

Reminders...

- You'll use this file for the entirety of this course. Save it in a place where you can easily access it over the upcoming weeks.
 - You can edit and save this document in Google Drive
 - If you download this document, keep it in a place you can find it later
- The content you put into this document will be used for later lessons
 - It is recommended that you do not skip any activity in any of the lessons
 - It is recommended that you update this document after every week of content and start with week 2
- Requirements:
 - Answer all the questions in this document
 - When complete, download this as a PDF document for submission in the peer review assignment.
 - Don't know how to download it as a PDF? You can find more information about downloading this by [clicking here](#).
 - **Remove this section before submitting**

Content

- **Week 2 Activity: Obtaining and Scrubbing Data**
- **Week 3 Activity: Exploring and Modeling Data**
- **Week 4 Activity: iNterpreting Data**

Activity:

Obtaining and Scrubbing Data

Anna owns a clothing boutique in New York, called BrightThreads. She sells a mix of clothing brands and chooses items for her store that she believes her clients will like. She also sells online.

Anna is working on long-term planning for the upcoming year at BrightThreads. Business has been going well, but she would really like to increase sales and potentially open up a second location in a different neighborhood. Next year, Anna would like to increase her total sales by 10%. This would be a very good year for Anna and BrightThreads, but it seems doable based on the last few quarters and with some hard work.

Using this information, answer the questions below regarding the obtain and scrub stages of the OSEMN process. Add your answers to the template below.

In this scenario, what is a SMART goal that would benefit from data analysis?

What is a Primary KPI that would be useful to analyze for this goal?

What relevant data would you gather in this scenario?

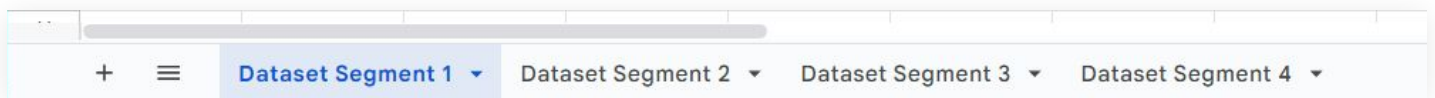
How do you imagine you could obtain this data? What sources would you gather data from? Specifically, note what kind of data (first-party, third-party) and what methods you might use (survey, web analytics).

Anna at BrightThreads has begun the process of gathering data to help analyze current sales.

She has collected data on recent online sales directly from the online storefront.

Access [this sample Customer Data](#) and click on Use Template in the upper right corner. You will need to be logged into a Google account to use this template.

Anna has isolated 4 different segments that each have issues that need to be fixed. You can access each segment in the four sheets in this one spreadsheet. Click on each sheet for a different segment of the dataset. You can click on the tabs at the bottom of the spreadsheet to move between sheets. Review the image below for a preview:



The four sheets are accessible by clicking the tabs at the bottom of the spreadsheet.

Using what you know about data validity, do you think the data Anna has gathered is valid? Why or why not?

What issue did you identify in segment 1 of the data?

What issue did you identify in segment 2 of the data?

What issue did you identify in segment 3 of the data?

What issue did you identify in segment 4 of the data?

Activity:

Exploring and Modeling Data

Anna from BrightThreads is exploring some data from last quarter's online sales. The data was gathered from the BrightThreads online store.

Access [BrightThread's online sales data](#) and click on Use Template in the upper right corner to access the dataset. Please note you will need to be logged into a Google account.

Review the following data and charts, then share what you can learn in the exploration stage of the OSEMN process.

Using this information, answer the questions below regarding the explore and model stages of the OSEMN process. Add your answers to the template below.

What are some things you can tell about this dataset? For instance, what does the size of the dataset tell you?

