DATA WAREHOUSE REPORT

Radosław Dawidowski; Aleksandra Hryncyszyn; Giacomo Magnocavallo

Introduction

Even though all businesses strive for constant improvement and raise in the revenue, some factors that have significant influence are hard to notice with a bare eye. Especially in large, complex companies, it is of utmost importance to have a clear sight on the goals and objectives that truly drive success. This is where Key Performance Indicators (KPIs) step in as invaluable tools for businesses seeking to navigate their course effectively. They can be separated into four categories. KPIs in the first one, ‘Customer’, measure and analyze customer-related metrics to understand satisfaction, loyalty, and the overall customer experience. These KPIs help businesses attract and retain customers, ultimately driving revenue growth. Second category, ‘Financial’, evaluates the financial health and performance of the business. Financial KPIs provide insights into profitability, revenue generation, and the efficient use of resources. The third category, ‘Internal Business Processes’, focuses on streamlining workflows, optimizing resource allocation, and enhancing overall business productivity. Lastly, ‘Innovation’ KPIs tackle transformations or introductions of new products or processes that might further improve the performance of the company.

To evaluate selected KPIs, data warehouses are and invaluable tool. Contrary to a standard database, they are specifically designed for efficient data analysis and reporting, providing a structured and optimized environment to handle large volumes of historical and real-time data. Those traits are achieved thanks to the star (or snowflake) design that stores all of the foreign keys and numerical data in a fact table, significantly shortening the path of extracting information from multiple tables and simplifying the SQL queries required to achieve that.

The dataset that was chosen for this project is named Sakila. Created by MySQL, it stands as a widely utilized and esteemed sample database within the realm of relational database management systems (RDBMS). Functioning as a simulated model of a DVD rental store, Sakila has been designed to offer a realistic and intricate representation of a rental business, encompassing various entities crucial to such an establishment. As an example, the table 'film' encapsulates details about available films, including title, description, release year, rental rate, and genre. Then, the 'customer' table holds information about the clientele, featuring names, addresses, and contact details. Moreover, 'rental' is a pivotal table that records the transactions involving rented films, linking customers with the films they've rented and capturing pertinent rental dates. The 'inventory' table is integral in tracking the available copies of films for rent, establishing connections with the 'film' table. 'Staff' contains details about the store's personnel, encompassing names, addresses, and other relevant information. Lastly, the 'payment' table meticulously documents the specifics of customer payments for rented films and ‘address’ stores the locations of both customers and staff.

Methodology

The project involved four crucial steps - choosing the dataset, selecting the key performance indicators (KPIs), designing the data warehouse suited to them, and finally the data visualization using Tableau. Firstly, the Sakila database was chosen, then two KPIs were selected per each of the categories – Customer, Financial, Business Processes, and Innovation. Subsequently, the structure of Sakila [Appendix] was meticulously investigated to extract only those entities that are necessary to visualize chosen performance indicators. As the relations are tightly intertwined and the selected KPIs aimed for a deep analysis, the number of tables didn't decrease a lot (from 16 to 14, excluding the facttable). However, the data warehouse design allowed for much simpler and readable relations between them: except for the 'film' and 'address', all of the tables were connected only to the facttable that stored the foreign keys and numerical information like 'amount' and 'rental_duration' [Figure 1]. Firstly, the previously planned design was implemented as an ER diagram in MySQL software, then using the 'Forward Engineer' function it was automatically converted into a data warehouse. Lastly, using the original Sakila database as a staging table, all of the data was extracted to the new data warehouse by means of an SQL query [Appendix]. Such functional data warehouse allowed for much simpler data retrieval SQL queries, further accelerating the extraction of meaningful conclusions.

Subsequently, queries for each of the selected KPIs were designed and executed. The returned data was saved as an .csv file and loaded to the Tableau software. Then, the most informative visualizations were chosen to design a final dashboard. Lastly, conclusions were drawn to make the best decision for the optimal development and revenue increase of the company

