SAKILA DVD MOVIE RENTAL

DEEPAK SHARMA

Abstract

The Sakila DVD Rental database, a representative sample database for educational and practice purposes, serves as a valuable resource for understanding relational database management systems (RDBMS) and honing skills in SQL queries. This abstract outlines a comprehensive analysis of the Sakila database, encompassing data modeling, performance optimization, and business intelligence aspects.

1. Data Modeling: The study delves into the structure of the Sakila database, examining its tables, relationships, and entities. An in-depth exploration of the schema provides insights into the design principles, normalization techniques, and the rationale behind the database architecture. The goal is to enhance understanding of how a real-world scenario is translated into a structured, normalized database model.

2. Performance Optimization: With an emphasis on database performance, the analysis explores strategies for optimizing SQL queries and indexing techniques. The objective is to identify and implement improvements to enhance the retrieval speed and overall efficiency of data operations. Practical examples and case studies demonstrate the impact of various optimization techniques on query execution times.

3. Business Intelligence: Leveraging the Sakila database, the study explores the application of business intelligence (BI) techniques to derive meaningful insights. Through the use of SQL queries, data visualization, and reporting tools, the analysis demonstrates how businesses can extract valuable information from the database to support decision-making processes. Topics include customer behavior analysis, revenue trends, and genre preferences.

This comprehensive exploration of the Sakila DVD Rental database aims to provide a valuable resource for database enthusiasts, students, and professionals seeking to deepen their understanding of data modeling, performance optimization, and business intelligence within the context of a relational database environment.

CONTENT

INTRODUCTION

➢ Project Overview:

The project aims to leverage the Sakila DVD Movie Rental database to extract valuable insights and drive informed decision-making within the context of a DVD rental business. Through a multifaceted approach encompassing data exploration, analysis, and visualization, the project seeks to address key business questions, optimize operational efficiency, and enhance customer satisfaction.

Objective:

The primary objective of the project is to utilize the Sakila database as a foundation for extracting actionable insights that can drive strategic decisions and operational improvements within the DVD rental business. By analyzing customer behaviour, rental trends, inventory management, and revenue streams, the project aims to identify opportunities for growth and optimization.

Scope:

The project scope encompasses various aspects of the DVD rental business, including customer segmentation, genre preferences, rental patterns, staff performance, and revenue analysis. Additionally, the project explores avenues for improving inventory management, enhancing customer experience, and maximizing revenue through targeted promotions and pricing strategies.

Methodology:

The project methodology involves several stages, starting with data exploration and preprocessing to familiarize stakeholders with the structure and content of the Sakila database. Subsequently, data analysis techniques, including SQL queries, statistical analysis, and data visualization, are employed to uncover patterns, trends, and correlations within the dataset. Finally, insights derived from the analysis are translated into actionable recommendations and strategic initiatives.

Key Deliverables:

The project will deliver a comprehensive set of deliverables, including:

Detailed data exploration and profiling report

Analysis of customer behavior, rental patterns, and revenue streams

Visualization of key metrics and trends through charts, graphs, and dashboards

Recommendations for operational improvements, marketing strategies, and pricing optimization

Executive summary highlighting key findings, insights, and actionable recommendations

Expected Outcomes:

The project's expected outcomes include:

Enhanced understanding of customer preferences and behavior

Identification of opportunities for revenue growth and operational efficiency

Improved inventory management and resource allocation

Data-driven decision-making processes informed by actionable insights

Alignment of business strategies with market trends and customer needs

Timeline and Resources:

The project timeline and resource allocation will be determined based on the scope and complexity of the analysis. Key resources required include access to the Sakila database, expertise in SQL queries and data analysis, and visualization tools for presenting findings effectively.

In summary, the project endeavors to unlock the full potential of the Sakila DVD Movie Rental database as a catalyst for driving business innovation, optimization, and growth. By harnessing the power of data analytics and business intelligence, the project aims to empower stakeholders with actionable insights that can shape the future trajectory of the DVD rental business.

➢ Introduction of dataset Table:

Sakila is structured around a set of interlinked tables, each representing a distinct entity within the DVD rental business. These tables are designed in accordance with relational database principles, with relationships established through primary and foreign keys. Key entities within the dataset include:

ACTOR

The actor table lists information for all actors.

The actor table is joined to the film table by means of the film_actor table.

Columns:

• actor_id: A surrogate primary key used to uniquely identify each actor in the table.

• first_name: The actor first name.

• last_name: The actor last name.

• last_update: When the row was created or most recently updated.

ADDRESS

The address table contains address information for customers, staff, and stores.

The address table primary key appears as a foreign key in the customer, staff, and store tables.

Columns:

• address_id: A surrogate primary key used to uniquely identify each address in the table.

• address: The first line of an address.

• address2: An optional second line of an address.

• district: The region of an address, this may be a state, province, prefecture, etc.

• city_id: A foreign key pointing to the city table.

• postal_code: The postal code or ZIP code of the address (where applicable).

• phone: The telephone number for the address.

• last_update: When the row was created or most recently updated.

• location: A Geometry column with a spatial index on it.

CATEGORY

The category table lists the categories that can be assigned to a film.

The category table is joined to the film table by means of the film_category table.

Columns:

• category_id: A surrogate primary key used to uniquely identify each category in the table.

• name: The name of the category.

• last_update: When the row was created or most recently updated.

CITY

The city table contains a list of cities.

The city table is referred to by a foreign key in the address table and refers to the country table using

a foreign key.

Columns:

• city_id: A surrogate primary key used to uniquely identify each city in the table.

• city: The name of the city.

• country_id: A foreign key identifying the country that the city belongs to.

• last_update: When the row was created or most recently updated.

COUNTRY

The country table contains a list of countries.

The country table is referred to by a foreign key in the city table.

Columns:

• country_id: A surrogate primary key used to uniquely identify each country in the table.

• country: The name of the country.

• last_update: When the row was created or most recently updated.

CUSTOMER

The customer table contains a list of all the customers.

The customer table is referred to in the payment and rental tables and refers to the address and

store tables using foreign keys.

Columns:

• customer_id: A surrogate primary key used to uniquely identify each customer in the table.

• store_id: A foreign key identifying the customer "home store." Customers are not limited to renting

only from this store, but this is the store at which they generally shop.

• first_name: The customer first name.

• last_name: The customer last name.

• email: The customer email address.

• address_id: A foreign key identifying the customer address in the address table.

• active: Indicates whether the customer is an active customer. Setting this to FALSE serves as an

alternative to deleting a customer outright. Most queries should have a WHERE active = TRUE clause.

• create_date: The date the customer was added to the system. This date is automatically set using a trigger during an INSERT.

• last_update: When the row was created or most recently updated.

FILM

The film table is a list of all films potentially in stock in the stores. The actual in-stock copies of each film

are represented in the inventory table.

The film table refers to the language table and is referred to by the film_category, film_actor,

and inventory tables.

Columns:

• film_id: A surrogate primary key used to uniquely identify each film in the table.

• title: The title of the film.

• description: A short description or plot summary of the film.

• release_year: The year in which the movie was released.

• language_id: A foreign key pointing at the language table; identifies the language of the film.

• original_language_id: A foreign key pointing at the language table; identifies the original

language of the film. Used when a film has been dubbed into a new language.

• rental_duration: The length of the rental period, in days.

• rental_rate: The cost to rent the film for the period specified in the rental_duration column.

• length: The duration of the film, in minutes.

• replacement_cost: The amount charged to the customer if the film is not returned or is returned in a damaged state.

• rating: The rating assigned to the film. Can be one of: G, PG, PG-13, R, or NC-17.

• special_features: Lists which common special features are included on the DVD. Can be zero or

more of: Trailers, Commentaries, Deleted Scenes, Behind the Scenes.

• last_update: When the row was created or most recently updated.

FILM_ACTOR

The film_actor table is used to support a many-to-many relationship between films and actors. For each

actor in a given film, there will be one row in the film_actor table listing the actor and film.

The film_actor table refers to the film and actor tables using foreign keys.

Columns:

• actor_id: A foreign key identifying the actor.

• film_id: A foreign key identifying the film.

• last_update: When the row was created or most recently updated.

FILM_CATEGORY

The film_category table is used to support a many-to-many relationship between films and categories.

For each category applied to a film, there will be one row in the film_category table listing the category

and film.

The film_category table refers to the film and category tables using foreign keys.

Columns:

• film_id: A foreign key identifying the film.

• category_id: A foreign key identifying the category.

• last_update: When the row was created or most recently updated.

INVENTORY

The inventory table contains one row for each copy of a given film in each store.

The inventory table refers to the film and store tables using foreign keys and is referred to by the rental table.

Columns:

• inventory_id: A surrogate primary key used to uniquely identify each item in inventory.

• film_id: A foreign key pointing to the film this item represents.

• store_id: A foreign key pointing to the store stocking this item.

• last_update: When the row was created or most recently updated.

LANGUAGE

The language table is a lookup table listing the possible languages that films can have for their language and original language values.

The language table is referred to by the film table.

Columns:

• language_id: A surrogate primary key used to uniquely identify each language.

• name: The English name of the language.

• last_update: When the row was created or most recently updated.

PAYMENT

The payment table records each payment made by a customer, with information such as the amount and

the rental being paid for (when applicable).

The payment table refers to the customer, rental, and staff tables.

Columns:

• payment_id: A surrogate primary key used to uniquely identify each payment.

• customer_id: The customer whose balance the payment is being applied to. This is a foreign key reference to the customer table.

• staff_id: The staff member who processed the payment. This is a foreign key reference to the staff table.

• rental_id: The rental that the payment is being applied to. This is optional because some payments

are for outstanding fees and may not be directly related to a rental.

• amount: The amount of the payment.

• payment_date: The date the payment was processed.

• last_update: When the row was created or most recently updated.

RENTAL

The rental table contains one row for each rental of each inventory item with information about who rented what item, when it was rented, and when it was returned.

The rental table refers to the inventory, customer, and staff tables and is referred to by the payment table.

Columns:

• rental_id: A surrogate primary key that uniquely identifies the rental.

• rental_date: The date and time that the item was rented.

• inventory_id: The item being rented.

• customer_id: The customer renting the item.

• return_date: The date and time the item was returned.

• staff_id: The staff member who processed the rental.

• last_update: When the row was created or most recently updated.

STAFF

The staff table lists all staff members, including information for email address, login information, and

picture.

The staff table refers to the store and address tables using foreign keys, and is referred to by the

rental, payment, and store tables.

Columns:

• staff_id: A surrogate primary key that uniquely identifies the staff member.

• first_name: The first name of the staff member.

• last_name: The last name of the staff member.

• address_id: A foreign key to the staff member address in the address table.

