



# **Universidad De Sonora**

**Maestría:** en Ciencia de datos

**Curso:** Introducción a la ciencia de datos

**Profesor:** Dr. Juan Pablo Soto

Proyecto Contando Historias con Datos

Alumno: Aarón Francisco Granich Armenta  
Expediente: A210214575

## Objetivo

Desarrollar una aplicación que permita analizar información contenida dentro de una base de datos en SQL.

Dentro de SQL generar una base de datos que contenga al menos una vista, un Procedimiento almacenado y una Función.

Una vez generado, vamos a utilizar Jupyter LAB para conectarnos a la base de datos por medio de la librería pymysql.

## Iniciamos con el Proyecto:

Primero vamos a iniciar explicando algunos conceptos y resolviendo unas preguntas

### ¿Qué es una base de datos?

- Una base de datos es una herramienta para recopilar y organizar información

### ¿Qué es SQL?

- SQL es un acrónimo en inglés para Structured Query Language. Un Lenguaje de Consulta Estructurado. Un tipo de lenguaje de programación que te permite manipular y descargar datos de una base de datos

### ¿Dentro de SQL que es un Esquema, una Vista, un procedimiento almacenado y una función?

- **Esquema** = El esquema es una entidad en el ámbito de la base de datos
- **Vista / View** = Las vistas en MySQL (VIEWS) son tablas virtuales. Solo muestran los datos que están almacenados en otras tablas (que sí son reales). Siendo así, crear vistas en MySQL significa mostrar información de una fuente de origen sin necesidad de mostrar ni exponer a la fuente en sí.

- **Procedimiento de almacenamiento** = Los procedimientos almacenados MySQL, también conocidos como Stored Procedure, se presentan como conjuntos de instrucciones escritas en el lenguaje SQL. Su objetivo es realizar una tarea determinada, desde operaciones sencillas hasta tareas muy complejas
- **Funcion** = Una función en MySQL es una rutina creada para tomar unos parámetros, procesarlos y retornar en un salida. Deben retornar en un valor con algún tipo de dato definido.

Ahora que ya tenemos claros los conceptos, iniciaremos explicando los pasos a seguir para generar una vista, un procedimiento almacenado y una funcion dentro de MySQL.

### **Descripción de nuestra base de datos:**

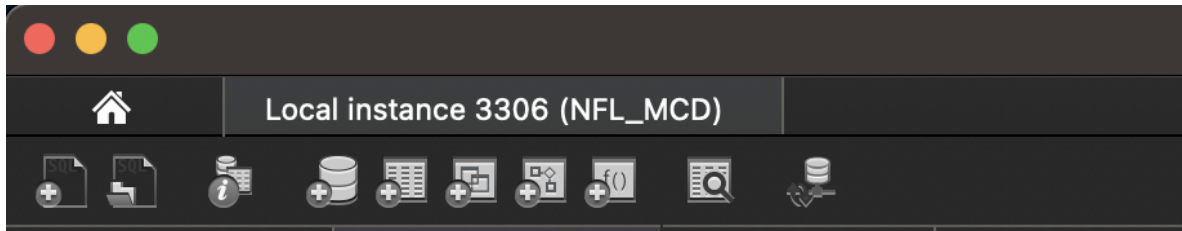
Para este proyecto utilizaremos una base de datos que puedes encontrar en la siguiente liga:

(<https://www.kaggle.com/datasets/timschutzang/dataset1>)

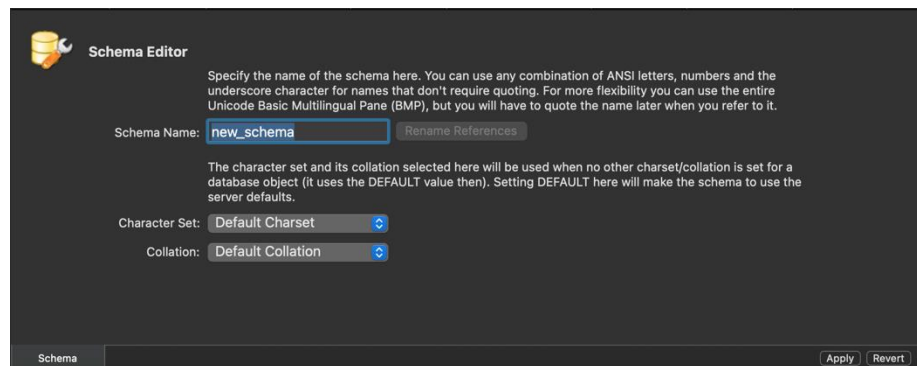
Esta base de datos nos da los resultados y estadísticas de cada equipo desde 1876 al 2016

