**Capstone Project**

**Movie Rental Analysis**

**Overview:**

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.

**Report and Presentation:**

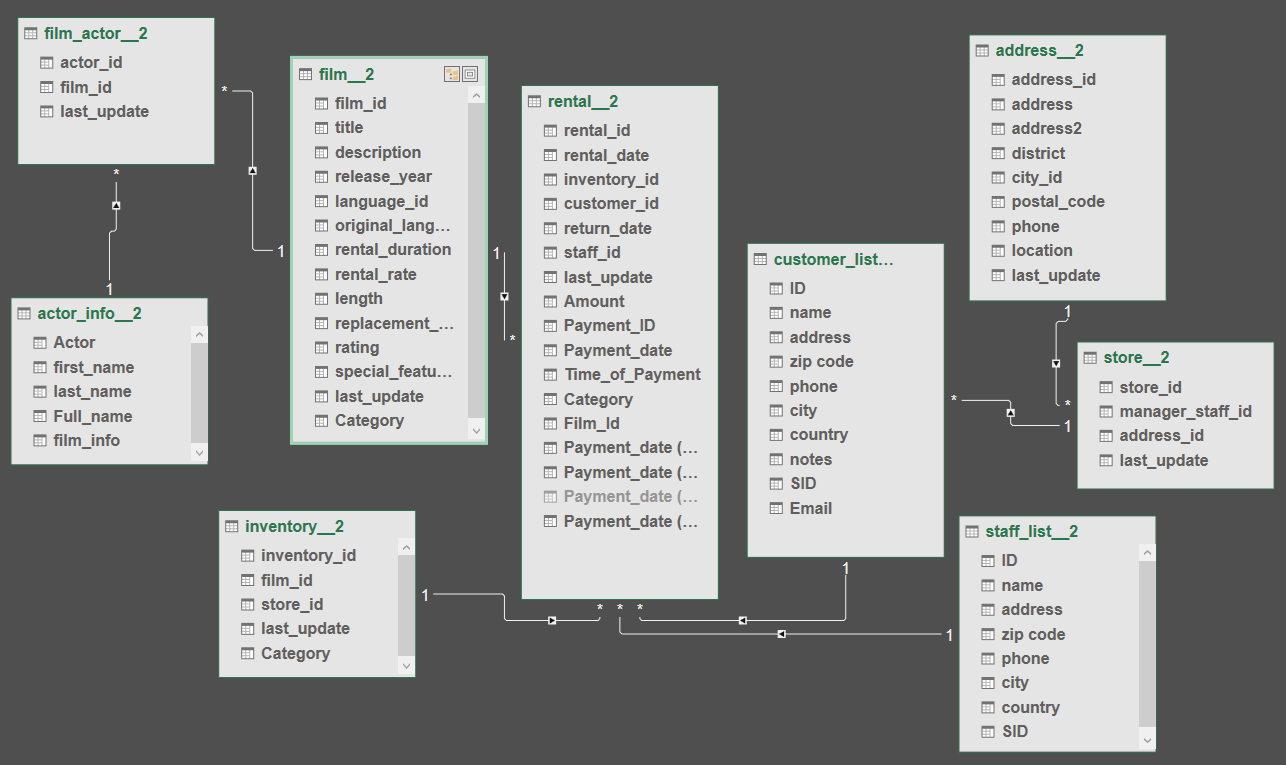
After conducting the Exploratory Data Analysis (EDA) and uncovering patterns, trends, and other insights, we will proceed to create an Excel dashboard to visualize and present the findings. This dashboard will serve as a powerful tool for the business to efficiently analyse their operations and make informed decisions based on the data-driven insights gathered from the EDA process. The visualization will facilitate a clear understanding of the business performance, allowing them to identify opportunities for improvement, optimize strategies, and enhance overall efficiency

