



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. Introdução	4
1.1. Contexto	4
1.2. Objetivos	6
1.3. Público-alvo	7
2. Modelo de Dados	8
2.1. Modelo Dimensional	8
2.2. Fatos e Dimensões	9
2.3. Melhorias Futuras	11
3. Integração, Tratamento e Carga de Dados	11
3.1. Fontes de Dados.....	11
3.2. Processos de Integração e Carga (ETL).....	13
4. Camada de Apresentação	15
4.1 Dashboard	15
4.2 Análises avançadas	23
5. Registros de Homologação	28
6. Conclusões	29
6.1 Proposta de Intervenção.....	31
6.2 Lições Aprendidas	32
7. Links	33
REFERÊNCIAS	34

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.

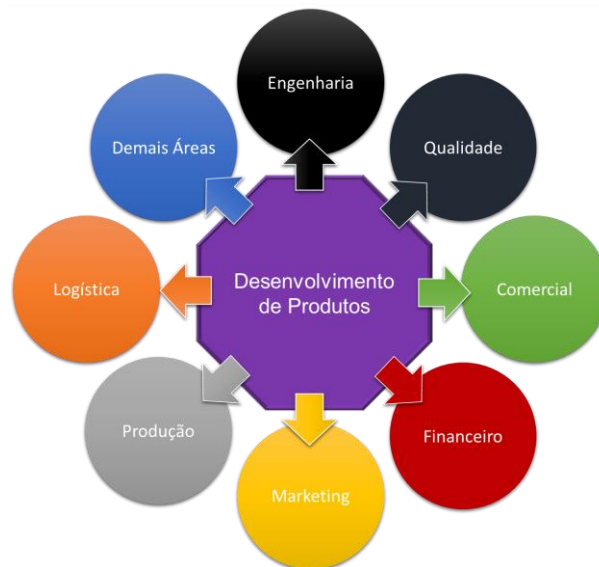


FIGURA 01 – ÁREAS DA EMPRESA ENVOLVIDAS NO DESENVOLVIMENTO DE PRODUTOS (ACERVO PESSOAL DO AUTOR)

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.

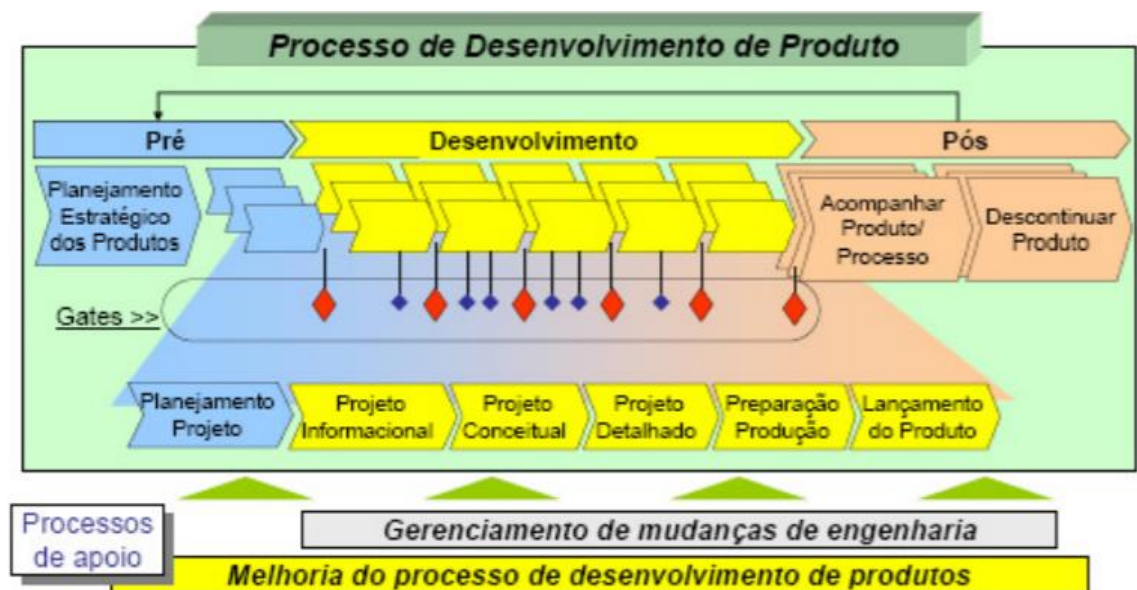


FIGURA 02 – PROCESSO DO DESENVOLVIMENTO DE PRODUTOS (ROZENFELD ET. AL., 2006)

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.

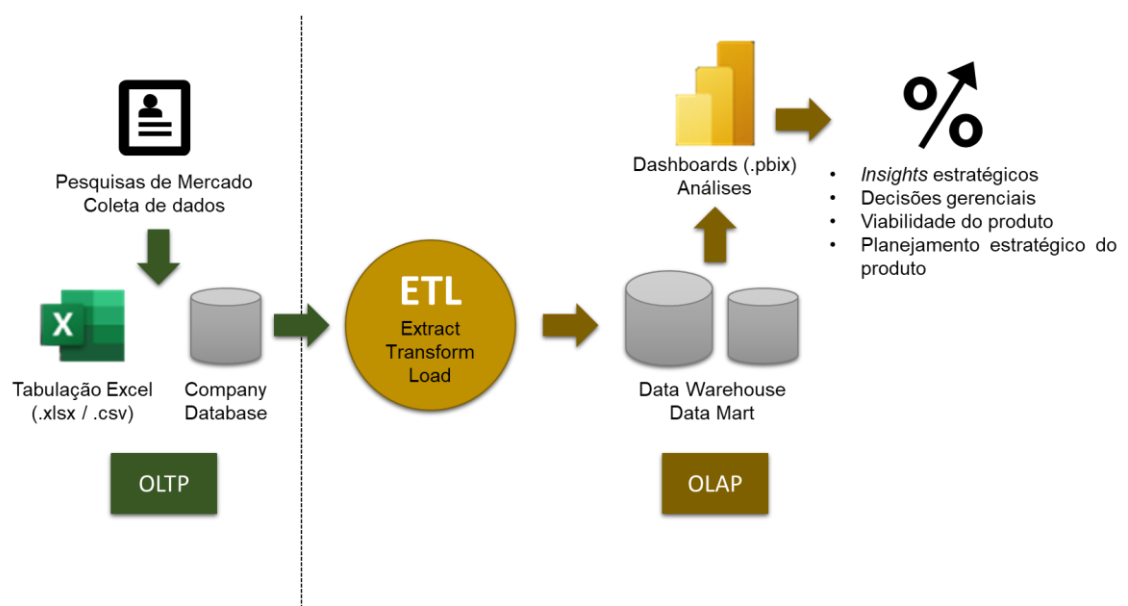


FIGURA 03 – ELEMENTOS DA SOLUÇÃO DE BI (ACERVO PESSOAL DO AUTOR)

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 pesquisas de mercado, coletando dados ao longo dos anos tais como avaliações de chocolates já pré-existent no mercado por usuários considerando as diversas sementes de cacau que o compõe, resultando em uma análise qualitativa considerando *ratings*, percentual de cacau, ingredientes e sabores. A partir destas

pesquisas 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 organizacionais.

Datasets extraídos 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 respondidas pelos usuários de chocolate, 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 – o público-alvo a nível operacional e tático - 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, o público-alvo a nível estratégico serão os colaboradores da Empresa X de 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 e objetivos organizacionais 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 deste trabalho: responder perguntas direcionadas para definição das características de um novo produto de maneira alinhada com os objetivos e metas do negócio, tais como:

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

As pesquisas de mercado usualmente são realizadas a partir de questionários respondidos pelos usuários de chocolates ao longo dos anos, seja via internet ou por coletas de campo - para este trabalho, foi utilizado um *dataset* fictício extraído do KAGGLE. 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 pelos usuários de chocolate, as informações pessoais dos usuários poderiam vir a ser dados considerados para se levantar no modelo dimensional. Entretanto, considerando o objetivo proposto para a análise, 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**.

