



INSTITUTO
SUPERIOR
DE CONTABILIDADE
E ADMINISTRAÇÃO
DO PORTO

Business Intelligence

Análise da empresa Northwind

Pós-Graduação em Gestão de Sistemas de Informação Empresariais
Unidade Curricular: Noções Avançadas de Bases de Dados e *Business Intelligence*

Maio 2020

Bruna Tomé | 12190209

João Pereira | 12190211

Rui Ribeiro | 6940007

Índice

Introdução	4
1. Definição e características de um <i>Datawarehouse</i>	5
1.1. Tabelas de Factos	7
1.2. Tabelas de Dimensão.....	7
1.3. Esquema em estrela	8
2. Análise do Negócio e Modelação do <i>Datawarehouse</i>	10
2.1. Estrutura da Base de Dados <i>Northwind</i>	10
2.2. Modelação do <i>Datawarehouse</i>	12
3. Processo de Integração ETL – <i>Extract, Transform and Load</i>	14
3.1. Projeto de <i>Integrations Services</i> SSIS	14
3.1.1. Conexões às Bases de Dados (<i>Connection Manager</i>).....	15
3.1.2. Definição do Controlo de Fluxos (<i>Control Flow</i>).....	15
3.1.3. Definição do Fluxo de Dados (<i>Data Flow</i>)	16
3.1.4. <i>Deploy</i> do <i>package</i> em ambiente produtivo	18
4. Processo de Criação do Cubo OLAP e de Análise – <i>Analysis Services, EXCEL e PowerBI</i>	21
4.1. <i>Analysis Services</i>	21
4.2. Cubo OLAP	21
4.3. Projeto de <i>Analysis Services</i> SSAS.....	21
4.3.1. Conexão ao <i>Datawarehouse</i>	22
4.3.2. Definição do Cubo OLAP.....	23
4.3.3. <i>Deploy</i> do cubo em ambiente produtivo.....	27
4.4. Visualização dos dados em EXCEL	29
4.5. Visualização dos Dados no <i>Power BI</i>	34
4.5.1. Elementos de Visualização	35
4.5.2. <i>Dashboards Online</i>	40
Conclusão	43
Bibliografia.....	44

Índice de Figuras

Figura 1 – Datawarehouse Framework (GARDNER, 1998).....	5
Figura 2 – Esquema em Estrela	8
Figura 3 - Diagrama ER da Base de Dados Northwind	11
Figura 4 - Definição do Datawarehouse	13
Figura 5 - Ligações configuradas no Connection Manager	15
Figura 6 - Definição do Controlo de Fluxos do Projeto	15
Figura 7 - Definição do Data Flow do Projeto	16
Figura 8 - 2154 registos processados pelo processo ETL	17
Figura 9 - Instrução SQL para obter 2154 registos no datawarehouse	17
Figura 10 - Instrução sql para obter o volume de vendas e quantidades vendidas no datawarehouse	18
Figura 11 - Instrução sql na base de dados Access	18
Figura 12 - Resultados obtidos no Access	18
Figura 13 - Deploy do projeto no SS Integrations Service	19
Figura 14 - Parâmetros de execução do package	20
Figura 15 - Resultados da execução periódica do package (jobs)	20
Figura 16 - Ligação ao Datawarehouse	22
Figura 17 - Identificação da tabela de factos e de dimensão na criação do cubo.....	23
Figura 18 - Cubo OLAP	24
Figura 19 - Total de Vendas e Quantidades Vendidas	25
Figura 20 - Total de Vendas e Quantidades Vendidas Por Ano	25
Figura 21 - Total de Vendas e Quantidades Vendidas Por Categoria de Produto	25
Figura 22 - Total de Vendas e Quantidades Vendidas Por País	26
Figura 23 - Autenticação no SQL Server Analysis Services	27
Figura 24 - Cubo implementado em ambiente de Produção no SQL Server	28
Figura 25 - Conexão do EXCEL para o Cubo OLAP	29
Figura 26 - Volume de Negócios e Quantidades por País e por Categoria.....	30
Figura 27 - Gráfico de Colunas das Vendas por País	31
Figura 28 - Gráfico Circular de Vendas por País	31
Figura 29 - Previsão das Vendas até 31/12/1996	32
Figura 30 - Previsão das Vendas nos USA até 31-12-1996	33
Figura 31 - Volume de Negócios por País	35
Figura 32 - Volume de Negócios por Categoria	36
Figura 33 - Top 10 dos Produtos Mais Vendidos.....	36
Figura 34 - Top5 Volume de Negócios por País e Por Ano	37
Figura 35 - Evolução do Peso das Quantidades Vendidas do Top1 dos Produtos.....	38
Figura 36 - Exemplo de Atributo da Dimensão Local	39
Figura 37 - Exemplo de Gráfico com Métricas Diferentes	39
Figura 38 - Dashboard Vista Browser.....	41
Figura 39 - Dashboard Vista Mobile.....	42

Introdução

Este trabalho foi realizado no âmbito da unidade curricular de Noções Avançadas de Base de Dados e *Business Intelligence*, da Pós-Graduação em Gestão dos Sistemas de Informação Empresariais, do Instituto Superior de Contabilidade e Administração do Porto e tem como principal objetivo aprofundar os conhecimentos adquiridos na disciplina construindo para tal uma solução de *Data Warehouse* e de *Analysis Services* tendo como fonte a base de dados de demonstração da empresa *Northwind*.

Vivemos numa era em que os dados são o bem mais precioso e a necessidade de extrair informação e conhecimento destes dados é cada vez mais premente.

Os dados estão em todo o lado. Em base de dados, em ficheiros EXCEL, na “nuvem”, nas ruas, nos telefones, as suas fontes são imensas. A capacidade de relacionar todos estes dados de uma forma fácil de utilizar e permitir aos utilizadores extrair aquilo que necessitem para os seus negócios é uma habilidade fundamental para qualquer analista do século XXI.

Desta forma pretende-se com este trabalho mostrar as mais valias que uma solução de *Business Intelligence* poderia trazer para o negócio potenciando a vertente na analítica, descritiva e preditiva no processo de criação de KPIs e no sempre difícil processo de tomada de decisão.

Neste trabalho foi utilizada a arquitetura de modelação de *Datawarehouse* proposta por Ralph Kimball do Kimball Group (Kimball, Reeves, Ross, & Thornthwaite, 1998), sendo que o mesmo está dividido em cinco capítulos distintos onde se abordam diferentes temas.

No capítulo 1 pretende-se contextualizar este trabalho, explicando sucintamente em que consiste um *Datawarehouse* e de que forma poderá ser útil para o negócio.

No capítulo 2 será feita a descrição das tarefas necessárias à modelação e criação do *Datawarehouse*, o esquema em estrela e as respetivas dimensões e factos mensuráveis.

No capítulo 3 serão explicados os processos de ETL, através da criação da solução de *Integration Services*, demonstrando brevemente os fluxos de dados e respetivas transformações de forma a carregar o *Datawarehouse* com a informação necessária, o mais agregado possível.

No capítulo 4 será construída a solução de *Analysis Services* e respetivo cubo. Será também feita uma breve análise descritiva e preditiva utilizando as ferramentas que a Microsoft disponibiliza para este efeito, nomeadamente o EXCEL e o *PowerBI*.

Finalmente, no capítulo 5 será feita uma conclusão evidenciando, de certa forma, as lições aprendidas com este trabalho.

1. Definição e características de um *Datawarehouse*

Um *Data Warehouse* é um conjunto de tecnologias destinadas a permitir aos agentes de tomada de decisão que o façam melhor e mais rapidamente. Muitos investigadores e profissionais compartilham a opinião de que uma arquitetura de *Data Warehouse* pode ser formalmente entendida como camadas de vistas materializadas umas sobre as outras. Uma arquitetura de *Data Warehouse* exhibe várias camadas de dados nas quais os dados de uma camada são derivados de dados da camada inferior. (Vassiliadis, 2000)

O *Datawarehouse* traduz-se num software que permite extrair os dados transacionais de uma empresa – dados de recursos humanos, contabilidade, ordens de fabrico, clientes, emails – limpa, transforma e valida estes dados, carregando-os posteriormente de uma forma inteligente, através de um processo denominado de ETL, para uma base de dados com um formato hierárquico e organizado em grupos de dimensões de análise e em grupos agregados de informação mensurável. Desta forma os dados poderão ser filtrados, agregados e apresentados das mais diversas formas, sempre contextualizados de uma forma multidimensional, permitindo a quem os observa tirar conclusões rápidas sobre o negócio e sobre as decisões a tomar.

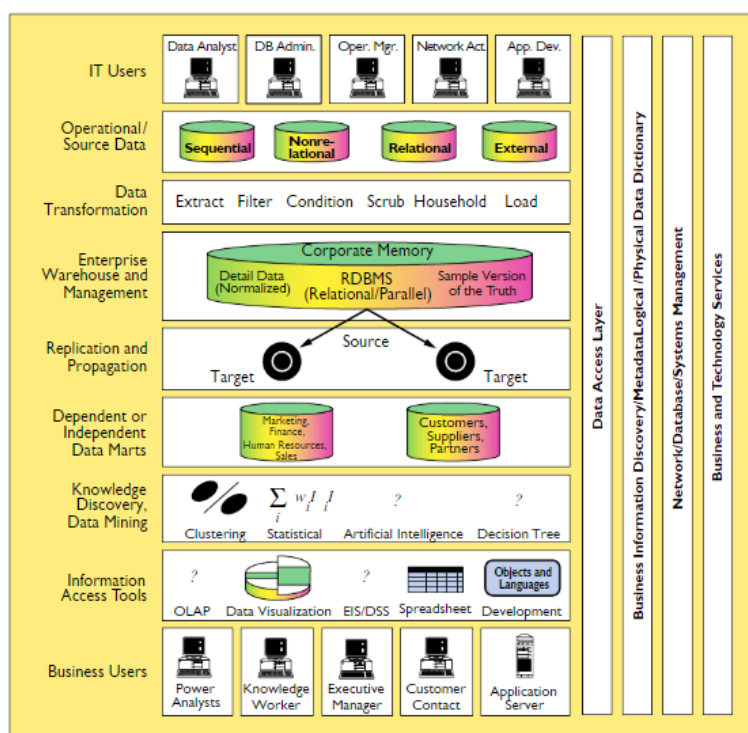


Figura 1 – Datawarehouse Framework (GARDNER, 1998)

Poderá ser colocada a questão da necessidade de um *Datawarehouse* se os dados já se encontram organizados em bases de dados relacionais.

Para um gestor que desconhece as vantagens que um sistema de *Business Intelligence* poderá trazer para o seu negócio, poderia facilmente ficar com ideia que este processo se limita a passar os dados de um sistema para o outro, aumentando a redundância dos dados e dos meios e consequentemente, dos custos inerentes ao processo.

No entanto, é fundamental transmitir e reforçar a ideia que, historicamente, tem-se vindo a comprovar que obter informação de diferentes fontes e sistemas poderá ser uma tarefa difícil ou impossível. Quantidades significativas de dados poderão ser de extração e leitura difíceis, e sem grande utilidade. Mostrar um relatório com 100 páginas de dados não traz nenhuma mais valia para a tomada de decisão.

Extrair a informação de vários sistemas e bases de dados poderá exigir conhecimentos técnicos muito elevados e normalmente as pessoas encarregues de efetuar análises não são especialistas em bases de dados nem em programação.

Outro fator muito importante é a heterogeneidade da informação dentro de uma organização. Efetivamente é comum que diferentes departamentos tenham os dados organizados de forma diferente. O mesmo indicador poderá ser quantificado e obtido de formas diferentes.

É, portanto, fácil de constatar as reais vantagens na implementação de uma solução de *Datawarehouse*, e podem traduzir-se no seguinte conjunto de características:

- **Acessibilidade** – permite uma fácil aprendizagem e utilização, e centraliza um vocabulário comum de BI para todos os utilizadores;
- **Rapidez** – os resultados deverão ser obtidos no mínimo tempo possível;
- **Consistência** – os resultados deverão ser credíveis e garantir que os dados que são obtidos em diferentes fontes são processados, limpos, transformados e validados, sem se perder ou adulterar a informação e ser mensuráveis na mesma unidade de medida;
- **Flexibilidade** – permitir á organização a mudança e escalabilidade do processo de ETL sem perder a informação existente.
- **Confidencialidade** – os dados sensíveis deverão ser guardados garantindo a segurança e controlo de acessos à informação
- **Utilidade** – os dados e informação obtidos deverão ser uteis ao processo de decisão. Este é, provavelmente, a característica mais importante de um sistema deste tipo.

Sendo assim, para a construção de um *Datawarehouse* deverão ser tidos em linha de conta as seguintes considerações e desafios:

- É necessário saber de que se trata o negócio, e quem são os utilizadores e de que tipo de informação precisam de uma solução de BI, e que tipo de decisões terão que tomar.

- Entregar informação de grande qualidade, relevante e de interpretação acessível. A informação deverá ser sempre precisa, confiável e adequada aos conhecimentos dos utilizadores.
- Adequar o ambiente de *Datawarehouse*, garantindo que funciona sem falhas e é carregado periodicamente.

Com um *Datawarehouse* bem implementado é possível assegurar e aumentar a participação da organização no mercado, aumentar significativamente o seu lucro e diminuir custos. Isso é consequência da flexibilidade com que estas organizações passam a tomar as decisões estratégicas. Os indicadores de gestão informam diariamente as variações do mercado e que decisões devem ser tomadas para que as perdas sejam minimizadas e o lucro aumentado. A diminuição de custos é evidente quando se compara os relatórios de empresas com e sem um *Datawarehouse* implementado. (Caldas, Sandelari, & Kovalski, 2006)

1.1. Tabelas de Factos

Os factos são as medidas ou métricas que se podem calcular ou contabilizar de um determinado modelo de negócio. Exemplos de factos são volume de vendas, quantidades vendidas, entre outras.

Os factos podem ser vistos e enquadrados com a ajuda de dimensões.

A tabela de factos é a principal tabela de um modelo dimensional, onde as medições numéricas de interesse da empresa estão armazenadas (Kimball, Reeves, Ross, & Thornthwaite, 1998). A palavra "facto" representa uma medida dos processos a serem modelados, como quantidades, valores e indicadores. A tabela de factos regista os factos que serão analisados. É composta por uma chave primária (formada por uma combinação única de valores de chaves de dimensão) e pelas métricas de interesse para o negócio. As dimensões indicam a forma como as medidas serão vistas, ou seja, são os aspetos pelos quais se pretende observar as métricas. A intersecção das chaves de dimensão define a granularidade da tabela de factos, e é importante que todas as medidas na tabela de factos tenham a mesma granularidade.

A granularidade diz respeito ao nível de detalhe ou de resumo contido nas unidades de dados existentes no *Datawarehouse* (Inmon, 1997) Quanto mais detalhe, mais baixo o nível de granularidade. Quanto menos detalhe, mais alto o nível de granularidade.

1.2. Tabelas de Dimensão

As dimensões fornecem o contexto de um modelo de negócio. São os elementos descritivos que definem os factos. Se, por exemplo, considerarmos o volume de vendas como um facto, poderemos ter como dimensões os produtos mais vendidos, as lojas que mais venderam, ou o tipo de clientes que mais compraram.

Os atributos das dimensões poderão aumentar a capacidade de descrever os factos. Atributos como, por exemplo, a idade dos clientes, sexo, entre outras, poderão aumentar a capacidade descritiva e de segmentação dos dados.

A qualidade da base de dados é proporcional à qualidade dos atributos de dimensões, portanto deve ser dedicado tempo e atenção a sua descrição, ao seu preenchimento e a garantia da qualidade dos valores em uma coluna de atributos (Kimball, Reeves, Ross, & Thornthwaite, 1998).

1.3. Esquema em estrela

O nome "estrela" está associado à disposição das tabelas no modelo, que consiste de uma tabela central, a tabela de fatos, que se relaciona com diversas outras tabelas, as tabelas de dimensão. Este tipo de esquemas são estruturas dimensionais que residem num sistema de base de dados relacional. Se o esquema estiver devidamente construído, a tabela de factos juntamente com as tabelas de dimensão assemelha-se a uma estrela (Santos & Ramos, 2006).

Este tipo de modelação das tabelas cumpre dois requisitos fundamentais de uma solução de BI:

- Os dados são compreendidos e contextualizados pelos utilizadores.
- Permite a execução de consultas de uma forma rápida e eficaz.

