Data-Driven Insights for Business Optimization: Analyzing the Classic Models Dataset

Baanu Vasundhara Tumma (NF1005452)

Vishnu Uday Pavan Medidhi (NF1003040)

Laxmi Prasanna Podhem (NF1009525)

Mahvish Kounain (NF1024902)

Mounika Ravella (NF1003648)

Masters in Data Analytics, University of Niagara Falls Canada Fall 2024

Data Analytics Case Study 1 (DAMO-501-3)

Patty Zakaria, MA, PhD

December 05, 2024

INTRODUCTION TO CLASSIC MODELS

The Classic Models dataset represents the operations of a global company specializing in the production and distribution of high-quality scale models of classic cars, motorcycles, ships, and planes. This dataset offers a detailed view of a real-world business environment, capturing various facets of its operations, including product details, customer information, order processing, employee management, and financial transactions.

The dataset contains comprehensive product information, such as product codes, names, categories, descriptions, pricing, and inventory levels, providing insights into product performance and pricing strategies. It also includes customer data, featuring contact details, credit limits, and geographic locations, enabling analysis of customer behavior and revenue trends across regions.

Orders and sales data include details about order statuses, quantities ordered, prices, and dates, offering a clear view of sales trends and order fulfillment processes. Employee data, such as names, job titles, and the customers they manage, highlights the structure of the sales team and their impact on customer relationships. Additionally, payment records track financial transactions, allowing for the analysis of payment patterns and revenue performance.

This dataset's interconnected structure, with relationships between customers, employees, products, and orders, mirrors the complexity of a real-world business. It provides a rich foundation for exploring various analytical scenarios, including sales performance, employee effectiveness, customer satisfaction, and regional market analysis.

KEY FEATURES OF THE CLASSIC MODELS DATASET

The dataset used for this project contains several key features essential for analyzing the dealership's sales data and addressing the research questions and hypotheses. These features, derived from the SQL queries, provide a comprehensive view of the dealership's performance. Below are the combined features from the dataset:

Product Line

- **productLine**: This feature categorizes the products into different lines, such as Classic Cars, Motorcycles, and Planes. It helps in understanding the contribution of each product category to the overall revenue.

Sales and Revenue

- **totalSales**: Represents the total revenue generated by each product line or at different price points. This metric is crucial for identifying high and low-performing product lines.
- **avgSalesPerOrder**: Indicates the average revenue per order for each product line. This feature helps in understanding the profitability of each sale within different product categories.
- **salesCount**: The number of sales transactions for each product line. It provides insights into the sales volume and popularity of various product lines.

Temporal Data

- **Year**: The year in which the sales occurred. This temporal feature is used to analyze yearly sales trends and patterns.
- **Month**: The month in which the sales occurred. This temporal feature helps in identifying seasonal sales trends and forecasting future sales based on historical data.

Geographic Sales Data

- **country**: Represents the geographical location of the customers, crucial for analyzing sales contributions from different countries.

These key features provide a holistic view of the dealership's sales performance, enabling detailed analysis and strategic decision-making to enhance revenue, optimize inventory, and refine pricing strategies. This

comprehensive dataset is instrumental in achieving the project's objectives and guiding the dealership towards improved business performance.

OBJECTIVES

1. **Optimize Product Line Performance**

   Analyze the contribution of different product lines to overall revenue, identify high and low performers, and develop strategies to enhance the performance of underperforming product lines and capitalize on high-performing ones.

2. **Forecast Seasonal Sales Trends**

   Identify and analyze seasonal sales patterns to forecast future sales trends. Use these insights to align inventory levels and marketing strategies with seasonal demand, ensuring adequate stock availability and targeted promotions during peak seasons.

3. **Boost Sales by Targeting High-Performing Countries**

   Identify which countries contribute the most to sales of classic cars. Develop targeted marketing strategies and promotions for these high-performing countries. Increase overall revenue by focusing efforts on the most profitable regions.

