

Trabalho Teórico-Prático de Avaliação

Sistemas de Informação II



Luís Henrique P. O. Travassos, nº2021136600

Rodrigo Ferreira Ramalho, nº2021139149

Conteúdo

1. Resumo	1
2. Introdução	2
3. Contextualização	3
4. Estado de arte genérica sobre datawarehouse	4
5. Exploração da ferramenta de BI Grafana	5
5.1. Introdução à ferramenta Grafana	5
5.2. Instalação da ferramenta Grafana	6
5.3. Apresentação da BD criada e usada	7
5.4. Ligação do SQL Server ao Grafana	9
5.5. Análise do fluxo de trabalho da ferramenta Grafana	11
5.6. Criação de Dashboards em Grafana	14
5.6.1. Dashboard de gêneros de filmes mais populares	14
5.6.2. Dashboard da receita total por cinema	15
5.6.3. Dashboard da idade média dos clientes por cinema	16
5.6.4. Dashboard de filmes em exibição por cinema	17
5.6.5. Dashboard da quantidade de vendas por funcionários	18
5.7. Comparações com a ferramenta BI Power BI	19
6. Criação de uma datawarehouse	21
6.1. Processo de negócio a modelar	22
6.2. Avaliação da Granularidade	23
6.3. Selecionar dimensões e atributos	24
6.4. Selecionar factos	25
6.5. Modelo em estrela criado	26
6.6. Cálculos do tamanho aproximado do Modelo	27
6.7. Tratamento do modelo em estrela no SQL Server	29
6.8. Vistas no Visual Studio Data Warehouse	29
6.8.1. Vista de Performance de Vendas por Loja	30
6.8.2. Vista de Tendências de Vendas Temporais	31
6.8.3. Vista de Eficiência de Promoções por Região	31
6.8.4. Vista de Análise de Desempenho de Categorias de Produtos	32
6.8.5. Vista de Performance de Vendas por Loja no mês de setembro	32
6.9. Dashboard no Power BI	33
6.9.1. Dashboard - Quantidade Vendida por Loja	33
6.9.2. Dashboard - Quantidade Vendida por Ano	34
6.9.3. Dashboard - Desconto por Produto e Cidade	34
6.9.4. Dashboard – Quantidade Vendida por Loja e Ano	35
6.9.5. Dashboard – Soma do Valor Total Faturado em Vendas	35
7. Conclusões	36
8. Referências	37

1. Resumo

Este relatório oferece uma análise detalhada da ferramenta de Business Intelligence (BI) Grafana, colocando-a em contraste com o Power BI e explorando suas aplicações em um contexto de data warehouse. Inicialmente, o documento fornece uma introdução abrangente ao tema e contextualiza a importância de ferramentas de BI no ambiente empresarial moderno. Segue-se uma discussão sobre o estado da arte em data warehouses, estabelecendo uma base teórica para a exploração subsequente do Grafana.

Nos capítulos dedicados ao Grafana, o relatório aborda desde a introdução à ferramenta até a instalação e a configuração da mesma. Uma base de dados (BD) específica é apresentada, usada para ilustrar a integração do Grafana com o SQL Server. O fluxo de trabalho da ferramenta é analisado, com ênfase na criação de dashboards e cinco dashboards são detalhadamente descritas.

Um capítulo crucial do relatório é dedicado às comparações com o Power BI, destacando as vantagens do Grafana como sendo open-source, mas também notando suas limitações, como a dificuldade de uso em comparação com o Power BI e a necessidade de um bom conhecimento de SQL, além de opções mais limitadas de gráficos e detalhamento.

As conclusões do relatório sintetizam as descobertas, salientando as situações em que o Grafana pode ser preferível ao Power BI, e vice-versa, dependendo das necessidades específicas de BI e do nível de habilidade técnica dos usuários. As referências utilizadas para compilar o relatório são listadas no final, fornecendo um recurso abrangente para pesquisa e análise adicional.

2. Introdução

No atual cenário de avanços tecnológicos e digitalização empresarial, as ferramentas de Business Intelligence (BI) emergem como componentes essenciais para apoiar a tomada de decisões e otimizar a gestão de dados. Este relatório foca na exploração da ferramenta Grafana, uma plataforma de BI alternativa ao amplamente utilizado Power BI. A relevância do Grafana em ambientes empresariais que requerem monitoramento em tempo real e dashboards personalizáveis é destacada na literatura de BI, como evidenciado por autores como Chen, Chiang e Storey (2012) em seu trabalho sobre a importância crescente de BI nas empresas.

A escolha do Grafana como objeto de estudo é motivada pela sua popularidade e funcionalidades únicas, aspectos sublinhados por pesquisas recentes sobre ferramentas de BI e análise de dados (Khan, 2017). Este estudo visa ampliar o espectro de conhecimento em ferramentas de BI, explorando soluções que podem ser mais adequadas para contextos empresariais específicos, um ponto de vista apoiado por Davenport (2013) em sua análise sobre a adaptação de ferramentas de BI às necessidades empresariais.

Neste relatório, a abordagem teórica é complementada por análises práticas, visando fornecer uma visão abrangente sobre a ferramenta Grafana. Isso inclui uma avaliação de suas funcionalidades, vantagens e desvantagens, bem como possíveis cenários de uso, conforme recomendado por estudos no campo de BI (Wixom, Yen, e Relich, 2010). O objetivo é fornecer não apenas um entendimento aprofundado do Grafana, mas também insights valiosos sobre a escolha e implementação de ferramentas de BI em organizações modernas, uma necessidade destacada pela pesquisa de Hopkins (2011) sobre tendências emergentes em BI.

3. Contextualização

Nos últimos anos, o mundo corporativo tem passado por uma transformação digital sem precedentes, uma mudança marcada pelo crescente volume de dados e pela necessidade de ferramentas capazes de transformar esta vasta quantidade de informação em insights acionáveis. Como apontado por Brynjolfsson e McAfee (2014) em sua análise sobre a era digital, as ferramentas de Business Intelligence (BI) ganharam destaque, servindo como uma ponte crucial entre dados brutos e decisões estratégicas nas organizações.

Conforme descrito por Davenport e Patil (2012), o BI envolve a coleta, integração, análise e apresentação de informações de negócios. Empresas de todos os tamanhos utilizam ferramentas de BI, conforme destacado por Mayer-Schönberger e Cukier (2013), para identificar tendências, detectar padrões de negócios e gerar insights baseados em dados. Essa capacidade de transformar dados em conhecimento aplicado é crucial para responder rapidamente às mudanças do mercado, antecipar tendências e melhorar a tomada de decisão, como observado por Brown, Chui e Manyika (2011).

A escolha da ferramenta de BI correta é um fator crítico para o sucesso desses esforços. Tradicionalmente, ferramentas como o Power BI da Microsoft dominam o mercado, como notado por Howson e Idoine (2017), mas com a crescente complexidade dos cenários de negócios e a necessidade de soluções mais personalizadas, outras ferramentas, como o Grafana, começaram a ganhar espaço. O Grafana, originalmente projetado para o monitoramento de métricas e dados em tempo real, destaca-se por sua flexibilidade e interface de usuário intuitiva, um ponto enfatizado por Turner, Gantz e Reinsel (2014) ao discutir as tendências em análise de dados.

A adoção de ferramentas como o Grafana reflete uma tendência mais ampla no campo de BI, uma busca por soluções que se adaptem às necessidades específicas e à infraestrutura de cada organização, conforme observado por Bughin, Chui e Manyika (2010). Este relatório busca entender como o Grafana se encaixa nesse cenário em evolução e qual seu potencial para atender às demandas contemporâneas de BI nas organizações.

4. Estado de arte genérica sobre datawarehouse

O conceito de Data Warehouse (DW) tem sido fundamental na evolução de como as organizações armazenam, acedem e analisam dados. Conforme Inmon (2005), conhecido como o 'pai da data warehouse', o DW surgiu nas últimas décadas do século XX como resposta aos desafios impostos pelo aumento exponencial do volume de dados e pela necessidade de sistemas que pudessem armazenar informações de forma consolidada e estruturada.

Um Data Warehouse é um repositório centralizado de dados, integrando informações de diversas fontes, como descrito por Kimball e Ross (2011) em suas abordagens sobre modelagem dimensional. Esses dados são processados e transformados através de um processo conhecido como ETL (Extração, Transformação e Carregamento), uma metodologia destacada por Vassiliadis et al. (2002).

Com o avanço da tecnologia e os métodos analíticos, os DWs tornaram-se ainda mais cruciais nas estratégias de BI das empresas, uma observação compartilhada por Davenport (2006) em sua análise sobre analytics e BI. Os DWs modernos são projetados para serem altamente escaláveis, adaptando-se ao crescimento contínuo do volume de dados, um aspecto ressaltado por Devlin (2009) ao explorar as arquiteturas de DW.

A introdução de tecnologias como data mining, machine learning e analytics avançada, discutida por Han, Pei e Kamber (2011), ampliou as capacidades dos DWs, permitindo análises preditivas e a identificação de padrões ocultos nos dados. Essa capacidade preditiva é crucial para o planejamento estratégico das organizações.

Adicionalmente, a era da cloud computing, como explorado por Marz e Warren (2015), trouxe um novo paradigma para o DW. As soluções de DW baseadas em nuvem, oferecidas por players como Amazon Web Services, Google Cloud Platform e Microsoft Azure, trazem vantagens em termos de escalabilidade, flexibilidade e custo, tornando o armazenamento e processamento de dados acessível a empresas de todos os tamanhos.

5. Exploração da ferramenta de BI Grafana

No capítulo seguinte será apresentada a exploração da ferramenta BI Grafana e uma análise/comparação da sua utilidade quando comparada com o a ferramenta BI PowerBI.

5.1. Introdução à ferramenta Grafana

Grafana, uma plataforma de visualização e análise de dados em código aberto, tem emergido como uma ferramenta de Business Intelligence (BI) notável, especialmente no contexto de monitoramento de dados em tempo real e visualização interativa. Originada em 2014 e desenvolvida pela Grafana Labs, esta ferramenta rapidamente ganhou popularidade devido à sua flexibilidade, capacidade de integração e interface de utilizador intuitiva.

Grafana distingue-se no universo de BI por sua abordagem focada na visualização de dados e monitoramento de performance. A ferramenta permite aos utilizadores criar dashboards ricos e interativos, que exibem dados de forma clara e acessível, facilitando a compreensão de tendências complexas e a realização de análises detalhadas. O que torna Grafana particularmente atraente é a sua capacidade de integrar-se com uma vasta gama de fontes de dados, como bancos de dados SQL, sistemas de monitoramento de infraestrutura e serviços em nuvem.

Um dos principais aspetos do Grafana é a sua natureza em código aberto, o que significa que é continuamente aprimorada por uma comunidade ativa de desenvolvedores. Essa característica assegura que a plataforma esteja sempre evoluindo, com novas funcionalidades e integrações sendo adicionadas regularmente. Além disso, a capacidade de personalizar e estender o Grafana através de plugins e APIs oferece uma flexibilidade sem precedentes, permitindo que os usuários moldem a ferramenta de acordo com suas necessidades específicas.

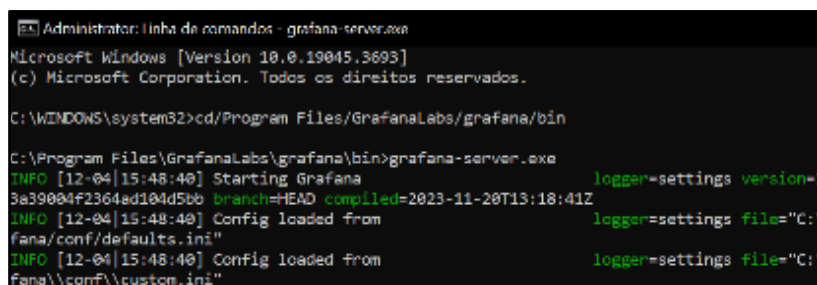
Em termos de usabilidade, o Grafana é conhecido por sua interface limpa e amigável, que simplifica a criação e o compartilhamento de dashboards. A ferramenta oferece uma variedade de opções de visualização, desde gráficos de linha e barra até mapas de calor e geomapas, possibilitando a representação de dados de maneiras variadas e informativas.

