

# Welcome at Banco Sabadell 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 its own badge where you can earn the title of **Developer Pioneer** recognizing your commitment and earned expertise.



Workshop Detailed:  
<https://github.com/operard/bancosabadell2022>

Lab instructions  
<https://bit.ly/bancosabadell2022>

# Organización del workshop

# Data Engineers



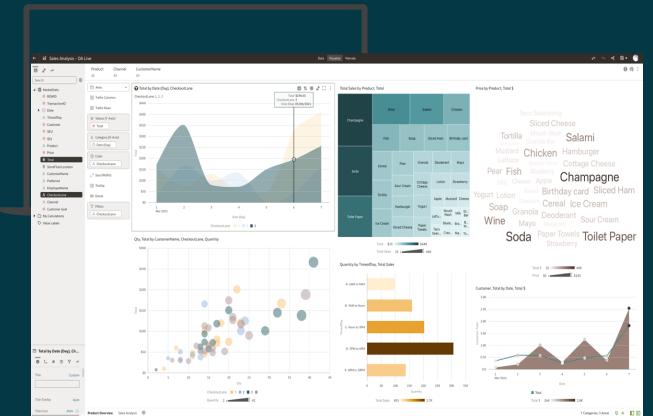
## Data Loading

# Data Scientists



## Model training

# Data Visualization



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

## Autonomous Data Warehouse

workshop1

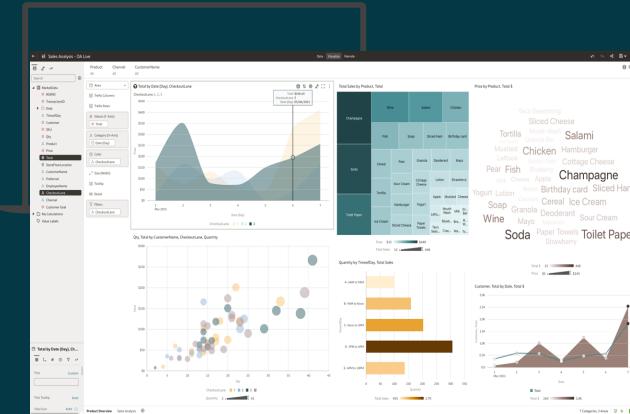
workshop2

workshop3

workshop...

workshop40

## Analytics Cloud



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

## Users

 Create	 Delete	<input type="checkbox"/> 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	

# 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



# Formula 1 Data



<https://www.formula1.com/>

The screenshot shows the Formula 1 website homepage for the 2022 Australian Grand Prix. At the top, there are links for FIA, F1™, F2™, and F3™. The main navigation includes AUTHENTICS, STORE, TICKETS, HOSPITALITY, EXPERIENCES, F1 TV, SIGN IN, and SUBSCRIBE. Below this, a banner for the "AUSTRALIA 2022" race features the date 08 - 10 April and a timer showing 00 DAYS, 21 HRS, 29 MINS. The main content area has a large image of two men smiling. A news article headline reads "'It's like arriving late to school' says Vettel, as he prepares for first race of 2022 in Melbourne". To the right, there are video thumbnails for "First victory main target for 2022, says Sainz" and "WATCH: Extended highlights of the 2003 Australian Grand Prix".

## 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/>

The screenshot shows the RaceFans website homepage for the 2022 Australian Grand Prix. The top navigation includes Log in, Register, Go Ad-Free, Predictions, and About. The main content area features several news articles:

- Pirelli explains thinking behind 'step' in Melbourne tyre compound choice** (2022 Australian Grand Prix) by Hazel Southwell and Claire Cottingham (1 comment)
- No "magic fix" for Mercedes' lack of pace in Australia** (RaceFans Round-up) by Hazel Southwell (6 comments)
- Albert Park's second sector "now basically a massive straight" after changes** (2022 Australian Grand Prix) by Keith Collantine (17 comments)

A sidebar on the right is titled "FOLLOW RACEFANS" and includes social media links for Twitter, Facebook, Instagram, YouTube, and RSS. Another sidebar titled "NEW ON RACEFANS" lists recent articles.

## DATA SETS

- Qualifying Fans Data
- Race Fans Data



© Renault

## 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 [ADVERT | BECOME A SUPPORTER & GO AD-FREE](#) 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

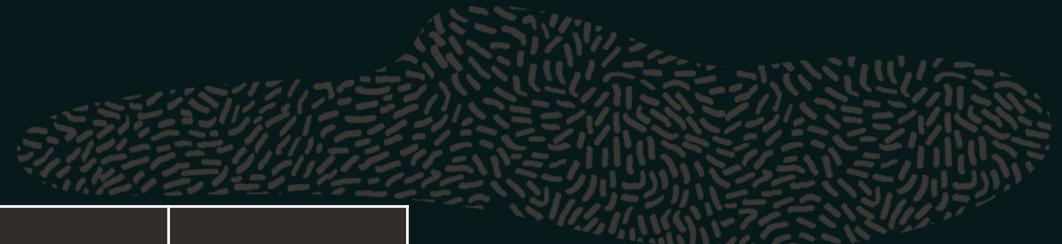
Lab 2 Basic data exploration

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



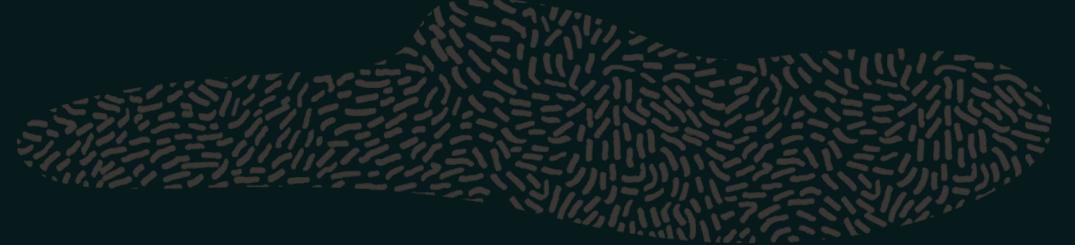
Name	Round	Wet weather	Over taken positions	Did-Not-Finish Count	...	Score (racefans.net)
2020 Hungarian Grand Prix	3	Y	222	1	...	6.1
2020 Styrian Grand Prix	2	N	131	3	...	6.6
2020 Austrian Grand Prix	1	N	92	9	...	8.2
2010 Malaysian Grand Prix	3	N	87	7	...	6.684
2010 Australian Grand Prix	2	Y	120	10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

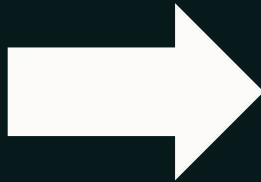
Name	Round	Wet weather	Over taken positions	Did-Not-Finish Count	...	Score (racefans.net)
2020 Hungarian Grand Prix	3	Y	222	1	...	6.1
2020 Styrian Grand Prix	2	N	131	3	...	6.6
2020 Austrian Grand Prix	1	N	92	9	...	8.2
2010 Malaysian Grand Prix	3	N	87	7	...	6.684
2010 Australian Grand Prix	2	Y	120	10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



Number of overtakes

?

