**PONTIFÍCIA UNIVERSIDADE CATÓLICA DE MINAS GERAIS  
NÚCLEO DE EDUCAÇÃO A DISTÂNCIA**

**Pós-graduação *Lato Sensu* em Business Intelligence e Analytics**

**Flávia Elias da Silva**

ANÁLISE DE VENDAS DA BASE DE DADOS NORTHWIND

Belo Horizonte

2021

**Flávia Elias da Silva**

**ANÁLISE DE VENDAS DA BASE DE DADOS NORTHWIND**

Trabalho de Conclusão de Curso apresentado ao Curso de Especialização em Business Intelligence e Analytics como requisito parcial à obtenção do título de especialista.

Belo Horizonte

2021 **RESUMO**

O resumo deve apresentar, de forma sucinta, os pontos relevantes do trabalho em um parágrafo único e espaço entre linhas de 1,5. O resumo deve conter a apresentação do problema, uma descrição sucinta da aplicação construída, os achados mais relevantes e as experiências adquiridas no processo de construção. O resumo também deve apresentar resultados de testes e de usabilidade. O resumo deve conter de 100 a 250 palavras.

Logo abaixo do resumo, devem-se pontuar as palavras-chave que representam o conteúdo do estudo. São separadas entre si com ponto final e finalizadas também com ponto final.

Atualmente empresas precisam estar preparadas para tomar decisões rápidas e assertivas a cerca dos assuntos relacionados ao negócio em geral. A ideia deste trabalho é apresentar, de forma simples e objetiva, como organizações podem se munir de informações analíticas para ter controle estratégico e operacional da visão de negócios. Foram construídos painéis que conversam com a camada estratégica da companhia passando pelos assuntos táticos e chegando na camada mais detalhista da operação. Neste projeto foi utilizado o famoso *dataset* da Northwind Traders. empresa fictícia criada pelo Microsoft com intuito de servir de material de estudo para estudantes, profissionais e demais pessoas interessadas em construção de análise de dados. A camada analítica avança para a parte de *Machine Learning* com informações sobre recomendação de produtos para os clientes da Northwind. Por fim, sugestões de estratégias são apresentadas a partir do entendimento analítico dos dados da empresa que parte desde a camada logística até passando pelos vendedores e demais funcionários da empresa.

Palavras-chave: Análise de dados. *Dashboard*. *Northwind*. Vendas.*Machine Learning.*

**SUMÁRIO**

[1. Introdução 6](#_Toc43168386)

[1.1. Contexto 6](#_Toc43168387)

[1.2. Público alvo 7](#_Toc43168388)

[2. Modelo de Dados 7](#_Toc43168389)

[2.1. Modelo Dimensional 7](#_Toc43168390)

[2.2. Fatos e Dimensões 8](#_Toc43168391)

[3. Processo ETL 11](#_Toc43168392)

[4. Camada de Apresentação 12](#_Toc43168393)

4.1.Painel estratégico.................................................................................................13

4.2. Painel tático.........................................................................................................14

4.3. Painel operacional...............................................................................................14

4.4. Machine Learning................................................................................................15

[5. Registros de Homologação 16](#_Toc43168394)

[5. Conclusões 17](#_Toc43168395)

[6. Links 19](#_Toc43168396)

[REFERÊNCIAS 20](#_Toc43168397)

# 1. Introdução

## Contexto

Para gerar maior vantagem competitiva em um mercado extremamente disputado, antecipar problemas e identificar necessidades do negócio e de clientes, empresas precisam cada vez mais se munir de conhecimento analítico que suporte a tomada de decisão em um ambiente corporativo e gerencial.

Essa necessidade, por mais que seja urgente de ser resolvida, não significa que seja fácil e que corporações já entendam como solucionar. Ao contrário, o que é mais visto dentro da estrutura empresarial são gestores que não sabem organizar dados para trabalharem em favor da resolução de tarefas e, muitas vezes, não conseguem medir a saúde do negócio por estarem rodeados de uma quantidade infinita de dados que separados e sem contexto, não são capazes de fazer absolutamente nada para auxiliar na tomada de decisão.

