Tredence

Capstone-SQL

Note: This document is for educational purpose only. This is being provided for learners in this Data Core Tech Essentials for Supply Chain Practice and not meant for circulation outside Tredence.

Business Scenario:

The **Sakila DVD Rental Company** is a nationwide chain operating multiple stores across different cities. It offers a large collection of films in various categories and languages. Despite having a decent customer base, the company has noticed **stagnating revenues, rising inventory costs, and inconsistent staff performance** across locations.

Leadership suspects inefficiencies in customer engagement, poor inventory utilization, and missed cross-selling opportunities. They have tasked the Data Analysis team (you) with a **deep-dive data analysis** to uncover patterns, trends, and operational bottlenecks.

You've been provided access to the company's central transactional database — the **Sakila Database**, which includes detailed records of customers, films, rentals, inventory, payments, stores, staff, and geographic information.

Your task is to **generate data-driven insights** using SQL and present your findings in a **storytelling-style PowerPoint deck** to business stakeholders who are not technical. Your analysis should highlight key opportunities to improve profitability, streamline operations, and enhance customer retention.

Data Dictionary – Sakila Database

Customer & Location

Table Name	Column Name	Description
customer	customer_id	Unique ID for each customer
	store_id	Store to which the customer belongs
	first_name	Customer's first name
	last_name	Customer's last name
	email	Customer email
	address_id	Links to address table
	create_date	Customer account creation date
address	address_id	Unique address ID
	address	Street address

Table Name	Column Name	Description
	district	Region/District name
	city_id	Foreign key to city
city	city_id	Unique city ID
	city	Name of the city
	country_id	Foreign key to country
country	country_id	Unique country ID
	country	Country name

Film & Category

Table Name	Column Name	Description
film	film_id	Unique ID for each film
	title	Film title
	description	Film description
	release_year	Year of release
	language_id	Primary language ID
	rental_duration	Days a customer can keep the film
	rental_rate	Rental cost
	length	Film duration in minutes
	rating	Film rating (e.g., G, PG, R)

Table Name	Column Name	Description
	special_features	Extras included
category	category_id	Unique category ID
	name	Category name (e.g., Comedy, Action)
film_category	film_id	FK to film
	category_id	FK to category
language	language_id	Unique ID for language
	name	Language name

Inventory & Rental

Table Name	Column Name	Description
inventory	inventory_id	Unique ID for each copy of a film
	film_id	FK to film
	store_id	Store that holds the inventory
rental	rental_id	Unique rental transaction ID
	rental_date	Date and time the rental occurred
	inventory_id	FK to inventory
	customer_id	FK to customer
	return_date	When the rental was returned
	staff_id	Staff member who processed the rental

Payments & Staff

Table Name	Column Name	Description
payment	payment_id	Unique payment transaction ID
	customer_id	FK to customer
	staff_id	FK to staff
	rental_id	FK to rental
	amount	Amount paid
	payment_date	Date and time of payment
staff	staff_id	Unique staff member ID
	first_name	Staff first name
	last_name	Staff last name
	store_id	Store staff belongs to
	email	Staff email
store	store_id	Unique store ID
	manager_staff_id	Staff member managing the store
	address_id	Address of the store

Data Import Guide – MySQL Workbench

Objective:

Import the Sakila database schema and data into MySQL Workbench using the provided SQL files: `sakila-schema.sql` and `sakila-data.sql`.

Pre-requisites:

- MySQL Workbench installed
- MySQL Server running
- Access to sakila-schema.sql and sakila-data.sql files

Step-by-Step Instructions:

Step 1: Open MySQL Workbench and Connect

1. Launch MySQL Workbench.
2. Connect to your local MySQL server (usually localhost with root user).

Step 2: Create Sakila Schema

1. Open a new SQL tab.
2. Execute:

```
CREATE DATABASE sakila;
```

4. Right-click the sakila schema in the SCHEMAS panel and choose **Set as Default Schema**.

Step 3: Load Schema (Tables and Relationships)

1. Go to **File > Open SQL Script**.
2. Select the sakila-schema.sql file.
3. Execute the script using the lightning bolt icon or press Ctrl + Shift + Enter.

Step 4: Load Data (Insert Records)

1. Open sakila-data.sql via **File > Open SQL Script**.
2. Execute the script to populate tables with sample data.

Step 5: Verify the Load

1. Refresh the SCHEMAS panel.
2. Expand the sakila schema and check that tables and data are visible.

Troubleshooting:

- If foreign key errors occur during data import, make sure you executed the schema file first.
- As a workaround, you can disable foreign key checks temporarily:

```
SET FOREIGN_KEY_CHECKS = 0;

-- your import queries here

SET FOREIGN_KEY_CHECKS = 1;
```

Business Objectives of the Capstone Project:

Customer Segmentation & Behavior

- Identify **top revenue-generating customers** and what differentiates them.
- Classify customers into **payment and activity tiers** for personalized marketing.
- Analyze **customer churn** and **retention behavior**.
- Determine **repeat rental behavior** and customer lifecycle patterns.

