

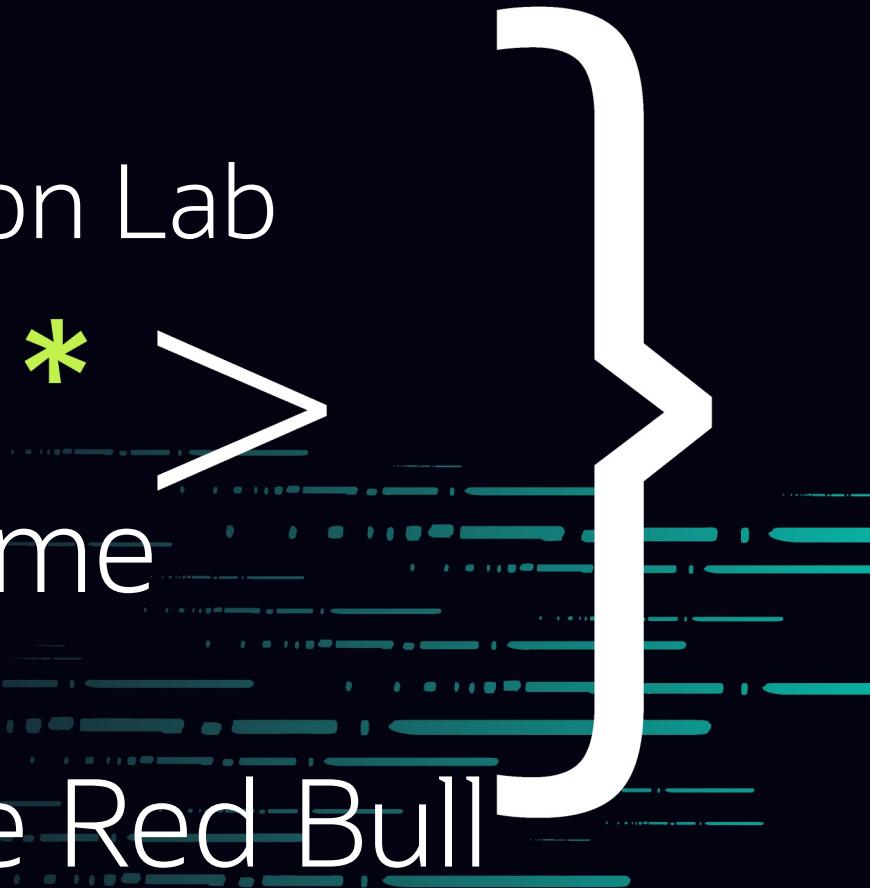
# Welcome at JFOKUS 2022 Hands-on Lab

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



# Welcome at JFOKUS 2022 Hands-on Lab

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





Copyright © 2022, Oracle and/or its affiliates

# 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





Sign in

CAR

DRIVER

TRACK

# MAX VERSTAPPEN



Netherlands

Rank

**3**



Wins  
**10**



Poles  
**3**



Podiums  
**42**



Fastest laps  
**10**

News lorem ipsum



Verstappen not getting distracted  
by Red Bull's 'favourites' tag



Hamilton vs Verstappen: The 3  
key factors that will define...



**33**

Driven car



Most recent races lorem



2020 Abu Dhabi  
Grand Prix



2020 Abu Dhabi  
Grand Prix



2020 Abu Dhabi  
Grand Prix

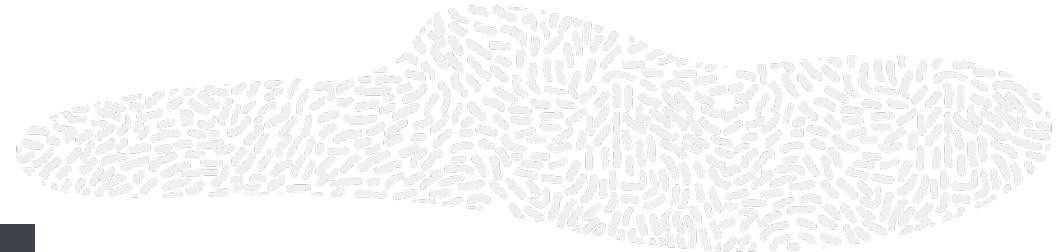
Next race in:

**5**  
DAYS

**20**  
MINUTES

**48**  
SECONDS

# 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](https://bit.ly/quiz_redbull_JFOKUS22)

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.



# The Team



Olivier Perard

Principal Data Scientist



Ettore Volpicelli

Developer Marketing Director EMEA

Oracle

<https://github.com/operard/jfokus2022>

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

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?



© 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](#)



