

Estudo do Índice S&P 500

Vitor Hugo Alves Rodrigues

15 de fevereiro de 2024

1 Ferramentas Usadas

Estamos usando o Python 3.11.2 com o Framework Pandas 2.2.0. Toda a análise usando o jupyter notebook.

2 Introdução ao Estudo

Nessa seção iremos fazer uma análise dos dados do índice S&P¹. É uma abreviação para Standard & Poor's 500. Um índice composto por quinhentos ativos (ações) da bolsa de valores da NYSE ou NASDAQ. Iremos carregar as bibliotecas para uso do pandas e exibição dos gráficos:

```
import pandas as pd
import numpy as np
```

Agora iremos fazer o carregamento do arquivo CSV que contém os dados que precisamos:

```
data_spx = pd.read_csv('spx.csv', parse_dates=['Date'], index_col='Date')
```

Acima estamos importando os dados usando a coluna 'Date' como data e indicando que está mesma coluna seja usada como índice das linhas. Os dados que temos acesso estão organizados de forma diária. De 01/02/1990 até 14/10/2011. Segue uma imagem dos dados que temos carregados:



SPX	
Date	
1990-02-01	328.79
1990-02-02	330.92
1990-02-05	331.85
1990-02-06	329.66
1990-02-07	333.75
...	...
2011-10-10	1194.89
2011-10-11	1195.54
2011-10-12	1207.25
2011-10-13	1203.66
2011-10-14	1224.58

Figura 1: Estrutura original dos dados

3 Análise: média mensal separado por ano

Aqui estamos buscando encontrar algum padrão nos dados de uma forma específica. Os dados são diários. Pegaremos a média de cada mês e faremos um gráfico de cada ano. Ou seja, teremos 12 pontos no gráfico para cada ano. Tentaremos descobrir algo com esse tipo de visualização.

Primeiro precisamos agrupar os dados por mês de toda a série história.

¹Mais informações em: https://pt.wikipedia.org/wiki/S&P_500

```
data_spx_mensal_media = data_spx.resample('MS').mean()
```

Com a função **'resample'** faremos uma reamostragem do intervalo de datas para que tem um intervalo mensal. Para isso passamos o código **'MS'** para isso. Teremos intervalo separados pelo primeiro dia de cada mês, pois **'MS'** vem de **MonthBegin**. Com isso a variável *data_spx_mensal_media* armazena os seguintes dados:

Date	
1990-02-01	330.45
1990-03-01	338.46
1990-04-01	338.18
1990-05-01	350.25
1990-06-01	360.39
...	
2011-06-01	1287.29
2011-07-01	1325.18
2011-08-01	1185.31
2011-09-01	1173.88
2011-10-01	1171.36

Figura 2: Dados da média mensal do indicador S&P

Para montar os gráficos iremos precisar de uma tabela que tenha os dados mensais de cada ano separados em colunas diferentes com os mesmos meses. No final essa tabela terá as seguintes colunas:

```
Index(['SPX_1990', 'Mês', 'SPX_1991', 'SPX_1992', 'SPX_1993', 'SPX_1994',
      'SPX_1995', 'SPX_1996', 'SPX_1997', 'SPX_1998', 'SPX_1999', 'SPX_2000',
      'SPX_2001', 'SPX_2002', 'SPX_2003', 'SPX_2004', 'SPX_2005', 'SPX_2006',
      'SPX_2007', 'SPX_2008', 'SPX_2009', 'SPX_2010', 'SPX_2011'],
      dtype='object')
```

Figura 3: Estrutura da nossa tabela com os dados juntos

```
data_spx_1990 = data_spx_mensal_media.loc['1990-01-01':'1990-12-01']
data_spx_1990.rename(columns='SPX':'SPX_1990', inplace=True)
data_spx_1990['Mês'] = np.array(data_spx_1990.index.strftime('%
```

Acima criamos um novo DataFrame para comportar os dados dos doze meses do ano de 1990. Alteramos o nome da coluna convenientemente para este ano. Também criamos uma coluna **'Mês'** que será usada como chave para fazer o merge com o DataFrame completo com todos os anos. A tabela ficou assim:

Date	SPX_1990	Mês
1990-02-01	330.452632	02
1990-03-01	338.465000	03
1990-04-01	338.178000	04
1990-05-01	350.250000	05
1990-06-01	360.386667	06
1990-07-01	360.030952	07
1990-08-01	330.746522	08
1990-09-01	315.406842	09
1990-10-01	307.124783	10
1990-11-01	315.290952	11
1990-12-01	328.749500	12

Figura 4: Dados da média mensal do indicador S&P

Essa trinca de comandos será feita para cada ano até 2011. Quando tivermos todos então iremos fazer o merge para juntar todas as colunas em uma única tabela.

```
data_merged_tempos = pd.merge(data_spx_1990, data_spx_1991, left_on='Mês',
                              right_on='Mês', how='outer')
```

```
data_merged_tempos = pd.merge(data_merged_tempos, data_spx_1992, left_on='Mês',
                               right_on='Mês', how='outer')
```