• picture: A BLOB containing a photograph of the employee.

• email: The staff member email address.

• store_id: The staff member "home store." The employee can work at other stores but is generally

assigned to the store listed.

• active: Whether this is an active employee. If employees leave, their rows are not deleted from this

table; instead, this column is set to FALSE.

• username: The user name used by the staff member to access the rental system.

• password: The password used by the staff member to access the rental system. The password should

be stored as a hash using the SHA2() function.

• last_update: When the row was created or most recently updated.

STORE

The store table lists all stores in the system. All inventory is assigned to specific stores, and staff and

customers are assigned a “home store”.

The store table refers to the staff and address tables using foreign keys and is referred to by the

staff, customer, and inventory tables.

Columns:

• store_id: A surrogate primary key that uniquely identifies the store.

• manager_staff_id: A foreign key identifying the manager of this store.

• address_id: A foreign key identifying the address of this store.

• last_update: When the row was created or most recently updated.

➢ Key Functionality of the Project:

The key functionalities of this Power BI project for SAKILA DVD MOVIE RENTAL:

Data Integration and Transformation:

Importing and integrating data from multiple tables (Customers, RENTAL, CATEGORY, etc.) into Power BI.

Performing data transformations such as cleaning, shaping, and merging to prepare the data for analysis.

Visual Data Exploration:

Creating interactive visualizations to explore key performance metrics such as sales trends, customer segmentation, inventory levels, and employee performance.

Utilizing various chart types (line charts, bar charts, pie charts, etc.) to present data in a meaningful and visually appealing manner.

Implementing slicers, filters, and drill-down capabilities to allow users to dynamically interact with the data and gain deeper insights.

Sales Analysis:

Analyzing sales data to identify trends, patterns, and opportunities for growth. Calculating key sales metrics such as total revenue, average order value, and sales by product category or region.

Visualizing sales performance over time and comparing it with historical data to track progress.

Customer Segmentation:

Segmenting customers based on demographic data (location, industry, etc.) to better understand their behavior and preferences.

Creating visualizations to show the distribution of customers across segments and analyze their purchasing patterns.

Identifying high-value customers and developing targeted marketing strategies to enhance customer engagement and retention.

Interactive Dashboards and Reports: Designing interactive dashboards with a user-friendly layout and intuitive navigation. Incorporating storytelling elements to guide users through the data and highlight key insights.
Enabling users to customize views, apply filters, and drill down into specific data points for deeper analysis.

Report Sharing and Collaboration:
Publishing the Power BI report to the Power BI Service for online access. Sharing the report with relevant stakeholders and teams within the organization. Collaborating on data analysis and decision-making by providing access to the report and facilitating discussions around the insights generate

➢ Aim and Objective:

Consolidate Data: Integrate and consolidate data from multiple tables within the SAKILA DATBASE to provide a comprehensive view of the company's operations.
Analyze Sales Patterns: Analyze sales data to identify trends, patterns, and fluctuations in sales performance over time, by product, region, and customer segment.
Understand Customer Behavior: Analyze customer data to understand buying behavior, preferences, and demographics, enabling targeted marketing strategies and improved customer engagement.
Optimize Inventory Management: Monitor inventory levels, analyze stock movement, and identify inventory trends to optimize inventory management processes and reduce carrying costs.

Evaluate Employee Performance: Assess employee sales performance, productivity, and effectiveness to identify top performers, training needs, and opportunities for improvement.

Provide Actionable Insights: Present insights derived from data analysis in a visually appealing and intuitive dashboard format, enabling stakeholders to make informed decisions and take proactive actions.

Foster Data-Driven Culture: Promote a culture of data-driven decision-making within the organization by providing access to timely and relevant data, empowering stakeholders at all levels to leverage data for strategic planning and operational improvements.

Facilitate Continuous Improvement: Continuously monitor and analyze key performance metrics, gather feedback from stakeholders, and iterate on the Power BI dashboard to ensure it remains relevant, insightful, and aligned with business objectives.

By achieving these objectives, the Power BI project aims to empower SAKIALA DVD MOVIE RENTAL with actionable insights, enabling them to make informed decisions, drive business growth, and maintain a competitive edge in the marketplace.

DATA CLEANING

Data cleaning is the process of identifying and correcting or removing errors, inconsistencies, and inaccuracies in data. It is an important step in data analysis as it ensures that the data is accurate, complete, and consistent. Data cleaning is used to improve the quality of data and to make it suitable for analysis. It involves identifying missing values, removing duplicates, correcting spelling errors, and standardizing data formats. Data cleaning is essential because it helps to ensure that the insights generated from the data are reliable and accurate.

Some common challenges faced during data cleaning include:

Missing Data: Dealing with missing values in the dataset, which may require imputation or removal of incomplete records.

Inconsistent

Data: Addressing inconsistencies in data formats, such as date formats, currency symbols, or units of measurement.

Duplicate Records: Identifying and removing duplicate entries, which can skew analysis and lead to inaccurate results.

Outliers: Handling outliers that can significantly impact statistical analysis and visualization.

Data Standardization: Ensuring that data is consistent and standardized across different sources or systems.

Data Validation: Verifying the accuracy and integrity of the data, which may involve cross-referencing with external sources or known benchmarks.

Data Transformation: Converting data into a suitable format for analysis, such as aggregating, pivoting, or normalizing data.

Data Quality: Ensuring the overall quality of the data, including accuracy, completeness, and reliability.

These challenges are common in the data cleaning process and require careful attention to detail to ensure the accuracy and integrity of the data for analysis.

Process of Cleaning in North Wind dataset tables:

- **Address table:**

  . Removed address2 column as it contained no values.

  .In district column empty values are replaced by N/A.

  .In PostalCode column the empty values are replaced by -1.

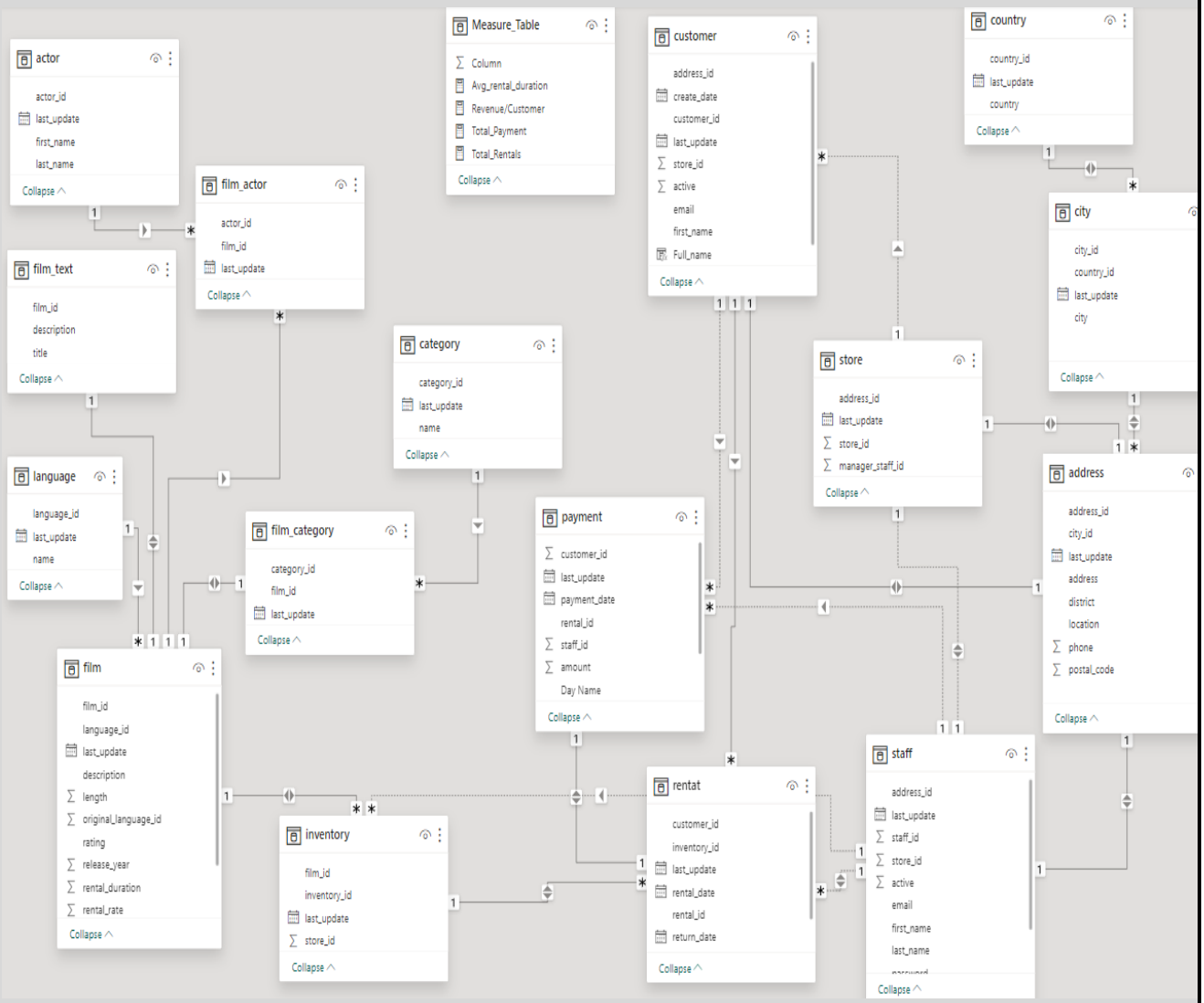
  .In Phone column the empty values are replaced by -1.

After cleaning all the table are used for the visualization in power BI and build a relationship between the tables:

All the tables are connected with each other in a one-to-many, one to one and many to one relationships.

