

Grand Prix Argentina - Race 1

¿Porqué una carrera de F1 es emocionante?

Es una pregunta tiene una respuesta diferente según a que persona le preguntemos y es un contexto ideal para hacer Data Science.

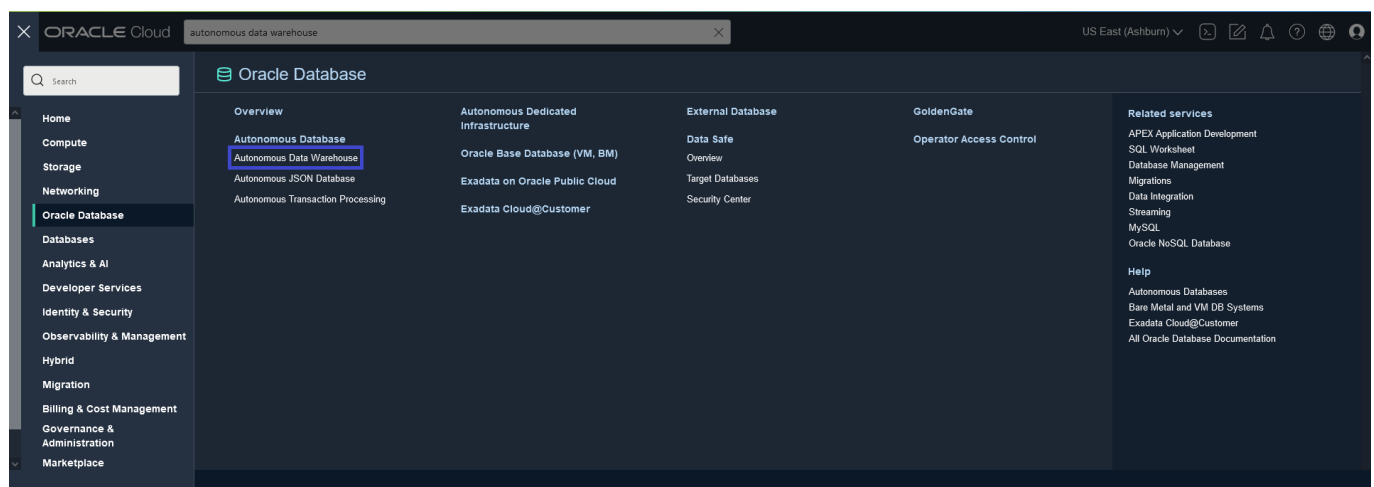
En este sentido, en esta carrera usaremos datos de la Evaluación de Experiencia de los Fanáticos de la Fórmula 1 (F1) y crearemos un experimento para entrenar 6 algoritmos de Machine Learning usando Auto Machine Learning (Auto ML) en Autonomous Data Warehouse (ADW) y con Ingeniería de Atributos entenderemos lo que la Inteligencia Artificial "piensa" de esto.

Pre-requisito: para comenzar, se asume que ya el participante tiene acceso a Oracle Cloud Infrastructure (OCI) y un ADW disponible.

Las (6) vueltas para llegar a la meta son:

1. Buscar nuestra consola de ADW [INICIO]

Desde el inicio en consola **Oracle Cloud Infrastructure (OCI)**, presionando el **Menú de Hamburguesa** ubicado arriba a la izquierda de la pantalla podemos ubicar todas las opciones de la nube:



Luego, en el menu que se despliega seleccionamos primero **Oracle Data Bases** y después **Autonomous Data Warehouse**.

Y allí veremos todas las instancias que tenemos creadas.

Oracle Cloud | autonomous data warehouse | US East (Ashburn)

Overview > Autonomous Database > Autonomous Databases

Autonomous Database

Autonomous Databases in PublicSector Compartment

Autonomous Database delivers fast performance and requires no database administration. It performs all routine database maintenance tasks while the system is running, without human intervention. Autonomous Databases located in the Oracle cloud can run on dedicated or shared infrastructure. [Learn more](#)

[Create Autonomous Database](#)

Display Name	State	Dedicated	OCPU	Storage	Workload type	Autonomous Data Guard	Created
ADWF1	Available	No	1	1 TB	Data Warehouse	—	Mon, Oct 24, 2022, 13:36:35 UTC

Displaying 3 Autonomous Databases < 1 of 1 >

Filters: Workload type: Data Warehouse, State: Any state

Tag filters: add | clear

no tag filters applied

Actualmente mi instancia se llama **ADWF1** y está Available, en este sentido, debes observar un escenario similar con el nombre de base de datos que hayas escogido.

Luego, en caso de que no veas tu actual instancia puedes verificar y ajustar:

- Tenancy. Ejemplo: US East (Ashburn).
- Compartment. Ejemplo: Public Sector.
- Workload Type. Ejemplo: Data Warehouse.

Luego dando click en la instancia **ADWF1** llegamos a los detalles:

Overview > Autonomous Database > Autonomous Database Details

ADWF1

[Database Actions](#) [DB Connection](#) [Performance Hub](#) [Manage Scaling](#) [More actions](#)

Autonomous Database information Tools Tags

General information

Database name: ADWF1
 Workload type: Data Warehouse
 Compartment: latinoamerica (root)/Argentina/PabloSierra/PublicSector
 OCID: ...h6aqq [Show](#) [Copy](#)
 Created: Mon, Oct 24, 2022, 13:36:35 UTC
 OCPU count: 1
 OCPU auto scaling: Enabled [ⓘ](#)
 Storage: 1 TB
 Storage auto scaling: Disabled [ⓘ](#)
 License type: Bring your own license (BYOL), Standard Edition
 Database version: 19c
 Lifecycle state: Available [Check database availability](#)
 Instance type: Paid
 Character set: AL32UTF8
 National character set: AL16UTF16
 Auto Start/Stop schedule: Disabled [Schedule](#)
 Mode: Read/Write [Edit](#)

Associated Services

Database Management: Not Enabled [Enable](#) [ⓘ](#)
 Operation Insights: Not Enabled [Enable](#)

APEX Instance

Infrastructure

Dedicated infrastructure: No

Autonomous Data Guard [ⓘ](#)

Status: Disabled [Enable](#)

Backup

Last automatic backup: Mon, Oct 24, 2022, 12:25:03 UTC
 Manual backup store: Not Configured

Network

Access type: Allow secure access from everywhere
 Access control list: Disabled [Edit](#)
 Mutual TLS (mTLS) authentication: Required [Edit](#) [ⓘ](#)

