**UNIVERSIDAD TECNICA PARTICULAR DE LOJA**

Ingeniería en Ciencias de la Computación

Fundamentos de Base de Datos



**Nombre:**

Cristhian Villamarin

**Fecha:**

03/01/02

**Docente:**

Nelson Piedra

**Octubre – Febrero**

**Catedra Integradora**

1. **Introducción**

Las bases de datos relacionales son un tipo de base de datos que almacena y proporciona acceso a puntos de datos relacionados entre sí. Las bases de datos relacionales se basan en el modelo relacional. Las columnas de la tabla contienen los atributos de los datos y cada registro suele tener un valor para cada atributo, lo que simplifica la creación de relaciones entre los puntos de datos.

Con estos conocimientos previos empezamos con nuestro análisis del archivo csv de base de datos. Comienza con la identificación de todos los campos que están en el archivo continuando con unas tablas en la cuales detallamos varias propiedades de cada uno de los campos

1. **Desarrollo**

A continuación, se enlista todos los campos de la base de datos en el orden en el que se encuentran en el archivo:

* Index
* Budget
* Genre
* Homepage
* Id
* Keywords
* Original\_language
* Original title
* Overview
* Popularity
* Production\_company
* Production\_countries
* Reléase\_date
* Revenue
* Runtime
* Spoken\_language
* Status
* Tagline
* Title
* Vote\_average
* Vote\_count
* Cast
* Crew
* Director

**2.1 Clasificación**

Enlistados todos los campos procedemos mediante tablas a clasificar cada uno con el dominio, la opcionalidad, multivaluado, tipo, y comentario de tal forma que sepamos las características y nos facilite la realización del trabajo al momento de llevarlo a la práctica. Esta clasificación está dividida por tablas las cuales representan cada uno de los atributos

* *Movie*

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Campos** | **Dominio** | **Opcionalidad** | **Multivaluado** | **Tipo** | **Comentario** |
| Id\_movie | Int | Not Null | No | Primary key | Identificación de Película |
| index | Varchar(5) | Not Null | No | Primary Key | Indicador |
| keyword | Varchar(10) | Not Null | Si |  | Palabra clave para encontrar la película |
| Budget | Int | Not Null | No |  | presupuesto |
| Homepage | Varchar(30) | Not Null | No |  | Página web |
| OriginalLanguage | Varchar(30) | Not Null | No |  | Lenguaje original |
| Popularity | Int |  | No |  | popularidad |
| Cast | Varchar(30) |  | No |  | Elenco |
| Title | Varchar(20) | Not Null | No |  | Titulo |
| Relase\_date | DATE |  | No |  | Fecha de grabación |
| Runtime | Int |  | No |  | Tiempo de duración |
| Revenue | Int | Not Null |  |  | Ingresos |
| OriginalTitle | Varchar(30) | Not Null | No |  | Titulo original |

* *Production\_Company*

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Campos** | **Dominio** | **Opcionalidad** | **Multivaluado** | **Tipo** | **Comentario** |
| Id\_company | Int | Not Null | No | Primary key | Identificación de la compañía |
| companyName | Varchar(30) | Not Null | No |  | Nombre de compania |

* *Production\_Countries*

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Campos** | **Dominio** | **Opcionalidad** | **Multivaluado** | **Tipo** | **Comentario** |
| Iso | Varchar(10) | Not Null | No | Primary Key | Abreviatura del país |
| countriesName | Varcahr(30) | Not Null | No |  | Nombre del país |

* *Spoken\_Language*

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Campos** | **Dominio** | **Opcionalidad** | **Multivaluado** | **Tipo** | **Comentario** |
| Iso | Varchar(10) | Not Null | No | Primary Key | Abreviatura del lenguaje |
| languageName | Varcahr(30) | Not Null | No |  | Nombre del lenguaje |

* Crew

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Campos** | **Dominio** | **Opcionalidad** | **Multivaluado** | **Tipo** | **Comentario** |
| Id\_crew | Int | Not Null | No | Primary Key | Abreviatura del país |
| crewName | Varcahr(30) | Not Null | No |  | Nombre del país |
| Gender | Int | Not Null | No |  | genero |
| Credit\_Id | Int | Not Null | No |  | Identificador de créditos |

* *Department*

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Campos** | **Dominio** | **Opcionalidad** | **Multivaluado** | **Tipo** | **Comentario** |
| Id\_department | Varchar(20) | Not Null | No | Primary Key | Identificación de |
| departmentCrew | Varchar(30) | Not null | No |  | Nombre de director |

* *Job*

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Campos** | **Dominio** | **Opcionalidad** | **Multivaluado** | **Tipo** | **Comentario** |
| Id\_job | Varchar(20) | Not Null | No | Primary Key | Identificación de |
| job\_name | Varchar(30) | Not null | No |  | Nombre de director |

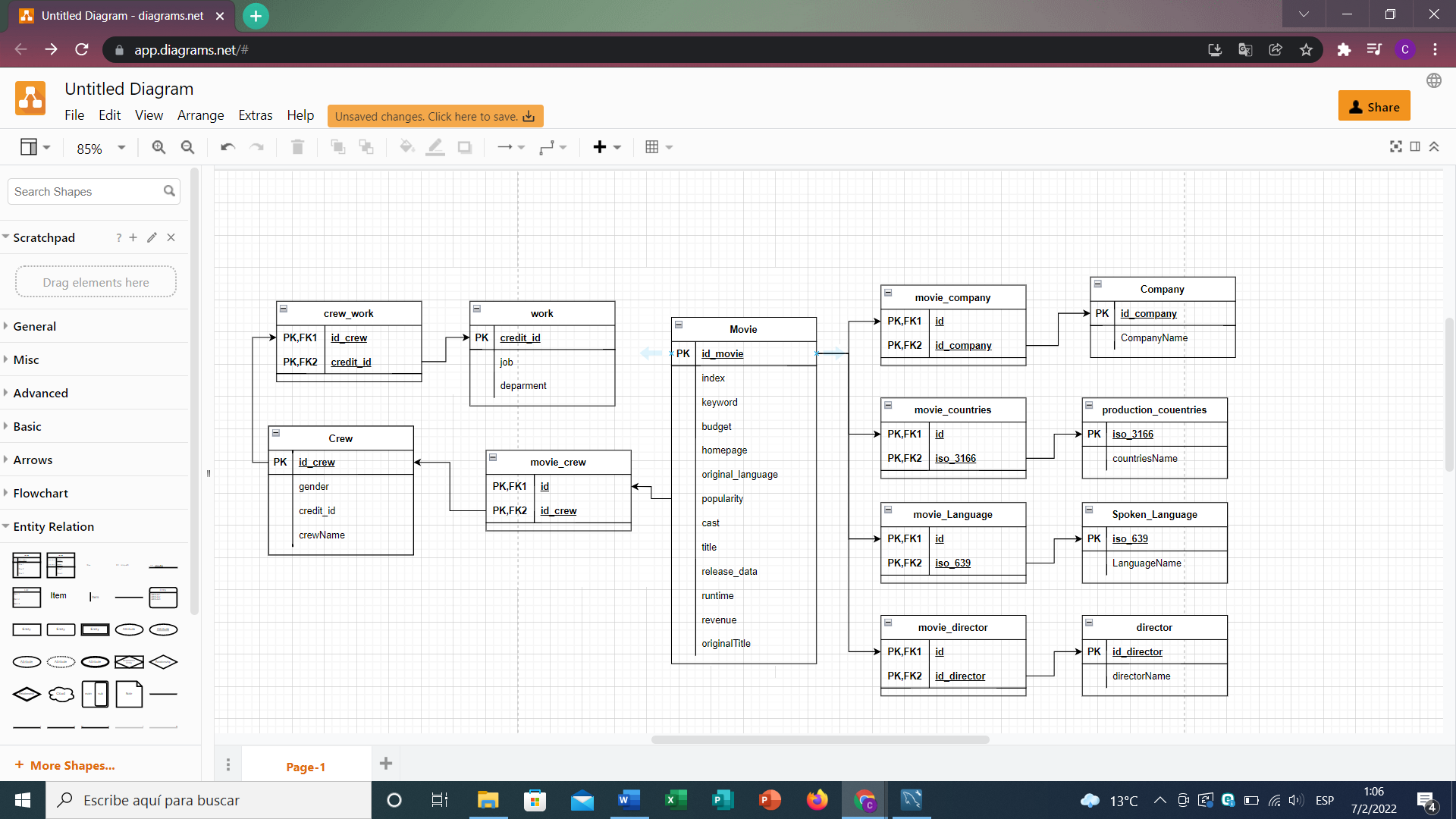
* *Cast*

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Campos** | **Dominio** | **Opcionalidad** | **Multivaluado** | **Tipo** | **Comentario** |
| Id\_cast | Varchar(20) | Not Null | No | Primary Key | Identificación de |
| castName | Varchar(30) | Not null | No |  | Nombre de director |

* *Director*

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Campos** | **Dominio** | **Opcionalidad** | **Multivaluado** | **Tipo** | **Comentario** |
| Id\_director | Varchar(20) | Not Null | No | Primary Key | Identificación de director |
| directorName | Varchar(30) | Not null | No |  | Nombre de director |

