ODTUG Kscope19

SEATTLE, WASHINGTON • JUNE 23-27

PLEASE FILL OUT YOUR EVALUATIONS

SEATTLE

Washington State Convention Center

vlamis™
SOFTWARE SOLUTIONS

Modern Machine Learning

With Oracle Analytics Cloud and Oracle Autonomous Data Warehouse Cloud

Tim Vlamis

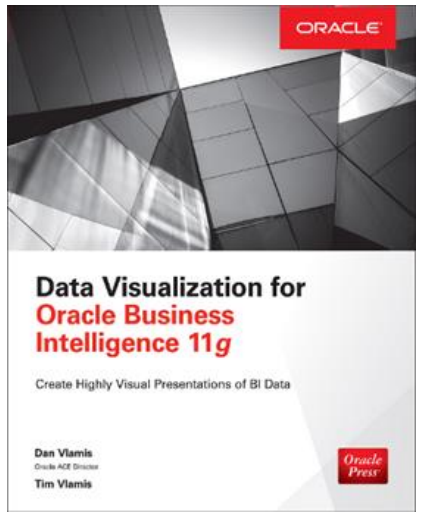
June 26, 2019

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”
- Co-author of book “Oracle Essbase & Oracle OLAP”
- Oracle University Reseller
- Oracle Gold Partner

ORACLE Gold Partner
Specialized
Oracle Business Analytics Cloud Platform



ORACLE ACE Director

ORACLE ACE

ORACLE EDUCATION RESELLER

ORACLE Gold Partner
Specialized
Oracle Business Intelligence Foundation Suite 11g

vlamis™ SOFTWARE SOLUTIONS

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 Techniques and Oracle R Enterprise Essentials Courses
- Professional Certified Marketer (PCM) from AMA
- MBA Kellogg School of Management (Northwestern University)
- BA Economics Yale University

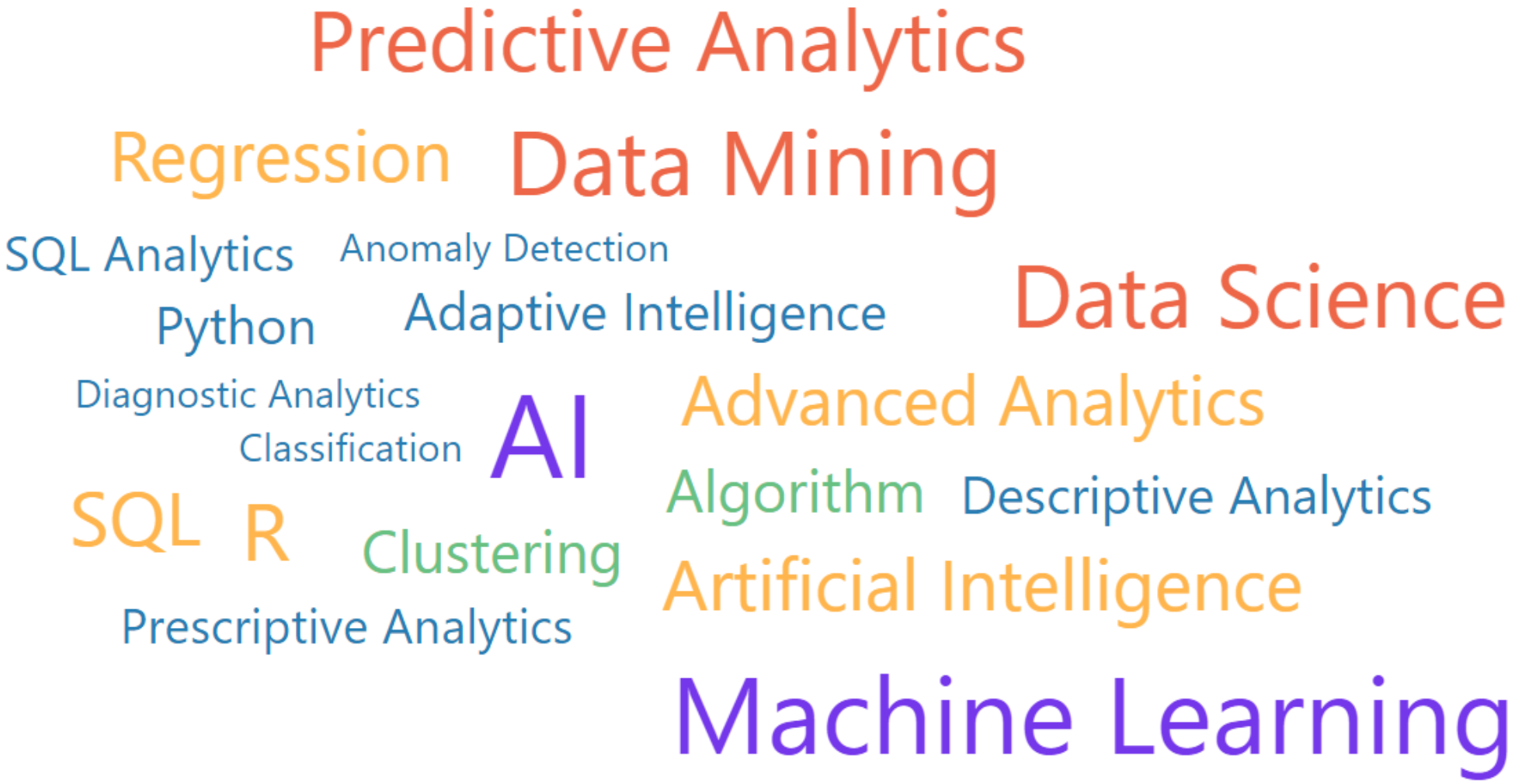
vlamis™ SOFTWARE SOLUTIONS

What is Machine Learning?

