

Relatório Adventure Works

Projeto de Dados
Março de 2024

1. Introdução

Neste relatório, estão sendo detalhados os passos necessários para a realização do projeto de dados da empresa Adventure Works (AW). Foi conduzida uma abrangente análise dos dados para extrair *insights* coerentes. Dessa forma, o projeto possibilita a utilização dos dados de forma estratégica para garantir o crescimento da empresa e o consolidar uma posição competitiva em relação aos concorrentes.

Portanto, a metodologia empregada, conhecida como Modern Data Stack (MDS), visa desenvolver a capacidade analítica da empresa, por meio da combinação de boas práticas, ferramentas específicas e tecnologias inovadoras para a criação de infraestruturas de dados flexíveis, adaptáveis e acessíveis.

1.1. Objetivos

1. Realizar a transformação e análise dos dados brutos da Adventure Works para extrair *insights* relevantes e coerentes.
2. Utilizar os *insights* obtidos para orientar decisões estratégicas para impulsionar o crescimento da empresa e fortalecer sua posição competitiva no mercado.
3. Desenvolver e implementar uma infraestrutura de dados flexível, adaptável e acessível, alinhada à metodologia Modern Data Stack (MDS).
4. Promover a cultura *data driven* e aumentar a capacidade analítica da empresa.

2. Planejamento do Projeto

Para implementação do projeto e criação de infraestruturas de dados foi utilizada uma combinação de boas práticas e ferramentas, definida como Modern Data Stack (MDS). Dessa forma, é possível combinar ferramentas que executam funções distintas para construir uma estrutura de dados moderna, mutável e mais independente. Logo, para execução do projeto foram seguidas etapas e atribuições específicas ao longo de

uma pipeline. A metodologia utilizada no planejamento e execução do projeto pode ser resumida no diagrama a seguir.



2.1. Coleta, integração e limpeza de dados

Nesta etapa, foram diagnosticados os sistemas utilizados pela Adventure Works para gerar dados relevantes para o negócio e que devem ser integrados à infraestrutura de dados. Foram coletados dados de múltiplas fontes, como, CRM, ERP, web analytics e sites. Logo, os ajustes necessários foram realizados para que os dados estejam preparados para a etapa seguinte.

2.2. Armazenamento e gerenciamento de dados

Para garantir armazenamento dos dados e consultas de grandes volumes de dados, foi utilizado o Data Warehouse **Google BigQuery**, que possibilita análises e o gerenciamento de informações em larga escala.

2.3. Transformação de dados

Para ganhos de eficiência no projeto, foi realizada a abordagem ELT (Extract-Load-Transform), na qual, os dados foram extraídos, carregados e em seguida, transformados. Trata-se de uma abordagem mais rápida e flexível para a transformação de dados brutos em dados modelados dentro do Data Warehouse.

A principal ferramenta utilizada para a transformação dos dados foi o **DBT** (data build tool). A tecnologia permite executar o processo de ELT para a transformação dos dados brutos em dados modelados, garantindo um código analítico, capaz de produzir resultados de maior qualidade, realizar testes e documentar consultas.

2.4. Business Intelligence e Data Analytics

Para fornecer a visualização, as análises de informações e gerar insights que agregam valor à tomada de decisão empresarial foi utilizada a ferramenta de *Business*

Intelligence **Power BI**. Com isso, os gestores podem analisar os principais indicadores, otimizar processos e tomar decisões assertivas.

2.5. Advanced Analytics

Por fim, para análises mais complexas, foi utilizado o conceito de *Machine Learning*. Fez-se uso de diversas bibliotecas nas linguagens **Python**, a fim de realizar a previsão sobre a demanda dos produtos em cada loja ter uma melhor estimativa sobre a necessidade de compra de matéria prima.

3. Transformação dos Dados

A etapa de transformação dos dados consiste em garantir a qualidade e utilidade das informações, de modo que, os dados brutos serão transformados em dados modelados. A priori, foi realizada a conexão com o Google BigQuery, para acessar os dados brutos armazenados.

