

Name: \_\_\_\_\_

PLEASE READ THIS PART CAREFULLY BEFORE STARTING THE EXAM.

1. Your submitted exam answers should be in **one document**, as a **Microsoft Word document**. Spreadsheet answers are not permitted.
2. Please show **all** of your work. Do not expect the grader to guess your reasoning. Your grade on the exam will depend on the clarity of your answers, the reasoning you have used, and the correctness of your answers.
3. There are **6 problems** in the exam and **100 points in total**.

\_\_\_\_\_

**GOOD LUCK!!**



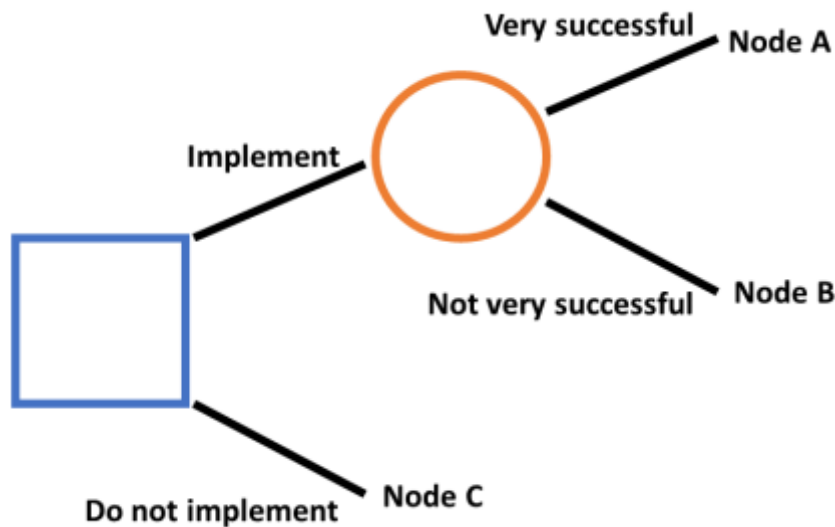
## **Problem 1 (20 points)**

### **Part 1 (12 points)**

Commonwealth Health Insurance has become interested in a new type of cancer screening. Early screening of cancer reduces the risk that cancers can develop, leading to greatly improved patient health and less costly procedures. Implementing this new screening procedure costs \$40 million every year. Staying with the current screening procedure is far less costly, only costing \$5 million every year. Hence, Commonwealth Health Insurance wants to be careful and conduct a rigorous analysis of this important decision.

Regardless, this new type of cancer screening has not been put in practice before, and hence, the final cost reduction is uncertain. Commonwealth Health Insurance has assessed that there is an 80% chance that the screening procedure is very successful, and a 20% chance that the screening is not very successful. In the case that the screening procedure is very successful, it is estimated that the yearly costs related to cancer are decreased by \$100 million. On the other hand, when it is not a very successful case, yearly costs are reduced by only \$30 million, which is the same as under the current procedure.

The decision tree associated with this problem is shown in Figure 1 below.



**Figure 1**

- (4 points) Fill in the decision tree, i.e., where needed fill in the probabilities, the end point values, and calculate the Expected Monetary Value (EMV). Please explain what is the optimal decision based on EMV.
- (4 points) Calculate the probability to be very successful for which the decisions to implement and not implement yield the same EMV.
- (4 points) Discuss the sensitivity of the decision from (a) using the outcome of (b).

Part 2 (8 points)

For each of the following statements determine if it is true or false. Please offer a **one-sentence explanation** of your answer.

