Capstone Week 4: Show Me the Model

You’ve come a long way in your Capstone Project. We gave you data to analyze and a specific question to answer: “Is there a difference between the number of conversions on the Facebook platform versus the AdWords platform?” In Week 3, you answered this question. You showed how the data reveals that I’m likely to see more conversions from my Facebook Ad versus my AdWords Ad. That’s great information to have—I think I’ll go with the Facebook Ad!

But there’s more to the story that I’d like to know. What will happen when I do go with the Facebook Ad? You’ve shown that I’ll do better than if I go with AdWords, but what do those “better” results really look like? Here’s where modeling comes into play. With the right modeling technique, you can show me what my results are likely to be, given the data we’ve seen in the past.

In this week’s readings and videos, we’ve shown you several models that you can run to answer questions like the one posed above, about what kind of conversion rates I can expect from my Facebook Ads. But which model you run for your analysis depends on the question you are trying to answer and the type of data you have to work with. So, for this week’s Capstone assignment, we want you to go through the process of *choosing a model that’s appropriate to the evaluation question and the Capstone Dataset* that you’ve been working with over the last three weeks. Specifically, now that we’ve determined (in Week 3) that Facebook Ad Clicks are most effective at generating Conversions, let’s get a clearer handle on this effectiveness by answering this question:



Your Capstone Week 4 assignment is to run a model that answers this question.

**1. Choosing a Model**

In  this week’s discussions we looked closely at three common types of models and how to run them using the Tableau software. Those models were: **Simple Linear Regression, Cluster Analysis, and Time Series Analysis**. We also discussed **Classification Analyses**, though more briefly and without running one in Tableau. For all these models, there are specific requirements that must be met for you to even be able to run them and obtain accurate results. This is why we have spent so much time considering how to choose the right model for the specific analysis you might be asked to run. We’ve collected all these considerations into a **“Choosing A Model”** checklist, included below, that you can use to help you decide what model to run for this week’s assignment.

The first consideration when choosing a model involves understanding that each type of model was designed to meet *a specific purpose*. So, the very first thing you want to do when choosing a model is to determine whether the purpose of a model matches the objective behind your analysis! For example, if you are asked for an estimate of what your company’s profits are projected to be over each of the next three months, a Time Series Analysis could be a good modeling technique for answering this question. That’s because Time Series Analysis models are designed for forecasting, i.e., predicting the value of some value (like profits) at future times (like each of the next three months).

After you’ve narrowed your choice of model according to purpose, you next have to consider *variable requirements*. For example, the variables you’re concerned with in the question given above are “Facebook Ad Conversions'' and “Facebook Ad Clicks.” You’ve learned how to characterize these variables by type (such as quantitative v. qualitative, and independent v. dependent). This information is important because each model requires variables of specific types. So, if you consult our checklist below, you can see that you need a quantitative independent variable and a quantitative dependent variable in order to run a Cluster Analysis.