Os modelos de dados foram organizados em camadas distintas, incluindo stagings, intermediates e marts. Na camada staging, os dados são recebidos da fonte e são realizados pequenos ajustes, de modo que as informações iniciais sejam preservadas. Na intermediate, os dados são refinados para atender necessidades específicas. Por fim, na camada marts, os dados são organizados em formato otimizado e são criadas as tabelas de fato, dimensão e agregadas, conforme, o modelo star schema.

Neste projeto, foram criadas as tabelas de dimensão “dim_creditcards”, “dim_customers”, “dim_dates”, “dim_locations”, “dim_products” e “dim_reasons”. Além disso, foi criada uma tabela agregada de vendas por região e vendedor e uma tabela fato de vendas, utilizada como ponto central para análises posteriores. Além disso, foi incluída uma documentação detalhada dos processos e a criação de testes de schema e de dados, para garantir a integridade e consistência dos dados.

Por fim, uma vez que os dados foram modelados de acordo com os requisitos propostos, esses foram disponibilizados para uma análise utilizando ferramentas de Business Intelligence, permitindo a visualização e exploração dos dados. Além disso, foram aplicadas técnicas de análises avançadas aos dados modelados e desenvolvido um modelo para previsão de demanda de produtos.

4. Partes Interessadas

O projeto é composto por diversas partes interessadas que possuem diferentes papéis e perspectivas sobre o projeto:

- Diretor de inovação (João Muller): Idealizador e patrocinador do projeto. Identifica o uso de dados como uma oportunidade estratégica para impulsionar a inovação e o crescimento da empresa.
- CEO (Carlos Silveira): Reconhece o diferencial estratégico do uso dos dados em longo prazo.
- Diretora Comercial (Silvana Teixeira): Acredita que os recursos investidos no projeto poderiam ser direcionados para ações promocionais com resultados de vendas imediatos. Além disso, questiona a utilidade da infraestrutura de dados para o departamento comercial.
- Gestor de Planejamento de Demanda (Luís Soares): Interessado em uma análise preditiva para melhorar o planejamento da produção.

5. Riscos e Contingências

Durante o desenvolvimento do projeto de dados, foram identificados riscos que podem afetar sua execução e os resultados esperados:

- Falta de adesão das partes interessadas: A falta de adesão e o questionamento sobre a utilidade da infraestrutura de dados pode atrasar os prazos e comprometer a qualidade dos resultados.
- Tempo de realização do projeto: O prazo para conclusão pode comprometer a entrega, visto, o extenso escopo previsto.
- Limitação de dados para análise preditiva: Dados históricos incompletos sobre a demanda de produtos em cada loja pode dificultar a previsão precisa e confiável.

Para mitigar os riscos e contingências e garantir a qualidade da entrega, pode-se implementar as seguintes estratégias:

- Engajamento das partes interessadas: Estabelecer uma comunicação clara e regular com todos os interessados fornecendo informações detalhadas sobre o desenvolvimento do projeto.
- Tempo de desenvolvimento: Manter um monitoramento contínuo, a fim de identificar rapidamente quaisquer problemas e impedimentos, o que permitirá a tomada de medidas proativas e uma entrega de qualidade no prazo estabelecido.
- Coleta adicional de dados: Realizar reuniões periódicas para entender e complementar as informações disponíveis.

6. Geração de Valor

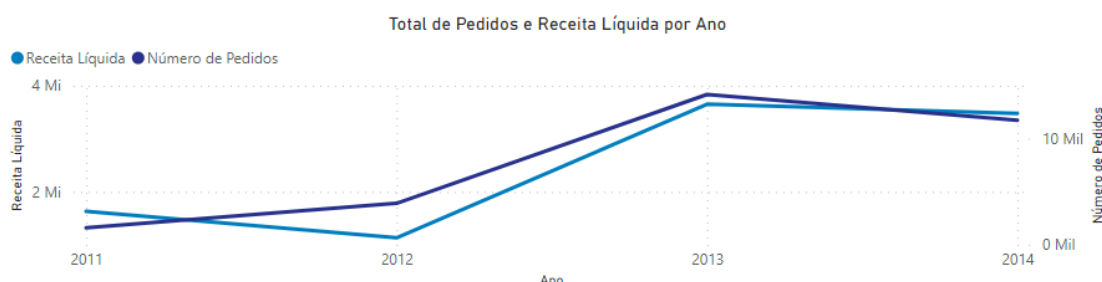
A geração de valor em um projeto de infraestrutura de dados é necessário, além de implementar sistemas e ferramentas, construir abordagem de analytics moderna e eficiente. Dessa forma, é possível agregar ao negócio:

- **Qualidade e Integridade dos Dados:** São fatores essenciais para agregar valor ao projeto e permitem tomadas de decisão mais assertivas.
- **Governança:** Informações centralizadas e acessíveis, permitindo o gerenciamento dos dados de forma íntegra.
- **Testes:** A realização de testes nos projetos de dados,, garantem a consistência e a confiabilidade dos resultados.
- **Acessibilidade dos Dados:** Garantir que os dados estejam acessíveis a todas as partes interessadas é fundamental.
- **Gestão de Riscos:** A identificação e gestão eficaz dos riscos são fundamentais para a geração de valor em um projeto de infraestrutura de dados.
- **Entrega de produto e cumprimento de metas:** Garantir a entrega do produto e o cumprimento das metas estabelecidas são aspectos fundamentais para assegurar a entrega de valor ao cliente.

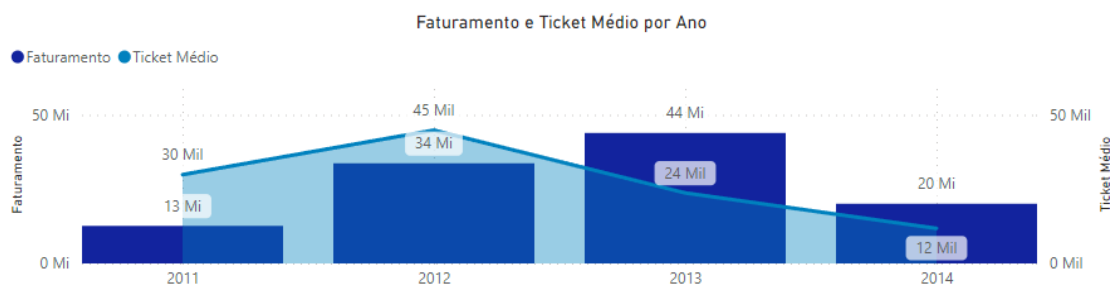
7. Análises de Dados

7.1. Business Intelligence

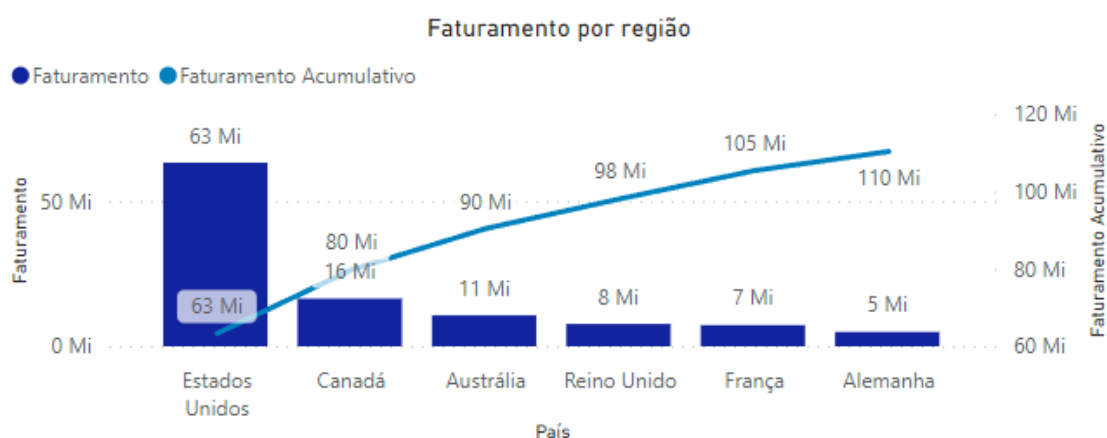
As vendas totais da Adventure Works, entre os anos de 2011 e 2024, atingiram 31465 pedidos, com concentração de vendas no ano de 2013. A receita líquida (descontando impostos e fretes) foi de US\$ 9,9 milhões:



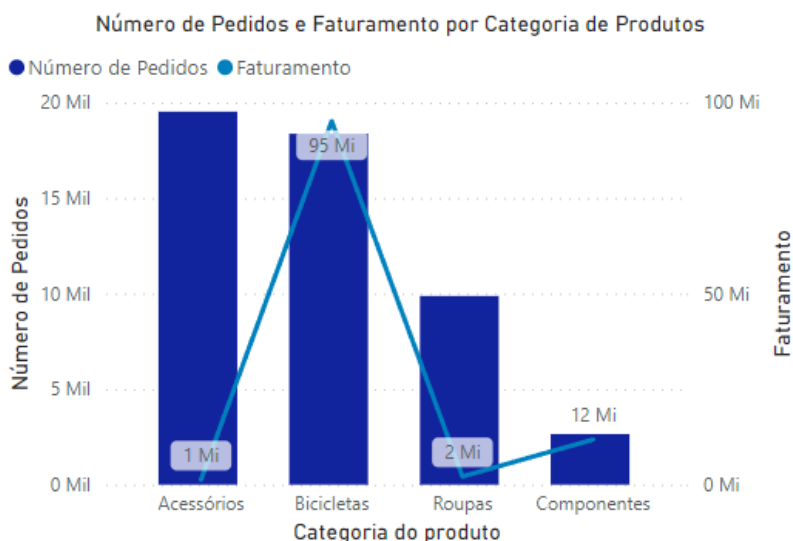
Em questões de faturamento, a empresa atingiu, aproximadamente, US\$110,37 milhões. Além disso, o ticket médio da empresa foi de US\$24,13 mil.



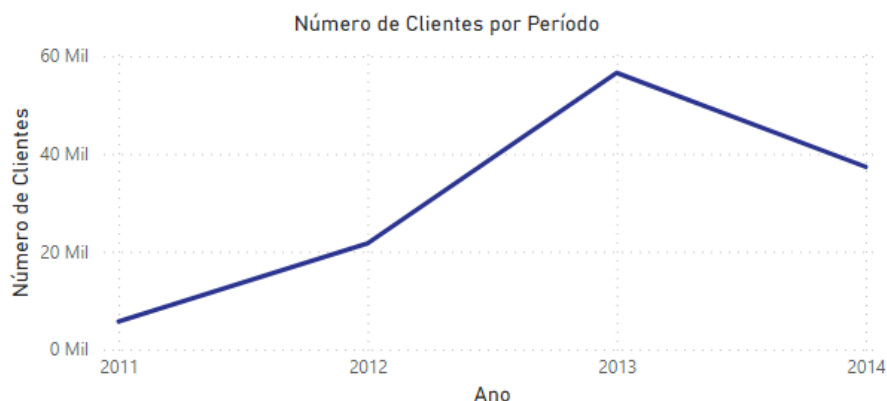
A empresa tem presença comercial em seis países distintos, abrangendo um total de 535 cidades. O país que possui a maior parcela de mercado é os Estados Unidos, com 12 mil pedidos e US\$63 milhões de faturamento. Vale destacar que o faturamento gerado pelos Estados Unidos supera o total obtido pelos demais países.



A categoria de produtos mais demandada foi a de acessórios, com um total de 19.524 pedidos. No entanto, apesar da menor quantidade de pedidos, a categoria de bicicletas se destacou por representar o maior faturamento, totalizando cerca de US\$95 milhões.



A empresa conta 19 mil clientes ao todo e possibilita compras realizadas em canais online ou presencial, sendo 87,9% dos pedidos realizados de forma online. O método de pagamento realizado pelos clientes é integralmente cartão de crédito.



Além disso, é possível observar que dentre os principais motivos para a efetivação da compra, tem-se o preço sendo o principal fator relatado pelos clientes. Seguido por promoções e qualidades. O que demonstra que esses são fatores que os clientes consideram ao efetuar a compra.

