

# DSO 570 Final Project Assignment Description

## Learning Objectives:

1. Model a complex business case as a precise mathematical problem, and clearly describe the assumptions.
2. Use data analysis or simulation to rigorously identify opportunities for improvement to the status quo.
3. Formulate an appropriate optimization problem that can be numerically solved using Python and Gurobi.
4. Implement the optimization formulation in Python, and supplying appropriate test inputs.
5. Communicate the findings using both technical and business language.

## Contents

1. Problem
2. Relevant Issues
3. Deliverables
4. Grading Rubrics
5. Ideas for Getting Started

### 1. Problem

Frito-Lay is a subsidiary of PepsiCo that manufactures and sells corn chips, potato chips, and other snack foods. It owns 35 brands, including Lay's, Fritos, Cheetos, Doritos, Ruffles, Sun Chips, Tostitos. Each brand has many different flavors, packaging, sizes. All together there are over 1400 SKUs (Stock Keeping Units). The company divides the US market into 10 regions, and each region has its own sales team and would like to consolidate the SKUs for each region. **The purpose of this project is to develop data analytical insights regarding the choice of SKUs for each region, as well as to create a prototype optimization tool that can help regional sales directors better select SKUs.** You should provide a quantitative estimate of the potential benefit from applying the insights you found or from using your tool.

The reason for limiting the number of SKUs are as follows. The company delivers products to retailers in two different formats:

- Large format (60% of the business): delivering entire cases of the same SKU to a retailer. Typically, these are large retailers like Walmart.
- Small format or "eaches" (40% of the business): delivering individual packages of items to a retailer. Typically, these are small stores with limited shelf-space.

The company views its own competitive edge as being able to service small stores across the nation, so the small format business is very important. However, the eaches module in distribution centers can support 250 distinct SKUs at a time. Effectively, each sales region can only support about 250 SKUs in small format. The limit on the number of SKUs for large format is more relaxed, but there is still a limit because space is limited in warehouses and there is managerial overhead and reduced inventory efficiency when there are many SKUs.

#### 1.1 Status Quo

The current process by which directors of regional sales teams select SKUs is highly manual. There is a spreadsheet-based tool that ranks current items based on velocity (speed of sales) and staleness (amount of product past expiration returned by retailers), and helps sales teams

identify SKUs that are not selling well or are being returned a lot. Based on this basic analysis and the regional team's intuition, they remove certain products and add new ones. Considerations include differential tastes of customers in that region, maintaining market share, minimizing cost, maximizing profit margins, curbing competitors, product innovation, and branding. Based on the regional director's decision, all of the distribution centers (DC) in their region would migrate to the new set of SKUs, although smaller warehouses (called bins) serving isolated or rural localities may make their own product selection decisions. This process occurs in a decentralized way across the sales regions. Certain regions may update SKUs twice a year, and others may do it less frequently.

## **2. Relevant Issues**

This is a complex problem with many facets. The following is a partial list of issues that may be related to the question of how to choose the right SKUs for each region.

### **2.1 Differential Customer Taste**

Demographics across the 10 regions are quite different. For example, taste preferences differ by racial groups. See [this visualization](#) or [this one](#) for the geographic distribution of racial groups in the US. As a result, hot and spicy does well in the southwest and midsouth regions, whereas cheesy dominates certain midwest regions. Regions with many demographic groups would need to have enough variety of SKUs that satisfy customers from each group.

### **2.2 Transportation Costs**

One driver of SKU decisions is profit margin, as certain slow selling products may nevertheless be profitable if they have high margins. What differentiates margins across regions is transportation costs. Certain products are only produced in certain plants, and must be shipped to other regions via trucks. Transportation is what differentiates costs for the same product in different regions. Unless there is a clear demand for a product or potential for growth, it may not make sense to be constantly transporting it from across the country, and perhaps a similar locally produced product can be a cost-saving substitute.

### **2.3 Production Costs**

Another aspect of profit margins is production costs, as certain types of products are in general more profitable than others. While detailed margins data is confidential, roughly speaking the margins are driven by the main ingredients, and whether it is manufactured by Frito-Lay itself or manufactured by another company and only packed by Frito-Lays. As a rule of thumb, dips and corn-based products have about a 15% margin, potato-based products 6-8%. Jerky, nuts, and cookies tend to be manufactured by other companies and packed by Frito-Lay, so the margins are only around 5%. Cost of goods sold represent around 33% of the revenue of goods sold.

