

Nomes: Guilherme Pretto, Samuel Matias Finkler e Thiago Sordi

Disciplina: Mineração de dados (ELC1098)

## Trabalho prático 2

### Limpeza e tratamento dos dados

Inicialmente, foi identificado que o arquivo com o nome “2020ClosePrice\_SAX240\_60\_10.csv” possuía 46 colunas, onde a primeira era do índice, a qual foi removida, e o restante seria de cada uma das ações, restando 45. Como temos somente 44 ações na lista, seria necessário remover uma das colunas. Analisando mais precisamente cada uma delas, identificamos que, com exceção da primeira, o restante tinha classes que correspondiam com os valores mostrados no arquivo “2020ClosePrice\_work.csv”. Além disso, a primeira coluna continha classes em ordem alfabética de “a” até “j” e um padrão onde a primeira classe possuía exatamente o mesmo número de linhas da última, a segunda, o mesmo número da penúltima e assim sucessivamente. Considerando essas características, optamos por excluir também essa primeira coluna, restando as 44, uma para cada ação.

Ainda com o dataset “2020ClosePrice\_SAX240\_60\_10.csv”, separamos entre acima da média para as classes de “f” até “j” e abaixo da média para as classes de “a” até “e”. Dessa forma, as classes acima da média receberam o valor 0 e as classes abaixo da média receberam o valor 1. Com esses valores, as regras foram geradas a partir das relações que estão abaixo da média, visto que o suporte mínimo é de 75%. Com esse suporte, não são geradas regras de ações acima da média.

### Tarefa mínima

Utilizando o APRIORI com suporte de 75% e confiança de 80%, podemos identificar alguns padrões interessantes:

	antecedents	consequents	antecedent support	consequent support	support	confidence	lift	leverage	conviction
0	(BRML3.SA)	(EMBR3.SA)	0.800000	0.783333	0.783333	0.979167	1.25000	0.156667	10.400000
1	(EMBR3.SA)	(BRML3.SA)	0.783333	0.800000	0.783333	1.000000	1.25000	0.156667	inf
2	(BRML3.SA)	(IRBR3.SA)	0.800000	0.816667	0.800000	1.000000	1.22449	0.146667	inf
3	(IRBR3.SA)	(BRML3.SA)	0.816667	0.800000	0.800000	0.979592	1.22449	0.146667	9.800000
4	(EMBR3.SA)	(IRBR3.SA)	0.783333	0.816667	0.783333	1.000000	1.22449	0.143611	inf
5	(IRBR3.SA)	(EMBR3.SA)	0.816667	0.783333	0.783333	0.959184	1.22449	0.143611	5.308333
6	(BRML3.SA, EMBR3.SA)	(IRBR3.SA)	0.783333	0.816667	0.783333	1.000000	1.22449	0.143611	inf
7	(BRML3.SA, IRBR3.SA)	(EMBR3.SA)	0.800000	0.783333	0.783333	0.979167	1.25000	0.156667	10.400000
8	(EMBR3.SA, IRBR3.SA)	(BRML3.SA)	0.783333	0.800000	0.783333	1.000000	1.25000	0.156667	inf
9	(BRML3.SA)	(EMBR3.SA, IRBR3.SA)	0.800000	0.783333	0.783333	0.979167	1.25000	0.156667	10.400000
10	(EMBR3.SA)	(BRML3.SA, IRBR3.SA)	0.783333	0.800000	0.783333	1.000000	1.25000	0.156667	inf
11	(IRBR3.SA)	(BRML3.SA, EMBR3.SA)	0.816667	0.783333	0.783333	0.959184	1.22449	0.143611	5.308333

**1º regra:** quando a ação “BRML3.SA” está abaixo da média, a ação “EMBR3.SA” também está. (suporte 78% e confiança de 97%)

**2º regra:** quando as ações “EMBR3.SA” e “IRBR3.SA” estão abaixo da média, a ação “BRML3.SA” também está. (suporte 78% e confiança de 100%)

**3° regra:** 2° regra: quando a ação “IRBR3.SA” está abaixo da média, as ações “BRML3.SA” e “EMBR3.SA” e também estão. (suporte 78% e confiança de 95%)

### **Ação “PETR4.SA”**

Para esta, invertemos os valores usados no APRIORI, utilizando 1 para as ações situadas acima da média e 0 para as ações abaixo da média, pois a ideia é obter o suporte mínimo onde a ação “PETR4.SA”, estando acima da média, influencie no aumento de outra ação. Dessa forma, com suporte de 50%, a ação “PETR4.SA”, influencia na alta da ação “MRVE3.SA”.

### **Outras análises**

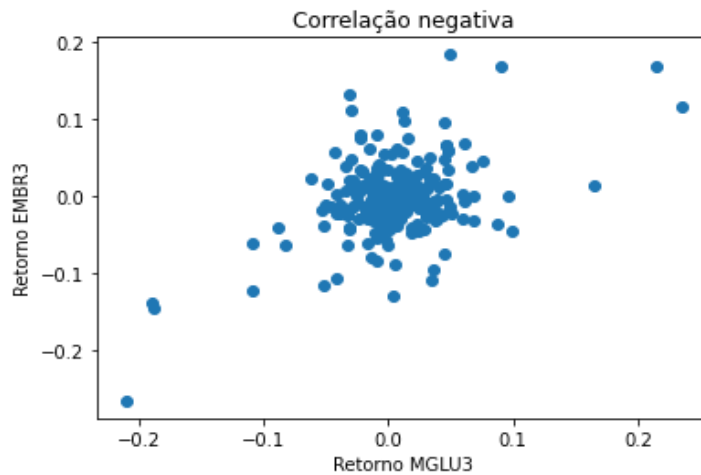
Para realizarmos as outras análises do trabalho, utilizamos apenas o dataset “2020ClosePrice\_work.csv”.

