



SIMON BUSINESS SCHOOL

**(CIS467.SA.AN3. SPRING A 2025 SIMON)
Data Management, Warehousing, and Visualization**

05 March 2025

Final Project

Team 3F

Dhvani Patel

Hrishabh Tomar

Kavita Adha

Osigbemhe Ikhanoba

Syed Ibnul Haq

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.

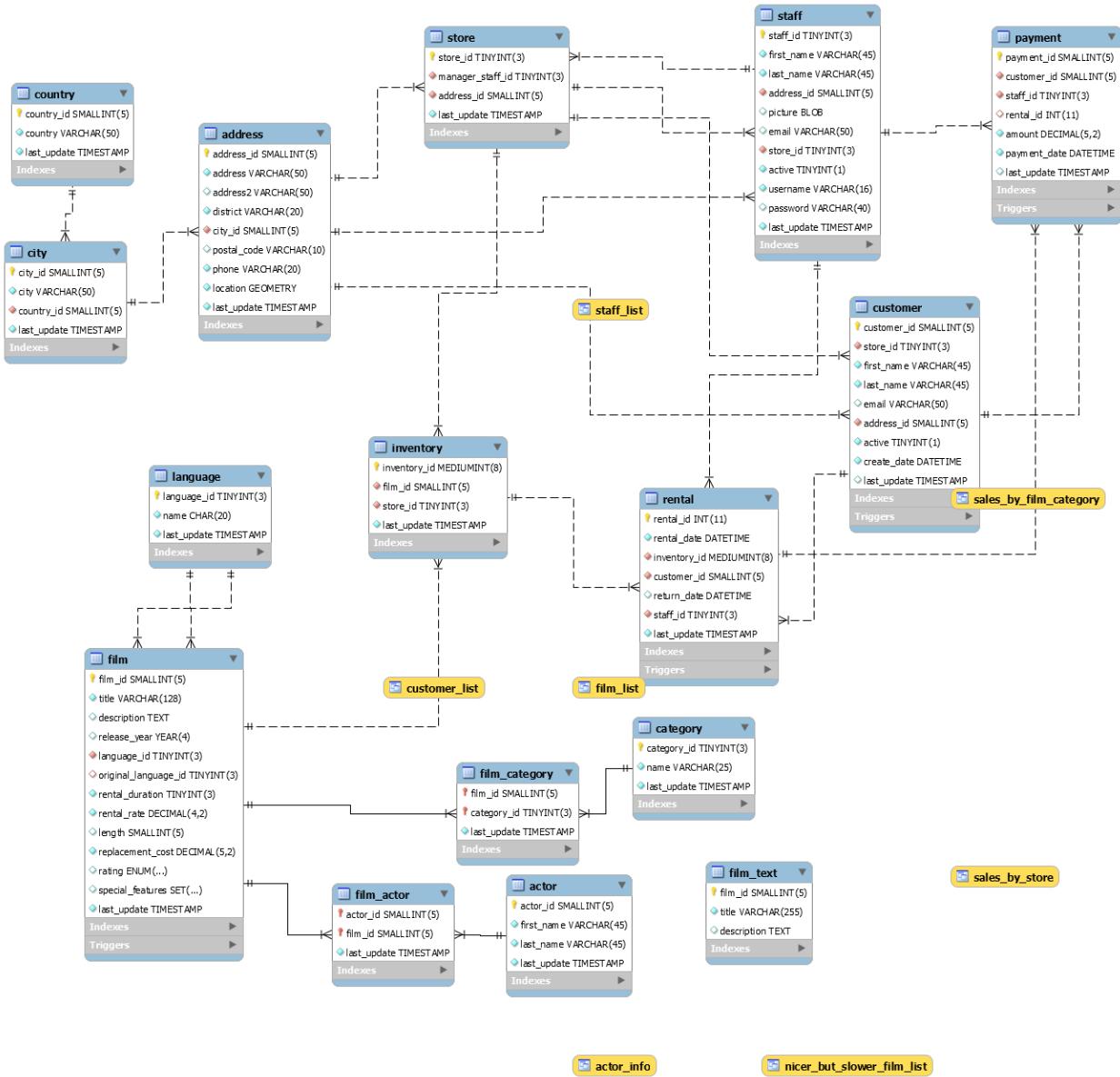
The script files sakila-schema.sql and sakila-data.sql create a database which contains tables (the database schema is below), with transactional data related to some company operations.

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.

Answer 1: DATA WAREHOUSE

Subject: Customer Insights and Strategic Analysis

Data warehouse Overview:

The Data Warehouse, named **CustomerInsightsDW**, is designed to aggregate and consolidate key customer-related metrics from the Sakila database into a single, denormalized table. This table gathers information from multiple source tables—such as **customer**, **rental**, **payment**, **film** and **film category**—to provide a holistic view of customer behavior. The primary focus is to capture customer segments according to revenue and consumer preferences for movies and genres to enable deep, actionable insights.

Key Data Elements Tracked

1. Customer Identification and Demographics:

- **Customer_ID and Customer_Name:** Uniquely identifies each customer and provides a human-readable name for reporting and analysis.
- **Location Information:** Includes **City**, and **Country** (sourced from the address, city, and country tables). This geographical data supports segmentation by region and enables location-based marketing strategies.

2. Transactional Behavior:

- **Total_Rentals:** Counts the number of rental transactions per customer. This metric helps assess customer engagement and frequency.

- **Total_Revenue:** Aggregates the sum of payments received from each customer. It reflects the overall revenue contribution of each customer.
- **Avg_Revenue_Per_Rental:** Calculates the average revenue per rental transaction, offering insight into spending behavior and pricing sensitivity.

3. Customer Preferences:

- **Highest_Rented_Category:** Identifies the film category that a customer rents most frequently. This helps in understanding genre preferences, which can be leveraged for personalized recommendations and targeted promotions.
- **Highest_Rented_Film:** Captures the most frequently rented film title for each customer. This adds another layer of insight into individual tastes and content popularity.

4. Operational Metrics:

- **Total_Late_Fees:** Sums any late fees incurred by the customer when films are returned past the expected rental period. This metric can indicate potential customer service issues or inform policies around rental returns.
- **First_Rental_Date and Last_Rental_Date:** Provide a timeline of customer activity, allowing you to analyze customer lifetime value, tenure, and recency of interactions.
- **Customer_Segment:** (Derived using NTILE over Total_Revenue) Classifies customers into 3 tiers—Platinum, Gold, and Regular. This segmentation is crucial for prioritizing high-value customers and tailoring loyalty programs.

Business Value and Use Cases:

1. Customer Segmentation & Targeted Marketing:

- By classifying customers into **Platinum** (top 20% revenue generators), **Gold** (next 20%), and **Regular** (remaining 60%), the business can target high-value customers with premium offers, rewards, or personalized communications. This segmentation facilitates effective marketing campaigns and helps in resource allocation.

2. Content and Catalog Optimization:

- Insights into the **Highest_Rented_Category** and **Highest_Rented_Film** help the company understand customer content preferences. This information is invaluable for decisions around inventory, licensing, and promotional strategies. For example, if a certain genre consistently drives revenue, the company might invest more in that category.

3. Regional Insights:

- Location data (district, city, country) allows the company to perform regional analysis, understand geographic trends in customer behavior, and tailor local marketing efforts. It

also aids in identifying regions with growth potential or those requiring additional support.

4. Performance Monitoring and Customer Lifetime Value (CLV):

- Tracking total rentals, total revenue, and average spending per rental provides insights into customer engagement and profitability. The data warehouse enables the calculation of CLV, allowing the company to identify which customers yield the most revenue over time and how their behavior evolves.

