS_CONV_TOLERANCE	.0001
6	SVMS_KERNEL_FUNCTION	SVMS_LINEAR
7	SVMS_OUTLIER_RATE	.01



# ML model training in Oracle Analytics

- Oracle Analytics offers several machine learning algorithms for supervised and unsupervised machine learning
  - Binary and Multivariate Classification
  - Regression
  - Clustering
  - Anomaly Detection\*
  - Frequent Itemsets & Association Rules\*
  - Time Series
- Limited parametrisation
- Train/test split and evaluation is done automatically
- Additional datasets are created for better explainability



LEAD SCORE - Create NeuralNetwork ML model

Lead Scor... → Train Binary Classifier → Save Model

### Train Binary Classifier

Model Training Script: **Neural Network for Classification**

- Target: **Converted**  
target, the target(label) to learn/predict
- Positive Class in Target: **Yes**  
Positive class in the target value. Default is Yes.
- Categorical Column Imputation: **Most Frequent**  
The mode method for categorical features to fill NA. Two options: mostFrequent and leastFrequent. Default is mostFrequent.
- Numerical Column Imputation: **Mean**  
The mode method for numeric features to fill NA. Four options: mean, max, min, median. Default is mean.
- Encoding Method: **Onehot**  
Encoding method.
- Maximum Null Value Percent: **80**  
Maximum Null Value Percent
- Train Partition Percent: **80**  
The percentage of original data used for training, default is 80%.
- Batch Size: **500**  
Batch Size to be used for model creation
- Optimizer Method: **Adam Optimizer**  
The optimizer method for cost minimization during creation of the neural network
- Activation Function: **Logistic Sigmoid**  
Type of activation function to be used for hidden layers of MLP

# Modeling in database and things that come very handy

```
CREATE TABLE DATA_4_MPO_HOUR_BAND
PARTITION BY LIST(CLANICA)
(
PARTITION PO VALUES ('651187'),
PARTITION VU VALUES ('651188'),
PARTITION BB VALUES ('651189'),
PARTITION SE VALUES ('651190'),
PARTITION EU VALUES ('651191'),
PARTITION VP VALUES ('651192'),
PARTITION KA VALUES ('651193'),
PARTITION ST VALUES ('651194'),
PARTITION MO VALUES ('651195'),
PARTITION TE VALUES ('651196'),
PARTITION ME VALUES ('651197'),
PARTITION LE VALUES ('651198'),
PARTITION PR VALUES ('651199')
)
AS
(SELECT DISTINCT p.k_clanica CLANICA,
p.k_clanica || p.k_ob
o.objekt_l0_clanicao_
c.clanica_l0_id || o.
a.artikal_l4_subcat_i
t.TIME_PART_OF_DAY HO
FROM BI_DW_PROD.F_PROD
BI_DW_PROD.D_ARTI
BI_DW_PROD.D_CLAN
BI_DW_PROD.D_OBJE
BI_DW_PROD.D_DATE
BI_DW_PROD.D_TIME
WHERE c.k_clanica = p.
AND p.k_objekt = o.k_o
AND a.k_artikal = p.k_
AND p.k_inv_date = d.k
AND p.k_inv_time = t.k
AND d.date_date > TO_C
AND a.artikal_l4_subca
```

## Create Association Rules model

```
%script

/* Drop model AR_F_PRODAJA */

BEGIN DBMS_DATA_MINING.DROP_MODEL('AR_4_MPO_HOUR_BAND');
EXCEPTION WHEN OTHERS THEN NULL; END;
/

DECLARE
v_setlst DBMS_DATA_MINING.SETTING_LIST;
BEGIN

v_setlst('ALGO_NAME') := 'ALGO_APRIORI_ASSOCIATION_RULES';
v_setlst('PREP_AUTO') := 'ON';
v_setlst('ASSO_MIN_SUPPORT') := '0.0001';
v_setlst('ASSO_MIN_CONFIDENCE') := '0.0001';
v_setlst('ASSO_MAX_RULE_LENGTH') := '2';
v_setlst('ODMS_ITEM_ID_COLUMN_NAME') := 'ARTIKAL';
v_setlst('ODMS_PARTITION_COLUMNS') := 'MPO,HOUR_BAND';
v_setlst('ODMS_MAX_PARTITIONS') := '3000';

DBMS_DATA_MINING.CREATE_MODEL2(
MODEL_NAME => 'AR_4_MPO_HOUR_BAND',
MINTNG_FUNCTION => 'ASSOCIATION',
DATA_QUERY => 'SELECT * FROM DATA_4_MPO_HOUR_BAND',
SET_LIST => v_setlst,
CASE_ID_COLUMN_NAME => 'RACUN');

END;
```

