```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
from sklearn.linear model import LinearRegression as lr
import statsmodels.api as sm
import statsmodels.formula.api as smf
salary = pd.read csv('CEO Salary Data.csv')
salary
                     CONAME
                               SALARY
                                        GVKEY YEAR PCEO \
     ADVANCED MICRO DEVICES
                              924.997
                                         1161
                                               2017 CEO
0
1
     SKYWORKS SOLUTIONS INC
                              850.000
                                         1327
                                               2017
                                                     CE0
2
         ANALOG DEVICES INC
                              878.000
                                         1632
                                               2017
                                                     CE0
3
                  APPLE INC
                             3057.692
                                         1690
                                               2017
                                                     CE0
4
      APPLIED MATERIALS INC
                             1000.000
                                         1704
                                               2017 CE0
                                                . . .
               FORTINET INC
                                               2017
89
                              450.882
                                       183377
                                                     CE0
90
           SPS COMMERCE INC
                              495.000
                                       183974
                                               2017
                                                     CE0
91
                              492.299
              MAXLINEAR INC
                                       184551
                                               2017
                                                     CE0
92
                              700.027
              BLACKBAUD INC
                                       260893
                                               2017
                                                     CE0
93 ULTRA CLEAN HOLDINGS INC
                              406.923 264416 2017 CEO
                                    TITLE
                                                 EXEC LNAME
EXEC FNAME \
                              CEO & Chair
                                                  Su, Ph.D.
0
Lisa
   CEO, President & Chairman of the Board
                                                    Griffin
Liam
                President, CEO & Director
                                                      Roche
Vincent
                           CEO & Director
                                                       Cook
Timothy
      President, CEO & Executive Director
                                                  Dickerson
Gary
. .
                                                                   . .
                  Founder, Chairman & CEO
89
                                                        Xie
Ken
                           CEO & Director
                                                      Black
90
Archie
    Co-Founder, Chairman, CEO & President Seendripu, Ph.D.
Kishore
                President, CEO & Director
                                                    Gianoni
92
Michael
                           CEO & Director Scholhamer
93
James
   GENDER
```