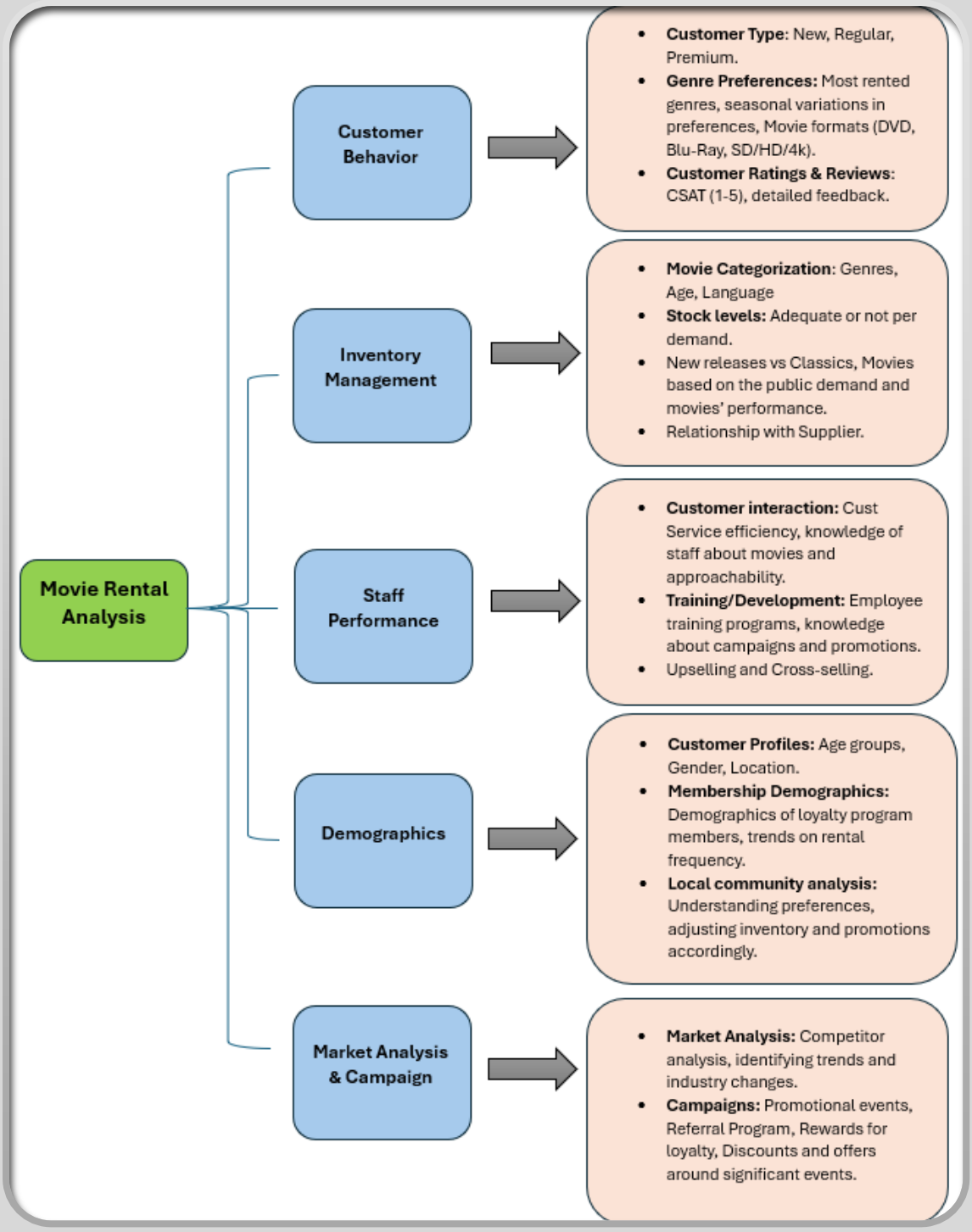
Overview of Transformed dataset

Entity Relationship Diagram



MECE-Breakdown:

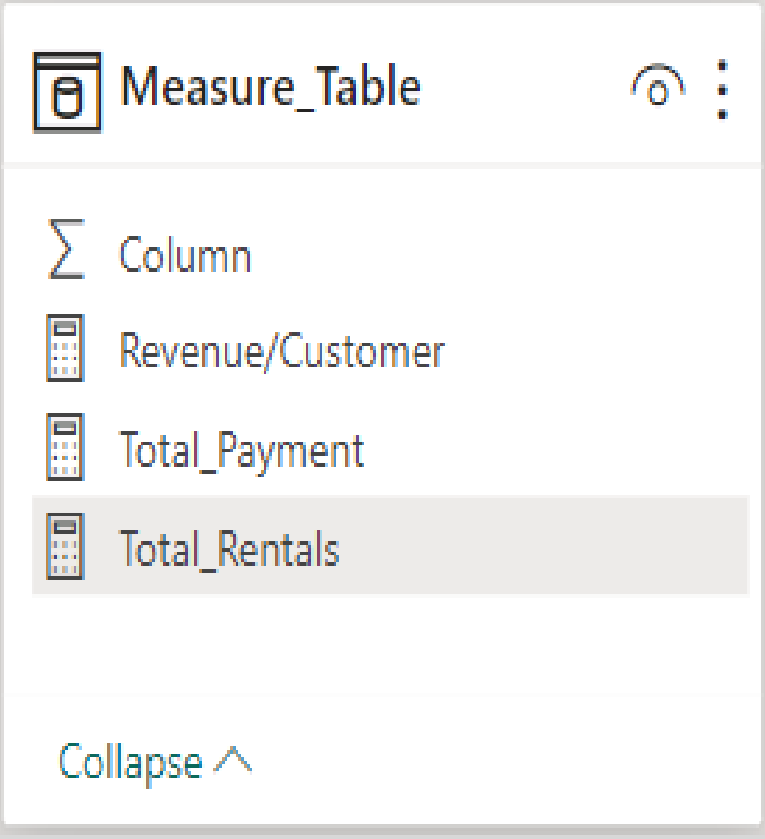
MECE stands for "Mutually Exclusive, Collectively Exhaustive." It's a principle often used in problem-solving and structuring information to ensure that all elements being considered are distinct from on another (mutually exclusive) and that together they cover all possible options (collectively exhaustive).



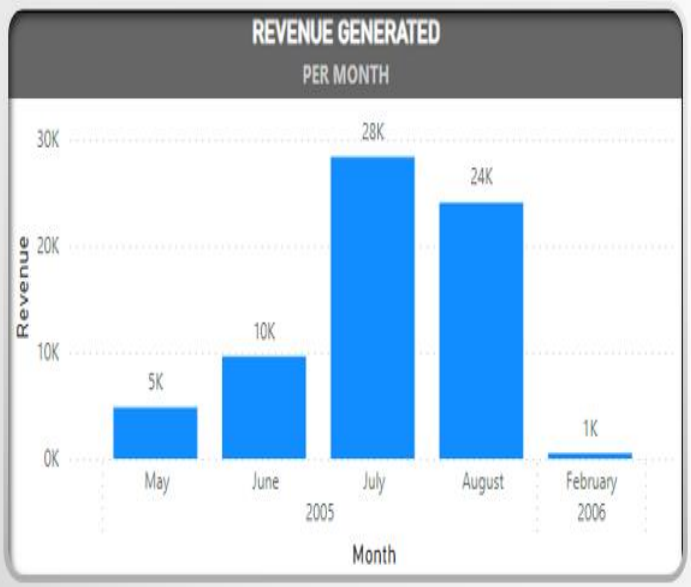
UNDERSTANDING & SOLVING POWER BI QUESTION

Questions of Power BI.

- **Before doing any analysis, I have created a table which contains all the Measures used in Analysis in Power BI**

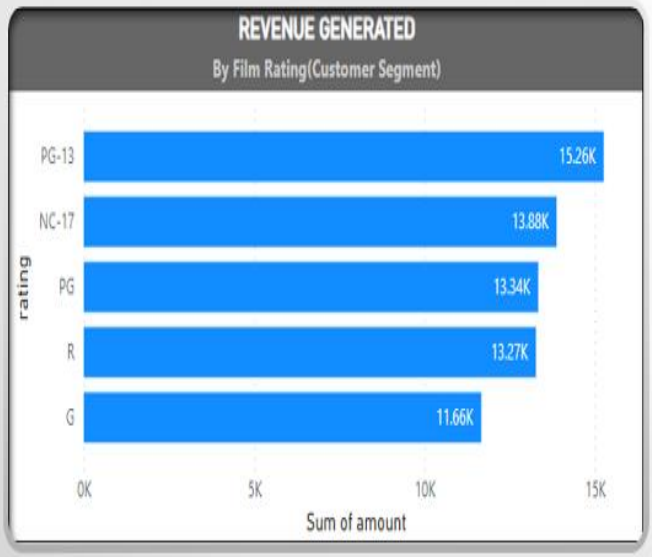


How does the sales revenue vary by month?



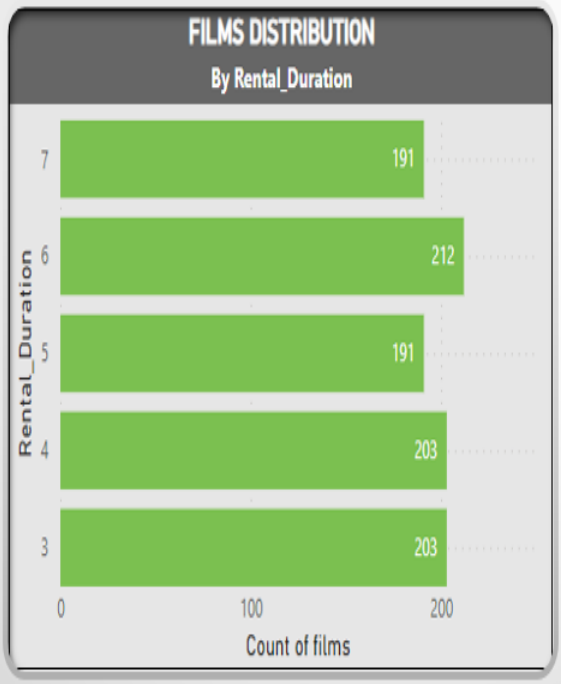
The above visual shows the variation in revenue for each available rental month.

Which customer segments generate the highest sales?



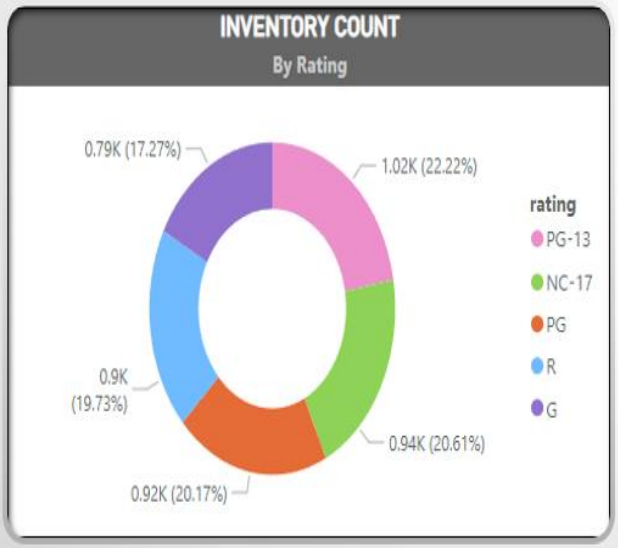
Considering that the film rating is synonymous to or directly corresponds to Customer Segment (basis of age), the greatest contributor to Sales is by PG-13 segment.

What is the distribution of films by rental duration?



