# Stock Prediction Application – Prompt A

**2. Project Recommendation** -- A project recommendation to convince those same senior, non- technical executives to implement the data product you are proposing to design to meet their business need. This is not a lengthy document and should have minimal technical jargon (or explanations if jargon or technical terms are used), but sufficient material to explain what will be accomplished. Around two to three pages should suffice. Remember, you’re establishing the context for your project and what it will accomplish for the client. The summary should

include *each* of the following and *be written in future tense*:

**Problem Summary** -- A summary of the problem that the application will solve. This should include a comprehensive description of the setting so it’s clear how the solution meets the client’s need(s). This is a scope statement complete with what will be included and incidental items that will not.

**Application Benefits** -- A description of how the application benefits the client and how the application will support a decision-making process in the context of the business need. Where does the product fit in filling an identified gap?  
**Application Description** -- A description, using technical details, of the data solution application and how it aligns with the client’s business priorities you’re planning to solve. **Data Description** -- A description of the data that will be used to construct the data product including its type (nominal, real numbers, etc...), origin, and structure. Identify

the independent and dependent variables. Comment on the anomalies of the dataset and the limitations of the data collected.

**Objective and Hypotheses** -- The objectives represent the primary desired outcomes for the project and application. State what you want the application to achieve in relation to the goals. The hypotheses should be structured as a proposed explanation for a

phenomenon(s) based on what the application will produce. State the “if” condition and the “then” outcome which is the prediction of general outcomes. You’ll describe the validity of your hypothesis in Prompt D.  
**Methodology** -- An outline of the project methodology. Choose an industry standard methodology (e.g. ADDIE, Agile, or SDLC) that you’ll use to manage your project. Describe why that methodology is appropriate and then indicate what parts of the project will align with the methodology phases.

**Funding Requirements**-- The project’s funding requirements. Taking into consideration environment, personnel, and any licensing or programming tools required what would be the overall cost involved. Be sure this amount matches the letter of transmittal.

**Stakeholders Impact** – Stakeholders are individuals or organizations with a vested interest in the project including being affected as a result of project execution or project completion. They may also influence the project's objectives and outcomes.

Stakeholders may be internal, or external (e.g. customers), or members of a community. Describe the impact of your solution might have on the project stakeholders.

**Data Precautions** – Identify the data from the dataset(s) that are sensitive and/or protected. Then review the general guidelines related to working with that data. Discuss the general guidelines related to working with the sensitive data. Precautions may include data preparation, data analysis, data storage, access to data, and data dissemination. Explain the relevance between the precautions and application treatment of said data. If these considerations do not apply to the project and the data you are using, review the ethical and legal precautions commonly used when working with sensitive data. Obvious situations would be health care (HIPAA), education (FERPA), or payment (PCI DSS).

**Developer’s Expertise** – Provide information (e.g. academic training and professional expertise) about the software developer and how he/she is qualified to complete the project in a timely fashion. Be sure to relate it to the solution you’re proposing. This information may be real or hypothetical to fit the project topic and scenario you have created.

# Stock Prediction Application - Prompt B

## Problem Statement –

The problem with GIC’s current configuration revolves around a lack of a predictive, flexible, and description solution. This inhibits GIC from optimizing adaptive trading strategies and being able to communicate the results to clients. Adopting a tool that can provide better throughput and potential improved accuracy in trading strategies.

## Customer Summary –

The proposed Stock Prediction Application (SPA) will provide the following opportunities for GIC’s lacking framework:

* Flexibility to evaluate a stock prediction by itself or against other stocks loaded into the dataset.

The SPA will grant users the option to either predict a stock based on itself (labeled, data only), or make a prediction based on other stock datasets. Additional datasets beyond the target stock will be considered independent or descriptive features. By including these capabilities within the SPA, the application sets itself apart by allowing the user to easily see how either positive or negative correlations effect the prediction of the stock. Each feature will also be provided its own scaler within the source code allowing the descriptive features to come from datasets with completely different range (I.E. predicting a penny stock based on Nasdaq).

* Visualization of predictions, correlated data, and hierarchical clustering.

The SPA will provide visualizations of the training data set predictions as well at the actual future predictions. The training and future prediction visualizations will be in the forms of active response linear graphs. These graphs, leveraged from the Plotly package, contain their own tools bars and scroll-over data display to allow for a richer and more dynamic user experience.

