

Pig User Defined Functions for Analyzing Stock Data

Chan-Ching Hsu

Purpose

Use Pig to analyze stock prices from 440 companies in the S&P 500. The data consists of a stock ticker symbol, a date, the open price, the high, the low, the closing price, and the volume traded that day. Analyze the dataset by creating a UDF (User Defined Function).

Dataset

A dataset of stock prices from the S&P 500. The dataset covers every day from Jan 1st 1980 – Jan 31st 2014. The data is in the following format

Ticker, Date, Open, High, Low, Close, Volume

The tickers can be obtained from:

https://en.wikipedia.org/wiki/List_of_S%26P_500_companies

For the purpose of price analysis we are only concerned with the **opening price**.

Task 1

Find the companies with highest stock price growth for the following time periods.

Time Periods:

Jan 1st, 1990 – Jan 3rd, 2000 (both days inclusive)

Jan 2nd, 2005 – Jan 31st, 2014 (both days inclusive)

Growth means the largest growth factor, which can be calculated by taking the ending price, divided by the starting stock price (i.e. \$1 at start to \$2 at end indicates a growth factor of 2). If a company doesn't have a price for the start date (for example, Google doesn't exist in 1982) then pick the next closest day **after** that date. If a company isn't listed on the end date, pick the next closest day **before** that date. If a company doesn't exist for more than 1 day in a time period then ignore it.

The results are as follows.

Period 1: (Ticker, Starting, Ending, Growth)

EMC	{ (EMC,19900102,3.38) }	{ (EMC,20000103,116.25) }	34.39349112426036
JDSU	{ (JDSU,19931117,8.0) }	{ (JDSU,20000103,183.75) }	22.96875
KLAC	{ (KLAC,19900326,8.75) }	{ (KLAC,20000103,112.25) }	12.82857142857143
QCOM	{ (QCOM,19911216,16.38) }	{ (QCOM,20000103,199.25) }	12.164224664224665
FFIV	{ (FFIV,19990604,10.25) }	{ (FFIV,20000103,121.25) }	11.829268292682928
CTSH	{ (CTSH,19980619,10.0) }	{ (CTSH,20000103,110.25) }	11.025
MWW	{ (MWW,19961213,14.25) }	{ (MWW,20000103,139.38) }	9.781052631578946
JBL	{ (JBL,19930503,7.5) }	{ (JBL,20000103,73.12) }	9.749333333333334
BBY	{ (BBY,19900102,6.0) }	{ (BBY,20000103,57.75) }	9.625
ADI	{ (ADI,19900102,9.87) }	{ (ADI,20000103,93.5) }	9.473150962512666

Period 2: (Ticker, Starting, Ending, Growth)

PCLN	{ (PCLN,20050103,23.72) }	{ (PCLN,20140131,1143.44) }	48.20573355817876
NFLX	{ (NFLX,20050103,12.6) }	{ (NFLX,20140131,402.31) }	31.92936507936508
CF	{ (CF,20050811,16.25) }	{ (CF,20140131,230.13) }	14.161846153846154
CMG	{ (CMG,20060126,45.0) }	{ (CMG,20140131,553.42) }	12.298222222222222
ISRG	{ (ISRG,20050103,40.02) }	{ (ISRG,20140131,409.92) }	10.24287856071964
AMZN	{ (AMZN,20050103,44.95) }	{ (AMZN,20140131,371.76) }	8.270522803114572
AAPL	{ (AAPL,20050103,64.78) }	{ (AAPL,20140131,495.18) }	7.644025933930226
MSI	{ (MSI,20050103,10.1) }	{ (MSI,20140131,64.12) }	6.348514851485149
GOOG	{ (GOOG,20050103,197.4) }	{ (GOOG,20140131,1171.35) }	5.933890577507598
CELG	{ (CELG,20050103,26.67) }	{ (CELG,20140131,150.8) }	5.654293213348332

Task 2

Calculate the 20-day moving average **for every company** and **for each day** between Oct 1st, 2013 and Oct 31st, 2013 (both days inclusive). Moving average for a particular company and for a particular date is defined as the average of the last available 20 trading days before that date for which stock prices are available for that company. For each company, find what day to start on to get a result for Oct 1st (20 trading days before).

The following is the results for General Electric, IBM, Intel, Microsoft and Apple's moving average.

GE: (Ticker, Date, Open Price, Moving Average Price)

GE	20131001	23.94	23.93
GE	20131002	24.14	23.961499999999997
GE	20131003	24.22	24.019
GE	20131004	24.18	24.062999999999995
GE	20131007	23.84	24.1145
GE	20131008	23.92	24.144
GE	20131009	23.84	24.156499999999998
GE	20131010	23.9	24.150999999999996
GE	20131011	24.32	24.137999999999998
GE	20131014	24.21	24.165499999999994
GE	20131015	24.27	24.173
GE	20131016	24.35	24.18
GE	20131017	24.38	24.174
GE	20131018	25.21	24.152999999999995
GE	20131021	25.87	24.186
GE	20131022	26.19	24.269499999999997
GE	20131023	25.99	24.368
GE	20131024	25.84	24.452499999999993
GE	20131025	26.07	24.5295
GE	20131028	25.88	24.627999999999993
GE	20131029	26.19	24.727999999999994
GE	20131030	26.29	24.840499999999995
GE	20131031	26.35	24.947999999999997

IBM: (Ticker, Date, Open Price, Moving Average Price)