Code:

```
USE sakila;
DROP TABLE IF EXISTS CustomerInsightsDW;
```

```
CREATE TABLE CustomerInsightsDW AS
WITH base AS (
    SELECT
        c.customer_id AS Customer_ID,
        CONCAT(c.first_name, ' ', c.last_name) AS Customer_Name,
        COUNT(r.rental_id) AS Total_Rentals,
        ROUND(SUM(p.amount), 2) AS Total_Revenue,
        ROUND(SUM(p.amount) / COUNT(r.rental_id), 2) AS Avg_Revenue_Per_Rental,
        (SELECT cat.name
         FROM category cat
         JOIN film_category fc2 ON cat.category_id = fc2.category_id
         JOIN inventory i2 ON fc2.film_id = i2.film_id
         JOIN rental r2 ON i2.inventory_id = r2.inventory_id
         WHERE r2.customer_id = c.customer_id
         GROUP BY cat.name
         ORDER BY COUNT(*) DESC
         LIMIT 1) AS Highest_Rented_Category,
        (SELECT f2.title
         FROM film f2
         JOIN inventory i2 ON f2.film_id = i2.film_id
         JOIN rental r3 ON i2.inventory_id = r3.inventory_id
         WHERE r3.customer_id = c.customer_id
         GROUP BY f2.title
         ORDER BY COUNT(*) DESC
         LIMIT 1) AS Highest_Rented_Film,
        ROUND(COALESCE(SUM(CASE
            WHEN DATEDIFF(r.return_date, r.rental_date) > f.rental_duration
            THEN (DATEDIFF(r.return_date, r.rental_date) - f.rental_duration) * 0.50
            ELSE 0
        END))) AS CLV
    FROM customer c
    JOIN rental r ON c.customer_id = r.customer_id
    JOIN payment p ON r.rental_id = p.rental_id
    JOIN film f ON p.film_id = f.film_id
)
SELECT *
FROM base;
```

```

        END), 0), 2) AS Total_Late_Fees,
        (SELECT MIN(rental_date) FROM rental WHERE customer_id = c.customer_id) AS
First_Rental_Date,
        (SELECT MAX(rental_date) FROM rental WHERE customer_id = c.customer_id) AS
Last_Rental_Date,
        ci.city AS Customer_City,
        co.country AS Customer_Country
FROM customer c
JOIN address a ON c.address_id = a.address_id
JOIN city ci ON a.city_id = ci.city_id
JOIN country co ON ci.country_id = co.country_id
LEFT JOIN rental r ON c.customer_id = r.customer_id
LEFT JOIN inventory i ON r.inventory_id = i.inventory_id
LEFT JOIN film f ON i.film_id = f.film_id
LEFT JOIN payment p ON r.rental_id = p.rental_id
LEFT JOIN film_category fc ON i.film_id = fc.film_id
GROUP BY
        c.customer_id,
        Customer_Name,
        ci.city,
        co.country
)
SELECT
        base.*,
        CASE
                WHEN NTILE(5) OVER (ORDER BY base.Total_Revenue) = 5 THEN 'Platinum'
                WHEN NTILE(5) OVER (ORDER BY base.Total_Revenue) = 4 THEN 'Gold'
                ELSE 'Regular'
        END AS Customer_Segment
FROM base;

```

C...	Customer_Name	Total_...	Total_...	Avg_...	Highest_Re...	Highest_Rente...	Tot...	First_Rental_D...	Last_Rental_...	Customer_...	Custom...	Custo...
1	MARY SMITH	32	118.68	3.71	Classics	PATIENT SISTER	12.50	2005-05-25 11:...	2005-08-22...	Sasebo	Japan	Gold
2	PATRICIA JOHNS...	27	128.73	4.77	Sports	DOORS PRESI...	23.00	2005-05-27 00:...	2005-08-23...	San Berna...	United S...	Gold
3	LINDA WILLIAMS	26	135.74	5.22	Action	RINGS HEART...	25.00	2005-05-27 17:...	2005-08-23...	Athenai	Greece	Platin...
4	BARBARA JONES	22	81.78	3.72	Horror	DOGMA FAMILY	9.00	2005-06-15 09:...	2005-08-23...	Myingyan	Myanmar	Regular
5	ELIZABETH BRO...	38	144.82	3.81	Classics	TOOTSIE PILOT	22.50	2005-05-29 07:...	2006-02-14...	Nantou	Taiwan	Platin...
6	JENNIFER DAVIS	28	93.72	3.35	Drama	ENGLISH BUL...	19.00	2005-05-25 08:...	2005-08-23...	Laredo	United S...	Regular
7	MARIA MILLER	33	151.67	4.60	Animation	TELEMARK HE...	26.50	2005-05-25 06:...	2005-08-21...	Kragujevac	Yugoslavia	Platin...
8	SUSAN WILSON	24	92.76	3.87	Drama	MALKOVICH PET	11.50	2005-05-30 03:...	2005-08-23...	Hamilton	New Zea...	Regular
9	MARGARET MOO...	23	89.77	3.90	Foreign	MULAN MOON	10.50	2005-05-27 05:...	2006-02-14...	Masqat	Oman	Regular
10	DOROTHY TAYLOR	25	99.75	3.99	Games	SNOWMAN RO...	17.50	2005-05-31 19:...	2005-08-22...	Esfahan	Iran	Regular
11	LISA ANDERSON	24	106.76	4.45	Games	BOOGIE AMELIE	19.50	2005-05-30 22:...	2006-02-14...	Sagamihara	Japan	Regular
12	NANCY THOMAS	28	103.72	3.70	Drama	ROCK INSTINCT	18.00	2005-05-30 23:...	2005-08-23...	Yamuna N...	India	Regular
13	KAREN JACKSON	27	131.73	4.88	Classics	DECEIVER BE...	23.50	2005-06-17 06:...	2005-08-22...	Osmaniye	Turkey	Gold
14	BETTY WHITE	28	117.72	4.20	Games	OUTBREAK DI...	19.00	2005-05-26 00:...	2006-02-14...	Citrus Heig...	United S...	Gold
15	HELEN HARRIS	32	134.68	4.21	Foreign	CANDIDATE P...	15.50	2005-06-18 21:...	2006-02-14...	Bhopal	India	Platin...
16	SANDRA MARTIN	28	118.72	4.24	Sports	FEUD FROGMEN	14.50	2005-05-27 03:...	2005-08-21...	Southend...	United K...	Gold
17	DONNA THOMPMS...	21	98.79	4.70	Foreign	SHREK LICENSE	15.00	2005-05-26 19:...	2005-08-23...	Elista	Russian...	Regular
18	CAROL GARCIA	22	91.78	4.17	Classics	YENTL IDAHO	9.00	2005-05-25 06:...	2005-08-20...	Kaduna	Nigeria	Regular
19	RUTH MARTINEZ	24	125.76	5.24	Sports	TALENTED HO...	22.00	2005-05-25 01:...	2005-08-23...	Kimberley	South Af...	Gold
20	SHARON ROBINS...	30	115.70	3.86	New	CHOCOLAT HA...	15.00	2005-05-26 07:...	2005-08-23...	Mardan	Pakistan	Regular
21	MICHELLE CLARK	35	155.65	4.45	Sci-Fi	LOSE INCH	29.50	2005-05-26 15:...	2006-02-14...	Tangail	Banglad...	Platin...
22	LAURA RODRIGU...	22	113.78	5.17	Foreign	ROBBERS JOON	17.00	2005-05-27 07:...	2006-02-14...	Salé	Morocco	Regular
23	SARAH LEWIS	30	119.70	3.99	Sci-Fi	MOURNING PU...	18.00	2005-05-25 21:...	2006-02-14...	Liepaja	Latvia	Gold
24	KIMBERLY LEE	25	95.75	3.83	Animation	PRINCESS GIA...	12.50	2005-05-31 01:...	2005-08-22...	Córdoba	Argentina	Regular
25	DEBORAH WALKER	29	115.71	3.99	Action	PANKY SUBMA...	11.50	2005-05-25 14:...	2005-08-23...	Shikarpur	Pakistan	Regular

