A close-up of a cd

Description automatically generated

CAPSTONE Project Movie Rental Analytics --report by Akarsh Sharma

**SaKILA DVD RENTAL ANALYSIS**

**OVERVIEW**

**Sakila DVD Rental is a fictional company that operates a database of DVD rentals, often used as a sample database in educational and testing environments. Analyzing Sakila DVD Rental data provides insights into various aspects of a rental business, such as customer behavior, inventory management, and revenue trends. This overview will highlight the key components and objectives of analyzing the Sakila DVD Rental database.**

**1. Database Structure:**

**The Sakila DVD Rental database comprises multiple interconnected tables, including information about customers, staff, films, rental transactions, inventory, and payments. Each table contains specific data relevant to the DVD rental business, making it a suitable dataset for a wide range of analysis and reporting.**

**2.Objectives of Analysis:**

**Analyzing the Sakila DVD Rental database can serve several objectives, including:**

**A. Customer Behavior: Understanding customer preferences, rental frequency, and the correlation between customer demographics and rental choices.**

**B. Inventory Management: Optimizing inventory levels by identifying popular and underutilized films, monitoring stock levels, and ensuring availability for high-demand titles.**

**C. Revenue Analysis: Exploring revenue trends, identifying peak rental periods, and assessing the impact of pricing strategies on overall earnings.**

**D. Geographic Insights: Examining the geographic distribution of customers and the performance of different rental stores.**

**PROCESS**

1. **Data acquisition from Github.**

**Obtained the valuable dataset from Github repository provided by Acciojob. The dataset contains detailed information of film inventory , actors , customers behaviors. Customers address, country, city , staffs, stores, way of payments .**

1. **Tools connections**

**Building connections between datasets provided in different datasets files ( csv Files ) .connections can be built from excel, & power BI . also data is connected with SQL , from this SQL query we extract different set of data by writing of required SQL queries.**

1. **Data Analysis:**

**a. Query the database: Use SQL queries to extract the relevant data needed for your analysis. Write queries to retrieve data from different tables, filter data, and perform joins as necessary.**

**b. Calculate key metrics: Compute the metrics and KPIs that align with our objectives. For instance, calculate rental frequency, film popularity, revenue per film, employee efficiency, and any other relevant metrics.**

1. **Exploratory data Analysis(EDA)**

**EDA is performed using excel and SQL workbench , use of this is dine as per required and writing queries is done by understanding the nature of the questions and required steps. Problem solving provides meaningful insights and patterns from the datasets for better business decision making required for analysis.**

1. **Understanding solutions and building dashboards in power BI**

**Use power bi for building relationship between different table provided in csv files . Understanding the process , and solving the problem statements provided in project with the use of different skills used in power bi. Also building dashboard is required to show the beautiful connection of data and results.**

1. **Creation of PowerPoint & documentation**

**Creation of detailed reports of project in word documents also buildings ppt that focus on insights details of all the process and insights of EDA ,power bi and respective dashboard.**

**SIGNIFICANCE**

**The significance of Sakila DVD Rental analysis lies in its ability to provide valuable insights into various aspects of a fictional DVD rental business. Although the Sakila database is not a real business, it serves as a powerful tool for educational and practical purposes, demonstrating the importance of data analysis and its potential impact on real-world businesses. Here are several key aspects of the significance of Sakila DVD Rental analysis:**

1. **Data-Driven Decision-Making:**

**Sakila analysis illustrates the power of data-driven decision-making. By examining customer behavior, inventory management, and revenue trends, individuals can better understand how data can inform strategic decisions in a real business setting.**

1. **Optimizing Business Operations:**

**Analysis of the Sakila database can help fictional DVD rental businesses optimize their operations. By identifying popular and underutilized films, understanding customer preferences, and streamlining inventory management, businesses can enhance their overall efficiency and profitability.**

1. **Customer Insights:**

**The analysis can provide insights into customer behavior, such as rental frequency, genre preferences, and geographic distribution. This knowledge can be applied to tailor marketing efforts, improve customer service, and build customer loyalty in real businesses.**

1. **Revenue Growth:**

**By evaluating revenue trends and assessing pricing strategies, businesses can identify opportunities for revenue growth. This might involve adjusting pricing for specific films, or optimizing the rental period.**

