

# First\_notebook

May 8, 2023

```
[33]: %matplotlib inline
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
sns.set(style = 'darkgrid')
```

```
[4]: df = pd.read_csv('/home/student/Desktop/classroom/myfiles/notebooks/fortune500.
↪csv')
```

```
[5]: df.head()
```

```
[5]:
```

	Year	Rank	Company	Revenue (in millions)	Profit (in millions)
0	1955	1	General Motors	9823.5	806
1	1955	2	Exxon Mobil	5661.4	584.8
2	1955	3	U.S. Steel	3250.4	195.4
3	1955	4	General Electric	2959.1	212.6
4	1955	5	Esmark	2510.8	19.1

```
[6]: df.tail()
```

```
[6]:
```

	Year	Rank	Company	Revenue (in millions)	\
25495	2005	496	Wm. Wrigley Jr.	3648.6	
25496	2005	497	Peabody Energy	3631.6	
25497	2005	498	Wendy's International	3630.4	
25498	2005	499	Kindred Healthcare	3616.6	
25499	2005	500	Cincinnati Financial	3614.0	

	Profit (in millions)
25495	493
25496	175.4
25497	57.8
25498	70.6
25499	584

```
[10]: df.columns = ('year', 'rank', 'company', 'revenue', 'profit')
```

```
[11]: df.head()
```

```
[11]:
```

	year	rank	company	revenue	profit
0	1955	1	General Motors	9823.5	806
1	1955	2	Exxon Mobil	5661.4	584.8
2	1955	3	U.S. Steel	3250.4	195.4
3	1955	4	General Electric	2959.1	212.6
4	1955	5	Esmark	2510.8	19.1

```
[12]: len(df)
```

```
[12]: 25500
```

```
[13]: df.dtypes
```

```
[13]: year          int64
rank          int64
company       object
revenue      float64
profit       object
dtype: object
```

```
[17]: non_numeric_profits = df.profit.str.contains('[^0-9.-]')
df.loc[non_numeric_profits].head()
```

```
[17]:
```

	year	rank	company	revenue	profit
228	1955	229	Norton	135.0	N.A.
290	1955	291	Schlitz Brewing	100.0	N.A.
294	1955	295	Pacific Vegetable Oil	97.9	N.A.
296	1955	297	Liebmann Breweries	96.0	N.A.
352	1955	353	Minneapolis-Moline	77.4	N.A.

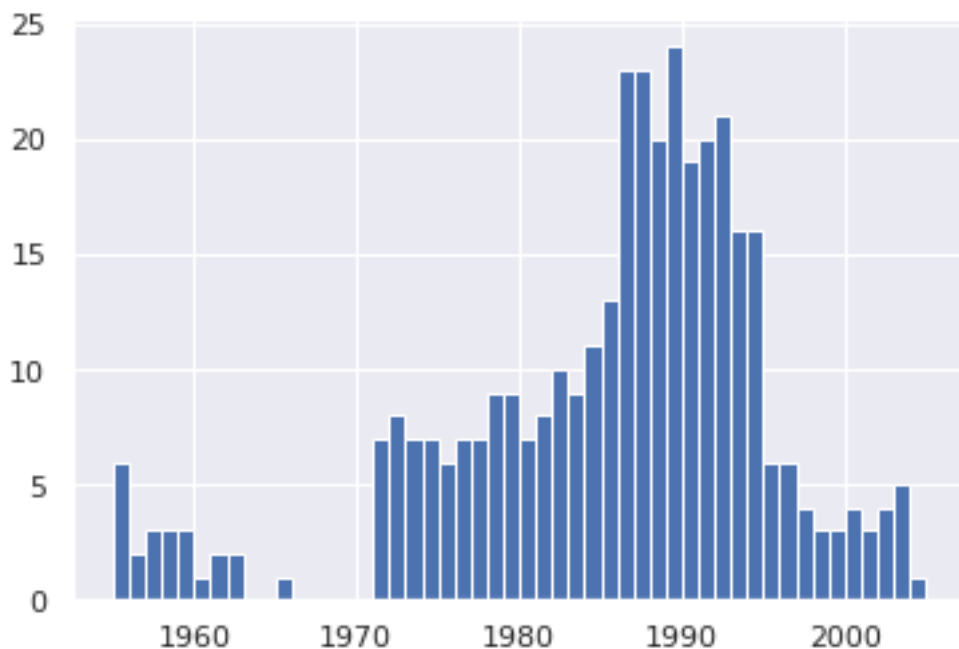
```
[19]: set(df.profit[non_numeric_profits])
```

```
[19]: {'N.A.'}
```

```
[20]: len(df.profit[non_numeric_profits])
```

```
[20]: 369
```

```
[34]: bin_sizes, _, _ = plt.hist(df.year[non_numeric_profits], bins=range(1955, 2006),
↪)
```



```
[35]: df = df.loc[-non_numeric_profits]
df.profit = df.profit.apply(pd.to_numeric)
```

```
[36]: len(df)
```

