



FACULDADE DE TECNOLOGIA DO IPIRANGA
CURSO DE BIG DATA PARA NEGÓCIOS

ARTUR STEIN
CAIO DIAMANTINO
CÁTIA URBONAS
HEITOR TERRA
TATIANA BARONI
TIAGO GOMES

PROJETO INTEGRADOR III
DATA WAREHOUSE

Construção de um modelo para e-commerce

SÃO PAULO
2024

RESUMO

O Projeto Integrador III aborda o processo de planejamento e construção de um *Data warehouse* (DW) para um ecossistema de e-commerce, com o objetivo de proporcionar um diferencial competitivo para os clientes que adotam esse modelo de negócio.

Em um cenário dinâmico e globalizado, o Big Data capacita a empresa a se tornar mais competitiva, otimizando suas operações e decisões estratégicas por meio de tecnologias avançadas de dados e Business Intelligence (BI).

A capacidade de processar e analisar grandes volumes de dados em tempo real tornou-se crucial para as empresas que lidam com enormes quantidades de dados gerados por transações, interações com clientes e atividades de marketing e vendas.

A criação do DW envolveu o desenvolvimento de dashboards e relatórios focados em vendas, a definição clara das etapas do projeto, a utilização de um modelo dimensional para otimização das consultas e a criação de dashboards interativos com Power BI para suportar decisões estratégicas.

A implementação de um DW, aliada a técnicas de BI, demonstrou ser capaz de proporcionar significativas melhorias na eficiência operacional e na qualidade das decisões estratégicas. Isso permite à empresa responder rapidamente às mudanças do mercado e às necessidades dos clientes, mantendo-se competitiva e inovadora.

A integração de DW e BI fornece uma base sólida para a inteligência competitiva, permitindo que a empresa obtenha insights valiosos e se posicione de maneira vantajosa no mercado.

ABSTRACT

The Integrative Project III addresses the planning and construction process of a *Data warehouse* (DW) for an e-commerce ecosystem, aiming to provide a competitive advantage for clients adopting this business model.

In a dynamic and globalized scenario, Big Data enables the company to become more competitive by optimizing its operations and strategic decisions through advanced data technologies and Business Intelligence (BI).

The ability to process and analyze large volumes of data in real-time is crucial for the company, which handles enormous amounts of data generated by transactions, customer interactions, and marketing and sales activities.

The creation of the DW involved the development of dashboards and reports focused on sales, the clear definition of project stages, the use of a dimensional model to optimize queries, and the creation of interactive dashboards with Power BI to support strategic decisions.

The implementation of a DW, combined with BI techniques, has demonstrated the ability to significantly improve operational efficiency and the quality of strategic decisions.

This enables the company to quickly respond to market changes and customer needs, maintaining its competitiveness and innovation. The integration of DW and BI provides a solid foundation for competitive intelligence, allowing the company to gain valuable insights and position itself advantageously in the market.

LISTA DE FIGURAS

Figura 1 Ecosistema Olist E-commerce	9
Figura 2 Data Source OLIST	11
Figura 3 Cronograma Trello	12
Figura 4 Data warehouse - Camada Única	14
Figura 5 Data warehouse de duas camadas	14
Figura 6 Data warehouse de três camadas	15
Figura 7 Banco de Dados x Serviço AWS	16
Figura 8 Resumo Analítico Amazon Redshift (AWS1. 2024).	17
Figura 9 Resumo analítico Snowflake	18
Figura 10 Solução de BI com Pentaho Open BI Suíte	20
Figura 11 Componentes da solução Pentaho	21
Figura 12 Dashboard construído com Power BI	22
Figura 13 Power BI - Resumo funcional	23
Figura 14 Tableau - Resumo funcional	24
Figura 15 Dashboard construído com Tableau	24
Figura 16 Estrutura Analítica do Projeto (EAP)	26
Figura 17 Arquitetura DW	27
Figura 18 Ferramenta CASE utilizada para desenho do Modelo de Dados	29
Figura 19 Área de Trabalho da Ferramenta – Power Designer	29
Figura 20 Detalhe do Modelo Dimensional	30
Figura 21 Tela PgAdmin – Postgres – Criação DW	30
Figura 22 Tabelas-base OLIST - Kaggle	31
Figura 23 Dashboard de Vendas	33
Figura 24 Dashboard de Entregas	35
Figura 25 Dashboard de Avaliações	38
Figura 26 Classification Report – XGBoost	41
Figura 27 Classification Report – Com ajuste GridSearch	41
Figura 28 Cluster Report	42
Figura 29 Dados de vendas diárias	43
Figura 30 Previsão de Vendas com LSTM	43

SUMÁRIO

1. INTRODUÇÃO	7
2. A EMPRESA	9
Descrição da empresa	9
Objetivo da empresa	10
Missão, visão e valores	10
Mercado de atuação	10
Massa de dados	10
Aplicações de negócios e sistemas	10
ERP e Hub de Integrações.	11
Marketing e Vendas	11
3. CRONOGRAMA DE DESENVOLVIMENTO	12
4. ARQUITETURA DE BUSINESS INTELLIGENCE (BI)	13
Arquitetura em camadas	13
Armazenamento	13
Extração	13
DW corporativo	14
5. PESQUISA DE BANCOS DE DADOS PARA DW	16
5.1. Amazon Redshift (software proprietário)	16
5.2. Snowflake (Open source)	18
6. PESQUISAS DE SOFTWARES PARA ETL	19
6.1. Microsoft SQL Server Integration Services (software proprietário)	19
6.2. Pentaho Data Integration (PDI) (Open Source)	20
7. PESQUISA DE SOFTWARES DE VISUALIZAÇÃO DOS DADOS	22
7.1. Microsoft Power BI (software proprietário)	22
7.2. Tableau (software proprietário)	23

8.	DESENVOLVIMENTO DO PROJETO	26
8.1.	Definição dos Requisitos do Projeto	26
8.1.1.	Provisionamento de Infraestrutura	26
8.1.2.	Desenvolvimento de Dashboards e Relatórios referentes a Vendas:	26
8.1.3.	Estrutura Analítica do Projeto (EAP)	26
8.1.4.	Cronograma do Projeto	27
8.1.5.	Termo de Abertura do Projeto	27
8.2.	Preparação da Infraestrutura	27
8.2.2.	Arquitetura do DW	27
8.2.3	Modelo de Dados do Projeto (Modelo Dimensional)	28
8.3.	Criação do DW	30
8.4.	Extração, Transformação e Carga de Dados (ETL)	31
8.5.	A origem dos dados:	31
8.6.	Análise dos Dados	31
8.7	Visualização dos Dados com Power BI	32
8.8.	Inteligência Competitiva aplicada ao Projeto	40
9.	CONCLUSÃO	44
	REFERÊNCIAS BIBLIOGRÁFICAS	45
	ANEXOS	47

1. INTRODUÇÃO

A formação do curso de Big Data para Negócios requer um entendimento diferenciado das diversas tecnologias que tornam a gestão dos negócios mais eficiente, objetiva e rápida.

Num mundo globalizado, onde as grandes organizações estão presentes em diferentes países, com suas peculiaridades culturais e situações diversas, torna-se cada vez mais complexo o entendimento das rotinas que envolvem a administração de empresas.

Para lidar com o grande volume de dados, produzidos por fontes diversas e descobrir as tendências com agilidade e segurança, surgem as ferramentas de Business Intelligence.

Considerando esse cenário, a proposta do Projeto Integrador III foi a construção de um *Data warehouse* (DW): um repositório de dados que visa aprimorar a qualidade das informações para a tomada de decisões estratégicas da empresa.

As disciplinas do semestre fornecem as bases necessárias para a elaboração da solução completa e estão distribuídas da seguinte maneira:

Parte técnica:

- Programação em Banco de Dados: gestão de documentos e programação em banco SQL.
- Gestão de Projetos: técnicas de desenvolvimento e administração de equipes, visando o alto desempenho, planejamento, agilidade e qualidade da entrega dos resultados.
- Inteligência Artificial: algoritmos de regressão linear, aplicados ao projeto.
- Laboratório de Programação: linguagem Python
- Redes: protocolos, endereçamento, internet.

Parte de negócios

- Inteligência competitiva: técnicas de distribuição da informação
- Estatística e Probabilidade II: interpretação de indicadores de desempenho, teste de hipóteses, análises comparativas.
- Inglês III: vocabulário técnico de apoio à leitura de documentos.

A Equipe

BIG DATA PARA NEGÓCIOS MÓDULO 3/ 2024	
Artur Stein Carmo RA:2041382311021	<ul style="list-style-type: none"> • Pesquisa e Desenvolvimento • Planejamento • Fundamentação BI – Negócios • Seleção da Base de Dados • Arquitetura de solução DW • Arquitetura de solução ETL • Codificação • Visualização de dados – Dashboard e BI • Inteligência Artificial • Documentação do projeto • Revisão do documento final
Caio da Silva Diamantino 2041382311027	
Catia Urbonas 2041382121040	
Heitor Terra 2041382221043	
Tatiana de Araujo Baroni RA 2041382311006	
Tiago Gomes de Barros RA 2041382311022	

A finalidade deste documento é, portanto, elaborar soluções que apoiem as tomadas de decisões e ofereçam uma visão global da empresa.

2. A EMPRESA

Atualmente, a área de e-commerce ocupa uma posição de destaque como novo modelo de negócios. Formas mais ágeis de operação, se aplicadas de forma ampla, permitem o crescimento rentável do negócio.

Segundo (ARBEX, 2024) o setor foi impulsionado em decorrência da Pandemia, com projeções de movimentação financeira que podem chegar a US\$ 34 tri em 2025.

Esse modelo de vendas com alta escalabilidade, favorece desde microempresas que desejam expandir seus negócios, até gigantes globais maduras com alta capilaridade logística.

Descrição da empresa

A empresa objeto desse estudo é a Olist, provedora de um ecossistema para alavancar vendas de produtos do varejo online. A Figura 1 demonstra o resumo de suas soluções.

Figura 1 Ecossistema Olist E-commerce



FONTE: Olist (2024)

Objetivo da empresa

A Olist tem como objetivo fornecer soluções completas, através de um gerenciamento centralizado, para que seus clientes tenham a melhor performance em seus negócios (OLIST, 2024).

Missão, visão e valores

A empresa (OLIST, 2024) define a cultura com base na descrição abaixo:

Missão

Alavancar negócios do varejo digital, dando oportunidades reais de crescimento, superando as barreiras presenciais e potencializando os resultados dos clientes.

Visão

Dedicação ao sucesso do cliente faz parte da realização buscada pela organização.

Valores

A ênfase no relacionamento entre cliente e empresa focando no aspecto humano do atendimento e na realização do sonho do cliente.

Mercado de atuação

E-commerce para varejo online.

Massa de dados

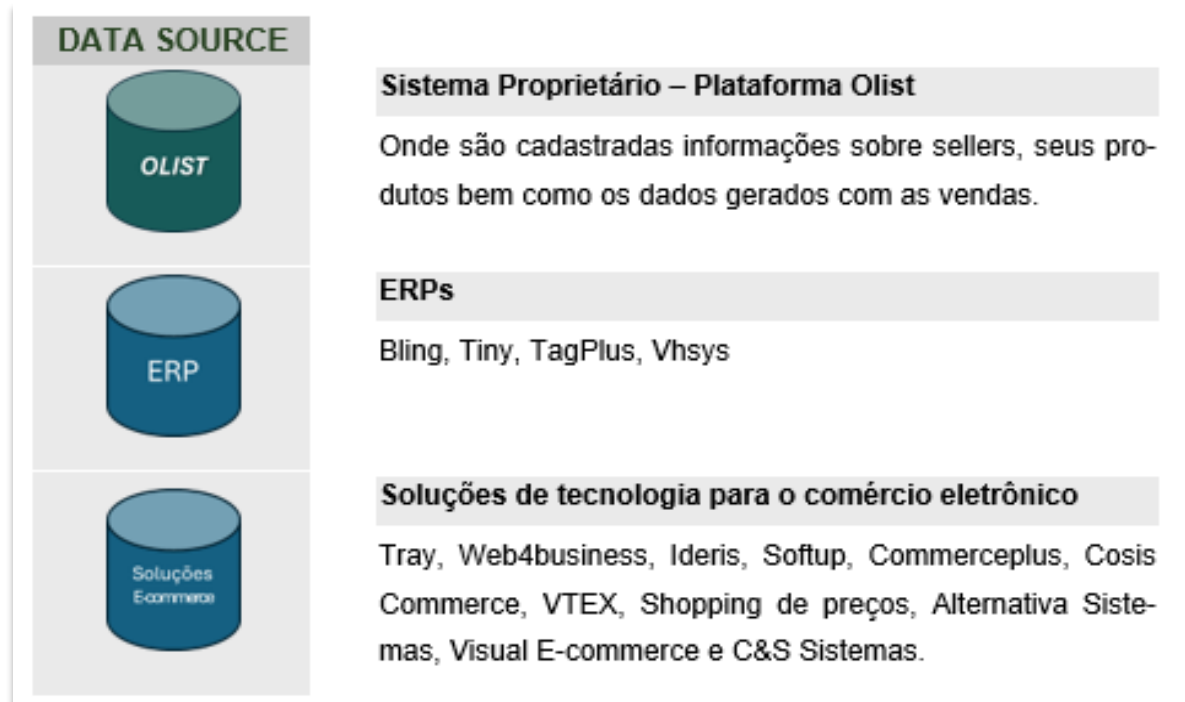
Para a elaboração do projeto, será utilizada a massa de dados disponibilizada pela plataforma Kaggle.

https://www.kaggle.com/datasets/olistbr/brazilian-ecommerce?select=product_category_name_translation.csv

Aplicações de negócios e sistemas

Para uma melhor compreensão das aplicações, as soluções são descritas na Figura 2 Data Source Olist.

