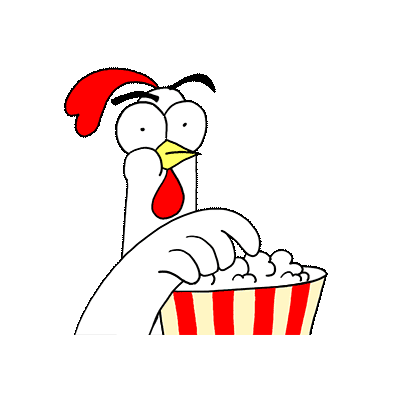
CHAKIT VERMA

saaa

SAKILA DVD MOVIE RENTAL

BY CHAKIT Verma

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
* Data cleaning
* MECE-Breakdown
* Understanding & solving Power BI Questions
* Dashboard
* Understanding & solving EDA Questions

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 data set 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:

**The actor Table.**

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.

**The address Table**

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.

**The category Table**

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.

**The city Table**

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.

**The country Table**

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.

**The customer Table**

The customer table contains a list of all 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.

**The film Table**

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.

**The film\_actor Table**

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.

**The film\_category Table**

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.

**The film\_text Table**

The film\_text table contains the film\_id, title and description columns of the film table

**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.

The contents of the film\_text table should never be modified directly. All changes should be made to the film table instead.

**The inventory Table**

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

**The language Table**

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.

**The payment Table**

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.

**The rental Table**

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.

**The staff Table**

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.

**The store Table**

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 cause it contain no value.

.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 in a one-to-many relationship.*

**Overview of Transformed dataset**

**Entity Relationship Diagram**

A screenshot of a computer screen

Description automatically generated

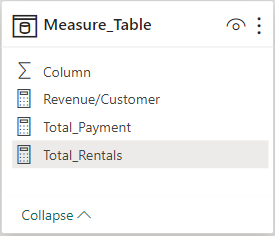
**MECE-Breakdown:**

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.**

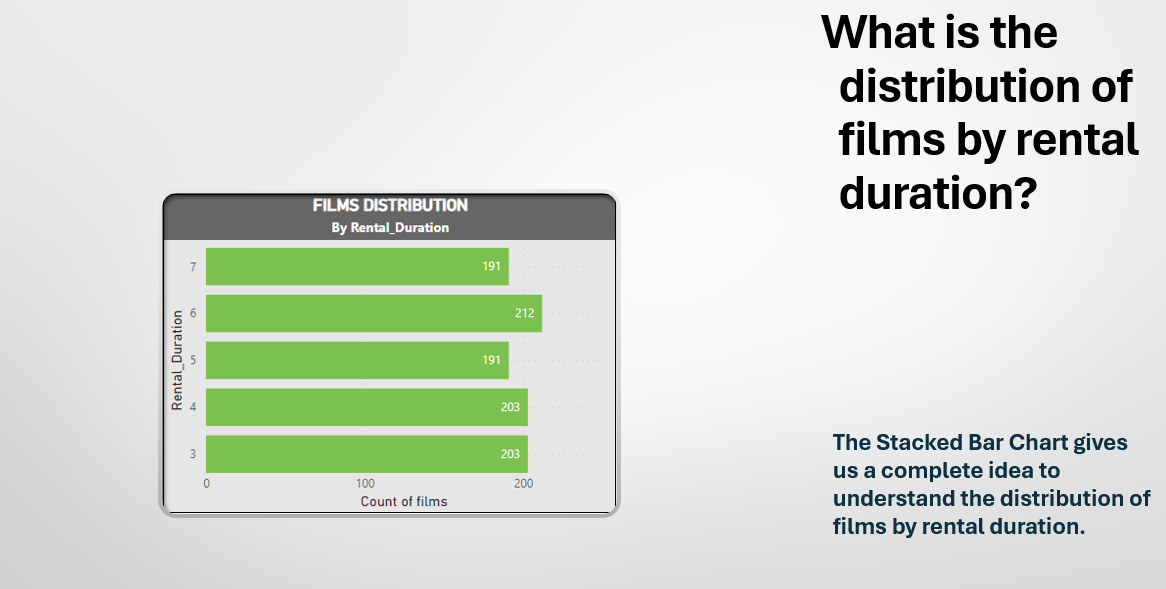
* **Before doing any analysis, I have created a measure table contain all measure that are used in in Analysis in power bi**

