A MINI PROJECT REPORT ON

**Analysing Amazon Inventory at Data Warehouse**

Submitted in partial fulfilment of the course

CSE-1006:Foundation of Data Analytics

Under the guidance of prof. Deepashikha Mishra

By

22BCE7831 SUDARSHAN VERMA

22BCE8793 G.DHATRI NARESH



VIT-AP UNIVERSITY

AMARAVATI

ANDHRAPRADESH,

INDIA

### **Abstract**

This data analysis project focuses on Amazon inventory at a data warehouse. The dataset includes 1300 columns and 50 products with IDs ranging from 1001 to 1050 and 20 warehouses ranging from W001 to W020. The columns include product ID, warehouse ID, invoice date, and sales count. The goal of this project is to analyse the sales data and identify trends and patterns in order to optimize inventory management and increase profits and prevents product production crisis in future by predicting sales count in future by using machine learning model.

## Introduction

Welcome to our data analysis project on Amazon inventory at a data warehouse. Our goal is to analyse the sales count data for 50 products with product IDs ranging from 1001 to 1050 and 20 warehouses ranging from W001 to W020. The dataset includes columns for product ID, warehouse ID, invoice date, and sales count.

The purpose of this project is to provide insights to the manufacturing industries on the sales patterns and trends of the products in the warehouses over a certain period of time. By analysing the data, we hope to identify any areas where improvements can be made in terms of inventory management and sales strategies.

## Problem Statement and Objectives

### Problem Statement

The problem is to analyse this dataset to identify trends and patterns in the sales count of products across different warehouses and over time. And to predict the sales count of a particular product in future by using machine learning model.

### Objectives

* Analyse the sales count of products across different warehouses and over time.
* Identify trends and patterns in the sales data.
* Provide insights and recommendations for improving inventory management and sales strategies.
* Predicting the sales of a particular product in future by using machine learning model

Working with dataset

* Data is collected from amazons website ,we have collected subset of data from the website which is necessary for our project
* Missing values in dataset are filled by the columns mean, median, mode .
* The modified dataset is stored in excel sheet
* In this project we have imported packages dplyr,readxl.ggplot2,tidyverse to perform smoother operations on data set.

Extracting data

* First five columns are extracted into the data
* Further subsets are extracted for sales analysis based on warehouse-id, invoice date, year etc.
* This involves removing any duplicate or irrelevant data, correcting any errors, and filling in any missing values

Data sorting

* Data is sorted by product-id by default .
* Data is grouped and filtered and then sorted for analysis Sales-Count of a particular product, warehouse wise, Year wise, moth wise analysis etc.
* Sorting the data also helps to identify any trends or patterns that may be present in the data.

PREDICTION OF ANALUSIS USING MACHINE LEARNING TECHNIQUE

### **Overview of Prediction Analysis with Machine Learning**

* Prediction analysis is a process of using machine learning algorithms to make predictions based on historical data. In the context of our Amazon inventory project, we can use prediction analysis to forecast future sales of products in different warehouses. This can help us optimize inventory management and ensure that we have enough stock on hand to meet demand.

### **Steps in Prediction Analysis with Machine Learning**

* **Data Preparation**: The first step is to prepare the data for analysis by cleaning, transforming, and normalizing it. This involves removing any missing or irrelevant data and ensuring that the data is in a format that can be used by machine learning algorithms.
* **Feature Selection**: The next step is to select the most relevant features from the dataset that can contribute to accurate predictions. This involves identifying the key variables that are likely to influence sales, such as product ID, warehouse ID, and invoice date.
* **Model Selection**: Once the features have been selected, the next step is to choose a machine learning algorithm that is appropriate for the data and the problem at hand. Common algorithms for prediction analysis include linear regression, decision trees, and neural networks.
* **Training and Testing**: The selected model is trained on a subset of the data and tested on another subset to evaluate its accuracy. This involves splitting the dataset into training and testing sets and using cross-validation techniques to ensure that the model is robust and not overfitting the data.
* **Prediction and Evaluatio**n: Once the model has been trained and tested, it can be used to make predictions on new data. The accuracy of the predictions can be evaluated by comparing them to actual sales data and calculating metrics such as root mean squared error (RMSE) and mean absolute error (MAE).

**Results:**

Our program will ask for product id first and after that user have to enter a number for whatever analysis user need out of inventory management.