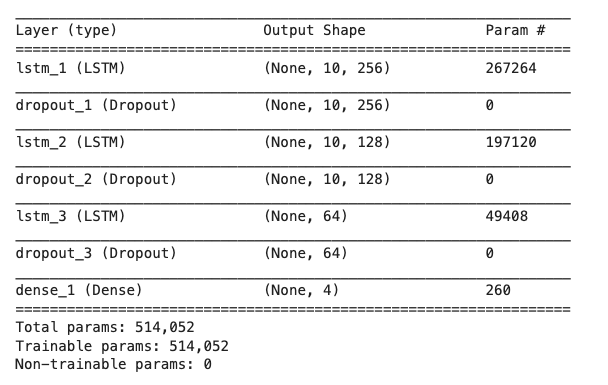
When two or more stock datasets are loaded into the SPA, the application will provide two additional graphs of processed correlated data:

* + Heatmap – this will display a graph in blocks of color, legend provided, that will represent the correlations between the datasets. This graph is also leveraged from the Plotly package and has an interactive toolbar and scroll-over displays of the data.
  + Dendrogram – this graph will chart the hierarchical clustering in a bottom-up approach. This means that the closest correlations will display as connected together and closest to the bottom. From there, every connection up the graph will indicate a weaker correlation. This graph is also a part of the Plotly package.

These visualization tools will set the SPA apart as a user-friendly and viable predictive tool and will allow also GIC to better communicate strategies to clients.

* Ease of model refinement

Leveraging a long short-term memory (LSTM) recurrent neural network (RNN), the source code behind the application will easily be configurable to add or subtract different types of layers to the model. Beyond this, parameters such as: batch size, sample size, number of epochs, step size, and number of days to predict will be able to be extracted to a front panel with little additional work required. For the initial development release of the software the RNN model with solely be located within the source code with the following architecture:



The SPA will be developed as a web application. Initially it will be released on <https://mybinder.org>, pending the approval and interested of GIC, the application would eventually be migrated to their website as a tool for employees and clients.

The application will provide the user a file upload to load the data from the desired datasets. Beyond that, the rest of the interactions with the user will be button clicks to: process data, train the model, and run the model predictions. The application will initially be designed to keep the RNN configuration parameters hidden from users. Keeping the parameters hidden will allow for consistency across datasets. The configuration parameters that will be implemented in the source code are shown in the table below:

|  |  |
| --- | --- |
| Parameter | Value |
| Input (look-back) | 10 |
| Features | Number of dataset columns |
| Epochs | 20 |
| Batch size | 50 |
| Days to predict | 5 |

Existing System Analysis –

GIC's existing system has been implementing a stock tool that is based on the current correlation of stocks based around linear regression. This tool can be used to make trend predictions; however, it is based mainly on correlation only. This measurement ultimately boils down to the equation of a straight line (y = mx+b) as the correlation slope. This kind of stock prediction can be seen as more ineffective because it tries to measure the non-linearity of the stock market within a linear measurement. The current tool is not optimized for creating stock trend prediction based on large datasets.

Implementing the SPA in place of GIC’s current software would provide the functionality of a LSTM RNN. RNNs can be much more effective at learning trends based on datasets because of the ability to remember past values and apply those values to the current measurement. Based on the output of the RNN compared against the validation data, the RNN will then adjust the weights of each neuron within the architecture before processing the next batch of data.

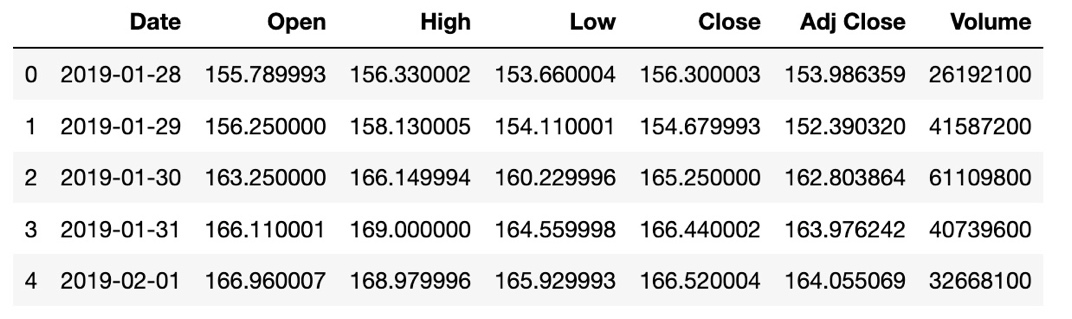
GIC’s present software is a license-based desktop application. This configuration prevents users from accessing the tool on different systems without downloading the application and purchasing another license file. The tool is also not available to clients who would like to self-invest and leverage GIC’s tool framework. Migrating this part of GIC’s toolset to the SPA will allow for employees to access the prediction application from anywhere with internet connectivity. The SPA web-based solution will also make it available for client use.

