ORACLE

Welcome at JFOKUS 2022 Hands-on Lab

Discover the best race of all time
Learn Analytics and
Machine Learning with Oracle Red Bull

Oracle Red Bull Racing

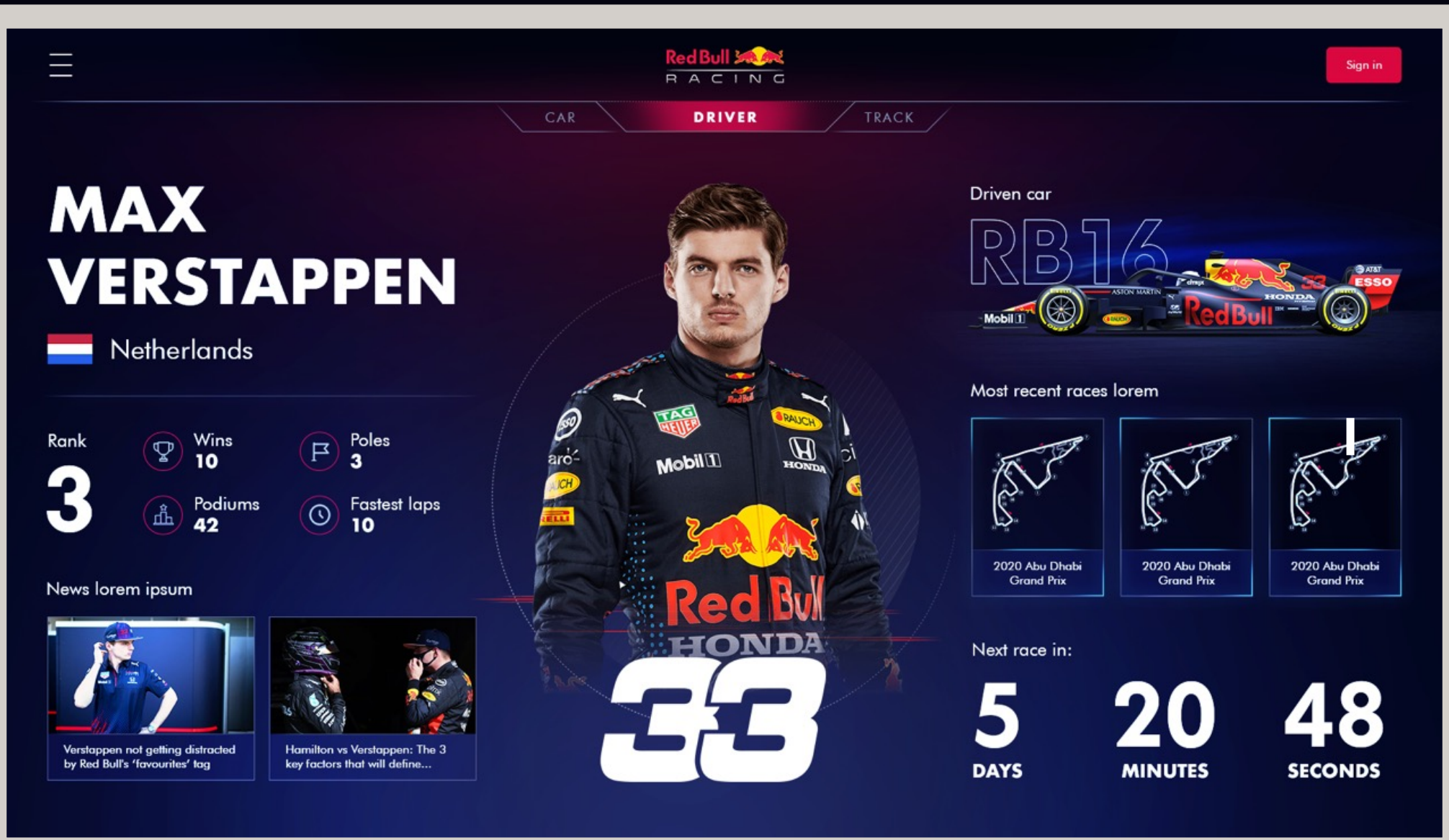
Team Performance

Red Bull Racing runs massive quantity of simulations before and during a race to optimize strategy for pit stops and tire selections.

Fan Experience

Red Bull Racing fan experience runs on Oracle ACX solutions to deliver unrivaled fan experience to their global fanbase around the world

PHOTO CREDIT: Red Bull Content Pool



Developer Masterclass Badge:

Take a short quiz on the topics covered in today's session and earn a badge to share to your social media.

Access the quiz here: **bit.ly/quiz_redbull_JFOKUS22**

This workshop is part of the **Modern AppDev Masterclass** series, and each session will have it's own badge where you can earn the title of **Developer Pioneer** recognizing your commitment and earned expertise.

The Team

Olivier Perard

Principal Data Scientist

https://www.jfokus.se/talks/1036

2 May 13:30 - 17:00 Room 23

What makes a race GREAT? We're sure you know the feeling sometimes a race is so exciting that you can't look away for one second while other times the result is clear from the beginning. What is that "magic" factor that makes a race exciting? Is it the amount of overtaking? Is it the battle for the number 1 position? Is it the weather conditions? Or are we missing some invisible factor that we can't quite understand yet?

The answer to the question lies hidden away in data. 70 Years of racing data to be precise! In this workshop you will have the opportunity to experiment with data using a combination of techniques to look for the answer. You will extract key pieces of information that define the essence of each race and visualize races and seasons using advanced techniques to determine how interesting each race is. You will learn the basics of ANALYTICS and MACHINE LEARNING and become familiar with the entire process from data preparation and visualization to model building prediction and evaluation.

You'll experience first-hand what challenges data scientists! So come join us in this fun workshop and solve the challenge!

Olivier Perard, *Oracle*

https://github.com/operard/jfokus2022

Lab instructions

https://bit.ly/jfokus22-f1

“Data is our lifeblood, we’re a data-hungry business.”

Christian Horner

Team Principal and CEO of Oracle Red Bull Racing.

“How can we make the sport even more interesting?”

What is it that makes a race great/not so great?

Rate the Race: RaceFans' Top 100 Races

These are the 100 best Formula 1 races since 2008 according to readers of RaceFans.

The ranking is compiled based on over 200,000 votes which were cast to give grand prix an average score out of ten. You can also view the ten least popular races since 2008.

You can join in Rate the Race after every grand prix by logging in with a RaceFans account. If you do not have one you can register an account here or read more about registering here.

1. 2012 Brazilian Grand Prix - 9.45/10
2. 2019 German Grand Prix - 9.44/10
3. 2011 Chinese Grand Prix - 9.24/10
4. 2014 Canadian Grand Prix - 9.19/10

- Paddock Diary: 2022 Emilia-Romagna Grand Prix
- I didn't appreciate how hard teams work at the back of the grid – Vettel
- DRS was activated too late in Emilia-Romagna Grand Prix, say drivers
- Imola podium shows McLaren's Melbourne performance was no one-off – Seidl
- Russell 'not getting comfortable' ahead of Hamilton, 'I know what he's capable of'
- Bottas escaped repeat of Monaco wheel nut misfortune on way to fifth

Use Machine Learning to predict race score

e.g. “2020 Turkish Grand Prix”

9.36

Machine Learning model score prediction

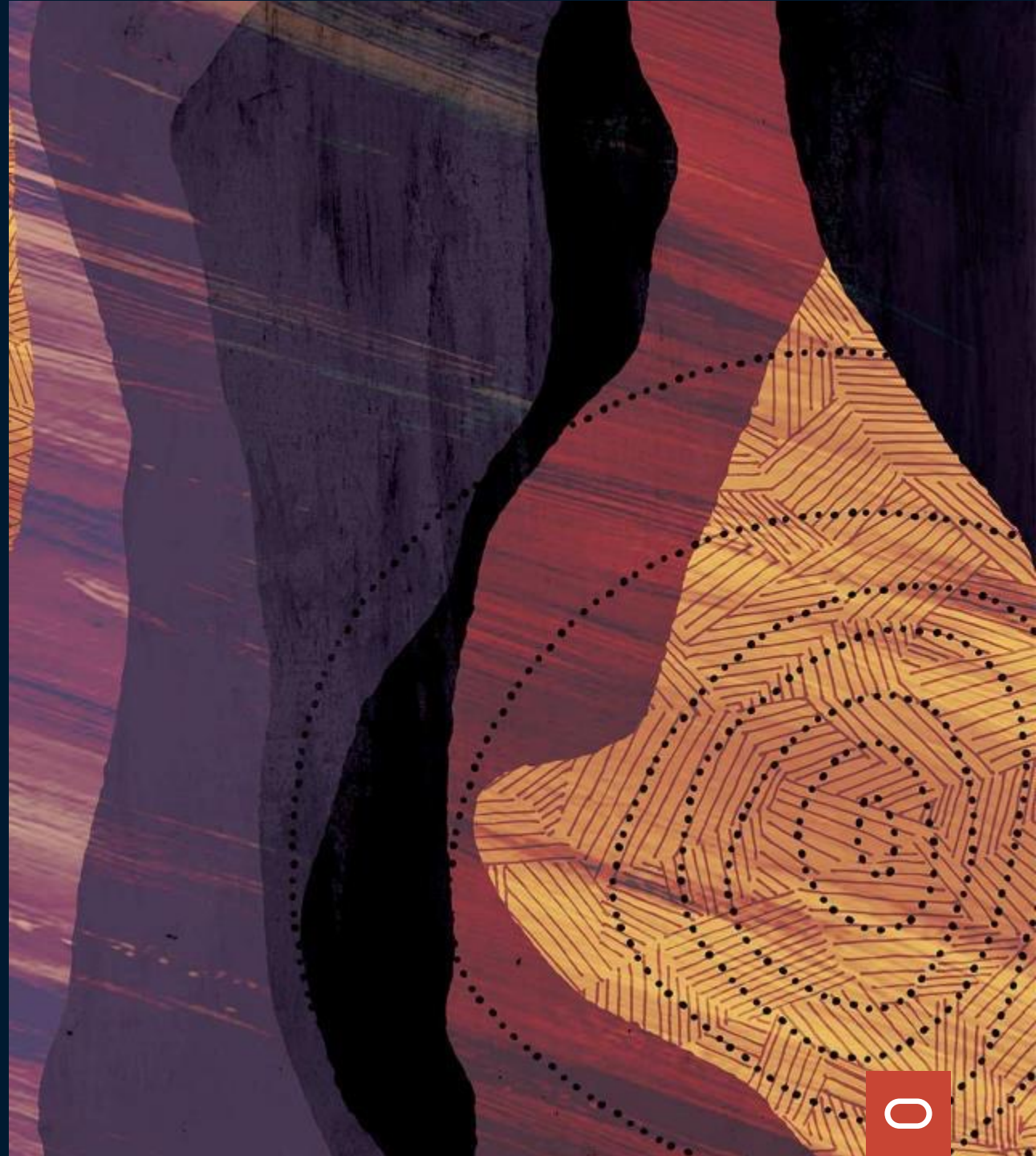
8.9

Actual score (racefans.net)

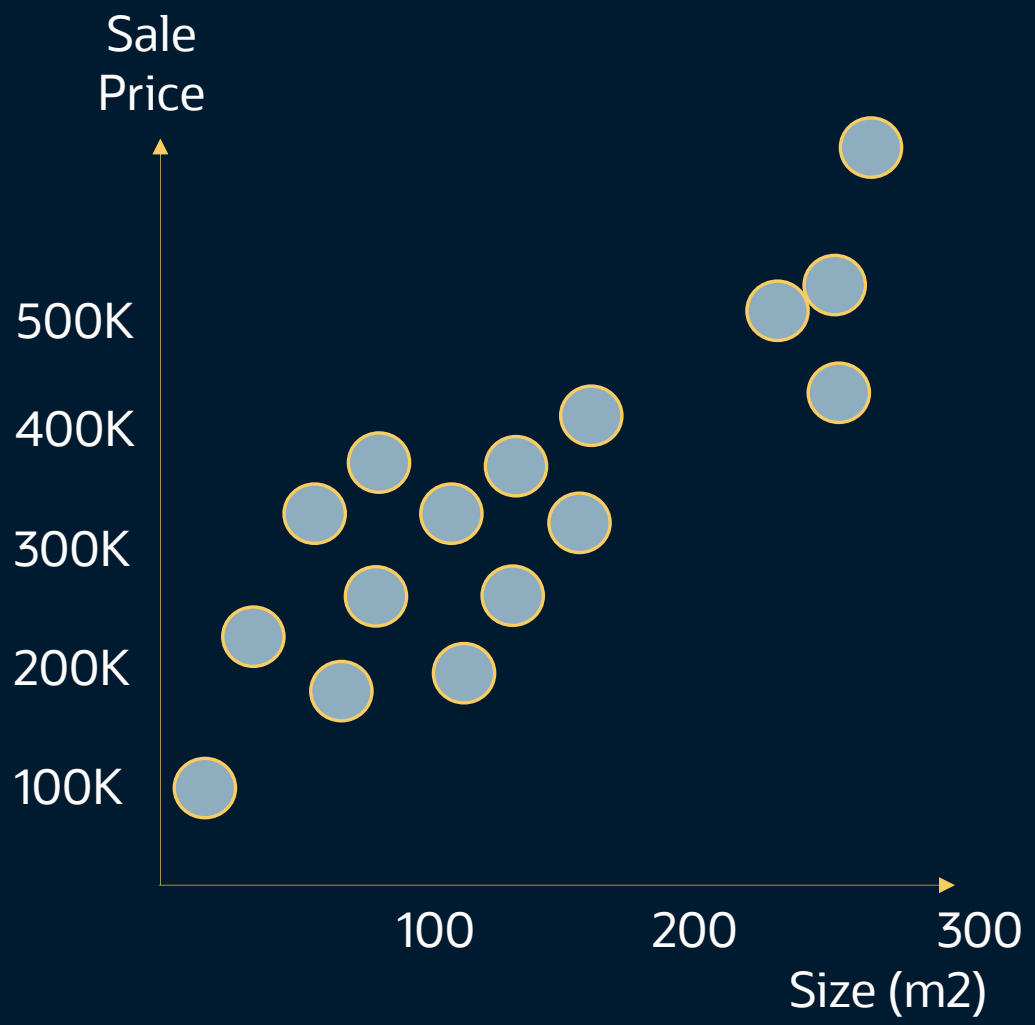
Agenda

Presentation	Introduction to the challenge + ML essentials
Lab 1	Loading the data
Lab 2	Basic data exploration
Lab 3	Advanced data exploration
Lab 4	Feature generation, machine learning, evaluation
Lab 5	Process F1 2022 Dataset from racefans.net

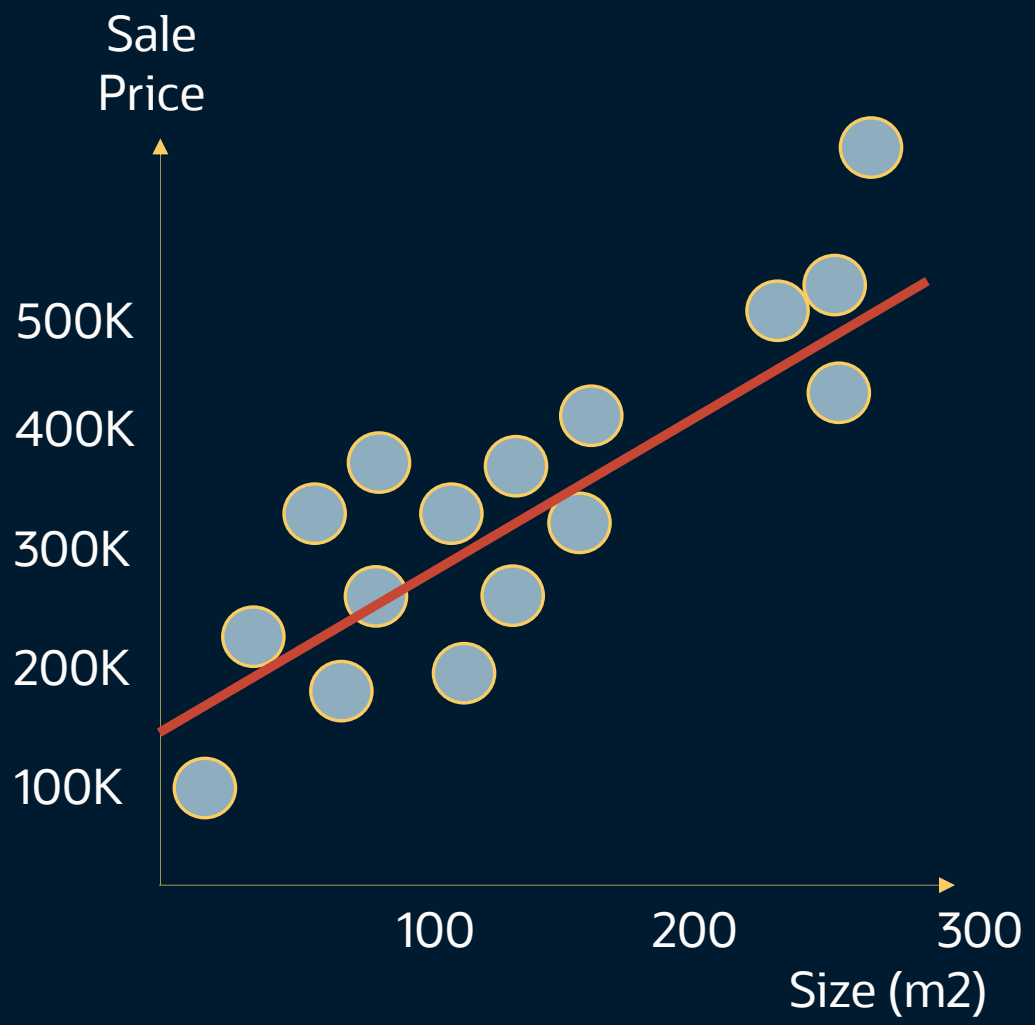
Machine Learning automatically sifts through large amounts of data to discover hidden patterns, new insights and make predictions.



