ANSWERING BUSINESS QUESTIONS FOR AN ONLINE VIDEO RENTAL COMPANY

A DATA ANALYSIS CASE STUDY

Dena Coduri

THE PROJECT

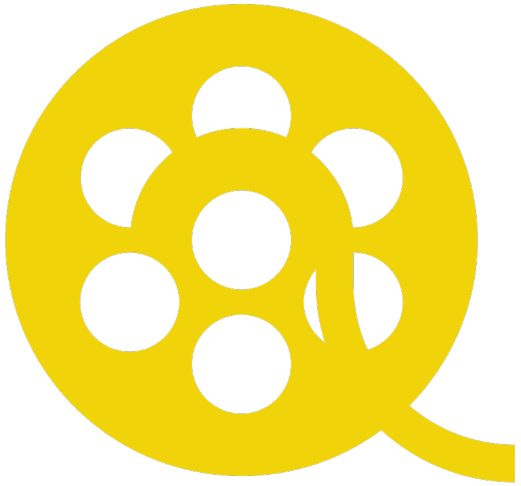
Company

Rockbuster Stealth LLC is a fictional movie rental company that used to have stores around the world. The Rockbuster Stealth management team is planning to use its existing movie licenses to launch an online video rental service in order to compete with Netflix and Amazon Prime.

Objective

Help the business intelligence department with the launch strategy for Rockbuster's new online streaming service and answer any ad-hoc questions from other departments along the way.

THE QUESTIONS



The Rockbuster Stealth Management Board has a series of business questions, and they expect data-driven answers that they can use for their strategy. Here are the main questions they would like to answer:

- Which movies contributed the most/least to revenue gain?
- What was the average rental duration for all videos?
- Which countries are Rockbuster customers based in?
- Do sales figures vary between geographic regions?

They have also asked for a data dictionary to be created for their SQL database.

THE PROCESS

The data used for this analysis was provided by Career Foundry.

Software used – SQL, Excel, Google Slides, Tableau

Minimal cleaning was required for this dataset. Consistency issues were fixed.

Steps

- Fix consistency issues in the data
- Explore the data using SQL queries and answer ad-hoc business questions
- Prepare an Excel file with important SQL queries/outputs for technical colleagues
- Create an interactive map in Tableau
- Create a data dictionary in Excel
- Create a presentation in Google Slides that answers the questions from the business

THE DELIVERABLE

The process for creating the 14-page data dictionary was straight-forward. I inspected every single table in the Rockbuster database, described every row, and showed how other tables were linked to and from the table at hand.

Dimension Table: city

Lists the name of the city associated with the city id number.

Columns

Key	Columns	Data Type	Description	To
🔑	city_id	SERIAL	Unique identification number for the city.	
	city	CHARACTER VARYING(50)	Name of the city.	
	country_id	SMALLINT	Unique identification number for the country.	↗
	last_update	TIMESTAMP(6) WITHOUT TIME ZONE	Date and time the record was last updated.	

