**INFO 3440 - Project Summary**

**Winter 2023**

**Description**

Who wants to win at Daily Fantasy Sports on Draftkings? Everybody? Anybody? Well, let’s be honest, this project isn’t going to make you win, but it is a fun **optimization** problem! And that’s really the point in this class.

Imagine that we were interested in joining one of the Draftkings’ National Hockey League (NHL) contests on Feb 11, 2023. There were fourteen real NHL games played amongst twenty-eight teams. The fantasy contest involves selecting a collection of players (a lineup) in these fourteen games and you “play” against other people that (potentially) selected different lineups. The winner of the contest is the person whose team accumulated the most fantasy points. The fantasy points are accumulated according to a set of rules (see Figure 1) that are related to how the players perform in their real-life games. Selecting a lineup is subject to a fictitious salary cap and several additional constraints that are outlined below. The **goal** of this project is to select an “optimal” lineup. What is “optimal?” That is for you to decide!

**Data**

Here are the data that you need to solve this problem. I will provide a data dictionary on Canvas related to each data set describing the column names.

* DKSalaries.csv – Data from Draftkings website including player names, teams, salary, position, and average points per game.
* Actual game data from Feb 11, 2023
  + skaters-actual.csv – Data from each of the seven games on Feb 11 related to the skaters (i.e., non-goalies).
  + goalies-actual.csv – Data from each of the seven games on Feb 11 related to the goalies.
* Season Statistics
  + skaters.csv – Data from the NHL season (October 7, 2022 through Feb 10, 2023) related to skaters.
  + goalies.csv - Data from the NHL season (October 7, 2022 through Feb 10, 2023) related to goalies.

**Objectives**

1. Use python to apply optimization models and solution techniques learned in class to solve a real-world problem.
2. You will have to do some exploratory data analysis using the historical data to carefully consider your objective function coefficients.
3. You will solve at least two optimization problems for this project. One problem will involve selecting a lineup based on historical information. The second problem will involve selecting the optimal lineup based on how the players performed during their actual games.
4. Write a report summarizing your results.

**Constraints**

1. The fantasy team must have nine players.
2. The salary of your nine players may not exceed $50,000.
3. You need players from at least three different teams.
4. The makeup of the team in terms of positions is:
   1. Two Centers (C),
   2. Three Wings (any combination of left (LW) or Right (RW)),
   3. Two Defenders (D),
   4. One Utility (C, W, or D), and
   5. One Goalie.

**Deliverables**

1. Project Jupyter notebook with embedded final report as an html file.

**Notes**

1. You absolutely do not need to create an account on Draftkings, or any other website, to solve this problem.
2. If you do not want to work on this project, you are welcome to select your own. See below for a description of a project proposal. I will have to approve your project prior to you beginning work. I don’t want you to waste your time if the project is not suitable for this class.



Figure . Draftkings scoring for an NHL contest

**Data Dictionary**

File: DKSalaries11Feb2023.csv

Columns/Variables:

* + - 1. Position – The player’s position on the ice.
      2. Name + ID – Player name plus their ID in Draftkings
      3. Name – Player name
      4. ID – Player ID in Draftkings
      5. Roster Position – The player’s position in the Draftkings game. Note that players that are, e.g., a left wing (LW) in real life are just listed as a wing (W) in the game. Players with a position other than goalie (G) can also be considered for the utility (Util) position.
      6. Salary – Player salary in the game. Remember that each “team” in Draftkings may not exceed $50,000 in total salary.
      7. Game Info – General information related to the game including teams that are playing, time of game, etc.
      8. TeamAbbrev – The player’s team name (abbreviated).
      9. AvgPointsPerGame – The average Draftkings points for the player up to this point in the season.