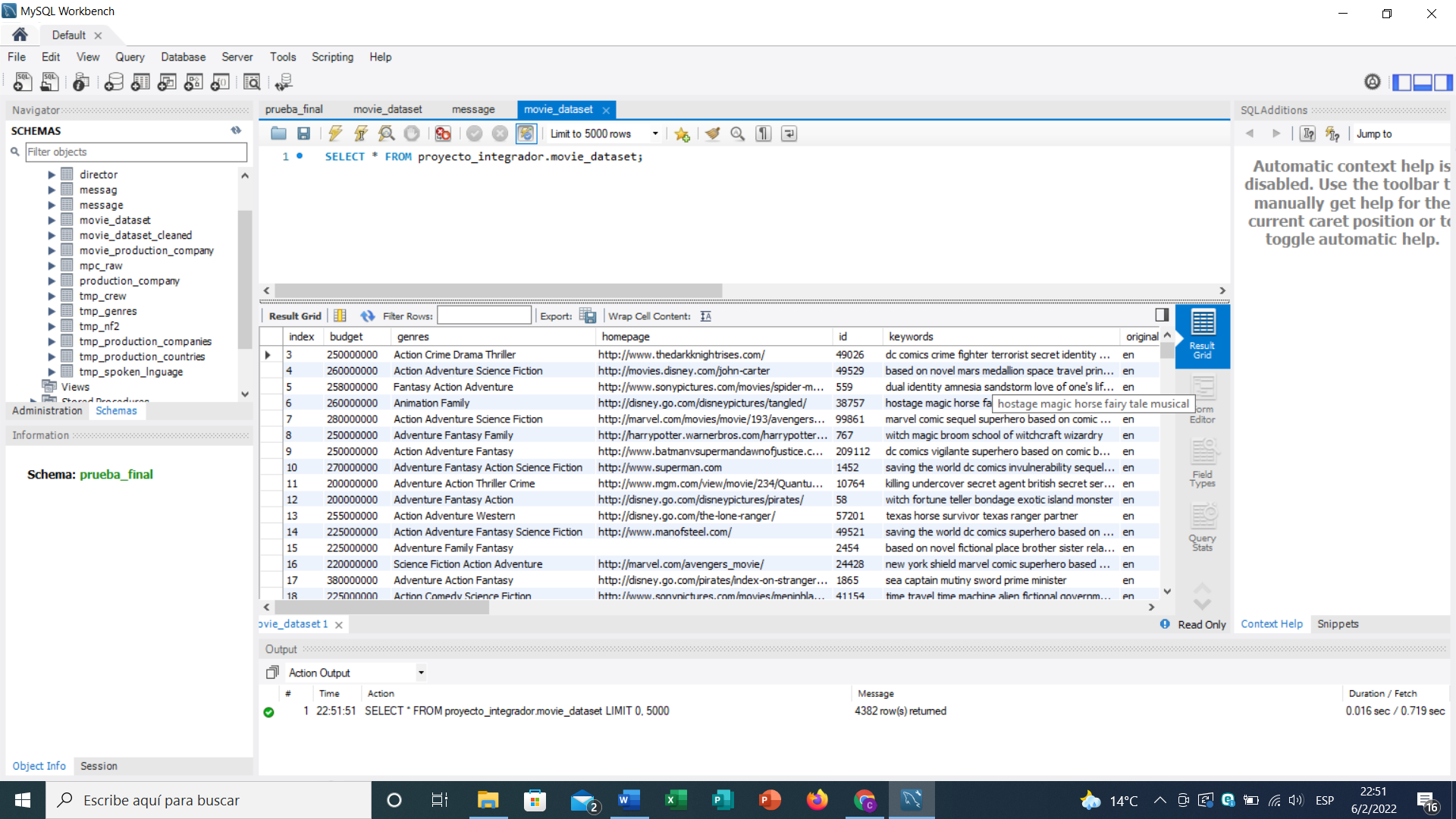
* 1. **Modelo E/R**

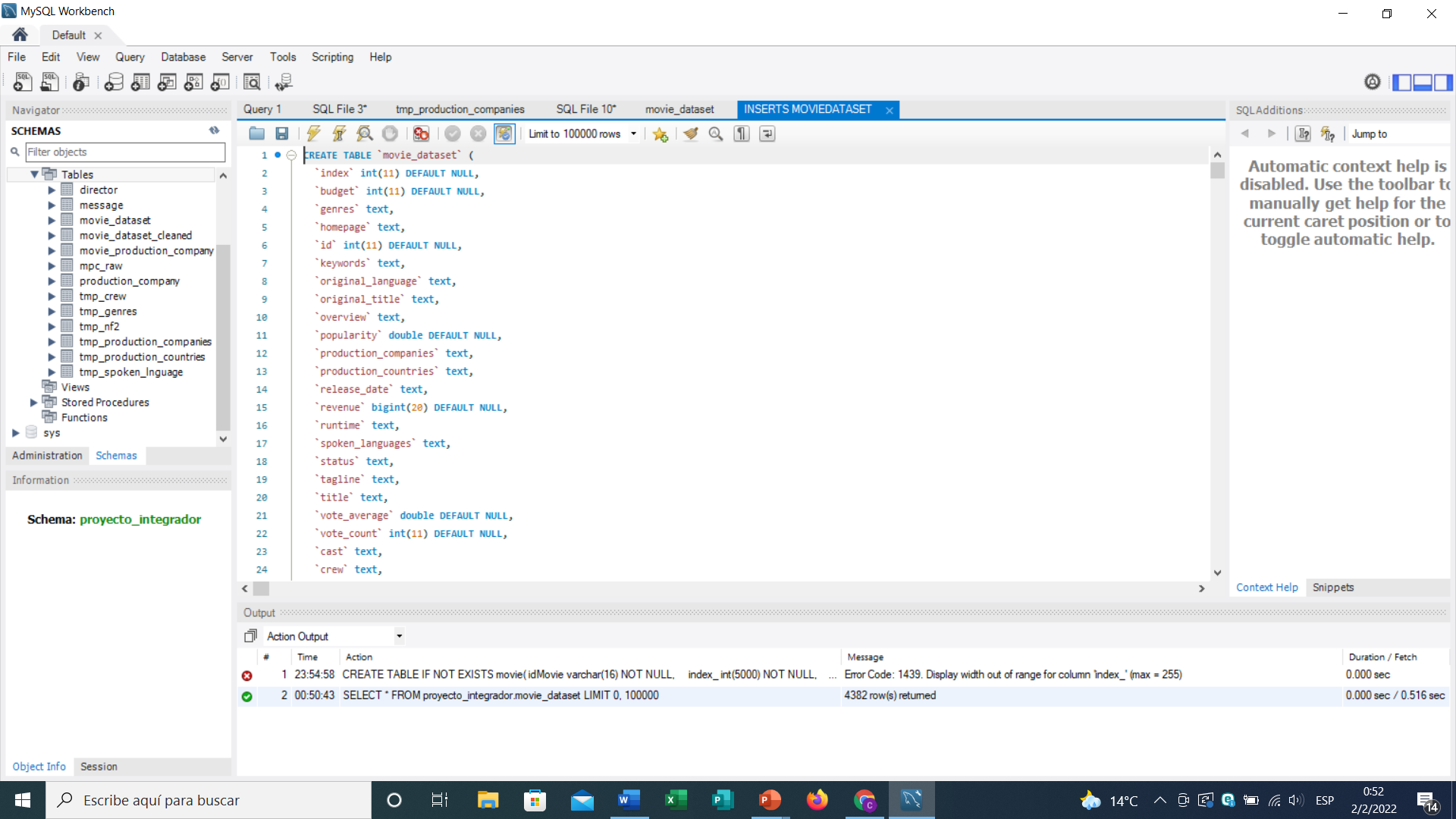


* 1. **Normalización**

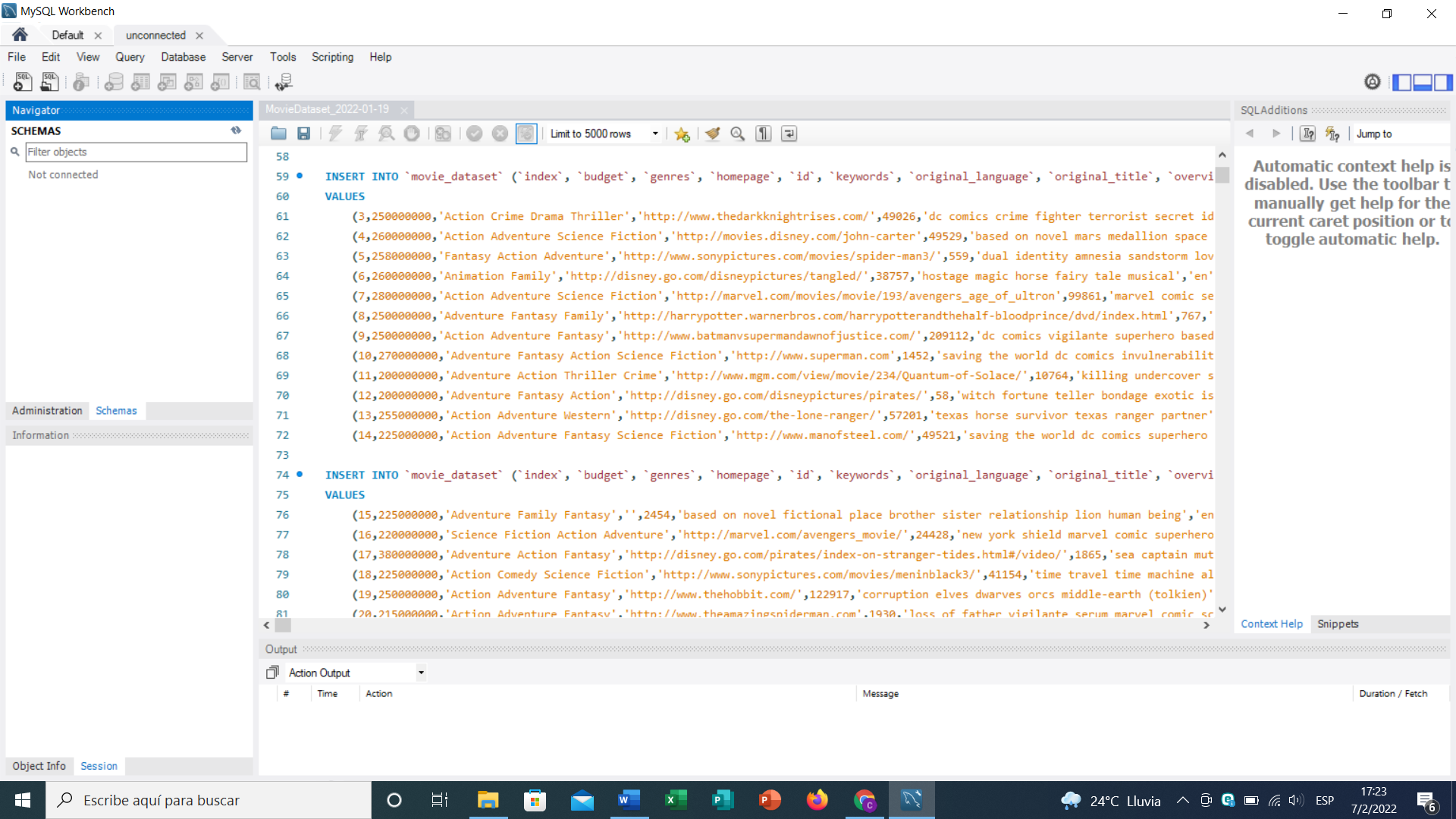
La normalización se refiere a un proceso el cual es manejado de tal forma que asignamos y organizamos los datos aplicando una serie de reglas a las relaciones obtenidas después estas reglas son utilizadas tanto para que los datos estén seguros y la base de datos sea más flexible al eliminar la redundancia y las dependencias incoherentes.