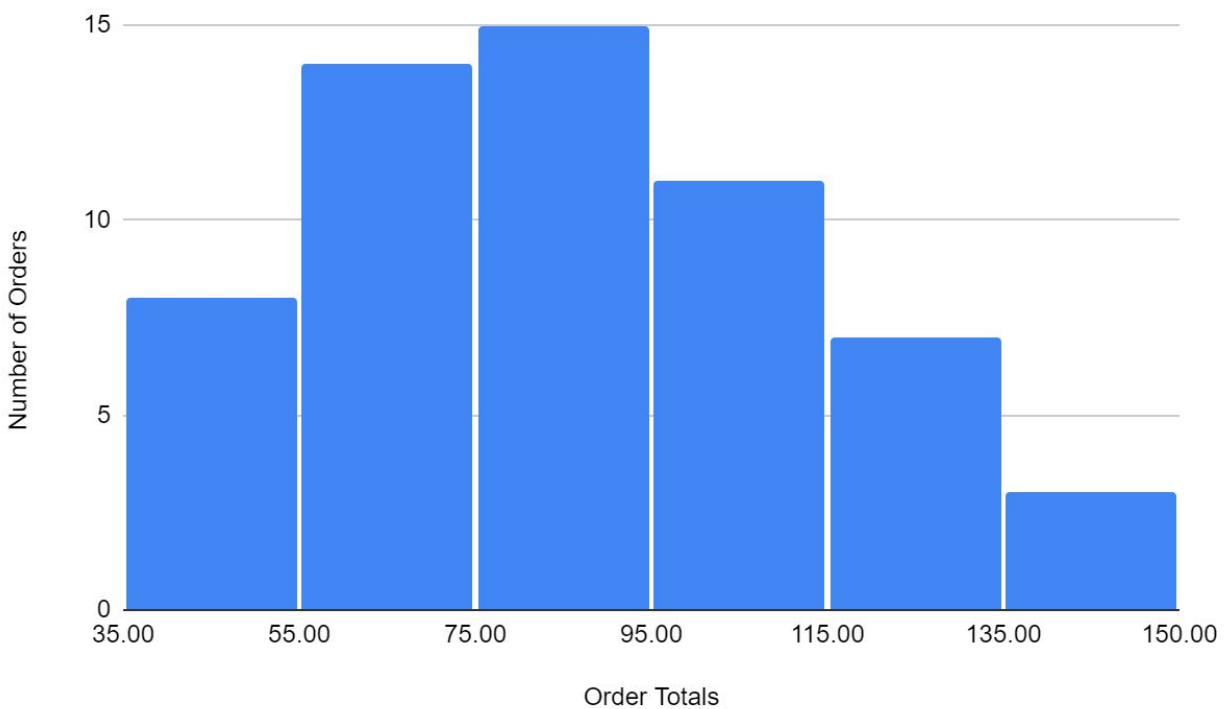
What kind of data is in this dataset? (Numerical, categorical, etc.)

**Reviewing this data, what is the minimum value in the order_total column?
What is the maximum value in order_total column?**

What kind of chart would you use to help visualize this data?

Based on what you have learned, would you add an additional column to this dataset using feature engineering? For instance, using the sales dates, would it be helpful to add in the day of the week data?

Anna has created the following chart to explore the relationship between order totals and the number of orders.



