Trevor Maxwell

DSC 680

2024-05-25

Forecasting National Sales with Machine Learning

Optimizing organizational planning through machine learning methods can aid the growth and efficiency of the organization's operations. Forecasting can be performed for several different functions such as sales, supply chain, finance, and many more. With proper planning in these functions, companies can have a better idea of the sales initiatives for the future year, how many parts to order for future months, or where to invest within the company. Forecasting through machine learning algorithms can also supplement business knowledge for an even stronger forecast for the business (Fildes, Goodwin, 2007).

The dataset used in this analysis contained order, shipping, customer, and sales information from 2015 through the end of 2018. This company had a wide array of product offerings from office supplies to furniture and technology. The data included products ranging from tables and chairs to phones and copiers. In terms of the customer attributes, the data consisted of their ID number, name, and segment, and the shipping information for the customer consisted of the shipping method, city, state, postal code, country, region, and shipping date. In terms of the order attributes, the data consisted of the product name and ID number, order date, and sale of the item.

A few attributes were cleaned before the exploratory data analysis phase could be performed. Firstly, both order date and shipping date attributes had various date formats in the values, and this was cleaned to have a single date format for both columns. Secondly, some attributes that provided no value to the exploration and forecasting model; thus, these were

removed. Some of these columns included the row, product, and ID columns, the customer and product names, and a few shipping address attributes.

Understanding the sales of each product is important for a company to understand. This allows the organization to gauge their best-performing products and least-performing products.

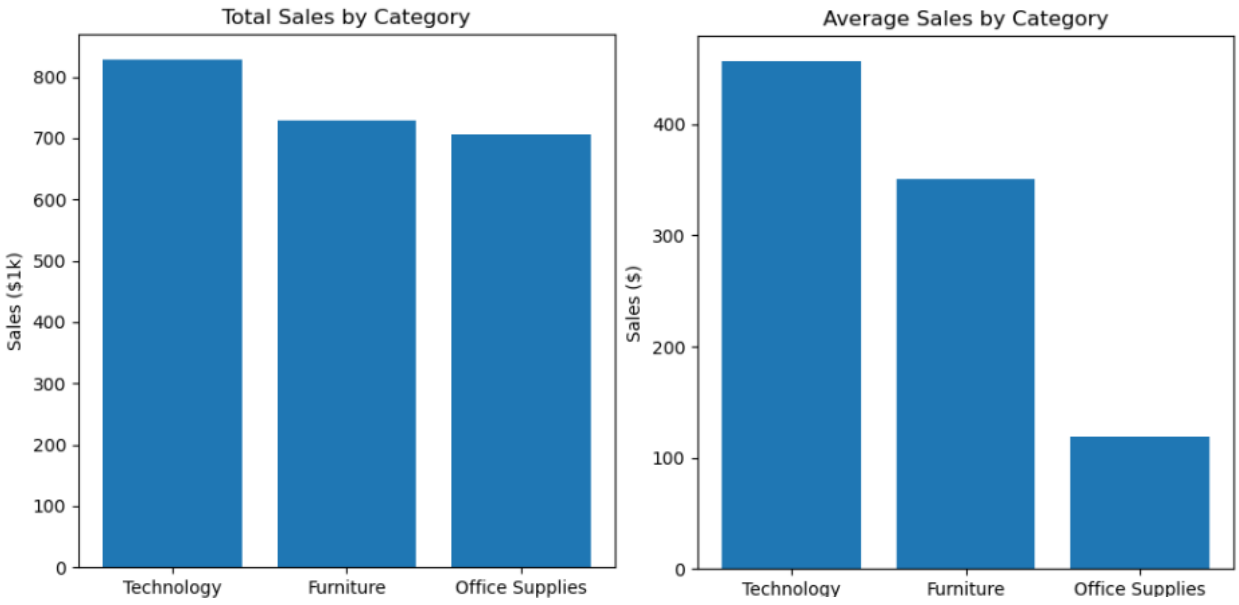


Fig 1: Total and average sales by product category

The technology products had the highest total and average sales out of all product categories. This indicates that technology products performed well. Office supplies have a total sales volume similar to furniture, however the average sales for this category were low. This indicates that the office supplies category still performed well due to the high volume of the products sold. It is also crucial to understand the market where the products were sold. One forecasting initiative by an electronics company determined what markets were most profitable and exited least profitable markets (Prevedere, 2022).

