# Predicting Beer Ratings with in-database machine learning and APEX (in ATP)

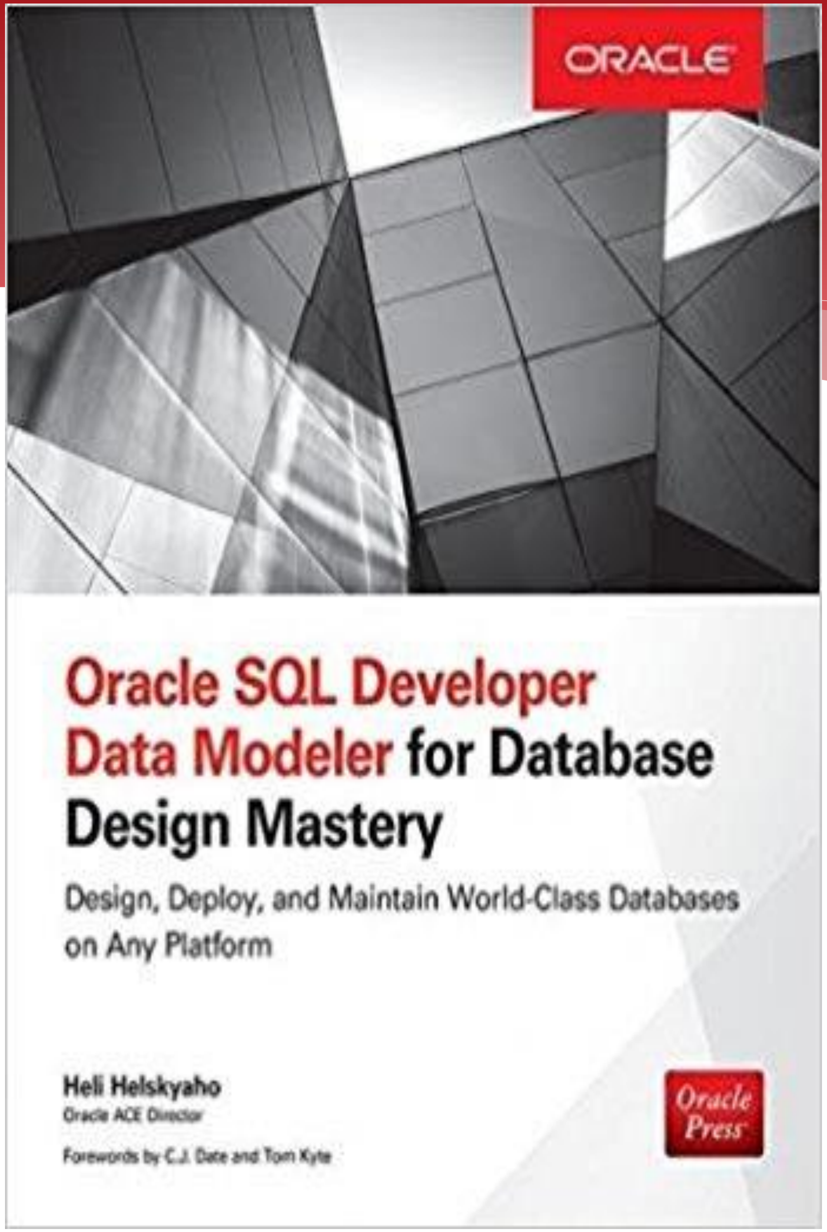
Heli Helskyaho
@HeliFromFinland

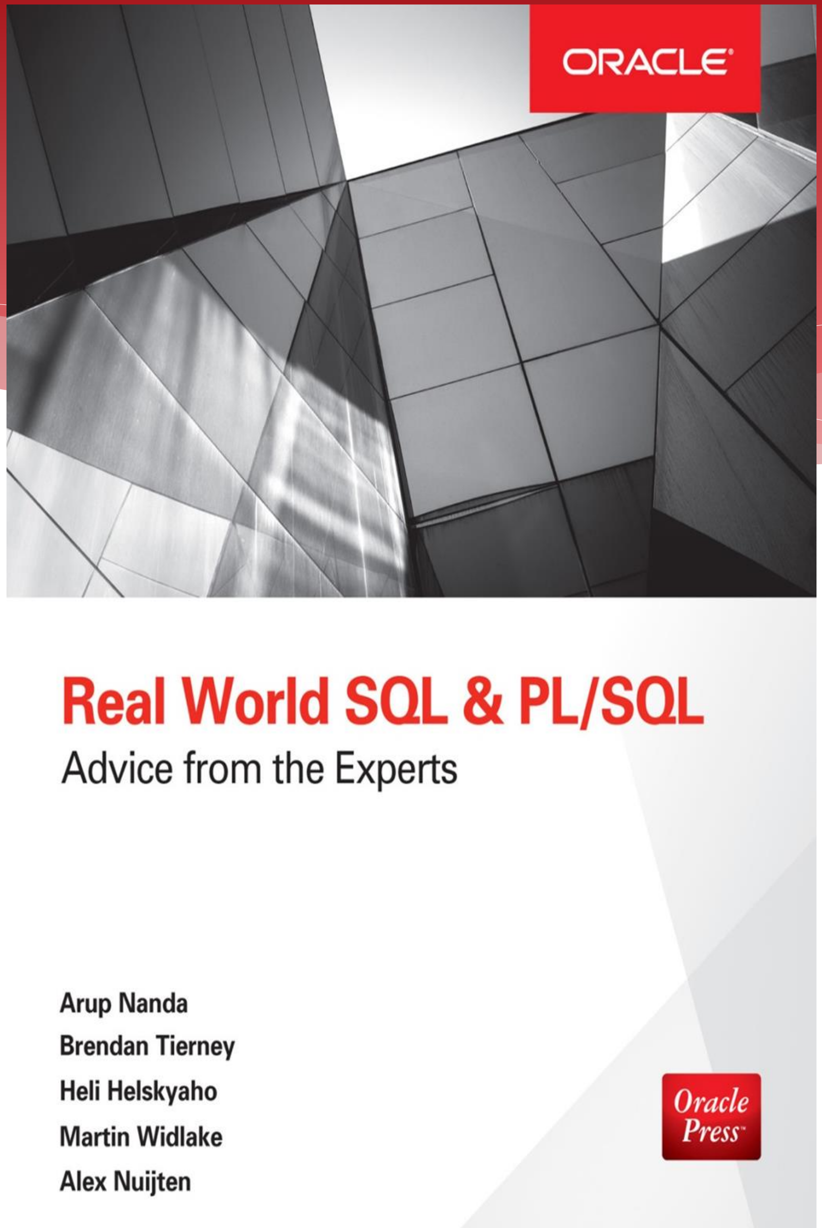
Miracle Finland Oy

# Introduction, Heli

* Graduated from University of Helsinki (Master of Science, computer science), currently a doctoral student, researcher and lecturer (databases, Big Data, Multi-model Databases, methods and tools for utilizing semi-structured data for decision making) at University of Helsinki
* Worked with Oracle products since 1993, worked for IT since 1990
* Data and Database!
* CEO for Miracle Finland Oy
* Oracle ACE Director, Oracle Groundbreaker Ambassador
* Ambassador for EOUC (EMEA Oracle Users Group Community)
* Listed as one of the TOP 100 influencers on IT sector in Finland (2015, 2016, 2017, 2018)
* Public speaker and an author
* Winner of Devvy for Database Design Category, 2015
* Author of the book Oracle SQL Developer Data Modeler for Database Design Mastery (Oracle Press, 2015), co-author for Real World SQL and PL/SQL: Advice from the Experts (Oracle Press, 2016)