Weather



Score

Circuit (location)

DNFs

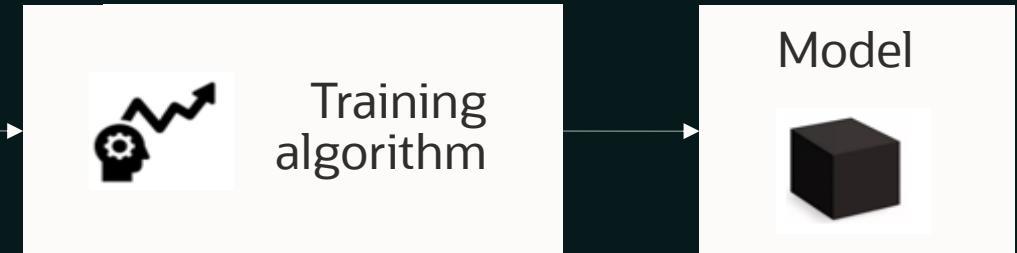
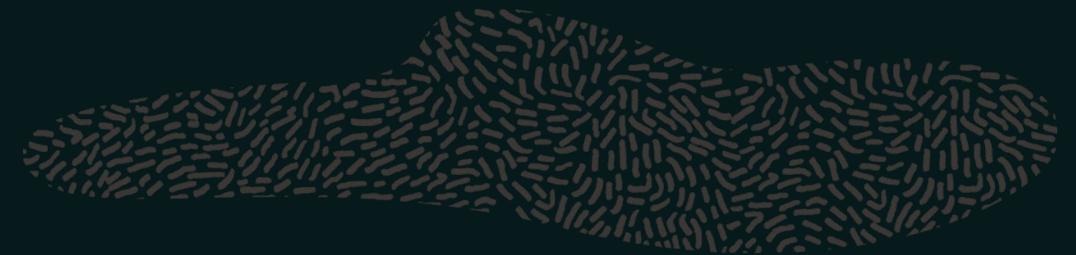
...

## 1. Train (data until end of 2019)

Input features train (factors)

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
...	...	...	...	...	...

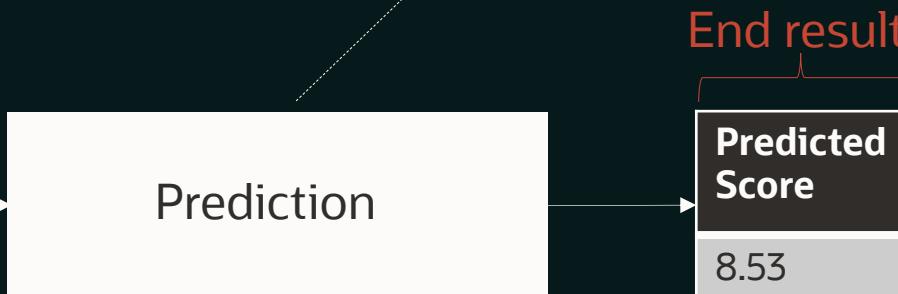
Target



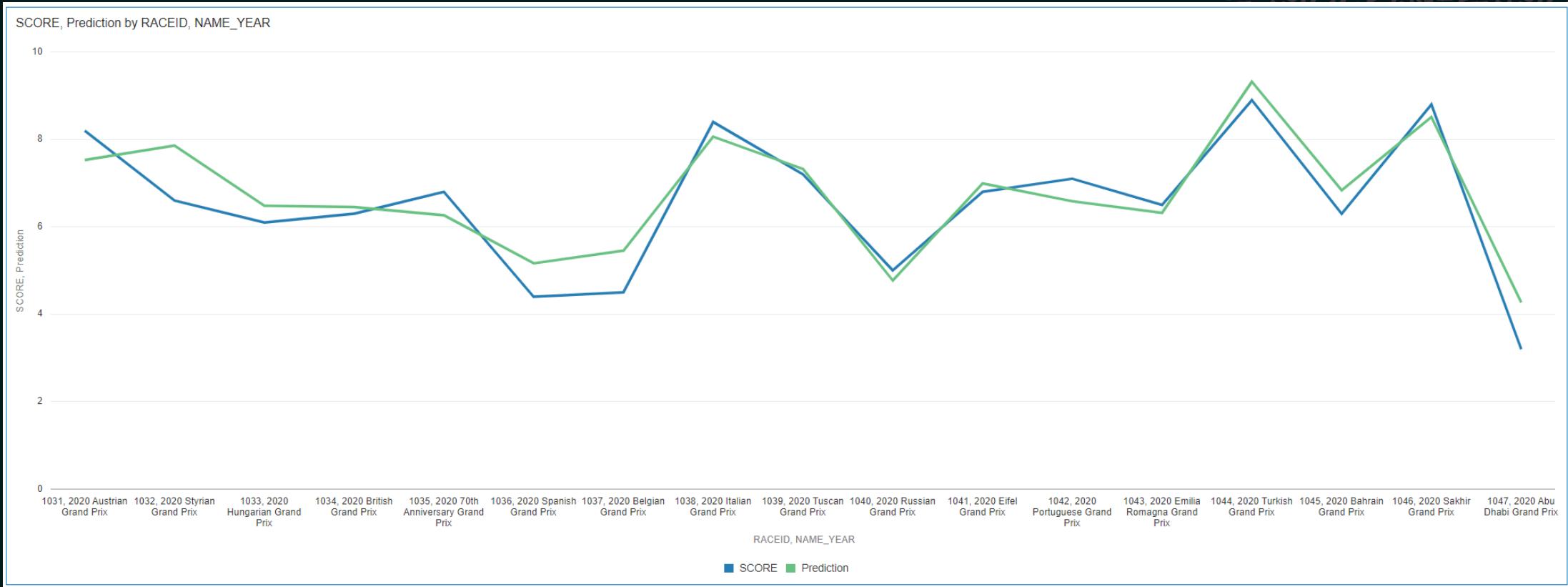
## 2. Test (data 2020)

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

Input features test (factors)