1. **Data Visualization Skills:**

**Working with Sakila data provides an opportunity to enhance data visualization skills. Creating meaningful charts, graphs, and reports from the data helps individuals convey their findings effectively and make data more understandable to stakeholders.**

1. **Risk Assessment:**

**By examining rental patterns and revenue trends, businesses can identify risks and opportunities for mitigating them. For example, they can anticipate periods of low rental activity and plan accordingly.**

1. **Geographic Insights:**

**Understanding the geographic distribution of customers and the performance of different rental stores can inform business expansion strategies. Companies can identify areas with untapped potential or underperforming locations.**

**TABLE OVERVIEW**

**The Sakila DVD Rental database comprises several interconnected tables, each containing specific data related to the operations of a fictional DVD rental business. Here is an overview of some of the key tables in the Sakila database:**

1. **Customer Table:**

**Purpose: Stores information about customers who rent DVDs, including their names, addresses, email addresses, and phone numbers.**

**Fields: customer\_id(PRIMARY KEY), store\_id, first\_name, last\_name, email, address\_id(FOREIGN KEY), active, create\_date, last\_update, and more.**

1. **Film Table:**

**Purpose: Contains details about the DVDs available for rental, such as film titles, descriptions, release years, rental durations, rental rates, and more.**

**Fields: film\_id(PRIMARY KEY), title, description, release\_year, language\_id, rental\_duration, rental\_rate, replacement\_cost, rating, special\_features, and more.**

1. **Inventory Table:**

**Purpose: Manages the inventory of DVDs in the rental stores. It associates films with specific stores and tracks their availability.**

**Fields: inventory\_id(PRIMARY KEYS), film\_id(FK), store\_id(FK), last\_update.**

1. **Rental Table:**

**Purpose: Records information about rental transactions, including the customer who rented the DVD, the rental date, return date, and other rental-specific details.**

**Fields: rental\_id(PK), rental\_date, inventory\_id(FK), customer\_id(FK), return\_date, staff\_id(FK), last\_update.**

1. **Payment Table:**

**Purpose: Tracks payment information for DVD rentals, including the payment date, rental charges, and payment amount.**

**Fields: payment\_id(PK), customer\_id(FK), staff\_id(FK), rental\_id, amount, payment\_date, last\_update.**

1. **Staff Table:**

**Purpose: Stores information about staff members working in the rental stores, including their first and last names, email addresses, and contact details.**

**Fields: staff\_id(PK), first\_name, last\_name, address\_id, email, store\_id, active, username, password, and more.**

1. **Store Table:**

**Purpose: Contains details about the rental stores, including their locations, addresses, and contact information.**

**Fields: store\_id, manager\_staff\_id, address\_id, last\_update.**

1. **Address Table:**

**Purpose: Manages address information for customers, stores, and staff. It includes details like street addresses, postal codes, and phone numbers.**

**Fields: address\_id(PK), address, address2, district, city\_id, postal\_code, phone, and last\_update.**

1. **City Table:**

**Purpose: Stores city-related information, including city names and their association with countries and regions.**

**Fields: city\_id, city, country\_id, last\_update.**

1. **Country Table:**

**Fields: country\_id, country, last\_update.**

**PROBLEM STATEMENTS OF EDA**

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

**A screenshot of a computer

Description automatically generated**

**>> Here after analysis of new and repeated customer , i find that there is no relation and no pattern in new and repeated one .**

**all the customers are repeated because there count of rental is always more than 1 .**

**A screenshot of a computer

Description automatically generated**

**FROM THE ABOVE ATTACHMENTS ,**

**So i can analysis the customer pattern , that how much they order or rent the dvd in a month.**

**there are only repeated customers so how much they rent in a month, we can analysis this with the data.**

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

**A screenshot of a graph

Description automatically generated**

**HERE WE Find rental count according to rental count.**

**For this data is extracted from SQL query and with the use of join , tables are joined together for extraction of the data and building relationships.**

**Also built a Chart for understanding the relationship in explained manner.**

**3. Are there correlations between staff performance and customer satisfaction?**

**A blue rectangle with white text

Description automatically generated**

**4. Are there seasonal trends in customer behavior across different locations?**

**A screenshot of a graph

Description automatically generated**

**Above extraction of year month and total revenue from the table payment. Chart is also created here to understand the vision.**