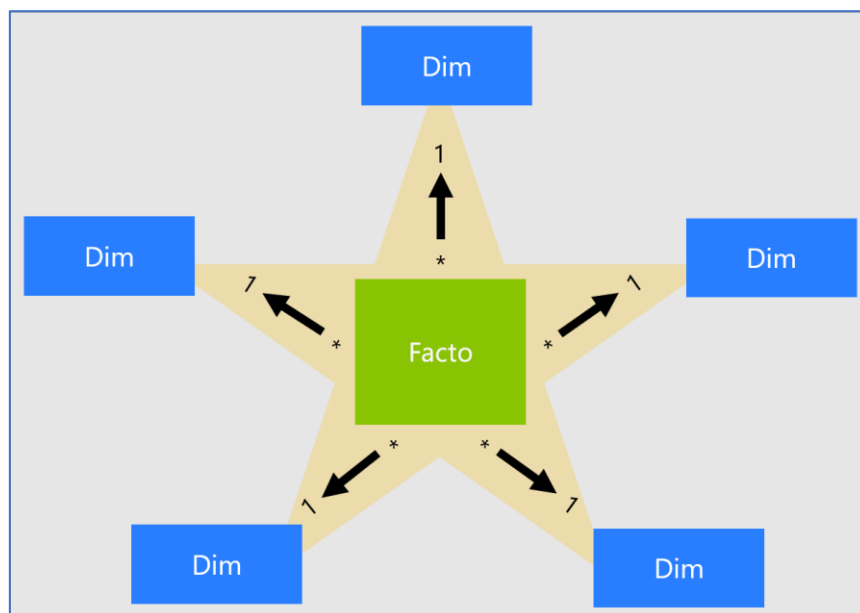


Figura 2 – Esquema em Estrela

Analisando o seguinte exemplo teórico de construção de um esquema em estrela supõem-se que numa reunião, o cliente descreve o seu negócio da seguinte forma:

“Nós vendemos gelados e outros produtos em várias lojas e pretendemos medir a nossa performance ao longo do tempo.”

A partir desta informação pode assumir-se alguma informação para a construção do esquema. Produtos, Lojas e Tempo são dimensões de análise óbvias. *Performance* é um termo que pode significar volume de vendas, lucro, ou quantidades vendidas.

Trata-se de um modelo simples, mas serve de procedimento de entrada para a análise de requisitos para a construção e modelação do *Datawarehouse*.

2. Análise do Negócio e Modelação do *Datawarehouse*

Foi decidido, pelos elementos do grupo, que seria fundamental aprender a fazer todo o processo de *Business Intelligence*, simulando o mais aproximado possível, uma situação em ambiente real. Ou seja, entender um modelo de negócio, estudar os dados, simular os KPIs que pudessem vir a ser necessários estudar, efetuar o processo ETL, e, por fim, conseguir disponibilizar aos utilizadores uma solução de análise para consulta, foi o desafio que nos propusemos abraçar, mesmo sabendo da dificuldade do processo.

Sendo assim, foi necessário escolher um modelo de negócio que fosse adequado ao trabalho, e estudar o modelo de dados que o suporta. Foram definidos três requisitos fundamentais nesta pesquisa, nomeadamente:

- O modelo de negócio deveria de ser simples, perçetível e semelhante a situações reais que pudéssemos vir a encontrar num futuro próximo;
- Os dados deveriam de estar acessíveis e organizados num sistema relacional;
- Os dados deveriam evidenciar transações, no mínimo, ao longo de três anos, de forma a ser possível a análise temporal e acompanhar a evolução dos indicadores.

Para o efeito, e tendo em consideração estes requisitos, foi escolhido o modelo de negócio *Northwind* e a sua base de dados em formato Access.

Este modelo de dados, apesar de ser fictício, assemelha-se e adequa-se perfeitamente a um ERP de pequenas dimensões, e poderia ser um exemplo real de uma empresa portuguesa.

2.1. Estrutura da Base de Dados *Northwind*

A *Northwind* é uma base de dados de demonstração utilizada pela Microsoft para demonstrar as funcionalidades dos seus produtos. Está disponível no formato SQL Server e em Access. Esta base de dados contém a informação das vendas de uma empresa fictícia, denominada “*Northwind Traders*” cujo modelo de negócio se traduz na importação e exportação de alimentos.

O diagrama de entidades e relacionamentos é o seguinte:

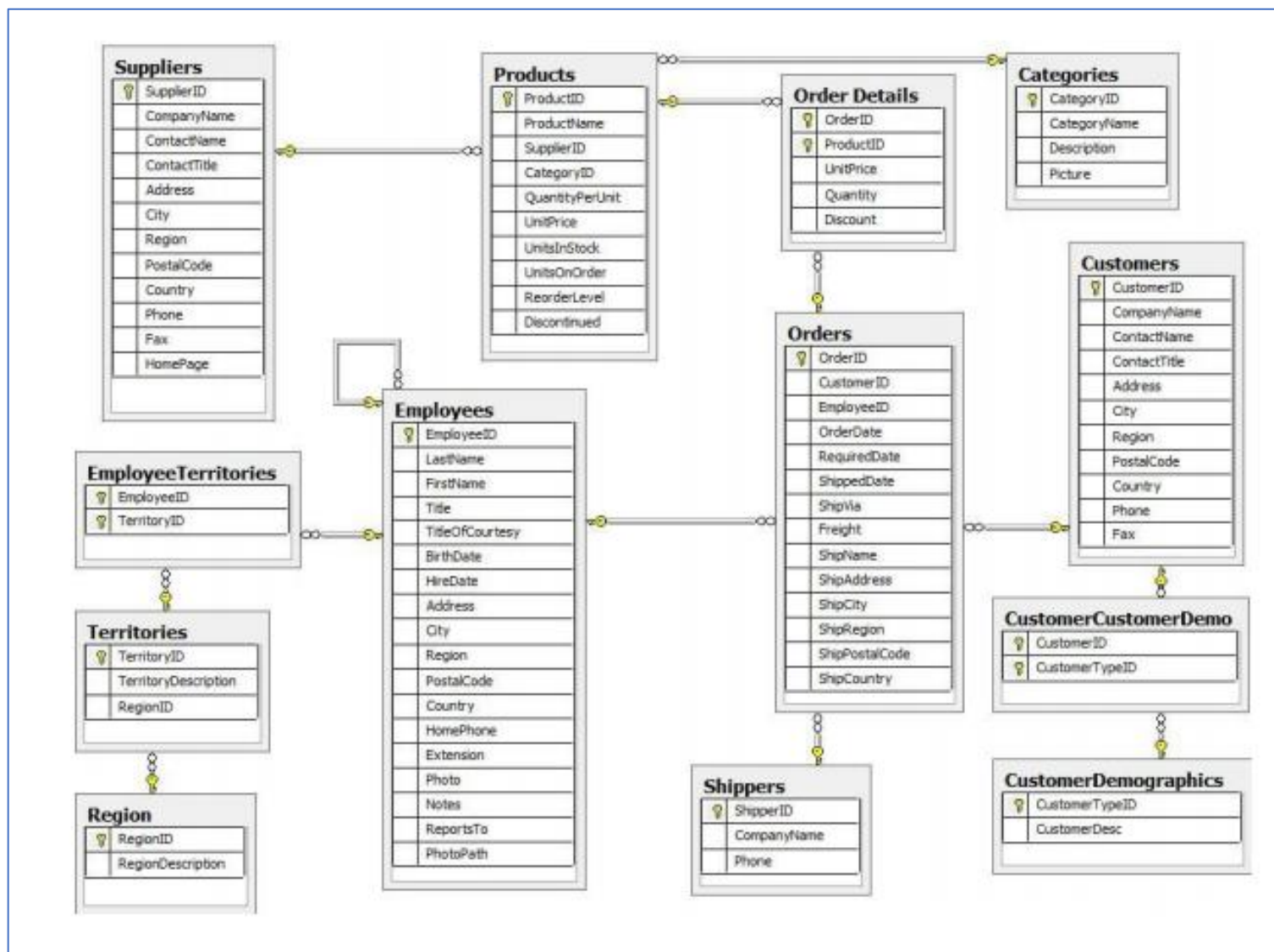


Figura 3 - Diagrama ER da Base de Dados Northwind

As tabelas principais, onde se baseia o modelo de negócios e onde incidirá este trabalho, são as seguintes:

- **Suppliers** – tabela com a lista de fornecedores das mercadorias e seus atributos;
- **Customers** – tabela com a lista de clientes que compram os produtos da *Northwind*, e seus atributos, nomeadamente a localização;
- **Products** – tabela com o inventário dos produtos e seus atributos, nomeadamente o seu preço unitário, e a categoria.
- **Categories** – tabela com a lista de todas as categorias de produtos existentes. Todos os produtos estão contextualizados com o identificador de categoria.
- **Order e Order_Details** – nestas duas tabelas estão evidenciadas as transações de vendas estabelecidas com os respetivos clientes e a empresa. Destacam-se os atributos da data da venda, quantidades vendidas, preço e desconto praticado.

Após o estudo dos dados pelos elementos deste grupo, foi facilmente perceptível quais os factos que poderiam ser medidos e que dimensões poderiam ser escolhidas para contextualizar essas métricas.

2.2. Modelação do *Datawarehouse*

Sendo assim, as perguntas que fizemos, e que nos ajudaram a modelar o *Datawarehouse* foram as seguintes:

- “Qual o volume de vendas por produto, categoria e local, ao longo do tempo?”
- “Qual a quantidade vendida por produto, categoria e local, ao longo do tempo?”

Foi respondendo a estas perguntas fundamentais que consolidamos a definição de como deveria de ser a estrutura do nosso armazém de dados. A construção do nosso esquema em estrela traduziu-se no seguinte:

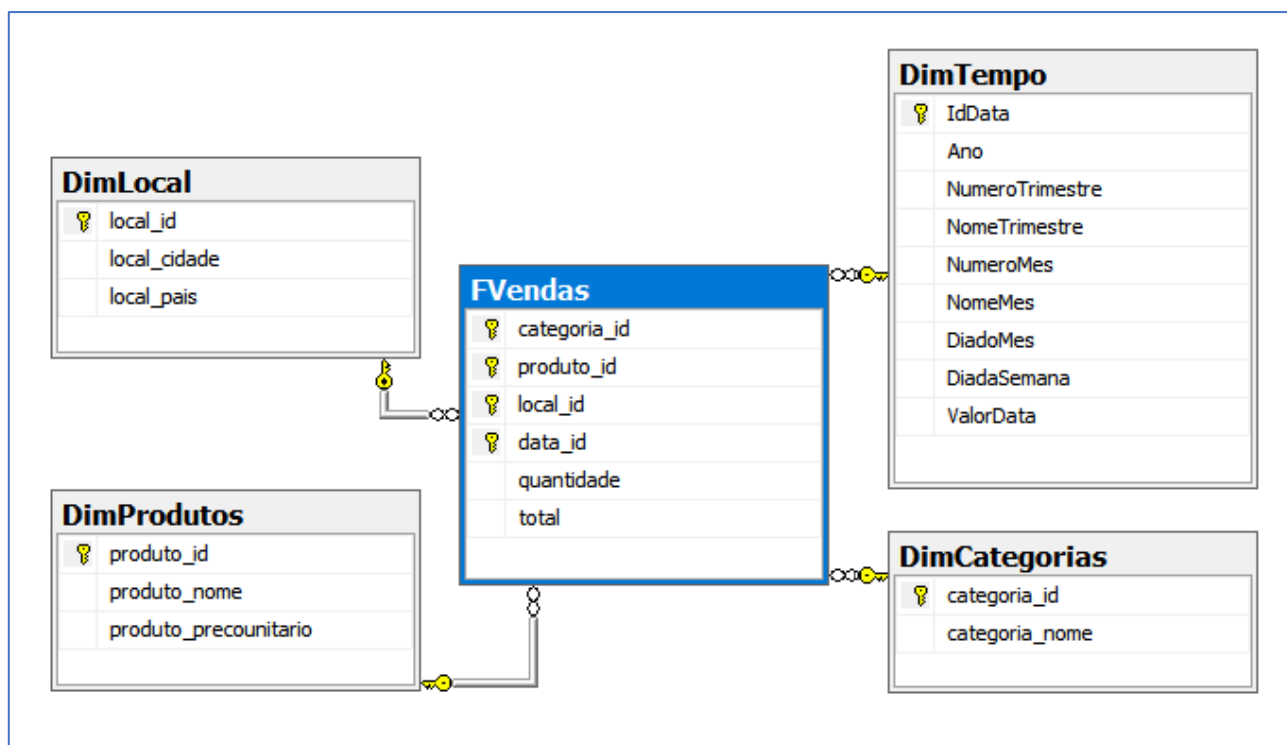


Figura 4 - Definição do Datawarehouse

Como pode ser observado neste diagrama de entidades e relacionamentos, foram definidas quatro dimensões de análise: Tempo, Local, Produto e Categoria e uma tabela de factos mensuráveis, nomeadamente o total do volume das vendas e a quantidade vendida.

3. Processo de Integração ETL – *Extract, Transform and Load*

Tendo o *Datawarehouse* definido, foi necessário definir uma estratégia para o abastecer com a informação transacional, existente no modelo de dados que suporta o negócio- Neste caso, a base de dados Access *Northwind*. Normalmente, neste tipo de situações é utilizado um processo ETL.

ETL é um tipo de “*data integration*” em três etapas (extração, transformação, carregamento) usado para combinar dados de diversas fontes. Neste processo, os dados são retirados (extraídos) de um sistema-fonte, convertidos (transformados) em um formato que possa ser analisado, e armazenados (carregados) em um armazém ou outro sistema. (SAS, 2020)

O ETL é um processo para extrair dados de um sistema de Bases de Dados, sendo esses dados processados, modificados, e posteriormente inseridos numa outra Base de dados. Estudos relatam que o ETL e as ferramentas de limpeza de dados consomem um terço do orçamento num projeto de *Datawarehouse*, podendo, no que respeita ao tempo de desenvolvimento de um projeto de *Datawarehouse*, chegar a consumir 80% desse valor. Outros estudos mencionam, ainda, que o processo de ETL tem custos na ordem dos 55% do tempo total de execução do projeto de *Datawarehouse* (Ferreira, Miranda , Abelha , & Machado, 2010)

Existem no mercado várias ferramentas de ETL. Neste trabalho em concreto foi escolhida o *SQL Server Integrations Services*.

3.1. Projeto de *Integrations Services SSIS*

O Projeto de *Integration Services* foi criado utilizando o *Visual Studio 2019*. A arquitetura deste projeto foi dividida nas seguintes tarefas fundamentais:

- Definição das conexões com as diferentes fontes de dados (*Connection Manager*);
- Definição do Controlo de Fluxos (*Control Flow*);
- Definição do Fluxo de Dados (*Data Flow*);
- *Deploy* do package em ambiente produtivo.

3.1.1. Conexões às Bases de Dados (*Connection Manager*)

Relativamente às conexões com as fontes de dados foram criadas quatro ligações, que permitem fazer a importação dos dados do Access (Jet 4.0), EXCEL e sua tabela de tempo, e as tabelas do *Datawarehouse* no SQL Server.

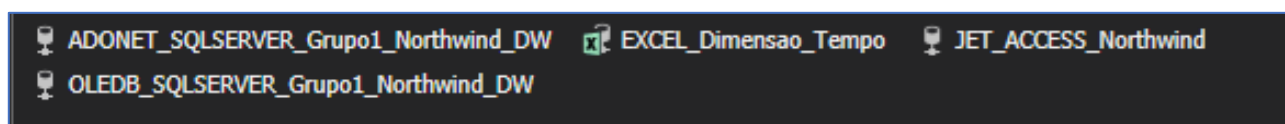


Figura 5 - Ligações configuradas no Connection Manager

3.1.2. Definição do Controlo de Fluxos (*Control Flow*)

No que diz respeito ao Controlo de Fluxos, este processo foi dividido em duas etapas: Eliminação dos registos no *Datawarehouse*, e posterior carregamento das tabelas de dimensão e de factos.

