# SAS Optimization Challenge: Case Study

## Title: HELP YOUR OPTIMIZATION PROFESSOR!!!!!

## Introduction, Data Details, and Overview

## Introduction

Your optimization professor is in a major pickle, and he needs your team's help to develop a final exam for his Decision Science course this upcoming semester.

For years, your professor has taught a Decision Science/Optimization course to undergraduate and graduate students, but this year's cohort is his ***most advanced group yet***. Every student enrolled is not just academically gifted, but they all have an insatiable curiosity and desire to be the best decision scientist possible.

In fact, they're so committed to learning that they're currently attending optimization classes on the weekend!

Knowing that Operations Research is the highest paying degree for graduates in the United States in 2023, your professor knows he'll need to craft both a challenging curriculum and final exam to prepare his students to be successful decision scientists upon graduation.

The problem is that your professor has completely lost track of time. He just got back from vacation and realized he doesn't have time to develop both the curriculum ***and*** the final exam for his upcoming course, so he's opting for a divide-and-conquer approach. He'll focus on developing the curriculum and course content. You and your team will focus on developing the final exam.

The final exam that you'll develop needs to contain **fifty (50) questions** from a large question bank developed and maintained by your professor over the years. The question bank currently contains one hundred and fifty-four (154) total questions.

Developed by: Jay Laramore, Sr. Analytical Training Consultant, SAS Institute

# Data Details

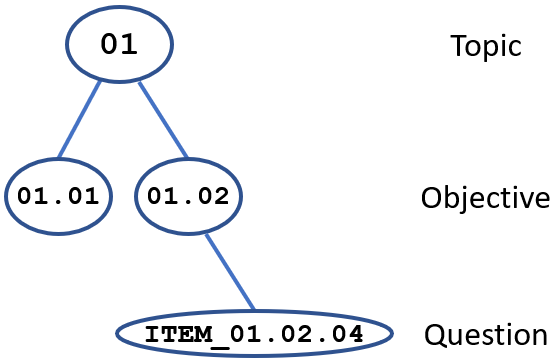
The first text file, **question_bank.txt**, contains the SAS® DATA step and datalines code for all 154 exam questions in the question bank. Every question rolls up to a specific *topic objective* (i.e., the **Objective** column), which rolls up to a specific *topic* (i.e., the **Topic** column).

For example, see question `ITEM_01.02.04` below.

| Question | Topic | Objective |
|---|---|---|
| ITEM_01.02.04 | 01 | 01.02 |

Question `ITEM_01.02.04` is the 4th question in *topic objective* 01.02, and 01.02 is the 2nd *topic objective* in *topic* 01.

Below is a picture of the hierarchical structure of question `ITEM_01.02.04`.




The question bank data show that t*opic objective* 01.02 contains eight (8) total questions, and *topic* 01 contains eighteen (18) total questions across two topic objectives: 01.01 and 01.02.

Question `ITEM_01.02.04` was chosen arbitrarily to illustrate the relationship among the **Question**, **Topic**, and **Objective** columns. The description above can be applied to every topic, objective, and question in the question bank data set.

The **Correct** column is the percentage of students who have correctly answered the question on previous exams. This value is inversely proportional to a question's difficulty, since a question's difficulty drops as this value approaches 100%.

For example, question `ITEM_01.02.04` has been answered correctly 85.7% of the time.

| Question | Topic | Objective | Correct |
|---|---|---|---|
| ITEM_01.02.04 | 01 | 01.02 | 0.857 |

The **Frenemy_Group** column groups questions together that have been identified as being extremely similar to one another. The term "frenemy" is the combination of the word "friend" + "enemy". Every frenemy group contains at least two questions.

A question is either assigned to one frenemy group, or not assigned to any frenemy group. In other words, a question isn't part of multiple frenemy groups.

For example, questions `ITEM_01.02.03` and `ITEM_01.02.04` are in the same frenemy group, 3.

| Question | Frenemy_Group | Topic | Objective | Correct |
|---|---|---|---|---|
| ITEM_01.02.03 | 3 | 01 | 01.02 | 0.817 |
| ITEM_01.02.04 | 3 | 01 | 01.02 | 0.857 |

Questions that are not in a frenemy group have a missing value in the **Frenemy_Group** column. There are forty (40) distinct frenemy groups containing 83 total questions. The remaining seventy-one (71) questions do not belong to a frenemy group.

---

The second text file, **objective_questions.txt**, contains the SAS® DATA step and datalines code for all sixteen (16) *topic objectives*. For each *topic objective*, the **Objective_Questions** column contains the total number of questions required to appear on the final exam.

For example, *topic objective* 01.02 must contain three (3) questions on the final exam.

| Objective | Objective_Questions |
|---|---|
| 01.02 | 3 |

Summing the **Objective_Questions** column across all *topic objectives* in this data set contains the desired number of final exam questions, fifty (50).

# Evaluation Criteria

This case study contains five tiers (Tier 1-5). Each tier builds on the model developed in the previous tier, so you're not advised begin the next tier before completing the current one.

You do not have to complete all five tiers, but you should attempt to complete as many tiers as you can.

You are required to submit the information below before advancing to the next tier, however ***only the evaluation from your highest completed tier will be evaluated for the competition***.

Therefore, you're encouraged to submit the information below once you feel confident in your solution and can defend your model's results.

---

Each tier will be evaluated based on the following criteria:

**Accuracy (0 - 60 points)**: How well does the model accomplish its objective while satisfying the conditions outlined in the Modeling Details section?

The Modeling Details section is provided in the PDF for each tier.

**Data/Model Separability (0 – 40 points)**: Is the OPTMODEL code production-ready?

---

To evaluate these two criteria, provide one (1) SAS program (.sas) containing all code, including the DATA step/datalines code to create the data sets, the OPTMODEL code, and any additional data processing, if applicable. The reviewer should be able to run your code ***as-is, without error*** to evaluate these two criteria.

## Frequently Asked Questions

**Q**: Can I use external resources (e.g., Google, ChatGPT, SAS Documentation, etc.) to help formulate and solve the model?

**A**: Yes. In the real world, when you are assigned to work on a key project, your client will never say to you, "Please solve our problem, but only use your graduate school textbook as a resource". The world is your oyster. Use it.

**Q**: Do I need to do any preprocessing of the two data sets before reading them into the OPTMODEL procedure?

**A**: No. For the first tier, Tier 1, the two data sets are ready to be read into the OPTMODEL procedure. Subsequent tiers may require preprocessing, but you'll be provided with all preprocessing code and instructions, so prior knowledge of SAS DATA step syntax ***is not required*** to successfully complete any tier.