Maintenance [ⓘ](#)

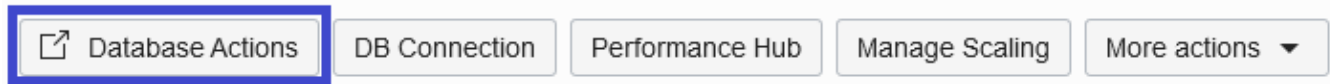
Patch level: Regular [ⓘ](#)
 Next maintenance: Sun, Oct 30, 2022, 10:00:00 UTC - 12:00:00 UTC [View History](#)
 Customer contacts: None [ⓘ](#) [Manage](#)

Data Safe [ⓘ](#)


Status: Not Registered [Register](#)

2. Crear un usuario de base de datos llamado F1 usando Database Actions.

ADWF1



En el siguiente paso nos concentraremos en las Data Base Actions, en el que haciendo click nos abrirá una nueva pestaña en el navegador y nos preguntara por nuestro usuario (ADMIN) y contraseña.

**ORACLE**
Database Actions

Username

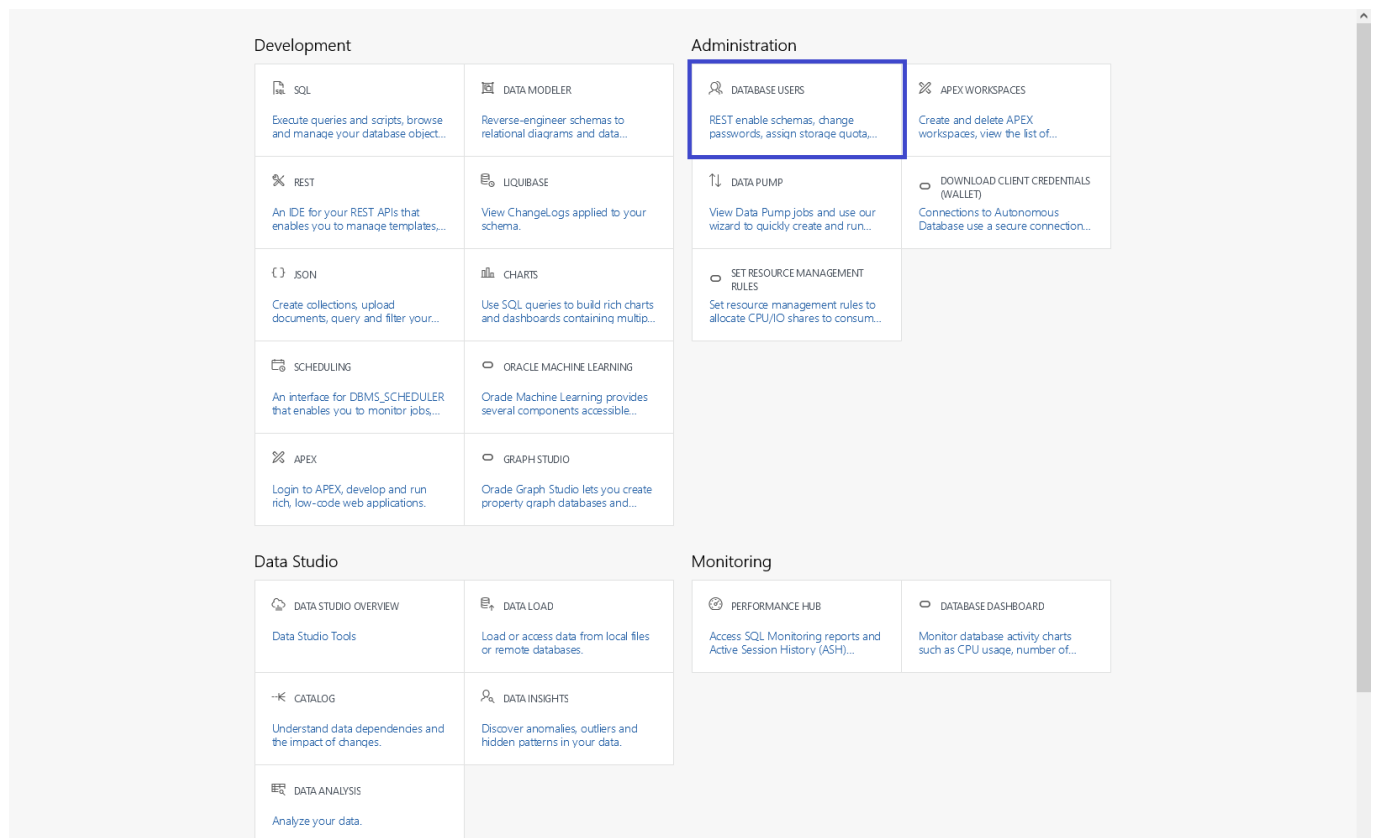
ADMIN

Password

••••••••••

Sign in

Luego, veremos la interfaz principal y seleccionaremos la opción **DATABASE USERS**.



En este apartado daremos click en el botón **+ Create New User**.

Y finalmente crearemos nuestro usuario **F1** con las opciones que vemos en pantalla:

Create User

User

5 Granted Roles

User Name *

F1

Quota on tablespace DATA

UNLIMITED

Password *

.....

Password Expired (user must change)

☐

Confirm Password *

.....

Account is Locked

☐

Graph ?

☐

OML ?

☒

Web Access ?

☒

Web access advanced features

☐ ?

☐ Show code

Create User

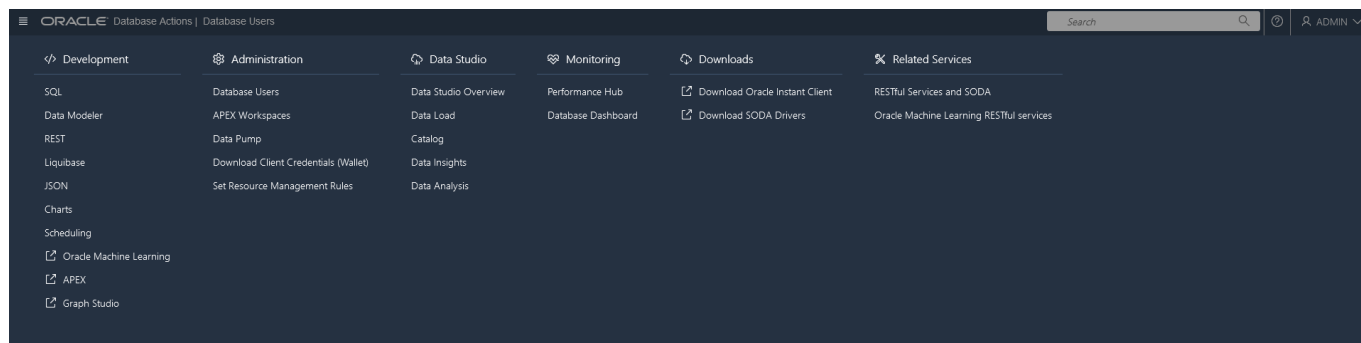
Cancel

Nota: en caso de que no veas estas pantallas, verifica que las pantallas emergentes PopUps de tu navegador este habilitadas, usualmente aparecen con un mensaje visible a la izquierda de la barra de navegación.

