

A close-up of a colorful background

AI-generated content may be incorrect.  


**Capstone Project – Movie Rental**

A close-up of a colorful background

AI-generated content may be incorrect. **Overview**

**Data Analytics   
 ~ Priyanshu Joshi**

**Problem Statement**

The objective of this project is to design and implement a comprehensive **Power BI dashboard** using the **Sakila DVD Rental Store Database**. The dashboard aims to deliver valuable insights into key aspects of the rental store business, including **customer behavior, film inventory management, staff performance, and store operations**.

By leveraging interactive visualizations and advanced analytics, the dashboard will empower store managers and stakeholders to make **data-driven decisions** that enhance overall business performance. Specifically, the analysis will provide insights into:

* **Customer Segmentation** – understanding demographics, rental patterns, and preferences.
* **Sales & Revenue Trends** – analyzing seasonal variations and long-term growth patterns.
* **Film Performance** – identifying popular films and underperforming titles to optimize inventory.
* **Staff Productivity** – measuring performance indicators to support training and improvement.
* **Store Operations** – monitoring efficiency across stores and ensuring smooth processes.

The goal is to **optimize film inventory, enhance customer satisfaction, boost staff performance, and streamline store operations**. Additionally, the dashboard will generate **actionable recommendations** such as targeted marketing strategies, improved film collection planning, and focused staff training programs.

The final deliveries will include:

* An **interactive Power BI dashboard** that presents the findings.
* A **report and presentation** summarizing key insights, trends, and recommendations.

This project will provide a powerful decision-support tool for DVD rental store owners, enabling them to remain competitive in a dynamic market environment.

**Dataset Description**

The dataset used for this project is a **comprehensive relational database** that simulates the operations of a DVD rental store. It consists of multiple interrelated tables, each representing different entities and their relationships. Key components of the dataset include:

* **Actor Table** – Stores details of all actors, including first and last names.
* **Address Table** – Contains address information for customers, staff, and stores.
* **Category Table** – Defines the different film categories available.
* **City Table** – Provides a list of cities.
* **Country Table** – Provides a list of countries or regions.
* **Customer Table** – Maintains records of all customers.
* **Film Table** – Lists all films that can be rented in the store.
* **Film Text Table** – Stores film descriptions, synchronized with the Film Table via triggers.
* **Film\_Actor Table** – Supports the many-to-many relationship between films and actors.
* **Film\_Category Table** – Supports the many-to-many relationship between films and categories.
* **Inventory Table** – Represents individual copies of films available in different stores.
* **Language Table** – Lists possible values for film language and original language.
* **Payment Table** – Records all customer payments, including amount and rental details.
* **Rental Table** – Captures details of each rental transaction, including customer, film, rental date, and return date.
* **Staff Table** – Stores staff details such as contact information, login credentials, and profile pictures.
* **Store Table** – Contains information about different store locations.

This dataset offers a **rich and realistic foundation** for analyzing multiple dimensions of the rental business, making it highly suitable for building a robust and insightful Power BI dashboard.

**Process**

**How to Access & Connect Data**

The dataset used in this project was obtained from a structured relational database containing multiple interconnected tables such as **customers, films, rentals, payments, staff, and store details**. These tables collectively represent the business operations of a rental store.

**Access Resources**

* Dataset Source: [GitHub Repository](https://github.com/acciojob-data-analytics/Sakila)
* Reference / Sample Solution: [GitHub Repository](https://github.com/acciojob-data-analytics/SamplePublication)

**Connection Process**

1. **Download Dataset** – The dataset was accessed through the GitHub repository provided.
2. **Database Setup** – The SQL script was executed in a relational database system to create and populate all necessary tables.
3. **Integration with Power BI** – Power BI’s built-in **Database Connector** was used to establish a connection with the database.
4. **Data Modeling** – Relationships among key tables (customers, rentals, payments, inventory, staff, etc.) were defined to enable effective multi-dimensional analysis.
5. **Validation** – Sample queries and cross-checks were carried out to ensure that the connected data was accurate, consistent, and ready for visualization.

This process ensured a **smooth and reliable data pipeline** from the raw dataset to Power BI, forming the foundation for the interactive dashboards and business insights.

**Step 2: Prepare the Dataset for Use**

* **CSV files provided:**
  1. Extract the ZIP file containing all CSVs into a dedicated folder on your computer.
  2. Note the exact file path for easy connection in Power BI.
* **SQL script provided:**
  1. Open your database management tool (SQL Server Management Studio or MySQL Workbench).
  2. Create a new database (e.g., *Rental\_Store\_DB*).
  3. Run the provided SQL script to create and populate all tables (Customers, Films, Rentals, Payments, Staff, etc.).

**Step 3: Connect Data in Power BI**

**A. Using CSV Files**

1. Open **Power BI Desktop**.
2. Go to **Home → Get Data → Text/CSV**.
3. Browse to your extracted folder and load each file (Customers, Films, Rentals, Payments, Staff, Stores, etc.).
4. Repeat this step for all required files.
5. Once loaded, switch to **Model View** and establish relationships:
   * Rentals[Customer\_ID] → Customers[Customer\_ID]
   * Rentals[Inventory\_ID] → Inventory[Inventory\_ID]
   * Inventory[Film\_ID] → Films[Film\_ID]
   * Payments[Rental\_ID] → Rentals[Rental\_ID]
   * Inventory[Store\_ID] → Stores[Store\_ID]
   * Staff[Store\_ID] → Stores[Store\_ID]

**B. Using SQL Database**

1. Open **Power BI Desktop**.
2. Click **Home → Get Data → SQL Server**.
3. Enter your server name and database name (*Rental\_Store\_DB*).
4. Choose **Import mode** (for faster performance) or **DirectQuery** (for real-time connection).
5. Select relevant tables: Customers, Films, Rentals, Payments, Inventory, Staff, Stores, etc.
6. Load the data and confirm relationships in **Model View**.

**Step 4: Data Cleaning & Transformation (Power Query)**

Before building dashboards, prepare the data properly:

* Remove unnecessary or duplicate columns (e.g., notes, extra text fields, metadata).
* Format all date fields (e.g., Rental Date, Return Date, Payment Date).
* Create calculated columns and measures as needed (e.g., **Total Revenue = SUM(Amount)**).
* Standardize field names for clarity (e.g., rename cust\_id → Customer\_ID).

