Sales Data Forecast

Problem

Evaluate the performance of the supermarket operations through its sales data and forecast future sales revenue.

Solution

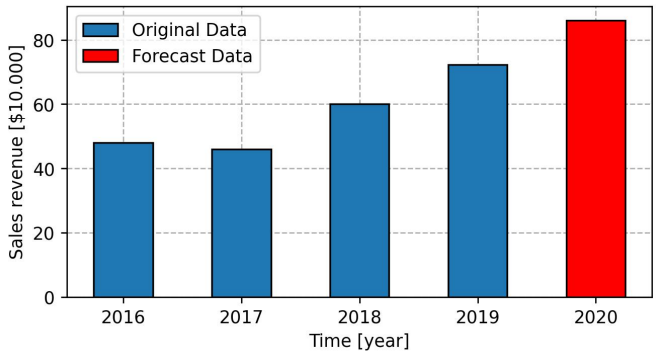
After an initial exploratory data analysis to get an overview of how the supermarket performed in the time span of the dataset, we are going to develop a model by decomposing the sales data into its seasonal and trend components using LOESS to understand the trend and possible patterns in the data.

Next Steps

1. Create a more complex model with SARIMA to try to improve the MAPE and R^2 metrics.
2. Retrieve information about the number of citizens and average income in the different Regions, States, and Cities to understand better the relationship between sales and location.

V. Ganci, GitHub:
https://github.com/Zipter?tab=repositories

Yearly sales revenue over the years and forecast



Key Insights

1. The West and East Regions had the highest sales revenue ($\sim$ \$710.000 and $\sim$ \$670.000). Follow the Central ($\sim$ \$490.000) and the South Region ($\sim$ \$390.000)
2. California was the State accounting for the highest percentage of the total revenue ($\sim 20\%$). Follow New York ($\sim 13\%$) and Texas ($\sim 7\%$).
3. New York City ($\sim 11\%$) and Los Angeles ($\sim 8\%$) distinctly accounted for the highest percentage of the total revenue.
4. While the office supply category dominated the order count ($\sim 60\%$ of the total orders), the technology category had the highest percentage of the total revenue ($\sim 36\%$). Office supply and furniture shared a similar total revenue percentage ($\sim 32\%$).
5. The phones and chairs sub-categories distinctly accounted for the highest percentage of the total revenue ($\sim 14\%$).
6. The Canon imageCLASS 2200 Advanced Copier was the item with the highest sales revenue ($\sim 3\%$ of the total revenue).
7. Except for a small drop in sales between 01.01.2016 and 01.01.2017, the sales increased in revenue yearly. With respect to the sales revenue of 2015, after 3 years the revenue increased by 50.5%.
8. The sales revenue shows a yearly seasonality, peaking between the months of October and January.
9. The model created by decomposing the weekly sales revenue through LOESS can explain 78.1% of the sales total variance ($R^2 = 0.781$). The model's predictions deviate on average by 10.6% from the actual values (MAPE = 10.6%).
10. The model forecasts that in 2019 the sales revenue will be $19.2 \pm 10.6\%$ larger than the ones in 2018.

Weekly sales revenue over the years and forecast