0

**FEMALE** 

```
MALE
1
2
      MALE
3
      MALE
4
      MALE
       . . .
89
      MALE
90
      MALE
91
      MALE
92
      MALE
93
      MALE
[94 rows x 9 columns]
company = pd.read csv('Company Data.csv')
company = company.drop(columns=["conm"])
company
                                PCC
      GVKEY
              fyear TICKER
                                               MV
0
       1161
               2017
                       AMD
                              10.28
                                       9940,7600
1
       1327
               2017
                              94.95
                                      18657.8900
                      SWKS
2
       1562
               2017
                     AMSWA
                              11.63
                                         390.5836
3
       1632
               2017
                              89.03
                                      33656.4668
                       ADI
4
       1690
                             169.23
               2017
                      AAPL
                                     790050.0981
153
     184551
               2017
                              26.42
                                       1780.7080
                       MXL
154
     187740
               2017
                      MTSI
                              32.54
                                       2866.4602
155
     187959
                      BCOV
                               7.10
                                        247.0658
               2017
                              94.49
156
     260893
               2017
                      BLKB
                                       4542.7012
                              23.09
                                        777.3249
157
     264416
               2017
                      UCTT
[158 rows x 5 columns]
market = pd.read csv('Nasdaq.csv')
market = market[["TICKER", "Sector", "Industry"]]
market
    TICKER
                 Sector
                                                           Industry
            Technology
0
      AAPL
                                            Computer Manufacturing
1
      ACCD
            Technology
                                                 Interactive Media
2
      ACIW
            Technology
                                                      EDP Services
3
                                  Industrial Machinery/Components
      ACLS
            Technology
4
      ADBE
            Technology
                         Computer Software: Prepackaged Software
       . . .
. .
228
       XRX
            Technology
                                                      EDP Services
229
      ZBRA
            Technology
                                    Computer peripheral equipment
230
        ΖI
                         Computer Software: Prepackaged Software
            Technology
231
        ZM
            Technology
                         Computer Software: Prepackaged Software
                                                      EDP Services
232
        ZS
            Technology
[233 rows x 3 columns]
```

```
joined = pd.merge(salary, # the "left" dataframe
                       company, # the "right" dataframe
                       how = 'left', # which observations to keep?
Here we are specifying that we keep the "left" dataset
                       on = 'GVKEY' # the join key
joined
                     CONAME
                              SALARY
                                       GVKEY
                                              YEAR PCEO \
                              924.997
0
     ADVANCED MICRO DEVICES
                                        1161
                                              2017 CEO
     SKYWORKS SOLUTIONS INC
                             850.000
                                        1327
                                              2017 CE0
1
2
         ANALOG DEVICES INC
                             878.000
                                              2017 CEO
                                        1632
3
                  APPLE INC
                             3057.692
                                        1690
                                              2017 CE0
4
      APPLIED MATERIALS INC
                             1000.000
                                        1704
                                              2017 CE0
           SPS COMMERCE INC
90
                             450.882
                                      183377
                                              2017
                                                   CE0
                             495.000
91
                                      183974
                                              2017 CE0
92
                             492.299 184551
                                              2017
                                                    CE0
                             700.027
93
              BLACKBAUD INC
                                      260893
                                              2017
                                                    CE0
94 ULTRA CLEAN HOLDINGS INC
                             406.923 264416 2017 CEO
                                   TITLE
                                                EXEC LNAME
EXEC FNAME \
                             CEO & Chair
                                               Su, Ph.D.
Lisa
   CEO, President & Chairman of the Board
                                                 Griffin
Liam
                President, CEO & Director
                                                     Roche
Vincent
                           CEO & Director
3
                                                      Cook
Timothy
      President, CEO & Executive Director
                                               Dickerson
Gary
. .
                  Founder, Chairman & CEO
90
                                                      Xie
Ken
                           CEO & Director
91
                                                     Black
Archie
    Co-Founder, Chairman, CEO & President Seendripu, Ph.D.
92
Kishore
                President, CEO & Director
93
                                                   Gianoni
Michael
94
                           CEO & Director Scholhamer
James
                           PCC
                                         MV
   GENDER
          fyear TICKER
            2017
                          10.28
                                 9940.7600
   FEMALE
                    AMD
0
     MALE
            2017
                   SWKS
                          94.95
                                 18657.8900
1
2
     MALE
            2017
                   ADI
                          89.03
                                 33656,4668
```

```
AAPL
                          169.23 790050.0981
3
      MALE
             2017
4
      MALE 2017 AMAT 51.12 59815.8000
                     . . .
              . . .
                    FTNT 43.69 7335.1141
SPSC 48.59 832.2009
90
      MALE
             2017
91
             2017
      MALE
                    MXL 26.42 1780.7080
BLKB 94.49 4542.7012
92
      MALE
             2017
93
             2017
      MALE
             2017 UCTT 23.09 777.3249
94
      MALE
[95 rows x 13 columns]
joined = pd.merge(joined, # the "left" dataframe
                        market,  # the "right" dataframe
how = 'left',  # which observations to keep?
Here we are specifying that we keep the "left" dataset
                        on = 'TICKER' # the join key
        )
joined
                               SALARY GVKEY YEAR PCEO \
                      CONAME
                                          1161
