

Projeto Blinkit MySQL: Documento Técnico

Súmano:

Súmano:-----	1
Introdução:-----	1
1. Demanda do Cliente:-----	2
2. KPI's de Interesse:-----	2
3. Objetivos específicos:-----	2
4. Requisições para a visualização:-----	3
5. As análises para os KPI's primários:-----	3
6. As análises para os objetivos específicos:-----	5

Introdução:

Este documento, intitulado “Projeto Blinkit MySQL”, foi desenvolvido com o objetivo de demonstrar e consolidar habilidades de análise de dados em SQL, especialmente aplicadas ao contexto de vendas e satisfação de clientes em uma plataforma de comércio eletrônico. O foco principal é a utilização prática de conceitos fundamentais, como limpeza de dados (identificação e padronização de valores inconsistentes), construção de consultas SQL para extração de indicadores (KPIs) e interpretação de resultados para gerar insights relevantes.

Ao longo do projeto, foram definidas métricas que incluem o total de vendas, a média de vendas, o número de itens vendidos e a média de avaliações dos clientes, bem como análises mais específicas, como vendas segmentadas pelo conteúdo de gordura dos produtos e pelo tipo de item. Essas análises fornecem uma visão aprofundada do desempenho e da satisfação do cliente, sendo fundamentais para subsidiar a tomada de decisões e identificar oportunidades de otimização no negócio.

As etapas de limpeza e padronização de dados, seguidas por consultas sistemáticas em SQL e a posterior visualização dos resultados, comprovam a aplicação de boas práticas na manipulação

e exploração de dados. Esse conjunto de processos e resultados evidencia a capacidade de geração de relatórios claros e objetivos, demonstrando competências valiosas para a atuação como analista de dados. Adicionalmente, no futuro esse projeto poderá servir como base para a criação de painéis em ferramentas como o Power BI para reforçar a importância de uma comunicação eficiente dos resultados e insights obtidos.

Em suma, o “Projeto Blinkit MySQL” apresenta uma solução completa para análise de dados em SQL, desde o tratamento inicial das bases até a comunicação dos insights de valor. Ele evidencia habilidades técnicas, senso analítico e comprometimento com a excelência, atuando como um importante marco no desenvolvimento profissional na área de Data Analytics.

1. Demanda do Cliente:

Realizar uma análise profunda da performance de vendas, satisfação dos clientes e de distribuição de inventário da plataforma BlinkIt, com o intuito de realizar insights cruciais e descobrir oportunidades de aumentar eficiência. Entregar as visualizações com KPI's e tabelas realizadas em MySQL.

2. KPI's de Interesse:

- Total de vendas
- Média das vendas
- Numero de itens
- Média das avaliações

3. Objetivos específicos:

- Total de vendas por conteúdo de gordura
- Total de vendas por tipo de item
- Conteúdo de gordura por estabelecimento para o total de vendas
- Total de vendas por idade de estabelecimento

4. Requisições para a visualização:

- Porcentagem de vendas por tamanho dos estabelecimentos
- Vendas por tier localização dos estabelecimentos
- Todas as métricas por tipo de estabelecimentos

5. As análises para os KPI's primários:

Vamos começar com a etapa de limpeza dos dados, uma vez que depois de uma breve análise exploratória dos dados foi possível detectar valores não constantes na coluna “Item Fat Content” e portanto devemos consertar e padronizar a coluna antes de iniciar as queries de análise.

Figura 1 - Limpeza dos dados

```
-- Primeira Etapa do projeto: Limpeza do dataset! --  
  
select *  
from blinkit_data;  
  
-- item fat content e regularizar essa coluna --  
update blinkit_data  
set `Item Fat Content` =  
case  
    when `Item Fat Content` in ('LF', 'low fat') then 'Low Fat'  
    when `Item Fat Content` = 'reg' then 'Regular'  
    else `Item Fat Content`  
end;  
  
select distinct `Item Fat Content`  
from blinkit_data;
```

Fonte: Autoria própria.

Após realizada essa simples limpeza, vamos começar com as partes mais fáceis das demandas do cliente, as KPI 's de interesse. Começando por o total do valor de vendas, nesse caso foi escolhido ser representado em milhões para facilitar a visualização.

Figura 2 - Total de Vendas em Milhões

```
-- Segunda Etapa: As análises do cliente --

-- Total de vendas --
select cast(sum(`Total Sales`)/1000000 as decimal(10,2)) as total_vendas_milhoes
from blinkit_data
;
```

Fonte: Autoria própria.

O resultado do query foi 1.54M de rúpias indianas. Seguindo a lista vem a média do valor das vendas:

Figura 3 - Média das Vendas

```
-- Média das vendas --

Select cast(avg(`Total Sales`) as decimal(10,2)) as media_vendas
from blinkit_data
;
```

Fonte: Autoria própria.

