



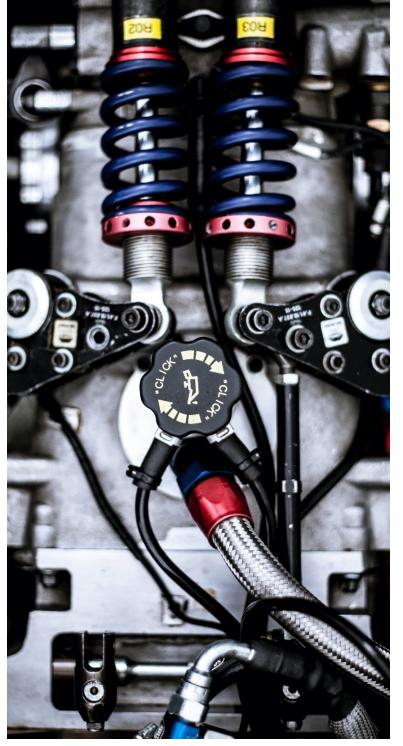




www.rittmanmead.com



A DATA AND ANALYTICS COMPANY



Data Engineering



Analytics



Data Science

Agenda



- Data Scientist
- Become a Data Scientist

Oracle Analytics Cloud

- Platform Services (PaaS)
- Delivered entirely in the cloud:
 - No infrastructure footprint
 - Flexibility
 - Simplified, metered licensing
- Several options to suit your needs:
 - •BYOL
 - Functionality bundled into 2 editions
 - Professional
 - Enterprise



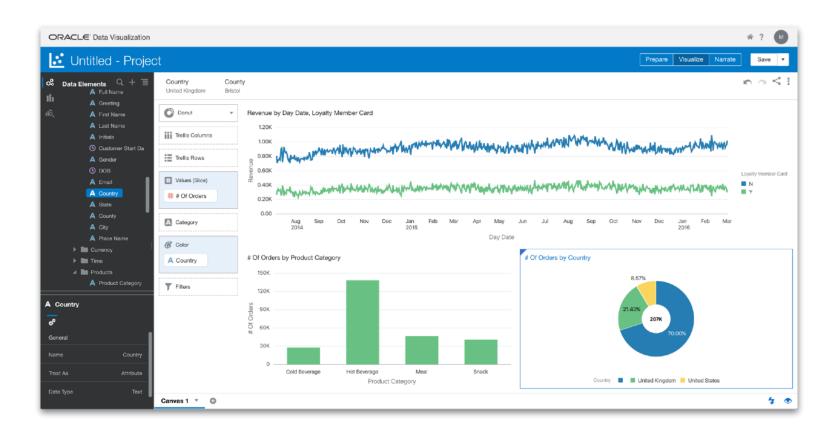
Functions

OAC supports **Every** type of analytics

Classic

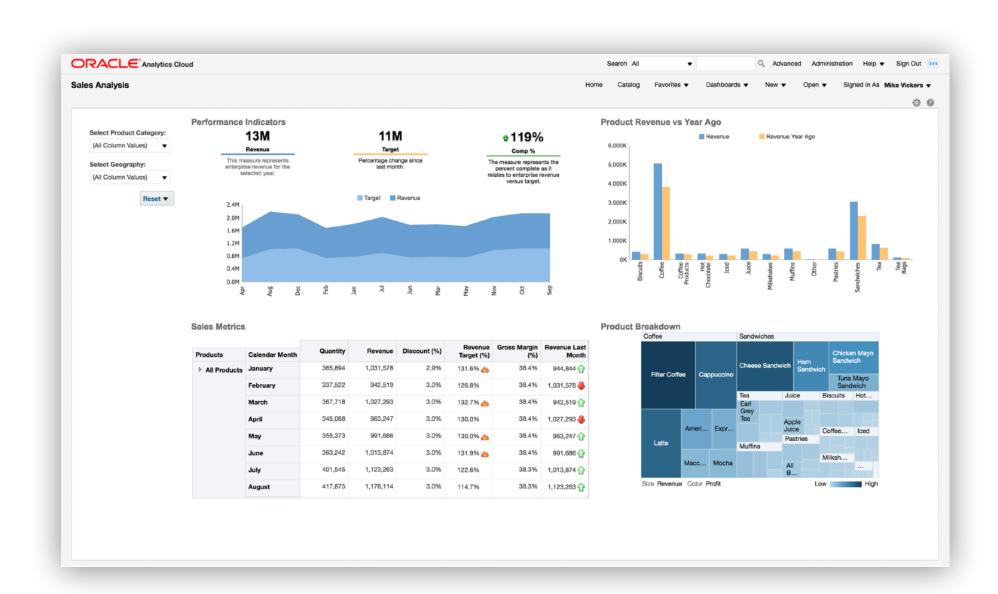
Sales Analysis | Performance Indicators | Per

Modern



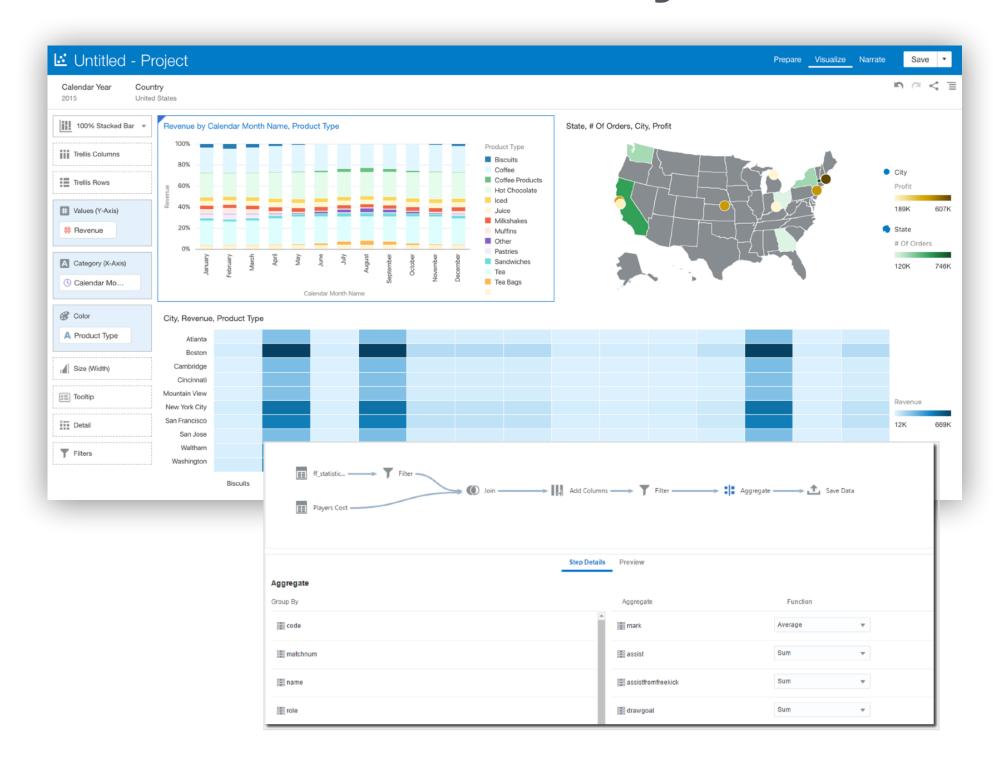
Classic Enterprise Bl

- Similar to OBIEE 12c
 - Centrally maintained & governed
 - Semantic model
- Interactive Dashboards
 - KPI measurement & monitoring
 - Guided navigation paths
- BI Publisher
 - Highly formatted, burst outputs
- Action Framework
 - Navigation actions
 - Scheduled agents



Modern Data Discovery

- Data Preparation
 - Acquire data
 - •Clean/Enrich
 - Transform
 - Repeatable Flows
- Data Visualisation
 - Create visual insights rapidly
 - Construct narrated storyboards
 - Share findings



Unified Analytics

Centralised Reporting Free Discovery

Unique Source of Truth

Raw Data
To Insights

Data
Enrichment
and
Cleaning

Specific Access
Control

Augmented Analytics

Data Enrichment Suggestions

Natural Language Processing

Explain

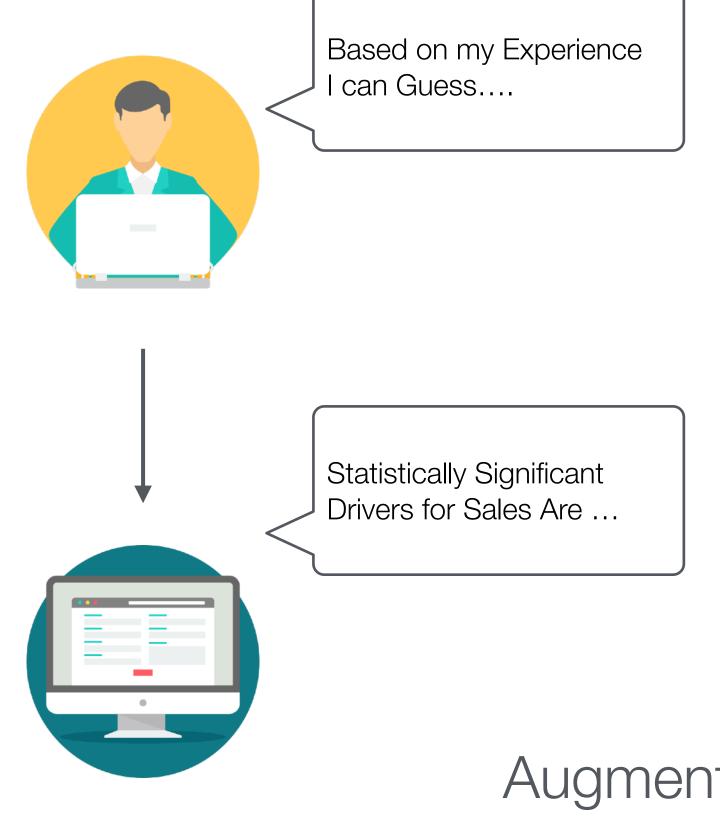


Advanced Machine Learning



Basic Operations





Augmented Analytics

Basic Operations





Before Starting.... Define the Problem!





