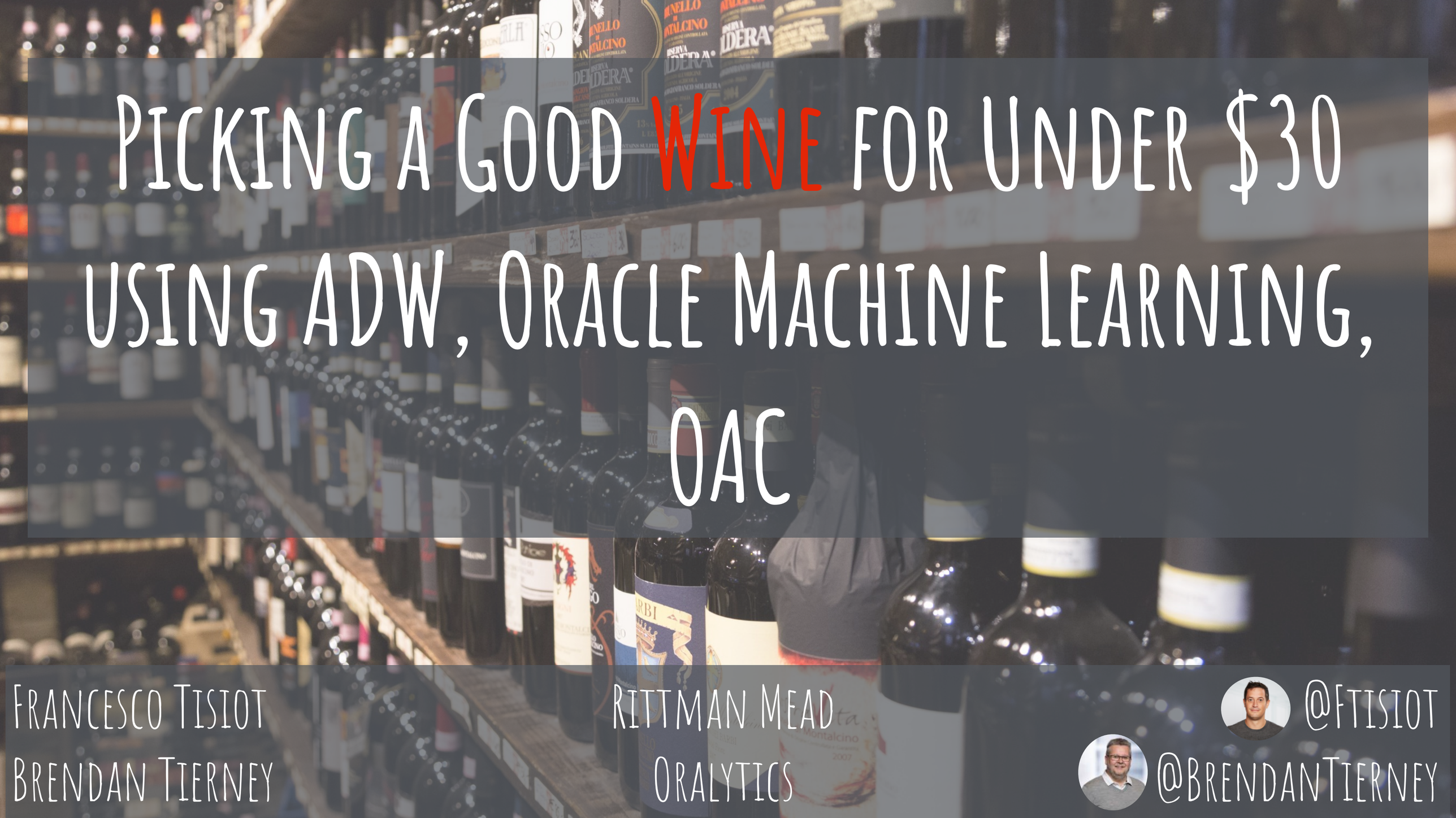


Picking a Good Wine for Under $30 using ADW, Oracle Machine Learning, OAC

Francesco Tisiot
Brendan Tierney

Rittman Mead
Oralytics

@Ftisiot

@BrendanTierney

Francesco Tisiot

Rittman Mead

@Ftisiot

Brendan Tierney

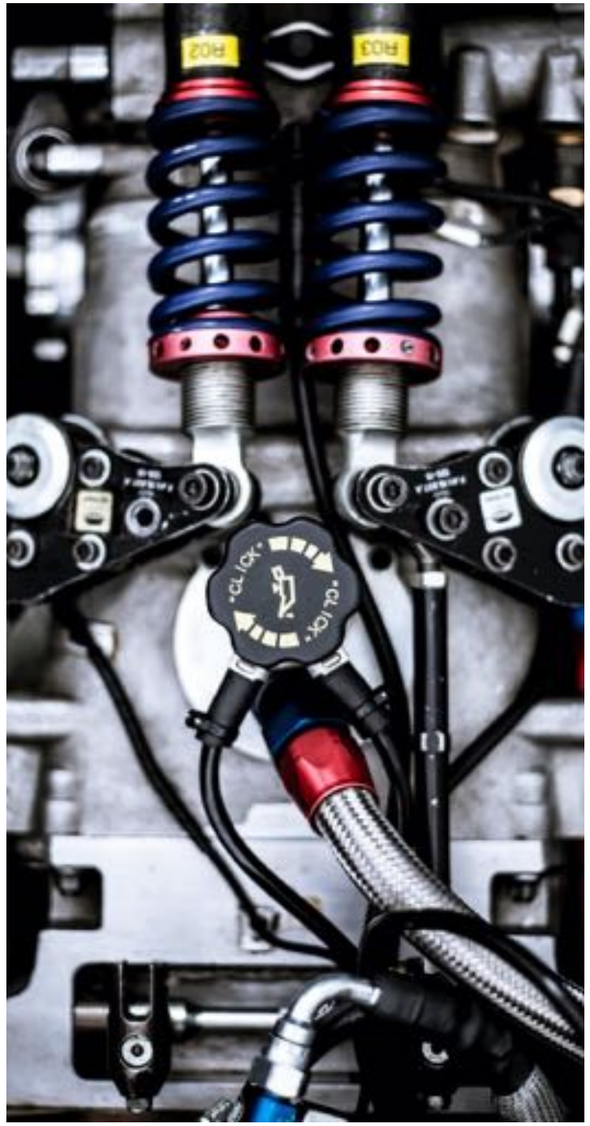
oralytics

@Brendantierney

rittmanmead

A DATA AND ANALYTICS COMPANY

INFO@RITTMANMEAD.COM
WWW.RITTMANMEAD.COM
@RITTMANMEAD



Data Engineering

Analytics

Data Science

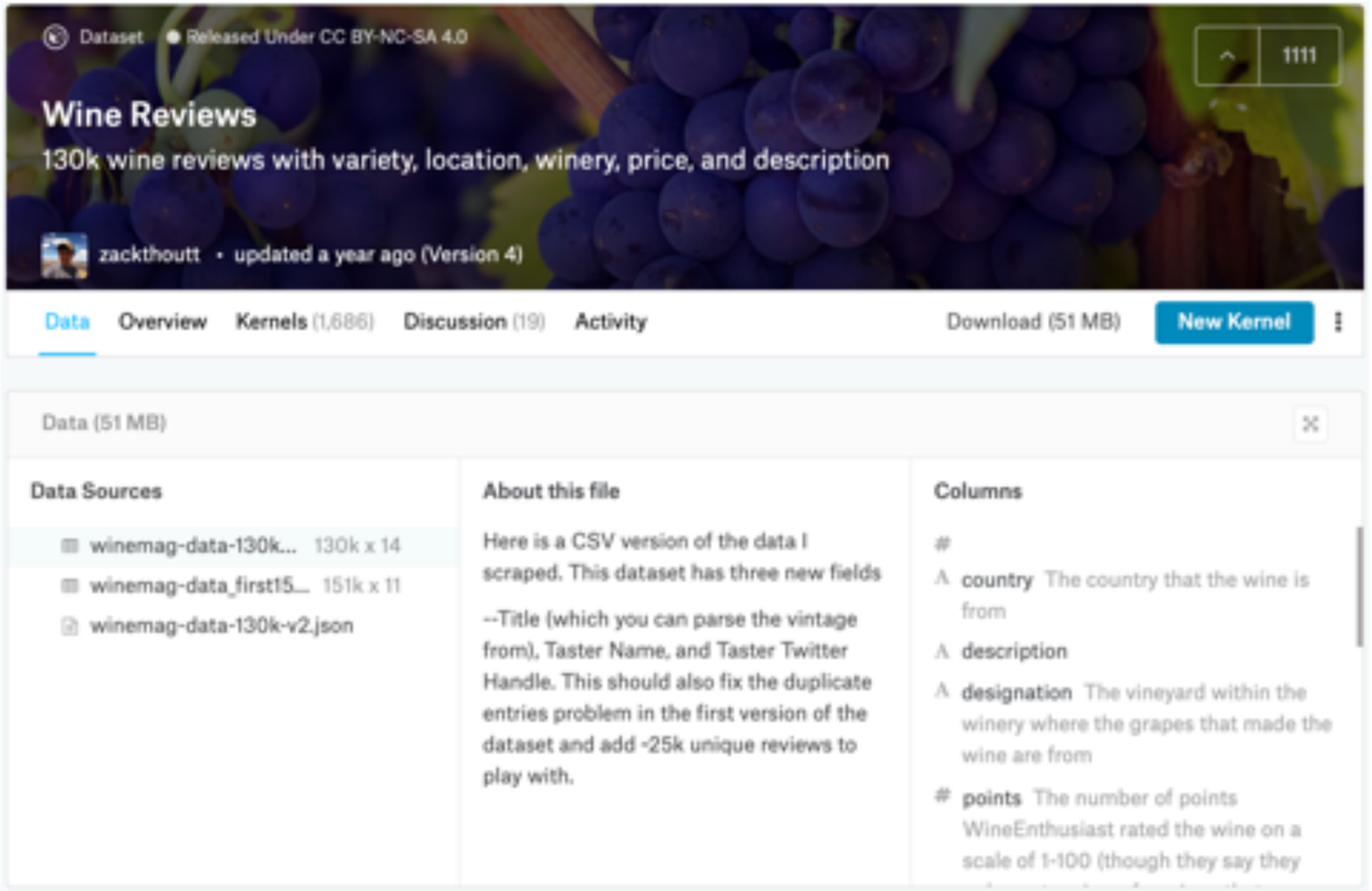

Can I have a glass of Wine?
Yes Sir, what kind of wine?
Tell me what you have there!
@Ftisiot
@BrendanTierney


Can I have a glass of Wine?
Yes Sir, what kind of wine?
Red Wine!
@Ftisiot
@BrendanTierney


Can I have a glass of Wine?
Yes,
Do you like this Nice English pinot noir??
HELL NO!
@Ftisiot
@BrendanTierney