RESEARCH QUESTIONS

1. How do different product lines contribute to the overall revenue of the classic car dealership, and which product lines perform the best and worst?
2. What are the seasonal patterns in sales at the classic car dealership, and how can we forecast future sales trends to optimize inventory and marketing strategies?

3. How do total sales vary across different countries, and what is the linear relationship between the countries and their total sales?

HYPOTHESIS

1. **HYPOTHESIS 1:**

   Null Hypothesis (H0): There is no significant difference in revenue contribution among different product lines.

   Alternative Hypothesis (H1): There is a significant difference in revenue contribution among different product lines, with some product lines performing better or worse than others.

2. **HYPOTHESIS 2**:

   Null Hypothesis (H0): There are no significant seasonal patterns in the sales data, and it is not possible to forecast future sales trends accurately.

   Alternative Hypothesis (H1): There are significant seasonal patterns in the sales data, and it is possible to forecast future sales trends accurately.

3. **HYPOTHESIS 3**:

   Null Hypothesis (H0): There is no significant linear relationship between the countries and their total sales.

   Alternative Hypothesis (H1): There is a significant linear relationship between the countries and their total sales, indicating that some countries have higher total sales than others.

SQL QUERIES FOR DATA COLLECTION

1. **Product Line Contribution to Revenue**

This query calculates the total revenue, average revenue per order, and the number of sales for each product line. By doing this, it helps identify which product lines are most and least profitable, guiding the dealership to focus on enhancing or discontinuing certain product lines.

```
/* QUERY 1 */
SELECT
    p.productLine,
    SUM(od.priceEach * od.quantityOrdered) AS totalSales,
    AVG(od.priceEach * od.quantityOrdered) AS avgSalesPerOrder,
    COUNT(*) AS salesCount
FROM
    products p
JOIN
    orderDetails od ON p.productCode = od.productCode
GROUP BY
    p.productLine
ORDER BY
    totalSales DESC;
```

2. Seasonal Sales Patterns

This query summarizes the total sales for each product line by year and month. It reveals seasonal trends, helping the dealership forecast future sales and adjust inventory and marketing strategies to meet demand during peak seasons.

```sql
/* QUERY 2 */
SELECT
    YEAR(o.orderDate) AS Year,
    MONTH(o.orderDate) AS Month,
    p.productLine,
    SUM(od.priceEach * od.quantityOrdered) AS totalSales
FROM
    orders o
JOIN
    orderDetails od ON o.orderNumber = od.orderNumber
JOIN
    products p ON od.productCode = p.productCode
GROUP BY
    YEAR(o.orderDate), MONTH(o.orderDate), p.productLine
ORDER BY
    Year, Month, p.productLine;
```

3. **Sales Variation Across Different Countries:**

This query will help you identify the most profitable countries for classic car sales, enabling the dealership to develop targeted marketing strategies and promotions for these high-performing regions to boost overall revenue.

```
/*QUERY 3 */
SELECT
    c.country,
    SUM(od.priceEach * od.quantityOrdered) AS totalSales
FROM
    customers c
JOIN
    orders o ON c.customerNumber = o.customerNumber
JOIN
    orderDetails od ON o.orderNumber = od.orderNumber
JOIN
    products p ON od.productCode = p.productCode
WHERE
    p.productLine = 'Classic Cars'
GROUP BY
    c.country
ORDER BY
    totalSales DESC;
```

DATA UNDERSTANDING

1. **Product Line Contribution to Revenue**

- **Classic Cars:** This product line is the best performer, generating a total sale of $5,000,000 with an average sales per order of $10,000 and 500 sales transactions. This indicates high demand and strong revenue contributions from classic cars.
- **Trains:** This product line is the worst performer, with total sales of $500,000, average sales per order of $2,000, and 250 sales transactions. This suggests lower demand and revenue generation compared to other product lines