0
      ADVANCED MICRO DEVICES 924.997
                                                 2017 CEO
      SKYWORKS SOLUTIONS INC 850.000
ANALOG DEVICES INC 878.000
1
                                           1327
                                                 2017 CE0
                                                 2017 CEO
2
                                         1632
3
                   APPLE INC 3057.692
                                           1690
                                                 2017 CE0
     APPLIED MATERIALS INC 1000.000
                                           1704 2017 CEO
4
           FORTINET INC 450.882 183377

SPS COMMERCE INC 495.000 183974

MAXLINEAR INC 492.299 184551

BLACKBAUD INC 700.027 260893
                                                  2017 CEO
90
91
                                                 2017 CE0
92
                                                 2017 CE0
93
                                                 2017 CE0
94 ULTRA CLEAN HOLDINGS INC
                               406.923 264416 2017 CEO
                                      TITLE EXEC LNAME
EXEC FNAME \
                               CEO & Chair Su, Ph.D.
Lisa
1 CEO. President & Chairman of the Board
                                                    Griffin
Liam
                President, CEO & Director
                                                       Roche
Vincent
                            CEO & Director
                                                         Cook
Timothy
4 President, CEO & Executive Director Dickerson
Gary
. .
90
                  Founder, Chairman & CEO
                                                         Xie
Ken
                                               Black
                            CEO & Director
91
Archie
```

```
92
     Co-Founder, Chairman, CEO & President Seendripu, Ph.D.
Kishore
                  President, CEO & Director
93
                                                        Gianoni
Michael
94
                             CEO & Director
                                                     Scholhamer
James
            fyear TICKER
                              PCC
                                             MV
    GENDER
                                                      Sector \
                            10.28
    FEMALE
             2017
                      AMD
                                      9940.7600
                                                 Technology
0
             2017
                            94.95
1
      MALE
                     SWKS
                                     18657.8900
                                                 Technology
      MALE
2
             2017
                            89.03
                                     33656.4668
                                                 Technology
                      ADI
3
             2017
                           169.23
                                                 Technology
      MALE
                     AAPL
                                    790050.0981
4
      MALE
             2017
                     AMAT
                            51.12
                                     59815.8000
                                                 Technology
       . . .
              . . .
                      . . .
                               . . .
      MALE
                            43.69
                                      7335.1141
90
             2017
                     FTNT
                                                 Technology
91
                            48.59
                                       832.2009
      MALE
             2017
                     SPSC
                                                 Technology
92
                     MXL
                            26.42
                                      1780.7080
                                                 Technology
      MALE
             2017
93
      MALE
             2017
                     BLKB
                            94.49
                                      4542.7012
                                                 Technology
94
             2017
                     UCTT
                            23.09
                                       777.3249
                                                 Technology
      MALE
                                     Industry
                              Semiconductors
0
1
                              Semiconductors
2
                              Semiconductors
3
                      Computer Manufacturing
4
                              Semiconductors
90
                   Security Systems Services
91
                                EDP Services
92
                              Semiconductors
    Computer Software: Prepackaged Software
93
94
                              Semiconductors
[95 rows x 15 columns]
returns = pd.read_csv("returns.csv")
returns
      GVKEY
               datadate fyear indfmt consol popsrc datafmt
                                                                 tic
curcd \
       1161 31/12/2017
                                             C
                           2017
                                   INDL
                                                     D
                                                           STD
                                                                 AMD
USD
                           2018
       1161
            31/12/2018
                                   INDL
                                             C
                                                           STD
                                                                 AMD
1
                                                     D
USD
2
       1327
             30/09/2017
                           2017
                                   INDL
                                             C
                                                     D
                                                           STD
                                                                SWKS
USD
             30/09/2018
3
       1327
                                             C
                                                     D
                                                           STD
                                                                SWKS
                           2018
                                   INDL
USD
                                             C
                                                     D
                                                           STD
                                                                 ADI
       1632 31/10/2017
                           2017
                                   INDL
USD
```

```
. .
                                                               MXL
199
     184551 31/12/2018
                          2018
                                 INDL
                                            C
                                                   D
                                                         STD
USD
200
    260893 31/12/2017
                                 INDL
                                            C
                                                         STD
                                                              BLKB
                          2017
                                                   D
USD
201
    260893 31/12/2018
                          2018
                                 INDL
                                            C
                                                   D
                                                         STD
