SQL Project Report – DVD Rental Store Analysis

1. Project Overview

2. Tools and Techniques

• Tool: MySQL
• Techniques: CTEs, Joins, Aggregation, Window Functions
• Dataset: Sakila Database (1,000 films, 599 customers, 16 categories, 16,044 rental transactions)

3. Business Problem

You are tasked with performing Exploratory Data Analysis (EDA) for a DVD Rental Store using the Sakila database. The goal is to gain insights into customer behavior, film popularity, rental trends, and revenue generation through SQL queries. The analysis aims to support decision-making processes to enhance customer experience, optimize inventory, and increase profitability.

The objectives of this analysis are:

1. Identify the **most rented films** and their categories.
2. Determine the **top customers** based on rental frequency and total payments.
3. Analyze **rental trends** over time, including busiest rental months and days.
4. Identify the **most profitable film categories** based on total revenue.
5. Evaluate **staff performance** by analyzing revenue generated per staff member.
6. Locate the **best-performing stores** based on revenue.

4. SQL Approach

Collected and explored the Sakila dataset, cleaned and validated data for accuracy, applied SQL techniques to solve key business problems, and derived insights to guide informed decisions.

1.DATA GATHERING

Imported the sakila-schema.sql to create the database structure and loaded transactional data from sakila-data.sql into MySQL. This established the foundational dataset used for analysis.

2.DATA UNDERSTANDING

To gain familiarity with the Sakila database, each table was explored to understand its structure, key attributes, and analytical relevance. The goal was to identify tables required for rental, revenue, and customer behavior analysis.
Key Observations:

1. **Actor:** Contains details of 200 actors including their IDs and names; supports linking films to cast members.
2. **Address:** Includes customer and staff addresses; not directly used in analysis.
3. **Category:** Defines 16 film categories, used to classify movies by genre.
4. **Customer:** Contains 599 customers, their store associations, and activity status.
   84 customers are active, while 15 are inactive.
5. **Film:** Core dataset with 1,000 films, including attributes like rental duration, rate, and replacement cost.
   Analysis of rental duration showed:
   6-day rentals are most common (21.2% of films)
   3–4-day rentals form roughly 40% combined, showing balanced duration distribution.
6. **Film_Actor / Film_Category:** Bridge tables linking films with actors and categories, essential for join operations.
7. **Inventory:** Lists available film copies at stores.
8. **Rental:** Contains 16,044 transactions with rental and return timestamps — key for trend and revenue analysis.
9. **Staff / Store:** Contain employee and branch information for performance-level insights.

Summary Insight:
The Sakila database offers a comprehensive relational structure connecting **customers, films**, and **transactions.** For this analysis, the most critical tables identified were **rental, customer, film, and payment,** supported by category and inventory for contextual analysis.

To ensure data accuracy, null value checks were performed on key tables — customer and rental. No null values were found in either table, confirming that both customer and rental datasets are complete and reliable for analysis.

3. DATA CLEANING, VALIDATION, AND PREPARATION

Data cleaning was performed to ensure data accuracy, consistency, and readiness for analysis. The following checks were conducted on key tables within the Sakila database:

Duplicate Check:
Used the ROW_NUMBER() function to detect duplicate records in the rental and film tables.
No duplicate rows were found, confirming data uniqueness and integrity.

Customer Activity Validation:
Compared total customers (599) with distinct customers in the rental table (599).
Result confirmed that all customers had at least one rental record.
Some inactive customers were identified, likely due to past rentals but no recent activity.

Table Consistency Check:
Verified record alignment between film and film_text tables using count comparison.
Both tables contained equal records, ensuring synchronization between film metadata and descriptions.

4.PROBLEM SOLVING

1.Identify the **most rented films** and their categories.

I. BASED ON FILMS

Table: Top 10 Films by Rental Count

Rank	Film Title	Rental Count
1	BUCKET BROTHERHOOD	34
2	ROCKETEER MOTHER	33
3	FORWARD TEMPLE	32
4	GRIT CLOCKWORK	32
5	JUGGLER HARDLY	32
6	RIDGEMONT SUBMARINE	32
7	SCALAWAG DUCK	32
8	HOBBIT ALIEN	31
9	RUSH GOODFELLAS	31
10	TIMBERLAND SKY	31

Insight:
Rentals are evenly distributed among the top films (31–34 each), showing balanced audience preferences without extreme bias toward a single title.

