ROCKBUSTER MOVIE RENTAL ANALYSIS PROJECT

DATA DICTIONARY

PREPARED BY SRILATHA KUMMARI

Table of Contents

1. Project Overview 2
2. Entity Relationship Diagram: 3
3. Fact Tables 4
   a. Rental 4
   b. Payment 5
4. Dimension Tables 5
   a. Film 5
   b. inventory 6

c. Language 6
d. Film Actor 7
e. Actor 7
f. Film Category 7
g. Category 8
h. Customer 8
i. Staff 9
j. Store 9
k. Address 9
l. City 10
m. Country 10

1. Project Overview

Rockbuster is a movie rental company that have stores around the world. They are facing competition from streaming services such as Netflix and Amazon Prime. So, the Rockbuster Stealth management team plans to use its existing movie licences to launch an online video rental service to stay competitive.

Objective:

- To identify the most loyal customers, whre they are located & what films shouls be in the inventory

Tools:

- SQL
- Excel
- Tableau
- Powerpoint

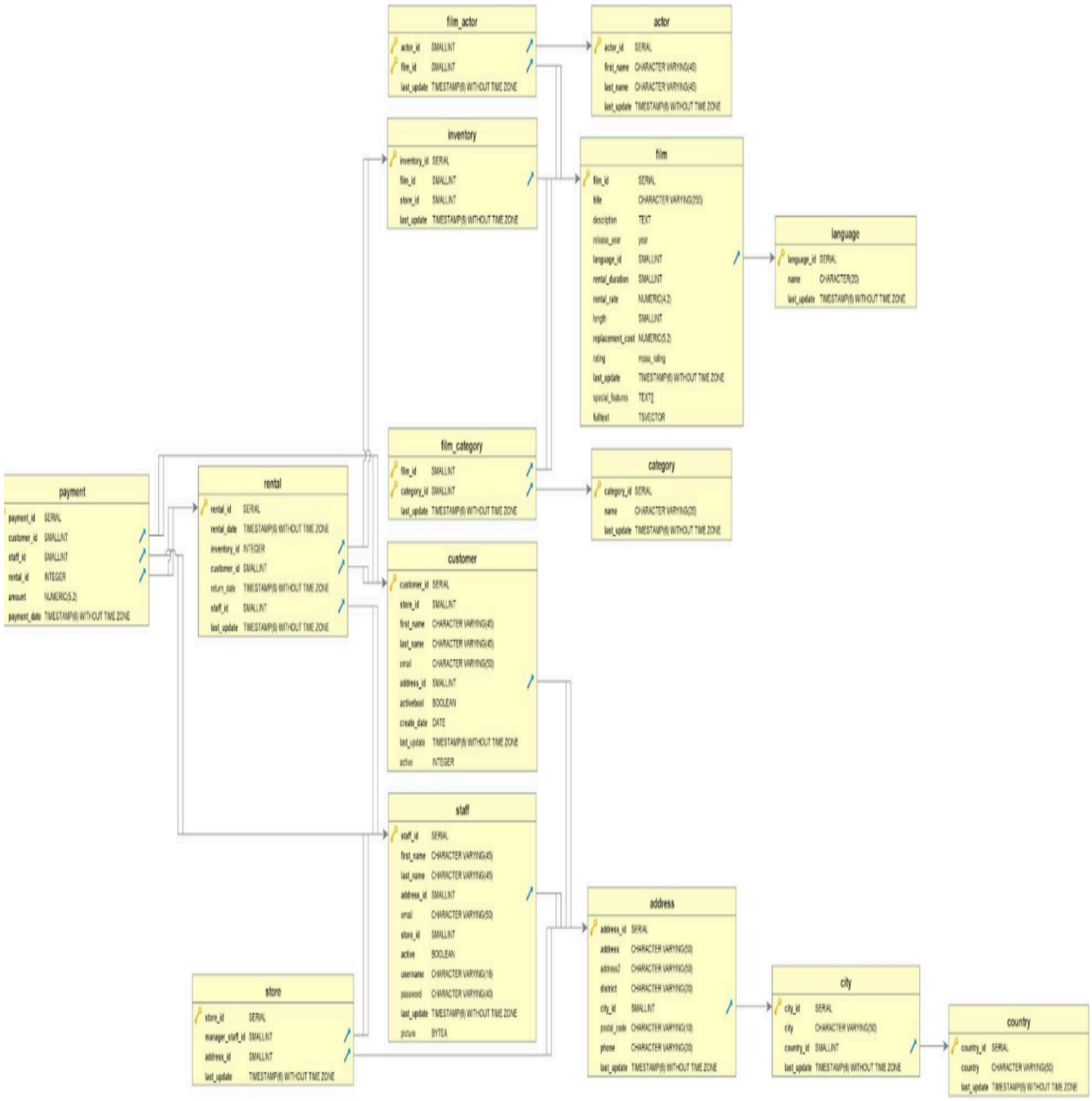
Data:

The data contained the information about inventory, customers, country, payment, city etc.

