



PONTIFÍCIA UNIVERSIDADE CATÓLICA DE MINAS GERAIS

NÚCLEO DE EDUCAÇÃO A DISTÂNCIA

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

RELATÓRIO TÉCNICO

DESENVOLVIMENTO DE PRODUTOS BASEADO EM ANÁLISE DE DADOS DE
MERCADO – BARRA DE CHOCOLATE

Marco Antônio de Oliveira Júnior

Belo Horizonte

2022

SUMÁRIO

- 1. Introdução4**
 - 1.1. Contexto4
 - 1.2. Objetivos.....6
 - 1.3. Público-alvo7
- 2. Modelo de Dados8**
 - 2.1. Modelo Dimensional8
 - 2.2. Fatos e Dimensões9
 - 2.3. Melhorias Futuras10
- 3. Integração, Tratamento e Carga de Dados11**
 - 3.1. Fontes de Dados.....11
 - 3.2. Processos de Integração e Carga (ETL).....12
- 6. Links13**
- REFERÊNCIAS14**

1. Introdução

1.1. Contexto

A criação de um novo produto é uma etapa do ciclo de vida de uma empresa que envolve diversos processos, sendo uma atividade de extrema importância para o bom desenvolvimento e sucesso dos negócios. Nesta atividade, diversas áreas da empresa são envolvidas, tais como engenharia, qualidade, comercial, financeiro, marketing, produção, logística, dentre outros, onde a participação, comunicação e interação entre elas se torna vital para que o produto seja bem-sucedido e alcance os objetivos e metas do negócio.



FIGURA 01 – ÁREAS DA EMPRESA ENVOLVIDAS NO DESENVOLVIMENTO DE PRODUTOS (ACERVO PESSOAL DO AUTOR)

A interação entre as áreas da empresa depende fundamentalmente da natureza do negócio e de seu mercado (Dierkes, 2003 *apud* FINEP, 2005). Podemos observar como exemplo o setor de *commodities*, onde a sua diferenciação competitiva tende a se concentrar nos custos de produção, haja vista preços e demandas já pré-estabelecidas pelo mercado. Nisto, a viabilidade e sucesso dos produtos dependerá de uma maneira mais expressiva de fatores ligados às áreas de produção, comercial e financeiro.

Tendo em vista a participação de diversas áreas no desenvolvimento de um produto, que contribuem com aspectos técnicos, estratégicos e comerciais,

usualmente esta atividade é dividida em fases conforme ilustrado na **Figura 02**. Estas fases podem ser estruturadas pelo processo de “*gates de decisão*”, onde, ao fim de cada fase, é realizada, com base nos resultados obtidos até então, uma rodada de tomada de decisões para verificar se o produto é viável e deve prosseguir para a próxima fase, ou se é inviável e o projeto deve ser finalizado.

