

Sumário

1. Introdução
2. Contexto e Visão Geral do Projeto
 - 2.1 Visão Geral da Empresa
 - 2.2 Partes Interessadas
 - 2.3 Objetivos
3. Estratégia de Dados e Abordagem de Transformação
 - 3.1 Fontes de Dados e Pipeline de Ingestão
 - 3.2 Análise e Qualidade dos Dados
 - Medidas Tomadas para Correção
 - 3.3 Tabelas Staging
 - 3.4 Transformação de Dados
 - 3.5 Governança de Dados e Validações
 - 3.6 Pipeline de Dados e Arquitetura
 - Ingestão de Dados
 - Transformação de Dados
 - Analytics
4. Data Warehouse e Design da Arquitetura
 - 4.1 Modelo Conceitual e Star Schema
 - 4.2 Tabelas Fato e Dimensão
 - 4.2.1 Tabela Fato: fact_order_item
 - 4.2.2 Tabelas Dimensão
 - 4.3 Processo de Transformação e Tabelas Intermediárias
 - 4.3.1 Tabelas Intermediárias (int)
 - 4.3.2 Tabelas Staging (stg)
 - 4.4 Relações e Agregações
 - 4.5 Escolha do Data Warehouse: Google BigQuery
 - 4.5.1 Por que Google BigQuery?
 - 4.5.2 Custos Envolvidos com BigQuery
 - 4.5.3 Benefícios de Custo e Operação
5. Apresentação dos Dados
 - 5.1 Relatório Gerencial
 - 5.2 Dashboard Operacional
 - 5.2.1 Análise por Loja
 - 5.2.2 Análise por Território de Vendas
 - 5.2.3 Análise de Clientes
 - 5.2.4 Análise de Pedidos
 - 5.2.5 Análise de Produtos
 - 5.2.6 Análise por Vendedor
6. Planejamento do Projeto de Dados e Ferramentas Utilizadas
 - 6.1 Fases do Projeto de Dados
 - 6.1.1 Descoberta e Diagnóstico Inicial (2 semanas)

- 6.1.2 Definição da Arquitetura e Ferramentas (1 semana)
- 6.1.3 Ingestão de Dados e Preparação (2 semanas)
- 6.1.4 Transformação e Limpeza de Dados (3 semanas)
- 6.1.5 Desenvolvimento de Dashboards e Relatórios (3 semanas)
- 6.1.6 Análise Preditiva e Modelos de Regressão (3 semanas)
- 6.1.7 Testes e Validação (2 semanas)
- 6.1.8 Treinamento e Documentação (1 semana)
- 6.1.9 Entrega e Lançamento do Projeto (1 semana)

6.2 Ferramentas Utilizadas

- 6.2.1 Google BigQuery (Data Warehouse)
- 6.2.2 dbt (Data Build Tool)
- 6.2.3 Power BI (Business Intelligence)
- 6.2.4 Python (Modelos de Machine Learning)
- 6.2.5 Google Cloud Platform (Infraestrutura)
- 6.2.6 GitHub (Controle de Versão)

6.3 Estimativa de Custos Totais

6.4 Considerações sobre o Tempo de Execução

7. Previsões e Modelos Utilizados

- 7.1 Previsão de Demanda para os Próximos 3 Meses
 - 7.1.1 Estruturação dos Dados
 - 7.1.2 Previsões com Séries Temporais
 - 7.1.3 Previsões com Modelos de Aprendizado de Máquina
- 7.2 Estimativa de Matéria-Prima: Previsão de Zíperes
 - 7.2.1 Definição do Problema
 - 7.2.2 Modelos Usados

7.3 Resultados por Território

- 7.4 Conclusão e Melhorias Implementadas
- 7.5 Localização dos Resultados das Previsões

8. Resultados Concretos e Impacto no Negócio

- 8.1 Melhorias na Previsão de Demanda
- 8.2 Melhoria no Planejamento e Distribuição de Matéria-Prima
- 8.3 Aumento na Eficiência Comercial e Estratégia de Vendas
- 8.4 Redução de Problemas Operacionais e Melhoria na Satisfação do Cliente
- 8.5 Melhoria na Organização e Acesso aos Dados
- 8.6 Utilidade dos Dashboards para Operações e Gerência

9. Desafios e Riscos Enfrentados no Projeto

- 9.1 Desafios Técnicos
 - Solução
- 9.2 Desafios de Negócio
 - Solução
- 9.3 Riscos do Projeto
 - 9.3.1 Acesso aos Dados
 - Solução
 - 9.3.2 Escalabilidade e Custos

Solução

9.4 Benefícios de Negócio com a Infraestrutura de Dados e Dashboards

Organização e Acesso a Dados

Facilidade de Acesso às Informações Tratadas

Utilidade dos Dashboards

Redução de Riscos e Melhoria da Eficiência Operacional

10. Conclusões

Conclusão Final

1. Introdução

A Adventure Works (AW) é uma empresa em rápido crescimento no setor de fabricação de bicicletas, reconhecida por sua vasta gama de produtos e uma base de clientes diversificada. Com mais de 500 produtos distintos e 31.000 pedidos processados, a empresa se destaca como uma líder no mercado de ciclismo.

Para sustentar seu crescimento e se diferenciar da concorrência, a AW busca transformar sua abordagem em uma organização orientada por dados, aproveitando o potencial de seus ativos de dados para tomar decisões mais informadas, aumentar a eficiência operacional e criar estratégias inovadoras.

Este relatório detalha o projeto de modernização da infraestrutura de dados da Adventure Works, que visa consolidar dados fragmentados e desatualizados em uma plataforma única e acessível. O projeto aborda a construção de um Data Warehouse robusto, a criação de dashboards interativos e a implementação de análises preditivas, com o objetivo de melhorar o planejamento de vendas e a tomada de decisões estratégicas.

Com o apoio de partes interessadas-chave dentro da organização, o projeto busca não apenas resolver problemas operacionais existentes, mas também criar novas oportunidades de crescimento, tornando a Adventure Works uma empresa verdadeiramente data-driven.

2. Contexto e Visão Geral do Projeto

2.1 Visão Geral da Empresa

A **Adventure Works (AW)** é uma empresa de fabricação de bicicletas, conhecida por sua vasta gama de produtos e base de clientes. Com mais de **500 produtos distintos, 20.000 clientes e 31.000 pedidos** processados, a AW se orgulha de ser uma líder no setor de ciclismo. Para manter seu ritmo de crescimento e se diferenciar da concorrência, a empresa busca utilizar seus dados de forma estratégica, transformando-se em uma organização **data-driven**.

A AW reconhece o valor dos dados para informar decisões de negócios, aumentar a eficiência operacional e criar estratégias inovadoras. No entanto, apesar de suas ambições, a empresa enfrenta desafios ao utilizar plenamente seus ativos de dados, devido a sistemas de dados fragmentados e métodos de análise desatualizados.

Atualmente, a infraestrutura de banco de dados da empresa inclui **68 tabelas** distribuídas em cinco esquemas: **Recursos Humanos (HR)**, **Vendas (Sales)**, **Produção (Production)** e **Compras (Purchasing)**. Esse grande volume de dados possui um potencial significativo para revelar insights valiosos para o negócio, mas as ferramentas e processos existentes limitam a capacidade da AW de extrair, analisar e agir com base nesses dados de forma eficaz.

2.2 Partes Interessadas

O projeto de transformação de dados da **Adventure Works** está sendo conduzido por stakeholders dentro da organização, cada um com diferentes perspectivas e níveis de apoio para a iniciativa.

- **João Muller**, Diretor de Inovação, é o principal patrocinador deste projeto. Ele vê os dados como um diferencial estratégico que pode impulsionar a Adventure Works para o futuro. João acredita que a infraestrutura moderna de dados permitirá que a empresa tome decisões melhores, otimize operações e identifique novas oportunidades de negócios.
- **Carlos Silveira**, CEO da Adventure Works, apoia a visão de João e considera o projeto um investimento essencial para o sucesso de longo prazo da empresa. Carlos está particularmente interessado em usar dados para aumentar a vantagem competitiva da empresa e melhorar a tomada de decisões em nível executivo. Ele espera que o projeto forneça insights sobre **tendências de vendas, comportamento dos clientes e oportunidades de mercado**.
- **Silvana Teixeira**, Diretora Comercial, é mais cética em relação à iniciativa. Ela questiona o retorno sobre o investimento (ROI) da plataforma de dados, acreditando que os fundos teriam sido melhor utilizados em atividades promocionais imediatas para impulsionar as vendas. Silvana já viu promessas semelhantes de outros sistemas, como **CRM** e **Web Analytics**, que não atingiram os resultados esperados. Ela precisa de insights claros e açãoáveis da infraestrutura de dados que possam beneficiar diretamente o departamento comercial, especialmente em relação a vendas e segmentação de clientes.
- **Nilson Ramos**, Diretor de TI, desempenha um papel fundamental ao garantir que o projeto tenha acesso aos dados necessários. Ele designou **Gabriel Santos**, um analista de TI, para gerenciar o acesso aos dados. No entanto, o tempo de Gabriel é limitado devido a outras responsabilidades, o que torna a **comunicação eficiente** e a coordenação essenciais para o sucesso do projeto.
- **Luís Soares**, Chefe de Planejamento de Demanda, demonstrou interesse em usar dados para otimizar o planejamento de produção. Atualmente, o departamento depende de **médias móveis de três meses** calculadas no Excel. Luís está aberto a adotar modelos preditivos para melhorar a previsão de demanda, mas só se comprometerá com mais investimentos se o projeto demonstrar um valor claro em comparação ao processo atual.

2.3 Objetivos

O objetivo principal deste projeto é transformar a **Adventure Works** em uma organização **data-driven** através da construção de uma infraestrutura robusta de dados que suporte a tomada de decisões estratégicas. O projeto foca inicialmente na **área de Vendas**, embora possa requerer dados de outros departamentos, como **Produção e Compras**.

Os objetivos específicos incluem:

- **Construção de um Data Warehouse:** Consolidar dados de várias fontes, permitindo análises abrangentes de vendas e clientes.
- **Criação de Tabelas Fato e Dimensão:** Essas tabelas serão projetadas para responder às **perguntas de negócios específicas** levantadas pela liderança, como desempenho de vendas por região, tendências de comportamento do cliente e eficiência operacional.
- **Implementação de Análises Preditivas:** Melhorar o planejamento de produção e inventário por meio de modelos preditivos, especialmente na previsão de demanda para os próximos três meses.
- **Desenvolvimento de Dashboards de Inteligência de Negócios:** Para executivos de alto nível e equipes operacionais, permitindo acesso em tempo real a métricas e insights-chave.

A fase inicial do projeto concentra seus esforços na **análise de vendas**, com dashboards voltados tanto para o **CEO Carlos Silveira**, que busca **insights estratégicos de alto nível**, quanto para a **Silvana Teixeira**, com foco em **métricas operacionais de vendas**.

O sucesso do projeto será medido pela sua capacidade de fornecer insights açãoáveis, melhorar a eficiência operacional e oferecer **capacidades preditivas** que demonstrem um valor claro em relação aos métodos manuais atuais.

3. Estratégia de Dados e Abordagem de Transformação

3.1 Fontes de Dados e Pipeline de Ingestão

A **Adventure Works** utiliza um banco de dados transacional PostgreSQL que armazena dados de diferentes áreas da empresa, como Vendas, Produção, Compras e Recursos Humanos, distribuídos em **68 tabelas**. Para esse projeto, o foco inicial foi na área de **vendas**, mas dados de outras áreas também foram integrados.

Os dados foram extraídos para o data warehouse (DW) utilizando **seeds no dbt**. Esse processo foi executado localmente e os dados foram carregados no **Google BigQuery** em um ambiente de staging, armazenando as tabelas em sua forma bruta.

A pipeline de ingestão seguiu o método **Extract-Load (EL)**, sem transformações antes do carregamento dos dados no DW. As tabelas brutas foram salvas no schema **stg_adventure_works**, representando a primeira etapa da construção da infraestrutura de dados.

3.2 Análise e Qualidade dos Dados

Após a ingestão inicial dos dados brutos, foi realizada uma análise detalhada da base de dados para identificar os principais problemas de qualidade que poderiam comprometer tanto as operações diárias quanto a estratégia de longo prazo da **Adventure Works**. As principais questões encontradas foram:

1. Campos Nulos em Áreas Críticas:

- **Problema:** Diversos campos que deveriam ser obrigatórios apresentavam valores nulos, como **informações de contato de clientes, endereços de entrega, e dados de produtos**. Esses campos são essenciais para operações como o atendimento ao cliente, a logística de entregas e o processamento de vendas.
- **Impacto:** A ausência desses dados impacta diretamente a eficiência operacional, podendo resultar em falhas nas entregas, perda de clientes, dificuldade na personalização de campanhas de marketing, e até erros no inventário. Para uma empresa que busca se diferenciar no atendimento ao cliente, essa falha pode levar a uma diminuição da satisfação e à perda de receita.

2. Endereços Incorretos e Incompletos:

- **Problema:** Foram identificados diversos endereços de clientes com erros ou dados faltantes, o que inviabiliza o envio de produtos corretamente e pode gerar retornos desnecessários.
- **Impacto:** Problemas com endereços incorretos geram **custos adicionais com logística** e, mais criticamente, podem prejudicar a **experiência do cliente**, afetando diretamente a retenção de clientes e a reputação da empresa. Além disso, podem ocorrer atrasos na entrega, o que afeta o cumprimento de prazos, aumentando as reclamações.

3. Dados de Clientes Sensíveis Expostos:

- **Problema:** Foi descoberto que informações altamente sensíveis, como **números de cartões de crédito e senhas de clientes**, estavam armazenadas sem qualquer tipo de criptografia. Essas informações estavam em texto simples, acessíveis a qualquer consulta direta no banco de dados.
- **Impacto:** A falta de criptografia para esses dados representa um **risco sério de segurança** e pode levar a violações da privacidade dos clientes, o que expõe a empresa a **multas regulatórias** (como as previstas pela **LGPD** e **GDPR**), além de prejudicar a confiança dos clientes. Uma violação desse tipo pode resultar em perda massiva de clientes e danos irreparáveis à marca da empresa.