Production costs are also higher when a plant is near capacity. This is because without machine downtimes, breakdowns happen, which are expensive, and renting additional space when warehouses are at capacity also cost money. You should assume that when a plant is over 95% capacity, there is at least an added 1% of production cost.

### **2.4 Inventory Costs**

For products with highly variable demand but limited production capacity, the company would need to store up sufficient inventory (called safety stock) to meet the demand. Storing inventory requires warehouse space and incurs a cost. Another type of inventory cost is

incurred when transporting many items over long distances, as this take up space in trucks and in the receiving warehouse. It is possible that a better choice of SKUs may reduce the need for high inventory levels and reduce such costs.

## **2.5 Competition**

FritoLays is the leader in salty snacks and would like to maintain its large market share. This requires having a sufficiently diverse offering and having flavors that appeal to all major demographics.

Another reason to include a certain product would be to match competitor offerings. If a competitor is hoping to gain a following in a particular niche segment, it may be a good strategy to offer a similar product just to limit competitor growth and cut into competitor profits. In fact, this may be a good strategy even if the product is not very profitable for FritoLays on its own, as limiting competitor growth allows the company to have higher profit margins in the long run.

## **2.6 Innovation and cannibalization**

A key to FritoLay's sustained growth is the focus on innovation. Frito-Lay as a company is constantly experimenting with introducing new products. Frito-Lay estimates that when it introduces a new SKU, only a fraction of the demand add to overall revenue, while the rest cannibalizes existing Fritolays products. For example, simply introducing a new package size of Fritos may not be increasing total sales, but reducing the sales of other package sizes. At any given time, approximately 1-2% of the total amount by volume, or 20 out of 250 SKUs by item, are new products being tested. As a rule of thumb,

- For a new flavor of an existing product, only about 10% of the demand is truly new, while the rest cannibalizes demand from existing flavors.
- For a new shape, or texture, about 20% is new sales.
- For a minority of new SKUs, which represent true departures from the existing portfolio, perhaps 30% is new sales.

While introducing a new product and having to later pull it out of market incurs cost, it turns out that buying back expired products is cheaper than conducting a detailed market analysis, so this is the preferred channel for product experimentation.

FritoLay may also innovate by acquiring other brands to diversify its portfolio. A recent acquisition is Bare snacks, which expands its presence into the baked fruit and vegetable snacks market. In this case, the company estimates that cannibalization of existing sales would be minimal, since the product targets a different customer segment.

## **2.7 Advertisements and Price Promotions**

While the choice of SKUs depends on customer demand, demand in turn demands on how the company markets the product and offers price promotions. Advertising and Marketing (A&M) decisions are made nationally and are not under the purview of the regional teams. Moreover, for Frito-Lay, A&M is relatively small, as most of the products are already quite established.

One promotional effort regional sales teams engage in is to work with retailers on price promotion campaigns. For example, Ralphs may discount Fritos for a week, and based on negotiation with the Fritolays sales teams, Fritolays may pay for part of the price discount. Retailers differ in their promotional strategies and promotional campaigns. Someone from the sales team, called a RSR, also visits retailers periodically and help them make ordering decisions to restock.

## 2.8 Brand, Reputation and Network Effects

Another reason to include certain slow-moving products is that the company would like to grow in that direction, as building a new brand and gaining a following in a new customer segment take time. The long term benefit of a certain product may not be visible in the short-run. For example, having a footing in one customer segment may make it easier for the company to venture further into this segment in the future, and this is only capitalized years down the road. Moreover, customers influence one another so it may take a while before building up a sufficiently large customer base, before a product can truly take off. (This is one reason FritoLays may be patient with certain new products in the health and wellness segment for example.)

## 3. Deliverables

The following 8 files are due on Blackboard by **Tuesday, April 30 at 6pm PST**.

