Processamento e Visualização de Dados (P01)

SISTEMAS DE APOIO À DECISÃO, 2021-22

**Hélder Carvalho (15310), José Cosgrove (18826), Miguel Silva (18829)**

# Introdução

“O recurso mais valioso do mundo já não é o petróleo, mas sim os dados” (The Economist, 2017).

Esta frase, retrata perfeitamente o mundo em que vivemos, e, portanto, todas as empresas sentem a necessidade de aproveitar ao máximo os dados que possuem, analisando e compreendendo o seu significado e como se relacionam.

Com o aumento da utilização de técnicas de agregação de dados em Armazéns de Dados/*Data Marts*, as empresas procuram não só padrões e tendências, mas também analisar financeiramente um específico intervalo de tempo.

A exibição dos dados segue uma série de medidas estatísticas que, em combinação com as dimensões e tabelas de factos, visam responder a um conjunto de questões colocadas pelo utilizador.

Adicionalmente, podem ser utilizados *Key Performance Indicators* (*KPI’s*) que, aos gestores, fornecem uma análise baseada em objetivos.

Este documento visa relatar o processo de processamento e visualização de dados do Armazém de Dados/*Data Mart* “*Adventure Works Cycles*”, desenvolvido no trabalho prático 1. Esta empresa fictícia é uma fabricante e vendedora de bicicletas e respetivos acessórios, com dois canais de vendas, online para milhares de clientes e retalho para algumas lojas. Existe também informação sobre todas as compras efetuadas a fornecedores, com objetivo de reposição de stocks. O valor médio de cada venda a retalho é muito mais elevado do que o valor médio de cada venda efetuada *online*. Cada compra e venda possui informação sobre o fornecedor (no caso de compra), cliente/loja (no caso de venda), funcionário que comprou ou vendeu, total base, total de taxas/impostos, custo de portes de envio, custo total, entre outras. Cada produto, de entre informações básicas como nome, número, etc., tem também informações sobre se é fabricado ou comprado, vendível ou não, custo, preço de venda, etc. Os fornecedores são classificados não só como preferenciais e ativos, mas também por *rating* de crédito.

# Aquisição e Preparação dos Dados

Os dados foram adquiridos a partir de duas fontes, sendo uma delas o (1) *Data Mart* “*Adventure Works Cycles*” desenvolvido no trabalho prático 1, e outra um (2) ficheiro *CSV* com informações referentes à “População urbana (% da população total)” (United Nations Population Division, 2018).

Relativamente à preparação dos dados, na fonte 1 foram apenas criadas algumas hierarquias e definidas algumas categorias de dados, enquanto que na fonte 2 foi necessária uma transformação mais complexa, com recurso ao *Power Query*.

