Capstone Project

Movie Rental Analytics

By

Busireddygari Sai Charan.

**Overview of the project**

We are using movie rental analytics for analysing the trends, and analysing various factors related to movie rental Analysis. The movie rental analysis dataset comprises data on movie rentals that occurred during a specific period in different countries. It encompasses a wide range of information, including details about films, actors, stores, staff, customers, sales and various other factors.

The central table in the dataset is the rental table, which functions as a transaction log, capturing comprehensive information about all rental transactions that took place throughout the specified period. The dataset contains several hundred entries for customers, films, and countries, distributed across multiple tables.

**Objective:**

In the provided dataset, which is distributed across multiple tables, the main objective is to identify and establish relationships between these tables. Once the relationships are established, Exploratory Data Analysis (EDA) will be conducted on the data.

The purpose of EDA is to uncover patterns, trends, and factors that impact the sales and volume of the movie rental business. By gaining valuable insights from the analysis, strategies can be developed to enhance sales and volume while reducing operational costs for the business.

**Data Dictionary:**

**Actor Table**

The actor table contains information for all the actors, including first name and last name of actors.

**Address Table**

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

**Category Table**

The category table contains the categories that can be assigned to films.

**City Table**

The city table contains a list of cities.

**Country Table**

The country table contains a list of countries or regions.

**Customer Table**

The customer table contains a list of all customers.

**Film Table**

The film table contains all the films that may be in stock in the store.

**Film\_text Table**

The content of the film\_text table is kept in synchrony with the film table by means of triggers on the film table INSERT, UPDATE, and DELETE operations.

**Film\_actor Table**

The film\_actor table is used to support many-to-many relationships between films and actors.

**Film\_category Table**

The film\_category table is used to support many-to-many relationships between films and categories.

**Inventory Table**

A row in the inventory table represents a copy of a given film in a given store.

**Language Table**

The language table lists all possible values for the film language and original language.

**Payment Table**

The payment table records every payment made by the customer, including information such as the amount and rent paid.

**Rental Table**

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

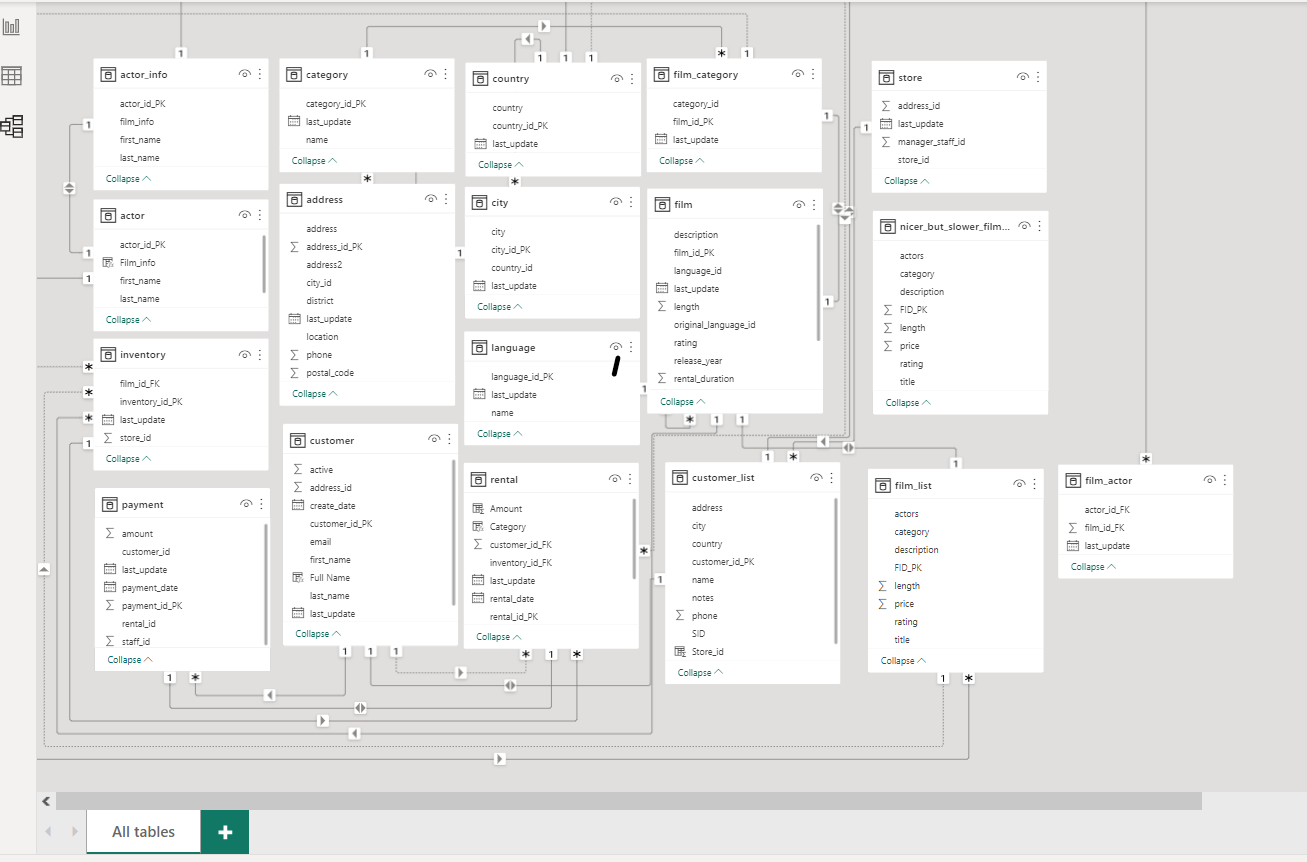
**Staff Table**

The staff table lists all staff information, including email addresses, login information, and pictures.

**Store Table**

The store table lists all stores in the system.

**E-R Diagram:**

****



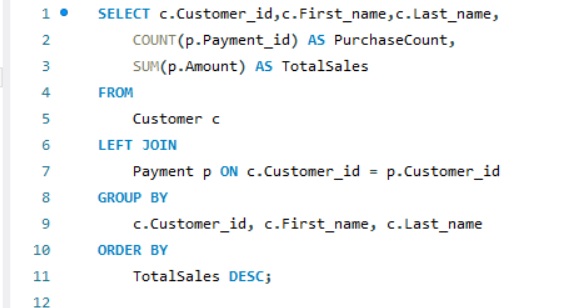
**Exploratory Data Analysis:**

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

(or) analyse the data to determine the count of purchases made by individual customers and the corresponding sales they have generated?

Since the provided dataset lacks information about customer on boarding dates and covers only a limited time span, we are unable to identify the purchasing pattern of new customers and repeat customers. However, we can analyse the data to determine the count of purchases made by individual customers and the corresponding sales they have generated.