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

WITH CustomerSpending AS (

SELECT

Customer_ID,

Customer_Name,

Total_Revenue AS total_spent

FROM CustomerInsightsDW

),

RankedCustomers AS (

```

SELECT
    cs.*,
    NTILE(100) OVER (ORDER BY total_spent DESC) AS percentile_rank
FROM CustomerSpending cs
)
SELECT
    rc.Customer_ID,
    rc.Customer_Name,
    rc.total_spent,
    CASE
        WHEN percentile_rank <= 10 THEN 'Platinum'
        WHEN percentile_rank <= 50 THEN 'Gold'
        ELSE 'Regular'
    END AS Customer_Tier
FROM RankedCustomers rc
ORDER BY rc.total_spent DESC;

```

Customer_ID	Customer_Name	total_spent	Customer_Tier
526	KARL SEAL	221.55	Platinum
148	ELEANOR HUNT	216.54	Platinum
144	CLARA SHAW	195.58	Platinum
178	MARION SNYDER	194.61	Platinum
137	RHONDA KENNEDY	194.61	Platinum
459	TOMMY COLLAZO	186.62	Platinum
469	WESLEY BULL	177.60	Platinum
468	TIM CARY	175.61	Platinum
236	MARCIA DEAN	175.58	Platinum
181	ANA BRADLEY	174.66	Platinum
176	JUNE CARROLL	173.63	Platinum
50	DIANE COLLINS	169.65	Platinum
259	LENA JENSEN	168.68	Platinum
522	ARNOLD HAVENS	167.67	Platinum
410	CURTIS IRBY	167.62	Platinum
403	MIKE WAY	166.65	Platinum
295	DAISY BATES	162.62	Platinum
209	TONYA CHAPMAN	161.68	Platinum
373	LOUIS LEONE	161.65	Platinum
470	GORDON ALLARD	160.68	Platinum
187	BRITTANY RILEY	159.72	Platinum
550	GUY BROWNLEE	159.68	Platinum
462	WARREN SHERROD	159.67	Platinum
267	MARGIE WADE	159.64	Platinum
293	MAE FLETCHER	158.69	Platinum
372	STEVE MACKENZIE	158.66	Platinum

Query 1: Customer Segmentation for Strategic Growth

This SQL Code Addresses

This query helps the company categorize customers based on their total spending, dividing them into Platinum (Top 10%), Gold (Next 40%), and Regular (Bottom 50%). The goal is to identify high-value customers, analyze their rental behaviors, and implement engagement strategies that maximize retention and revenue.

Insights-

- Platinum customers generate the highest revenue and are crucial for long-term profitability, making them a key focus for loyalty programs.
- Gold customers have a strong spending pattern but room for increased engagement, presenting an opportunity for targeted marketing.
- Regular customers rent infrequently, indicating a higher risk of churn and requiring focused re-engagement efforts.

Business Decisions-

- Enhance retention for Platinum customers by offering VIP perks, such as priority rentals, early access to new releases, and personalized recommendations, ensuring they continue renting frequently.
- Encourage Gold customers to move up to Platinum by introducing subscription plans, spending-based discounts, and loyalty rewards to increase their lifetime value.
- Re-engage Regular customers by offering promotions, sending reminders, or bundling offers to encourage repeat rentals and improve customer retention.

This segmentation provides a clear roadmap for management to optimize customer engagement, allocate resources efficiently, and drive long-term revenue growth.

-- Query 2

```
WITH LateReturns AS (
    SELECT
        Customer_ID,
        Customer_Name,
        Total_Rentals,
        Total_Late_Fees,
        ROUND(Total_Late_Fees / NULLIF(Total_Rentals, 0), 2) AS Avg_Late_Fee_Per_Rental,
        NTILE(100) OVER (ORDER BY Total_Late_Fees DESC) AS Percentile_Rank
    FROM CustomerInsightsDW
    WHERE Total_Late_Fees > 0
)
SELECT
    lr.Customer_ID,
```

```

lr.Customer_Name,
lr.Total_Rentals,
lr.Total_Late_Fees,
lr.Avg_Late_Fee_Per_Rental,
CASE
    WHEN lr.Percentile_Rank <= 10 THEN 'High Risk'
    WHEN lr.Percentile_Rank <= 50 THEN 'Moderate Risk'
    ELSE 'Low Risk' -- Bottom 50%
END AS Late_Fee_Risk_Category
FROM LateReturns lr
ORDER BY lr.Total_Late_Fees DESC
LIMIT 25;

```

Customer_ID	Customer_Name	Total_Rentals	Total_Late_Fees	Avg_Late_Fee_Per_Rental	Late_Fee_Risk_Category
526	KARL SEAL	45	41.50	0.92	High Risk
469	WESLEY BULL	40	36.00	0.90	High Risk
148	ELEANOR HUNT	46	34.50	0.75	High Risk
137	RHONDA KENNEDY	39	34.00	0.87	High Risk
187	BRITTANY RILEY	28	33.00	1.18	High Risk
506	LESLIE SEWARD	35	33.00	0.94	High Risk
373	LOUIS LEONE	35	32.50	0.93	High Risk
207	GERTRUDE CASTILLO	34	31.00	0.91	High Risk
181	ANA BRADLEY	34	30.50	0.90	High Risk
468	TIM CARY	39	30.50	0.78	High Risk
144	CLARA SHAW	42	30.00	0.71	High Risk
178	MARION SNYDER	39	30.00	0.77	High Risk
21	MICHELLE CLARK	35	29.50	0.84	High Risk
82	KATHRYN COLEMAN	26	29.50	1.13	High Risk
436	TROY QUIGLEY	30	29.50	0.98	High Risk
259	LENA JENSEN	32	29.50	0.92	High Risk
176	JUNE CARROLL	37	28.50	0.77	High Risk
50	DIANE COLLINS	35	28.50	0.81	High Risk
295	DAISY BATES	38	28.50	0.75	High Risk
522	ARNOLD HAVENS	33	28.50	0.86	High Risk
293	MAE FLETCHER	31	28.00	0.90	High Risk
479	ZACHARY HITE	31	28.00	0.90	High Risk
39	DEBRA NELSON	29	27.50	0.95	High Risk
221	BESSIE MORRISON	28	27.50	0.98	High Risk
459	TOMMY COLLAZO	38	27.50	0.72	High Risk

Query 2: Identifying Late Return Trends & Customer Risk Segmentation

This SQL Code Addresses-

This query identifies customers who frequently return rentals late and ranks them based on total late fees incurred. By segmenting them into High Risk (Top 10%), Moderate Risk (Next 40%), and Low Risk (Bottom 50%), the company can take proactive measures to reduce late returns and optimize revenue collection.

