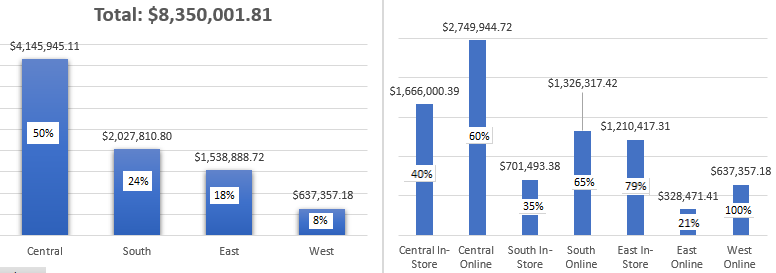
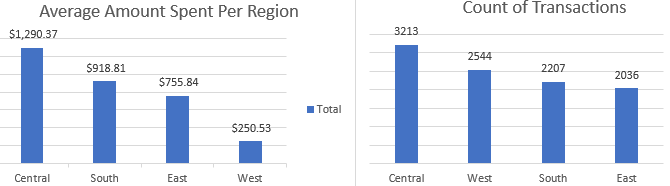
**Customer Purchasing Data Report**

A critical component of any successful business is determining how to keep and attract new customers. The last decade there has been an incredible amount of data available that if collected, can give invaluable insight into current and potential customer preferences. The goal of this task was to perform data analysis for Blackwell Electronics and utilize data mining methods on the sales data that was provided to determine customer preferences and leverage these insights into attracting new customers.

Danielle Sherman, the CTO of Blackwell Electronics provided me with the sales data that contained 10,000 lines of data. From this data set, I found important general information regarding buying patterns that I will explain throughout this paper.

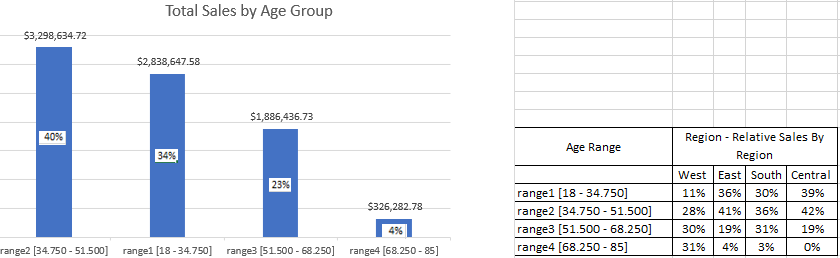


The above graph on the left, we can see that most of the sales occurred in the central region (50% of Total Sales) and the lowest was in the west (8% of Total Sales). The above picture on the right, gives us a breakdown of sales that occurred Online or In-Store according to their Region. Important to note, that In-Store sales accounted for 43% and Online 57% of all sales. In Central Region (which contained 50% of total sales), we can see that 40% of the sales occurred In-Store, while the other 60% occurred Online. Conversely, we can see that in West (Lowest sales region at 8%), all purchases occurred online. Regarding Danielle Sherman’s request to investigate average amount spent per region, we can see on the graph Labeled ““Average Amount

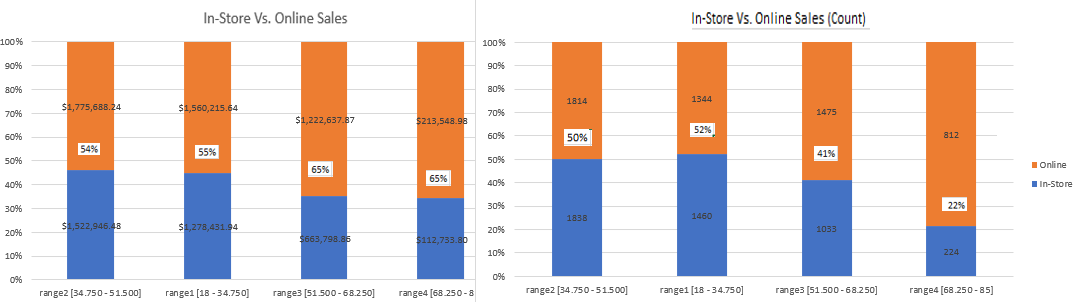


Spent Per Region”, the rank order matches which regions spent the most (Central > South > East > West). However, the graph to the right of it labeled “Count of Transactions” does not mirror this phenomenon. Although region West had the least sales at 8% of the total sales, this region had the 2nd most transactions (via solely online). One obvious conclusion we can come with is that Customers in West region purchased many items online that were most economical which I will get back to later.

Customer age is another factor I data mined. From the data set, I segmented the data for 4 age groups which are: Range1 [18 - 34.750], range2 [34.750 - 51.500], range3 [51.500 - 68.250], and range4 [68.250 - 85]. I did this so we can easily target and separate marketing campaigns from each other. Below is a screenshot of the mined data visually displayed. We can see general trends across age groups and regions.



One example, the graph on the left labeled “Total Sales by Age Group”, we can see that Range4 age group did poorly on sales at 4%. Moreover, the data set screenshotted on the right illustrates more in depth these sales broken down by region. Range4 age group essentially performed in West. On the other hand, our top performing age group, which is range2, accounted for 40% of all sales. Next closest is Range1 which accounted for 34% of all sales. Danielle also wanted insight into if there was any correlation between age of a customer and if the transaction was made online or in the store. The graphs below break down these sales by age group. As previously disclosed, although majority of sales occurred online, not all transactions occurred online.



It looks like the younger generation (Range1 [18-34.75]) did most of their transactions in store. As we move into the next segments of age ranges (Ranges 2-4, i.e. older generations), each age range is doing less and less transactions In-Store ([34.75-51.50 @ 50%], [51.5-68.25 @ 41%] and [68.25-85] @ 22%] and rather on-line.

Further analysis was made on segmenting the data of amount spent. These were seperated by ranges from range1 [5.23 - 1003.220], range2 [1003.220 - 2001.210], and range3 [2001.210 - 2999.20]. The below graph is a stacked 100% representation of the data retrieved. In our best performing region (Central), we can see

that the amount spent in the range 2 category accounted for over half of the sales that occurred (51%) and all ranges accounted for a significant percentage of the sales. In East Region, the range that was the most expensive [2001.21-2999.20] falls out. Lastly, in the west region, all purchase was in the most economical range [5.23-1003.22].

Given all the data that we have, we can make predictions on potential new customers. We can predict with 83% accuracy of predicting whether a new customer will be shopping online (74%) or in store (96%). Additionally, we can predict with a combined 71% accuracy whether a new customer will spend either in the most economical range [5.23 - 1003.220] (at 82%) or economical range [1003.220 – 2001.21] (at 64%). Lastly, Danielle wanted to know if we can predict region, in which we can at a combined 59% accuracy. We can predict if a new customer will be from South Region (at 49%), West Region (at 100%), East (at .15%) and Central (at 71%). With these insights, Blackwell Electronics can now move forward with more confidence regarding customer meta data and strategize their marketing campaigns to keep and attract customers.