4. Dados Duplicados:

- **Problema:** Durante a análise, foram identificados muitos registros duplicados, especialmente relacionados a **clientes e pedidos de venda**.
- **Impacto:** A duplicação de dados leva a **inconsistências nas análises e relatórios**, causando distorções em métricas críticas, como total de vendas, número de clientes ativos, e inventário. A tomada de decisões baseada em dados duplicados ou imprecisos pode comprometer iniciativas estratégicas, como promoções e planejamento de estoque.

5. Campos Faltantes em Produtos:

- **Problema:** Informações importantes sobre os produtos, como **modelos, categorias e descrições**, estavam ausentes em muitos registros.
- **Impacto:** Isso afeta diretamente a **precisão das previsões de demanda** e dificulta a criação de análises detalhadas sobre o desempenho de produtos, além de atrapalhar iniciativas de marketing personalizadas, prejudicando o crescimento e a eficácia nas vendas.

6. Dados Incoerentes nas Relações Entre Tabelas:

- **Problema:** Relações entre tabelas importantes, como entre **clientes e pedidos**, e entre **vendedores e territórios de vendas**, apresentavam falhas. Em muitos casos, registros de uma tabela não possuíam correspondência na tabela relacionada.
- **Impacto:** Esses problemas comprometem a capacidade de gerar **relatórios confiáveis e dashboards** que dependem de dados interconectados. Isso afeta diretamente a **eficiência operacional e a tomada de decisões estratégicas**, como a alocação de vendedores para territórios ou a análise do comportamento do cliente.

7. Falta de Anonimização de Dados Pessoais:

- **Problema:** Dados pessoais, como **nomes completos, endereços de e-mail, e números de telefone**, estavam armazenados sem qualquer medida de anonimização ou pseudonimização.
- **Impacto:** Essa falta de anonimização viola as **diretrizes de privacidade e segurança de dados**, expondo a empresa a riscos legais e financeiros. A conformidade com leis de proteção de dados é fundamental para evitar multas pesadas e garantir a confiança dos clientes no manejo de suas informações.

Medidas Tomadas para Correção

Com base nos problemas identificados, foram aplicadas as seguintes soluções:

- **Preenchimento de Valores Nulos:** Utilizamos técnicas de imputação e completude de dados para preencher campos nulos essenciais, como endereços e informações de produtos.
- **Correção de Endereços:** Foram implementadas rotinas de verificação e correção de endereços com base em APIs de geolocalização.
- **Remoção de Dados Duplicados:** Processos de deduplicação foram aplicados para garantir a integridade dos registros, especialmente nas tabelas de clientes e vendas.
- **Criptografia de Dados Sensíveis:** Foi implementada a criptografia dos dados de clientes, incluindo senhas e informações de pagamento, garantindo que as informações sensíveis estejam protegidas.
- **Validação e Auditoria de Relações entre Tabelas:** Validações de consistência foram aplicadas para garantir que todas as relações entre tabelas fossem coerentes e confiáveis.

Essas correções não só melhoraram a **qualidade dos dados**, mas também aumentaram a **confiabilidade dos relatórios e dashboards**, permitindo que a Adventure Works tome **decisões informadas** com base em dados precisos e protegidos.

3.3 Tabelas Staging

Após a etapa de limpeza e validação, as tabelas staging foram criadas no **schema stg_adventure_works** usando o **dbt**. As tabelas staging são responsáveis por organizar os dados brutos em um formato que facilita as transformações posteriores. As principais tabelas staging criadas incluem:

- **stg_address**
- **stg_address_type**
- **stg_business_entity**
- **stg_customer**
- **stg_product**
- **stg_sales_order_detail**
- **stg_sales_person**, entre outras.

Essas tabelas consolidaram os dados necessários para a criação dos modelos intermediários, conhecidos como **tabelas intermediárias**, onde ocorreram as principais transformações de dados.

3.4 Transformação de Dados

As transformações de dados foram realizadas nas **tabelas intermediárias (int)** no schema **dev_adventure_works**, a partir das tabelas staging. Nessas transformações, foram aplicadas diversas validações, correções e ajustes para garantir a integridade e consistência dos dados. As tabelas intermediárias criadas incluem:

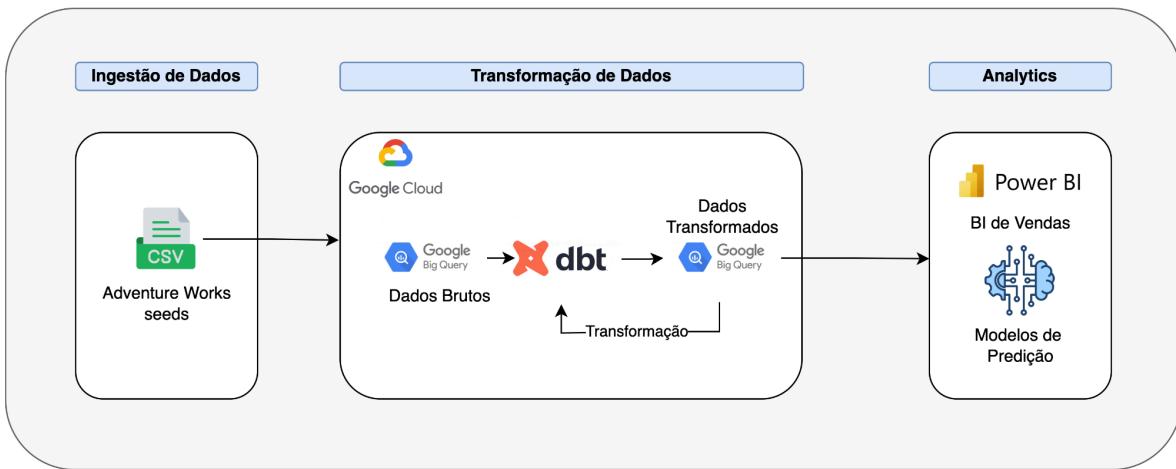
- **int_customer_info**: Consolidou informações detalhadas sobre os clientes.
- **int_product_info**: Reuniu dados relacionados aos produtos, como categorias e modelos.
- **int_sales_person_info**: Organizou informações dos vendedores e seus históricos de quotas.
- **int_sales_summary**: Agregou dados de vendas para criar um resumo com métricas essenciais, com granularidade de item por pedido.
- **int_sales_territory_info**: Reuniu informações das regiões de vendas.
- **int_special_offer_info**: Centralizou os dados relacionados a ofertas especiais.
- **int_store_info**: Organizou as informações das lojas da Adventure Works.

3.5 Governança de Dados e Validações

Para garantir a qualidade dos dados, foi implementada uma estratégia robusta de **governança de dados**. Isso incluiu:

- **Validações de esquema e testes de dados**, assegurando que os dados carregados e transformados estavam corretos e completos.
- **Documentação de toda a base de dados** dentro do dbt, detalhando cada campo e sua função.
- Aplicação de regras de **criptografia** e **anonimização** para proteger informações sensíveis de clientes, como senhas e dados de pagamento.

3.6 Pipeline de Dados e Arquitetura



A imagem apresentada ilustra o **pipeline de dados** utilizado para o projeto da **Adventure Works**, demonstrando como os dados são extraídos, transformados e, finalmente, utilizados para gerar insights de negócio através de **BI** e **modelos preditivos**.

Ingestão de Dados

A etapa inicial do pipeline começa com a **ingestão de dados**. Neste caso, os dados foram extraídos a partir de **seeds da Adventure Works** no formato **CSV**. Esses seeds contêm dados brutos que são carregados diretamente para o ambiente do **Google Cloud** na ferramenta **BigQuery**. Nessa fase, os dados ainda estão em sua forma original e não passaram por nenhum processo de limpeza ou transformação.

Transformação de Dados

Após a ingestão dos dados no BigQuery, entra a fase de **transformação de dados** utilizando o **dbt (Data Build Tool)**. O dbt é responsável por transformar os dados brutos em um formato mais estruturado e utilizável, por meio de scripts que:

- Realizam a **limpeza de dados**, removendo duplicidades e corrigindo valores errados ou nulos.
- Aplicam **validações** para garantir que os dados estejam consistentes e prontos para análise.
- Organizam os dados em **modelos intermediários** e **dimensões**, criando tabelas transformadas que facilitam análises avançadas e futuras utilizações.

Os dados transformados são, então, armazenados novamente no **Google BigQuery** como **dados transformados**, prontos para serem consumidos pelas ferramentas de análise e modelos.

Analytics

A etapa final do pipeline é o uso dos dados transformados para **Analytics**. Existem duas principais saídas dessa etapa:

1. BI de Vendas no Power BI:

- As tabelas transformadas são conectadas ao **Power BI** para gerar dashboards de Business Intelligence (BI), oferecendo insights de vendas, desempenho por região, comportamento dos clientes, entre outros.

2. Modelos de Predição:

- Além do BI, os dados transformados são utilizados para alimentar **modelos preditivos**, responsáveis por gerar previsões, como demanda de produtos nos próximos três meses. Isso ajuda a equipe de Planejamento de Demanda a tomar decisões mais embasadas, comparando com os métodos anteriores baseados em médias móveis.

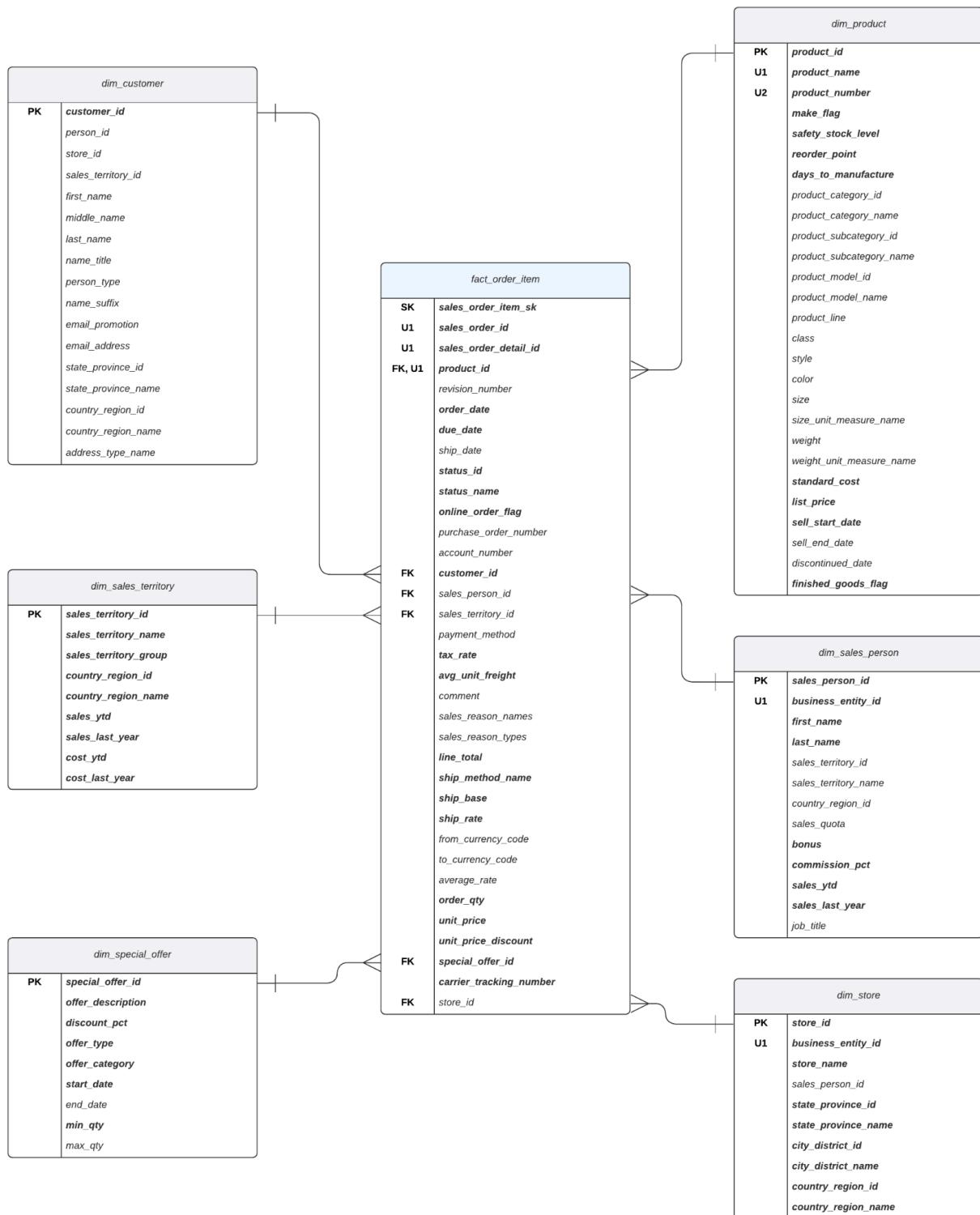
Esse pipeline, como mostrado na imagem, permite com que a **Adventure Works** utilize seus dados de forma estratégica, alinhando os objetivos do projeto com a transformação da empresa em uma organização **data-driven**.

4. Data Warehouse e Design da Arquitetura

4.1 Modelo Conceitual e Star Schema

O **Data Warehouse da Adventure Works** foi projetado utilizando um modelo **Star Schema**, uma das abordagens mais comuns em data warehouses devido à sua simplicidade e eficiência para consultas analíticas. O **star schema** tem como principal característica a utilização de uma **tabela fato** no centro, conectada a várias **tabelas dimensão**.

