**1. Introduction**

**Objective**

The objective of this analysis is to examine and gain insights from Amazon sales data. This includes understanding sales trends, distribution of quantities across categories, sales performance by country, and distribution of fulfillment methods and sales channels. The ultimate goal is to provide actionable recommendations to improve sales strategies and operational efficiency.

**Dataset Description**

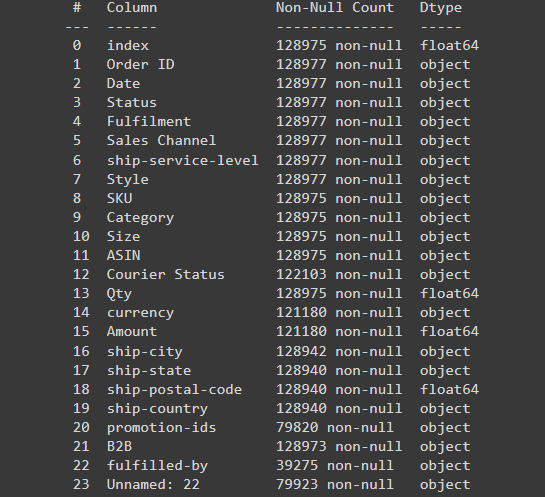
* **Source**: Amazon sales data.
* **Number of Records**: 128,799
* **Columns**:
  + **index**: Unique identifier for each row.
  + **Order ID**: Unique identifier for each order.
  + **Date**: Date when the order was placed.
  + **Status**: Current status of the order (e.g., Shipped, Pending, Returned).
  + **Fulfilment**: Method of fulfillment (e.g., Amazon, Merchant).
  + **Sales Channel**: Channel through which the sale was made (e.g., Online, Offline).
  + **ship-service-level**: Level of shipping service (e.g., Standard, Expedited).
  + **Style**: Style of the product.
  + **SKU**: Stock Keeping Unit, a unique identifier for each product.
  + **Category**: Category of the product (e.g., Electronics, Clothing, Books).
  + **Size**: Size of the product.
  + **ASIN**: Amazon Standard Identification Number, a unique identifier for each product on Amazon.
  + **Courier Status**: Current status of the shipment (e.g., Delivered, In Transit, Out for Delivery).
  + **Qty**: Quantity of products ordered.
  + **currency**: Currency in which the transaction was made (e.g., USD, EUR).
  + **Amount**: Total amount of the transaction.
  + **ship-city**: City to which the order was shipped.
  + **ship-state**: State to which the order was shipped.
  + **ship-postal-code**: Postal code of the shipping address.
  + **ship-country**: Country to which the order was shipped.
  + **B2B**: Indicates if the transaction was Business to Business (B2B).
  + **Month**: Month in which the order was placed.

**2. Data Cleaning**

**Initial State of the Data**

The initial dataset consists of 128,977 records and 24 columns. Below is the initial state of the data, including data types and the number of missing values.

**Data Types and Missing Values**



**Summary Statistics for Numerical Variables:**

A screenshot of a computer screen

Description automatically generated

**Summary Statistics for Categorical Variables:**

A screenshot of a computer

Description automatically generated

**Distribution of Numerical Columns**  
  
A graph with a blue line

Description automatically generated

**Distribution of Categorical Columns**

**A close-up of a graph

Description automatically generated**

**Missing Values in Columns as percentage:**

**A screenshot of a computer

Description automatically generated**

N**umber of Missing Values:**

**A screenshot of a computer

Description automatically generated**

Ater removing the 3 Columns that had a lot of missing values percentage and imputing the numerical and categorical columns.

1. **imputing in numerical using Median:**

A screenshot of a computer

Description automatically generated

1. **imputing Using Mode:**

**A screenshot of a computer

Description automatically generated**

1. **In Categorical Using Mode:**

**A screen shot of a computer program

Description automatically generated**

1. **Result After Cleaning:**

**A screenshot of a computer program

Description automatically generated**

1. **Data Type conversion:**

**A screenshot of a computer program

Description automatically generated**

1. **Treatment of the Outliers:**

**A screenshot of a computer

Description automatically generated**

**3. Data Visualization:**

Creating insightful diagrams using Matplotlib and Seaborn:

* **Like Distribution of Qty and Distribution of Amount**

**A graph of a number of data

Description automatically generated with medium confidence**

* **Count Plot of Status and Count Plot of Category**

**A screenshot of a graph

Description automatically generated**

* **Trend Chart of Amount Over Time**

**A graph showing a line of blue lines

Description automatically generated with medium confidence**

* **Correlation Heatmap between Quantity and Amount**

**A graph with a red square

Description automatically generated**

* **Add Some Visual Analysis like:**
  + **Monthly Sales Trend:**

**A graph with a line going up

Description automatically generated**

* + **Top 10 Selling Categories**

**A graph of sales

Description automatically generated**

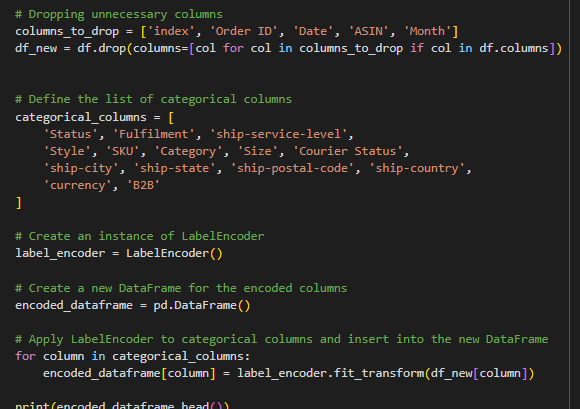
* + **Top Selling Products:**
  + **Sales Distribution by Country**

**A map of the world

Description automatically generated**

**4. Machine Learning Models:**

* **Drop unnecessary columns and encode the categorical columns**

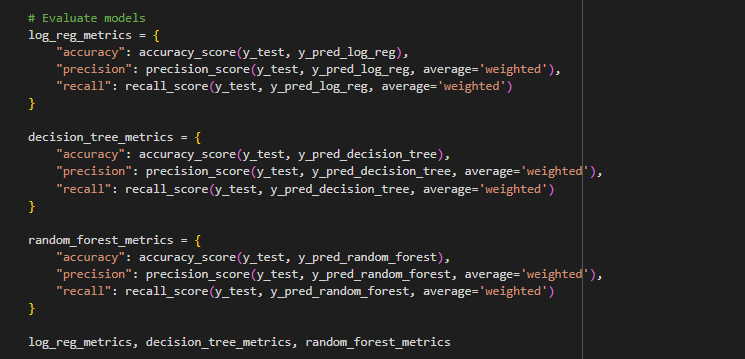
****

* **Create the machine learning models:**

**A screen shot of a computer program

Description automatically generated**

* **Evaluate the Accuracy of each Model:**

****

**A computer screen shot of numbers

Description automatically generated**

* **Cross-Validation for each Model:**

**A computer screen shot of white text

Description automatically generated**

****