**A screenshot of a computer

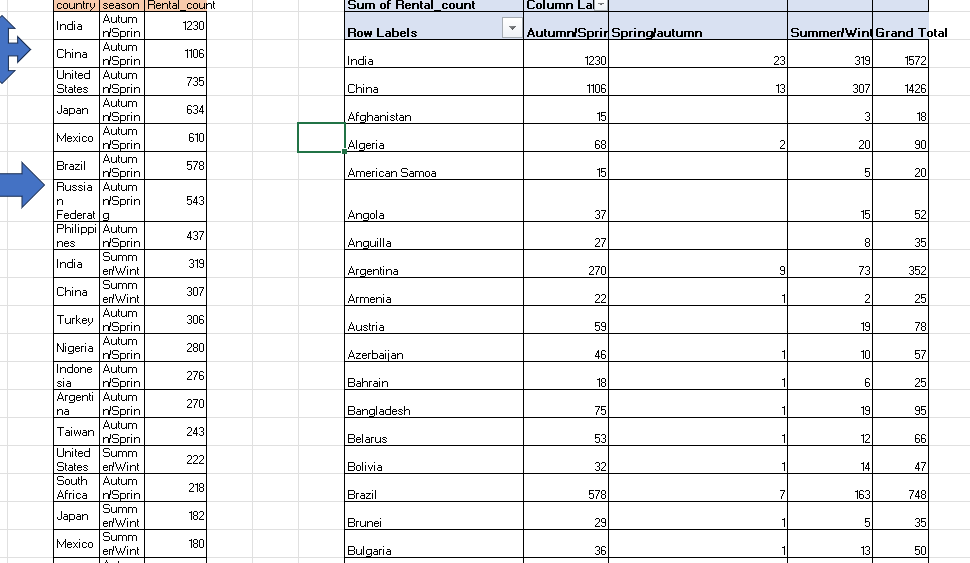
Description automatically generated**

**First , with the help of sql queries extract all the country in regions wise , so that regions and geographical wise data can be understand**

**After doing that extract the season according to month**

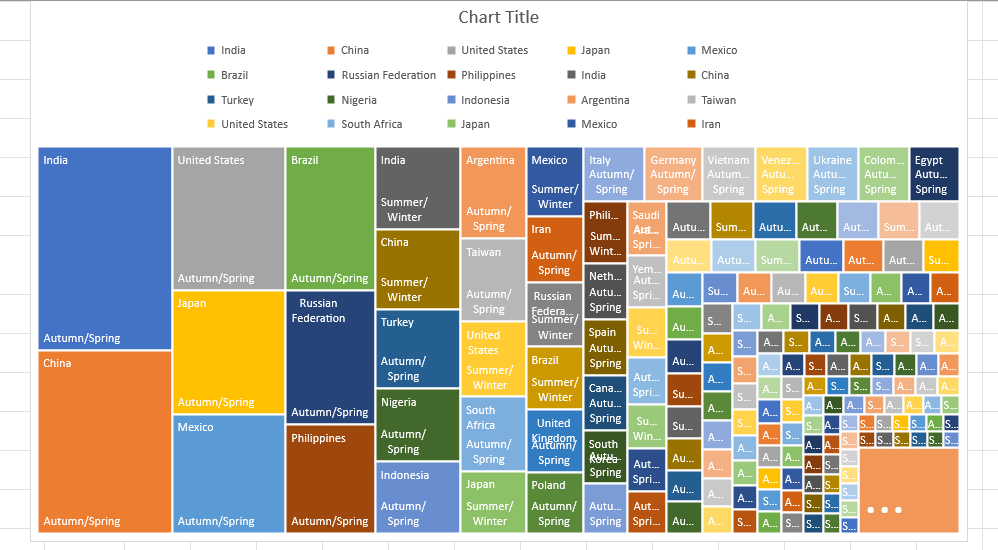
**i.e. month no 5 and 6 are summer / winter and 7 and 8 are autumn / winter according to the hemisphere .**

**calculate the rental count according to the country and season . India is t 1st . as we seen in the table build above .**

