**EDA Summary**

Project 200 team

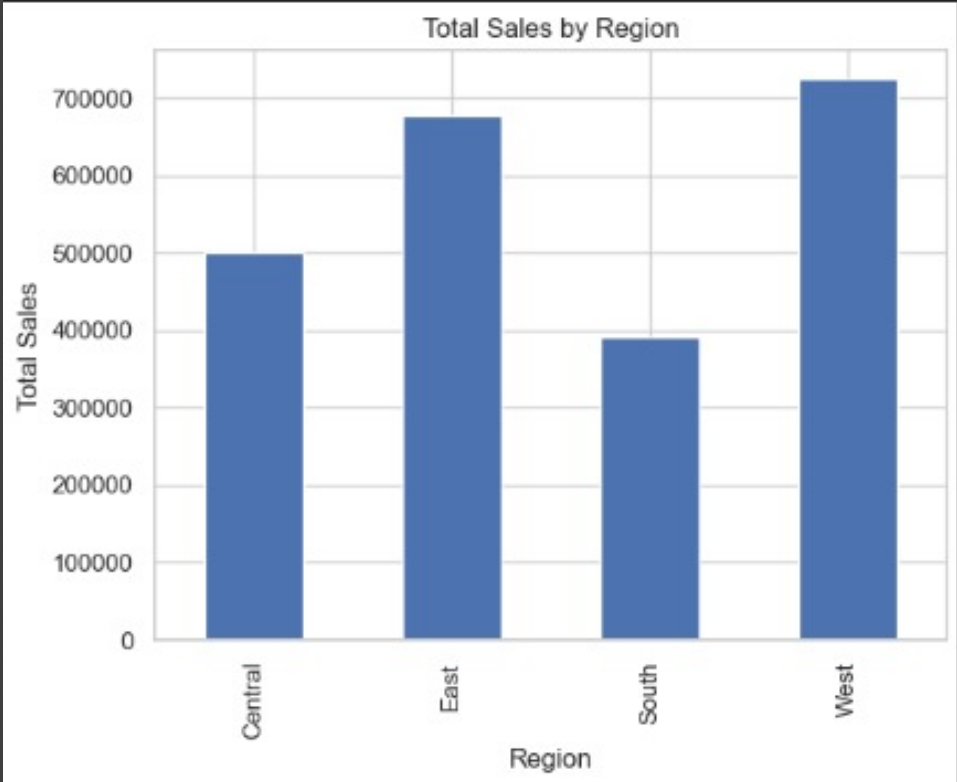
Westcliff University

DATA 200

Professor Regmi

June 2, 2025

**EDA Summary**



The chart shows the amount made in sales for each region. Once more, the West takes first place in total sales, followed by the East. More orders are placed in the Central region, but their total sales are not much greater, which suggests that the orders from Central are usually less valuable. The South has the fewest orders, but they are very valuable, showing that it earns more from each transaction.

A graph of numbers and a number of orders

AI-generated content may be incorrect.

The chart shows the number of orders placed in the regions of West, East, Central and South. West comes first and East follows in the number of orders. Well, the Central region makes less manpower demands and the South turns in the fewest orders. Therefore, the regions where the West and East lie tend to have higher numbers of customer transactions which means more sales.

A graph of sales by region

AI-generated content may be incorrect.

You can see from the chart how much each region sells on average through a single order. Sales levels in the West are higher than in the East and South, which have almost the same median sales. Median sales are the lowest in the Central part of the country. So, while Central gets numerous orders, the usual amount each order represents is quite low. Alternatively, in the South, because the number of orders is low, the median sales are quite high, suggesting larger and more important sales.

A graph with numbers and lines

AI-generated content may be incorrect.

According to the chart, the distribution of sales amounts is clearly right-skewed (positively skewed). Most sales, according to this statistic, are cheaper ones at the lower end of the value range. As the sales value rises, the number of these sales declines quickly which means large sales do not happen often. Even so, there’s a very long part of the graph on the right-hand side, proving that a few sales are much higher than the rest, possibly due to rare but major deals. In many retail or e-commerce businesses, there are many small transactions that make up the majority of income and only a small number of expensive products really influence the total revenue. It could be beneficial to use logarithmic transformation on this data to help the distribution be better distributed for statistics or visual research.

A graph showing sales by category

AI-generated content may be incorrect.

The box plot reveals how sales are divided among Office Supplies, Furniture and Technology. There are lots of different sale values in each category and the midpoint is shown with a central horizontal line. The lines going across the boxes mark the IQR (interquartile range) which makes up the middle half of the data. Whiskers on the boxplot show the area where typical sales values are found, which is 1.5 times the IQR. Each group shows several markers above the whiskers; these are sales numbers that stand out as being much higher than average. Technology Sales exhibits the highest and most repeated outliers, which suggests that while the majority of transactions are average, some can be very large. Even though they are outliers, Office Supplies and Furniture see lessen extreme trends than is seen in Technology. While the sizes of sales distributions vary, the main ones (excluding large deviations) look similar among categories. All in all, high-value sales happen in every category, although there are more and bigger transactions in the Technology segment.

A screenshot of a graph

AI-generated content may be incorrect.

It provides a visual example of how the different variables (such as Postal Code, Discount, Profit, Profit Ratio, Quantity and Sales) interact. The gradient from blue (negative) to red (positive) represents how strong the correlation is. Darker colors stand for stronger links. An important point is that there is a strong negative relationship between Discount and Profit Ratio (-0.86) which means higher discounts lead to a lower profit ratio. Discount and Profit are negatively correlated by a moderate amount (0.22), showing that when discounts are high, profit tends to be lower. Research shows that Sales and Profit have a moderate positive correlation (0.48), showing that as sales increase, profits are usually higher too. There is not much significance in the weak correlations among variables like Quantity and Profit (0.066) and Profit Ratio and Sales (0.0034). In general, the heatmap clearly illustrates the way different variables affect each other, and the impact of discounts causes the biggest drop in profitability.

A graph showing sales trend

AI-generated content may be incorrect.

The line chart shown represents the sales trend over time from early 2013 to early 2017. The x-axis denotes the timeline by date, while the y-axis represents the total sales values. Overall, the graph reveals a fluctuating sales pattern with several prominent spikes and troughs throughout the period. A particularly large spike can be observed around mid-2013, where sales peaked significantly above 25,000, marking the highest point on the chart. Although the sales trend does not show a perfectly steady growth, there appears to be a gradual increase in sales activity over the years, especially noticeable from 2016 onwards, where both the frequency and intensity of higher sales points become more consistent. These repeated spikes may indicate periodic high-sales events, such as promotional campaigns, holidays, or end-of-quarter pushes. Despite the fluctuations, the overall trend suggests increasing engagement and revenue generation as time progresses.