Rule Based

Italy or France -> Good

Rest of the World -> Bad

Price >= 10 Euros -> Good

Price < 10 Euros -> Bad

Price > 30 & Production Zone = Veneto & -> 6.5

TEP

ask

Experience

Performance

Estimate Wine Good/Bad

Corpus of Wines
Descriptions with Ratings

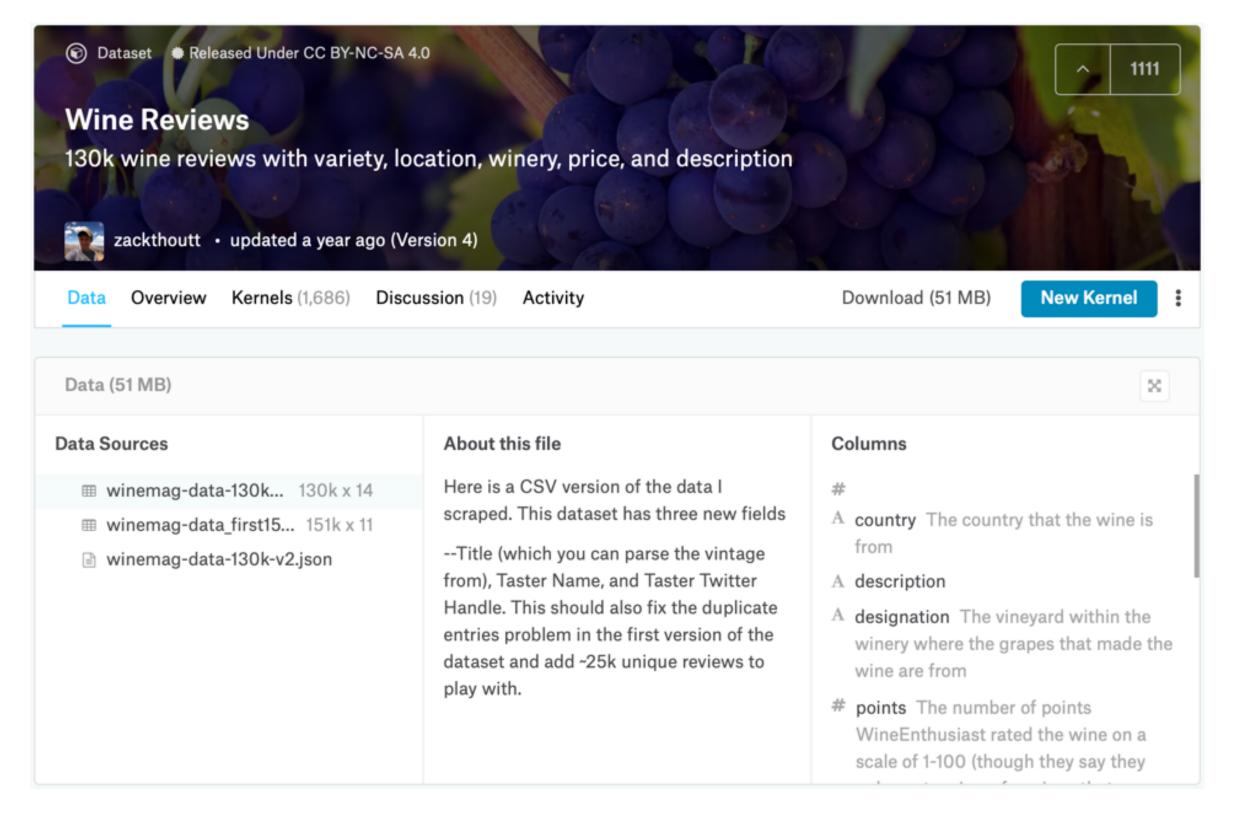
Accuracy

Accuracy

Predicted Value

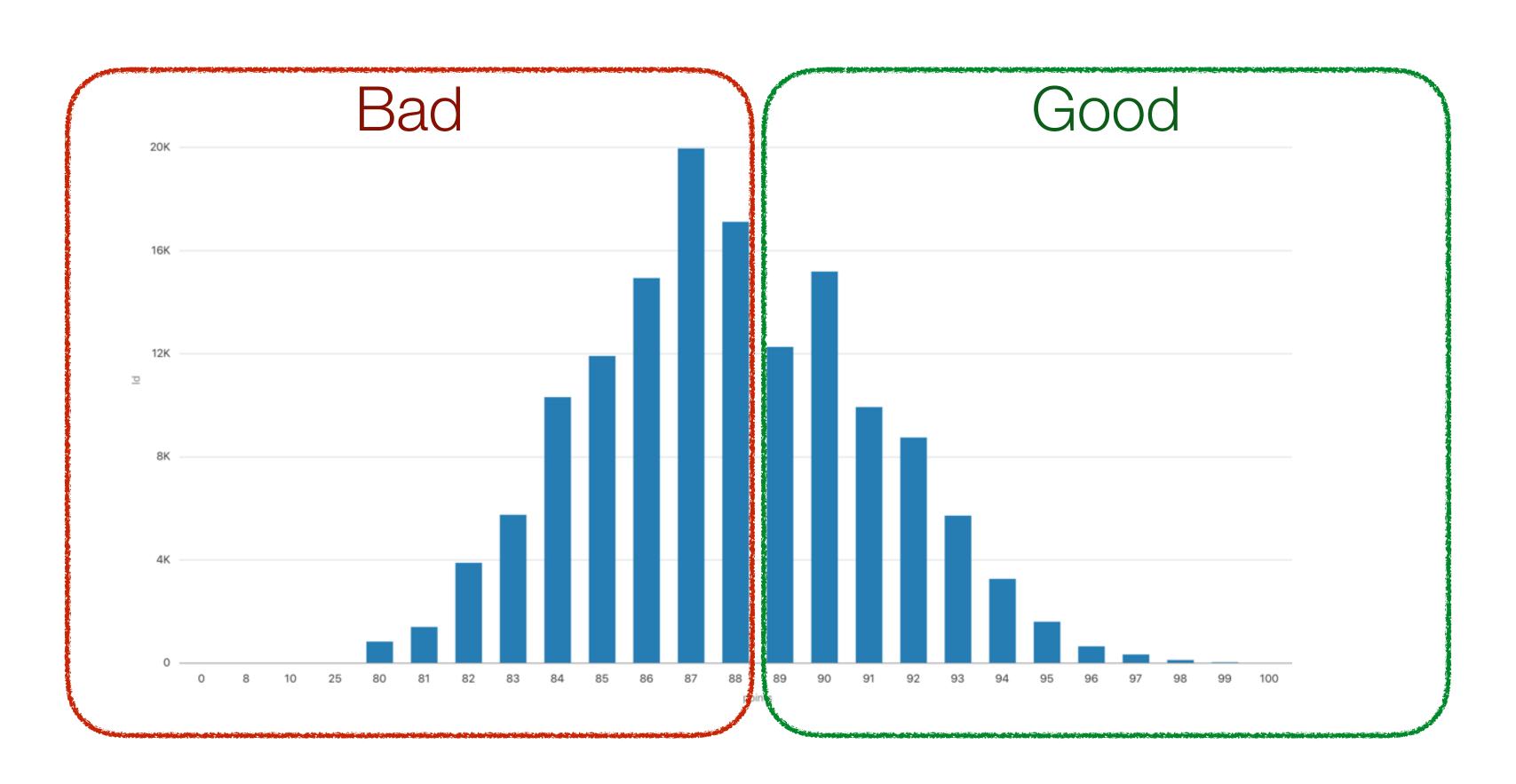
		Good	Bad
Value	Good		
Real	Bad		

Dataset



The Data

А	В	С	D	F		F	G	Н	I	J	K	
Id	▼ country	▼ description ▼	designation 🔻	points	rice	•	province	▼ region_1 ▼	▼ region_2 ▼	▼ variety ▼	▼ winery ▼	₩
	0 US	This tremendous 100	009 Martha's Vineyard	96	ۇ	235	5 California	Napa Valley	Napa	Cabernet Sauvignon	n Heitz	
	1 Spain	Ripe aromas of fig, b'	bla Carodorum Selecció	/ <mark>≥</mark> r 96	ۇ	110	Northern Spain	Toro		Tinta de Toro	Bodega Carmen Rodi	dríguez
	2 US	Mac Watson honors	rs ti Special Selected Late	t€ 96	ا ز	90	O California	Knights Valley	Sonoma	Sauvignon Blanc	Macauley	
	3 US	This spent 20 months	ns Reserve	96	ۇ	65	5 Oregon	Willamette Valley	Willamette Valley	Pinot Noir	Ponzi	
	4 France	This is the top wine fr	frc La BrVªlade	95	ا د	66	6 Provence	Bandol		Provence red blend	Domaine de la Bég	gude
	5 Spain	Deep, dense and pure	∡re Numanthia	95	ا ز	73	Northern Spain	Toro		Tinta de Toro	Numanthia	
	6 Spain	Slightly gritty black-fr	frı San Romv°n	95	ا ز	65	Northern Spain	Toro		Tinta de Toro	Maurodos	
	7 Spain	Lush cedary black-fr	frui Carodorum Vönico Cı	CI 95	غ	110	Northern Spain	Toro		Tinta de Toro	Bodega Carmen Rodi	dríguez
	8 US	This re-named vineya	yar Silice	95	ا ز	65	5 Oregon	Chehalem Mountains	s Willamette Valley	Pinot Noir	Bergström	
	9 US	The producer source	ces Gap's Crown Vineyar	ard 95	غ	60	O California	Sonoma Coast	Sonoma	Pinot Noir	Blue Farm	
