

UNIVERSIDAD DE SONORA

División de ciencias exactas y naturales



Maestría en ciencia de datos

Alumna:
Ballesteros Valenzuela, Diana Laura.

Introducción a la ciencia de datos y sus metodologías:
Proyecto final

Noviembre del 2022

Índice

1. Introducción
2. Objetivos
3. Fuente de datos
4. Desarrollo
5. Referencias

1. Introducción.

Las bases de datos tienden a tener una gran relevancia tanto a nivel personal como a nivel empresarial, y se consideran una de las mayores aportaciones que ha dado la informática.

Hoy en día, cualquier organización por pequeña que sea, debe contar con una base de datos en la que pueda organizar sus datos, pero no basta con solo tenerla, hay que saber gestionarla.

SQL Server Management Studio (SSMS para abreviar) es un entorno de desarrollo integrado para administrar cualquier infraestructura SQL. Se utiliza para acceder, administrar, configurar y desarrollar todos los componentes de SQL server y SQL database. Microsoft lo ha optimizado a lo largo de los años y es un programa de administración de servidores y bases de datos muy popular.

El presente trabajo es para la materia “Introducción a la ciencia de datos y sus metodologías” impartida por el Dr. Juan Pablo Soto en la maestría de ciencia de datos de la Universidad de Sonora.

Nos centraremos en el uso de una base de datos SQL, la cual administraremos mediante SSMS y desde jupyter notebook con la ayuda del lenguaje Python.

2. Objetivos.

Dentro de los objetivos que se realizaron en el presente trabajo fue:

- Creación de base de datos.
- Creación de conexiones
- Importación de datos
- Creación de una vista
- Creación de procedimiento almacenado y
- Creación de una función.

Herramientas utilizadas:

- Microsoft SQL Server
- Anaconda
- Jupyter notebook
- Python
- SQL

3. Fuente de datos

Los datos utilizados fueron obtenidos de Kaggle.

Los accidentes aéreos son extremadamente raros, especialmente en los últimos años. Sin embargo, en situaciones de tragedia, a menudo se consigue obtener un registro de varios parámetros de vuelo, así como de las palabras pronunciadas en la cabina justo antes del desastre.

El conjunto de datos seleccionado presenta el registro de la última frase pronunciada por los pilotos del avión antes del accidente, de las 82 catástrofes para las que se obtuvieron tales resultados, así como la fecha de la catástrofe, la aerolínea y el número de vuelo.

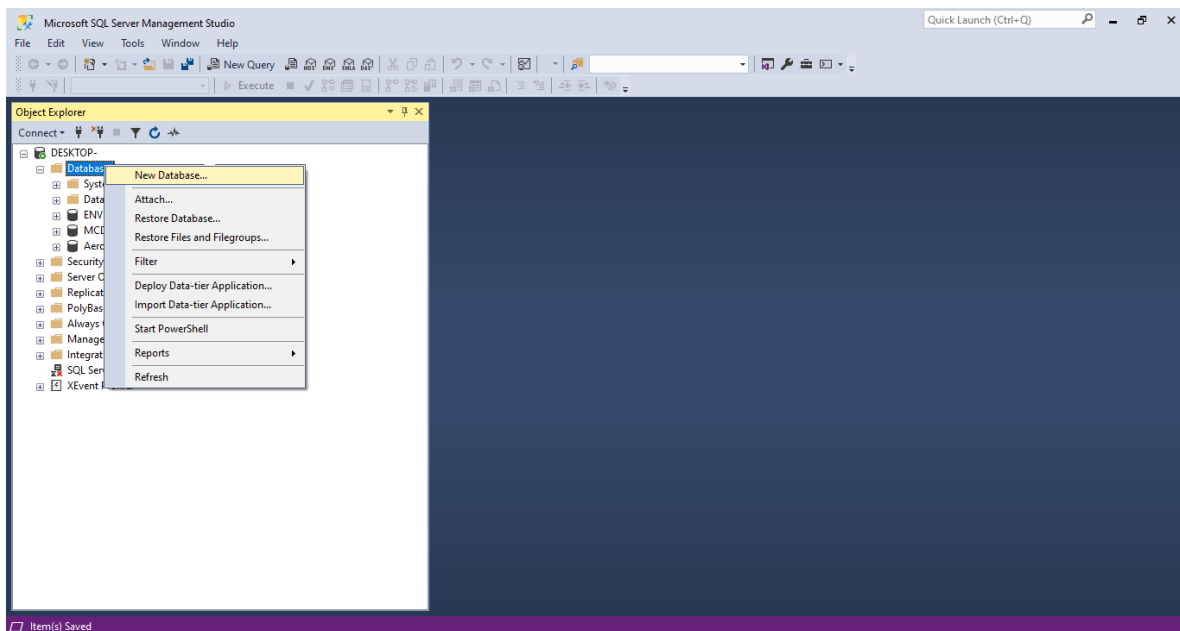
3.1 Diccionario de datos

- Id: numérico único para cada renglon de nuestros datos.
- Date: Fecha en la que ocurrió el suceso.
- Airline: Aerolínea de la que sucedió el suceso.
- Flight_number: número del vuelo.
- Last_words: Última oración que dijeron los pilotos del avión antes del accidente.