**Data Dictionary:**

Source Tables:

|  |  |  |
| --- | --- | --- |
| Table Name | Important Columns | Significance |
| Rental- Transaction table containing information about all the rental transaction | Rental id | Unique number representing each rentals |
| Amount | Amount payed by the customer for rentals |
| Payment date | Date of payment for the rentals by the customer |
| Customer- This lookup table contains information about the customers | ID | Unique ID of the customer |
| Name | Name of the customer |
| City | City of residence of the customer |
| Country | Country of residence of the customer |
| SID | Store ID associated with the customer |
| Inventory- This lookup table contains information about the inventory maintained in the stores | Inventory Id | Unique ID associated with each inventory |
| Category | Category of the films present in the given inventory |
| Film id | Id of the film present in the inventory |
| Store Id | Id of the store to which the inventory belongs |
| Film- This is lookup table about films containing information about film id, title, duration etc. | Film Id | Unique ID associated with the film |
| Title | Title of the film |
| Rental duration | Describes the rental duration for the film |
| Rental rate | Rate of the rental film |
| Rating | Rating of the film |
| Film Actor- This is a transaction table for the film actors and the films they appeared in. | Actor ID | ID associated with the actor |
| Film ID | ID associated with the film |
| Actor Info – This is a lookup table containing information about the films | Actor ID | ID associated with the actor |
| Full name | Full name of the actor |
| Store- this is a lookup table about the store | Store ID | ID associated with the store |
| Staff- This is lookup table containing information about the staff | ID | ID associated with the staff |
| Name | Name of the staff |
| Address- This is a lookup table about the address | Address ID | ID associated with the staff |

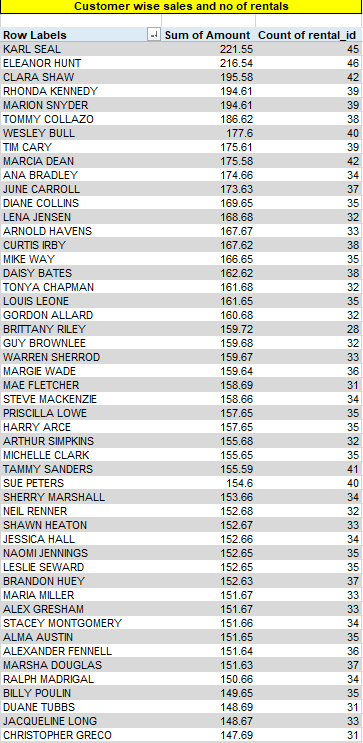
**ER Diagram:**



**Exploratory Data Analysis:**

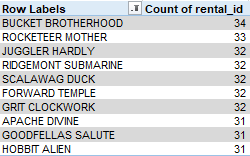
1. What are the purchasing patterns of new customers versus repeat customers?

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

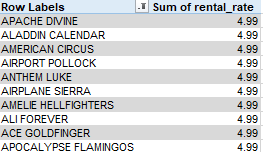


1. Which films have the highest rental rates and are most in demand?

Top 10 Films with highest demand



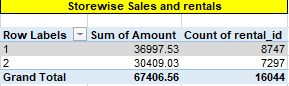
Top 10 Films with highest rental rates



From the above data one can see that films with high rental rates have less demand.

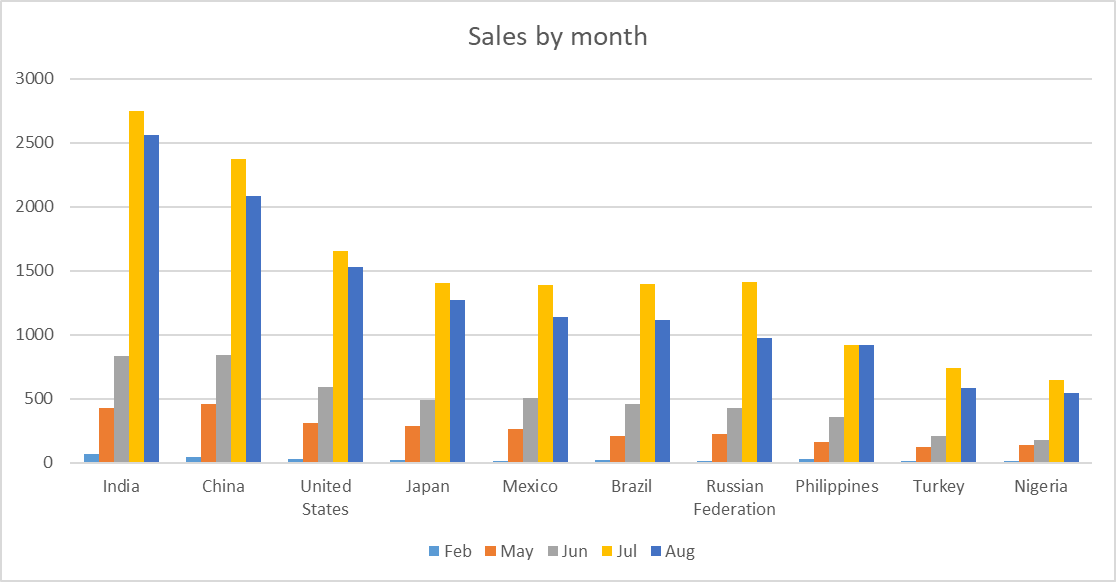
1. Are there correlations between staff performance and customer satisfaction?

