

RELATÓRIO TÉCNICO

1º Avaliação

Disciplina: Séries Temporais

Professor: Vinícius

Aluno: Júlio Henri

Matrícula: 201807840011

1. Introdução

O presente trabalho é dividido em duas partes, onde a primeira parte tem por objetivo apresentar características de uma série temporal, tais como tendência, sazonalidade e ruído por meio de simulação, bem como explica-las e dar exemplos.

Enquanto na segunda parte do projeto será realizado uma análise de uma série temporal real, ilustrando com gráficos, fazendo a exploração dos dados, separando os dados de treino e de teste e fazendo a comparação de dois modelos estatísticos cedidos pelo professor.

2. Materiais e Métodos

Os dados dos tópicos **3. Tendência** ao **6. Estacionariedade** foram gerados através do Python. Para isso, foram utilizadas as bibliotecas Pandas, Numpy e Random que geraram valores pseudoaleatórios para a exemplificação dos termos citados, além disso, os gráficos foram construídos pelas bibliotecas Matplotlib e Seaborn. Por outro lado, do tópico **7. A série temporal** em diante os dados analisados foram disponibilizados pelo professor.

Os dados entregues são de natureza quantitativa e compreendem as informações de diversas séries temporais, tais arquivos (no formato .csv) contém valores de temperatura, umidade, consumo de algo, bebida, e o período no tempo para cada série temporal.

A análise estatística será inicialmente realizada por meio de análise exploratória de dados, objetivando fornecer uma visão geral do estudo e delinear possíveis características estruturais que possam ser avaliadas em maiores detalhes com o auxílio de testes estatísticos.

A análise exploratória de dados será conduzida por meio de testes e gráficos. Além dos gráficos de linhas, que serão utilizados para visualizar o perfil da amostra, serão apresentados também histogramas, os quais são apropriados para perceber a distribuição dos dados. Nos gráficos de linha, há a obtenção visual da estacionariedade, no qual é comprovada (e explicada) posteriormente com um teste estatístico.

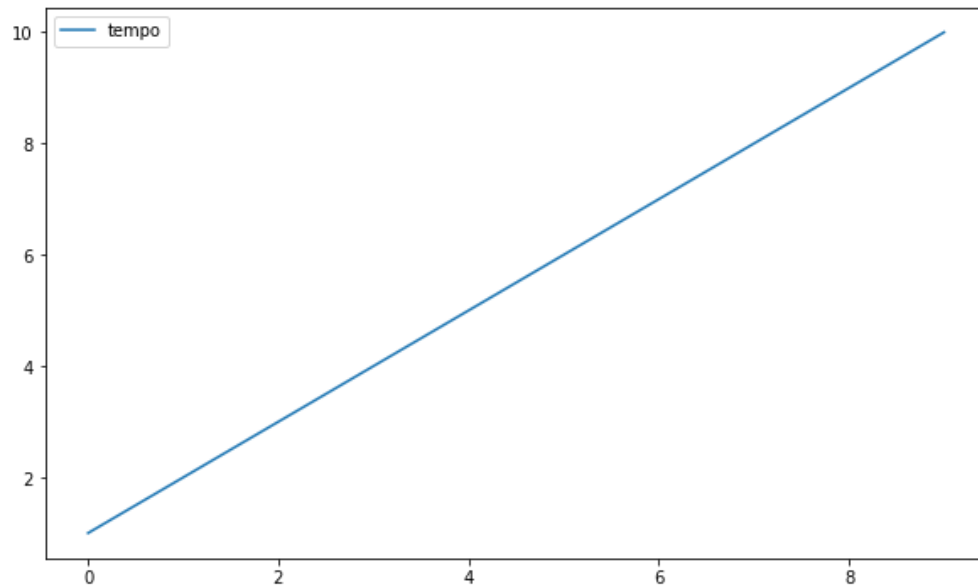
3. Tendência

A tendência é um conceito importante quando se trata de séries temporais e, nada mais é, que a elevação ou diminuição dos dados observados em um tempo especificado. A tendência indica se a série temporal cresce, decresce ou se mantém constante.

3.1 Exemplos

No primeiro exemplo é apresentado dados com tendência positiva, logo, conforme o tempo aumenta, os dados também aumentam

Figura 1: Exemplo de dados com tendência



O gráfico abaixo mostra uma tendência negativa, isso é confirmado ao observar os valores dos dados caírem conforme o tempo aumenta.

Figura 2: Exemplo de dados com tendência

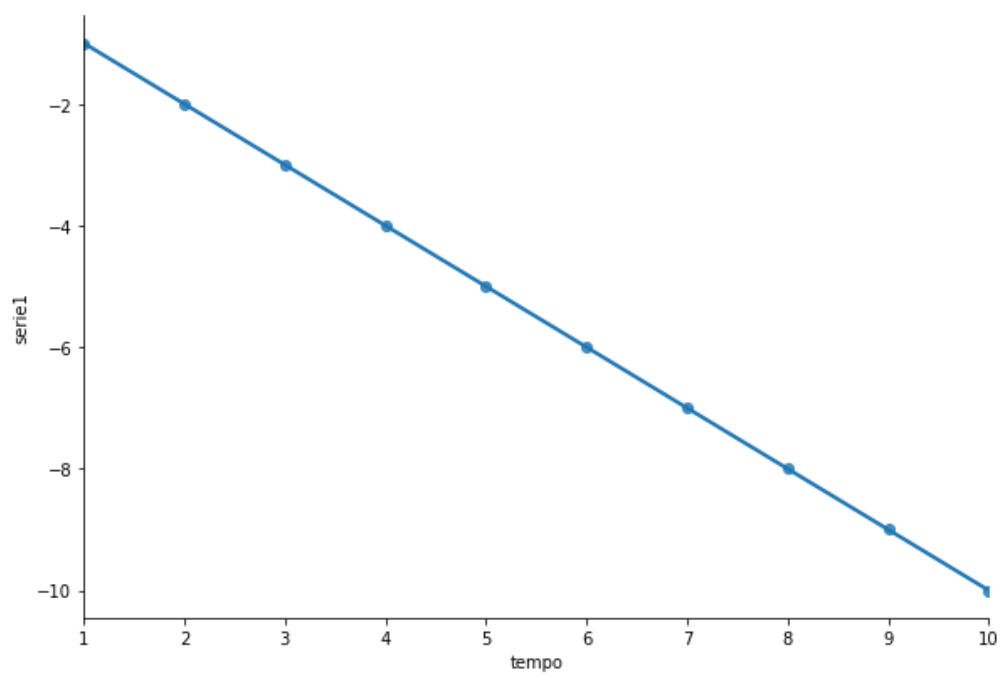


Figura 3: Exemplo de dados com tendência

A Figura 3 expõe dados com padrão de crescimento no período, de forma exponencial.

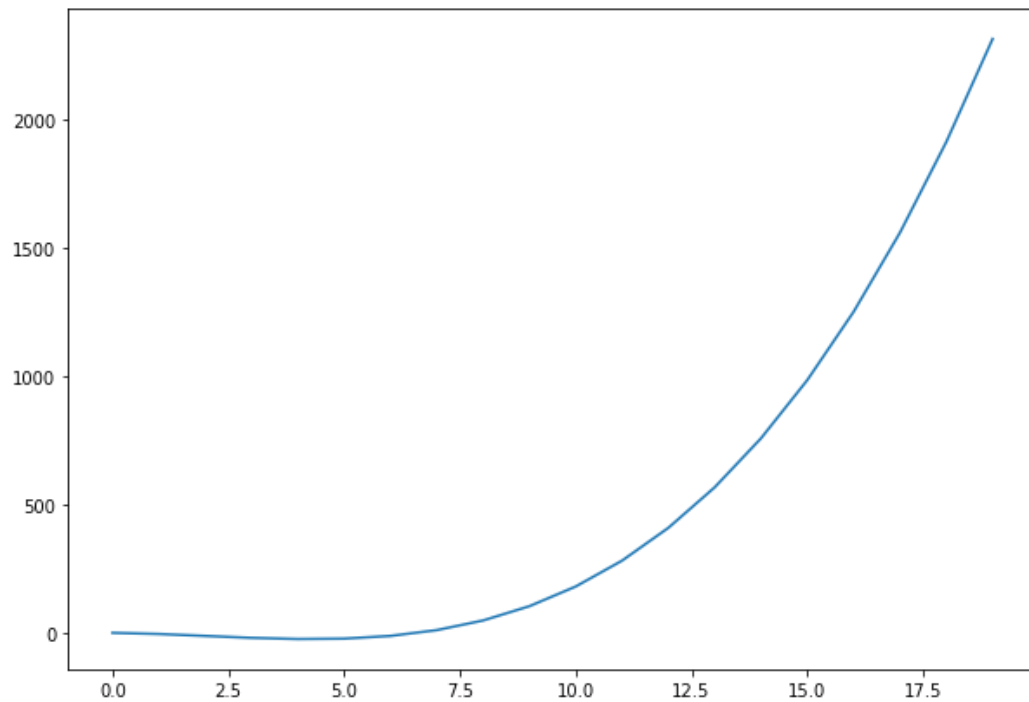
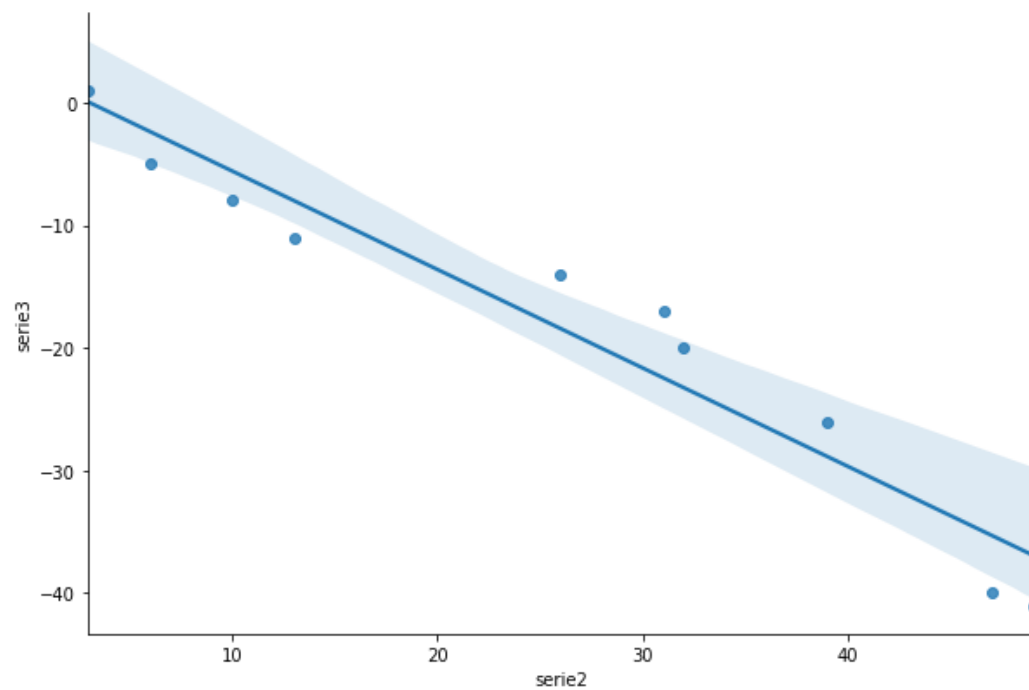