Insights-

- High-risk customers (Top 10%) have significant late fees, indicating they frequently return rentals past the due date, impacting inventory availability and potential revenue loss.
- Moderate-risk customers (Next 40%) return movies late occasionally, contributing to late fee revenue but also showing behavioral patterns that could be improved.

- Low-risk customers (Bottom 50%) rarely return rentals late, making them reliable renters who maintain smooth inventory turnover.

Business Decisions-

- For high-risk customers, stricter policies such as higher penalties, mandatory deposits, or enhanced reminders can help reduce late returns.
- For moderate-risk customers, implementing grace periods, small incentives for timely returns, or discount-based penalties may encourage better return habits.
- For low-risk customers, the company could reward responsible renters by offering perks like waived late fees or bonus rental credits to enhance loyalty.

By analyzing late return behavior, management can optimize rental policies, improve cash flow, reduce inventory bottlenecks, and ensure a more efficient rental system.

-- Query 3

```
WITH CountryRevenue AS (
    SELECT
        Customer_Country,
        COUNT(Customer_ID) AS Total_Customers,
        SUM(Total_Revenue) AS Total_Revenue,
        ROUND(SUM(Total_Revenue) / COUNT(Customer_ID), 2) AS
        Avg_Revenue_Per_Customer
    FROM CustomerInsightsDW
    GROUP BY Customer_Country
)
SELECT
    cr.Customer_Country,
    cr.Total_Customers,
    cr.Total_Revenue,
    cr.Avg_Revenue_Per_Customer,
    RANK() OVER (ORDER BY cr.Total_Revenue DESC) AS Revenue_Rank
FROM CountryRevenue cr
ORDER BY cr.Total_Revenue DESC;
```

Customer_Country	Total_Customers	Total_Revenue	Avg_Revenue_Per_Customer	Revenue_Rank
India	60	6628.28	110.47	1
China	53	5798.74	109.41	2
United States	36	4110.32	114.18	3
Japan	31	3470.75	111.96	4
Mexico	30	3307.04	110.23	5
Brazil	28	3200.52	114.30	6
Russian Federation	28	3045.87	108.78	7
Philippines	20	2381.32	119.07	8
Turkey	15	1662.12	110.81	9
Nigeria	13	1511.48	116.27	10
Indonesia	14	1510.33	107.88	11
Argentina	13	1434.48	110.34	12
Taiwan	10	1209.95	121.00	13
South Africa	11	1204.15	109.47	14
Iran	8	950.75	118.84	15
United Kingdom	9	922.81	102.53	16
Poland	8	877.97	109.75	17
Italy	7	831.11	118.73	18
Germany	7	831.04	118.72	19
Vietnam	6	746.28	124.38	20
Ukraine	6	730.42	121.74	21
Colombia	6	709.41	118.24	22
Egypt	6	694.39	115.73	23
Venezuela	7	683.30	97.61	24
Spain	5	606.58	121.32	25
Canada	5	593.63	118.73	26
Netherlands	5	586.66	117.33	27
Grand Total	-	5714.05	111.66	28

Query 3: Revenue Contribution by Country

This SQL Code Addresses-

This query analyzes revenue distribution across different countries, ranking them based on total revenue, customer count, and average revenue per customer. The goal is to identify high-value markets, assess regional performance, and optimize business expansion strategies.

Insights-

- Top-ranked countries generate the highest revenue, showcasing strong market penetration and customer engagement.
- Some mid-tier countries have high customer counts but lower revenue per customer, suggesting an opportunity to increase spending per user through targeted incentives.
- Lower-ranked countries contribute minimal revenue, indicating potential underperformance or limited market reach.

Business Decisions-

- Invest in top-performing countries by reinforcing localized promotions, premium content, and exclusive rental offers to sustain growth.
- Optimize pricing and marketing efforts in mid-tier regions to encourage higher spending per customer.
- Evaluate the viability of low-revenue countries, deciding whether to expand presence or shift focus to more profitable markets.

By utilizing these insights, management can allocate resources strategically, optimize international market performance, and drive sustainable revenue growth

-- Query 4

```
WITH GenrePreferences AS (
    SELECT
        Customer_ID,
        Customer_Name,
        Highest_Rented_Category AS Genre,
        SUM(Total_Rentals) AS Total_Rentals,
        SUM(Total_Revenue) AS Total_Genre_Revenue,
        RANK() OVER (PARTITION BY Customer_ID ORDER BY SUM(Total_Rentals) DESC,
        SUM(Total_Revenue) DESC) AS Genre_Rank
    FROM CustomerInsightsDW
    GROUP BY Customer_ID, Customer_Name, Highest_Rented_Category
)
SELECT
    Customer_ID,
    Customer_Name,
    Genre,
    Total_Rentals,
    Total_Genre_Revenue
FROM GenrePreferences
WHERE Genre_Rank = 1
ORDER BY Total_Rentals DESC, Total_Genre_Revenue DESC;
```

Customer_ID	Customer_Name	Genre	Total_Rentals	Total_Genre_Revenue
148	ELEANOR HUNT	Sci-Fi	46	216.54
526	KARL SEAL	Animation	45	221.55
144	CLARA SHAW	Drama	42	195.58
236	MARICA DEAN	Foreign	42	175.58
75	TAMMY SANDERS	Family	41	155.59
469	WESLEY BULL	Games	40	177.60
197	SUE PETERS	Comedy	40	154.60
137	RHONDA KENNEDY	Games	39	194.61
178	MARION SNYDER	Travel	39	194.61
468	TIM CARY	Animation	39	175.61
459	TOMMY COLLAZO	Family	38	186.62
410	CURTIS IRBY	Family	38	167.62
295	DAISY BATES	Family	38	162.62
5	ELIZABETH BROWN	Classics	38	144.62
176	JUNE CARROLL	Animation	37	173.63
366	BRANDON HUEY	Sports	37	152.63
257	MARSHA DOUGLAS	Documentary	37	151.63
198	ELSIE KELLEY	Documentary	37	141.63
267	MARGIE WADE	New	36	159.64
439	ALEXANDER FENNELL	Family	36	151.64
348	ROGER QUINTANILLA	Animation	36	146.64
29	ANGELA HERNANDEZ	Children	36	140.64
380	RUSSELL BRINSON	Foreign	36	136.64
354	JUSTIN NGO	Documentary	36	129.64
50	DIANE COLLINS	Drama	35	169.65
403	MIKE WAY	Sci-Fi	35	166.65
373	LOUIS LEONE	Animation	35	161.65

Query 4: Identifying Customer Genre Preferences for Targeted Engagement

This SQL Code Addressess-

This query identifies each customer's most rented genre, helping the company understand viewing preferences and tailor content recommendations to enhance engagement and retention.

Insights-

- Customers have clear genre preferences, with some genres contributing higher revenue despite lower rental counts (indicating premium appeal).
- High-frequency renters in a specific genre present opportunities for personalized promotions and exclusive offers.
- Lesser-rented genres may need reevaluation, as they could either be underperforming or niche-interest categories.

Business Decisions-

- Enhance customer retention by providing personalized recommendations and exclusive offers for favorite genres.
- Increase revenue per customer by introducing genre-specific promotions and bundled offers.
- Optimize inventory and content strategy by ensuring high-demand genres are prioritized while reassessing underperforming categories.

These insights help management boost customer engagement, refine marketing strategies, and maximize profitability.