## Paso 1. Generamos un nuevo Esquema (Schema)

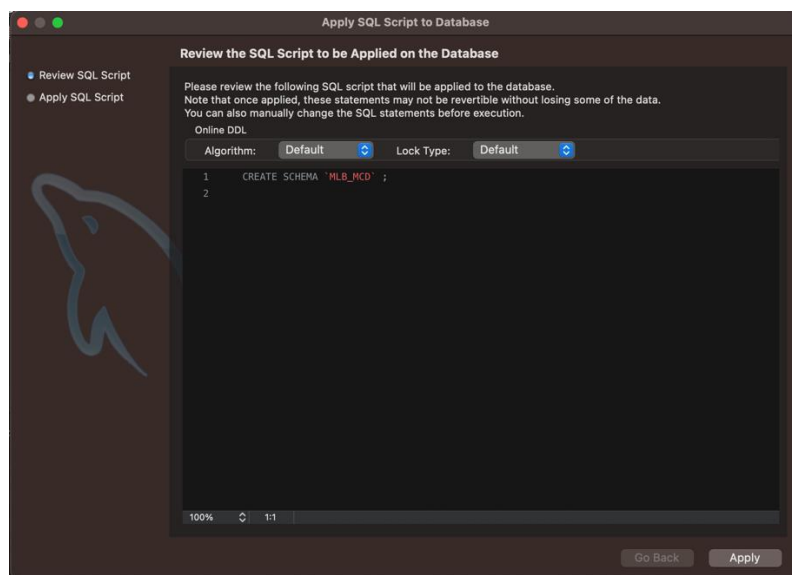
Para Generarlo tenemos que dar click en el icono



Te Va aparecer esta ventana, tienes que colocarle el nombre y después dar click en Apply.