Figura 4: Exemplo de dados com tendência

Novamente é exposto um gráfico de dados aleatórios e com tendência de decrescimento.



4. Sazonalidade

Sazonalidade refere-se a flutuações periódicas fixas que tem um comportamento que se repete em um período específico, esse padrão pode ser por conta da estação do ano, datas comemorativas ou qualquer outro fator externo que se repita e afete a série. É comumente identificada somente observando o gráfico da série ou por meio de autocorrelações relevantes. Por exemplo, uma série temporal que marca a temperatura do país por hora e constata o pico dos dados às 12 horas todo dia.

Nesse sentido, é importante salientar que a sazonalidade pode ser dividida em sazonalidade determinística e sazonalidade estocástica. A sazonalidade determinística acontece quando o padrão dos dados é constante, enquanto a sazonalidade estocástica refere-se quando o componente sazonal não é constante, ou seja, ela oscila junto ao período. Convém também explicitar que período é um intervalo de tempo específico.

Um teste comum para verificar a sazonalidade é o Teste de *Kruskal-Wallis*, que pode ser realizado pela biblioteca *Scipy* no programa Python. O teste em questão testa a hipótese nula de que a média da população dos grupos são iguais, é uma versão da ANOVA não paramétrica.

4.1 Exemplos

Figura 5: Exemplo de dados com sazonalidade

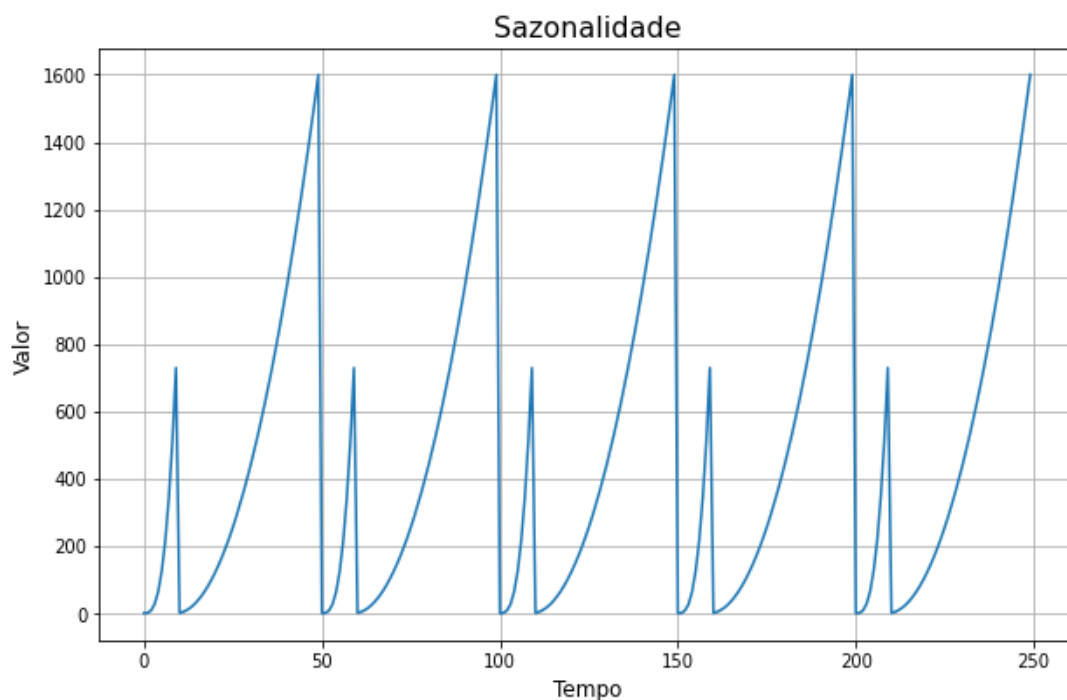


Figura 6: Exemplo de dados com sazonalidade

O gráfico abaixo mostra uma sazonalidade mais comum com repetição aproximadamente a cada 50 valores de tempo.

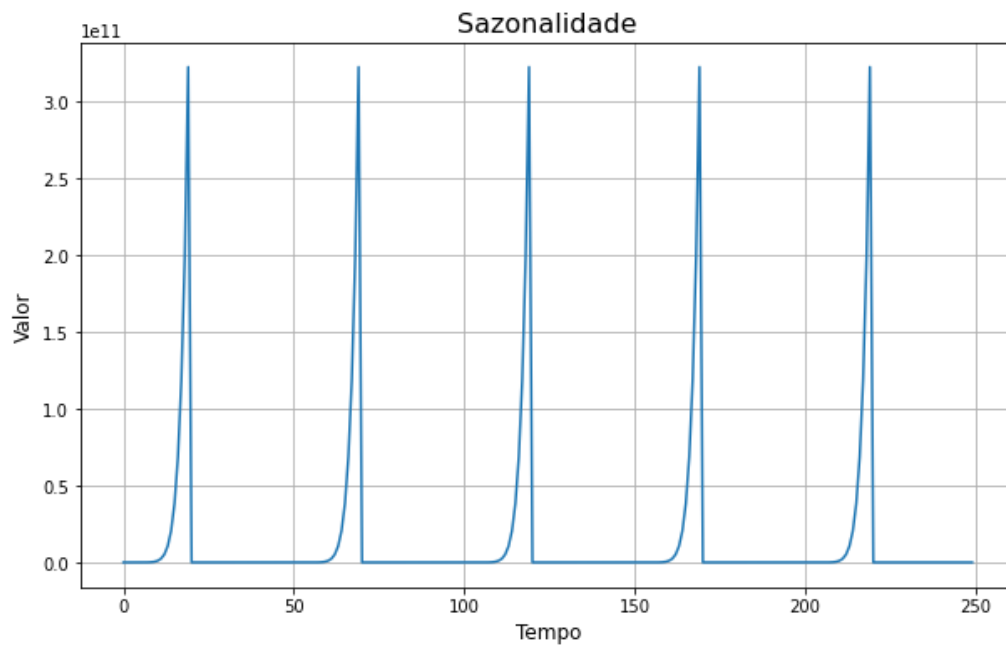


Figura 7: Exemplo de dados com sazonalidade

O terceiro exemplo, visualmente, difere dos anteriores mas continua apresentando a sazonalidade nos dados.