@Ftisiot
@BrendanTierney

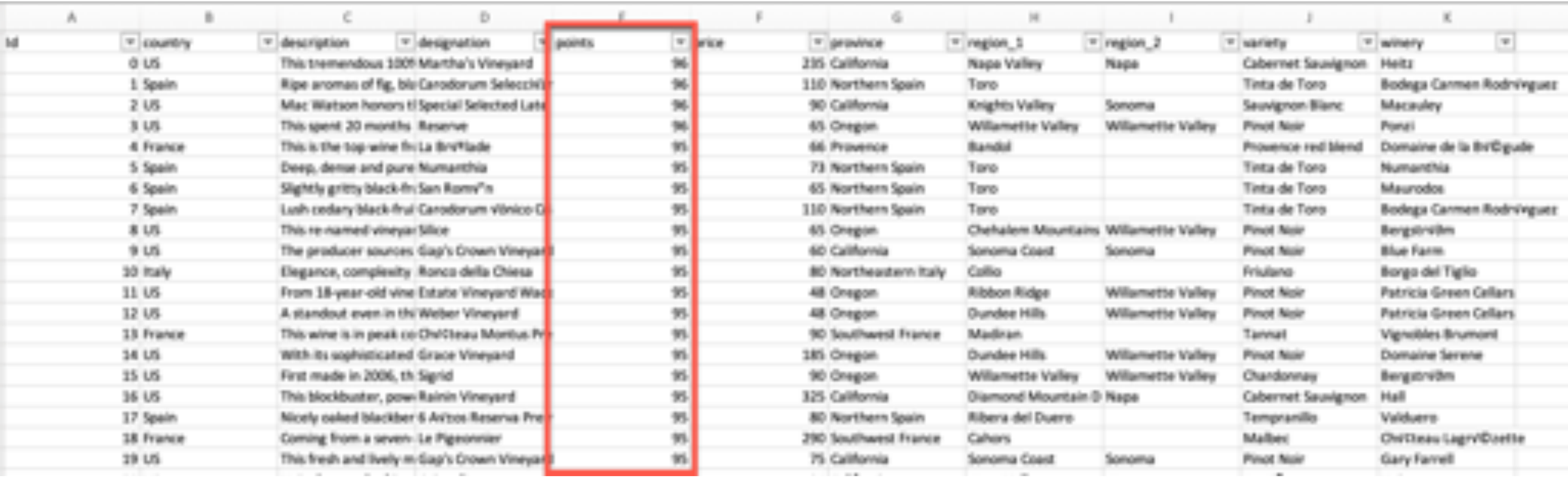
Dataset



https://www.kaggle.com/zynicide/wine-reviews


@Ftisiot
@BrendanTierney

The Data



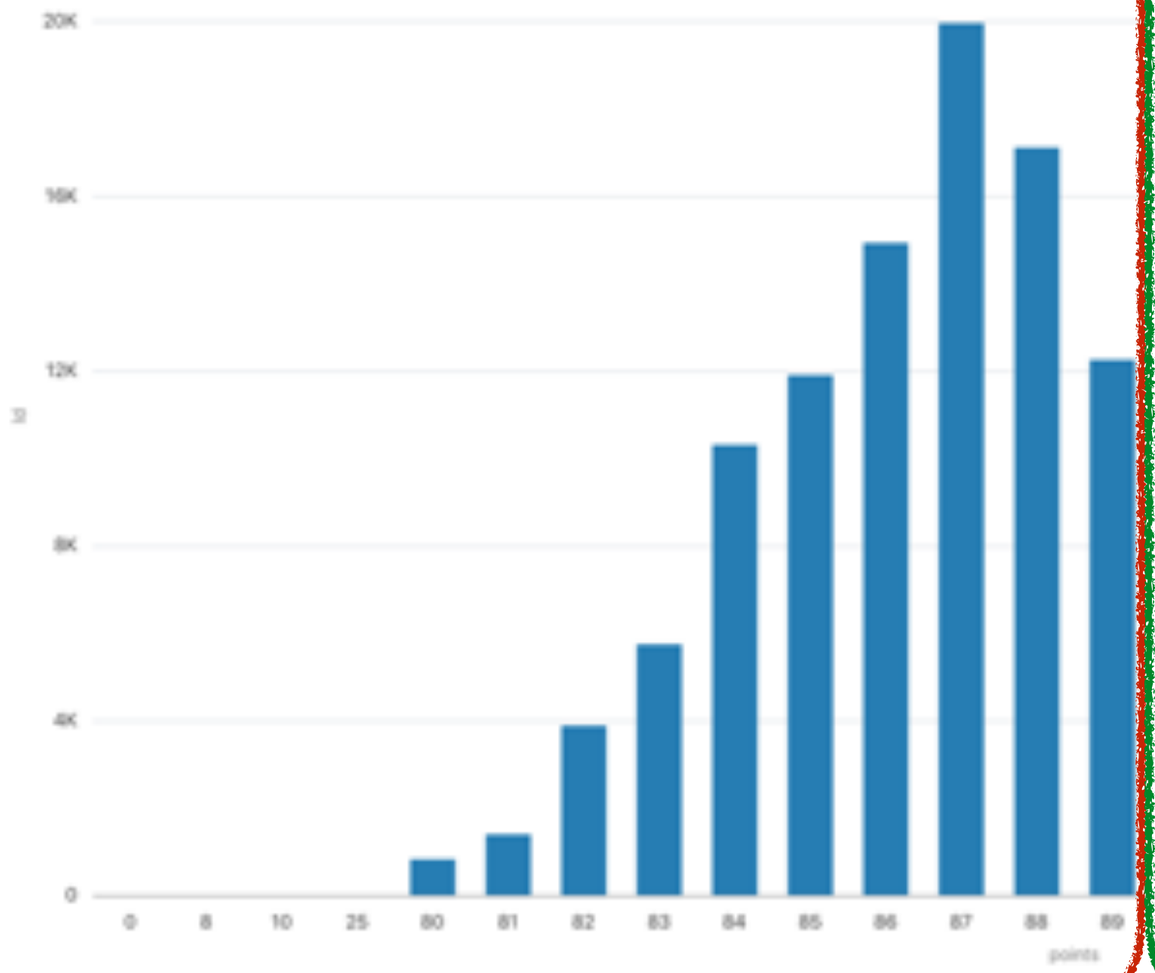
	A	B	C	D	E	F	G	H	I	J	K
	id	country	description	designation	points	rice	province	region_1	region_2	variety	winery
	0	US	This tremendous 100%	Martha's Vineyard	96	235	California	Napa Valley	Napa	Cabernet Sauvignon	Heitz
	1	Spain	Ripe aromas of fig, blo	Carodorum Selecci	96	110	Northern Spain	Toro		Tinta de Toro	Bodega Carmen Rodríguez
	2	US	Mac Watson honors t	Special Selected Late	96	90	California	Knights Valley	Sonoma	Sauvignon Blanc	Macauley
	3	US	This spent 20 months	Reserve	96	65	Oregon	Willamette Valley	Willamette Valley	Pinot Noir	Ponzi
	4	France	This is the top wine fr	La Brï¹lade	95	66	Provence	Bandol		Provence red blend	Domaine de la Bï©gude
	5	Spain	Deep, dense and pure	Numanthia	95	73	Northern Spain	Toro		Tinta de Toro	Numanthia
	6	Spain	Slightly gritty black-fr	San Roma"n	95	65	Northern Spain	Toro		Tinta de Toro	Maurodos
	7	Spain	Lush cedary black-frui	Carodorum vïnico C	95	110	Northern Spain	Toro		Tinta de Toro	Bodega Carmen Rodríguez
	8	US	This re-named vineyar	Silice	95	65	Oregon	Chehalem Mountains	Willamette Valley	Pinot Noir	Bergstrïm
	9	US	The producer sources	Gaja's Crown Vineyard	95	60	California	Sonoma Coast	Sonoma	Pinot Noir	Blue Farm
	10	Italy	Elegance, complexity	Ronco della Chiesa	95	80	Northeastern Italy	Collio		Friulano	Borgo del Tiglio
	11	US	From 18-year-old vine	Estate Vineyard Wac	95	48	Oregon	Ribbon Ridge	Willamette Valley	Pinot Noir	Patricia Green Cellars
	12	US	A standout even in thi	Weber Vineyard	95	48	Oregon	Dundee Hills	Willamette Valley	Pinot Noir	Patricia Green Cellars
	13	France	This wine is in peak co	Chï¹teau Montus Pr	95	90	Southwest France	Madiran		Tannat	Vignobles Brumont
	14	US	With its sophisticated	Grace Vineyard	95	185	Oregon	Dundee Hills	Willamette Valley	Pinot Noir	Domaine Serene
	15	US	First made in 2006, th	Sigrid	95	90	Oregon	Willamette Valley	Willamette Valley	Chardonnay	Bergstrïm
	16	US	This blockbuster, pow	Rainin Vineyard	95	325	California	Diamond Mountain D	Napa	Cabernet Sauvignon	Hall
	17	Spain	Nicely oaked blackber	6 Avitos Reserva Pre	95	80	Northern Spain	Ribera del Duero		Tempranillo	Valduero
	18	France	Coming from a seven-	Le Pigeonnier	95	290	Southwest France	Cahors		Malbec	Chï¹teau Lagrï©zette
	19	US	This fresh and lively m	Gaja's Crown Vineyard	95	75	California	Sonoma Coast	Sonoma	Pinot Noir	Gary Farrell

