#### BUSINESS CASE 2: SALE MARKETING PROJECT

#### Goal

The primary goal of this sales data project is to leverage historical sales data to gain insights into sales trends, optimize marketing strategies, and improve overall business performance. By analyzing various aspects of the sales records, we aim to enhance decision-making, increase revenue, and improve customer satisfaction.

#### Objectives

1. **Identify Sales Trends:**
   * Analyze sales data to identify seasonal patterns, growth rates, and other significant trends over time.
   * Understand how sales, revenue, and profit fluctuate across different periods (e.g., monthly, quarterly).
2. **Optimize Product Performance:**
   * Determine which products (Item Type) are high-performing in terms of units sold, total revenue, and total profit.
   * Identify underperforming products and investigate potential reasons for their lower sales.
3. **Enhance Regional and Country Insights:**
   * Examine sales performance across different regions and countries to identify top-performing markets.
   * Understand regional and country-specific trends and preferences to tailor marketing strategies accordingly.
4. **Improve Sales Channel Effectiveness:**
   * Compare the performance of different sales channels (Online vs. Offline) in terms of revenue, profit, and units sold.
   * Identify the most effective sales channels and allocate resources to maximize their impact.
5. **Refine Customer Insights and Segmentation:**
   * Segment customers based on purchasing behavior, order priority, and other relevant factors.
   * Develop targeted marketing campaigns for different customer segments to increase customer engagement and loyalty.
6. **Evaluate and Optimize Marketing Campaigns:**
   * Analyze the impact of different marketing efforts on sales, revenue, and profit.
   * Identify the most effective marketing campaigns and strategies to optimize marketing spend.
7. **Forecast Future Sales:**
   * Develop predictive models to forecast future sales based on historical data.
   * Use these forecasts to plan inventory, staffing, and marketing efforts more effectively.
8. **Increase Operational Efficiency:**
   * Identify inefficiencies in the sales process and recommend improvements.
   * Optimize inventory management by aligning it with sales trends and forecasts to prevent overstocking or stockouts.
9. **Generate Actionable Insights:**
   * Provide actionable insights and strategic recommendations based on data analysis.
   * Support decision-making with clear and concise reports that highlight key findings and opportunities for growth.
10. **Deliver High-Quality Reports and Code:**
    * Ensure the analysis is well-documented, with clean and reproducible code.
    * Produce concise reports that summarize the key insights, model parameters, and results effectively.

By achieving these objectives, the project aims to provide a comprehensive understanding of the sales dynamics, enabling the business to make data-driven decisions that drive growth and efficiency.

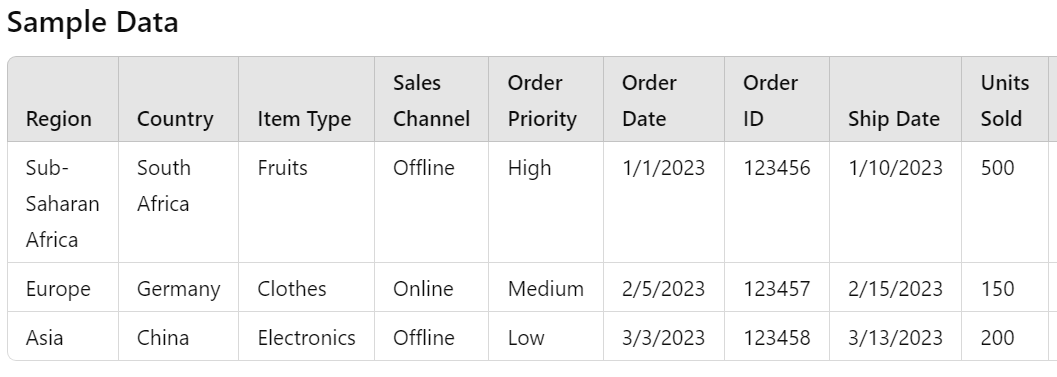
**Data Description**

The provided dataset in "100000 Sales Records.csv" contains information about sales transactions. Below is a detailed description of each column in the dataset:

| **Column Name** | **Description** |
| --- | --- |
| **Region** | The geographic area where the sales transaction occurred (e.g., Sub-Saharan Africa, Europe, Asia). |
| **Country** | The country where the sales transaction occurred. |
| **Item Type** | The type of product sold (e.g., Fruits, Clothes, Electronics). |
| **Sales Channel** | The channel through which the sale was made (e.g., Online, Offline). |
| **Order Priority** | The priority level of the order (e.g., High, Medium, Low). |
| **Order Date** | The date when the order was placed. |
| **Order ID** | A unique identifier for the order. |
| **Ship Date** | The date when the order was shipped. |
| **Units Sold** | The number of units sold. |
| **Unit Price** | The retail price per unit of the product. |
| **Unit Cost** | The cost per unit of the product. |
| **Total Revenue** | The total revenue from the sales transaction (calculated as Units Sold \* Unit Price). |
| **Total Cost** | The total cost associated with the sales transaction (calculated as Units Sold \* Unit Cost). |
| **Total Profit** | The total profit from the sales transaction (calculated Total Revenue -Total Cost). |

**Column Significance**

1. **Region and Country:**
   * Analyzing sales data by region and country to understand geographic market performance.
2. **Item Type:**
   * Identifying best and worst-performing product types.
3. **Sales Channel:**
   * Evaluating the effectiveness of different sales channels (Online vs. Offline).
4. **Order Priority:**
   * Understanding the impact of order priority on sales and shipping efficiency.
5. **Order Date and Ship Date:**
   * Analyzing time-based data to identify seasonal trends and shipping timelines.
6. **Units Sold, Unit Price, and Unit Cost:**
   * Detailed assessment of quantities sold, pricing, and costs to calculate financial metrics.
7. **Total Revenue, Total Cost, and Total Profit:**
   * Key financial metrics to assess overall business performance.