MIRACLE
Miracle Finland Oy


ORACLE
Oracle SQL Developer
Data Modeler for Database
Design Mastery
Design, Deploy, and Maintain World-Class Databases
on Any Platform
Heli Helskyaho
Oracle ACE Director
Forewords by C.J. Date and Tom Kyte
Oracle Press


ORACLE
Real World SQL & PL/SQL
Advice from the Experts
Arup Nanda
Brendan Tierney
Heli Helskyaho
Martin Widlake
Alex Nuijten
Oracle Press

# 500+ Technical Experts Helping Peers Globally

ORACLE ACE Program

ORACLE ACE Director

ORACLE ACE

ORACLE ACE Associate

---

**3 Membership Tiers**

- Oracle ACE Director
- Oracle ACE
- Oracle ACE Associate

bit.ly/OracleACEProgram

**Connect:**

oracle-ace_ww@oracle.com

Facebook.com/oracleaces

@oracleace

Oracle Groundbreakers

Nominate yourself or someone you know: acenomination.oracle.com

# Why in-database?

* The *DATA*
  * Why to move the data?
    * It usually is a lot of data…
    * No hassle with moving the data
  * Preparing the data is the hard part (80/20)
    * Doing it in the database is the easiest way
    * Database is the natural environment for handling data

# Why in-database?

* The *SKILLS* needed already acquired
  * You already know SQL and PL/SQL

# Why in-database?

* The database has *APEX*
  * You can use APEX for visualizing the data and the process

Miracle Finland Oy

# Why in-database?

* The deployment
  * table, view, PL/SQL Package, Function, Procedure,… is easily used with many technologies
* …

# In-database Machine Learning

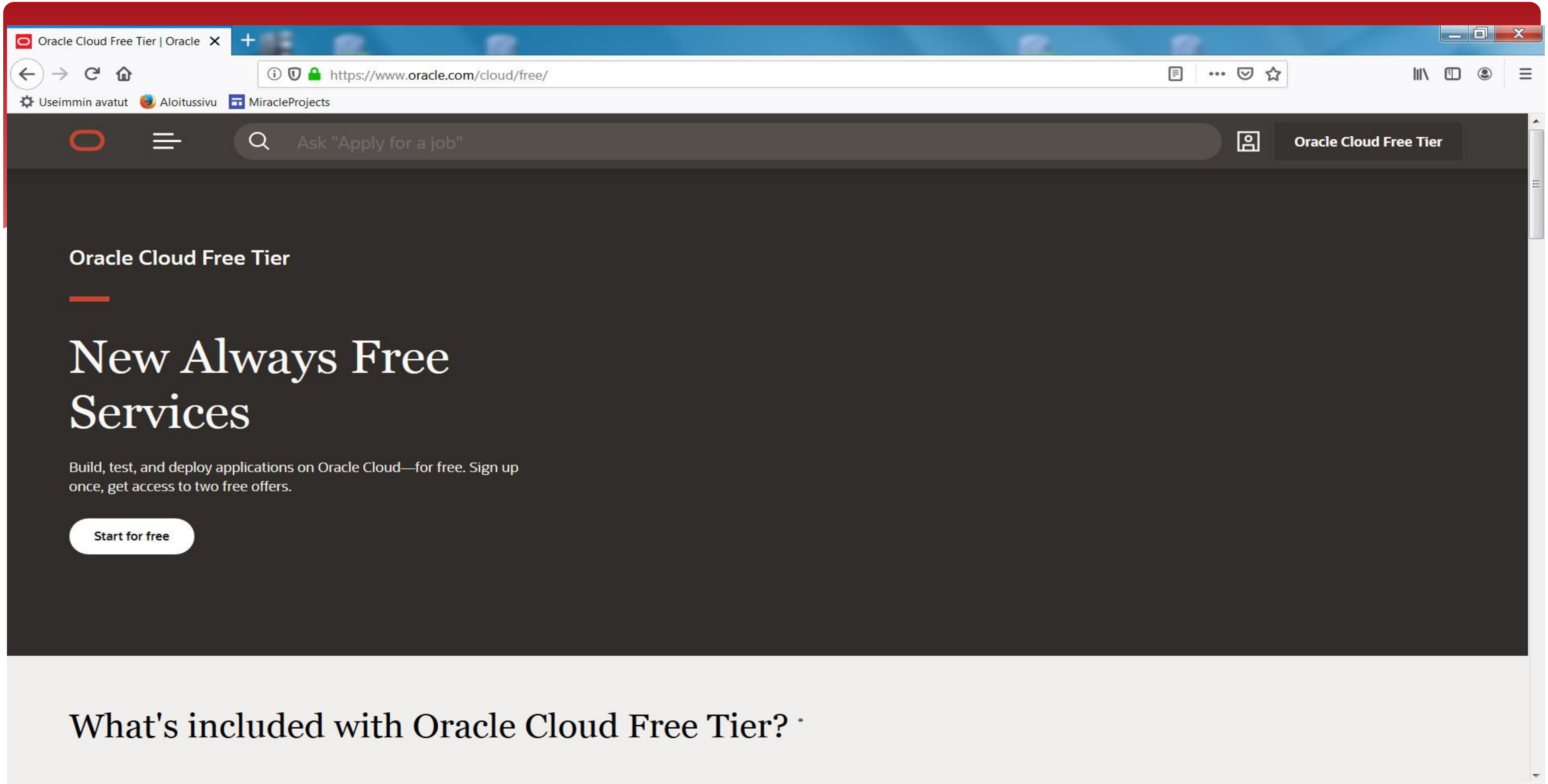
* “Oracle Machine Learning” is a Zeppelin based SQL notebook that is available with ADW and ATP
* Oracle Database Advanced Analytics (OAA) =
  * Oracle DB + Oracle Data Mining (ODM) (+Data Miner GUI in Oracle SQL Developer) +
  * Oracle R Enterprise (ORE)
* Predictive Analytics with Oracle Data Mining (ODM)
* Predictive Queries with Oracle Analytic Functions