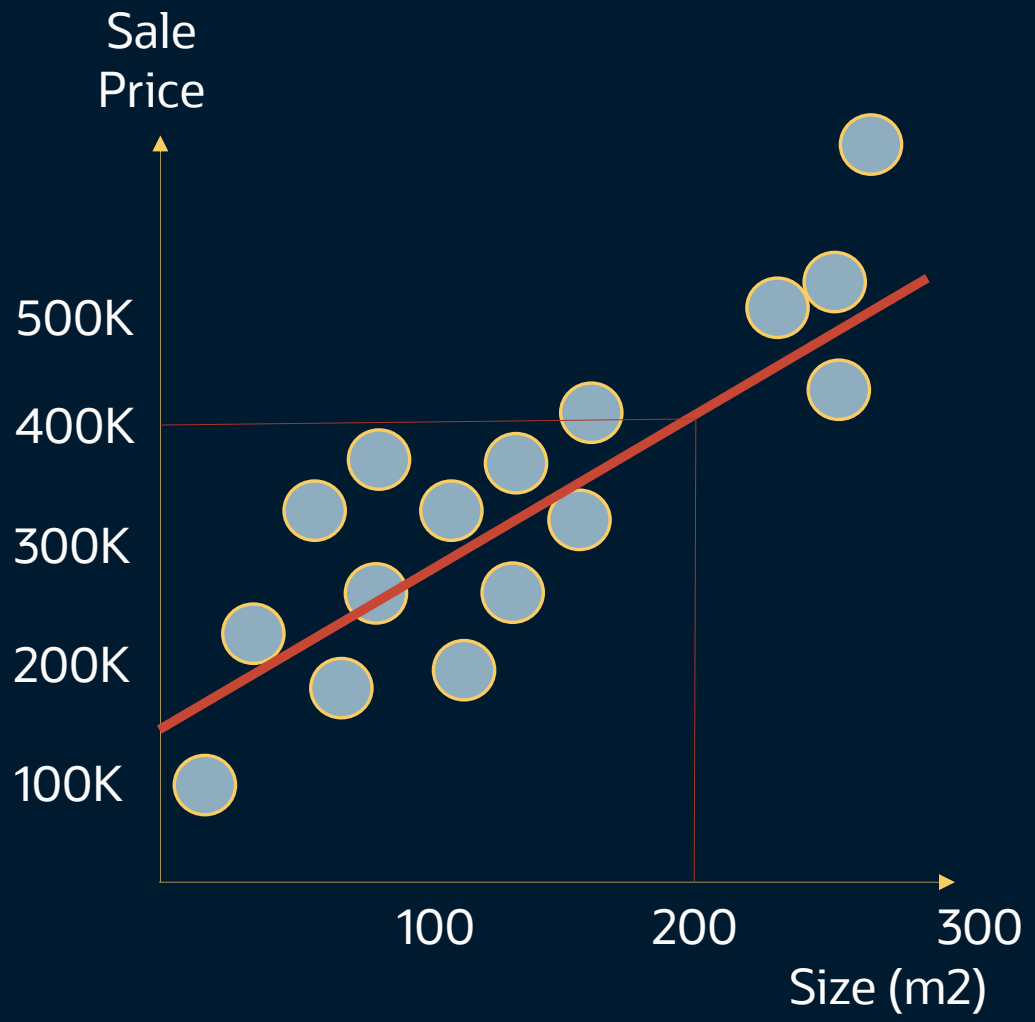
Example of a model

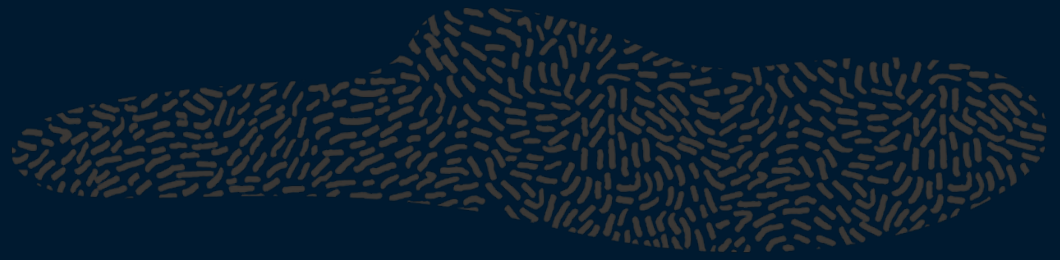


Example of a model



Example of a model





1 input feature:
Size => Sales Price

Many input features:
Size + Year Built + Neighborhood + Number of Bathrooms + … => Sales Price

Data Available for the Challenge

Name	Round	Wet weather	Over taken positions	Did-Not-Finish Count	...	Score (racefans .net)
2020 Hungarian Grand Prix	3	Y	222	1	...	6.1
2020 Styrian Grand Prix	2	N	131	3	...	6.6
2020 Austrian Grand Prix	1	N	92	9	...	8.2
2010 Malaysian Grand Prix	3	N	87	7	...	6.684
2010 Australian Grand Prix	2	Y	120	10	...	8.638
2010 Bahrain Grand Prix	1	N	76	9	...	4.587
...	...	...	...	...	...	...

Data Available for the Challenge

Name	Round	Wet weather	Over taken positions	Did-Not-Finish Count	…	Score (racefans .net)
2020 Hungarian Grand Prix	3	Y	222	1	…	6.1
2020 Styrian Grand Prix	2	N	131	3	…	6.6
2020 Austrian Grand Prix	1	N	92	9	…	8.2
2010 Malaysian Grand Prix	3	N	87	7	…	6.684
2010 Australian Grand Prix	2	Y	120	10	…	8.638
2010 Bahrain Grand Prix	1	N	76	9	…	4.587
…	…	…	…	…	…	…

Data Available for the Challenge

Name	Round	Wet weather	Over taken positions	Did-Not-Finish Count	...	Score (racefans .net)
2020 Hungarian Grand Prix	3	Y	222	1	...	6.1
2020 Styrian Grand Prix	2	N	131	3	...	6.6
2020 Austrian Grand Prix	1	N	92	9	...	8.2
2010 Malaysian Grand Prix	3	N	87	7	...	6.684
2010 Australian Grand Prix	2	Y	120	10	...	8.638
2010 Bahrain Grand Prix	1	N	76	9	...	4.587
...	...	...	...	...	...	...

Number of overtakes

Weather

Circuit (location)

DNFs

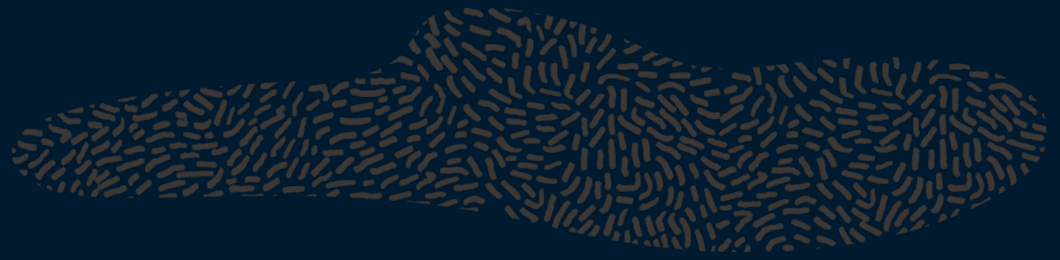
...

Score

1. Train (data until end of 2019)

Input features train (factors) | Target

Name	Round	Wet weather	Over taken positions	Did-Not-Finish Count	Score (racefans .net)
2020 Hungarian Grand Prix	3	Y	222	1	6.1
2020 Styrian Grand Prix	2	N	131	3	6.6
2020 Austrian Grand Prix	1	N	92	9	8.2
2010 Malaysian Grand Prix	3	N	87	7	6.684
2010 Australian Grand Prix	2	Y	120	10	8.638
2010 Bahrain Grand Prix	1	N	76	9	4.587
...	...	...	...	...	...



2. Test (data 2020)

Name	Round	Wet weather	Over taken positions	Did-Not-Finish Count
2021 Bahrain	1	N	239	4
...	...	...	...	...

Input features test (factors)

End result

Predicted Score
8.53

The end goal

Oracle Autonomous Database for analytics and data warehousing

Eliminates virtually all the complexities of operating a data warehouse

Easy

Easy to start, easy to drop, provision in minutes, self-optimizing, self-tuning

Scalable

From workgroup, to department and enterprise, ADB scales with ease

Fast

Leverage decades of Exadata database optimizations, no tuning required

Oracle Machine Learning

Bringing the algorithms to the data

Available and secure

Fully encrypted, policy-based access, high availability, automated backups

High-concurrency

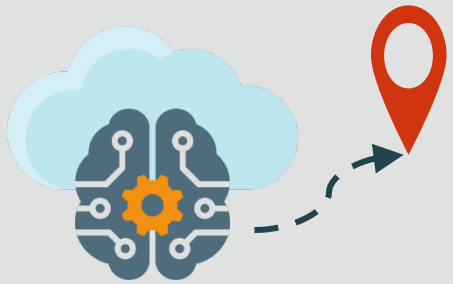
Supports multi-user access and high-concurrency workload demands

Low cost

Elastic scalability, pay only for what is used, scale up or down as needed

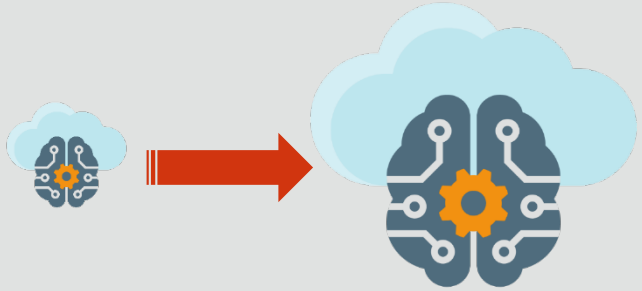
Trend: AI & ML - Analytics Machine Learning & AI

Increase productivity, Achieve enterprise goals, Innovate more



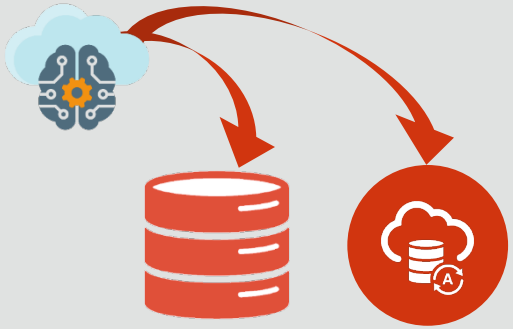
Automated

Get better results faster with less effort – even non-expert users & advanced Data Scientist



Scalable

Handle big data volumes using parallel, distributed algorithms – no data movement



Production-ready

Deploy and update data science solutions faster with integrated ML platform in Real-Time

CLOUD SOLUTION

- Analytics
- Oracle Machine Learning
- Data Science Service

Performance and scalability

In-Database Built-in algorithms

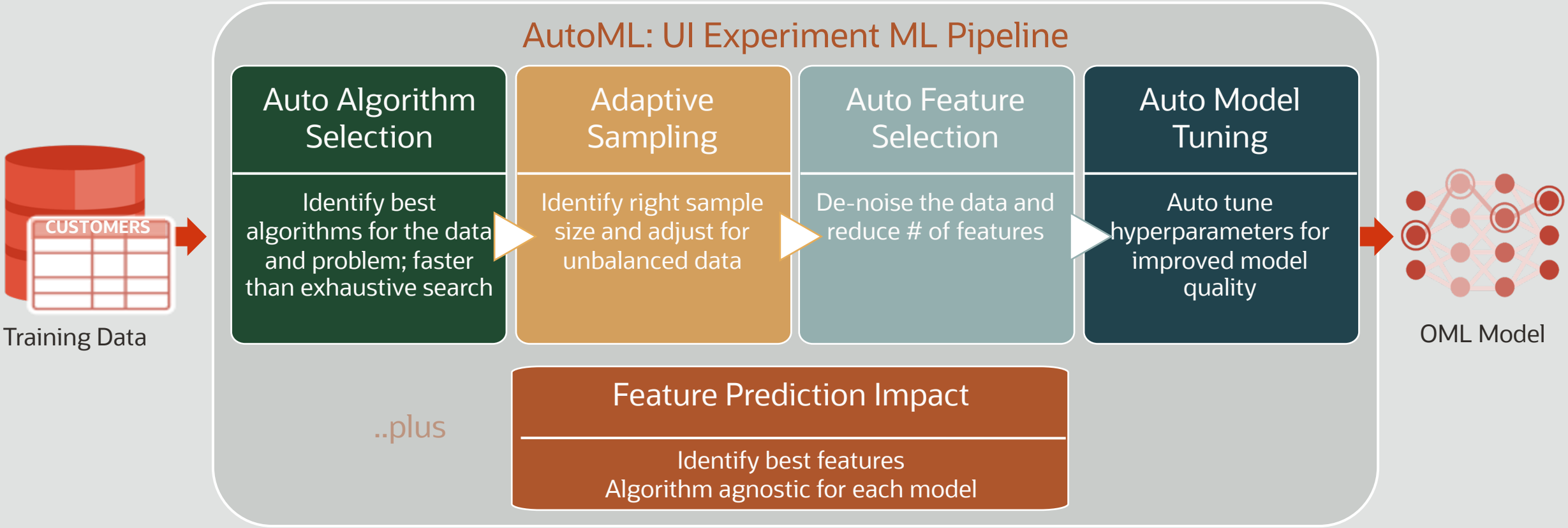
Auto-ML

GPU-ready

Best Open Source

Auto Machine Learning

Faster and Easier Machine Learning for Data Scientists, Non-Experts and Developers