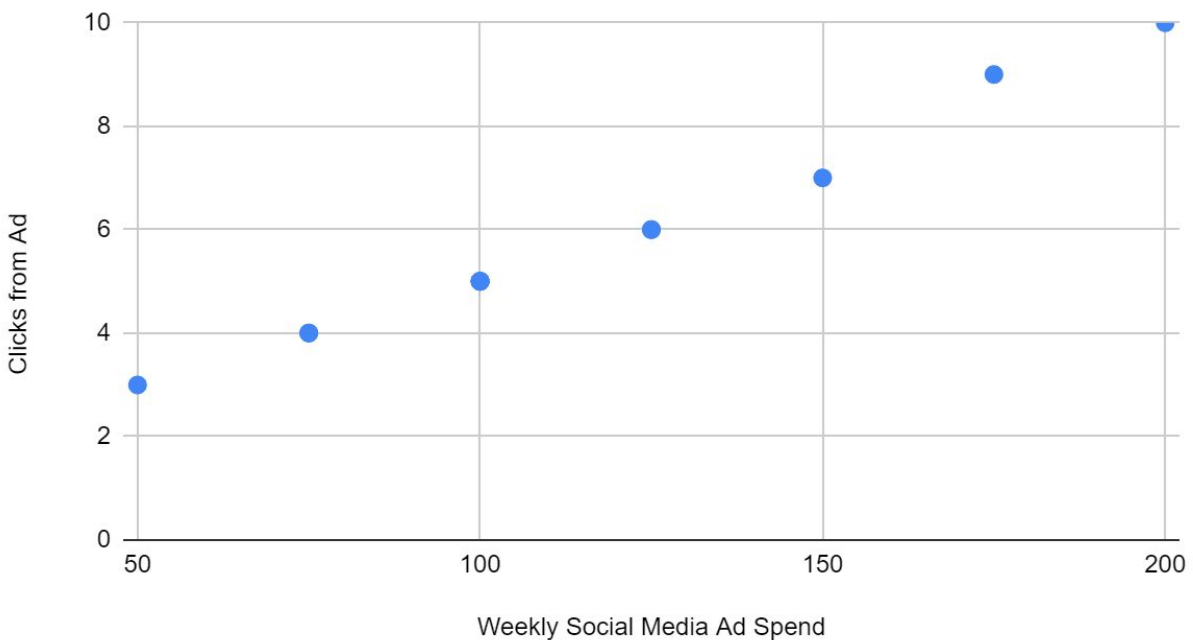
Based on the data in this chart, what would be a good title for this chart?

What does this chart tell you about the number of orders in relation to the amount someone spends per order?

What range do most of the orders tend to be in?

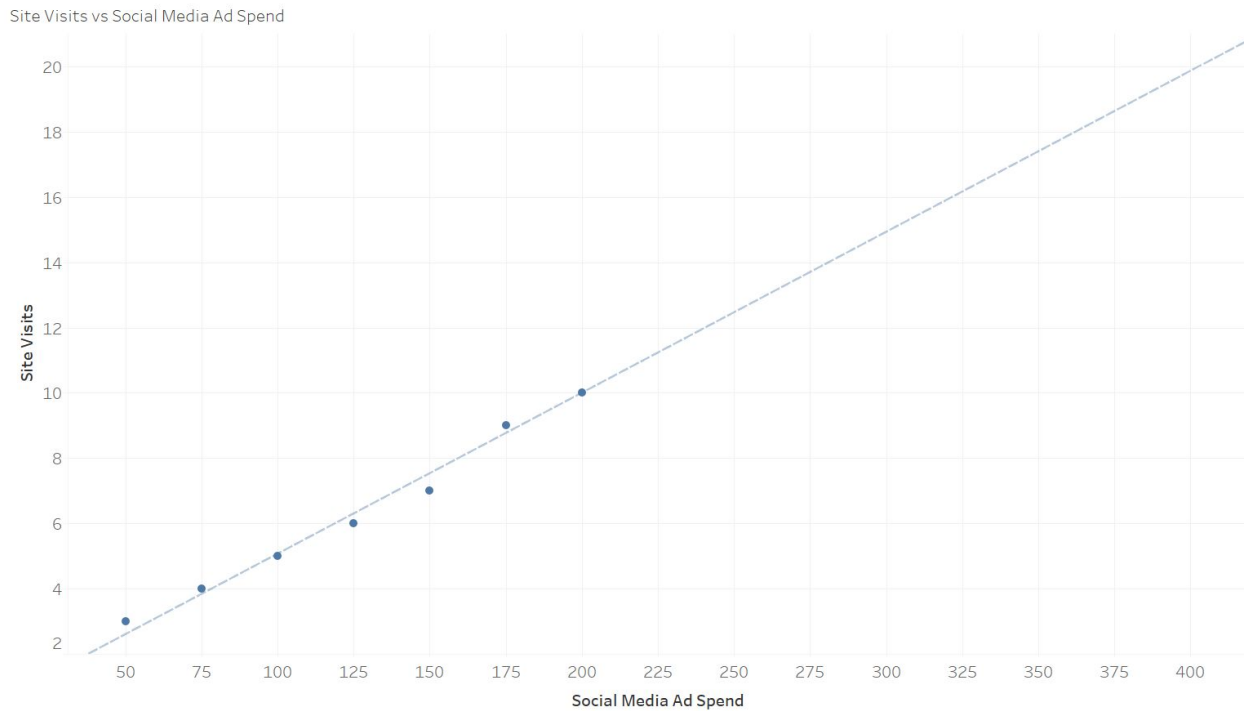
Anna has also been analyzing data on the amount of money she spends on social media ads and how many clicks to the BrightThreads website they are generating.

