

Business case: Understanding profitability in the US financial market

Eduardo Rodríguez Gil - A01274913

This is business case prepared for the Statistics Module (Bloque 1) of the Advanced AI concentration

1 Case Description

You were hired as a data scientist in an important mutual fund firm in the department of financial analysis. The firm has been doing financial analysis and financial forecast for several years. You were hired to come up with alternative approaches to do descriptive analytics in order to find better future alternatives for forecasting methods.

You have to analyze historical quarterly financial statements of all US public firms listed in the New York Exchange and NASDAQ. You will receive this dataset in a .csv format.

You have to carefully read the data dictionary to understand each variable and the dataset to understand its structure.

2 Business Questions

All your data and statistical analysis has to be tailored to respond the following questions:

```
# Importamos las librerias que vamos a utilizar.  
import numpy as np  
import pandas as pd  
import pandas_datareader as pdr  
import seaborn as sb  
import matplotlib.pyplot as plt  
import plotly.express as px
```

```
# Importamos los datos del dataframe del .csv  
df_us2022 = pd.read_csv('us2022q2a.csv') # US 2022 data frame
```

```
df_firms = pd.read_csv('usfirms2022.csv') # US firms 2022 data frame
```

df_us2022

	firm	q	revenue	cogs	sgae	otheropexp	extraincome	finexp	i
0	A	2000q1	NaN	NaN	NaN	NaN	NaN	NaN	
1	A	2000q2	2485000.0	1261000.0	1010000.0	0.0	42000.0	0.0	
2	A	2000q3	2670000.0	1369000.0	1091000.0	0.0	28000.0	0.0	
3	A	2000q4	3372000.0	1732000.0	1182000.0	0.0	10000.0	0.0	
4	A	2001q1	2841000.0	1449000.0	1113000.0	0.0	-6000.0	0.0	
...	
82114	DCO	2000q2	NaN	NaN	NaN	NaN	NaN	NaN	
82115	DCO	2000q3	NaN	NaN	NaN	NaN	NaN	NaN	
82116	DCO	2000q4	NaN	NaN	NaN	NaN	NaN	NaN	
82117	DCO	2001q1	NaN	NaN	NaN	NaN	NaN	NaN	
82118	DCO	2001q2	NaN	NaN	NaN	NaN	NaN	NaN	

82119 rows × 20 columns



df_firms

	Ticker	Name	N	Class	Country\nof Origin	Type of Asset	Sector NAICS\nlevel 1	Exchange / Src
0	FLWS	1 800 Flowers Com Inc	1	Com A	US	Stock	Retail Trade	NASDAQ
1	TXG	10x Genomics, Inc	2	Com A	US	Stock	Manufacturing	NASDAQ
2	GOED	1847 Goedeker Inc	3	Com	US	Stock	Retail Trade	AMEX

```
# Unimos la dos tablas para convertirlo en un solo dataframe
us2022_merged = df_us2022.merge(df_firms, left_on='firm', right_on='Ticker')

us2022_merged
```

	firm	q	revenue	cogs	sgae	otheropexp	extraincome	finexp	i
0	A	2000q1	NaN	NaN	NaN	NaN	NaN	NaN	
1	A	2000q2	2485000.0	1261000.0	1010000.0	0.0	42000.0	0.0	
2	A	2000q3	2670000.0	1369000.0	1091000.0	0.0	28000.0	0.0	
3	A	2000q4	3372000.0	1732000.0	1182000.0	0.0	10000.0	0.0	
4	A	2001q1	2841000.0	1449000.0	1113000.0	0.0	-6000.0	0.0	

▼ 2.1 General questions:

By industry, what is the composition of US public firms in terms of firm size, sales performance and profitability?

Why some firms are more profitable than others? Which factors/variables from financial statements are related to stock returns?