- Data stays where it is, in database
- SQL can be used at what it is the best
- Automatic data preparation is supported out-of-the-box
- Model partitioning allows building several models at the same time

# Oracle Analytics: scoring is done automatically



LEAD SCO... Apply Model Save Data

### Apply Model

Model QBX Lead Score NeuralNetwork

#### Outputs

Create	Output	Column Name
<input checked="" type="checkbox"/>	PredictedValue	PredictedValue
<input checked="" type="checkbox"/>	PredictionConfidence	PredictionConfidence

#### Parameters

Maximum Null Value Percent: 80  
 Compute lift and gain: Yes  
 Target column to compute lift: Converted  
 Positive class to compute lift: Yes

### LEAD SCORE NeuralNetwork

Binary Classification Model

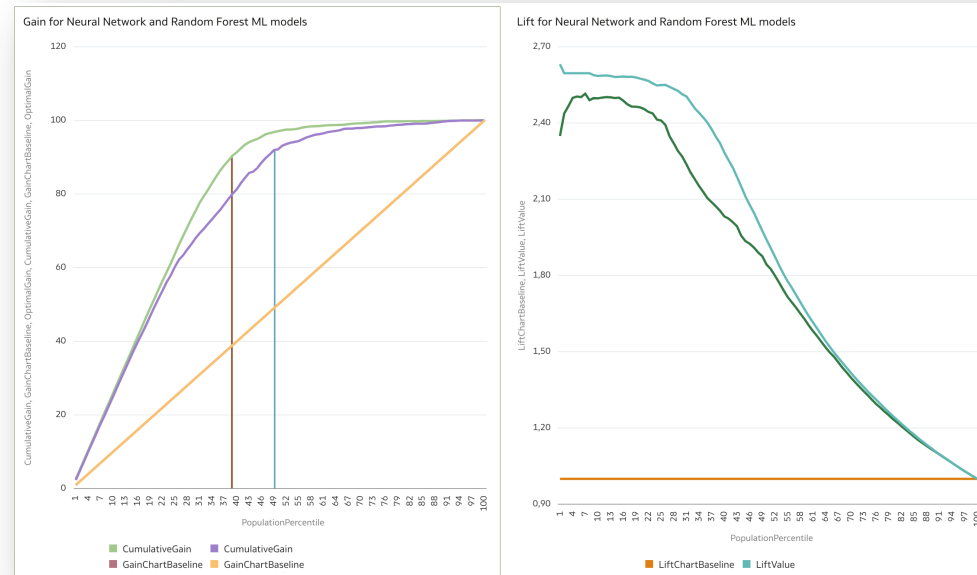
Positive Label for Converted: Yes

F1 Value: 0,90  
 Model Accuracy: 93 %  
 Precision: 92 %  
 Recall: 89 %  
 False Positive Rate: 5 %

#### Predicted Values

	Yes	No	Total
Yes	621	78	699 (38 %)
No	57	1092	1149 (62 %)
Total	678 (37 %)	1170 (63 %)	1848 (100 %)