****



A screenshot of a computer

Description automatically generated

****

**A screenshot of a graph

Description automatically generated**

**A screen shot of a computer

Description automatically generated**

**A screenshot of a graph

Description automatically generated**

**A screenshot of a computer

Description automatically generated**

A screenshot of a computer

Description automatically generated

**A screenshot of a computer

Description automatically generated**

**A screen shot of a computer

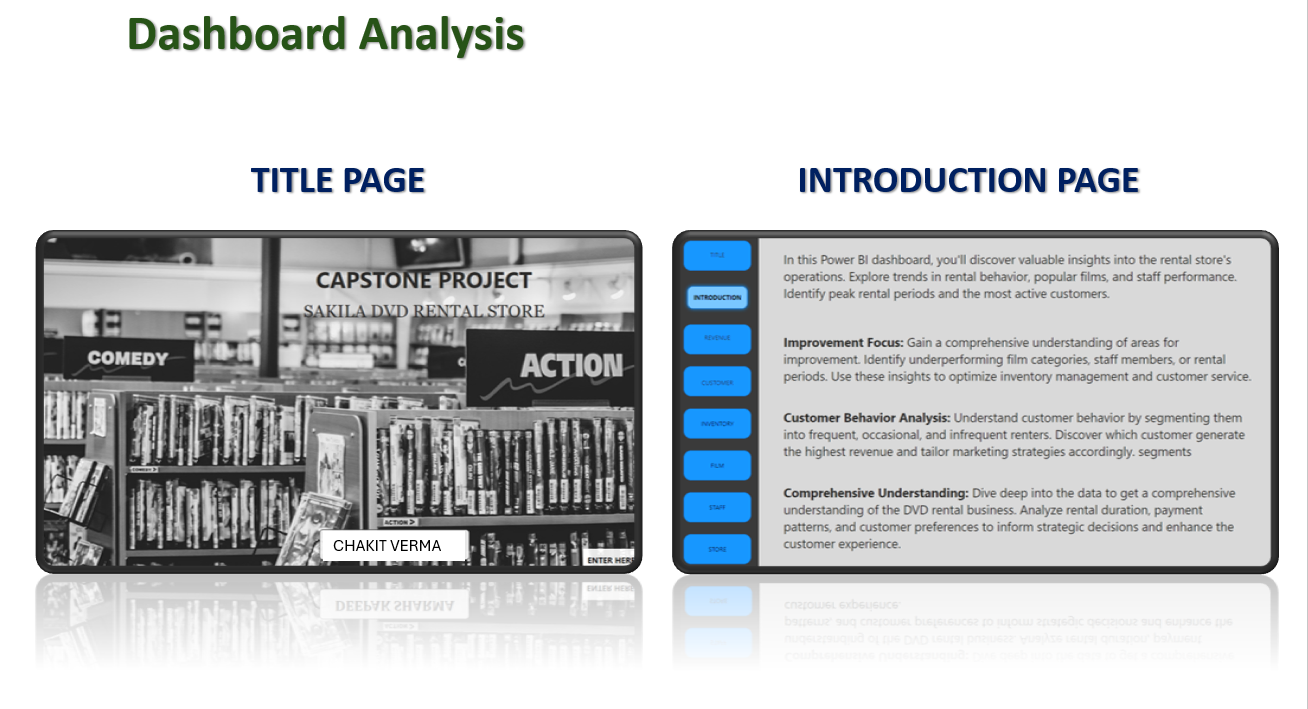
Description automatically generated**

