**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>

A computer screen shot of a diagram

AI-generated content may be incorrect.

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.**

**Answer 1:**

**Code:**

USE sakila;

DROP TABLE IF EXISTS film\_summary;

CREATE TABLE film\_summary AS

SELECT

YEAR(r.rental\_date) AS rental\_year,

MONTH(r.rental\_date) AS rental\_month,

f.film\_id,

f.title,

fc.category\_id,

c.name AS category\_name,

COUNT(r.rental\_id) AS total\_rentals,

ROUND(SUM(p.amount), 2) AS total\_revenue,

AVG(DATEDIFF(r.return\_date, r.rental\_date)) AS avg\_rental\_duration,

f.replacement\_cost,

a.actor\_count,

cn.country AS rental\_country,

ct.city AS rental\_city

FROM film f

JOIN film\_category fc ON f.film\_id = fc.film\_id

JOIN category c ON fc.category\_id = c.category\_id

JOIN inventory i ON f.film\_id = i.film\_id

JOIN rental r ON i.inventory\_id = r.inventory\_id

LEFT JOIN payment p ON r.rental\_id = p.rental\_id

LEFT JOIN (

SELECT fa.film\_id, COUNT(fa.actor\_id) AS actor\_count

FROM film\_actor fa

GROUP BY fa.film\_id

) a ON f.film\_id = a.film\_id

JOIN customer cu ON r.customer\_id = cu.customer\_id

JOIN address ad ON cu.address\_id = ad.address\_id

JOIN city ct ON ad.city\_id = ct.city\_id

JOIN country cn ON ct.country\_id = cn.country\_id

GROUP BY rental\_year, rental\_month, f.film\_id, f.title, fc.category\_id, c.name,

f.replacement\_cost, a.actor\_count, cn.country, ct.city;

**Screenshot:**

**A table with numbers and numbers

Description automatically generated**

**Description:**

The Film Data Warehouse tracks film performance, rental trends, and revenue insights to help Sakila optimize movie selection, pricing, and inventory management. It consolidates key metrics like rental counts, revenue, average rental duration, and regional preferences to support data-driven decision-making.

**Key Insights Tracked**

* Film Popularity – Identifies the most and least rented films by time and location.
* Revenue Analysis – Tracks top-grossing films and best-performing categories.
* Rental Duration – Measures how long customers keep each film.
* Actor Influence – Identifies actors starring in high-rented films.
* Regional Trends – Finds which film categories are popular in different countries.

**Business Benefits**

1. Better Film Selection – Stock high-demand films, remove underperformers.
2. Optimized Pricing – Adjust rental rates for high-grossing films.
3. Improved Customer Experience – Personalize film recommendations by region.

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.**

**Answer 2:**

**Query 1: Top 5 Most Profitable Categories**

**Question:** What are the Top 5 film categories that generate the highest total revenue for the company?

**Code:**

SELECT category\_name, SUM(total\_revenue) AS total\_earnings

FROM film\_summary

GROUP BY category\_name

ORDER BY total\_earnings DESC

LIMIT 5;

**Screenshot of output:**

**A white and black text box

Description automatically generated with medium confidence**

**Description:** Understanding which film categories contribute the most revenue allows the company to optimize its film inventory, marketing strategies, and future acquisitions. If certain genres consistently generate higher revenue, the company can prioritize investments in these categories, negotiate better licensing deals, and design targeted marketing campaigns to further boost revenue. Additionally, analyzing the most profitable categories provides insight into customer preferences, which can guide subscription-based recommendations or promotional bundling strategies**.**

**Findings:**

* The Sports category leads in total earnings, suggesting high demand and potentially high rental frequency. This may indicate that customers have a strong preference for sports-related content, making it a key focus area for future investments.
* Sci-Fi and Animation also perform well, reinforcing the idea that fantasy and visually engaging content attract a significant audience.
* Drama and Comedy remain strong performers, highlighting the broad appeal of emotionally driven and entertaining content.
* Interestingly, traditional blockbuster categories like Action or Horror do not appear in the top 5, suggesting a unique customer preference in this dataset.