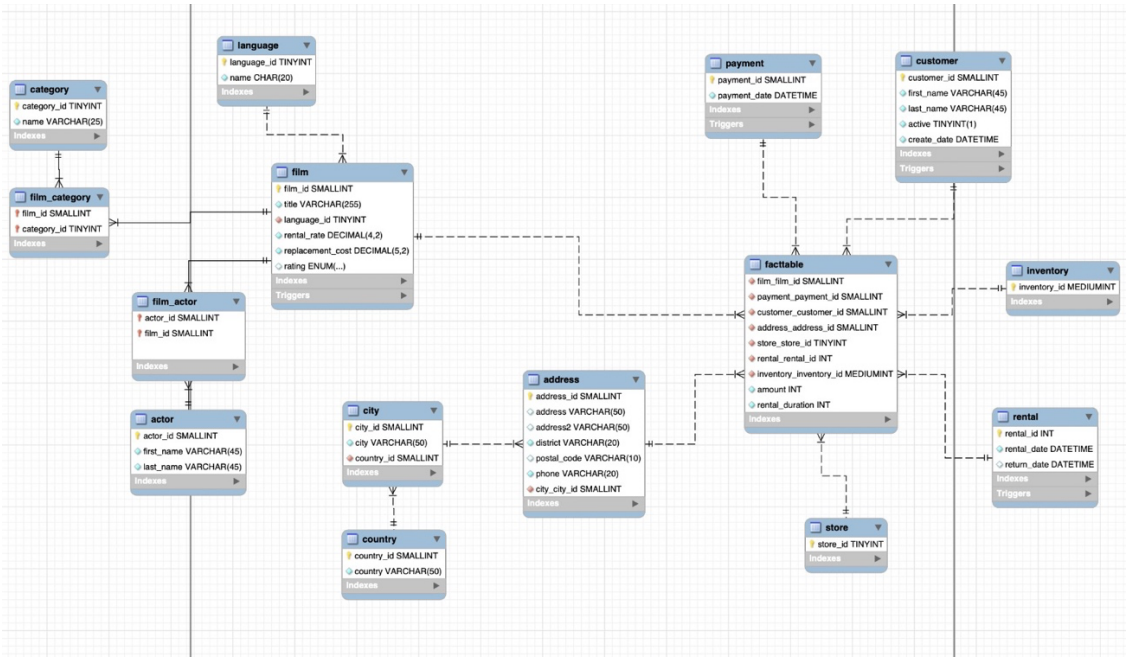
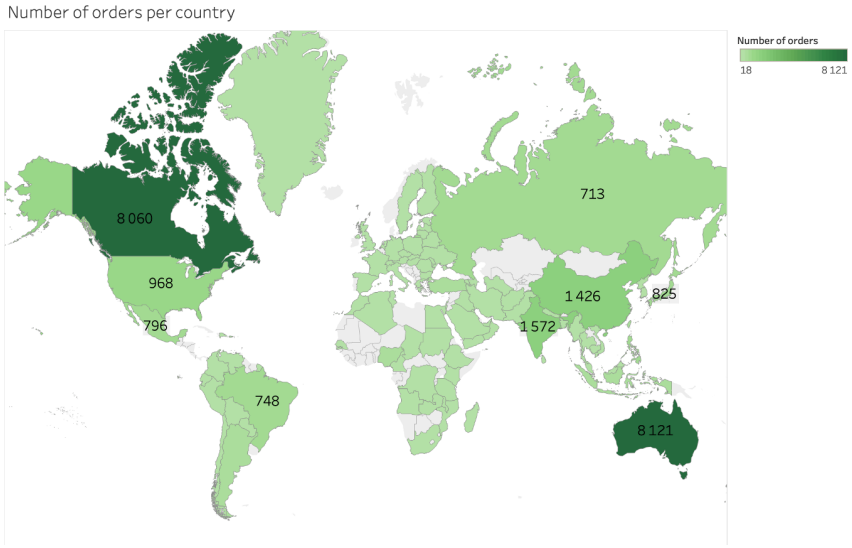


Figure 1 – the final data warehouse design

Results and Discussion

The first Key Performance Indicator revolves around the customer aspect – which new languages might appeal to customers? Recognizing that the company exclusively offers films in English despite having a diverse global customer base, a strategic evaluation was undertaken to enhance the appeal of the offer to non-English speaking countries. To determine the most promising new language to introduce, a map graph [Figure 2] was employed to visually represent the distribution of orders originating from each country. This form was chosen as the most natural to human perception without the requirement of reading the name of the country. The darkest shade of green, representing the highest number of orders, marked two counties where the stores are located – Canada and Australia. However, there were also eight different countries that totaled up to more than the threshold of 700 orders – USA, Mexico, Brazil, Russia, China, Japan, and India. To summarize that information, a table was designed that contains the official language of a country and the number of the orders from it [Figure 1]



language	orders
hindi	1572
mandarin	1426
japanese	825
spanish	796
portugese	748
russian	713

Figure 2 - visualization of orders per each country as well as which languages apart from English are spoken in the countries which the highest number of orders

It can be concluded that the most promising change would be the introduction of films in Hindi, Mandarin, Japanese, or Spanish. It would not only expand the attractiveness of the films offer but also become accessible to much larger group of people.

The second KPI in the Customer category was: are there any customers that can be distinguished for their loyalty and rewarded to promote such behaviors? To assess that, a bar graph was used to plot the number of orders of each customer and sort them in decreasing order [Figure 3].

