Rodrigo Rodriguez

07/06/2024

Applied Data Science

DSC680 (SUMMER)

Milestone4: Whitepaper:

Analyzing Marketing Lift of a Targeted Local Campaign

**Business Problem**

WM (Waste Management) is a Fortune 200 company that provides waste and recycling management services across the U.S. and Canada. Sales come in the form of new service contracts and will be referred to as “orders” through the remainder of this project. The company has had a recent push to increase the number of online orders. A recent marketing campaign has been launched with the goal of increasing online orders for a specific targeted area. The marketing team now must report on the lift that advertising had on online orders in the targeted area.

Identifying the effect that advertising has on sales has been a problem that many companies have tackled. One approach is to establish a baseline of the dependent variable via a modeling technique (Shapiro & Masand, 1999). Then actual data is compared to the baseline and the lift is calculated. This is the approach that will be followed to solve the question of marketing’s effect on orders. The effectiveness of this campaign will influence the possibility of similar types of campaigns launching in other targeted areas. The success of the campaign will ultimately be determined by order lift, and subsequently, return on investment.

**Background and History**

WM is the largest provider of waste services in North America. They have historically had most of their order volume come through offline channels such as call-centers or local sales agents. A recent push has been made by the company to increase the share of orders received through the online e-commerce channel. The online channel has a lower cost to serve which ultimately makes it a more desirable channel for generating orders. The recent local marketing campaign focuses on increasing both offline and online orders. However, the dataset that WM has given permission to share for this project consists of only online orders. Although the data is only for the online channel, the methods for calculating lift are the same regardless of sales channel.

Marketing strategies have historically been measured without advanced analytics techniques (ML modeling). Points of comparison have been temporal rather than analytical, meaning that year-over-year or month-over-month comparisons were used to gauge marketing effectiveness. Although this approach may work directionally, a ML approach will provide more accurate points of comparison and will enable higher precision decision making.

**Data Explanation**

The data that is available for this project is a simple time-series data. There are no exogenous variables. While there are features that count order by line of business, the only inputs that the model will use are the date variable ‘VISIT\_START\_WEEK’ and the order variable ‘IS\_ORDER\_CONFIRM\_IND\_SUM’. Figure (1) shows the variables and data types available in the dataset. The data contains robust historical order records for the last three years ranging from May 2021 to May 2024.

A screenshot of a computer program

Description automatically generated

*Figure (1)*

**Data Prep/Cleaning**

The data was obtained from internal WM systems and was pulled, plus transformed, in a Snowflake SQL environment. The transformation done was the aggregation of daily records to weekly records. Based on domain knowledge, it is known that seasonality is present at the daily level with weekends having low sales volume and Monday’s having large sales volume. Aggregating to weekly level removes the daily seasonality from the data. As for other data preparation, the only other tasks were to change the date variable to datetime datatype and set as index for the data frame.

**Methods**

The lack of exogenous variables narrowed the focus of initial EDA to plotting the time-series data. A seasonal component was observed when plotting the data. As highlighted in Figure (2) there are yearly seasonal cycles of large spikes in order volumes starting in Spring and winding down in Summer followed by drops in order volumes in late Fall and Winter. This seasonality follows a normal industry sales cycle.

A graph with lines and numbers

Description automatically generated

*Figure (2)*

With the presence of seasonality in a 52-week cycle confirmed during EDA, a SARIMA model was chosen as the primary model to test. For additional comparison models an ARIMA and Exponential Smoothing model will also be tested. An ARIMA model will most likely not perform well due to the lack of accounting for seasonality, but it would still be interesting to observe what happens when trying to ft an ARIMA model. Aside from seasonality, weekly fluctuations can be observed throughout the year. It is possible that these fluctuations may be caused by external factors that we cannot account for in this project. Our solution to mitigate the influence of external factors is to feed a robust three-year dataset to the model to train and learn trends that may be influencing these fluctuations. Lastly, descriptive statistics revealed a total of 157 weekly records, a mean weekly order volume of 22, and a weekly order standard deviation of 8.3.

**Analysis**

The training and test sets were split using a manually listed date. Test dates range from 2/12/24 to 5/27/24. The dates specified ensured a 90/10 training and test split. A 90/10 split was used to feed as much data to the model as possible so the model can better learn the seasonality present. Additionally, the marketing campaign began on the week of 4/15/24 and thus, the test period partially included the presence of advertising. This fact will result in the performance metrics having to be evaluated with this contextual information in mind.