Figura 2 Data Source OLIST



FONTE: Adaptado pelos autores (2024)

ERP e Hub de Integrações.

Vnda : *Cash-back*, Indicações premiadas, combinações promocionais

Marketing e Vendas

Entre as vantagens oferecidas (OLIST, 2024), estão:

Exposição online

- Múltiplos canais
- Visibilidade em dobro
- Gestão centralizada

Agilidade Operacional

- Produtos pré-configurados
- Serviço de coleta
- Integrações

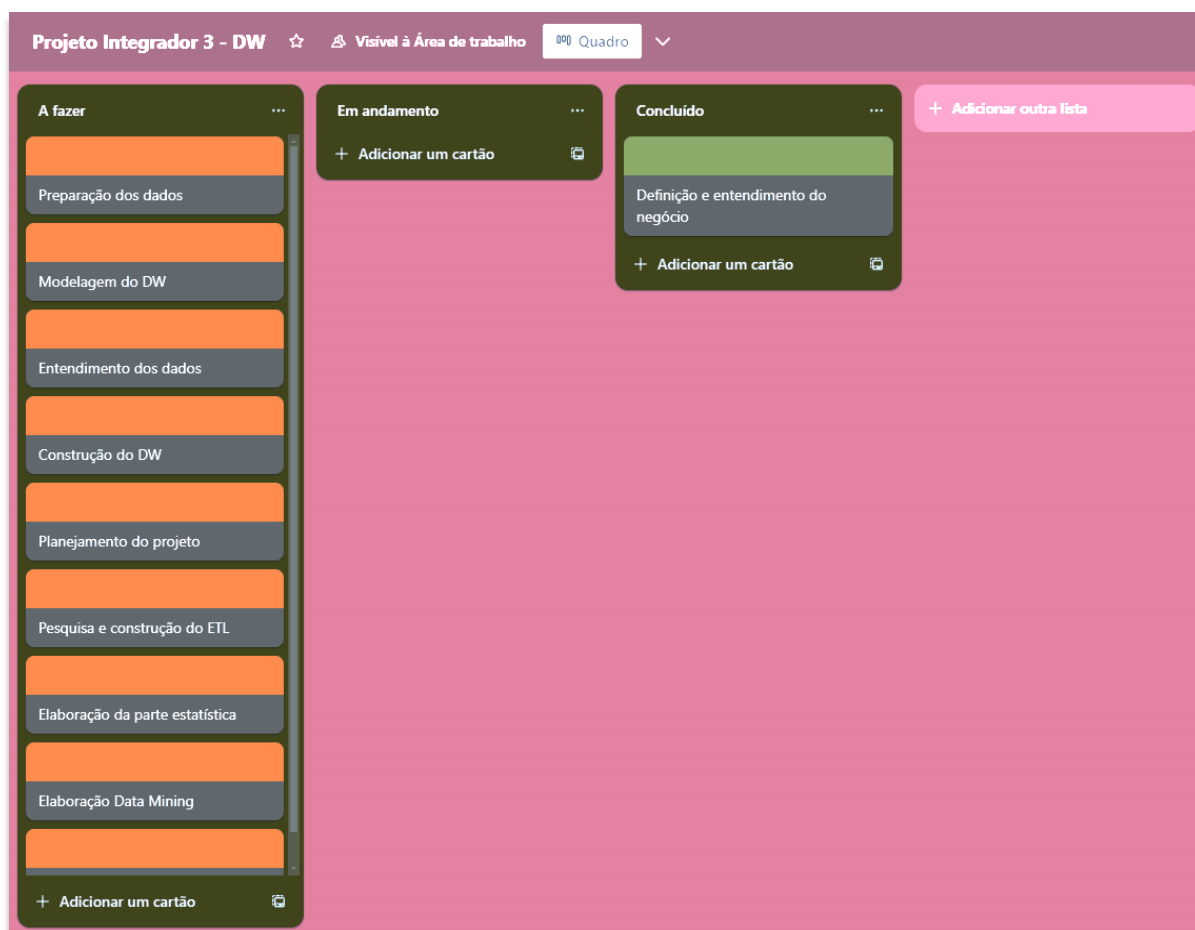
Inteligência de Mercado

- Incentivo Olist (Subsídio)
- Ferramenta de precificação
- Negociações exclusivas

3. CRONOGRAMA DE DESENVOLVIMENTO

Inicialmente feito através da ferramenta Trello, foi substituído pelo Microsoft Planner, disponibilizado pela plataforma Teams, para cumprir os requisitos da disciplina de Gestão de Projetos de maneira mais completa.

Figura 3 Cronograma Trello



FONTE: Autores (2024)

4. ARQUITETURA DE BUSINESS INTELLIGENCE (BI)

Conforme definição do provedor ([AWS 1,2024](#)), *data warehouse* é um repositório central de informações, para o qual os dados fluem de sistemas transacionais, bancos de dados relacionais e de outras fontes.

Analistas de negócios, profissionais de TI e tomadores de decisões, acessam os dados por meio de ferramentas de *Business Intelligence* (BI), clientes SQL e outras aplicações de análise.

Arquitetura em camadas

Composição das camadas:

- **Superior:** Cliente do *Front-end*: apresenta os resultados através de ferramentas de relatórios, análises e mineração de dados
- **Intermediária:** consiste no mecanismo de acesso e análise de dados
- **Inferior:** servidor do BD, responsável por carregar e armazenar os dados.

Armazenamento

Segundo ([AWS1](#), 2024), os dados são armazenados de dois modos:

- I. Dados acessados com frequência: armazenamento muito rápido (como unidades SSD);
- II. Dados acessados com pouca frequência: armazenamento de objetos baratos, como o Amazon S3.

Extração

A arquitetura de um *data warehouse* típico define a disposição dos dados em diferentes bancos de dados ([ASTERA](#),2024). Para extrair informações valiosas de dados brutos, uma estrutura moderna de *data warehouse* identifica a técnica mais eficaz para organizar e limpar os dados.

Usando um modelo dimensional, o *data warehouse* extrai e converte os dados brutos na área de preparação, em uma estrutura de armazenamento de consumíveis simples, para fornecer uma inteligência de negócios valiosa.

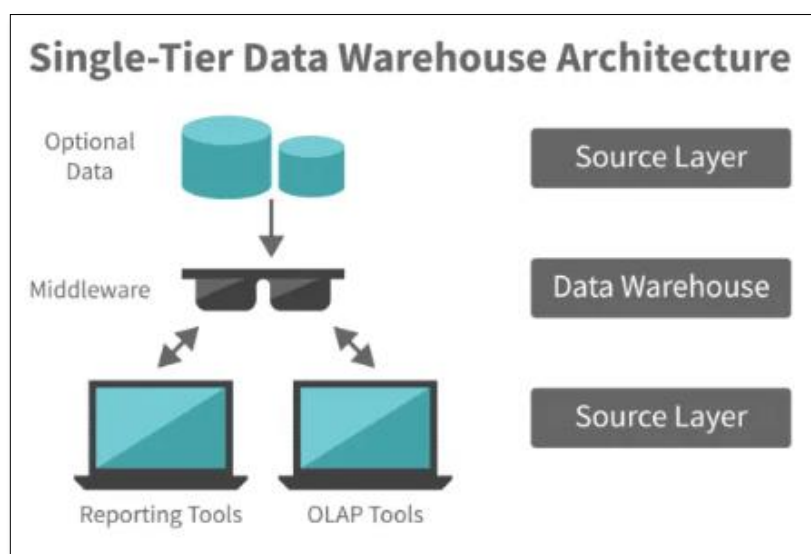
Além disso, ao contrário de um armazém de dados na nuvem, um modelo de *data warehouse* tradicional requer servidores locais para que todos os componentes do warehouse funcionem.

DW corporativo

Ao projetar um *data warehouse* corporativo, existem três tipos diferentes de modelos a serem considerados ([ASTERA,2024](#)).

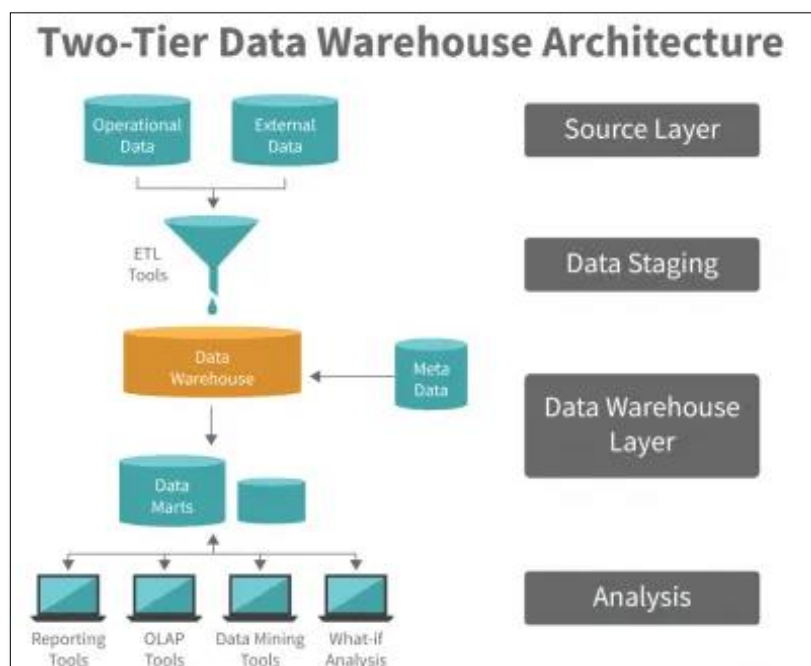
1. Camada única, figura 3;
2. Duas camadas, figura 4;
3. Três camadas, figura 5.

Figura 4 *Data warehouse* - Camada Única



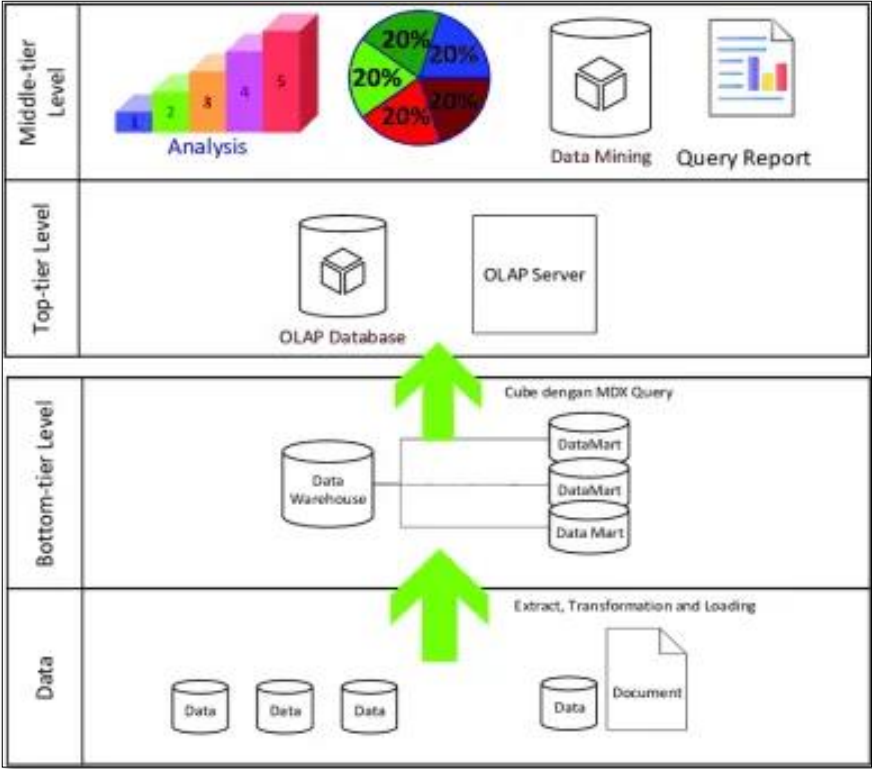
FONTE: Astera (2024)

Figura 5 *Data warehouse* de duas camadas



FONTE: Astera (2024)

Figura 6 Data warehouse de três camadas



FONTE: Astera (2024)

A compreensão da arquitetura de BI direcionou as pesquisas de ferramentas que atendessem melhor ao projeto em cada etapa do desenvolvimento.

5. PESQUISA DE BANCOS DE DADOS PARA DW

Bancos de dados para *data warehouse* (DW) são sistemas de gerenciamento de banco de dados (DBMS) especialmente projetados para a análise de dados e suporte a processos de tomada de decisão em larga escala.

Diferentemente dos bancos de dados transacionais (OLTP), que são otimizados para operações de inserção, atualização e exclusão de dados, os bancos de dados para DW são otimizados para consultas complexas e agregações em grandes volumes de dados (OLAP - Online Analytical Processing).

A seguir citamos dois bancos de dados com estas características.

5.1. Amazon Redshift (software proprietário)

A Amazon Web Services ([AWS3](#), 2024), plataforma de serviços de computação em nuvem oferecida pela Amazon, possui os serviços ilustrados pela Figura 7:.

