Movie Rental Analytics

Data Analysis

Capstone Project

Tanu Shree

Project Overview

The objective of this project is to create a comprehensive Power BI dashboard using the Sakila DVD Rental Store Database, providing valuable insights into the rental store business. The analysis will focus on customer behavior, film inventory management, staff performance, and store operations. The goal is to enable data-driven decision-making and improve overall business performance. The Power BI dashboard will offer insights into customer segmentation, sales trends, film performance, staff productivity, and store revenue. The primary aim is to optimize film inventory, enhance customer satisfaction, improve staff performance, and streamline store operations. The dashboard will provide actionable recommendations for targeted marketing campaigns, film collection enhancements, and staff training initiatives to improve business performance. The final deliverables will include a report and presentation showcasing the dashboard's findings and recommendations. The Power BI dashboard will serve as a valuable tool for rental store owners to make informed decisions and achieve success in the competitive DVD rental market.

Objective

The objective of the Movie Rental Analytics project is to analyze historical movie rental data to extract actionable business insights. The project focuses on identifying trends in customer behavior, staff performance, film categories, rental revenue, and geographic distribution. Using Power BI, Excel, and SQL, the goal is to create a visually compelling and data-driven decision support system for stakeholders.

The Process

1. Data Acquisition from Sakila:

The dataset was sourced from the Sakila open-source database, accessed in .csv format and/or via MySQL. It contains transactional-level data regarding customers, films, rentals, payments, and staff.

---

2. Data Transformation & Cleaning:

Performed using Power Query in Power BI:

- Removed nulls, blanks, and duplicate records.

- Standardized date formats and currency.

- Merged related tables using keys for customer_id, film_id, rental_id, etc.

---

3. Tool Integration:

Connected Power BI to the MySQL Sakila database (or imported CSVs). Integrated:

- Excel: For pivot summaries and dashboard planning.

- SQL Workbench: For complex queries, joins, and views.

---

4. Problem Statement Definition:

Defined 15 business questions across categories such as:

- Revenue and Rentals

- Customer Demographics

- Staff & Store Performance

- Film and Inventory Analysis

---

5. Exploratory Data Analysis (EDA):

Conducted using SQL views and Power BI visuals. EDA helped identify:

- Peak rental months

- Most rented films

- Countries contributing highest revenue

- Customer segmentation by city/country

---

6. Insightful Power BI Dashboard Creation:

Built a dark-themed interactive dashboard segmented into:

- Home

- Customer Analysis

- Store & Staff

- Film & Category

- Inventory & Revenue

Each section answered 2-4 analytical questions visually using:

- Donut charts

- Bar/Column graphs

- Slicers

- Geo Maps

---

7. Documentation & PowerPoint Report:

- All business questions answered using visual storytelling.

- Added conclusions, key metrics, and trend observations.

- PPT + Word report created for submission and portfolio.

Data Dictionary

- **Customer: Details of all rental customers**

  - customer_id, first_name, last_name, email, address_id, active, store_id, create_date

- **Store: Information about physical store locations**

  - store_id, address_id, manager_staff_id

- **Address: Full address of customers and stores**

  - address_id, address, city_id, postal_code, phone

- **City / Country: Hierarchical location details**

  - city_id, city, country_id | country_id, country

- **Film: Movie data**

  - film_id, title, release_year, language_id, rental_duration, rental_rate, length, replacement_cost, rating

- **Category / Film_Category: Genre classification**

  - category_id, name | linking: film_id, category_id

- **Inventory: Film availability per store**

  - inventory_id, film_id, store_id

- **Rental: Rental transactions**

  - rental_id, rental_date, inventory_id, customer_id, return_date

- **Payment: Payment details for rentals**
  - **payment_id, customer_id, staff_id, rental_id, amount, payment_date**
- **Staff: Store employees**
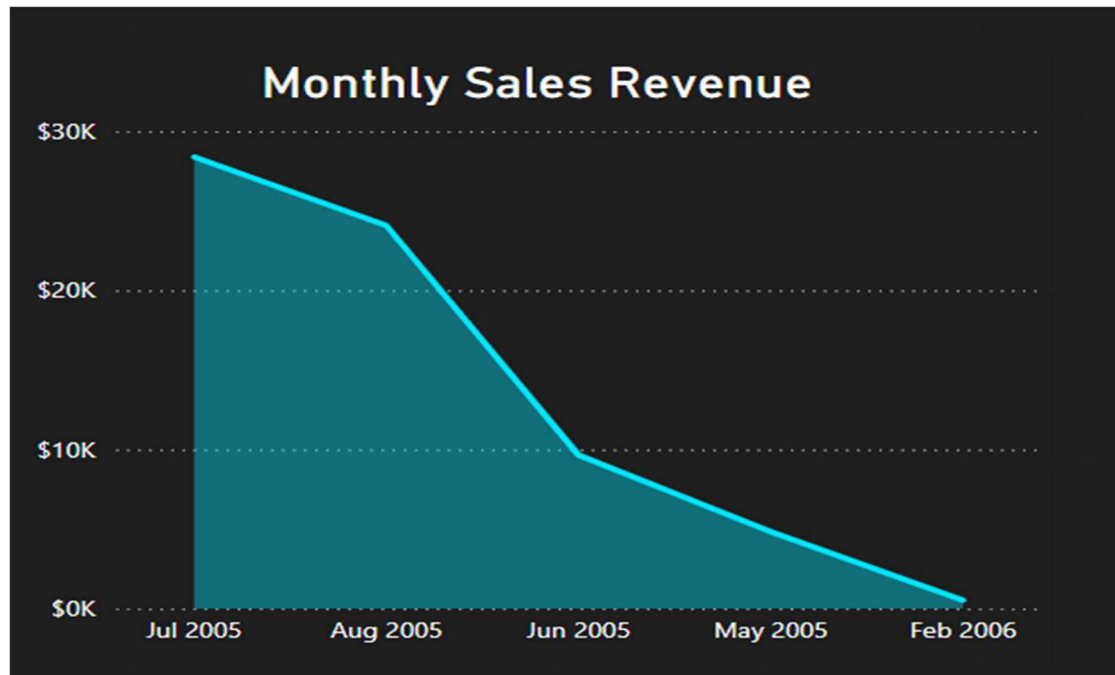 - **staff_id, first_name, last_name, address_id, store_id**

ER Diagram
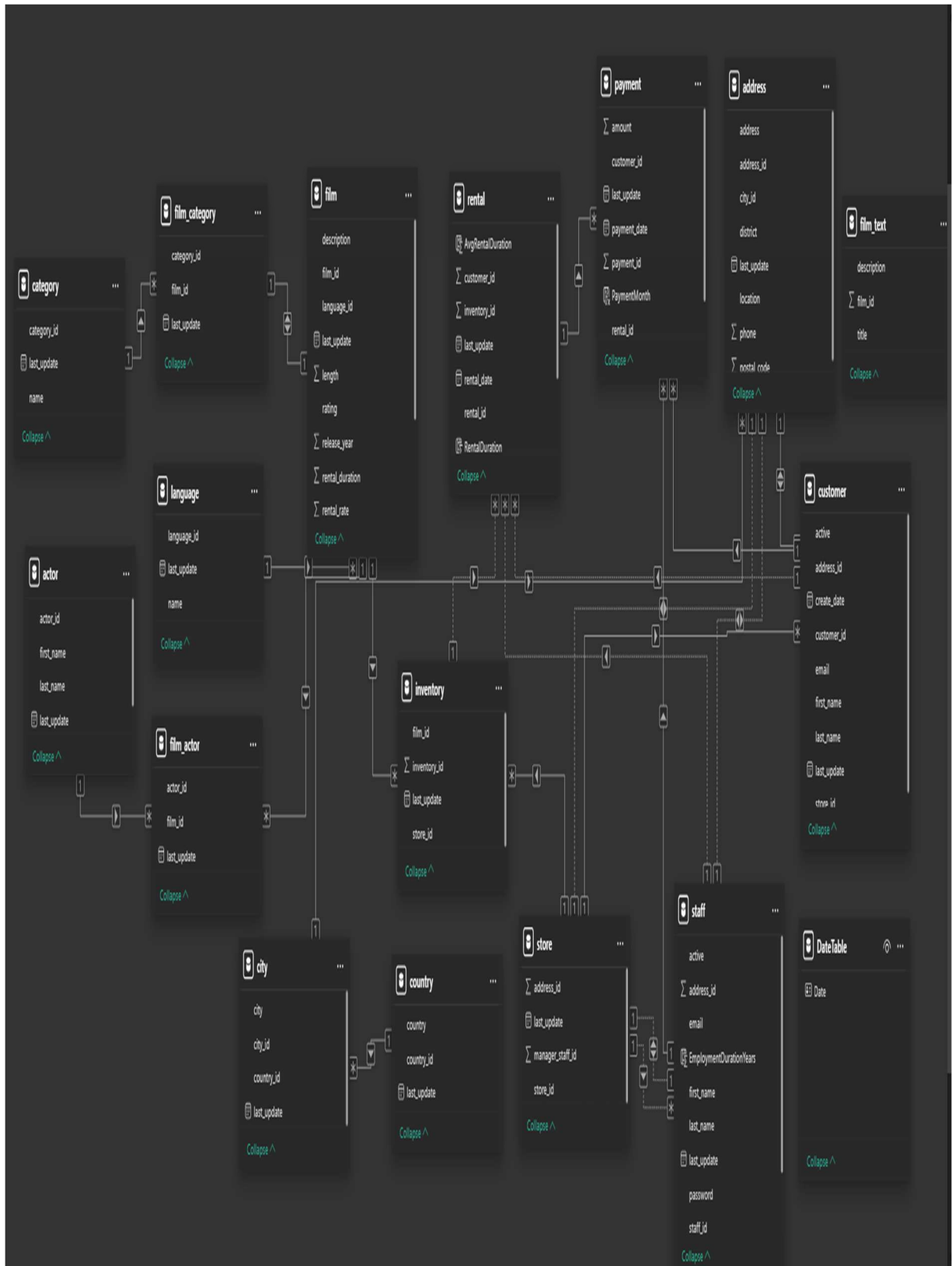
# Power BI Problem Statements

How does the sales revenue vary by month?

