

Mercado Siderúrgico Brasileiro 2022

2 DE SETEMBRO

Análise exploratória dos dados
Por Kauê Mandarino



Sumário

A questão de negócio.

PASSO 01

Meu cliente, uma start-up, do setor de delivery de alimentação no Brasil, teve um lucro de R\$ 1mi. em 2021, e quer aplicar uma percentagem desse lucro (30% a 60%) na bolsa de valores, mais especificamente no setor siderúrgico do Brasileiro, pois, embora sem conhecer o setor a fundo, tem ciência que este, é um setor no qual o Brasil se destacou em 2021. Nossa análise vem, com insights, baseados em dados, amparar a empresa nesta tomada ou não de decisão. O motivo de o cliente escolher uma empresa brasileira para o aporte financeiro, além do lucro, é ético, visando devolver valor a sociedade brasileira, que é o ambiente em que sua empresa está inserida, fomentando assim, a economia local, como uma maneira de devolver o que lhe foi dado.

O cliente quer que eu apresente o cenário de mercado e projeções previstas para o curto e médio prazo (2 e 5 anos).

Nosso cliente quer investir, mas antes quer saber se isto é rentável, ou não, segundo os nossos insights amparados nos dados.

Entendimento do negócio

PASSO 02

No primeiro ciclo, fiz um estudo, breve e introdutório, para me contextualizar no mercado, e confirmar com dados as seguintes hipóteses:

H1: O mercado siderúrgico brasileiro é grande e rentável.

H2: A crise financeira, devido a pandemia de COVID-19 acabou.

H3: O setor esta aquecido.

No segundo ciclo, fiz um [estudo](#) bem mais imersivo e abrangente, que me contextualizou no cenário, levantou informações valiosas, entre as quais destaco: o surgimento da indústria; sua transformação na década de 40; as privatizações na década de 90; o que é o aço e qual é a sua cadeia produtiva; o desgaste ambiental devido a queima de carvão, e nos operários do setor, devidos as altas temperaturas envolvidas no processo; cenário de exportações; as taxações do governo Trump; o fim das taxas em julho deste ano(2022), a competitividade tecnológica do setor, que no passado nos fez perder para o Reino Unido; a importância da reciclagem no mercado; as maiores empresas do setor; e a nova tendência de mercado, o aço verde, com zero emissão de gases do efeito estufa. Nesta etapa já surgiram insights, que explorarei mais abaixo.

Coleta e limpeza de dados

PASSO 03 E 04

Inicialmente, li sobre o assunto, e em seguida, levantei uma série estatística, “[Performance-Mensal 2022.05.xls](#)”, junto ao instituto do aço e ao ministério da economia, que continha, entre outros, dados semiestruturados, referentes à produção, importação e exportação brasileira, de minério de ferro e seus subprodutos antes e após a pandemia, podendo assim, ter um panorama e contextualizado de mercado e validar as hipóteses. Utilizei o Excel para o processo de ETL: limpar, extrair e estruturar os dados, que posteriormente foram carregados para análise no Google Data Studio. Nesta etapa, durante o processo de ETL, eu desenvolvi um [extrator de dados](#): uma caixa de seleção, em que você escolhe o dado desejado em uma lista, e o algoritmo busca os valores nas tabelas semiestruturadas, para retornar o valor estruturado automaticamente, já pronto para exportação em .csv. Em seguida (segundo cliço de entrega), já ciente que:

Utilizei o Python (Pandas, Matplotlib.pyplot e Pandas_datareader), para criar um [algoritmo que minerou, limpou e validou](#) uma série estatística de dados, referente aos valores das ações na BOVESPA, das empresas supracitadas. A fonte foi o [Yahoo Finance](#). Defini nesta etapa, o intervalo de pré e pós COVID (01-2016 até 07-2022).

Explicando o funcionamento do extrator de dados ([clique aqui](#) para uma melhor visualização).



Kauê Mandarinino
DATA SCIENTIST

Extrator de dados

Entendendo a lógica por trás da tabela

SISTEMA AUTOMATIZADO DE EXTRAÇÃO DE DADOS SEMI ESTRUTURADOS

Escolha o indicador:

Mês: 2016

Produção / Production
Aço Bruto / Crude Steel
1. Produção Laminados / Rolled Products (*)
Planos / Flat Products
Longos / Long Products
Semiacabados P/ Vendas / Semifinished Products for Sale (*)
Placas / Slabs
Blocos e Tarugos / Ingots, Blooms and Billets

Escolha o indicador

Escolha um indicador na lista para fazer o ETL, escolhi aqui o indicador "Aço Bruto"

Localização automática dos dados.

Através da lógica, mapeei as colunas semiestruturadas, da tabela com os dados brutos, o sistema vai até elas e busca os dados do indicador selecionado.