Figura 7 Banco de Dados x Serviço AWS

DB	EXEMPLO	SERVIÇO AWS
Relacional	Aplicações tradicionais, planejamento de recursos empresariais (ERP), gerenciamento de relacionamento com o cliente (CRM), <i>e-commerce</i>	<ul style="list-style-type: none">▪ Amazon Aurora▪ Amazon RDS▪ Amazon Redshift
Chave – Valor	Aplicativos da Web de alto tráfego, sistemas de comércio eletrônico, aplicações de jogos	<ul style="list-style-type: none">▪ Amazon DynamoDB
Em memória	Armazenamento em cache, gerenciamento de sessões, tabelas de classificação de jogos, aplicações geoespaciais	<ul style="list-style-type: none">▪ Amazon ElastiCache▪ Amazon MemoryDB for Redis
Documento	Gerenciamento de conteúdo, catálogos, perfis de usuários	<ul style="list-style-type: none">▪ Amazon DocumentDB Compatível com o MongoDB
Coluna ampla	Aplicações industriais de grande escala para manutenção de equipamentos, gerenciamento de frota e otimização de rotas	<ul style="list-style-type: none">▪ Amazon Keyspaces
Grafo	Deteccção de fraudes, redes sociais, mecanismos de recomendação	<ul style="list-style-type: none">▪ Amazon Neptune
Séries temporais	Aplicações de Internet das Coisas (IoT), DevOps, telemetria industrial	<ul style="list-style-type: none">▪ Amazon Timestream
Ledger	Sistemas de registro, <i>supply chain</i> , registros, transações bancárias	<ul style="list-style-type: none">▪ Amazon Ledger Database Services (QLDB)

FONTE: Adaptado de [AWS2](#) (2024)

O Amazon Redshift ([AWS 3, 2024](#)), solução de *data warehouse* integrada ao ecossistema da Amazon Web Services, foi projetada para lidar com grandes volumes de dados e migrações de banco de dados em larga escala. As características são demonstradas na figura 8:

Figura 8 Resumo Analítico Amazon Redshift ([AWS1, 2024](#)).

Tecnologia de processamento	<ul style="list-style-type: none"> ▪ Processamento paralelo massivo da Parcel (MPP); ▪ Reduz o tempo de execução de comandos; ▪ Permite operações em bilhões de linhas de dados de uma só vez
Escalabilidade	<ul style="list-style-type: none"> ▪ significativa.
Capacidade	<ul style="list-style-type: none"> ▪ Suporta <i>clusters</i> de até 16 petabytes; ▪ Análises avançadas; ▪ Grandes volumes de dados de logs ou feeds em tempo real, ou através do Amazon Kinesis Data Firehouse.
Compatibilidade	<ul style="list-style-type: none"> ▪ compatível com ampla variedade de aplicativos
Conexões	<ul style="list-style-type: none"> ▪ ODBC e JDBC
Início	<ul style="list-style-type: none"> ▪ derivado da versão 8.0.2 do PostgreSQL: Versão Beta: 2012 Versão completa: 2013
APLICAÇÕES DE NEGÓCIOS	
<i>Business Intelligence</i>	<ul style="list-style-type: none"> ▪ Análise de grandes volumes de dados para obter insights de negócios.
Análise de Logs	<ul style="list-style-type: none"> ▪ Processamento e análise de grandes volumes de dados de log para monitoramento e detecção de anomalias.
<i>Big Data Analytics</i>	<ul style="list-style-type: none"> ▪ Integração com outras ferramentas de big data para processamento e análise de dados em grande escala.

Fonte: resumo elaborado por Tatiana Baroni ([AWS 3, 2024](#))

5.2. Snowflake (Open source)

O Snowflake, um *data warehouse* em nuvem, funciona em várias plataformas, incluindo AWS, Azure e Google Cloud.

Conhecido por sua arquitetura única, separa armazenamento e computação, permitindo escalabilidade independente ([SNOWFLAKE](#), 2024).

Foi criado para atender às demandas de empresas que lidam com grandes volumes de dados e precisam de uma solução escalável, flexível e de alto desempenho.

As características principais incluem:

Figura 9 Resumo analítico Snowflake

Processamento	<ul style="list-style-type: none">• Arquitetura de múltiplos <i>clusters</i>:<ul style="list-style-type: none">○ Cluster de Armazenamento○ Cluster Dados processados• Simultâneo, sem comprometer o desempenho
Capacidade	<ul style="list-style-type: none">• Dimensiona automaticamente a capacidade de computação e armazenamento• Adaptação a picos de demanda ou alterações nas cargas de trabalho sem comprometer o desempenho.
Escalabilidade	<ul style="list-style-type: none">• Alta
Compatibilidade	<ul style="list-style-type: none">• Multiplataformas• Suporte total à linguagem SQL
APLICAÇÕES DE NEGÓCIOS	
<i>Business Intelligence</i>	<ul style="list-style-type: none">▪ Análise de grandes volumes de dados para obter insights de negócios.
Segurança:	<ul style="list-style-type: none">▪ recursos avançados de criptografia de dados em repouso e em trânsito;▪ autenticação multifator;▪ controle de acesso baseado em função;▪ compartilhamento seguro.

Fonte: resumo elaborado por Tatiana Baroni ([AWS 3](#),2024)

6. PESQUISAS DE SOFTWARES PARA ETL

Ferramentas de ETL (Extract, Transform, Load) são softwares que facilitam o processo de extração, transformação e carregamento de dados de várias fontes para um destino, geralmente um data warehouse.

Processo fundamental para a integração de dados, garante que informações de diferentes sistemas sejam consolidadas, formatadas e disponibilizadas de forma coerente para análise e produção de relatórios. Aqui estão os exemplos de duas ferramentas utilizadas no processo ETL:

6.1. Microsoft SQL Server Integration Services (software proprietário)

Segundo documentação da ferramenta ([MICROSOFT](#), 2024), a plataforma SQL Server Integration Services (**SSIS**) possibilita criar a integração e soluções de transformações de dados.

Resolve problemas de negócios complexos:

- Copiar ou baixar arquivos
- Carregar data warehouses
- Limpar e explorar dados
- Gerenciar objetos e dados do SQL Server

Composição:

- Conjunto avançado de tarefas e transformações internas.
- Ferramentas gráficas para a criação de pacotes.
- Banco de dados do Catálogo do SSIS para armazenar, executar e gerenciar pacotes.

Vantagens:

- ETL de ampla variedade de fontes: XML, arquivos simples e dados relacionais
Carrega dados em um ou mais destinos.
- Criar soluções sem escrever uma única linha de código.
- Programar o modelo de objeto extensivo do Integration Services.
- Criar programaticamente pacotes e tarefas personalizadas de código.

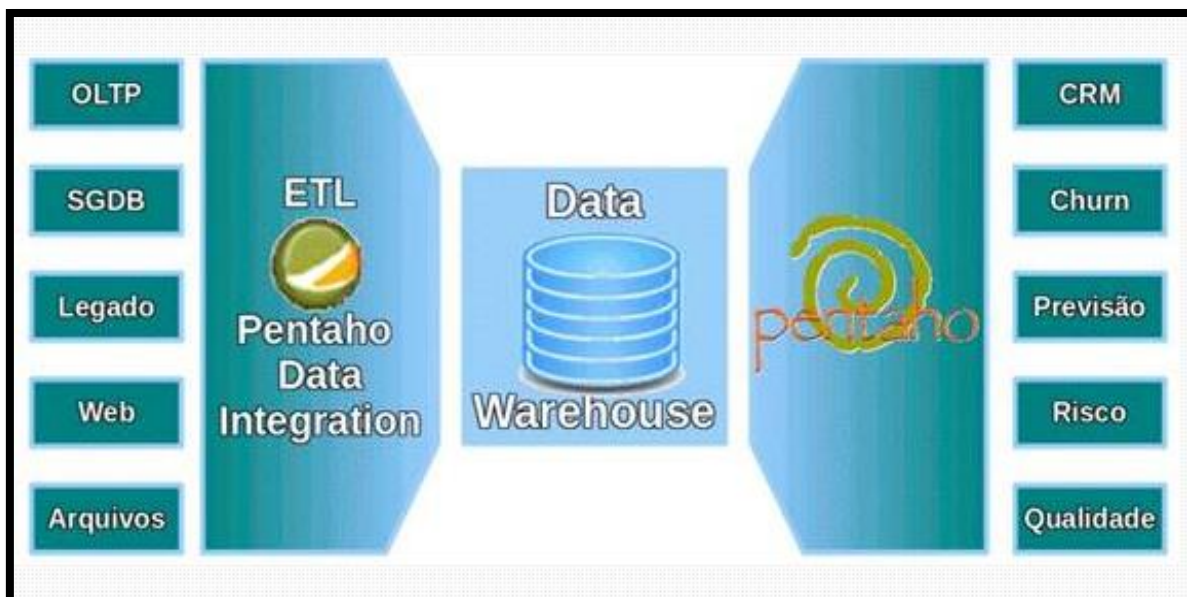
6.2. Pentaho Data Integration (PDI) (Open Source)

De acordo com (ROLDÁN,2013) a solução de ETL Pentaho Data Integration permite atender a requisitos de extração, transformação e carga de dados.

Possui um ambiente de design gráfico intuitivo, de arrastar e soltar.

Desenvolvido pela Pentaho Corporation, que agora faz parte do grupo Hitachi Vantara, o Pentaho se destaca por sua flexibilidade, escalabilidade e capacidade de fornecer soluções completas de BI, demonstrada pela figura 8.

Figura 10 Solução de BI com Pentaho Open BI Suíte



FONTE: DEVMEDIA (2024)

Vantagens

- Custo acessível por ser open source;
- Capacidade de customização;
- Flexibilidade para integração com infraestruturas de TI existentes;
- Reconhecimento de qualidade no mercado pela Consultoria Gartner;
- Acessibilidade através da web.

A composição dessa solução de BI ([VOITTO, 2024](#)) será descrita pela figura 11:

Figura 11 Componentes da solução Pentaho

Pentaho BI Server	Servidor web de aplicação que provê a arquitetura e infraestrutura para a construção de soluções de BI. Além de possibilitar a exibição de relatórios, visões OLAP e dashboards.
Pentaho Meta Editor (PME)	Permite agrupar campos de tabelas que possuem correlações, criando visões de negócios independentes, mesmo quando da mesma tabela. Visual, capaz de mapear qualquer fonte de dados através de um driver JDBC.
Pentaho Schema Workbench (PSW)	Cria cubos OLAP que serão exibidos na PUC (Console de Usuário), com uma interface visual que facilita a navegação entre as definições do cubo, permitindo a criação de métricas, dimensões e hierarquias.
Pentaho Data Integration (PDI)	Também conhecido como Kettle: solução para integração de dados, destinada a processos de ETL, além de facilitar a captura, limpeza e armazenamento de dados em <i>Data Warehouses</i> , usando um formato uniforme e consistente para usuários finais. Operações visuais podem ser realizadas por usuários em geral.
Pentaho Business Analytics	Esta plataforma permite acessar, integrar, manipular, visualizar e analisar com segurança seus dados, independentemente de onde estejam armazenados. Auxilia na descoberta, análise e visualização de dados.
Pentaho Reporting Designer	Ferramenta de criação: permite que profissionais criem relatórios detalhados e com qualidade de impressão, conectando-se a qualquer fonte para qual exista um driver JDBC.
Pentaho Dashboards Designer	Possibilita de maneira simples a criação do seu dashboard, com controles de filtro dinâmicos, permitindo que os usuários alterem os detalhes do painel.
Pentaho Data Mining	Pesquisa de dados e padrões, usando o Waikato Environment for Knowledge Analysis (Weka) que consiste em algoritmos de aprendizado de máquinas (machine learning) para um amplo conjunto de tarefas de mineração de dados.

FONTE: Adaptado de [VOITTO](#) (2024)

7. PESQUISA DE SOFTWARES DE VISUALIZAÇÃO DOS DADOS

As visualizações de indicadores de desempenho nos negócios, são parte fundamental para *Business Intelligence*.