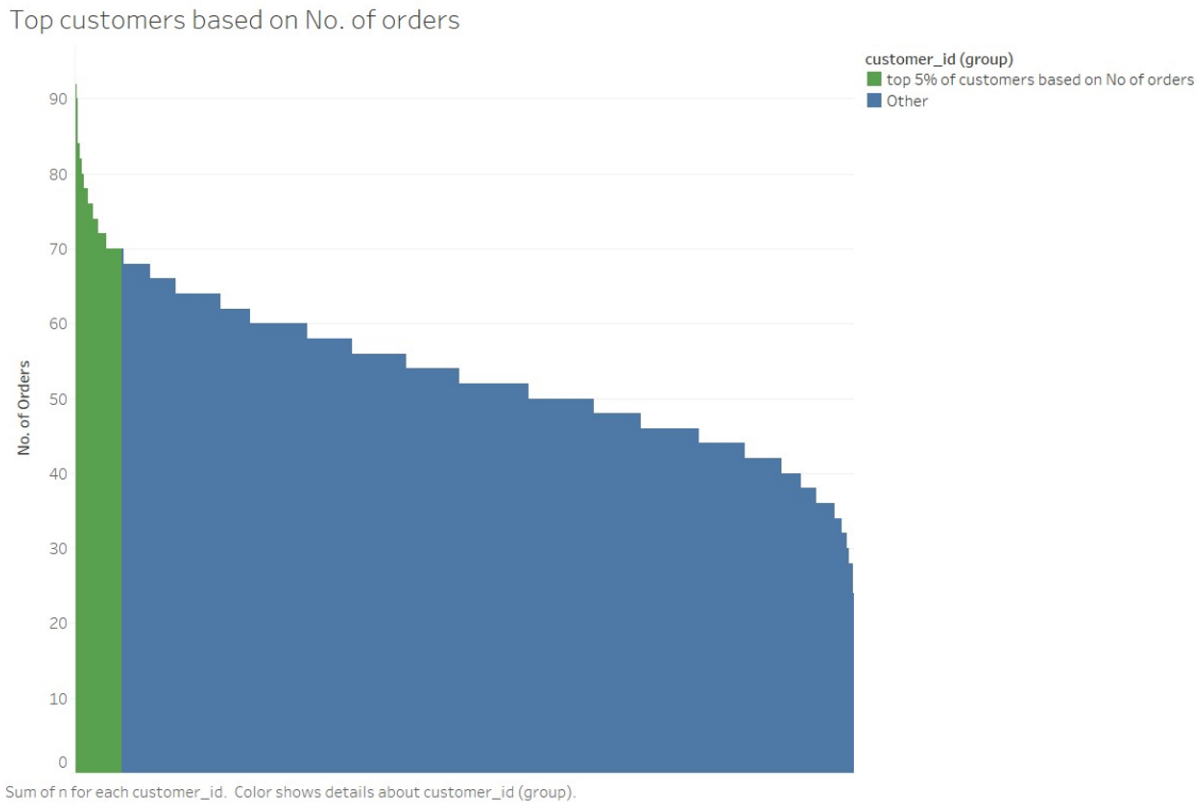


Figure 3 - graph of the number of orders of each customer, sorted in decreasing order. Green color distinguishes the most loyal 5% of the customers that should be rewarded

It was observed that the distribution forms a 'staircase' shape, with many customers having the same number of orders forming one 'step'. However, on the left side on the graph (the highest numbers of orders), the steps become narrower, indicating less people per each number of orders. Thus, approximately 5% of the most loyal customers were chosen and marked in green – they will receive a loyalty card entitling to discounts. Such an incitive might motivate more of the customers to stay loyal.

Another KPI, this time in the 'Finance' category, was: Which rental duration is the most profitable? This insight was not straightforward to visualize. Firstly, the durations of each order were counted and visualized in a bar plot [Figure 4]. It can be observed that the most popular is 7 days rental, followed by 2 days rental. However, there is also a significant number of orders being kept for 8 or even 9 days. Subsequently, the revenue of each order was divided by its rental duration and again visualized as a bar plot per each rental duration category. It can be clearly observed that the daily profit from the 1-day rentals is approximately three times as high as the daily profit from rentals longer than 2 days. This implies that shorter rental periods should be incentivized, or the longest ones prohibited. This approach aims to optimize profit maximization for each rented film, enabling the business to extract the highest possible financial return.

