Modern Machine Learning
With Oracle Analytics Cloud and Oracle Autonomous Data Warehouse Cloud

Tim and Dan Vlamis

- March 14, 2019

Vlamis Software Solutions

# Happy Pi Day March 14, 2019!

vlamis
SOFTWARE SOLUTIONS

Vlamis Software Solutions

- Vlamis Software founded in 1992 in Kansas City, Missouri
- Developed 200+ Oracle BI and analytics systems
- Specializes in Oracle-based:
  - Enterprise Business Intelligence & Analytics
  - Analytic Warehousing
  - Data Mining and Predictive Analytics
  - Data Visualization
- Multiple Oracle ACEs, consultants average 15+ years
- www.vlamis.com (blog, papers, newsletters, services)
- Co-authors of book "Data Visualization for OBI 11g"
- Co-author of book "Oracle Essbase & Oracle OLAP"
- Oracle University Partner
- Oracle Gold Partner
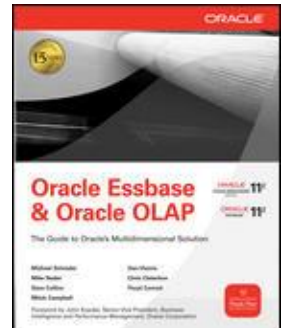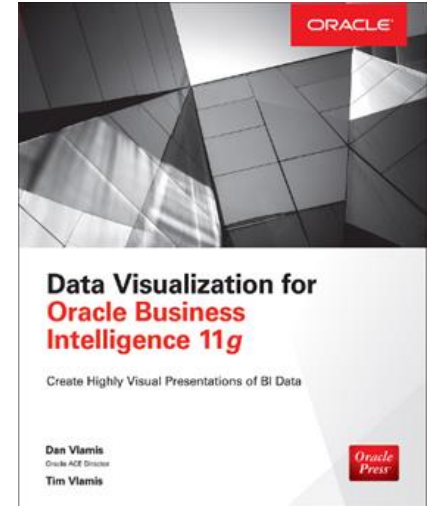
Presenter Background

Dan Vlamis – President
- Founded Vlamis Software Solutions in 1992
- 30+ years in business intelligence, dimensional modeling
- Oracle ACE Director
- Developer for IRI (expert in Oracle OLAP and related)
- BIWA Board Member since 2008
- BA Computer Science Brown University
- @dvlamis

Tim Vlamis – Vice President & Analytics Strategist
- 30+ years in business modeling and valuation, forecasting, and scenario analyses
- Oracle ACE
- Instructor for Oracle University's Predictive Analytics, Data Mining, and Oracle R Enterprise Courses
- Professional Certified Marketer (PCM) from AMA
- MBA Kellogg School of Management (Northwestern University)
- BA Economics Yale University
- @TimVlamis

Vlamis Presentations at AnD Summit

Presenter	Location	Time	Title
Derek Hayden Tim Vlamis	Room 203	Tuesday 11:15am-12:05pm	Billboards to Dashboards: How OUTFRONT Media is Using OAC to Analyze Marketing
Mike Caskey Dan Vlamis	Room 104	Tuesday 2:30-3:20pm	Architecting for Analytics
Jonathan Clark	Room 203	Tuesday 3:35-4:05pm	Automating Pay-As-You-Go Oracle Analytic and Other Cloud Instances
Cathye Pendley Derek Hayden	Room 105	Wednesday 1:00-1:50pm	Building Modern Analytic Map Views in Oracle Analytics Cloud
Tim Vlamis Dan Vlamis	Room 202	Thursday 2:30-3:20pm	Modern Machine Learning with OAC and ADWC

What is Machine Learning?

- The application of advanced analytic algorithms which automatically update their predictions over time.

vlamis™
SOFTWARE SOLUTIONS

Many Words Used for Similar Concepts

Predictive Analytics

Regression Data Mining

SQL Analytics Anomaly Detection

Python Adaptive Intelligence Data Science

Diagnostic Analytics Advanced Analytics

Classification AI

SQL R Algorithm Descriptive Analytics

Clustering Artificial Intelligence

Prescriptive Analytics

Machine Learning

vlamis
SOFTWARE SOLUTIONS

Implications of Systems That "Learn"

CRISP – DM Phases
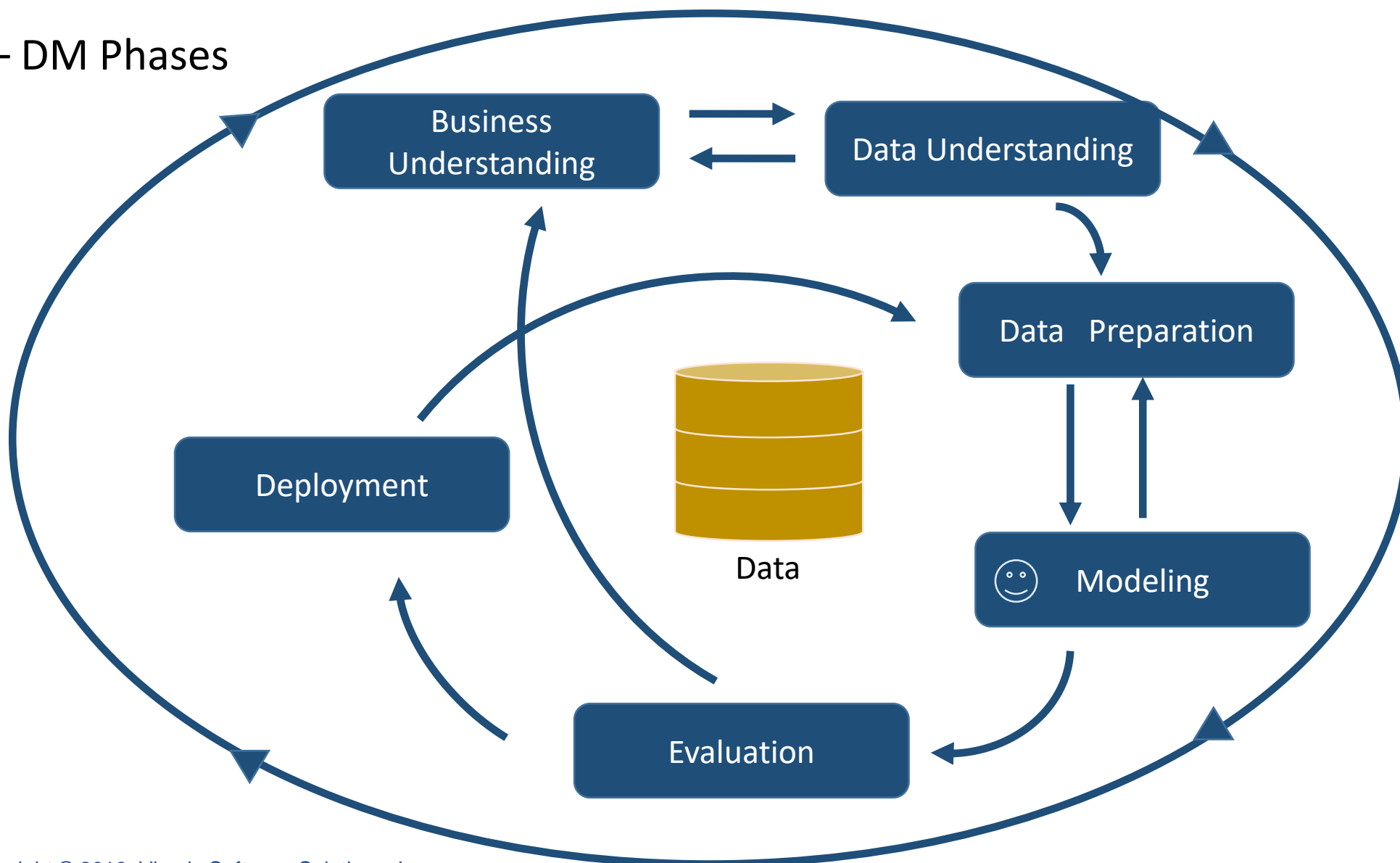
Implications of Systems that "Learn"