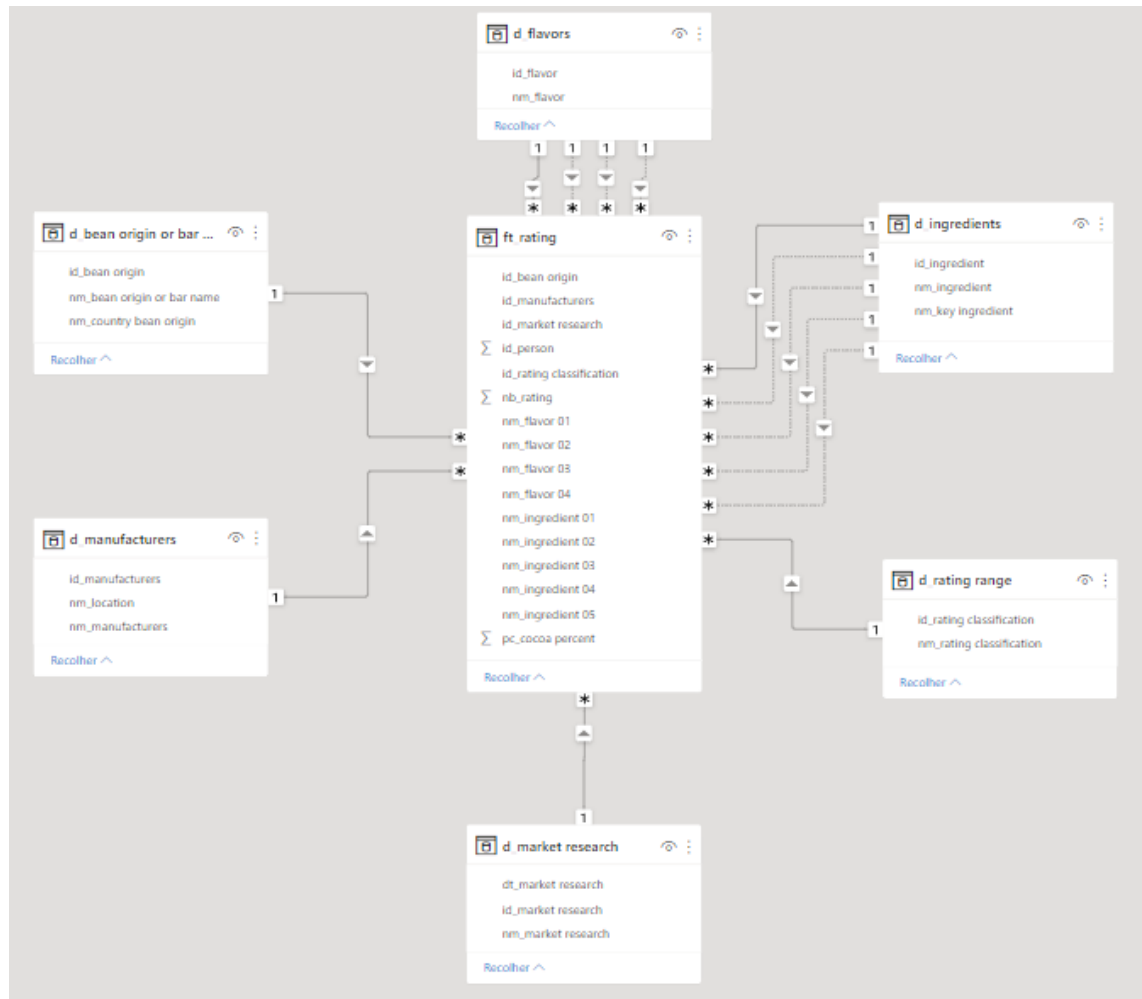


FIGURA 04 – DIAGRAMA DO MODELO DIMENSIONAL (ACERVO PESSOAL DO AUTOR)

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
vl_minimun range	Valor mínimo do range	integer	Nota de 0 a 5	
vl_maximum range	Valor maximo do range	integer	Nota de 0 a 5	
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	Ingredientes concatenados da barra de chocolate	string	Nome	
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	Sabores marcantes concatenados 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. Abaixo, podemos observar as possíveis melhorias já identificadas:

- **dt_market research:** aumentar a granulometria para dia, mês e ano visando melhor análise temporal.
- **nm_location:** aumentar a granulometria para cidade, estado e país visando melhor análise geográfica.
- **nm_bean origin or bar name:** segregar as informações “barra de cacau” de “local da semente” e aumentar a granulometria para cidade, estado e país visando melhor análise geográfica.
- **nm_country bean origin:** aumentar a granulometria para cidade, estado e país visando melhor análise geográfica.
- **nm_rating classification:** aumentar a granulometria para um *range* de 0,5 pontos visando análises mais profundas.

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 de maneira fictícia pela Empresa X e tabuladas em formato .csv / .xlsx pela equipe de desenvolvimento ETL – originalmente, o *dataset* foi extraído do KAGGLE. 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 06** ilustra a fonte de

dados original, que apresenta os dados das pesquisas de mercados tabulados em formato .csv / .xlsx contendo dados de rótulos da pesquisa de mercado, das barras de chocolate avaliadas e de suas respectivas sementes / barras de cacau e fabricantes, e, por fim, da avaliação feita pelos usuários de chocolate, contendo o *rating* e os sabores marcantes. A tabela apresenta mais de 2000 linhas, com dados distribuídos entre os anos de 2006 e 2021.

REF	Company (Manufacturer)	Company Location	Review Date	Country of Bean Origin	Specific Bean Origin or Bar Name	Cocoa Percent	Ingredients	Most Memorable Characteristic	Rating
5	Jacque Torres	U.S.A.	2006	Ghana	Trinitario Treasure	71%	5- B,S,C,V,L	gritty, unrefined, off notes	2,0
15	Green & Black's (ICAM)	U.K.	2006	Blend	Dark	70%	5- B,S,C,V,L	mild rich, basic, roasty	2,5
15	Guitard	U.S.A.	2006	Colombia	Chucuri	65%	5- B,S,C,V,L	cream, sweet, floral, vanilla	3,0
15	Neuhaus (Callebaut)	Belgium	2006	Blend	West Africa	73%	5- B,S,C,V,L	poor aftertaste	2,0
15	Neuhaus (Callebaut)	Belgium	2006	Sao Tome	Sao Tome	75%	5- B,S,C,V,L	grassy, earthy, burnt	2,8
15	Scharffen Berger	U.S.A.	2006	Blend	Extra Dark	82%	5- B,S,C,V,L	dry, bitter, poor aftertaste	2,0
15	Scharffen Berger	U.S.A.	2006	Blend	Bittersweet	70%	5- B,S,C,V,L	cherry, mild bitter	3,5
24	Bonnat	France	2006	Blend	Caribbean, Trinite	75%	3- B,S,C	spicy, wood, long lasting	3,8
24	Michel Cluizel	France	2006	Blend	Carre Amer	60%	4- B,S,C,V	sweet, crisp nibs	2,0
24	Michel Cluizel	France	2006	Blend	Carre Grand Noir	85%	4- B,S,C,V	bitter coffee flavor	2,0
24	Michel Cluizel	France	2006	Sao Tome	Tamarina	70%	4- B,S,C,V	cream, licorice, herb	3,0
24	Michel Cluizel	France	2006	Dominican Republic	Los Ancones P.	67%	4- B,S,C,V	cream, raisin, licorice	3,8
24	Michel Cluizel	France	2006	Madagascar	Mangro P.	65%	4- B,S,C,V	grapefruit, spicy	3,8
24	Michel Cluizel	France	2006	Papua New Guinea	Maralumi P.	64%	4- B,S,C,V	tropical fruit, mild smoke	3,8
24	Neuhaus (Callebaut)	Belgium	2006	Venezuela	Occumare	71%	5- B,S,C,V,L	pastey, complex, coffee, vanilla	3,0
24	Original Hawaii Chocolate Factory	U.S.A.	2006	U.S.A.	Hawai'i Island, Kona Estate, Hualalai Mtn	60%	5- B,S,C,V,L	muddy, pastey, sweet, vanilla	3,0
32	Bonnat	France	2006	Venezuela	Occumare, Puerto Cabello	75%	3- B,S,C	sandy, cream, berry	3,8
32	El Rey	Venezuela	2006	Venezuela	Carenero Superior, Gran Saman	70%	5- B,S,C,V,L	gritty, chalky, earthy, sour	2,8
32	Pralus	France	2006	Jamaica	Jamaique	75%	4- B,S,C,L	cream, spicy	3,8
32	Santander (Compania Nacional)	Colombia	2006	Colombia	Colombian Semi Dark	53%	5- B,S,C,V,L	too sweet, candy-like	2,0
32	Santander (Compania Nacional)	Colombia	2006	Colombia	Colombian	65%	5- B,S,C,V,L	sweet, cherry, marshmallow	3,0
32	Santander (Compania Nacional)	Colombia	2006	Colombia	Colombian Dark	70%	5- B,S,C,V,L	rustic, cherry, marshmallow	3,8
40	Pralus	France	2006	Vanuatu	Vanuatu	75%	4- B,S,C,L	smoke dominants, mild fruit	3,0
40	Scharffen Berger	U.S.A.	2006	Blend	Kumasi Sambirano, Ghana & Madagascar	68%	5- B,S,C,V,L	fruit, smoke, burnt	2,0
48	Bonnat	France	2006	Venezuela	Maracaibo, El Rosario	75%	3- B,S,C	cream, fatty, floral	4,0
48	Felchlin	Switzerland	2006	Madagascar	Madagascar, Grand Cru	64%	5- B,S,C,V,L	basic cocoa, short length	3,0
48	Felchlin	Switzerland	2006	Venezuela	Maracaibo Clasificado	65%	5- B,S,C,V,L	cream, floral, complex	3,0
48	Felchlin	Switzerland	2006	Ecuador	Arriba	72%	5- B,S,C,V,L	fatty, rich, spicy, nut	3,5
56	Felchlin	Switzerland	2006	Bolivia	Alto Beni, Cru Savage	68%	5- B,S,C,V,L	complex, raspberry, cocoa	4,0
56	Scharffen Berger	U.S.A.	2006	Jamaica	Jamaica a l'ancienne	70%	5- B,S,C,V,L	fresh nibs, chocolate base off	2,0
56	Scharffen Berger	U.S.A.	2006	Blend	Semisweet	62%	5- B,S,C,V,L	tobacco, sweet, dairy	3,0
63	Dolfin (Belcolade)	Belgium	2006	Blend	Noir	70%	3- B,S,L	acidic, astringent, unpleasant	1,5
63	Valrhona	France	2006	Blend	Caribbean, Caraibe	66%	5- B,S,C,V,L	sligh bitter, wood	3,0