	10 Italy	Elegance, complexit	ity Ronco della Chiesa	95	غ ا	80	Northeastern Italy	y Collio		Friulano	Borgo del Tiglio	
	11 US	From 18-year-old vir	ine Estate Vineyard Wac	act 95	غ	48	8 Oregon	Ribbon Ridge	Willamette Valley	Pinot Noir	Patricia Green Cellars	ırs
	12 US	A standout even in th	thi Weber Vineyard	95	ا ف	48	8 Oregon	Dundee Hills	Willamette Valley	Pinot Noir	Patricia Green Cellars	ırs
	13 France	This wine is in peak r	co Ch√¢teau Montus Pr	Pr 95	غ	90	Southwest France	Madiran		Tannat	Vignobles Brumont	
	14 US	With its sophisticated	ed Grace Vineyard	95	غ ا	185	5 Oregon	Dundee Hills	Willamette Valley	Pinot Noir	Domaine Serene	
	15 US	First made in 2006, th	th Sigrid	95	خ	90	O Oregon	Willamette Valley	Willamette Valley	Chardonnay	Bergström	
	16 US	This blockbuster, pow	w Rainin Vineyard	95	غ ا	325	5 California	Diamond Mountain D	⊘ Napa	Cabernet Sauvignon	n Hall	
	17 Spain	Nicely oaked blackb	ber 6 AV±os Reserva Pre	e r 95	خ ا	80	Northern Spain	Ribera del Duero		Tempranillo	Valduero	
	18 France	Coming from a seven	∠n- Le Pigeonnier	95	خ	290	Southwest France	Cahors		Malbec	Château Lagrézet	ette
	19 US	This fresh and lively	y m Gap's Crown Vineyar	ard 95	خ ا	75	5 California	Sonoma Coast	Sonoma	Pinot Noir	Gary Farrell	
									1		1	



Become a Data Scientist with OAC

Connect

Clean

Transform & Enrich

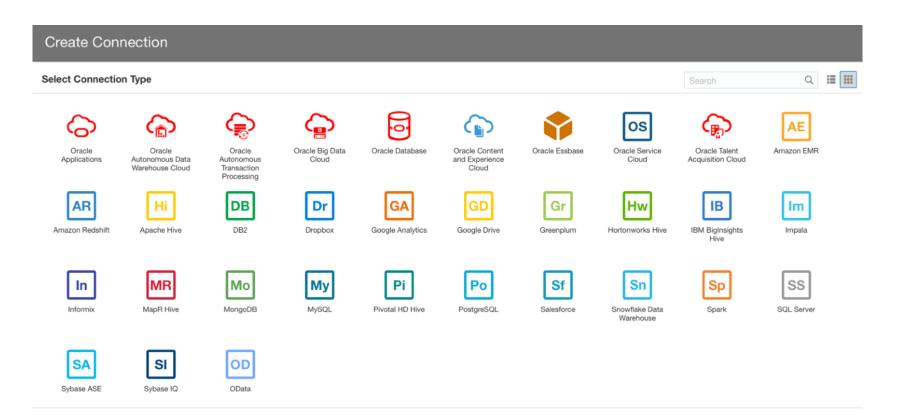
Analyse

Train & Evaluate

Predict

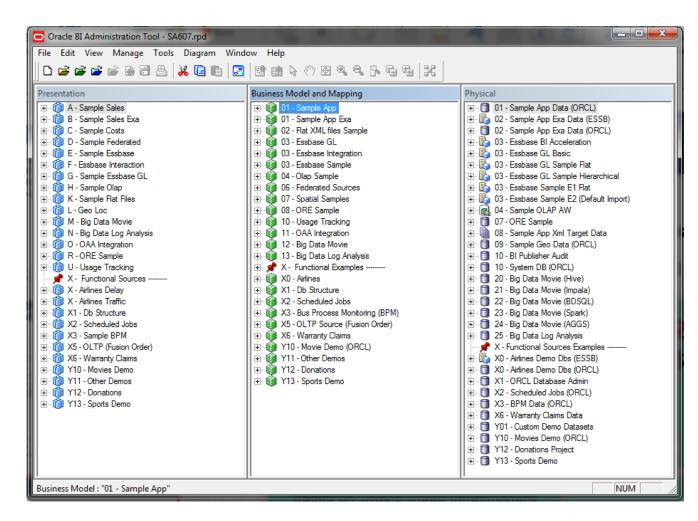
Connection Options in OAC

Cancel



External Data Sources

Pre-Defined Data Models



Cleaning What?







Irrelevant Observations





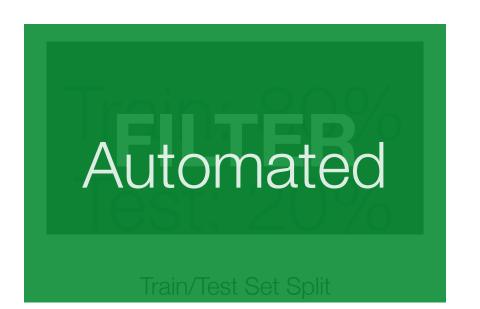


Labelling Columns

Handling Outliers



Aggregation



Feature Engineering

Location -> ZIP Code

Additional Data Sources?