@ftisiot
@BrendanTierney

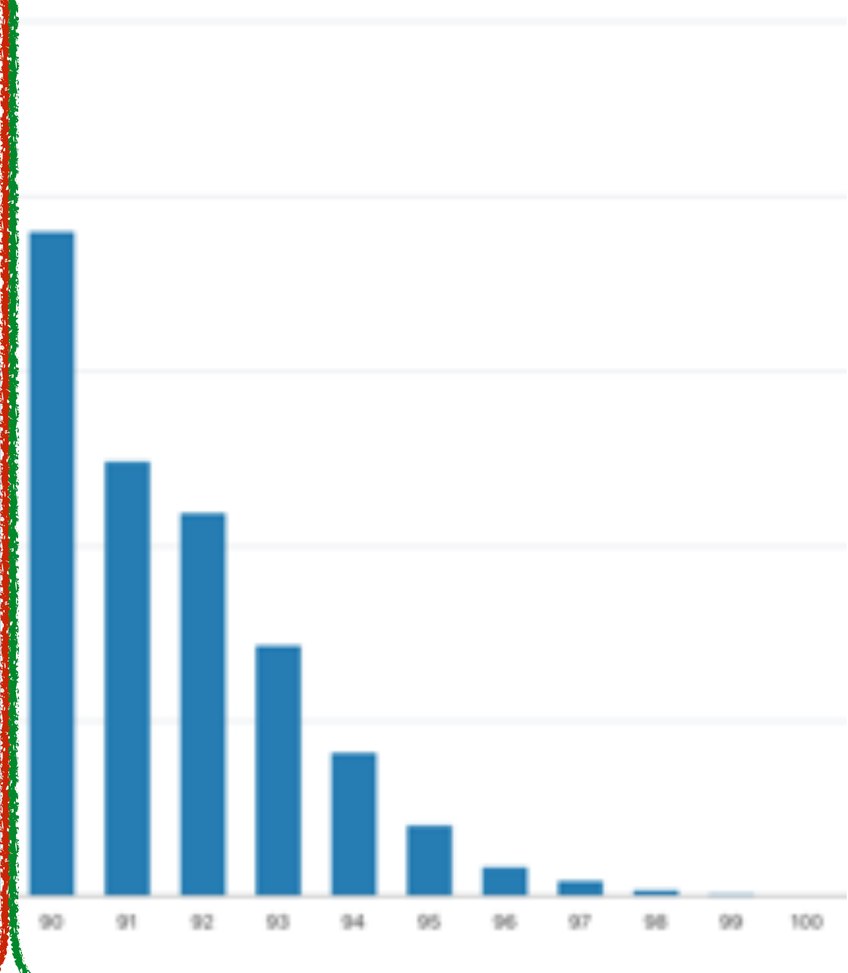
@ftisiot
@BrendanTierney

WINE SCORE

Bad



Good

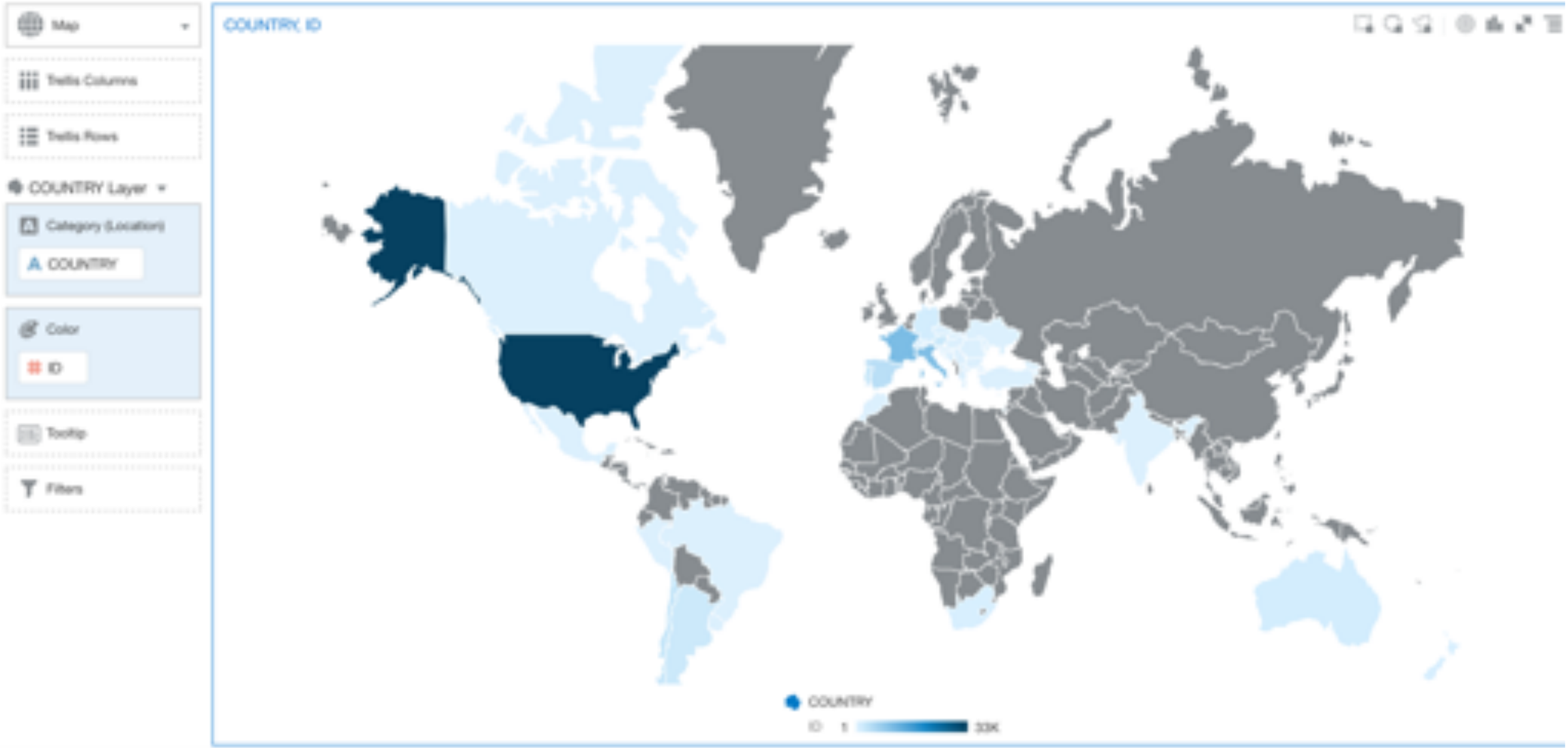


WINE DESCRIPTION

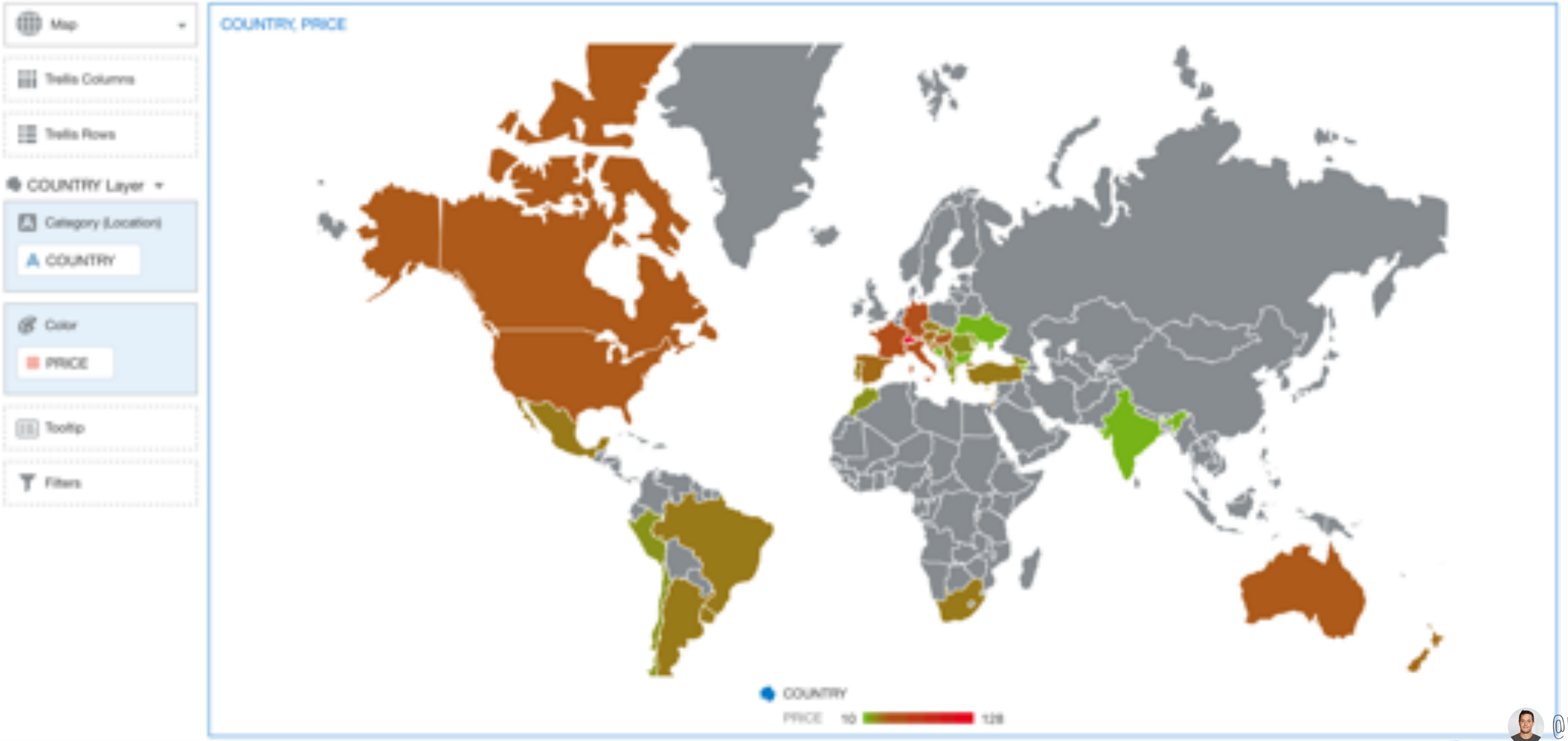
@Ftisiot
@BrendanTierney

WINE SAMPLES BY COUNTRY

@Ftisiot
@BrendanTierney



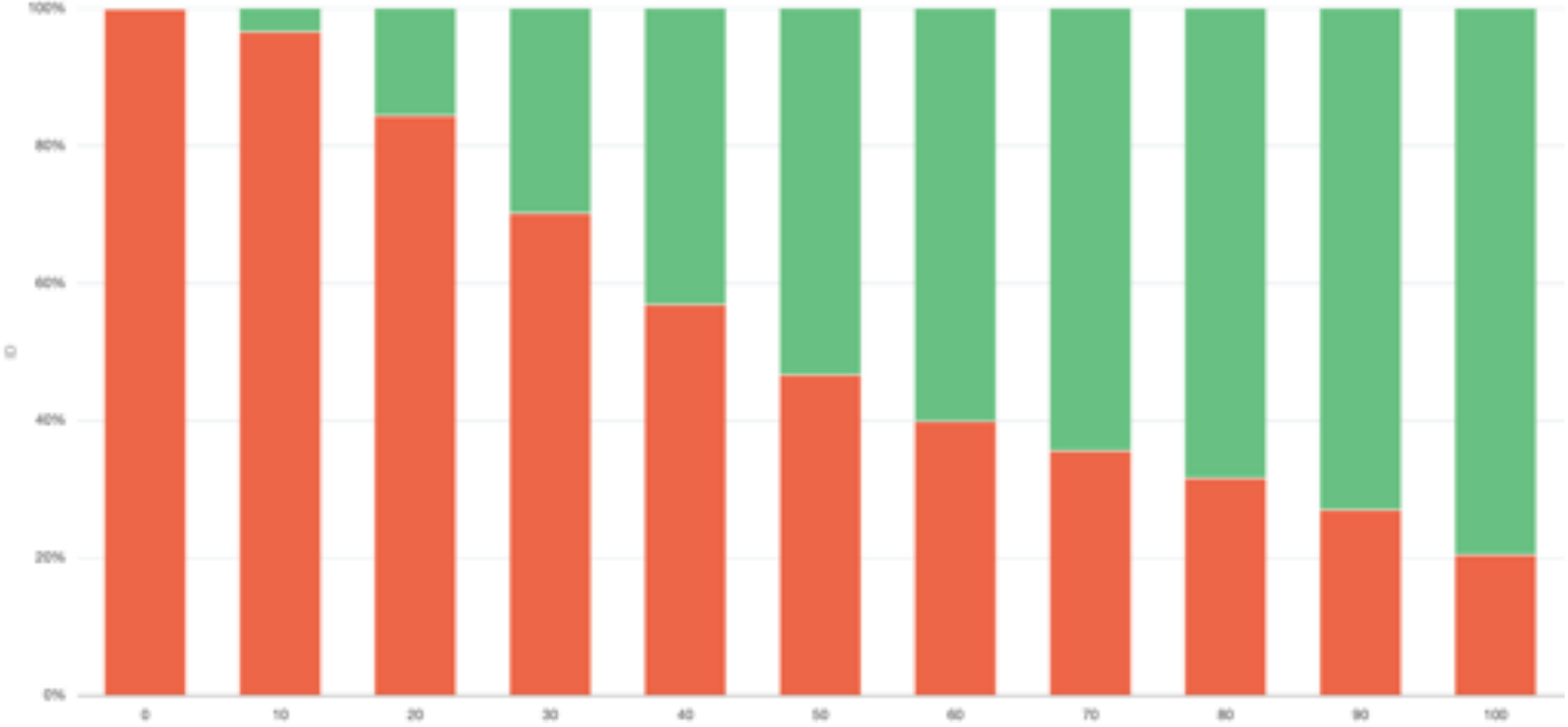
WINE PRICE BY COUNTRY



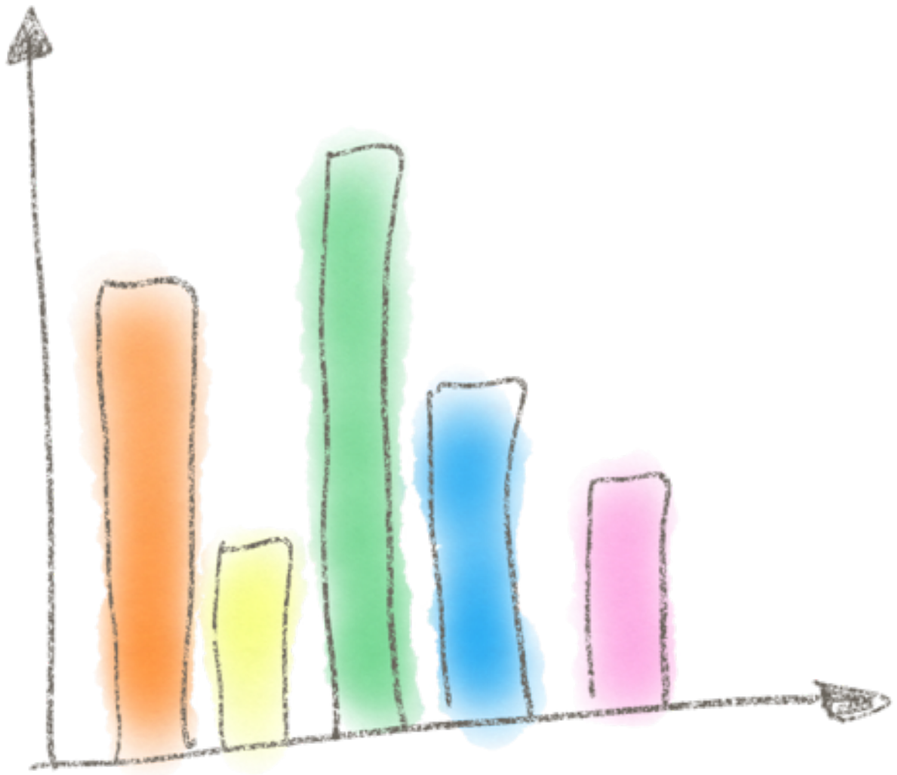
@FTISIOT
@BRENDANTIERNEY

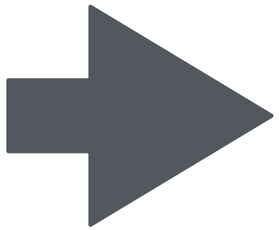
@ftisiot
@BrendanTierney

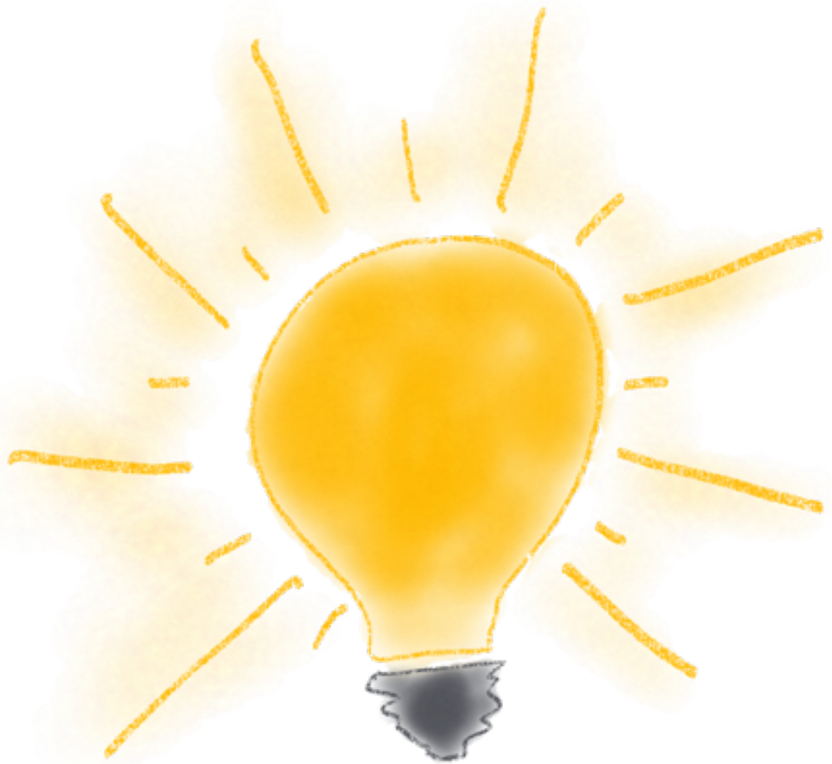
GOOD / BAD WINE BY PRICE



From insights to PREDICTIONS



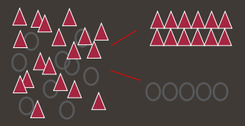


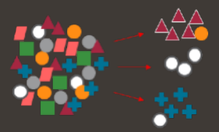


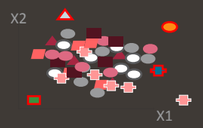

@Ftisiot

@BrendanTierney

Oracle Machine Learning Algorithms

Classification

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

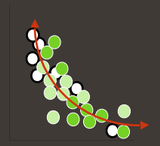
Clustering

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

Anomaly Detection


One-Class SVM

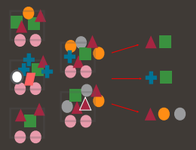
Time Series

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 & AI

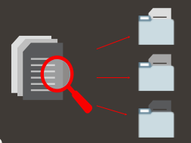
Association Rules

A priori/ market basket

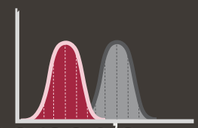
Predictive Queries
Predict, cluster, detect, features

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.

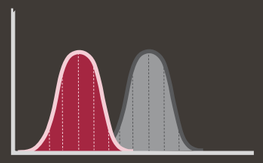
R Packages
Third-party R Packages through Embedded Execution
Spark MLlib algorithm integration

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

•Includes support for Partitioned Models, Transactional, Unstructured, Geo-spatial, Graph data. etc,

Statistical Functions and Analytical SQL

Statistical Functions



Descriptive statistics
(e.g. median, stdev, mode, sum, etc.)

Hypothesis testing
(t-test, F-test, Kolmogorov-Smirnov test, Mann Whitney test, Wilcoxon Signed Ranks test

Correlations analysis
(parametric and nonparametric e.g. Pearson's test for correlation, Spearman's rho coefficient, Kendall's tau-b correlation coefficient)

Ranking functions

Cross Tabulations with Chi-square statistics

Linear regression

ANOVA (Analysis of variance)

Test Distribution fit
(e.g., Normal distribution test, Binomial test, Weibull test, Uniform test, Exponential test, Poisson test)

Statistical Aggregates
(min, max, mean, median, stdev, mode, quantiles, plus x sigma, minus x sigma, top n outliers, bottom n outliers)

Analytical SQL

SQL Windows

SQL Aggregate functions

LAG/LEAD functions

SQL for Pattern Matching

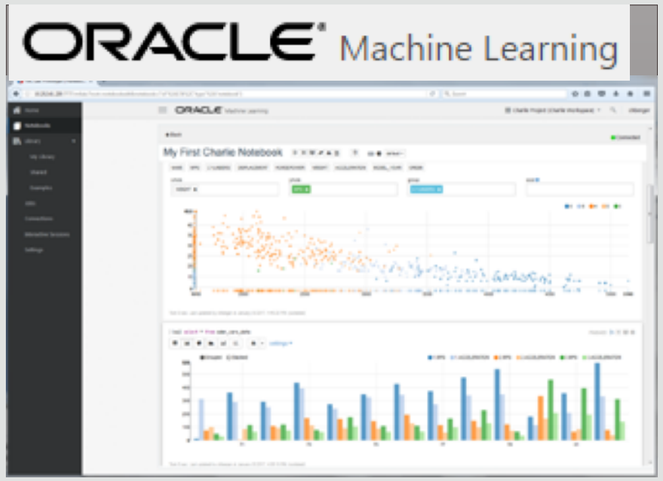
Additional approximate query processing: APPROX_COUNT, APPROX_SUM, APPROX_RANK

Regular Expressions

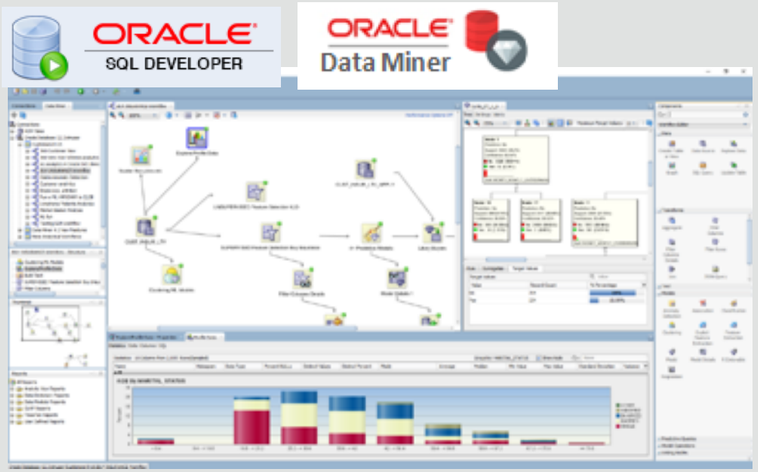
Oracle Machine Learning

Multiple Languages UIs Supported for End Users & Apps Development

R & Python Data Scientists

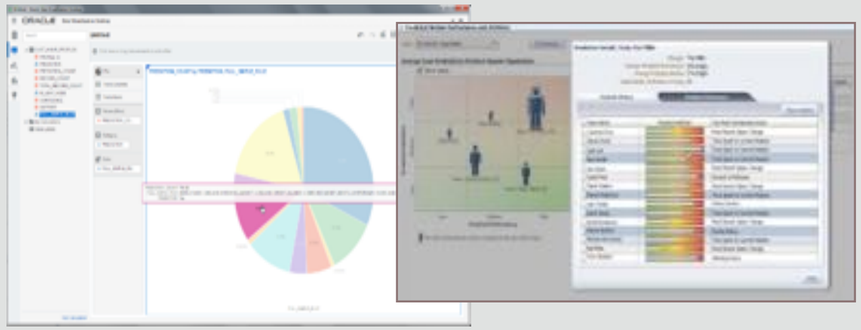


Notebook Users & DS Teams



“Citizen” Data Scientists

DBAs



Application Developers

CREATE ML MODEL

SQL>_

```
BEGIN
      DBMS_DATA_MINING.CREATE_MODEL(
        model_name           => 'Wine_CLASS_MODEL',
        mining_function      => dbms_data_mining.classification,
        data_table_name      => ''Wine_TRAIN_DATA',
        case_id_column_name  => 'ID',
        target_column_name   => 'POINTS_BIN',
        settings_table_name => 'Wine_build_settings');
   END;
   /
```


@FTISIOT


@BRENDANTIERNEY

APPLY ML MODEL

SQL>_

```
SELECT PREDICTION_PROBABILITY(
          Wine_CLASS_MODEL,
          'GT_90_POINTS'
          USING
            25 as PRICE,
            ‘MALBEC' as VARIETY,
            ‘SPAIN' as COUNTRY
            )
 FROM dual;
```


@FTISIOT


@BRENDANTIERNEY

Picking a Good Wine for <$20 with ADW+OML

default

FINISHED

Selecting a wine to bring to a party that is likley to be good and is also inexpensive!

We've all been there. We are invited to a party and want to bring something to contribute. A bottle of wine is always a good option. But which one? There are so many! And we don't want to break the bank for our party gift. Let's try to use the Autonomous Database and Oracle Machine Learning to find a likely tasty and likley to be rated greater than 90 Points (GT_90_Points) adult beverage that is also affordable (Under $35).

Original data comes from Kaggle WineReviews130K data = 130k wine reviews with variety, location, winery, price, and description etc. Wine Reviews and Twitter Handle were removed. Points was binned to GT_90 and LT_90 https://www.kaggle.com/zynicide/wine-reviews

Steps

1. Explore the data
2. Define the Target Attribute (GT_90_Points vs. LT_90_Points)
3. Find the Key Attributes that Most Influence GT_90_Points_Bin)
4. Build a Machine Learning Model to Predict Good Wines to Buy
5. Save our results and further investigage our wine findings using the Oracle Analytics Cloud

Took 2 sec. Last updated by BRENDAN at August 22 2020, 12:01:38 PM.

FINISHED

Took 0 sec. Last updated by BRENDAN at August 22 2020, 12:01:41 PM.

Let's Explore the WineReviews130K data

FINISHED

```
%sql

-- Explore the WineReviews130K data
```

Wine Reviews Data

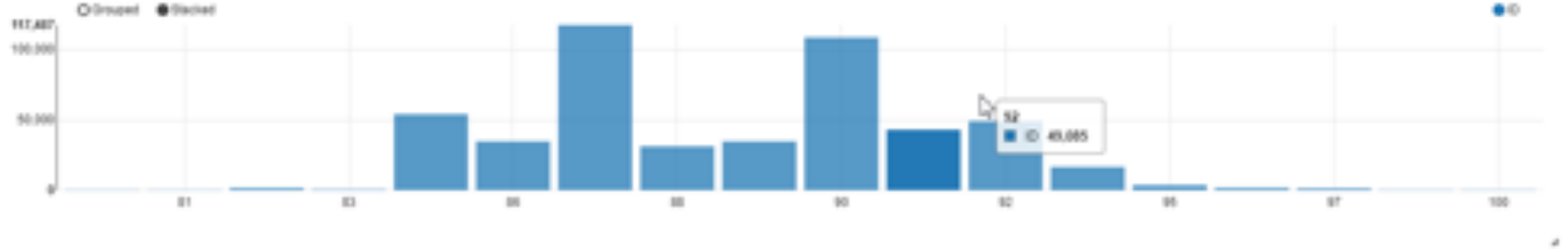
default

Wine Points Ratings Distribution

FINISHED

```
%sql
select * from winereviews100k;
```

settings



Took 0 sec. Last updated by CHARLIE at April 10 2018, 2:11:50 PM. (outdated)

Display Count of Wines by Country in Descending Order (Top 10)

FINISHED

```
%sql
select count(ID) as TOTAL, COUNTRY from winereviews100k group by COUNTRY order by count(ID) desc fetch first 10 rows only;
```

settings

Took 0 sec. Last updated by CHARLIE at April 04 2018, 12:17:41 PM. (outdated)

Connected

Picking a Good Wine for <$20 with ADW+…

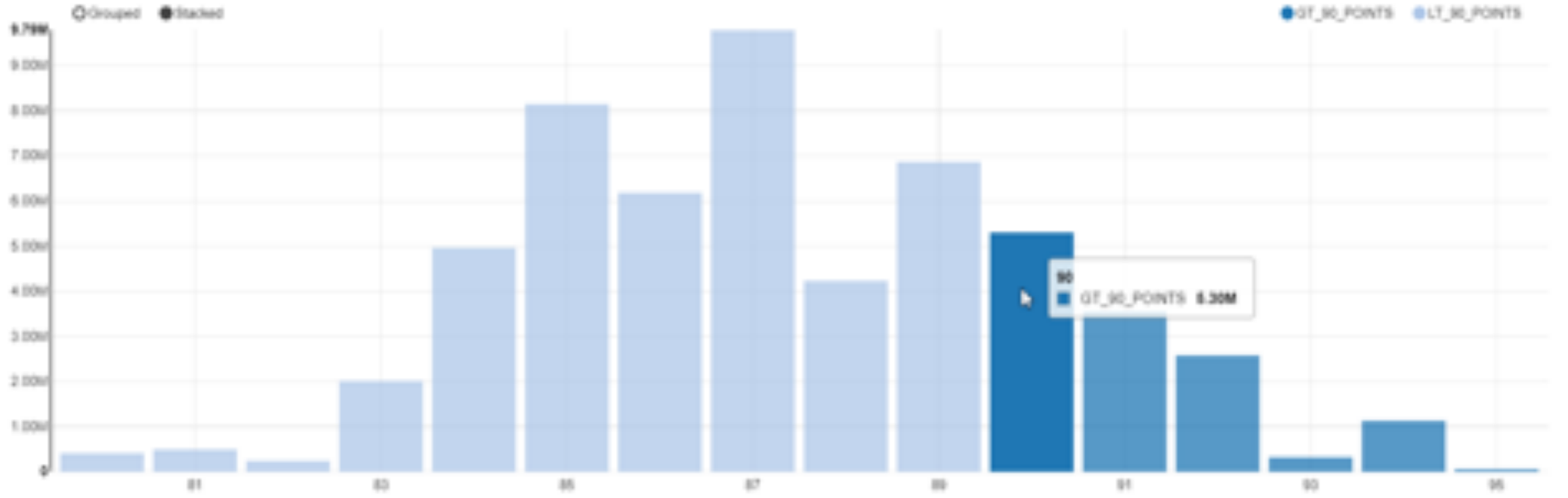
default

Wine Points Ratings Distribution; We Can Explore our Data Graphically using Simple Graphs

FINISHED

```
%adj
select * from WineReviews130K;
```

settings



Took 0 sec. Last updated by CHRISJE at July 19 2019, 6:07:14 PM.