- Machine Learning implies a far higher level of involvement and connectedness.
- Being in production implies being at the center of the system.
- Deployment becomes even more important.

Four Realms of Analytics

	Past	Future
Probability Based	**Diagnostic Analytics**	**Predictive Analytics**
Rules Based	**Descriptive Analytics**	**Prescriptive Analytics**

vlamis
SOFTWARE SOLUTIONS

Many Options for ML in OAC

- Work in OAC Data Visualization Project
  - One-click options
  - Use built-in scripts via My Calculations
  - Automated data enrichment in Prepare tab
  - Use Explain for data profiling and unsupervised learning
  - Upload and call custom scripts
- Train and apply models in Data Flows
- Work in "edit formula" in Criteria Tab in Answers Analysis
- Apply custom scripts in Repository
- Connect to Oracle Advanced Analytics in Oracle Database Cloud Service (High Performance or Extreme Performance)
  - Oracle Data Mining
  - Oracle R Enterprise
- Connect to Oracle Machine Learning in Autonomous Data Warehouse Cloud Service

vlamis™
SOFTWARE SOLUTIONS

Many Options for ML in OAC

- 4 Different interfaces with OAC
- 4 Different places with OAC Data Visualization Projects
- Custom script in R or Python
- Integrate with Oracle Advanced Analytics, Oracle Spatial and Graph in Oracle Database Cloud Service
- Integrate with Oracle Machine Learning and Oracle Spatial and Graph in Autonomous Data Warehouse Cloud Service
- Integrate with ORAAH and Oracle Spatial and Graph in Big Data Cloud Service

vlamis
SOFTWARE SOLUTIONS

Many Options for ML in OAC

- 4 Different interfaces with OAC
  - Data Visualization Project
  - Data Flows train and apply models
  - Machine Learning Inspect Models
  - Classic interface Answers & Dashboards and .rpd evaluate function
- 4 Different Places in DV Projects
  - Visualization on Canvas right click
  - "My Calculations" custom script
  - "Explain"
  - Prepare machine learning enrichments

vlamis
SOFTWARE SOLUTIONS

Tradeoffs Abound

- Explanatory power and transparency vs. Accuracy
- Automated data prep vs. conscious data shaping choices
- Clear visualizations vs. multi-dimensional transforms and relationships
- Ease of use vs. computational understanding

vlamis™
SOFTWARE SOLUTIONS

Demo of Single-click ML with use cases and limitations

- Clustering
- Outliers
- Trend
- Forecast

Single-click OAC ML Don'ts

- Do not overinterpret
- Do not be afraid to use
- Do not spend a lot of time when you have more powerful tools

vlamis ™
SOFTWARE SOLUTIONS

Explain feature has power

- Demo of "Explain" with use cases and limitation

OAC Data Flows

- Requires some knowledge of data mining/machine learning
- Build/train models in data flows
- Use classification for single/multi-class predictions
  - Churn models of customer loyalty
  - Predict buy/not buy specific products
- Use regression models for prediction of continuous values
  - Lifetime customer value prediction
  - Next year customer purchases
- Use clustering to segment members into groups
  - Customer segmentation based on history/buying behavior
- Use sentiment analysis to understand comments

vlamis™
SOFTWARE SOLUTIONS

Building Models in Data Flows (demo)

- Training and testing supervised models
- Classification methods and algorithms in OAC
- Interpretation of results
    - Confusion matrix
    - Accuracy, precision, and recall
    - ROC curve
- Visualizing classification results
- Applying classification models
- Use cases

vlamis
SOFTWARE SOLUTIONS

Other ML Models in OAC Data Flows

- Sentiment analysis
- Regression
- Clustering

vlamis ™
SOFTWARE SOLUTIONS

Oracle ML in Autonomous Data Warehouse

- In-database machine learning
  - Don't move the data
  - Extreme power and scalability
  - Extreme flexibility and extendibility
- Zeppelin Notebooks
  - Great for collaboration
  - Powerful tool in the hands of knowledgeable
  - Built-in visualization capability
  - Script development and management
- Oracle Data Mining Algorithms

vlamis
SOFTWARE SOLUTIONS

Demo of Oracle Machine Learning in ADW

vlamis ™
SOFTWARE SOLUTIONS

Oracle Data Miner in Database Cloud Service

- GUI for building predictive analytics workflows
- Build scripts for oracle database without coding
- Powerful built-in visualizations for interpretation

vlamis
SOFTWARE SOLUTIONS

Demo of Oracle Data Miner

vlamis
SOFTWARE SOLUTIONS

Questions to ask yourself

- Do we have people currently on staff who want to execute R or Python models inside the BI system?

