

Use-Case Data Analyst Intermediate

Para resolver este use-case deves suportar-te através das seguintes ferramentas que podes encontrar na pasta deste enunciado. Estas são:

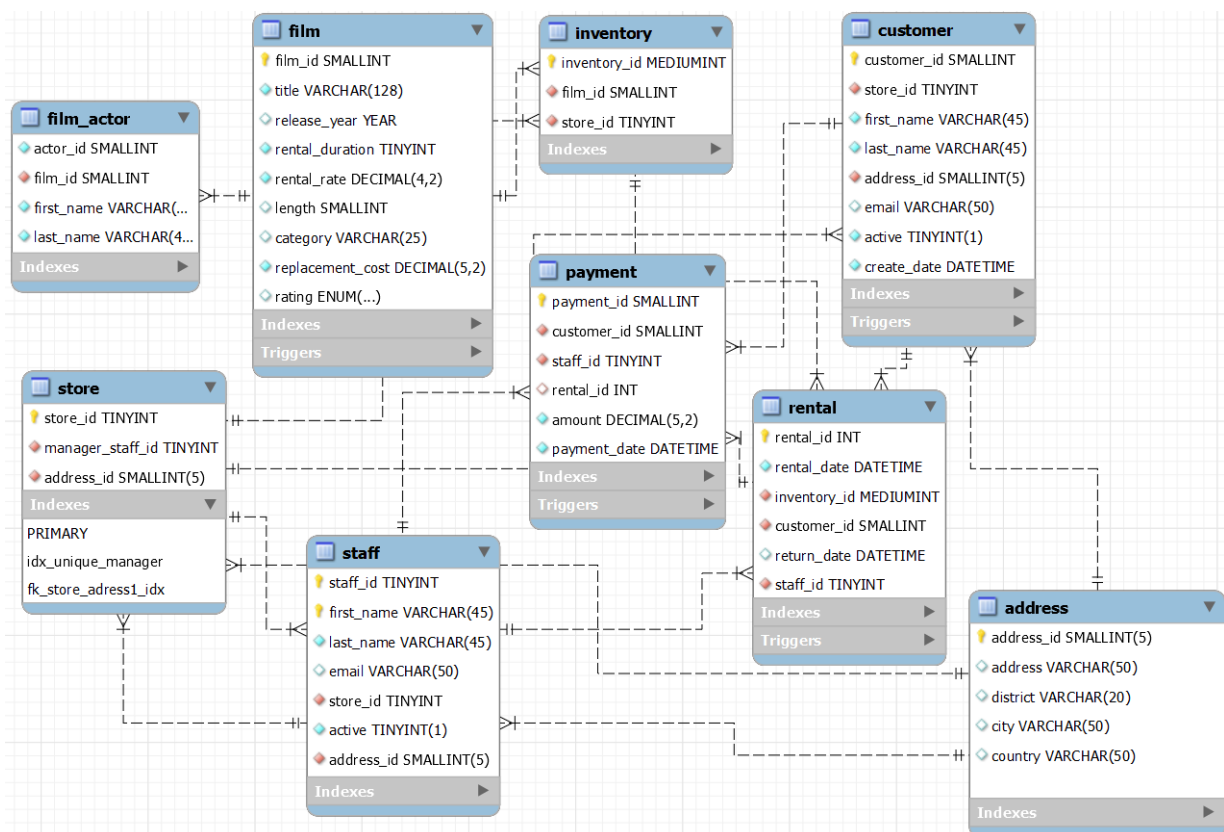
- Guia de instalação de software necessário.
- Ficheiro .SQL para criação da base de dados do use-case
- Ficheiros .CSV com as tabelas e respetivos dados do use-case
- Excel com a descrição da base de dados usada no use-case.

Para responder às questões presentes no enunciado deves formular e entregar um relatório onde deves incluir, conforme necessário:

- Raciocínio utilizado para obter a resposta.
- Código que suporte o teu raciocínio.
- Imagens, gráficos ou outputs.
- Conclusões.

Enunciado:

A empresa Sakila tem um negócio de aluguer de filmes suportado por uma base de dados com as seguintes tabelas e ligações:



No esquema anterior é possível observar as respetivas tabelas e os seus atributos, o tipo de dados de cada atributo tal como as ligações e tipo de ligações existentes entre tabelas. A representação de ligações segue a notação Crow's Foot (<https://www.freecodecamp.org/news/crows-foot-notation-relationship-symbols-and-how-to-read-diagrams/>).