                                                              BLKB
USD
202
    264416 31/12/2017
                          2017
                                 INDL
                                            C
                                                   D
                                                         STD
                                                             UCTT
USD
203
     264416 31/12/2018
                          2018
                                 INDL
                                            C
                                                   D
                                                         STD UCTT
USD
    costat
            prcc c
                        mkvalt
                                 returns
0
         Α
             10.28
                     9940.7600 0.795720
             18.46 18552.3000
1
         Α
             94.95 18657.8900 -0.294155
2
3
             67.02 16091.9540
         Α
                                     NaN
4
         Α
             89.03 33656.4668 -0.035943
              . . .
             17.60
                     1224.0976
199
         Α
                                     NaN
200
             94.49
                     4542.7012 -0.334321
         Α
             62.90
201
         Α
                     3054.8643
                      777.3249 -0.633175
202
             23.09
         Α
203
              8.47
                      330,8890
         Α
                                     NaN
[204 rows x 13 columns]
returns = returns [["GVKEY", "fyear", "returns"]]
returns
      GVKEY
             fyear
                    returns
0
       1161
              2017 0.795720
1
       1161
              2018
                         NaN
2
       1327
              2017 -0.294155
3
       1327
              2018
                         NaN
4
       1632
              2017 -0.035943
        . . .
               . . .
199
     184551
              2018
                         NaN
200
     260893
              2017 -0.334321
201
     260893
              2018
                         NaN
202
     264416
              2017 -0.633175
203
     264416
              2018
                         NaN
[204 rows x 3 columns]
                            # the "left" dataframe
df = pd.merge(joined,
                                  # the "right" dataframe
                        returns,
                        how = 'left', # which observations to keep?
Here we are specifying that we keep the "left" dataset
```

```
on = 'GVKEY' # the join key
        )
df
                                           GVKEY
                        CONAME
                                  SALARY
                                                   YEAR PCEO \
       ADVANCED MICRO DEVICES
                                 924.997
                                            1161
0
                                                   2017
                                                         CE0
       ADVANCED MICRO DEVICES
                                 924.997
                                             1161
1
                                                   2017
                                                         CE0
2
       SKYWORKS SOLUTIONS INC
                                 850.000
                                            1327
                                                   2017
                                                         CE0
3
       SKYWORKS SOLUTIONS INC
                                 850.000
                                             1327
                                                   2017
                                                         CE0
4
           ANALOG DEVICES INC
                                 878,000
                                             1632
                                                   2017
                                                         CE0
                                                    . . .
                                 492.299
                                                   2017
185
                 MAXLINEAR INC
                                          184551
                                                         CE<sub>0</sub>
186
                 BLACKBAUD INC
                                 700.027
                                          260893
                                                   2017
                                                         CE0
                 BLACKBAUD INC
                                 700.027
187
                                          260893
                                                   2017
                                                         CE0
                                 406.923
188
     ULTRA CLEAN HOLDINGS INC
                                          264416
                                                   2017
                                                         CE0
189
     ULTRA CLEAN HOLDINGS INC
                                 406.923
                                          264416
                                                   2017
                                                         CE0
                                        TITLE
                                                      EXEC LNAME
EXEC FNAME \
                                  CEO & Chair
                                                       Su, Ph.D.
Lisa
                                  CEO & Chair
1
                                                       Su, Ph.D.
Lisa
     CEO, President & Chairman of the Board
                                                         Griffin
Liam
3
     CEO, President & Chairman of the Board
                                                         Griffin
Liam
                   President, CEO & Director
                                                           Roche
Vincent
185
      Co-Founder, Chairman, CEO & President Seendripu, Ph.D.
Kishore
                   President, CEO & Director
186
                                                         Gianoni
Michael
                   President, CEO & Director
187
                                                         Gianoni
Michael
                              CEO & Director
                                                      Scholhamer
188
James
189
                              CEO & Director
                                                      Scholhamer
James
     GENDER
             fyear x TICKER
                                 PCC
                                              MV
                                                       Sector \
                 2017
                              10.28
                                       9940.7600
                                                   Technology
0
     FEMALE
                         AMD
                              10.28
                                       9940.7600
                                                   Technology
1
     FEMALE
                 2017
                         AMD
2
       MALE
                 2017
                        SWKS
                              94.95
                                      18657.8900
                                                   Technology
3
                              94.95
                                      18657.8900
                                                   Technology
       MALE
                 2017
                        SWKS
                                                   Technology
4
       MALE
                 2017
                         ADI
                              89.03
                                      33656.4668
```