La tabla tal y como fue entregada sin ningún tipo de normalización o procedimiento es la siguiente:



****

Una vez teniendo la tabla creada procedemos a hacer la inserción de los datos a la misma:

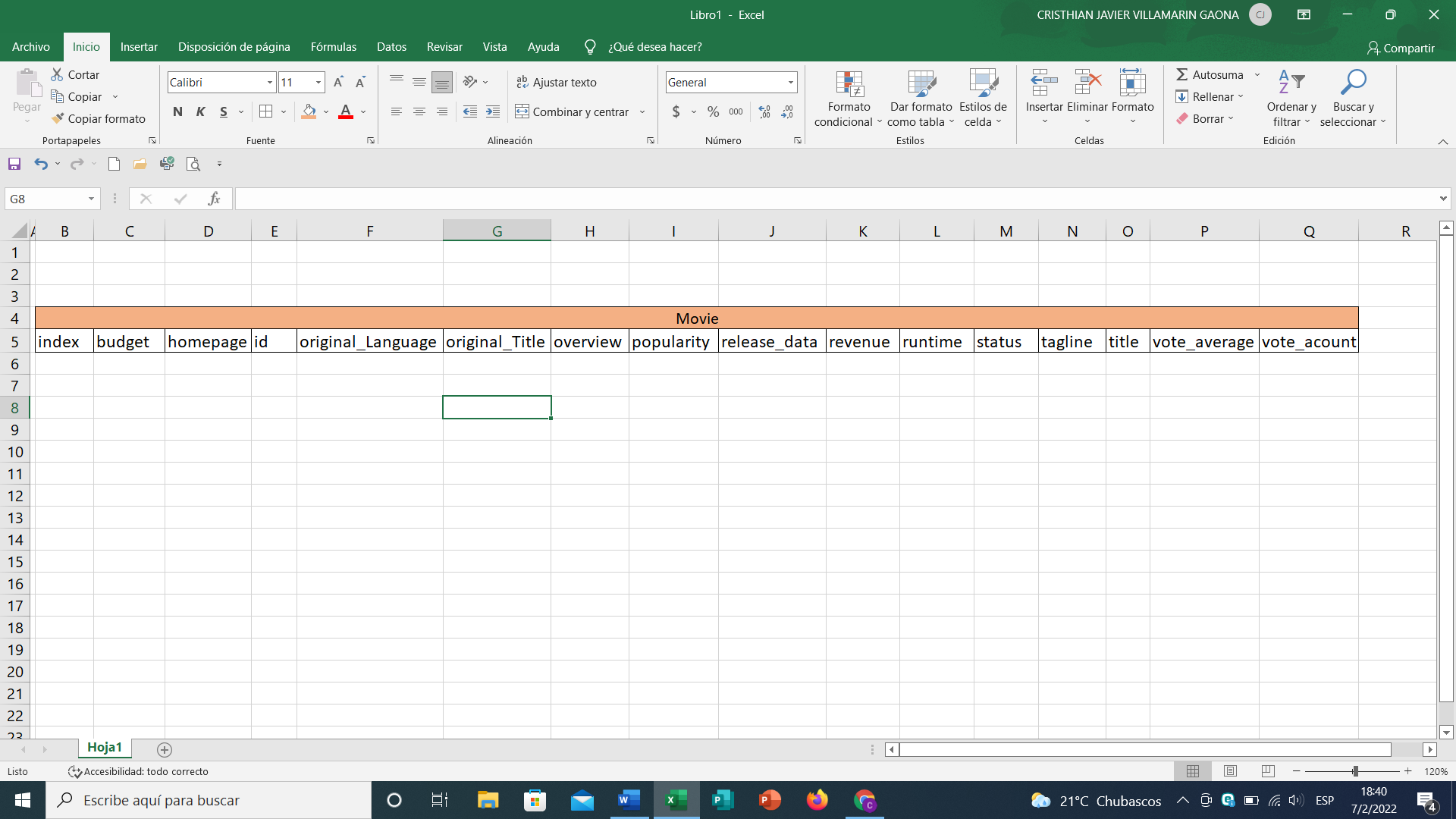


Teniendo toda la tabla “movie” creada y con todos los datos insertados continuamos a realizar la normalización, la cual consiste en los tres siguientes pasos:

* + 1. **Primera Forma Normal**

La primera forma normal consiste en eliminar los grupos repetidos de tablas individuales identificar cada tabla con su clave primaria y tener en cuenta que todos los datos sean atómicos

Para aplicar estas reglas a nuestra tabla movie se lo realiza separando los campos que no son atómicos y los que son multivaluados y convertirlos en tablas separadas de la tabla inicial y de esta manera tendríamos aplicada la primera forma normal



|  |  |
| --- | --- |
| production\_company | |
| id\_PC | CompanyName |

|  |  |
| --- | --- |
| production\_countries | |
| iso\_3166 | CountiresName |

|  |  |
| --- | --- |
| spoken\_language | |
| iso\_639 | LanguageName |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Crew | | | | | |
| Id\_Crew | credit\_id | name | gender | department | job |

|  |  |
| --- | --- |
| director | |
| id\_director | directorName |

Se tiene en cuenta que también se separan los campos multivaluados como keyword genres y cast:

|  |  |
| --- | --- |
| Keywords | |
| id\_movie | keywordsName |

|  |  |
| --- | --- |
| genres | |
| id\_movie | genresName |

|  |  |
| --- | --- |
| Cast | |
| id\_movie | castName |

|  |  |
| --- | --- |
| movie\_countries | |
| id\_movie | ido\_3166 |

El siguiente paso seria crear las relaciones de estas entidades con la tabla movie :

|  |  |
| --- | --- |
| movie\_company | |
| id\_movie | id\_company |

|  |  |
| --- | --- |
| movie\_Language | |
| id\_movie | iso\_639 |

|  |  |
| --- | --- |
| movie\_director | |
| id\_movie | id\_director |

|  |  |
| --- | --- |
| movie\_crew | |
| id\_movie | id\_crew |

* + 1. **Segunda Forma Normal**

Comprobamos que la primera forma normal se cumple en todas las entidades y con esta información podemos avanzar a la segunda forma de normalización la cual se basa en las

dependencias funcionales completas que deben tener todos los campos de una entidad.

Después de un profundo análisis se ha llegado a la conclusión que después de cumplir la primera forma normal todos los campos que hemos creado cumplen también con la segunda forma normal por lo que no es necesario la creación de ninguna otra entidad, relación u otros campos adicionales.

* + 1. **Tercera Forma Normal**

La tercera forma normal debe de realizarse solo si se cumplen las dos anteriores formas y que consiste en la revisión y solucion de las dependencias funcionales transitivas que ocurren cuando un campo no clave depende de una clave primaria, pero a su vez otros campos no claves dependen de este primer campo no clave.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Crew | | | | | |
| Id\_Crew | credit\_id | name | gender | department | job |

Después de analizar las tablas se llego a la siguiente solución:

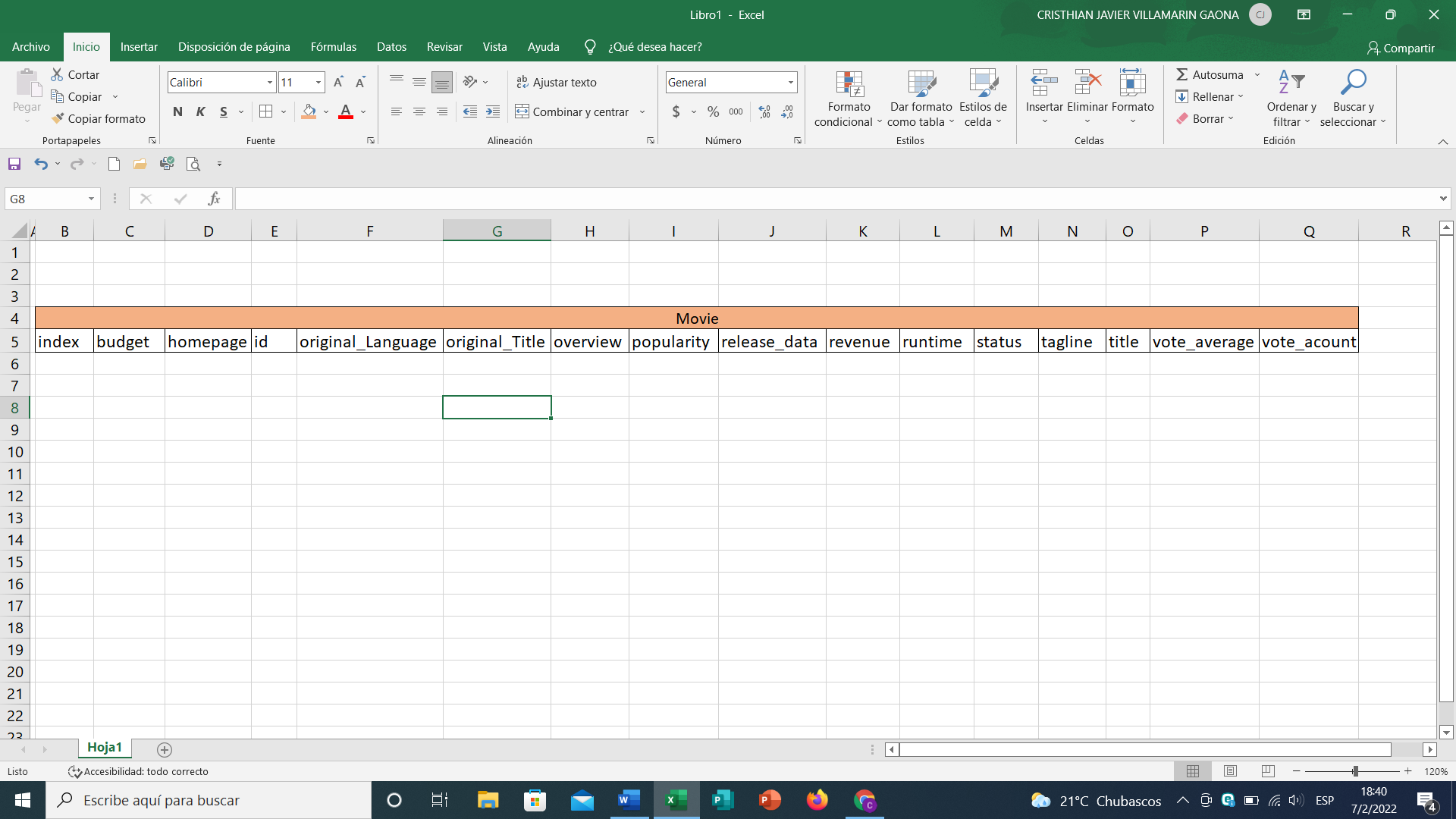
|  |  |  |
| --- | --- | --- |
| Crew | | |
| id\_crew | gender | crewName |

