

Faculdade

XPe



RELATÓRIO

PROJETO
APLICADO

XP Educação
Relatório do Projeto Aplicado

Análise de Vendas de Bebidas Alcoólicas em Iowa: Insights para o Varejo

Lorival Souza Macêdo

Orientador(a): Marcos Prochnow

17/03/2025



LORIVAL SOUZA MACÊDO

XP EDUCAÇÃO

RELATÓRIO DO PROJETO APLICADO

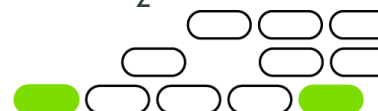
Análise de Vendas de Bebidas Alcoólicas em Iowa: Insights para o Varejo

Relatório de Projeto Aplicado desenvolvido para fins de conclusão do curso DE Ciências de Dados.

Orientador (a): Marcos Prochnow

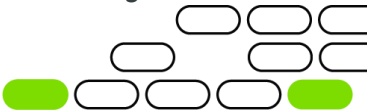
Salvador

17/03/2025



Sumário

1. CANVAS do Projeto Aplicado	4
Desafio	5
1.1.1 Análise de Contexto	5
1.1.2 Personas	6
1.1.3 Benefícios e Justificativas	7
1.1.4 Hipóteses	8
1.2 Solução	9
1.2.1 Objetivo SMART	9
1.2.2 Premissas e Restrições	11
1.2.3 Backlog de Produto	13
2. Área de Experimentação	112.1 Sprint 1
	16
2.1.1 Solução	16
Evidência do planejamento:	16
Evidência da execução de cada requisito:	16
Evidência dos resultados:	16
2.1.2 Lições Aprendidas	16
2.2 Sprint 2	17
2.2.1 Solução	17
Evidência do planejamento:	17
Evidência da execução de cada requisito:	17
Evidência dos resultados:	17
2.2.2 Lições Aprendidas	17
2.3 Sprint 3	18
2.3.1 Solução	18
Evidência do planejamento:	18
Evidência da execução de cada requisito:	18
Evidência dos resultados:	18
2.3.2 Lições Aprendidas	18
3. Considerações Finais	203.1 Resultados
	19
3.2 Contribuições	19
3.3 Próximos passos	22



1. CANVAS do Projeto Aplicado

Abaixo segue a Figura 1, que representa a estrutura do Projeto Aplicado conforme o proposto pelo curso de Ciência de Dados.



Figura 1 - Canvas do Projeto Aplicado.

1.1 Desafio

1.1.1 Análise de Contexto

O mercado de bebidas alcoólicas é um setor altamente lucrativo, mas sujeito a flutuações sazonais, variações regionais e mudanças nos padrões de consumo. Compreender esses fatores é essencial para distribuidores e varejistas maximizarem seus lucros e otimizarem a gestão das vendas dos produtos e estoques.

A base de dados Iowa Liquor Retail Sales, de domínio público no Bigquery, contém registros detalhados das vendas de bebidas alcoólicas no estado de Iowa. Esses dados incluem informações sobre categorias de produtos, volume de vendas, preços, localização das lojas e tendências ao longo do tempo. No entanto, a análise direta dessas informações pode ser complexa, exigindo um processo estruturado de tratamento e visualização para gerar insights estratégicos.

O objetivo deste projeto é explorar essa base para identificar padrões de consumo, sazonalidade e margens de lucro, possibilitando recomendações para otimização de estoques e precificação. Para isso, será utilizada uma abordagem

baseada em Bigquery para processamento e Looker para visualização interativa, permitindo que gestores do setor tomem decisões mais assertivas.

A análise incluirá a identificação de produtos mais e menos vendidos, períodos de alta demanda, lojas com maior participação de mercado e tendências que impactam a performance do setor. Esses insights poderão apoiar distribuidores e investidores na maximização de oportunidades e mitigação de riscos, tornando a gestão do varejo de bebidas alcoólicas mais eficiente e estratégica.



Figura 2 - Matriz CSD

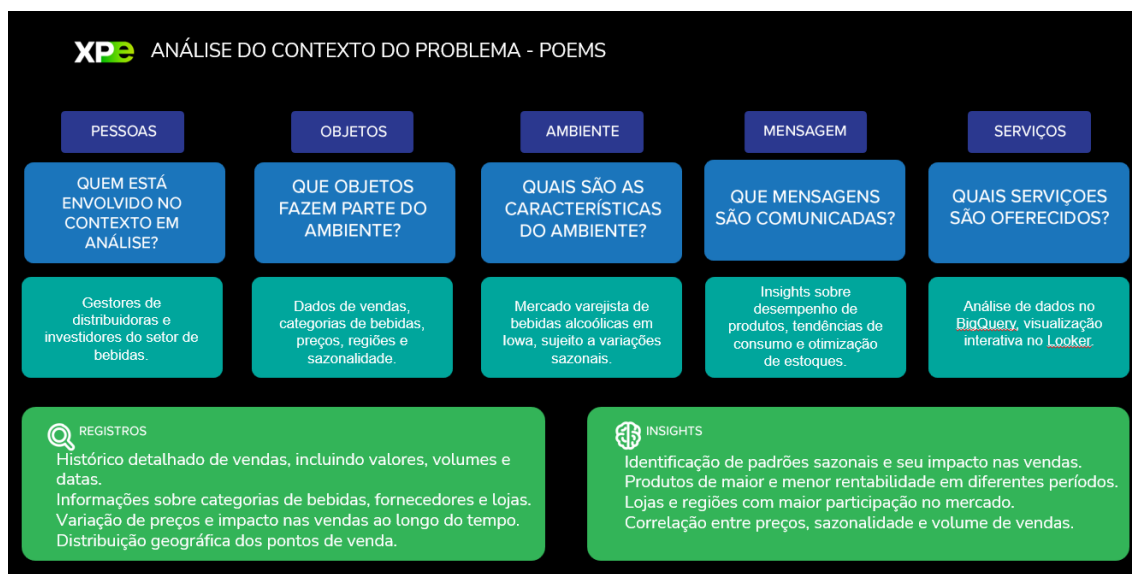


Figura 3 - POEMS

1.1.2 Personas

A persona referenciada neste projeto será Lorival Macêdo, 45 anos, Gestor de Vendas em Distribuidora de Bebidas há 5 anos e atua na área há 25 anos. Formado em Administração de Empresas, hoje é responsável por decisões estratégicas distribuição.

