**Detailed Description of the US Regional Sales Dataset**

The dataset provides an extensive collection of US regional sales data, capturing a diverse range of details pertaining to sales transactions and customer interactions across different sales channels. With 7992 rows and 15 columns, it offers a comprehensive overview of the sales landscape, encompassing various aspects from order and product details to sales performance metrics. This dataset is valuable for analyzing sales patterns, optimizing inventory management, improving customer relationships, and forecasting future sales.

**Feature Descriptions**

1. OrderNumber: This feature uniquely identifies each sales order with a string value, such as "SO 000101".

2. Sales Channel: This categorical feature indicates the channel through which the sale was made, such as In-Store, Online, Distributor, or Wholesale.

3. WarehouseCode: A string feature representing the warehouse code from where the product was shipped, for example, "WARE-UHY1004".

4. ProcuredDate: This date feature (formatted as DD-MM-YYYY) shows when the product was procured, such as "31-12-2017".

5. OrderDate: A date feature (formatted as DD-MM-YYYY) indicating when the order was placed, like "31-05-2018".

6. ShipDate: This date feature (formatted as DD-MM-YYYY) represents when the product was shipped, for instance, "14-06-2018".

7. DeliveryDate: A date feature (formatted as DD-MM-YYYY) showing when the product was delivered, such as "19-06-2018".

8. CurrencyCode: A string feature indicating the currency in which the transaction was made, like "USD".

9. \_SalesTeamID: An integer feature identifying the sales team responsible for the order, e.g., 6.

10. \_CustomerID: An integer feature representing the customer who placed the order, for instance, 15.

11. \_StoreID: An integer feature identifying the store where the order was placed, such as 259.

12. \_ProductID: An integer feature representing the product sold, e.g., 12.

13. Order Quantity: An integer feature indicating the quantity of the product ordered, like 5.

14. Discount Applied: A float feature representing the discount applied to the order, such as 0.075.

15. Unit Cost: A float feature showing the cost price of one unit of the product, for example, 1,001.18.

16. Unit Price: A float feature representing the selling price of one unit of the product, like 1,963.10.

**Potential Usecases:**

This dataset can be used for various analytical and machine learning tasks, including:

1. Sales Analysis: Understanding sales patterns across different channels and regions.

2. Inventory Management: Optimizing stock levels based on sales trends.

3. Customer Segmentation: Identifying different customer segments based on purchasing behavior.

4. Revenue Forecasting: Predicting future sales and revenue.

5. Discount Effectiveness: Analyzing the impact of discounts on sales.

**Applicable Machine Learning Algorithms**

**Use of Classification Algorithms** (e.g., Logistic regression, random forest, SVM, decision tree, KNN, and ANN):

- Classify high-value vs. low-value orders using features such as unit price and discount applied.

- Predict the likelihood of a product being sold through a particular channel.

**Use of Regression Algorithms** (e.g., Linear regression, random forest, SVM, decision tree, KNN, and ANN):

- predict sales (or revenue) based on order quantity, discount, and unit price.

- forecast delivery times using procurement and shipping dates.

**Use of Unsupervised Learning Algorithms** (e.g., Kmeans clustering):

- To segment customers into clusters based on similar attributes, which can be useful for targeted marketing and personalized services.

**What is expected?**

1. Understand the objective of the project.
2. Preprocessing the data (look for missing values, distribution of the features, transform some of the features), not necessary to follow all the steps.
3. Try to find the correlation between the features and ignore some of them (feature selection) if needed with the techniques that’ll be taught in the class.
4. Split the data into training, validation, and test (might require iterations to get better results).
5. Build at least 3 different machine learning models with the training data and validate which one would be best for testing.
6. Finally, use proper metrics to estimate which model has performed better.