Rodrigo Rodriguez

07/03/2024

Applied Data Science

DSC680 (SUMMER)

Week 5 Milestone1: Project Proposal and Data Selection: Project B

Analyzing Marketing Lift of a Targeted Local Campaign

Topic

My second project involves analyzing online sales of zip codes targeted by a local marketing campaign to estimate marketing lift. The factors included will be order quantity and transaction date.

Introduction

This project is an active project at my company, Waste Management, and has been shared with permission. WM is a Fortune 200 company that provides waste and recycling management services. Sales are made both through offline and online channels. For this project we will be focusing on a dataset containing online sales only. Recently a local campaign was launched for certain zips in the Milwaukee area. The goal of the campaign was to increase the volume of orders received within the targeted zips.

Problem Statement

The goal of this project is to determine the effect that marketing has on order volume during the campaign period and thus to determine the lift on orders that the

marketing campaign had. I plan to solve this business problem by leveraging a time-series predictive model. Based on domain knowledge, I know that the dataset has seasonal periods. For this reason, I plan to leverage a SARIMA model (Seasonal Autoregressive Integrated Moving Average).

Datasets

The dataset is obtained internally from my employer (WM) with permission. The dataset is a time-series dataset that contains e-commerce (online channel) order volumes by day, as well as indicator variables that indicate how many orders were from each line of business. This data will be used to build predictive models to understand a baseline of expected orders for the targeted zip.

Methods

The methods I will use follow the established Machine Learning phases. These include the following:

- **EDA**: Visualizing and summarizing the data to identify patterns and relationships.
- **Data Preparation**: Cleaning the dataset, handling missing values, and encoding categorical variables.
- **Data Processing**: Transforming the data if needed, vectorization, feature engineering, etc.
- **Modeling**: Applying ML modeling techniques.
- **Model Evaluation**: Measure the effectiveness of the models, MAE, MSE, etc.
- **Interpreting Results**: Analyzing model coefficients to understand the impact of different variables.

For this project I plan to test a SARIMA model. Based on the contextual knowledge I have on this dataset I know that the dataset contains seasonality thus a SARIMA model works well.

The goal of the project is to use this dataset to establish a baseline and compare the estimated orders without marketing to actual orders with marketing. Figure (1) shows a similar analysis that this project is based on. The analysis in Figure (1) estimates website hits compared to actual hits after a marketing campaign (Shapiro & Masand, 1999).

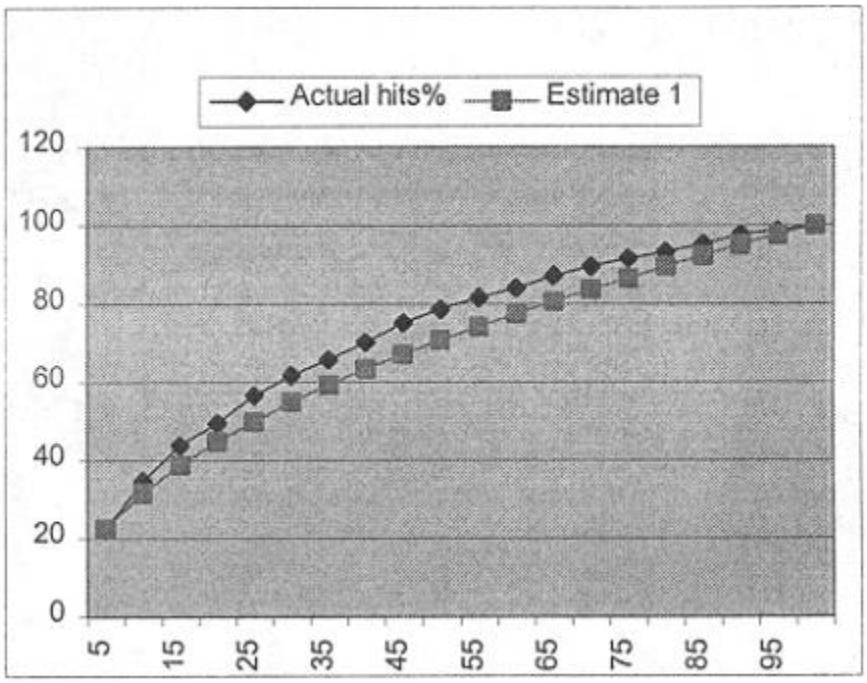

Figure (1)

Ethical Considerations

We are aware that seasonality is present in the data. However, additional careful examination of the data should be done to ensure that there are no further biases that could skew the results. Biases such as, fluctuations due to holidays, or other external events. Additionally, the limitations of the model should be clearly communicated to ensure that stakeholders understand the potential margin of error and uncertainty in the

predictions. Lastly, we must ensure that any outcomes are used to make ethical business decisions that do not exploit customers or lead to unfair practices.

Challenges / Issues

Luckily data collection will not be a challenge for this project since all datasets are structured and stored in our data warehouse. A challenge will be handling the seasonality component of the data during modeling. A grid search will likely have to be used to determine the best hyperparameters to be used for the SARIMA model.

References

I plan to reference academic sources on how best to test marketing lift. I have found one source thus far that has been referenced in Figure (1). This reference pertains to estimating marketing campaign benefits and modeling lift (Shapiro & Masand, 1999).

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://doi.org/https://dl.acm.org/doi/10.1145/312129.312225