- Características Comportamentais:
 - Focado em resultados, sempre buscando otimizar a margem de lucro.
 - Valoriza a análise de dados para tomar decisões mais precisas.
 - Proativo na busca por soluções que melhorem a performance da equipe e o controle de estoques.
- Desafios e Obstáculos:
 - Tem dificuldades em entender e prever as flutuações sazonais do mercado de bebidas.
 - Enfrenta desafios com a falta de ferramentas precisas para otimização de estoque e previsões de vendas.
- Necessidades e Motivação:
 - Busca soluções que permitam prever a demanda com mais precisão, ajudando na gestão de estoques e evitando excessos ou faltas de produtos.
 - Necessita de dados mais detalhados sobre as preferências regionais de consumo e sazonalidade para melhorar a distribuição dos produtos.



Figura 4 - Mapa de Empatia

1.1.3 Justificativas

Conforme demonstrado anteriormente, gestores de vendas em distribuidoras de bebidas enfrentam desafios significativos na previsão de demanda e na otimização de estoques. A falta de ferramentas precisas e a dificuldade em interpretar padrões de consumo sazonais impactam diretamente a margem de lucro e a eficiência da distribuição.

Para solucionar esse problema, este projeto aplicará técnicas de Ciência de Dados para identificar padrões de consumo, sazonalidade e preferências regionais. Com isso, será possível gerar insights estratégicos que auxiliem os gestores a prever demandas com maior precisão, evitando excessos ou faltas de produtos.

Além disso, a análise aprofundada dos dados permitirá a identificação de oportunidades de melhoria na gestão de estoques e na distribuição, reduzindo desperdícios e aumentando a rentabilidade. Dessa forma, os gestores poderão tomar decisões embasadas em dados concretos, otimizando processos e garantindo maior competitividade no mercado.

A Tabela 1 apresentar o mapeamento das ações de Lorival, conforme a metodologia Blueprint, permitindo entender melhor o problema.

Itens	Detalhamento
Objetivos	Melhorar a previsão de demanda e otimizar a gestão de estoques na distribuidora de bebidas.
Atividades	Coletar e analisar dados de consumo, desenvolver modelos descritivos e gerar insights para suporte à decisão.
Questões	Como prever com precisão a demanda sazonal? Quais fatores influenciam o consumo em diferentes regiões?
Barreiras	Falta de dados organizados, resistência à adoção de novas tecnologias e variações imprevisíveis do mercado.
Itens	Ações do cliente
Funcionalidades	Dashboard interativo, relatórios personalizados e alertas sobre tendências de demanda.
Interação	Acesso via plataforma online, filtros personalizados e notificações automatizadas.
Mensagem	Insights baseados em dados para otimizar a distribuição e evitar desperdícios.
Onde ocorre	Aplicação web, e-mails com alertas e reuniões estratégicas.
Tarefas pendentes	Refinar a análise descritiva, integrar novas fontes de dados e treinar equipe na utilização da ferramenta.
Tarefas escondidas	Processamento de dados, limpeza de informações e ajuste de algoritmos.



Processo de suporte	Atendimento ao cliente, suporte técnico e atualizações do sistema.
Saída desejável	Melhor previsibilidade de demanda, redução de custos operacionais e aumento da margem de lucro.

Tabela 1 - Blueprint



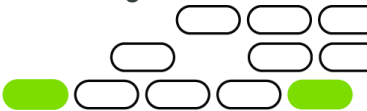
Figura 5 - Explicação de Proposição de Valor

1.1.4 Hipóteses

Já identificamos que um dos principais desafios enfrentados por gestores de vendas em distribuidoras de bebidas é a dificuldade em prever a demanda e equilibrar os estoques de forma eficiente. Muitas decisões são baseadas em experiência e intuição, o que pode levar a excessos ou faltas de produtos, impactando a margem de lucro e a satisfação dos clientes.

Além disso, a ausência de ferramentas analíticas avançadas limita a capacidade de identificar padrões de consumo e sazonalidade, dificultando a otimização da distribuição. Isso gera desperdícios e aumenta custos operacionais, reduzindo a competitividade da empresa no mercado.

Uma hipótese relevante para esse estudo é que a utilização de análise descritiva e inteligência de dados pode fornecer previsões mais precisas, permitindo uma melhor alocação de produtos e uma gestão de estoque mais eficiente. Com relatórios detalhados e insights baseados em dados reais, os gestores poderão tomar decisões mais embasadas, reduzindo desperdícios e maximizando os lucros.



Observação	Descrição da hipótese	Criterização			Somatório	Priorização	Grau de risco
		G	U	T			
Falta de previsibilidade na demanda	Dificulta na tomada de decisão e perdas financeiras	5	5	4	14	Alta	Alto
Falta de integração entre os sistemas de dados	Informações isoladas dificultam análises precisas e reduzem eficiência operacional	4	5	4	13	Alta	Alto
Dificuldade dos usuários em interpretar os dados	A equipe tem dificuldade em analisar os relatórios, impactando decisões estratégicas	4	4	5	13	Média-Alta	Médio
Decisões tomadas sem embasamento estatístico	O uso de intuição em vez de dados concretos compromete a eficiência das estratégias	3	4	5	12	Média	Médio
Falta de automação nos relatórios gerenciais	Relatórios manuais consomem tempo e aumentam risco de erros	3	3	5	11	Média	Baixo

Tabela 2 - Tabela de priorização

1.2 Solução

1.2.1 Objetivo SMART

Este estudo tem como objetivo fornecer informações estratégicas e conhecimento aprofundado para gestores de vendas no setor de distribuição de bebidas, ajudando-os a tomar decisões mais assertivas. A análise de dados sobre padrões de consumo e sazonalidade será organizada para oferecer insights que otimizem a gestão de estoques e a distribuição de produtos.

A expectativa é que a empresa consiga reduzir em pelo menos 20% as perdas com excesso ou falta de produtos, melhorando o planejamento e a eficiência logística. Isso será possível ao prever com maior precisão as variações sazonais e as preferências regionais de consumo, permitindo um controle mais eficaz do estoque.

