Text, whiteboard

Description automatically generated

Production – Applicability of Analytics

Business Analytics

**Course Instructor**: Prof. Arun Phadke

**Course Name**: Business Analytics

**Group Details: -**

|  |  |
| --- | --- |
| J V V Sriram Sasank | D010 |
| Abhijay Singh | D019 |
| Ritik Chopra | D033 |
| Ashwin Mahajan | D044 |
| Tanvi Kaur Bhatia | D045 |
| Sasank Hota | D060 |

Table of Contents

[Introduction: 3](#_Toc120135884)

[Pre-Existing problems: 3](#_Toc120135885)

[Problem Statement: 3](#_Toc120135886)

[Data Set Information: 4](#_Toc120135887)

[Data-Set Description: 4](#_Toc120135888)

[Pre-Processing and Cleaning of data: 5](#_Toc120135889)

[ABC Analysis: 5](#_Toc120135890)

[Outlier Treatment: 7](#_Toc120135891)

[Predictive Modelling: 8](#_Toc120135892)

[Confusion Matrix: 10](#_Toc120135893)

[ROC Curve: 12](#_Toc120135894)

[Managerial Contributions and Conclusion: 13](#_Toc120135895)

[Limitations: 13](#_Toc120135896)

[References: 14](#_Toc120135897)

# Introduction:

Every manufacturing industry frequently faces high uncertainty in production planning and supply chain management. With live product tracking and data collection across sectors, companies have enormous opportunities to adapt to a data-driven approach. Companies are adopting a data-driven approach to reduce uncertainty and minimise losses. Companies are desperate for an efficient machine learning model that can comprehend interconnected trends and predict future trends with greater accuracy.

# Pre-Existing problems:

The finished products of the manufacturing industry frequently have multiple raw materials involved in the production process. Each item must be meticulously tracked and accounted for in both the inventory and the manufacturing process. Improper handling reduces the efficiency of the final product and raises costs. They must be stocked in the inventory in accordance with their usage and rate of use. As a result, industries have a constant problem with efficient inventory management.

# Problem Statement:

In this study we aim to do descriptive analysis by creating a dashboard for analysing the

historical data. So that non-developer can understand the inventory data in a better way.

We also aim to solve multiple-item inventory management problem by doing ABC analysis

which clusters the inventory in to most used to less used. We tend to implement predictive

analysis using Random Forest Classifier and identify the items that will be sold and will not

be sold. So that they can be removed /stocked according to the prediction.

# Data Set Information:

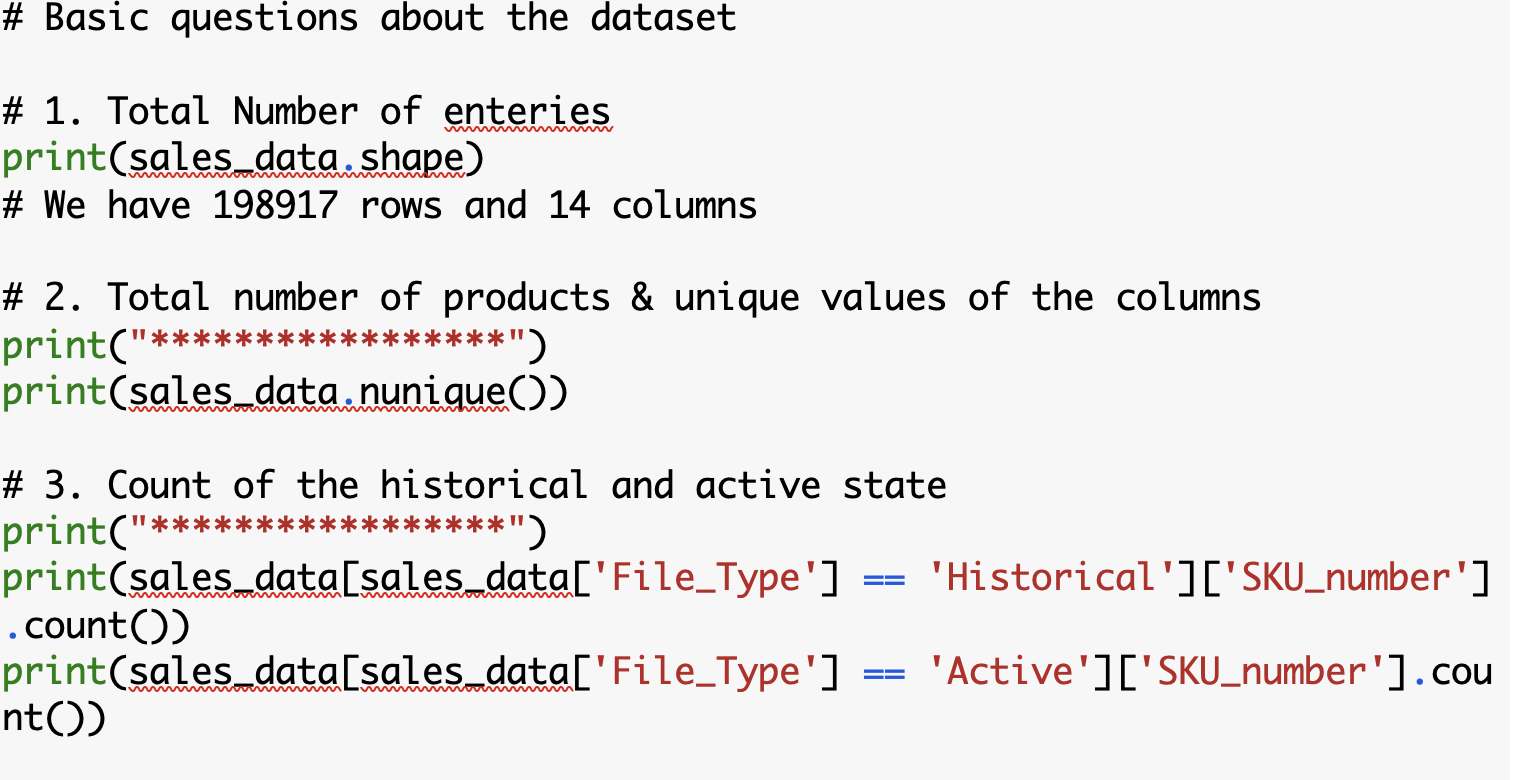
Dataset is open-source data and downloaded from [www.kaggle.com](http://www.kaggle.com) .

|  |  |
| --- | --- |
| Order | Order Id for Indexing for the order data. |
| File\_Type | Historical Sales data and Active Inventory |
| SKU\_number | This is the unique identification ID for each product |
| SoldFlag | 1= sold in last 6 months  0 = Not Sold |
| SoldCount | Categorical attributes of the sales. |
| MarketingType | Categorizes between two kind of marketing campaigns |
| ReleaseNumber | Attribute of Sales |
| New\_Release\_Flag | Flag for identifying future release |
| StrengthFactor | Strength of the product |
| PriceReg | Price of the product |