Distribution of rental durations

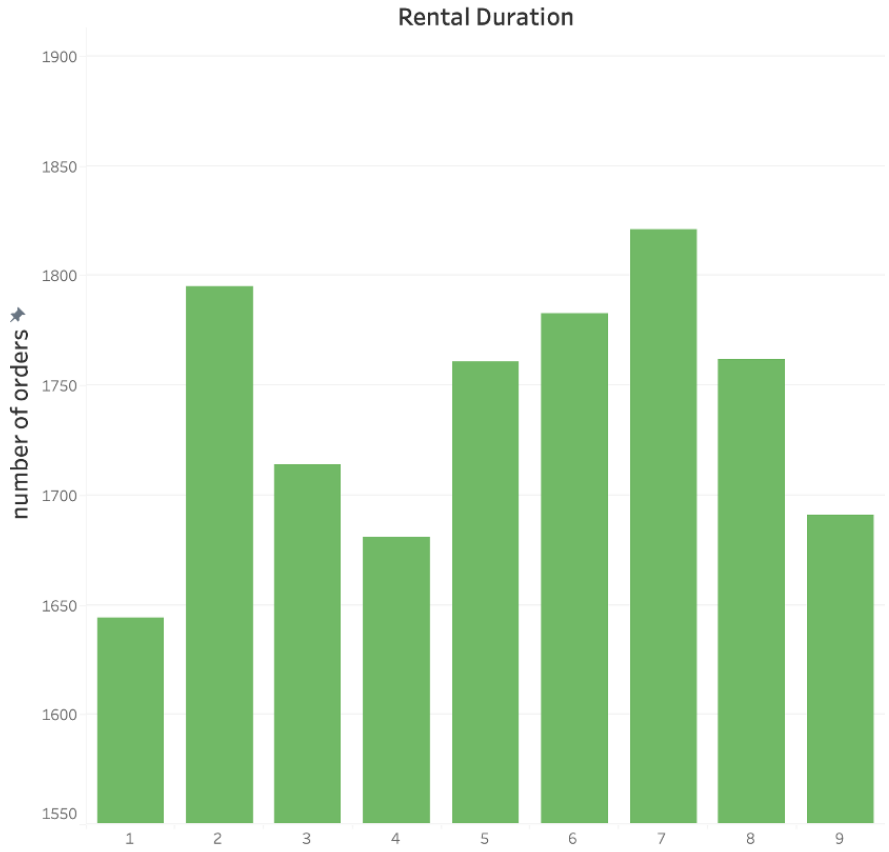


Figure 4 - number of orders per each rental duration

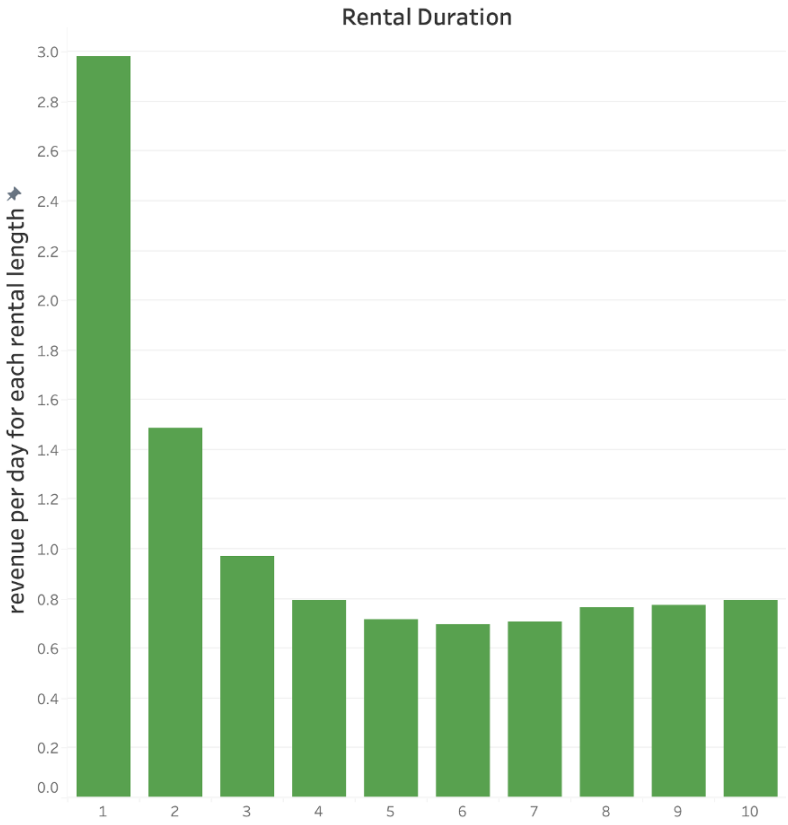


Figure 5 - revenue per day for each rental duration

Another KPI in the same category was: which film categories bring the most revenue? Again, a bar plot of the number of orders in each category, sorted in the decreasing order, was used. Through this visual representation, it becomes evident that the sports, sci-fi, animation, and drama categories emerged as the highest revenue-generating genres. This analysis guides strategic decision-making, suggesting opportunities to expand the film catalogue in directions that are particularly appealing to customers, enhancing overall profitability and customer satisfaction. To delve deeper into this analysis, the visualization not only incorporated revenue data but also introduced the count of films within the inventory,

