|  |
| --- |
| **Data Set: Advertising** |
| **Question / Problem:**  1. Calculate the correlation coefficient (r value of the two variables)  2. Create a scatter plot with trendline, regression equation and r2 value  3. What percentage of the dependent is explained by the Newspaper advertising budget  4. If the Radio advertising budget is increased by €3527 what would the estimated sales increase by? 5. Comment on the general use of the model for estimating the sales increase based on the newspaper advertising budget |
| **Variables:**  **• ID -** product id (nominal - identifier)  **• Newspaper -** advertising expenditure allocated to newspaper in €1000s (numerical) **– Independent Variable**  • **Sales-** of a product in 200 different markets in 1000s of units (numerical) **– Dependent Variable** |
| **Techniques used:**   1. **Correlation Analysis**:    * **Purpose**: To determine the strength and direction of the linear relationship between Newspaper advertising expenditure and Sales.    * **Approach**: Calculated the correlation coefficient rr using a statistical tool (e.g., Excel’s CORREL function or similar). 2. **Scatter Plot with Trendline**:    * **Purpose**: To visualize the relationship between Newspaper advertising expenditure and Sales.    * **Approach**:      + Plotted Sales (y-axis) against Newspaper expenditure (x-axis).      + Added a **trendline** to represent the linear relationship.      + Displayed the **regression equation** and R2R^2-value on the chart for interpretability. 3. **Regression Analysis**:    * **Purpose**: To develop a linear regression equation for predicting Sales based on Newspaper advertising expenditure.    * **Regression Equation Format**:   Sales=b0+b1×Newspaper\text{Sales} = b\_0 + b\_1 \times \text{Newspaper}   * b0b\_0: Intercept (predicted Sales when Newspaper = 0). * b1b\_1: Slope (rate of change in Sales for each unit increase in Newspaper expenditure).  1. **Explanation of** R2R^2:    * **Purpose**: To determine how much of the variation in Sales is explained by Newspaper advertising expenditure.    * **Interpretation**: An R2R^2 value of 5.21% indicates a very weak explanatory power, suggesting that factors other than Newspaper expenditure have a greater influence on Sales. 2. **Predictive Model**:    * Used the regression equation to estimate the change in Sales when the Radio advertising budget increases by €3527. |
| **Analysis and Visualisation:**   * **Q1**   The correlation coefficient rr indicates a weak positive relationship between Newspaper advertising and Sales, confirming limited dependence.  **A white grid with black text  Description automatically generated**  **A close up of a text  Description automatically generated**  **A close up of a number  Description automatically generated**   * **Q2** * The scatter plot shows the individual data points for Sales vs. Newspaper expenditure. * The trendline represents the best-fit line for the data, with the equation and R2R^2 value annotated. * The points are widely scattered around the trendline, confirming the weak relationship indicated by the low R2R^2.   **A graph with blue dots  Description automatically generated**   * **Q3**       Only 5.21% of the variation in Sales can be attributed to Newspaper advertising expenditure. The remaining 94.79% is influenced by other factors.   * **Q4**       The low R² value means that Newspaper ads alone are not a strong predictor of Sales. Other factors (like TV or Radio ads) likely have a bigger impact on Sales.  Using the regression equation and the given slope b1b\_1 (from the trendline):   * ΔSales=b1×ΔRadio\Delta \text{Sales} = b\_1 \times \Delta \text{Radio}. * For ΔRadio=€3527\Delta \text{Radio} = €3527, calculate ΔSales\Delta \text{Sales}. |
| **Considerations:**  **Q5:** The data shown in the Scatter plot is grouped together making it hard to read. Using a Histogram or a Bar chart would be better to help visualise the relationship between Sales and Newspaper ads.  A histogram or bar chart could be used to compare aggregated data, but they wouldn’t illustrate the linear relationship as effectively.  Due to the low R2R^2, the model is unreliable for predicting Sales based solely on Newspaper expenditure.  The model isn't very reliable for estimating sales increases based on the newspaper advertising budget. Here’s why:  **Strengths**   1. **Easy to Use**: The model gives a quick way to estimate sales based on newspaper spending without needing complex tools. 2. **Basic Insight**: It shows that there’s a small link between newspaper ads and sales, which can help you understand trends at a high level.   **Weaknesses**   1. **Weak Connection**: Only 5.21% of sales changes can be explained by newspaper spending. This means other factors have a bigger influence. 2. **Not Accurate for Decisions**: The model isn’t precise enough to rely on for major business decisions. It might lead to wrong conclusions. 3. **Too Simple**: It doesn’t consider how other advertising channels, seasons, or market differences might impact sales.   **How to Improve**   1. **Add More Data**: Include other factors like TV, radio, and online ads to get a clearer picture of what drives sales. 2. **Try Non-Linear Models**: Sales might not increase evenly with ad spending, so testing more flexible models could reveal better patterns. 3. **Focus on What Works**: Newspaper ads don’t seem to have much impact. It might be better to invest more in other channels like TV or radio.   **Conclusion**  The model is fine for basic insights but not great for accurate predictions or strategic decisions. To make better estimates, more data is needed and a more detailed approach. |