Passo a Passo

exemplo de como você pode fazer isso:

1. Faça o download do conjunto de dados Brazilian E-Commerce Public Dataset by Olist do Kaggle e salve-o em um bucket do Amazon S3.

2. Crie um cluster do Amazon Redshift e configure-o para acessar o bucket do Amazon S3.

3. Use o COPY command do Amazon Redshift para carregar os dados do conjunto de dados do bucket do Amazon S3 para o cluster do Amazon Redshift.

4. Agora você pode usar o SQL para consultar e analisar os dados no cluster do Amazon Redshift.

exemplo de como você pode documentar todo o processo do seu projeto de engenharia de dados passo a passo:

1. \*\*Introdução\*\*: Descreva o objetivo do seu projeto e o conjunto de dados que você está usando (Brazilian E-Commerce Public Dataset by Olist do Kaggle).

2. \*\*Configuração\*\*: Descreva como você configurou o ambiente para o seu projeto, incluindo a criação de um bucket do Amazon S3 e um cluster do Amazon Redshift.

3. \*\*Carregamento de Dados\*\*: Descreva como você carregou os dados do conjunto de dados do bucket do Amazon S3 para o cluster do Amazon Redshift usando o COPY command.

4. \*\*Análise de Dados\*\*: Descreva como você usou o SQL para consultar e analisar os dados no cluster do Amazon Redshift. Inclua exemplos de consultas e resultados.

5. \*\*Conclusão\*\*: Resuma os resultados da sua análise e discuta as conclusões que você tirou dos dados.

exemplo de introdução para o seu projeto de engenharia de dados:

Neste projeto, vamos explorar o conjunto de dados Brazilian E-Commerce Public Dataset by Olist do Kaggle. Este conjunto de dados contém informações sobre pedidos feitos em uma plataforma de comércio eletrônico brasileira entre 2016 e 2018. Vamos usar as ferramentas Amazon S3 e Amazon Redshift para armazenar e analisar os dados. Nosso objetivo é entender melhor o comportamento dos clientes e identificar padrões e tendências nas vendas.