

Análise Interpretativa do Pipeline de Previsão para a Série Temporal *AirPassengers*

Análise da Execução do Código

29 de outubro de 2025

Conteúdo

1	Introdução	3
2	Explorando os Dados: O Ponto de Partida	3
2.1	Inspeção Visual da Série	3
2.2	Decomposição da Série	4
3	A Questão da Estacionariedade	4
3.1	Provas Estatísticas	4
4	Metodologia de Modelagem e Validação	5
4.1	Divisão e Validação	5
4.2	Os Contendores	6
5	Análise de Desempenho	6
5.1	Interpretação das Métricas	6
5.2	A Análise Gráfica (Figura 4)	6
6	Diagnóstico dos Resíduos	7
6.1	SARIMA (Linha Superior)	8
6.2	Random Forest (Linha Inferior)	8
7	Conclusão	9

1 Introdução

Este relatório apresenta uma análise dos resultados do pipeline de modelagem de séries temporais executado sobre o clássico conjunto de dados *AirPassengers*. O objetivo não é apenas reportar as métricas, mas interpretar o **porquê** de cada etapa e **o que** os resultados nos dizem sobre os modelos testados.

O pipeline executou uma análise exploratória, testes de estacionariedade, e comparou o desempenho de um modelo estatístico (SARIMA) contra um modelo de **machine learning** (RandomForest). A validação foi realizada usando a metodologia de *walk-forward*, que simula a operação de um modelo em um cenário real.

2 Explorando os Dados: O Ponto de Partida

A primeira etapa de qualquer análise é entender profundamente os dados. O conjunto carregado contém 144 observações mensais, de Janeiro de 1949 a Dezembro de 1960.

2.1 Inspeção Visual da Série

O gráfico da série temporal (Figura 1) é revelador. Duas características saltam aos olhos imediatamente:

1. **Tendência (Trend):** Há uma clara tendência de crescimento ao longo dos anos. O número de passageiros não é estático; ele aumenta consistentemente década após década.
2. **Sazonalidade (Seasonality):** Existe um padrão que se repete a cada ano, com picos nos meses de verão (meio do ano) e vales nos meses de inverno.

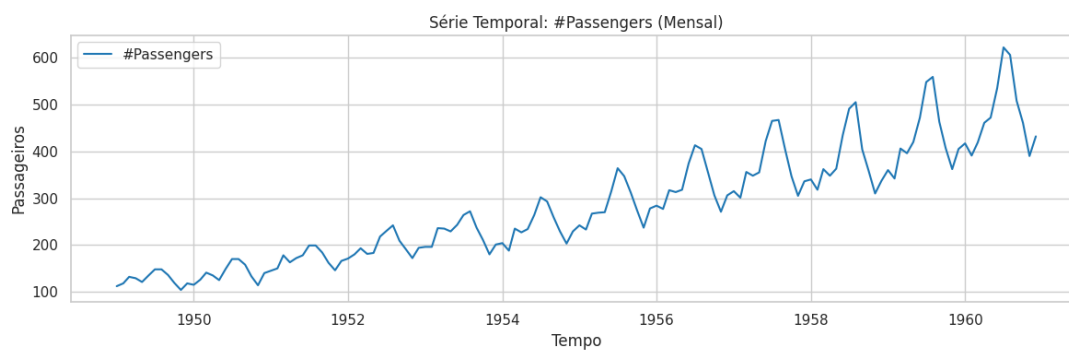


Figura 1: *Série temporal original AirPassengers (1949-1960).*

Um ponto crucial, muitas vezes ignorado, é que a **variância da sazonalidade parece aumentar com o tempo**. Os picos estão ficando mais altos e os vales também (embora em menor grau). Isso sugere que o efeito sazonal não é aditivo (ex: "todo verão vendemos +50"), mas sim **multiplicativo** (ex: "todo verão vendemos 20% a mais que a média").

2.2 Decomposição da Série

A escolha de um modelo de decomposição multiplicativo (Figura 2) foi acertada, justamente para lidar com essa variância crescente. A decomposição nos permite isolar os componentes:

- **Trend (Tendência):** Mostra a curva de crescimento de longo prazo, suavizando os picos e vales sazonais.
- **Seasonal (Sazonal):** Isola perfeitamente o padrão de 12 meses.
- **Resid (Resíduos):** O que sobra. Idealmente, isso deve ser apenas ruído aleatório. O fato de os resíduos flutuarem em torno de 1.0 (e não de 0) é a característica de um modelo multiplicativo.

