

Análise de Vendas de Produtos Eletrônicos (2019 - 2020)

Objetivo

Este estudo consiste em uma análise exploratória a respeito das vendas de uma loja fictícia de produtos eletrônicos no período completo de 2019 até Janeiro de 2020. O objetivo principal integra a obtenção de insights por meio da análise do comportamento das vendas desta loja fictícia, possibilitando o aumento das vendas e por consequência o aumento do lucro.

Com os dados presentes, será criado um dashboard, utilizando o Power BI, para responder algumas questões importantes e um relatório em formato de apresentação.

Obtenção dos dados

Os dados que serão utilizados neste estudo foram obtidos no Github do Keith Galli.

Os dados podem ser acessado por meio do link:

<https://github.com/KeithGalli/Pandas-Data-Science-Tasks> (<https://github.com/KeithGalli/Pandas-DataScience-Tasks>)

Canal do Keith Galli no YouTube:

<https://www.youtube.com/c/KGMIT> (<https://www.youtube.com/c/KGMIT>)

Apresentação dos dados

Antes de começarmos o estudo é interessante observarmos os dados com que iremos trabalhar. Os meses em análise estão divididos em 12 arquivos **.csv**. Dentro

do Power BI faremos a importação de cada arquivo e os combinaremos para em seguida começar o processo de transformação dos dados.

	ABC 123	Order ID	ABC 123	Product	ABC 123	Quantity Ordered	ABC 123	Price Each	ABC 123	Order Date	ABC 123	Purchase Address
		<div><div>Válidos</div><div>100%</div><div>0%</div><div>0%</div></div>		<div><div>Válidos</div><div>100%</div><div>0%</div><div>0%</div></div>		<div><div>Válidos</div><div>100%</div><div>0%</div><div>0%</div></div>		<div><div>Válidos</div><div>100%</div><div>0%</div><div>0%</div></div>		<div><div>Válidos</div><div>100%</div><div>0%</div><div>0%</div></div>		<div><div>Válidos</div><div>100%</div><div>0%</div><div>0%</div></div>
1		176558		USB-C Charging Cable		2		11.95		04/19/19 08:46		917 1st St, Dallas, TX 75001
2												
3		176559		Bose SoundSport Headphones		1		99.99		04/07/19 22:30		682 Chestnut St, Boston, MA 02215
4		176560		Google Phone		1		600		04/12/19 14:38		669 Spruce St, Los Angeles, CA 90001
5		176560		Wired Headphones		1		11.99		04/12/19 14:38		669 Spruce St, Los Angeles, CA 90001
6		176561		Wired Headphones		1		11.99		04/30/19 09:27		333 8th St, Los Angeles, CA 90001
7		176562		USB-C Charging Cable		1		11.95		04/29/19 13:03		381 Wilson St, San Francisco, CA 94016
8		176563		Bose SoundSport Headphones		1		99.99		04/02/19 07:46		668 Center St, Seattle, WA 98101
9		176564		USB-C Charging Cable		1		11.95		04/12/19 10:58		790 Ridge St, Atlanta, GA 30301
10		176565		Macbook Pro Laptop		1		1700		04/24/19 10:38		915 Willow St, San Francisco, CA 94016

Como podemos observar, possuímos as seguintes colunas:

- **Order ID:** Representa o código da compra
- **Product:** Representa o produto comprado
- **Quantity Ordered:** Representa a quantidade encomendada
- **Price Each:** Representa o preço unitário dos produtos em dólar
- **Order Date:** Representa a data e o horário da compra
- **Purchase Address:** Representa o endereço onde a compra foi realizada

Tratamento dos dados

Nesta seção iremos conhecer um pouco mais sobre os dados com os quais estamos trabalhando e faremos os tratamentos necessários para que nossa análise não seja prejudicada ao longo deste estudo.

O primeiro tratamento será alterarmos a data de acordo com a localidade, como essa loja fictícia está localizada nos Estados Unidos, o formato da data não se enquadra no formato que usamos no Brasil. E ao fazermos isso, verificaremos que as linhas em branco retornarão valor "null".

