**RELATÓRIO TÉCNICO**

**2° Avaliação**

**Disciplina: Séries Temporais**

**Professor: Vinícius**

**Aluno: Júlio Henri**

**Matrícula: 201807840011**

**1. Introdução**

O presente trabalho é a segunda avaliação da disciplina de Séries Temporais do curso de Estatística da Universidade Federal do Pará e tem por objetivo apresentar os conhecimentos desenvolvidos durante o curso relacionados aos modelos de séries temporais e suas características.

**2. Materiais e Métodos**

Os dados do trabalho foram obtidos através de um arquivo no formato .csv no site <https://www.kaggle.com/c/demand-forecasting-kernels-only/data>. Foi utilizado este site devido a facilidade na obtenção dos dados e pela possibilidade de submeter sua previsão, de modo que consiga ver o quão bom seu modelo foi e comparar com o modelo de outras pessoas.

A base de dados foi construída a partir de itens de lojas durante 5 anos. Todos os 913000 registros em questão são de natureza quantitativa, contendo 4 variáveis: “date” que compreende a data que foi obtido o registro, “store” diz o número da loja, “item” que equivale ao número do item desta loja e “sales” que é o número de vendas, dado uma data, loja e item.

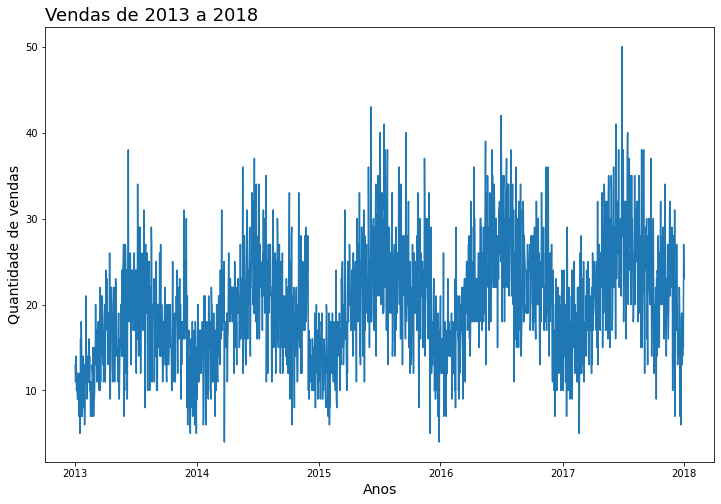
A análise estatística será inicialmente realizada por meio de análise exploratória de dados, objetivando fornecer uma visão geral do estudo, de modo que se possa ter os melhores resultados de previsão de dados o possível posteriormente.

Para a visualização de gráficos, será mostrado gráficos de linhas, que serão utilizados para visualizar os dados no tempo, serão apresentados também boxplot, que é apropriado para perceber outliers nos dados, bem como sua simetria e dispersão. Nos gráficos de linha, há a obtenção visual de diversas características de uma série temporal, no qual é comprovada (e explicada) posteriormente com um teste estatístico.

**3. Sazonalidade da série temporal e separação em dados de treino e teste**

*Disclaimer: não foi possível plotar o gráfico de todos os dados devido a grande quantidade*

*de registros, desse modo, foi selecionado para modelar somente o item 1 da loja 1.*

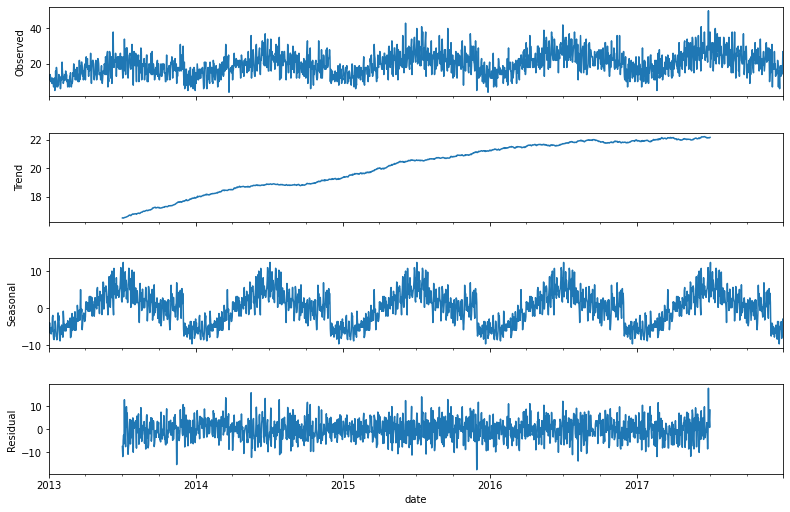
 **Figura 01:** Quantidade de vendas do item 1 na loja 1, nos anos de 2013 a 2018.