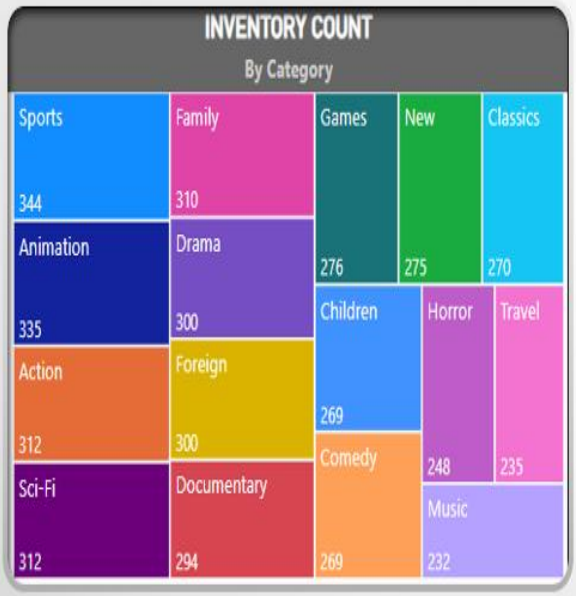
The Stacked Bar Chart gives us a complete idea to understand the distribution of films by rental duration.

How does the inventory vary by film rating?

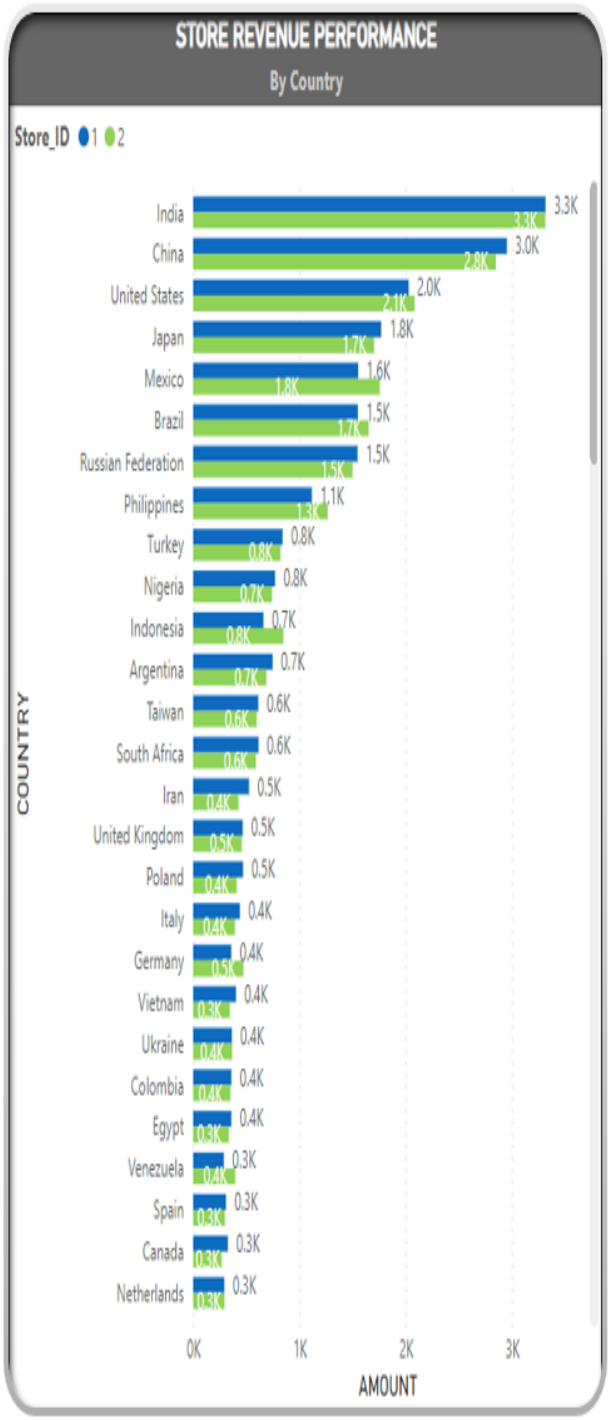


The slices of this donut chart give us a complete idea to understand partitions and percentage of inventory divided by rating of films.

What is the breakdown of film categories in the inventory?



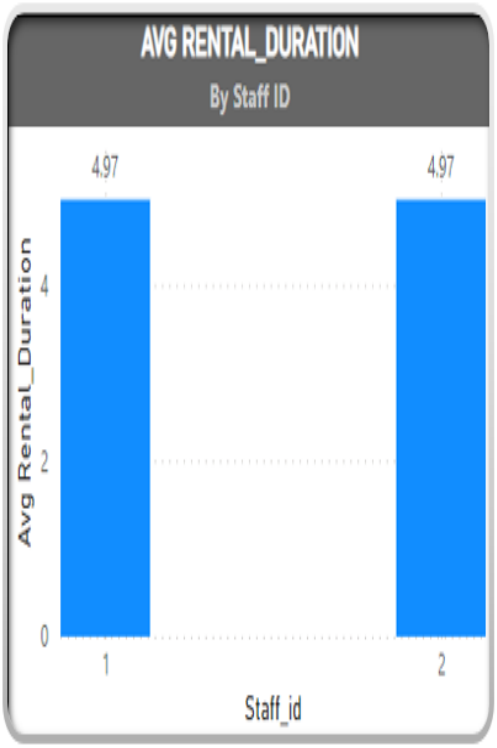
Each part of the Tree map gives us a complete idea to understand the count of inventory for each film category.



How does the store performance vary by location?

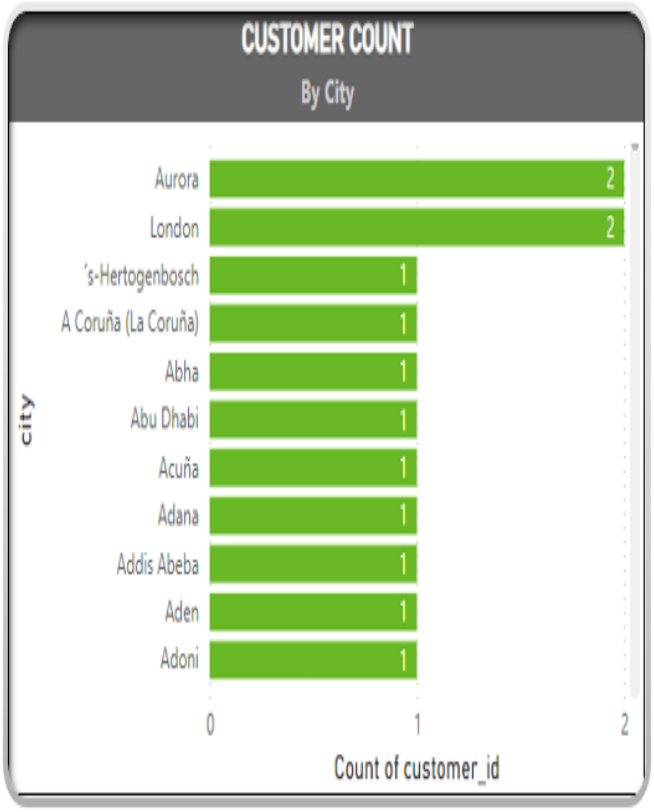
The Clustered Bar Chart gives us a complete idea to understand the earning revenue structure based on country for each store.

What is the average rental duration by staff member?

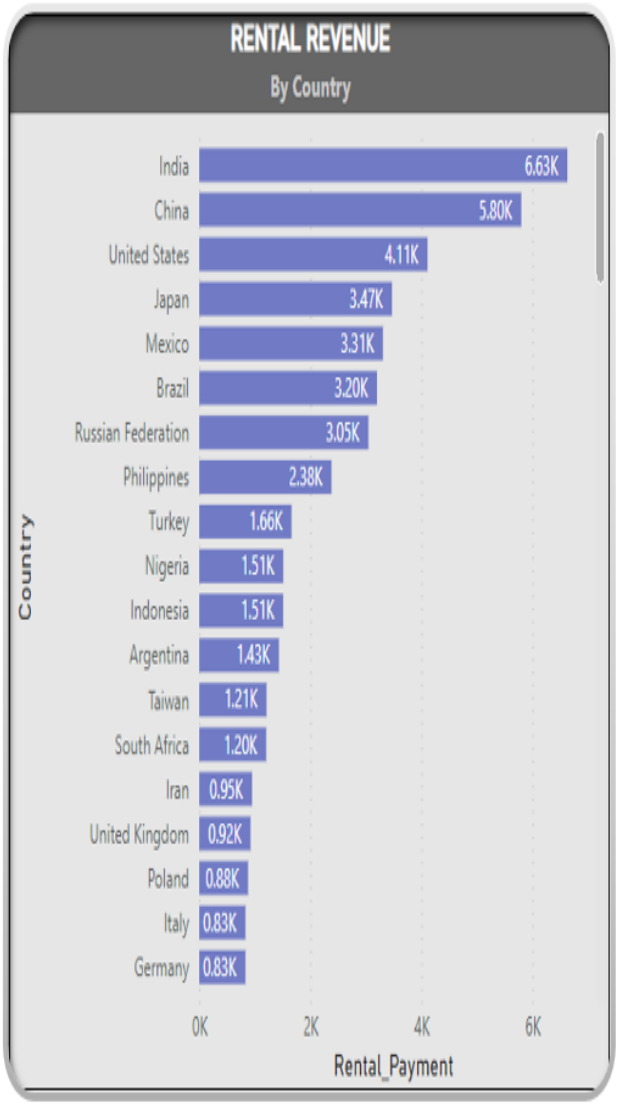


This Column Chart gives us a complete idea that the average rental duration of each staff are very same.

What is the distribution of customers across different cities?



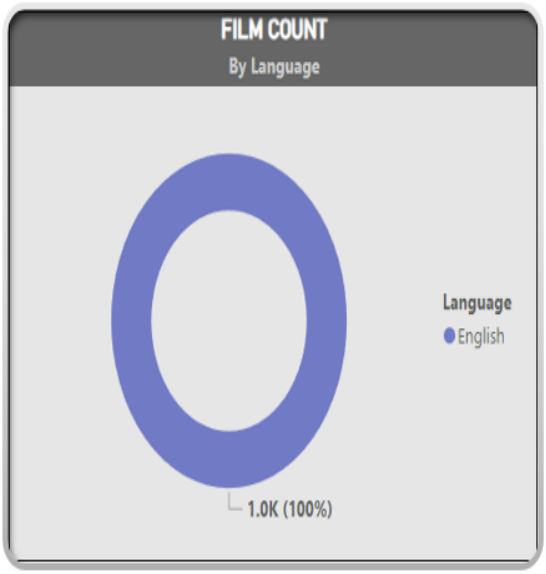
The Stacked Bar Chart gives us the count of customers for each city.