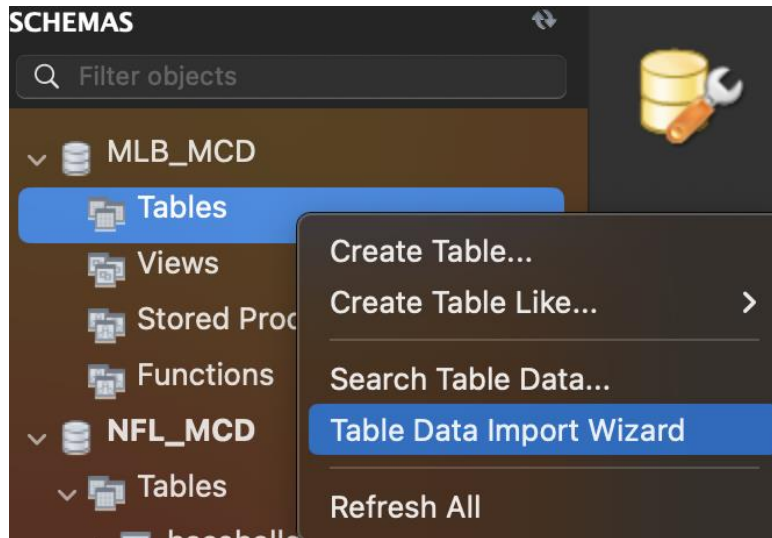


Nuevamente das click en Apply y tu nuevo esquema se va generar

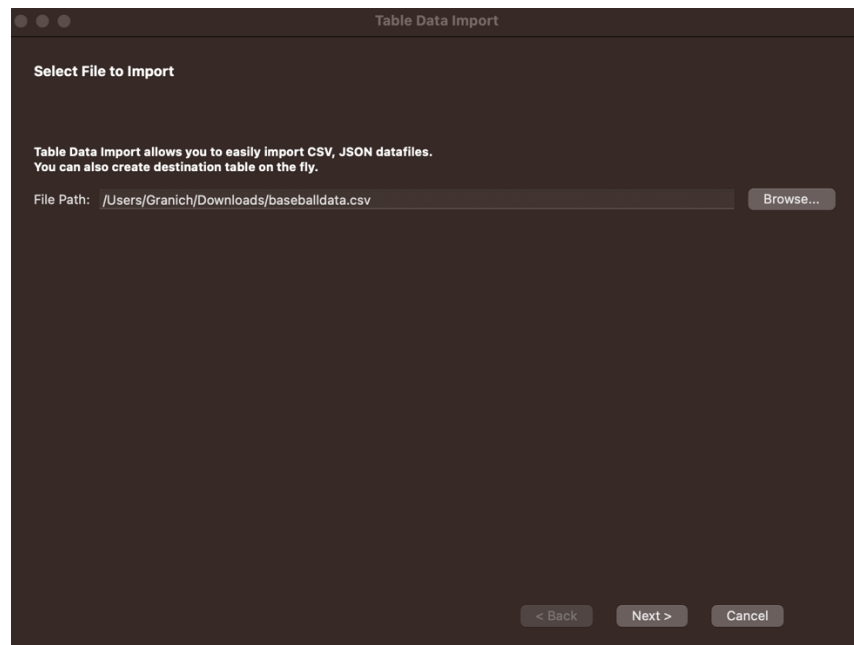


## Paso 2. Importamos una tabla dentro de nuestra nueva esquema

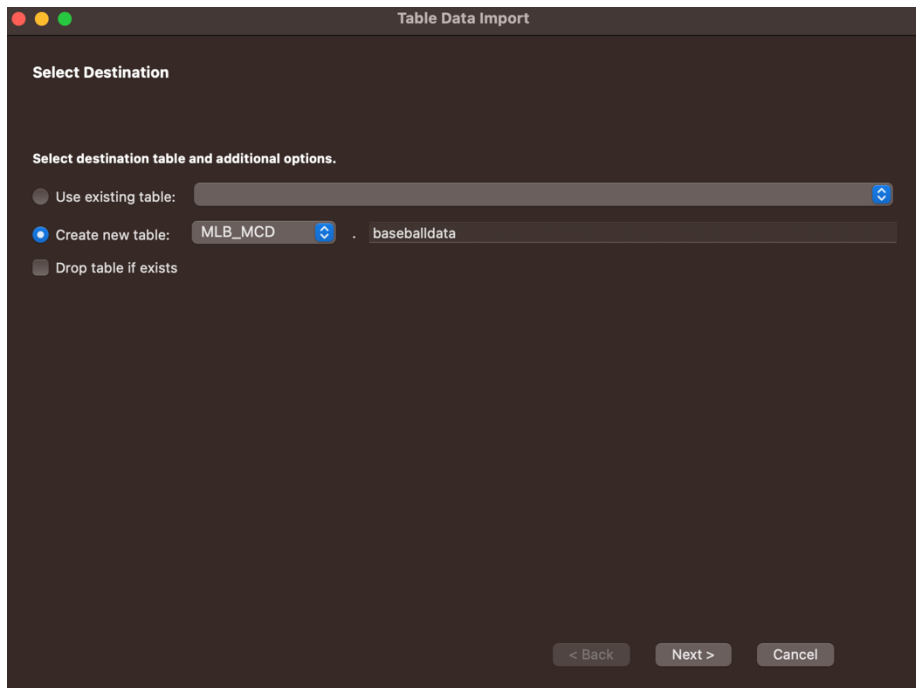
Vamos a irnos a Tables -> Click Derecho -> Table Data Import Wizard



Seleccionamos la ubicación del archivo CSV, Browse->Next>



Damos next



**Table Data Import**

**Select Destination**

Select destination table and additional options.

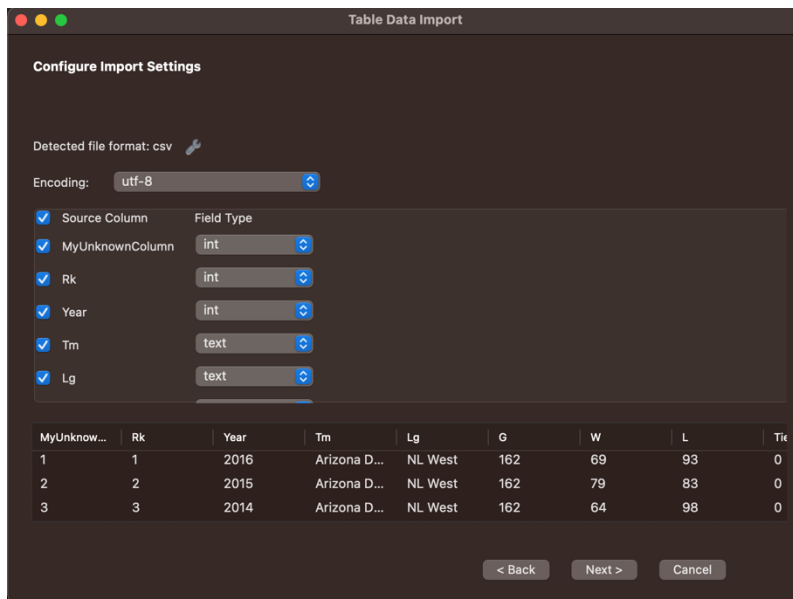
☐ Use existing table:

☒ Create new table:  .

☐ Drop table if exists

< Back 
 > Next 
 > Cancel

Revisamos que el tipo de dato este correcto y damos next



**Table Data Import**

**Configure Import Settings**

Detected file format: csv

Encoding: utf-8

Source Column	Field Type
<input checked="" type="checkbox"/> MyUnknownColumn	int
<input checked="" type="checkbox"/> Rk	int
<input checked="" type="checkbox"/> Year	int
<input checked="" type="checkbox"/> Tm	text
<input checked="" type="checkbox"/> Lg	text

