Super-store Sales Marketing Analysis

Summer Lieberman
Jessica Maffe
James Santagati
Owen Fernandez

May 9, 2024

Introduction

For our final project, we utilized a data set about Superstore sales. With Tableau and Python, we utilized A/B Testing and Forecasting to interpret the data and find trends in the data that weren't otherwise explicitly stated for us. Our results are stated in the paper below. We will also discuss different segment results, compare Tableau and Python in terms of interpretations, as well as discuss our results in depth.

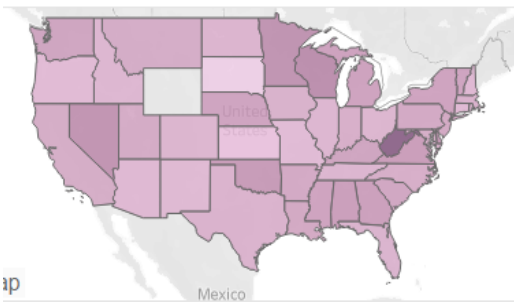
Python A/B Testing Results

We ran three A/B tests in python to determine any significant relationship. The first Python A/B testing results comparing segment vs segment had no significance. Two codes were then run comparing segment v. state. First being which state had the greatest significance. The second being a greater lesser test to find out the least significant states as well. The results showed that Consumer's greatest significant state was New York with a P-value of about 0.011. The least significant state was Illinois with a P-value of about 0.045. The A/B testing for Corporate showed the greatest significance state was Indiana with a P-value of about 0.000000058. The least significant state was Texas with a P-value of about 0.029. The last results were for Home Office, with Florida being the greatest significance state with a P-value of about 0.00114. The least significant state was again Texas with a P-value of about 0.036. These results tell us that these states have the highest significant sales by segment.

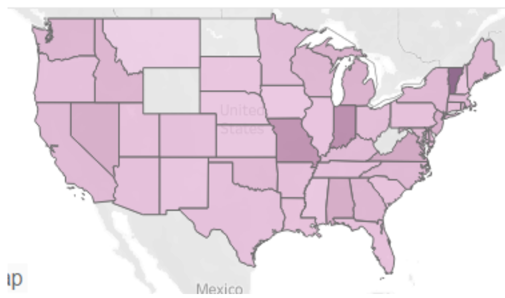
Initial Tableau State v Segment Results

After seeing the results of the A/B testing of state v. segment, we decided to create a visual comparing the same in Tableau. In Tableau we used average sales to determine what state for which segment seems to perform the best. Once this visual was created, we noticed a big discrepancy between what Python determined to be the most significant states and which states Tableau determined to have the most on average sales. We applied a filter to the EDA dashboard to easily toggle between the segments on the visuals. Results are as follows: Consumer- West Virginia, Corporate- Vermont, Home Office- Wyoming. This is obviously very different from the Python testing results.

Consumer:



Corporate:



Home Office:

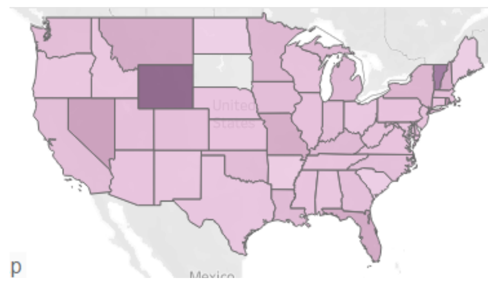
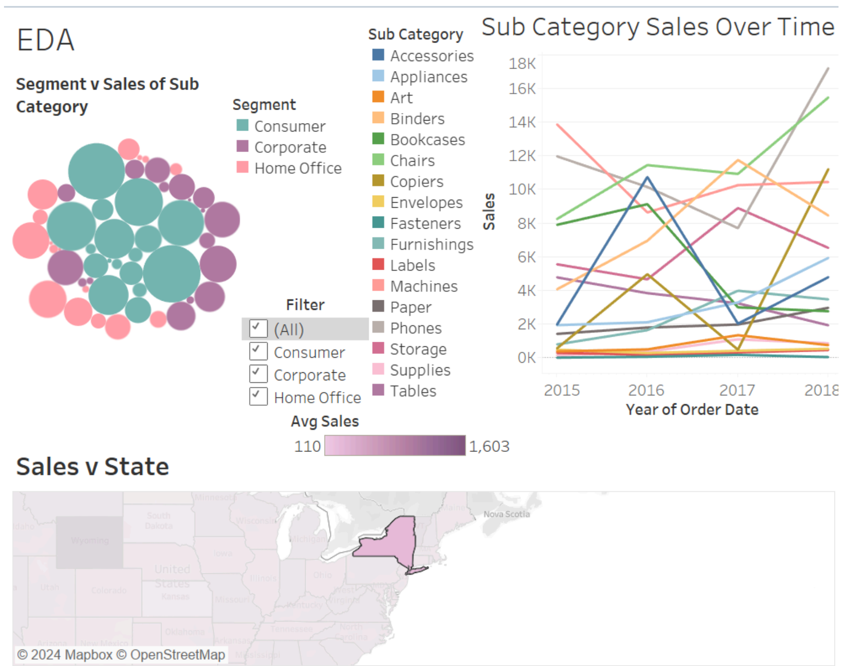


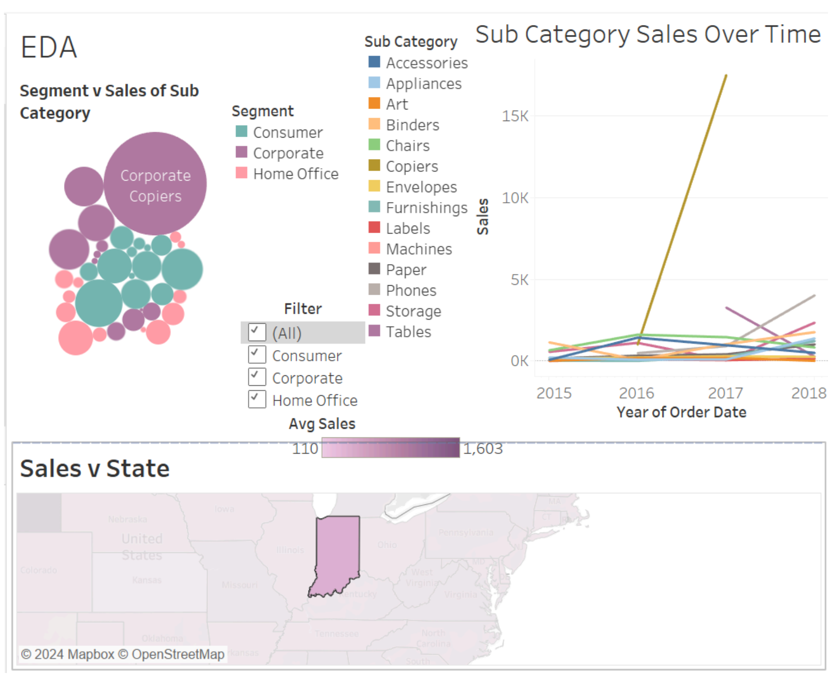
Tableau v. Python Comparison

As previously stated the results of the Python A/B testing were Consumer- New York, Corporate- Indiana, and Home Office- Florida. While the most average sales were Consumer- West Virginia, Corporate- Vermont, and Home Office- Wyoming. While at first glance this may seem contradictory, the results are just giving us different information. The Python testing is just simply telling us which state has the most significance. So while New York may have less sales on average for the Consumer segment than West Virginia, this may just mean the sales in New York are more significant/ higher profit than West Virginia. This would hold true for the other segments as well. We added another filter to our EDA dashboard to be able to see the differences in visuals by state. We used this to be able to see what segment dominated for the states Python came back with and what products are working best for them. The results of this comparison did show that those segments dominated their state. The Home Office dominated the state of Florida as seen in the visuals below. This also allowed us to determine what products are dominating in that state as well.

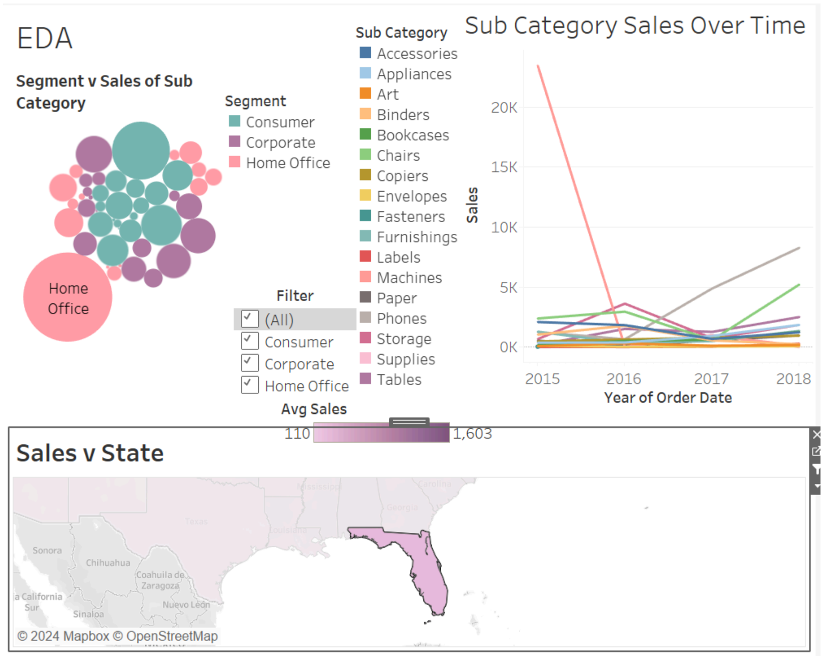
New York:



Indiana:

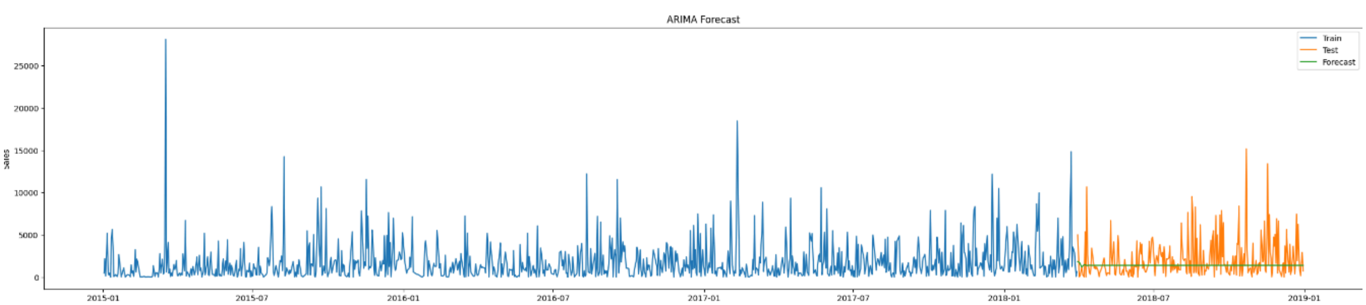


Florida:

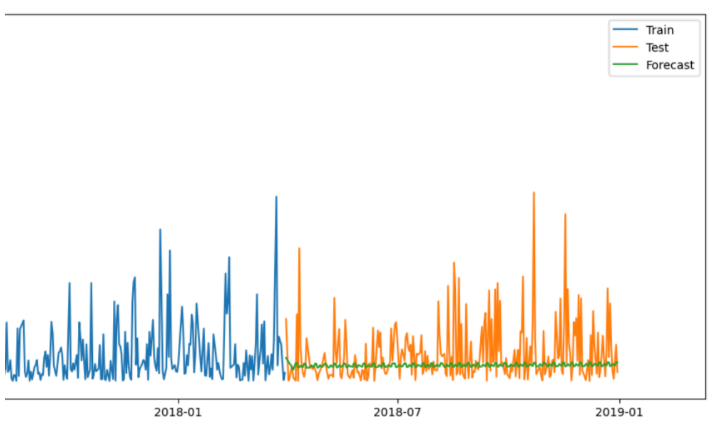


Python Forecast

For our python forecast, we ran two codes to determine a forecast. This first code produced the result as follows:



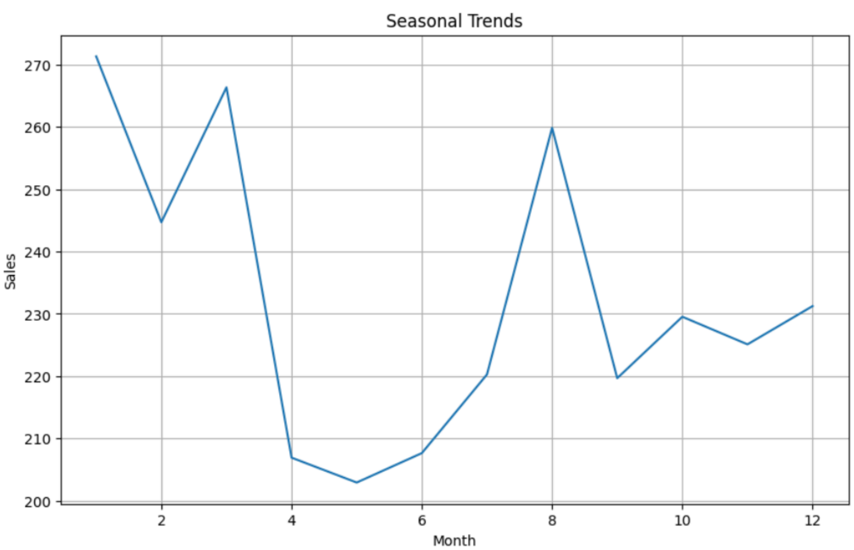
Not much could be interpreted from this forecasted, so we refined the code and produced as follows:



The spikes show increases in sales. So, Superstore would need to account for this increased demand by increasing inventory during those times. This would ensure smooth functioning of the business.

Python Seasonality

In terms of seasonality, we utilized Python to determine what months sales are the highest in and interpreted our findings. What we found was that sales were the highest in January, second highest in March, and third highest in August. From these results, we can interpret that sales will most likely follow a trend in the following years, with sales spiking in these months stated. We also ran a code to determine the moving average using python. Our moving average findings showed a similar pattern of seasonality.



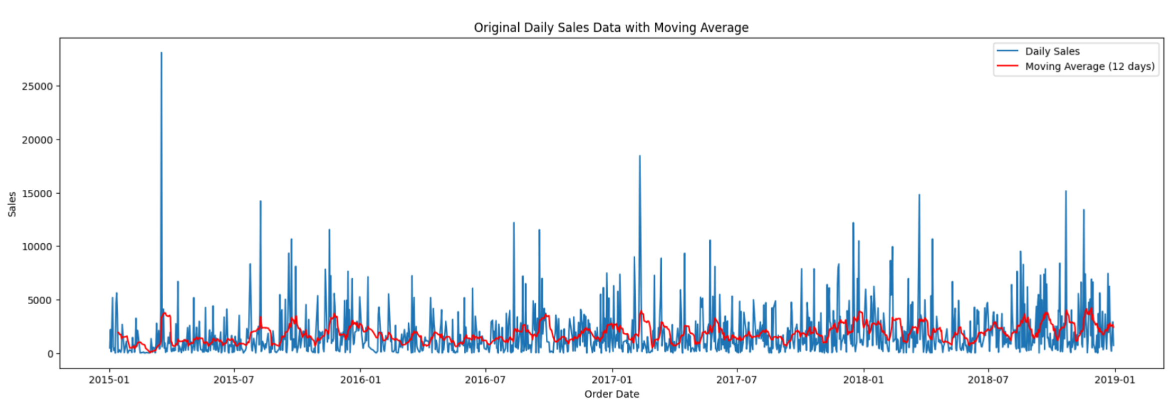
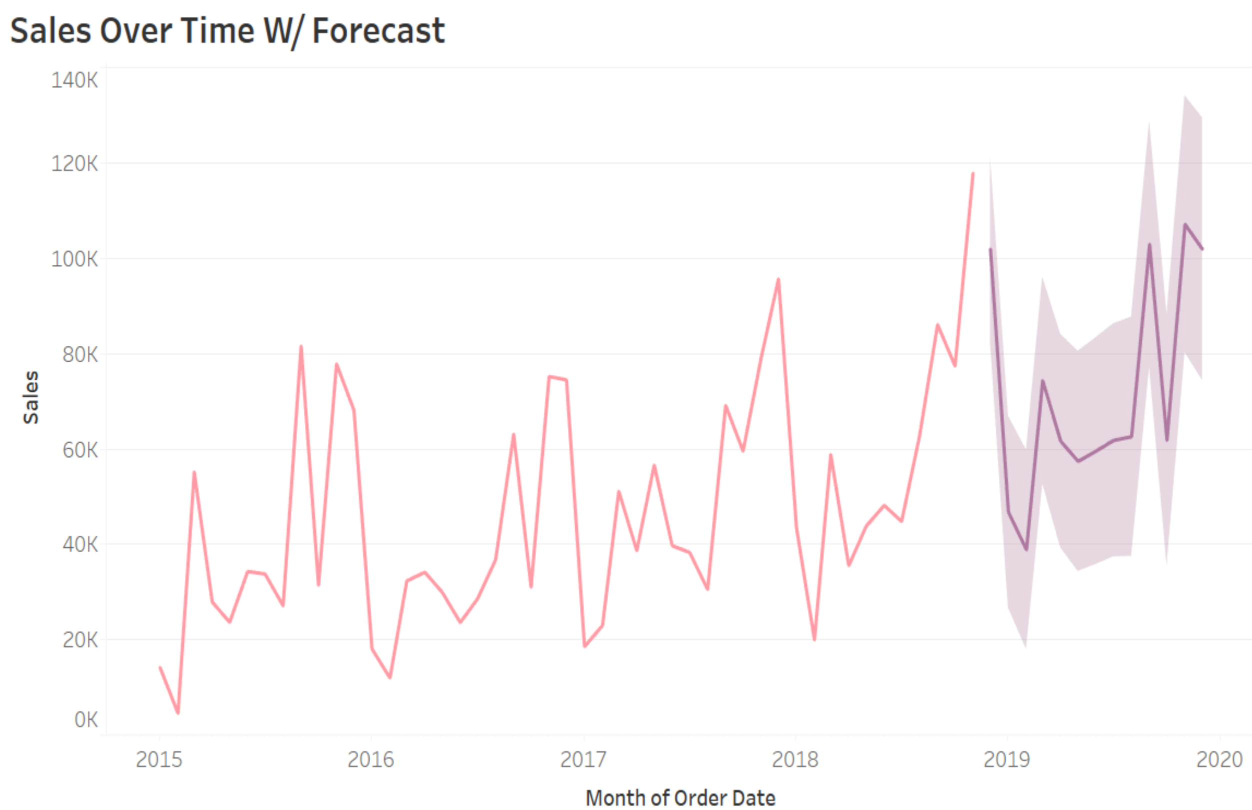


Tableau Forecast

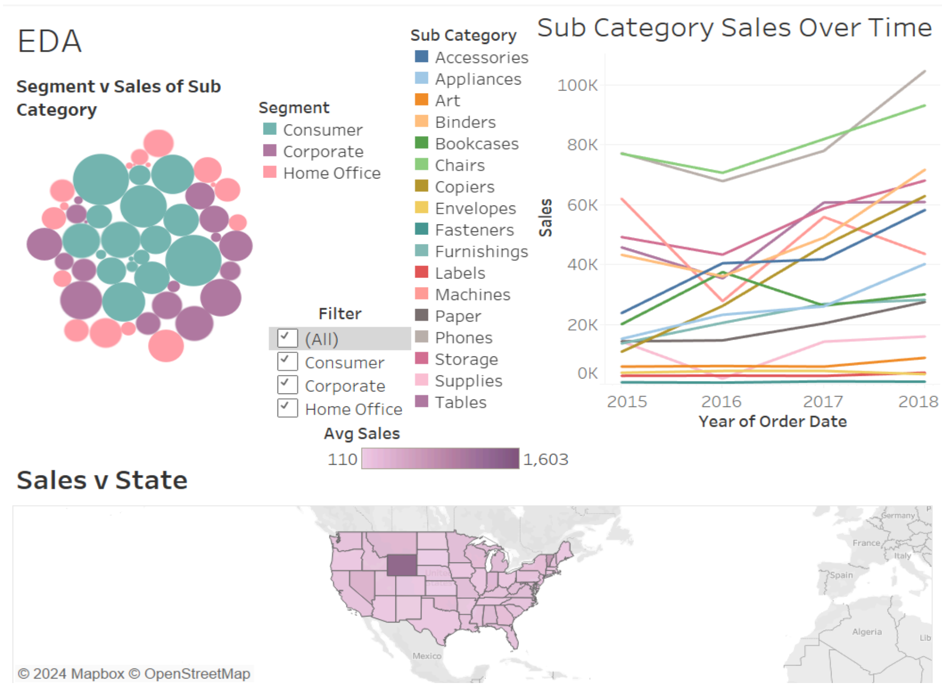
Our analysis revealed consistent findings between our Tableau and refined Python forecasts. Both visualizations identified sales spikes occurring during specific months of the year. Additionally, our examination across various products within the Tableau dashboard exhibited similar seasonal trends. The estimated sales figures displayed comparable spikes across different subcategories, affirming the reliability of our results.



Dashboards 1, 2, & 3

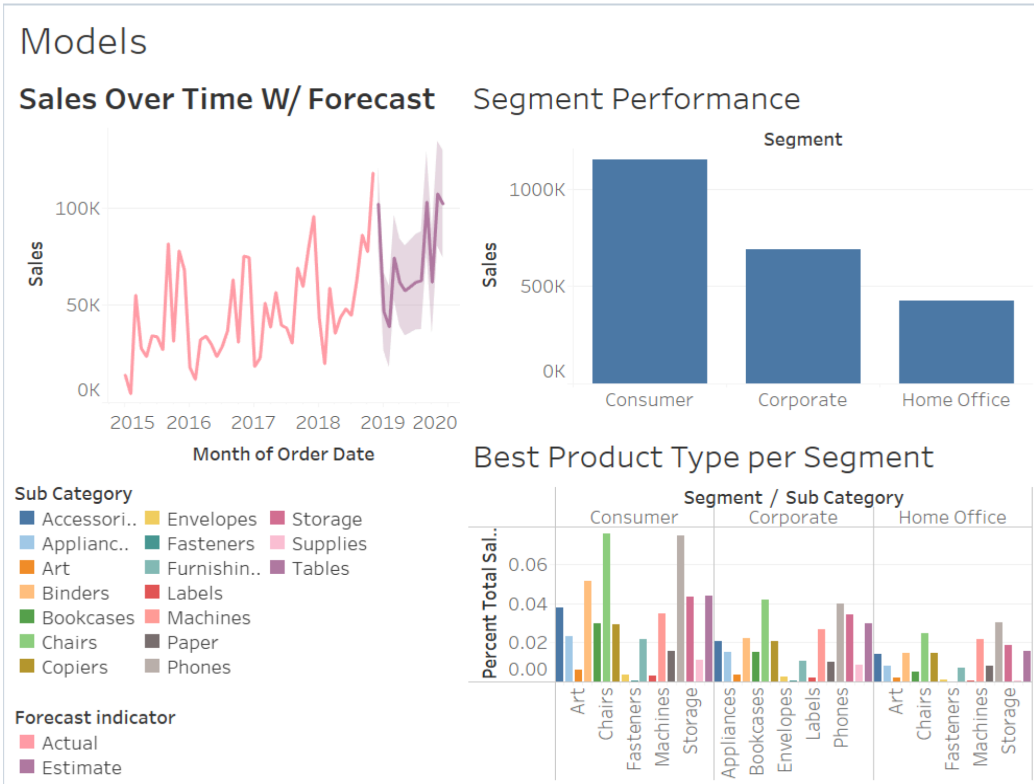
On Tableau we created three dashboards to determine any relationships that could potentially be influenced by a marketing strategy to increase sales.

Our first EDA dashboard determined a relationship based on state, segment, and sub category.



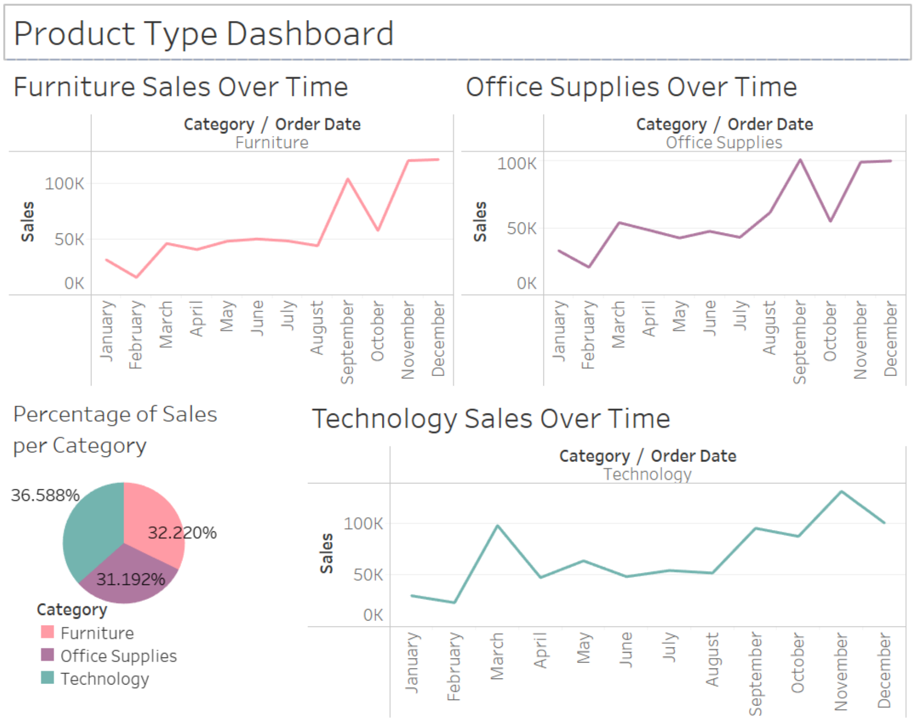
As previously mentioned we were able to determine insights regarding which segments and sub categories of products worked best for which state using the filters applied. This can be used to tailor a marketing strategy to each specific state. On this dashboard we are also able to see which category of products is outperforming the others in general. As well as which are performing the worst. Based off of the Sub Category Sales Over Time visual you can see that phones and chairs are the best performing products. While art, envelopes, labels and fasteners are performing the worst. Especially fasteners with sales near 0.

The second dashboard we created to further visualize the machine learning models we performed in Python.



The forecast as previously stated follows a similar pattern and seasonality of the Python forecast, as seen below. Some other insights that we can draw from this is that Consumer clearly outperforms the other two segments in regards to how many sales. But as we saw from the first significance test we ran there was no significant difference between the 3 segments. While Consumer out performs in number of sales this may mean the significance of type of sales even the segments out. You can also see what best product category works for which segment based on total percent of sales. Again phone and chairs dominate for each segment but we can also determine that binders are third best for Consumers, storage for Corporate, and machines for Home Office. It can also be determined which products may work for one segment but not another. Allowing us to further tailor our marketing strategy based on which market we are targeting.

Third dashboard we created with the intent of analyzing the main three categories of products which are split into: Furniture, Office Supplies, and Technology.



Again you can see a similar seasonality pattern from Python among the different product types. These graphs do steadily increase over time as sales increase across the years but Fall and Winter remain the better performing seasons, with a big spike in March. Furniture and Office Supplies both see a very big spike in September most likely due to back to school seasons. Sales again pick up in November for all products most likely due to Black Friday and the Christmas season as well. Technology sees a much bigger spike in March than the other products so researching this month may be worthwhile to adjust strategies accordingly. Product categories as seen in the pie chart are mostly evenly split among percentage of sales.

Recommendations

Based on the analysis we performed we have three main recommendations. First based on the results of both the Python A/B testing and the Tableau state comparison, it would not make sense to completely pivot and put most of the marketing resources into the three most significant states. While yes they potentially have more significant sales for the segments and are worth being invested too; they still do not produce the most sales on average compared to the other states. Instead all six should be invested into rather than one set over the other. The dashboard filters on the EDA dashboard can be used to tailor marketing strategies to each campaign. Based on which segments and products perform the best in each. Although the main six should have more of a priority.

Second recommendation would be to tailor strategies based on the seasonality. As it was determined Fall and Winter were the best performing seasons. August-September potentially due to the back to school season saw significant spikes in sales. Marketing strategies/ messages can be tailored to this during those seasons. March and November were most significant for specifically sales of technology potentially due to Black Friday and preparation of the Christmas season. We recommend looking into why March is so significant as well. November and December are also very significant for Office Supplies and Furniture; most likely due to the same reasons of Black Friday and Christmas.

Third and final recommendation is to focus resources more on marketing phones and chairs as they are the products that perform best across all segments. Messaging should be tailored to the specific segments. While those are the best performing products, marketing strategies should also be tailored around the other products that may only perform well for one segment and not the others. Such as binders which generated a lot of sales in the Consumer segment but not in the Corporate segment. It also may be worth it to completely drop the worst performing products such as fasteners that have near 0 sales. Or at least not invest any marketing resources into them. This will reduce costs and increase the amount of resources that can go to better performing products.