Os primeiros resultados deverão ser percebidos em até três meses, à medida que os insights forem aplicados na gestão. O impacto será maximizado conforme o gestor implementar as recomendações na sua estratégia de distribuição e otimização de estoque.



1.2.2 Escopo do Projeto

Este estudo foi desenvolvido considerando algumas premissas e restrições identificadas ao longo do levantamento de necessidades e definições do escopo do projeto.

Premissas

1. O Looker Studio suportará a implementação da solução sem limitações críticas.
2. Consultas SQLs serão otimizadas para melhor desempenho e extração de insights.
3. O gestor da distribuidora, bem como toda a empresa, terá fácil acesso ao dashboard para realizar o acompanhamento.
4. O acesso aos dados será possível com acesso gratuito.
5. A interatividade do dashboard trará agilidade para acompanhamento estratégico, analítico e operacional.

Caso essas premissas não sejam verdadeiras, o projeto poderá enfrentar dificuldades na obtenção de insights precisos, na adoção por parte dos usuários ou na viabilidade técnica de implementação.

Restrições

1. Tempo para projeto: O tempo é um fator limitante para o desenvolvimento de uma solução robusta com análises profundas. Cabendo a este projeto apresentar um Mínimo Produto Viável (MVP).
2. Base de dados bruta: A quantidade de dados no banco é elevada e impossibilita o consumo sem prejuízo na performance.

Diante dessas restrições, será necessário adotar estratégias para mitigação de riscos e alternativas para garantir que os objetivos do projeto sejam alcançados dentro do prazo e escopo definidos.



1.2.3 Cronograma de Ações Planejadas

Nesta etapa, será descrito detalhadamente cada passo que deve ser realizado para a elaboração da proposta final deste trabalho. Utilizamos o [Trello \(Figura 6\)](#) para demonstrar os passos do backlog do produto, o que está em andamento e o que já está concluído.

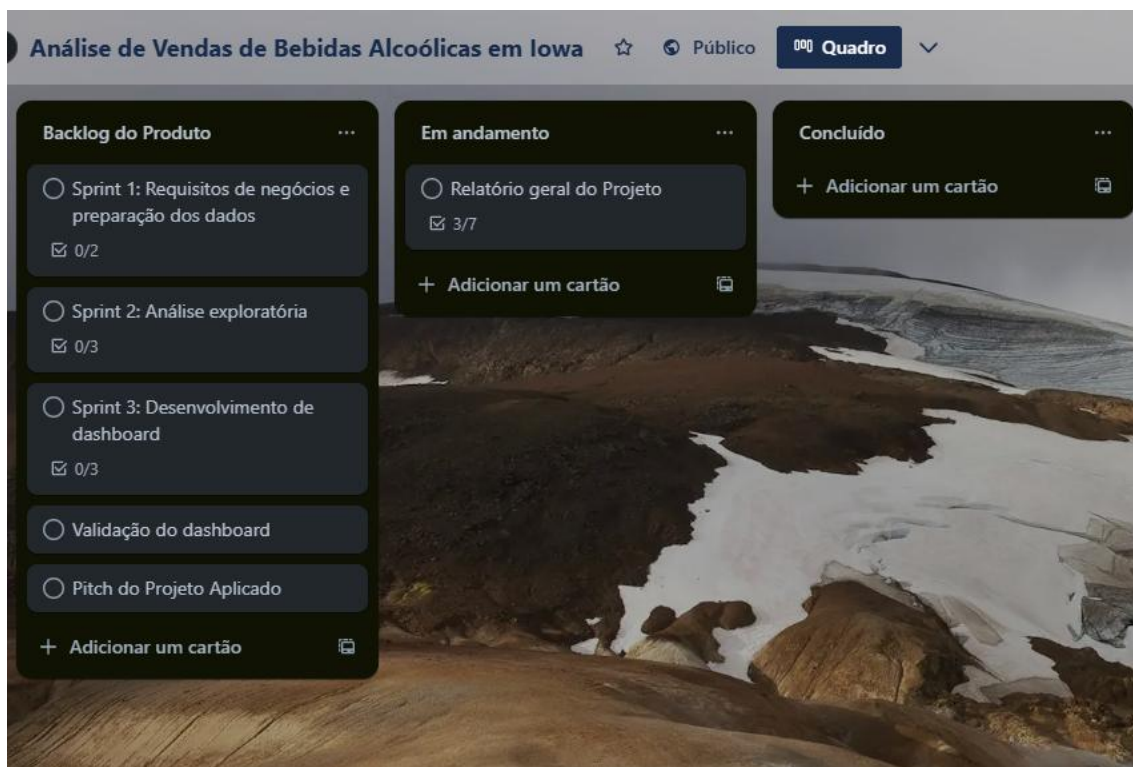


Figura 6 - Trello

2. Área de Experimentação

Esta seção tem a finalidade de mostrar a execução das atividades propostas no Cronograma. Serão mostrados em detalhes todas as etapas, divididas em 03 (três) Sprints. Cada um deles representa um conjunto de tarefas a serem executadas, de forma que, juntos, representam a solução do problema proposto.

A ferramenta escolhida para isso foi o Trello, onde através dele, todas as tarefas foram divididas. Cada cartão representa uma etapa ou Sprint. Dentro de cada um, estão as anotações e checklist das atividades.



2.1 Sprint 1

2.1.1 Solução

- Evidência da execução de cada requisito:

Para estudar a base, foram utilizados o BigQuery e a Documentação da base de dados, para identificar as colunas que podem ser usadas para montar o Dashboard. Com isso, foi montada a consulta inicial que retorna as colunas selecionadas, sendo retirada a coluna “state_bottle_cost”, pois ela traz o valor que o Estado paga aos fornecedores. O Estado realiza a compra das bebidas e revende para as lojas, apresentando assim o controle comercial sobre as bebidas alcoólicas no estado de Iowa. Estes valores comerciais não serão necessários nesta abordagem.

```
select date, store_name, city, store_location, county, category_name,  
vendor_name, item_description, pack, bottle_volume_ml, state_bottle_cost,  
bottles_sold, volume_sold_liters  
from `bigquery-public-data.iowa_liquor_sales.sales`
```

Avaliando as colunas e identificando oportunidades de insights, foi elaborado um rascunho dos requisitos de Dashboard. Essa etapa refere-se a uma entrevista com o cliente, onde se entende as dores do mesmo em identificar oportunidades de negócios e como uma visualização de dados pode ajudá-lo, nisso são listados indicadores importantes, formas de visualização e análises esperadas. Cabe aqui a individualidade do Analista/Cientista de dados em gerar valor onde o cliente não viu, a fim de aumentar o impacto com o produto de dados. O rascunho da visualização pode ser feito a mão ou em algum software como PowerPoint e não necessariamente a visualização listada lá será igualmente aplicada, pois no tratamento de dados pode ser necessário utilizar uma outra, servindo assim como guia inicial. É apresentado na Figura 7 o rascunho do painel.