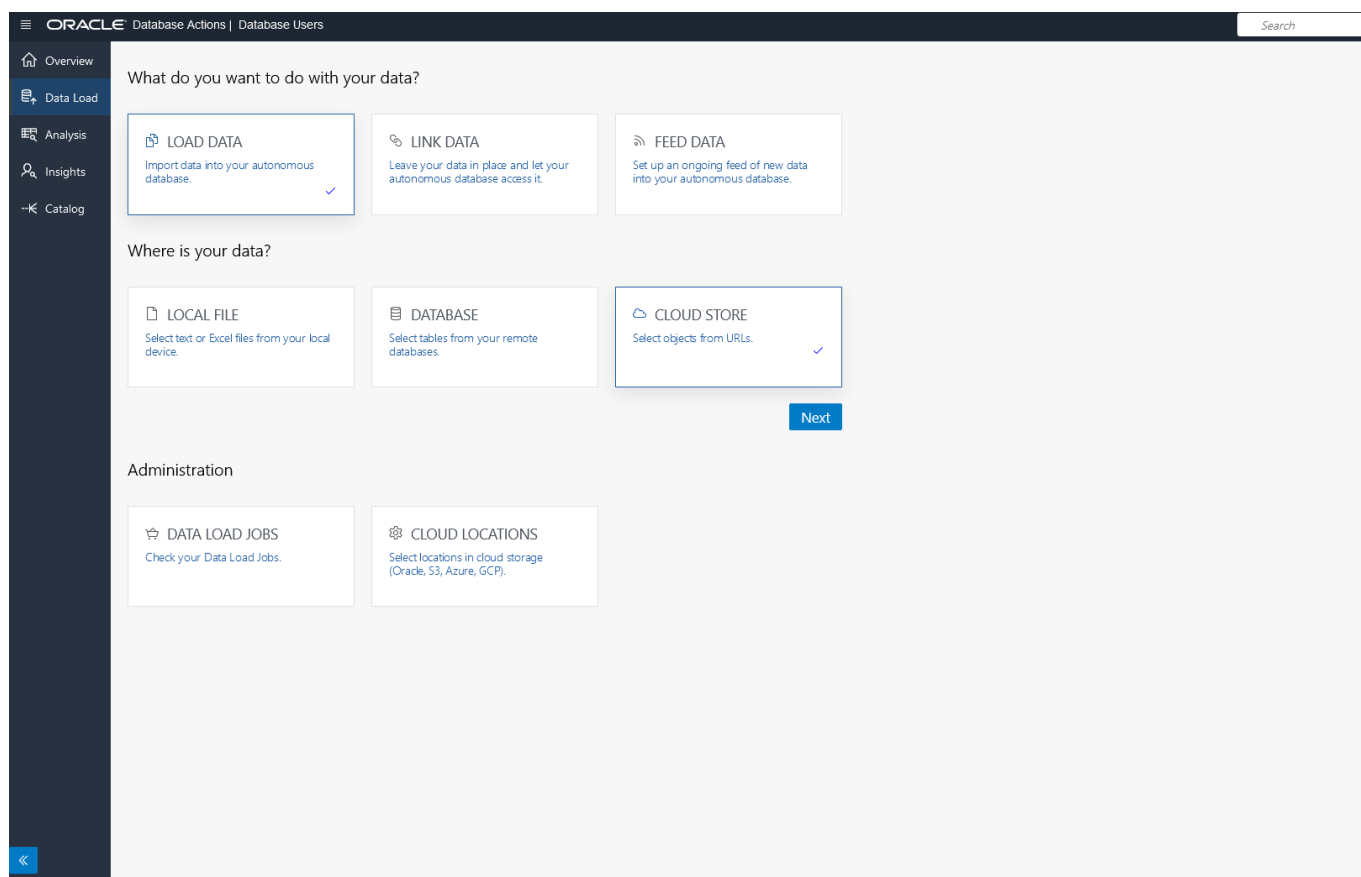
2. Cargar un set de datos

Existen varias formas de hacer este paso, y el camino seleccionado apunta a que un usuario gestor de la base de datos (ADMIN) en su labor de data managment cargue los datos y otorgue un acceso al científico de datos (F1).

Entonces, aun con el usuario ADMIN abrimos el menu de la hamburguesa que está arriba a la izquierda y de la solapa **Data Studio** seleccionamos **Data Load**.



