

# Low-Level Design Document (LLD) for Sales Management Project

- KAINAT NAQVI

## 1. Introduction

This Low-Level Design Document (LLD) provides a detailed breakdown of the technical implementation, data flow, and core functionalities of the Sales Management project, which aims to analyze sales trends and key performance metrics using an Amazon sales dataset. The data is ingested from a CSV file, processed using Python libraries, and analyzed for insights.

## 2. Data Ingestion

- **Data Source:** CSV file sourced from PwSkills.
- **Platform:** The dataset is loaded and processed in Google Colab.
- **File Details:**
  - **File Name:** Amazon Sales Records.csv
  - **Sheet/Tab:** Single sheet.
  - **Format:** Comma-Separated Values (CSV).
  - **Data Size:** One dataset containing several columns related to sales records such as Region, Country, Item Type, Sales Channel, Total Revenue, etc.

## 3. Data Preprocessing

- **Date Parsing:** Custom date parsing was implemented to convert mixed-format date columns (**Order Date**, **Ship Date**) to a standardized **datetime** format. This allows the extraction of month and year for time-series analysis.
  - Extracted **Year** and **Month** for further grouping and analysis.
- **Missing Values Handling:** The dataset was checked for missing values, with no null values found. No imputation was necessary.
- **Feature Engineering:**
  - **New Columns Created:** **Year**, **Month**, **Profit Margin (%)**.
  - **Date Parsing:**
    - **Order Date & Ship Date:** Parsed from strings to **datetime** format using Python's **pandas** library. This ensures consistency in date representation, facilitating time-series analysis.
  - **Extracting New Time Features:**
    - **Year:** Extracted from **Order Date** to analyze yearly trends.
    - **Month:** Extracted from **Order Date** to explore seasonal and monthly patterns.

- **Year-Month:** Combined the year and month fields to create a **Year-Month** feature for more granular trend analysis.
- **Profit Margin Calculation:**
  - **Profit Margin %:** Created a new column using the formula:  $\text{Profit Margin \%} = \left( \frac{\text{Total Profit}}{\text{Total Revenue}} \right) \times 100$
- This helps in evaluating profitability across various dimensions like product categories and regions.
- **Categorical Grouping:**
  - **Sales Channel:** Categorized sales as either **Online** or **Offline** to facilitate comparison between these two channels.
  - **Order Priority:** Grouped into **High**, **Medium**, **Low**, and **Critical** to examine how order urgency impacts revenue and profit.
- **Revenue per Unit Sold:**
  - Calculated as:  $\text{Revenue per Unit Sold} = \frac{\text{Total Revenue}}{\text{Units Sold}}$
- This feature reveals the efficiency and pricing strategies for each product category.

This feature engineering allows deeper insights into sales trends and profitability factors.

#### 4. Data Transformation

- **Groupings for Key Metrics:**
  - **Yearly Sales Metrics:** Total Revenue, Total Profit, Profit Margin, and Units Sold were grouped by **Year**.
  - **Monthly Sales Metrics:** Grouped by **Month**.
  - **Yearly-Monthly Sales Metrics:** Grouped by both **Year** and **Month** for a granular time-series analysis.

#### 5. Exploratory Data Analysis (EDA)

The following analyses were performed:

- **Sales Trends:**
  - **Yearly:** Revenue and profit trends were visualized by year.
  - **Monthly:** Grouped by month to identify seasonal patterns.
  - **Yearly-Monthly:** A detailed breakdown of sales trends across year and month combinations.
- **Key Metrics Analysis:**

- **Total Sales:** Aggregated Total Revenue.
- **Average Order Value:** Calculated as **Total Revenue** divided by the number of unique orders.
- **Revenue per Unit Sold:** **Total Revenue** / **Units Sold**.
- **Category Performance:**
  - **Item Type:** Total Revenue, Total Profit, and Units Sold were grouped by **Item Type** to identify high-performing product categories.
- **Geographical Performance:**
  - Sales and profitability analysis was grouped by **Region** and **Country**.
- **Sales Channel Performance:**
  - Analyzed and compared performance between **Online** and **Offline** sales channels.
  - Hypothesis testing was performed to statistically compare sales channel profitability.
- **Order Priority Impact:**
  - Total Revenue, Total Profit, and Units Sold were grouped and analyzed by **Order Priority** (High, Medium, Low, Critical).

## 6. Visualizations

Several visualizations were created to understand trends and patterns:

- **Time Series Analysis:** Line charts displaying **Total Revenue** trends by year, month, and year-month combinations.
- **Heatmaps:** Heatmap for profit and revenue analysis across regions, sales channels, and product categories.
- **Bar Charts:** Category performance, geographical performance, and sales channel comparison were visualized using bar charts.
- **Box Plots:** Profit margin distribution across product categories, regions, and sales channels.

## 7. Statistical Analysis

- **Hypothesis Testing:**
  - **Sales Channel Comparison:** An independent t-test was performed to check for significant differences in profit between **Online** and **Offline** sales.
  - **Item Type Comparison:** ANOVA test was performed to identify significant differences in profit across different **Item Types**.

## 8. Technology Stack

- **Programming Language:** Python
- **Libraries Used:**
  - **Data Processing:** `pandas`, `numpy`
  - **Date Parsing:** `dateutil`
  - **Visualization:** `matplotlib`, `seaborn`
  - **Statistical Analysis:** `scipy.stats`
- **Platform:** Google Colab

## 9. Outputs and Findings

- **Trends Identified:** Significant trends in sales and profitability were observed, with certain months and years outperforming others.
- **Top Performing Regions and Categories:** Some regions, like `Sub-Saharan Africa`, showed higher profitability. Product categories like `Cosmetics` were highly profitable.
- **Sales Channel Insights:** There was no significant difference in profit between `Online` and `Offline` channels.
- **Order Priority Impact:** High priority orders were associated with higher revenue.

## 10. Conclusion

The LLD outlines the detailed technical steps taken to implement the Sales Management project, from data ingestion, preprocessing, and transformation to performing exploratory data analysis and statistical testing. The system is designed using Python in a modular and structured way to ensure scalability and ease of future modifications.