

Aluno: Patrick Soares Mendes

MBA em Big Data e Machine Learning

Unidade Curricular 16 - Ferramentas de Business Intelligence

DISPONIBILIDADE DOS ARQUIVOS: Devido a limitação de tamanho de arquivo da Ferramenta EAD, o projeto com dashboard + base de dados foi inserida em meu repositório no GitHub:

[psoares07/Projetos-Analise-de-Dados: Repositório dedicado a armazenar, documentar e acompanhar os projetos criados.](#)

Situação de Aprendizagem – Projeto de Análise de dados

Descrição da atividade

Objetivo:

Nesta atividade, você deverá conceber e desenvolver um projeto completo de análise de dados, aplicando os conceitos estudados ao longo da disciplina. O desafio consiste em transformar dados brutos em insights estratégicos por meio de um dashboard interativo e informativo, utilizando técnicas de ETL, modelagem de dados, indicadores (KPI's) e storytelling.

Atividade

Utilizando a **base de dados fictícia de uma empresa de delivery**, siga os passos abaixo para estruturar seu projeto:

1. Estudo dos Dados:

- Avalie a estrutura das tabelas disponíveis e identifique as correlações entre elas.
- Identifique **tabelas fato e dimensão** para a correta modelagem dos dados.

2. Planejamento:

- Defina os principais **indicadores e KPI's** que serão apresentados no dashboard.
- Escolha os tipos de gráficos e visualizações que melhor representam os dados.

3. Processo de ETL:

- Implemente a **extração, transformação e carga (ETL)** dos dados, garantindo sua limpeza, validação e integridade para análise.

4. Modelagem de Dados:

- Crie os relacionamentos entre tabelas de forma adequada.
- Desenvolva cálculos e medidas para compor a análise dos indicadores.

5. Composição Visual e Storytelling:

- Estruture o **dashboard**, organizando gráficos, filtros e informações de forma clara e intuitiva.
- Certifique-se de que a apresentação dos dados "**conte uma história**", permitindo ao observador compreender a análise de forma objetiva.

Critérios de Avaliação:

- **Qualidade da Análise de Dados (10 pontos)** → Estudo detalhado dos dados e correta identificação de tabelas e relacionamentos.
- **Execução do Processo de ETL (10 pontos)** → Implementação eficiente da extração, transformação e carga de dados.
- **Modelagem e Construção de Indicadores (10 pontos)** → Definição correta de KPI's, medidas e cálculos.
- **Visualização e Storytelling (10 pontos)** → Estruturação do dashboard, clareza das informações, escolha de gráficos adequados e harmonia visual.

Entrega:

- **Formato:** Arquivo do projeto salvo no formato adequado à ferramenta utilizada + relatório explicativo em PDF contendo os principais insights e justificativas para as escolhas feitas.
- **Envio:** Plataforma do curso.

Projeto – Dashboard de Performance Delivery Center

Estudo dos dados

O projeto foi elaborado a partir do *dataset* [Delivery Center: Food & Goods orders in Brazil](#), disponível no Kaggle.

O *dataset* é composto por 370.000 mil pedidos. As tabelas podem ser encontradas na tabela 1 abaixo:

Tabelas	Descrição
Channels	Este dataset possui informações sobre os canais de venda (marketplaces) onde são vendidos os good e food de nossos lojistas.
Deliveries	Este dataset possui informações sobre as entregas realizadas por nossos entregadores parceiros.
Drivers	Este dataset possui informações sobre os entregadores parceiros. Eles ficam em nossos hubs e toda vez que um pedido é processado, são eles fazem as entregas na casa dos consumidores.
Hubs	Este dataset possui informações sobre os hubs do Delivery Center. Entenda que os Hubs são os centros de distribuição dos pedidos e é dali que saem as entregas.
Orders	Este dataset possui informações sobre as vendas processadas através da plataforma do Delivery Center.
Payments	Este dataset possui informações sobre os pagamentos realizados ao Delivery Center.
Stores	Este dataset possui informações sobre os lojistas. Eles utilizam a Plataforma do Delivery Center para vender seus itens (good e/ou food) nos marketplaces.