Para atender os requisitos levantados, os dados foram tratados usando o próprio BigQuery, utilizando SQL estruturado em CTE (Common Table Expression), conforme a seguir:

```
with  
tb_location as (/*garantir de pegar o registro de localização mais recente, uma  
vez que a loja pode mudar o endereço  
além disso, registros de Latitude e Logitude costumam ter variações nas casas  
decimais de um registro pra outro*/  
select store_name, store_location  
from(  
select store_name, store_location,  
row_number() over (partition by store_name order by date desc) as  
rank_store  
from `bigquery-public-data.iowa_liquor_sales.sales`  
) as base  
where rank_store = 1  
,  
tb_sales as (  
select date_trunc(date, MONTH) AS date,  
case  
when bottles_sold < 0 then 'Refund'  
else 'Sale'
```



```
end as transaction_type, store_name, city, county,
category_name, vendor_name, item_description,
round(sum(abs(state_bottle_cost * bottles_sold)),2) as sales_cost,
round(sum(abs(volume_sold_liters)),2) as volume_sold_liters
from `bigquery-public-data.iowa_liquor_sales.sales`
group by 1, 2, 3, 4, 5, 6, 7, 8
)
select tb_sales.*, tb_location.store_location
from tb_sales
left join tb_location
on tb_sales.store_name = tb_location.store_name
```

Esta consulta foi carregada com sucesso no Looker Studio, conforme apresentado na Figura 8, finalizando assim a primeira sprint. Com isso, os dados estão mais limpos e tratados, o que facilitará os trabalhos nas próximas etapas. Na etapa de Análise exploratória podem surgir pequenos tratamentos a fim de atingir o objetivo de apresentar os indicadores.

- Evidência dos resultados:

A seguir, o rascunho da visualização segundo os requisitos levantados estudando a base inicial. Assim, são traçados os objetivos da visualização, onde será realizado alguma mudança caso julgue necessário. O painel apresenta apenas dois campos de filtros, enquadrando em um painel Analítico/Tático. Essa tomada de decisão foi aplicada para que pudesse explorar os cenários em cima de uma única distribuidora por vez.



Figura 7 - Rascunho do dashboard



A seguir, apresentação da carga de dados da consulta, sendo realizada com sucesso. Assim, pode-se realizar a etapa de Análise exploratória dos dados para a elaboração dos indicadores levantados na etapa de requisitos.

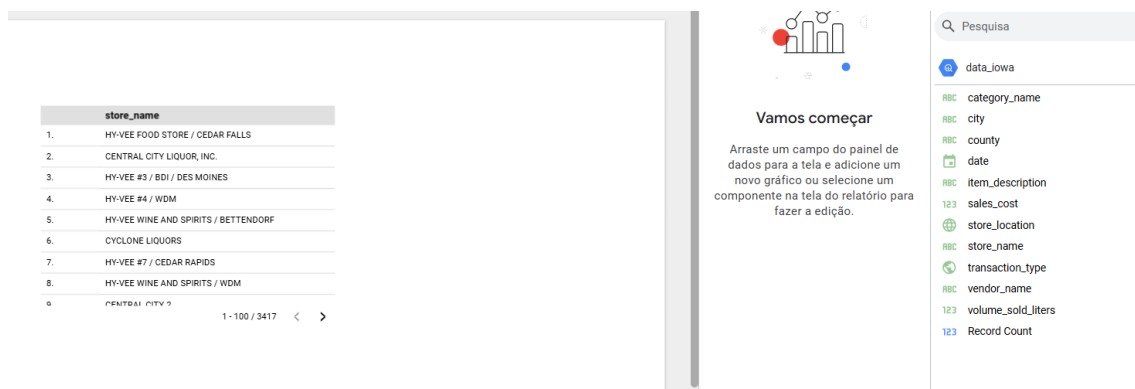


Figura 8 - Carga de dados no Looker

2.1.2 Retrospectiva da Sprint

Na Sprint 1 foi observada a importância de elaborar requisitos do produto de dados antes de começar a tratar os dados em si. Isso poupa trabalho e retrabalho com o tratamento, além disso, facilitará a montagem dos dados no dashboard. Essa etapa é crucial no projeto, exigindo muita empatia com o cliente, quem vai usar os dados, para resolver os problemas dele com o produto e não apenas fornecer um Relatório arrumado.

É importante destacar que uma boa escrita de SQL auxilia na performance dos dados, sendo utilizado aqui as estruturas CTEs que carregam tabelas temporárias na consulta. Observa-se que realizou um agrupamento de dados, isso reduz a quantidade de linhas da tabela final, o que torna leve a carga de dados no Looker. A utilização de uma CTE a parte para coletar a “store_location” foi importante pois registros de latitude e longitude podem variar com os registros.

Conclui-se assim, a sprint 1 com sucesso.