**A screen shot of a computer

Description automatically generated**

**A screen shot of a bar chart

Description automatically generated**



A screenshot of a computer

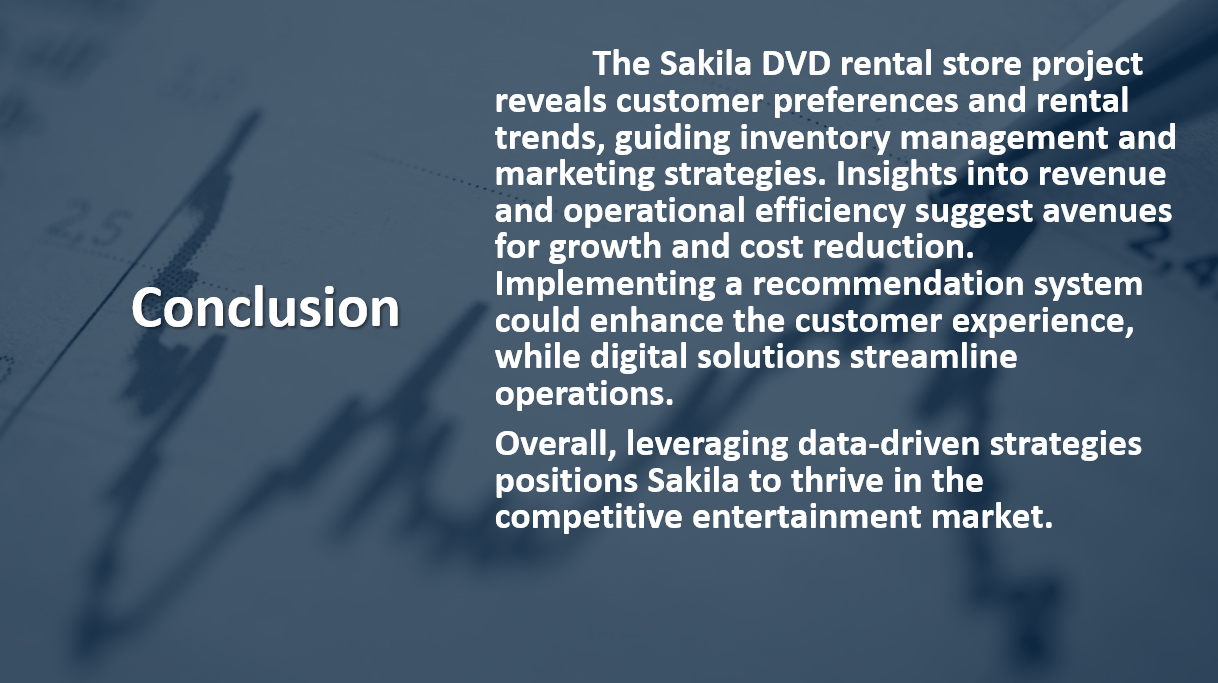
Description automatically generated

**A screenshot of a computer

Description automatically generated**

**A screenshot of a computer

Description automatically generated**

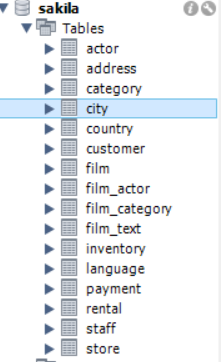
****

Understanding & solving EDA Question’s

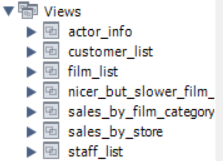
**Question’s of EDA**

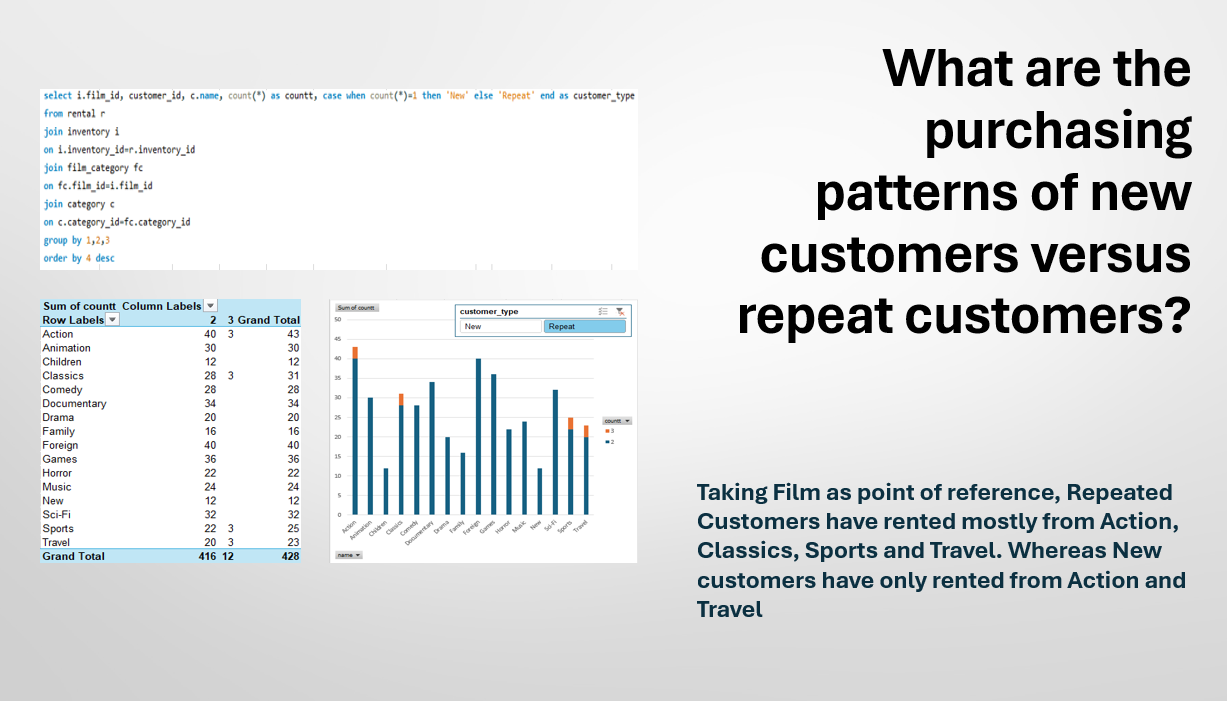
**Before starting solving the EDA questions :**