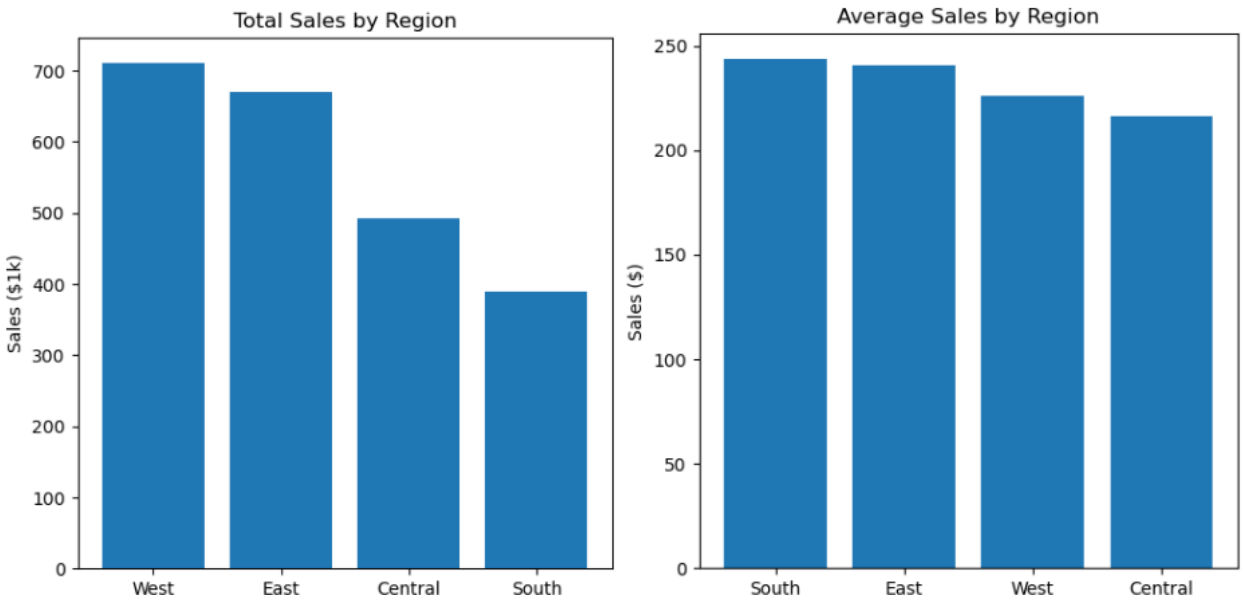


Fig 2: Total and average sales by market

The west region had the highest total sales amount with an average sales amount in the range of the other three regions. This indicates the west region was a well-performing market. The south market on the other hand had a significantly lower total sales amount compared to the rest which means this market did not perform well.

Before the forecasting models were deployed, the dataset was transformed. The sales were on an individual basis meaning each sale for each person was a separate record. The dataset was grouped by the date to provide the total sales on each day. Additionally, the forecast was monthly, thus the dataset was again grouped by each month to provide the total sales monthly.

After this transformation, the data was split into training and test datasets. The forecast was for future year sales, therefore the training set contained monthly volumes from 2015 through 2017, and the test set contained the 2018 monthly volumes.

Three models were fitted to the training set to forecast the 2018 data: ARIMA, SARIMA, and XGBoost. The predicted values were compared to the true values of the test set and the Root Mean Square Error (RMSE) was calculated. Based on the RMSE, the best performing model was

the XGBoost with an RMSE of 18,575. The forecasted and true values of the XGBoost model can be observed below.