Data –

The data processed by the application will be based around the .csv structure of historic stock data downloaded from <https://finance.yahoo.com>. Thus, all datasets to be processed by the application must have a file type of .csv. Example datasets will be provided in the “Data\_Samples” folder of the proposal package. Leveraging the Pandas Python package, the application will be able to read the .csv and load it into a Pandas DataFrame. The DataFrame will be the base object of all data manipulation for the application. An example of how this will be performed is shown in the code snippet below:



The previous code will output the following:



The function “.head()” will output the first 5 rows of the DataFrame as displayed above.

*The SPA will assume that the user if the user is providing multiple datasets that each dataset is in the same format and contains the same number of rows*. For consistency it will be recommended to pull stock data from Yahoo! Finance with the same timeline for each stock. This will ensure that the datasets are formatted in the same way. Since the application will make a 5-day prediction the “Date” column of the dataset will need to be in timesteps of days formatted as: YYYY-MM-DD. The feature from each dataset that will be used in the prediction is “Close” column. The collaborative DataFrame built from all the loaded datasets will contain an indexed date for each timestep and the “Close” column from each of the datasets passed in the SPA. Because of this, *it will be necessary that datasets passed into the SPA have “Date” and “Close” columns*. The data will be partitioned in to the training dataset and the full dataset.

After the data has been processed and partitioned it will be passed into the model. LSTM RNNs require that input be in the shape of tensors. Tensors are 3-dimensional datasets in following structure:

A close up of a logo

Description automatically generated

*(*TensorFlow. 2020. *Time series forecasting.* time\_series[digital image]. Retrieved from https://www.tensorflow.org/tutorials/structured\_data/time\_series)

In order for the model to train and validate properly, the data will have to be shaped in a way that allows for progressive validation. This mean the next value after the lookback timesteps will be the value added to the validation dataset for that series of time steps. A rudimentary example of the data series would look like this:

|  |  |
| --- | --- |
| Training Data | Validation (number to be predicted) |
| [[1], [2], [3]] | [4] |
| [[2], [3], [4]] | [5] |
| [[3], [4], [5]] | [6] |

Preparing and shaping the data for the model will be one of the most important steps of the SPA. The DataFrame will be shaped into the tensors will the following algorithm:



The “shape\_data” function will return arrays built using the NumPy Python package. The package will allow for array reshaping within the source code. The algorithm shown above will return the training data (x) in the shape of a 3-dimensional NumPy array which will be the tensor model input. The validation (y) data will be in the shape of a 1-dimensional NumPy array. After the data is shaped into the proper datasets it will be ready to input into the model.

Project Methodology –

The SPA will be developed in an Agile environment. Developing in an Agile environment will allow for rapid revision of the product and constant improvement and debugging. The product development process will begin by research on the best way to implement the RNN. Once the decision on the model to use has been made, focus will shift into locating data resources and data formatting. Different types of independent features will be experimented with to determine the best route for optimal prediction accuracy. The project process will then roll into the development process, the application will go through iterative updates and added features based on the requirements provided by GIC. The application functionality will be segmented out into different runnable portions and will be worked on by different developers. This divide-and-conquer strategy will streamline the development process to ensure that deliverable deadlines are met.

Unit testing will consist of functional testing and accuracy optimization. Refining the configuration of the learning model will be a large portion of the unit testing. This will consist of running many tests with varying numbers of: days to predict, epochs, batch sizes, and timeseries look-backs. The process will attempt to settle on a model configuration that is generic and effective across many different kinds of datasets.

The integration of the application will ensure the accuracy of the graphs against the data being passed in. Varying numbers of properly formatted datasets will also be run over several iterations to verify that the application can perform as necessary under different constraints. The application will then be validated on several browsers and operating systems to ensure its interoperability across platforms.

The SPA system test will run using a stripped-down interface built on <https://mybinder.org> and the Voila package to render the source code dashboard. This will allow for a real-world test of the application and validate the functionality on deployment. Acceptance of the application on the development end will consist of the successful functionality of the application. Since the stock market is largely unpredictable it will be determined by the customer if the software meets the standards put forth.