Conclusion:-

The monthly revenue trend reveals a **sharp peak in July 2005**, where sales reached approximately **$30K**, followed by a noticeable decline in subsequent months. August 2005 still recorded significant revenue, but from **June 2005 onward**, a continuous downward trajectory is observed, with the lowest revenue occurring in **February 2006**. This suggests strong seasonal demand in mid-year months, possibly due to summer holidays or promotional campaigns, while year-end months experienced reduced customer activity. Understanding this trend is crucial for planning marketing strategies and inventory during high and low seasons.

Which customer segments generate the highest sales?
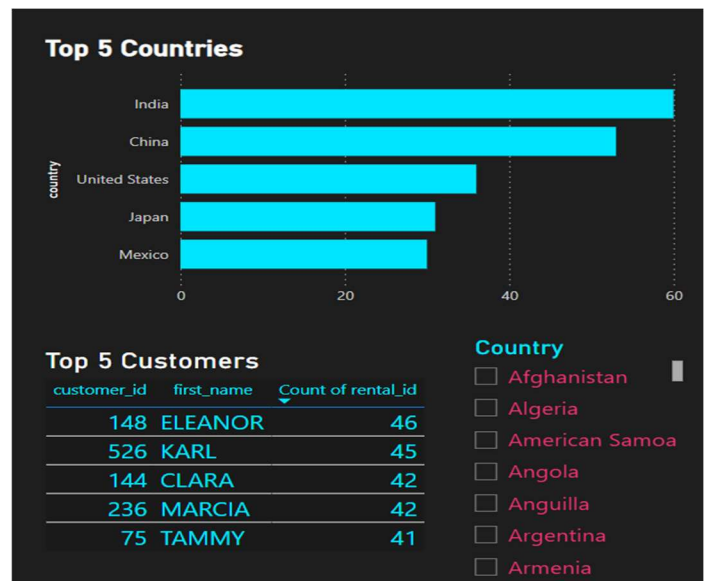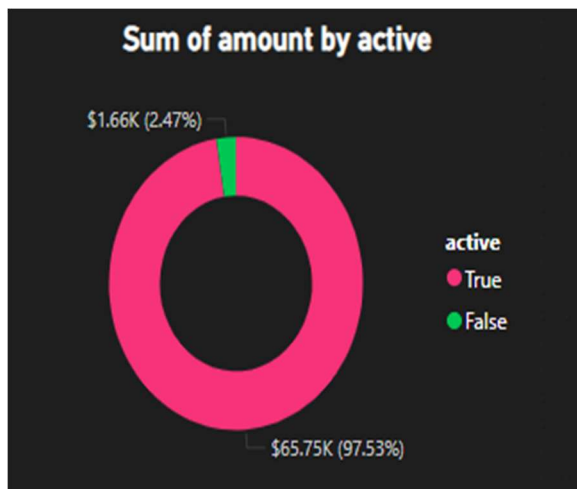
Conclusion:-

From the customer analytics dashboard, **India and China** emerged as the top-performing countries in terms of customer count and revenue contribution. **India leads with the highest number of active customers**, followed closely by China and the United States. At the individual level, **Eleanor, Karl, and Clara** are the most frequent renters, with Eleanor completing **46 rentals**, indicating high engagement. Notably, **97.53% of the total revenue** is generated by **active customers**, showing that customer engagement and retention directly impact business performance. This highlights the importance of nurturing top customers and focusing efforts in countries with high rental activity.

What is the distribution of films by rental duration?
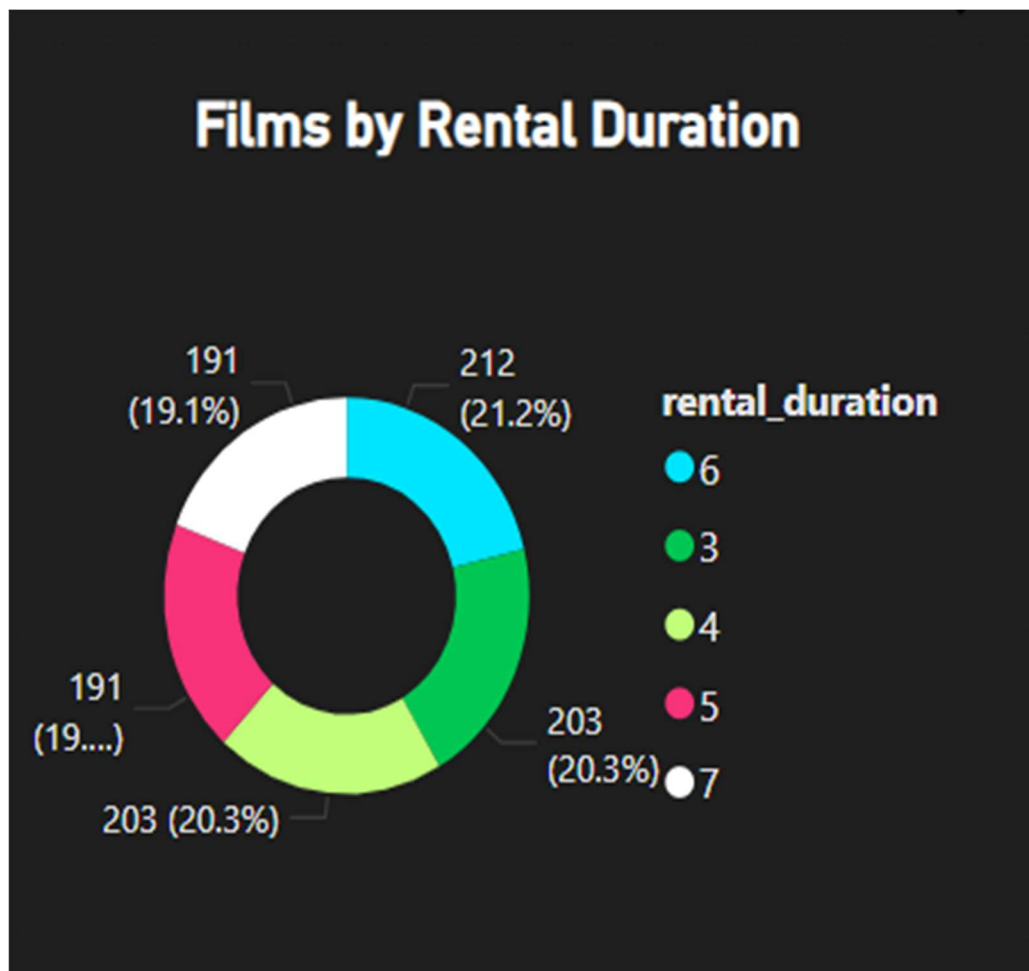
Conclusion:-

The film inventory reveals a diverse spread of rental durations, but a majority of films are rented for **5 to 7 days**, with the most common durations being **6 days (21.2%)**, **5 days (20.3%)**, and **7 days (19.1%)**. Shorter rental durations (e.g., 3–4 days) have significantly lower representation, indicating that customers prefer medium-term rentals that provide value without long-term commitment. This insight can guide decisions on pricing models and inventory stocking, ensuring more films fall into the preferred duration brackets.

How does the inventory vary by film rating?
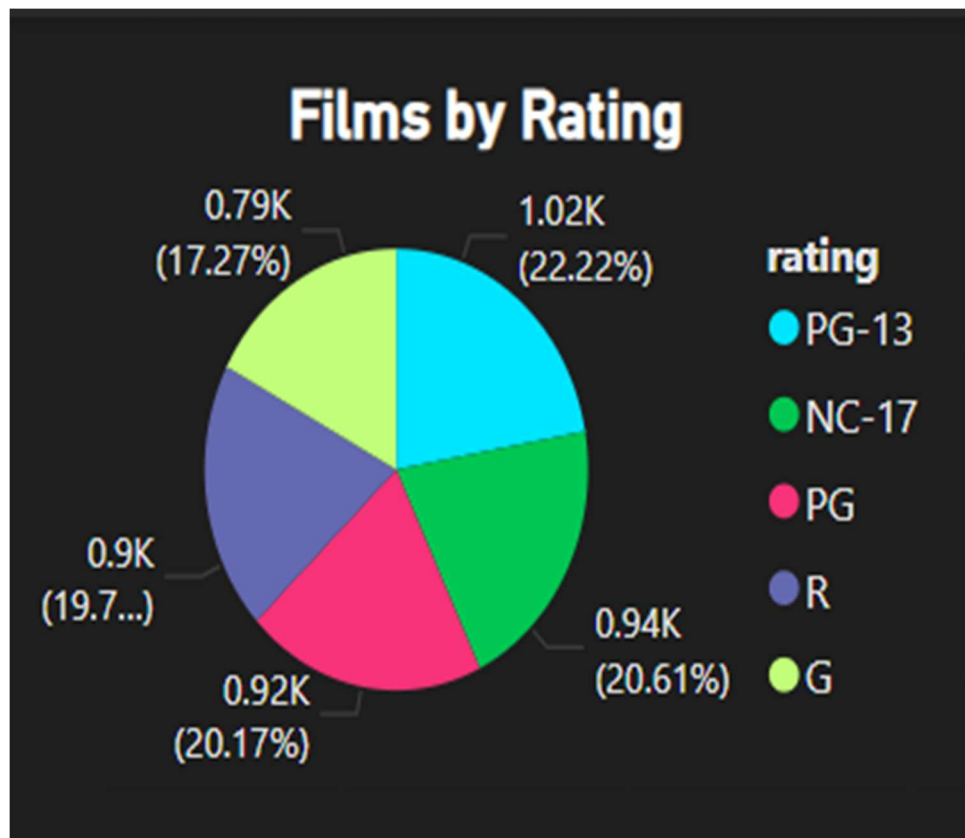
Conclusion:-

The inventory is evenly distributed across major film ratings, with **PG-13 (22.22%)**, **R (20.61%)**, and **PG (20.17%)** being the top three categories. These three ratings collectively account for over **60% of the inventory**, showing a strong tilt toward mature and family-friendly content. The presence of **G-rated and NC-17 films** is comparatively lower. This distribution reflects a balanced content strategy aimed at attracting both adult viewers and families, which is appropriate given the observed rental patterns and customer demographics.

