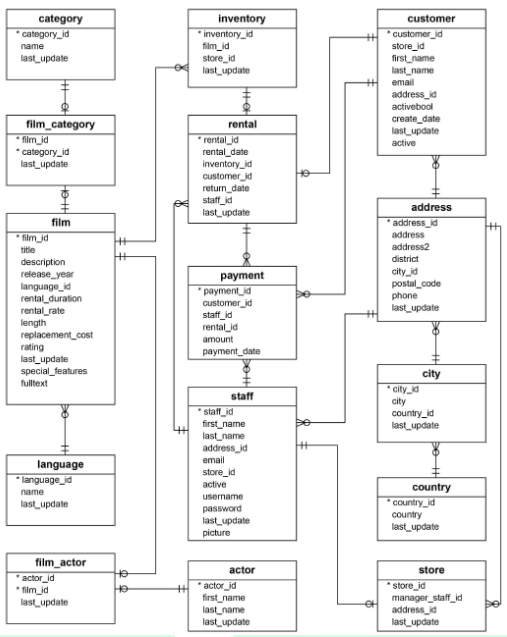
CAPSTONE PROJECT – DVD RENTAL BUSINESS ANALYSIS

-SHREYAS REDDY KANTHAREDDY

Problem Statement: Design and implement a data-driven solution using the DVD rental database to perform descriptive analysis showcasing mean, median, distribution and variability. This involves preparing the dataset by joining various tables in SQL and, visualizing the insights through an interactive dashboard prepared using Power BI or tableau. The goal is to simulate a real-world business scenario to develop hands-on expertise in data preparation, SQL, and business intelligence tools like Power BI or Tableau.

Entity-Relation model:



Approach:

1. Downloaded the compressed zip file of the DVD rental dataset.
2. Prepare and clean the data using SQL to create a consolidated table for analysis.

* First, I studied the entity relationship diagram which is in the previous slide.
* There were 15 tables in the entire data set.
* By using the SQL functions on the 15 tables, I created a final table consisting of 26 columns.

1. Extracted the table in the CSV format.
2. Uploaded the CSV file into Power Bi and started to find patterns and trends.

SQL:

* There were 15 tables in the dataset as I mentioned earlier.
* Tables:

1. Payment
2. Rental
3. Inventory
4. Staff
5. Actor
6. Customer
7. Address
8. City
9. Country
10. Store
11. Category
12. Film Category
13. Film
14. Language
15. Film Actor

* I used the command called JOIN to join all 15 tables on their respective primary and foreign keys to create a final table.
* In the final table, there were 26 columns which are later used to discover the patterns and trends.
* Extracted the data into CSV file.

POWER BI:

* Uploaded the CSV file into Power Bi to analyze the trends and patterns.
* I classified the data into segments for better understanding.

1. Location Analysis
2. Revenue Analysis
3. Customer Analysis
4. Rental Analysis
5. Film and Actor Analysis

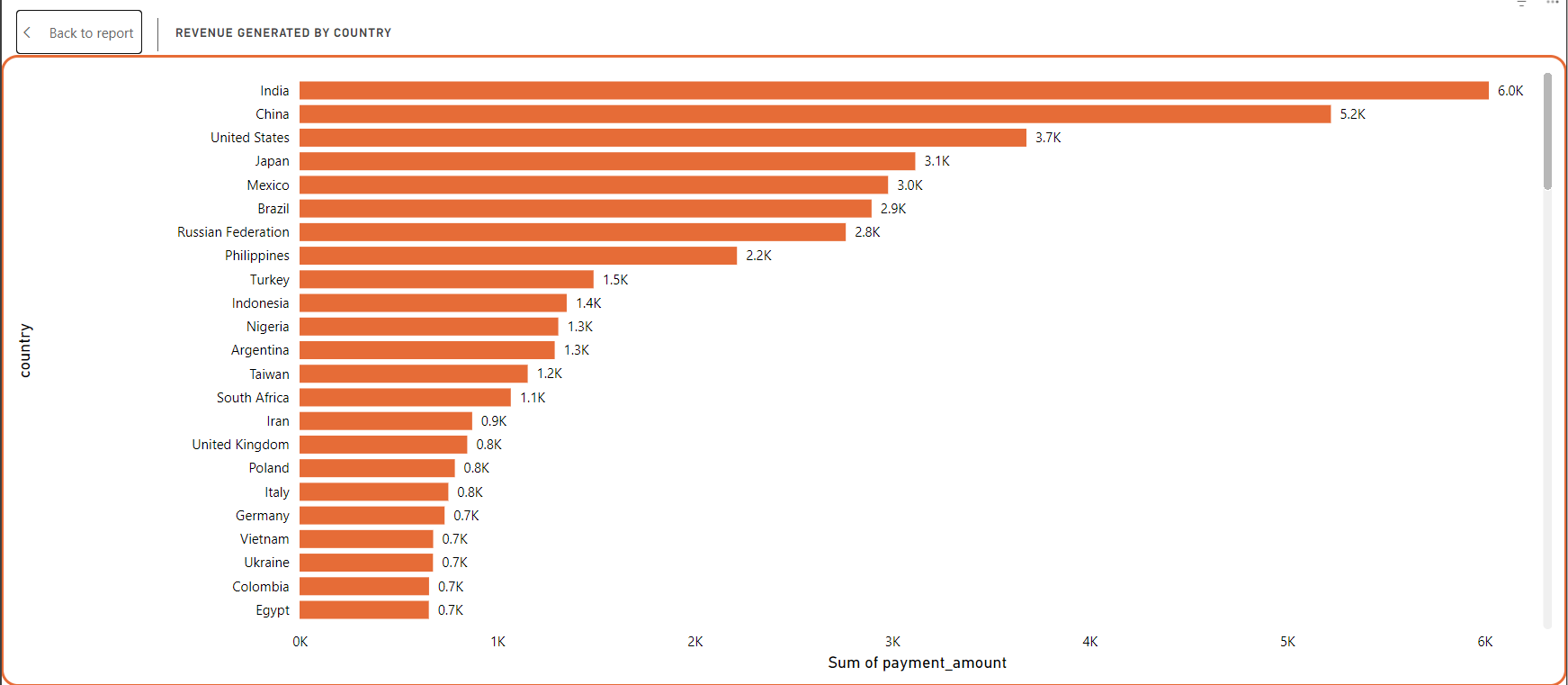
* By using the final table, I have analyzed the above elements.

1. Location Analysis:

A screenshot of a computer

Description automatically generated

* There are 108 countries, 597 cities, and 376 districts in the dataset.
* I have used 2 slicers to filter the dashboard.
* Revenue generated by countries:

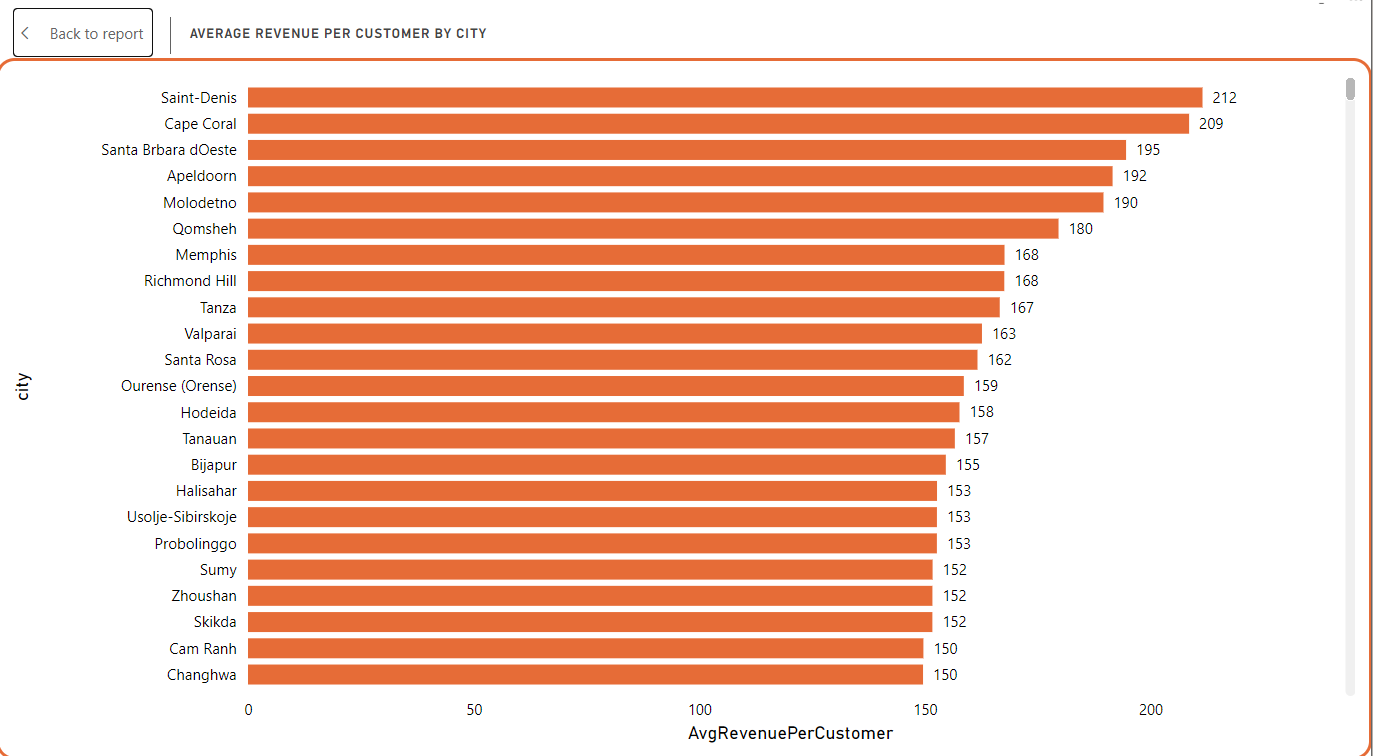


* The first chart is about the revenue generated from each country.
* I have used a bar chart to analyze this task.
* It is seen that India is generating the highest revenue of $6022, followed by China and USA with $5222 and $3680 respectively.
* The top three countries contribute 25% of the total revenue.
* District with most customers:

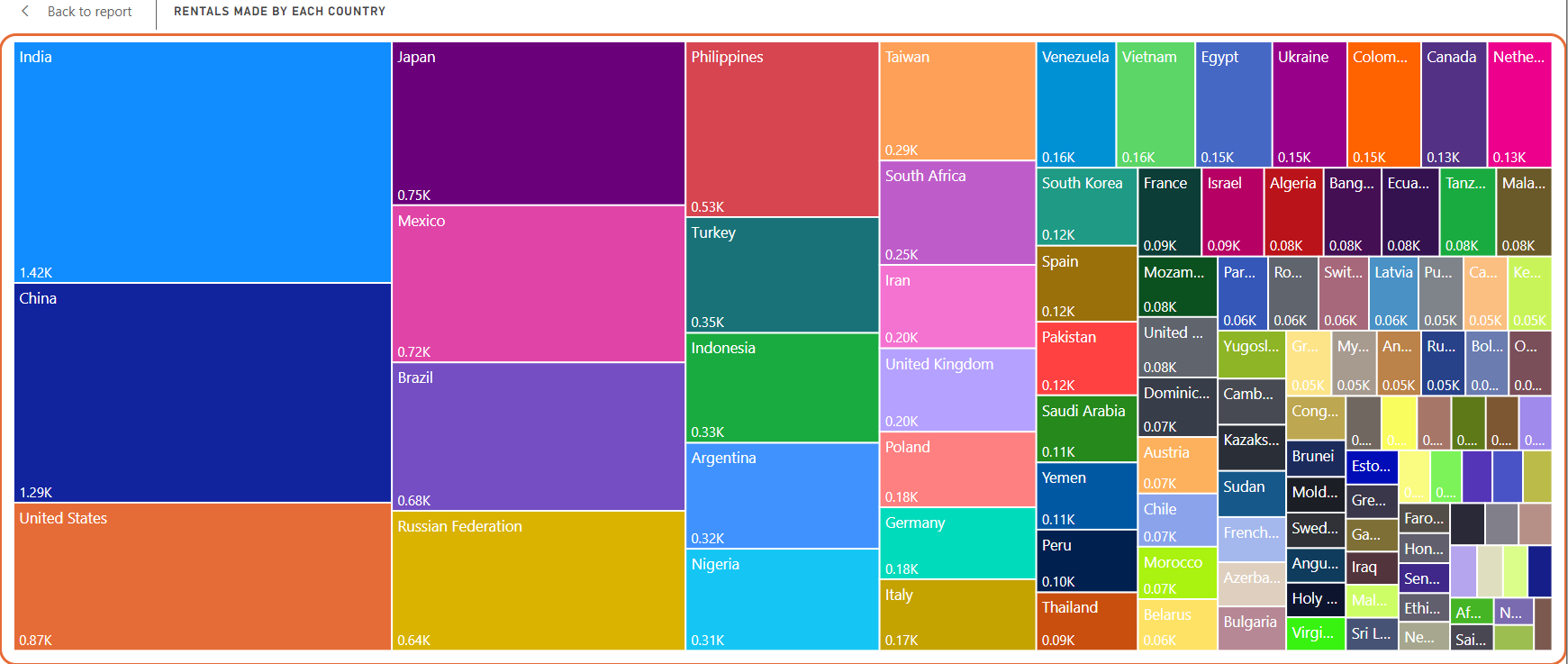
A graph of a bar graph

Description automatically generated with medium confidence

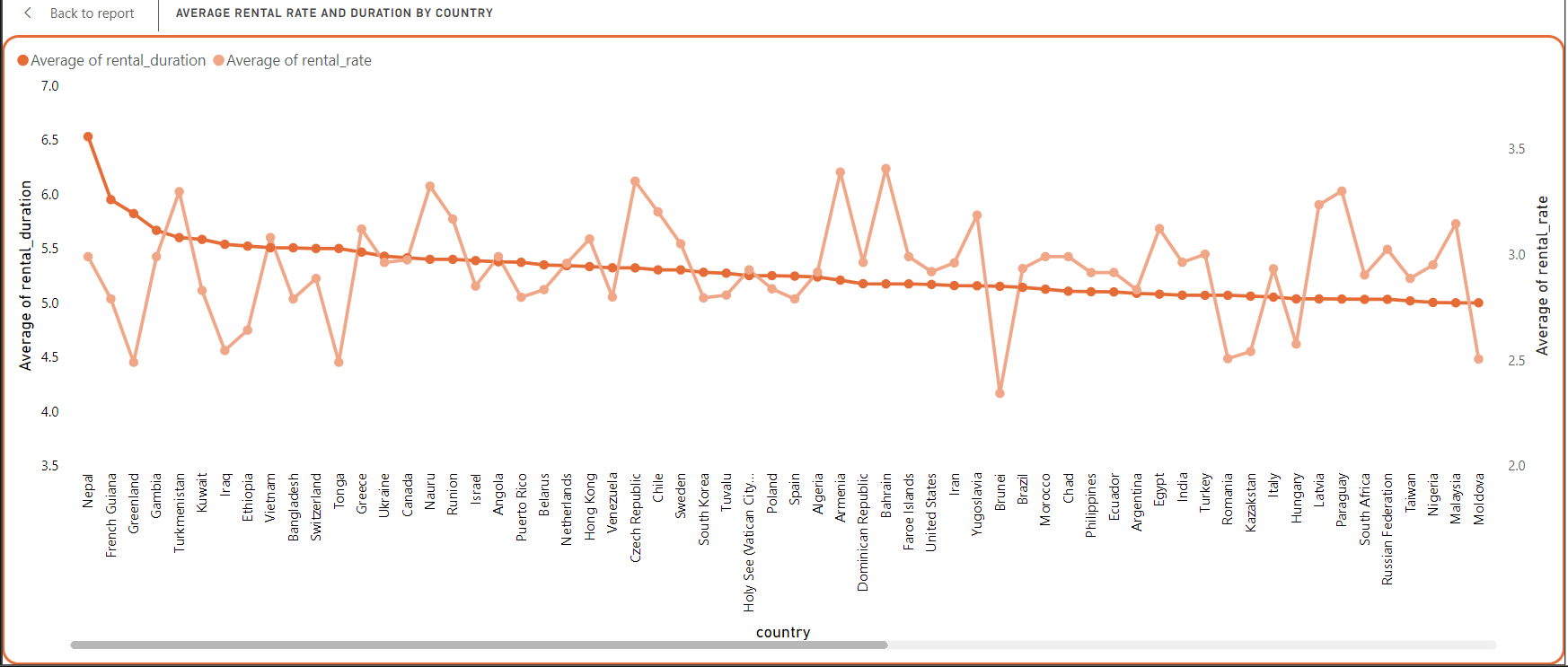
* A column chart is used to analyze the district with most number of customers who rent the DVD.
* Most number of customers are from Buenos Aries in Argentina with 251 customers followed by California in USA with 224.
* Average amount spent by customer in each city:



* This bar chart shows the average amount spent by customers on rentals in each city.
* Saint-Denis has the highest numbers of customers who spend on rentals with $212, followed by Cape Coral in USA with $209.
* The average ranges between $28 to $212.
* Rentals made by each Country:



* There are 14,558 rentals in total.
* I have used treemap chart to analyze the total rentals made by each country.
* India has the highest rentals with 1419 rentals followed by China and USA with 1291 and 870 rentals respectively.
* America Samoa, Tonga, and Nepal has the least rentals with 15,16, and 17 respectively.
* Average rental rate and duration by country:



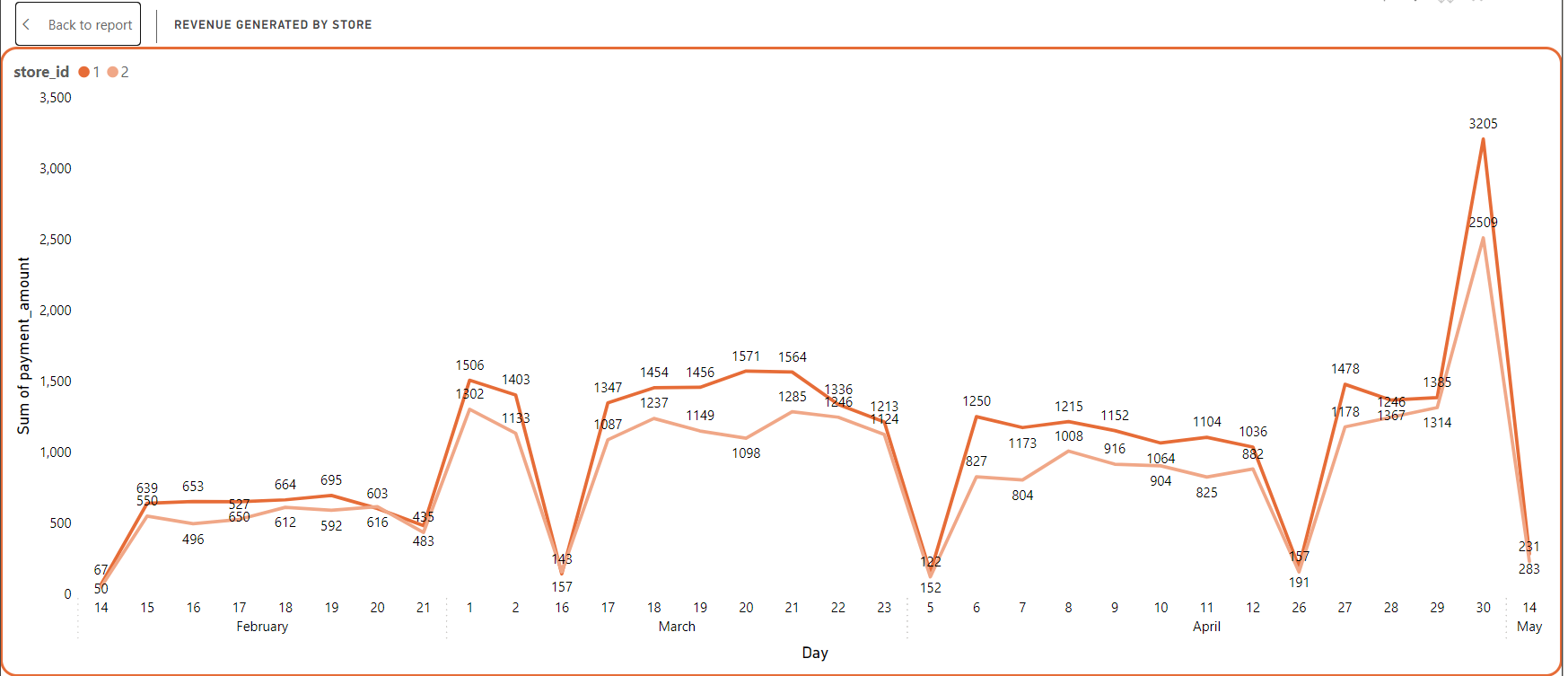
* A line chart is used to analyze the average rental rate and duration in different countries.
* The first line depicts average rental duration, and the second line depicts average rental rates.
* By using a country slicer, we can easily identify both the parameters instantly.

