

CIS 467 final group project (due by Wednesday, March 5 at 11:59 PM).

This is a group project (total 300 points). The groups have been created on Blackboard. Please make only one submission per group and put all your team members' full names into this Word document and submit this Word document. Please also submit a Tableau Workbook file .twb into the Final Project folder on Blackboard together with the Excel file of your Data Warehouse which you uploaded to Tableau and used for visualizations.

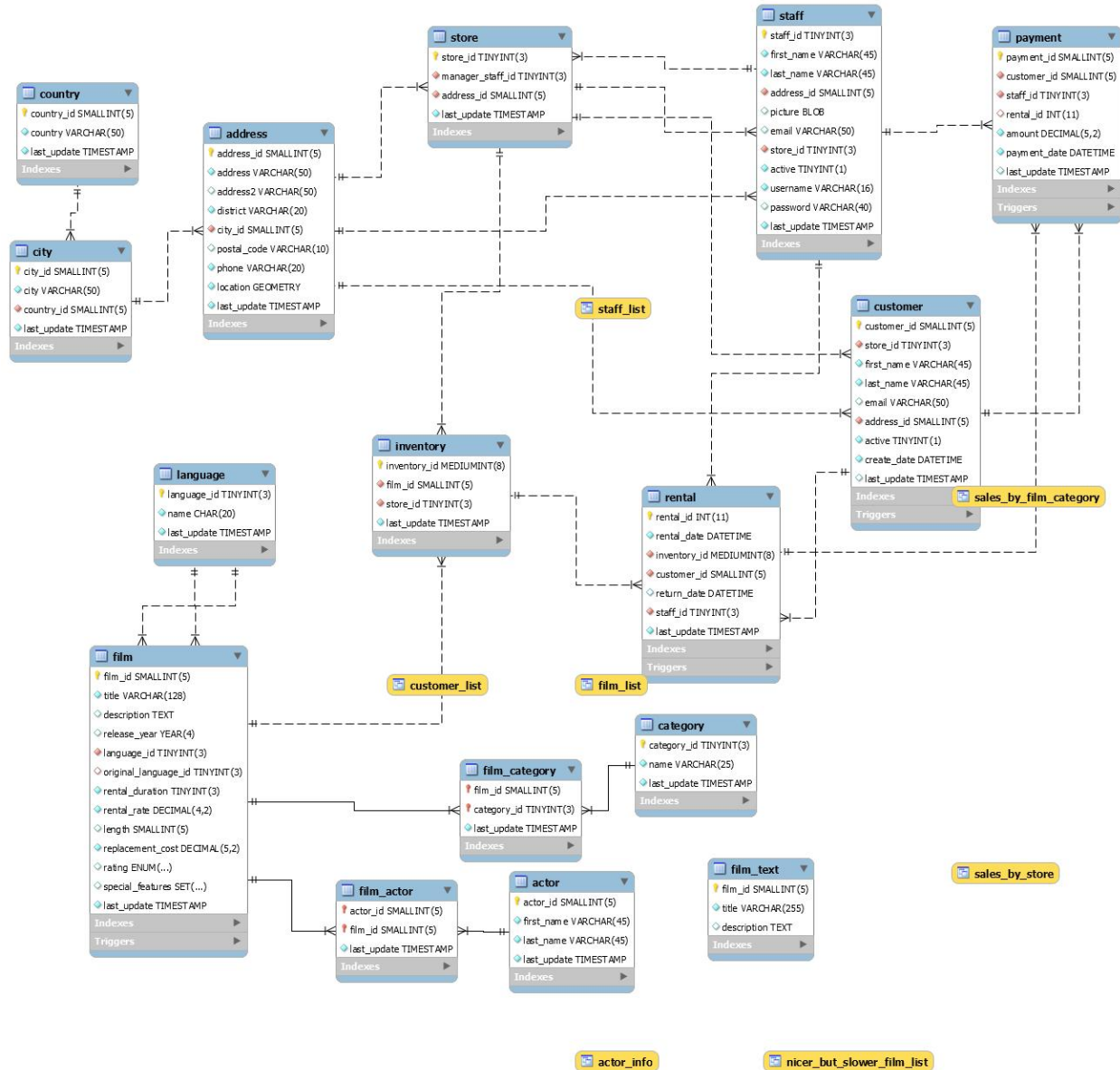
The script files sakila-schema.sql and sakila-data.sql create a database which contains tables (the database schema is below), with transactional data related to some company operations.

Please check early that you can create the sakila database on your machine. First, run this code - sakila-schema.sql, and second, run this code - sakila-data.sql.

Very Important! All three parts of the final project should be on One topic/subject of the data warehouse. For example, if you decide to track customers as your topic/subject, part 1 (Data Warehouse), part 2 (Queries) and part 3 (visualizations) should only be related to customers and should NOT include any other topics.