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

vlamis™
SOFTWARE SOLUTIONS

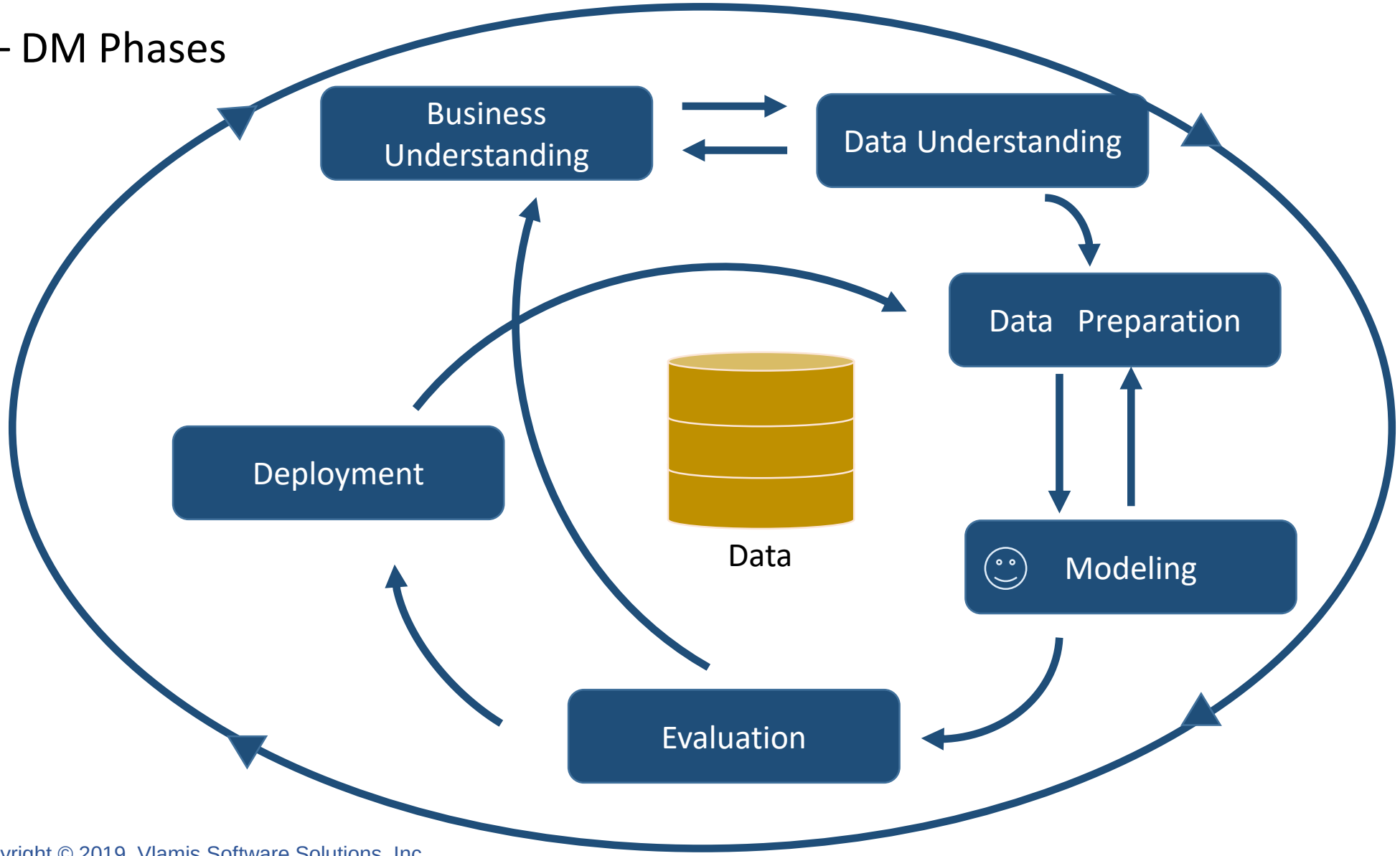
Many Words Used for Similar Concepts



vlamis™
SOFTWARE SOLUTIONS

Implications of Systems That "Learn"