|  |  |  |
| --- | --- | --- |
| Work | | |
| credit\_id | department | job |

|  |  |
| --- | --- |
| Crew\_work | |
| id\_crew | credit\_id |

* + 1. **Resultado Final de la Normalizacion**

Una vez aplicada las tres formas normales hemos conseguido un resultado que ha satisfecho todas las incongruencias de la tabla principal y como desenlace final de este proceso de normalización hemos conseguido las siguientes entidades, campos y tablas:



|  |  |
| --- | --- |
| production\_company | |
| id\_PC | CompanyName |

|  |  |
| --- | --- |
| production\_countries | |
| iso\_3166 | CountiresName |

|  |  |
| --- | --- |
| spoken\_language | |
| iso\_639 | LanguageName |

|  |  |
| --- | --- |
| director | |
| id\_director | directorName |

|  |  |
| --- | --- |
| Keywords | |
| id\_movie | keywordsName |

|  |  |
| --- | --- |
| genres | |
| id\_movie | genresName |

|  |  |
| --- | --- |
| Cast | |
| id\_movie | castName |

|  |  |  |
| --- | --- | --- |
| Crew | | |
| id\_crew | gender | crewName |

|  |  |  |
| --- | --- | --- |
| Work | | |
| credit\_id | department | job |

|  |  |
| --- | --- |
| movie\_company | |
| id\_movie | id\_company |

|  |  |
| --- | --- |
| movie\_countries | |
| id\_movie | ido\_3166 |

|  |  |
| --- | --- |
| movie\_Language | |
| id\_movie | iso\_639 |

|  |  |
| --- | --- |
| movie\_director | |
| id\_movie | id\_director |

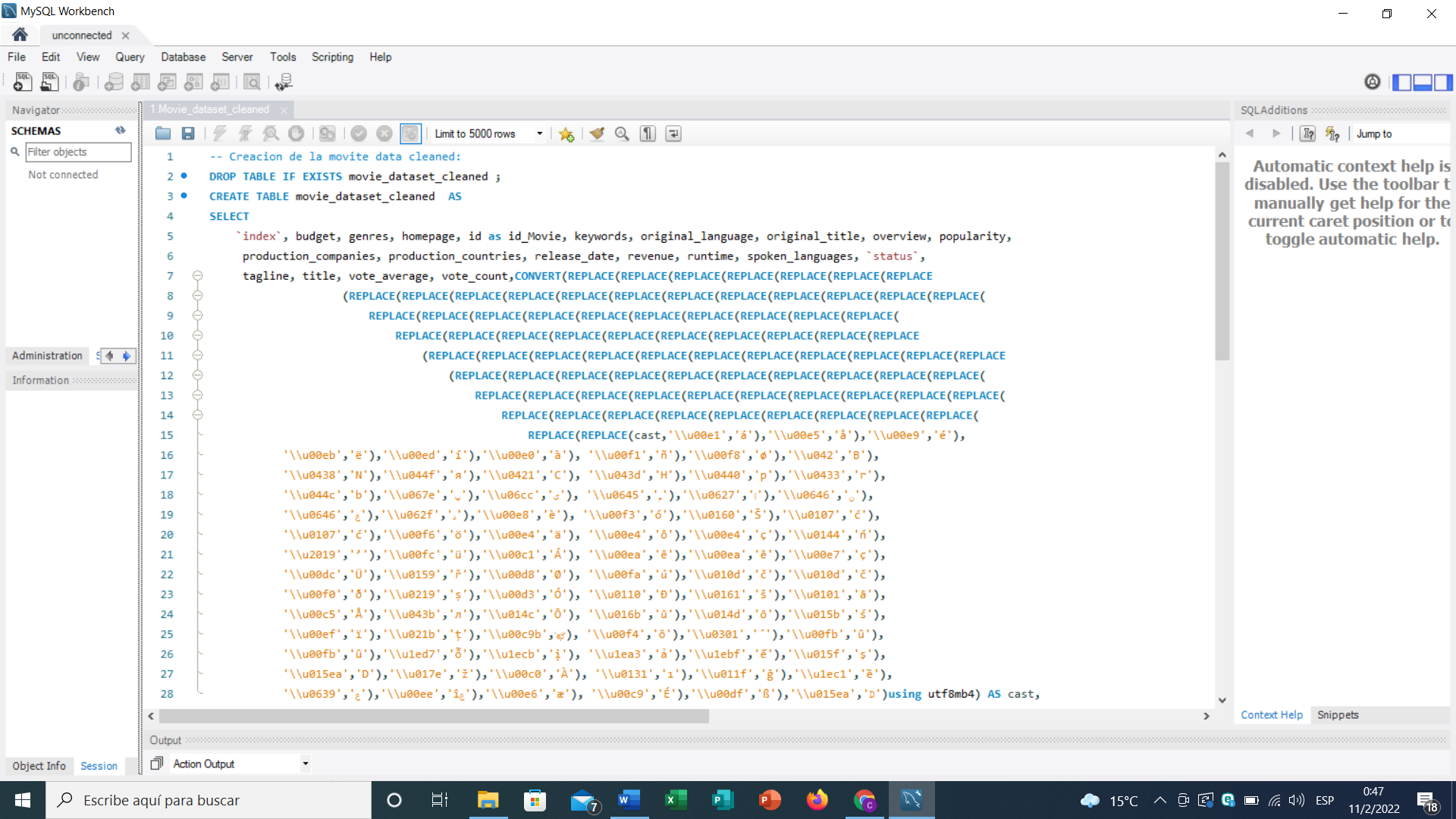
|  |  |
| --- | --- |
| movie\_crew | |
| id\_movie | id\_crew |