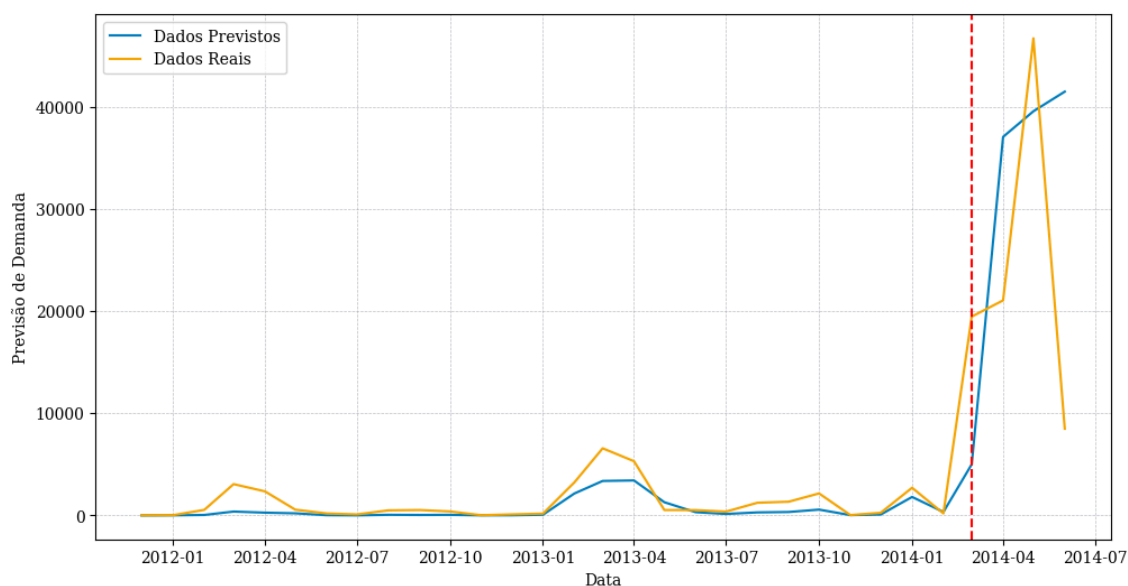


Portanto, é possível identificar as regiões que demandam maior atenção, compreender os fatores que influenciam as decisões de compra dos clientes e detectar aqueles clientes que estão inativos. A partir dessas análises, é possível desenvolver estratégias para reativar esses clientes e atrair novos, visando aprimorar o desempenho das vendas e aumentar o lucro da empresa.

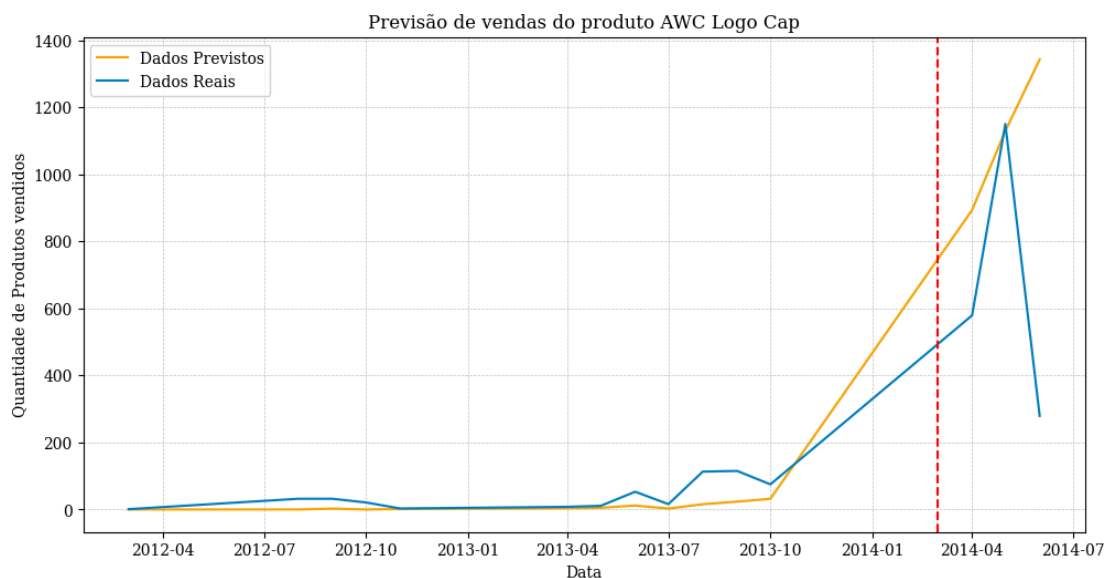
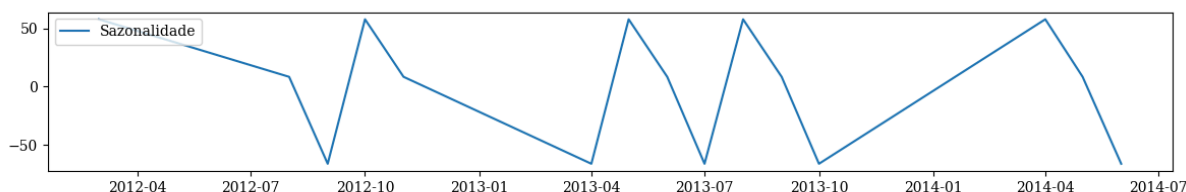
7.2. Previsão de Demanda