Since there is no provided column for measuring customer satisfaction, 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



From the table above, it's evident that store 1 demonstrates a slightly superior performance compared to store 2.

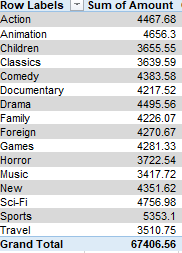
1. Are there seasonal trends in customer behaviour across different locations?



As evident from the chart, there is a noticeable rise in sales starting from May, peaking in July, and then gradually declining once again.

1. Are certain language films more popular among specific customer segments?

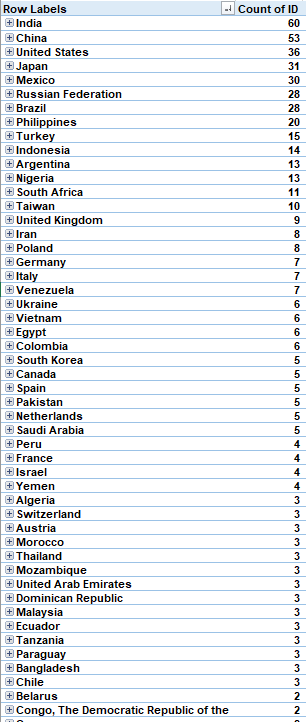
Since the provided dataset exclusively comprises English-language films, conducting language-based analyses is not feasible. Nonetheless, we can proceed to analyse the sales figures associated with various film categories.



Based on the table above, it is evident that films belonging to the sports category exhibit the highest sales figures.

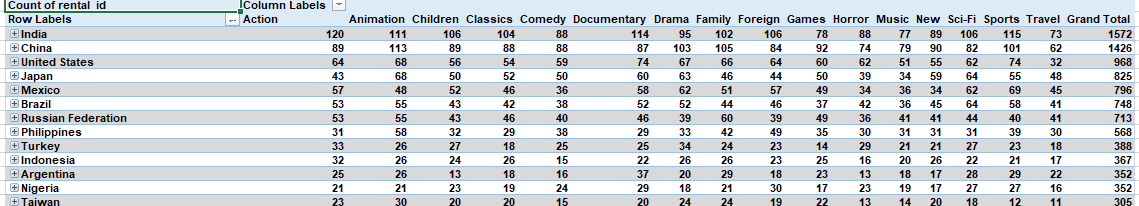
1. How does customer loyalty impact sales revenue over time?

Due to the absence of customer on boarding dates in the provided dataset and its limited timeframe, assessing the impact of customer loyalty over time is unfeasible. Nevertheless, we can proceed to analyse the distribution of customers among different countries



Top 8 countries have almost 47 % of the customers.

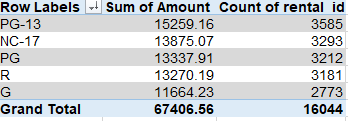
1. Are certain film categories more popular in specific locations?



The data does not indicate the presence of any such relationship.

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

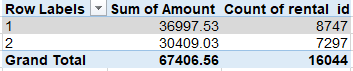
Since there is a lack of information regarding customer ratings, addressing the problem at hand remains unattainable. Nevertheless, we can proceed to assess the sales performance of films based on their respective ratings.