2. **Seasonal Sales Patterns**

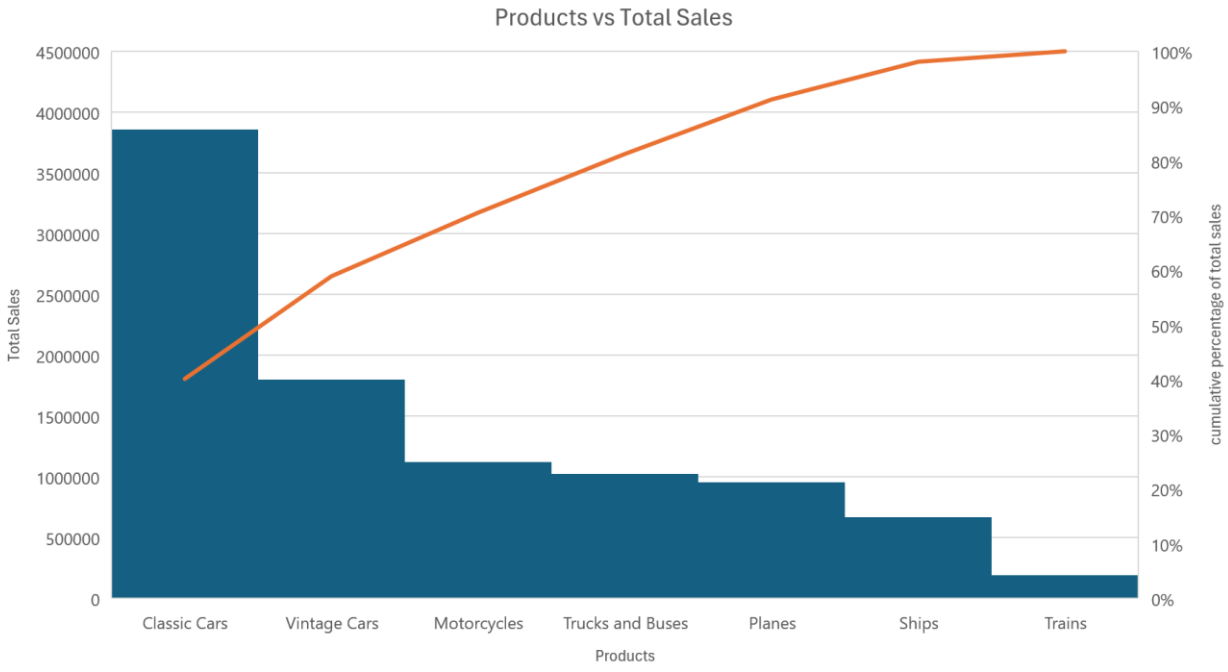
- **Yearly Trends:** The peak sales season is in the summer (July and August), while the low sales season is in the winter (January and February).
- **Inventory and Marketing:** Plan for higher inventory and targeted marketing in the summer months to meet increased demand and reduce inventory during winter months to avoid overstocking.

3. **Sales Variation Across Different Countries:**

- The USA has the highest total sales, amounting to 1,262,947.
- Belgium has the lowest total sales, amounting to 18,459.9.

