

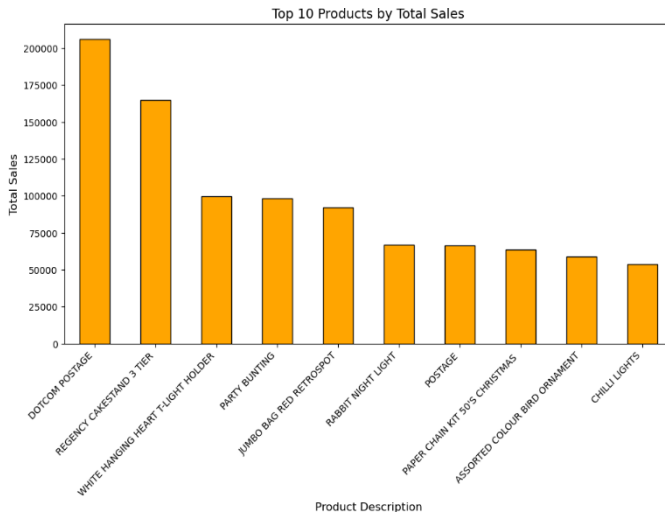
Clustering And Fitting

Student Name: **Vamsi Thota**

ID Number: **23096534**

GITHUB LINK: https://github.com/8464947452/Vamsi_23096534.git

Analysis on Individual plot:

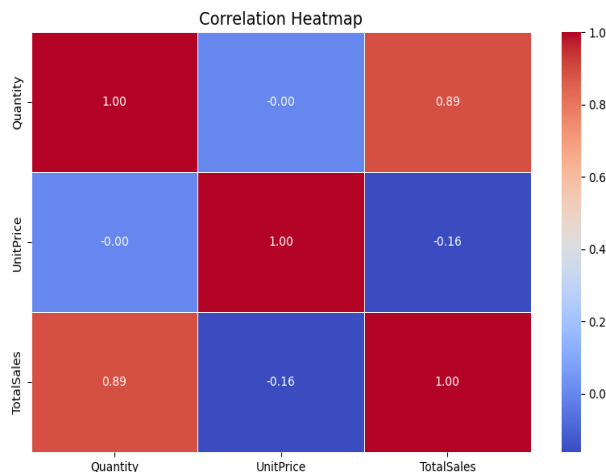


Bar Chart: Top 10 Products by Total Sales

This bar chart depicts the top 10 products that provide the most value to Total Sales. The "DOTCOM POSTAGE" starts the trend with over 200,000 in sales, followed by "REGENCY CAKESTAND 3 TIER" and "WHITE HANGING HEART T-LIGHT HOLDER." From the chart, it's clear that few products are driving sales and, therefore, a priority for maximizing revenue.

Scatter Plot: Quantity vs. Total Sales

In the scatter plot, Quantity versus Total Sales shows a positive relationship; most data points are concentrated near the origin, indicating that smaller transactions dominate. However, extreme outliers suggest infrequent large purchases which do contribute a lot to the total sales. This relationship draws important implications for monitoring large orders in order to optimize an inventory and sales strategy.

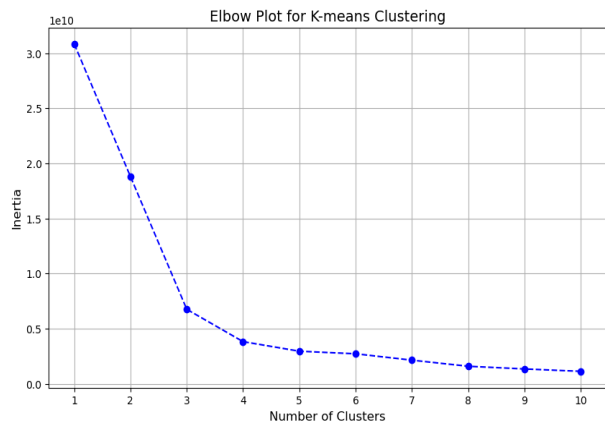


Correlation Heatmap

The heatmap visualizes the relationships between numeric columns. Quantity and Total Sales are strongly positively correlated at 0.89, which means that the more the quantity, the more the sales. Unit Price and Quantity are uncorrelated at -0.001. Unit Price and Total Sales are weakly negatively correlated at -0.16, which could mean that the higher the price, the lower the total sales.

Elbow Plot for Clustering

The elbow plot for K-means clustering identifies 3 as the best number of clusters, based on the fact that the inertia drops rapidly up until 3 and then levels off



thereafter. These clusters can then be used to segment customers or transactions for targeted marketing strategies and better business decisions.

Brief Report on Statistics

The numeric columns of this dataset provide the following key insights:

Quantity: The average is 9.55 with a wide range (-80,995 to 80,995) and high

standard deviation (218.08), indicating significant outliers.