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

****

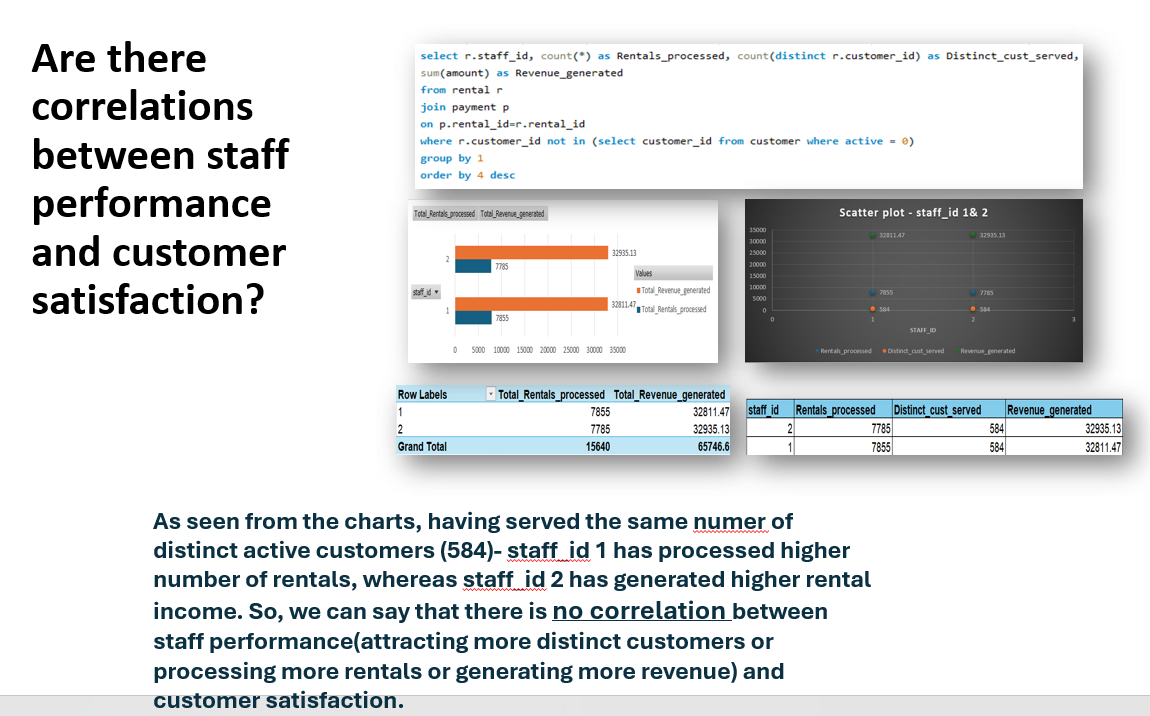
* **After loading created a view in SQL server of each table that are necessary for solving the EDA Questions.**

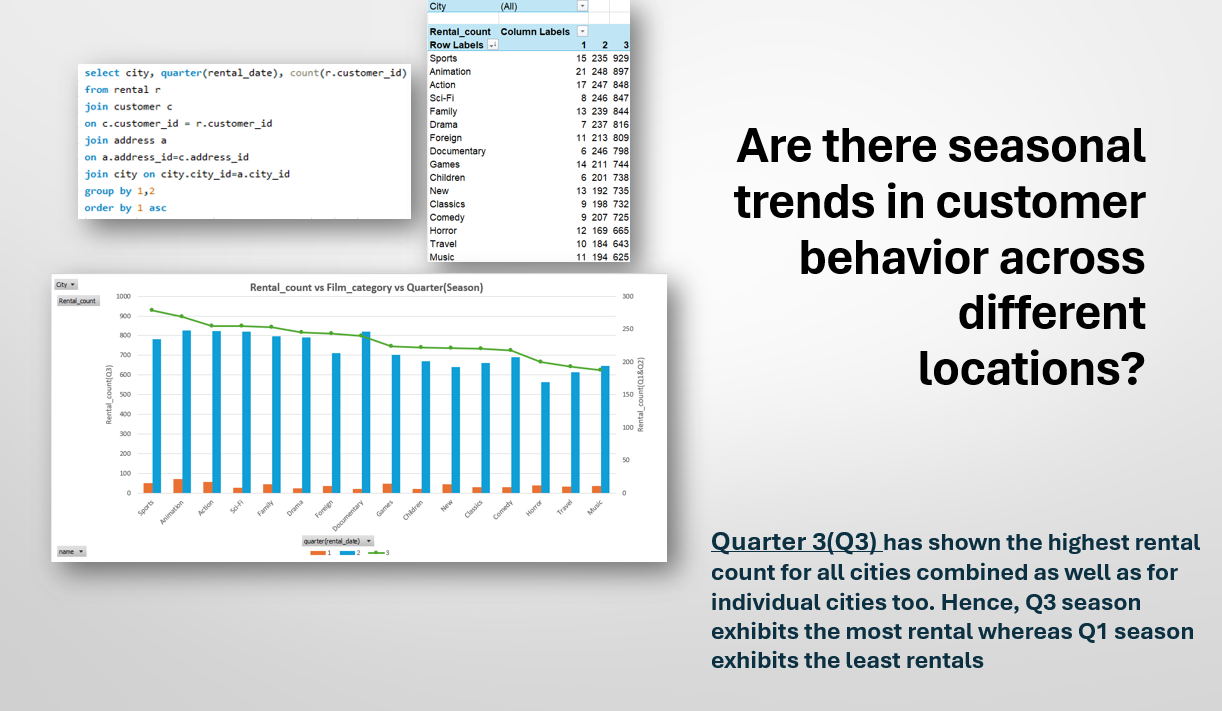
****

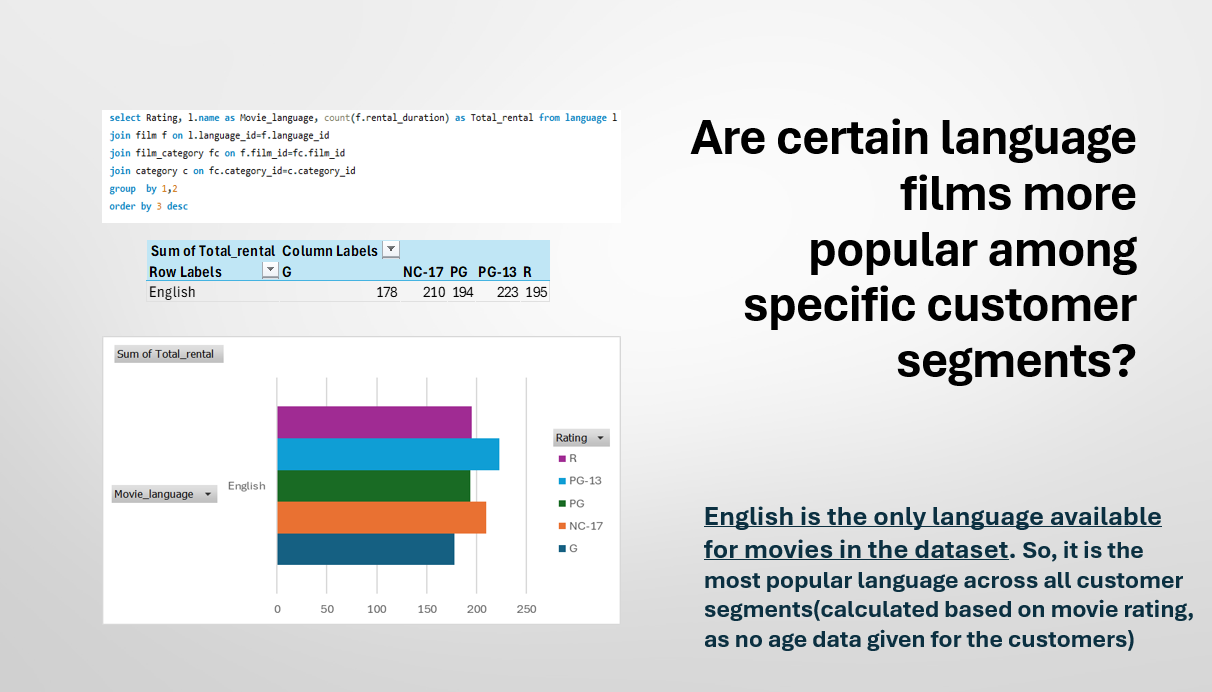
****

A screenshot of a computer

Description automatically generated



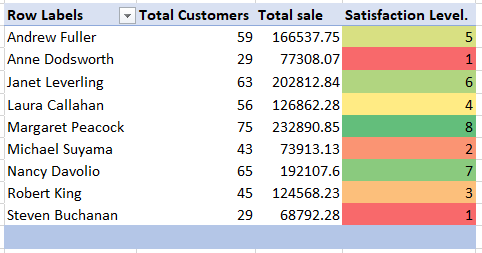


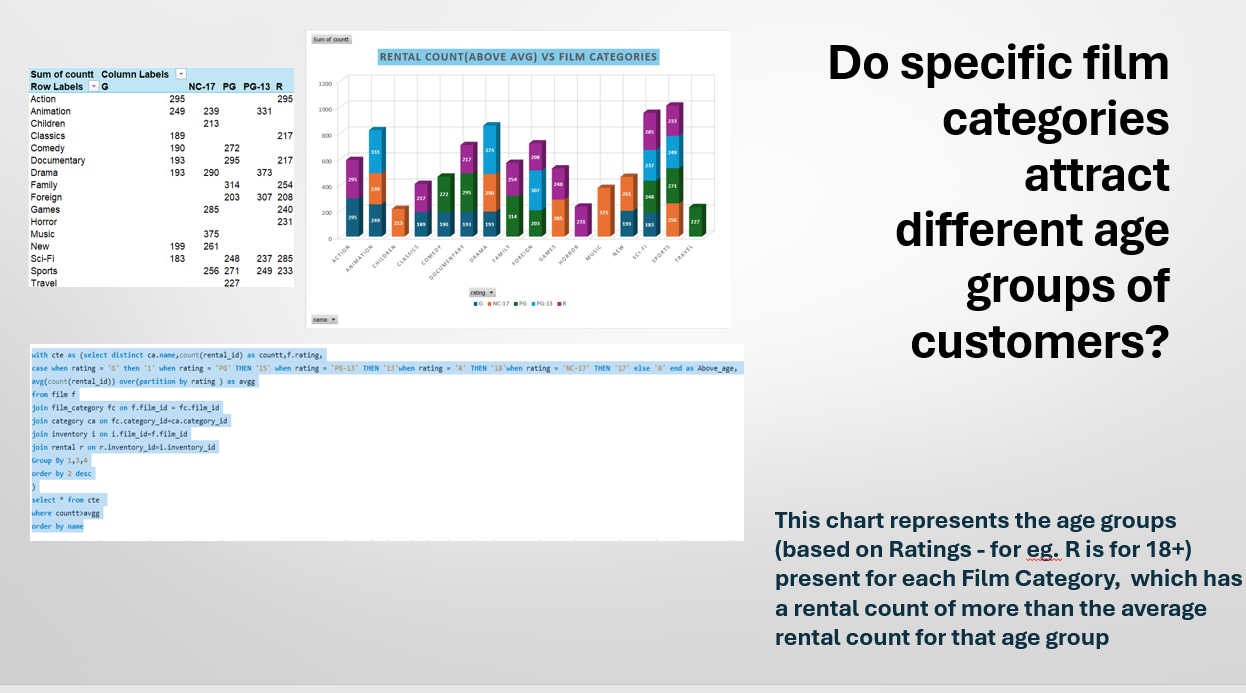


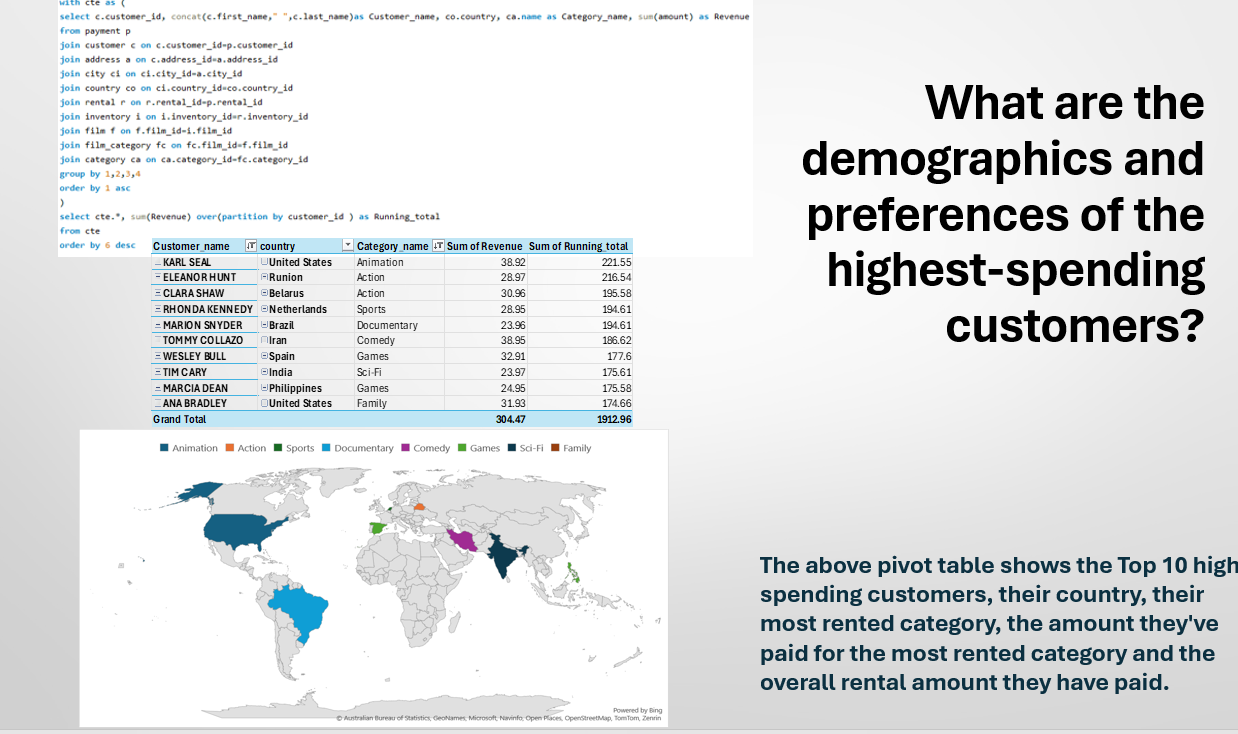


