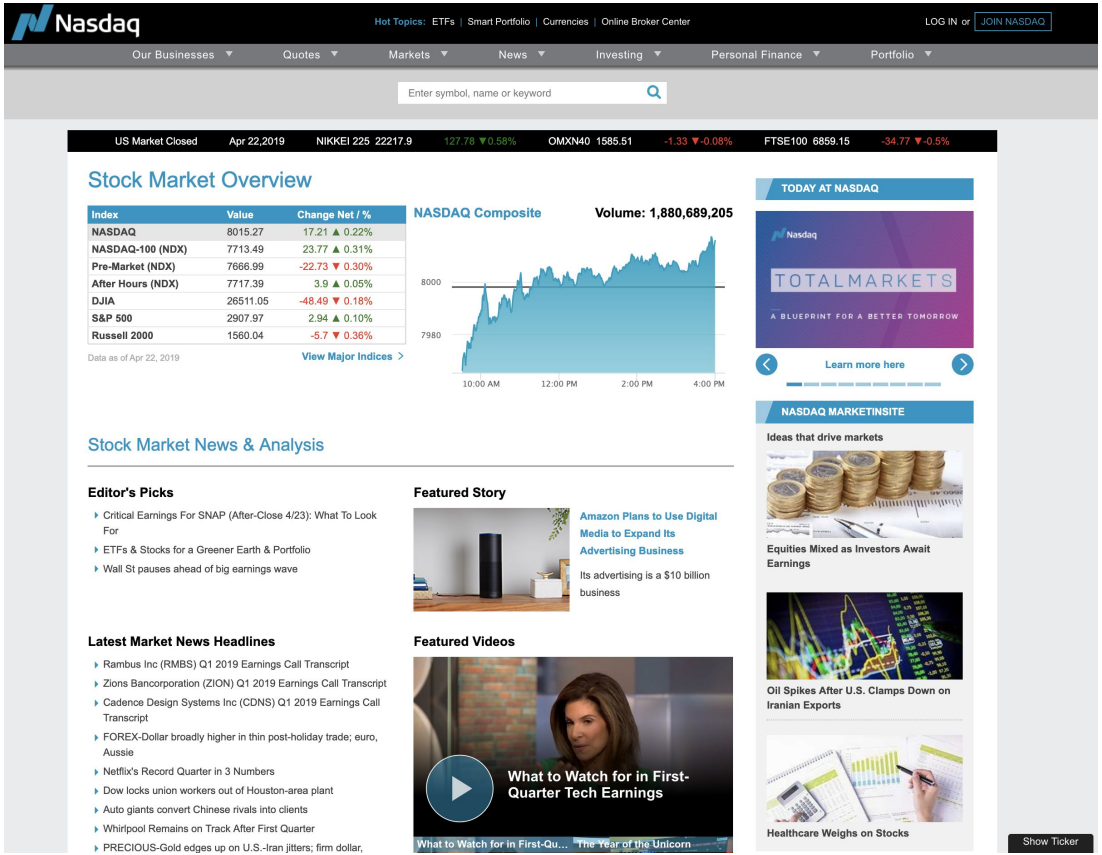
Webscraping Nasdaq.com With Scrapy

Cameron Wisdom

Background

- 2nd largest stock exchange in the world by Market Cap
- World's first electronic Stock Market
- Notable listings include: Microsoft, Oracle, Apple, Cisco, and Dell
- Helped modernize the IPO



Home > Company List > Companies By Industry > **NASDAQ**

Filters

Exchange
NASDAQ
NYSE
AMEX

Region
Africa
Asia
Australia and South Pacific
Central America and Caribbean
Europe
South America
Middle East
North America

Industry
Basic Industries
Capital Goods
Consumer Durables
Consumer Non-Dur.
Consumer Services
Energy
Finance
Healthcare
Miscellaneous
Public Utilities
Technology
Transportation

