ORACLE

Welcome at VidaCaixa 2022 Hands-on Lab

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

Objetivos del workshop

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.

Workshop Detailed:

https://github.com/operard/vidacaixa2022

Lab instructions

https://bit.ly/vidacaixa2022

Organización del workshop

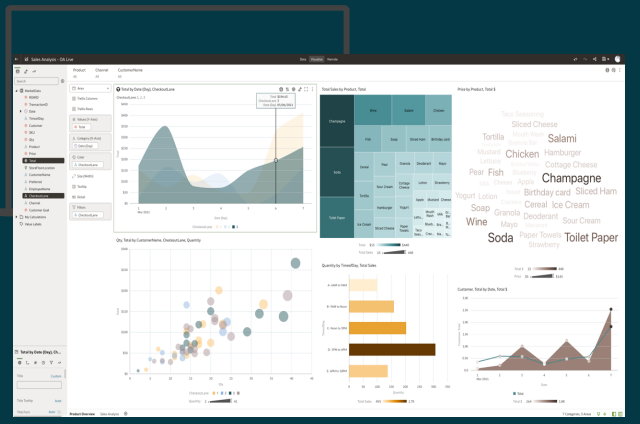
Data Engineers

Data Loading

Data Scientists

Model training

Data Visualization



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

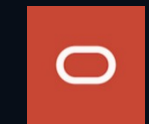
Users

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

	Data Engineers	Cloud USER (OAC)	PASS	ADB USER	PASS
Paula Álvarez González	palvarez@vidacaixa.es	palvarez	Vidacaixa2022$$	WORKSHOP1	Oracle_12345
Alberto Tejedor González	atejedor@vidacaixa.es	atejedor	Vidacaixa2022$$	WORKSHOP2	Oracle_12345
Carlos Almansa Rosales	calmansa@vidacaixa.es	calmansa	Vidacaixa2022$$	WORKSHOP3	Oracle_12345
Cristina Eslava Damea	ceslava@vidacaixa.es	ceslava	Vidacaixa2022$$	WORKSHOP4	Oracle_12345
David Lencina Hors	dlencina@vidacaixa.es	dlencina	Vidacaixa2022$$	WORKSHOP5	Oracle_12345
Isabel Gil Rodríguez	igil@vidacaixa.es	igil	Vidacaixa2022$$	WORKSHOP6	Oracle_12345
Jordi Madrenas Ciurana	jmadrenas@vidacaixa.es	jmadrenas	Vidacaixa2022$$	WORKSHOP7	Oracle_12345
Josep López Pérez	jlopez@vidacaixa.es	jlopez	Vidacaixa2022$$	WORKSHOP8	Oracle_12345
Manuel Luna Feijoo	mluna@vidacaixa.es	mluna	Vidacaixa2022$$	WORKSHOP9	Oracle_12345
Marina Barrios Jurado	mbarrios@vidacaixa.es	mbarrios	Vidacaixa2022$$	WORKSHOP10	Oracle_12345
Carmen Gimeno Olmos	carmen.gimeno@vidacaixa.es	carmen.gimeno	Vidacaixa2022$$	WORKSHOP11	Oracle_12345
Roberto Lorenzo	rlorenzo@vidacaixa.es	rlorenzo	Vidacaixa2022$$	WORKSHOP12	Oracle_12345
	Data Scientists				
Jordi Vidal Araujo	jvidal@vidacaixa.es	jvidal	Vidacaixa2022$$	WORKSHOP1	Oracle_12345
Christian Castells Gort	ccastells@vidacaixa.es	ccastells	Vidacaixa2022$$	WORKSHOP2	Oracle_12345
David Lozano Jiménez	dlozano@vidacaixa.es	dlozano	Vidacaixa2022$$	WORKSHOP3	Oracle_12345
Iñaki García Galera	igarcia@vidacaixa.es	igarcia	Vidacaixa2022$$	WORKSHOP4	Oracle_12345
Lluís Riudavets Martí	lriudavets@vidacaixa.es	lriudavets	Vidacaixa2022$$	WORKSHOP5	Oracle_12345
Marta Castellano Rodríguez	mcastellano@vidacaixa.es	mcastellano	Vidacaixa2022$$	WORKSHOP6	Oracle_12345
Marta Garcia Torres	mgarciat@vidacaixa.es	mgarciat	Vidacaixa2022$$	WORKSHOP7	Oracle_12345
Raúl Hernández Marruedo	rhernandez@vidacaixa.es	rhernandez	Vidacaixa2022$$	WORKSHOP8	Oracle_12345
Arnau Colomer Domingo	acolomer@proveedorexterno.vidacaixa.es	acolomer	Vidacaixa2022$$	WORKSHOP9	Oracle_12345
Esther Mayol Puyol	emayol@vidacaixa.es	emayol	Vidacaixa2022$$	WORKSHOP10	Oracle_12345
	Data Visualizers				
Francesc Massip Sabaté	fmassip@vidacaixa.es	fmassip	Vidacaixa2022$$	WORKSHOP1	Oracle_12345
Juan Antonio Millán Méndez	jamillan@vidacaixa.es	jamillan	Vidacaixa2022$$	WORKSHOP2	Oracle_12345
Lara Bejarano Martín	lbejarano@vidacaixa.es	lbejarano	Vidacaixa2022$$	WORKSHOP3	Oracle_12345
Cristina Bellido Andújar	cbellido@vidacaixa.es	cbellido	Vidacaixa2022$$	WORKSHOP4	Oracle_12345
Silvia Galup Russines	sgalup@vidacaixa.es	sgalup	Vidacaixa2022$$	WORKSHOP5	Oracle_12345

How to use the Formula 1 Data…



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

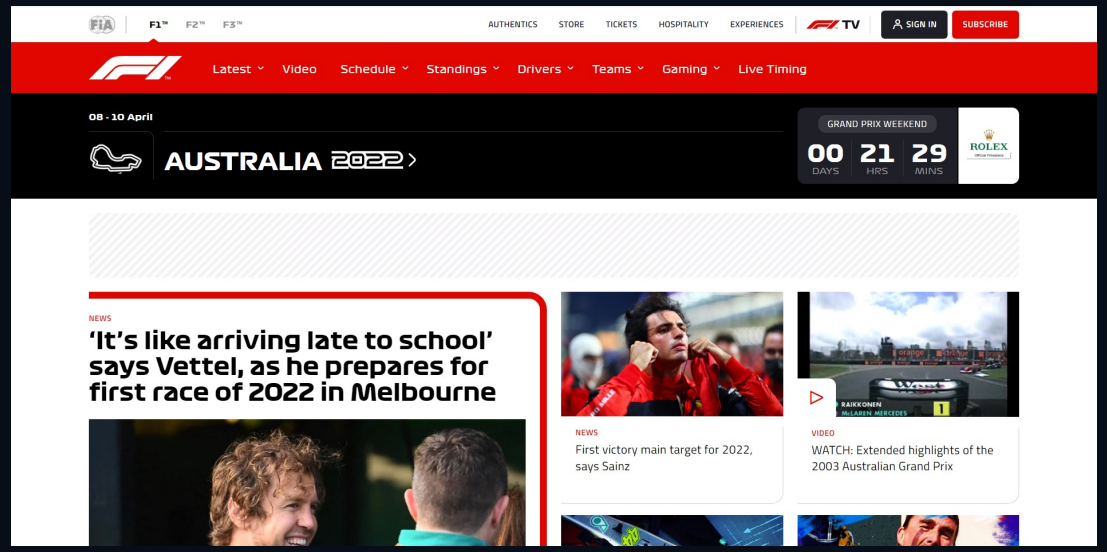
PHOTO CREDIT: Red Bull Content Pool

Formula 1 Data

ORACLE Red Bull RACING

CLOUD PARTNER

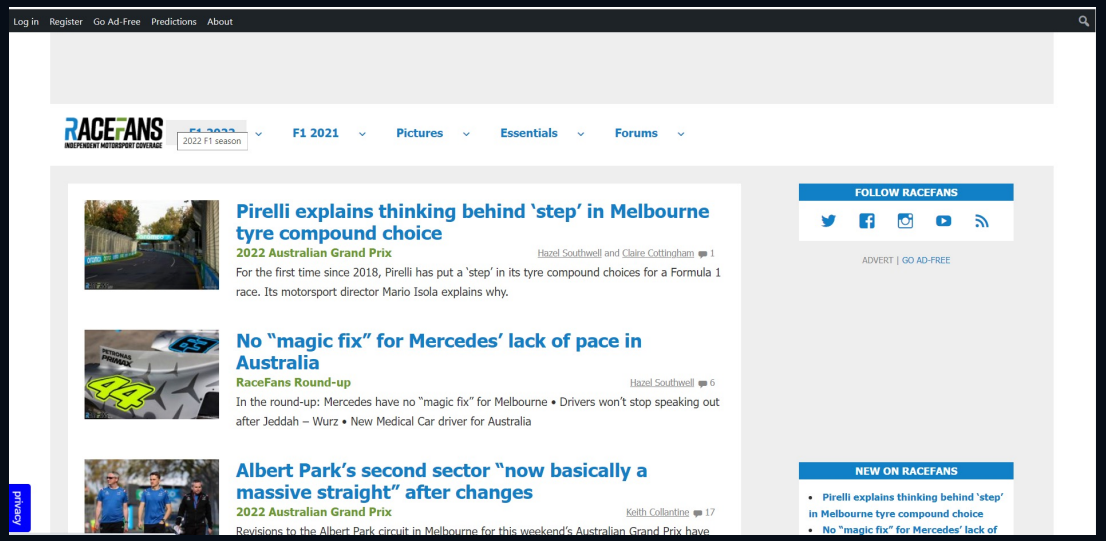
https://www.formula1.com/



DATA SETS

- Driver's & Constructor's historical lap time and competitive super times
- Circuit Performance
- Practice (FP1, FP2, FP3)
- Qualifying Data
- Race Data

https://www.racefans.net/



DATA SETS

- Qualifying Fans Data
- Race Fans Data

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

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

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

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 1: Loading the data

¿Quien tiene este rol?