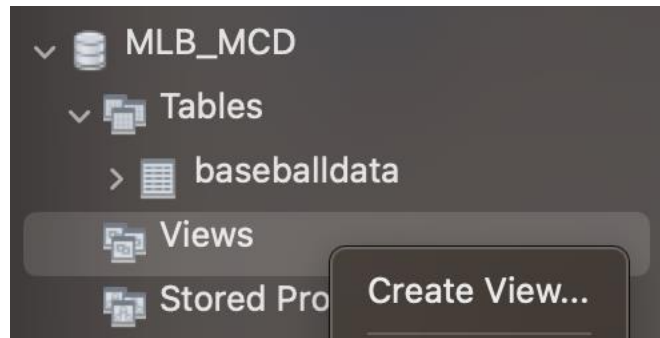
MyUnknown...	Rk	Year	Tm	Lg	G	W	L	Tie
1	1	2016	Arizona D...	NL West	162	69	93	0
2	2	2015	Arizona D...	NL West	162	79	83	0
3	3	2014	Arizona D...	NL West	162	64	98	0

< Back 
 > Next 
 > Cancel

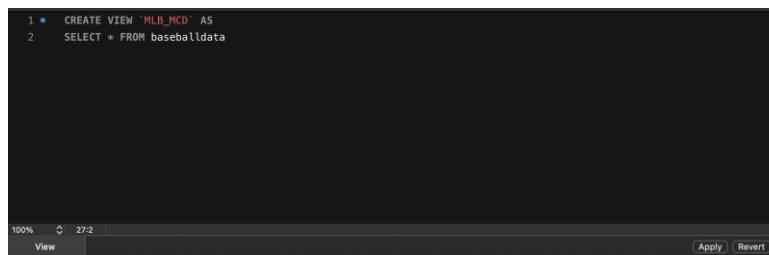
Después esperamos que los datos se importen y terminamos ya tenemos nuestro nuevo esquema.

### Paso 3. Creamos una nueva vista

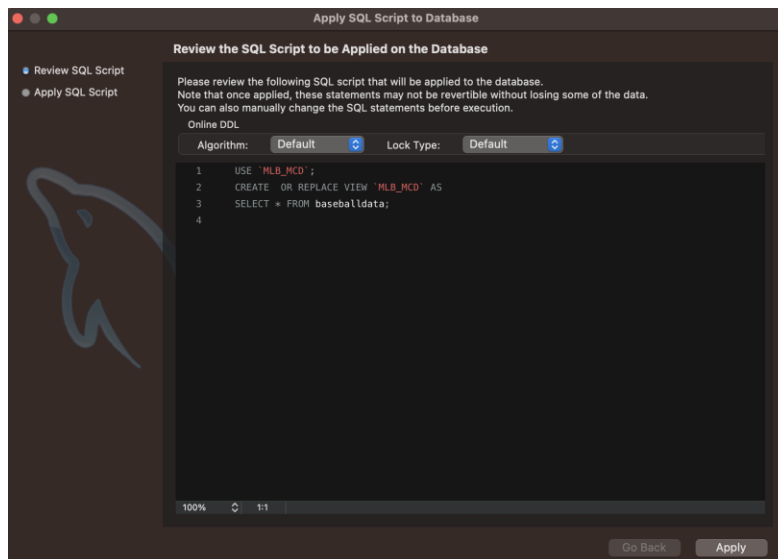
Damos Click derecho en Views -> Create View



Seleccionamos de donde vamos hacer una vista en este caso de nuestra tabla baseballdata