# 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

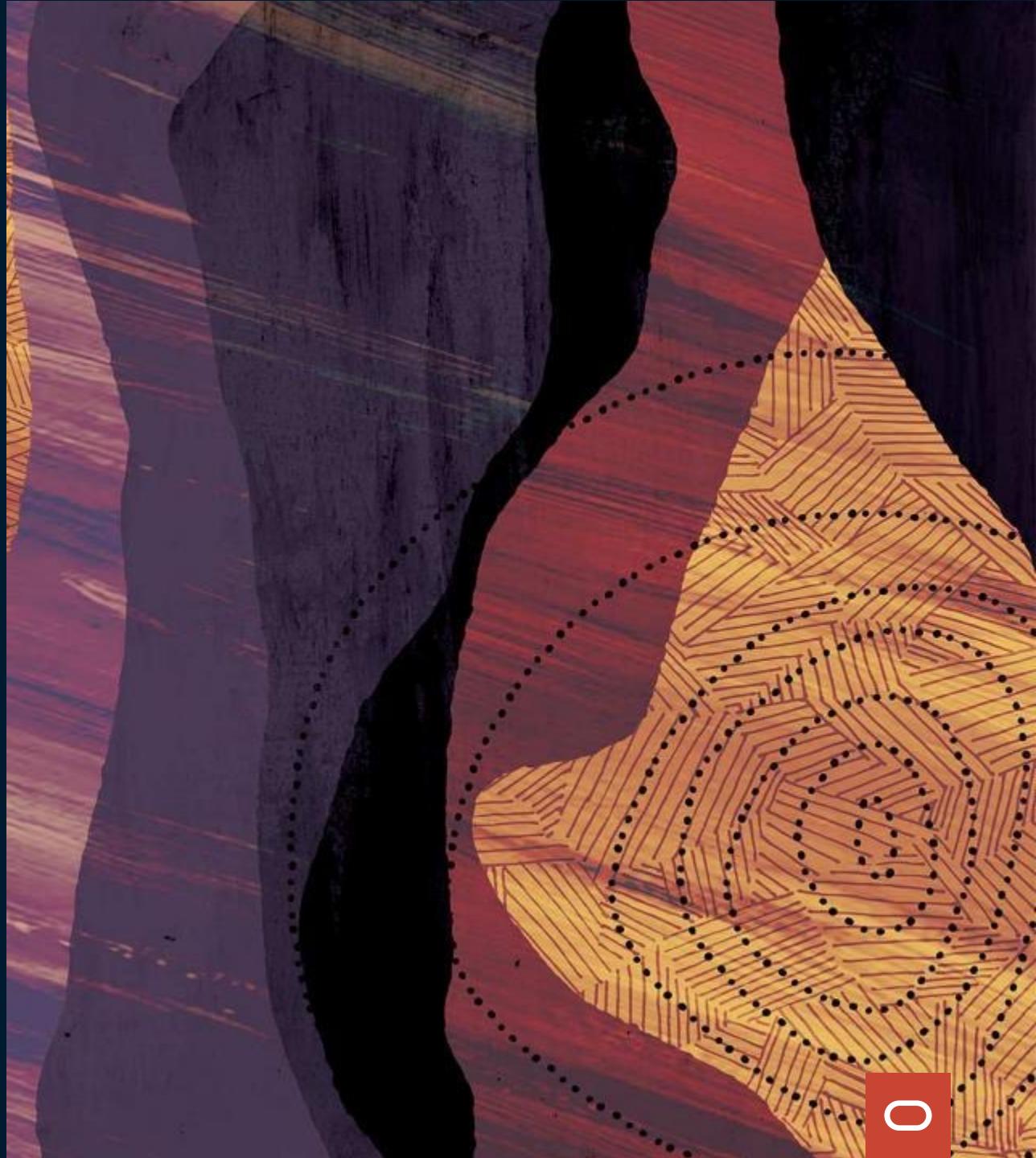
Lab 2 Basic data exploration

Lab 3 Advanced data exploration

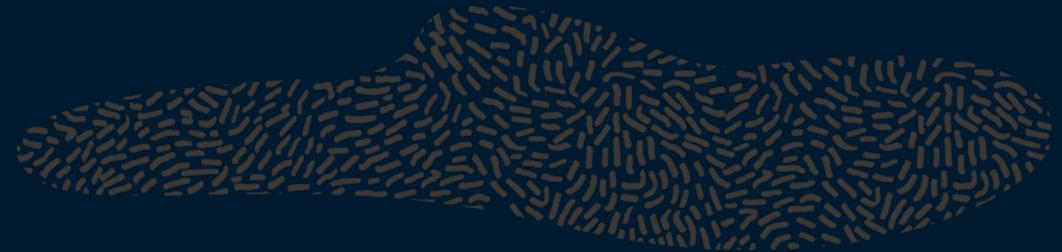
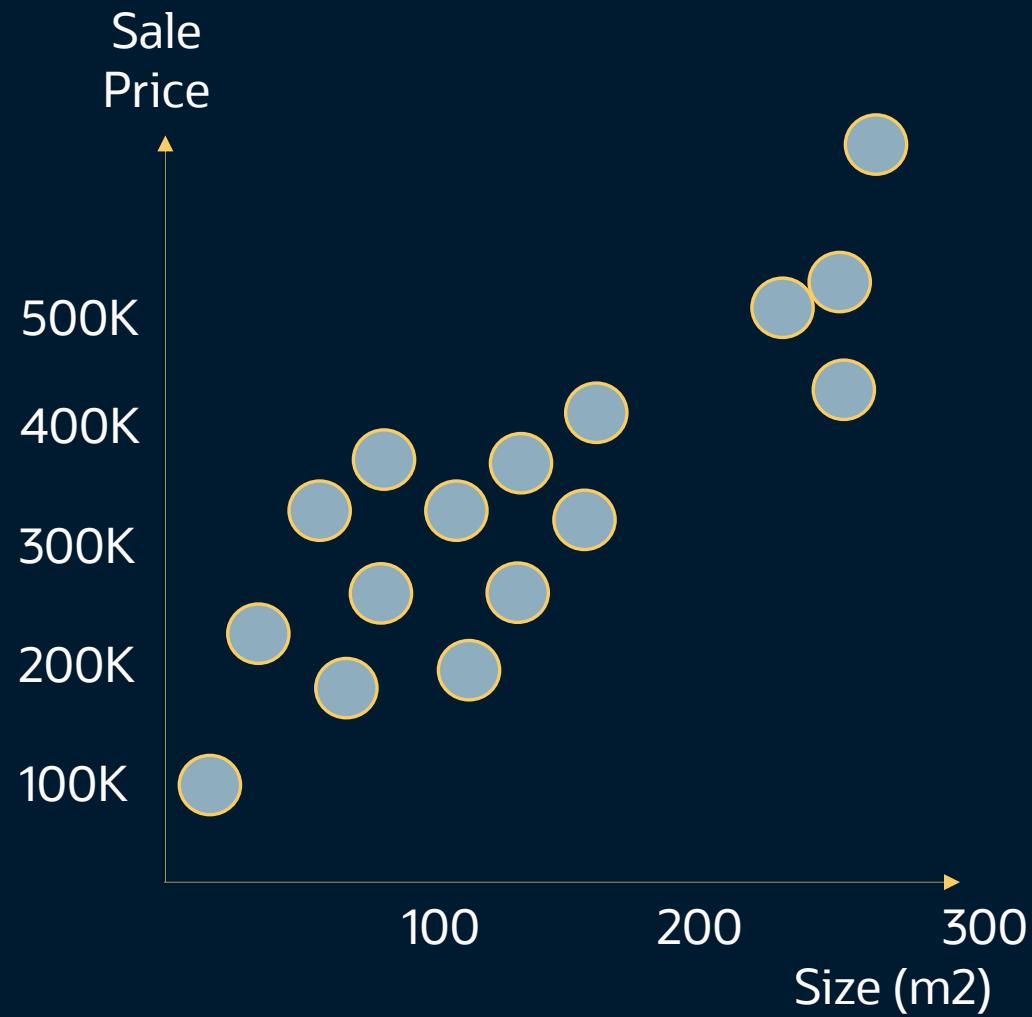
Lab 4 Feature generation, machine learning, evaluation