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/vidacaixa2022

Lab 1: Loading the data

Rol: Data Engineer

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/vidacaixa2022

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

¿Quien tiene este rol?

Explore the data and look for clues in the data that are of value to predict the SCORE of a race.

Exploratory Data Analysis refers to the critical process of performing initial investigations on data so as to discover patterns, to spot anomalies, to test hypothesis and to check assumptions with the help of summary statistics and graphical representations.

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

Lab instructions

https://bit.ly/vidacaixa2022

Lab 2: Basic Data Exploration

Rol: Data Scientist

Using Notebooks + JupyterLab.

Explore the data and look for clues in the data that are of value to predict the SCORE of a race.

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

Lab instructions
https://bit.ly/vidacaixa2022

Rol: Data Visualizer

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/vidacaixa2022

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

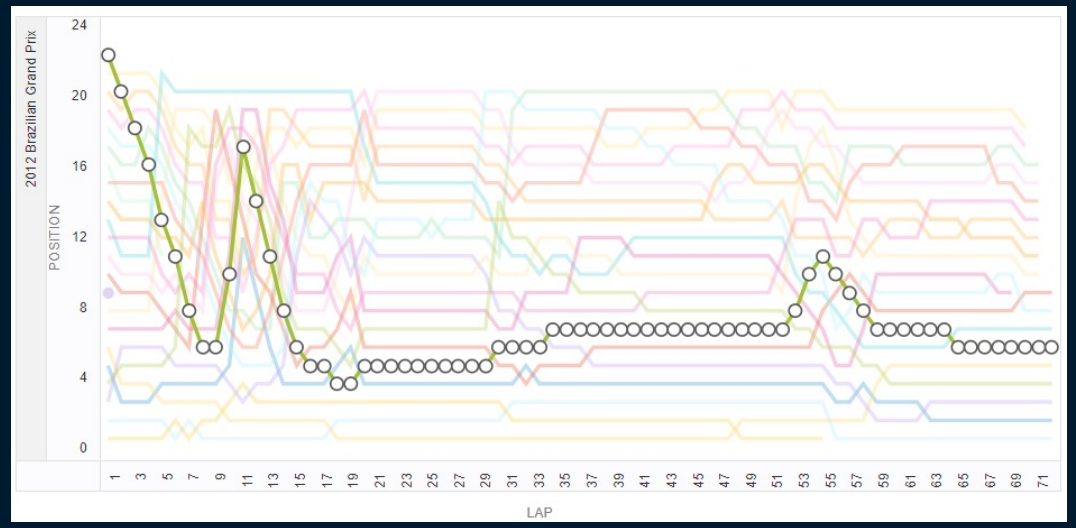
Lab 3: Advanced Data Exploration

¿Quien tiene este rol?

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/vidacaixa2022

Lab 3: Advanced Data Exploration

Rol: Data Scientist

Using Notebooks + JupyterLab

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/vidacaixa2022

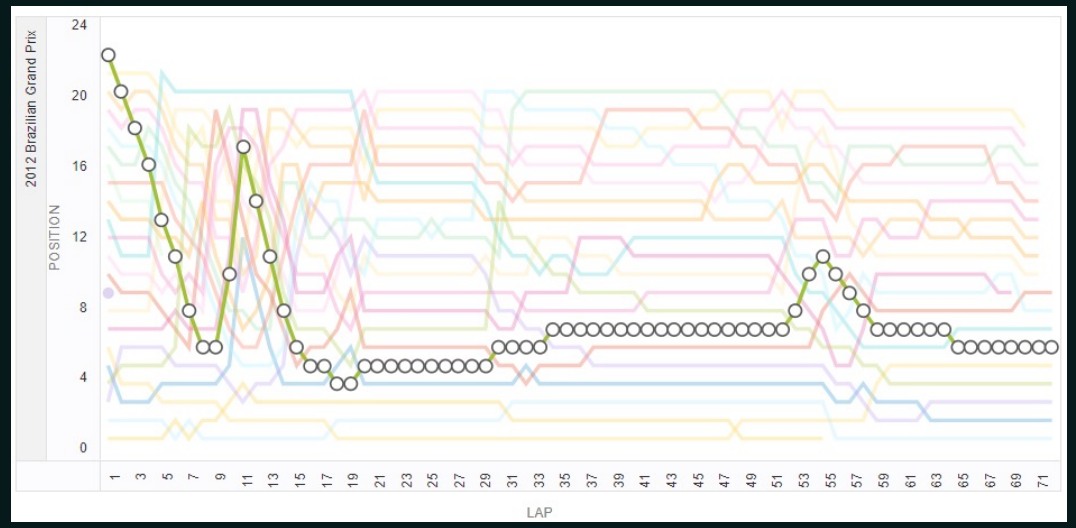
Lab 3: Advanced Data Exploration

Rol: Data Visualizer

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/vidacaixa2022

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

¿Quien tiene este rol?