Visualmente a série temporal aparenta ter um padrão nos dados que se repete com o tempo e isso é chamado de sazonalidade, com elevação na quantidade de vendas nos meses do meio de cada ano e para haver uma leve tendência nos registros.

Os dados foram separados de maneira a dividir os 1074 primeiros registros (de 301/01/2013 até 31/08/2017) para treino do modelo e os 92 restantes para formar o grupo de teste (de 01/10/12017 até 31/12/2017).

**3. Características dos dados**

Para melhor compreensão dos dados se faz necessário a decomposição da série, de forma que se pode observar suas características principais e por meio delas obter informações importantes para um futuro modelo de previsão. Dito isto, abaixo é apresentando vários gráficos de linhas nos dados de treino:

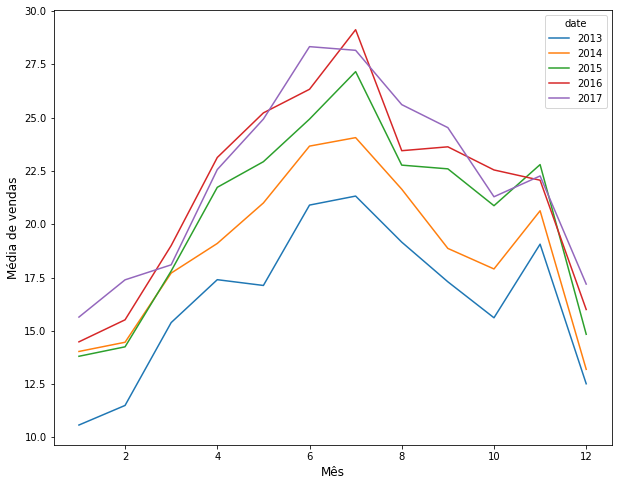
**Figura 02:** Dados de venda de 2013 a 2017

No gráfico de linhas mais acima é mostrado os dados reais, logo abaixo a tendência desses dados, que 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 e nesse caso, a tendência dos dados é positiva, o que indica o crescimento dos dados no tempo.

Mais abaixo é possível observar a sazonalidade dos dados, que novamente aparenta crescer durante o meio de cada ano e abaixar nos meses de início e fim, isto é, o período sazonal é anual. Por fim é exposto os resíduos ou variações irregulares e é uma variável na qual não se pode prever, são variações inexplicadas dos dados, que pode ser de mudança de governante da cidade, nova loja vendendo outros produtos próximos, queda do dólar e/ou diversas outras situações.

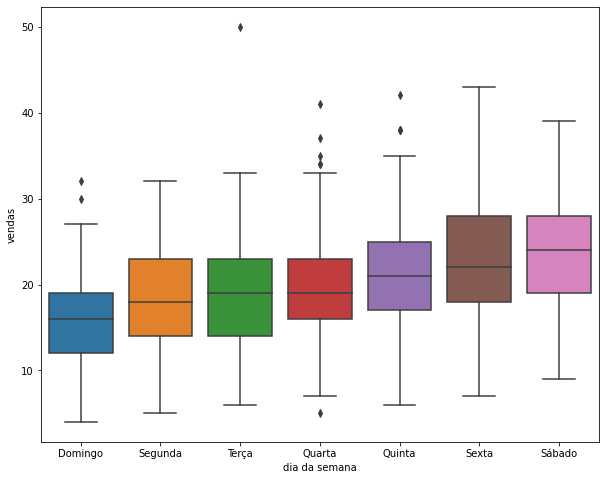
A partir da figura 03 abaixo mostra a média de vendas de todos os anos, por mês, e sendo assim, destaca-se a sazonalidade, o crescimento das vendas desde o início do ano até o mês 7, após isso o decrescimento e ainda há uma pequena subida em novembro antes de cair novamente.