What is the breakdown of film categories in the inventory?
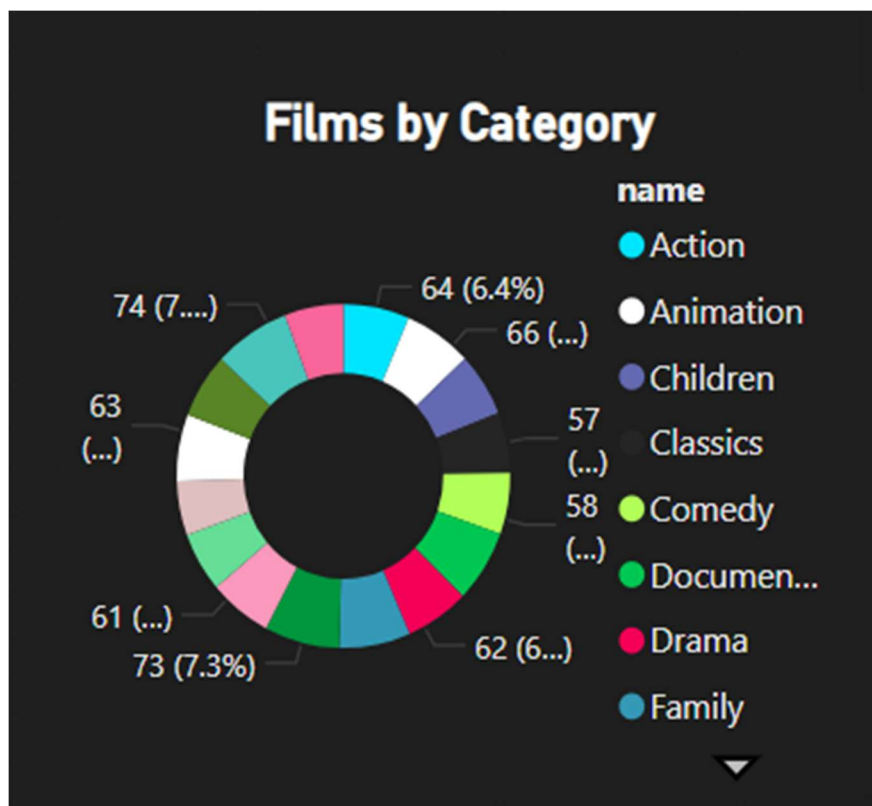
Conclusion:-

A category-wise breakdown reveals that **Action, Drama, and Comedy** are the leading genres in the inventory, both in terms of **volume and rental rates**. These categories also generate higher revenue and experience more frequent rentals, indicating strong viewer preference. Niche genres such as **Foreign films, Documentary, and Games** are present but contribute less to the overall rental share. This implies the business strategy aligns with popular genre demand, ensuring inventory is optimized for the most-watched categories, while still maintaining variety for niche audiences.
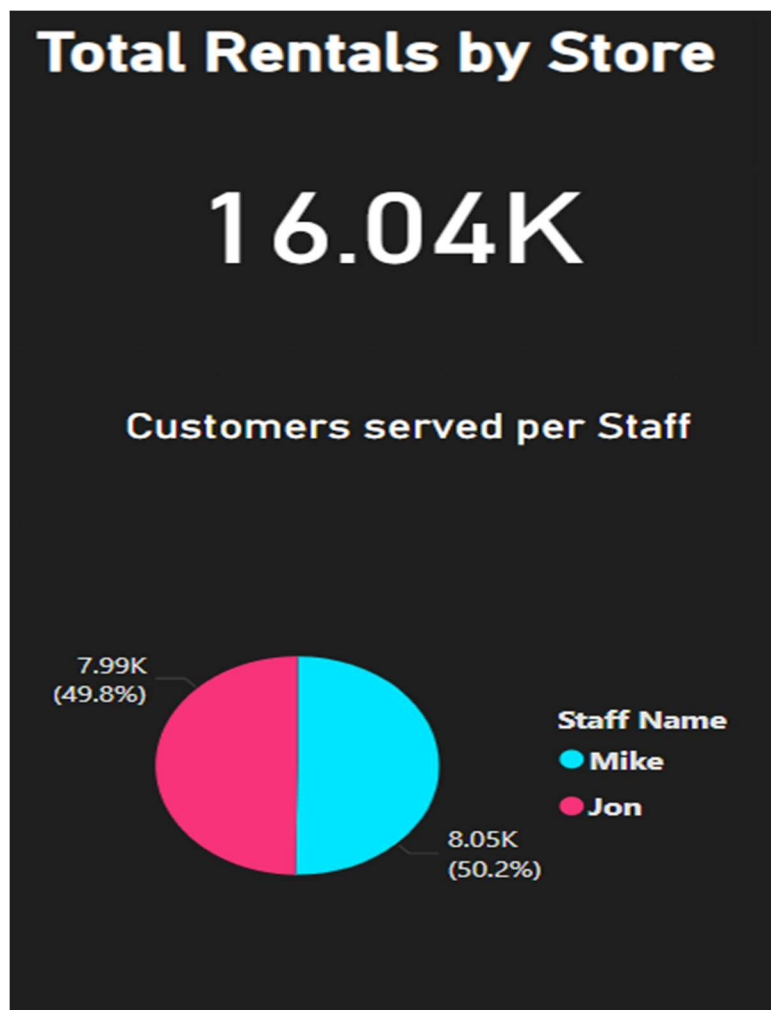
What is the distribution of staff by employment duration?
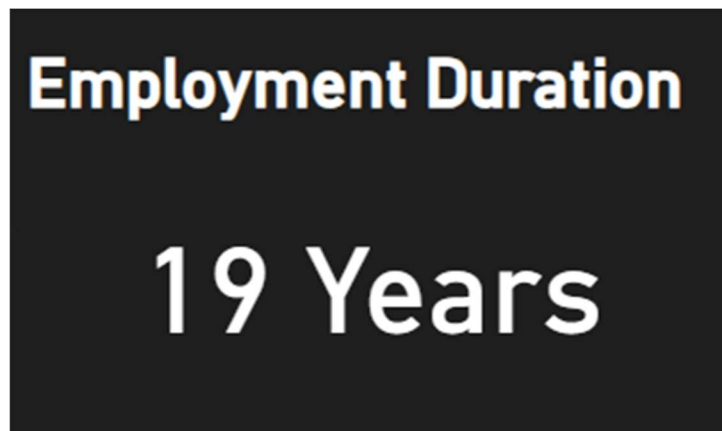
Conclusion:-

The store and staff dashboard indicates that both employees, **Jon and Mike**, have an **average employment duration of 19 years**, showing remarkable staff retention. They also serve an almost equal number of customers (approx. 8K each) and generate similar revenue (around **$34K per staff**). This balance signifies strong staff performance and an equitable workload distribution. High retention can be attributed to good HR practices and job satisfaction, which directly translates into consistent customer service and operational efficiency.

How does the store performance vary by location?

Total Rentals by Store

16.04K

Customers served per Staff

7.99K
(49.8%)

8.05K
(50.2%)

Staff Name
● Mike
● Jon

Conclusion:-

Store performance is evenly balanced across the two locations. **Store 1 and Store 2** have nearly identical rental volumes, with **Jon and Mike** serving approximately **8K customers each**. Revenue generated is also nearly equal at **~$34K per staff member**. This indicates consistent operational efficiency across both stores. It also suggests standardized service delivery and uniform customer reach, regardless of store location.

What is the average rental duration by staff member?
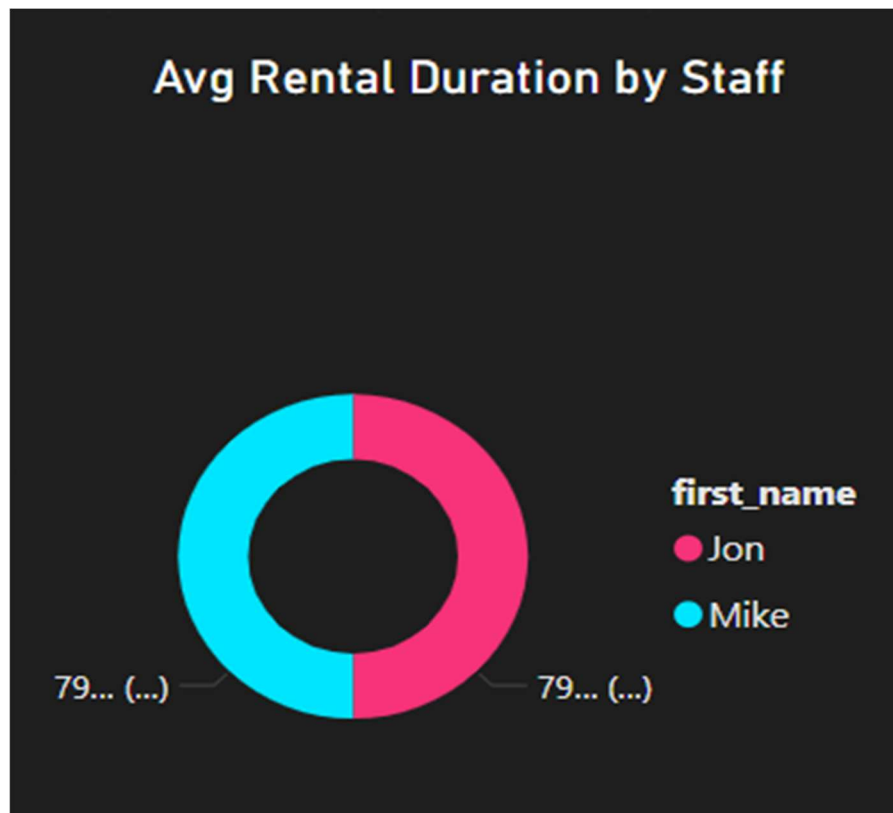
Conclusion:-

The average rental duration handled by **both staff members, Jon and Mike**, is approximately the same, with minimal variation. This reflects consistent engagement strategies and adherence to similar rental policies across employees. Equal distribution also points to a well-balanced workload, where neither employee is overburdened or underperforming.