Correct Prediction: 0 to 100%  
 Incorrect Prediction: 0 to 100%

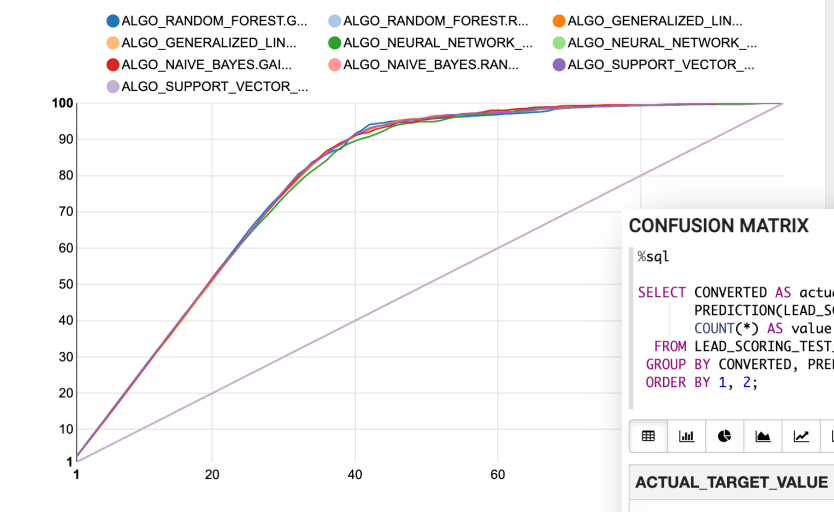


# OML: scoring is done in traditional way – using SQL



```
-- Score the data and compute lift
BEGIN
DBMS_DATA_MINING.APPLY('LEAD_SCORING_NN_MODEL',
'LEAD_SCORING_TEST_DATA',
'PROSPECT_ID',
'LEAD_SCORING_NN_RESULT');
DBMS_DATA_MINING.COMPUTE_LIFT('LEAD_SCORING_NN_RESULT',
'LEAD_SCORING_TEST_DATA',
'PROSPECT_ID',
'CONVERTED',
'LEAD_SCORING_NN_LIFT',
'Yes',
'PREDICTION',
'PROBABILITY',100);
```

```
%sql
SELECT A.ALGO_NAME, A.QUANTILE_NUMBER, ROUND(A.GAIN_CUMULATIVE*100,2) GAIN_CUMULATIVE, B
.RANDOM_GUESS
FROM LEAD_SCORING_ALL_LIFT_DATA A
, (SELECT ALGO_NAME, QUANTILE_NUMBER, QUANTILE_NUMBER RANDOM_GUESS FROM LEAD_SCORING_ALL_LIFT_DATA)
B
WHERE A.QUANTILE_NUMBER = B.QUANTILE_NUMBER AND A.ALGO_NAME = B.ALGO_NAME
```



**CONFUSION MATRIX**

```
%sql
SELECT CONVERTED AS actual_target_value,
PREDICTION(LEAD_SCORING_NN_MODEL USING *) AS predicted_target_value,
COUNT(*) AS value
FROM LEAD_SCORING_TEST_DATA
GROUP BY CONVERTED, PREDICTION(LEAD_SCORING_NN_MODEL USING *)
ORDER BY 1, 2;
```

ACTUAL_TARGET_VALUE	PREDICTED_TARGET_VALUE	VALUE
No	No	1670
No	Yes	129
Yes	No	147
Yes	Yes	1000

```
%sql
SELECT 'LEAD_SCORING_NN_MODEL' Model,
(TP+TN)/(TP+TN+FP+FN) Accuracy,
TP/(TP+FN) Precision,
TP/(TP+FP) Recall,
2 * (TP/(TP+FN)*TP/(TP+FP))/(TP/(TP+FN)+TP/(TP+FP)) F1_Score,
FN/(FN+TP) False_Positive_Rate
FROM (
with
tp as (SELECT tp_value FROM (SELECT CONVERTED AS actual_target_value,
PREDICTION(LEAD_SCORING_NN_MODEL USING *) AS predicted_target_value,
COUNT(*) AS tp_value
FROM LEAD_SCORING_TEST_DATA
GROUP BY CONVERTED, PREDICTION(LEAD_SCORING_NN_MODEL USING *)
ORDER BY 1, 2)
WHERE actual_target_value = 'Yes' and predicted_target_value = 'Yes'),
fp as (SELECT fp_value FROM (SELECT CONVERTED AS actual_target_value,
PREDICTION(LEAD_SCORING_NN_MODEL USING *) AS predicted_target_value,
COUNT(*) AS fp_value
FROM LEAD_SCORING_TEST_DATA
GROUP BY CONVERTED, PREDICTION(LEAD_SCORING_NN_MODEL USING *)
ORDER BY 1, 2)
WHERE actual_target_value = 'No' and predicted_target_value = 'Yes'),
fn as (SELECT fn_value FROM (SELECT CONVERTED AS actual_target_value,
PREDICTION(LEAD_SCORING_NN_MODEL USING *) AS predicted_target_value,
COUNT(*) AS fn_value
FROM LEAD_SCORING_TEST_DATA
GROUP BY CONVERTED, PREDICTION(LEAD_SCORING_NN_MODEL USING *)
ORDER BY 1, 2)
WHERE actual_target_value = 'Yes' and predicted_target_value = 'No'),
tn as (SELECT tn_value FROM (SELECT CONVERTED AS actual_target_value,
PREDICTION(LEAD_SCORING_NN_MODEL USING *) AS predicted_target_value,
COUNT(*) AS tn_value
FROM LEAD_SCORING_TEST_DATA
GROUP BY CONVERTED, PREDICTION(LEAD_SCORING_NN_MODEL USING *)
ORDER BY 1, 2)
WHERE actual_target_value = 'No' and predicted_target_value = 'No')
SELECT tp.tp_value TP, fp.fp_value FP, fn.fn_value FN, tn.tn_value TN
FROM tp, fp, fn, tn)
```

```
%sql
SELECT VIEW_NAME, VIEW_TYPE
FROM USER_MINING_MODEL_VIEWS
WHERE MODEL_NAME = 'LEAD_SCORING_RF_MODEL'
ORDER BY VIEW_NAME;
```