It will ask to enter the numbers to show the following outputs:

press 1 for over all products sales analysis

press 2 for over all products sales report

press 3 for for your product sales report

press 4 for your product sales report year wise

press 5 for your product sales analysis year wise

press 6 for your product month wise sales report

press 7 for your product month wise sales analysis

press 8 for sales report of all products warehouse wise

press 9 for sales analysis of all products warehouse wise

press 10 for your product sales report warehouse wise

press 11 for your product sales analysis warehouse wise

press 12 for your product sales report from a particular warehouse id.

press 13 to know your product sales by invoice date.

press 14 to know your product sales report by invoice date followed by warehouse

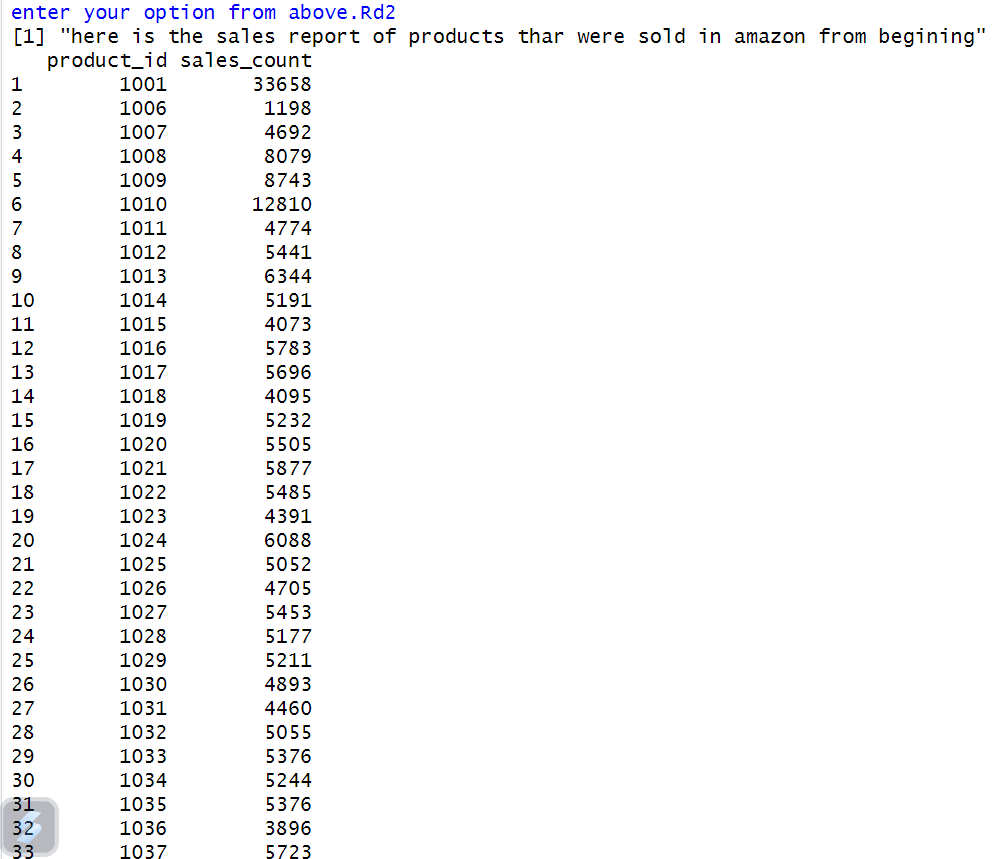
press 15 to know your product sales in future by entering date

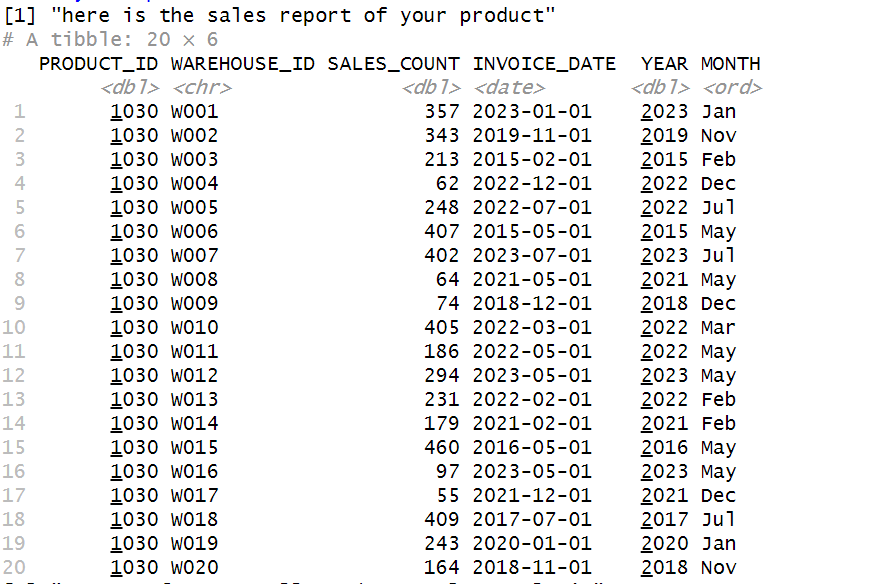
press 16 to know your product sales in future by entering year