**Figura 03**: Média de vendas nos anos de 2013 a 2017, por mês.



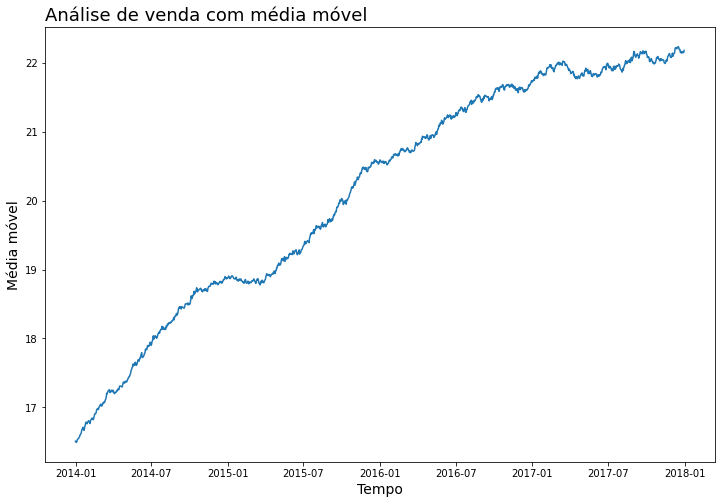
Ainda explorando os dados, é apresentado a figura 04 e figura 05, para melhor entendimento dos dados de acordo com o tempo.

**Figura 04:** Média de vendas de todos os anos de acordo com os dias da semana.

****

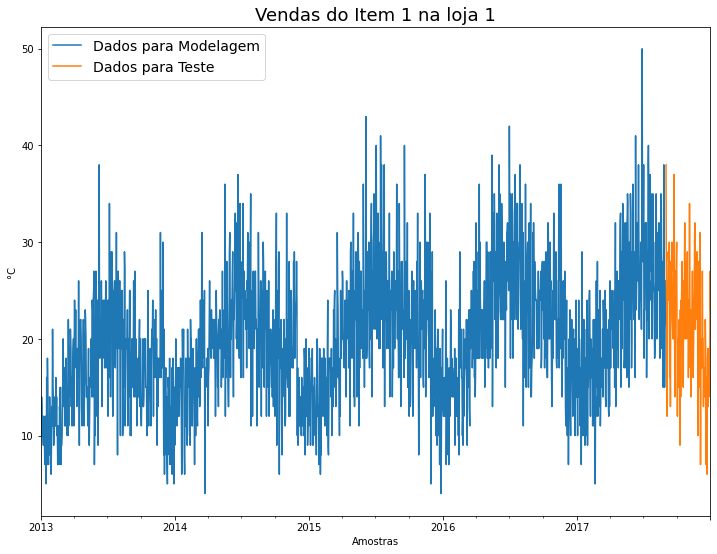
Visualizando os dados acima é possível concluir uma crescente nos dados conforme o passar dos dias da semana, com o a mediana do pico de produtos vendidos no sábado, além da observação de outliers que não são um bom sinal para alguns tipos de modelo de séries temporais.

Enquanto na Figura 05 destaca-se a tendência por meio da utilização da análise de dados com médias móveis com um valor alto, confirmando a tendência crescente dos dados quando diminuído seus resíduos e sazonalidade.

**Figura 05:** Média de vendas de todos os anos de acordo com os dias da semana.

Logo abaixo, a Figura 06 mostra a separação de dados em modelagem e teste.