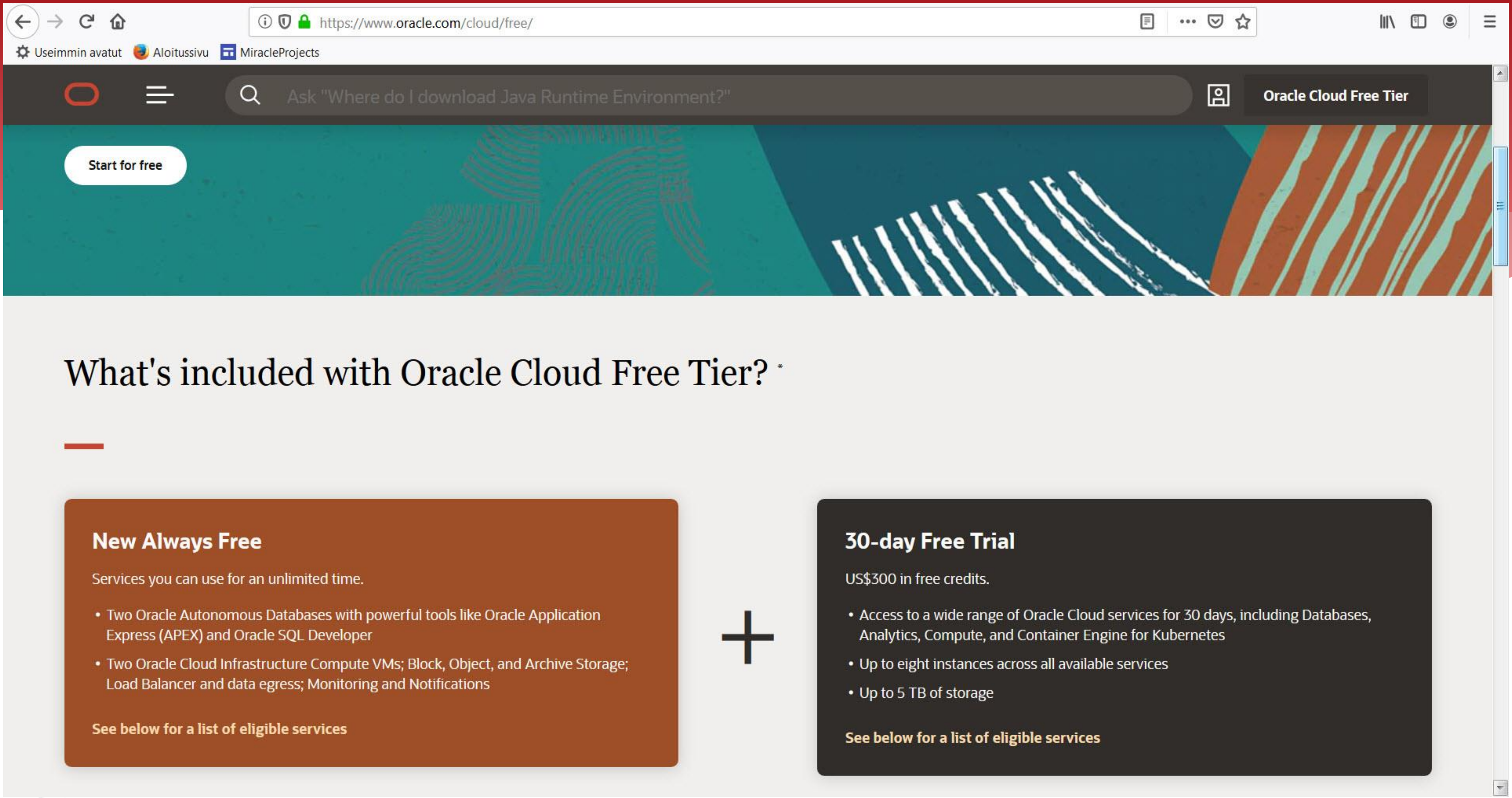
MIRACLE Miracle Finland Oy

# Advanced analytics (and ODM)

* is a **licensed product**
  * in the EE database separately licensed
  * in the Cloud included in: Database Service either High Performace Package or Extreme Performance Package

  * Make sure to check the licenses before using


Oracle Cloud Free Tier | Oracle
https://www.oracle.com/cloud/free/
Useimmin avatut
Aloitussivu
MiracleProjects
Ask "Apply for a job"
Oracle Cloud Free Tier
Oracle Cloud Free Tier
New Always Free Services
Build, test, and deploy applications on Oracle Cloud—for free. Sign up once, get access to two free offers.
Start for free
What's included with Oracle Cloud Free Tier? *


https://www.oracle.com/cloud/free/
Useimmin avatut
Aloitussivu
MiracleProjects
Ask "Where do I download Java Runtime Environment?"
Oracle Cloud Free Tier
Start for free
What's included with Oracle Cloud Free Tier? *
New Always Free
Services you can use for an unlimited time.
• Two Oracle Autonomous Databases with powerful tools like Oracle Application Express (APEX) and Oracle SQL Developer
• Two Oracle Cloud Infrastructure Compute VMs; Block, Object, and Archive Storage; Load Balancer and data egress; Monitoring and Notifications
See below for a list of eligible services
+
30-day Free Trial
US$300 in free credits.
• Access to a wide range of Oracle Cloud services for 30 days, including Databases, Analytics, Compute, and Container Engine for Kubernetes
• Up to eight instances across all available services
• Up to 5 TB of storage
See below for a list of eligible services