Tabela 1 – *Features* do dataset, retirado da [página no Kaggle](#)

São classificadas a tabela **Orders** como tabela fato, e as demais como dimensão. Pode-se ver na imagem abaixo:

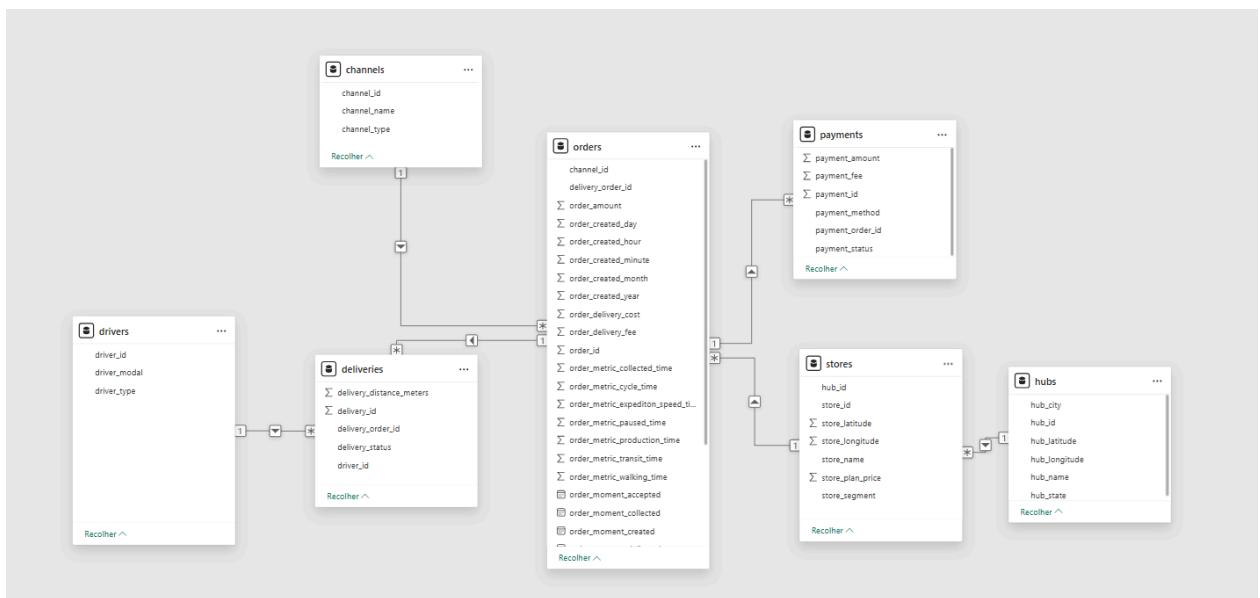


Figura 1 – Composição das tabelas do Dataset.

Identificou-se que a base opera sob uma lógica relacional, onde transações centrais se conectam a diversas entidades de cadastro através de chaves estrangeiras (*ids*).

Podemos classificar a modelagem da seguinte forma:

Tabela Fato (F): A tabela *orders* foi eleita como a tabela fato principal, pois contém os eventos granulares, métricas quantitativas e chaves de ligação. As tabelas *payments* e *deliveries* atuam como fatos auxiliares ou extensões da fato principal.

Tabelas Dimensão (D): As tabelas *stores* (lojas), *hubs* (centros de distribuição), *channels* (canais de venda) e *drivers* (entregadores) foram

classificadas como dimensões, pois fornecem o contexto descritivo (quem, onde, como) para as transações

Planejamento

Definiu-se o escopo do *dashboard* focado em duas perspectivas: Estratégica (Financeira) e Tática (Operacional).

- Definição das KPIs:
 - Volume: Faturamento Bruto, Quantidade de Pedidos Válidos.
 - Eficiência: Ticket Médio, Taxa de Cancelamento.
 - Logística: Tempo Médio de Ciclo (Segmentado obrigatoriamente entre *Food* e *Good* devido à disparidade de natureza).
- Escolha das Visualizações:
 - Cartões (Big Numbers): Para destacar os números macro no topo.
 - Gráfico de Área: Para demonstrar tendência temporal e crescimento.
 - Gráfico de Barras: Para análise de Pareto (Ranking de Hubs).
 - Scatter Plot (Dispersão): Para correlacionar Distância vs. Tempo de Entrega.
 - Matriz Hierárquica: Para detalhamento granular (Drill-down) com alertas visuais.

Processo de ETL (Extração, Transformação e Carga)

A etapa de tratamento no Power Query foi crítica para garantir a integridade dos dados, corrigindo falhas de importação e ruídos estatísticos.