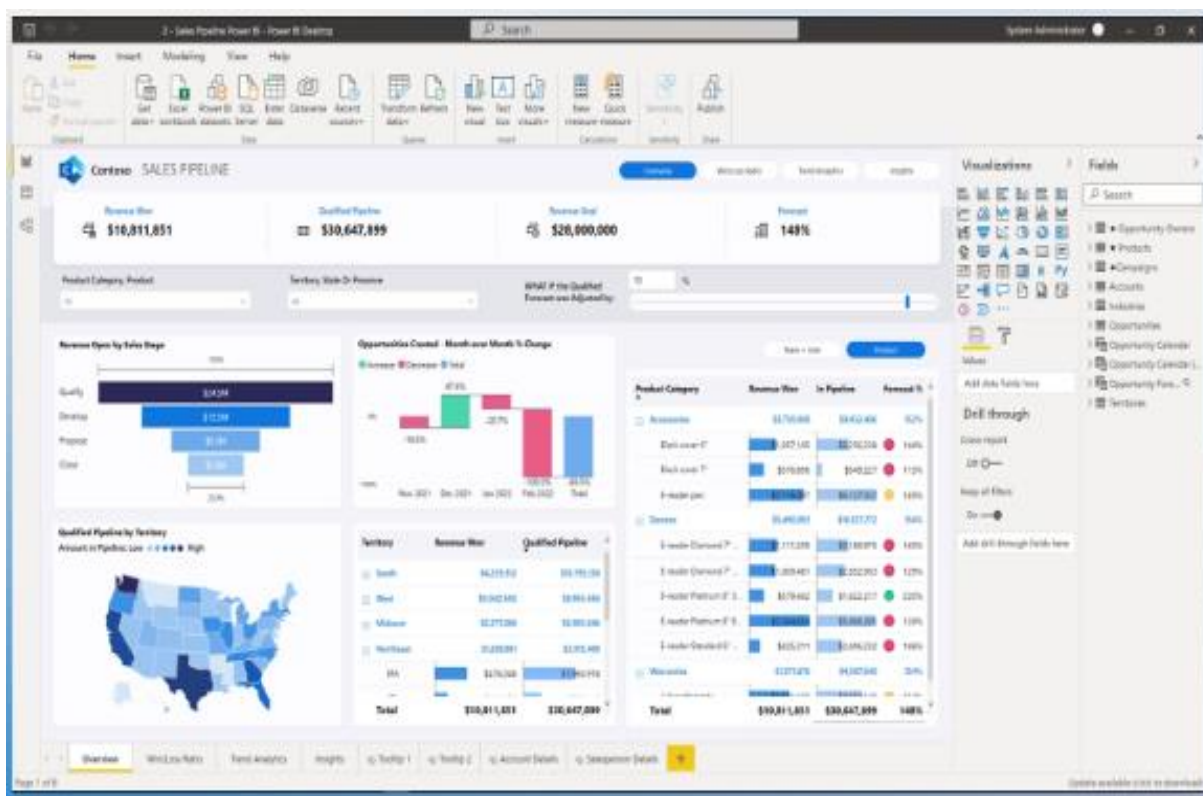
A disponibilização das informações analisadas na forma de Dashboards, permite o acompanhamento dos resultados da empresa, oferecendo *feedback* mais preciso para tomada de decisões.

Nesse capítulo, comentamos sobre dois desses recursos.

7.1. Microsoft Power BI (software proprietário)

O Power BI ([MICROSOFT](#), 2024), uma ferramenta de visualização customizada de dados da Microsoft, será exemplificada pela figura 12, e a figura 13 resume suas principais características.

Figura 12 Dashboard construído com Power BI



FONTE: [MICROSOFT](#) 2 (2024)

Figura 13 Power BI - Resumo funcional

Conectividade <ul style="list-style-type: none"> • Ampla e variada: arquivos locais, banco de dados relacional, nuvem etc. 	Visualizações interativas <ul style="list-style-type: none"> • gráficos: barras, linhas, pizza; • mapas; • tabelas dinâmicas e muito mais.
Compatibilidade <ul style="list-style-type: none"> • Excel, SQL Server, Azure SQL • Database, Salesforce, SharePoint, Google Analytics Dynamics 365, etc. 	Dashboards <ul style="list-style-type: none"> • Compartilháveis; • Atualizações em tempo real; • Várias visualizações numa tela.
Segurança de Dados <ul style="list-style-type: none"> • controles de acesso baseados em função; • criptografia de dados; • auditoria de atividades; • em conformidade com padrões regulatórios. 	Análise de Dados <ul style="list-style-type: none"> • segmentação de dados; • filtragem; • ordenação; • agrupamento; • hierarquias; • medidas calculadas

FONTE: Autores (2024)

7.2. Tableau (software proprietário)

A ferramenta ([TABLEAU](#), 2024) surgiu de um projeto de ciência da computação na Universidade Stanford em 2011, como uma plataforma de análise visual de dados.

Em 2019, foi adquirida pela empresa *Salesforce*.

Diferenciais

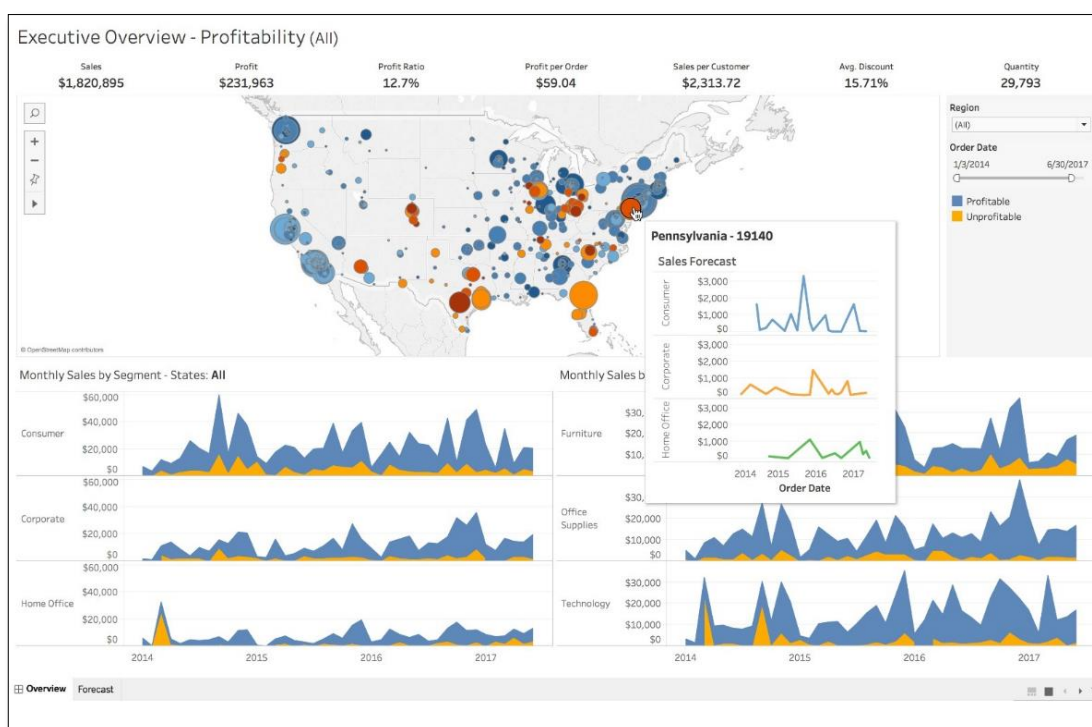
- **Colaborativa:** comunidade, fóruns e programas permanentes;
- **Reciprocidade:** todos os membros contribuem tanto on-line quanto off-line, compartilhando conexões, experiências e práticas recomendadas;
- **Investimento contínuo:** foco exclusivo na área de BI e análise.
- **Inovação:** fortemente impulsionada pelo feedback do cliente.

Figura 14 Tableau - Resumo funcional

Conectividade <ul style="list-style-type: none"> • Ampla e variada: arquivos locais, banco de dados relacional, nuvem etc. 	Dashboards <ul style="list-style-type: none"> • Atualizações em tempo real • Várias visualizações numa tela. • Integração: exploração de dados visual ilimitada. Não interrompe o fluxo da análise.
Compatibilidade <ul style="list-style-type: none"> • ampla e variada: qualquer tipo de dados, praticamente qualquer sistema. 	
Segurança de Dados <ul style="list-style-type: none"> • conta de logon no banco de dados • autenticação por meio do Windows NT • filtros de usuário 	Análise de dados <ul style="list-style-type: none"> • segmentação de dados; • filtragem; • ordenação; • agrupamento; • hierarquias; • medidas calculadas
Visualizações interativas <ul style="list-style-type: none"> • gráficos de barras, linhas, pizza; • mapas; • tabelas dinâmicas e muito mais. 	

FONTE: Autores (2024)

Figura 15 Dashboard construído com Tableau

FONTE: [Tableau](#) (2024)

7.3. Vantagens do Power BI em relação ao Tableau:

- Integração com o ecossistema Microsoft
- Custo: Possui uma versão gratuita com recursos limitados
- Interface intuitiva e amigável
- Inteligência Artificial (IA) integrada:
 - análise de dados automatizada
 - detecção de anomalias
 - sugestões de insights: ajuda na busca de padrões e tendências
- Integração com serviços em nuvem.

Importante notar que a escolha entre o Power BI e o Tableau depende das necessidades específicas da organização, das habilidades da equipe e do orçamento disponível.

8. DESENVOLVIMENTO DO PROJETO

Baseados nos resultados das pesquisas, seguimos para o desenvolvimento da solução, cujas etapas são objeto desse capítulo, e a documentação complementar da codificação segue na íntegra nos ANEXOS componentes desse trabalho.

8.1. Definição dos Requisitos do Projeto

8.1.1. Provisionamento de Infraestrutura

Implementação DW: pode incluir a configuração de instâncias de computação, armazenamento de dados, redes, serviços de BD e outros recursos operacionais.

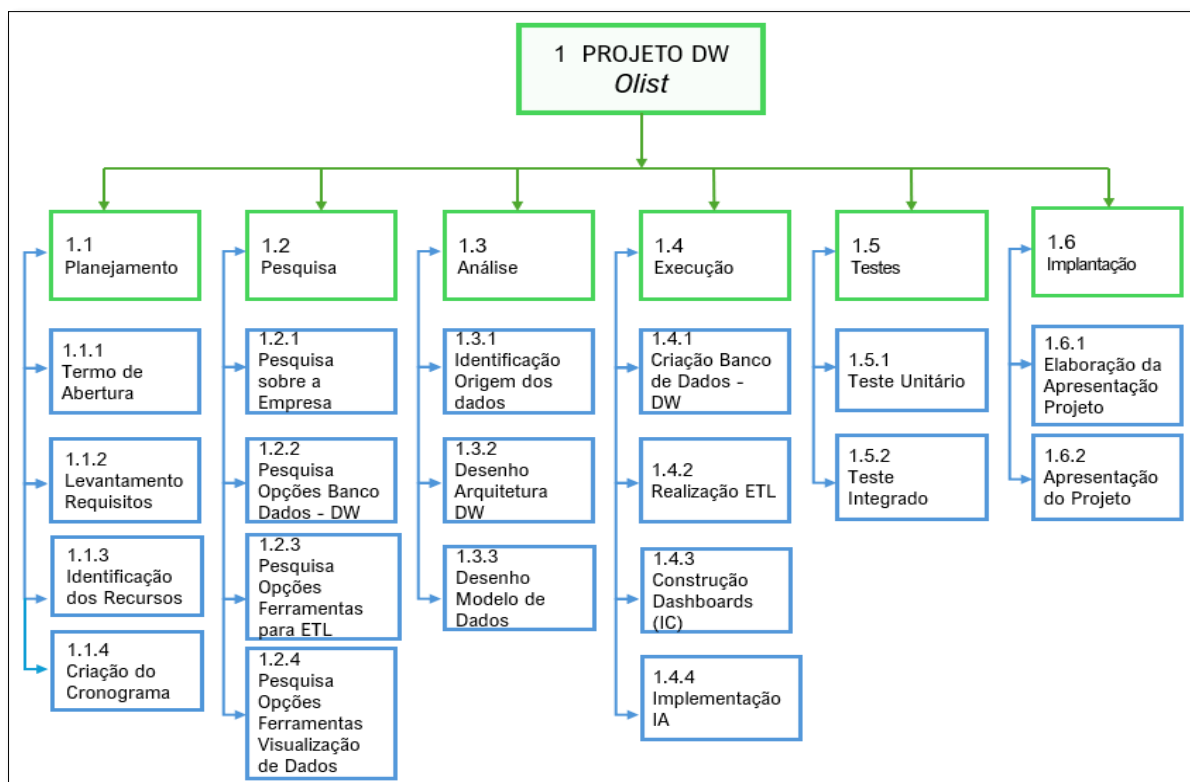
8.1.2. Desenvolvimento de Dashboards e Relatórios referentes a Vendas:

A composição de dashboards envolveu conceitos de Inteligência competitiva, estatística, programação em banco de dados, além da aplicação de algoritmos de Big Data para o tratamento de dados, codificação e desenvolvimento de IA.

8.1.3. Estrutura Analítica do Projeto (EAP)

A Gestão de Projetos baseou-se na Estrutura Analítica do Projeto (EAP) (PMBOK,2024). Segue a decomposição hierárquica detalhada na figura 16.

Figura 16 Estrutura Analítica do Projeto (EAP)



FONTE: Autores (2024)

8.1.4. Cronograma do Projeto

Há várias ferramentas que auxiliam a equipe na organização de tarefas em equipe, para facilitar o compartilhamento de informações e acompanhar os prazos importantes a serem cumpridos durante todo o processo.

Os detalhes estão no Anexo A, componente do documento.

8.1.5. Termo de Abertura do Projeto

O documento formaliza o início do projeto e foi disponibilizado no Anexo B desse trabalho.

8.2. Preparação da Infraestrutura

A implementação de um *Data warehouse* (DW) requer uma infraestrutura adequada para garantir que o sistema possa armazenar, gerenciar e processar grandes volumes de dados de forma eficiente e confiável.

