



DSCI6840 Final Exam Spring 2024

Submission Instructions:

1. This Word document. Write all your answers in this document. Only this Word document is to be graded for Case 1, 2, 3, and 5.
2. An Excel file for Case 4. This Excel file is also to be graded.
3. Supporting documents:
 - 1) One single R script breaking down into Four sections for Case 1, Case 2, Case 3, and Case 5.
 - 2) Your .txt file for Case 2.

Deadline: 6PM, May 12th, Sunday.

Late submission will be graded as 0.

Visual aesthetics and accuracy are the two key KPIs affecting your grade in all five cases.

Notes:

Final Exam is individual work. This exam is to further deepen your skills to apply what we have learned to solve real life procurement and SCM projects.

If you work with others on the final Exam, that is fine. But make sure you write your own answers and do your own analysis. Same answers with the exact everything from different individuals will get both parties penalized.



Case 1. Dovrex Spend Analysis (20%)

Instructions:

Read the case. Read the associated Excel file. Read each Excel worksheet. Make sure you understand the meaning of each column. (This instruction applies to all five cases)

Use the file "[1. Dovrex Spend Analysis.xlsx](#)" to enhance the statistical and data visualization solutions for Dovrex.

Prep:

- Know your data:
- Remove **NA** from quantity and value column.
- Remove "**Division Details Unavailable**" observations from the corresponding column.
- Save the new file.
- Use this new file in all the tasks.

Task 1:

Create a spend tree for Level 1 and Level 2 products only. Attach your spend tree in below. Use a separate landscape page if your tree is too wide.

[Briefly explain your tree and findings.](#)

Task 2:

Create a butterfly spend chart for Top Suppliers contributing to the cumulative 50% spend measured by Level 2 category only. Attach your butterfly chart in below.

[Briefly explain your chart and findings.](#)

Task 3:

What are the top 10 suppliers in terms of degree and betweenness? Attach your result in below.

[Explain what these two terms stand for in plain language.](#)

Task 4:

Make a proper network graph between suppliers and buyers assigning size of nodes as "Spend" and width of edges as "Quantity". You need to know your data to find which column is supplier and which column is buyer.

Do NOT show labels for your nodes. Use `vertex.label=NA` in your code wherever applicable.

Only do analysis for spend > 50,000 from your spend column. Attach your graph in below.



What is your observation? Explain to a business audience in business language about your chart.

Task 5:

Make a community detection graph assigning size of nodes as “Spend” and width of edge as “Quantity”. Only do spend for spend > 1 million from your spend column.

Keep all labels shown (you may see a litter clutter on your graph but that’s okay) by removing `vertex.label=NA` from your code. Attach your graph in below.

What is your observation? Explain to a business audience in business language about your findings.

Case 2. Altimus Supplier Selection (15%)

Instructions:

Read the case to understand the case background. Read Exhibit 2 in details. Make sure you understand Exhibit 2 and its relationship with the case.

Task:

Use AHP to analyze Exhibit 2 – which country/supplier should Altimus source from?

Do this in R. Paste your chart generated by R in below.

What is your observation? Explain to a business audience in business language.

Case 3. Artificial Intelligence Indian Railways (20%)

Task 1:

Use the file [3. Artificial Intelligence Indian Railways.xlsx](#) to create a word cloud for the Nonstock items Indian Railways has purchased. Attach your word cloud in below.

Requirement: Remove some apparently unnecessary words when creating your word cloud.

Explain your results of Word Cloud to a business audience, such as what is a word cloud, how the word cloud was constructed, what is the meaning of different size and color of words, and etc.

What is your observation regarding how Indian Railways named their products from your word Cloud?



Task 2:

The case mentioned clustering texts to improve search accuracy. We did not have time in class to go over this. But it simply requires a couple line of codes.

Copy past the following codes to R and run. Paste your cluster plot in below.

```
#### CLUSTERING ANALYSIS to DETECT THEMES ####
# get term frequency (number of times a term appears in the doc/total terms)
tdm.tf <- weightTf(corpdtm) # corpdtm is the file from your previous step using
TermDocumentMatrix function.

# Remove words using term frequency
tdm.tf <- removeSparseTerms(tdm.tf, 0.98) # lower number keep less words
tf.matrix <- as.matrix(tdm.tf)

# Create distance matrix
dist.matrix = proxy::dist(tf.matrix, method = "euclidean")

# plot clusters
hfit <- hclust(d = dist.matrix, method = "ward.D")
plot(hfit, hang = -1)
```

Read your cluster plot carefully. Roughly, how many big clusters do you see from this Hierarchical Clustering algorithm? How many total clusters at the lowest bottom level do you see?

How many lowest level clusters from this analysis make sense to you? Use one or two lowest level clusters that make sense to you and explain the rationale to Indian Railways.

Case 4. Highway Expansion (15%)

Task 1:

Use the file [4. Highway Expansion_Template.xlsx](#) to find an optimal solution by allocating supplier capacity to demand of Granular A and B in the six areas. One worksheet “Template” was already created and all necessary setup was completed to facilitate your work.

Answer the following questions:

1. What are your total costs using Solver compared with the current proposed total costs? How much did you save, if applicable?
2. Does each supplier exceed the 50% rule?
3. Does your Solver allocation allow 3% discount rule for Supplier 3? How much is your total cost after applying the 3% discount on Supplier 3?



Case 5. Indian Railways Part 1 and Part 2 (30%)

Task 1:

Use the file “[5. Part1.xlsx](#)” to visualize the electricity demand for the two stations of Barasat TSS and Sonarpur FP.

What are your observations from your chart?

Task 2:

Use file “[5. Part2_train.xlsx](#)” and “[5. Part2_test.xlsx](#)” to create a Total Loss Demand Histogram Chart (including all passenger classes) for the training and testing data in Part 2 of the case study. You can use Excel to do this task.

What are your observations from your histogram chart?

Task 3:

Test forecasting accuracy using HW, ETS, and ARIMA for Barasat TSS and Sonarpur FP (keep the last 12 months data as your testing dataset) using files of “[5. Barasat2020.csv](#)” and “[5. Sonarpur2020.csv](#)”. Plot the forecasting result in one single chart and attach your chart in below.

Then, select the best performing model to forecast electricity demand in 2020 for Barasat TSS and Sonarpur FP.

Attach your forecasting chart and your forecasted values for 2020 for the two stations in below. Briefly explain why did you select your specific model to do your forecasting.

Task 4:

To predict special train demand, first train your model using file “[5. Part2_Train.xlsx](#)” by comparing the model performance among linear regression, decision tree, and random forest.

Which model would you choose to predict special train demand? Why?

Task 5:

Use the model selected from the previous step to test your model performance on the data “[5. Part2_Test.xlsx](#)”.

How is your model performing in predicting special train demand? What is your recommendation to trigger the scheduling of a special train? What kind of data/variables should you tell Indian Railways to keep collecting for future model use?