**Project Proposal – Milestone 2**

Felipe Rodriguez and Christopher Botts

Bellevue University

DSC 630 Predictive Analytics

Professor Andrew Hua

**Electronic Products Profitability Assessment**

Growth in the electronics sector is primarily driven by innovation and accelerated by consumer spending on a global level (Beers, 2022). This growth has induced market entry by new retailers. The rise of online shopping has also fragmented the market and created a highly competitive market (Beers, 2022). An analysis of this market could prove beneficial and will uncover the possibilities for investments and areas of concern.

**The Data**

This data is from a retail company that sells a variety of electronic products in the United States. This data was obtained from Kaggle and contains sales data by consumer for the year 2019. Products include small ticket items such as batteries, as well as more expensive products, such as MacBooks.

The history of sales and consumer information collected by this company provide the opportunity to analyze the sales of specific product categories in the electronics market. Furthermore, this dataset includes information about product costs that can be used to find profit margin. This enables us to determine not only which products bring in the most revenue, but also which products are the most profitable. The data also includes the quantity of items sold and the purchasing address.

**What we hope to learn**

By analyzing the data, we hope to uncover which products are most profitable or carry the most interest among our consumers. The products with the leading profits will be further analyzed to forecast sales and see how the performance will do over time. This project will also observe the items with the least profits and interest. This analysis will show where improvements can be made and determine if products need to be discontinued or if additional investment is needed. The models explored will allow us to forecast profitability over time. With the products identified and studied, a recommendation can be provided to the stakeholders. This analysis would be most useful to stakeholders and will allow them to make decisions that can allocate company resources and future investments to the products analyzed. Insights can be used to drive R&D investments to make informed innovation decisions. Product managers would also be able to utilize this information to make marketing decisions. By using the results, product managers can determine what the keys marketing differences are for products that perform better than others. If there are notable areas that can be improved, these adjustments can be made, and the new results can be analyzed.

**The Models – Analysis and Evaluation**

The plan for research is to conduct a Time Series Forecast. The specific model that will be used is the Autoregressive Moving Average or ARMA. ARMA “uses a combination of past values and white noise in order to predict future values” (Pierre, 2022). Prior to implementing and running the model, the individual sales of each product will be analyzed and the highest selling versus least selling will be selected and used in the model. This model will allow us to see which products will perform better or worse in the future. To confirm that an ARMA model will be effective for this data, an augmented Dickey-Fuller test (ADF) will be performed to determine the stationarity of the time series. Assuming the time series is stationary, we will proceed with an ARMA model. Otherwise, we will implement an ARIMA model, so that trends can be removed, and the time series can be made stationary.

When building the ARMA modeling, we will determine the order (p, q) of the autoregressive component(p) and the moving average component(q). This can be accomplished by analyzing the autocorrelation and partial autocorrelation functions of the time series, but we can use a grid search with Python to find the ideal order. The dataset will be split into testing and training set and models will be built using the training set and values from the specified grid. The performance of each model will be evaluated and compared using Bayesian Information Criterion (BIC); we are looking for the model with the lowest BIC value. The goal of utilizing BIC is to ensure the model is not overfitted but includes enough parameters to capture patterns in product sales. After deciding the order for ARMA model, we will use maximum likelihood estimation (MLE) to determine the coefficients. This will provide the maximum likelihood estimate – the point at which the observed data in our dataset is the most probable*.*

**Risks and Ethical Concerns**

This analysis has ethical concerns we want to note. This data only contains the electronic division data, which is not all encompassing of the entire company. As we explore this data and publish it, it does not contain or reflect any changes in cost or sale price, meaning that no seasonal discounts or promotions will be included in the data analysis. When consumers look at our data, they will not be able to identify target dates where there are promotions and price reductions.

**Contingency Plan**

When conducting research, if there are any issues or roadblocks that we cannot overcome, there is a contingency plan in place. Since the data includes purchasing address, the alternative that can be explored is identifying the most profitable locations, allowing for a geographical analysis. By studying the data geographically, we can make marking assumptions and determine which areas to focus marketing in. This discovered information can be relayed to the marketing team for further analysis and exploration to understand the differences between the analyzed areas. The same models can be used for this contingency plan; however, the layout will be by geographical location as opposed to by product.

Reference:

Beers, B. (2022, September 6). *Electronics Sector*. Investopedia.

<https://www.investopedia.com/ask/answers/042915/what-electronics-sector.asp>

Cornelius, V. (2023, August 24). Sales Orders. Kaggle. <https://www.kaggle.com/datasets/vincentcornlius/sales-orders>

Pierre, S. (2022, November 4). *A guide to time series forecasting in Python*. Built In. https://builtin.com/data-science/time-series-forecasting-python