- A tabela *film* contém informação referente à seleção de filmes que a empresa disponibiliza.
- A tabela *film_actor* contém informação referente aos atores que aparecem nos filmes.
- A tabela *inventory* permite saber que filmes é que se encontram em quais lojas.
- A tabela *customer* contém informações pessoais dos clientes da empresa.
- A tabela *store* contém informação referente às várias lojas que a empresa tem.
- A tabela *staff* contém informação pessoal e profissional dos colaboradores da empresa.
- A tabela *rental* contém o registo dos arrendamentos feitos na empresa.
- A tabela *payment* contém o registo dos pagamentos efetuados à empresa.
- A tabela *address* contém a lista de moradas das lojas, colaboradores e clientes da empresa.

Para uma descrição mais específica das colunas de cada tabela deves consultar o ficheiro excel fornecido em conjunto com enunciado “Descrição de Tabelas - Use Case.xlsx”

Questões:

Questão 1 - Store Analysis

Questão 1.1

A Sakila pretende saber quais são os filmes mais rentáveis em cada uma das suas lojas.

Através de uma consulta de SQL responde à seguinte questão que te foi colocada como Data Analyst da Sakila:

- **Quais são os 10 filmes com o maior rendimento total de aluguer para cada loja da empresa?**

Na consulta que irás efetuar, deves incluir o identificador de cada loja, o nome completo (numa única coluna) do manager de cada loja, o título do filme e o rendimento total dos alugueres do filme para a loja respetiva.

Questão 1.2

A Sakila pretende saber mais informações relativamente às suas lojas. Lista o **identificador**, **rendimento**, **número de clientes distintos** e **rendimento por cliente** de cada loja.

Questão 2 – Category Analysis

A Sakila pretende saber como é que o seu negócio é afetado pelo tipo de filmes que aluga.

Para tal pretende que obtenhas diferentes métricas e efetues diferentes análises para ajudar nas suas tomadas de decisão.

Questão 2.1

Recolhe a informação referente ao **rendimento total** por **categoria**. Cria uma visualização adequada para expor esta informação à empresa.

Para recolher a informação poderás recorrer a uma query de SQL ou a código R. Para a visualização, qualquer software que o permita.

Questão 2.2.1

Mostra à empresa qual é o **custo médio de aluguer de filmes por categoria**.

Para recolher a informação poderás recorrer a uma query de SQL ou a código R.

Questão 2.2.2

A Sakila pretende saber como é que aumentar a rentabilidade dos alugueis em categorias específicas afetaria o rendimento da empresa.

Mostra à empresa qual seria o impacto caso o preço de aluguer médio de filmes das categorias *Action*, *Drama* e *New* fosse **0.25** (valor monetário) mais alto. Mostra quanto seria o aumento nas receitas para as categorias afetadas, quanto seria o aumento e qual seria a receita total da empresa e qual seria a percentagem de crescimento da empresa a nível de receitas.

Para te ajudar nos cálculos destas métricas recolhe a informação referente ao número de alugueres por categoria.

Questão 2.3

A Sakila pretende efetuar uma campanha de marketing específica para cada loja. Esta campanha pretende mitigar o declínio das vendas nesta época do ano.

A campanha consiste num **desconto de 15%** no preço de aluguer de filmes no **top 3 categorias de cada loja** e teria a duração de um mês. Com base na média de vendas por mês, efetua uma previsão do impacto monetário da campanha no negócio caso as vendas aumentassem 5% no mês da campanha **(inclui apenas o ano de 2005 para efetuar a previsão)**.

Para responder ao pedido recorre a uma ou mais queries de SQL

Questão 2.4

Identifica em que mês do ano **2005** é que a evolução percentual dos valores de rendimento das categorias ***Comedy*** e ***Drama*** se distanciou mais.

Para extração de dados utiliza uma ou mais queries de **SQL** e a ferramenta de extração de dados do **MySQLWorkbench**.

Cria uma **visualização** para justificar a tua resposta.

Questão 3 – Actor Analysis

Questão 3.1.1

Mostra à Sakila quais são **os 10 atores que apareceram em mais filmes**. Para o fazer recorre ao R.

Questão 3.1.2

Mostra à Sakila quais são **os 10 atores mais populares nos alugueres**. Para o fazer recorre ao R.

Questão 3.2.1

A Sakila gostaria de saber quais são os atores que indiretamente **geraram mais rendimento** para a empresa de modo a efetuar futuras parcerias. Recolhe o **id**, **primeiro** e **último nome** e o **rendimento gerado**. Para o fazer deves recorrer ao R.

Questão 3.2.2

A Sakila gostaria de conjugar as parcerias com atores com as campanhas de marketing que pretende fazer, para tal precisa de informação suficiente para saber que ator é que deve representar cada campanha.