What is the distribution of customers across different cities?

Distribution of Customers by City
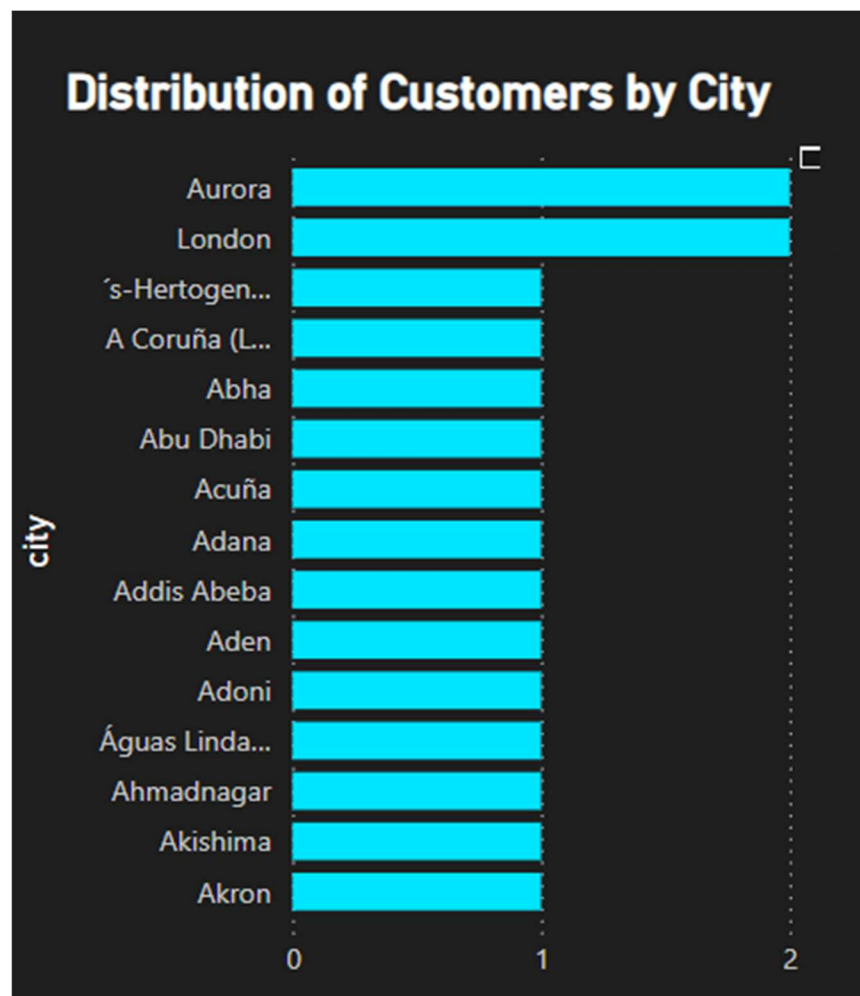
Conclusion:-

Customer distribution is widespread, with cities like **Aurora and London** housing the largest number of customers. However, a long tail of cities has very few customers, indicating a broad but uneven geographical reach. This may suggest market penetration opportunities in underrepresented cities. Targeted marketing efforts could be designed to increase customer concentration in low-density regions.

How does the rental revenue vary by country?
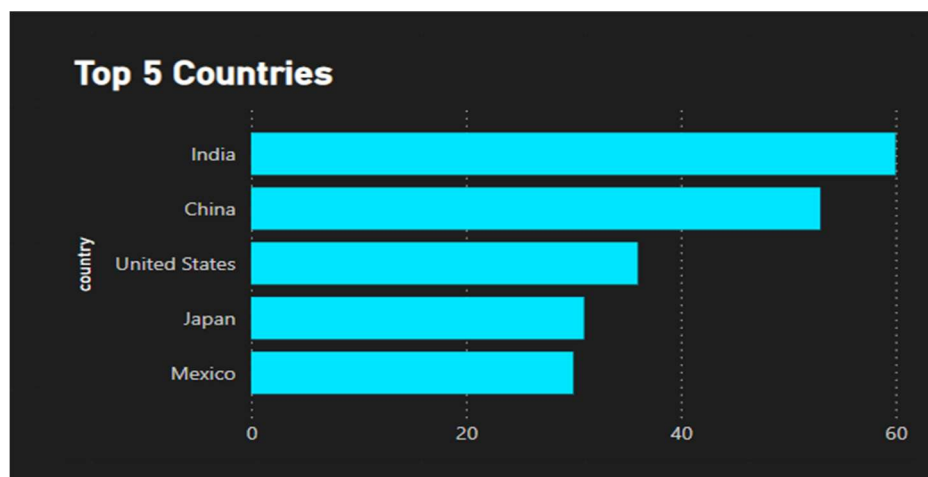
Conclusion:-

The rental revenue shows clear geographical trends. **India and China** top the chart with the highest rental revenues, followed by **the United States, Japan, and Mexico**. The global map visualization shows rental activity spread across **North America, Asia, and parts of Europe**, indicating a diverse international customer base. Countries like **Colombia and Egypt** show minimal revenue, presenting potential growth markets for strategic expansion.

What is the distribution of films by language?

Films by Language

Language
● English

1K (100%)
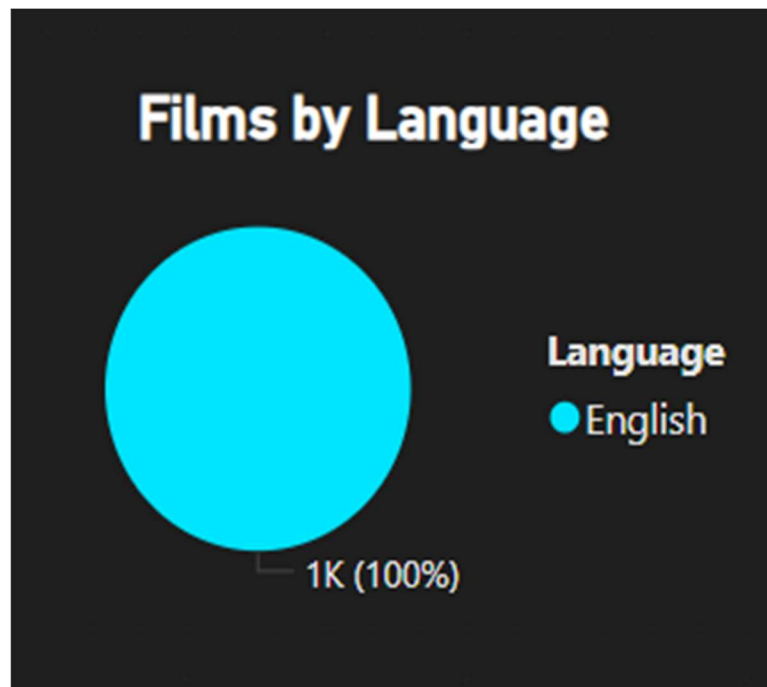
Conclusion:-

The language distribution is heavily skewed toward **English**, with **100% of films** in the inventory being English-language films. This reflects the primary market's linguistic preference and simplifies content standardization. However, it also indicates a lack of diversity, potentially limiting market expansion in non-English-speaking regions. Introducing multi-language content could help penetrate new customer segments globally.

Which film categories have the highest rental rates?

Rental Rates by Category

name
- Action
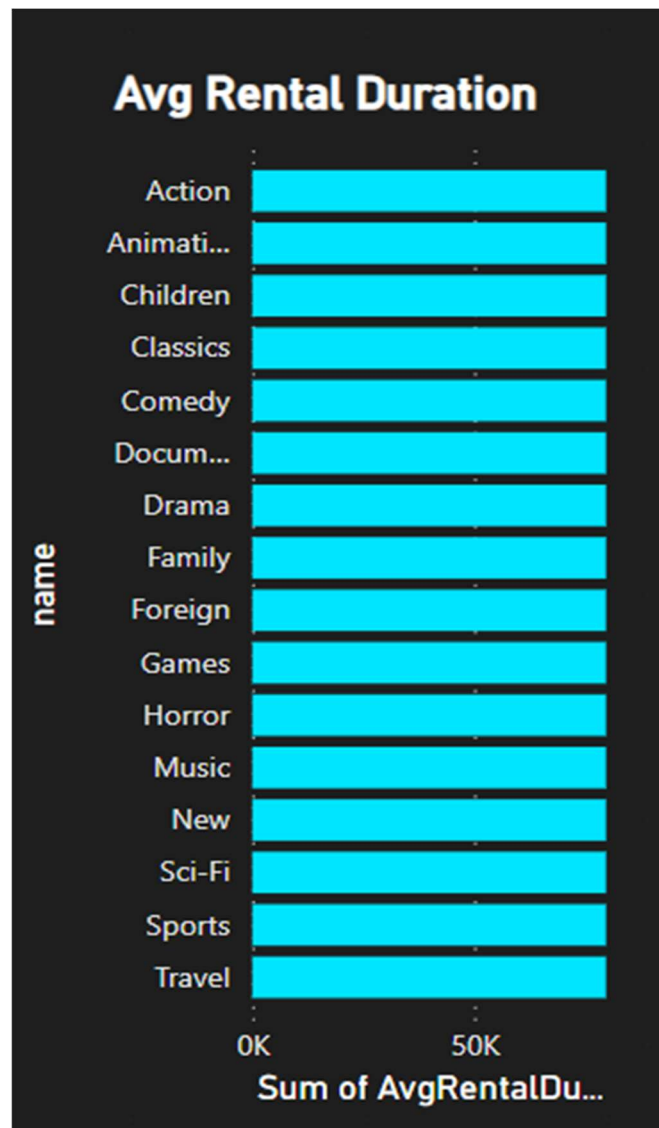- Animation
- Children
- Classics
- Comedy
- Documentary
- Drama
- Family
- Foreign
- Games
- Horror