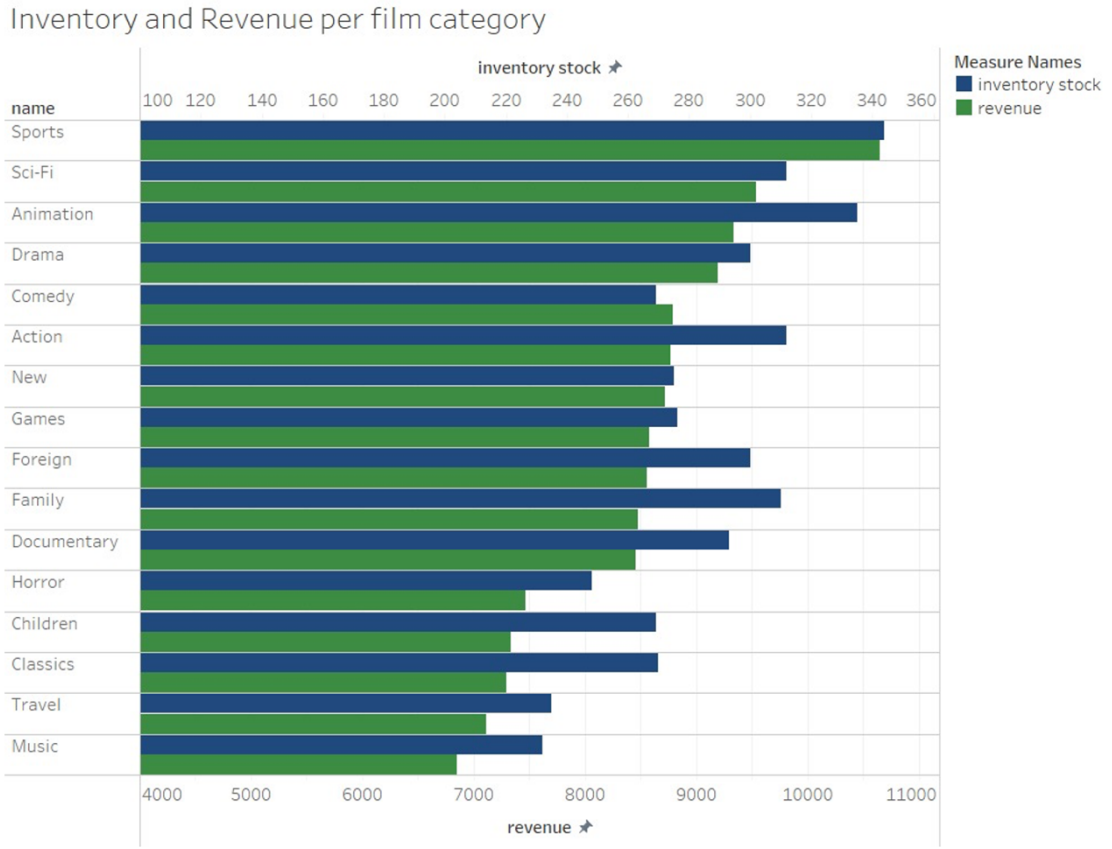


Figure 6 - number of films in inventory as well as revenue they bring per each category

represented in blue [Figure 6]. This comprehensive view reveals that certain categories, such as Child and Classics, have extensive catalog sizes but yield comparatively lower revenue. This insight signals an opportunity for strategic optimization, suggesting a consideration for limiting films in these categories and reallocating resources toward genres that demonstrate higher profitability.

The third KPIs category tackled internal business processes. The first chosen KPI was: How to adjust the inventory to the orders popularity? Again, to properly visualize this insight, firstly a bar plot was selected to show both the number of orders of each film (green) as well as their copies in stock (blue) [Figure 7]. While sorted by the decreasing number of orders, it can be clearly noticed how there are a lot of single blue stripes – those are films that are overstocked.