FIGURA 06 – FONTE DE DADOS ORIGINAL (ACERVO PESSOAL DO AUTOR)

A **Figura 04** apresentada no **item 2.1** ilustra o modelo dimensional adotado, bem como as tabelas de fatos e dimensões estruturadas durante o processo de ETL. A **Figura 07** ilustra a carga da base de dados original supracitada, em formato .csv / .xlsx, extraída a partir de um *dataset* extraído do KAGGLE.

Configurações da fonte de dados

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

☒ Fontes de dados no arquivo atual ☐ Permissões globais

Pesquisar configurações da fonte de dados

c:\users\moliveira\downloads\t...atasets\chocolate_ratings.xlsx

FIGURA 07 – 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 *software Microsoft Power BI* conforme indicado acima. Com os dados carregados, os responsáveis pelo ETL partem para a transformação dos dados, que é realizada no *Power BI* a partir do *Power Query*. No *Power Query*, foram estruturadas as seguintes etapas:

1. Tratamento da fonte de dados original carregada, onde foram realizados alguns tratamentos básicos e usuais, tais como decodificação de dados, conversão de atributos, agregação de valores, criação de chaves identificadoras e verificação da integridade dos dados. Estes tratamentos são necessários para garantia da qualidade, formatação e integridade dos dados de acordo com as necessidades de análise. Por exemplo, diversas colunas foram renomeadas, reorganizadas, reclassificadas, padronizadas, tiveram o seu tipo alterado e passaram por verificações de erros. Por fim, para as colunas “*Ingredients*” e “*Flavors*”, foi necessária a modificação da estrutura dos dados original, já que elas estavam com dados concatenados em uma única coluna. Nisto, estas 2 colunas foram desmembradas em outras 9 colunas a partir de operações de divisão por delimitador e limpeza de texto conforme necessidade de análise. As etapas aplicadas estão ilustradas na **Figura 08**;

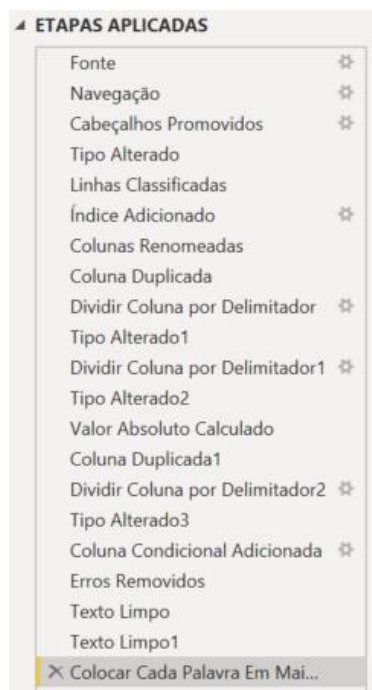


FIGURA 08 – PRIMEIRA ETAPA DE TRANSFORMAÇÃO DOS DADOS (ACERVO PESSOAL DO AUTOR)

2. Com a base de dados original tratada, foi executada uma etapa de otimização da estrutura relacional a partir do seu desmembramento em tabelas de dimensões com chaves identificadoras e da tabela fato relacionada às de dimensões 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**. As tabelas de dimensões foram construídas a partir da base de dados original tratada, utilizando operações básicas, onde as colunas desnecessárias foram excluídas e as remanescentes foram renomeadas, reorganizadas, reclassificadas, padronizadas e tiveram valores duplicados excluídos – a exclusão de dados duplicados é essencial para otimização da estrutura relacional, removendo dados desnecessários. Já a tabela fato foi estruturada, além da utilização operações básicas de organização de dados já citadas, a partir do *Rating* das avaliações (fato), que foi ligado às suas respectivas dimensões por meio da mescla de consultas utilizando chaves estrangeiras conforme **Figura 09**;

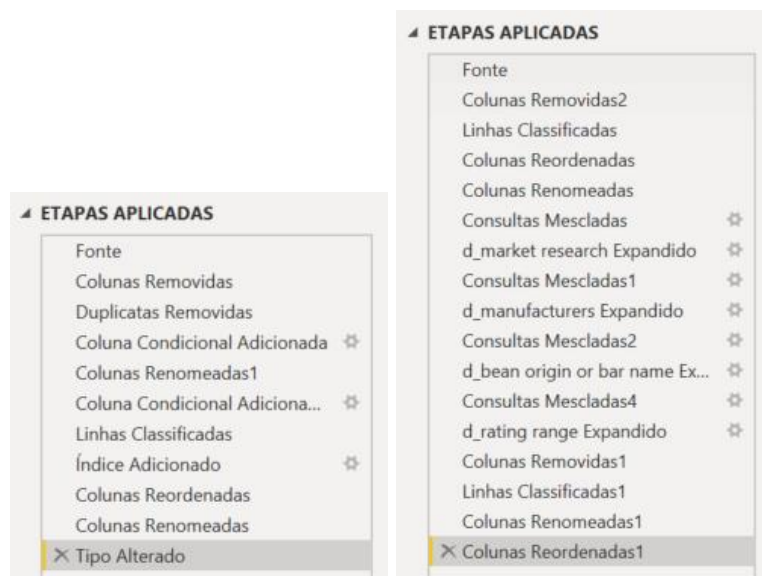


FIGURA 09 – SEGUNDA ETAPA DE TRANSFORMAÇÃO DOS DADOS (ACERVO PESSOAL DO AUTOR)

3. Por fim, independente da etapa de homologação de dados que será executada posteriormente, foi realizada uma conferência básica 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 10**.

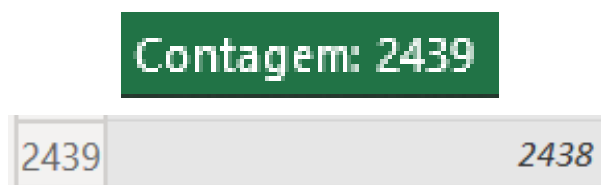


FIGURA 10 – TERCEIRA ETAPA DE TRANSFORMAÇÃO DOS DADOS (ACERVO PESSOAL DO AUTOR)

4. Camada de Apresentação

4.1 Dashboard

Para apresentação e visualização dos dados foi utilizado o *Power BI*, uma ferramenta de *self-service BI* que permite a criação de visualizações interativas e aplicação de recursos de *Business Intelligence* de uma maneira simples, intuitiva e

de fácil acesso. A concepção do *dashboard* considera fundamentos da pirâmide organizacional, onde foram estruturados três painéis:

- **Estratégico:** concebido para alta administração, com os principais indicadores e dados resumidos e com visualizações objetivas, visando tomadas de decisões gerenciais.
- **Tático:** concebido para nível de gerência, com os demais indicadores e dados considerando ferramentas OLAP (*Online Analytical Processing*) que permitam análises hierárquicas e históricas dos dados, como filtros de segmentação, *drilling* etc., visando tomadas de decisões gerenciais e melhoria de processos
- **Operacional:** concebido para nível de supervisão e operação, com os dados dispostos em sua maior granulometria visando análises cruzadas para tomadas de decisões e otimizações de atividades rotineiras.

