

Instituto Tecnológico y de Estudios Superiores de Monterrey

Inteligencia Artificial Avanzada para la Ciencia de Datos

Módulo 1. Big Data.

Samuel Méndez Villegas – A01652277

23 de noviembre del 2022

Data Warehouse

La base de datos que se utilizó para esta actividad fue la llamada “PKDD’99 Financial”, la cual contiene información acerca de 606 préstamos exitosos, y 76 no exitosos junto con toda la información y transacciones relacionadas. La base de datos contiene 8 distintas tablas y puede ser accedida desde: <https://relational.fit.cvut.cz/dataset/Financial>

Creación de Data Warehouse

Para guardar la base de datos se utilizó el servicio local “mySQL” en donde se cargó la base de datos. Esto se logró conectándose al servidor especificado en la liga anterior.

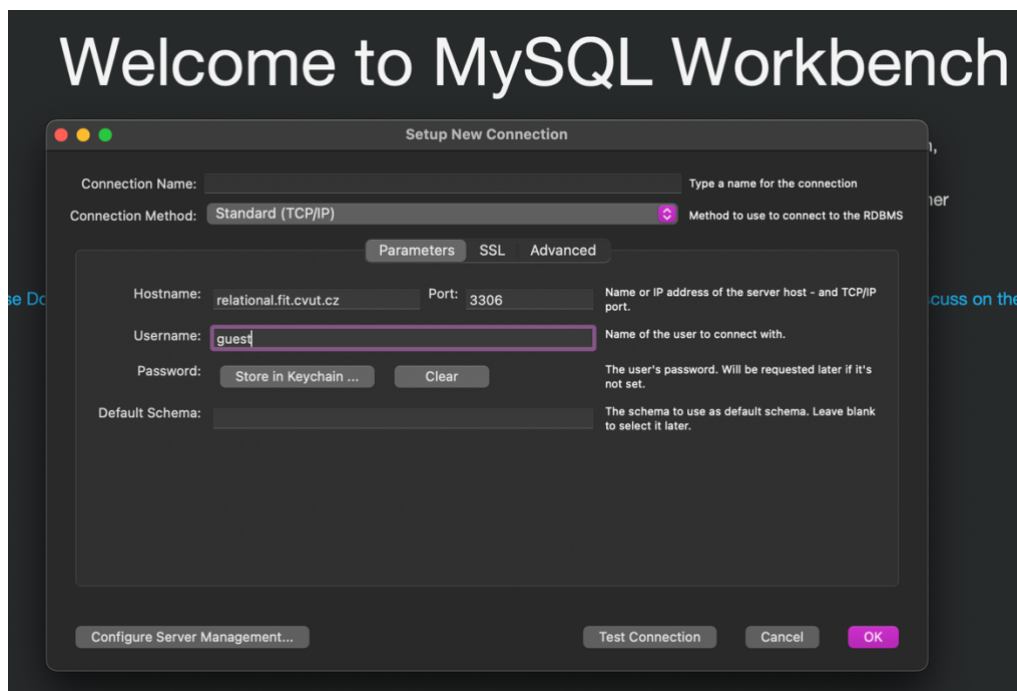


Figura 1. Conexión al servidor en donde se encuentra la base de datos a utilizar

Una vez hecho lo anterior, se comprobó la correcta conexión visualizando una de las 8 tablas distintas de la base de datos.

