Introduction

The retail industry has always been a dynamic and competitive field, with clothing retailers like CC's constantly looking for ways to improve their performance. One of the biggest challenges that CC's is currently facing is the low gross margins and excessive inventory going on clearance in some of their stores. We with be analyzing transaction-level data for two selected product classes (tops and dresses) in 10 different stores across the chain in 2019, with the goal of making recommendations to improve gross margin and reduce inventory sold at clearance.

We have utilized a three-step approach using Python. Starting with Exploratory Data Analysis (EDA) to gain insights into the data, followed by Clustering Analysis to identify patterns and segment the data. Finally, we used a supervised machine learning technique, Logistic Regression Model, to predict whether an item will go on clearance or not.

Exploratory Data Analysis (EDA)

Sales and Gross Profit Performance: Store 19 generated the highest sales amounting to $239,191.94 and the highest gross profit of $151,730.00). In contrast, Store 13 recorded the lowest sales of $80,598.48 and the lowest gross profit, totaling $51,317.70 (Figure 1).

Seasonal Gross Margin Variation: We observed that the highest gross margin (GM) of 68% occurred in Store 13 during the summer season, while the lowest GM of 57% was recorded in Store 11 during Winter. This finding suggests that seasonal variation might be a contributing factor to the fluctuating gross margins across different stores.

Product Classes and Gross Margin Performance: Cardigans, Wraps, and Sweaters" category had the highest GM in Store 16 at 64.58% and the lowest GM in Store 3 at 59.64%. Notably, the "Blouse" subclass recorded the highest number of units sold (11,191 units) with a GM of 61.00%, whereas the "Misc. Fashion Tops" subclass had the lowest number of units sold (2 units) with a GM of 38.60% (Figure 2).

Store Performance by Subclass: Store 14, the highest-margin store, had a GM of 69.12% for the Kimono subclass with 50 units sold. In contrast, Store 21, the lowest-margin store, reported a GM of -53.33% for the "Misc. Fashion Tops" subclass, with only 1 unit sold.

Clearance Performance: The top 3 stores with the highest GM on clearance items were Stores 27 (74%), 21 (73%), and 20 (72%) (Figure 3). Conversely, the bottom 3 stores with the lowest GM on clearance items were Stores 27 (45%), 11 (52%), and 20 (54%).
Store 19 had the highest number of clearance products (2230), while Store 13 had the lowest (575) (Figure 4). The highest clearance by season was observed in Fall at Store 19 and lowest clearance at Store 13 (3.43%) and in Winter at Store 13 (4.44%).

Cluster Analysis Approach

Our team decided to use cluster analysis on the merged dataset to gain insights into the performance and factors driving clearance sales. The aim was to segment the stores or products into groups with similar characteristics, which can help provide insights into their performance and the factors driving clearance sales. **K-means clustering with a K value of 2** was utilized, which gave a promising result.

The cluster analysis revealed two clusters that were clearly distinguished based on their percentage of clearance items. **Cluster 1 showed a higher percentage of clearance items and Cluster 2 showed a lower percentage of clearance items**.

Our team observed key differences between the two clusters in terms of the impact variables. Cluster 1 had a low average unit cost, low average sale amount, and low average gross margin. It had a high inventory of XS, XXS, and XL sizes, a high inventory of product types such as blouses, bodysuits, long-sleeve knits, and high clearance products. Additionally, it had high sales duringsummer and spring.

On the other hand, Cluster 2 had a high average unit cost, high average sale amount, and high average gross margin. It had a low inventory of XS, XXS, and XL sizes, and a high inventory of product types such as button-downs, cardigans, short sleeve knits, woven tanks, and vests. It had low clearance products and low sales during summer and spring.

Cluster Analysis Recommendations

Based on our cluster analysis, we have identified some key recommendations:

For Cluster 1, which has a higher percentage of clearance items and lower gross margin, we recommend that stores keep a close eye on **their excess inventory of XS and XXS sizes, which are typically associated with low margins, higher clearance, or both.** These stores should aim to decrease their inventory levels of these sizes.