O resultado foi 180,87 rúpias indianas. Para descobrir o número de itens vendidos, é bem simples no caso deste dataset, já que cada linha significa a compra de um item no e-commerce e então precisamos apenas saber a quantidade linhas da tabela. Podemos fazer isso de várias maneiras mas para manter o padrão de realizar queries no MySQL iremos fazer assim:

Figura 4 - Número de Itens

```
-- Numero de itens vendidos --
select count(*)
from blinkit_data
;
```

Fonte: Autoria própria.

Para finalizar as queries com intuito de realizar as KPI's, veremos a média das avaliações dada pelo clientes da plataforma para cada item comprado:

Figura 5 - Média das Avaliações

```
-- Média das avaliações dos clientes pelo item comprado --  
  
select cast(avg(Rating) as decimal(10,2)) as media_avaliacao  
from blinkit_data;
```

Fonte: Autoria própria.

O resultado da média foi 3,96 estrelas.

6. As análises para os objetivos específicos:

Iniciando essa parte do projeto por a ordem apresentada nas demandas do cliente, veremos a soma dos valores de venda por tipo de conteúdo de gordura dos itens, ou seja, para “Low Fat” e para “Regular”. Mas como o próximo objetivo, o valor das vendas total por tipo do item, é bem semelhante (as queries seguem o mesmo princípio) a esse irei agrupar os dois e mostrar o resultado juntos.

Figura 6 - Total por Conteúdo de Gordura

```
-- Objetivos Especificos --  
  
-- Soma dos valores de vendas por conteudo de gordura do item --  
select `Item Fat Content`, cast(sum(`Total Sales`) as decimal (10,2) )as total_vendas_gordura  
from blinkit_data  
group by `Item Fat Content`  
order by total_vendas_gordura desc  
;  
  
-- podemos inclusive entregar informações adicionais, com base no que fizemos para os KPI's iniciais --  
  
select `Item Fat Content`,  
       cast(sum(`Total Sales`) as decimal (10,2) )as total_vendas_gordura,  
       cast(avg(`Total Sales`) as decimal(10,2)) as media_vendas,  
       count(*) as qntd_itens,  
       cast(avg(Rating) as decimal(10,2)) as media_avaliacao  
from blinkit_data  
group by `Item Fat Content`  
order by total_vendas_gordura desc  
;
```

Fonte: Autoria própria

Tabela 1 - Tabela de Output de Vendas por Gordura

	Item Fat Content	total_vendas_gordura	media_vendas	qntd_itens	media_avaliacao
►	Low Fat	958567.17	173.75	5517	3.96
	Regular	582962.30	193.93	3006	3.95

Fonte: Autoria própria

Figura 7 - Total por Tipo do Item

```
-- Total do valor de vendas pelo tipo do item --

select `Item Type`,
       cast(sum(`Total Sales`) as decimal (10,2) )as total_vendas_tipo,
       cast(avg(`Total Sales`) as decimal(10,2)) as media_vendas,
       count(*) as qntd_itens,
       cast(avg(Rating) as decimal(10,2)) as media_avaliacao
from blinkit_data
group by `Item Type`
order by total_vendas_tipo desc
;
```

Fonte: Autoria própria

Tabela 2 - Tabela de Output do Total por Tipo de Item

	Item Type	total_vendas_tipo	media_vendas	qntd_itens	media_avaliacao
▶	Fruits and Vegetables	204778.44	166.22	1232	3.94
	Frozen Foods	173070.78	202.19	856	3.96
	Snack Foods	169296.19	141.08	1200	3.95
	Baking Goods	165590.19	255.54	648	3.98
	Dairy	141099.36	206.89	682	3.96
	Canned	133428.00	205.59	649	3.99
	Household	125477.91	137.89	910	4.00
	Health and Hygiene	91358.69	175.69	520	3.97
	Soft Drinks	83268.21	187.12	445	3.91
	Meat	82152.73	193.30	425	4.00
	Breads	54608.37	217.56	251	3.86
	Others	34089.48	201.71	169	3.97
	Starchy Foods	31603.54	213.54	148	3.92
	Hard Drinks	23704.30	110.77	214	3.87
	Breakfast	20403.20	185.48	110	3.93
	Seafood	7600.07	118.75	64	3.92

Fonte: Autoria própria

Agora para realizar o próximo objetivo específico, foi feita uma alteração na visualização para melhor entendimento da tabela. Usando como base os códigos criados anteriormente foi realizado uma tabela pivot para demonstrar a o valor total das vendas por tier das localizações do BlinkIT e também por o tipo de conteúdo de gordura do item vendido, então temos:

Figura 8 - Total do Valor de Vendas por Tier e Conteúdo de Gordura

```
-- Conteúdo de gordura por estabelecimento para o total de vendas--
-- Para isso iremos realizar uma pivot table, para facilitar a visualização. --
-- Podemos nos basear no código do objetivo passado e realizar algumas mudanças: --

select `Outlet Location Type`,
cast(sum(case
when `Item Fat Content` = 'Low Fat' then `Total Sales` else 0 end)as decimal(10,2)) as `Baixa Gordura`,
cast(sum(case
when `Item Fat Content` = 'Regular' then `Total Sales` else 0 end)as decimal(10,2)) as `Regular`
from blinkit_data
group by `Outlet Location Type`;
```

