

# Relatório da análise de vendas

## Teste prático Kardbank

Isabela Maria Evangelista

Isabelaevangelista17@gmail.com / (11) 987879636 /  
<https://www.linkedin.com/in/isabela-evangelista-27461829a/>

### Sumário

A base de dados .....	2
Ferramentas .....	3
Análise exploratória .....	4
Média, desvio padrão e quartis .....	4
Histogramas .....	4
Análises gerais .....	7
Quantidade de vendas .....	7
Total de vendas .....	9
Análises específicas .....	11
Vendas por região .....	11
Produtos mais vendidos .....	11
Análise de preços e suas correlações .....	12
Variação das vendas ao longo do tempo .....	13
Modelo de regressão linear .....	15
Conclusões .....	16

## A base de dados

O conjunto de dados utilizado nesse teste e análise foi originado do arquivo 'sales.csv' e conta com 1000 linhas e 6 colunas: Date, Region, Product, Price, Quantity Sold e Sales. Foi mantido o idioma original das informações nas análises e visualizações, porém nesse documento serão tratadas em português.

## Ferramentas

A principal ferramenta utilizada para análise foi um Jupyter Notebook, anexo no repositório do GitHub, via Google Colab. A linguagem de programação Python foi a escolhida, contando com as bibliotecas Pandas, Matplotlib e Seaborn por exemplo. Outras ferramentas utilizadas e consultadas para a execução desse projeto foram Microsoft Excel e Power BI, com dashboard a ser anexado no repositório do GitHub também.

## Análise exploratória

### Média, desvio padrão e quartis

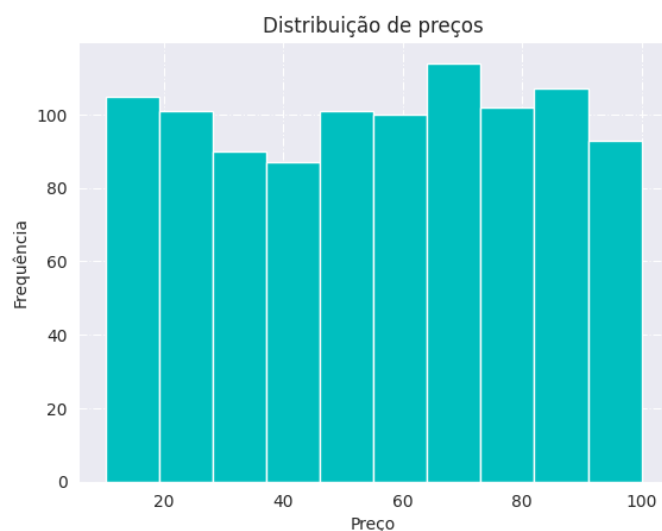
Por meio do comando `'dados.describe()'` foi possível acessar as informações de média, desvio padrão e quartis de forma rápida e fácil, como nota-se na tabela abaixo.

	Date	Price	Quantity Sold	Sales
count	1000	1000.000000	1000.000000	1000.000000
mean	2024-05-14 12:00:00	55.417260	10.103000	562.475330
min	2023-01-01 00:00:00	10.290000	1.000000	11.730000
25%	2023-09-07 18:00:00	32.347500	5.000000	208.627500
50%	2024-05-14 12:00:00	56.520000	10.000000	448.470000
75%	2025-01-19 06:00:00	77.345000	15.000000	829.900000
max	2025-09-26 00:00:00	99.950000	19.000000	1891.260000
std	NaN	25.902891	5.547357	428.443008

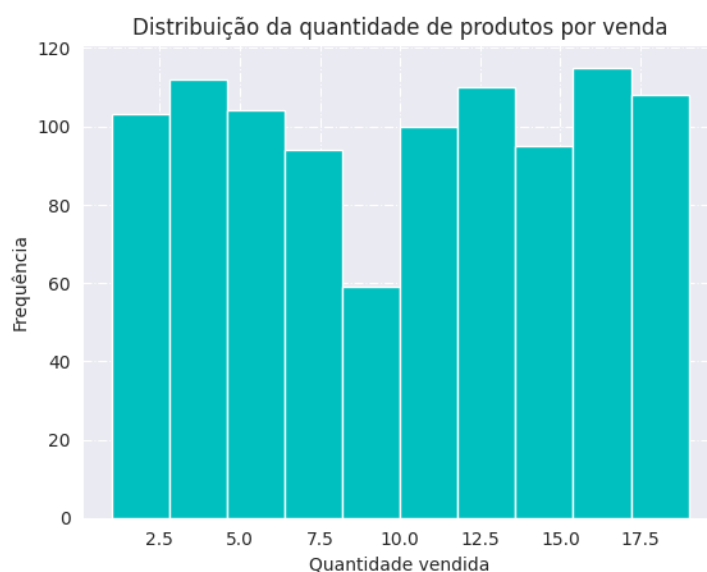
As informações dos quartis (min, 25%, 50%, 75% e max) mostram como cada variável está distribuída no todo, e a distribuição desses campos é bem estável no geral. A localização do primeiro e terceiro quartis em Price, Quantity Sold e Sales é equilibrada em relação à mediana, mínimo e máximo, o que indica que nenhum desses tende para nenhum dos extremos.

