**Time Series Stock Analysis**

**Francis Odo**

**(Preliminary DRAFT – Work in progress)**

**Background**

Trading stocks in the stock market is a decision-based activity that relies heavily on current performance information and the prediction of future stock performance. Data is compiled daily for all active trading sessions Monday through Friday, weekly, monthly and yearly. Collected data consists of Opening Price, Lowest Price, Highest Price, Closing Price, Trading Volume and Stock Name. The information can be used for many reasons depending on the interest of participants in the stock market or anyone interested.

**Objective**

The objective is to demonstrate the application of every key aspect of data analysis techniques working with Time Series Stock Data collected over a period of five years. The analysis provides answers to key performance index questions through classification using Neural Network and Deep Learning predictive algorithms.

**Question to which answers are offered**

* Trend - Is there an upward or downward trend?
* Pattern - Is there a pattern over a period of time?
* Seasonality or Stationarity – If mean and variance are constant or joint distribution remain the same?

**Development Environment** Python Pandas Matplotlib Jupyter Notebook TensorFlow Data file (mod1\_stocks\_5yr.csv)

**Exploratory and Preprocessing**

The first step in this analysis is to inspect the data to be analyzed, in other words, get familiar with the structure and characteristics of this particular data. Having a good understanding of the data is crucial to determining how to proceed with the analysis of the data.

Our Time Series Stock Market Data consists of seven columns and sixty thousand rows. Due to processing resource challenges the data is significantly reduced to a three thousand and seventy seven rows.

Every raw data consists of wrong or incomplete or null information, which does not make any sense to the computer in terms of analysis. In the process, the data has to be cleaned. Following that cleaning procedure, the NAN, Nas and Nulls are dropped from the data. Further the date format was changed and set as index.

Column objects or features of this data are: Date – Date of trading Open – Opening trading stock price for the day High – The highest trading stock price of the day Low – The lowest trading stock price of the day Close – The closing stock price of the day Volume – The total stock trading volume of the day Name – The Ticker symbol for the company stock

In order to ensure our data sample is not overly or extremely skewed, it needs to be normalized. This step in the process converts all input data to what the computer can interpret, zeros and ones. Furthermore, the data is standardized using the Python Pandas StandardScaler.

**Code Plan**

An outline of the code plan

**Summary**

Summarization of key performance metrics

**Visualization**

Graphical representation using 2D and 3D plots in Pandas matplotlib library.

**Risks**