**Figura 06:** Separação dos dados em modelagem e teste.

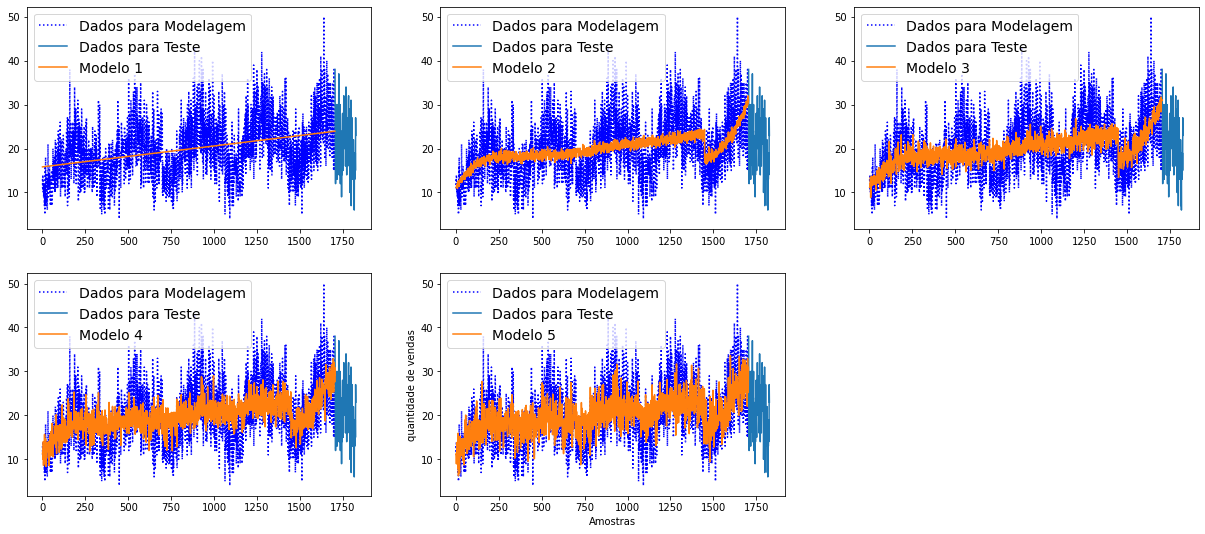


**4. Modelagem e comparação de modelos**

Para a modelagem dos dados da série temporal foi utilizado o método dos mínimos quadrados ordinários que se traduz na soma da diferença das amostras já obtidas com as amostras estimadas de acordo com a equação de regressão com os parâmetros ao quadrado, de uma maneira que tais parâmetros obtidos reduzam essa diferença.

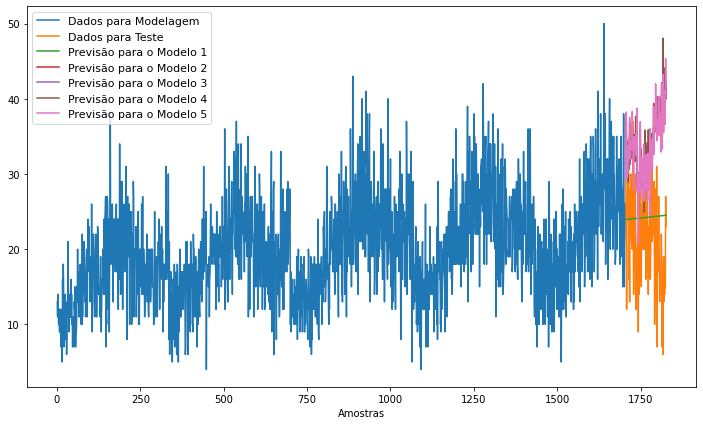
Dito isto, foi criado um gráfico com 5 modelos de regressão, sendo alterado somente o grau do polinômio de cada modelo. Sendo assim, o primeiro modelo é de ordem 1, o segundo de ordem 30, o terceiro de ordem 90, o quarto de ordem 180 e por último de ordem 300. O resultado visual dos modelos polinomiais é apresentado a seguir.

**Figura 07:** Comparação dos modelos polinomiais

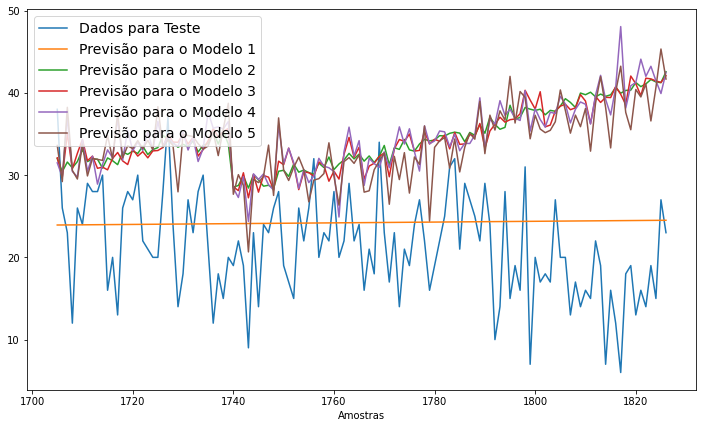


Após modelar a série com os dados de treino, é realizado o teste dessa sério, isto é, o modelo previu os dados separados para teste de acordo com os parâmetros já entregues a ele.