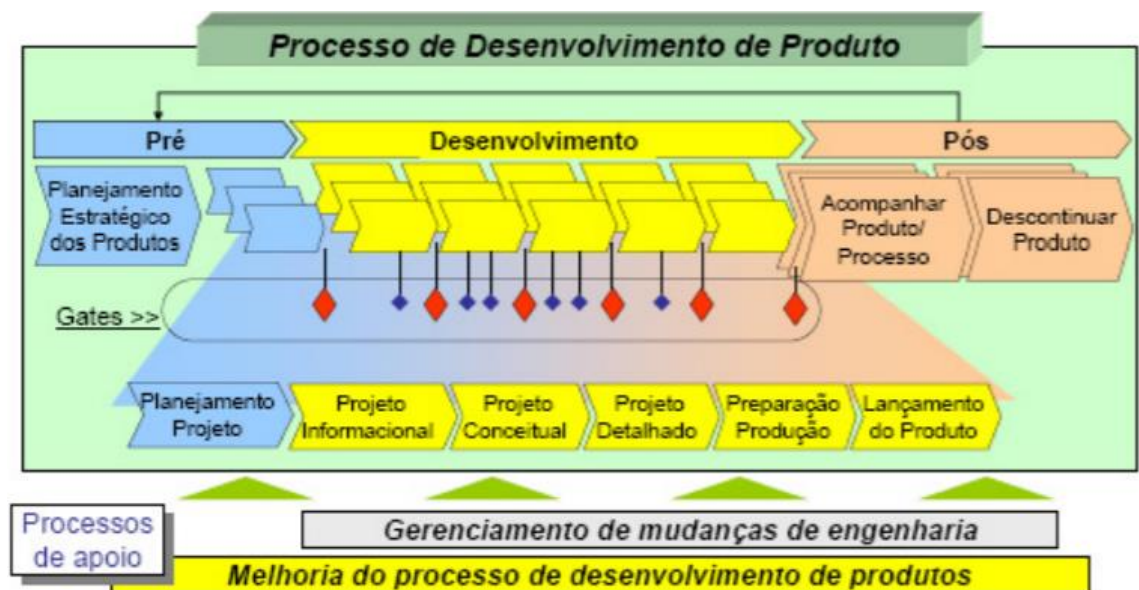


FIGURA 02 – PROCESSO DO DESENVOLVIMENTO DE PRODUTOS (ROZENFELD ET. AL., 2006)

Dentre as fases do desenvolvimento de produto, na fase inicial de pré-desenvolvimento, usualmente são realizadas pesquisas de mercado com o intuito de abastecer o planejamento estratégico a partir do levantamento de dados direcionados para definição das características do produto, tais como: tendências, aceitação, demandas, preferências, hábitos de consumo, público-alvo, entre outros (Faria *et. al.*, 2008). Neste sentido, a aplicação de técnicas de *Business Intelligence* (BI) e *Business Analytics* (BA) para coleta, armazenamento, tratamento, visualização e proteção dos dados das pesquisas de mercado se torna um poderoso aliado para fomentar as tomadas de decisão estratégicas em consonância com os objetivos e metas organizacionais, maximizando a agregação de valor do produto. A partir do correto manejo dos dados de pesquisa de mercado, é possível transformá-los de uma maneira mais precisa e confiável em conhecimentos produtivos que irão gerar ações e análises prescritivas, sugerindo estratégias mais eficientes para que os envolvidos na criação do produto possam tomar decisões.

1.2. Objetivos

O objetivo deste trabalho é desenvolver uma solução de BI conforme ilustrado na **Figura 03** para aplicação em dados coletados em pesquisas de mercado para apoio à fase de pré-desenvolvimento de produtos. Nisto, serão desenvolvidos processos de modelagem de dados, ETL (*Extract, Transform, Load*), visualização de dados, *machine learning* e homologação de dados com o intuito de se estruturar artefatos que possam ser utilizados por organizações para apoio à fase de pré-desenvolvimento de produtos, principalmente para tomadas de decisão estratégicas.