- (d) (2 points) The table below describes the probability distribution of a random variable  $X$ . The values of  $X$  are given in the first row. The probabilities,  $p_1, p_2, p_3, p_4, p_5$ , and  $p_6$  are in the second row. For example,  $P(X = 6) = p_4$ .

|             |       |       |       |       |       |       |
|-------------|-------|-------|-------|-------|-------|-------|
| Value       | 7     | 3     | 13    | 6     | 20    | 21    |
| Probability | $p_1$ | $p_2$ | $p_3$ | $p_4$ | $p_5$ | $p_6$ |

- i. It is possible that the expected value of  $X$  is 13.
  - ii. It is possible that the expected value of  $X$  is 23.
- (e) (2 points) A resident of Boston is chosen completely at random. Consider the following two events:
- i. The person selected is a teacher
  - ii. The person selected is a teacher and is a vegetarian
- The probability of event (ii) can never exceed that of event (i).
- (f) (2 points) In a plot of a regression output, corresponding to a simple linear regression (OLS) model with one explanatory variable, it is possible that all of the training data points are above the regression line.
- (g) (2 points) In an optimization problem, the optimal solution is always on the boundary.

## **Problem 2 (15 Points)**

Jacob has recently opened a new apparel store close to the towns of Bern and Oulli. Bern and Oulli together have a total population of 10,000 out of which 4,000 are from Bern and 6,000 are from Oulli. Everyday multiple customers enter the store, but Jacob is interested in counting the number of times the first customer comes from Bern. Each person is equally likely to stop at the store on any given day. Moreover, this likelihood is independent and identical for different days.

- a) (2 Point) What is the probability that the first person of the day comes from Bern?

For the first 10 days, Jacob wants to know how many times the first arrival will come from Bern. Let  $Y$  denote the number of days so that the first arrival comes from Bern.

For parts (b)-(e) you can use the answer that you calculated for part (a). (You will not be penalized in case your answers to later parts change due to a calculation error in part (a).)

- b) (3 Points) What is the distribution of  $Y$ ? What is the mean of  $Y$  and standard deviation of  $Y$ ?
- c) (3 Points) What is the probability that none of the first arrivals happened from Bern?
- d) (3 Points) What is the probability that Jacob saw an equal number of first arrivals from both Bern and Oulli in the first 10 days?
- e) (4 Points) Write  $Y$  as a sum of random variables and use the Central Limit Theorem to calculate the probability that Jacob saw at least 3 and at most 5 days on which the first arrival was from Bern.

(PLEASE SHOW US THE FORMULAS YOU USED AS WELL AS THE FINAL NUMBERS FROM YOUR CALCULATIONS)

### **Problem 3 (15 Points)**

In the past year, Winter Parks has opened a new recreation park on the shores of Lake Summer. Entrance to the park is free, but there are several paid attractions as well as a membership option with additional benefits. Winter Parks wants to offer coupons to their customers to encourage them to use the attractions in the new park. Steven has been tasked with analyzing which customers should be targeted. To do this, Steven wants to predict how much each customer would spend at Lake Summer. The following is a list of the variables that Steven gathered about several households that have visited before:

- *expsum*: expenditures when visiting Lake Summer
- *visits*: number of visits to Lake Summer
- *ski*: indicator of whether the customer waterskied, 0 if not waterskiing and 1 if waterskiing
- *income*: annual household income
- *feesum*: annual member fee for Lake Summer, 0 if not paid and 1 if paid

The initial model regresses *expsum* on all the available independent variables.

| <i>Regression Statistics</i> |             | <b>(Model 1)</b> |  |  |  |
|------------------------------|-------------|------------------|--|--|--|
| Multiple R                   | 0.420358565 |                  |  |  |  |
| R Square                     | 0.176701323 |                  |  |  |  |
| Adjusted R Square            | 0.162805987 |                  |  |  |  |
| Standard Error               | 36.45922964 |                  |  |  |  |
| Observations                 | 242         |                  |  |  |  |