### **Correlação entre ações**

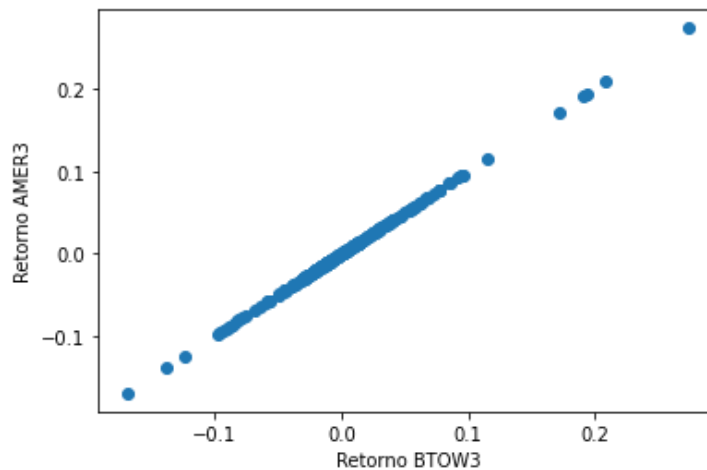
Com o dataframe contendo cada ação como um atributo bastou gerar uma matriz de correlação (nativa do Pandas) e então aplicado um heatmap para melhor visualização.

Nessa matriz de correlação o coeficiente de correlação varia entre -1 e 1. Quando próximo de 1 significa que há uma correlação positiva forte, ou seja, quando uma ação BTOW3.SA subir ela tende a subir junto com uma ação AMER3.SA que possui uma correlação extremamente forte com ela (ou também PETR4.SA com PETR3.SA). Quando o coeficiente é próximo de -1 significa que é uma correlação negativa forte, ou seja, quando o valor de uma ação sobe, o valor de outra desce (como MGLU3.SA com EMBR3.SA) ou vice-versa. Enquanto o coeficiente próximo de 0 significa que não há nenhuma correlação linear (CCRO3.SA com AMER3.SA é algo próximo disso).

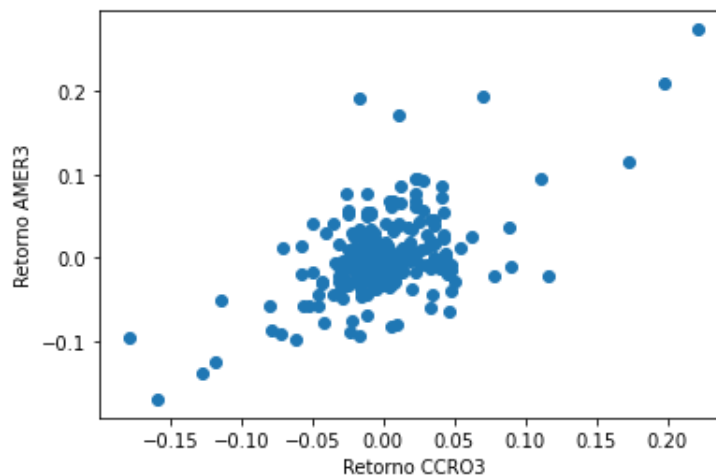
Para visualizar alguns casos mais específicos, foi utilizada uma função (nativa do Pandas) `pct_change()`, que calcula a alteração percentual da linha imediatamente anterior por padrão. Isso é útil para comparar a porcentagem de mudança em uma série temporal, que é o nosso caso.



Aqui temos uma correlação negativa da nossa matriz, não é extremamente forte, na verdade, tende mais ao neutro, mas conseguimos mesmo assim notar a tendência quando uma aumenta seu percentual de retorno, a outra tende a ficar com valor negativo.



Enquanto nesse caso temos uma correlação positiva extremamente forte, ou seja, as duas aumentam e diminuem juntas.



Por fim, temos um caso neutro, não totalmente neutro, pois seu coeficiente de correlação é um pouco maior do que 0. Mas notamos que não tem necessariamente um padrão entre nossos valores, na sua grande maioria.