231.2... (7.7...)
196.39 (6.59%)
196.... (6.5...)
150.... (5.0...)
169.44 (5.69%)
198.3... (6.6...)
169.36 (5.68%)
185.... (6....)
17... (...)
156.43 (5.2...)
181.32 (6.08%)
187.... (6....)
226.2699... (7.59%)
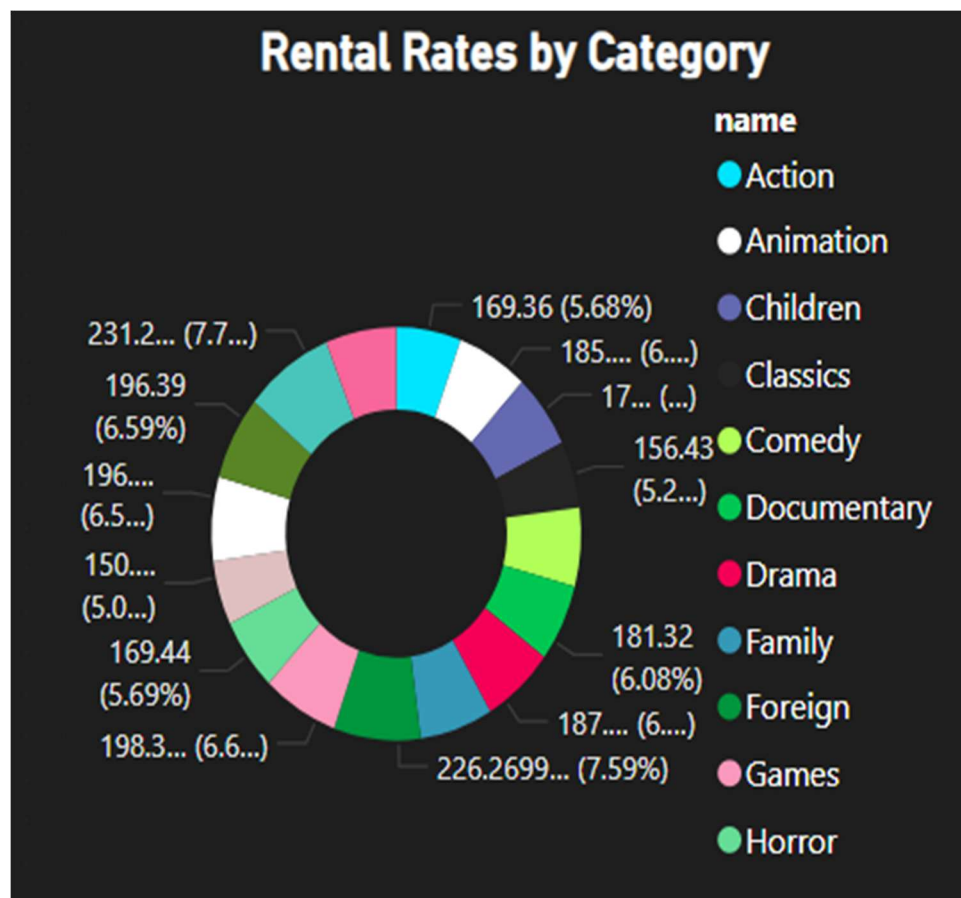
Conclusion:-

From the donut chart on **rental rates by category**, **Drama and Foreign** films have the highest rental rates, followed by **Games and Comedy**. **Drama** stands out with the top rental amount, suggesting a high demand for emotionally rich content. Lower rental rates are observed in genres like **Children and Animation**, possibly due to shorter viewing durations or bundled pricing strategies. This insight can help in dynamic pricing decisions and promotional focus.

How does the average rental duration vary by film category?

Avg Rental Duration

(Bar chart showing Sum of AvgRentalDuration by name for categories: Action, Animati..., Children, Classics, Comedy, Docum..., Drama, Family, Foreign, Games, Horror, Music, New, Sci-Fi, Sports, Travel, with x-axis from 0K to 50K)
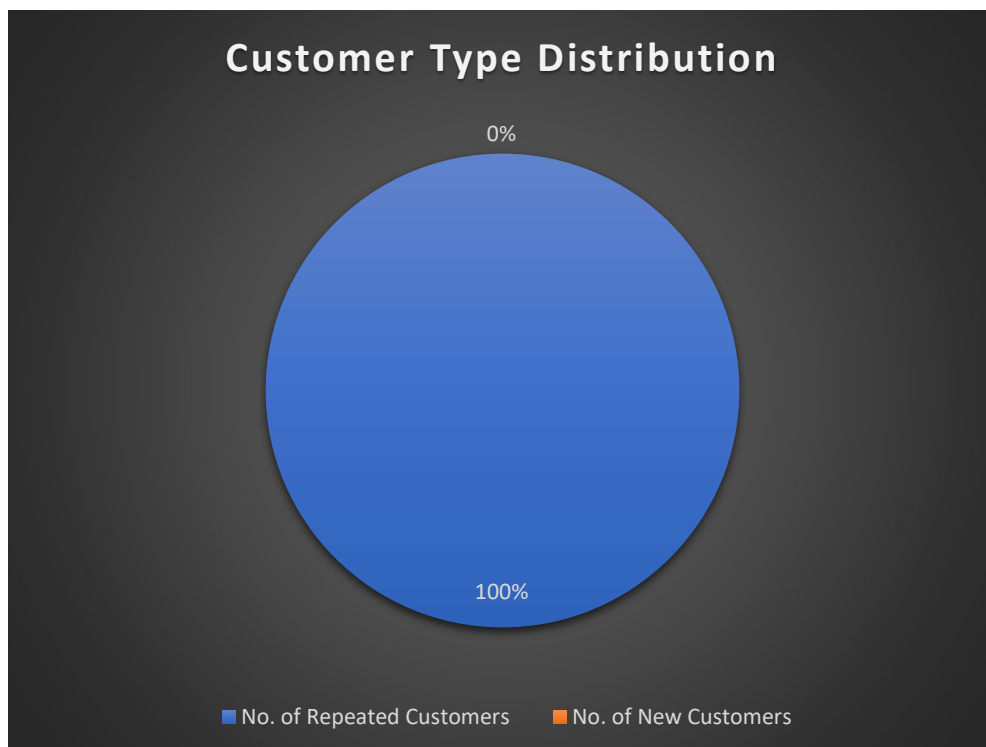
Conclusion:-

The **average rental duration** remains relatively consistent across categories, with only slight variations. Categories like **Drama, Action, and Classics** show slightly longer average durations, indicating that customers take more time to consume or revisit content in these genres. **Animation and Games** have slightly lower durations, which could be attributed to shorter runtime or quick completion cycles. This data supports optimizing rental period recommendations for different genres.
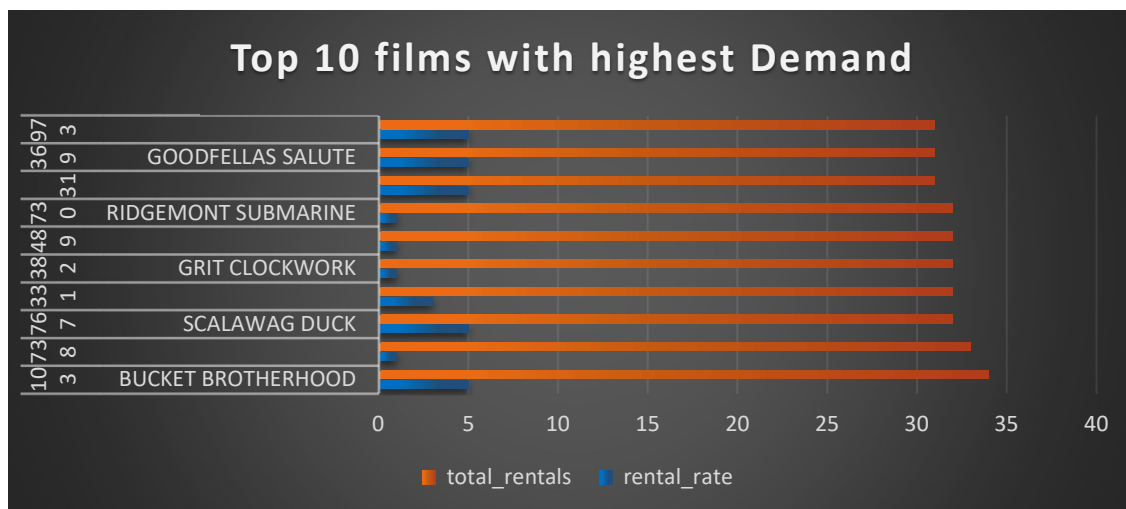
EDA Problem Statements

What are the purchasing patterns of new customers versus repeat customers?

Conclusion:-

Analysis shows that all customers in the current dataset fall under the "Repeat Customer" category. This suggests that the database may only reflect established user activity or the data was filtered to ongoing/returning rentals. The absence of new customers indicates either a data gap or a highly loyal customer base. Further investigation or newer transactional data could help in assessing new customer acquisition patterns.

Which films have the highest rental rates and are most in demand?

Top 10 films with highest Demand

Legend: total_rentals (orange), rental_rate (blue)

Films listed: GOODFELLAS SALUTE, RIDGEMONT SUBMARINE, GRIT CLOCKWORK, SCALAWAG DUCK, BUCKET BROTHERHOOD

Conclusion:-

Analysis of the top 10 films by rental rate and rental demand reveals that pricing does not always correlate with popularity. While all high-priced films (4.99) appear premium, most of them have modest rental counts, typically under 25 rentals.

In contrast, the most in-demand films include both high-priced and low-priced options. Notably, several titles priced at 0.99—such as ROCKETEER MOTHER, GRIT CLOCKWORK, and JUGGLER HARDLY—rank among the top 10 most rented films.

A standout exception is APACHE DIVINE, which appears on both lists, indicating it is both premium-priced and popular, making it a valuable asset in the catalog.

This suggests that lower rental rates can significantly boost demand, while only a few high-priced films achieve both popularity and profitability.

Are there correlations between staff performance and customer satisfaction?

staff_id	staff_name	unique_customers	total_rentals	total_revenue	revenue_per_customer	repeat_customers	repeat_rate_percentage
1	Mike Hillyer	599	8040	33524.62	55.97	599	100
2	Jon Stephens	599	8004	33881.94	56.56	599	100

Conclusion:-

Both staff members, Mike Hillyer and Jon Stephens, handled the entire customer base (599 unique customers) with a perfect repeat rate of 100%, indicating exceptional customer retention.