Além da divisão conforme pirâmide organizacional, que visa segregar as visualizações conforme público-alvo, foram utilizadas técnicas de *Data Storytelling* mirando uma disposição dos dados que combinasse as suas respectivas visualizações com o contexto em que estão inseridos e com a narrativa desejada. As cores do *Dashboard* foram selecionadas estrategicamente buscando um maior contraste que facilitasse as visualizações, bem como para caracterização do produto em análise – foram utilizados tons de branco, marrom e dourado majoritariamente, tendo vista as cores dos chocolates, das sementes de cacau e contrastes.

Painel Estratégico

O painel Estratégico, conforme já descrito, visa apresenta de maneira objetiva os principais indicadores relacionados com o objetivo organizacional do *Dashboard*. Dentro do contexto deste trabalho, tendo em vista os objetivos descritos no **item 1.2** e as perguntas-alvo do *Dashboard* descritas no **item 2.1**, este painel mira responder aos principais questionamentos que irão fomentar, de maneira objetiva, a criação da nova barra de chocolate: qual o percentual de cacau mais bem avaliado pelos usuários, quais as melhores sementes / barras de cacau e qual a relação da qualidade das sementes / barras com sua localidade.

A **Figura 11** ilustra este painel, que apresenta instantaneamente os principais indicadores da pesquisa de mercado em formato de indicador velocímetro:

percentual de Cacau e *Rating* médio. Estes indicadores foram pensados para serem exibidos de maneira iterativa a partir de filtros de segmentação, que determinam faixas de *Rating* e de tempo, bem como por *Cards* quantitativos, que se alteram conforme filtro de segmentação escolhido para passar noções de tamanho amostral ao público-alvo. Para reuniões específicas, é possível que estes filtros sejam congelados em indicadores do *Power BI* para transmissão da narrativa desejada.



FIGURA 11 – VISUALIZAÇÃO DO PAINEL ESTRATÉGICO (ACERVO PESSOAL DO AUTOR)

Além disto, tendo em vista a seleção das melhores sementes / barras de cacau e a criação de um nexos casual com sua localidade, temos um gráfico de barras com o *Rating* médio de cada semente / barra, bem como um mapa com bolhas que crescem conforme *Rating* médio. Como temos um *Rating* com duas casas decimais, os gráficos foram equipados com um Controle deslizante de *zoom*, que permite ao usuário controlar a granularidade da informação.

A **Figura 12** ilustra a utilização dos filtros de segmentação, onde temos filtradas apenas as duas melhores classes de *Rating* (em uma escala de 0 a 5, foram filtrados valores maiores do que 3,5) em um tamanho amostral de 957 avaliações. É possível observar um aumento do *Rating* médio para 3,63, enquanto o percentual de cacau médio se mantém em 71%, indicando ser a média mais bem avaliada pelos usuários. Além disto, temos uma segmentação das sementes / barras de cacau mais

bem avaliadas pelos usuários, com um indicativo de que elas se concentram na linha do equador (América Central e do Sul, África, sul da Ásia e Oceania), com preferência para a zona tropical do planeta. Ao limitarmos a segmentação para avaliações após 2019, notamos um tamanho amostral reduzido de 168 avaliações – quanto menor o tamanho amostral, menor o nível de confiança da pesquisa de mercado. Temos a manutenção do percentual médio de cacau, indicando manutenção da preferência dos usuários ao longo do tempo, bem como agrupamento das sementes / barras de cacau mais bem avaliadas para a América Central e do Sul, indicando que, atualmente, as melhores sementes se encontram nesta região.



FIGURA 12 – FILTROS DE SEGMENTAÇÃO DO PAINEL ESTRATÉGICO (ACERVO PESSOAL DO AUTOR)

Tendo em vista tomadas de decisão a nível estratégico, temos indicadores objetivos sugerindo que barras de chocolate com sementes da América Central e do Sul, com percentual de cacau próximo a 71%, tendem a ser mais bem avaliadas e preferidas pelos usuários de chocolate. Reitera-se ser uma sugestão preliminar, já que é necessário explorar melhor os filtros de segmentação e os demais painéis para conclusões mais completas – tal ponto será mais bem detalhado no vídeo de apresentação do *Dashboard* que compõe este trabalho e no **item 6 - Conclusões**.

Painel Tático

O painel Tático, conforme supracitado, apresenta visualizações com recursos que possibilitam análises hierárquicas e históricas mais detalhadas visando, além da convergência com as metas organizacionais, melhorias mais granulares em tomadas de decisões gerenciais e em processos. Nisto, nesta aba, além dos recursos de tamanho amostral e segmentação presentes no painel Estratégico, temos também

gráficos e mapas mais detalhados que exploram informações que contêm maior grau de subjetividade, porém que podem induzir análises de causa e efeito gerando *insights* sobre a relação dos ingredientes e sabores com a preferência dos usuários – estes *insights* fomentarão a definição das características da barra de chocolate, e, conseqüentemente, o seu processo de fabricação, respondendo às perguntas finais estabelecidas no **item 2.1**. Ademais, o painel apresenta informações sobre fabricas concorrentes, que podem ser utilizadas como *benchmarking*.

A **Figura 13** ilustra as visualizações de ingredientes deste painel, que contam com gráficos de barras para demonstração do *Rating* médio tanto dos ingredientes individuais, quanto do *mix* de ingredientes. Além disto, o painel apresenta um gráfico de faixas para demonstrar a evolução do *Rating* de cada ingrediente individual com o passar do tempo. A preferência por gráficos de barra foi dada em função da grande quantidade de elementos por eixo.

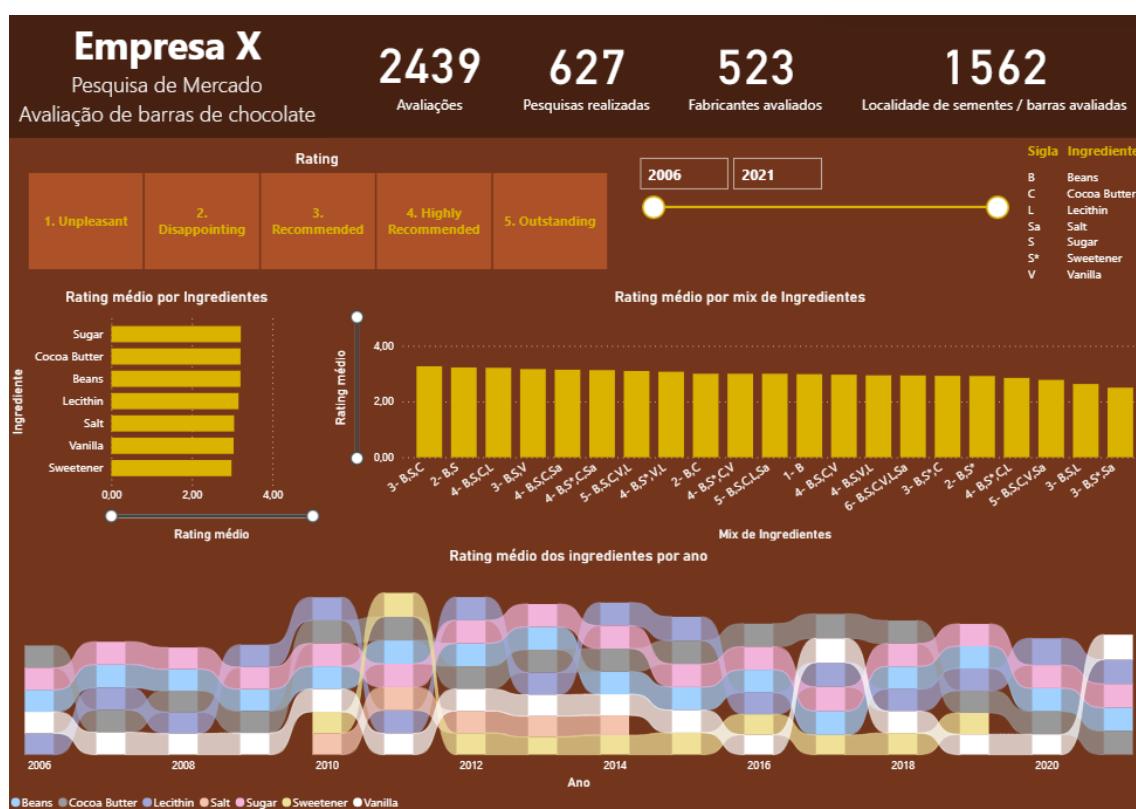


FIGURA 13– VISUALIZAÇÕES DE INGREDIENTES DO PAINEL TÁTICO (ACERVO PESSOAL DO AUTOR)

Para operacionalizar o cálculo do *Rating* dos sabores e ingredientes isolados, foi realizada uma operação de pivoteamento das colunas em linhas para que cada um dos elementos tivesse o seu respectivo *Rating* atribuído em uma linha

específica. Esta operação foi necessária em função da maneira como os dados foram inseridos na base de dados original, onde, conforme já descrito no **item 3.2**, eles se apresentavam concatenados em uma única coluna.