Films with a PG-13 rating stand out in terms of both sales and rentals.

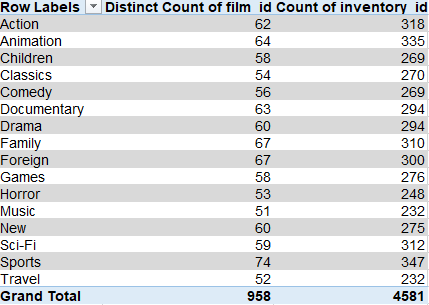
1. How does the proximity of stores to customer’s impact rental frequency?

Given the presence of two stores, one in Canada and the other in Australia, and considering that rentals are occurring globally, determining the impact of store proximity on rental frequency is unfeasible. However, we can proceed to analyse the sales and rentals generated by each individual store.



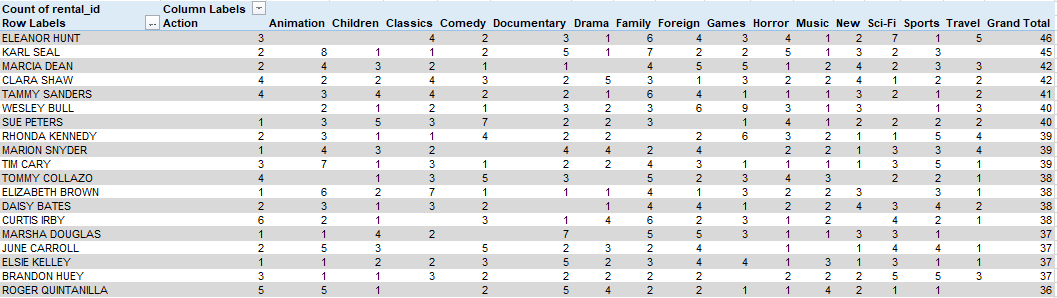
1. Do specific film categories attract different age groups of customers?

Since there is a lack of data pertaining to customer ages, conducting the intended analysis is not feasible. Instead, we can proceed to examine the distribution of films across different categories and assess the corresponding inventory levels.



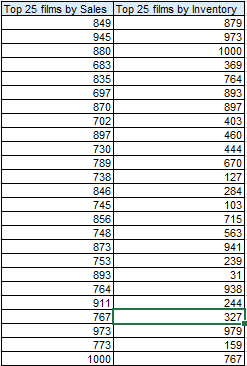
1. What are the demographics and preferences of the highest-spending customers?

Due to the absence of customer demographic information within the dataset, we are unable to ascertain specific customer preferences. Nonetheless, we can proceed to identify category-wise preferences among high-spending customers.



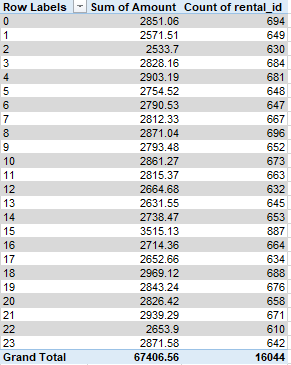
1. How does the availability of inventory impact customer satisfaction and repeat business?

The dataset lacks information concerning customer satisfaction. However, from the available data, it is evident that a sufficient inventory availability directly correlates with increased sales and repeat business.



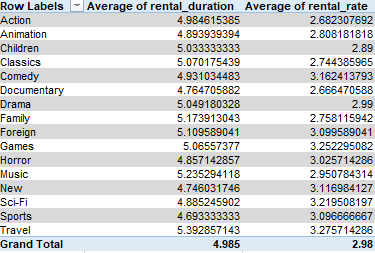
1. What are the busiest hours or days for each store location, and how does it impact staffing requirements?

This distribution displays the sales and rentals on an hourly basis. The image clearly indicates that the peak sales period falls within the timeframe of 15:00 to 16:00.



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

Due to the absence of customer demographic data, establishing a relationship is not feasible. Instead, we can proceed to analyse the average rental duration and rental rate across different film categories.



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

Given that the dataset exclusively contains films in the English language, we are unable to identify the specified relationship. Nonetheless, we can proceed to analyse the distribution of customers across different stores.