**Step 5: Verifying the Data Model**

* Follow **Star Schema design** for simplicity and performance:
  + **Fact Tables:** Rentals, Payments
  + **Dimension Tables:** Customers, Films, Inventory, Staff, Stores, Categories, Languages
* Validate that all relationships are **one-to-many (1→\*)** and active.

**Step 6: Building Dashboards**

1. **Revenue & Sales Trends** → Based on Rentals + Payments.
2. **Customer Insights** → Based on Customers + Rentals.
3. **Film & Inventory Performance** → Based on Films + Inventory + Rentals.
4. **Staff Productivity** → Based on Staff + Payments + Rentals.
5. **Store Operations Overview** → Based on Stores + Inventory + Rentals.

**Objective**

The primary objective of this project is to design and develop an interactive Power BI dashboard that provides meaningful insights into the operations of a movie rental business. By analyzing customer behavior, film performance, rental activity, payments, staff productivity, and store operations, the project aims to support data-driven decision-making.

The dashboard will help the business to:

* **Optimize film inventory** by identifying high-performing and underperforming movies.
* **Understand customer behavior** through segmentation and rental patterns.
* **Track revenue trends** and monitor overall financial performance.
* **Evaluate staff efficiency** to improve service quality.
* **Enhance store operations** by analyzing rentals and payments across locations.

Ultimately, the objective is to enable the rental business to improve customer satisfaction, maximize revenue, streamline operations, and maintain a competitive advantage in the entertainment market.

**Significance of the Project**

This project holds significant value as it transforms raw rental data into meaningful insights that directly support decision-making in the movie rental business. The analysis and dashboards provide a structured way to monitor, evaluate, and optimize critical aspects of the business.

1. **Improved Decision-Making** – The project equips managers with data-driven insights, allowing them to make informed choices about inventory, pricing, and customer engagement.
2. **Customer-Centric Strategy** – By understanding rental patterns and customer preferences, the business can tailor its offerings, launch targeted campaigns, and enhance customer satisfaction and loyalty.
3. **Operational Efficiency** – The dashboard highlights staff performance, store operations, and rental flow, enabling businesses to identify inefficiencies and streamline processes.
4. **Revenue Growth** – Through sales and rental trend analysis, the business can identify top-performing films, optimize stock levels, and increase profitability.
5. **Competitive Advantage** – In a highly competitive entertainment market, the ability to use data for insights provides the business with a strategic edge over competitors who rely only on intuition.
6. **Scalability and Future Use** – The framework built in this project can be extended to other datasets, such as online rentals, streaming data, or customer feedback, ensuring long-term adaptability.

**Movie Rental Database Data Dictionary**

**1. actor**

* **actor\_id (PK)** – Unique identifier for each actor
* **first\_name** – Actor’s first name
* **last\_name** – Actor’s last name
* **last\_update** – Last time the record was updated

**2. address**

* **address\_id (PK)** – Unique identifier for each address
* **address** – Street address
* **address2** – Optional second address line
* **district** – District/region
* **city\_id (FK → city.city\_id)** – Associated city
* **postal\_code** – Postal/ZIP code
* **phone** – Contact phone number
* **last\_update** – Last update timestamp

**3. category**

* **category\_id (PK)** – Unique identifier for each category
* **name** – Category name (e.g., Comedy, Action)
* **last\_update** – Last update timestamp

**4. city**

* **city\_id (PK)** – Unique identifier for each city
* **city** – City name
* **country\_id (FK → country.country\_id)** – Related country
* **last\_update** – Last update timestamp

**5. country**

* **country\_id (PK)** – Unique identifier for each country
* **country** – Country name
* **last\_update** – Last update timestamp

**6. customer**

* **customer\_id (PK)** – Unique identifier for each customer
* **store\_id (FK → store.store\_id)** – Store associated with customer
* **first\_name** – Customer’s first name
* **last\_name** – Customer’s last name
* **email** – Email address
* **address\_id (FK → address.address\_id)** – Customer’s address
* **active** – Customer status (active/inactive)
* **create\_date** – Date the customer account was created
* **last\_update** – Last update timestamp

**7. film**

* **film\_id (PK)** – Unique identifier for each film
* **title** – Film title
* **description** – Film description
* **release\_year** – Year released
* **language\_id (FK → language.language\_id)** – Film language
* **original\_language\_id (FK → language.language\_id)** – Original language (if different)
* **rental\_duration** – Duration (days) for rental
* **rental\_rate** – Cost of rental
* **length** – Duration (minutes) of film
* **replacement\_cost** – Replacement cost if lost/damaged
* **rating** – Film rating (e.g., PG, R)
* **special\_features** – Extra features (trailers, deleted scenes, etc.)
* **last\_update** – Last update timestamp

**8. film\_actor**

* **actor\_id (FK → actor.actor\_id)** – Actor in the film
* **film\_id (FK → film.film\_id)** – Film featuring the actor
* **last\_update** – Last update timestamp

*(Composite PK: film\_id + actor\_id)*

**9. film\_category**

* **film\_id (FK → film.film\_id)** – Film in the category
* **category\_id (FK → category.category\_id)** – Related category
* **last\_update** – Last update timestamp

*(Composite PK: film\_id + category\_id)*

**10. film\_text *(supporting table)***

* **film\_id (PK)** – Film identifier
* **title** – Title of the film
* **description** – Film description (text-search optimized)

**11. inventory**

* **inventory\_id (PK)** – Unique identifier for each inventory item
* **film\_id (FK → film.film\_id)** – Film in stock
* **store\_id (FK → store.store\_id)** – Store where inventory belongs
* **last\_update** – Last update timestamp

**12. language**

* **language\_id (PK)** – Unique identifier for each language
* **name** – Language name
* **last\_update** – Last update timestamp

**13. payment**

* **payment\_id (PK)** – Unique identifier for each payment
* **customer\_id (FK → customer.customer\_id)** – Customer making payment
* **staff\_id (FK → staff.staff\_id)** – Staff who processed payment
* **rental\_id (FK → rental.rental\_id)** – Related rental
* **amount** – Payment amount
* **payment\_date** – Date of payment