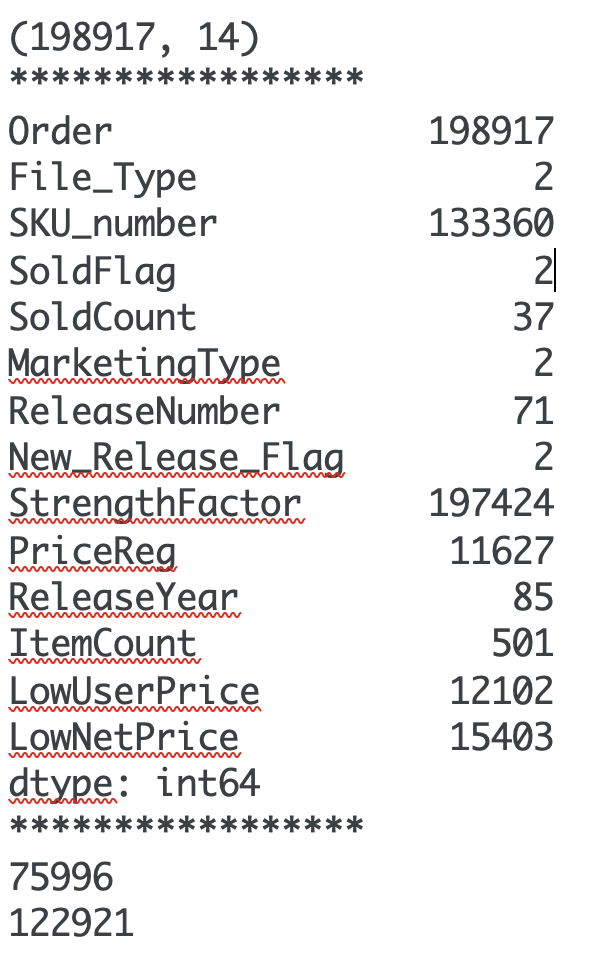
# Data-Set Description:

Below code snippet is used to analyse the data in detail like number of entries , Total number of products & unique values of the columns, Count of historical and active state

CODESNIPPET:

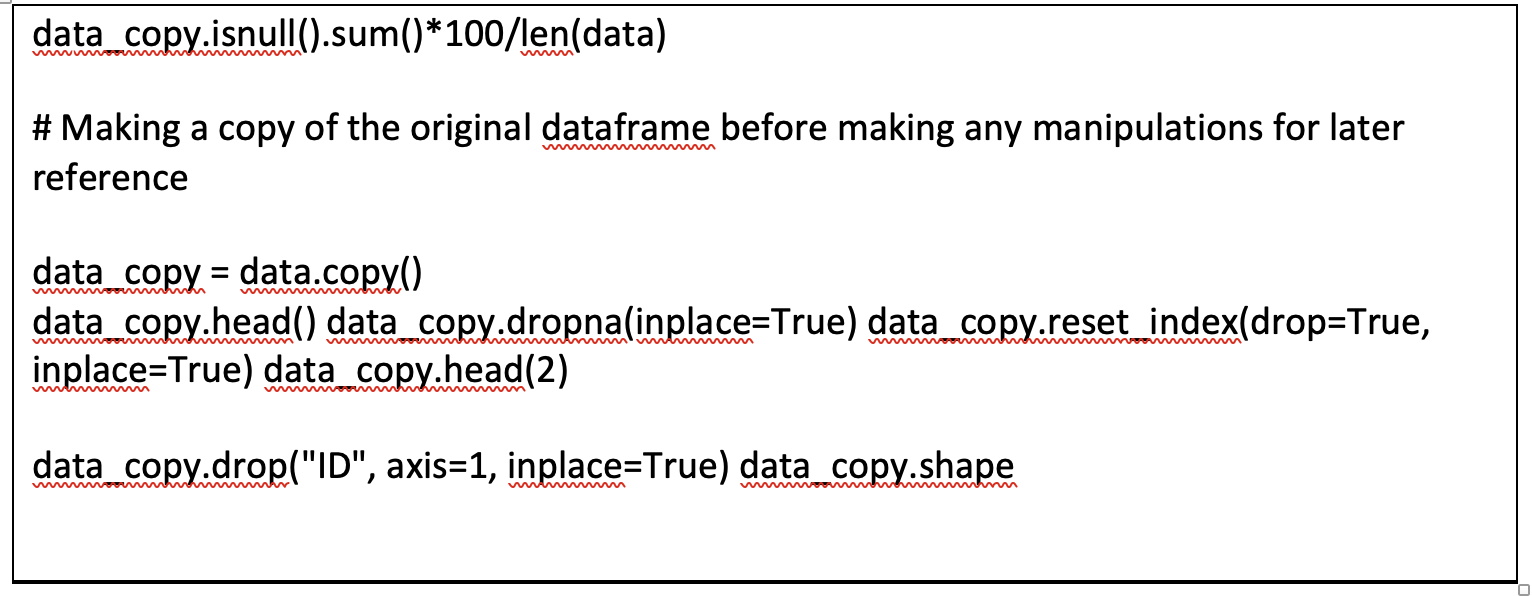


**OUTPUT:**



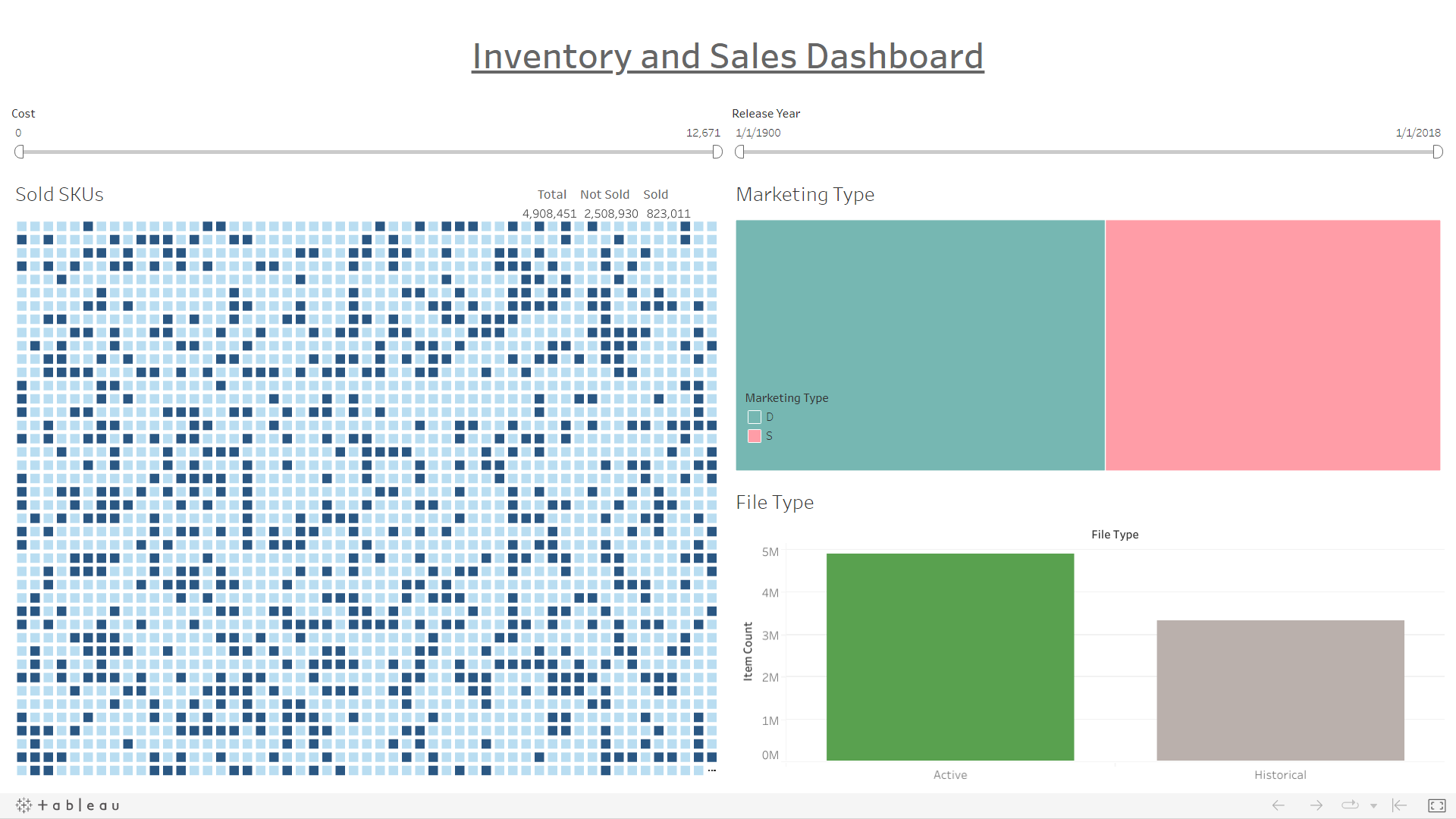
# 

# Pre-Processing and Cleaning of data:



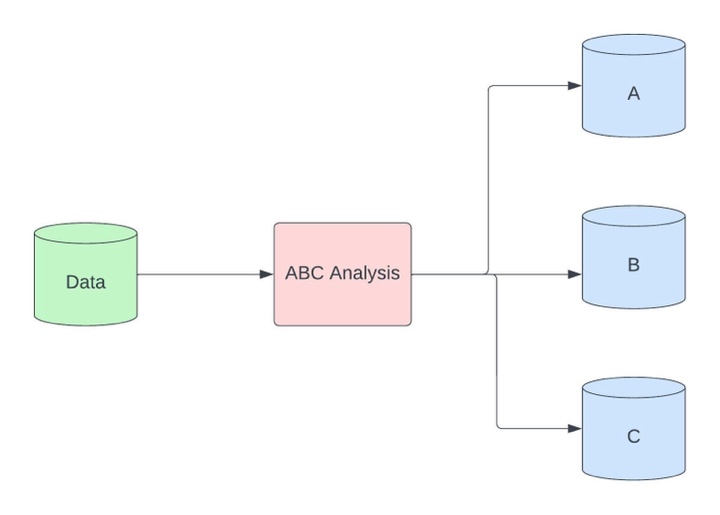
# Descriptive Analysis and Live Dashboard:

As stated in the problem statement, for this section we have utilized the data visualization tool – Tableau to create a live dashboard. The dashboard consists of three main sections – the SKU units sold, the marketing type used and the file type of the SKUs. To help analyse a situation there are two filters one on cost and one on the year to filter out the kind of SKU’s needed to analyse. The Dashboard gives an insight as to unit sold, which marketing type they are and which file type they are. There is a link attached to the references for the live dashboard.



# 

# ABC Analysis:



Every organization has inventory filled different kind of raw materials and products. These products are stocked in the inventory according to their usage. So, these products need to be managed separately. ABC analysis helps to categorise the products according their usage so that inventory can be stocked according to the usage. Products are categorised in A,B,C.

Category A:

This category has 15-20% of the quantity of products and they represent 80% of the value.

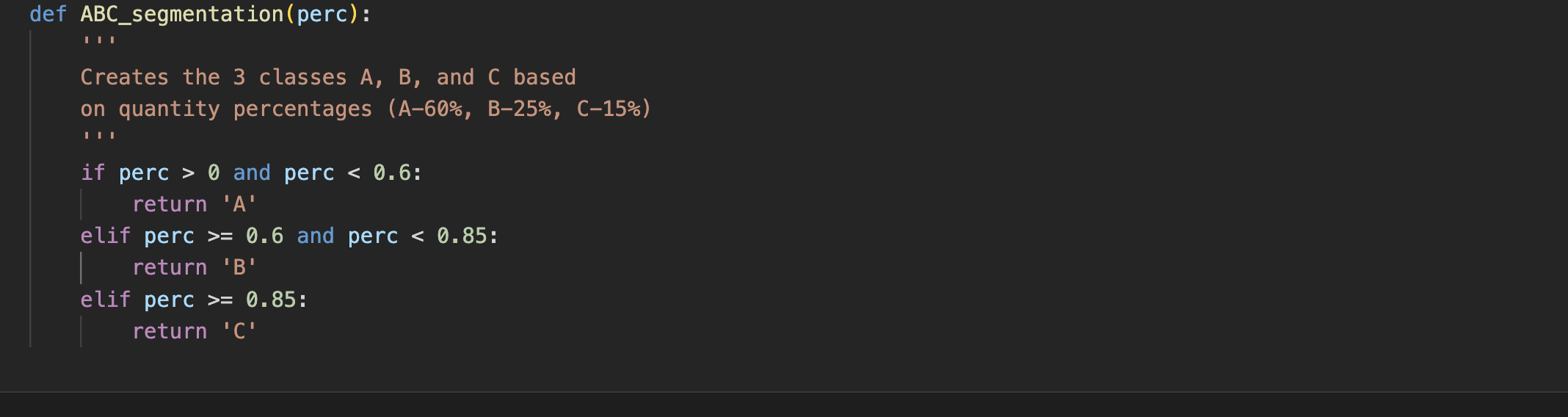
So, High valued/consumed products comes under this category.

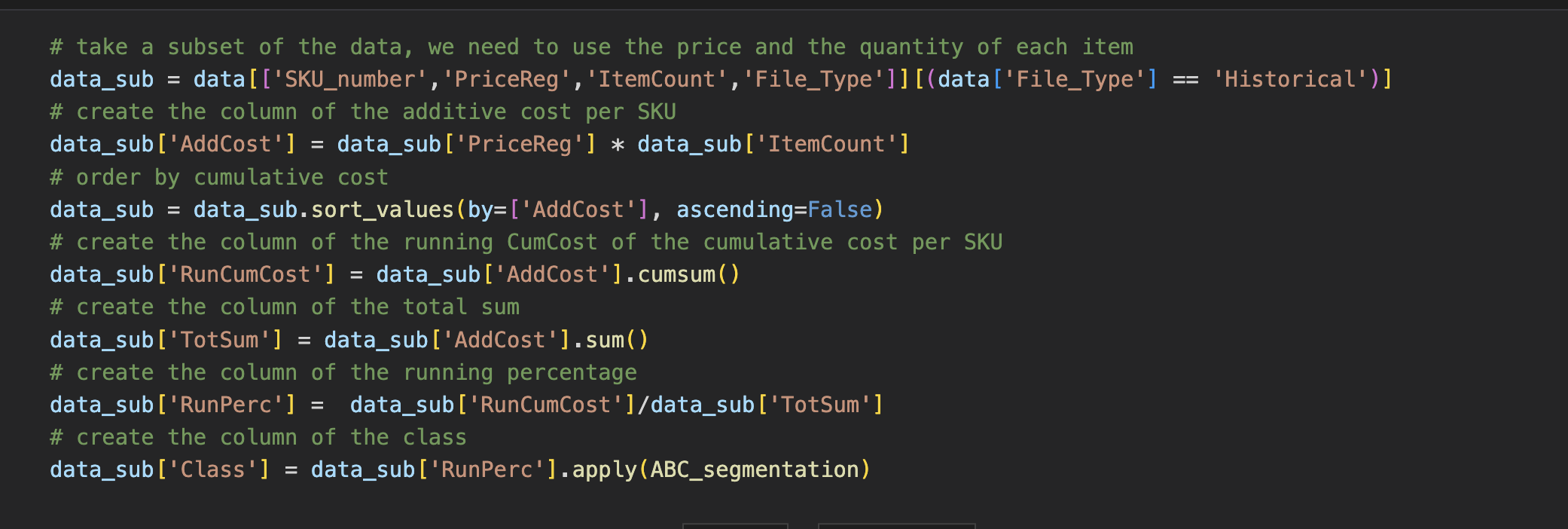
Category B:

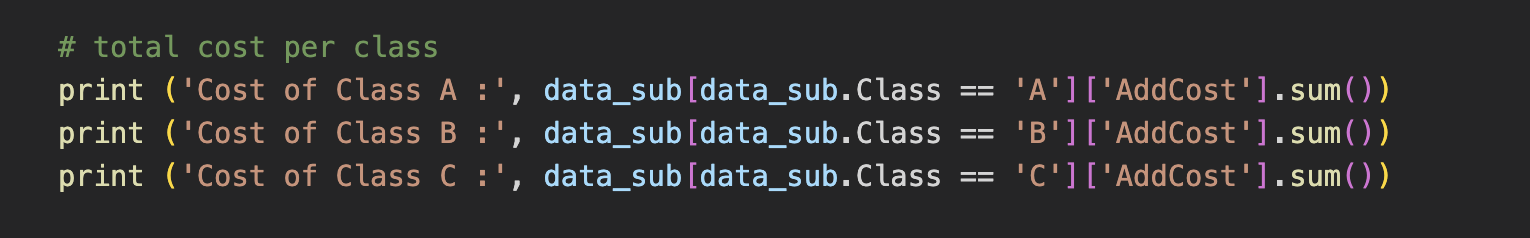
This category has 30-35% of the quantity of products and they represent nearly 15% of the value. Medium valued/ consumption products comes under this category.

Category C:

This category has 50% of the quantity of products and they represent 5% of the value. Low valued/ Consumed products comes under this category.







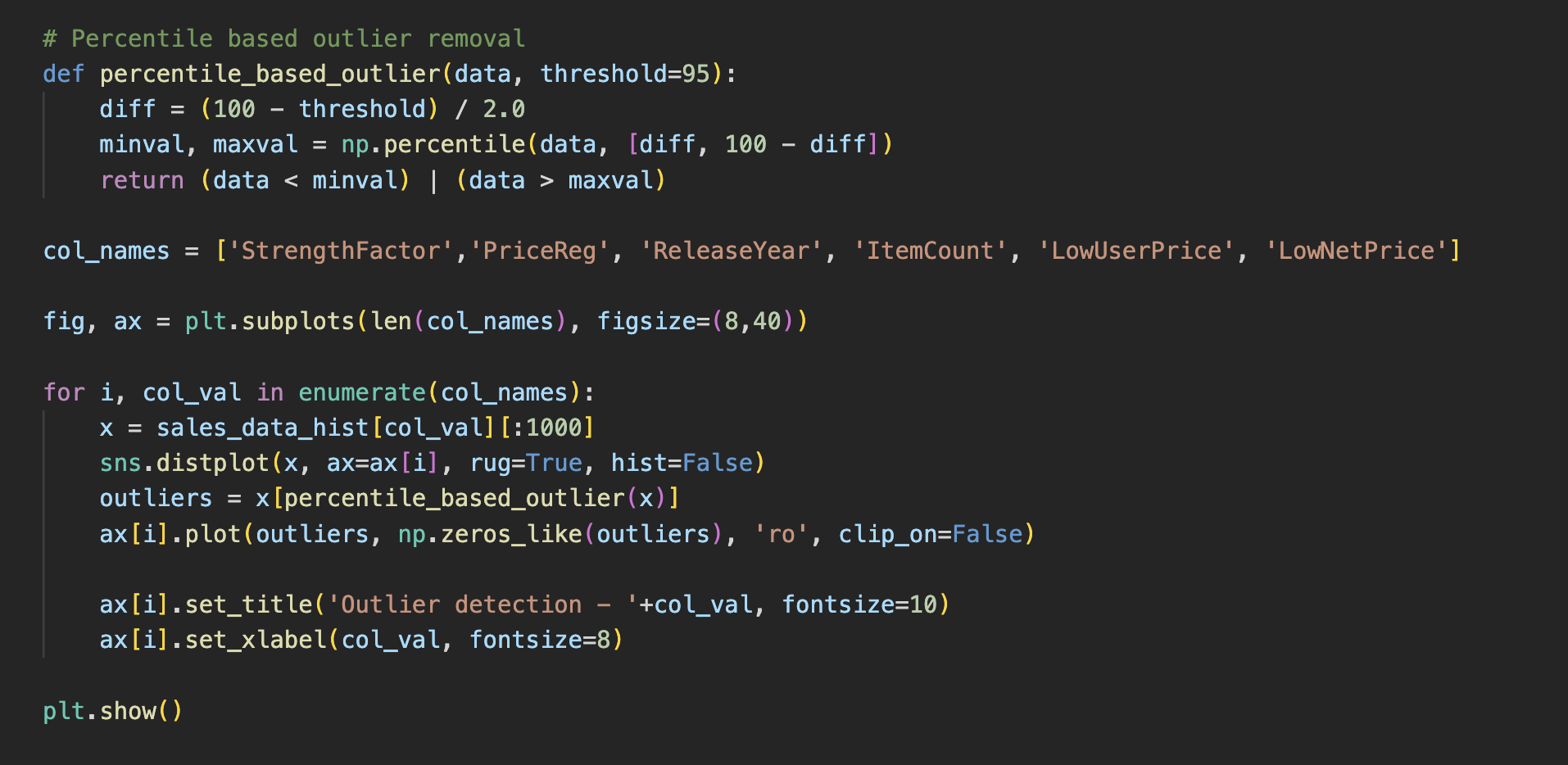
Output:

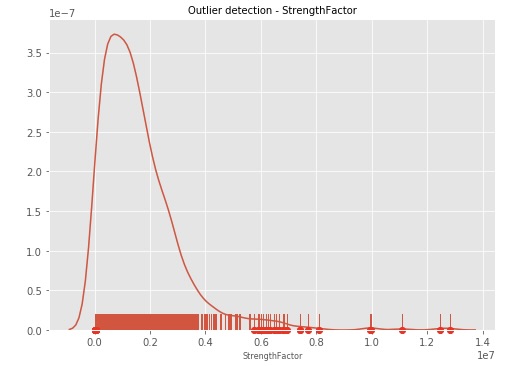
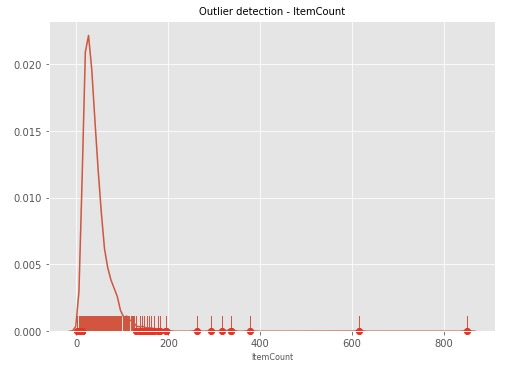
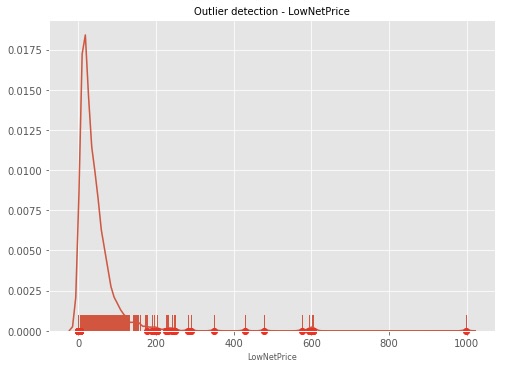
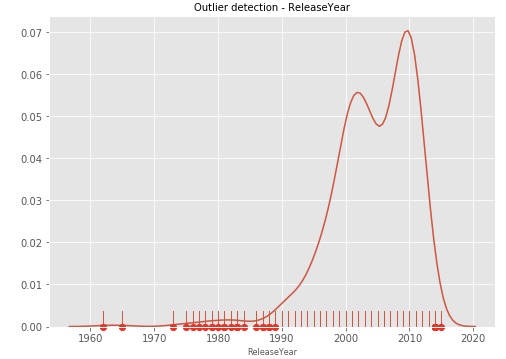
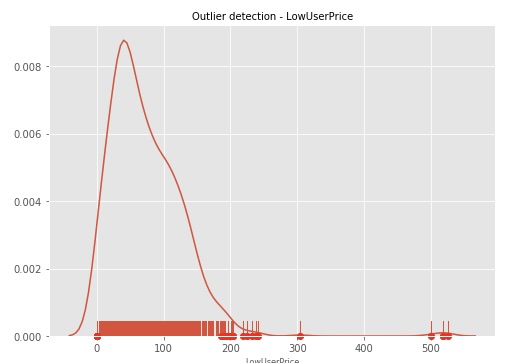
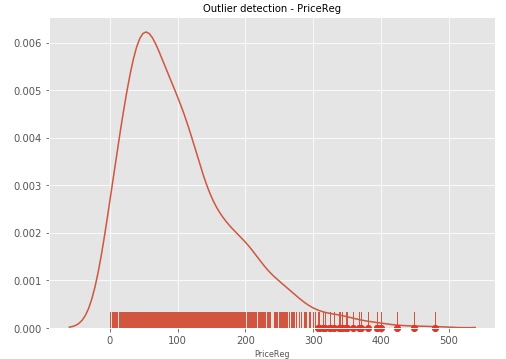
|  |
| --- |
| Cost of Class A : 205577451.4 |
| Cost of Class B : 85658215.25999999 |
| Cost of Class C : 51396361.129999995 |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Class | Total SKUs | Total Cost ($) | Quantity Percentage | Cost Percentage |
| A | 16295 | 205577451 | 21% | 60% |
| B | 22288 | 85658215 | 30% | 25% |
| C | 37413 | 51396361 | 49% | 15% |
| Total | 75996 | 342632027 | 100% | 100% |

# Outlier Treatment:

Many machine learning algorithms are highly sensitive to outliers in the data. They affect the prediction accuracy while prediction. So in this method all the outliers are removed based on the percentile for each attribute.



****

# Predictive Modelling:

In this predictive modelling, we predict the products that need to stocked in the active inventory. We accomplish this by using SKU which is an unique identifier for each product.

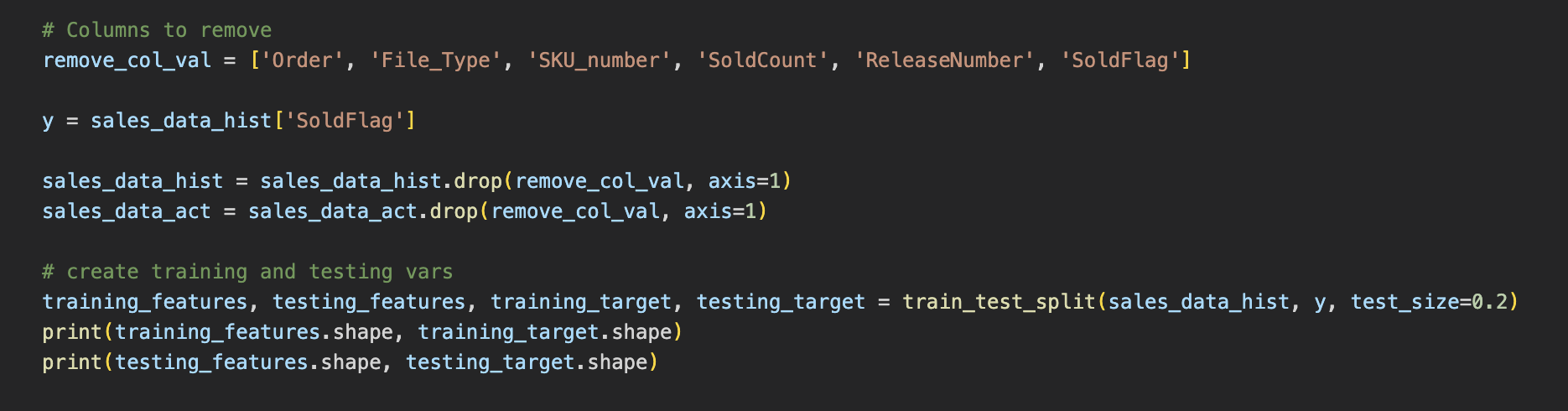
In this method initially we convert the marketing into a categorical variable as mentioned in the code below. Then to apply the model on the active inventory, we split the inventory by active and historical state.. Then we create testing and training parameters. As the data in the testing and training is skewed we may develop improper modelling. So, We implement

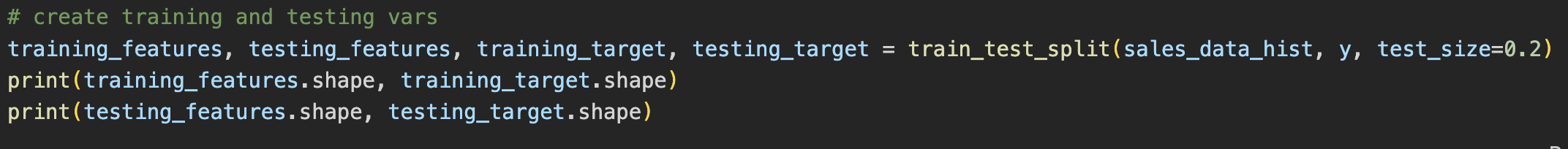
Synthetic Minority Over-sampling Technique to balance the data in which it balances the minority class to match the majority class. After the sampling has been done. Random forest classifier, Logistic regression, Decision tree are used to analyse the active inventory for true label and predicted label for sold and not sold. A confusion Matrix has been built to better understand the model accuracy.