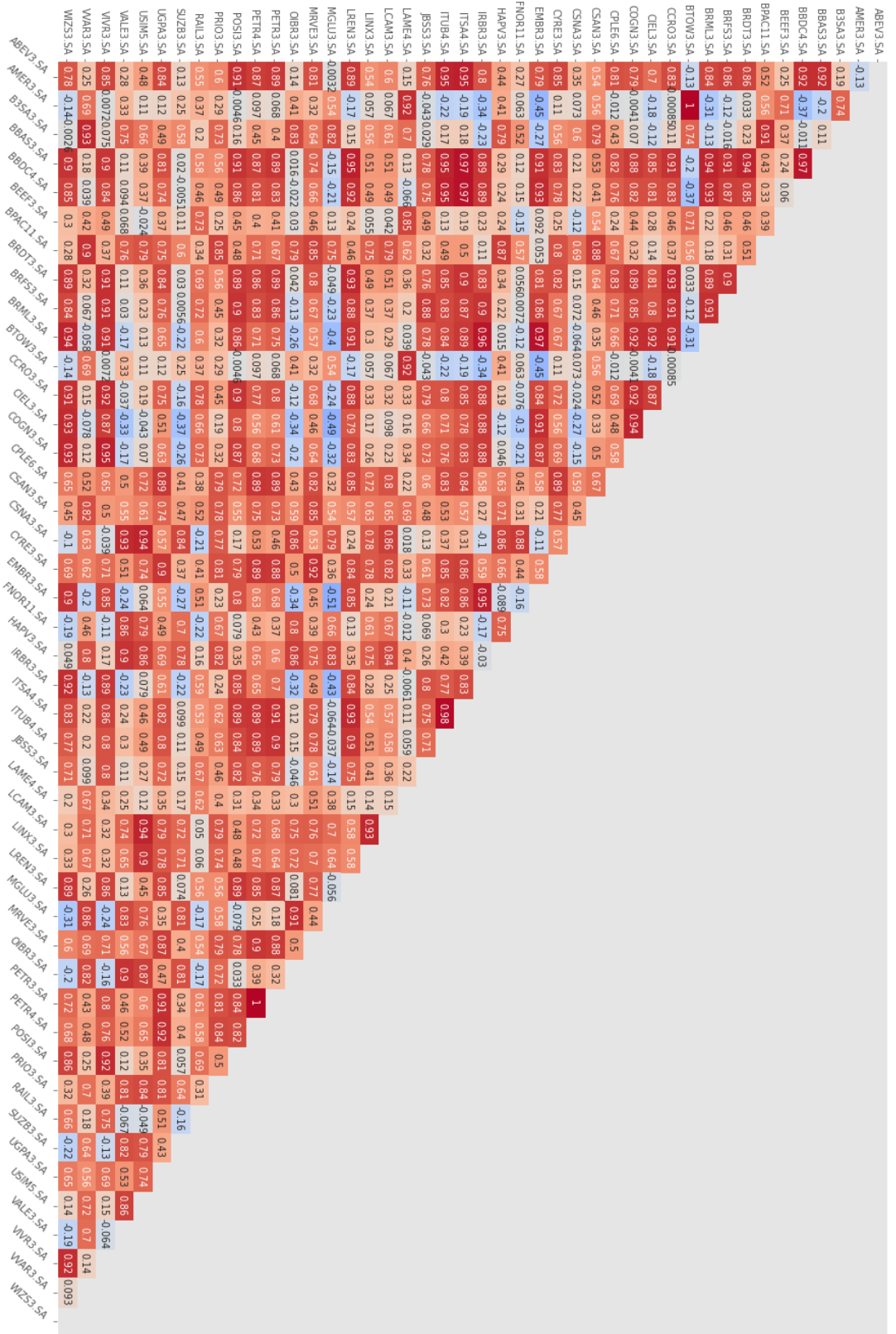
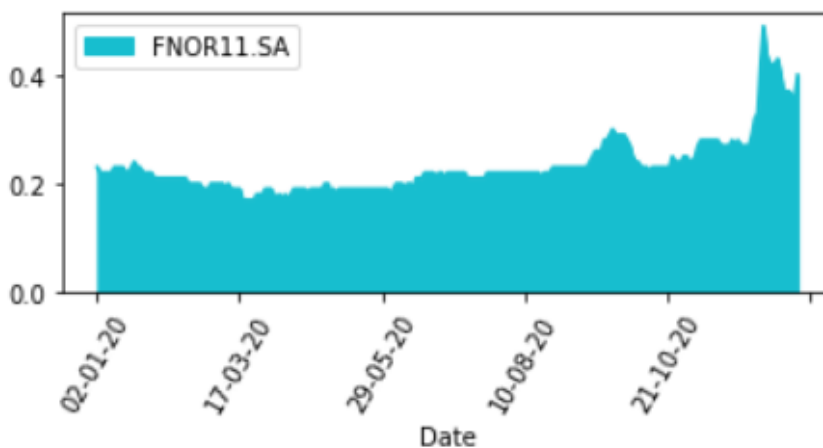
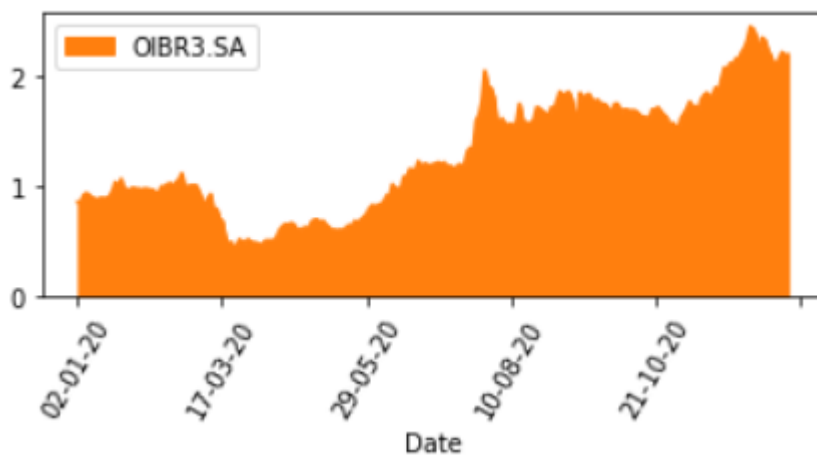
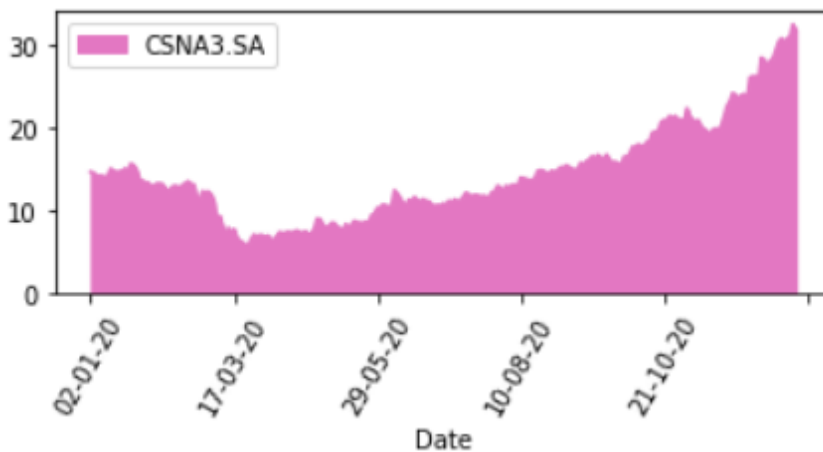