An ARIMA, Exponential Smoothing, and SARIMA models were fitted on the training dataset. Since the SARIMA model was the primary model for this project, a grid search was performed to identify the optimal set of hyperparameters that balanced model complexity and performance. This ultimately helps prevent the model from becoming too complex and overfitting the training data. List (1) shows the performance metrics for each model and Figures (3), (4) and (5) show the plots of the ARIMA, Exponential Smoothing, and SARIMA predictions.

As expected, we can quickly rule out the ARIMA model. The lack of a seasonal component restricts the model to only observing the most recent order trend when making predictions, resulting in a large mean square error (MSE). Since MSE penalizes larger errors more significantly than smaller ones, when comparing multiple models, the model with the lower MSE is generally considered more accurate (Pratt, McCabe, Cortes-Jimenez, & Blake, 2010).

The performance metrics and plot of both the Exponential Smoothing and SARIMA models appear to both be a good fit at first glance. However, one model is clearly superior when considering the fact of advertising being present in the test period.

1) ARIMA MAE - 15.47

ARIMA MSE - 289.83

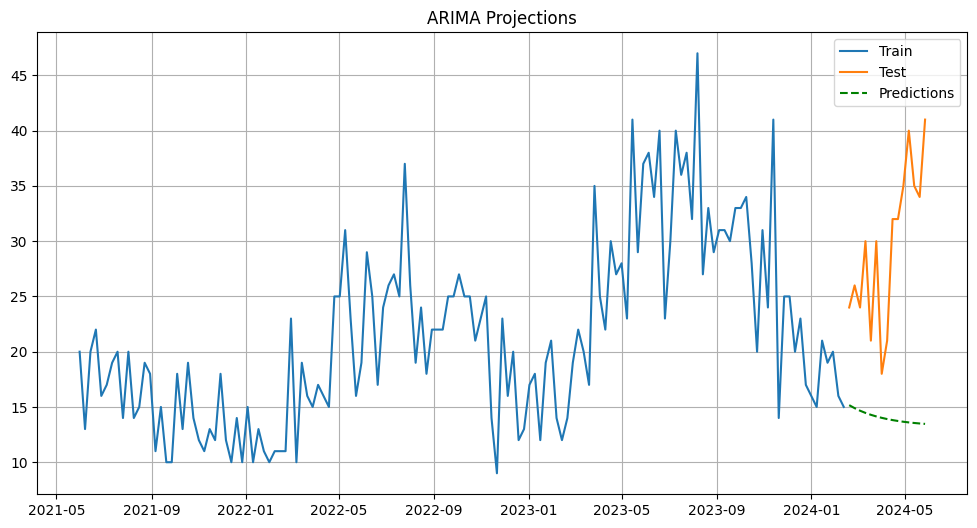
2) Exponential Smoothing MAE - 4.86

Exponential Smoothing MSE - 27.7

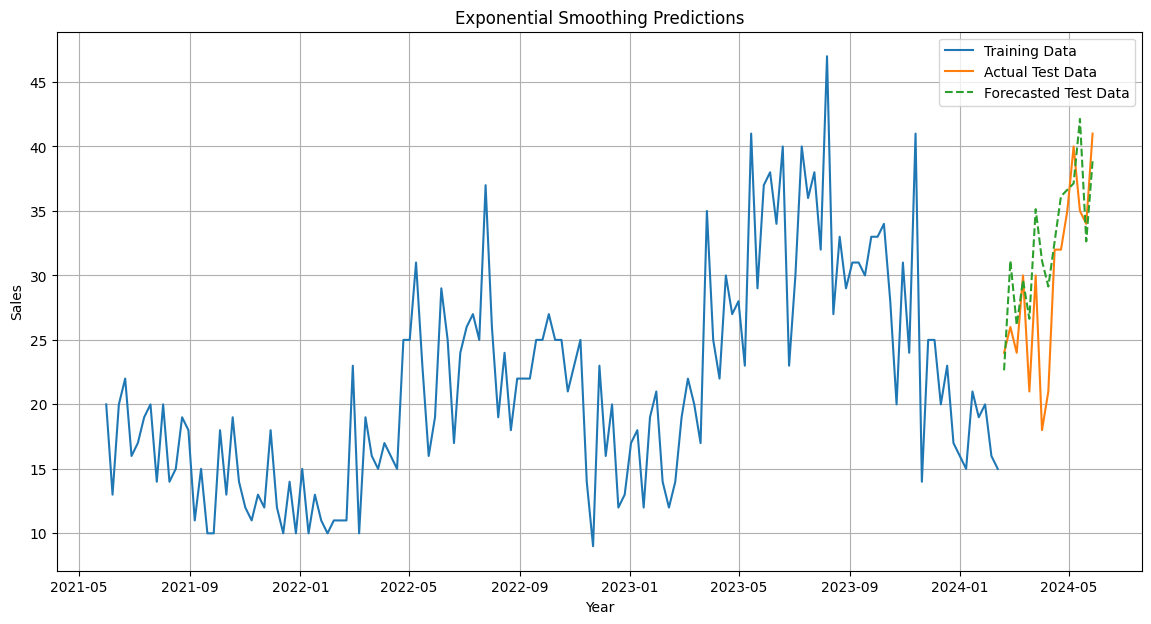
3) SARIMA MAE - 25.11

SARIMA MSE - 4.1

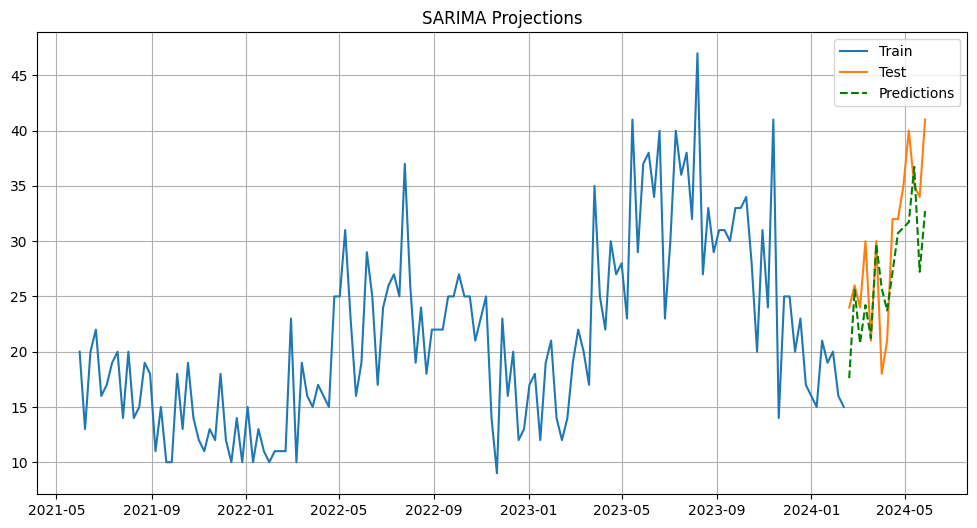
*List (1)*



*Figure (3)*



*Figure (4)*



*Figure (5)*