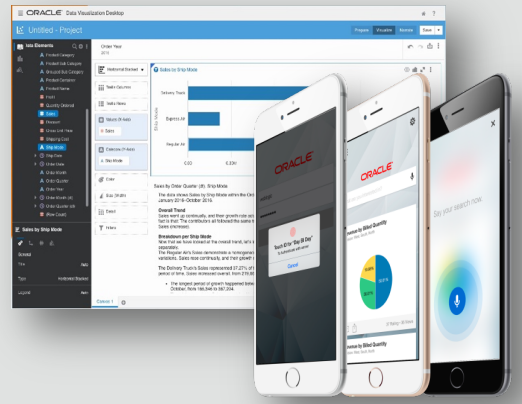


ENABLES NON-EXPERT USERS TO LEVERAGE MACHINE LEARNING

Oracle Analytics Cloud

ML Democratization to End Users or Citizen Data Scientist

ML/AI Enabled



Governed Analytics		Self-Service Analytics		Augmented Analytics	
Dashboards	Pixel-Perfect Reports	Data Preparation	Data Visualization	Voice & Chatbot	Natural Language
Semantic Models	Query Federation	Heterogeneous Data Sources	Collaboration	Data Enrichment	1-Click Explain
Role-Based Access Control	Excel Integration	Storytelling	Mobile	Adaptive & Personalized	Machine Learning

Machine Learning

Evolve towards“cititzen data scientist”

ML/AI Enabled

Embedded algorithms: Parametrize and use

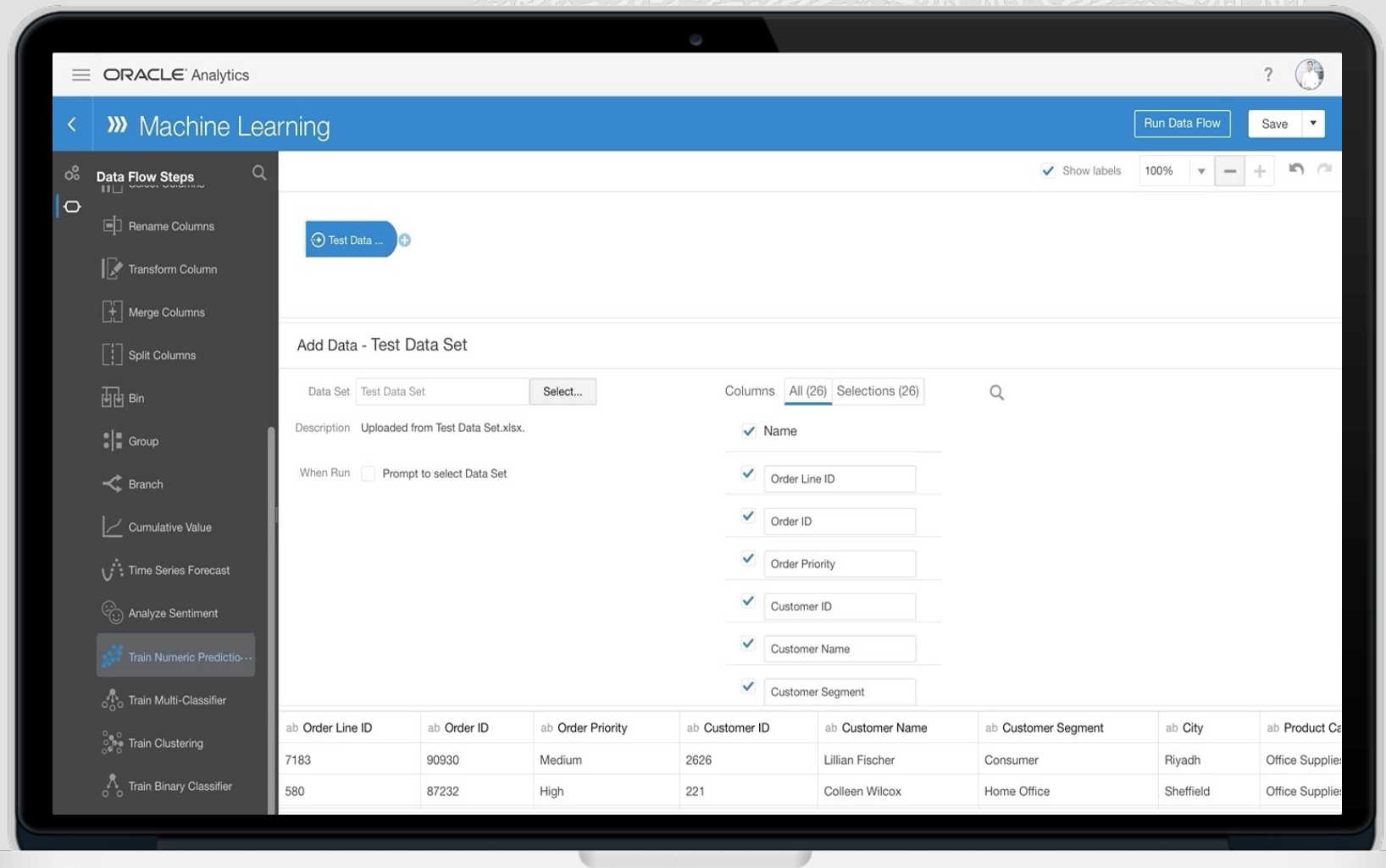
- CART for Numeric Prediction training
- Elastic Net Linear Regression for model training
- Linear Regression for model training
- Random Forest for Numeric model training

- Hierarchical Clustering for model training
- K-Means Clustering for model training

- CART for model training
- Naive Bayes for Classification
- Neural Network for Classification
- Random Forest for model training
- SVM for Classification

- CART for model training
- Logistic Regression for model training
- Naive Bayes for Classification
- Neural Network for Classification
- Random Forest for model training
- SVM for Classification

OAC consumes ML algorithms from Oracle Data Base



Augmented Analytics

AI for knowledge democratization

ML/AI Enabled

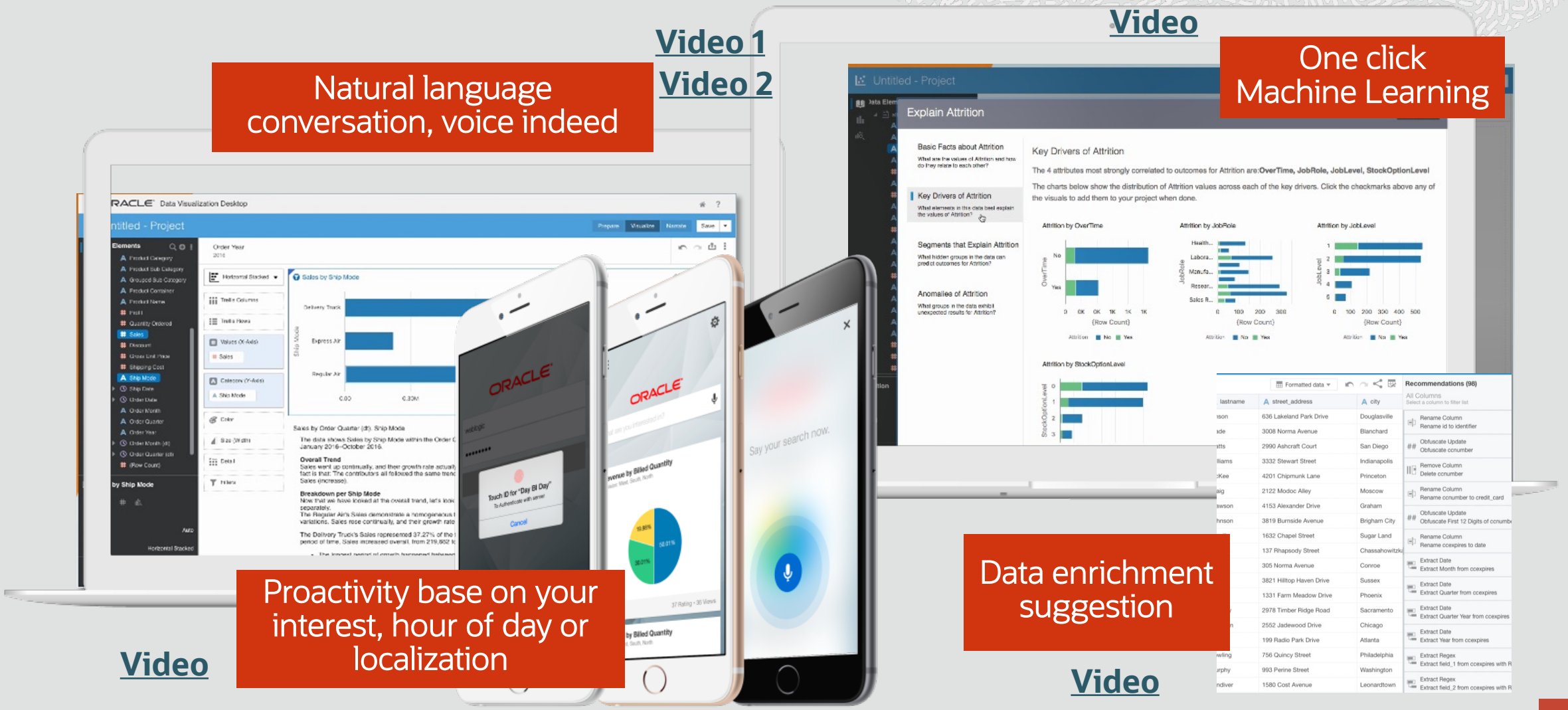
Video 1

Video 2

Natural language conversation, voice indeed

Video

One click Machine Learning



Proactivity base on your interest, hour of day or localization

Video

Data enrichment suggestion

Video

Agenda

Presentation Introduction to the challenge + ML essentials

Lab 1 Loading the data

Lab 2 Basic data exploration

Lab 3 Advanced data exploration

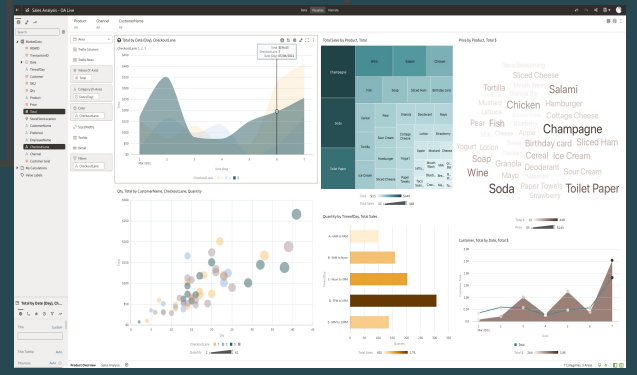
Lab 4 Feature generation, machine learning, evaluation

Lab 5 Process F1 2022 Dataset from racefans.net

Autonomous Data Warehouse

Data Loading
Model training

Analytics Cloud

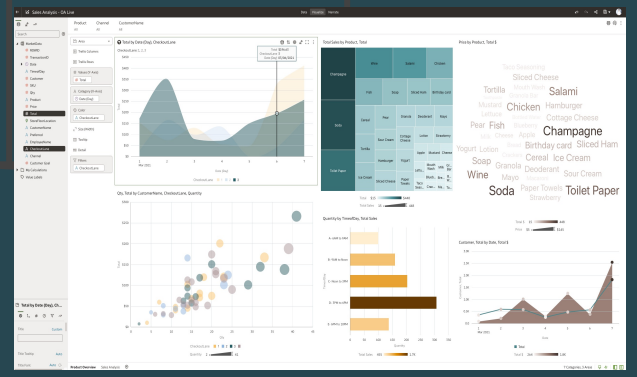


Data Exploration
Applying model (making predictions)
Verifying quality of model

Autonomous Data Warehouse

- workshop1
- workshop2
- workshop3
- workshop...
- workshop40

Analytics Cloud



Data Exploration
Applying model (making predictions)
Verifying quality of model

Users

Create | Delete | Show All Users | Search...

User Name	Full Name	Role	Email	Created On	Status
ADMIN		System Administrator		1/27/20 11:34 PM	Open
WORKSHOP1		Developer		5/1/22 3:30 PM	Open
WORKSHOP10		Developer		5/1/22 3:30 PM	Open
WORKSHOP11		Developer		5/1/22 3:30 PM	Open
WORKSHOP12		Developer		5/1/22 3:30 PM	Open
WORKSHOP13		Developer		5/1/22 3:30 PM	Open
WORKSHOP14		Developer		5/1/22 3:30 PM	Open
WORKSHOP15		Developer		5/1/22 3:30 PM	Open
WORKSHOP16		Developer		5/1/22 3:30 PM	Open
WORKSHOP17		Developer		5/1/22 3:30 PM	Open
WORKSHOP18		Developer		5/1/22 3:30 PM	Open
WORKSHOP19		Developer		5/1/22 3:30 PM	Open
WORKSHOP2		Developer		5/1/22 3:30 PM	Open

Firstname	Surname	Email	Workshop USER
Jordi Llorens	Lemonade Software Development SL.	jordillorens@lemonade.be	**WORKSHOP1**
Robert Bergqvist	KG Knutsson AB	robert.bergqvist@kgk.se	**WORKSHOP2**
Ashis Patra	Nordnet Bank AB	ashis.patra@nordnet.se	**WORKSHOP3**
Mohammadmahdi Amini	Nordnet Bank AB	Mohammadmahdi.Amini@nordnet.se	**WORKSHOP4**
Magnus Angermund	Högskolan i Gävle	magnus.angermund@hig.se	**WORKSHOP5**
Pär Lindström	CSN	par.lindstrom@csn.se	**WORKSHOP6**
Jan Hallonsten	Predictly AB	jan@predictly.se	**WORKSHOP7**
Rickard Audulv	CSN	rickard.audulv@csn.se	**WORKSHOP8**
Dan Sjöström	CSN	dan.sjostrom@csn.se	**WORKSHOP9**
Adrián Erik Clarenbeek Sánchez	Omegapoint	adrian.clarenbeek@omegapoint.se	**WORKSHOP10**
Anne-Marie Daniels	FM	annemarie.daniels@gmail.com	**WORKSHOP11**
Henrik Eklund	ITS/Umeå universitet	henrik.eklund@umu.se	**WORKSHOP12**
Mehmet Alp Guler	MAG Interactive	mehmetalpguler@gmail.com	**WORKSHOP13**
Carlos Lane Ferraro	Polismyndigheten	carlos.lane-ferraro@polisen.se	**WORKSHOP14**
Daniel Rickardsson	Sogeti Sverige AB	daniel.rickardsson@sogeti.se	**WORKSHOP15**
Mattias Lönnerberg	Thunderkick AB	mattias.lonnerberg@thunderkick.com	**WORKSHOP16**
Ilian Corneliussen	Thunderkick AB	ilian.corneliussen@thunderkick.com	**WORKSHOP17**
Adam Olsson	Thunderkick AB	adaols80@gmail.com	**WORKSHOP18**
Martin Hallgren	Folkhälsomyndigheten	martin.hallgren@folkhalsomyndigheten.se	**WORKSHOP19**
Jimmi Wimmersjö	jimwim AB	jimmi.wimmersjo@bahnhof.se	**WORKSHOP20**
Meeri Lembinen	Progmatic OÜ	meeri@progmatic.ee	**WORKSHOP21**
Emil Fenenko	Progmatic OÜ	emil@progmatic.ee	**WORKSHOP22**
Rudy Akram Kakus Kaka	Scania CV AB IT	rudy.akram@scania.com	**WORKSHOP23**
Dmitry Gorohov	Zettle	dmitry.gorohov@gmail.com	**WORKSHOP24**

Lab 1: Loading the data

Get ready for analysis by first loading the data into the Autonomous Data Warehouse.

Every individual has their own schema WORKSHOP<x>. Use the username + password provided to you individually.

Database tables have already been created, you are going to fill them with

data.

Lab instructions
https://bit.ly/jfokus22-f1

Agenda

Presentation	Introduction to the challenge + ML essentials
Lab 1	Loading the data
Lab 2	Basic data exploration
Lab 3	Advanced data exploration
Lab 4	Feature generation, machine learning, evaluation
Lab 5	Process F1 2022 Dataset from racefans.net

Lab 2: Basic Data Exploration

Finish setting up Oracle Analytics Cloud. Then get to know the data and look for clues in the data that are of value to predict the SCORE of a race.