### Histogramas

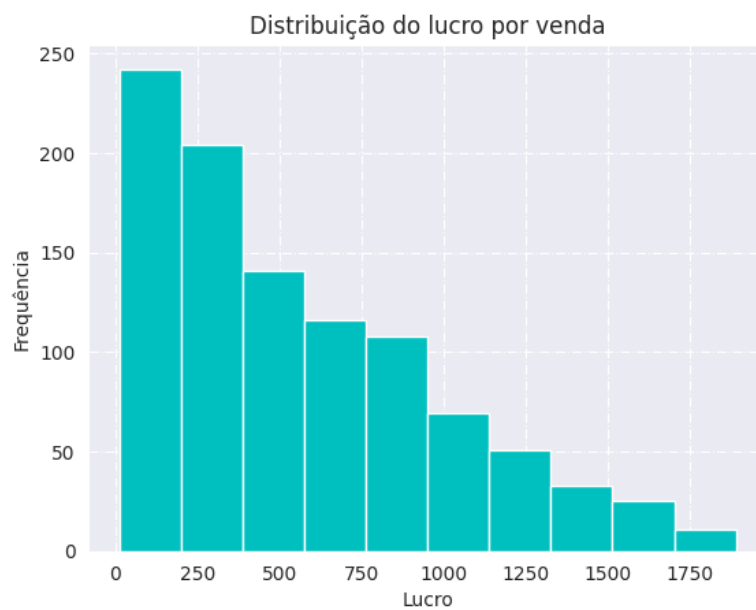
Em relação a distribuição dos preços dos produtos, pelo histograma abaixo é possível notar que estes preços variaram bastante dentro do conjunto de dados, já que, com preços de \$0 a \$100, todos tiveram pelo menos 80 ocorrências e menos de 120. Isso indica que há grande variedade de preços nos produtos da empresa, o que pode ser benéfico, já que é um fator favorável a um alcance maior de público.



Já a quantidade de vendas teve uma variação um pouco maior, mas ainda assim não tão grande. Segundo gráfico abaixo, as vendas do período foram feitas, em sua maioria, em grande ou pequena quantidade, sendo que as vendas de 10 unidades, por exemplo, foram minoria. Essa estatística não necessariamente indica alguma tendência específica, mas reforça a teoria de que a empresa atende vários tipos de público.



Em contrapartida, a distribuição do lucro por venda é mais heterogênea e marca claramente como as vendas de grande valor são minoria, enquanto que as vendas de baixo lucro aparecem 5 vezes a mais que as primeiras no conjunto de dados.

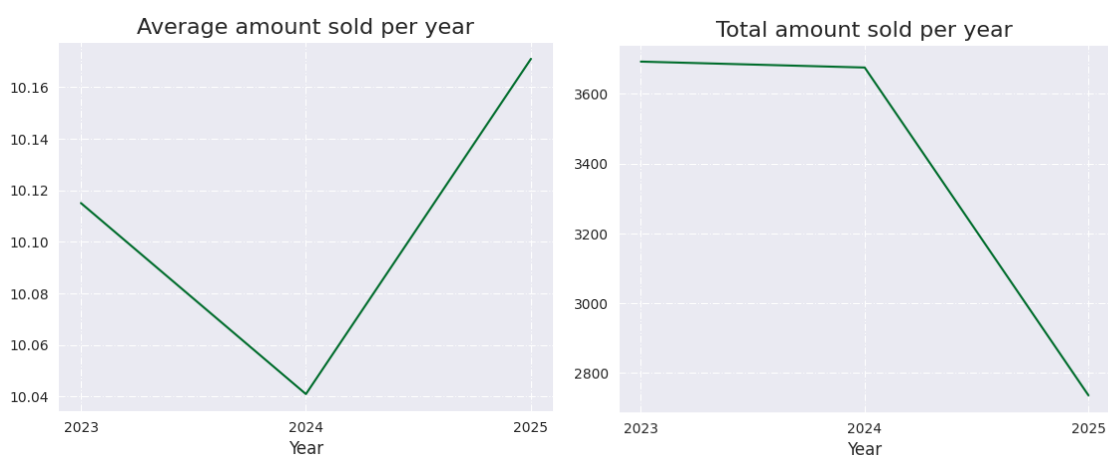


Isso pode indicar que os produtos mais caros não têm valor agregado suficiente para que os clientes invistam neles, ou uma preferência geral pelos produtos mais baratos, ou ainda, somente uma preferência dos clientes em fazer compras menores.