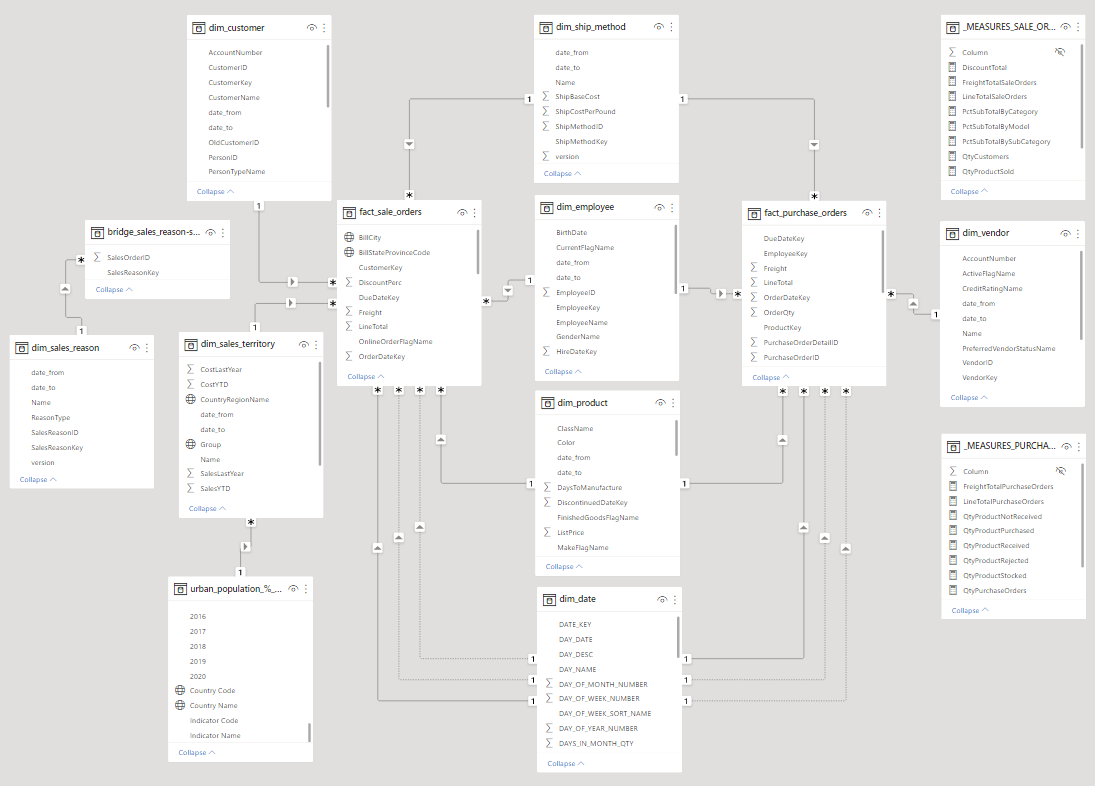


Figura 1 – Modelo de Dados do Data Mart “Adventure Works Cycles”

## Transformação dos dados da fonte “População urbana (% da população total)”

Esta fonte possui dados sobre a percentagem de população que vive em zonas urbanas para cada país entre 1960 e 2020.

Por isso, é constituída por diversas colunas, uma referente ao país, outra com a abreviatura do país e uma para cada ano.

Porém, para que o *Power BI* apresentasse as percentagens corretamente, cada coluna representativa de um ano tinha de ser convertida para percentagem, porém existia um problema, os valores não estavam num formato que o *Power BI* conseguisse converter automaticamente, mais concretamente, o *Power BI* necessitava não só de valores em que o separador decimal fosse um ponto (e não uma virgula), mas também que esses valores estivessem compreendidos entre 0 e 1 (ex.: 0 = 0%, 0.5 = 50%, 1 = 100%) isto porque o *Power BI*, quando transforma uma coluna em percentagem, multiplica os seus valores por 100.

Para isso, recorrendo ao *Power Query*, foram substituídas todas as virgulas por pontos (linha 1) e de seguida convertidas as colunas para o tipo de dados “Número” (linha 2). Após isto, com recurso à função “Transformar Colunas”, foi dividido cada valor por 100 (para compreender os valores entre 0 e 1) (linha 3). Por fim, foram convertidas novamente as colunas, mas desta vez para o tipo de dados “Percentagem” (linha 4).

1. #"Replaced Value" = Table.ReplaceValue(#"Removed Errors", ".", ",", Replacer.ReplaceText,{“coluna”, “coluna”, …}),
2. #"Changed Type2" = Table.TransformColumnTypes(#"Replaced Value",{{"coluna", type number}, {"coluna", type number}, …}),
3. #"Divided Column" = Table.TransformColumns(#"Changed Type2", List.Transform({"coluna", "coluna", …}, each {\_, (inner) => inner / 100, type number})),
4. #"Changed Type3" = Table.TransformColumnTypes(#"Divided Column",{{"coluna", Percentage.Type}, {"coluna", Percentage.Type}, …}})

# Modelação e Processamento dos Dados

## Novas Relações Entre Tabelas

Apenas foi criada uma nova relação entre tabelas, mais concretamente, entre a tabela da fonte de dados externa e a tabela da dimensão territórios de venda (*dim\_sales\_territory*). Esta relação foi criada para que seja possível associar a percentagem de população urbana a cada território de venda.

## Hierarquias

O objetivo das hierarquias é associar uns elementos a outros. Isto permite que, nas visualizações, se possa utilizar a funcionalidade *drill-down*.

Dimensão Data

A hierarquia da Dimensão Data tem como objetivo agrupar:

* O dia no mês;
* O mês no trimestre;
* O trimestre no ano.

Dimensão Cliente

A hierarquia da Dimensão Cliente tem como objetivo filtrar os clientes mediante os seguintes agrupamentos:

* Zonas em países;
* Países em continentes.