****

**From above results we have list of data of country , regions and rental counts.**

**Built pivot table with this table and also create chart to understand the rental count according to country and regions**

****

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

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

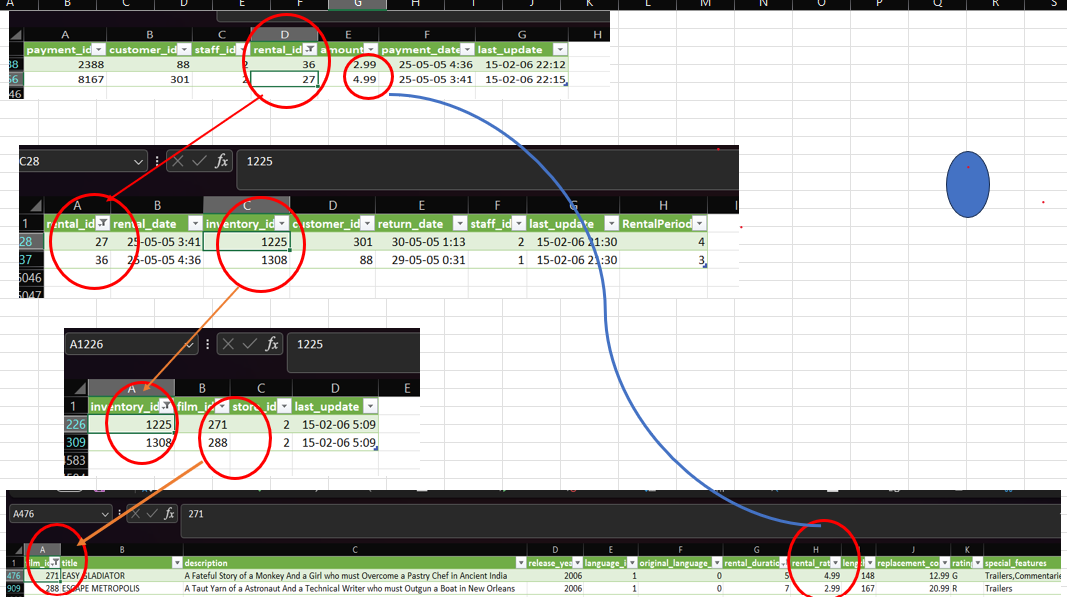
**A screenshot of a computer

Description automatically generated**

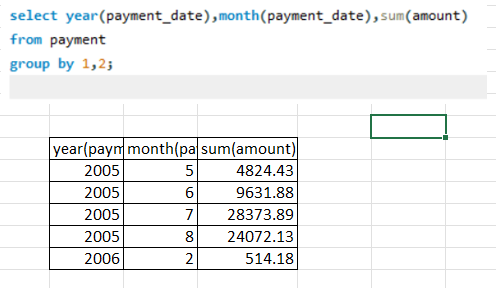
**there is only one language i.e. English language is available in**

**language id, so we cannot extract the above result.**

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

****

**We can not extract the sales revenue with shows the loyalty of the customer as we seen from the above results**

****

**For this we have extract year wise ,month wise total amount as sum( payment ) from the payment table**

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

****

**Extract the country according to regions wise**

**After that**

**A screenshot of a computer program

Description automatically generated**

**A graph on a graph

Description automatically generated**

**A screenshot of a computer

Description automatically generated**

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

**A screenshot of a computer

Description automatically generated**

**Built a problem statements , count of inventory id according to the store wise and rating wise?**

**Here is the answers .**

**A screenshot of a graph

Description automatically generated**

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

**A screenshot of a computer

Description automatically generated**

**Firstly extract the data from the tables after joining the different table .such as address , city and country.**

**2 store are located in these two country Australia and Canada ,**

**So I analysis the data for thesr two country.**

**A screenshot of a computer

Description automatically generated**

**Also create treemap chart to show the data clearly.**

**A chart with different colors and text

Description automatically generated**

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

**A screenshot of a computer

Description automatically generated**

**Extract the data according to the age category , for this we extract age group according to the rating given in dataset , Use SQL query to extract the data I join different table .**