O gráfico de faixas, ainda sem a aplicação de filtros, já permite alguns *insights*: há uma queda no *Rating* de barras com açúcar e manteiga de cacau com o passar de anos, enquanto a preferência dos usuários por lecitina e baunilha aumenta – estes são aditivos alimentares mais saudáveis. Para ilustrar a aplicação das segmentações deste painel, conforme **Figura 14**, foram segmentados os *Ratings* acima de 3,5 avaliados após 2019, similar à segmentação feita no painel Estratégico, porém com uma segmentação complementar de barras que possuem baunilha e lecitina. A **Figura 14** ilustra o efeito destes filtros para as visualizações de sabores, que, além dos mesmos gráficos de barras presentes nas visualizações dos ingredientes, apresenta também um gráfico de nuvem de palavras. Este gráfico, configurado para separar as palavras compostas, fornece uma visualização extra mais intuitiva de uma análise complementar de sabores, que além do seu *Rating*, considerando também a sua recorrência – no filtro em exemplo, para chocolates com lecitina e baunilha, temos boas avaliações de sabores de cacau, picantes e frutados, independente do sabor ser suave (*mild*) ou forte (*rich*).



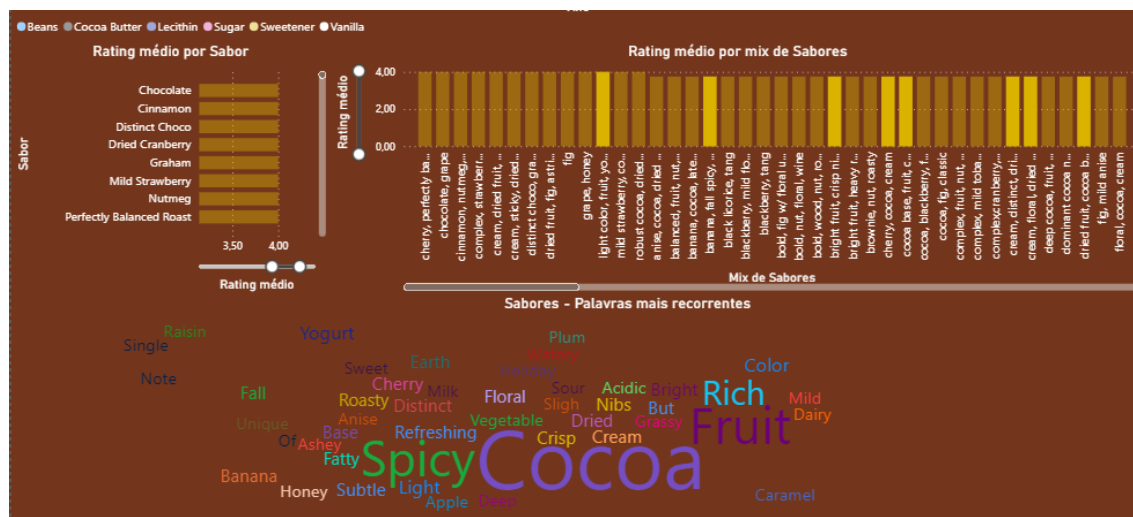


FIGURA 14 – FILTROS DE SEGMENTAÇÃO DO PAINEL TÁTICO (ACERVO PESSOAL DO AUTOR)

Por fim, a **Figura 15** apresenta as visualizações para fabricantes concorrentes a partir da segmentação para as avaliações realizadas após 2019 com *Ratings* acima de 3,5. Esta visualização, além do gráfico de barras fornecendo uma maior granularidade dos fornecedores conforme *Rating*, temos um mapa com bolhas que crescem conforme avaliação da barra de chocolate dos fornecedores. Filtrando os cinco fornecedores mais bem avaliados, podemos notar que eles se concentram na França, U.S.A e Vietnã respectivamente.

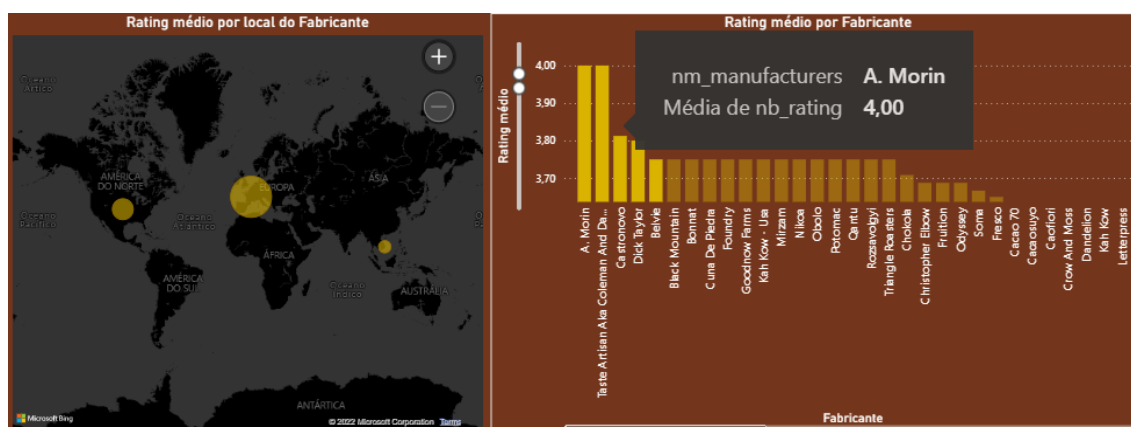


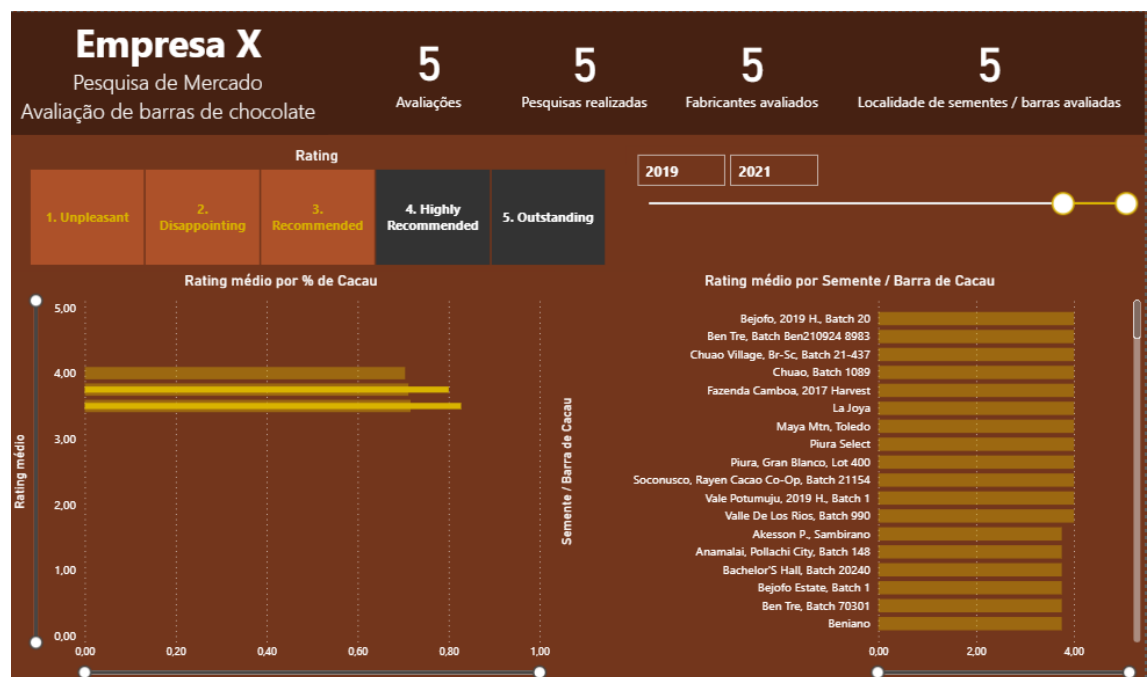
FIGURA 15 – FILTROS DE SEGMENTAÇÃO DO PAINEL ESTRATÉGICO (ACERVO PESSOAL DO AUTOR)

Painel Operacional

O painel Operacional apresenta visualizações em sua maior granulometria, tendo como foco o máximo uso de interações e filtros de segmentação para que o público-alvo a nível de supervisão e operação possa realizar análises rotineiras

visando otimizações em seus trabalhos com maior agilidade. Para isto, este painel apresenta, além dos recursos de tamanho amostral e segmentação presentes nos painéis Estratégico e Tático, diversas tabelas contendo todos os fatos e dados do modelo dimensional, onde temos como foco análises cruzadas em sua maior granulometria.