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/vidacaixa2022

Lab 4: Feature Generation, ML Training and Evaluation

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

1: Feature Generation → Rol: Data Engineer

2: Train the ML model

3: Predict the scores

4: Evaluate the accuracy of the model

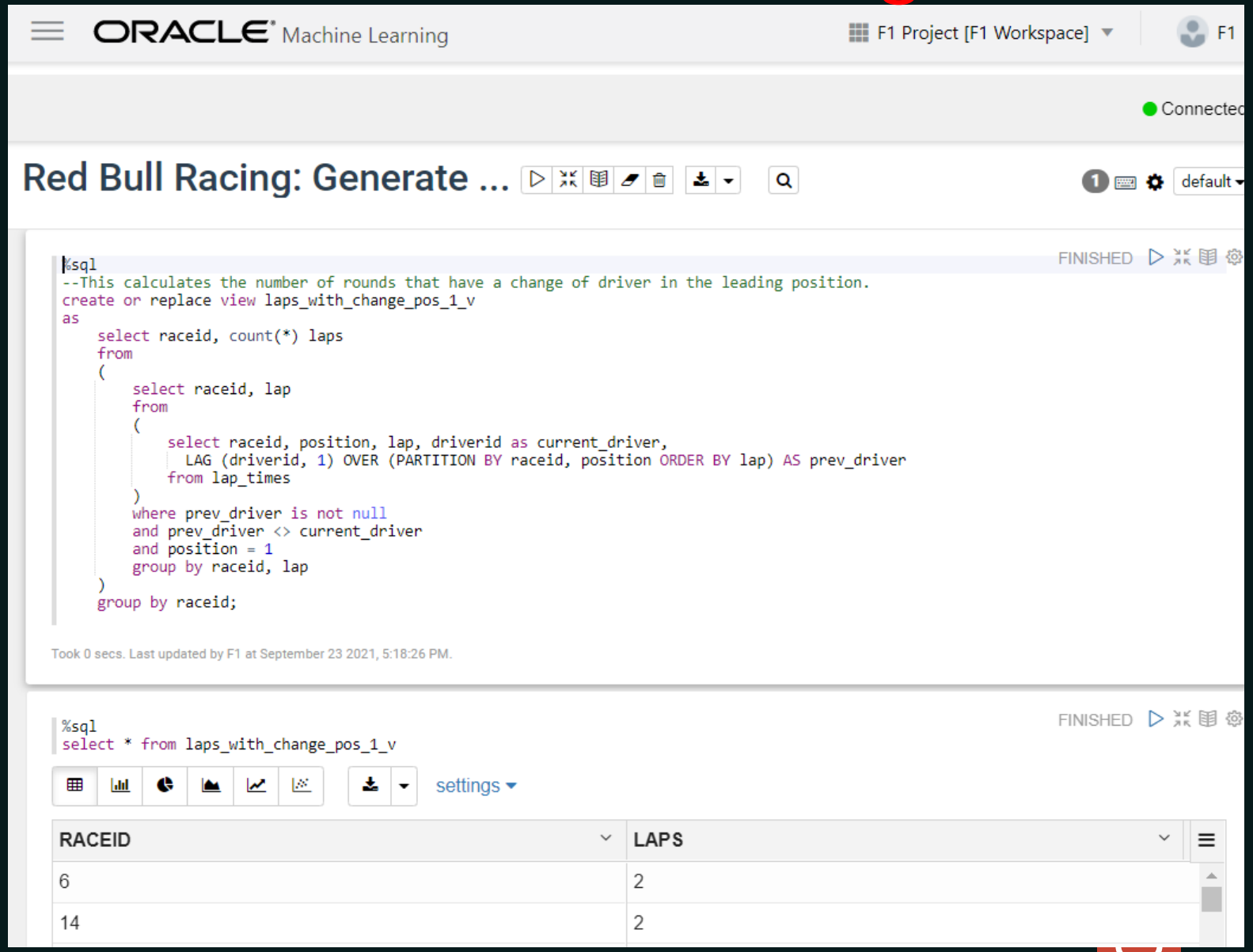
Rol: Data Scientist

Lab instructions
https://bit.ly/vidacaixa2022

Task 1: Feature Generation

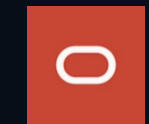
Rol: Data Engineer

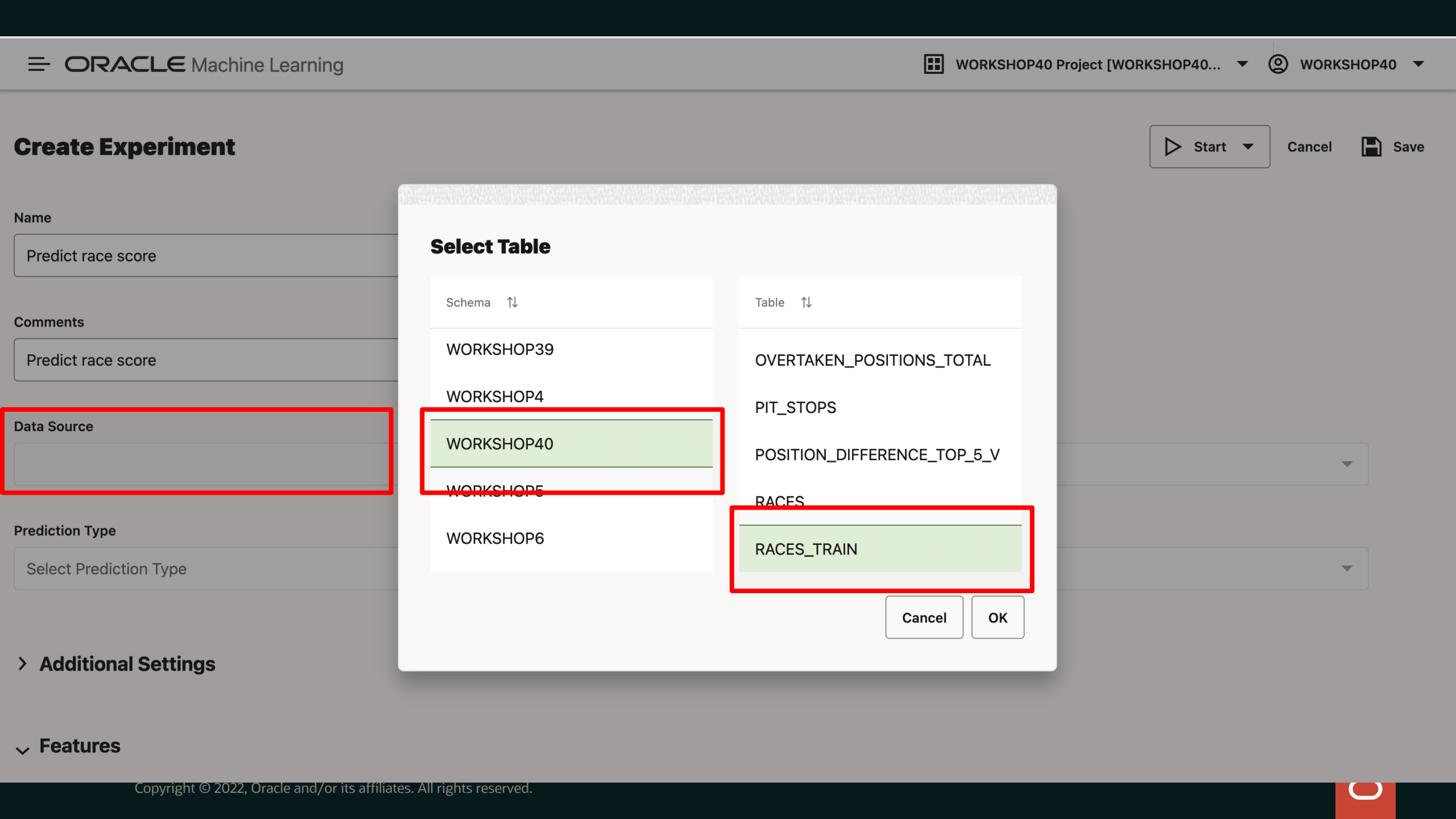
1. Using Notebook
2. Using SQL Script



Rol: Data Scientist

Task 2: Using ML Algorithms …





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

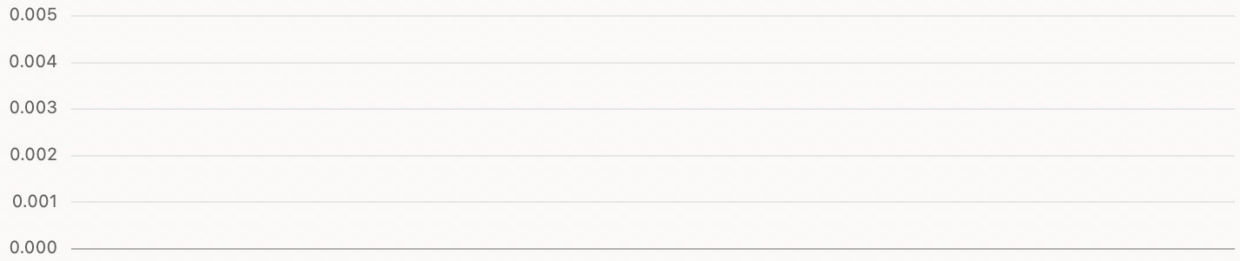
<- Experiments

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

