Capstone Week 2: Understanding Your Data Samples

In this week’s videos and readings you learned some important questions to ask about your data samples. The answers to these questions can be essential to determining what kind of analysis you will ultimately be able to perform with your data! For example, you learned how to create histograms for your data samples, which help you determine what kind of shape best describes the distribution of your data. It’s important to know this because many analyses require that your data has a specific kind of distribution shape (for example, a *normal distribution* curve). Similarly, in later weeks we’ll be learning several types of analyses that require specific types of variables. So, it’s important that you are able to determine what types of variables you are working with. For this Capstone Week 2 assignment, we want to use the skills you’ve learned to better understand the distributions and variables in your Capstone Dataset spreadsheet.

**1. Create Histograms for ‘Clicks’ and ‘Conversions’ Data**

A *distribution* shows the probability of possible outcomes occurring within your dataset.

Distributions are represented by a distribution curve. This week we learned how to visualize the distribution of values by creating **histograms**. Histograms are a type of bar chart, where each bar represents a value range and the number of data values that fall into that range. We call these ranges “buckets.” The height of each bar in your histogram will be determined by the number of data values that fall into each “bucket.” So, the bars in your histogram just represent the frequency for each value or value range in your chart. Review this week’s reading, “Distribution,” for a refresher on how to create a histogram in Excel and Google Sheets.

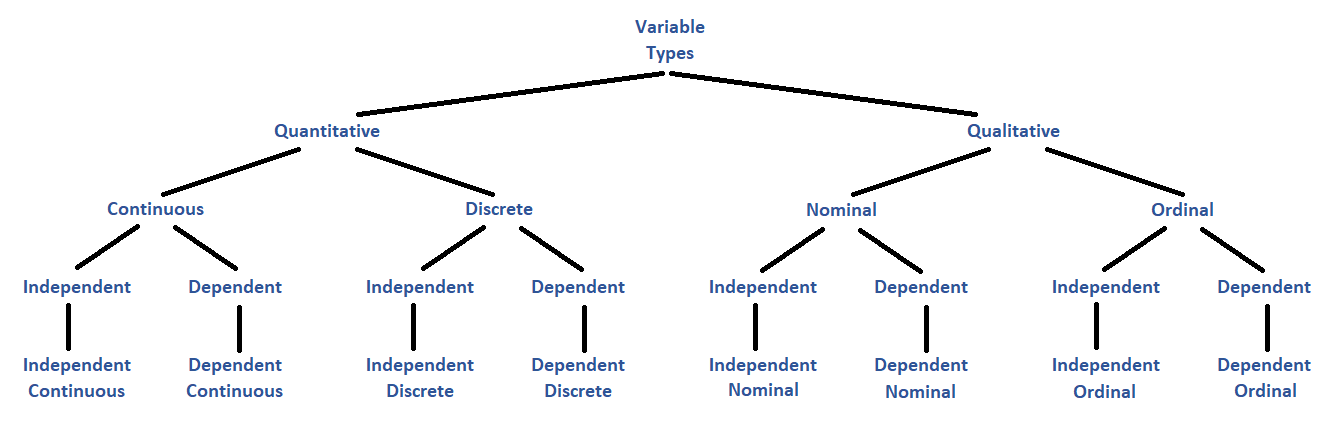
In our Capstone Project we have begun to focus on our data for “Clicks” and “Conversions” from each of the ad campaigns (Facebook and AdWords) in our Capstone Dataset. For this first part of our Capstone Week 2 assignment, we want to determine whether our data for “Clicks” and “Conversions” falls under a *normal distribution curve*. To do this, use your spreadsheet software to create two histograms—one for the “Clicks” data and one for the “Conversions” data.

* For your “Clicks” data create a histogram with a *bucket size of 10.*
* For your “Conversions” data create a histogram with a *bucket size of 2.*
* Evaluate your histograms: *Do they show a Normal distribution?*

In our example below, we show you histograms for the Facebook Ad Clicks and the Facebook Ad Conversions. **For your Capstone Week 2 assignment create histograms for *AdWords Ad Clicks* and *AdWords Ad Conversions,* then determine whether they fall under a normal distribution curve.** You will cut-and-paste copies of your AdWords Ad histograms into your Capstone Slide Deck slides for week 2.

**2. Determine What Variable Types You’re Working With**

In this week’s videos and readings we learned about eight different types of variables you might deal with when running your data analyses. We summed them up with this chart:



Now, the question of “Independent” versus “Dependent” cannot be answered until you have set up a hypothesis test for your data, which we will do next week. But you can determine whether your variables are “Quantitative,” “Qualitative,” “Continuous,” “Discrete,” “Nominal,” and “Ordinal” just by looking at the data values and considering how they were collected (or calculated, as the case may be). Review this week’s reading, “Variable Types Review,” for a refresher on how to determine these variable types.