**SQL Query:**



**Chart to analyse:**

* The total\_sales starts at its highest value and shows a general decreasing trend as we move along the horizontal axis.
* The purchase\_count remains relatively same and much lower compared to the total sales throughout the chart.

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

SQL Query:

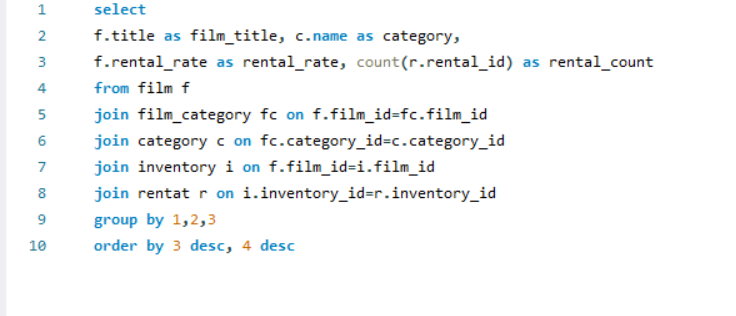


Chart to visualize:

Insights:

* The "Sports" category has the highest count,
* The lowest counts are in the "Horror" and "Classics" .

3) Are there correlations between staff performance and customer satisfaction? (or)

Analysing the performance of individual stores through metrics such as sales and rental counts.?

Since there is no provided column for measuring customer satisfaction and staff performance, responding to this query is not feasible. However, an alternative approach would involve evaluating the performance of individual stores through metrics such as sales and rental counts.

SQL query:

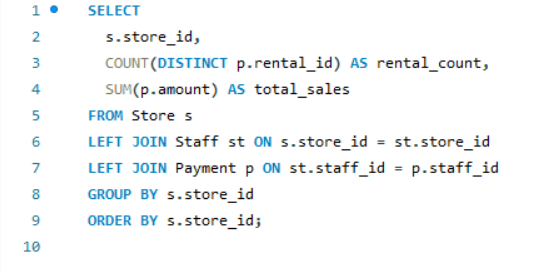


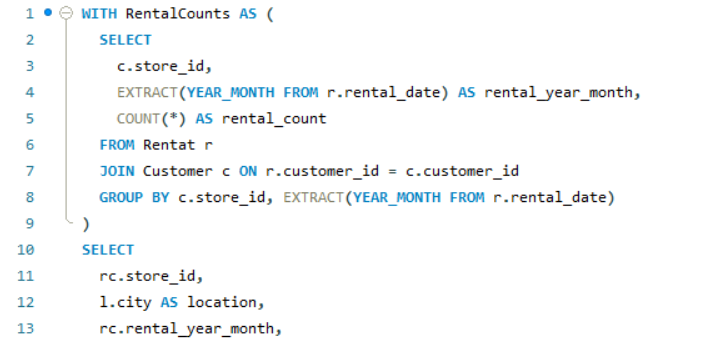
Chart to visualize

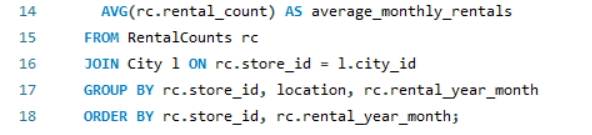
Insights:

* The "total\_sales" category has the highest values for both Series1 and Series2
* .The "rental\_count" category shows a moderate number of counts for both series, with Series1 being slightly higher than Series2.
* .The "store\_id" category has the lowest values for both series

4)Are there seasonal trends in customer behaviour across different locations?

**SQL query:**





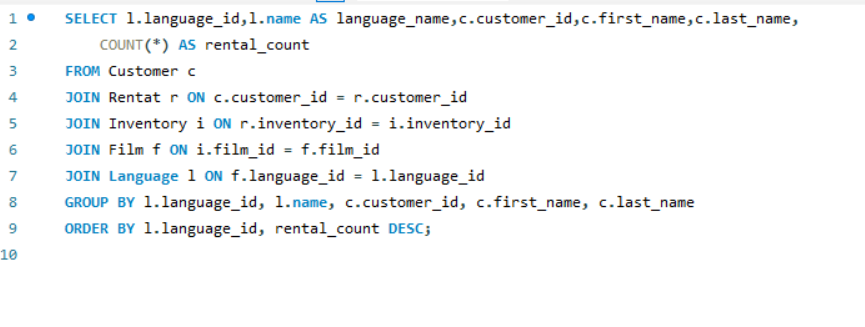
**Charts to visualize:**

**Insights:**

* The location A Coruña (La Coruña) has a higher average monthly rental figure than Abha.
* This indicates that on average, each month, A Coruña has more rental activity compared to Abha.

5) Are certain language films more popular among specific customer segments?

**SQL Query:**



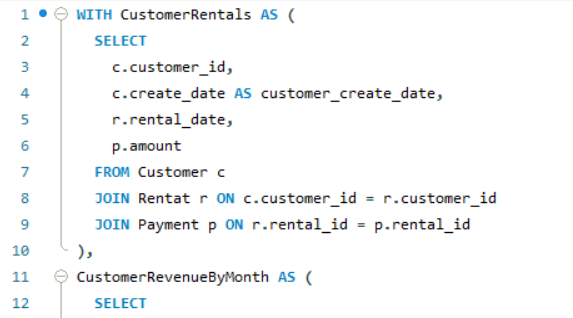
**Chart to visualize:**

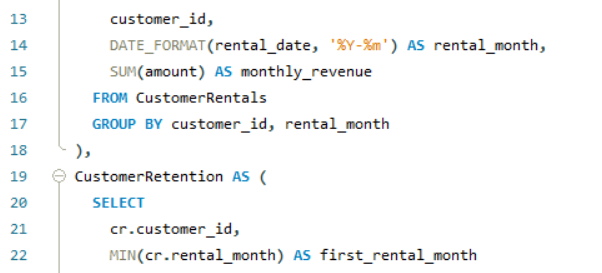
**Insights:**

* The language name count is labeled with numbers from 1 to over 500, which likely correspond to individual languages, though they are not explicitly named. The languages appear to be ordered by the number of rentals, starting with the most popular.

6)How does customer loyalty impact sales revenue over time?