▼ 2.2.1 About descriptive statistics:

▼ 2.2.1.1 Considering the most recent financial quarter of the dataset:

```
# Obtenemos el financial quarter del dataset
us2022_quarter = us2022_merged[us2022_merged['q'] == '2022q2']
```

- Show how many firms by industry there are in the sample

```
# Aqui se muestra el número de industrias que hay en el actualmente en el dataframe
df_firms['Sector NAICS\nlevel 1'].value_counts()
```

Manufacturing	1567
Finance and Insurance	703
Information	263
Retail Trade	152
Professional, Scientific, and Technical Services	145
Administrative and Support and Waste Management and Remediation Services	133
Mining, Quarrying, and Oil and Gas Extraction	104
Wholesale Trade	79
Utilities	77
Transportation and Warehousing	69
Accommodation and Food Services	69
Real Estate and Rental and Leasing	68
Health Care and Social Assistance	64
Construction	45
Arts, Entertainment, and Recreation	22
Other Services (except Public Administration)	16
Agriculture, Forestry, Fishing and Hunting	16
Educational Services	14
-	2

Name: Sector NAICS\nlevel 1, dtype: int64

- For each industry (and for all industries), what can you say about the typical firm size in terms of market value and book value? How much these variables change within each industry? How firm size (in market value) is distributed?

Para poder realizar esto ocupamos las formulas que nos proporciono para poder hacer las Variable calculations.

```
# Obtenemos los Firm Size measures de Book value
us2022_merged['Book value'] = us2022_merged['totalassets'] - us2022_merged['totalliabilities']

# Obtenemos los Firm Size measures de Market value
us2022_merged['Market value'] = us2022_merged['originalprice'] * us2022_merged['sharesoutstan

# Obtenemos el Profit Margin measures con Operating profit
us2022_merged['Operating profit'] = us2022_merged['revenue'] - us2022_merged['cogs'] - us2022.

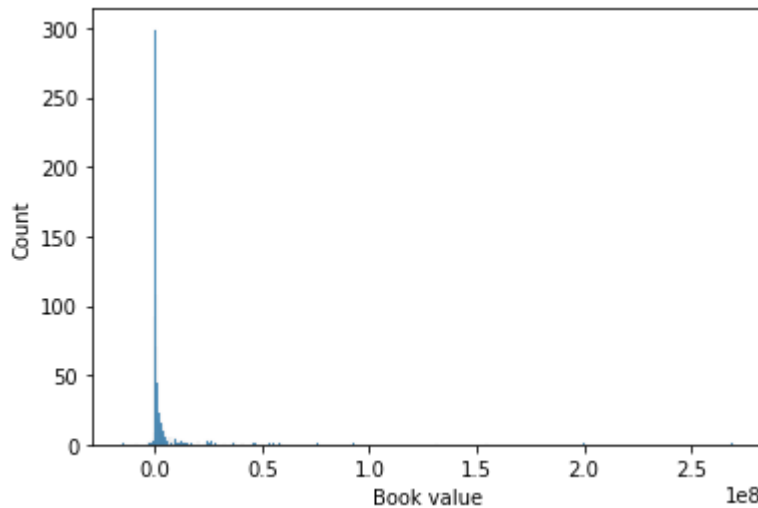
# Obtenemos el Profit Margin measures con Operating profit margin
us2022_merged['Operating profit margin'] = us2022_merged['Operating profit'] / us2022_merged[

# Obtenemos el Net Income
us2022_merged['Net Income'] = us2022_merged['Operating profit'] - us2022_merged['incometax']
```

```
# Obtenemos el Profit margin
us2022_merged['Profit margin'] = us2022_merged['Net Income'] / us2022_merged['revenue']
```

```
# Mostramos un histograma para ver la frecuencia del Book value
sb.histplot(us2022_quarter['Book value'])
```