1: Create a connection from Analytics to the Data Warehouse

2: Add the datasets to Analytics Cloud

3: Investigate the usability of various variables: overtakes, DNFs, weather

Lab instructions
https://bit.ly/jfokus22-f1

Agenda

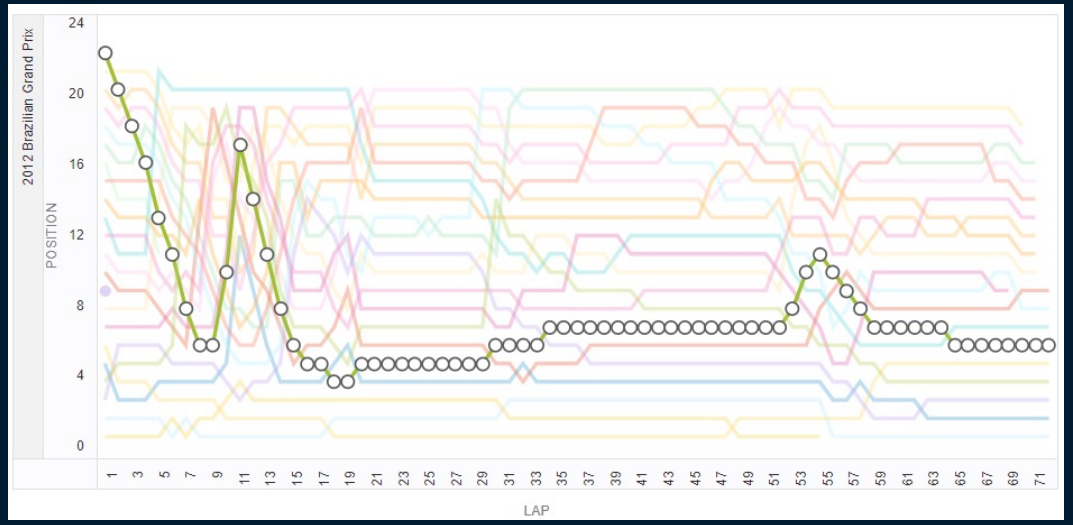
Presentation	Introduction to the challenge + ML essentials
Lab 1	Loading the data
Lab 2	Basic data exploration
Lab 3	Advanced data exploration
Lab 4	Feature generation, machine learning, evaluation
Lab 5	Process F1 2022 Dataset from racefans.net

Lab 3: Advanced Data Exploration

Go beyond the data that's visible at first sight, and look for additional clues by digging deeper into the data.

Identify Feature Generation candidates.

Make the most of the data that we have!



Lab instructions
https://bit.ly/jfokus22-f1

Agenda

Presentation	Introduction to the challenge + ML essentials
Lab 1	Loading the data
Lab 2	Basic data exploration
Lab 3	Advanced data exploration
Lab 4	Feature generation, machine learning, evaluation
Lab 5	Process F1 2022 Dataset from racefans.net

Lab 4: Feature Generation, ML Training and Evaluation

In this lab we will finally get to the actual Machine Learning part!

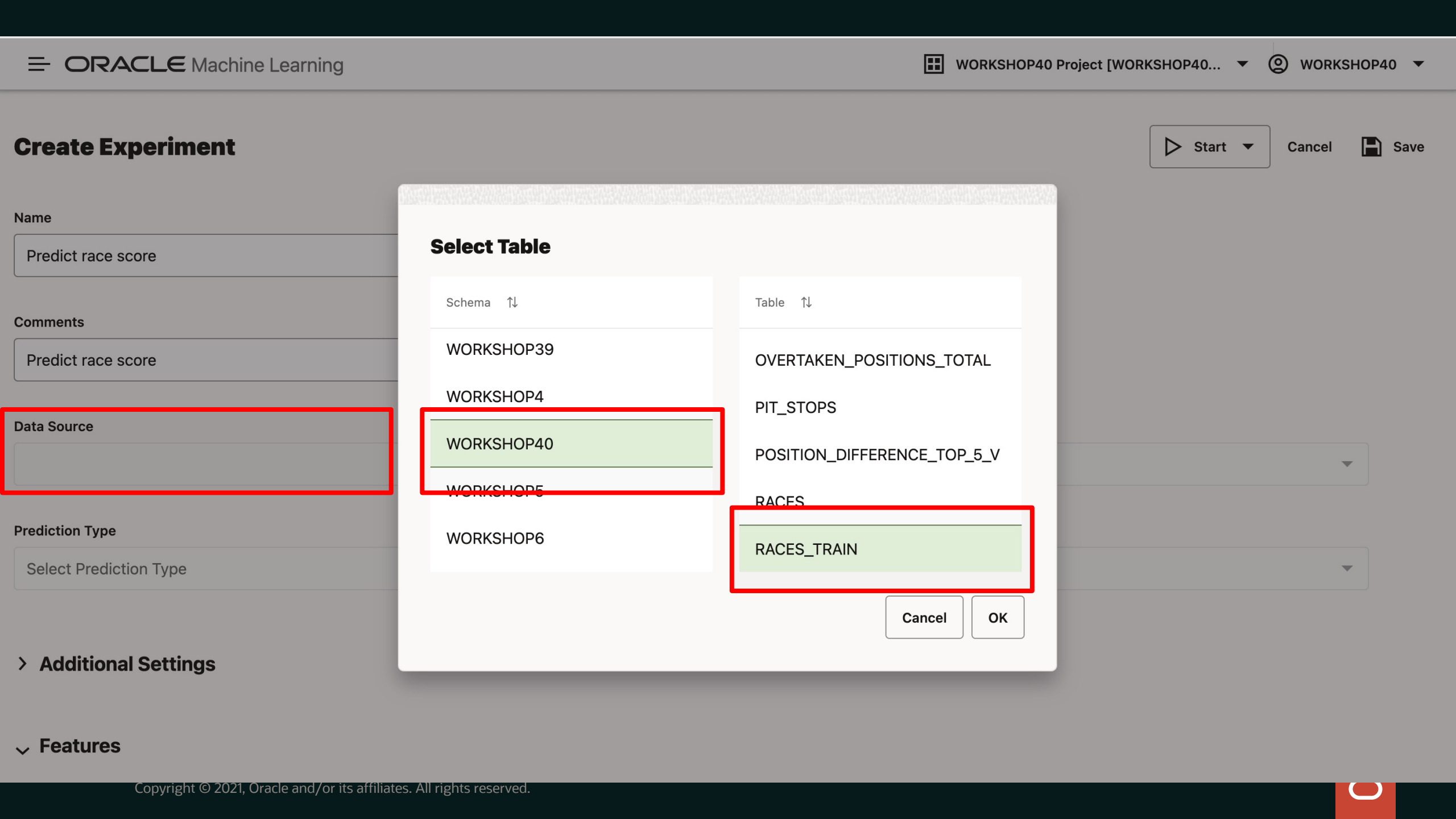
1: Feature Generation

2: Train the ML model

3: Predict the scores

4: Evaluate the accuracy of the model

Lab instructions
https://bit.ly/jfokus22-f1


ORACLE Machine Learning
WORKSHOP40 Project [WORKSHOP40...
WORKSHOP40
Create Experiment
Start
Cancel
Save
Name
Predict race score
Comments
Predict race score
Data Source
Prediction Type
Select Prediction Type
Additional Settings
Features
Select Table
Schema
Table
WORKSHOP39
WORKSHOP4
WORKSHOP40
WORKSHOP5
WORKSHOP6
OVERTAKEN_POSITIONS_TOTAL
PIT_STOPS
POSITION_DIFFERENCE_TOP_5_V
RACES
RACES_TRAIN
Cancel
OK
Copyright © 2021, Oracle and/or its affiliates. All rights reserved.

Create Experiment

Start | Cancel | Save

Name

Predict race score

Comments

Predict race score

Data Source

WORKSHOP40.RACES_TRAIN

Predict

Select Prediction Target

Required

Prediction Type

Select Prediction Type

Case ID

Select Case ID

Additional Settings

Features

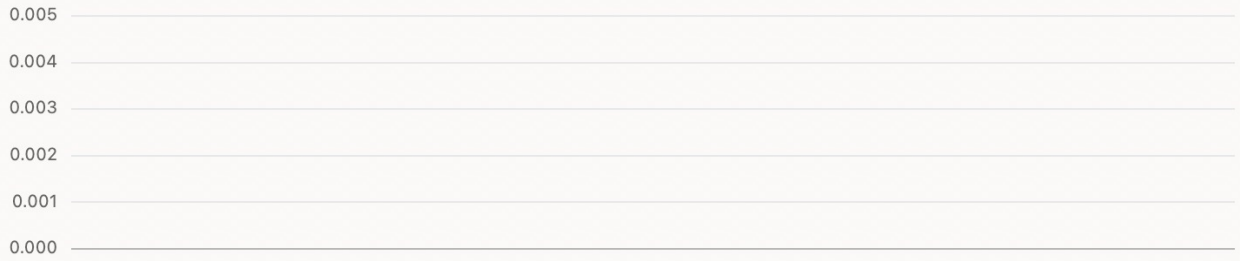
<- Experiments

Predict race score

Running
Edit
Stop

Settings

Negative Mean Absolute Error



Leader Board

Deploy Rename Create Notebook Metrics

Algorithm	Model Name	Negative Mean Absolute Error
Support Vector Machine (Linear)	SVML_552C7B93B5	-0.6360
Generalized Linear Model (Ridge ...	GLMR_7969DE44D0	-0.6348
Support Vector Machine (Gaussi...	SVMG_DA5C86CCD7	-0.6346
Generalized Linear Model	GLM_0418DDD523	-0.6345

Running: 0h 0m

- Initialization — Completed
- Algorithm Selection — Completed
- Adaptive Sampling — Completed
- Feature Selection — Completed
- Model Tuning — Running
- Generalized Linear Model — Completed

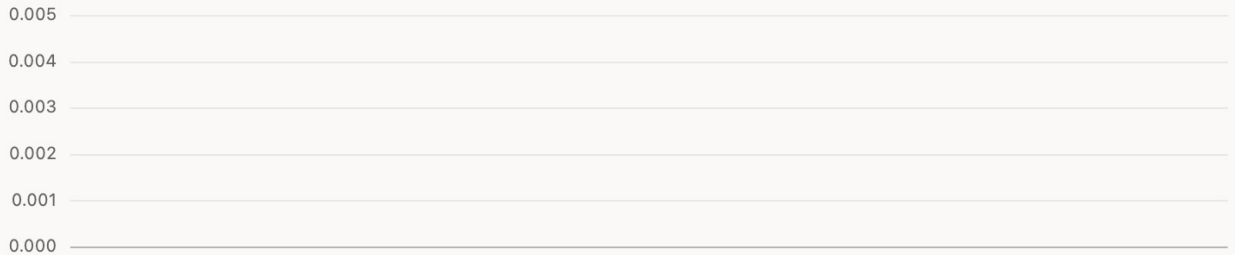
<- Experiments

Predict race score

Completed | Edit | Start

Settings

Negative Mean Absolute Error



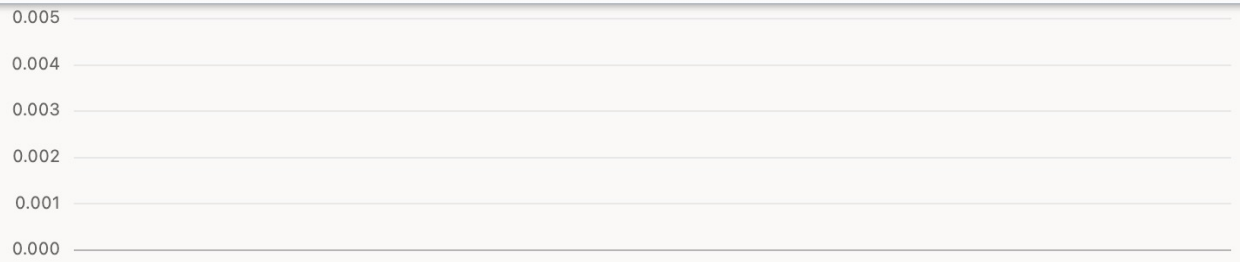
Leader Board

Deploy | Rename | Create Notebook | Metrics

Algorithm	Model Name	Negative Mean Absolute Error
Neural Network	NN_F28CEE823D	-0.6774
Support Vector Machine (Linear)	SVML_552C7B93B5	-0.6360
Generalized Linear Model (Ridge ...	GLMR_7969DE44D0	-0.6348
Support Vector Machine (Gaussi...	SVMG_DA5C86CCD7	-0.6346

Completed: 0h 1m

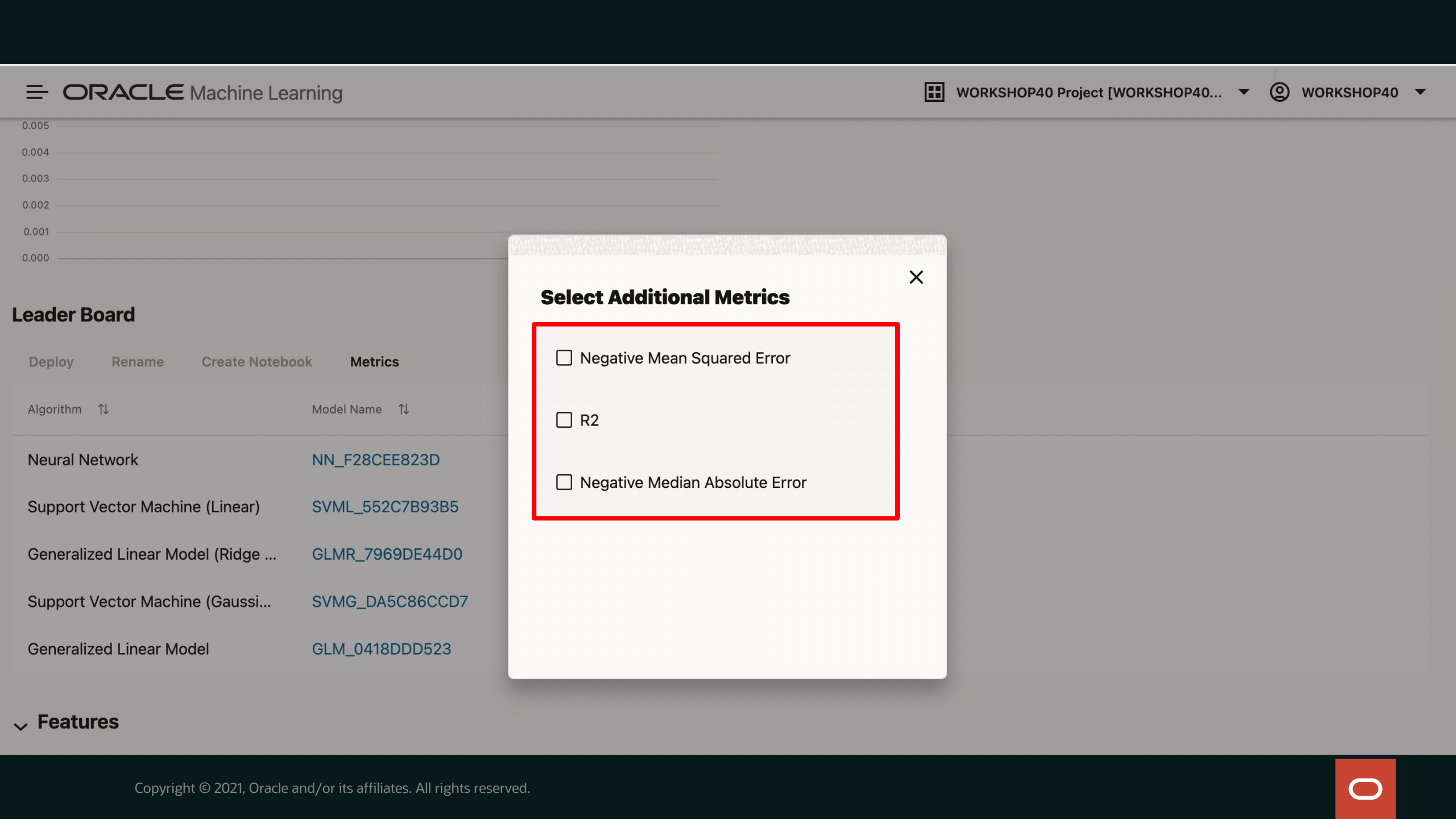
- Initialization — Completed
- Algorithm Selection — Completed
- Adaptive Sampling — Completed
- Feature Selection — Completed
- Model Tuning — Completed
- Generalized Linear Model — Completed