This dataset provides a comprehensive view of the company's sales operations, including geographic, product type, sales channel, and financial metrics. By analyzing these factors, the company can make strategic decisions to optimize sales and marketing efforts, improve revenue, and increase profitability.

#### REQUIREMENTS:

#### Exploratory Data Analysis (EDA)

1. **Data Understanding and Cleaning:**
   * What are the key variables in the dataset? (e.g., Region, Country, Item Type, Sales Channel, Order Priority, Order Date, Order ID, Ship Date, Units Sold, Unit Price, Unit Cost, Total Revenue, Total Cost, Total Profit)
   * Are there any missing values or outliers that need to be addressed?
2. **Price and Volume Curves:**
   * How are the Unit Price and Units Sold curves plotted and interpreted?
   * What insights can be drawn from the price and volume trends?
3. **Time-Based Analysis:**
   * How do sales, revenue, and profit change over time (e.g., monthly, quarterly)?
   * What are the seasonal trends or patterns observed in the data based on Order Date and Ship Date?
4. **Regional and Country Analysis:**
   * How do sales, revenue, and profit vary across different regions (e.g., Region) and countries (e.g., Country)?
   * What are the top-performing regions and countries in terms of Total Revenue and Total Profit?

#### Model Development and Evaluation

1. **Model Selection and Parameters:**
   * What models are being considered for predicting sales, revenue, or profit?
   * How are the best model parameters chosen (e.g., through hyperparameter tuning)?
2. **Training and Testing:**
   * What data splitting method is used to train and test the models (e.g., train-test split, cross-validation)?
   * How are the models evaluated based on the training and testing results?
3. **Metrics for Evaluation:**
   * What are the key metrics used to evaluate the models (e.g., accuracy, precision, recall, F1 score)?
   * How are these metrics interpreted and used to compare the performance of different models?

#### Evaluation Criteria

1. **Analytical Skill:**
   * How effectively do the teams analyze the data and interpret the results?
   * Are the insights from the EDA meaningful and actionable?
2. **Model Performance:**
   * How well do the models perform based on the chosen metrics?
   * Is there evidence of thoughtful hyperparameter tuning to optimize model performance?
3. **Overall Quality:**
   * How well is the report structured and written?
   * Is the code clean, well-documented, and easy to understand?

### Summary of Key Requirements

Each team is required to:

* Conduct a thorough EDA, including plotting and interpreting Unit Price and Units Sold curves, and analyzing sales data by time (Order Date, Ship Date), region (Region), and country (Country) based on Total Revenue and Total Profit.
* Develop and evaluate models, selecting the best parameters through hyperparameter tuning and splitting data for training and testing.
* Prepare a concise report (less than 3 A4 pages) focusing on EDA, chosen model parameters, and training/testing results.
* Submit the report and code to the organizers by the specified deadline.
* Ensure their work is evaluated based on analytical skill, model performance, and the overall quality of their report and code.

By addressing these requirements and questions, teams can ensure their analysis is comprehensive, their models are well-tuned, and their findings are clearly communicated.

**Evaluation Criteria Table**

Below is the table for the evaluation criteria along with their corresponding percentage points. This table provides a clear breakdown of how each aspect of the project will be assessed.

| **Evaluation Criteria** | **Description** | **Percentage Points** |
| --- | --- | --- |
| **Exploratory Data Analysis (EDA)** | Quality and depth of data exploration, including plots and interpretations. | 25% |
| **Price and Volume Analysis** | Accuracy and insights from the analysis of price and volume curves. | 10% |
| **Time-Based Analysis** | Insights and trends identified from time-based data (monthly, quarterly analysis). | 10% |
| **Regional and Country Analysis** | Analysis of sales data by region and country, including identification of top-performing areas. | 10% |
| **Model Selection and Parameters** | Appropriateness of model selection, parameter tuning, and explanation of chosen parameters. | 15% |
| **Training and Testing Results** | Evaluation of model performance based on training and testing results, including metric analysis. | 15% |
| **Report Quality** | Clarity, conciseness, and structure of the report, focusing on EDA and model results. | 10% |
| **Code Quality and Documentation** | Cleanliness, organization, and documentation of the code for reproducibility. | 5% |

**Detailed Breakdown**

1. **Exploratory Data Analysis (EDA) - 25%**
   * Thoroughness in examining the data.
   * Quality and relevance of plots (e.g., histograms, line charts).
   * Interpretation and actionable insights from the data.
2. **Price and Volume Analysis - 10%**
   * Accuracy in plotting Unit Price and Units Sold.
   * Insights and trends identified from the price and volume curves.
3. **Time-Based Analysis - 10%**
   * Identification of sales trends over time.
   * Analysis of seasonal patterns and growth rates.
4. **Regional and Country Analysis - 10%**
   * Detailed analysis of sales performance by region and country.
   * Identification of top-performing and underperforming regions/countries.
5. **Model Selection and Parameters - 15%**
   * Justification for chosen models.
   * Effectiveness of hyperparameter tuning.
   * Explanation of why specific parameters were selected.
6. **Training and Testing Results - 15%**
   * Methodology for splitting data (e.g., train-test split, cross-validation).
   * Performance metrics (e.g., accuracy, precision, recall, F1 score) and their interpretation.
7. **Report Quality - 10%**
   * Clarity and structure of the report.
   * Conciseness (less than 3 A4 pages).
   * Focus on key findings from EDA and model results.
8. **Code Quality and Documentation - 5%**
   * Clean and organized code.
   * Proper documentation and comments for reproducibility.
   * Ease of understanding and following the analysis process.