# 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/bancosabadell2022>

## 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/bancosabadell2022>

# 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/bancosabadell2022>

## 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/bancosabadell2022>

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/bancosabadell2022>

# 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

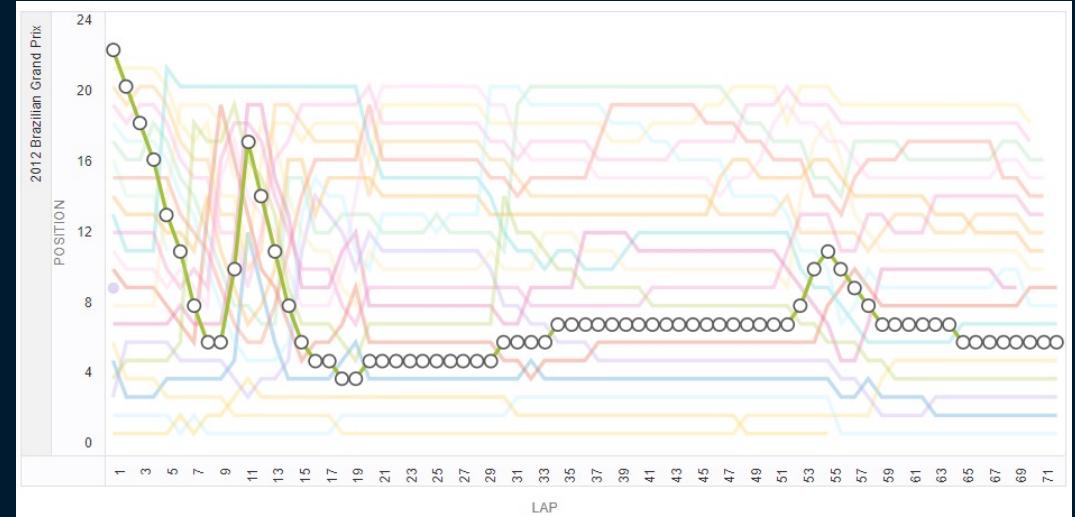
## Lab 3: Advanced Data Exploration

¿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/bancosabadell2022>

## 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/bancosabadell2022>

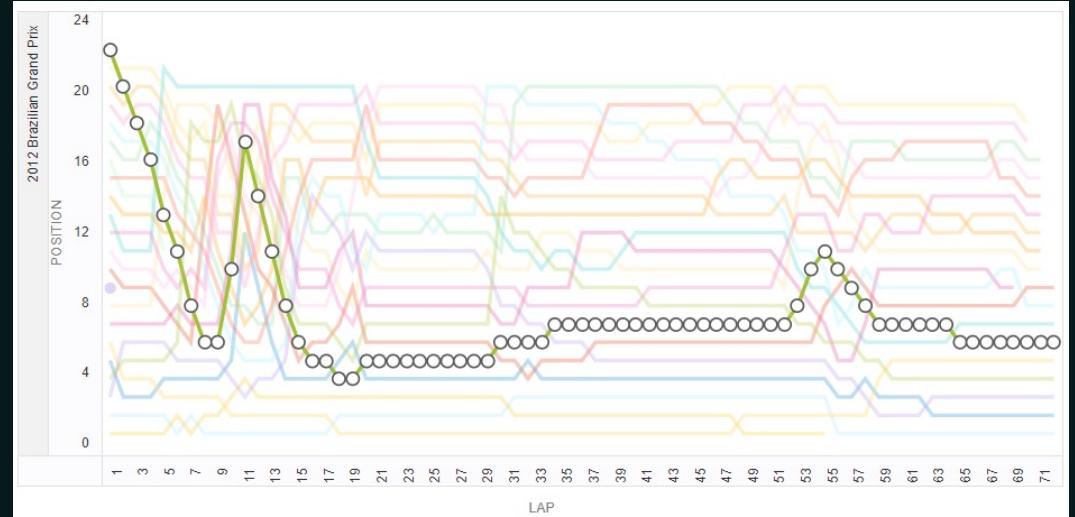
## Lab 3: Advanced Data Exploration

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/bancosabadell2022>

# 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/bancosabadell2022>

# 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/bancosabadell2022>

## Task 1: Feature Generation

1. Using Notebook
2. Using SQL Script

Role: Data Engineer

The screenshot shows the Oracle Machine Learning interface with two code snippets in a notebook.

**Top Snippet:**

```
%sql
--This calculates the number of rounds that have a change of driver in the leading position.
create or replace view laps_with_change_pos_1_v
as
  select raceid, count(*) laps
  from
  (
    select raceid, lap
    from
    (
      select raceid, position, lap, driverid as current_driver,
      LAG(driverid, 1) OVER (PARTITION BY raceid, position ORDER BY lap) AS prev_driver
      from lap_times
    )
    where prev_driver is not null
    and prev_driver <> current_driver
    and position = 1
    group by raceid, lap
  )
  group by raceid;
```

FINISHED

Took 0 secs. Last updated by F1 at September 23 2021, 5:18:26 PM.

