ROCKBUSTER PERFORMANCE ANALYSIS

A relational database was analyzed utilizing **PostgreSQL**, **Excel** and **Tableau** to assess quarterly performance and produce a targeted marketing strategy.

This case study details the analytical process undertaken and the skills developed throughout. For immediate access to the findings and queries, please click the links below:

PRESENTATION | GITHUB | DATA DICTIONARY

BACKGROUND

Rockbuster is a global movie rental company that previously had physical locations in the face of streaming services like Netflix and Amazon Prime. Rockbuster has launched an online video rental service to stay competitive.

NEED

A performance analysis and data-driven insights for their preceding three months to inform their future company strategy.

QUESTIONS ADDRESSED:

Which countries are customers based in?

Do sales figures vary between geographic regions?

Which movies contribute the most and least to revenue gain?

What is the average rental duration for all videos?

Where are the customers with the highest life-time values located?

PROCESS IN A NUTSHELL:

PostGreSQL — A relational database was imported into PostGreSQL. Desired data were queried and exported as .csv files.

Excel — Exported .csv files were imported into Excel for further integration and aggregation.

Tableau — Excel spreadsheets were imported into Tableau to create visualizations.

Adobe Illustrator — Visualizations were imported into Adobe Illustrator to create the presentation.

DATA PREPARATION:

To understand the fabric of the database and the relationships of the tables, an Entity Relationship Diagram was created using Dbvisualizer (click image). Fact tables were boxed in green to distinguish them from the dimension tables. The contents of each table were organized in a data dictionary, with primary and foreign keys specified (see data dictionary). Erroneous or less-than-ideal data types were altered and standardized in PostGreSQL: date quarry and consistency were checked for duplicated data and misspellings of nominal values of each variable were checked by using COUNT(*), GROUP BY, ORDER BY and HAVING COUNT(*) > 1.

DATA EXPLORATION:

Queries were crafted to extract necessary data to address the stated questions/needs. Complex queries were broken down into smaller sub-queries, solving problems like customer, country, or film-specific examples below.

To address questions beyond the scope of the project, subqueries and Common Table Expressions (CTEs) were written. These queries, along with the queries written for this project, can be found on Github.

In addition to the designated questions, an analysis comparing the revenues, customer volume, and transaction counts of the two sister locations were performed but ultimately, there were no conclusive findings.

The queried data were exported from PostGreSQL as .csv and imported into Excel where some variables were integrated and further aggregated. Between revenue, customer volume and transaction counts, a strong correlation of 0.998 was identified. Thus revenue was utilized as the primary determinant of the company's performance for the remainder of the analysis.

```
-- Top 5 customers in the top 10 cities who paid the highest total amount.

SELECT
    b.customer_id,
    b.first_name,
    b.last_name,
    SUM(b.amount) AS total_amount,
    d.city,
    e.country
FROM payment a
    JOIN customer b ON b.customer_id = a.customer_id
    JOIN address c ON b.address_id = c.address_id
    JOIN city d ON c.city_id = d.city_id
    JOIN country e ON d.country_id = e.country_id
WHERE city IN (
    'Aurora',
    'Atlixco',
    'Bhopal',
    'Brest',
    'Cabanatuan',
    'Cayenne',
    'Ciudad Lerdo',
    'Ebolowa',
    'Fengyuan',
    'Garland'
)
GROUP BY
    b.city,
    b.customer_id,
    b.first_name,
    b.last_name
ORDER BY total_amount desc
LIMIT 5
```

CHALLENGES:

Initially, I viewed the targeted nature of this project's analysis as restrictive, focusing on exercising queries to produce outputs without engaging in critical analysis/thinking. However, I found this assumption to be incorrect. Having a comprehensive understanding of the company's needs, goals, existing information and decision-making requirements, I couldn't assess whether a more extensive analysis would yield new or useful insights. This realization not only prevented unnecessary analyses and improved time efficiency, but also opened the door to proposing additional analyses in the future, should they be proven applicable.

During my explanation of analyses beyond the project's scope, I observed that addressing these questions required a different database design. Recognizing these limitations, a redesigned structure can be considered to expand future analyses.

PRESENTATION HIGHLIGHTS:

DATA DESCRIPTION

Between Feb – May: 561,312 gross sales (in US Dollars)

Spanning 14,596 transactions (no NULL/Nullvalues)

Across 584 customers

Across 109 countries

Spanning 1,000 films, 17 genres, 5 rental durations, 3 rental rates, 1 language

Through 2 store locations

Descriptive analysis paints an overall picture of the scope and performance of Rockbuster

GLOBAL | MEASUREMENTS

Revenue, customer volume and transaction count showed a strong correlation with a coefficient of 0.998 but of 1.00. This means that countries with high customer volume and low high transaction count are diamonds.