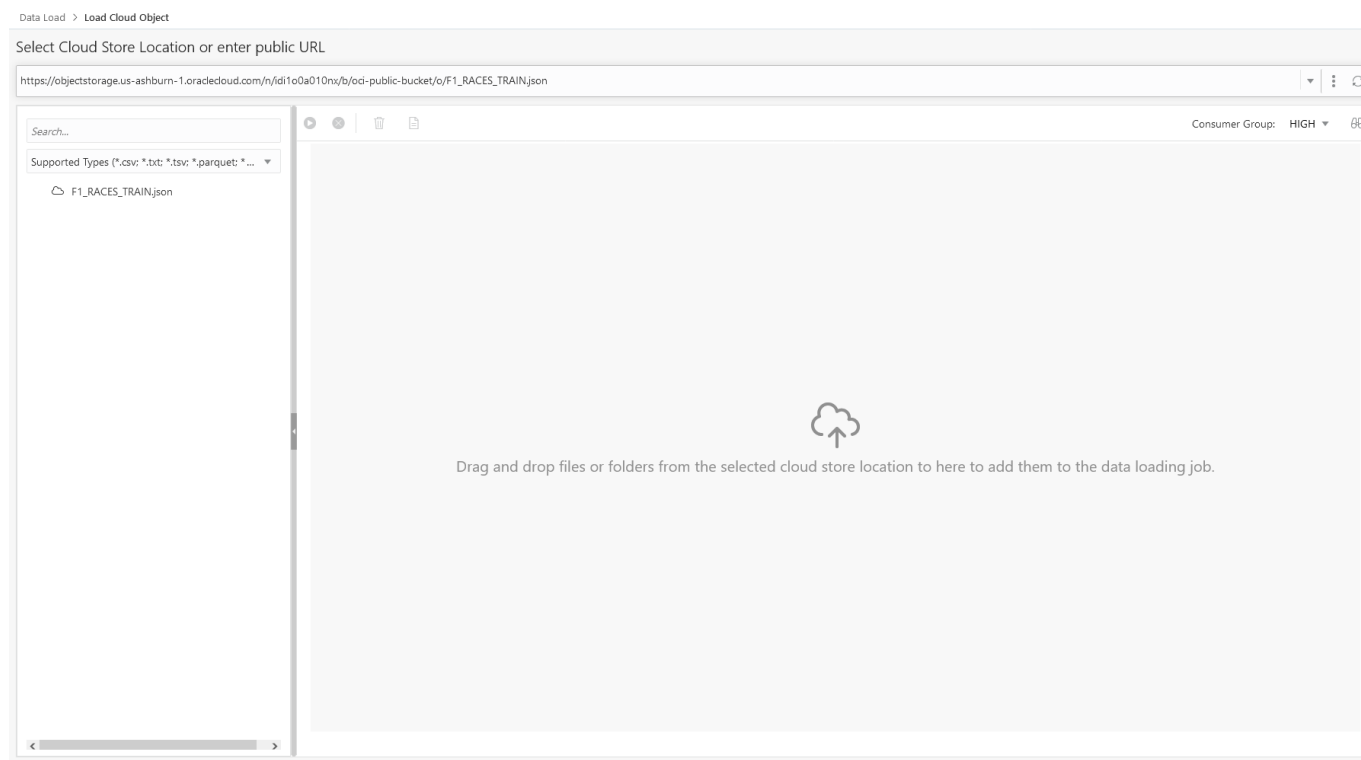
Y luego seleccionamos las opciones **LOAD DATA**, **CLOUD STORE** y damos click en **Next**.



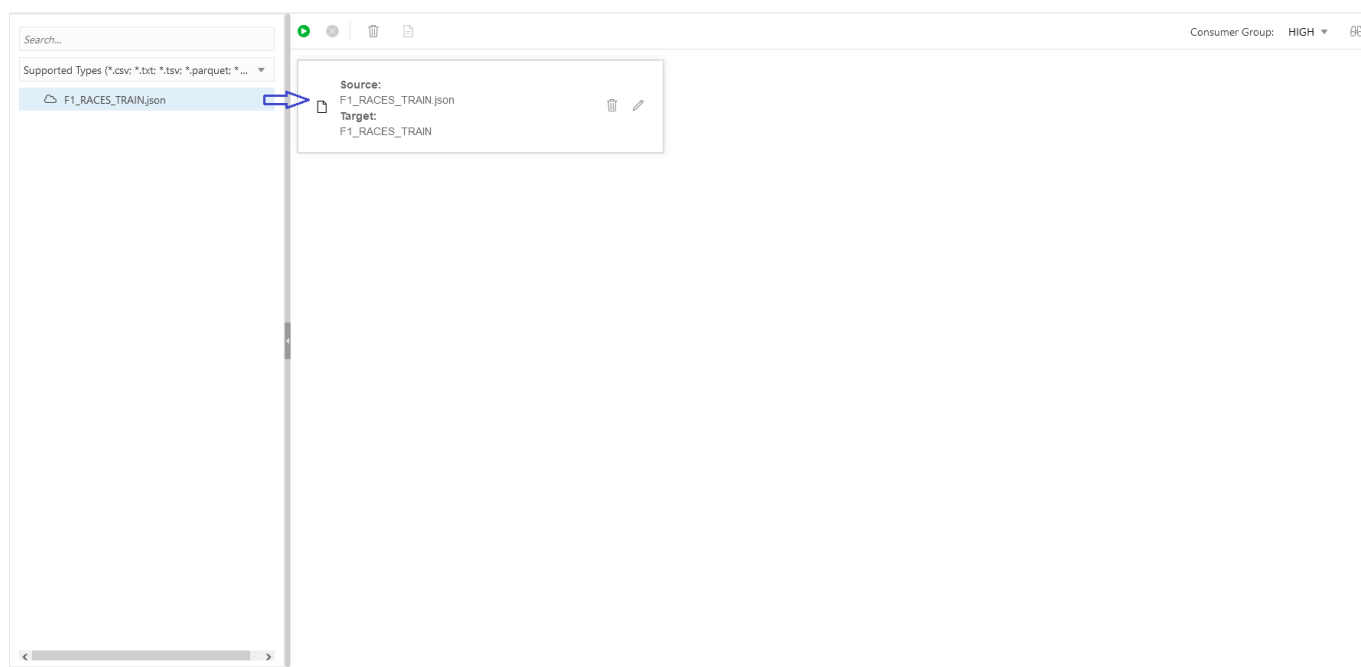
Aquí usaremos el siguiente:

- **ENLACE_PUBLICICO:** https://objectstorage.us-ashburn-1.oraclecloud.com/n/idi1o0a010nx/b/oci-public-bucket/o/F1_RACES_TRAIN.json

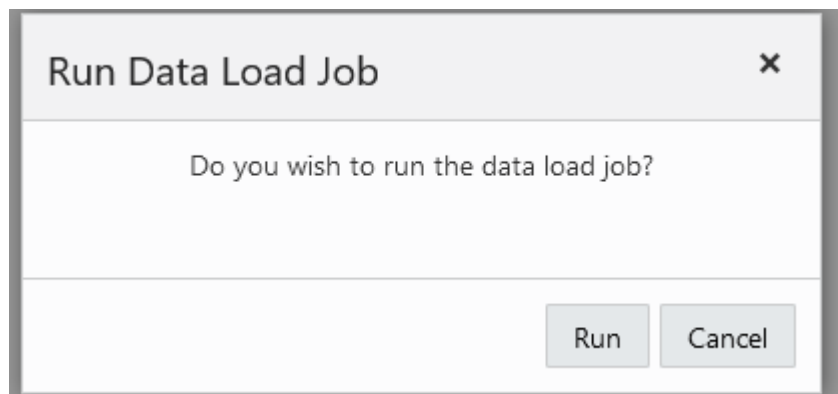
Y lo copiaremos en lo copiaremos en la barra:



Y con el click izquierdo del mouse, arrastraremos el set de datos **F1_RACES_TRAIN** del panel izquierdo al panel derecho.