Project Outcomes –

The project outcome will result with two types of deliverables: development process deliverables, and final outcome deliverables. The development process deliverables will consist milestone achievement and schedule, application piece-parts and runnable code segments. The action items recorded in the scrums will be recorded and charted for the customer as an additional form of accountability to meet delivery deadline. runnable code segments will be available in the project source control repository via GitHub. These deliverables will be scheduled to be delivered every two weeks at the GIC customer requirement sync-ups.

Final outcome deliverables will consist of a functional web application, sample datasets, and complete source code. The source code will be source controlled in GitHub to allow for optimal flexibility and redundancy.

## Implementation Plan --

The implementation process will begin once the application has reached initial acceptable functionality. This will begin with the customer “development release”. The development release will consist of initial user testing and identification of potential software bugs. An opportunity for additional customer requirements will also be derived during this release. After the development release the application will be taken back to development to add additional features and work out identified bugs.

After the development revisions have been made, the “alpha” software release will be delivered to the customer. The alpha release will consist of the following deliverables: a functional application, source code repository, and an operation manual. During this time the application will go through extensive black box user testing by GIC. For one month, GIC will use the application regularly and log additional bugs during this time. Once the alpha release has lapsed the application bug log will be brought back to application development. The bugs will be fixed and the application will go through regression testing to validate the bug fixes did not break any other parts of the application execution.

Application deployment will then roll in the beta release of the software. For a period 3-6 months the application will be regularly used by GIC and GIC clients. All bugs will be recorded by GIC for the beta period. After the beta period the bug fixes will be implemented, regression tested and pushed to the repository. The final release will follow the bug fixes and software implement following the beta period. The deliverables of the final release will include: final SPA, all raw source code files, operations manual, sample datasets, and included software support for the following 6 months.

## Evaluation Plan –

The evaluation plan for the SPA will include functional acceptance and measurement satisfaction. Functional acceptance will come from the application being able to perform the tasks specified in the requirements. These requirements include: the description and processing of datasets loaded into the application, prediction of stock based on selected features, and visualization of the data from the previously listed requirements.

The second part of evaluation will be determined in measurement satisfaction. The nature of the stock market is impossible to completely and accurately predict because of infinite amount of potential independent features. The metric that will be used in the application and provided to GIC is the “root mean squared” (RMS) statistical measurement. The RMS formula is shown as follows:

### Root Mean Squared Formula

A picture containing object

Description automatically generated

(Wikipedia. 2020. *Root Mean Squared.* RMS [digital image]. Retrieved from https://en.wikipedia.org/wiki/Root\_mean\_square)

RMS will be used to calculate the difference between the prediction of the target feature and the actual value of the target feature. This information will available and printed out after the first iteration of the training data. The RMS will not be calculated for the actual future prediction because there will be no data to validate it against.

## Resources and Costs –

GIC will not be hosting its own server but rather hosting via a third-party; this cost will be reflected in the price breakdown table. GIC already hosts its website so factoring in URL and webpage will not be reflected in the pricing. GIC already has a baseline infrastructure, the SPA will be integrated into the current structure which will save GIC money in cost of hardware and other software tools.

Required costs:

|  |  |
| --- | --- |
| Requirement | Cost |
| Server space | $50 (monthly) |
| Final release build software | $250 (one-time) |

## Environment Costs --

The application leverages open source IDEs, software libraries, and software package; because of this, there will not be any additional cost for the software to develop the application. The only environment cost to GIC will be server space, this pricing is reflected in the table located “[Resources and Costs](#_Resources_and_Costs)” section of this document.

## Human Resource Requirements –

Pricing based on base rate of $35 an hour.

|  |  |  |
| --- | --- | --- |
| Requirement | Time (Hrs.) | Cost |
| RNN Planning | 20 | $700 |
| Data Sourcing and Planning | 20 | $700 |
| Development, bug fixes, and feature additions | 120 | $4,200 |
| Meetings and Milestone Tracking | 15 | $525 |
| Testing (unit, regression, acceptance) | 40 | $1,400 |
| Documentation | 30 | $1,050 |
| **Totals** | **245 hrs.** | **$8,575** |

## Timeline and Milestones–

|  |  |  |
| --- | --- | --- |
| Start Date | End Date | Milestone |
| 2/10/2020 | 2/15/2020 | RNN Planning & Data Sourcing |
| 2/17/2020 | 3/9/2020 | Initial Development |
| 3/9/2020 | 3/16/2020 | Initial Testing |
| 3/16/2020 | 3/30/2020 | Development Release |
| 3/30/2020 | 4/27/2020 | Alpha Release |
| 4/27/2020 | 7/27/2020 | Beta Release |
| 7/27/2020 | 8/31/2020 | Final Release |

# Stock Prediction Application – Prompt C

## Data Methods –

The data methods that will be within the Stock Prediction Application (SPA) include a descriptive method and a non-descriptive method.