Wine Reviews Data

default

Compare Wines from Different Countries

FINISHED

```
%sql
select * from wineReviews130K where COUNTRY = ${COUNTRY='Germany',
    'Germany'|'Spain'|'Chile'|'US'|'Argentina'|'Italy'|'Australia'};
```

COUNTRY

'US'

settings

California Virginia Oregon Washington New York
Michigan Idaho Texas Pennsylvania America

Took 4 sec. Last updated by CHARLIE at May 26 2018, 2:26:59 PM.

Compare Wines from Different Countries

FINISHED

```
%sql
select * from wineReviews130K where COUNTRY = ${COUNTRY='Germany',
    'Germany'|'Spain'|'Chile'|'US'|'Argentina'|'Italy'|'Australia'};
```

COUNTRY

'Italy'

settings

Table

Sicily & Sardinia Tuscany Piedmont
Northeastern Italy Central Italy Southern Italy
Veneto Italy Other Lombardy
Northwestern Italy

Took 10 sec. Last updated by CHARLIE at April 17 2018, 2:26:32 PM. (outdated)

Compare Wines from Different Countries

FINISHED

```
%sql
select * from wineReviews130K where COUNTRY = ${COUNTRY='Germany',
    'Germany'|'Spain'|'Chile'|'US'|'Argentina'|'Italy'|'Australia'};
```

COUNTRY

'Germany'

settings

Mosel Rheinhessen Rheingau Franken
Ahr Nahe Pfalz Baden
Württemberg Mosel-Saar-Ruwer Germany Mittelrhein

Took 0 sec. Last updated by CHARLIE at April 10 2018, 1:43:02 PM.

Create New Table (or View) WineReviews130KTarget from WineReviews130K but without POINTS Attribute

FINISHED

```
%script

/*  Clean out old table  */
BEGIN
    EXECUTE IMMEDIATE 'DROP Table wineReviews130KTarget';
  EXCEPTION
    WHEN OTHERS THEN NULL;
END;
/

/*  Create wineReviews130KTarget table  */
```

ORACLE Machine Learning

Charlie Project [Charlie Workspace]   CHARLIE

Connected

Wine Reviews Data

default

Build Attribute Importance Model

FINISHED

```
%script

BEGIN
   DBMS_DATA_MINING.CREATE_MODEL(
     model_name            => 'GOOD_WINE__AI',
     mining_function       => DBMS_DATA_MINING.ATTRIBUTE_IMPORTANCE,
     data_table_name       => 'WineReviews130KTarget',
     case_id_column_name   => 'ID',
     target_column_name    => 'POINTS_BIN',
     settings_table_name   => 'AI_Import_Mode_Settings');
END;
/
```

```
PL/SQL procedure successfully completed.
---------------------------
```

Took 3 sec. Last updated by CHARLIE at April 10 2018, 2:15:17 PM.

Show Top Attributes that Most Influence Good (GT_90_POINTS) Wines

FINISHED

```
%sql
SELECT *
FROM TABLE(DBMS_DATA_MINING.GET_MODEL_DETAILS_AI('GOOD_WINE__AI'))
ORDER BY RANK;
```

settings

Wine Reviews Data

default

Show Lift Chart for Model Evaluation

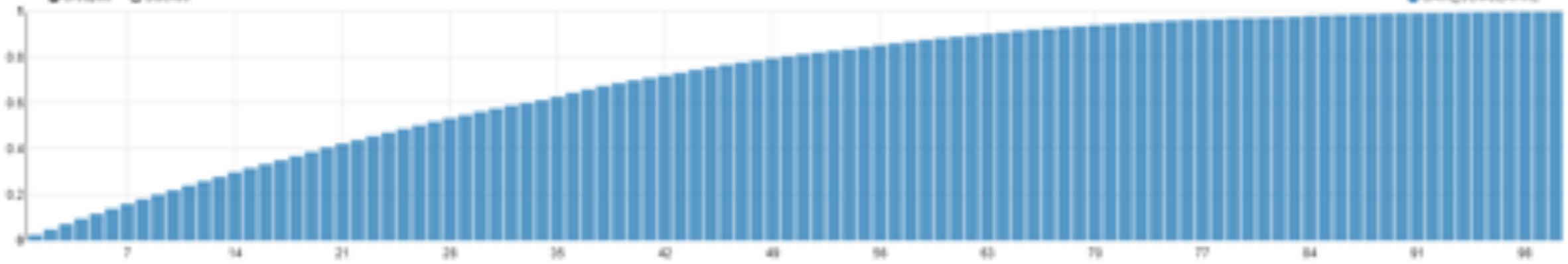
FINISHED

```
%sql
SELECT QUANTILE_NUMBER, GAIN_CUMULATIVE from WINE_LIFT_TABLE;
```

settings

Grouped Stacked

GAIN_CUMULATIVE



Took 0 sec. Last updated by CHARLIE at April 02 2018, 1:50:19 PM. (outdated)

Let's Explore our Wines

FINISHED

```
%sql
SELECT * FROM WINE_APPLY_RESULT WHERE PREDICTION = '${PREDICTION=LT_90_POINTS,GT_90_POINTS|LT_90_POINTS}';
```

PREDICTION

LT_90_POINTS

ID	PREDICTION	PROBABILITY	COST
7318	LT_90_POINTS	0.9961268289485274	0.0038731710514726

Wine Reviews Data

default

Show the Most Likely to be Best Wines (Highest Probability of "GT_90_ POINTS"

FINISHED

```
%md
SELECT ID, Price, Country, Designation, Province, Variety, prediction, probability
  FROM (SELECT ID, Price, Country, Designation, Province, Variety,
               PREDICTION_SET(SVM_CLASS_MODEL USING *) pset
          FROM WineReviews150KTarget
         WHERE ID > 10000 and Price < 15) T,
       TABLE(pset) S
ORDER BY prediction, probability desc, ID;
```

settings

Took 0 sec. Last updated by CHARLIE at May 03 2019, 3:01:27 PM. (outdated)

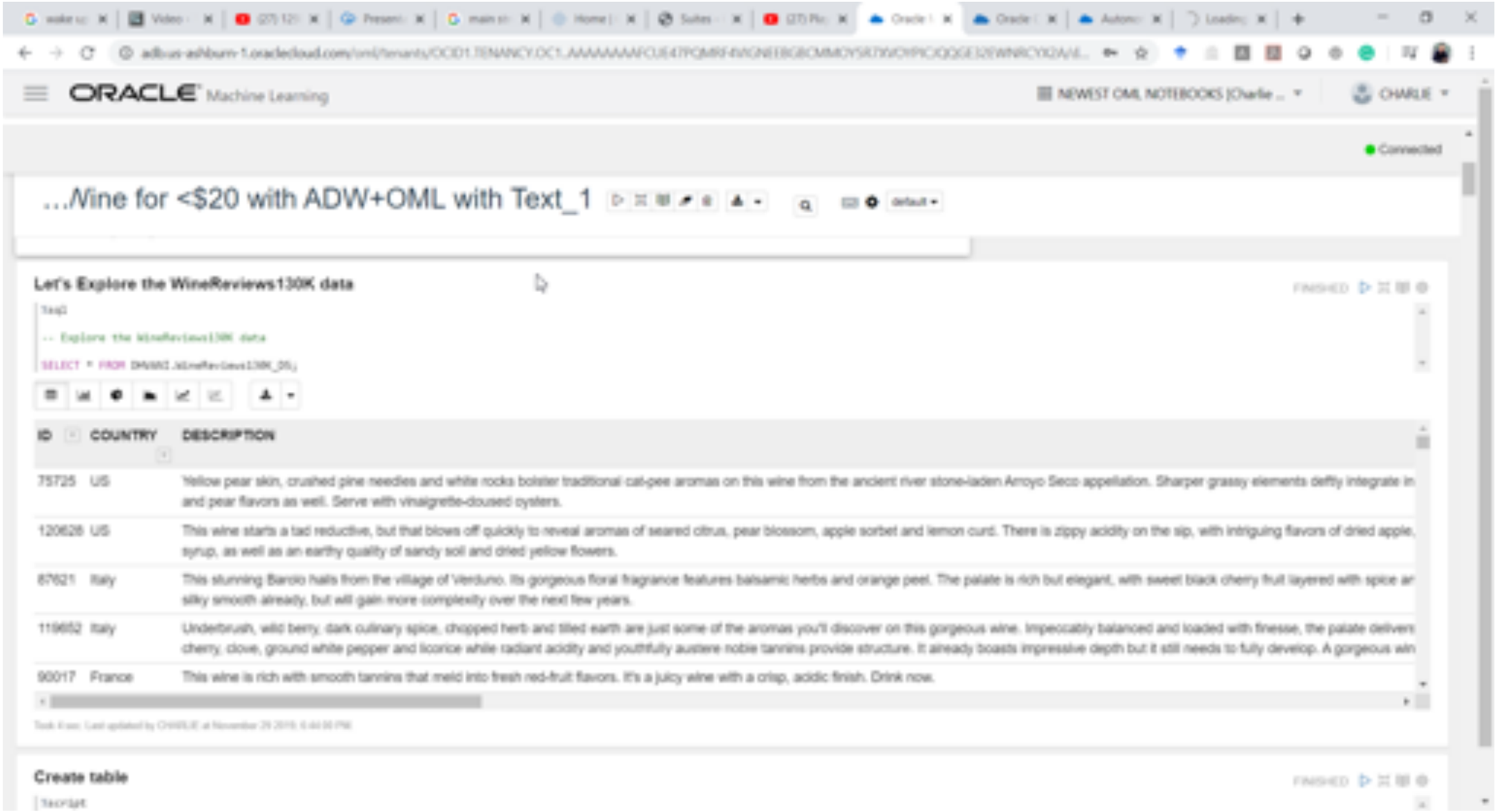
FINISHED

```
%md
### See how easy machine learning is inside the Oracle Autonomous Database?  Celebrate with a fine glass of wine!
### Let's now look further using Oracle Analytics Cloud!
```

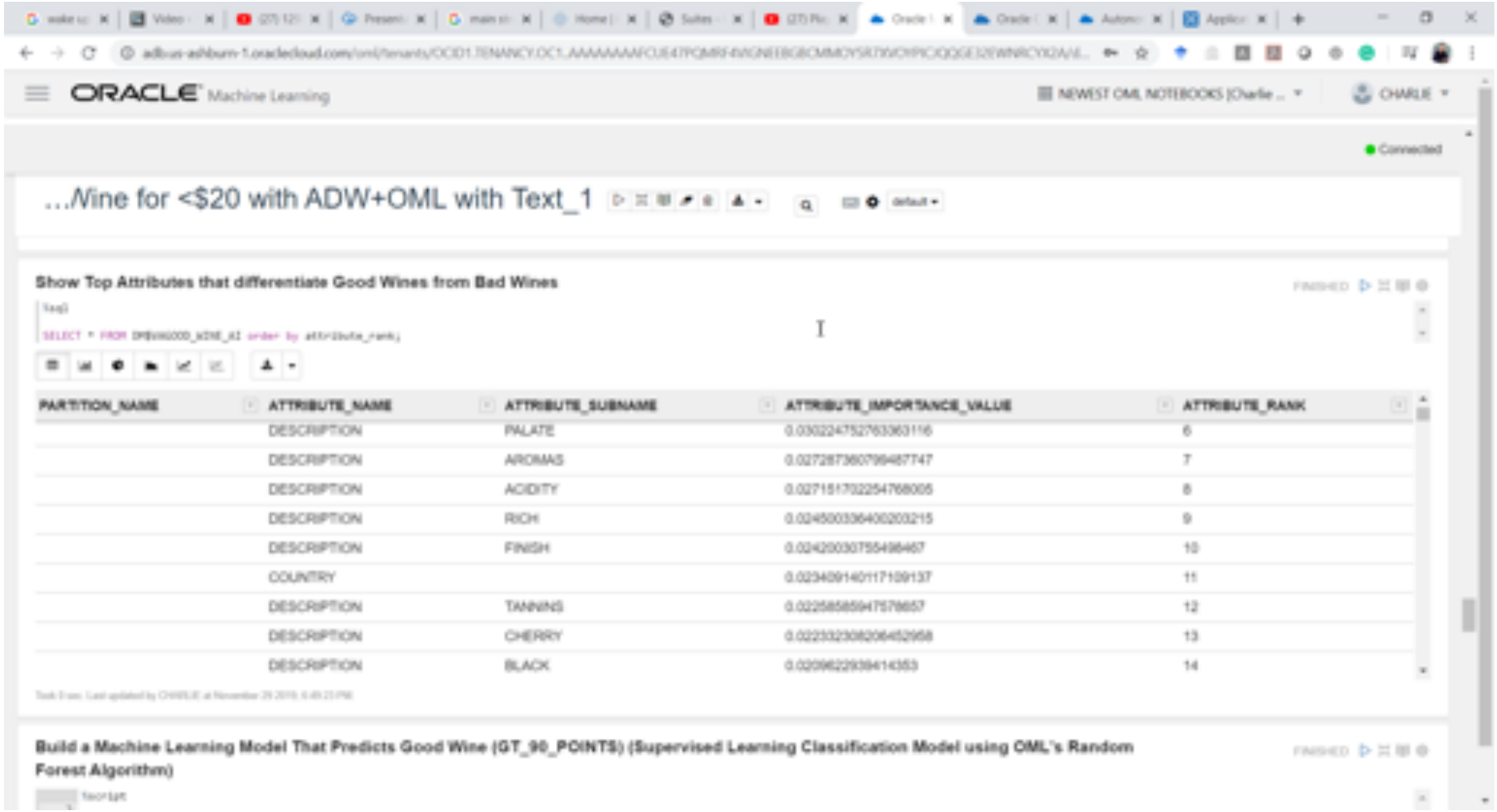
See how easy machine learning is inside the Oracle Autonomous Database? Celebrate with a fine glass of wine!

Let's now look further using Oracle Analytics Cloud!