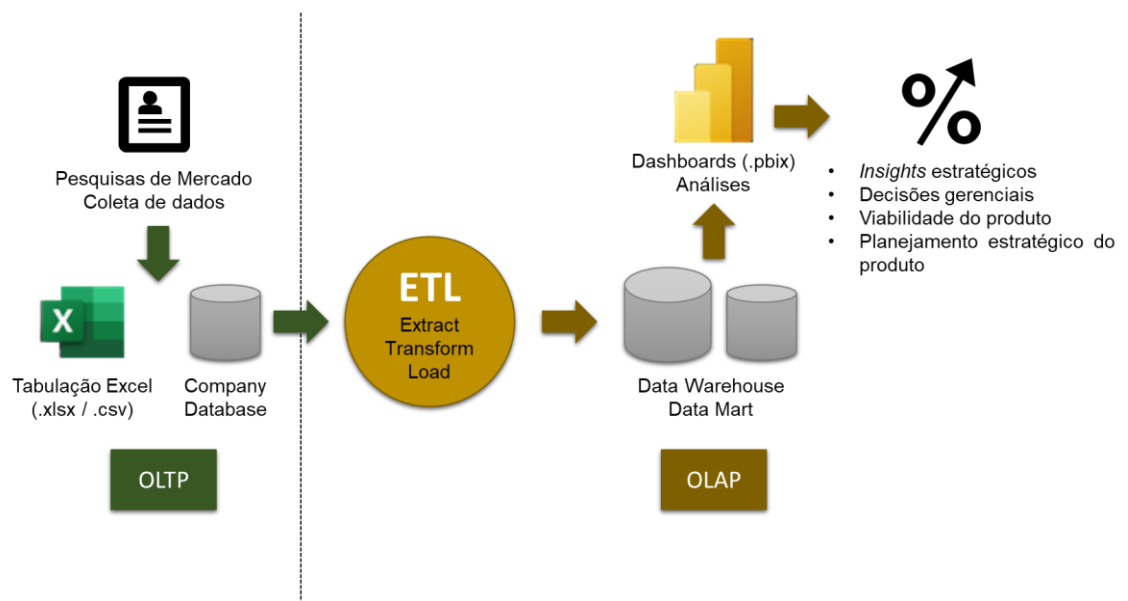


FIGURA 03 – ELEMENTOS DA SOLUÇÃO DE BI (ACERVO PESSOAL DO AUTOR)

Para atingir ao objetivo, este trabalho irá apresentar um estudo de caso criado a partir de uma empresa de chocolate fictícia (Empresa X), que se encontra na fase de pré-desenvolvimento de uma nova barra de chocolate. O chocolate é um produto fabricado com base em sementes de cacau fermentadas e torradas, sendo altamente consumido pela nossa sociedade. No Brasil, o consumo chega a até 2 quilos por ano por pessoa, sendo consumido por aproximadamente 75% da população, onde 56% são mulheres e 35% não o trocariam por outro alimento.

Para apoio ao desenvolvimento de seu produto, a Empresa X procedeu com uma pesquisa de mercado, coletando dados tais como avaliação de chocolates já pré-existent no mercado considerando as diversas sementes de cacau que o compõe, resultando em uma análise qualitativa considerando *ratings*, percentual de cacau, ingredientes e sabores. A partir desta pesquisa de mercado, a Empresa X espera

estruturar dados que respondam perguntas para direcionar o desenvolvimento do seu produto.

Com os resultados da estruturação dos dados, a Empresa X espera obter melhores parâmetros para desenvolver um produto que seja mais bem aceito pelo público e possa obter maior sucesso comercial em consonâncias com suas metas e objetivos.

Datasets extraído do Kaggle, acesso em 22/01/2022:

<https://www.kaggle.com/andrewmvd/chocolate-ratings>

1.3. Público-alvo

O produto objeto deste trabalho visa fomentar as tomadas de decisão em nível estratégico para o desenvolvimento de produtos. Neste sentido, o público-alvo seriam os colaboradores da Empresa X envolvidos na fase de pré-desenvolvimento do produto.

A coleta dos dados é feita a partir de pesquisas de mercado, usualmente realizadas a nível operacional pela área de Marketing das empresas. Assim, os processos de modelagem de dados, ETL, visualização de dados, *machine learning* e homologação de dados desenvolvidos neste trabalho poderão apoiar os colaboradores desta área na transformação dos dados e realização de análises de uma maneira mais rápida, precisa, confiável e segura, gerando relatórios dinâmicos para suporte às análises gerenciais.

Após a transformação dos dados e realização das análises, com os relatórios dinâmicos concluídos, os usuários finais serão os colaboradores da Empresa X a nível gerencial, que utilizarão estes produtos para tomada de decisões estratégicas quanto às características do novo produto. O objetivo é que as análises possam responder perguntas que direcionem as tomadas de decisão estratégicas de uma maneira mais precisa e proativa para obtenção de um produto que esteja alinhado com os objetivos da companhia, sendo mais bem aceito pelo mercado e contribuindo para o atingimento das metas da empresa.