**Query 2: Top 10 films that generated the highest total revenue**

**Question:** Which are the Top 10 films that generated the highest total revenue for the company?

**Code:**

SELECT title, category\_name, SUM(total\_revenue) AS Total\_Revenue

FROM film\_summary

GROUP BY title, category\_name

ORDER BY Total\_Revenue DESC

LIMIT 10;

**Screenshot:A white background with black lines

Description automatically generated with medium confidence**

**Description:**

Identifying the top 10 highest revenue-generating films helps the company prioritize its most valuable assets. These films are proven revenue drivers, meaning they are critical for customer retention, targeted marketing, and future licensing negotiations. By analyzing these top performers, the company can also look for patterns -such as common genres, popular cast members, or release timing-to guide future content acquisition and pricing strategies. Additionally, these films can be featured in premium collections, or used as anchor titles in bundled offers to help drive demand for lesser-known films.

**Findings:**

* The Top 10 includes films from diverse genres, including Music, Documentary, Comedy, Sci-Fi, Sports, Drama, and Foreign, confirming that high revenue is not limited to a single type of content.
* The Music film "TELEGRAPH VOYAGE" leads the list, showing that music-themed content has strong revenue potential, possibly driven by niche fan loyalty.
* Comedy films appear twice in the list, confirming that lighthearted content appeals to a broad audience and has strong repeat rental potential.
* Sci-Fi also appears twice, indicating that sci-fi content might justify higher licensing investment given its strong revenue impact.
* The presence of a Foreign film (INNOCENT USUAL) demonstrates the importance of maintaining culturally diverse content to serve international audiences or niche cultural interests.
* The overlap between this list and earlier queries (such as most-rented films) suggests that films with both high rental volume and higher price points tend to dominate revenue — showing that premium pricing for popular titles works well.

**Query 3: Actors that have starred in the most high-revenue films**

**Question:** Which actors have appeared in the most high-revenue films, and how much total revenue have their films generated?

**Code:**

SELECT fa.actor\_id, a.first\_name, a.last\_name, COUNT(fa.film\_id) AS num\_films, SUM(fs.total\_revenue) AS total\_earnings

FROM film\_actor fa

JOIN actor a ON fa.actor\_id = a.actor\_id

JOIN film\_summary fs ON fa.film\_id = fs.film\_id

GROUP BY fa.actor\_id, a.first\_name, a.last\_name

ORDER BY total\_earnings DESC

LIMIT 10;

**Screenshot:A screenshot of a computer

Description automatically generated**

**Description:**

Understanding which actors consistently appear in top-performing films helps the company refine its content acquisition and marketing strategies. If certain actors are proven revenue drivers, the company can prioritize acquiring films featuring these actors and leverage their popularity in promotional campaigns. Highlighting films with popular actors in featured sections or curated collections can also attract customers and boost rentals, directly impacting revenue. This is particularly relevant in the entertainment industry, where star power strongly influences consumer choices.

**Findings:**

* Gina Degeneres leads the list, with her films generating over $3,400 in total revenue, showing her consistent presence in popular titles.
* The list features a mix of male and female actors, indicating that both genders contribute equally to revenue-driving content.
* The number of films each actor appeared in is relatively high, meaning these actors have broad appeal across multiple genres and film types.
* Actors like Matthew Carrey and Meryl Keitel are also strong performers, meaning films with these actors should be prioritized in future licensing decisions.

**Query 4: Top cities that rent the most films**

**Question:** Which cities rent the highest number of films, and how do they compare in terms of rental activity?

**Code:**

SELECT rental\_city, SUM(total\_rentals) AS total\_rentals

FROM film\_summary

GROUP BY rental\_city

ORDER BY total\_rentals DESC

LIMIT 10;

**Screenshot:A white and grey striped background

Description automatically generated with medium confidence**

**Description:**

Understanding which cities generate the highest rental volumes helps the company prioritize marketing efforts, localized promotions, and inventory distribution. Cities with high rental activity could be targeted with special discounts, loyalty programs, or curated film collections tailored to local tastes. This insight also helps identify high-potential markets where the company could expand its presence or invest in more localized content.