Na prática, o Grafana é amplamente utilizado em cenários de monitoramento de TI, como acompanhamento de servidores, redes e aplicações web. No entanto, seu uso não se limita a estes casos. Cada vez mais, empresas de diferentes setores estão adotando o Grafana para monitorar indicadores de desempenho chave (KPIs), analisar métricas de vendas, e até mesmo para monitorar dados de sensores em ambientes de Internet das Coisas (IoT).

5.2. Instalação da ferramenta Grafana

A instalação do Grafana em sistemas operacionais Windows é um processo direto, permitindo aos utilizadores começarem rapidamente a explorar as suas funcionalidades de visualização e análise de dados. Existem duas maneiras principais de instalar o Grafana no Windows: usando o instalador do Windows ou através do arquivo binário autônomo do Windows. No nosso caso usamos o instalador, logo iremos focar neste:

1. Escolha da Versão e Edição do Grafana:
 - a. Aceda a página de download do Grafana.
 - b. Selecione a versão do Grafana que deseja instalar. Por padrão, a versão mais recente é selecionada.
 - c. Escolha entre as edições “Enterprise”, que é a versão recomendada com funcionalidades adicionais desbloqueáveis por licença, e “Open Source”, que é idêntica em funcionalidade, mas sem as características empresariais. Nós escolhemos a versão “Open Source”
2. Instalação via Windows Installer:
 - a. Clique em “Download the installer” para transferir o instalador.
 - b. Execute o instalador abrindo o arquivo transferido.
3. Inicialização do Grafana:
 - a. Execute o grafana-server.exe, localizado no diretório bin, preferencialmente através da linha de comando. Na figura 1 mostramos a ligação do servidor Grafana no CMD:



```
Administrator: Linha de comandos - grafana-server.exe
Microsoft Windows [Version 10.0.19045.3693]
(c) Microsoft Corporation. Todos os direitos reservados.

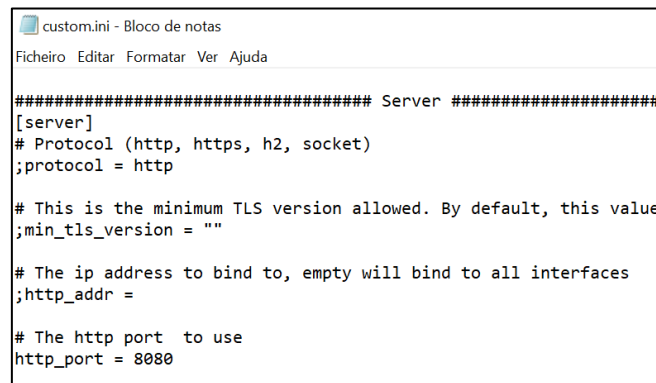
C:\WINDOWS\system32>cd/Program Files/GrafanaLabs/grafana/bin

C:\Program Files\GrafanaLabs\grafana\bin>grafana-server.exe
INFO [12-04|15:48:40] Starting Grafana              logger=settings version=1
3a39084f2364ad104d5bb branch=HEAD compiled=2023-11-20T13:18:41Z
INFO [12-04|15:48:40] Config loaded from              logger=settings file="C:\
Grafana\conf\defaults.ini"
INFO [12-04|15:48:40] Config loaded from              logger=settings file="C:\
Grafana\conf\custom.ini"
```

Figura 1 - Ligação servidor Grafana no CMD

4. Aceder ao Grafana:
 - a. Para iniciar o Grafana, abra o navegador e vá para o endereço <http://localhost:3000/>, que é o padrão. No nosso caso não funcionou por falta de permissões no Windows.
 - b. Se necessário, altere a porta padrão do Grafana. Por padrão, ele usa a porta 3000, que pode exigir permissões adicionais no Windows.
5. Mudança da Porta Padrão:
 - a. No diretório conf, copie “sample.ini” para “custom.ini”. Lembre-se de sempre editar o “custom.ini”, nunca o “defaults.ini”.
 - b. No arquivo “custom.ini”, descomente a opção de configuração “http_port”.

- c. Altere a porta para 8080 ou outra de sua escolha. A porta 8080 geralmente não requer privilégios extras no Windows. Na figura 2 mostramos o ficheiro “custom.ini” alterado de forma a configurar o http_port:



```
custom.ini - Bloco de notas
Ficheiro  Editar  Formatar  Ver  Ajuda

##### Server #####
[server]
# Protocol (http, https, h2, socket)
;protocol = http

# This is the minimum TLS version allowed. By default, this value
;min_tls_version = ""

# The ip address to bind to, empty will bind to all interfaces
;http_addr =

# The http port to use
http_port = 8080
```

Figura 2 - custom.ini configurado

Seguindo estes passos, a instalação do Grafana em uma máquina Windows é simplificada, permitindo que o utilizador rapidamente comece a criar dashboards e a visualizar dados de maneira eficaz.

5.3. Apresentação da BD criada e usada

De forma a termos um problema concreto com o que trabalhar pedimos um enunciado de um problema ao ChatGPT, com base nos problemas das aulas práticas da cadeira de Sistemas de Informação II, sendo este:

Considere uma rede de cinemas que opera em diversas cidades. Esta rede deseja melhorar a gestão das suas operações e a análise de dados sobre as preferências dos espectadores, a performance dos diferentes filmes, e o desempenho das vendas de bilhetes e concessões. Para tal, é necessário desenvolver uma base de dados em formato estrela que permita uma análise multidimensional desses aspetos.

Cada cinema na rede possui várias salas, e cada sala exhibe diferentes filmes ao longo do dia. Os espectadores compram bilhetes para sessões específicas e, muitas vezes, adquirem também produtos nas concessões, como pipoca e bebidas. A base de dados deve ser capaz de armazenar informações detalhadas sobre cada venda, incluindo o filme assistido, a sala, a data e hora da sessão, os itens de concessão comprados, e o total gasto pelo cliente.

Além disso, a rede de cinemas realiza frequentemente promoções, como descontos em dias específicos da semana, preços especiais para lançamentos de filmes, ou ofertas nas concessões. A base de dados deve ser capaz de registar essas promoções e associá-las às vendas correspondentes para análises futuras de eficácia de marketing.

Os dados demográficos dos clientes, como idade e sexo, também são coletados (respeitando as normativas de privacidade) no momento da compra de bilhetes, oferecendo uma oportunidade para análises demográficas detalhadas e personalização de ofertas.

O objetivo é estruturar a base de dados de modo que a equipe de análise possa facilmente extrair insights sobre as preferências dos espectadores, a popularidade de diferentes filmes e horários, o desempenho de vendas de itens de concessão, e a eficácia das estratégias de marketing. As informações devem ser organizadas para facilitar consultas rápidas e eficientes, permitindo que a rede de cinemas tome decisões informadas para otimizar as suas operações e estratégias de marketing.

De forma a responder a este enunciado criamos um modelo em estrela, figura 3, com o seguinte formato:

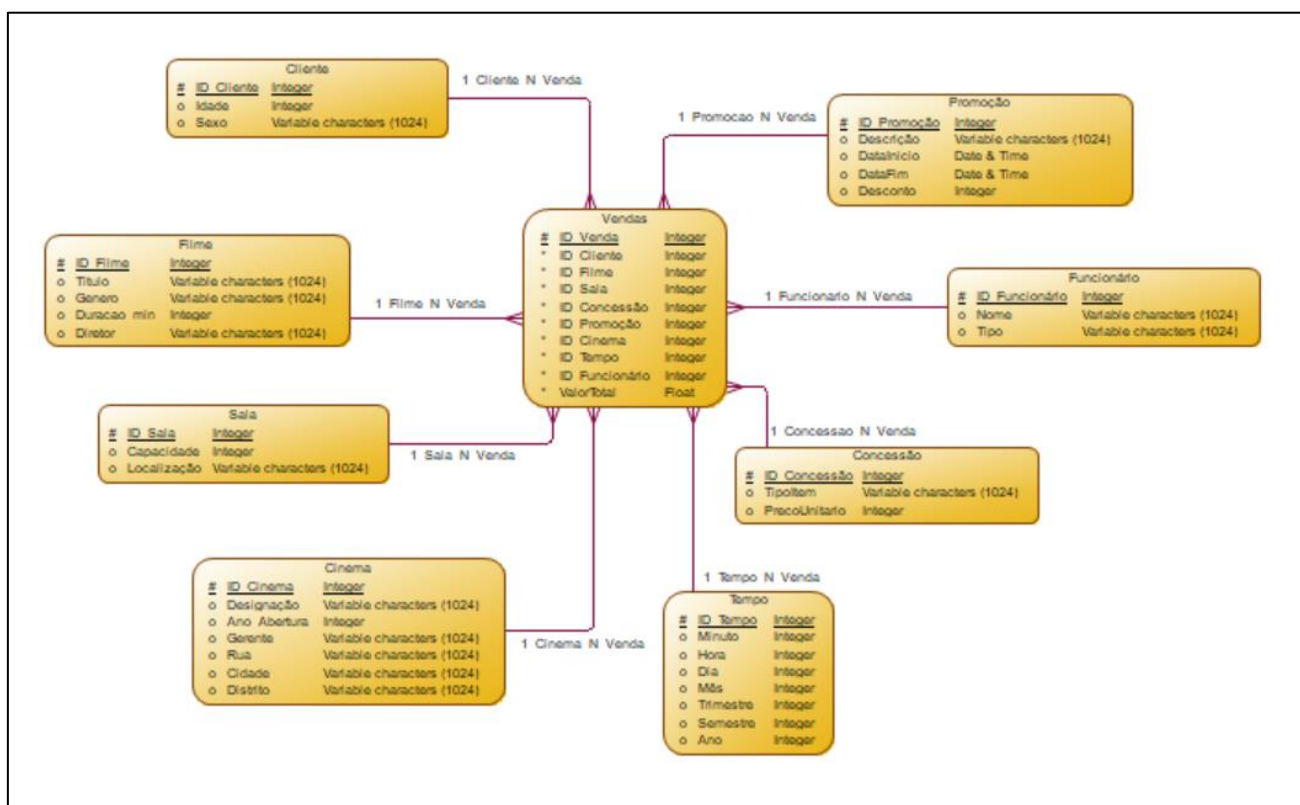


Figura 3 - Modelo de BD em estrela

Após a criação desta no Sql Server Management Studio povoamo-la com 200 entradas em cada entidade, todos os códigos SQL destas podem ser encontrados na pasta “codigosSQL”.

5.4. *Ligação do SQL Server ao Grafana*

A integração do Grafana com um banco de dados SQL Server é um processo simples e direto, permitindo a visualização e análise de dados armazenados no SQL Server através do Grafana. Siga estes passos para conectar a sua fonte de dados SQL Server ao Grafana:

- Acesso ao Grafana:
 - Abra o seu navegador web e navegue até a instância do Grafana, normalmente localizada em <http://localhost:3000>, a menos que você tenha configurado uma porta diferente.
- Login no Grafana:
 - Faça login no Grafana usando suas credenciais.
- Adicionar Fonte de Dados:
 - Clique no ícone da Casa no menu lateral esquerdo para abrir o menu de Conexões. Em seguida, clique em “Add new connection”.
 - Procure por "SQL Server" entre os plugins de fonte de dados disponíveis.
 - Clique em "Microsoft SQL Server" para selecioná-lo.
- Configurar a Conexão com o SQL Server:
 - Forneça um nome para a sua fonte de dados.
 - No campo "Host", insira o nome do host ou o endereço IP do seu SQL Server.
 - Insira o nome do banco de dados ao qual deseja se conectar.
 - Forneça as credenciais (nome de usuário e senha) para acessar o banco de dados.
 - Configurações opcionais como "Modo SSL" e "Nome da Instância" podem ser ajustadas, se necessário.
- Testar a Conexão:
 - Clique no botão “Save & Test” para testar a conexão com o banco de dados SQL Server.
 - Se o teste falhar, um erro comum é a falta de habilitação do TCP/IP no SQL Server.