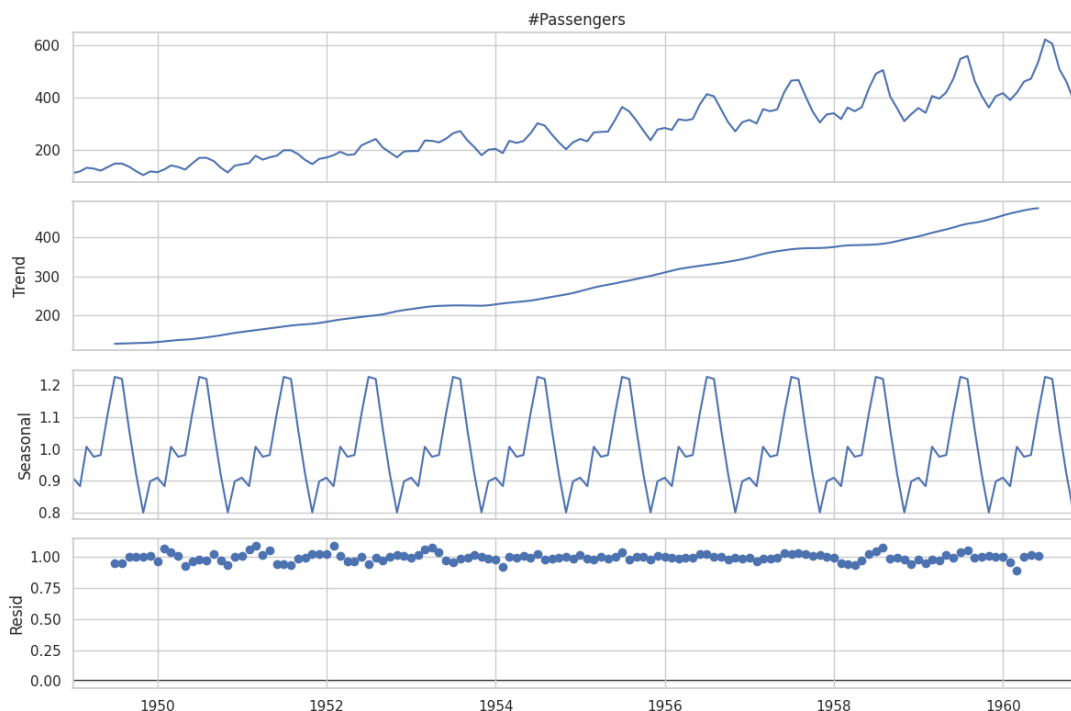


Figura 2: *Decomposição multiplicativa da série.*

3 A Questão da Estacionariedade

Modelos estatísticos como o ARIMA exigem que a série seja **estacionária**, ou seja, que sua média, variância e autocorrelação não mudem ao longo do tempo. Nossos dados, como vistos, violam claramente essa premissa (a média cresce!).

3.1 Provas Estatísticas

A análise visual é confirmada pelos gráficos de autocorrelação (Figura 3) e pelos testes estatísticos formais.