|  |  |
| --- | --- |
| Crew\_work | |
| id\_crew | credit\_id |

1. **CREACION DEL DDL**

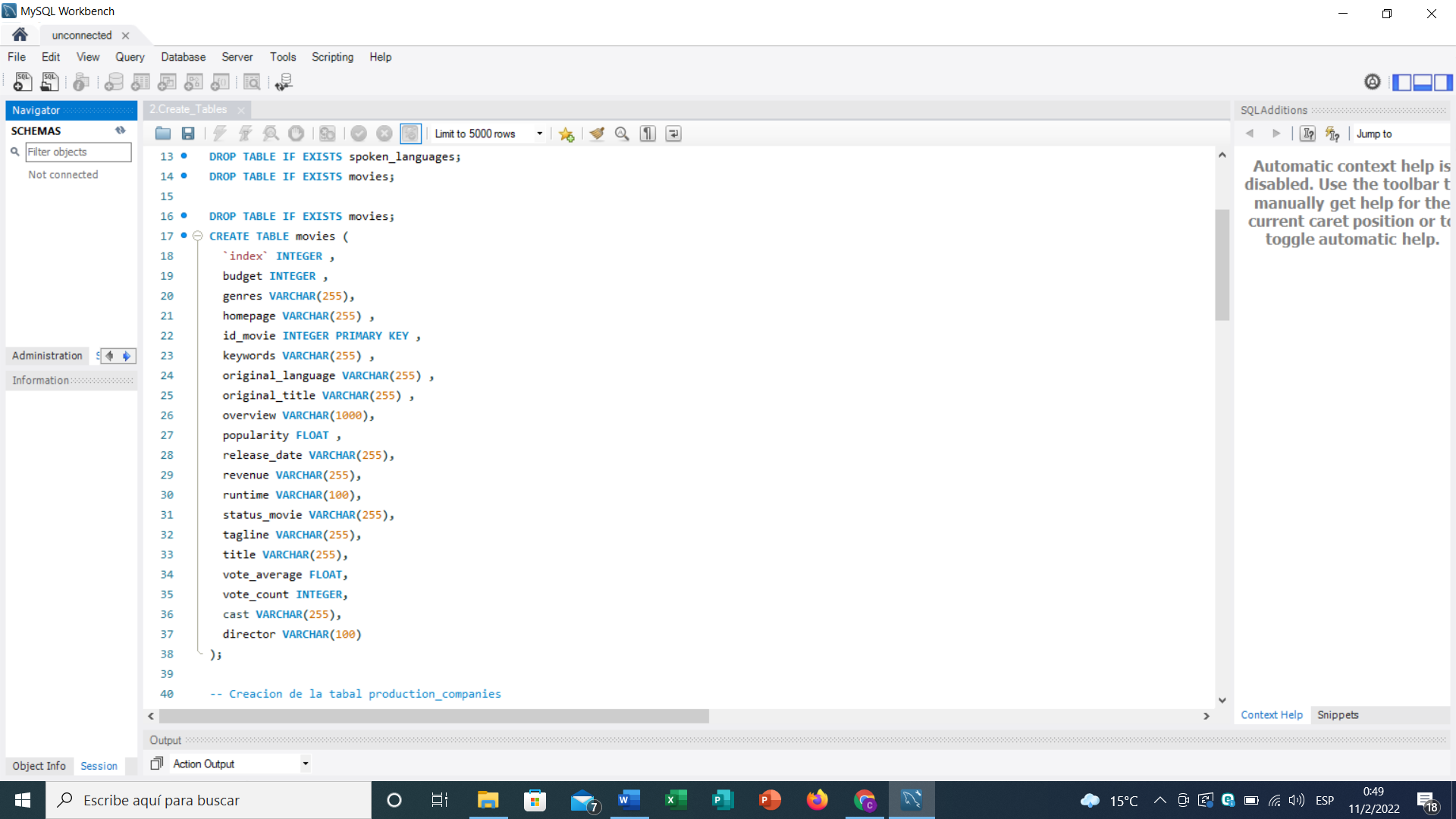
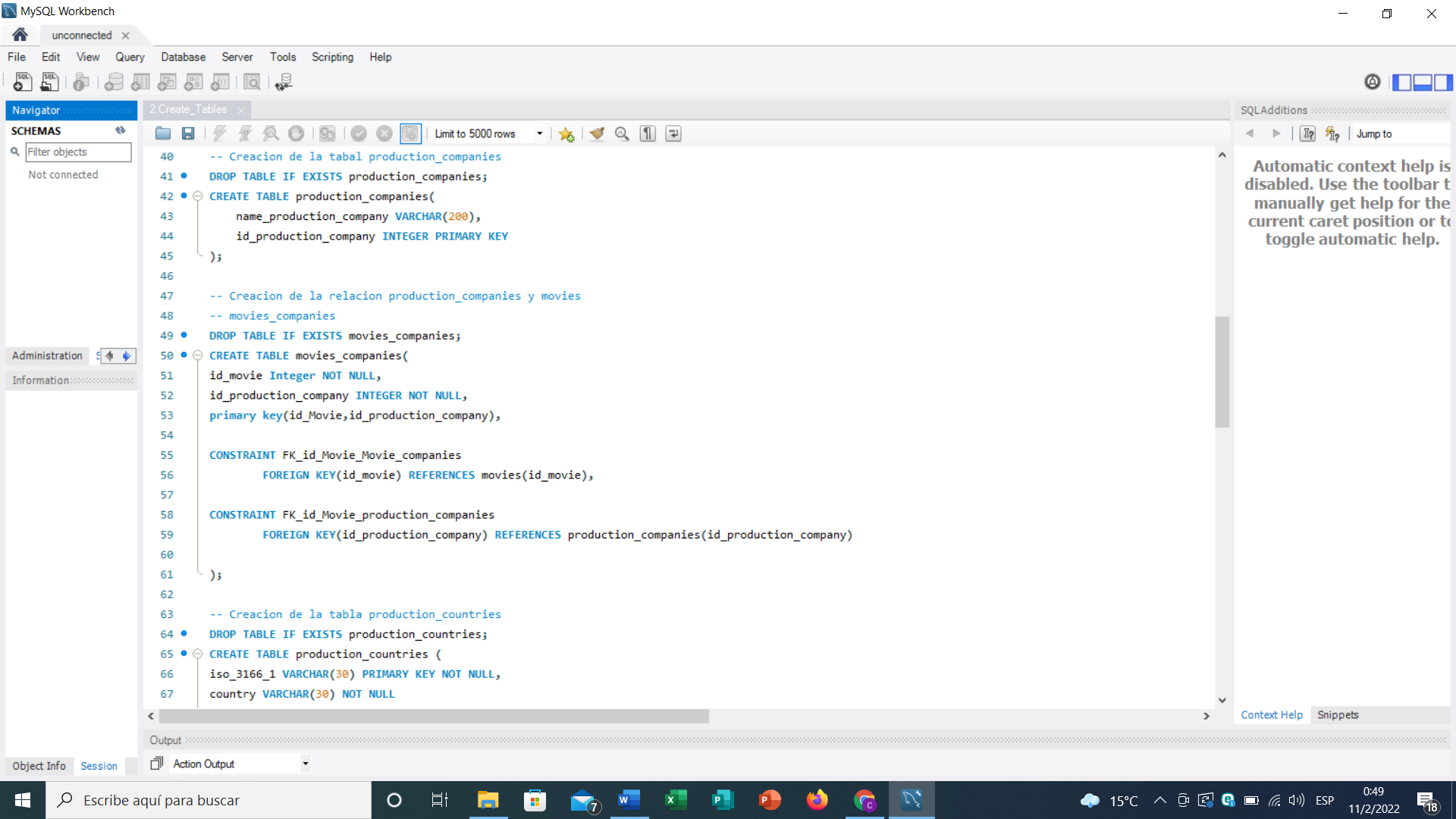
**3.1 Creación de movie\_dataset cleaned**

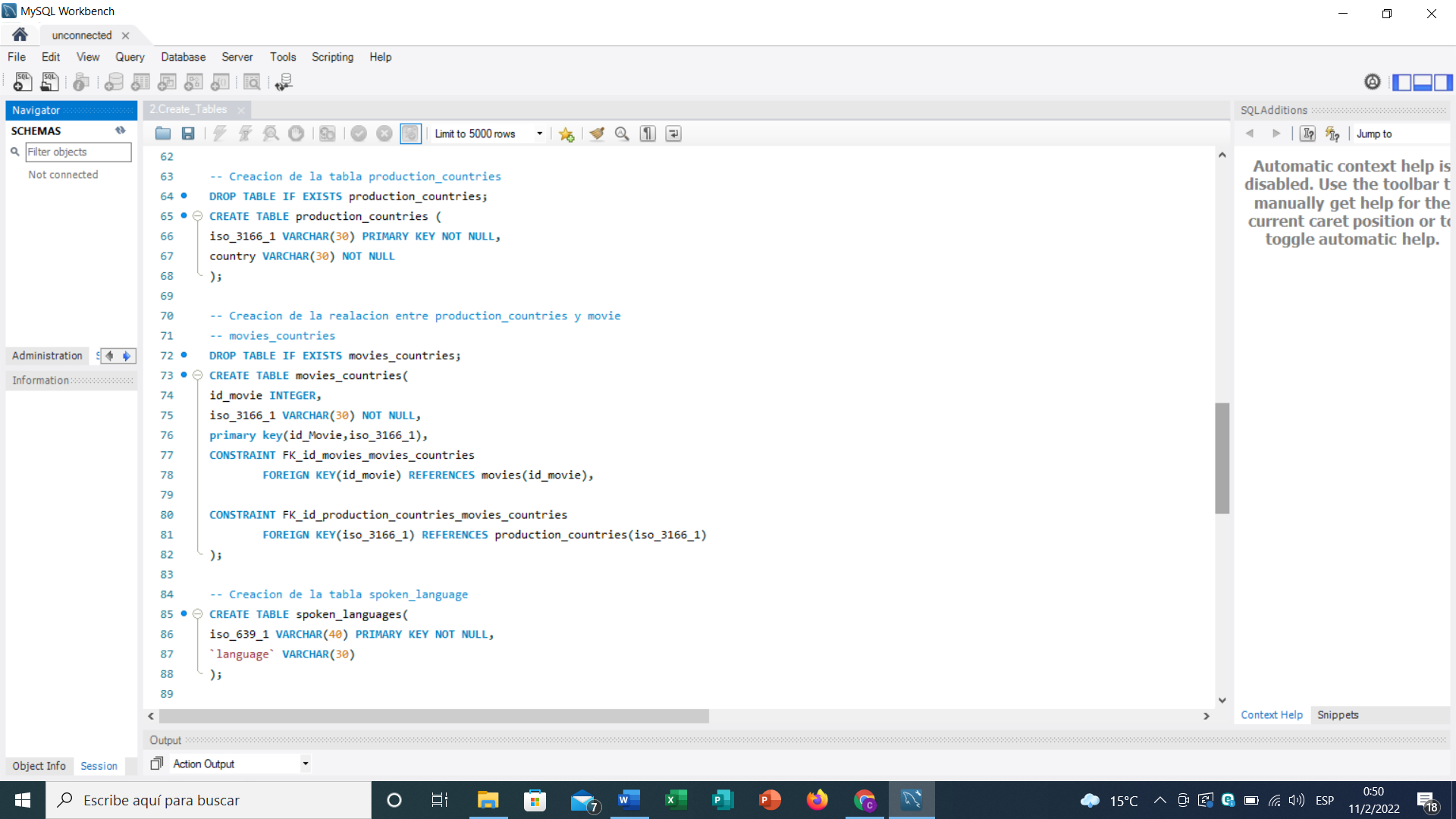
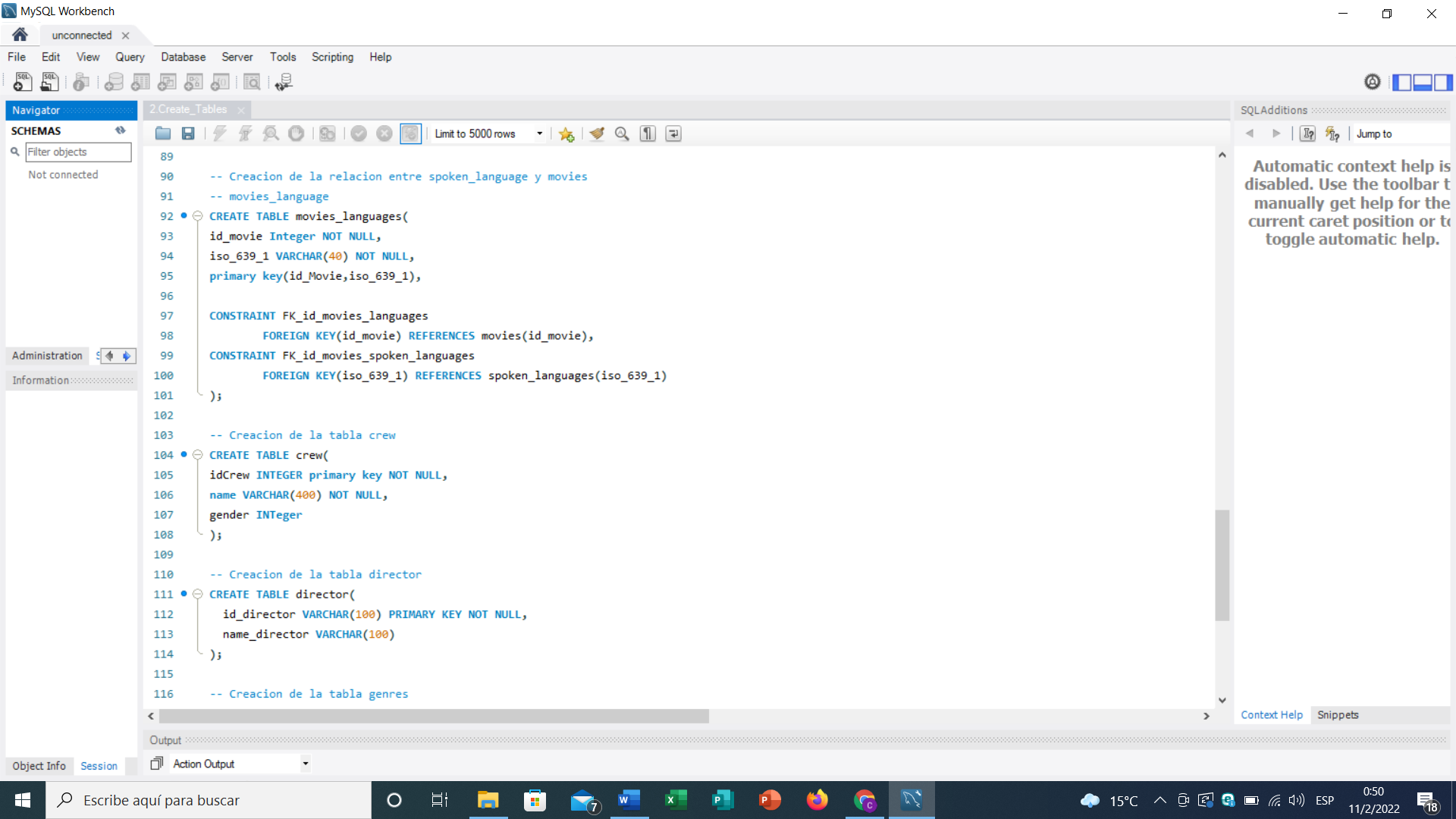
Se creo una tabal movie dataset cleaned donde hemos limpiado la tabla movie dataset para poder manejarla de una mejor manera