Por exemplo, um executivo, entre muitas outras coisas, precisa decidir quem contratar, demitir ou treinar, além de quando baixar preços ou lançar novos produtos. Se uma dessas decisões falhar, toda a empresa é comprometida. Portanto, não é prudente acreditar no achismo: é preciso se basear em informações para uma boa tomada de decisão. (MOVIDESK BLOG, 2020)

Neste trabalho será apresentado uma estrutura de visualização de informações gerenciais como exemplo de modelo útil para auxiliar a camada gerencial da empresa. A Northwind Traders é uma empresa fictícia que vende alimentos de especialidade no atacado para varejistas em todo o mundo e simula o gerenciamento de diversos aspectos de uma companhia como pedidos, clientes, fornecedores, entregas, dentro outras informações. A base de dados desta empresa fictícia surgiu inicialmente como recurso de modelo de uso do Microsoft Access e ainda está disponível para uso nas versões atuais do sistema.

Neste *dataset* é possível organizar diversas estruturas analíticas para auxilio da tomada de decisão de líderes de operações, passando por gerências intermediárias e chegando em áreas de alto nível como diretorias e conselhos administrativos.

## 1.2. Público alvo

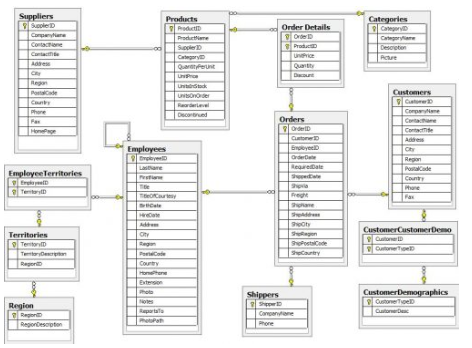
O público alvo da solução que será apresentada são pessoas que trabalham como líderes de operações logísticas e precisam visualizar rapidamente o andamento dos processos que suportam as vendas da empresa, gerentes de áreas em uma segunda camada que controlam etapas com nível de complexidade mais abrangente e também diretores e membros de conselhos administrativos que acompanham as métricas da companhia com menor nível de detalhe, mas com necessidade de maior assertividade dos números. A ideia é que independentemente do nível gerencial, a solução seja fácil e intuitiva de ser usada principalmente no que tange as visualizações dos dashboards, não sendo necessário um conhecimento técnico aprofundado e sim que saiba apenas acessar as ferramentas de visualização de dados. Também não é preciso ter experiência em análise de dados volumosos ou complexos porque as informações serão apresentadas de uma forma que gere entendimento rápido do que está acontecendo na empresa.

# 2. Modelo de Dados

## 2.1. Modelo Dimensional

A base de dados utilizada neste trabalho possui como modelo dimensional o *Star Schema* onde existe uma tabela fato ligada as várias dimensões do *dataset*.

**Figura 1 -** Diagrama dimensional



Fonte: Artigo do Medium[[1]](#footnote-1)