A propriedade **how** foi usada porque em 1990 não temos valor para o mês de Janeiro (01). Então precisamos que tudo que não seja encontrado seja incluído. Para os valores 'NA' depois podemos usar a função **fillna(0)** para zerar os valores. Fazemos isso para todos os anos até 2011. A seguir uma parte de como seria a tabela final:

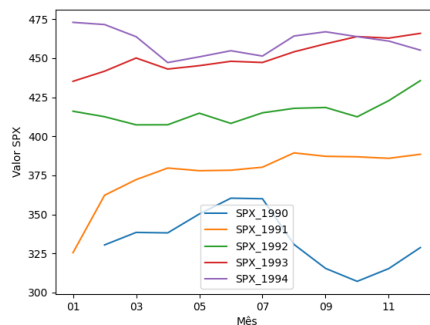
Mês	SPX_1990	SPX_1991	SPX_1992
01	0.00	325.49	416.08
02	330.45	362.26	412.56
03	338.46	372.28	407.36
04	338.18	379.68	407.41
05	350.25	377.99	414.81
06	360.39	378.29	408.27
07	360.03	380.23	415.05
08	330.75	389.40	417.93
09	315.41	387.20	418.48
10	307.12	386.88	412.50
11	315.29	385.92	422.84
12	328.75	388.51	435.64

Figura 5: Estrutura da nossa tabela com os dados juntos

Agora faremos a plotagem dessas informações nos gráficos. Separamos em gráficos diferentes para ver se algum padrão entre os anos se apresenta.

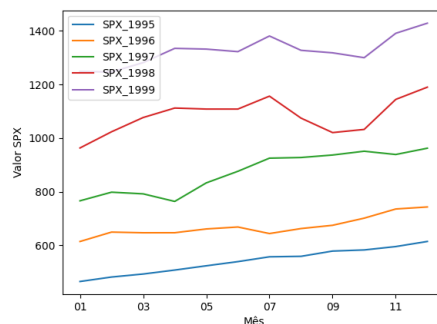
```
data_merged_tempos.loc[:, ['Mês', 'SPX_1990', 'SPX_1991', 'SPX_1992',
                             'SPX_1993', 'SPX_1994']].plot(x='Mês', ylabel='Valor SPX')
```

Para plotar o gráfico usaremos a função do DataFrame **plot**. Iremos fazer uma visualização das colunas que queremos para que o gráfico não fique muito poluído com informação demais. Acima selecionamos o intervalos de 1990 até 1994. Abaixo o gráfico:



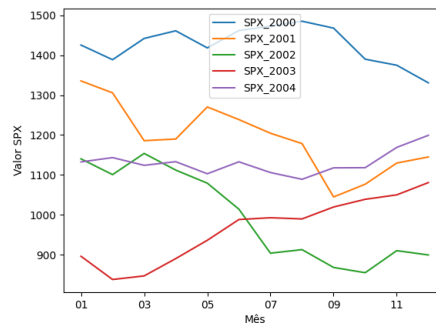
```
data_merged_tempos.loc[:, ['Mês', 'SPX_1995', 'SPX_1996', 'SPX_1997',
                             'SPX_1998', 'SPX_1999']].plot(x='Mês', ylabel='Valor SPX')
```

Intervalo de 1995 até 1999. Abaixo o gráfico:



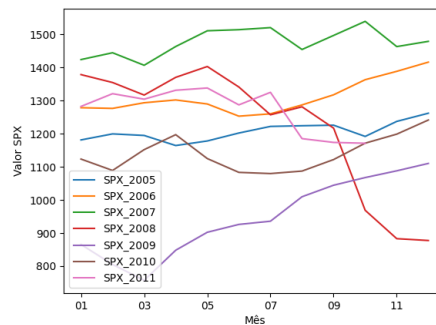
```
data_merged_tempos.loc[:, ['Mês', 'SPX_2000', 'SPX_2001', 'SPX_2002',  
'SPX_2003', 'SPX_2004']].plot(x='Mês', ylabel='Valor SPX')
```

Intervalo de 2000 até 2004. Abaixo o gráfico:



```
data_merged_tempos.loc[:, ['Mês', 'SPX_2005', 'SPX_2006', 'SPX_2007',  
'SPX_2008', 'SPX_2009', 'SPX_2010', 'SPX_2011']].plot(x='Mês', ylabel='Valor  
SPX')
```

Intervalo de 2005 até 2011. Abaixo o gráfico:



Há um comportamento "caótico" em quase todos os períodos dos gráficos mas um deles se mostrou mais "comportado". O período de 1995 até 1999 tem um crescimento muito parecido em todos os anos. Acredito ter identificado um grupo que tem algo diferente nos outros períodos. Poderíamos analisar outros índices para ver se o comportamento segue esse padrão. Ou tentar procurar dentro deste índices ações que mais pesaram no comportamento. Assim rastreando mais a fundo do porque esse fenômeno. Como estamos falando de um índice, uma ação ou algumas delas tiveram registros diferentes. Então podemos ir atrás de alguma para saber o que houve. Também uma análise de acontecimentos gerais no país pode responder.