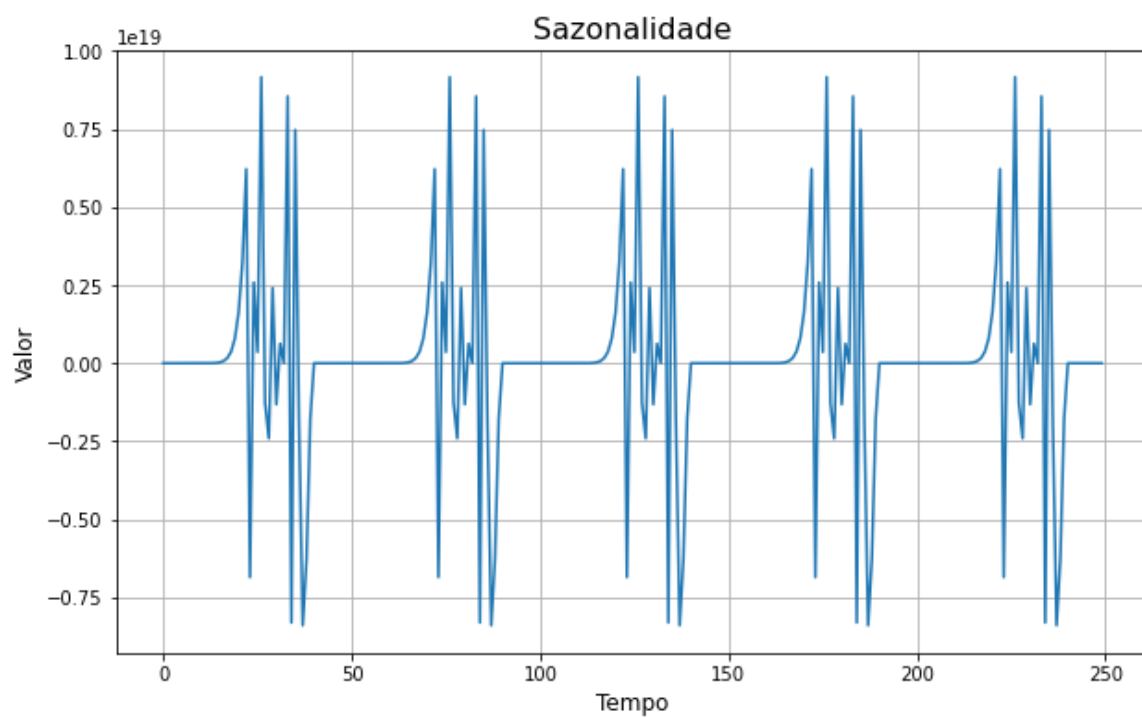
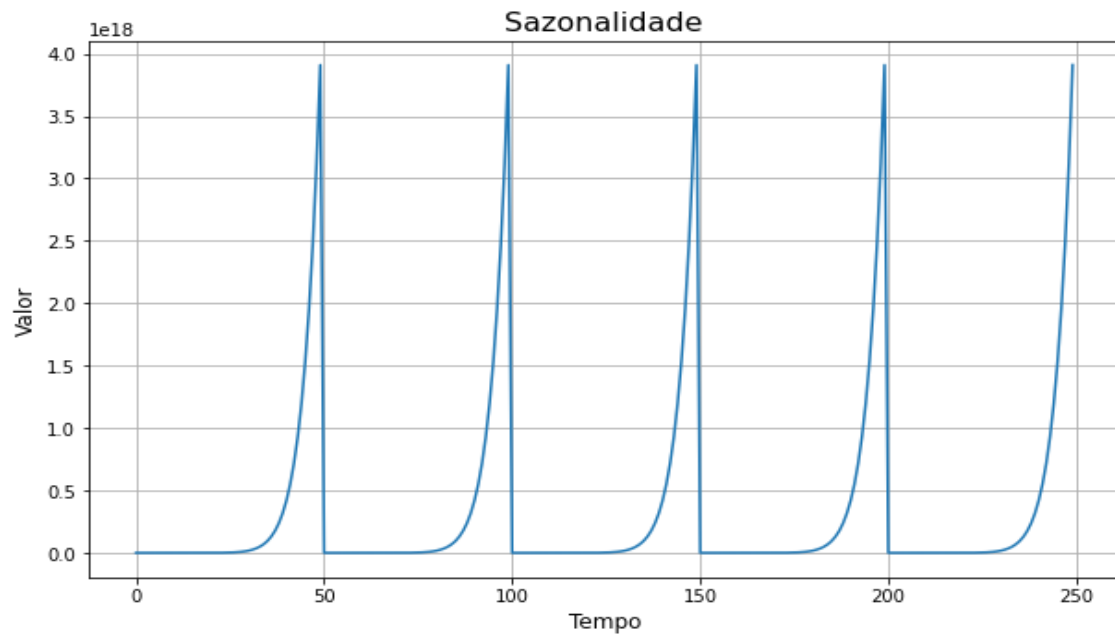


Figura 8: Exemplo de dados com sazonalidade



5. Componente Aleatório

Também chamado de ruído, o componente aleatório é uma variável na qual não se pode prever, são variações inexplicadas dos dados, que pode ser de mudança de governante da cidade, novas obras, modificação climática ou diversas outras situações.

É importante ressaltar o termo ruído branco, que é uma série de variáveis independentes e identicamente distribuídas, onde não há autocorrelação significativa e a variância é constante entre as variáveis. Há uma variante deste termo, que é o ruído branco gaussiano, que é aquele no qual as variáveis aleatórias seguem uma distribuição Normal.

Figura 9: Exemplo série temporal com ruído branco

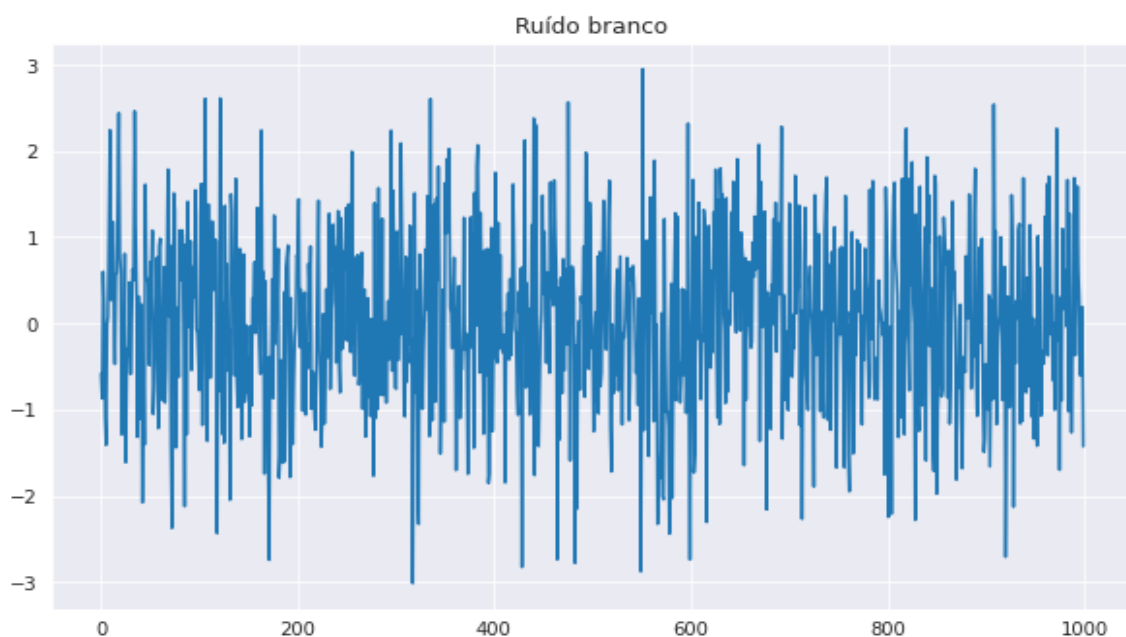
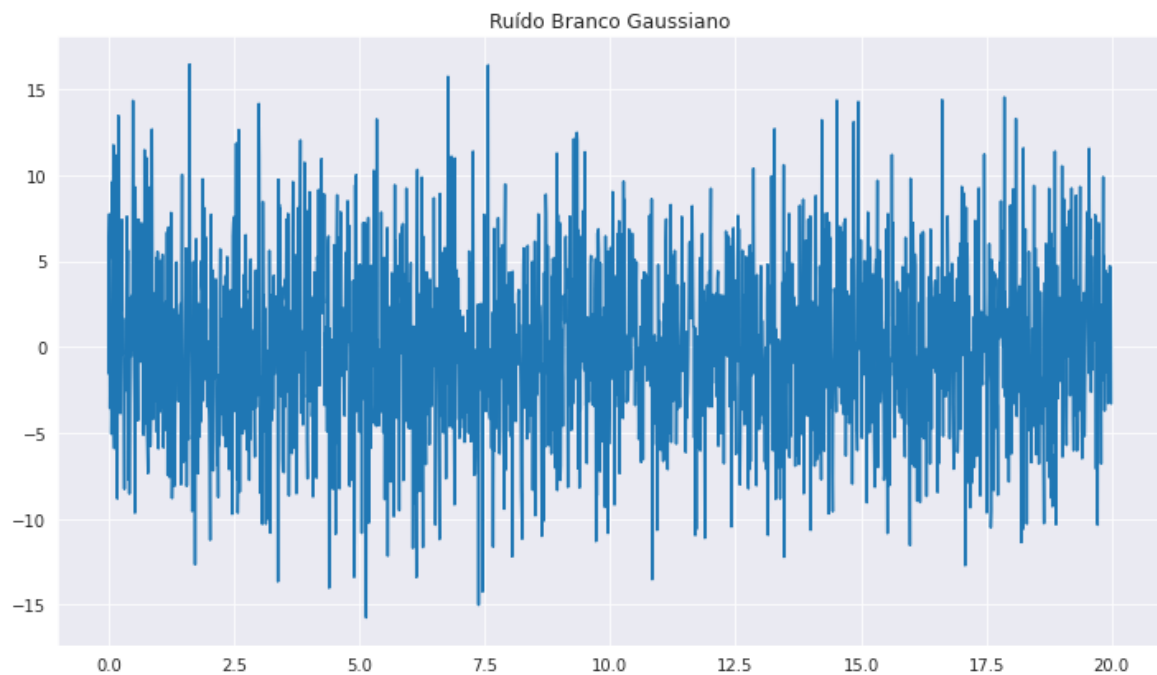