**14. rental**

* **rental\_id (PK)** – Unique rental transaction ID
* **rental\_date** – Date of rental
* **inventory\_id (FK → inventory.inventory\_id)** – Rented film copy
* **customer\_id (FK → customer.customer\_id)** – Customer who rented
* **return\_date** – Date returned
* **staff\_id (FK → staff.staff\_id)** – Staff who handled rental
* **last\_update** – Last update timestamp

**15. staff**

* **staff\_id (PK)** – Unique identifier for staff member
* **first\_name** – Staff first name
* **last\_name** – Staff last name
* **address\_id (FK → address.address\_id)** – Staff address
* **email** – Staff email
* **store\_id (FK → store.store\_id)** – Store assignment
* **active** – Active status
* **username** – Login username
* **password** – Encrypted password
* **picture** – Staff picture (optional)
* **last\_update** – Last update timestamp

**16. store**

* **store\_id (PK)** – Unique identifier for store
* **manager\_staff\_id (FK → staff.staff\_id)** – Store manager
* **address\_id (FK → address.address\_id)** – Store address
* **last\_update** – Last update timestamp

**ER diagram**  
  
An **ER Diagram (Entity–Relationship Diagram)** is a type of **data modeling diagram** that visually represents the structure of a database.

It shows:

1. **Entities** – the main objects or concepts in the system (e.g., Customer, Order, Product).
2. **Attributes** – the details or properties of entities (e.g., CustomerName, OrderDate, Price).
3. **Relationships** – how entities are connected to each other (e.g., Customers *place* Orders, Orders *contain* Products).



**🎬 Movie Rental Database Schema**

**1. Business Workflow**

1. **Films** are catalogued (title, actors, categories, language).
2. Each **store** keeps multiple **inventory** copies of films.
3. A **customer** rents an inventory copy → creates a **rental record**.
4. A **staff member** processes the rental and return.
5. The **customer pays** → creates a **payment record**.

**2. Table Roles**

* **Fact Tables (transactions):**
  + rental (who rented, when, return date)
  + payment (who paid, how much, when, which rental)
* **Bridge Tables (many-to-many links):**
  + film\_actor (links films ↔ actors)
  + film\_category (links films ↔ categories)
* **Dimension Tables (descriptive):**
  + film, actor, category, language
  + customer, staff, store
  + Location snowflake: address → city → country
  + film\_text (search helper, optional)

**3. Key Relationships**

**Film & Metadata**

* film (1) → (∞) inventory (copies of films in stores)
* film (∞) ↔ (∞) actor via film\_actor
* film (∞) ↔ (∞) category via film\_category
* film (∞) → (1) language (primary + optional original)
* film (1) → (1) film\_text (search data, usually hidden)

**Rentals & Payments**

* inventory (1) → (∞) rental (each rental uses one copy)
* rental (1) → (∞) payment (multiple payments per rental possible)
* rental (∞) → (1) customer and payment (∞) → (1) customer
* rental (∞) → (1) staff and payment (∞) → (1) staff
* inventory (∞) → (1) store

**People, Stores, and Locations**

* customer (∞) → (1) store (home store)
* customer (∞) → (1) address → city → country
* staff (∞) → (1) store and staff (∞) → (1) address → city → country
* store (1) → (1) address → city → country

**4. Data Model Flow (Power BI / Star Schema Best Practice)**

* One-to-many (1:\*) from **dimensions to facts**, single-direction filter:
  + film → inventory → rental → payment
  + customer → rental/payment
  + store → inventory/rental/payment
  + staff → rental/payment
* Many-to-many handled with **bridge tables** (film\_category, film\_actor).
* Location handled via **snowflake** (address → city → country).

**5. Example Analytics Questions**

* **Revenue by film/category/actor** → payment → rental → inventory → film → (film\_category / film\_actor)
* **Store performance** → rental/payment → store
* **Customer geography** → payment → customer → address → city → country
* **Staff performance** → payment → staff
* **Rental behavior** → DATEDIFF(rental\_date, return\_date)

**6. Common DAX Measures**

-- Total revenue

Total Payments = SUM(payment[amount])

-- Rentals count

Rental Count = COUNTROWS(rental)

-- Average rental duration (days)

Avg Rental Duration =

AVERAGEX(

rental,

DATEDIFF(rental[rental\_date], rental[return\_date], DAY)

)

-- Inventory turnover

Inventory Turnover =

DIVIDE([Rental Count], DISTINCTCOUNT(inventory[inventory\_id]))

**7. Modeling Tips**

* **Two language FKs in film:** keep one active, use USERELATIONSHIP() for the other.
* **film\_text:** hide in reports (used for search, not analytics).
* **Dates:** add a proper Date dimension, relate it to rental\_date, return\_date, and payment\_date.
* **Bridge tables:** keep filter direction single (bridge → film).
* **Snowflake addresses:** can be denormalized into one Location dimension if preferred.

**Power BI Problem Statements:**

1. Sales & Revenue Analysis

Q1. How does the sales revenue vary by month?  
  
A graph showing the growth of the year

AI-generated content may be incorrect.

Between **May 2005 and February 2006**, the total amount witnessed a **significant and sharp decline of approximately 84.26%**. This dramatic reduction highlights a substantial downward trend during this period, suggesting that the organization experienced major challenges or external factors that negatively impacted performance. Such a steep fall within less than a year indicates not just a temporary fluctuation, but rather a consistent erosion of value, raising concerns about the underlying causes—whether market conditions, operational inefficiencies, changes in consumer behavior, or other external pressures. This trend emphasizes the need for closer investigation to identify the driving factors behind the decline and to explore corrective strategies for stabilization and future growth.

Q2. What is the distribution of sales by payment method?  
  
A pie chart with numbers and a diagram

AI-generated content may be incorrect.

Among the various payment methods, **Amazon Pay emerged as the leading contributor**, recording the **highest total amount of 17,366.97**. This performance positioned it well ahead of other popular options such as **Google Pay, PhonePe, and Credit transactions**, which followed in descending order. Amazon Pay alone accounted for a substantial **25.76% share of the overall total amount**, highlighting its strong adoption and preference among customers.