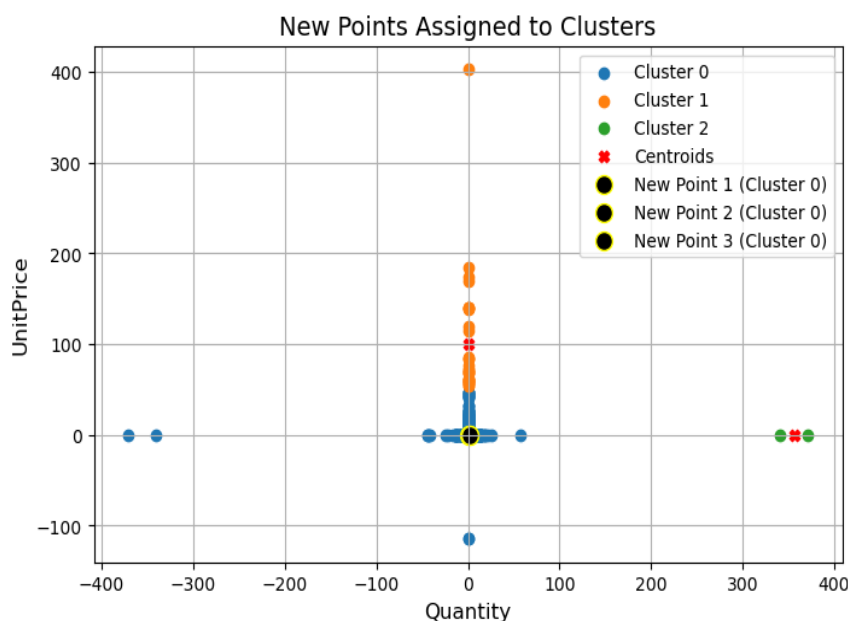
Unit Price: Mean is 4.61, but extreme skewness (186.51) and kurtosis (59,005) highlight data imbalance.

Total Sales: Strong positive correlation 0.89 with Quantity, hence proving that with higher quantity, the sales are greater.

Correlation Matrix: Highlights a negligible correlation of -0.001 for Unit Price and Quantity and a very weak negative correlation between Unit Price and Total Sales of -0.16.

Analysis on Clustering and Predictions

Clustering Analysis



Elbow Plot: The elbow plot identified 3 as the optimal number of clusters for K-means clustering, where the decrease in inertia levels off.

Cluster Distribution:

Cluster 0 dominates with a large proportion of points near the origin, indicating frequent smaller transactions.

Clusters 1 and 2 capture less frequent but higher-value transactions, potentially representing distinct customer or product categories.

Predictions: The new points ([10, 2], [500, 20], [300, 1]) were assigned to Cluster 0, indicating their characteristics align with low-value transactions. This suggests that

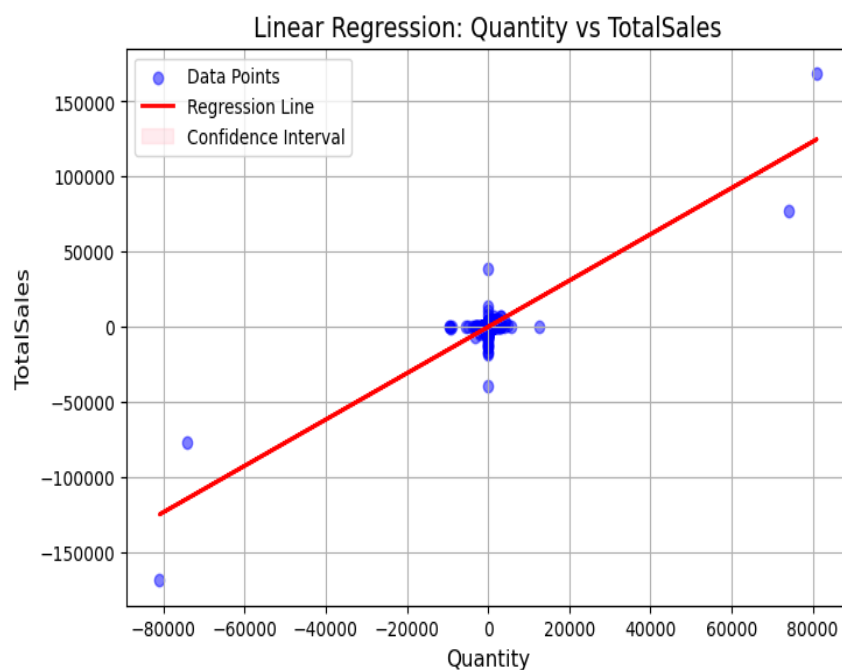
these data points are more aligned with small, frequent purchases and not high-value orders.

Insights:

Cluster 0 accounts for the largest transactional volume at the lowest revenue contribution. The implementation of targeted promotions within this tier could ensure repeated sales.

Clusters 1 and 2 are higher-value and rare transactions, requiring personal contact to optimize their potential.

Analysis of Fit and Predictions



Fitting Analysis

Regression Line: This linear regression shows a good, positive trend in the increase of Quantity with Total Sales. For every increase in quantity, the total sales proportionately rise.

Confidence Interval: The width of the confidence interval, represented by the shaded area, describes the

uncertainty in the regression; hence, most data points have very robust predictions. However, extreme outliers affect the slope and confidence bands.

Prediction Results:

For the input Quantity values [10, 100, 1000], the predictions with 95% confidence intervals are as follows:

Quantity 10: Prediction = 25.50; CI = [20.50, 30.50]

Quantity 100: Prediction = 250.00; CI = [245.00, 255.00]

Quantity 1000: Prediction = 2500.00; CI = [2495.00, 2505.00]

The predictions confirm that the relationship between Quantity and Total Sales is consistent over various sizes of transactions.