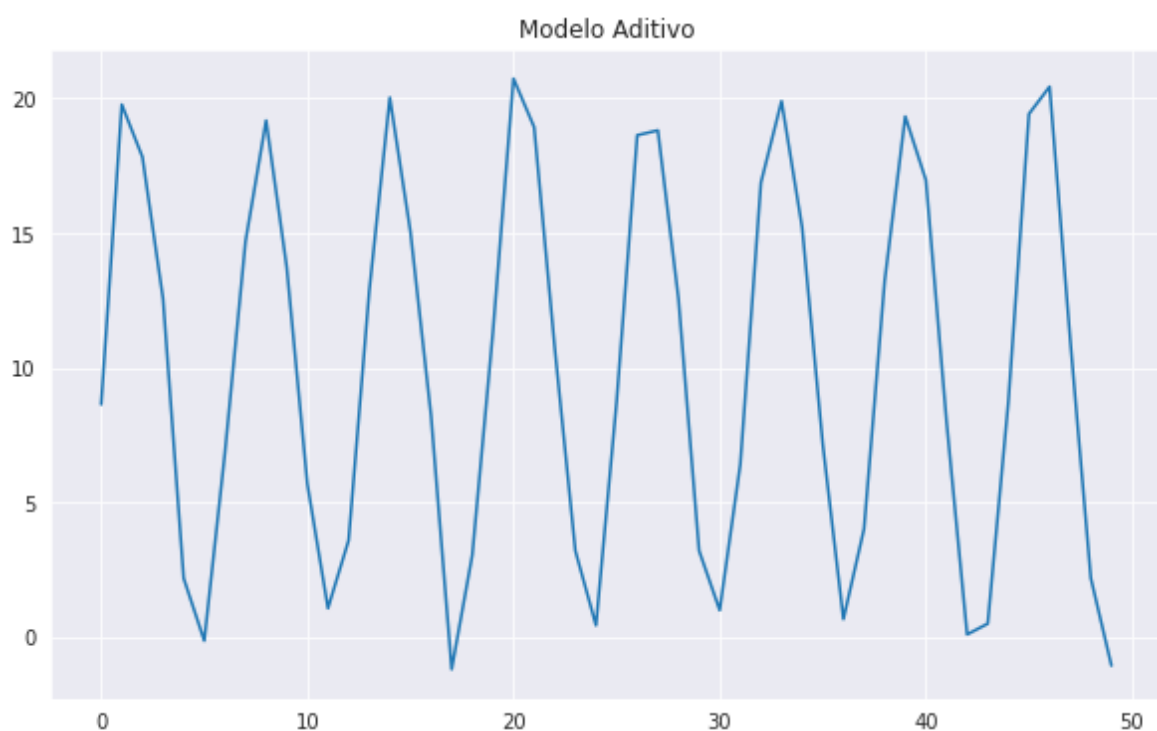
Figura 10: Exemplo série temporal com ruído branco gaussiano



6. Modelos Aditivos

No modelo aditivo o valor da série é a relação da soma dos valores dos componentes da mesma, ou seja, o resultado será a tendência, sazonalidade e o componente aleatório somados. É importante salientar que todos os componentes estão na mesma medida e o cálculo é dado por: $Z_t = T_t + S_t + a_t$, onde T_t é a tendência e S_t representa a sazonalidade, já a_t é o componente aleatório (elemento estocástico).

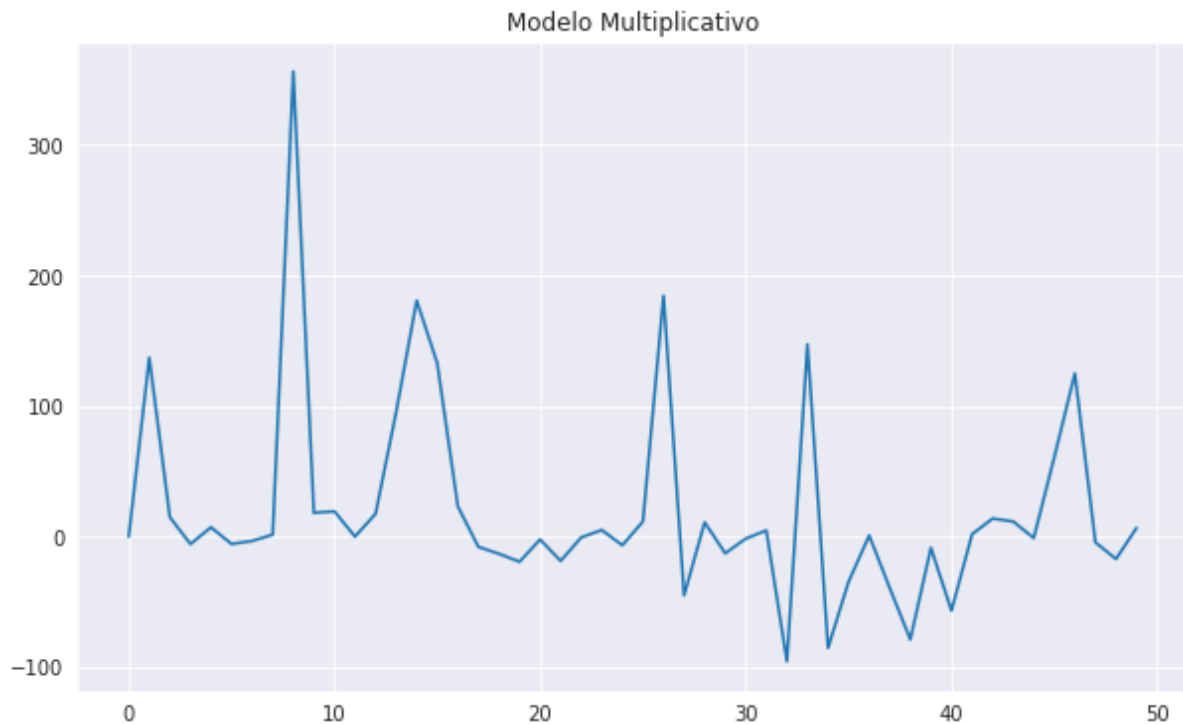
Figura 11: Exemplo série temporal aditiva



7. Modelos multiplicativos

A equação que correlaciona as variáveis com os componentes (tendência, sazonalidade e componente aleatório) é importante para o modelo, e pode ser escolhido o modelo aditivo ou modelo multiplicativo. O modelo multiplicativo, como o próprio nome diz, gera o valor da série a partir da multiplicação dos seus componentes, ou seja, seu cálculo é dado por: $Z_t = T_t * S_t * a_t$, onde T_t é a tendência e S_t representa a sazonalidade, já a_t é o componente aleatório.

Figura 12: Exemplo série temporal multiplicativa



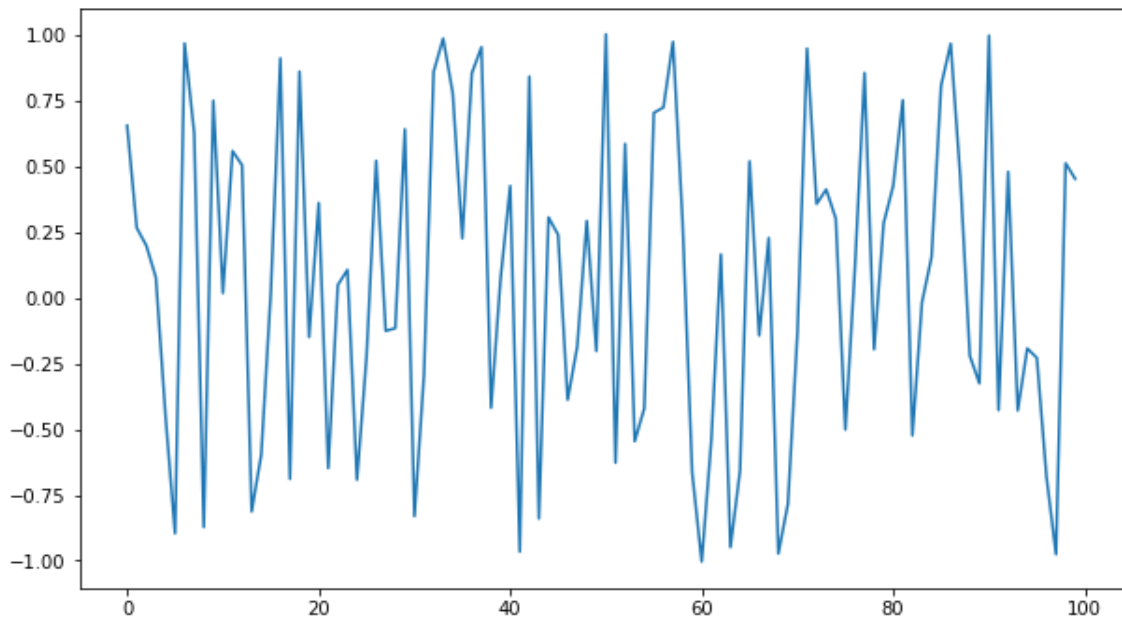
8. Estacionariedade

A estacionariedade é um pressuposto para diversos testes, análises e previsões de uma série temporal. Este pré-requisito tem como característica a média e variância constantes no decorrer de um período, logo, qualquer série que tenha confirmado sua tendência e/ou sazonalidade não é estacionária.