Leader Board

Deploy | Rename | Create Notebook | **Metrics**

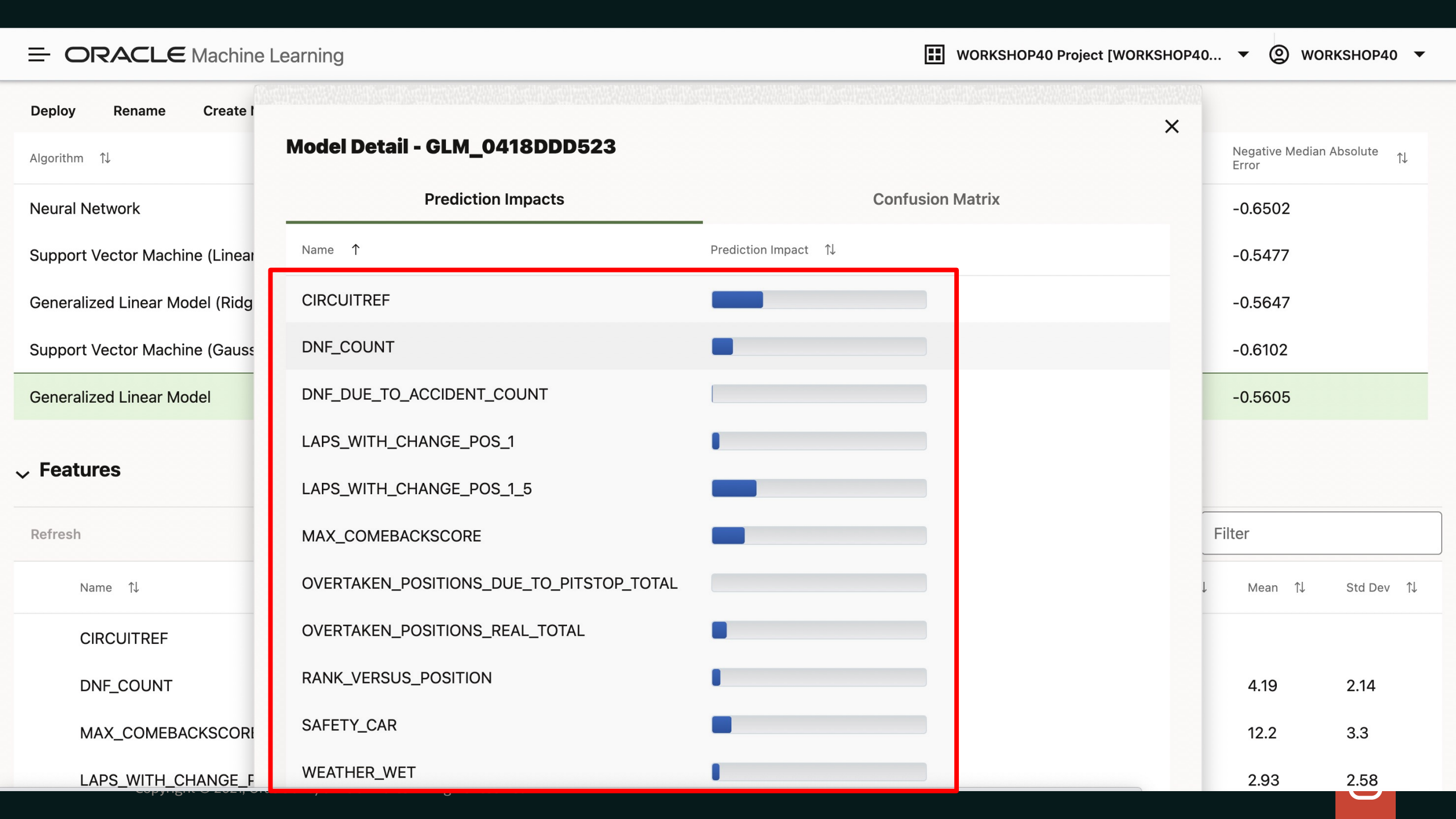
Algorithm	Model Name	Negative Mean Absolute Error
Neural Network	NN_F28CEE823D	-0.6774
Support Vector Machine (Linear)	SVML_552C7B93B5	-0.6360
Generalized Linear Model (Ridge ...	GLMR_7969DE44D0	-0.6348
Support Vector Machine (Gaussi...	SVMG_DA5C86CCD7	-0.6346
Generalized Linear Model	GLM_0418DDD523	-0.6345

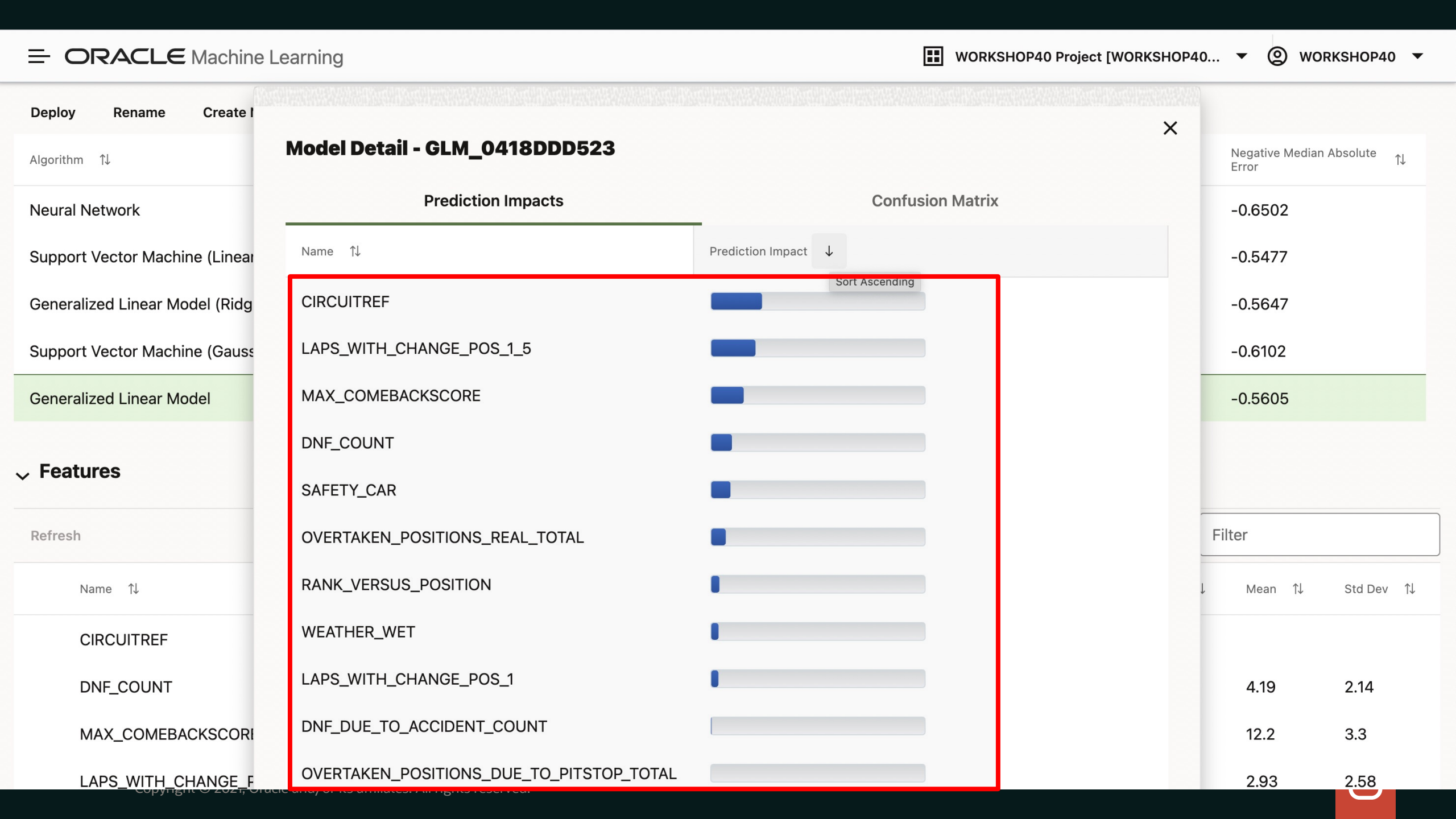


Leader Board

Deploy Rename Create Notebook **Metrics**

Algorithm	Model Name	Negative Mean Absolute Error	Negative Mean Squared Error	R2	Negative Median Absolute Error
Neural Network	NN_F28CEE823D	-0.6774	-0.7624	0.4804	-0.6502
Support Vector Machine (Linear)	SVML_552C7B93B5	-0.6360	-0.6458	0.5599	-0.5477
Generalized Linear Model (Ridge ...	GLMR_7969DE44D0	-0.6348	-0.6331	0.5685	-0.5647
Support Vector Machine (Gaussi...	SVMG_DA5C86CCD7	-0.6346	-0.6079	0.5857	-0.6102
Generalized Linear Model	GLM_0418DDD523	-0.6345	-0.6333	0.5684	-0.5605


ORACLE Machine Learning
WORKSHOP40 Project [WORKSHOP40...
WORKSHOP40
Deploy
Rename
Create I
Algorithm
Neural Network
Support Vector Machine (Linear
Generalized Linear Model (Ridg
Support Vector Machine (Gauss
Generalized Linear Model
Negative Median Absolute Error
-0.6502
-0.5477
-0.5647
-0.6102
-0.5605
Features
Refresh
Name
CIRCUITREF
DNF_COUNT
MAX_COMEBACKSCORE
LAPS_WITH_CHANGE_F
Filter
Mean
Std Dev
4.19
2.14
12.2
3.3
2.93
2.58
Model Detail - GLM_0418DDD523
Prediction Impacts
Confusion Matrix
Name
Prediction Impact
CIRCUITREF
DNF_COUNT
DNF_DUE_TO_ACCIDENT_COUNT
LAPS_WITH_CHANGE_POS_1
LAPS_WITH_CHANGE_POS_1_5
MAX_COMEBACKSCORE
OVERTAKEN_POSITIONS_DUE_TO_PITSTOP_TOTAL
OVERTAKEN_POSITIONS_REAL_TOTAL
RANK_VERSUS_POSITION
SAFETY_CAR
WEATHER_WET


ORACLE Machine Learning
WORKSHOP40 Project [WORKSHOP40...
WORKSHOP40
Deploy
Rename
Create I
Algorithm
Neural Network
Support Vector Machine (Linear
Generalized Linear Model (Ridg
Support Vector Machine (Gauss
Generalized Linear Model
Negative Median Absolute Error
-0.6502
-0.5477
-0.5647
-0.6102
-0.5605
Features
Refresh
Name
CIRCUITREF
DNF_COUNT
MAX_COMEBACKSCORE
LAPS_WITH_CHANGE_P
Filter
Mean
Std Dev
4.19
2.14
12.2
3.3
2.93
2.58
Model Detail - GLM_0418DDD523
Prediction Impacts
Confusion Matrix
Name
Prediction Impact
Sort Ascending
CIRCUITREF
LAPS_WITH_CHANGE_POS_1_5
MAX_COMEBACKSCORE
DNF_COUNT
SAFETY_CAR
OVERTAKEN_POSITIONS_REAL_TOTAL
RANK_VERSUS_POSITION
WEATHER_WET
LAPS_WITH_CHANGE_POS_1
DNF_DUE_TO_ACCIDENT_COUNT
OVERTAKEN_POSITIONS_DUE_TO_PITSTOP_TOTAL

Leader Board

Deploy | Rename | Create Notebook | Metrics

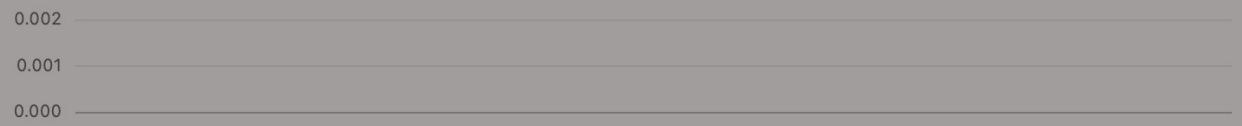
Algorithm	Model Name	Negative Mean Absolute Error	Negative Mean Squared Error	R2	Negative Median Absolute Error
Neural Network	NN_F28CEE823D	-0.6774	-0.7624	0.4804	-0.6502
Support Vector Machine (Linear)	SVML_552C7B93B5	-0.6360	-0.6458	0.5599	-0.5477
Generalized Linear Model (Ridge ...	GLMR_7969DE44D0	-0.6348	-0.6331	0.5685	-0.5647
Support Vector Machine (Gaussi...	SVMG_DA5C86CCD7	-0.6346	-0.6079	0.5857	-0.6102
Generalized Linear Model	GLM_0418DDD523	-0.6345	-0.6333	0.5684	-0.5605

Features

Refresh

Filter

Name	Importance	Type	Percent NULLs	Distinct Values	Min	Max	Mean	Std Dev
WEATHER_WET		VARCHAR2	0	2				



Leader Board

Deploy Rename Create Notebook Metrics

Algorithm	Model Name	R2	Negative Median Absolute Error
Neural Network	NN_F28CEE823D	0.4804	-0.6502
Support Vector Machine (Linear)	SVML_552C7B93B5	0.5599	-0.5477
Generalized Linear Model (Ridge ...	GLMR_7969DE44D0	0.5685	-0.5647
Support Vector Machine (Gaussi...	SVMG_DA5C86CCD7	0.5857	-0.6102
Generalized Linear Model	GLM_0418DDD523	0.5684	-0.5605

Create Notebook

Create a notebook based on selected model and experiment's settings

Notebook Name:

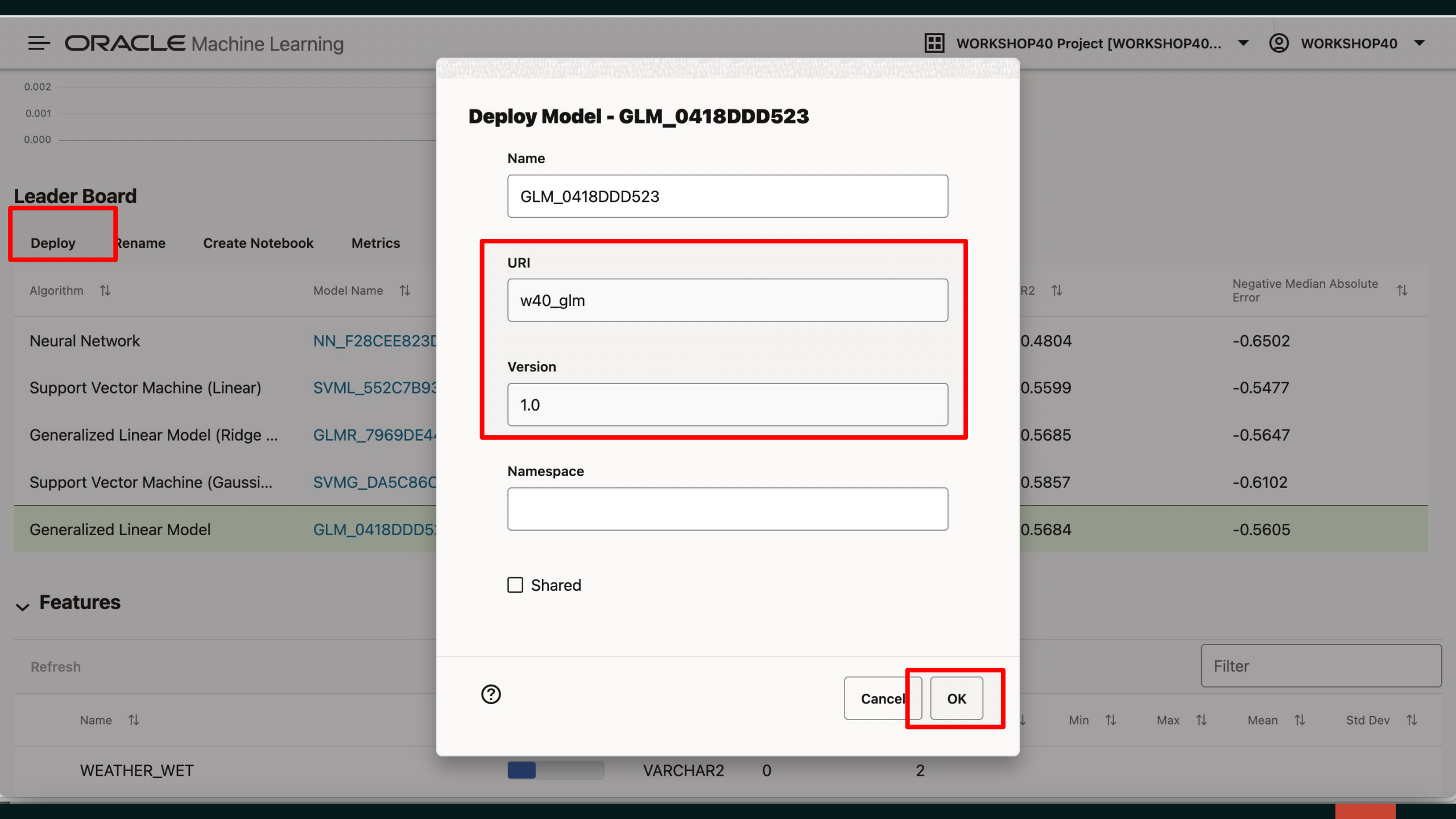
Predict race score GLM (1)