Figura do Heatmap

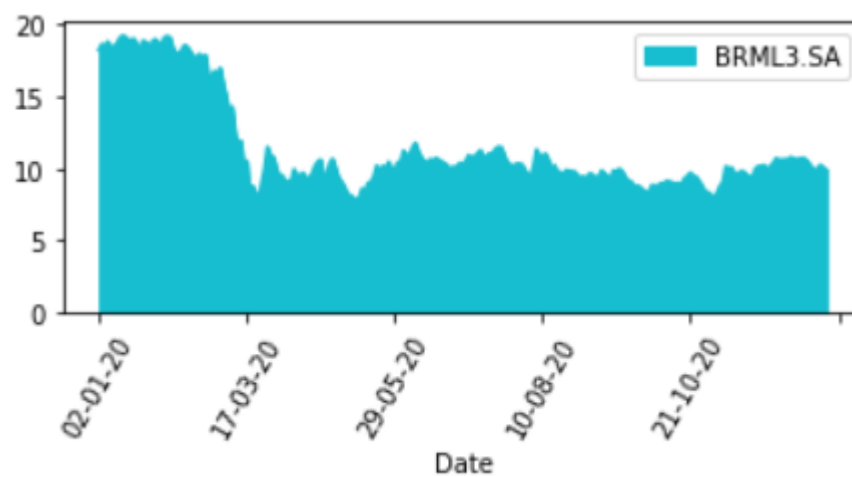
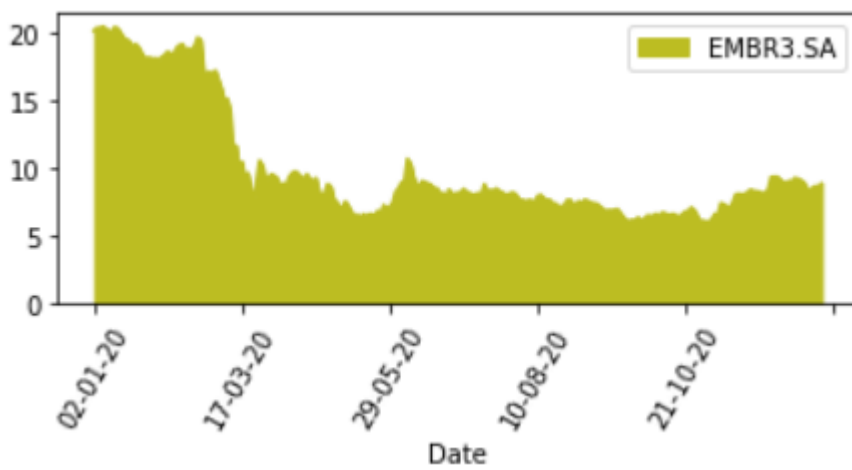
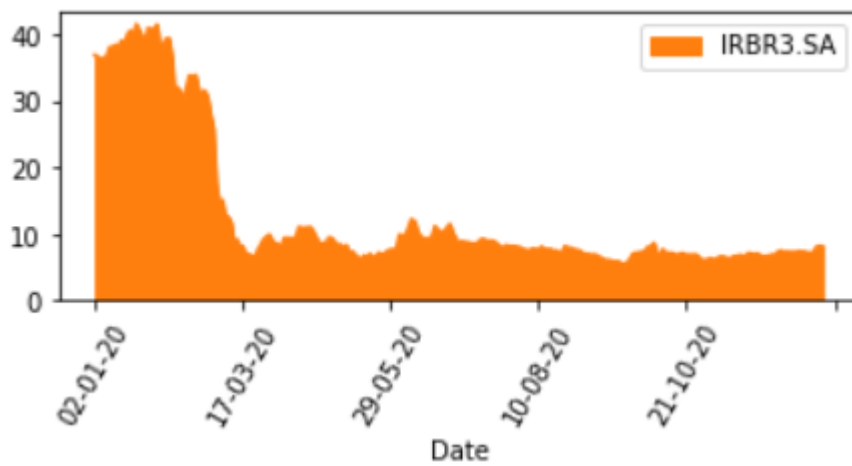
## Análise gráfica

Observar o desempenho positivo da CSNA3.SA, OIBR3.SA, FNOR11.SA, e o desempenho negativo da IRBR3.SA, EMBR3.SA, BRML3.SA (ambas estabilizaram no valor da queda). Quase todas ações tiveram uma queda no início da pandemia no Brasil, algumas voltaram ao seu valor normal, outras acima, e algumas se mantiveram no valor de queda ou menos.