**Bottom Snippet:**

```
%sql
select * from laps_with_change_pos_1_v
```

FINISHED

RACEID	LAPS
6	2
14	2

Rol: Data Scientist

## Task 2: Using ML Algorithms ...

## Create Experiment

▶ Start ▾

Cancel

Save

Name

Predict race score

Comments

Predict race score

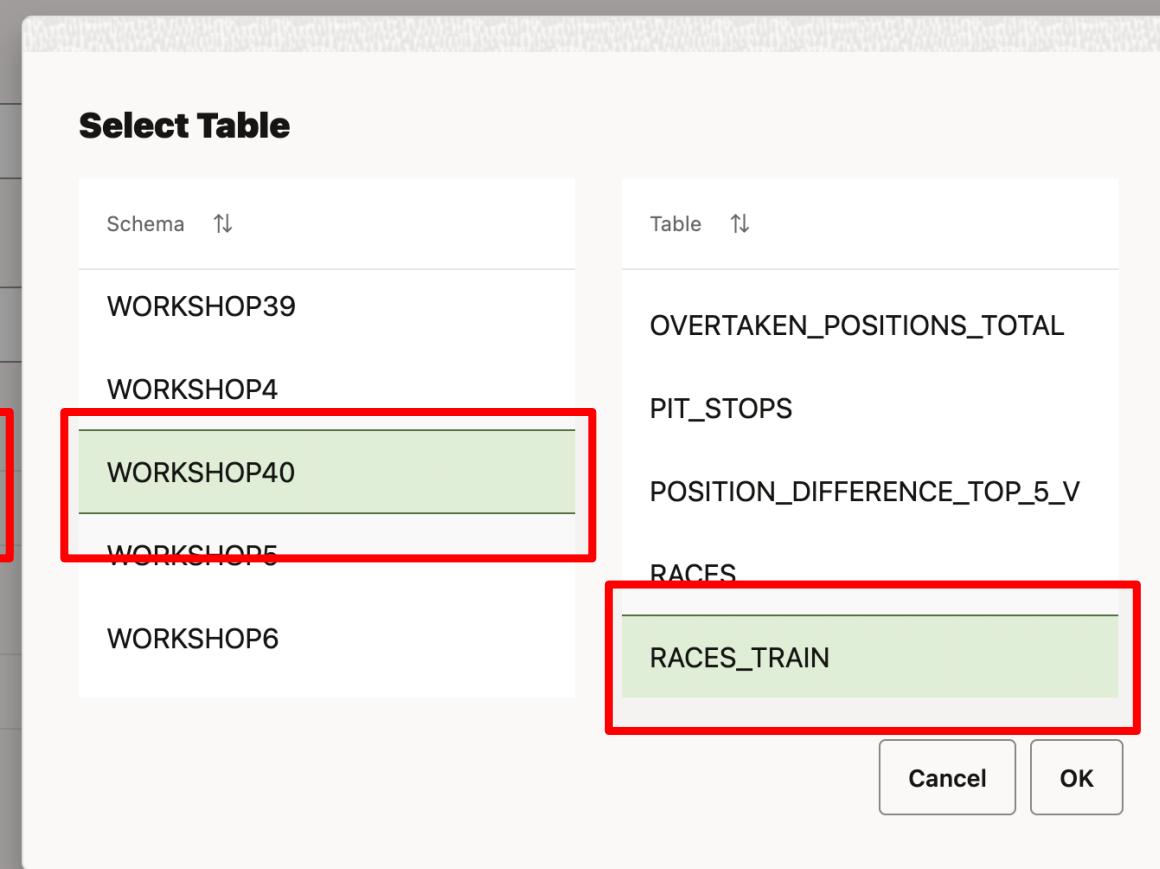
Data Source

Prediction Type

Select Prediction Type

› Additional Settings

▼ Features



## 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**

&lt;- Experiments

## Predict race score

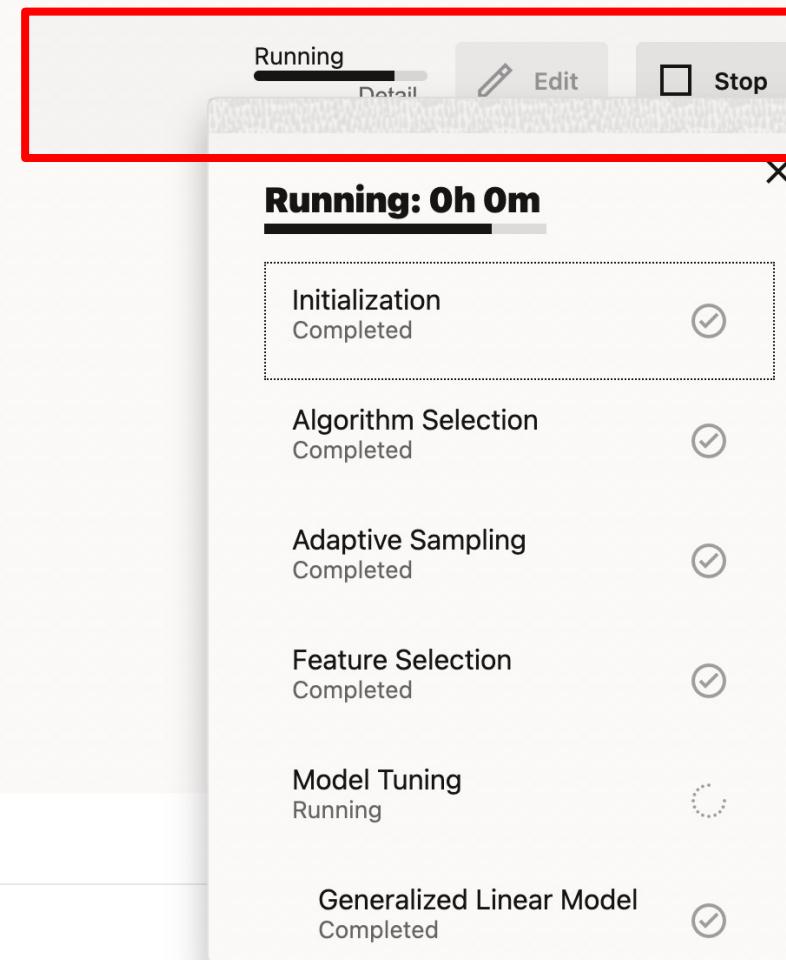
&gt; Settings

### 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



&lt;- Experiments

## Predict race score

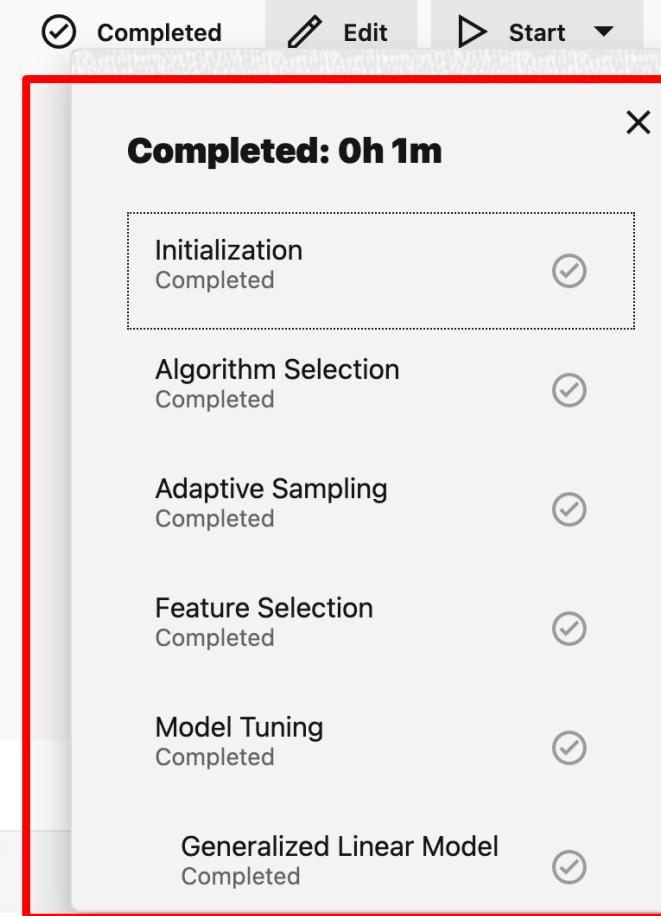
&gt; Settings

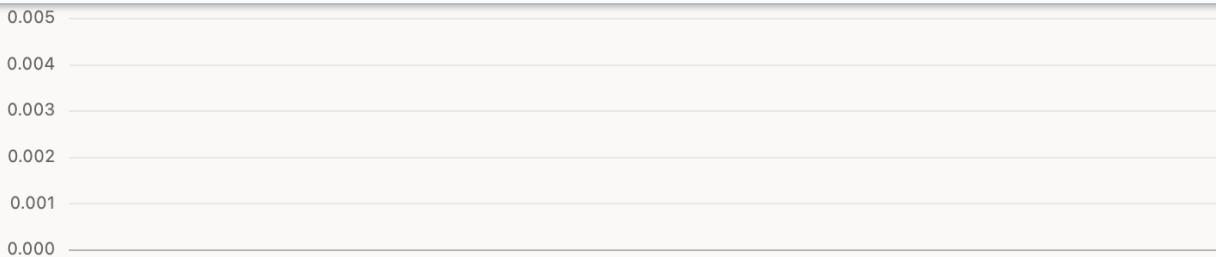
### Negative Mean Absolute Error



### Leader Board

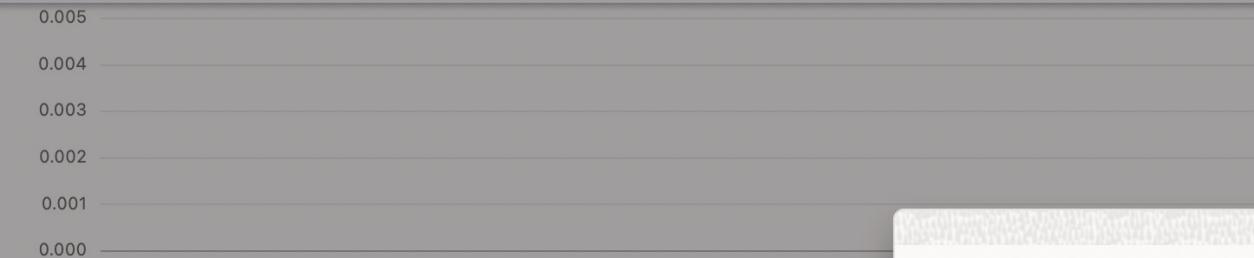
Deploy	Rename	Create Notebook	Metrics
Algorithm	Model Name	Negative Mean Absolute Error	
Neural Network	NN_F28CEE823D	-0.6774	
Support Vector Machine (Linear)	SVML_552C7B93B5	-0.6360	
Generalized Linear Model (Ridge ...)	GLMR_7969DE44D0	-0.6348	
Support Vector Machine (Gaussi...	SVMG_DA5C86CCD7	-0.6346	





## Leader Board

Deploy	Rename	Create Notebook	Metrics
Algorithm	↑↓	Model Name	↑↓
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

Deploy   Rename   Create I

Algorithm ↑

Neural Network

Support Vector Machine (Linear)

Generalized Linear Model (Ridge)

Support Vector Machine (Gaussian)

Generalized Linear Model

Features

Refresh

Name ↑

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

### Model Detail - GLM\_0418DDD523

**Prediction Impacts**

Name ↑	Prediction Impact ↑↓
CIRCUITREF	<div style="width: 30%;"></div>
DNF_COUNT	<div style="width: 10%;"></div>
DNF_DUE_TO_ACCIDENT_COUNT	<div style="width: 0%;"></div>
LAPS_WITH_CHANGE_POS_1	<div style="width: 5%;"></div>
LAPS_WITH_CHANGE_POS_1_5	<div style="width: 20%;"></div>
MAX_COMEBACKSCORE	<div style="width: 15%;"></div>
OVERTAKEN_POSITIONS_DUE_TO_PITSTOP_TOTAL	<div style="width: 0%;"></div>
OVERTAKEN_POSITIONS_REAL_TOTAL	<div style="width: 5%;"></div>
RANK_VERSUS_POSITION	<div style="width: 5%;"></div>
SAFETY_CAR	<div style="width: 10%;"></div>
WEATHER_WET	<div style="width: 5%;"></div>