Mês	2016	2017	2018	2019	2020	2021	2022
1	2516	2868	2914	3021	2740	3023	2928
2	2453	2612	2766	2747	2722	2867	2713
3	2529	2887	3113	2878	2722	2802	2988
4	2322	2927	2995	2970	1929	3037	2923
5	2612	2952	2724	2826	2256	3125	
6	2577	2673	2970	2883	2148	3069	
7	2841	2852	3086	2505	2626	3091	
8	2961	2981	2914	2537	2759	3203	
9	2788	2986	3081	2394	2645	3108	
10	2966	3095	3224	2690	2854	2968	
11	2673	3075	2911	2657	3018	3134	
12	2404	2868	2709	2462	2944	2644	

Limpeza e estruturação

Automaticamente o sistema retornará os dados limpos e estruturados, na aba "CSV". Ele ainda vai pré validar os dados, destacando em vermelho os valores destoantes da amostra.

Exportando

Agora é só salvar como CSV.

Salvar como

- Pasta de Trabalho do Excel (*.xlsx)
- Pasta de Trabalho Habilitada para Macro do Excel (*.xlsm)
- Pasta de Trabalho Binária do Excel (*.xlsb)
- Pasta de Trabalho do Excel 97-2003 (*.xls)
- CSV UTF-8 (Delimitado por vírgulas) (*.csv)
- Dados XML (*.xml)
- Página da Web de Arquivo Único (*.mht;*.mhtml)



1- Produção de aço bruto 2016~2022.csv

1	Month, 2016, 2017, 2018, 2019, 2020, 2021, 2022
2	1, 1629, 1799, 1923, 1907, 2043, 2160, 2024
3	2, 1715, 1722, 1861, 1865, 1898, 2078, 1881
4	3, 1771, 1962, 2130, 2060, 1903, 2101, 2144
5	4, 1650, 1906, 2038, 1942, 1278, 2247, 2032
6	5, 1744, 1837, 2026, 2010, 1432, 2437, 2032
7	6, 1790, 1820, 1918, 1801, 1550, 2248, 2032
8	7, 1929, 1873, 1947, 1894, 1657, 2286, 2032
9	8, 1903, 1963, 1906, 1895, 1846, 2332, 2032
10	9, 1827, 1881, 2107, 1969, 1972, 2181, 2032
11	10, 1953, 2126, 2129, 1887, 2177, 2177, 2032
12	11, 1794, 2008, 2054, 1774, 1956, 1946, 2032
13	12, 1416, 1792, 1657, 1482, 2098, 1771, 2032

E pronto!

Os dados estão preparados para as análises!



Ainda nesta etapa de coleta e limpeza de dados, gerei este gráfico de com o Python:

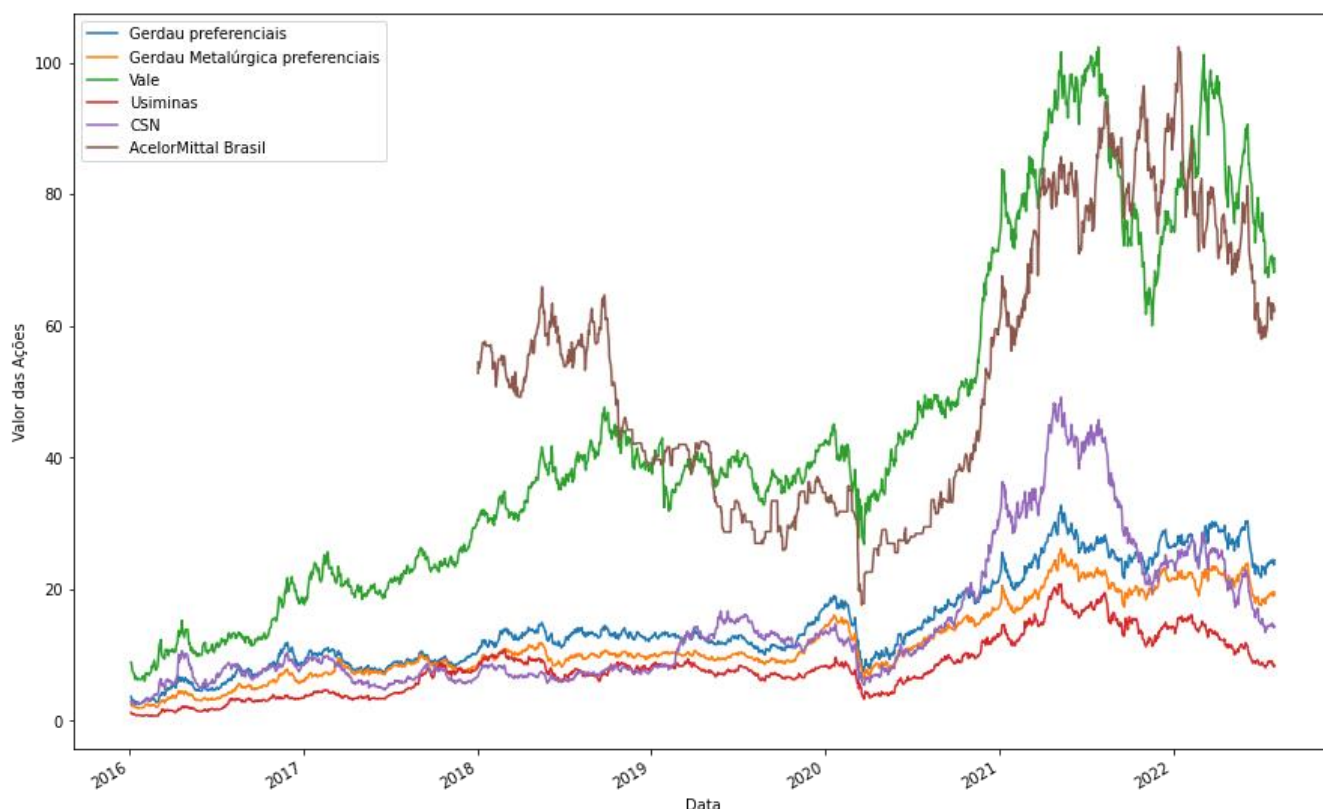


Gráfico com as cotações das ações das maiores empresas siderúrgicas brasileiras de 2016 até o segundo semestre de 2022

Insight 2. Salvo as proporções, as empresas apresentaram, nos mesmos períodos, picos de oscilação equivalentes, ou seja, fatores externos geraram no período, consequências comuns às empresas analisadas, embora os picos sejam comuns, Vale e Gerdau, tiveram um desempenho superior.

Insight 3. A AcelorMittal Brasil, não apresentou crescimento aceitável no valor de suas ações, não acompanhando o desempenho do setor, isto é um indicativo contra o investimento em suas ações, talvez isto tenha ocorrido por eles terem trabalhado seu valor mais em outras carteiras fora da BOVESPA.

Passo 05

ANÁLISE EXPLORATÓRIA DOS DADOS.

Nesta etapa do processo, vou organizar as hipóteses e validá-las com os dados e informações levantadas ao longo do estudo. Vou utilizar a letra **H** para representar as hipóteses e a letra **I** para os insights.

HIPÓTESES:

H0 – As empresas com potencial de investimento segundo a requisição do cliente, são Gerdau e CSN.

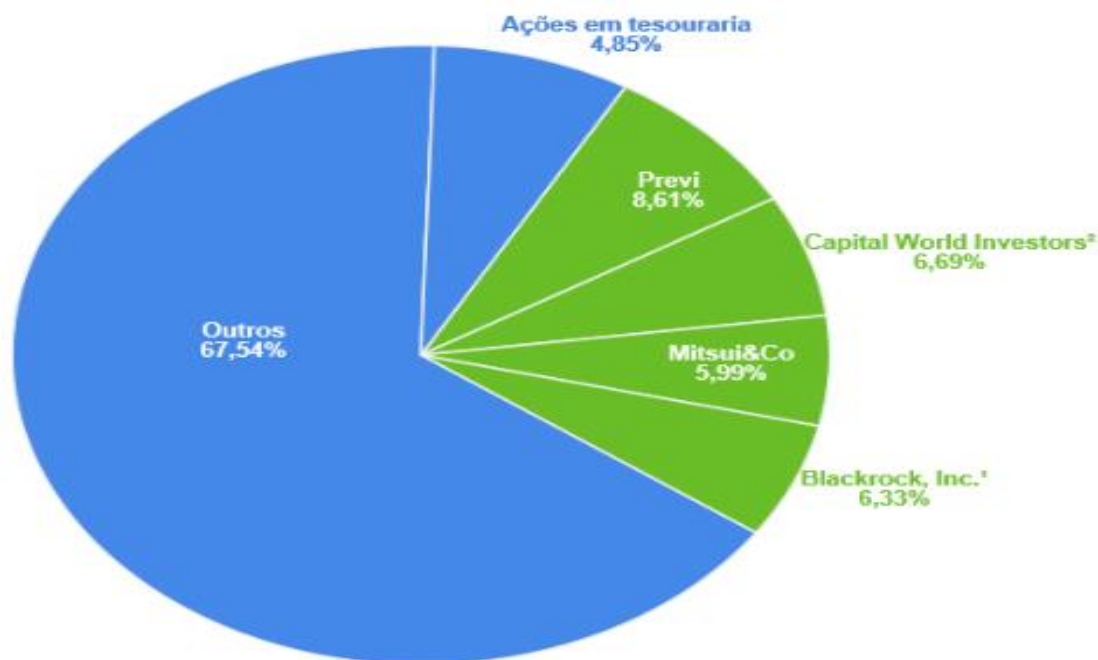
Validação H0: Segundo o levantamento e estudo de dados, as grandes empresas do setor, que teriam potencial para serem escolhidas ao aporte financeiro são: Gerdau, Vale, Usiminas, CSN e ArcelorMittal Brasil. Como visto em nosso estudo de mercado, a partir de 1988 tivemos privatizações que abriram nossas empresas ao setor privado, e hoje em dia, destas supracitadas, a únicas genuinamente brasileiras são Gerdau e CSN. Observemos as informações abaixo, referente aos quadros societários das empresas, para amparar este argumento..