## 2.2. Fatos e Dimensões

Foram usadas as seguintes dimensões e fato para realização desse trabalho

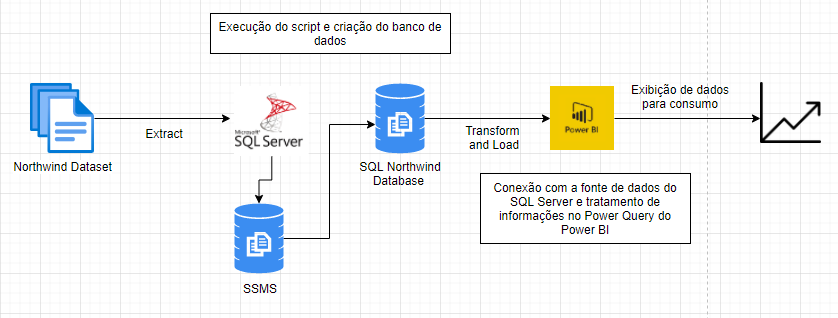
|  |  |  |
| --- | --- | --- |
| Tabela | Tipo de tabela | Descrição |
| Categories | Dimensão | Apresenta nome da categoria de produtos, id de categoria e descrição da categoria |
| Customers | Dimensão | Apresenta informações completas dos clientes como nome da empresa, nome para contato, titulo do contato, endereço, cidade, região, código postal, telefone e fax |
| Customer Demographic | Dimensão | Apresenta informações demográficas dos clientes como a descrição do cliente e também o tipo de cliente |
| CustomerCustomerDemo | Dimensão | Apresenta o id de cliente e o id do tipo de cliente |
| Employees | Dimensão | Apresenta informações dos empregados da Northwind Traders como ID, Nome e data de nascimento, data de contratação, informações de contato e endereço e para quem o empregado reporta dentro da organização |
| EmployeesTerritories | Dimensão | Possui o ID do empregado e o ID do território que ele trabalha |
| OrderDetails | Dimensão | Apresenta as informações detalhadas dos pedidos realizados na Northwind: ID do pedido, ID do produto, valor pago e quantidade de itens comprados por ID de produto. |
| Order | Fato | A tabela fato apresenta os seguintes registros: Order ID como PK, ID do cliente, ID do empregado, Data do pedido, Endereço de entrega, valor do frete. |
| Products | Dimensão | Tabela que apresenta todas as informações dos produtos vendidos pela Northwind Traders incluindo o estoque como ID de produto, nome do produto, preço por unidade, categorias de produtos, fornecedor, itens em estoque, itens descontinuados, itens solicitados. |
| Region | Dimensão | Informa as regiões que estão presentes os empregados/vendedores da empresa com ID e nome da região |
| Shippers | Dimensão | Essa tabela informa as transportadoras que são usadas nas entregas de pedidos com nome e id. |
| Suppliers | Dimensão | Apresenta os fornecedores de produtos da Northwind com nome, informações de contato, endereço de localização. |

# 3. Processo ETL

Os dados utilizados no trabalho fazem parte do conhecido *dataset* daNorthwind disponibilizado pela Microsoft. O *script* foi extraído do repositório do *Github* para ser carregado no Microsoft SQL Server. O script foi executado através do SQL Server Management Studio onde previamente tinha sido criado um servidor para utilização e onde o banco de dados com todas as tabelas acima foi criado.

O processo de carregamento dos dados foi realizado através do Power BI onde foi conectado ao servidor do SQL Server para exibir as fontes de dados relacionadas. O tratamento dos dados foi realizado pelo Power Query que serviu para criar campos calculados e colunas customizadas para otimizar a criação dos dashboards e análises.

**Figura 2** – Fluxo de ETL



Fonte: Criado pelo autor na ferramenta online Draw.iO

# 4. Camada de Apresentação

**4.1 - Painel estratégico**

O painel estratégico foi pensando para reunir *KPIs* mais relevantes em relação ao contexto da companhia e que pudesse suportar de maneira ampla a necessidade de metrificação da saúde do negócio por parte da área de diretoria da empresa e também membros do conselho administrativo ou investidores.

No painel é apresentado 3 módulos de visualização:

Módulo 1 – Faturamento

Reúne informações de faturamento da Northwind Traders

* Valor em vendas: valor somado de todas vendas da companhia
* Ticket médio por venda: campo calculado para mostrar o valor que cada pedido gera em média. Cálculo: faturamento / quantidade de vendas.
* Valor em vendas por ano: mostrar um panorama das vendas em reais de cada ano desde o início da operação
* Faturamento por ano: KPIque apresenta o faturamento do ano vigente, 1998, desde janeiro até o período atual (*year to date*). Junto deste KPI é mostrado a meta de faturamento com o valor em reais e a porcentagem que resta para atingir o objetivo.
* Projeção de faturamento nos próximos 5 anos baseado em dados histórico das vendas.