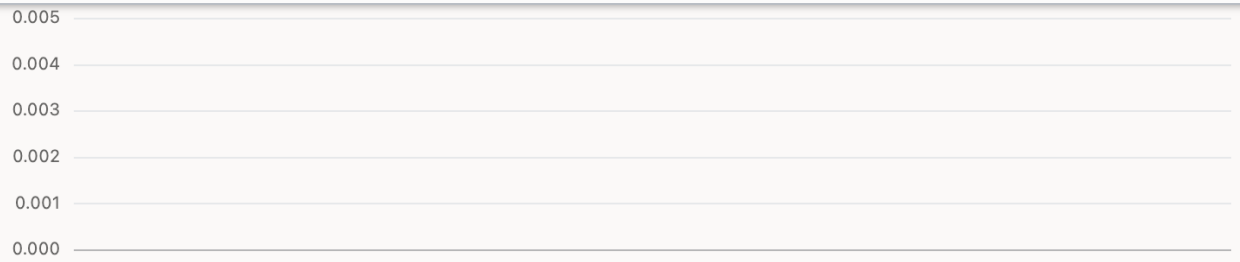
Leader Board

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

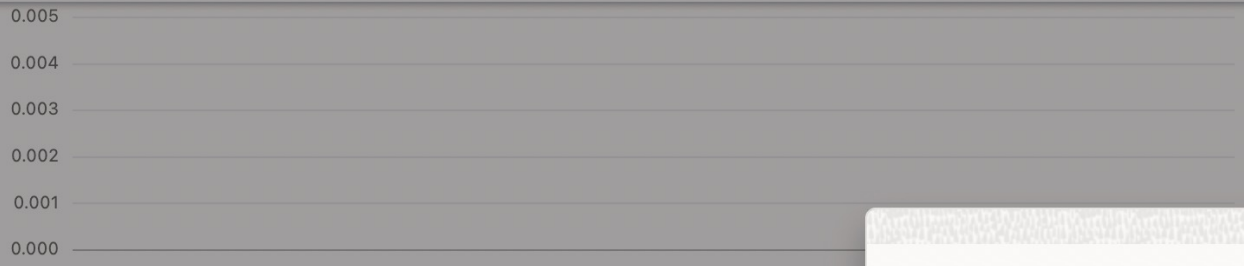
- Initialization — Completed
- Algorithm Selection — Completed
- Adaptive Sampling — Completed
- Feature Selection — Completed
- Model Tuning — Completed
- Generalized Linear Model — Completed



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
Neural Network	NN_F28CEE823D
Support Vector Machine (Linear)	SVML_552C7B93B5
Generalized Linear Model (Ridge ...	GLMR_7969DE44D0
Support Vector Machine (Gaussi...	SVMG_DA5C86CCD7
Generalized Linear Model	GLM_0418DDD523

Select Additional Metrics

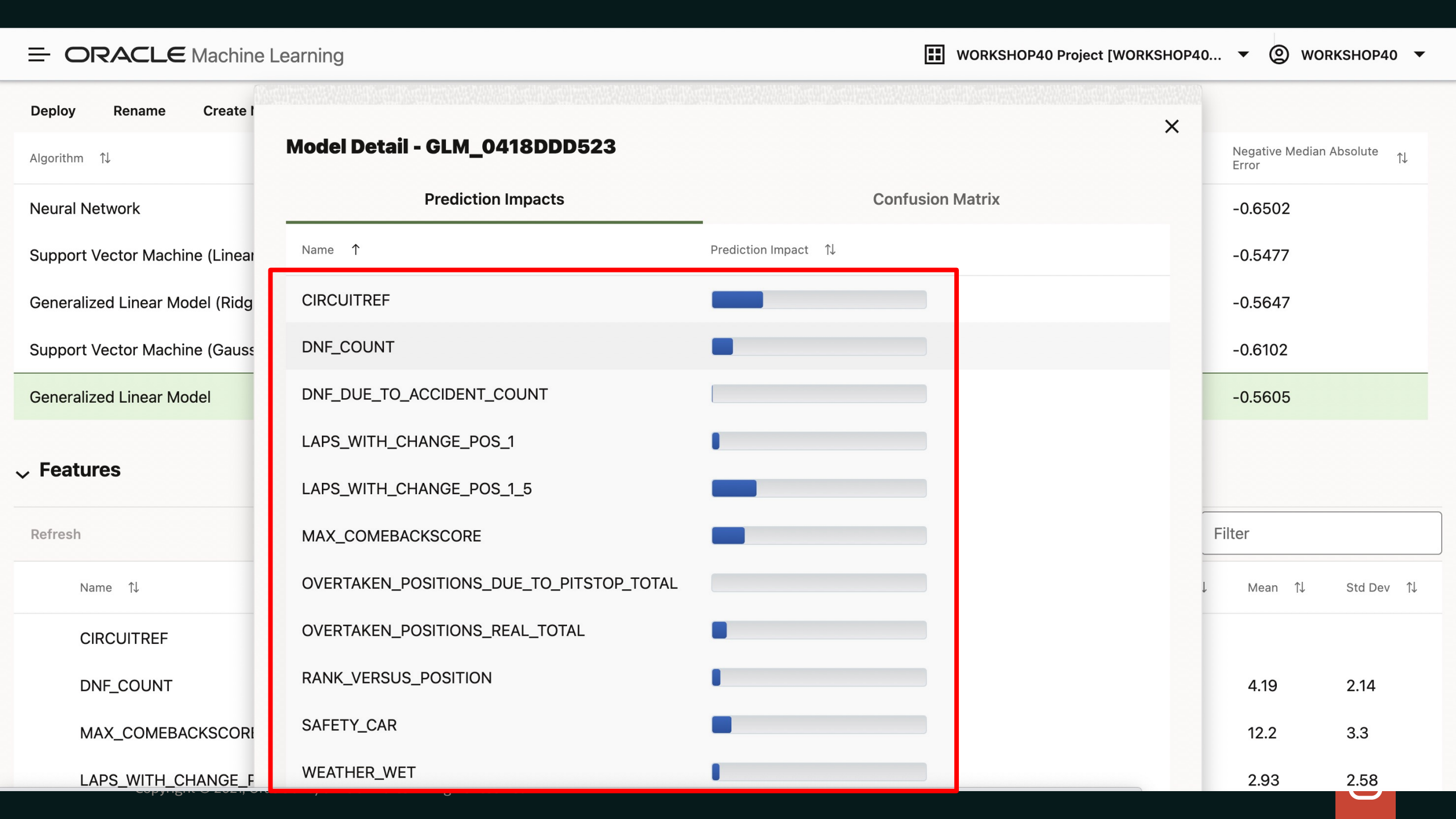
- [ ] Negative Mean Squared Error
- [ ] R2
- [ ] Negative Median Absolute Error