**Findings:**

* The top-ranking cities show strong demand for film rentals, suggesting these locations could be targeted for tailored marketing campaigns.
* Interestingly, the list includes a mix of large, globally known cities (like London) and smaller, regional cities, showing that demand exists across a diverse geographic footprint.
* Cities such as Aurora and London show particularly strong activity, meaning these markets could benefit from exclusive promotions like early access to new films or local loyalty rewards.
* Regional cultural preferences might also influence rental trends — for example, localized content (such as films set in these cities or dubbed in local languages) might perform especially well.

**Query 5: Most rented film in each category**

**Question:** Which film holds the highest number of rentals within each film category?

**Code:**

SELECT fs1.category\_name, fs1.title, fs1.total\_rentals

FROM (

SELECT category\_name, title, SUM(total\_rentals) AS total\_rentals

FROM film\_summary

GROUP BY category\_name, title

) fs1

JOIN (

SELECT category\_name, MAX(total\_rentals) AS max\_rentals

FROM (

SELECT category\_name, title, SUM(total\_rentals) AS total\_rentals

FROM film\_summary

GROUP BY category\_name, title

) fs2

GROUP BY category\_name

) fs2

ON fs1.category\_name = fs2.category\_name AND fs1.total\_rentals = fs2.max\_rentals

ORDER BY fs1.category\_name ASC;

**Screenshot:A white and grey striped background

Description automatically generated with medium confidence**

**Description:**

Knowing which film has the most rentals within each category allows the company to understand which content truly captures customer attention across genres. This is crucial for shaping content acquisition, marketing promotions, and customer recommendations. It also helps to identify films that could be featured in highlighted sections (such as “Most Popular Films by Genre”) or used for bundling strategies to boost rentals for lower-performing films. Additionally, by identifying these top performers, management can analyze their common attributes (cast, story, release year) to guide future licensing decisions.

**Findings:**

* No single film dominates across all genres, which is a positive sign that the company’s catalog offers diverse and genre-specific hits rather than being overly reliant on a few blockbusters.
* Family films stand out, with three different films sharing the top spot, signaling that family-friendly content has broad appeal and consistent engagement.
* Some categories (like Action and Games) have ties, with two films sharing the same rental count, which could indicate fierce competition between similarly popular titles.
* ZORRO ARK, a Comedy film, appears here as well — reinforcing its earlier presence in the Top Revenue Films query. This confirms it’s not just high-priced, but genuinely popular.
* The diversity of titles and genres represented in this list highlights the importance of maintaining a broad content library, as no single genre or type of film consistently dominates across the board.

**Query 6: Highest-grossing film in each category**

**Question:** Which film generates the highest total revenue in each film category?

**Code:**

SELECT fs1.category\_name, fs1.title, fs1.total\_revenue

FROM (

-- Total revenue per film within each category

SELECT category\_name, title, SUM(total\_revenue) AS total\_revenue

FROM film\_summary

GROUP BY category\_name, title

) fs1

JOIN (

-- Maximum revenue per category

SELECT category\_name, MAX(total\_revenue) AS max\_revenue

FROM (

SELECT category\_name, title, SUM(total\_revenue) AS total\_revenue

FROM film\_summary

GROUP BY category\_name, title

) fs2

GROUP BY category\_name

) fs2

ON fs1.category\_name = fs2.category\_name AND fs1.total\_revenue = fs2.max\_revenue

ORDER BY fs1.category\_name ASC;

**Screenshot:A white and grey striped background

Description automatically generated with medium confidence**

**Description:**

This query helps the company identify the top revenue contributor in every genre, giving a clear view of which specific films drive the most financial success in each segment. This is crucial for both content curation and future licensing strategies. Understanding these top performers also helps with targeted promotions — for example, featuring these films in “Best of Genre” collections. Additionally, knowing which films dominate within each category helps evaluate whether the company is investing in the right types of content within each genre.