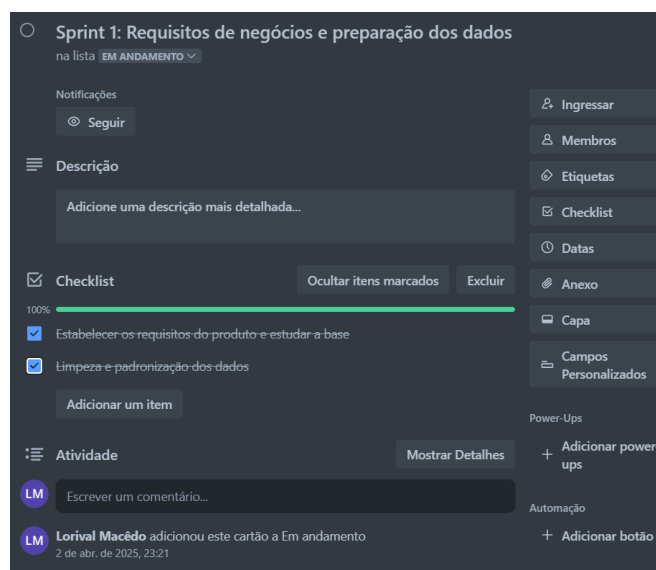


Figura 9 - Conclusão Sprint 1

2.2 Sprint 2

2.2.1 Solução

- Evidência da execução de cada requisito:

Na montagem dos indicadores levantados nos requisitos, foi possível conhecer a base a fundo, o volume de dados, possíveis erros e ajustes necessários na query ou no tratamento final dentro do Looker.

Dos indicadores que serão apresentados, a quantidade de transações de vendas não havia sido adicionada aos requisitos, bem como a receita por transação. Estes são KPIs importantes para uma análise gerencial. A contagem de transações veio agregada na query, a qual será apresentada nesta sessão.

Na identificação dos dados geográficos dos condados em Iowa, foi avaliado a necessidade de ajustar na query os nomes, como por exemplo, de “POLK” para “POLK COUNTY, IA” assim ajustou o reconhecimento do Looker para os condados do estado de forma adequada, porém, ainda assim, identificou-se o condado “EL PASO” fora de Iowa, próximo ao Texas. A cidade identificada para este condado foi “Colorado Springs”. Realizando pesquisas na internet, confirmou-se que não existe este condado em Iowa, sendo assim retirado da base, o que impactou em apenas 2 registros. Registros assim podem ter sido por erros na gravação do dado.

A seguir, apresenta-se a query modificada para os ajustes adicionais:

```
with
tb_location as (/*garantir de pegar o registro de localização mais recente, uma
vez que a loja pode mudar o endereço
além disso,registros de Latitude e Logitude costumam ter variações nas casas
decimais de um registro pra outro*/
select store_name, store_location
from(
select store_name, store_location,
row_number() over (partition by store_name order by date desc) as
rank_store
from `bigquery-public-data.iowa_liquor_sales.sales`
) as base
where rank_store = 1
),
tb_sales as (
select date_trunc(date, MONTH) AS date,
case
when bottles_sold < 0 then 'Refund'
else 'Sale'
end as transaction_type, store_name, city, concat(county, ' COUNTY, IA') as
county, -- AJUSTE REALIZADO
category_name, vendor_name, item_description,
round(sum(abs(state_bottle_cost * bottles_sold)),2) as sales_cost,
round(sum(abs(volume_sold_liters)),2) as volume_sold_liters,
count(*) as transaction_count -- NOVA COLUNA ADICIONADA
from `bigquery-public-data.iowa_liquor_sales.sales`
where county not in ('EL PASO') -- FILTRO DE EXCLUSÃO ADICIONADO
group by 1, 2, 3, 4, 5, 6, 7, 8
)
select tb_sales.*, tb_location.store_location
```




```
from tb_sales
left join tb_location
on tb_sales.store_name = tb_location.store_name
```

Nesta etapa de geração dos KPIs segundo os requisitos levantados na etapa anterior, foram necessários criar alguns campos calculados, adicionar cards numéricos ou tabelas para verificar a correta aplicação de cálculos. Além disso, a decisão de escolha de categorias como “category_name” e “item_description” passaram pela análise de quantos campos terão as visualizações. Nesta etapa as visualizações são meramente temporárias, podendo mudar para a versão final o tipo de apresentação.

A seguir, campos calculados que foram criados para gerar indicadores:

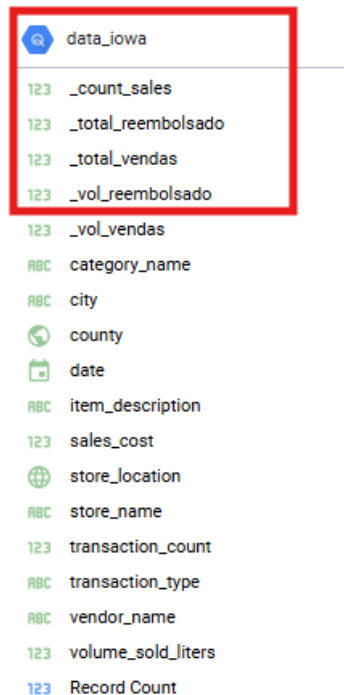
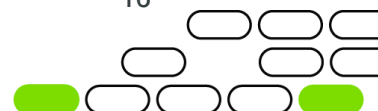


Figura 10 - Geração de campos calculados

- Evidência dos resultados:

A seguir, a apresentação do dashboard com os indicadores desejados, dentro da etapa exploratória dos dados. Observa-se pela taxa de churn e pela retenção da receita (lucro bruto dividido pela receita), que as distribuidoras não precisam se preocupar com o nível do reembolso de forma tão urgente. Há uma grande aceitação das vendas realizadas, havendo casos mais isolados de reembolso. Além disso, a tendência temporal é de aumento consistente, garantido às distribuidoras receita para o futuro. Dentro da linha temporal, observa-se sazonalidades que serão melhores estudadas mais a frente.

Na tabela com “store_name” com lucro, ticket médio e volume vendido, observa-se que as duas primeiras lojas compram mais que o dobro da terceira maior compradora, ambas são do condado Polk, condado que gerou maior lucro acumulado. A quantidade de lojas é elevada, o que torna ruim a visualização no gráfico. Adiantando-se assim, a possibilidade de um tratamento direcionado em agrupar as



lojas por franquias que apresenta um trabalho árduo e suscetível a erros, além de direcionar a análise para uma outra visão.