SQL Query:





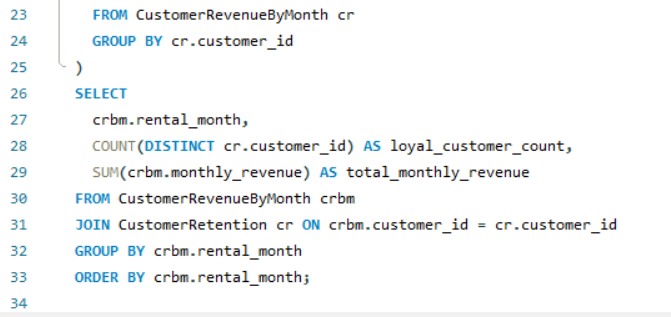


Chart to visualize:

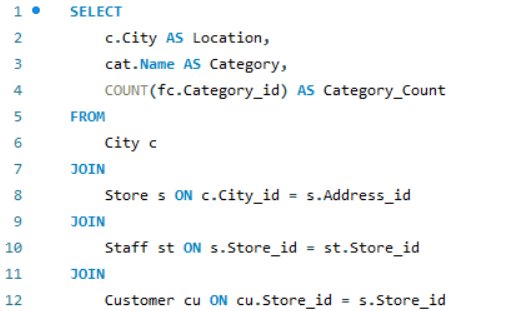
Insights:

* The highest revenue recorded was in the summer months of July and August 2005, with July being the peak.
* The number of loyal customers remains consistent at 599 in July and August 2005 but is significantly lower in February 2006 with only 158.

7) What is the distribution of staff by employment duration?

(or) \*are certain film categories more popular in specific locations?

SQL Query:



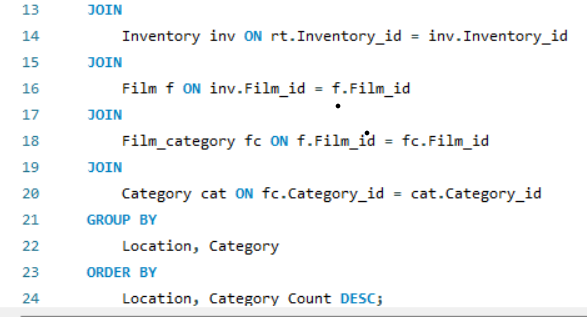


Chart to visualize:

Insights:

* Both locations have a similar distribution across the categories.
* Action and Drama categories have the highest counts in both locations.
* The counts for each category are very similar between A Coruña and Abha
* Categories like Documentary and Classics have lower counts

8) How does the store performance vary by location?

or the sales performance of films based on their respective ratings.

SQL Query:

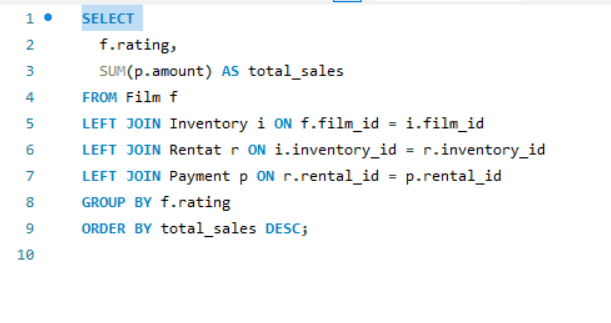


Chart to visualize:

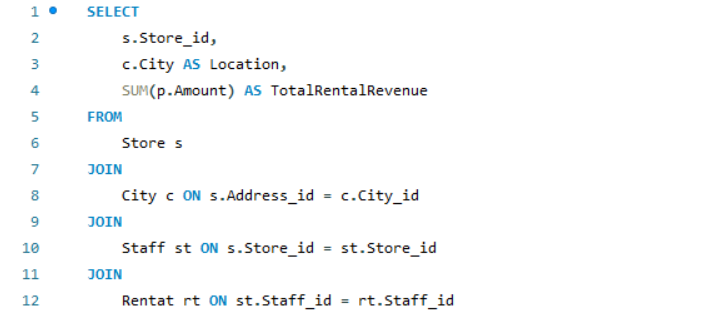
Insights:

* PG-13 rated films have the highest sales,
* Mature Content is Common: When combined, the sales of R and NC-17 rated films make up 40% of total sales, indicating a substantial market for movies with mature content.
* Family-Friendly Films: G and PG films together account for 37% of sales, showing a significant market presence for family-friendly content.

9) What is the average rental duration by staff member?

(or) calculate the rental generated for particular location by store?

SQL Query



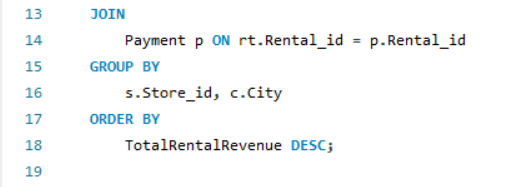


Chart to visualize:

Insights:

* Equal Revenue Share: Both locations contribute equally to the total rental revenue.

10)find the distribution of films across different categories and assess the corresponding inventory levels.

SQL Query:

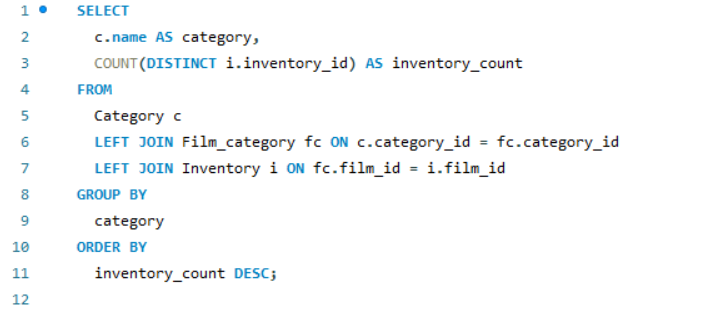


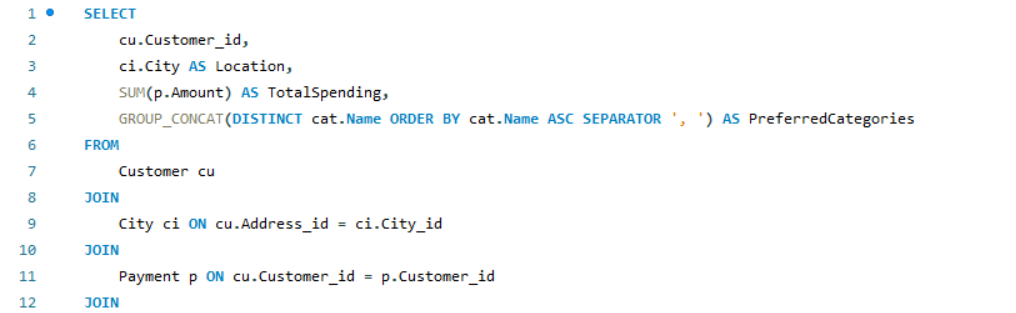
Chart to visualize:

Insights:

* Uniform Distribution: The inventory is relatively evenly distributed across the categories.
* Sports Category: The Sports category has the highest inventory count.
* Music Category: The Music category has the lowest inventory count

11) What are the location and preferences of the top 10 spending customers?

