Company XYZ Retail (Initial Analysis)

Questions that are solvable by classification:

1. What are the products that **sell well** or **not**?
   1. According to gender/age/nationality/time of year/location
2. If a customer is **satisfied** or **not**?
   * What was the medium that was used to address a customer’s concern?
     + Email
     + By Phone
     + Public Statement
     + User ratings/reviews
     + Social Media Platforms (Facebook Messenger, Viber, Discord, etc)
   * How long did it take for customer support to answer
     + Less than a day
     + Less than 3 days
     + Less than a week
     + More than a week
   * Was it resolved or not?
     + At what degree was this resolved?
     + Very Good, satisfactory/resolved, terrible/did not solve
   * Positive / negative words in the customer’s review of customer service
     + Positive Examples: quick reply, accommodating, helpful, concise, straight to the point, excellent, kind, polite
     + Negative Examples: bossy, confusing, unhelpful, rude, demeaning
     + Neutral Examples: okay, satisfactory
3. How is the company performing compared to the overall industry and major competitors?
   1. Price analysis of company products against that of competitors
   2. Overall, are the prices of the clothes/shoes expensive?

**Sources of Data**

* Company Data
  + Location/Time/Date of sales
* Customer Data (assume retail membership)
* Company App for customer reviews
* Individual Customer Information of online customers
* Phone Calls
* Public Statements
* Overall Market Data for the retail clothing industry
* Prices of major competitors
* QR code for reviews of physical purchases

**Tabular Data - Raw**

| **Product** | **Personal Customer Information** | **Location / Branch** | **Total Price of Products Purchased** | **Date of Purchase** |
| --- | --- | --- | --- | --- |
| Name of Product | Sex/Age/Nationality/Address of customer  This data will be used to profile the customers who most often buy a specific product. | Classify each purchase according to what branch it took place | Data is the sum of the prices of all products bought by the consumer | Recorded by the cashier at each store |

**Tabular Data - Vectorized**

| **x1** | **x2** | **x3** | **x4** | **x5** | **x6** | **x7** | **x8** | **x9** |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Customer\_id randomly generated | One hot encoded into a vector with length equal to the number of products of the store | x3 = [Male, Female]  Binary | x4 = [12 and below, 13-20, 21-39, 40-59, 60 +]  One hot encode | x5 = [Filipino, American, Japanese, etc.]  One hot encode | x6=[NCR, Luzon, Visayas, Mindanao]  Vector of length 4 | One hot encoded  x7 = [branch\_1, branch\_2, …] | Price ranges 0-399, 400-799, 800-1199, 1200+  All in Pesos  (input 1-4 instead of raw data for Pesos; 1 for 0-399, 4 for 1200+) | Month: One hot encoded per month |

**Tabular Data - Raw**

| **Level of Satisfaction** | **Length of Review** | **Number of Employees** |
| --- | --- | --- |
| Level of customer satisfaction  The customer rates his/her level of satisfaction with the order through the company’s online app, email, website, or physical surveys in physical stores. | Number of words in the customer’s review of the product.  Filter out short reviews and focus on evaluating long reviews from customers to get the public perception of a product. | Get the number of cashiers per store. |

**Tabular Data - Vectorized**

| **x10** | **x11** | **x12** |
| --- | --- | --- |
| Rated from 1 to 5 stars | 1 if number of words is greater than or equal to 50,  0 if number of words is less than 50  No review means 0. | Current hired workers: 1-5, 6-10, 10+  Converted to numerical form (1-3) |

**Classification Applications**

| **Name of Application** | **Stakeholders** | **Key Performance Indicators** |
| --- | --- | --- |
| 1. Level of Satisfaction with the customer service  1-5 | Customer Relations: After each inquiry, the customer can send a review through the company’s app where they rate the person they talked to from a scale of 1 to 5. Using this data, the company can decide to lay off employees which give off a bad impression to clients. | Average of 4.8 out of 5 rating for all service employees.  90% rate of success for customer inquiries |
| 2. Manpower Optimization  Predict the additional number of workers needed to hire in each branch  1-5; 6-10; >10 workers to hire | Logistics Department: Add more queue lines to lessen the wait time for customers at the counter. We use the data on the number of customers in a branch per month.  Human Resources: Hire more cashiers depending on the season | Less than 500 customers served per worker per month. |
| 3. Product Profiling  Binary: sell well or not | Purchase Order:  To determine whether to continue ordering a specific product or to lessen orders; also take note of whether there has been an increase or decrease in sales for each product. Purchase orders should vary between branches because different branches have different target markets. | 50 pieces of each clothing item should be sold at each branch per month.  More than 80% of all ordered products must be sold within a year. Those not sold should be put on sale to clear up inventory. |

**Clustering Applications**

| **Name of Application** | **Stakeholders** | **Key Performance Indicators** |
| --- | --- | --- |
| 1. Customer Segmentation using online data | Marketing Department:  Group types of customers based on the products they most often buy. We will cluster based on age, gender, and location attributes and group together similar customers. This will help in identifying the target market and ideal customer for certain products, which will make advertising and marketing more personalized and effective. | Strong social media engagement of company advertisement and product placements on Facebook, Youtube, and TikTok  A 50% increase in purchases for customers after getting personalized ads. |
| 2. Location Optimization of Online Customers  Cluster based on the demography of online customers | Logistics Department:  We will use the data on the address of the online customers to find the geographic locations with a strong online consumer base. We will then check if there are any nearby physical stores. In order to save costs, these physical stores will be closed to avoid extra costs since most of the customers buy through online means anyway. This will also optimize online visits and sales. | 30% increase in website visits; 20% decrease in operating expenses |
| 3. Season/Weather/holidays (types of products bought)  Group products based on what month/holiday/season it is likely to sell  Example:  month/season/holiday data: February has Valentines, December has Christmas, August is the rainy season | Marketing Department & Production:  Based on the number of purchases of specific products during certain months, we could identify the products which sell better. For example, during Christmas, sweatshirts and jackets may be more popular among customers. We could also create and sell seasonal limited edition products for certain holidays.  marketing: ads, promos/sales based on the season and year  production: see the needs per season and adjust/create new products | 20% increase in sale items bought and a 50% increase in seasonal items bought |