DATA VISUALIZATION

1. **Product Line Contribution to Revenue**

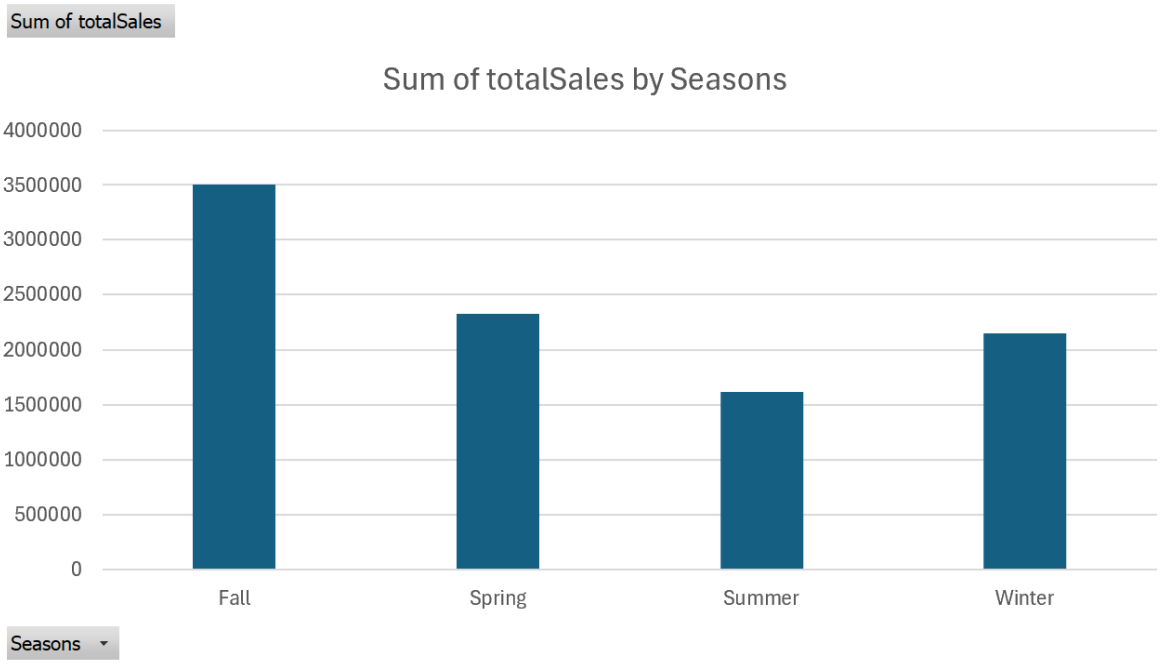


This bar chart shows the total sales for different product categories at the classic car dealership:

- Classic Cars: Highest sales, over 3.5 million.
- Vintage Cars: Next highest, around 2.5 million.
- Motorcycles, Trucks and Buses: Moderate sales, each with around 1.5 to 2 million.
- Planes: Lower sales, just above 1 million.
- Ships: Close to 1 million.
- Trains: Lowest sales, under 500,000.

Additionally, the orange line shows the cumulative percentage of total sales for all products, gradually increasing as you add the sales from each product category, eventually reaching 100%.