Order Date
<div> <div>Válidos</div> <div>- %</div> </div> <div> <div>Erro</div> <div>< 1%</div> </div> <div> <div>Vazio</div> <div>- %</div> </div>
19/04/2019 08:46:00
null
07/04/2019 22:30:00
12/04/2019 14:38:00
12/04/2019 14:38:00
30/04/2019 09:27:00
29/04/2019 13:03:00

Com a data no formato adequado podemos inserir colunas que serão úteis como, dia, mês, ano, hora e nome do mês.

123 Dia	123 Mês	123 Ano	123 Hora	A8C Nome do Mês
<div> <div>Válidos</div> <div>- %</div> </div> <div> <div>Erro</div> <div>< 1%</div> </div> <div> <div>Vazio</div> <div>- %</div> </div>	<div> <div>Válidos</div> <div>- %</div> </div> <div> <div>Erro</div> <div>< 1%</div> </div> <div> <div>Vazio</div> <div>- %</div> </div>	<div> <div>Válidos</div> <div>- %</div> </div> <div> <div>Erro</div> <div>< 1%</div> </div> <div> <div>Vazio</div> <div>- %</div> </div>	<div> <div>Válidos</div> <div>- %</div> </div> <div> <div>Erro</div> <div>< 1%</div> </div> <div> <div>Vazio</div> <div>- %</div> </div>	<div> <div>Válidos</div> <div>- %</div> </div> <div> <div>Erro</div> <div>< 1%</div> </div> <div> <div>Vazio</div> <div>- %</div> </div>
19	4	2019	8	abril
null	null	null	null	null
7	4	2019	22	abril
12	4	2019	14	abril
12	4	2019	14	abril
30	4	2019	9	abril
29	4	2019	13	abril

O próximo passo é tratarmos o endereço, neste caso, queremos separar toda a cadeia de texto para obtermos a cidade juntamente com o Estado. E o motivo disso é que existem cidades com o mesmo nome em Estados diferentes, então se separarmos apenas a cidade, nossa análise de vendas pode ficar distorcida pois vendas de cidades com o mesmo nome seriam somadas.

Para isso, basta adicionarmos uma coluna de exemplo baseada na coluna de endereço retirando o nome de cada cidade e o Estado.

ABC Personalizar	
● Válidos	99%
● Erro	0%
● Vazio	< 1%
Dallas, TX	
Boston, MA	
Los Angeles, CA	
Los Angeles, CA	
Los Angeles, CA	
San Francisco, CA	

Removeremos as colunas que não nos interessam mais, as linhas em branco e os erros. Esses erros vieram da combinação dos arquivos. Ao serem combinados, os cabeçalhos de cada arquivo foram inseridos como dados, e somente os nomes da colunas do primeiro arquivo foram promovidos a cabeçalho automaticamente pelo Power BI. Isso pode ser visto por meio de um filtro.

ABC 123 Order ID	ABC 123 Product	ABC 123 Quantity Ordered	ABC 123 Price Each	Order Date
● Válidos 100%	● Válidos 100%	● Válidos 100%	● Válidos 100%	● Válidos - %
● Erro 0%	● Erro 0%	● Erro 0%	● Erro 0%	● Erro 100%
● Vazio 0%	● Vazio 0%	● Vazio 0%	● Vazio 0%	● Vazio - %
Order ID	Product	Quantity Ordered	Price Each	Error
Order ID	Product	Quantity Ordered	Price Each	Error
Order ID	Product	Quantity Ordered	Price Each	Error
Order ID	Product	Quantity Ordered	Price Each	Error
Order ID	Product	Quantity Ordered	Price Each	Error
Order ID	Product	Quantity Ordered	Price Each	Error
Order ID	Product	Quantity Ordered	Price Each	Error

Trataremos os preços que da mesma forma que as datas, estão em um formato não adequado para nós aqui do Brasil. O separador de decimal utilizado é o ponto e substituiremos pela vírgula.