Oracle Cloud Infrastructure

# New Free Tier

oracle.com/gbtour

## Always Free

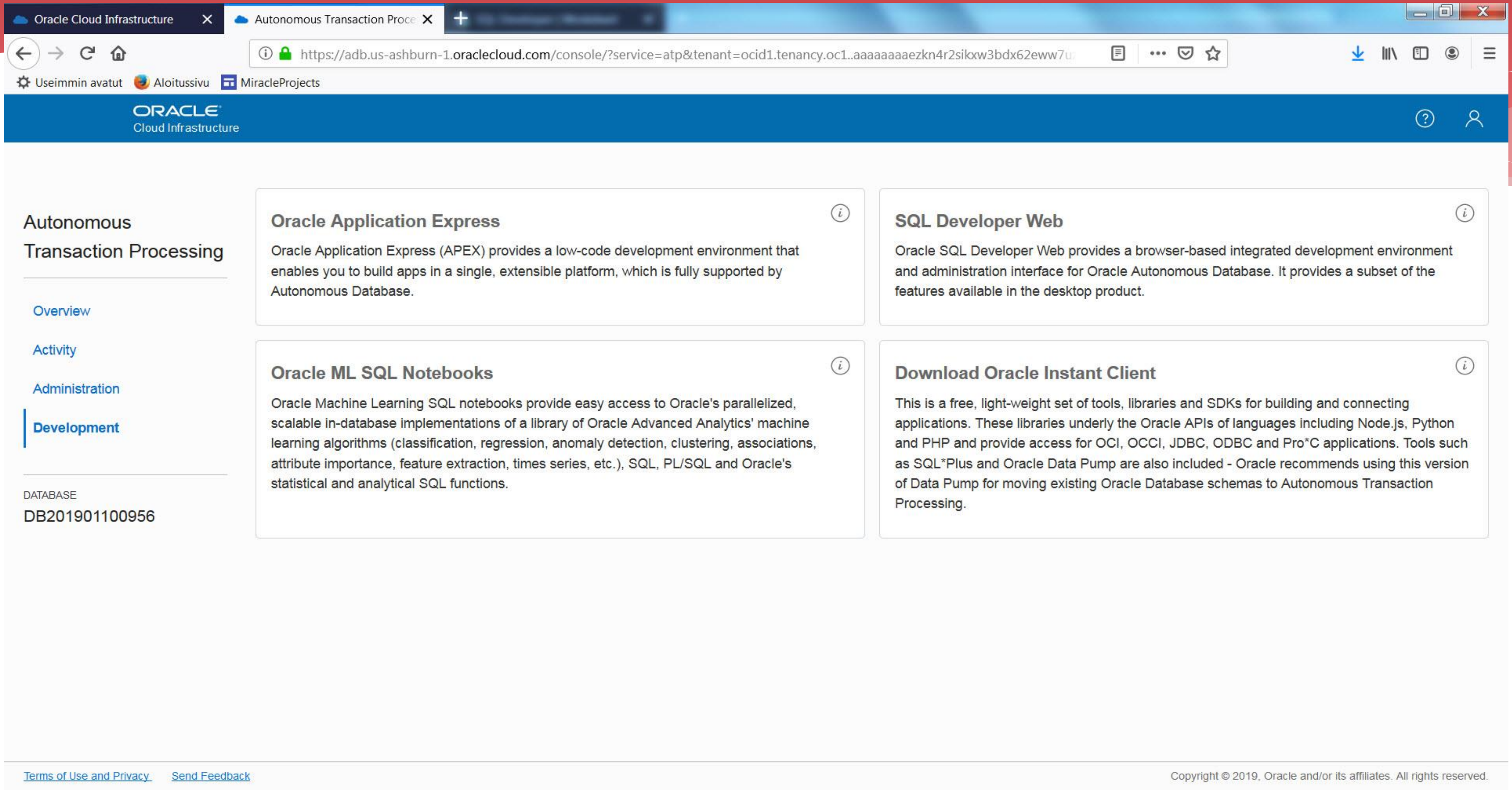
Services you can use for unlimited time

+

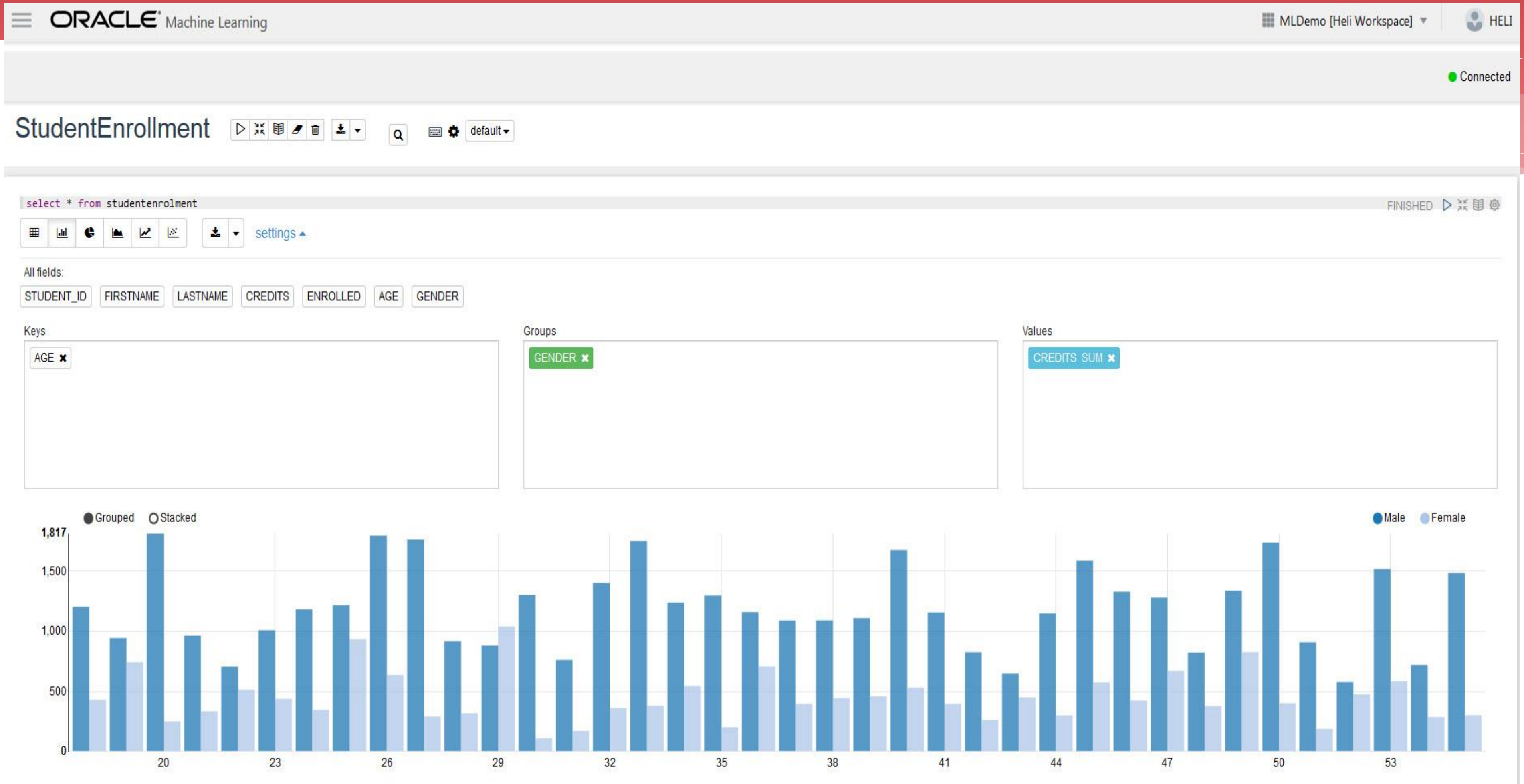
## 30-Day Free Trial

Free credits you can use for more services

# APEX, ORDS, SQL Developer in ATP and ADW



# Oracle Machine Learning in ATP and ADW




MIRACLE
Miracle Finland Oy

# Oracle DB and Python

* Python
* Cx_Oracle (pip install)
* Oracle Instant Client
  * same version 32/64 as your Python
  * add to the path
  * 19 will not work with Windows 7, 18 does