-- Query 5

```
WITH GenrePreferences AS (
    SELECT
        Customer_ID,
        Highest_Rented_Category AS Genre
    FROM CustomerInsightsDW
    GROUP BY Customer_ID, Highest_Rented_Category
),
RankedGenres AS (
    SELECT
        gp.Customer_ID,
        ci.Customer_Name,
        gp.Genre,
        RANK() OVER (PARTITION BY gp.Customer_ID ORDER BY COUNT(*) DESC) AS
        Genre_Rank
    FROM GenrePreferences gp
    JOIN CustomerInsightsDW ci ON gp.Customer_ID = ci.Customer_ID
    GROUP BY gp.Customer_ID, ci.Customer_Name, gp.Genre
)
SELECT
    Customer_ID,
    Customer_Name,
    Genre
FROM RankedGenres
WHERE Genre_Rank = 1
ORDER BY Customer_Name;
```

Customer_ID	Customer_Name	Genre
375	AARON SELBY	Travel
367	ADAM GOOCH	Foreign
525	ADRIAN CLARY	Family
217	AGNES BISHOP	Drama
389	ALAN KAHN	Sports
352	ALBERT CROUSE	Music
568	ALBERTO HENNING	Documentary
454	ALEX GRESHAM	Sports
439	ALEXANDER FENNELL	Family
423	ALFRED CASILLAS	Drama
567	ALFREDO MCADAMS	Travel
51	ALICE STEWART	Action
152	ALICIA MILLS	Comedy
548	ALLAN CORNISH	Children
412	ALLEN BUTTERFIELD	Action
228	ALLISON STANLEY	Foreign
196	ALMA AUSTIN	Animation
467	ALVIN DELOACH	Children
40	AMANDA CARTER	Games
139	AMBER DIXON	Games
32	AMY LOPEZ	Family
181	ANA BRADLEY	Family
515	ANDRE RAPP	Classics
81	ANDREA HENDERSON	Horror
333	ANDREW PURDY	Games
582	ANDY VANHORN	Sci-Fi
503	ANGEL BARCLAY	Sci-Fi
29	ANGELA HERNANDEZ	Children

Query 5: Customer Genre Preferences for Personalized Marketing This SQL Code Addressess

This query identifies each customer's most preferred movie genre by ranking genres based on rental frequency per customer. The goal is to enhance personalization, improve customer experience, and drive targeted marketing efforts.

Insights-

- Customers have distinct genre preferences, which can be leveraged for tailored recommendations and targeted promotions.
- Some genres are consistently preferred across multiple customers, indicating high-demand content categories.
- Lesser-engaged genres may be niche markets or underperforming categories, requiring further analysis.

Business Decisions-

- Increase customer retention by delivering personalized content recommendations based on individual preferences.
- Drive higher engagement and revenue by offering genre-specific promotions, such as discounted rentals or exclusive content bundles.
- Optimize content acquisition and inventory by prioritizing popular genres while reassessing demand for lesser-rented categories.

These insights enable management to refine content strategies, enhance user satisfaction, and maximize revenue through data-driven personalization.

-- Query 6

WITH FilmPreferences AS (

SELECT

Customer_ID,
Highest_Rented_Film AS Film_Title,
SUM(Total_Revenue) AS Total_Film_Revenue,
SUM(Total_Rentals) AS Total_Film_Rentals,
ROUND(SUM(Total_Revenue) / NULLIF(SUM(Total_Rentals), 0), 2) AS
Revenue_Per_Rental

FROM CustomerInsightsDW

GROUP BY Customer_ID, Highest_Rented_Film

),

RankedFilms AS (

```
SELECT
    fp.Customer_ID,
    ci.Customer_Name,
    fp.Film_Title,
    fp.Total_Film_Revenue,
    fp.Total_Film_Rentals,
    fp.Revenue_Per_Rental,
    RANK() OVER (PARTITION BY fp.Customer_ID ORDER BY fp.Revenue_Per_Rental
DESC) AS Film_Rank
FROM FilmPreferences fp
JOIN CustomerInsightsDW ci ON fp.Customer_ID = ci.Customer_ID
)
SELECT
    Customer_ID,
    Customer_Name,
    Film_Title,
    Total_Film_Rentals,
    Total_Film_Revenue,
    Revenue_Per_Rental
FROM RankedFilms
WHERE Film_Rank = 1
ORDER BY Revenue_Per_Rental DESC, Total_Film_Revenue DESC;
```

Customer_ID	Customer_Name	Film_Title	Total_Film_Rentals	Total_Film_Revenue	Revenue_Per_Rental
187	BRITTANY RILEY	EAGLES PANKY	28	159.72	5.70
433	DON BONE	FIDELITY DEVIL	25	133.75	5.35
321	KEVIN SCHULER	TITANS JERK	22	116.78	5.31
259	LENA JENSEN	CENTER DINOSAUR	32	168.68	5.27
542	LONNIE TIRADO	ELIZABETH SHANE	18	94.82	5.27
311	PAUL TROUT	CHICKEN HELLFIGHTERS	23	120.77	5.25
19	RUTH MARTINEZ	TALENTED HOMICIDE	24	125.76	5.24
3	LINDA WILLIAMS	RINGS HEARTBREAKERS	26	135.74	5.22
285	MIRIAM MCKINNEY	GOODFELLAS SALUTE	26	135.74	5.22
22	LAURA RODRIGUEZ	ROBBERS JOON	22	113.78	5.17
181	ANA BRADLEY	BUCKET BROTHERHOOD	34	174.66	5.14
293	MAE FLETCHER	REIGN GENTLEMEN	31	158.69	5.12
583	MARSHALL THORN	PHILADELPHIA WIFE	23	117.77	5.12
508	MILTON HOWLAND	DARKNESS WAR	25	127.75	5.11
522	ARNOLD HAVENS	DOUBLE WRATH	33	167.67	5.08
432	EDWIN BURK	SUPER WYOMING	23	116.77	5.08
571	JOHNNIE CHISHOLM	CALIFORNIA BIRDS	24	121.76	5.07
200	JEANNE LAWSON	AUTUMN CROW	27	136.73	5.06
209	TONYA CHAPMAN	GRINCH MASSAGE	32	161.68	5.05
429	FREDERICK ISBELL	DADDY PITTSBURGH	21	105.79	5.04
509	RAUL FORTIER	CLOSER BANG	20	100.80	5.04
82	KATHRYN COLEMAN	STRANGELOVE DESIRE	26	130.74	5.03
470	GORDON ALLARD	BINGO TALENTED	32	160.68	5.02
137	RHONDA KENNEDY	DAWN POND	39	194.61	4.99
178	MARION SNYDER	WIFE TURN	39	194.61	4.99
550	GUY BROWNLEE	LOST BIRD	32	159.68	4.99
591	KENT ARSENAULT	SLACKER LIAISONS	27	134.73	4.99
477	DAN PAINÉ	STAR OPERATION	22	109.78	4.99
449	OSCAR AQUINO	DROP WATERFRONT	20	99.80	4.99
315	KENNETH GOODEN	PINOCCHIO SIMON	17	84.83	4.99

Query 6: Identifying Top Customer Film Preferences for Targeted Engagement

This SQL Code Addresses-

This query determines each customer's most rented film and evaluates total revenue, rental frequency, and revenue per rental to identify high-demand content and personalized marketing opportunities.

Insights-

- Customers show strong individual film preferences, allowing for precise content recommendations.
- Certain films generate higher revenue per rental, indicating premium appeal.
- High-rental, low-revenue films suggest opportunities for upselling premium versions or bundles.

Business Decisions-

- Personalized film recommendations can improve engagement and retention.
- High-value films can be monetized through premium pricing, exclusive editions, or early access.
- Content acquisition can be optimized by prioritizing films with strong engagement.