### 3.1 Project report (.docx or .pdf file)

A 10-15 page report in Word or PDF format, with the main body targeted to a regional sales director, and a technical appendix targeted to an operations research analyst (both parts are included in the page count, but an optional title page is excluded). The report should answer the main questions: **what are the biggest opportunities for improvement and how much are the potential gains from applying your optimization?**

The **main body** of the report should be written in language appropriate for a non-technical but business savvy executive, with minimal mathematical notation or technical jargon. It must contain the following five components:

**1. Executive Summary:** Concise overview of the main insights of your analysis. This includes the biggest opportunity for improvement, the estimated potential gain from your optimization, and your recommendation to the business. Be brief and clear.

**2. Opportunity for Improvement:** What is the main weakness or inefficiency in the current process of selecting SKUs that you can identify, and which you choose to focus on in your project? You must illustrate this weakness or inefficiency either by analyzing existing data, or by conducting a simulation analysis with reasonable assumptions. (Either data analysis or simulation can suffice, but you cannot simply claim the weakness based on words alone; simulation can be used if the data you need is not available.)

**3. Optimization Methodology:** What does the optimization tool (see 3.3) you created do and how can it be used in this business case? This section should describe the optimization tool in enough detail and precision so another analyst can run it on different data of the same format, and understand what are your objectives and constraints. In particular, you should clearly describe

- **i)** At a conceptual level, what are the input data needed for your optimization? (This should match the types of data in the sample input that you provide, see 3.4)
- **ii)** What are the outputs of the optimization? (This should match types of data in the sample output that you provide, see 3.5)
- **iii)** What are your decision variables, objectives, and constraints in words? (You should defer the precise formulation using mathematical notation to Appendix A1.)
- **iv)** How do you envision your tool being used by the end user? You should describe both the command needed to execute your tool, as well as how you imagine the user interacting with the tool.

**4. Optimization Results:** Use easy to understand tables or graphs (not large printouts or screenshots that are hard to read) to illustrate the result of using your optimization tool. In this

section, you need to quantify the potential gains from applying your optimization methodology. Furthermore, you need to provide qualitative intuition on where the gains come from and why your outputs make sense. You should also illustrate what qualitative insights can be learned from using your optimization. For any claims you make, you must rigorously justify them based on sound reasoning and reference to supporting analysis (rather than simply making claims out of thin air). It is important not to oversell your result, and not make bold claims without sufficient justification. You would be graded on the appropriateness of your presentation of results, as well as the rigour of your analysis.

**5. Discussion:** This section should discuss the following

- **i) Appropriateness of Methodology:** Why do you think your proposed methodology is appropriate? In particular, you need to explain the reason behind your choice of input and output data, as well as your choice of objective and constraints. What are potential alternative approaches and why did you choose what you proposed? You should also summarize the main assumptions in words, and acknowledge the main weaknesses of your analysis. (Detailed discussion of technical issues are in Appendix A2.)
- **ii) Final Recommendation:** In light of your analysis, what do you suggest FritoLay to pursue as next steps? Beside stating your recommendations, you must justify them using sound reasoning, while exercising business sense, which involves taking into account other important business issues outside of your mathematical formulation.

The **technical appendix** should be written in language appropriate for a technical audience, using correct mathematical notation and precise technical terms. It include the following two components.

**A1. Mathematical Formulation:** A precise statement of your optimization formulation using correct mathematical notation. You should define the decision variables, the data variables, and use them to express the objectives and constraints.

**A2. Discussion of Technical Details:** Imagine the reader is a saavy analyst who is sceptical of the technical rigour your results, you should use this section to anticipate the issues that the reader sees, address them to the best of your ability. (Potential objections a technically saavy reader may have include independence of random variables, linearity of objective and constraints, lack of statistical significance, confusion of correlation and causation, not accounting for trends, misinterpretation of data, etc.) You should critically examine all of the assumptions underlying your analysis, explain your reasoning behind them, and acknowledge what are the weakest points and the biggest gaps, which you would like to fix if you had more time or resources.

### 3.2 Documentation of all tables and plots (.ipynb file)

A Jupyter notebook which provides sufficient information so that a trained analyst can understand what exactly went into each numerical estimates, tables and plots in your project report, and be able to reproduce them if needed. If you conducted this analyst in Python, your Jupyter notebook should show the code that created them, and explains any technical details or underlying assumptions. You should make this notebook easy to navigate using headers in Markdown cells, and clearly indicate which block of code corresponds to which plot. If your report contained figures or plots created using another program (i.e. R or Excel), you should include a section for those analysis and include a text description of how exactly you created the figure or plot or made your calculations.