## Análises gerais

### Quantidade de vendas

Entre o total e a média de vendas anuais é clara a discrepância de valores. No ano de 2024 a média foi a menor, porém o total não apresentou queda em relação ao ano interior, o que indica mais vendas mais caras ou de produtos mais caros. Já em 2025 foi o contrário, a média aumentou e o total caiu. Essa movimentação é igualmente benéfica e maléfica. Isso porque aumentar a média de vendas anuais é algo que traz estabilidade para os negócios, porém deve ser feita uma análise mais aprofundada do porque o total diminuiu: se foram os produtos mais baratos, se foram quantidades menores, se foram mais vendas porém com valor menor, entre outros.



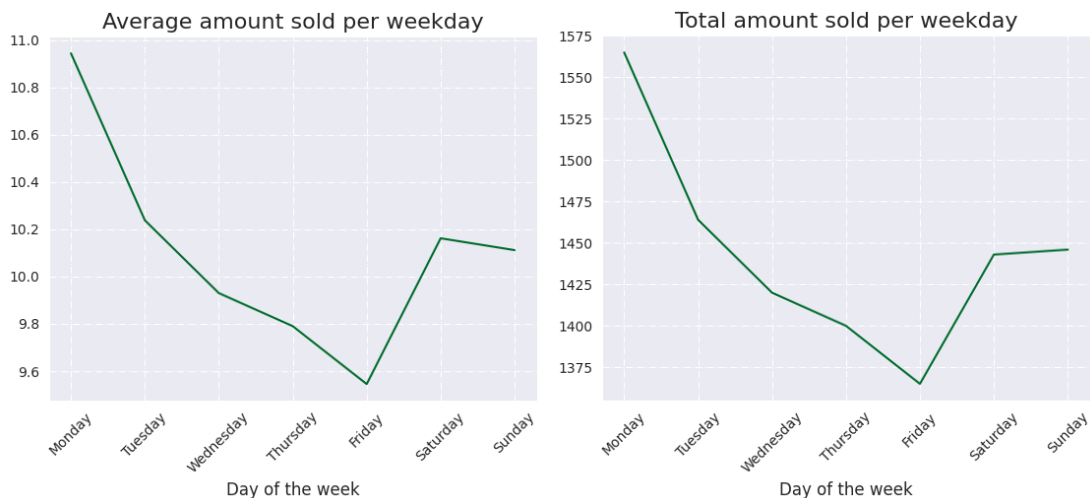
A variação mensal mostra os pontos altos e baixos do ano, sendo que a diferença entre total e média de vendas também existe, apesar de mais discreta. Nos meses finais do ano, as vendas despencam, possivelmente pela temporada de feriados em que os clientes tem outras prioridades de gastos. Isso logo em seguida de um período relativamente logo de alta. Dependendo do produto, isso pode indicar uma tendência de alto consumo preventivo dos clientes, que decidem comprar o que precisam antes da temporada de feriados, por motivos de organização de gastos ou viagens, que impossibilitam a manutenção das compras dos produtos dessa empresa. Esse é um ponto que desperta uma outra possível análise interior da organização da empresa, nesse caso para entender se esse movimento se deve aos feriados e comemorações, em que as pessoas estão ocupadas demais para pensar na compra desse tipo de produto, ou se a compra destes é algo dificultoso com mais fatores da época envolvidos, como a distância, menor disponibilidade financeira ou até entrega dos produtos.

Essa tendência provavelmente é o que impacta a queda nas vendas no mês 4, abril, porque acaba desregulando e tirando o ritmo de compras que deveria ser reestabelecido. O aumento de vendas do mês 6, junho, até o mês 9, setembro, é o que permite a diminuição de vendas dos meses 10 a 12, outubro a dezembro, posteriormente gerando a necessidade de compras no começo do ano e logo em seguida caindo novamente pelo alto volume das últimas compras. Essa situação acaba se tornando cíclica e, apesar de acrescentar no fator

previsibilidade, prejudica as finanças da empresa ao longo prazo, já que variâncias tão grandes assim podem gerar inseguranças financeiras.