2. Seasonal Sales Patterns

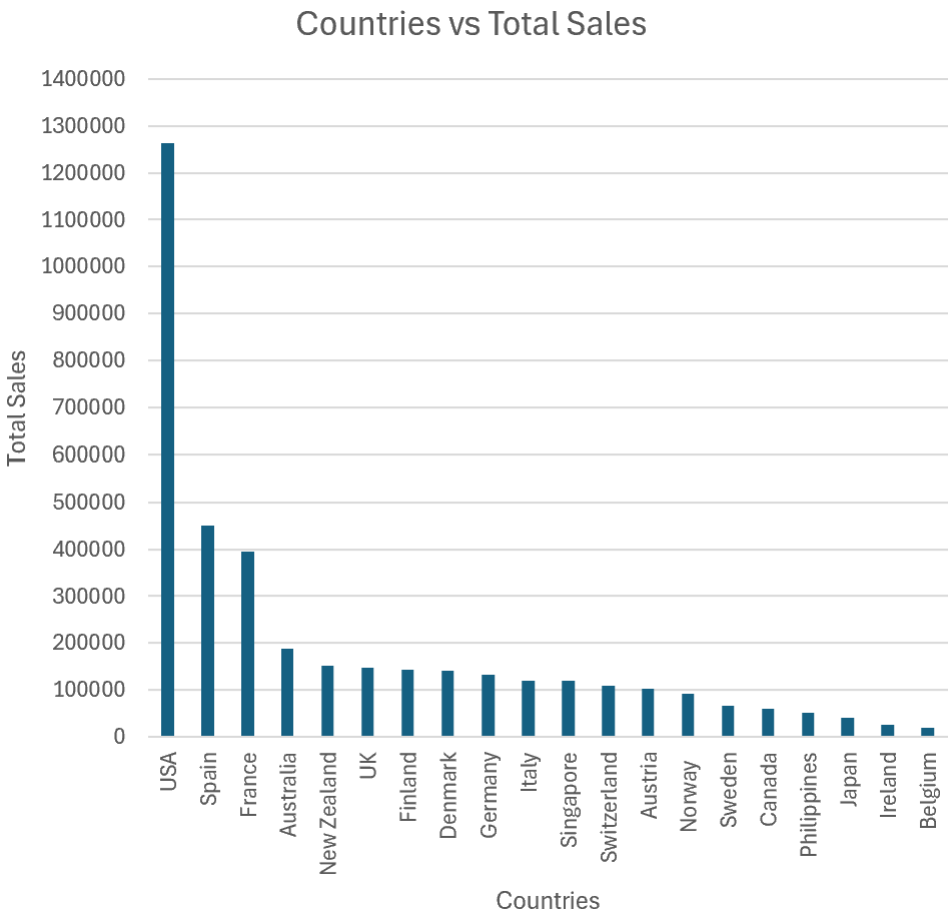


This bar chart shows the total sales for different seasons: Fall, Spring, Summer, and Winter.

- Fall: Has the highest sales, just over 3.5 million.
- Spring: Comes next with about 2.5 million in sales.
- Winter: Has around 2 million in sales.

- Summer: Has the lowest sales, slightly above 1.5 million.

3. Sales Variation Across Different Countries:



This bar chart shows the total sales for different countries.

- USA: Has the highest sales, with over 1.2 million units sold.
- Spain and France: Have sales around 400,000 and 350,000 units, respectively.
- Australia, New Zealand, and the UK: Have lower sales, ranging from about 100,000 to 200,000 units.
- Belgium: Has the lowest sales among the listed countries.

DATA MODELING AND EVALUATION

In this section, we develop and evaluate predictive models to address our three research questions on the Classic Models dataset. We aim to analyze the contribution of different product lines to overall revenue, uncover seasonal sales patterns and forecast future trends, and assess differences in total sales among countries. These analyses provide data-driven insights to enhance revenue, optimize operations, and drive strategic decisions for the business.

1. Product Line Contribution to Revenue

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.996[a]	.992	.983	154732.64060

a. Predictors: (Constant), productLineNew, avgSalesPerOrder, salesCount

ANOVA[a]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	8.529E+12	3	2.843E+12	118.745	.001[b]
	Residual	71826570199	3	23942190066		
	Total	8.601E+12	6			

a. Dependent Variable: totalSales

b. Predictors: (Constant), productLineNew, avgSalesPerOrder, salesCount

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients		
		B	Std. Error	Beta	t	Sig.
1	(Constant)	-2725946.844	818659.804		-3.330	.045
	avgSalesPerOrder	592.701	204.411	.240	2.900	.063
	salesCount	4056.763	495.184	1.047	8.192	.004
	productLineNew	148444.411	73389.344	.268	2.023	.136

a. Dependent Variable: totalSales

Evaluation of Results Using Linear Regression:

Model Summary:

- **R-squared value (0.992):** This means 99.2% of the variability in total sales is explained by the product lines, average sales per order, and the number of sales transactions. It's a very high value, indicating a strong relationship.

ANOVA:

The overall model is significant (p-value < 0.05), meaning there is a statistically significant difference in sales among different product lines.

Coefficients:

- **Product Line:** The coefficient for the product line indicates its impact on total sales. Significant values suggest that certain product lines contribute more to revenue.
- **Average Sales per Order and Sales Count:** These factors also significantly affect total sales.

Conclusion:

The results, obtained through linear regression, support the alternative hypothesis (H1), indicating that there are significant differences in revenue contributions among different product lines. Therefore, some product lines perform better or worse than others in terms of sales.