Módulo 2 – Pedidos

Reúne KPIs relacionados as vendas da Northwind Traders

* Categorias mais vendidas: resumo das 3 principais categorias de alimentos vendidos
* Países com mais pedidos: resumo dos 3 principais países que mais realizam compras com a Northwind
* Valor em frete: valor somado do gasto que a companhia tem com fretes/entregas
* Número de pedidos (com meta): KPI que apresenta a quantidade total de pedidos realizados desde o início da operação junto com uma meta de 3000 mil vendas somando todos os anos

Módulo 3 – Clientes

Apresenta dados de clientes da Northwind Traders

* Ticket médio por cliente: quanto em média cada cliente compra em valor de produto. Cálculo: faturamento total / por quantidade de clientes.
* Principais clientes em valor de vendas: principais clientes que realizam compras ordenados pela soma do valor comprado
* Quantidade de clientes: quantidade de clientes que a companhia possui com uma meta de obter 300 clientes ao todo

**4.2 - Painel tático**

A camada tática do painel apresenta informações que contribuem para a tomada de decisão e para que a empresa atinja os objetivos de médio prazo. Esse painel também serve para avaliar a eficácia e erros da estratégia da Northwind.

O painel apresenta dois gráficos de pizza mostrando a porcentagem de entregas realizadas por transportadora além de informações do frete gasto por transportadora. Essa informação ajuda a entender como está o desempenho da parte de logística dos pedidos e seu impacto no faturamento. Também é exibido a média de *Leadtime* das vendas atuai, ou seja, quanto tempo em média está levando para um pedido ser entregue ao cliente. Em seguida, é exibido um gráfico em cascata com os números de faturamento mês a mês e uma lista de todos os clientes ordenada pelos clientes que mais compram para os que menos compram. Um gráfico em barras com o faturamento por país e uma *treemap* com a visualização das categorias vendidas aparece na parte inferior do painel. Por último, é apresentado um gráfico de barras com a quantidade de pedidos mensais.

**4.3 - Painel operacional**

O painel operacional reúne dados mais detalhados da área de vendas e de logística para suportar o dia a dia e tomada de decisão dos gerentes e analistas operacionais.

É exibido na parte superior os filtros que servem para segmentar as camadas dos pedidos do painel como número da ordem, mês de vendas, ano de vendas, clientes, transportadoras e vendedor.

Na camada de visualização é exibida a quantidade de itens vendidos, itens que existem no estoque e itens que foram solicitados para reabastecer o estoque junto aos fornecedores. Essa visão serve para deixar informado o time operacional sobre as necessidades de reabastecimento de produtos confrontado com as vendas e, assim, não deixar que falte produtos que tem alta saída nos pedidos.

Também é exibida uma visualização de pedidos entregue por cidade, quantidade de vendas realizadas por cada vendedor e quantidade de pedidos realizados por clientes da empresa.

**4.4 – Aplicação de técnicas de *Machine Learning***

Para realizar as aplicações de técnicas de *machine learning* neste trabalho, foi utilizada a interface do Orange Data Mining. Esse programa consegue exibir diversos recursos para análise de aprendizagem sem especificamente ter experiência com códigos ou linguagens de programação. As técnicas escolhidas foram regras de associação e clusterização.

