**Analysis of Sales Report of a Clothes Manufacturing Outlet**

**Introduction & Objective**

A high-end fashion retail store seeks expand its product offerings by understanding current market and sales drivers. Using two months of sales and product attribute data, this analysis aims to:

* Predict which dresses should be recommended (Task 1)
* Forecast future sales trends (Task 2)
* Assess how style, season, material, price affect sales (Task 3)
* Identify leading factors influencing sales (Task 4)
* Evaluate ratings impact sales (Task 5)

**2. Data Preparation & Cleaning**

* **Data sources:**
* Attribute DataSet.xlsx: Dress attributes (style, price, rating, size, season, etc.)
* Dress Sales.xlsx: Sales per dress on multiple dates
* **Cleaning steps:**
* Removed duplicates by aggregating sales and keeping unique attribute rows.
* Converted all relevant columns to factors or numerics as appropriate.
* Fixed typos and standardized levels in categorical variables.
* Merged datasets on Dress\_ID.
* Created a Total\_Sales variable by summing all sales columns per dress.

**3. Exploratory Data Analysis (EDA)**

* **Style Distribution:**  
  The most common style was "Casual," followed by "Sexy" and "Party." Other styles were less frequent.
* **Price Distribution:**  
  Most dresses were priced as "Average" or "Low," with fewer in the "High," "," or "very-high" categories.
* **Ratings Distribution:**  
  Ratings were skewed toward higher values, but a significant number of dresses had a rating of zero.
* **Season Distribution:**  
  Dresses were distributed across all seasons, with Summer and Winter being most common.

**4. Task 1: Recommendation Prediction Model**

A Random Forest classifier was built to predict whether a dress would be recommended (binary outcome) based on its attributes (style, price, size, season, neckline, sleeve length, material, fabric type, decoration, pattern type, and rating).

* **Model Performance:**
* Accuracy: 61%
* AUC: 0.60
* Sensitivity: 59%, Specificity: 64%
* Balanced Accuracy: %
* **Interpretation:**  
  The model performs only slightly better than random guessing. Product attributes alone do not strongly predict recommendations, suggesting that other unasured factors may play a role.

**5. Task 2: Sales Trend Prediction**

A time series ARIMA model was fitted to the sales history for each dress. For example, forecasting was performed for one sample dress, projecting sales for three additional periods.

* **Interpretation:**  
  The method provides a basic approach for short-term inventory planning, but accuracy may limited due to the short time window and variability in sales.

1. Task 3: Influence of Style, Season, Material, Price on Sales

* **ANOVA Results:**
* **Price** significantly affects sales (p = 0032).
* **Style**, **Season**, and **Material** were not significant at p < 0.05.
* **Multiple Regression:**
* R²: 0.105 (low explanatory power)
* Only some style levels (e.g., 'sexy') were individually significant.
* **Interpretation:**  
  Price is a key driver of sales, while style, season, and material have weaker effects in this dataset. The overall explanatory power of these variables is low.

**7. Task 4: Leading Factors Affecting Sales**

To identify product attributes most strongly influence total sales, a random forest regression model used and variable importance was examined.

* **Top 5 predictors (by IncNodePurity):**

1. **NeckLine** – The neckline design of the dress has the highest impact on sales.
2. **PatternType** – The pattern or print type (e.g., animal, dot, solid) is also highly influential.
3. **Decoration** – Features like ruffles, lace, embroidery contribute significantly to sales.
4. **Rating** – Dresses with higher ratings tend to sell more, though the effect is moderate.
5. **Style** – The overall style category (e.g., Casual, Sexy, Vintage) is also important.

* **Business meaning:**  
  This suggests that design-related features (neckline, pattern, decoration) are more critical for driving sales than price or size. Management should focus on these aspects when planning inventory and marketing.

**. Task 5: Does Rating Affect Sales?**

The relationship between dress rating and total sales was analyzed using correlation and regression.

* **Correlation:** The Pearson correlation coefficient between rating and total sales is weakly positive (r = 0.18).
* **Regression:**  
  Linear regression shows the relationship is statistically significant (R² = 0.12, p < 0.001), but only about 12% of the variation in sales can be explained by rating alone.
* **Interpretation:**  
  Although higher-rated dresses tend to sell more, the effect is weak. Other factors play a much larger role in driving sales performance.

**9. Conclusions & Recommendations**

* Product recommendations cannot be reliably predicted attributes alone; consider incorporating customer or behavioral data.
* Price is a significant driver of sales, but design features such as NeckLine and PatternType are even more influential according to machine models.
* Ratings have a statistically significant but modest effect on sales; focus on quality but also on other differentiators.
* Inventory planning should prioritize styles and materials with proven sales, but also monitor market trends and customer preferences.
* Further data collection (.g., marketing activities, promotions, online engagement) could improve predictive modeling.

**10. Limitations**

* Short sales period (two months) limits long-term trend analysis.
* Potential subjectivity in recommendation labels.
* Some variables (e.g., waiseline, pattern type) may need further cleaning or consolidation.