* Descriptive Method

The descriptive method that will be used in the SPA in hierarchical clustering. The clustering is based off a correlation table that is constructed using the Pandas package correlation function. The function returns a DataFrame that consists of the correlation values of the DataFrame containing all the descriptive and target features. The correlation is then partitioned into hierarchical clusters and visualized in graphs.

* Non-descriptive Method

The non-descriptive method that will be used in the SPA is long short-term memory (LSTM) recurrent neural network (RNN). The RNN will be the method used for making all predictions within the application. LSTM RNNs function by have an extra layer of neurons within the model that act a feedback loop to the input neuron. The input neurons contain vector inputs, one input coming from data in the dataset, the other input coming from the loopback of the hidden neuron that contains the previous iterations of data.

Datasets –

The example datasets that will be used for the SPA come from Yahoo! Finance. They are public files with a constantly updated and extensible library of information. Yahoo! Finance historical data is the recommended data source for the SPA. All datasets downloaded from “Historical Data” of Yahoo! Finance, provided they contain the same date range, will be compatible with the SPA without any additional user processing.

## Analytics –

The SPA will receive its analytics as a result of the RNN model predictions. The RNN will learn throughout 20 epochs and make a prediction based on what it has learned. The RNN model learns by starting at an arbitrary prediction, after the model makes a prediction the neurons inside the RNN are assigned different weights, these weights help the RNN refine its learning process by evaluating the input data through calibrated neurons and data that has already passed through the RNN. The loopback nature of the LSTM allows the model to hold onto values and make a more accurate prediction. The SPA will output a prediction of the target stock price for the next 5 days.

## Data Cleaning –

The processing and cleaning of the datasets will begin by dropping all null rows in the DataFrame, this will aid in ensuring accuracy throughout the measurements. After this, each individual feature (column) of the DataFrame will be pulled out and scaled to a relative number between 0 and 1. The MinMaxScaler functionality from the SciKit-Learn Python package will be leveraged to achieve such scaling. A Python dictionary will contain the scaler objects for each feature, the dictionary will be keyed by the stock name previously uploaded. The process of training the data will happen over two datasets: the training dataset, and the full dataset. The training dataset will consist of the first 80% of the full data, the model will train and be validated against this data. Once the data is scaled and reassembled into the main DataFrame it will partitioned into the training dataset. The following code snippet will be used to perform the data cleaning:

## Data Visualization –

The hierarchical clustering descriptive method of the SPA will have two different kind of visualizations for the data: a heatmap, and a dendrogram. These visualizations will help the user to better interpret how the datasets are correlated to each other. The caveat with SPA will be that that in order it to run hierarchical cluster analysis on the data, there must be more than two stocks uploaded to the application. If the user processes data on a single stock the application will output that it is not capable of running correlations on less than two stocks. Below are examples of the hierarchical clustering visualizations:

Heatmap:

A screenshot of a cell phone

Description automatically generated

Dendrogram:

A screenshot of a social media post

Description automatically generated

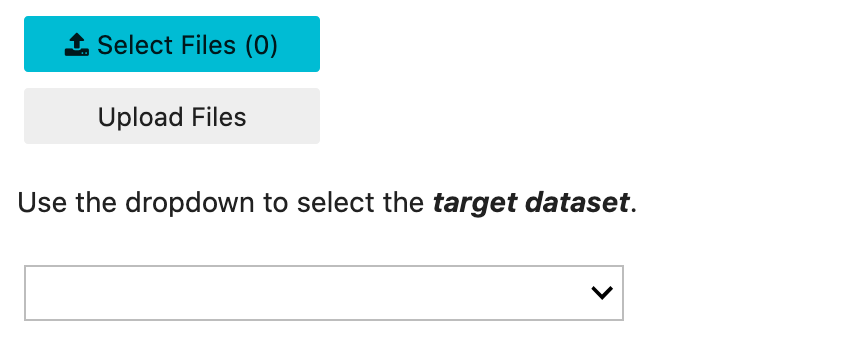
Linear graphing will be used to visualization the output from the non-descriptive model. The graph will contain two plots: a real value plot, and a predicted value plot. The predicted value is only 5 days and thus will be much shorter than the real value plot. The user will be able to, if desired, to select a box on the graph that will some in the selected section. The linear graph will used on the training data results as well as the 5-day future forecast results. Below is an example of what these linear graphs will look like:

Linear Prediction Graph:



Real-Time Queries–

The SPA GUI dashboard will allow user to run real time queries by selecting a file after from the file browser that appear when the “Select File” button is pressed. When the application run locally the response time will run very quick; however, when the application will be hosted on Binder the response time can be lagging. Nonetheless the application will still allow the user to select files, run descriptive correlations, graph the data, and make predictions on the data in real time. The file selection portion of the application is shown below:



## Adaptive Element –

The adaptive element in the SPA will be contained within the LSTM RNN. The more that the RNN model will train on datasets, the more accurate and calibrated it will become. The happens by adjust the weights of each of the neurons based on the output validated again the training set. There are two iterations of training that will occur within running the program, one of the training dataset and one on the full dataset. RNNs models can be over-trained on training data which mean that the weights are shift just to cater to predict the training data only. Over-training can lead to more deceptive accuracy during the training and poor results in the actual prediction. It is for this reason the application only allows the model to run through the training dataset once before training on the full dataset.

Outcome Accuracy–

The outcome accuracy of the measurement will come from the RMS measurement that is output as a result of the first iteration of training; a review of description of the RMS can be found in the “[Evaluation Plan](#_Root_Mean_Squared)” portion of this document. This measurement will indicate the accuracy of the prediction. The results can vary based on several different factors such as: the calibrated weights of the model, the correlation of the datasets loaded, and the cumulative timespan of the dataset. GIC as the client will ultimately determine if the accuracy of the measurement is suitable for their needs.

Security Measures–

The tool is intended to be open-access meaning that there will be no login necessary to access the application. Some of the security measures in the application revolve around trying to perform functionality that is not suitable at the time. These issues are handled by disabling and enabling controls only as the application is ready for the usage. This will prevent out of order sequencing ultimately resulting in the application crashing.

Product Health Monitoring–

The SPA will be a very simple application from the user perspective. The application products health monitoring will come from that user prompt that will occur on button clicks. For instance if the user tries to upload a file even though a file has not been select yet, the application will output “Select a file”. This kind of functionality is echoed throughout the application that acts as the product health monitoring.

## Dashboard –

The SPA dashboard is a simple user friendly way to interact with RNN model an make predictions on a stock. The dashboard consists of the following sections:

* Example output of what to expect from the application
* A file upload section
* A train model section
* A run prediction section

These components provide the necessary functionality to operate the SPA.

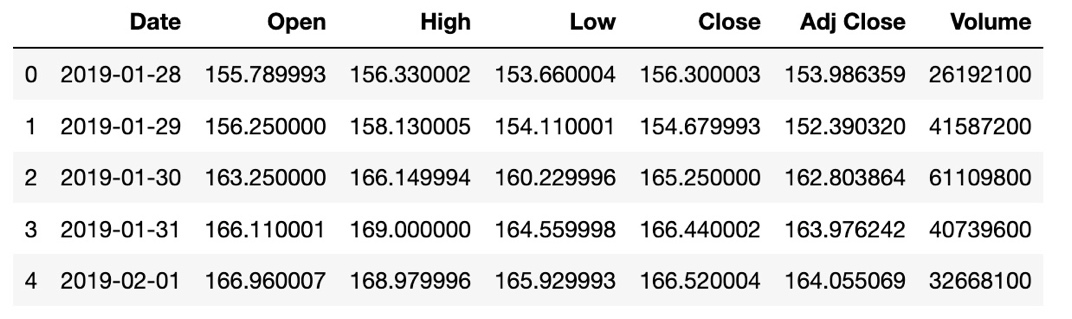
# Stock Prediction Application – Prompt D

## Project Purpose –

The purpose of the Stock Prediction Application (SPA) project was to provide Growing Investment Company (GIC) with a tool that offered a flexible, scalable, and easily interpretable predictive solution. The application was designed to predict a stock trend based on correlations to other datasets (stocks) and visualize the output. GIC’s previous toolset was not meeting the company constraints and needed to be overhauled. The SPA offered a predictive solution that met GIC’s needs and constraints.