These insights help maximize customer satisfaction, optimize content strategy, and drive revenue growth.

-- Query 7

WITH RentalCounts AS (

```
SELECT
    MONTHNAME(First_Rental_Date) AS rental_month,
    COUNT(*) AS rental_count
FROM CustomerInsightsDW
GROUP BY rental_month
```

UNION ALL

```
SELECT
    MONTHNAME(Last_Rental_Date) AS rental_month,
    COUNT(*) AS rental_count
FROM CustomerInsightsDW
GROUP BY rental_month
```

)

SELECT

```
    rental_month,
    SUM(rental_count) AS total_rentals
FROM RentalCounts
GROUP BY rental_month
ORDER BY total_rentals DESC
LIMIT 25;
```

rental_month	total_rentals
May	520
August	441
February	158
June	78
July	1

---- Query 7: Identifying Monthly Rental Trends for Demand Forecasting This SQL Code Addresses-

This query analyzes rental activity by month, identifying trends in customer engagement and seasonal demand. By examining both first and last rental dates, it helps determine peak and low-demand periods for strategic planning.

Insights-

- Certain months have higher rental activity, suggesting peak seasons for movie rentals.
- Some months show low engagement, indicating periods where demand drops.
- Seasonal trends can influence marketing campaigns and content availability.

Business Decisions-

- Optimize promotional campaigns by offering discounts or exclusive content during low-rental months.
 - Increase content availability during peak months to maximize revenue and prevent shortages.
 - Plan strategic marketing efforts around seasonal trends to boost engagement and retention.

These insights help align inventory, marketing, and pricing strategies with customer behavior, ensuring higher profitability and sustained engagement.

```
WITH GenreRevenue AS (
    SELECT
        Customer_Country AS country,
        Highest_Rented_Category AS genre,
        SUM(Total_Revenue) AS total_revenue
    FROM CustomerInsightsDW
    GROUP BY Customer_Country, Highest_Rented_Category
)
SELECT
    gr.country,
    gr.genre,
    gr.total_revenue
FROM GenreRevenue gr
ORDER BY gr.total_revenue DESC;
```

country	genre	total_revenue
China	Drama	989.52
China	Sports	887.89
India	Action	867.00
Japan	Sci-Fi	785.10
India	Sci-Fi	653.50
India	Foreign	652.43
India	Sports	641.60
China	Classics	601.48
Brazil	Sci-Fi	595.45
India	Documentary	562.59
United States	Documentary	544.66
India	Drama	526.66
India	Animation	509.79
Brazil	Sports	505.83
Mexico	Sci-Fi	504.85
India	New	488.92
China	Animation	470.81
Philippines	Animation	458.99
Russian Federation	Family	452.01
United States	Animation	444.96
China	New	422.98
China	Action	416.98
Mexico	Sports	409.95
Philippines	Foreign	402.00
Argentina	Action	396.10
United States	Family	387.11
United States	Games	383.02
Russian Federation	Documentary	381.10
Japan	Children	381.04
China	Music	378.10

Query 8: Analyzing Genre Revenue by Country for Market Optimization

This SQL Code Addresses-

This query examines which movie genres generate the highest revenue in each country, helping identify regional content preferences and optimize market-specific strategies.

Insights-

- Certain genres dominate revenue in specific countries, indicating strong local preferences.
- Some countries generate high overall revenue, but specific genres contribute disproportionately.
- Regional genre demand varies, suggesting opportunities for tailored content marketing.

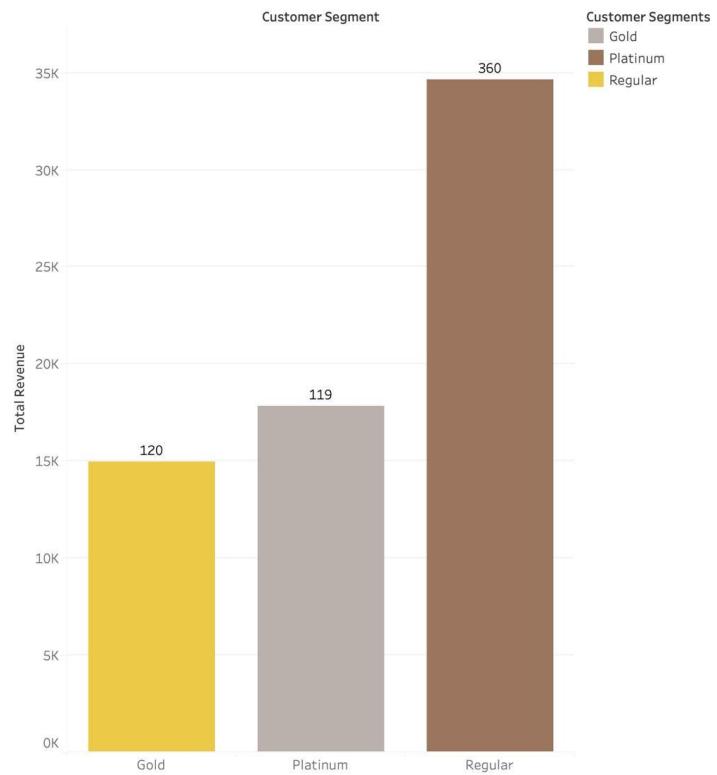
Business Decisions-

- Customize content offerings by prioritizing popular genres in each country.
- Optimize marketing campaigns by targeting users with region-specific recommendations.
- Strategically allocate content acquisition budgets to focus on high-performing genres by country.

By leveraging these insights, management can enhance localization strategies, improve customer engagement, and drive higher international revenue.

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.

Customer Segmentation by Revenue

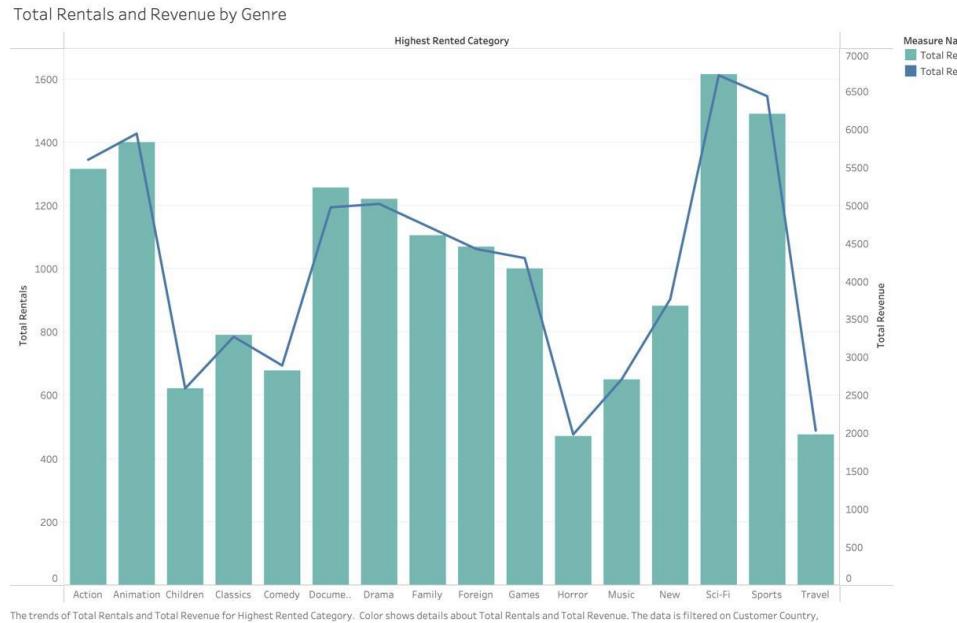


Sum of Total Revenue for each Customer Segment. Color shows details about Customer Segments. The marks are labeled by distinct count of Customer ID. The data is filtered on Customer Country, which keeps 108 of 108 members. The view is filtered on Customer Segments, which keeps multiple members.