* Oracle DB/ATP/ADW (TNS_ADMIN, instant client… /network/admin)
  * Sqlnet.ora
  * Tnsnames.ora
  * Wallet

Miracle Finland Oy

# Data Dictionary Views for ODM

*Table 2-1 Data Dictionary Views for Oracle Data Mining*

| View Name | Description |
|---|---|
| ALL_MINING_MODELS | Provides information about all accessible mining models |
| ALL_MINING_MODEL_ATTRIBUTES | Provides information about the attributes of all accessible mining models |
| ALL_MINING_MODEL_PARTITIONS | Provides information about the partitions of all accessible partitioned mining models |
| ALL_MINING_MODEL_SETTINGS | Provides information about the configuration settings for all accessible mining models |
| ALL_MINING_MODEL_VIEWS | Provides information about the model views for all accessible mining models |
| ALL_MINING_MODEL_XFORMS | Provides the user-specified transformations embedded in all accessible mining models. |

# Machine Learning in short

* Supervised learning
  * *Regression*, how much? How many?, numbers
  * *Classification*, Is it this or that?, groups/classes
  * *Anamoly detection (classification)*, Is this weird?
* Unsupervised learning
  * *Clustering*, Which groups? Group them in 5 groups
* Reinforcement learning
  * which action? feedback positive/negative

Miracle Finland Oy

# Machine Learning in short

* Use the right *Features*
  * with right Algorithms
    * to build the right *Models*
      * that achieve the right *Tasks*

# Machine Learning

1. Defining the Task, understanding the Task
2. Collecting the data, understanding the data
   1. Attributes (Features, Columns)
   2. Exploring
   3. Preparing the data/Transforming the data
3. Creating a Model
   1. Model (Function)
   2. Algorithm
4. Train, test (validate)
5. Evaluate the models
6. Scoring and Deployment

# Beer rating

# Define the Task

* The Task: ***Predict the overall rating for a beer***
  * ***Find the beers that will most likely be rated as 5***
  * ***Find the brewers that most likely will make beer rated as 5***

* We will use *Supervised Learning* and *Classification*, our *target* attribute is OVERALL (values 1-5)

# Collecting the data

* www.kaggle.com/c/beer-ratings/data

* *Beer_train*, all know input and output
* (*Beer_test*, only known input, output unknown)

* In the demo we will use *Beer_train but we call it Beer_data*
* For deployment we will use *Beer_Test*

MIRACLE Miracle Finland Oy

# The data

* For Supervised Learning the data must be divided in *two* sets:
  * one for *training*  the other one for *testing* the model really works (validation).
  * In the demo we will split *Beer_data in to two sets:*
    * *Beer_training_data*
    * *Beer_testing_data*

* The data requirements for ODM:
  * Data must be stored in a *single table or view* (***a case table***)
  * Each record must be stored in a *separate row as* **a case**
  * Each case can (optionally) be identified by a *unique case ID*

Miracle Finland Oy

# The data for Supervised Learning

* Beer_data -> *Beer_training_data, Beer_testing_data*

```
-- create training set
CREATE TABLE Beer_training_data
AS SELECT * FROM Beer_data
WHERE ORA_HASH (IDIndex, 99, 5) < 65;

-- create testing set
CREATE TABLE Beer_testing_data
AS SELECT * FROM Beer_data
WHERE ORA_HASH(IDIndex, 99, 5) >= 65;
```

# A quick look to APEX (demo)

* Beer_Train
* Beer_Training_Data
* Beer_Testing_Dat

# Oracle PL/SQL Packages for Data Mining

* DBMS_PREDICTIVE_ANALYTICS
  * Routines for performing *predictive analytics*
* DBMS_DATA_MINING_TRANSFORMING
  * Routines for *transforming the data* for mining models
* DBMS_DATA_MINING
  * Routines for *creating and managing mining models*

# DBMS_PREDICTIVE_ANALYTICS

* routines that perform an automated data mining known as *predictive analytics*
* no need to be aware of model building or scoring
  * All mining activities are handled internally by the procedure.

# DBMS_PREDICTIVE_ANALYTICS

* EXPLAIN
  * *ranks attributes* in order of influence in explaining the target column
* PREDICT
  * *predicts the value of a target* column based on values in the input data
* PROFILE
  * *generates rules* that describe the cases from the input data

Miracle Finland Oy

# DBMS_PREDICTIVE_ANALYTICS

```
SQL> describe dbms_predictive_analytics

PROCEDURE EXPLAIN
 Argument Name                  Type                    In/Out Default?
 ------------------------------ ----------------------- ------ --------
 DATA_TABLE_NAME                VARCHAR2                IN
 EXPLAIN_COLUMN_NAME            VARCHAR2                IN
 RESULT_TABLE_NAME              VARCHAR2                IN
 DATA_SCHEMA_NAME               VARCHAR2                IN     DEFAULT

PROCEDURE PREDICT
 Argument Name                  Type                    In/Out Default?
 ------------------------------ ----------------------- ------ --------
 ACCURACY                       NUMBER                  OUT
 DATA_TABLE_NAME                VARCHAR2                IN
 CASE_ID_COLUMN_NAME            VARCHAR2                IN
 TARGET_COLUMN_NAME             VARCHAR2                IN
 RESULT_TABLE_NAME              VARCHAR2                IN
 DATA_SCHEMA_NAME               VARCHAR2                IN     DEFAULT

PROCEDURE PROFILE
 Argument Name                  Type                    In/Out Default?
 ------------------------------ ----------------------- ------ --------
 DATA_TABLE_NAME                VARCHAR2                IN
 TARGET_COLUMN_NAME             VARCHAR2                IN
 RESULT_TABLE_NAME              VARCHAR2                IN
 DATA_SCHEMA_NAME               VARCHAR2                IN     DEFAULT
```

# EXPLAIN

```
BEGIN
    DBMS_PREDICTIVE_ANALYTICS.EXPLAIN(
        data_table_name      => 'Beer_training_data',
        explain_column_name  => 'overall',
        result_table_name    => 'beer_explain');
END;
/
```

Miracle Finland Oy

# EXPLAIN

Welcome Page | BEER_EXPLAIN - Structure | HR 12.2 | BEER_EXPLAIN

Columns | Data | Model | Constraints | Grants | Statistics | Triggers | Flashback | Dependencies | Details | Partitions | Indexes | SQL

Sort.. | Filter: | Actions...