- **Gerdau**

Gerdau é a maior empresa brasileira produtora de aço, e também a empresa que mais recicla aço no mundo. Criada em 1901 no Rio Grande do Sul, a companhia hoje esta presente em mais de 10 países, e conta com mais de 30 mil colaboradores. A Gerdau é de propriedade de Jorge Gerdau, empresário brasileiro, e atual presidente do conselho de administração da empresa.

- **Vale**

Segundo o [site da Vale](#), consultado dia 17-08-2022, temos que: a vale tem uma base de acionista global e diversificada



Ações da Vale em 31 de julho de 2022

Tomamos por exemplo a empresa Black Rock Inc. que detém 6,33% das ações da Vale. Este Grupo está ligado ao grande capital internacional, sendo uma das empresas que controla grande parte da economia mundial, tendo o patrimônio perto dos US\$ 10 tri. Um valor mais que 5x maior do que o PIB brasileiro em 2021 US\$ 1,8 tri. A presença da Black Rock Inc, dentre outros grandes grupos de investidores internacionais, aponta-nos que **sim, é um mercado lucrativo, não o fosse, grandes empresas não estariam aportando um grande volume de capital neste mercado**, porém isto torna a Vale uma empresa global, eliminando-a da lista de empresas a serem escolhidas, visto que nosso cliente deseja investir em uma empresa nacional, para que o dinheiro fique no Brasil, rendendo retorno ao país em que sua empresa esta inserida.



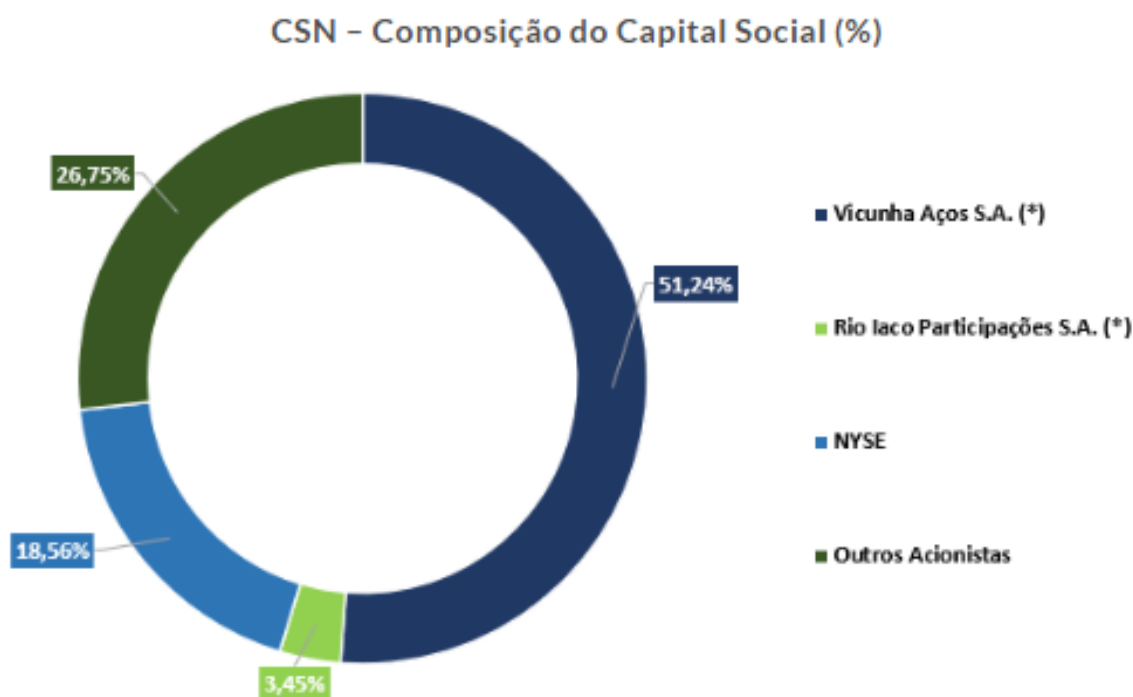
Quadro com empresas ligadas à Black Rock Inc. (fonte)

- **Usiminas**

Segundo [matéria publicada na gazeta do povo](#), embora o nome sugira o contrário, argentinos e japoneses são os donos da Usiminas, tendo vários japoneses, inclusive na composição de sua diretoria, [segundo a própria USIMINAS](#). Logo também por não ser genuinamente brasileira, não está adequada a solicitação de investimento do meu cliente.