- Where do you want to shape data sets for machine learning/predictive analytics?

- Do we currently have clean, consistent, accurate data?

- Do we have an executive champion who understands that systems with have to be grown over time?

- Do we want to start with training or a defined use case?

vlamis ™
SOFTWARE SOLUTIONS

Advice

- Start simply and build
- Start (don't wait)
- Make sure data is clean and consistent
- Don't believe you need a lot of data
- Your best data is likely your internal data
- Promote solid interpretation and understanding of models and results
- Be careful of months (days and weeks often work better) with forecast
- Negative values can throw off some models
- Outliers can have very large effects

vlamis ™
SOFTWARE SOLUTIONS

Thank You!!

- Questions??

- Tim Vlamis tvlamis@vlamis.com
- Dan Vlamis dvlamis@vlamis.com

vlamis ™
SOFTWARE SOLUTIONS

My First Notebook

▷ ✕ 📖 🖊 🗑 ⬇   🔍   ⌨ ⚙ default ▾

Graph the data as a bar chart of EDUCATION

FINISHED ▷ ✕ 📖 ⚙

```
%sql
/*  Bar chart of EDUCATION (Keys), Cust_ID Sum) (Values), nothing in Groups   */

SELECT * FROM SH.SUPPLEMENTARY_DEMOGRAPHICS where ROWNUM <1000;
```

⊞ 📊 ◔ 📈 📉 📉   ⬇ ▾   settings ▾

● Grouped ○ Stacked   ● CUST_ID
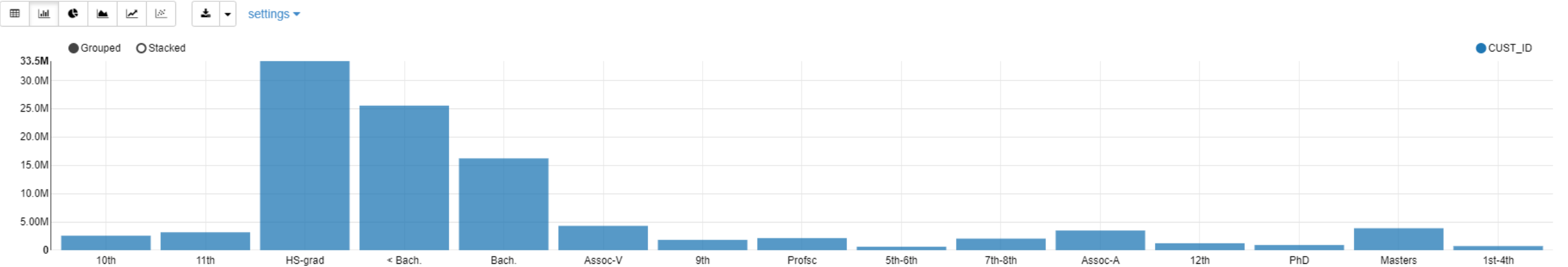
Took 0 sec. Last updated by CHARLIE at October 18 2018, 2:22:22 PM. (outdated)

Graph EDUCATION

FINISHED ▷ ✕ 📖 ⚙

```
%sql
/*  Create pie chart of EDUCATION (Keys), NA (Groups), CUST_ID (SUM) (Values)  */

SELECT * FROM SH.SUPPLEMENTARY_DEMOGRAPHICS where rownum < 1000
```

⊞ 📊 ◔ 📈 📉 📉   ⬇ ▾   settings ▾

● 10th ● 11th ● HS-grad ● < Bach. ● Bach. ● Assoc-V ● 9th ● Profsc ● 5th-6th ● 7th-8th ● Assoc-A ● 12th
● PhD ● Masters ● 1st-4th
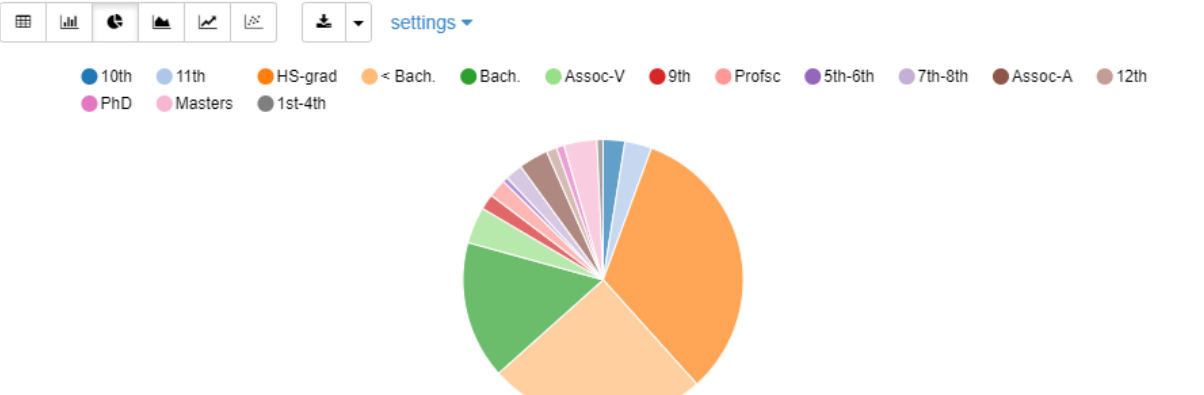
Graph HOUSEHOLD_SIZE