Company List Home | Name | Region | Industry

NASDAQ Companies

Index Page 1 of 15

Exchange: NASDAQ
Region: North America

All Countries | All NASDAQ Securities

Displaying 200 of 2980 results Download this list

Name	Symbol	Market Cap	Country	IPO Year	Subsector
Apple Inc.	AAPL	$964.42B	United States	1980	Computer Manufacturing
AAPL Stock Quote	AAPL Ratings	AAPL Stock Report			
Microsoft Corporation	MSFT	$949.51B	United States	1986	Computer Software: Prepackaged Software
MSFT Stock Quote	MSFT Ratings	MSFT Stock Report			
Amazon.com, Inc.	AMZN	$927.05B	United States	1997	Catalog/Specialty Distribution
AMZN Stock Quote	AMZN Ratings	AMZN Stock Report			
Alphabet Inc.	GOOGL	$871.6B	United States	n/a	Computer Software: Programming, Data Processing
GOOGL Stock Quote	GOOGL Ratings	GOOGL Stock Report			
Alphabet Inc.	GOOG	$868.18B	United States	2004	Computer Software: Programming, Data Processing
GOOG Stock Quote	GOOG Ratings	GOOG Stock Report			
Facebook, Inc.	FB	$518.12B	United States	2012	Computer Software: Programming, Data Processing
FB Stock Quote	FB Ratings	FB Stock Report			
Intel Corporation	INTC	$264.61B	United States	n/a	Semiconductors
INTC Stock Quote	INTC Ratings	INTC Stock Report			
Cisco Systems, Inc.	CSCO	$248.01B	United States	1990	Computer Communications Equipment
CSCO Stock Quote	CSCO Ratings	CSCO Stock Report			

This lovely table is dynamic and changes throughout each day during trading hours.

Interestingly enough, it also omits cells if the information is not available. yay!

summary page

Home > Quotes > **AAPL**

Apple Inc. Common Stock (AAPL) Quote & Summary Data

$204.53* **0.67** **0.33%**

*Delayed - data as of Apr. 22, 2019 - Find a broker to begin trading AAPL now

Exchange:NASDAQ
Industry: Technology
Community Rating: Bullish

Edit Symbol List
Symbol Lookup

USIG | QQQ | RMR | STLD | BOKF | NFE | FEYE | CELG | OPES | MLZ | ADBE | AMZN | ACST
AAON | TURN | PIH | FLWS | QFIN | TPNL | TWOU | SRCE | JFKKU | YI | AAPL

Save Stocks Refresh

SYMBOL LIST VIEWS
FlashQuotes
InfoQuotes

STOCK DETAILS
Summary Quote
Real-Time Quote
After Hours Quote
Pre-market Quote
Historical Quote
Option Chain

CHARTS
Basic Chart
Interactive Chart

COMPANY NEWS

Key Stock Data

Best Bid / Ask	N/A / N/A	P/E Ratio	16.82
1 Year Target	217.5	Forward P/E (1y)	17.87
Today's High / Low	$ 204.94 / $ 202.34	Earnings Per Share (EPS)	$ 12.16
Share Volume	19,439,545	Annualized Dividend	$ 2.92
50 Day Avg. Daily Volume	26,720,185	Ex Dividend Date	2/8/2019
Previous Close	$ 203.86	Dividend Payment Date	2/14/2019
52 Week High / Low	$ 233.47 / $ 142	Current Yield	1.43 %
Market Cap	964,416,218,400	Beta	1.02

Intraday Chart

The right tools for the job

- Scrapy
- Pycharm
- Github Desktop
- ROBOTSTXT_OBEY = False
- NordVPN
- Caffeine
- Selector Gadget
- Jupyter notebook
- “Hey Siri, how do I export graphs from jupyter notebook? Asking for a friend”