. . .

. . .

. . .

. . .

. . .

```
185
       MALE
                 2017
                          MXL
                                26.42
                                        1780.7080
                                                     Technology
                                94.49
186
       MALE
                 2017
                         BLKB
                                        4542.7012
                                                     Technology
                                94.49
187
       MALE
                 2017
                         BLKB
                                        4542.7012
                                                     Technology
188
       MALE
                 2017
                         UCTT
                                23.09
                                         777.3249
                                                     Technology
189
       MALE
                                23.09
                                         777.3249
                 2017
                         UCTT
                                                     Technology
                                       Industry fyear y
                                                              returns
                                 Semiconductors
                                                      2017
0
                                                            0.795720
1
                                 Semiconductors
                                                      2018
                                                                  NaN
2
                                                      2017 -0.294155
                                 Semiconductors
3
                                 Semiconductors
                                                      2018
                                                                  NaN
4
                                                      2017 -0.035943
                                 Semiconductors
                                                       . . .
185
                                 Semiconductors
                                                      2018
                                                                  NaN
186
     Computer Software: Prepackaged Software
                                                      2017 -0.334321
     Computer Software: Prepackaged Software
                                                      2018
187
                                                                  NaN
188
                                                      2017 -0.633175
                                 Semiconductors
189
                                 Semiconductors
                                                      2018
                                                                  NaN
[190 rows x 17 columns]
df = df.dropna()
df
                         CONAME
                                    SALARY
                                              GVKEY
                                                     YEAR PCEO
0
       ADVANCED MICRO DEVICES
                                   924.997
                                               1161
                                                      2017
                                                            CE0
2
                                                      2017
       SKYWORKS SOLUTIONS INC
                                   850.000
                                               1327
                                                            CE0
4
                                   878.000
            ANALOG DEVICES INC
                                               1632
                                                      2017
                                                            CE<sub>0</sub>
                      APPLE INC
6
                                  3057.692
                                               1690
                                                      2017
                                                            CE0
8
        APPLIED MATERIALS INC
                                  1000.000
                                               1704
                                                      2017
                                                            CE0
. .
                                                . . .
                                                      . . .
                                                             . . .
                  FORTINET INC
180
                                   450.882
                                             183377
                                                      2017
                                                            CE<sub>0</sub>
182
              SPS COMMERCE INC
                                   495.000
                                             183974
                                                      2017
                                                            CE<sub>0</sub>
                                   492.299
184
                 MAXLINEAR INC
                                             184551
                                                      2017
                                                            CE<sub>0</sub>
186
                 BLACKBAUD INC
                                   700.027
                                             260893
                                                      2017
                                                            CE<sub>0</sub>
                                   406.923
188
     ULTRA CLEAN HOLDINGS INC
                                             264416
                                                      2017
                                                            CE<sub>0</sub>
                                          TITLE
                                                        EXEC LNAME
EXEC FNAME \
                                   CEO & Chair
                                                         Su, Ph.D.
0
Lisa
     CEO, President & Chairman of the Board
                                                           Griffin
Liam
                   President, CEO & Director
                                                              Roche
Vincent
                                CEO & Director
                                                               Cook
Timothy
        President, CEO & Executive Director
                                                         Dickerson
Gary
. .
                                            . . .
```

```
Founder, Chairman & CEO
                                                              Xie
180
Ken
182
                               CEO & Director
                                                            Black
Archie
184
      Co-Founder, Chairman, CEO & President Seendripu, Ph.D.
Kishore
186
                   President, CEO & Director
                                                         Gianoni
Michael
188
                               CEO & Director
                                                      Scholhamer
James
     GENDER
             fyear x TICKER
                                  PCC
                                                 MV
                                                          Sector
0
     FEMALE
                 2017
                         AMD
                                10.28
                                          9940.7600
                                                     Technology
2
       MALE
                 2017
                        SWKS
                                94.95
                                         18657.8900
                                                     Technology
4
       MALE
                 2017
                         ADI
                                89.03
                                         33656.4668
                                                     Technology
6
       MALE
                 2017
                        AAPL
                               169.23
                                       790050.0981
                                                     Technology
8
       MALE
                        AMAT
                                51.12
                                        59815.8000
                 2017
                                                     Technology
                          . . .
                                          7335.1141
180
       MALE
                 2017