**Confusion Matrix**

Negative Median Absolute Error ↑↓

-0.6502
-0.5477
-0.5647
-0.6102
-0.5605
Filter ↓
Mean ↑ Std Dev ↑
4.19 2.14
12.2 3.3
2.93 2.58

Deploy   Rename   Create I

Algorithm ↑↓

Neural Network

Support Vector Machine (Linear)

Generalized Linear Model (Ridge)

Support Vector Machine (Gauss)

Generalized Linear Model

Features

Refresh

Name ↑↓

CIRCUITREF

DNF\_COUNT

MAX\_COMEBACKSCORE

LAPS\_WITH\_CHANGE\_POS\_1\_5

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

### Model Detail - GLM\_0418DDD523

**Prediction Impacts**

Name ↑↓	Prediction Impact ↓
CIRCUITREF	Sort Ascending
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	

**Confusion Matrix**

Negative Median Absolute Error
-0.6502
-0.5477
-0.5647
-0.6102
-0.5605
Filter
↓ Mean ↑↓ Std Dev ↑↓
4.19 2.14
12.2 3.3
2.93 2.58

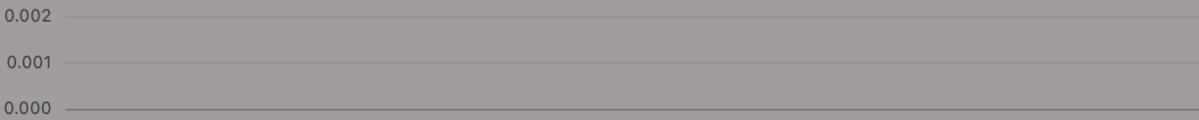


## Leader Board

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								
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 ↑

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

### 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

**Leader Board****Deploy**   **Rename**   **Create Notebook**   **Metrics**

Algorithm ↑   Model Name ↑

Neural Network   NN\_F28CEE823D

Support Vector Machine (Linear)   SVML\_552C7B93

Generalized Linear Model (Ridge ...)   GLMR\_7969DE44

Support Vector Machine (Gaussi...)   SVMG\_DA5C86C

Generalized Linear Model   GLM\_0418DDD523

**Features**

Refresh

Name ↑

WEATHER\_WET

**Deploy Model - GLM\_0418DDD523****Name**

GLM\_0418DDD523

**URI**

w40\_glm

**Version**

1.0

**Namespace** Shared

Cancel

OK

Filter

VARCHAR2 0

2

## Models

### User Models      Deployments

Deploy    Delete

<input type="checkbox"/>	Name ↑	Owner ↑	Algorithm ↑	Creation Date ↑
<input type="checkbox"/>	GLM_0418DDD523	WORKSHOP40	Generalized Linear Model	01/05/2024
<input type="checkbox"/>	GLMR_7969DE44D0	WORKSHOP40	Generalized Linear Model	01/05/2024
<input type="checkbox"/>	NN_F28CEE823D	WORKSHOP40	Neural Network	01/05/2024
<input type="checkbox"/>	SVMG_DA5C86CCD7	WORKSHOP40	Support Vector Machines	01/05/2024
<input type="checkbox"/>	SVML_552C7B93B5	WORKSHOP40	Support Vector Machines	01/05/2024