- Correção de Localidade (Regional Settings): As colunas financeiras (*order_amount*) e de tempo (*order_metric_cycle_time*) utilizavam o padrão americano (ponto decimal). A leitura automática multiplicava os valores incorretamente (ex: 62.7 virava 627). Portanto, foi realizada a conversão manual substituindo ponto por vírgula antes da tipagem decimal.
- **Outliers:** Identificou-se o registro "*HUBLESS SHOPPING*" com métricas irrealis (Distância > 270km e Tempo > 9.600 min). Portanto, foi realizado a exclusão desse registro e gerado um alerta para a gestão.
- **Padronização:** Conversão de colunas de *timestamp* para o formato de Data/Hora permitindo a criação correta do relacionamento com o calendário.

Modelagem de Dados

A modelagem seguiu o esquema Star Schema (Estrela) para otimizar a performance e facilitar a criação de medidas DAX.

- Foram estabelecidos relacionamentos Um-para-Muitos (1:*) entre as chaves primárias das Dimensões e as chaves estrangeiras de Fato *orders*
- Criação da tabela *dCalendario* via DAX (*CALENDARAUTO*) para garantir continuidade temporal e permitir análises no tempo, com colunas de ordenação customizada para meses.
- Os Cálculos e Medidas foram organizados em um grupo chamado *Medidas*. As principais funções utilizadas foram:
 - *Calculate*: Garantindo medidas aplicando filtros explícitos (*order_status = "FINISHED"*). O objetivo de utilizar essa função é para obter o faturamento total somente dos pedidos finalizados.
 - *All*: Utilizado para calcular a Distribuição de Vendas (*Market Share*).
 - *DIVIDE*: Utilizado para evitar erros de divisão por zero, como pode ocorrer no cálculo do *Ticket Médio* e Taxas.

Composição Visual e *Storytelling*

A estrutura do dashboard pode ser dividida em duas seções (narrativas):

Página 1 - Visão Executiva (O "Quê"): Apresenta a leitura macro da saúde financeira. A leitura segue o padrão "Z", iniciando pelos KPIs macro, passando pela concentração de receita por Hub (Pareto), a dependência de canais (Risco) e, por fim, a tendência de faturamento mensal.

Página 2 - Visão Operacional (O “Porquê”): Identificação de gargalos. Destaque central é o Gráfico de Dispersão (Distância x Tempo), que comprova a baixa correlação entre distância e atrasos. Logo abaixo, uma Matriz Hierárquica que exibe detalhadamente cada Hub e Loja, evidenciando sua situação de acordo com a formatação condicional: Alerta Vermelho para entregas acima de 60 minutos.

Insights

Alguns insights rápidos podem ser obtidos:

1. A análise de dispersão (Gráfico de Dispersão) revela um ponto de atenção crítico no Golden Shopping: apesar de ser o líder em faturamento, ele situa-se no quadrante de gargalo, com tempos de ciclo acima da média, indicando que sua operação interna não escala na mesma proporção da demanda. Em contrapartida, o Fortran Shopping estabelece-se como *benchmark* operacional, mantendo agilidade mesmo em entregas de longa distância. Nota-se ainda uma alta variabilidade nos tempos de OWN CHANNEL (Canal Próprio), sugerindo

que, ao contrário dos *Marketplaces* padronizados, a logística própria da empresa carece de processos mais robustos.

2. O gráfico de composição de receita evidencia que os *Marketplaces* dominam o faturamento (aprox. 79%), enquanto o OWN CHANNEL representa uma fatia minoritária, isto é um risco crítico. Embora os *marketplaces* garantam volume e aquisição de clientes, essa dependência diminui as margens de lucro devido às taxas de comissão relacionadas à plataforma utilizada, além de deixar a empresa vulnerável a mudanças de algoritmo. O insight açãoável é a urgência de campanhas para migrar clientes fiéis para o app próprio, visando maior rentabilidade.
3. A análise temporal demonstra que o mês de **março** não foi apenas um pico isolado, mas o momento de consolidação da base de clientes no período analisado, superando significativamente os meses de janeiro e fevereiro. Esse desempenho estabelece um novo patamar de faturamento para a empresa. O próximo passo agora é analisar se esse crescimento foi orgânico ou impulsionado por fatores externos (sazonalidade), utilizando esse patamar como meta mínima para o planejamento do próximo trimestre.