- Solução, figura 4 como guia, para Erro de Conexão TCP/IP:
 - Vá até iniciar > Microsoft SQL Server > Ferramentas de Configuração > SQL Server Configuration Manager.
 - Abra o "SQL Server Network Configuration" e, em seguida, "Protocols for SQLEXPRESS". Verifique se o protocolo TCP/IP está habilitado. Se não estiver, habilite-o.
 - Nas propriedades do TCP/IP, remova todos os "TCP Dynamic Ports" e adicione o valor 1433 a todas as portas TCP.
 - Reinicie os Serviços do SQL Server.

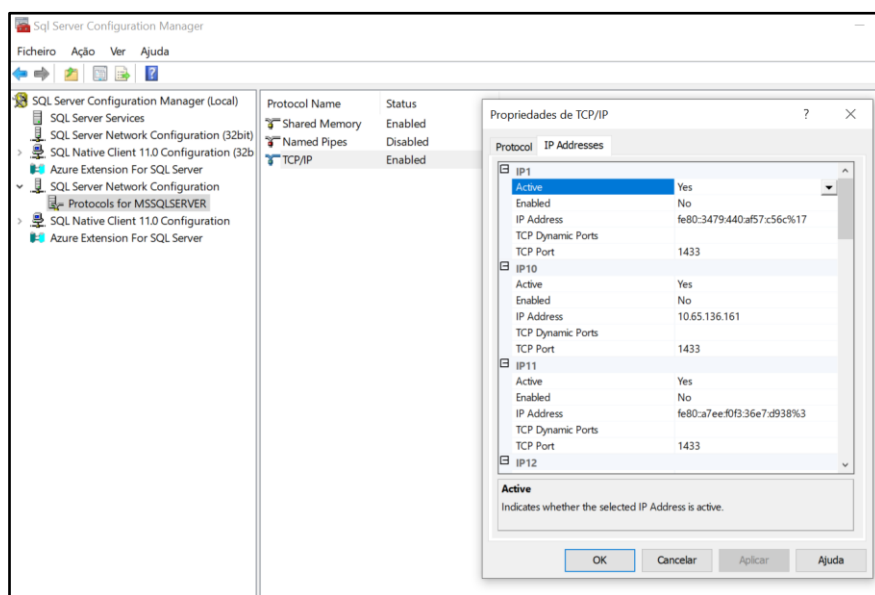


Figura 4 - Configuração TCP/IP no SQL Server Configuration

- Permissão de Usuário:
 - Verifique se o usuário tem as permissões necessárias para acessar e consultar o banco de dados.
- Conclusão:
 - Se o teste for bem-sucedido, você verá uma mensagem de confirmação. Agora, você pode começar a criar dashboards e painéis no Grafana utilizando dados do seu banco de dados SQL Server.

Esses passos garantem que você possa ligar eficientemente o Grafana ao SQL Server, permitindo a exploração e visualização avançada de dados para análises mais profundas e decisões informadas.

5.5. Análise do fluxo de trabalho da ferramenta Grafana

Neste subcapítulo iremos percorrer o “fluxo de trabalho” da ferramenta BI Grafana.

A primeira coisa que vemos, figura 5, quando abrimos o servidor Grafana é a tela inicial da interface do utilizador. Na parte superior, há um cabeçalho de boas-vindas que diz "Welcome to Grafana". Abaixo, o layout é dividido em três colunas principais.

Na primeira coluna à esquerda, sob o título "Basic", há um guia passo a passo destacado para ajudar os novos usuários a terminar de configurar a instalação do Grafana. Este guia contém etapas iniciais como a conexão á fontes de dados e a criação de painéis.

Na coluna central, há um "Tutorial" sobre "Data Source and Dashboards" que oferece aos usuários um tutorial sobre os fundamentos do Grafana. Este tutorial é projetado para orientar os usuários através de todo o processo e cobre os passos de adicionar uma "Data source" e criar "Dashboards".

Na terceira coluna, à direita, existem dois painéis. O primeiro painel, marcado como "Complete", parabeniza o usuário por adicionar sua primeira fonte de dados e oferece um link para a documentação. O segundo painel, sob o título "Dashboards", incentiva o usuário a criar seu primeiro painel, com um link para aprender como fazer isso na documentação.

Abaixo dos painéis principais, na parte inferior da tela, existe uma seção "Dashboards" com subseções para "Starred dashboards" e "Recently viewed dashboards", que permitem aos utilizadores marcar os seus painéis favoritos ou aceder rapidamente os painéis visualizados recentemente.

No canto inferior direito, há atualizações do blog do Grafana, sugerindo que a interface também mantém os usuários informados sobre as últimas notícias e atualizações da comunidade Grafana.

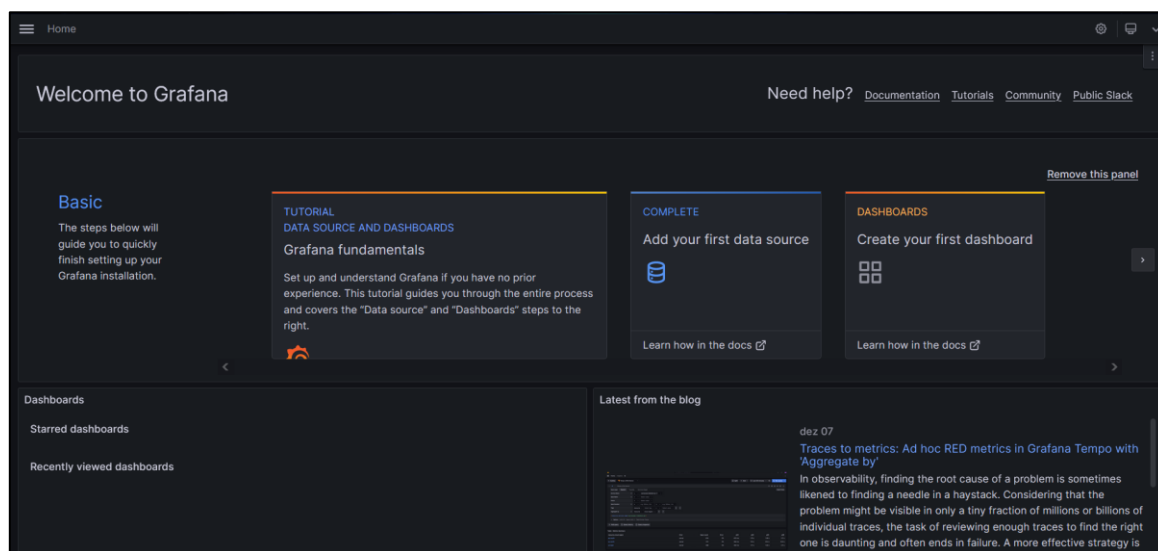


Figura 5 - Tela inicial do servidor Grafana

Avançando para a criação de uma dashboard, figura 6, vemos uma interface de edição de painel dentro da ferramenta de BI Grafana. Na parte superior, há uma barra de navegação com os caminhos "Home > Dashboards > New dashboard > Edit panel", indicando que o utilizador está atualmente a editar um novo painel em um dashboard.

O corpo principal da tela está dividido em duas seções:

- **Seção Esquerda:** Exibe uma área de trabalho onde os utilizadores podem configurar e personalizar as suas consultas de dados. Há uma barra de título "Panel Title", seguida de uma área de visualização que atualmente exibe "No data", indicando que nenhuma consulta foi executada ou não há dados para exibir. Abaixo disso, há várias opções de configuração:
 - Query: É a aba selecionada, onde se pode definir a fonte de dados (neste caso, "Microsoft SQL Server") e configurar a consulta SQL. Existem opções para ajustar o formato da visualização (definido como "Table"), filtragem, agrupamento e ordenação de dados.
 - Transform data: Uma aba onde transformações adicionais nos dados podem ser aplicadas.
 - Alert: Onde alertas podem ser configurados com base nos dados do painel.
 - Há também opções para escolher o dataset e a tabela de onde os dados serão puxados, bem como para selecionar as colunas e as agregações.
- **Seção Direita:** Contém configurações e opções de personalização para o painel, como:
 - Panel options: Para ajustes básicos do painel.
 - Tooltip: Configurações para informações sobre ferramentas que aparecem ao passar o mouse.
 - Legend: Opções para a legenda do gráfico.
 - Axis: Configurações para os eixos do gráfico.
 - Graph styles: Para personalizar a aparência do gráfico.
 - Standard options: Opções padrão para todos os painéis.
 - Data links: Onde links para dados externos podem ser configurados.
 - Value mappings: Para mapear valores de dados para texto ou cores.
 - Thresholds: Definir limites para alertas e marcações visuais nos dados.

Na parte inferior, temos botões para "Run query" que executará a consulta, "Builder" para construir a consulta visualmente e "Code" para editar a consulta em formato de texto puro. Além disso, o "Query Inspector" é uma ferramenta que pode ser usada para depurar e inspecionar as consultas executadas.

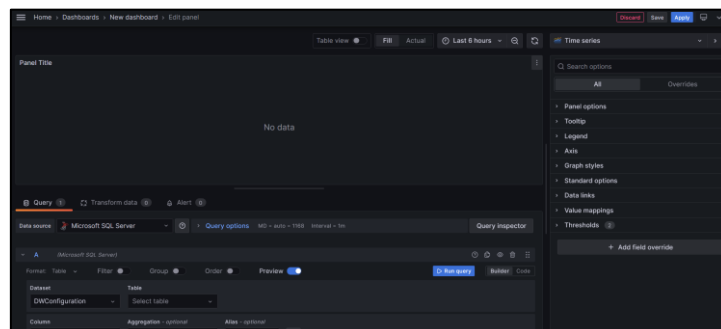


Figura 6 - Página de criação de Dashboards

Observando com mais atenção a tabela lateral podemos ver que existem diversas formas de visualização, cada uma tendo a sua “mecânica” com os dados e a forma de tratar e observar estes. Na figura 7 podemos ver um exemplo dessa tabela lateral e as diferentes formas de visualização:

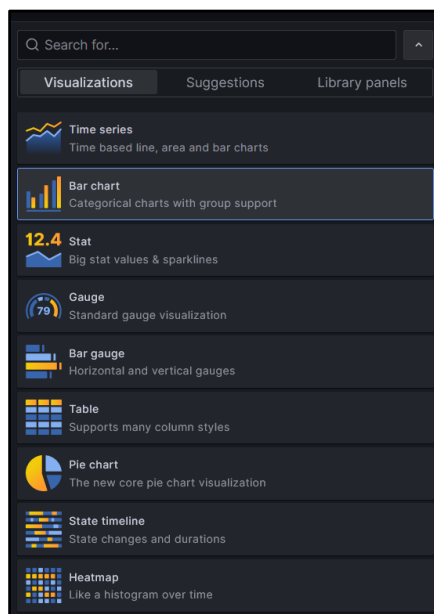


Figura 7 - Tabela lateral com diferentes formas de visualização

Existem duas formas de colocar dados na dashboard, ou com um “builder” integrado ou escrevendo o código SQL da Query. O “builder” é bastante limitado, portanto não iremos usar esse durante a construção das dashboards. Na figura 8 mostramos a tabela de construção de Queries SQL:

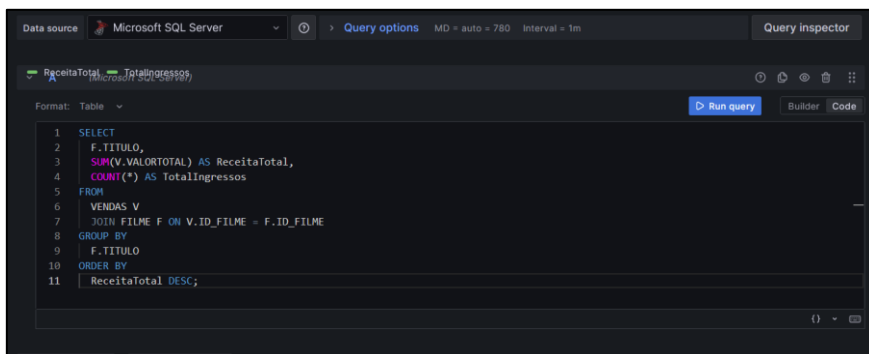


Figura 8 - tabela de construção de Queries SQL

5.6. Criação de Dashboards em Grafana

Iremos desenvolver 5 dashboards na ferramenta BI Grafana com o objetivo de analisar e tirar conclusões sobre esta.