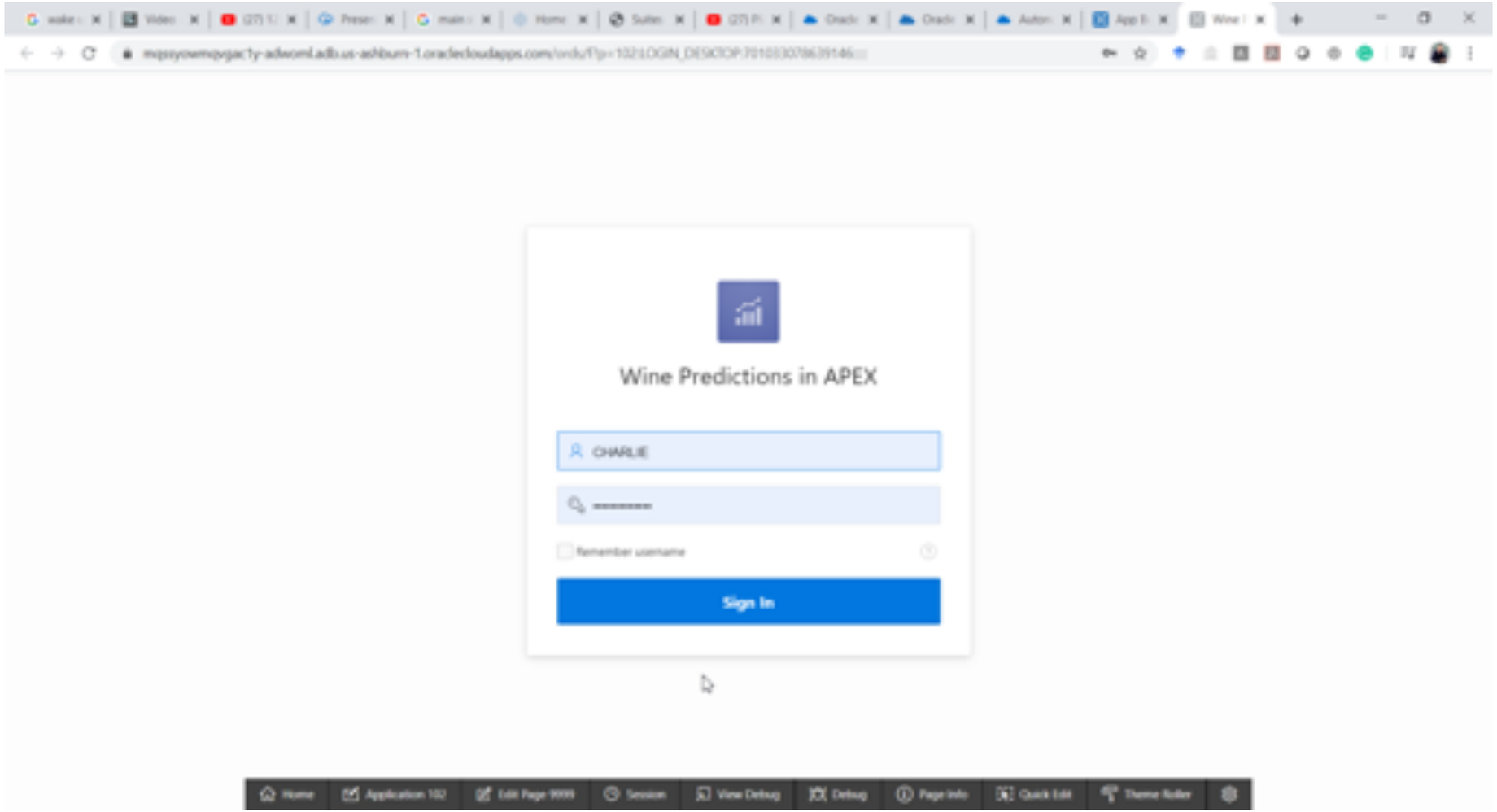
Optional: Include Wine Reviews



Optional: Include Wine Reviews

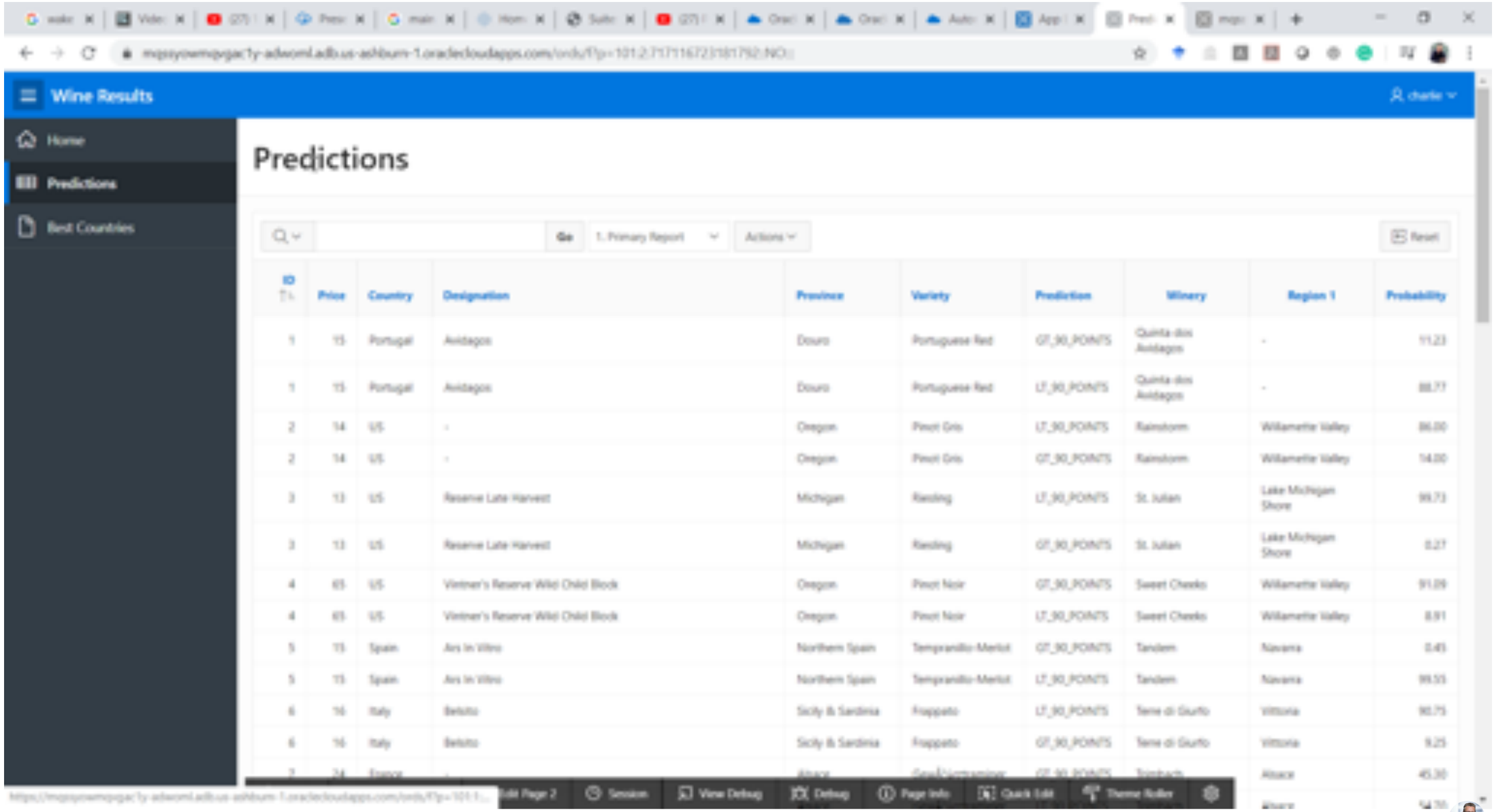


Oracle APEX



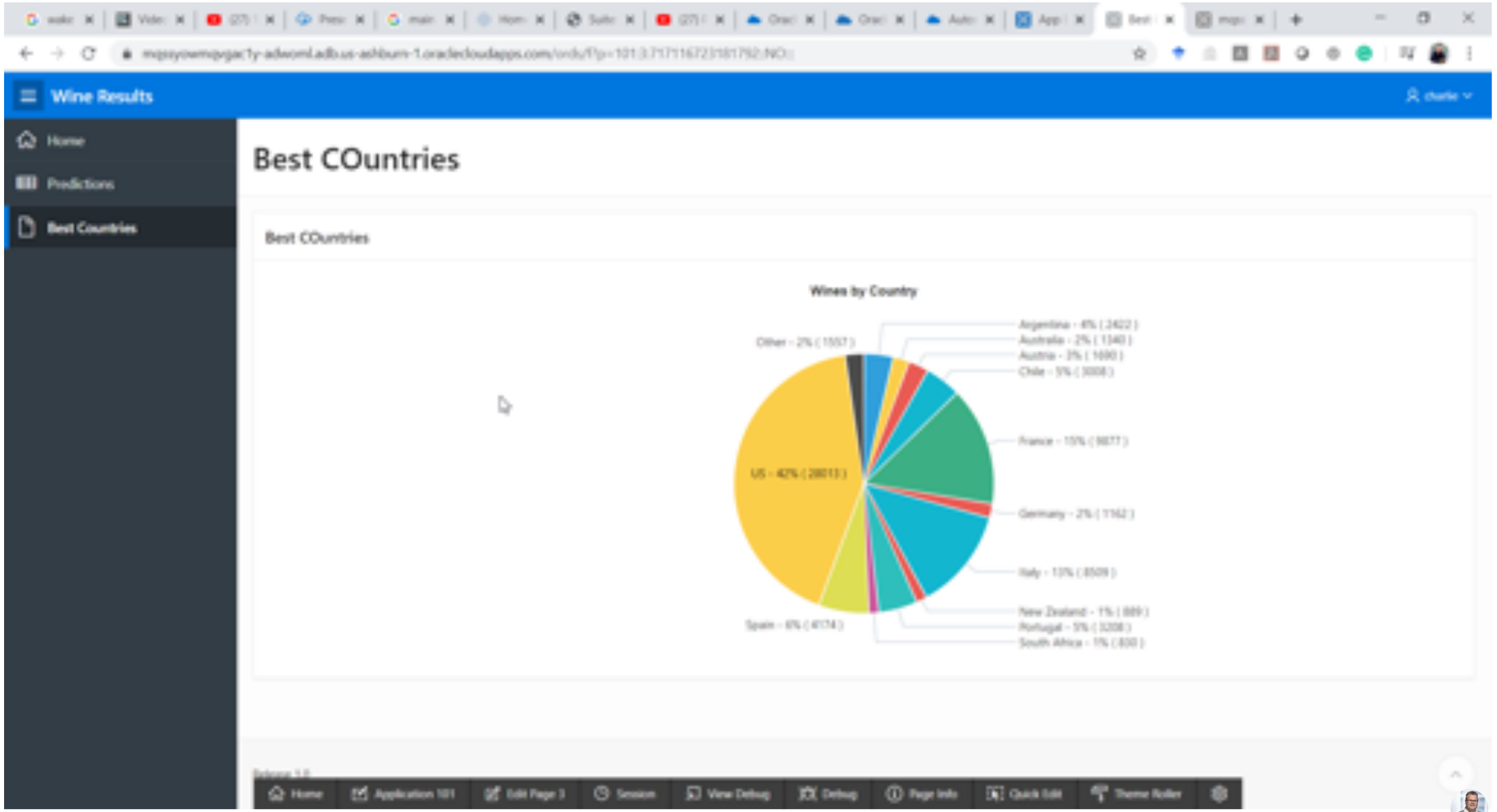
@Ftisiot
@BrendanTierney

Oracle APEX



@FTISIOT
@BRENDANTIERNEY

Oracle APEX

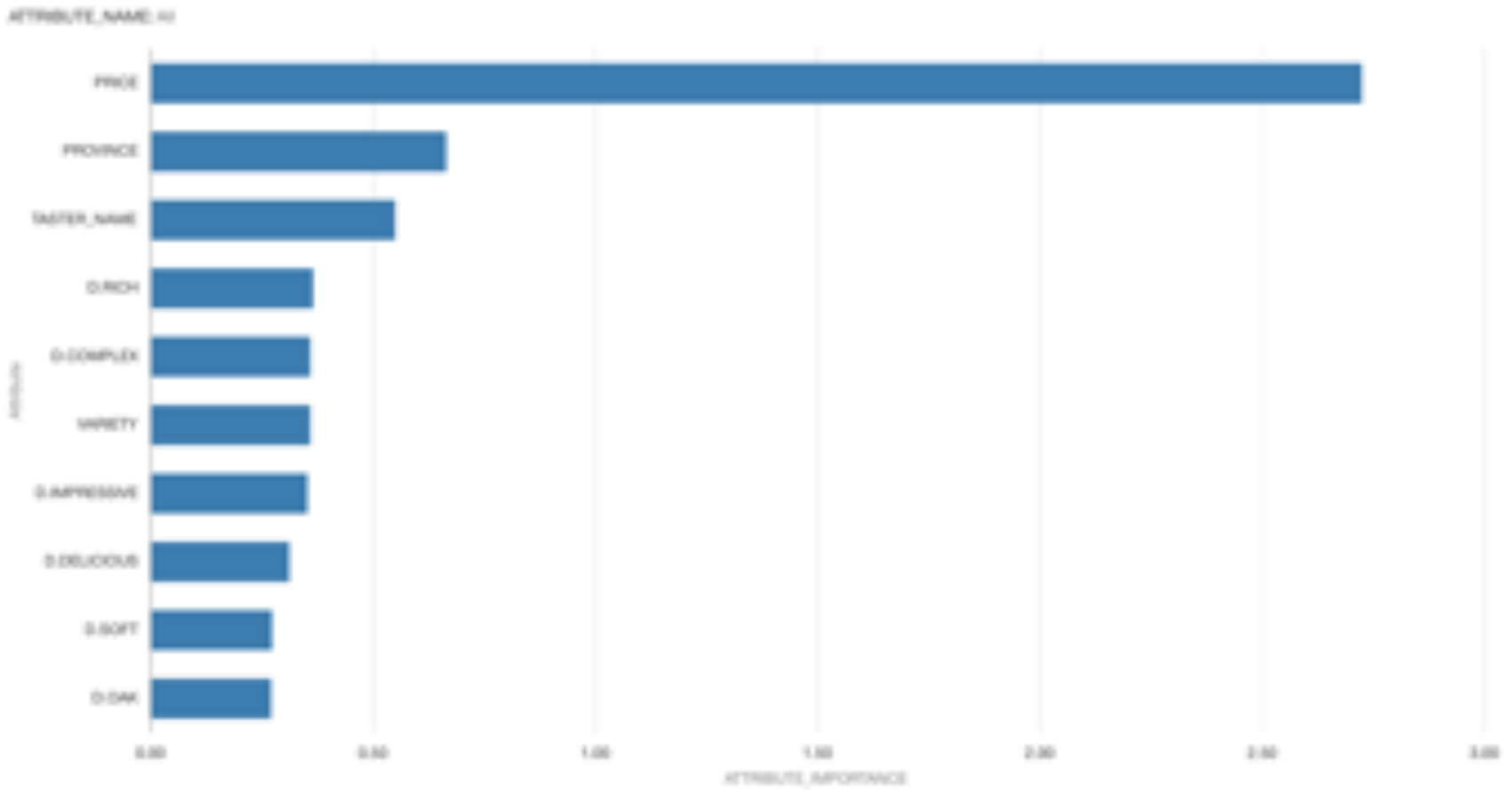


@Ftisiot
@BrendanTierney

Wine Attributes Importance

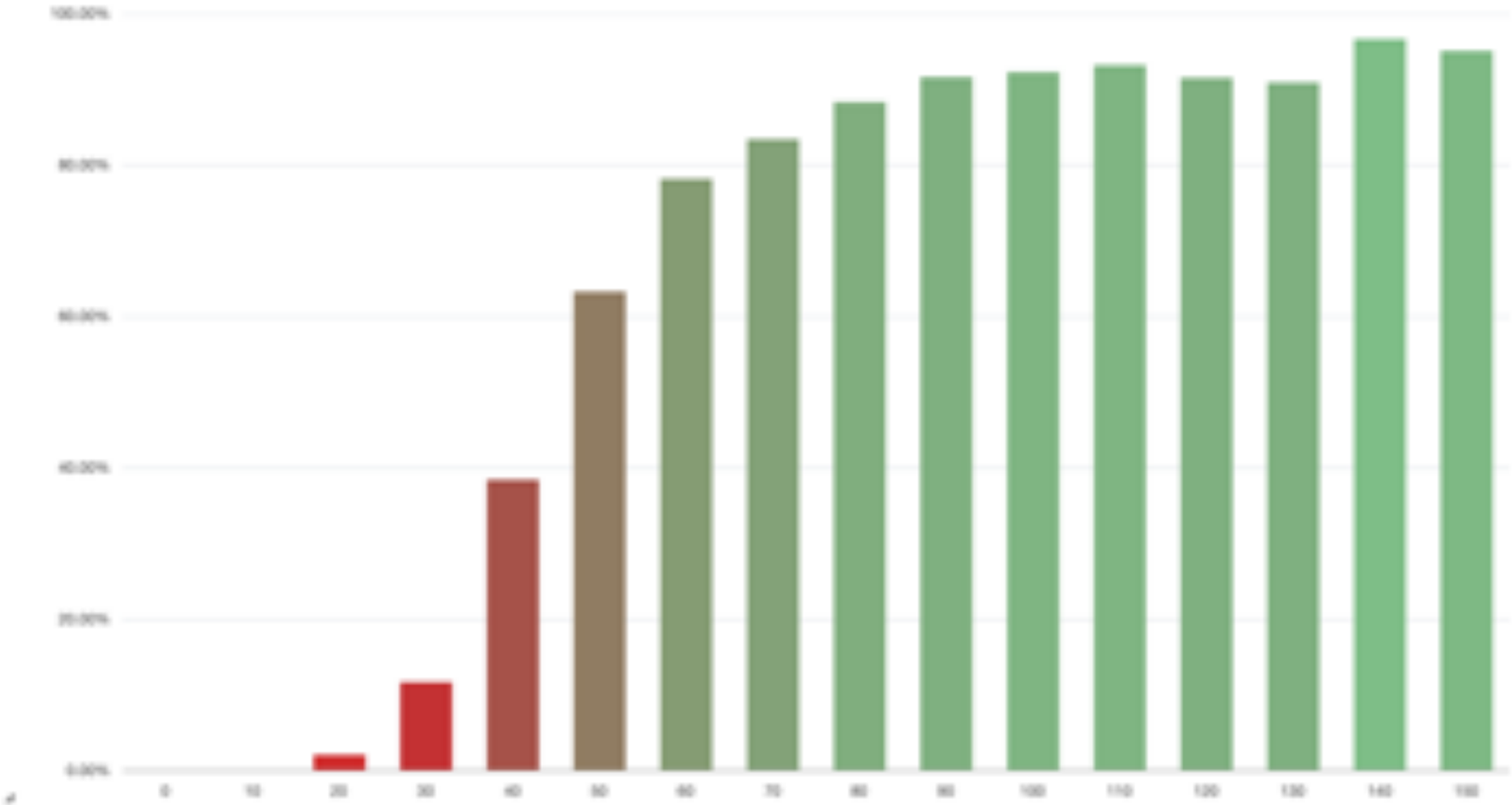
Top 10 ATTRIBUTE_IMPORTANCE ATTRIBUTE_NAME: All

What Makes a Good Wine?