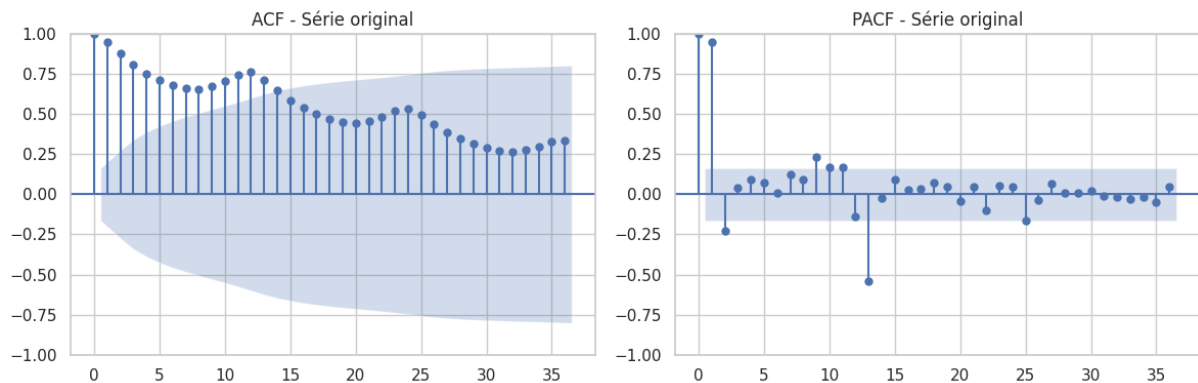


Figura 3: Funções de Autocorrelação (ACF) e Autocorrelação Parcial (PACF) da série original.

- **Gráfico ACF (esquerda):** O decaimento *extremamente lento* das barras é o sinal clássico de uma forte tendência (não-estacionariedade). Os "dentes" no decaimento indicam a sazonalidade.
- **Testes Formais:** O pipeline executou dois testes com hipóteses opostas, uma prática excelente para confirmação.
 - **ADF ($p=0.9919$):** Falha categoricamente em rejeitar a hipótese nula de que a série **não é estacionária**.
 - **KPSS ($p=0.0100$):** Rejeita fortemente a hipótese nula de que a série **é estacionária**.

Conclusão Inequívoca: A série original não é estacionária e precisa de tratamento (transformação e/ou diferenciação) antes da modelagem.

4 Metodologia de Modelagem e Validação

O pipeline comparou dois mundos: o estatístico (SARIMA) e o de **machine learning** (Random Forest).

4.1 Divisão e Validação

Os dados foram divididos com os últimos 24 meses (1959-1960) servindo como conjunto de teste. Mais importante, foi usada a validação **Walk-Forward (passo=1)**.

Isso significa que, para prever cada um dos 24 pontos de teste, os modelos foram **completamente re-treinados** usando todos os dados disponíveis até o mês anterior. Esta é uma simulação precisa de como um modelo seria usado na prática (re-treinando à medida que novos dados chegam) e é muito mais confiável do que um simples split "treino-teste".

4.2 Os Contendores

1. **SARIMA (SARIMAX):** O modelo estatístico. O pipeline realizou uma busca (grid-search) por AIC e encontrou a ordem ótima: `order=(1, 0, 0)` e `seasonal_order=(1, 0, 1, 12)`.
 - **Ponto-chave:** O modelo foi ajustado aos dados **transformados por log** (para estabilizar a variância) e com um componente de `trend='c'` (constante), que ajuda a capturar o crescimento exponencial implícito.
2. **RandomForest (com lags):** A abordagem de ML. O modelo não "vê" o tempo, então criamos *features* para ele: os 12 valores anteriores (lags), médias e desvios-padrão móveis (janela de 12), e o número do mês (para capturar a sazonalidade).
3. **Naive Sazonal (Baseline):** Nosso ponto de referência. Ele simplesmente prevê que o valor deste mês será igual ao valor do mesmo mês no ano anterior (ex: prev. de Jan/1959 = real de Jan/1958). Um modelo só é útil se superar esse baseline.

5 Análise de Desempenho

A fase de testes revela um vencedor claro. As métricas de erro (quanto menor, melhor) estão resumidas na Tabela 1.

Tabela 1: *Métricas de Erro no Conjunto de Teste (Walk-Forward, $h=24$).*

Modelo	MAE	RMSE	MAPE (%)
SARIMA (SARIMAX)	11.15	15.31	2.44%
RandomForest (lags)	23.84	32.76	5.02%
Naive Sazonal	47.58	49.99	10.52%

MAE: Erro Médio Absoluto. RMSE: Raiz do Erro Quadrático Médio. MAPE: Erro Percentual Médio Absoluto.

5.1 Interpretação das Métricas

O **SARIMA foi o vencedor indiscutível**. Seus erros (MAE, RMSE, MAPE) são, em média, **menos da metade** dos erros do RandomForest. Ambos os modelos superaram confortavelmente o baseline Naive Sazonal, provando seu valor. O MAPE de apenas 2.44% para o SARIMA é um resultado excelente.