Site Visits vs. Social Media Ad Spend

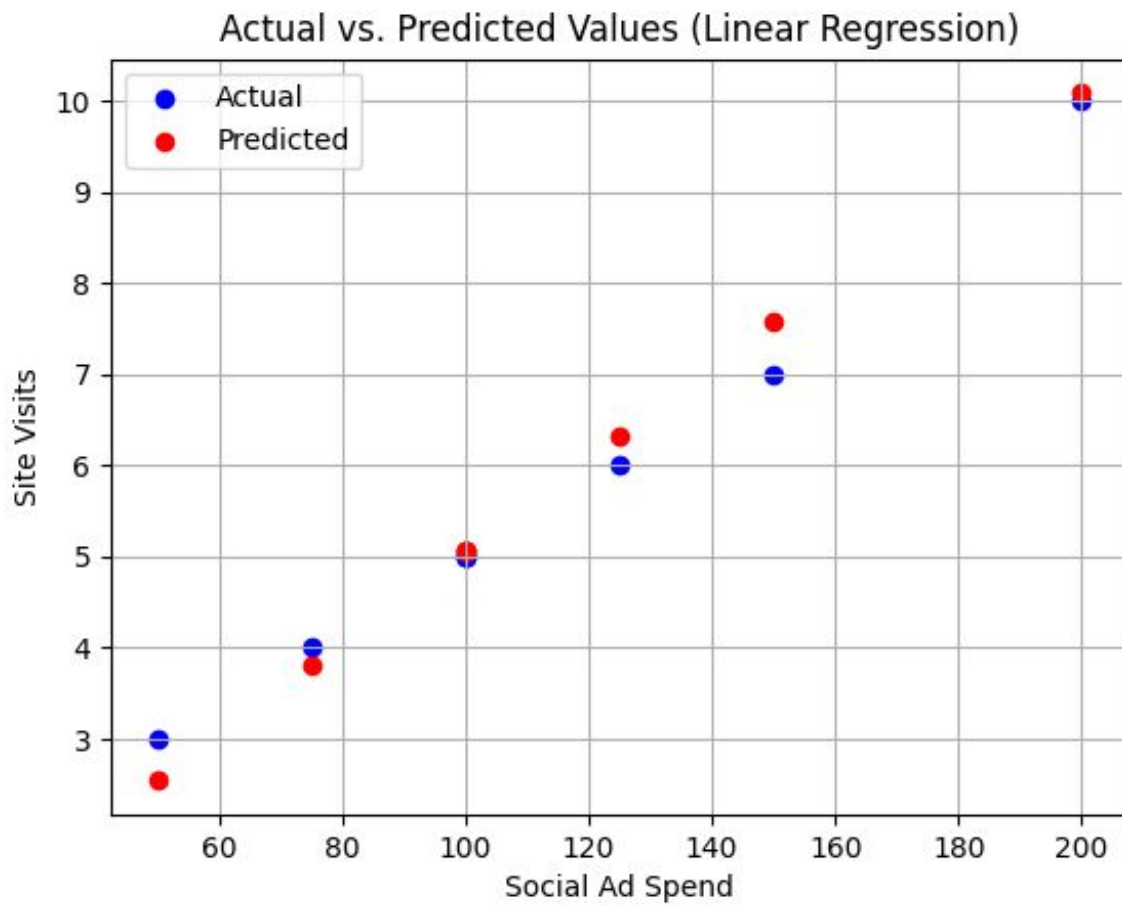


Do you notice any correlations between the variables in this chart? If so, how would you describe them?