### 3.3 Optimization tool (.py file)

A Python module containing your optimization tool, which implements the optimization formulation in your project report, and using which one can reproduce the optimization results. This module should contain a function called “optimize” with at least two input parameters:

- inputFile: the path to the input file.
- outputFile: the path to write the output.

If your code requires multiple input files, then you can update the parameters appropriately. Having this function would allow a user to input your module from Jupyter notebook and run your code without understanding your code.

In addition, the module need to be able to be directly executed from command line. The following example teaches you how to do this. Suppose you write a module called “optimize.py” and you include the above code at the end:

```
if __name__=='__main__':  
    import sys  
    print (f'Running {sys.argv[0]} using argument list {sys.argv}')
```

Then one can run it in command line as follows.

```
ipython optimize.py input.xlsx output.xlsx
```

This would yield the output

```
Running optimize.py using argument list ['optimize.py','input.xlsx','output.xlsx']
```

Instead of printing the arguments supplied by the command line, you can simply call your optimization function.

Your submitted Python module will be graded on correctness and ease of being understood. For the latter, you should include appropriate docstrings to all functions, comments to explain the logic, and reasonable variable names.

### 3.4 Sample input (.xlsx or .csv file)

One or more sample input files that can be fed into your optimization code. Your optimization code should be flexible enough such that one can use it simply by changing the input file, and not have to alter anything in the code. You should avoid hardcoding any parameters into the code, but try to read all parameters from the input files.

### 3.5 Sample output (.xlsx or .csv file)

The result of running your optimization code given the input files you provided. This needs to be reproducible in that running your code again with the input files you submitted should yield the same output file on another computer.

### 3.6 Documentation of optimization tool (.txt, .pdf or .docx)

A text file describing how to use your optimization module. You should explain how a user can run your code from the command line, as well as what functions to import into Jupyter notebook. You should give example inputs to illustrate both usages. This needs to be sufficiently clear so someone with a very basic knowledge of Python can use your tool.

In addition, you should explain in detail the format and significance of every part of your sample input and output data. This needs to be in sufficient data so a trained analyst can create the correct input data that would work with your tool, and interpret your results correctly.

### 3.7 Presentation slides (.pptx or .pdf)

The slide deck that you will use during the final presentation. You will not be able to change your slides on the day of presentation.

### 3.8 Contributions of team members (.docx or .pdf file)

A one page description of the individual contribution of each team member to the project, as well as the percentage that each contributed. This may be used to adjust individual grades at the end of the course.

## 4. Grading Rubrics

### 4.1 Project report (20% of course grade)

The report is out of a maximum of 36 points, with 3 points being distributed in each of the following 12 categories. What is needed to obtain a perfect score in each category is described below. Within each category, there is a 0.5 point deduction for every minor issue and a 1 point deduction for every major issue.

**1. Executive summary:** The executive summary clearly and concisely summarizes the main findings of the report, including the inefficiency found, the potential gain from optimization, and the recommendation to the business. The language is polished and format is easy to read for a busy executive.

**2. Opportunity for improvement:** Clear and precise description of what is the inefficiency in the status quo you found. The description of the opportunity for improvement should be compelling and significant.

**3. Rigour in justifying opportunity for improvement:** The identified opportunity for improvement is convincingly illustrated using either data analysis or simulation. Moreover, the analysis is rigorous and free from errors and statistical fallacies. The way you conducted the data analysis or set up the simulation is clearly explained so another trained analyst can reproduce the findings if needed.

**4. Clarity of methodology:** The required inputs, outputs, as well as the decision variables, objective, and constraints are clearly described using English (not mathematical notation). Furthermore, the methodology is appropriate for the problem and capture the most important issues.

**5. Clarity of results:** The optimization results and insights are clearly presented. If tables, figures or plots are used, they are easy to understand, well labeled, and clearly described. Furthermore, the potential gain from using your optimization tool is clearly articulated, as well as qualitative intuition on where the gains may come from and why they make sense.

**6. Rigour of results:** The interpretation of the numerical results is correct and any claim of potential gain or insight found is convincingly justified by the presented supporting evidence. The result is not oversold, and the logic of the analysis and its explanation is correct.

