Capstone Week 1: Getting to Know the Data

**Welcome to the Capstone Project!**

We’ve covered a lot of information about descriptive statistics this week. You’ve learned how to use spreadsheets to calculate some key measures of central tendency and dispersion, and to analyze frequency and correlation data. Now we want you to have an opportunity to apply what you’ve learned to some real-world data, similar to what you’re likely to encounter in life as a marketing analyst! That’s the purpose of our Capstone Project.

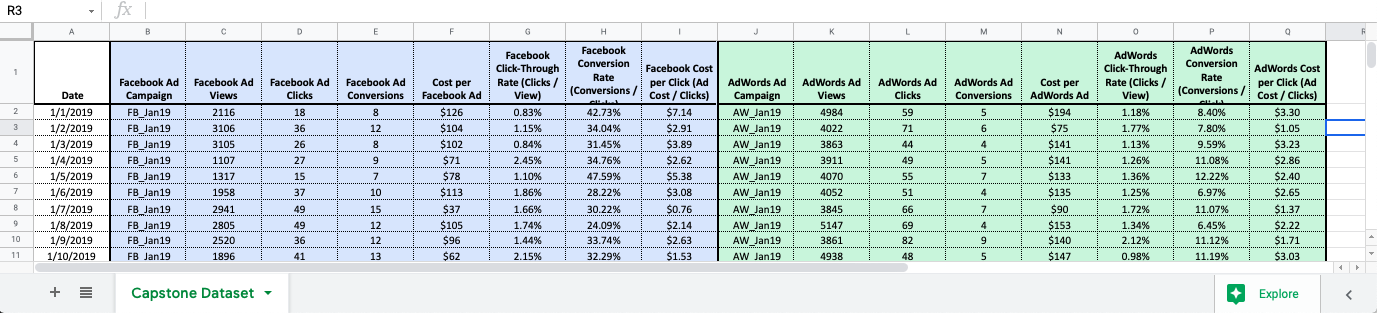
At the end of each week’s activities, we will have one of these Capstone readings for which you will apply that week’s learnings to a Capstone Dataset that we’ve prepared for you. After you run your analyses for the week’s Capstone reading, we’d like you to cut-and-paste your results into a slide-deck that we are also supplying to you. Once you have completed four weeks of Capstone activities, you will have compiled a full slide deck of information that shows you running a completed data analysis of the kind you might be asked to run for a marketing analytics job. At the end of this four week Capstone Project you will submit your completed slide deck for peer review, so that you can compare your work and share feedback with one another.

To get started, you will need to download a copy of the Capstone Dataset spreadsheet.

**Directions for how to download the Capstone Dataset.**

[Click on this link](https://docs.google.com/spreadsheets/d/111Qu0KWR5ILzmmB2TVZTwesUI-zkNW7m/edit?usp=sharing&ouid=112542094875848186692&rtpof=true&sd=true) to go to the Capstone Dataset spreadsheet. You can create a copy of the file in Google Drive and fill in your answers there or download it to your computer to work on it locally. This dataset is what you’ll use for the entire capstone project and it’s a good idea to check it out now. Don’t know how to download this file? You can find more information about downloading this by [clicking here](https://support.google.com/docs/answer/49114?hl=en&co=GENIE.Platform%3DDesktop#zippy=%2Cdownload-a-copy-of-a-file).

You will be working with this spreadsheet for all four weeks of Capstone readings, so **be sure to save your own copy of the Capstone Dataset spreadsheet on your computer or in your Google Drive** to go back to each week! Here’s what the first several rows of data should look like in the spreadsheet:

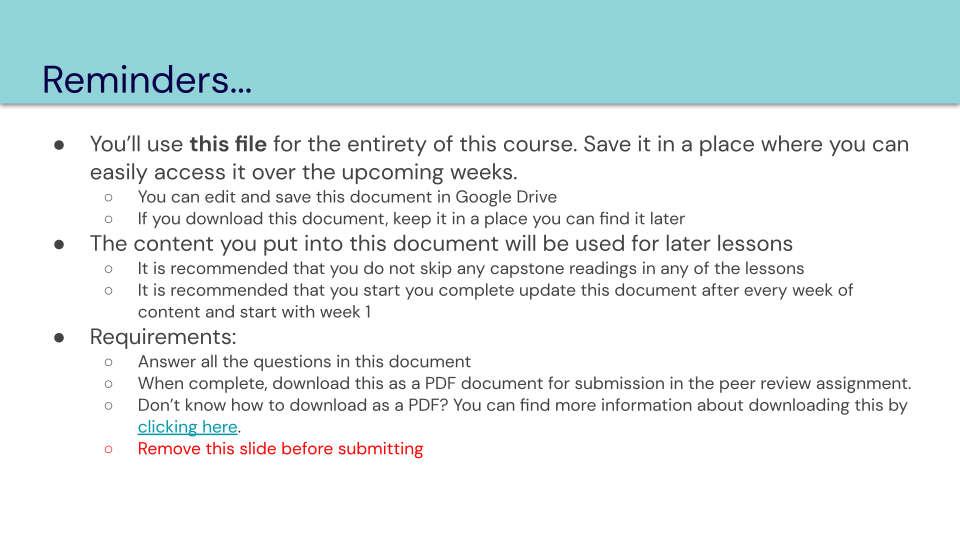


Each week you will also be cut-and-pasting your results into your ongoing Capstone Slide Deck. So, you will also need to download a copy of this slide deck.

**Directions for how to download the Capstone Slide Deck.**

[Click on this link](https://docs.google.com/presentation/d/1RPdw8A_66qE3aaGPHi2RozXSFsO9aROEm0jPqNuR51I/edit?usp=sharing) to go to the template slide deck file. You can create a copy of the file in Google Drive and fill in your answers there. This slide deck will be what you submit at the end of your capstone project, and it’s a good idea to check it out now. Don’t know how to download this file? You can find more information about downloading this by [clicking here](https://support.google.com/docs/answer/49114?hl=en&co=GENIE.Platform%3DDesktop#zippy=%2Cdownload-a-copy-of-a-file).

Again, you will be working with your slide deck each week and making changes to it. **Be sure to save your own copy of  the Capstone Slide Deck on your computer or in your Google Drive** to work with each week! There should be 17 slides in the deck you are downloading, and here’s what the first slide in the deck should look like:



Once you’ve downloaded your own copies of the dataset and slide deck, let’s get to work!

**Week 1: Getting to Know the Data**

This week’s videos and readings were all about descriptive statistics. You learned about measures of central tendency, measures of dispersion, frequency tables, scatter plots, and correlation coefficients. These all can be important tools for understanding your marketing data. For this Capstone Week 1 assignment, we want to use these tools to better understand the data in your Capstone Dataset spreadsheet.

When you open the Capstone Dataset spreadsheet you can see that it is a collection of data comparing the performance of two separate ad campaigns: a Facebook Ad campaign (in the blue colored columns of the sheet) and an AdWords Ad campaign (in the green colored columns). The data is collected for every day of the year 2019, so you have 365 lines of campaign data to analyze! Ultimately, we want to determine whether one of these ad campaigns is more effective, in terms of conversions, compared to the other. But before we address that question, we should get a better understanding of the conversion data for each of the campaigns.

To better understand how effective our campaigns are at producing conversions, we’re going to begin by analyzing the relationship between a campaign’s number of clicks and the corresponding number of conversions. Let’s do this by calculating (1.) the measures of central tendency, (2.) some measures of dispersion, and (3.) the frequency and correlation data for a campaign’s clicks and conversions.

We’ve learned how to do all of these things with our spreadsheet software. Below, we show you the results we get for the Facebook Ad data. We started by creating a new sheet in our Capstone Dataset spreadsheet and copy-and-pasting the “Date,” “Facebook Ad Clicks,” and “Facebook Ad Conversions” data into the first three columns of our new sheet. This makes it easier to work with that specific data for the calculations that follow!