A tabela fato escolhida para esse modelo é a **fact_order_item**, que armazena dados detalhados de transações de vendas, como IDs de pedidos, produtos vendidos, quantidades e valores relacionados. Em torno dessa tabela-fato, temos diversas **tabelas-dimensão** que contêm informações adicionais para enriquecer as análises, como detalhes de clientes, produtos, territórios de venda, vendedores, ofertas especiais e lojas.



4.2 Tabelas Fato e Dimensão

4.2.1 Tabela Fato: fact_order_item

A tabela **fact_order_item** é a parte principal do modelo star schema. Ela armazena cada linha de pedido de venda, com informações específicas de cada item comprado, suas quantidades e os valores envolvidos. Os principais campos incluem:

- **`sales_order_item_sk`**: A chave substituta única para identificar cada item de pedido.
- **`sales_order_id`**: O ID do pedido de venda.
- **`product_id`**: A chave estrangeira referenciando a dimensão de produto.
- **`customer_id`**: A chave estrangeira referenciando a dimensão de cliente.
- **`sales_person_id`**: A chave estrangeira referenciando a dimensão de vendedor.
- **`sales_territory_id`**: A chave estrangeira referenciando a dimensão de território de vendas.
- **`order_date`**: A data em que o pedido foi realizado.
- **`ship_date`**: A data em que o pedido foi enviado.
- **`unit_price`**: O preço unitário do item.
- **`order_qty`**: A quantidade de itens vendidos.
- **`total_due`**: O valor total do pedido.
- **`tax_rate`**: A taxa de imposto aplicada.

Essa tabela é preenchida a partir de diversas tabelas intermediárias (**int**) que consolidam os dados brutos vindos das tabelas **staging**.

4.2.2 Tabelas Dimensão

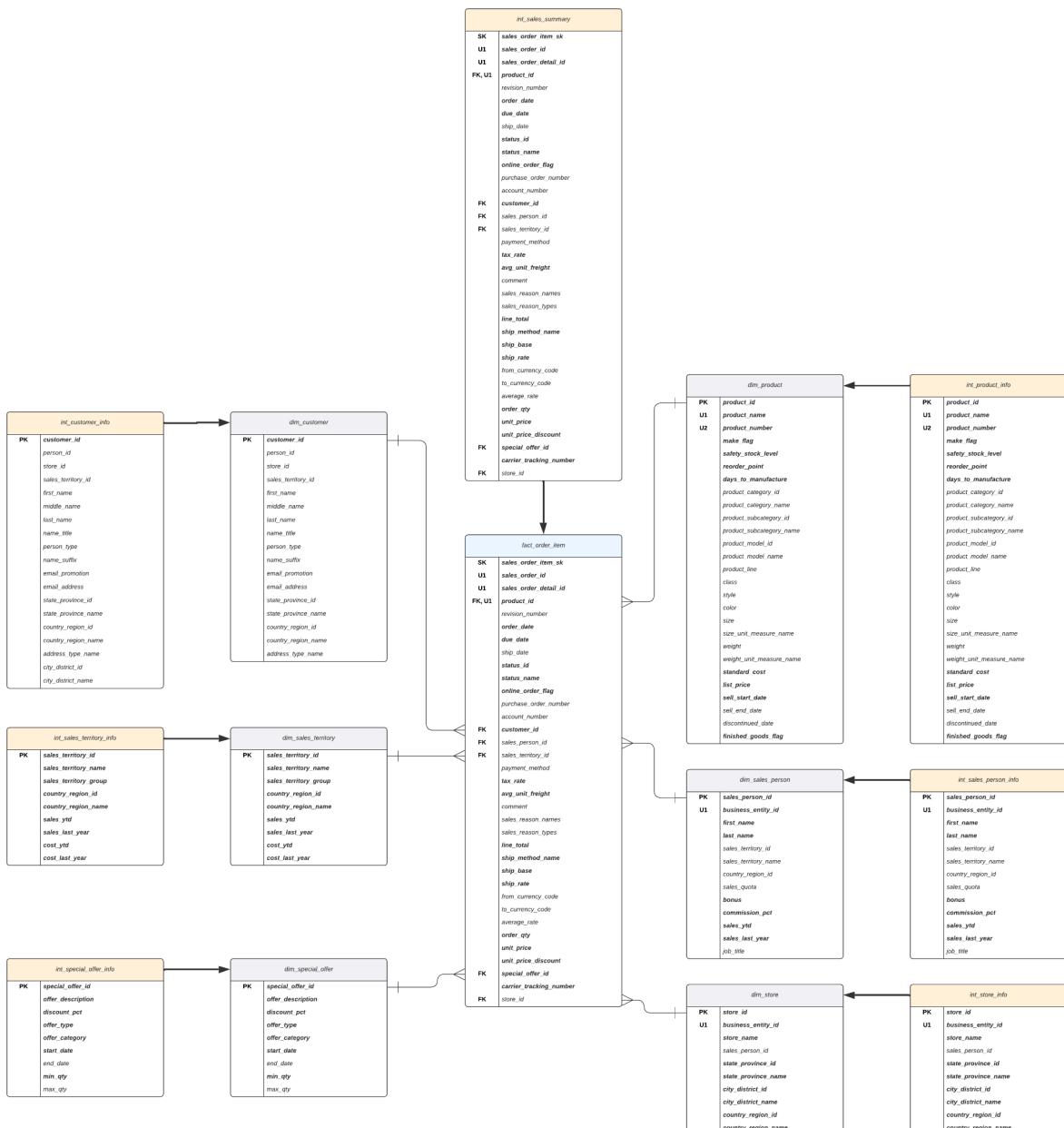
As **tabelas dimensão** fornecem os detalhes adicionais necessários para analisar as transações de vendas em diferentes perspectivas:

- **`dim_customer`**: Armazena informações detalhadas sobre os clientes. Campos como **`customer_id`, `first_name`, `last_name`, `email_address`, `city`, `state_province_name`, e `country_region_name`** permitem realizar análises sobre a localização dos clientes, comportamentos de compra e segmentação para campanhas de marketing.
- **`dim_product`**: Contém os detalhes de cada produto, como **`product_id`, `product_name`, `product_category_id`, `product_subcategory_id`, `standard_cost`, e `list_price`**. Essa dimensão possibilita analisar o desempenho de produtos em diferentes categorias e subcategorias, além de compreender melhor os custos envolvidos na produção.
- **`dim_sales_person`**: Representa os vendedores da **Adventure Works**. Campos como **`sales_person_id`, `first_name`, `last_name`, `territory_id`, `bonus`, `commission_pct`, e `sales_quota`** permitem avaliar o desempenho dos vendedores, comparando suas vendas com suas metas de comissão e bônus.
- **`dim_sales_territory`**: Fornece detalhes sobre os territórios de vendas. Contém campos como **`sales_territory_id`, `sales_territory_name`, `country_region_name`, `sales_ytd`, e `cost_ytd`**, que permitem análises por regiões geográficas, incluindo o desempenho de vendas e os custos associados.
- **`dim_special_offer`**: Armazena dados sobre as ofertas especiais aplicadas às vendas, como **`special_offer_id`, `offer_description`, `discount_pct`, e `offer_type`**. Isso facilita análises sobre o impacto de promoções e descontos nas vendas totais.

- **dim_store**: Representa as lojas da **Adventure Works**. Campos como **store_id**, **store_name**, **city_district_id**, **state_province_id**, e **sales_territory_id** permitem analisar o desempenho por loja, seja por região ou em nível local, ajudando a identificar quais locais geram mais receita.

4.3 Processo de Transformação e Tabelas Intermediárias

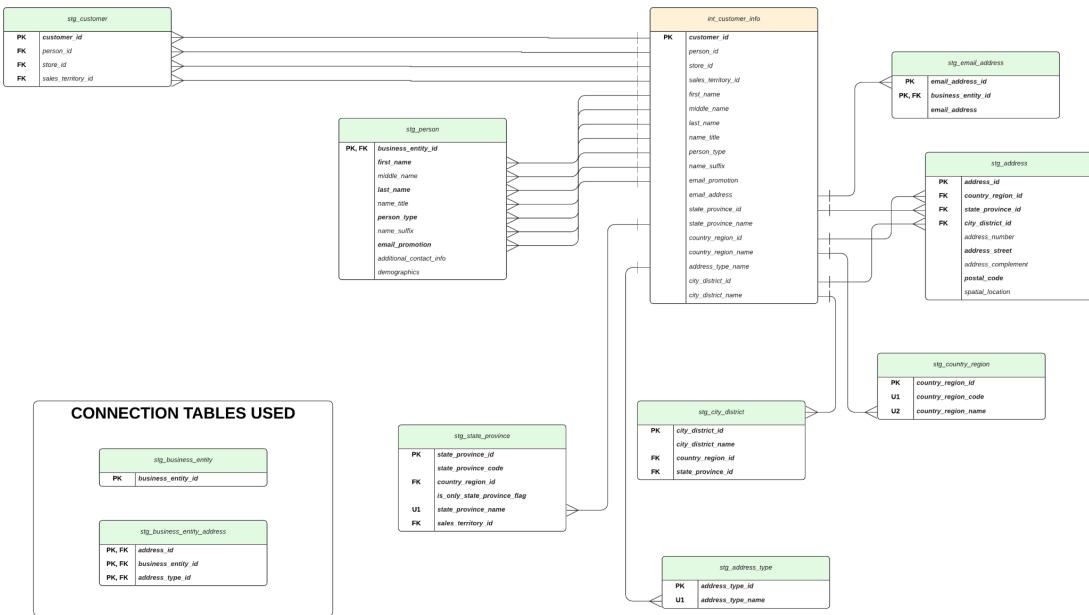
A construção do **star schema** foi baseada em um processo de transformação que utilizou tabelas **intermediárias (int)** para agregar e consolidar os dados brutos das tabelas **staging (stg)**. As **tabelas intermediárias** foram fundamentais para realizar as transformações necessárias, incluindo a correção de dados faltantes, eliminação de duplicatas e preenchimento de valores nulos.



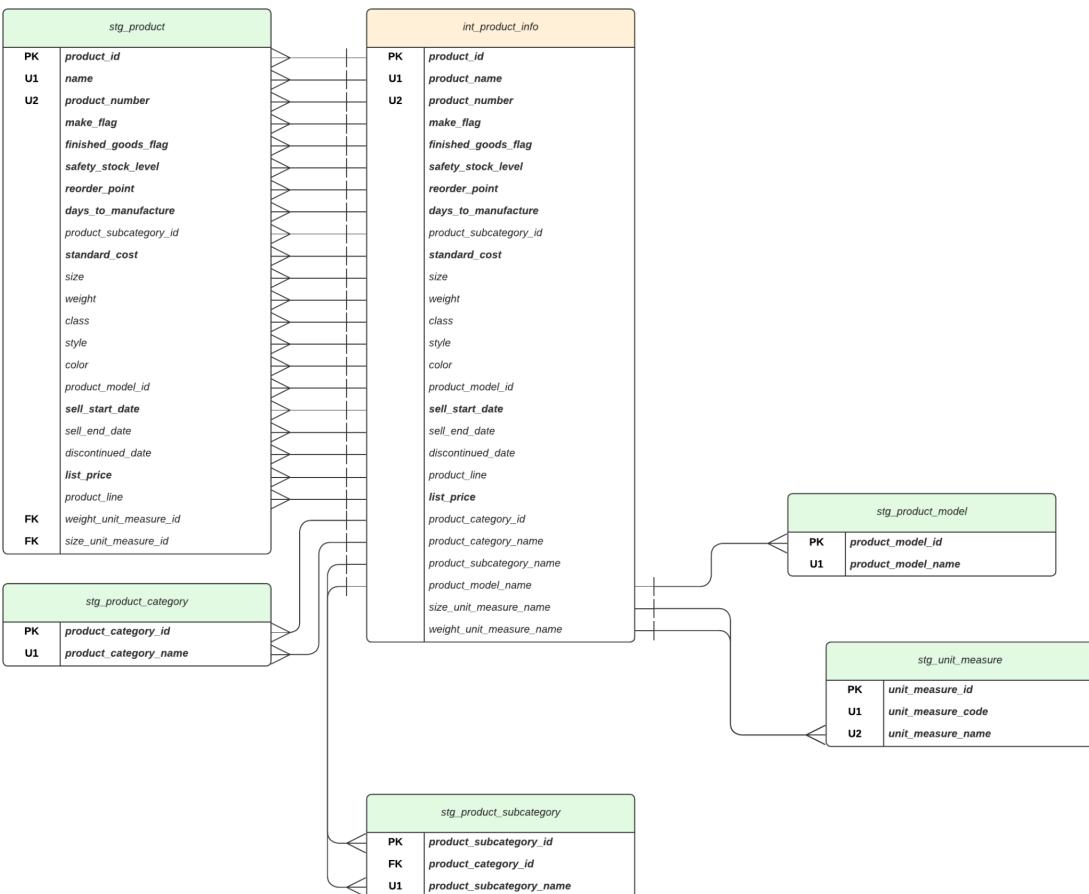
4.3.1 Tabelas Intermediárias (int)

As principais tabelas intermediárias criadas foram:

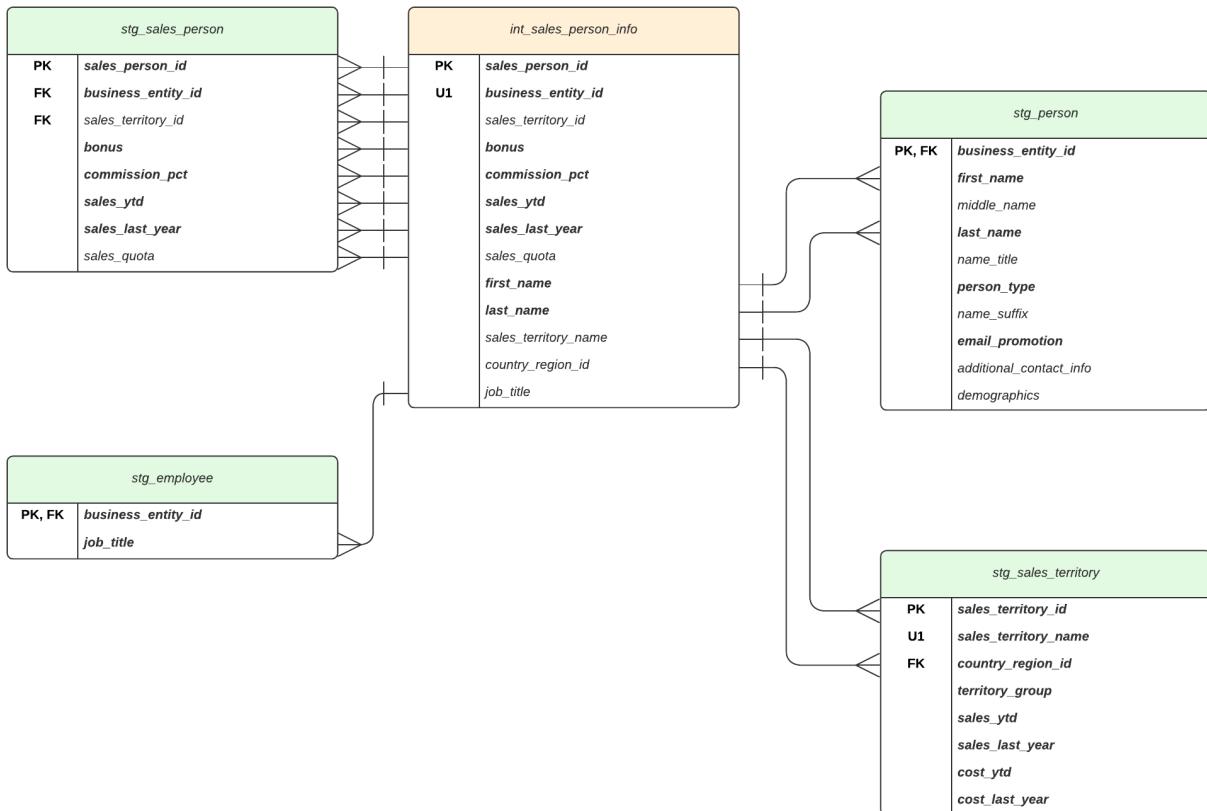
- **int_customer_info**: Construiu a dimensão **dim_customer**, consolidando dados de clientes de várias fontes e validando informações críticas como endereços e promoções por e-mail.



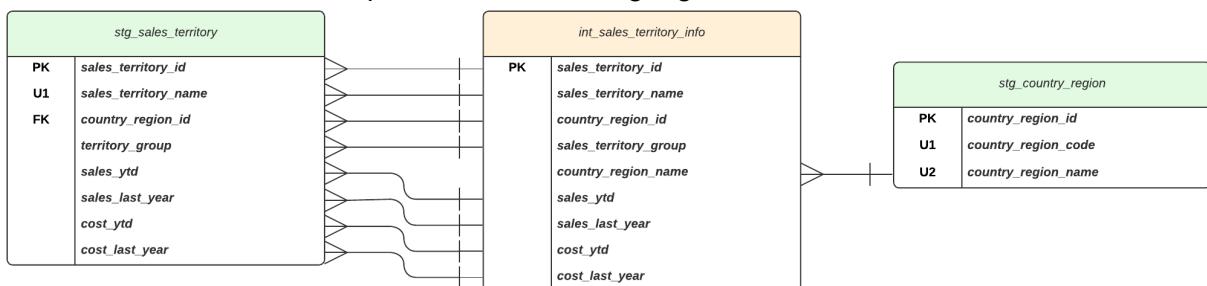
- **int_product_info**: Usada para construir a dimensão **dim_product**, agregando informações de produtos, incluindo categorias e modelos.



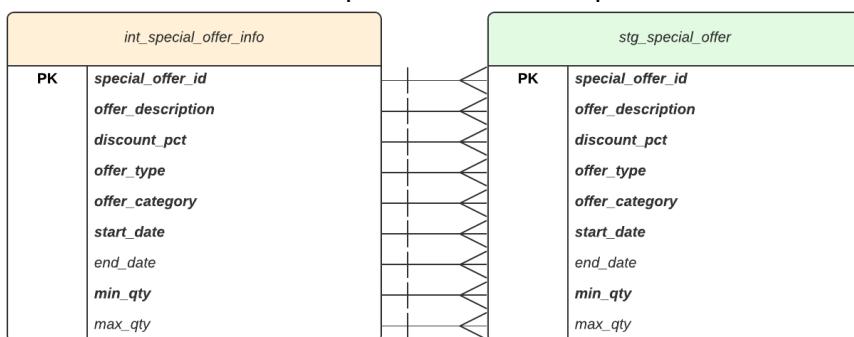
- **int_sales_person_info**: Construiu a dimensão **dim_sales_person**, associando dados de vendedores com suas metas e territórios de vendas.



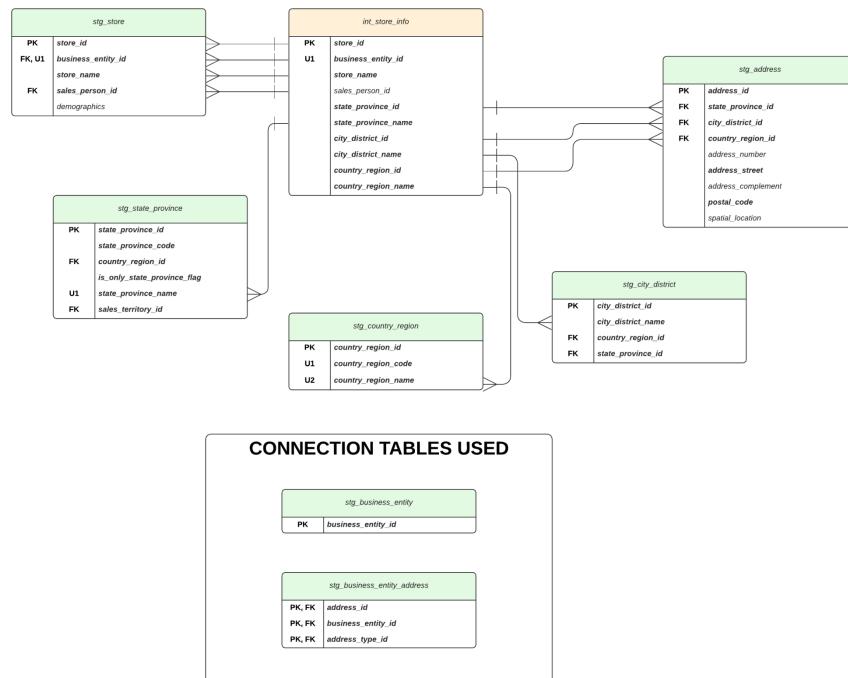
- **int_sales_territory_info**: Populou a dimensão **dim_sales_territory**, organizando os dados de território e permitindo análises geográficas.



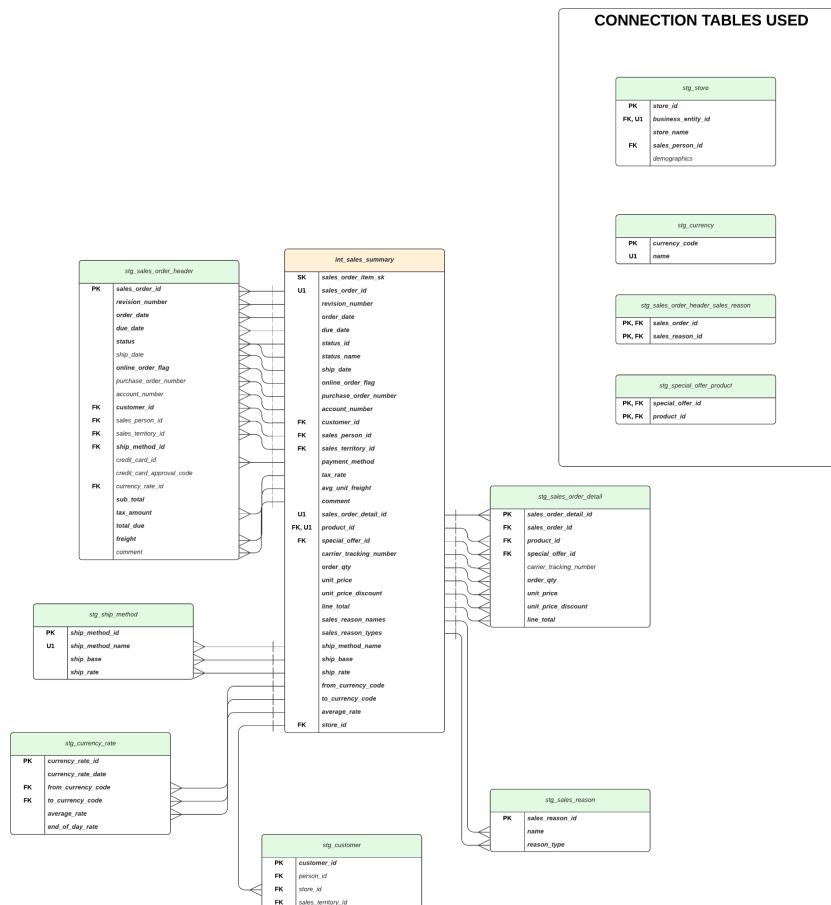
- **int_special_offer_info**: Agregou dados para a dimensão **dim_special_offer**, facilitando o rastreamento do impacto de ofertas especiais nas vendas.



- **int_store_info**: Construiu a dimensão **dim_store**, integrando dados de localização das lojas e relacionando com seus territórios de vendas.

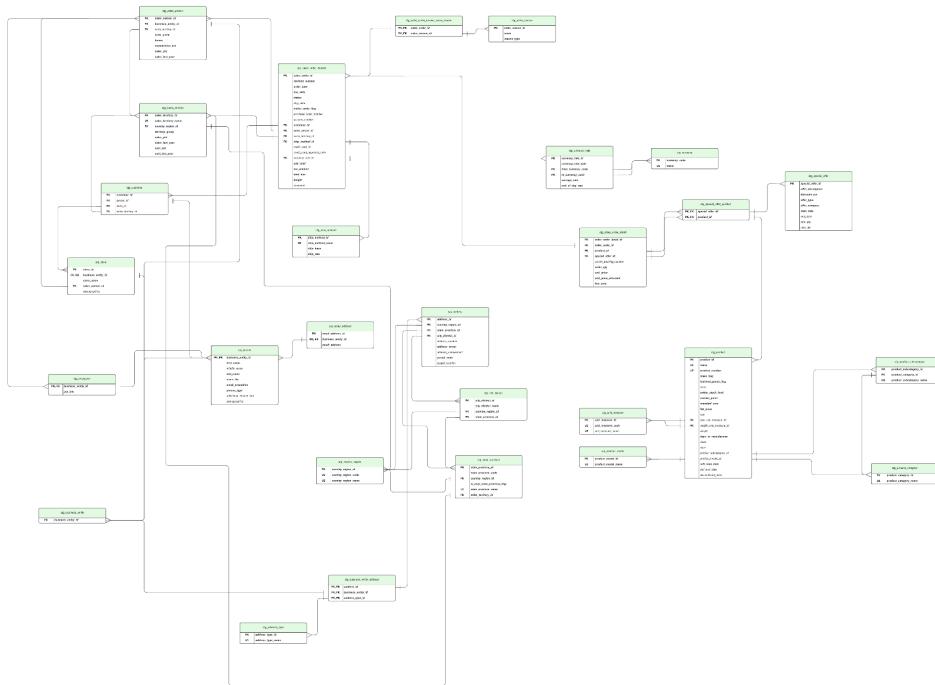


- **int_sales_summary**: Construiu a tabela fato, integrando todos os dados com relação a itens dentro de um pedido.

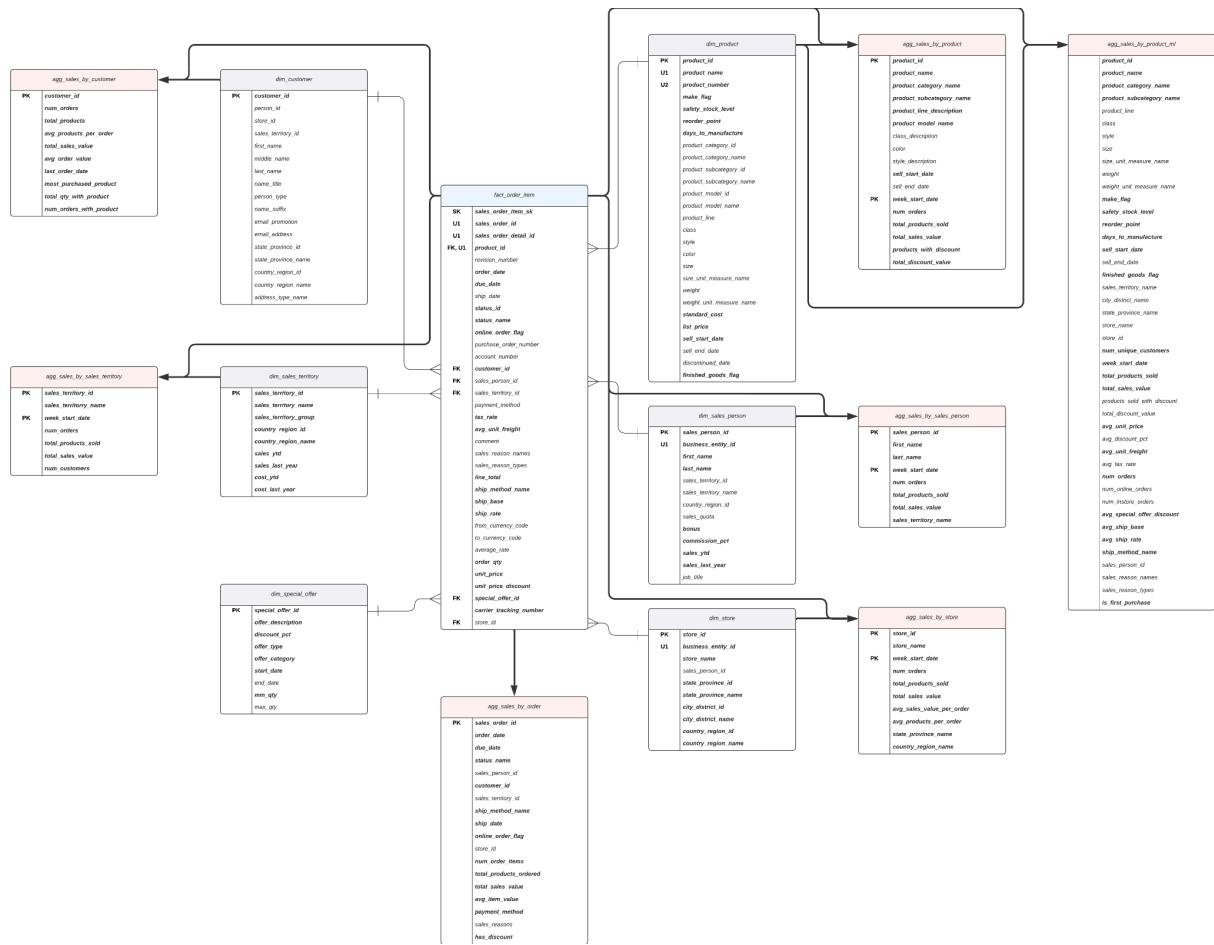


4.3.2 Tabelas Staging (stg)

As tabelas staging foram extraídas diretamente dos seeds do dbt e armazenadas no Google BigQuery. Essas tabelas brutas incluem dados como **stg_customer**, **stg_product**, **stg_sales_order_header**, **stg_sales_order_detail**, entre outras. Elas fornecem os dados iniciais que passam pelo processo de transformação para alimentar o modelo final.