CRISP – DM Phases

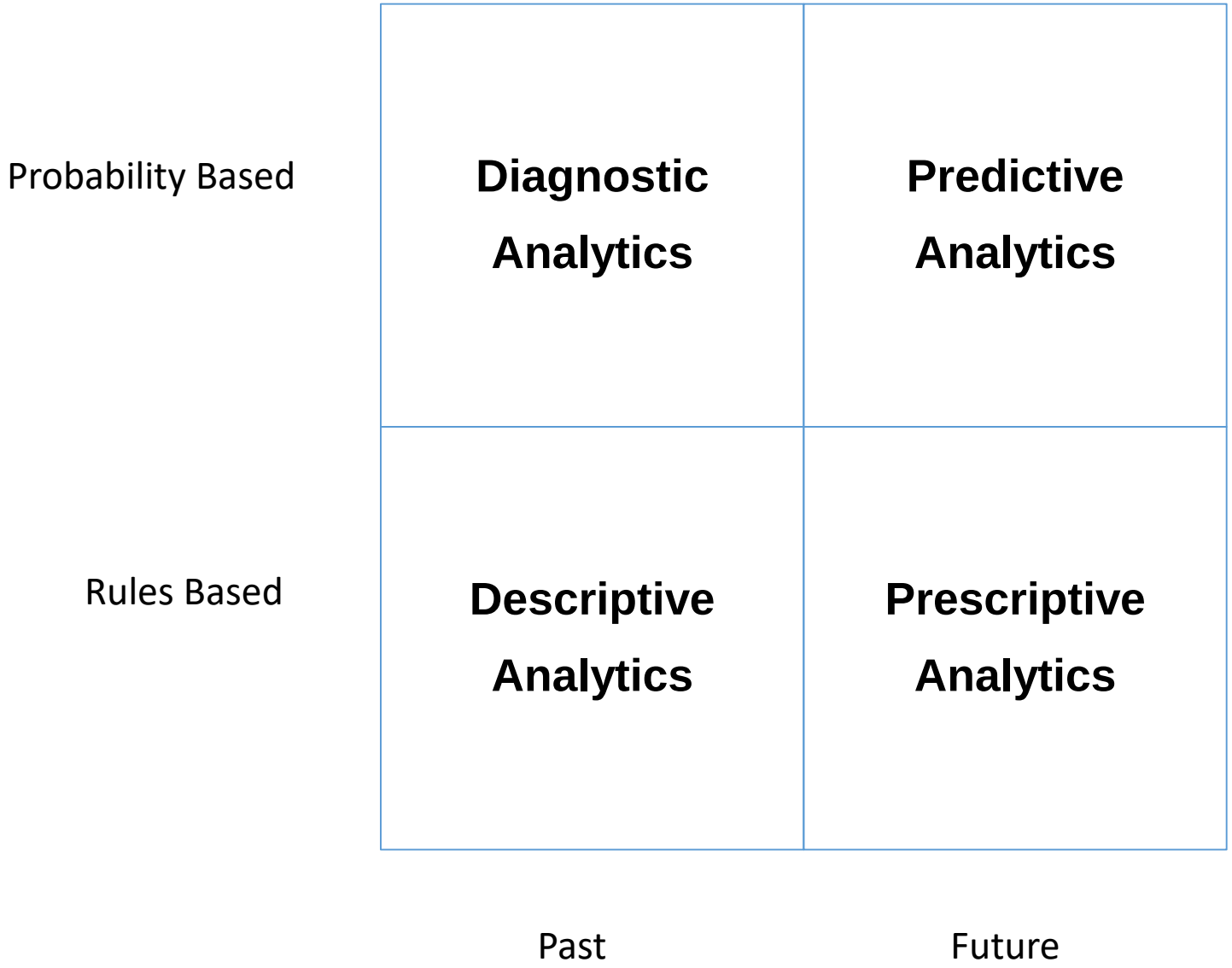


vlamis™
SOFTWARE SOLUTIONS

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



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

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

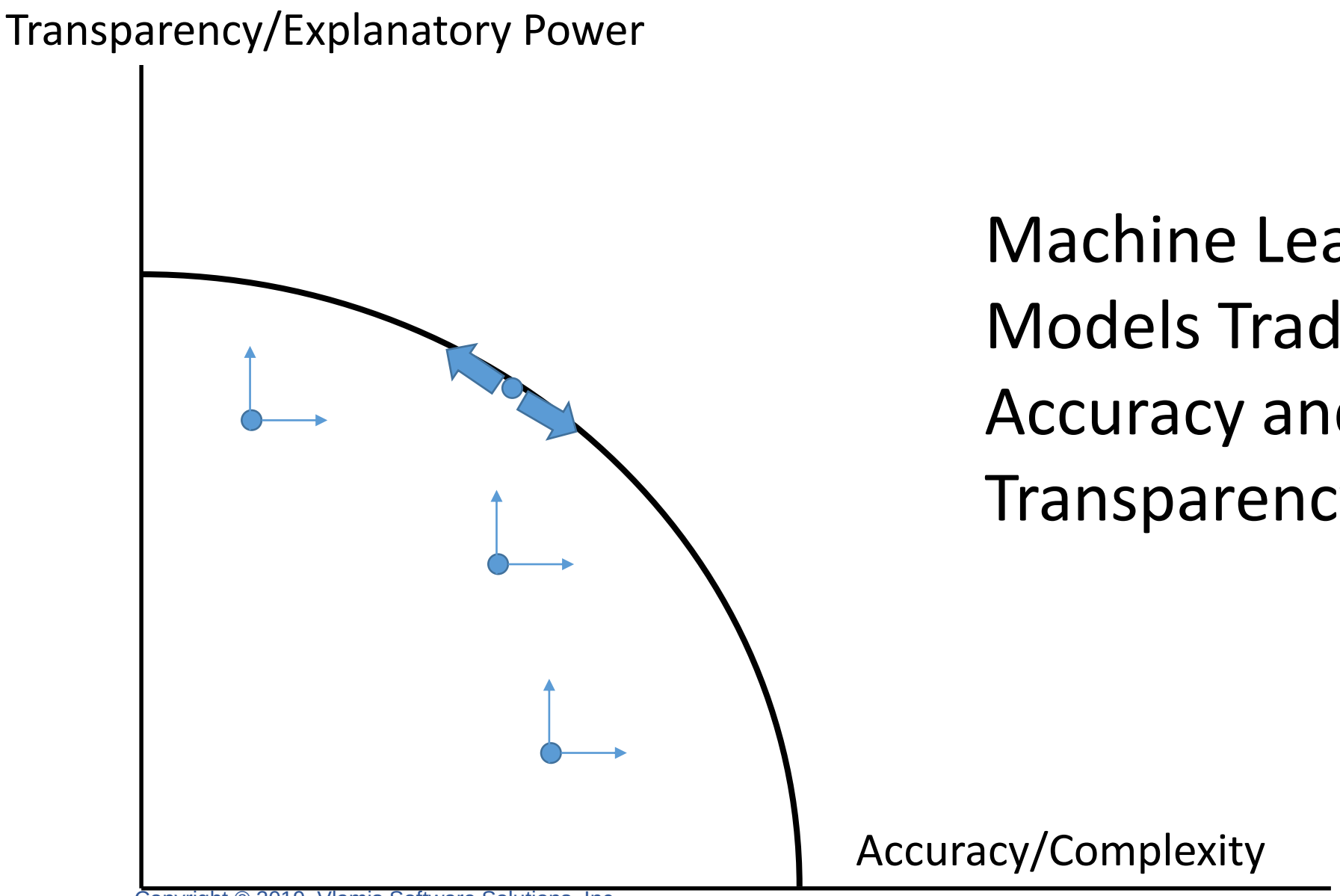
Many Options for ML in OAC

- 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

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

Horizon Function – Tradeoff Visualization



Machine Learning Models Tradeoff Accuracy and Transparency.

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

Explain feature has power

- Demo of “Explain” with use cases and limitation

Explain Inspection Demerits

Add Selected

Basic Facts about Inspection Demerits

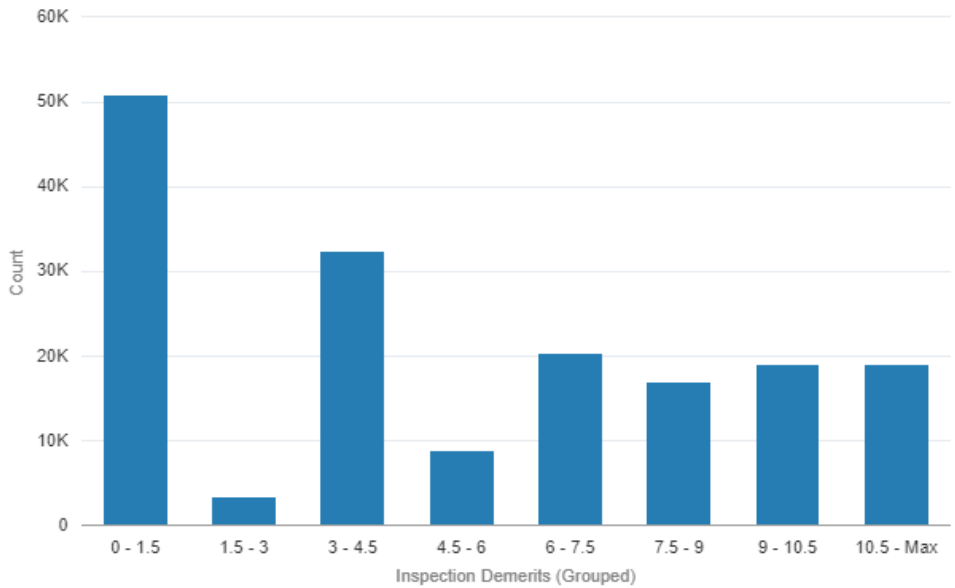
What are the values of Inspection Demerits and how do they relate to each other?