*Note: for your Capstone Week 1 assignment, you are doing these calculations for “AdWords Ad Clicks” and “AdWords Ad Conversions.”*

**1. Calculate Measures of Central Tendency**

For this first set of calculations, let’s determine the *mean*, the *median*, and the *mode* for both the number of “Clicks” and the number of “Conversions.” Recall from our reading on “Measures of Central Tendency” how to calculate these measures in Excel or Google Sheets:

* You can calculate the *mean* using the **=AVERAGE** formula. In our example below, we displayed the result for “Clicks” in cell E2, and for “Conversions” in cell E6.
* You can calculate the *median* using the **=MEDIAN** formula. In our example below, we displayed the result for “Clicks” in cell E3, and for “Conversions” in cell E7.
* You can calculate the *mode* using the **=MODE** formula. In our example below, we displayed the result for “Clicks” in cell E4, and for “Conversions” in cell E8.

In our example, we show you the results for the Facebook Ad Clicks and the Facebook Ad Conversions. **For your Capstone Week 1 assignment calculate the *mean*, *median*, and *mode* for *AdWords Ad Clicks* and *AdWords Ad Conversions*.** You will enter your AdWords Ad results into your Capstone Slide Deck slides for week 1.

**2. Calculate the Standard Deviations**

Now that we have determined the “middle” values for our ad campaigns let’s get a sense of how much variance from the middle value we should expect from our data. We’ll do this by calculating the *standard deviation* for our “Clicks” data and our “Conversions” data. This tells us how “spread out” our data is. Recall from our reading on “Measures of Dispersion” how to make these calculations in Excel or Google Sheets:

* You can calculate a *standard deviation* using the **=STDEV** formula. In our example below, we displayed the result for “Clicks” in cell E10, and for “Conversions” in cell E11.

In our example, we show you the results for the Facebook Ad Clicks and the Facebook Ad Conversions. **For your Capstone Week 1 assignment calculate the *standard deviations* for the *AdWords Ad Clicks* and *AdWords Ad Conversions* data.** You will enter your AdWords Ad results into your Capstone Slide Deck slides for week 1.