A **Figura 16** ilustra um exemplo de análise cruzada com alta granulometria: nela, temos segmentados os *Ratings* acima de 3,5 avaliados após 2019. Conforme demonstrado no painel Estratégico, o percentual médio de Cacau para este grupo é de 72%. Entretanto, fruto da alta granulometria, é possível observar que este grupo apresenta cinco barras de chocolate bem avaliadas com percentual de cacau acima de 80%. Ao também agrupá-las, podemos observar diversas informações em um nível de detalhes maior: temos as sementes / barras que compõe estes chocolates, com a sua respectiva localização e *Rating* médio, os sabores mais marcantes dessas barras, os respectivos ingredientes e *mix* de ingredientes, bem como os fabricantes com sua respectiva localidade.



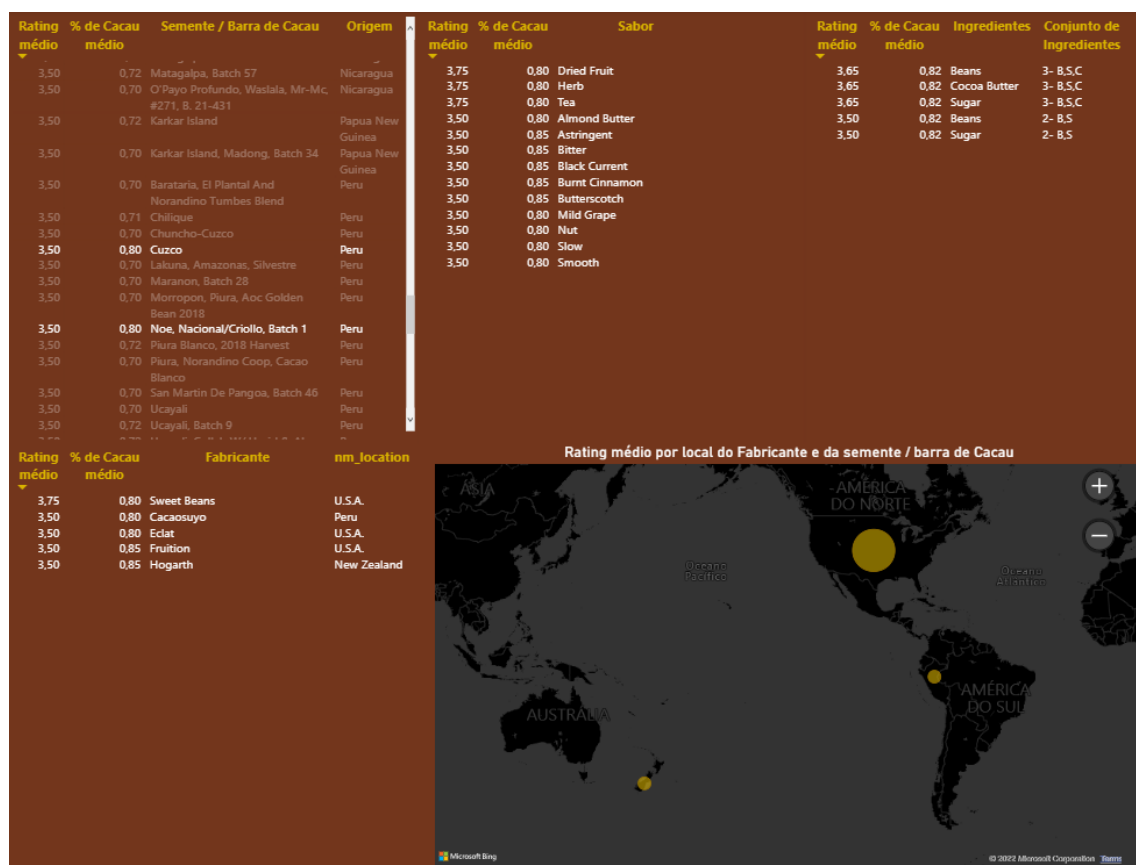


FIGURA 16 – FILTROS DE SEGMENTAÇÃO DO PAINEL OPERACIONAL (ACERVO PESSOAL DO AUTOR)

Nisto, observa-se que este painel apresenta inúmeras opções de análises cruzadas que podem ser utilizadas pelo público-alvo para otimizações rotineiras e análises mais profundas.

4.2 Análises avançadas

A solução de BI desenvolvida neste trabalho tem como intuito suportar o processo de pré-desenvolvimento de produtos, associando aspectos de barras de chocolate, como percentual de Cacau, fabricante, ingredientes e sabores, a avaliações realizadas por usuários (*Ratings*). Neste sentido, considerando os parâmetros de cada barra e as respectivas avaliações feitas pelos usuários, uma das formas de se aprimorar o processo de decisão das características do produto em fase de pré-desenvolvimento parte da utilização de técnicas de *Machine Learning* e *Deep Learning* com o intuito de se prever se determinado produto, ainda não criado, será bem avaliado ou não pelos usuários.

Para isto, foi desenvolvido um *Script* utilizando a linguagem *Python* com o objetivo de se treinar uma Rede Neural de Aprendizado Profundo para realizar a

predição do *Rating* de uma barra de chocolate a partir de seus ingredientes, percentual de Cacau e localização da semente. Abaixo, é possível observar as bibliotecas utilizadas para o desenvolvimento do *Script* – em específico para o desenvolvimento e treinamento do modelo de *Machine Learning*, foi utilizada a biblioteca *Tensorflow*.

```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
from sklearn.preprocessing import MinMaxScaler
from sklearn.model_selection import train_test_split
import tensorflow as tf
```

FIGURA 17 – BIBLIOTECAS UTILIZADAS PARA MACHINE LEARNING (ACERVO PESSOAL DO AUTOR)

A base de dados utilizada se trata da mesma utilizada para criação da solução de *Business Intelligence*, em formato .csv. Inicialmente, o *Script* realiza uma etapa de verificação da integridade dos dados, observando se existem valores nulos ou se o tipo dos dados se apresenta no formato correto. Por fim, as demais colunas, que não apresentam dados de interesse para a análise, são removidas.

	Country of Bean Origin	Cocoa Percent	Ingredients	Rating
0	Ghana	0.71	5- B,S,C,V,L	2.0
1	Blend	0.70	5- B,S,C,V,L	2.5
2	Colombia	0.65	5- B,S,C,V,L	3.0
3	Blend	0.73	5- B,S,C,V,L	2.0
4	Sao Tome	0.75	5- B,S,C,V,L	2.8
...
2434	Peru	0.74	3- B,S,C	3.2
2435	Philippines	0.70	3- B,S,C	3.2
2436	Papua New Guinea	0.70	3- B,S,C	3.5
2437	Indonesia	0.72	2- B,S	3.5
2438	Malaysia	0.72	2- B,S	3.5

2439 rows × 4 columns

FIGURA 18 – BASE DE DADOS TRATADA PARA MACHINE LEARNING (ACERVO PESSOAL DO AUTOR)

Com a base de dados definidas, tendo a coluna “*Rating*” como variável de interesse (y), as demais colunas com dados categóricos são convertidas em variáveis “*dummys*” com indicadores binários (x) conforme ilustrado na **Figura 19**.