**After this**

**A screenshot of a computer

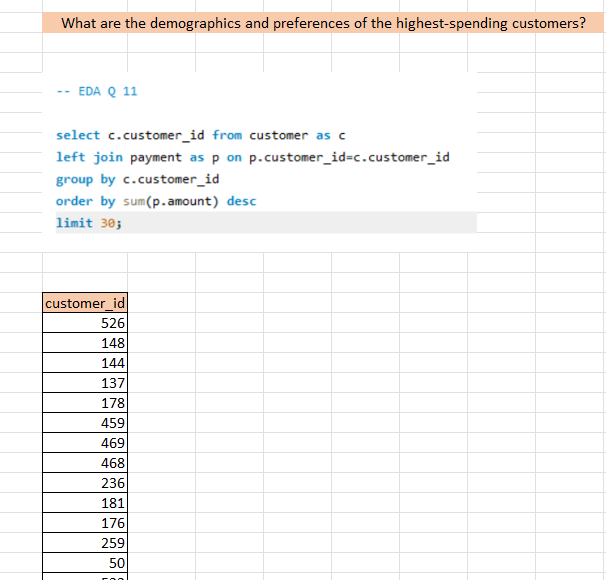
Description automatically generated**

**According to the age group , we can extract the count of rental id . also extract name wise rental count after joining different tables. Also create the line chart to understand the rental count according to name of category .**

**A screenshot of a graph

Description automatically generated**

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

****

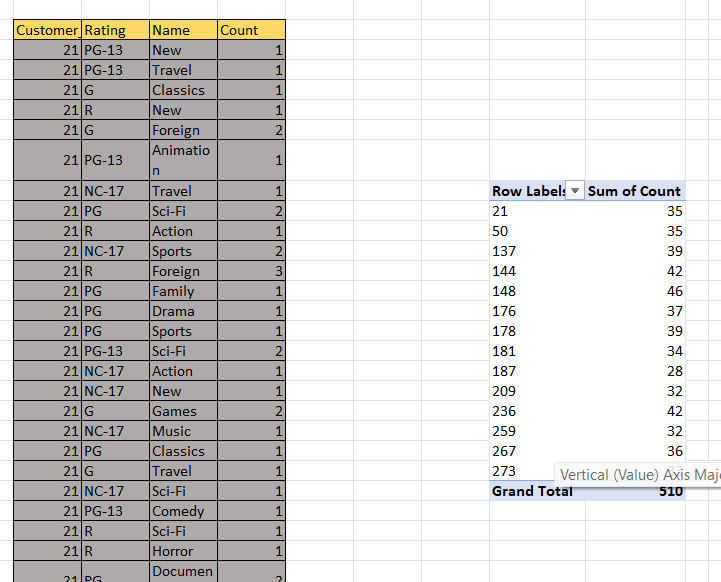
**In this question with the use of sql query I extract the customer id ,order by sum of amount .**

**After this**

**A screenshot of a computer

Description automatically generated**

**Extract the customer id , rating, name , and count of rental id after joininf the dieffrent table like rental , inventory , film category , film , category , and customer.**

****

**Create the pivot table from extracted data and after that create a chart on the basis of customer id and sum of count**

