In [25]:
```
import sqlite3
import pandas as pd

csvfile = "resources/phase3/simfin/income.csv"

df = pd.read_csv(csvfile, sep=';')
df.columns = df.columns.str.replace(' ', '')

import os
os.chdir('/home/ec2-user/SageMaker/')

db_file = "simfim.sqlite"
conn = sqlite3.connect(db_file)
table_name = "tblSimFin"

df.to_sql(table_name, conn, if_exists='replace', index=False)
```

In [26]:
```
querySimFinCount = """
select count(*) from tblSimFin
"""

dfSimFinCount = pd.read_sql_query(querySimFinCount, conn)

print(dfSimFinCount)
```

```
   count(*)
0     45312
```

In [27]:
```
index = df.index
number_of_rows = len(index)
print(number_of_rows)
```

```
45312
```

In [34]:
```
queryData = """
select * from tblSimFin
"""

dftblSimFin = pd.read_sql_query(queryData, conn)
dftblSimFin.head(100)
```

Out[34]:

	Ticker	SimFinId	Currency	FiscalYear	FiscalPeriod	ReportDate	PublishDate	Shares(Basic)	Shares(Diluted)	Revenu
0	A	45846	USD	2011	Q2	2011-04-30	2011-06-07	346250000.0	354500000.0	6.156000e+0
1	A	45846	USD	2011	Q3	2011-07-31	2011-09-07	346500000.0	355750000.0	6.463000e+0
2	A	45846	USD	2011	Q4	2011-10-31	2011-12-16	347000000.0	355000000.0	6.615000e+0
3	A	45846	USD	2012	Q1	2012-01-31	2012-03-05	347250000.0	354250000.0	6.731000e+0
4	A	45846	USD	2012	Q2	2012-04-30	2012-06-04	347500000.0	354000000.0	6.787000e+0
5	A	45846	USD	2012	Q3	2012-07-31	2012-09-05	347500000.0	353000000.0	6.819000e+0
6	A	45846	USD	2012	Q4	2012-10-31	2012-12-20	348000000.0	353000000.0	6.858000e+0
7	A	45846	USD	2013	Q1	2013-01-31	2013-03-06	347750000.0	353000000.0	6.903000e+0
8	A	45846	USD	2013	Q2	2013-04-30	2013-06-03	347000000.0	351750000.0	6.902000e+0
9	A	45846	USD	2013	Q3	2013-07-31	2013-09-06	344750000.0	349250000.0	6.831000e+0
10	A	45846	USD	2013	Q4	2013-10-31	2013-12-19	341000000.0	345000000.0	3.894000e+0
11	A	45846	USD	2014	Q1	2014-01-31	2014-03-05	337500000.0	341500000.0	3.222000e+0
12	A	45846	USD	2014	Q2	2014-04-30	2014-06-04	334500000.0	338500000.0	2.478000e+0
13	A	45846	USD	2014	Q3	2014-07-31	2014-09-02	333250000.0	337250000.0	1.835000e+0
14	A	45846	USD	2014	Q4	2014-10-31	2014-12-22	333000000.0	338000000.0	4.048000e+0
15	A	45846	USD	2015	Q1	2015-01-31	2015-03-10	333750000.0	338000000.0	4.066000e+0
16	A	45846	USD	2015	Q2	2015-04-30	2015-06-05	334000000.0	338000000.0	4.041000e+0
17	A	45846	USD	2015	Q3	2015-07-31	2015-09-02	333500000.0	337000000.0	4.046000e+0
18	A	45846	USD	2015	Q4	2015-10-31	2015-12-21	333000000.0	335000000.0	4.038000e+0
19	A	45846	USD	2016	Q1	2016-01-31	2016-03-08	331250000.0	333500000.0	4.040000e+0
20	A	45846	USD	2016	Q2	2016-04-30	2016-06-07	329250000.0	331250000.0	4.096000e+0
21	A	45846	USD	2016	Q3	2016-07-31	2016-09-07	327500000.0	329750000.0	4.126000e+0
22	A	45846	USD	2016	Q4	2016-10-31	2016-12-20	326000000.0	329000000.0	4.202000e+0
23	A	45846	USD	2017	Q1	2017-01-31	2017-03-08	324250000.0	327500000.0	4.241000e+0
24	A	45846	USD	2017	Q2	2017-04-30	2017-06-06	323000000.0	326750000.0	4.324000e+0
25	A	45846	USD	2017	Q3	2017-07-31	2017-09-06	322000000.0	326250000.0	4.394000e+0
26	A	45846	USD	2017	Q4	2017-10-31	2017-12-21	322000000.0	326000000.0	4.472000e+0
27	A	45846	USD	2018	Q1	2018-01-31	2018-03-06	322250000.0	325250000.0	4.616000e+0
28	A	45846	USD	2018	Q2	2018-04-30	2018-05-31	322500000.0	325500000.0	4.720000e+0
29	A	45846	USD	2018	Q3	2018-07-31	2018-08-30	322250000.0	325000000.0	4.809000e+0
...	...	...	...	...	...	...	...	...	...	.
70	AAL	68568	USD	2018	Q1	2018-03-31	2018-04-26	481262750.0	483392250.0	4.298400e+1
71	AAL	68568	USD	2018	Q2	2018-06-30	2018-07-26	474441500.0	476305500.0	4.352200e+1
72	AAL	68568	USD	2018	Q3	2018-09-30	2018-10-25	468380000.0	470026000.0	4.420300e+1
73	AAL	68568	USD	2018	Q4	2018-12-31	2019-02-25	464236000.0	465660000.0	4.454100e+1
74	AAL	68568	USD	2019	Q1	2019-03-31	2019-04-26	459149500.0	460367750.0	4.472400e+1
75	AAN	441241	USD	2011	Q1	2011-03-31	2011-05-04	80866500.0	81839000.0	1.913734e+0
76	AAN	441241	USD	2011	Q2	2011-06-30	2011-08-04	80418250.0	81486500.0	1.951435e+0
77	AAN	441241	USD	2011	Q3	2011-09-30	2011-11-07	79434000.0	80647750.0	1.984016e+0
78	AAN	441241	USD	2011	Q4	2011-12-31	2012-02-29	78101000.0	79339000.0	2.012578e+0
79	AAN	441241	USD	2012	Q1	2012-03-31	2012-05-08	77071500.0	78323000.0	2.065909e+0
80	AAN	441241	USD	2012	Q2	2012-06-30	2012-08-08	76131750.0	77343000.0	2.121828e+0
81	AAN	441241	USD	2012	Q3	2012-09-30	2012-11-07	75805500.0	76987500.0	2.166607e+0
82	AAN	441241	USD	2012	Q4	2012-12-31	2013-02-22	75820000.0	76826000.0	2.212827e+0
83	AAN	441241	USD	2013	Q1	2013-03-31	2013-05-03	75767250.0	76711500.0	2.219841e+0
84	AAN	441241	USD	2013	Q2	2013-06-30	2013-08-02	75760750.0	76614000.0	2.231767e+0
85	AAN	441241	USD	2013	Q3	2013-09-30	2013-11-01	75829000.0	76553500.0	2.239481e+0
86	AAN	441241	USD	2013	Q4	2013-12-31	2014-02-24	75747000.0	76390000.0	2.234631e+0
87	AAN	441241	USD	2014	Q1	2014-03-31	2014-05-09	74923750.0	75467500.0	2.227044e+0
88	AAN	441241	USD	2014	Q2	2014-06-30	2014-08-08	74010000.0	74469750.0	2.338989e+0
89	AAN	441241	USD	2014	Q3	2014-09-30	2014-11-05	73069750.0	73465750.0	2.500183e+0
90	AAN	441241	USD	2014	Q4	2014-12-31	2015-03-02	72384000.0	72723000.0	2.695033e+0
91	AAN	441241	USD	2015	Q1	2015-03-31	2015-05-07	72396500.0	72715750.0	2.931424e+0
92	AAN	441241	USD	2015	Q2	2015-06-30	2015-08-06	72478000.0	72807500.0	3.037983e+0
93	AAN	441241	USD	2015	Q3	2015-09-30	2015-11-09	72539500.0	72911500.0	3.107259e+0
94	AAN	441241	USD	2015	Q4	2015-12-31	2016-02-29	72568000.0	73043000.0	3.179756e+0
95	AAN	441241	USD	2016	Q1	2016-03-31	2016-05-06	72597250.0	73133500.0	3.212369e+0
96	AAN	441241	USD	2016	Q2	2016-06-30	2016-08-04	72644500.0	73212000.0	3.232673e+0
97	AAN	441241	USD	2016	Q3	2016-09-30	2016-11-02	72650000.0	73242750.0	3.233961e+0
98	AAN	441241	USD	2016	Q4	2016-12-31	2017-02-24	72354000.0	73013000.0	3.207716e+0
99	AAN	441241	USD	2017	Q1	2017-03-31	2017-05-03	72025000.0	72805250.0	3.197843e+0