Visual 1: Customer Segmentation by Revenue

Business Description:

This bar chart shows **total revenue** generated by three customer segments—**Gold**, **Platinum**, and **Regular**. Each bar's height and label represent how much that segment contributes to overall sales. Management can use these insights to **pinpoint high-value segments**, allocate **marketing budgets** effectively, and **develop targeted retention strategies** for each group.

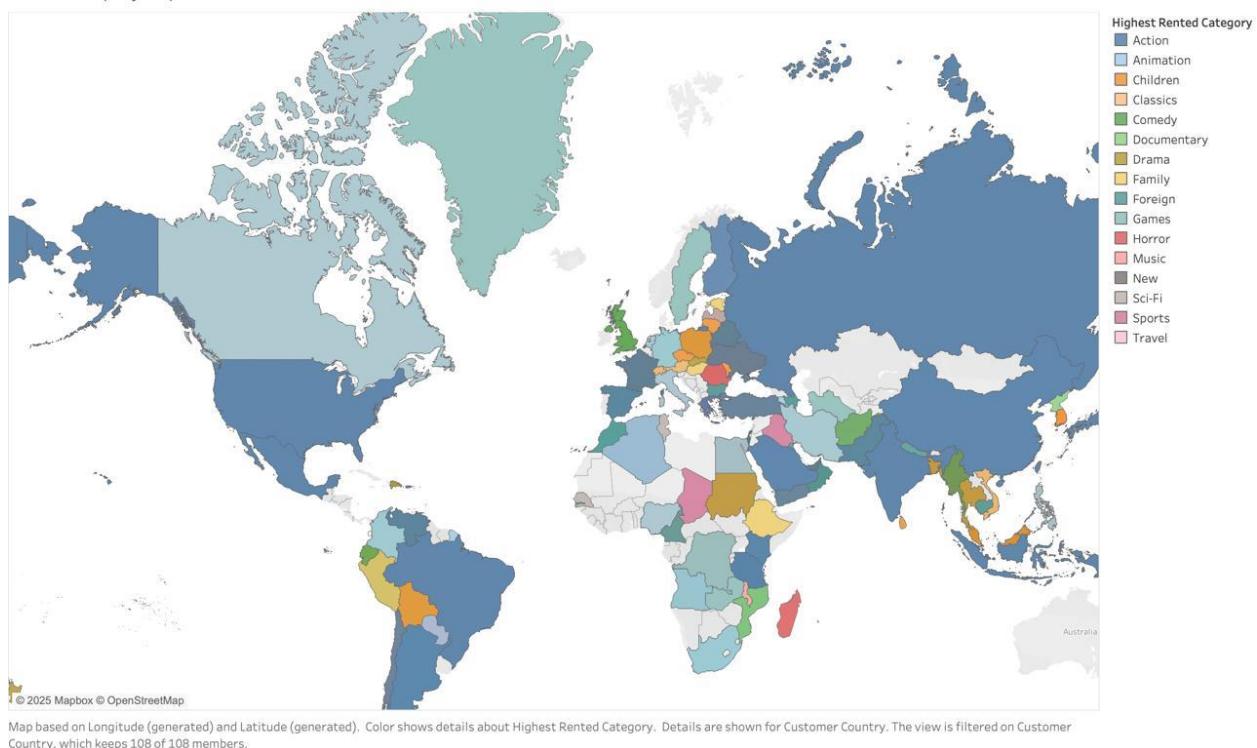


Visual 2: Total Rentals and Revenue by Genre

Business Description:

This combined bar-and-line chart compares **total rentals** (bars) and **total revenue** (line) for each film genre. It highlights which genres attract the most rentals versus which drive the highest revenue. Management can use this insight to **prioritize inventory**, **tailor marketing**, and **adjust pricing** to optimize both popularity and profitability of each genre.

World Map by Top Genre and Revenue

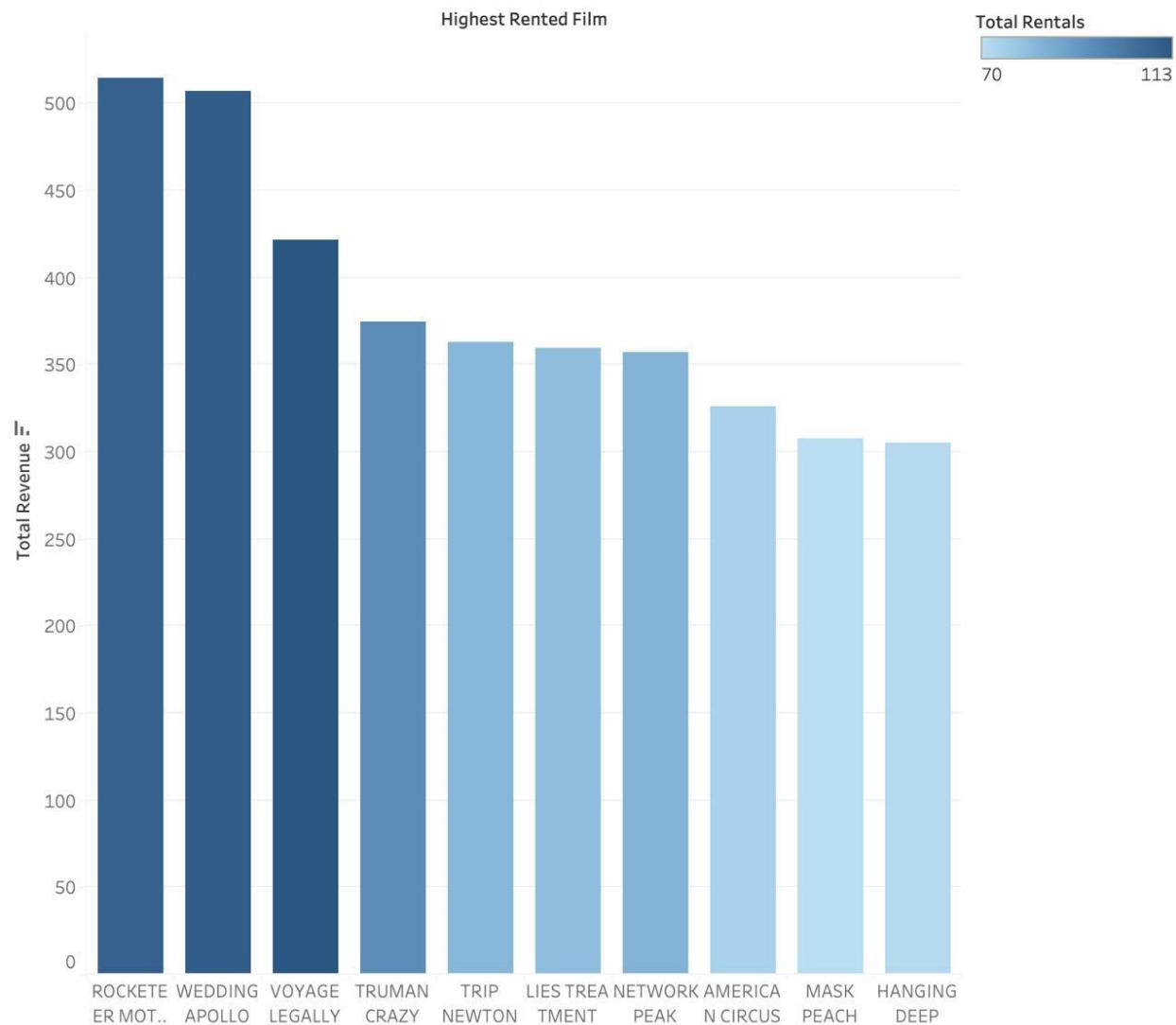


Visual 3: World Map

Business Description:

This world map shows each country's **most rented genre** (by color) and the **associated revenue** (size or shade). By revealing regional preferences and income levels, management can **tailor content offerings** to local tastes, **focus marketing efforts** on high-performing regions, and **expand** into areas with untapped potential.