**3. Create a Frequency Table for Conversions; Analyze the Correlation between Clicks and Conversions**

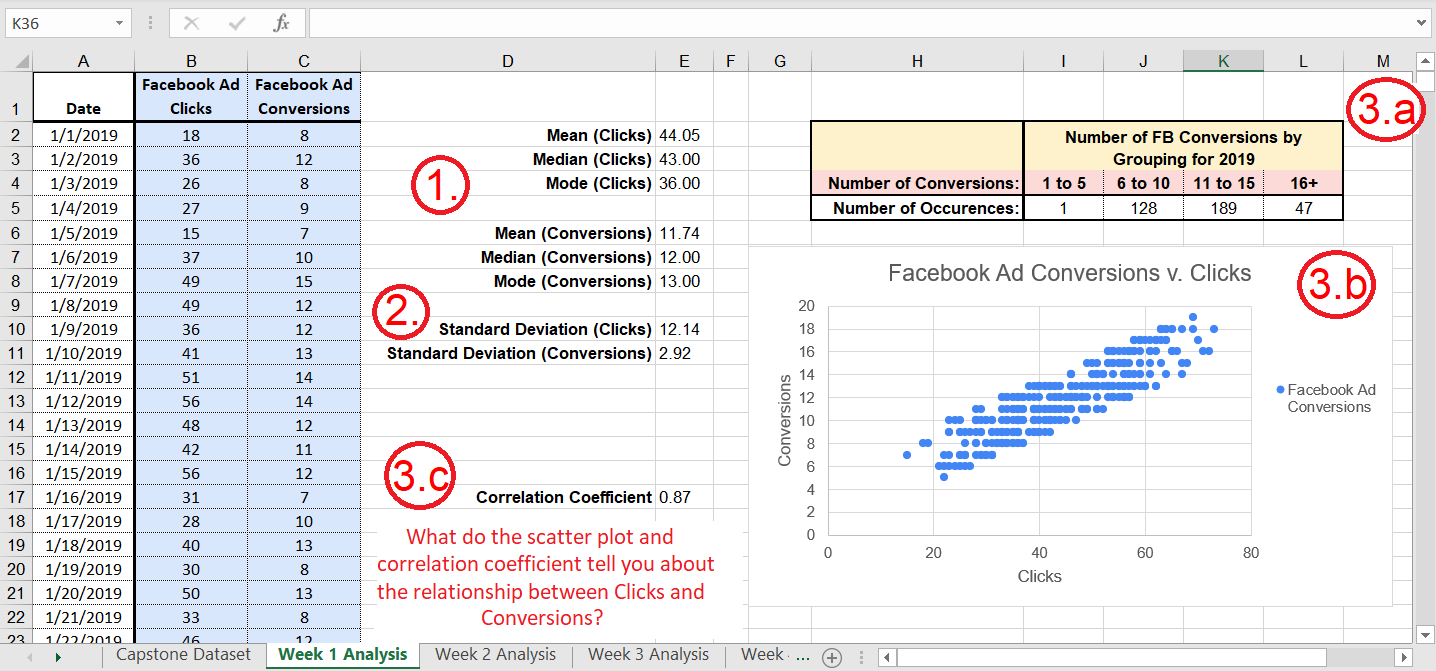
Finally, let’s get a clearer picture on how much correlation there really is between our ad clicks and the number of conversions we get through the ad. Do more clicks on the ad really lead to more sales? We’ll address this question by, first, creating a frequency table that tells us exactly how many days of the year we recorded specific numbers of conversions: less than 6 conversions, 6 - 10 conversions, 11 - 15 conversions, and greater than 15 conversions (see example below). Then, we’ll analyze how much these conversions correlate to the number of clicks by creating a scatter plot for Conversions v. Clicks, and by calculating the correlation coefficient for these two variables. Recall from our reading on “Frequency, Contingency, and Scatter Plots” how to perform these tasks in Excel or Google Sheets:

* You can create a *frequency table* by using the **=COUNTIF** and **=COUNTIFS** formulas to count the number of occurrences for each grouping of conversions (i.e., “<6” occurrences, “>5” but “<11” occurrences, “>10” but “<16” occurrences, and “>15” occurrences). In our example below, you can find our frequency table in the area highlighted with the red “3a.”
* You can create a *scatter plot* by using the “insert chart” functions in your spreadsheet software. (Refer to your software documentation for more information.)  In our example below, you can find our scatter plot in the area highlighted with the red “3b.”
* You can calculate a *correlation coefficient* using the **=CORREL** formula. In our example below, we displayed the result for “Clicks” in cell E10, and for “Conversions” in cell E17.

In our example, we show you these items for the Facebook Ad Clicks and the Facebook Ad Conversions. **For your Capstone Week 1 assignment create a *frequency table* for the *AdWords Ad Clicks* data (with the groupings described above); then create a *scatter plot* and calculate the *correlation coefficient* for the *AdWords Ad Clicks* and *AdWords Ad Conversions* data.** You will enter your AdWords Ad results into your Capstone Slide Deck slides for week 1.

**Your Turn!**

Here’s what we got when we completed all the above tasks for the Facebook Ad Clicks and Facebook Ad Conversions data from the Capstone Dataset spreadsheet. (Question: what does our correlation coefficient and scatter plot tell you about the relationship between clicks and conversions for the Facebook Ad?)



Now it’s your turn! **Complete these tasks for the *AdWords Ad Clicks* and *AdWords Ad Conversions* data from the Capstone Dataset spreadsheet. Enter your results into your Capstone Slide Deck.** *What does your correlation coefficient and scatter plot tell you about the relationship between clicks and conversions for the AdWords Ad?*

**Completed**