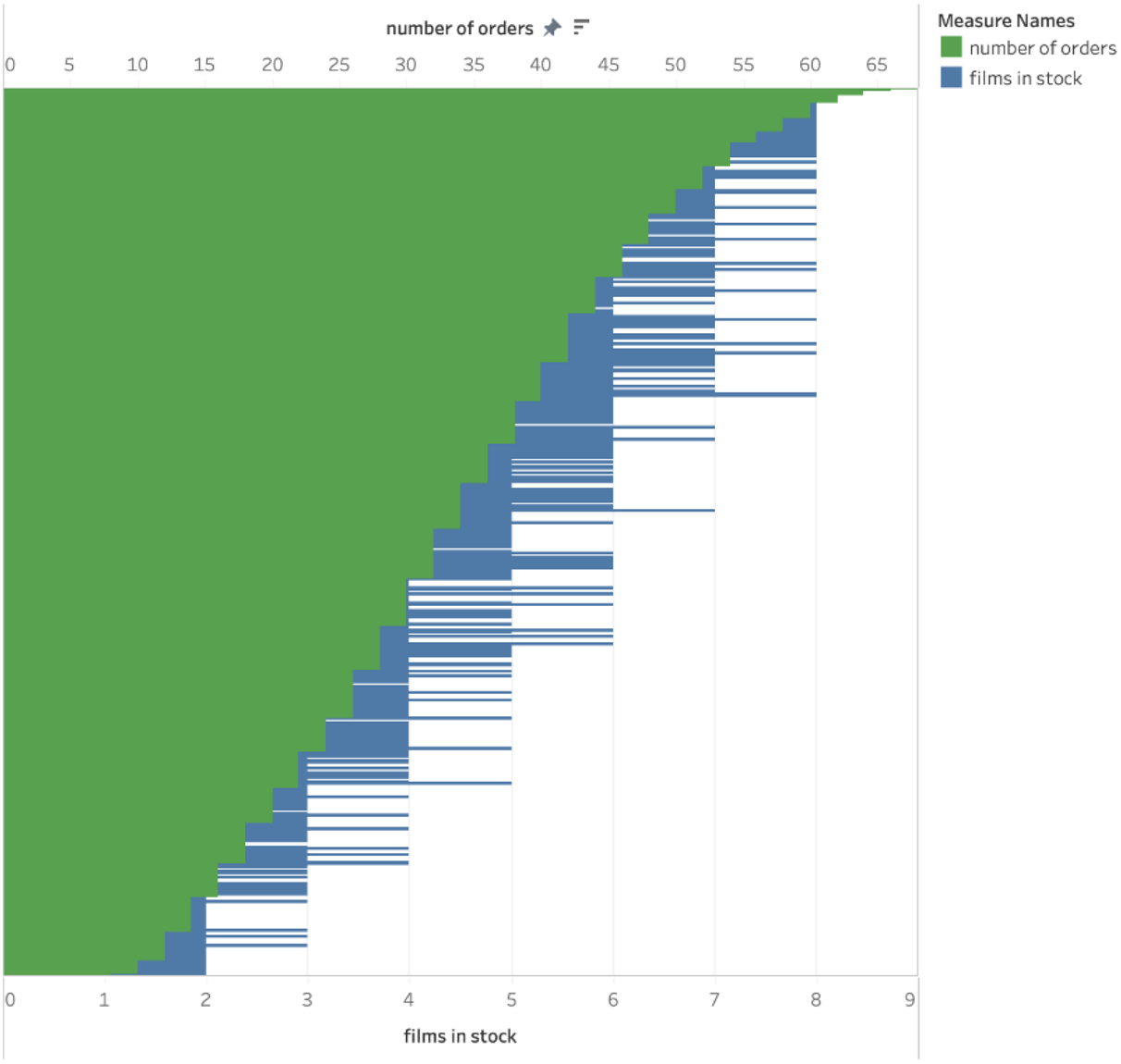


Figure 7 - number of copies as well as number of order of each film

Then, the ratio of those two values was calculated for each film. To visualize how often each value occurs, an area plot was employed [Figure 8]. There are three most intense spikes – at values 6, 7, and 8 orders per film copy. However, there are also a lot of examples above (understocked films), and below (overstocked), in total ranging from 4 orders/copy to as much as 10 orders/copy. Addressing this imbalance holds the potential to optimize inventory management, ensuring a more balanced distribution of film copies and, consequently, enhancing overall operational efficiency.

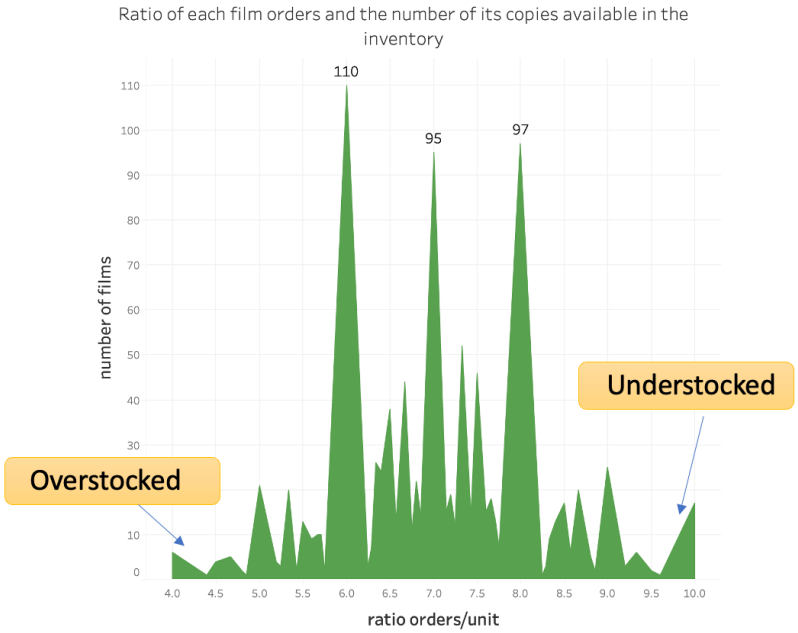


Figure 8 - number of films with certain values of ratio film orders per copy in the inventory

Another KPI within this category was: Which days are the most important for the business? Rental operations, thus both the renting as well as handing in the film were counted for each day of the week. Surprisingly, the findings revealed strikingly similar values across all days, implying that each day of the week holds equal importance for the business. This insight underscores the continuous and consistent demand for rental services, emphasizing the need for operational readiness and resource allocation throughout the entire week to meet customer needs effectively.