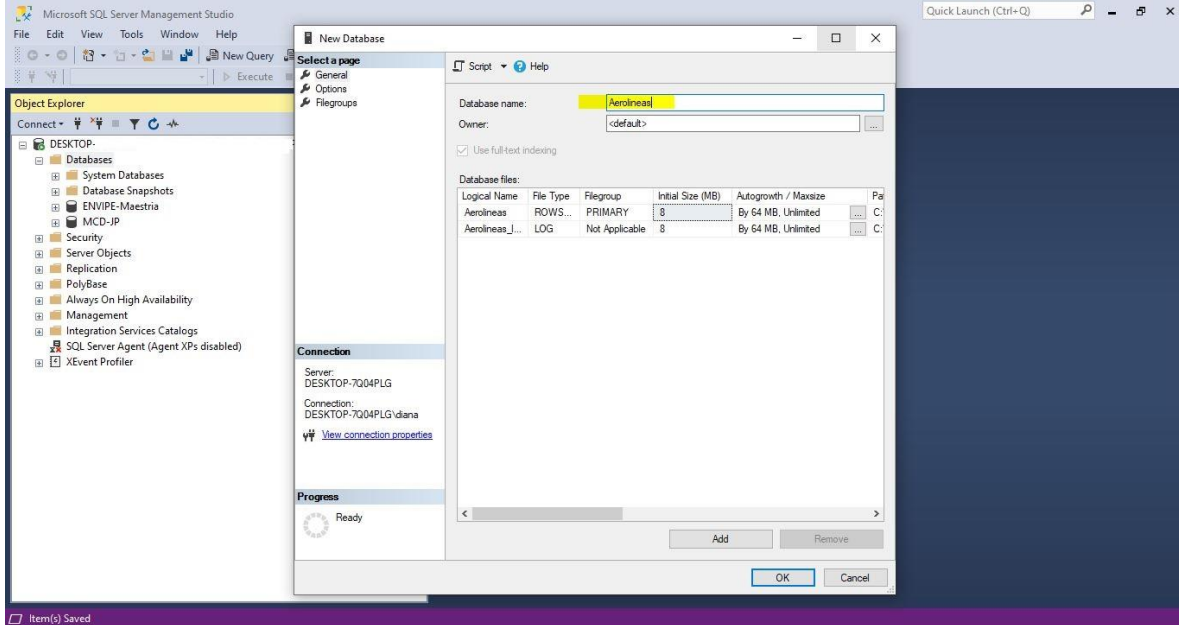
4. Desarrollo

4.1. Creación de base de datos

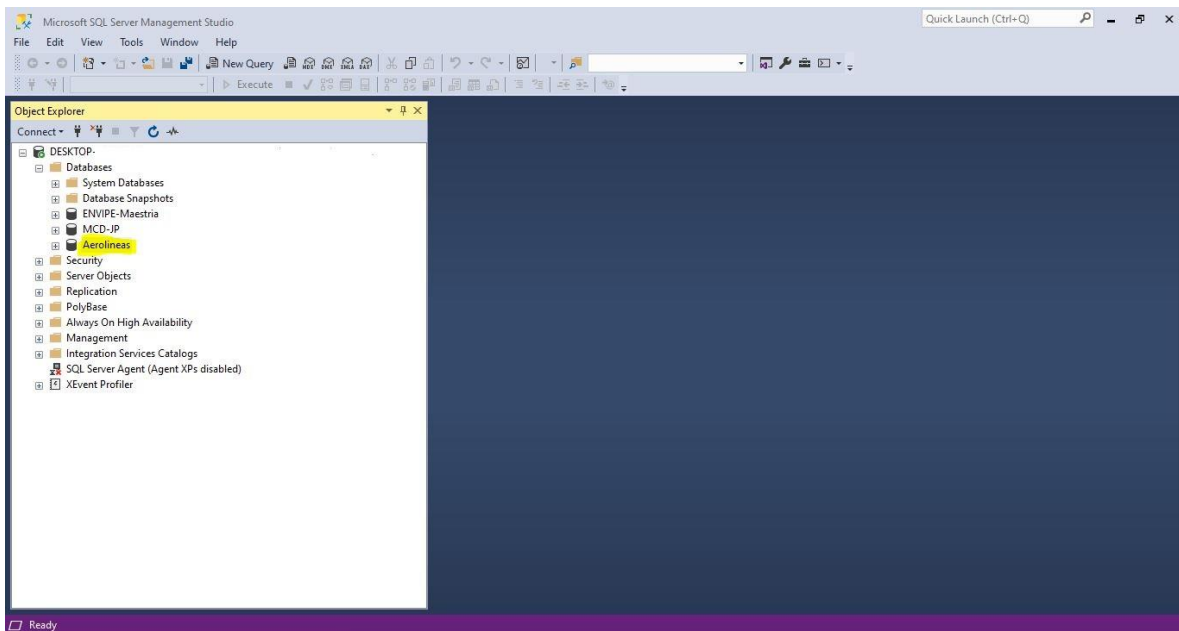
Para crearla, abrimos MySQL con nuestras credenciales y damos clic derecho en "Databases"



Nos aparecerá un recuadro donde le tenemos que dar un nombre a la base de datos.



Damos clic en "ok" y nuestra base de datos nueva nos debería de aparecer de lado izquierdo junto con las demás bases de datos (en dado caso que tengamos)



Y listo, ahora podemos comenzar con el proceso de subir nuestros datos con ayuda de python.

4. 2. Conectandonos a base de datos

Importamos librerías necesarias
import pandas as pd
import pyodbc

Leemos el archivo que queremos convertir a sql

```
df= pd.read_csv("Last_worlds_2022_11_13.csv", encoding='latin1')
```

4.2.1 Conectandonos a la base de datos

El primer paso es conectarnos con python al servidor de SQL

```
conn = pyodbc.connect('Driver={SQL Server};'  
                      'Server=;' #Nombre de mi servidor local  
                      'Database=Aerolineas;' #Nombre de la base de datos  
                      'Trusted_Connection=yes;')
```

```
cursor = conn.cursor()
```

4.3 Creación de tabla para nuestros datos

El siguiente paso es crear una tabla donde importaremos nuestros datos del archivo SQL

```
cursor.execute("""  
    CREATE TABLE llamadas (  
    id int primary key,  
    Date nvarchar(50),  
    Airline nvarchar(50),  
    Flight_number nvarchar(50),  
    Last_words nvarchar(120))  
    """)
```

OJO: Tener en cuenta que cada vez que ejecutemos este código se creará una tabla, por lo que si ejecuta el mismo código (con el mismo nombre de la tabla) mas de una vez le dará error.