VIEW_NAME	VIEW_TYPE
DM\$VLEAD_SCORING_RF_MODEL	Variable Importance
DM\$VCLLEAD_SCORING_RF_MODEL	Scoring Cost Matrix
DM\$VGLLEAD_SCORING_RF_MODEL	Global Name-Value Pairs
DM\$VSLLEAD_SCORING_RF_MODEL	Computed Settings
DM\$VTLLEAD_SCORING_RF_MODEL	Classification Targets
DM\$VWLEAD_SCORING_RF_MODEL	Model Build Alerts

# Deployment



- In Oracle Analytics, models are ready to deploy instantly
- ML models that were created in database must be registered with Oracle Analytics first

← **Select a Model to Register**

Search

Type	Name		Name	
</>	GLM_HOUSING	Jul 28, 2021	LEADSCORING_BESTMODEL_RF	
</>	LEADSCORING_BESTMO...	Feb 3, 2022	Description	
</>	LEAD_SCORE_GLM	Jan 13, 2021	▶ Model Info	
</>	LEAD_SCORE_NNET	Apr 15, 2021	▶ Input Columns	
</>	LEAD_SCORE_SVM	Jan 13, 2021	▶ Output Columns	
</>	LEAD_SCORING_DECISI...	Apr 22, 2021	▲ Parameters	
</>	LEAD_SCORING_DT_MO...	Aug 5, 2021	target	CONVERTED
</>	LEAD_SCORING_GLM_M...	Aug 5, 2021	ALGO_NAME	ALGO_RANDOM
</>	LEAD_SCORING_NB_MO...	Aug 5, 2021	PREP_AUTO	ON
</>	LEAD_SCORING_RF_MO...	Aug 5, 2021	TREE_TERM_MINPCT_NODE	0.05
			TREE_TERM_MINREC_SPLIT	20
			ODMS_RANDOM_SEED	0
			TREE_IMPURITY_METRIC	TREE_IMPURITY
			CLAS_MAX_SUP_BINS	32
			CLAS_WEIGHTS_BALANCED	OFF
			TREE_TERM_MINPCT_SPLIT	0.1

Cancel Register

# From here on, the steps are the same ...



The screenshot shows a workflow with three steps: 'LEAD SCO...', 'Apply Model', and 'Save Data'. The 'Apply Model' step is expanded, showing the following configuration:

- Model:** LEAD SCORING Neural Network OAML
- Outputs:**
  - PredictedValue: Prediction Neural Network
  - PredictionConfidence: Prediction Confidence Neural Network
- Parameters:** (collapsed)
- Inputs:** (collapsed)

- Create a new data flow and Apply Model to generate a new “predicted data” dataset
- Use „predicted data” dataset in a new workbook

The screenshot shows the 'Create Scenario - Select Model' dialog box. The search filter is set to 'OAML'. A list of models is displayed with columns for Type, Name, Target, and Author.