100 rows × 27 columns

In [32]:
```
#What is the average revenue of all the years past 2013?
queryAvgRev = """
select avg(Revenue) from tblSimFin
where FiscalYear > 2013
"""

resultAvgRev = pd.read_sql_query(queryAvgRev,conn)
print("The average revenue since 2014 has been: \n\n %s" % resultAvgRev)
```

```
The average revenue since 2014 has been:

    avg(Revenue)
0  7.209511e+09
```

In [33]:
```
#Which single day had the highest net income, and what was that date?
queryHighIncome = """
select ReportDate, max(NetIncome) from tblSimFin
"""

resultHighIncome = pd.read_sql_query(queryHighIncome, conn)
print("The highest daily income has been: \n\n %s" % resultHighIncome)
```

```
The highest daily income has been:

    ReportDate  max(NetIncome)
0  2018-09-30     61964000000
```

In [43]:
```
#Which quarter for what Ticker had the lowest revenue?
queryLowQuarterlyRev = """
select FiscalYear, FiscalPeriod, Ticker, min(Revenue) from tblSimFin
"""

resultLowQuarterlyRev = pd.read_sql_query(queryLowQuarterlyRev, conn)
print("The Lowest quarterly revenue was: \n\n %s" % resultLowQuarterlyRev)
```

```
The Lowest quarterly revenue was:

    FiscalYear FiscalPeriod Ticker  min(Revenue)
0         2015           Q4   AVXL           0.0
```