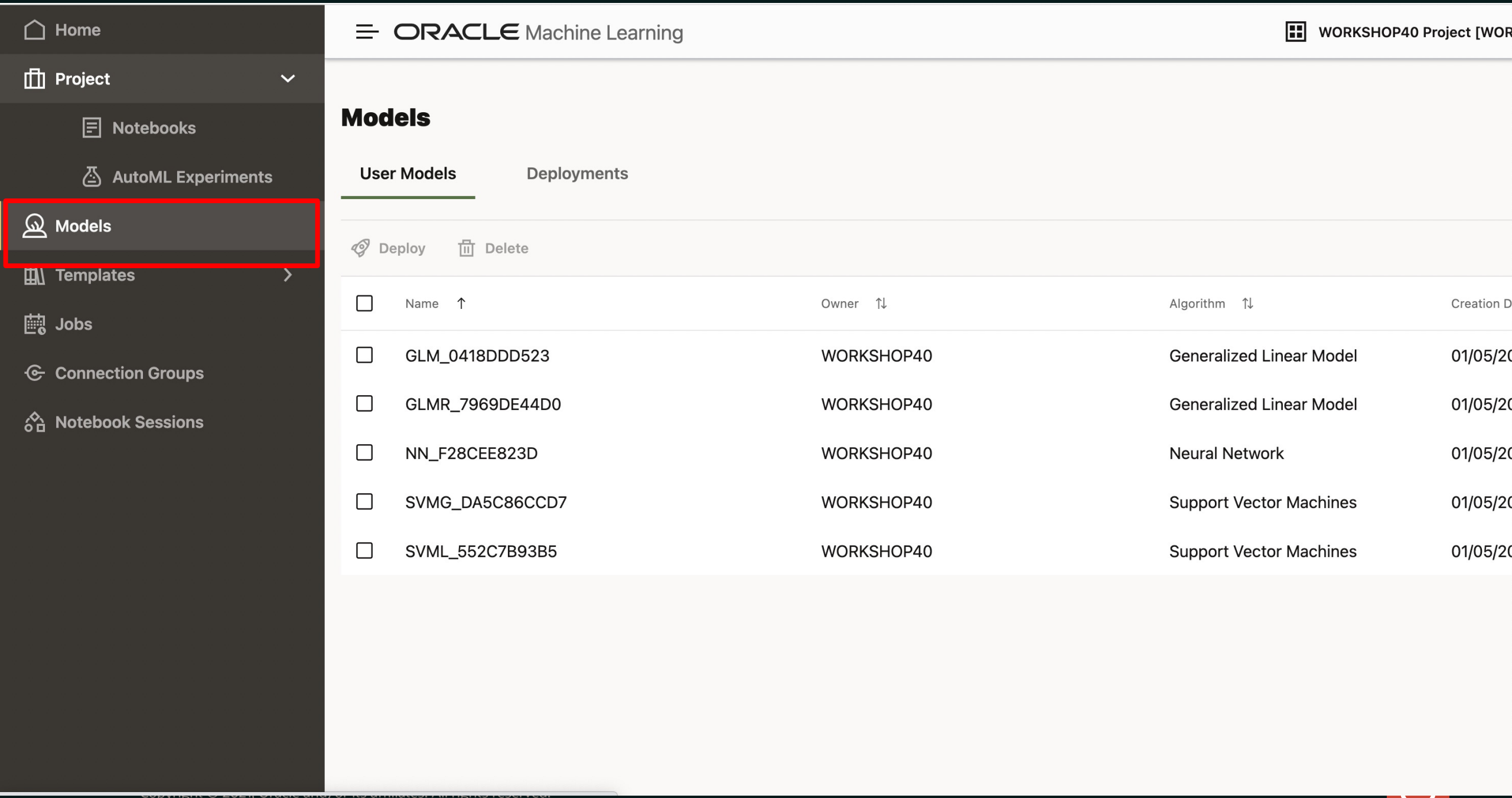
Cancel OK

Features

Refresh Filter

Name	Importance	Type	Percent NULLs	Distinct Values	Min	Max	Mean	Std Dev
WEATHER_WET		VARCHAR2	0	2				


ORACLE Machine Learning
WORKSHOP40 Project [WORKSHOP40...
WORKSHOP40
0.002
0.001
0.000
Leader Board
Deploy
Rename
Create Notebook
Metrics
Algorithm
Model Name
R2
Negative Median Absolute Error
Neural Network
NN_F28CEE823D
0.4804
-0.6502
Support Vector Machine (Linear)
SVML_552C7B93
0.5599
-0.5477
Generalized Linear Model (Ridge ...
GLMR_7969DE44
0.5685
-0.5647
Support Vector Machine (Gaussi...
SVMG_DA5C86C
0.5857
-0.6102
Generalized Linear Model
GLM_0418DDD5
0.5684
-0.5605
Features
Refresh
Filter
Name
Min
Max
Mean
Std Dev
WEATHER_WET
VARCHAR2
0
2
Deploy Model - GLM_0418DDD523
Name
GLM_0418DDD523
URI
w40_glm
Version
1.0
Namespace
Shared
Cancel
OK


Home
Project
Notebooks
AutoML Experiments
Models
Templates
Jobs
Connection Groups
Notebook Sessions
ORACLE Machine Learning
WORKSHOP40 Project [WOR
Models
User Models
Deployments
Deploy
Delete
Name
Owner
Algorithm
Creation D
GLM_0418DDD523
WORKSHOP40
Generalized Linear Model
01/05/20
GLMR_7969DE44D0
WORKSHOP40
Generalized Linear Model
01/05/20
NN_F28CEE823D
WORKSHOP40
Neural Network
01/05/20
SVMG_DA5C86CCD7
WORKSHOP40
Support Vector Machines
01/05/20
SVML_552C7B93B5
WORKSHOP40
Support Vector Machines
01/05/20

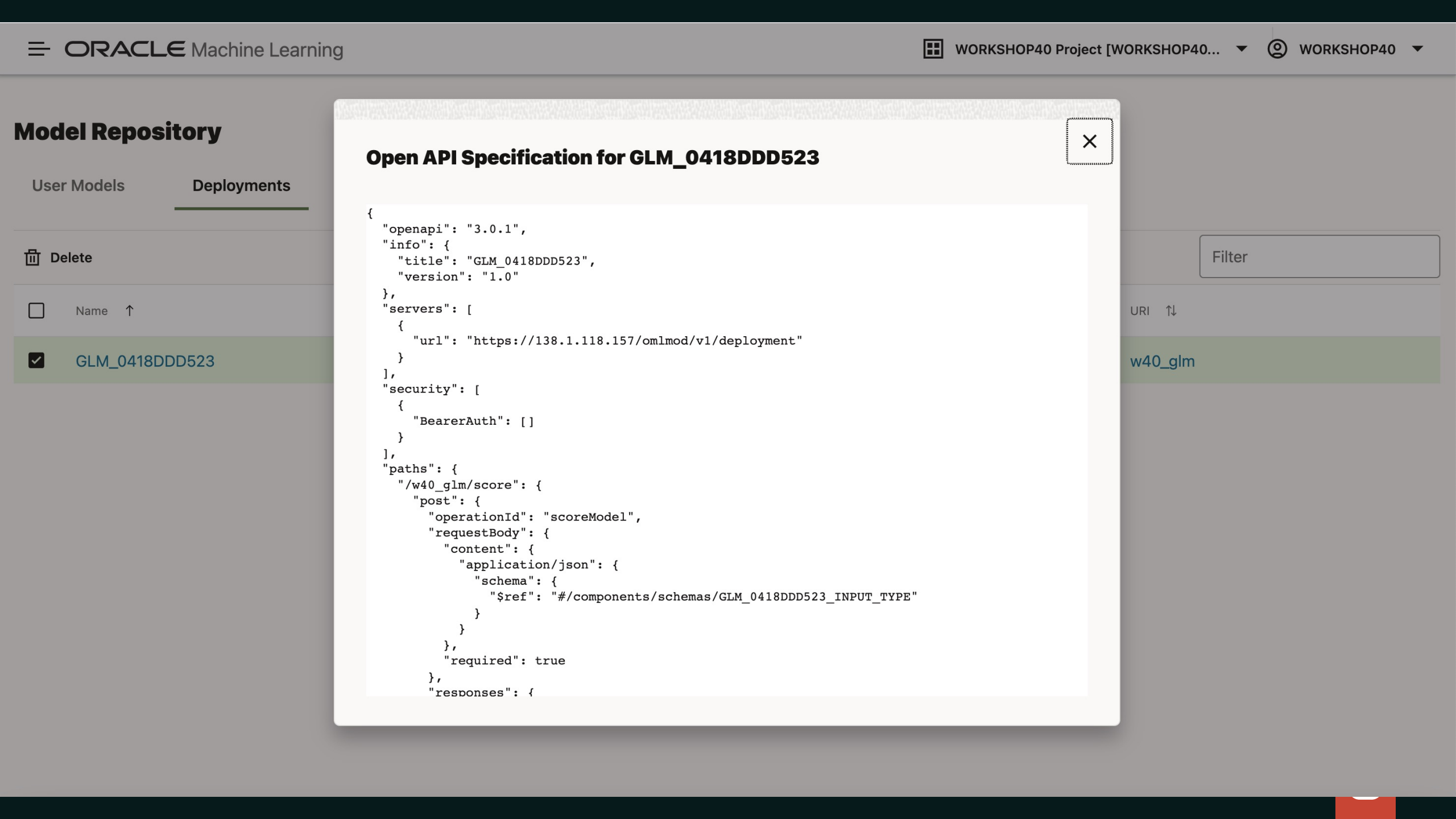
Model Repository

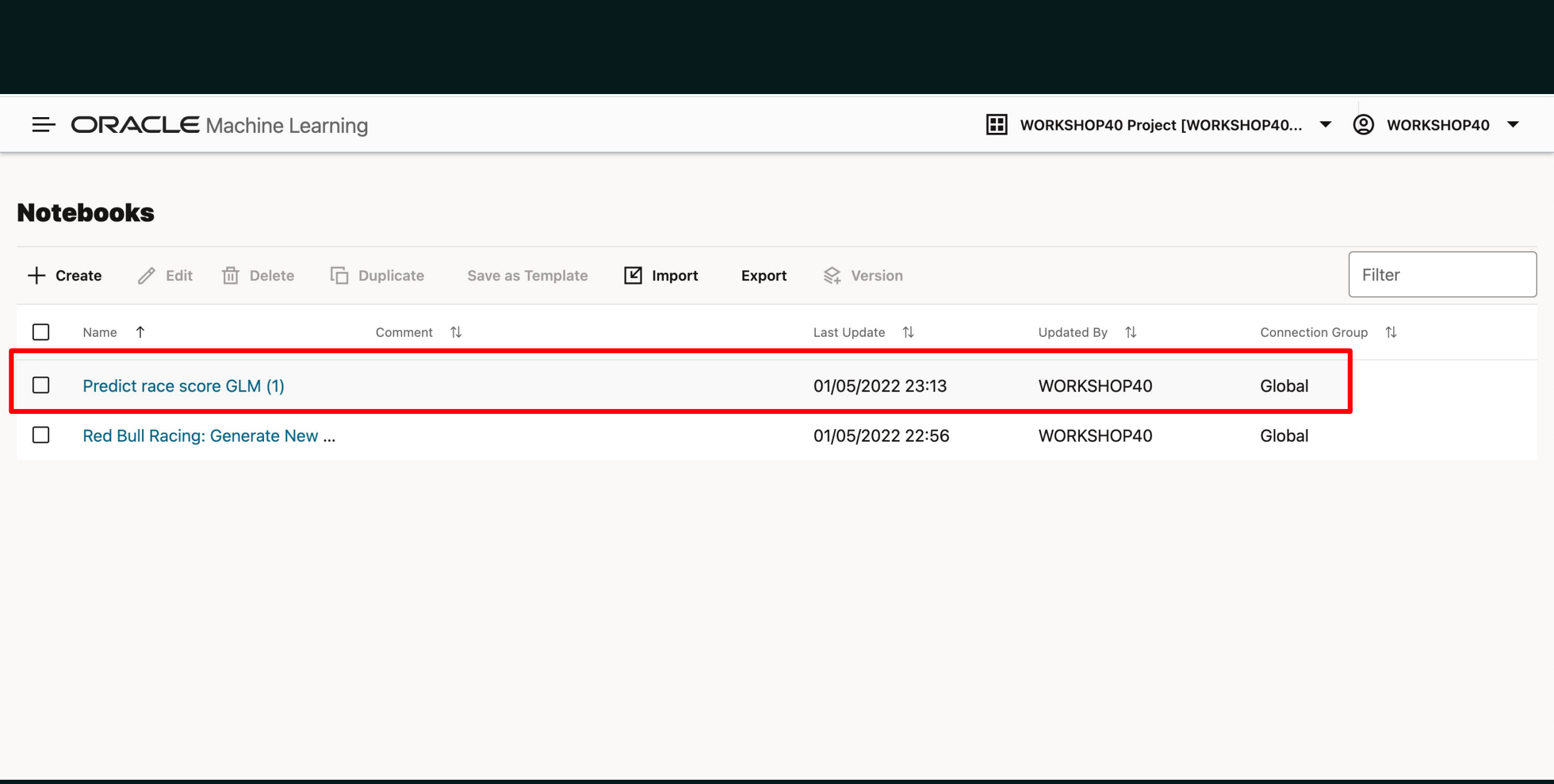
User Models | Deployments

Delete

Filter

Name	Shared	Version	Namespace	Owner	Deployed Date	URI
GLM_0418DDD523		1.0		WORKSHOP40	01/05/2022 ...	w40_glm


ORACLE Machine Learning
WORKSHOP40 Project [WORKSHOP40...
WORKSHOP40
Model Repository
User Models
Deployments
Delete
Filter
Name
URI
GLM_0418DDD523
w40_glm
Open API Specification for GLM_0418DDD523
{
"openapi": "3.0.1",
"info": {
"title": "GLM_0418DDD523",
"version": "1.0"
},
"servers": [
{
"url": "https://138.1.118.157/omlmod/v1/deployment"
}
],
"security": [
{
"BearerAuth": []
}
],
"paths": {
"/w40_glm/score": {
"post": {
"operationId": "scoreModel",
"requestBody": {
"content": {
"application/json": {
"schema": {
"$ref": "#/components/schemas/GLM_0418DDD523_INPUT_TYPE"
}
}
},
"required": true
},
"responses": {



Predict race score GLM (1)