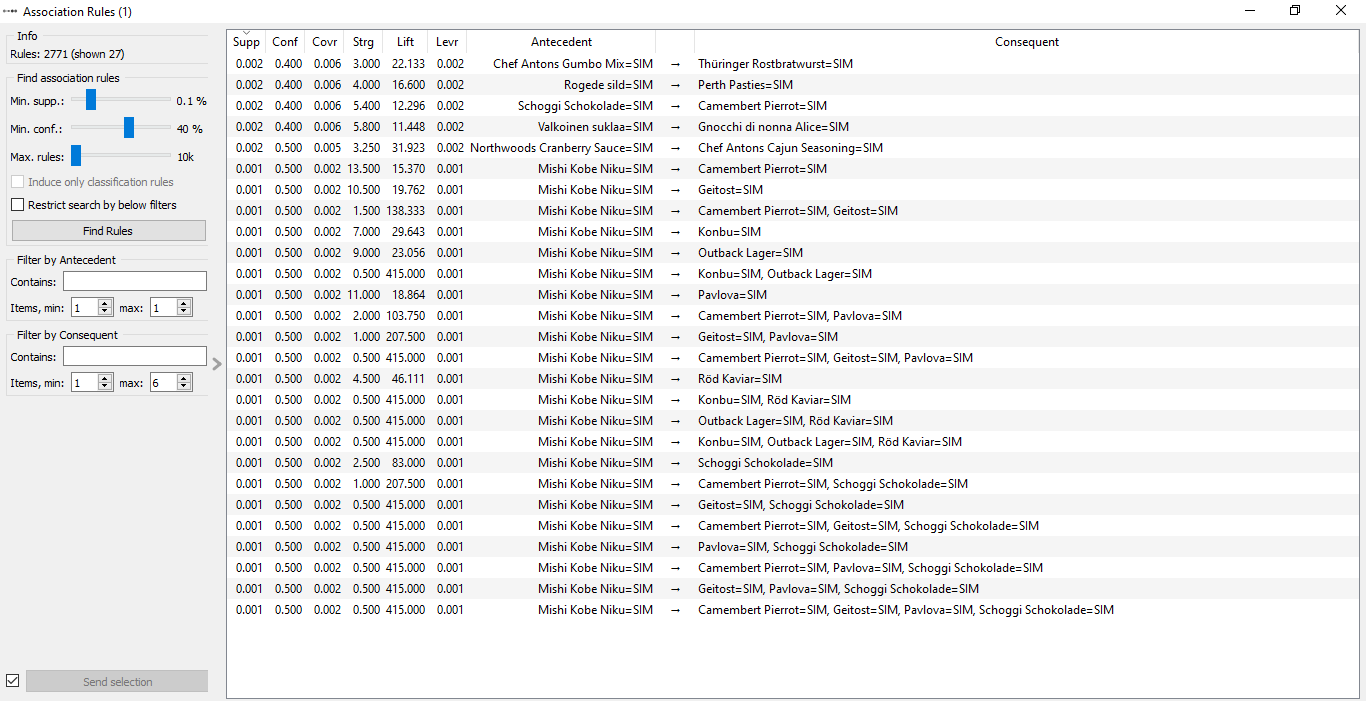
Foi importado um arquivo em formato CSV para dentro do Orange Data Mining que continha a relação de todos os pedidos realizados e os produtos vendidos com a sinalização de qual produto tinha sido vendido para cada identificador de ordem.

Na análise com regras de associação, era buscado padrões de consumo dos clientes da Northwind. A ideia era entender se haviam produtos que usualmente eram comprados em conjunto com outros produtos (*Market basket analysis*) e, assim, utilizar essas informações para aprimorar a recomendação de produtos no momento das vendas. Para essa mineração foi utilizado o algoritmo *Apriori*.

O nível de suporte era bastante baixo, só sendo possível encontrar uma maneira de extrair informações relevantes com ele em 0,1%. Isso aconteceu porque a maioria dos pedidos, mesmo com mais de um item, eram extremamente variados no tange os produtos vendidos.

O nível de confiança também precisou ficar abaixo dos 50% para apresentar informações que pudessem ser analisadas. Apesar da baixa nos índices, foi possível extrair insights relevantes.

**Figura 3** – Regras de associação



Fonte: Orange Data Mining

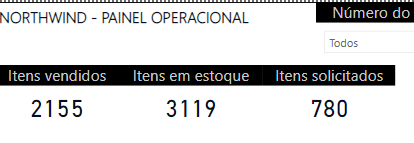
O mais relevante desses insights foi a presença do item *Mishi Kobe Niku*, um tipo de carne, na maioria das regras encontradas pelo algoritimo. Esse produto somente aparece em dois pedidos do *dataset* mas sempre acompanhado de diversos outros produtos.

Agrupando esses itens em suas categorias entendesse que sempre que o item carne *Mishi Kobe Niku* é comprado, também é consumido produtos das categorias de Lacticínios, Frutos do Mar, Bebidas e Confeitos. Assim, é possível recomendar os produtos que apareceram nas regras ou produtos de mesma categoria.