Renomearemos as colunas traduzindo seus nomes e finalizaremos o tratamento escolhendo o tipo de dado que mais se adequa aos dados com que estamos trabalhando.

A ^B _C ID_Produto	A ^B _C Produto	t ² ₃ Quantidade	1.2 Preço	Data da Compra	t ² ₃ Dia
● Válidos 100%	● Válidos 100%	● Válidos 100%	● Válidos 100%	● Válidos 100%	● Válidos 100%
● Erro 0%	● Erro 0%	● Erro 0%	● Erro 0%	● Erro 0%	● Erro 0%
● Vazio 0%	● Vazio 0%	● Vazio 0%	● Vazio 0%	● Vazio 0%	● Vazio 0%
176558	USB-C Charging Cable	2	11,95	19/04/2019 08:46:00	19
176559	Bose SoundSport Headphones	1	99,99	07/04/2019 22:30:00	7
176560	Google Phone	1	600	12/04/2019 14:38:00	12
176560	Wired Headphones	1	11,99	12/04/2019 14:38:00	12
176561	Wired Headphones	1	11,99	30/04/2019 09:27:00	30
176562	USB-C Charging Cable	1	11,95	29/04/2019 13:03:00	29
176563	Bose SoundSport Headphones	1	99,99	02/04/2019 07:46:00	2
176564	USB-C Charging Cable	1	11,95	12/04/2019 10:58:00	12
176565	Macbook Pro Laptop	1	1700	24/04/2019 10:38:00	24

A partir daqui poderemos fechar e aplicar, e iniciarmos a construção do dashboard.

Construção do Dashboard

Antes de começarmos a construir os visuais, vamos nos atentar a quais perguntas queremos responder. Então para esse estudo iremos buscar responder as seguintes perguntas:

- Qual o mês de maior faturamento ?;
- Quais são os 10 produtos mais vendidos ?;
- Qual a cidade com o maior faturamento ?;
- Como as vendas se comportam em relação ao horário ?.

Medidas e Tabela Calendário

Para responder as perguntas acima iremos criar algumas medidas como o total de faturamento, o percentual de vendas, a quantidade total vendida e um ranking para os produtos. E para auxiliar o trabalho com datas nos visuais, criaremos uma tabela de calendário para criar uma relação com a coluna de data da nossa base de dados.

A tabela calendário será criada por meio da função DAX CALENDARAUTO(),

```

1 Tabela Calendário =
2 ADDCOLUMNS(
3     CALENDARAUTO(),
4     "Dia num",DAY([Date]),
5     "Dia nome",FORMAT([Date], "dddd"),
6     "Dia Semana", WEEKDAY([Date]),
7     "Semana Num", WEEKNUM([Date]),
8     "Mês Num",MONTH([Date]),
9     "Mês Nome",FORMAT([Date], "mmm"),
10    "Trimestre",QUARTER([Date]),
11    "Ano",YEAR([Date])
12 )

```

O faturamento total será a soma de cada linha da multiplicação da quantidade pelo preço, utilizando a função DAX SUMX().

```

1 Faturamento Total = sumx(Sales_Data, Sales_Data[Quantidade] * Sales_Data[Preço])

```

O percentual de vendas será a divisão entre o faturamento, seja ele filtrado por mês, ano, cidade, pelo faturamento total retirando a aplicação dos filtros. E aqui utilizaremos a função DAX DIVIDE() para evitar problemas com divisões por zero, e no denominador usaremos a função DAX CALCULATE() para realizar a somatória, filtrando pela seleção completa de nossos dados por meio da função DAX ALL().

```

1 Percentual Vendas =
2 DIVIDE(
3     sumx(Sales_Data, [Quantidade] * [Preço]),
4     CALCULATE(sumx(Sales_Data, [Quantidade] * [Preço]), ALL(Sales_Data))
5 )

```

A Quantidade total será calculada de forma bem simples apenas pela soma da coluna com as quantidades.

```

1 Quantidade Total = sum(Sales_Data[Quantidade])

```

Para o ranking dos produtos usaremos a função DAX RANKX(), selecionando todos os produtos da nossa tabela com a função DAX ALLSELECTED() para ignorar os filtros, e usando como expressão a medida já criada acima com a quantidade total.

```
1 Rank Produtos =  
2 RANKX(  
3 | ALLSELECTED(Sales_Data[Produto]),  
4 | [Quantidade Total]  
5 )
```

Selecionando Visuais

Este é o passo onde testaremos e selecionaremos os visuais que melhor representam nossos dados e respondem nossas perguntas.