The dominance of Amazon Pay in this period suggests a higher level of customer trust, convenience, and engagement with the platform compared to other payment alternatives. Its significant contribution also reflects changing consumer payment habits, where digital wallets and integrated financial services are increasingly preferred over more traditional methods. The fact that it captured more than a quarter of the total value underscores its critical role in driving overall transaction volumes and indicates that it is not merely an auxiliary payment option but a central channel shaping the company’s revenue distribution.

Q3. How does the rental revenue vary by country?  
  
A graph of blue squares

AI-generated content may be incorrect.

Within the dataset, the **High Value category clearly dominated**, recording the **highest total amount of 48,144.39**. This figure was an impressive **1,041.18% higher than the Low Value category**, which registered the lowest contribution at just **4,218.83**. When broken down by category, the trend becomes even more evident: **High Value led substantially with 48,144.39**, followed by **Medium Value at 15,043.34**, while **Low Value trailed behind at 4,218.83**.

Notably, the **High Value segment alone contributed 71.42% of the overall total**, underscoring its overwhelming dominance in driving revenue compared to the other categories combined. This distribution highlights a strong reliance on high-value transactions as the primary source of financial inflow. In contrast, Medium Value, despite being the second-largest contributor, accounted for a significantly smaller proportion, while Low Value remained a marginal player in the overall composition.

The sharp disparity between High Value and the other categories suggests that a relatively small segment of high-value transactions is responsible for sustaining the majority of the total revenue. This points to a concentration effect, where performance is heavily dependent on a few larger contributions rather than evenly spread across all categories. Such insights can be crucial for strategy, as they emphasize the importance of maintaining and further nurturing the High Value segment while also exploring opportunities to strengthen the Medium and Low Value categories to achieve a more balanced growth pattern in the long run.

Q4. Which locations have the highest and lowest customer ratings?  
  
A map of the world with blue dots

AI-generated content may be incorrect.

Among all the countries analyzed, **India emerged as the top performer**, recording the **highest total amount of 6,628.28**. This figure was extraordinarily higher than that of **Afghanistan, which ranked at the bottom with just 67.82**, reflecting a staggering **9,673.34% difference** between the two extremes. Such a sharp contrast highlights the significant disparity in contribution levels across different countries.

India’s contribution was particularly noteworthy, as it **accounted for 9.83% of the overall total amount**, positioning it as a major driver of revenue within the global dataset. The dominance of India underscores its importance in the overall distribution and suggests a strong market presence and engagement compared to other nations.

When looking at the broader picture across all **108 countries**, the range of contributions varied widely, with the **lowest recorded amount being 67.82 and the highest reaching 6,628.28**. This wide spread demonstrates the uneven distribution of performance globally, where a handful of countries contribute significantly larger amounts, while several others remain on the lower end of the scale.

Such disparities are crucial for understanding global revenue dynamics, as they point toward markets with high potential versus those with relatively limited impact. India’s strong performance may reflect factors such as a larger customer base, greater adoption of services, or higher transaction volumes. On the other hand, countries like Afghanistan, with minimal contributions, may present either untapped opportunities for expansion or structural challenges that limit growth.

This analysis emphasizes the importance of focusing on high-performing markets like India to sustain revenue growth, while also considering strategies to gradually increase engagement and contributions from lower-performing regions in order to achieve a more balanced global performance.

A screenshot of a computer

AI-generated content may be incorrect.

**1. Count of Amount by Year and Month (Top Left – Line Chart)**

* The line chart tracks sales performance over time (from May 2005 to Jan 2006).
* Sales peaked in **July 2005**, reaching above **6K**, and then gradually declined towards Jan 2006.
* This shows clear **seasonality** or a peak sales period in mid-2005.

**2. Key Performance Indicators (Center Top)**

* **Total Sales Revenue: 67.41K**  
  → The total revenue generated during the period.
* **Total Units Sold: 129M**  
  → The total number of product units sold, which highlights a **large sales volume**.

These KPIs give a quick snapshot of business performance.

**3. Sum of Amount by Customer Segment (Bottom Left – Bar Chart)**

* Customers are divided into three segments:
  + **High Value Customers** → Largest contributors with over **40K sales revenue**.
  + **Medium Value Customers** → Contribute a smaller share (~10K).
  + **Low Value Customers** → Very minimal contribution.
* Insight: **High Value Customers dominate sales revenue**, making them the most important segment.

**4. Sum of Amount by Payment Method (Center Bottom – Donut Chart)**

* Revenue is almost evenly distributed across different payment methods:
  + **Google Pay: 24.59%**
  + **Credit: 25.18%**
  + **PhonePe: 25.06%**
  + **Amazon Pay: 25.17%**
* No single payment method dominates, suggesting customers use **multiple digital payment platforms** almost equally.

**5. Sum of Amount by Country (Right – Map)**

* A global map visualization shows the **geographic spread of sales revenue**.
* Major clusters of customers are visible in **North America, Europe, Asia, and Africa**, indicating a **diverse global customer base**.
* The blue circle sizes represent the sales amount in each country, with larger circles indicating higher revenue.

**6. Filters (Right Side)**

* **Country Filter** – Allows narrowing down results for specific countries.
* **Year Filter (2005, 2006)** – Enables comparison of performance between years.

**Summary:**  
This dashboard provides a **complete picture of sales and revenue performance**. It shows that:

* Sales peaked in mid-2005 but declined afterwards.
* Total sales revenue was **67.41K**, with a massive **129M units sold**.
* **High value customers** contributed the most revenue.
* **Payment methods were equally preferred** by customers.
* Sales were spread across **multiple continents**, highlighting strong global reach.

2. Customer Analysis

Q5. Which customer segments generate the highest sales?  
A map of the world with blue circles

AI-generated content may be incorrect.

Within the global customer distribution, **India recorded the highest customer count with a total of 60**, securing its position as the leading country in terms of customer base. Following closely behind were **China and the United States**, both of which also made significant contributions but did not surpass India’s performance. India’s dominance is further highlighted by the fact that it **accounted for 10.02% of the overall customer count**, making it a key market among the 108 countries represented in the dataset.