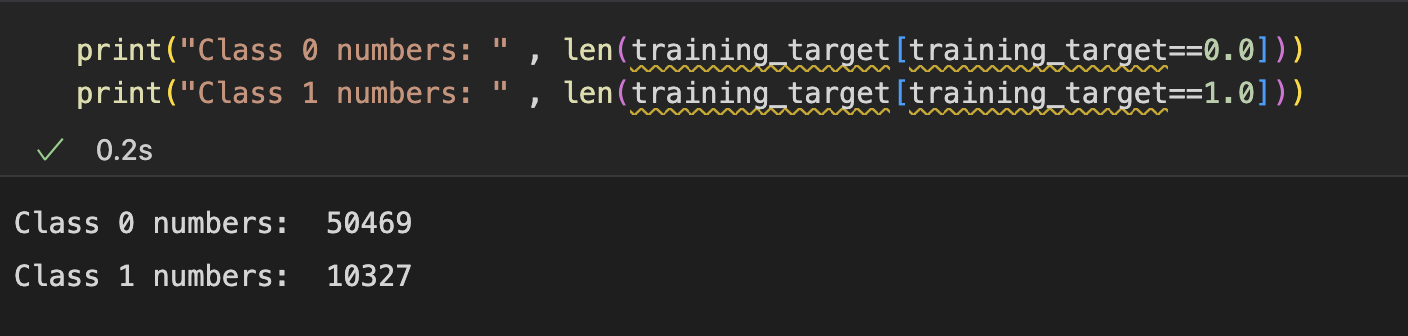
This model will give predict the SKU that will be sold and that will not be sold.

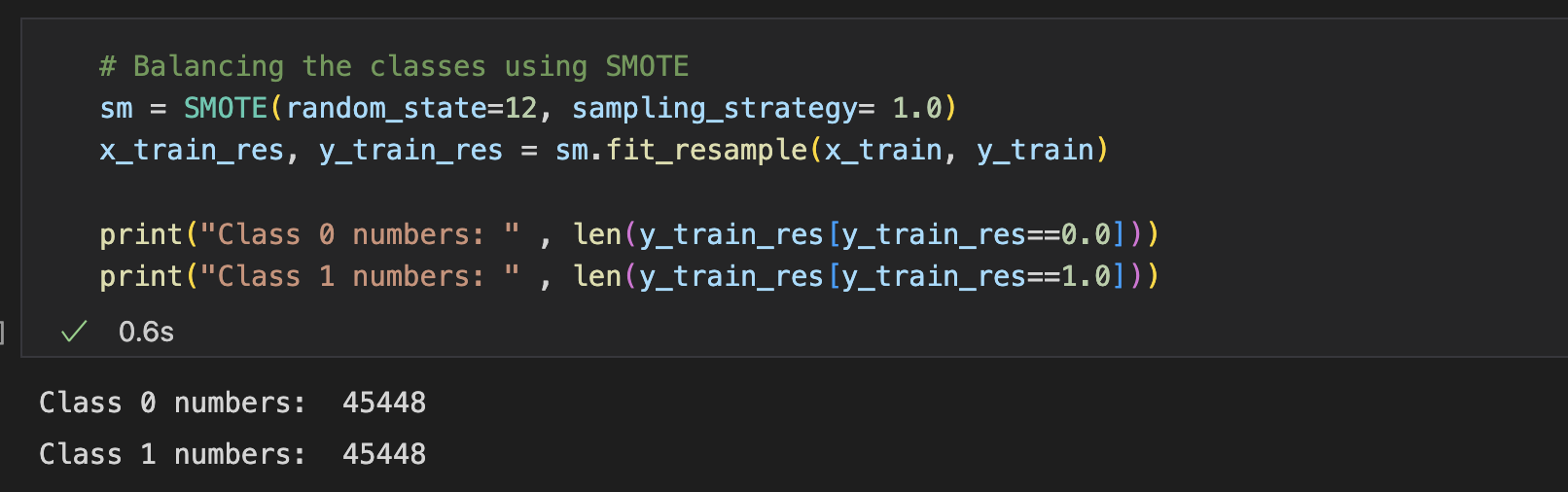
The results are in the figures given below