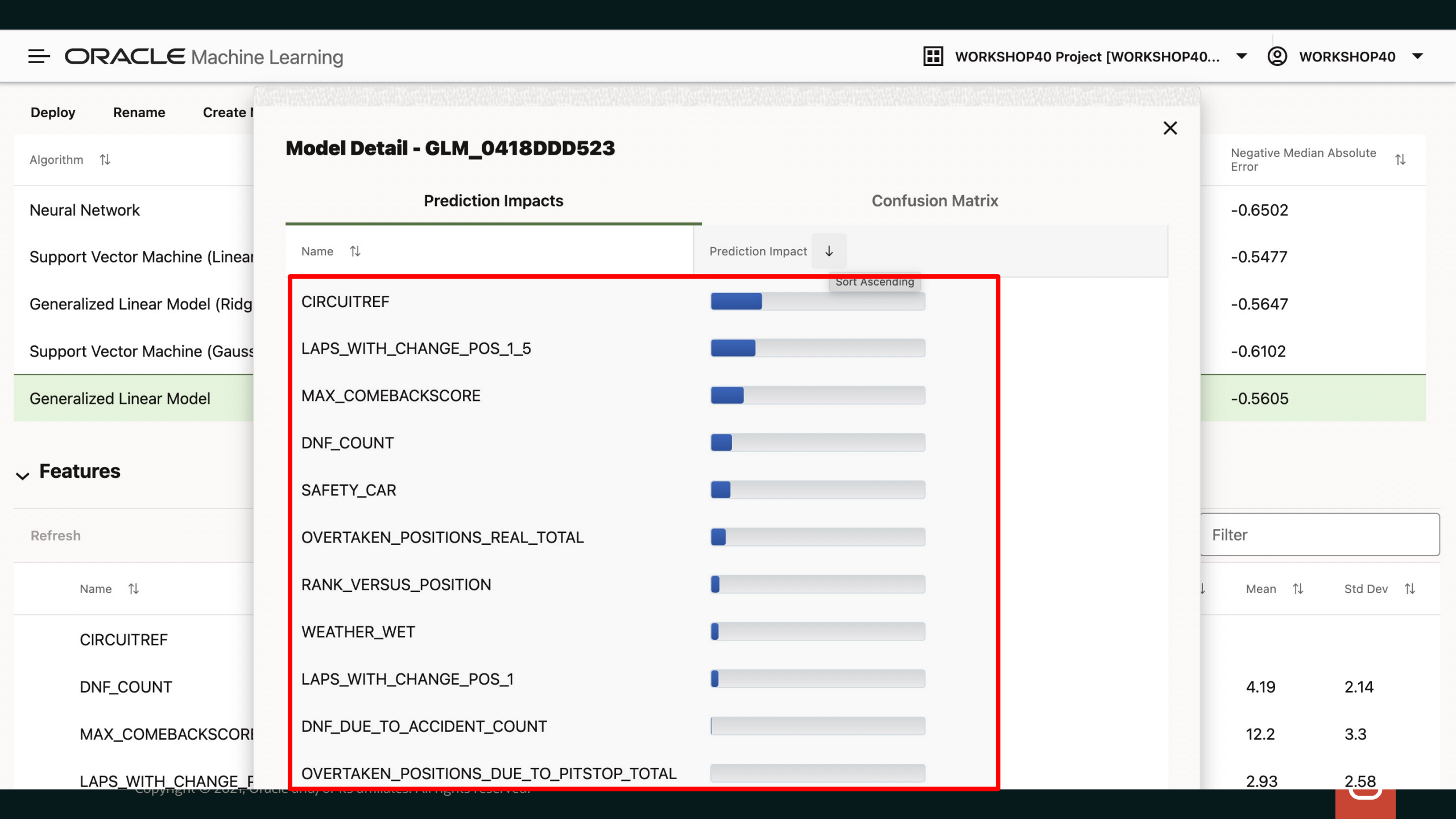
Features

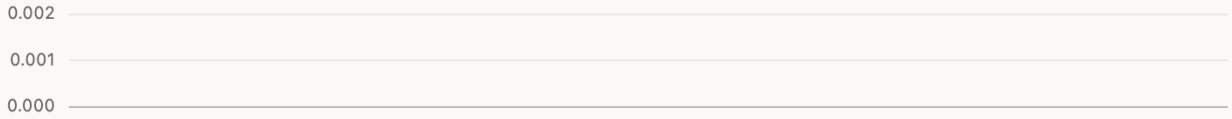
Leader Board

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

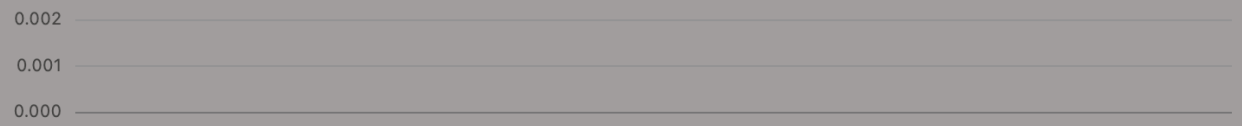
Algorithm	Model Name	Negative Mean Absolute Error	Negative Mean Squared Error	R2	Negative Median Absolute Error
Neural Network	NN_F28CEE823D	-0.6774	-0.7624	0.4804	-0.6502
Support Vector Machine (Linear)	SVML_552C7B93B5	-0.6360	-0.6458	0.5599	-0.5477
Generalized Linear Model (Ridge ...	GLMR_7969DE44D0	-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:

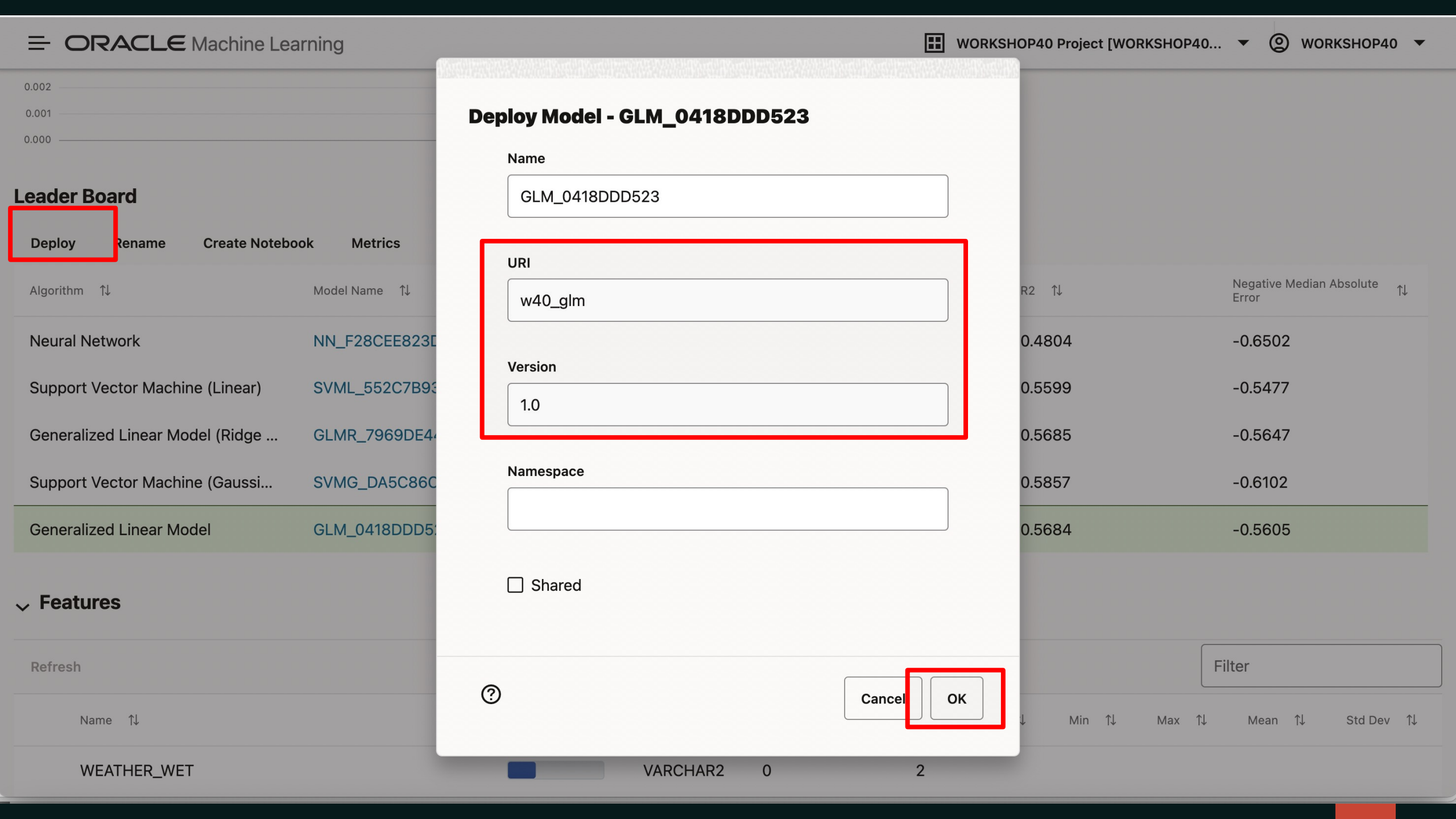
Predict race score GLM (1)

