**Assessment – Data Scientist**

\*There is no strict constraint on the analytical tool. But Excel, R, and Python are preferred.  
  
\*Hint: Please note that Assessment data2 has 3.6 million rows so opening it in excel might result in data loss.  
  
**QUESTION 1**

Attached, please find an Excel worksheet titled “Q1.Rev\_Projection” with a mock-up revenue history. Based on the provided data points, please project the monthly revenue for 20X4 using 3 different methods. For each method, include your projection in numbers, formula and/or code, and provide an explanation of your approach. Additionally, select the projection you believe is the most accurate and explain your reasoning.  
  
**QUESTION 2**

Our small parcel shipping costs have been increasing as a percentage of revenue every year. Attached, please find an Excel workbook with UPS shipment data for the same week 2017 vs 2018 (tab Q2.2017 and tab Q2.2018). We shipped out of two warehouses for these two weeks, New Jersey and Ohio. Supplyhouse had a new contract with UPS with better pricing for the 2018 week. However, as a % of revenue, UPS spend was higher year over year.  
  
• Using the provided data sets (and any other resources you require), identify some reasons why we are experiencing these results. Please include your analysis, showing the Excel formulas and/or code that led you to these conclusions.  
• Given this limited information, how would you identify areas of opportunity for decreasing our overall shipping costs? Please explain your approach.  
  
**QUESTION 3**

Attached, please find “Assessment Data2.csv” with a mock-up order history.  
  
Please note:  
• Line item: Every product on an order is referenced as a line item. You can also define it as a combination of order ID and product ID. E.g., order1\_productA and order2\_productA are two unique line items.  
• Product ID could duplicate under the same order due to a combination of box sales and individual sales of different quantities.  
 o As long as both the product ID and the order ID is the same, the duplicated products should be counted as 1 line item. E.g., order1\_productA\_qty3 and order1\_productA\_qty1 is 1 line item.  
 o The total quantity sold should factor in both the box sales and the individual sales. E.g., order1\_productA\_qty3 and order1\_productA\_qty1 sum up to 4 units sold.  
  
To better prepare for the work volume in the fulfillment center, please:

• Predict daily line items and daily quantity sold in the year 2004. Share your formula and/or code, and explain your methods.

• Summarize:

o The trends you found in the mock-up sales history.  
 o The outlier days in the mock-up dataset, and briefly explain how you handled them.