

Order Volume Forecasting Analysis – Executive Summary

For online e-commerce companies, a common problem is knowing how much inventory to stock and how many people to staff the warehouse. These questions could become very easy to answer if there were an accurate estimate of how many orders would come in on a given day. Needing to know a predicted order volume gives the research question: *To what extent can daily order volume be predicted?* The hypothesis is that an ARIMA model trained on sufficient historical data can forecast daily order volumes with greater than 80% accuracy.

Data Analysis Process

A time series of historical order volumes over about 8 years is collected from the database. That series is prepared for analysis by backfilling missing values, removing outliers, and standardizing the data types of the variables in anticipation of using them for training and testing. This preparation is done in a Python environment using the Pandas library. Then, a series of tests are done on the data to determine the proper input variables for the ARIMA model, which will be the backbone of the analysis. The tests performed on the data include the AD Fuller test for the integral component, autocorrelation functions for the moving average and autoregressive components, and decomposition for the seasonal components. Given the recommended inputs from each test, a model can use the training data from our input dataset and forecast the next steps. These next steps are then compared with the actual results from the test set, and an RMSE is calculated to determine the model's efficacy in answering the research question.

The Findings

The model's results using these configurations failed to reject the null hypothesis and answered the research question that daily order volume could only be predicted to within an average error rate of 108 order units per day using RMSE. Given that fewer than 108 orders are even placed on some days, this is an unacceptably high error rate and implies that the model failed to find any significant pattern within the data through which it could predict an accurate forecast.

Limitations of ARIMA

These results are primarily a failure of ARIMA, which *can* discover simple patterns rather well but commonly struggles to detect any helpful pattern in highly dynamic or complex spaces, as it appears the data was to it. The basis of this inability to comprehend extreme patterns boils down to a requirement of ARIMA that the data be stationary. Stationary status requires a couple of different things. Firstly, no trends are present, which is handled using differencing. Secondly, no autocorrelation exists in the dataset. This second requirement is difficult to detect and correct if it does exist in the set. Oftentimes, accommodating autocorrelation of any kind in the dataset results in data loss, which impacts results in other ways.

Proposed Actions

Further attempts should be made to produce an accurate model to predict daily order volume within a reasonable error rate. More optimizations can still be made using ARIMA to try and remove the autocorrelation inherent to the dataset, and if successful, it would likely result in a much more accurate model. If possible, it will be the cheapest and easiest route to success. But if it fails, a deep learning approach will be needed for an accurate forecast. A Recurrent Neural Network model (RNN) could likely predict the required outcome in this use case very accurately.

Benefits of the Study

This study was unsuccessful in its goal to produce a model that accurately predicted daily order volume. But it did lay the groundwork for future studies to do better, and in answering the research question and providing results of specific model contexts, it gave any future studies into this topic a big head start. The potential benefits of those future studies (if successful) are primarily financial. Knowing the order volume of a day allows the company to stock inventory properly, ensuring it does not maintain excess stock (losing money) or sell out of stock early (missing out on making money).

Works Cited

No sources were used for this report.