# **Predictive Analytics and Optimization in Amazon E-commerce Fulfillment and Sales**

## **ECOMMERCE BUSINESS REQUIREMENT DOCUMENT**

### **Executive Summary**

This project aims to make better use of Amazon’s e-commerce order data to improve how operations are run. By analyzing patterns in how orders are fulfilled, how products are delivered, what customers are buying, and how they behave, we can uncover valuable insights. Using these findings, we plan to build predictive models and practical tools that help reduce delivery problems, increase revenue, and make inventory and logistics planning more effective.

### **Project Objective**

To analyze Amazon order and fulfillment data to:

* Discover trends in sales, fulfillment, and delivery outcomes.
* Predict risks like failed deliveries or returns.
* Optimize decisions related to shipping, inventory, and marketing.

### **Business Requirements**

|  |  |
| --- | --- |
| **ID** | **Requirement Description** |
| BR1 | Identify top-selling product categories and their performance by region. |
| BR2 | Evaluate order delivery performance by fulfillment type and courier status. |
| BR3 | Identify regional trends and customer types. |
| BR4 | Predict whether an order will be delivered, returned, or cancelled. |
| BR5 | Support inventory and logistics planning using predictive analytics. |

### **Functional Requirements**

|  |  |
| --- | --- |
| **ID** | **Functionality** |
| FR1 | Perform data cleaning and transformation for inconsistent/missing values |
| FR2 | Visualize order status breakdowns across regions and categories |
| FR3 | Generate reports on revenue, quantity sold, and delivery delays |
| FR4 | Build ML models to classify/predict order outcomes (delivered/returned/cancelled) |
| FR5 | Enable time-series forecasting for future sales volume |
| FR6 | Segment regions/customers/products for targeted operations |

### **Non-Functional Requirements**

|  |  |
| --- | --- |
| **ID** | **Functionality** |
| NFR1 | System must process at least 10,000 rows efficiently |
| NFR2 | Reports and visualizations must be generated within 5 seconds |
| NFR3 | Data must be stored securely and comply with data protection policies |
| NFR4 | ML models should target a minimum 80% accuracy (baseline) |

### **Success Metrics**

|  |  |  |
| --- | --- | --- |
| **Success Area** | **Criteria** | **How It Will Be Measured** |
| **Data Quality** | Data is cleaned, consistent, and usable for analysis | Minimal missing values, categorical columns are standardized |
| **Exploratory Analysis** | Key trends in sales, returns, fulfillment, and regions are discovered | Clear visualizations and summaries showing top products, regions, delays, etc. |
| **Model Accuracy** | Models predict delivery success or return/cancellation effectively | At least **80% accuracy** in classification models |
| **Sales Forecasting** | Sales predictions are reliable for top product categories | Error metrics like **MAE** or **RMSE** are within acceptable thresholds |
| **Insights Generation** | Actionable business insights are derived (e.g., best fulfillment method) | Recommendations supported by data with clear justification |

### **Assumptions**

1. **Data Completeness**

The dataset contains all necessary fields to perform order fulfillment, sales, and shipping analysis — including product, region, and delivery status attributes.

1. **Data Representativeness**

The historical data reflects actual business operations across multiple fulfillment types and regions and is not artificially sampled or test data.

1. **Consistent Labeling (with Cleaning Required)**

Categorical values such as Status, Courier Status, and fulfilled-by may require standardization (e.g., fixing typos, missing values), but their categories are generally interpretable.

1. **Missing/Null Handling is Feasible**

Some columns (e.g., promotion-ids, Unnamed: 22) may contain missing or irrelevant values and can be dropped or imputed without affecting model performance.

### **Scope**

 Exploratory analysis of sales, fulfillment, and delivery data

 Detection of patterns in cancellations, returns, and delays

 Classification of delivery outcomes based on product and shipping features

 Sales forecasting by category, region, or time

 Feature engineering for SKU, fulfillment channel, courier service, and location

### **Risks and Mitigations**

|  |  |  |
| --- | --- | --- |
| **Risk** | **Description** | **Mitigation Strategy** |
| Data Incompleteness | Some records may have missing or inconsistent values | Clean the data using imputation or by removing incomplete rows |
| Inconsistent Labeling | Variations in naming conventions (e.g., "standard" vs "Standard") can cause confusion | Standardize and normalize values during preprocessing |
| Class Imbalance | Fewer records for returns or cancellations may reduce model performance | Use resampling techniques (oversampling/under sampling) to balance the dataset |
| Spurious Correlations | Some patterns may seem important but are not actually meaningful | Validate patterns using cross-validation and model explainability tools |

### **Glossary**

**Dataset Name: Amazon\_Order\_Data.csv**

https://www.kaggle.com/code/mehakiftikhar/amazon-sales-dataset-eda

| **Column Name** | **Description** |
| --- | --- |
| **index** | System-generated index column (may not be needed for analysis) |
| **Order ID** | Unique identifier for each customer order placed on the platform |
| **Date** | Date the order was placed (format: MM/DD/YY or similar) |
| **Status** | Final order outcome (e.g., "Shipped", "Cancelled", "Delivered to Buyer") |
| **Fulfilment** | Indicates whether the order was fulfilled by Amazon or by the Merchant |
| **Sales Channel** | Platform or channel where the sale occurred (e.g., "Amazon.in") |
| **ship-service-level** | Shipping speed/service selected (e.g., "Standard", "Expedited") |
| **Style** | Style or product model name (can be used for grouping) |
| **SKU** | Stock Keeping Unit – unique product identifier used by the seller |
| **Category** | Product category (e.g., "Kurta", "Top", "Western Dress") |
| **Size** | Size of the product (e.g., "L", "XXXL", numeric sizes) |
| **ASIN** | Amazon Standard Identification Number – unique identifier across Amazon's catalog |
| **Courier Status** | Current delivery stage (e.g., "Shipped", "Out for Delivery", "Returned to Seller") |
| **Qty** | Quantity of units ordered in the transaction |
| **currency** | Currency of the transaction (usually "INR" for India) |
| **Amount** | Total order amount (revenue generated) |
| **ship-city** | Destination city for shipping |
| **ship-state** | Destination state for shipping |
| **ship-postal-code** | Postal code/ZIP of the delivery address |
| **ship-country** | Destination country (likely "IN" for India) |
| **promotion-ids** | Promotion or discount codes applied to the order (if any) |
| **B2B** | Indicates whether the order is Business-to-Business (True/False) |
| **fulfilled-by** | Shipping/handling method (e.g., "Easy Ship", Fulfilled by Amazon) |
| **Unnamed: 22** | Likely an empty or unused column (can be dropped) |