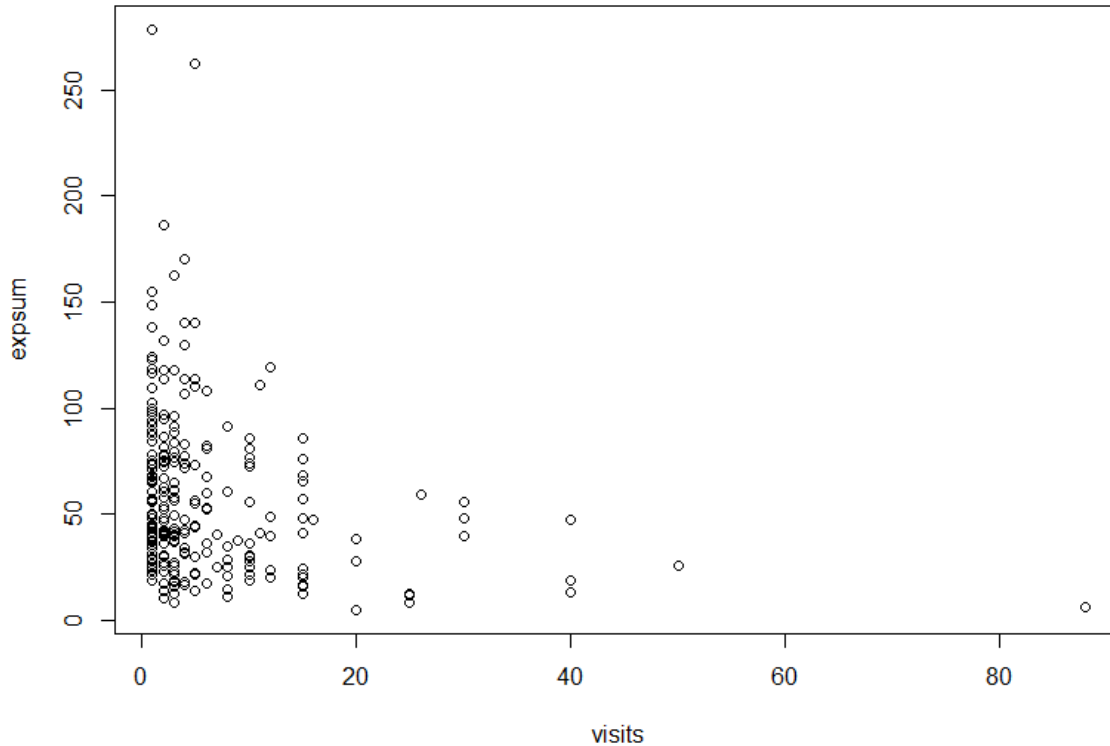
  

|           | <i>Coefficients</i> | <i>Standard Error</i> | <i>t Stat</i> | <i>Lower 95%</i> | <i>Upper 95%</i> |
|-----------|---------------------|-----------------------|---------------|------------------|------------------|
| Intercept | 37.22597296         | 6.318887743           | 5.891222391   | 24.77761238      | 49.67433353      |
| visits    | -0.968530065        | 0.265269904           | -3.651111762  | -1.491118144     | -0.445941986     |
| ski       |                     | 4.920334635           | 3.55243513    | 7.785992236      | 27.17234698      |
| income    | 4.965190085         | 1.448287511           | 3.428317961   | 2.112028916      | 7.818351255      |
| feesum    | 0.05710018          | 10.66093741           | 0.005356019   | -20.94520248     | 21.05940284      |

First, Steven wants to analyze the regression output to see which variables are useful predictors.

- (a) (2 Points) Calculate the missing coefficient of the variable *ski* in Model 1; use the output of Model 1.
- (b) (2 Points) For each of the variables *visits* and *feesum* explain if they are insignificant and if the variable is insignificant, explain why; use the output of Model 1. Discuss whether the coefficients for the independent variables *visits* and *feesum* make sense in Model 1.

Years ago, Steven took a class and he remembers that his professor told him that it is a good custom to go back to the data and plot it. Figure 1 shows the dependent variable *expsum* and the independent variable *visits* for each data point.



**Figure 1**

(c) (2 Points) Do you suspect that a linear equation describes the relationship between *visits* and *expsum*. Given your answer above, how would you improve the model, if at all.