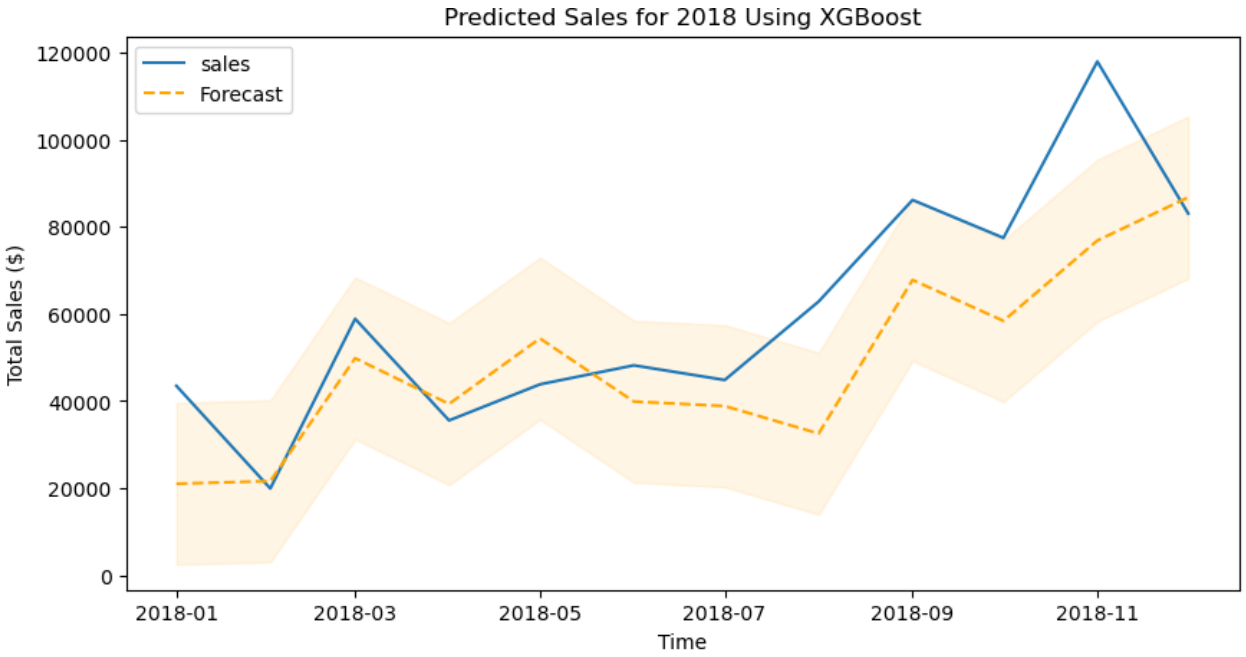


Fig 3: Forecasted and actual sales for 2018

With this forecast, the company will be able to predict the sales volumes across all products and markets for the future year. Organizational planning will benefit from this forecast by gaining the ability to create key performance indicators and sales initiatives to ensure the company meets or exceeds the projected volumes. With forecasting sales, the organization is also able to manage supply chain demands for the products that are offered. Overall, the company can use forecasting to aid the growth and success of the business.

Limitations/Challenges

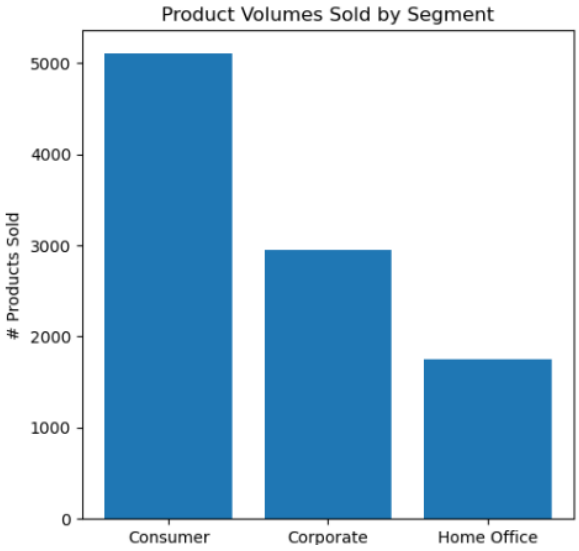
A few challenges arose throughout this analysis. The data was on an individual basis, therefore the data needed to be grouped, and some features were lost for one of the models. It would have been beneficial to have additional features for the XGBoost model which could make it more accurate. For example, if the data had the number of products sold in each market for each product in each month, that could aid the model's accuracy.