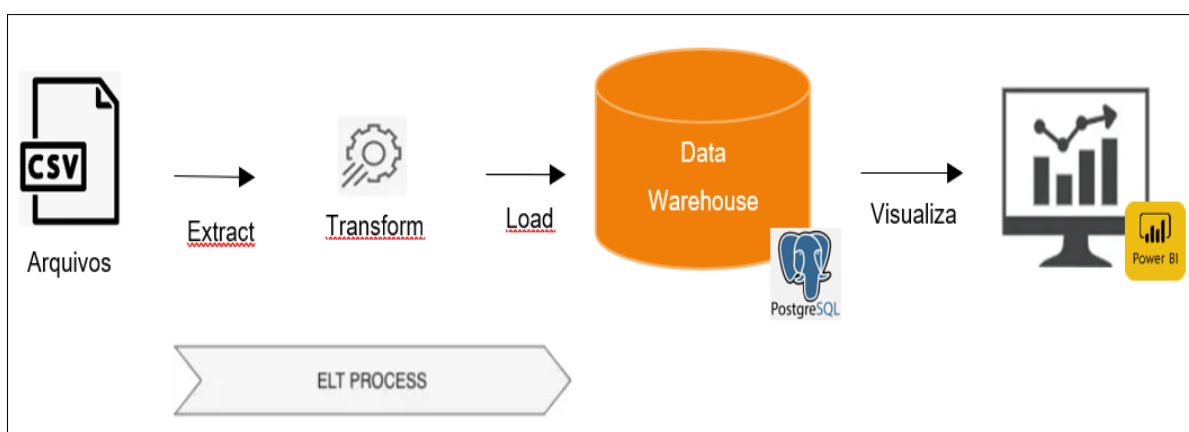
Sua criação será descrita nessa fase.

8.2.2. Arquitetura do DW

Utilizamos o PostgreSQL, SGBD utilizado na disciplina de Programação em Banco de Dados I.

A figura 17 ilustra a arquitetura escolhida para implementação do DW:

Figura 17 Arquitetura DW



Fonte: Autores (2024)

8.2.3 Modelo de Dados do Projeto (Modelo Dimensional)

A modelagem dimensional, técnica predominante no projeto lógico de bancos de dados, tem como objetivo apresentar os dados de forma intuitiva, ao mesmo tempo que atende às necessidades de acesso com alto desempenho (Kimball, 1998).

Composição:

- tabela de fatos:
- um conjunto de tabelas menores denominadas tabelas de dimensão.

Nessa estrutura chamada "modelo estrela", os dados são desnormalizados para evitar junções entre tabelas, reduzindo assim o tempo de consulta.

Modelo:

- quase todos os tipos de dados empresariais podem ser visualizados como cubos de dados.
- As células do cubo contêm os valores medidos, enquanto os lados do cubo definem as dimensões naturais dos dados (Harrison, 1998). Por exemplo, um cubo de dados pode conter dimensões como produto, mercado e tempo.
- Podem abranger de 4 a 15 dimensões, refletindo a complexidade dos dados empresariais do mundo real.

Dimensão

- conjunto de atributos textuais que estão fortemente correlacionados entre si e descrevem características específicas de um objeto.

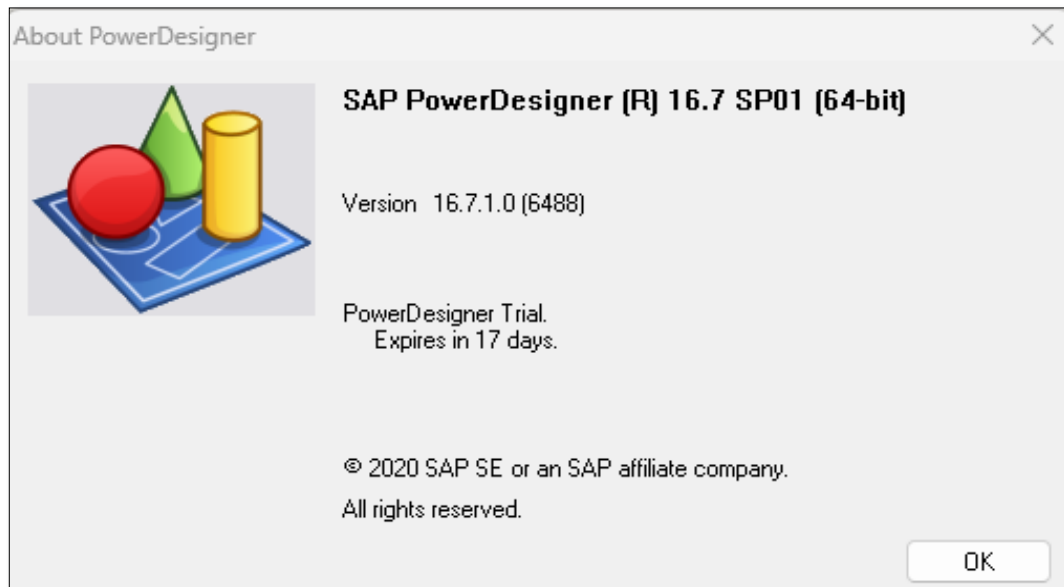
Exemplo de dimensão: no BD de varejo, dimensões comuns incluem produto, armazém, cliente, promoção e tempo.

Fato

- resume as relações entre diversas dimensões.
- Sua chave é uma combinação das chaves primárias das tabelas de dimensão.
- Tipicamente, contém um ou mais fatos numéricos que ocorrem na combinação de chaves que define cada registro.

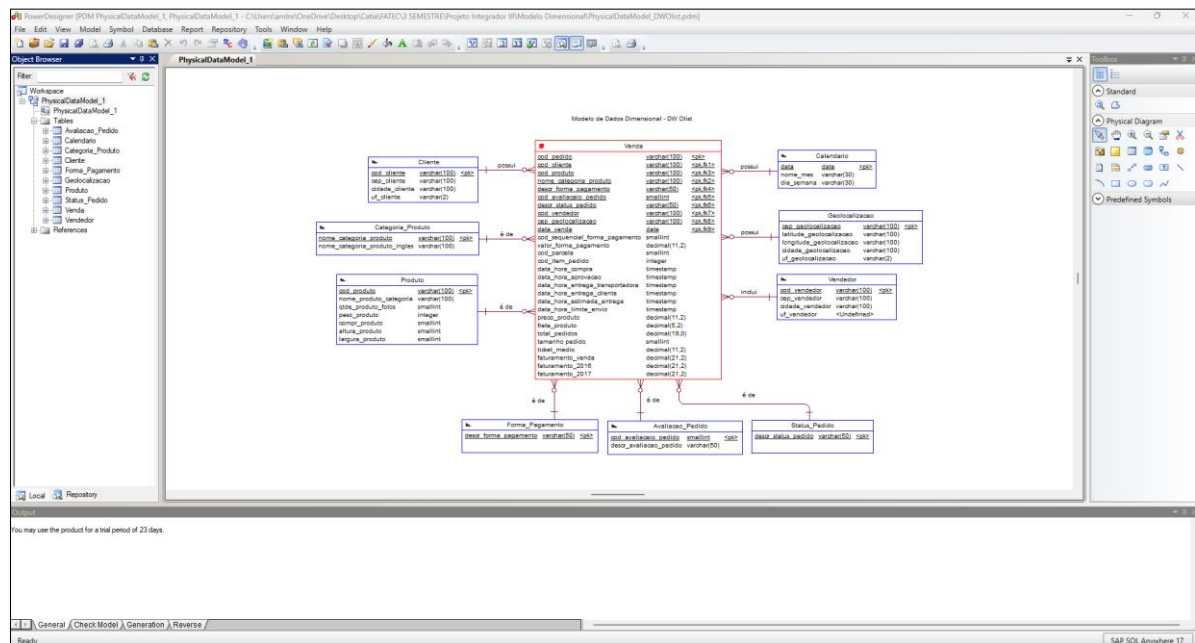
Foi utilizada a ferramenta SAP Power Designer para a construção do Modelo Dimensional de Dados: As figuras 18, 19 e 20 demonstram detalhes do Modelo

Figura 18 Ferramenta CASE utilizada para desenho do Modelo de Dados'



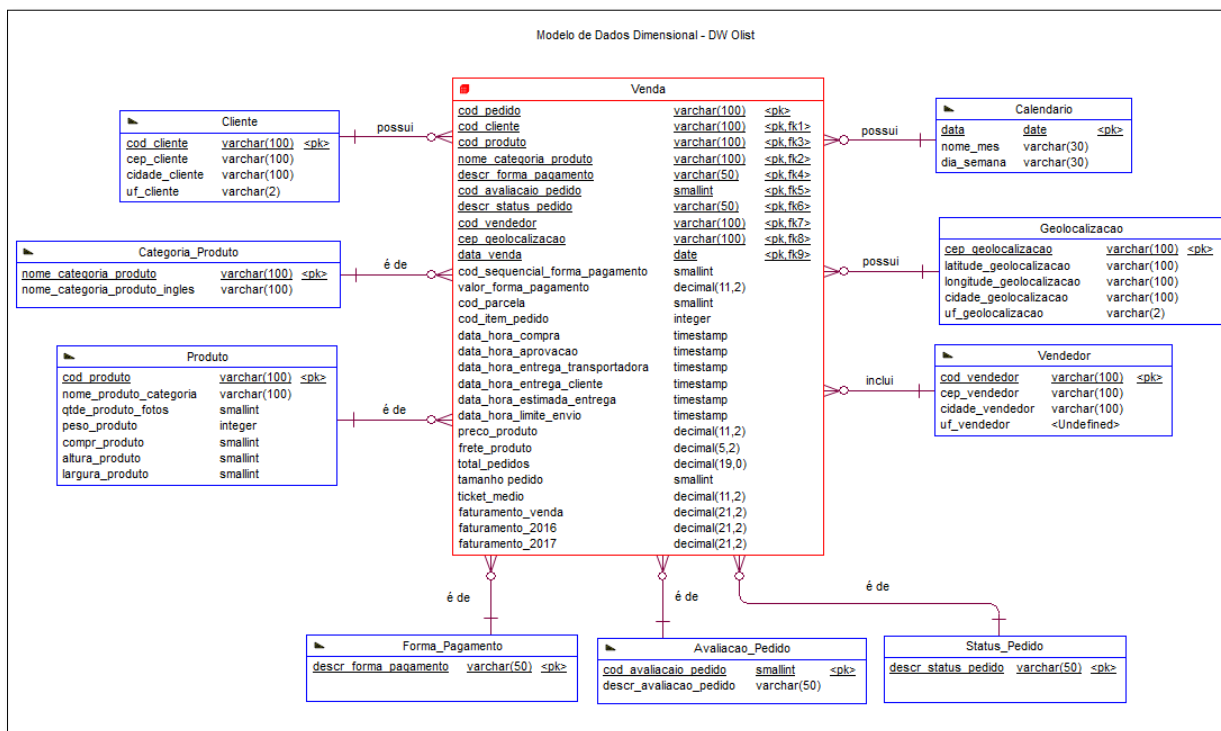
Fonte: ferramenta case Power Designer (2024)

Figura 19 Área de Trabalho da Ferramenta – Power Designer



Fonte: Ferramenta case Power Designer (2024)

Figura 20 Detalhe do Modelo Dimensional

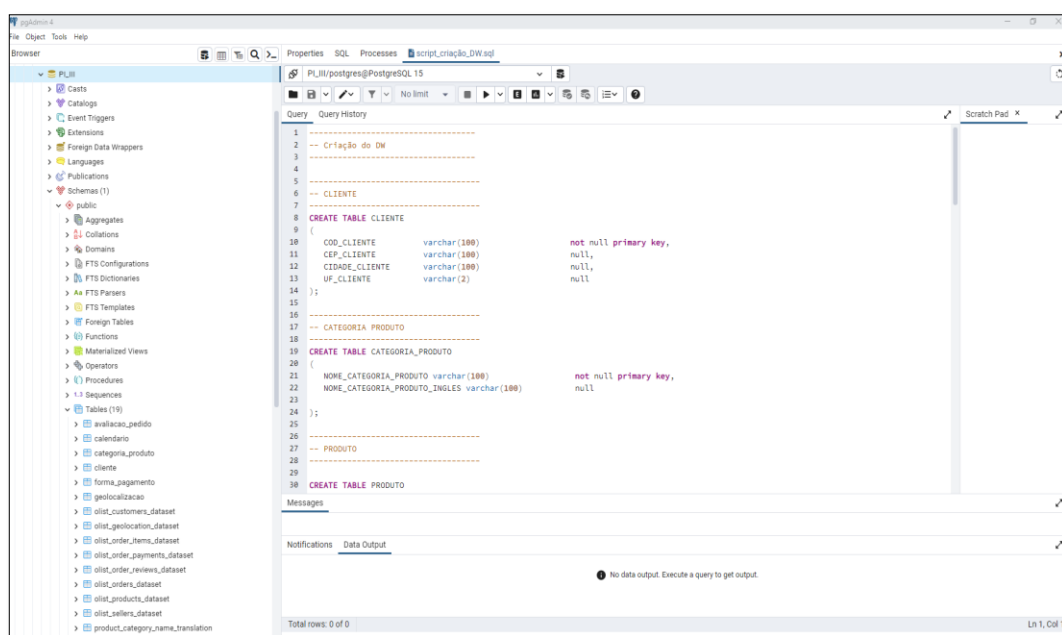


Fonte: Ferramenta case Power Designer (2024)

8.3. Criação do DW

As tabelas foram criadas no BD Postgres utilizando a interface da ferramenta PgAdmin. O script completo está no Anexo C.

Figura 21 Tela PgAdmin – Postgres – Criação DW



Fonte: PgAdmin – Postgres (2024)

8.4. Extração, Transformação e Carga de Dados (ETL)

A ETL foi realizada no PgAdminm em SQL. Detalhes do script, no ANEXO D.