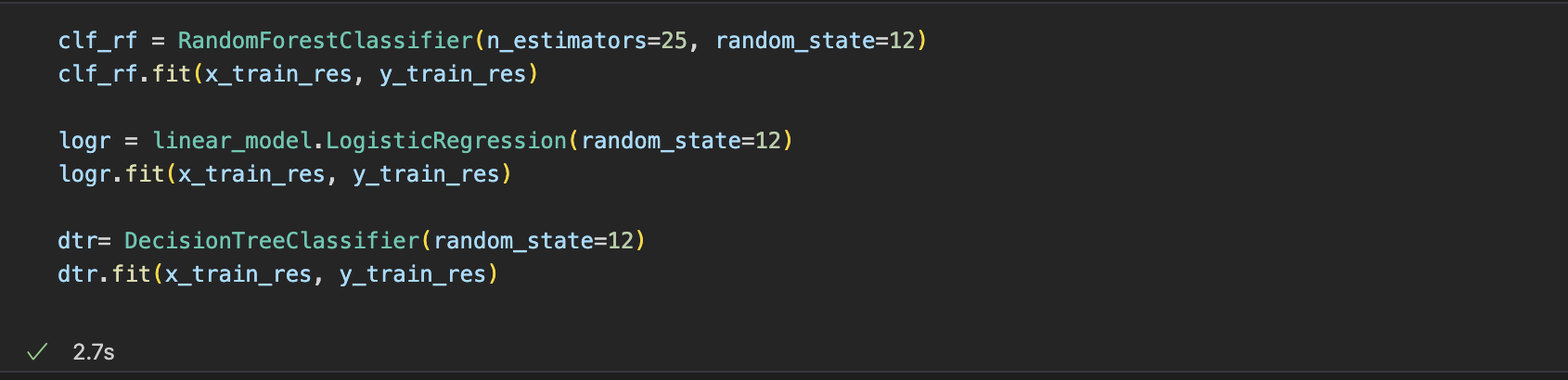


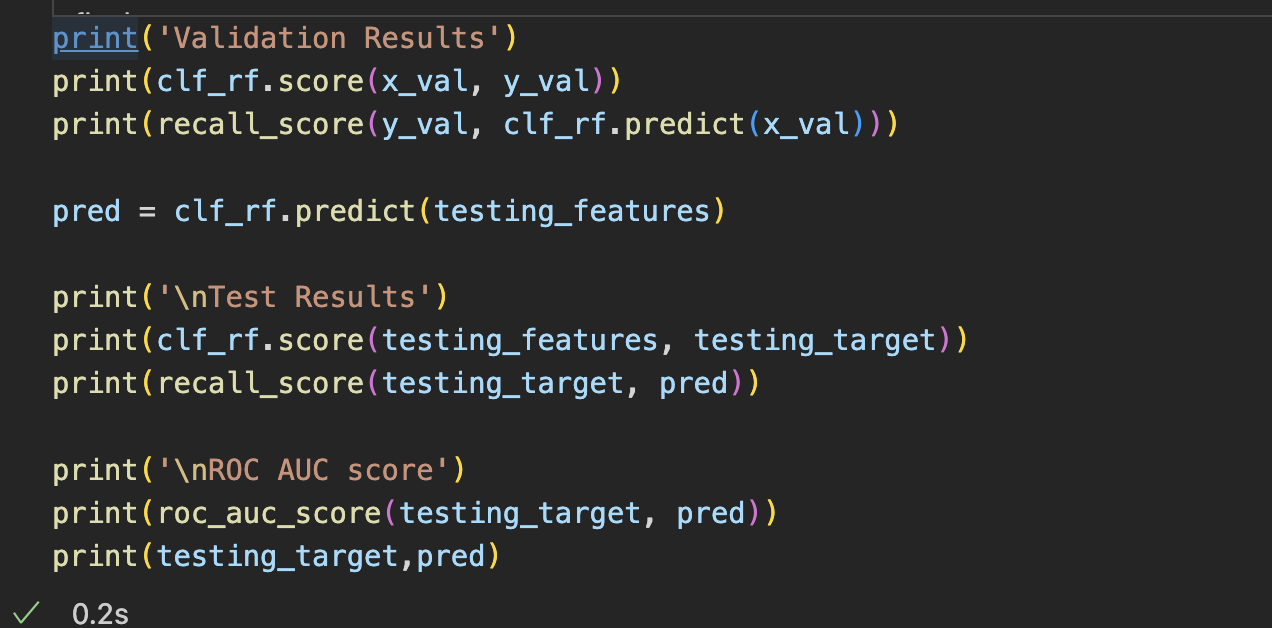












Output:

|  |  |  |  |
| --- | --- | --- | --- |
|  | **Validation Results** | Accuracy Scores | ROC-AUC score |
| Random Forest | 0.73651 | 0.7477 | 0.6657 |
| Logistic Regression | 0.69483 | 0.5807 | 0.5900 |
| Decision Tree | 0.68892 | 0.6923 | 0.6138 |