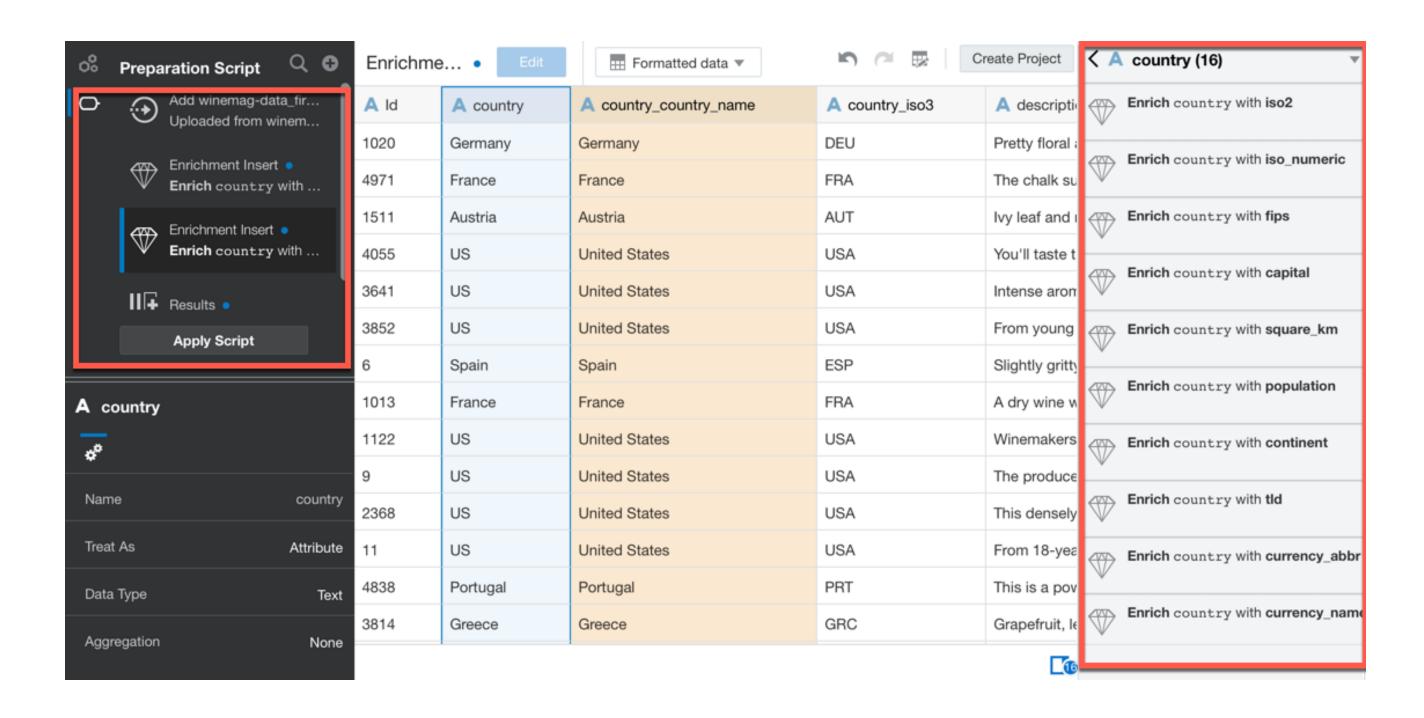
Name -> Sex

2 Locations -> Distance

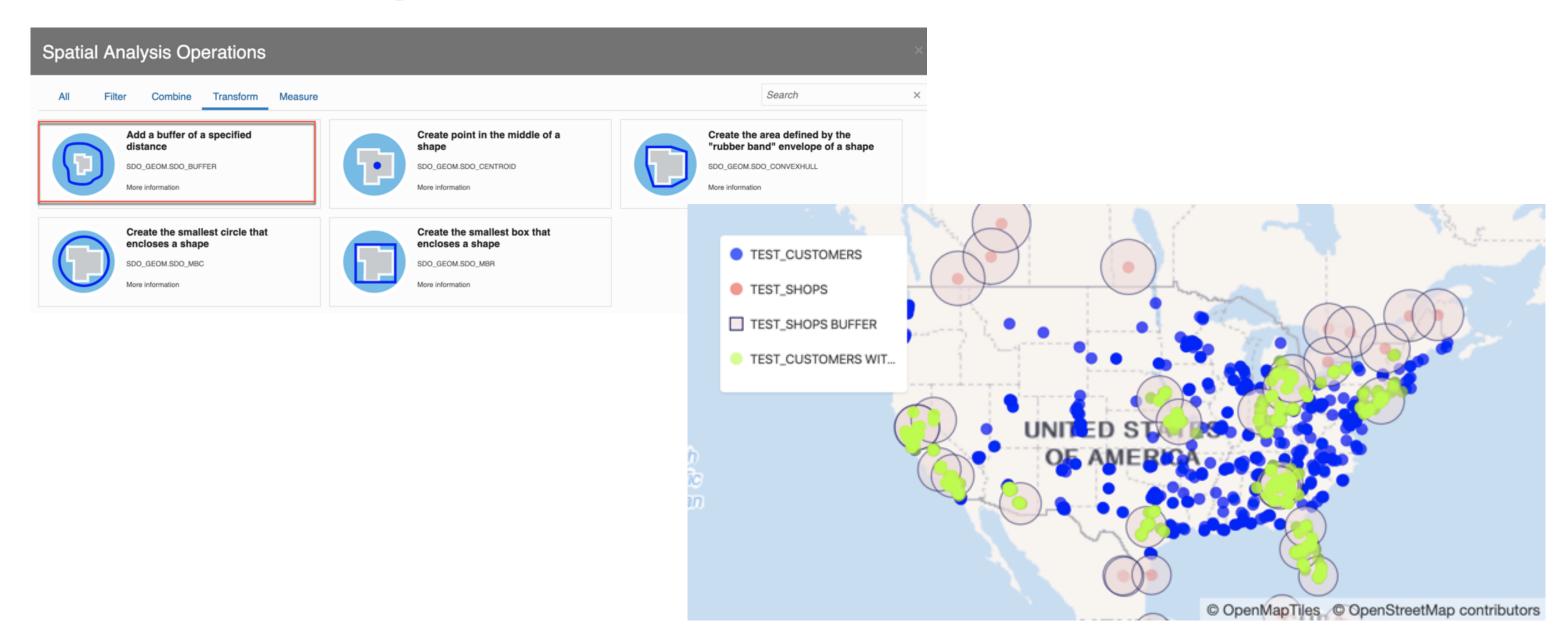
Data Flow

Day/Month/Year -> Date

Data Preparation Recommendations



Spatial Enrichment



Oracle Spatial Studio

http://ritt.md/spatial-studio

Data Overview

Results

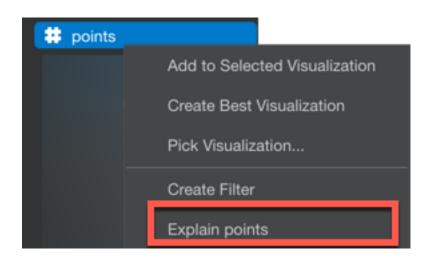


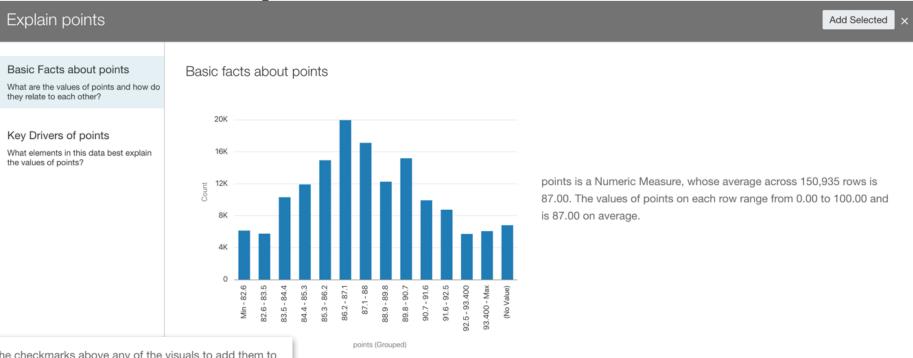


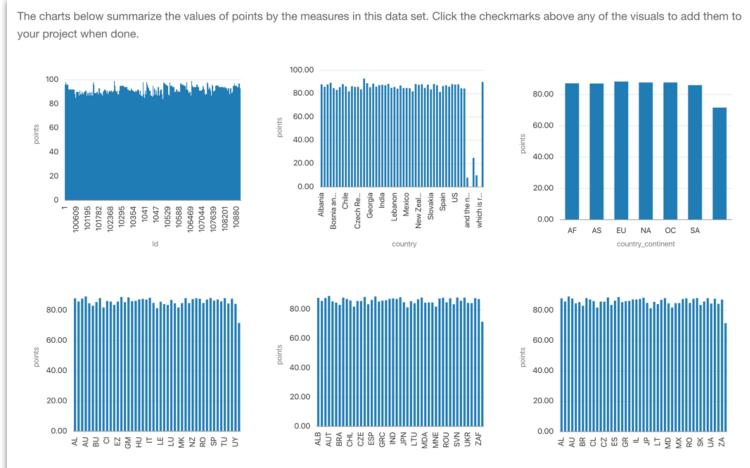


Data Element	Data Type	Treat As	Aggregation	Sample Values
Id	varchar(80)	A Attribute	none	1470; 817; 1028; 632; 3689; 4148; 2576; 963; 4979; 281
country	varchar(137)	A Attribute	none	US; France; Italy; Spain; Portugal; Germany; Argentina; Chile; Austria; Greece
country_continent	varchar(4000)	A Attribute	none	NA; EU
country_fips	varchar(4000)	A Attribute	none	US; IT; FR
country_iso3	varchar(4000)	A Attribute	none	USA; FRA; ITA
country_iso_numeric	number	# Measure	sum	840; 380; 250
country_iso2	varchar(4000)	A Attribute	none	US; IT; FR
description	varchar(1247)	A Attribute	none	This elegant wine combines subtle nutmeg and cardamom aromas with crisp app
designation	varchar(122)	A Attribute	none	Reserve; Estate; Reserva; Riserva; Estate Bottled; Vieilles Vignes; Crianza; Classic
points	number	# Measure	sum	90; 89; 88; 87; 91; 86; 92; 93; 85; 94
price	varchar(15)	A Attribute	none	25; 20; 40; 18; 60; 30; 28; 35; 50; 15
province	varchar(53)	A Attribute	none	California; Oregon; Bordeaux; Tuscany; Piedmont; Washington; Northern Spain; M
region_1	varchar(75)	A Attribute	none	Willamette Valley; Napa Valley; Barolo; Brunello di Montalcino; Russian River Valle
region_2	varchar(35)	A Attribute	none	Central Coast; Sonoma; Willamette Valley; Napa; Columbia Valley; Mendocino/La
variety	varchar(53)	A Attribute	none	Pinot Noir; Chardonnay; Bordeaux-style Red Blend; Cabernet Sauvignon; Red Ble
winery	varchar(84)	A Attribute	none	Tarara; Heron Hill; Byron; Bergström; Herdade do Rocim; Rusack; Sarah's Viney

Analyse - Explain





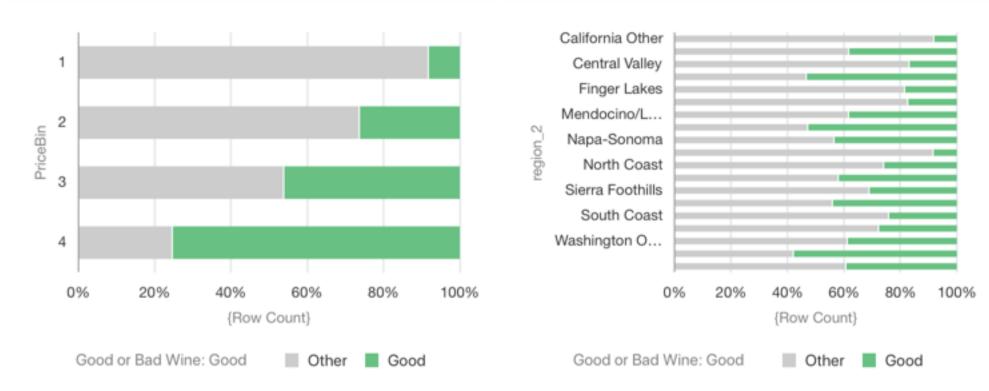


Explain - Key Drivers

Key Drivers of Good or Bad Wine

Based on Good or Bad Wine: Good the 2 attributes that are most strongly correlated are: PriceBin, region_2

The charts below show the distribution of Good or Bad Wine values across each of the key drivers. Click the checkmarks above any of the visuals to add them to your project when done.