5.6.1. Dashboard de gêneros de filmes mais populares

Este dashboard tem como objetivo apresentar visualmente a divisão de gêneros dos filmes mais populares na rede de cinemas. O gráfico de pizza é uma representação eficaz para mostrar a distribuição de diferentes gêneros de filmes de forma clara e concisa. Ele permite identificar facilmente quais gêneros são os mais populares entre os espectadores.

- Código SQL usado na dashboard 1, figura 9:

```
1 SELECT
2   GENERO,
3   COUNT(*) AS NUMERO_DE_FILMES
4 FROM
5   FILME
6 GROUP BY
7   GENERO
8 ORDER BY
9   NUMERO_DE_FILMES DESC
```

Figura 9 - código SQL da Dashboard 1

- Dashboard 1, figura 10:

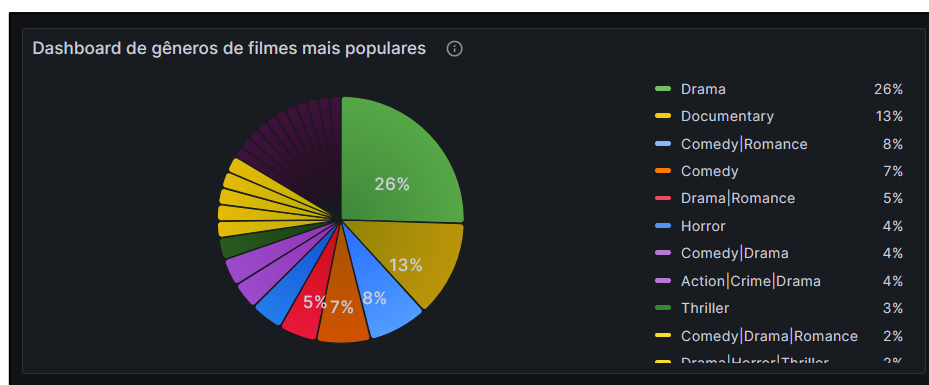


Figura 10 - Dashboard 1, Gêneros de filmes mais populares

5.6.2. Dashboard da receita total por cinema

Esta dashboard exibe a receita total por cinema em um formato simples e direto. Ele fornece uma visão geral das receitas geradas por cada cinema na rede de cinemas. Cada cinema é listado e acompanhado pelo valor total da receita que gerou. Essa visualização permite uma comparação rápida das receitas entre diferentes cinemas, destacando aqueles que estão a ter um desempenho particularmente bom em termos de geração de receita.

- Código SQL usado na dashboard 2, figura 11:

```
1 SELECT TOP 50
2   C.DESIGNACAO AS CINEMA,
3   SUM(V.VALORTOTAL) AS RECEITA_TOTAL
4 FROM
5   CINEMA C
6   INNER JOIN VENDAS V ON C.ID_CINEMA = V.ID_CINEMA
7 GROUP BY
8   C.DESIGNACAO
9 ORDER BY
10  RECEITA_TOTAL DESC
```

Figura 11 - código SQL da Dashboard 2

- Dashboard 2, figura 12:

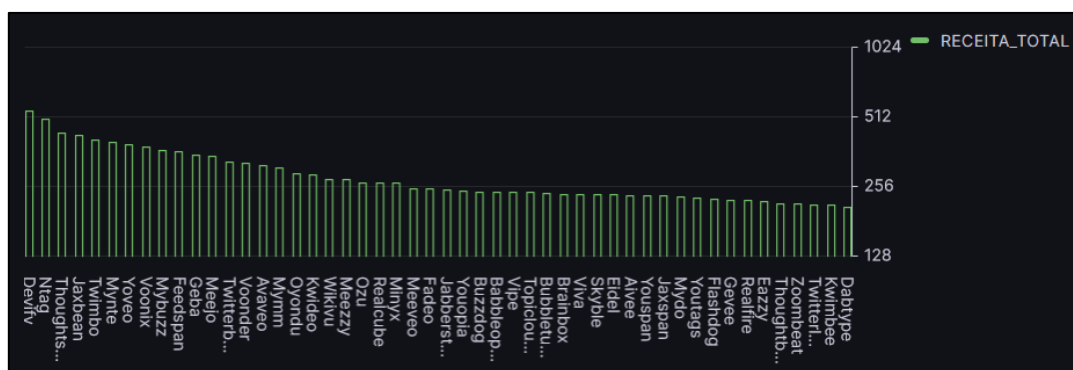


Figura 12 - Dashboard 2, Receita total por cinema

5.6.3. Dashboard da idade média dos clientes por cinema

Este dashboard fornece uma visão geral da idade média dos clientes por cinema na rede. A análise da idade média dos clientes é essencial para entender o público de cada cinema e adaptar as estratégias de marketing e programação de filmes de acordo com as preferências demográficas.

- Código SQL usado na dashboard 3, figura 13:

```
1 SELECT
2   TOP 50 C.DESIGNACAO AS CINEMA,
3   AVG(CL.IDADE) AS IDADE_MEDIA
4 FROM
5   CLIENTE CL
6   INNER JOIN VENDAS V ON CL.ID_CLIENTE = V.ID_CLIENTE
7   INNER JOIN CINEMA C ON V.ID_CINEMA = C.ID_CINEMA
8 GROUP BY
9   C.DESIGNACAO
10 ORDER BY
11   IDADE_MEDIA DESC;
```

Figura 13 - código SQL da Dashboard 3

- Dashboard 3, figura 14:

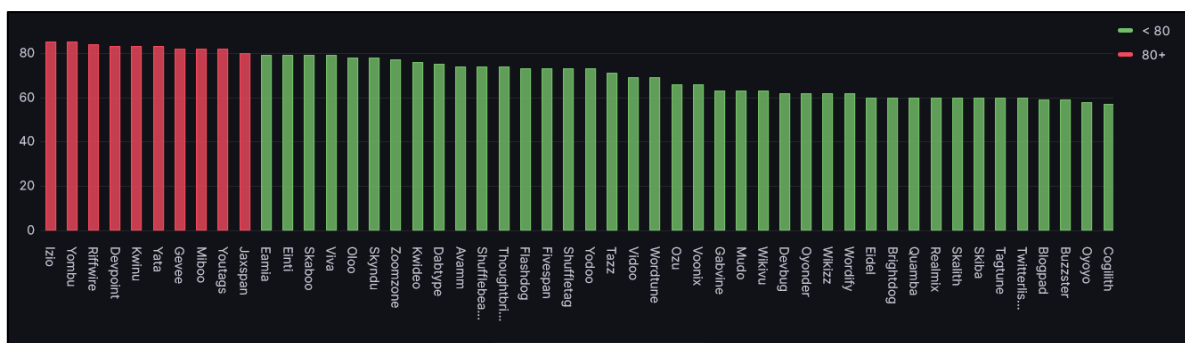


Figura 14 - Dashboard 3, idade média dos clientes por cinema

5.6.4. Dashboard de filmes em exibição por cinema

Este dashboard fornece uma visão em tempo real dos filmes em exibição em cada cinema da rede. Isso é útil para que a equipe de operações e os espectadores saibam quais filmes estão disponíveis em cada local neste exato momento.

- Código SQL usado na dashboard 4, figura 15:

```
1  SELECT DISTINCT
2  |    F.TITULO AS FILME,
3  |    C.DESIGNACAO AS CINEMA
4  FROM
5  |    VENDAS V
6  INNER JOIN
7  |    CINEMA C ON V.ID_CINEMA = C.ID_CINEMA
8  INNER JOIN
9  |    FILME F ON V.ID_FILME = F.ID_FILME
10 ORDER BY
11 |    CINEMA;
```

Figura 15 - Código SQL da Dashboard 4

- Dashboard 4, figura 16:

Dashboard de filmes em exibição por cinema ⓘ	
FILME	CINEMA
Russia's Toughest Prisons (National Geographic)	Abatz
American Outlaws	Ailane
Three Crowns of the Sailor (Les trois couronnes du ...	Aimbu
Circle of Deception, A	Ainyx
Behind the Sun (Abril Despedaçado)	Aivee
Day Night Day Night	Aivee

Figura 16 - Dashboard 4, filmes em exibição por cinema

5.6.5. Dashboard da quantidade de vendas por funcionários

Este dashboard exibe uma tabela que classifica os funcionários com base na quantidade de vendas realizadas em cada cinema da rede. Ele fornece uma visão geral do desempenho de vendas de cada funcionário e permite identificar os funcionários mais produtivos em cada local. Isso é útil para reconhecer o esforço individual dos funcionários e tomar decisões informadas sobre reconhecimento e incentivos.

- Código SQL usado na dashboard 5, figura 17:

```
1  SELECT
2    C.DESIGNACAO AS CINEMA,
3    F.NOME AS FUNCIONARIO,
4    COUNT(V.ID_VENDA) AS VENDAS_REALIZADAS
5  FROM
6    VENDAS V
7    JOIN CINEMA C ON V.ID_CINEMA = C.ID_CINEMA
8    JOIN FUNCIONARIO F ON V.ID_FUNCIONARIO = F.ID_FUNCIONARIO
9  GROUP BY
10   C.DESIGNACAO,
11   F.NOME
12  ORDER BY
13   C.DESIGNACAO ASC,
14   VENDAS_REALIZADAS DESC;
```

Figura 17 - Código SQL da Dashboard 5

- Dashboard 5, figura 18:

CINEMA	FUNCIONARIO	VENDAS_REALIZADAS
Abatz	Ethyl Whartonby	1
Ailane	Miles Kainz	1
Aimbu	Correy Manger	1
Ainyx	Leanora Ainsbury	1
Aivee	Demetria St. Leger	1
Aivee	Wendell Postan	1

Figura 18 - Dashboard 5, quantidade de vendas por funcionários

5.7. Comparações com a ferramenta BI Power BI

O Grafana é uma ferramenta de Business Intelligence (BI) open-source conhecida pela sua flexibilidade e capacidade de integração com diversas fontes de dados. Sua natureza open-source é uma grande vantagem, oferecendo personalização e economia. No entanto, seu uso eficaz requer conhecimento avançado de SQL, representando um desafio para usuários menos técnicos.

Em contraste, o Power BI da Microsoft é conhecido por sua interface intuitiva e facilidade de uso, tornando a análise de dados acessível até mesmo para aqueles com habilidades técnicas limitadas. Possui uma ampla gama de visualizações e integrações avançadas, incluindo capacidades de inteligência artificial.