4.4 Relações e Agregações



O modelo **star schema** permite uma eficiente exploração dos dados com consultas otimizadas. A tabela **fact_order_item** contém as chaves estrangeiras para cada uma das dimensões, permitindo agregações como:

1. **agg_sales_by_customer:** Esta tabela agrupa métricas de vendas por cliente, como o número de produtos comprados, o valor total gasto, o número de pedidos realizados e a média de produtos por pedido. Com essas informações, a empresa pode identificar clientes de maior valor, comportamentos de compra recorrentes e oportunidades de personalização de campanhas.
2. **agg_sales_by_sales_territory:** A tabela consolida os dados de vendas por território de vendas, permitindo que a empresa avalie o desempenho por região. Métricas como o total de vendas, número de produtos vendidos e o número de clientes por território são cruciais para identificar regiões mais lucrativas e direcionar estratégias de expansão ou melhoria em regiões subdesenvolvidas.
3. **agg_sales_by_product:** Esta tabela é responsável por agrregar as métricas de vendas de produtos, como o total vendido, o valor de descontos aplicados, a receita total e o estoque médio disponível. É ideal para identificar os produtos mais vendidos e entender o impacto de descontos nas vendas de cada item.

4. **agg_sales_by_product_ml**: Voltada para análises de machine learning, essa tabela oferece métricas detalhadas e estruturadas para alimentar algoritmos de ML, como previsão de demanda ou recomendação de produtos. Inclui dados como tendências de vendas ao longo do tempo, categorias de produto, e outras variáveis essenciais para modelagem preditiva.
5. **agg_sales_by_store**: Essa tabela agrupa dados de vendas por loja, incluindo o número total de vendas, quantidade de produtos vendidos e o ticket médio por cliente. É uma ferramenta essencial para gerentes de lojas que desejam monitorar e melhorar o desempenho local.
6. **agg_sales_by_sales_person**: Aqui são agregadas as métricas de desempenho por vendedor, facilitando a análise de metas alcançadas, produtos vendidos e comissões geradas. Essa visão detalhada auxilia a empresa na gestão de incentivos e na identificação de vendedores de alta performance ou aqueles que precisam de suporte adicional.
7. **agg_sales_by_order**: A tabela de vendas agregadas por pedido oferece uma visão consolidada de cada transação, mostrando o número de itens, valor total do pedido, descontos aplicados, e métodos de pagamento utilizados. Isso permite uma análise precisa do comportamento de compra, fornecendo insights sobre preferências de clientes e otimizando o processo de checkout.

4.5 Escolha do Data Warehouse: Google BigQuery

Para o projeto de modernização da infraestrutura de dados da **Adventure Works**, foi escolhida a plataforma **Google BigQuery** como o **Data Warehouse**. A decisão foi baseada em fatores que equilibram o custo, a facilidade de uso e a escalabilidade, levando em consideração as necessidades atuais e futuras da empresa. Abaixo, explicamos por que essa escolha foi feita, os benefícios para o negócio e os custos envolvidos.

4.5.1 Por que Google BigQuery?

Google BigQuery foi a melhor opção para atender às necessidades de crescimento e transformação da **Adventure Works**, pelos seguintes motivos:

- **Escalabilidade**: A empresa está em franco crescimento, com mais de **31.000 pedidos** e uma base de **20.000 clientes**. BigQuery é uma solução que acompanha esse crescimento sem a necessidade de investimentos em servidores físicos. Como é uma plataforma que se ajusta automaticamente ao volume de dados e consultas, ela garante que a **Adventure Works** possa crescer sem interrupções ou limitações técnicas.
- **Rapidez nas Análises**: BigQuery é conhecido por sua velocidade ao processar grandes volumes de dados. Isso significa que a equipe de vendas e gestão da **Adventure Works** pode obter insights em tempo real sobre as vendas, comportamento dos clientes e desempenho das campanhas promocionais. Esse benefício é essencial para a tomada de decisões rápidas e baseadas em dados.

- **Integração com Ferramentas de Negócio:** Como o BigQuery se integra facilmente com ferramentas como **Power BI** (que foi usado para os dashboards de vendas e operacionais), isso reduz a complexidade e o tempo de implementação. Ou seja, a empresa pode começar a extrair valor de seus dados em menos tempo, sem a necessidade de grandes ajustes ou treinamentos.
- **Segurança:** A segurança dos dados é uma prioridade, especialmente considerando que foram identificados problemas de exposição de dados sensíveis, como informações de clientes. O BigQuery oferece segurança de nível empresarial, criptografando os dados e garantindo que apenas pessoas autorizadas possam acessar informações críticas. Isso ajuda a **Adventure Works** a proteger sua reputação e a evitar penalidades ligadas a violações de privacidade.

4.5.2 Custos Envolvidos com BigQuery

O modelo de preços do **Google BigQuery** é bastante flexível, permitindo que a **Adventure Works** pague apenas pelo que realmente usa. Isso é especialmente útil para uma empresa em expansão, pois evita altos investimentos iniciais e permite ajustar os custos conforme o volume de dados cresce. Os principais componentes de custo incluem:

- **Armazenamento de Dados:** A empresa paga pelo espaço de armazenamento utilizado, com os primeiros **10 GB por mês** sendo gratuitos. Atualmente, a **Adventure Works** usa entre **50 e 100 GB** de armazenamento, o que resulta em custos razoáveis e controlados.
- **Consultas:** As consultas no BigQuery são cobradas com base no volume de dados processados em cada análise. Isso significa que a empresa só paga quando efetivamente utiliza o sistema para gerar relatórios ou obter insights. Esse modelo permite uma gestão eficiente dos custos, ajustando-se ao volume de dados analisado, sem surpresas desagradáveis.

4.5.3 Benefícios de Custo e Operação

Em termos de **custo-benefício**, o **Google BigQuery** permite que a **Adventure Works** tenha uma solução de ponta para **armazenamento e análise de dados** sem precisar investir em uma infraestrutura física complexa, como servidores ou grandes equipes de TI. Os custos são previsíveis e diretamente proporcionais ao uso, o que torna fácil ajustar o orçamento à medida que o volume de dados e análises cresce.

Além disso, a empresa economiza tempo e recursos ao centralizar todos os dados em um único lugar, permitindo que as equipes de vendas, marketing e gestão tenham acesso instantâneo às informações necessárias para a tomada de decisões estratégicas, o que contribui para **aumentar a eficiência operacional e otimizar os resultados de negócios**.

5. Apresentação dos Dados

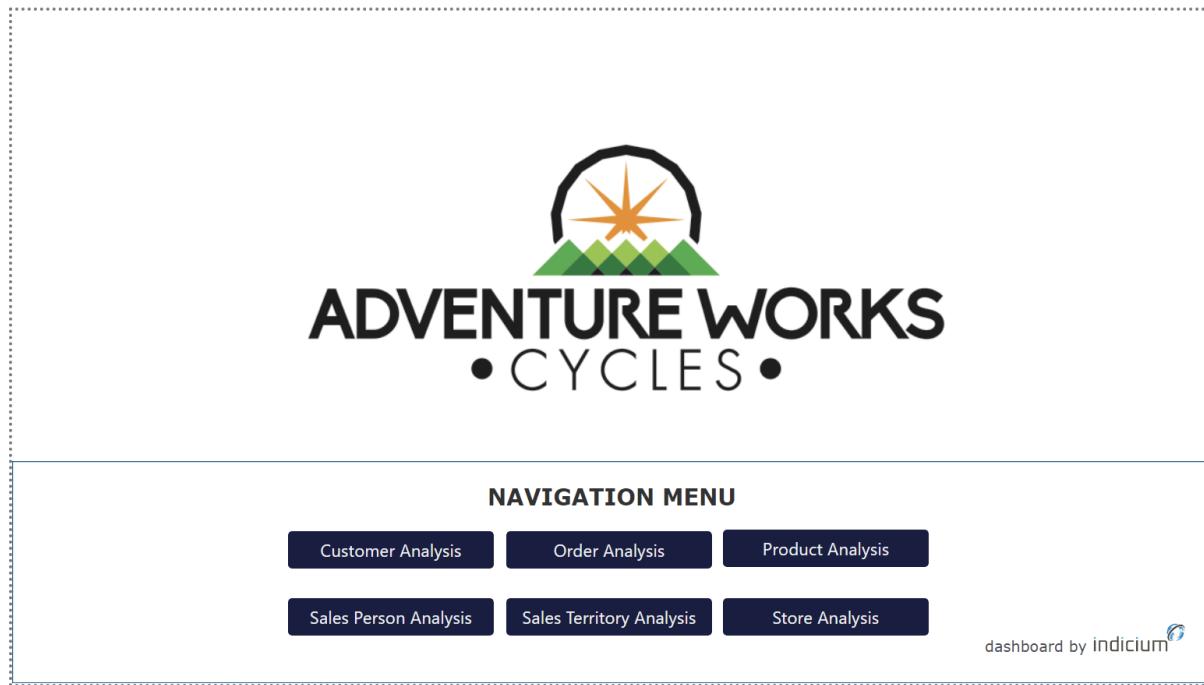
5.1 Relatório Gerencial

O **Relatório Gerencial** criado para o **CEO Carlos Silveira** da **Adventure Works** foi projetado para fornecer uma visão de alto nível dos principais KPIs da empresa, facilitando a tomada de decisões estratégicas. As principais métricas apresentadas incluem:

- **Total de Pedidos:** Indicando o número total de transações realizadas ao longo do período.
- **Valor Total das Vendas:** O valor total gerado pelas vendas, essencial para monitorar o crescimento das receitas.
- **Ticket Médio:** Um cálculo do valor médio por pedido, importante para entender o comportamento de compra dos clientes.

O relatório também permite que o CEO visualize o **desempenho das vendas por trimestre**, facilitando o acompanhamento de tendências ao longo do tempo e fornecendo insights sobre sazonalidades nas vendas.

5.2 Dashboard Operacional



O **Dashboard Operacional** foi criado para fornecer uma visão mais granular das operações de vendas e está dividido em várias seções, cada uma com foco em uma área específica. As principais funcionalidades incluem:

5.2.1 Análise por Loja

Este dashboard permite à equipe de vendas monitorar o desempenho das lojas, com informações detalhadas sobre:

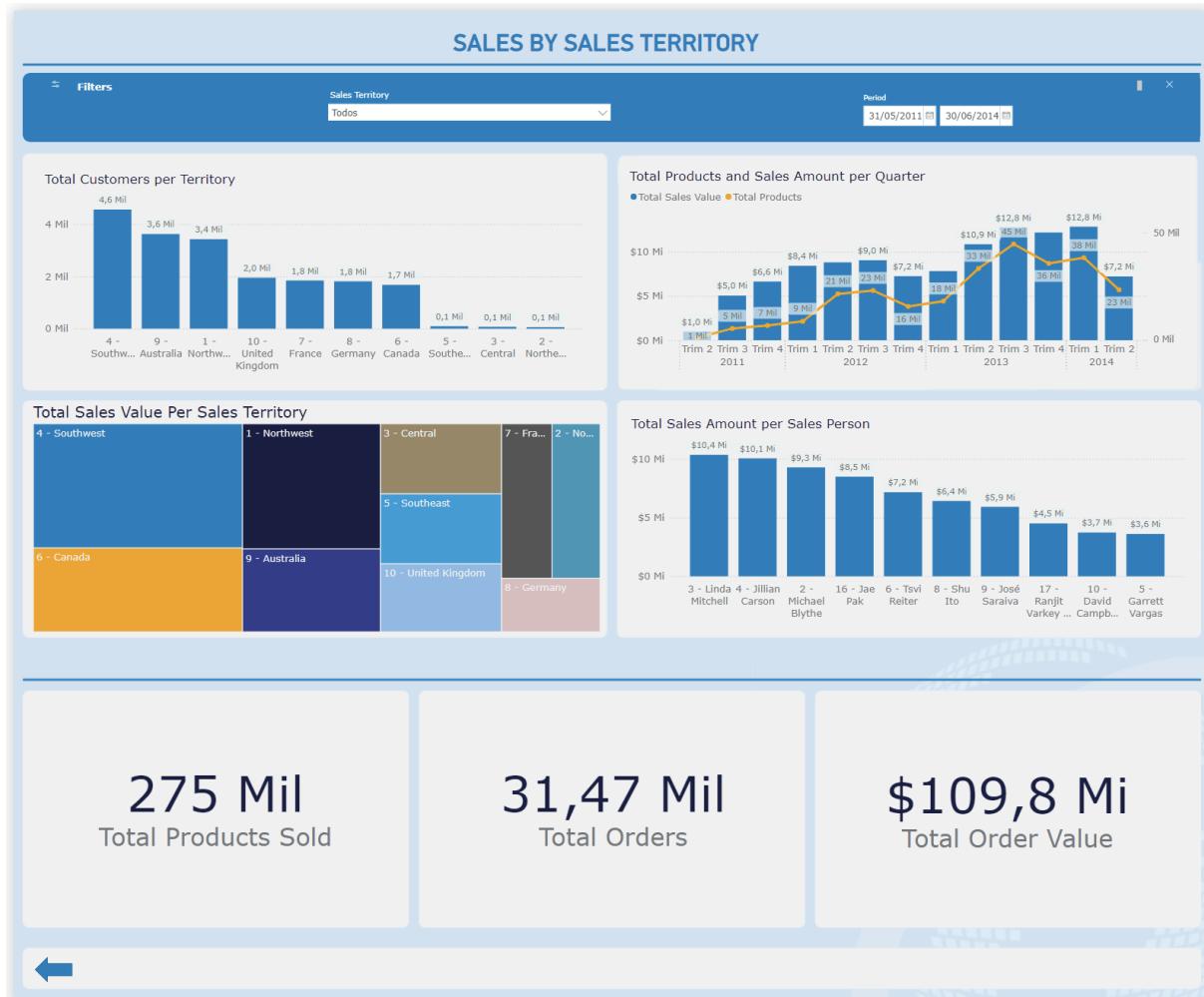
- **Total de Produtos Vendidos por Ano e Estado/Província:** Gráficos que mostram o volume de produtos vendidos ao longo dos anos e a distribuição geográfica das vendas.
- **Valor Total das Vendas por Trimestre:** Um gráfico de linha que destaca a performance trimestral, ajudando a identificar períodos de pico ou queda nas vendas.
- **Indicadores Chave (KPIs):** Como o número total de pedidos, valor total das vendas e ticket médio.



5.2.2 Análise por Território de Vendas

Esse dashboard foca nas vendas por território, oferecendo insights sobre:

- **Cientes por Território:** Um gráfico que mostra o número de clientes em cada território.
- **Vendas por Território e por Pessoa:** Um gráfico de barras que detalha tanto as vendas por território quanto as vendas por vendedor, permitindo identificar os melhores desempenhos.
- **Distribuição do Valor das Vendas:** Um mapa de árvore que exibe a distribuição do valor das vendas por território.



5.2.3 Análise de Clientes

Este dashboard se concentra no comportamento dos clientes, com os seguintes insights:

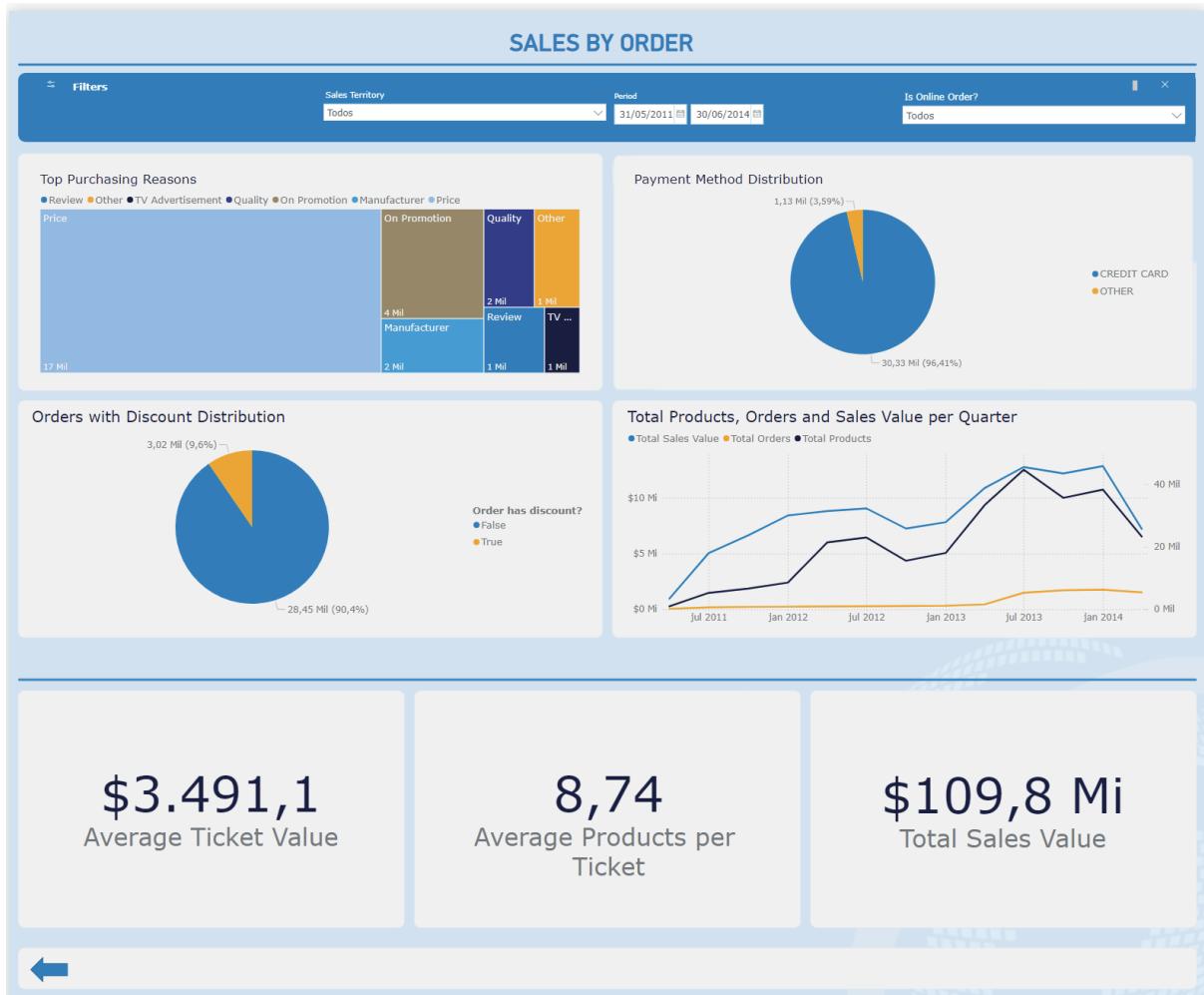
- **Número Total de Pedidos:** Um indicador claro de quantos pedidos foram feitos por todos os clientes.
- **Principais Clientes:** Gráficos que mostram os clientes que mais compraram, com base no valor total dos pedidos.
- **Ticket Médio e Produtos por Pedido:** Indicadores que mostram o número médio de produtos por pedido e o valor médio dos pedidos.



5.2.4 Análise de Pedidos

O dashboard de análise de pedidos fornece uma visão detalhada sobre os motivos de compra e a distribuição dos métodos de pagamento:

- **Motivos de Compra:** Um gráfico que detalha os principais motivos pelos quais os clientes realizaram compras, como preço, qualidade ou promoções.
- **Distribuição por Método de Pagamento:** Um gráfico de pizza que exibe como os clientes pagaram suas compras, seja por cartão de crédito ou outros métodos.
- **Distribuição de Descontos:** Um gráfico que indica o percentual de pedidos que receberam descontos e a sua relação com o volume total de vendas.



5.2.5 Análise de Produtos

Essa seção do dashboard foca no desempenho dos produtos vendidos:

- **Valor de Vendas por Categoria de Produto:** Um gráfico que mostra o valor total das vendas por categoria de produto, permitindo à equipe de gestão entender quais categorias estão gerando mais receita.
- **Histórico de Vendas de Produtos:** Um gráfico de linha que exibe o histórico de vendas dos produtos mais vendidos, permitindo monitorar as tendências de vendas ao longo do tempo.



5.2.6 Análise por Vendedor

O desempenho de vendas por vendedor é monitorado neste dashboard, que inclui:

- **Vendas por Trimestre:** Um gráfico de barras mostrando o valor de vendas por trimestre, permitindo monitorar o desempenho ao longo do ano.
- **Vendas por Vendedor:** Um gráfico de barras que destaca os vendedores com maior volume de vendas, ajudando a equipe de gestão a reconhecer os principais contribuintes para o sucesso da empresa.



6. Planejamento do Projeto de Dados e Ferramentas Utilizadas

O projeto de dados da **Adventure Works** foi estruturado em várias fases que cobrem desde a ingestão e transformação dos dados até a análise preditiva e o desenvolvimento de dashboards para uso estratégico e operacional. A seguir, apresentamos uma descrição detalhada do cronograma, ferramentas, custos e as previsões realizadas no contexto do projeto.

6.1 Fases do Projeto de Dados

Abaixo estão descritas as principais fases do projeto.

6.1.1 Descoberta e Diagnóstico Inicial (2 semanas)

Objetivo: Identificar as principais questões de negócios e mapear as fontes de dados da Adventure Works.

Atividades:

- Reuniões com stakeholders (diretores, gerentes).
- Identificação das necessidades de dados e problemas.
- Avaliação da estrutura atual dos dados.

Entregáveis: Documento de diagnóstico inicial e mapeamento das fontes de dados.

6.1.2 Definição da Arquitetura e Ferramentas (1 semana)

Objetivo: Escolher as ferramentas e definir a arquitetura técnica necessária para o projeto.

Atividades:

- Escolha de ferramentas como **Google BigQuery**, **dbt** e **Power BI**.
- Definição da infraestrutura e estimativa de custos.

Entregáveis: Documento de arquitetura e planejamento técnico.

6.1.3 Ingestão de Dados e Preparação (2 semanas)

Objetivo: Carregar os dados no data warehouse e prepará-los para transformações.

Atividades:

- Extração e carga dos dados no **Google BigQuery**.
- Criação das tabelas **staging** usando **dbt**.
- Validação inicial dos dados.

Entregáveis: Dados carregados no BigQuery, tabelas staging criadas.

6.1.4 Transformação e Limpeza de Dados (3 semanas)

Objetivo: Transformar os dados brutos em dados prontos para análise.

Atividades:

- Modelagem de dados para criação de tabelas de fatos e dimensões (**star schema**).
- Aplicação de transformações para corrigir inconsistências e remover duplicatas.
- Testes de dados e validação.

Entregáveis: Tabelas de fatos e dimensões prontas para análise, documentação e validação de dados.

6.1.5 Desenvolvimento de Dashboards e Relatórios (3 semanas)

Objetivo: Desenvolver dashboards interativos para gerência e operação.

Atividades:

- Criação de dashboards no **Power BI** para análise estratégica e operacional.
- Configuração de filtros dinâmicos por loja, território e produto.
- Testes de usabilidade e desempenho.

Entregáveis: Dashboards interativos, relatórios agregados.

6.1.6 Análise Preditiva e Modelos de Regressão (3 semanas)

Objetivo: Prever a demanda futura de produtos e ajustar a distribuição de matéria-prima, utilizando técnicas de machine learning e regressão.

Atividades:

- Aplicação de modelos de previsão para estimar a demanda dos próximos 3 meses em cada loja.
- Análise da **sazonalidade** de determinados produtos para ajustar a produção.
- Avaliação de diferentes **modelos de regressão** (linear e log-log) para comparar a performance.
- Cálculo da demanda de **zippers** para o novo fornecedor de luvas.

- Análise do crescimento de demanda em centros de distribuição (EUA vs. países estrangeiros).
- **Métricas de Avaliação:** Utilizamos **MAE (Mean Absolute Error)** e **RMSE (Root Mean Square Error)** para avaliar os modelos preditivos.

Entregáveis:

- Previsões de demanda por loja e sazonalidade de produtos.
- Relatório com a comparação dos modelos de regressão e justificação do modelo escolhido.
- Estimativa de demanda de matéria-prima (zippers).
- Relatório de crescimento de demanda por centros de distribuição.

6.1.7 Testes e Validação (2 semanas)

Objetivo: Garantir que os dados, dashboards e previsões estejam corretos e alinhados aos objetivos do negócio.

Atividades:

- Validação com stakeholders e ajuste das métricas.
- Testes de performance no Google BigQuery e dashboards no Power BI.

Entregáveis: Dashboards validados, previsões revisadas e correções aplicadas.

6.1.8 Treinamento e Documentação (1 semana)

Objetivo: Treinar a equipe e fornecer documentação detalhada.

Atividades:

- Sessões de treinamento para gerência e equipe operacional.
- Documentação de uso dos dashboards e dos modelos preditivos.

Entregáveis: Documentação e treinamento da equipe.

6.1.9 Entrega e Lançamento do Projeto (1 semana)

Objetivo: Entregar o projeto finalizado e garantir o uso correto dos dashboards e previsões.

Atividades:

- Apresentação final do projeto.
- Acompanhamento pós-lançamento.

Entregáveis: Sistema de dados em produção e dashboards em uso.

6.2 Ferramentas Utilizadas

Para garantir o sucesso do projeto, foram utilizadas diversas ferramentas, cada uma com seu papel específico no fluxo de trabalho:

6.2.1 Google BigQuery (Data Warehouse)

Função: Armazenamento e processamento de dados de grande escala.

Benefícios: Alta escalabilidade, integração nativa com o Power BI e suporte a consultas SQL de alto desempenho.

Custos:

- **Armazenamento:** Aproximadamente \$0.02 por GB ao mês.
- **Consultas:** Aproximadamente \$5 por TB de dados processados.

6.2.2 dbt (Data Build Tool)

Função: Transformação de dados e criação de modelos de dados (star schema).