Before Steven could make these changes, he was interrupted by his colleague who handed him last year's customer expenditure data for the park at Lake Weather. He decided to include the variable *expwea*: expenditures when visiting Lake Weather.

| <i>Regression Statistics</i> |             |
|------------------------------|-------------|
| Multiple R                   | 0.972935477 |
| R Square                     | 0.946603442 |
| Adjusted R Square            | 0.945472159 |
| Standard Error               | 9.304725896 |
| Observations                 | 242         |

**(Model 2)**

|           | <i>Coefficients</i> | <i>Standard Error</i> | <i>t Stat</i> | <i>Lower 95%</i> | <i>Upper 95%</i> |
|-----------|---------------------|-----------------------|---------------|------------------|------------------|
| Intercept | -0.534059455        | 1.737704045           | -0.307336256  | -3.957452594     | 2.889333683      |
| visits    | -0.346415744        | 0.068534161           | -5.054643404  | -0.481432622     | -0.211398865     |
| ski       | 2.331074143         | 1.282283939           | 1.817907932   | -0.195110941     | 4.857259227      |
| income    | 0.582727703         | 0.377174043           | 1.544983578   | -0.160330377     | 1.325785783      |
| feesum    | -6.084416517        | 2.722804123           | -2.234614112  | -11.44852267     | -0.720310367     |
| expwea    | 0.979709191         | 0.016795002           | 58.3333779    | 0.946621914      | 1.012796467      |

Additionally, Steven calculates the correlation.

| <b>Table 1</b> | <i>expsum</i> | <i>visits</i> | <i>ski</i> | <i>income</i> | <i>feesum</i> | <i>expwea</i> |
|----------------|---------------|---------------|------------|---------------|---------------|---------------|
| expsum         | 1             |               |            |               |               |               |
| visits         | -0.2431       | 1             |            |               |               |               |
| ski            | 0.2664        | 0.0427        | 1          |               |               |               |
| income         | 0.3024        | -0.1375       | 0.2713     | 1             |               |               |
| feesum         | -0.0575       | 0.2210        | 0.0093     | -0.0496       | 1             |               |
| expwea         | 0.9676        | -0.1632       | 0.2491     | 0.2732        | -0.0047       | 1             |

Steven looks at both the model and the correlation table and finds ways for improvement. In particular, he wants to remove a variable from the model.

- (d) (2 Points) Explain which variable could be removed from the model first; use the output of Model 2 and Table 1.
- (e) (2 Points) Discuss whether there is multicollinearity between the independent variables.
- (f) (2 Points) Write explicitly the multiple linear regression equation describing expenditures at Lake Summer corresponding to Model 2.
- (g) (3 Points) Explain which variable is the most useful in predicting expenditures at Lake Summer; use specific numbers from the output of both Model 1 and Model 2.



#### **Problem 4 (15 Points)**

Datatronics is a consumer analytics firm that offers licenses for two different types of software packages: DAP (Data Analytics Package) and DMP (Data Modeling Package). While the first package provides data analytics tools for clients, the second focuses on modeling support and aids clients' decision making process.

Licensing agreements involve setup and initial support that Datatronics must provide to its customers. The company has two different customer support centers for providing assistance: one in the Philippines (P) and another in the United States (U) to serve its customers.

The company is considering to offer at most 750 licenses of DAP (which is the demand for DAP in the next quarter), and at most 950 licenses of DMP (which is the demand for DMP in the next quarter). Because of the limited customer support personnel available, U can only support up to 800 licenses (of either kind), and P is limited to supporting 1,000 licenses (of either kind). The two facilities employ different workforce which translates into different customer hour requirements, as well as earnings per serviced customer. The relevant information is summarized in the table below. Labor is measured in hours. For example, customer support of 1 DAP client from the center at P requires 30 hours, and the total amount of labor available over the next quarter will be 17,500 hours.