Links to

Table	Join
country	city.country_id = country.country_id

Links from

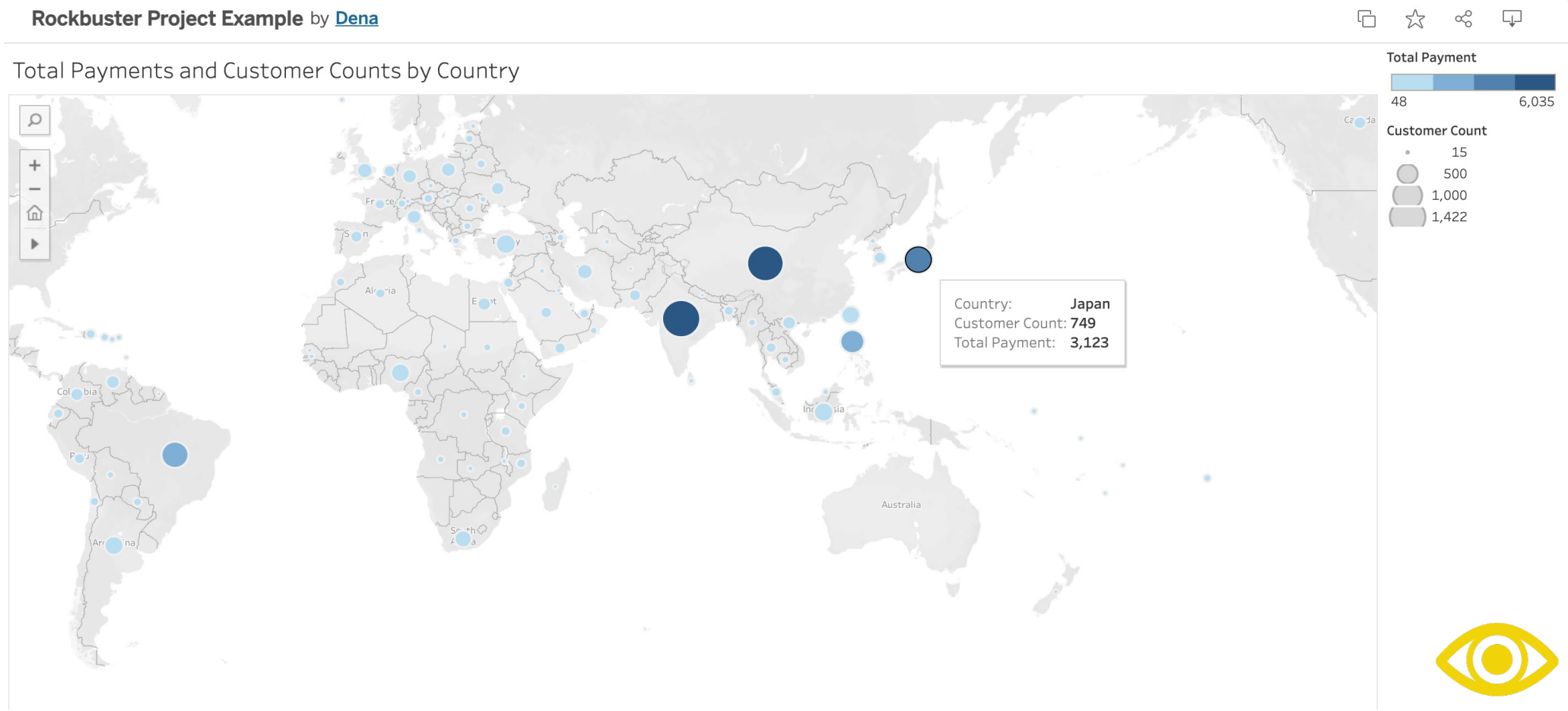
Table	Join
address	city.city_id = address.city_id