Nesse sentido, há o Teste de *Dickey-Fuller* (ADF) que indica, estatisticamente, se o teste é ou não estacionário. A biblioteca *statsmodels* do Python tem um pacote chamado *adfuller* (ADF) e ele entrega o resultado do P-Valor para valores críticos de 1%, 5% e 10%, além da estatística do teste.

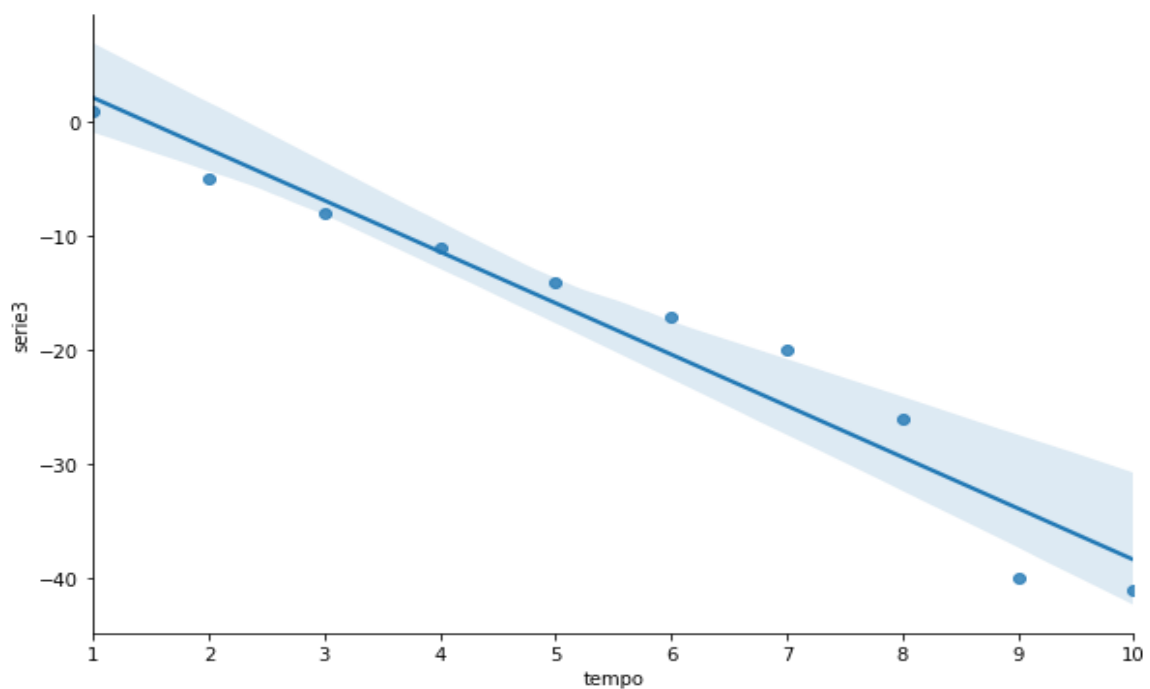
8.1 Exemplos

Figura 13: Exemplo de série temporal estacionária



Como explicado anteriormente, é possível notar que há uma variação entre os dados mas, visualmente, os dados parecem estar variando em torno de uma média e não têm sazonalidade e tendência. Para confirmar esta hipótese será feito o teste ADF e a análise do gráfico de autocorrelação.

Figura 14: Exemplo de série temporal não-estacionária

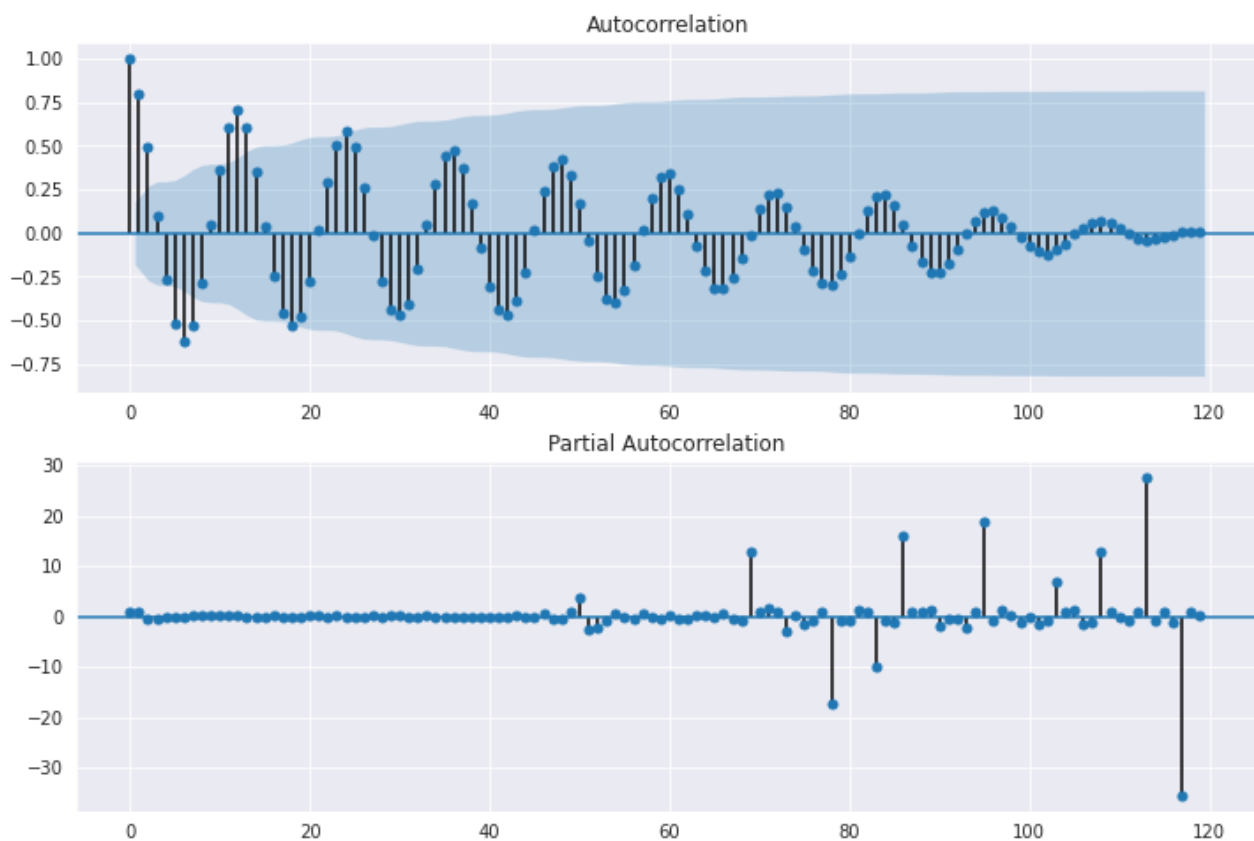


8.3 Autocorrelação

A existência da auto correlação é um pressuposto para diversos modelos de previsão de séries temporais, pois através da auto correlação é possível saber se há correlação de um determinado tempo anterior com um período recente, ou seja, as variáveis serão auto correlacionadas quando elas têm variação similar em relação ao desvio padrão.

Analizando o gráfico de autocorrelação abaixo, é possível notar que os lags não apresentam queda rápida e que há um comportamento cíclico a cada 5 lags, o que indica uma sazonalidade na série temporal. Já em relação ao gráfico de autocorrelação parcial, há picos a partir do lag 65, o que indica que a correlação para esses lags, é diferente de zero.

Figura 15: Gráfico de autocorrelação e autocorrelação parcial da série temporal de Ubatuba.