	Cocoa Percent	Australia	Belize	Blend	Bolivia	Brazil	Burma	Cameroon	China	Colombia	...	4- B.S*.CV	4- B.S*.VL	4- B.S.C.L	4- B.S.C.Sa	4- B.S.C.V	4- B.S.VL	5- B.S.C.LSa	5- B.S.C.VL	5- B.S.C.VSa	6- B.S.C.VLSa
0	0.500000	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.0	0.0	0.0
1	0.482759	0.0	0.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.0	0.0	0.0
2	0.396552	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.0	0.0	0.0
3	0.534483	0.0	0.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.0	0.0	0.0
4	0.568966	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.0	0.0	0.0
...
2434	0.551724	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
2435	0.482759	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
2436	0.482759	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
2437	0.517241	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
2438	0.517241	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0

2439 rows x 83 columns

FIGURA 19 – MATRIZ “X” TRATADA PARA MACHINE LEARNING (ACERVO PESSOAL DO AUTOR)

Com as variáveis “x” e “y” definidas, o *Script* inicia o treinamento do modelo. Com o intuito de se otimizar o modelo, algumas variáveis podem ser alteradas ao longo do treinamento visando maior acurácia. A **Figura 20** apresenta as camadas de entrada, escondidas e de saída definidas para o modelo. Para treinamento e teste, foram definidas aleatoriamente amostras de 50% do volume de dados.

Model: "model"

Layer (type)	Output Shape	Param #
input_1 (InputLayer)	[(None, 83)]	0
dense (Dense)	(None, 8)	672
dense_1 (Dense)	(None, 8)	72
dense_2 (Dense)	(None, 1)	9
Total params: 753		
Trainable params: 753		
Non-trainable params: 0		

FIGURA 20 – ARQUITETURA DO MODELO DE MACHINE LEARNING (ACERVO PESSOAL DO AUTOR)

A função de ativação escolhida para as camadas foi a “*Rectified Linear Unit*” (*ReLU*), que se trata de uma função linear por partes que produzirá a entrada diretamente se for positiva, caso contrário, produzirá zero. A *ReLU* é uma função de

ativação “padrão” para muitos tipos de Redes Neurais por tornar mais fácil o treinamento do modelo que a usa, geralmente obtendo melhor desempenho.

Já a função de perda utilizada foi a “Mean Squared Error” (MSE), sendo a função de perda mais comumente utilizada para regressões. O MSE é utilizado quando se acredita que seus dados de destino, condicionados à entrada, são normalmente distribuídos em torno de um valor médio e quando é importante penalizar muito os valores discrepantes.

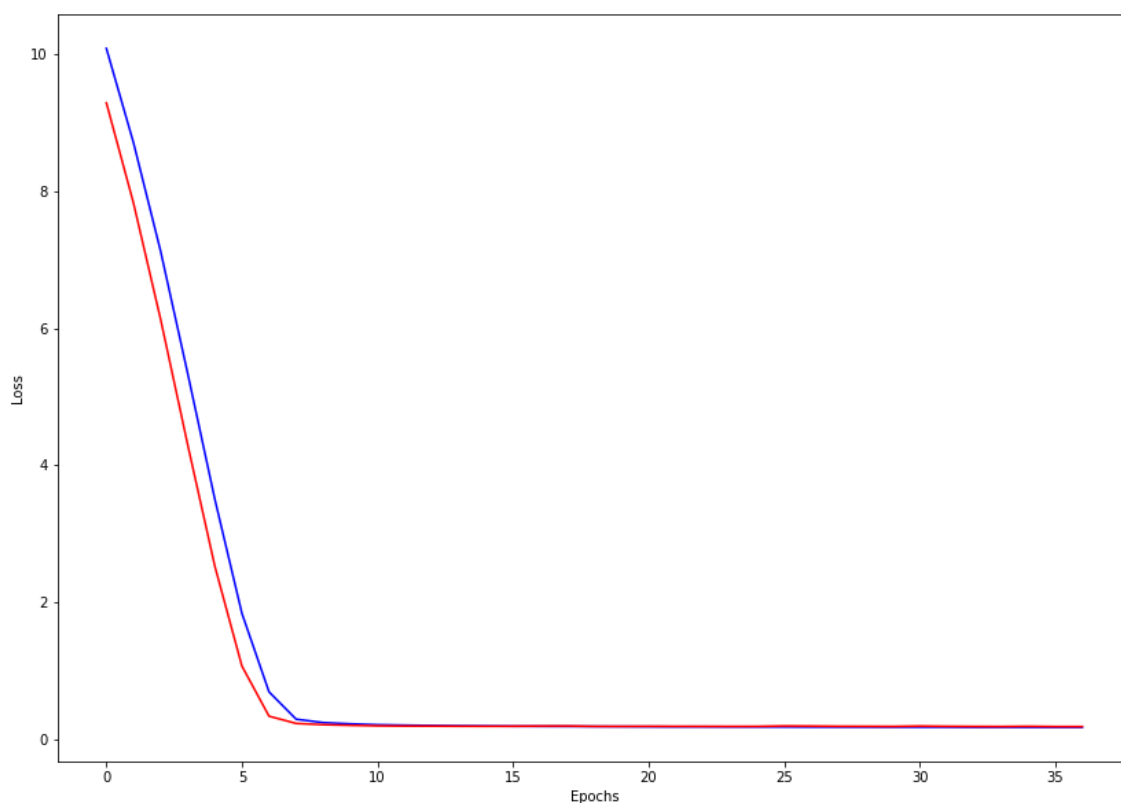
Por fim, a **Figura 21** ilustra os parâmetros finais definidos para treinamento do modelo, onde estes foram sendo testados e alterados por diversas vezes ao longo do treinamento até que se atingisse a menor perda. O *Batch Size* (tamanho do lote) se refere ao número de amostras usadas numa iteração. A alteração do *Batch Size* não representou melhorias na acurácia, onde foi mantido o valor padrão de 30. Já a *Learning Rate* (taxa de aprendizagem) indica em que ritmo os pesos do modelo são atualizados na rede com respeito a perda de gradiente. Notou-se que, para este modelo, o decréscimo da *Learning Rate* para 0.001 melhorou a sua acurácia.

```
epochs=37

history = model.fit(
    x_train,
    y_train,
    validation_split=0.2,
    epochs=epochs,
    batch_size=30,
    shuffle = True,
    verbose=2
)
```

FIGURA 21 – TREINAMENTO DO MODELO DE MACHINE LEARNING (ACERVO PESSOAL DO AUTOR)

Os *Epochs* (época) descrevem o número de vezes que o algoritmo vai passar pelo treinamento. Para este modelo foi utilizado o valor de 37, que foi definido a partir da observação do gráfico de *Perdas x Epochs*, definindo-se a taxa ótima de passagens pelo treinamento que resultaria em menor perda (ilustrada abaixo na **Figura 22**).



+ Code

+ Markdown

```
np.argmax(history.history['val_loss'])
```

36

FIGURA 22 – PERDA X EPOCHS DO MODELO DE MACHINE LEARNING (ACERVO PESSOAL DO AUTOR)

A avaliação do modelo resultou em uma perda de aproximadamente 19% para as amostras de teste, sendo considerado um bom resultado. Abaixo, na **Figura 23**, é possível observar os resultados da predição do lote de testes (de tamanho de 50% do volume de dados), onde observamos uma acurácia aceitável para o modelo.

prediction	y_test
array([[2.8365834],	1090 2.5
[3.168179],	1492 3.2
[3.1986709],	238 2.8
...,	60 4.0
[3.2802076],	718 3.2
[3.324985],	...
[3.2618902]], dtype=float32)	169 2.8
	1616 3.2
	1177 3.5
	1223 3.8
	1074 3.8
	Name: Rating, Length: 1220, dtype: float64

FIGURA 23 – PREDIÇÃO DO MODELO X LOTE DE TESTE (ACERVO PESSOAL DO AUTOR)

Estando o modelo definido, sua arquitetura foi salva no formato .h5, bem como o *Script* foi salvo em formato .ipynb. A partir deste modelo, a Empresa X pode utilizá-lo para predição do *Rating* de um determinado produto em fase de pré-desenvolvimento que tenha os seus parâmetros de percentual de Cacau, localidade da semente e ingredientes definidos. Além disto, o modelo pode ser continuamente alimentado com novos dados de pesquisa de mercado para melhoria do aprendizado a partir de um aumento do espaço amostral.

5. Registros de Homologação

Para homologação dos dados apresentados no *Dashboard*, foram desenvolvidas consultas à base de dados original a partir do *software Microsoft Excel*, considerando que, conforme solução de *Business Intelligence* adotada, não há um servidor SQL específico para realização de consultas. Neste sentido, a partir de uma tabela dinâmica criada via base de dados original, foram levantados os principais valores e quantitativos dos fatos e dimensões utilizados – quantidades de avaliações, pesquisas, fabricantes, sementes / barras de cacau, *mix* de sabores e *mix* de ingredientes, além das médias gerais do percentual de cacau e do *rating* (Figura 24).