Fonte: Autoria própria

Tabela 3 - Tabela de Output das Vendas por Tier e Conteúdo de Gordura

	Outlet Location Type	Baixa Gordura	Regular
►	Tier 1	267810.94	164633.84
	Tier 3	375069.97	237613.62
	Tier 2	315686.27	180714.84

Fonte: Autoria própria

Como o projeto foi feito em MySQL, que não possui a keyword PIVOT como outras databases, a query foi feita de maneira não ortodoxa mas o resultado alcançado é igual. Para finalizar os objetivos específicos realizou-se a análise de vendas a partir do ano de inauguração dos estabelecimentos:

Figura 9 - Total de Vendas por Idade do Estabelecimento

-- Total de vendas por idade de estabelecimento--

```
select `Outlet Establishment Year`,
       cast(sum(`Total Sales`) as decimal (10,2)) as total_vendas
from blinkit_data
group by `Outlet Establishment Year`
order by `Outlet Establishment Year` asc;
```

Fonte: Autoria própria

Tabela 4 - Tabela de Output do Total de Vendas por Idade

	Outlet Establishment Year	total_vendas
►	1998	269220.02
	2000	172468.83
	2010	172705.25
	2011	95581.24
	2012	164234.12
	2015	167797.80
	2017	160947.52
	2020	167655.79
	2022	170918.91

Fonte: Autoria própria

7. As análises requisitada para as visualizações:

O cliente fez algumas demandas em que a visualização dos dados tem que ser a prioridade, portanto foi realizado queries no MySQL com o intuito de deixar a visualização dos “outputs” o mais eficiente e clara o possível. Começamos com a demanda de porcentagem de vendas por tamanho dos estabelecimentos:

Figura 10 - Porcentagem de Vendas por Tamanho dos Estabelecimentos

```
-- Requisitos de visualização --

-- Porcentagem de vendas por tamanho dos estabelecimentos --

SELECT
  `Outlet Size`,
  CAST(SUM(`Total Sales`) AS DECIMAL(10,2)) AS total_vendas,
  CAST((SUM(`Total Sales`) * 100.0 / SUM(SUM(`Total Sales`)) OVER())) AS DECIMAL(10,2)) AS porcentagem_vendas
FROM blinkit_data
GROUP BY `Outlet Size`
ORDER BY total_vendas DESC;
```

Fonte: Autoria própria

Tabela 5 - Tabela de Output da Porcentagem de Vendas por Tamanho dos Estabelecimentos

	Outlet Size	total_vendas	porcentagem_vendas
▶	Medium	662661.82	42.99
	Small	558934.41	36.26
	High	319933.24	20.75

Fonte: Autoria própria

Após realizar a porcentagem, verificou-se o valor total das vendas por o tier de estabelecimento:

Figura 11 - Total de Vendas por o Tier de Estabelecimento

```
-- Vendas por tipo de estabelecimento --

SELECT `Outlet Location Type`,
  CAST(SUM(`Total Sales`) AS DECIMAL(10,2)) AS total_vendas
FROM blinkit_data
GROUP BY `Outlet Location Type`
ORDER BY total_vendas DESC;
```

Fonte: Autoria própria

Tabela 6 - Tabela de Output de Vendas por o Tier de Estabelecimento

	Outlet Location Type	total_vendas
►	Tier 3	612683.58
	Tier 2	496401.11
	Tier 1	432444.78

Fonte: Autoria própria

E então foi realizado a análise de todas as métricas de KPI's iniciais por tipo de estabelecimento:

Figura 12 - Todas as Métricas por Tipo de Estabelecimentos

```
SELECT `Outlet Type`,
CAST(SUM(`Total Sales`) AS DECIMAL(10,2)) AS total_vendas,
CAST(AVG(`Total Sales`) AS DECIMAL(10,0)) AS media_vendas,
COUNT(*) AS qtnd_itens,
CAST(AVG(`Rating`) AS DECIMAL(10,2)) AS media_avaliacao,
CAST(AVG(`Item Visibility`) AS DECIMAL(10,2)) AS visibilidade_itens
FROM blinkit_data
GROUP BY `Outlet Type`
ORDER BY total_vendas DESC;
```

Fonte: Autoria própria

Tabela 7 - Tabela de Output de Todas as Métricas por Tipo de Estabelecimento

	Outlet Type	total_vendas	media_vendas	qtnd_itens	media_avaliacao	visibilidade_itens
►	Supermarket Type1	1005809.31	180	5577	3.95	0.06
	Grocery Store	191086.65	176	1083	3.98	0.10
	Supermarket Type3	173714.60	186	935	3.95	0.06
	Supermarket Type2	170918.91	184	928	3.95	0.06

Fonte: Autoria própria