Dimensão Produto

A hierarquia da Dimensão Produto tem como objetivo filtrar os produtos mediante os seguintes agrupamentos:

* Medelo em subcategoria;
* Subcategoria em categoria.

## Medidas Calculadas

O *Power Bi* tem a capacidade de detetar automaticamente as colunas que podem ser utilizadas como medidas, mas, muitas vezes, é necessário criar medidas personalizadas para realizar cálculos mais específicos e complexos. Por esse motivo existe a linguagem *Data Analysis Expressions* (*DAX*). Esta linguagem é similar à utilizada no *Microsoft Excel*, porém é talhada especificamente para trabalhar com bases de dados relacionais.

Nas tabelas abaixo são apresentadas as medidas calculadas e suas expressões *DAX* relacionadas com cada tabela de factos.

Tabela 1 – Medidas Calculadas das Vendas

|  |  |  |
| --- | --- | --- |
| Nome da Medida | Descrição | Expressão *DAX* |
| QtySaleOrders | Quantidade total de vendas | DISTINCTCOUNT ( fact\_sale\_orders[SalesOrderID] ) |
| QtyProductSold | Quantidade de produtos vendidos | SUM ( fact\_sale\_orders[OrderQty] ) |
| QtyCustomers | Quantidade total de clientes | DISTINCTCOUNT ( fact\_sale\_orders[CustomerKey] ) |
| LineTotalSaleOrders | Total líquido de vendas em $ | SUM( fact\_sale\_orders[LineTotal] ) |
| TaxAmtTotalSaleOrders  (Ng, 2017) | Total de taxas de vendas em $ | SUMX (  DISTINCT ( fact\_sale\_orders[SalesOrderID] ),  FIRSTNONBLANK ( fact\_sale\_orders[TaxAmt], 0 )  ) |
| FreightTotalSaleOrders  (Ng, 2017) | Total de transporte de vendas em $ | SUMX (  DISTINCT ( fact\_sale\_orders[SalesOrderID] ),  FIRSTNONBLANK ( fact\_sale\_orders[Freight], 0 )  ) |
| TotalDueTotalSaleOrders  (Ng, 2017) | Total bruto de vendas em $ | SUMX (  DISTINCT ( fact\_sale\_orders[SalesOrderID] ),  FIRSTNONBLANK ( fact\_sale\_orders[TotalDue], 0 )  ) |
| DiscountTotalSaleOrders | Total de desconto de vendas em $ | SUMX (  fact\_sale\_orders,  fact\_sale\_orders[UnitPrice] \* fact\_sale\_orders[UnitPriceDiscount] \* fact\_sale\_orders[OrderQty]  ) |
| PctSubTotalByCategory  (Microsoft Corporation, sem data) | Percentagem do total líquido de vendas por categoria de produto | DIVIDE (  [LineTotalSaleOrders],  CALCULATE (  [LineTotalSaleOrders],  ALLSELECTED (  dim\_product[ProductCategoryName]  )  )  ) |
| PctSubTotalBySubCategory  (Microsoft Corporation, sem data) | Percentagem do total líquido de vendas por subcategoria de produto | DIVIDE (  [LineTotalSaleOrders],  CALCULATE (  [LineTotalSaleOrders],  ALLSELECTED (  dim\_product[ProductSubcategoryName]  )  )  ) |
| PctSubTotalByModel  (Microsoft Corporation, sem data) | Percentagem do total líquido de vendas por modelo de produto | DIVIDE (  [LineTotalSaleOrders],  CALCULATE (  [LineTotalSaleOrders],  ALLSELECTED (  dim\_product[ProductModelName]  )  )  ) |