Oracle Machine Learning AutoML UI - Experiment - Generated Notebook

READY

Get proxy object for selected data

READY

```
%python

import oml

columns = '"CIRCUITREF"' , '"SAFETY_CAR"' , '"OVERTAKEN_POSITIONS_REAL_TOTAL"' , '"LAPS_WITH_CHANGE_POS_1"' , '"LAPS_WITH_CHANGE_POS_1_5"' , '"DNF_COUNT"' , '"DNF_DUE_TO_ACCIDENT_COUNT"' ,
    '"MAX_COMEBACKSCORE"' , '"OVERTAKEN_POSITIONS_DUE_TO_PITSTOP_TOTAL"' , '"WEATHER_WET"' , '"RANK_VERSUS_POSITION"' , '"SCORE"' , '"RACEID"'
schema='"WORKSHOP40"'
table='"RACES_TRAIN"'

column = ','.join(columns)
query = 'SELECT ' + column + ' FROM ' + schema + '.' + table + ' where ' + '"SCORE"' + ' is not null'

build_data = oml.sync(query=query)
z.show(build_data)
```

Prepare training data

READY

```
%python

X_train = build_data.drop('SCORE')
y_train = build_data[:,'SCORE']
```

Build MODEL_NAME_TITLE model

READY

```
%python
```

Build MODEL_NAME_TITLE model

READY

```
%python

glm_settings = {
  'GLMS_NUM_ITERATIONS' : '30' , 'GLMS_RIDGE_REGRESSION' : 'GLMS_RIDGE_REG_DISABLE' , 'GLMS_SOLVER' : 'GLMS_SOLVER_CHOL' , 'ODMS_DETAILS' : 'ODMS_ENABLE'
  }

glm_mod = oml.glm('regression',**glm_settings)

glm_mod = glm_mod.fit(X_train, y_train , case_id = 'RACEID')
```

Show model details

READY

```
%python

glm_mod
```

Data for scoring

READY

```
%python

# using build data for prediction
```

Score data

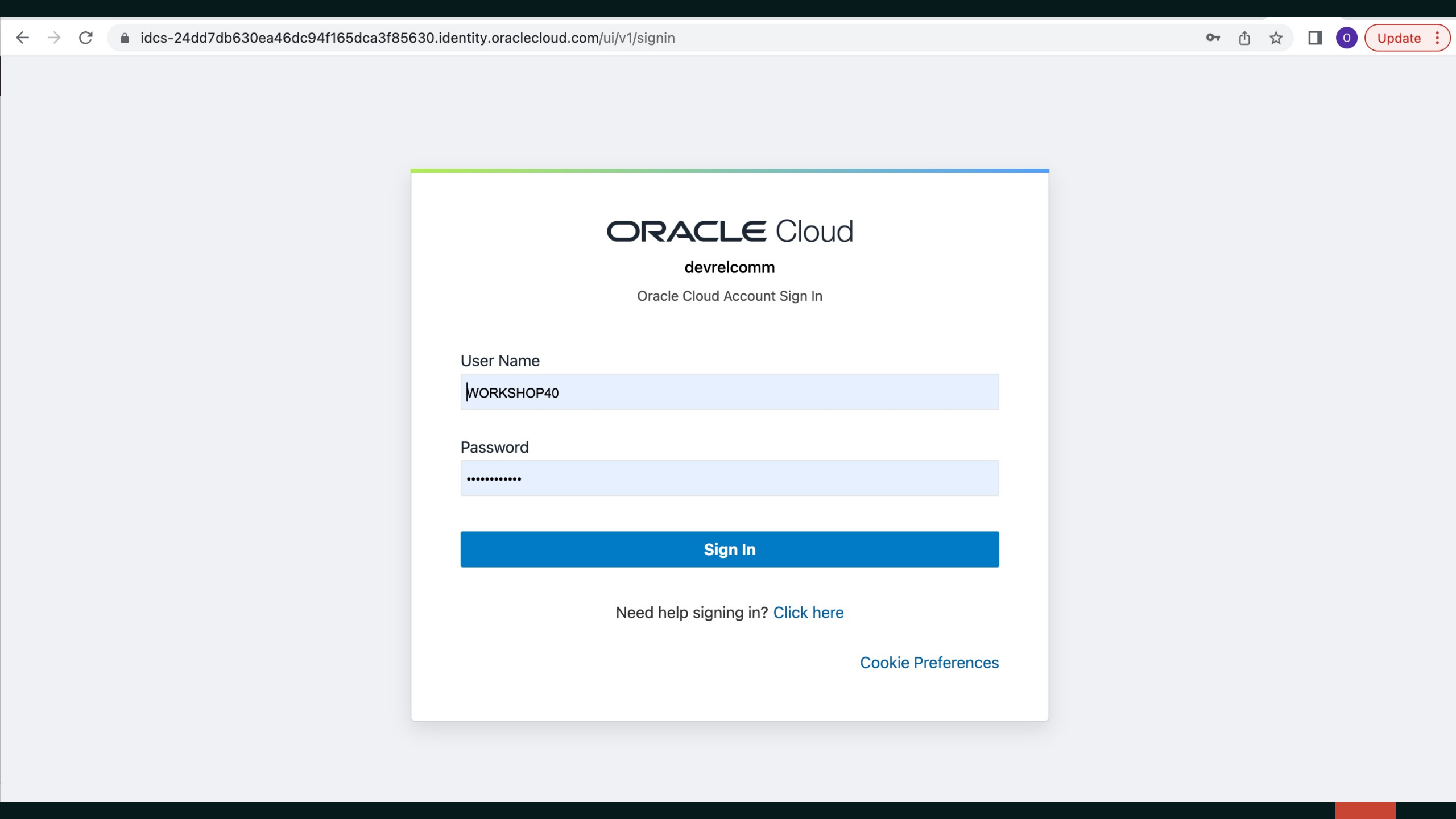
READY

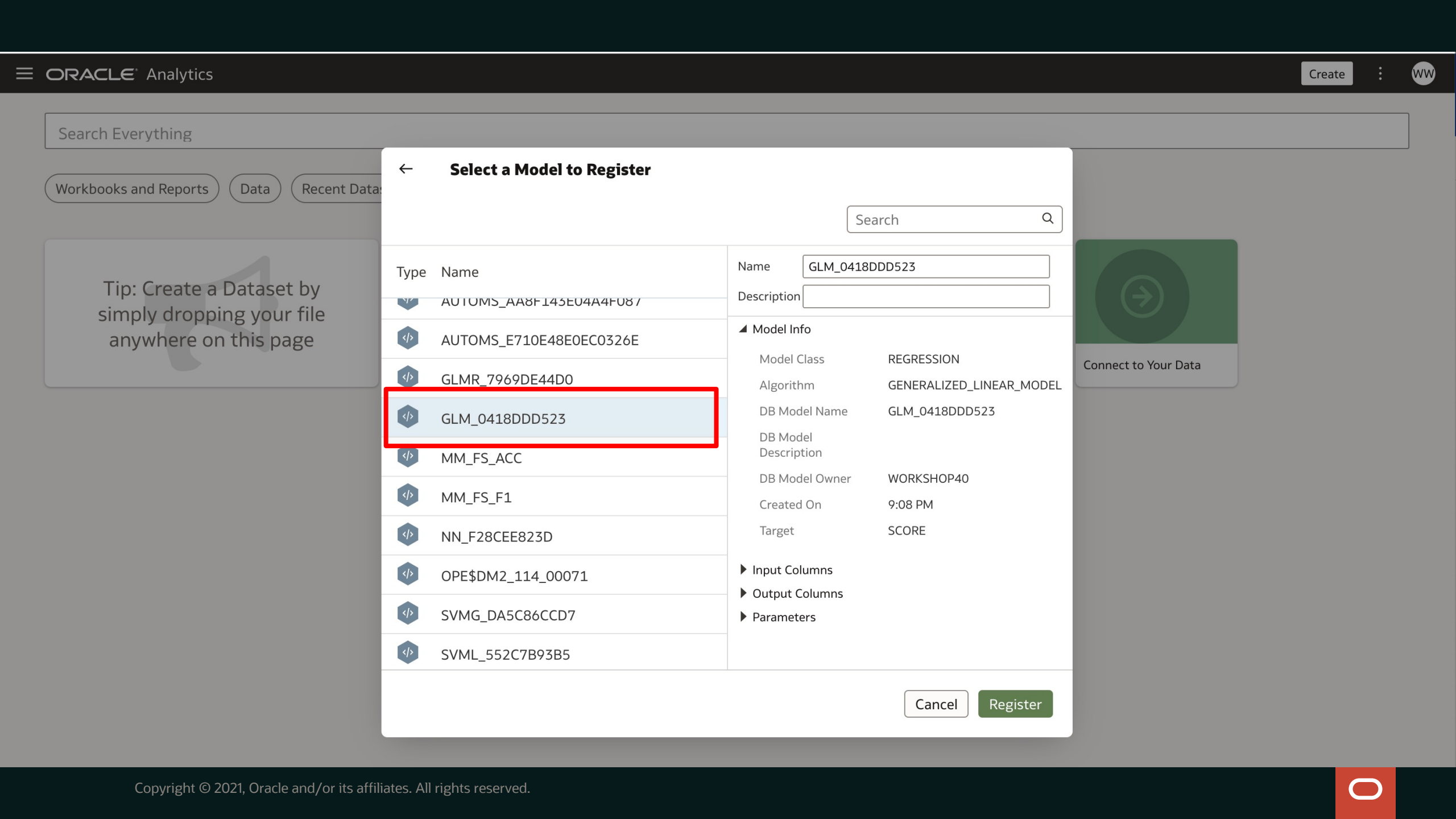
```
%python
mod_predict = glm_mod.predict(build_data ,supplemental_cols = build_data[:, ['SCORE']]).pull()
y_true = mod_predict['SCORE']
y_pred = mod_predict['PREDICTION']
```

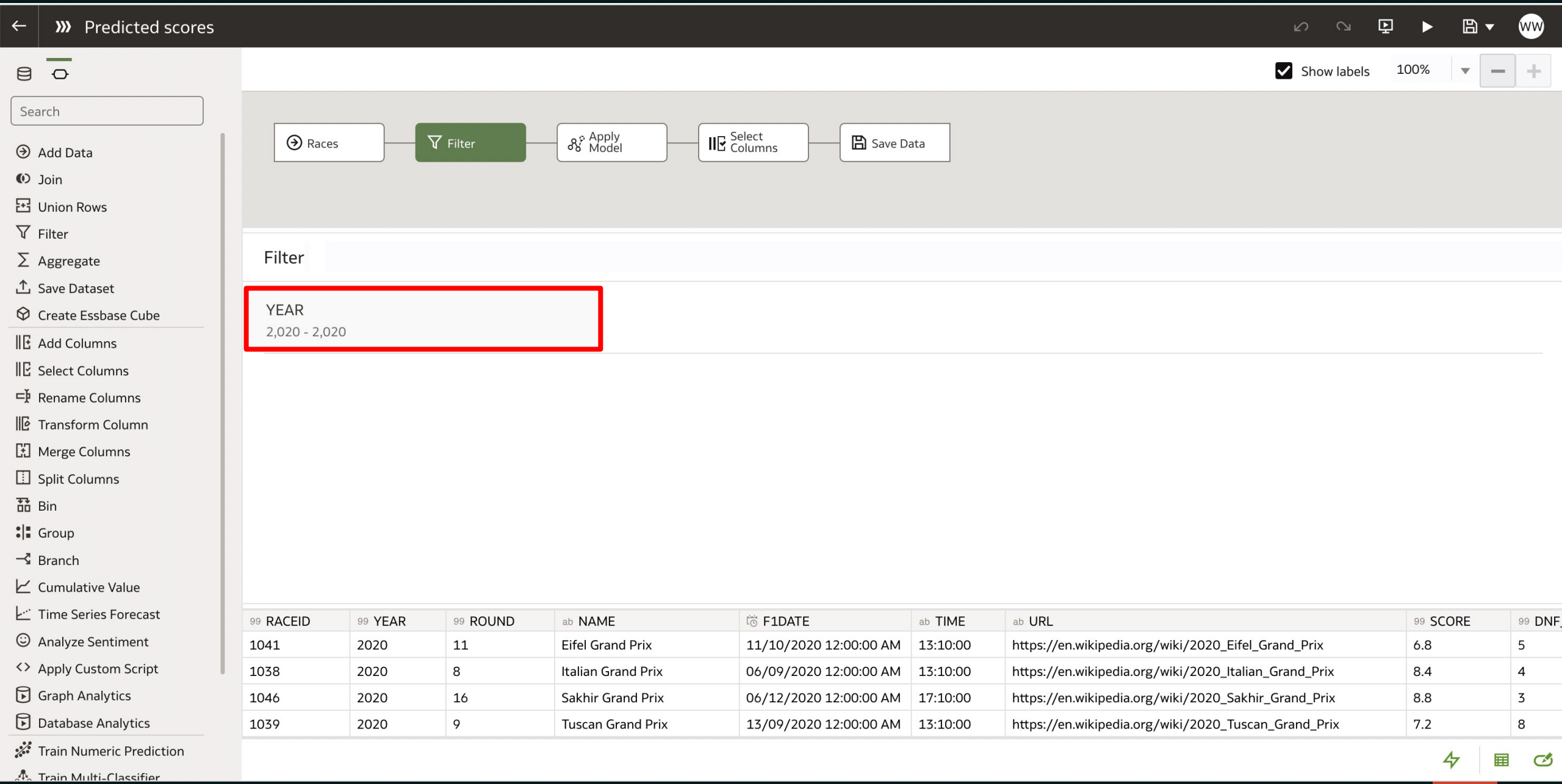
Show model quality metric

READY

```
%python
import sklearn as skl
metric_score = skl.metrics.mean_absolute_error(y_true, y_pred)
print(metric_score.round(4))
```


idcs-24dd7db630ea46dc94f165dca3f85630.identity.oraclecloud.com/ui/v1/signin
Update
ORACLE Cloud
devrelcomm
Oracle Cloud Account Sign In
User Name
WORKSHOP40
Password
Sign In
Need help signing in? Click here
Cookie Preferences





Check prediction accuracy

Data | Visualize | Present

Search

- Predicted scores
  - RACEID
  - NAME_YEAR
  - SCORE
  - Prediction
- My Calculations
- Value Labels

SCORE, Prediction by R...