**Findings:**

* Several films appear as top performers in more than one category — for example, "INNOCENT USUAL" tops both the Family and Foreign categories, and "WIFE TURN" leads in both Documentary and New Releases. This highlights that some films transcend category boundaries, making them valuable multi-segment assets.
* Music, Documentary, and Sci-Fi lead with the highest individual film revenues, indicating that these genres have standout blockbuster-level content.
* Animation, Children, and Games categories all share a top film ("SLEEPLESS MONSOON" and "MONSOON ARIZONA") with the exact same revenue, which could indicate a bundled licensing deal or crossover appeal between categories.
* Comedy’s top earner, "ZORRO ARK," surpasses many other categories’ leaders, showing strong revenue potential for Comedy content.
* These findings are valuable for both targeted promotions (such as “Top in Genre” collections) and for licensing renewal decisions, ensuring the company retains rights to these critical films.
* The presence of some less traditional genres (like Documentary and Travel) in the top list signals that the company’s customer base may appreciate diverse content, not just mainstream titles.

**Query 7: Countries Generating Highest Revenue**

**Question:** Which countries generate the highest total revenue from film rentals?

**Code:**

SELECT rental\_country, SUM(total\_revenue) AS revenue

FROM film\_summary

GROUP BY rental\_country

ORDER BY revenue DESC

LIMIT 10;

**Screenshot:A white and grey striped background

Description automatically generated with medium confidence**

**Description:**

Identifying the top revenue-generating countries allows the company to focus its marketing, content acquisition, and regional strategies more effectively. Countries with higher revenue contributions may warrant localized content curation, targeted advertising campaigns, and premium service offerings. Understanding the revenue contribution from each country also helps the company prioritize investments in customer service, localized user interfaces, and culturally relevant promotions. This insight is particularly important for global expansion strategies, ensuring resources are focused on the most valuable markets.

**Findings:**

* India and China lead the list, highlighting their strong demand for film rentals and the significant revenue potential of these markets.
* The United States and Japan are also high performers, indicating mature markets with strong revenue contributions, which could benefit from ongoing content refreshes and premium offerings.
* The presence of emerging markets like Brazil, the Philippines, and Nigeria suggests that there is strong potential for growth in these regions. Tailored local content and affordable pricing strategies could help unlock further revenue.
* Countries with strong cultural film industries (like India and Mexico) might prefer local-language content alongside international titles, emphasizing the importance of localized content strategies.
* This geographic revenue analysis gives management a clear roadmap for regional investment decisions, ensuring marketing budgets, content spending, and customer engagement strategies are aligned with each country's revenue potential.

**Query 8: Revenue Contribution by Budget Category**

**Question:** How do films of different budget categories (low, mid, high) contribute to total revenue, and how many films fall into each budget category?

**Code:**

SELECT

CASE

WHEN replacement\_cost < 10 THEN 'Low Budget (<$10)'

WHEN replacement\_cost BETWEEN 10 AND 20 THEN 'Mid Budget ($10-$20)'

ELSE 'High Budget (>$20)'

END AS budget\_category,

COUNT(film\_id) AS num\_films,

SUM(total\_revenue) AS total\_revenue

FROM film\_summary

GROUP BY budget\_category

ORDER BY total\_revenue DESC;

**Screenshot:A screenshot of a computer

Description automatically generated**

**Description:**

Understanding the relationship between film budget and revenue generation helps the company make smarter content acquisition and pricing decisions. If high-budget films consistently drive the majority of revenue, management might prioritize acquiring more premium titles. On the other hand, if lower-budget films deliver strong returns relative to their cost, it could highlight the importance of maintaining a diverse content library that includes indie films, older titles, or niche content. This analysis also helps the company balance investment across different types of content to optimize profitability.

**Findings:**

* High-budget films generate the majority of total revenue, confirming that big-budget titles (likely blockbusters or premium content) drive customer demand and justify their higher acquisition costs.
* Mid-budget films also contribute significantly, indicating that moderately priced films still have strong commercial appeal, making them an important part of a balanced content strategy.
* Low-budget films account for a large number of titles but contribute relatively little revenue, which suggests they serve niche audiences or long-tail demand. This implies that while they may be valuable for content diversity or filling gaps in the catalog, they are unlikely to be core revenue drivers.
* This analysis supports a tiered content acquisition strategy, where the company invests heavily in premium blockbusters, maintains a solid base of mid-range content, and selectively acquires low-budget films that meet specific audience needs.