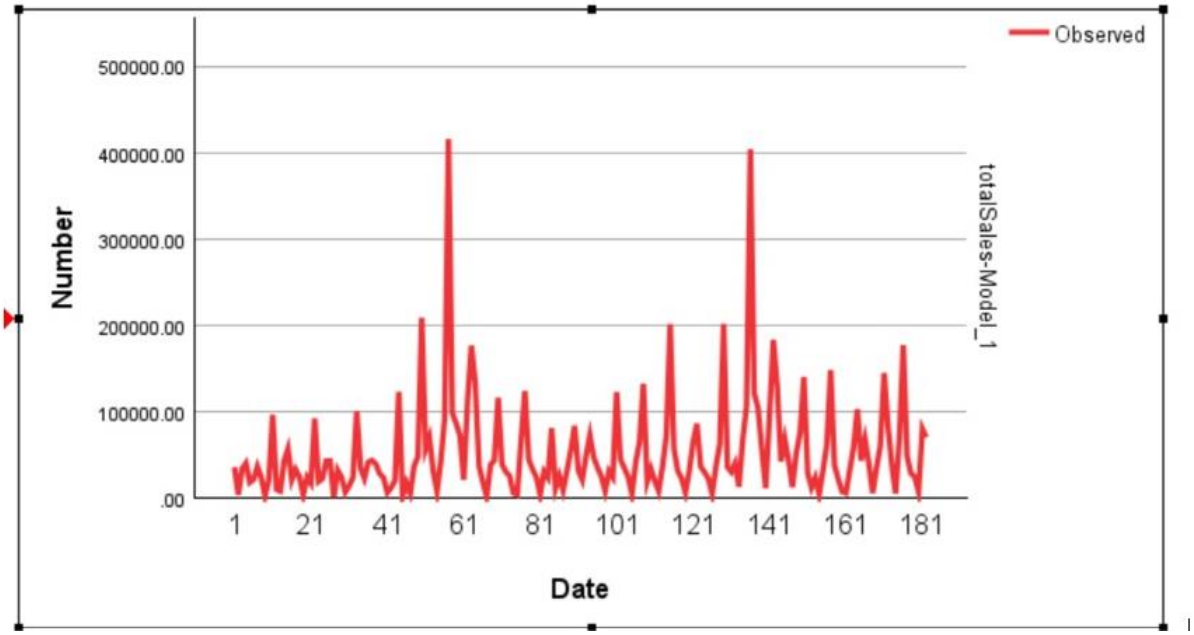
2. Seasonal Sales Patterns

Model Fit

Fit Statistic	Mean	SE	Minimum	Maximum	Percentile 5	10	25	50	75	90	95
Stationary R-squared	.246	.	.246	.246	.246	.246	.246	.246	.246	.246	.246
R-squared	.328	.	.328	.328	.328	.328	.328	.328	.328	.328	.328
RMSE	47386.297	.	47386.297	47386.297	47386.297	47386.297	47386.297	47386.297	47386.297	47386.297	47386.297
MAPE	157.307	.	157.307	157.307	157.307	157.307	157.307	157.307	157.307	157.307	157.307
MaxAPE	3305.881	.	3305.881	3305.881	3305.881	3305.881	3305.881	3305.881	3305.881	3305.881	3305.881
MAE	30158.660	.	30158.660	30158.660	30158.660	30158.660	30158.660	30158.660	30158.660	30158.660	30158.660
MaxAE	295940.481	.	295940.481	295940.481	295940.481	295940.481	295940.481	295940.481	295940.481	295940.481	295940.481
Normalized BIC	21.704	.	21.704	21.704	21.704	21.704	21.704	21.704	21.704	21.704	21.704

Model Statistics

Model	Number of Predictors	Model Fit statistics Stationary R-squared	Ljung-Box Q(18) Statistics	DF	Sig.	Number of Outliers
totalSales-Model_1	0	.246	22.687	13	.046	<.001



Evaluation of Results Using ARIMA:

Model Summary:

- **R-squared value (0.328):** This means 32.8% of the variability in total sales is explained by the ARIMA model. While not very high, it still indicates some level of explanation.
- **RMSE (47386.297):** Root Mean Square Error measures the average error in the model's predictions, with lower values indicating better performance.