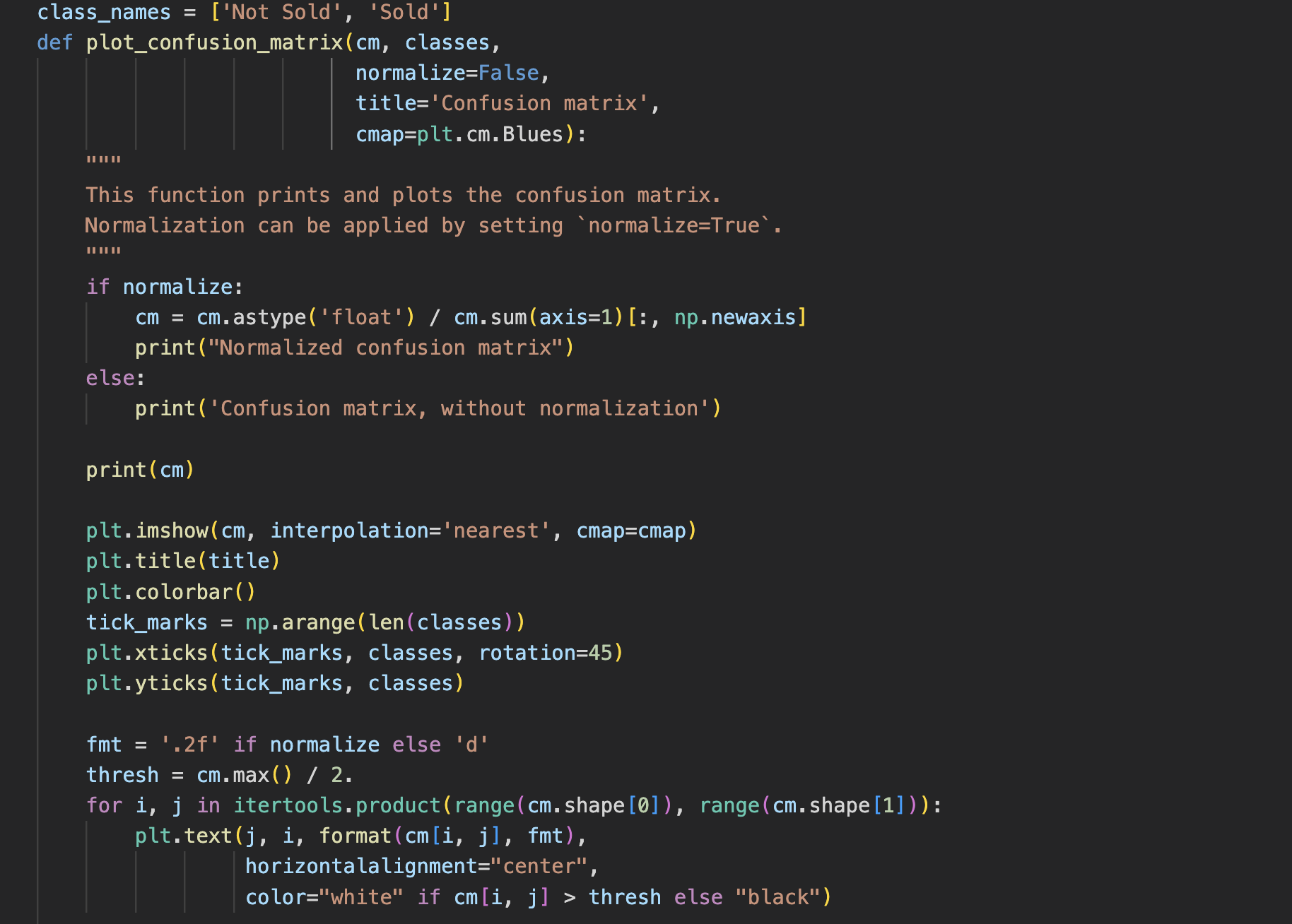
**Final Result:**

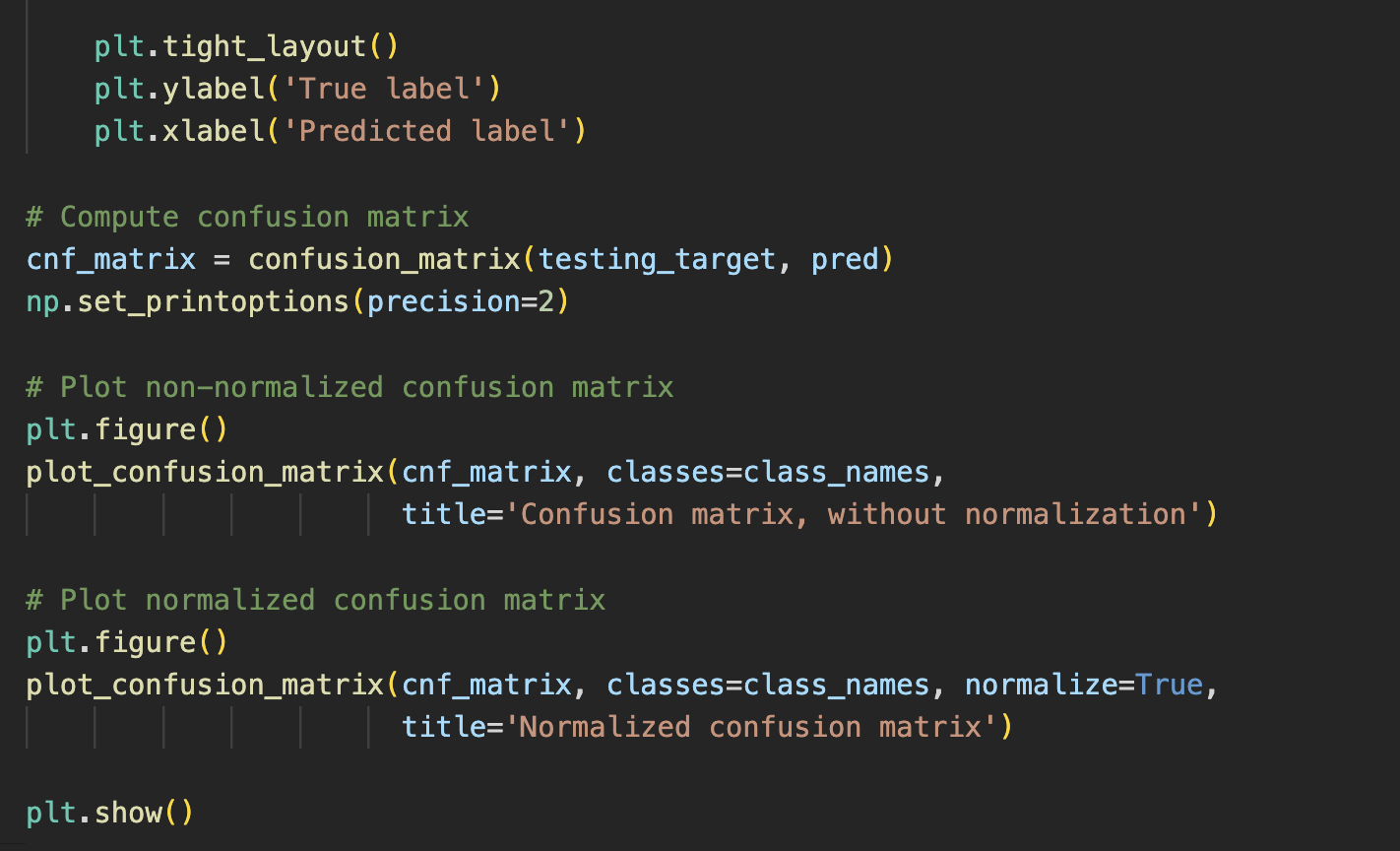
From the algorithms we have used , RF gave an accuracy of 0.75 which is the better accuracy. So, this can be used in the final model for prediction.

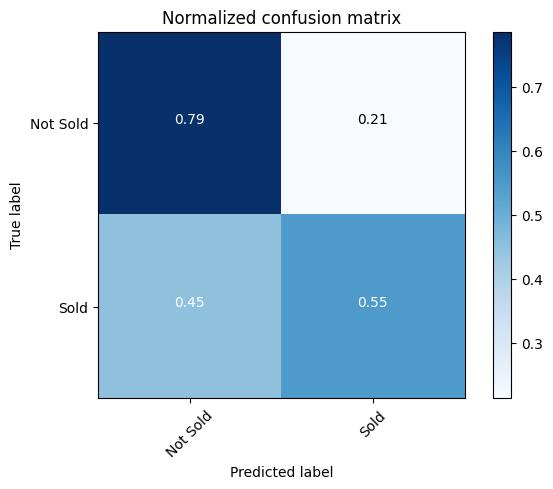
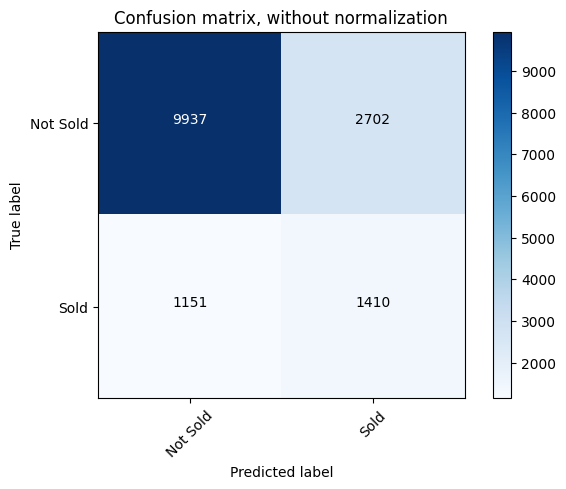
# 

# Confusion Matrix:

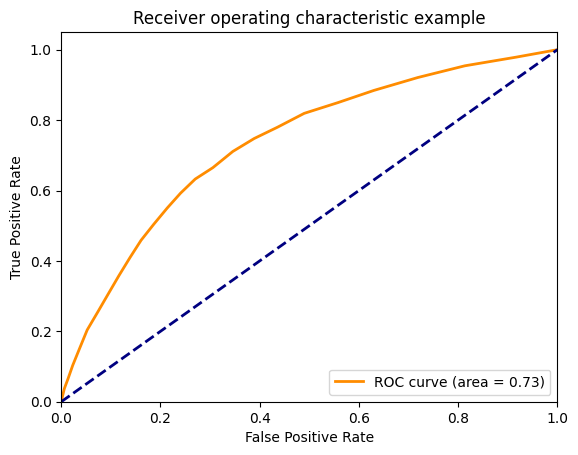
With this confusion matrix , we can understand the performance of our random forest algorithm. It Plots the predicted class and true class values over a 4 quadrant matrix and gives the number of TP,FP,TN,FN.

****

****

****

# ROC Curve:

****

# Managerial Contributions and Conclusion:

With more data and using them to train these models we will be able to improve the learning curve and get a better fit and improve accuracy. The following methods can be used to improve accuracy of the machine learning models:

1. Add more data
2. Feature Engineering
3. Feature Selection
4. Using multiple algorithms
5. Algorithm Tuning
6. Ensemble methods

Using the above methods, a manager in today’s Manufacturing industry will be able to leverage this information to predict accurately which segment of inventory will be fast moving and which products will be sold faster and stock the inventory based on the prediction. The manager should also capture more relevant information with regards to each customer like “Entry Date”, ”Exit Date”, “Location” etc to feed more relevant information to the models used. So, that the model can be implemented broadly across the scenarios.

# Limitations:

In this model Random Forest Classifier is used to predict and classify into will be sold and will not be sold. This model gives an ROC-AUC score of 0.73 which is poor. This can be solved by using more data to get a better fit, cleaning and transforming data to a better degree or even using a better algorithm that fits the purpose more accurately.

# References:

<https://github.com/SasankJayavarapu/Business-Analytics>

<https://www.kaggle.com/datasets/flenderson/sales-analysis>

<https://ieeexplore.ieee.org/document/9002289>

<https://public.tableau.com/views/BusinessAnalytics_Group_08/Dashboard1?:language=en-US&publish=yes&:display_count=n&:origin=viz_share_link>