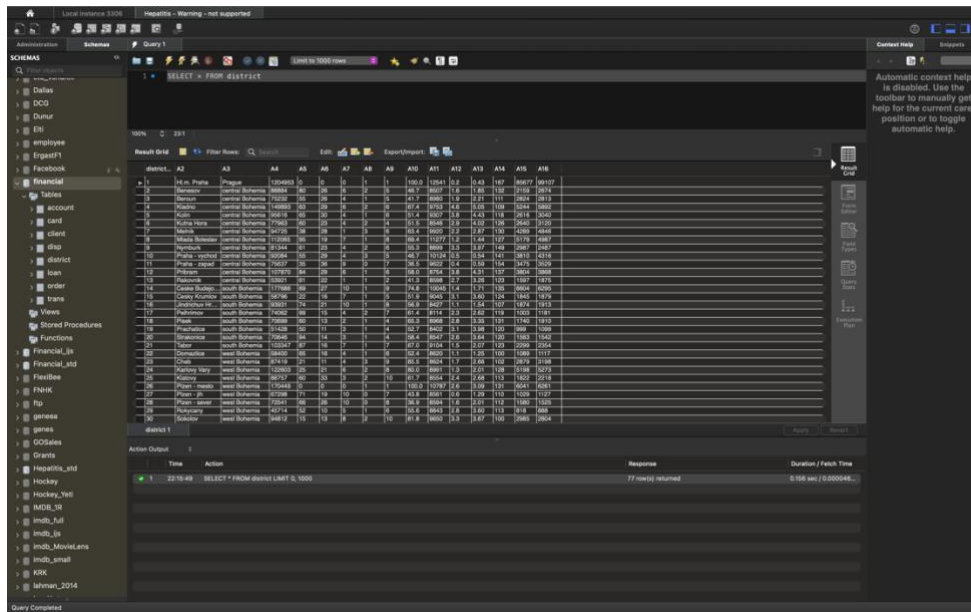


Figura 2. Correcta conexión a la base de datos.

Esquema de relación entre tablas

MySQL ofrece como herramienta el poder visualizar el diagrama entidad relación de cualquier base de datos. Para este caso, el diagrama es el siguiente:

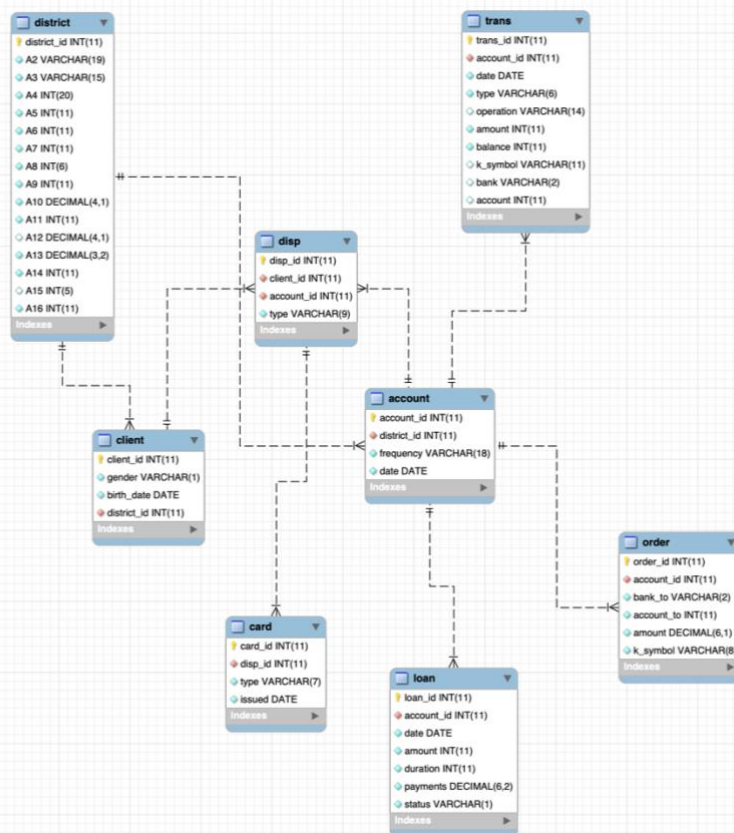


Figura 3. Diagrama entidad relación de la base de datos

Uso de Tableau para visualizar información relevante

Posteriormente, se abrió Tableau Desktop y se le dio clic a conectar a un servidor, en donde se seleccionó MySQL. De igual forma, se pidieron las credenciales respectivas para conectarse a la base de datos.