Tabela 2 – Medidas Calculadas das Compras

|  |  |  |
| --- | --- | --- |
| Nome da Medida | Descrição | Expressão *DAX* |
| QtyPurchaseOrders | Quantidade total de compras | DISTINCTCOUNT (  fact\_purchase\_orders[PurchaseOrderID]  ) |
| QtyProductPurchased | Quantidade total de produtos comprados | SUM ( fact\_purchase\_orders[OrderQty] ) |
| QtyProductReceived | Quantidade total de produtos recebidos | SUM ( fact\_purchase\_orders[ReceivedQty] ) |
| QtyProductNotReceived | Quantidade total de produtos não recebidos | [QtyProductPurchased] - [QtyProductReceived] |
| QtyProductStocked | Quantidade total de produtos colocados em stock | SUM ( fact\_purchase\_orders[StockedQty] ) |
| QtyProductRejected | Quantidade total de produtos rejeitados | SUM ( fact\_purchase\_orders[RejectedQty] ) |
| QtyVendors | Quantidade total de fornecedores | DISTINCTCOUNT (  fact\_purchase\_orders[VendorKey]  ) |
| LineTotalPurchaseOrders | Total líquido de compras em $ | SUM( fact\_purchase\_orders[LineTotal] ) |
| TaxAmtTotalPurchaseOrders  (Ng, 2017) | Total de taxas de compras em $ | SUMX (  DISTINCT (  fact\_purchase\_orders[PurchaseOrderID]  ),  FIRSTNONBLANK (  fact\_purchase\_orders[TaxAmt], 0  )  ) |
| FreightTotalPurchaseOrders  (Ng, 2017) | Total de transporte de compras em $ | SUMX (  DISTINCT (  fact\_purchase\_orders[PurchaseOrderID]  ),  FIRSTNONBLANK (  fact\_purchase\_orders[Freight], 0  )  ) |
| TotalDueTotalPurchaseOrders  (Ng, 2017) | Total bruto de compras em $ | SUMX (  DISTINCT (  fact\_purchase\_orders[PurchaseOrderID]  ),  FIRSTNONBLANK (  fact\_purchase\_orders[TotalDue], 0  )  ) |

# Visualização dos Dados

## Perfil de Utilizador

O principal utilizador final será a equipa de gestão de topo ou a equipa de marketing, que terão acesso a informações gerais e especificas acerca das vendas por tipo, país, trimestre, produto, categoria de produto, subcategoria de produto, modelo de produto, etc. Adicionalmente, terá também informações sobre as compras por trimestre, fornecedor, produto, produtos recebidos e não recebidos, produtos colocados em stock e rejeitados, etc.

Assim, é possível prever padrões e tendências, como também analisar financeiramente específicos intervalos de tempo, com o objetivo final de tomar as decisões mais acertadas para o futuro da empresa.

## *Dashboards*

Os objetivos dos *dashboards* são aumentar a compreensão sobre os clientes, analisar as vendas, compras, fornecedores, material vendido/comprado, etc.

