15.071A: The Analytics Edge **Fall 2019**
Dartboard Team Case

Out: November 05; Due: November 20, 11:59 pm.

Guidelines:

– *This is a team assignment. Work with your project team and submit a single report per team on Canvas (in pdf format). Please make sure to write down the names of all team members in your submission.*

– *Please show your computational work (e.g., through screenshots and/or code samples).*

– *For all optimization problems, please select a relative termination tolerance of 0.0001, i.e., any algorithm should not terminate until it finds a solution within 0.01% of the true optimal solution.*

– *When asked to "formulate an optimization model", no computation is required. It involves writing the model mathematically. This is a required step toward implementing the model and getting a solution.*

Read the case "*Supply/Distribution Chain Planning at Dartboard Corporation 2.0*," available on Canvas.

To support the analysis, you have access to the following datasets on Canvas:

- `Dartboard_historical.csv`: This file spans the three-year period 2012-2014. Each row corresponds to a week during this period and a county in Dartboard's Northeast region (determined by its FIPS code). The dataset consists of 11 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; and Dartboard's actual online sales amount, in dollars.

- `Dartboard_future.csv`: This file spans the "future" 2.5-year planning horizon, from mid-2015 to end-2017. Each row corresponds to a week during this horizon and a county in Dartboard's Northeast region (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. Population and income figures were obtained from official governmental projections.

Part A. Demand Forecasting [25 pts]

a. Split the historical data into a training set, comprising all observations that fall within the 2012-2013 period, and a test set, comprising all observations that fall in 2014. Fit the regression model described in the case to the training set. Report and interpret the coefficients of your model. Make predictions on the test set and report out-of-sample performance (as measured by the R^2). **[15 pts]**

Hint: You can either use the `log()` or the `log10()` function in `R`. Just be consistent when it comes to making sales projections.

Hint: Out-of-sample performance should measure the difference between actual and predicted sales.

b. Project your demand forecast on the "future" 2.5-year planning horizon, using the `Dartboard_future.csv` file. Write a paragraph, supported by a couple of visualization, to highlight the main takeaways of your demand projections—in view of the main business problem at hand. You may find it useful to report the predicted sales in the last 8 weeks of 2017. **[10 pts]**

Part B. Determining the Distribution Center Footprint [60 pts]

c. Formulate an optimization model that determines the number, locations, and size of new DCs from the list of 17 potential sites listed in Exhibit 2 of the case, as well as which counties are served by which DCs. The data are available on Canvas, in the file `Dartboard_DCs.csv`. **[15 pts]**

Your formulation must reflect the following:

- The DC that serves each county depends on the set of constructed DCs and their capacities.
- Each county must be served exclusively by a single DC.
- The existing DCs cannot be modified.
- The company's objective is to minimize the total cost of distribution over the planning horizon spanning mid-2015 to end-2017. Total cost is the sum of the upfront DC construction costs and the transportation costs incurred throughout all 130 weeks of the planning horizon.
- The capacity plan must be sufficient to avoid stockouts during the planning horizon (according to your demand forecast). Recall that Dartboard's average inventory turnover is 8 weeks, and that it was deemed sufficient to ensure that sufficient capacity is available to cover Dartboard's demand over the last 8 weeks of 2017.

d. Implement and solve your optimization model in Julia. Report the new DCs (if any) and which counties are served by which DC in your solution. What is the optimal cost of distribution? [**15 pts**]

Hint: You will need to compute distances between counties and DCs. We will consider here the geodesic distance, available in the `Geodesy` package. You can compute the distance between a point of latitude `lat1` and longitude `long1` and a point of latitude `lat2` and longitude `long2` as follows:

```
distance(LLA(lat1,long1,0.0),LLA(lat2,long2,0.0))/1609.34
```

e. In order to evaluate your recommendation, propose an alternative sequential approach as follows. In a first problem, find the cheapest set of DCs (and their capacities) that will enable Dartboard to avoid stockouts during the planning horizon (still, according to your demand forecast). In a second problem, given that set of DCs, optimize which counties are served by which DC. What is the optimal cost of distribution under this alternate approach? Comment on your results. [**15 pts**]

Hint: To avoid numerical errors, you can inflate DC capacities by 0.1% in the second problem.

f. Dartboard is considering one of the following three projects:

A. An overhaul of its procurement and inventory management systems, which is expected to reduce the number of weeks of carried inventory from 8 weeks to 7 weeks.

B. A renegotiation of its contracts with construction companies, which is expected to reduce the DC construction costs by 10% (including both fixed and variable costs).

C. A renegotiation of its truck leases, which is expected to reduce the transportation costs by $0.20 per mile (from $1.55 per mile down to $1.35 per mile).

Unfortunately, everyone at Dartboard is very busy and the company is only able to handle one of these three projects. Which one do you recommend Dartboard to prioritize? [**15 pts**]

Part C. Memo [15 pts]

g. Write a short memo (at most one page) addressed to Rama Anand to synthesize your recommendations, the intuition behind them, and any supporting information. [**15 pts**]

Appendix: A Simpler Optimization Model

In this appendix, we provide, as a starting point, a formulation of an optimization model that addresses a simpler problem: how to determine which counties to serve from each of the three existing DCs in order to minimize transportation costs between January 2012 and December 2013, while preventing stockouts during the last 8 weeks of 2013. We have provided the script file **`simpler_model.jl`** on Canvas that implements this model, and thus demonstrates some useful syntax and functions.

We define the following optimization model.

Sets

$$\begin{aligned} i = 1, 2, 3 :& \quad \text{distribution centers (DCs)} \\ j = 1, \cdots, 765 :& \quad \text{counties} \end{aligned}$$

Parameters

C_i : capacity of DC i (in pallets)

D_j : total demand from county j over the 2.5-year planning horizon (in pallets)

PD_j : peak demand from county j over the last 8 weeks of the planning horizon (in pallets)

d_{ij} : (geodesic) distance between DC i and county j (in miles)

tc : unit transportation cost (in \$ per pallet per mile)

Decision variables

$$x_{ij} = \begin{cases} 1 & \text{if DC } i \text{ serves county } j \\ 0 & \text{otherwise} \end{cases}$$

Formulation

$$\text{minimize} \quad tc \times \sum_{i=1}^{3} \sum_{j=1}^{765} D_j \times d_{ij} \times x_{ij} \tag{1}$$

$$\text{subject to} \quad \sum_{j=1}^{765} PD_j \times x_{ij} \leq C_i, \quad \forall i = 1, 2, 3 \tag{2}$$

$$\sum_{i=1}^{3} x_{ij} = 1, \quad \forall j = 1, \cdots, 765 \tag{3}$$

$$\mathbf{x} \text{ binary} \tag{4}$$

Equation (1) expresses the model's objective of minimizing total transportation costs, expressed as the sum-product of the unit transportation cost, the demand from the counties, and the distance traveled from the relevant DCs to the counties. Equation (2) imposes capacity constraints: each DC serves counties such that the total peak demand from all these counties will not exceed the DC's capacity. Equation (3) ensures that each county is served by exactly one DC. Equation (4) defines the domain of definition of the variables.