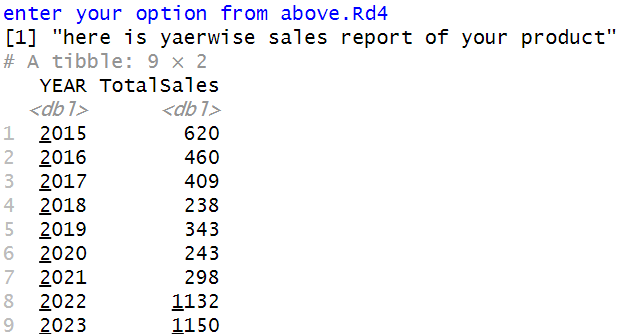
press 17 to know your product sales in future in a specific year in a specific month

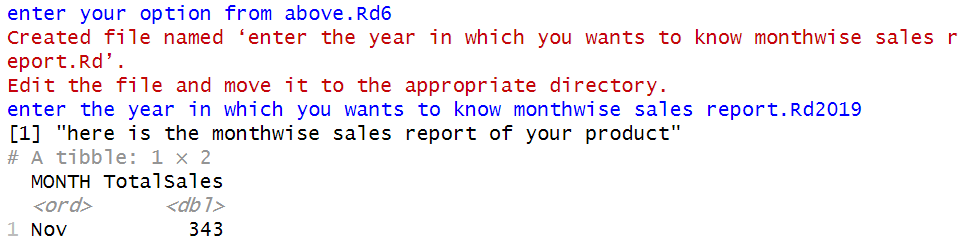
press 0 to exit

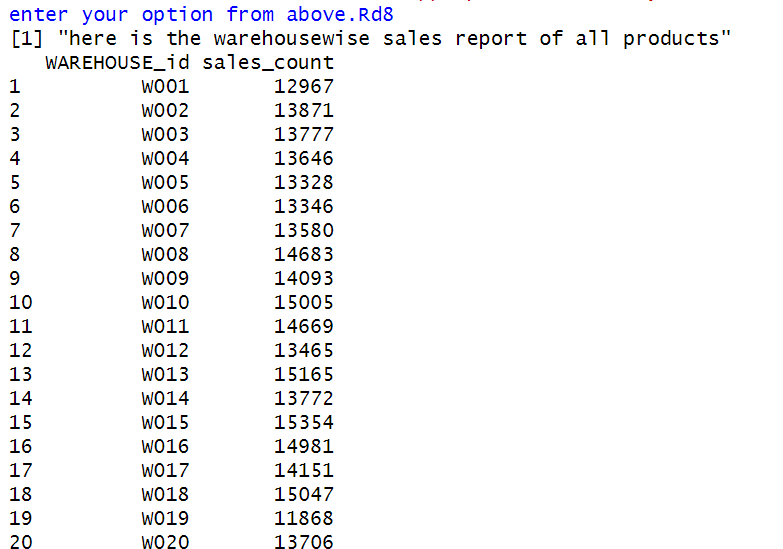
**Sample outputs:**

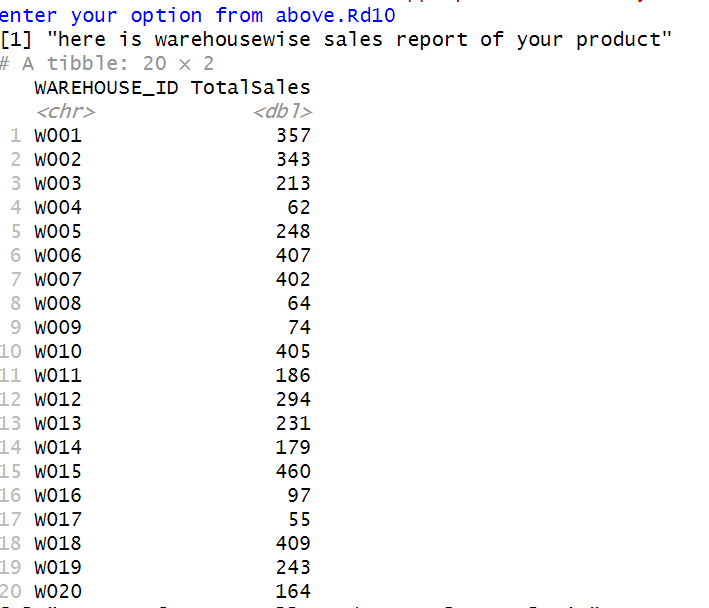
****

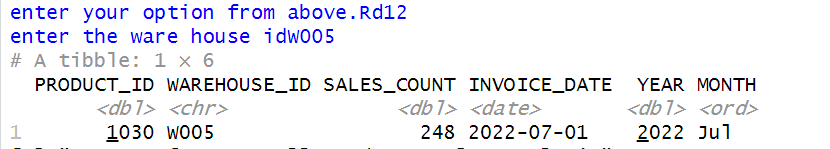


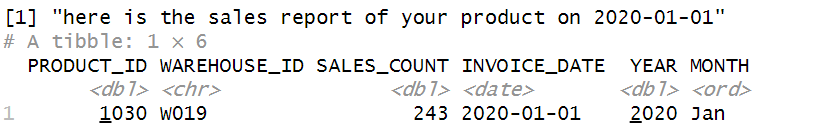


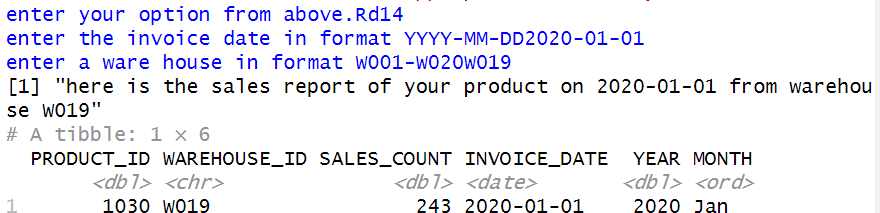