# 5. Registros de Homologação

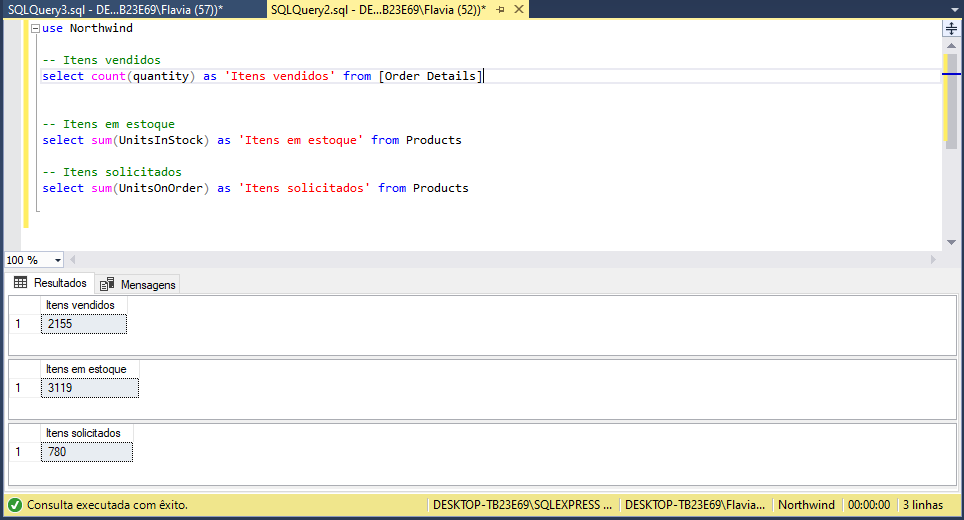
Para homologar o trabalho foram realizadas consultas na base de dados da Northwind presente no SQL Server e essas consultas foram comparadas com algumas visualizações do Power BI. As primeiras consultas se referem as quantidades de produtos vendidos, em estoque e também produtos que foram solicitados ao fornecedor para reabastecer o estoque:

**Figura 4** – Visualizações de homologação 1



Fonte: Power BI

**Figura 5** – Query no SQL MS 1



Fonte: *SQL Management Studio*

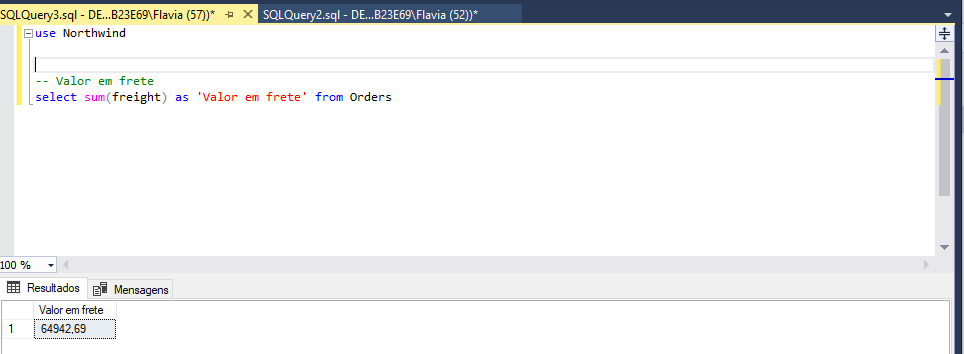
Outro exemplo é o valor de frete que na camada de apresentação aparece o arredondado em R$64.940,00 e no SQL Server é apresentado o mesmo valor, porém sem o arredondamento:

**Figura 6** – Visualização de homologação 2



Fonte: Power BI

**Figura 7** – Query no SQL MS 2



Fonte: *SQL Management Studio*

# 5. Conclusões

Realizando análises dos dados levantado no *dataset* da Northwind foi possível extrair alguns insights importantes que podem nortear a estratégia da empresa. O primeiro deles é sobre o entendimento do faturamento. Apesar da quantidade pequena de pedidos em três anos de operação (2155 pedidos), é possível entender que o que faz diferença é o valor de cada venda, atingindo um ticket médio de quase 15 mil reais por pedido. A chave do faturamento pode estar em aumentar a carteira de clientes ou incentivo de vendas para clientes que fazem compras de valores mais altos. Outro ponto interessante em relação ao faturamento é sobre a previsão de vendas durante os próximos anos. A previsão apresenta um aumento de faturamento de quase R$500.000,00 por ano consolidando o potencial lucrativo da companhia.