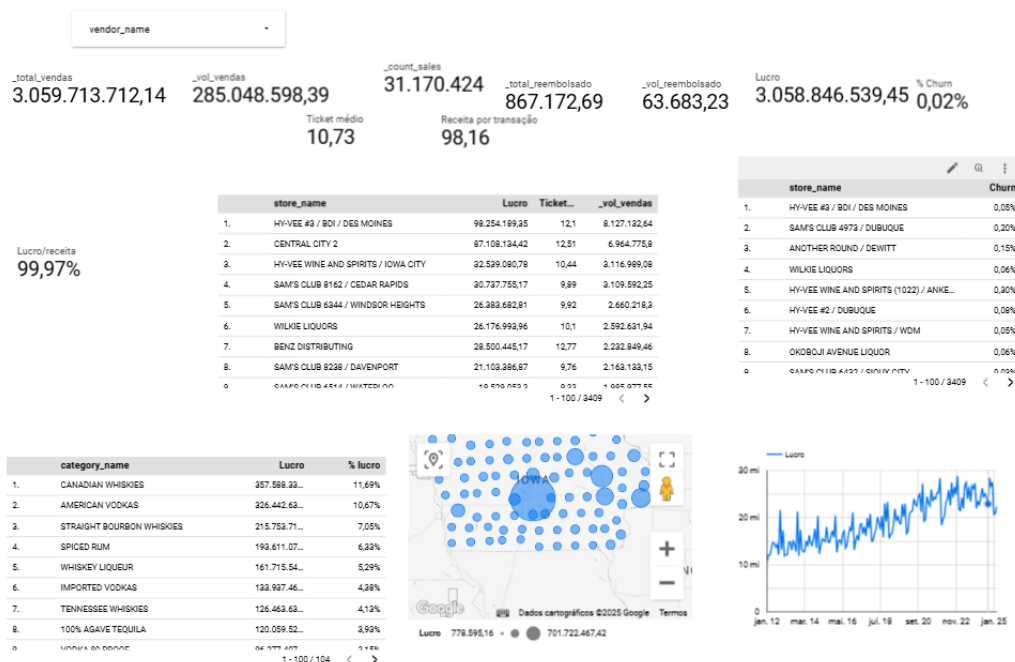


Figura 11 - Preview de indicadores

2.2.2 Retrospectiva da Sprint

Na Sprint 2 foi avaliado que com a análise exploratória, consegue-se absorver insights iniciais bem como identificar ajustes adicionais ao tratamento de dados, não havendo problema se for resolvido na query ou dentro do próprio Looker. Além disso, observou-se que esta base possui uma grande população de itens como “store_name” e “category_name” o que para trazer uma visualização enxuta de poucas categorias exige um trabalho manual árduo de identificação e agregações.

Conclui-se assim, a sprint 2 com sucesso.

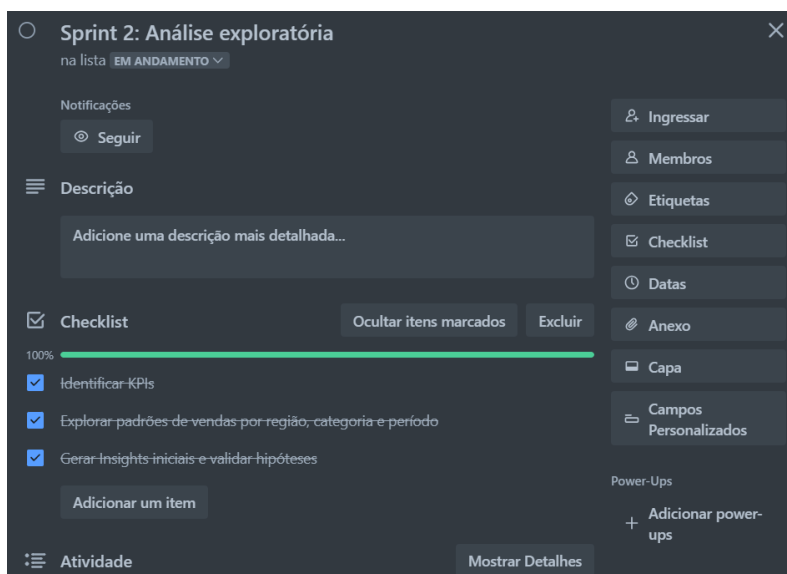


Figura 12 - Conclusão da Sprint 2

2.3 Sprint 3

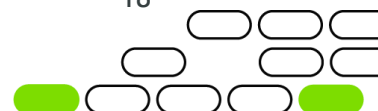
2.3.1 Solução

- Evidência da execução de cada requisito

Nesta etapa, o objetivo era criar as visualizações, refinando os dados, criando segmentos e explorando insights. Com esta etapa o acesso ao Bigquery apresentou sua limitação, pois o Looker atualizava os dados com as diversas interações, assim, o limite de extração dos dados no Bigquery atingia com facilidade. Essa limitação tornava o dashboard sem atualização, porém é diária, podendo ser usado no dia seguinte. A fim de tornar usual o relatório, filtrou-se para apenas o distribuidor “Diageo Americas”, escolhido de forma aleatória. Além disso, o histórico de dados foi restringido para o último mês carregado, nesta base, na data do dia 24 de abril de 2025, possui os dados do mês anterior, março de 2025. Os dados de latitude e longitude foram removidos, uma vez que nem estavam no plano de utilizar, devido a alta granularidade de lojas. A seguir a query final simplificada comparada as versões anteriores

```
WITH max_date_select AS (
  SELECT MAX(date) AS date_max
  FROM `bigquery-public-data.iowa_liquor_sales.sales`
),
tb_sales AS (
  SELECT
    s.date,
    CASE
      WHEN s.bottles_sold < 0 THEN 'Refund'
      ELSE 'Sale'
    END AS transaction_type,
    s.store_name,
    s.city,
    CONCAT(s.county, ' COUNTY, IA') AS county,
    s.category_name,
    s.vendor_name,
    s.item_description,
    ROUND(SUM(ABS(s.state_bottle_cost * s.bottles_sold)), 2) AS sales_cost,
    ROUND(SUM(ABS(s.volume_sold_liters)), 2) AS volume_sold_liters,
    COUNT(*) AS transaction_count
  FROM `bigquery-public-data.iowa_liquor_sales.sales` s
  CROSS JOIN max_date_select
  WHERE s.county NOT IN ('EL PASO')
    AND s.vendor_name = 'DIAGEO AMERICAS'
    AND s.date BETWEEN DATE_TRUNC(date_max, MONTH) AND LAST_DAY(date_max)
  GROUP BY 1, 2, 3, 4, 5, 6, 7, 8
)
SELECT *
FROM tb_sales
```

O indicador de lucro foi retirado do valor geral na parte superior do relatório, pois o valor do reembolso não é tão relevante assim para gerar grande diferença ao descontar. Dessa forma, o indicador lucro/receita não se fez tão necessário, abrindo a chance do churn ser explorado de forma mais visual com um gráfico de velocímetro. Enquanto isso, os valores gerais “Receita por loja” e “Lojas” (quantidade) foram adicionados para somar à análise.