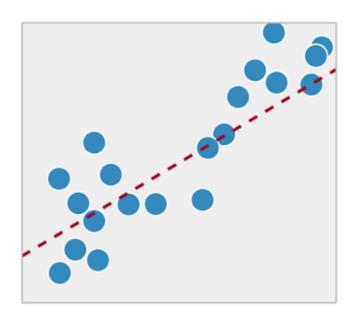


Train - What Problem are we Trying to Solve?

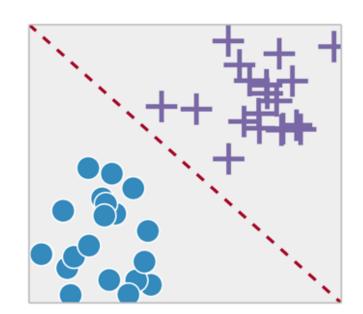
Supervised

"I want to predict the value of Y, here are some examples"

Regression



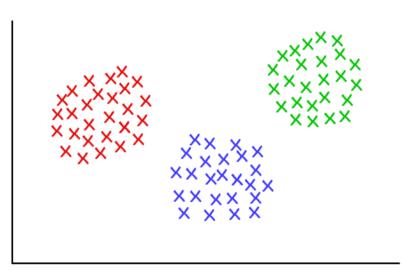
Classification



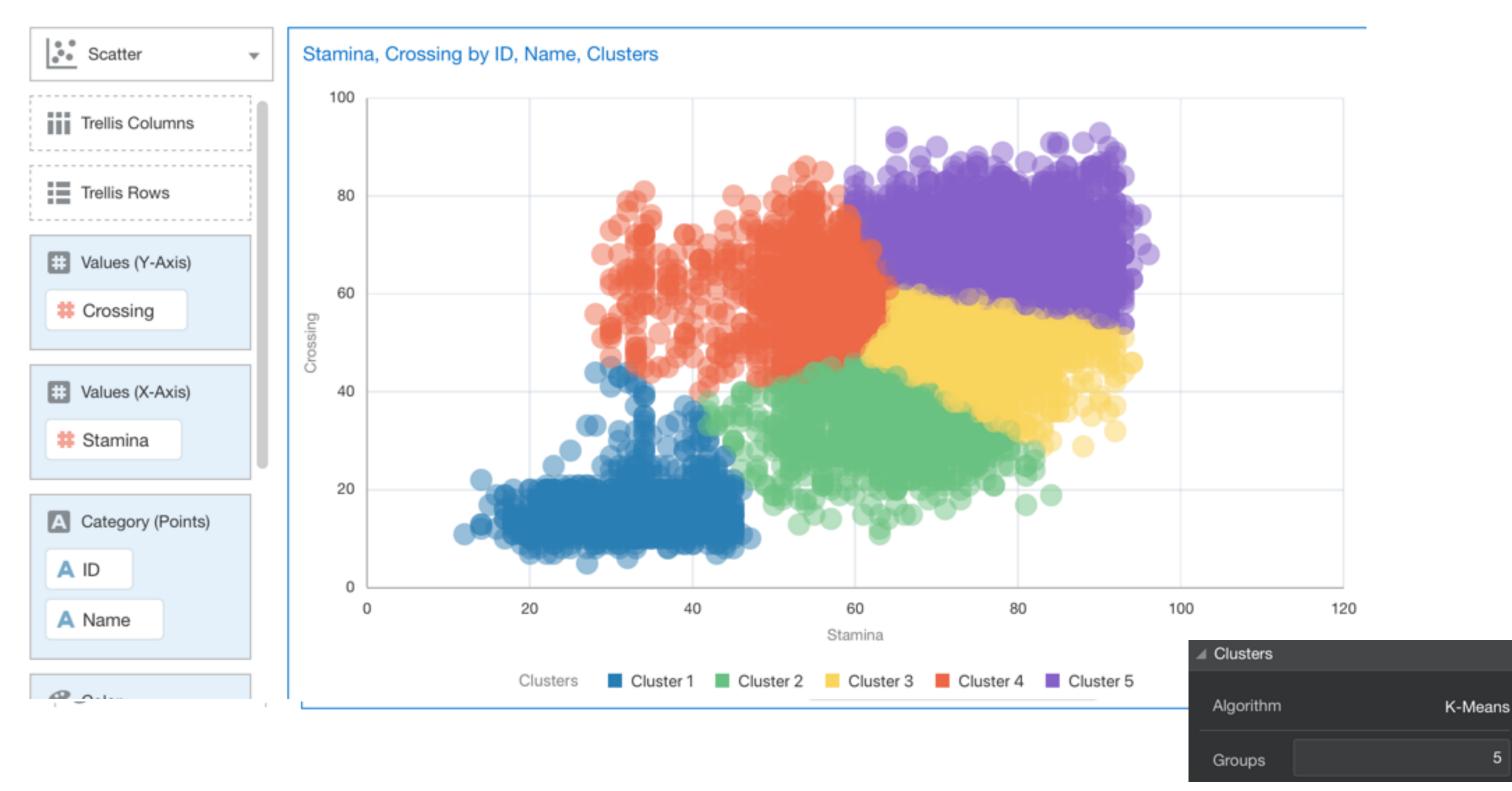
Unsupervised

"Here is a dataset, make sense out of it!"

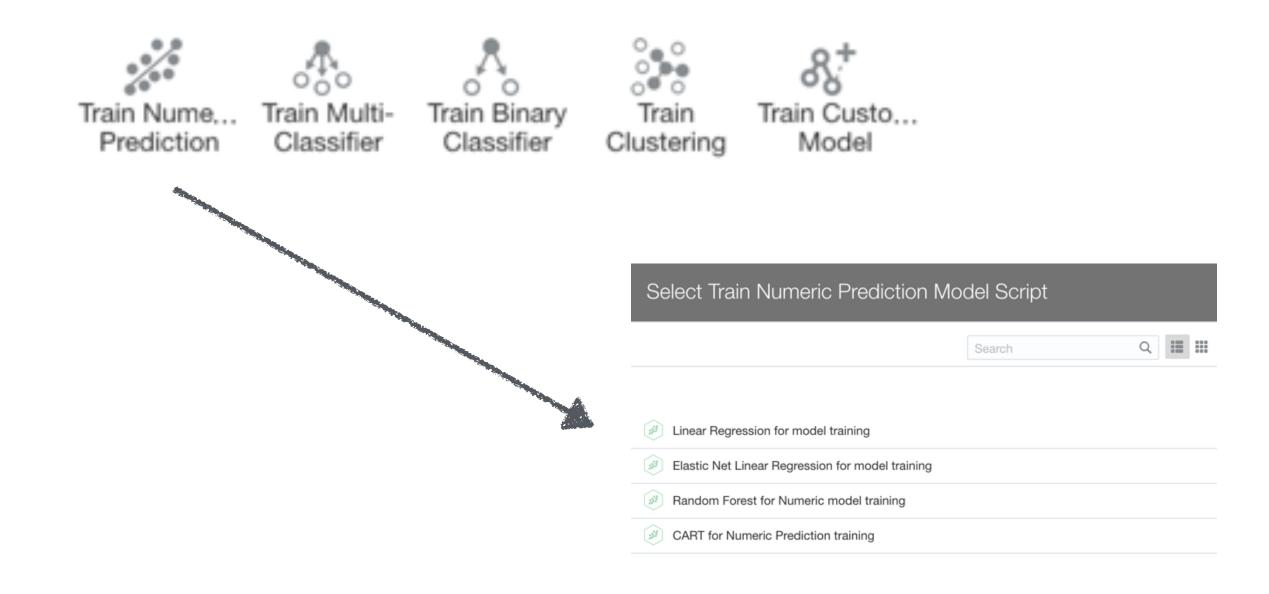
Clustering



Model Training - Easy Models

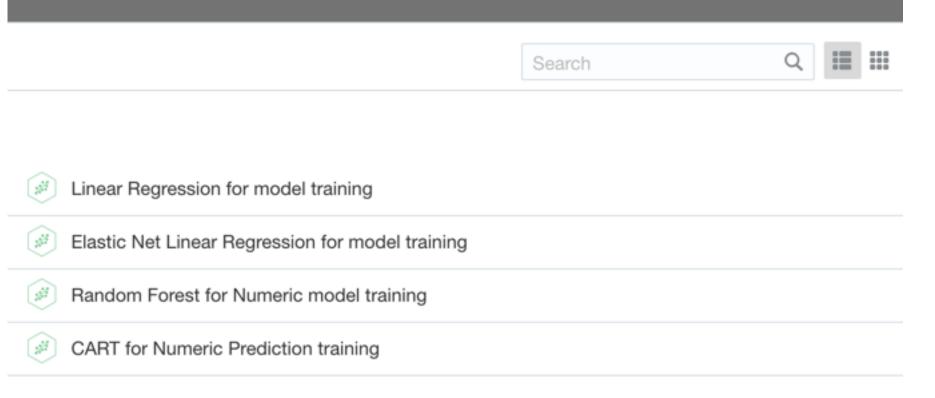


DataFlow Train Model

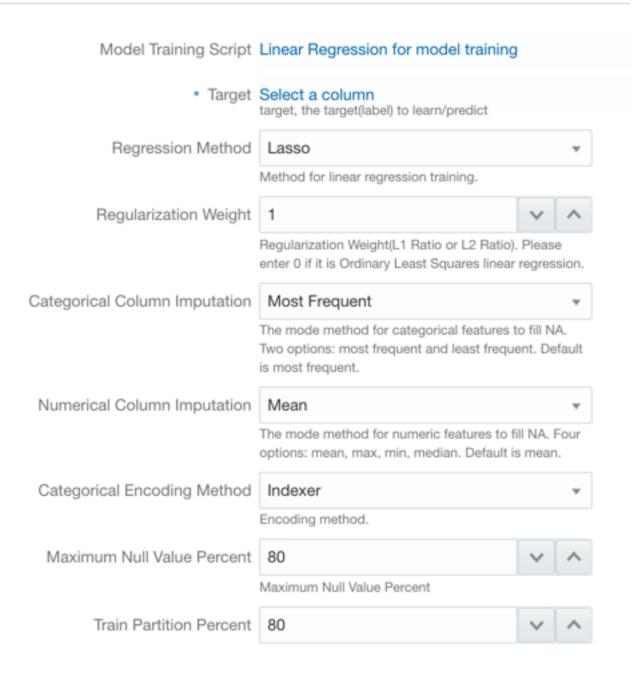


Which Model - Parameters To Pick?