SQL Query:



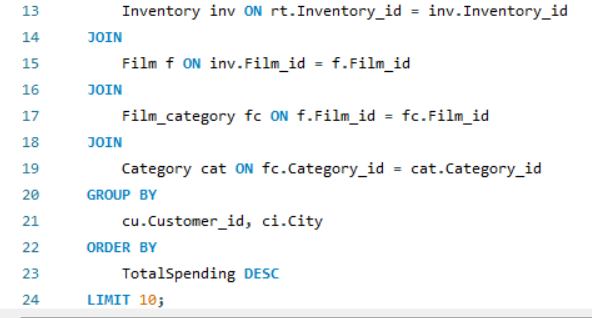


Chart to visualize:

Insights:

* customer id 1,2 are very high when comparing with other customer id’s.
* Customer id 10 is very low with comparing with other customers.

12) top 10 films by sales?

**SQL Query:**

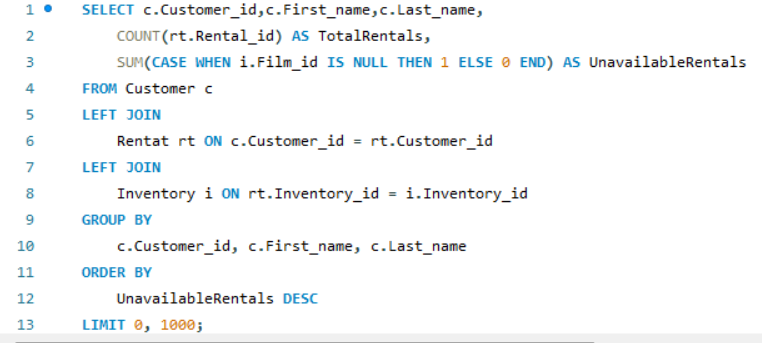


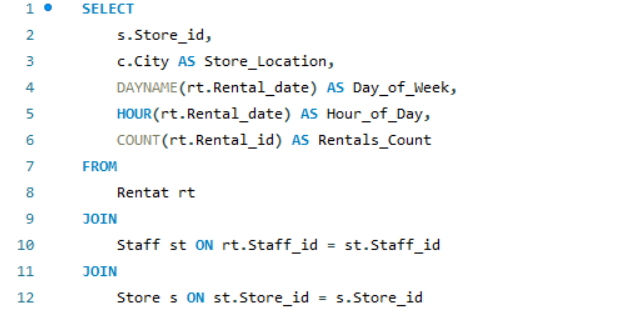
Chart to visualize:

Insights:

* Highest Sales: The item with the highest total sales is represented by the identifier "1000 ZORRO ARK," with sales of 231.73 units.
* Lowest Sales: On the lower end, "403 HARRY IDAHO" shows the smallest portion of the total sales at 190.78 units.

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

SQL Query:



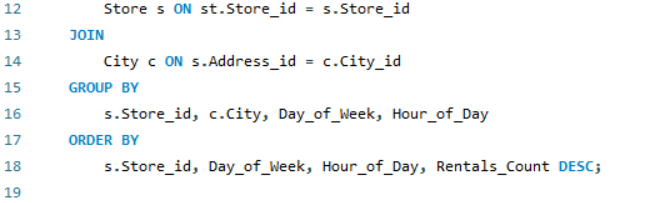


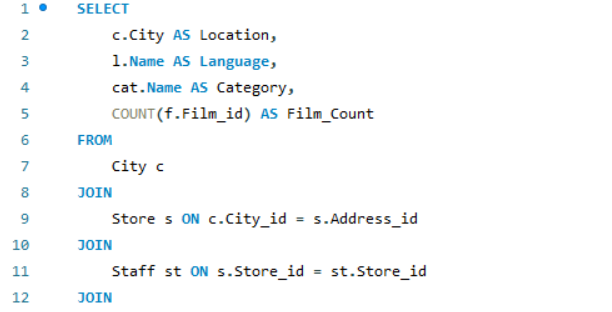
Chart to visualize:

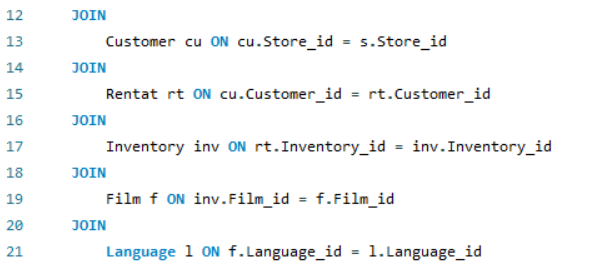
* Insights:

Both locations have a consistent pattern across the period represented

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

SQL Query:





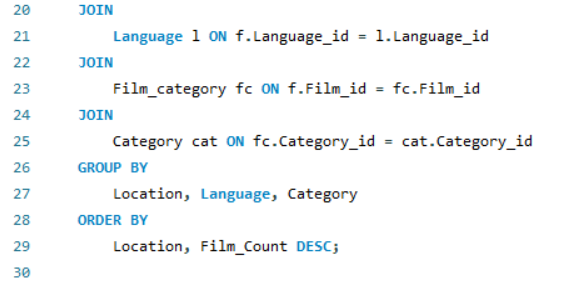
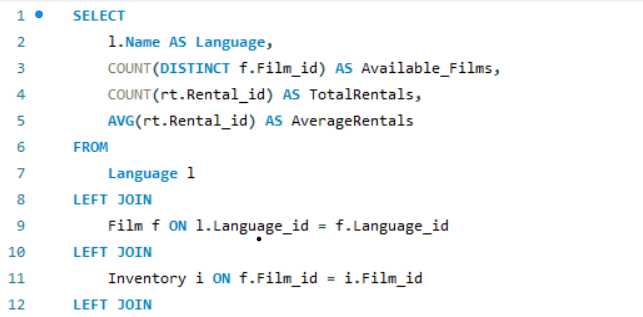


Chart to visualize:

Insights:

* Both locations have a significant number of films across all.
* Categories like Action, Animation, and Drama show some of the highest inventory counts for both locations.

