BUAN 5210 Quick Project 1
Professor Green Due: Thursday, Jan 31, 5:00pm

You have recently begun an internship for an online office supply store, Office Mate. They have been in business since 2012 and have good success. However, they are beginning to face more competition from Amazon and several other online retailers, so they need to determine how to become more profitable. Your manager, Chris Lee, has come to you and a co-worker to complete an analysis and come up with two recommendations to improve profitability. The specifics of the task are outlined below. Chris will be sharing your findings with the board of directors and other managers in the form of a memorandum, so it is critical that your writing and graphs are clear to people without a technical background.

The goal of this assignment is to explore and wrangle a data set with the tidyverse packages, find and discuss two ways to increase profits, and support those findings with memo and plots from ggplot2. You may work in teams of two. Use the data "5210_W19_QP1_Data.csv", which includes following variables:

1. Order ID: Each order ID is unique, but may contain several products
2. Ship Mode: Shipping method
3. Customer ID: Each customer has a unique ID, but can make more than one order
4. Region: Sales region in the United States (Central, East, South, West)
5. Segment: Customer segment (Consumer, Corporate, Home Office)
6. Category: Type of product sold (Furniture, Office Supply, Technology)
7. Sub-category: Detailed type of product sold
8. Product Name: Name of product sold
9. Product ID: Product ID code
10. Quantity = number of units of the specific product sold for that order
11. Revenue = Price * Quantity (though the dataset does not include Price)
12. Discount: Percent discount from full Price
13. Profit = Revenue – Cost (not including shipping or tax, both passed directly to customer)

Each observation is for sale of a specific product

Submit your findings in 2 documents described below, a memo and technical appendix.

Your Task:

1. Complete a basic and detailed EDA in an R Notebook. The basic EDA is what we have discussed in class, univariate/multivariate/non-graphical/graphical EDA. A detailed EDA is when you dig deeper into the data to answer questions you raised in the basic EDA. Use a ***variety*** of tidyverse data verbs to select, filter, group, create new, and summarize variables—be adventurous! Use exploratory visualizations in base and/or ggplot2. Render this R Notebook in HTML with a table of contents to show all code and output—this is your Technical Appendix. Make sure your code is clear and documented with code comments and comments/questions regarding your findings. Remember code comments are in the code chunk and data comments are after the output (but leave a blank line between output and data comment). You may want to label sections by EDA step and/or the question you are exploring to organize the R Notebook. Your file should contain a Table of Contents as discussed in Day 1 of class. This is one file to be turned in; the Technical Appendix in HTML titled "last_name-last_name.html". Save the image of the file so you can load it in your memo, as demonstrated in the first day of class.

2. Determine your two preferred findings to improve profitability and produce a well-titled and labeled graph in ggplot2 to support each finding. Though you may create many graphs in your EDA, you will only want to use a two of your graphs to illustrate your findings. It is important the graph is easy to understand so make sure it is not complicated or cluttered, with clear titles and axis labels. I expect you to have at least two nicely formatted graphs; quality is more important than quantity. Write up a four-paragraph summary of the project: an introduction (why is the analysis important), a paragraph and nicely formatted graph for each finding, and a conclusion/summary. Render these paragraphs and the graphs as an html memo with text and graphs (no code) from RMarkdown (use my example from Day 1 as a template). Write the memo for a lay audience—no code, no technical graphs (boxplots and correlation graphs are technical), and no jargon. The memo with graphs should be shorter than approximately 2 pages (html does not have page breaks so you will have to use your judgement).

Quick Project Grade Rubric

Category - Percent of grade	Score	**Comments**
Introductory Paragraph – 3 points Why important, brief findings, clear, no jargon, Memo format		
Finding 1 – 3 points Description, Visual, no jargon		
Finding 2 – 3 points Description, Visual, no jargon		
Conclusion – 1 points Clear, no jargon		
Technical Appendix – 12 points Clear, labeled, complete, documented, variety of commands, appropriate use of techniques, TOC		
Total	**/20**	