                        FTNT
                                43.69
                                                     Technology
       MALE
                        SPSC
                                48.59
182
                 2017
                                          832.2009
                                                     Technology
       MALE
                                          1780.7080
                                                     Technology
184
                 2017
                         MXL
                                26.42
       MALE
                                94.49
                                          4542.7012
                                                     Technology
186
                 2017
                        BLKB
       MALE
                                23.09
                                          777.3249
188
                 2017
                        UCTT
                                                     Technology
                                      Industry
                                                 fyear_y
                                                            returns
0
                                Semiconductors
                                                    2017
                                                           0.795720
2
                                Semiconductors
                                                    2017 -0.294155
4
                                Semiconductors
                                                    2017 -0.035943
6
                       Computer Manufacturing
                                                    2017 -0.067896
8
                                Semiconductors
                                                    2017 -0.359546
                    Security Systems Services
180
                                                    2017
                                                           0.612039
182
                                  EDP Services
                                                    2017
                                                           0.695411
184
                                Semiconductors
                                                    2017 -0.333838
186
     Computer Software: Prepackaged Software
                                                    2017 -0.334321
                                                    2017 -0.633175
188
                                Semiconductors
[95 rows x 17 columns]
df.describe()
            SALARY
                                                                PCC
                              GVKEY
                                       YEAR
                                              fyear x
                                                                     \
         95.000000
                         95.000000
                                       95.0
                                                 95.0
                                                          95.000000
count
                      71676.568421
        680.593042
                                     2017.0
                                               2017.0
                                                          87.234947
mean
std
        405.147769
                      67102.532898
                                        0.0
                                                  0.0
                                                         152.838782
          0.001000
                       1161.000000
                                     2017.0
                                                           4.200000
min
                                               2017.0
        460.625000
                      15095.000000
                                               2017.0
25%
                                     2017.0
                                                          23.745000
                      31843.000000
                                               2017.0
                                                         48.900000
50%
        600.000000
                                     2017.0
                                     2017.0
75%
        846.154000
                     129151.000000
                                               2017.0
                                                          99.375000
```

```
3057.692000 264416.000000 2017.0 2017.0 1053.400000
max
                 MV
                     fyear_y
                               returns
          95.000000
                        95.0
                              95.000000
count
mean
       45840.681017
                      2017.0
                              -0.037420
       149594.009840
std
                         0.0
                               0.304431
          77.846400
                      2017.0
                              -0.633175
min
        1102.177750
25%
                      2017.0
                              -0.258670
        2631.105600
                      2017.0
                              -0.067896
50%
75%
       11325.945900
                      2017.0 0.137018
max
      790050.098100
                      2017.0 0.808134
df = df.drop(columns=["fyear_y", "fyear_x"])
df
                      CONAME
                                SALARY
                                         GVKEY YEAR PCEO \
0
      ADVANCED MICRO DEVICES
                               924.997
                                          1161
                                                2017
                                                      CE0
2
      SKYWORKS SOLUTIONS INC
                              850.000
                                          1327
                                                2017
                                                      CE<sub>0</sub>
                               878.000
                                                2017
4
          ANALOG DEVICES INC
                                          1632
                                                      CE0
6
                   APPLE INC
                              3057.692
                                                2017 CE0
                                          1690
       APPLIED MATERIALS INC
8
                              1000.000
                                          1704
                                                2017
                                                      CE0
                FORTINET INC
                               450.882
                                        183377
                                                2017
180
                                                      CE0
182
            SPS COMMERCE INC
                               495.000
                                        183974
                                                2017
                                                      CE0
                               492.299
184
               MAXLINEAR INC
                                        184551
                                               2017
                                                     CE0
                                        260893
                              700.027
186
               BLACKBAUD INC
                                               2017
                                                      CE0
                               406.923
                                        264416 2017 CEO
188 ULTRA CLEAN HOLDINGS INC
                                     TITLE
                                                  EXEC LNAME
EXEC FNAME \
                               CEO & Chair
                                            Su. Ph.D.
0
Lisa
    CEO, President & Chairman of the Board
                                                    Griffin
Liam
                 President, CEO & Director
                                                       Roche
Vincent
                            CEO & Director
                                                        Cook
Timothy
       President, CEO & Executive Director
                                                 Dickerson
Gary
. .
                  Founder, Chairman & CEO
                                                         Xie
180
Ken
182
                            CEO & Director
                                                       Black
Archie
     Co-Founder, Chairman, CEO & President Seendripu, Ph.D.
184
Kishore
                 President, CEO & Director Gianoni
186
```