- **CSN – Companhia de siderúrgica nacional**

Como visto, fundada por Getúlio Vargas em 1941, com capital estadunidense, emprestado em uma barganha por uma base aérea na segunda guerra mundial. Presidida por um brasileiro, essa sim esperamos que seja nacional, devido ao nome, devido à história. Porém:



Composição Acionária CSN [fonte](#)

De propriedade de Benjamin Steinbruch, banqueiro carioca, formado pela fundação Getúlio Vargas, fundador do grupo Vicunha, maior grupo têxtil da América Latina, ex gestor executivo do banco Safra, e Ex presidente da FIEP. Mesmo com seu capital aberto na bolsa de Nova York, tem sua maior parte brasileira, logo entra para lista como uma das empresas potenciais para o aporte de meu cliente.

- **ArcelorMittal Brasil**

Embora segundo apontado no nosso estudo de mercado, destaque-se em atitudes ecológicas, trabalhistas e com grande governança corporativa, o grupo ArcelorMittal tem

sede em Luxemburgo e é de propriedade do Indiano Lakshmi Mittal, logo será retirada de nossa lista.

H1: Investir em ações da Gerdau e CSN, no curto e médio prazo (2 e 5 anos) é financeiramente rentável.

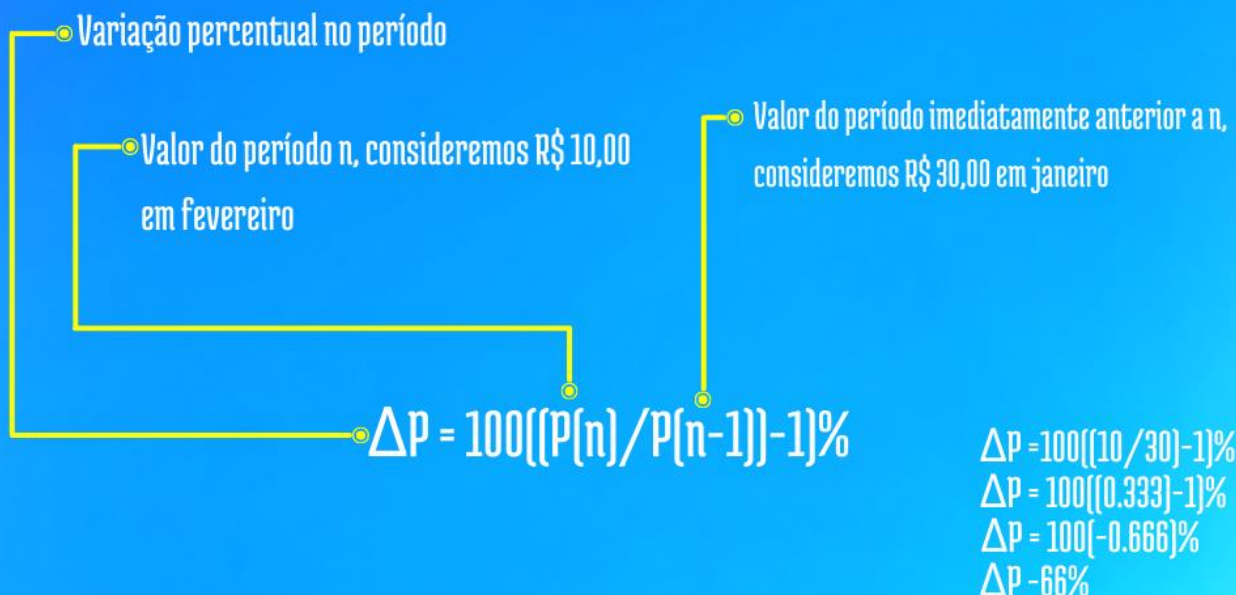
Validação H1: Aqui levantamos um ponto importante: no próximo ciclo, irei aprofundar mais no estudo do mercado financeiro, que é o real assunto do nosso problema. Porém é sabido que o investimento rentável é aquele, no qual o capital investido, rendeu um valor acima da inflação, e acima de um investimento livre de risco, tomemos como investimento livre de risco, o tesouro SELIC, que é aquele investimento em que se adquire letras do tesouro nacional (LTF), ou seja, se empresta para o governo, que tende sempre à adimplência, visto que o pagamento do rendimento é feito em moeda, que o governo mesmo imprime. Embora estejamos enfrentando um momento atípico, em que a taxa Selic ficou negativa (-0,68% em julho de 2022), em parte devido a baixa nos preços dos combustíveis: a chamada deflação de demanda. Tomemos ainda assim, por referência o rendimento do tesouro direto Selic, para observamos como esta taxa se relacionou com as cotações das ações das empresas **GGBR4**(Gerdau preferenciais), **GGBR3**(Gerdau Metalúrgica) e **CSNA3**(Companhia Siderúrgica Nacional) na Bovespa. Para isto utilizei o Python, para criar um [notebook Jupyter\(clique aqui para acessar\)](#), no qual gerei um série histórica, com os valores das cotações das ações na **Bovespa**, das empresas supracitadas, de janeiro/2013 até julho/2022.