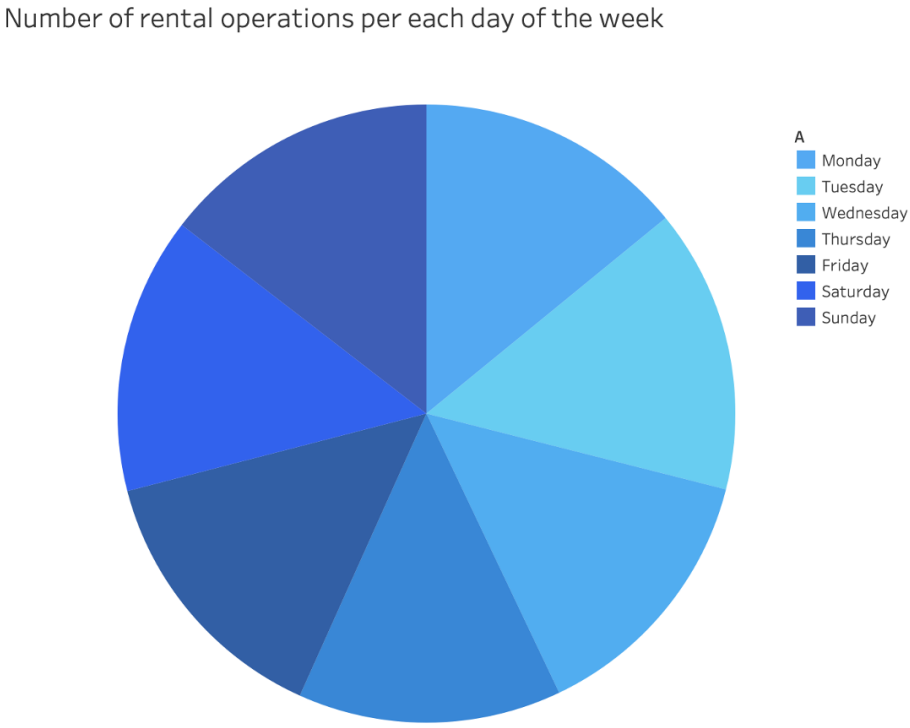


Figure 9 - number of rental operations per day of the week

The last KPIs category was Innovation. Within this category, the first selected KPI was: Where should another shop be opened? To investigate that, again a map was used, with the revenue value reflected in the shade of green. Again, in the locations of currently operating shops, Canada and Australia, the revenue was the highest. However, also India and China were observed to generate approximately $6000 in annual revenue despite extended shipping times due to the absence of physical stores in these regions. Thus, it can be concluded that those locations of significant market interest would be a strategic location for a new shop. Even though the implementation would require meticulous planning and innovative market-entry strategies to mitigate shipping constraints, this expansion aligns with the growth objectives and presents a promising opportunity for substantial revenue increase.

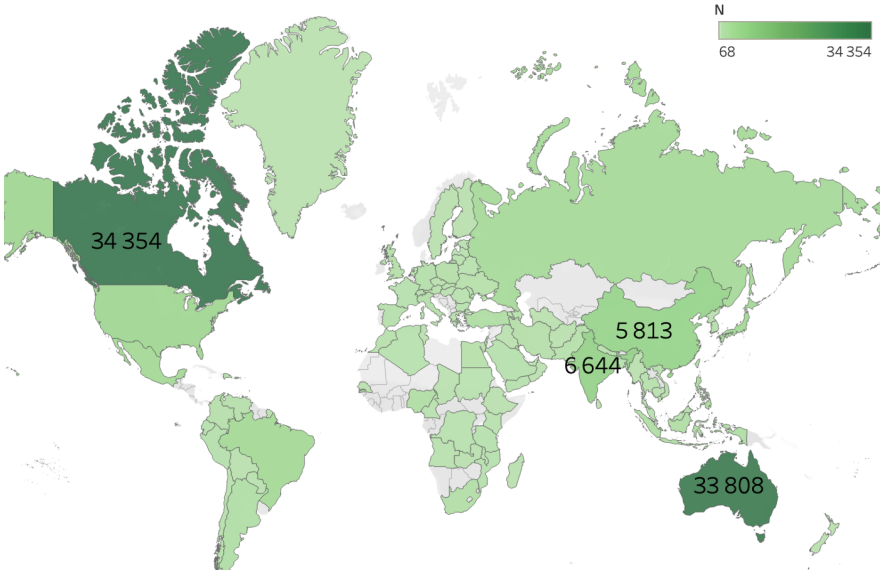


Figure 10 - revenue per country

Last KPI within this category was: which actors result in the most revenue? While variations in revenue among actors exist, they generally aren't significant. However, one standout is Gina Degeneres; movies featuring her contributions accumulate nearly $7000 in total revenue. This noteworthy difference prompts consideration for increasing the film stock featuring her, suggesting a potentially lucrative strategy for increased profitability.

