Revenue Management- Part 1

Subscription on bamboo products

Presented By:-

Aastha Valecha (2019ME20858)
Nitin Katariya (2019ME10819)
Shubham Sharma (2019ME20933)

Supervisors:-

Prof. Nomesh Bolia

Mechanical Engineering Department

Indian Institute of Technology Delhi

November 2022

Introduction

- Bamboo is a very fast growing, renewable and easy-to-grow resource.
- India has more than 1600 species across the world. India has 136 species of bamboo
- India having the largest area under bamboo, China has more developed bamboo processing industries than India
- Bamboo is mostly grown in North East Region. Different kinds of bamboo product hairbrush, toothbrush, pen, earbuds, tea cups, straw, bottle,etc .
- Subscription-based models have been growing on today's market. These models assume that customers pay fee for access to a product.
- There are many different types of subscription models that are currently used by many high tech companies

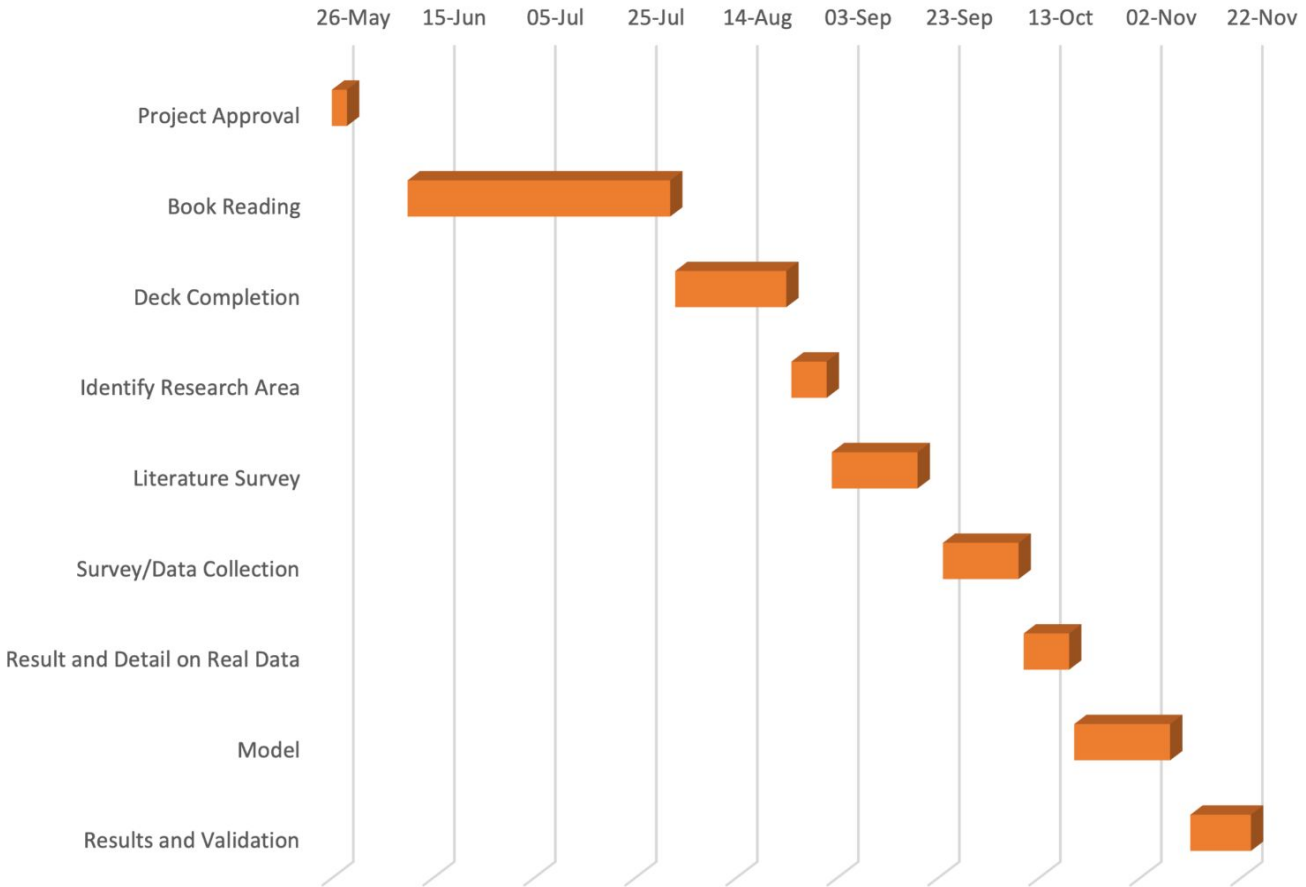
Literature Survey

- It is important to understand the use of bamboo products for eco-friendly in order to promote its application
- This plant is beneficial to the environment and very thankful to those who wish to buy high quality goods
- This saves energy, reduce CO2 emission and boost carbon storage
- Bamboo is a typical renewable and green material due to its vast volume and quick development
- Many Barriers are facing like lack of industrialization in the manufacturing process, high material costs, and a shortage of skilled workers
- The cost of production is quite expensive because there is currently no industry for processing bamboo resources

Project objectives

- To Understand the motive and barriers of subscription on bamboo based products
- To Identify customer motive and product for Subscription
- To understand various parameters involved in creating subscription of bamboo products
- Model the willingness of customers to pay using classification techniques
- Create awareness about this product among the people
- consumers behaviors towards subscription-based products and services

Gantt Chart


26-May
15-Jun
05-Jul
25-Jul
14-Aug
03-Sep
23-Sep
13-Oct
02-Nov
22-Nov
Project Approval
Book Reading
Deck Completion
Identify Research Area
Literature Survey