METODOLOGIA DA ANÁLISE

Inicialmente realizei a análise comparativa, do valor das [ações x Selic em Excel](#), buscando a criação de uma metodologia protótipo, para servir direcionamento à etapa posterior, no qual serão realizados os mesmos passos do protótipo, só que em **Python**, assim será possível o escalonamento, a automatização, e a reutilização dos algoritmos para necessidades posteriores, como por exemplo, analisar um período maior, ou valores de ações de outros ativos. Embora existam outros métodos científicos financeiros, já bem estabelecidos, para quantificar o retorno esperado de um ativo, como o **CAPM**(modelo de precificação de ativos de capital), e o **WACC**(custo médio ponderado de capital), que também aplicarei posteriormente em nossa análise, inicialmente, para realizar a comparação entre o rendimento das ações e o valor da Selic, criei uma fórmula que calcula a taxa de variação percentual no período.



Fórmula de cálculo da variação percentual no período



Logo temos que houve uma variação de -66% no período de janeiro-fevereiro

Fórmula autoral de cálculo de variação percentual

O cálculo da variação percentual é interessante pois é possível ser feito tanto em cima do valor das ações(reais), quanto em cima da taxa Selic(percentagem), igualando duas medidas que até então eram diferentes, e, por conseguinte, permitindo a comparação linear de ambas.



Kauê Mandarinino
DATA SCIENTIST

Análise Ações x Selic

Protótipagem do modelo de análise em Excel

Banco de Dados Cotações

Índice	Data	GGBR4	GGBR3	CSNA3
20	1/31/2013	12.901	10.6189	6.635826
38	2/28/2013	12.14776	9.989786	6.172862
58	3/28/2013	11.36466	9.643923	5.635824
80	4/30/2013	11.54951	9.366996	5.079046
101	5/31/2013	9.663493	8.356131	4.244311
121	6/28/2013	9.30065	7.732018	3.826942
143	7/31/2013	10.80386	8.682053	4.218626

Validação do filtro mês: Mês 1, Ano 2013

1. Banco de Dados Cotações.csv 2. Query Filtrada 3. Estruturando D

Modo de filtro: Acessibilidade: Investigar

Aba 1: Banco de dados

Banco de dados com as datas e valores das ações das empresas escolhidas, de 2013 até 2022, esses dados foram obtidos em Python na etapa de mineração

Aba 2: Query filtrada

Utilizando a filtragem do Excel filtrei apenas os últimos dias uteis no mês.

Data	GGBR4	GGBR3	CSNA3
1/31/2013	12.901	10.6189	6.63
2/28/2013	12.14776	9.989786	6.17
3/28/2013	11.36466	9.643923	5.63
4/30/2013	11.54951	9.366996	5.07
5/31/2013	9.663493	8.356131	4.24

Valores das Ações

Índice	CSNA3	2013	2014	2015	2016	2017	2018	2019	2020	2021
1	CSNA3	6.635826	7.826395	3.138779	2.700161	8.980243	8.544836	8.511428	12.06138	28.48308
2	CSNA3	6.172862	6.945139	3.963687	4.048053	9.31457	7.891727	10.91466	10.4532	32.29723
3	CSNA3	5.635824	7.090333	4.214105	5.543648	7.083117	6.84209	13.55987	6.500238	35.51016
4	CSNA3	5.079046	6.168014	6.243407	10.21648	6.025704	6.82654	12.60998	8.377517	46.08825
5	CSNA3	4.244311	6.18963	4.773912	5.084917	5.33372	5.001025	15.14519	9.609256	42.05369
6	CSNA3	3.826942	6.773285	4.035278	6.080129	5.748208	6.11123	15.30085	10.00656	41.77061
7	CSNA3	4.218626	8.250439	3.82169	4.607029	5.924627	7.137543	15.13603	11.47756	44.39136
8	CSNA3	5.581767	7.090333	5.721235	6.686588	6.756563	7.326505	13.14903	14.25091	34.02408
9	CSNA3	6.212178	5.001025	3.04784	7.067507	7.471872	7.793798	12.36993	15.45957	28.04107
10	CSNA3	5.054317	5.944639	3.374304	8.350459	6.554411	7.585722	11.05289	19.24482	22.23375
11	CSNA3	8.329568	4.359402	4.276306	9.664452	5.722475	7.401606	11.74348	22.02754	21.23821
12	CSNA3	10.05751	4.071177	3.11004	8.435985	6.515535	7.376573	13.19272	29.84165	24.39076

Aba 3: Estruturando os dados

Mapeei, estruturei os dados em forma matricial e os validei manualmente, conferindo os valores em amarelo.

Aba 4: Ações e variações

Nesta aba estruturei os dados dos valores das ações, e criei uma fórmula para calcular a variação percentual/mês a mês

$$\Delta P = 100((P(n)/P(n-1)) - 1)\%$$

Levando em conta a cotação do último dia útil do mês.