Benefícios: Facilita a gestão de transformações, documentação e testes de dados.

Custos: Uso local, sem custos diretos além da infraestrutura.

6.2.3 Power BI (Business Intelligence)

Função: Criação de dashboards e visualizações interativas.

Benefícios: Relatórios dinâmicos, integração com BigQuery, facilidade de compartilhamento.

Custos:

- **Power BI Pro:** \$9.99 por usuário/mês.
- **Power BI Premium:** Dependendo do volume de dados e usuários.

6.2.4 Python (Modelos de Machine Learning)

Função: Aplicação de modelos de previsão de demanda e regressão.

Benefícios: Flexibilidade para construir e ajustar modelos, uso de bibliotecas como `scikit-learn` e `statsmodels`.

Custos: Sem custo direto, exceto tempo de desenvolvimento.

6.2.5 Google Cloud Platform (Infraestrutura)

Função: Infraestrutura de suporte (Compute Engine e Storage).

Custos:

- **Compute Engine:** Usado para rodar scripts e processar dados.
- **Storage:** Backup de dados e scripts do projeto.

6.2.6 GitHub (Controle de Versão)

Função: Controle de versão e armazenamento de código e documentação.

Custos: Plano gratuito, sem custos adicionais.

6.3 Estimativa de Custos Totais

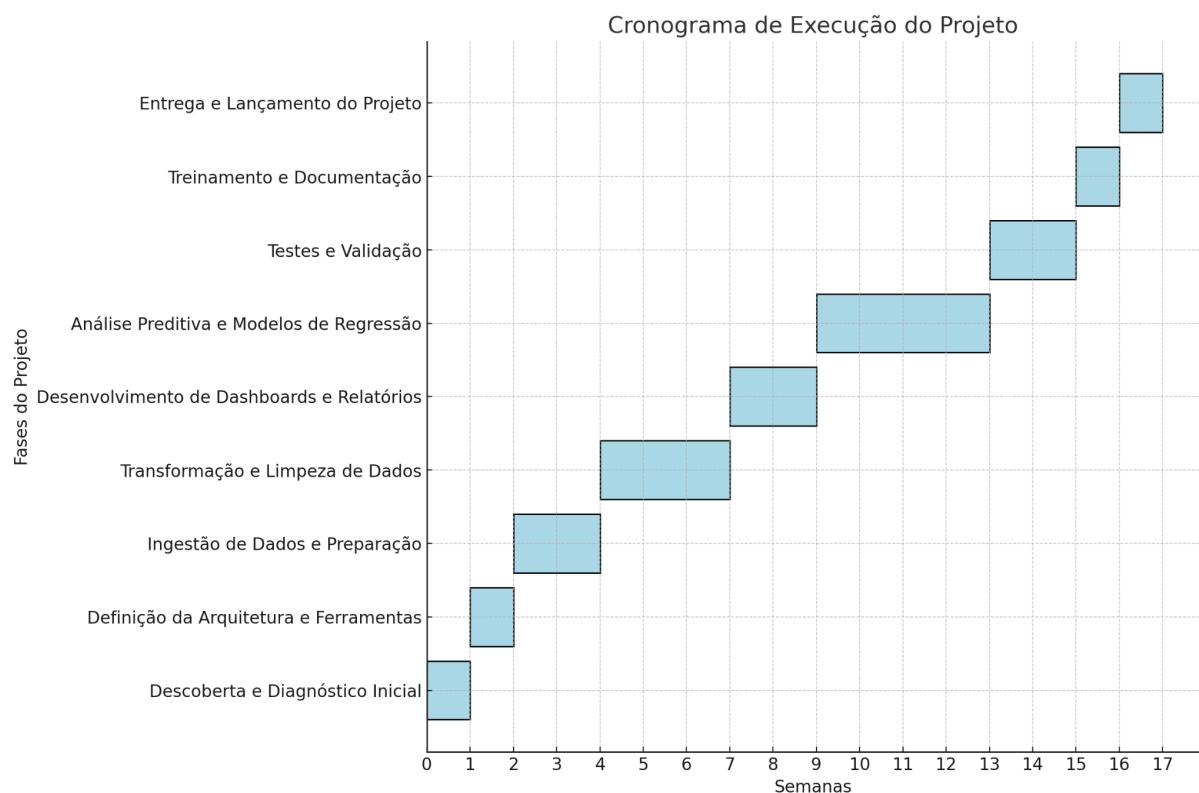
Os custos variam conforme o uso da infraestrutura e das ferramentas. Abaixo está uma estimativa aproximada dos custos:

Item	Custo Estimado
Google BigQuery	\$300 - \$500/mês dependendo do uso
Power BI (5 usuários)	\$49.95/mês
Infraestrutura GCP (Compute + Storage)	\$200/mês
dbt (versão local)	Sem custo direto

Item	Custo Estimado
GitHub (plano gratuito)	\$0
Modelo de Machine Learning (Python)	Sem custo direto

6.4 Considerações sobre o Tempo de Execução

Com base no cronograma ajustado, o projeto terá duração estimada entre **15 e 18 semanas**, como mostrado no diagrama abaixo. Cada etapa e sua duração pode ser vista nesse diagrama.



7. Previsões e Modelos Utilizados

Nesta seção, vamos detalhar o processo de previsão de demanda e a estimativa de matéria-prima, explicando passo a passo como as previsões foram geradas e os modelos escolhidos. O foco será nas abordagens usadas, incluindo séries temporais e algoritmos de aprendizado de máquina, como Random Forest e XGBoost, e nas ferramentas utilizadas para cada etapa do processo.

7.1 Previsão de Demanda para os Próximos 3 Meses

7.1.1 Estruturação dos Dados

Antes de gerar as previsões, foi necessário preparar os dados. O processo incluiu as seguintes etapas:

1. Limpeza de Dados:

- Foram removidos valores nulos e outliers dos dados históricos de vendas para garantir a qualidade das previsões.

2. Agregação dos Dados:

- Não havia dados suficientes por **store_id**. Portanto, os dados foram agrupados por **território de vendas (sales_territory_name)**, agrupando todos os produtos em suas respectivas regiões.

3. Divisão em Conjuntos de Treinamento e Teste:

- O histórico de vendas foi dividido em dois conjuntos: **treinamento** (dados passados usados para treinar os modelos) e **teste** (dados mais recentes usados para validar a performance dos modelos).

7.1.2 Previsões com Séries Temporais

O modelo de **séries temporais** foi o primeiro a ser implementado. Esse modelo foi escolhido devido à sua capacidade de capturar tendências e padrões sazonais com base nos dados históricos.

- **Ferramenta utilizada:** A biblioteca **statsmodels** foi usada para implementar o modelo de séries temporais (ARIMA/SARIMA), permitindo a inclusão de padrões sazonais nas previsões.
- **Técnica utilizada:** O modelo capturou as seguintes características:
 - **Tendência:** Direção geral das vendas ao longo do tempo (aumento ou queda).
 - **Sazonalidade:** Picos e vales nas vendas que ocorrem de forma cíclica (por exemplo, em épocas de feriados ou promoções anuais).
- **Resultado da previsão (Exemplo):**
 - Produto: **Sport-100 Helmet, Black** (ID: 708)
 - Território: **Southwest**
 - Previsão para os próximos 3 meses em semanas:

Semana	Quantidade
Semana 1	6.56
Semana 2	5.62
Semana 3	1.57

Semana	Quantidade
Semana 4	1.05
Semana 5	4.15
Semana 6	4.41
Semana 7	1.83
Semana 8	15.44
Semana 9	2.75
Semana 10	4.22
Semana 11	3.50
Semana 12	1.22
Semana 13	15.01

7.1.3 Previsões com Modelos de Aprendizado de Máquina

Além das séries temporais, foram aplicados dois modelos de **aprendizado de máquina: Random Forest e XGBoost**. Esses modelos foram escolhidos por sua capacidade de lidar com grandes volumes de dados e por sua robustez em prever dados complexos, considerando múltiplas variáveis de entrada.

- **Random Forest:**

- Modelo de árvore de decisão que combina diversas árvores para reduzir a variância e melhorar a precisão das previsões.
- **Ferramenta Utilizada:** Biblioteca **scikit-learn**.
- O modelo foi treinado com dados de vendas passadas, considerando variáveis como produto, território, e características dos produtos.

- **XGBoost:**

- Modelo de **Boosting** que combina árvores de decisão de forma sequencial para otimizar a precisão.
- **Ferramenta Utilizada:** Biblioteca **XGBoost**.
- O XGBoost se destacou por fornecer melhores resultados em termos de precisão de previsão, sendo o modelo escolhido para a etapa final.

Comparação dos Modelos: Ambos os modelos de aprendizado de máquina foram avaliados com base em métricas como **RMSE** (Root Mean Square Error) e **MAE** (Mean Absolute Error). O **XGBoost** apresentou melhor desempenho geral, especialmente em previsões de longo prazo e para produtos com maior volume de vendas.

7.2 Estimativa de Matéria-Prima: Previsão de Zíperes

7.2.1 Definição do Problema

Um novo fornecedor de luvas foi contratado para fornecer zíperes, e era necessário prever a quantidade de zíperes que precisaria ser solicitada para atender à demanda dos próximos três meses. Como cada par de luvas requer **dois zíperes**, o desafio era prever quantos pares de luvas seriam vendidos para então calcular a quantidade de zíperes necessária.

7.2.2 Modelos Usados

Os mesmos modelos de aprendizado de máquina (Random Forest e XGBoost) foram aplicados para a previsão da demanda de luvas, e a quantidade de zíperes foi calculada a partir dessas previsões.

- **Resultado Final:** O modelo **XGBoost** foi escolhido para essa previsão, e estimou-se que aproximadamente **142.973 zíperes** seriam necessários para atender à demanda nos próximos três meses.

7.3 Resultados por Território

Além das previsões por produto, uma análise foi feita para determinar quais territórios apresentaram maior crescimento de demanda durante o período de previsão. A **região Southwest** foi a que demonstrou o maior crescimento em relação aos outros territórios. Isso permitirá à empresa ajustar suas estratégias de distribuição de maneira mais eficiente, priorizando as regiões de maior crescimento.

7.4 Conclusão e Melhorias Implementadas

Com base nas previsões realizadas, foi possível identificar padrões de sazonalidade e ajustar as operações de distribuição de produtos, bem como melhorar a previsão de compra de matéria-prima. O uso de modelos de aprendizado de máquina, como **Random Forest** e **XGBoost**, mostrou-se superior ao modelo de séries temporais para as previsões de demanda.

- **Modelos Utilizados:** Séries Temporais, Random Forest e XGBoost.
- **Ferramentas:** scikit-learn, XGBoost, statsmodels.
- **Previsão de Zíperes:** Aproximadamente **142.973 zíperes** serão necessários nos próximos três meses.
- **Crescimento por Território:** A região **Southwest** apresentou o maior crescimento de demanda.

Essa abordagem permitiu otimizar as operações logísticas, garantindo que a demanda seja atendida com eficiência e que os recursos de produção sejam bem gerenciados, minimizando custos e evitando o excesso de inventário.

7.5 Localização dos Resultados das Previsões

Todas as previsões detalhadas acima, incluindo as previsões de demanda para os próximos três meses, a estimativa de zíperes, e os resultados de cada território, estão documentadas

no arquivo **demand-forecast.csv** e nos notebooks Jupyter. Esses arquivos estão localizados na **pasta notebooks** do repositório Git utilizado neste projeto.

Os dados incluem as previsões feitas pelos modelos de séries temporais, oferecendo uma visão clara da previsão de vendas para diferentes produtos em diversos territórios.

Assim, o arquivo **demand-forecast.csv** centraliza todos os resultados utilizados nas análises de séries temporais, sendo uma importante referência para futuras atualizações ou ajustes nas previsões.

8. Resultados Concretos e Impacto no Negócio

O projeto de modernização da infraestrutura de dados da **Adventure Works (AW)** trouxe benefícios mensuráveis em várias áreas críticas do negócio, resolvendo problemas específicos e criando novas oportunidades de crescimento. Com a implementação de uma plataforma de dados moderna, foi possível atingir objetivos claros que diretamente influenciam as operações, vendas, e a tomada de decisões estratégicas.

8.1 Melhorias na Previsão de Demanda

Uma das principais dores enfrentadas pela **AW** era a ineficiência no planejamento de demanda. O modelo anterior, baseado em médias móveis de três meses no Excel, resultava em erros frequentes de previsão, gerando problemas como falta de estoque ou excesso de produtos em inventário.

Com a introdução de modelos preditivos baseados em **aprendizado de máquina (Random Forest e XGBoost)**, foi possível prever a demanda com muito mais precisão. Isso permitiu:

- **Redução de Estoques Excedentes:** Com previsões mais acuradas, a AW pôde reduzir os níveis de estoque, evitando o acúmulo de produtos que não estavam sendo vendidos, resultando em economias significativas em armazenamento e logística.
- **Aumento na Disponibilidade de Produtos:** Ao prever corretamente os picos de demanda, a AW pôde garantir que os produtos populares estivessem disponíveis nos momentos certos, melhorando a satisfação do cliente e maximizando as vendas.

Por exemplo, a previsão para o produto **Sport-100 Helmet, Black** no território **Southwest** permitiu que a equipe de planejamento ajustasse a produção e distribuição de acordo com as necessidades do mercado, o que levou a um aumento direto nas vendas nesse território.

8.2 Melhoria no Planejamento e Distribuição de Matéria-Prima

A nova infraestrutura de dados também foi crucial para otimizar a aquisição de matéria-prima. Com a previsão da demanda de **luvas** e a necessidade de **zíperes** (zíperes por par de luvas), a **AW** pôde ajustar sua cadeia de suprimentos e evitar tanto a falta quanto o excesso de materiais.