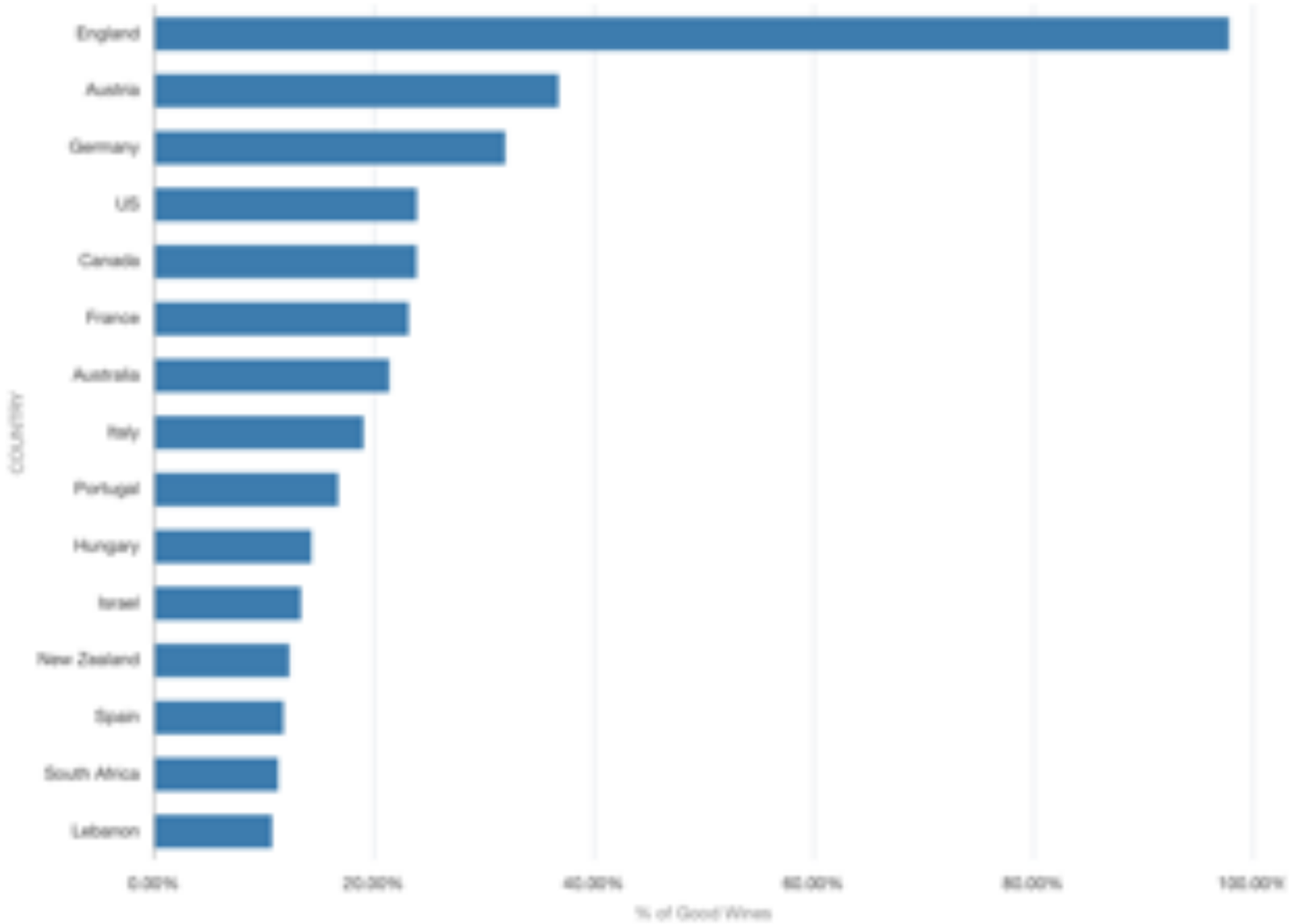
@Ftisiot
@BrendanTierney

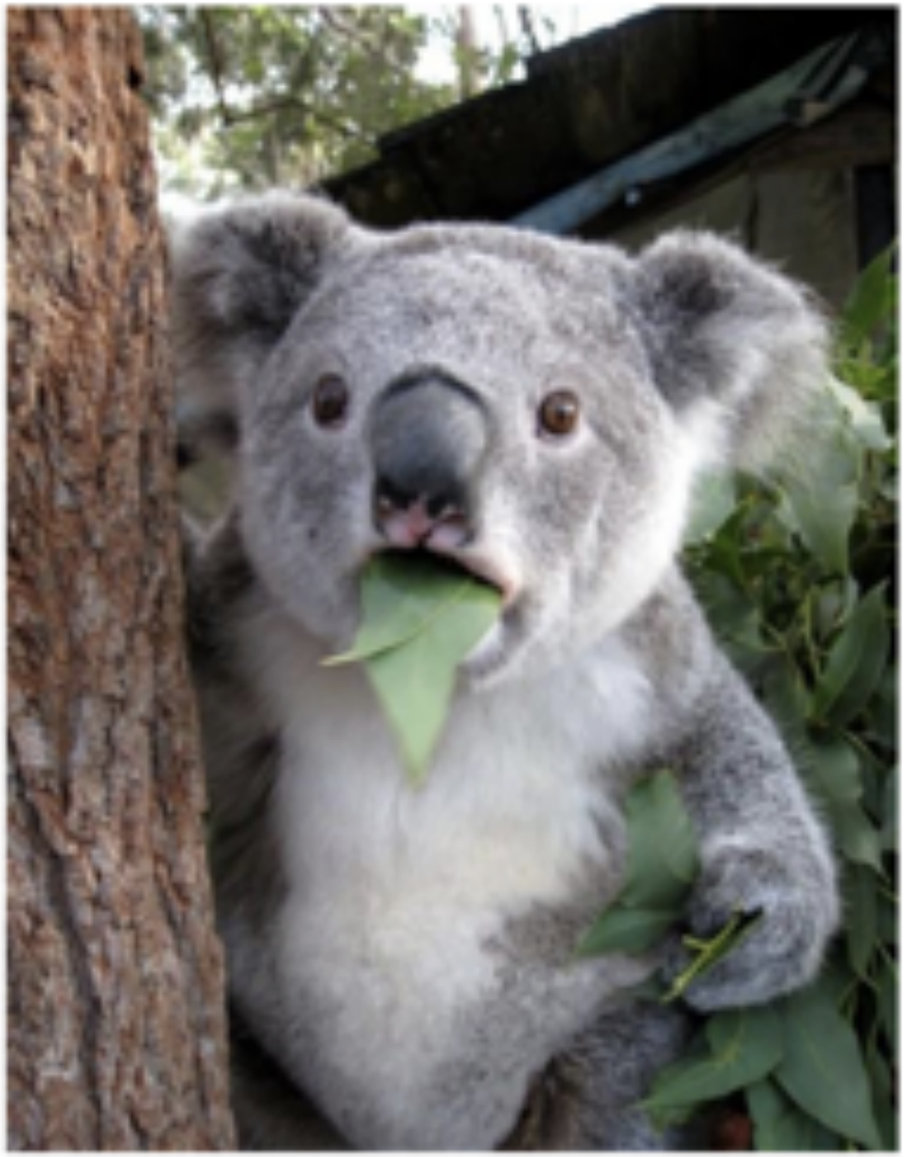
How Much Should I Spend for a Good Wine?



@Ftisiot
@BrendanTierney

Which Country Makes the Best Wines?



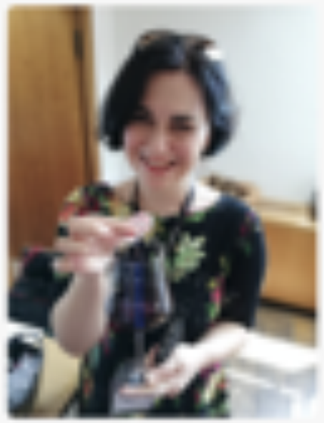


@Ftisiot
@BrendanTierney

Why England?????

@Ftisiot
@BrendanTierney

Who Rated in England?



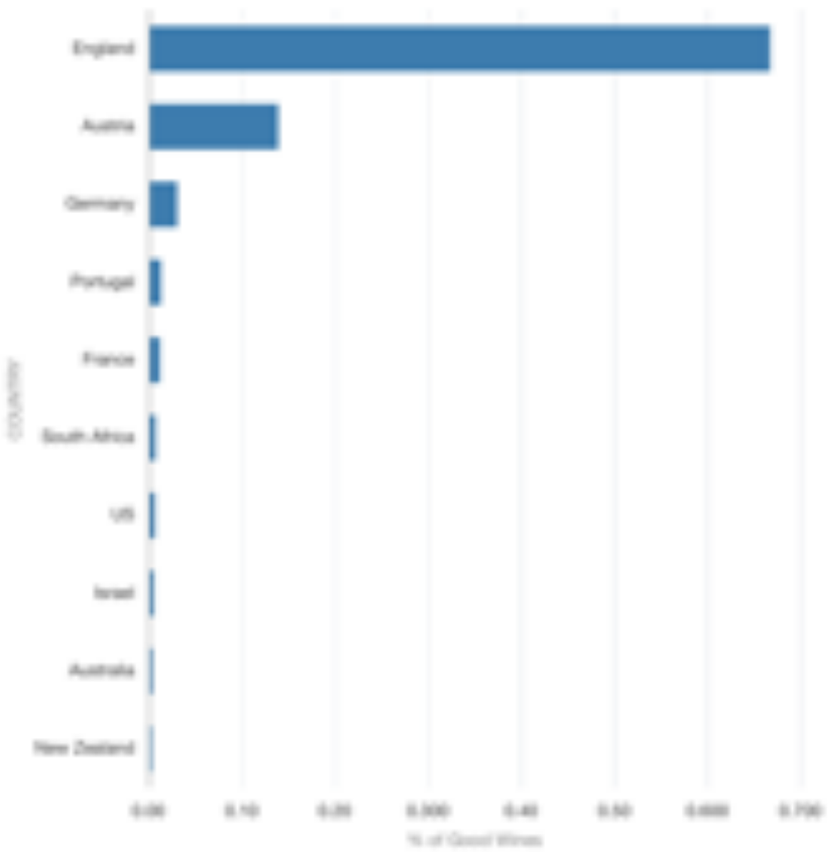
Anne Krebiehl MW

Date an MW: 2014
Based in: UK

What Should I Choose With Less Than 30$?

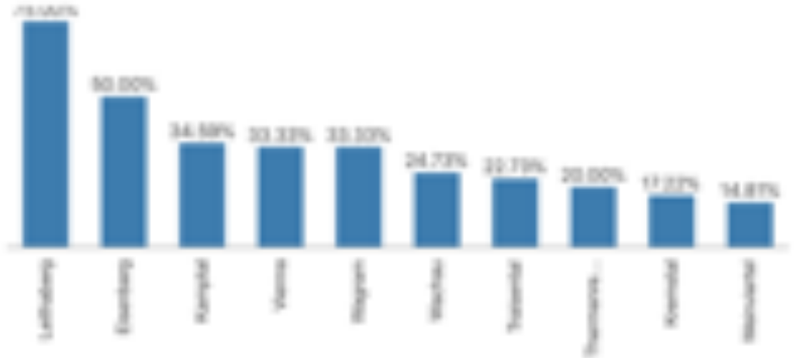
Top Countries

Top 10 % of Good Wines



Province

Top 10 % of Good Wines COUNTRY Austria



Variety

Top 10 % of Good Wines COUNTRY Austria

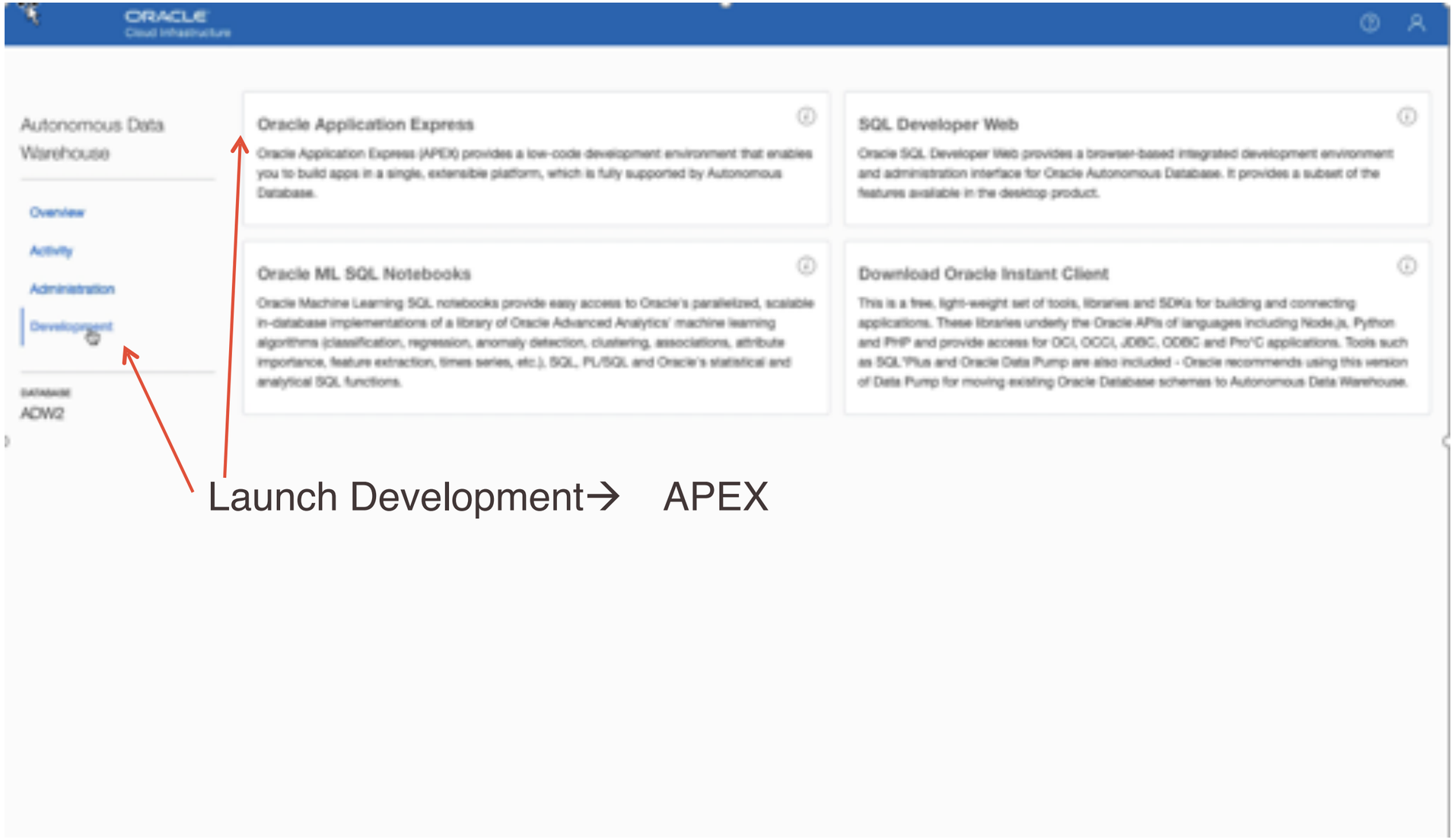
Good Wines

COUNTRY Austria

DESIGNATION	PRICE	PROBABILITY
1000-Eimerberg Smaragd	29	72%
13	13	98%
Alwagg	28	68%
Alte Point Smaragd	29	60%
Alte Reben	17	94%
Alte Reben	18	85%
Alte Reben	19	70%
Alte Reben	21	53%
Alte Reben	28	57%
Alte Reben Reserve	22	50%
Alte Reben Reserve	25	65%
Altenberg	15	88%
Altenberg	25	80%
Altes Kreuz	17	60%
Altweingarten	12	92%
Am Berg	19	74%
Am Berg	22	74%
Ambau	13	69%
Anning	19	60%
Anninger	15	92%
Anninger	19	57%
Apfelton Rouge	25	54%
Aristos	28	60%

@Ftisiot
@BrendanTierney

ML Model Deployment via ORDS REST API



@Ftisiot
@BrendanTierney

Autonomous Data Warehouse

Overview
Activity
Administration
Development

DATABASE
ADW2

Download Client Credentials (Wallet)

Connections to Autonomous Data Warehouse use a secure connection. Your existing tools and applications will need to use this wallet file to connect to your Autonomous Data Warehouse instance. If you are familiar with using an Oracle Database within your own data center, you may not have previously used these secure connections.

