## Joe Marco - Springboard Capstone Project 2017



### What is the problem I want to solve?

* Overall business problems around risk avoidance, brand loyalty, and new customer engagement/experience. These focus on the following questions.
* Predict back orders based off historical data 8 weeks prior?
* Identify “high value” products (SKUs) based off repetitive high inventory need over time
* How many “potential issues” actually equated into actual backorders? What is the distinction for when a potential issue, might be low, medium, high risk based off correlated entities.
* Correlation between lead time and ultimate backorder? How many days lead time am I at high risk of backorder for a specific SKU or SKUs
* Based off of Performance avgs can I predict which SKUs are using the same source? Can I shift from one source to another to save cost and business issues?
* More…

### Who is your client and why do they care about this problem? In other words, what will your client DO or DECIDE based on your analysis that they wouldn’t have otherwise?

My Client(s) could be consisted of the following business groups who each have specified decisions based off the data:

1. Supply Chain: produce more products in high demand with high boackorder frequency, or reduce production of low need, low risk of backorder. Saving costs and time
2. Marketing: How to plan marketing campaigns in accordance with current state backorder problem? I do not want to market for a product that is going to go on backorder
3. Sourcing/Procurement: How do I identify which source(s) of product are putting my business at risk by consistently having a low average performance, and who else can do what they do but better?
4. Inventory management - decisions based on high or low inventory models

### What data are you going to use for this? How will you acquire this data?

* <https://www.kaggle.com/tiredgeek/predict-bo-trial>
* The data (fortunately) is easily accessible and downloadable. It is 23.44MB in size with over 1MM rows.

### In brief, outline your approach to solving this problem (knowing that this might change later)

* Inspect the data and definitions of the data
* build a model based on available stock, lead time and the source performance
  + could use help where to get started
* build model based on high inventory, high lead-time products
  + could use help where to get started
* Other models depending on complexity of the above

### What are your deliverables? Typically, this would include code, along with a paper and/or a slide deck.

* R Markdown slideshow with end to end work
* Includes intro to me, my background, my proposal, code throughout, did I answer any of the above questions, if so how? What was the business value? & A lessons learned
* Potential website with R Markdown if I can figure out the html based configs in R Studio & Github.

### Real World Value and Interpretation

These methodologies can be applied in the real world work environment, and at my current employment in numerous ways, a few are listed below: \* Can I predict when raw materials are going to be at risk of backorder based off sales for product and demand of consumer? \* Can i identify which of my suppliers may be consistently a "high risk" when it comes to providing materials or products late? \* Is there a way to identify simply based off legacy product backorder information, which new products (based off inventory, lead time needed, source/suppliers) can be predicted to be at risk for a backorder issue? \* Additional real world applications