India China and the United States make up 25% of the total revenue.

	Revenue	Customer	Transaction	Title
China	$36,389.26	302	2,811	25%
India	$35,267.18	297	2,697	24%
United States	$34,776.43	591	5,421	23%
Japan	$12,834.94	119	1,232	9%
Mexico	$3,924.00	30	283	3%
Brazil	$3,419.45	28	254	2%
Russian Federation	$3,178.48	28	232	2%
Philippines	$3,086.64	30	273	2%
Turkey	$2,636.65	15	150	2%
Indonesia	$2,601.39	14	128	2%

GENRE RANKING BASED ON REVENUE

	Revenue	Rental Count	Average Rental Rate	Average Replacement Cost	Average Rental Duration
Action	$4,375.85	864	2.82	20.29	4.77
Animation	$4,351.62	816	2.92	20.67	5
Children	$4,245.66	945	2.82	20.54	4.87
Classics	$4,086.39	939	2.78	20.75	4.84
Comedy	$4,384.85	941	2.94	20.42	4.65
Documentary	$4,272.66	951	2.92	20.43	4.75
Drama	$4,781.52	862	2.92	20.74	4.71
Family	$4,232.79	1,096	2.75	20.32	4.94
Foreign	$4,270.67	893	2.92	20.31	4.69
Games	$4,562.30	969	3.12	20.29	4.65
Music	$3,071.52	830	2.92	20.46	4.83
New	$4,351.62	940	2.92	20.99	4.77
Sci-Fi	$4,756.98	961	2.92	21.10	4.75
Sports	$4,892.19	1,179	2.82	20.77	4.92
Travel	$3,632.05	757	2.92	20.87	4.62

Horror generated the lowest revenue at $3,722.54

Thriller ranked at the bottom, making up 1.5% of the revenue, at the rather old but then-hot-ticket rental duration of 6 days

- Mixture of bar charts, tables and colored buckets accommodates different ways people take in information. Bar charts provide a quick picture of how each category compares to one another
- Tables provide deeper insights with specific values (cold used to emphasize the highs and lows
- Color buckets highlighting pertinent takeaway message on each slide

Mixture of bar charts, tables and colored buckets accommodates different ways people take in information. Bar charts provide a quick picture of how each category compares to one another. Tables provide deeper insights with specific values (cold used to emphasize the highs and lows. Color buckets highlighting pertinent takeaway message on each slide

REGIONAL | MEASUREMENTS

Same as the global measurements, there is a strong correlation between revenue, customer volume and transaction count within the six regions.

Below are the revenue, customer volume, transaction count and revenue rate of each region.

Asia made up nearly 50% of the revenue and nearly 3X the revenue of the runner-up, Europe.

	Revenue	Customer	Transaction	Title
Asia	$61,438	510	4,753	49%
Europe	$21,610	179	1,679	17%
North America	$21,227	176	1,648	17%
South America	$7,141	59	554	6%
Oceania	$6,966	58	540	6%
Canada	$6,837	57	531	5%

CASE/END statement used to conditionally assign region names to preceeding country names. This provided free insights into regional performance, underscoring Asia as the top performing region with nearly 50% higher revenue compared to the runner-up region, Europe.

GLOBAL | REVENUE x TRANSACTIONS

RECOMMENDATIONS

Scatterplot allowed the countries to naturally fall into four distinct performance clusters. With the addition of average lines for both revenue and number of transactions, and based on the current positions and growth potential, each cluster received distinct marketing recommendations.

CONCLUSION:

This project served as a valuable illustration of what I might encounter with ad-hoc assignments. This emphasized the importance of performing a specific, client-focused analysis without assuming that a fully comprehensive analysis is always necessary. This ran into amplified client satisfaction but also optimized time and resource allocation.

I enjoyed the seamless use of multiple tools - PostGreSQL, Excel, Tableau and Illustrator - throughout the analysis and in compiling the final report. Illustrator allowed for complete customization of the visuals, and enhance presentation. I enjoyed the combination of the bar chart with the table to highlight insights in different graphics, as well as the the opportunity to provide specific recommendations for each category.

SKILLS ACQUIRED:

SQL

- Understanding of database schemas: Entity Relationship Diagram, schema types, primary/foreign keys
- Create, Read, Update, Delete
- syntax and formatting best practices
- indexing
  - JOINs
  - subqueries
- CASE/END conditional logic
- Common Table Expressions
- descriptive analysis
- aggregation
- Extraction, Transformation and Loading

Data Cleanin | Data Analyst | Information Designer | cecynthiamillo@gmail.com