Achievement 3 Project Brief: Rockbuster Stealth Data Analysis Project

CF

Introduction

Rockbuster Stealth LLC is a movie rental company that used to have stores around the world. Facing stiff competition from streaming services such as Netflix and Amazon Prime, the Rockbuster Stealth management team is planning to use its existing movie licenses to launch an online video rental service in order to stay competitive.

You've been hired as a data analyst by Rockbuster Stealth's business intelligence (BI) department to help with the launch strategy for the new online video service. The BI department helps other departments, from inventory to customer insights, with data-related queries. Your first task is to load all of Rockbuster's data into a relational database management system (RDBMS). Then, you'll use SQL to analyze the data and answer any ad-hoc business questions that other departments may have.

Before you can begin your analysis and answer more complex business questions, you'll need to acquire a good understanding of the various data points. You'll eventually compile the results of your analysis into an easily digestible format, which will be presented to the Rockbuster Stealth Management Board.

Key Questions and Objectives

The Rockbuster Stealth Management Board has asked a series of business questions and they expect data-driven answers that they can use for their 2020 company strategy. Here are the main questions they'd 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?
- Where are customers with a high lifetime value based?
- Do sales figures vary between geographic regions?

CF

Context

To answer the questions posed by the different departments, you'll ==query the data using SQL== (short for structured query language). Throughout this Achievement, you'll learn SQL incrementally, and you'll use it to answer increasingly complex business questions for each task. The results of your analysis will be presented to Rockbuster management, so you'll need to ==visualize the data== in an easy-to-consume manner. In other words, you'll build on what you've learned in previous Achievements.

Although you'll be working with data from the online video streaming industry for this project, the skills you develop could be applied to any industry and will highlight your ability to solve problems for a variety of clients. The aim is to give you hands-on experience with SQL. By the time you finish your project, you'll be able to use SQL to perform various kinds of query-based analyses.

Data Set

In this Achievement, you'll be using a data set that contains information about Rockbuster's film ==inventory, customers==, and ==payments==, among other things. The first thing you'll need to do is load the data set into the PostgreSQL database. Keep in mind the following points regarding the data set:

- It's around 3MB and contains several files.
- A relationship exists between two tables if a column name is present in both tables.

To install PostgreSQL and load the Rockbuster data set into it, follow the instructions in Exercise 3.1: Introduction to Relational Databases carefully.

- [Download the Rockbuster data set](#)
- [Download PostgreSQL Database](#) (for Windows, macOS X, and Linux)

Project Deliverables

At the end of each Exercise, you'll complete a task and submit a deliverable that directly contributes to your final project—in this case, an analysis that answers business questions put forward by the management team.

Your tutor will review and give you feedback on your deliverables for tasks 3.1 to 3.9. Your final project, which you'll submit at the end of Exercise 10, will be reviewed by your mentor, who will also give you feedback on how your project relates to the work of a data analyst.

Final Analysis Criteria

Your final project will be evaluated on your ability to:

- Write moderately complex SQL queries to answer business questions.
- Present your SQL results to business managers by creating visualizations and telling a compelling story.
- Present your SQL results to your technical colleagues using Excel and by creating a data dictionary.
- Create a professional project that you can add to your portfolio and show to employers.

Learning Goals

Below is a breakdown of the learning goals for each Exercise in Achievement 3. They state what you'll achieve after you've read the Exercise and completed the task.

Exercise 1: Introduction to Relational Databases

- Explain the difference between relational databases and non-relational databases
- Explain the importance of online analytical processing databases (OLAP) and relational database management systems (RDBMS)

- Set up a database environment using the PostgreSQL client GUI tool to begin an analysis

Exercise 2: Data Storage & Structure

- Explain data storage structures, including keys and indexes
- Identify common data types in relational databases
- Explain the different types of relational database schemas and their components
- Extract an entity relationship diagram and create a first draft of a data dictionary

Exercise 3: SQL for Data Analysts

- Identify the syntax of common SQL commands, including basic CRUD operations
- Explain commonly used constraints for creating tables
- Write SQL commands in PostgreSQL to answer some basic business questions

Exercise 4: Database Querying in SQL

- Explain the importance of querying in SQL
- Identify SQL syntax for ordering, limiting, and grouping data
- Explain the basics of database migration and ETL
- Write SQL queries to organize and sort data

Exercise 5: Filtering Data

- Identify commonly used operators for filtering data
- Differentiate between the WHERE and HAVING clauses
- Explain CASE statements in SQL
- Filter and order data using the WHERE and HAVING clauses

Exercise 6: Summarizing & Cleaning Data in SQL

- Identify dirty data and describe ways of cleaning it
- Explain the DSP model and the use of the VIEW statement
- Create a data profile of summary statistics using SQL

Exercise 7: Joining Tables of Data

- Identify join types in SQL
- Explain how multiple joins are used
- Use SQL to join tables

Exercise 8: Performing Subqueries

- Explain the order of SQL query writing
- Define subqueries and their various applications in SELECT commands
- Write subqueries to answer complex business questions

Exercise 9: Common Table Expressions

- Explain how CTEs are used
- Differentiate between a subquery and a CTE
- Rewrite subqueries as CTEs

Exercise 10: Presenting SQL Results

- Create visualizations of SQL results
- Create a presentation of findings

CF