Data Analysis Bootcamp Project One-Analyzing Sales Data

For this project our group selected a set of sales data from an office supplier to analyze, which included information about order dates, locations, prices, etc. We set out 4 main questions to examine and analyzed the data to answer.

The questions were: Where are the most orders being placed? Are different categories of products more popular? Are there specific products that are more popular at different times of the year? Is there a difference between sales dollars and order numbers?

In looking at where most orders are being placed, the number of orders was first grouped by the state where the order was delivered, then further grouped by city. When plotted, both of these showed that the largest number of orders tended to occur in areas with high populations.

Next, the data was analyzed to find out which categories of items were most ordered. A plot of the categories of items ordered vs. a sum of the orders including those items showed that most of the categories had lots of orders, with two notable exceptions being copiers and machines. This could be due to the fact that these items tend to be both high cost, and long lasting as compared to other more consumable supplies.

The next analysis was done to examine what sales looked like at different times of the year. To do this, a plot was made comparing the sum of sales (in dollars) vs. the month of the sale. This showed that the two most popular months to order were March and October. Examining the timing of orders further, this data was broken down into the categories of items and their contribution to this sales pattern. This analysis showed that the item most consistently contributing to sales across the year is Tables. There are other categories that may have a larger one month share of the sales, but the Tables category is consistent across all months.

Finally, we examined differences between sales dollars and order numbers. The first way this was examined was by comparing the sales across the four US regions (North, East, South, West) when using the number of sales vs. sales in dollars. This showed that there was a noticeable difference between this two methods of measuring 'sales'. This questions was examined another way, by comparing the number of items per order vs. the average cost of the order. This showed that as the number of items in the order went up, so did the average cost of the order. This seems to indicate that there is not a clear pattern of extremely expensive one item orders, nor large bulk orders of only the cheapest items.

This project was done with a limited amount of time and only the data already contained in the dataset that we used. We identified a number of further points of analysis that we would pursue with more time. One unexplored area is the type of customer. The dataset categorizes the customers between consumer, corporate, and home office. Our analysis did not examine these but more trends could be found by using these categories overlaid onto our original analysis to look at things like whether certain customer types had more predictable order patterns, for example a corporate customer may be more aligned to traditional fiscal year and financial

quarter trends than a home office customer. There is also more that could be answered within the analysis of where orders were being placed. If we were able to examine the orders per capita of the city or state, would this show any outliers or trends within the sales? Finally, would we be able to find repeat customers within the data set and examine individual trends within their orders such as which time of year they order and what type of item they order, etc.