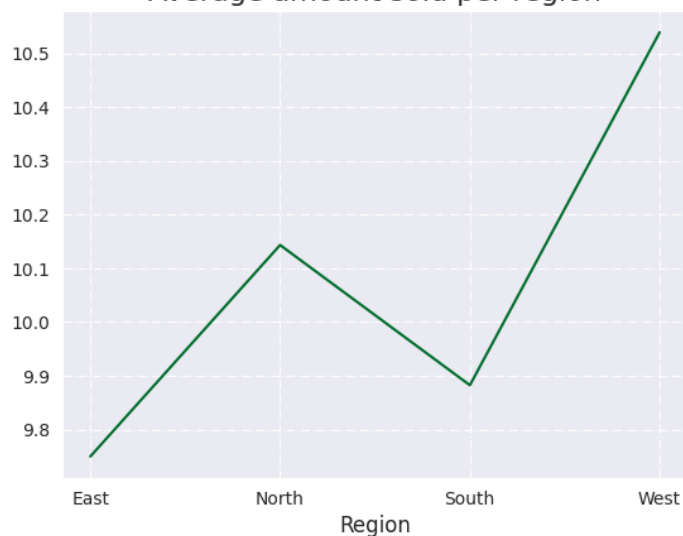
Já a variância semanal é mais comum e normalmente não indica grandes problemas. A empresa apresenta melhor desempenho na segunda-feira e pior desempenho da sexta-feira, como mostram os gráficos abaixo. Novamente, nesse caso vale a análise da organização empresarial para descobrir se essa queda no último dia útil da semana se deve a preferências pessoais ou algum procedimento das vendas que é mais complicado no final da semana.



O gráfico abaixo foi produzido pensando em visualizar a distribuição de vendas por região, e a região oeste é a que de longe tem melhor desempenho. Isso porque, além do valor da média de vendas ser o mais alto dentre as 4 regiões da cidade, uma média maior indica mais estabilidade de vendas, o que aponta uma melhor situação a longo prazo.



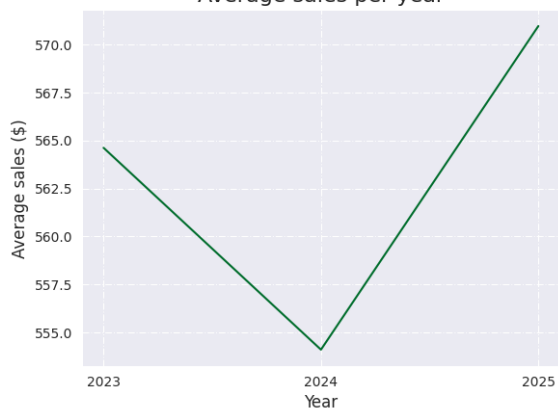
Average amount sold per region



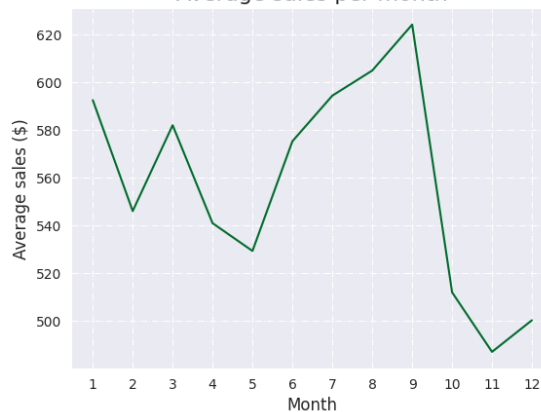
## Total de vendas

Os indicadores do total de vendas, os mesmos da quantidade de vendas, são todos muito parecidos com os primeiros. Com os gráficos a seguir, isso tudo só reforça as hipóteses em relação aos ciclos de compras e ao desempenho da região oeste.