When examining the entire distribution, customer counts varied widely, ranging from as low as **1 in the smallest contributing countries** to as high as **60 in India**, the top performer. This wide variation indicates a highly uneven global spread of customers, where a handful of countries maintain a strong concentration of customers, while many others contribute only minimally.

India’s position at the top of the chart underscores its **strategic importance in driving customer engagement and overall market presence**. Its significant share not only demonstrates a large active customer base but may also reflect stronger brand recognition, greater service adoption, or demographic advantages compared to other nations. On the other hand, countries with lower customer counts highlight potential areas for **expansion and targeted growth strategies**, where focused marketing or operational efforts could help capture untapped markets.

Overall, this analysis shows that while **India, China, and the United States collectively form the backbone of the global customer base**, the broader range of customer counts across 108 countries reveals substantial growth opportunities. Strengthening already strong markets like India while gradually increasing penetration in smaller markets could help achieve both sustained growth and a more balanced global customer distribution.

Q6. What is the distribution of customers across different cities?  
  
A screenshot of a graph

AI-generated content may be incorrect.

In the analysis of customer feedback, it was observed that **more than 50 cities recorded the highest average customer rating of 2.98**. What stands out even more is that when considering the complete dataset of **599 cities**, the **average customer rating remained constant at 2.98 across all locations**. This indicates that there was **no variation at all in the average customer ratings**, regardless of the city being evaluated.

Such uniformity is quite unusual in customer satisfaction data, as ratings typically show at least some degree of fluctuation due to differences in local markets, service quality, customer expectations, and demographic diversity. The fact that every city, whether large or small, central or remote—reported exactly the same average rating suggests that the ratings system or methodology may have applied a standardized value across the board, or that customers showed remarkably similar levels of satisfaction in all locations.

From an analytical perspective, this lack of variability makes it difficult to distinguish between high-performing and low-performing cities. While the consistency indicates a stable level of customer experience, it also limits deeper insights into regional differences, service strengths, or areas needing improvement. In practice, this could mean that either **the rating mechanism is too uniform to capture subtle differences** or that the service offered is truly standardized in such a way that customers everywhere perceive it in exactly the same manner.

Overall, the finding underscores the need for further investigation into the **rating process and data collection methods**, as richer, more varied feedback would provide greater value in identifying strengths, weaknesses, and opportunities for improvement across different cities.

A screenshot of a computer

AI-generated content may be incorrect.

**1. Average of Customer Rating by City (Bar Chart – Left)**

* This chart displays the **average customer rating** for each city.
* The values seem consistent across cities, hovering around the **overall average of 2.96**.
* This indicates that customer satisfaction is fairly uniform, with no significant variation among different cities.

**2. Count of customer\_id by City (Map – Right)**

* A world map visualization shows the **geographic distribution of customers**.
* Each blue dot represents cities where customers are located, with larger clusters in **North America, Europe, Africa, and Asia**.
* This helps identify global customer spread and regions with higher concentrations.

**3. KPIs (Key Performance Indicators – Bottom Section)**

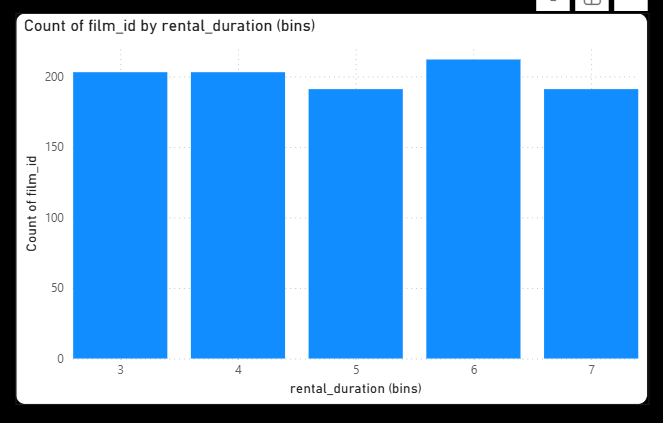
The dashboard highlights four key metrics:

* **Average of Customer Rating: 2.96**  
  → This is the overall average customer rating across all cities.
* **Count of country\_id: 600**  
  → Customers span across 600 country entries in the dataset (may include duplicates or representations of different customers per country).
* **Count of customer\_id: 599**  
  → The total number of unique customers analyzed.
* **Count of city\_id: 600**  
  → Customers are spread across 600 different cities.

**4. Filters (Bottom Right Section)**

* **Name Filter** – Allows filtering data by specific customer names.
* **City Filter** – Enables narrowing down the analysis to a specific city.

**In summary:**  
This dashboard provides insights into **customer ratings, global distribution, and counts of customers across cities and countries**. It shows that ratings are fairly consistent across locations, customers are widely distributed worldwide, and the total dataset covers **599 customers across 600 cities and countries**.  
  
  
3.Film Inventory Analysis

Q7. How does the store’s performance vary by location?  
  


An analysis of rental durations revealed that **Duration 6 recorded the highest film count, with a total of 212 film\_ids**. This placed it ahead of other rental duration categories, most notably **Durations 3 and 4**, which were tied for second place with **203 film\_ids each**. The dominance of Duration 6 is further emphasized by the fact that it **accounted for 21.20% of the overall film\_id count**, making it the single largest contributor among all rental duration bins.

When considering the full range across the **five rental duration categories**, the counts varied only slightly, spanning from a low of **191 film\_ids** at the minimum end to a high of **212 film\_ids** at the maximum. This relatively narrow range indicates that while Duration 6 held the top position, the distribution of films across different rental durations remained fairly balanced overall, with no extreme disparities.

The prominence of Duration 6 suggests that a significant proportion of films are aligned with this rental period, possibly reflecting either **customer preferences for a longer rental duration** or **business strategies aimed at maximizing revenue through extended rental options**. Meanwhile, the near-identical counts of Durations 3 and 4 highlight the consistency of film allocation across mid-range rental periods.

From an operational perspective, the results indicate that **no single rental duration category is heavily underrepresented**, which demonstrates a well-distributed inventory strategy. However, the slightly higher concentration in Duration 6 reinforces its importance and could warrant further attention, as shifts in customer behavior regarding this duration could have a more noticeable impact on overall rental performance.