**Figura 08:** Previsões de teste das vendas do item 1 na loja 1, nos anos de 2013 a 2018

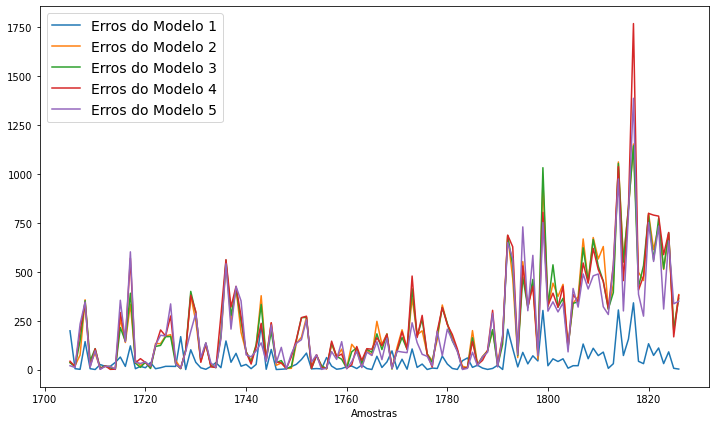


**Figura 09:** Previsões de teste das vendas do item 1 na loja 1, nos anos de 2013 a 2017.



As figuras 08 e 09 tratam do mesmo assunto, enquanto a figura mais acima apresenta os dados de uma maneira mais geral, desde os dados que foram usados para treino, teste e os modelos, no abaixo destaca-se as previsões e dados de teste para comparação. A partir da análise desse gráfico é notório dizer que mesmo o Modelo 1 sendo o mais simples, os outros modelos preveem os dados mais erroneamente.

**Figura 10:** Erros das previsões de teste das vendas do item 1 na loja 1, nos anos de 2013 a 2017.

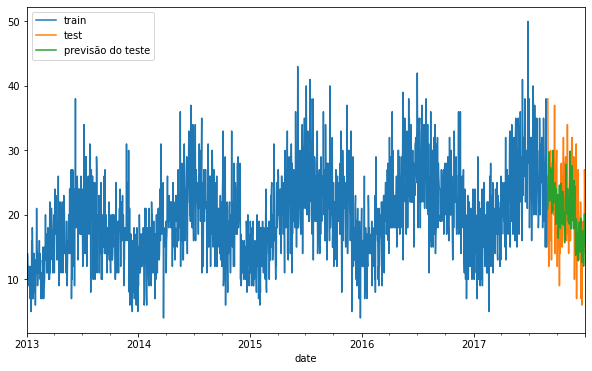


Comprovando a suposição de antes, aqui é a mostrado que os erros do Modelo 1 são menores do que outros, mesmo sendo o modelo mais simples.

**5. Predição dos dados através de método de suavização**

O método utilizado para a previsão através de suavização foi o Holt-Winters, pois é o método indicado para séries em que há tendência e sazonalidade. Então, com auxílio do programa Python foi possível otimizar o modelo para que possa ser escolhido os melhores valores para constantes de suavização. Foi usado 0.42 para constante de sazonalidade e 0.526 para constantes de tendência de nível de suavização.

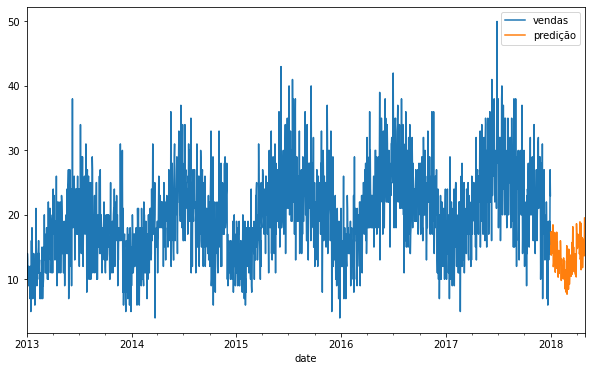
**Figura 11:** Previsão de teste para o modelo de suavização das vendas do item 1 na loja 1, nos anos de 2013 a 2017.



O gráfico acima destaca a previsão do teste em relação aos dados de testes e apresenta pontos aparentemente aproximados. Então é realizada a validação do modelo por meio do erro médio quadrático para compreender a diferença entre os dados reais e a previsão. O resultado foi 6.341 e quanto menor esse valor, mais eficaz é o modelo, logo, o resultado apresentado é satisfatório para o estudo.

Finalmente, é executado a previsão para dados de 4 meses futuros, o que resulta no gráfico abaixo.

**Figura 11:** Previsão de suavização das vendas do item 1 na loja 1 para os 4 primeiros meses de 2018.



**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 se realizou 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.