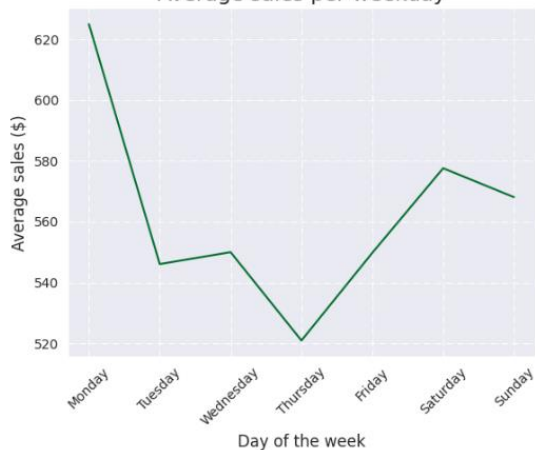
Average sales per year



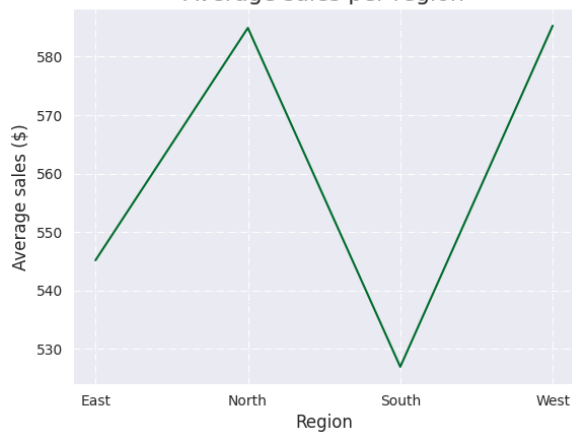
Average sales per month



Average sales per weekday



Average sales per region

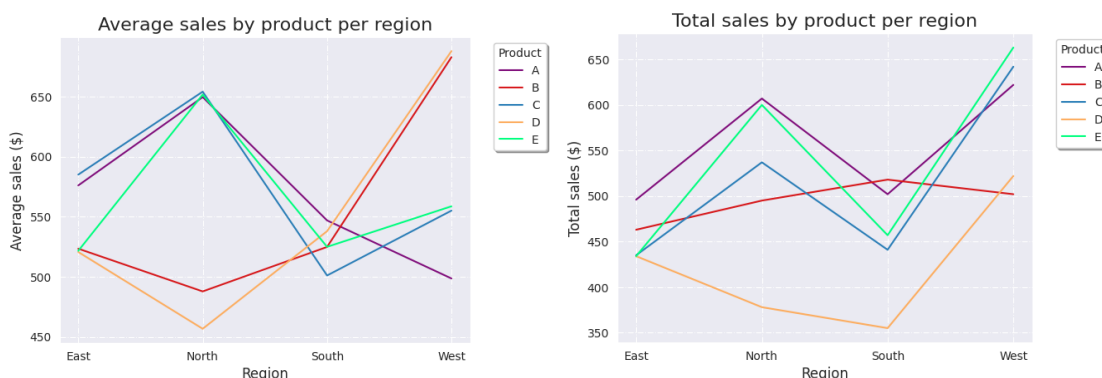


Perante a análise geral da quantidade e do valor total de vendas, se faz necessário investir na análise organizacional para entender e melhorar as questões cíclicas da distribuição de vendas dos produtos dessa empresa.

## Análises específicas

### Vendas por região

Foram plotados os seguintes gráficos para analisar as questões referentes às vendas em relação as regiões da cidade:



Como mostrado anteriormente, a região oeste é a mais lucrativa para a empresa. Os produtos, em média, mais vendidos são B e D, porém os produtos A, C e E disparam em quantidade nessa região.

A variância entre a média e o lucro total nas vendas de cada produto aponta que os produtos B e D provavelmente são os que tem preço mais baixo, porque são os mais vendidos e também os que dão menos lucro. A grande quantidade de lucro que os produtos A, C e E geram na região oeste em contrapartida com as suas médias de venda indicam que seus preços são mais altos e, portanto, carregam a maior parte da responsabilidade de deixar a região oeste no topo de vendas.

Essa diferença entre esses dois grupos se apresenta muito fortemente na região norte também, que é a segunda mais lucrativa e tem média de vendas altíssimas dos produtos A, C e E.

Sendo assim, é possível concluir que, pelo menos por regiões, os produtos mais lucrativos são os A, C e E, mas seus preços altos podem ser impeditivos para as compras em grandes quantidades, já que a região norte apresenta as médias de vendas desses produtos mais altas, mas continua não sendo a região mais lucrativa.

### Produtos mais vendidos

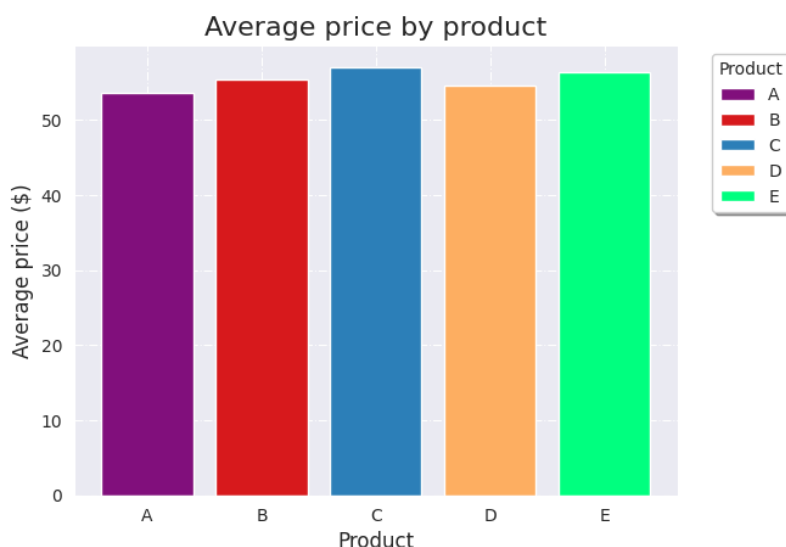
De acordo com soma simples realizada no próprio Jupyter Notebook, os produtos mais lucrativos e mais vendidos são A, C e E respectivamente.

Em quantidade de vendas, esses produtos atingiram a máxima de 2227, 2154 e 2055 unidades vendidas, nessa ordem.

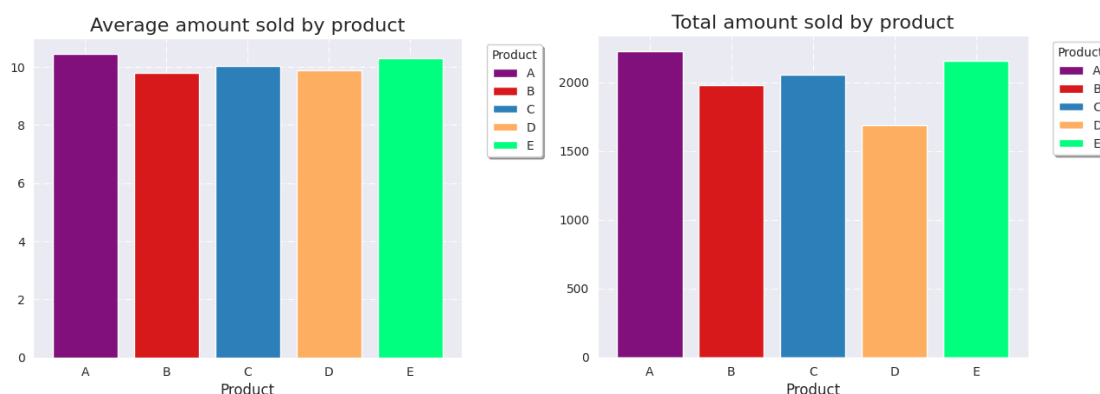
Já em total de vendas, o produto A vendeu um total de \$120588,59, o produto C um total de \$118729,01, e o produto E, \$117639,92.

## Análise de preços e suas correlações

A média dos preços dos produtos é observada no seguinte gráfico:



Os preços se mantiveram na média dos \$50 durante os 3 anos de vendas da empresa. O produto C é o que tem o maior preço médio, e o produto A tem o menor. Sendo o produto A um dos três mais lucrativos e tendo menor preço médio, isso indica que o que lhe faz tão lucrativo pode ser o preço baixo, o que gera mais vendas em valores totais mais altos, como o gráfico abaixo da quantidade média vendida confirma.

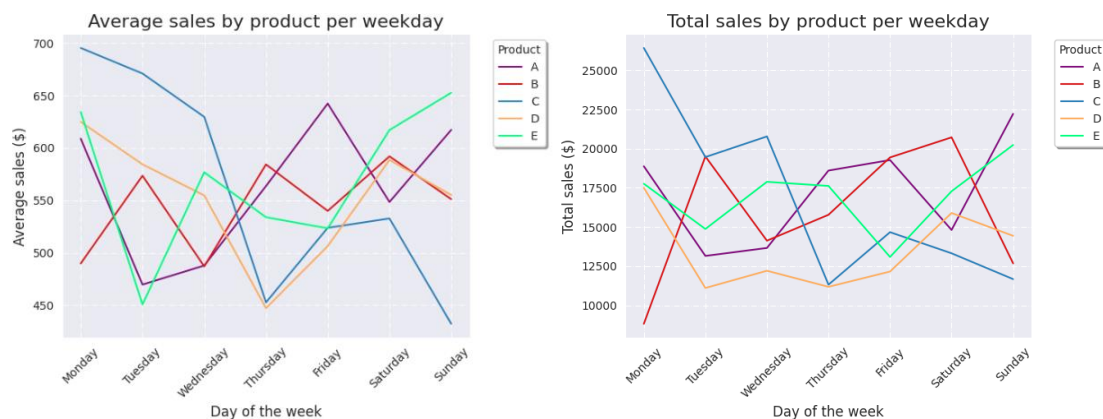


O produto D pode ser confirmado como o produto menos lucrativo da empresa, já que tem preço baixo, quantidade média de vendas baixa e o menor lucro total de vendas. Esse produto é um possível candidato para ser deixado de lado, caso um dia a empresa decida focar somente nos produtos mais lucrativos. O produto B é o que tem menor média de quantidade vendida, e também se apresentou um dos produtos menos lucrativos, mas ainda assim gera mais lucro que o produto D.

Esses gráficos provam que a correlação entre o preço e a quantidade vendida de cada produto é pequena, já que o produto A tem maior quantidade vendida e lucro enquanto tem o menor preço, e o produto D tem um dos menores preços e é o menos vendido.

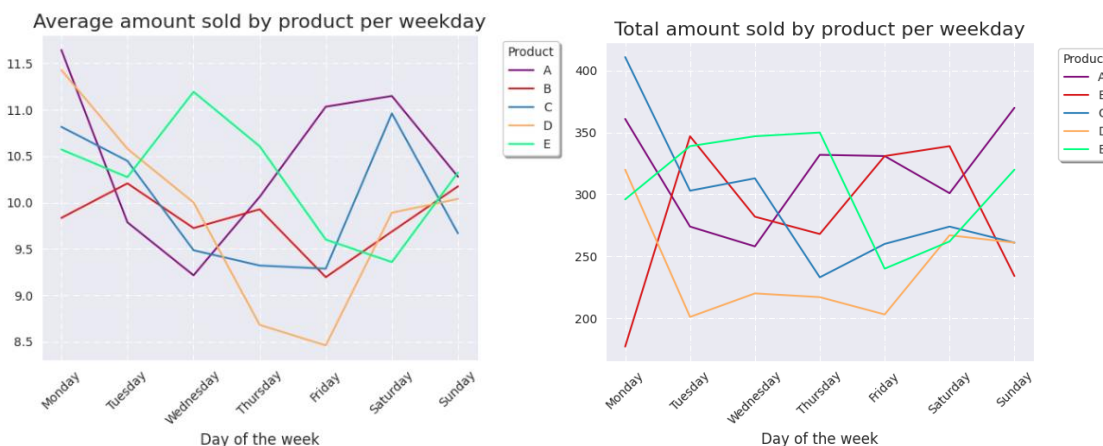
## Variação das vendas ao longo do tempo

Como a análise geral anterior já havia mostrado, as vendas têm variação semanal significativa, sendo segunda-feira o melhor dia de vendas e sexta-feira o pior.



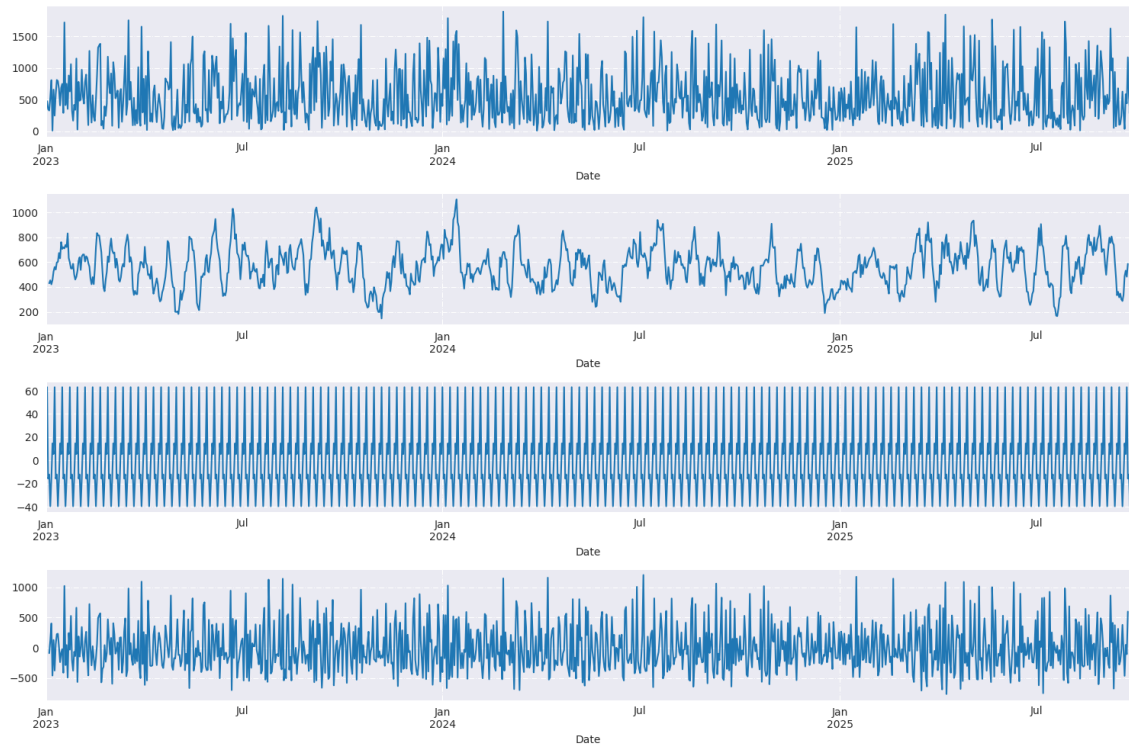
Pelos gráficos acima, nota-se que a hipótese dada na análise geral foi parcialmente confirmada, já que segunda-feira continua sendo o melhor dia de vendas no geral, mas a sexta-feira não é o pior dia para todos os produtos, inclusive é o dia que tem melhor média de vendas do produto A.