## Desempenho positivo



## Desempenho negativo



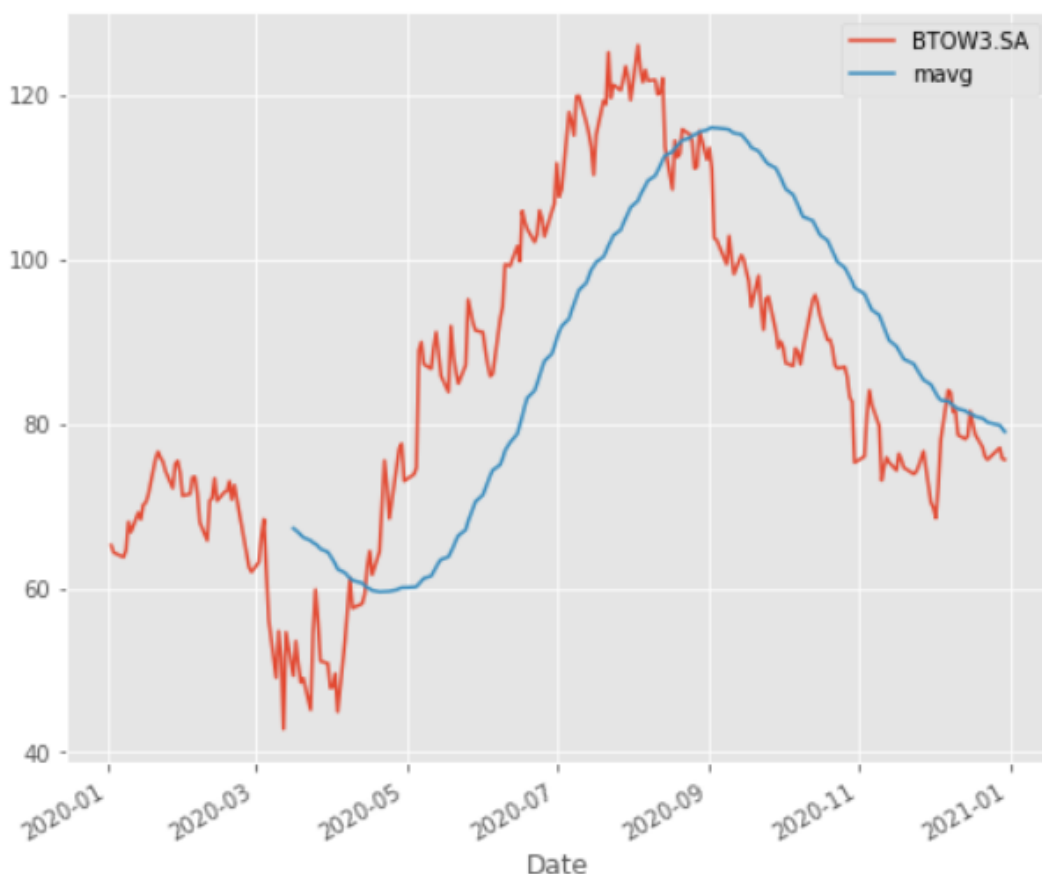
## Performance de cada ação

### Média móvel - determinar a tendência

A média móvel suaviza o preço criando um preço médio constantemente atualizado. É interessante para remover alguns "ruídos" do gráfico de preços. Além de que, essa média móvel pode atuar como uma "resistência", ou seja, das tendências de alta e baixa pode se esperar que seguirá a tendência e não desvie do ponto de resistência.

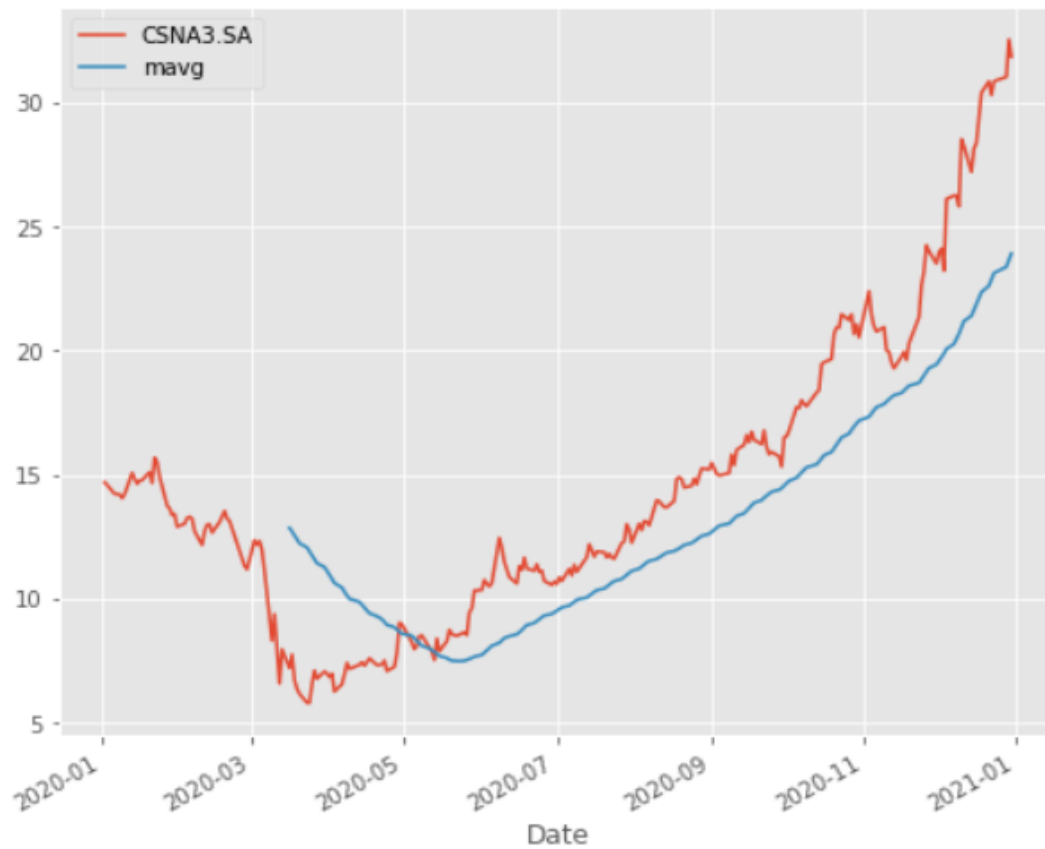
Será calculada a média móvel dos últimos 50 dias (50 janelas) de preços de ações, e obterá a média para cada média móvel da janela.