FIGURA 24 – DADOS LEVANTADOS EM EXCEL PARA HOMOLOGAÇÃO (ACERVO PESSOAL DO AUTOR)

Para automatização do processo de homologação, estes dados foram inseridos no *Power BI* da solução construída a partir de uma aba oculta, tendo em vista que a homologação deve estar visível apenas para consulta da equipe de ETL. Esta aba apresenta uma consolidação dos dados gerados pelo *Power BI* com os dados gerados pela tabela dinâmica do *Excel*, onde, conforme **Figura 25**, é possível

observar a convergência dos dados e homologar os resultados apresentados no *Dashboard*.



FIGURA 25 – HOMOLOGAÇÃO DO DASHBOARD (ACERVO PESSOAL DO AUTOR)

Cabe reiterar que, apesar da homologação disposta acima, a qualidade, integridade e padronização dos dados foram observadas de maneira criteriosa em todas as etapas propostas para a solução de *Business Intelligence*.

6. Conclusões

A partir do desenvolvimento da solução de *Business Intelligence* objeto deste trabalho, que foi construída desde a estruturação do Modelo Dimensional conforme perguntas alvo até a criação da máscara do *Dashboard*, foi possível identificar diversos fatores chave ligados ao objetivo deste trabalho, que consiste em apoiar a Empresa X na fase de pré-desenvolvimento de produtos. De maneira inicial, a partir do desenho do Modelo Dimensional e do tratamento da base de dados, desde a extração até a carga, foi possível levantar diversos pontos de melhoria na construção e alimentação da base de dados a partir das pesquisas de mercado que podem levar a análises mais completas e precisas – a maioria destas lacunas se encontram na granulometria e no registro dos dados.

Observando os dados gerados pelo *Dashboard*, foi possível identificar de uma maneira clara e objetiva que as barras de chocolate mais bem avaliadas pelos usuários apresentam percentual de cacau por volta de 71%. Além disto, para seleção das melhores sementes e barras de cacau para confecção da barra de chocolate, os dados indicam que os melhores fornecedores se encontram nas zonas

tropicais, principalmente na América Central, do Sul e Nova Zelândia. Além da localização, é possível identificar no *Dashboard* a lista de nominal das sementes e barras com seu respectivo *Rating* médio.



FIGURA 26 – % DE CACAU MÉDIO E LOCALIZAÇÃO DAS SEMENTES / BARRAS DE CACAU (ACERVO PESSOAL DO AUTOR)

Ao explorarmos de uma maneira mais detalhado os parâmetros avaliados, é possível notar uma evolução histórica da preferência dos usuários por ingredientes mais saudáveis, tal como Lecitina ou Baunilha, em detrimento da Manteiga de Cacau ou do Açúcar. Apesar da preferência dos usuários por barras de chocolate com 71% de cacau e ingredientes mais saudáveis, existem diversas barras de chocolate com sabores e características diferentes com excelentes *Ratings*, abrindo espaço para a pesquisa, criação e desenvolvimento de inúmeros produtos com características variadas e não usuais.

Neste sentido, nota-se que a solução de *Business Intelligence*, combinada com o modelo de *Machine Learning*, permite ao público-alvo explorar diversas opções para a criação de novos produtos, sejam eles com características usuais ou diferenciadas, predizendo o seu sucesso perante os usuários. Tais análises fomentam as decisões estratégicas, táticas e operacionais da Empresa X, permitindo

análises e avaliações de fornecedores e concorrentes (ambiente externo, oportunidades e ameaças) bem como de seu próprio produto (ambiente interno, fortalezas e fraquezas), gerando vantagem competitiva e agregando valor.

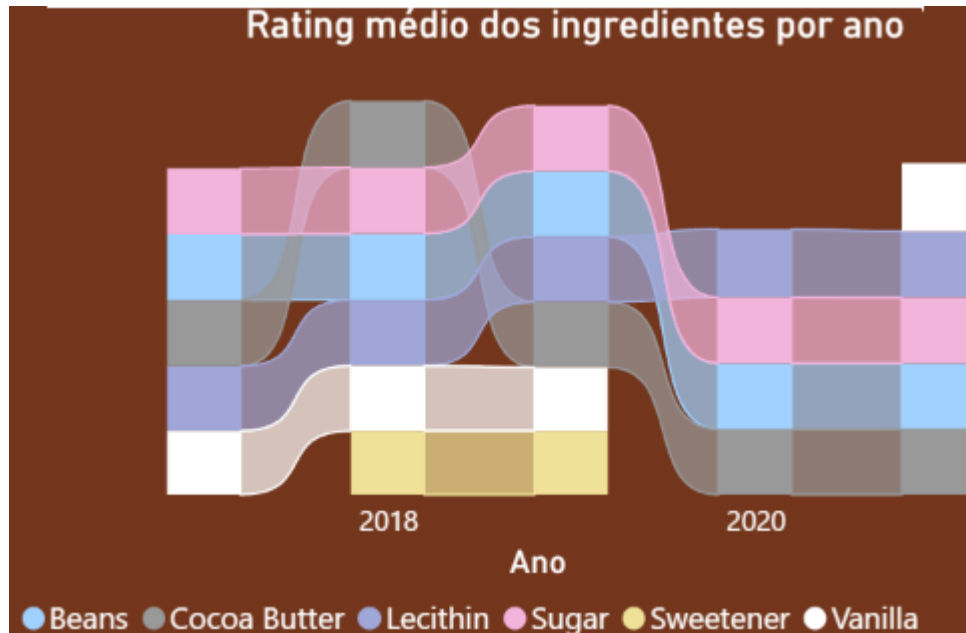


FIGURA 27 – EVOLUÇÃO DA AVALIAÇÃO DOS INGREDIENTES (ÁCERVO PESSOAL DO AUTOR)

6.1 Proposta de Intervenção

Considerando as principais conclusões geradas no trabalho, é possível tecer propostas de intervenção para a Empresa X nas seguintes esferas abaixo:

- **Estratégica:** Dentre as perguntas respondidas pelo trabalho, torna-se clara uma maior preferência dos usuários por barras com 71% de cacau e ingredientes mais saudáveis. Tais fatores devem direcionar o desenvolvimento de novos produtos com estas características, se utilizando também de sementes e barras de cacau provindas de fornecedores localizados em zonas tropicais. Em relação aos concorrentes, é possível monitorá-los tendo em vista suas principais localizações, em sua maioria na América Central, do Sul, Europa e Nova Zelândia.
- **Tática:** Apesar das características mais usuais observadas, é possível que a Empresa X realize explorações e criações mais detalhadas com o intuito de se identificar lacunas de mercado e atingir usuários

específicos. A partir das diversas combinações de percentual de cacau, sabores e ingredientes, seria possível identificar nichos de mercado não explorados e ainda não povoado por concorrentes conforme a sua localização. Estes produtos, após selecionados, seriam submetidos ao algoritmo de *Machine Learning*, que auxiliaria no processo de decisão de sua viabilidade através da predição de seu *Rating* médio.

- **Operacional:** Por fim, a nível operacional, seria possível atuar em todos os processos envolvidos da solução de *Business Intelligence*, desde a captação de dados através de pesquisas de mercado, buscando melhorar os dados que alimentam as análises, até a seleção minuciosa de características dos produtos, fornecedores e concorrentes através das inúmeras possibilidades que o *Dashboard* trás.

6.2 Lições Aprendidas

De uma maneira geral, a principal lição aprendida neste trabalho é que a utilização de metodologias e boas práticas para o desenvolvimento de soluções de *Business Intelligence* torna a sua criação muito mais rápida, fácil e intuitiva. Usualmente, após a expansão das soluções de *self-service BI*, as soluções passaram a ser criadas de uma maneira empírica, fazendo com que o seu resultado tenha maior probabilidade de gerar informações erradas, apresentar lentidão, ou não estar preparada para o carregamento de dados atualizados por exemplo. Ao se prover de metodologias e boas práticas de *Business Intelligence* e *Data Science* em todas as etapas, temos soluções homologadas, confiáveis, contínuas, atualizáveis, eficientes e alinhadas com os objetivos propostos.

Considerando o trabalho desenvolvido, foram identificados diversos pontos de melhoria. Além dos pontos já citado no **item 2.3**, para as análises avançadas, seria possível enrobustecer mais o modelo criado a partir de um tratamento mais profundo da base de dados, incluindo parâmetros como ingredientes e sabores de maneira individualizada. O aumento dos parâmetros de análise tornaria as predições ainda mais confiáveis. Também seria possível selecionar parâmetros, tais como função de ativação, função de perda, taxa de aprendizado, *batch size* e *epochs* de uma maneira mais personalizada, aumentando a acurácia do modelo de predição.

7. Links

1. *Datasets* extraídos 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.