Para o planejamento da demanda da organização, foi realizada uma análise com o modelo AutoARIMA. Por meio de análises preditivas, é possível prever a demanda dos

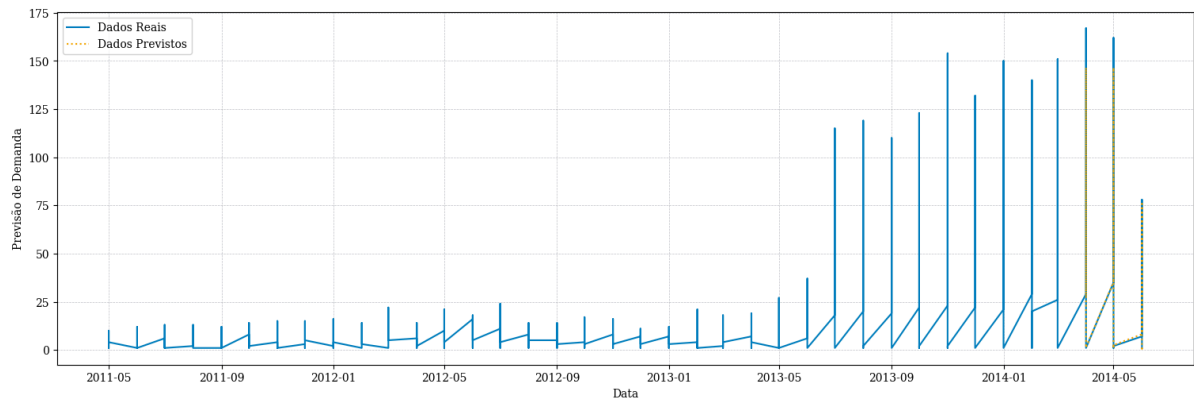
próximos 3 meses para cada produto em cada loja. Essa previsão permite ajustar estrategicamente a distribuição de produtos e obter estimativas mais precisas sobre a necessidade de reposição de matéria prima. Assim, realizou-se a previsão de demanda para determinar as quantidades de produtos necessários nos próximos 3 meses.



Além disso, foi analisada a sazonalidade das vendas do produto AWC Logo Cap. Essa análise permite identificar padrões recorrentes nos dados de vendas ao longo do tempo. Com isso, é possível ajustar estratégias de produção e distribuição para melhor atender à demanda sazonal e otimizar o desempenho geral das vendas.