<matplotlib.axes._subplots.AxesSubplot at 0x7f14f2342710>

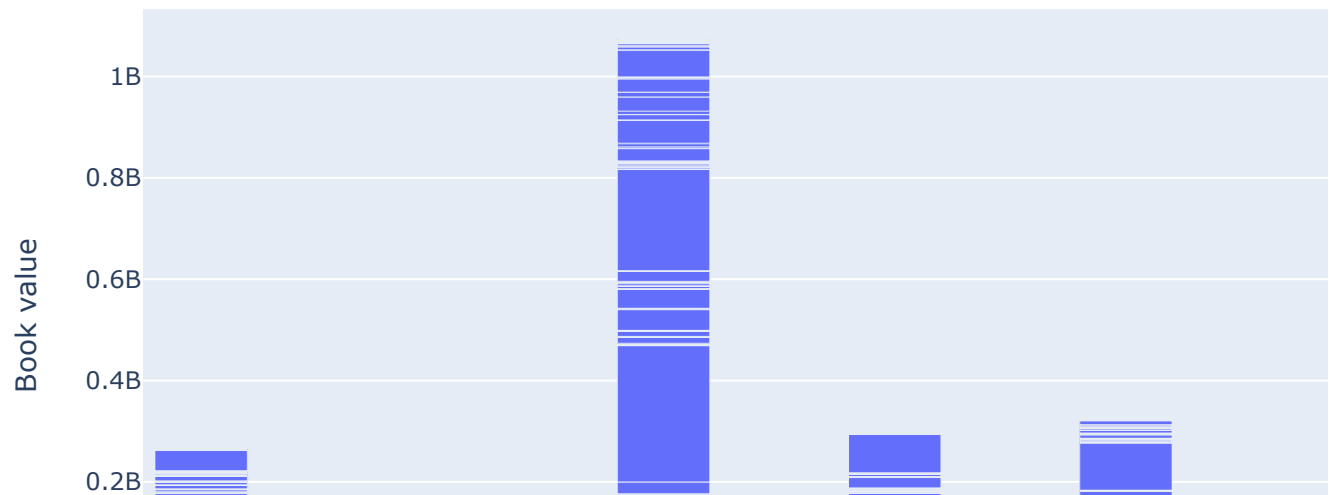


```
# Imprimimos valores descriptivos para tener una mejor perspectiva de los datos del histogram
print(us2022_quarter['Book value'].describe())
```

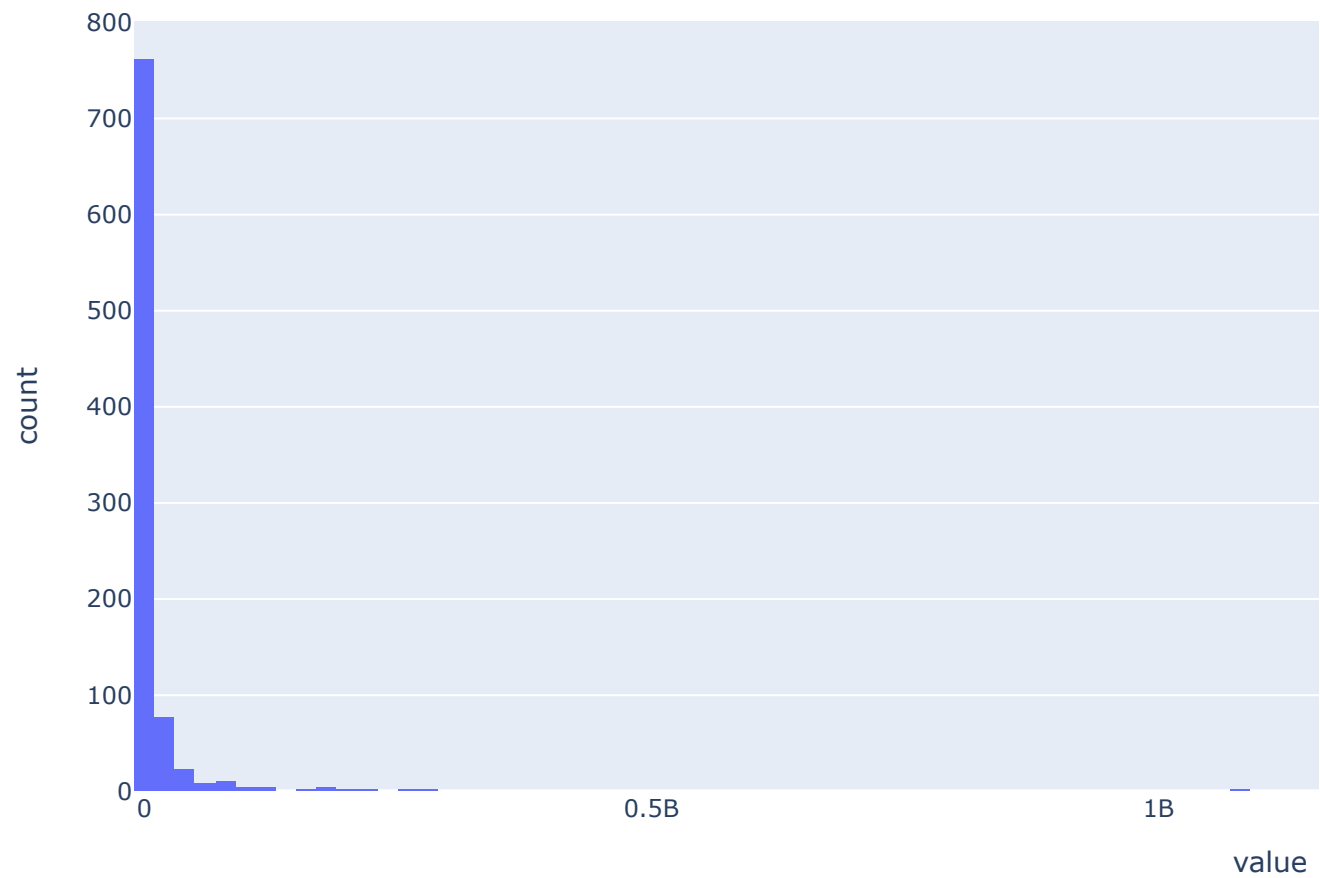
```
count    8.690000e+02
mean      3.477157e+06
std       1.524353e+07
min      -1.479100e+07
25%       1.012760e+05
50%       4.235880e+05
75%       1.746849e+06
max       2.691180e+08
Name: Book value, dtype: float64
```

```
# Mostramos el gráfico para revisar los datos de Book value de cada industria y podemos darnos
px.bar(us2022_quarter, x = 'Sector\nEconomatica', y = 'Book value', title = "Book values por
```