File: skatersactual11Feb2023.csv (skatersseason2022-2023.csv)

Columns/Variables:

1. player – player name
2. team – The player’s team abbreviation
3. pos – player position
4. gp – games played in the data set (note that this should only be one in this data set)
5. g – goals in the game (Note: a hat trick is 3 or more goals!)
6. a – assists in the game
7. p – points in the game (g + a)
8. pim – player’s penalty minutes
9. x – the player’s +/- rating for the game (essentially the # of goals that player’s team scored while he was on the ice minus the # of goals that the player’s opponent scored while he was on the ice)
10. toi\_gp – time on ice per game played
11. TOIES – time on ice in even strength (e.g., both teams have the same number of skaters on the ice) situations
12. TOIPP - time on ice in power play (e.g., this skater’s team has more skaters on the ice than the opposition) situations
13. TOISH - time on ice in shorthanded (e.g., this skater’s team has less skaters on the ice that the opposition) situations
14. ESG - even strength goals
15. PPG – power play goals
16. SHG – shorthanded goals
17. GWG – game winning goals
18. OTG – overtime (non-shootout) goals
19. SOG – shootout goals
20. ESG - even strength assists
21. PPG – power play assists
22. SHG – shorthanded assists
23. GWG – game winning assists
24. OTG – overtime (non-shootout) assists
25. ESG - even strength points
26. PPG – power play pointss
27. SHG – shorthanded points
28. GWG – game winning points
29. OTG – overtime (non-shootout) points
30. PPPPCT – percentage of points scored on the power play
31. GGP - player goals per games played
32. AGP – player assists per game played
33. pgp – player points per games played (equal to p in actual file)
34. s – shots on goal
35. shpct – shooting percentage (goals/shots)
36. hits
37. blocked shots

File: goaliesactual11Feb2023.csv (and goaliesseason2022-2023.csv)

Columns/Variables:

1. name – player name
2. team – player team
3. gp – games played in the data set (note that this should only be one in the actuals data set)
4. gaa – goals against average (goals allowed per 60 minutes played)
5. svpct – save percentage
6. w – wins
7. l – losses
8. ga – goals allowed
9. sv – saves
10. sog – shots on goal faced
11. so – shutouts
12. toi – time on ice
13. g – goals in the game
14. a – assists in the game
15. p – points in the game (g + a)
16. pim – player’s penalty minutes

**Project Proposal**

**(From Professor Keeling’s class)**

**Summary:** Your proposal will be, at most, a one-page document that describes the problem that you intend to work on. The purpose of the proposal is to make sure that you’ve identified (i) a reasonable target application, (ii) an optimization model that is appropriate for the application, and (iii) adequate data sources to ensure your success.

**(i) Target Application.** In your proposal, identify a single application, your reason or motivation for choosing it, and why it is amenable to optimization modeling. You might start by brainstorming about real-world scenarios you’ve encountered that seem amenable to the types of optimization modeling we have done in this class, especially LP and IP. If you’re having trouble getting started, go back through the examples we’ve worked in class and try to think of how you might extend these to situations you’ve experience first hand.

**(ii) Optimization Model.** Once you’ve targeted a specific application, write down the corresponding decision variables, objective function and types of constraints. Be as detailed as you can -- if you can write down the exact mathematical formulation, great! If not, describe in English as specifically as you can. What kind of model do you intend to use (e.g., product mix, covering constraints, facility location, logistics)? Will you be using binary constraints? Integer constraints? Consider writing a “toy” model in Excel with a small number of decision variables and constraints, and imaginary data, to help flesh out what your full scale model might look like.

**(iii) Data.** Your model will need data in the form of objective coefficients, constraint coefficients, and constraint RHS values. Where will these data come from? Will you, compile data from the web, use existing data you have or can collect, or create synthetic (fake) data? What will you have to do to get the data into a format that you can read into a pyomo model? Does this seem realistic for you to accomplish?

**Proposal Evaluation.** The purpose of this project is to get you as close to building a “real-world” optimization model as possible. As such, the closer you get in terms of the target application, modeling techniques and sophistication, and realistic data sources, the more favorable your evaluation. Don’t be too ambitious because you might make the project too difficult to execute. But try to be creative in coming up with an application area that interests you and seems feasible as a project that can be completed over the course of 3 weeks.

NOTE: It is understood that as you actually work on the final parts of your project that your topic may morph into something different – that is OK.