Anomalies of Inspection Demerits

What groups in the data exhibit unexpected results for Inspection Demerits?

About Explain pages

Basic facts about Inspection Demerits



Inspection Demerits is a Numeric Measure, whose sum across 170,463 rows is 979,790.00. The values of Inspection Demerits on each row range from 0.00 to 250.00 and is 5.00 on average.

The charts below summarize the values of Inspection Demerits by the measures in this data set. Click the checkmarks above any of the visuals to add them to your project when done.

vlamis SOFTWARE SOLUTIONS

Explain Inspection Demerits

Add Selected

Basic Facts about Inspection Demerits

What are the values of Inspection Demerits and how do they relate to each other?

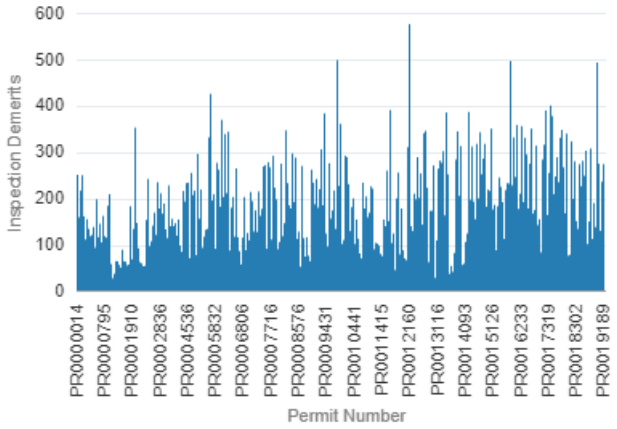
Anomalies of Inspection Demerits

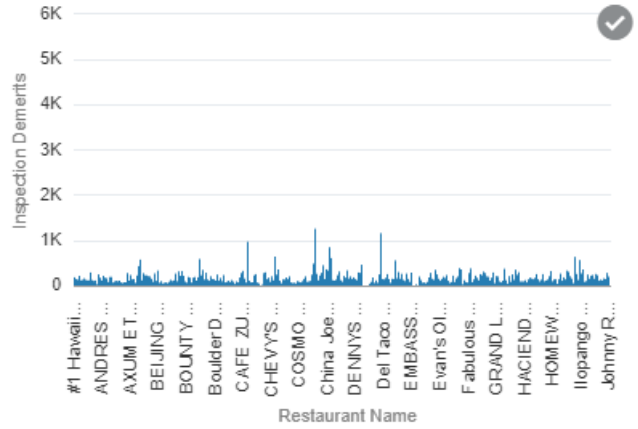
What groups in the data exhibit unexpected results for Inspection Demerits?

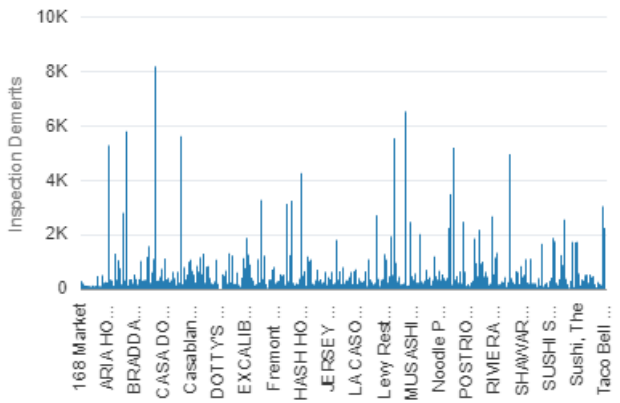
About Explain pages

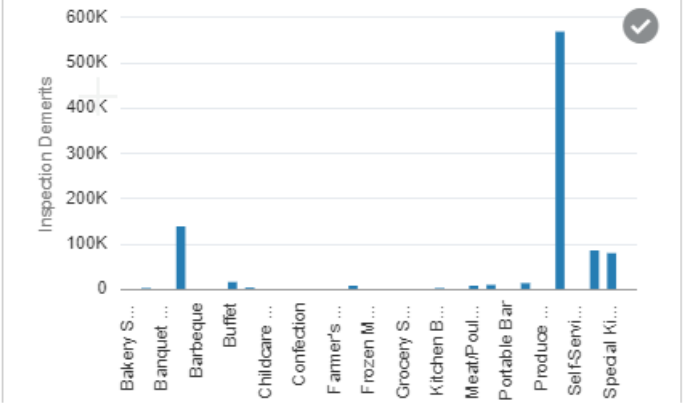
The charts below summarize the values of Inspection Demerits by the measures in this data set. Click the checkmarks above any of the visuals to add them to your project when done.

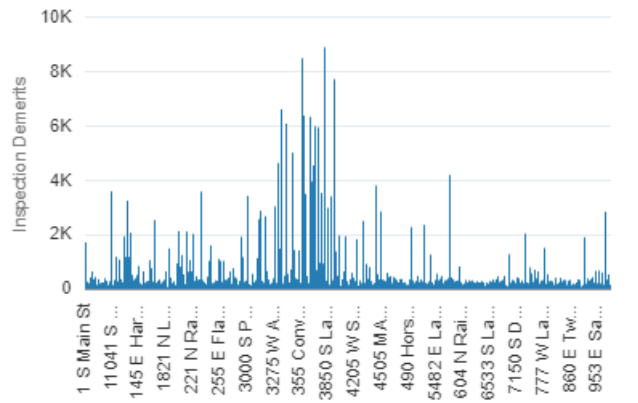
Too Much Data. Please add more filters.
More Detail











vlamis SOFTWARE SOLUTIONS

Explain Inspection Demerits

Add Selected

Basic Facts about Inspection Demerits

What are the values of Inspection Demerits and how do they relate to each other?