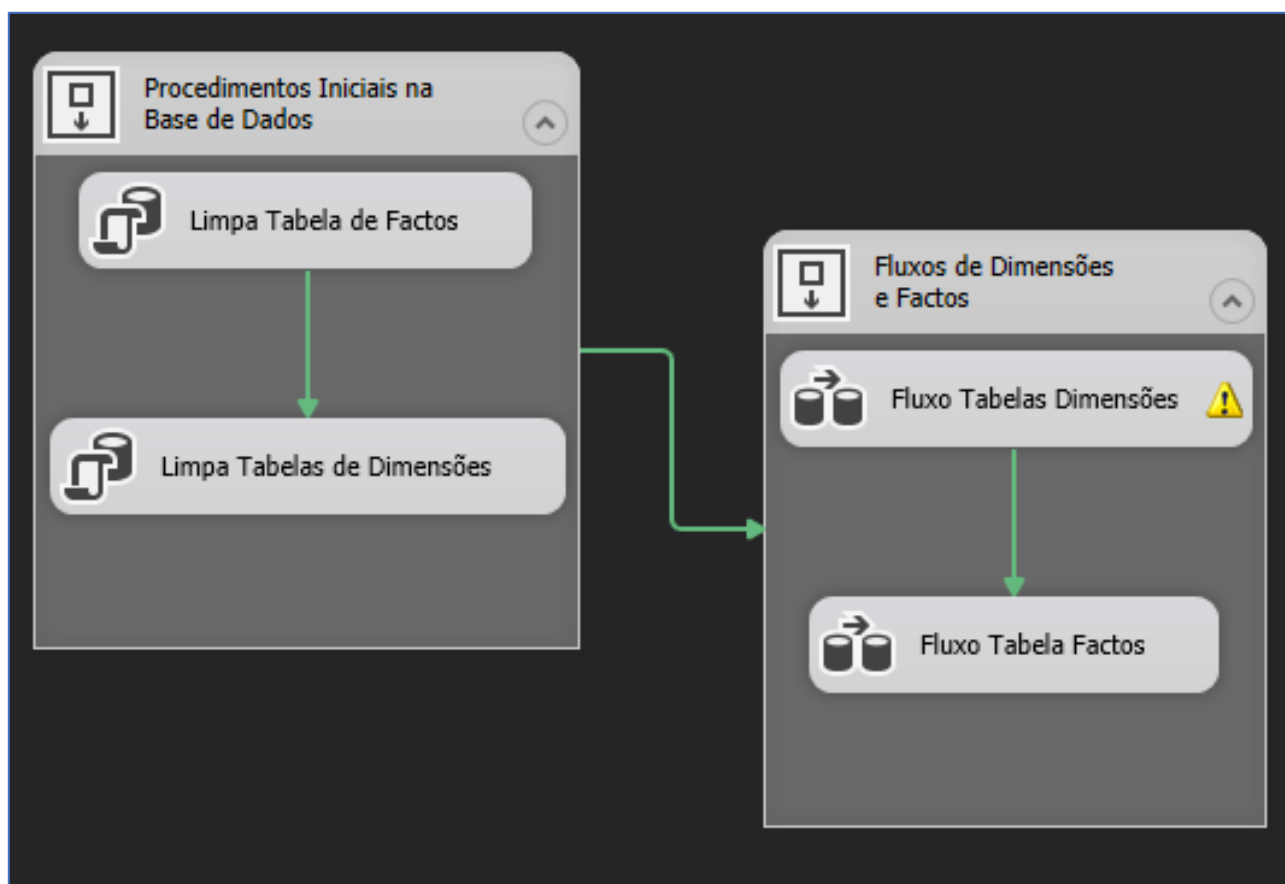


Figura 6 - Definição do Controlo de Fluxos do Projeto

3.1.3. Definição do Fluxo de Dados (*Data Flow*)

O Processo mais detalhado é o de Fluxo de Dados. Efetivamente é nesta etapa que o processo ETL de facto ocorre. É no processo ETL que os dados são extraídos, transformados e finalmente depositados no seu destino, o armazém de dados. Este processo foi efetuado para cada uma das tabelas de dimensão e para a tabela de factos.

No caso em concreto deste trabalho, pode-se dizer que o processo de Fluxo de Dados da tabela de Factos foi bastante complexo, e foi sem dúvida, um desafio aliciante para os elementos deste grupo. O processo de tentativa e erro fez com que os conhecimentos de *Integration Services* e suas ferramentas ficassem bem presentes, o que não aconteceria se tivéssemos o *Datawarehouse* pré-carregado.

Definitivamente, foi nesta etapa que confirmamos que a nossa escolha pelo desenvolvimento de raiz deste processo tornou-se uma mais valia.

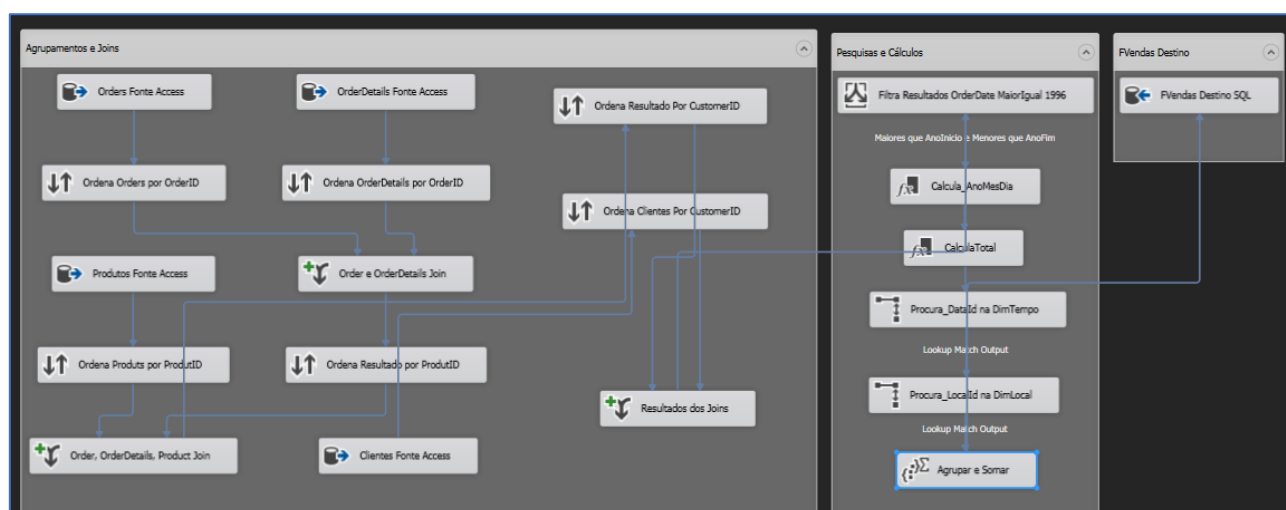


Figura 7 - Definição do Data Flow do Projeto

O processo poderá ser mais facilmente confirmado pelo projeto de SSIS que acompanha este relatório, mas de uma forma geral foi dividido em três grandes processos:

- Processo *Merge and Joins* das diferentes tabelas existentes nas diferentes fontes – neste processo foram utilizados essencialmente os componentes de Ordenação e de *Merge Joins*.
- Processos de transformação e de cálculos de dados – neste processo foram calculados os valores totais de faturação e foram definidos os Ids da dimensão tempo e da dimensão Locais. Para o efeito, foram essencialmente utilizados os componentes de *Derived Columns* e de *Lookup*
- Processo de agregação da informação e respetivo *Load* para o *Datawarehouse*.

Após a execução destes três grandes processos de fluxos de dados foi necessário confirmar a consistência de dados. Sendo assim, numa primeira instância foi garantido que todos os dados que foram processados no processo ETL estariam no *Datawarehouse*. Neste caso em concreto seria expectável que a tabela de factos estivesse carregada com 2154 registos, após todas as etapas de transformação e agregação.

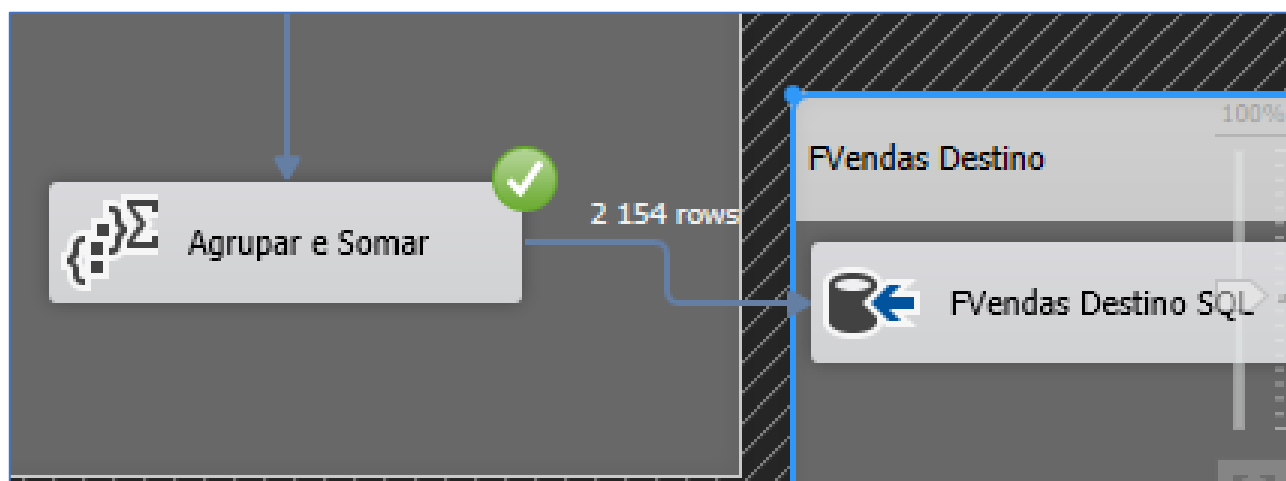


Figura 8 - 2154 registos processados pelo processo ETL

```
SQLQuery1.sql - WIN1...40G\yuiribeiro (53))* X
1  /***** Script for SelectTopNRows command from SSMS *****/
2  SELECT count(*)
3  FROM [Grupo1_NorthWind_DW].[dbo].[FVendas]
```

100 %

Results Messages

	(No column name)
1	2154

Figura 9 - Instrução SQL para obter 2154 registos no datawarehouse

Outra análise importante, para garantir a credibilidade dos dados, foi confirmar que os cálculos efetuados no processo ETL foram efetuados sem erros. Desta forma, foi essencial garantir que o total de vendas fossem iguais, tanto no sistema de origem como no destino.

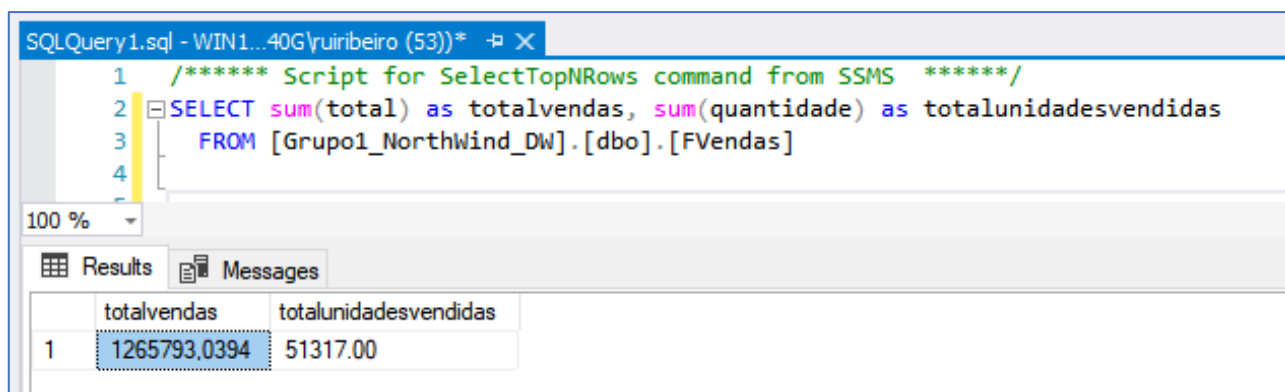


Figura 10 - Instrução sql para obter o volume de vendas e quantidades vendidas no datawarehouse

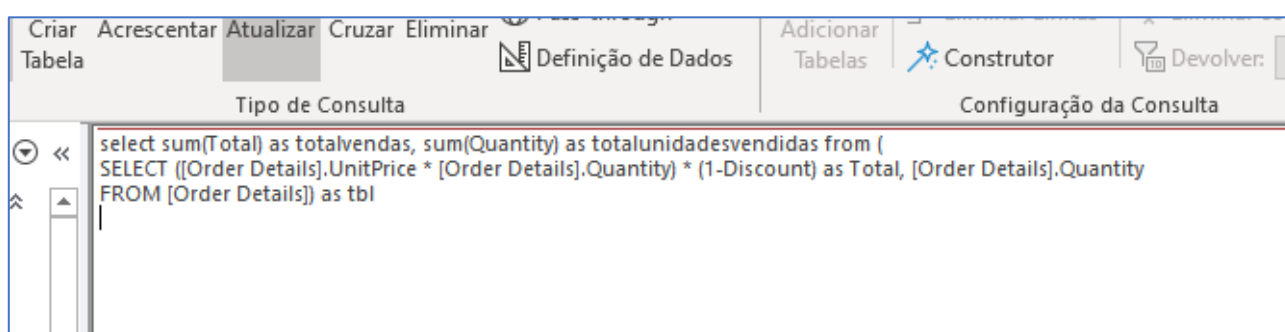


Figura 11 - Instrução sql na base de dados Access

totalvendas	totalunidadesvendidas
1265793,0383	51317

Figura 12 - Resultados obtidos no Access

Como se pode observar, os valores bateram certo. Com estas análises conseguiu-se mitigar o risco da incoerência da informação e garantir que as análises que possam vir a ser efetuadas sejam corretas.

3.1.4. Deploy do package em ambiente produtivo

A tarefa final neste capítulo foi a disponibilização destas funcionalidades no servidor de *Integrations Services* do *SQL Server*, de forma a ser automaticamente e periodicamente executado, garantindo que os dados ficam integrados todos os dias. Normalmente este processo decorre durante um período não produtivo (madrugada) para limitar ao máximo o impacto na performance dos ambientes produtivos.

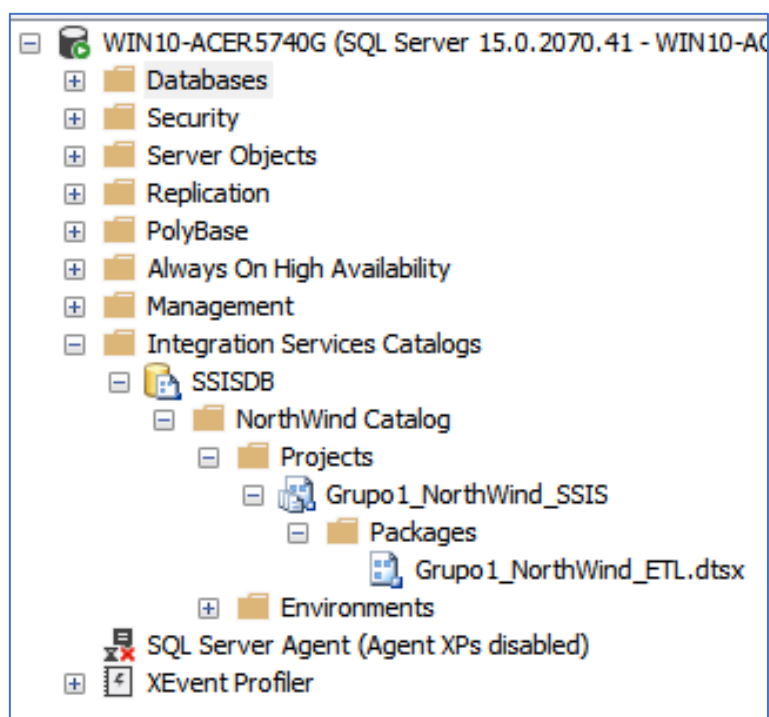


Figura 13 - Deploy do projeto no SS Integrations Service

Foi definida uma tarefa recorrente (job) que permite a execução automática do processo. O resultado obtido por este job pode ser constatado através de um relatório disponibilizado no *SQL Server*.