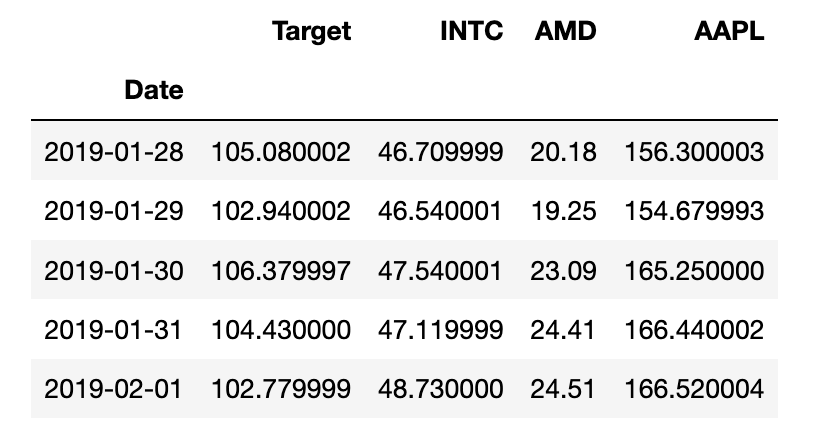
Datasets–

The raw datasets used for the SPA were historic data .csv downloads from Yahoo! Finance. The are the baseline datasets recommend for the application. When a file is uploaded to the application the raw is interpreted as:



The DataFrame above is a single dataset. The SPA is optimized when using multiple datasets. The application pulls out the “Close” column from each of the uploaded datasets and construct a collaborative “master” DataFrame with the pertinent features.

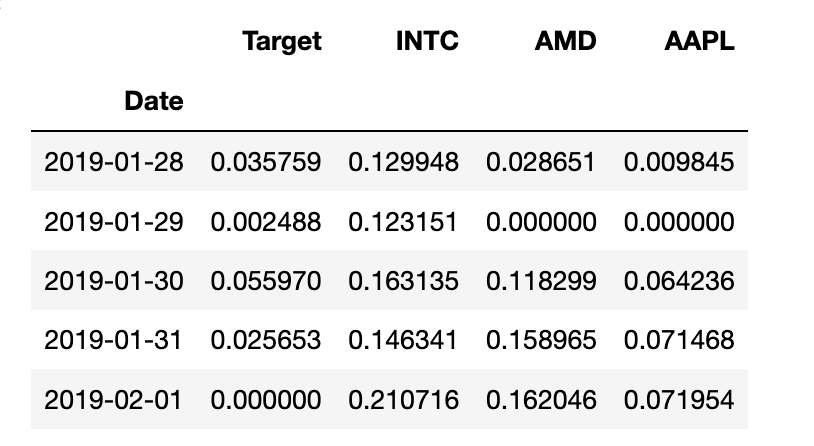
When construct, the master DataFrame is interpreted as:



The data is then scaled using the following function:



The master DataFrame is then represented as:



Data Product Code –

The SPA uses the descriptive and non-descriptive method to analyze the dataset and provide descriptive outputs.

Implementation of the descriptive hierarchical clustering method is centered around the correlation of the all the of the datasets. Once the correlations are processed, the application with order them and visualize the results. The functionality of this is displayed in the following code:



The other method that provided the predictive functionality for the application was wrapped inside the recurrent neural network (RNN) model. The following code snippet shows how the predictive functionality was accomplished within the application.