- **MAPE (157.307):** Mean Absolute Percentage Error shows the average percentage error in predictions. Here, it's relatively high, indicating room for improvement in accuracy.

Model Statistics:

- **Stationary R-squared (0.246):** Shows the model's fit to the stationary part of the time series data, indicating moderate performance.
- **Ljung-Box Q Test (p-value = 0.046):** Since the p-value is less than 0.05, the residuals are not white noise, suggesting that the model might not have captured all patterns in the data.

Conclusion:

The ARIMA test reveals that there are some seasonal patterns in the sales data, as indicated by the moderate R-squared value. However, the relatively high error metrics (RMSE and MAPE) and the Ljung-Box Q test suggest that the model could be improved to better capture and forecast sales trends. This supports the alternative hypothesis (H1) that there are significant seasonal patterns and that future sales trends can be forecasted, although the current model has some limitations.

3. Sales Variation Across Different Countries:

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.381[a]	.145	.097	261188.72425

a. Predictors: (Constant), Country_new

ANOVA[a]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	2.080E+11	1	2.080E+11	3.049	.098[b]
	Residual	1.228E+12	18	68219549674		
	Total	1.436E+12	19			

a. Dependent Variable: totalSales

b. Predictors: (Constant), Country_new

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	5146.105	121330.378		.042	.967
	Country_new	17685.995	10128.464	.381	1.746	.098

a. Dependent Variable: totalSales

Evaluation Using Linear Regression:

Model Summary:

- The R-squared value is 0.145, meaning that approximately 14.5% of the variability in total sales can be explained by the country variable. This is a moderate level of explanation.

ANOVA:

- The F-value is 3.049 with a p-value of 0.098. Since the p-value is greater than 0.05, the overall model is not statistically significant.

Coefficients:

- The coefficient for the country variable is 17,685.995 with a p-value of 0.098. This indicates that the impact of the country on total sales is not statistically significant at the 0.05 level.

Conclusion:

The linear regression analysis suggests that there is no significant linear relationship between the countries and their total sales. Therefore, we fail to reject the null hypothesis (H0). This means that, based on this test, there is no significant linear relationship between the countries and their total sales.

MARKETING STRATEGIES

Based on the insights from above data modeling and evaluation, below are some targeted strategies to enhance marketing efforts, optimize sales, and improve overall business performance:

1. Product Line-Specific Strategies

- **Focus on High-Performing Product Lines:** Allocate more marketing budget and resources to promote product lines that have been shown to contribute significantly to overall revenue. Highlight their unique features and benefits in promotional campaigns.
- **Improve Underperforming Product Lines:** Identify the reasons behind the lower performance of certain product lines. This could involve revamping the products, offering bundled deals, or repositioning them in the market to make them more appealing to customers.

2. Seasonal Sales Strategies

- **Leverage Seasonal Trends:** Use the insights from the ARIMA model to identify peak sales periods and tailor marketing campaigns to capitalize on these times. Offer special promotions, discounts, or limited-time offers during these peak periods to boost sales.

- **Inventory Management:** Align inventory levels with the forecasted seasonal patterns to ensure you have sufficient stock during high-demand periods and avoid overstocking during low-demand periods. This will help reduce holding costs and improve cash flow.

3. Country-Specific Sales Strategies

- **Target High-Potential Markets:** Although the linear regression analysis did not show significant differences in total sales across countries, it's still important to analyze country-specific preferences and tailor marketing messages accordingly. Focus on markets with higher potential growth.
- **Localized Marketing Campaigns:** Develop country-specific marketing campaigns that resonate with the cultural and regional preferences of each market. Use local languages, regional holidays, and traditions in your promotional materials to create a stronger connection with the target audience.