- **Estimativa de 142.973 zippers** necessários para os próximos três meses foi calculada com base nas previsões de vendas das luvas, garantindo uma produção eficiente e sem desperdícios.
- **Melhoria na Relação com Fornecedores:** O planejamento preciso da quantidade de zippers possibilitou uma negociação mais eficaz com os fornecedores, resultando em melhores condições de preço e prazos de entrega.

Essa eficiência no planejamento logístico não só reduziu os custos, mas também permitiu que a AW mantivesse a continuidade da produção sem interrupções, especialmente durante picos sazonais.

8.3 Aumento na Eficiência Comercial e Estratégia de Vendas

O projeto trouxe também uma transformação na forma como a equipe comercial da **AW** gerencia suas operações e identifica oportunidades de crescimento. Com as ferramentas de **Business Intelligence (BI)** implementadas, foi possível gerar dashboards que oferecem insights detalhados sobre o comportamento dos clientes, a performance de vendas, e a distribuição de produtos.

Os principais ganhos foram:

- **Identificação de Territórios com Maior Potencial de Crescimento:** As análises mostraram que o território **Southwest** teve o maior crescimento de demanda nos últimos três meses, permitindo que a equipe de vendas concentrasse esforços nessa região para maximizar as oportunidades.
- **Melhora no Desempenho de Vendedores:** Com o monitoramento detalhado de KPIs por vendedor e por território, a AW conseguiu identificar os vendedores com melhor desempenho, ajustando metas e oferecendo suporte personalizado para aqueles que estavam abaixo do esperado.
- **Otimização de Promoções e Ofertas:** Ao utilizar dados preditivos para entender a resposta do mercado a diferentes promoções, a equipe de vendas pôde ajustar suas ofertas e campanhas promocionais, resultando em um **aumento direto nas vendas e na margem de lucro**.

8.4 Redução de Problemas Operacionais e Melhoria na Satisfação do Cliente

A ausência de uma plataforma de dados moderna anteriormente impactava diretamente na experiência do cliente e na eficiência das operações da **AW**. Problemas como a falta de produtos em estoque ou erros no processamento de pedidos resultavam em uma experiência insatisfatória para os clientes.

Com a modernização da infraestrutura de dados e a implementação de dashboards operacionais, a **AW** conseguiu:

- **Reducir o Tempo de Resposta ao Cliente:** Com uma visão em tempo real das vendas, dos estoques e dos pedidos, a equipe operacional pôde processar e responder aos pedidos dos clientes com muito mais agilidade, o que resultou em um **aumento na satisfação do cliente** e na redução de reclamações.

- **Aumento no Ticket Médio:** O uso de insights para personalizar ofertas para clientes de alto valor ajudou a aumentar o ticket médio por pedido, proporcionando ganhos financeiros diretos.

8.5 Melhoria na Organização e Acesso aos Dados

A criação de um **Data Warehouse** baseado no **Google BigQuery** e o uso de **dbt** para transformação de dados permitiram uma organização significativa dos dados da **AW**. Antes, os dados estavam dispersos em 68 tabelas espalhadas por cinco esquemas diferentes (Vendas, Produção, Compras, Recursos Humanos). Essa fragmentação dificultava a extração de informações de valor para os gestores e operadores.

Com a nova arquitetura, a **AW** passou a usufruir de:

- **Centralização de Informações:** A consolidação dos dados em um ambiente único permitiu que todas as equipes tivessem acesso a informações organizadas, facilitando o uso dos dados para tomadas de decisões rápidas e precisas.
- **Facilidade no Acesso aos Dados Tratados:** Ao utilizar o modelo **star schema**, que organiza os dados em tabelas fato e dimensões, ficou mais fácil acessar informações críticas, como total de vendas por região, desempenho de produtos e análises de clientes.
- **Dados Confiaíveis e Consistentes:** A aplicação de **testes de dados e validações automáticas** com o **dbt** garantiu que as informações utilizadas pelos dashboards e relatórios fossem confiáveis, eliminando erros que antes eram recorrentes nas análises.

8.6 Utilidade dos Dashboards para Operações e Gerência

Os dashboards implementados trouxeram **valor imediato** tanto para a gerência quanto para a operação. A diferença na granularidade das informações ofereceu uma **visão estratégica** para os executivos e uma **ferramenta operacional robusta** para os times que lidam diretamente com vendas e clientes.

- **Dashboard Gerencial:** Oferece uma visão de alto nível com **indicadores chave de performance (KPIs)** como total de vendas, ticket médio e tendências trimestrais, permitindo ao **CEO Carlos Silveira** acompanhar o desempenho da empresa e tomar decisões informadas rapidamente.
- **Dashboard Operacional:** Fornece informações detalhadas sobre **vendas por pedido, região, e cliente**, com recortes que permitem à equipe de vendas entender melhor os períodos de pico, ajustar campanhas de marketing e oferecer suporte direcionado aos clientes de maior valor.

Esses dashboards **reduziram significativamente o tempo necessário para gerar relatórios**, facilitando a tomada de decisão em tempo real e permitindo que as equipes se concentrem em estratégias de crescimento, em vez de gastar tempo extra com consolidação e análise manual de dados.

9. Desafios e Riscos Enfrentados no Projeto

9.1 Desafios Técnicos

Um dos maiores desafios enfrentados durante a execução deste projeto foi a integração de dados fragmentados e desatualizados de diversos sistemas que a Adventure Works utiliza. Esses sistemas, ao longo do tempo, geraram dados não padronizados e duplicados, o que dificultou o processo de transformação. **Campos críticos, como dados de clientes, produtos e vendedores, estavam faltando ou corrompidos** em várias tabelas, o que impactava diretamente a capacidade de gerar relatórios e análises precisas.

Solução

Para contornar esses problemas, foi essencial aplicar técnicas de deduplicação e limpeza de dados, além de **adotar um processo rigoroso de validação** com regras definidas para garantir que as informações armazenadas fossem consistentes e confiáveis. O uso de ferramentas como **dbt** permitiu automatizar parte desse processo, assegurando uma governança de dados adequada e uma pipeline de ETL eficiente.

9.2 Desafios de Negócio

Do ponto de vista de negócio, o principal desafio foi demonstrar o **valor tangível de uma infraestrutura moderna de dados** para as partes interessadas que tinham dúvidas, como a **Diretora Comercial, Silvana Teixeira**. Ela acreditava que o valor do projeto não seria superior ao de ações promocionais imediatas e que o retorno sobre o investimento (ROI) poderia ser questionável, especialmente considerando o fracasso de iniciativas passadas, como a implementação de CRM e Web Analytics, que não geraram os resultados esperados.

Solução

A criação de dashboards **personalizados para cada departamento** (gerencial e operacional) foi essencial para endereçar essas preocupações. No caso de Silvana Teixeira, o dashboard operacional forneceu insights detalhados sobre **vendas por região, por cliente, por método de pagamento e por periodicidade**, permitindo que ela enxergasse o impacto direto nas vendas e na segmentação de clientes. Ao demonstrar como os dados fornecem uma visão clara sobre as oportunidades de venda e o comportamento dos clientes, foi possível convencer a diretoria comercial do **potencial de longo prazo** da plataforma de dados.

9.3 Riscos do Projeto

9.3.1 Acesso aos Dados

A **limitação de tempo do analista de TI Gabriel Santos**, que tinha uma agenda restrita, representou um risco de atraso no projeto. Sem um acesso eficiente e contínuo aos dados, o processo de construção da infraestrutura poderia ser seriamente comprometido.

Solução

Uma comunicação coordenada e eficiente foi estabelecida para garantir que o tempo de Gabriel fosse utilizado de maneira produtiva, com reuniões objetivas e tarefas bem definidas para que não houvesse perda de tempo. O planejamento adequado das interações com a equipe de TI ajudou a minimizar este risco.

9.3.2 Escalabilidade e Custos

Com o crescimento acelerado da Adventure Works, havia uma preocupação de que a infraestrutura montada para o projeto não fosse suficientemente escalável ou que os custos associados à manutenção de uma solução em nuvem como o **Google BigQuery** fossem inviáveis no longo prazo.

Solução

Foi escolhida uma arquitetura de dados altamente escalável, utilizando **Google BigQuery**, com uma estratégia de **pagamento sob demanda**, onde a empresa paga apenas pelo que usa. Isso garantiu uma relação custo-benefício vantajosa, onde a Adventure Works não precisa investir em infraestrutura física, e os custos são proporcionais ao volume de dados processado.

9.4 Benefícios de Negócio com a Infraestrutura de Dados e Dashboards

Organização e Acesso a Dados

A construção de um **Data Warehouse robusto** e a organização dos dados em **tabelas-fato e dimensão** permitiram à Adventure Works transformar dados brutos e fragmentados em informações estruturadas e acionáveis. Isso facilitou o acesso a dados de diferentes áreas, como **vendas, clientes e produtos**, em um só lugar. A empresa passou a poder tomar decisões mais informadas e fundamentadas, evitando erros devido a dados inconsistentes ou incompletos.

Facilidade de Acesso às Informações Tratadas

Com a implementação de dashboards em ferramentas como **Power BI**, foi possível fornecer **informações acessíveis em tempo real** tanto para os executivos, como o CEO Carlos Silveira, quanto para as equipes operacionais. Esse acesso centralizado a relatórios e métricas críticas reduziu significativamente o tempo gasto em processos manuais e consultas SQL, além de fornecer uma visão consolidada da **performance da empresa**.

Utilidade dos Dashboards

Os **dashboards gerenciais e operacionais** criados permitiram insights valiosos:

- **Dashboard Operacional:** Permitiu à equipe de vendas **monitorar a performance por loja, território de vendas e cliente**, ajudando a identificar os principais produtos vendidos, regiões com maior demanda e clientes de maior valor. A análise detalhada sobre **métodos de pagamento e motivos de compra** ajudou a afinar as estratégias promocionais e de marketing.

Redução de Riscos e Melhoria da Eficiência Operacional

Com dados de vendas e estoque consolidados e previsões mais precisas para demanda futura, a empresa conseguiu **melhorar o planejamento de produção e otimizar o estoque**, minimizando desperdícios e evitando faltas de produtos. A integração dos dados de diversas fontes permitiu um **monitoramento proativo** e reduziu riscos operacionais, como falhas de abastecimento e acúmulo de estoque.

10. Conclusões

A implementação de uma plataforma de dados moderna na **Adventure Works** trouxe uma transformação significativa em diversos aspectos do negócio, permitindo que a empresa adotasse uma abordagem **data-driven** para a tomada de decisões estratégicas e operacionais.

1. Resolução de Desafios de Negócio

- O projeto abordou diretamente as **dores e preocupações** de diferentes stakeholders, como a necessidade de justificar o investimento para a **Diretora Comercial, Silvana Teixeira**, ao demonstrar o impacto direto dos dados nas vendas e na segmentação de clientes.
- A implementação da infraestrutura de dados ajudou a resolver **problemas operacionais críticos**, como a falta de integração entre sistemas e a baixa qualidade de dados, facilitando a extração de insights precisos e confiáveis para as áreas de **vendas, planejamento de demanda, e estoque**.

2. Transformação em Empresa Data-Driven

- A Adventure Works conseguiu transformar sua abordagem de negócios de uma estrutura transacional limitada para uma estratégia baseada em **dados centralizados**, acessíveis, e altamente organizados. O **data warehouse** e os **dashboards gerenciais** agora fornecem uma visão clara e abrangente do desempenho da empresa, ajudando a identificar **oportunidades de crescimento** e áreas de melhoria.
- Com os dados bem organizados, limpos e validados, a empresa ganhou uma visão holística do seu **pipeline de vendas, comportamento do cliente, e distribuição de produtos**, permitindo **ajustes rápidos** em estratégias de marketing e produção.

3. Impacto dos Dashboards

- O **dashboard gerencial** forneceu insights estratégicos para a liderança da empresa, como o **CEO Carlos Silveira**, ajudando na tomada de decisões fundamentadas e baseadas em indicadores de desempenho claros e acessíveis.
- O **dashboard operacional** permitiu que as equipes de vendas e operação tivessem uma visão detalhada das métricas críticas, como **vendas por região, por produto, e por cliente**, facilitando a identificação de áreas com

maior potencial de crescimento e ajustes necessários nas campanhas de marketing e promoções.

4. Previsões e Planejamento Futuro

- A adoção de **modelos preditivos** para previsão de demanda representou uma melhoria significativa no planejamento da produção e na alocação de recursos. A empresa agora consegue se antecipar a **tendências sazonais**, ajustando estoques e produção de acordo com a demanda projetada.
- O projeto também destacou o potencial de **modelos de regressão e séries temporais** para aprimorar ainda mais a previsão de demanda e a otimização do inventário, minimizando o risco de falta de produtos ou excesso de estoque.

5. Facilidade de Escalabilidade

- A escolha de uma solução escalável como o **Google BigQuery** garantiu que a infraestrutura de dados da Adventure Works pudesse crescer conforme a empresa expandisse suas operações. O modelo de pagamento sob demanda oferece flexibilidade, permitindo que os custos sejam ajustados de acordo com o volume de dados e processamento.

6. Futuro da Cultura Data-Driven

- A Adventure Works, agora equipada com uma infraestrutura robusta e uma **plataforma de análise de dados moderna**, está preparada para enfrentar os desafios futuros com maior agilidade e precisão. As decisões baseadas em dados ajudarão a empresa a continuar crescendo, identificando novas oportunidades e ajustando suas operações com base em **insights açãoáveis**.

Conclusão Final

Em suma, o projeto não apenas entregou uma solução técnica de alto nível, mas também **mudou a cultura organizacional da Adventure Works**, tornando-a uma empresa verdadeiramente orientada por dados. Isso não apenas facilita a operação diária, mas também coloca a empresa em uma posição estratégica para competir e se destacar no mercado de bicicletas em constante evolução. A combinação de **análises preditivas**, **dashboards detalhados**, e a **escalabilidade do Google BigQuery** proporcionou à empresa uma base sólida para seu crescimento contínuo, superando as expectativas e justificando o investimento inicial.