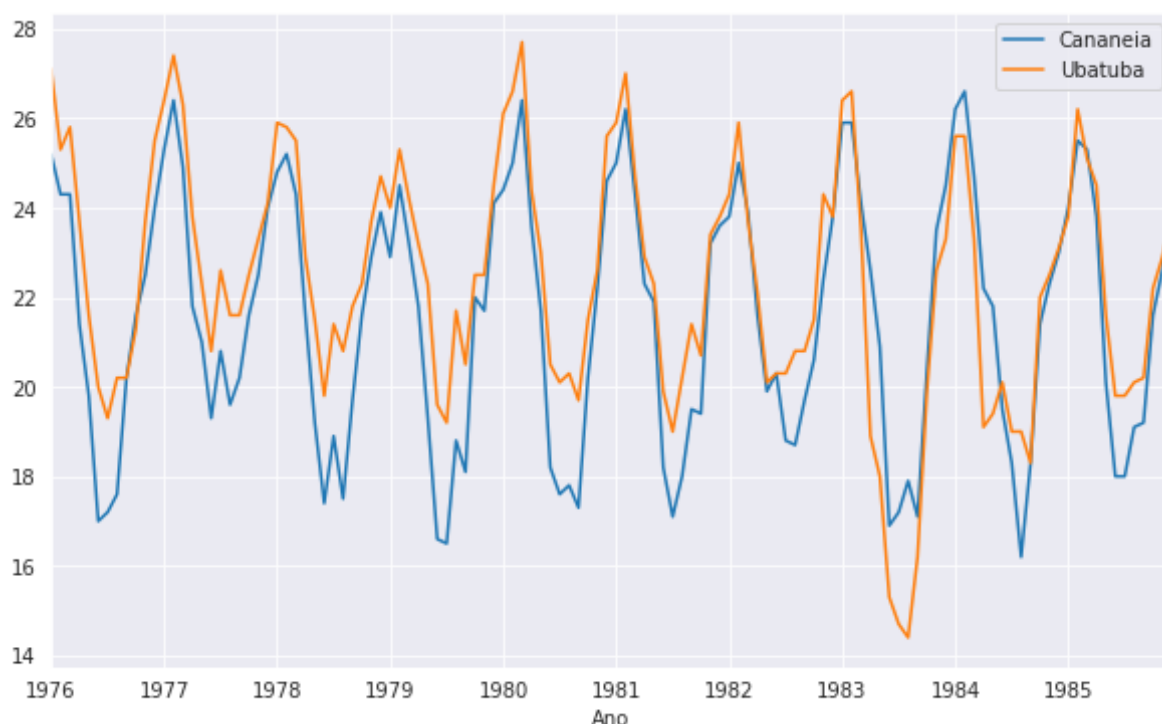
9. A série temporal

Dentre as diversas séries temporais de dados que foram disponibilizadas pelo professor, a escolhida foi um arquivo com 3 colunas e 121 registros chamado de “a1_temperatura.csv”, tais dados são compostos pelo período de tempo, na coluna “Ano”, uma coluna nomeada

“Cananeia” (um município no litoral de São Paulo) com valores de temperatura na cidade, além disso, há a última coluna que contém a temperatura do município de Ubatuba.

Para abrir o banco de dados, utilizei da biblioteca *Pandas* do Python e após manipular a coluna “Ano” para que fosse mais apropriada para analisar, foi possível a realização do gráfico abaixo. Ao observar a figura 13, é razoável dizer que as séries temporais das duas cidades são semelhantes, com as maiores variações ocorrendo entre 1976 e 1977 e em 1983 e 1984.

Figura 16: Temperatura das cidades de Cananeia e Ubatuba nos anos de 1976 a 1985.

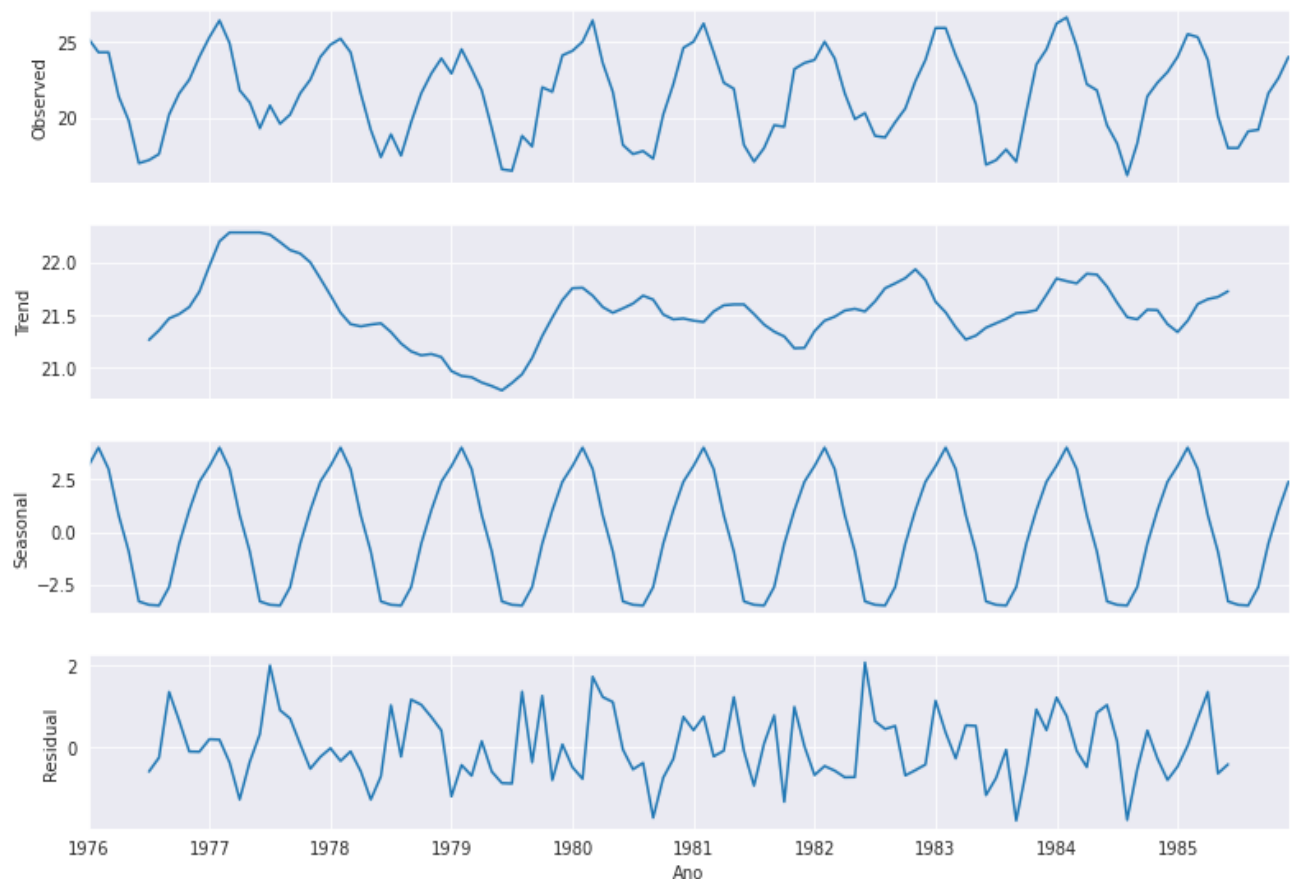


10. Análise e teste estatístico

Este tópico se refere a verificação das características da série temporal, isto é, observar se há ou não tendência, sazonalidade e observar o comportamento dos resíduos dos dados. Para tanto, utilizei novamente da biblioteca *statsmodels* que fornece tais informações em formato de gráfico, o que facilita a visualização.

Dessa maneira, se faz necessário também o teste estatístico. Foi escolhido o Teste de Dickey-Fuller Aumentado pela facilidade de uso e de interpretabilidade de seus resultados, além de haver uma biblioteca específica para tal no Python, também do *statsmodels*.

Figura 17: Série temporal da temperatura de Cananeia decomposta.



A partir da decomposição da série temporal, pode-se observar características importantes sobre os dados. O gráfico mais acima mostra os dados observados de fato, já o abaixo mostra a tendência dos dados (explicada anteriormente) e exibe que a tendência oscila com o tempo.

No gráfico posterior é possível a considerar a sazonalidade. O padrão nesse caso, aparentemente, é de crescimento nos meses iniciais e finais do ano enquanto nos meses do meio ocorre um decrescimento da temperatura, por fim é apresentado os resíduos da série.

10.1 Teste de *Dickey-Fuller*

Para a realização de modelos de previsão de séries temporais, um pressuposto importante é a estacionariedade dos dados, para isso, utiliza-se do teste de *Dickey-Fuller*, que é um teste intuitivo e o mais popular para verificar a estacionariedade dos dados.

Além disso, é importante ressaltar a interpretação do teste. Se o P-Valor apresentar um resultado acima do valor crítico escolhido (1%, 5% ou 10%) então o processo não é estacionário (tem raiz unitária), caso contrário, é estacionário (não tem raiz unitária).