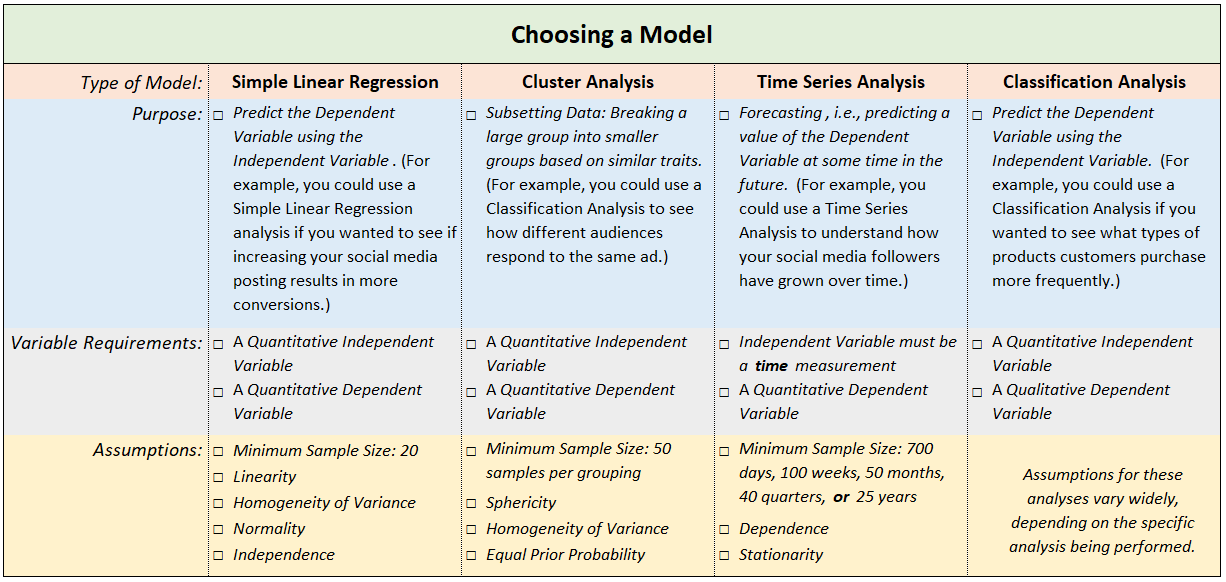
Finally, once you’ve determined a model you’d like to run, based on purpose and variable requirements, you must look at your data to make sure it meets a set of certain *assumptions*. Your data must meet these assumptions in order to give you accurate results! For example, you might determine that Simple Linear Regression fits your analysis in terms of purpose and variable types, but does your distribution for these variables meet the “normality” assumption, under that model in the “Choose A Model” checklist? (Notice that much of the work that you’ve done in previous weeks of the Capstone project will help you determine whether your data meets such assumptions.) Refer again to this week’s readings and videos for a refresher on what all of these assumptions mean.

For your Week 4 Capstone question:



Refer to the “Choose A Model” checklist below and determine:

* Which model’s purpose best addresses the question at hand.
* Which model’s variable requirements are met by the question at hand.
* Which model’s assumptions are met by the Capstone Dataset data.



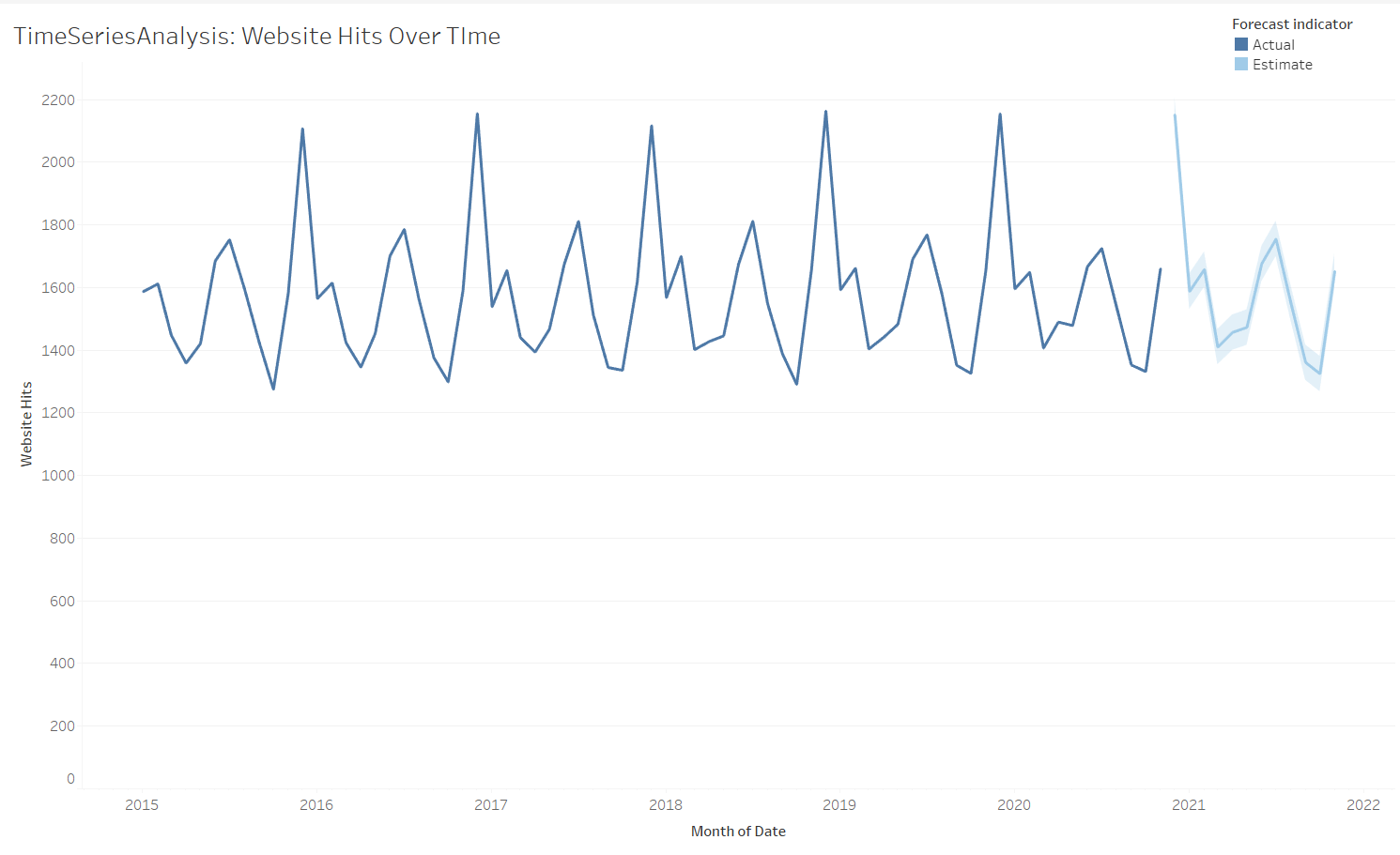
**In the Capstone Slide Deck slides for week 4, include the model you’ve selected and your reasoning behind that choice.** If you’d like, you may [**download a PDF of this checklist here**](https://drive.google.com/uc?export=download&id=1eSgb4ZhF7pYnvLDw7--4gmkjePF2v7MF), fill it in, then paste an image of it into your slide deck.