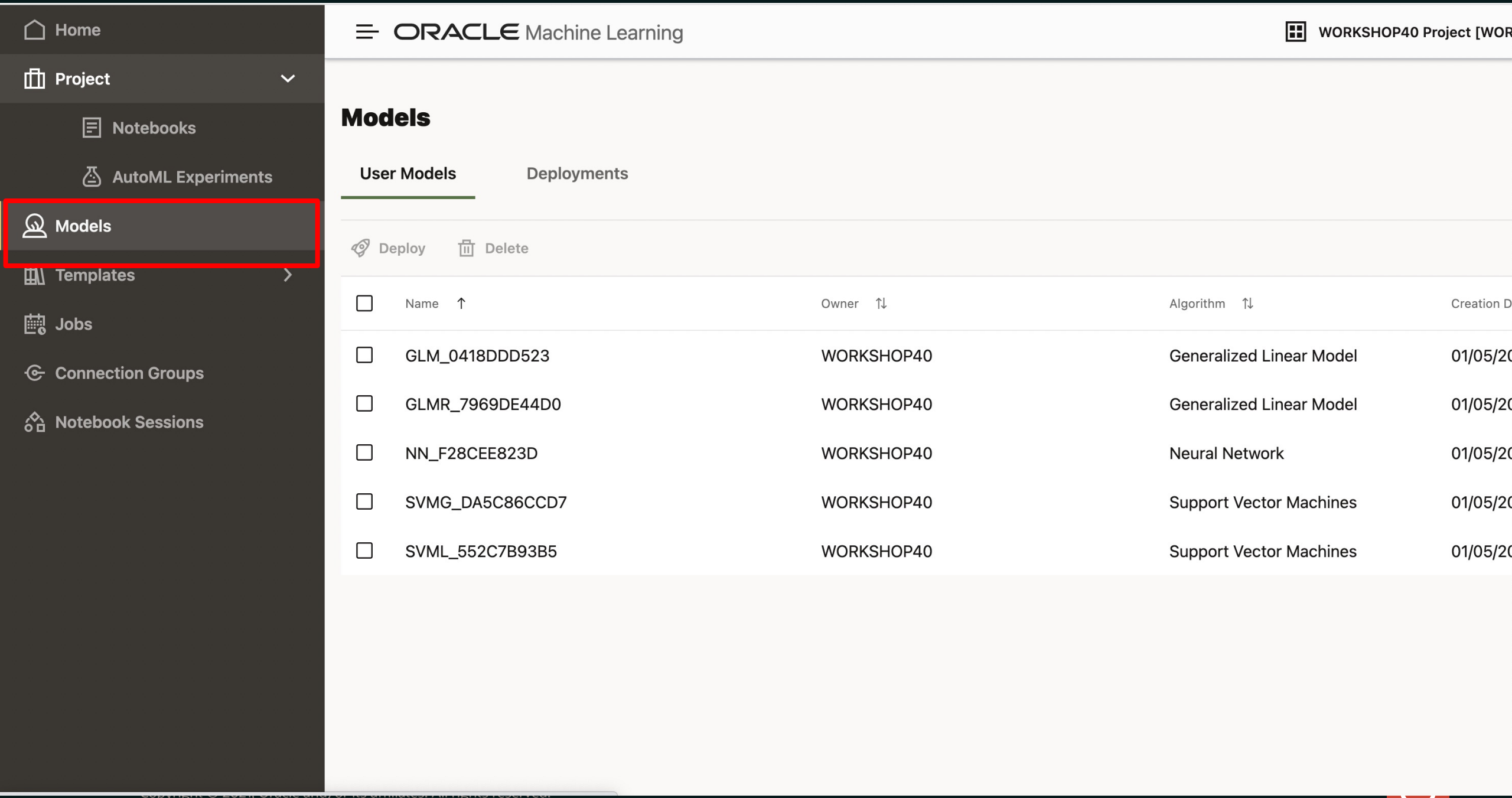
Cancel OK

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

Model Repository

User Models | Deployments

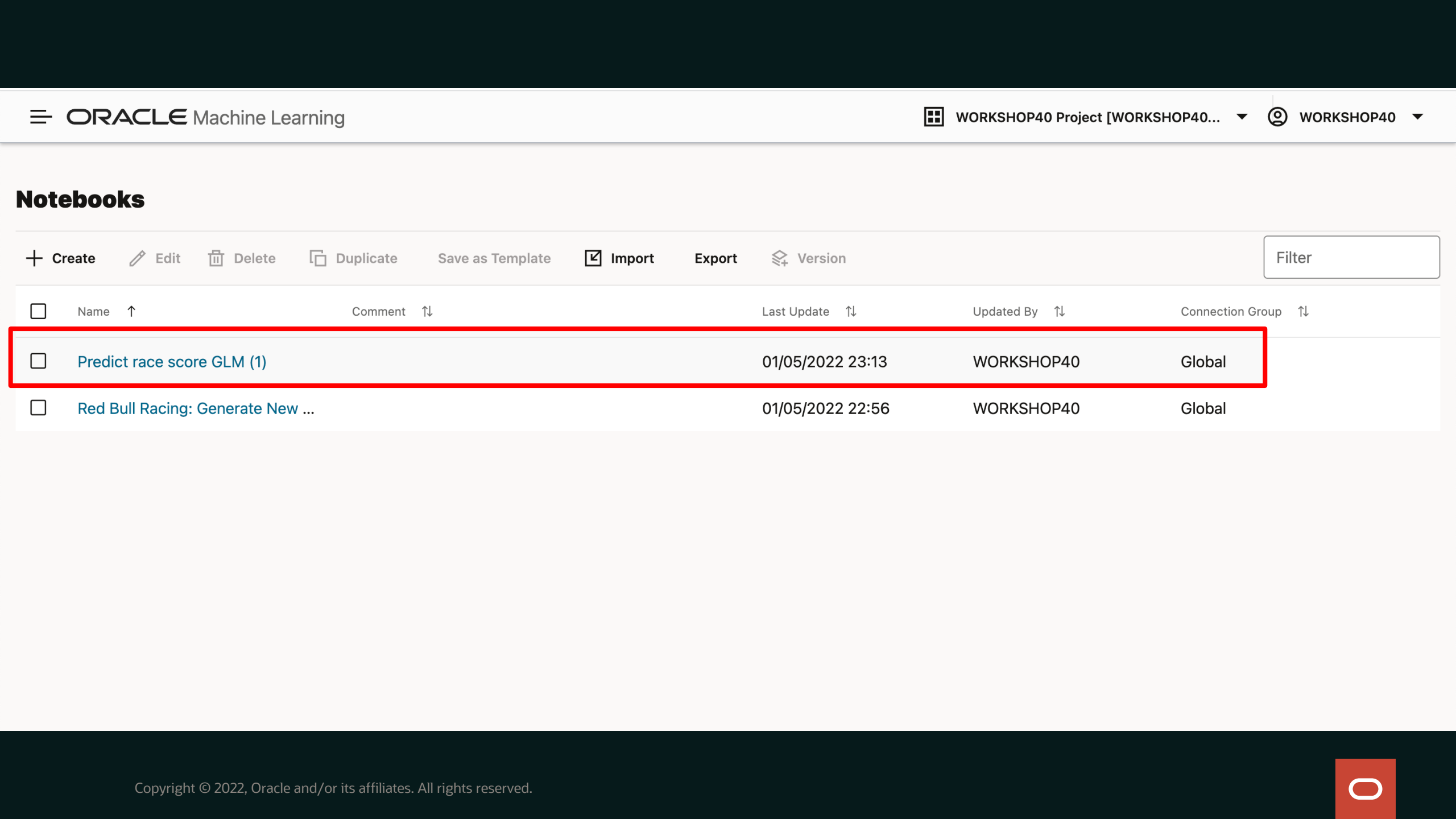
Delete

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

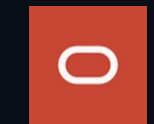
READY

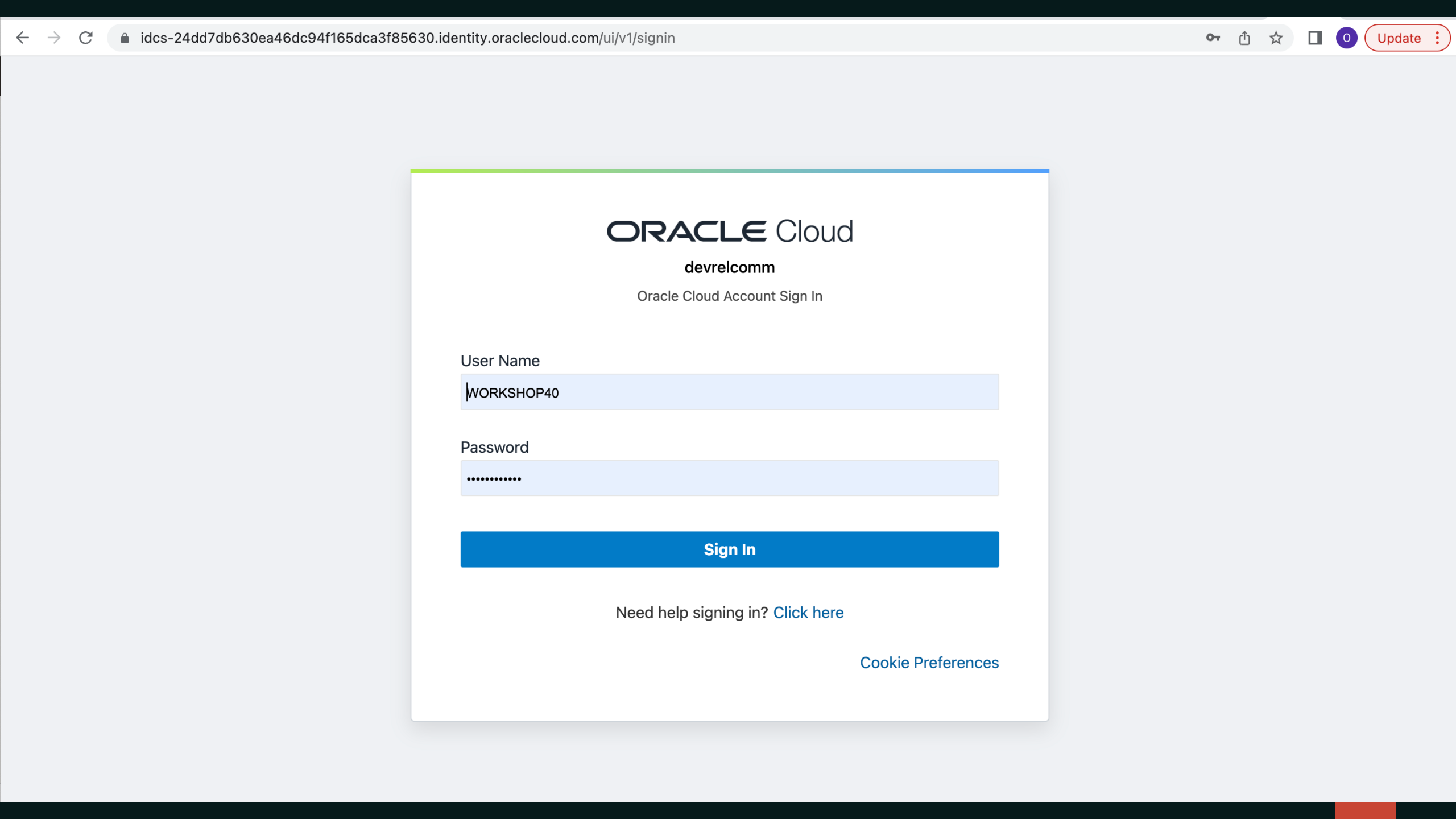
```
%python
import sklearn as skl
metric_score = skl.metrics.mean_absolute_error(y_true, y_pred)
print(metric_score.round(4))
```