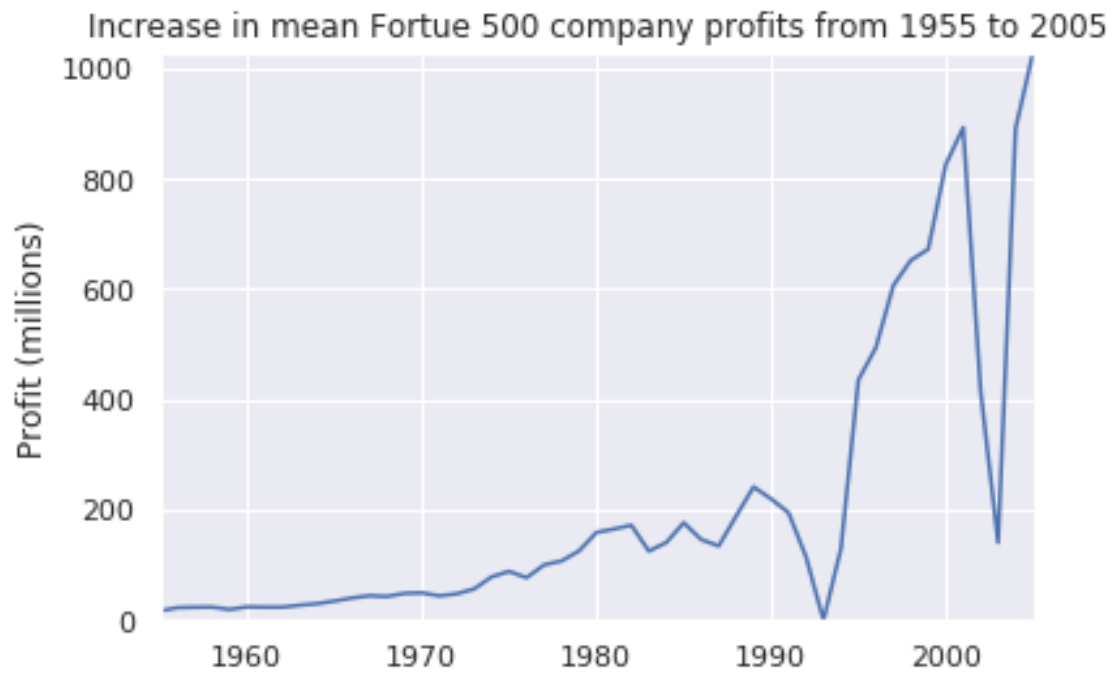
```
[36]: 25131
```

```
[37]: df.dtypes
```