Devido à grande quantidade de lojas na base de dados, a análise de Lucro x Ticket médio ficou focada no TOP 10 de lucro entre lojas, sendo esse o cohort para o churn. Enquanto isso, para a análise com gráfico de Pareto para o Lucro ou Volume vendido para cada item dos produtos, a visualização estará apresentando o TOP 20, o que será suficiente para apresentar os 80% do impacto avaliado. O gráfico de mapa com bolhas ficou bem ajustado com o tratamento apresentando boa visualização de insights. Já a tendência histórica, ao longo do tempo, precisou ficar ajustada para dias, ao invés de mês com a redução da base de dados, o que para este MVP atende bem.

- Evidência dos resultados

A seguir o relatório final com os ajustes informados anteriormente e com toda a arrumação profissional: [Dashboard comercial](#). Ele pode servir como estratégico pois o único filtro para interação com o usuário é alternar algumas visualizações de “Lucro” para “Volume vendido”, além da interação entre gráficos permitida pela ferramenta.



Figura 13 - Dashboard estratégico

A seleção entre Lucro e Volume precisou ser em caixas, pois a ferramenta apresentou limitação para criar botões para este controle.

As informações ficaram bem claras e com fácil leitura para analisar de forma estratégica. Os insights serão debatidos nas conclusões, apresentando a utilidade de um produto de dados como este.

2.3.2 Retrospectiva da Sprint

Nesta sprint enfrentou-se o principal problema com recursos gratuitos, o limite de extração de dados. Para isso, os dados foram restringidos em 1 mês apenas e para apenas 1 distribuidora de bebidas. Isso permitiu criar as visualizações, testar os resultados e vai permitir que o relatório seja atualizado com maior frequência sem

perder as conexões dos dados. Com o cumprimento da criação das visualizações, os insights serão registrados nas considerações finais, para então encerrar a sprint 3.

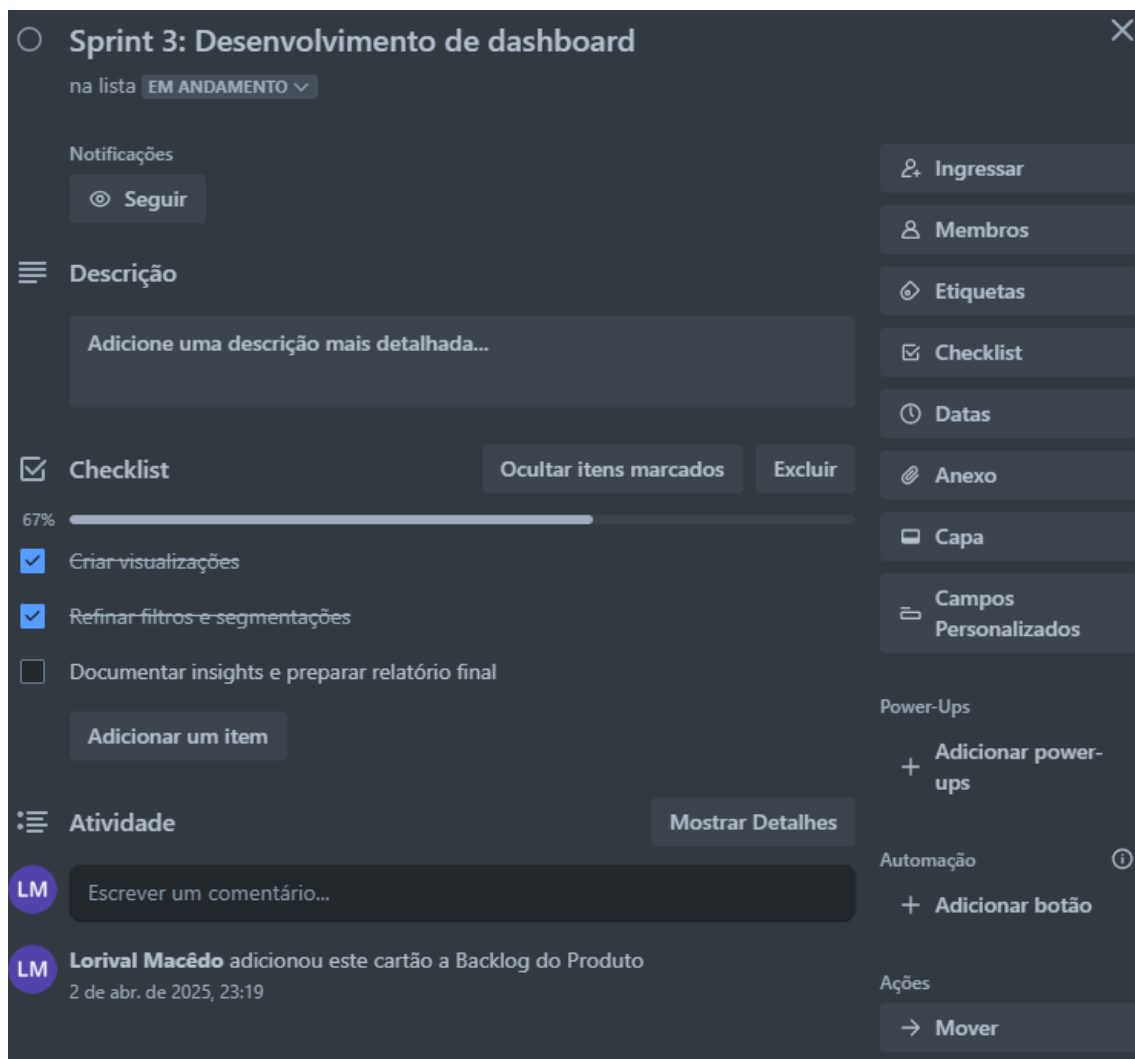


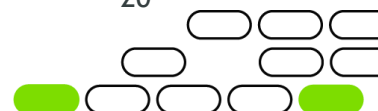
Figura 14 - Progresso Sprint 3

3. Considerações Finais

3.1 Resultados

Os dados coletados estão em fonte pública, disponíveis via BigQuery. Isso foi de grande consideração pois com o uso de ferramentas gratuitas permitiu trabalhar com Análise de Dados e geração de BI de ponta a ponta. Utilizando esses dados no Looker os principais desafios encontrados foram:

- Limitação de extração de dados diária
- Personalização do Looker frente ao Power BI ferramenta com maior potencial



O dashboard respondeu bem, trazendo uma análise enxuta e eficiente, podendo explorar insights. Assim, serve como modelo de visualização para projetos em diversos problemas de negócios, principalmente para o varejo de bebidas. Com o Looker, o acesso online ao relatório é fácil e gratuito, representando grande vantagem em custo-benefício no mercado de ferramentas de BI.

Da visualização a seguir, podemos retirar de insights quais são as 10 lojas que retornaram mais lucro para a distribuidora “Diageo Americas”, é interessante notar que com várias lojas a receita por litro de bebida (ticket médio) é abaixo do valor geral. Isso representa que algumas lojas comprem grandes volumes de bebidas mais baratas, o que pode estar relacionado à fatores socioculturais dos clientes.

Uma grande oportunidade se mostra com a HY-VEE #3, pois ela possui um churn de volume de bebidas de 1,31% o que pode parecer pouco, mas clicar nesta coluna, nota-se que representam 117 litros o que resultou em US\$ 2447 reembolsado em um mês. Isso ao longo de um ano acumulará bom valor financeiro, precisando assim, ser controlado a qualidade a ser entregue. Esse valor do churn se mostra elevado, pois do TOP 10 lojas que retornaram maior lucro, as outras lojas tiveram valores aproximadamente nulos.



O gráfico de Pareto no canto inferior esquerdo apresenta informações extremamente importantes, que apenas esses 20 itens dos produtos (são mais de 100),

são suficientes para representar 80% do lucro ou 80% do volume, de acordo com o filtro desejado.

Com o gráfico de mapa, pode-se observar o lucro ou volume para cada condado, destacando que o condado Polk teve os maiores valores.

O gráfico de área traz a métrica diariamente, nele é fácil notar que os dias da semana Sábados e Domingos são bem reduzidos, com os sábados sendo nulos. Isso pode estar relacionado a uma medida da própria distribuidora ou legislação de Iowa, uma vez que o estado determina certos controles na região.

Diversos insights podem ser extraídos interagindo com os gráficos, o que supre uma demanda Analítico/Tático, bem como servir como um painel estratégico sem realizar interações.

O dashboard atingiu o objetivo servir insights de forma clara e objetiva para gestão. A base de dados é volumosa, com cada categoria tendo alta variedade, isso representou um bom desafio, sendo necessário tomadas de decisões para ajustes técnicos e estratégicos.

3.2 Próximos passos

Este dashboard pode ser melhorado com as seguintes medidas:

- Adição de uma página como relatório de análises mais detalhadas, focando até na operação.
- Adição de storytelling de dados, com textos instrutivos para quem for utilizar o dashboard.
- Abranger mais dados históricos, o que precisará entender como aumentar a base sem impactar no acesso do Looker ao Bigquery de forma gratuita.

Estas são melhorias que serão implementadas em um segundo momento, principalmente porque será um produto a ser utilizado no portfólio profissional, dado ao valor desenvolvido neste projeto e aplicação da teoria em um problema de negócio real.

Com este material, o projeto é encerrado com sucesso, atendendo os requisitos avaliativos e com satisfação pelo amadurecimento na jornada.



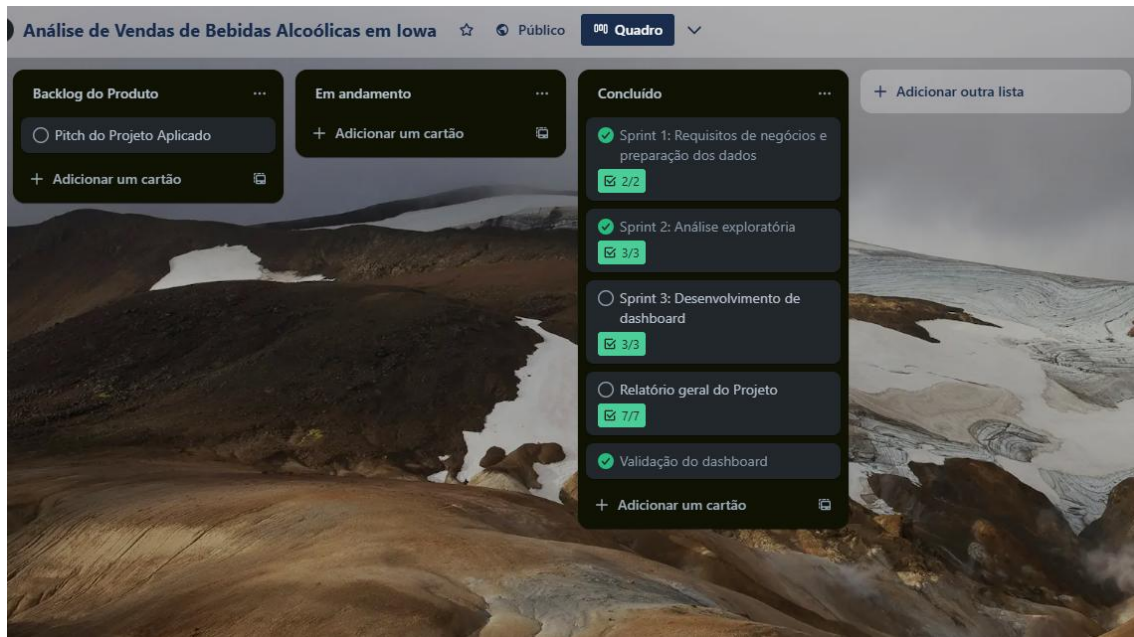


Figura 15 - Encerramento do projeto