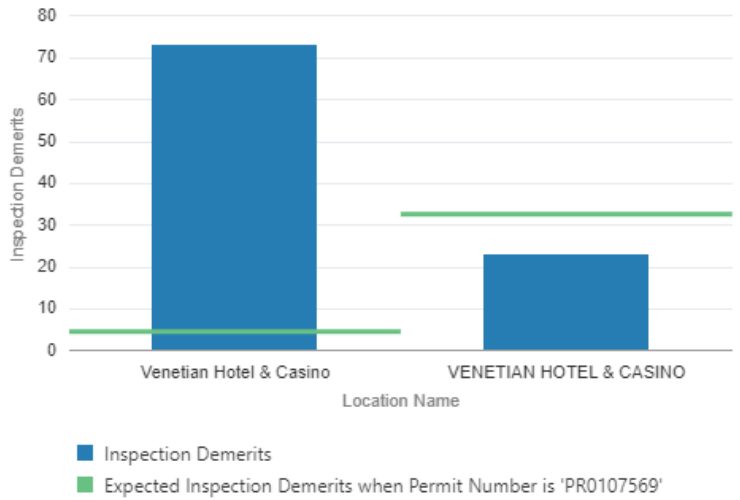
Anomalies of Inspection Demerits

What groups in the data exhibit unexpected results for Inspection Demerits?

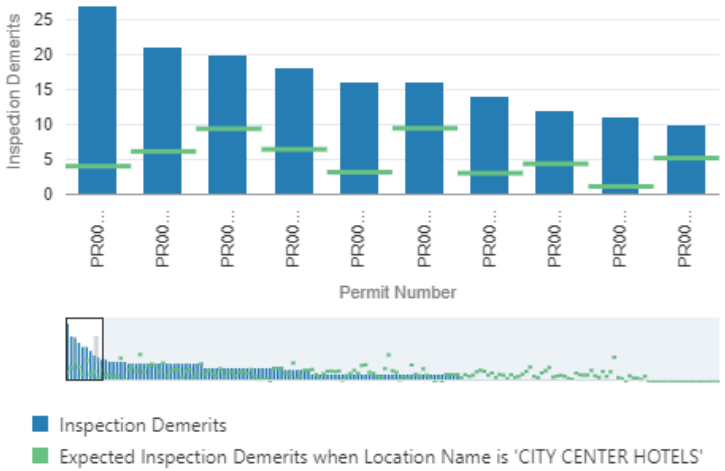
Anomalies of Inspection Demerits

231 combinations of 22 dimensions are being analyzed. Here are the top outliers for Inspection Demerits

When Permit Number is PR0107569, we expected Inspection Demerits for Location Name: Venetian Hotel & Casino to be 4.65, however, it is 73.00, representing a difference of 68.35.



When Location Name is CITY CENTER HOTELS, we expected Inspection Demerits for Permit Number: PR0023967 to be 4.08, however, it is 27.00, representing a difference of 22.92.



About Explain pages

Searching for anomalies

Cancel

4 updates available

Refresh View

vlamis SOFTWARE SOLUTIONS

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

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

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

Demo of Oracle Machine Learning in ADW

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

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?

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

Thank You!!

- Questions??

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

vlamis™
SOFTWARE SOLUTIONS

My First Notebook

learning-sql-notebooks-for-the-oracle-autonomous-data-warehouse-cloud on Oracle Machine Learning blog.

Took 0 sec. Last updated by CHARLIE at October 18 2018, 2:18:52 PM.

Show all tables

FINISHED

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

OWNER	TABLE_NAME	TABLESPACE_NAME	CLUSTER_NAME	IOT_NAME	STATUS	PCT_FREE	PCT_USED	INI_TRANS	MAX_TRANS	INITIAL_EXTENT	NEXT_EXTENT	MIN_EXTENTS	MAX_EXTEI
SH	TIMES	SAMPLESCHEMA			VALID	0		1	255	65536	1048576	1	2147483645
SH	PRODUCTS	SAMPLESCHEMA			VALID	0		1	255	65536	1048576	1	2147483645
SH	PROMOTIONS	SAMPLESCHEMA			VALID	0		1	255	65536	1048576	1	2147483645
SH	COUNTRIES	SAMPLESCHEMA			VALID	0		1	255	65536	1048576	1	2147483645
SH	CHANNELS	SAMPLESCHEMA			VALID	0		1	255	65536	1048576	1	2147483645
SH	COSTS	SAMPLESCHEMA			VALID	0		1	255	65536	1048576	1	2147483645
SH	SALES	SAMPLESCHEMA			VALID	0		1	255	65536	1048576	1	2147483645
SH	SUPPLEMENTARY_DEMOGRAPHICS	SAMPLESCHEMA			VALID	0		1	255	65536	1048576	1	2147483645

Took 1 sec. Last updated by USER01 at March 12 2019, 12:20:52 AM.

Display table

FINISHED

```
%sql
/*  Display SUPPLEMENTARY_DEMOGRAPHICS table  */

SELECT * FROM SH.SUPPLEMENTARY_DEMOGRAPHICS;
```

CUST_ID	EDUCATION	OCCUPATION	HOUSEHOLD_SIZE	YRS_RESIDENCE	AFFINITY_CARD	BULK_PACK_DISKETTES	FLAT_PANEL_MONITOR	HOME_THEATER_PACKAGE	BOOKKEEPING_APPLICATION	PRINTER_SUPPL
101860	HS-grad	Other	1	1	0	1	1	0	1	1

vlamis™ SOFTWARE SOLUTIONS

My First Notebook

default

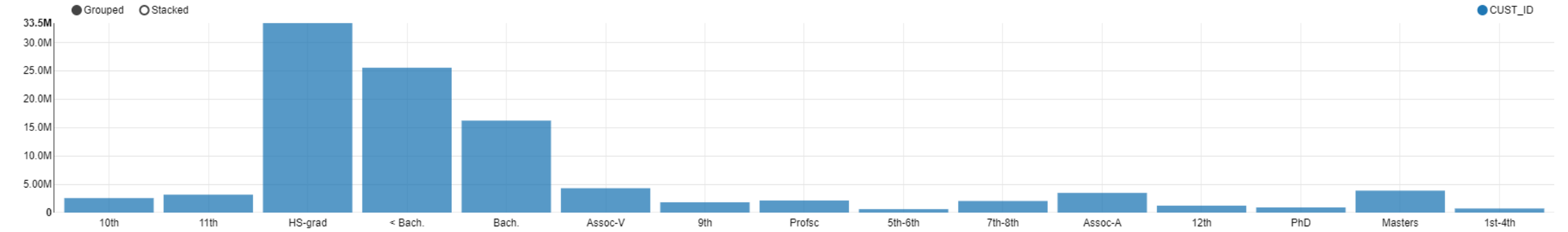
Graph the data as a bar chart of EDUCATION