Foi escolhida a ação BTOW3.SA por causa do seu valor acima das demais, e alta variação dentro do período indo dos 65 reais até 126 (pico), e depois voltando aos 75.

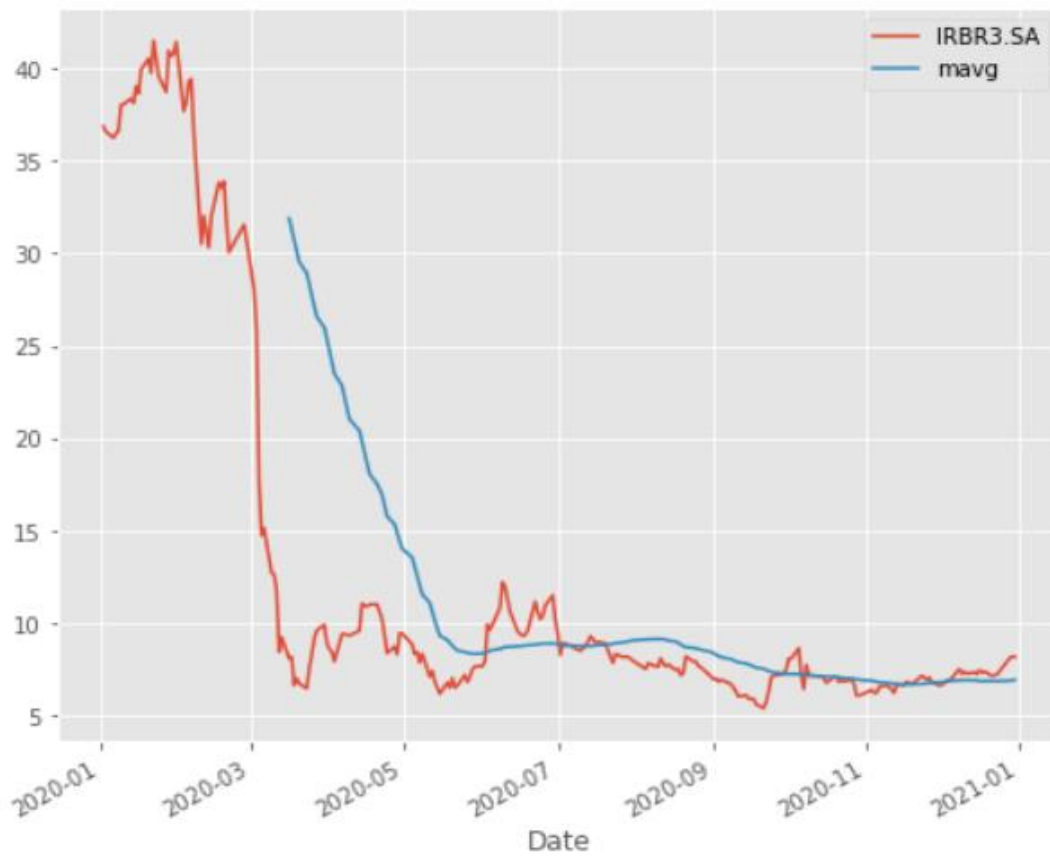


Este gráfico, da Média Móvel, nos mostra a tendência crescente de alta ou baixa do preço das ações. Dessa forma, podemos ter noção dos momentos de operação, como comprar quando as ações estiverem em baixa e vender quando as ações estiverem em alta. No caso **BTOW3.SA** está tendendo a baixa.

Outro exemplo, tendendo a alta:

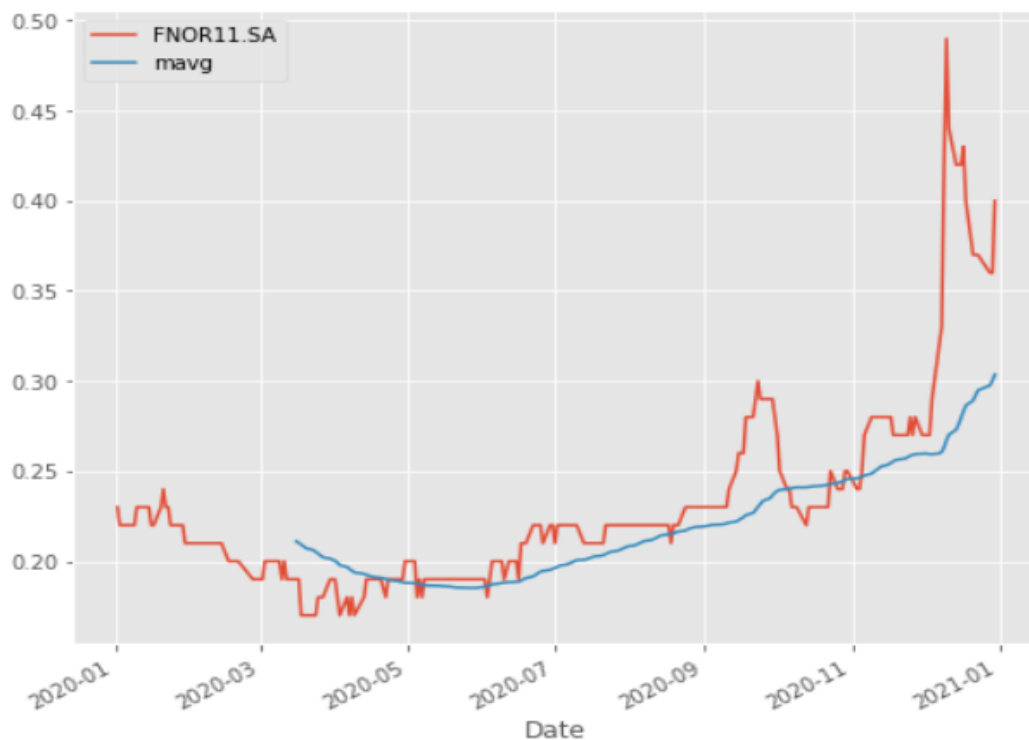


O próximo apresenta uma leve queda, mas parece manter uma certa constância.





Outro exemplo, desta vez, assim como na primeira, tendendo a alta

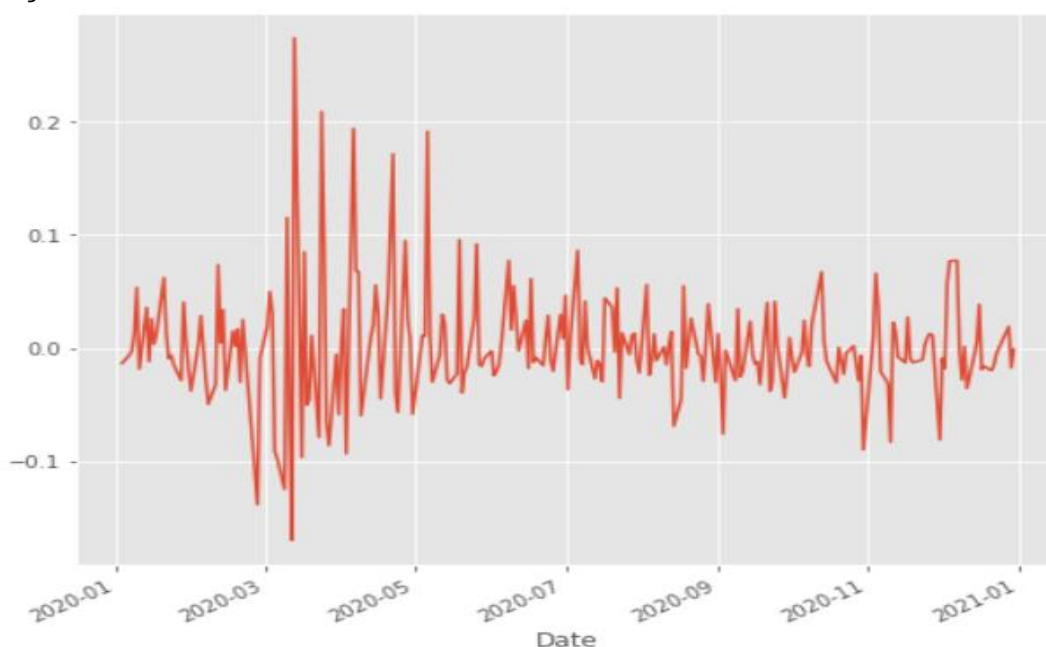


**Desvio de retorno** — para determinar o risco e o retorno

Retorno Esperado usa a média, ou valor esperado, da distribuição de probabilidade dos retornos do investimento. O retorno esperado de uma carteira é calculado multiplicando o peso de cada ativo pelo seu retorno esperado e somando os valores de cada investimento —

<https://www.investopedia.com/terms/e/expectedreturn.asp>.