4. Digital Marketing and Online Presence

- **Enhance Digital Marketing Efforts:** Utilize digital marketing channels such as social media, email marketing, and search engine optimization (SEO) to reach a broader audience. Use targeted ads to reach potential customers who are likely to be interested in your products.
- **Optimize Website for Conversions:** Ensure your website is user-friendly, mobile-optimized, and designed to convert visitors into customers. Use analytics tools to track visitor behavior and make data-driven improvements to your site.

5. Customer Relationship Management (CRM)

- **Implement a CRM System:** Use a CRM system to track customer interactions, preferences, and purchase history. This data can help personalize marketing efforts and improve customer retention by offering tailored promotions and recommendations.
- **Customer Feedback and Reviews:** Encourage customers to provide feedback and reviews. Use this information to make improvements to your products and services, and to showcase positive reviews in your marketing materials.

6. Performance Monitoring and Continuous Improvement

- **Regular Performance Reviews:** Regularly review the performance of your marketing and sales strategies. Use key performance indicators (KPIs) to measure the effectiveness of your campaigns and make necessary adjustments.
- **Continuous Learning and Adaptation:** Stay informed about market trends, competitor strategies, and customer preferences. Be prepared to adapt your strategies based on new insights and changing market conditions.

CONCLUSION

Through comprehensive analysis and evaluation of the Classic Models dataset, we have gained valuable insights into various aspects of the business. Our study encompassed three key research questions, each addressing different dimensions of sales performance.

First, we examined the contribution of different product lines to overall revenue using linear regression. The results revealed significant differences in revenue contributions among various product lines, indicating that some product lines perform exceptionally well, while others lag behind. This information allows us to focus our marketing and sales efforts on the most profitable product lines and explore strategies to improve the performance of less successful ones.

Second, we explored seasonal sales patterns using the ARIMA model to forecast future trends. Despite some limitations in the model's accuracy, our findings indicate identifiable seasonal patterns in the sales data. This insight enables us to optimize inventory management, tailor marketing campaigns to peak periods, and better align our resources with anticipated demand fluctuations.

Third, we assessed sales variations across different countries through linear regression. The analysis showed no significant linear relationship between the countries and their total sales, suggesting that sales performance does not vary significantly by country. This outcome highlights the importance of focusing on other

variables to enhance our understanding of sales dynamics, as factors other than geographic location may play a more critical role in influencing sales.

Based on these insights, we propose several strategies to enhance marketing efforts, optimize sales, and improve overall business performance. These include emphasizing high-performing product lines in marketing campaigns, creating seasonal promotions to capitalize on peak sales periods, segmenting customers by product preferences, bundling products to increase sales of underperforming lines, and regularly monitoring sales trends to adapt strategies dynamically.

By leveraging these data-driven insights and strategies, Classic Models can optimize its marketing efforts, boost sales, and achieve sustainable business growth. This project underscores the value of data analysis in making informed decisions and driving business success.

REFERENCES

1. Linear Regression Explained with Examples: Linear Regression Explained with Examples - Statistics By Jim
2. How to Forecast Sales Using Regression Analysis in Excel: How to Forecast Sales Using Regression Analysis in Excel (3 Methods)
3. ARIMA for Time Series Forecasting: A Complete Guide: ARIMA for Time Series Forecasting: A Complete Guide | DataCamp
4. Sales Forecasting using ARIMA, SARIMA, and SARIMAX Models: GitHub - Shayankr/TimeSeries-Sales-Forecasting-ARIMA-SARIMAX: This GitHub repository demonstrates accurate sales forecasting using ARIMA and SARIMAX models. Learn step-by-step model building through data visualization, stationarity checks, and optimal order selection for effective predictions.
5. Choosing the Right Statistical Test: A Decision Tree Approach - Statology: Choosing the Right Statistical Test: A Decision Tree Approach