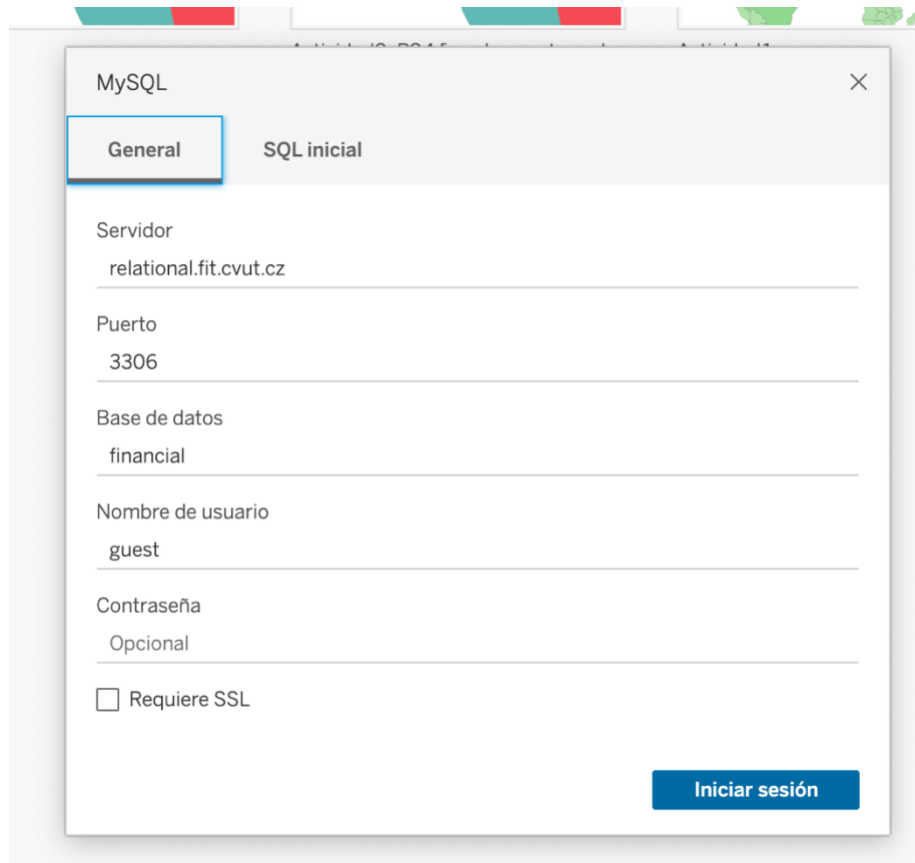


Figura 4. Conexión de Tableau Desktop con el servidor en MySQL.

Una vez establecida la conexión, se realizó el siguiente dashboard con información relevante:

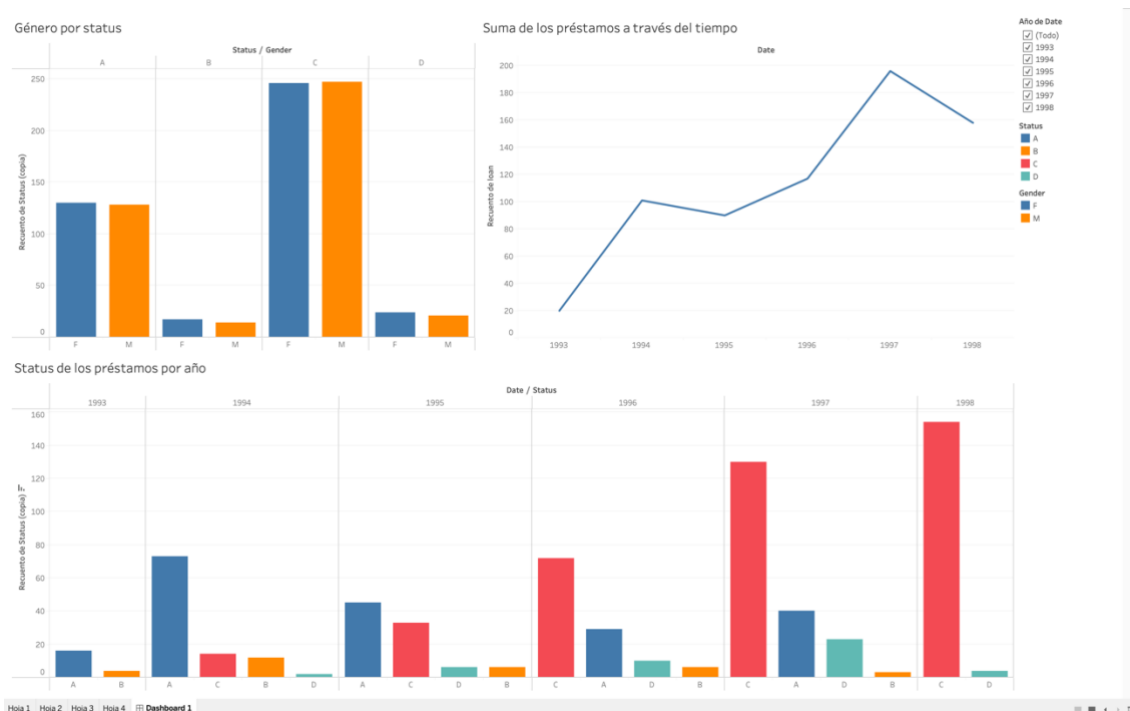


Figura 5. Dashboard generado en Tableau

Una vez generado el dashboard, se publicó el tablero en Tableau Cloud tal y como se hizo en la actividad 2. Al final de este documento se anexa la liga de donde se puede acceder.

Relevancia de la información mostrada

La información que se muestra en el dashboard es relevante para analizar el status de los préstamos que se han concedido a través del tiempo. Según el gráfico de barras, se puede visualizar que el préstamo C es el que más presencia tienen en la mayoría de los años, seguido de A.

Esta información puede de igual forma ser de bastante relevancia a la hora de tomar decisiones, ya que se tiene un panorama más amplio del comportamiento de los datos.

De igual forma, el analizar el perfil de las personas que piden los préstamos es importante, ya que, dependiendo de este, se puede establecer una relación entre por qué a cierto grupo de personas se les está otorgando o no un crédito. En este caso, por ejemplo, se observa claramente que el género no es un diferenciador.

Liga a Tableau Cloud: https://prod-useast-a.online.tableau.com/#/site/datawarehousesamuelmendez/workbooks/763242?origin=card_share_link

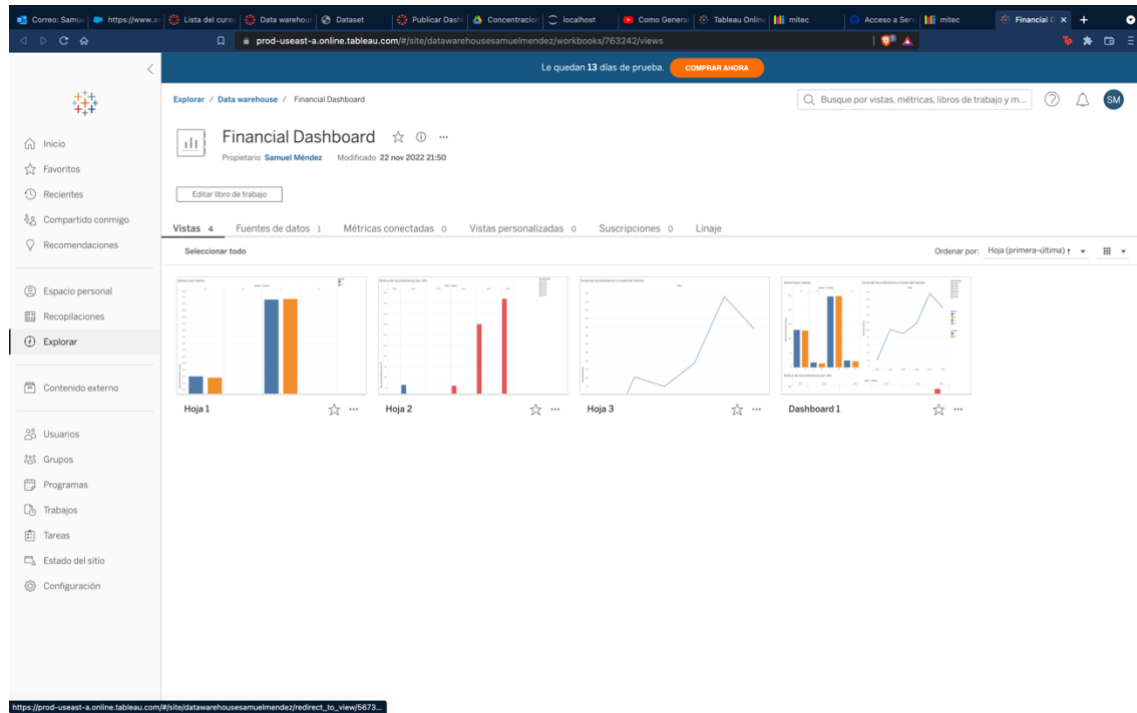


Figura 6. Publicación del dashboard en Tableau Cloud