FINISHED

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

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

settings



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

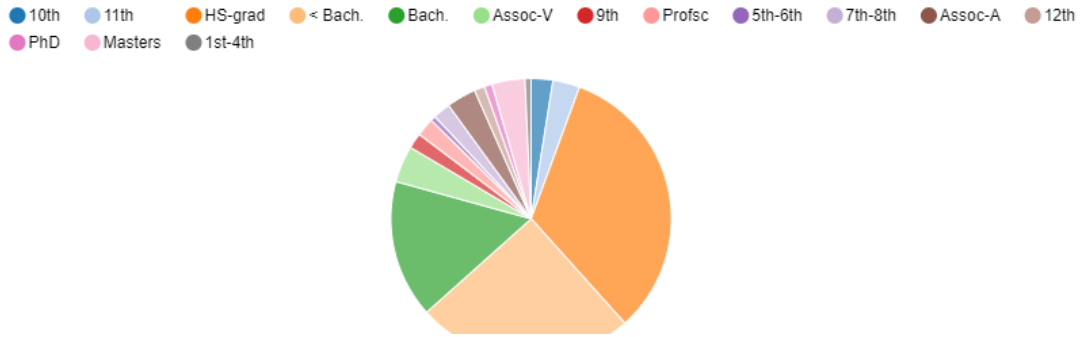
Graph EDUCATION

FINISHED

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

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

settings



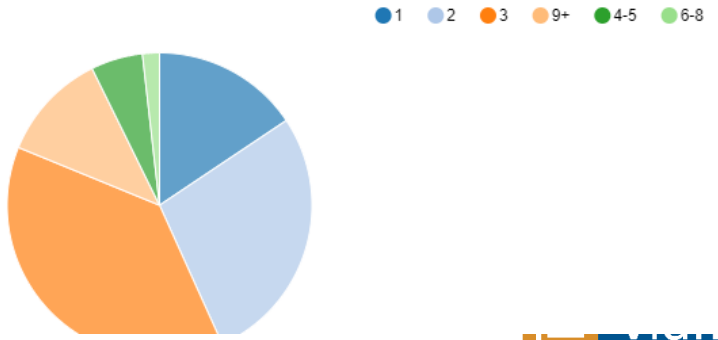
Graph HOUSEHOLD_SIZE

FINISHED

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

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

settings



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.

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

FINISHED

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.

```
%script
```

FINISHED

Display CUSTOMERS360 table

PENDING

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

outdated

Clustering

default

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.
---------------------------
```

rag data to add a filter

Edit Calculation

Name: Customer Cluster KM 6