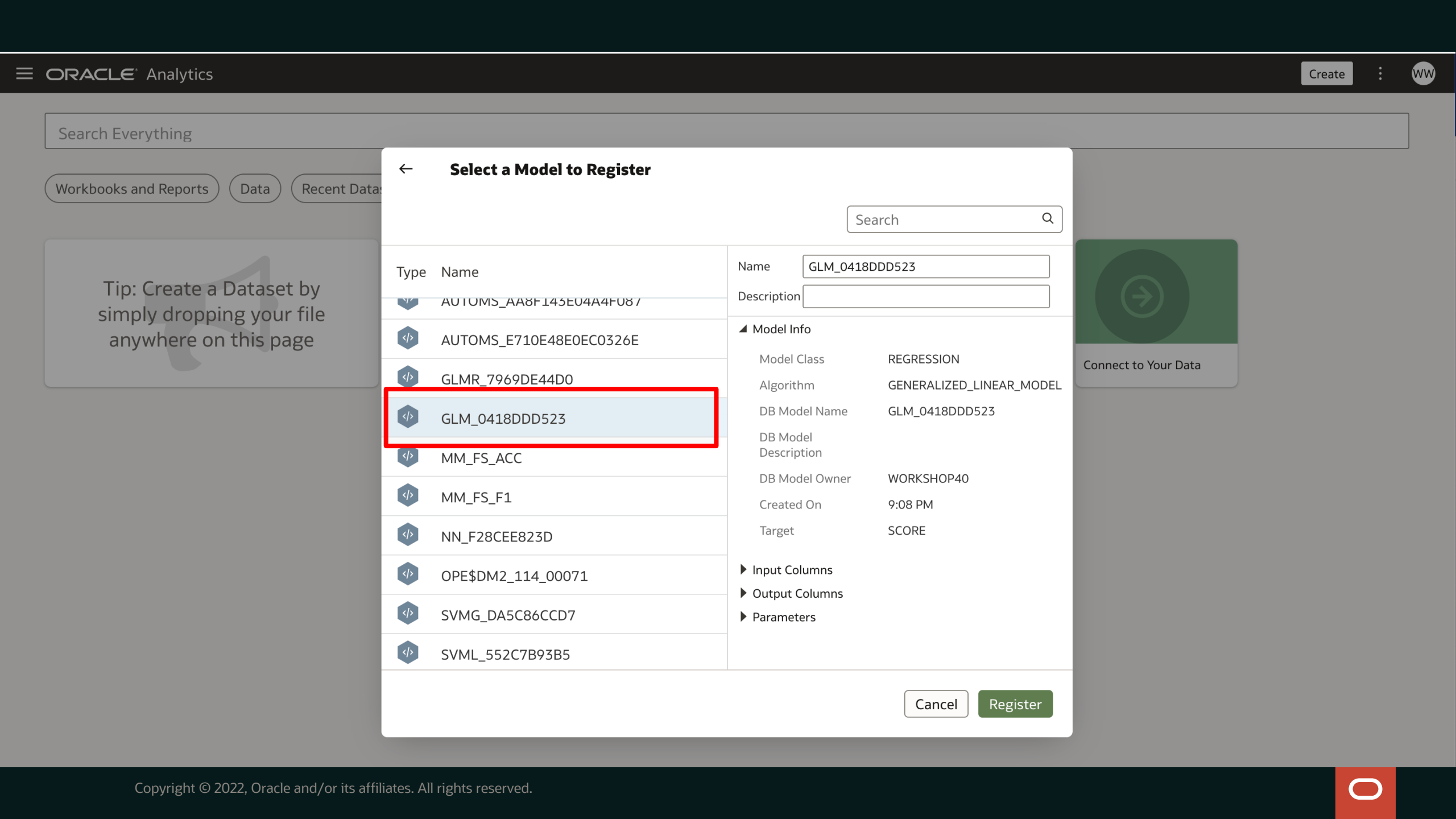
Rol: Data Visualizer

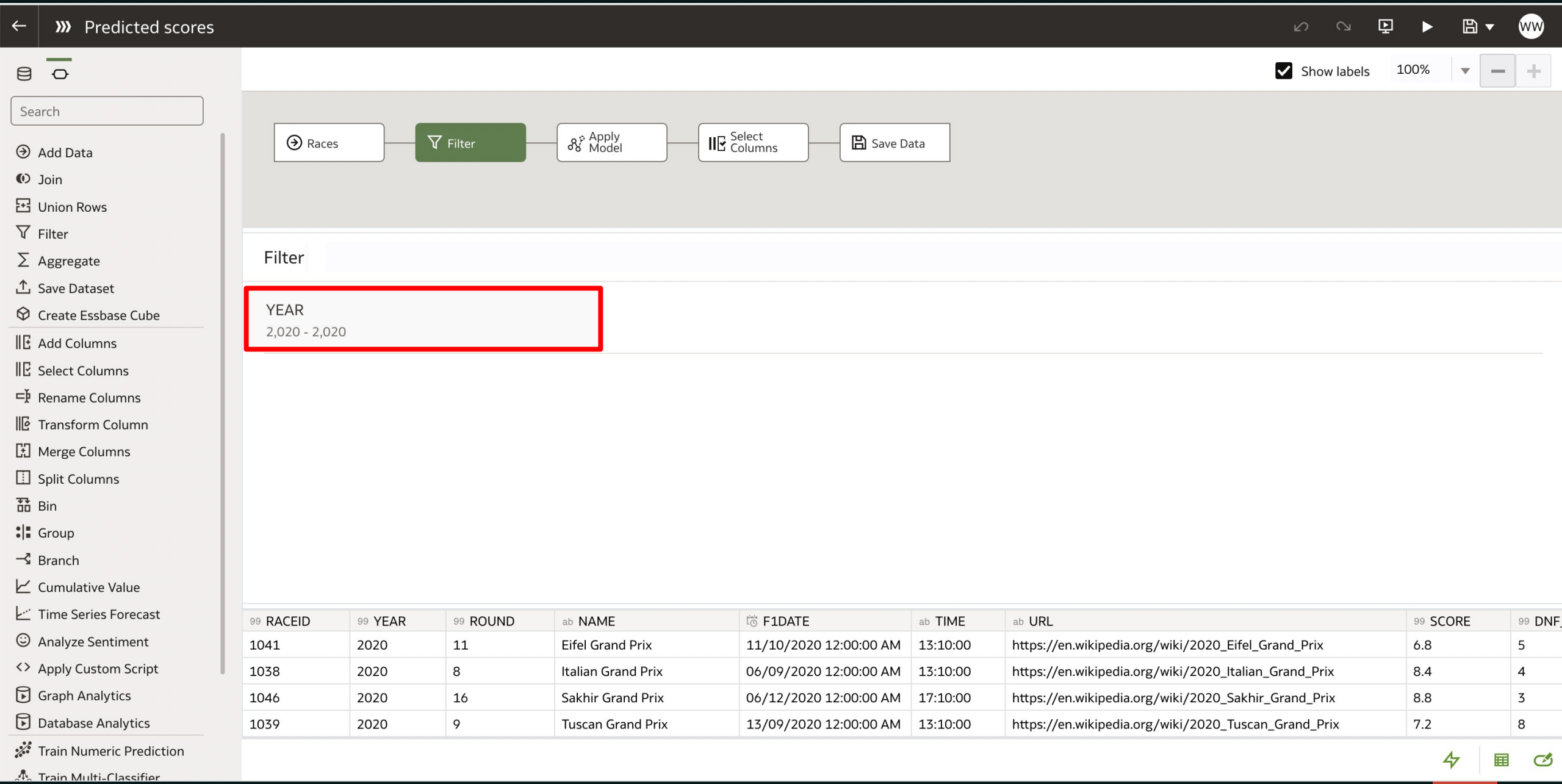
Rol: Data Scientist

Tasks 3,4,5,6: Using Data Visualization …




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

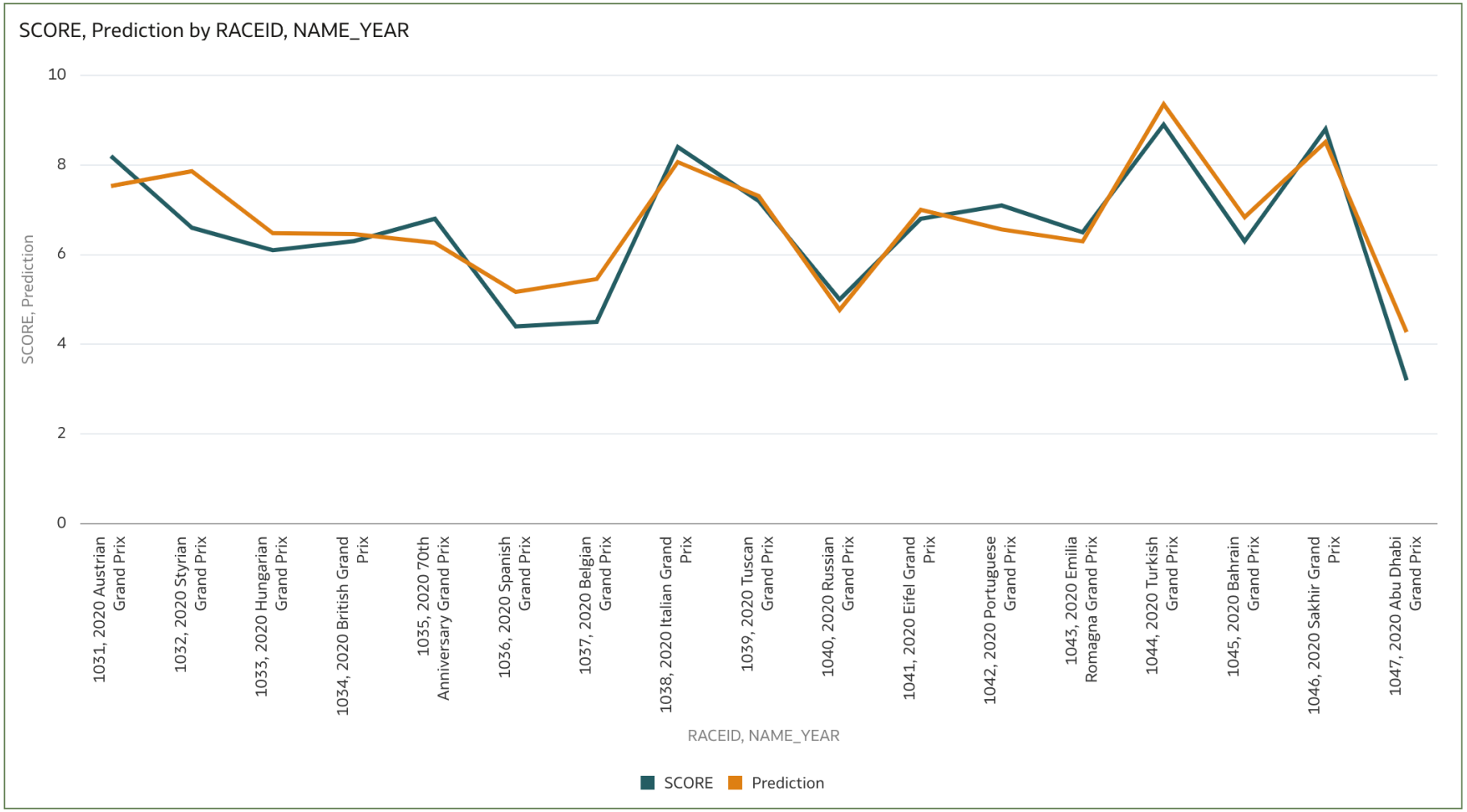
Size (Width)