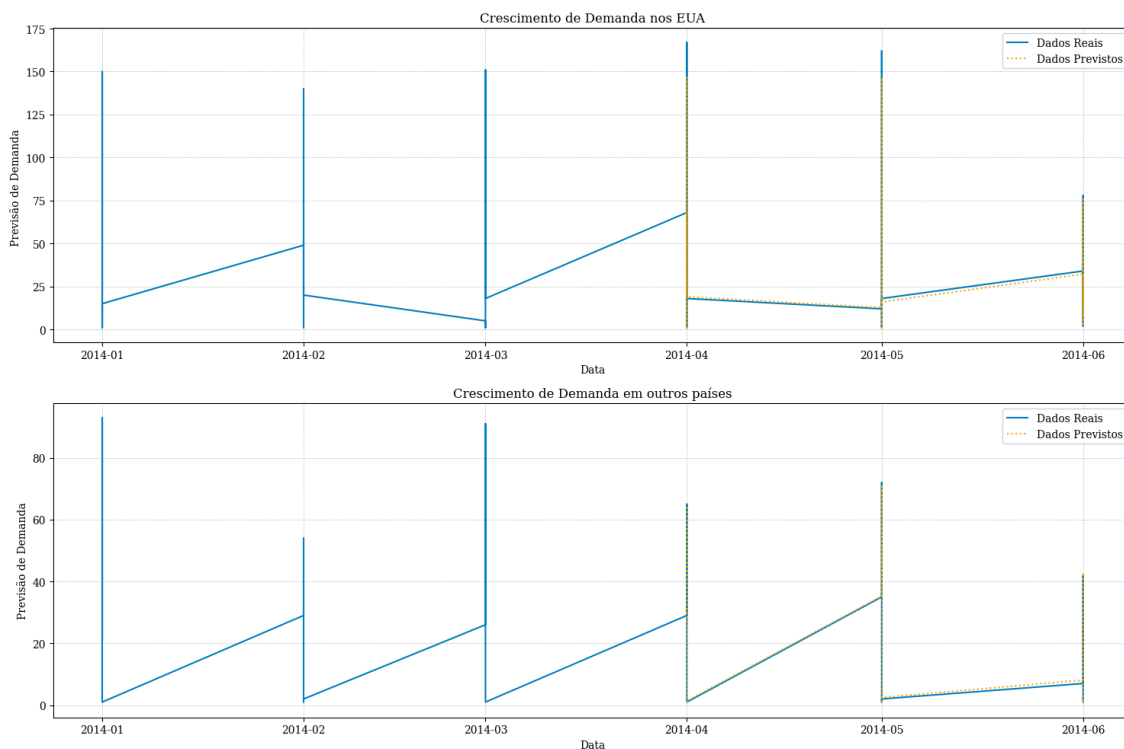
Para garantir a precisão e confiabilidade dos resultados, adotamos a abordagem de regressão utilizando o modelo XGBoost. Este modelo demonstrou resultados altamente significativos na previsão de demanda.



Em seguida, foi avaliado o desempenho do modelo por meio de métricas, como MSE (Erro Quadrático Médio), RMSE (Raiz do Erro Quadrático Médio) e MSA (Erro Absoluto Médio). Os resultados obtidos com o XGBoost apresentaram valores significativamente melhores para as métricas demonstrando sua efetividade.

	RMSE	MSE	MAE
AutoARIMA	148,43	22030,77	16,06
XGBoost	1,13	1,28	0,63

Além disso, foi analisado o crescimento da demanda nos últimos três meses em diferentes centros de distribuição. Pela análise é possível constatar uma tendência de queda na demanda para o próximo trimestre. Além disso, os centros de distribuição localizados nos EUA demonstraram uma demanda inferior em comparação com os novos centros de distribuição estabelecidos em outros países ao redor do mundo.



Por fim, determinou-se que para um novo fornecedor de luvas, deve pedir 620 zíperes para os próximos 3 meses, sabendo que são necessários 2 zíperes por par de luvas.

8. Entregas

- Projeto no GitHub: <https://bit.ly/4c72N3B>
- Modelo Conceitual: <https://bit.ly/3wL5Azp>
- Dashboard : <https://bit.ly/3T4pWv4>
- Relatório: <https://bit.ly/3P9uoaw>
- Notebook de desenvolvimento - Previsão de Demanda: <https://bit.ly/49GGr7E>