**2. Run Your Model**

Now that you have determined the model that will best address your question, it’s time to run it!

In this week’s readings and videos, you learned how to use Tableau to run three of the model types described above: Simple Linear Regression, Cluster Analysis, Time Series Analysis. (Hint: one of these is the model that best addresses the question at hand.) We recommend that you use Tableau to run your model. However, if you have other tools to run your model, feel free to use those instead. But for this exercise, you are required to produce a visualization chart, similar to the one below, which shows the results of your analysis. *You will need to take a screenshot of your visualization and paste it into your Capstone Slide Deck slides for week 4*. As we’ve seen from this week’s discussions, Tableau is an excellent tool for producing such visualizations.

Here’s an example of the visualization we created in Tableau for a Time Series Analysis:



Your visualization chart should have a structure similar to this one. However, *the above chart is just for example purposes*. Unless you are using a Time Series Analysis model, your chart details are likely to vary significantly from this one. (Each model produces a distinct chart.)

Finally, remember that the purpose of running models is to be able to analyze specific results (such as *predicting* a value for a specific future date, or *segmenting* data to see the habits of a specific group, etc.). Once you have run your model for this week’s assignment, you should be able to read a specific result for Facebook Ad Conversions v. Facebook Ad Clicks directly from your visualization chart! Refer again to this week’s readings and videos for a refresher on how to do this.

For this second part of the Week 4 Capstone assignment, run the model you chose above for the Capstone Dataset data.

* Produce the appropriate visualization chart for your results, and cut-and-paste a copy of it into your Capstone Slide Deck slides for week 4.
* Using your chart, state what your model shows to be the *expected number of Facebook Ad Conversions* for a day of *50 Facebook Clicks*.

**Enter your results for these tasks into your Capstone Slide Deck slides for week 4**.

**Well done!**

After completing the tasks from this Week 4 Capstone assignment, you’ve now completed a rather thorough analysis of the Clicks to Conversions data from your Capstone Dataset spreadsheet. You’ve given me a detailed understanding of the Clicks and Conversions data for the Facebook and AdWords platforms (in your Week 1 and Week 2 assignments), you were able to show me which platform is most effective for turning clicks into conversions (in your Week 3 assignment), and now you’ve shown me what kind of conversion rate I can expect from my winning platform (Week 4). Nice work!