Vendas

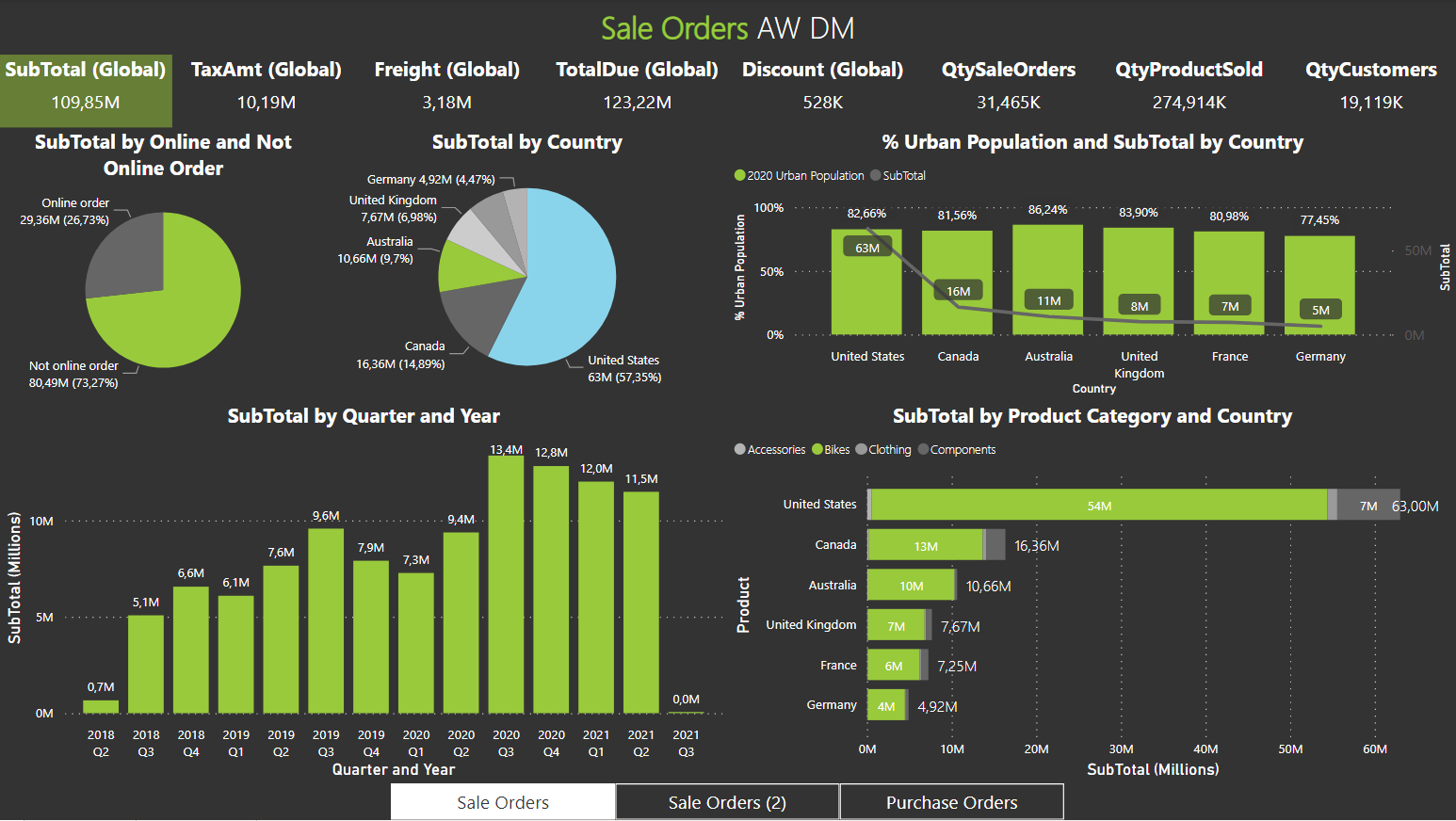


Figura 2 – Dashboard Vendas

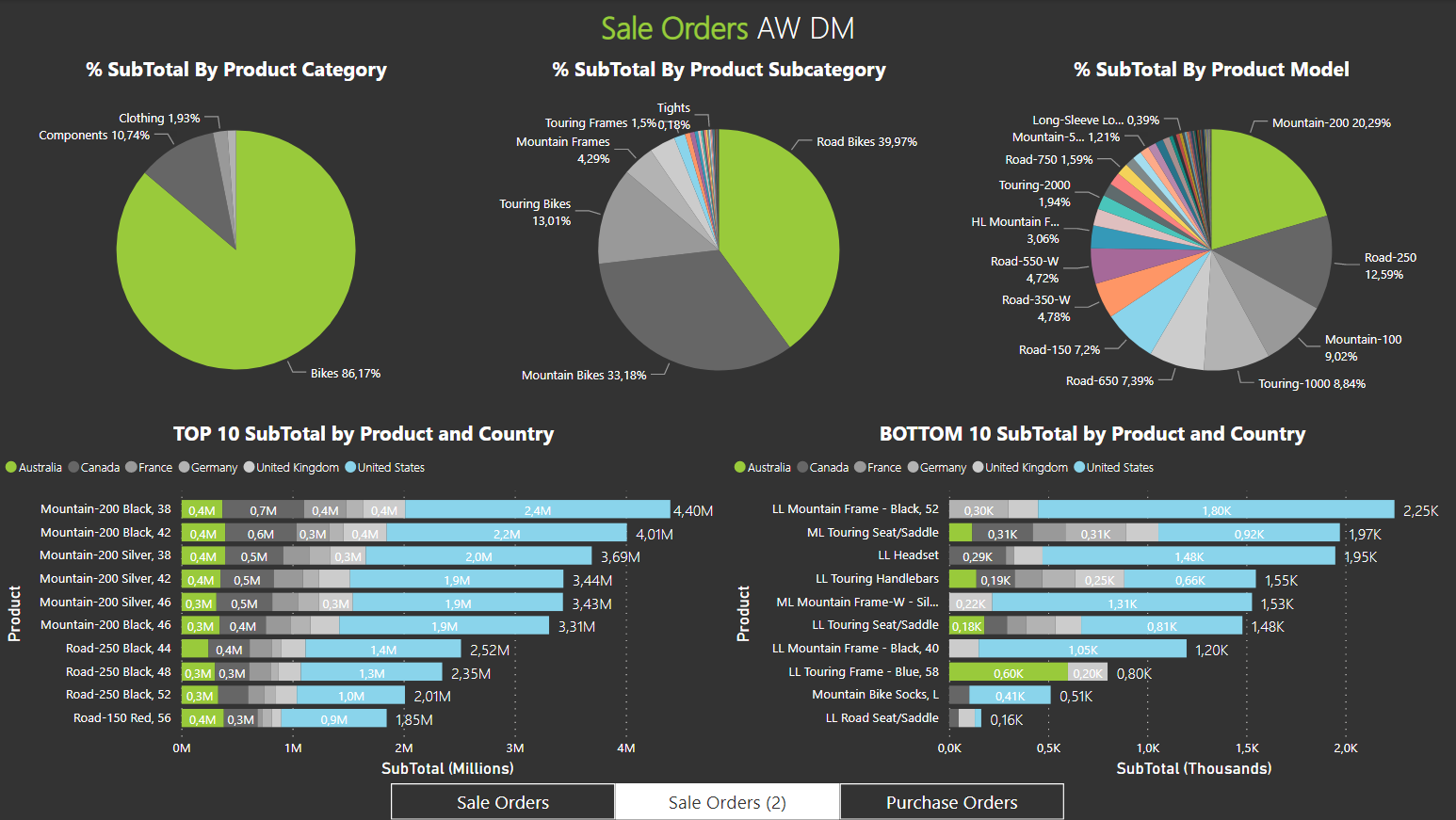


Figura 3 – Dashboard Vendas (2)

Para a venda de produtos foram criados dois *dashboards*, onde é possível, através de filtros, restringir a informação por tipo de venda (*online* ou não), país, trimestre, produto, categoria de produto, subcategoria de produto, modelo de produto e, por fim, produtos mais e menos vendidos.

* **Valores globais e quantidades**: pode-se ver o total líquido, total de taxas/impostos, total de transporte, total bruto, total de descontos, quantidade de vendas, quantidade de produtos vendidos, quantidade de clientes;
* **Subtotal por tipo de venda (*online* ou não)**: cerca de 73% do valor total das vendas não é proveniente de vendas *online* (são vendas de retalho efetuadas a lojas) e os restantes 26% são provenientes de vendas *online* (efetuadas a clientes individuais);
* **Subtotal por país**: cerca de 57% do valor total das vendas é proveniente dos EUA, 15% do Canada, 10% da Austrália, 7% do Reino Unido, 7% da França e 4% da Alemanha;
* **Subtotal e % de população urbana por país**: neste gráfico pode-se ver a percentagem de população urbana de cada país juntamente com o seu total de vendas e, facilmente se constata que todos os países pra os quais se vendem produtos têm uma percentagem de população urbana elevada (acima de 77%), logo é provável que existam mais utilizadores da bicicleta como transporte;
* **Subtotal por trimestre e ano**: este gráfico mostra o valor total de vendas por trimestre de todos os anos. Através da hierarquia Data é possível fazer *drill-up* e *drill-down* para visualizar o valor das vendas agregado também por ano e por trimestre de um específico ano;
* **Subtotal por categoria de produto e país**: neste gráfico pode-se ver o valor de vendas de cada categoria de produto em cada país e verificar que a categoria que gera mais receita em todos os países é a “Bicicletas”;