Top Movies by Revenue



Sum of Total Revenue for each Highest Rented Film. Color shows sum of Total Rentals. The data is filtered on Customer Country, which keeps 108 of 108 members. The view is filtered on Highest Rented Film, which keeps 10 of 428 members.

Visual 4: Top Movie Revenue

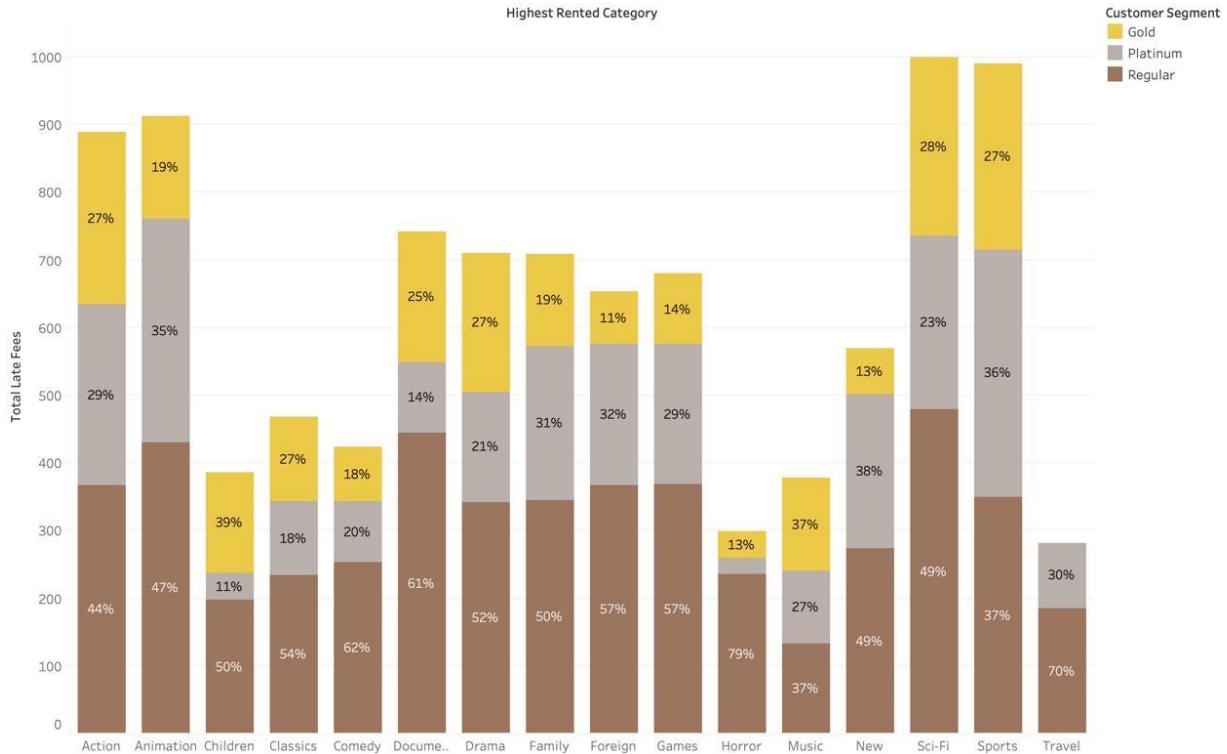
Business Description:

This bar chart ranks the **top movies** by their **total revenue** (bar height) and also shows **total rentals** (color scale). It helps management **pinpoint the most profitable titles**, understand which films drive high engagement, and **optimize marketing or inventory** around these top performers.

Visual 5:

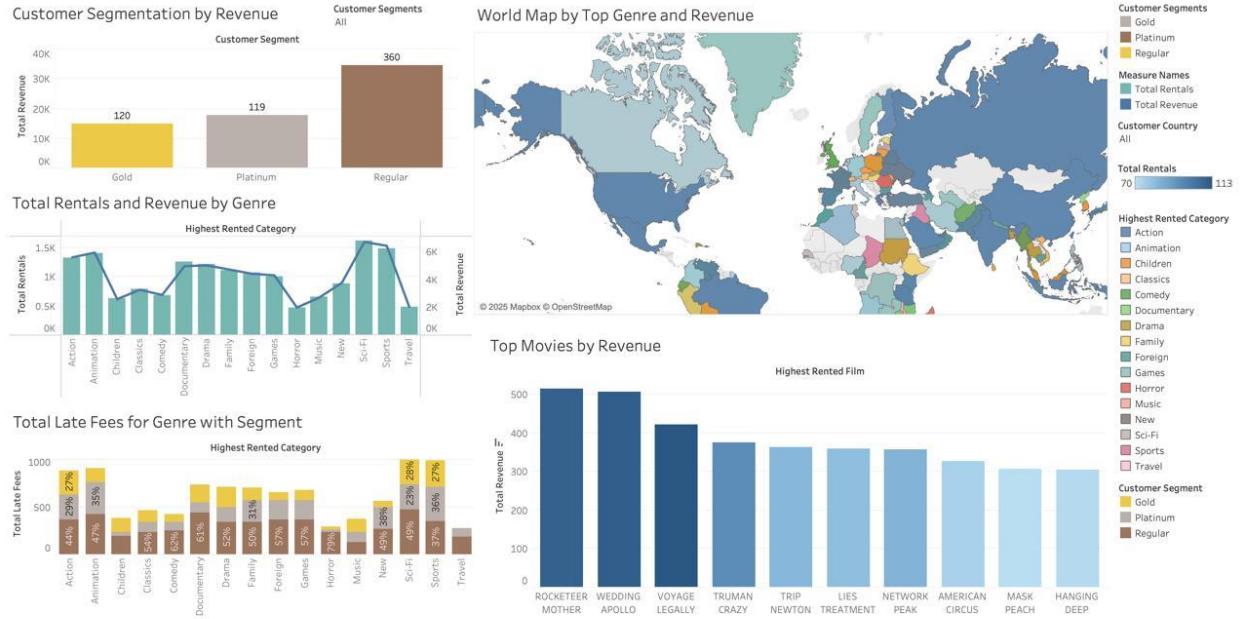
Business Description:

Total Late Fees for Genre with Segment



Sum of Total Late Fees for each Highest Rented Category. Color shows details about Customer Segment. The marks are labeled by % of Total Total Revenue. The data is filtered on Customer Country, which keeps 108 of 108 members.

This stacked bar chart displays total late fees across different film genres, segmented by Gold, Platinum, and Regular customers. It reveals which genres incur the most late fees and which customer tiers are most prone to late returns, enabling management to refine rental policies, target reminders, or offer incentives that reduce overdue rentals and boost customer satisfaction.



Dashboard: We added two filters for Customer segment and Country filter.

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.