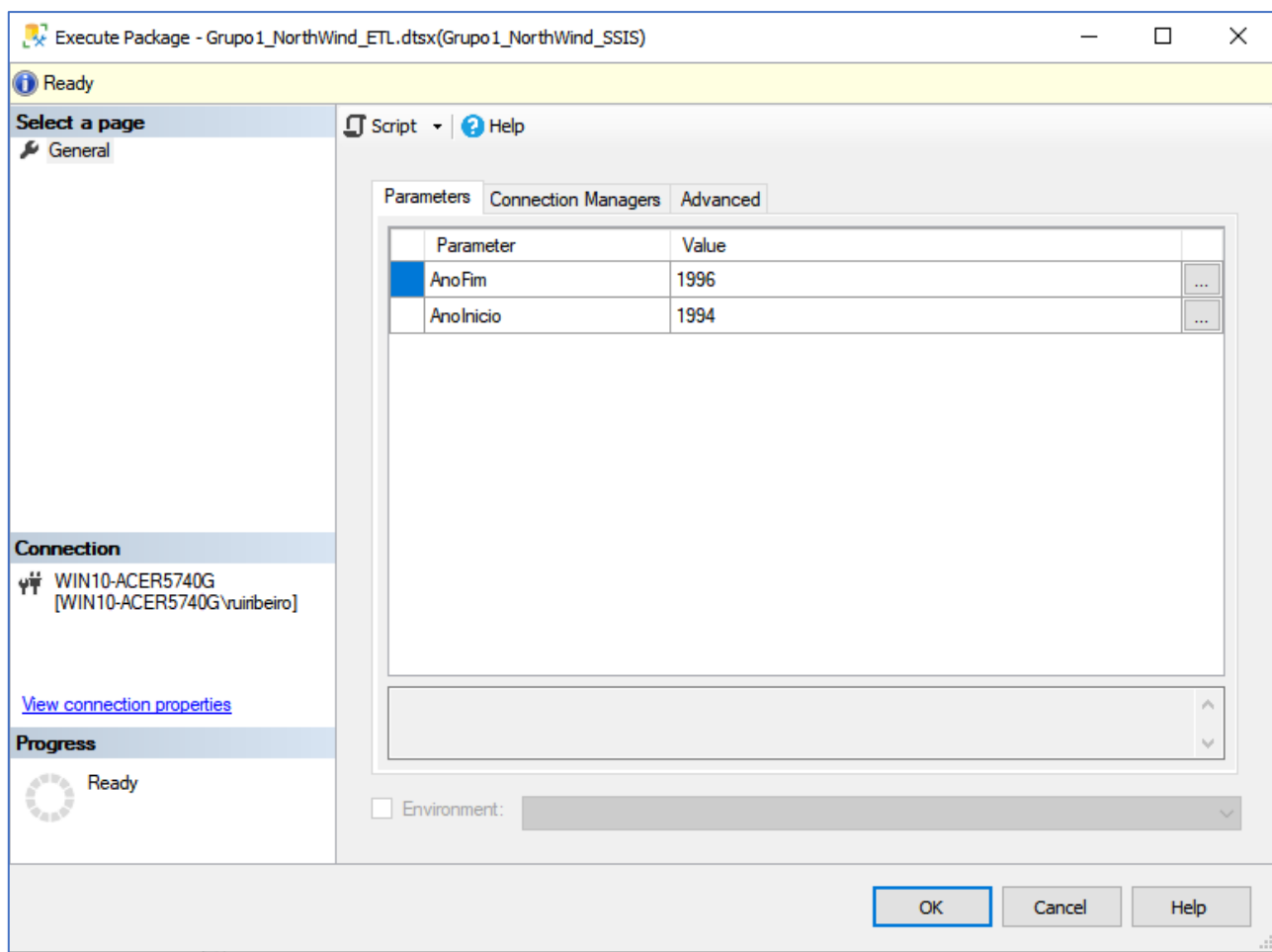


Figura 14 - Parâmetros de execução do package

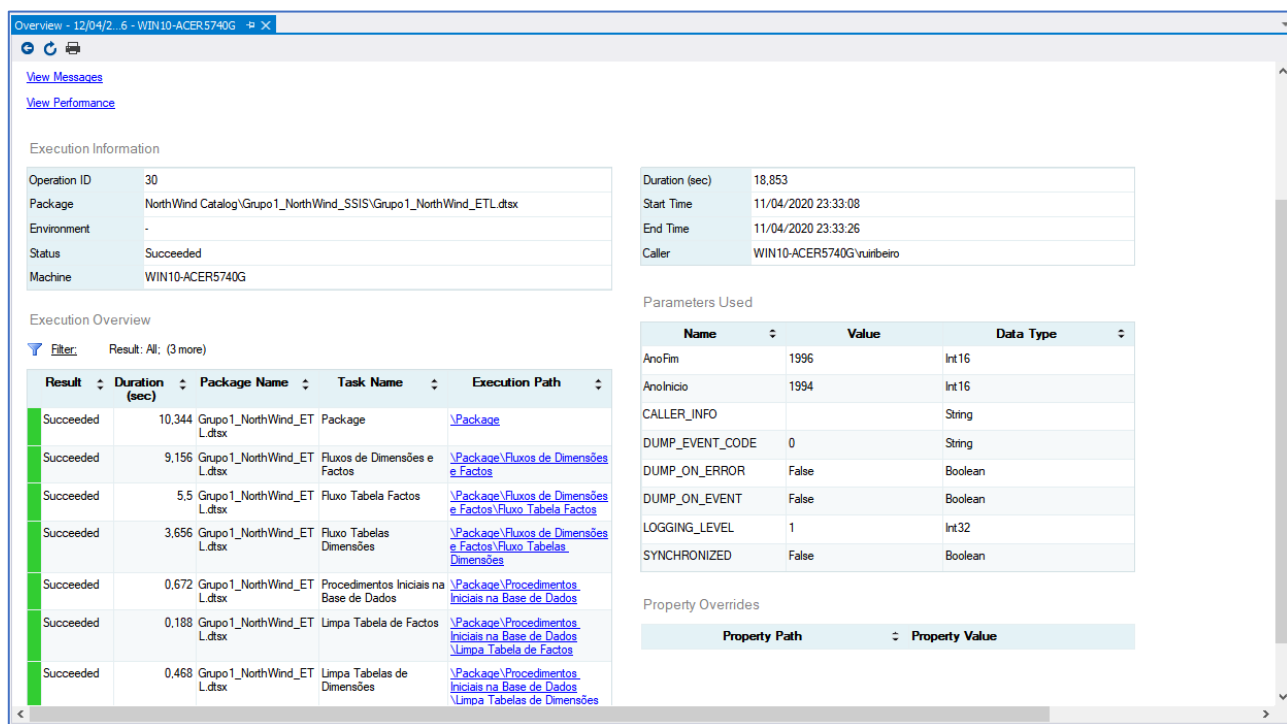


Figura 15 - Resultados da execução periódica do package (jobs)

4. Processo de Criação do Cubo OLAP e de Análise – *Analysis Services, EXCEL e PowerBI*

4.1. *Analysis Services*

Analysis Services é uma base de dados online analítica de processamento (OLAP) fornecida pela Microsoft. É um tipo de base de dados altamente otimizada e escalada para todo o tipo de consultas e cálculos comuns num ambiente empresarial comum. Apesar de fazer o mesmo que qualquer base de dados relacional, a sua estrutura difere em muitos aspetos. Na maioria dos casos será mais fácil desenvolver uma solução de *Business Intelligence* em combinação com uma base de dados relacional, como por exemplo o SQL Server, do que optar apenas pela utilização deste último. O *Analysis Services* não substitui a necessidade de uma base de dados relacional nem um bem *Datawarehouse* bem desenhado. (Ferrari, Russo, & Webb, 2012).

4.2. Cubo OLAP

Um cubo OLAP é uma estrutura de dados que elimina as limitações de bases de dados relacionais, fornecendo uma análise rápida de dados. Cubos OLAP podem apresentar grandes quantidades de soma de dados enquanto também fornecem aos utilizadores acesso pesquisável a quaisquer pontos de dados para que os dados podem ser agregados, segmentados e repartidos conforme necessários para lidar com a grande variedade de perguntas relevantes à área de um utilizador de interesse ser possível. (Descrição geral dos cubos OLAP do Service Manager para análise avançada, 2019).

4.3. Projeto de *Analysis Services SSAS*

A ferramenta utilizada para a criação do projeto de *Analysis Services* foi o *Visual Studio 2019*. A facilidade de integração do *Visual Studio* com o *SQL Server* é uma vantagem fundamental para a criação deste tipo de projetos.

A arquitetura deste projeto foi dividida nas seguintes tarefas fundamentais:

- Definição das conexões com o *Datawarehouse*;
- Definição do Cubo OLAP;
- *Deploy* do cubo em ambiente produtivo.

4.3.1. Conexão ao Datawarehouse

À semelhança do projeto de *Integrations Services*, o primeiro passo foi definir a ligação com a nossa fonte de dados para processamento e análise. Neste caso, o *Datawarehouse*, que se encontrava definido na base de dados SQL Server.

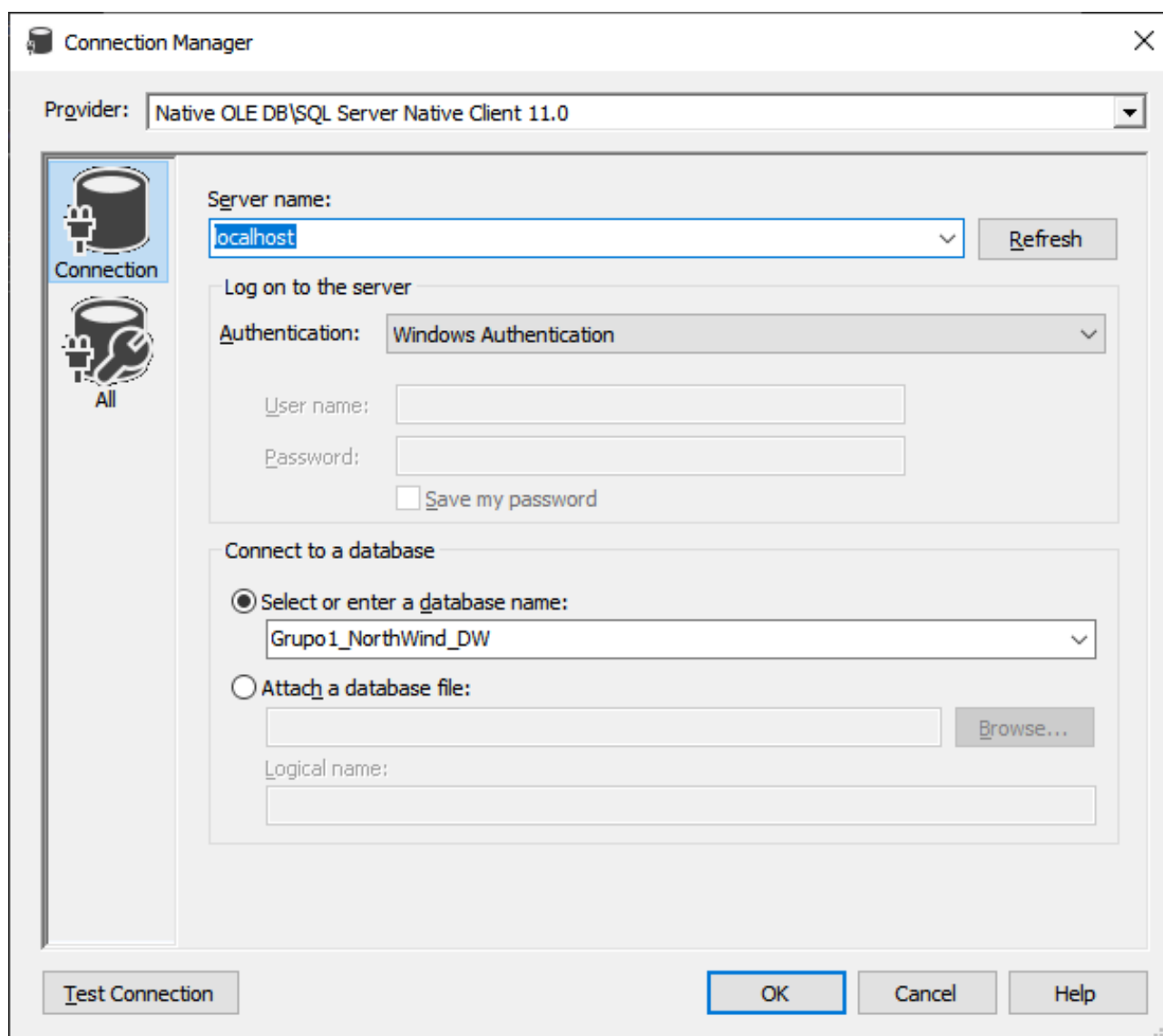


Figura 16 - Ligação ao Datawarehouse

4.3.2. Definição do Cubo OLAP

O Cubo OLAP foi definido através do assistente de criação disponibilizado pelo *Visual Studio*. Foram utilizadas as tabelas de dimensão e de fatos previamente criadas no *Datawarehouse*. O processo é bastante simples, já que indicando ao assistente qual a tabela de factos, e tendo o *Datawarehouse* bem definido, a identificação das respetivas tabelas de dimensão foi um processo automático.

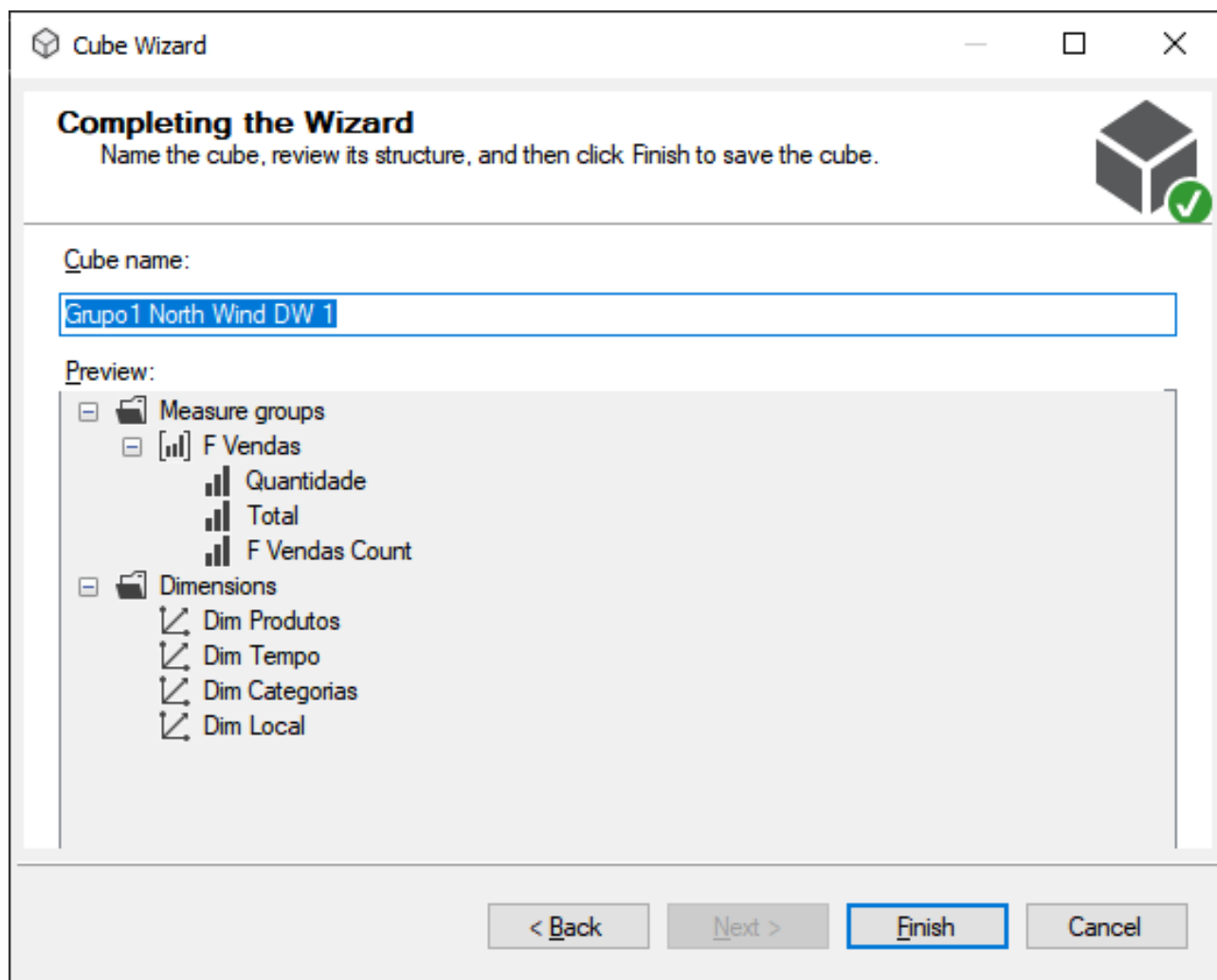


Figura 17 - Identificação da tabela de factos e de dimensão na criação do cubo

Após a conclusão de todas as etapas do assistente, foi criado o respetivo cubo.