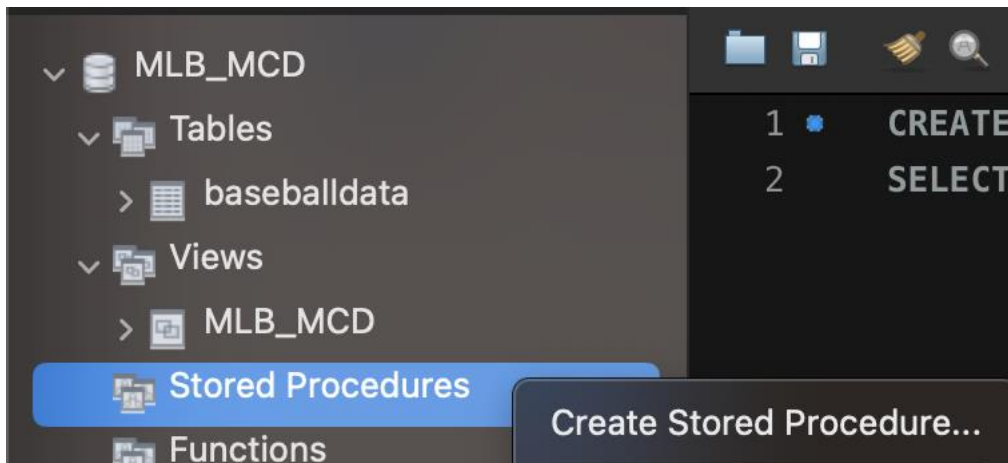
Damos Apply



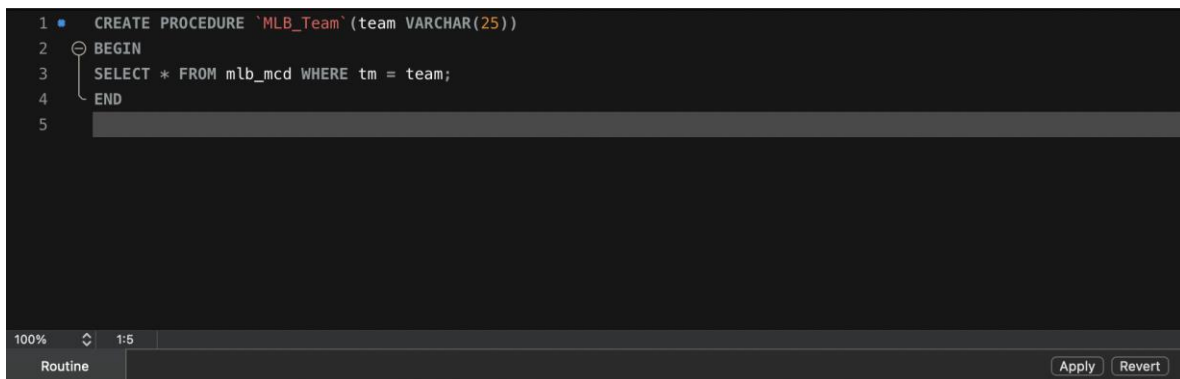
Y listo tenemos nuestra nueva Vista

#### Paso 4. Creamos un nuevo Procedimiento de almacenamiento

Mismo Procedimiento damos click derecho en stored procedures -> Create Stored Procedure

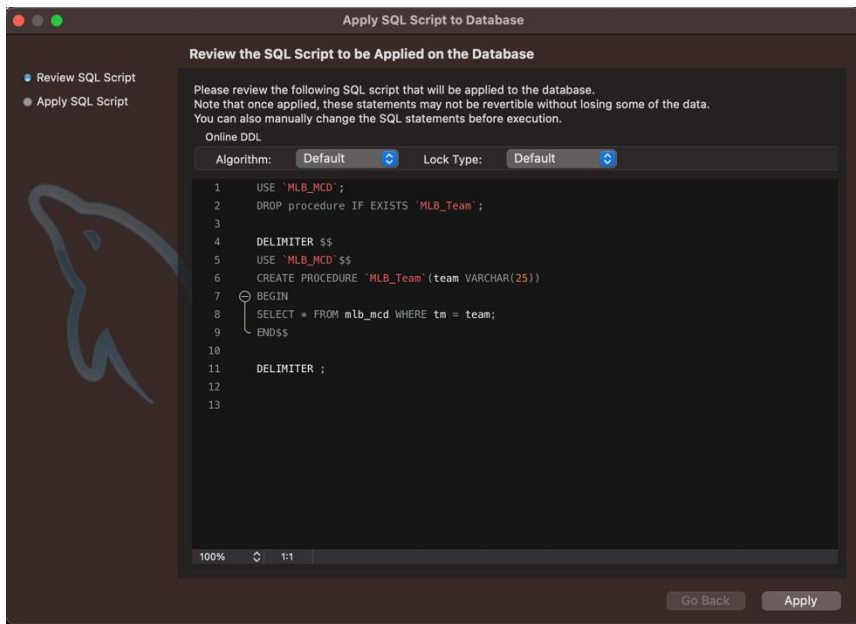


En este caso lo que queremos hacer es que al momento de darle un nombre del equipo nos arroje la informacion de ese equipo

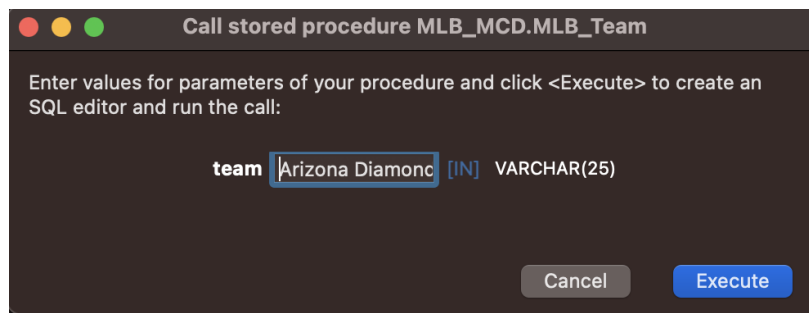


Damos Apply





Damos apply y corremos nuestro procedimiento, colocamos el nombre del equipo y ejecutamos



```
1 call MLB_MCD.MLB_Team('Arizona Diamondbacks');
2
```

100% 1:1

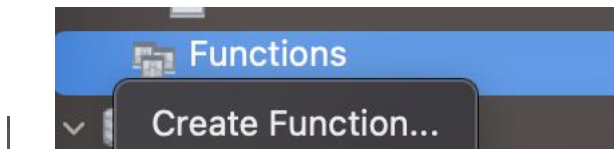
Result Grid Filter Rows: Search Export:

	MyUnknownColu...	Rk	Year	Tm	Lg	G	W	L	Ties	W.L.	pythW.L.	Finish	GB	Playoffs	R	RA	Attendanc
3		3	2014	Arizona Diamondbacks	NL West	162	64	98	0	0.395	0.415	5th of 5	30			615 742	2,073,730
4		4	2013	Arizona Diamondbacks	NL West	162	81	81	0	0.5	0.493	2nd of 5	11			685 695	2,134,895
5		5	2012	Arizona Diamondbacks	NL West	162	81	81	0	0.5	0.53	3rd of 5	13			734 688	2,177,617
7		7	2010	Arizona Diamondbacks	NL West	162	65	97	0	0.401	0.428	5th of 5	27			713 836	2,056,697
8		8	2009	Arizona Diamondbacks	NL West	162	70	92	0	0.432	0.462	5th of 5	25			720 782	2,128,765
9		9	2008	Arizona Diamondbacks	NL West	162	82	80	0	0.506	0.509	2nd of 5	2			720 706	2,509,924
11		11	2006	Arizona Diamondbacks	NL West	162	76	86	0	0.469	0.491	4th of 5	12			773 788	2,091,685
12		12	2005	Arizona Diamondbacks	NL West	162	77	85	0	0.475	0.406	2nd of 5	5			696 856	2,059,424
13		13	2004	Arizona Diamondbacks	NL West	162	51	111	0	0.315	0.333	5th of 5	42			615 899	2,519,560