**Also add slicer for rating and name ( category).**

**A screenshot of a computer

Description automatically generated**

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

**A blue and white sign

Description automatically generated**

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

**A screenshot of a computer

Description automatically generated**

**First extract the data on the days wise for each store .**

**A graph of a bar chart

Description automatically generated with medium confidence**

**After this extract the data according to the hour wise .**

**A screenshot of a computer

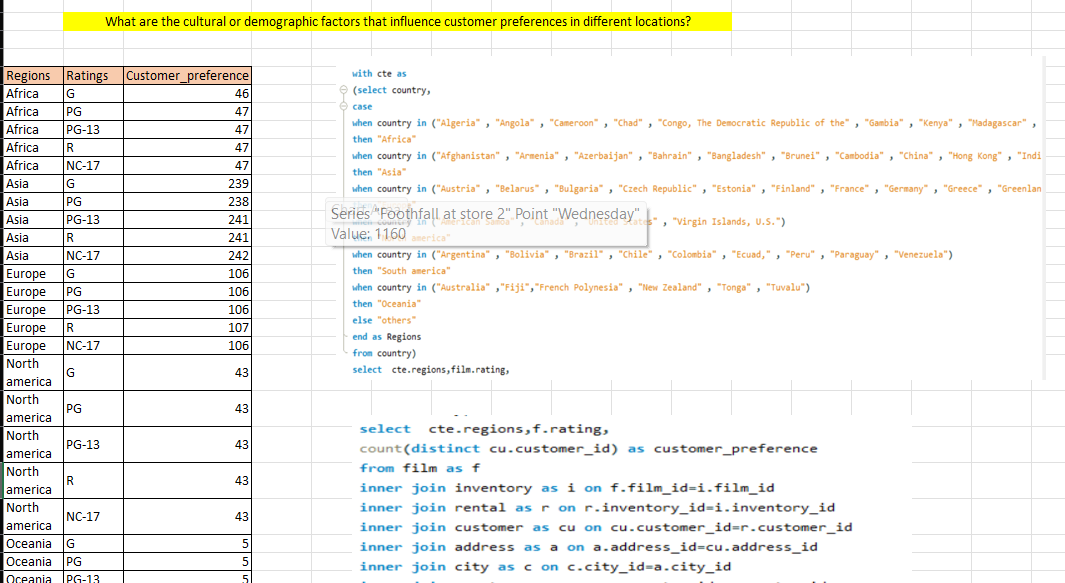
Description automatically generated**

**Here is the chart for the above result.**

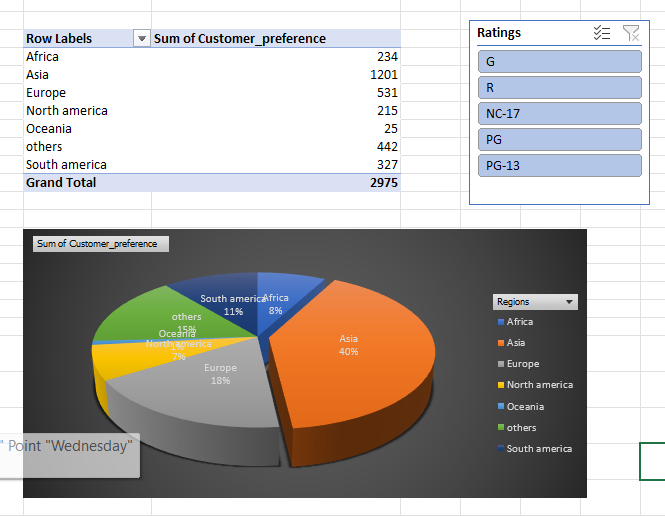
**A chart with multiple colors

Description automatically generated with medium confidence**

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

****

**Using sql queries ,1st extract the country as region wise , this is due to the demographic factors . using this query I can extract count of customers with their rating and regions wise . using join for different table I can extract that data . after this create a pivot table and with this create chart for clear understanding .**