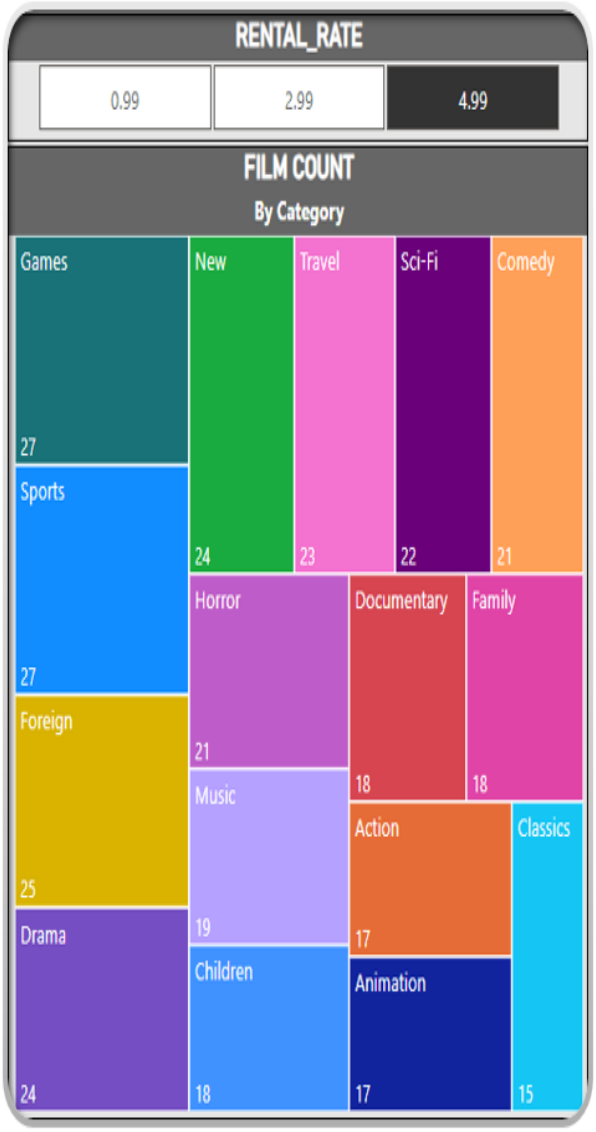
How does the rental revenue vary by country?

This Clustered Bar Chart helps us to understand the Revenue generated for the mentioned countries.

What is the distribution of films by language?

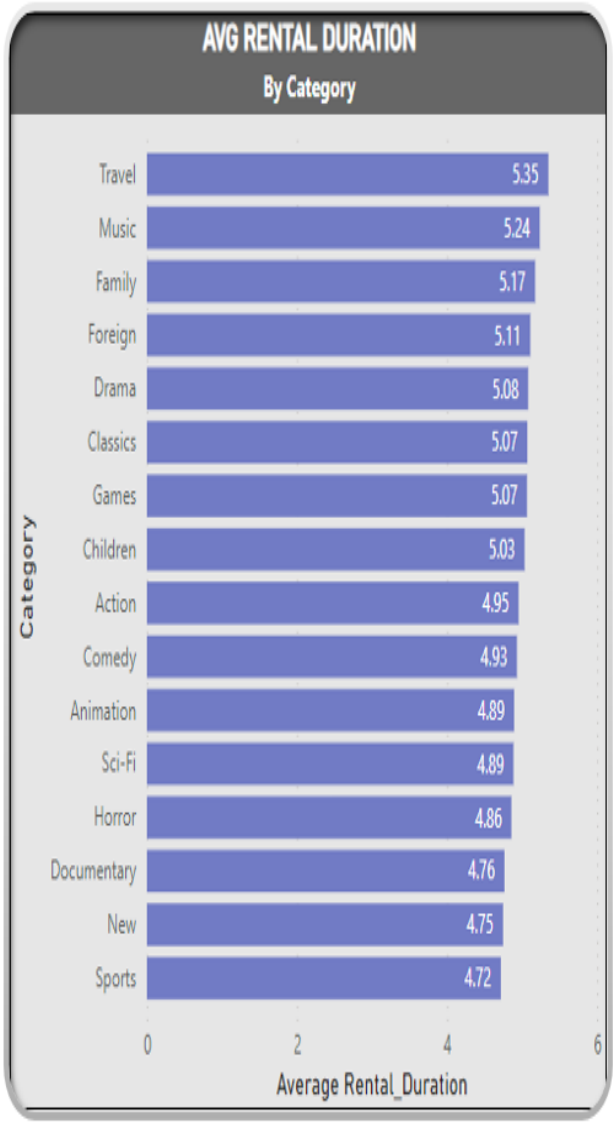


As clearly shown on the Donut Chart, all the films which were rented are in English language.



Which film categories have the highest rental rates?

This Tree Map shows that the movie category(s) with the highest rental rate (4.99) are Games and Sports.

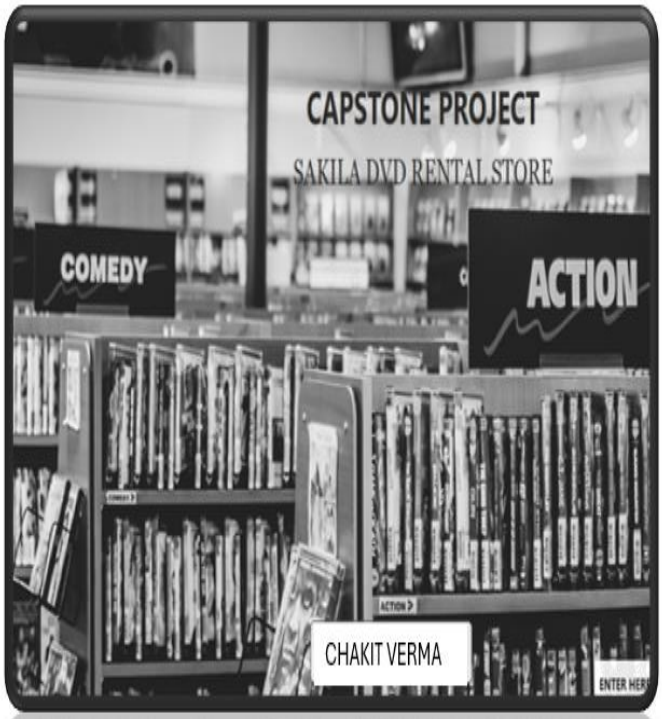


How does the average rental duration vary by film category?

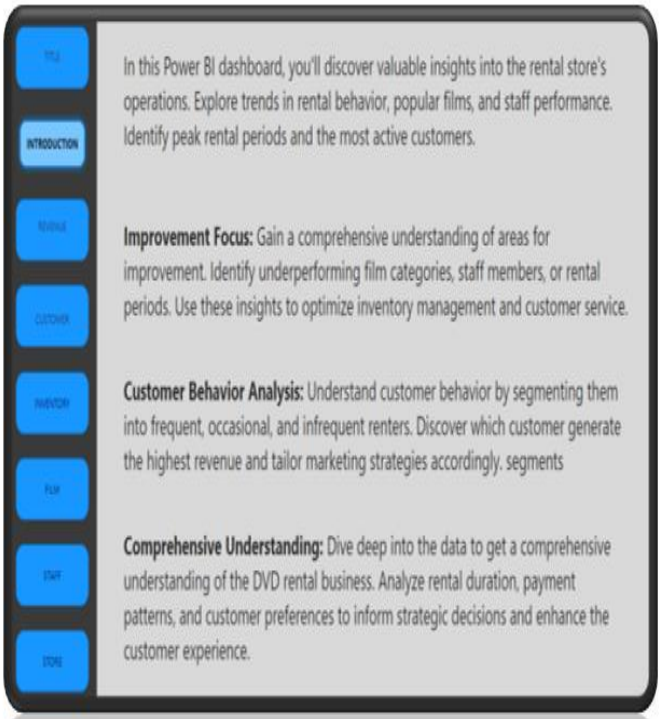
This Bar Chart gives us the average rental duration for each film category.

Dashboard Analysis

TITLE PAGE



INTRODUCTION PAGE



Dashboard Analysis

REVENUE SUMMARY

CUSTOMER SUMMARY

Dashboard Analysis

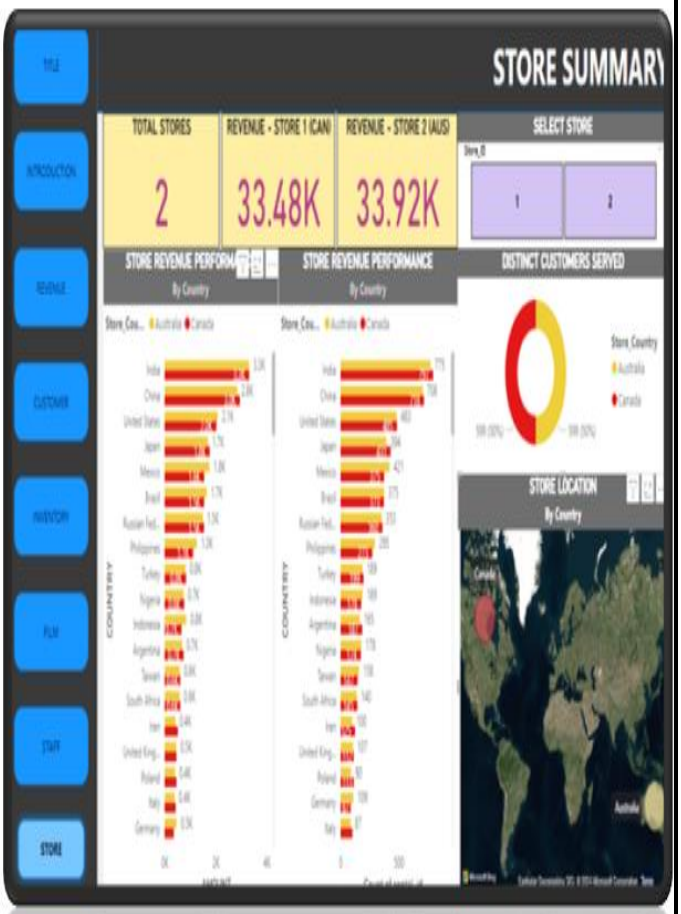
INVENTORY SUMMARY

FILM SUMMARY

Dashboard Analysis

STAFF SUMMARY

STORE SUMMARY



Conclusion

The Sakila DVD rental store project reveals customer preferences and rental trends, guiding inventory management and marketing strategies. Insights into revenue and operational efficiency suggest avenues for growth and cost reduction. Implementing a recommendation system could enhance the customer experience, while digital solutions streamline operations.