* **% do subtotal por categoria de produto**: neste gráfico pode-se constatar novamente que a categoria que gera mais receita é a “Bicicletas” com 86%, seguida da “Componentes” com 11%, depois a “Roupas” com 2% e por fim a “Acessórios” com 1%. Ao interagir com este gráfico, clicando nas categorias, os dois gráficos seguintes serão atualizados e mostradas apenas as subcategorias e modelos da categoria clicada;
* **% do subtotal por subcategoria de produto**: este gráfico mostra todas as subcategorias de produtos existentes e suas respetivas percentagens de receita, porém, quando selecionada uma categoria no gráfico anterior, este gráfico atualiza-se e mostra as percentagens de receita das subcategorias pertencentes à categoria selecionada;
* **% do subtotal por modelo de produto**: em semelhança ao gráfico anterior, este gráfico mostra todos os modelos de produtos existentes e suas respetivas percentagens de receita, porém, quando selecionada uma categoria no 2º gráfico anterior, este gráfico atualiza-se e mostra as percentagens de receita dos modelos pertencentes à categoria selecionada e, quando selecionada uma subcategoria no gráfico anterior, este gráfico atualiza-se e mostra as percentagens de receita dos modelos pertencentes à subcategoria selecionada;
* **TOP 10 subtotal por produto e país**: este gráfico mostra os 10 produtos que mais geram receita e a respetiva fatia pertencente a cada país (Tutorial Gateway, sem data). Todos os produtos geram mais receita nos EUA;
* **BOTTOM 10 subtotal por produto e país**: este gráfico mostra os 10 produtos que menos geram receita e a respetiva fatia pertencente a cada país (Tutorial Gateway, sem data). Quase todos os produtos geram mais receita nos EUA, porém existem alguns que nunca foram vendidos para pelo menos um país.

