**DS-670 Assignment 2 Mohamed Mohamar**

**Data Summary**

The original full data set that I am using was provided by Professor Robert Finn last semester during our DS-640 course. It is a stock market data set. It contains data about Income Statement, Cash Flow Statement, Balance Sheet, and Metrics and Ratios. Here is a list of the all the indicators or factors, as columns, included in the original data set:

**Income Statement**

**Indicator Code Name**

REVENUE Revenues

REVENUEUSD Revenues (USD)

COR Cost of Revenue

GP Gross Profit

RND Research and Development Expense

SGNA Selling, General and Administrative Expense

OPEX Operating Expenses

OPINC Operating Income

EBIT Earnings before Interest & Taxes (EBIT)

EBITUSD Earnings before Interest & Taxes (USD)

INTEXP Interest Expense

TAXEXP Income Tax Expense

CONSOLINC Consolidated Income

NETINCNCI Net Income to Non-Controlling Interests

NETINC Net Income

PREFDIVIS Preferred Dividends Income Statement Impact

NETINCCMN Net Income Common Stock

NETINCCMNUSD Net Income Common Stock (USD)

NETINCDIS Net Income from Discontinued Operations

EPS Earnings per Basic Share

EPSUSD Earnings per Basic Share (USD)

EPSDIL Earnings per Diluted Share

SHARESWA Weighted Average Shares

SHARESWADIL Weighted Average Shares Diluted

DPS Dividends per Basic Common Share

**Cash Flow Statement**

**Indicator Code Name**

NCFO Net Cash Flow from Operations

DEPAMOR Depreciation, Amortization & Accretion

SBCOMP Share Based Compensation

NCFI Net Cash Flow from Investing

CAPEX Capital Expenditure

NCFBUS Net Cash Flow - Business Acquisitions and Disposals

NCFINV Net Cash Flow - Investment Acquisitions and Disposals

NCFF Net Cash Flow from Financing

NCFDEBT Issuance (Repayment) of Debt Securities

NCFCOMMON Issuance (Purchase) of Equity Shares

NCFDIV Payment of Dividends & Other Cash Distributions

NCFX Effect of Exchange Rate Changes on Cash

NCF Net Cash Flow / Change in Cash & Cash Equivalents

**Balance Sheet**

**Indicator Code Name**

ASSETS Total Assets

ASSETSC Current Assets

ASSETSNC Assets Non-Current

CASHNEQ Cash and Equivalents

CASHNEQUSD Cash and Equivalents (USD)

RECEIVABLES Trade and Non-Trade Receivables

INTANGIBLES Goodwill and Intangible Assets

INVENTORY Inventory

LIABILITIES Total Liabilities

LIABILITIESC Current Liabilities

LIABILITIESNC Liabilities Non-Current

DEBT Total Debt

DEBTUSD Total Debt (USD)

DEBTC Debt Current

DEBTNC Debt Non-Current

DEFERREDREV Deferred Revenue

DEPOSITS Deposit Liabilities

INVESTMENTS Investments

INVESTMENTSC Investments Current

INVESTMENTSNC Investments Non-Current

PAYABLES Trade and Non-Trade Payables

PPNENET Property, Plant & Equipment Net

TAXASSETS Tax Assets

TAXLIABILITIES Tax Liabilities

EQUITY Shareholders Equity

EQUITYUSD Shareholders Equity (USD)

RETEARN Accumulated Retained Earnings (Deficit)

ACCOCI Accumulated Other Comprehensive Income

**Metrics & Ratios**

**Indicator Code Name Available Dimensions (old API)**

ASSETTURNOVER Asset Turnover ART, MRT

ASSETSAVG Average Assets ART, MRT

BVPS Book Value per Share ARQ, MRQ, ARY, MRY

CURRENTRATIO Current Ratio ARQ, MRQ, ARY, MRY

DE Debt to Equity Ratio ARQ, MRQ, ARY, MRY

DIVYIELD Dividend Yield

EBITDA Earnings before Interest, Taxes & Depreciation Amortization ARY, ARQ, ART, MRY, MRQ, MRT

EBITDAUSD Earnings before Interest, Taxes & Depreciation Amortization (USD) ARY, ARQ, ART, MRY, MRQ, MRT