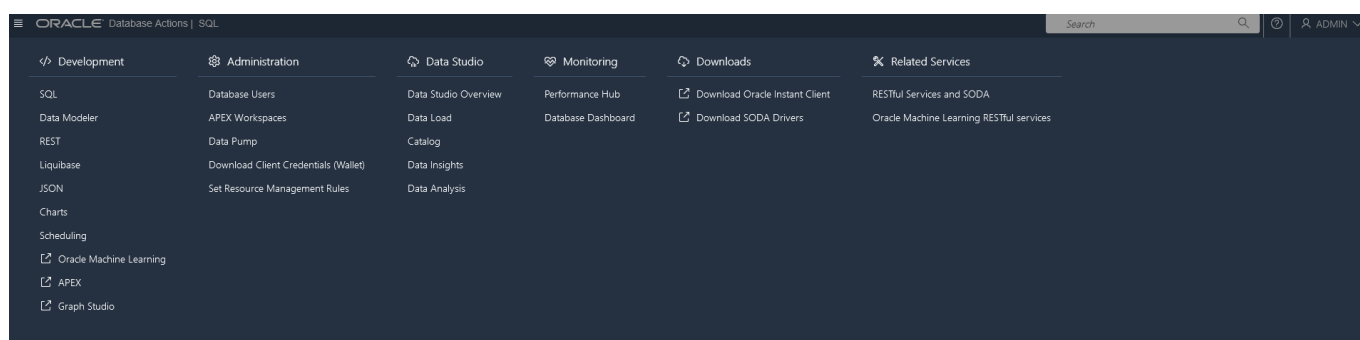
Luego daremos play al **botón verde** para ejecutar nuestro job, que consiste en, tomar los datos **F1_RACES_TRAIN** en formato **json** alojados en un **Object Storage Público** e insertarlos en una nueva Tabla **F1_RACES_TRAIN** en el esquema ADMIN.



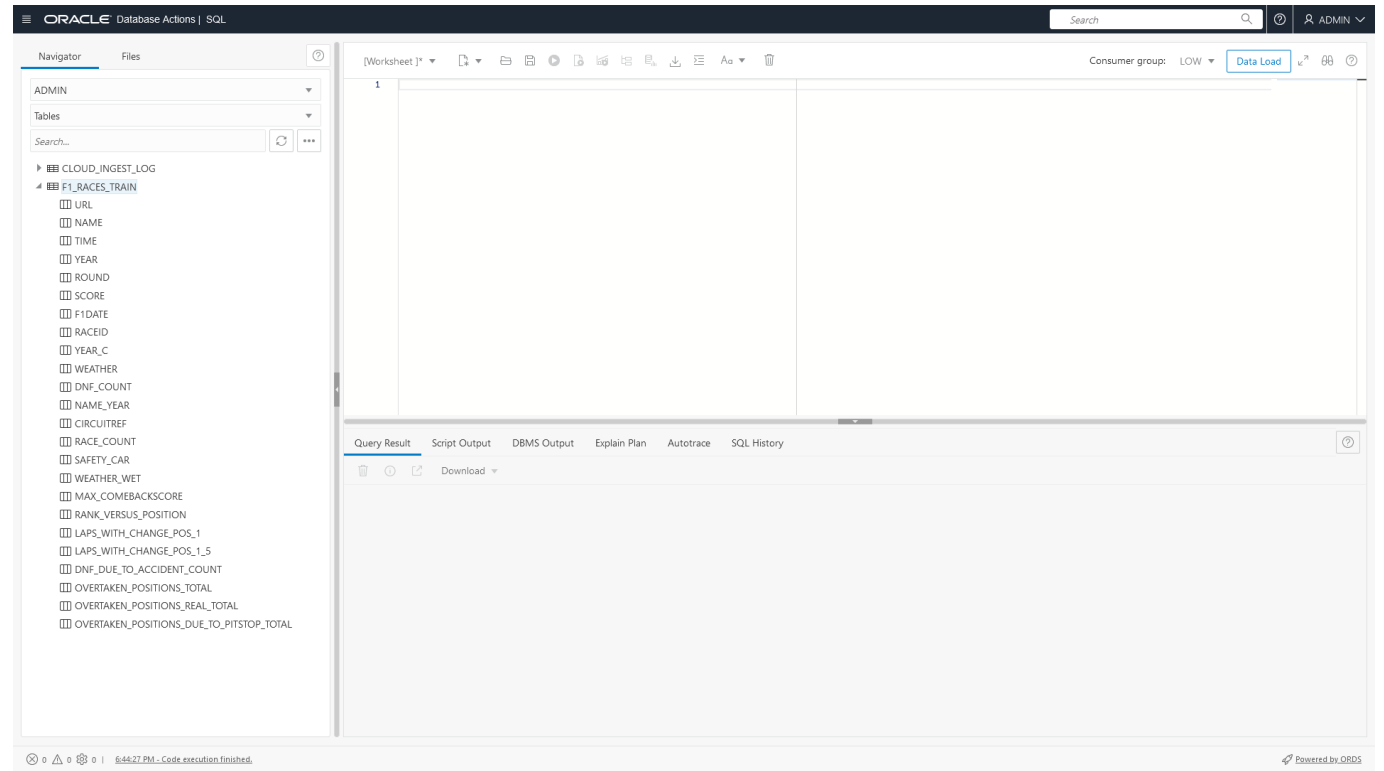
Y una vez que termine el job, veremos un cuadro de confirmación:



Finalmente con el mismo usuario ADMIN, vamos a verificar la tabla que se creó en el ADW. Para ello desde el **<Development>** seleccionamos **SQL**.