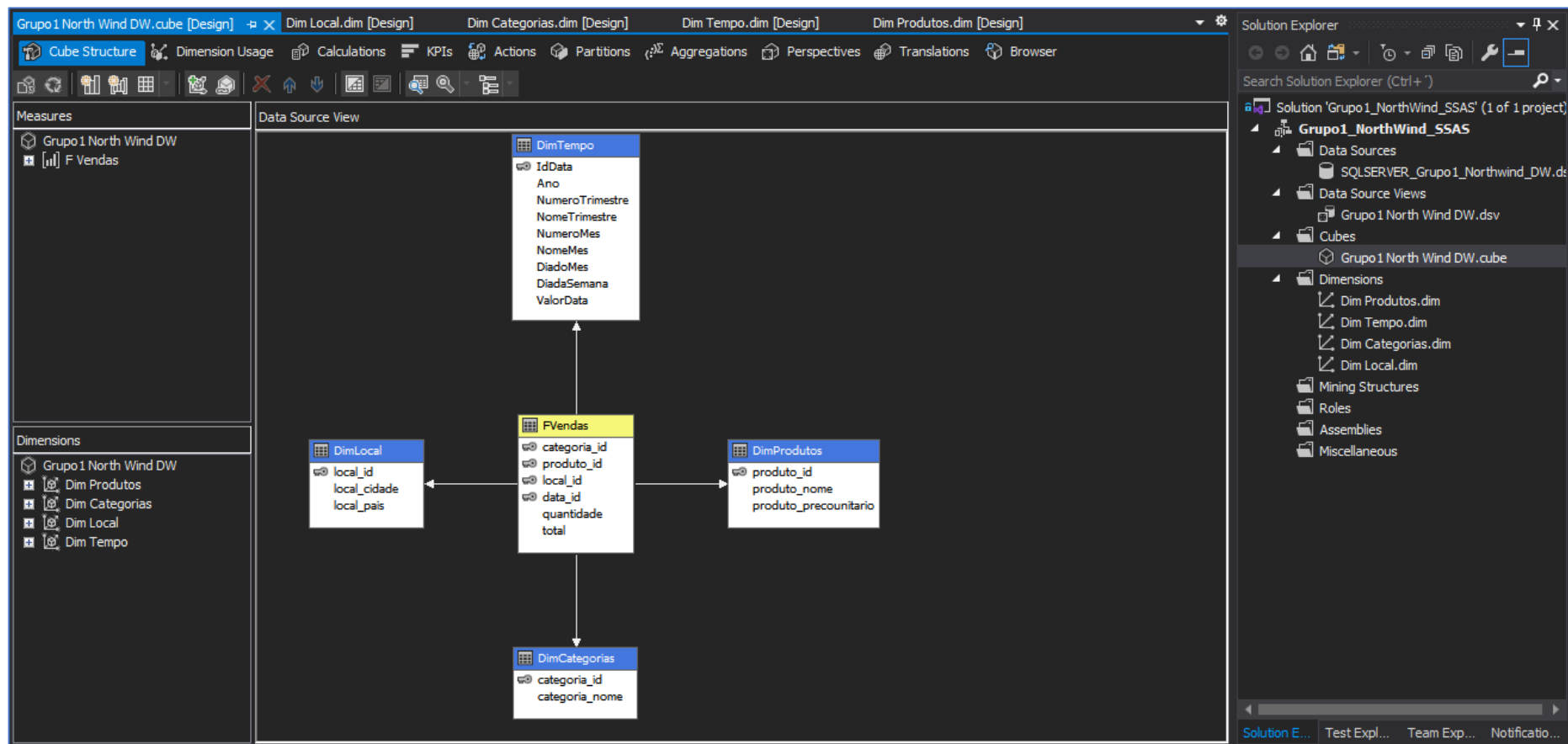


Figura 18 - Cubo OLAP

Desde logo foi possível começar a fazer as primeiras análises de dados agregados e KPIs, relativos ao modelo de Negócio da *Northwind Traders*, como por exemplo:

Quantidade	Total
51317	1265793,0394

Figura 19 - Total de Vendas e Quantidades Vendidas

Ano	Quantidade	Total
1994	7381	162844,3398
1995	25007	590926,4055
1996	18929	512022,2941

Figura 20 - Total de Vendas e Quantidades Vendidas Por Ano

Categoria Nome	Quantidade	Total
Beverages	9532	267868,1805
Condiments	5298	106047,0852
Confections	7906	167357,2249
Dairy Products	9149	234507,2849
Grains/Cereals	4562	95744,5875
Meat/Poultry	4199	163022,359
Produce	2990	99984,5799
Seafood	7681	131261,7375

Figura 21 - Total de Vendas e Quantidades Vendidas Por Categoria de Produto

Local Pais	Quantidade	Total
Argentina	339	8119,1
Austria	5167	128003,8387
Belgium	1392	33824,8549
Brazil	4247	106925,7765
Canada	1984	50196,2904
Denmark	1170	32661,0223
Finland	885	18810,0525
France	3254	81358,3226
Germany	9213	230284,6335
Ireland	1684	49979,905
Italy	822	15770,1549
Mexico	1025	23582,0776
Norway	161	5735,15
Poland	205	3531,95
Portugal	533	11472,3625
Spain	718	17983,2001
Sweden	2235	54495,14
Switzerland	1275	31692,6588

Figura 22 - Total de Vendas e Quantidades Vendidas Por País

4.3.3. Deploy do cubo em ambiente produtivo.

Após os testes terem sido efetuados com sucesso em ambiente de desenvolvimento (*Visual Studio*), foi necessário implementar a solução em ambiente produtivo. Neste caso, numa instância do *Analysis Services* do *SQL Server*.

Sendo assim, com o destino da implementação, definido nas propriedades do projeto, e após executar o processamento do cubo, podemos verificar que o *deploy* do cubo foi efetuado com sucesso no nosso ambiente de produtivo. Poderá se acedido com o *SQL Server Management Studio*, autenticando na instância ativa de *Analysis Services*.

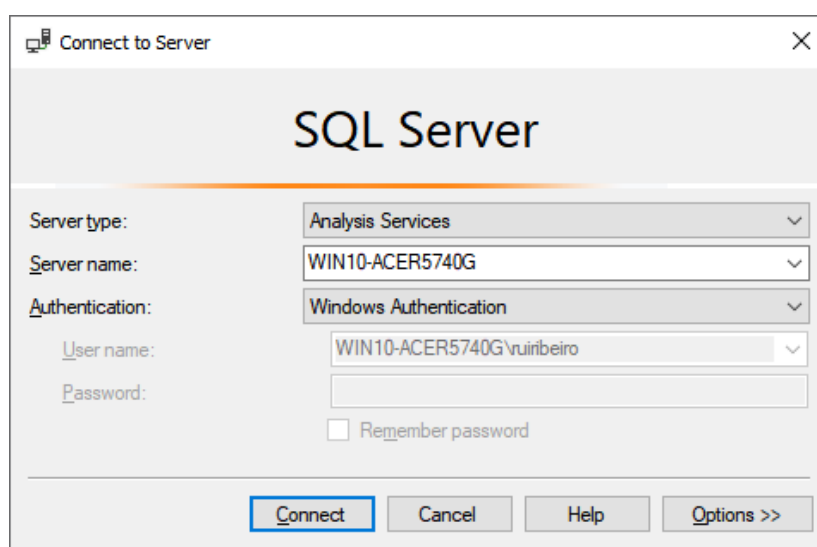


Figura 23 - Autenticação no SQL Server Analysis Services

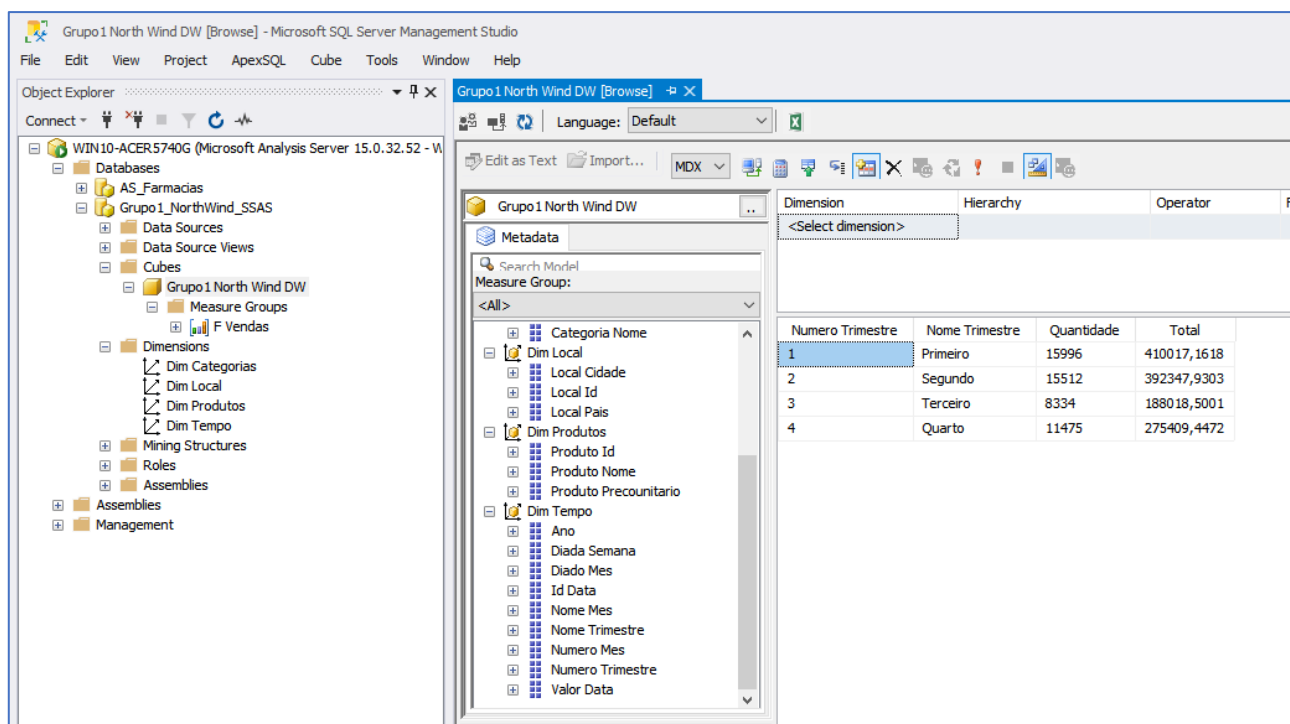


Figura 24 - Cubo implementado em ambiente de Produção no SQL Server

4.4. Visualização dos dados em EXCEL

Para visualizar e analisar os dados, de forma a transformá-los em informação útil para a decisão, são necessárias ferramentas com capacidades visuais mais adequadas do que os *Analysis Services do SQL Server*.

O EXCEL, o verdadeiro “canivete suíço” de gestão de dados de qualquer organização, ainda é uma opção muito útil para esta tarefa.

Sendo assim, aproveitando o cubo previamente implementado em ambiente produtivo, e com o objetivo de conhecermos as potencialidades, decidiu-se fazer algumas análises nesta ferramenta.

Neste sentido, o primeiro passo foi estabelecer uma conexão ao Cubo OLAP.

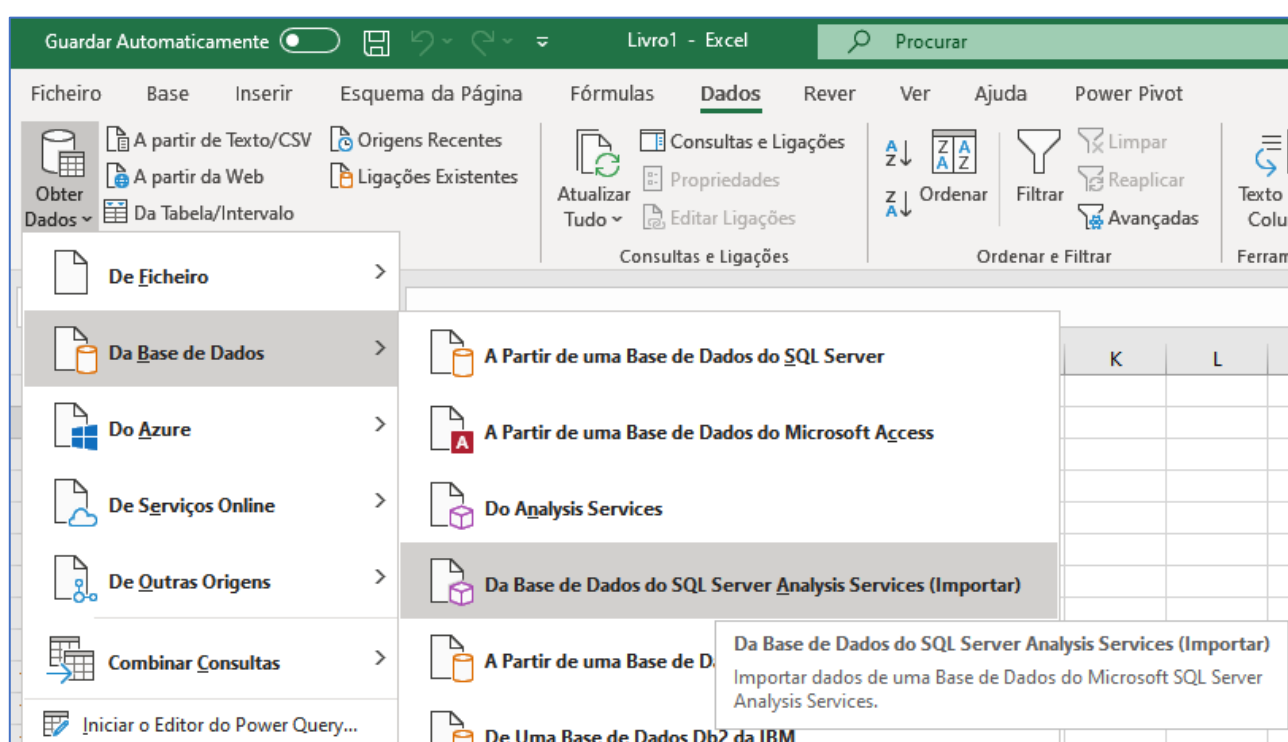


Figura 25 - Conexão do EXCEL para o Cubo OLAP

Tendo os dados carregados no EXCEL, torna-se muito fácil construir tabelas dinâmicas que construam visualizações adequadas ao modelo de negócio.

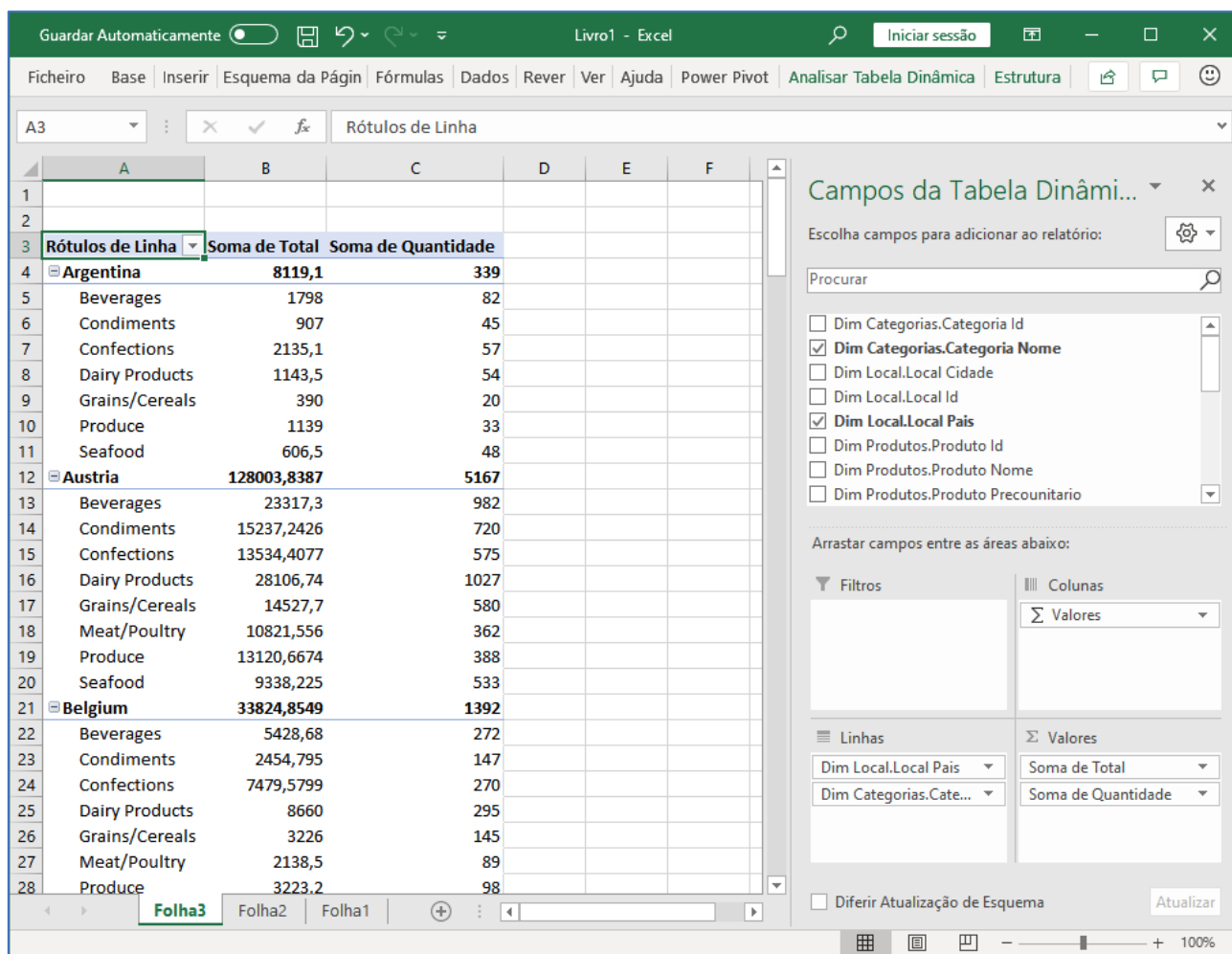


Figura 26 - Volume de Negócios e Quantidades por País e por Categoria

Outra opção que o EXCEL tem, na visualização da informação, é a capacidade de criar gráficos ricos em funcionalidades e com um aspeto muito apelativo.