Embora o Grafana seja eficiente para visualizações em tempo real e monitoramento de dados, ele não oferece a mesma variedade e sofisticação em gráficos e detalhamento de dados que o Power BI. Assim, enquanto o Grafana é ideal para cenários específicos que necessitam de monitoramento intensivo de dados, o Power BI pode ser mais adequado para utilizadores que procuram uma ferramenta de BI fácil de usar com funcionalidades mais abrangentes. A escolha entre Grafana e Power BI depende das necessidades específicas da organização e do nível de experiência técnica dos utilizadores.

Para exemplificar melhor esta ideia abaixo mostramos a Dashboard 5, anteriormente feita em Grafana, página 18, agora feita em Power BI, figura 19:

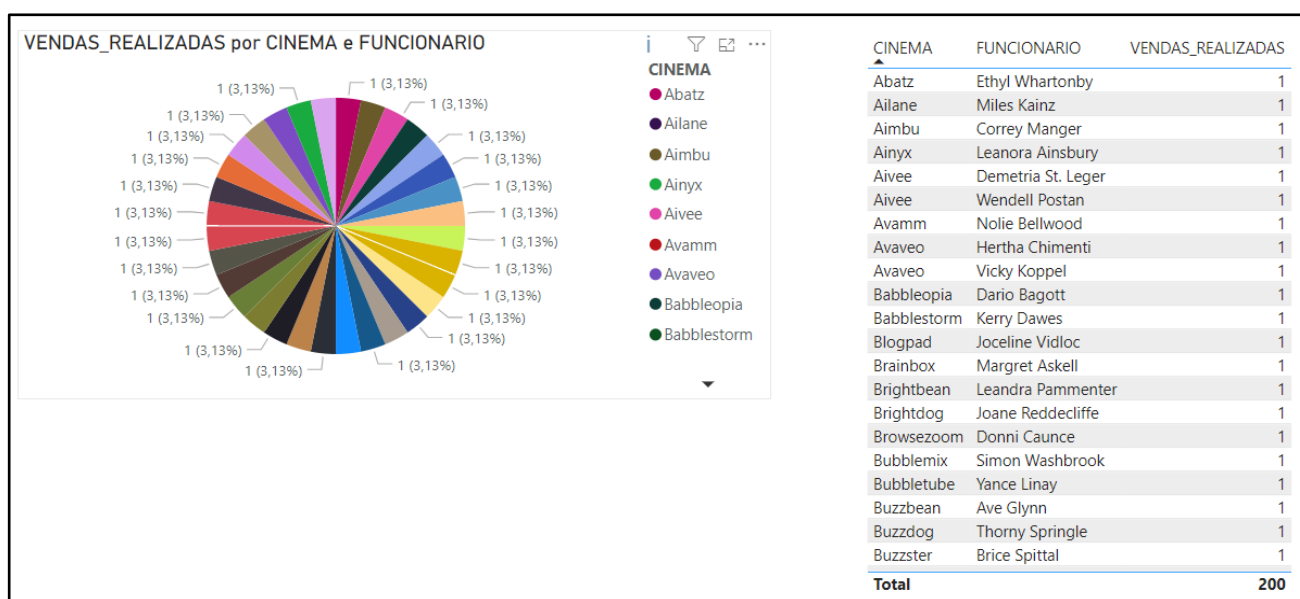


Figura 19 - Dashboard 5, quantidade de vendas por funcionários, em Power BI

Para destacar a facilidade de utilização da ferramenta Power BI em comparação com o Grafana, apresentamos, na Figura 20, a configuração da dashboard 5. Enquanto no Grafana foi necessário utilizar uma query em sql para apresentar os resultados do estudo, no Power BI essa tarefa é significativamente simplificada, exigindo apenas a seleção dos atributos desejados de cada tabela para o estudo, e à criação de uma medida com o propósito de contabilizar o número de vendas efetuadas.

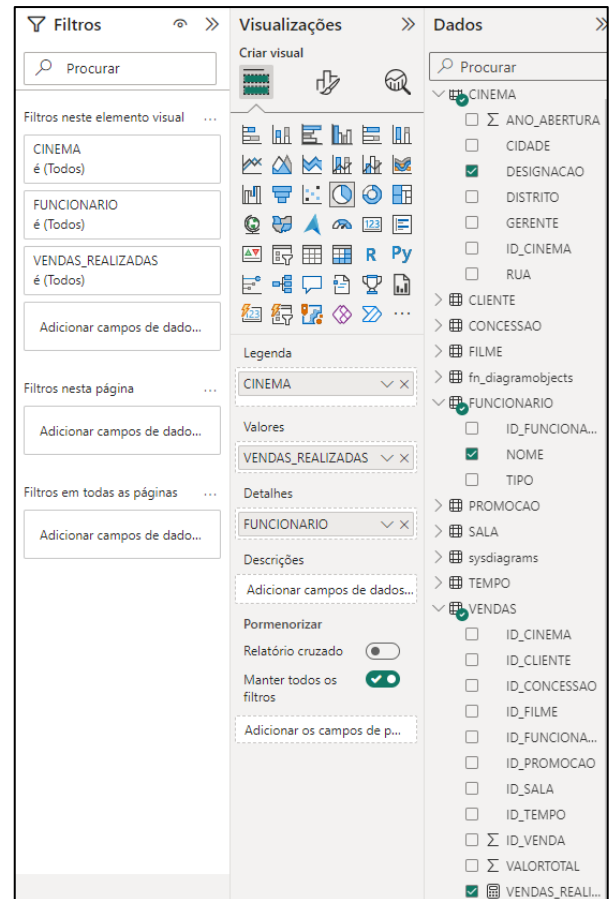


Figura 20 -Configuração da Dashboard 5

6. Criação de uma datawarehouse

Com base no seguinte enunciado da segunda parte da ficha de trabalho iremos realizar toda uma lista de tarefas com o objetivo final de estudar o caso de uma cadeia de lojas de retalho e criar uma DW para esta:

Uma cadeia de lojas de retalho tem estabelecimentos lojas em todo o território nacional. Cada uma das lojas é um moderno e típico supermercado com um conjunto completo de departamentos que incluem mercearia, alimentos congelados, frescos, carne, padaria, licores e vinhos, artigos de limpeza, máquinas e eletrodomésticos, maquinaria, artigos para casa e jardim, etc. Cada loja tem aproximadamente 60000 produtos individuais nas suas prateleiras.

Os produtos individuais são chamados SKUs (stock keeping units). Cerca de 40000 destes SKUs provêm de fabricantes externos e têm códigos de barras impressos na embalagem do produto. Esses códigos de barras são denominados UPCs (universal product codes). Os UPCs estão ao mesmo nível de granularidade dos SKUs individuais. Cada embalagem diferente de um produto tem um UPC separado e então é um SKU separado. Os restantes 20000 SKUs provêm de departamentos como talho, padaria, frescos e não dispõem de códigos UPC reconhecidos a nível nacional. A loja atribui códigos SKU a esses produtos e coloca etiquetas de códigos de barras nesses itens. Embora os códigos de barras não sejam UPCs, são, decerto, códigos SKU. Os dados são recolhidos em diversos locais na loja, mas principalmente nas caixas registadoras (TPVs), quando os clientes adquirem os produtos. Os TPVs usam scanners para ler os códigos de barras, sendo o local onde as compras dos clientes são medidas. O armazém, onde recebem os produtos dos fabricantes ou outros grandes distribuidores, é outro ponto interessante de recolha de dados.

Na loja, a gestão está especialmente interessada com a logística das encomendas, armazenamento nas prateleiras e vendas dos produtos, maximizando os lucros em cada loja. O lucro final provém de sobrecarregar tanto quanto possível cada produto, baixando os custos de aquisição e sobretaxas, e, ao mesmo tempo, atrair o máximo de clientes possível.

Algumas das decisões de gestão mais importantes têm a ver com preços e promoções. Quer a gestão da loja, quer os escritórios centrais, passam muito tempo a lidar com os preços e lançar promoções. As promoções numa loja incluem reduções temporárias de preços, cupões em jornais e cartazes na loja. A forma mais direta e eficaz de criar um grande aumento nas vendas de produtos é baixar o preço de forma dramática. Mas, infelizmente, uma tão grande redução de preço não é sustentável, pois que, provavelmente estarão a ser vendidos a um preço abaixo do preço do custo. Por esse motivo, a visibilidade de todas as formas de promoção é uma parte importante das operações de análise realizadas na loja.

6.1. Processo de negócio a modelar

Ao abordar a criação de uma Data Warehouse, o processo de negócio escolhido para modelar deve ser aquele que tem o maior impacto operacional e estratégico. Para a cadeia de lojas de retalho em questão, o processo de gestão de vendas e promoções foi identificado como o mais crítico devido à sua influência direta na rentabilidade e na atração de clientes. Este processo foi selecionado pois aborda as questões de negócio mais urgentes, tais como:

- Análise de rentabilidade por produto e loja.*
- Efetividade das campanhas promocionais.*
- Dinâmica de preços e o seu impacto nas vendas.*
- Comportamento de compra dos clientes em resposta a promoções.*

Os dados para este processo estão prontamente disponíveis e são captados em pontos críticos do funcionamento da loja, como as caixas registadoras e o armazém. Estes dados formam a espinha dorsal das operações analíticas e são essenciais para a tomada de decisões informadas que maximizem os lucros.

6.2. Avaliação da Granularidade

A granularidade refere-se ao nível de detalhe que os dados representam no modelo dimensional de uma Data Warehouse. A escolha da granularidade é uma decisão crítica no design do modelo, pois impacta diretamente a capacidade de análise e o desempenho do sistema. Para a cadeia de lojas de retalho em questão, a granularidade foi cuidadosamente avaliada para garantir um equilíbrio entre detalhe analítico e eficiência de armazenamento e processamento.

No modelo em estrela desenvolvido, a granularidade foi definida ao nível da transação individual de venda. Cada registo na tabela de factos Vendas captura dados de uma única transação, incluindo:

- A identificação da loja onde a transação ocorreu.
- O SKU do produto vendido.
- O momento exato da venda, com precisão até à data e hora.
- Detalhes da promoção aplicada, se houver.

Esta granularidade permite análises profundas e específicas, tais como:

- Desempenho de vendas de produtos individuais por dia ou hora.
- Impacto imediato de promoções específicas nas vendas.
- Compreensão do fluxo de vendas intra-diário para otimização de pessoal e operações.
- Análise de tendências de consumo a nível de item individual, que pode ser correlacionada com eventos específicos ou promoções.

Ao mesmo tempo, a granularidade foi limitada para evitar a inclusão de detalhes operacionais desnecessários que poderiam sobrecarregar o sistema sem agregar valor significativo à análise. Por exemplo, não se incluem detalhes ao nível do método de pagamento ou dos segundos exatos da transação, pois estes não foram considerados relevantes para as questões de negócio prioritárias.

A granularidade escolhida está, portanto, alinhada com as necessidades de negócio mais prementes da cadeia de lojas e com a capacidade do sistema de prover insights valiosos sem comprometer o desempenho.

6.3. Selecionar dimensões e atributos

A construção de um modelo dimensional robusto e eficaz para a nossa Data Warehouse requer uma seleção criteriosa das dimensões e dos seus atributos. As dimensões são as perspectivas segundo as quais analisaremos os nossos factos de negócio, e os atributos são os detalhes específicos que caracterizam cada uma dessas perspectivas.