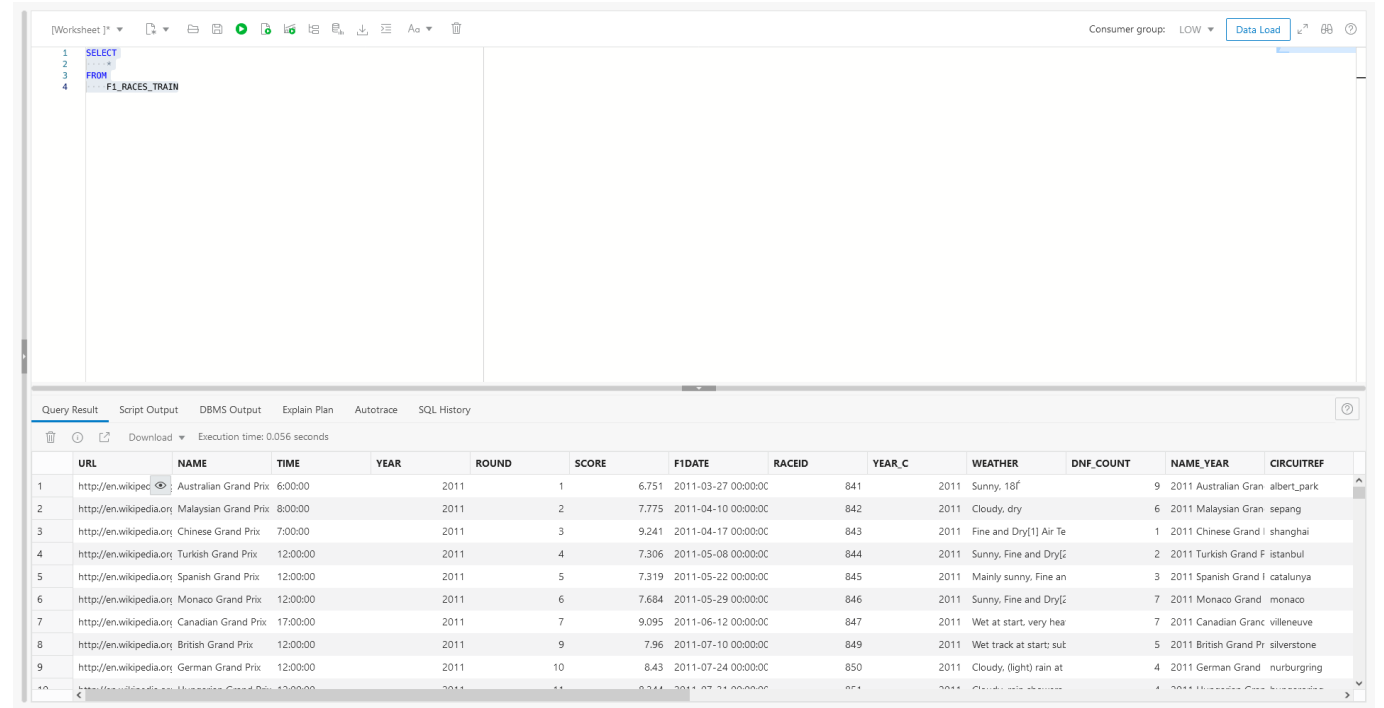


Y allí vemos a la izquierda la tabla **F1_RACES_TRAIN**.



Podemos hacer un preview de la tabla con:

```
SELECT
*
FROM
F1_RACES_TRAIN
```



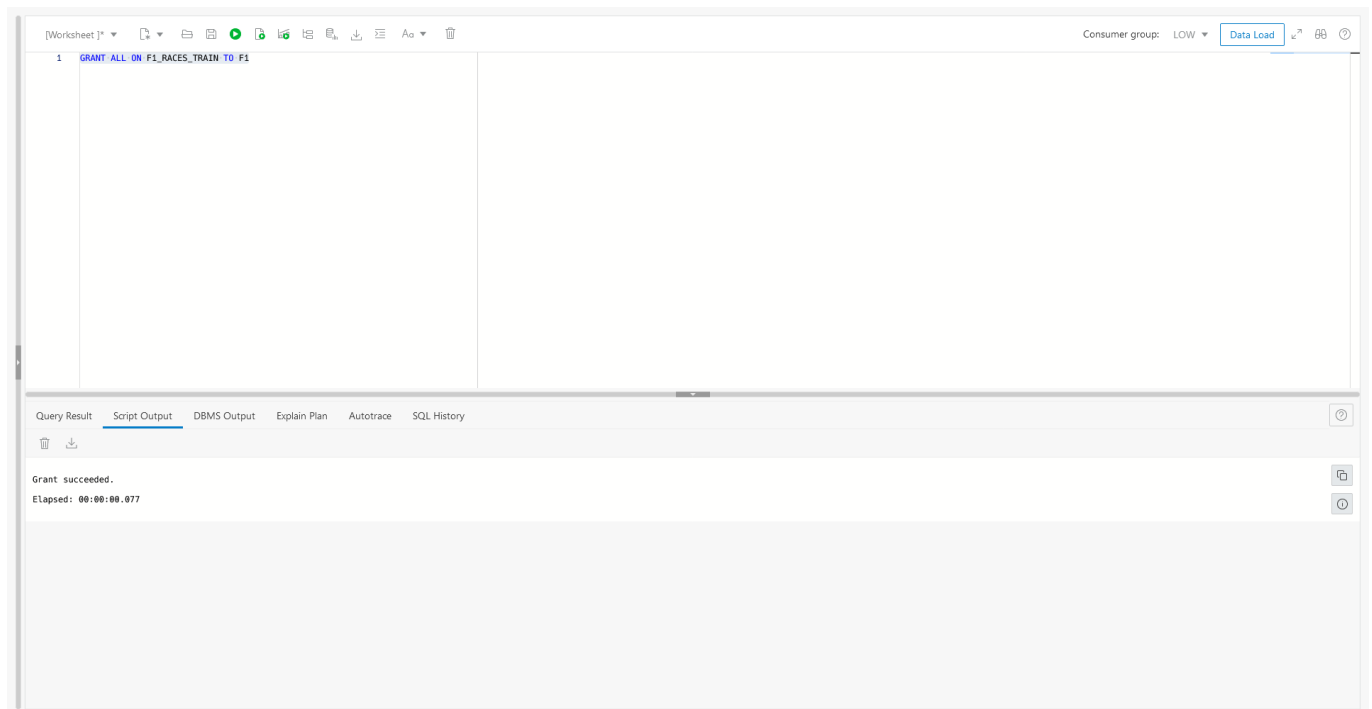
Esta tabla **F1_RACES_TRAIN** fué el producto de funcionar otros sets de datos de Fórmula 1 y hacer Ingeniería de Atributos, y ese contexto es suficiente para nuestro objetivo, en este sentido si quieres todos los detalles como un opcional después de terminar la carrera, el link se encuentra al final de este documento.

Debes saber, que:

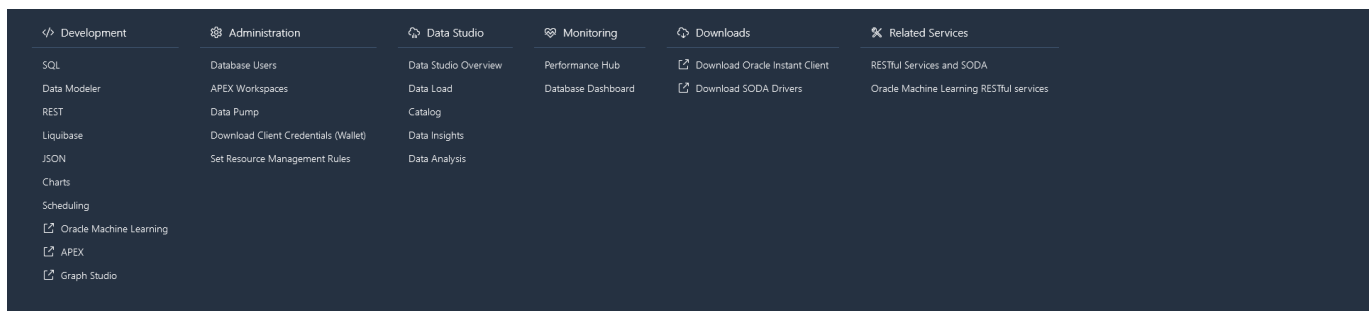
- Cada **fila** representa una carrera.
- Nuestro **TARGET** será el **SCORE**, que representa que tan "buena" fué la experiencia según una encuesta a los aficionados y será lo que queremos predecir.

En la siguiente vuelta necesitaremos usar al usuario **F1**, como la tabla **F1_RACES_TRAIN** la creamos desde el usuario **ADMIN**, debemos otorgar permisos de lectura al usuario **F1**, por lo tanto, ubicados en la misma pantalla borramos el Query anterior y ejecutamos el siguiente comando:

```
GRANT ALL ON F1_RACES_TRAIN TO F1
```



Finalmente, buscaremos el link que le compartiremos al usuario **F1** seleccionado **Oracle Machine Learning**.



Y en la siguiente pantalla debemos logear como usuario **F1**.

Database name:
ADWF1

Sign in with your Oracle Machine Learning Database User credentials

USERNAME

F1

PASSWORD

••••••••••••••••

Sign In

3. Crear un Experimento de Auto Machine Learning (Auto-ML)

Ahora, llegó el momento de hacer Auto ML!

Recordando que ahora estamos con el usuario **F1** Iniciaremos creando un **AutoML Experiments**:

The screenshot shows the Oracle Machine Learning web interface. On the left is a dark sidebar with navigation links: Home, Project, Notebooks, AutoML Experiments, Models, Templates, and Jobs. The main content area has a header 'ORACLE Machine Learning'. Below it are two sections: 'How Do I?' and 'Quick Actions'. 'How Do I?' contains links for Get Started, Use AutoML, Deploy Models, Create Notebooks, Create Jobs, Manage Permissions, and Try It. 'Quick Actions' contains links for AutoML, Models, Scratchpad, Notebooks, Jobs, and Examples. At the bottom is a 'Recent Activities' section showing 'Nothing to Display'.

Damos click en el botón **+ Create**:

AutoML Experiments

+ Create Edit Delete Duplicate Start Stop Filter

Name	Comment	Created On	Created By	Status
No data to display.				

Seleccionamos:

- Name: Argentina Grand Prix
- Data Source: ADMIN.F1_RACES_TRAIN
- Predict: SCORE
- Prediction Type: Regression
- Case ID: RACEID

Create Experiment

Start

Cancel

Save

Name

Carrera Argentina

Comments

Data Source

ADMIN.F1_RACES_TRAIN

Predict

SCORE

Prediction Type


Regression

Case ID

RACEID

En Additional Settings seleccionaremos las siguientes opciones:

▼ Additional Settings

 Reset

Maximum Top Models

5



Maximum Run Duration (Hours)

8



Database Service Level

Low



Model Metric

Median Absolute Error



Algorithms



Name ↕



Generalized Linear Model



Generalized Linear Model (Ridge Regression)



Neural Network



Support Vector Machine (Gaussian)



Support Vector Machine (Linear)

El más importante es la Métrica debemos seleccionar **Median Absolute Error (MAE)**, el cual es un indicador de la imprecisión que puede llegar a tener nuestros modelos.

"Por ejemplo, si el modelo predice 100 y tenemos un MAE de 2, significa que el modelo el valor "real" podría ser 102 o 98."

En cuanto a las Features, debes selccionar solo las que están marcadas:

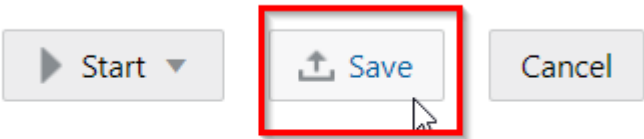
Features

Refresh

Search...