**3.2 Creación de todas las tablas**

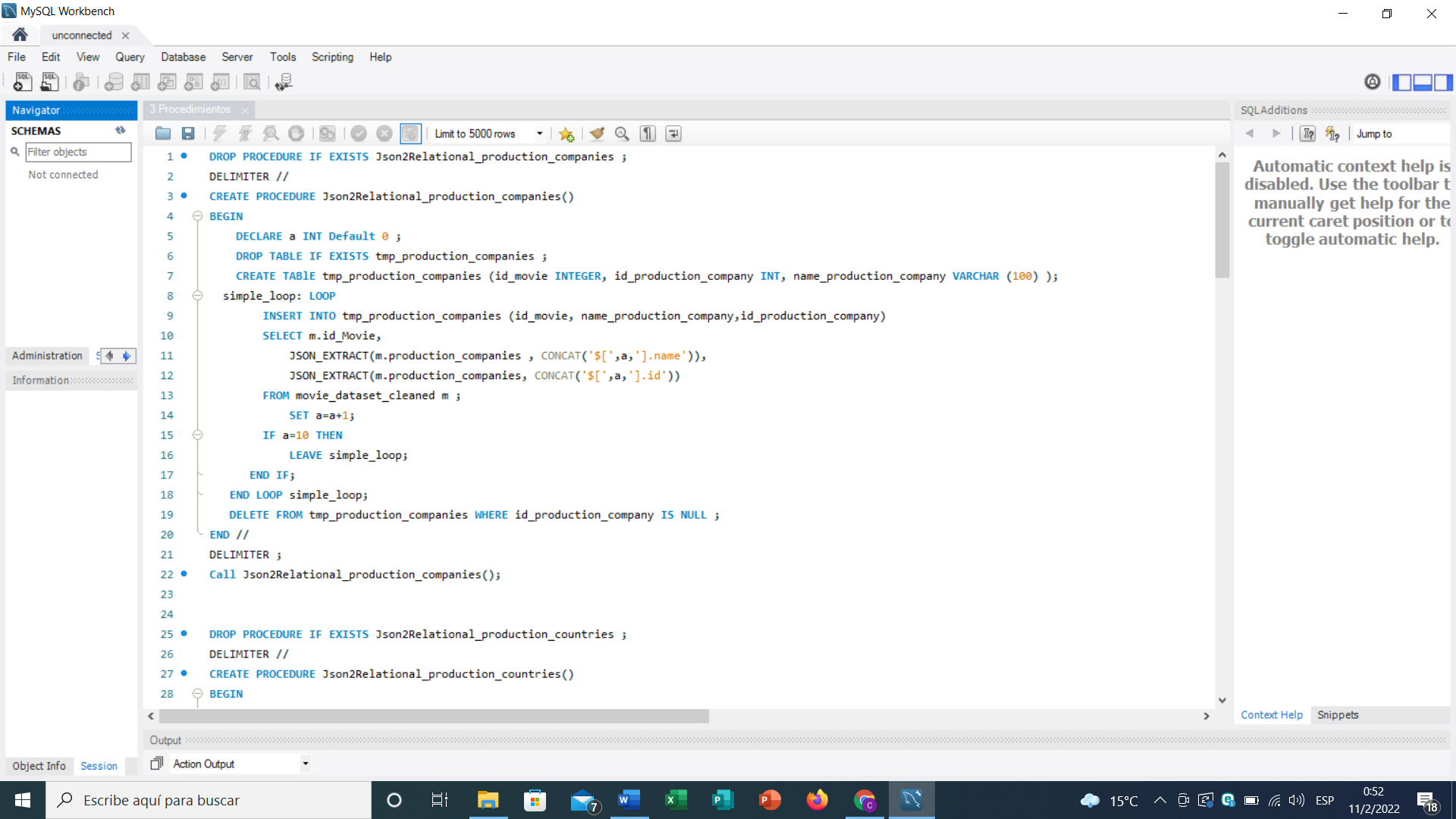
Se crearon todas las tablas resultantes de la normalización aplicada y se las paso a script de mysql workbench