Michael

188 Jame	es		CEO & Director	Scholhamer
	GENDER TICKER	PCC	MV	Sector \

	0 = 110 = 11	0			5555	
0	FEMALE	AMD	10.28	9940.7600	Technology	
2	MALE	SWKS	94.95	18657.8900	Technology	
4	MALE	ADI	89.03	33656.4668	Technology	
6	MALE	AAPL	169.23	790050.0981	Technology	
8	MALE	AMAT	51.12	59815.8000	Technology	
180	MALE	FTNT	43.69	7335.1141	Technology	
182	MALE	SPSC	48.59	832.2009	Technology	
184	MALE	MXL	26.42	1780.7080	Technology	
186	MALE	BLKB	94.49	4542.7012	Technology	
188	MALE	UCTT	23.09	777.3249	Technology	

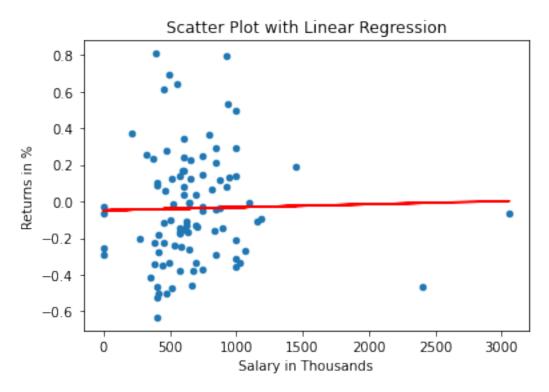
0 2 4	Industry Semiconductors Semiconductors Semiconductors	-0.294155
6	Computer Manufacturing	
8	Semiconductors	
180 182 184 186 188	Security Systems Services EDP Services Semiconductors Computer Software: Prepackaged Software Semiconductors	0.695411 -0.333838 -0.334321

[95 rows x 15 columns]

df.describe()

,	SALARY	GVKEY	YEAR	PCC	MV
count	95.000000	95.000000	95.0	95.000000	95.000000
mean	680.593042	71676.568421	2017.0	87.234947	45840.681017
std	405.147769	67102.532898	0.0	152.838782	149594.009840
min	0.001000	1161.000000	2017.0	4.200000	77.846400
25%	460.625000	15095.000000	2017.0	23.745000	1102.177750
50%	600.000000	31843.000000	2017.0	48.900000	2631.105600
75%	846.154000	129151.000000	2017.0	99.375000	11325.945900
max	3057.692000	264416.000000	2017.0	1053.400000	790050.098100

```
returns
       95.000000
count
       -0.037420
mean
std
        0.304431
min
       -0.633175
25%
       -0.258670
50%
       -0.067896
75%
        0.137018
        0.808134
max
fit = np.polyfit(df['SALARY'], df['returns'], deg=1)
df['fit'] = fit[0] * df['SALARY'] + fit[1]
df.plot(kind='scatter', x='SALARY', y='returns')
plt.plot(df['SALARY'], df['fit'], color='red')
# set the labels and title
plt.xlabel('Salary in Thousands')
plt.ylabel('Returns in %')
plt.title('Scatter Plot with Linear Regression')
# show the plot
plt.show()
```



results = smf.ols("returns ~ SALARY + MV + PCC", data=df).fit()

## OLS Regression Results

=======	:========		====	=====		.========
====== Dep. Varia	ıble:	retur	ns	R-squ	ared:	
0.001 Model:		0	lLS	Adj. I	R-squared:	
-0.032 Method:		Least Squar	es	F-sta	tistic:	
0.02656 Date:	Mo	on, 20 Feb 20	23	Prob	(F-statisti	.c):
0.994 Time:		13:50:	17	Log-L:	ikelihood:	
-21.270 No. Observ	ations:		95	AIC:		
50.54 Df Residua	ıls:		91	BIC:		
60.76 Df Model:			3			
Covariance	e Type:	nonrobu	ıst			
=======	:=======			=====	=======	:========
======	coef	std err		t	P> t	[0.025
0.975]						
	-0.0552	0.071	- 0	.774	0.441	-0.197
	2.351e-05	8.86e-05	0	. 265	0.791	-0.000
0.000 MV	-6.179e-08	3.48e-07	- 0	. 178	0.859	-7.53e-07
6.29e-07 PCC	5.341e-05	0.000	0	.168	0.867	-0.001
0.001			====	=====		
====== Omnibus: 1.767		6.9	31	Durbi	n-Watson:	
Prob(Omnib 6.571	ous):	0.0	31	Jarqu	e-Bera (JB)	:
5kew: 0.0374		0.6	33	Prob(.	JB):	
0.0374 Kurtosis:		3.2	237	Cond.	No.	