5.2 A Análise Gráfica (Figura 4)

O gráfico das previsões conta a história por trás dos números:

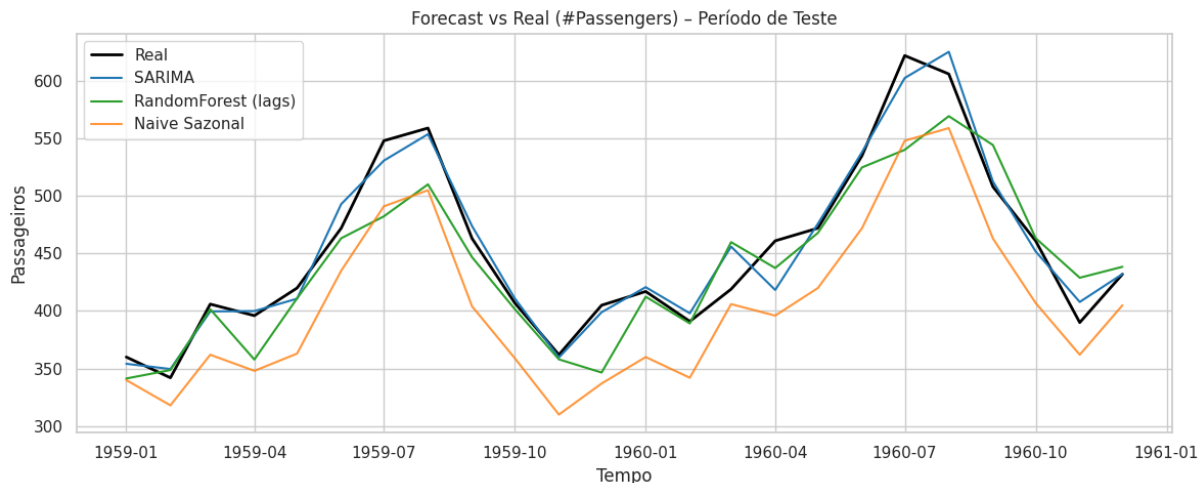


Figura 4: Comparação das previsões vs. valores reais no período de teste.

- **Real (Preto):** Os dados reais que tentamos prever.
- **SARIMA (Azul):** Esta linha "abraça" os dados reais. Ela não apenas acerta o padrão sazonal, mas também captura perfeitamente a magnitude crescente dos picos.
- **RandomForest (Verde):** O RF entendeu a sazonalidade (sobe e desce nos meses certos, graças à *feature* "mês"), mas falhou em extrapolar a tendência de crescimento. Suas previsões parecem "achatadas", subestimando os picos e superestimando os vales.
- **Naive Sazonal (Laranja):** Como esperado, ele segue o padrão sazonal, mas fica consistentemente abaixo do real, pois não tem nenhum mecanismo para aprender sobre a tendência de crescimento ano a ano.

6 Diagnóstico dos Resíduos

Um bom modelo deve deixar para trás apenas "ruído branco"—erros aleatórios e imprevisíveis. Se houver padrões nos erros, significa que o modelo deixou de capturar alguma informação.