Book values por industria



Mostramos un histograma para ver la frecuencia del Market value
`px.histogram(us2022_quarter['Market value'])`

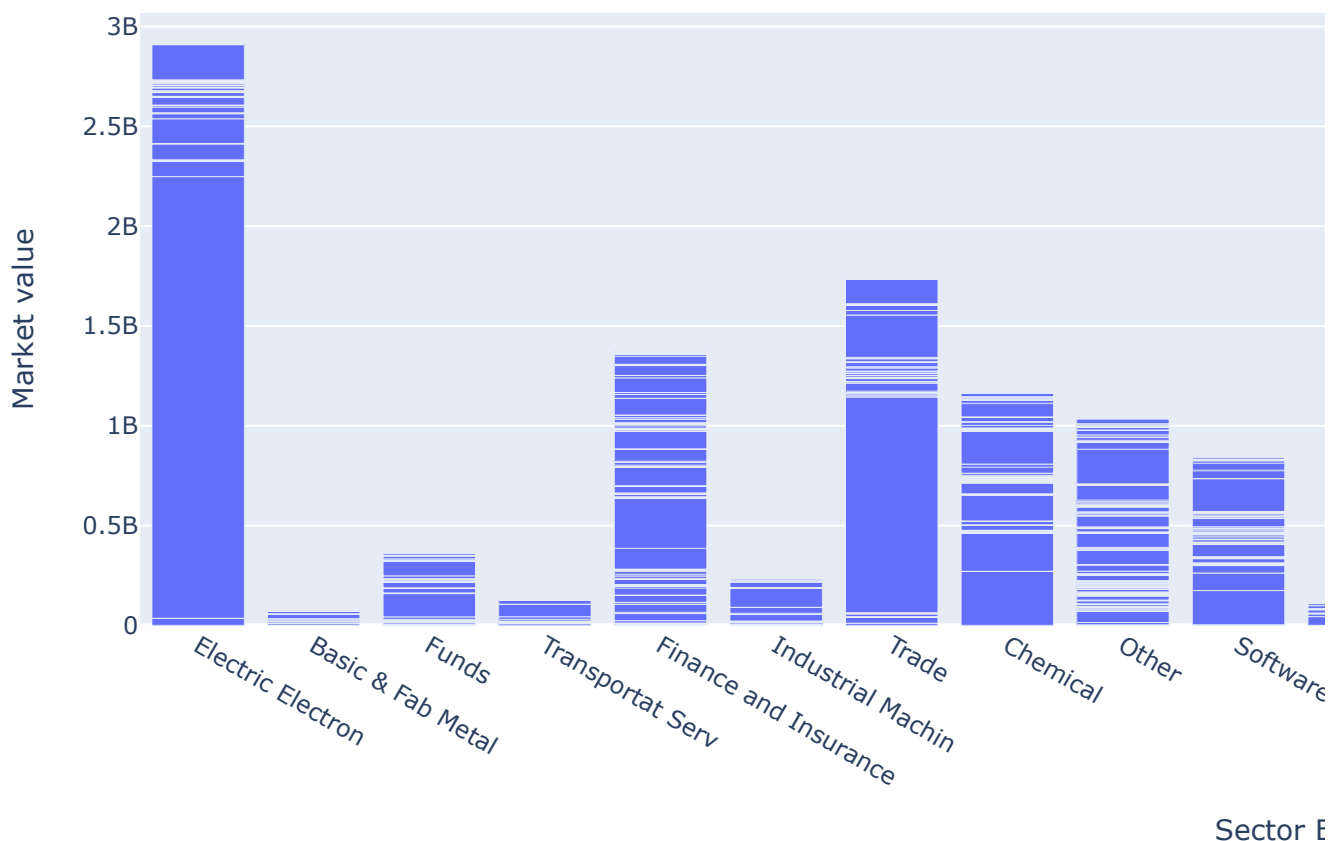


```
# Imprimimos valores descriptivos para tener una mejor perspectiva de los datos del histogram
print(us2022_quarter['Market value'].describe())
```

```
count      8.990000e+02
mean       1.221618e+07
std        8.595164e+07
min        3.490000e+01
25%        1.694179e+05
50%        9.956875e+05
75%        4.471550e+06
max        2.212838e+09
Name: Market value, dtype: float64
```

