15.071: The Analytics Edge
Spring 2019

Homework Assignment #8
Dartboard Corporation 2.0

Due at 5pm on April 26, 2019

This is a team deliverable. Work with your designated project team. Each team must post its completed deliverable on Canvas.

Read the case "**Supply/Distribution Chain Planning at Dartboard Corporation 2.0**," available on Canvas, and perform the analyses described in the following two problems.

You need to upload the following files to Canvas by the due date: i) your team's typed answers to Problems 1 and 2 below, ii) th R Script file you used to answer Problem 1, and iii) the Julia script file you used to answer Problem 2.

Problem 1: Demand Forecasting (50 points)

For this problem you will need the following two data files available on canvas:

- `Dartboard_historical.csv`: This file spans the three-year period 2012-2014. Each row corresponds to a specific week during this period and a specific county in Dartboard's Northeast region (uniquely determined by its FIPS code). The dataset consists of 12 fields: County FIPS Code; state name; county name; county latitude; county longitude; year; county income per capita in dollars; county population; number of weeks elapsed since the beginning of 2012; season indicator (1-13); and Dartboard's actual online sales amount, in dollars, for that week and that county.
- `Dartboard_future.csv`: This file spans the 'future' 2.5-year planning horizon, from mid-2015 to end-2017. Each row corresponds to a specific week during this horizon and a specific county in Dartboard's Northeast region (uniquely determined by its FIPS code). The dataset consists of the same fields as the file `Dartboard_historical.csv` except that sales figures are not known because they are in the future since the events of the case take place in January 2015. Population and income figures were taken from official government projections.

a) (25 points) Split the historical data into a training set, consisting of all observations that fall within the two-year period 2012-2013, and a test set, consisting of all observations that fall within the year 2014. Fit the following regression model to the training set (this is the model described in the case):

$$\log\left(\frac{d_{t,c}}{pop_{t,c}}\right) = a_0 + a_1 \log\left(inc_{t,c}\right) + b_0\ t + \sum_{s=1}^{13} b_s\ u_{t,s}.$$

(for the logarithm, use R's `log10` function). Provide a copy of the output of your regression model (a printout of the output of the **`summary`** function). Do not remove any non-significant variables from your model.

b) (5 points) What do the values of the seasonality regression coefficients $(b_1, \ldots, b_{13})$ indicate about the seasonality of Dartboard's online retail sales?

c) (8 points) Assess the model's performance on the test set by calculating the mean absolute percentage error (MAPE) of the test set predictions. By 'error' we refer to the difference between actual sales and predicted sales (not the log of sales per capita).

d) (12 points) Use the regression model to predict 'future' weekly sales for each county. That is, compute a sales prediction for each observation in the `Dartboard_future.csv` file.

   i) What is Dartboard's predicted total online retail sales in the Northeast region from mid-2015 to end-2017?

   ii) What is Dartboard's predicted total online retail sales in the Northeast region in the last 8 weeks of 2017?

   iii) What is Dartboard's predicted total online retail sales in Suffolk County, MA, in the last 8 weeks of 2017?

Problem 2: Determining the Most Cost-Efficient Distribution Center Footprint (50 points)

In this problem you are asked to formulate and solve, using Julia, an optimization problem that solves for the optimal number, locations, and sizes of new DCs from the list of 17 potential sites listed in Exhibit 2 of the case. The file `Dartboard_DCs.csv`, available on Canvas, is an electronic version of this exhibit.

Your optimization model must incorporate the following:

- An objective function that minimize the total cost of distribution over the planning horizon spanning the period from mid-2015 to end-2017. Total cost is the sum of (fixed and variable) DC construction costs and transportation costs incurred throughout all 130 weeks of the planning horizon.

- The capacity plan must be sufficient to avoid stockouts at any period during the planning horizon (according to your demand forecast). Note that Dartboard's average inventory turnover is 8 weeks. Note also that demand growth and seasonality patterns imply that the peak 8-week period of demand occurs at the end of the planning horizon. Therefore, it is sufficient for your model to

ensure that sufficient capacity is available at the DCs to cover the sum of Dartboard's weekly demand over the last 8 weeks of 2017.

- Each county must be served exclusively by a single DC.

a) (30 points) Formulate and solve your optimization model in Julia. What is the optimal cost? What is the optimal DC footprint?

b) (5 points) Currently, Dartboard carries on average 8 weeks of inventory. Obviously, the fewer days that items are in inventory, the more Dartboard can save on capacity expansion costs. If Dartboard reduces the number of weeks of inventory it carries, from 8 weeks to 7 weeks, what would be the estimated cost savings over the planning horizon?

c) (5 points) Which has more impact on the total cost of the DC capacity plan over the planning horizon: a +10% increase in DC variable costs across the board, or a +10% increase in transportation costs ($ per mile) across the board? Explain why?

d) (10 points) Suppose that prior to Rama's arrival and MCG's consulting engagement, the capacity expansion plan was to open, at full capacity, five new DCs in Buffalo, Chillicothe, Knoxville, Lancaster, and Worcester. If we consider this to be the baseline expansion plan, how would you estimate the monetary value that optimization analytics has brought to Dartboard?

Note: To help your team set up and solve this problem in Julia, we have provided the script file **`sample_script.jl`** on Canvas to demonstrate some useful syntax and functions. The script formulates and solves a simpler problem: how to allocate counties to each of the three existing DC in order to minimize transportation costs between January 2012 and December 2013, while ensuring the allocation does not result in stockouts based on peak period demand during the last 8 weeks of *2013*. We emphasize that the planning horizon (2012-2013) of the simplified problem is different from the planning horizon you are asked to consider in Problem 2. We also note that the simplified problem does not consider any of the potential new DCs.