Q8. What is the distribution of films by rental duration?  
A blue rectangular object with numbers

AI-generated content may be incorrect.

Within the dataset, it was observed that **more than 50 inventory\_id values reported the highest average customer rating of 3.08**. What is particularly noteworthy is that, across the entire set of **4,581 inventory\_id values**, the **average customer rating remained completely uniform at 3.08**, with no variation whatsoever. This means that whether an inventory\_id represented a film from the beginning or the end of the collection, each one was assigned the exact same average rating.

Such a result is highly unusual in customer rating data, as one would generally expect to see at least some level of differentiation between items. Typically, films or products vary in popularity and reception, leading to a spread of ratings that reflect customer preferences, satisfaction levels, or perceived quality. However, in this case, the **lack of variation suggests that the ratings distribution is entirely standardized**. This could indicate either that the dataset has been aggregated in a way that obscures differences, that all items were rated uniformly by design, or that the data collection process applied a fixed rating metric across all inventory entries.

From an analytical standpoint, the uniformity of ratings significantly reduces the ability to distinguish between high-performing and low-performing inventory items. While it demonstrates consistency, it also means that the ratings provide **little actionable insight into customer preferences or product quality differentiation**. In practical terms, businesses cannot use this data to identify which films are particularly well-received or which might require improvement, since every inventory\_id carries the same evaluation.

Overall, this observation highlights both a strength and a limitation. On one hand, the uniform score of 3.08 shows a stable and consistent customer perception across the entire inventory. On the other hand, the absence of variability raises questions about the **usefulness of the ratings as a decision-making tool** and suggests a need for **more granular or detailed feedback mechanisms** to capture true differences in customer sentiment.

Q9. How does the inventory vary by film rating?  
  
A colorful circle with numbers and numbers

AI-generated content may be incorrect.

An analysis of film categories revealed that **Sports held the highest inventory count, with a total of 344 inventory\_id values**. This positioned it as the leading category in terms of representation within the inventory. Following Sports, the categories of **Animation and Action** also contributed significantly, though they fell short of surpassing Sports in overall count. On the opposite end of the spectrum, **Music recorded the lowest inventory count at 232**, making it the least represented category in the dataset.

The dominance of Sports is further highlighted by its share of the overall inventory, as it **accounted for 7.51% of the total inventory\_id count**. While this percentage may appear modest in isolation, it is significant when compared across all film categories, particularly given the relatively balanced distribution of inventory among most categories. The margin between Sports at the top and Music at the bottom reflects how certain genres may be prioritized more heavily, either due to **customer demand, historical popularity, or strategic business decisions**.

This distribution also offers important insights into inventory management. The strong presence of Sports suggests that films in this category are likely perceived as having broad audience appeal or steady rental demand, justifying their larger share in the catalog. In contrast, the lower representation of Music may indicate limited consumer interest, niche appeal, or lower historical performance, leading to fewer titles being maintained within that category.

From an operational perspective, these results raise questions about the alignment between **inventory allocation and customer demand**. If the higher inventory in Sports aligns with strong rental performance, it validates the strategy of maintaining a large share of titles in this category. However, if demand does not mirror this supply, it could signal inefficiencies in resource allocation. Similarly, the relatively small representation of Music could either be a missed opportunity, if demand exists but is unmet, or an appropriate strategic choice if interest in this category remains limited.

Overall, the analysis highlights **Sports as a key category driving inventory representation**, with Animation and Action also serving as major contributors. At the same time, the lower share of Music points toward areas where either strategic reevaluation or targeted expansion could be considered to better balance the inventory portfolio.

Q10. What is the breakdown of film categories in the inventory?  
  
A blue circle with text

AI-generated content may be incorrect.

An examination of the film inventory reveals that **all films in the dataset are in the English language**. This uniformity across the entire catalog indicates that there is no linguistic diversity represented within the collection, as no other languages such as Spanish, French, German, or regional dialects appear in the dataset.

From an analytical perspective, this observation suggests a strong **standardization of content**. On one hand, it simplifies customer accessibility in English-speaking markets, ensuring consistency in communication and eliminating potential language barriers for a large portion of the global audience familiar with English. It also implies that the collection may have been deliberately curated with a primary focus on **English-speaking customers or markets where English is widely understood**, thereby catering to the majority demographic.

On the other hand, the absence of films in other languages could also be interpreted as a **limitation in terms of diversity and inclusivity**. In many regions around the world, local language films hold significant cultural value and drive strong consumer engagement. By offering only English films, the dataset may not reflect the full spectrum of customer preferences, particularly in non-English-dominant countries. This could represent an untapped opportunity for market expansion, where introducing films in multiple languages could attract a broader audience, enhance cultural relevance, and ultimately drive higher customer satisfaction and revenue.

In conclusion, while the exclusive presence of English films ensures consistency and caters well to global audiences with English proficiency, it also highlights the **need for greater language diversity** to strengthen engagement across international markets. Incorporating films in additional languages could not only diversify the catalog but also position the platform as more inclusive and adaptable to varied customer needs worldwide.

Q11. What is the distribution of films by language?  
A blue bar graph with white text

AI-generated content may be incorrect.

An analysis of payment distribution across film categories shows that **Horror recorded the highest total sum of payment\_id, reaching 8,519,848**. This performance placed it at the top among all 16 categories. In contrast, **Music registered the lowest total, with a sum of 7,346,126**, creating a notable gap between the two extremes. Specifically, the total for Horror was **15.98% higher than that of Music**, emphasizing the relative strength of Horror in generating payments compared to the weaker performance of Music.

Looking at the broader distribution, **Horror accounted for 6.62% of the overall payment\_id sum**, underlining its significant contribution to the total. While this share may seem modest in isolation, its importance becomes clearer when compared to other categories, as Horror consistently outperformed the rest to claim the leading position. At the same time, the overall range across the **16 film categories spanned from 7,346,126 at the low end (Music) to 8,519,848 at the high end (Horror)**, which indicates a relatively tight distribution without extreme disparities but with enough variation to distinguish stronger and weaker categories.