1. Revenue Analysis:

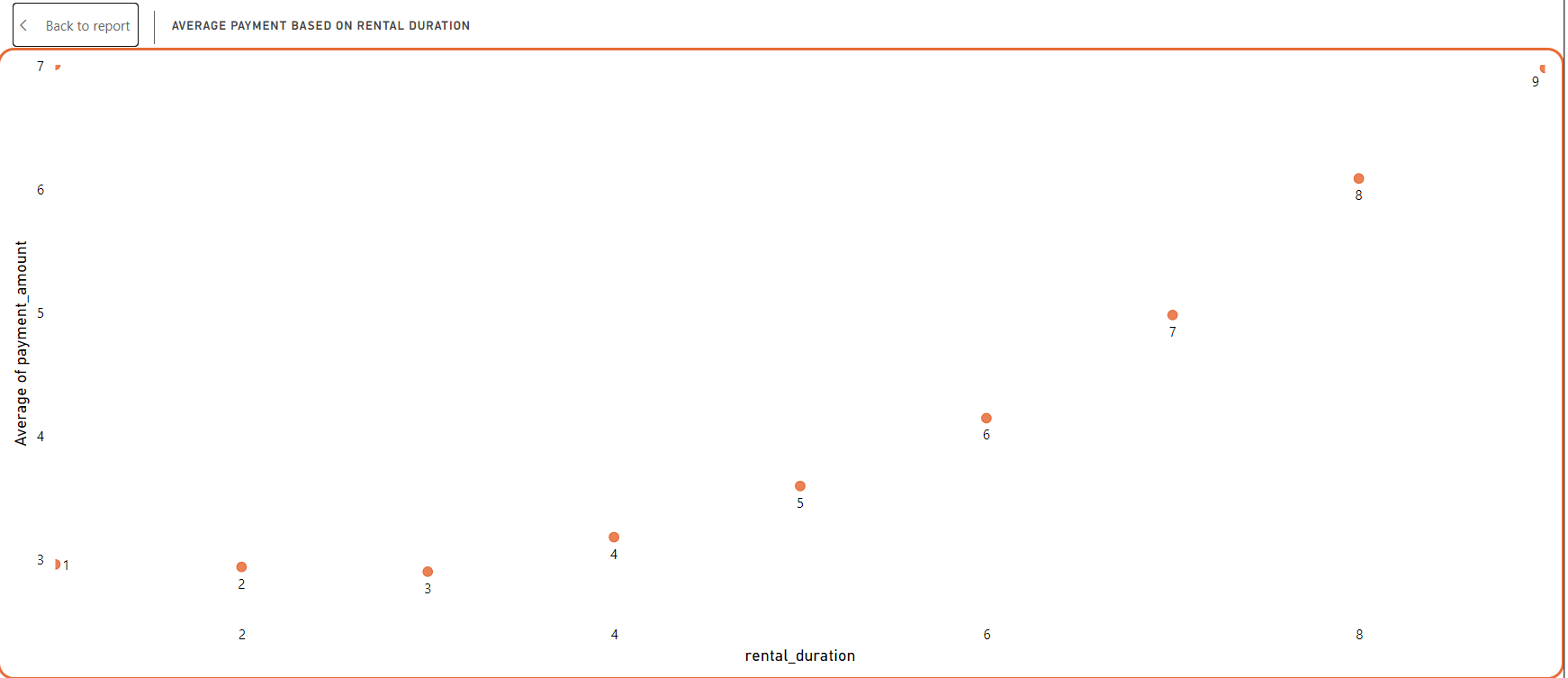
A screenshot of a graph

Description automatically generated

* Total revenue collected $61,120, average rental rate across all the countries is $2.94, and the total number of payments done by the customers is 14558.
* Revenue generated by the stores:



* There are two stores in the data set.
* A line chart is used to analyze this task.
* $61,120 is the total amount generated by the stores.
* It is seen that store 1 generated more revenue or given more sales when compared to store 2.
* Store 1 generated $33503 and store 2 generated $27617.
* Average payment based on rental duration:

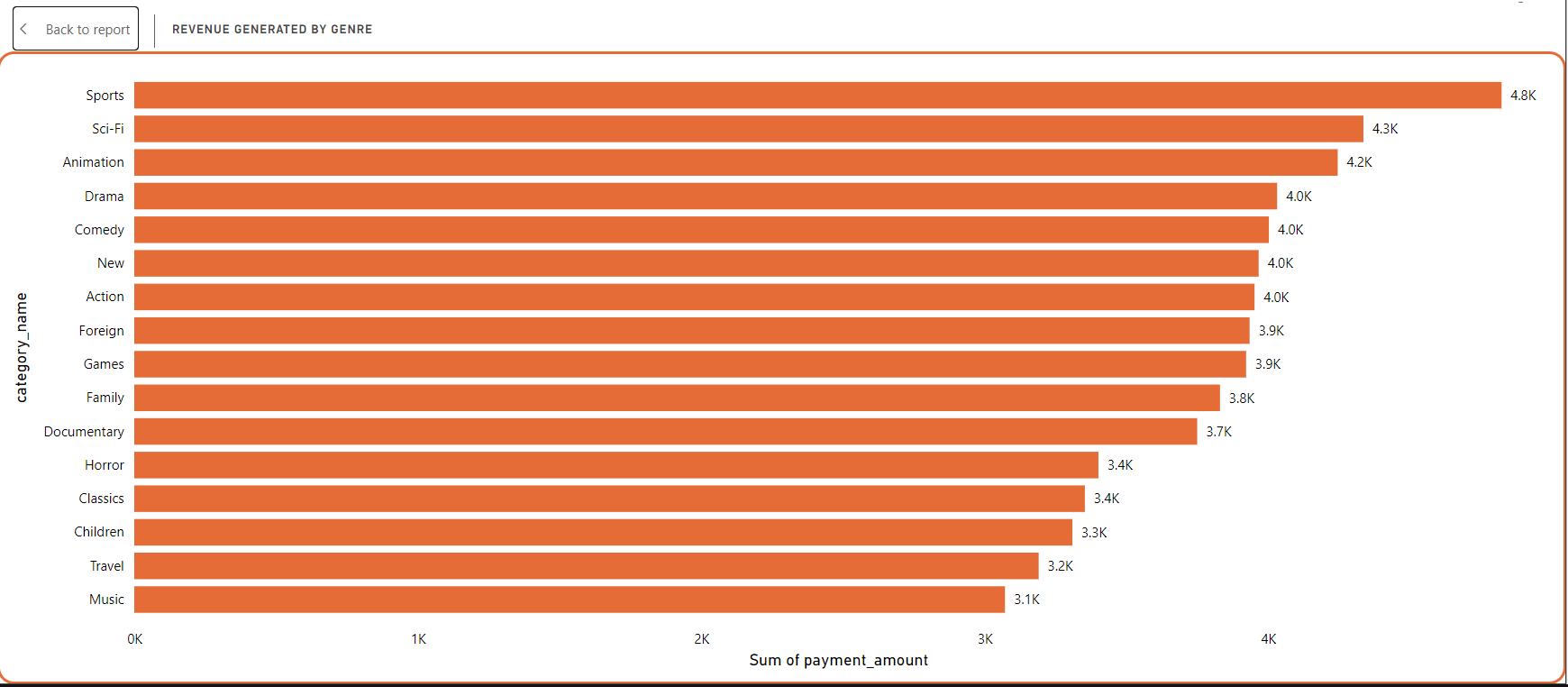


* A scatter plot chart is used to know the average payment based on the rental duration.
* It is seen that the average rental payment keeps increasing if the rental duration increases.
* When the DVD is rented for 1 day the avg payment is $2.96 and the duration is 9 days the payment is $6.98 almost $7.
* Revenue earned based on rental duration:

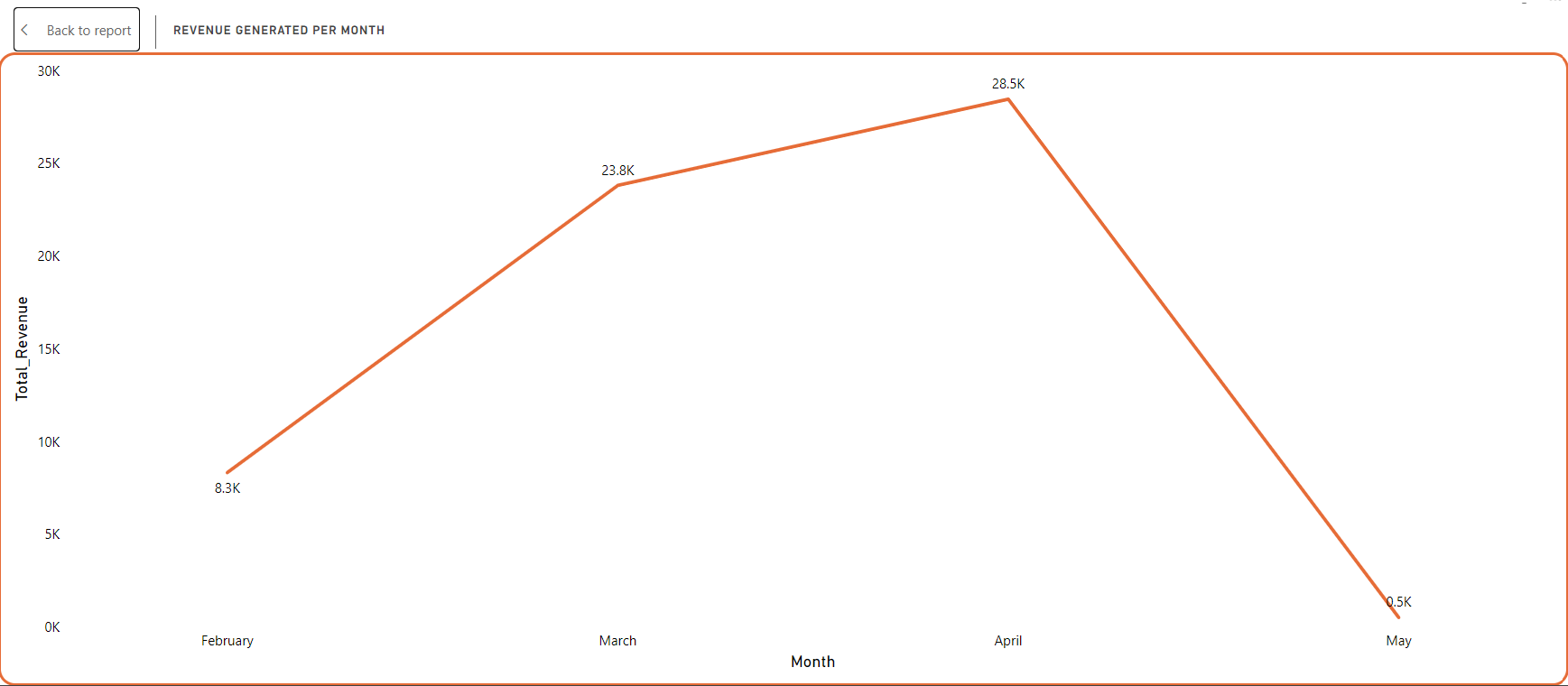
A colorful pie chart with white text

Description automatically generated

* A pie chart is used to analyze the task.
* Customers who have rented the DVD for 9 days are being charged more. $11,188 which is almost 18.46% of the total revenue collected.
* Followed by customers who have rented for 8 and 7 days with $9905 and $8258.
* Revenue earned based on genre (film category):



* A bar chart is utilized to analyze the company's revenue generated from renting movies of specific genres.
* Sports genre movies earn the highest amount of $4823.
* Followed by Sci-Fi, animation, drama, and comedy.
* Genres such as music, travel, and children earn the least when compared to other genres which ranges between $3000-$3300.
* Revenue generated in each month:



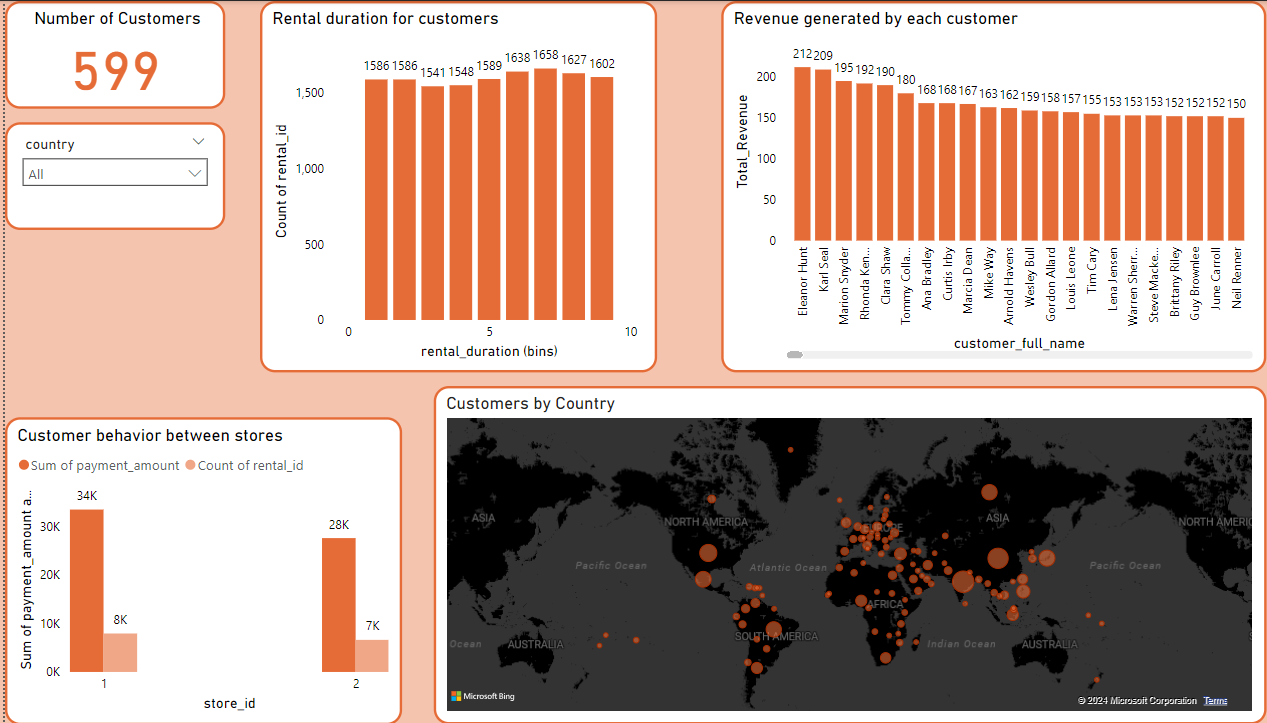
* A line graph is used to know the pattern.
* April has the most number of rentals with 6,736 and the amount is $28,462, followed by march with 5,629 rentals and the amount is $23,813.
* There is a sudden dip in the chart in May with only 182 rentals, fetching $514.
* Revenue generated by staff:

A screenshot of a computer

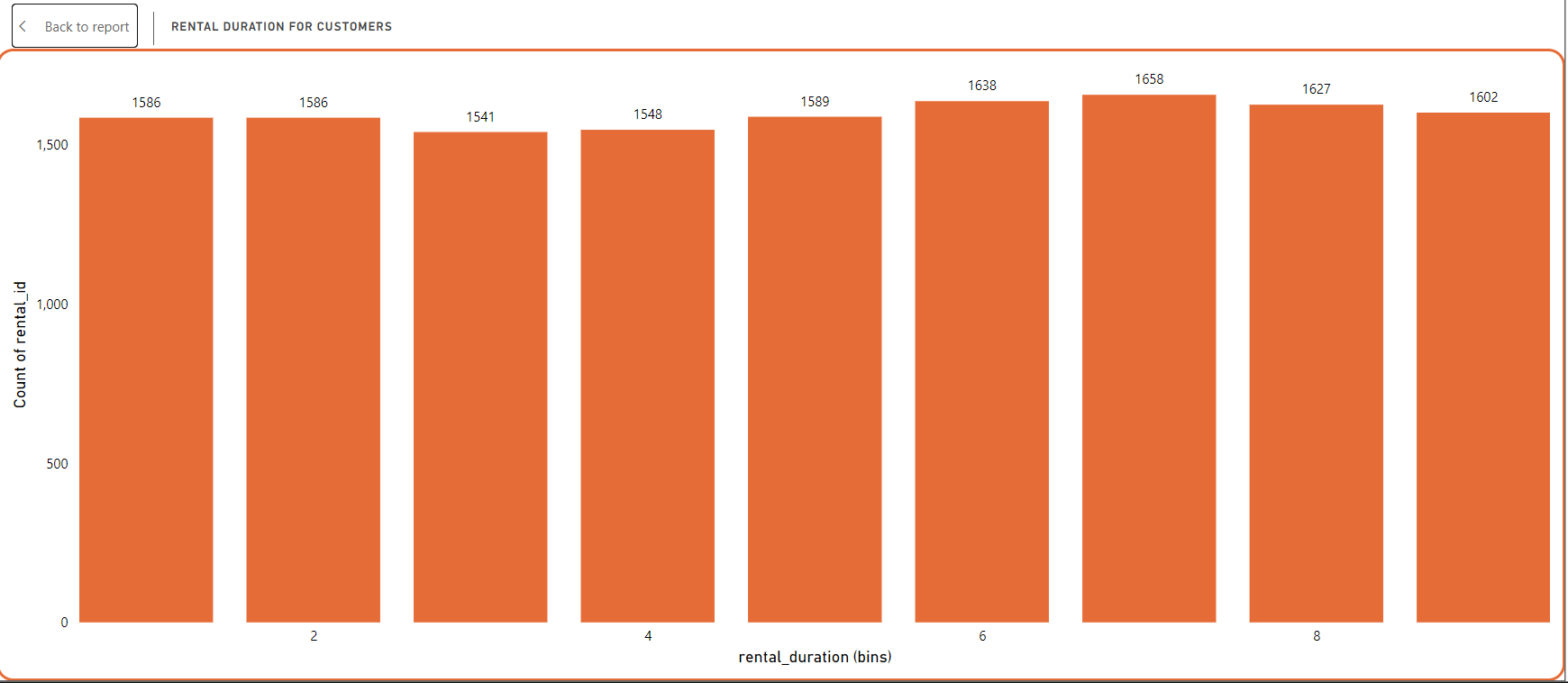
Description automatically generated

* A column chart is used to know which staff generated the most revenue.
* Jon Stephens has generated a bit more revenue for the company compared to Mike Hillyer.