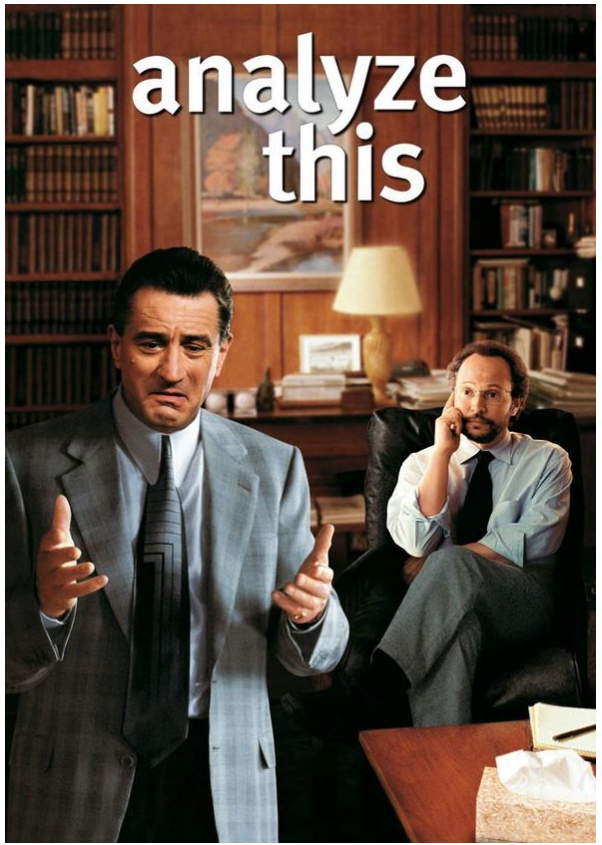
NordVPN

Desktop

PC

```
In [6]: df_clean.sort_values('marketCap', ascending=False)
```

symbol	industry	marketCap	price	previousClose	todaysHigh	todaysLow
AAPL	Technology	964416218400	204.53	203.86	204.940	202.3400
MSFT	Technology	949513136077	123.76	123.37	124.000	122.5700
AMZN	Consumer Services	927052126326	1887.31	1,861.69	1.000	888.4200
GOOGL	Technology	871596895849	1253.76	1,241.47	1.000	254.3405
GOOG	Technology	868176578781	1248.84	1,236.37	1.000	249.0900
FB	Technology	518122757477	181.44	178.28	181.665	178.2500
INTC	Technology	264611677758	58.82	58.49	58.830	57.8800
CSCO	Technology	248010241519	56.34	56.40	56.600	56.1000
CMCSA	Consumer Services	190090425924	42.00	42.43	42.405	41.8550



That's... Interesting

```
In [6]: df_clean.sort_values('marketCap', ascending=False)
```

symbol	industry	marketCap	price	previousClose	todaysHigh	todaysLow
AAPL	Technology	964416218400	204.53	203.86	204.940	202.3400
MSFT	Technology	949513136077	123.76	123.37	124.000	122.5700
AMZN	Consumer Services	927052126326	1887.31	1,861.69	1.000	888.4200
GOOGL	Technology	871596895849	1253.76	1,241.47	1.000	254.3405
GOOG	Technology	868176578781	1248.84	1,236.37	1.000	249.0900
FB	Technology	518122757477	181.44	178.28	181.665	178.2500
INTC	Technology	264611677758	58.82	58.49	58.830	57.8800
CSCO	Technology	248010241519	56.34	56.40	56.600	56.1000
CMCSA	Consumer Services	190090425924	42.00	42.43	42.405	41.8550

In [14]:
```
group.mean()
```

Out[14]:

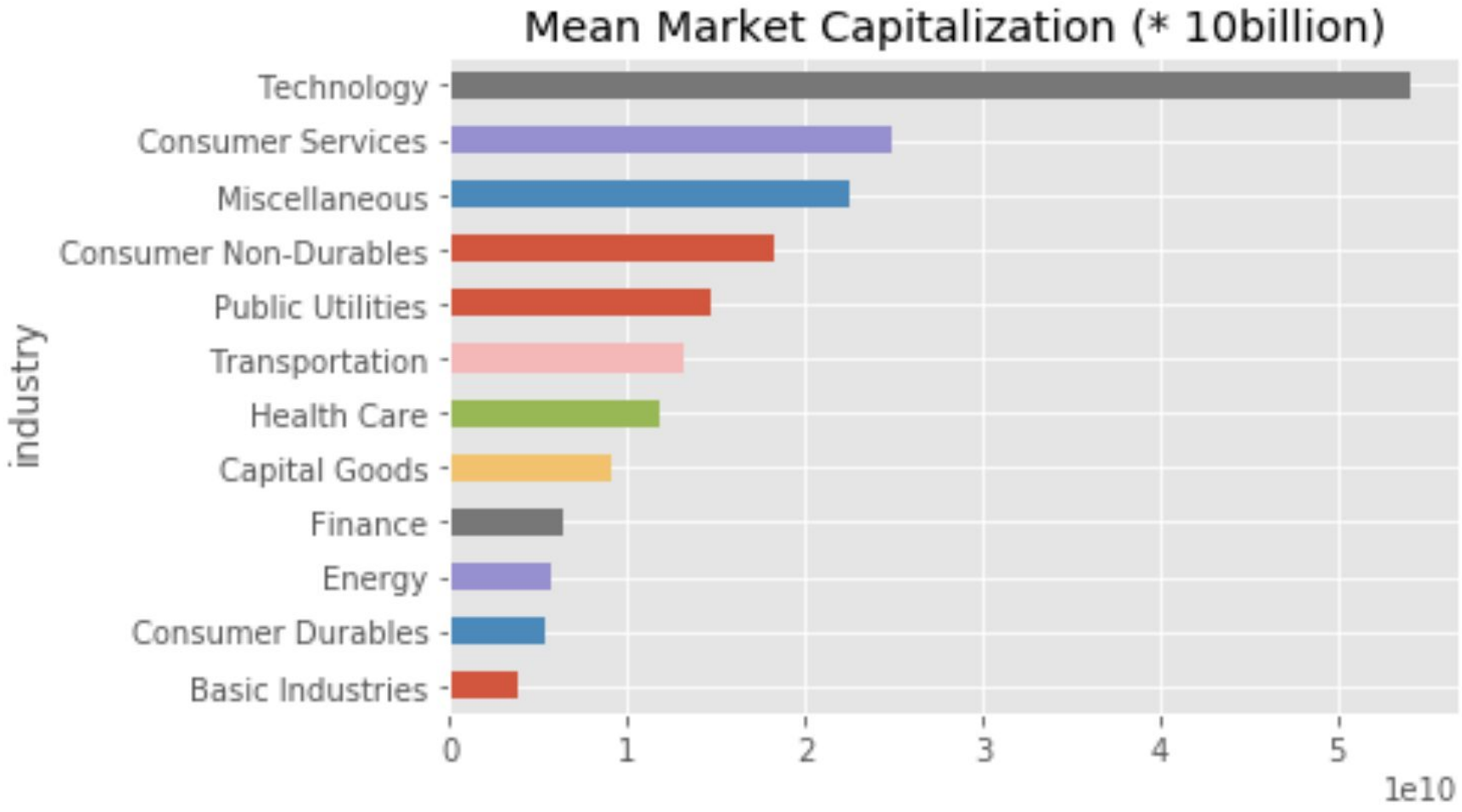
industry	marketCap	price	todaysHigh	todaysLow	fiftytwoHigh	fiftytwoLow	earnings_per_share	currentYield	beta	sharesOutstanding
Basic Industries	3.761036e+09	69.185556	70.638889	68.847500	84.066111	52.333344	2.337778	1.283333	0.912222	8.245970e+07
Capital Goods	9.055919e+09	84.813214	85.817111	83.930836	101.963943	58.364286	2.767500	0.844286	1.180000	1.100325e+08
Consumer Durables	5.314096e+09	93.416667	94.403333	92.326844	101.027111	65.512778	3.576667	1.171111	0.797778	8.944829e+07
Consumer Non-Durables	1.832956e+10	85.802083	86.977592	85.346183	97.115729	57.967125	0.952917	1.266250	0.645000	2.837596e+08
Consumer Services	2.485852e+10	97.546107	80.289259	88.070543	86.492536	56.725920	2.321250	1.585729	0.699896	2.974208e+08
Energy	5.763832e+09	56.311429	56.577143	55.010729	66.592857	40.468571	2.391429	3.108571	0.681429	1.141373e+08
Finance	6.315856e+09	69.354881	70.129214	68.786404	80.850967	52.168068	4.183571	2.478929	0.763571	1.436143e+08
Health Care	1.190516e+10	90.089014	91.374204	88.454168	120.312254	68.541017	0.920000	0.287465	1.173239	1.344428e+08
Miscellaneous	2.249201e+10	218.074286	86.886429	146.074150	98.850000	73.248343	7.487857	0.327143	1.123571	2.312697e+08
Public Utilities	1.463406e+10	45.182500	45.375000	44.857500	47.396175	34.150625	0.541250	2.871250	0.245000	2.631069e+08
Technology	5.420462e+10	104.790000	84.618118	87.062749	95.656540	61.612581	2.776198	0.567273	1.131736	4.112223e+08
Transportation	1.311387e+10	80.390000	80.836915	79.605977	92.710000	63.255669	4.855385	0.987692	0.917692	1.944754e+08

```
group.agg(['count', 'sum', 'min', 'max', 'mean', 'std'])['marketCap']
```