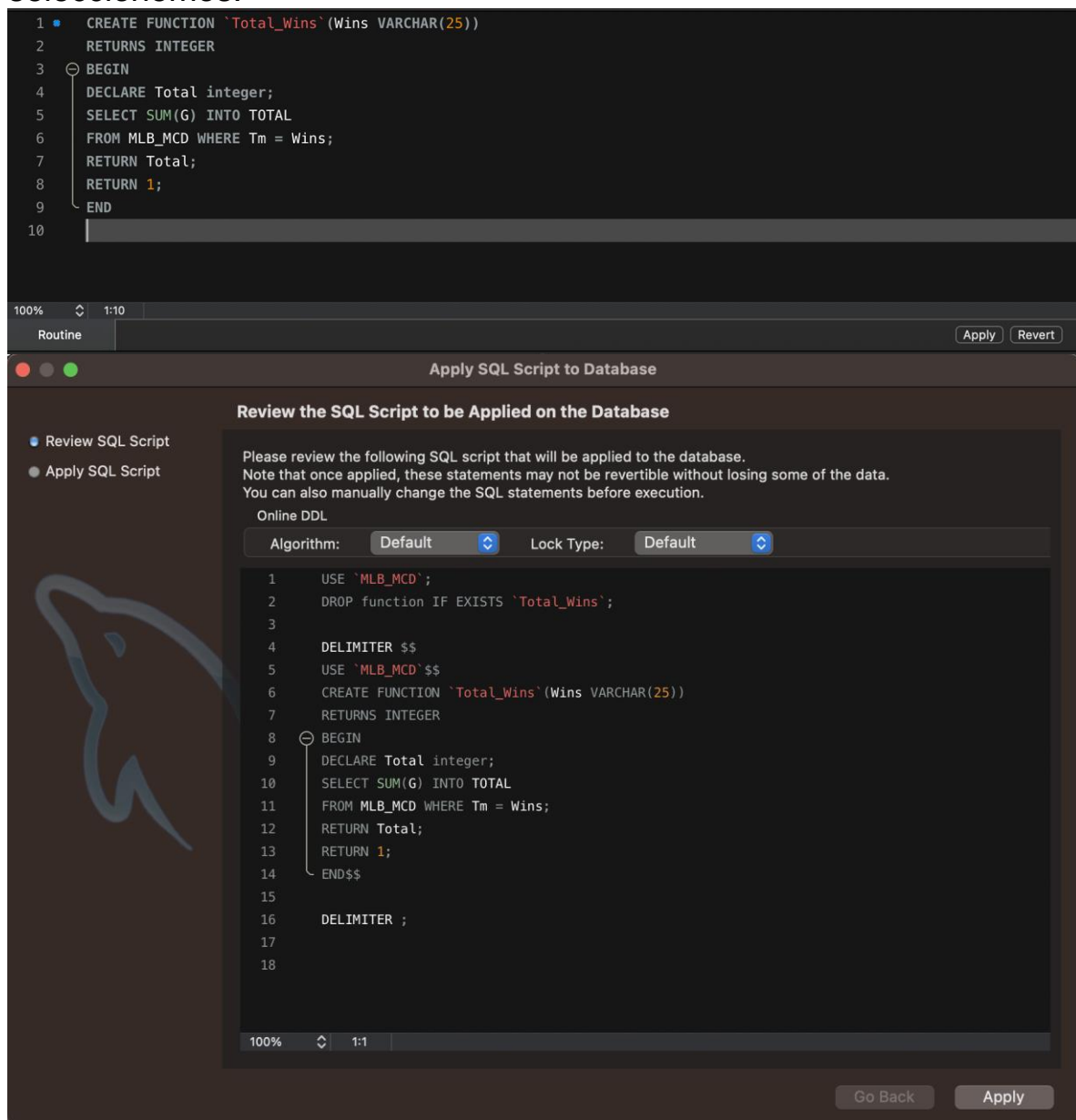
Result Grid Form Editor

## Paso 5. Creamos una nueva Función

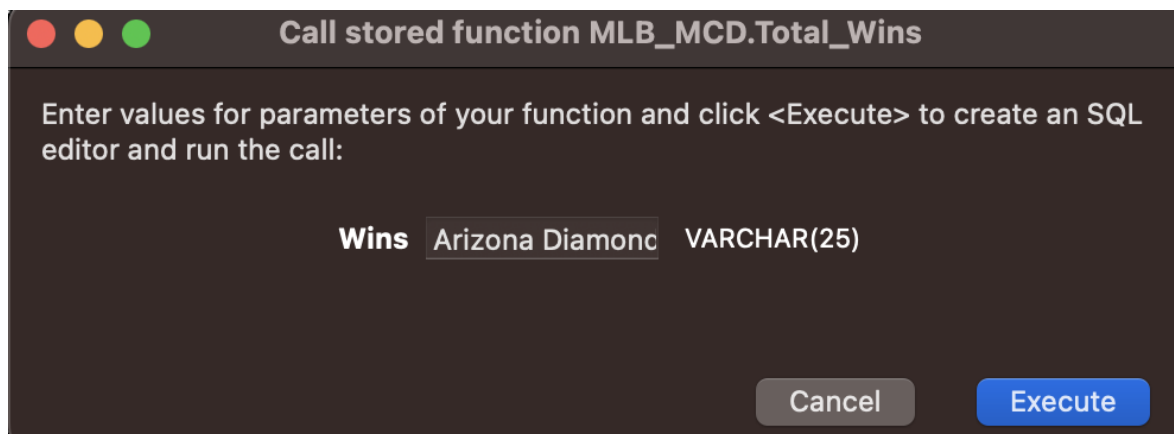
Para crear una función hacemos lo mismo, damos click derecho en Functions -> Create Fuction



Nuestra función sumara las cantidades de victoria del equipo que seleccionemos.

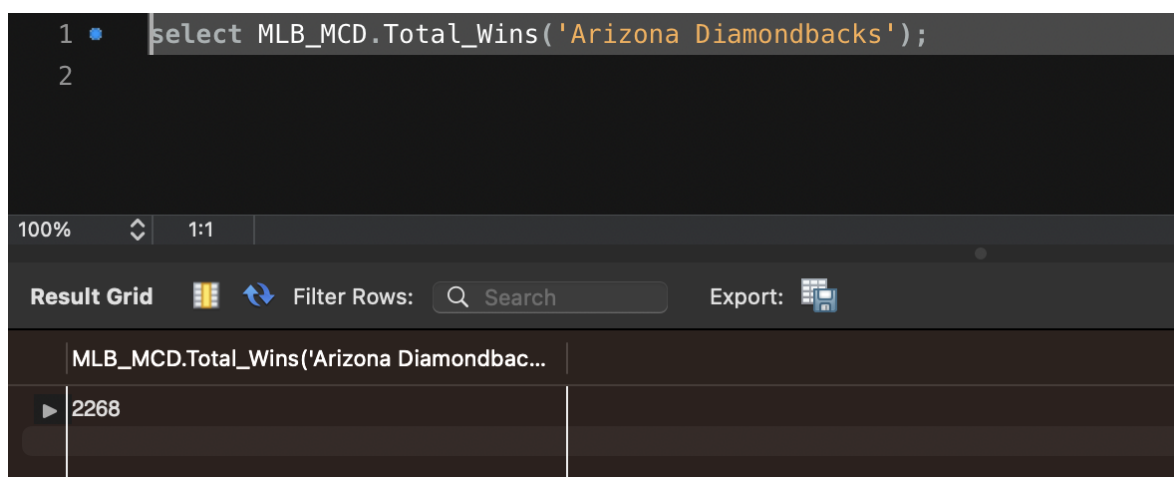


Le decimos que equipo queremos saber la cantidad de victorias y ejecutamos



A dialog box titled "Call stored function MLB\_MCD.Total\_Wins". It contains a text area with the instruction: "Enter values for parameters of your function and click <Execute> to create an SQL editor and run the call:". Below this, there is a label "Wins" followed by a text input field containing "Arizona Diamond" and a data type "VARCHAR(25)". At the bottom right, there are two buttons: "Cancel" and "Execute".