2. Modelo de Dados

2.1. Modelo Dimensional

Foram observados os processos de negócio da Empresa X com o intuito de se estruturar um modelo dimensional que estivesse alinhado com os objetivos do projeto: responder perguntas direcionadas para definição das características de um novo produto de maneira alinhada com os objetivos e metas do negócio, tais como:

- Quais os fabricantes mais bem avaliados pelos usuários?
- Quais os sabores mais bem avaliados pelos usuários?
- Quais ingredientes compõem os chocolates mais bem avaliados pelos usuários?
- Qual o percentual de cacau mais bem avaliado pelos usuários?
- Quais as melhores sementes de cacau?
- Existe relação entre a qualidade da semente de cacau e sua localidade?
- As preferências dos usuários variaram ao longo dos anos?

As pesquisas de mercado usualmente são realizadas a partir de questionários que serão respondidos pelo público-alvo, seja via internet ou por coletas de campo. Estes dados são coletados em formato .csv / .xlsx e armazenados no *database* da Empresa X. Assim, foram estruturadas métricas (fatos) organizadas com demais entradas (dimensões) de maneira alinhada com os processos da empresa para responder as perguntas objeto da análise.

Por exemplo, considerando que as pesquisas de mercado são preenchidas pelo público-alvo, as informações dos usuários poderiam vir a ser dados considerados para se levantar no modelo dimensional. Entretanto, considerando o objetivo do projeto, é possível concluir que tais dados não são necessários, fazendo com que a pesquisa de mercado possa ser conduzida de modo anônimo. Neste sentido, foram definidos os fatos e dimensões necessários conforme demonstrado no modelo dimensional da **Figura 4**.

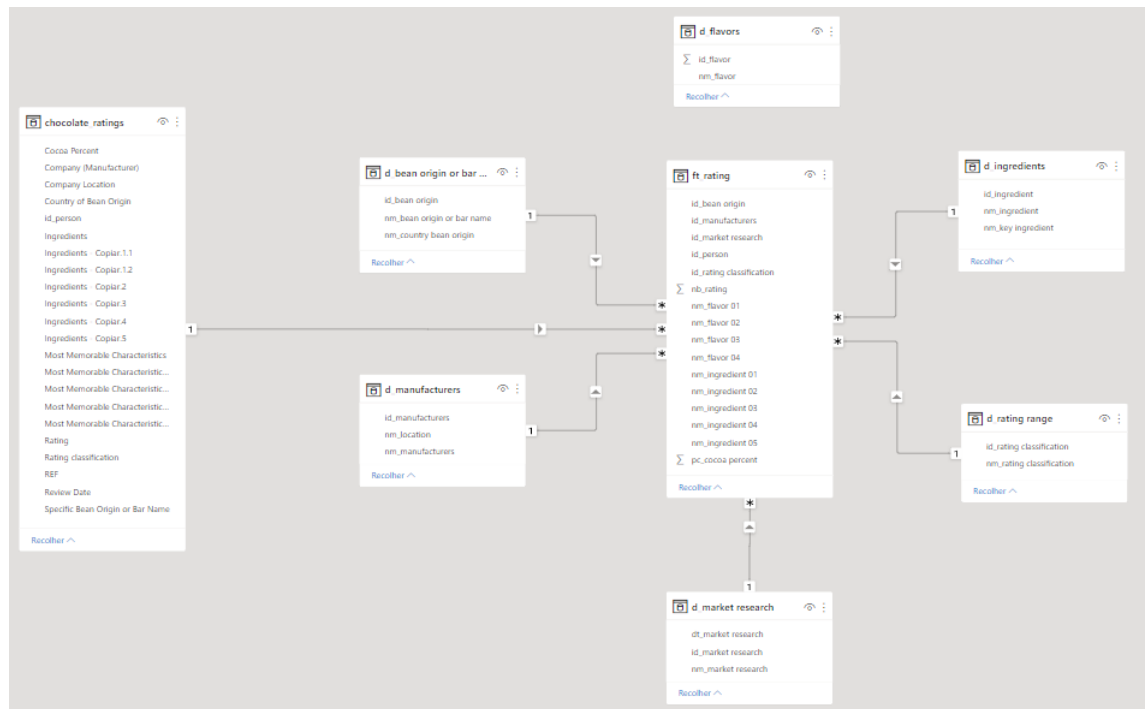


FIGURA 04 – DIAGRAMA DO MODELO DIMENSIONAL (ACERVO PESSOAL DO AUTOR)

O diagrama do modelo dimensional oferece uma visão sobre como os dados serão organizados para atender às necessidades identificadas por meio dos objetivos do projeto. Tendo em vista a simplicidade dos fatos e dimensões estabelecidas, bem como aspectos de performance do sistema, foi definido o modelo dimensional em esquema Estrela. Este modelo se mostra mais adequado considerando a abordagem relacional dos dados, bem como a simplificação da visualização dimensional, otimizando o sistema.