| | ATTRIBUTE_NAME | ATTRIBUTE_SUBNAME | EXPLANATORY_VALUE | RANK |
|---|---|---|---|---|
| 1 | TASTE | (null) | 0,1369959018036639046488620435087567350302 | 1 |
| 2 | PALATE | (null) | 0,1033038106866505612466700086317885929612 | 2 |
| 3 | AROMA | (null) | 0,08791526829406451506523536256475774282 | 3 |
| 4 | APPEARANCE | (null) | 0,0597994415746425094084278968669637963851 | 4 |
| 5 | IDINDEX | (null) | 0,0514743174015554311246566591668580953978 | 5 |
| 6 | STYLE | (null) | 0,0487182850810693894469596131763997020818 | 6 |
| 7 | BEERID | (null) | 0,0447651730492723089511095141894142199002 | 7 |
| 8 | ABV | (null) | 0,0282744648145592050251039059166762689924 | 8 |
| 9 | BREWERID | (null) | 0,0273217823948269630249887964806943058078 | 9 |
| 10 | TEXT | (null) | 0 | 10 |
| 11 | PROFILENAME | (null) | 0 | 10 |
| 12 | NAME | (null) | 0 | 10 |
| 13 | GENDER | (null) | 0 | 10 |
| 14 | AGEINSECONDS | (null) | 0 | 10 |
| 15 | BIRTHDAYRAW | (null) | 0 | 10 |
| 16 | TIMEUNIX | (null) | 0 | 10 |
| 17 | TIMESTRUCT | (null) | 0 | 10 |
| 18 | BIRTHDAYUNIX | (null) | 0 | 10 |

Miracle Finland Oy

# PREDICT

```
DECLARE
p_accuracy NUMBER(10,9);
BEGIN
  DBMS_PREDICTIVE_ANALYTICS.PREDICT(
       accuracy                => p_accuracy,
       data_table_name         =>'Beer_training_data',
       case_id_column_name     =>'idindex',
       target_column_name      =>'overall',
       result_table_name       => 'Beer_predict');
  DBMS_OUTPUT.PUT_LINE('Accuracy: ' || p_accuracy);
END;
/
```

Accuracy: .24618951 (a measure of improved maximum average accuracy versus a naive model's maximum average accuracy)

# PREDICT

| IDINDEX | PREDICTION | PROBABILITY |
|---|---|---|
| 0 | 2 | .42799808967620911 |
| 1 | 3 | .53057676003601528 |
| 2 | 3 | .51728627054274079 |
| 3 | 4 | .45808326842381863 |
| 4 | 4 | .47447188708319082 |
| 5 | 4 | .64845475978174982 |
| 6 | 4 | .65424415909731026 |
| 7 | 3 | .45898266110126107 |
| 9 | 4 | .51363296269020753 |

Miracle Finland Oy

# How did it predict?

* Total no of rows: 37 303
* Correct predictions: 25 421
* Not correct predictions: 11 882

# PROFILE

* creates a Decision Tree model
  * to identify the characteristics of the *attributes* that predict the target
  * creates rules (expressed in XML as if-then-else statements) that describe the *decisions* that affect the prediction
* PROFILE returns XML that is derived from the model details generated by the algorithm.

# PROFILE

```
BEGIN
    DBMS_PREDICTIVE_ANALYTICS.PROFILE(
         DATA_TABLE_NAME    => ’Beer_training_data',
         TARGET_COLUMN_NAME => 'overall',
         RESULT_TABLE_NAME  =>
'beer_profile_result');
END;
/
```

# PROFILE

| BEER_PROFILE_RESULT | | |
|---|---|---|
| **PROFILE_ID** | **RECORD_COUNT** | **DESCRIPTION** |
| 1 | 50 | [SYS.XMLTYPE] |
| 2 | 100 | [SYS.XMLTYPE] |
| 3 | 34 | [SYS.XMLTYPE] |
| 4 | 462 | [SYS.XMLTYPE] |
| 5 | 566 | [SYS.XMLTYPE] |
| 6 | 279 | [SYS.XMLTYPE] |
| 7 | 2445 | [SYS.XMLTYPE] |
| 8 | 1624 | [SYS.XMLTYPE] |
| 9 | 1522 | [SYS.XMLTYPE] |
| 10 | 653 | [SYS.XMLTYPE] |
| 11 | 8024 | [SYS.XMLTYPE] |
| 12 | 3557 | [SYS.XMLTYPE] |

# Now to the process itself

# Preparing the data

* This is usually the most difficult and time consuming part of machine learning… 80/20…

# Attributes

- **Data attributes**
  - columns in the data set used to build, test, or score a model
- **Model attributes**
  - the data representations used internally by the model
- The **target attribute**
  - in supervised learning contains the *known values* of output and to which the predictions are compared to
- *Identify* the columns (features) to include in the case table

Miracle Finland Oy

# Understanding the data

* go through the data and each attribute and feature
* datatype (numeric, category, ...)
  * note Palate/aroma/… is not numeric, it is categorial!
* type, expectation, missing, outliers, comment
* linear plot to find data to be removed (outlier)
* a correlation plot to show the correlation
* remove nulls on very significant features (somehow)
* transform features
* …

Miracle Finland Oy

# Understanding the data

* A demo with APEX or Oracle Machine learning (ATP)

Miracle Finland Oy

# Understanding the data

* The Target is **OVERALL**
* 0 value in Overall (1-5?)?
  * Remove?
  * Accept (0-5)?
  * …

# A Feature/Attribute

* Is a Feature/Attribute relevant?
* Is a Feature/Attribute independent? (the area vs no of rooms)
* Is a Feature/Attribute simple? (how to compare? GPS coordinates not good, Convert for intance distance from water)

# Let’s talk about some attributes in our example

* APPEARANCE,
* AROMA,
* PALATE,
* TASTE,
* BEERID,
* TEXT,
* TIMESTRUCT,
* TIMEUNIX,
* AGEINSECONDS,
* BIRTHDAYRAW,
* BIRTHDAYUNIX,
* PROFILENAME,
* NAME,
* GENDER

# Missing Data

* Missing Values or Sparse Data?
  * Missing values
    * some attribute values are *unknown*
    * missing values in columns with a simple data type
  * Sparse data
    * values that are *assumed to be known*, although they are not represented in the data
    * missing values in nested columns

Miracle Finland Oy

# Transforming the data

* Creating Nested Columns
  * if you want to include transactional data etc.
* Converting Column Data Types
  * Age -> Child, Adult
* Business and Domain-Sensitive Transformations
  * Date of birth -> age
* Text Transformation (a text column must be in a table, not a view)
* …
* Write SQL expressions for any transformations not handled by ADP

MIRACLE Miracle Finland Oy

# DBMS_DATA_MINING_TRANSFORM

# DBMS_DATA_MINING_TRANSFORM

Understand the routines of DBMS_DATA_MINING_TRANSFORM package.