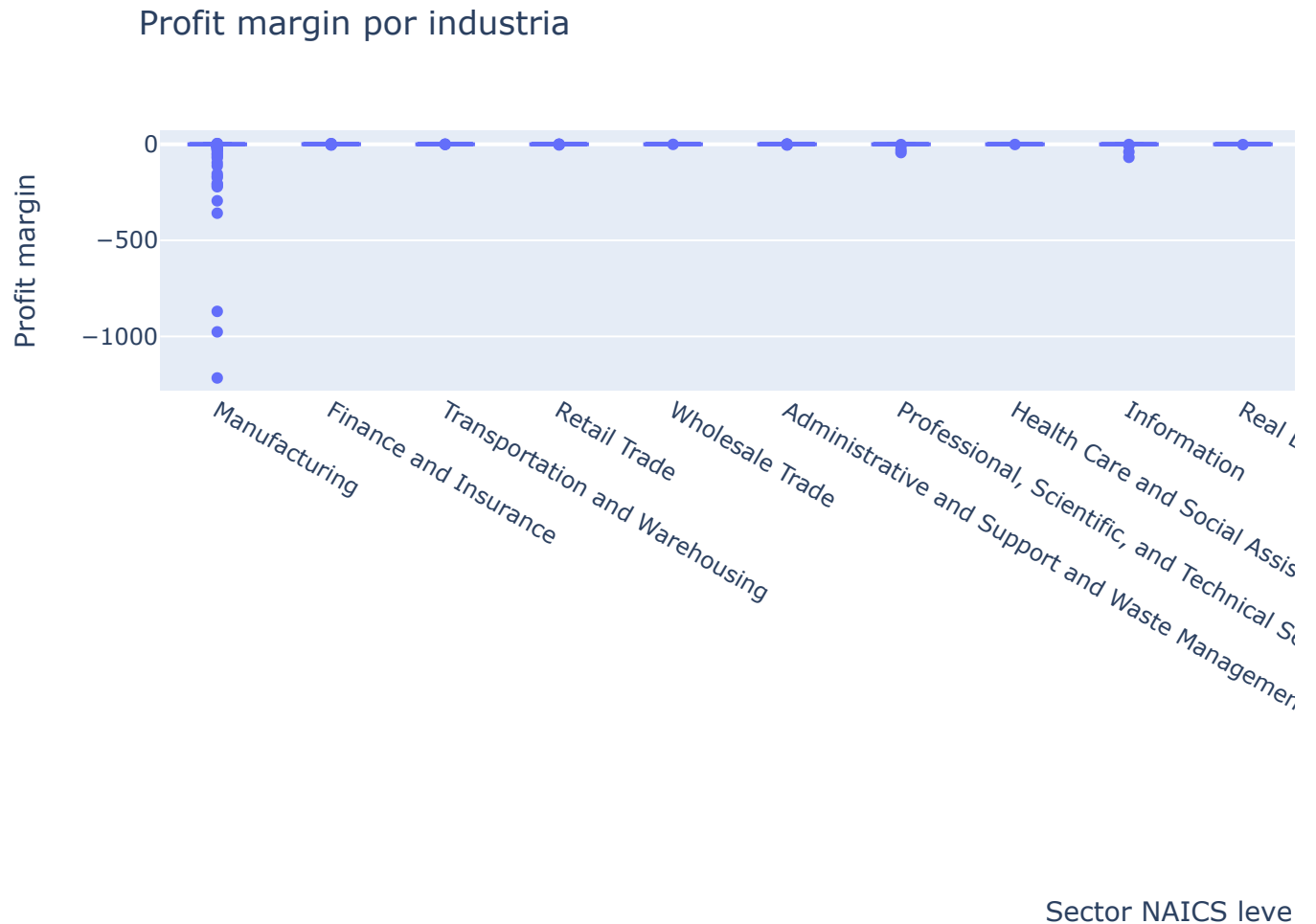
```
# Mostramos el gráfico para revisar los datos de Market value de cada industria y podemos dar
px.bar(us2022_quarter, x = 'Sector\nEconomatica', y = 'Market value', title = "Market values
```