2.2. Fatos e Dimensões

A **Figura 05** descreve as tabelas de fato e as dimensões definidas para o projeto:

Dimensão: d_market research				
Coluna	Descrição	Tipo	Granularidade	Observações
id_market research	Chave sequencial gerada automaticamente	integer		
nm_market research	Nome da pesquisa de mercado realizada	integer	Usuário	Uma linha por usuário
dt_market research	Data da pesquisa de mercado	string	Ano	Dimensão com alta granulometria, considerar melhorias futuras

Dimensão: d_manufacturers				
Coluna	Descrição	Tipo	Granularidade	Observações
id_manufacturers	Chave sequencial gerada automaticamente	integer		
nm_manufacturers	Nome do fabricante do chocolate	string	Fabricante	Lista de fabricantes das barras de chocolate avaliadas
nm_location	Local do fabricante do chocolate	string	País	Dimensão com alta granulometria, considerar melhorias futuras

Dimensão: d_bean origin or bar name				
Coluna	Descrição	Tipo	Granularidade	Observações
id_bean origin	Chave sequencial gerada automaticamente	integer		
nm_bean origin or bar name	Local da semente ou barra de cacau	string	Nome ou País	Local da semente ou barra de cacau utilizada no chocolate avaliado. Granulometria variada, onde a mais alta é o País; considerar melhorias futuras.
nm_country bean origin	Local de origem da semente ou barra de cacau	string	País	País de origem da semente ou barra de cacau utilizada no chocolate avaliado

Dimensão: d_ingredients				
Coluna	Descrição	Tipo	Granularidade	Observações
id_ingredients	Chave sequencial gerada automaticamente	integer		
nm_key ingredient	Identificador do ingrediente	string	Nome	Abreviação dos principais ingredientes da barra de chocolate avaliada
nm_ingredient	Nome do ingrediente	string	Nome	Nome dos principais ingredientes da barra de chocolate avaliada

Dimensão: d_flavors				
Coluna	Descrição	Tipo	Granularidade	Observações
id_flavor	Chave sequencial gerada automaticamente	integer		
nm_flavor	Nome do sabor	string	Nome	Nome dos sabores mais marcantes da barra de chocolate avaliada

Dimensão: d_rating range				
Coluna	Descrição	Tipo	Granularidade	Observações
id_rating classification	Chave sequencial gerada automaticamente	integer		
nm_rating classification	Nome do range de classificação	string	Nota de 0 a 5	Ranges de classificação considerando um nível maior de granulometria para fomentar análises

Dimensão: ft_rating				
Coluna	Descrição	Tipo	Granularidade	Observações
id_person	Chave sequencial gerada automaticamente	integer		
nb_rating	Classificação da barra de chocolate	real	Nota de 0 a 5	Nota decimal de 0 a 5 com duas casa decimais de precisão da barra de chocolate avaliada
pc_cocoa percent	Percentual de cacau da barra de chocolate	real	Percentual	Percentual de cacau com duas casa decimais de precisão da barra de chocolate avaliada
nm_ingredient 01	Ingrediente 01 da barra de chocolate	string	Nome	
nm_ingredient 02	Ingrediente 02 da barra de chocolate	string	Nome	
nm_ingredient 03	Ingrediente 03 da barra de chocolate	string	Nome	
nm_ingredient 04	Ingrediente 04 da barra de chocolate	string	Nome	
nm_ingredient 05	Ingrediente 05 da barra de chocolate	string	Nome	
nm_flavor 01	Sabor marcante 01 da barra de chocolate	string	Nome	
nm_flavor 02	Sabor marcante 02 da barra de chocolate	string	Nome	
nm_flavor 03	Sabor marcante 03 da barra de chocolate	string	Nome	
nm_flavor 04	Sabor marcante 04 da barra de chocolate	string	Nome	
id_market research	Chave estrangeira	integer		
id_manufacturers	Chave estrangeira	integer		
id_bean origin	Chave estrangeira	integer		
id_rating classification	Chave estrangeira	integer		