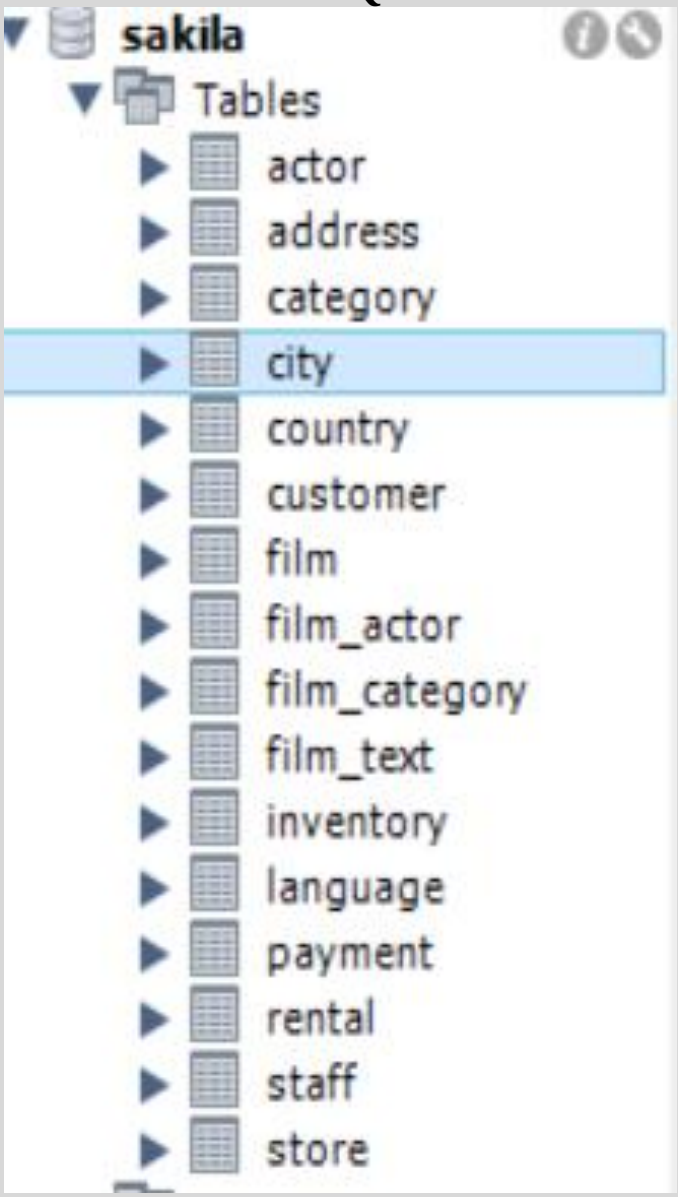
Overall, leveraging data-driven strategies positions Sakila to thrive in the competitive entertainment market.

UNDERSTANDING & SOLVING EDA QUESTIONS

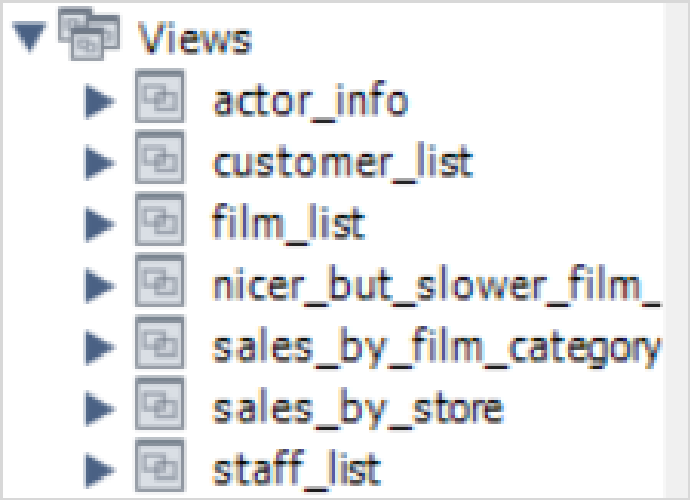
Questions of EDA

Before starting solving the EDA questions :

- **Load the Dataset SQL file into the SQL server.**



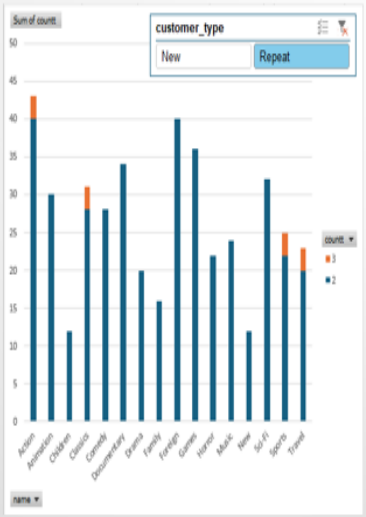
- **After loading created a view in SQL server of each table that are necessary for solving the EDA Questions.**



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

```sql
select i.film_id, customer_id, c.name, count(*) as countt, case when count(*)=1 then 'New' else 'Repeat' end as customer_type
from rental r
join inventory i
on i.inventory_id=r.inventory_id
join film_category fc
on fc.film_id=i.film_id
join category c
on c.category_id=fc.category_id
group by 1,2,3
order by 4 desc
```

Sum of countt	Column Labels		
Row Labels	2	3	Grand Total
Action	40	3	43
Animation	30		30
Children	12		12
Classics	28	3	31
Comedy	28		28
Documentary	34		34
Drama	20		20
Family	16		16
Foreign	40		40
Games	36		36
Horror	22		22
Music	24		24
New	12		12
Sci-Fi	32		32
Sports	22	3	25
Travel	20	3	23
Grand Total	416	12	428

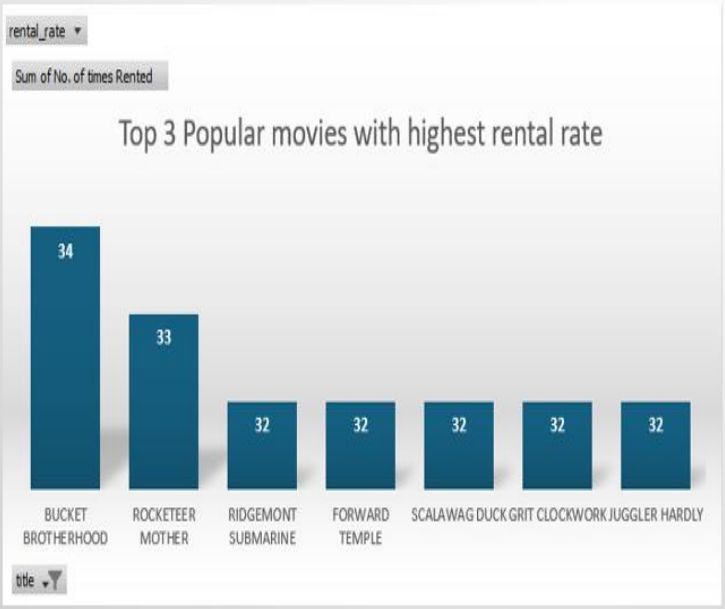


Taking Film as point of reference, Repeated Customers have rented mostly from Action, Classics, Sports and Travel. Whereas New customers have only rented from Action and Travel

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

```sql
select title, rental_rate, count(*) as "No. of times Rented"
from film f
join inventory i
on i.film_id=f.film_id
join rental r
on r.inventory_id=i.inventory_id
group by 1,2
order by 2 desc , 3 desc
```

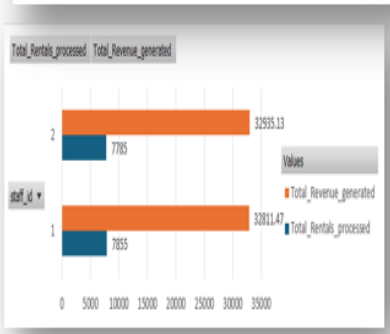
rental_rate	(All)
Row Labels	**Sum of No. of times Rented**
BUCKET BROTHERHOOD	34
ROCKETEER MOTHER	33
RIDGEMONT SUBMARINE	32
FORWARD TEMPLE	32
SCALAWAG DUCK	32
GRIT CLOCKWORK	32
JUGGLER HARDLY	32
Grand Total	**227**



The movie with the highest rental rate : 4.99 and which was rented the most (in demand) is Bucket Brotherhood

Are there correlations between staff performance and customer satisfaction?

```sql
select r.staff_id, count(*) as Rentals_processed, count(distinct r.customer_id) as Distinct_cust_served,
sum(amount) as Revenue_generated
from rental r
join payment p
on p.rental_id=r.rental_id
where r.customer_id not in (select customer_id from customer where active = 0)
group by 1
order by 4 desc
```



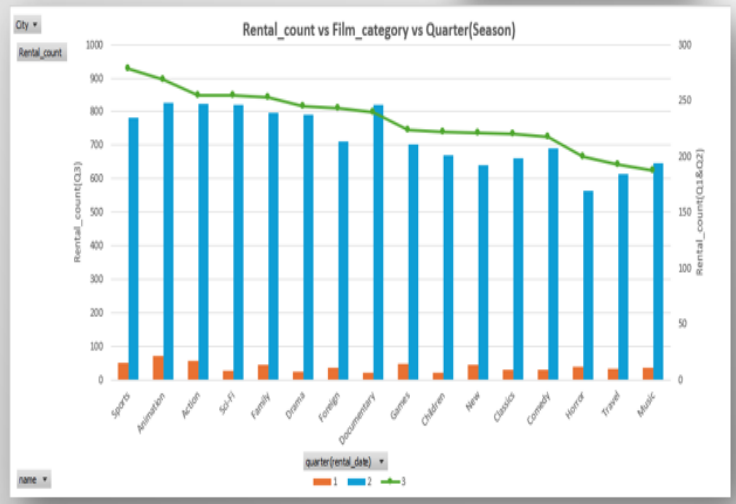
Row Labels	Total_Rentals_processed	Total_Revenue_generated
1	7855	32811.47
2	7785	32935.13
Grand Total	15640	65746.6

staff_id	Rentals_processed	Distinct_cust_served	Revenue_generated
2	7785	584	32935.13
1	7855	584	32811.47

As seen from the charts, having served the same numer of distinct active customers (584)- staff_id 1 has processed higher number of rentals, whereas staff_id 2 has generated higher rental income. So, we can say that there is no correlation between staff performance(attracting more distinct customers or processing more rentals or generating more revenue) and customer satisfaction.

```sql
select city, quarter(rental_date), count(r.customer_id)
from rental r
join customer c
on c.customer_id = r.customer_id
join address a
on a.address_id=c.address_id
join city on city.city_id=a.city_id
group by 1,2
order by 1 asc
```

City	(All)		
Rental_count	Column Labels		
Row Labels	1	2	3
Sports	15	235	929
Animation	21	248	897
Action	17	247	848
Sci-Fi	8	246	847
Family	13	239	844
Drama	7	237	816
Foreign	11	213	809
Documentary	6	246	798
Games	14	211	744
Children	6	201	738
New	13	192	735
Classics	9	198	732
Comedy	9	207	725
Horror	12	169	665
Travel	10	184	643
Music	11	194	625

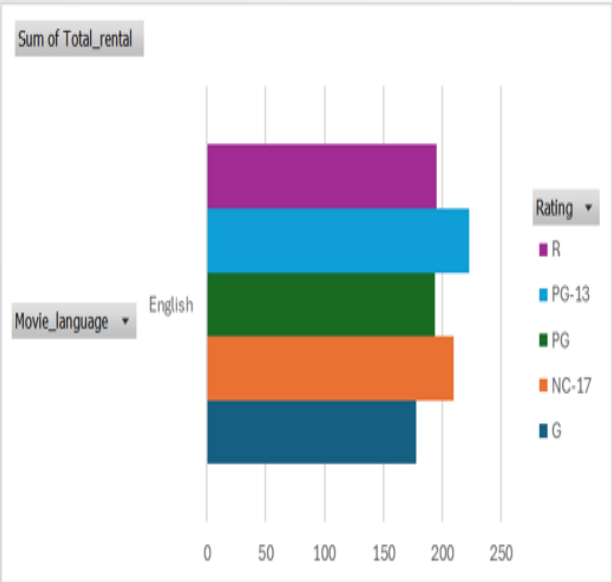


Are there seasonal trends in customer behavior across different locations?

Quarter 3(Q3) has shown the highest rental count for all cities combined as well as for individual cities too. Hence, Q3 season exhibits the most rental whereas Q1 season exhibits the least rentals

```
select Rating, l.name as Movie_language, count(f.rental_duration) as Total_rental from language l
join film f on l.language_id=f.language_id
join film_category fc on f.film_id=fc.film_id
join category c on fc.category_id=c.category_id
group  by 1,2
order by 3 desc
```

Sum of Total_rental	Column Labels				
Row Labels	G	NC-17	PG	PG-13	R
English	178	210	194	223	195



Are certain language films more popular among specific customer segments?

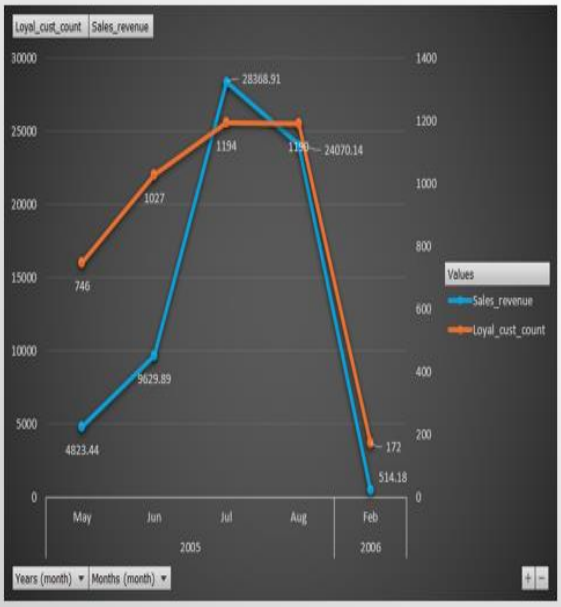
English is the only language available for movies in the dataset. So, it is the most popular language across all customer segments(calculated based on movie rating, as no age data given for the customers)

```sql
SELECT DATE_FORMAT(p.payment_date, '%Y-%m') AS month, s.store_id,
    COUNT(distinct c.customer_id) AS loyal_customer_count,
    SUM(p.amount) AS total_sales_revenue
FROM payment p
JOIN customer c ON p.customer_id = c.customer_id
join rental r on p.rental_id=r.rental_id
join staff s on s.staff_id=r.staff_id

GROUP BY 1,2
ORDER BY 1;
```

store_id: 1 | 2

Row Labels	Loyal_cust_count	Sales_revenue
2005	4157	66892.38
May	746	4823.44
Jun	1027	9629.89
Jul	1194	28368.91
Aug	1190	24070.14
2006	172	514.18
Feb	172	514.18
Grand Total	4329	67406.56



How does customer loyalty impact sales revenue over time?

As shown in the chart, with the increase in loyal customers - the sales revenue also increases over time.

Are certain film categories more popular in specific locations?

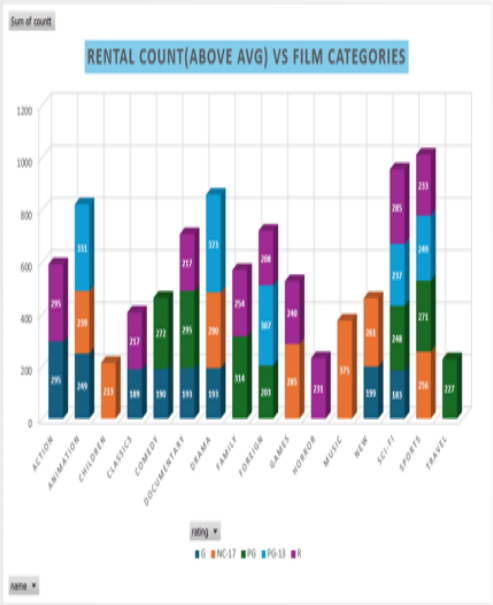

```sql
with cte as (
select country, City, cc.name as Category_Name, count(*) as countt,
dense_rank() over(partition by Country order by count(*) desc) as RK
from rental r
join inventory i
on i.inventory_id=r.inventory_id
join film f
on f.film_id=i.film_id
join film_category fc
on fc.film_id=f.film_id
join category cc
on cc.category_id=fc.category_id
join customer c
on c.customer_id = r.customer_id
join address a
on a.address_id=c.address_id
join city on city.city_id=a.city_id
join country on country.country_id=city.country_id
group by 1,2,3
order by 2 asc, 4 desc
)
select Country,City, Category_Name, countt
from cte
where RK=1
```

Row Labels	Max(Rented)
Turkey	45
Colombia	28
United Arab Emirates	24
Bolivia	24
Japan	24

The above graph shows the Top 3 Value Countries (based on max rental times in a country) for the selected movie category

Do specific film categories attract different age groups of customers?

Sum of countt	Column Labels			
Row Labels	G	NC-17	PG	PG-13	R
Action	295			295
Animation	249	239		331
Children		213		
Classics	189			217
Comedy	190		272	
Documentary	193		295	217
Drama	193	290		373
Family			314	254
Foreign			203	307	208
Games		285		240
Horror				231
Music		375		
New	199	261		
Sci-Fi	183		248	237	285
Sports		256	271	249	233
Travel			227	

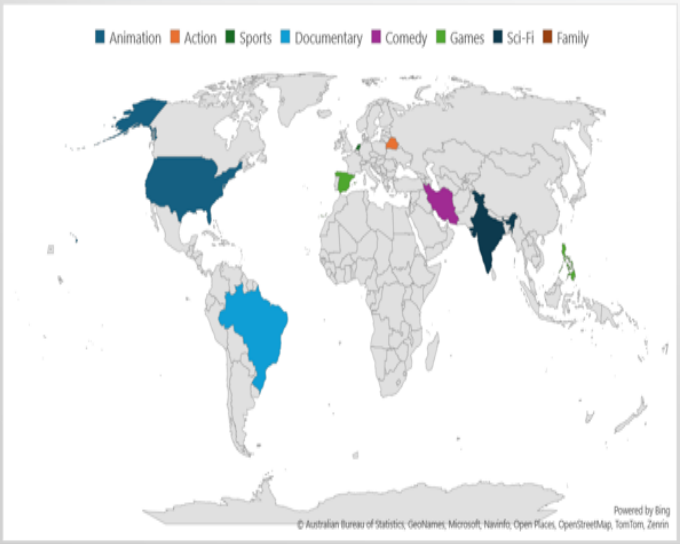



```
with cte as (select distinct ca.name,count(rental_id) as countt,f.rating,
case when rating = 'G' then '1' when rating = 'PG' THEN '15' when rating = 'PG-13' THEN '13'when rating = 'R' THEN '18'when rating = 'NC-17' THEN '17' else '0' end as Above_age,
avg(count(rental_id)) over(partition by rating ) as avgg
from film f
join film_category fc on f.film_id = fc.film_id
join category ca on fc.category_id=ca.category_id
join inventory i on i.film_id=f.film_id
join rental r on r.inventory_id=i.inventory_id
Group By 1,3,4
order by 2 desc
)
select * from cte
where countt>avgg
order by name
```

This chart represents the age groups (based on Ratings - for eg. R is for 18+) present for each Film Category, which has a rental count of more than the average rental count for that age group

```sql
with cte as (
select c.customer_id, concat(c.first_name," ",c.last_name)as Customer_name, co.country, ca.name as Category_name, sum(amount) as Revenue
from payment p
join customer c on c.customer_id=p.customer_id
join address a on c.address_id=a.address_id
join city ci on ci.city_id=a.city_id
join country co on ci.country_id=co.country_id
join rental r on r.rental_id=p.rental_id
join inventory i on i.inventory_id=r.inventory_id
join film f on f.film_id=i.film_id
join film_category fc on fc.film_id=f.film_id
join category ca on ca.category_id=fc.category_id
group by 1,2,3,4
order by 1 asc
)
select cte.*, sum(Revenue) over(partition by customer_id ) as Running_total
from cte
order by 6 desc
```

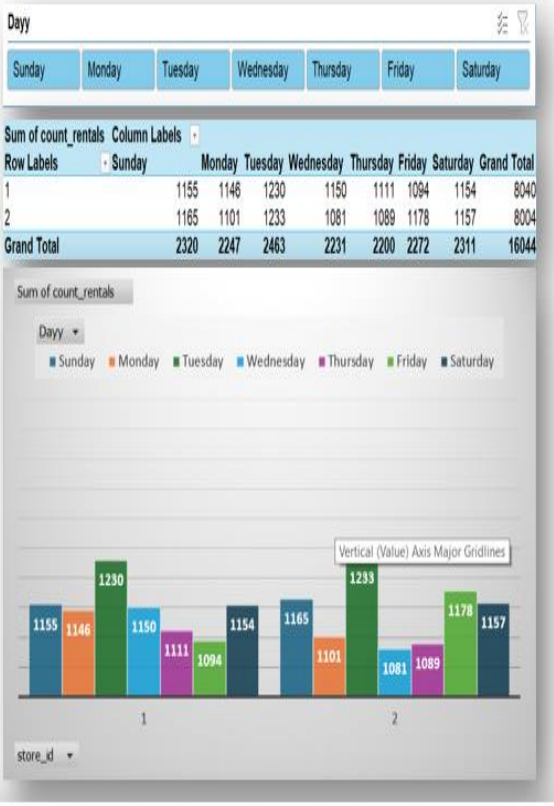
Customer_name	country	Category_name	Sum of Revenue	Sum of Running_total
KARL SEAL	United States	Animation	38.92	221.55
ELEANOR HUNT	Runion	Action	28.97	216.54
CLARA SHAW	Belarus	Action	30.96	195.58
RHONDA KENNEDY	Netherlands	Sports	28.95	194.61
MARION SNYDER	Brazil	Documentary	23.96	194.61
TOMMY COLLAZO	Iran	Comedy	38.95	186.62
WESLEY BULL	Spain	Games	32.91	177.6
TIM CARY	India	Sci-Fi	23.97	175.61
MARCIA DEAN	Philippines	Games	24.95	175.58
ANA BRADLEY	United States	Family	31.93	174.66
Grand Total			304.47	1912.96



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

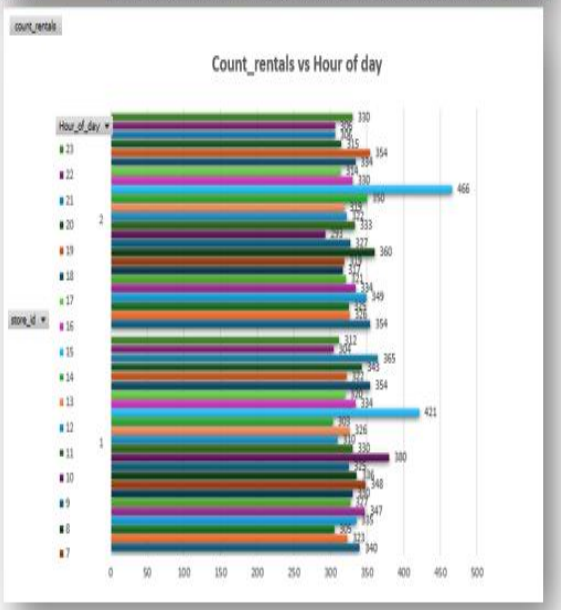
The above pivot table shows the Top 10 high spending customers, their country, their most rented category, the amount they've paid for the most rented category and the overall rental amount they have paid.

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



Sum of count_rentals	Column Labels						
Row Labels	Sunday	Monday	Tuesday	Wednesday	Thursday	Friday	Saturday	Grand Total
1	1155	1146	1230	1150	1111	1094	1154	8040
2	1165	1101	1233	1081	1089	1178	1157	8004
Grand Total	2320	2247	2463	2231	2200	2272	2311	16044

count_rentals	Column Labels																							
Row Labels	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23
1	340	323	305	335	347	327	330	348	336	325	380	330	310	326	303	421	334	320	354	322	343	365	304	312
2	354	326	325	349	334	321	317	319	360	327	293	333	322	319	350	466	330	314	334	354	315	306	308	330
Grand Total	694	649	630	684	681	648	647	667	696	652	673	663	632	645	653	887	664	634	688	676	658	671	610	642




```
WITH cte AS (
    SELECT
        city, store_id, Dayname(r.rental_date) as Dayy, hour(r.rental_date) as Hourr,
        COUNT(*) AS count_rentals,
        AVG(COUNT(*)) OVER ( partition by city) AS avg_rental_per_store
    FROM
        rental r
        JOIN staff s ON s.staff_id = r.staff_id
        JOIN address a ON s.address_id = a.address_id
        JOIN city c ON c.city_id = a.city_id
    GROUP BY
        1,2,3,4
        order by 1 asc, 5 desc
)
select Cte.*, case when count_rentals > 1.10*avg_rental_per_store then "understaffed" when count_rentals < 0.70*avg_rental_per_store then "overstaffed"
else "appropriately staffed" end as Staff_availability
from cte
```

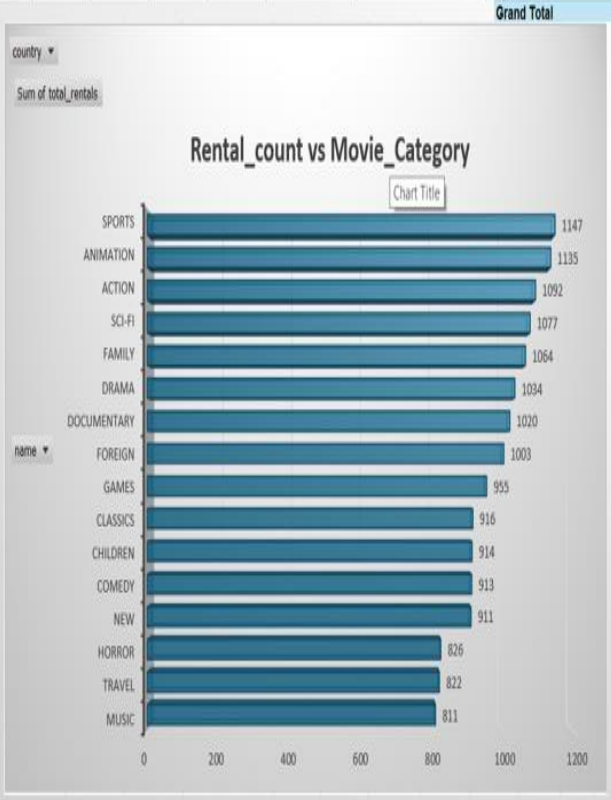
The above pivot tables show the count of rentals as per day of the week and count of rentals as per the hour of day. For both store 1 and 2, Tuesday is the busiest with the greatest number of rentals and the busiest hour of day is 15(on Tuesday as well). Based on the rental count values over different hours(when compared with the average rentals/store), we have classified staffing into three categories: Understaffed, overstaffed and appropriately staffed

```sql
SELECT c.customer_id, c.first_name, c.last_name, c.address_id,
    cc.city, co.country, ca.name, c.active, COUNT(*) AS total_rentals
FROM customer c
    join address a on c.address_id = a.address_id
    join city cc on a.city_id = cc.city_id
    join country co on cc.country_id=co.country_id
    join rental r on r.customer_id=c.customer_id
    join inventory i on r.inventory_id=i.inventory_id
    join film f on i.film_id=f.film_id
    join film_category fc on f.film_id=fc.film_id
    join category ca on ca.category_id=fc.category_id
GROUP BY 1,2,3,4,5,6,7,8
ORDER BY total_rentals DESC
```

active: 0 | 1

country: (All)

Row Labels	Sum of total_rentals
Music	811
Travel	822
Horror	826
New	911
Comedy	913
Children	914
Classics	916
Games	955
Foreign	1003
Documentary	1020
Drama	1034
Family	1064
Sci-Fi	1077
Action	1092
Animation	1135
Sports	1147
Grand Total	15640



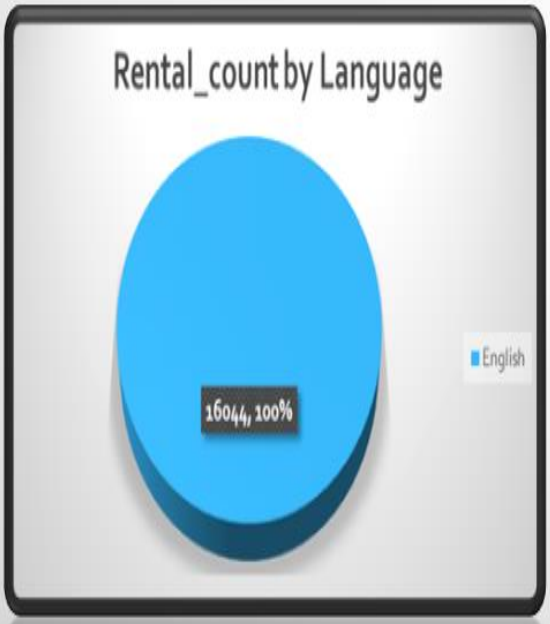
What are the cultural or demographic factors that influence customer preferences in different locations?

Upon selecting the country (from filter), we can see the Rental_Count distribution across different film categories.

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

```sql
SELECT l.name AS language, COUNT(r.rental_id) AS rental_count,
    AVG(DATEDIFF(r.return_date, r.rental_date)) AS avg_rental_duration
FROM rental r
JOIN inventory i ON r.inventory_id = i.inventory_id
JOIN film f ON i.film_id = f.film_id
JOIN language l ON f.language_id = l.language_id
GROUP BY 1
ORDER BY 2 DESC;
```

language	rental_count	avg_rental_duration
English	16044	5.0252



As seen from the Pie-chart, 100% of the movies rented were in English(as that's the only language available in the dataset). So, insufficient data to evaluate the impact of language over customer satisfaction and rental frequency.

Conclusion:

Movie Rental analytics presents a compelling avenue for extracting valuable insights and driving strategic decision-making in the entertainment industry. Through the analysis of rental patterns, customer behavior, and market trends, businesses can unlock opportunities for revenue growth, operational efficiency, and customer satisfaction.

Insightful Decision-Making:

Movie rental analytics enables businesses to make data-driven decisions that are grounded in a deep understanding of customer preferences and market dynamics. By analyzing rental trends, genre preferences, and viewer demographics, companies can tailor their offerings to meet the evolving needs and preferences of their audience.

Operational Efficiency:

Analytics empowers businesses to optimize their operations, from inventory management to distribution strategies. By identifying high-demand titles, predicting future rental trends, and optimizing pricing and promotion strategies, companies can streamline their processes and maximize resource utilization.

Enhanced Customer Experience:

Understanding customer behavior through analytics allows businesses to personalize their offerings and enhance the overall customer experience. By recommending relevant titles, offering targeted promotions, and providing seamless rental experiences, companies can build loyalty and drive repeat business.

Competitive Advantage:

Movie rental analytics provides companies with a competitive edge in a crowded marketplace. By staying ahead of industry trends, identifying emerging opportunities, and responding quickly to changing customer preferences, businesses can differentiate themselves from competitors and capture market share.

Continuous Improvement:

Analytics is not a one-time endeavor but rather an ongoing process of continuous improvement. By collecting and analyzing data on an ongoing basis, businesses can adapt to evolving market conditions, refine their strategies, and stay ahead of the competition in an ever-changing landscape.

In conclusion, movie rental analytics offers immense potential for unlocking insights, driving efficiency, and delivering exceptional value to customers. By leveraging the power of data, companies can navigate the complexities of the entertainment industry with confidence and chart a course for sustainable growth and success.