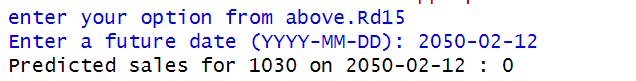


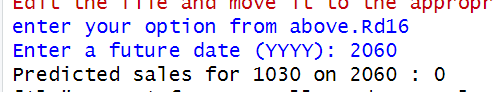


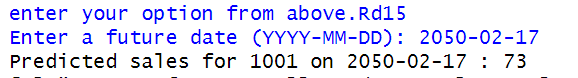


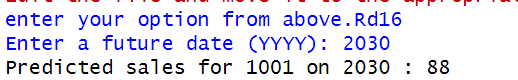


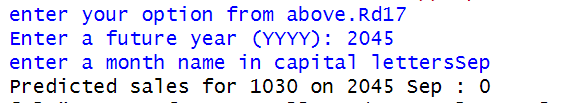






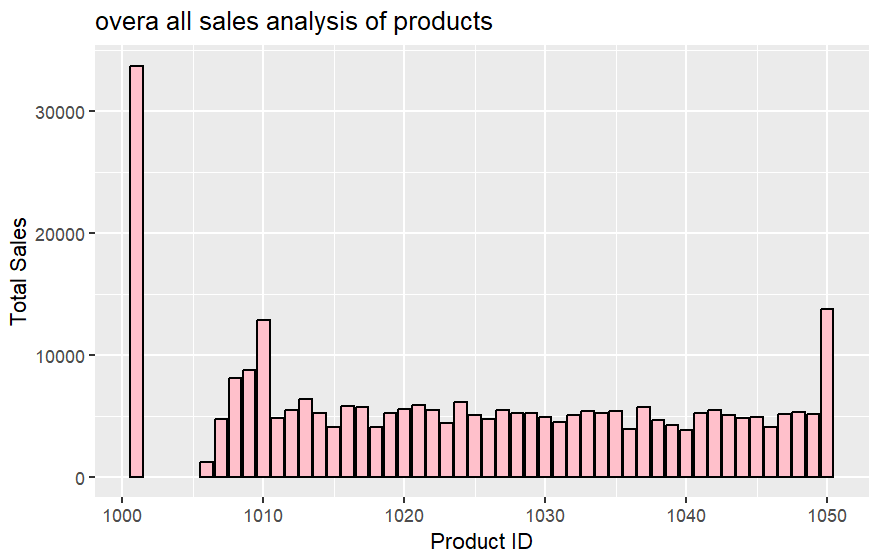


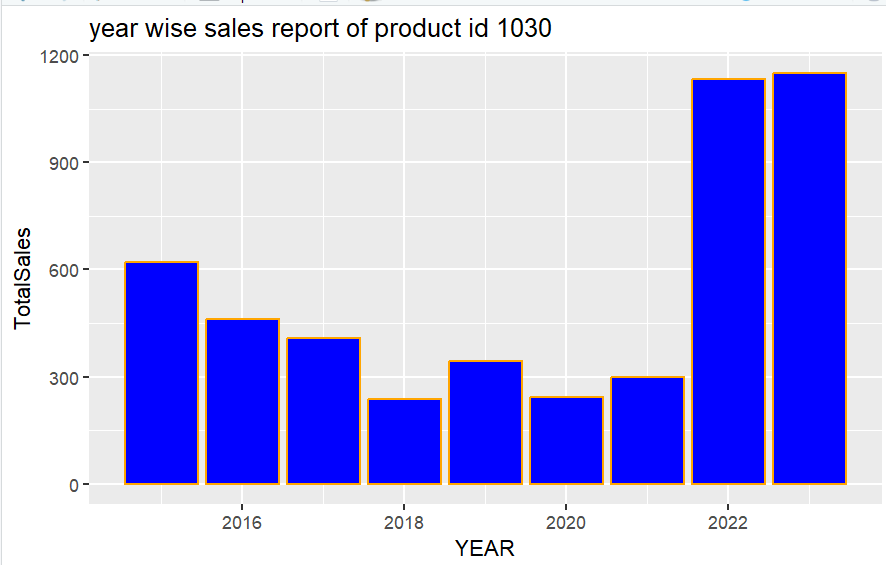


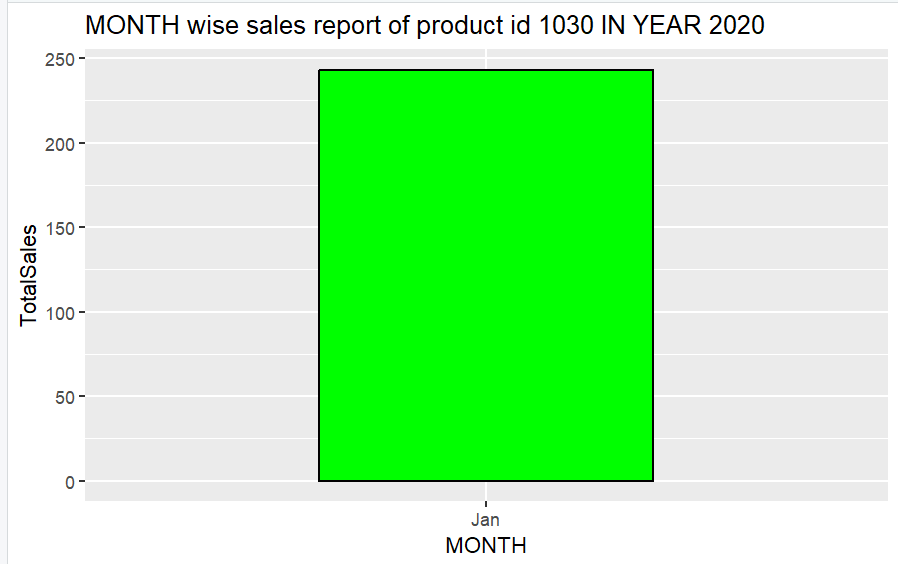


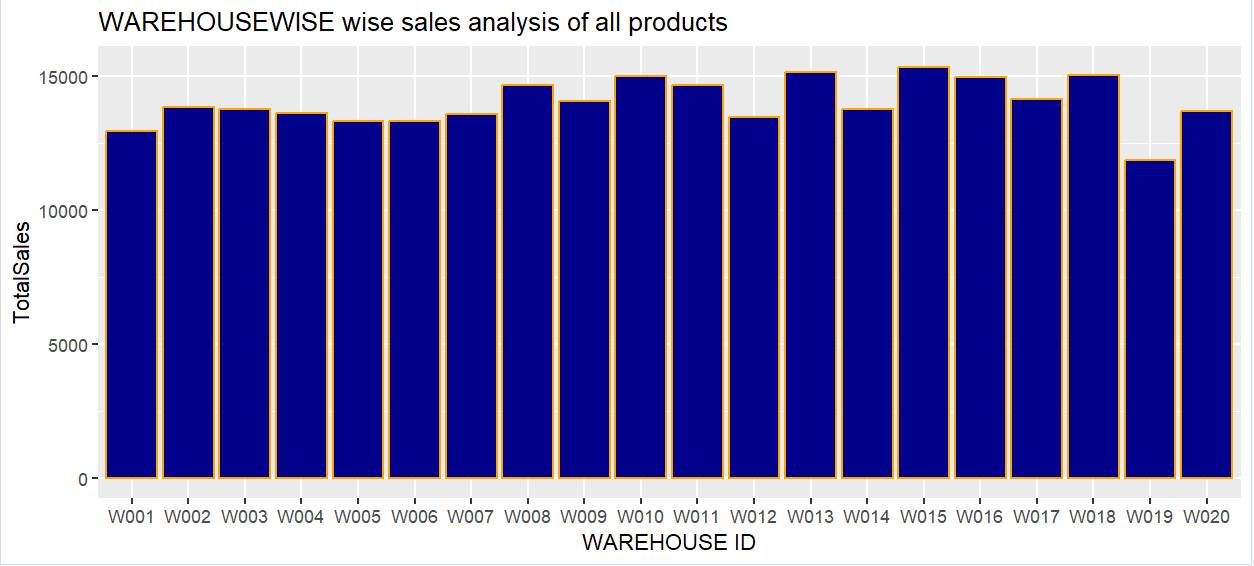