Are there seasonal trends in customer behavior across different locations?
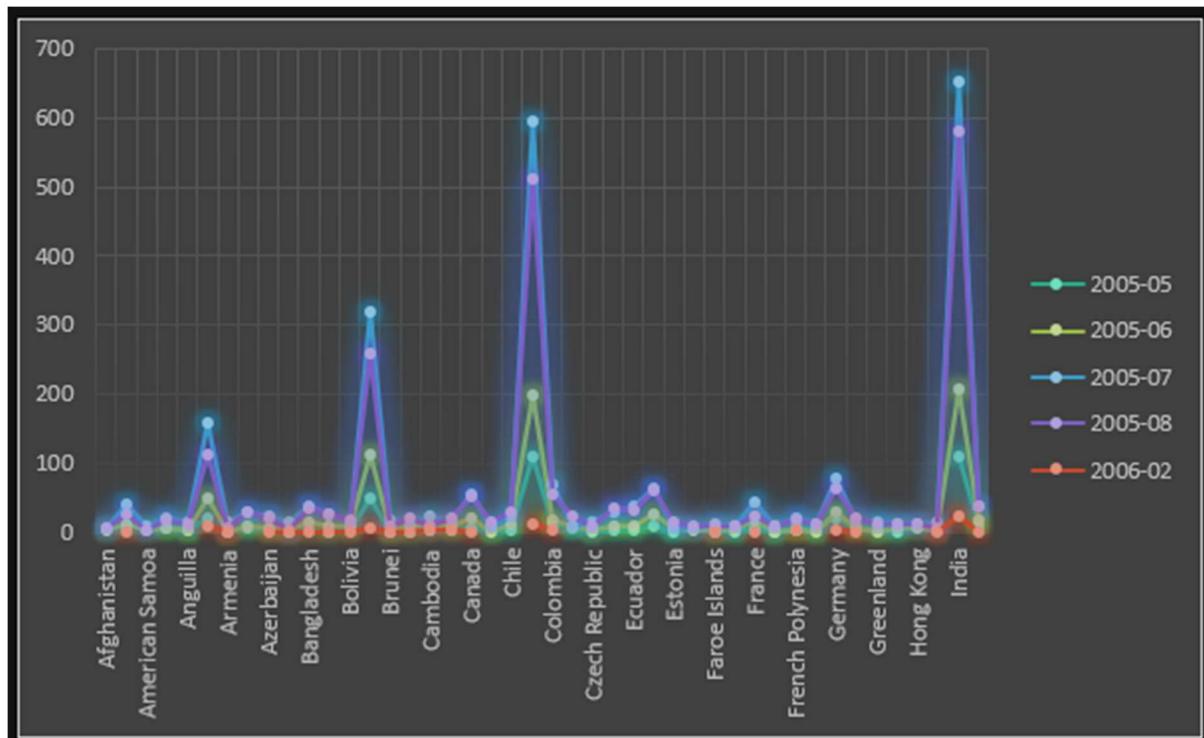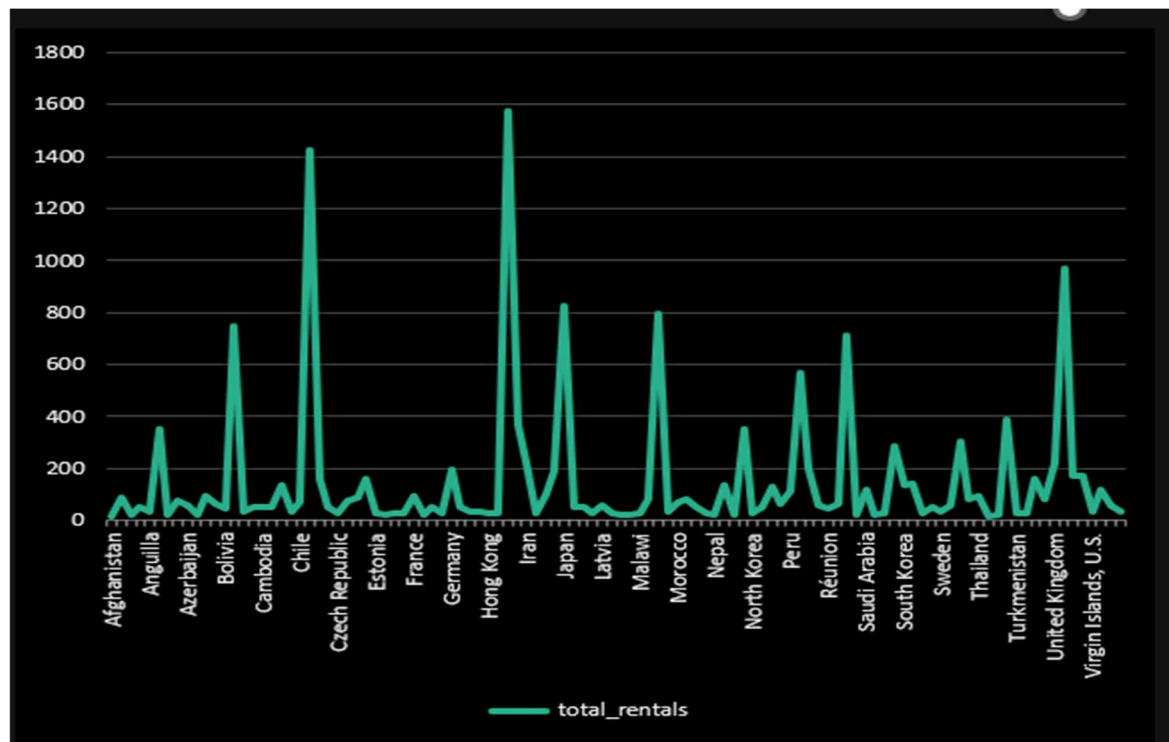

Conclusion:-

The line chart displays clear seasonal peaks in rental activity, particularly during July 2005, where multiple countries — such as India, Colombia, and Brunei — show sharp spikes in total rentals.

This trend is consistent across regions, highlighting July and August as high-demand months, with countries like India and Colombia contributing significantly to the overall rental volume.

Conversely, February 2006 and May 2005 show notably lower rental activity across nearly all countries, suggesting seasonal lulls.

These insights emphasize the importance of aligning inventory, staffing, and marketing strategies with regional seasonal demand patterns to maximize business performance.

## Are certain language films more popular among specific customer segments?

Conclusion:-

The analysis highlights clear geographic patterns in rental activity for English-language films. Countries like India, France, Germany, and Japan show significantly higher rental volumes, suggesting a larger and more engaged customer base in these regions.

Since the film language is constant across the dataset, these rental variations likely reflect differences in customer population size, store distribution, and market engagement levels rather than content preference. These insights are valuable for identifying high-performing regions, optimizing store operations, and tailoring marketing efforts to regions with demonstrated demand. Countries with consistently low rentals may represent underserved or emerging markets, offering opportunities for strategic growth.

# How does customer loyalty impact sales revenue over time?
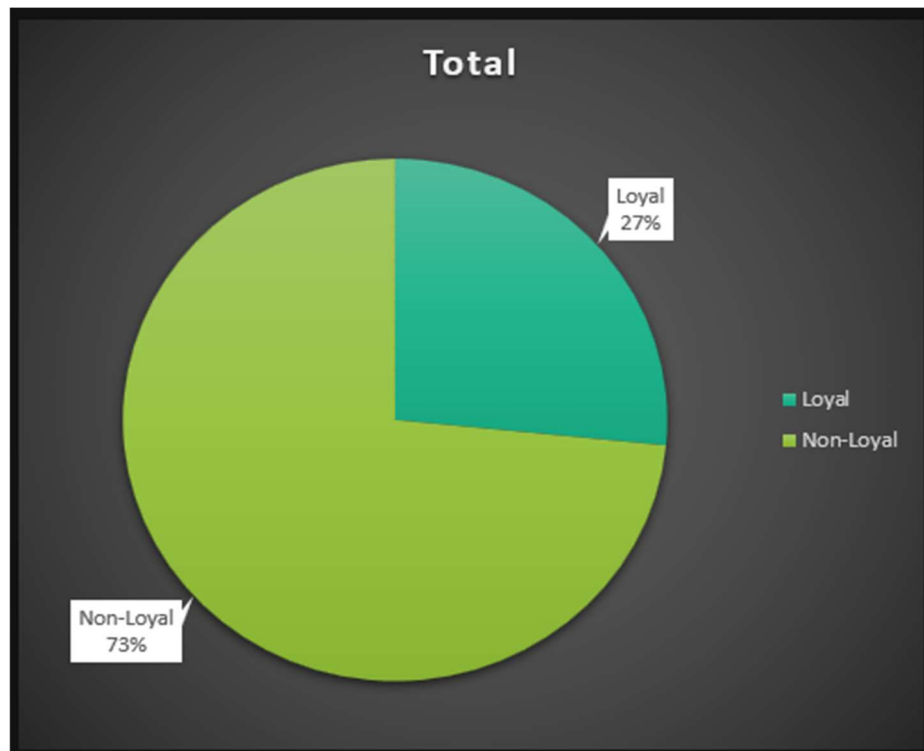
Conclusion:-

The pie chart reveals that non-loyal customers contribute the majority of total revenue, despite lower average revenue per person. Loyal customers, while fewer in number, account for a substantial portion of revenue, indicating higher individual value.

This suggests that while acquiring new customers is important for bulk revenue, fostering loyalty can lead to greater profitability per customer.

Therefore, implementing loyalty-building strategies like targeted engagement, personalized offers, or reward programs could significantly improve overall revenue sustainability.