Lo siguiente es insertar nuestros datos del archivo excel en la tabla de productos que creamos

```
for row in df.itertuples():  
    cursor.execute("""  
        INSERT INTO llamadas (id, Date, Airline, Flight_number, Last_words)  
        VALUES (?, ?, ?, ?, ?)  
        """,  
        row.id,
```

```

        row.Date,
        row.Airline,
        row.Flight_number,
        row.Last_words
    )
conn.commit()

```

Con ayuda python, lo que haremos también será un query para que nos diga los valores que hay en la base de datos ya después de insertar todos los valores de nuestro archivo.

NOTA: Haré el query de "SELECT" función para poder llamarla varias veces (ya que mas adelante agregaremos valores y eliminaremos, y será más fácil solo llamar a la función para ver el contenido de la base de datos).

Solo llamaremos a 3 valores cuando hagamos SELECT; será a el id, la fecha y la aerolínea (por ningún motivo especial, solo para no llamar a todos los valores de cada renglón).

```

def SELECT():
    sql_select_query = "SELECT * FROM llamadas"
    cursor.execute(sql_select_query)
    records = cursor.fetchall()

    print("\nImprimiendo valores de la base de datos")
    for row in records:
        print("Id = ", row[0], )
        print("Fecha = ", row[1])
        print("Aerolinea = ", row[2], "\n")

```

SELECT() #Llamando a la función

Insertamos valores en nuestra tabla de productos

Insertaremos un dato falso que después eliminaremos para no alterar nuestros datos reales.

```

sql_insert_query = """INSERT INTO llamadas (id, Date, Airline, Flight_number,
Last_words) VALUES (87, '27 Nov 2022', 'Diana Airways', '01', 'Necesitamos ayuda')
"""

```

```

# cursor = conn.cursor()
cursor.execute(sql_insert_query)
# connection.commit()
print(cursor.rowcount, "valor insertado a la base de datos exitosamente!")

```

Ahora, eliminaremos un registro de la base de datos

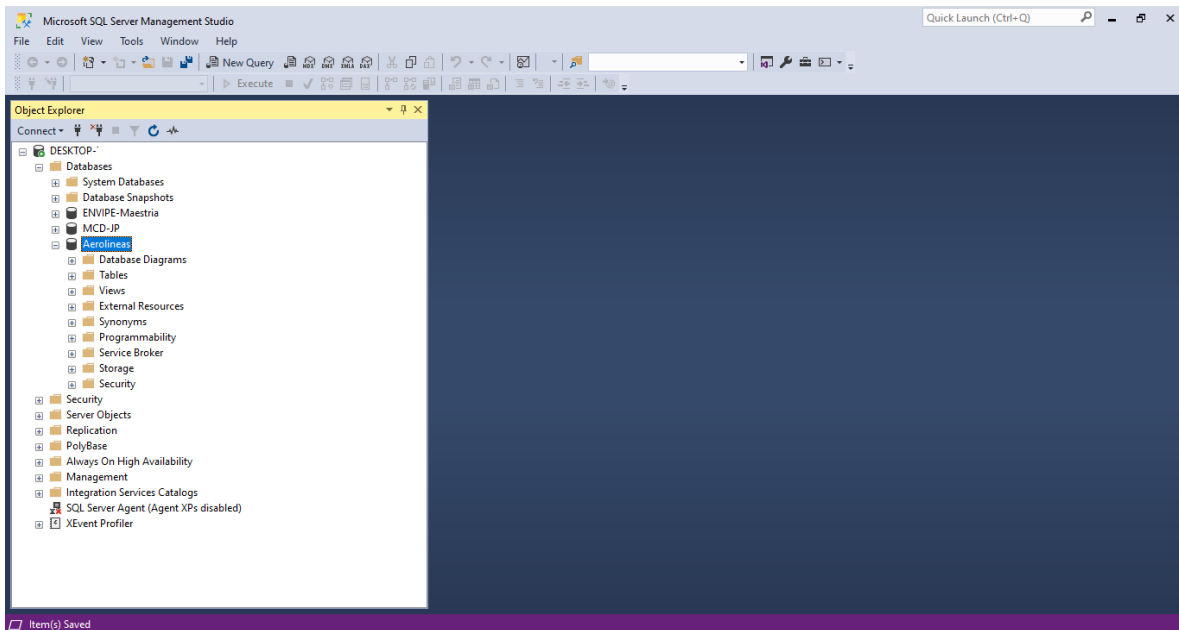
```
sql_Delete_query = """Delete from llamadas where id = 87"""  
cursor.execute(sql_Delete_query)  
print(cursor.rowcount, ' valor eliminado exitosamente!')
```

Y por último, cerramos la conexión a base de datos.