Select Train Numeric Prediction Model Script



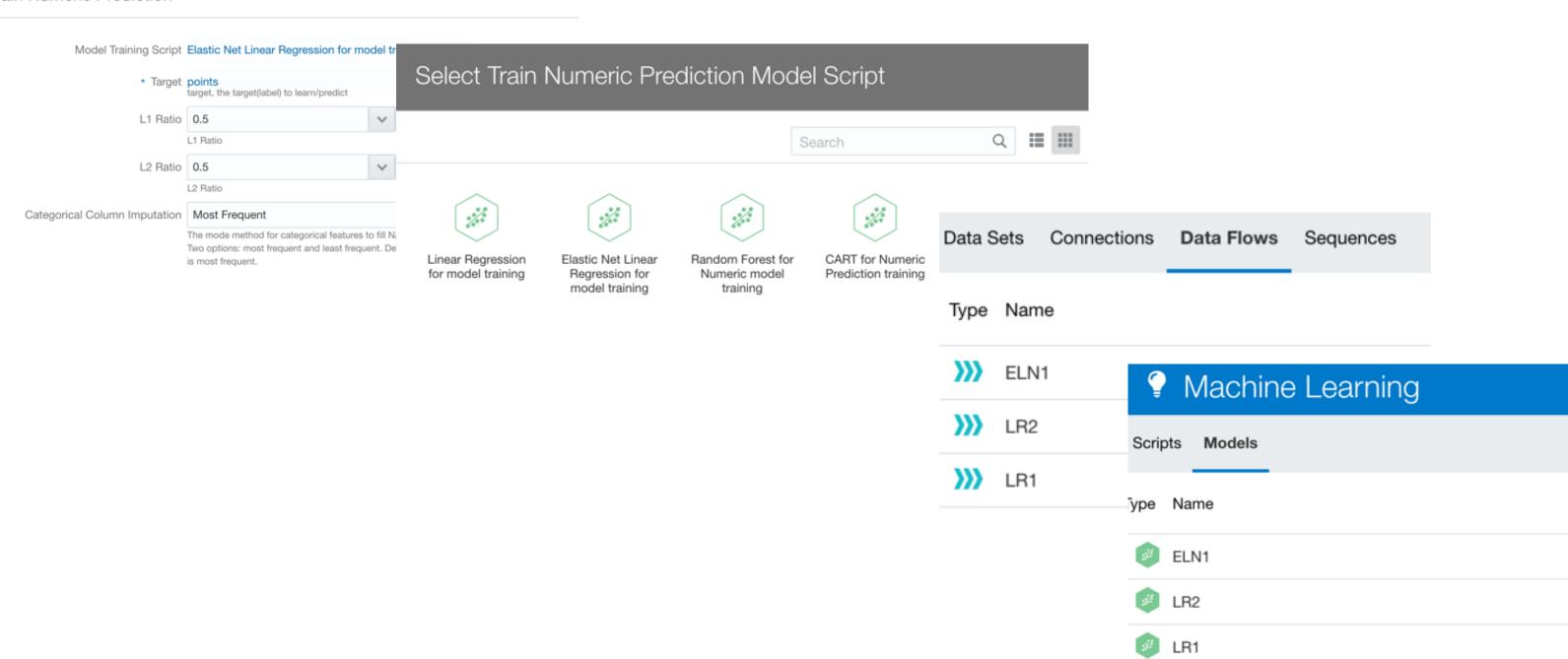
Train Numeric Prediction



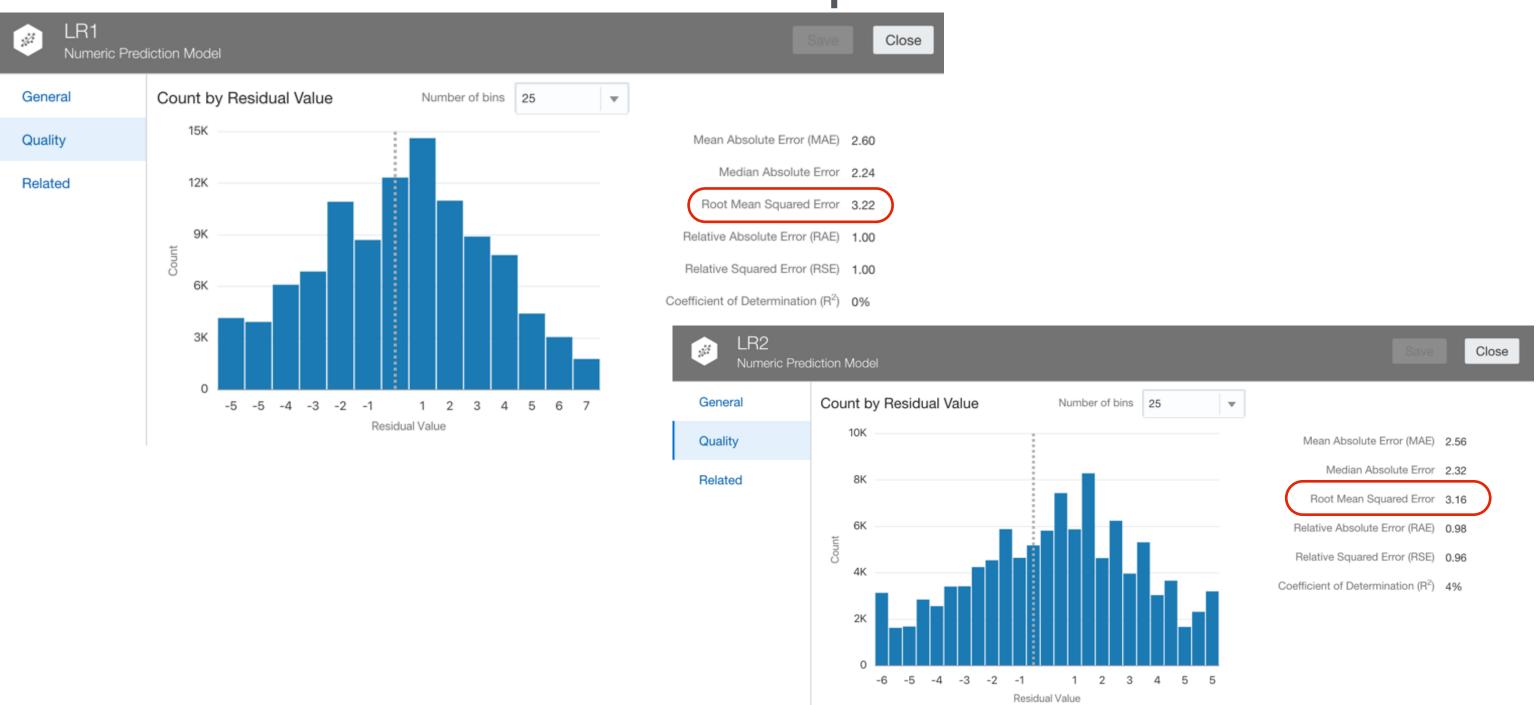
Select, Try, Save, Change, Try, Save



Train Numeric Prediction

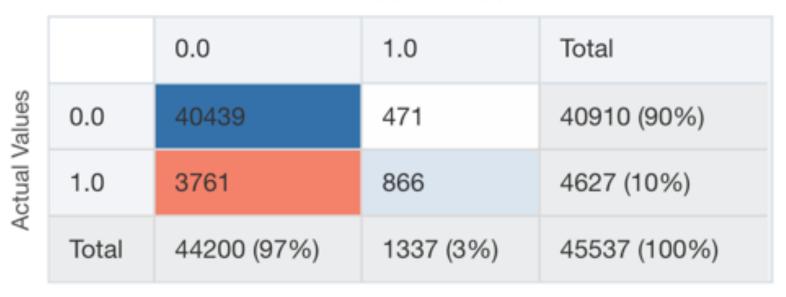


Compare

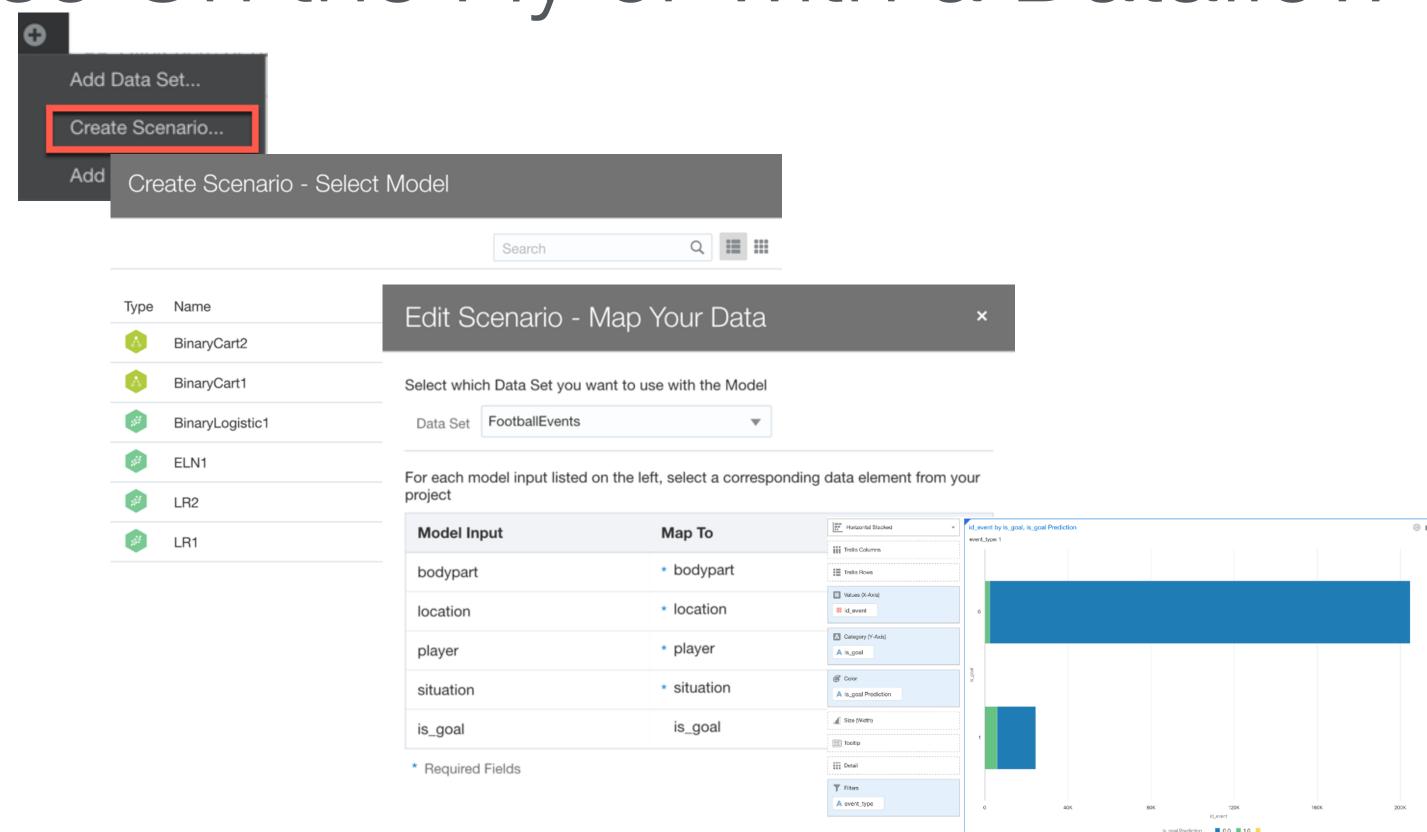


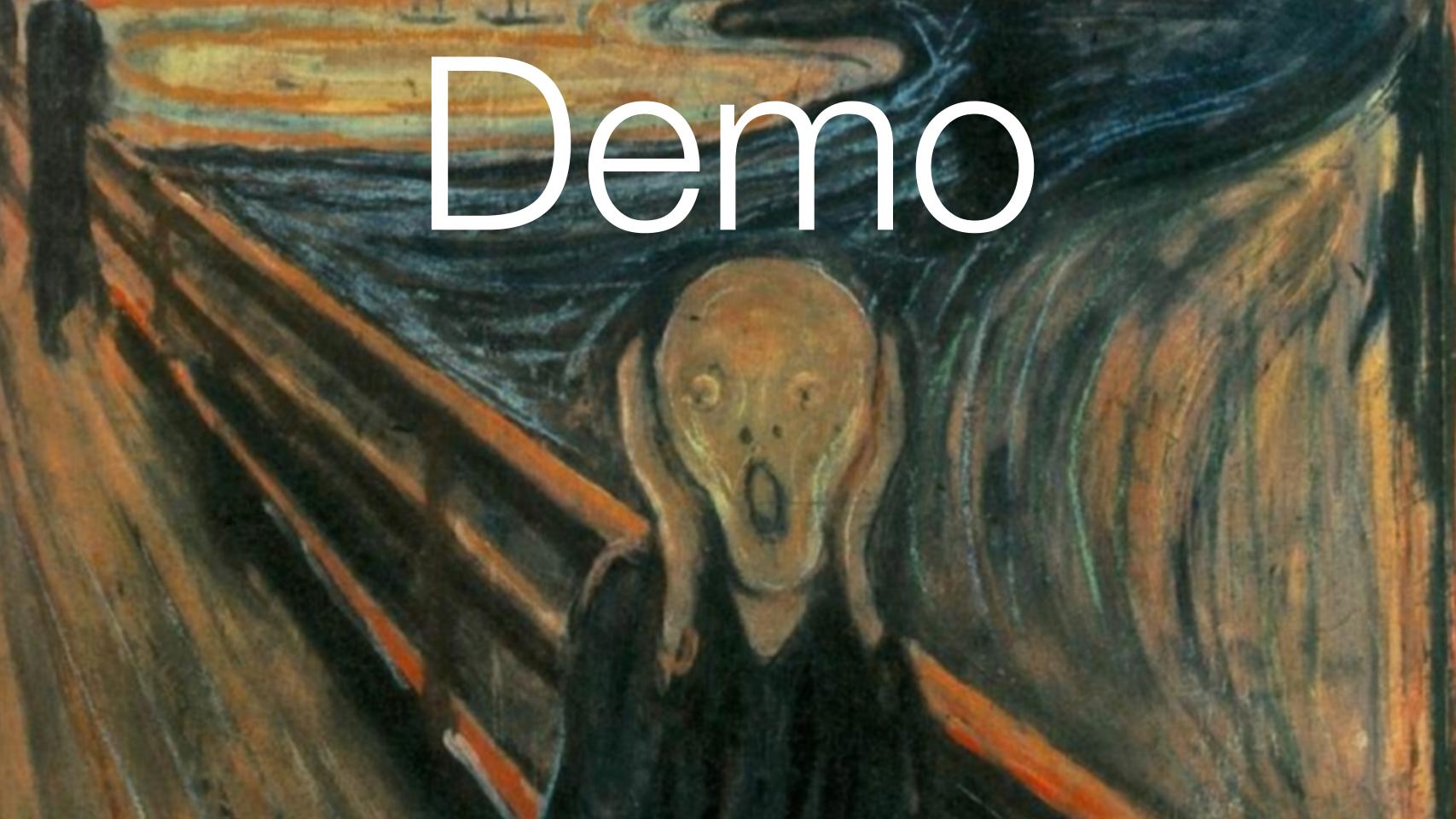
Compare - Classification

Predicted Values



Use On the Fly or with a Dataflow





Congratulations!



...You are now a Data Scientist!

ML Production Deployment

Data Scientist

ML -> Data

Oracle Machine Learning







Oracle Machine Learning

OML4SQL

Oracle Advanced Analytics
SQL API

OML4R

Oracle R Enterprise R API

OML4Py*

Python API

OML Notebooks

with Apache Zeppelin on Autonomous Database

Oracle Data Miner

Oracle SQL Developer extension

OML4Spark

Oracle R Advanced Analytics for Hadoop

OML Microservices*

Supporting Oracle Applications Image, Text, Scoring, Deployment, Model Management









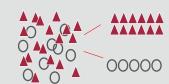


Oracle Machine Learning Algorithms



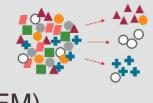
CLASSIFICATION

Naïve Bayes
Logistic Regression (GLM)
Decision Tree
Random Forest
Neural Network
Support Vector Machine
Explicit Semantic Analysis



CLUSTERING

Hierarchical K-Means
Hierarchical O-Cluster
Expectation Maximization (EM)



ANOMALY DETECTION

One-Class SVM



TIME SERIES

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

REGRESSION

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

ATTRIBUTE IMPORTANCE

Minimum Description Length Principal Comp Analysis (PCA) Unsupervised Pair-wise KL Div CUR decomposition for row & Al

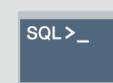


PREDICTIVE QUERIES

Predict, cluster, detect, features

SQL ANALYTICS

SQL Windows
SQL Patterns
SQL Aggregates



FEATURE EXTRACTION

Principal Comp Analysis (PCA) Non-negative Matrix Factorization Singular Value Decomposition (SVD) Explicit Semantic Analysis (ESA)

TEXT MINING SUPPORT

Algorithms support text
Tokenization and theme extraction
Explicit Semantic Analysis (ESA) for
document similarity

STATISTICAL FUNCTIONS

Basic statistics: min, max, median, stdev, t-test, F-test, Pearson's, Chi-Sq, ANOVA, etc.