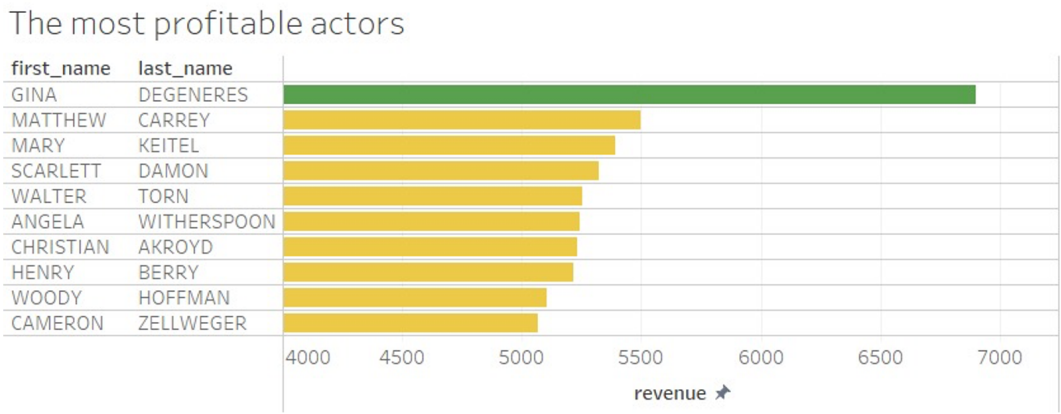


Figure 11 - revenue per actor

Conclusions

The Sakila dataset analysis allowed to extract insights essential for strategic decision-making. Notably, diversifying language offerings, particularly in Hindi, Mandarin, Japanese, and Spanish, stands as a promising tactic to broaden the film catalog and tap into new markets. From a customer perspective, targeting the top 5% of loyal customers lays the groundwork for a lucrative loyalty rewards program, fostering loyalty and potentially drawing in new customers through incentivized referrals. Financially, profitability appears linked to rental durations, highlighting the need to encourage shorter rentals, particularly 1-day durations, for heightened daily profits. Moreover, genre-based revenue disparities suggest the potential for optimizing inventory toward lucrative genres like sports, sci-fi, animation, and drama, potentially boosting overall profitability. In terms of innovation, expanding into India and China markets presents a promising growth opportunity, supported by existing revenue streams and market interest despite shipping challenges. Furthermore, leveraging Gina Degeneres' performance in films could be a strategic move to drive increased profitability based on her notably higher contribution to total revenue.

In conclusion, the analysis underscores clear directives for market expansion, customer loyalty reinforcement, profit optimization through rental strategies, and inventory focus refinement. These insights offer a roadmap for strategic decision-making aimed at revenue growth and operational efficiency enhancement within the company.

Appendix

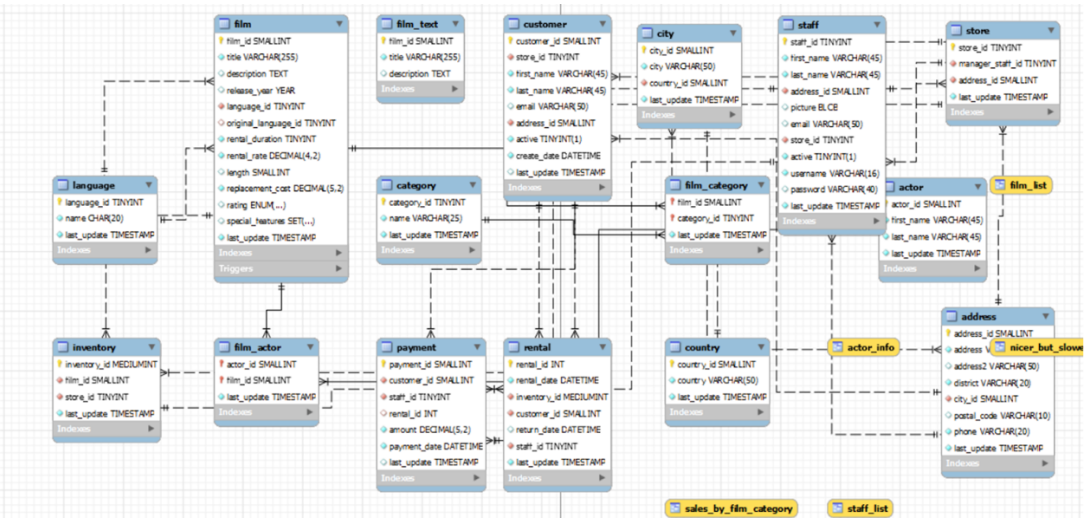


Figure 12 - original Sakila database design

film_film_...	payment_payment...	customer_customer...	address_address...	store_store_...	rental_rental...	inventory_inventory...	amount	rental_durati...
663	1	1	2	2	76	3021	3	7
875	2	1	2	2	573	4020	1	6
228	4	1	2	2	1422	1021	1	4
316	10	1	2	2	4526	1443	6	5
814	12	1	2	2	5244	3726	5	6
243	16	1	2	2	7841	1092	5	3
341	18	1	2	2	8074	1558	1	5
3	22	1	2	2	10437	14	5	7
709	23	1	2	2	11299	3232	4	7
315	24	1	2	2	11367	1440	1	4
204	26	1	2	2	12250	921	1	4
317	30	1	2	2	14825	1449	2	4
663	1	1	5	2	76	3021	3	7
875	2	1	5	2	573	4020	1	6

*Figure 13 - result from a select * from facttable query*