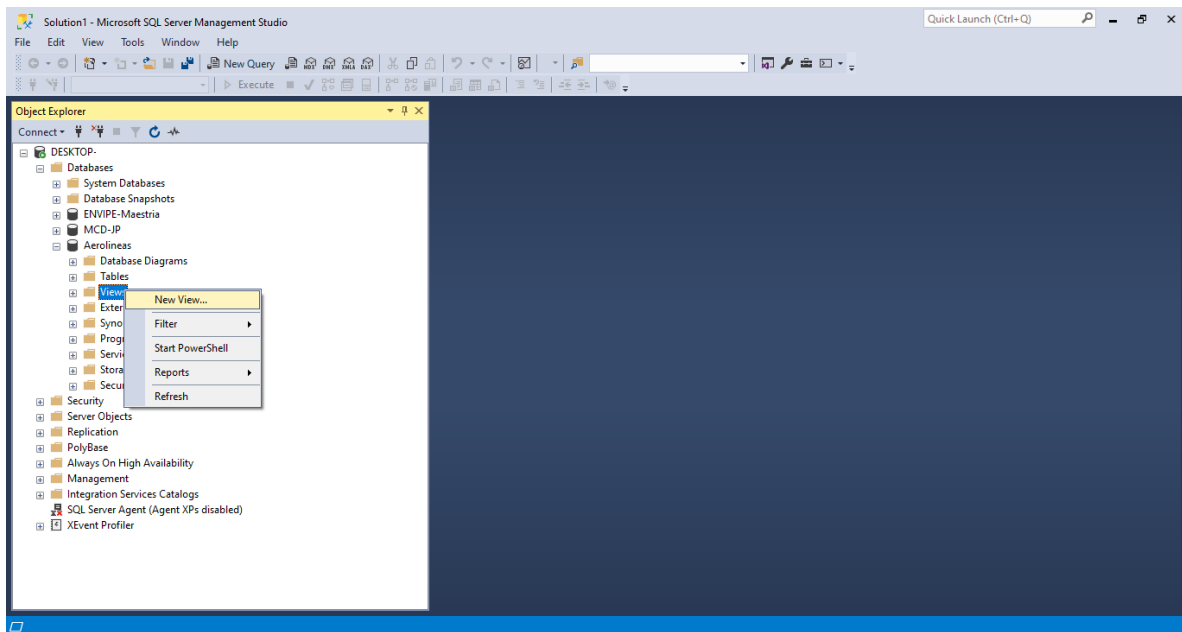
```
cursor.close() #Cerramos conexión
```

4.4 Creación de una vista de nuestra base de datos

Lo primero que tenemos que hacer es expandir nuestra base de datos (haciendo clic en el cuadrito con un más).

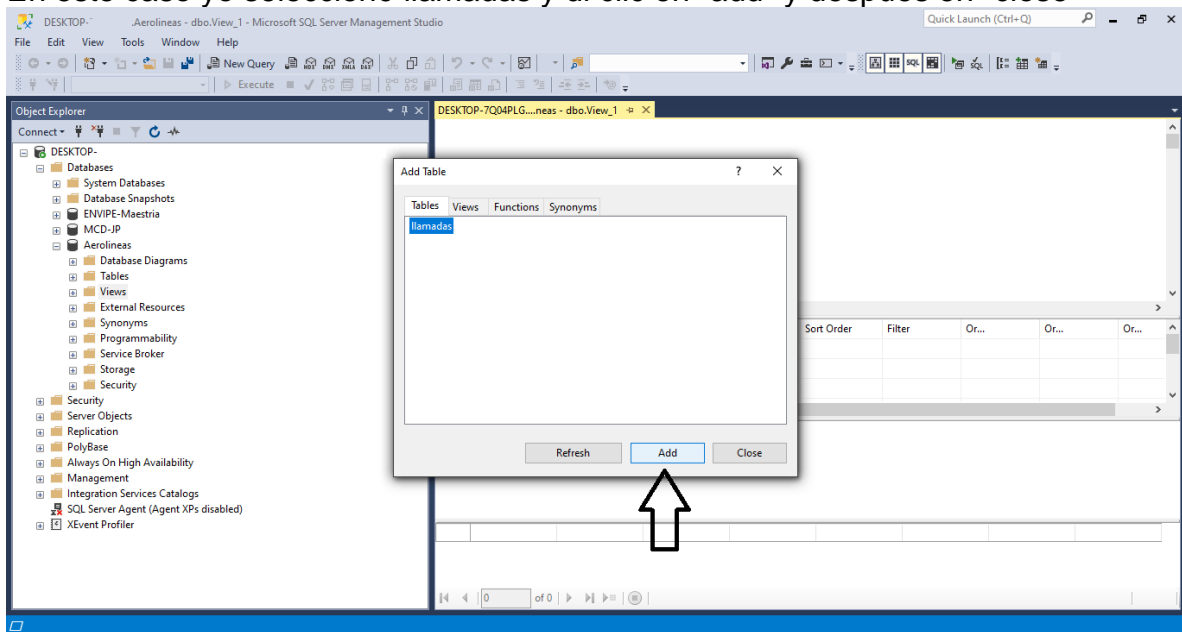


Daremos clic derecho en "view" y seleccionaremos "new view"