Ethical Considerations

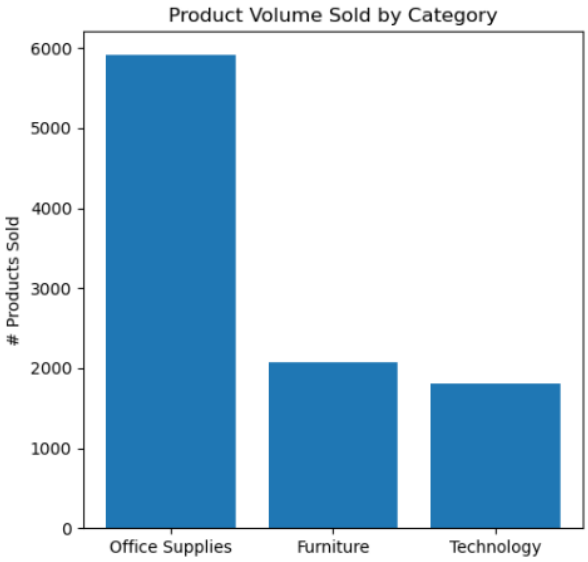
A few ethical considerations include the accuracy of the forecast. If the forecast is not accurate enough, the organization could deploy inaccurate planning for its operations and hinder the business. It is also important to be transparent and state the margins of error predicted in the forecast.

10 Questions an Audience Would Ask You

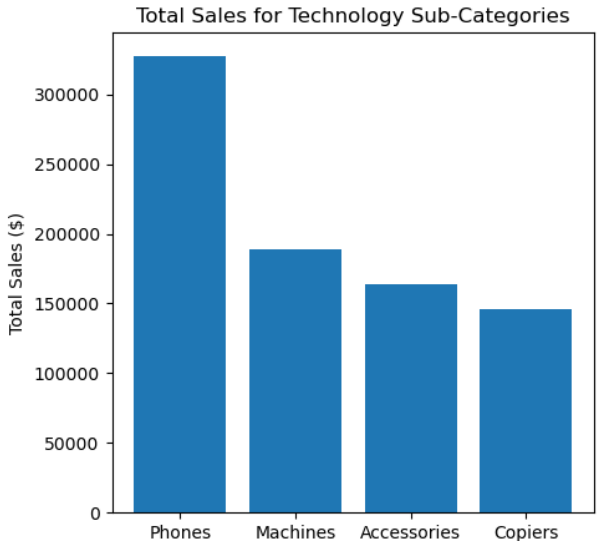
1. What is the number of products sold for each segment?



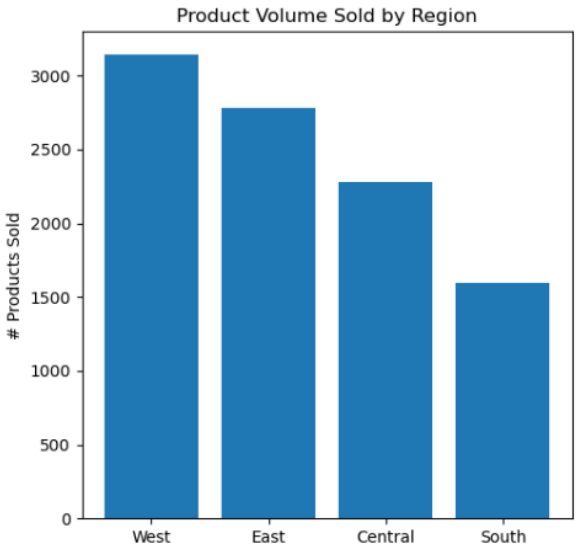
2. What is the number of products sold for each category?



3. What are the total sales by sub-category for the technology category?



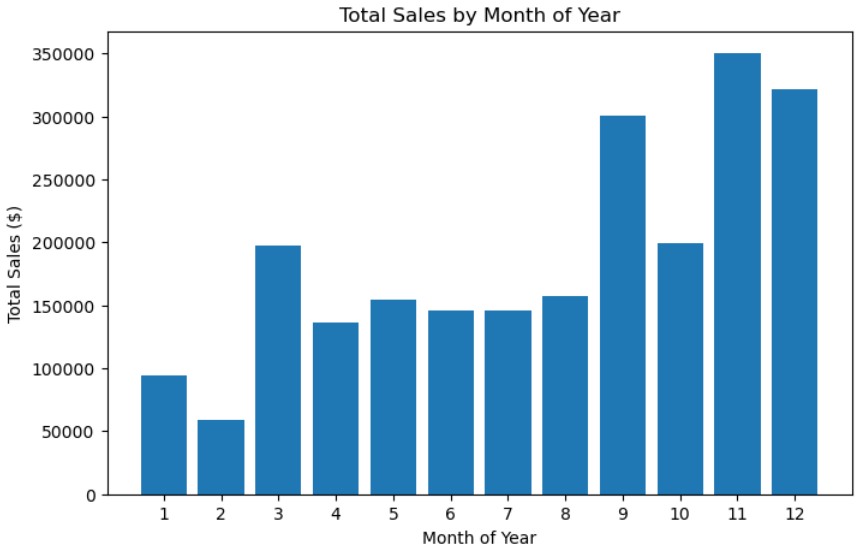
4. What is the number of products sold for each region?