Valor Ações	Variação GGBR4
Mes/Ano	Mes/Ano
2013	2013
2014	2014
2015	2015
2016	2016
2017	2017
2018	2018
2019	2019
2020	2020
2021	2021

Aba 5: Cotações Selic

Nesta aba retirei manualmente as cotações da taxa Selic no site do Governo Federal, e incidi sobre elas a mesma metodologia de cálculo de variação percentual mês a mês da aba anterior. Então tive um indicador em comum na mesma medida para quantificar a evolução percentual das ações x o valor da taxa Selic, conseguindo assim identificar qual teve maior rentabilidade no período.

Abas 6 e 7

De posse dessas informações, plotei um comparativo com escala de cores, e também gráficos para melhor compreensão do significado dos indicadores.

RENDIMENTO TOTAL	Últimos 3a 2019~2021	Últimos 5a 2017~2021	Últimos 9a 2013~21
Rendimento Selic	17.37%	33.14%	77.20%
Rendimento Médio das Ações	126.71%	151.51%	192.96%
GGBR4	111.66%	135.34%	167.15%
GGBR3	118.14%	160.45%	155.41%
CSNA3	150.33%	158.74%	256.31%

SELIC X MÉDIA DAS AÇÕES



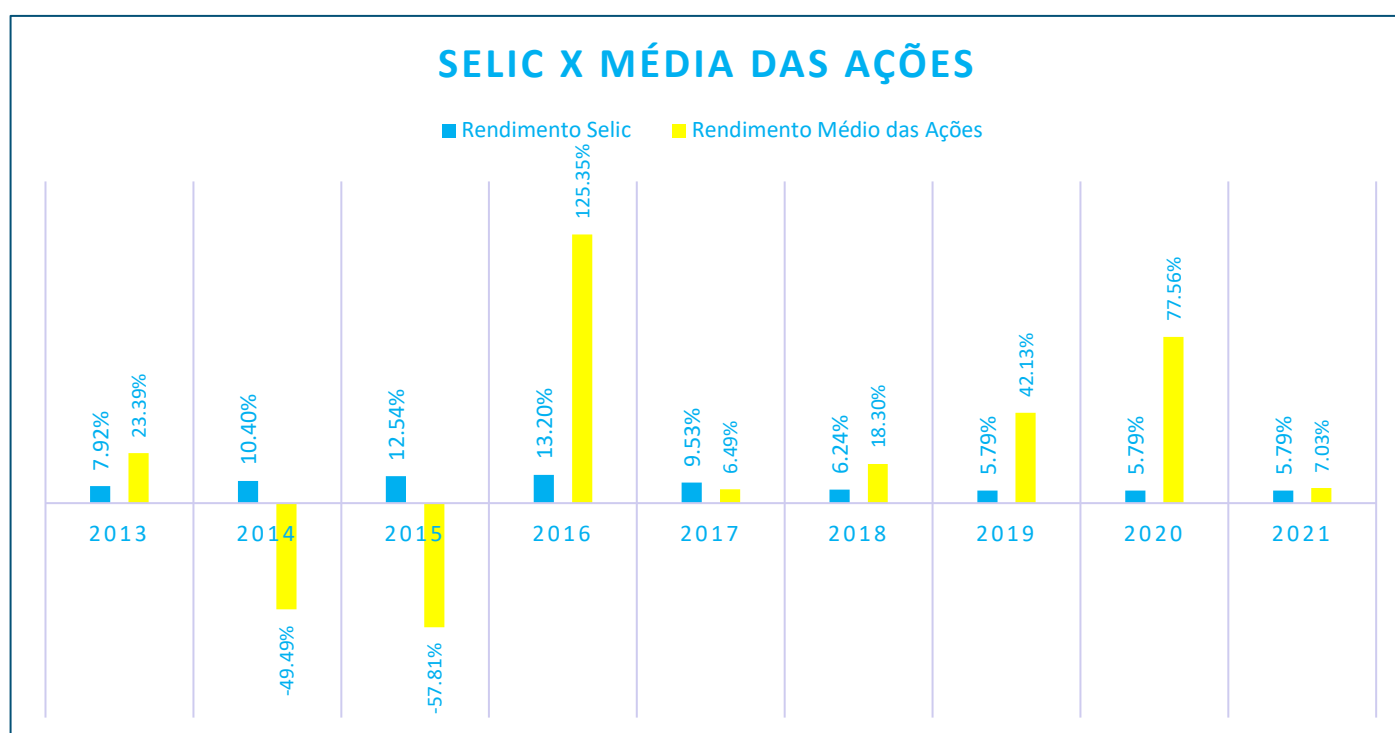
Ainda com o protótipo consegui gerar esta escala de cor:

RENDIMENTO TOTAL	Últimos 3a 2019~2021	Últimos 5a 2017~2021	Últimos 9a 2013~21
Rendimento Selic	17.37%	33.14%	77.20%
Rendimento médio das ações	126.71%	151.51%	192.96%
GGBR4	111.66%	135.34%	167.15%
GGBR3	118.14%	160.45%	155.41%
CSNA3	150.33%	158.74%	256.31%

Comparativo em escala de cor do rendimento Selic x valor das ações,
segundo o método de variação percentual.