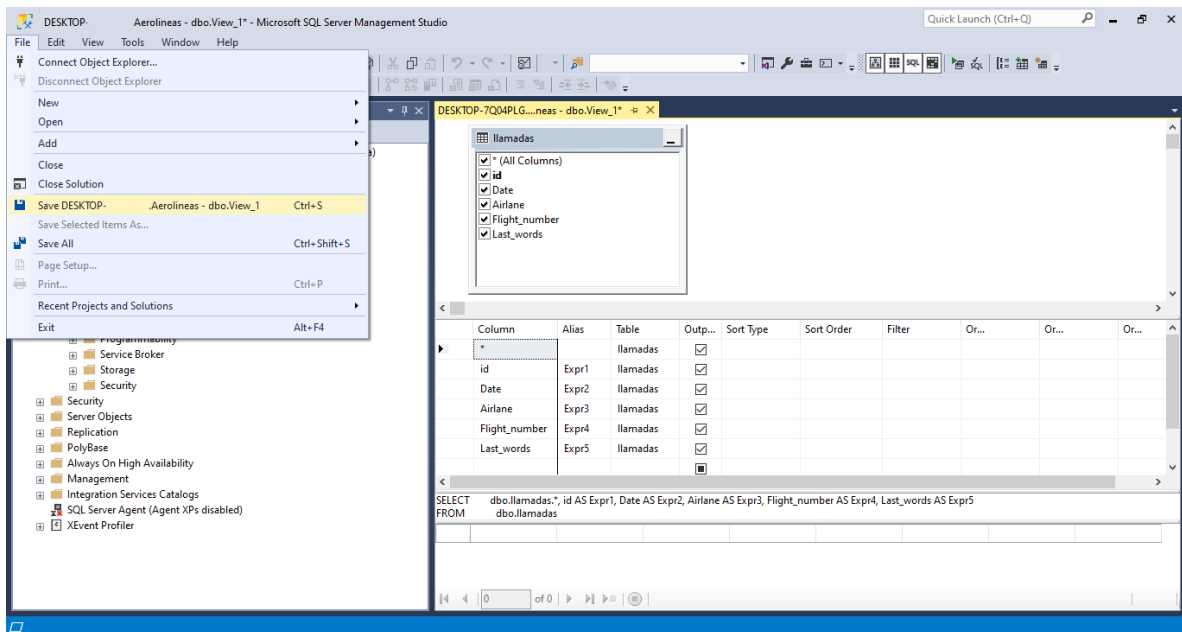


Seleccionamos la tabla de donde queremos hacer una vista de los datos

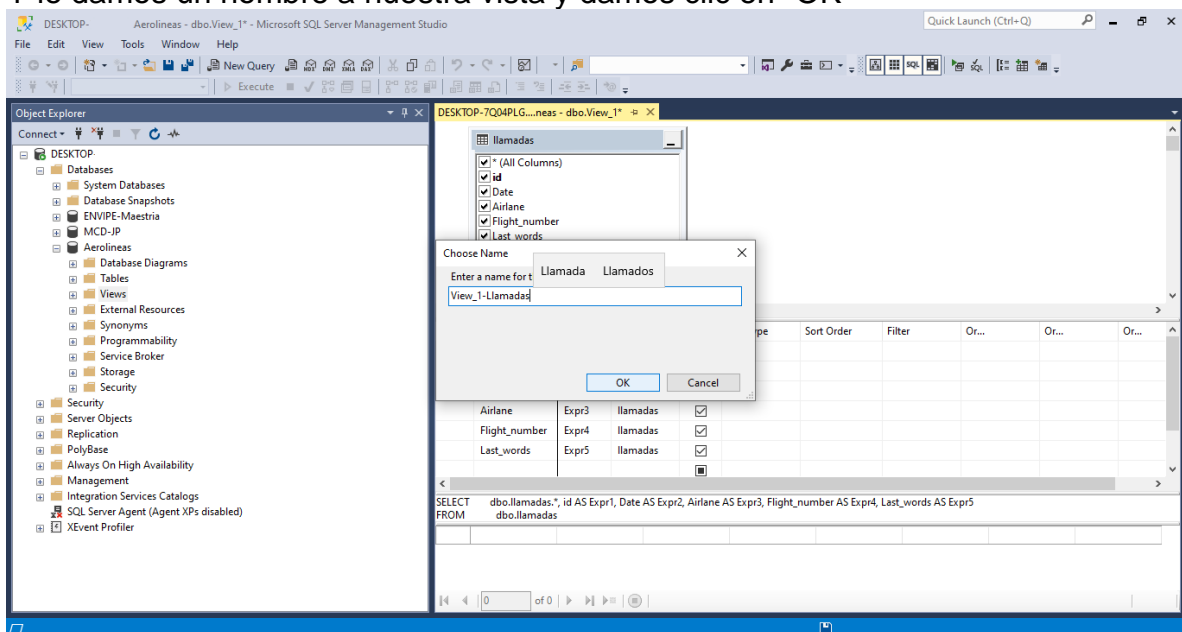
En este caso yo seleccioné llamadas y di clic en "add" y despues en "close"



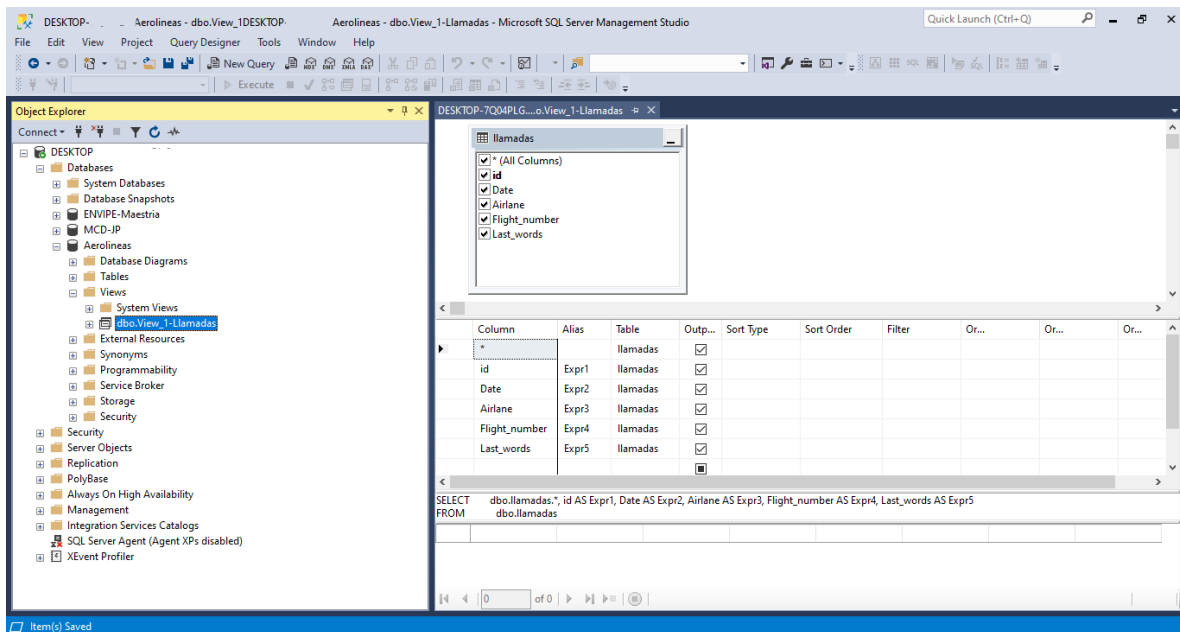
Despues en seleccionamos las variables que queremos y damos clic en la pestaña de "File" y damos clic en save view.