## Model Repository

User Models

Deployments



Delete

Filter

<input type="checkbox"/>	Name ↑	Shared ↑↓	Version ↑↓	Namespace ↑↓	Owner ↑↓	Deployed Date ↑↓	URI ↑↓
<input type="checkbox"/>	GLM_0418DDD523		1.0		WORKSHOP40	01/05/2022 ...	w40_glm

## Model Repository

User Models

Deployments



Name ↑

 GLM\_0418DDD523

### 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": {}  
      }  
    }  
  }  
}
```

Filter

URI ↑↓

w40\_glm

## Notebooks

<a href="#">Create</a> <a href="#">Edit</a> <a href="#">Delete</a> <a href="#">Duplicate</a> <a href="#">Save as Template</a> <a href="#">Import</a> <a href="#">Export</a> <a href="#">Version</a>								<a href="#">Filter</a>		
<input type="checkbox"/>	Name	↑	Comment	↑↓	Last Update	↑↓	Updated By	↑↓	Connection Group	↑↓
<input type="checkbox"/>	<a href="#">Predict race score GLM (1)</a>				01/05/2022 23:13		WORKSHOP40		Global	
<input type="checkbox"/>	<a href="#">Red Bull Racing: Generate New ...</a>				01/05/2022 22:56		WORKSHOP40		Global	

# Predict race score GLM (1)



## Oracle Machine Learning AutoML UI - Experiment - Generated Notebook

READY ▶ ✎ 📄⚙️

### Get proxy object for selected data

READY ▶ ✎ 📄⚙️

```
%python\n\nimport oml\n\ncolumns = '"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" ,\n    "MAX_COMEBACKSCORE" , "OVERTAKEN_POSITIONS_DUE_TO_PITSTOP_TOTAL" , "WEATHER_WET" , "RANK_VERSUS_POSITION" , "SCORE" , "RACEID"\nschema='WORKSHOP40'\ntable='RACES_TRAIN'\n\ncolumn = ','.join(columns)\nquery = 'SELECT ' + column + ' FROM ' + schema + '.' + table + ' where ' + '"SCORE"' + ' is not null'\n\nbuild_data = oml.sync(query=query)\nz.show(build_data)
```

### Prepare training data

READY ▶ ✎ 📄⚙️

```
%python\n\nX_train = build_data.drop('SCORE')\ny_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 ...

# ORACLE Cloud

devrelcomm

Oracle Cloud Account Sign In

User Name

Password

**Sign In**

Need help signing in? [Click here](#)

[Cookie Preferences](#)

Search Everything

Workbooks and Reports

Data

Recent Data

Tip: Create a Dataset by  
simply dropping your file  
anywhere on this page

## Select a Model to Register

Type	Name
▼	AUTOMS_AA8F145E04A4F087
◀/▶	AUTOMS_E710E48E0EC0326E
◀/▶	GLMR_7969DE44D0
◀/▶	GLM_0418DDD523
◀/▶	MM_FS_ACC
◀/▶	MM_FS_F1
◀/▶	NN_F28CEE823D
◀/▶	OPE\$DM2_114_00071
◀/▶	SVMG_DA5C86CCD7
◀/▶	SVML_552C7B93B5

 SearchName Description 

### Model Info

Model Class	REGRESSION
Algorithm	GENERALIZED_LINEAR_MODEL
DB Model Name	GLM_0418DDD523
DB Model Description	
DB Model Owner	WORKSHOP40
Created On	9:08 PM
Target	SCORE

### ▶ Input Columns

### ▶ Output Columns

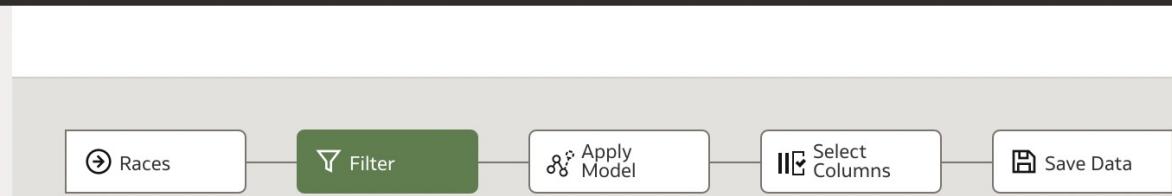
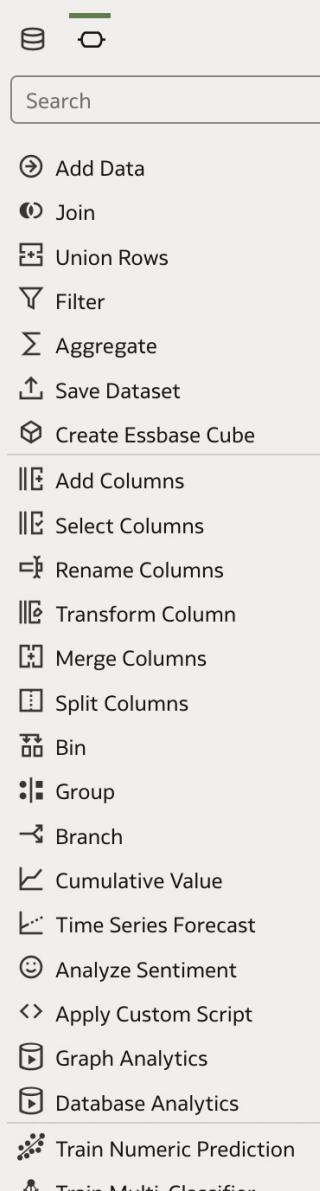
### ▶ Parameters