R PACKAGES

Third-party R Packages through Embedded Execution Spark MLlib algorithm integration



MODEL DEPLOYMENT

SQL—1st Class Objects Oracle RESTful API (ORDS) OML Microservices (for Apps)



Oracle Machine Learning

Machine Learning Notebook for Autonomous Data Warehouse Cloud

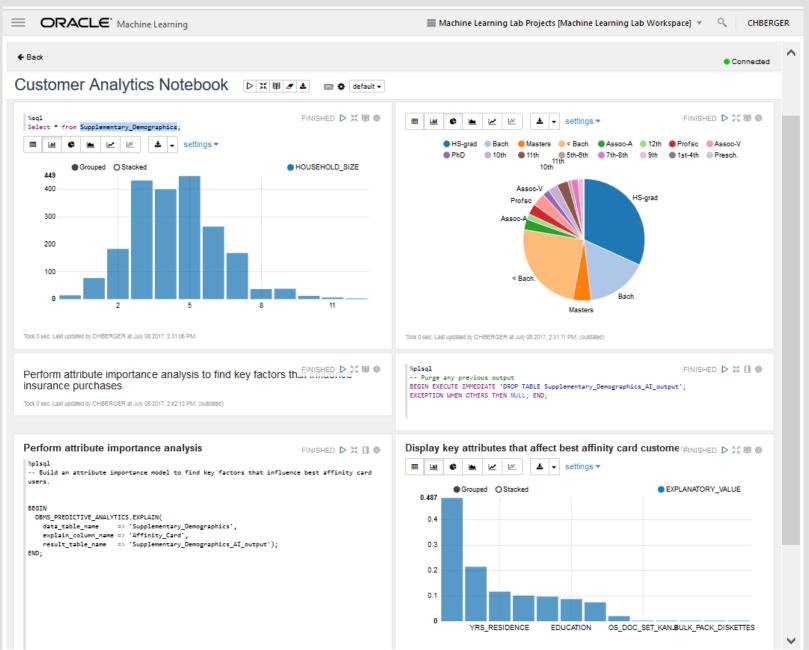


Key Features:

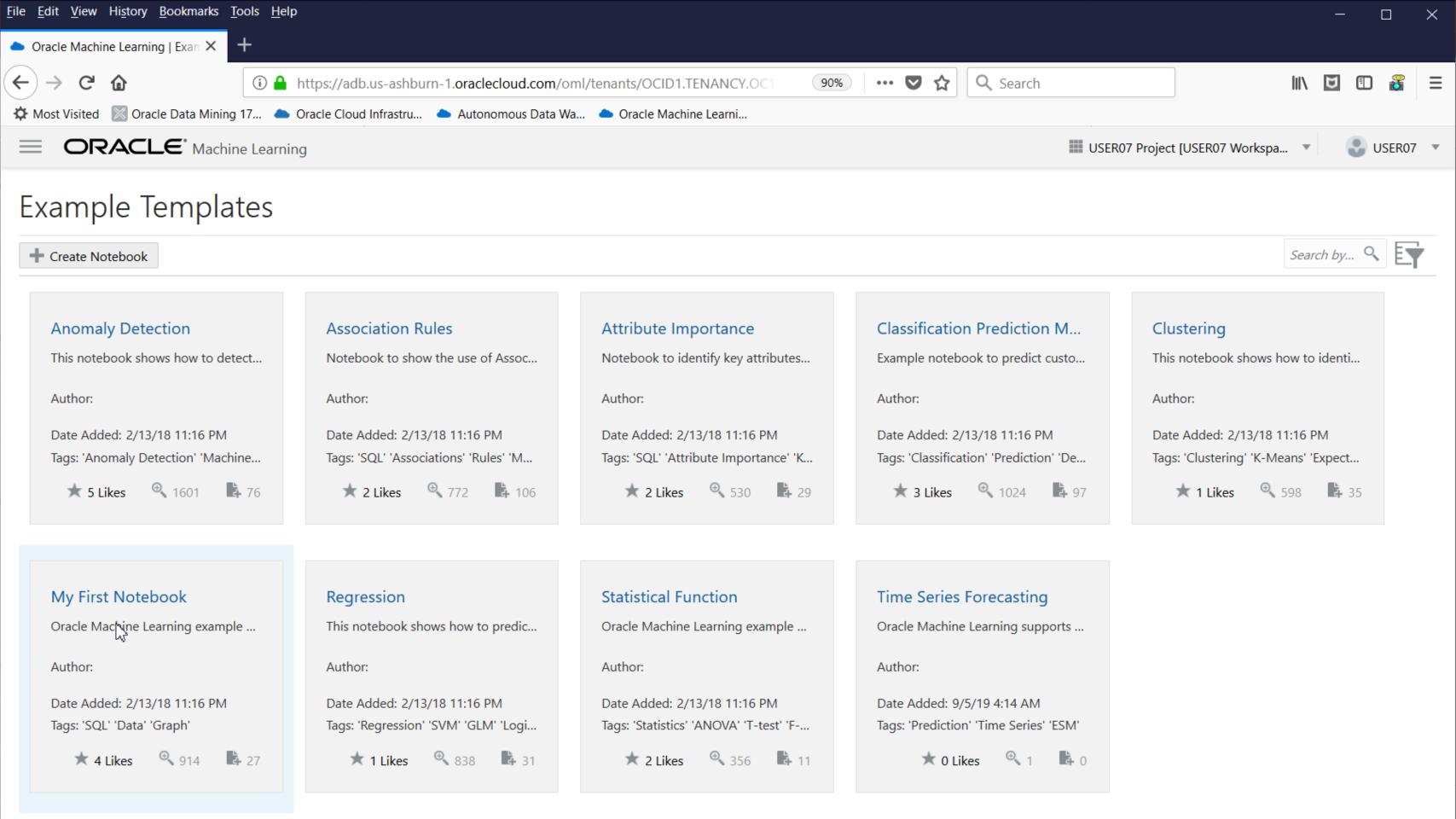
Collaborative UI for data scientists

Packaged with Autonomous Data Warehouse Cloud Easy access to shared notebooks, templates, permissions, scheduler, etc. SQL ML algorithms API

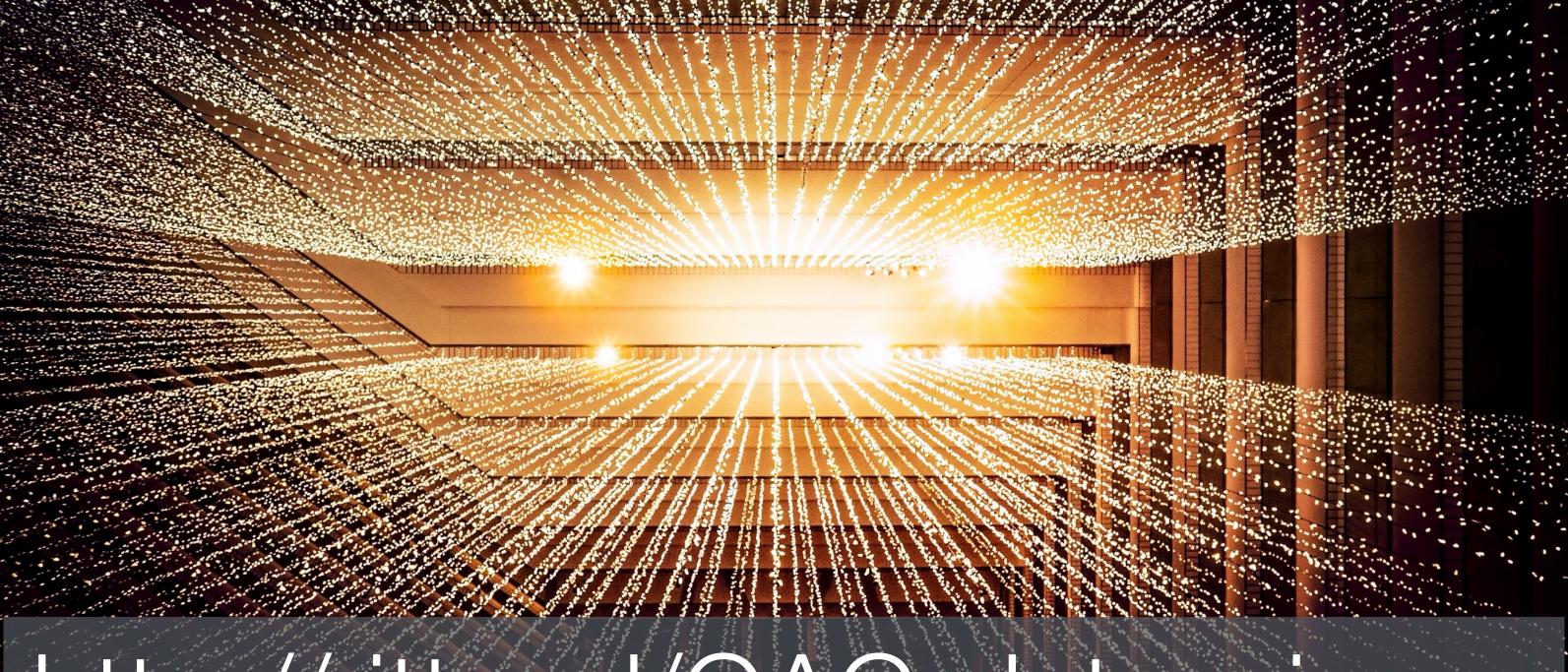
Supports deployment of ML analytics







Become a Data Scientist with OAC



http://ritt.md/OAC-datascience