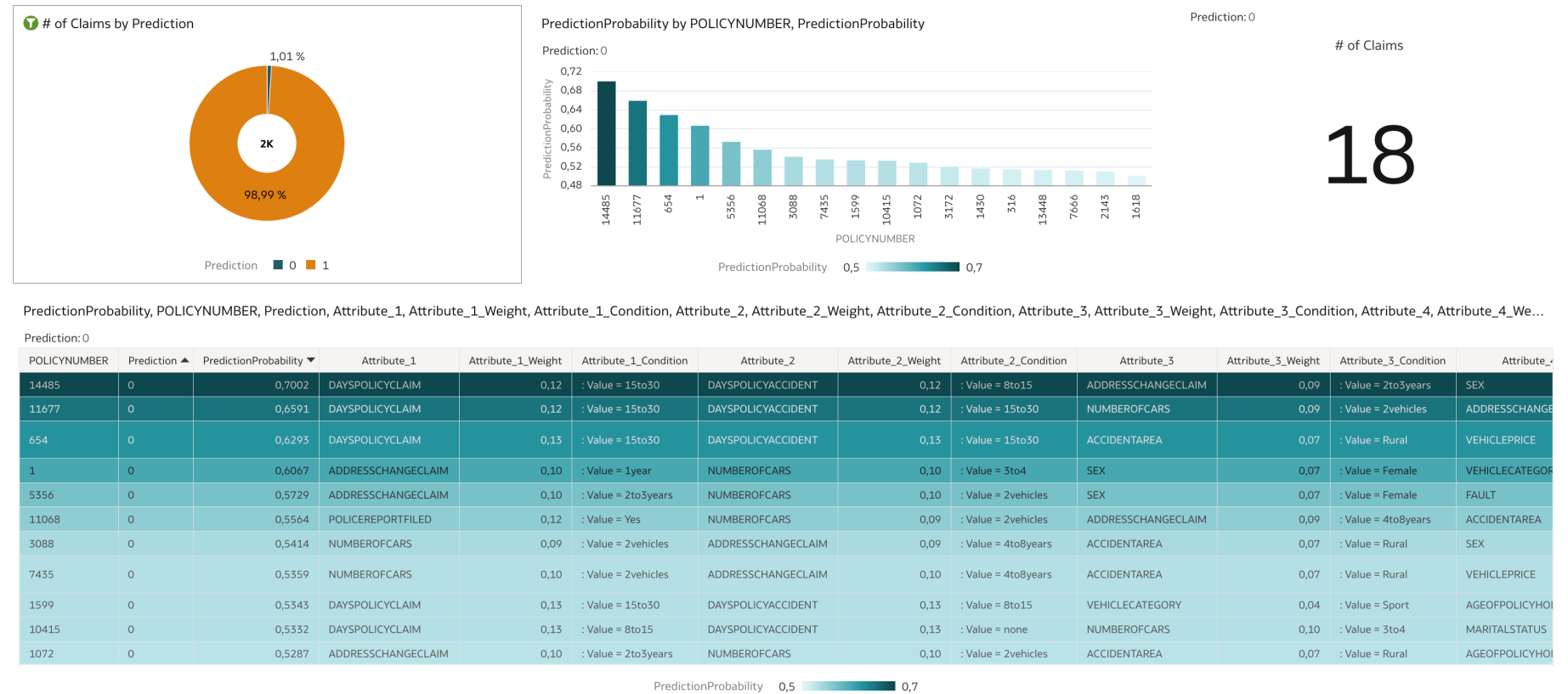
Type	Name	Target	Author
<input checked="" type="checkbox"/>	LEAD SCORING Neural Network OAML	Converted	ziga.vaupot@qubix.com
<input checked="" type="checkbox"/>	LEAD SCORING Random Forest OAML	Converted	ziga.vaupot@qubix.com
<input checked="" type="checkbox"/>	LEAD SCORING SVM-RBF OAML	Converted	ziga.vaupot@qubix.com
<input checked="" type="checkbox"/>	LEAD SCORING CART OAML	Converted	ziga.vaupot@qubix.com
<input checked="" type="checkbox"/>	LEAD SCORING Logistic Regression OAML	Converted	ziga.vaupot@qubix.com
<input checked="" type="checkbox"/>	LEAD SCORING Naive Bayes OAML	Converted	ziga.vaupot@qubix.com

Buttons: Cancel, OK

- Create a new workbook and add Scenario, which is ML model application on-the-fly
- This works only with Oracle Analytics ML models

# End results

Click here or drag data to add a filter



PredictionProbability, POLICYNUMBER, Prediction, Attribute\_1, Attribute\_1\_Weight, Attribute\_1\_Condition, Attribute\_2, Attribute\_2\_Weight, Attribute\_2\_Condition, Attribute\_3, Attribute\_3\_Weight, Attribute\_3\_Condition, Attribute\_4, Attribute\_4\_We...