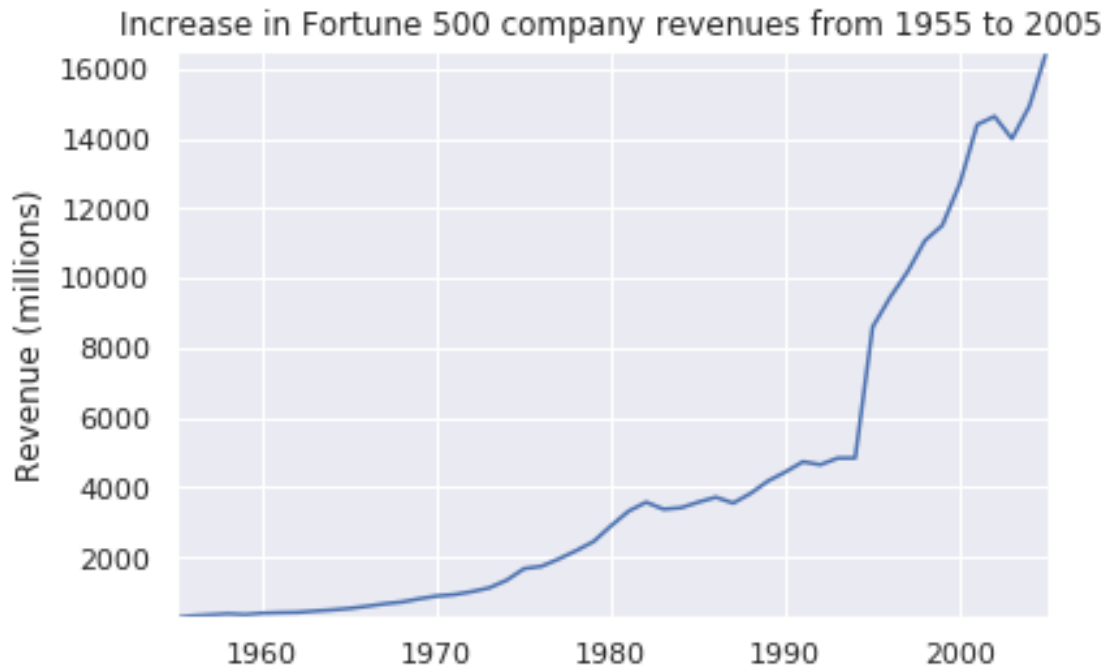
```
[37]: year          int64
rank          int64
company        object
revenue        float64
profit         float64
dtype: object
```

```
[38]: group_by_year = df.loc[:, ['year', 'revenue', 'profit']].groupby('year')
avgs = group_by_year.mean()
x = avgs.index
y1 = avgs.profit
def plot(x, y, ax, title, y_label):
    ax.set_title(title)
    ax.set_ylabel(y_label)
    ax.plot(x, y)
    ax.margins(x = 0, y = 0)
```

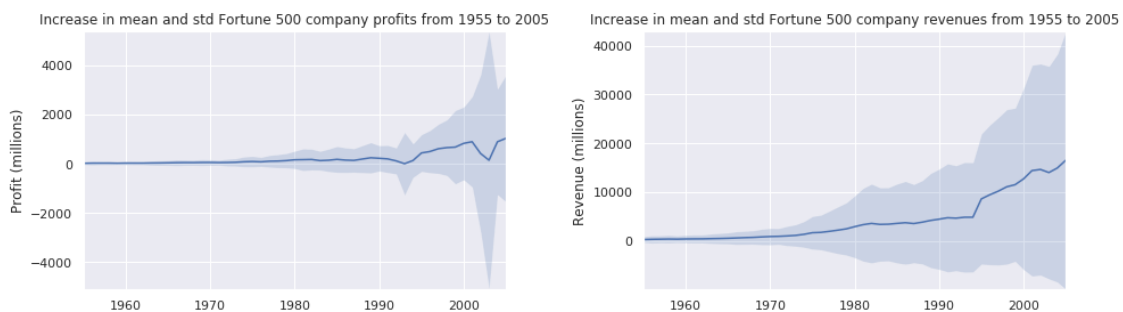
```
[40]: fig, ax = plt.subplots()
      plot(x, y1, ax, 'Increase in mean Fortue 500 company profits from 1955 to 2005', 'Profit (millions)')
```



```
[41]: y2 = avgs.revenue
      fig, ax = plt.subplots()
      plot(x, y2, ax, 'Increase in Fortune 500 company revenues from 1955 to 2005', 'Revenue (millions)')
```



```
[47]: def plot_with_std(x, y, stds, ax, title, y_label):
        ax.fill_between(x, y - stds, y + stds, alpha = 0.2)
        plot(x, y, ax, title, y_label)
    fig, (ax1, ax2) = plt.subplots(ncols= 2)
    title = 'Increase in mean and std Fortune 500 company %s from 1955 to 2005'
    stds1 = group_by_year.std().profit.values
    stds2 = group_by_year.std().revenue.values
    plot_with_std(x, y1.values, stds1, ax1, title % 'profits', 'Profit (millions)')
    plot_with_std(x, y2.values, stds2, ax2, title % 'revenues', 'Revenue (millions)')
    fig.set_size_inches(14,4)
    fig.tight_layout()
```



[ ]: