

Relatório A1

Séries Temporais

Gerardo Mikael Do Carmo Pereira

Professor: Thiago Guerrera Martins

RIO DE JANEIRO
2025

Conteúdo

1	Introdução	3
2	Desenvolvimento	3

1 Introdução

Este trabalho tem como objetivo analisar o comportamento de uma série temporal referente ao volume observado ao longo de várias semanas. A partir da visualização dos dados e do estudo das funções de autocorrelação e autocorrelação parcial (ACF e PACF), buscamos identificar padrões de tendência, estacionariedade e possíveis modelos adequados para previsão.

2 Desenvolvimento

Análise inicial

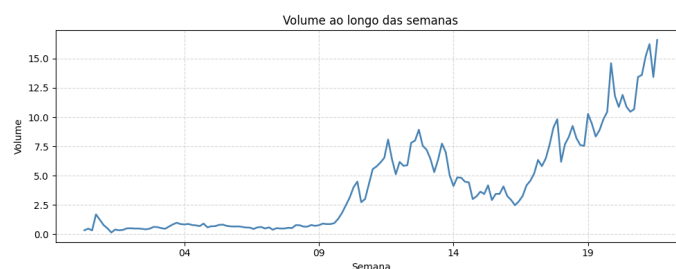


Figura 1

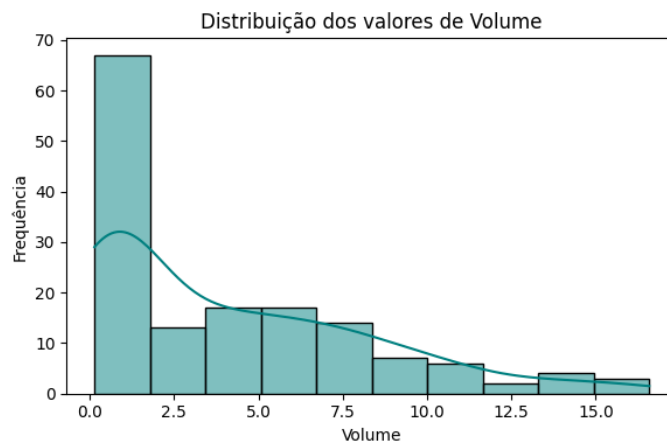


Figura 2

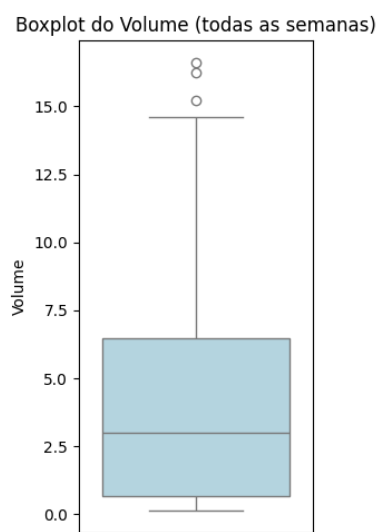


Figura 3

O histograma e o boxplot indicam que a distribuição da variável volume é marcadamente assimétrica positiva. A maior parte das observações concentra-se em valores baixos, como evidenciado pela alta frequência de volumes entre 0 e 2.5 no histograma. O boxplot corrobora essa observação, mostrando que a mediana está mais próxima do primeiro quartil, e a presença de uma longa cauda superior com múltiplos outliers. Essa forte assimetria sugere que modelos lineares que pressupõem normalidade dos resíduos podem não performar bem com a variável na escala original, apontando para a necessidade de aplicar uma transformação para estabilizar a variância e aproximar a distribuição de uma simetria.

A análise da série temporal, que plota o volume em função da semana, revela que os dados não são estacionários. Há um comportamento distinto em dois regimes: nas primeiras 9 semanas, o volume se mantém em um patamar baixo e com pouca variabilidade. A partir da semana 9, inicia-se uma clara tendência de crescimento, acompanhada por um aumento na volatilidade. A existência dessa tendência é o fator mais crítico para a modelagem violando a independência das observações. Isso justifica a necessidade de decompor a série ou de criar co-variáveis baseadas no tempo para capturar esse efeito e modelá-lo adequadamente.

Análise de autocorrelação

Foram gerados os gráficos da Função de Autocorrelação (ACF) e da Função de Autocorrelação Parcial (PACF). Ferramentas usadas para identificar a presença de tendência, sazonalidade e a natureza dos processos estocásticos subjacentes.

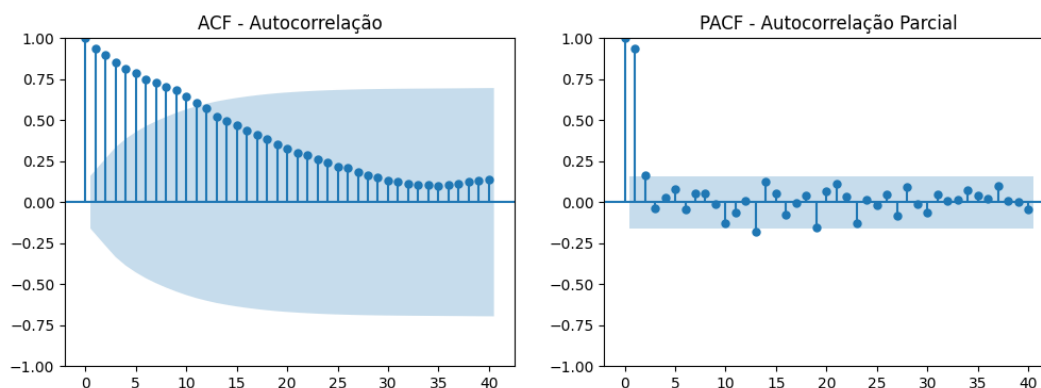


Figura 4

Gráfico ACF (Autocorrelação): O gráfico ACF à esquerda exibe um decaimento lento e quase linear. As autocorrelações são muito altas e positivas para um grande número de lags, permanecendo estatisticamente significativas (fora do intervalo de confiança azul) para todos os 40 lags exibidos. Este padrão é a assinatura clássica de uma série temporal não-estacionária, confirmando a forte tendência de crescimento observada no gráfico de linha. A intuição é que, devido à tendência, uma observação em um ponto t será muito similar à observação em $t-1$, $t-2$, etc., simplesmente porque ambas estão na mesma trajetória ascendente, gerando uma alta correlação que só diminui muito lentamente com o tempo.

Análise do Gráfico PACF (Autocorrelação Parcial): O gráfico PACF, por outro lado, nos mostra a correlação entre a série e um de seus lags após remover o efeito dos lags intermediários. O resultado é diferente: observa-se um pico muito significativo e positivo no lag 1, seguido por um corte abrupto. Após o primeiro lag, os valores da PACF caem imediatamente para dentro do intervalo de confiança, tornando-se não-significativos. Este comportamento sugere que, uma vez que a influência do valor imediatamente anterior $t-1$ é considerada, os valores mais antigos não adicionam uma informação nova e significativa para prever o valor atual.

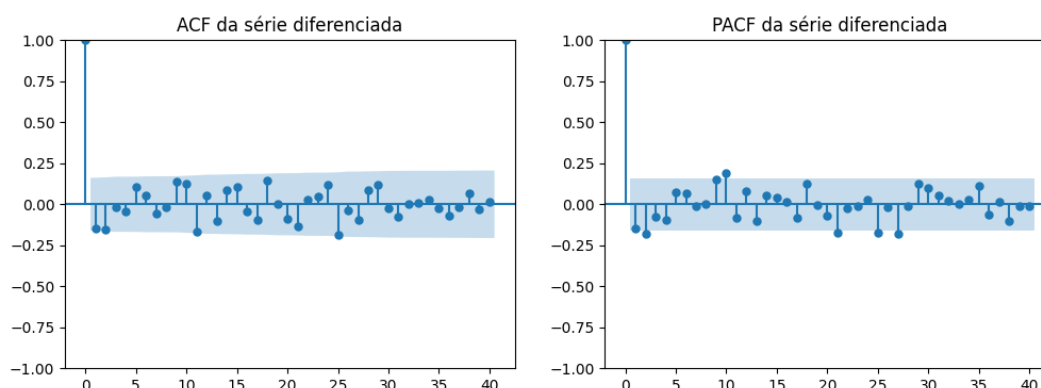


Figura 5

Após aplicar uma primeira diferenciação para remover a tendência, os novos gráficos ACF e PACF tiveram uma transformação bem-sucedida, pois ambos os gráficos

agora mostram que todas as autocorrelações estão dentro do intervalo de confiança (ou muito perto disso), indicando que a série se tornou estacionária.

O resultado se assemelha a um white noise, o que é uma forte evidência de que a série de volume original segue um processo de random walk.

Isso estabelece o modelo ARIMA(0, 1, 0) como um excelente e robusto modelo baseline para o projeto. A partir dele, qualquer modelo mais complexo deve provar seu valor superando a simples previsão de que o próximo valor será igual ao último observado.

ARIMA(0,1,0)

SARIMAX Results						
=====						
Dep. Variable:	volume	No. Observations:	150			
Model:	ARIMA(0, 1, 0)	Log Likelihood	-205.792			
Date:	Mon, 06 Oct 2025	AIC	413.584			
Time:	01:14:02	BIC	416.588			
Sample:	0	HQIC	414.804			
	- 150					
Covariance Type:	opg					
=====						
	coef	std err	z	P> z	[0.025	0.975]

sigma2	0.9272	0.061	15.217	0.000	0.808	1.047
=====						
Ljung-Box (L1) (Q):		3.33	Jarque-Bera (JB):	113.56		
Prob(Q):		0.07	Prob(JB):	0.00		
Heteroskedasticity (H):		32.52	Skew:	0.13		
Prob(H) (two-sided):		0.00	Kurtosis:	7.27		
=====						
--- Previsão para os próximos 5 períodos ---						
Último valor observado: 16.59						
Valores Previstos:						
150	16.59					
151	16.59					
152	16.59					
153	16.59					
154	16.59					
Name: predicted_mean, dtype: float64						
Intervalos de Confiança (95%):						
	lower volume	upper volume				
150	14.702720	18.477280				
151	13.920983	19.259017				
152	13.321136	19.858864				
153	12.815441	20.364559				
154	12.369914	20.810086				

Figura 6

A análise do modelo baseline ARIMA(0, 1, 0) revela um resultado misto. Por um lado, o teste de Ljung-Box (Prob(Q) = 0.07) indica que o modelo foi bem-sucedido em remover a autocorrelação dos resíduos, significando que a diferenciação capturou a dependência temporal dos dados. Contudo, os testes de diagnóstico expõem falhas importantes: os resíduos não seguem uma distribuição normal (teste Jarque-Bera com Prob(JB) = 0.00), principalmente devido a uma alta curtose que sugere a presença de outliers. Além disso, o teste de heterocedasticidade (Prob(H) = 0.00) confirma que a variância dos erros não é constante.

A previsão gerada, que exhibe uma tendência de queda em vez de se manter no último valor observado, mostra que o modelo ajustado foi, na verdade, uma "caminhada aleatória com drift negativo". Em conclusão, este modelo baseline cumpriu seu papel ao demonstrar que, embora a diferenciação trate a correlação, ela é insuficiente para lidar com a variância crescente e os valores extremos da série. Esses resultados validam a necessidade de avançar para uma abordagem mais robusta, como um modelo de

regressão sobre a variável transformada ($\log(\text{volume})$), para tratar essas questões de forma mais eficaz.

Regressão da tendencia no log

ARIMA(0,1,0)

```

--- Resumo do Modelo de Regressão ---
=====
                        OLS Regression Results
=====
Dep. Variable:          log_volume    R-squared:                0.665
Model:                  OLS           Adj. R-squared:            0.659
Method:                 Least Squares   F-statistic:              116.0
Date:                  Mon, 06 Oct 2025   Prob (F-statistic):       1.70e-28
Time:                  01:46:48         Log-Likelihood:          -53.861
No. Observations:      120             AIC:                    113.7
Df Residuals:          117             BIC:                    122.1
Df Model:               2
Covariance Type:       nonrobust
=====
                        coef    std err          t      P>|t|      [0.025    0.975]
-----
const                0.1907      0.103      1.845     0.068     -0.014     0.396
week_num              0.0120      0.004      2.981     0.003     0.004     0.020
week_squared          2.869e-05    3.27e-05     0.879     0.381    -3.6e-05    9.34e-05
=====
Omnibus:                 8.753   Durbin-Watson:           0.134
Prob(Omnibus):           0.013   Jarque-Bera (JB):        7.593
Skew:                    0.532   Prob(JB):                0.0225
Kurtosis:                2.380   Cond. No.                1.88e+04
=====

--- Comparação de Desempenho no Conjunto de Teste ---
RMSE do Modelo Baseline (ARIMA): 5.4411
RMSE do Modelo de Regressão: 1.8997
-----
MAE do Modelo Baseline (ARIMA): 4.6803
MAE do Modelo de Regressão: 1.4191

```

Figura 7

A conclusão desta análise clara, a abordagem de regressão linear para modelar a tendência em $\log(\text{volume})$ é superior ao modelo baseline de random walk. Isso é comprovado pela redução de quase 60% no erro de previsão (RMSE), que caiu de 4.68 para 1.90. O modelo final indica que o volume possui uma forte tendência de crescimento linear, e a transformação logarítmica foi essencial para estabilizar os dados.

Apesar do melhor desempenho preditivo, a análise dos resíduos do modelo de regressão deixa espaço pra melhoria. O teste Durbin-Watson (0.134) aponta que ainda existe autocorrelação nos erros, uma característica temporal que o modelo atual não capturou.

Este resultado estabelece um modelo robusto como ponto de partida. Para a próxima fase o foco será refinar este modelo, utilizando mais dados para corrigir a autocorrelação residual e aprimorar ainda mais a precisão das previsões.