Additionally, Store 13, which is currently part of Cluster 1, has the fewest XS and XXS size products but a good gross margin (63.8%). We recommend that other stores emulate Store 13's size combination across their products and aim to reach the same level of gross margin.

Further analysis of the product subclasses shows that blouses and short sleeve knits account for about 60% of the products put on clearance in Cluster 1. Therefore, we recommend that CCs **consider reallocating these product types between stores to help reduce clearance inventory.**

Logistic Regression Approach and Analysis

To conduct our analysis, we decided to use a supervised machine-learning technique:

Logistic Regression Model, to predict whether an item will go on clearance or not. The preliminary results of our model were not promising, so we decided to use the clusters previously defined by the Clustering Analysis and run the regression in each one of them. **The independent variables that we used in the model are: "gross_margin", dummies for the season, product classes, and size. and for colors. On the other hand, our dependent variable is "product clearance".** Based onour regressions, Cluster 1 got "gross_margin" as the significant variable (Figure 5). and Cluster 2 got"gross_margin", Spring, Winter, Cardigans, and Sweaters as significant (Figure 6). We can also find the good-fit measures in (Figure 7) for both models.

In this way, the highlight from our model is that “gross_margin” has a **negative coefficient which means that if it increases**, the probability of a product being on clearance decreases. Moreover, “Spring” and “Winter” have **positive coefficients**, meaning that products sold in Spring or Winter are more likely to be on clearance than those sold in other seasons. Additionally, **“Cardigans” and “Sweaters” also have a positive coefficient**, which means that these types of products are more likely to be on clearance than products of other categories.

Following this line, our recommendation for the company based on the logistic model is that stores with low gross margins and high product clearance can focus their offering of clearance products on the seasons of **Spring and Winter and decrease their clearance items in Fall and Summer.** Onthe other hand, it is also recommended that the stores with **low gross margins reduce their inventory of Cardigans and Sweaters and allocate them to stores with higher gross margins.**

Business Recommendations

1. Reduce the proportion of XS/XXS size product classes throughout the stores and primarily allocate most XS/XXS size product classes to store 21, rather than store 3.

2. Decrease the blouse inventory and reallocate a portion of it to store 14 from store 3.

3. Short sleeve knit inventory should be diminished, with a portion being reassigned to store 21 from store 3.

This inventory redistribution aims to optimize inventory management and enhance profitability. Stores receiving the reallocated inventory appear to be better suited for handling specific sizes and product subtypes, possibly due to factors like demographic differences or local market conditions.

By transferring inventory to locations with a higher likelihood of sales, the company can prevent both overstocking and understocking at individual stores, thus mitigating the costs associated with lost sales and surplus inventory. Moreover, prioritizing stores with higher gross margins allows the company to maximize profitability across its retail network. Overall, this approach to inventory management aims to bolster the company's financial performance while catering to customer demand more effectively.