```
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)
```

Validate | Save | Cancel

Cluster

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

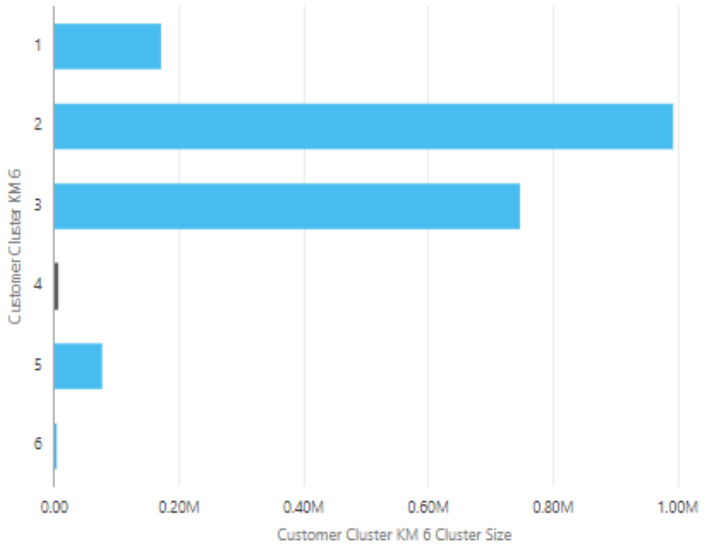
CLUSTER({dimension_expr1 , ... dimension_exprN}, {expr1, ... exprN}, output_column_name, options, [runtime_binded_options])

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'

vlamis™ SOFTWARE SOLUTIONS

to add a filter

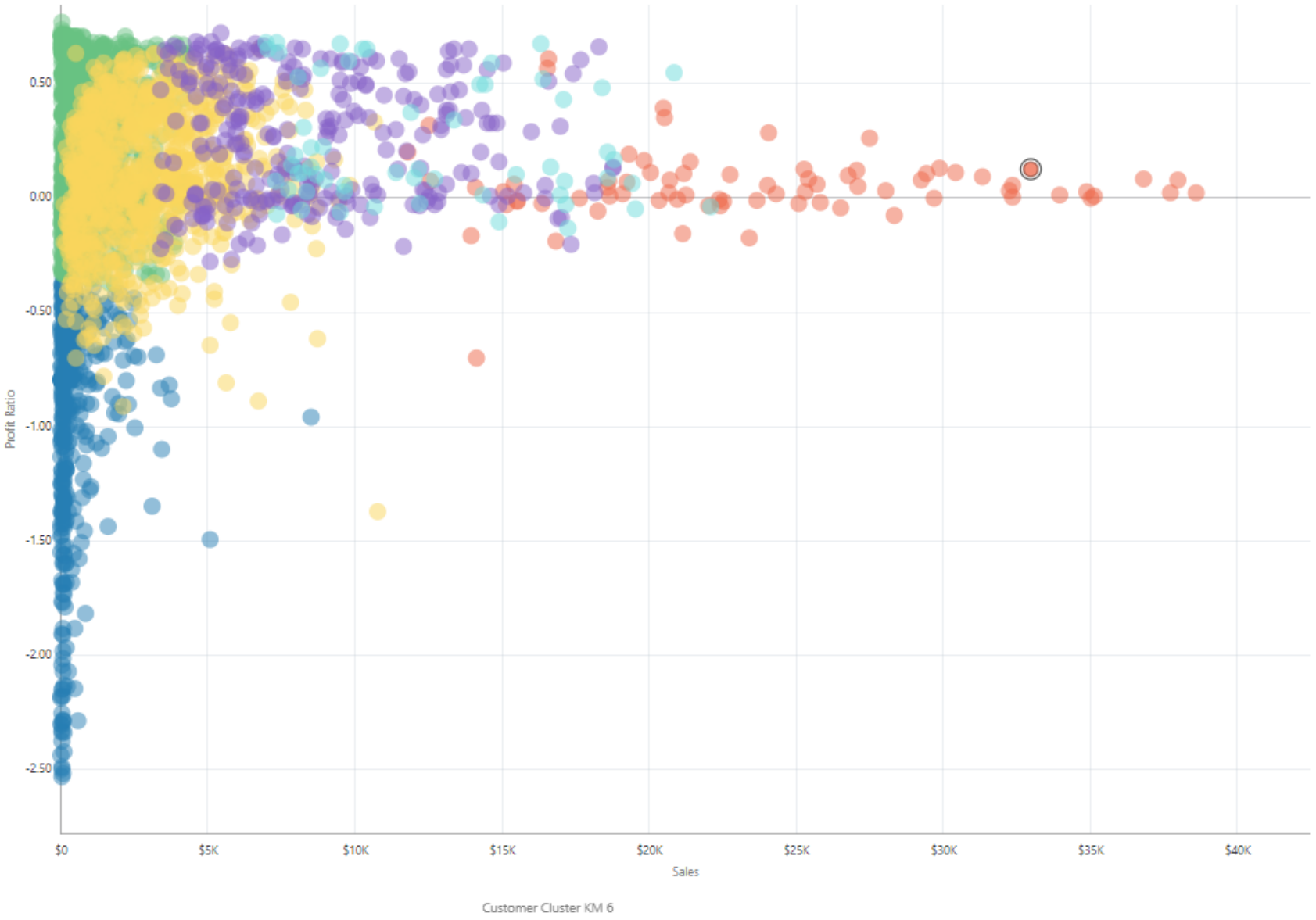
Customer Cluster KM 6 Cluster Size by Customer Cluster KM 6



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

Top 20 Customer Cluster KM 6 Distance from Center

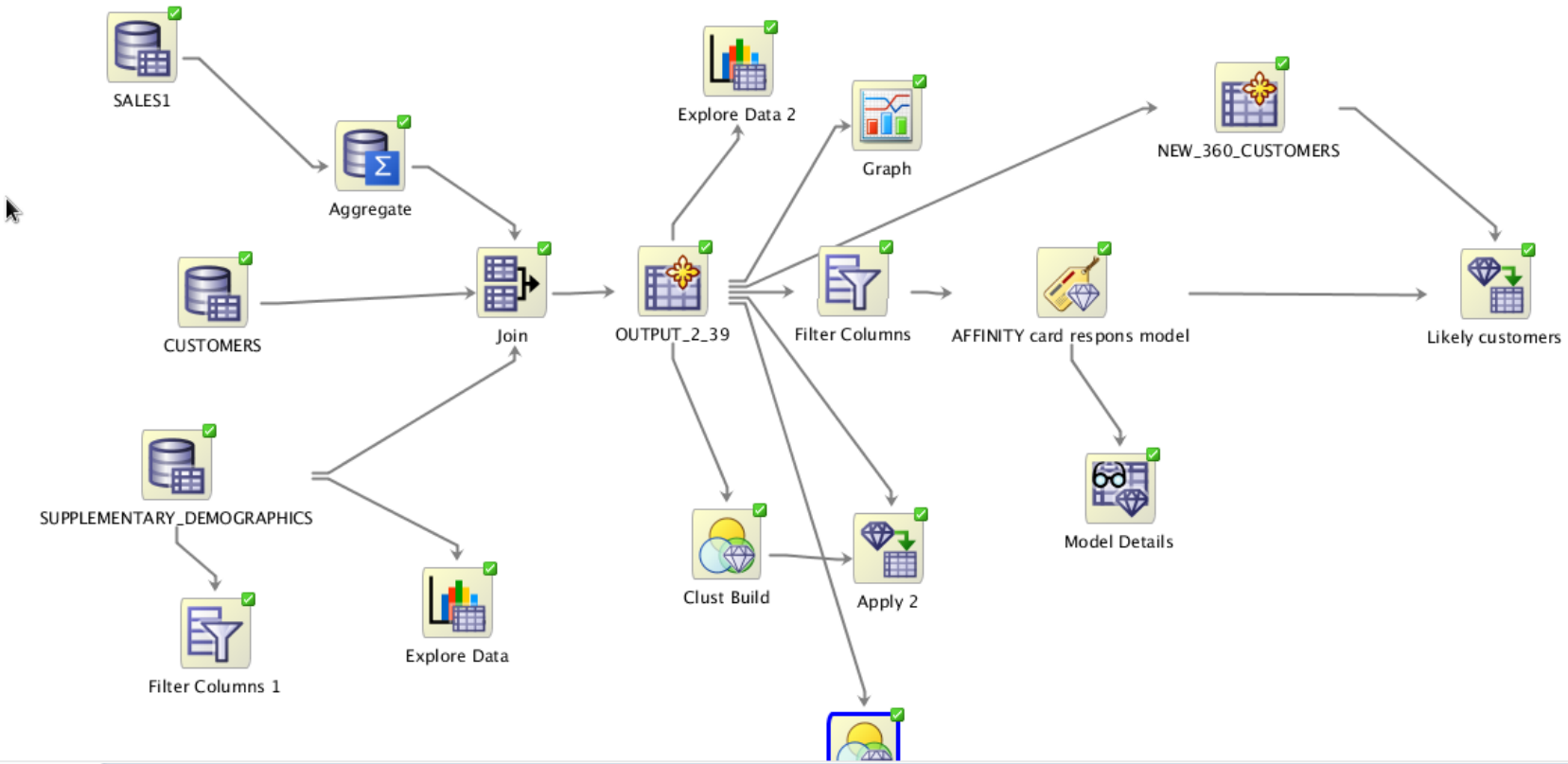
Sales, Profit Ratio by Customer Name, Customer Cluster KM 6



Run Diagram Team Tools Window Help

Welcome Page | Oracle 18c dmuser~1.sql | Oracle 18c Admin~1.sql | 360 Degree Customer View | Explore Data | OUTPUT_2_39 | CLUS_KM_2_2 | AFFINITY card respons model

100% Performance Options Off



Components

Workflow Editor

Data
Transforms
Text
Models

Anomaly Detection | Association | Classificat... | Clustering | Explicit Feature Extraction | Feature Extraction | Model | Model Details | R Extensible | Regression