Cancel

Register



Connect to Your Data



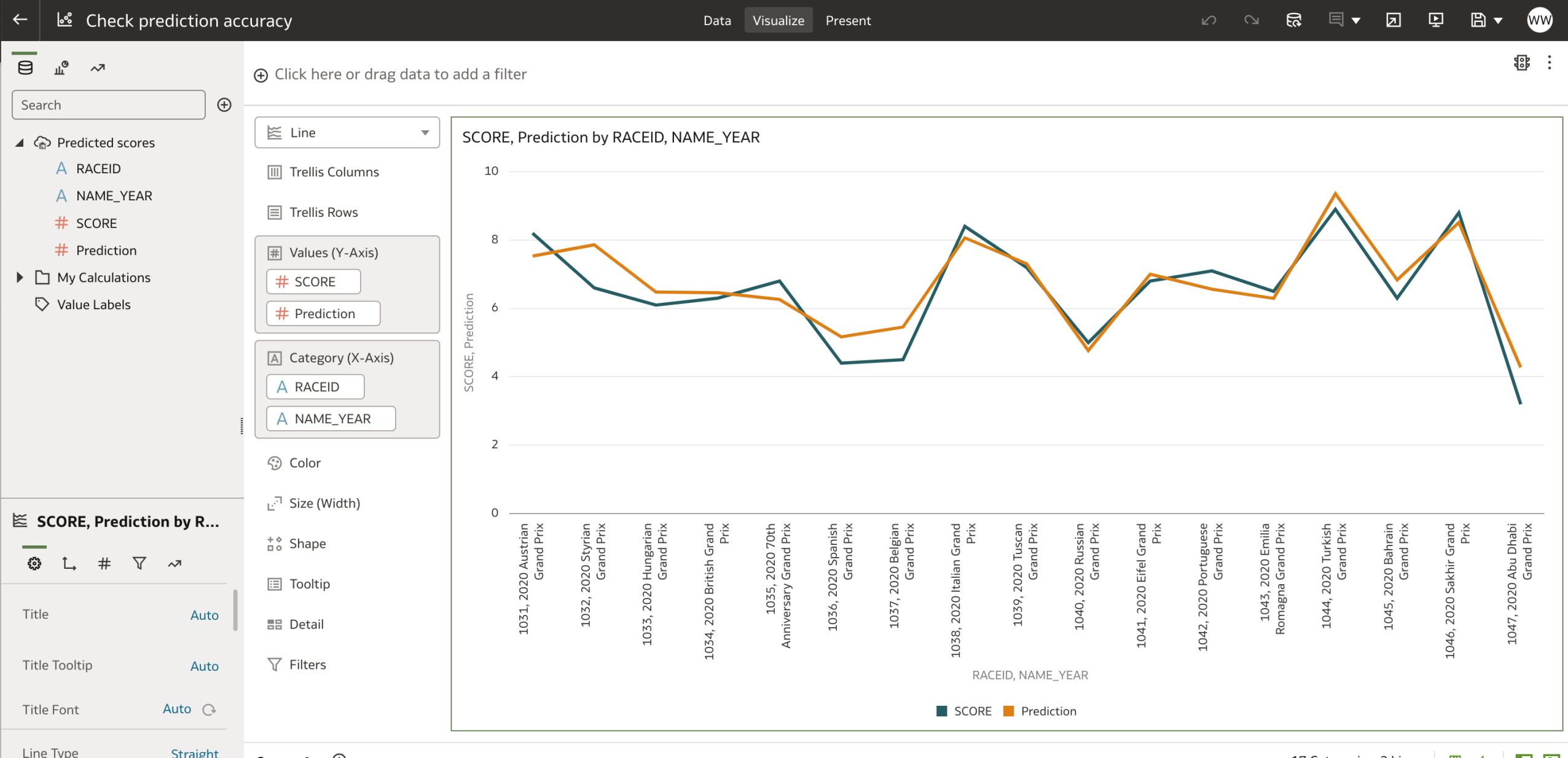
## Filter

## YEAR

2,020 - 2,020

99 RACEID	99 YEAR	99 ROUND	ab NAME	F1DATE	ab TIME	ab URL	99 SCORE	99 DNF
1041	2020	11	Eifel Grand Prix	11/10/2020 12:00:00 AM	13:10:00	<a href="https://en.wikipedia.org/wiki/2020_Eifel_Grand_Prix">https://en.wikipedia.org/wiki/2020_Eifel_Grand_Prix</a>	6.8	5
1038	2020	8	Italian Grand Prix	06/09/2020 12:00:00 AM	13:10:00	<a href="https://en.wikipedia.org/wiki/2020_Italian_Grand_Prix">https://en.wikipedia.org/wiki/2020_Italian_Grand_Prix</a>	8.4	4
1046	2020	16	Sakhir Grand Prix	06/12/2020 12:00:00 AM	17:10:00	<a href="https://en.wikipedia.org/wiki/2020_Sakhir_Grand_Prix">https://en.wikipedia.org/wiki/2020_Sakhir_Grand_Prix</a>	8.8	3
1039	2020	9	Tuscan Grand Prix	13/09/2020 12:00:00 AM	13:10:00	<a href="https://en.wikipedia.org/wiki/2020_Tuscan_Grand_Prix">https://en.wikipedia.org/wiki/2020_Tuscan_Grand_Prix</a>	7.2	8





# 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/bancosabadell2022>

## 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

Lab instructions

<https://bit.ly/bancosabadell2022>

4: Evaluate the accuracy of the model for 2021

Rol: Data Visualizer

Navigator Worksheets

WORKSHOP40

Tables

Search...

- ▶ DM\$P5GLMR\_7969DE44D0
- ▶ DM\$P5GLM\_0418DDD523
- ▶ DM\$P5NN\_F28CEE823D
- ▶ DM\$P5OPE\$DM2\_114\_00071
- ▶ DM\$P5SVMG\_DA5C86CCD7
- ▶ DM\$P5VML\_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	score
1	1064	2021	13	Dutch Grand Prix		9/5/2021, 12:00:00	13:00:00	<a href="http://en.wikipedia.org">http://en.wikipedia.org</a>	
2	1071	2021	19	Brazilian Grand Pri		11/14/2021, 12:00:00	17:00:00	<a href="http://en.wikipedia.org">http://en.wikipedia.org</a>	
3	1059	2021	7	French Grand Prix		6/20/2021, 12:00:00	13:00:00	<a href="http://en.wikipedia.org">http://en.wikipedia.org</a>	
4	1061	2021	10	British Grand Prix		7/18/2021, 12:00:00	14:00:00	<a href="http://en.wikipedia.org">http://en.wikipedia.org</a>	
5	1065	2021	14	Italian Grand Prix		9/12/2021, 12:00:00	13:00:00	<a href="http://en.wikipedia.org">http://en.wikipedia.org</a>	
6	1070	2021	18	Mexican Grand Pri		11/7/2021, 12:00:00	19:00:00	<a href="http://en.wikipedia.org">http://en.wikipedia.org</a>	
7	1066	2021	15	Russian Grand Prix		9/26/2021, 12:00:00	12:00:00	<a href="http://en.wikipedia.org">http://en.wikipedia.org</a>	
8	1055	2021	4	Spanish Grand Prix		5/9/2021, 12:00:00	13:00:00	<a href="http://en.wikipedia.org">http://en.wikipedia.org</a>	
9	1067	2021	16	Turkish Grand Prix		10/10/2021, 12:00:00	12:00:00	<a href="http://en.wikipedia.org">http://en.wikipedia.org</a>	

Navigator

Worksheets



WORKSHOP40

Tables

Search...