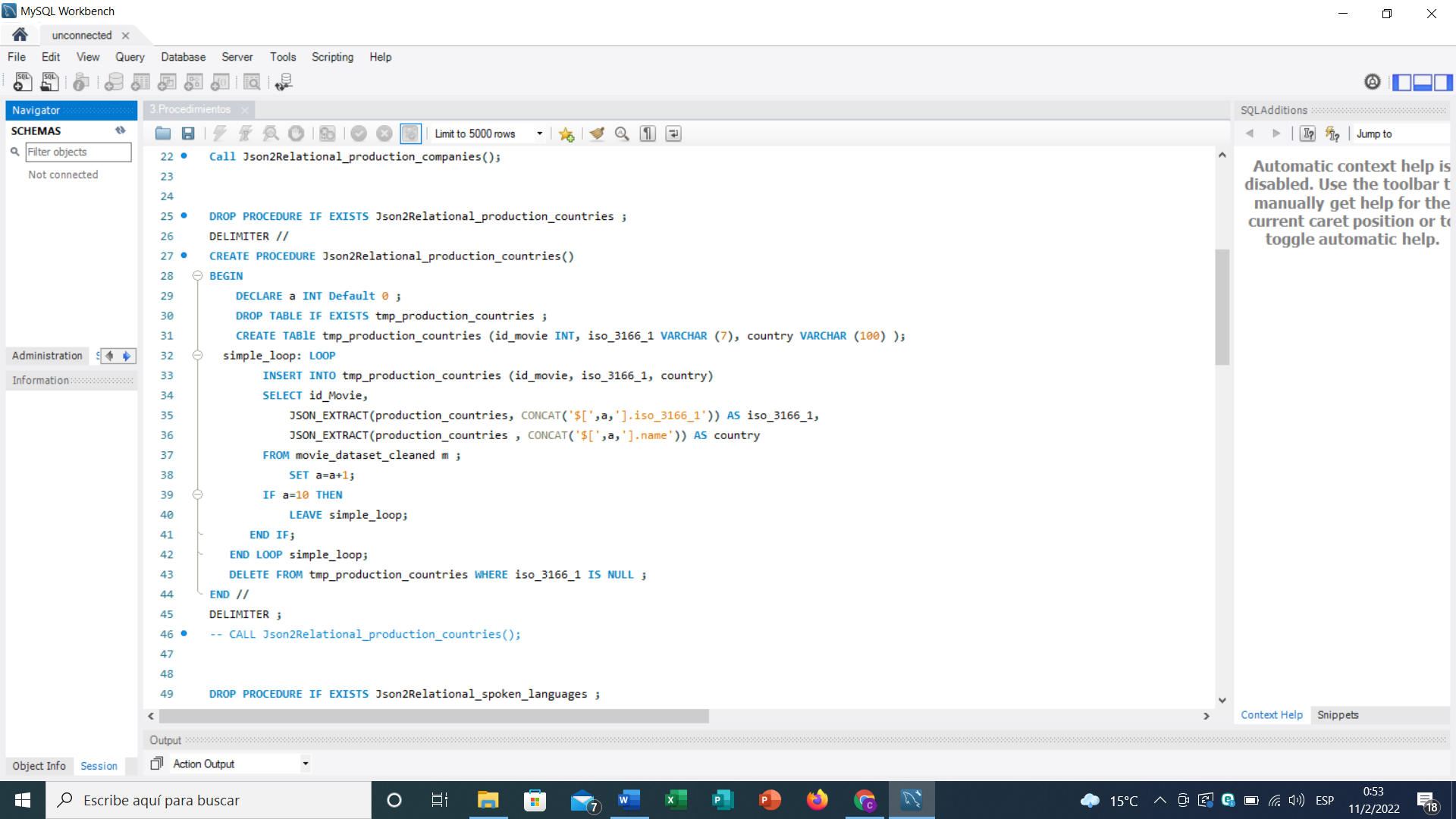
 

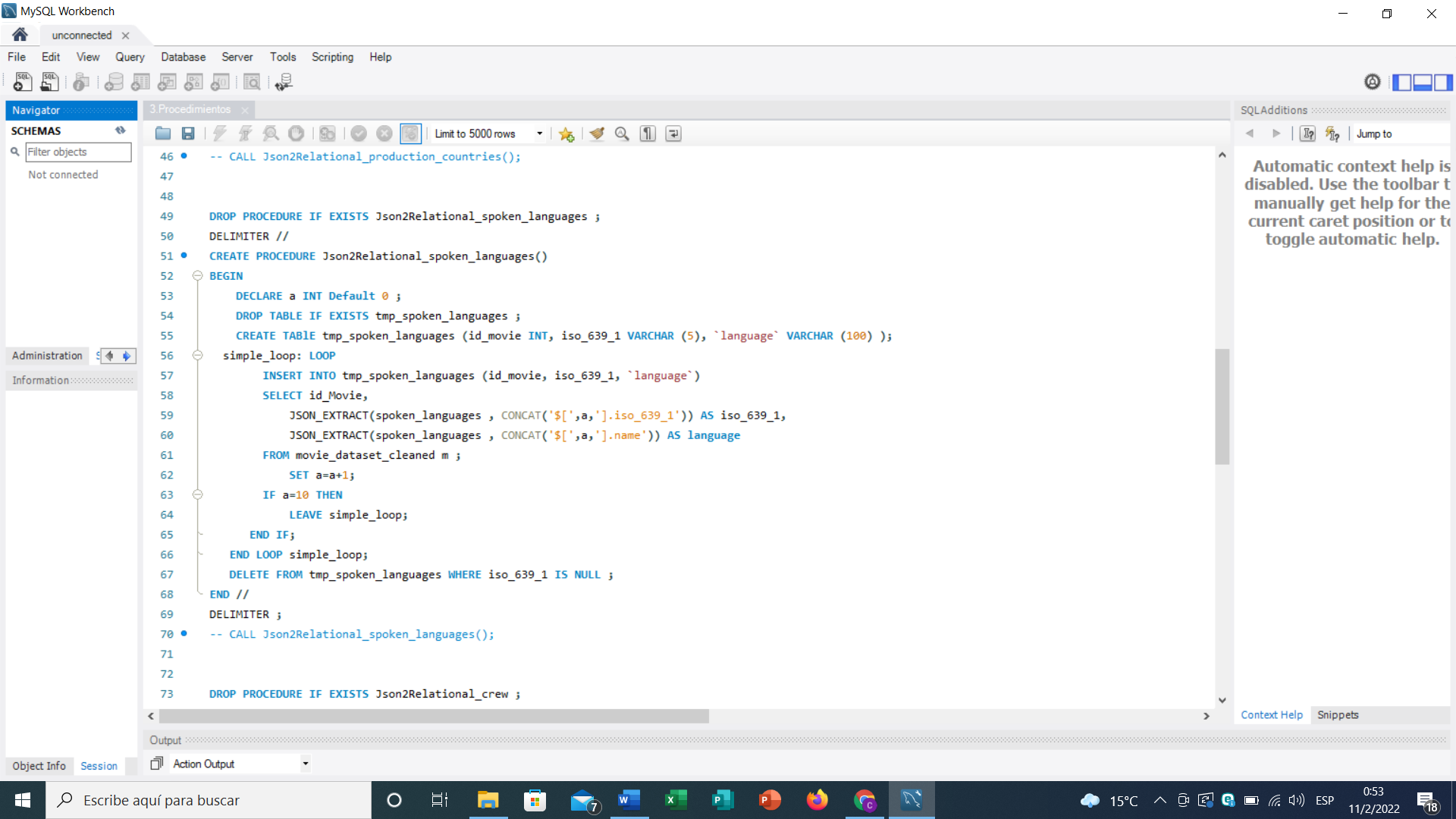
 

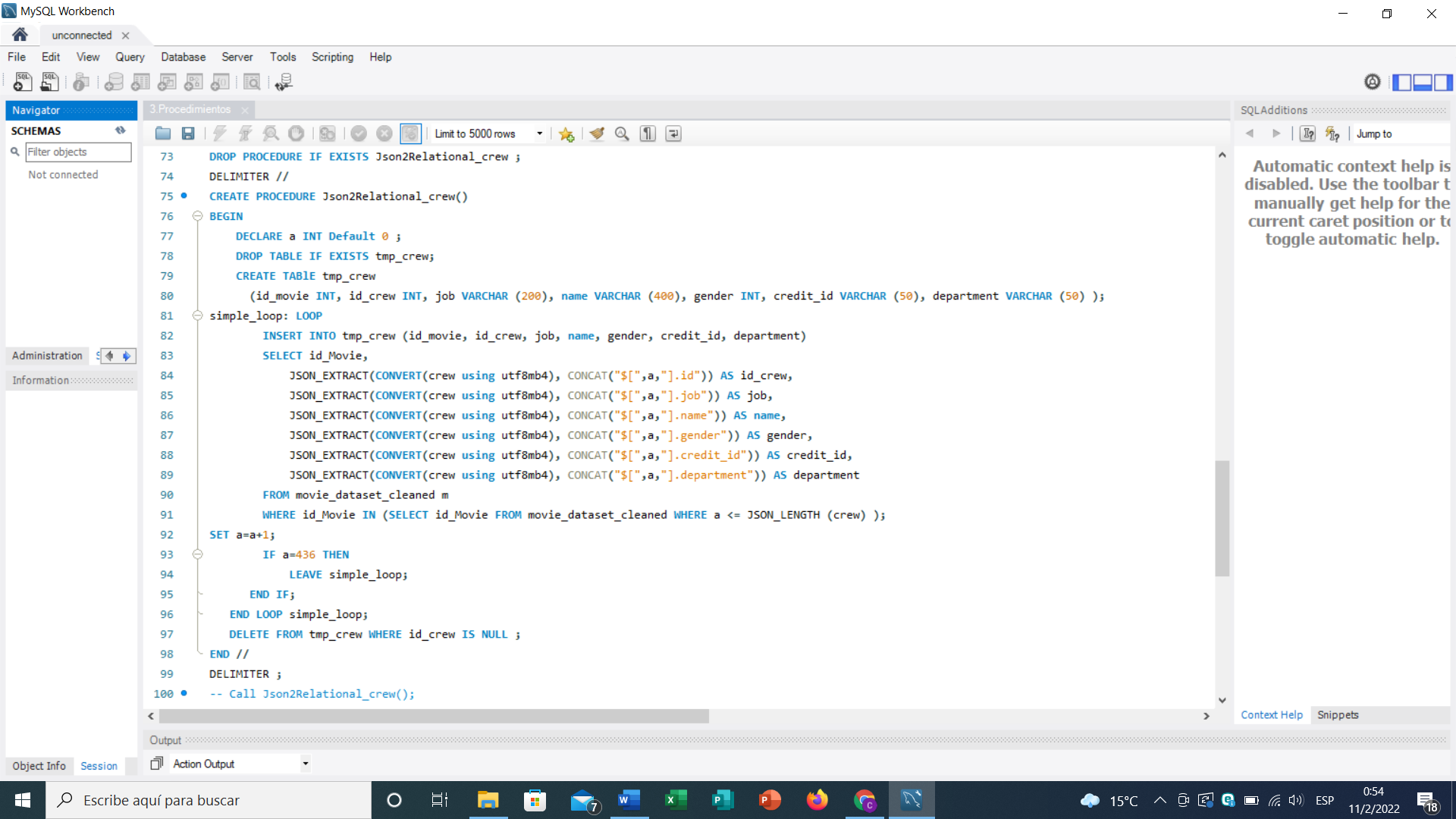
**3.3 Creación de los Metodos**

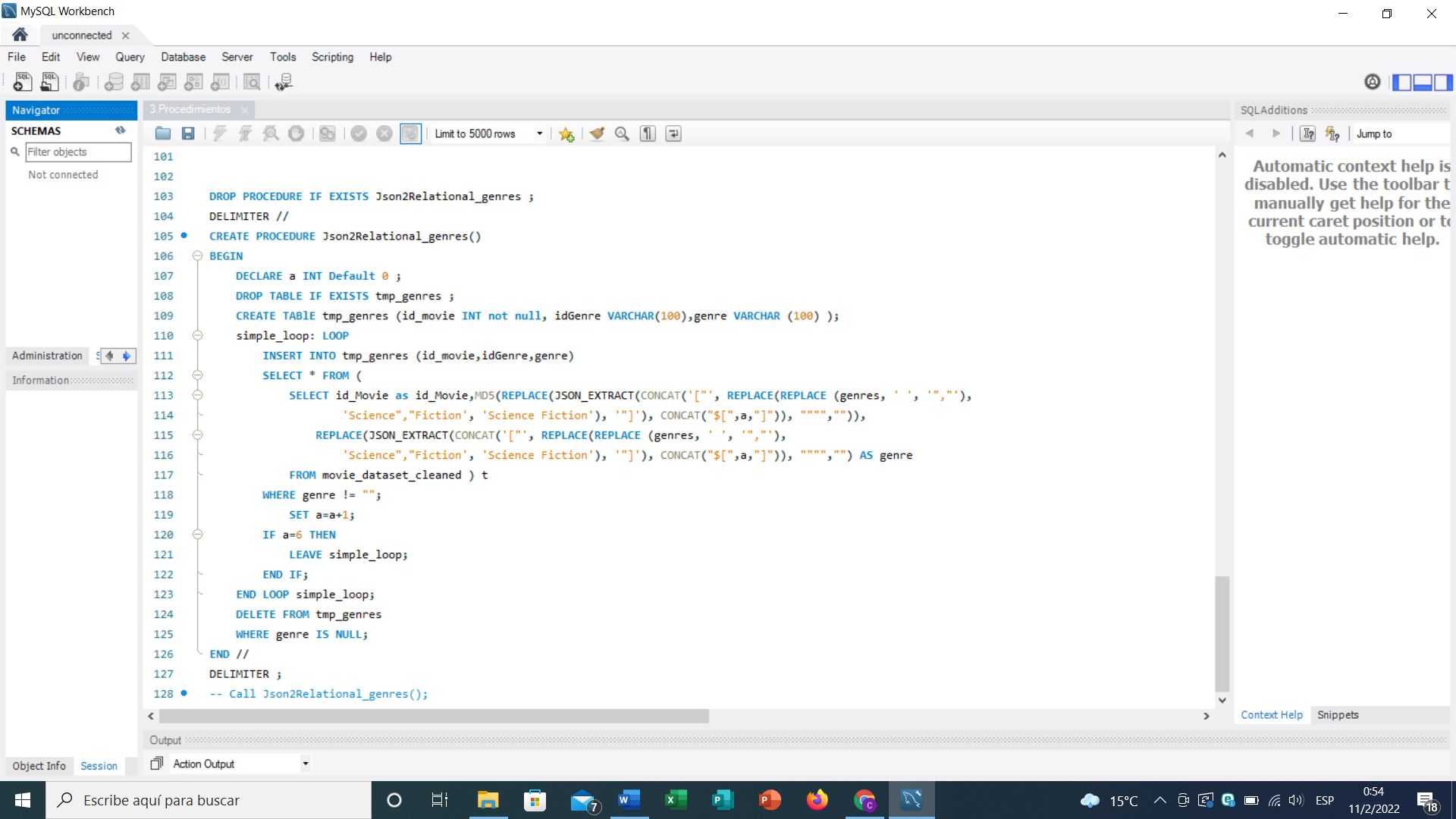
A continuación se muestra la creación de los métodos los cuales nos fueron de mucha ayuda para trabajr con tablas como crew las cuales tenían los datos en tipo json





