**Executive Summary:**

*Can Amway lower the MAPE for it’s revenue forecast model by at least 1% by moving from a univariate to a multivariate model?*

To address this question, this project will…

* analyze 50 features (listed in Table 1 at the end of this proposal) that could potentially enable Amway to have a lower [mean average percentage error (MAPE)](https://www.statology.org/mape-excel/) across the 3, 6, and/or 12 month time horizons
* for features with potential to lower the MAPE, an additive approach will be used to create an “optimal” multivariate model (optimal = a model that produces a MAPE at least 1% lower than the current univariate model)

An organization with a strong revenue prediction capability is able to more confidently (1) make cash flow decisions, (2) react sooner to markets with declining revenue, and (3) more pro-actively identify markets expected to grow.

**Background:**

[Amway](https://www.amway.com/en_US/about-amway) is a multi-level marketing (MLM) company, with an active presence in over 50 countries. Like all companies, the ability to accurately predict future revenue is highly valued, but also very challenging. Amway desires to improve accuracy by achieving at least a 1% lower MAPE. Amway’s annual revenue exceeds $8B, so an improvement of even a single percentage point could enable Amway’s predictions to be millions of dollars more accurate.

**Previous Forecasting Work:**

The Data Science team developed and implemented a univariate time series model that uses the ARIMA algorithm to produce a 3, 6, and 12 month revenue forecast. The feature used to predict future revenue is historical revenue, and the evaluation approach is MAPE.

**Data:**

*Description:*

The variables/features (see Table 1 for a list) were identified by talking with a number of analysts across Amway to identify what techniques they use to create revenue forecasts.

The data available for this study is from 2012 to 2021 and covers over 50 countries that Amway has operations in. The label (revenue) is tracked at a monthly level so…

* total instances available for each country is around 100 (8.5 years \* 12 months = 100 months of data).
* total instances = 100 \* 50 countries = approximately 5,000 rows of data

*Data Wrangling:*

1. This data has some missing values and will require limited transformations to bring the data to a state where model building can start.

*Data Storytelling:*

1. Analyze the features...
   1. look for correlation between each feature and the label
   2. look for correlation between features
   3. evaluate if the data has a linear, gaussian, or other distribution type
2. Various visualizations of the data (boxplots, residplots, histograms, etc...)

# Models/Machine Learning:

1. [Granger Causality](https://www.analyticsvidhya.com/blog/2021/08/granger-causality-in-time-series-explained-using-chicken-and-egg-problem/?utm_source=feedburner&utm_medium=email&utm_campaign=Feed%3A+AnalyticsVidhya+%28Analytics+Vidhya%29) will be used to analyze each feature’s potential ability to predict revenue. Granger produces a p-value, so features that have a p-value < 0.05 will be considered candidates for the multi-variate model.
2. [ARIMA](https://www.analyticsvidhya.com/blog/2018/08/auto-arima-time-series-modeling-python-r/) is the algorithm used to produce the current univariate forecast. Since ARIMA is not able to do multivariate time series, [Prophet](https://facebook.github.io/prophet/) and/or [VAR](https://www.machinelearningplus.com/time-series/vector-autoregression-examples-python/) will be used to build the multivariate model.

# Measuring Results:

1. Mean Absolute Percentage Error (MAPE) will be used to evaluate the predictive capability of the multi-variate model.
2. If two or more models produce a similar MAPE, the model with the least number of features will probably be selected.
3. The multi-variate model will be considered an improvement if it consistently generates a MAPE that is at least 1% lower than the current univariate model produces for the top 10 revenue countries.

# Table 1: List of Features

Common terms: ABO = Amway Business Owner

PV = point value (obtained when purchases are made)