Unique Keys

city_id

country_id

THE VISUALS

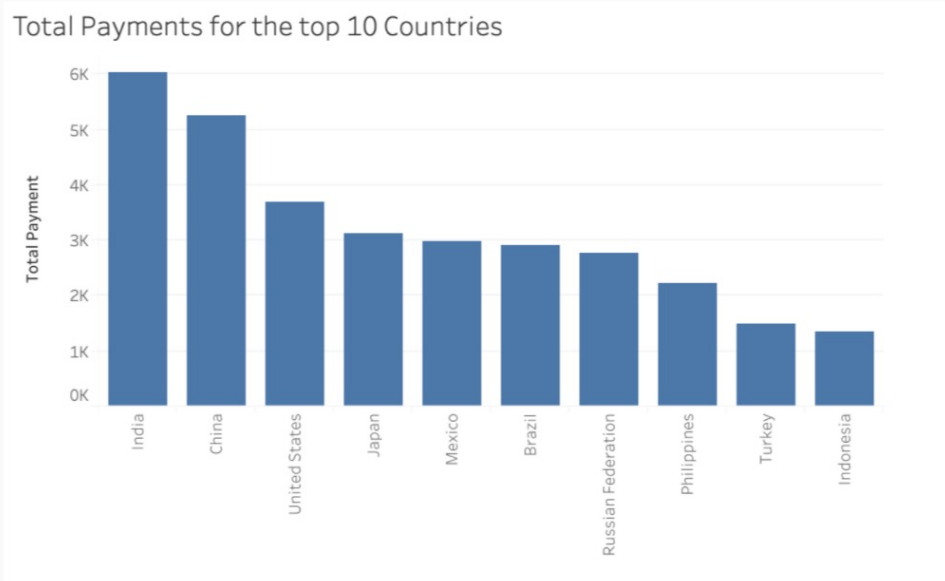


An interactive map was created in Tableau so the business intelligence department could easily see where most of the customers are located and the total payments for each country.

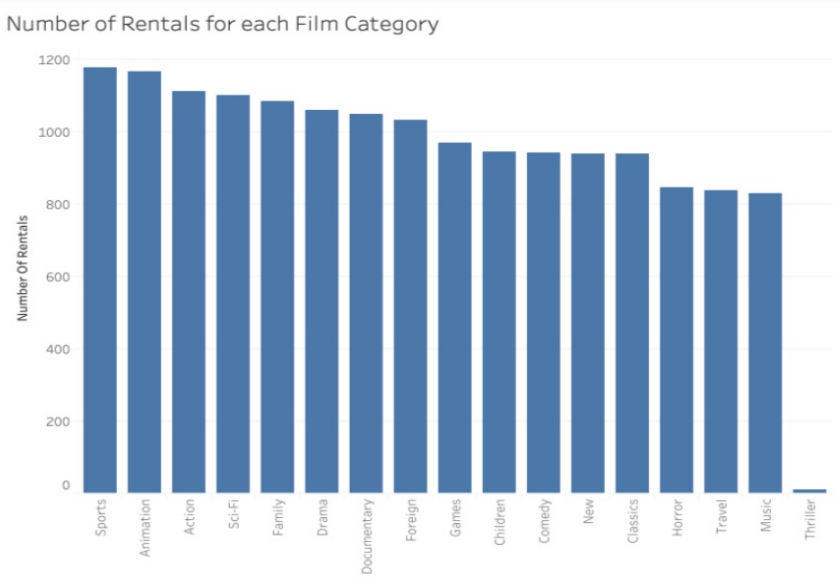
Link to interactive Tableau map

THE INSIGHTS

Which countries are Rockbuster customers based in?



Which film categories are the most popular?



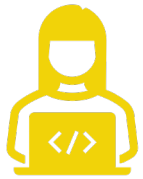
Bar charts were created in Tableau to show the top countries and film categories.

- Rockbuster's new online streaming service should launch first in the top two countries: India and China
- The most popular film categories should be released first: Sports and Animation

THE FEEDBACK

My original SQL queries for the top 10 cities and top 5 customers were hard-coded, meaning I had included specific country/city names in my queries.

Although the queries were correct, my mentor challenged me to rewrite them so that as the top countries/cities changed over time, the queries would still be valid.



```sql
SELECT C.city, D.country,
       COUNT(customer_id) AS customer_count
FROM customer A
     INNER JOIN address B ON A.address_id = B.address_id
     INNER JOIN city C ON B.city_id = C.city_id
     INNER JOIN country D ON C.country_ID = D.country_ID

WHERE country IN
          (SELECT D.country
          FROM customer A
               INNER JOIN address B ON A.address_id = B.address_id
               INNER JOIN city C ON B.city_id = C.city_id
               INNER JOIN country D ON C.country_ID = D.country_ID
          GROUP BY country
          ORDER BY COUNT(customer_id) DESC
          LIMIT 10)

GROUP BY city, country
ORDER BY customer_count DESC, city ASC
LIMIT 10
```

Link to GitHub Repository with SQL queries and outputs

THE RETROSPECTIVE

- I plan to avoid hardcoding my SQL queries in the future, so that they remain useful as the data changes. Reusing old queries makes more sense than rewriting them every time.
- Although the assignment was to find the top 5 cities within the top 10 countries, and the top 5 customers within those top 5 cities, I would like to add to this project by looking up the top 5 cites and customers overall.