**Conclusion**

As mentioned, at first glance both the Exponential Smoothing and SARIMA models appear to be a good fit. The Exponential Smoothing model has the lowest mean absolute error (MAE) 4.9, meaning that the predictions were the closest to real test data. However, the key metric to observe is MSE since it measures the overall stability of the model and is highly sensitive to errors. The SARIMA model had the lowest MSE at 4.1 and a moderately high 25.1 MAE.

Since advertising partially ran during the test period, then a higher MAE and low MSE for the SARIMA model contextually make sense. The SARIMA’s lower MSE indicates that the model is generally stable, and the higher MAE is likely to occur because of the effect of advertising on orders. With this context in mind, we can now better evaluate the performance metrics for the Exponential Smoothing model. A low MAE and high MSE indicate that the model is overestimating orders. It is important to note that the Exponential Smoothing model did not have hyper parameters tuned. If hyper parameters were to be tuned, then it is possible that the performance improves. However as is, the SARIMA model is the better performing model to establish a baseline of estimated orders without advertising.

Using the SARIMA model’s predictions as a baseline we can compare the predicted orders against actual orders to calculate lift (advertising campaign’s effect on orders). The campaign started on the week of 4/15/24, from the weeks of 4/15/24 to 5/27/26, we can estimate that the campaign generated 31 incremental orders, as seen on Figure (6).