| Service Center<br>Country | Software | Earnings<br>(per license) | Labor in hours<br>(per license) | Available labor |
|---------------------------|----------|---------------------------|---------------------------------|-----------------|
| P                         | DAP      | \$1000                    | 30                              | 17,500          |
|                           | DMP      | \$1350                    | 40                              |                 |
| U                         | DAP      | \$600                     | 20                              | 15,000          |
|                           | DMP      | \$800                     | 20                              |                 |

Datatronics has recently hired Anne, a undergrad for its internship program. Anne, has formulated a linear program and solved it using Excel's solver tool which gave the following output. (Some of the output is intentionally left blank)

Note that the variable UDAP and UDMP stand for the DAP and DMP licenses that should be serviced from USA and PDAP and PDMP are the licenses that should be serviced from Philippines.

Variable Cells

| Cell    | Name          | Final Value | Reduced Cost | Objective Coefficient | Allowable Increase | Allowable Decrease |
|---------|---------------|-------------|--------------|-----------------------|--------------------|--------------------|
| \$B\$8  | UDAP Variable | 0           | -183.3333333 | 600                   |                    |                    |
| \$B\$9  | UDMP Variable | 750         | 0            | 800                   | 1E+30              | 183.3333333        |
| \$B\$10 | PDAP Variable | 316.6666667 | 0            | 1000                  | 12.5               | 137.5              |
| \$B\$11 | PDMP Variable | 200         | 0            | 1350                  | 183.3333333        | 16.66666667        |

Constraints

| Cell    | Name                  | Final Value | Shadow Price | Constraint R.H. Side | Allowable Increase | Allowable Decrease |
|---------|-----------------------|-------------|--------------|----------------------|--------------------|--------------------|
| \$B\$16 | DAP demand Constraint | 316.6666667 | 0            | 750                  | 1E+30              | 433.3333333        |
| \$B\$17 | DMP demand Constraint |             | 16.66666667  | 950                  | 237.5              | 200                |
| \$B\$18 | U labor Constraint    | 15000       | 39.16666667  | 15000                | 1000               | 4750               |
| \$B\$19 | P labor Constraint    | 17500       | 33.33333333  | 17500                | 13000              | 9500               |
| \$B\$20 | P capacity Constraint | 516.6666667 |              | 1000                 | 1E+30              | 483.3333333        |
| \$B\$21 | U capacity Constraint | 750         | 0            | 800                  | 1E+30              | 50                 |

- (3 Points) Write the linear constraint corresponding to the DAP demand and linear constraint corresponding to the capacity constraint in the US.
- (3 Points) What is the optimal number of DAP licenses that should be assigned to the servicing center in the US?
- (3 Points) Is the demand constraint for DAP binding? Is the demand constraint for DMP binding?
- (3 Points) What is the shadow price associated with the capacity constraint at the support center in the Philippines?
- (3 Points) Datatronics can contract for 1,000 hours of additional labor in the Philippines, at a cost of \$29 per hour, including benefits, overhead, etc. Is this worth doing?

### **Problem 5 (20 Points)**

Donatello is graduating and his friends, Leonardo, Rafael, Michelangelo, and Splinter are coming for his graduation ceremony.

They will arrive a day early, and will have time to tour the city. Donatello decided to plan a fun day in the city for them. He began by composing a list of attractions around the city:

| <b>Attraction</b>  | <b>Approx. time</b> | <b>Cost</b>                                    | <b>Category</b>  | <b>Level of fun</b> |
|--|---------------------|--|------------------|---------------------|
| <b>Freedom Trail</b>                                     | 180 minutes         | free of charge                                 | Outdoor activity | 7                   |
| <b>Boston Public Garden</b>                              | 80 minutes          | free of charge                                 | Outdoor activity | 10                  |
| <b>Charles River Esplanade</b>                           | 40 minutes          | free of charge                                 | Outdoor activity | 6                   |
| <b>Boston Tea Party Ships &amp; Museum</b>               | 150 minutes         | \$28 per ticket (\$140 for the whole gang)     | Museum           | 11                  |
| <b>Back Bay</b>  | 60 minutes          | free of charge                                 | Outdoor activity | 6                   |
| <b>Old North Church</b>                                  | 30 minutes          | free of charge                                 | Tour             | 4                   |
| <b>Boston Duck Tours</b>                                 | 80 minutes          | \$39.5 per ticket (\$197.5 for the whole gang) | Tour             | 8                   |
| <b>Samuel Adams Brewery</b>                              | 60 minutes          | free of charge                                 | Tour             | 12                  |
| <b>Museum of Fine Arts</b>                               | 150 minutes         | \$25 per ticket (\$125 for the whole gang)     | Museum           | 9                   |
| <b>John F. Kennedy Presidential Museum &amp; Library</b> | 150 minutes         | \$14 per ticket (\$70 for the whole gang)      | Museum           | 7                   |
| <b>Museum of Science</b>                                 | 150 minutes         | \$25 per ticket (\$125 for the whole gang)     | Museum           | 9                   |
| <b>Harvard University</b>                                | 60 minutes          | free of charge                                 | University       | 5                   |

Unfortunately, they will not have time to visit all of the attractions, as they only have 10 hours. Since Donatello never took this class, he came to you for help.

- (a) (4 points) Formulate the problem as a discrete linear optimization problem to maximize the total fun during the limited time available. What are the decision variables? What is the range for each variable? What is the objective function? What are the constraints?
- (b) (8 points) Donatello sent the itinerary to his friends, and received a list of requests. Model each of the requests listed below as linear constraints
- (2 points) Rafael said that if they go to more than 3 museums, then they have to go to the Samuel Adams Brewery.
  - (2 points) Michelangelo loves to play outside, and therefore, asked that in total they will spend at least 2 hours in outdoor activities.

- (2 points) Splinter realized that the costs are getting high. He asked you to make sure that they are not spending more than \$300 (in total for Donatello and his friends).
- (2 points) Leonardo asked to visit at least one university or at least two museums.

You successfully modeled all of the requests, and shared the itinerary with your classmate. He mentioned a website that had a better estimation of the time each attraction takes. In particular, this website modeled the time at each attraction as a normal random variable, and provides the mean and standard deviation of each attraction, as well as the correlation between the different attractions. The data from the website is provided below.

| <b>Attraction</b>  | <b>Mean</b>    | <b>Standard deviation</b> |
|--|----------------|---------------------------|
| <b>Freedom Trail</b>                                     | 180<br>minutes | 20                        |
| <b>Boston Public Garden</b>                              | 80 minutes     | 30                        |
| <b>Charles River Esplanade</b>                           | 40 minutes     | 30                        |
| <b>Boston Tea Party Ships &amp; Museum</b>               | 150<br>minutes | 15                        |
| <b>Back Bay</b>  | 60 minutes     | 60                        |
| <b>Old North Church</b>                                  | 30 minutes     | 5                         |
| <b>Boston Duck Tours</b>                                 | 80 minutes     | 5                         |
| <b>Samuel Adams Brewery</b>                              | 60 minutes     | 5                         |
| <b>Museum of Fine Arts</b>                               | 150<br>minutes | 100                       |
| <b>John F. Kennedy Presidential Museum &amp; Library</b> | 150<br>minutes | 50                        |
| <b>Museum of Science</b>                                 | 150<br>minutes | 50                        |
| <b>Harvard University</b>                                | 80 minutes     | 20                        |

(The correlation table is at the end of the question)

- (c) (4 points) If the itinerary includes the Freedom Trail, Boston Public Garden and Charles River Esplanade, what is the standard deviation of the length of the trip? Add a (maybe non-linear) constraint that assures that the standard deviation of the total trip is at most 60 minutes.

You used your favorite solver, and solved the problem. After looking at your suggested schedule, Donatello realized that he forgot to account for the distance between the different attractions. He gave you the distances table below.

(Note that the distances table can be found at the end of the question)

(d) (4 points) Rewrite your constraint from part (a) to account for the traveling time. Make sure to start your day and end it at the Marriott Cambridge hotel. What are the decision variables? What is the range for each variable? What are the new constraints?