industry	count	sum	min	max	mean	std
Basic Industries	9	33849321579	1886513907	7196221880	3.761036e+09	1.951039e+09
Capital Goods	28	253565731354	1853150250	46876535166	9.055919e+09	1.184787e+10
Consumer Durables	9	47826860242	2196061140	14831101831	5.314096e+09	3.867154e+09
Consumer Non-Durables	24	439909356066	1999148410	177557403815	1.832956e+10	3.766769e+10
Consumer Services	96	2386418356430	1877523694	927052126326	2.485852e+10	9.760059e+10
Energy	7	40346821948	2278058593	18279225246	5.763832e+09	5.692920e+09
Finance	84	530531937126	1817316436	62543095876	6.315856e+09	8.185923e+09
Health Care	71	845266632315	1844836201	108628832833	1.190516e+10	1.964403e+10
Miscellaneous	14	314888118301	2087202245	125201900898	2.249201e+10	3.648870e+10
Public Utilities	8	117072480495	1991650262	62678326326	1.463406e+10	2.134160e+10
Technology	121	6558758780551	1839192081	964416218400	5.420462e+10	1.705887e+11
Transportation	13	170480267408	1936184855	63778280154	1.311387e+10	1.652140e+10

$24.1 billion

Mean Market Capitalization value of the top 500 listings on Nasdaq

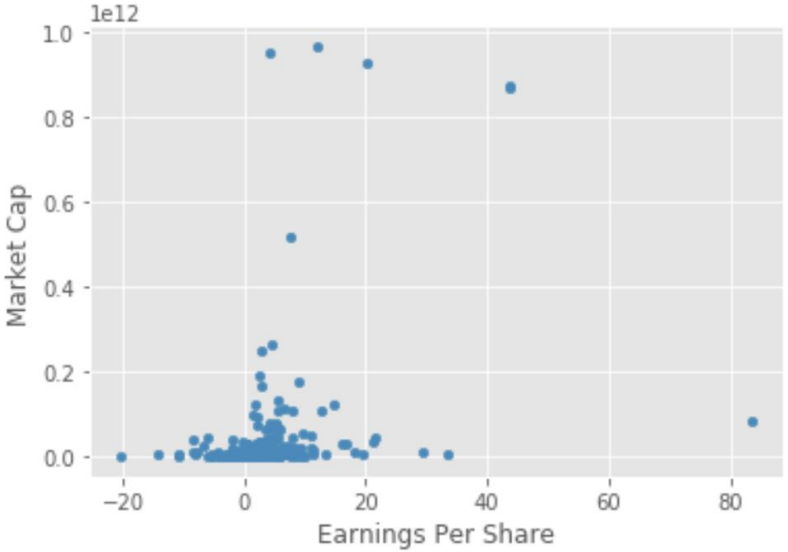

Mean Market Capitalization (* 10billion)
industry
Technology
Consumer Services
Miscellaneous
Consumer Non-Durables
Public Utilities
Transportation
Health Care
Capital Goods
Finance
Energy
Consumer Durables
Basic Industries
0
1
2
3
4
5
1e10

```
corr = df_clean.corr()
corr.style.background_gradient(cmap='coolwarm').set_precision(2)
```