Set Resource Management Rules

Set resource management rules to allocate CPU/IO shares to consumer groups and to cancel SQL statements based on their runtime and amount of IO.

Set Administrator Password

Set or reset your database administrator user's (ADMIN) password and when locked unlock your administrator user account on Autonomous Data Warehouse.

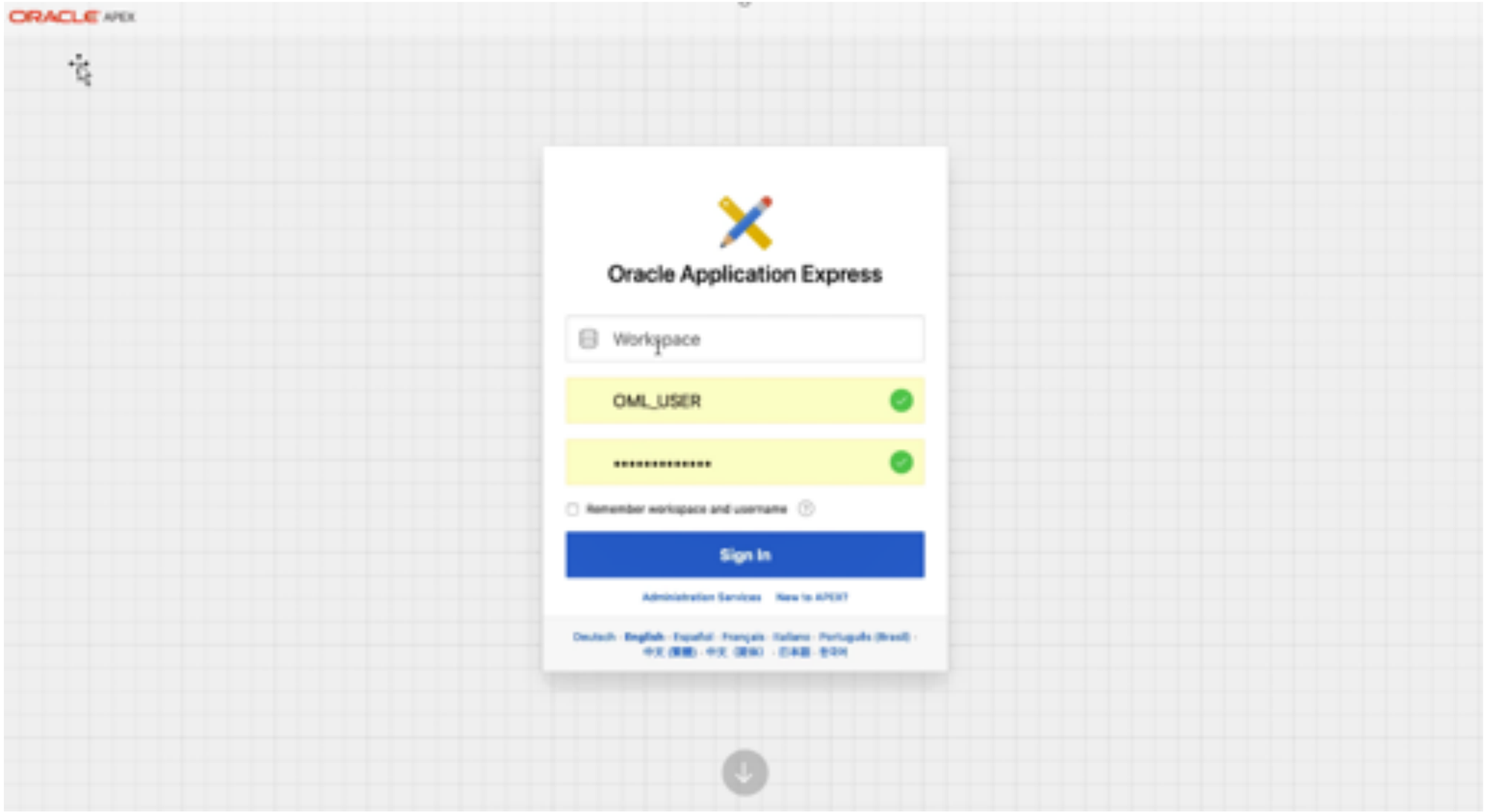
Manage Oracle ML Users

Create new Oracle Machine Learning user accounts and manage the credentials for existing Oracle Machine Learning users.

Send Feedback to Oracle

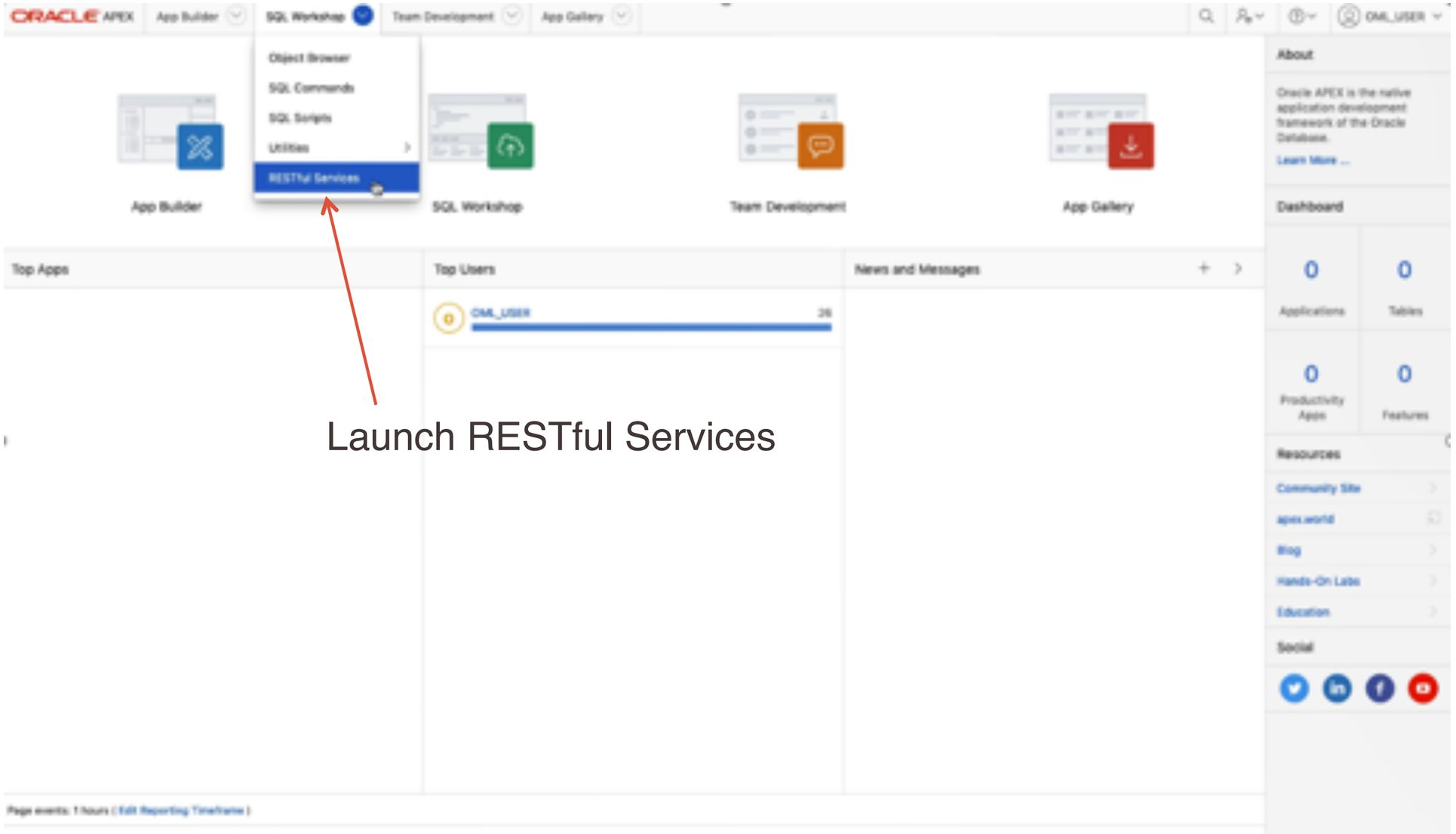
Use our CloudCustomerConnect forum to provide feedback about the service to Oracle, post questions, connect with experts, and share your thoughts and ideas. Click here to link to the forum.

ML Model Deployment via ORDS REST API



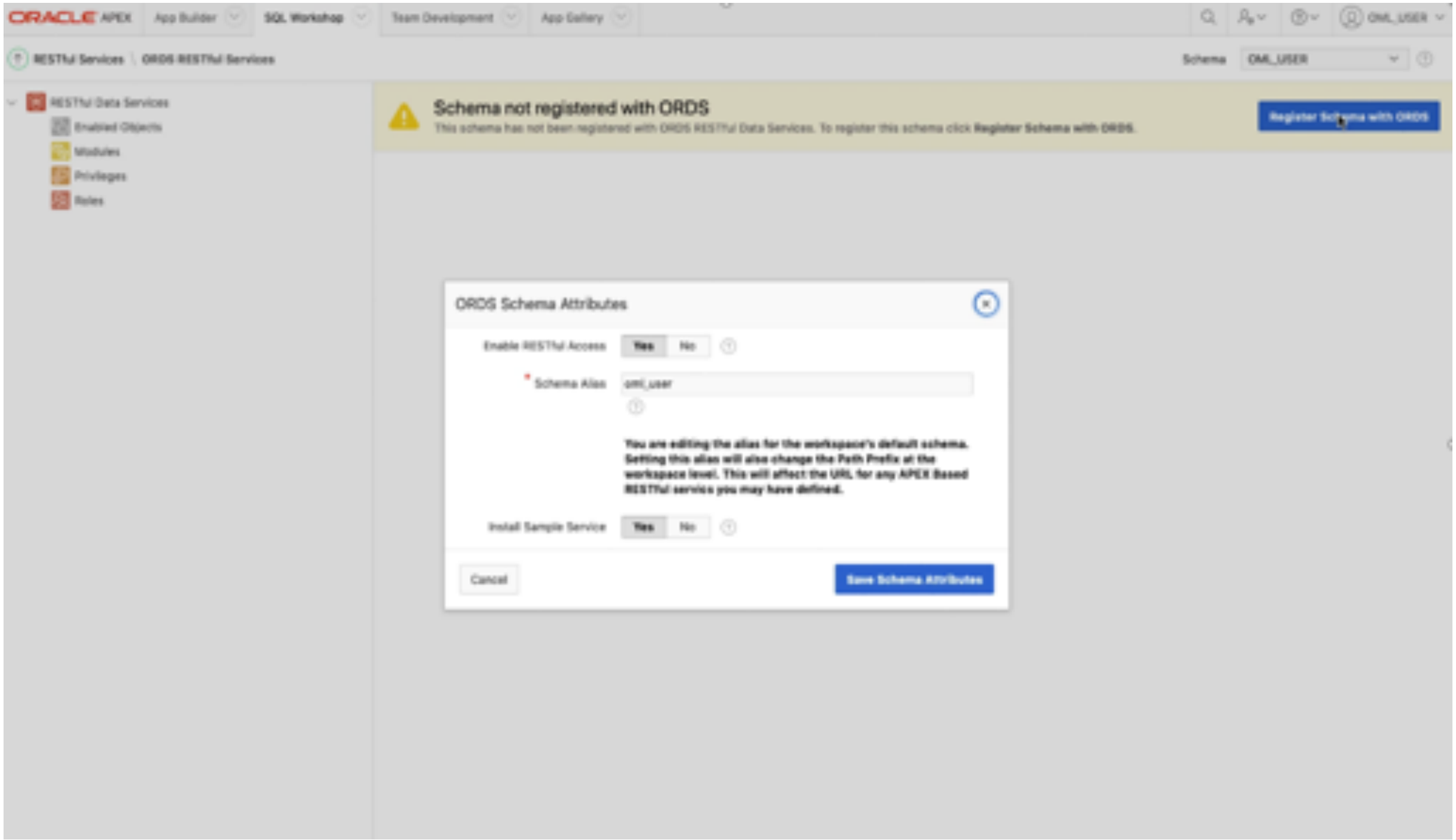
@ftisiot
@BrendanTierney

ML Model Deployment via ORDS REST API



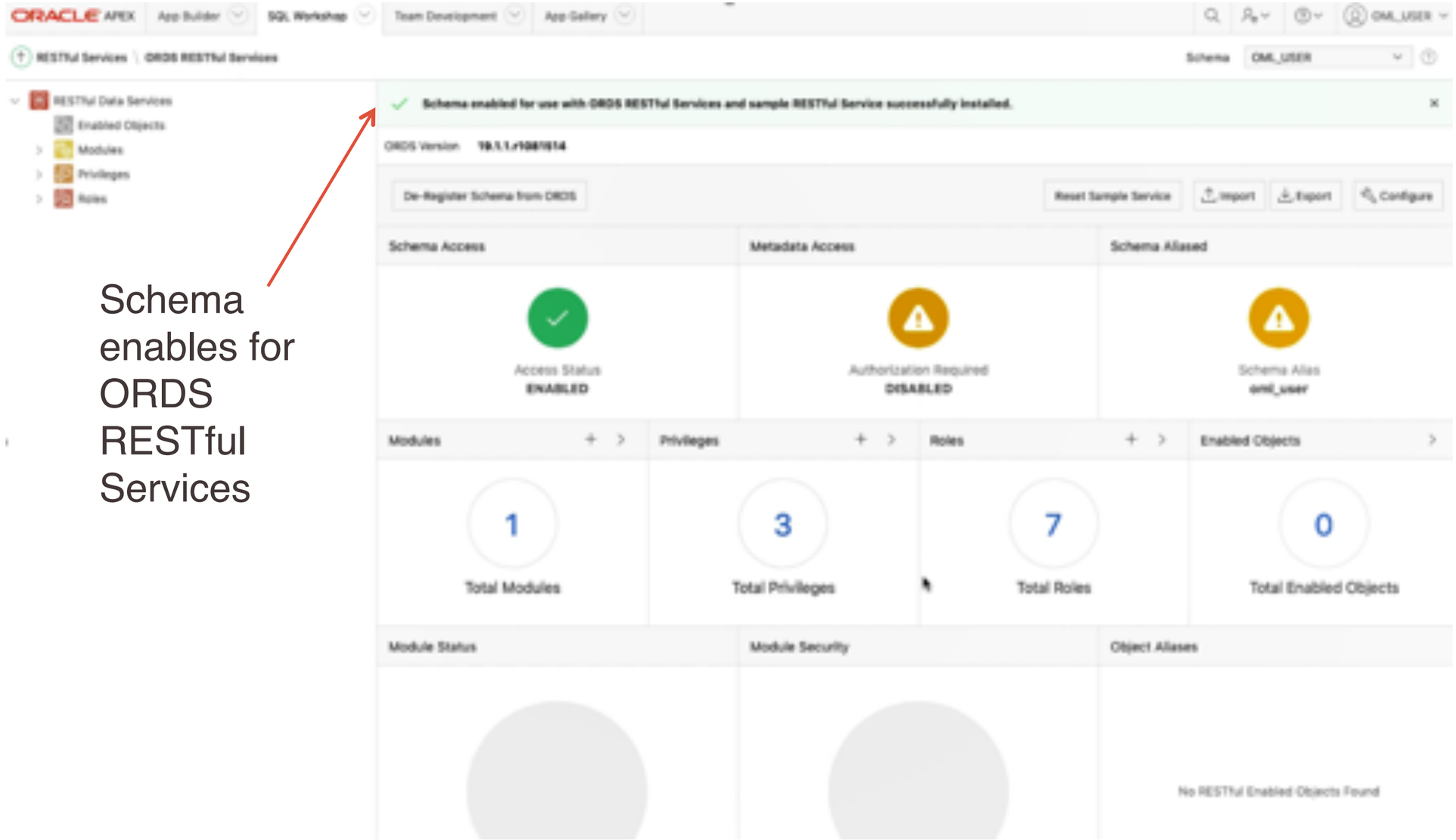
@Ftisiot
@BrendanTierney

ML Model Deployment via ORDS REST API



@Ftisiot
@BrendanTierney

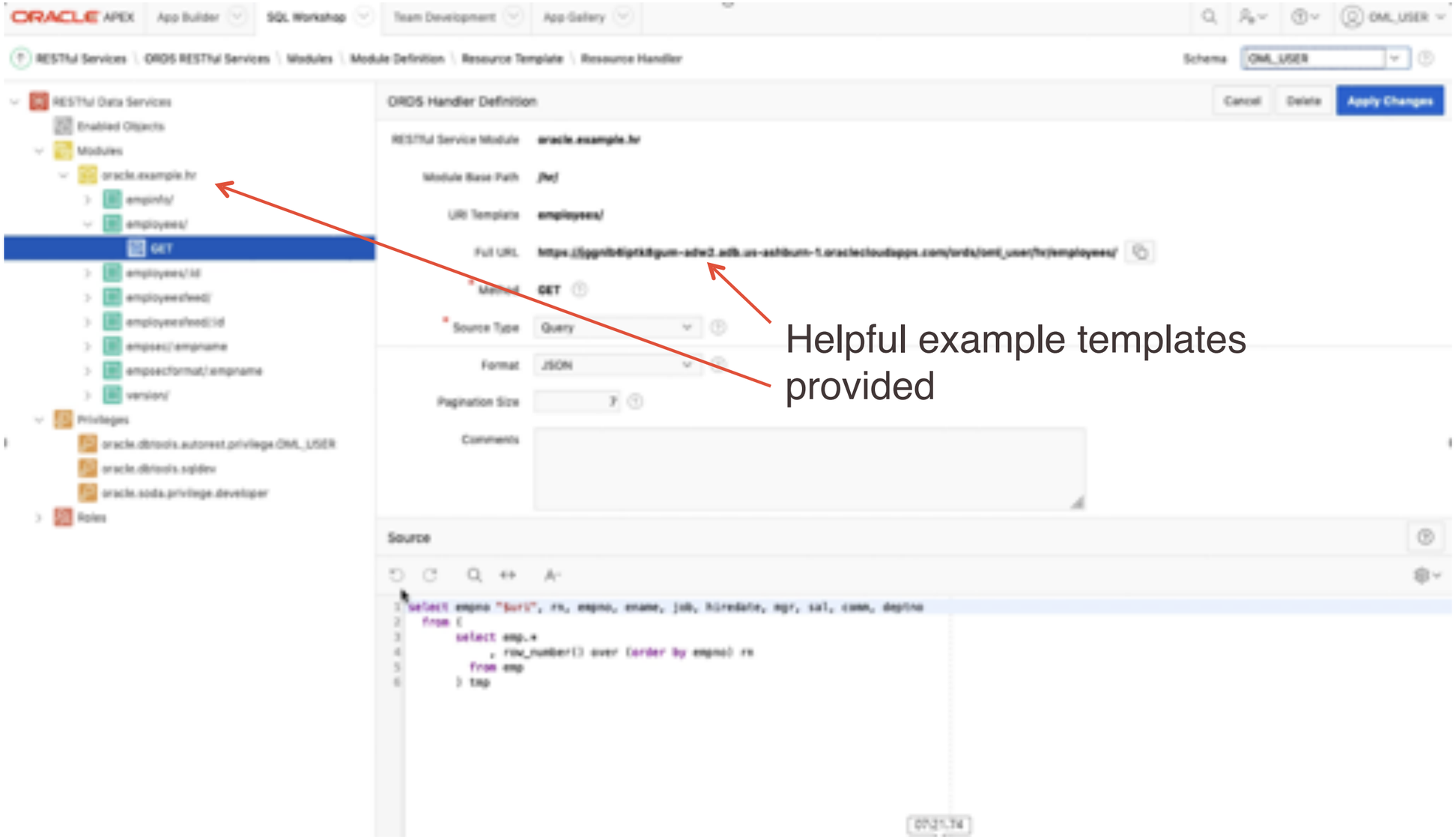
ML Model Deployment via ORDS REST API



Schema enables for ORDS RESTful Services

@Ftisiot
@BrendanTierney

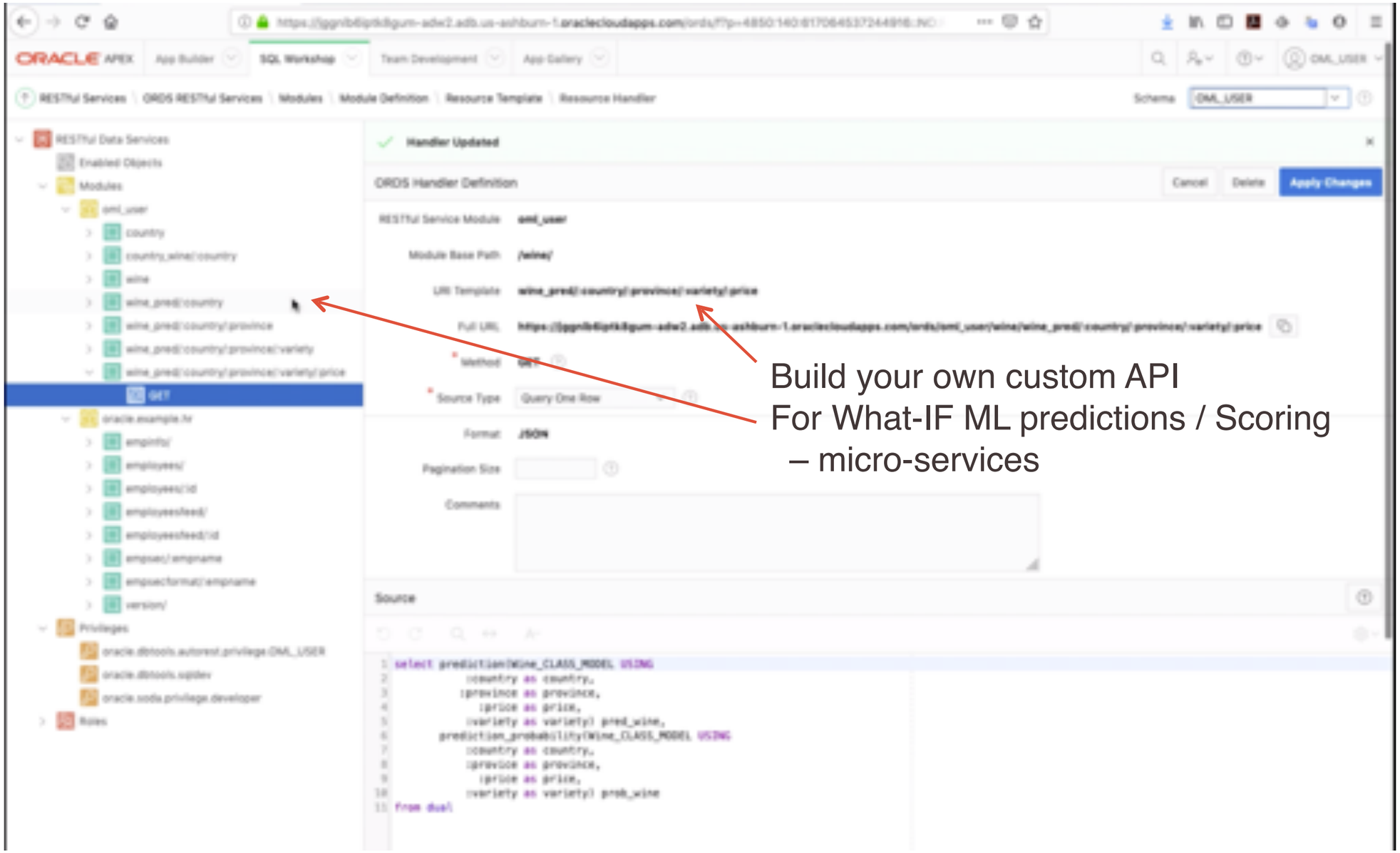
ML Model Deployment via ORDS REST API



Helpful example templates provided

@Ftisiot
@BrendanTierney

ML Model Deployment via ORDS REST API

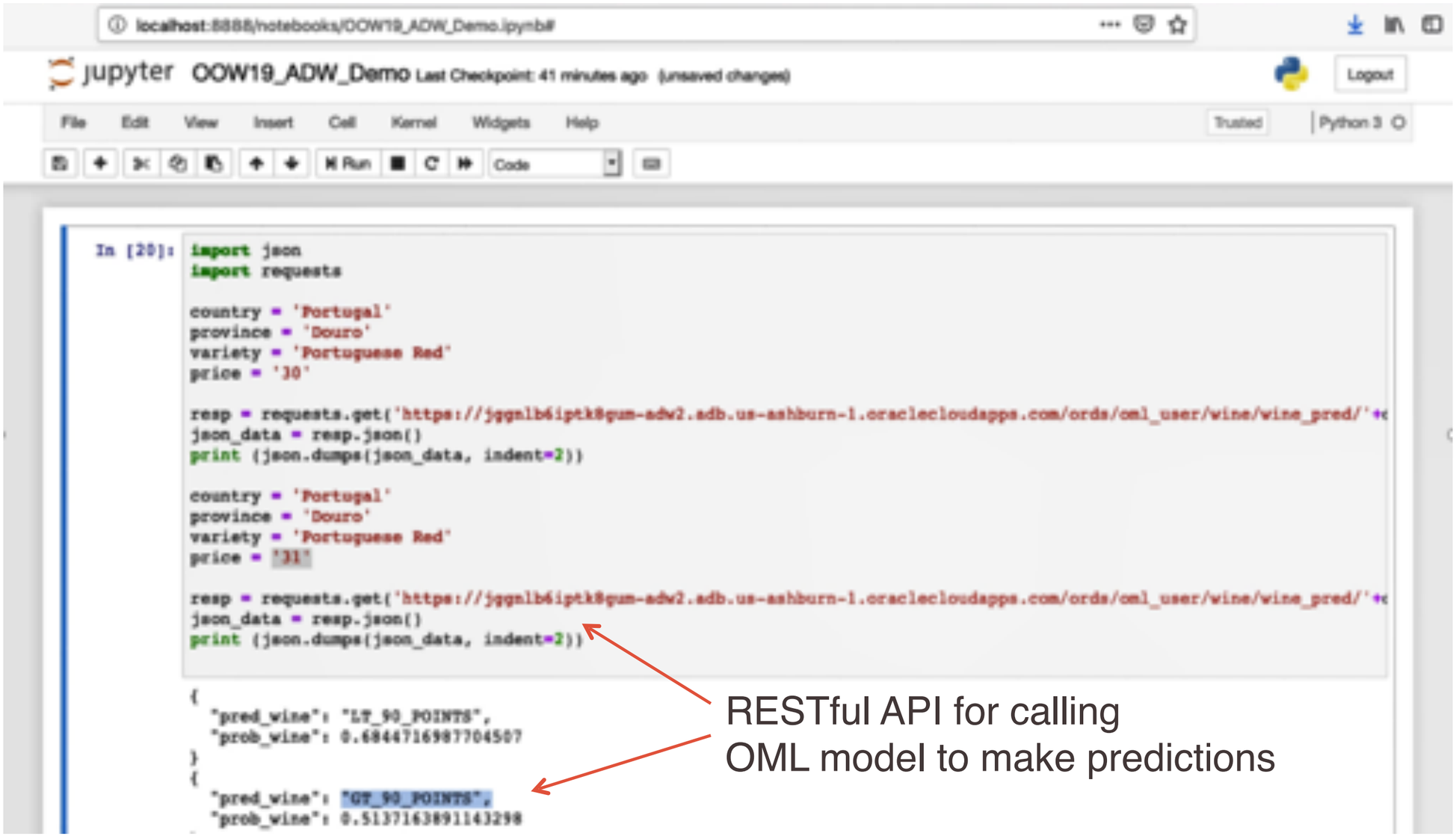


Build your own custom API
For What-IF ML predictions / Scoring
– micro-services


@Ftisiot


@BrendanTierney

ML Model Deployment via ORDS REST API



@Ftisiot
@BrendanTierney

Real-time

Wine Recommendation App + OpenDiningTable App

Get expert opinions before choosing a wine

Discover new wines and find your next favorite

@Ftisiot

@BrendanTierney

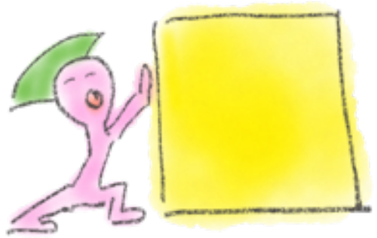
Summary

@FtisIoT
@BrendanTierney

ML in database

SECURITY

GOVERNANCE



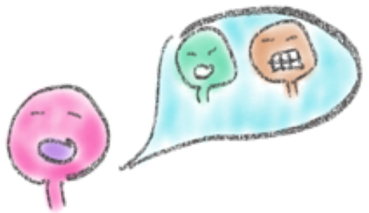
PROD
DEPLOYMENT

KNOWLEDGE
SHARING

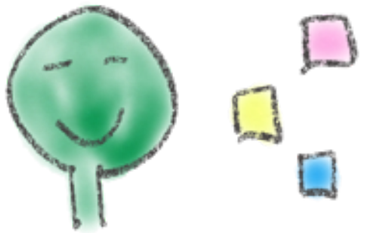
Analytics

DATA
EXPLORATION

VISUALISATIONS

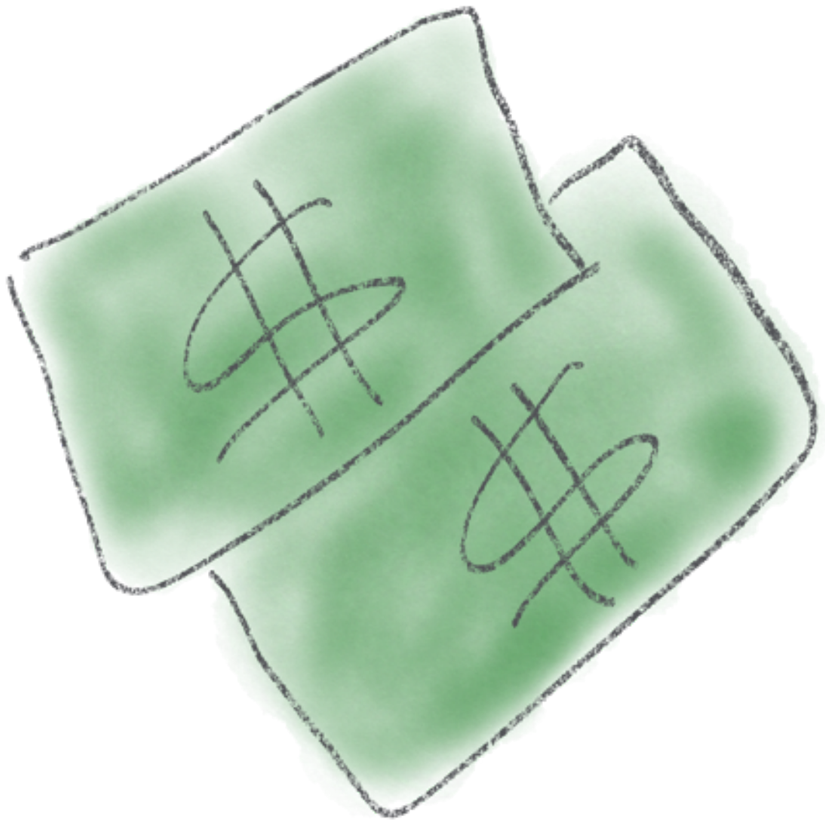


STORYTELLING



EASY REUSE

What about the money?



FREE!

@Ftisiot
@BrendanTierney

Picking a Good Wine for Under $30 using ADW, Oracle Machine Learning, OAC

Francesco Tisiot
Brendan Tierney

Rittman Mead
Oralytics


@Ftisiot


@BrendanTierney