****

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

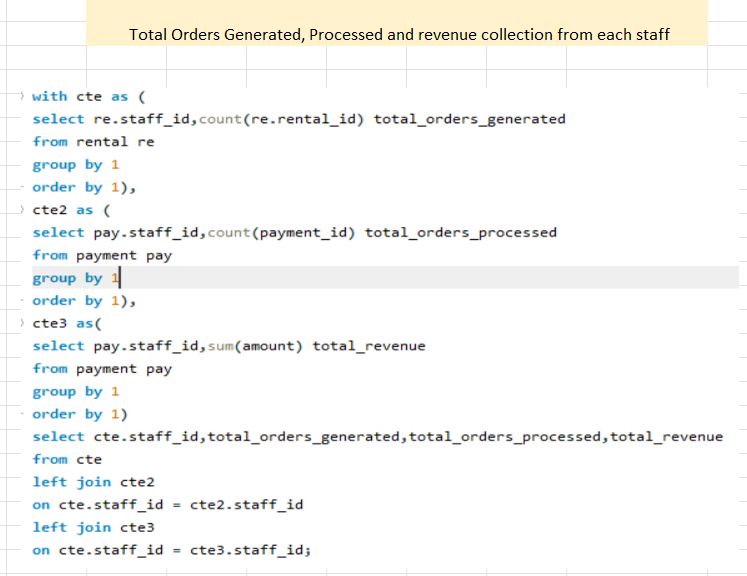
**A blue and white text on a grid

Description automatically generated**

**Additional problem statement in EDA**

**1.Calculate the Total Orders Generated, Processed and revenue collection from each staff**

**Staff performance**

****

**A screenshot of a graph

Description automatically generated**

**2.what are the Contribution to orders based upon different rental\_rate by each staff.**

**A computer screen shot of a program

Description automatically generated**

**A screenshot of a graph

Description automatically generated**

**PROBLEM STATEMENT ON POWER BI**

**1st of all understanding the data medeling**

**And relation building between different table**

**A screenshot of a computer

Description automatically generated**

**1 How does the sales revenue vary by month?**

**A graph of blue squares

Description automatically generated with medium confidence**

**Doing Cluster column chart for sum of amount in different month and year .**

**2. What is the distribution of sales by payment method?**

**A blue circle on a pink background

Description automatically generated**

**RentalDate = RELATED(rental[rental\_date])**

**Payment\_Mode = if(DATEDIFF(payment[RentalDate],payment[payment\_date],DAY)=0,"Prepaid","Postpaid")**

**Using above function we can crated payment mode in the additional column**

**A blue circle on a pink background

Description automatically generated**

**Pie chart for payment mode and sum of amount.**

**3.** **Which customer segments generate the highest sales?**

**A blue circle with a blue center

Description automatically generated**

**Pie chart for count of payment id with active**

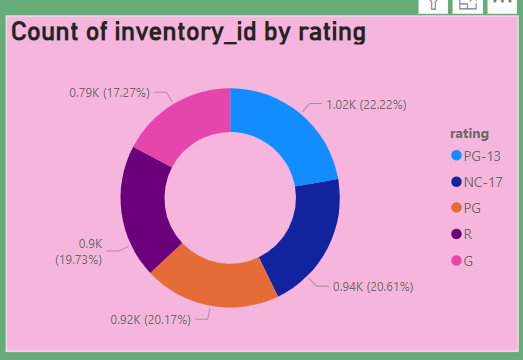
**4.** **What is the distribution of films by rental duration?**

**A colorful pie chart with text

Description automatically generated**

**Pie chart for film id vs rental duration .**

**5.** **How does the inventory vary by film rating?**

****

**6.** **What is the breakdown of film categories in the inventory?**

**A screenshot of a computer

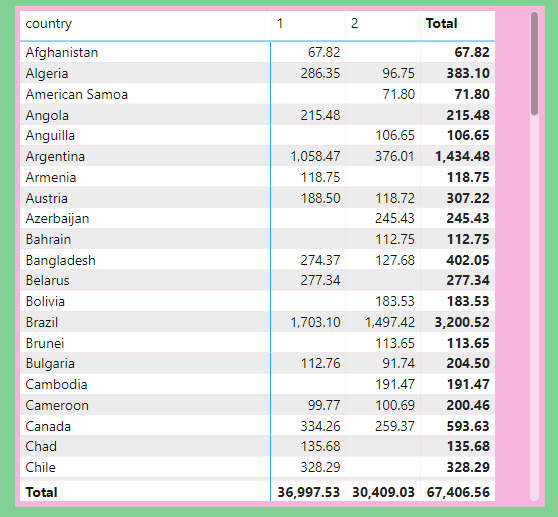
Description automatically generated**

**7.** **What is the distribution of staff by employment duration?**

**A pink and green rectangle with black text

Description automatically generated**

**8.** **How does the store performance vary by location?**

****

**9.** **What is the average rental duration by staff member?**

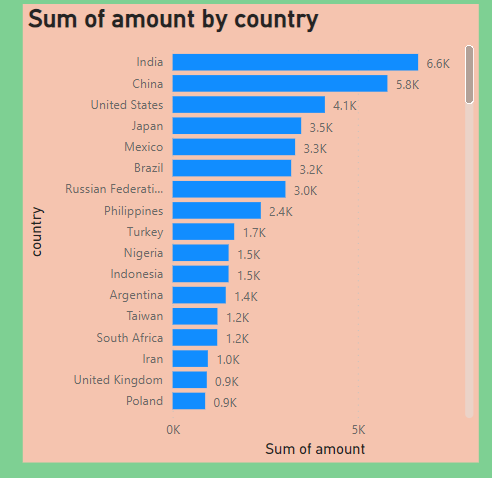
**A purple and green rectangle with black text