The DBMS_DATA_MINING_TRANSFORM package contains routines that perform data transformations such as binning, normalization, and outlier treatment. The package includes routines for:

- Specifying transformations in a format that can be embedded in a mining model.
- Specifying transformations as relational views (external to mining model objects).
- Specifying distinct properties for columns in the build data. For example, you can specify that the column must be interpreted as unstructured text, or that the column must be excluded from Automatic Data Preparation.

# Normalization

* Normalization
  * A technique for *reducing the range of numerical data*
  * (Model View Details: DM$VN for Normalization)

# Outlier Treatment

* A value is considered an *outlier* if it deviates *significantly* from most other values in the column
* Outliers can have a skewing effect on the data
* Outliers can interfere with the effectiveness of transformations such as normalization or binning.
* problematic or perfectly valid data?

# Creating a Model

# DBMS_DATA_MINING

Understand the routines of `DBMS_DATA_MINING` package.

The `DBMS_DATA_MINING` package contains routines for creating mining models, for performing operations on mining models, and for querying mining models. The package includes routines for:

- Creating, dropping, and performing other DDL operations on mining models
- Obtaining detailed information about model attributes, rules, and other information internal to the model (model details)
- Computing test metrics for classification models
- Specifying costs for classification models
- Exporting and importing models
- Building models using Oracle's native algorithms as well as algorithms written in R

# Creating a Model

* Choose the mining function
* Choose the algorithm
* Create and populate the settings table

# Choose the Model (Mining Function)

* Supervised Learning
  * Regression
  * Classification
  * Anomaly Detection
  * (Feature Selection)
* Unsupervised Learning
  * Clustering
  * Anomaly Detection
  * Association
  * (Feature Extraction)

MIRACLE
Miracle Finland Oy

# Choosing the Algorithm

* Decision Tree (classification)
* Naive Bayes (classification)
* Generalized Linear Models (regression and classification)
* Support Vector Machines (classification, regression, and anomaly detection)
* k-Means (clustering)
* O-Cluster (clustering)
* Minimum Description Length (for calculating attribute importance)
* Apriori (for calculating association rules)
* Non-Negative Matrix Factorization, NMF (feature extraction)
* … Each version brings more algorithms to choose from…

*Table 5-3 Data Mining Algorithms*

| ALGO_NAME Value | Algorithm | Default? | Mining Model Function |
|---|---|---|---|
| ALGO_AI_MDL | Minimum Description Length | — | attribute importance |
| ALGO_APRIORI_ASSOCIATION_RULES | Apriori | — | association |
| ALGO_CUR_DECOMPOSITION | CUR Decomposition | | Attribute Importance |
| ALGO_DECISION_TREE | Decision Tree | — | classification |
| ALGO_EXPECTATION_MAXIMIZATION | Expectation Maximization | | |
| ALGO_EXPLICIT_SEMANTIC_ANALYS | Explicit Semantic Analysis | — | feature extraction<br>classification |
| ALGO_EXPONENTIAL_SMOOTHING | Exponential Smoothing | — | time series |
| ALGO_EXTENSIBLE_LANG | Language used for extensible algorithm | — | All mining functions are supported |
| ALGO_GENERALIZED_LINEAR_MODEL | Generalized Linear Model | — | classification and regression |
| ALGO_KMEANS | *k*-Means | yes | clustering |
| ALGO_NAIVE_BAYES | Naive Bayes | yes | classification |
| ALGO_NEURAL_NETWORK | Neural Network | — | classification |
| ALGO_NONNEGATIVE_MATRIX_FACTOR | Non-Negative Matrix Factorization | yes | feature extraction |
| ALGO_O_CLUSTER | O-Cluster | — | clustering |
| ALGO_RANDOM_FOREST | Random Forest | — | classification |
| ALGO_SINGULAR_VALUE_DECOMP | Singular Value Decomposition (can also be used for Principal Component Analysis) | — | feature extraction |
| ALGO_SUPPORT_VECTOR_MACHINES | Support Vector Machine | yes | default regression algorithm<br>regression, classification, and anomaly detection (classification with no target) |

MIRACLE
Miracle Finland Oy

# Create a model

* CREATE_MODEL procedure in the DBMS_DATA_MINING package

```
PROCEDURE CREATE_MODEL(
                  model_name            IN VARCHAR2,
                  mining_function       IN VARCHAR2,
                  data_table_name       IN VARCHAR2,
                  case_id_column_name   IN VARCHAR2,
                  target_column_name    IN VARCHAR2 DEFAULT NULL,
                  settings_table_name   IN VARCHAR2 DEFAULT NULL,
                  data_schema_name      IN VARCHAR2 DEFAULT NULL,
                  settings_schema_name  IN VARCHAR2 DEFAULT NULL,
                  xform_list            IN TRANSFORM_LIST DEFAULT NULL);
```

# A Settings table

```
CREATE TABLE Beer_settings_DT (
setting_name  VARCHAR2(30),
setting_value VARCHAR2(4000));


BEGIN
 INSERT INTO Beer_settings_DT VALUES
 (dbms_data_mining.algo_name, dbms_data_mining.algo_decision_tree);
 …
END;
/
```

# Other possible settings

* Cost table and matrix (Decision Tree model)
* Prior Probabilities (Naive Bayes)
* Class Weights (Logistic Regression or Support Vector Machine)
* …

# Create a new model

```
BEGIN
  DBMS_DATA_MINING.CREATE_MODEL(
    model_name          => 'Beer_DT',
    mining_function     => dbms_data_mining.classification,
    data_table_name     => 'Beer_training_data',
    case_id_column_name => 'IDIndex',
    target_column_name  => 'Overall',
    settings_table_name => 'Beer_settings_DT');
END;
/
```

# Model Signature

* The **set of data attributes** that are used to build a model

# Model Signature

```
SELECT attribute_name, attribute_type
FROM TABLE(DBMS_DATA_MINING.GET_MODEL_SIGNATURE('BEER_DT'))
ORDER BY attribute_name;
```

| | ATTRIBUTE_NAME | ATTRIBUTE_TYPE |
|---|---|---|
| 1 | ABV | NUMBER |
| 2 | BREWERID | NUMBER |
| 3 | STYLE | VARCHAR2 |

# Testing and Evaluating the model

* Test: the model with new data (known input, known output)
* Evaluation: depends on the chosen metrics

# Accuracy

* In our task we have chosen that only accuracy is important (this is a simple demo)

# Evaluation

* Demo with APEX or ATP
  * (Confusion matrices, Beer model comparison,  Beer evaluation of models)

* What algorithm is the best?