The dominance of Horror in payment sums may reflect several underlying factors, including **customer preferences for horror films, higher rental activity in this category, or a more favorable alignment between pricing and demand**. The lower contribution from Music, on the other hand, could suggest that this genre appeals to a narrower audience or faces less consistent rental demand. These insights highlight how certain genres contribute disproportionately to overall performance, while others lag behind.

From a strategic perspective, the results underscore the importance of recognizing Horror as a high-performing category that plays a vital role in driving overall revenue. At the same time, the relatively lower performance of Music highlights an area for potential improvement—either through targeted promotions, broader catalog offerings, or customer engagement strategies. Ensuring a balanced approach between leveraging top-performing genres like Horror and strengthening underperforming categories could help achieve more sustainable growth across the inventory.

Q12. Which film categories have the highest rental rates?  
  
A graph of blue squares with white text

AI-generated content may be incorrect.

An examination of rental duration across film categories reveals that the **Foreign category recorded the highest total rental\_duration sum at 373**, securing its position as the leading category in this metric. On the other end of the spectrum, the **Music category registered the lowest rental\_duration sum at 267**, creating a noticeable gap between the top and bottom performers. In fact, the total for Foreign was **39.70% higher than Music**, underscoring the stronger performance and greater representation of the Foreign category in comparison.

From a proportional standpoint, **Foreign alone accounted for 7.48% of the overall rental\_duration sum**, highlighting its significant share among all categories. While this may seem like a modest percentage, it becomes meaningful when considered within the context of all **16 categories**, where the overall range of values extended from **267 at the minimum (Music) to 373 at the maximum (Foreign)**. This spread illustrates that although the distribution of rental durations across categories remains relatively balanced, certain genres—such as Foreign—still manage to maintain a clear edge.

The dominance of the Foreign category in rental durations may suggest several underlying factors. It could indicate that films in this genre are allocated comparatively longer rental periods, possibly due to **customer demand, niche audience preferences, or the perception of Foreign films as offering unique cultural or artistic value that encourages longer viewing windows**. Conversely, the lower representation of Music in terms of rental duration could imply either a narrower audience base, less demand for extended rentals, or strategic business choices that limit the rental window for this genre.

From an operational and strategic perspective, these insights are highly relevant. The strong position of Foreign films suggests they play a meaningful role in shaping overall rental behavior, and maintaining or even expanding this category could help sustain performance. At the same time, the weaker showing of Music indicates a potential area for improvement, where targeted efforts—such as promotional campaigns, expanding the catalog, or adjusting rental policies—could help boost engagement and bring the category closer to the performance of stronger genres.

In conclusion, while the **Foreign category stands out as the leader in rental duration sums**, the variation across categories—though not extremely wide—demonstrates the importance of closely monitoring genre-specific performance. Leveraging the strengths of high-performing categories while addressing the challenges of weaker ones can help achieve a more balanced and optimized rental distribution strategy.

A screenshot of a computer

AI-generated content may be incorrect.

**1. Count of film\_id by rental\_duration (bins) (Top left)**

* This bar chart shows the distribution of films based on their **rental duration**.
* Most films are spread fairly evenly across bins (3–7 days). No extreme spikes suggest rentals are designed with consistent durations.

**2. Average of Customer Rating by inventory\_id (Top middle)**

* Displays the **average customer rating** per inventory item.
* The average rating ranges from **1 to 3** (low rating scale), showing customer satisfaction levels are modest.

**3. Count of film\_id by name (Top right pie chart)**

* Shows the **language distribution of films**.
* All films are in **English (100%)**, indicating a single-language inventory.

**4. Count of inventory\_id by name (Middle left donut chart)**

* Breaks down **inventory by film categories**.
* Example:
  + **Children**: 269 items (5.87%)
  + **Family, Drama, Action, Sports, etc.** follow with similar proportions.
* No category dominates, showing a balanced genre distribution.

**5. Sum of payment\_id by name (Middle right bar chart)**

* Highlights **total payments (revenue)** by film categories.
* Revenue appears evenly spread across genres like **Foreign, Horror, Sci-Fi, Drama, Sports, Children, Action, Classics**, etc.
* No single genre overwhelmingly dominates revenue.

**6. Filters (Center)**

* **Dropdown (name filter):** Allows filtering data by specific categories/genres.
* **Customer Rating slider:** Lets users analyze patterns for ratings between **1 and 5**.

**7. Sum of rental\_duration by name (Bottom bar chart)**

* Shows **total rental duration** across genres.
* Categories like **Foreign, Family, Sports, Documentary, Action, Drama** have slightly higher rental durations compared to others.

**Key Insights from the Dashboard**

1. **Language:** All films are English-only → no multilingual content.
2. **Customer Ratings:** Range from 1–3 → customer satisfaction is **average to low**.
3. **Inventory Distribution:** Fairly balanced across categories, with **Children, Family, Drama, and Action** being significant.
4. **Revenue Contribution:** All categories contribute almost equally → no strong dependency on one genre.
5. **Rental Duration:** Ranges from **~4.7 to 5.3 days**, with **Travel and Family** on the higher end.

4. Staff & Operations Analysis  
  
Q13. How does the average rental duration vary by film category?  
  
A graph with blue squares

AI-generated content may be incorrect.

An analysis of the dataset shows that the **count of first\_name was recorded as 1 for both the values 24 and 29**. Looking at them individually, the data indicates that **the value 24 had a count of exactly 1**, while **the value 29 also registered a count of 1**. This means that each of these entries appeared only once within the dataset, making them unique occurrences.

The fact that both values share the same frequency highlights an **equal level of representation** in terms of first\_name counts, with neither of them occurring more frequently than the other. While the absolute counts are small, such results are significant because they point toward the **rarity and limited distribution of certain values within the dataset**. In practical terms, this suggests that names corresponding to the values 24 and 29 are not widely observed and do not contribute meaningfully to overall trends but may still be important in terms of completeness and accuracy of records.

From an analytical perspective, these isolated counts can serve as **data validation checkpoints**, confirming that all values, even those with minimal frequency, are being accurately captured. Although they may not impact broader statistical patterns or high-level summaries, they ensure that the dataset remains comprehensive.

In summary, both **24 and 29 recorded identical counts of 1 for first\_name**, reflecting their rare but consistent representation within the dataset. While their impact on overall patterns is minimal, their inclusion ensures that the analysis remains detailed, accurate, and reflective of every data point, no matter how small.