Store & Inventory Analysis

- Compare store-level performance in terms of **revenue, inventory usage**, and **rental frequency**.
- Identify **underutilized inventory** and suggest reallocation opportunities.
- Highlight **best-performing films** by store and category to drive stocking decisions.
- Recommend **restocking and inventory planning** strategies based on demand.

Film Demand & Categorization

- Determine which **film categories and ratings** generate the most revenue and engagement.
- Analyze **rental frequency** across different **film lengths, languages, and ratings**.
- Identify **films and categories with low or no rental activity** for potential removal.
- Track **seasonal and monthly trends** in film rentals.

Staff & Operational Performance

- Evaluate **staff contribution to revenue** and **customer handling efficiency**.
- Analyze **repeat customer interactions per staff** to track relationship-building.
- Compare **weekday vs. weekend performance** per staff and store.
- Identify **high performers** to model training and operations best practices.

Overall Business Trends & Recommendations

- Detect **monthly revenue trends** and **growth/decline rates**.
- Recommend **data-backed interventions** to improve rental conversion, retention, and profitability.
- Provide a **prioritized action plan** for the business to implement changes in marketing, operations, and stocking strategy.

Final Outcome:

You will deliver:

- A **PowerPoint presentation** with data-backed storytelling (8–10 slides).
- A **SQL query workbook** containing documented queries mapped to business questions.
- (Optional) A **Data Dictionary** explaining key fields, joins, and assumptions.

Customer Segmentation & Behavior (10 Questions)

1. **Identify the top 10 customers who have paid the highest total rental fees across all stores. Include their customer ID, full name, email, and total payment amount.**
2. **Group customers into three payment tiers — Low (< 100), Medium (100–200), and High (> 200) — based on their total payments. Count how many customers fall into each tier.**
3. **Calculate the average number of rentals made per customer who joined in each month of the last year.**
4. **Identify customers who rented more than 10 times in the past 3 months and calculate the average payment amount for this group.**
5. **Determine the city with the highest average payment per customer, considering only customers with at least 5 rentals.**
6. **Calculate the number of active vs. inactive customers, where active customers are defined as those who rented at least once in the last 6 months.**
7. **Find the percentage of customers who returned to rent again within 30 days of their first rental.**
8. **List the top 5 cities with the most customers who have not rented anything in the last 12 months.**
9. **Identify customers who rented the same film more than once and calculate how many such repeat rentals exist per customer.**
10. **Find the average time gap (in days) between successive rentals for each customer and rank the top 10 with the shortest average gap.**

Store & Inventory Analysis (10 Questions)

11. **Determine which store has generated the highest total revenue and provide the breakdown of revenue per staff member for that store.**
12. **List the top 5 rented films per store based on rental count. Show store ID, film title, and number of rentals.**
13. **Calculate the number and percentage of inventory items per store that have never been rented.**
14. **Find the film with the longest average rental duration per store and show the average in days.**
15. **Identify the most common rental day of the week for each store.**
16. **Which store has the lowest average rentals per inventory item? Show store ID, total inventory, total rentals, and average rentals per item.**
17. **Find the average number of films available per category per store.**
18. **List all stores with more than 50 films that haven't been rented in the last 6 months.**
19. **Determine the monthly rental growth or decline (in percentage) for each store over the last 6 months.**
20. **Which store has the highest customer repeat rate (customers with 2+ rentals from the same store)?**

Film Demand & Categorization (10 Questions)

21. **Identify the top 5 film categories with the highest total revenue. Show category name and revenue.**
22. **For each film rating (G, PG, PG-13, R, NC-17), calculate the average rental frequency and average rental duration.**
23. **Find the 10 most frequently rented films in the past year. Include film title, rental count, and total revenue.**
24. **Which actors appear in the highest number of top 50 rented films?**
25. **Determine the rental conversion rate per film — number of rentals divided by number of inventory copies available. List top and bottom 5 films.**
26. **List all films that have never been rented, along with their category and language.**
27. **Which film categories have the highest average revenue per rental transaction?**
28. **Find films that were rented more than 5 times in a single day — list film title, date, and number of rentals.**
29. **What percentage of rentals were of films longer than 120 minutes?**

30. **Determine the correlation between film rental rate and actual rental count. Group films into rental rate buckets and compute rental averages.**

Staff & Operational Performance (10 Questions)

31. **Identify the staff member who generated the most revenue in the last 6 months. Include their ID, name, and total amount.**
32. **For each staff member, calculate the average number of rentals processed per month.**
33. **Find the staff member with the most active customers (customers with more than 3 rentals handled by the same staff).**
34. **Determine the percentage of total revenue generated by each staff member. Present as percentage share.**
35. **Calculate the average payment amount processed per transaction by each staff member.**
36. **List the staff members who processed rentals for the highest number of unique customers.**
37. **What is the most common film category rented per staff member?**
38. **Which staff member handled the most rentals on weekends (Saturday and Sunday)?**
39. **Compare average revenue generated by staff during weekdays vs. weekends.**
40. **Find the average delay between rental and return per staff member and identify the staff with the lowest average delay.**