II. RENTALS BY GENRE

category	rental_count	rental_percentage
Action	1112	6.9
Animation	1166	7.3
Children	945	5.9
Classics	939	5.9
Comedy	941	5.9
Documentary	1050	6.5
Drama	1060	6.6

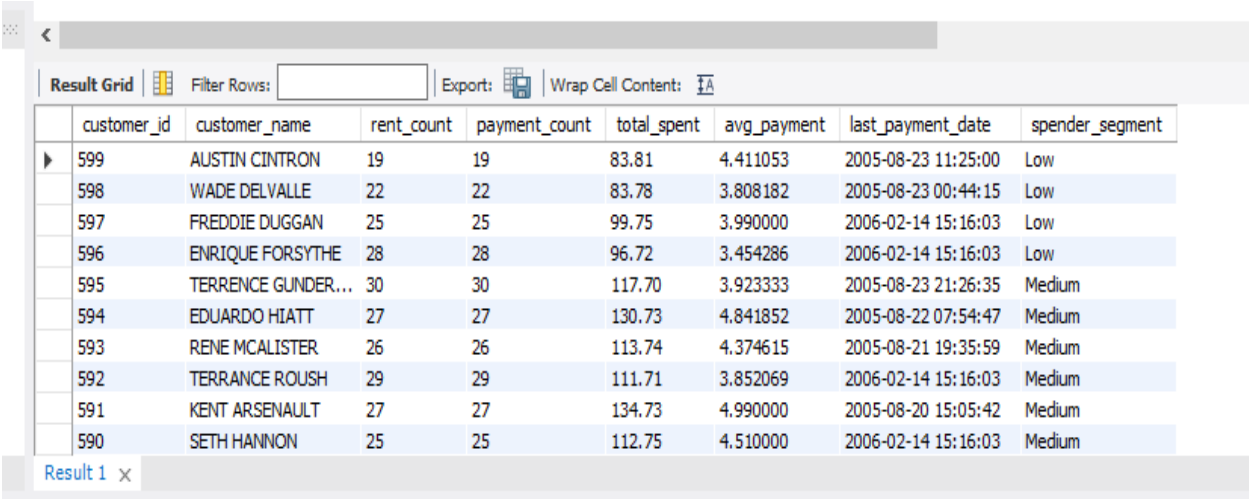
Family	1096	6.8
Foreign	1033	6.4
Games	969	6
Horror	846	5.3
Music	830	5.2
New	940	5.9
Sci-Fi	1101	6.9
Sports	1179	7.3
Travel	837	5.2

INSIGHT:
Sports, Animation, and Action lead with around 7% of total rentals each,
while the top five genres together capture over 35% of demand—showing audiences prefer energetic and family-oriented content,
with steady but lower interest in niche genres like Horror, Travel, and Music.y entertainment.

2.Determine the top customers based on rental frequency and total payments.

Using an SQL query, we generated a list of customers classified into three spending behavior categories — **High, Medium, and Low** — based on their total payments and rental frequency.
For deeper analysis, we used Excel to summarize the count and percentage of customers in each spending category, as shown below.

result table demo:



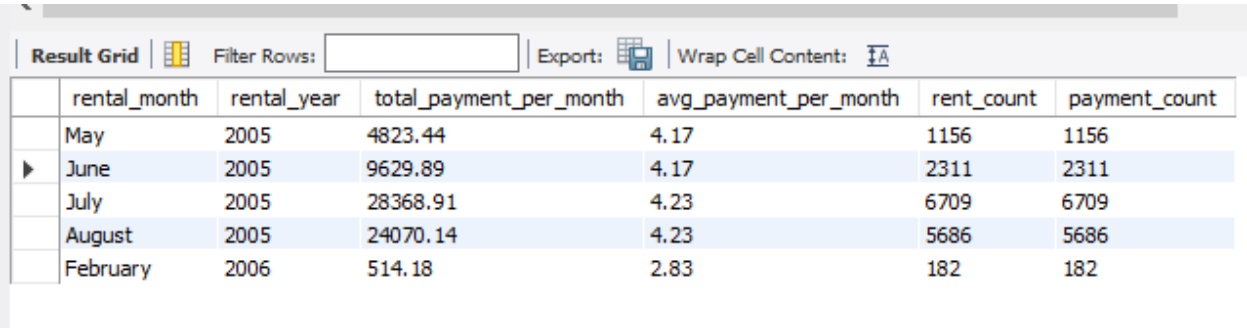
customer_id	customer_name	rent_count	payment_count	total_spent	avg_payment	last_payment_date	spender_segment
599	AUSTIN CINTRON	19	19	83.81	4.411053	2005-08-23 11:25:00	Low
598	WADE DELVALLE	22	22	83.78	3.808182	2005-08-23 00:44:15	Low
597	FREDDIE DUGGAN	25	25	99.75	3.990000	2006-02-14 15:16:03	Low
596	ENRIQUE FORSYTHE	28	28	96.72	3.454286	2006-02-14 15:16:03	Low
595	TERRENCE GUNDER...	30	30	117.70	3.923333	2005-08-23 21:26:35	Medium
594	EDUARDO HIATT	27	27	130.73	4.841852	2005-08-22 07:54:47	Medium
593	RENE MCALISTER	26	26	113.74	4.374615	2005-08-21 19:35:59	Medium
592	TERRANCE ROUSH	29	29	111.71	3.852069	2006-02-14 15:16:03	Medium
591	KENT ARSENAULT	27	27	134.73	4.990000	2005-08-20 15:05:42	Medium
590	SETH HANNON	25	25	112.75	4.510000	2006-02-14 15:16:03	Medium

Excel data analysis result:

Spending Category	Count of Customers	Percentage (%)
High	2	0.33 %
Low	204	34.06 %
Medium	393	65.61 %

Customer Insights (Summary):
Medium spenders (66%) show average rental and payment activity.
Low spenders (34%) have minimal engagement and can be targeted for reactivation.
High spenders (0.33%) are few but contribute significantly to revenue.
Action: Focus on retaining high spenders and upgrading medium customers through offers or loyalty programs.