FIGURA 05 – FATOS E DIMENSÕES (ACERVO PESSOAL DO AUTOR)

2.3. Melhorias Futuras

Considerando a base de dados extraída das pesquisas de mercado, é possível notar que algumas dimensões, como por exemplo a data da pesquisa, apresentam granulometria alta. Assim, constatada esta lacuna, é interessante que os profissionais envolvidos nos processos de *Business Intelligence* apresentem um *feedback* para os envolvidos nos processos transacionais de pesquisa de mercado para possíveis padronizações e melhorias na granulometria da coleta de dados. Este processo de

feedback e melhoria contínua é essencial para que os dados levantados possam estar cada vez mais alinhados com as necessidades das análises.

3. Integração, Tratamento e Carga de Dados

3.1. Fontes de Dados

A fonte de dados está ligada diretamente às pesquisas de mercado, que são coletadas pela Empresa X e tabuladas em formato .csv / .xlsx pela equipe de desenvolvimento ETL. É importante que os dados sejam tabulados pela equipe de desenvolvimento ETL, já que durante este processo são observados aspectos tais como integridade, limpeza e padronização dos dados.

A **Figura 04** supracitada no **item 2.1** ilustra o modelo dimensional adotado, bem como as bases de dados estruturadas durante o processo de ETL. A **Figura 06** ilustra a carga da base de dados original, em formato .csv / .xlsx, extraída a partir das pesquisas de mercado.

Configurações da fonte de dados

Gerenciar configurações para fontes de dados que você conectou usando o Power BI Desktop.

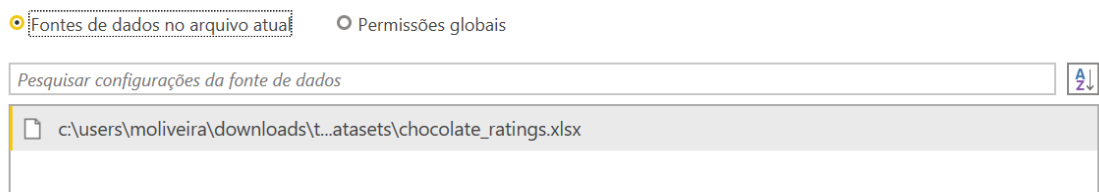


FIGURA 06 – FONTE DE DADOS CARREGADA NO POWER QUERY (ACERVO PESSOAL DO AUTOR)

É importante ressaltar que todo o processo de ETL foi estruturado de maneira a se automatizar o processamento da fonte de dados considerando a sua atualização a partir da realização de novas pesquisas de mercado. Neste sentido, basta a simples atualização do arquivo tabulado em formato .csv / .xlsx com acréscimo de novas linhas de dados, desde que mantida a estrutura original da tabela, para que todos os processos de ETL sejam executados novamente de maneira padronizada e automática, permitindo a visualização dos dados incluídos.

3.2. Processos de Integração e Carga (ETL)

O processo de ETL se inicia com a extração dos dados, feita através da carga da tabela em formato .csv / .xlsx no *Power BI*. Com os dados carregados, os responsáveis pelo ETL partem para a transformação dos dados, que é realizada no *software Microsoft Power BI* a partir do *Power Query*. No *Power Query*, foram estruturadas as seguintes etapas:

1. Tratamento dos dados carregados considerando decodificação de dados, conversão de atributos, agregação de valores, criação de chaves identificadoras, verificação da integridade dos dados e modificação da estrutura dos dados conforme necessidades de análise. As etapas aplicadas estão ilustradas na **Figura 07**;