**PLOT :**

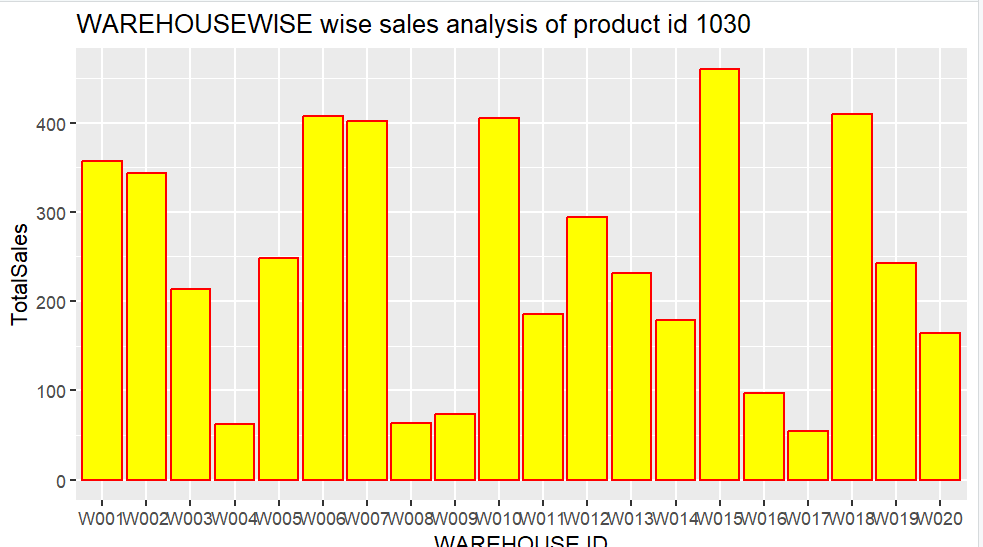
**Based on the information fetched from data set, program will prepare graphical analysis of the data, to make it easier to understand for user.**











## Conclusion

In conclusion, the data analysis project on Amazon inventory at our data warehouse has provided valuable insights into the sales patterns of products across warehouses. Through the analysis of a dataset with 1300 columns, we were able to identify trends in sales count and invoice dates, as well as variations in product popularity across different warehouses.

These findings can inform future inventory management strategies and help optimize product distribution across our warehouses. We are confident that the insights gained from this project will lead to improved efficiency and profitability for the company.