Prediction: 0

POLICYNUMBER	Prediction	PredictionProbability	Attribute_1	Attribute_1_Weight	Attribute_1_Condition	Attribute_2	Attribute_2_Weight	Attribute_2_Condition	Attribute_3	Attribute_3_Weight	Attribute_3_Condition	Attribute_4
14485	0	0,7002	DAYSPOLICYCLAIM	0,12	: Value = 15to30	DAYSPOLICYACCIDENT	0,12	: Value = 8to15	ADDRESSCHANGECLAIM	0,09	: Value = 2to3years	SEX
11677	0	0,6591	DAYSPOLICYCLAIM	0,12	: Value = 15to30	DAYSPOLICYACCIDENT	0,12	: Value = 15to30	NUMBEROFCARS	0,09	: Value = 2vehicles	ADDRESSCHANGECLAIM
654	0	0,6293	DAYSPOLICYCLAIM	0,13	: Value = 15to30	DAYSPOLICYACCIDENT	0,13	: Value = 15to30	ACCIDENTAREA	0,07	: Value = Rural	VEHICLEPRICE
1	0	0,6067	ADDRESSCHANGECLAIM	0,10	: Value = 1year	NUMBEROFCARS	0,10	: Value = 3to4	SEX	0,07	: Value = Female	VEHICLECATEGORY
5356	0	0,5729	ADDRESSCHANGECLAIM	0,10	: Value = 2to3years	NUMBEROFCARS	0,10	: Value = 2vehicles	SEX	0,07	: Value = Female	FAULT
11068	0	0,5564	POLICEREPORTFILED	0,12	: Value = Yes	NUMBEROFCARS	0,09	: Value = 2vehicles	ADDRESSCHANGECLAIM	0,09	: Value = 4to8years	ACCIDENTAREA
3088	0	0,5414	NUMBEROFCARS	0,09	: Value = 2vehicles	ADDRESSCHANGECLAIM	0,09	: Value = 4to8years	ACCIDENTAREA	0,07	: Value = Rural	SEX
7435	0	0,5359	NUMBEROFCARS	0,10	: Value = 2vehicles	ADDRESSCHANGECLAIM	0,10	: Value = 4to8years	ACCIDENTAREA	0,07	: Value = Rural	VEHICLEPRICE
1599	0	0,5343	DAYSPOLICYCLAIM	0,13	: Value = 15to30	DAYSPOLICYACCIDENT	0,13	: Value = 8to15	VEHICLECATEGORY	0,04	: Value = Sport	AGEOFPOLICYHOLDER
10415	0	0,5332	DAYSPOLICYCLAIM	0,13	: Value = 8to15	DAYSPOLICYACCIDENT	0,13	: Value = none	NUMBEROFCARS	0,10	: Value = 3to4	MARITALSTATUS
1072	0	0,5287	ADDRESSCHANGECLAIM	0,10	: Value = 2to3years	NUMBEROFCARS	0,10	: Value = 2vehicles	ACCIDENTAREA	0,07	: Value = Rural	AGEOFPOLICYHOLDER



# AutoML

Bringing machine learning self-service to database (?)



ORACLE Machine Learning ML\_USER Project [ML\_USER Works... ML\_USER

## TEST LEAD SCORING

Experiment Settings Edit

Balanced Accuracy

Leader Board

Deploy	Rename	Create Notebook	Metrics
Algorithm	Model Name	Balanced Accuracy	
Random Forest	RF_840275841F	0.9395	
Naive Bayes	NB_A4284C0801	0.9208	
Support Vector Machine (Linear)	SVML_1949F4AACB	0.9146	
Neural Network	NN_BBFCCFFB08	0.9132	
Decision Tree	DT_A38FE4FF8F	0.8668	

Features

Refresh

Name	Importance	Type	Percent NULLs	Distinct Values	Min	Max
TAGS	<div style="width: 100%;"></div>	VARCHAR2				
TOTAL_TIME_SPENT_ON_WEBSITE	<div style="width: 100%;"></div>	NUMBER				

Running: 0h 2m

- Initialization Completed ✓
- Algorithm Selection Completed ✓
- Adaptive Sampling Completed ✓
- Feature Selection Completed ✓
- Model Tuning Completed ✓
- Random Forest Completed ✓
- Neural Network Completed ✓

- Automated ML model creation
  - Algorithm selection
  - Adaptive sampling
  - Feature extraction
  - Hyperparameter optimisation
- Basic understanding of machine learning process
  - However data (training dataset) preparation is required!
- Model registration with Oracle Analytics

# Final Thoughts

---

- Oracle Analytics can support business users/analysts by generating „good enough“ ML models. For more, in-database OML is required.
- Business users are often prepared to compensate speed of development with accuracy of predictions
- AutoML will bring ML in Oracle Analytics even closer to Business users
- ... but still: data needs proper preparation, collected, and the right question has to be asked

# Qubix

Enabling the Predictive Enterprise

Qubix