Entrando no nível de pedido é possível observar um aumento vendas no entre dezembro a abril. Isso pode estar associado as festividades que acontecem com maior frequência nestes períodos do ano. Com essa observação, é possível aumentar a quantidade de vendedores no período, como vendedores temporários, para ajudar impulsionar a saída de vendas e aumentar o faturamento que irá balancear a queda das vendas que acontece entre maio e novembro de todos os anos.

Outro ponto de partida para evolução das vendas pode estar associado ao tempo de entrega dos pedidos. A média atual está em quase 9 dias para realizar a entrega de um pedido e levando em consideração o tipo de cliente que realiza essas compras, (varejistas que consomem produtos de atacadistas), esse tempo alto de entrega pode gerar complexidade no fluxo de vendas desses clientes, já que eles precisam constantemente reabastecer o estoque. A ideia central aqui é iniciar um estudo que leve a redução do prazo entrega como uma vantagem competitiva da Northwind perante o mercado, entregando produtos com o menor tempo de entrega ou criando um tipo de entrega *express* para alguns clientes em determinados países como um piloto para testar a estratégia. Aumentar a quantidade de centros de distribuição pode ajudar a otimizar o prazo das entregas também.

Na análise de *Machine Learning* foi possível observar que um único produto (Mishi Kobe Niku) é frequentemente comprado em associação com diversos outros produtos sendo um item importante para ser usada como chave de recomendação nas vendas. Outro fato importante sobre o item Mishi Kobe Niku é que ele possui um valor unitário alto, cada peça sai por R$97,00. Apesar de sair em baixa quantidade, o seu valor alto pode servir de uso para incentivar a venda de produtos que estão dentro de uma calda longa no catálogo da Northwind. A venda de poucos itens de um produto com valor unitário alto já consegue gerar um aumento significativo no faturamento.

Existem diversas oportunidades que podem ser mapeadas a partir dos dados da Northwind. Estudos mais aprofundados de recomendação de produtos e também projetações de vendas segmentadas por países. Estudos de comportamento dos varejistas também é uma maneira interessante de expandir a camada informativa dos consumidores da empresa além de possivelmente alavancar o aumento de pedidos.

Também é possível avançar as análises no que tange os funcionários da empresa, entendendo como cada vendedor pode ser melhor alocado dentro das regiões de alcance da Northwind além de avaliar se é possível um programa de premiação vantajoso para os vendedores que tiveram melhor performance de vendas.

.

# 6. Links

* Link da apresentação do trabalho: <https://www.youtube.com/watch?v=B10yJBiTobw>
* Repositório com *script* para criação do banco de dados no SQL Server: <https://github.com/microsoft/sql-server-samples/tree/master/samples/databases/northwind-pubs>
* *Script* utilizado para criação do bancos de dados no SQL Server: <https://raw.githubusercontent.com/microsoft/sql-server-samples/master/samples/databases/northwind-pubs/instnwnd.sql>
* Regras de associação (Orange Notebook Flow): <https://raw.githubusercontent.com/sflaviaelias/TCC-Northwind/main/Regras%20de%20associa%C3%A7%C3%A3o.ows>
* Link do dashboard no Power BI para ser acesso na web (disponível para pessoas com acesso ao domínio PUC Minas) <https://app.powerbi.com/links/tSTM92OM7P?ctid=14cbd5a7-ec94-46ba-b314-cc0fc972a161&pbi_source=linkShare>
* *Dashboard* do Power BI para *download* no Google Drive: <https://drive.google.com/file/d/1DmfzPE0vpvFolg3QHowbZ-HlOcIGE6c1/view?usp=sharing>
* Link do Github com o projeto completo: <https://github.com/sflaviaelias/TCC-Northwind>

# REFERÊNCIAS

Fuchs, Andriano. **Por que a coleta de dados é tão importante para a tomada de decisão?** Movidesk BLOG, 2020. Disponível em: <https://conteudo.movidesk.com/tomada-de-decisao/>. Acesso em: 10 de jun 2021.

1. Disponível em: https://medium.com/@kmsbmadhan/dimensional-modelling-visualization-of-northwind-database-beaac7fecb20 [↑](#footnote-ref-1)