**To save some time, please consider only the first 5 attractions (Freedom Trail, Boston Public Garden, Charles River Esplanade, Boston Tea Party Ships & Museum, and Back Bay) when you write your answer to part (d).**

Hints:

- The variables should be in the form of  $X_{i,j}$  that would indicate that they went from attraction  $i$  to attraction  $j$ .
- Make sure that you are leaving the hotel, and that you are going back to the hotel.
- Make sure that if you go to an attraction, you also leave it.
- Don't forget to link the new variables to the old ones.



Distances table

|                          |                          |               |                      |                         |                                 |          |                  |                   |                      |                     |                           |                   |                    |
|--------------------------|--------------------------|---------------|----------------------|-------------------------|---------------------------------|----------|------------------|-------------------|----------------------|---------------------|---------------------------|-------------------|--------------------|
| Marriott Cambridge hotel | Marriott Cambridge hotel | Freedom Trail | Boston Public Garden | Charles River Esplanade | Boston Tea Party Ships & Museum | Back Bay | Old North Church | Boston Duck Tours | Samuel Adams Brewery | Museum of Fine Arts | J. F. K. Museum & Library | Museum of Science | Harvard University |
| 0                        |                          |               |                      |                         |                                 |          |                  |                   |                      |                     |                           |                   |                    |
| 8                        | 0                        |               |                      |                         |                                 |          |                  |                   |                      |                     |                           |                   |                    |
| 7                        | 6                        | 0             |                      |                         |                                 |          |                  |                   |                      |                     |                           |                   |                    |
| 6                        | 17                       | 10            | 0                    |                         |                                 |          |                  |                   |                      |                     |                           |                   |                    |
| 11                       | 20                       | 20            | 25                   | 25                      | 0                               |          |                  |                   |                      |                     |                           |                   |                    |
| 10                       | 20                       | 25            | 17                   | 13                      | 0                               |          |                  |                   |                      |                     |                           |                   |                    |
| 15                       | 23                       | 25            | 27                   | 10                      | 17                              | 0        |                  |                   |                      |                     |                           |                   |                    |
| 13                       | 20                       | 15            | 14                   | 14                      | 5                               | 17       | 0                |                   |                      |                     |                           |                   |                    |
| 20                       | 20                       | 20            | 20                   | 20                      | 20                              | 0        |                  |                   |                      |                     |                           |                   |                    |
| 12                       | 40                       | 8             | 30                   | 14                      | 24                              | 18       | 21               | 20                | 0                    |                     |                           |                   |                    |
| 20                       | 20                       | 20            | 20                   | 20                      | 20                              | 20       | 20               | 20                | 20                   | 0                   |                           |                   |                    |
| 23                       | 25                       | 25            | 30                   | 10                      | 11                              | 24       | 12               | 20                | 11                   | 20                  | 0                         |                   |                    |
| 12                       | 15                       | 12            | 12                   | 19                      | 13                              | 21       | 17               | 20                | 13                   | 20                  | 19                        | 0                 |                    |

### **Problem 6 (15 points)**

The primary goal of the course has been to teach you some important analytics tools that we believe can make a difference in making decisions based on data. We would like to ask you to think back to your current or last job before coming to this class. Please identify a project, activity, task, or assignment that you worked on at that job where you would now have analyzed the problem differently given the knowledge you acquired from this semester.

- a) (4 points) Describe the project/activity/assignment.
- b) (4 points) Describe either the data that you had available for the project or that you now wish you had developed in order to complete the project.
- c) (7 points) What modeling tool(s) from this class would you have used on this project, and why do you think these tools would have been effective?

Try to be **as concise as possible**; we strongly suggest that you limit your answer to **each of these questions to approximately one paragraph**. Do not spend way too much time on this problem. (**Note:** If your job involved tasks for which the use of data and quantitative analysis was not relevant at all, then answer this question by instead discussing one of your possible job opportunities for an upcoming engagement)