Tratando de médias, o produto A tem desempenho igualmente bom tanto na segunda quanto na sexta-feira, com baixa significativa na terça e quarta-feira, e seu lucro acompanha. Já o produto C, que é o que gera maior lucro, é mais bem desenvolvido no começo da semana. Os produtos C e D apresentam grande queda na média e no total na quinta-feira, e os outros produtos têm desempenho mediano nesse dia, o que indica que quinta-feira seria o pior dia de vendas da empresa, e não sexta-feira, como apontado anteriormente, mesmo que o produto A, o mais lucrativo, apresente aumento nos lucros nesse dia. Segunda-feira continua sendo o melhor dia da semana pois mais produtos apresentam lucros e médias altas, inclusive o produto A.



Em relação à quantidade vendida, a tendência é a mesma, com segunda-feira como o dia de maiores quantidades vendidas e o fim da semana com as menores.

Pensando em análise sazonal do total de vendas foram gerados os seguintes gráficos, que contam com, respectivamente, evolução observada, tendência, sazonalidade e resíduos.



No caso desse conjunto de dados, não é possível observar nenhuma tendência na evolução do total de vendas, o que se deve às variações anuais e mensais apresentadas anteriormente. Caso as vendas estejam realmente se estabilizando, como a análise das médias anuais indicou, futuramente será possível gerar uma análise de tendência nas vendas.

A sazonalidade também não traz grandes conclusões, porque não há um padrão específico de variação de compras, já que há muitas diferenças de um mês para o outro, por exemplo, e isso não gera sazonalidade o suficiente para análise.

## Modelo de regressão linear

Foi criado um modelo de regressão linear com base nos preços e na região para prever o total de vendas geral e avaliada sua performance usando MAE (Mean Absolute Error, ou média absoluta dos erros), MSE (Mean Square Error, ou média quadrática dos erros) e o Coeficiente  $R^2$ .

No primeiro modelo treinado, as métricas apontaram respectivamente, 4,622, 29,002 e -0,001. Esses valores não são bons e mostram que o modelo não se desenvolveu bem e não pode ser usado para prever as vendas com precisão.

Em seguida foi treinado um segundo modelo de regressão linear, dessa vez usando o método `StandardScaler()` para ajudar na normalização dos dados e tentar um desempenho melhor. Infelizmente, as métricas foram ruins novamente: MAE de 4,622 e  $R^2$  de -0,211.

Analisando essas execuções com essas métricas de resultado e juntando com as análises anteriores, é possível concluir que os modelos de previsão não serão muito eficazes nesse conjunto de dados que não possui tendência ou sazonalidade nos seus valores no momento. Em uma série temporal, como é o caso, esses fatores são métricas para que o analista consiga perceber que o erro pode não ser o modelo de previsão, e sim o conjunto de dados. Futuramente, caso a média de vendas estabilize, pode ser gerada certa sazonalidade, e assim, possibilitar o uso de modelos de previsão, talvez o ARIMA, um modelo mais sofisticado e mais adequado para séries temporais.

## Conclusões

Em suma, após todas as análises citadas acima, é possível concluir que a empresa tem um bom desempenho geral.

Em relação a produtos, os produtos A e C devem se tornar prioridades no ciclo evolutivo, e deve ser considerada a possibilidade de parar as vendas do produto B, que não traz lucro o suficiente para a empresa.

Em relação a regiões, a região oeste deve receber investimentos extras para mais lucro a curto prazo, e as regiões leste e sul devem receber investimentos iguais para que possam evoluir e igualar suas vendas com as vendas da região oeste.

E, por fim, como sugestão, a análise organizacional deve ser feita para identificar a causa raiz das diferenças nas vendas em diferentes meses do ano, porque isso pode se tornar um problema no futuro.