3.Analyze **rental trends** over time, including busiest rental months and days.



rental_month	rental_year	total_payment_per_month	avg_payment_per_month	rent_count	payment_count
May	2005	4823.44	4.17	1156	1156
June	2005	9629.89	4.17	2311	2311
July	2005	28368.91	4.23	6709	6709
August	2005	24070.14	4.23	5686	5686
February	2006	514.18	2.83	182	182

Insight:
Rental activity peaked in July–August 2005 with the highest payments and rent counts, showing peak demand mid-year. Activity dropped sharply by February 2006, indicating an off-season trend.

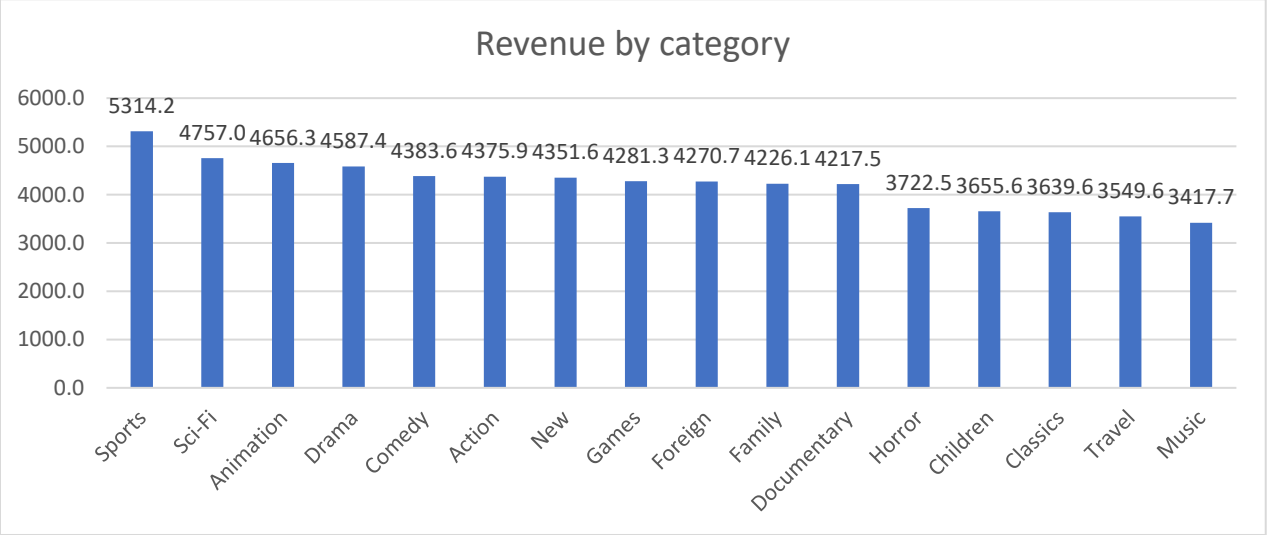
4. Identify the **most profitable film categories** based on total revenue

Total Revenue Generated: 67406.66

Insights:

Sports, Sci-Fi, and Animation lead with over 21% of total revenue.
Drama and Comedy show stable mid-range performance.
Music and Travel earn the least, indicating lower audience interest.

Query Output in graph format:



5. Evaluate **staff performance** by analyzing revenue generated per staff member

Query Output:

.

staff_id	staff_name	total_revenue	staff_rank
2	Jon Stephens	33,881.94	1
1	Mike Hillyer	33,524.62	2

Insight:

Jon Stephens achieved the highest total revenue (33,881.94), slightly surpassing Mike Hillyer (33,524.62), indicating both staff have nearly equal strong sales performance.

6. Evaluate Store **performance** by analyzing revenue generated.

Query output:

store_id	total_revenue	store_rank
2	33,881.94	1
1	33,524.62	2

Insight:

Store 2 leads with a total revenue of 33,881.94, slightly higher than Store 1's 33,524.62, showing nearly equal performance between both stores.

Final Insights & Suggestions:

- Data is clean and consistent across all tables.
- Sports, Sci-Fi, and Animation generate **21%+** of total revenue.
- Medium spenders (66%) dominate; high spenders drive major revenue.
- Rentals peak in **July–August**, lowest in **February**.
- Both stores and staff perform almost equally.
- **Suggestions:** Boost off-season rentals, target medium spenders, and promote top-performing genres for higher returns.