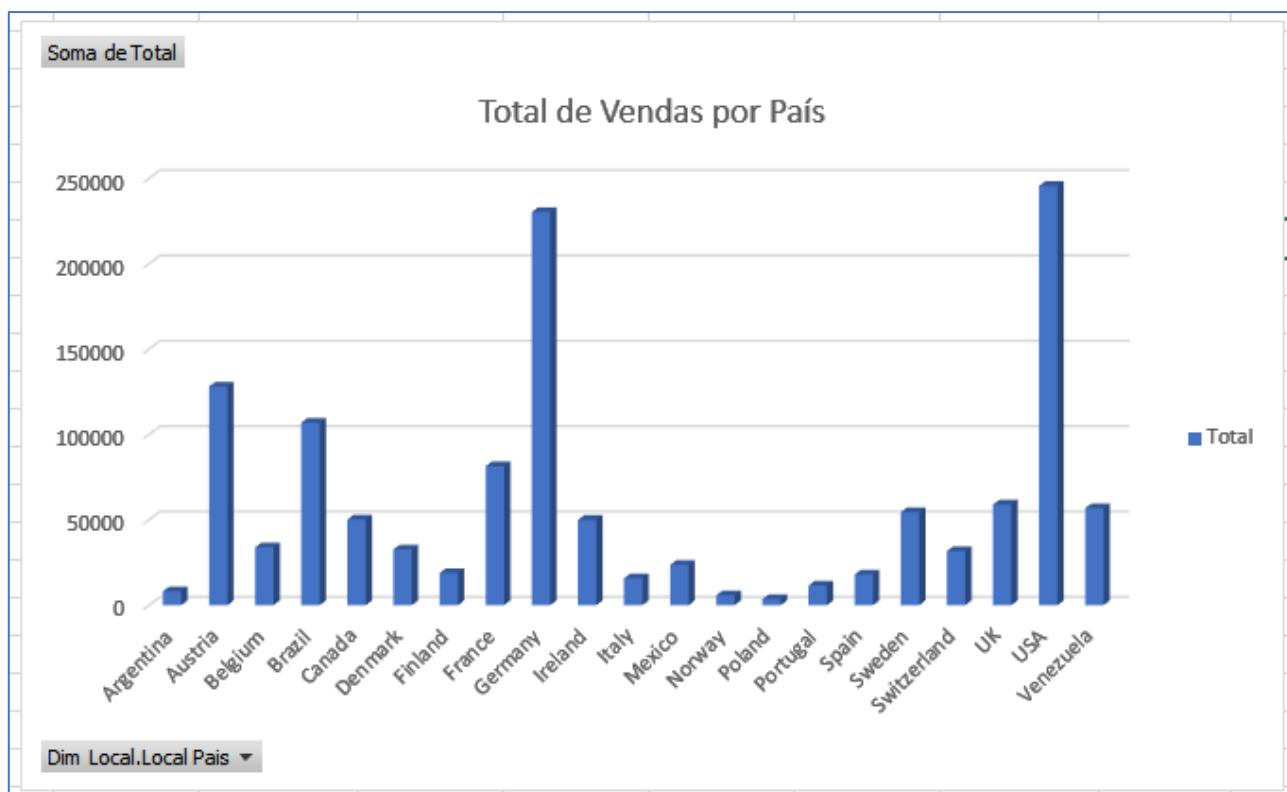


Figura 27 - Gráfico de Colunas das Vendas por País

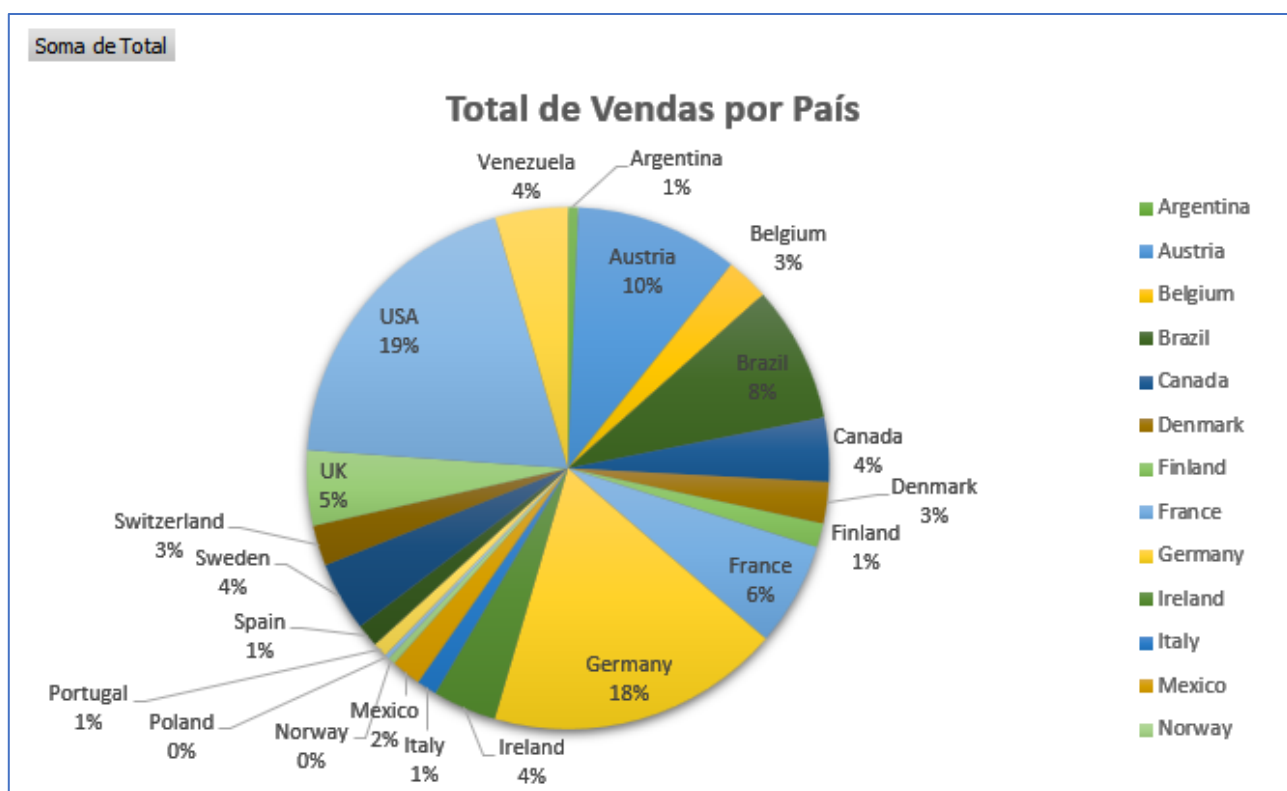


Figura 28 - Gráfico Circular de Vendas por País

Além da visualização da informação, o EXCEL tem a possibilidade de calcular cenários e projetar previsões. Desta forma, aproveitamos a ferramenta para fazer uma previsão do volume de vendas até 31/12/1996, com base nos totais de cada ano e cada mês, e respectiva sazonalidade dos valores.

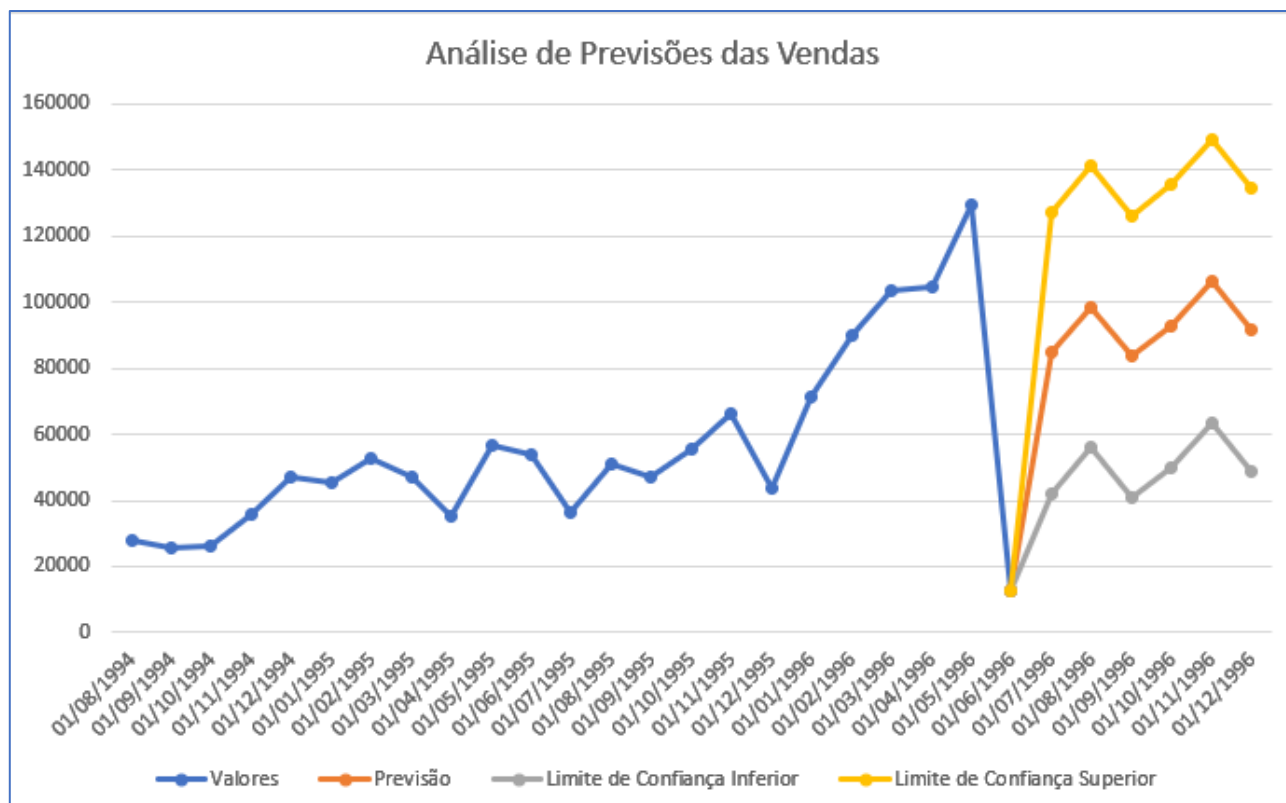


Figura 29 - Previsão das Vendas até 31/12/1996

O mesmo tipo de análise poderá rapidamente ser ajustado para mostrar previsões para apenas um País (ou outra dimensão qualquer), utilizando para isso os filtros que foram adicionados na tabela dinâmica. Na imagem seguinte podemos ver uma previsão para os USA.

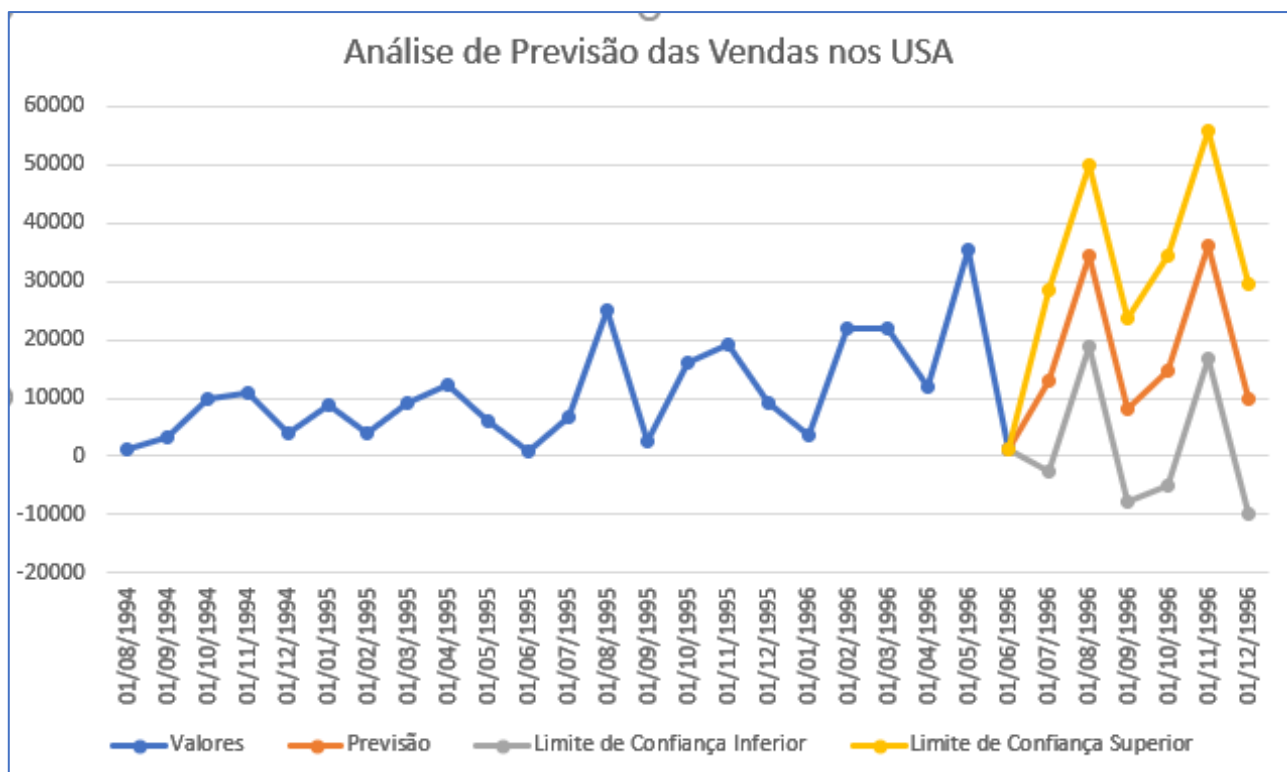


Figura 30 - Previsão das Vendas nos USA até 31-12-1996

4.5. Visualização dos Dados no *Power BI*

De uma forma resumida, pode definir-se o *Power BI* como um sistema de *Business Intelligence* que recolhe dados, que poderão estar mais ou menos estruturados e que poderão ser provenientes de uma ou várias fontes, e os transforma em informação útil à tomada de decisão, sob a forma de diversos tipos de visualizações, como gráficos, mapas, cartões, tabelas, matrizes, entre outros. (Nogueira, 2018)

A solução *PowerBI* está distribuída pelos seguintes componentes:

- ***Power BI Desktop*** – é a aplicação que funciona no computador. Tal como qualquer outra aplicação desktop, o trabalho é guardado localmente em ficheiros, neste caso com a extensão *.pbix*, que podemos armazenar no nosso computador ou enviar para outros destinatários.
- ***Power BI Service*** – funciona integralmente na nuvem, não sendo necessário transferir ficheiros para o computador. Tudo o que é preciso para começar a utilizar a aplicação é uma conta *Power BI* e fazer o respetivo login em <https://powerbi.microsoft.com>
- ***Power BI Mobile*** – permite a visualização dos elementos de visualização através dos dispositivos móveis, smartphone ou tablets. Pode ser obtida através das plataformas *Google Play*, *iTunes* ou *Microsoft Store*, conforme o sistema operativo do dispositivo.