Skills:

- Descriptive Analysis
- Data Cleaning
- Data Grouping
- SQL Joins
- Common Table Expressions (CTEs)
- Data Visualiation
- Storry Telling

2 Entity Relationship Diagram(ERD):



3. Fact Tables

a. Rental

Key	Columns	Data Type	Description	Links

	Rental_id	SERIAL	Primary key, integer, Unique value to identify each rental transaction.	 Payment
	Rental_date	TIMESTAMP(6) WITHOUT TIME ZONE	Temporal Data, Used to identify date and the time of the rent started.	
	Inventory_id	INTEGER	Foreign key, larger integer, unique value to identify inventory.	 Inventory
	Customer_id	SMALLIANT	Foreign key, small integer, unique value identify each customer.	 Customer
	Return_date	TIMESTAMP(6) WITHOUT TIME ZONE	Temporal Data, used to identify the date and time of the item retured.	
	Staff_id	SMALLIANT	Foreign key, small integer, unique value to identify personnel member.	 Staff
	Last_update	TIMESTAMP(6) WITHOUT TIME ZONE	Temporal data, used to identify the last changes and updates made to the record.	

b. Payment

Key	Columns	Data Type	Description	Links
	Payment_id	SERIAL	Primary key, integer, Unique value to identify payment.	
	Customer_id	SMALLINT	Foreign key, small integer, Unique value to identify customer.	 Customer
	Staff_id	SMALLINT	Foreign key, small integer, unique value to identify staff.	 Staff
	Rental_id	INTEGER	Foreign key, larger integer, unique value to identify the rental transaction.	 Rental
	Amount	NUMERIC(5,2)	Monetary amount, numeric value with 5 digits and with 2 digits after the decimal.	
	Payment_date	TIMESTAMP(6) WITHOUT TIME ZONE	Temporal data, used to identify the payment date and time.	

4. Dimension Tables

a. Film

Key	Columns	Data Type	Description	Links

	Film_id	SERIAL	Primary key, unique number to identify each film.	⟸ film_actor film_category invenrory
	Title	CHARACTER VARYING(255)	Film title, fixed length character with max 255 characters.	
	description	TEXT	Film synopsis with unlimited length.	
	Release_year	YEAR	Integer, the year in which the film was released.	
	Language_id	SMALLIANT	Small integer, Unique number to identify the language of the film.	⟹ language
	Rental_duration	SMALLIANT	Small integer, the number of days the item is rented for.	
	Rental_date	NUMERIC(4,2)	Rating, number with maxx 4 digits and 2 digits after decimal.	
	length	SMALLINT	Small integer, number to show the length of the film(minutes)	
	Replacement_cost	NUMERIC(5,2)	The costs that the customer has to pay in case of loss or damage caused by him with max 5 digits, including 2 digits after decimal.	
	rating	Mpaa_rating	Rating to the film.	
	Last_update	TIMESTAMP(6) WITHOUT TIME ZONE	Temporal data, store the data of the last update.	
	Special_features	TEXT[[	Extra videos related to the film.	
	fultext	TSVECTOR	Text of the movie	

b. inventory

Key	Columns	Data Type	Description	Links
	Inventory_id	SERIAL	Primary key, larger integer, unique value to identify inventory.	⟸ Rental
	Film_id	SMALLINT	Foreign key, unique number used to identify the film.	⟹ Film Store
	Store_id	SMALLINT	Small integer, unique value to identify the store.	
	Last_update	TIMESTAMP(6) WITHOUT TIME ZONE	Temporal data, used to identify the last changes and updates made to the record.	

c. Language

	Columns	Data Type	Description	Links

	Language_id	SERIAL	Primary key, Unique value to identify language	⟸ Film
	name	CHARACTER(20)	The name of the language of the film audio.	
	Last_update	TIMESTAMP(6) WITHOUT TIME ZONE	Temporal data, store the data of the last update.	

d. Film Actor

Key	Columns	Data Type	Description	Links
	Actor_id	SMALLINT	Unique number to identify actor.	⟹ Actor
	Film_id	SMALLINT	Unique number to identify the film.	⟹ Film
	Last_update	TIMESTAMP(6) WITHOUT TIME ZONE	Temporal data, store the data of the last update.	