Market values por industria



- For each industry (and for all industries), what can you say about profit margin of firms? show a) descriptive statistics of profit margin and b) plot(s) to illustrate how profit margin changes across industries.


```
px.box(us2022_quarter, x = 'Sector NAICS\nlevel 1', y = 'Profit margin', title = "Profit margi
```



- Which are the biggest 10 US firms in terms of market value and how far they are from the typical size of a US firm?

```
# Obtenemos las 10 mayores firms de US en terminos del Market value
us10_firms_markValue = us2022_quarter.sort_values("Market value", ascending = False).head(10)
```

```
us10_firms_markValue
```

	firm	q	revenue	cogs	sgae	otheropexp	extraincome	fi
809	AAPL	2022q2	82959000.0	47074000.0	12809000.0	0.0	-10000.0	
18173	AMZN	2022q2	121234000.0	66424000.0	51403000.0	90000.0	-5557000.0	425
78332	CVX	2022q2	68762000.0	46321000.0	4563000.0	1759000.0	-80000.0	129
1079	ABBV	2022q2	14583000.0	4170000.0	7290000.0	-172000.0	-1584000.0	532
34013	BAC	2022q2	14975000.0	2531000.0	0.0	0.0	-5552000.0	
68882	COST	2022q2	52596000.0	46355000.0	4450000.0	0.0	-19000.0	-36
2063	ABT	2022q2	11257000.0	4933000.0	3948000.0	0.0	82000.0	106
73022	CSCO	2022q2	12835000.0	4714000.0	4511000.0	0.0	166000.0	-25

- Which are the biggest 10 US firms in terms of book value and how far they are from the typical size of a US firm?

4583 ADBE 2022q2 4386000.0 539000.0 2318000.0 0.0 -9000.0 28

Obtenemos las 10 mayores firms de US en terminos del Book value

```
us10_firms_bookValue = us2022_quarter.sort_values("Book value", ascending = False).head(10)
```

✈️

```
us10_firms_bookValue
```

	firm	q	revenue	cogs	sgae	otheropexp	extraincome	fi
34013	BAC	2022q2	14975000.0	2531000.0	0.0	0.0	-5552000.0	
49680	C	2022q2	15630000.0	3666000.0	0.0	0.0	-6235000.0	
78332	CVX	2022q2	68762000.0	46321000.0	4563000.0	1759000.0	-80000.0	129
18173	AMZN	2022q2	121234000.0	66424000.0	51403000.0	90000.0	-5557000.0	425
64170	CMCSA	2022q2	30016000.0	20181000.0	3468000.0	0.0	-742000.0	968
78062	CVS	2022q2	80636000.0	76067000.0	0.0	0.0	33000.0	583
71492	CRM	2022q2	7411000.0	2045000.0	5346000.0	0.0	-49000.0	
809	AAPL	2022q2	82959000.0	47074000.0	12809000.0	0.0	-10000.0	
15473	AMD	2022q2	6550000.0	3522000.0	2508000.0	-6000.0	0.0	25
67230	COF	2022q2	7095000.0	578000.0	0.0	0.0	-3953000.0	