4.5.1. Elementos de Visualização

Para este trabalho recorreu-se á aplicação *Power BI Desktop* para a construção dos elementos de visualização mais adequados ao processo de decisão numa empresa, dos quais se apresentam alguns exemplos:

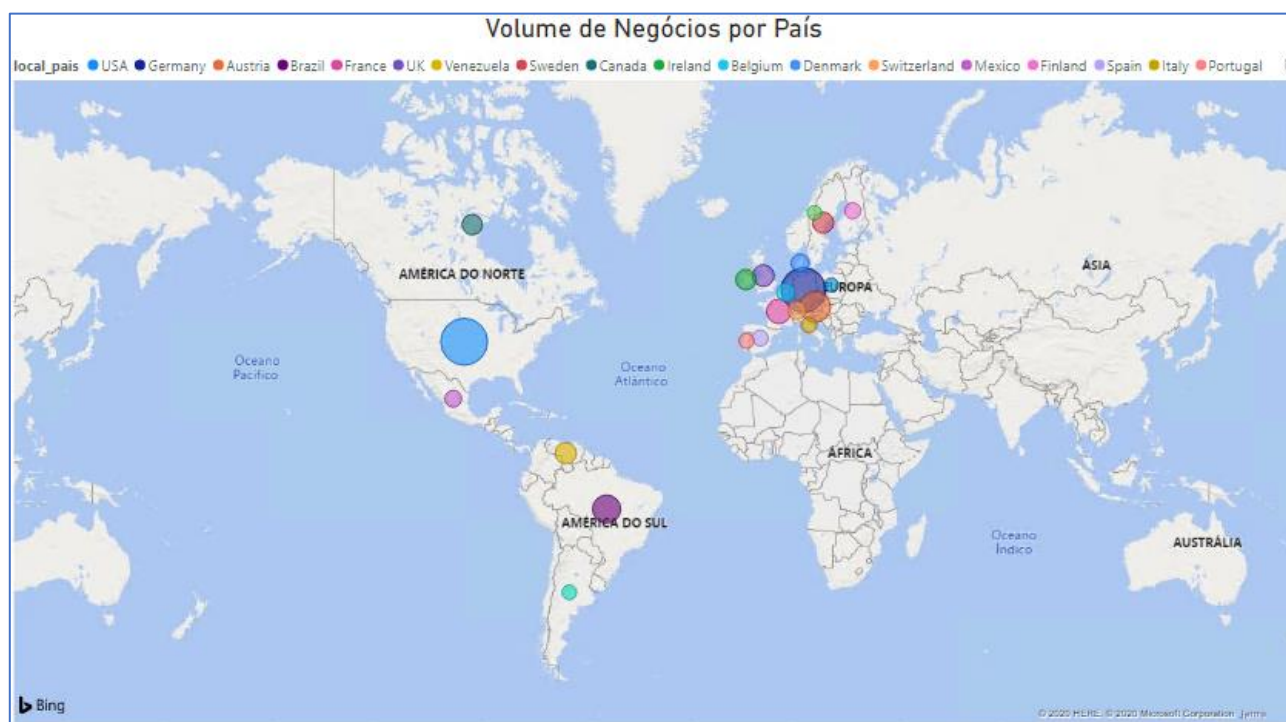


Figura 31 - Volume de Negócios por País

O mapa é um elemento de visualização é muito útil para ter uma perceção geográfica dos factos. Neste caso é possível perceber a dispersão geográfica do volume de negócios e quais os países para os quais a *Northwind Traders* mais vende. USA e Alemanha são, sem margem de dúvida, a origem dos melhores clientes.

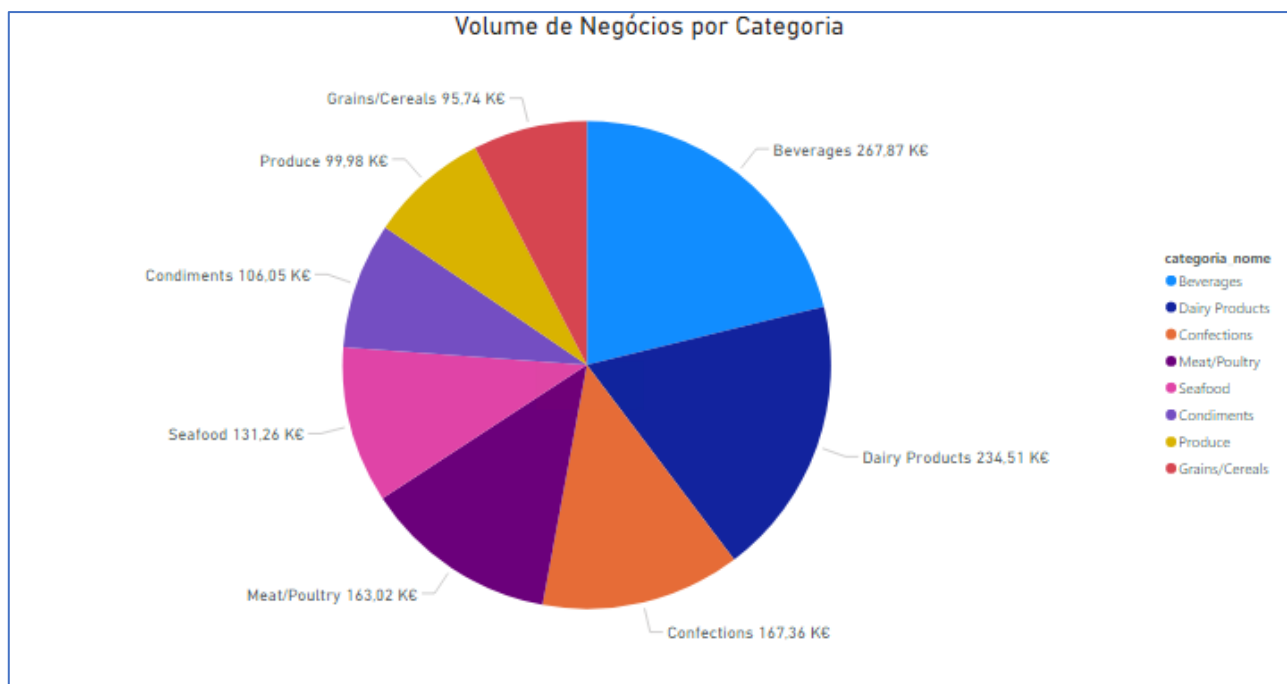


Figura 32 - Volume de Negócios por Categoria

Este gráfico circular permite rapidamente perceber quais as categorias de produtos que mais são vendidos. As “Beverages” e as “Dairy Products” são as categorias cujos produtos tem mais saída.

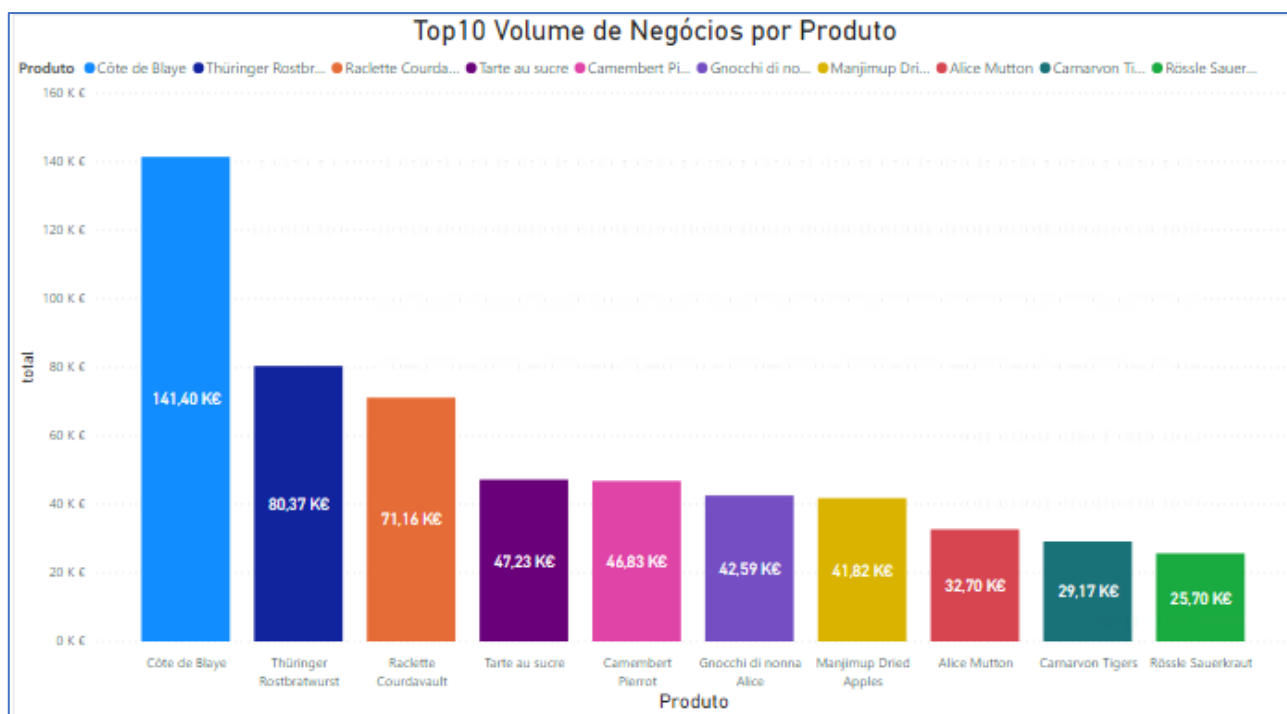


Figura 33 - Top 10 dos Produtos Mais Vendidos

Outra opção extremamente útil é a visualização dos dados com mais ocorrência. A possibilidade de ordenar e mostrar um Top10 dos produtos mais vendidos traduz-se numa mais valia, no que diz respeito a análise da informação. Seria praticamente impossível construir um gráfico de barras com todos os produtos que foram vendidos no período em análise. Neste gráfico também se percebe que o produto “*Côte de Blayde*” é o campeão das vendas, com 141,40K €, que se traduz quase no dobro do produto que se apresenta no segundo lugar.

Os gráficos anteriores apresentavam a informação agregada apenas por uma dimensão. Por local, categoria ou por produto. Apesar de se tratar de informação relevante, é fundamental adicionar uma camada de análise muito importante na gestão estratégica da procura de vantagem competitiva de qualquer organização: a dimensão tempo.

De facto, é com a análise da evolução dos indicadores e dos KPI ao longo do tempo (análise descritiva), que a organização poderá antecipar o futuro (análise preditiva) e desenhar assim as suas estratégias, baseada na melhor informação possível. Ou seja, produzir inteligência de negócio *Business Intelligence*.

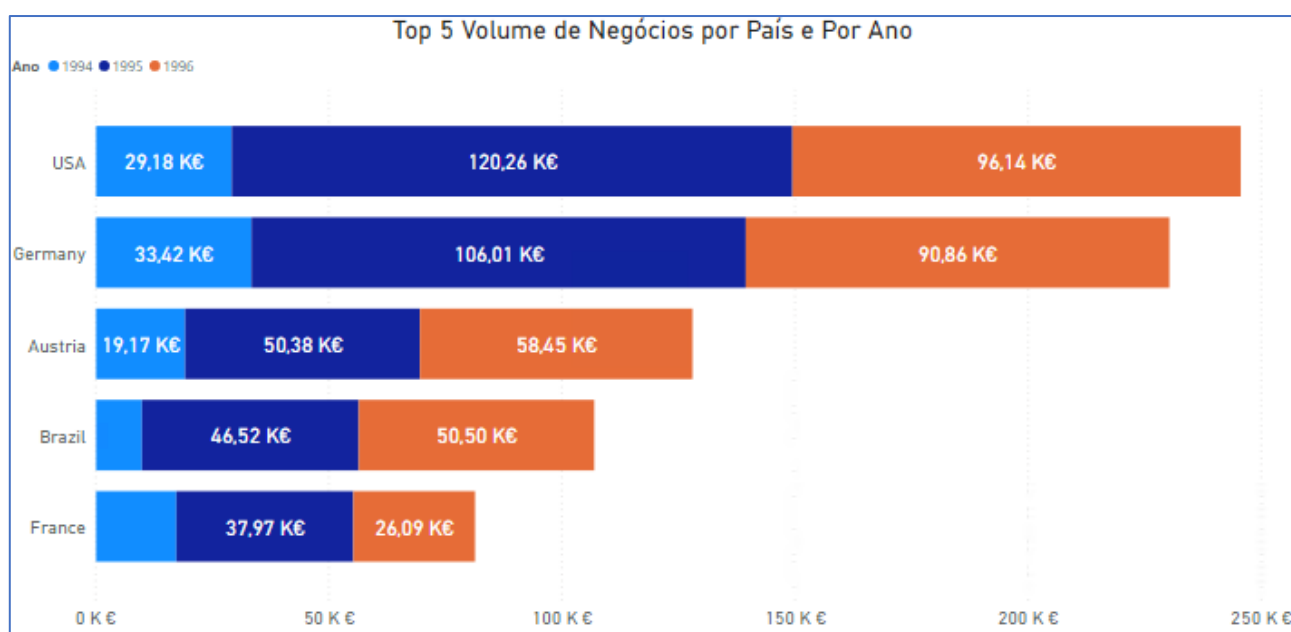


Figura 34 - Top5 Volume de Negócios por País e Por Ano

Neste exemplo, retrata-se um gráfico de barras empilhadas, relativa ao Top5 dos países com maior volume de negócios, e permite conjugar duas dimensões: Tempo com o Local. Efetivamente, com esta informação consegue-se perceber que, apesar dos USA serem o país que apresentam o maior peso no volume de negócios, no primeiro ano, foi a Alemanha que ocupou essa posição. Esta informação não seria perceptível se não tivéssemos adicionado a dimensão tempo neste elemento de visualização.

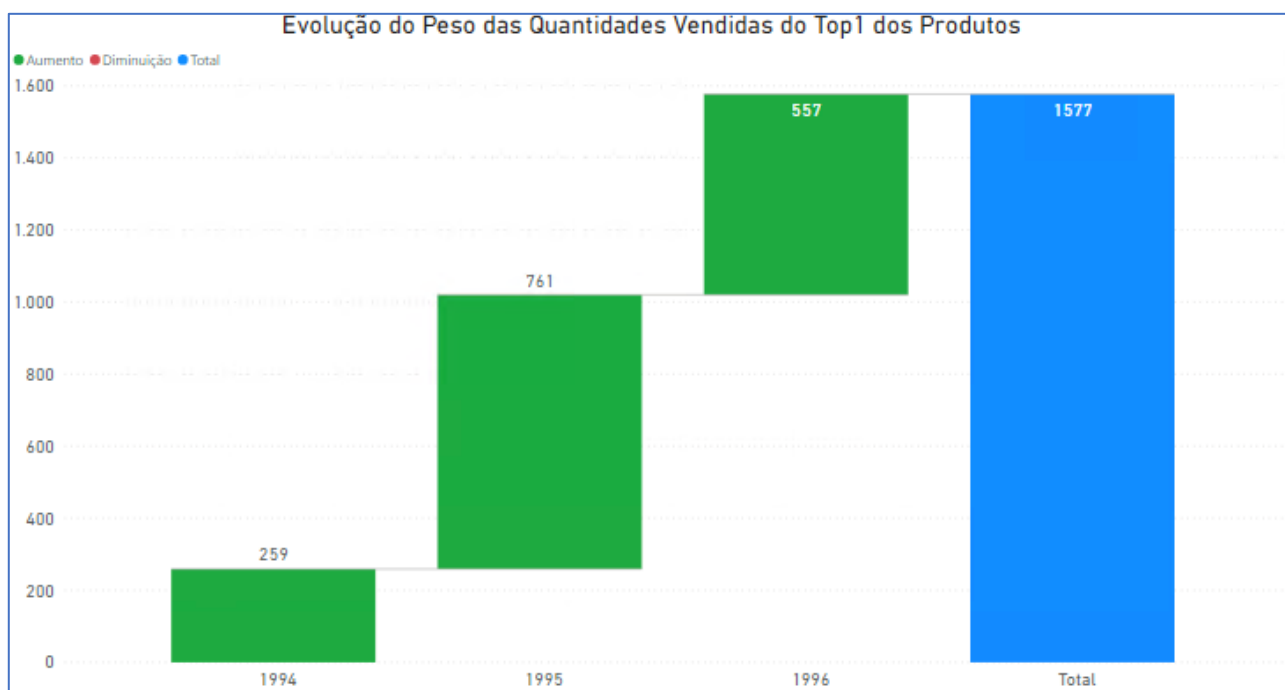


Figura 35 - Evolução do Peso das Quantidades Vendidas do Top1 dos Produtos

A análise da evolução do tempo das quantidades vendidas da “Cash Cow” (Wikipedia, 2020) de uma organização será um dos indicadores a acompanhar com mais interesse e cuidado.

Efetivamente o produto “Côte de Blaye” representa 25,30% de peso no Top10 dos produtos mais vendidos, e sendo assim este gráfico apresenta a evolução, em cada ano, do número de quantidades vendidas.

Percebe-se que foi o ano de 1995 que mais contribuiu para o total de 1577 unidades vendidas. Mais um exemplo da conjugação das dimensões para obter informação relevante.