FINISHED ▷ ✕ 📖 ⚙

```
%sql
/*  Create pie chart of HOUSEHOLD_SIZE (Keys), NA (Groups), CUST_ID (SUM) (Values)  */

SELECT * FROM SH.SUPPLEMENTARY_DEMOGRAPHICS where rownum < 1000
```
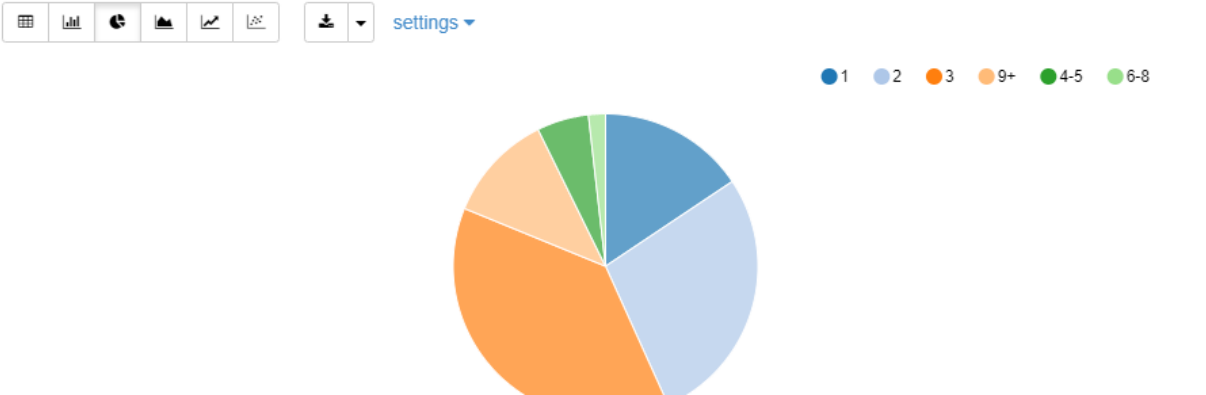
⊞ 📊 ◔ 📈 📉 📉   ⬇ ▾   settings ▾

● 1 ● 2 ● 3 ● 9+ ● 4-5 ● 6-8

Vlamis
SOFTWARE SOLUTIONS

Clustering

▷ ⤬ 📖 🖌 🗑 ⬇   🔍   ⌨ ⚙ default ▾

Took 1 sec. Last updated by USER01 at March 11 2019, 11:52:34 PM.

Took 0 sec. Last updated by USER01 at March 11 2019, 11:52:34 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

Took 0 sec. Last updated by USER01 at March 11 2019, 11:52:35 PM.

Clean up previously existing tables and create Data table

FINISHED ▷ ⤬ 📖 ⚙

```
%script
/*  Clean up and drop any CUSTOMERS360 table if previously exists for notebook reproducibility  */
BEGIN
    EXECUTE IMMEDIATE 'DROP Table CUSTOMERS360';
  EXCEPTION
    WHEN OTHERS THEN NULL;
END;
/

/*  JOIN selected attributes from SH.CUSTOMERS and SH.SUPPLEMENTARY_DEMOGRAPHICS tables to create better 360 view of customer */
Create table CUSTOMERS360 as SELECT a.CUST_ID, a.CUST_GENDER, a.CUST_MARITAL_STATUS, a.CUST_YEAR_OF_BIRTH, a.CUST_INCOME_LEVEL, a.CUST_CREDIT_LIMIT, b.EDUCATION, b.AFFINITY_CARD, b.HOUSEHOLD_SIZE, b.OCCUPATION, b.YRS_RESIDENCE, b.Y_BOX_GAMES
    FROM SH.CUSTOMERS a, SH.SUPPLEMENTARY_DEMOGRAPHICS b
    WHERE a.CUST_ID = b.CUST_ID;

PL/SQL procedure successfully completed.
--------------------------
Table CUSTOMERS360 created.
--------------------------
```

Took 3 sec. Last updated by USER01 at March 12 2019, 12:02:05 AM.

FINISHED ▷ ⤬ 📖 ⚙

```
%script
```

Display CUSTOMERS360 table

PENDING ▐▌ ⤬ 📖 ⚙

```
%sql
SELECT *FROM CUSTOMERS360;
```

outdated

Create a Model Settings table

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

/*  Create a Build Setting (DT) for K-Means Model Build  */

BEGIN
v_sql := 'CREATE TABLE km_sh_sample_settings (setting_name  VARCHAR2(30),setting_value VARCHAR2(4000))';
EXECUTE IMMEDIATE v_sql;
DBMS_OUTPUT.PUT_LINE (v_sql ||': succeeded');
EXCEPTION
WHEN OTHERS THEN
DBMS_OUTPUT.PUT_LINE (v_sql ||': drop unneccessary - no table exists');
END;
/

CREATE TABLE km_sh_sample_settings (setting_name  VARCHAR2(30),setting_value
VARCHAR2(4000)): succeeded
PL/SQL procedure successfully completed.
---------------------------
```

Took 0 sec. Last updated by USER01 at March 11 2019, 11:52:44 PM.

Define the Model Settings

```
%script

/*  Create Model Settings table  */

BEGIN
    INSERT INTO km_sh_sample_settings (setting_name, setting_value) VALUES
    (dbms_data_mining.kmns_distance,dbms_data_mining.kmns_euclidean);

    INSERT INTO km_sh_sample_settings (setting_name, setting_value) VALUES
    (dbms_data_mining.prep_auto,dbms_data_mining.prep_auto_on);
    -- Other examples of overrides are:
    -- (dbms_data_mining.kmns_iterations,3);
    -- (dbms_data_mining.kmns_block_growth,2);
    -- (dbms_data_mining.kmns_conv_tolerance,0.01);
    -- (dbms_data_mining.kmns_split_criterion,dbms_data_mining.kmns_variance);
    -- (dbms_data_mining.kmns_min_pct_attr_support,0.1);
    -- (dbms_data_mining.kmns_num_bins,10);
END;
/

PL/SQL procedure successfully completed.
---------------------------
```
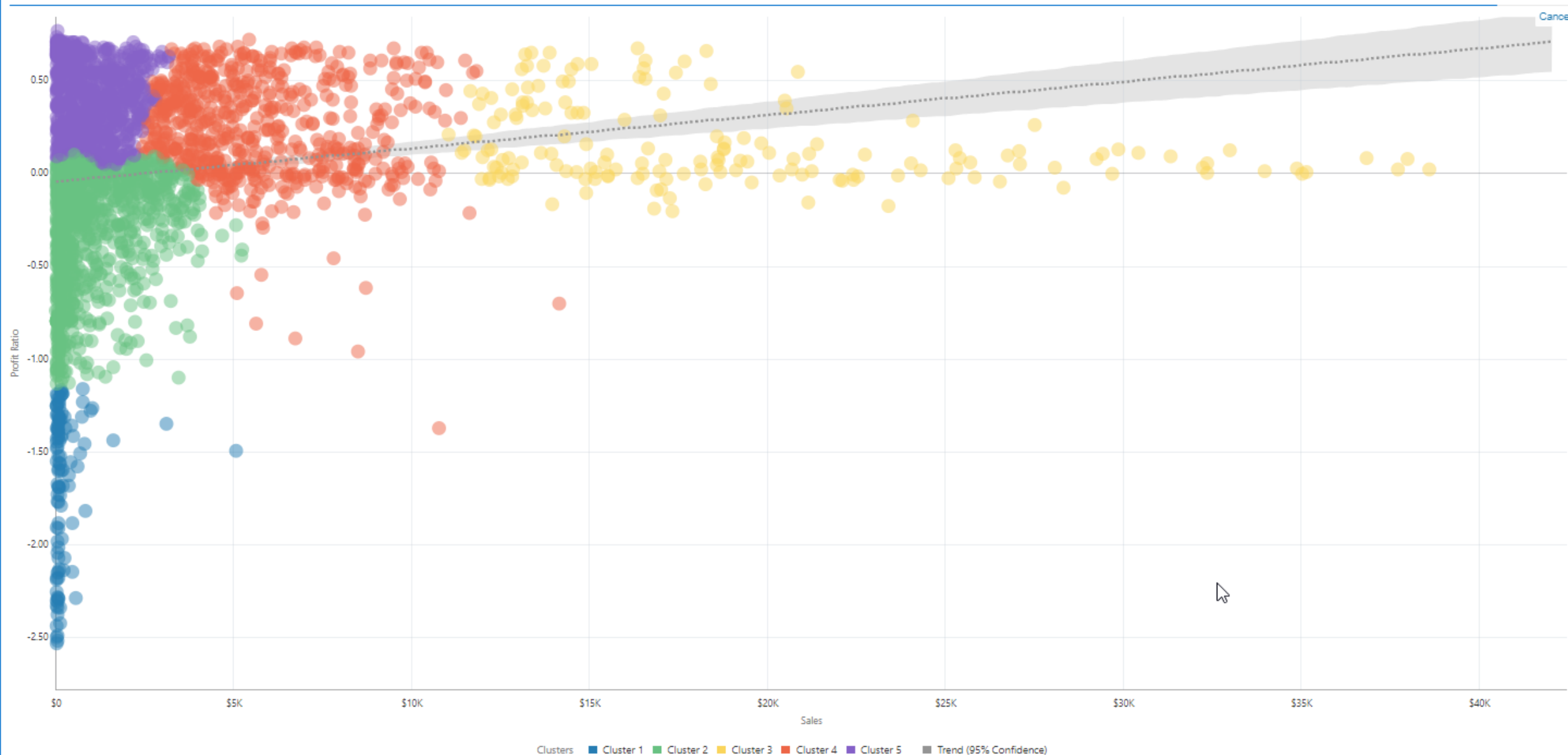
vlamis™
SOFTWARE SOLUTIONS

Sales, Profit Ratio by Customer Name, Clusters

Profit Ratio axis labels: 0.50, 0.00, -0.50, -1.00, -1.50, -2.00, -2.50

Sales axis labels: $0, $5K, $10K, $15K, $20K, $25K, $30K, $35K, $40K

Clusters: ■ Cluster 1 ■ Cluster 2 ■ Cluster 3 ■ Cluster 4 ■ Cluster 5 ■ Trend (95% Confidence)

Edit Calculation

Name   Customer Cluster KM 6 $f_{(x)}$

```
CLUSTER((Customer Name), (Sales,
(cast(Profitas double)/Sales),(Sales/# of
Orders),# of Orders ), 'clusterId',
'algorithm=k-
means;numClusters=%1;maxIter=%2;useRandomSe
ed=FALSE;enablePartitioning=TRUE', 6, 10)
```

Cluster

This function groups a set of records into groups based on one or more input expressions using K-Means or Hierarchical Clustering.

CLUSTER((dimension_expr1, ... dimension_exprN), (expr1, ... exprN), output_column_name, options, [runtime_binded_options])

dimension_expr represents a list of dimensions , e.g. (productID, companyID), to be clustered.
expr represents a list of dimension attributes or measures to be used to cluster the dimension_expr.
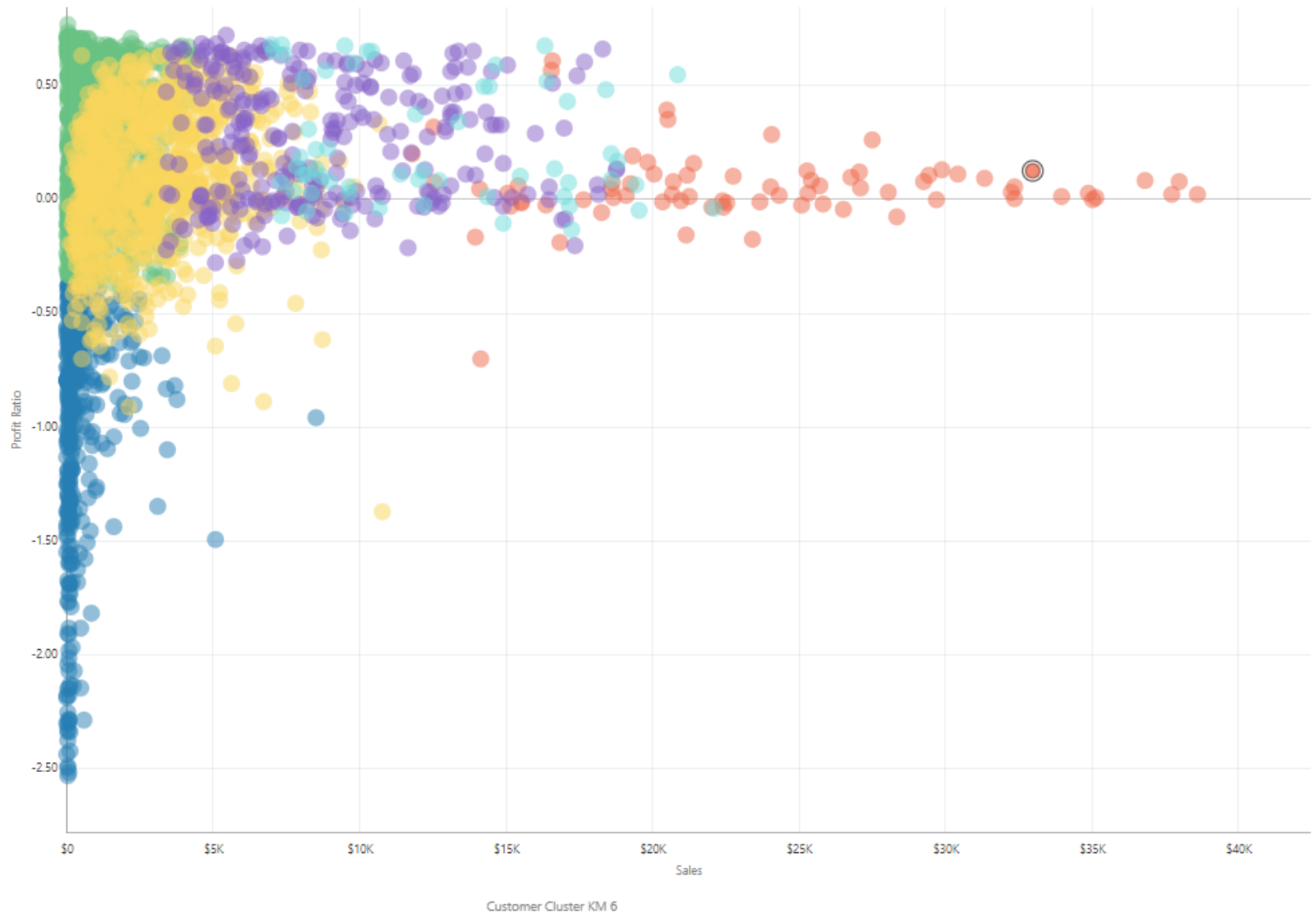output_column_name is the output column. The valid values are "clusterId",
'clusterName' 'clusterDescription'

Validate   Save   Cancel

vlamis ™
SOFTWARE SOLUTIONS

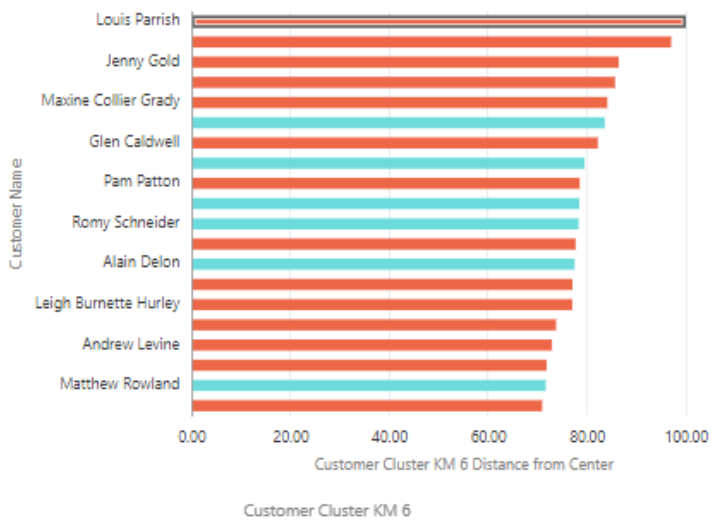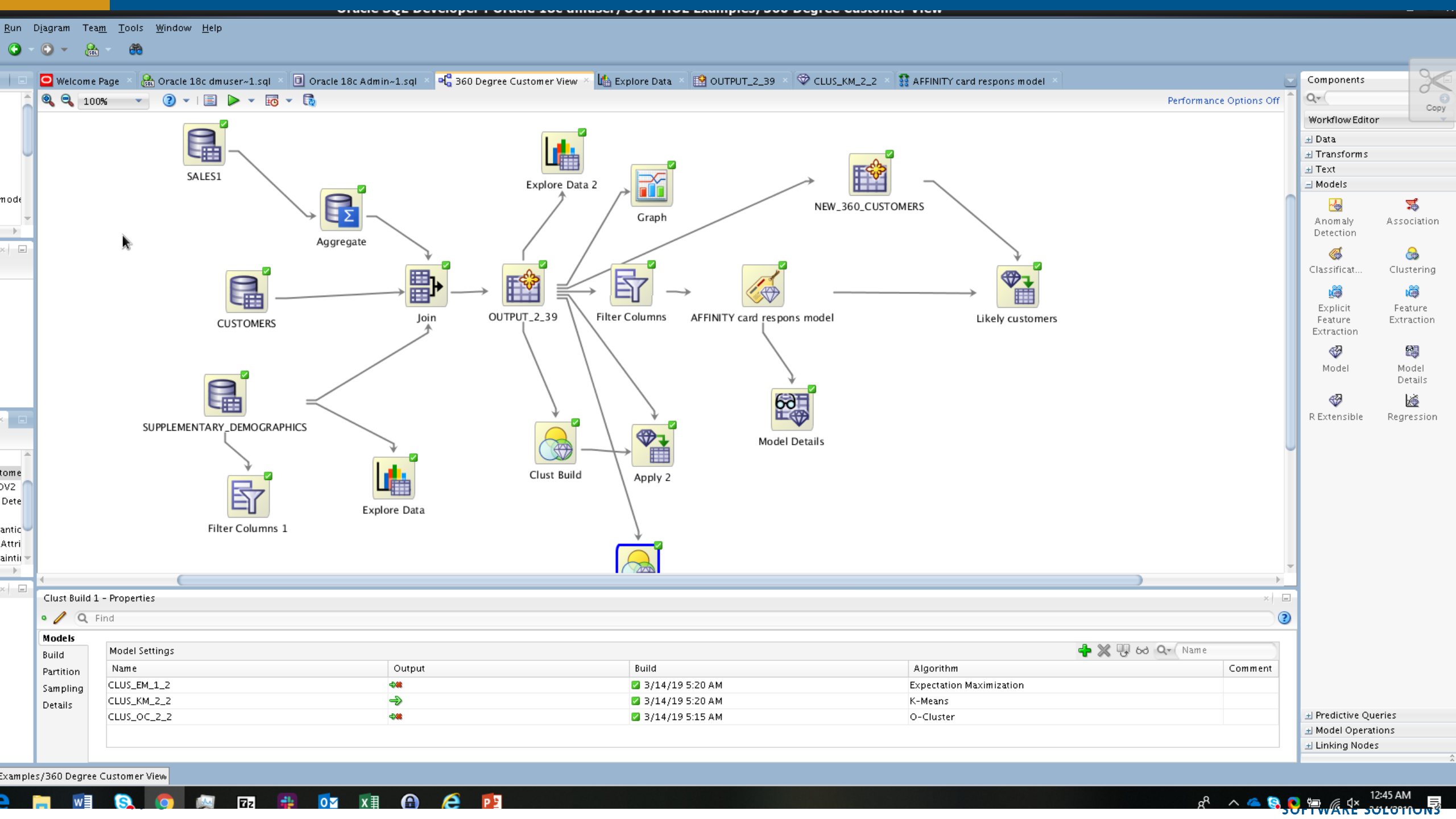**Customer Cluster KM 6 Cluster Size by Customer Cluster KM 6**
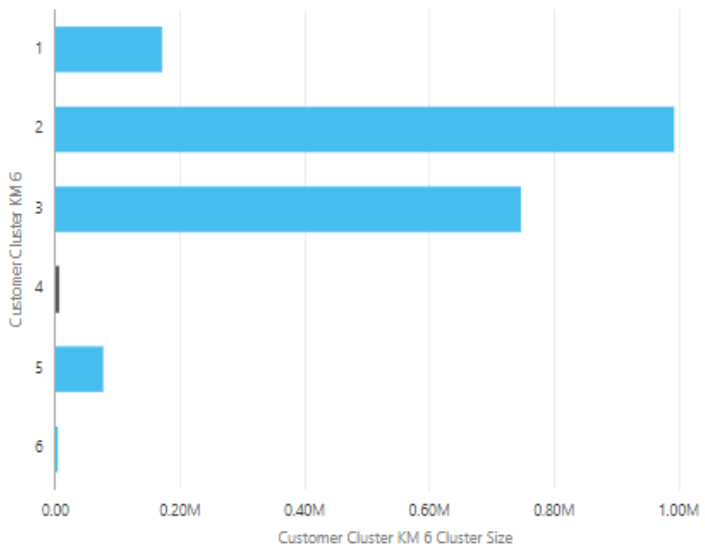
Sales, Profit Ratio by Customer Name, Customer Cluster KM 6

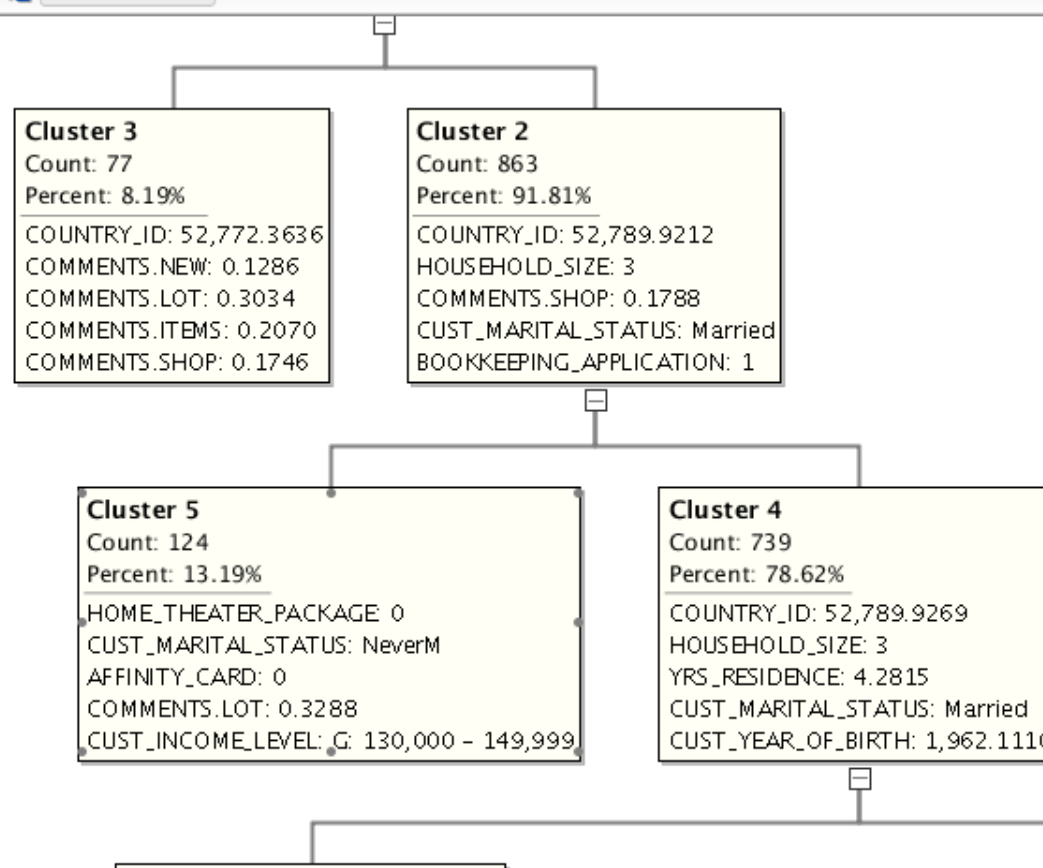**Customer Cluster KM 6 Distance from Center by Customer Name, Customer Cluster...**

Top 20 Customer Cluster KM 6 Distance from Center

Tree | Cluster | Compare | Settings | Alerts

100% | Top Attributes: 5 | Query Settings... | Save Rules...

Cluster 3
Count: 77
Percent: 8.19%

COUNTRY_ID: 52,772.3636
COMMENTS.NEW: 0.1286
COMMENTS.LOT: 0.3034
COMMENTS.ITEMS: 0.2070
COMMENTS.SHOP: 0.1746

Cluster 2
Count: 863
Percent: 91.81%

COUNTRY_ID: 52,789.9212
HOUSEHOLD_SIZE: 3
COMMENTS.SHOP: 0.1788
CUST_MARITAL_STATUS: Married
BOOKKEEPING_APPLICATION: 1

Cluster 5
Count: 124
Percent: 13.19%

HOME_THEATER_PACKAGE 0
CUST_MARITAL_STATUS: NeverM
AFFINITY_CARD: 0
COMMENTS.LOT: 0.3288
CUST_INCOME_LEVEL: G: 130,000 – 149,999

Cluster 4
Count: 739
Percent: 78.62%

COUNTRY_ID: 52,789.9269
HOUSEHOLD_SIZE: 3
YRS_RESIDENCE: 4.2815
CUST_MARITAL_STATUS: Married
CUST_YEAR_OF_BIRTH: 1,962.1110

Centroid | Rule

| Name	Importance	Mode	Mean	Variance	Support
HOME_THEATER_PACKAGE	0.3676	0			124
Y_BOX_GAMES	0.3182	1			115
FLAT_PANEL_MONITOR	0.2703	0			119