Faz uma consulta que inclua o **id do ator**, **primeiro e último nome**, a **categoria em que o ator mais aparece** e o **rendimento gerado pelo ator nessa categoria**.

Por fim identifica quais seriam os atores escolhidos para representar campanhas que atribuíssem descontos nas categorias: **Drama**, **Sci-Fi** e **Action**. (escolhe 1 ator para cada categoria).

Para responder a esta questão recorre a uma ou mais queries de SQL e ao SQLWorkbench.

Questão 4 – Customer Analysis

Questão 4.1

A Sakila reparou que alguns clientes não estavam a cumprir a cláusula de retorno de alugueres de **6 dias**. A empresa precisa de identificar estes clientes de modo a aplicar uma caução nos seus próximos alugueres.

Questão 4.1.1

Cria uma lista que para cada cliente contenha: **id**, **primeiro e último nome** e **tempo que em média demora a devolver os alugueres (em dias)**. Para o fazer recorre a uma ou mais queries de SQL.

Questão 4.1.2

Cria uma coluna na tabela **customer** chamada **blacklist** para alojar informação que identifique se este cliente tem um **tempo médio de devolução de aluguer** superior a **6 dias** ou não. A coluna deve ser criada com o valor default **“No”** e depois deves proceder à atualização da coluna com base nos critérios já mencionados. Para o fazer recorre a uma ou mais queries de SQL.

Questão 4.2

A Sakila tem clientes de vários países. Porém, a seleção de filmes que tem em inventário apenas está traduzida em inglês. A Sakila precisa que analyses os dados dos seus clientes e concluas **quais são os 5 idiomas adicionais** para os quais se deve arranjar traduções.

Para o fazer deves recorrer ao R.

Questão 4.3

A Sakila gostaria de implementar um sistema de ranking dos seus clientes. O seu objetivo é dividir a sua carteira de clientes em **três ranks diferentes**. Cliente **Bronze**, **Silver** ou **Gold**.

Para implementar este novo sistema a empresa precisa que atribuas um rank a cada cliente com base no seu gasto total na empresa.

Questão 4.3.1

Através da análise da distribuição do consumo total dos clientes da empresa **define critérios** para atribuir os diferentes **ranks**.

Para exportação de dados poderás recorrer a uma ou mais queries de SQL e a ferramenta de exportação do MySQLWorkbench.

Questão 4.3.2

Para que a análise posterior seja facilitada é necessário que o **rank de cada cliente esteja disponível na base de dados**.

Com base nos **critérios que definiste** anteriormente insere na tabela de **clientes** a informação referente ao **rank de cada cliente**.

Para qualquer alteração feita a tabelas da base de dados deverás usar uma ou mais queries de SQL e o MySQL Workbench.

Questão 4.3.3

Cria uma lista de recomendação de filmes para cada tipo de cliente de forma a aumentar a margem de lucro da Sakila. Recomenda **3** filmes para cada tipo de cliente com base no número de alugueres.

Para efeito de consulta da base de dados deves recorrer a uma ou mais queries de SQL.

Questão 4.4

A Sakila pretende ter conhecimento da categoria preferida de cada cliente na sua base de dados. Sendo o critério para definir categoria preferida **como dinheiro gasto em filmes dessa categoria** procede à criação da coluna **favorite_category** e insere a informação pretendida na mesma.

Para o fazer recorre a uma ou mais queries de SQL e ao SQLWorkbench para gestão da base de dados.

Questão 5 – Cost Analysis

Questão 5.1

A Sakila pretende começar a cobrar uma taxa de atraso caso os clientes demorem mais do que **6 dias** a devolver os filmes que alugam.

A empresa pretende ter uma previsão de quanto dinheiro irá ganhar/perder com esta nova medida **num período de atividade igual** ao que contêm na sua base de dados.

Sabendo que perderá **3%** dos seus clientes com esta medida e cobraria **1€ por atraso** na devolução **calcula se vale a pena ou não avançar com a medida e quanto dinheiro ganharia/perderia** ao avançar caso os valores do próximo período de atividade fossem iguais ao anterior.

Para responder à questão debes usar uma ou mais queries de SQL.

Questão 5.2

A Sakila está em processo de formalizar um novo seguro para as suas lojas. O seguro cobre custos de reposição de stock caso uma das lojas da empresa seja completamente destruída por causas naturais.

Sabendo que o seguro cobra uma anualidade de **2,5%** do valor total de reposição de stock e que no segundo ano esse valor recebe um desconto de **20%** efetua uma consulta que **para cada loja** inclua a **identificação**, o **valor total de reposição de stock**, **preço do seguro no primeiro ano** e o **preço do seguro no segundo ano**.

Para responder à questão debes usar uma ou mais queries de SQL.