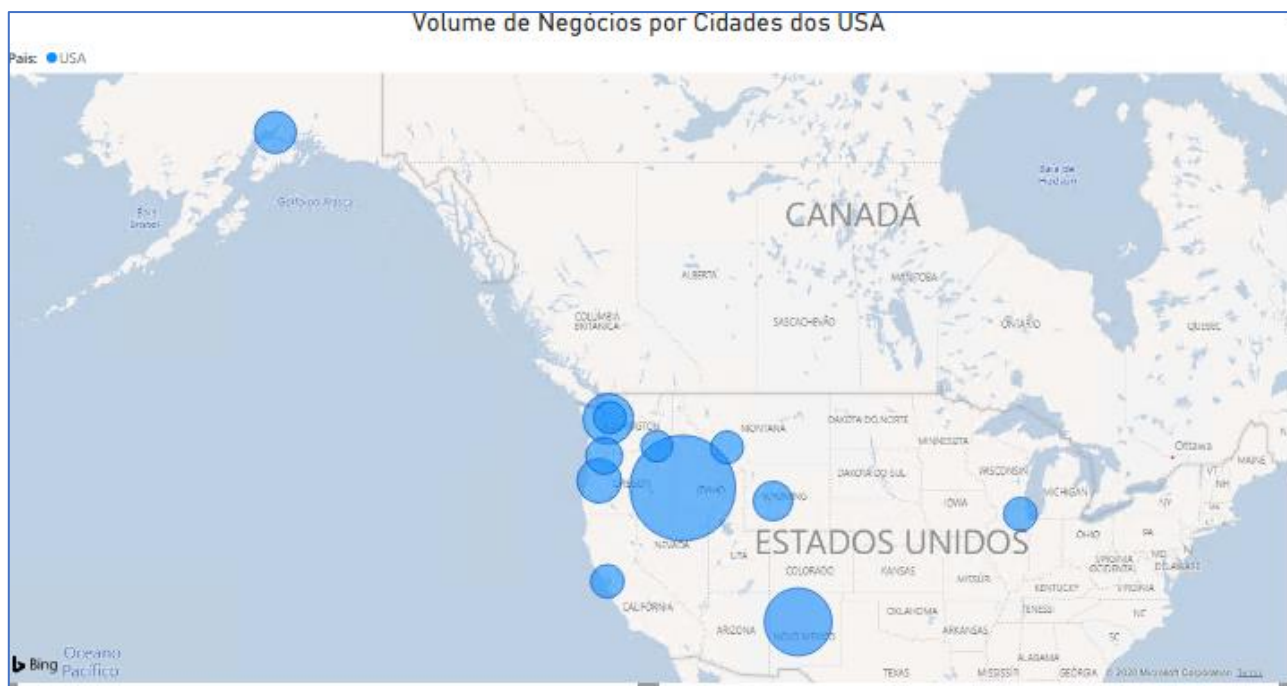


Figura 36 - Exemplo de Atributo da Dimensão Local

Neste segundo mapa de visualização, apresenta-se um exemplo da possibilidade de agrupar e filtrar a informação por atributo das dimensões. Efetivamente a dimensão Local do nosso *Datawarehouse* tinha uma dimensão Cidade, como um dos seus atributos. Neste elemento é possível então, mostrar as cidades com mais volume de vendas relativas ao País com mais vendas (USA).

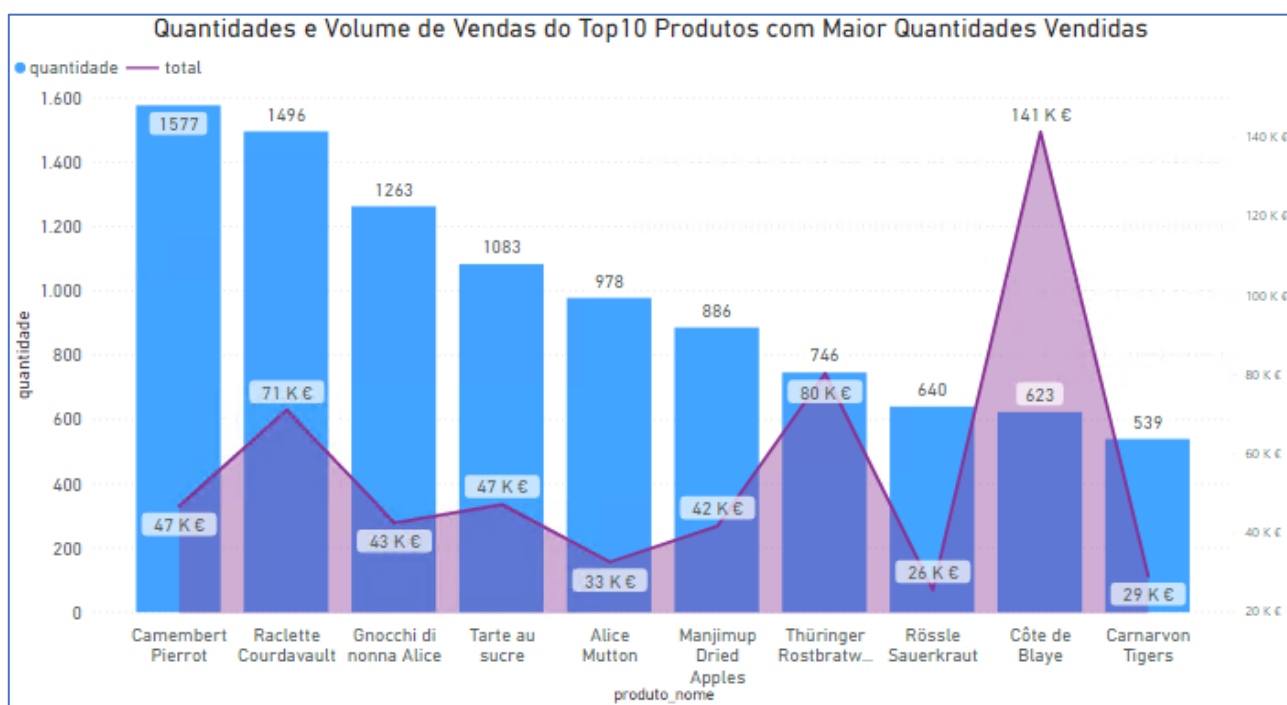


Figura 37 - Exemplo de Gráfico com Métricas Diferentes

Neste gráfico de colunas, existe outra funcionalidade que permite visualizar a informação de uma forma distinta dos elementos anteriores, ou seja, permite ter 2 métricas no mesmo eixo, com amplitudes diferentes. Neste caso em concreto agrupou-se a dimensão produto e observam-se as métricas volume de vendas (linha) e as quantidades vendidas (colunas). A leitura deste gráfico mostra a pertinência da informação, já que como se pode constatar, apesar do produto “*Côte de Blaye*” ser o que tem o maior volume de negócios, não é o que vende em maior quantidade. Por outro lado, o “*Camembert Pierrot*”, apesar de ter um volume de negócios discreto, é o que apresenta maior quantidade de unidades vendidas. Em termos de gestão comercial, uma alteração no preço unitário deste produto poderia originar um impacto considerável no volume de negócios. É uma informação estratégica fundamental.

4.5.2. *Dashboards Online*

Após a construção dos diferentes elementos visuais, foi construída uma ferramenta de análise da informação que permite de uma forma acessível, fácil e intuitiva a qualquer utilizador, independentemente dos seus conhecimentos analíticos, consumir os dados, agregando e filtrando, conforme as suas necessidades específicas. Desta forma, foi construído um *Dashboard* e disponibilizado no *PowerBi Cloud*. Esta infraestrutura na nuvem permite o acesso através de um browser ou de dispositivos móveis, melhorando assim a acessibilidade da informação.

Sendo assim, foram compilados alguns elementos que consideramos essenciais a fazer parte do *dashboard*, bem como os filtros necessários para as diversas interações dos utilizadores.

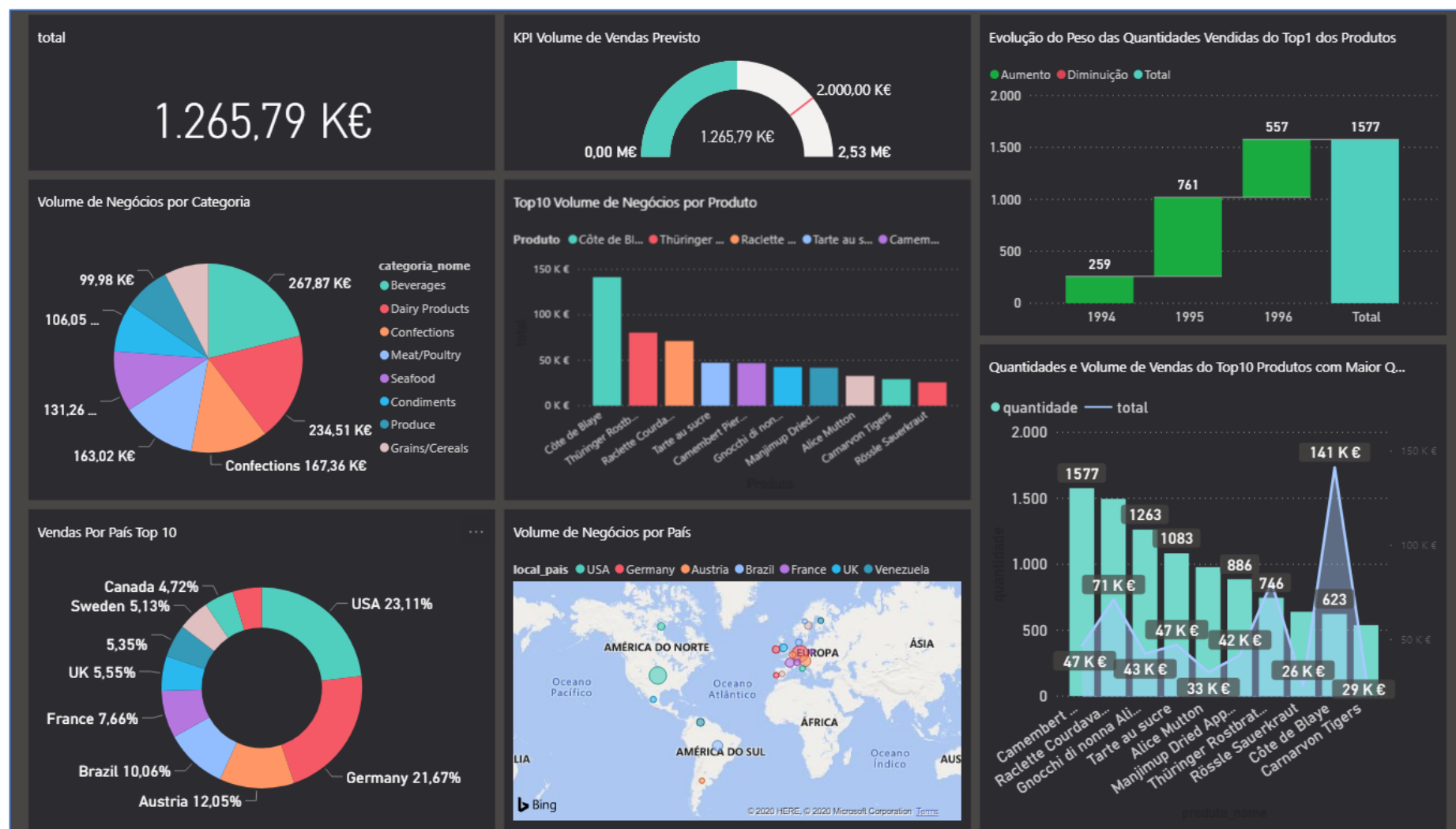


Figura 38 - Dashboard Vista Browser

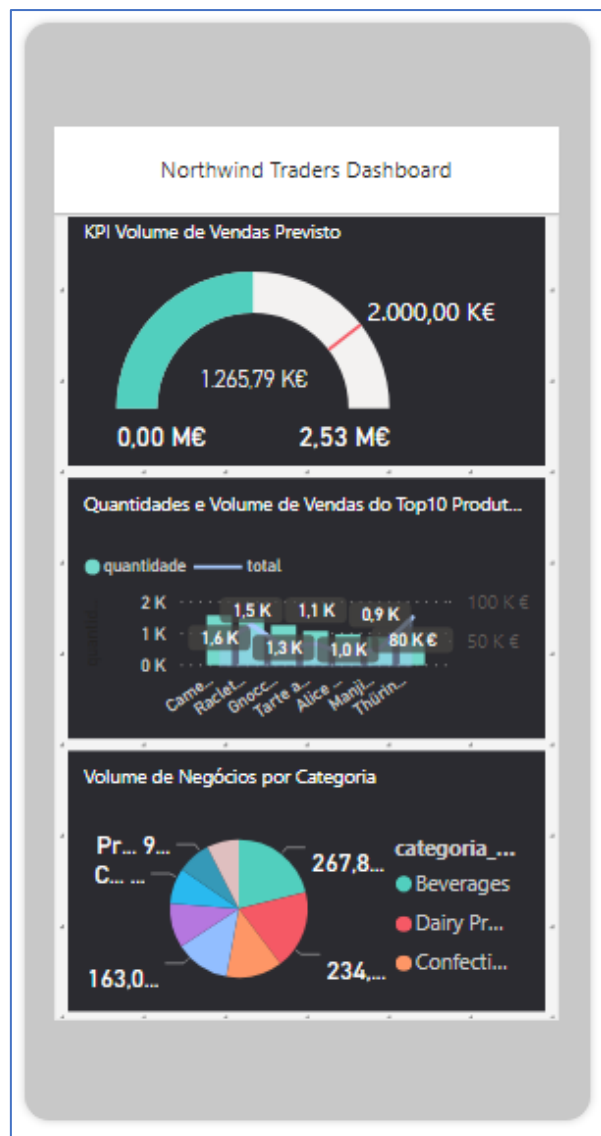


Figura 39 - Dashboard Vista Mobile

Conclusão

Sendo a informação o bem mais valioso dos tempos atuais este trabalho permitiu perceber o quão importante é a sua qualidade, atualidade, precisão, integridade, relevância, isenção de erros e consistência.

O tratamento dos dados que envolvem o negócio de qualquer organização é já entendido por muitos como uma mais valia e mesmo como um fator distintivo da capacidade de adaptação e flexibilidade das organizações ao mercado.

O uso de ferramentas informáticas de apoio ao tratamento e análise dos dados faz com que as organizações sejam mais ágeis nas tomadas de decisão, sejam estas estratégicas, táticas ou operacionais. O uso eficiente destas ferramentas faz com que os decisores de qualquer organização possam rapidamente analisar os dados atualizados por forma a fundamentar as decisões a tomar.

Todavia é igualmente importante o envolvimento de toda a organização na implementação destas ferramentas para que os seus *outputs* sejam os esperados e os mais pertinentes para o negócio. É essencial que a gestão de topo retire dos dados a melhor informação possível para suportar a viabilidade do negócio ou mesmo mudanças do paradigma da organização.

Constatou-se durante a realização deste trabalho que o uso de ferramentas de *Power BI* proporciona a visualização da informação (*dashboards*, gráficos, relatórios, entre outros) de uma forma mais interativa e detalhada, permitindo ao utilizador a interpretação quase intuitiva dos dados.

A complexidade do desenvolvimento de um *Datawarehouse* e as análises realizadas através das diferentes ferramentas permitiu aos elementos do grupo compreender melhor o paradigma dados vs informação e principalmente a capacidade que estes aplicativos têm para processar quantidades infinitas de dados.

A importância de um bom diagnóstico e uma boa análise permite tomar decisões mais alinhadas com as necessidades das organizações e do mercado.

As estratégias utilizadas para a realização do trabalho enquanto grupo como o *Jira*, o quadro *Kanban* e os recursos utilizados dotaram-nos de competências como a organização e gestão de tempo que vão muito além do esperado para este trabalho.

Fica presente em nós a importância do *Business Intelligence* e do *Data Mining* e da premência de nos mantermos atualizados neste tema de relevância presente e futura.

Agradecemos à Professora Célia Gonçalves o desafio que nos lançou.

Bibliografia

- Caldas, M. P., Sandelari, L., & Kovalski, J. L. (2006). Aplicações sobre uma Data Warehouse no ambiente das organizações e suas vantagens. *XIII SIMPEP*, (pp. 1-8). Bauru, SP, Brasil.
- Descrição geral dos cubos OLAP do Service Manager para análise avançada*. (06 de Maio de 2019). Obtido de <https://docs.microsoft.com/pt-pt/system-center/scsm/olap-cubes-overview?view=sc-sm-2019>
- Ferrari, A., Russo, M., & Webb, C. (2012). *Microsoft SQL Server 2012 Analysis Services: The BISM Tabular Model*. Microsoft.
- Ferreira, J., Miranda, M., Abelha, A., & Machado, J. (Setembro de 2010). O Processo ETL em Sistemas Data Warehouse. *INForum 2010 - II Simpósio de Informática*, pp. 757–765.
- GARDNER, S. R. (Setembro de 1998). Building the Data Warehouse. *COMMUNICATIONS OF THE ACM*, pp. 52-60.
- Inmon, W. (1997). *Managing the Data Warehouse*. John Wiley & Sons, Inc.
- Kimball, R., Reeves, L., Ross, M., & Thornthwaite, W. (1998). *The Data Warehouse lifecycle toolkit. Developing, and Deploying Data Warehouses*. Wiley.
- Nogueira, N. (2018). *Power BI para Gestão e Finanças*. FCA - Editora de Informática, Lda.
- Santos, M. Y., & Ramos, I. (2006). *Business Intelligence: tecnologias da informação na gestão de conhecimento*. FCA - Editora de Informática, Lda.
- SAS. (Abril de 2020). *ETL*. Obtido de SAS: https://www.sas.com/pt_br/insights/data-management/o-que-e-etl.html
- Vassiliadis, P. (Junho de 2000). Data Warehouse Modeling and Quality Issues. Atenas, Grécia: NATIONAL TECHNICAL UNIVERSITY OF ATHENS.
- Wikipedia. (18 de Abril de 2020). *Matriz BCG*. Obtido de Wikipedia: https://pt.wikipedia.org/wiki/Matriz_BCG