If you use Chat GPT, please use the “Share” button (looks like ‘upward arrow’) in the right corner of ChatGPT chat, and ‘copy link’ and share the link to that chat in this Word document and briefly explain how you used it for your Final Project (for each question if you used it). No points will be taken off for using ChatGPT (it is allowed to use it for Final Project) but you are required to share the link to a chat if you used it.

More information on how to share a chat here: <https://help.openai.com/en/articles/7925741-chatgpt-shared-links-faq>



Please put all your work into **this single Word doc** and also submit a **Tableau Workbook file .twb** and the **Excel file of your Data Warehouse** that you used for **Tableau visualizations**. Please see instructions for Tableau below in question 3.

1. **(121 points)** Design and create a data warehouse for the provided database. The decisions about which fields to include and how to aggregate the data are left to you. You do not need to include every single data point from the tables given. Use your judgement as to what will be interesting/useful for the organization. But please make sure that you pull (combine) data from **at**

least six tables and compute relevant aggregate statistics. Please compute relevant aggregate statistics for each table that you join. **In your queries later in part 2, you may join your Data Warehouse with other tables to answer useful questions.** Please see many examples from class lectures and you may adapt those codes for your purpose (for this dataset).

Submit a screenshot of the first 25 rows of your data warehouse (paste into this Word document) and the SQL code that you used to create it. Please copy and paste your SQL code into this Word document. If your PC does not show 25 rows of data, please submit what you have (i.e., rows you can see on a screenshot) with a comment that you cannot show 25 rows of data. Please add a full description of what your Data Warehouse will be tracking for a company. Please treat this assignment as a business case. So, the more you describe the better. Please also create an Excel file (Export from MySQL) of your data warehouse and use it for part 3 – Tableau visualizations.

2. (104 points) Create eight SQL queries on your data warehouse (not on the original dataset) that answer interesting questions. At least 6 queries should be more complex queries. For example, more complex queries could include Joins, a Group By, UNION elements or a subquery or use some aggregate functions and summary calculations and conditional logic codes (see examples in the class lectures' slides). **If needed, you may join your Data Warehouse with other tables (which are not a part of Data Warehouse) to answer useful questions.**

Submit a copy of each query SQL code (paste into this Word document), and the screenshot of each query results (or the first 25 rows if it is longer or how many rows you can get on your PC) and full description of the question your SQL code was addressing and what you found in the results. The question that each query answers should be useful for a company to make decisions and act upon.

3. (75 points) Create five Tableau individual visualizations (graphs) on your data warehouse (plus one dashboard as discussed below) with valuable information to present findings to senior management of the company. Save each visualization as a png file (as I show in class, and we will also practice in the lab 5) and paste each individual visualization png file into this Word document with the full explanation of what the visualizations show, how they are useful to a company and how company management could make decisions based on what you show. Finally, combine those five visualizations into one Dashboard (as I show in class, and we will also practice in the lab 5), and save this Dashboard as a png file and paste the Dashboard into this Word document.

Important! You might visualize some queries from part 2 or create completely new visualizations. If for queries (in part 2), you used data warehouse together with (joining) some additional tables, you need to create a new source Excel file for Tableau, and submit it as described below in the next paragraph (please explain in your submission what source you used). Alternatively, you may do visualizations only on Data Warehouse Excel file (from part 1) and submit that file as a source file for Tableau (see below). In general, we need one Excel source file for Tableau to check your visualizations and the Tableau workbook (see below).

Please also save the whole Tableau project as a Tableau Workbook file .twb (In Tableau use File - Save as) and submit to the Final Project folder on Blackboard together with this Word document and together with the Excel file of your Data Warehouse which you uploaded to Tableau and used for visualizations.

General grading criteria: Your completed work will be evaluated using the criteria below. I encourage you to use your creativity and other business skills (communication, presentation, critical thinking) in addition to the data management concepts and the SQL and Tableau skills that we have covered in CIS467.

High score	Score between high and good	Good/medium score	Low score
All required parts of the final project are complete and technically correct. Queries are useful/interesting and provide valuable information for senior management to act upon. Not just random queries. Tableau visualizations provide interesting useful information based on which senior management of the company can make important decisions.	All required parts of the final project are complete and technically correct (with possibly a few minor errors). Queries are useful/interesting and provide valuable information for senior management to act upon. Not just random queries (with possibly a few minor errors). Tableau visualizations provide interesting useful information based on which senior management of the company can make important decisions (with possibly a few minor errors).	Some required parts of the final project are missing and/or there are more significant errors. Some queries appear random and do not answer any useful/interesting questions. Tableau visualizations are very simple but may still provide interesting useful information based on which senior management of the company can make important decisions.	The final project has large portions missing and/or major conceptual errors. Most/all queries (if any) appear random and do not answer any useful/interesting questions. Tableau visualizations are very simple and do not provide interesting useful information based on which senior management of the company can make important decisions.