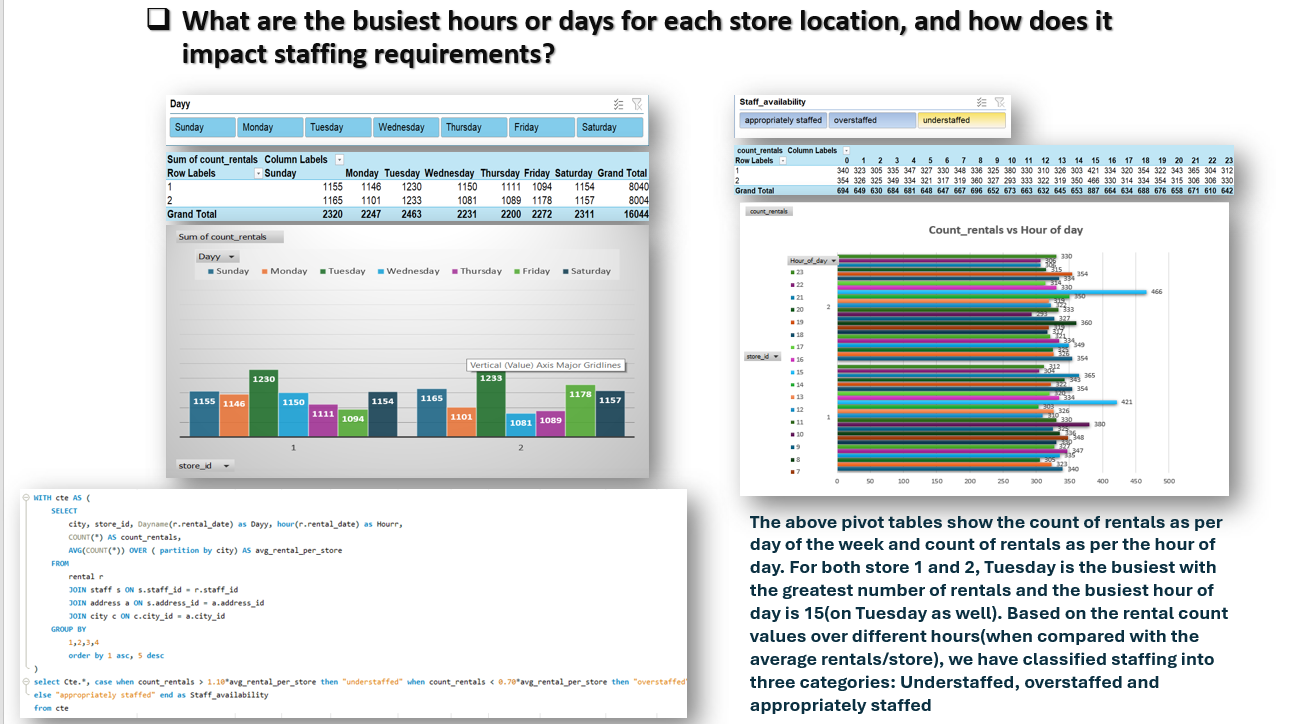
A screenshot of a computer

Description automatically generated



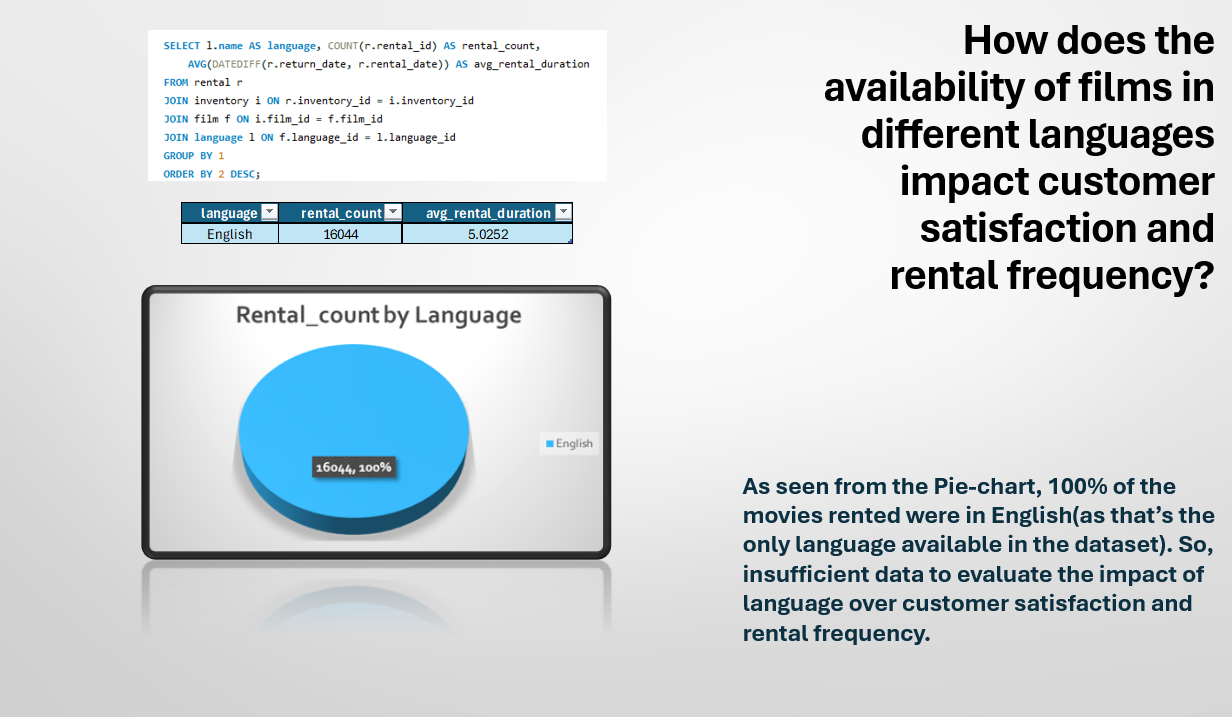






A screenshot of a computer

Description automatically generated



**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.