Figure 1

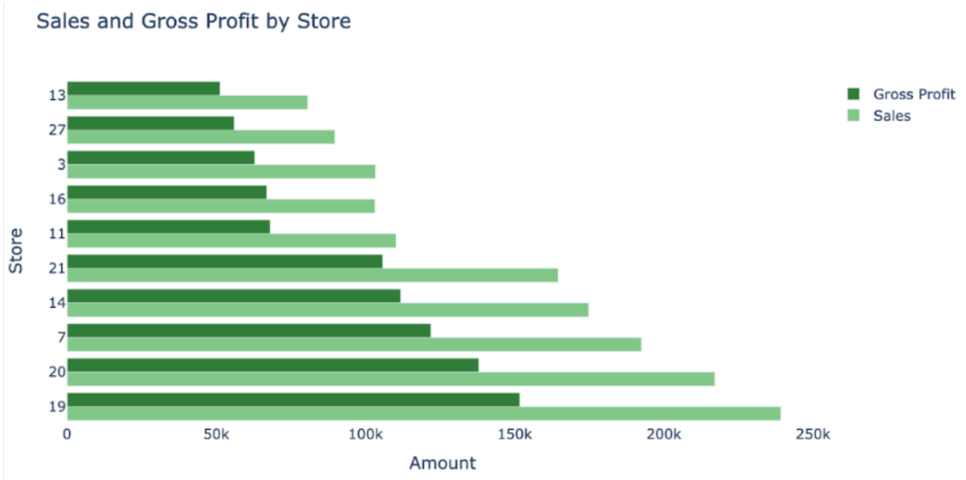

Figure2.

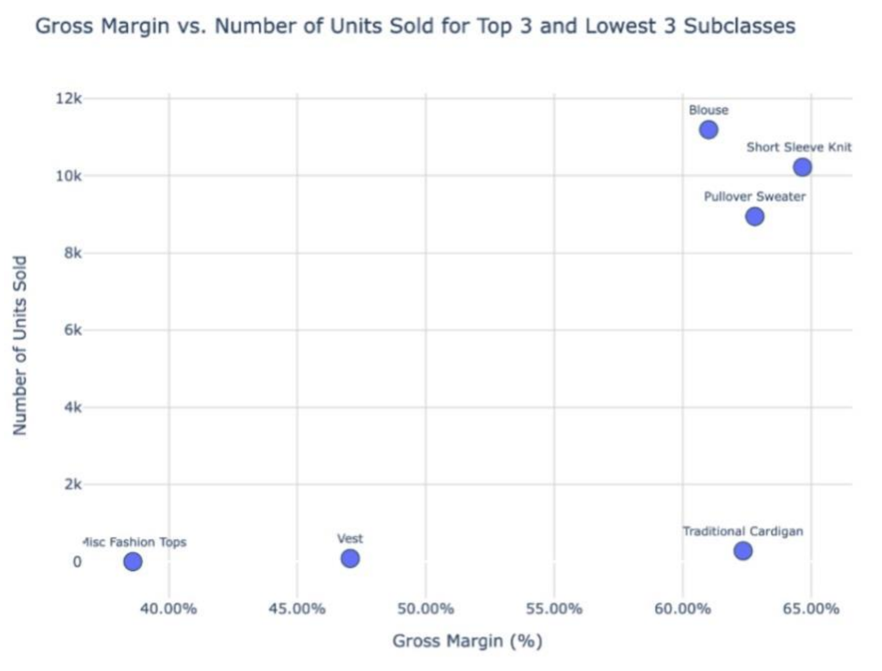

Figure 3.

Figure 4.