# Evaluation

* In our simple example none of the models was very good but Decision Tree was a little bit better than others
* We are now happy with our model and ready to deploy it

# Scoring and Deployment

* Deployment is implementing the models in the target environment
* Deployment
  * Moving a model from the database where it was built to the database where it will be used (export/import)
  * With scoring data either for real-time  or batch results
  * Extracting model details to produce reports (clustering rules, decision tree rules, … )

MIRACLE
Miracle Finland Oy

# Apply the Model

```
BEGIN
       DBMS_DATA_MINING.APPLY (
       model_name      => 'beer_DT',
       data_table_name     => 'beer_test',
       case_id_column_name => 'idindex',
       result_table_name   => 'beer_result_table_DT');
END;
/
```

Miracle Finland Oy

BEER_RESULT_TABLE_DT

Indexes Model Constraints Grants Statistics UI Defaults Triggers Dependencies SQL REST

t Rows Insert Row

| IDINDEX | PREDICTION | PROBABILITY | COST |
|---|---|---|---|
| 13803 | 4 | .61222339304531082 | .38777660695468918 |
| 13803 | 5 | .25594192717480391 | .74405807282519609 |
| 13803 | 3 | .11017445264020606 | .88982554735979391 |
| 13803 | 2 | .01803067556492214 | .98196932443507789 |
| 13803 | 1 | .00362955157475704544 | .99637044842524292 |
| 13803 | 0 | 0 | 1 |
| 13960 | 4 | .61222339304531082 | .38777660695468918 |
| 13960 | 5 | .25594192717480391 | .74405807282519609 |
| 13960 | 3 | .11017445264020606 | .88982554735979391 |

# Real-time scoring a prediction, single record scoring

*What is the probability for beer 43548 to get overall 5?*

```
SELECT PREDICTION_PROBABILITY(Beer_DT, 5 USING *) beer_overall_prob
FROM beer_test
WHERE idindex = 43658;
```

1.5087463556851313E-001

```
SELECT PREDICTION_PROBABILITY(Beer_DT, 5 USING STYLE) beer_overall_prob
FROM beer_test
WHERE idindex = 43658;
```

2.4140018157974377E-001

USING: Predictors and/or Expressions (only if

# Real-time scoring a prediction, batch scoring

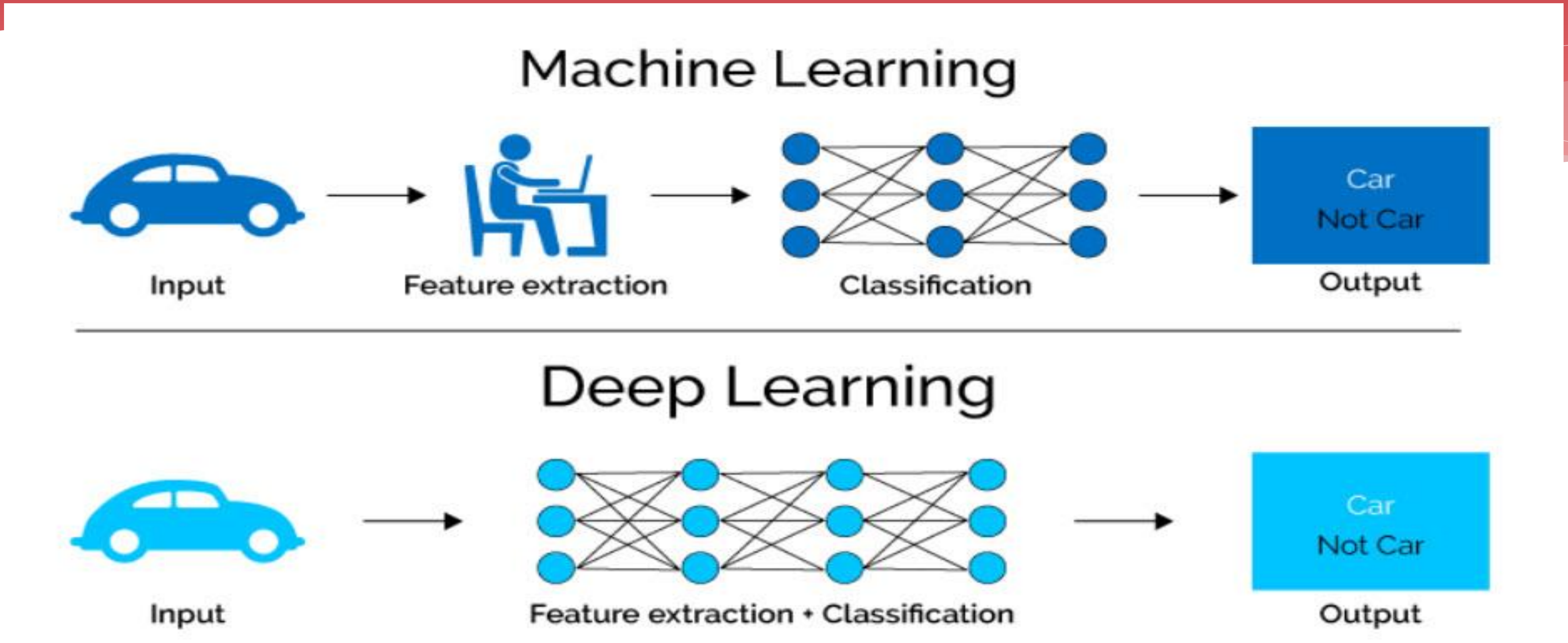
What are the brewers that I should contact (making most of the overall 5 beers)?

```
SELECT brewerid, count(*) as cnt
FROM beer_test
WHERE PREDICTION(Beer_DT USING *) = 5
group by brewerid
order by cnt desc;
```

MiRACLE
Miracle Finland Oy

# The deep learning version

# What is deep learning?




https://towardsdatascience.com/why-deep-learning-is-needed-over-traditional-machine-learning-1b6a99177063

# Predicting Beer Ratings with Deep Learning

* A quick demo in ATP

# Conclusion

* Oracle PL/SQL Packages for Data Mining
  * DBMS_PREDICTIVE_ANALYTICS
    * Routines for performing *predictive analytics*
  * DBMS_DATA_MINING_TRANSFORMING
    * Routines for *transforming the data* for mining models
  * DBMS_DATA_MINING
    * Routines for *creating and managing mining models*

# Conclusion

The Process

1. Defining the Task, understanding the Task
2. Collecting the data, understanding the data
3. Attributes (Features, Columns)
4. Preparing the data/Transforming the data
5. Creating a Model
    1. Model (Function)
    2. Algorithm
6. Evaluate the models
7. Scoring and Deployment

# THANK YOU!

QUESTIONS?

Email: heli@miracleoy.fi
Twitter: @HeliFromFinland
Blog: Helifromfinland.com

Miracle Finland Oy