Y le damos un nombre a nuestra vista y damos clic en "OK"

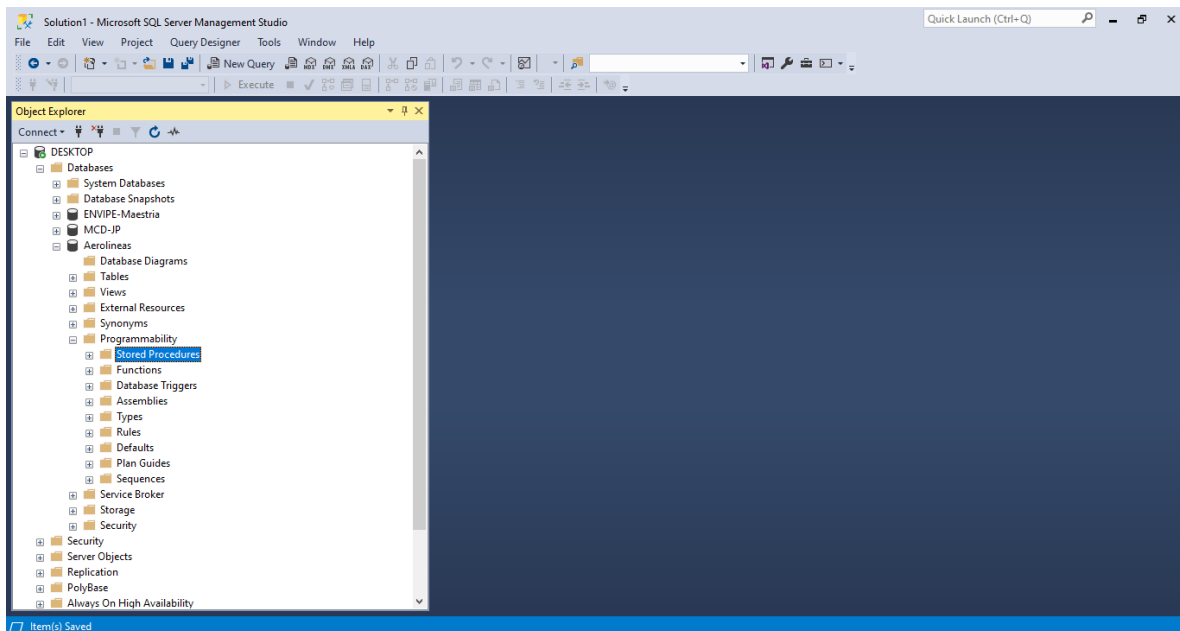


Si damos refresh veremos que aparece nuestra vista

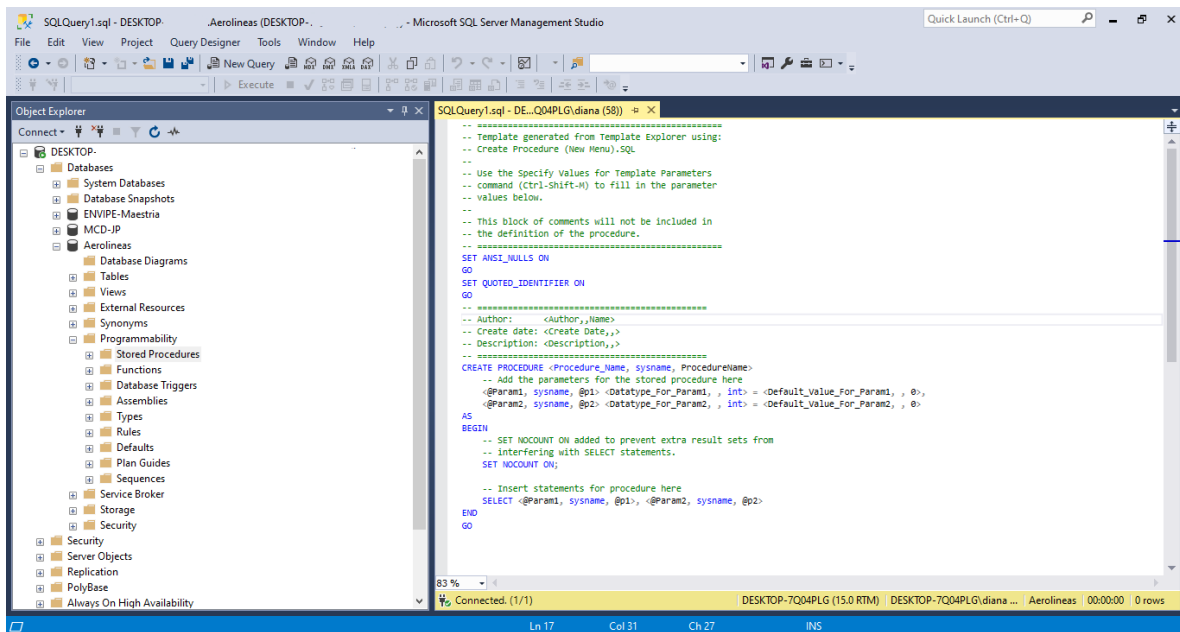


4.5 Creando un procedimiento almacenado

Extendemos dentro de nuestra base de datos "programmability" y dentro damos clic derecho en stored prodecures y damos clic en "New".