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

SQL Query:  


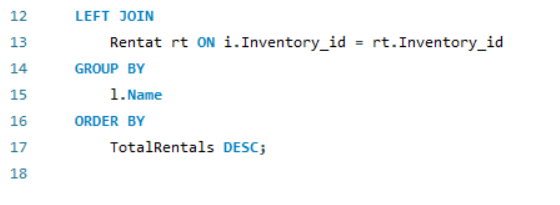


Chart to visualize:

Insights:

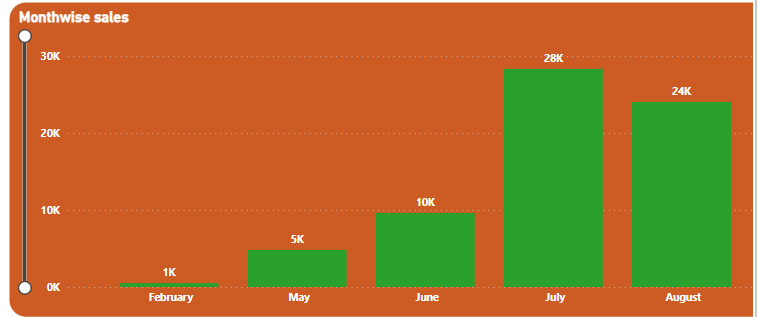
* The category for English-language movies shows a significantly higher
* French and German movies have notable counts in one of the metrics, potentially "Total Rentals," indicating a healthy demand or supply for these languages.



**Microsoft Power Bi**

1. How does the sales revenue vary by month?

Chart to visualize:

****

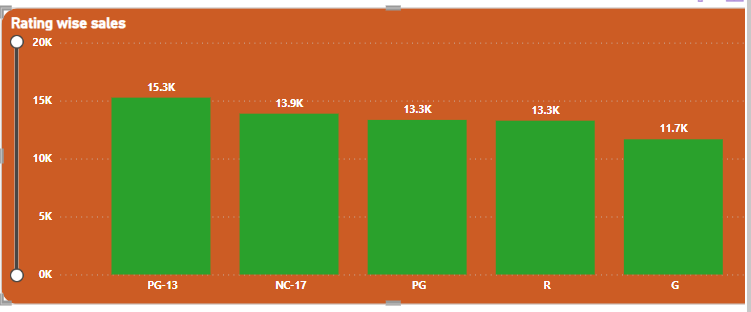
**Insights:**

* The sales begin at a low point in February .
* There is a slight increase in May, with sales .
* Sales dip again in June.
* There is a significant increase in July.
* In August, there is a slight decrease from July's peak**.**

2) What is the distribution of sales by payment method?

There is an insufficient data for solving this question. There is lack of data for payment method. So, I am adding another question.

Distribution of rating by sales?



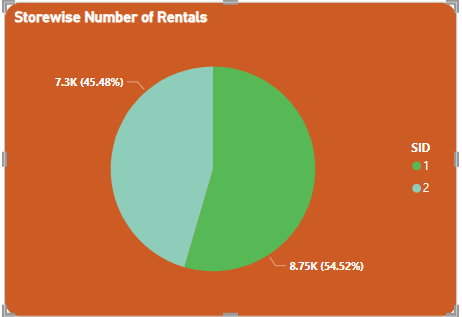
Insights:

* PG-13: This category has the highest sales.
* NC-17: The next category, NC-17, shows slightly lower sales.

3) Which customer segments generate the highest sales?

There is an insufficient data for solving this question. There is lack of data for customer segment. So, I am adding another question?

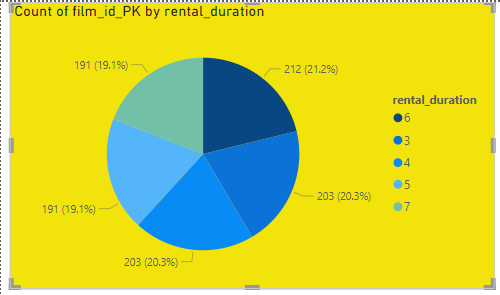
How will be the store performance by rentals?



Insights:

* SID 2: This store has a slightly larger share of the rentals at 8.75K,
* SID 1: The other store, SID 1, has a total of 7.3K rentals,

4) What is the distribution of films by rental duration?



Insights:

* The most common rental duration is 6 days, accounting for 21.2% of the rentals with a count of 212.
* The 3-day and 5-day rental durations have an equal share of 20.3% each, both with a count of 203.
* The rental durations of 4 days and 7 days each have a share of 19.1%, with a count of 191 for each.

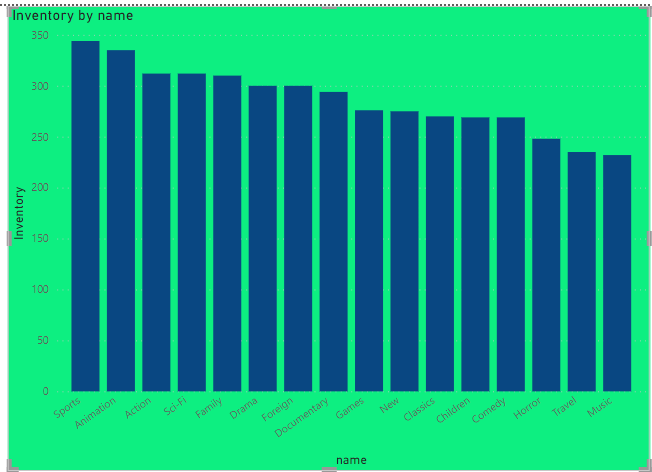
5) How does the inventory vary by film rating?



Insights:

* The PG-13 category appears to have the highest inventory
* NC-17 and PG categories have slightly lower inventory levels than PG-13
* The R-rated category's inventory is slightly less than NC-17 and PG
* The G-rated category has the lowest inventory among the listed ratings.

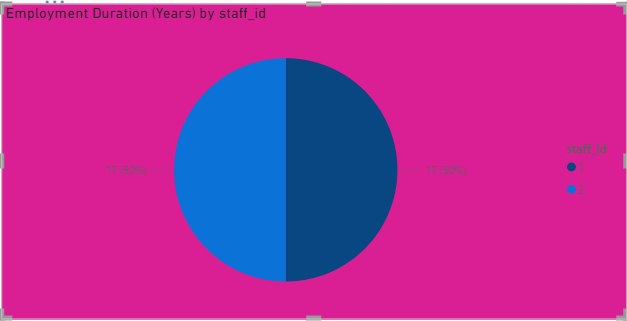
6) What is the breakdown of film categories in the inventory?



Insights:

* Sports Category: This category has the highest inventory.
* Animation and Action: These categories follow closely behind sports
* Consistent Inventory Levels: The inventory levels across Sci-Fi, Family, Drama, Foreign, Documentary, Games, New, and Classics categories are relatively consistent
* Lower Inventory: The inventory for Children, Comedy, Horror, Travel, and Music genres is slightly lower.

7) What is the distribution of staff by employment duration?