Lab 5 Process F1 2022 Dataset from racefans.net

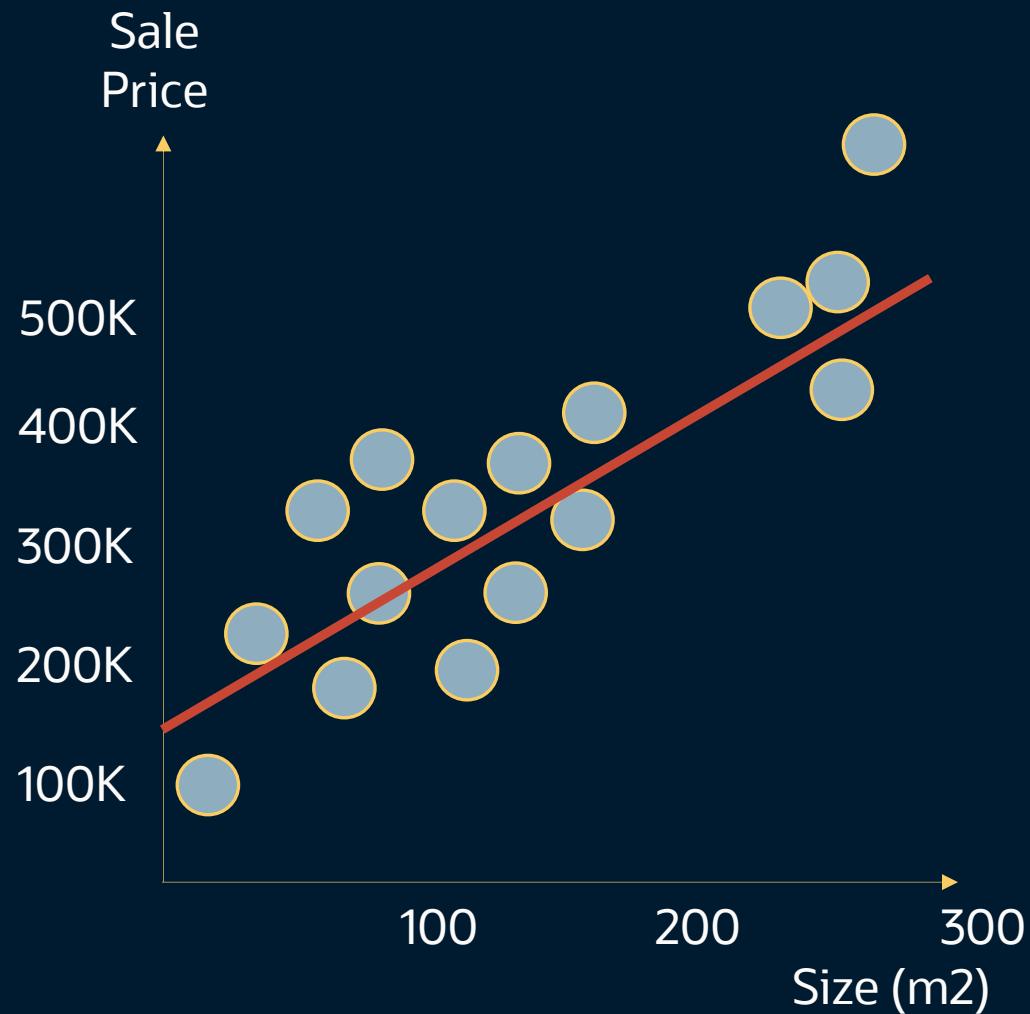
Machine Learning automatically sifts through large amounts of data to discover hidden patterns, new insights and make predictions.



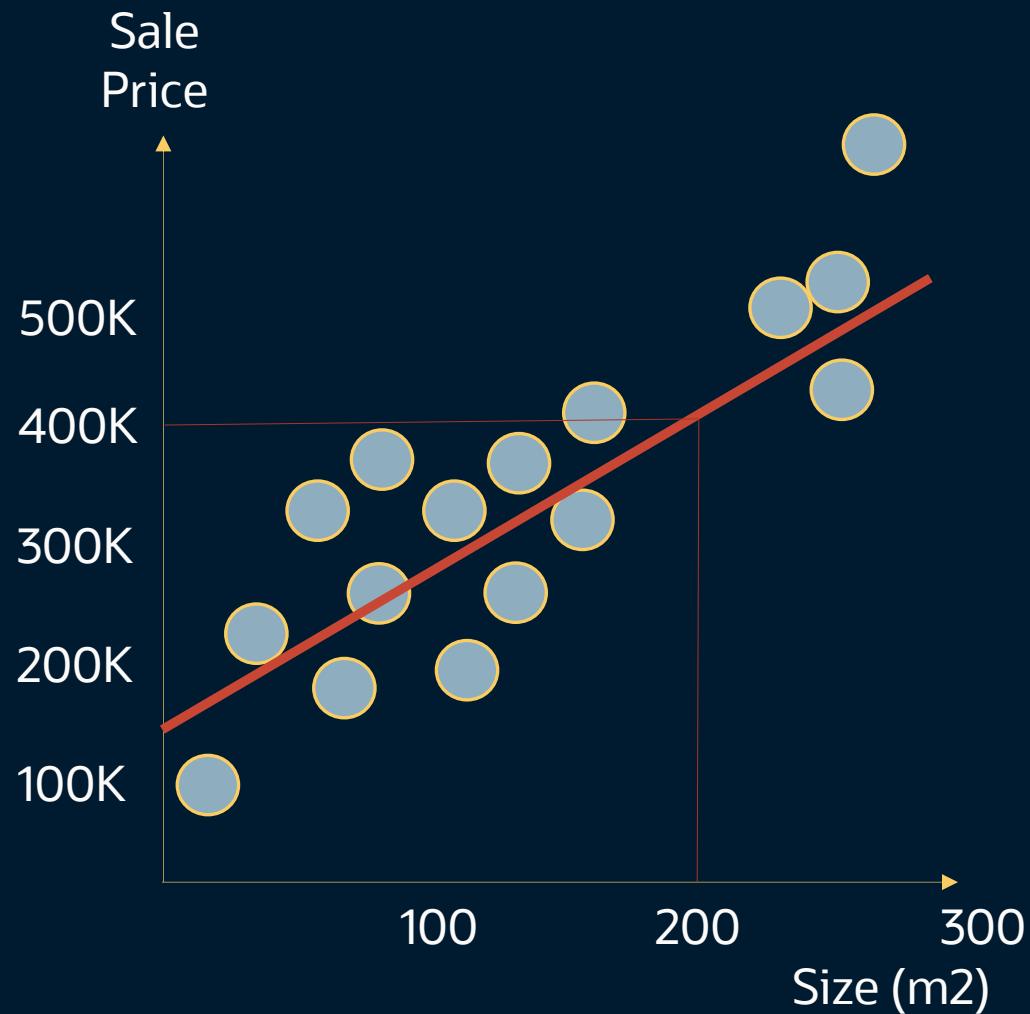
## Example of a model



## 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

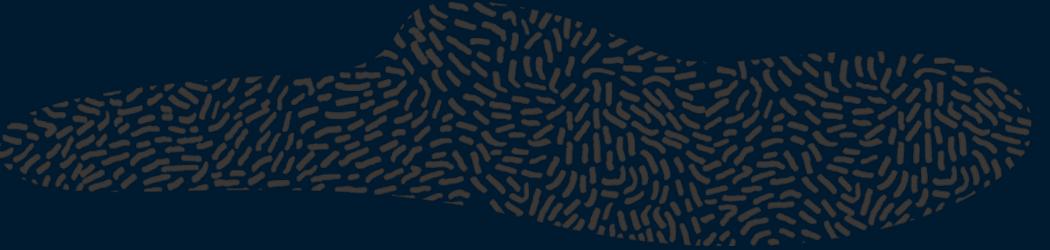
Name	Round	Wet weather	Over taken positions	Did-Not-Finish Count	...	Score (racefans.net)
2020 Hungarian Grand Prix	3	Y	222	1	...	6.1
2020 Styrian Grand Prix	2	N	131	3	...	6.6
2020 Austrian Grand Prix	1	N	92	9	...	8.2
2010 Malaysian Grand Prix	3	N	87	7	...	6.684
2010 Australian Grand Prix	2	Y	120	10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