**7. Discussion of methodology:** The reason behind the choice of input and output data, as well as the decision variables, objectives, and constraints are clearly described and the logic is sound. Potential alternatives are discussed and the main assumptions are clearly and thoroughly articulated. The weaknesses of the analysis are clearly acknowledged.

**8. Discussion of recommendation:** The recommendation to FritoLay is clearly stated and explained. The issues that are not captured by your model are accounted for, and your recommendation and explanation makes business sense (showing street smarts).

**9. Correctness of mathematical formulation:** The input data, decision variables, objectives, and constraints are clearly expressed by appropriate and correct mathematical notation. Furthermore, the mathematical formulation correctly optimizes the English description in the main body.

**10. Discussion of technical details:** The potentially problematic technical details in the analysis are identified and precisely described. The reasoning behind these technical choices are clearly presented and the remaining weak points and gaps are acknowledged. There are no large issues in the technical details that are not mentioned.

**11. Appropriate language:** The main body of the report uses appropriate language for a non-technical but business savvy executive, with minimal mathematical notation or technical jargon. The technical appendix uses language appropriate to a technical audience, using precise mathematical and technical language.

**12. Polish:** Overall look is professional. The formatting should be clean and consistent. Titles of sections, plots, and tables should be descriptive and appropriately sized. All figures should have descriptive axis labels that are clearly readable. Lack of unnecessary, distracting, or unprofessional looking visuals. The font size is not too small and the margins and spacing are sufficient. Mathematical notation and equations are rendered nicely. Writing is free of spelling and grammatical errors. Respects the page limit.

#### 4.2 Python code and documentation (5% of course grade)

The optimization tool (3.3), documentation (3.2 and 3.6), and data files (3.4 and 3.5) are graded together using the following rubric. There is a maximum of 18 points, with 3 points for each of the following categories. What is needed to obtain a perfect score in each category is described below. Within each category, there is a 0.5 point deduction for every minor issue and a 1 point deduction for every major issue.

**1. Correctness:** Code implements what is claimed without syntax or logical errors. The code can be run on another computer without error. Everything written in the documentation (see 3.6) works. Code correctly runs on the sample input (3.4) and produces the sample output (3.5).

**2. Scalability:** Code is able to be scaled up to the size of inputs that is needed for the actual business context, within a reasonable runtime, while obtaining a good solution. The optimization formulation is reasonably efficient (cannot do the same thing with many fewer decision variables or constraints.)

**3. Usability:** The inputs needed to run the code can be plausibly obtained by the end user within a reasonable amount of effort. The functionality of the code matches the need of the end user. While the expectations of the code is that it is only a prototype, it should be something that FritoLay would be interested to develop into a tool that they use in making decisions.

**4. Readability:** The code includes appropriate docstrings for all functions, and enough comments so a trained analyst can understand it easily. The variable names should be appropriate and there should not be extraneous logic that has little functionality but may confuse the reader.

**5. Completeness of documentation:** Both the documentation of plots and figures (see 3.2) and the documentation of the optimization tool (see 3.6) contain all of the required components: explanation/code with enough detail to reproduce every numerical estimate, table or plot in the main body; instructions with examples of how to use the optimization tool both from Jupyter notebook and from the command line; explanation of the format and significance of every piece of the sample input and output data.

**6. Completeness of data files:** All of the input files needed to run the code are attached to the submission (or is in the Dropbox folder of the project dataset). Moreover, the sample output that would be created by the code is also given. All of these should match the high-level description in the project report and the detailed description in the documentation (see 3.6).



### 4.3 Final presentation (5% of course grade)

The presentation is out of 30 points, with 3 points for each of the below categories. What is needed to obtain a perfect score in each category is described below. Within each category, there is a 0.5 point deduction for every minor issue and a 1 point deduction for every major issue.

**1. Addressing the main question:** Presentation clearly addresses the question of what are the biggest opportunities for improvement and how much are the potential gains from applying your optimization; presents sufficient evidence or arguments to support the answer.

**2. Business insights:** Presentation provides clear insights that may be valuable to staff and data analysts in FritoLays. Moreover, the insights are effectively communicated, and the reasonings are compelling.

**3. Structure:** Present information in a logical, interesting, orderly sequence. The main message of the talk and the outline is clear from the start. Use of signposting language to guide the audience through the presentation.