5. Can this forecast be split to forecast sales in each market?
    a. **Yes, this current forecast is for all sales, but new forecasting models can be created for each market.**
6. What were the important features of the best model (XGBoost)?

	importances
month	89.9%
year	10.1%

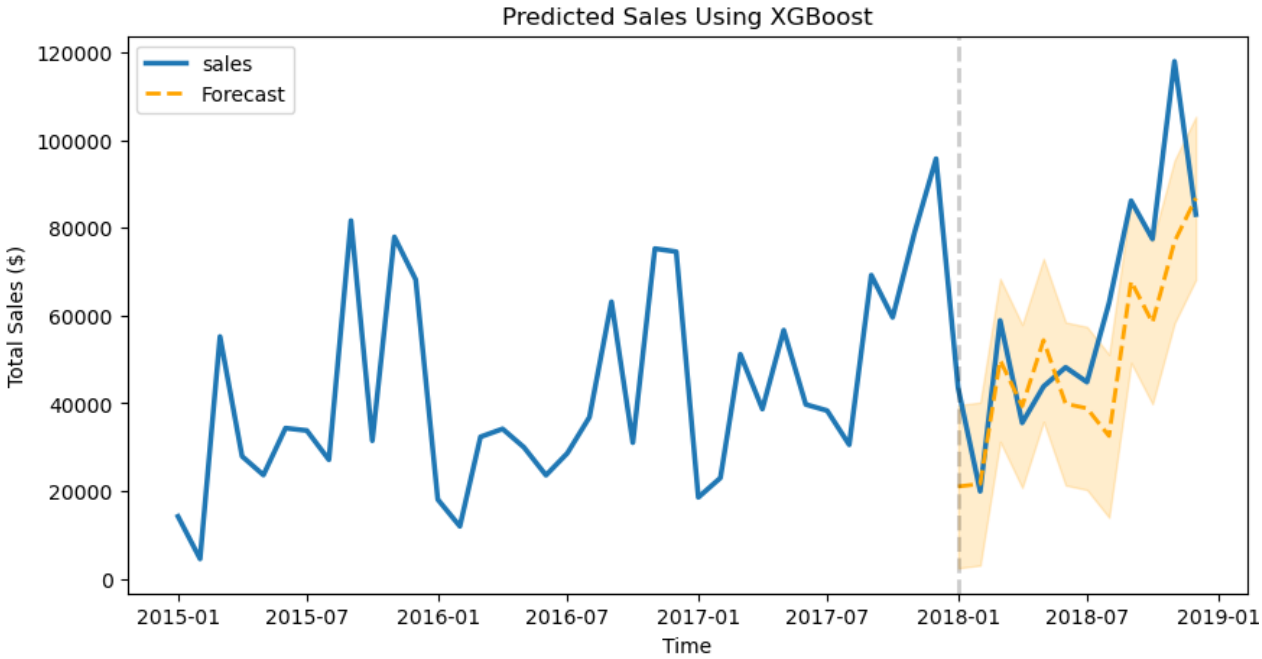
7. What are the total sales by month of the year?



8. What is the most popular shipping method?

9. Are there any states that have significantly higher sales compared to the rest?

10. Can we see the sales volumes from 2015 through 2018 along with the forecasted volumes?



References

Fildes, Robert, and Paul Goodwin. “Against your better judgment? how organizations can improve their use of management judgment in forecasting.” *Interfaces*, vol. 37, no. 6, 1 Dec. 2007, pp. 570–576, https://doi.org/10.1287/inte.1070.0309.

Prevedere. “3 Real-World Market Demand Forecasting Success Stories.” *Prevedere*, 20 Dec. 2022, prevedere.com/3-real-world-market-demand-forecasting-stories/.