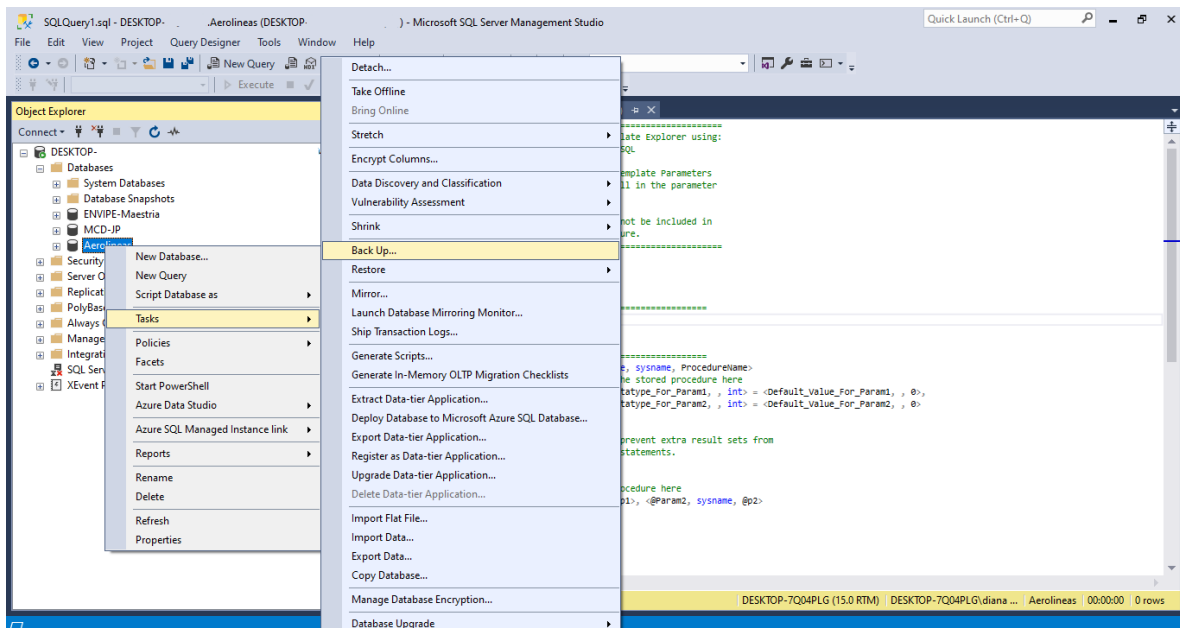
Y nos saldrá código de nuestro procedimiento almacenado



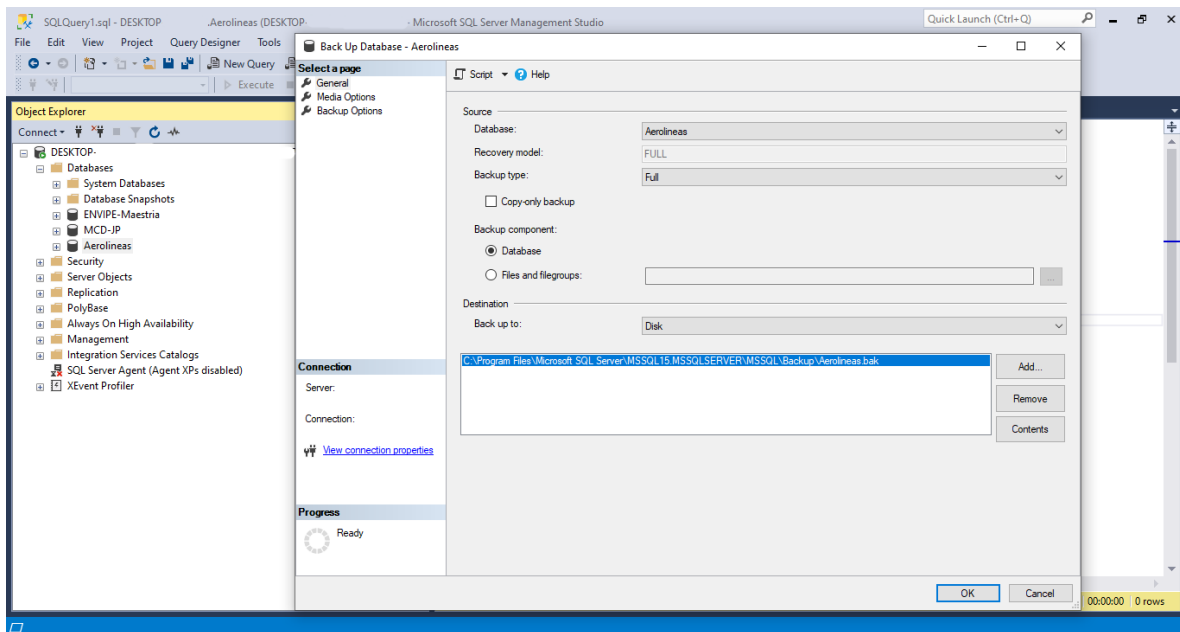
4.6 Backup

Y por último haremos un backup de nuestra base de datos para tenerla en nuestro github.

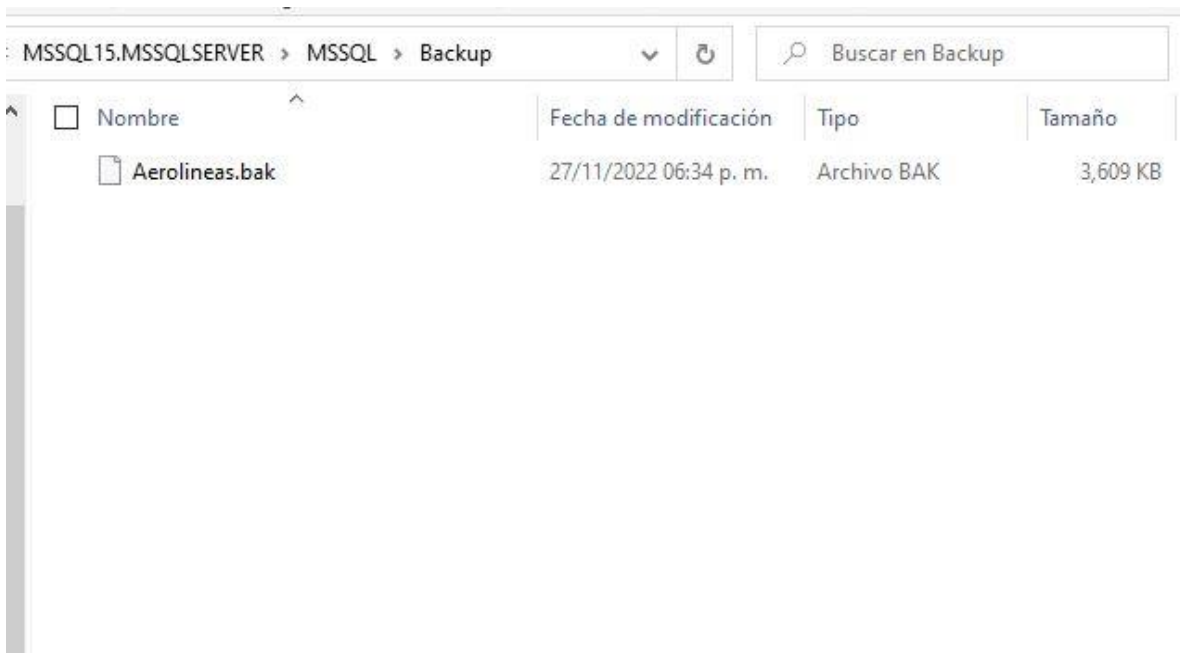
Daremos clic derecho a nuestra base de datos y daremos clic en "tasks" despues daremos clic en "Back up"