No nosso modelo em estrela para a gestão de vendas e promoções, identificamos as seguintes dimensões e os seus atributos correspondentes:

- **Dimensão Loja:**
 - ID_Loja: O identificador único de cada loja dentro da cadeia.
 - Nome_Loja: O nome comercial da loja.
 - Rua, Cidade, Distrito: Localização detalhada de cada loja, permitindo análises geográficas.
- **Dimensão Produto:**
 - ID_Produto: O identificador único de cada SKU.
 - Nome_Produto: A denominação do produto.
 - Categoria: A classificação do produto, como mercearia, eletrodomésticos, etc.
 - UPC: O código de barras universal quando aplicável.
- **Dimensão Tempo:**
 - ID_Tempo: Uma chave única para cada unidade de tempo.
 - Data, Ano, Mês, Trimestre, Semestre: Unidades de tempo para análise temporal.
 - Dia da Semana, Dia do Mês: Permite analisar padrões de venda semanais e mensais.
- **Dimensão Promoção:**
 - ID_Promocao: Identificador único de cada promoção.
 - Descricao_Promocao: Descrição textual da promoção.
 - Data_Inicio, Data_Fim: Período de vigência da promoção.
 - Desconto: Valor ou percentagem do desconto oferecido.

Estes atributos foram escolhidos para fornecer uma visão abrangente e multifacetada das operações de venda e promoção, possibilitando uma análise detalhada das estratégias de preço e promoção e do seu impacto nas vendas. As dimensões e atributos foram concebidos para suportar uma ampla variedade de consultas analíticas, desde a compreensão do desempenho das vendas até a otimização de estratégias de marketing e inventário.

6.4. Selecionar factos

Em uma Data Warehouse, os factos representam as medidas quantitativas em torno das quais as análises são conduzidas. A seleção adequada dos factos é crítica, pois eles formam o núcleo da análise de negócios, permitindo aos gestores avaliar o desempenho, identificar tendências e tomar decisões informadas. No nosso modelo dimensional, os seguintes factos foram meticulosamente escolhidos para fornecer insights valiosos no processo de gestão de vendas e promoções:

- **Valor Total de Vendas:**
 - Racional: Este é o principal indicador de desempenho, refletindo a receita gerada. Facilita a análise do sucesso das vendas em diferentes níveis, como por loja, região ou período.
- **Quantidade Vendida:**
 - Racional: Fornece insights sobre o volume de produtos vendidos, o que é crucial para compreender a popularidade dos itens e a eficácia da gestão de estoque.
- **Número de Transações:**
 - Racional: Uma métrica importante que ajuda a entender o tráfego de clientes e a frequência de compras, indicando a eficácia das estratégias de atração de clientes.
- **Desconto Aplicado:**
 - Racional: Crucial para avaliar o impacto das promoções no comportamento de compra e na rentabilidade das vendas.
- **Margem de Lucro:**
 - Racional: Indica a lucratividade de cada item vendido, permitindo uma análise detalhada da contribuição de cada produto para os lucros gerais.
- **Custo dos Bens Vendidos (COGS):**
 - Racional: Essencial para entender os custos associados aos produtos vendidos e para calcular a margem de lucro de maneira precisa.

Estes factos foram selecionados para alavancar o processo de negócio escolhido - gestão de vendas e promoções - com dados que são críticos para a cadeia de lojas de retalho. Eles são imediatamente disponíveis através dos sistemas de ponto de venda e gestão de estoque, garantindo que as análises sejam baseadas em dados atuais e precisos. Além disso, a combinação desses factos permite uma análise abrangente que vai desde a performance financeira até a eficácia operacional, dando suporte ao processo decisório em todos os níveis da organização.

6.5. Modelo em estrela criado

A construção de um modelo em estrela é o ponto de partida para a organização e a análise de dados multidimensionais em uma Data Warehouse. O modelo em estrela é composto por uma tabela de factos central, que regista eventos de negócios mensuráveis, e um conjunto de tabelas de dimensão que categorizam e descrevem os factos. Os atributos mais significativos foram cuidadosamente selecionados para cada tabela para fornecer insights valiosos e apoiar decisões estratégicas. A figura 21 representa o modelo:

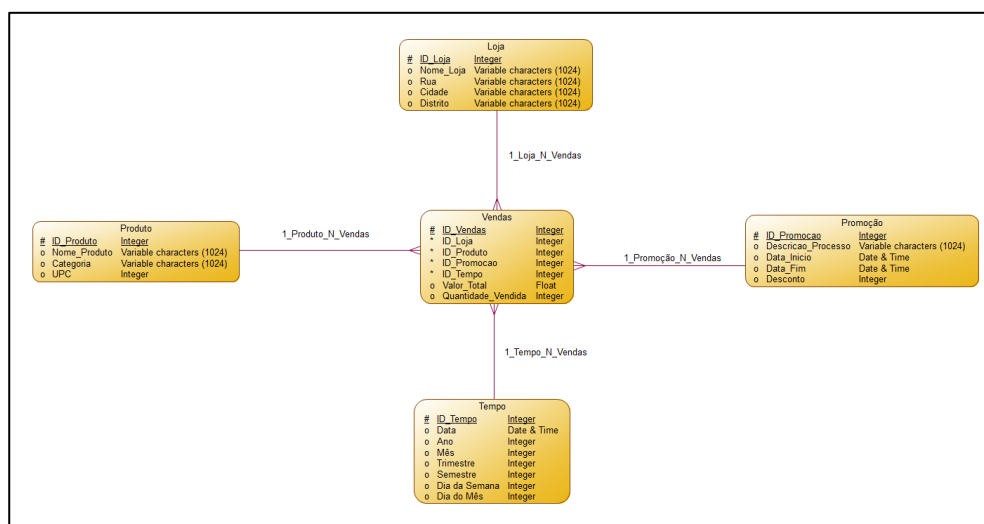


Figura 21 - Modelo estrela da segunda parte do trabalho

O dicionário de dados é um componente essencial do modelo em estrela, pois fornece uma descrição detalhada das tabelas, colunas e relações. Para cada entrada no dicionário, incluímos o nome da coluna, o tipo de dados, uma descrição breve e as relações de chave estrangeira, se aplicáveis. Isso não apenas facilita a compreensão do modelo para os usuários finais, mas também serve como um guia para a manutenção e a expansão do sistema.

Tabela de Factos 'Vendas':

- ID_Vendas: Chave primária que identifica de forma única cada venda.
- ID_Loja: Chave estrangeira que referencia a dimensão Loja.
- ID_Produto: Chave estrangeira que referencia a dimensão Produto.
- ID_Tempo: Chave estrangeira que referencia a dimensão Tempo.
- ID_Promocao: Chave estrangeira que referencia a dimensão Promoção.
- Valor_Total, Quantidade_Vendida: Métricas de negócio que representam o resultado financeiro e quantitativo das vendas.

Dimensões:

- **Dimensão Loja:**
 - ID_Loja, Nome_Loja, Rua, Cidade, Distrito: Descrevem a localização e identificação de cada loja.
- **Dimensão Produto:**
 - ID_Produto, Nome_Produto, Categoria, UPC: Identificam o produto e categorizam sua natureza e origem.
- **Dimensão Tempo:**
 - ID_Tempo, Data, Ano, Mês, Trimestre, Semestre, Dia da Semana, Dia do Mês: Capturam a dimensão temporal das vendas.
- **Dimensão Promoção:**
 - ID_Promocao, Descricao_Promocao, Data_Inicio, Data_Fim, Desconto: Caracterizam as promoções e permitem análises do impacto das mesmas nas vendas.

6.6. Cálculos do tamanho aproximado do Modelo

De seguida iremos efetuar os cálculos necessários para saber, aproximadamente, quanto irá ocupar o modelo:

Cálculo Parâmetros Físicos (EXEMPLO)

HEADER FIXO (HF)= 84 bytes

HEADER VARIÁVEL (HV)= 5 bytes por cada registo

(2 bytes header do registo + 1 byte com nº de colunas) + (2 bytes no Row directory)

PCTFREE = 10 (ENTRE 1 A 25)

PCTUSED = 70 (ENTRE 40 A 95)

TAMANHO DO BLOCO (T.B.) = 4096 Bytes (por defeito)

TAMANHO MEDIO DO REGISTO (T.M.R.) =

SOMA(Tamanho médio dos campos)

+ 5 bytes por registo (Header Variável)

+ 1 byte por cada coluna do registo

ESPAÇO LIVRE NO BLOCO (E.L.B.) =

Tamanho do Bloco * (100 - PCTFREE) / 100 - Header Fixo

Nº DE REGISTOS POR BLOCO (N.R.B.) =

Espaço Livre do Bloco / Tamanho Médio do Registo (ARREDONDADO PARA BAIXO)

Cálculo do NEXT

Nº DE BLOCOS (N.B. Previstos) =

Nº Registos Previstos / Nº Registos por Bloco (ARREDONDADO PARA CIMA)

ESPAÇO NEXT DA TABELA (E.N.T.) =

Nº Blocos Previstos * Tamanho do Bloco

*INT = 4 bytes, VARCHAR(255) = 50 bytes, date&time = 8 bytes

Tamanho médio dos campos da tabela Loja:

TAMANHO MEDIO DO REGISTO (T.M.R.) = $4 + 50 + 50 + 50 + 50 + 5 + 5 = 214$ bytes

ESPAÇO LIVRE NO BLOCO (E.L.B.) = $4096 * (100 - 10) / 100 - 84 = 4012,90$ bytes

Nº DE REGISTOS POR BLOCO (N.R.B.) = $4012,90 / 214 = 18$

Cálculo do NEXT

Nº DE BLOCOS (N.B. Previstos) = $25 / 18 = 2$ *25 lojas

ESPAÇO NEXT DA TABELA (E.N.T.) = $2 * 4096 = 8192$ bytes

Tamanho médio dos campos da tabela Produto:

TAMANHO MEDIO DO REGISTO (T.M.R.) = $4 + 50 + 50 + 4 + 5 + 4 = 117$ bytes

ESPAÇO LIVRE NO BLOCO (E.L.B.) = $4096 * (100 - 10) / 100 - 84 = 4012,90$ bytes

Nº DE REGISTOS POR BLOCO (N.R.B.) = $4012,90 / 117 = 34$

Cálculo do NEXT

Nº DE BLOCOS (N.B. Previstos) = $5000 / 34 = 148$ *5000 produtos por dia

ESPAÇO NEXT DA TABELA (E.N.T.) = $148 * 4096 = 593910$ bytes

Tamanho médio dos campos da tabela Tempo:

TAMANHO MEDIO DO REGISTO (T.M.R.) = $4 + 8 + (4 * 6) + 5 + 8 = 49$ bytes

ESPAÇO LIVRE NO BLOCO (E.L.B.) = $4096 * (100 - 10) / 100 - 84 = 4012,90$ bytes

Nº DE REGISTOS POR BLOCO (N.R.B.) = $4012,90 / 49 = 81$

Cálculo do NEXT

Nº DE BLOCOS (N.B. Previstos) = $(365 * 5) / 81 = 23$ *5 anos de dados

ESPAÇO NEXT DA TABELA (E.N.T.) = $23 * 4096 = 94208$ bytes

Tamanho médio dos campos da tabela Promoção:

TAMANHO MEDIO DO REGISTO (T.M.R.) = $4 + 50 + 8 + 8 + 4 + 5 + 5 = 84$ bytes

ESPAÇO LIVRE NO BLOCO (E.L.B.) = $4096 * (100 - 10) / 100 - 84 = 4012,90$ bytes

Nº DE REGISTOS POR BLOCO (N.R.B.) = $4012,90 / 84 = 47$

Cálculo do NEXT

Nº DE BLOCOS (N.B. Previstos) = $3000 / 47 = 64$ *3000 combinações de promoções

ESPAÇO NEXT DA TABELA (E.N.T.) = $64 * 4096 = 262144$ bytes

Tamanho médio dos campos da tabela Vendas:

TAMANHO MEDIO DO REGISTO (T.M.R.) = $4 + (4 * 4) + 8 + 4 + 5 + 7 = 44$ bytes

ESPAÇO LIVRE NO BLOCO (E.L.B.) = $4096 * (100 - 10) / 100 - 84 = 4012,90$ bytes

Nº DE REGISTOS POR BLOCO (N.R.B.) = $4012,90 / 44 = 91$

Cálculo do NEXT

Nº DE BLOCOS (N.B. Previstos) = $(1000 * 20 * 365 * 5 * 25) / 91 = 10027473$

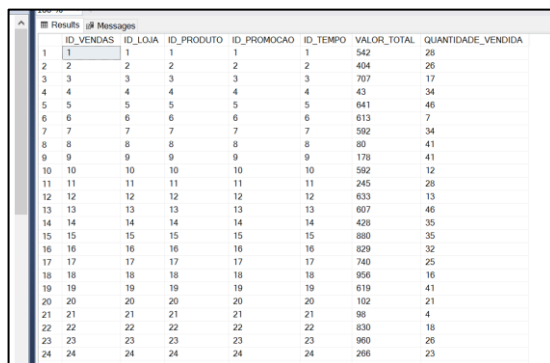
*1000 talões com 20 linhas cada, por dia, por 5 anos, por 25 lojas

ESPAÇO NEXT DA TABELA (E.N.T.) = $10027473 * 4096 = 4,107 * 10^9$ bytes = 39167.40 Mbytes

Tamanho total calculado para o modelo = 39167,40 Mbytes

6.7. Tratamento do modelo em estrela no SQL Server

De forma a podermos trabalhar o problema no Visual Studio e no PowerBI foi necessário criarmos a BD na SQL Server e popular esta. Nas figuras 22 e 23 podemos ver a Base de dados no SQL Server e alguns dos dados embutidos nesta:



ID_VENDAS	ID_LOJA	ID_PRODUTO	ID_PROMOCAO	ID_TEMPO	VALOR_TOTAL	QUANTIDADE_VENDIDA
1	1	1	1	1	542	28
2	2	2	2	2	404	25
3	3	3	3	3	707	17
4	4	4	4	4	43	34
5	5	5	5	5	641	46
6	6	6	6	6	613	7
7	7	7	7	7	592	34
8	8	8	8	8	80	41
9	9	9	9	9	178	41
10	10	10	10	10	592	12
11	11	11	11	11	245	28
12	12	12	12	12	633	13
13	13	13	13	13	607	46
14	14	14	14	14	428	35
15	15	15	15	15	880	35
16	16	16	16	16	829	32
17	17	17	17	17	740	25
18	18	18	18	18	956	16
19	19	19	19	19	619	41
20	20	20	20	20	102	21
21	21	21	21	21	98	4
22	22	22	22	22	830	18
23	23	23	23	23	960	26
24	24	24	24	24	266	23
25	25	25	25	25	521	25

Figura 22

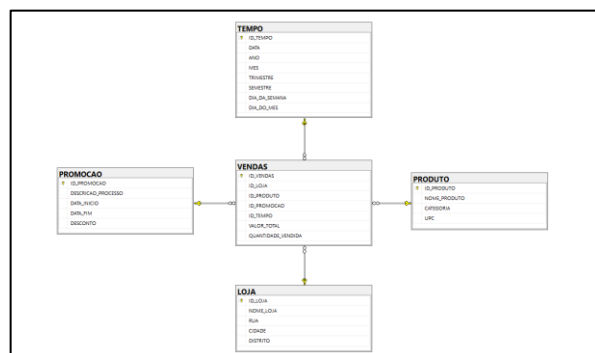


Figura 22

6.8. Vistas no Visual Studio Data Warehouse

Com o intuito de apoiar o estudo de um gestor, foram escolhidas cinco vistas no Visual Studio Data Warehouse que proporcionam uma visão abrangente e estratégica do desempenho e das tendências de negócios.

Cada perspectiva foi selecionada com base na sua capacidade de fornecer informações valiosas para apoiar as decisões do gestor no âmbito do processo de negócio.

6.8.1. Vista de Performance de Vendas por Loja

A vista, como ilustra a Figura 22, permite que gestores identifiquem as lojas que estão superando as expectativas em termos de volume de vendas e valor obtido nessas transações. Essas lojas de alto desempenho podem servir como referências para a implementação de boas práticas em outras unidades.

Ao ter uma visão detalhada do desempenho individual de cada loja, os gestores podem adaptar estratégias para atender às necessidades específicas de cada localidade. Isso envolve ajustes em campanhas de marketing, promoções especiais e eventos promocionais direcionados.

A capacidade de comparar o desempenho entre diferentes lojas fornece informações valiosas sobre tendências e padrões, auxiliando os gestores a compreender as variáveis que influenciam o sucesso de cada unidade. Essa abordagem integrada permite uma gestão mais eficaz, pois os gestores podem tomar decisões informadas com base em uma compreensão abrangente do desempenho global das lojas.

ID LOJA	NOME LOJA	QUANTIDADE VENDIDA	VALOR TOTAL
1	Skyvu	28	542
2	Twitterbeat	26	404
3	Youopia	17	707
4	Jabbersphere	34	43
5	Tekfly	46	641
6	Realfire	7	613
7	Dazzlesphere	34	592
8	Quimm	41	80
9	Jabbertype	41	178
10	Vimbo	12	592
11	Oyundu	28	245
12	Edgedclub	13	633
13	Oyoba	46	607
14	Fadeo	35	428
15	Buzzbean	35	880
16	Voonte	32	829
17	Skyble	25	740
18	Wikibox	16	956
19	Miboo	41	619
20	Feedbug	21	102
21	Rhynyx	4	98
22	Youtags	18	830
23	Meejo	26	960
24	Edgedclub	23	266

Figura 23

6.8.2. Vista de Tendências de Vendas Temporais

A vista, como mostrado na Figura 23, permite uma análise visual e temporal crucial para os gestores identificarem desempenho em lojas e compreenderem tendências ao longo do tempo.

Permite compreender variações nas vendas ao longo do tempo, o que é fundamental para estratégias de negócios mais eficazes maximizando oportunidades de vendas, como por exemplo, aumentar stocks em períodos de alta demanda.

Ao compreender as tendências, os gestores otimizam recursos, evitam excesso de stock em períodos lentos e garantem disponibilidade nos momentos de maior procura.

ID TEMPO	DATA	MES	TRIMESTRE	QUANTIDADE VENDIDA
1	2023...	4	1	28
2	2023...	3	3	26
3	2023...	11	1	17
4	2023...	6	2	34
5	2023...	7	3	46
6	2023...	8	1	7
7	2023...	8	2	34
8	2023...	7	1	41
9	2023...	6	3	41
10	2022...	6	2	12
11	2023...	8	1	28
12	2023...	3	3	13
13	2023...	8	3	46
14	2023...	6	1	35
15	2023...	10	1	35
16	2023...	2	2	32
17	2023...	7	1	25
18	2023...	11	1	16
19	2023...	4	1	41
20	2023...	5	3	21
21	2023...	6	3	4
22	2023...	7	1	18
23	2023...	4	2	26
24	2023...	8	2	23

Figura 24

6.8.3. Vista de Eficiência de Promoções por Região

Ao categorizar promoções por cidade, os gestores identificam tendências e padrões regionais, ajustando ofertas com base nas respostas a descontos, produtos ou métodos promocionais.

Com resultados regionais, os gestores personalizam estratégias, otimizando o orçamento promocional para áreas de maior retorno.

Essa abordagem estratégica promove uma gestão flexível, adaptando-se às necessidades específicas de cada localidade.

CIDADE	ID PROMOCAO	DESCONTO	QUANTIDADE VENDIDA
Abilene	168	74	6
Albany	1	37	38
Albany	42	35	13
Albany	67	75	12
Albany	167	22	30
Arlington	62	60	44
Arlington	83	38	23
Asheville	124	48	29
Asheville	184	74	43
Atlanta	100	35	17
Atlanta	191	57	4
Austin	69	14	3
Baltimore	29	25	20
Baton...	66	62	26
Beaver...	193	25	35
Bellevue	111	14	49
Berkeley	187	52	39
Bethesda	31	46	2
Bethesda	93	38	24
Birming...	24	40	23
Birming...	96	19	4
Birming...	120	71	45
Bloom...	37	25	16
Bloom...	183	35	8

Figura 25

6.8.4. Vista de Análise de Desempenho de Categorias de Produtos

A vista, conforme ilustra a Figura 25, oferece uma perspectiva específica e detalhada sobre as vendas de produtos de categoria 190E. Essa vista é essencial para otimizar a gestão desses itens. Ao apresentar informações detalhadas por loja (ID_Loja), a vista possibilita uma análise individual do desempenho dos produtos 190E em cada estabelecimento. Essa abordagem é valiosa para ajustar estratégias de acordo com os momentos de maior procura e padrões de consumo locais.

As métricas, como Quantidade_Vendida e Valor_Total, oferecem uma visão abrangente do desempenho desse produto. Isso capacita os gestores a tomar decisões mais informadas sobre estratégias de promoção e gestão de estoque, utilizando os dados detalhados fornecidos por essa vista específica.

Dimension	Hierarchy	Operator	Filter Expression
PRODUTO	CATEGORIA	Equal	{ 190E }
<Select dimension>			
ID LOJA	CATEGORIA	QUANTIDADE VENDIDA	VALOR TOTAL
124	190E	29	269

Figura 26

6.8.5. Vista de Performance de Vendas por Loja no mês de setembro

Esta vista, como mostrado na Figura 26, concentra-se nas vendas específicas do mês de setembro. Isso permite uma análise detalhada do desempenho das lojas durante esse período específico, identificando padrões sazonais que possam influenciar as vendas.

Além disso, inclui métricas como receita total e quantidade de produtos vendidos que permitem avaliar o desempenho de cada loja.

Portanto, a vista permite a identificação rápida de lojas que se destacaram durante o mês de setembro. Essas lojas podem servir como modelos para estratégias bem-sucedidas que podem ser replicadas em outras unidades.

Dimension	Hierarchy	Operator	Filter Expression	
TEMPO	MES	Equal	{ 9 }	
<Select dimension>				
ID LOJA	MES	NOME LOJA	QUANTIDADE VENDIDA	VALOR TOTAL
29	9	Teklist	15	873
37	9	Rhynoodie	15	526
48	9	Quatz	6	401
65	9	Buzbean	22	137
66	9	Mynyx	26	374
100	9	Feednation	17	152
102	9	Shuffletag	9	991
107	9	Truideo	19	76
111	9	Gevee	49	558
112	9	Zoovee	47	292
118	9	Browsezoom	8	623
125	9	Thoughtap...	23	981
128	9	LiveTube	30	317
137	9	Twinder	35	628
139	9	Eadel	30	161
145	9	Tantbee	17	42
146	9	Topichots	30	941
186	9	Zoombeat	41	687

Figura 27

6.9. Dashboard no Power BI

Aproveitando a potência do Power BI, os nossos dashboard's oferecem análises sólidas e intuitivas com base no modelo em estrela previamente estabelecido. Cada visualização foi cuidadosamente concebida para fornecer percepções valiosas sobre o processo de negócio em questão.

Desde a quantidade vendida por loja até análises detalhadas de descontos por produto e cidade, exploraremos cinco vistas distintas, realçando a flexibilidade e utilidade de cada uma. Este conjunto de dashboards tem como objetivo capacitar gestores, analistas e equipas de negócios a tomarem decisões informadas e estratégicas.

6.9.1. Dashboard - Quantidade Vendida por Loja

Este painel no Power BI, conforme ilustrado na Figura 27, proporciona uma visão detalhada da quantidade de produtos vendidos em diversas lojas, permitindo uma filtragem eficaz por cidade, nome da loja e ano.