Total de Victorias del equipo desde 1876 hasta 2016



The image shows a SQL editor window with the following SQL query:

```
1 select MLB_MCD.Total_Wins('Arizona Diamondbacks');
2
```

Below the editor is a "Result Grid" showing the output of the query. The grid has a single column with the header "MLB\_MCD.Total\_Wins('Arizona Diamondbac...". The first row contains the value "2268".

MLB_MCD.Total_Wins('Arizona Diamondbac...
2268

# Paso 6. Vamos a utilizar la librería pymysql de phyton para conectarnos a SQL y realizar algunas consultas y DF con Phyton.

Iniciamos Importando las Librerias necesarias

```
In [33]: import pymysql
import matplotlib.pyplot as plt
import pandas as pd
```

Generamos la coneccion con la libreria de pymysql

```
In [ ]: connection = pymysql.connect(host='localhost',
                                     user='root',
                                     password='YourPassword',
                                     db='NFL_MCD')
```

Vaoms a revisar el DF por medio de un SQL Query

```
In [35]: dfr=pd.read_sql_query("SELECT * FROM NFL_MCD.baseballdata", connection, index_col='MyUnknownColumn')
dfr.head(10)
```

Out[35]:

	Rk	Year	Tm	Lg	G	W	L	Ties	W.L.	pythW.L.	...	R	RA	Attendance	BatAge	PAge	X.Bat	X.P	Top.Player	Managers	cu	
MyUnknownColumn																						
	1	1	2016	Arizona Diamondbacks	NL West	162	69	93	0	0.426	0.424	...	752	890	2,036,216	26.7	26.4	50	29	J.Segura (5.7)	C.Hale (69-93)	Arizona Diamondbacks
	2	2	2015	Arizona Diamondbacks	NL West	162	79	83	0	0.488	0.504	...	720	713	2,080,145	26.6	27.1	50	27	P.Goldschmidt (8.8)	C.Hale (79-83)	Arizona Diamondbacks
	3	3	2014	Arizona Diamondbacks	NL West	162	64	98	0	0.395	0.415	...	615	742	2,073,730	27.6	28.0	52	25	P.Goldschmidt (4.5)	K.Gibson (63-96) and A.Trammell (1-2)	Arizona Diamondbacks
	4	4	2013	Arizona Diamondbacks	NL West	162	81	81	0	0.500	0.493	...	685	695	2,134,895	28.1	27.6	44	23	P.Goldschmidt (7.1)	K.Gibson (81-81)	Arizona Diamondbacks
	5	5	2012	Arizona Diamondbacks	NL West	162	81	81	0	0.500	0.530	...	734	688	2,177,617	28.3	27.4	48	23	A.Hill (5.0)	K.Gibson (81-81)	Arizona Diamondbacks
	7	7	2010	Arizona Diamondbacks	NL West	162	65	97	0	0.401	0.428	...	713	836	2,056,697	26.8	27.9	48	28	C.Young (5.4)	A.Hinch (31-48) and K.Gibson (34-49)	Arizona Diamondbacks

Generamos un nuevo DF

```
In [36]: Wins= dfr.groupby("Tm").sum()["W"].reset_index()
Wins
```

Out[36]:

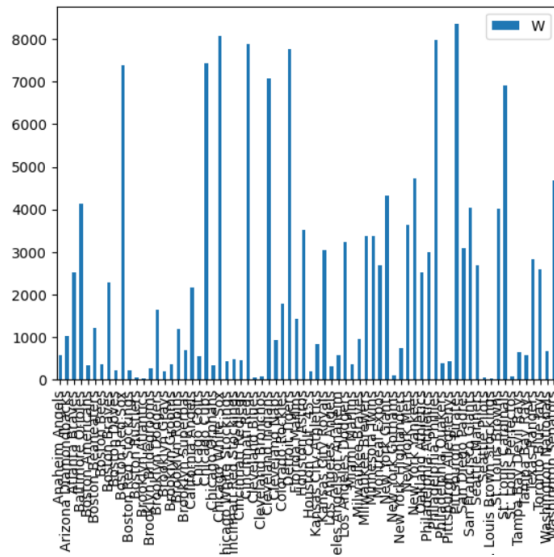
	Tm	W
0	Anaheim Angels	572
1	Arizona Diamondbacks	1029
2	Atlanta Braves	2533
3	Baltimore Orioles	4133
4	Boston Americans	342
...	...	...
67	Tampa Bay Rays	582
68	Texas Rangers	2843
69	Toronto Blue Jays	2604
70	Washington Nationals	661
71	Washington Senators	4676

72 rows × 2 columns

In [40]:

```
Wins.plot.bar(x="Tm", y="W")
```

Out[40]:



Código completo en:

[https://github.com/AARONGRANICH/MCD\\_ICDYM/blob/main/MCD\\_SQL.ipynb](https://github.com/AARONGRANICH/MCD_ICDYM/blob/main/MCD_SQL.ipynb)