Predictive Queries
Model Operations
Linking Nodes

Clust Build 1 - Properties

Find

Models

Build
Partition
Sampling
Details

Model Settings

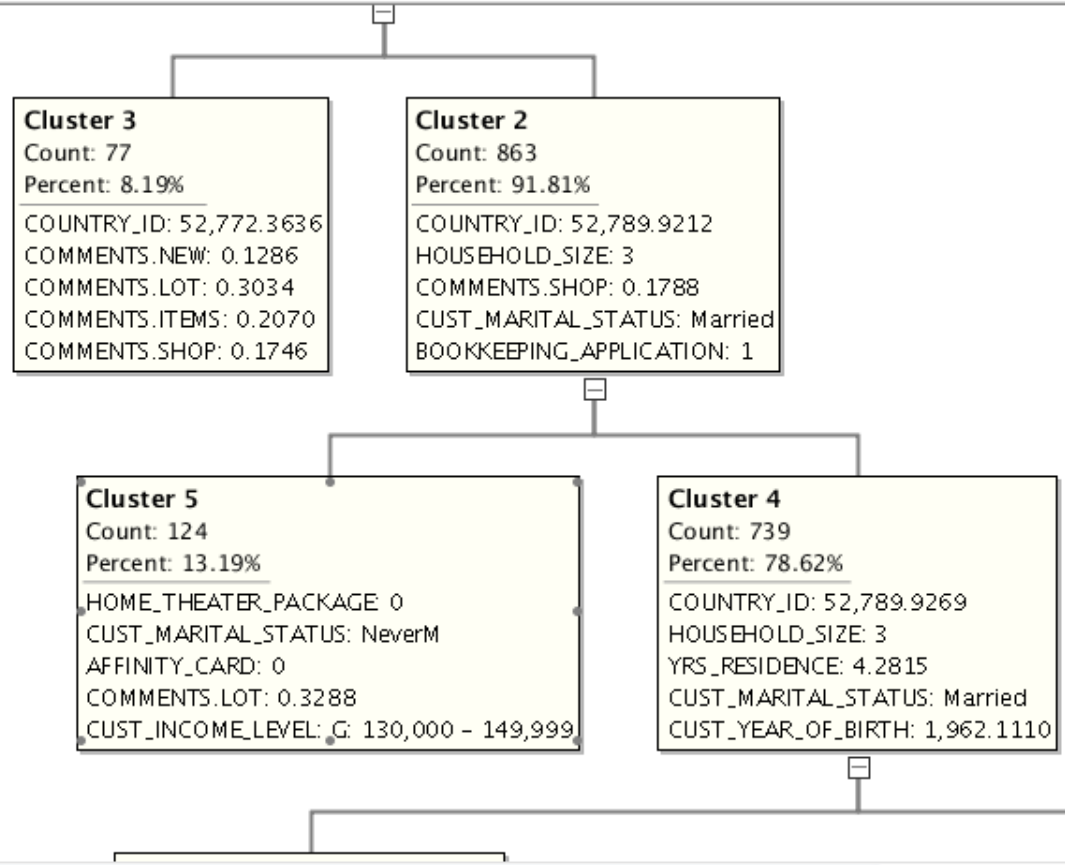
Name	Output	Build	Algorithm	Comment
CLUS_EM_1_2		3/14/19 5:20 AM	Expectation Maximization	
CLUS_KM_2_2		3/14/19 5:20 AM	K-Means	
CLUS_OC_2_2		3/14/19 5:15 AM	O-Cluster	

Examples/360 Degree Customer View

12:45 AM

Tree | Cluster | Compare | Settings | Alerts

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



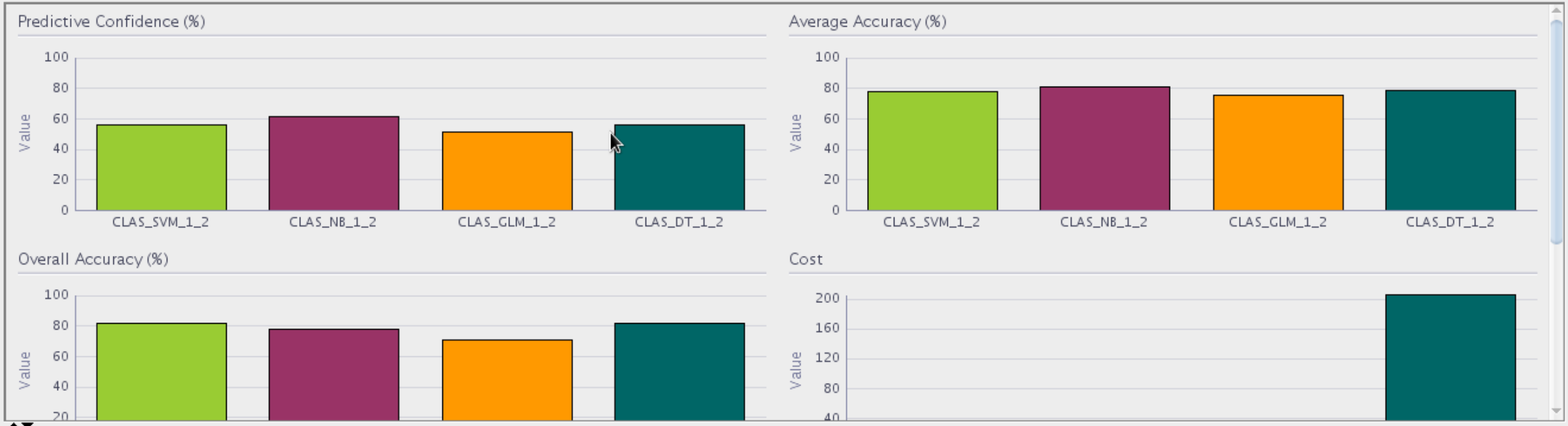
Centroid | Rule

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

...ge | Oracle 18c dmuser~1.sql | Oracle 18c Admin~1.sql | 360 Degree Customer View | CLUS_KM_1_2 | Explore Data | OUTPUT_2_39 | CLUS_KM_2_2 | AFFINITY card respons model

Performance | Performance Matrix | ROC | Lift | Profit

Measure: All Measures Sort By: Name Descending



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?

Tim Vlamis

tvlamis@vlamis.com

vlamis™
SOFTWARE SOLUTIONS

ODTUG Kscope19

SEATTLE, WASHINGTON • JUNE 23-27

PLEASE FILL OUT YOUR EVALUATIONS

SEATTLE

Washington State Convention Center