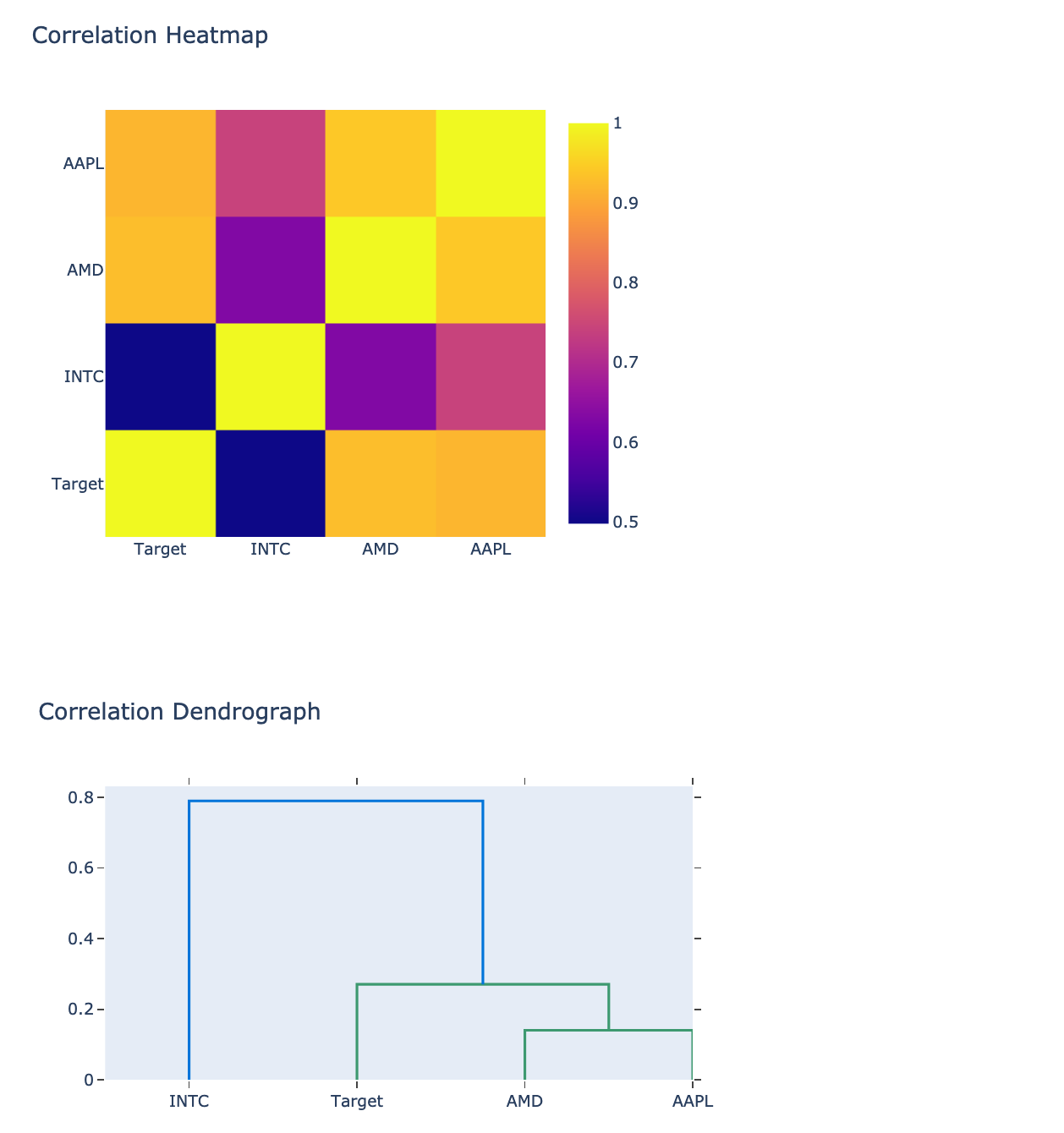


## Hypothesis Verification –

Hypothesis verification of the application was approved by GIC. The SPA provides the functionality to the user to perform predictive methods. The results of the SPA will vary greatly based on the types and relations of the datasets input to the application. Because of this, the verification metric, root mean squared, is assessed on an individual measurement basis. Thus, GIC verified the hypothesis by validating the functionality of the application.

## Effective Visualizations and Reporting –

Visualizations used in the application enhance the user experience and provide clearer results. Specific format of visualization were chosen to provide active response and interactions with the user, such as display graph or chart values on scroll over. The Plotly visualizations that were used also include tool bars that enable the user to select specific portions the graph, download an image of the graph, as well as many other features. Below is an example of the descriptive method graphs output by the application:



## Accuracy analysis –

Root mean squared (RMS) is the accuracy metric used for the SPA. RMS is a statistical standard for measuring difference between predicted values and actual value. The RMS value is output at the end of the model training on the training dataset. An example output from the application is shown below:

**Application Testing** – clearly describe the different levels of testing you used to confirm the functionality of your application. These could include unit, integration, system, and acceptance. Also, did you conduct any beta tests or usability tests for your project?

Summarize those tests and explain how results were used to modify/calibrate the product during its development.  
**Application Files** – provide a comprehensive inventory of the files required to execute your application. Include a clear description of the interdependencies and file hierarchy. Describe how those files will be organized for submission. Include those files in the submission.

**User’s Guide** – include a brief manual concerning the installation and use of your application. Be sure to describe all steps necessary to establish an environment capable of running your application. Provide clear, concise steps of how a user would execute the application and produce the results you’ve described in your documentation. Carefully consider and describe the procedural aspect of the application including know areas where certain crucial steps can affect the performance of the application. You must ensure that anybody can run the application. Please include details on the

technology context that is required for your application to properly execute. **Summation of Learning Experience** – describe how your prior experience assisted you with the capstone project then review the additional professional development you had to pursue to complete the capstone. What assistance did you seek out from other individuals? How has the experience contributed to your concept of life-long learning?