▼ 2.2.1.2 Considering the whole history of financial data for all firms:

- How can you measure firm profitability that can be used to compare performance among firms of different sizes? Select and justify at least 3 measures and show descriptive statistics

```
us2022_merged.corr()['totalassets']
```

```

revenue          0.333495
cogs              0.179930
sgae              0.209718
otheropexp       0.037647
extraincome      -0.682078
finexp           0.061794
incometax        0.234134
totalassets      1.000000
totalliabilities 0.998120
shortdebt        0.342227
longdebt         0.898848
stockholderequity 0.884258
adjprice         -0.002741
originalprice    0.046352
sharesoutstanding 0.646396
fiscalmonth      -0.000798
year             -0.032821
cto              -0.000816
N                0.012467
Book value       0.884259
Market value     0.315519
Operating profit  0.677114
Operating profit margin 0.004278
Net Income       0.701521
Profit margin    0.004407
Name: totalassets, dtype: float64

```

Para obtener el rendimiento entre firms podemos obtener el dato con una función de correlación

```
us2022_merged.corr()['Operating profit']
```

```

revenue          0.590156
cogs              0.435373
sgae              0.206746
otheropexp       0.081137
extraincome      -0.458654
finexp           0.136946
incometax        0.568639
totalassets      0.677114
totalliabilities 0.655998
shortdebt        0.183939
longdebt         0.664445
stockholderequity 0.750023
adjprice         -0.003311
originalprice    0.134275
sharesoutstanding 0.652116
fiscalmonth      -0.009582
year             -0.034102
cto              -0.003799
N                -0.003575
Book value       0.750018
Market value     0.658490
Operating profit  1.000000
Operating profit margin 0.006105
Net Income       0.951768

```

```
Profit margin          0.006232
Name: Operating profit, dtype: float64
```

```
# Aqui podemos observar que el revenue tiene una mejor correlación
us2022_merged.corr()['revenue']
```

```
revenue          1.000000
cogs              0.914765
sgae              0.582169
otheropexp        0.328325
extraincome       -0.080037
finexp            0.329195
incometax          0.441100
totalassets        0.333495
totalliabilities   0.297745
shortdebt          0.173205
longdebt           0.372473
stockholderequity  0.562425
adjprice           -0.005063
originalprice       0.288482
sharesoutstanding   0.434002
fiscalmonth         -0.001056
year               -0.046721
cto                0.002773
N                  -0.009127
Book value         0.562421
Market value       0.688222
Operating profit    0.590156
Operating profit margin 0.007746
Net Income         0.506595
Profit margin       0.007998
Name: revenue, dtype: float64
```

- Calculate and explain earnings per share deflated by price.

```
# Obtenemos las earnings per share
earnings_per_share = us2022_quarter["revenue"]
earnings_per_share.sort_values( ascending = True )
```

```
58320    -252794.0
36263         0.0
36353         0.0
37613         0.0
70772         0.0
...
68972         NaN
73562         NaN
74552         NaN
74912         NaN
78152         NaN
Name: revenue, Length: 910, dtype: float64
```

earnings_per_share

89	1607000.0
179	3644000.0
269	10900.0
359	13422000.0
449	44669.0
	...
81392	830543.0
81482	289409.0
81572	0.0
81662	572700.0
81752	853200.0

Name: revenue, Length: 910, dtype: float64

▼ 2.2.2 About statistical modeling

- You have to select a group of firms according to their general industry classification:
- Using your subset of firms that belong to your industry, which factors (variables) might be related to annual stock return one quarter in the future? Select at least 3 factors and briefly explain why you think might be related to stock returns.
- Design and run a multiple regression model to examine whether your selected factors and earnings per share deflated by price can explain/predict annual stock returns. You have to control for industry and firm size. To control for these variables you have to include them as extra independent variables in the model
- Interpret your model
- Adjustments to your model. If there is one or more independent variables (factors or control variables) that were not significant, drop them from your model. You have to run and interpret your final model.

Productos pagados de Colab - Cancela los contratos aquí

✓ 0 s se ejecutó 22:57

