

Sales Analysis

A look at company sales across the years 2015-2018.

The manager wants to take a look at the sales during the 2015-2018 timeframe. She wants to know the top 10 projects by value and by total number of orders. They want to see that broken down by category, subcategory, and segment. They'd also like to see the top sellers to a specific customer.

On top of this, the manager said she'd like to see how each category and segment did in each region. She would like to be able to split that out specifically by category, sub-category, region, and segment. She's interested in finding out whether they need to pay more attention to any particular region, or see if a region is doing poorly despite past efforts.

In discussions with the manager, I asked if she'd also like to see information on a state-by-state basis, and she thought that was a good idea. She would like the top 5 states by total order value, and she'd like to see the number of orders for each of those states as well. She'd like to be able to break that down by year, category, sub-category, region, and segment.

I thought it may also be useful for her to add the bottom 5 states, so I decided to add that with the same ability to view that by year, category, sub-category, region, and segment.

Before I could do any of this, I needed the data. For the data for this portfolio project, I found a dataset on kaggle.com that worked well for my analysis. I downloaded that .csv file and loaded it into the online version of Excel.

One of the first things I noticed when I opened the data file was that the order dates and ship dates were not in an expected format. Upon further review, I realized that part of those columns were in the expected month/day/year format, some were in the day/month/year format. On top of that, the ones in the day/month/year format were saved as text, not a date.

Because of this, I created a calculated field to accept a date from all of the Order Date field. Using that calculated field and splitting the Order Date (using Left, Mid, and Right), I was able to correct the dates that were problematic. I sorted the Order Date field, which moved all the incorrectly formatted fields to the top of the table. I was able to then copy the corrected date from the calculated field and paste the value directly into the original Order Date field. I then deleted the calculated column, since I no longer needed it.

I also created a calculated field to get just the year from the Order Date to simplify the slicing process later on.