US Decision Intelligence Data Challenge

Please complete the following data challenge over the next few days and prepare for a 30-minute presentation followed by 15 to 20 minutes of Q&A.

Context

Modeling sales data can be challenging especially to attribute features to decompose sales and predict the impact of those features. Events like Weather, Holidays and specific major events come once a year, and so does the chance to see how strategic decisions impacted the bottom line. In addition, promotions are known to affect sales, the challenge is to predict which products will be affected and to what extent.

Content

You are provided with historical sales and inventory data at store level from a retail company across the country - each store contains a number of departments. The company also runs several promotional events throughout the year. These promotions precede prominent holidays, like Black Friday, Labor Day, tax season, Christmas, etc. Within the Excel Sheet, there are 3 Tabs – Stores, Features, and Sales/Inventory

Stores

Anonymized information at store level cardinality, indicating the type and size of store

Features

Contains additional data related to the store, department, and regional activity for the given dates.

- Store - the store number
- Date - the week
- Temperature - average temperature in the region
- Fuel_Price - cost of fuel in the region

- Promotion 1-5 - anonymized data related to promotional markdowns. MarkDown data is only available after Nov 2011, and is not available for all stores all the time. Any missing value is marked with an NA
- CPI - the consumer price index
- Unemployment - the unemployment rate
- IsHoliday - whether the week is a special holiday week

Sales & Inventory

Historical sales data, which covers to 2010-02-05 to 2012-02-02. Within this tab you will find the following fields:

- Store - the store number
- Region - the region number
- Date - the week
- Weekly_Sales - sales for the given department in the given store
- IsHoliday - whether the week is a special holiday week

The Task

1. Create a model which predicts the weekly sales for each department in each store between 2012-02-03 and 2012-11-01. Report the prediction error as weighted mean absolute error between predicted sales and actual sales, where the errors in holiday weeks are weighted five times more than the non-holiday weeks
2. Model the effects of promotions on holiday weeks
3. Provide recommended actions based on the insights drawn, with prioritization placed on largest business impact
4. Model the effects of supply vs demand

Recommendations

- Feel free to use any tool of your preference but be ready to explain your thought process from data prep, to modeling to data viz.
- Do not feel constrained by these data sets, feel free to add other contextual data like weather, raw material shortages, etc. so feel free to be creative to enrich your “Causal Inference” model.
- Feel free to represent your insights however you see fit (ie. any combination of charts, text, etc.)

- Create a short presentation derived from analyzing the dataset, the presentation should tackle the following areas to the best of your ability:
    - In developing your slides, assume the intended audience is the Senior leadership of the retail company, and within the audience you will have technical and non-technical people, so consider using a business-friendly language while explaining your model and overall approach
    - Profile the provided data sets (e.g. datatypes, data distribution, missing values etc.)
    - Be prepared to explain your methodology
- Feel free to use any means to obtain the answers (Python, R, Tableau, Keynote, etc.) and be prepared to share your work afterward for further review.
- You do not need to limit your presentation to the answers to the questions above.