Title: Auto

Title Tooltip: Auto

Title Font: Auto

Line Type: Straight

Click here or drag data to add a filter

Line

Trellis Columns

Trellis Rows

Values (Y-Axis)
- SCORE
- Prediction

Category (X-Axis)
- RACEID
- NAME_YEAR

Color

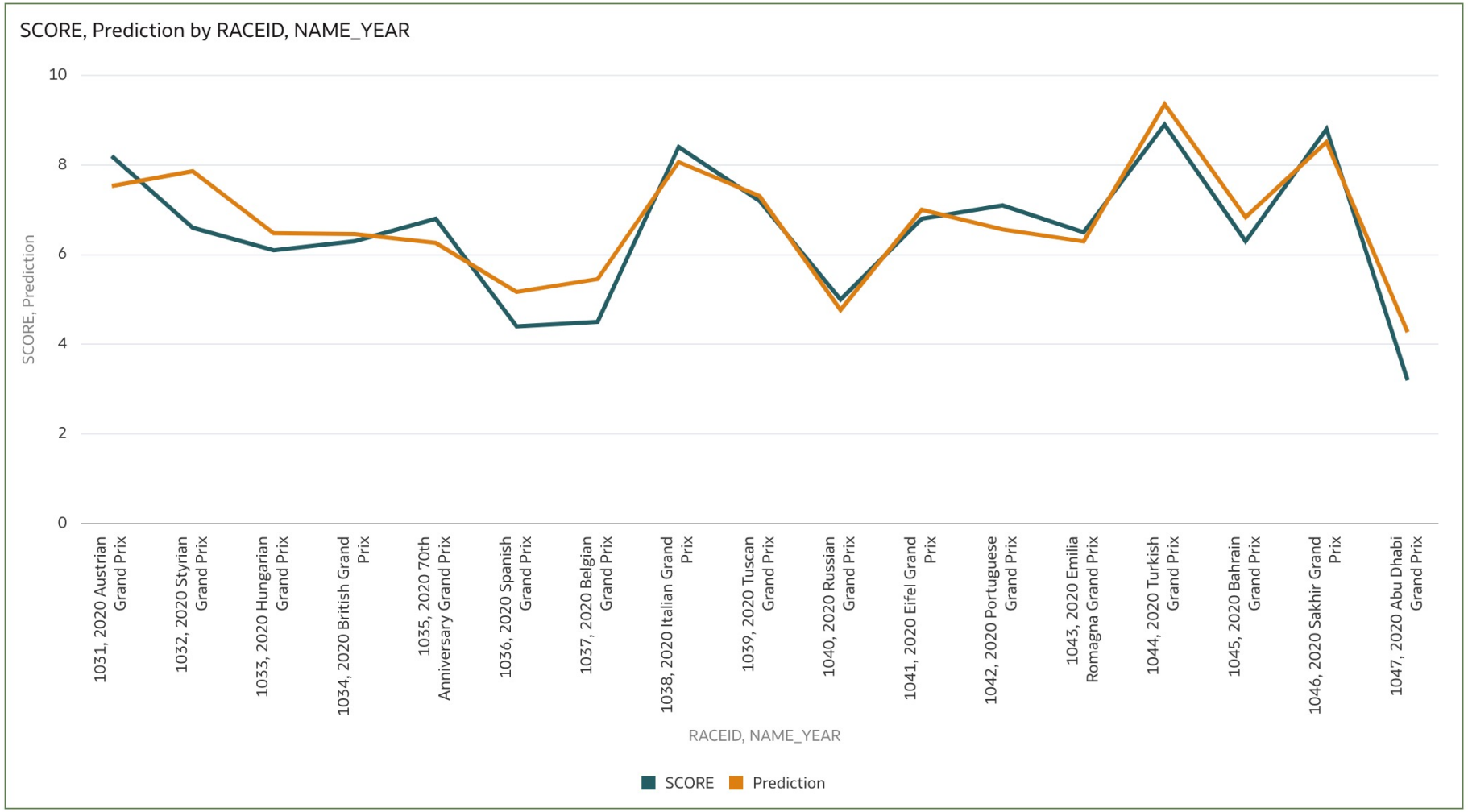
Size (Width)

Shape

Tooltip

Detail

Filters



Canvas 1

17 Categories, 2 Lines

Agenda

Presentation	Introduction to the challenge + ML essentials
Lab 1	Loading the data
Lab 2	Basic data exploration
Lab 3	Advanced data exploration
Lab 4	Feature generation, machine learning, evaluation
Lab 5	Process F1 2022 Dataset from racefans.net

Lab 5: Process to execute same model for F1 2021 Fans

In this lab we will finally get to the actual Machine Learning part!

1: Check in the database dataset for 2021

https://www.racefans.net/2021/12/25/the-f1-races-of-2021-ranked-by-racefans-readers-including-the-lowest-score-ever/

https://www.racefans.net/rate-the-race/f1-fanatic-top-100/

2: How to Update dataset

3: Predict the scores for 2021

4: Evaluate the accuracy of the model for 2021

Lab instructions
https://bit.ly/jfokus22-f1

Navigator | Worksheets

WORKSHOP40

Tables

Search...

- DM$P5GLMR_7969DE44D0
- DM$P5GLM_0418DDD523
- DM$P5NN_F28CEE823D
- DM$P5OPE$DM2_114_00071
- DM$P5SVMG_DA5C86CCD7
- DM$P5SVML_552C7B93B5
- DM$PBGLMR_7969DE44D0
- DM$PBGLM_0418DDD523
- DM$PBOPE$DM2_114_00071
- DM$PCGLMR_7969DE44D0
- DM$PCGLM_0418DDD523
- DM$PCOPE$DM2_114_00071
- DM$PDGLMR_7969DE44D0
- DM$PDGLM_0418DDD523
- DM$PDNN_F28CEE823D
- DM$PDOPE$DM2_114_00071
- DM$PDSVMG_DA5C86CCD7
- DM$PDSVML_552C7B93B5
- DM$PEGLMR_7969DE44D0
- DM$PEGLM_0418DDD523

[Worksheet]*

```
1  select * from RACES WHERE YEAR = 2021;
```

Query Result | Script Output | DBMS Output | Explain Plan | Autotrace | SQL History | Data Loading

Download | Execution time: 0.013 seconds

	raceid	year	round	name	f1date	time	url	scor
1	1064	2021	13	Dutch Grand Prix	9/5/2021, 12:00:00	13:00:00	http://en.wikipedia.	
2	1071	2021	19	Brazilian Grand Pri	11/14/2021, 12:00:(	17:00:00	http://en.wikipedia.	
3	1059	2021	7	French Grand Prix	6/20/2021, 12:00:0	13:00:00	http://en.wikipedia.	
4	1061	2021	10	British Grand Prix	7/18/2021, 12:00:0(	14:00:00	http://en.wikipedia.	
5	1065	2021	14	Italian Grand Prix	9/12/2021, 12:00:0	13:00:00	http://en.wikipedia.	
6	1070	2021	18	Mexican Grand Pri	11/7/2021, 12:00:0(	19:00:00	http://en.wikipedia.	
7	1066	2021	15	Russian Grand Prix	9/26/2021, 12:00:0	12:00:00	http://en.wikipedia.	
8	1055	2021	4	Spanish Grand Prix	5/9/2021, 12:00:00	13:00:00	http://en.wikipedia.	
9	1067	2021	16	Turkish Grand Prix	10/10/2021, 12:00:(	12:00:00	http://en.wikipedia.	

Navigator | Worksheets

WORKSHOP40

Tables

Search...

- DM$P5GLMR_7969DE44D0
- DM$P5GLM_0418DDD523
- DM$P5NN_F28CEE823D
- DM$P5OPE$DM2_114_00071
- DM$P5SVMG_DA5C86CCD7
- DM$P5SVML_552C7B93B5
- DM$PBGLMR_7969DE44D0
- DM$PBGLM_0418DDD523
- DM$PBOPE$DM2_114_00071
- DM$PCGLMR_7969DE44D0
- DM$PCGLM_0418DDD523
- DM$PCOPE$DM2_114_00071
- DM$PDGLMR_7969DE44D0
- DM$PDGLM_0418DDD523
- DM$PDNN_F28CEE823D
- DM$PDOPE$DM2_114_00071
- DM$PDSVMG_DA5C86CCD7
- DM$PDSVML_552C7B93B5
- DM$PEGLMR_7969DE44D0
- DM$PEGLM_0418DDD523

[Worksheet]*

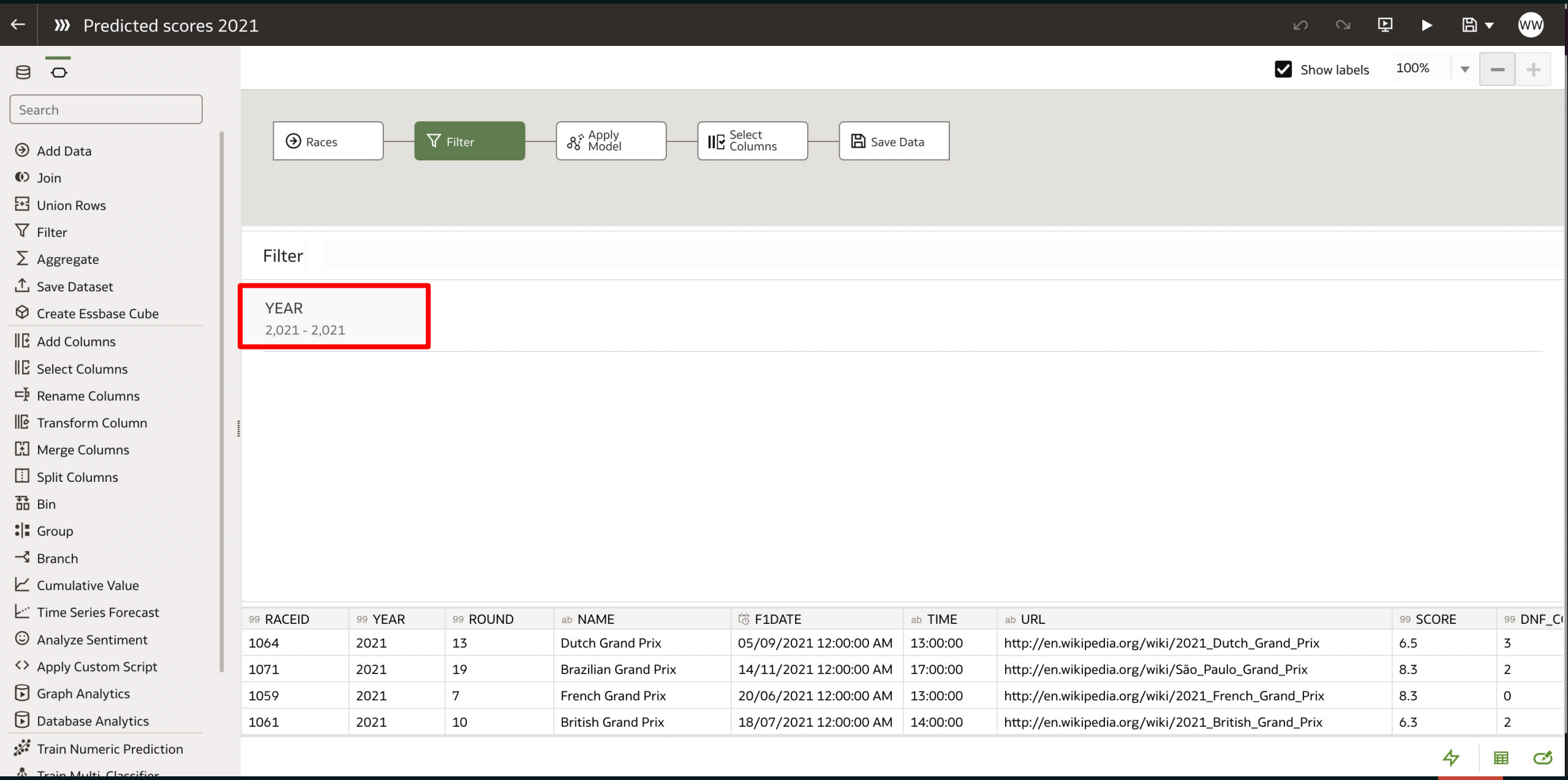
```
1  select * from RACES WHERE YEAR = 2021;
```

Query Result | Script Output | DBMS Output | Explain Plan | Autotrace | SQL History | Data Loading

Download

Execution time: 0.013 seconds

		round	name	f1date	time	url	score	dnf_count
1	2021	13	Dutch Grand Prix	9/5/2021, 12:00:00	13:00:00	http://en.wikipedia.	(null)	
2	2021	19	Brazilian Grand Pri	11/14/2021, 12:00:(	17:00:00	http://en.wikipedia.	(null)	
3	2021	7	French Grand Prix	6/20/2021, 12:00:0	13:00:00	http://en.wikipedia.	(null)	
4	2021	10	British Grand Prix	7/18/2021, 12:00:0(	14:00:00	http://en.wikipedia.	(null)	
5	2021	14	Italian Grand Prix	9/12/2021, 12:00:0	13:00:00	http://en.wikipedia.	(null)	
6	2021	18	Mexican Grand Pri	11/7/2021, 12:00:0(	19:00:00	http://en.wikipedia.	(null)	
7	2021	15	Russian Grand Prix	9/26/2021, 12:00:0	12:00:00	http://en.wikipedia.	(null)	
8	2021	4	Spanish Grand Pri	5/9/2021, 12:00:00	13:00:00	http://en.wikipedia.	(null)	
9	2021	16	Turkish Grand Prix	10/10/2021, 12:00:(	12:00:00	http://en.wikipedia.	(null)	



Check prediction accuracy 2021

Data | Visualize | Present

Search

Predicted scores 2021
- RACEID
- NAME_YEAR
- SCORE
- Prediction

My Calculations

Value Labels

Click here or drag data to add a filter

Line

Trellis Columns

Trellis Rows

Values (Y-Axis)
- SCORE
- Prediction

Category (X-Axis)
- RACEID
- NAME_YEAR

Color

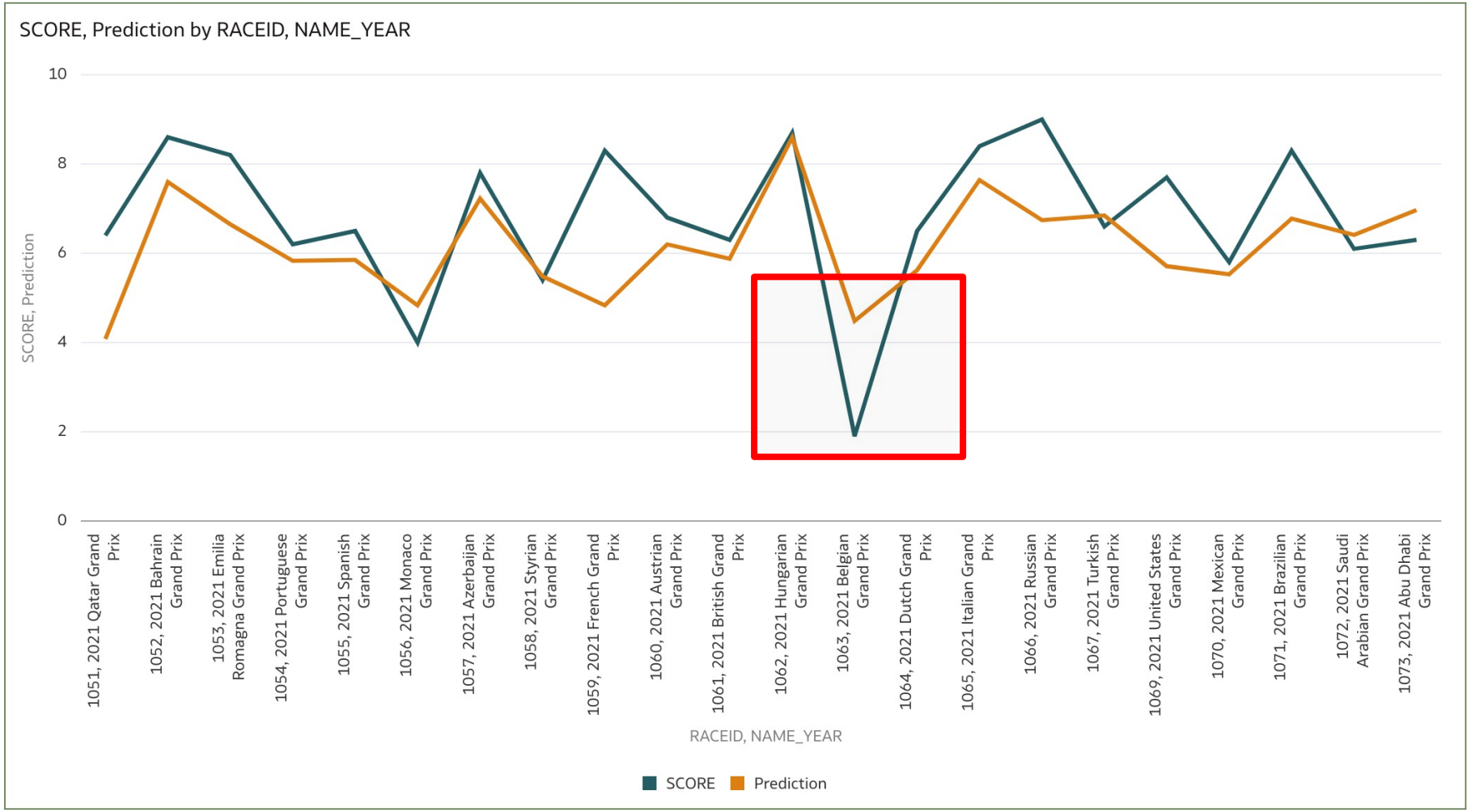
Size (Width)

Shape

Tooltip

Detail

Filters



SCORE, Prediction by R...

Title	Auto
Title Tooltip	Auto
Title Font	Auto
Line Type	Straight

Canvas 1

22 Categories, 2 Lines

Developer Masterclass Badge:

Take a short quiz on the topics covered in today's session and earn a badge to share to your social media.

Access the quiz here: **bit.ly/quiz_redbull_JFOKUS22**

This workshop is part of the **Modern AppDev Masterclass** series, and each session will have it's own badge where you can earn the title of **Developer Pioneer** recognizing your commitment and earned expertise.

Thank You


ORACLE