PART I:

1. A statistics professor is interested in helping students improve their understanding of key concepts. She thinks that requiring homework may help students better master necessary concepts. Before making this decision she decides to offer optional homework assignments. She compares the test scores of students who complete the homework with those who do not. The homework group obtains significantly higher scores.
   1. Does this suggest that homework increases comprehension? Why or why not?

**Answer:** Since the experimental group obtained “significantly higher scores”, we would be tempted to think that “homework increases comprehension”. However, there are many variables that might be omitted, which if included in the analysis would have given us a better understanding of the reason behind a specific student’s academic performance. I would have to control for such variables and compare students’ performance before making that suggestion. Additionally, since the assignment was optional, students who better understood the material would take advantage of the opportunity; while those who did not comprehend the concept would avoid taking the optional assignment. Factors such as socioeconomic status and parents’ education level are some of the factors that might aid in making a potent analysis in this survey and yet, we do not have this data for the students. The data provided is not enough to conclude whether or not “homework increases comprehension”.

1. The ASPCA is trying to increase donations via its latest fundraising drive. It has recently launched a social media campaign on Twitter posting a picture and accompanying story of dog who overcame a difficult situation and found a loving forever home. Users can click on the post to view the story, as well as the amount of money and number of donors to the campaign. **The goal of the ASPCA campaign is to both attract new donors, as well as increase the donations of previous donors.**

Use the data provided to run an appropriate statistical analysis looking at the effectiveness of the Twitter campaign.

1. What statistical analysis did you run and why? Was the campaign effective? If so, can we determine a causal relationship between our variables of interest?

**Answer:** I generated a descriptive analysis table since we are mainly concerned with the donations data prior to and post campaign. I determined that the campaign was effective since there is an increase in donors and in donations from 2nd time donors as well. The total contribution of 2nd time donors was higher than the total pre-campaign donations. The average donation was also larger. I conclude that the campaign was effective based on these results. However, I must note that there was not a significant relationship between comments and retweets for the post-campaign donations. The r-squared was .1; a 1.0 is considered a good indicator of a relationship. I did not detect a causal relationship between post-campaign donations and the other variables.

1. Are there alternative explanations or exogenous factors that the company should consider? If so, name them, and describe how they would impact the interpretation of the results.

**Answer:** The company can consider targeted ads to specific audiences it determines tend to respond the most to its posts. On average, gender1 contributed more in both donations than any other gender. Additionally, science and technology workers also donated more in average than any other professions. Other exogenous factors the company can consider are also unemployment rates, which would affect whether or not people might donate and if they do, how much. It can also consider the time of day and week it makes the posting. For example, creating the post during the middle of the day might receive less responders since people tend to be at work during that timeframe. Another factor could be the pandemic; people tend to spend more time on their social media and might be in a spending mood since they get bored at home with not much to spend on.

1. You have to brief your boss, who is not familiar with research design and statistics. Provide a short summary describing the most important information in terms understood by a lay audience. Please also include descriptive information about the users.

**Answer:** Donations significantly increased after the Twitter campaign. The average donation increased by $54.2 and most people donate around $35 dollars compared to the pre-campaign where most people did not make any contributions. The minimum donation was $10, and the maximum was $240. Before the campaign, the maximum was $100, and the minimum was 0. We had 40% less donors prior to the campaign; that is a significant increase. With a total donation increase of $5,420 can conclude that the campaign was effective, with a caveat on causal relationship. There seems to be no relationship between the prior donation and the post-campaign one

1. Based on your results, what recommendations would you provide your boss and other stakeholders regarding future campaigns?

**Answer:** We should conduct one more campaign so we can gather some more data. To make accurate predictions for future campaigns, we will need some more observations. However, for now I think we can suggest that the success stories encourage people to donate more since they have a positive feedback from their expenditure. Also, allowing people to see current statistics of the donations give them confidence in the company’s mission and dedication to dogs’ wellbeing; if that is the case, we might see more donations next time.

PART II:

The attached spreadsheet provides quarterly sales data (Q1 2012-Q3 2014). Use this data to complete the tasks below. To allow us to better follow your work process, please begin a new tab for each task, labeling the tab with the task number (e.g., ‘Task 1’). When all of the tasks are complete, please use your work to create a PowerPoint presentation detailing your findings. Your presentation should create a polished narrative linking together each of the tasks to tell an overarching story.

Task 1: Using a pivot table, restructure the data to show details on each **product line split out by retailer type.** For each product line include the following:

1. Number of transactions (consider each row as a single transaction)
2. Volume sold
3. Total revenue

Once the table is created highlight the most popular item for each retailer type and select an appropriate way to represent the data.

Task 2: Restructure the data to display the total number of transactions by order method type over time. Select an appropriate way to graphically display the data and write a short 3-5 sentence conclusion explaining the trend.

Task 3: Using a bubble chart, graphically display revenue, quantity, and gross margin aggregated by product line. Format the chart to be as clean as possible and include a label for each bubble. Write a short explanation detailing what the graph reveals about the data. Also include 1-2 sentences about your structural choices- how you determined what data should go one each axis- and why those were the best choices.