*Figure (6)*

**Assumptions**

Since this is real field data and not obtained in a controlled environment, we must assume that advertising has an effect on orders and that no other external factors (such as competition) are affecting the performance of the model during the test period. With these assumptions the project can show advertising’s effect on orders by comparing predicted orders and actual orders during the test period.

**Limitations**

The main limitation of the dataset is the exclusion of exogenous variables related to the marketing campaign. If these variables were available, then a SARIMAX could have been used to leverage these predictors and possibly further improve performance.

**Challenges**

The primary challenges included the presence of advertising in the test period and lack of exogenous marketing variables in the dataset. As previously mentioned, the addition of exogenous marketing variables would open the opportunity to test a SARIMAX model. Also, the presence of advertising during the test period makes the evaluation of performance metrics more challenging. Metrics cannot be taken at face value and must be evaluated with the context of the advertising present in the training period in mind.

**Future Uses / Additional Applications**

The marketing team has expressed interest in potentially launching similar campaigns in other locally targeted areas. This creates the opportunity to scale this analysis to future local campaigns, or to similar types of campaigns. This approach could be used to establish a baseline for other metrics besides orders. For example, the study conducted by Shapiro & Masand used a similar modeling approach to establish a baseline of estimated website visits without advertising and comparing the estimated visits to actual visits in the presence of advertising (Shapiro & Masand, 1999).

Lastly a future application that can stem from this analysis is to create cumulative gain charts to visualize the lift curve. Gain charts were not used in this analysis due to the low volume of campaign weeks currently available. However, with additional weeks of advertising campaign data, a cumulative gain chart could be created similarly to the one shown in Figure (6) (Jaffery & Liu, 2009.)

A graph with red and blue lines

Description automatically generated

*Figure 6*

**Recommendations**

My recommendation is to move forward with using the SARIMA model to establish a baseline for estimated orders without advertising. Secondly, I would recommend solving the issues detailed in the challenges section. Moving forward, the training data should only include normal circumstances to better understand model performance. Future predictions should not predict orders too far out. A SARIMA model utilizes the previous period’s predictions to predict the next period. For this reason, errors exponentially grow the farther out you try to predict. This effect limits the time range that predictions can be made. Predictions should be done for a few time periods only to minimize error. Lastly, if possible, I would include marketing variables in the dataset to leverage a SARIMAX model.

**Implementation Plan**

Before implementing I would advise that the improvements laid out in the recommendations section be made. Afterwards, a continuous feed of new data is required to be fed to the model. This ensures that the model is continuously adjusting to any changing trends in order volume. The SQL script used to pull and transform the data is written in a way that the only change needed to scale is the update the list of targeted zips. With the SQL script written for scalability, the datasets will be produced in an identical format to the dataset used for this project. The model code could then also be scaled with data for other regions, further speeding the data prep/cleaning/ and modeling phase. At that point data quality checks would need to be created to ensure that the scaled data flow and modeling process does break at any point in the pipeline.

**Ethical Assessment**

The datasets contain no personal identifiable information and data can be considered field data not obtained in a controlled environment. The model’s limitations and assumptions must be clearly communicated to the stakeholders to establish realistic expectations and avoid any misunderstanding. Additionally, any business decision made from the results of the model should be done so ethically and not exploit customers or lead to unfair business practices.

**Sources**

Shapiro, G. P., & Masand, B. (1999). Estimating Campaign Benefits and Modeling Lift. *KDD99: The First Annual International Conference on Knowledge Discovery in Data*, 185-193. https://dl.acm.org/doi/10.1145/312129.312225

Jaffery, T., & Liu, S. X. (2009, March). *Measuring campaign performance by using cumulative gain and lift chart*. In SAS Global Forum (Vol. 19). https://scsug.org/SCSUGProceedings/2008/papers/app/Tariq\_Jaffery\_and\_Shirle\_%20Liu.pdf

Pratt, S., McCabe, S., Cortes-Jimenez, I., & Blake, A. (2010). *Measuring the Effectiveness of Destination Marketing Campaigns: Comparative Analysis of Conversion Studies*. Journal of Travel Research, 49(2), 179-190. https://doi.org/10.1177/0047287509336471