Compras

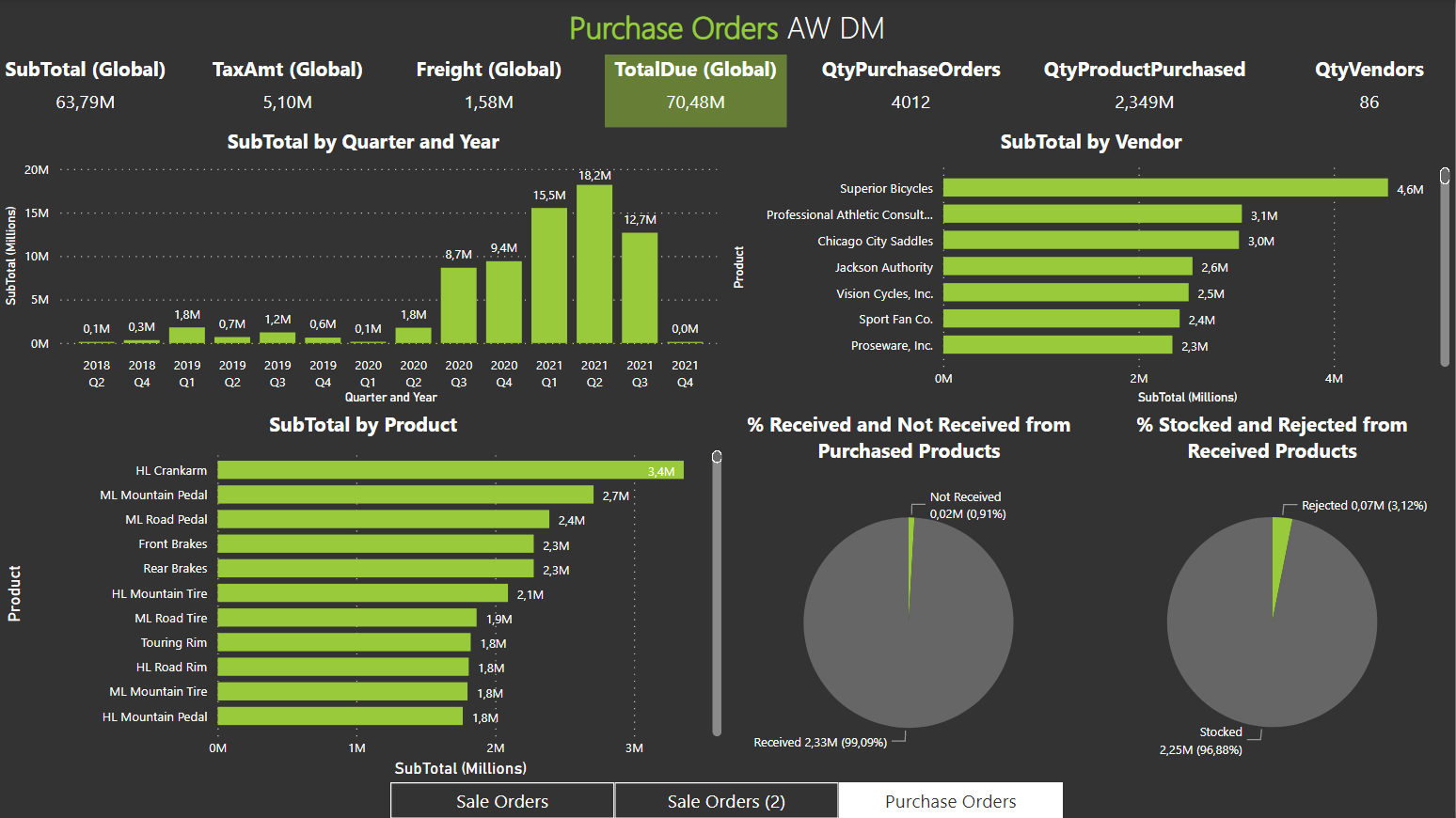


Figura 4 – Dashboard Compras

Para a compra de produtos foi criado um *dashboards*, onde é possível, através de filtros, restringir a informação por trimestre, fornecedor, produto, % de produtos recebidos e não recebidos e, por fim, % de produtos colocados em stock e rejeitados.

* **Valores globais e quantidades**: pode-se ver o total líquido, total de taxas/impostos, total de transporte, total bruto, quantidade de compras, quantidade de produtos comprados, quantidade de fornecedores;
* **Subtotal por trimestre e ano**: este gráfico mostra o valor total de compras por trimestre de todos os anos. Através da hierarquia Data é possível fazer *drill-up* e *drill-down* para visualizar o valor das compras agregado também por ano e por trimestre de um específico ano;
* **Subtotal por fornecedor**: este gráfico mostra o valor total das compras efetuadas a cada fornecedor;
* **Subtotal por produto**: este gráfico mostra o valor total das compras efetuadas por produto;
* **% de produtos recebidos e não recebidos do total de produtos comprados**: este gráfico mostra a percentagem de produtos recebidos (99%) e não recebidos (1%) tendo em conta o total de produtos comprados;
* **% de produtos colocados em stock e rejeitados do total de produtos recebidos**: este gráfico mostra a percentagem de produtos colocados em stock (97%) e rejeitados (3%) tendo em conta o total de produtos recebidos.

## 

# Conclusão

“O recurso mais valioso do mundo já não é o petróleo, mas sim os dados” (The Economist, 2017).

Neste mundo, o sucesso de uma empresa está diretamente relacionado com os dados que possui, a forma que estes são agregados, tratados e, acima de tudo, a qualidade dos mesmos.

As *dashboards* fornecem informação tanto generalizada como específica sobre o negócio. Essa informação pode e deve ser usada como vantagem estratégica da empresa.