Para valores como o faturamento total, o percentual de vendas e a quantidade total vendida, o visual de cartão nos serve muito bem pois todos são valores únicos.

No caso da distribuição do faturamento mês a mês, um gráfico de linhas é uma boa opção por se tratar de uma série ao longo do tempo.

Para o faturamento por cidade, um gráfico de colunas pode ser utilizado mostrando uma coluna para cada cidade.

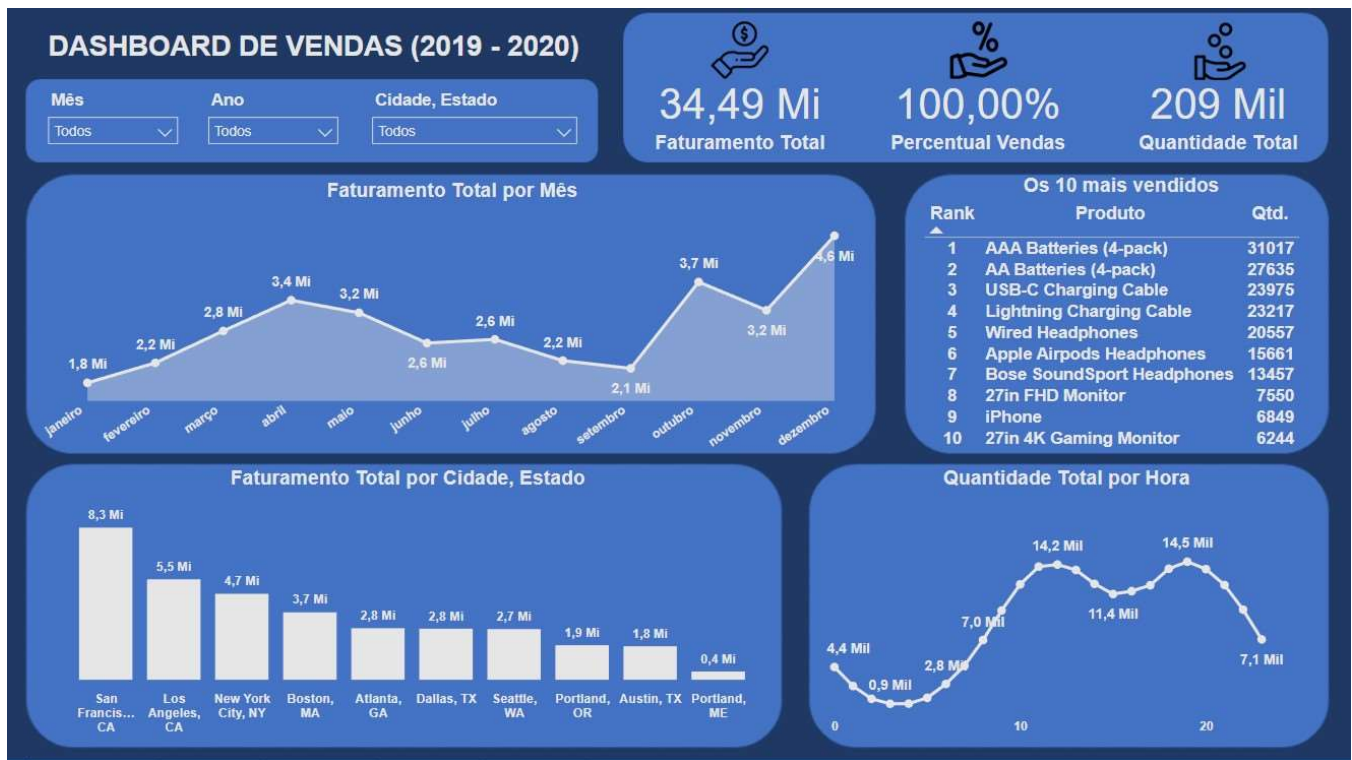
Em uma tabela podemos inserir o ranking dos 10 produtos mais vendidos, o que permite fácil visualização de maneira mais enxuta.