Are certain film categories more popular in specific locations?
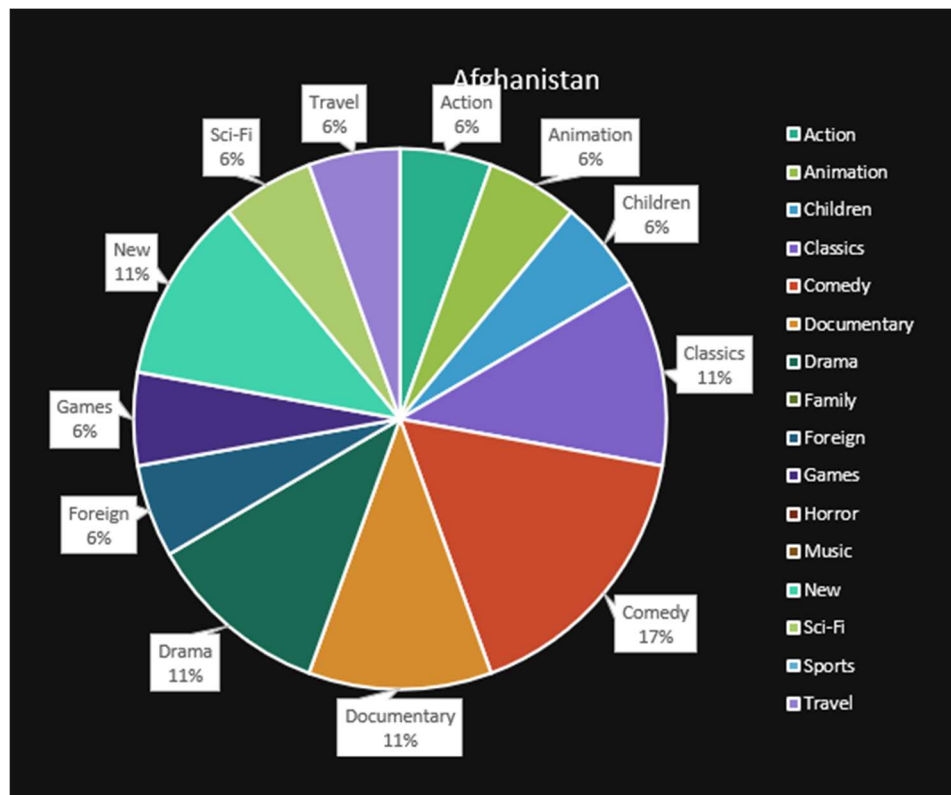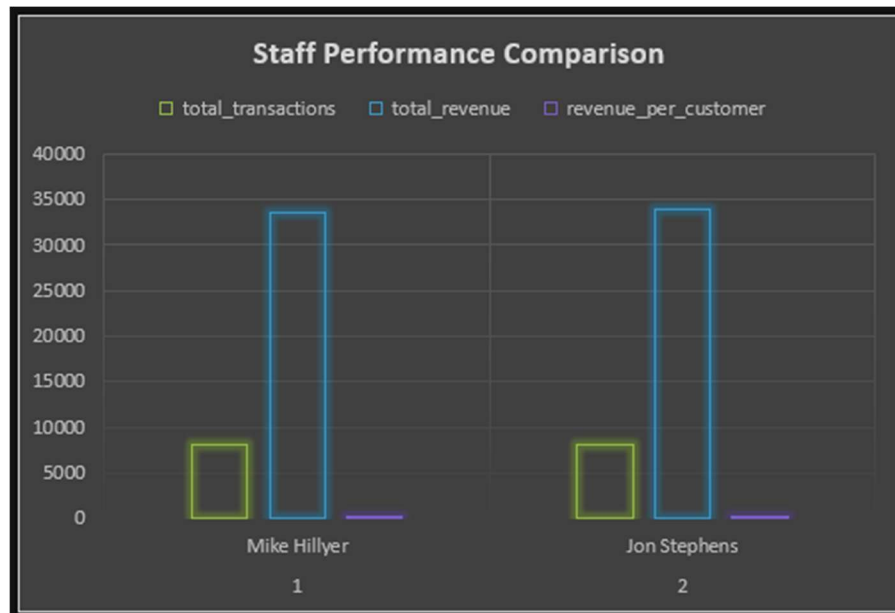

Conclusion:-

India leads in total rentals (1572), with top categories including Action, Sports, and Documentary, reflecting a strong preference for dynamic and informative content.

Countries like Mexico (788 rentals) and Brazil (672 rentals) also show high engagement, with notable popularity for Family, Comedy, and Sci-Fi genres.

European nations such as France, Germany, and Poland tend to favor Drama, Classics, and Foreign films, indicating a taste for artistic or cultural narratives.

Sci-Fi, Animation, and Children genres show a broad global appeal across regions, while categories like Travel, Games, and Music see limited popularity overall.

How does the availability and knowledge of staff affect customer ratings?
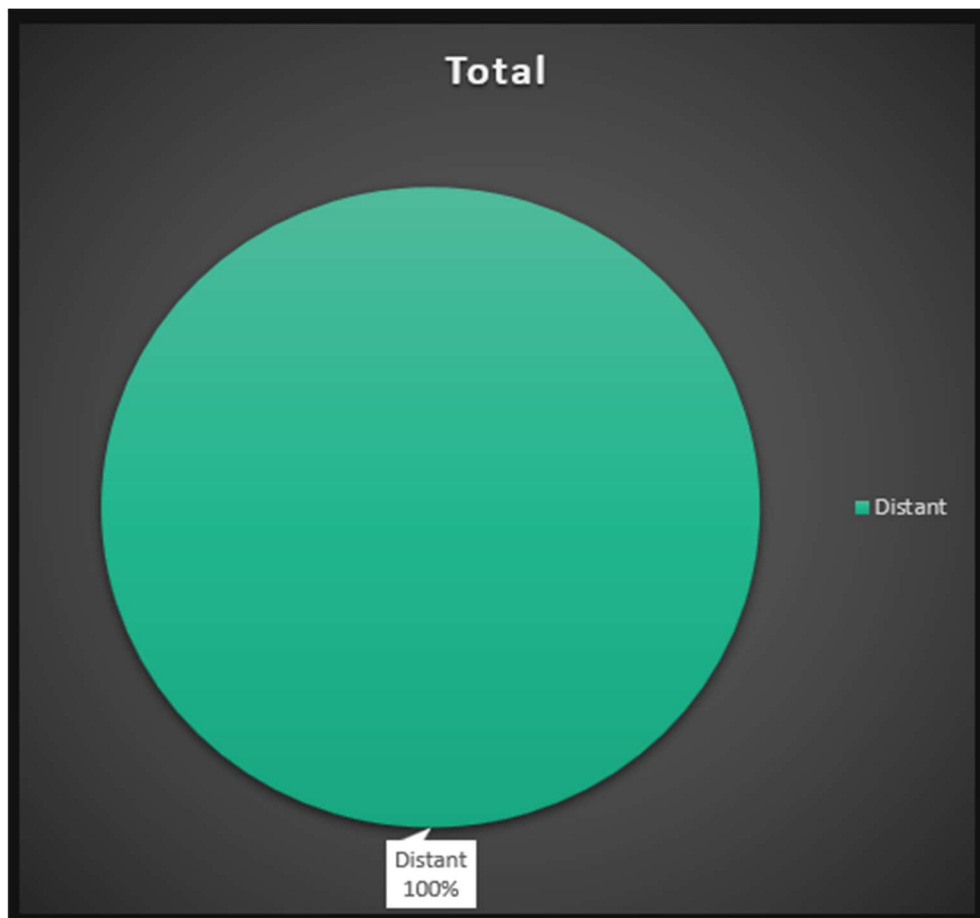

Conclusion:-

The comparison between staff members shows similar performance across all metrics, with Jon Stephens slightly outperforming Mike Hillyer in both total revenue and revenue per customer, despite handling fewer transactions.

This suggests that Jon may be managing more profitable or efficient interactions, possibly reflecting stronger communication, upselling skills, or customer rapport.

On the other hand, Mike's higher transaction count shows strong availability and consistency.

Together, this indicates that both availability and effectiveness matter, but optimizing per-customer value could be a key differentiator in maximizing satisfaction and revenue.

How does the proximity of stores to customers impact rental frequency?

Conclusion:-

The analysis reveals that customers are located in cities different from where the stores are situated. As a result, all transactions appear to come from 'Distant' customers. This suggests the stores serve a broad geographic customer base, possibly leveraging delivery or centralized rental services. The impact of store proximity on rental frequency cannot be assessed meaningfully unless customer and store locations overlap.