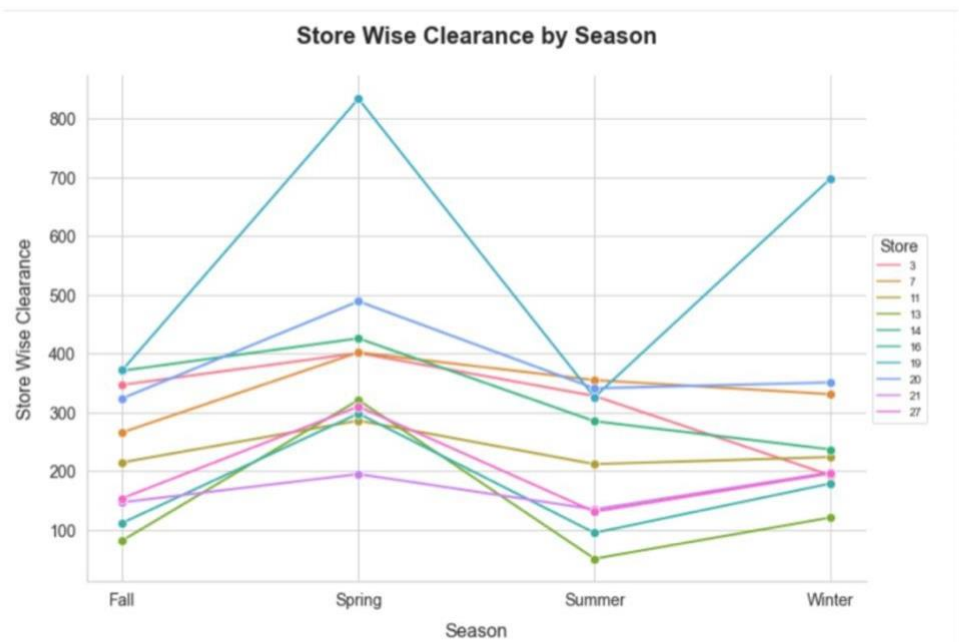

Figure 5. Cluster 1 and Cluster 2.

Cluster	Product on Clearance	Avg.Sales Amt.	Avg.Gross Margin.	Inventory (XS, XXS, XL)	Inventory (M/L)	Avg.unit Cost	Popular Product Types
1	High	Low	Low	High	Low	Low	Blouse, Bodysuits, Long Sleeve Knit
2	Low	High	High	Low	High	High	Button Down, Cardigans, Short Sleeve Knits, Woven Tanks, Vest

Figure 6. Log Model: Cluster 1

```
gross_margin: -9.733 (p-value=0.000*)
size_XXS: 0.828 (p-value=0.408)
size_L: 0.579 (p-value=0.563)
size_M: 0.469 (p-value=0.639)
size_S: 0.421 (p-value=0.674)
size_XS: 0.614 (p-value=0.539)
size_XL: 0.842 (p-value=0.400)
size_S/M: 0.000 (p-value=1.000)
size_M/L: -0.290 (p-value=0.772)
Spring: 0.633 (p-value=0.526)
Summer: 0.196 (p-value=0.845)
Winter: 0.389 (p-value=0.697)
Blouse: -0.033 (p-value=0.974)
button_down: -0.095 (p-value=0.924)
others_class: -0.426 (p-value=0.670)
cardigans: 1.257 (p-value=0.209)
sweater: -0.868 (p-value=0.385)
Blue: 0.369 (p-value=0.712)
Earthy: -0.029 (p-value=0.977)
Green: 1.081 (p-value=0.280)
Other_Color: 0.267 (p-value=0.789)
Pink/Red: 0.142 (p-value=0.887)
Purple: 0.699 (p-value=0.485)
Yellow/Orange: -0.169 (p-value=0.866)
```

Figure 7. Log Model: Cluster 2

```
gross_margin: -11.914 (p-value=0.000*)
size_XXS: 1.148 (p-value=0.251)
size_L: 0.122 (p-value=0.903)
size_M: -0.049 (p-value=0.961)
size_S: -0.071 (p-value=0.943)
size_XS: 0.042 (p-value=0.966)
size_XL: 0.182 (p-value=0.855)
size_S/M: -0.373 (p-value=0.709)
size_M/L: 0.133 (p-value=0.894)
Spring: 3.482 (p-value=0.000*)
Summer: 0.659 (p-value=0.510)
Winter: 2.369 (p-value=0.018*)
Blouse: 1.194 (p-value=0.233)
button_down: 1.568 (p-value=0.117)
others_class: 1.717 (p-value=0.086)
cardigans: 2.096 (p-value=0.036*)
sweater: 1.900 (p-value=0.057)
Blue: -0.102 (p-value=0.919)
Earthy: -0.303 (p-value=0.762)
Green: 1.792 (p-value=0.073)
Other_Color: -0.183 (p-value=0.855)
Pink/Red: -0.759 (p-value=0.448)
Purple: 0.172 (p-value=0.863)
Yellow/Orange: -0.743 (p-value=0.458)
```

Figure 8. Model Performance Measures

Cluster 1	Cluster 2
Measures: F1 Score: 0.81 Accuracy: 0.91 Precision: 0.94 Recall: 0.71	Measures: F1 Score: 0.81 Accuracy: 0.96 Precision: 0.93 Recall: 0.70