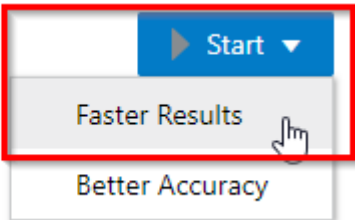
<input type="checkbox"/> Name	Type	Percent NULLs	Distinct Values	Min	Max	Mean	Std Dev
<input checked="" type="checkbox"/> CIRCUITREF	VARCHAR2	0	27				
<input checked="" type="checkbox"/> DNF_COUNT	NUMBER	0	12	0	11	4.07	2.21
<input checked="" type="checkbox"/> DNF_DUE_TO_ACCIDENT_COUNT	NUMBER	0	7	0	6	1.35	1.63
<input checked="" type="checkbox"/> LAPS_WITH_CHANGE_POS_1	NUMBER	0	11	0	10	2.91	2.56
<input checked="" type="checkbox"/> LAPS_WITH_CHANGE_POS_1_5	NUMBER	0	27	4	34	14.3	6.08
<input checked="" type="checkbox"/> MAX_COMEBACKSCORE	NUMBER	0	17	5	22	12.19	3.27
<input type="checkbox"/> NAME	VARCHAR2	0	26				
<input type="checkbox"/> NAME_YEAR	VARCHAR2	0	177				
<input checked="" type="checkbox"/> OVERTAKEN_POSITIONS_DUE_TO_PITSTOP_TOTAL	NUMBER	0	81	0	295	42.23	39.91
<input checked="" type="checkbox"/> OVERTAKEN_POSITIONS_REAL_TOTAL	NUMBER	0	139	5	341	127.11	72.3
<input type="checkbox"/> OVERTAKEN_POSITIONS_TOTAL	NUMBER	0	134	35	422	169.34	80.52
RACEID	NUMBER	0	174	841	1030	936.29	56.61
<input type="checkbox"/> RACE_COUNT	NUMBER	0	1	1	1	1	0
<input checked="" type="checkbox"/> RANK_VERSUS_POSITION	NUMBER	0	32	0	36	11.07	7.51
<input type="checkbox"/> ROUND	NUMBER	0	21	1	21	10.44	5.92
<input checked="" type="checkbox"/> SAFETY_CAR	NUMBER	0	2	0	1	0.58	0.7
<input checked="" type="radio"/> SCORE	NUMBER	0	167	3.5	9.449	6.86	1.34
<input type="checkbox"/> TIME	VARCHAR2	0	24				
<input type="checkbox"/> URL	VARCHAR2	0	178				
<input type="checkbox"/> WEATHER	VARCHAR2	0	116				
<input checked="" type="checkbox"/> WEATHER_WET	VARCHAR2	0	2				
<input type="checkbox"/> YEAR	NUMBER	0	9	2011	2019	2015.09	2.76
<input type="checkbox"/> YEAR_C	VARCHAR2	0	9				

Regresamos al inicio de la página y guardamos nuestro expermiento:

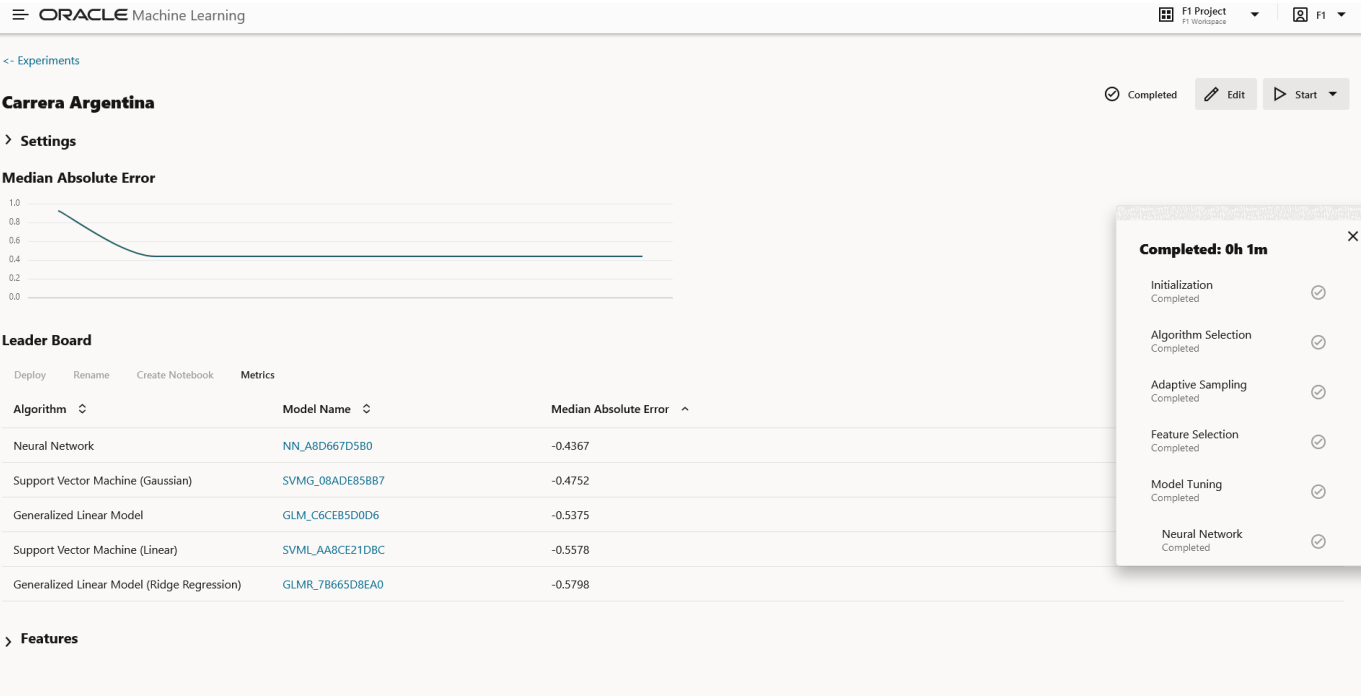


4. Iniciar Entrenamiento

El siguiente paso es iniciar el experimento con los resultados más rápidos.



Y estaré atento a que el entrenamiento termine:



5. Encontrar el top 3 de atributos más importantes [FINAL]

Desplegamos la pestaña de Features:

Features								
Refresh								
Name	Importance	Type	Distinct Values	Min	Max	Mean	Std Dev	
F1DATE		VARCHAR2	164					
CIRCUITREF		VARCHAR2	27					
DNF_DUE_TO_ACCIDENT_COUNT		NUMBER	7	0	6	1.41	1.63	
SAFETY_CAR		NUMBER	2	0	1	0.6	0.69	
LAPS_WITH_CHANGE_POS_1_5		NUMBER	26	4	34	14.43	6.12	
DNF_COUNT		NUMBER	12	0	11	4.19	2.14	
LAPS_WITH_CHANGE_POS_1		NUMBER	11	0	10	2.93	2.58	
MAX_COMEBACKSCORE		NUMBER	17	5	22	12.2	3.3	
OVERTAKEN_POSITIONS_DUE_TO_PITSTOP_TOTAL		NUMBER	78	0	295	42.02	40.09	

Vemos que le más importante fue:

1. F1DATE
2. CIRCUITREF
3. DNF_DUE_TO_ACCIDENT_COUNT



Nota Final

Este laboratorio fue diseñado a partir los [LiveLabs](#) en [Title Learn Analytics and Machine Learning with Red Bull Racing](#)