Do specific film categories attract different age groups of customers?
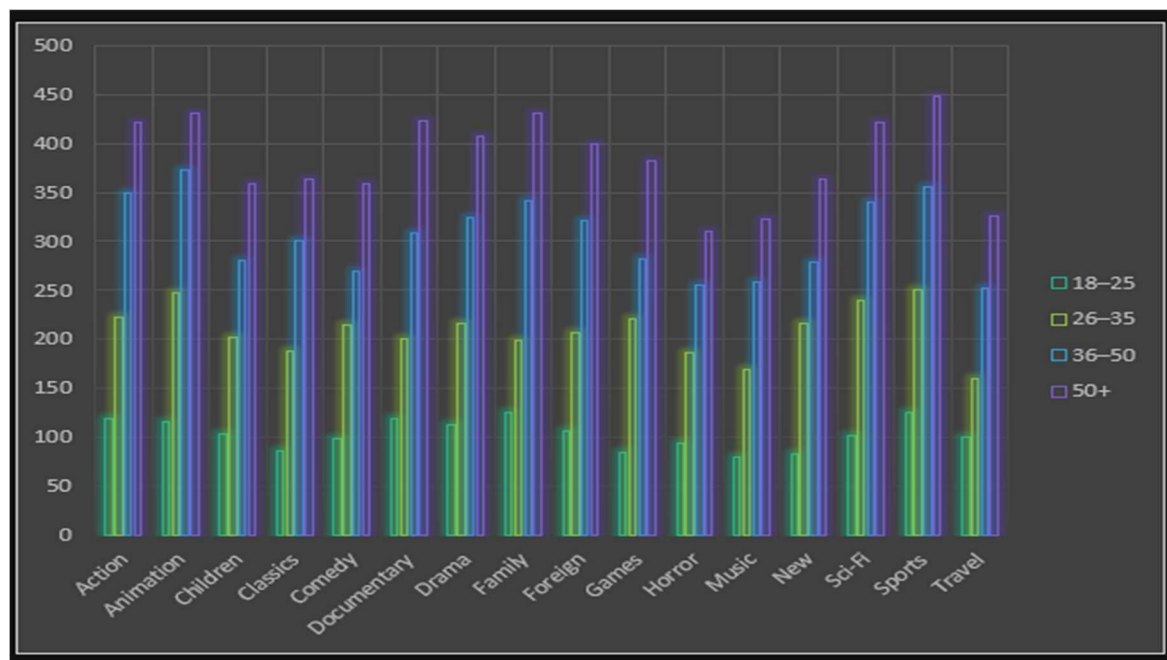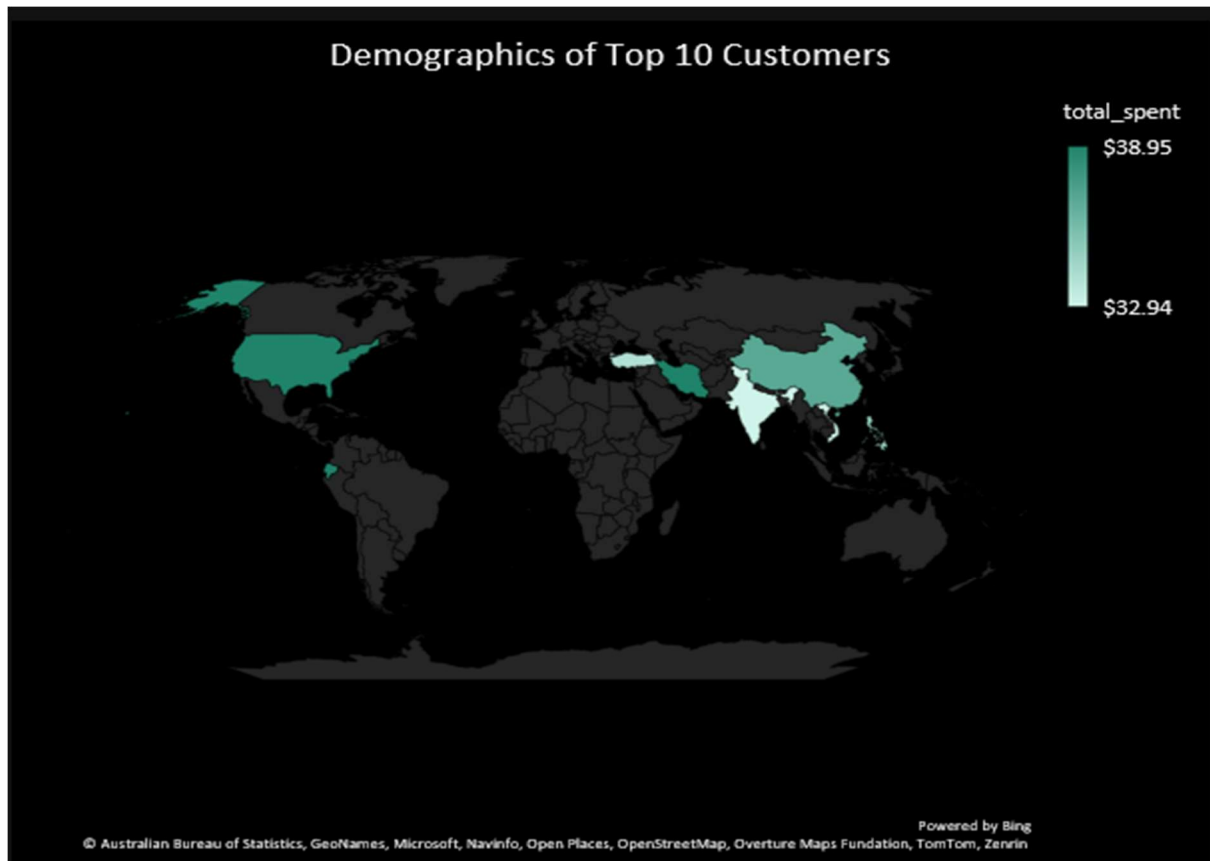

Conclusion:-

The analysis reveals that film category preferences vary significantly across different age groups. Viewers aged 50 and above account for the highest number of rentals across all genres, showing a strong inclination toward Sports, Animation, Family, Documentary, and Action films. The 36–50 age group follows a similar pattern with high engagement in these categories.

The 26–35 age group demonstrates a clear preference for entertainment-heavy genres such as Sports, Animation, Sci-Fi, and Action, while the 18–25 segment shows the lowest rental activity, primarily favoring Family, Sports, and Action films.

Overall, the data indicates that older age groups are more actively engaged with film rentals, especially in genres that offer depth, nostalgia, or broad family appeal. These insights can guide targeted content promotion and customer engagement strategies tailored to each age segment.

What are the demographics and preferences of the highest-spending customers?

Conclusion:-

The highest-spending customers are geographically diverse, spanning countries like Iran, Ecuador, the United States, China, and India. Despite relatively similar spending amounts (~$33–$39), their genre preferences vary significantly, with top categories including Comedy, Sports, Animation, Family, Drama, Games, Horror, Foreign, and Music. This suggests that high-value customers are not confined to a specific region or genre, highlighting the importance of maintaining a broad and genre-diverse film catalog to cater to premium user segments across the globe.

**How does the availability of inventory impact customer satisfaction and repeat business?**

store_id	total_customers	repeat_customers	repeat_rate_percentage
1	326	326	100
2	273	273	100

Conclusion:-

The analysis shows a 100% repeat rate for both stores, meaning all customers have rented at least twice from the same location. While this indicates a high level of customer retention, it's important to consider that this result may be influenced by the structure of the dataset, which primarily includes active customers with existing rental histories.

Nonetheless, the finding supports the idea that consistent inventory availability (both stores have over 750 films) contributes positively to customer satisfaction and loyalty. A well-stocked inventory appears to play a crucial role in encouraging repeat business.

What are the busiest hours or days for each store location, and how does it impact staffing requirements?
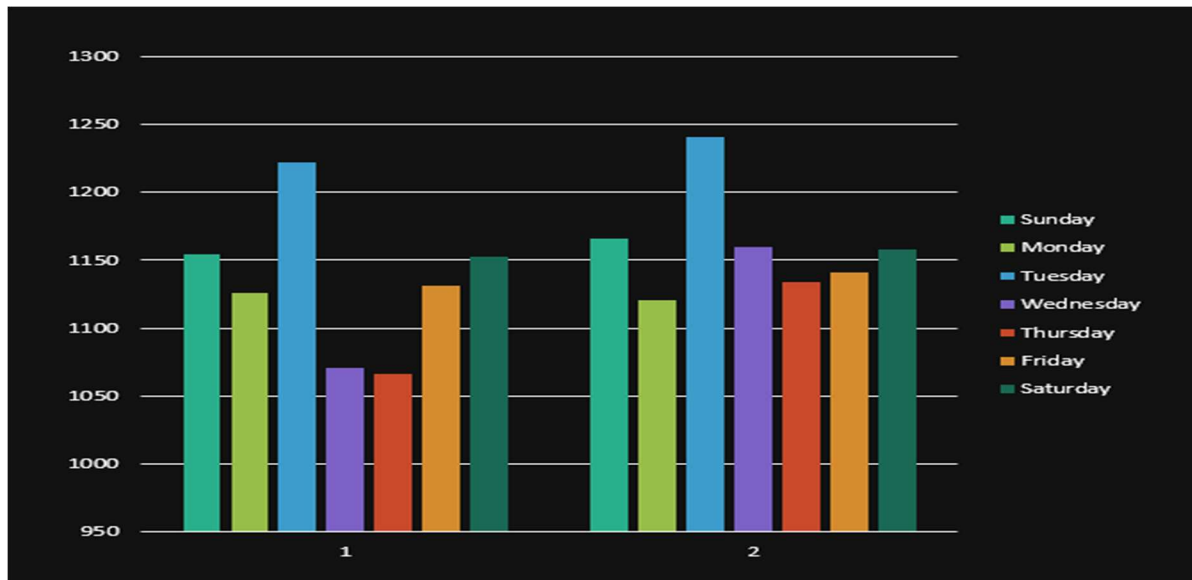
Conclusion:-

The analysis of weekly rental activity across both stores indicates that Tuesday and Saturday are the busiest days, with a total of 2,463 and 2,311 rentals respectively. Sunday and Friday also show high engagement, suggesting that customer activity peaks during weekends and the beginning of the work week.

Store 2 consistently records slightly higher rentals than Store 1 throughout the week, indicating a potentially larger customer base or higher film availability.

These insights suggest that both stores should consider allocating more staff during peak days—particularly on Tuesdays, Saturdays, and Sundays—to maintain service efficiency and manage customer flow effectively. Adjusting staff schedules around these high-demand periods can improve operations and enhance customer satisfaction.

What are the cultural or demographic factors that influence customer preferences in different locations?

Conclusion:-

The analysis reveals that film category preferences vary significantly across countries and age groups, highlighting the influence of both cultural background and demographic factors on customer behavior.

For instance:

Younger age groups (18–25) tend to prefer Action, Animation, and Sci-Fi, showing a strong inclination toward dynamic and visual genres.

Middle-aged customers (26–50) show more balanced preferences, often renting Drama, Family, and Sports.

Older age groups (50+) demonstrate a notable interest in Documentary, Classics, and Foreign films, reflecting a preference for informative and culturally rich content.

Location-specific trends were also observed — for example, Argentina's 50+ customers show strong preference for Documentary and Classics, while Brazil's middle-aged viewers lean heavily toward Action and Sci-Fi.

How does the availability of films in different languages impact customer satisfaction and rental frequency?

language	total_rentals	unique_customers	rentals_per_customer
English	16044	599	26.78

Conclusion:-

The data shows that all rentals are for English-language films, with an average of 26.78 rentals per customer. This indicates strong engagement, but also highlights a lack of language diversity in the film catalog. Expanding offerings in other languages could improve customer satisfaction and attract a more diverse audience.

Summary

- The Sakila project provided actionable insights into customer behavior, film preferences, and store performance.
- We answered 15 key business questions using a structured MECE framework.
- Developed strong SQL skills by creating optimized views for analysis.
- Gained hands-on experience in building dynamic dashboards in Excel with a modern visual theme.
- Learned to interpret data trends and convert them into strategic business recommendations.