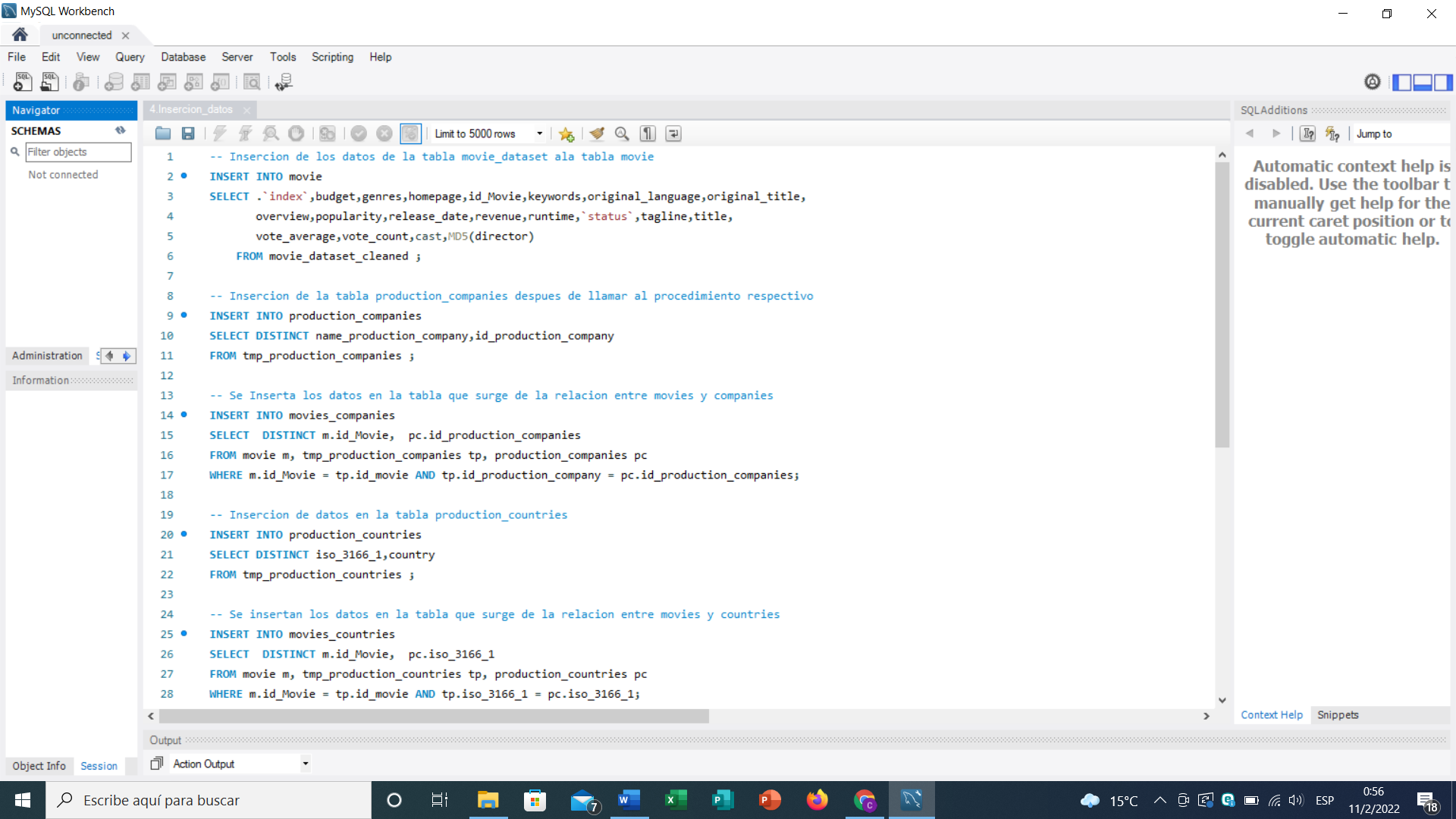


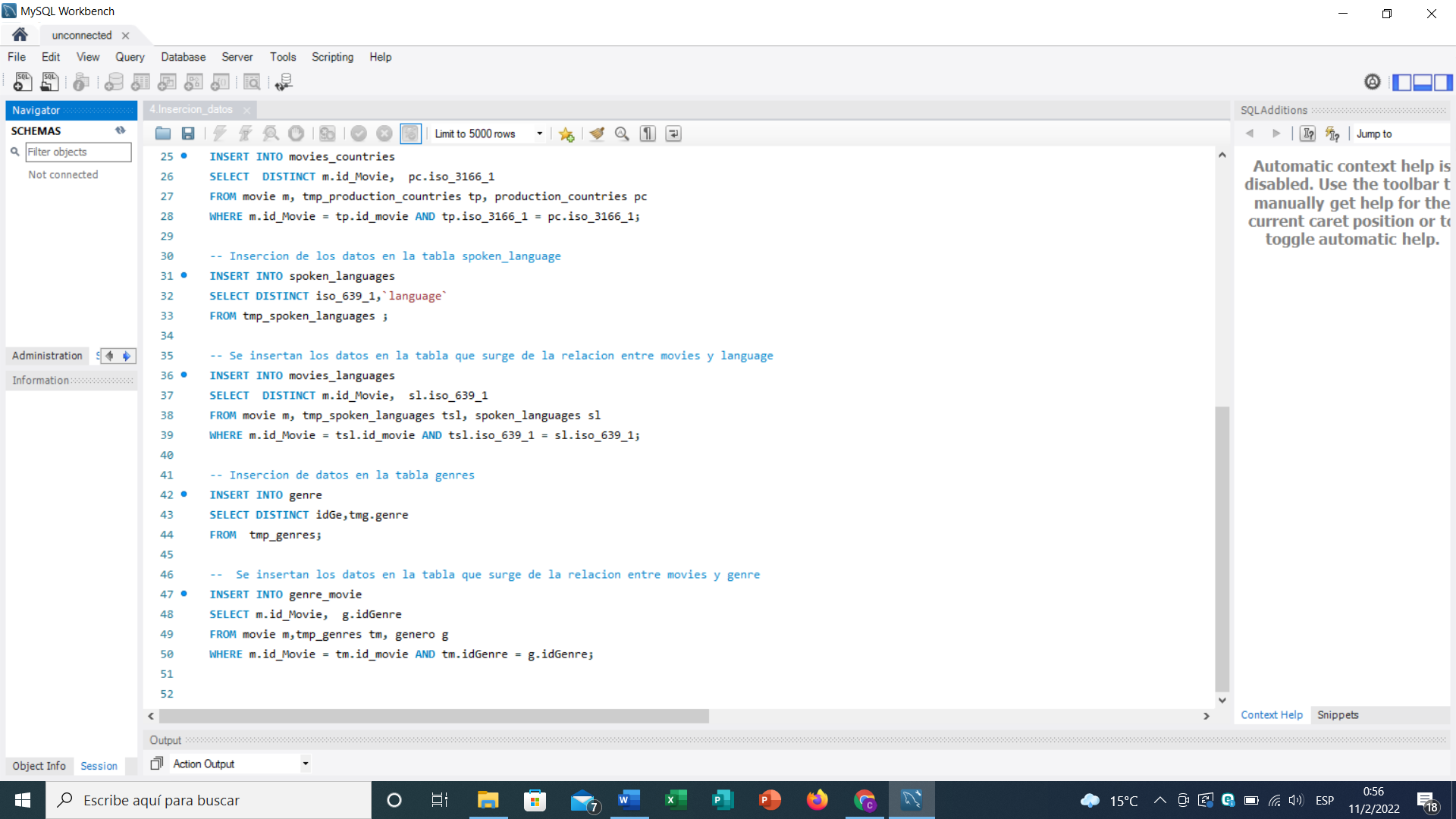




**3.4 Creación de los Metodos**

Por ultimo se ha hecho la inserción de los datos en las respectivas tablas que hemos creado para hacer esta inserción hemos utilizado el código que se muestra a continuación





1. **Conclusiones**

* Hemos podido concluir que después de un análisis completo del proyecto se han conseguido los resultados deseados y que el resultado final satisface todos los objetivos que nos hemos propuesto a lo largo del bimestre
* Por ultimo cabe afirmar que en este proyecto se a puesto e evidencia todo lo aprendido por lo cual se ha registrado nuestro aprendizaje a través de el resultado obtenido a través del proyecto