Onde podemos verificar, destacado em vermelho, que no nos períodos analisados: últimos 3, 5 e 9 anos, as ações tiveram um rendimento percentual notadamente maior que o rendimento seguro da Selic. Anotei esta informação, e continuei a análise gráfica em busca de mais inferências de informações.

Abaixo, ainda no protótipo, plotei um gráfico geral em que se pode comparar visualmente a variação percentual das ações x o valor da Selic.



Podemos notar que em 2014 e 2015 o rendimento das ações teve uma queda expressiva, e entender a causa dessa queda é entender a relação de funcionamento da valorização e desvalorização das ações do setor, devido a isto, ampliando um pouco mais, ainda a validação da hipótese H1, voltei a etapa de estudo e identifiquei que neste

período o país estava em crise, devido à uma somatória de políticas económicas do governo, como liberação abrupta do preço da gasolina, aumento na taxa de juros para lidar com o choque inflacionário decorrente da liberalização, houve em consequência disto, aumento no desemprego e contração na renda(mercado interno), isto somado a valorização do real, culminou no baixo rendimento do setor, logo temos um insight interessante, que não tinha sido notado até agora: somos o segundo maior exportador de minério do ferro do mundo, porém para a exportação estar em alta, o real tem que estar desvalorizado em relação ao dólar. Conseguimos definir uma relação de casualidade linear de grande peso, o valor do dólar interfere no valor da ação.

Insight: Existe uma relação linear de casualidade entre o valor do dólar e o valor das ações siderúrgicas.

Este insight é muito interessante, pois pode posteriormente, junto com outras características, tal como, o poder de compra do brasileiro (mercado interno), dar origem a um algoritmo de regressão linear, em que podemos ter um arcabouço científico para a predição dos valores das ações. Posterior ao término do escopo deste projeto de insights, vou conectá-lo com um projeto de Machine Learning, para criar um modelo preditivo com base nessa relação de casualidade entre o câmbio e o valor das ações, diante disso, faz se ainda mais necessária a transição deste modelo prototipado em Excel, para um modelo mais eficiente em Python no qual poderemos filtrar facilmente um período maior de dados, e também, escalonar esta lógica para outros ativos .

Visando que esta etapa não se torne demasiadamente longa, deixo uma curiosidade abaixo e encerro o segundo ciclo por aqui.

Curiosidade: no processo de automação Python do protótipo supracitado, um dos requisitos funcionais, era que o sistema recebe-se um input, com um intervalo de datas, e me retorna-se uma lista com os últimos dias úteis de cada mês, porém não encontrei na comunidade uma biblioteca fizesse isto, então, criei eu mesmo esta biblioteca. A [documentação](#) e o repositório já estão disponíveis no meu [GitHub](#).

```
testando.py x  calendario_dias_uteis.py x  __init__.py x
import calendario_dias_uteis as cdu

# Vou chamar a função e informar o período e a
# biblioteca vai me retornar uma lista com todos
# os dias últimos dias uteis de cada mes no período
lista = cdu.ultimo_util(2000, 2022)
print(lista)
```

testando (2) x

```
'30/12/2011', '31/1/2012', '29/2/2012', '30/3/2012', '30/4/2012', '31/5/2012'
'30/11/2012', '31/12/2012', '31/1/2013', '28/2/2013', '29/3/2013', '30/4/2013'
'31/10/2013', '29/11/2013', '31/12/2013', '31/1/2014', '28/2/2014', '31/3/2014'
'30/9/2014', '31/10/2014', '28/11/2014', '31/12/2014', '30/1/2015', '27/2/2015'
'31/8/2015', '30/9/2015', '30/10/2015', '30/11/2015', '31/12/2015', '29/1/2016'
'29/7/2016', '31/8/2016', '30/9/2016', '31/10/2016', '30/11/2016', '30/12/2016'
'30/6/2017', '31/7/2017', '31/8/2017', '29/9/2017', '31/10/2017', '30/11/2017'
'31/5/2018', '29/6/2018', '31/7/2018', '31/8/2018', '28/9/2018', '31/10/2018'
'30/4/2019', '31/5/2019', '28/6/2019', '31/7/2019', '30/8/2019', '30/9/2019',
'31/3/2020', '30/4/2020', '29/5/2020', '30/6/2020', '31/7/2020', '31/8/2020',
'26/2/2021', '31/3/2021', '30/4/2021', '31/5/2021', '30/6/2021', '30/7/2021',
'31/1/2022', '28/2/2022', '31/3/2022', '29/4/2022', '31/5/2022', '30/6/2022',
'30/12/2022']
```

Biblioteca Python `calendario_dias_uteis`, criada para auxiliar no processo de análise exploratório dos dados.