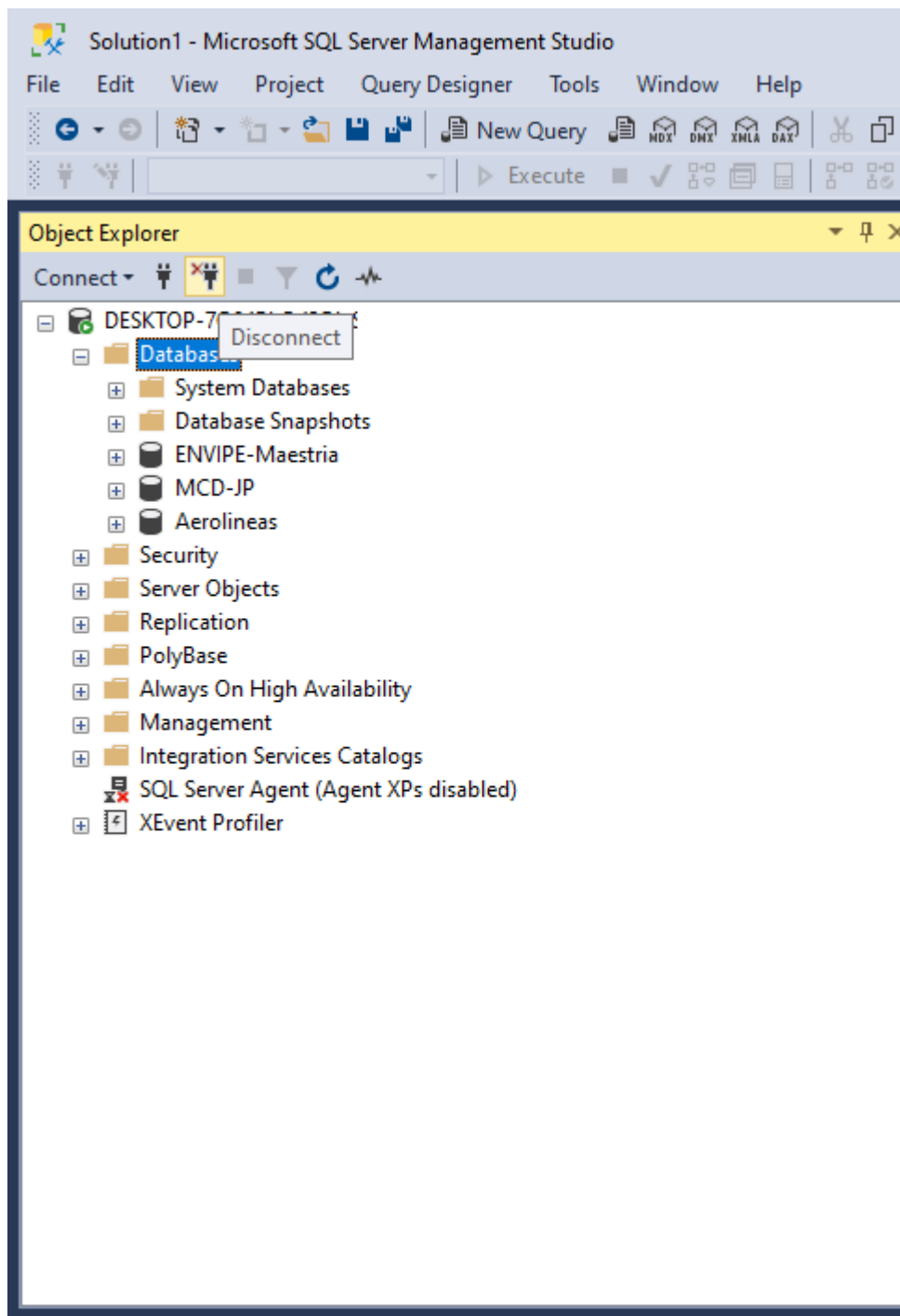
Nos aparecerá un reacuadro, donde nos dice donde se va a guardar nuestro archivo .bak (es el back up de la base de datos) y damos clic en "ok".



Y verificamos que el back up esté en la dirección que nos aparecía



Por último, nos desconectamos



5. Referencias

Last words before the plane crash. (2022, 13 noviembre). Kaggle.

<https://www.kaggle.com/datasets/michau96/last-words-before-the-plane-crash>

Softtrader, M. (2022, 16 marzo). *¿Qué es SQL Server Management Studio y qué puedo hacer con él?* Softtrader. <https://softtrader.es/blog-microsoft/que-es-sql-server-management-studio/>

CREATE PROCEDURE (Transact-SQL) - SQL Server. (2022, 26 septiembre).

Microsoft Learn. <https://learn.microsoft.com/es-es/sql/t-sql/statements/create-procedure-transact-sql?view=sql-server-ver16>