EBITDAMARGIN EBITDA Margin ART, MRT

EBT Earnings before Tax

EQUITYAVG Average Equity ART, MRT

EV Enterprise Value

EVEBIT Enterprise Value over EBIT ART, MRT

EVEBITDA Enterprise Value over EBITDA ART, MRT

FCF Free Cash Flow ARY, ARQ, ART, MRY, MRQ, MRT

FCFPS Free Cash Flow per Share ARY, ARQ, ART, MRY, MRQ, MRT

FXUSD Foreign Currency to USD Exchange Rate

GROSSMARGIN Gross Margin ART, MRT

INVCAP Invested Capital ARQ, MRQ, ARY, MRY

INVCAPAVG Invested Capital Average ART, MRT

MARKETCAP Market Capitalization

NETMARGIN Profit Margin ART, MRT

PE Price Earnings Damodaran Method ART, MRT

PE1 Price to Earnings Ratio ART, MRT

PS1 Price to Sales Ratio ART, MRT

PS Price Sales Damodaran Method ART, MRT

PB Price to Book Value ARQ, MRQ, ARY, MRY

ROIC Return on Invested Capital ART, MRT

SPS Sales per Share ART, MRT

PAYOUTRATIO Payout Ratio ART, MRT

ROA Return on Average Assets ART, MRT

ROE Return on Average Equity ART, MRT

ROS Return on Sales ART, MRT

TANGIBLES Tangible Asset Value ARQ, MRQ, ARY, MRY

TBVPS Tangible Asset Book Value per Share ARQ, MRQ, ARY, MRY

WORKINGCAPITAL Working Capital

EVENT Material Corporate Events

PRICE Share Price (Adjusted Close)

SHAREFACTOR Share Factor

SHARESBAS Shares (Basic)

I am only using a subset of the data corresponding to the ARQ (As Reported Quarterly) listings, meaning where Dimension = ARQ. The Rows are the quarterly calendar date Time Series. The Columns are twenty chosen indicators along with the calculated Returns and Log of returns.

My list of twenty chosen indicators or factors is the following:

1. CURRENTRATIO Current Ratio ARQ, MRQ, ARY, MRY
2. DE Debt to Equity Ratio ARQ, MRQ, ARY, MRY
3. DIVYIELD Dividend Yield
4. FCFPS Free Cash Flow per Share ARY,MRQ, ART, MRT,

MRQ, MRY

1. GROSSMARGIN Gross Margin ART, MRT
2. PAYOUTRATIO Payout Ratio ART, MRT
3. SPS Sales per Share ART, MRT
4. NETMARGIN Profit Margin ART, MRT
5. BVPS Book Value per Share ARQ, MRQ, ARY, MRY
6. EVEBITDA Enterprise Value over EBITDA ART, MRT
7. PE1 Price to Earnings Ratio ART, MRT
8. PS1 Price to Sales Ratio ART, MRT
9. TBVPS Tangible Asset Book Value per Share ARQ, MRQ,

ARY, MRY

1. PRICE Share Price (Adjusted Close)
2. PS Price Sales Damodaran Method ART, MRT
3. EBITDAMARGIN EBITDA Margin ART, MRT
4. PB Price to Book Value ARQ, MRQ, ARY, MRY
5. REVENUE Revenues
6. EPS Earnings per Basic Share
7. PE Price Earnings Damodaran Method ART, MRT

I first used an R for loop to compute the returns from price using the formula: Returns = P(i)/P(i-1). I added it to the data set as separate column. After selecting the twenty indicators along with two columns: ticker and calendar date, I then used another R for loop to compute the log of returns and added it to the data set as a separate column. Because there are some zeros in the Returns column, using the log of returns rendered some R infinite (Inf) values in the Logreturns column. I converted those infinite values into “NAs”, before cleaning the data set from all the “NAs”.

I finally normalized all the factors using the following formula:

Normalized factor = (factor – mean (factor))/sd (factor), where mean is the mean of the column and sd is the standard deviation of the column.

I created a new R data frame consisting of all the twenty normalized factors, along with the calendar date column, the ticker column, the returns column and the log of returns column.

I found out the number of date in the data. And using an R for loop, I extracted all the data corresponding to a particular calendar date and created a data frame for each calendar date.