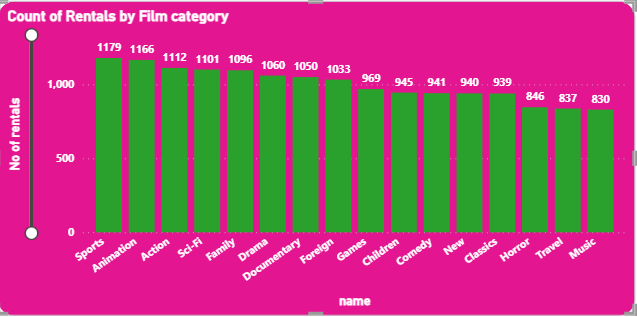
Insights:

* Both staff members have the same employment duration, which totals to 34 years combined, given that each has a duration of 17 years, making up 50% of the total each.

8) How does the store performance vary by location?

There is an insufficient of data for solving this question. We have lack of data for store performance.so we are adding question?

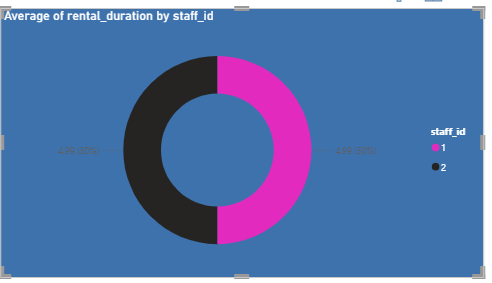
How is the distribution of rental by category?



Insights:

* Sports Category: This category has the highest number of rentals
* Animation and Action: These categories follow closely with 1,166 and 1,112 rentals.
* Consistent Rental Counts: The rental counts for Sci-Fi, Family, Drama, Documentary, and Foreign genres are relatively close, ranging from 1,101 to 1,033 rentals,
* Lower Rental Counts: The genres on the lower end of the rental count include Music, Travel, Horror, and Classics, with Music being the least rented category at 830.

9) What is the average rental duration by staff member?



Insights:

* Both staff members have identical average rental durations, with each contributing to 50% of the average duration represented by
* This equal distribution suggests that both staff members are equally efficient or that the rental duration policy is consistently applied across the board for rentals handled by either staff member.

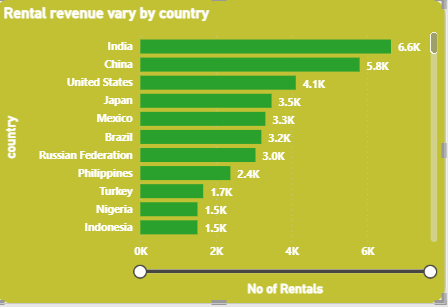
10) What is the distribution of customers across different cities?



Insights:

* The customers are globally distributed, with a significant presence
* There are dense concentrations of customers in Europe, parts of Asia, and the coastal areas of North America.

11) How does the rental revenue vary by country?



Insights:

* India: India leads with the highest number of rentals at 6.6K (6,600 rentals).
* China: China follows with 5.8K (5,800 rentals).
* United States: The United States is listed with 4.1K (4,100 rentals).

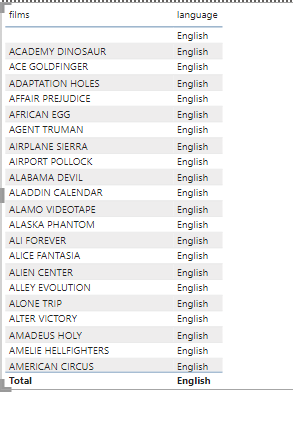
12) Which locations have the highest and lowest customer ratings?



Insights:

* North America has the highest customer rating
* Australia has the lowest customer rating

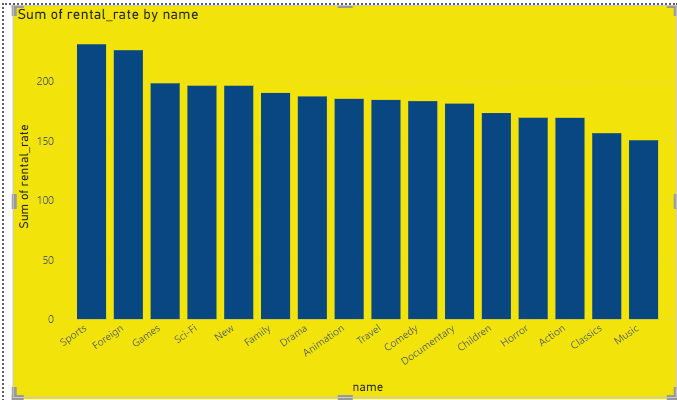
13) What is the distribution of films by language?



Insights:

* The distribution of films are only by English language. There is no other language.

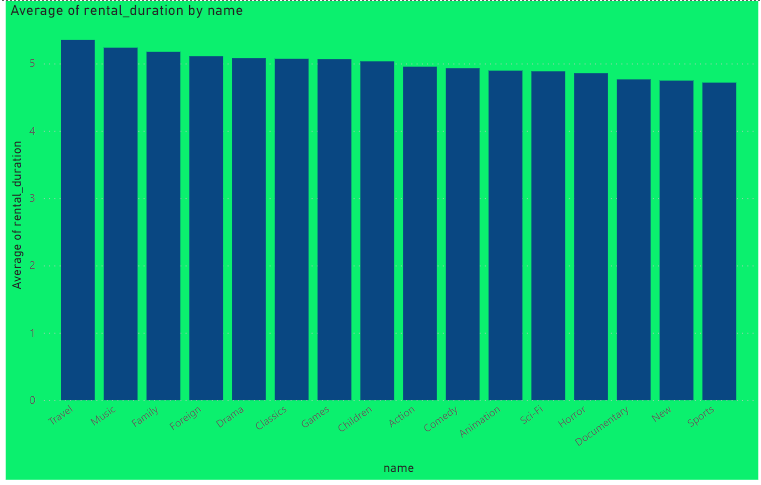
14) Which film categories have the highest rental rates?



Insights:

* By visualizing the chart sports and foreign film categories have highest rental rates.