Esta abordagem simplifica a análise específica do desempenho de vendas em lojas particulares, em diversas localidades e ao longo do tempo. Essa visualização é particularmente útil para gestores de lojas e equipas de vendas, possibilitando a identificação de padrões sazonais, comparação de desempenho entre lojas e a tomada de decisões direcionadas para otimizar resultados.



Figura 28

6.9.2. Dashboard - Quantidade Vendida por Ano

Este painel, conforme mostrado na Figura 28, concentra-se na quantidade total de produtos vendidos ao longo dos anos, oferecendo uma visão abrangente das tendências de vendas ao longo do tempo.

Com a capacidade de filtrar por data e nome da loja, os utilizadores podem realizar análises comparativas e identificar variações significativas nas vendas. Este recurso é valioso para gestores e analistas de negócios que desejam compreender a evolução do desempenho de vendas em diferentes períodos e contextos específicos das lojas.

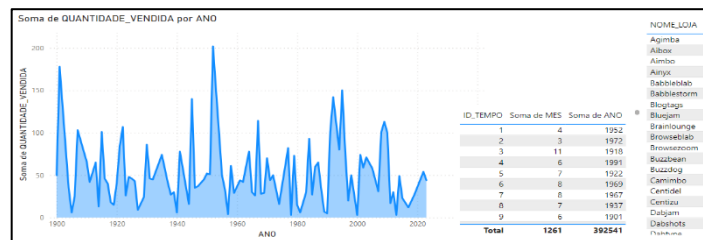


Figura 29

6.9.3. Dashboard - Desconto por Produto e Cidade

Esta dashboard, como mostra a Figura 29, centrado na análise dos descontos aplicados a produtos em diversas cidades, este painel permite filtrar dados por ano e promoção.

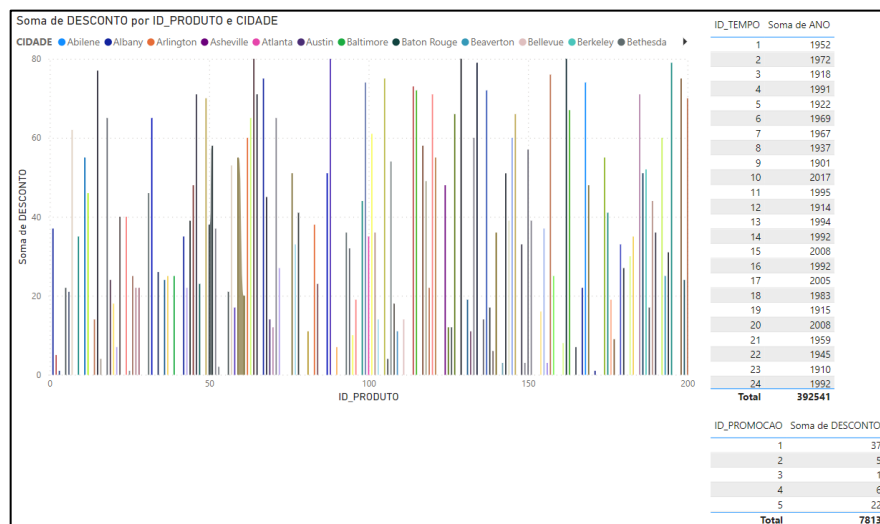


Figura 30

Isso fornece insights valiosos sobre a eficácia das promoções em áreas geográficas específicas, facilitando a adaptação de estratégias de desconto com base no comportamento do consumidor em diferentes localidades e períodos específicos.

Ideal para equipas de marketing e estrategistas de preços, este painel ajuda a otimizar a eficácia das campanhas promocionais.

6.9.4. Dashboard – Quantidade Vendida por Loja e Ano

Esta dashboard, mostrada na Figura 30, foca na quantidade de produtos vendidos em lojas específicas ao longo dos anos, com a capacidade de filtrar por ano, nome da loja e promoção.

Esta perspectiva permite uma análise aprofundada do desempenho de vendas em lojas individuais, possibilitando a identificação de padrões sazonais, correlações entre promoções e vendas, e a formulação de estratégias específicas para impulsionar o desempenho em lojas particulares em momentos específicos. É uma ferramenta valiosa para gerentes de lojas e analistas de vendas.

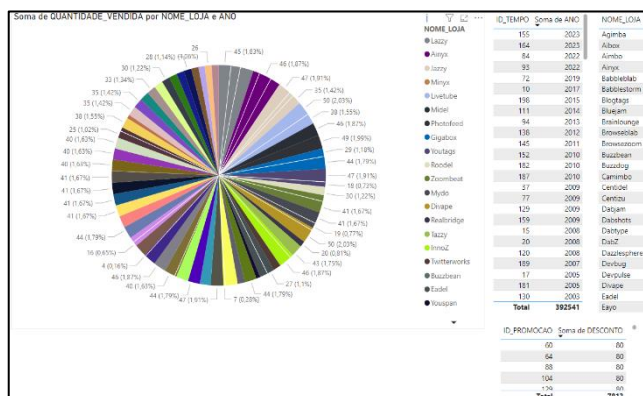


Figura 31

6.9.5. Dashboard – Soma do Valor Total Faturado em Vendas

Esta dashboard, conforme ilustrada pela Figura 31, destaca a agregação do valor total faturado por loja durante um mês específico.

A sua operacionalidade permite uma filtragem minuciosa, facultando a seleção da loja e do mês desejados. Além da visualização principal da soma total, são mostrados o valor médio e máximo do faturamento, proporcionando uma compreensão abrangente do desempenho financeiro da loja no período escolhido.

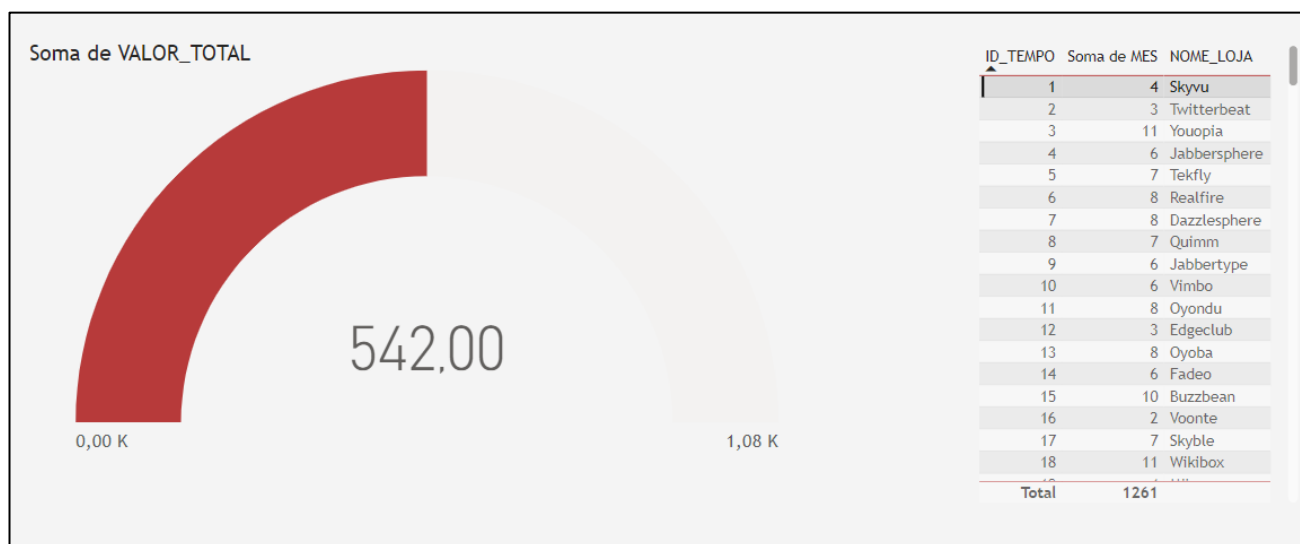


Figura 32

7. Conclusões

A jornada para compreender e implementar um Data Warehouse e ferramentas de Business Intelligence (BI) para análise de dados é vasta e complexa, mas fundamental para organizações que buscam transformar dados em decisões estratégicas informadas. Ao longo deste trabalho, exploramos o conceito e a aplicação prática de Data Warehousing, com uma ênfase particular na criação de um modelo em estrela no SQL Server e na visualização de dados através do Grafana e do Power BI.

A partir da contextualização e revisão do estado da arte, estabelecemos uma base teórica sólida que informou as etapas subsequentes da implementação prática. A ferramenta Grafana revelou-se uma plataforma poderosa para a visualização de dados, permitindo-nos criar dashboards intuitivos e detalhados que destacam métricas chave como gêneros de filmes mais populares, receitas por cinema, e a demografia dos clientes.

A criação de uma Data Warehouse foi abordada metodicamente, desde a seleção do processo de negócio a modelar até a avaliação da granularidade necessária para análises significativas. As dimensões e factos foram cuidadosamente escolhidos para refletir as necessidades analíticas do negócio, culminando na construção de um modelo em estrela que serve como a espinha dorsal para o armazenamento e recuperação de dados.

Os cálculos do tamanho aproximado do modelo garantiram que a implementação fosse realizada com eficiência de armazenamento em mente, enquanto as simulações de operações de ETL reforçaram a aplicabilidade do modelo no mundo real. As vistas desenvolvidas no Visual Studio Data Warehouse forneceram insights adicionais, alinhando os dados armazenados com as necessidades informacionais de um gestor de topo.

Comparativamente, a utilização do Power BI demonstrou como diferentes ferramentas de BI podem ser empregues para alcançar objetivos semelhantes em termos de visualização de dados e suporte à decisão. Cada ferramenta tem seus pontos fortes e contextos preferenciais de uso, e a escolha entre elas pode depender de uma variedade de fatores, incluindo a familiaridade do usuário, a complexidade dos dados e as necessidades específicas da análise.

Em suma, este trabalho não só forneceu uma compreensão profunda dos componentes técnicos e teóricos de Data Warehousing e BI, mas também destacou a importância de apresentar dados de uma forma que seja intuitiva e acionável para os decisores. O conhecimento adquirido e as habilidades desenvolvidas ao longo deste estudo são inestimáveis para a aplicação futura em iniciativas de dados em qualquer setor.

8. Referências

Referências consultadas pela equipa, incluindo o uso do ChatGPT para assistência na elaboração deste relatório:

- Livros e Textos Acadêmicos sobre Business Intelligence e Data Warehousing:
 - Business Intelligence e Analytics: Exploração do impacto do Big Data e Analytics nas empresas (Chen et al., 2012; Davenport, 2013; Wixom et al., 2010; Hopkins, 2011).
 - Big Data e seu Impacto: Discussões sobre como o Big Data está transformando negócios e sociedade (Khan, 2017; Mayer-Schönberger & Cukier, 2013; Brown et al., 2011; Turner et al., 2014).
 - Data Science e Data Warehousing: Importância crescente do papel do cientista de dados (Davenport & Patil, 2012) e fundamentos do Data Warehousing (Inmon, 2005; Kimball & Ross, 2011; Devlin, 2009).
 - Análise e Modelagem de Dados: Técnicas e conceitos em data mining e modelagem (Han et al., 2011; Vassiliadis et al., 2002).
 - Tecnologias Emergentes e Tendências: Exploração de tendências tecnológicas emergentes, incluindo cloud computing e Internet das Coisas (IoT) (Brynjolfsson & McAfee, 2014; Bughin et al., 2010; Marz & Warren, 2015).
 - Business Analytics: Estratégias para maximizar o valor da análise de negócios (Davenport, 2006; Howson & Idoine, 2017).
- Fichas Práticas da cadeira Sistema de Informação II, 2023/24;
- PDFs Teóricos da cadeira Sistema de Informação II, 2023/24;
- Documentação Oficial e Recursos Online do Grafana:
 - Fonte: Site Oficial do Grafana
 - URL: <https://grafana.com/>