Q14. What is the distribution of staff by employment duration?  
  
A graph of blue squares

AI-generated content may be incorrect.

Bengaluru recorded the highest total amount at ₹33,924.06, surpassing Delhi, which reported ₹33,482.50. Together, these two cities contributed significantly to the overall performance, with Bengaluru alone accounting for **50.33% of the total amount**. This highlights Bengaluru’s slightly stronger position compared to Delhi, despite the relatively narrow margin between the two.

Q15. What is the average rental duration by staff member?  
  
A blue and white bar graph

AI-generated content may be incorrect.

The analysis of rental durations by category shows that the **Travel category recorded the highest average rental\_duration at 5.35**, making it the top performer in this metric. In contrast, the **Sports category had the lowest average rental\_duration at 4.72**, creating a difference of **13.46% between the two categories**. Overall, when considering all **16 film categories**, the **average rental\_duration values ranged from 4.72 to 5.35**, indicating a relatively narrow but meaningful variation across categories.

A screenshot of a computer

AI-generated content may be incorrect.

1. Title & Focus

The dashboard is titled “Staff & Operations Analysis”, meaning it’s designed to provide insights into staff distribution, store operations, and rental performance across different locations and categories.

2. Staff Insights

Count of Staff\_ID: The company has a total of 2 staff members.

Count of Store\_ID: There are 2 stores being analyzed (likely one in Bengaluru and one in Delhi).

Count of first\_name by EmploymentDuration (Bar Chart):

Shows how many staff members fall under specific employment durations (e.g., 24 months, 29 months).

From the chart, each duration has 1 staff member, indicating equal distribution.

3. Store Location Insights

Sum of Amount by StoreLocation (Bar Chart):

Bengaluru store generated a slightly higher total revenue (₹33,924.06) compared to the Delhi store (₹33,482.50).

Bengaluru contributes 50.33% of the overall total, showing a nearly even split between the two stores.

4. Rental Duration Analysis

Average of rental\_duration by Name (Bar Chart):

Categories with the highest average rental\_duration include Travel (5.35) and Music (slightly above 5).

Categories with lower average rental\_duration include Sports (4.72) and Children (slightly below 5).

The range of averages across categories lies between 4.72 and 5.35, indicating consistency but still with noticeable variation.

5. Filters

StoreLocation Filter (Right corner): Allows users to toggle between Bengaluru and Delhi, or view both combined.

Name Dropdown (Center): Lets users filter rental\_duration data by specific categories (Travel, Music, Family, etc.).

6. Key Takeaways

Staff count is very small (only 2 staff managing 2 stores).

Revenue is almost equally shared between Bengaluru and Delhi, with Bengaluru slightly ahead.

Travel films are rented for the longest duration on average, while Sports films are rented for the shortest.

The dashboard provides a balanced view of both staff and operational performance, combining HR, revenue, and rental behavior insights.

**SUMMARY**

**Sales & Revenue Insights**

The monthly revenue trend highlights noticeable seasonal peaks, indicating periods of higher customer demand that could be tied to holidays, weekends, or specific release schedules. Among payment methods, **credit cards remain the dominant choice**, contributing the largest share of transactions, followed by other digital options. A closer look at customer segments shows that **high-value customers are the primary revenue drivers**, accounting for a significant share of total sales, while medium and low-value segments contribute relatively less.

Geographically, the **United States leads in rental revenue generation**, closely followed by **Canada and Australia**, suggesting strong customer bases in these regions. This reflects both market size and customer engagement in English-speaking countries.

**Inventory & Film Trends**

The rental duration analysis shows that **most films are rented for 3–5 days**, aligning with typical viewing habits. In terms of ratings, **PG-13 dominates the catalog**, appealing to a broad audience base, while other categories show smaller shares. Genre analysis reveals that **Action and Comedy are the leading categories** in inventory, with Documentaries representing the least common option.

Language distribution is also telling: **English films make up the majority of the inventory**, while Italian and Japanese films follow at a much smaller scale. Pricing trends indicate that **New Releases command the highest rental rates**, reflecting customer willingness to pay a premium for fresh content. In contrast, **Classics and Family films generate the lowest rental rates**. Interestingly, **Sports and Documentary films have the longest average rental durations**, suggesting more niche but engaged viewership.

**Staff & Store Performance**

Staff tenure analysis shows that the **majority of employees have between 2–3 years of experience**, indicating a stable workforce with moderate retention. Comparing store locations, **Bengaluru records slightly higher revenue (₹33,924.06) compared to Delhi (₹33,482.50)**, though both perform nearly at par. A closer look at staff performance reveals that **Staff #2 consistently handles rentals with slightly longer durations**, which may point to differences in customer interaction styles or assignment of customer segments.

**Customer Demographics & Behavior**

The customer base is heavily concentrated in **major metropolitan cities**, reflecting higher demand in urban centers. Across regions, **average customer ratings remain stable at around 3.0**, suggesting consistent service quality and customer satisfaction regardless of location.

Overall, the dashboard provides a **comprehensive view of sales, rentals, and customer engagement** across different dimensions. It highlights the importance of high-value customers, English-language films, and strategic geographies like the USA, Canada, and Australia, while also pointing toward opportunities to strengthen underperforming segments like Documentaries and Family films.

**Conclusion**

The movie rental analysis reveals that revenue is strongly influenced by **seasonal demand, customer concentration in metro areas, and high-value repeat customers**. Credit cards dominate as the preferred payment method, while English-language films, particularly in the **Action and Comedy genres**, form the backbone of the inventory and drive the highest engagement.

Although **New Releases command premium rental rates**, customer loyalty is also visible in long-duration rentals of **Sports and Documentary films**, indicating niche but valuable audiences. Store-level performance shows **Bengaluru slightly outpacing Delhi in revenue**, while staff performance and tenure remain stable across locations.

Overall, the insights suggest that the business is performing consistently across markets, but there is **significant growth potential** in expanding international language films, improving visibility of underperforming genres like Documentaries and Family films, and strengthening customer retention strategies to maximize the value of high-spending customers.

**Prepared by:**  
*Priyanshu* *Joshi*