Description automatically generated**

**10.** **What is the distribution of customers across different cities?**

****

**11.** **How does the rental revenue vary by country?**

****

**12.** **Which locations have the highest and lowest customer ratings?**

**A green and pink rectangle with black text

Description automatically generated**

**13.** **What is the distribution of films by language?**

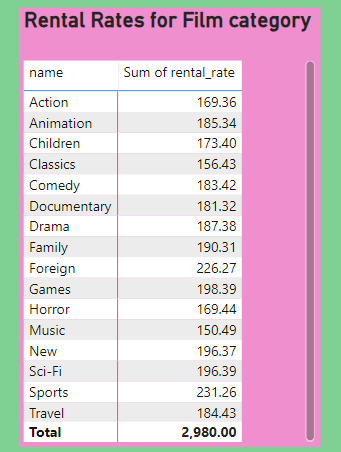
**A screenshot of a computer

Description automatically generated**

**As we know , here is only one language in film inventory also we can only focus one the English language due to lack of the data given in dataset .**

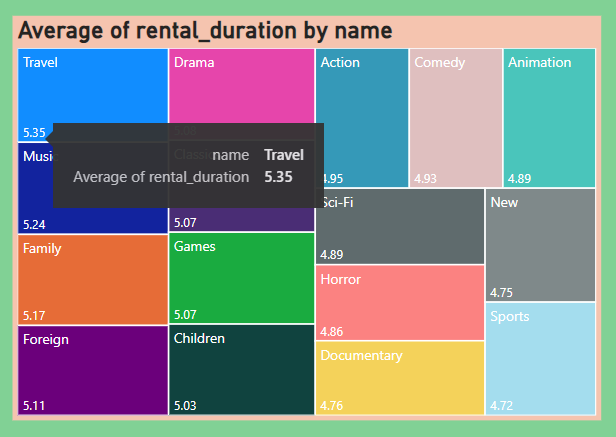
**So take pie chart for language as name.**

**14.** **Which film categories have the highest rental rates?**

****

**Taking name from category and sum of rental rate from the film table , here we can see that different film category having different rental rate.**

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

****

**Describing the Dashboard in power bi**

**A screen shot of a computer

Description automatically generated**

**For dashboard, 1st describe the overview of the detailed description of the problem statement linked with film analysis.**

**In the film analysis, we focus on the comprehensive evolution of the performance and focus on pattern of trends .**

**In SALE REVENUE ANALYSIS , focus on the financial evaluation and inner structure of the business analysis . also explain decision making process and focus on the improvement of the sales and revenue.**

**In STORE & INVENTORY ANALYSIS , we focus on the management of inventory , insight of the business performance as well as the store analysis.**

**FILM ANALYSIS DASHBOARD**

**A screenshot of a computer screen

Description automatically generated**

**ALL THE DETAILED information of the analysis related to the film shown in singal dashboard , giving detail information about the industry and also clear the vision to understand the dataset.**

**SALES REVENUE ANALYSIS DASHBOARD**

**A close-up of a graph

Description automatically generated**

**Problem statement related with the sales revenue analysis , all are combined in a single dashboard and giving the detail information about the sales revenue in better manner . Dashboard gives the great insight of all the information here.**

**STORE & INVENTARY ANALYSIS**

**A screenshot of a computer

Description automatically generated**

**Problem statement related with the STORE & INVENTORY ANALYSIS , all are combined in a single dashboard and giving the detail information about the store & inventory as well as revenue in better manner . Dashboard gives great insight of all the information here.**

**CONCLUSION**

**The "Sakila DVD rental database" is a popular sample database used for practicing SQL and data analysis. It contains various tables related to a hypothetical DVD rental store, such as customer data, rental history, film details, and staff information. Without specific details about the analysis, I can offer a general conclusion that could apply to various analyses of this dataset:**

1. **Customer Behavior**
2. **Film Insights**
3. **Staff Performance**
4. **Revenue and Profitability**
5. **Inventory Management**
6. **Seasonal Trends**
7. **Customer Segmentation**