Team Members:

Bhavyanshu Rajendra Kadu, bhavyans@buffalo.edu
Shubham Shankar Sathe, ssathe@buffalo.edu
Rahul Haribhau Shelke, rahulhar@buffalo.edu

Proposed Project Title:

Inventory Management using Binary Classifiers in Python

Data Sources:

Historical Sales and Active Inventory Dataset: This data source will contain historical sales data for the retail firm's inventory. The data will include information such as product ID, release year, price reg, and quantity sold. This data will enable us to identify which products are not selling well and should be considered for removal from the inventory.

The dataset is of the firm that has a large inventory, but only a small percentage of products sell each year, with many products only having a single sale annually. The sales and growth team wants to use historical sales data and active inventory to build a binary classifier that will identify which products should be retained for sale and which products should be removed.

Data Source Link: https://www.kaggle.com/datasets/flenderson/sales-analysis
API Link: The dataset does not have an API link by itself so we will use a Kaggle package to create an API for our account using the Kaggle package in Python and we will dynamically load the dataset into our code.

Analysis Plan:

1) As part of the analysis strategy, a binary classifier model will be created that will categorize the inventory's items as either "to be retained" or "to be removed." The retail company's past sales data and current inventories will be used to conduct the analysis.
2) To ensure that the data are accurate and in the right format, data pretreatment will be the first step in the analysis plan. In this step, duplicates and missing data will be eliminated, categorical data will be transformed, and the data will be scaled.
3) The data will then be divided into training and testing sets. The binary classifier model will be trained using the training set, and its performance will be assessed using the testing set.We will explore and select different classification algorithms that are best suited for this type of data. We may consider algorithms such as logistic regression, decision trees, and Random forest classifiers. The algorithm will be selected based on its accuracy, R2 score, and any other methods necessary for calculating error.
4) After choosing the algorithm, we will use the practice data set to train the binary classifier model. Then, we will assess the model's performance using the testing data set to gauge its precision, R2 score, and F1 score.

5) As a last step, we will utilize the trained binary classifier model to assign the items in the inventory to one of two categories: "to be retained" or "to be removed." A list of the product IDs that must be kept and a different list of the product IDs that must be deleted will be provided.

Our team will be able to undertake this research thanks to the source data we've chosen because it includes both active inventory and historical sales information. The historical sales data will enable us to pinpoint the products that aren't doing well, and the active inventory data will notify us of the items that are now in stock. These two data sources can be used to create a binary classifier model that can correctly categorize the inventory's items as "to be retained" or "to be removed."

Motivation:

The primary goal of this research is to help a retail company's inventory be as efficient as possible. The company can cut inventory holding costs, boost sales and profitability, and make better use of its resources by deciding which products should be kept and which ones should be dropped.

Another intriguing aspect of this investigation is the development of a binary classifier model. The analysis offers a great chance to apply various classification algorithms and assess their effectiveness because binary classification is a fundamental machine-learning problem.

The analysis may also offer insightful information about consumer behavior and preferences. The company can better understand what customers want by determining which products aren't selling well, and they can then change their inventory.

In conclusion, the analysis is crucial for any retail company that wishes to stay effective, lucrative, and competitive. It offers a chance to use machine learning methods on actual data, optimize inventories, and learn more about client behavior.