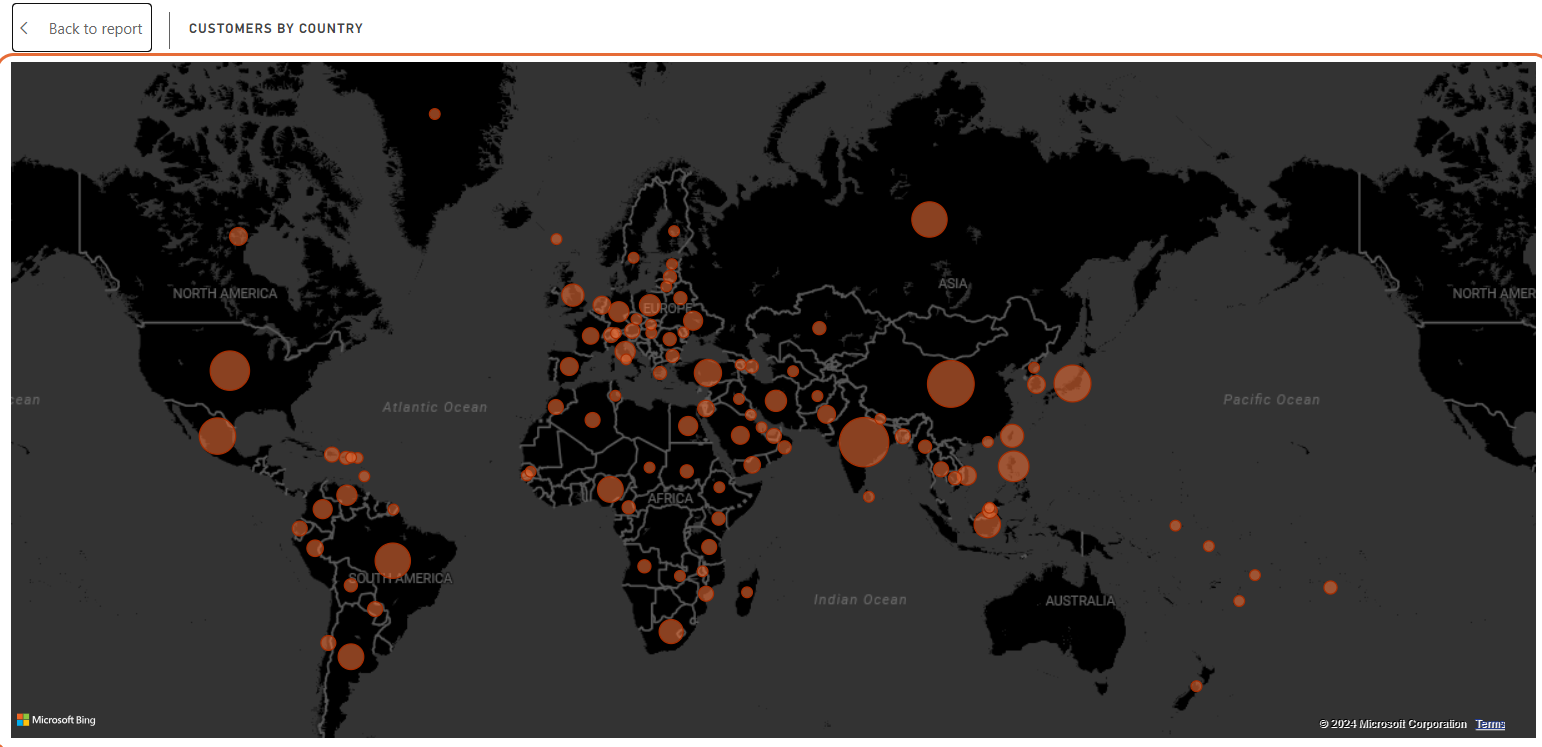
1. Customer Analysis:

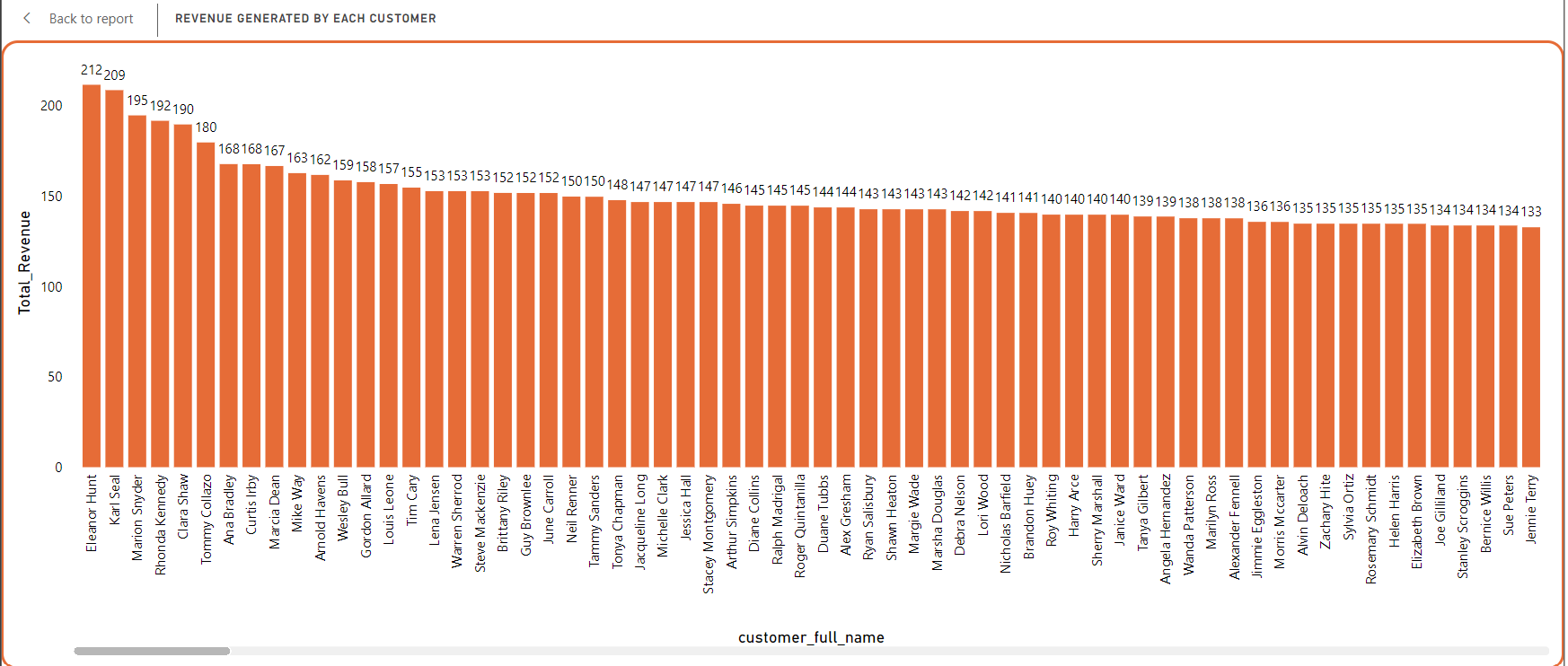


* There are 599 customers in the data set.
* Rental duration taken by customers:



* Used histogram to analyze the rental duration taken by the customers.
* Bins for 6 and 7 rental days show slightly higher counts with 1638 and 1658 respectively, indicating these durations are more popular among customers.
* Customers in each country:

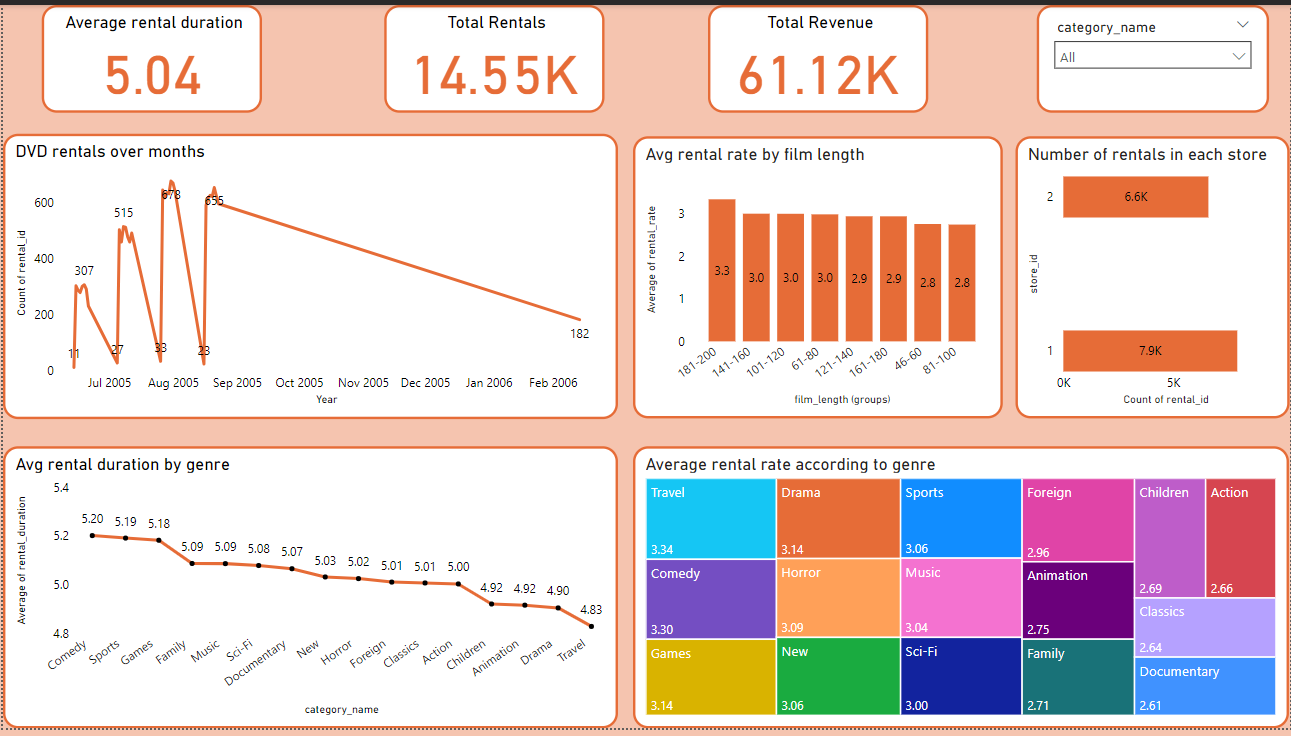


* Used map chart to represent number of customers who rents the DVD across the world.
* There are 599 total customers.
* India has the highest customers with 60 in number which is almost 10% of the total strength.
* Followed by China and USA with 53 and 36 in numbers respectively.
* Revenue generated by each customer:  
  
* Used column to interpret the task.
* By using country slicer, we can find the customers who have spent a lot on renting the DVD.
* Customer business based on stores:



* I have used clustered column chart to analyze this task.
* It is seen that store 1 has more customers than store 2.
* Store 1 has 7938 DVD rentals in total which earns $33,503 to the company.
* Store 2 has 6620 DVD rentals in total which earns $27,617 to the company.

1. Rental Analysis:



* The average rental duration from the customers 5 days. The total rental in numbers is 14558.
* DVD rentals over the months:

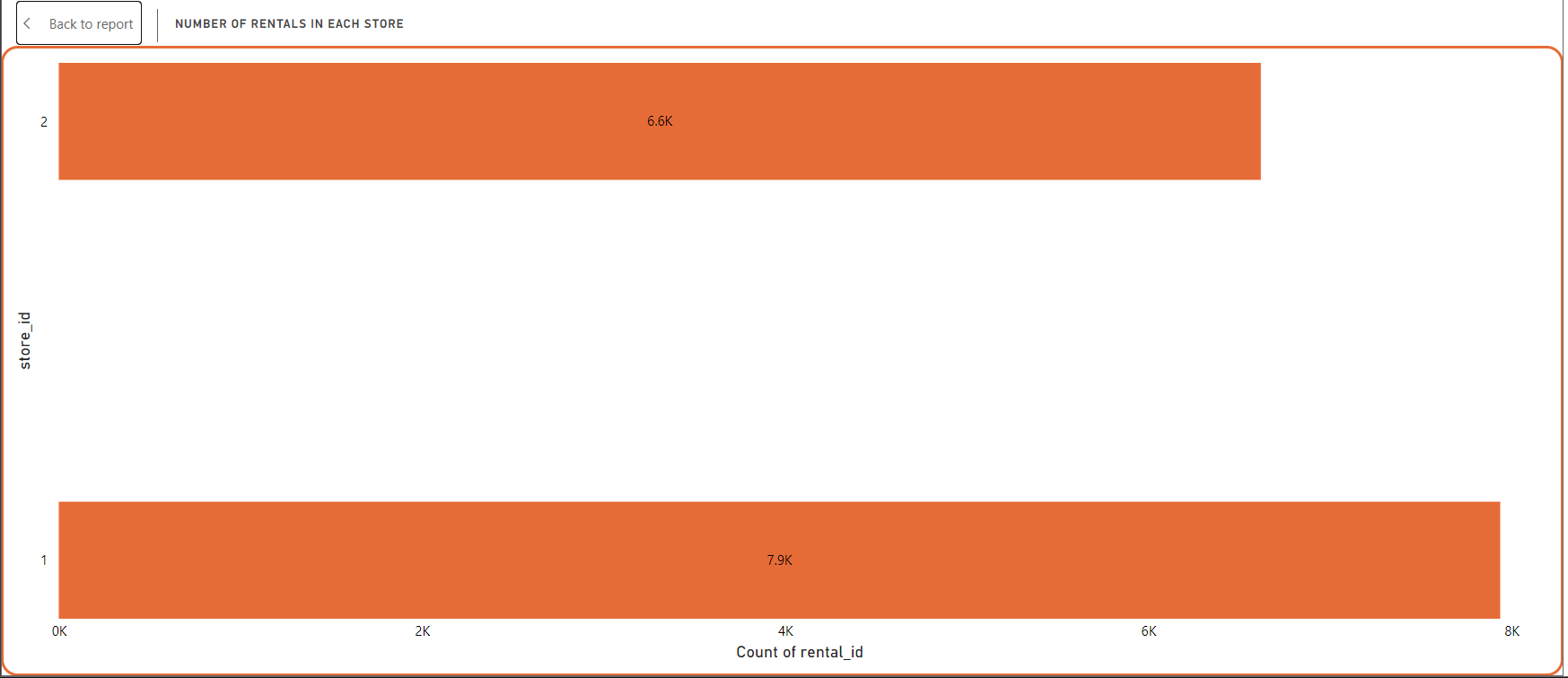


* A line graph is used to check the DVD rentals over the months.
* Between July 6th and August 21st customers have rented the most.
* Average rental rate by film length:

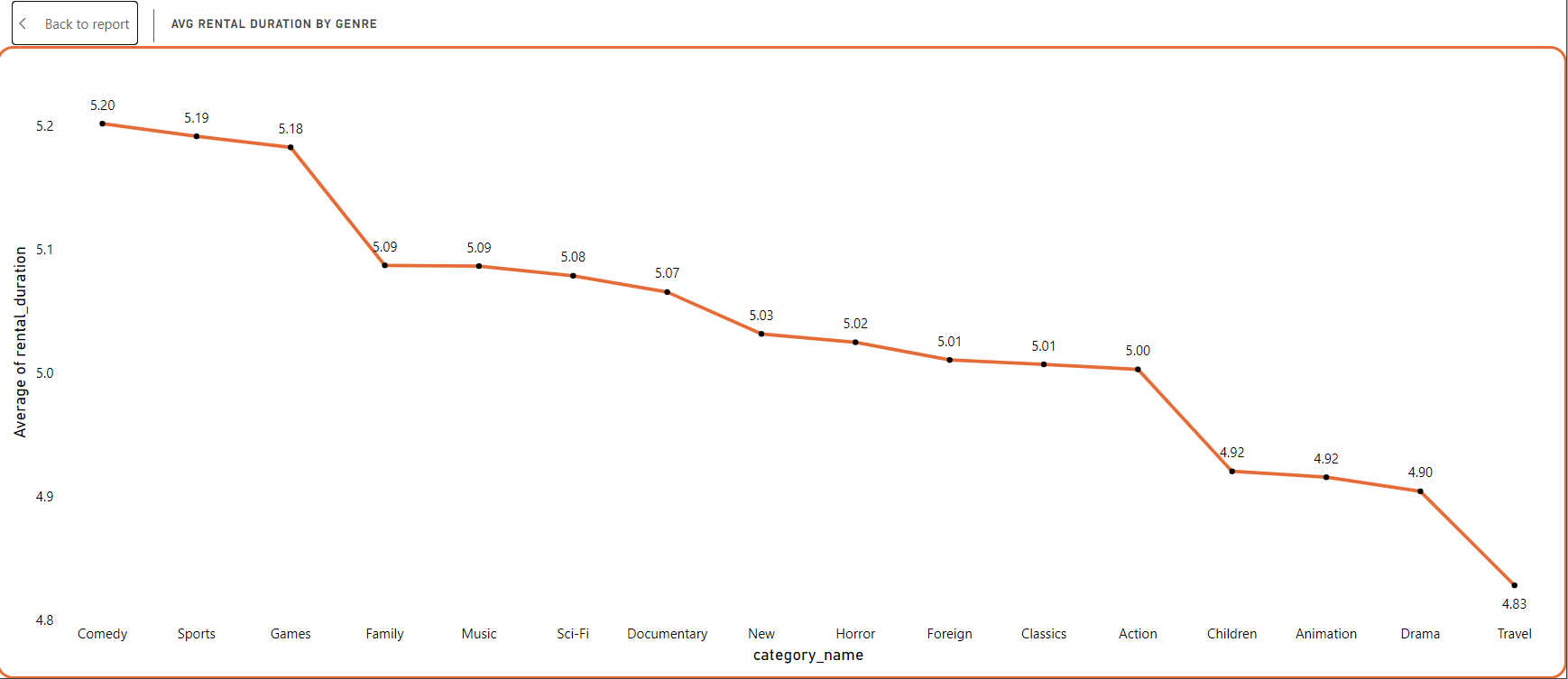
A screenshot of a computer screen

Description automatically generated

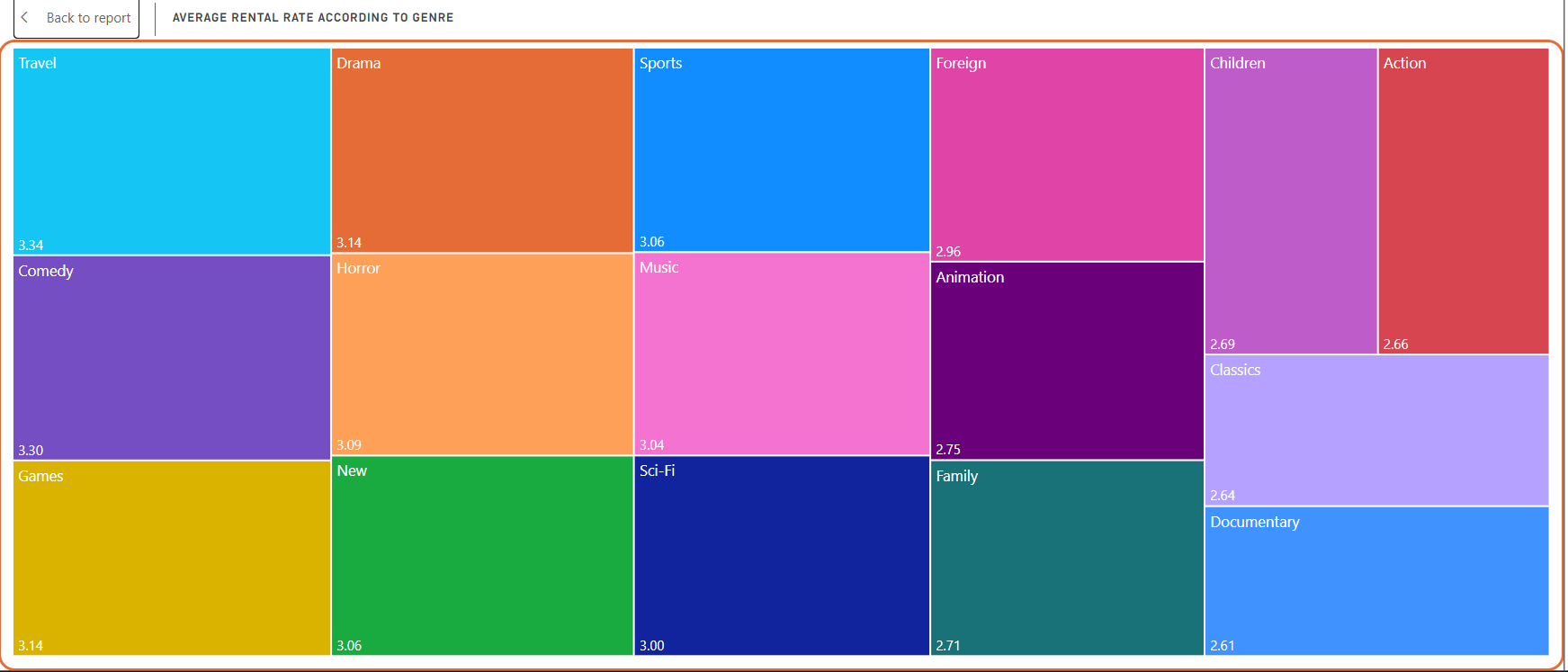
* I have used column chart to interpret the task.
* Created bins for film length with the interval of 20 mins (example: 181-200 mins, 141-160 mins).
* I did not much changes in average rental rate when the movie increases.
* Number of rentals in each store:



* Used a bar chart to know the number of DVD rentals in each store.
* Store 1 has 7938 DVD rentals.
* Store 2 has 6620 DVD rentals.
* Average rental duration based on genre (Film Category):



* I have used a line chart to analyze the average rental duration based on genres.
* The average rental duration varies between 5.20 and 4.83 days.
* Comedy has the longest rental duration, and travel has the least rental duration.
* Average rental rate based on genre:



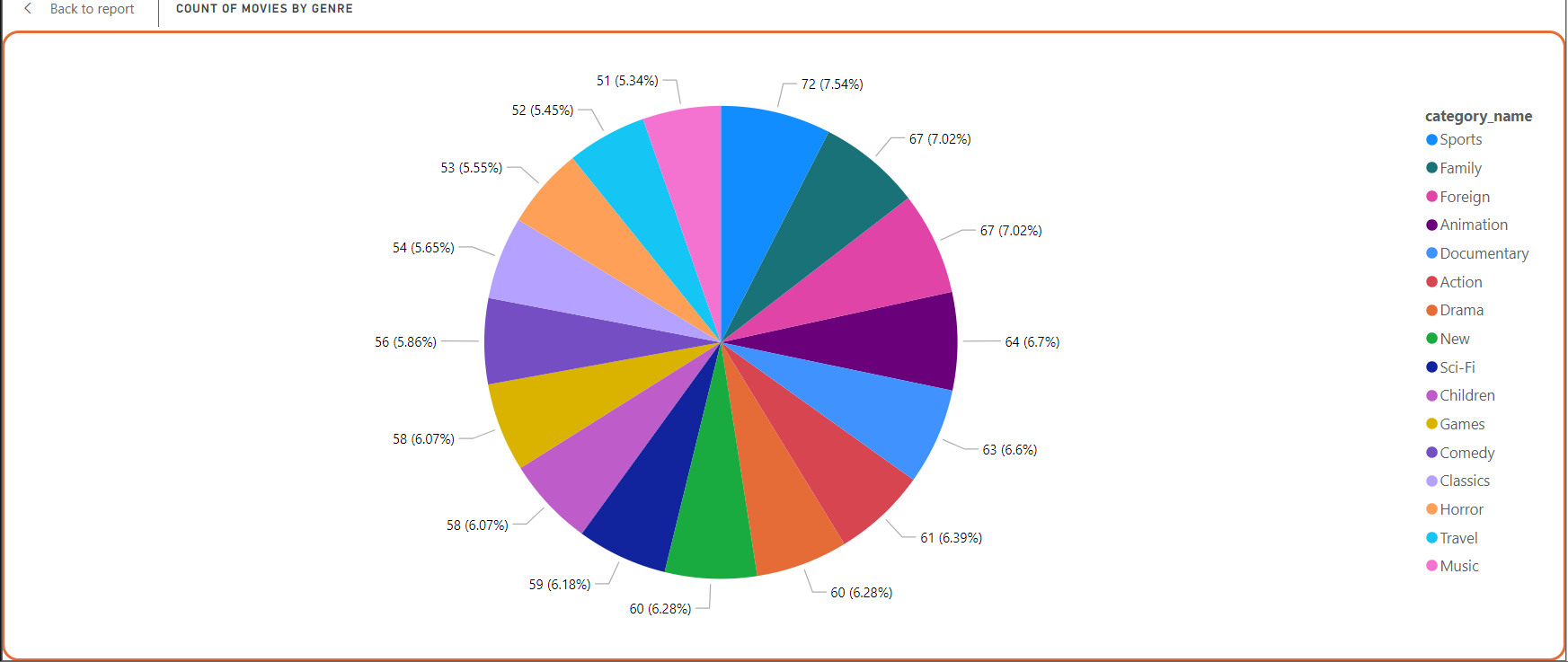
* I have used a treemap to analyze the average rental rate between the genres given in the data set.
* The genre ‘Travel’ has the highest average rental rate of $3.34, followed by ‘Comedy’ with $3.30.
* The genre ‘Documentary’ has the least rental rate of $2.61.

1. Film and Actor Analysis:

A screenshot of a computer

Description automatically generated

* There are 138 actors in total, 955 total films, and 16 film categories.
* Count of movies by genre:

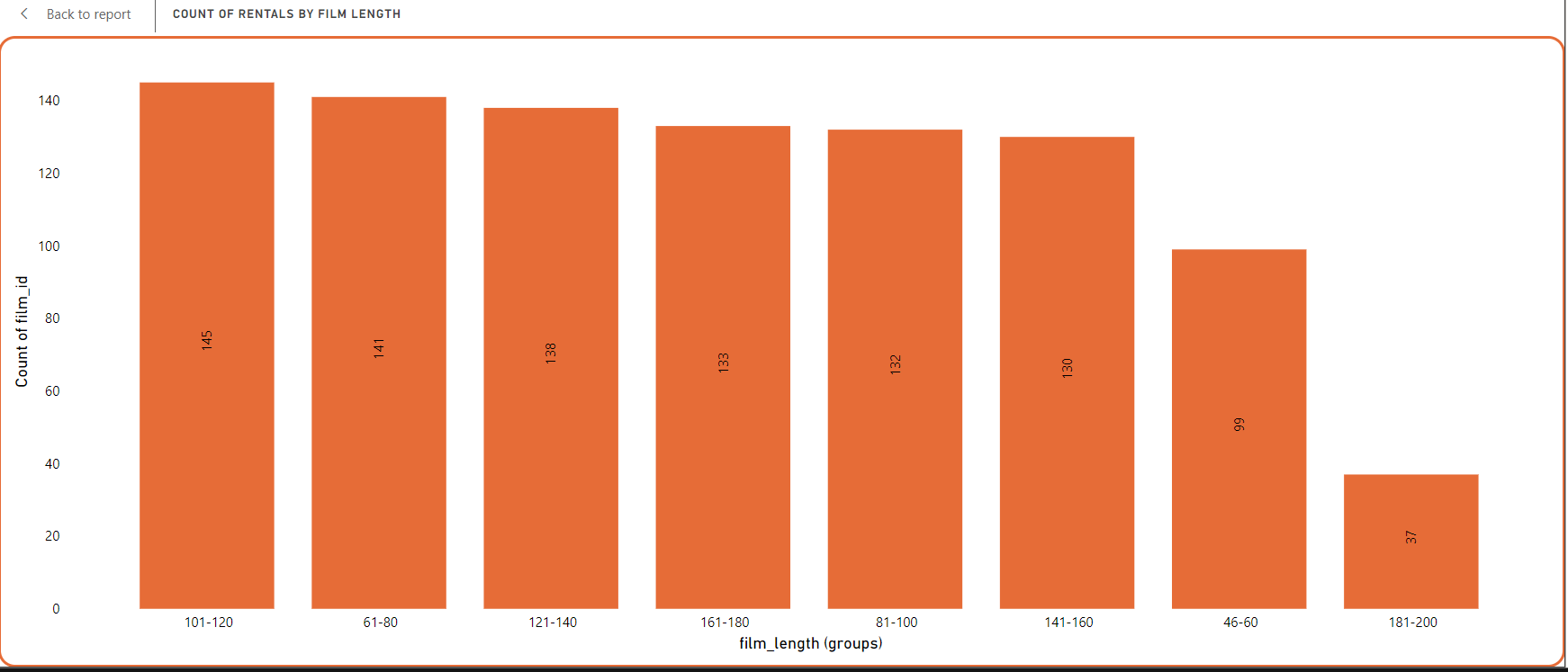


* A pie chart has been used to analyze the number of movies based on genre.
* ‘Sports’ genre has the most number of movies with 72 movies.
* Followed by ‘Family’ and ‘Foreign’. Both combined there are 134 movies.
* Movies based on ‘Music’ have the least number of movies with 51.
* Top 20 actors who contributed the most for the revenue:

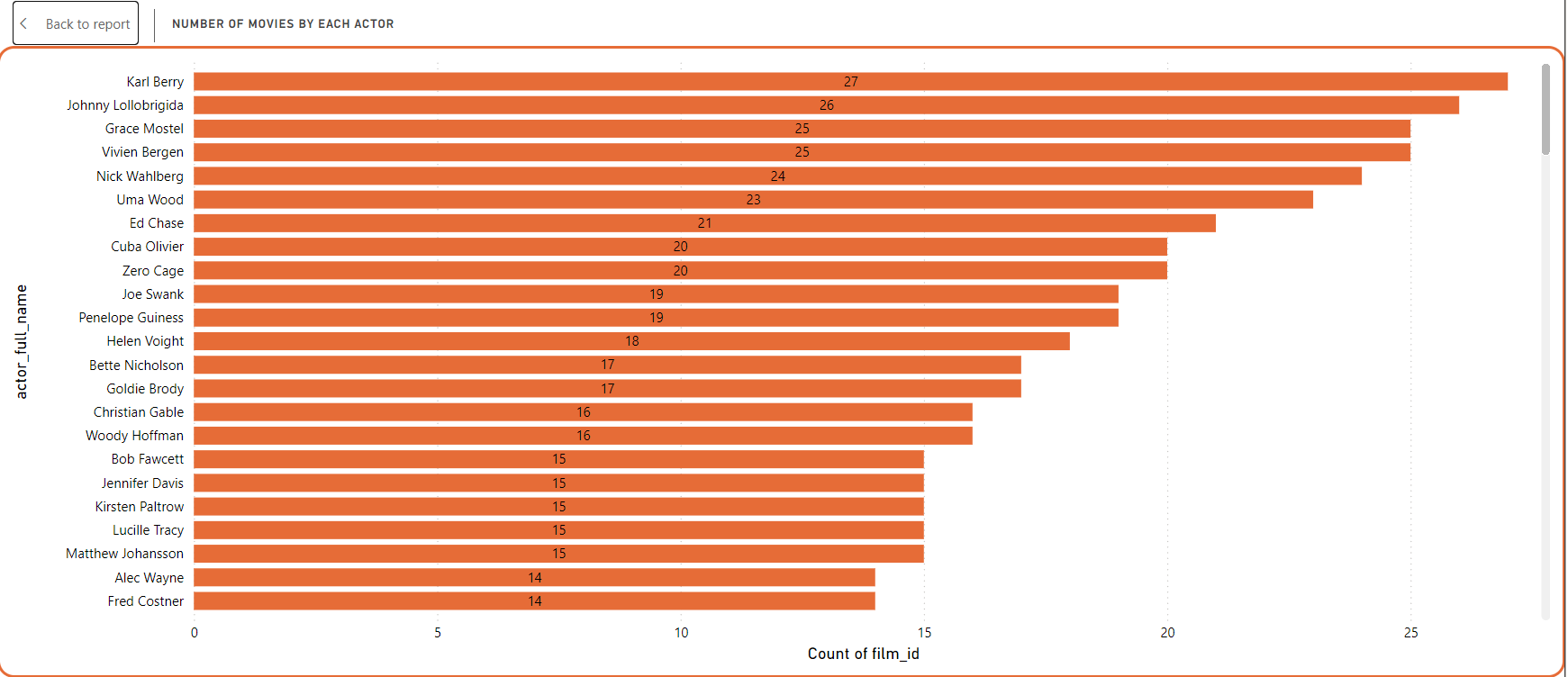
A graph with numbers and letters

Description automatically generated with medium confidence

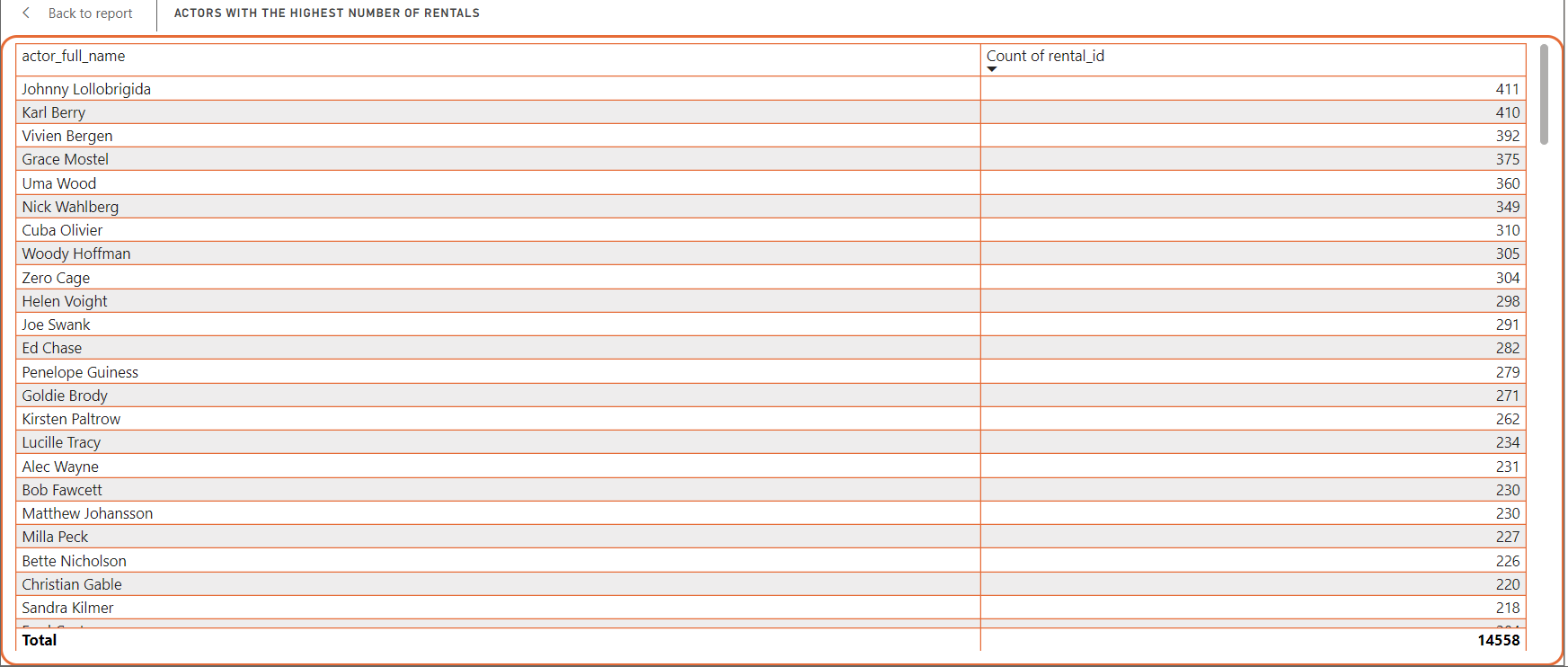
* Used an area chart to know which actor contributed the most for the revenue.
* Here we get to know fan bases among the customers.
* Vivien Bergen’s movies are rented the most by the customers which in return the company earned $1,769.
* I have used a filter to know the top 20 actors.
* Count of rentals based on film length:



* I have used a column chart for this task.
* Created bins for film length with the interval of 20 mins (example: 181-200 mins, 141-160 mins).
* Movie length between 101-120 mins consist of 145 movies.
* Movie length between 181-200 mins consist of 37 movies.
* Number of movies by each actor:



* Used a bar chart to find the count of movies of each actor.
* Karl Berry has acted in 27 movies and stands in 1st place.
* There are only 9 actors who have acted in more than 20 movies.
* There are 74 actors who have the movie count below 5.
* Actors with the Highest Number of Film Rentals:



* I have used the Table chart for this task.
* Actors at the top of this list are likely associated with high-demand films.
* The actor with the highest rentals is Johnny Lollobrigida, with 411 rentals, followed closely by Karl Berry with 410 rentals.
* Other notable actors include Vivien Bergen (392 rentals) and Grace Mostel (375 rentals).
* The total rentals across all actors listed amount to 14,558.

CONCLUSION:

These are the patterns and trends which I found out using the final table. This project evaluated the DVD rental dataset using Power BI to discover insights about revenue, rentals, customer behaviour, film, and actor performance, providing a comprehensive view of business operations and client preferences. The analysis identified chances for increasing revenue by focusing on successful genres and films while customizing marketing approaches for cities that have valuable customers. It emphasized the importance of connecting with regular renters through loyalty programs and enhancing the customer experience in areas that are not performing well. Furthermore, adjusting inventory according to rental trends and local demand can further improve operational effectiveness.

The findings from the analysis provide practical suggestions, including highlighting well-known actors and genres, concentrating on successful cities with tailored marketing initiatives, and refining rental lengths to enhance profits. Implementing these tactics can assist the company in aligning its services more effectively with customer tastes and market needs.