Tabela 1: Resultados do Teste de *Dickey-Fuller*

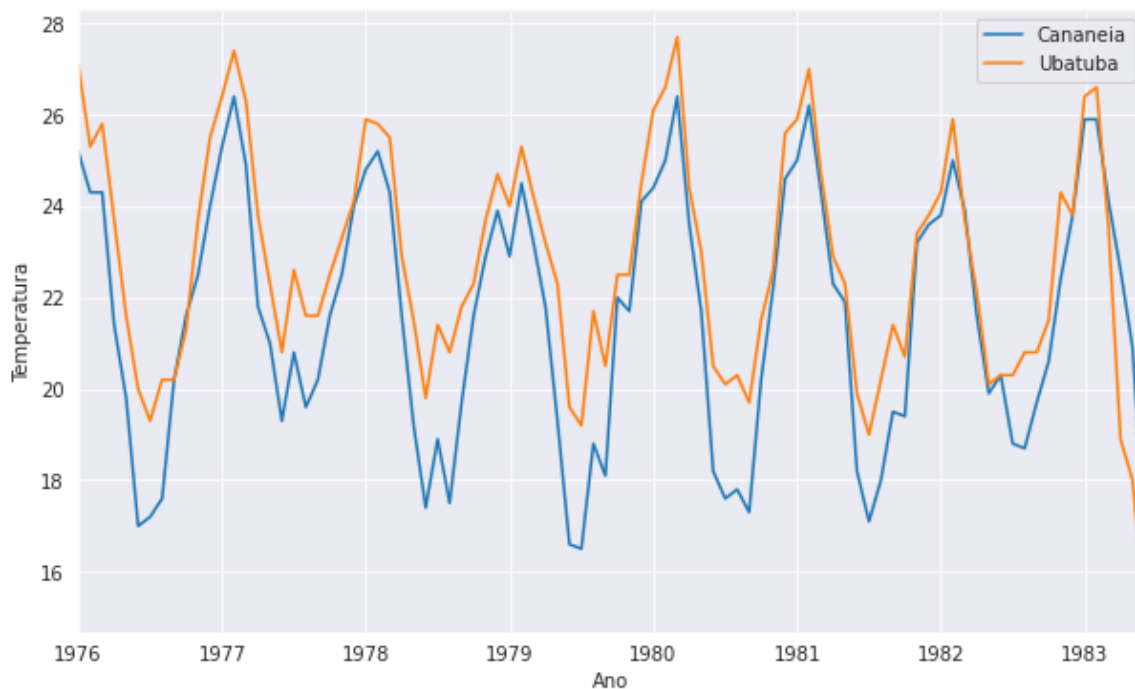
Valor Crítico	P -Valor Ubatuba
0,01	
0,05	0,427
0,1	

Como demonstrado na Tabela 1, o P-Valor de Ubatuba é maior do que todos os valores críticos apresentados, logo, é factível dizer que a série não é estacionária para qualquer valor crítico.

11. Separação dos dados

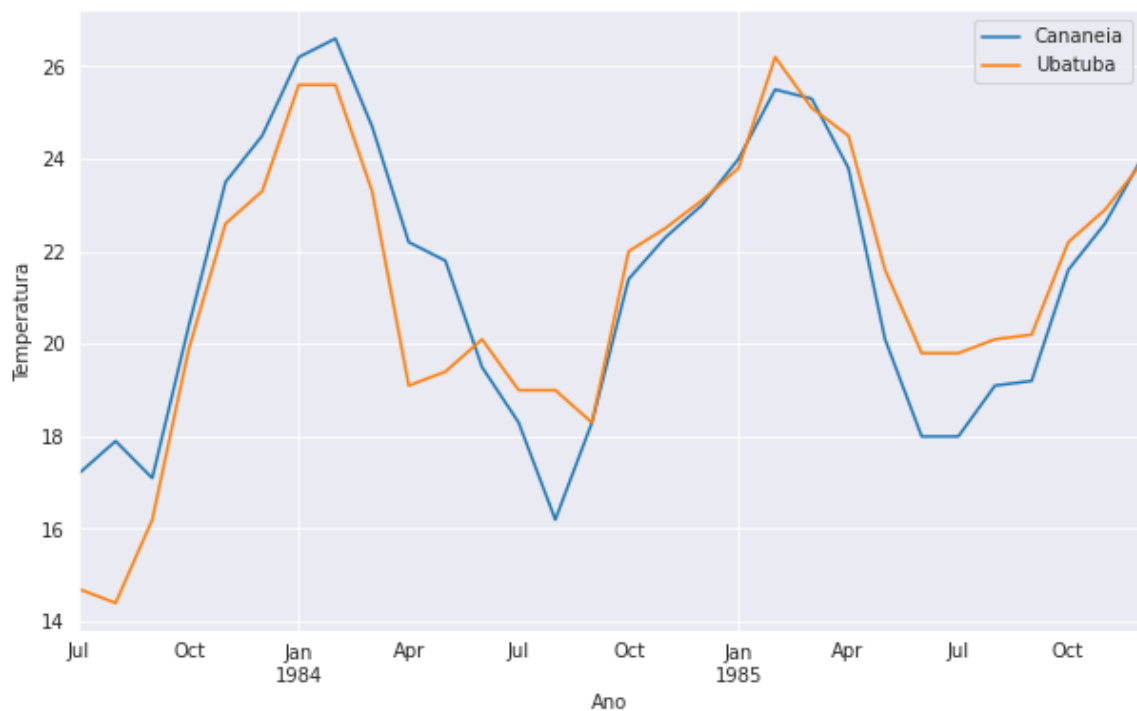
Os dados foram separados de maneira a dividir os 90 primeiros registros (de 31/01/1976 até 30/06/1983) para treino do modelo e os 30 restantes para formar o grupo de teste (de 31/07/1983 até 21/12/1985). Abaixo é apresentado o gráfico do banco de dados de treino.

Figura 18: Temperatura das cidades de Cananeia e Ubatuba de janeiro de 1976 até junho de 1983.



A Figura 19 demonstra a temperatura das cidades em questão nos dados que serão usados para o teste, ou seja, dados para validar o modelo e posteriormente confirmar a eficácia desse modelo.

Figura 19: Temperatura das cidades de Cananeia e Ubatuba de julho de 1983 até dezembro de 1985.



12. Comparação dos modelos

Após a separação dos dados, foi criado um modelo simples de previsão que prevê sempre que a temperatura da próxima amostra será igual a anterior, é apresentado na Figura 20 o treino do modelo e posteriormente, na Figura 21, o gráfico de teste do modelo.