Neste trabalho podemos afirmar que foram cumpridos todos os objetivos que nos foram impostos.

O processo de obtenção dos dados foi simples devido ao facto de estes terem sido preparados no trabalho prático 1. Apesar disso, a preparação e combinação dos dados com a fonte externa deu um pouco mais de trabalho devido à necessidade de converter e transformar os valores originais da fonte externa em valores que o *Power BI* conseguisse processar corretamente.

O desenvolvimento de todas as medidas calculadas e *dashboards* foi uma aventura devido ao facto de nunca termos utilizado linguagem *DAX* e *Power BI*.

Por fim, este projeto não só permite a visualização de dados do Armazém de Dados/*Data Mart* “*Adventure Works Cycles*”, mas também foi importante para o nosso futuro profissional porque nos permitiu aprender a desenvolver soluções analíticas que muita uteis são para as empresas.

# 

# Bibliografia

Microsoft Corporation. (sem data). *When to use Calculated Columns and Calculated Fields*. Obtido 28 de Dezembro de 2021, de https://support.microsoft.com/en-us/office/when-to-use-calculated-columns-and-calculated-fields-ca18d63a-5b6d-4000-8ca2-14d2aadbb734

Ng, F. (2017). *SUM Distinct values for first occurance in Power BI*. stackoverflow. https://stackoverflow.com/a/43549011/13356013

The Economist. (2017). The world’s most valuable resource is no longer oil, but data | The Economist. *Leaders*. https://www.economist.com/leaders/2017/05/06/the-worlds-most-valuable-resource-is-no-longer-oil-but-data

Tutorial Gateway. (sem data). *Power BI Top 10 Filters*. Obtido 27 de Dezembro de 2021, de https://www.tutorialgateway.org/power-bi-top-10-filters/

United Nations Population Division. (2018). *Urban population (% of total population)*. https://data.worldbank.org/indicator/SP.URB.TOTL.IN.ZS