In the Capstone Dataset you are given data for a total of sixteen variables: Facebook Ad Views, Facebook Ad Clicks, Cost per Facebook Ad, Facebook Click-Through Rate, Facebook Conversion Rate, Facebook Cost per Click; and, AdWords Ad Views, AdWords Ad Clicks, Cost per AdWords Ad, AdWords Click-Through Rate, AdWords Conversion Rate, AdWords Cost per Click. For this second part of your Capstone Week 2 assignment, we want to determine the variable types *for all the AdWords variables in the Capstone Dataset*.

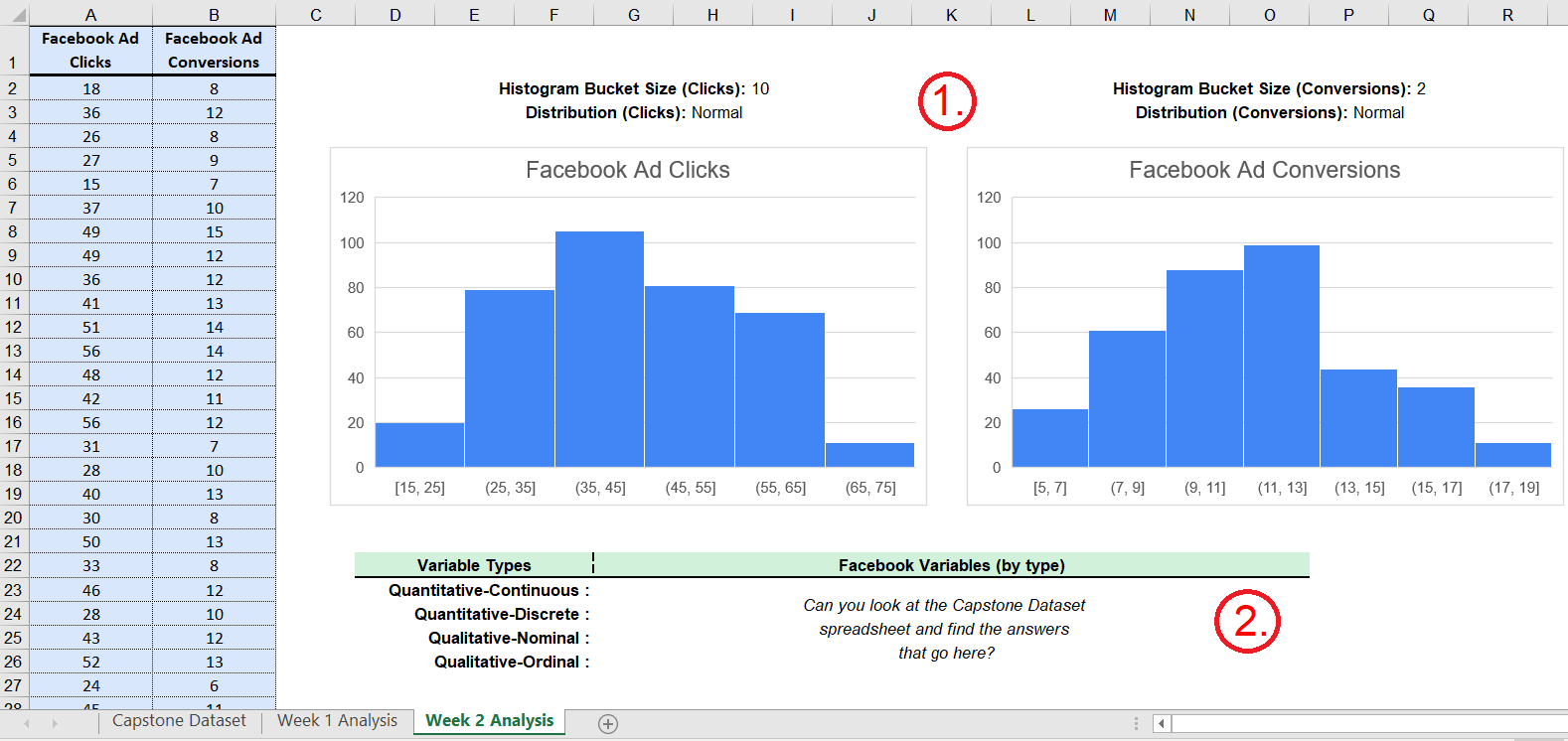
**Place each of these variables: (1.) *AdWords Ad Views*, (2.) *AdWords Ad Clicks*, (3.) *AdWords Ad Conversions,* (4.) *Cost per AdWords Ad*, (5.) *AdWords Click-Through Rate*, (6.) *AdWords Conversion Rate*, (7.) *AdWords Cost per Click*, into one of the following variable types.**

* **Quantitative-Continuous**
* **Quantitative-Discrete**
* **Qualitative-Nominal**
* **Qualitative-Ordinal**

You will enter your classifications into your Capstone Slide Deck slides for week 2.

**Your Turn!**

Here’s what we got when we completed the above tasks for the Facebook Ad data from the Capstone Dataset spreadsheet.



Now it’s your turn! Can you *complete these tasks for the AdWords Ad data* from the Capstone Dataset spreadsheet?