======

3.50e+05

## Notes:

- [1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
- [2] The condition number is large, 3.5e+05. This might indicate that there are

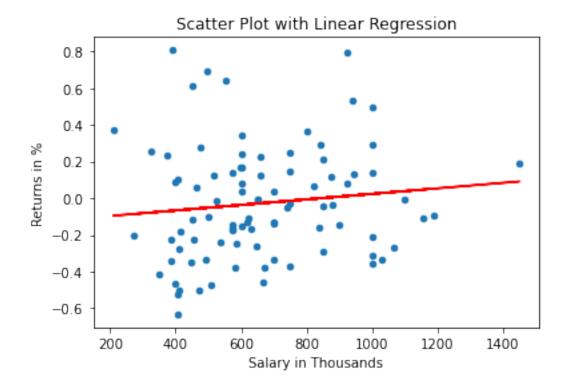
strong multicollinearity or other numerical problems.

```
df.to_csv(r'C:\Users\yuxua\Desktop\Merged Data.csv', index = False)
df = df.drop(df[df['SALARY'] > 2000 ].index)
df = df.drop(df[df['SALARY'] < 10].index)
fit = np.polyfit(df['SALARY'], df['returns'], deg=1)
df['fit'] = fit[0] * df['SALARY'] + fit[1]

df.plot(kind='scatter', x='SALARY', y='returns')
plt.plot(df['SALARY'], df['fit'], color='red')

# set the labels and title
plt.xlabel('Salary in Thousands')
plt.ylabel('Returns in %')
plt.title('Scatter Plot with Linear Regression')

# show the plot
plt.show()</pre>
```



print(results.summary())

## OLS Regression Results

			====	=====		========
====== Dep. Varia	ble:	returi	ns	R-squ	uared:	
0.001 Model:		01	LS	Adj.	R-squared:	
-0.032 Method:		Least Square	es	F-sta	atistic:	
0.02656 Date:	ı	Mon, 20 Feb 202	23	Prob	(F-statistic	):
0.994 Time:		14:00:5	50	Log-L	_ikelihood:	
-21.270 No. Observ	ations:		95	AIC:		
50.54 Df Residua			91	BIC:		
60.76	(5:	`	3	DIC:		
Df Model:			3			
Covariance	Type:	nonrobus	st			
	========		====			========
======	coef	std err		t	P> t	[0.025
0.975]						
	-0.0552	0.071	- 0	. 774	0.441	-0.197
	2.351e-05	8.86e-05	0	. 265	0.791	-0.000
	-6.179e-08	3.48e-07	- 0	. 178	0.859	-7.53e-07
6.29e-07 PCC 0.001	5.341e-05			. 168	0.867	
======	=======					=======
Omnibus: 1.767					In-Watson:	
Prob(Omnib	0.03	31	Jarqu	ue-Bera (JB):		

3.50e+05

0.633 Prob(JB):

3.237 Cond. No.

Kurtosis: 3.50e+05

6.571

Skew: 0.0374

Notes:

- [1] Standard Errors assume that the covariance matrix of the errors is correctly specified.[2] The condition number is large, 3.5e+05. This might indicate that
- [2] The condition number is large, 3.5e+05. This might indicate that there are strong multicollinearity or other numerical problems.

## df.describe()

,	SALARY	GVKEY	YEAR	PCC	MV
count	89.000000	89.000000	89.0	89.000000	89.000000
mean	665.153292	71729.280899	2017.0	85.180337	34101.435880
std	237.137403	67712.425552	0.0	157.250155	122122.606578
min	210.000000	1161.000000	2017.0	4.200000	77.846400
25%	475.000000	16710.000000	2017.0	22.120000	1024.909600
50%	620.000000	31607.000000	2017.0	47.160000	2631.105600
75%	842.308000	133288.000000	2017.0	89.030000	11168.314000
max	1450.000000	264416.000000	2017.0	1053.400000	729439.025200

	returns	fit
count	89.000000	89.000000
mean	-0.026824	-0.026824
std	0.309084	0.035744
min	-0.633175	-0.095431
25%	-0.246311	-0.055487
50%	-0.053590	-0.033630
75%	0.149886	-0.000121
max	0.808134	0.091478