8.5. A origem dos dados:

As tabelas obtidas ([KAGGLE](#),2024) foram detalhadas na figura 23.

Figura 22 Tabelas-base OLIST - *Kaggle*

olist_customers_dataset: <ul style="list-style-type: none"> • customer_id • customer_unique_id • customer_city • customer_state 	olist_order_reviews_dataset: <ul style="list-style-type: none"> • review_id • order_id • review_score
olist_order_items_dataset: <ul style="list-style-type: none"> • order_id • order_item_id • product_id • seller_id • shipping_limit_date • price • freight_value 	olist_orders_dataset: <ul style="list-style-type: none"> • order_id • customer_id • order_status • order_purchase_timestamp • order_approved_at • order_delivered_carrier_date • order_delivered_customer_date • order_estimated_delivery_date
olist_order_payments_dataset: <ul style="list-style-type: none"> • order_id • payment_sequential • payment_type • payment_value 	olist_sellers_dataset: <ul style="list-style-type: none"> • seller_id • seller_city • seller_state
olist_products_dataset: <ul style="list-style-type: none"> • product_id • product_category_name 	

FONTE: Lay-out adaptado pelos Autores (2024)

8.6. Análise dos Dados

A análise de dados foi beneficiada com sua visualização através de dashboards construídos em Power BI. A partir deles poderão ser extraídos insights valiosos que podem ser usados para melhorar processos, prever tendências, identificar problemas e oportunidades e tomar decisões.

8.7 Visualização dos Dados com Power BI

A visualização no Power BI demonstra três situações de negócio:

1. Análise do histórico de Vendas
2. Status de entrega
3. Análise da experiência de compra

CRIANDO COLUNAS

dias_vendedor_transportadora	dias_transportadora_cliente	dias_cliente_previsto	horas_processamento_pedido	dias_aprovacao_cliente
1	3	19	2	4
1	17	8	26	18
3	6	22	24	9
4	3	27	0	7
3	3	24	0	6
3	3	24	0	6
2	6	23	0	8
2	7	19	0	9
2	4	17	0	6
2	3	11	0	5
4	3	18	25	7
3	3	20	2	6
3	2	18	0	5
3	9	14	0	12

FÓRMULAS:

```

dias_vendedor_transportadora = DATEDIFF((vendas[data_hora_aprovacao]),
(vendas[data_hora_entrega_transportadora]), DAY)

dias_transportadora_cliente = DATEDIFF((vendas[data_hora_entrega_transportadora]),
(vendas[data_hora_entrega_cliente]), DAY)

dias_cliente_previsto = DATEDIFF((vendas[data_hora_entrega_cliente]),
(vendas[data_hora_estimada_entrega]), DAY)

horas_processamento_pedido = DATEDIFF((vendas[data_hora_compra]),
(vendas[data_hora_aprovacao]), HOUR)

dias_aprovacao_cliente = DATEDIFF((vendas[data_hora_aprovacao]),
(vendas[data_hora_entrega_cliente]), DAY)

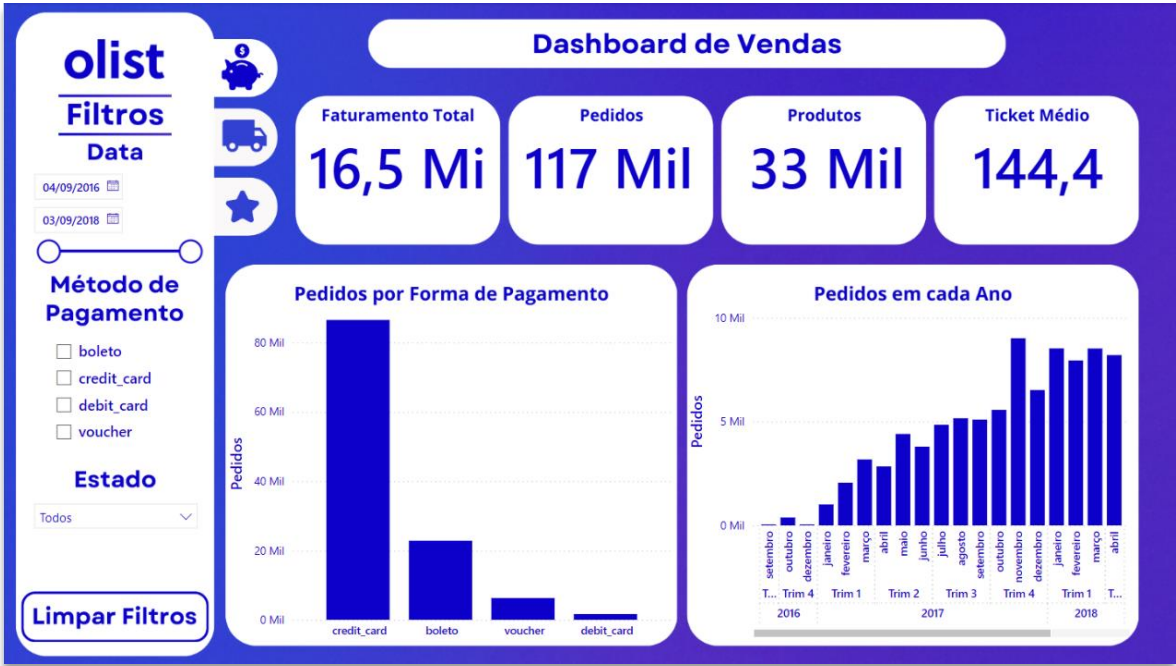
```

As fórmulas criadas servirão para criar medidas a serem utilizadas na criação do dashboard.

ANÁLISE DE VENDAS

O objetivo geral deste dashboard da figura 24, foi analisar os dados do negócio, visando compreender seu desempenho.

Figura 23 Dashboard de Vendas



FONTE: Autores (2024)

A seguir, serão apresentados detalhes de cada informação coletada no dashboard.

<div><div>Faturamento Total</div><div>16,5 Mi</div></div> <div>COLUNA CALCULADA <i>Medida:</i> Valor_arrecadado = SUMX (vendas, vendas[preco_produto] + vendas[frete_produto]).</div> <div>INDICADOR DE NÉGOCIO<ul style="list-style-type: none">• Visualização rápida do desempenho financeiro da empresa.• Monitoramento de metas de receita</div>	<div><div>Pedidos</div><div>117 Mil</div></div> <div>COLUNA CALCULADA <i>Medida:</i> Contagem de linhas da tabela de pedidos.</div> <div>INDICADOR DE NÉGOCIO<ul style="list-style-type: none">• Monitora o volume de vendas,• Identifica tendências de demanda• Avalia a eficiência operacional.</div>
---	---

<div data-bbox="320 293 730 584"> <p>Ticket Médio</p> <p>144,4</p> </div> <div data-bbox="252 600 587 633"> <p>COLUNA CALCULADA</p> </div> <div data-bbox="252 651 786 768"> <p><i>Medida:</i> ticket_médio = SUM (vendas [preco_produto])/DISTINCTCOUNT (vendas[cod_pedido])</p> </div> <div data-bbox="252 813 635 846"> <p>INDICADOR DE NÉGOCIO</p> </div> <div data-bbox="252 869 799 1111"> <ul style="list-style-type: none"> • Analisa o valor médio das vendas por pedido • Identifica tendências de consumo • Mede a eficácia das estratégias de vendas. • Direciona campanhas de marketing. </div>	<div data-bbox="951 293 1331 584"> <p>Produtos</p> <p>33 Mil</p> </div> <div data-bbox="868 600 1203 633"> <p>COLUNA CALCULADA</p> </div> <div data-bbox="868 667 1418 741"> <p><i>Medida:</i> Contagem de linhas da tabela de produtos.</p> </div> <div data-bbox="868 813 1251 846"> <p>INDICADOR DE NÉGOCIO</p> </div> <div data-bbox="868 869 1350 1016"> <ul style="list-style-type: none"> • Visão da variedade de produtos • Visão de disponibilidade. • Planejamento de compras </div>
--	---

<div data-bbox="252 1317 914 1888"> <p>Pedidos por Forma de Pagamento</p> <table border="1"> <thead> <tr> <th>Forma de Pagamento</th> <th>Quantidade de Pedidos (Mil)</th> </tr> </thead> <tbody> <tr> <td>credit_card</td> <td>~85</td> </tr> <tr> <td>boleto</td> <td>~25</td> </tr> <tr> <td>voucher</td> <td>~5</td> </tr> <tr> <td>debit_card</td> <td>~2</td> </tr> </tbody> </table> </div>	Forma de Pagamento	Quantidade de Pedidos (Mil)	credit_card	~85	boleto	~25	voucher	~5	debit_card	~2	<div data-bbox="951 1305 1315 1339"> <p>GRÁFICO EM COLUNAS</p> </div> <div data-bbox="951 1357 1430 1514"> <p>Eixo X: Tipo das formas de pagamento</p> <p>Eixo Y: Contagem da quantidade de pedidos</p> </div> <div data-bbox="951 1525 1337 1559"> <p>INDICADOR DE NÉGOCIO</p> </div> <div data-bbox="951 1576 1430 1910"> <ul style="list-style-type: none"> • Preferências dos clientes • Identificação de preferências em meios de pagamento • Gestão de fluxos de caixa e processos financeiros • Auxilia na tomada de decisões estratégicas para melhorar a experiência do cliente. </div>
Forma de Pagamento	Quantidade de Pedidos (Mil)										
credit_card	~85										
boleto	~25										
voucher	~5										
debit_card	~2										

GRÁFICO EM COLUNAS

Eixo X: Data do pedido

Eixo Y: Contagem de pedidos

INDICADOR DE NÉGOCIO

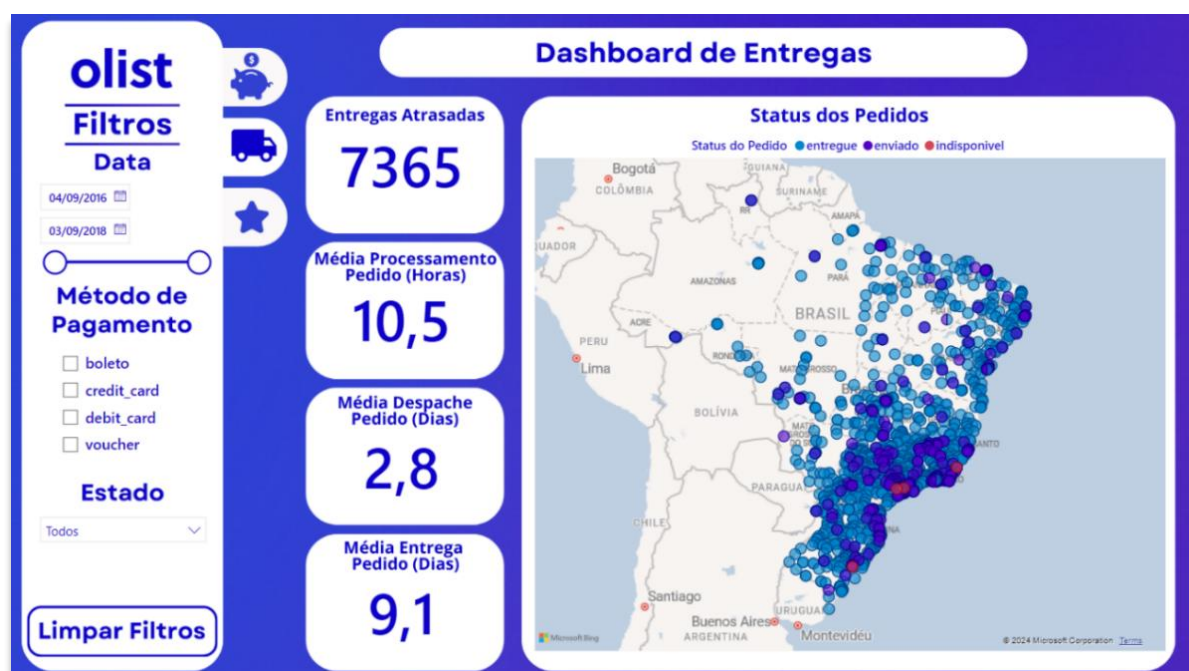
- Analisa o crescimento e sazonalidade das vendas.
- Auxilia na definição de estratégias de longo prazo.
- Avalia o impacto de campanhas e eventos específicos.
- Previsão de demanda



ANÁLISE DE DADOS LOGÍSTICOS

Analisar o desempenho logístico, permite acompanhar a capilaridade de entregas e o desempenho dos serviços terceirizados. A seguir na figura 24 e demais imagens, serão detalhadas todas as informações coletadas no dashboard.

Figura 24 Dashboard de Entregas



Fonte: Autores (2024)

Entregas Atrasadas**7365****FILTRO**

Medida: entregas_atrasadas =
COUNTROWS (FILTER(vendas,
vendas[dias_cliente_previsto] < 0))

INDICADOR DE NÉGOcio

- Monitora a eficiência logística
- Identifica problemas na cadeia de suprimentos.
- Informação de eventos adversos.

**Média Despache
Pedido (Dias)****2,8****COLUNA CALCULADA**

Medida: Média da coluna calculada de
tempo de despacho do pedido, do ven-
dedor a transportadora.

INDICADOR DE NÉGOcio

- Monitora a eficiência da logística de distribuição.
- Identifica atrasos no processo
- Apura prazos de entrega.

**Média Processamento
Pedido (Horas)****10,5****COLUNA CALCULADA**

Medida: Média da coluna calculada de
tempo de processamento dos pedidos.