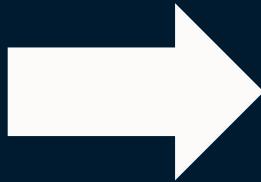
Name	Round	Wet weather	Over taken positions	Did-Not-Finish Count	...	Score (racefans.net)
2020 Hungarian Grand Prix	3	Y	222	1	...	6.1
2020 Styrian Grand Prix	2	N	131	3	...	6.6
2020 Austrian Grand Prix	1	N	92	9	...	8.2
2010 Malaysian Grand Prix	3	N	87	7	...	6.684
2010 Australian Grand Prix	2	Y	120	10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



Number of overtakes

?

Weather



Score

Circuit (location)

DNFs

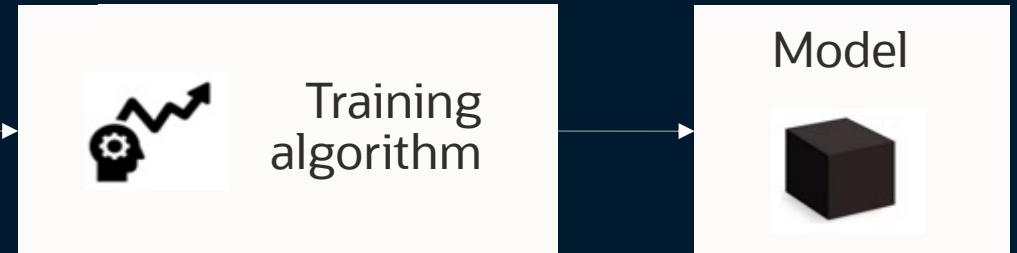
...

## 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



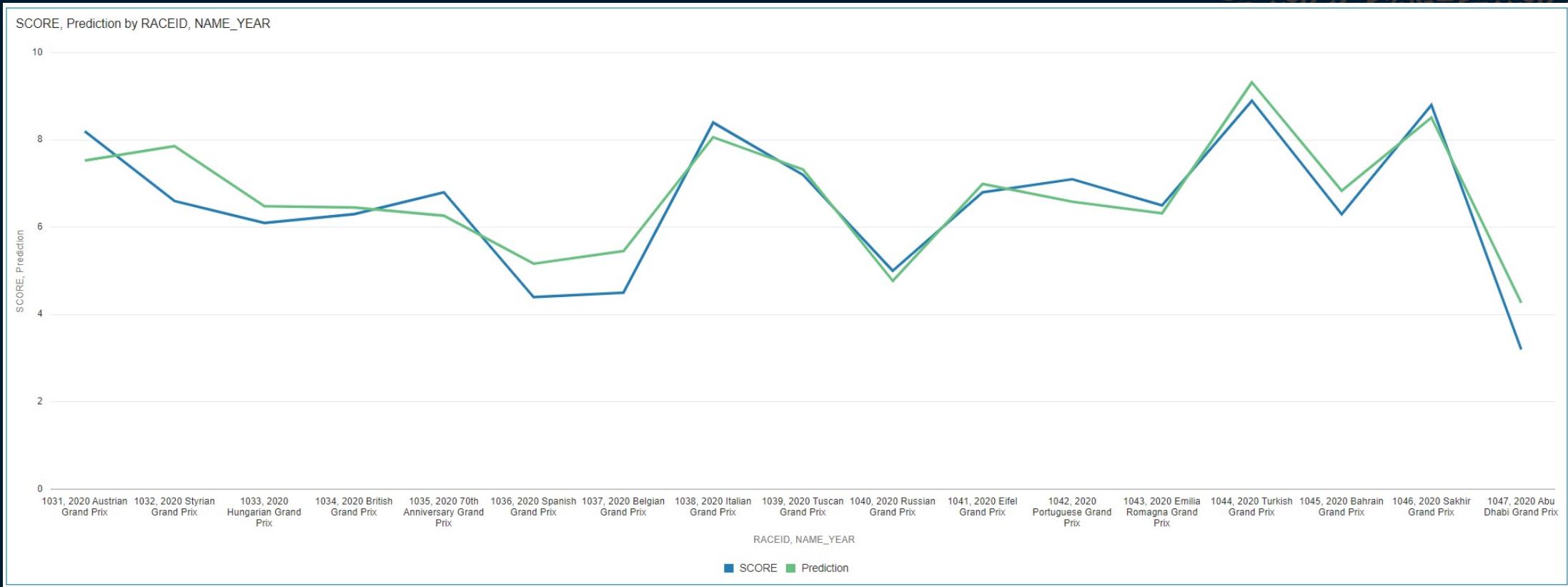
## 2. Test (data 2020)

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

Input features test (factors)



# The end goal



# 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



## 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



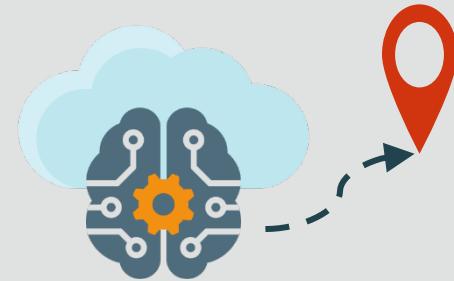
## Oracle Machine Learning

Bringing the algorithms to the data



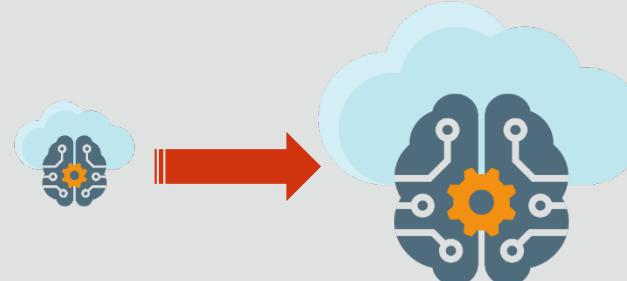
# Trend: AI & ML - Analytics Machine Learning & AI