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/vidacaixa2022

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

Rol: Data Engineer

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

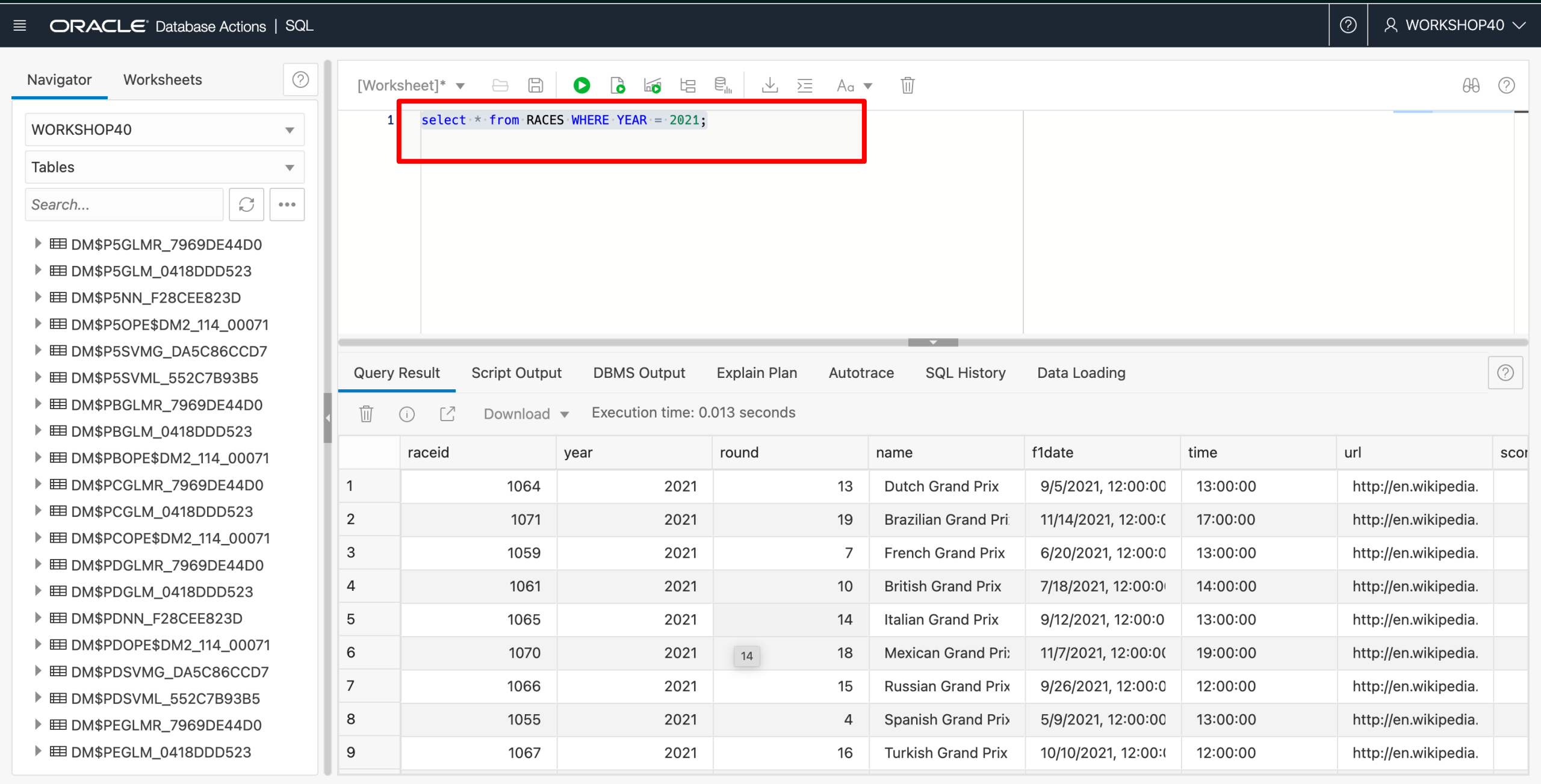
Rol: Data Scientist

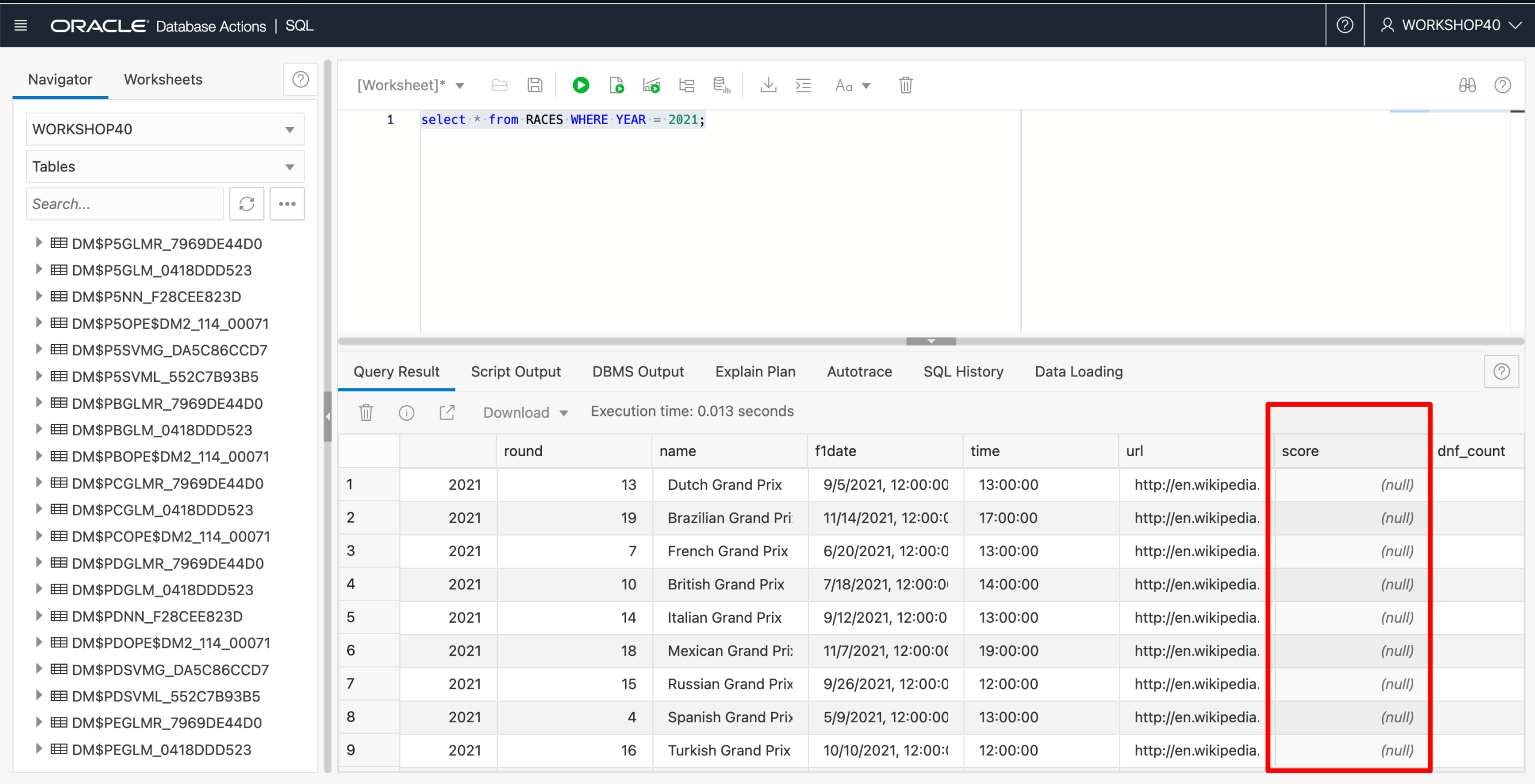
4: Evaluate the accuracy of the model for 2021

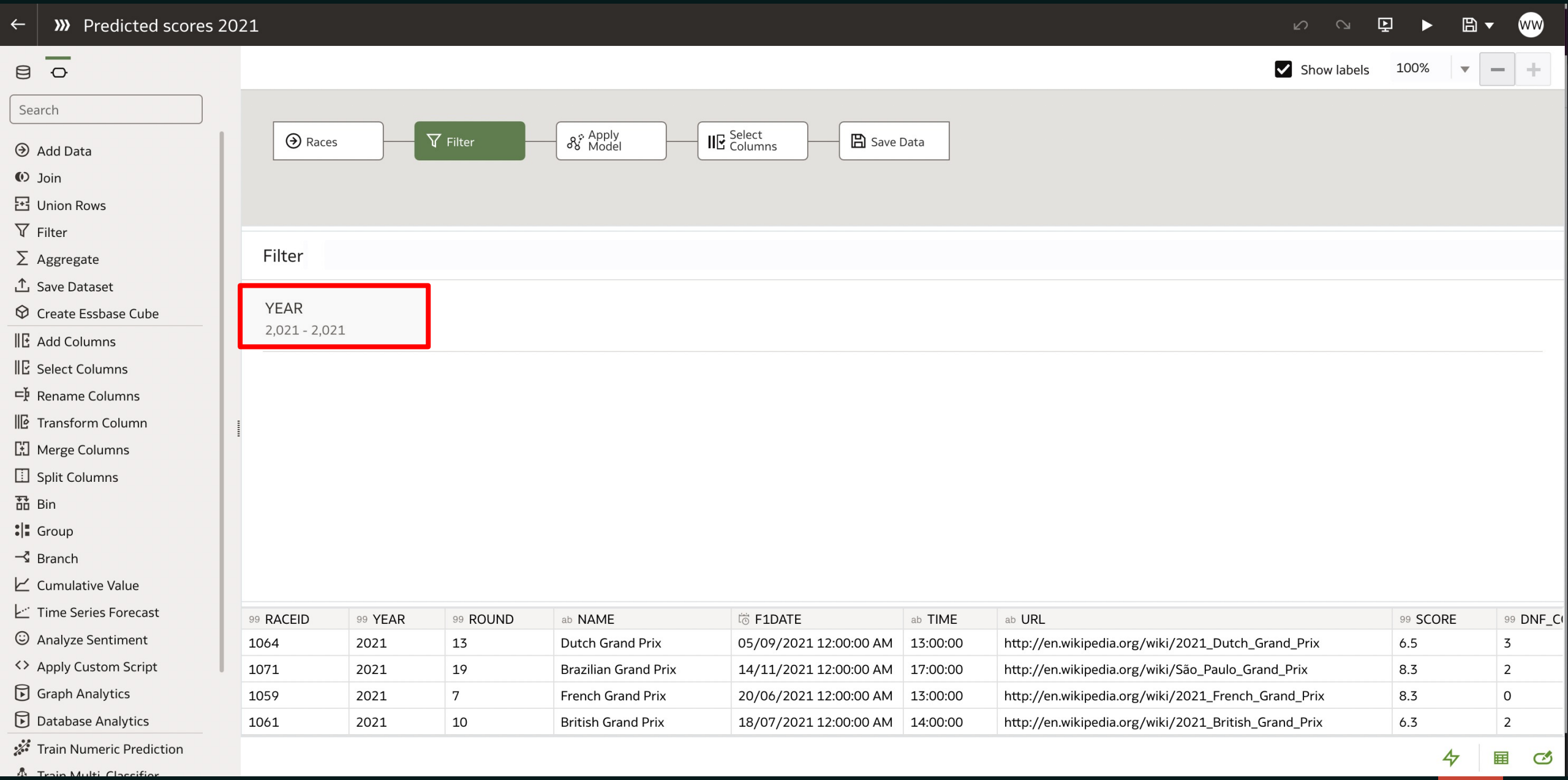
Lab instructions

https://bit.ly/vidacaixa2022

Rol: Data Visualizer







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

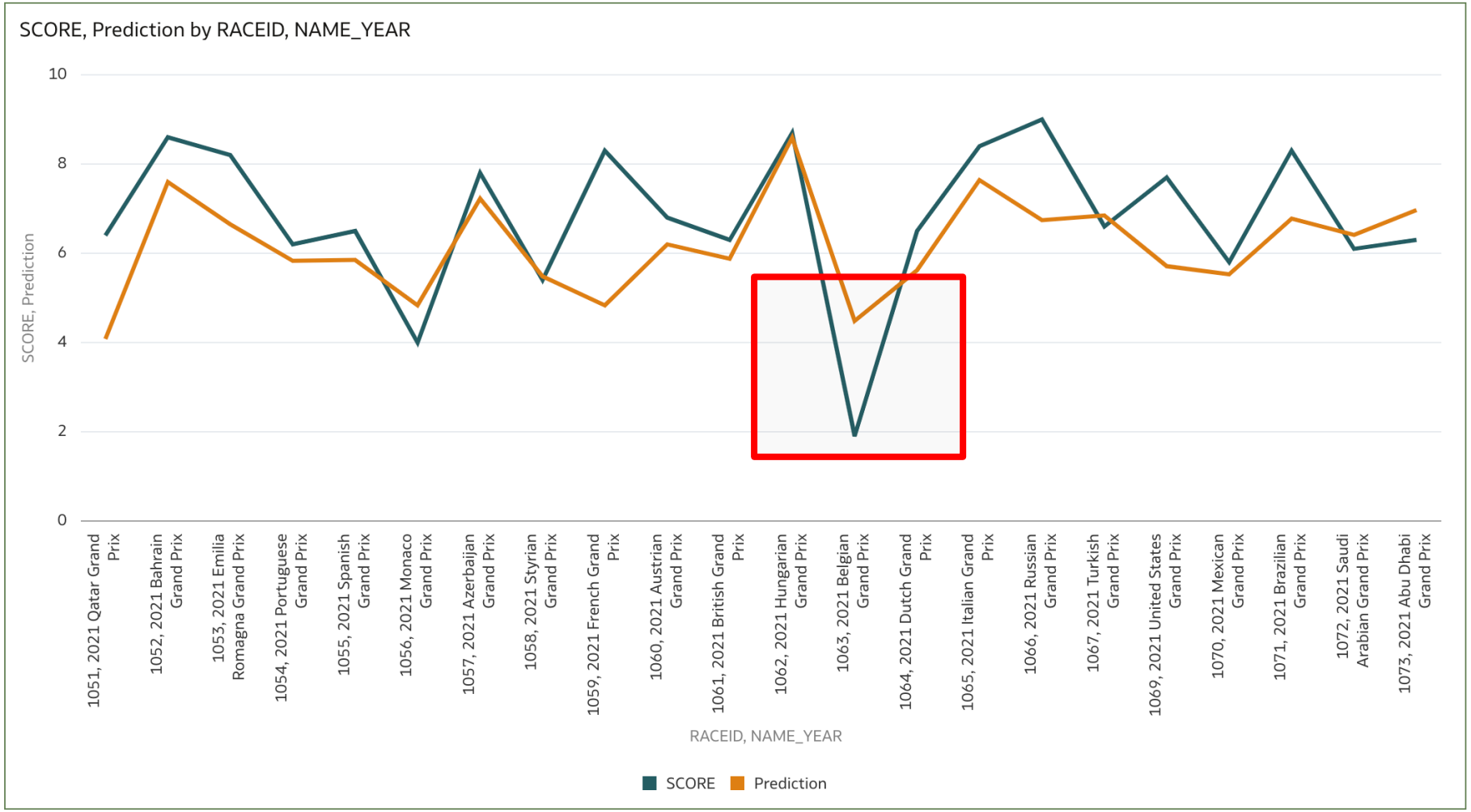
Size (Width)

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

Thank You


ORACLE