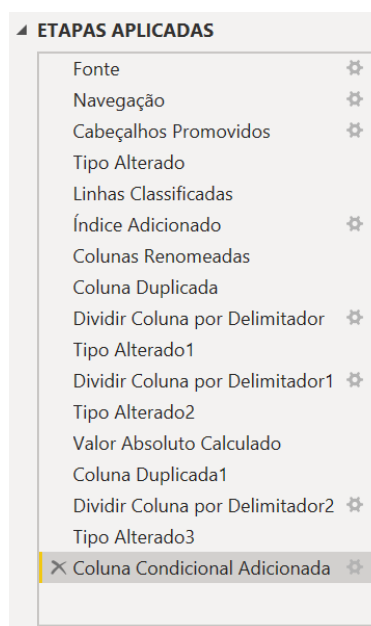


FIGURA 07 – PRIMEIRA ETAPA DE TRANSFORMAÇÃO DOS DADOS (ACERVO PESSOAL DO AUTOR)

2. Otimização da estrutura relacional a partir da criação de tabelas de dimensões com chaves identificadoras e da tabela fato a partir de chaves estrangeiras, dando origem ao modelo dimensional em esquema estrela conforme proposto e já ilustrado na **Figura 04** supracitada no **item 2.1**. A tabela fato foi estruturada a partir da mescla de consultas considerando as chaves identificadoras conforme **Figura 08**;

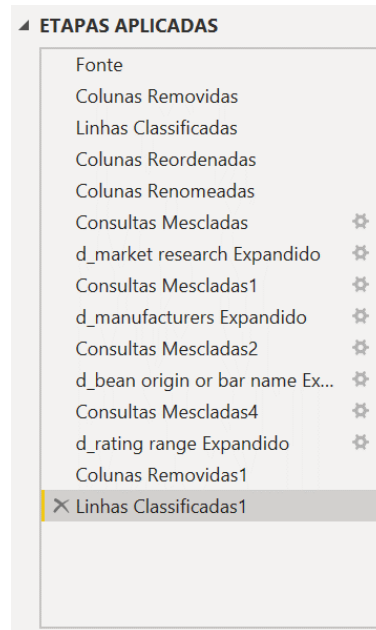


FIGURA 08 – SEGUNDA ETAPA DE TRANSFORMAÇÃO DOS DADOS (ACERVO PESSOAL DO AUTOR)

3. Conferência da integridade dos dados a partir da contagem do número de linhas da tabela fato em comparação com a fonte de dados original e verificação da integridade dos dados transformados considerando a existência de células vazias, despadronizadas etc. A verificação da contagem de linhas está evidenciada na **Figura 09**.

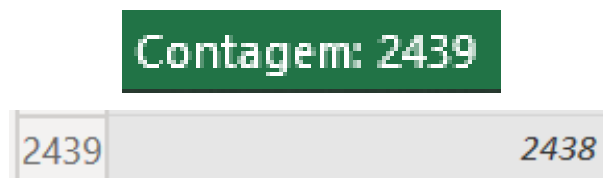


FIGURA 09 – TERCEIRA ETAPA DE TRANSFORMAÇÃO DOS DADOS (ACERVO PESSOAL DO AUTOR)

6. Links

1. Datasets extraído do Kaggle, acesso em 22/01/2022:

<https://www.kaggle.com/andrewmvd/chocolate-ratings>

2. Repositório GITHUB:

<https://github.com/marcoaojunior/TCC-BI-BA-PUC>

REFERÊNCIAS

FARIA, Adriana Ferreira et. al. **Processo de desenvolvimento de novos produtos: uma experiência didática**. Rio de Janeiro: XXVIII Encontro Nacional de Engenharia de Produção, p. 1-14, 2008.

FINEP. **Manual de Oslo – Diretrizes para coleta e interpretação de dados sobre inovação**. Rio de Janeiro: ARTI/FINEP, 3 ed, 2005.

ROZENFELD, Henrique et. al. **Gestão de Desenvolvimento de Produtos: uma referência para a melhoria do processo**. São Paulo: Saraiva, 2006.