BULK_PACK_DISKETTES	0.2244	0			110
CUST_YEAR_OF_BIRTH	0.1217		1,979.06451613	16.80881196	120
YRS_RESIDENCE	0.0904		2.26612903	0.83103855	124
CUST_INCOME_LEVEL	0.0885	G: 130,000 – 149,999			117
COMMENTS.LOT	0.0874		0.32879328	0.00020443	15

Performance | Performance Matrix | ROC | Lift | Profit
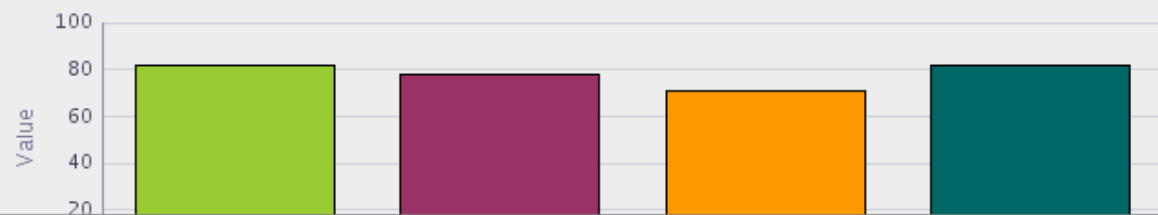
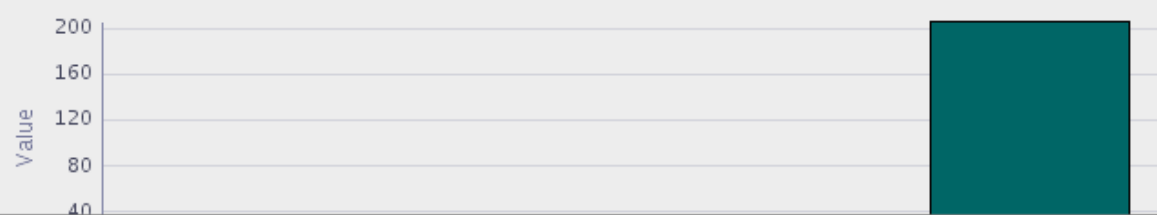Measure: All Measures   Sort By: Name   Descending

Predictive Confidence (%)
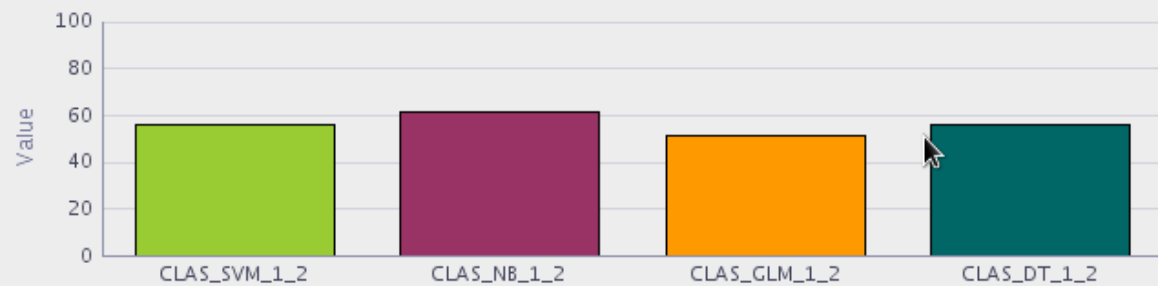
Average Accuracy (%)
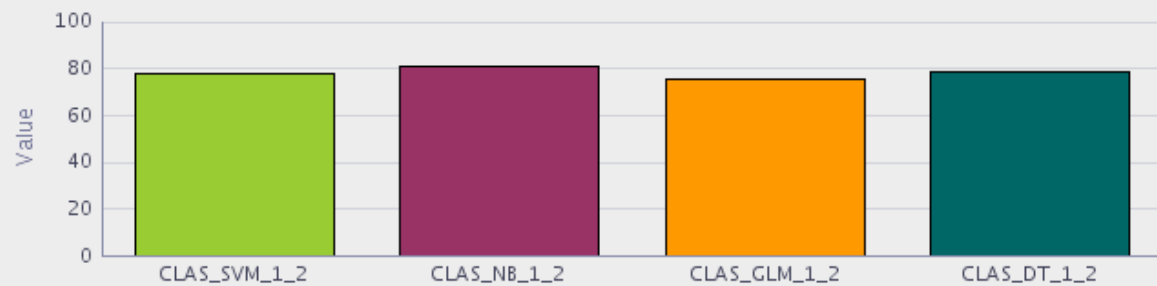
Overall Accuracy (%)

Cost

Models

Name	Predictive Confidence %	Overall Accuracy %	Average Accuracy %	Cost	Algorithm	Creation Date
CLAS_DT_1_2	56.2715	81.7391	78.1357	204.96894410	Decision Tree	10/16/18 10:00 PM
CLAS_GLM_1_2	51.2016	70.4348	75.6008		Generalized Linear Mo...	10/16/18 10:00 PM
CLAS_NB_1_2	61.2716	77.6087	80.6358		Naive Bayes	10/16/18 10:00 PM
CLAS_SVM_1_2	55.9849	81.5217	77.9925		Support Vector Machine	10/16/18 10:00 PM

Thank You!

Questions?

Dan and Tim Vlamis

dvlamis@vlamis.com

tvlamis@vlamis.com

vlamis ™

SOFTWARE SOLUTIONS