**4. Visual aids:** Slides have few words and can be easily read by audience; visuals improve the presentation and coordinate with the points of speech; effective use of highlighting techniques; no extraneous graphics, animations and distracting visuals; avoid complex looking math formula.

**5. Time control:** Complete the presentation within the time allotted, spending enough time on the important points rather than rushing through many things; gracefully end the presentation when time is up rather than require intervention from moderator. (When time is short, you should skip less important points if appropriate rather than trying to rush through and lose the audience.)

**6. Elocution:** Projecting voice so audience can hear at all times; speaking clearly and understandably; correct, precise pronunciation of terms; not speaking too fast; avoiding repeated use of filler words like “um,” “er,” “like,” etc; engaging voice with appropriate variation in tone, emphasis and pauses.

**7. Body language:** Relaxed and confident poise; no distracting mannerisms or fidgeting; appropriate pointing, gestures and movement to enhance articulation.

**8. Connect with audience:** Attempted to engage audience members from all sections of the room through eye contact and enthusiasm; demonstrates a strong positive feeling about the topic during the entire presentation.

**9. Accessibility to non-technical audience:** Explain content in such a way that a senior executive without much technical training can follow along the basic logic. When appropriate, rewording technical terms in a non-technical way to aid understanding.

**10. Correct technical content:** Whenever technical terms or mathematical notation is used, they are used correctly; the logic of arguments presented is rigorous and free from fallacies.

## 5. Ideas for Getting Started

The size and complexities of the data (see [detailed description of data here](#)) may seem overwhelming. As with most real world analytics project, much of the available data may not be directly relevant, and the data you really need may be missing. Hence, it is important to first form in your mind a conceptual picture of what are the most important components to the problem, and what you would like to focus on, rather than try to take in all of the available data without such a mental picture.

Below are ideas of potential directions (you are not limited to these options but these are given to inspire your own thinking):

- **Define what a “good” portfolio of SKU looks like:** The first step of optimization is to measure what is “good.” One way to start is to go through the list of issues in section

2 and come up with a reasonable metric to quantify each. There are many issues so you should first identify ones you believe are the most important. For example, you may decide that the ideal selection of SKUs for a region should have products that are predicted to sell well, have low supply chain costs, as well as be sufficiently diverse to cover the population's needs. Consider how you would measure each of these, and try to approximate each of these measures using linear objectives or constraints in your optimization formulation. It is okay that your metric may require data beyond what is given in the project, as long as you can argue that this data can be obtained without too much effort by the company.

- **Find examples of SKUs which you might argue for addition or removal:** the status quo at FritoLay chooses products based on sales volume (velocity) and percentage of sales that are returned (stale rates). One approach is to dig into the available data and find evidence for SKUs and regions in which the current policy is not good. For example, you may want to identify certain products with very high transportation cost and low production margins, and try to make a case that it might not make sense to include this product at the region. Another example is to find SKU that you think should be added to a certain region, and make a case for the addition (due to good performance in other regions that you argue have similar tastes, and reasonable transportation costs).
- **Simulating data you do not have:** Beside analyzing existing data (which can be very challenging due to the high level of aggregation), an alternative approach for identifying opportunities for improvement is to use simulation to illustrate stylized examples in which the current policy would perform very badly. For example, products that have very unpredictable demand spikes may warrant stores to have larger inventory to be able to meet the spikes, but this may also result in products expiring (high stale rates). There might be demand distributions in which it is more profitable to incur a high stale rates rather than keeping inventories low and incurring high lost sales, and this may challenge the status quo of using stale rates to filter products. It would be helpful to identify using the data the type of products that may exhibit such demand distributions.
- **Supplementing with other data sources:** one way to make your submission stand out is to find your own datasets to supplement your analysis. As an example, the data from FritoLay does not segment consumers from each region or study the demographics of regions. You can find alternative sources of demographic data for various states and try to use that to guide the SKU decisions.
- **Creating numerical product attributes:** Much of the information about products in the data set are hidden in the descriptions, and some are dispersed in multiple files (for example, the transportation data has product shape and flavor, which you may want to merge with the other files). It might be helpful for your analysis to create numerical attributes, such as whether a product is spicy, cheesy, healthy, etc. An example of how you might use these attributes is to help quantify what a "diverse" portfolio mean. Another use would be to estimate what attributes are most important for each region.