- ▶ DM\$P5GLMR\_7969DE44D0
- ▶ DM\$P5GLM\_0418DDD523
- ▶ DM\$P5NN\_F28CEE823D
- ▶ DM\$P5OPE\$DM2\_114\_00071
- ▶ DM\$P5SVMG\_DA5C86CCD7
- ▶ DM\$P5VML\_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, 12:00:00	17:00:00	http://en.wikipedia.	(null)	
3	2021		7	French Grand Prix	6/20/2021, 12:00:00	13:00:00	http://en.wikipedia.	(null)	
4	2021		10	British Grand Prix	7/18/2021, 12:00:00	14:00:00	http://en.wikipedia.	(null)	
5	2021		14	Italian Grand Prix	9/12/2021, 12:00:00	13:00:00	http://en.wikipedia.	(null)	
6	2021		18	Mexican Grand Pri	11/7/2021, 12:00:00	19:00:00	http://en.wikipedia.	(null)	
7	2021		15	Russian Grand Prix	9/26/2021, 12:00:00	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, 12:00:00	12:00:00	http://en.wikipedia.	(null)	



- Search
- Add Data
- Join
- Union Rows
- Filter
- Aggregate
- Save Dataset
- Create Essbase Cube
- Add Columns
- Select Columns
- Rename Columns
- Transform Column
- Merge Columns
- Split Columns
- Bin
- Group
- Branch
- Cumulative Value
- Time Series Forecast
- Analyze Sentiment
- Apply Custom Script
- Graph Analytics
- Database Analytics
- Train Numeric Prediction
- Train Multi Classifier

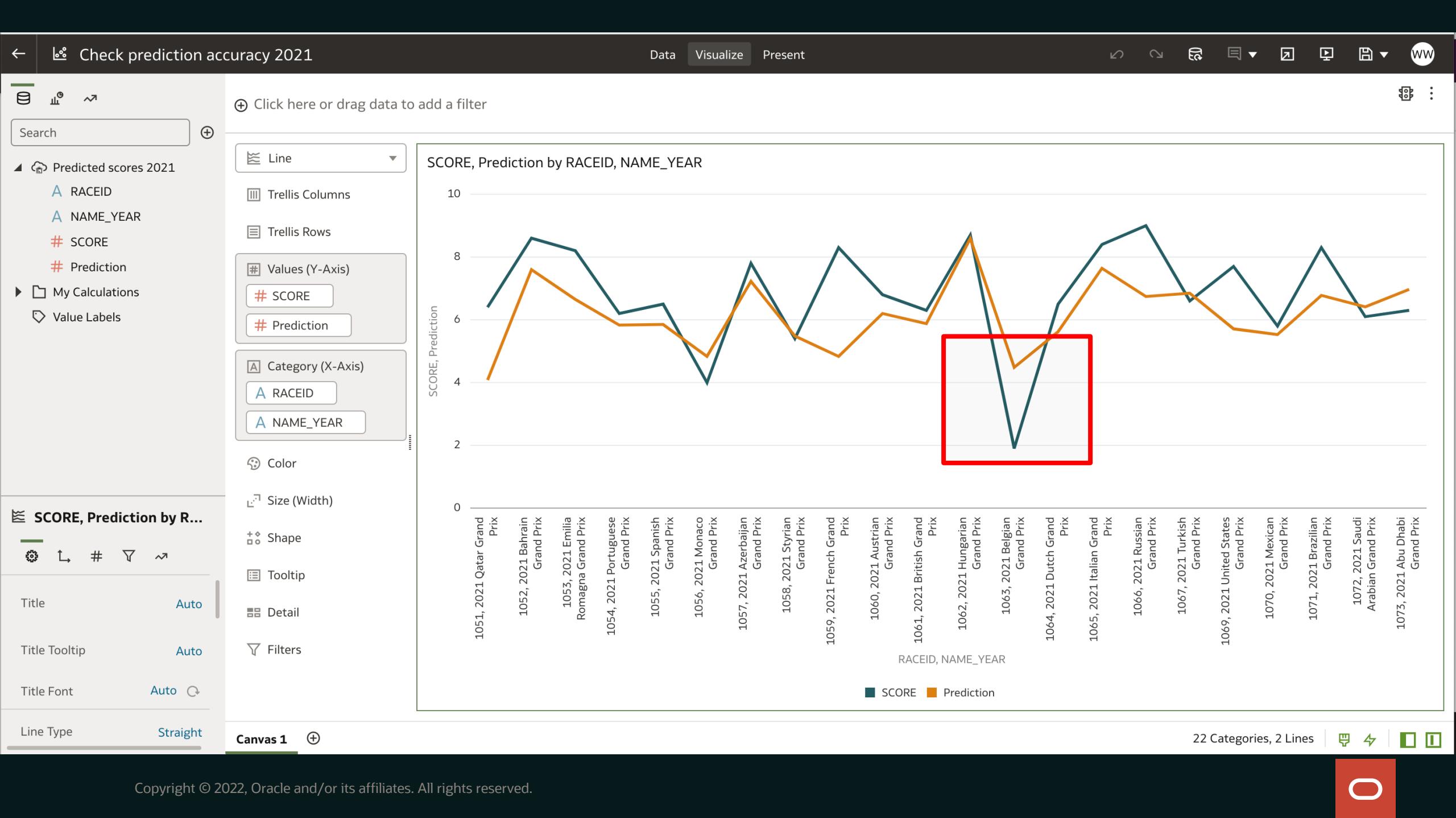


## Filter

YEAR  
2,021 - 2,021

99 RACEID	99 YEAR	99 ROUND	ab NAME	⌚ F1DATE	ab TIME	ab URL	99 SCORE	99 DNF_CNT
1064	2021	13	Dutch Grand Prix	05/09/2021 12:00:00 AM	13:00:00	<a href="http://en.wikipedia.org/wiki/2021_Dutch_Grand_Prix">http://en.wikipedia.org/wiki/2021_Dutch_Grand_Prix</a>	6.5	3
1071	2021	19	Brazilian Grand Prix	14/11/2021 12:00:00 AM	17:00:00	<a href="http://en.wikipedia.org/wiki/S%C3%A3o_Paulo_Grand_Prix">http://en.wikipedia.org/wiki/S%C3%A3o_Paulo_Grand_Prix</a>	8.3	2
1059	2021	7	French Grand Prix	20/06/2021 12:00:00 AM	13:00:00	<a href="http://en.wikipedia.org/wiki/2021_French_Grand_Prix">http://en.wikipedia.org/wiki/2021_French_Grand_Prix</a>	8.3	0
1061	2021	10	British Grand Prix	18/07/2021 12:00:00 AM	14:00:00	<a href="http://en.wikipedia.org/wiki/2021_British_Grand_Prix">http://en.wikipedia.org/wiki/2021_British_Grand_Prix</a>	6.3	2





# Thank You

---

ORACLE