e. Actor

Key	Columns	Data Type	Description	Links
	Actor_id	SERIAL	Primary key, unique number to identify the actor.	⟸ film_actor
	First_name	CHARACTER VARYING(45)	Actor's first name, max 45 characters.	
	Last_name	CHARACTER VARYING(45)	Actor's last name, max 45 characters.	
	Last_update	TIMESTAMP(6) WITHOUT TIME ZONE	Temporal data, store the data of the last update	

f. Film Category

Key	Columns	Data Type	Description	Links
	Film_id	SMALLINT	Unique number to identify the film.	⟹ Film

	Category_id	SMALLINT	Unique number to identify film category.	 Category
	Last_update	TIMESTAMP(6) WITHOUT TIME ZONE	Temporal data, store the data of the last update	

g. Category

Key	Columns	Data Type	Description	Links
	Category_id	SERIAL	Primary key, unique number to identify film category.	 film_categorry
	name	CHARACTER VARYING(25)	The name of the film category, ma 25 characters.	
	Last_update	TIMESTAMP(6) WITHOUT TIME ZONE	Temporal data, store the data of the last update	

h. Customer

Key	Columns	Data Type	Description	Links
	Customer_id	SERIAL	Primary key, unique number to identify the customer.	 Rental payment
	Store_id	SMALLINT	Small integer, unique number to identify the store.	
	First_name	CHARACTER VARYING(45)	Customer's first name, ma 45 characters.	
	Last_name	CHARACTER VARYING(45)	Customer's last name, ma 45 characters.	
	email	CHARACTER VARYING(50)	Cuatomer's email address, ma 50 characters.	
	Address_id	SMALLINT	Customer's address	 Address
	activebool	BOOLEAN	True or false statement to indicate customer's account status.Boolean data holds 3 possible values (true, false or null)	
	Create_date	DATE	The date when the customer's account was created.	
	Last_update	TIMESTAMP(6) WITHOUT TIME ZONE	Temporal data, store the data of the last update	
	active	INTEGER	Integer, a number to indicate if the customer's account is active.	

i. Staff

Key	Columns	Data Type	Description	Links
	Staff_id	SERIAL	Primary key, unique number to identify the personnel.	 Rental Payment Store
	First_name	CHARACTER VARYING(45)	Customer's first name, ma 45 characters.	
	Last_name	CHARACTER VARYING(45)	Customer's last name, ma 45 characters.	
	Address_id	SMALLINT	Employee's address	 Address
	email	CHARACTER VARYING(50)	Employee's email, ma 50 characters	
	Store_id	SMALLINT	Unique number used to identify the store	
	active	BOOLEAN	True/False statement to indicate the employee's statement.	
	username	CHARACTER VARYING(16)	Employee's username, max 16 characters.	
	password	CHARACTER VARYING(40)	Employee's password for internal account, max 40 characters.	
	Last_update	TIMESTAMP(6) WITHOUT TIME ZONE	Temporal data, store the data of the last update	
	Picture	BYTEA	Employee's picture. Bytea is used to store raw binary data like images.	

j. Store

Key	Columns	Data Type	Description	Links
	Store_id	SERIAL	Primary key, unique number to identify the store.	 Inventory Staff
	Manager_staff_id	SMALLINT	Unique number to identify the manager staff.	 Staff
	Address_id	SMALLINT	Unique value to identify the store address.	address
	Last_update	TIMESTAMP(6) WITHOUT TIME ZONE	Temporal data, store the data of the last update	

k. Address

Key	Columns	Data Type	Description	Links

	Address_id	SERIAL	Primary key, unique value to identify address.	⟸ Cusromer Store Staff
	address	CHARACTER VARYING(50)	Any address-line 1, City and street name.	
	Address 2	CHARACTER VARYING(50)	Address in the database line-2 such as house number, apartment number.	
	district	CHARACTER VARYING(20)	Name of the district, Max 20 characters.	
	City_id	SMALLINT	Unique value to identify the city.	⟹ City
	Postal_code	CHARACTER VARYING(10)	Postal code, Max 10 characters.	
	Phone	CHARACTER VARYING(20)	Phone number related to address, Max 20 characters.	
	Last_update	TIMESTAMP(6) WITHOUT TIME ZONE	Temporal data, store the data of the last update	

l. City

Key	Columns	Data Type	Description	Links
	City_id	SERIAL	Primary key, unique number to identify the city.	⟸ Address
	city	CHARACTER VARYING(50)	City name in the address, Max 50 characters.	
	Country_id	SMALLINT	Unique number to identify country.	⟹ Country
	Last_update	TIMESTAMP(6) WITHOUT TIME ZONE	Temporal data, store the data of the last update	

m. Country

Key	Columns	Data Type	Description	Links
	Country_id	SERIAL	Primary key, unique number used to identify country.	⟸ City
	Country	CHARACTER VARYING(50)	Country name in the address.	
	Last_update	TIMESTAMP(6) WITHOUT TIME ZONE	Temporal data, store the data of the last update	

Primary Key

Foreign Key

Links to

Links From