IBM	20131001	185.34	188.9305
IBM	20131002	185.54	189.016
IBM	20131003	184.7	189.11399999999998
IBM	20131004	184.17	189.18149999999997
IBM	20131007	181.85	189.15749999999997
IBM	20131008	181.89	189.06599999999997
IBM	20131009	179.37	188.8005
IBM	20131010	183.17	188.42749999999998
IBM	20131011	185.25	188.03799999999998
IBM	20131014	185.41	187.73999999999998
IBM	20131015	185.74	187.32549999999998
IBM	20131016	185.42	186.94149999999996
IBM	20131017	173.84	186.58249999999998
IBM	20131018	174.8	185.5655
IBM	20131021	174.42	184.6075
IBM	20131022	173.35	183.82549999999998
IBM	20131023	175.09	182.9465
IBM	20131024	176.43	182.19400000000002
IBM	20131025	178.43	181.5105
IBM	20131028	177.04	180.9885
IBM	20131029	177.62	180.5625
IBM	20131030	181.69	180.17649999999998
IBM	20131031	179.65	179.98399999999998

INTEL: (Ticker, Date, Open Price, Moving Average Price)

INTC	20131001	22.94	23.215
INTC	20131002	22.71	23.2515
INTC	20131003	22.82	23.268
INTC	20131004	22.6	23.275
INTC	20131007	22.71	23.2705000000000002
INTC	20131008	22.87	23.27
INTC	20131009	22.51	23.259999999999998
INTC	20131010	22.99	23.24
INTC	20131011	22.83	23.2505
INTC	20131014	23.09	23.2395
INTC	20131015	23.4	23.2125
INTC	20131016	23.26	23.198
INTC	20131017	23.58	23.168499999999998
INTC	20131018	23.86	23.149499999999996
INTC	20131021	23.94	23.1455
INTC	20131022	24.17	23.1575
INTC	20131023	23.93	23.186999999999998
INTC	20131024	23.86	23.194499999999998
INTC	20131025	24.11	23.199999999999996
INTC	20131028	24.18	23.244999999999997
INTC	20131029	24.4	23.317999999999998
INTC	20131030	24.53	23.391
INTC	20131031	24.52	23.482

MICROSOFT: (Ticker, Date, Open Price, Moving Average Price)

MSFT	20131001	33.35	32.491499999999995
MSFT	20131002	33.36	32.5715
MSFT	20131003	33.88	32.67
MSFT	20131004	33.69	32.809
MSFT	20131007	33.6	32.928000000000004
MSFT	20131008	33.31	33.047
MSFT	20131009	33.07	33.1175
MSFT	20131010	33.31	33.1425
MSFT	20131011	33.68	33.172
MSFT	20131014	33.9	33.2175
MSFT	20131015	34.67	33.2435
MSFT	20131016	34.6	33.306
MSFT	20131017	34.45	33.386500000000005
MSFT	20131018	34.82	33.434999999999995
MSFT	20131021	34.98	33.5055
MSFT	20131022	35.02	33.627500000000005
MSFT	20131023	34.35	33.735
MSFT	20131024	33.82	33.828
MSFT	20131025	35.88	33.887000000000001
MSFT	20131028	35.61	34.037000000000006
MSFT	20131029	35.63	34.167500000000004
MSFT	20131030	35.53	34.2815
MSFT	20131031	35.66	34.39

GOOGLE: (Ticker, Date, Open Price, Moving Average Price)

GOOG	20131001	880.25	884.5290000000001
GOOG	20131002	882.73	885.8235000000001
GOOG	20131003	888.0	886.9425000000001
GOOG	20131004	875.0	887.6675000000001
GOOG	20131007	867.45	887.2955
GOOG	20131008	865.32	886.4815000000001
GOOG	20131009	856.28	885.2465
GOOG	20131010	863.84	883.6335000000001
GOOG	20131011	866.03	881.9555000000003
GOOG	20131014	866.66	880.5320000000002
GOOG	20131015	875.76	879.0550000000001
GOOG	20131016	885.87	878.4725000000001
GOOG	20131017	892.99	878.4485000000001
GOOG	20131018	976.58	877.7985000000001
GOOG	20131021	1011.46	881.708
GOOG	20131022	1005.0	887.4735000000001
GOOG	20131023	1001.0	893.3985
GOOG	20131024	1031.87	899.1210000000001
GOOG	20131025	1028.82	906.7995000000001
GOOG	20131028	1015.2	914.4995000000001
GOOG	20131029	1019.1	921.8055
GOOG	20131030	1037.43	928.7479999999999
GOOG	20131031	1028.93	936.483

APPLE: (Ticker, Date, Open Price, Moving Average Price)

AAPL	20131001	478.45	482.77250000000004
AAPL	20131002	485.63	482.0400000000001
AAPL	20131003	490.51	481.34350000000006
AAPL	20131004	483.86	480.8565
AAPL	20131007	486.56	480.12749999999994
AAPL	20131008	489.94	479.20550000000003
AAPL	20131009	484.64	478.39250000000004
AAPL	20131010	491.32	479.274
AAPL	20131011	486.99	480.41499999999996
AAPL	20131014	489.83	481.29749999999996
AAPL	20131015	497.51	482.73900000000003
AAPL	20131016	500.79	485.2165
AAPL	20131017	499.98	487.09700000000004
AAPL	20131018	505.99	488.56100000000004
AAPL	20131021	511.77	489.9605
AAPL	20131022	526.41	490.744
AAPL	20131023	519.0	492.3205
AAPL	20131024	525.0	493.81050000000005
AAPL	20131025	531.32	495.76050000000004
AAPL	20131028	529.04	498.1375
AAPL	20131029	536.27	500.72700000000003
AAPL	20131030	519.61	503.61800000000005
AAPL	20131031	525.0	505.31700000000001