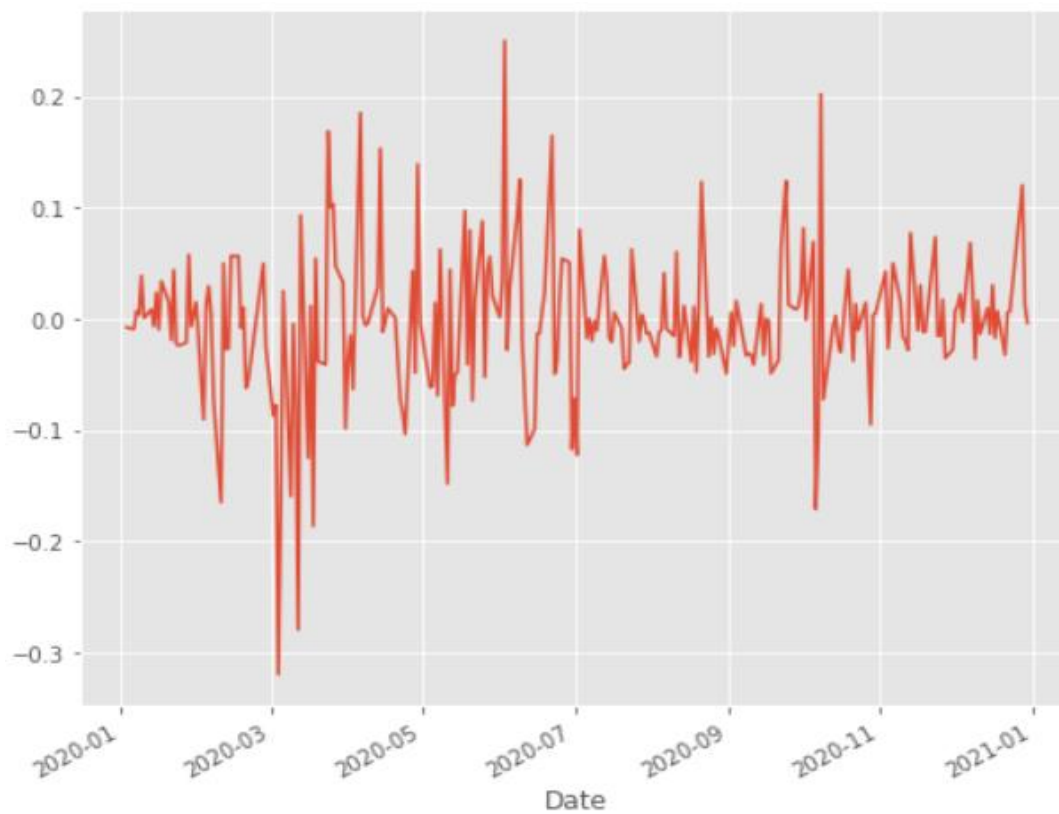
**Ação BTOW3.SA:**



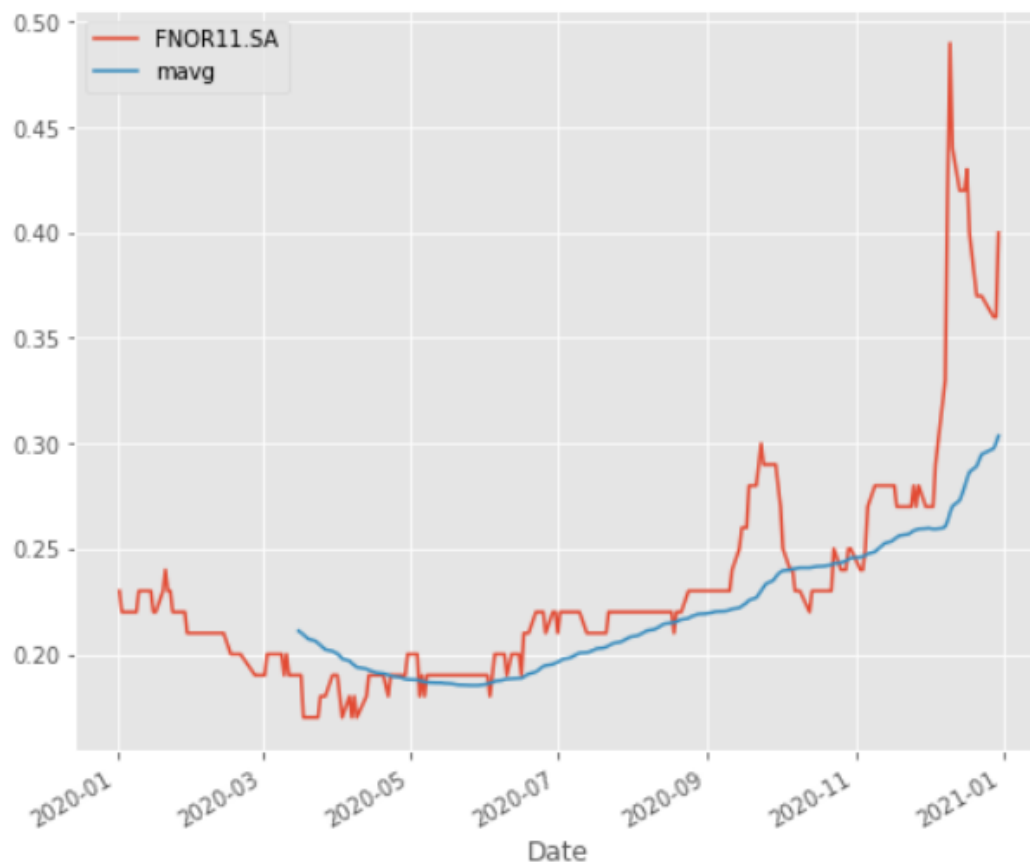
**Ação CSNA3.SA:**



**Ação IRBR3.SA:**



### Ação FNOR11.SA:



Obviamente, as ações ideais devem ter um retorno alto e estável. Nesses gráficos conseguimos notar que ações como da **BTOW3.SA**, **IRBR3.SA** e **CSNA3.SA** tiveram quedas de mais de 10% algumas vezes ao longo do ano, claro que tirando a **IRBR3.SA**, as outras duas tiveram momentos de alta que compensaram as quedas. Mas temos ações como o caso da **FNOR11.SA**, onde teve apenas dois momentos que ultrapassou minimamente a linha dos 10% de queda, mas sua valorização foi muito boa.