Considerando a quantidade vendida por hora como uma série temporal, podemos também utilizar o gráfico de linhas para representar este visual.

As segmentações dos dados poderão ser realizadas por mês, ano e pela localização (cidade, Estado).

Dashboard

Após termos nossos visuais prontos, podemos editar nosso dashboard da maneira que quisermos e ao final, teremos algo parecido com a figura a seguir.



Agora podemos responder as perguntas em nosso relatório.

Insights

Para facilitar a construção do relatório usaremos indicadores criados para melhorar a apresentação.



1. Qual o mês de maior faturamento ?



Ao observarmos o gráfico, inferimos que o mês de dezembro foi o mês com o maior valor de faturamento, representando um total de \$ 4.613.443,34. E uma possível explicação para isso ter ocorrido é porque em dezembro acontece o Natal, portanto nessa data a procura por presentes aumenta consideravelmente.

2. Quais são os 10 produtos mais vendidos ?



A tabela mostra que produtos como **AAA batteries (4-pack)** e **AA batteries (4pack)** são os mais vendidos no geral, e uma simples resposta para isso é porque são mais baratos do que produtos como **Macbook Pro Laptop** por exemplo.

Podemos dentro dessa pergunta, criar um gráfico inserindo em colunas as quantidades vendidas de cada produto, e em uma linha, a média de preço dos mesmos.



Dessa forma confirmamos a hipótese acima de que produtos mais baratos são comprados em maiores quantidades.

Porém existem casos onde, por exemplo, o **Macbook Pro Laptop** possui um preço bem superior do que a **LG Washing Machine** e mesmo assim foi mais vendido. Aqui é importante entender o padrão de consumo, onde provavelmente existe uma demanda maior de um laptop do que uma máquina de lavar.

Esta sobreposição pode ainda não evidenciar tudo, mas é um bom indicativo da relação da vendas e dos preços de cada produto.