	marketCap	price	todaysHigh	todaysLow	fiftytwoHigh	fiftytwoLow	earnings_per_share	currentYield	beta	sharesOutstanding
marketCap	1	0.57	0.081	0.36	0.071	0.27	0.39	-0.032	0.14	0.61
price	0.57	1	0.41	0.86	0.4	0.63	0.75	-0.15	0.14	0.012
todaysHigh	0.081	0.41	1	0.78	0.98	0.91	0.31	-0.21	0.13	-0.028
todaysLow	0.36	0.86	0.78	1	0.76	0.84	0.63	-0.2	0.14	-0.012
fiftytwoHigh	0.071	0.4	0.98	0.76	1	0.91	0.3	-0.22	0.18	-0.041
fiftytwoLow	0.27	0.63	0.91	0.84	0.91	1	0.55	-0.19	0.14	-0.01
earnings_per_share	0.39	0.75	0.31	0.63	0.3	0.55	1	-0.0056	0.048	0.052
currentYield	-0.032	-0.15	-0.21	-0.2	-0.22	-0.19	-0.0056	1	-0.28	0.075
beta	0.14	0.14	0.13	0.14	0.18	0.14	0.048	-0.28	1	0.095
sharesOutstanding	0.61	0.012	-0.028	-0.012	-0.041	-0.01	0.052	0.075	0.095	1

Colors indicate correlation relationship (+ or -), darker hues indicate correlation strength

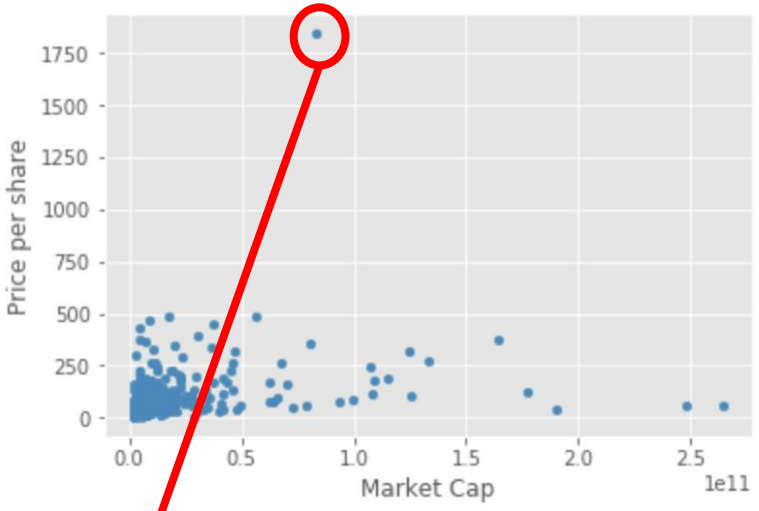
WARNING: Outliers will skew your data




```
df_NoOutliers.plot(kind='scatter',
x='earnings_per_share', y='marketCap')
plt.xlabel('Earnings Per Share')
plt.ylabel('Market Cap')
```

```
Text(0, 0.5, 'Market Cap')
```

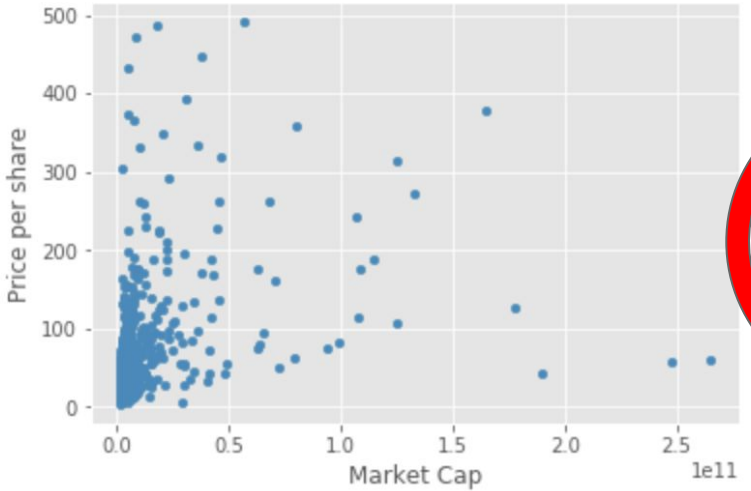
Outliers 2: the skewed dataset strikes back