Increase productivity, Achieve enterprise goals, Innovate more



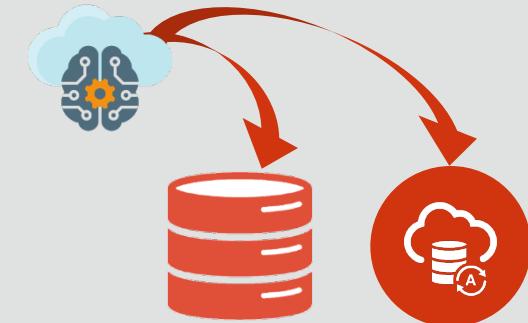
## Automated

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



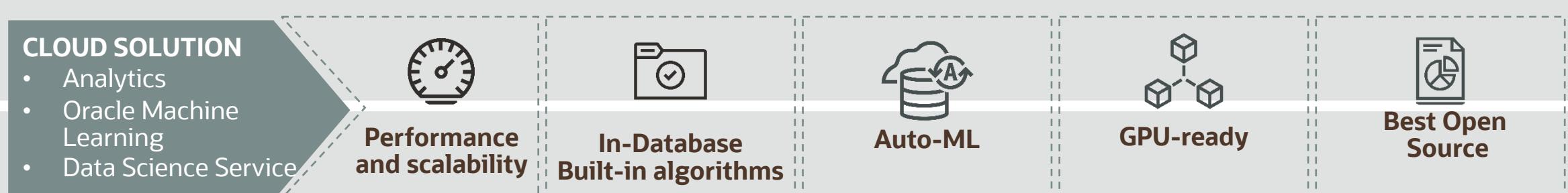
## Scalable

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



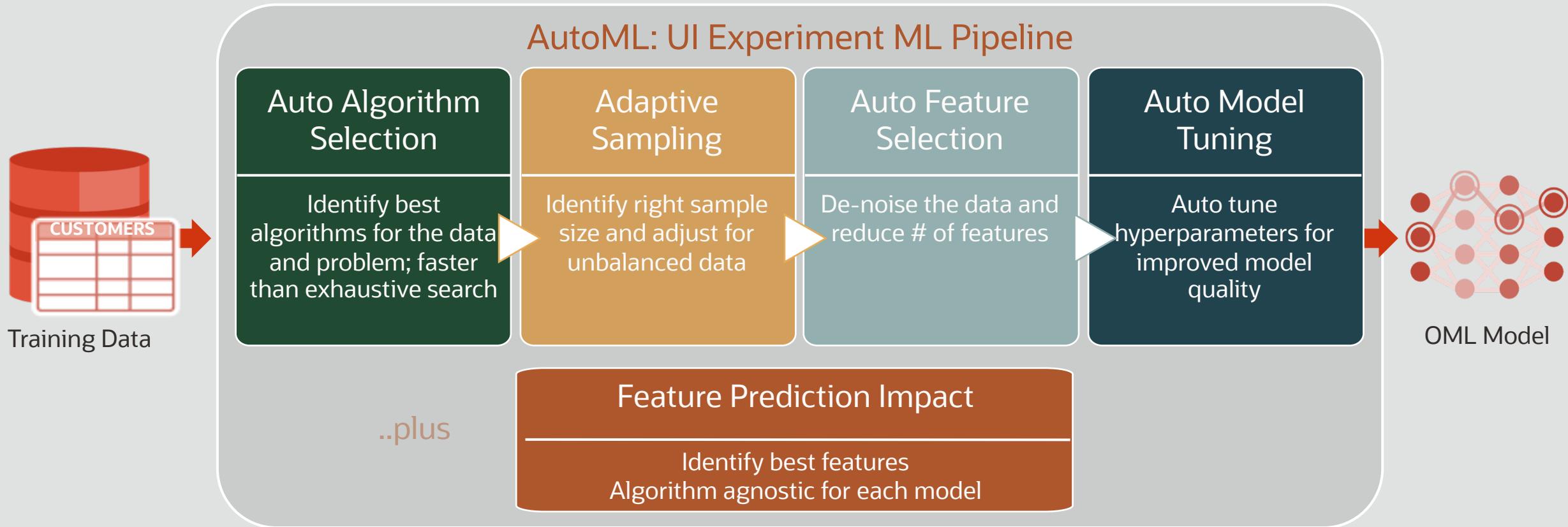
## Production-ready

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



# Auto Machine Learning

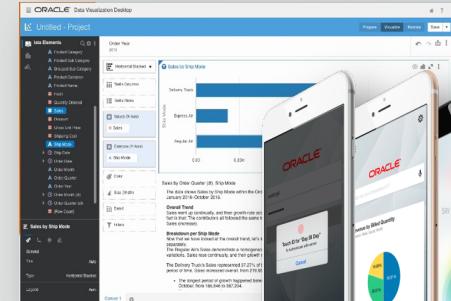
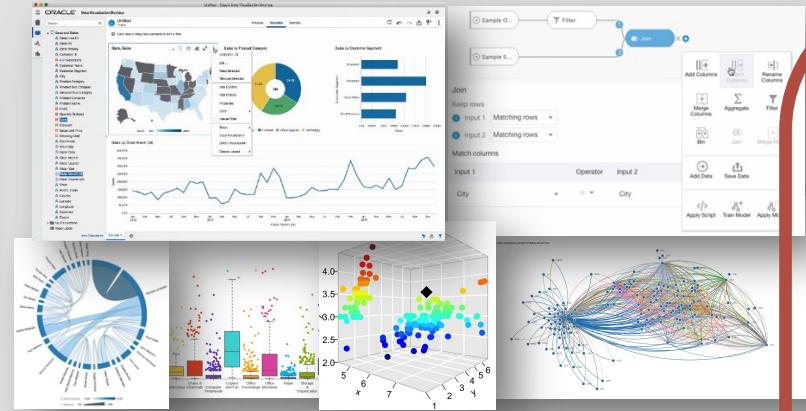
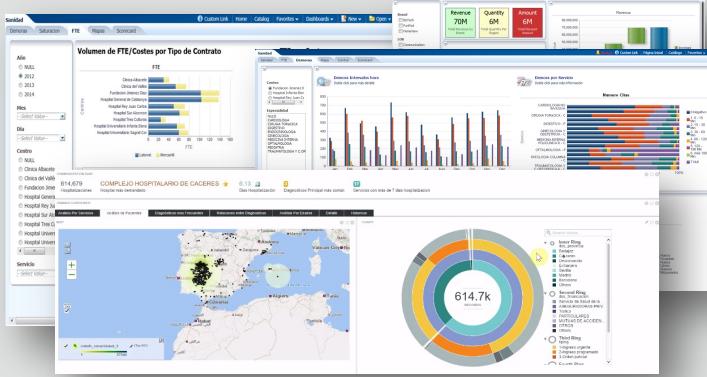
Faster and Easier Machine Learning for Data Scientists, Non-Experts and Developers