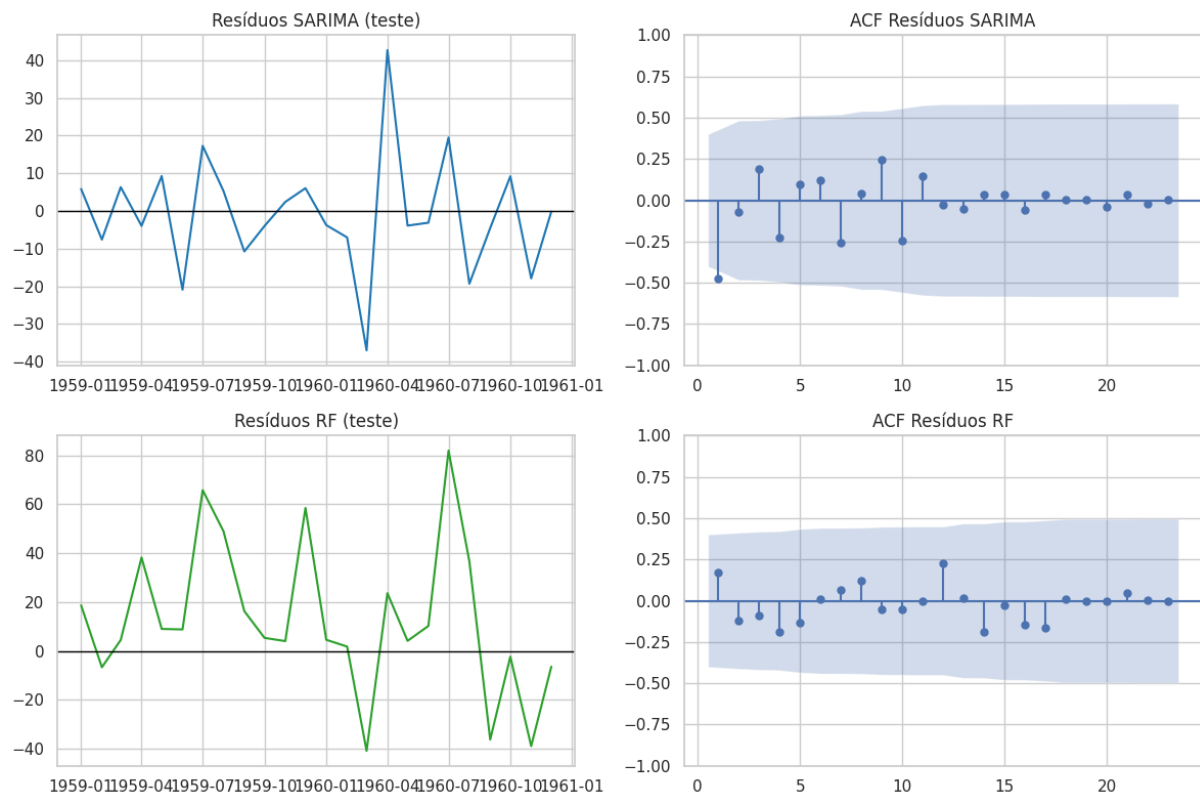


Figura 5: Diagnóstico dos resíduos (erros) dos modelos SARIMA e RF.

6.1 SARIMA (Linha Superior)

- **Gráfico de Resíduos (esquerda):** Os erros parecem aleatórios, flutuando em torno do zero, exceto por uma grande subestimação no início de 1960.
- **ACF dos Resíduos (direita):** Este é o gráfico mais importante. **Nenhuma barra sai da área azul (intervalo de confiança).** Isso significa que não há autocorrelação estatisticamente significativa nos erros.
- **Teste Ljung-Box ($p=0.1003$):** Este teste formal confirma o gráfico. Como o p-valor é maior que 0.05, **falhamos em rejeitar a hipótese nula** de que os resíduos são independentes e não correlacionados (ou seja, são ruído branco).

Veredito do SARIMA: Excelente. O modelo extraiu toda a estrutura previsível dos dados.

6.2 Random Forest (Linha Inferior)

- **Gráfico de Resíduos (esquerda):** Os erros *não são* aleatórios. Eles mostram um padrão claro: são negativos nos picos (subestimação) e positivos nos vales (superestimação).
- **ACF dos Resíduos (direita):** Vemos várias barras (lags 1, 12, 13) saindo da zona de confiança, confirmando que os erros de hoje estão correlacionados com os erros passados.

- **Teste Durbin-Watson (1.389):** Um valor longe de 2.0, indicando autocorrelação positiva, o que bate com o pico no lag 1 do ACF.

Veredito do RF: Subajustado (Underfit). O modelo falhou em capturar toda a estrutura da série, deixando padrões previsíveis nos resíduos.

7 Conclusão

O pipeline foi executado com sucesso e forneceu uma comparação clara.

Para esta série temporal, que é dominada por uma tendência forte e uma sazonalidade multiplicativa, o modelo estatístico **SARIMA (especificamente, SARIMAX do statsmodels)** foi **dramaticamente superior** à abordagem de *machine learning* (Random Forest).

O modelo SARIMA, ajustado sobre dados transformados por log e com a ordem $(1,0,0) \times (1,0,1,12)$, não apenas produziu os menores erros (MAPE de 2.44%) mas, crucialmente, foi o único modelo a passar nos testes de diagnóstico, produzindo resíduos que se assemelham a ruído branco. Isso indica um modelo bem especificado, confiável e que capturou a essência dos dados.