3. Qual a cidade com o maior faturamento ?

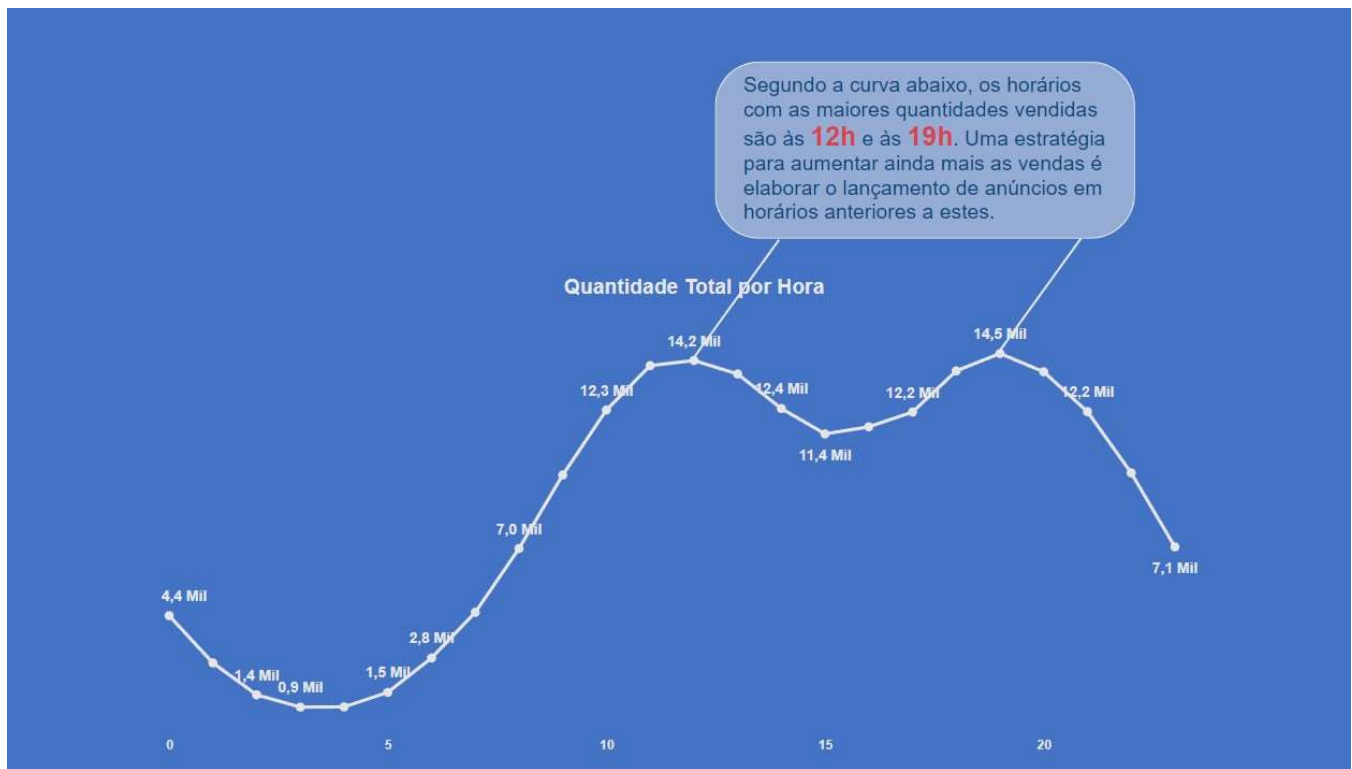
San Francisco é a cidade com o maior faturamento, representando 23,95% do total de vendas. Localizada perto da região do **Vale do Silício**, isso justifica a demanda elevada de produtos eletrônicos.

Faturamento Total por Cidade, Estado



Observando o gráfico vemos que a cidade de San Francisco no Estado da Califórnia é a cidade que mais consumiu produtos eletrônicos. Dentre os motivos que podem justificar esse grande consumo, é o fato do Silicon Valley (Vale do Silício) precisar mais de produtos eletrônicos, ou as pessoas em San Francisco terem mais acesso ou condições financeiras de adquirir produtos eletrônicos, ou até mesmo o engajamento em anúncios das empresas Tech seja melhor em San Francisco.

4. Como as vendas se comportam em relação ao horário ?



Como podemos analisar no gráfico, o primeiro pico aconteceu às 12h, e o segundo pico às 19h. Então possivelmente os melhores horários para exibir mais anúncios seriam provavelmente às 11h, e às 18h.

Sugestões

Em estudos futuros é interessante aumentar a base de dados, refinar melhor a colheita desses dados, treinar algoritmos para identificar mais padrões de consumo que visem aumentar a receita da companhia e procurar estratégias de mercado para potencializar as vendas.

Conclusão

Por meio deste estudo foi possível saber em qual mês há maior demanda, ou seja, para qual mês os estoques deverão estar preparados. Também descobrimos a cidade que consome mais produtos, o que implica onde está a maior receita da loja,

então é necessário uma atenção especial em estratégias de marketing que envolvem esta cidade.

Além disso, observamos quais os horários mais quentes de compras, os quais poderão ser utilizados para publicação de anúncios e promoções.

E assim finalizamos nosso dashboard e nosso relatório

Referências

GALLI, Keith. Solving real world data science tasks with Python Pandas!. Youtube, 13 jan. 2020. Disponível em: <https://www.youtube.com/watch?v=eMOA1pPVUc4&t=3083s>. Acesso em: 30 out. 2021.