ENABLES NON-EXPERT USERS TO LEVERAGE MACHINE LEARNING

# Oracle Analytics Cloud

## ML Democratization to End Users or Citizen Data Scientist



### Governed Analytics

Dashboards

Pixel-Perfect Reports

Data Preparation

Data Visualization

Semantic Models

Query Federation

Heterogeneous Data Sources

Collaboration

Role-Based Access Control

Excel Integration

Storytelling

Mobile

### Augmented Analytics

Voice & Chatbot

Natural Language

Data Enrichment

1-Click Explain

Adaptive & Personalized

Machine Learning

# Machine Learning

Evolve towards “cititizen data scientist”



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

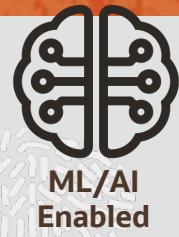
OAC consumes ML  
algorithms from Oracle  
Data Base

The screenshot displays the Oracle Analytics Cloud (OAC) interface for Machine Learning. On the left, a sidebar lists various data flow steps and machine learning algorithms. The main workspace shows a 'Machine Learning' project with a 'Test Data ...' button. Below it, a configuration panel for 'Add Data - Test Data Set' is shown, listing columns from an uploaded Excel file. A preview table at the bottom shows two rows of data.

Order Line ID	Order ID	Order Priority	Customer ID	Customer Name	Customer Segment	City	Product Category
7183	90930	Medium	2626	Lilian Fischer	Consumer	Riyadh	Office Supplies
580	87232	High	221	Colleen Wilcox	Home Office	Sheffield	Office Supplies

# Augmented Analytics

AI for knowledge democratization



Natural language  
conversation, voice indeed

**Video**

Proactivity base on your interest, hour of day or localization

**Video 1**

**Video 2**

**Video**

One click Machine Learning

**Video**

Data enrichment suggestion

# 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

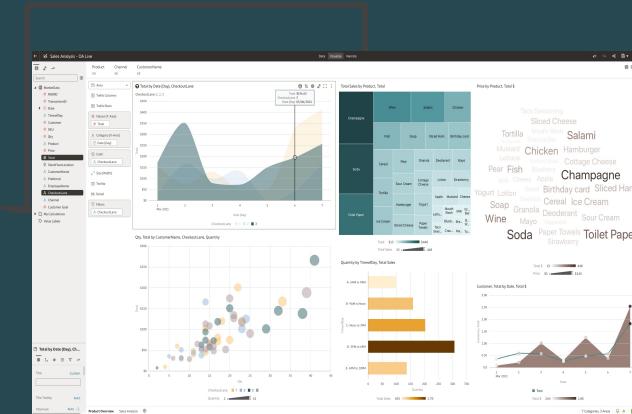
Lab 5 Process F1 2022 Dataset from racefans.net

## Autonomous Data Warehouse



Data Loading  
Model training

## Analytics Cloud



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

## Autonomous Data Warehouse

workshop1

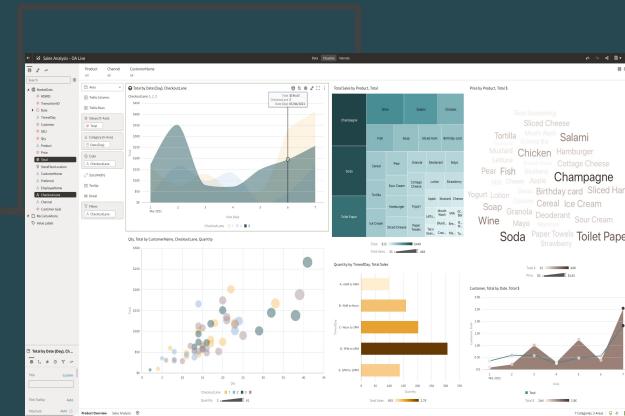
workshop2

workshop3

workshop...

workshop40

## Analytics Cloud



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

## Users

<span>+ Create</span> <span>- Delete</span> <input type="checkbox"/> Show All Users						<input type="text"/> 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	

<b>Firstname</b>	<b>Surname</b>	<b>Email</b>	<b>Workshop USER</b>
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 Lönnberg	Thunderkick AB	mattias.lonnerberg@thunderkick.com	WORKSHOP16
Ilian Corneliusen	Thunderkick AB	ilian.corneliusen@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

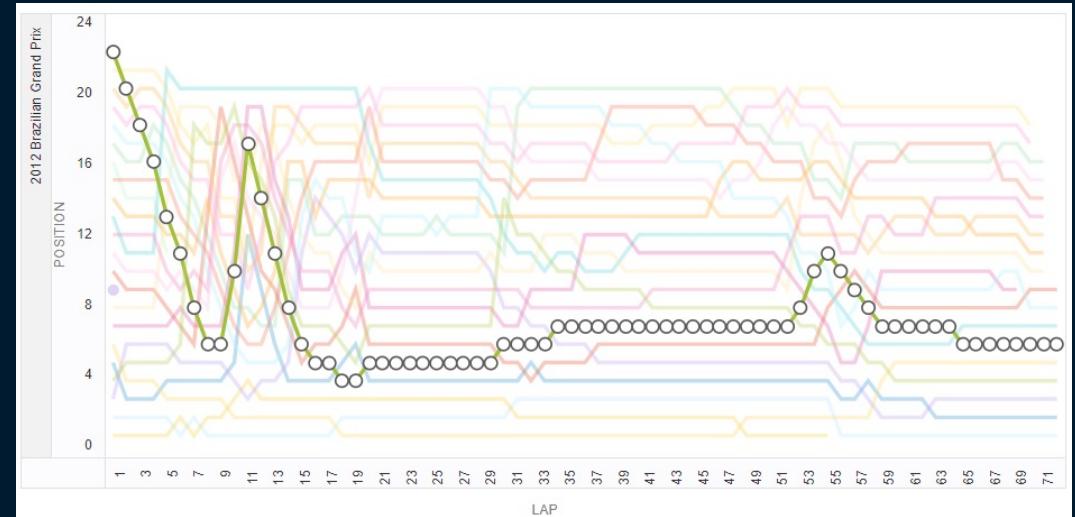
Lab 5 Process F1 2022 Dataset from racefans.net