INDICADOR DE NÉGOcio

- Mede a eficiência operacional
- Identifica gargalos no processo de atendimento.
- Analisa os fluxos de trabalho
- Mede o desempenho interno da organização.

**Média Entrega
Pedido (Dias)****9,1****COLUNA CALCULADA**

Medida: Média da coluna calculada de
tempo de entrega do pedido ao cliente.

INDICADOR DE NÉGOcio

- Avalia a eficiência da cadeia de suprimentos
- Permite identificar atrasos, melhorar a precisão dos prazos estimados e ajustar estratégias logísticas conforme necessário.

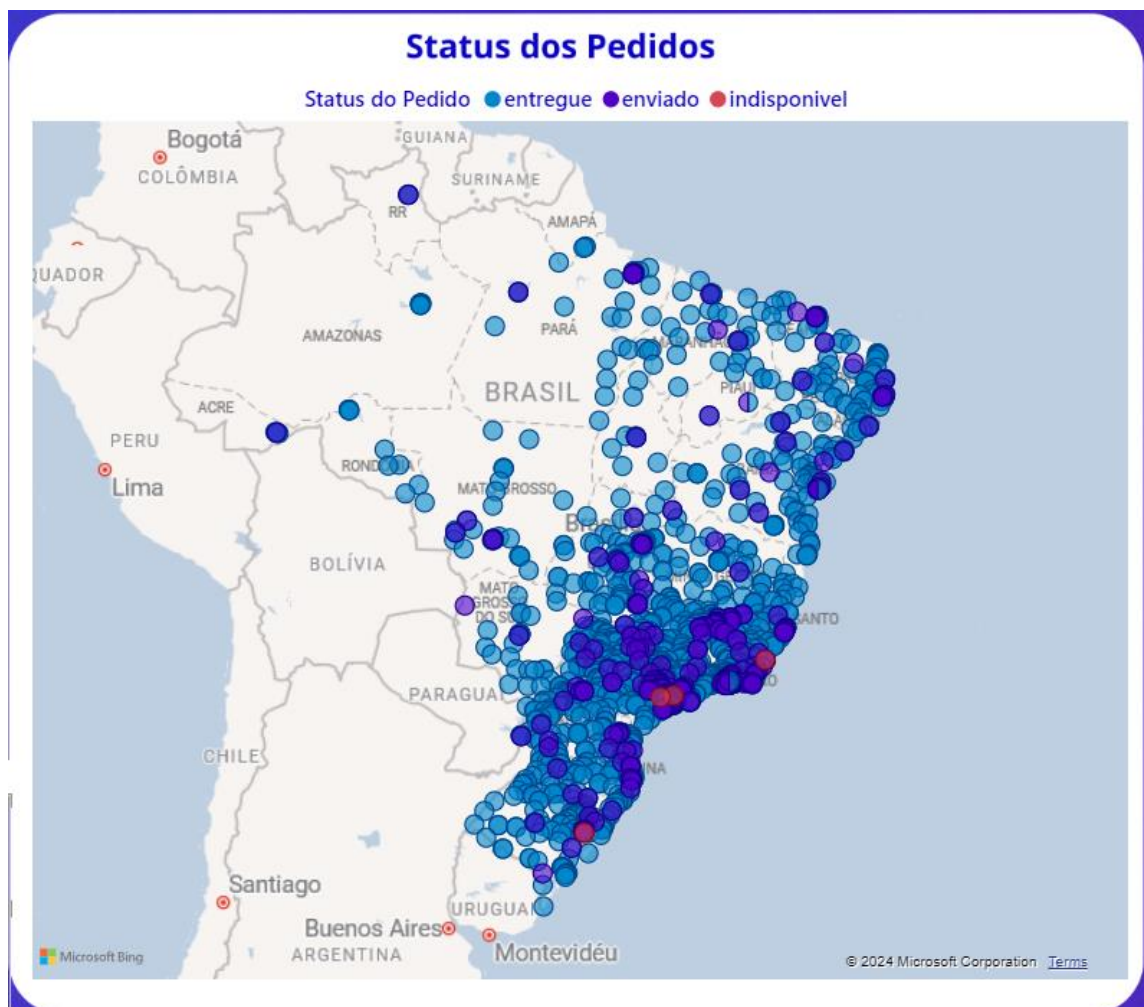


GRÁFICO EM MAPA

Localização: Latitude e Longitude

Bolhas: Status dos pedidos

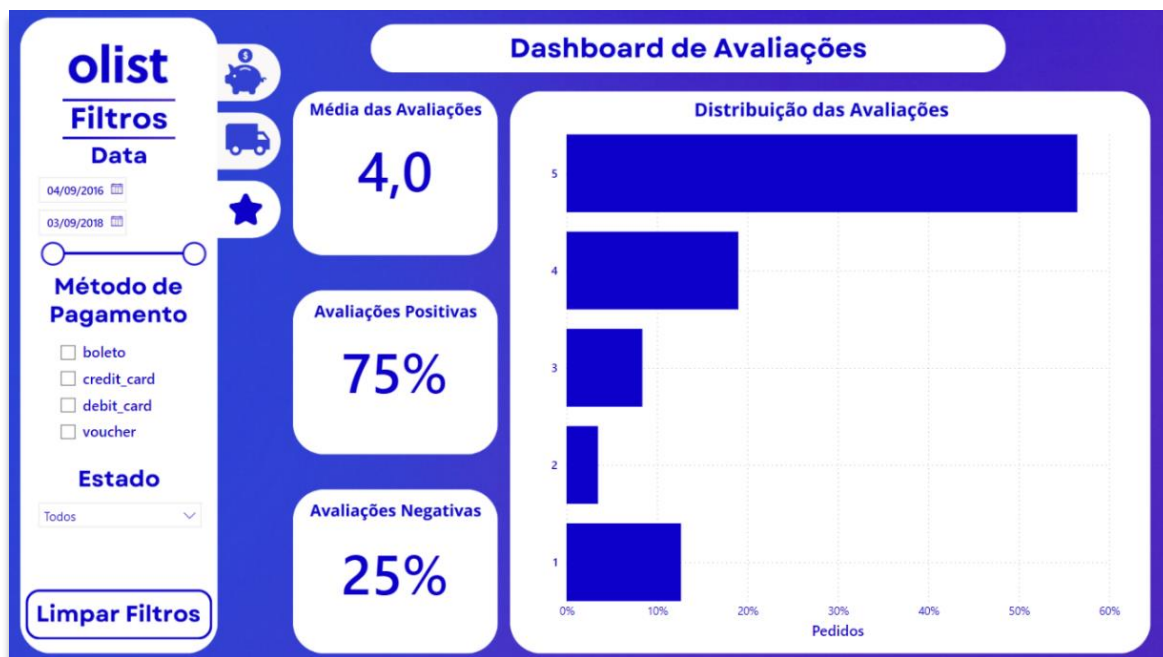
INDICADOR DE NÉGOCIO

- Visualização geográfica do progresso das entregas
- Identifica áreas com maior concentração de pedidos pendentes ou concluídos.
- Mapeia padrões regionais de demanda
- Contribui com a eficiência logística.
- Acompanha desempenho de prestadores de serviços terceirizados

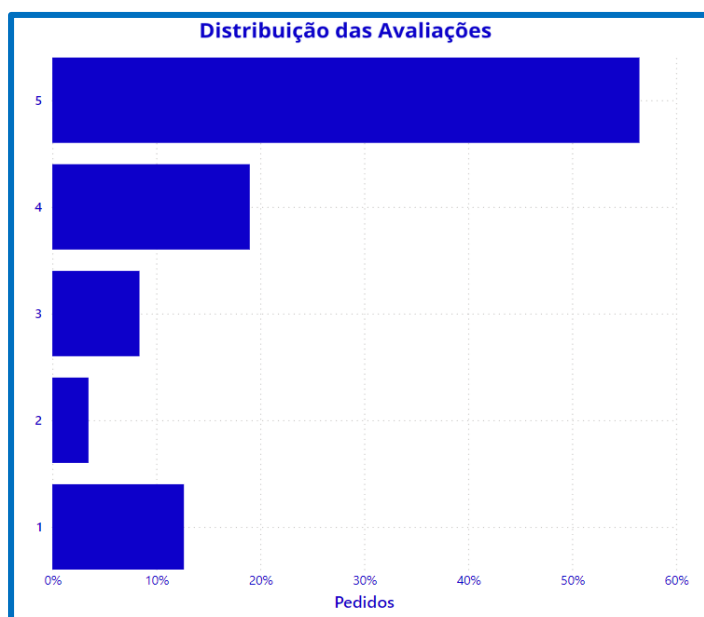
AVALIAÇÕES: SATISFAÇÃO DO CLIENTE

Para analisar a qualidade da prestação de serviços, a visualização traz os indicadores que medem a satisfação do cliente.

Figura 25 Dashboard de Avaliações



FONTE: Autores (2024)



Eixo X: Quantidade de pedidos

Eixo Y: Nota da avaliação

INDICADOR DE NÉGOCIO

- Mede a percepção dos clientes sobre produtos ou serviços.
- Visão da variação das notas
- Destaca áreas de excelência
- identifica de padrões de comportamento dos clientes
- Direciona de esforços para aprimorar a qualidade e a experiência do usuário.

<div data-bbox="359 235 694 510"> <p>Avaliações Positivas</p> <p>75%</p> </div> <div data-bbox="245 524 360 557">FILTRO</div> <div data-bbox="245 580 798 887"> <p><i>Medida:</i> PercentualBoasAvaliacoes = VAR TotalPedidos = COUNTROWS (vendas) VAR BoasAvaliacoes = COUNTROWS (FILTER (vendas, vendas[cod_avaliacao_pedido] = 4 vendas[cod_avaliacao_pedido] = 5)) RETURN DIVIDE (BoasAvaliacoes, TotalPedidos)</p> </div> <div data-bbox="245 958 635 992">INDICADOR DE NÉGOCIO</div> <div data-bbox="245 1014 804 1205"> <ul style="list-style-type: none"> • Avaliações de satisfação do cliente. • Acompanhar a qualidade do serviço prestado • Identificar áreas de excelência e oportunidades de melhoria. </div>	<div data-bbox="986 235 1321 510"> <p>Avaliações Negativas</p> <p>25%</p> </div> <div data-bbox="874 524 989 557">FILTRO</div> <div data-bbox="874 580 1426 925"> <p><i>Medida:</i> PercentualAvaliacoesNegativas = VAR TotalPedidos = COUNTROWS(vendas) VAR AvaliacoesNegativas = COUNTROWS (FILTER (vendas, vendas [cod_avaliacao_pedido] = 3 vendas[cod_avaliacao_pedido] = 2 vendas[cod_avaliacao_pedido] = 1)) RETURN DIVIDE (AvaliacoesNegativas, TotalPedidos)</p> </div> <div data-bbox="874 958 1262 992">INDICADOR DE NÉGOCIO</div> <div data-bbox="874 1003 1428 1227"> <ul style="list-style-type: none"> • Identifica pontos fracos no atendimento ao cliente • Acelera a resolução de problemas • Mapeia tendências negativas • Auxilia na implementação de estratégias corretivas. </div>
--	--

<div data-bbox="667 1303 1005 1579"> <p>Média das Avaliações</p> <p>4,0</p> </div> <div data-bbox="555 1585 898 1619">COLUNA CALCULADA</div> <div data-bbox="555 1624 1117 1697"> <p><i>Medida:</i> Média das notas de avaliação dos pedidos</p> </div> <div data-bbox="555 1742 948 1776">INDICADOR DE NÉGOCIO</div> <div data-bbox="555 1787 1117 1977"> <ul style="list-style-type: none"> • Nota média a satisfação do cliente ao longo do tempo. • Identifica áreas de melhoria • Destaca pontos fortes no atendimento. • Monitora a reputação da marca. </div>
--

8.8. Inteligência Competitiva aplicada ao Projeto

Nos últimos anos, a crescente disponibilidade de dados e o avanço das tecnologias de armazenamento e processamento têm revolucionado a forma como as organizações tomam decisões estratégicas.

Nesse contexto, a Inteligência Competitiva (IC) emerge como um componente essencial para o sucesso empresarial, permitindo que as empresas colem, processem e analisem informações relevantes sobre o mercado, concorrentes e tendências, a fim de obter insights valiosos para direcionar suas ações.

O Data Warehouse, importante componente de sistemas de IC, integra dados de diversas fontes internas e externas, consolidando-os em um único repositório para análise e consulta. Essa abordagem proporciona às empresas uma visão unificada e completa de suas operações, facilitando a geração de insights e a tomada de decisões de maneira concisa e direta.

Os gestores conseguem, portanto, concentrar suas ações na administração do negócio, desenvolvendo estratégias conforme o cenário apresentado, a fim de obter vantagem competitiva no mercado, além de permanecer constantemente atualizados.

8.9. Inteligência Artificial (IA) Aplicada ao Projeto

O projeto apresenta ainda, as técnicas de Machine Learning aplicadas a análise de dados de clientes, com o propósito de aprimorar a compreensão de seus comportamentos e, conseqüentemente, impulsionar estratégias de marketing direcionadas e eficazes.

Para atingir esse fim, empregamos métodos como Ensemble para classificar a satisfação dos clientes e Clustering para agrupar clientes com base em dados RFM (Recency, Frequency, Monetary).