Figura 20: Valores reais dos dados e valores previstos no treino, pelo modelo

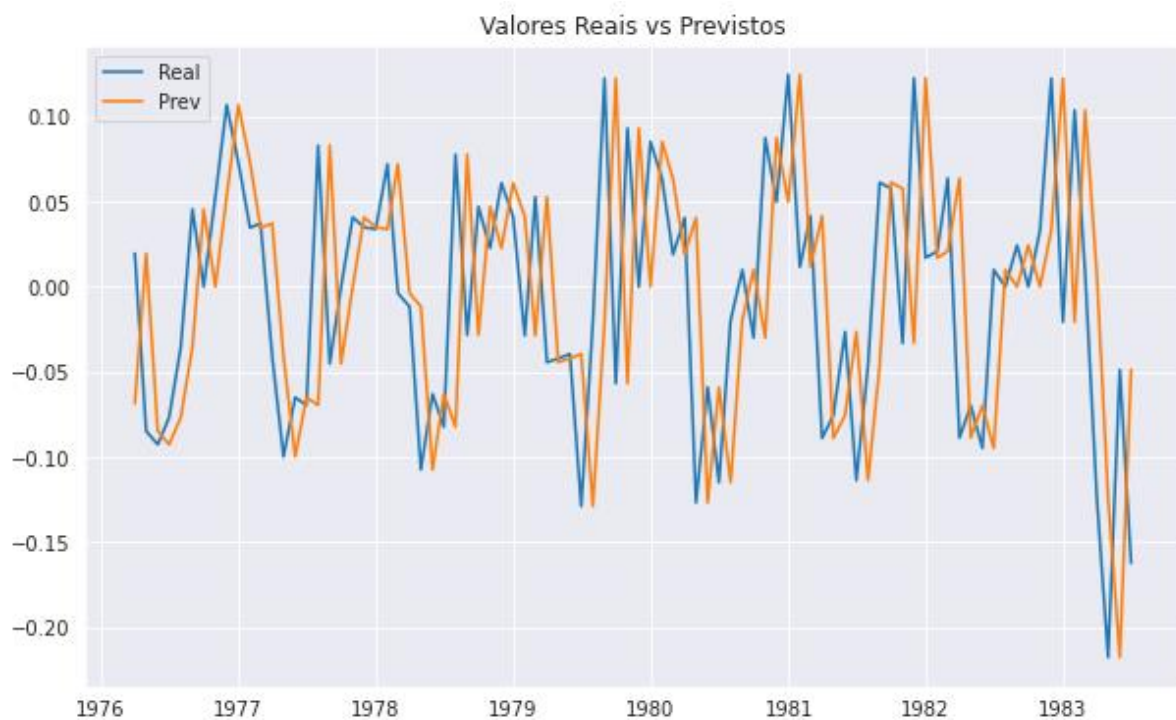
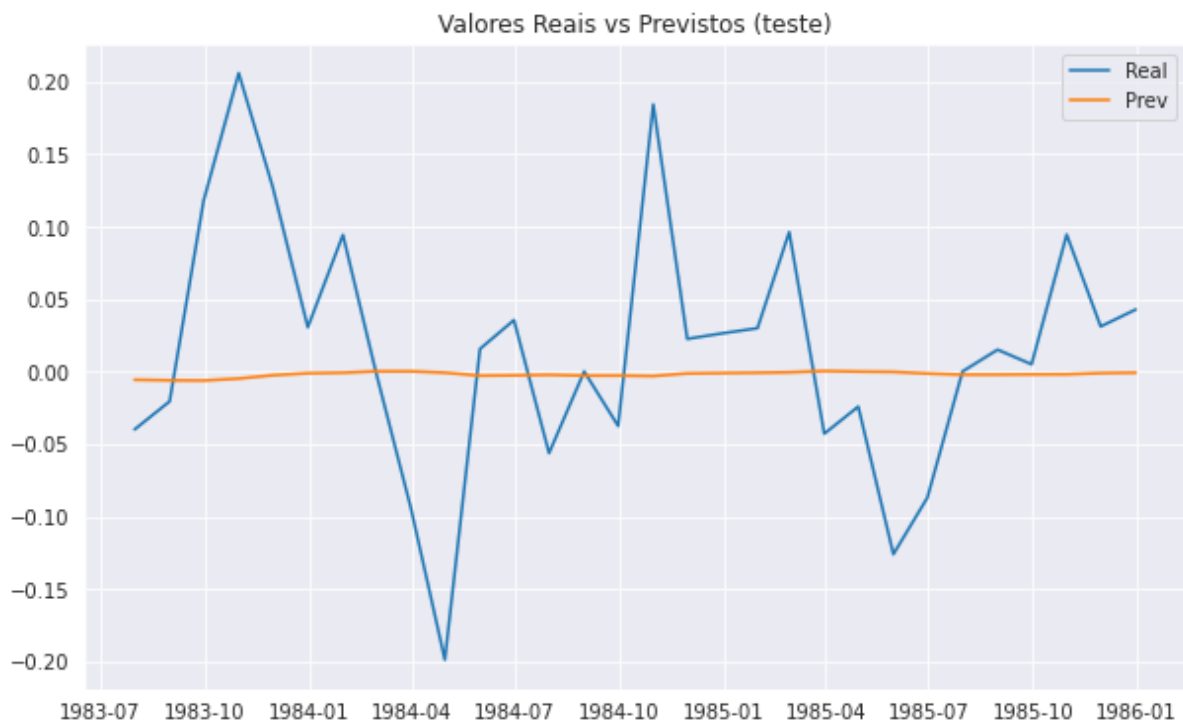
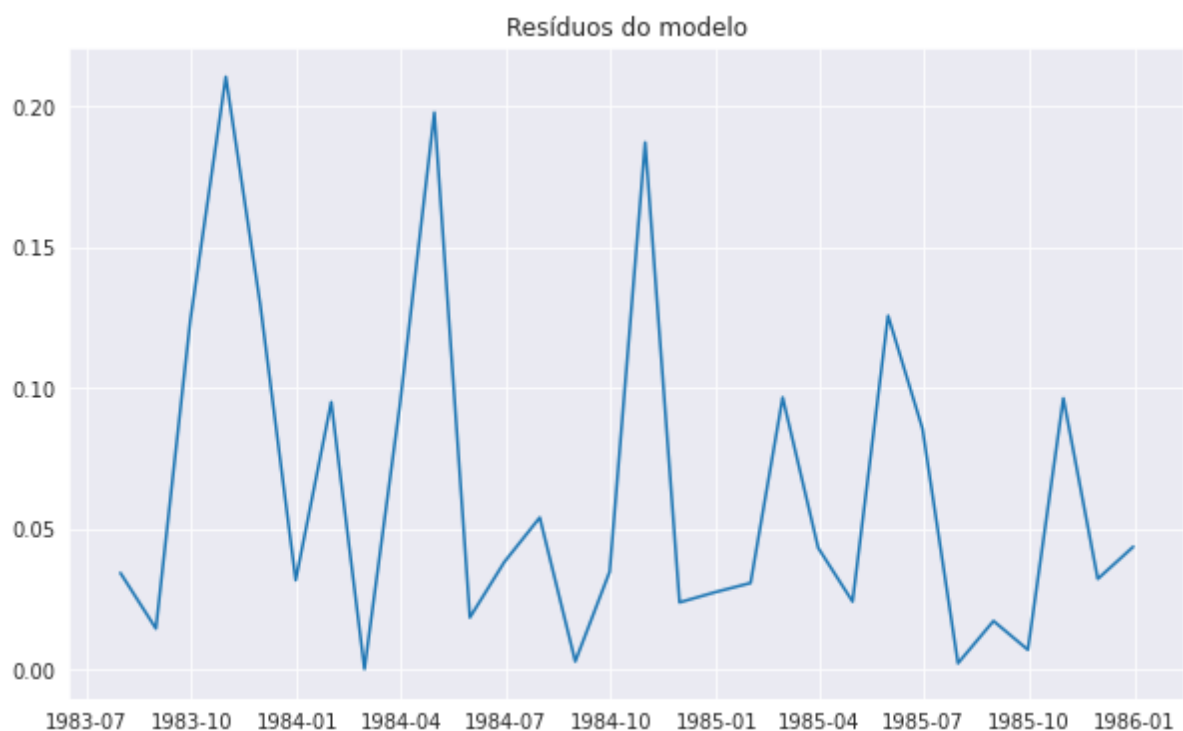


Figura 20: Valores reais dos dados e valores previstos no teste, pelo modelo



Depois de realizar a previsão do modelo, é notável que o modelo não prevê os futuros dados com eficiência. Além disso, é apresentado o gráfico dos resíduos do modelo.

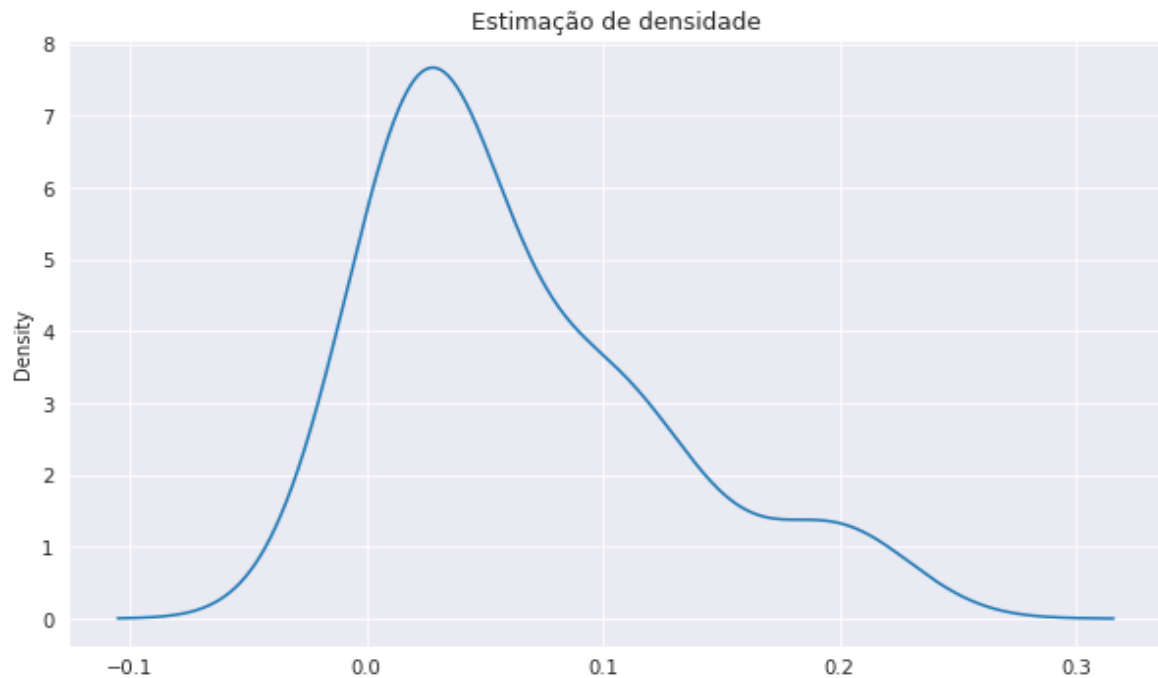
Figura 21: Resíduos do teste do modelo apresentado



Os resíduos mostram valores entre 0 e 0,2 para as amostras, sem alternância sistemática entre os dados, isto é, é um gráfico com dados aleatórios e próximo de zero. Por meio do gráfico

de estimação de densidade abaixo, é possível verificar que os dados, aparentemente, se aproximam de uma distribuição normal mas para obter certeza, iria ser necessário um teste estatístico.

Figura 21: Densidade dos resíduos do teste do modelo apresentado



13. Conclusão

Este trabalho teve o intuito de explicar as principais características uma série temporal, tais quais são a tendência, sazonalidade e componente aleatório, bem como demonstrar de que maneira se dá um modelo aditivo e um modelo multiplicativo dos dados temporais, por meio de simulação.

Além disso, não só foi apresentado o conceito de estacionariedade (e feito uma comparação visual entre um exemplo de um modelo estacionário e um não-estacionário) como também realizou-se uma análise com um banco de dados real entregue pelo professor.

Para explorar tal banco de dados utilizou-se da análise das características supracitadas, separado os dados para a modelagem do mesmo e, por fim, a comparação dos modelos. Para tanto, foi usado um modelo simples que não obteve bons resultados na base de dados de teste e concluído com a apresentação dos resíduos.