# 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>

## 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 ↑

WORKSHOP39

WORKSHOP4

WORKSHOP40

WORKSHOP5

WORKSHOP6

Table ↑

OVERTAKEN\_POSITIONS\_TOTAL

PIT\_STOPS

POSITION\_DIFFERENCE\_TOP\_5\_V

RACES

RACES\_TRAIN

Cancel

OK

## Create Experiment

▷ Start ▾ Cancel  Save

**Name**

Predict race score

**Comments**

Predict race score

**Data Source**

WORKSHOP40.RACES\_TRAIN

**Prediction Type**

Select Prediction Type

**Predict**

Select Prediction Target



Required

**Case ID**

Select Case ID



### › Additional Settings

### ⌄ Features

&lt;- Experiments

## Predict race score

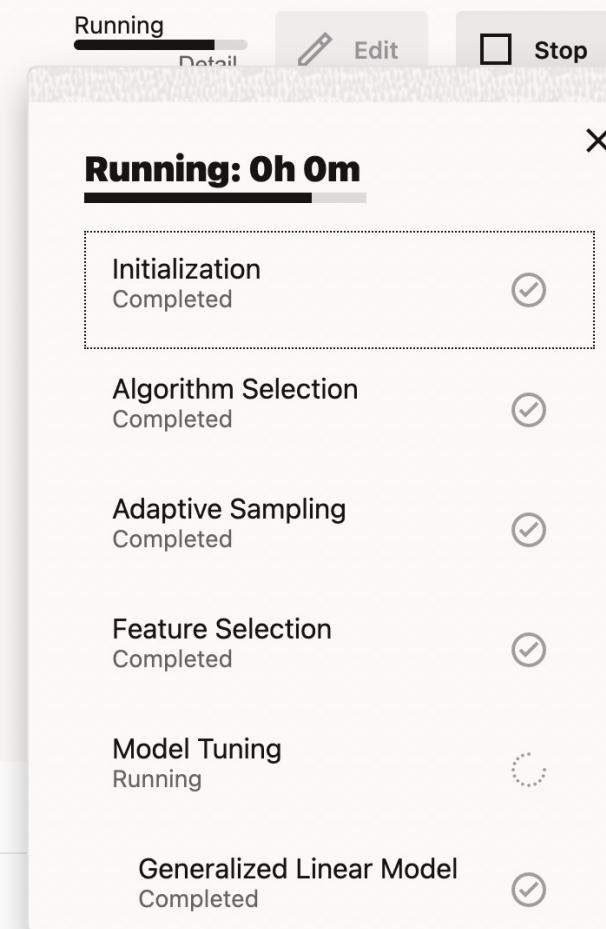
&gt; Settings

### 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

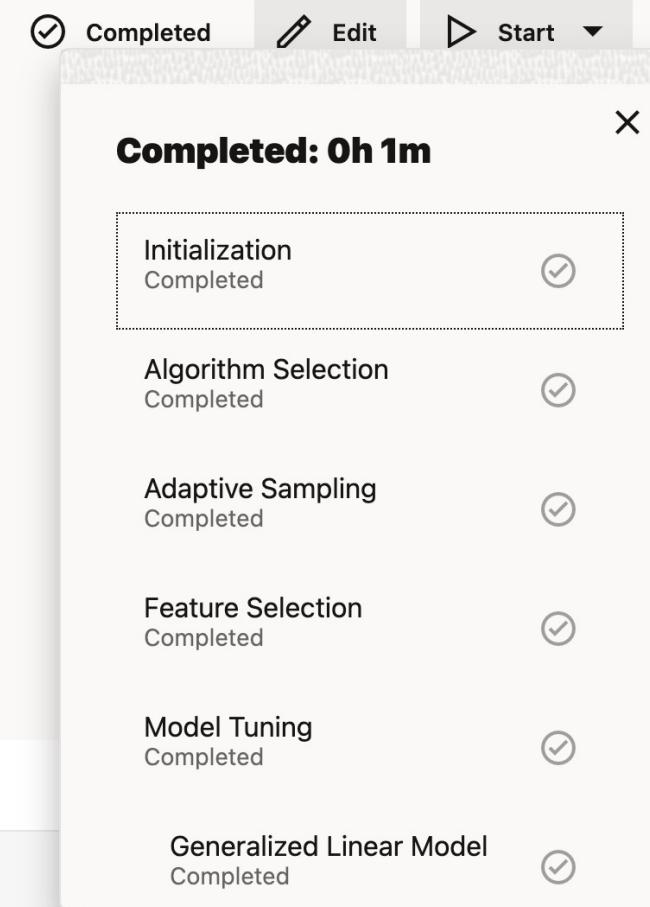
&gt; Settings

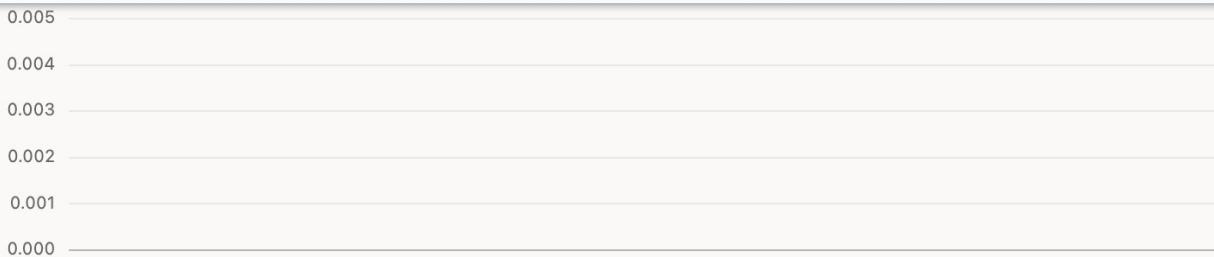
### Negative Mean Absolute Error



### Leader Board

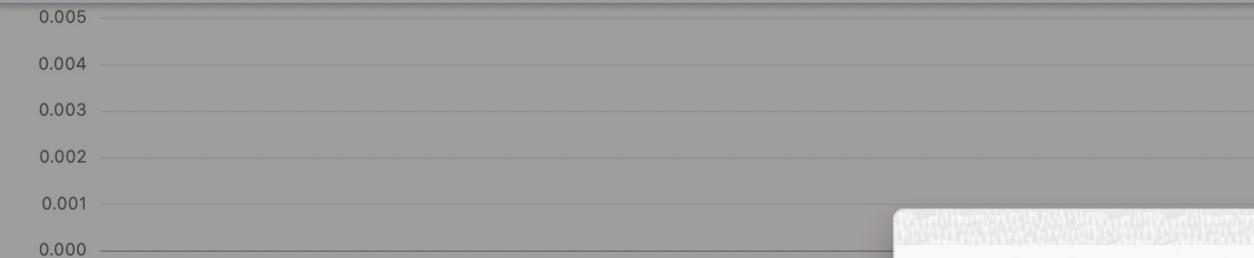
Deploy	Rename	Create Notebook	Metrics
Algorithm	Model Name	Negative Mean Absolute Error	
Neural Network	NN_F28CEE823D	-0.6774	
Support Vector Machine (Linear)	SVML_552C7B93B5	-0.6360	
Generalized Linear Model (Ridge ...)	GLMR_7969DE44D0	-0.6348	
Support Vector Machine (Gaussi...	SVMG_DA5C86CCD7	-0.6346	





## Leader Board

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

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

MAX\_COEMBACKSCORE

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: 80%;">██████████</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_COEMBACKSCORE	<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 ↑↓
CIRCUITREF	0.00	0.00	0.00	0.00	0.00		4.19	2.14
DNF_COUNT	0.00	0.00	0.00	0.00	0.00		12.2	3.3
MAX_COEMBACKSCORE	0.00	0.00	0.00	0.00	0.00		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

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

0.002  
0.001  
0.000

## 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

**Deploy Model - GLM\_0418DDD523**

Name \_\_\_\_\_

GLM\_0418DDD523

UR

w40\_glm

Version

1.0

## Namespace

Shared



Canc

OK

Name ↑↓

Name ↑↓

## WEATHER\_WET

## 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"/>	<a href="#">GLM_0418DDD523</a>				1.0				WORKSHOP40		01/05/2022 ...		<a href="#">w40_glm</a>	

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

# 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** Search 

Type

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

Name

 GLM\_0418DDD523

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

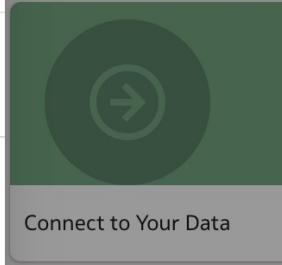
## ▶ Input Columns

## ▶ Output Columns

## ▶ Parameters

Cancel

Register



Connect to Your Data

# 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\$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	http://en.wikipedia.org	
2	1071	2021		19	Brazilian Grand Prix	11/14/2021, 12:00:00	17:00:00	http://en.wikipedia.org	
3	1059	2021		7	French Grand Prix	6/20/2021, 12:00:00	13:00:00	http://en.wikipedia.org	
4	1061	2021		10	British Grand Prix	7/18/2021, 12:00:00	14:00:00	http://en.wikipedia.org	
5	1065	2021		14	Italian Grand Prix	9/12/2021, 12:00:00	13:00:00	http://en.wikipedia.org	
6	1070	2021	14	18	Mexican Grand Prix	11/7/2021, 12:00:00	19:00:00	http://en.wikipedia.org	
7	1066	2021		15	Russian Grand Prix	9/26/2021, 12:00:00	12:00:00	http://en.wikipedia.org	
8	1055	2021		4	Spanish Grand Prix	5/9/2021, 12:00:00	13:00:00	http://en.wikipedia.org	
9	1067	2021		16	Turkish Grand Prix	10/10/2021, 12:00:00	12:00:00	http://en.wikipedia.org	



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, 12:00:0	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, 12:00:0	12:00:00	http://en.wikipedia.	(null)	

## 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.

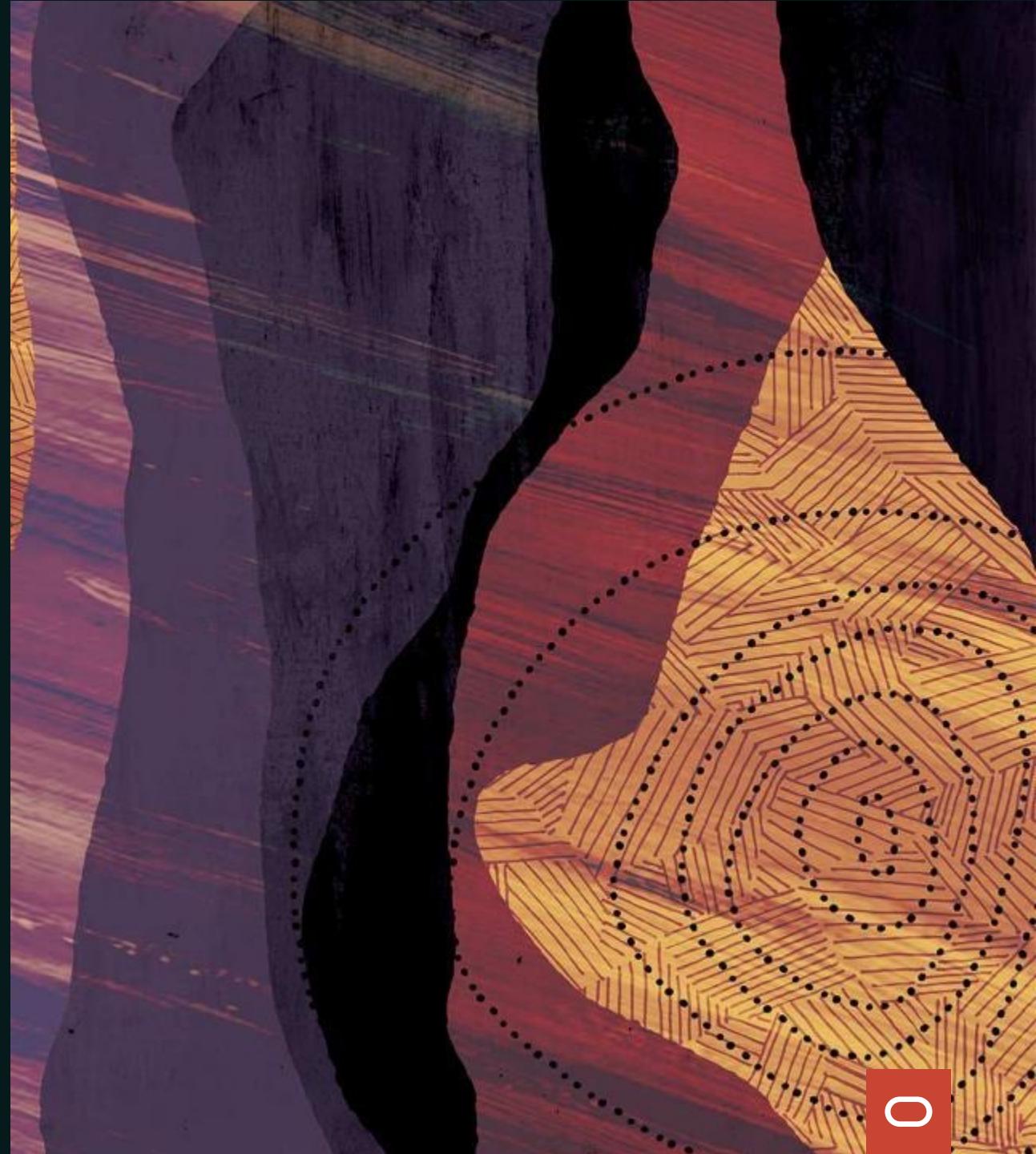
Access the quiz here: [bit.ly/quiz\\_redbull\\_JFOKUS22](https://bit.ly/quiz_redbull_JFOKUS22)

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.



# Thank You

---



ORACLE