Para avaliar a satisfação dos clientes, utilizamos o XGBoost, um algoritmo de Ensemble reconhecido por sua eficácia em problemas de classificação e regressão.

O XGBoost ajusta os parâmetros do modelo sequencialmente, empregando o gradiente descendente e técnicas de regularização para produzir previsões precisas e robustas.

Os resultados iniciais revelaram uma acurácia de 86% no conjunto de treino e 81% no conjunto de teste. Visando aprimorar esses resultados, realizamos uma otimização de hiperparâmetros por meio do GridSearch, o que resultou em uma acurácia de 91% no conjunto de treino e 82% no conjunto de teste, representando uma melhora de 5% e 1%, respectivamente.

Figura 26 Classification Report – XGBoost

CR de Treino				
	precision	recall	f1-score	support
0	0.96	0.76	0.85	69338
1	0.80	0.97	0.88	69338
accuracy			0.86	138676
macro avg	0.88	0.86	0.86	138676
weighted avg	0.88	0.86	0.86	138676
CR de Teste				
	precision	recall	f1-score	support
0	0.69	0.30	0.42	5183
1	0.82	0.96	0.88	17335
accuracy			0.81	22518
macro avg	0.75	0.63	0.65	22518
weighted avg	0.79	0.81	0.78	22518

FONTE: Autores (2024)

Figura 27 Classification Report – Com ajuste GridSearch

CR de Treino				
	precision	recall	f1-score	support
0	0.99	0.83	0.90	69338
1	0.85	0.99	0.92	69338
accuracy			0.91	138676
macro avg	0.92	0.91	0.91	138676
weighted avg	0.92	0.91	0.91	138676
CR de Teste				
	precision	recall	f1-score	support
0	0.75	0.35	0.47	5183
1	0.83	0.97	0.89	17335
accuracy			0.82	22518
macro avg	0.79	0.66	0.68	22518
weighted avg	0.81	0.82	0.80	22518

FONTE: Autores (2024)

Para segmentar os clientes com base nos dados RFM, aplicamos métodos de Clustering, especialmente o K-means, em conjunto com o método do cotovelo para determinar o número ideal de clusters.

RFM (Recency, Frequency, Monetary) - técnica de segmentação de clientes usada em marketing e análise de dados, classifica clientes com base em três métricas: Recência (tempo desde a última interação), Frequência (número de interações) e Valor Monetário (total gasto).

Combinando essas métricas, o RFM segmenta os clientes em grupos, permitindo a personalização de estratégias de marketing e o direcionamento para clientes com maior potencial de retorno. Isso nos permitiu identificar quatro grupos distintos de clientes: clientes perdidos/baixos gastadores, grandes gastadores, novos entrantes/pequenos gastadores e clientes fiéis.

Esses clusters proporcionam informações valiosas para personalizar estratégias de marketing, permitindo o direcionamento de esforços para os segmentos de clientes mais relevantes e com maior potencial de retorno.

Figura 28 Cluster Report

	Recency	Frequency	Monetary	
	mean	mean	mean	count
Cluster				
0	417.0	1.0	121.0	27599
1	191.0	1.0	519.0	22348
2	139.0	1.0	82.0	39570
3	219.0	2.0	486.0	2736

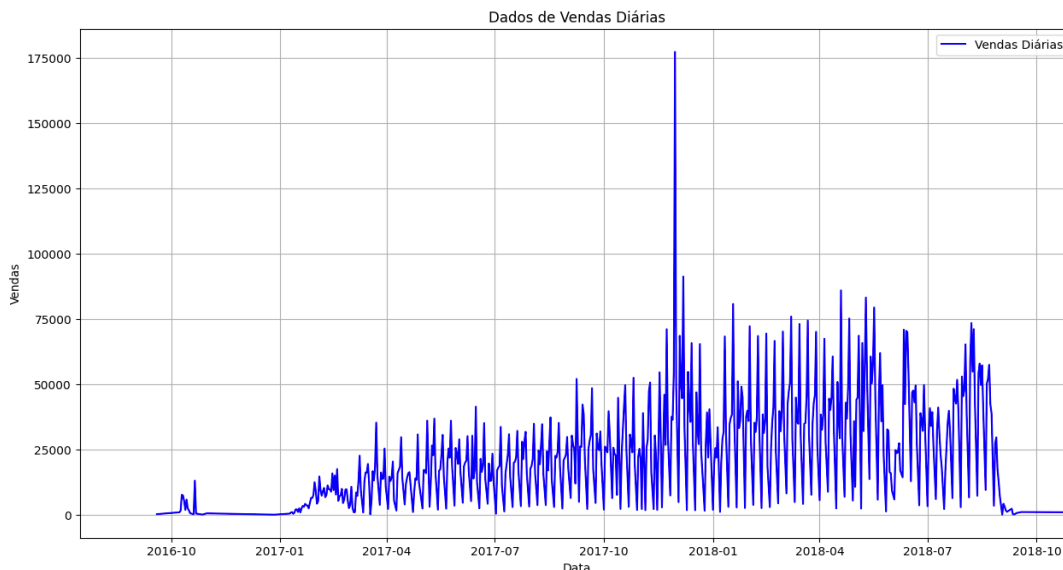
FONTE: Autores (2024)

Para prever as quantidades de vendas de um período, utilizamos um modelo de Redes Neurais Recorrentes (RNN), especificamente uma Long Short-Term Memory (LSTM), implementado com a API Sequencial do Keras.

LSTM é uma variação avançada de RNNs projetada para aprender dependências de longo prazo em dados sequenciais, o que a torna ideal para séries temporais como as vendas.

Primeiro, preparamos os dados históricos de vendas, que foram normalizados para garantir que todos os valores estivessem dentro de um intervalo específico. Em seguida, organizamos os dados em sequências temporais, onde um conjunto de valores passados (entradas) foi usado para prever valores futuros (saída).

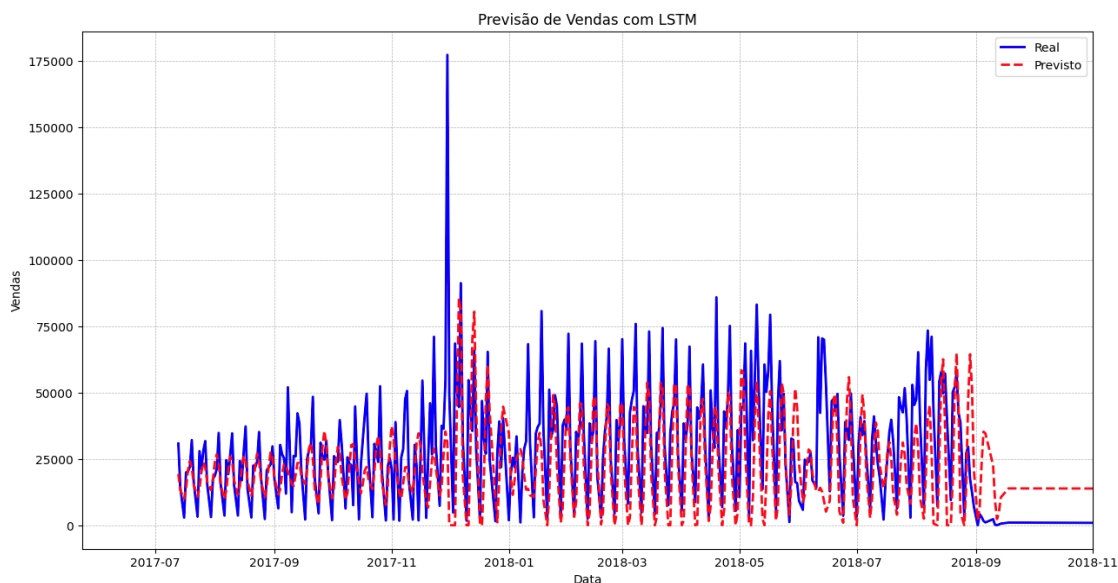
Figura 29 Dados de vendas diárias



FONTE: Autores (2024)

Construímos o modelo LSTM utilizando a API Sequencial do Keras. O treinamento do modelo foi realizado ajustando os pesos através da minimização do erro entre as previsões do modelo e os valores reais. Levando a esse resultado:

Figura 30 Previsão de Vendas com LSTM



FONTE: Autores (2024)

9. CONCLUSÃO

Este projeto teve como objetivo principal proporcionar uma infraestrutura robusta para a gestão e análise de dados, possibilitando a extração de insights valiosos e a tomada de decisões mais informadas.

A equipe optou pelo Microsoft Planner para cronograma, a fim de cumprir requisitos específicos de Gestão de Projeto, para melhorar a organização do trabalho.

Optamos por fazer o ETL e a criação do DW via pgAdmin porque já dominávamos a linguagem e a ferramenta (Open Source) para fazê-lo, enquanto, se optássemos pelas ferramentas pesquisadas levaríamos tempo aprendendo seu funcionamento e/ou pagaríamos pelo serviço utilizado.

A construção de um *Data warehouse* para a empresa Olist, fundamentada nos conhecimentos adquiridos nas disciplinas de Inteligência Competitiva, Estatística, Programação em Banco de Dados e Inteligência Artificial II revela-se uma iniciativa estratégica de grande relevância.

Ao longo do desenvolvimento do Data Warehouse, foram aplicadas técnicas de Inteligência Competitiva para identificar e monitorar tendências de mercado, concorrentes e comportamento do consumidor.

Essas informações foram integradas no Data Warehouse, permitindo que a Olist se mantenha competitiva e inovadora em um ambiente de negócios dinâmico e em constante transformação.

A integração dos conhecimentos de Inteligência Competitiva e IA não só enriqueceu o processo de desenvolvimento do Data Warehouse, como também estabeleceu um diferencial competitivo sustentável para a empresa.

A partir deste ponto, a Olist está melhor equipada para enfrentar os desafios do mercado e capitalizar oportunidades com maior precisão e agilidade.

REFERÊNCIAS BIBLIOGRÁFICAS

ARBEX, Gabriela. **E-commerce global pode movimentar US\$ 34 tri em 2025**. Disponível em <https://forbes.com.br/forbes-tech/2021/01/e-commerce-global-pode-movimentar-us-34-tri-em-2025-retailtechs-americanas-brf-muito-mais/> Acesso em 29/05/2024.

ASTERA. **O que é arquitetura de data warehouse?** Disponível em <https://www.astera.com/pt/knowledge-center/data-warehouse-architecture/> . Acesso em 26/05/2024.

AWS 1. **O que é um data warehouse?** Disponível em <https://aws.amazon.com> . Acesso em 21/03/2024.

AWS 2. **Bancos de Dados na nuvem AWS**. Disponível em: <https://aws.amazon.com/pt/products/databases>. Acesso em 21/03/2024.

AWS 3. **Redshift** Disponível em: <https://aws.amazon.com/pt/redshift/> . Acesso em 20/05/2024

DEVMEDIA. **Pentaho BI** Disponível em <https://www.devmedia.com.br/> Acesso em 15/04/2024

GOMEZ, José María Gómez, RLDAN María Carina. **Pentaho for Developers: A beginner's guide to the Pentaho business analytics platform"** (2013).

INTEGRATION SERVICES. Disponível em: <https://learn.microsoft.com/pt-br/sql/integration-services/sql-server-integration-services?view=sql-server-ver16> Acesso em 15/04/2024.

KAGGLE. **Kaggle**. Disponível em <https://www.kaggle.com/datasets/olistbr/> Acesso em 10/03/2024.

KIMBAL, R. et al. **The Data warehouse Lifecycle Toolkit**: Expert Methods for Designing, Developing, and Deploying Data Warehouses. John Wiley & Sons, New York, 1998. Harrison, T.H. Intranet data warehouse, Berkeley, 1998.

MEDIUM. **Modelagem multidimensional**, disponível em [Medium – Where good ideas find you.](https://medium.com/pt-br/ideas-find-you) / Acesso em 21/03/2024.

MICROSOFT. **Power BI disponível em:** <https://www.microsoft.com/pt-br/power-platform/products/power-bi> Acesso em: 14/04/2024.

OLIST. **Olist – sobre nós** -Disponível em <https://olist.com> , acesso em 24/03/2024.

PMBOK. **Guia do Conhecimento em Gerenciamento de Projetos**. PMI, Global Standad, 2021.

PENTAHO. **Dashboards Pentaho**, disponível em: <https://pentaho-bi-suite.blogspot.com/2013/12/pentaho-cdedashboard-complete-example.html> Acesso em 16/04/2024

SAS. **Visualização de dados**, disponível em: https://www.sas.com/pt_br/insights/big-data/data-visualization.html .Acesso em 20/04/2024.

SNOWFLAKE. **Snowflake para Data Warehouse**. Disponível em [Snowflake para Data warehouse | Cargas de trabalho Snowflake/](#). Acesso em 23/03/2024.

TABLEAU. **Software de Análise e Business Intelligence**. Disponível em: <https://www.tableau.com/pt-br>. Acesso em 20/05/2024.

VOITTO. **O que é Pentaho** Disponível em <https://www.voitto.com.br/> . Acesso em 15/04/2024

ANEXOS

[PI3 DW Olist entrega ANEXOS.pdf](#)

https://app.powerbi.com/groups/me/reports/7ccc0924-32eb-40e4-8e04-8d83f763220a?ctid=cf72e2bd-7a2b-4783-bdeb-39d57b07f76f&pbi_source=linkShare

[ANEXO MACHINE LEARN.pdf](#)