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.

**Dashboard:**

Data-Driven Insights for Strategic Decision-Making in Film Rentals

The following visualizations provide key insights into rental performance, revenue distribution, geographic trends, and consumer preferences. These findings can assist management in making informed data driven decisions to optimize operations, refine marketing strategies, and enhance content acquisition efforts.

**A close-up of a chart

Description automatically generated**

**Visualization 1: Geographic** **Revenue** **Insights**  
This visualization highlights total revenue generated across different countries. The size of each bubble represents the revenue volume, while distinct colors indicate different rental locations.

**Business Impact:**

Market Expansion: Identifying high-revenue countries enables strategic expansion and targeted investments.

Underperforming Regions: Understanding why certain regions generate lower revenue can inform pricing strategies and promotional efforts.

Localized Marketing: Developing region-specific campaigns can enhance customer engagement and rental performance.

**Management** **Decisions**:  
Invest in top-performing countries through marketing and content curation.  
Address barriers in low-revenue regions by assessing competition, pricing, and availability.  
Consider regional partnerships or exclusive content offerings for high-growth areas.

**A white page with colorful text

Description automatically generated with medium confidence**

**Visualization 2: Monthly** **Rental** **Trends** **by** **Category**  
This stacked bar chart visualizes rental volume over time, segmented by movie category. The distribution helps assess demand fluctuations across different months.

**Business Impact:**

Seasonality Analysis: Identifies peak rental months, enabling proactive inventory and marketing adjustments.

Category Trends: Detects shifts in genre popularity over time, guiding content acquisition and licensing strategies.

Demand Forecasting: Supports predictive analytics to optimize stock management and promotional efforts.

**Management** **Action** **Items**:  
Allocate advertising budgets effectively by capitalizing on peak months.  
Expand the catalog for high-demand genres while reconsidering underperforming categories.  
Implement discounted pricing or promotions during off-peak months to sustain revenue.

**A graph of a rental trend

Description automatically generated with medium confidence**

**Visualization3: Rental Performance by Genre**This treemap/heatmap categorizes rentals by genre, with larger sections representing more popular categories.

**Business Impact:**

Content Strategy Optimization: Understanding genre popularity assists in licensing and production decisions.

Personalized Recommendations: Using this data, the company can refine recommendation algorithms to enhance customer experience.

Revenue Optimization: Investing in high-performing genres ensures a greater return on content acquisitions.

**Management Decisions:**Prioritize licensing and marketing efforts for top genres.  
Discontinue or enhance marketing for underperforming categories.  
Use data-driven insights to improve recommendation algorithms for user engagement.

**A blue squares with white text

Description automatically generated**

**Visualization 4: Top 5 Most Rented Films**This visualization highlights the five most rented films, with bubble size representing total rentals.

**Business Impact:**

Customer Engagement: Popular films indicate audience preferences, which can inform promotional strategies.

Recommendation Systems: Data can be leveraged to suggest films similar to high-performing titles.

Licensing and Negotiations: Understanding which films perform best strengthens negotiation positions with studios and distributors.

**Management Decisions:**Feature top-performing films prominently on the platform.  
Consider exclusive licensing agreements for high-demand movies.  
Use this data to market similar films and increase engagement.

**A screen shot of a chart

Description automatically generated**

**Visualization 5: Top** **10** **Revenue**-**Generating** **Films**  
This pie chart illustrates the top ten films contributing the most to revenue.

**Business** **Impact**:

Maximizing Profitability: Identifying high-revenue films allows for targeted premium pricing strategies.

Sequel and Spin-Off Potential: High-performing movies may justify investments in sequels or similar content.

Content Acquisition: Prioritizing acquisitions of similar films ensures sustained revenue growth.

**Management Decisions:**Implement premium pricing for the highest-grossing films.  
Invest in sequels, spin-offs, or similar films based on revenue data.  
Strategically bundle lower-performing films with top rentals to enhance sales.

**A pie chart with numbers and a graph

Description automatically generated**