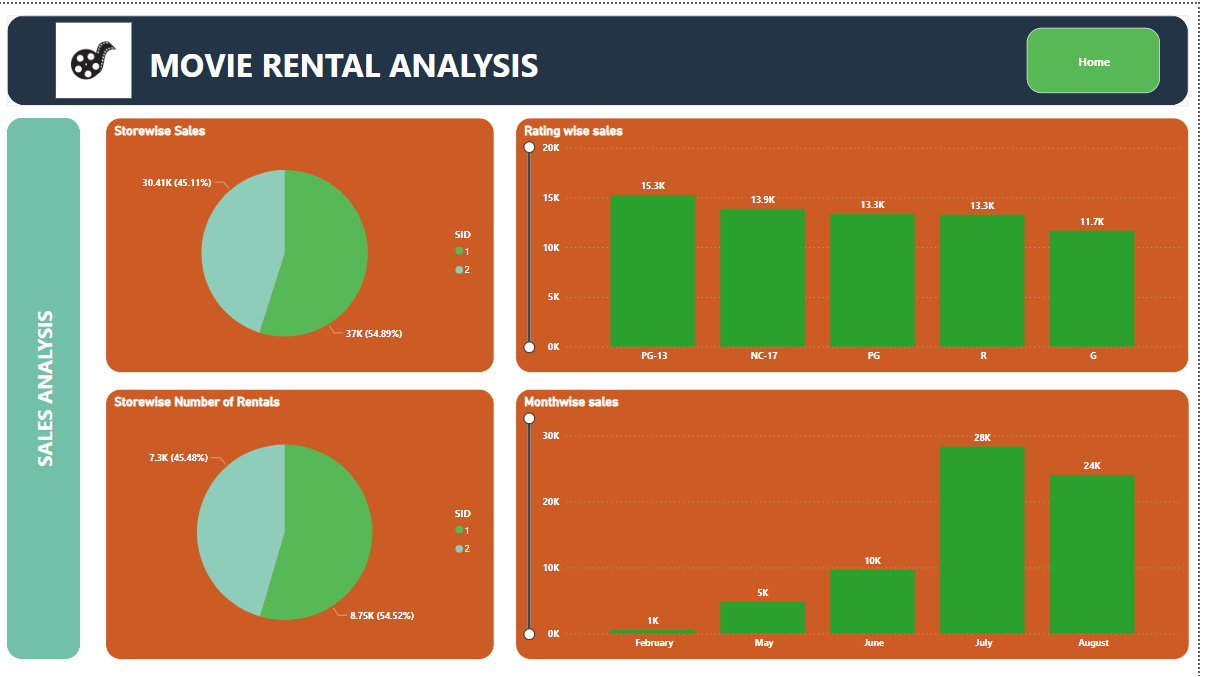
15) How does the average rental duration vary by film category?



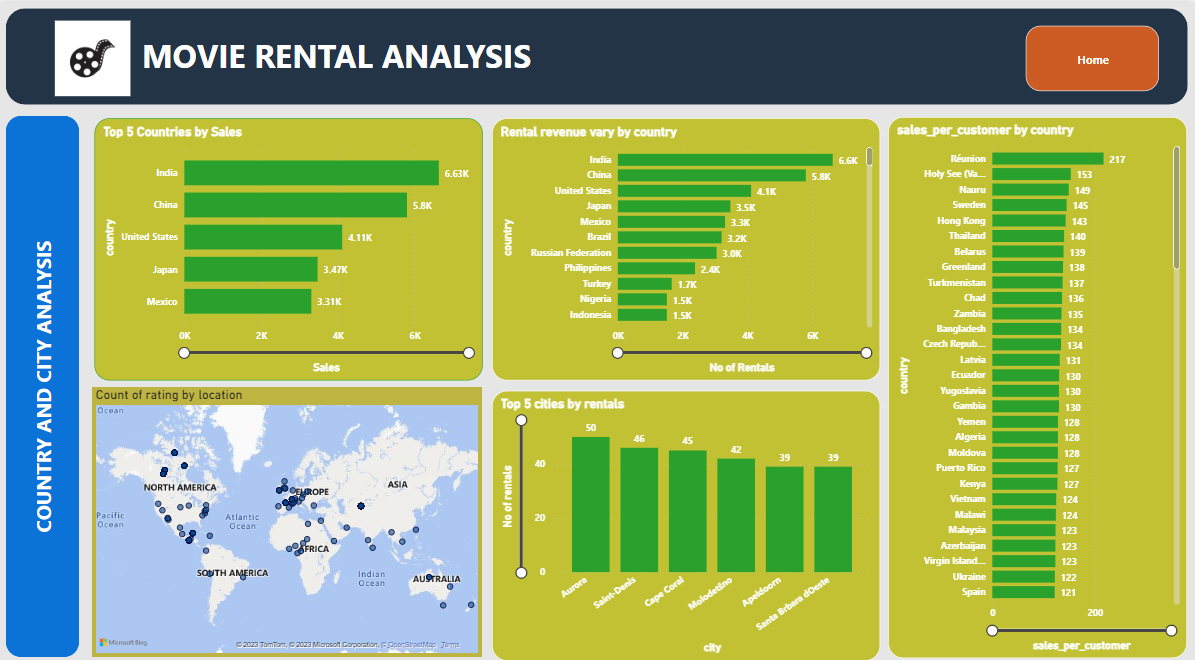
Insights:

* Travel has high average rental duration.
* Sports has low average rental duration.

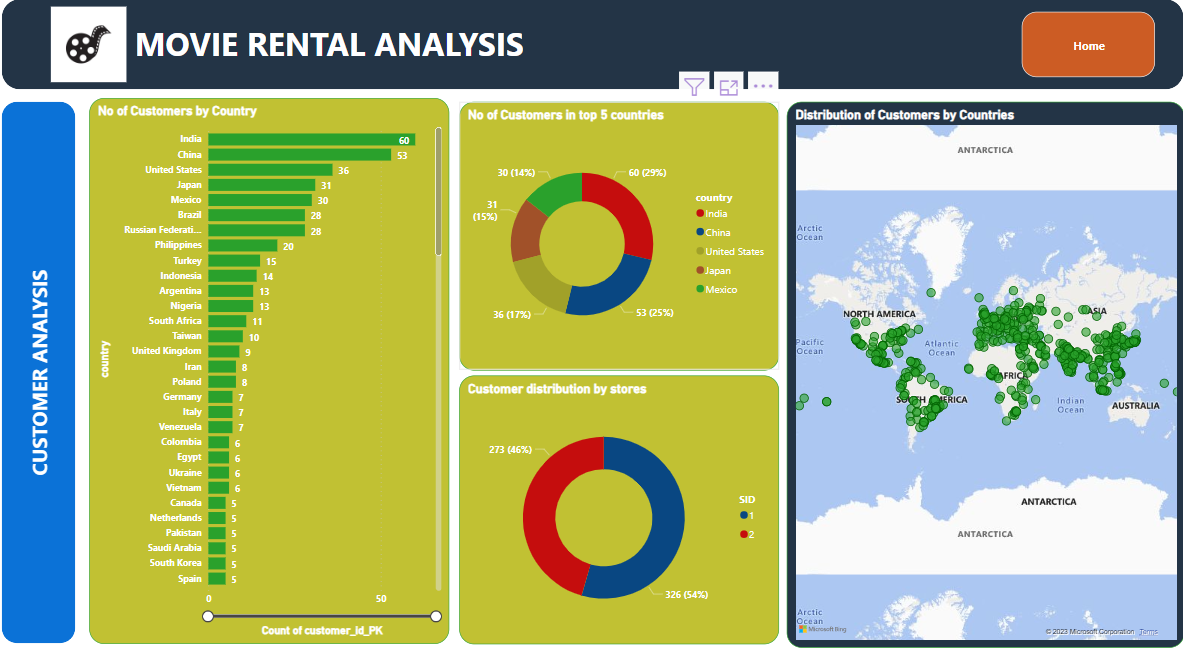
Sales Analysis:



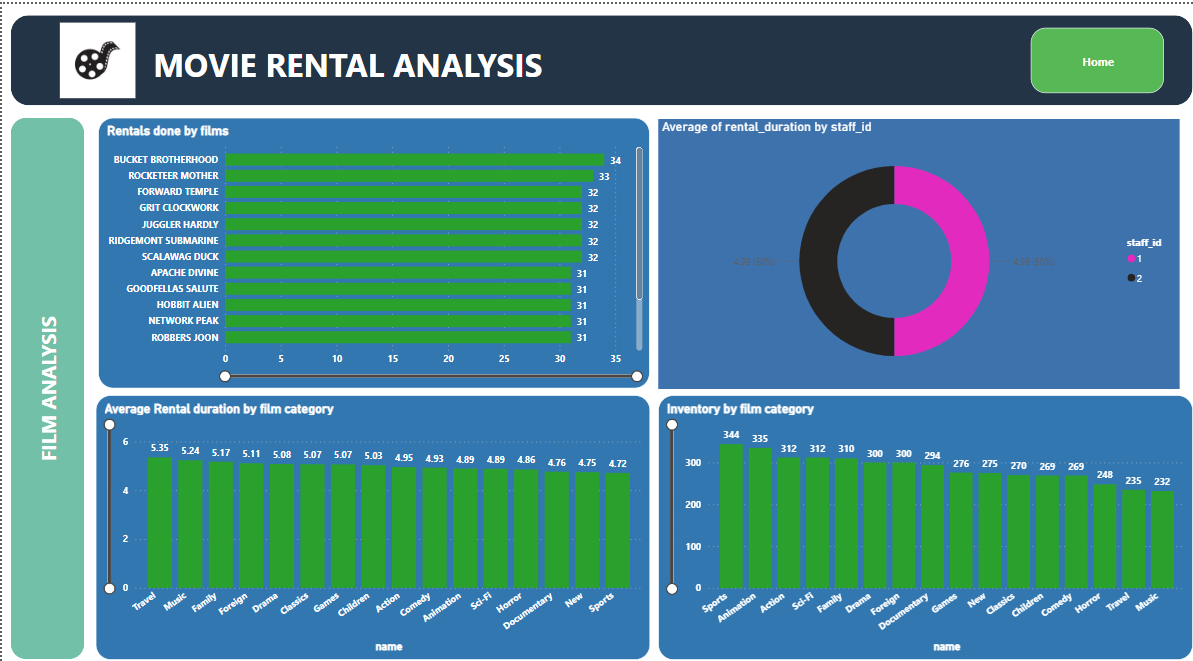
Country and city analysis:



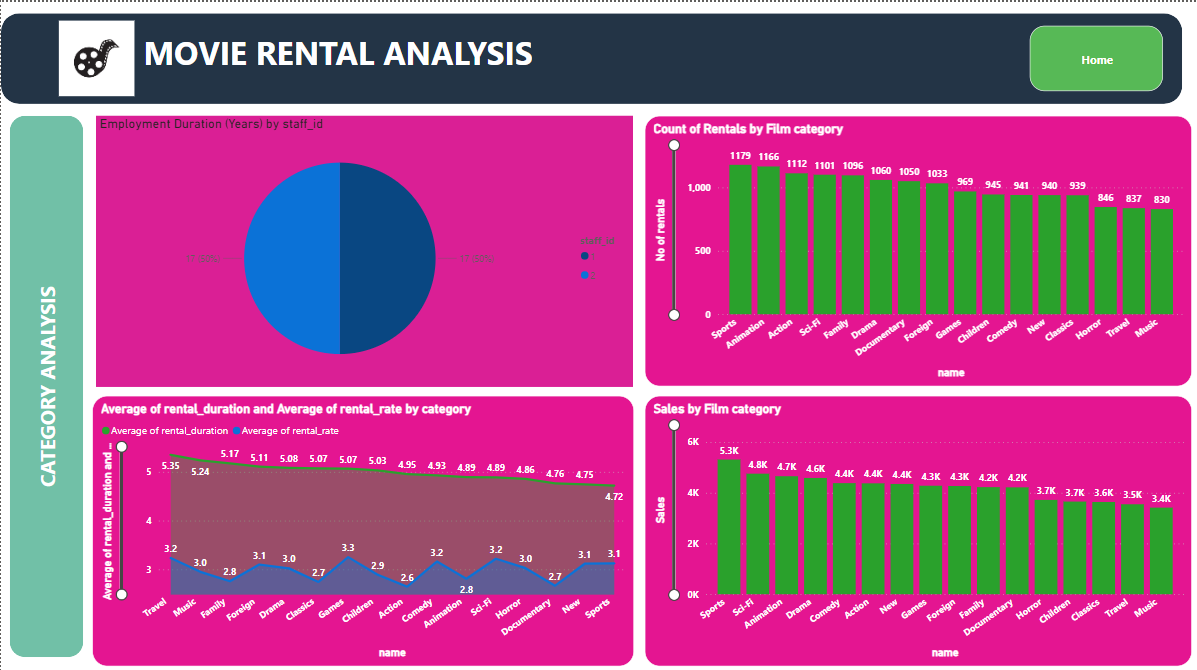
Customer Analysis:



Film Analysis:



Category Analysis:



Conclusion:

* In this project, we analyzed data from a DVD rental company we decided to call ‘Rent A Film’ to find insights about the customers and their preference. We got 3 major conclusions:
* The company has sport-loving customers and they would be advisable to stock more sport-related films to increase total sales compared to music-related movies. It would be a good idea to increase the average rental rate of sport genre films since it is not a major factor in renting for the customers. This, in turn, increases total revenue. However further analysis needs to be done to conclude on this
* There is potential to have an extra source of revenue through a fee on late film returns.
* Rent A Film has a presence in 108 countries with India been the largest market in people and revenue. Additionally, 20% of the countries they have presence in contribute to 80% of the total customer base.