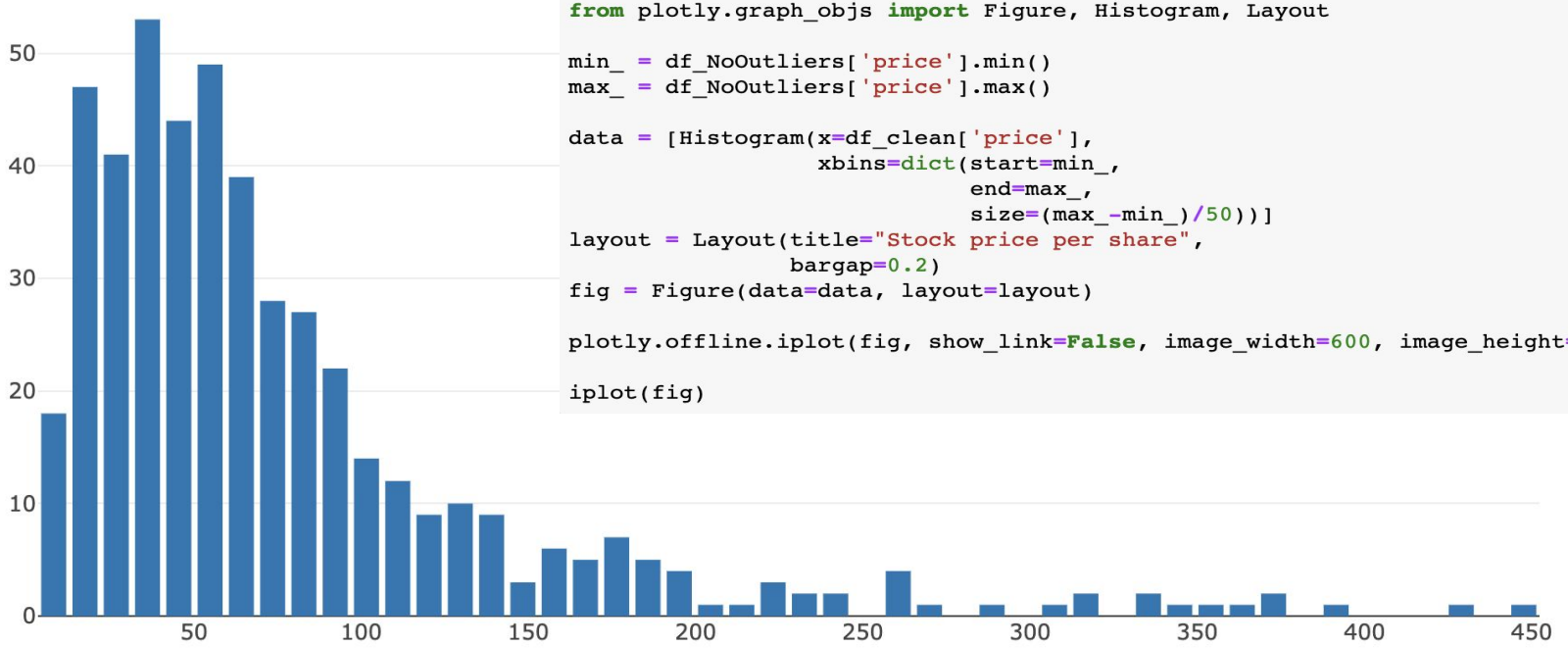


symbol	industry	marketCap	price	previousClose	to
BKNG	Miscellaneous	83011117063	1844.17	1,844.31	

```
df_NoOutliers2.plot(kind='scatter',x='marketCap', y='price')
plt.xlabel('Market Cap')
plt.ylabel('Price per share')
```

```
Text(0, 0.5, 'Price per share')
```






```
from plotly.graph_objs import Figure, Histogram, Layout

min_ = df_NoOutliers['price'].min()
max_ = df_NoOutliers['price'].max()

data = [Histogram(x=df_clean['price'],
                  xbins=dict(start=min_,
                             end=max_,
                             size=(max_-min_)/50))]
layout = Layout(title="Stock price per share",
                bargap=0.2)
fig = Figure(data=data, layout=layout)

plotly.offline.iplot(fig, show_link=False, image_width=600, image_height=400)

iplot(fig)
```

"Look at that neato right-skewed distribution!" - Stats Professor probably

fin

Any questions?