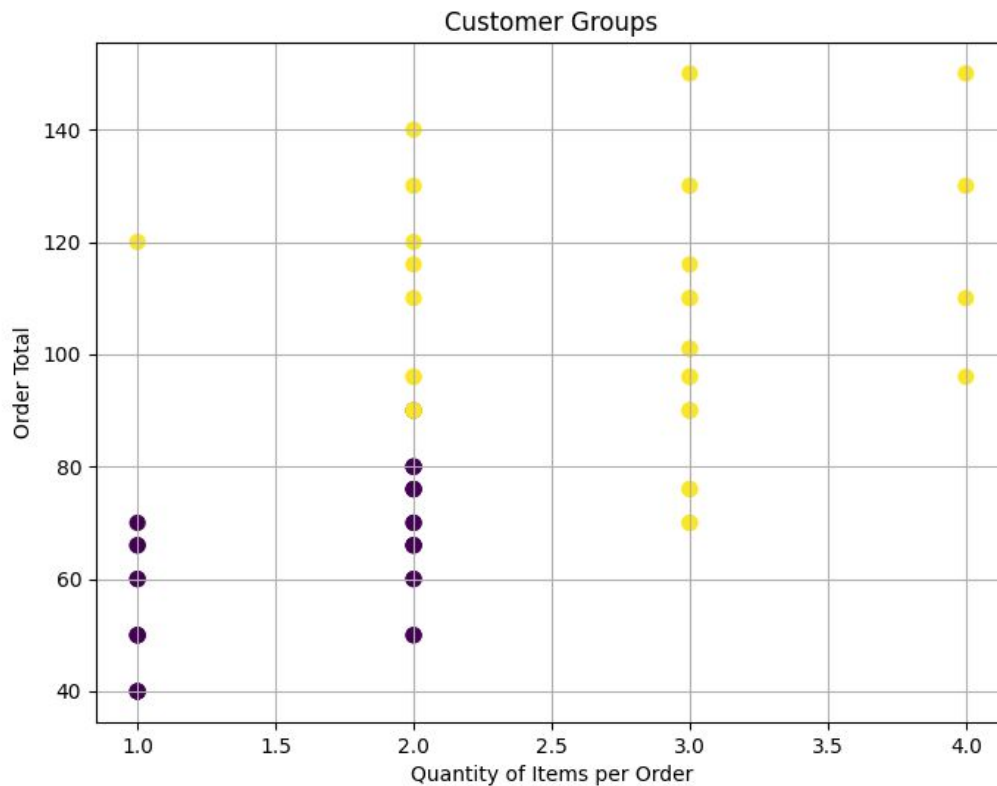
Anna has learned a lot while exploring the data she has gathered. Now, it's time to model some of this data.



Reviewing this linear regression model, roughly how many site visits can be expected if the marketing budget is increased to \$250?



Review this linear regression model which shows the actual data values and the values predicted by the model when given a test set. Do you think that this model is sufficient for general use for this data? Why or why not?



Review this clustering model. A clustering algorithm has been used and identified two groups. How would you describe the two different customer groups? Why?

You are trying to forecast BrightThreads sales in the coming quarter – what model might you use? Why did you choose this?

Activity:

iNterpreting Data

Anna has learned many things using data analysis. She has prepared a presentation to show to BrightThreads stakeholders. As a reminder, her goal is to grow sales by 10% in the upcoming year, and this presentation will cover what she's learned and how she plans to accomplish this goal.

Access [Anna's presentation](#).

Review the presentation, then share your thoughts on Anna's interpretation of the data at the end of OSEMN process.

Using this information, answer the questions below regarding the interpret stage of the OSEMN process. Add your answers to the template below.

What was the objective for this analysis?

How can Anna apply this in a business context?

What slides in the presentation covered the methods used in the project?

What slides in the presentation included visualization of the project?

What slides in the presentation offered recommendations after the project?

In your opinion, what parts of the presentation were the setup, buildup, climax, and conclusion? Why?