Survey/Data Collection
Result and Detail on Real Data
Model
Results and Validation

Methodology

- Identifying Characteristic Barriers
  - Analysis of barrier by ISM Method
  - Establishment of an SSIM(Structural Self-Interaction Matrix)
  - Establishment of an Adjacency Matrix
  - Establishment of an Reachability Matrix
- Level Partition of Barriers
  - Reachability Set
  - Antecedent Set
  - Intersection Set

- Modelling willingness to pay
  - Sample and data Collection
  - Pre-processing and parameter investigation
  - Implemented Logistic Regression and Support Vector Machine classifier
  - Mathematical modelling of LR and SVM
  - Re-implemented these classifiers from scratch
  - Evaluated the results predicted via these classifiers
  - Data Analysis
- Classifier Implementation Methodology
  - Used sklearn to do preliminary evaluation of willingness of customers to pay
  - Mathematical Modelling of the two classifiers
  - Implementation of two models from scratch

Theory

ISM

It is a modeling technique in which the specific relationships of the variables and the overall structure of the system under consideration are portrayed in a digraph model.

SSIM(Structural Self-Interaction Matrix)

V: parameter i will lead to parameter j;
A: parameter j will lead to parameter i;
X: parameter i and j will lead to each other; and
O: parameters i and j are unrelated.

Adjacency Matrix

- A contextual relationship of “direct influence” was used to indicate that one barrier directly influenced another barrier.
  - If P has a direct influence on Q, then (P, Q) entry :1, else 0
  - If Q has a direct influence on P, then the (Q, P) entry 1, else 0
  - If P has a direct influence on Q, and at same time Q has direct influence on P, then entries (P, Q) & (Q, P) be1.

Reachability

The adjacency has been converted into a binary matrix, called the initial reachability matrix by substituting V, A, X and O by 1 and 0 as per the case.

Analysis

Table1:Establishment of an SSIM

	B15	B14	B13	B12	B11	B10	B9	B8	B7	B6	B5	B4	B3	B2	B1
B1	O	O	O	O	O	V	O	O	O	O	O	O	O	O	
B2	O	O	O	O	O	O	X	O	O	O	O	O	O		
B3	O	O	O	O	O	V	O	V	O	O	O	O			
B4	V	V	V	V	O	O	V	O	O	V	V				
B5	O	O	V	A	O	O	O	O	O	O					
B6	O	A	A	O	O	A	V	V	V						
B7	O	O	O	O	V	O	O	O							
B8	O	O	O	O	V	O	O								
B9	O	O	O	O	V	X									
B10	O	O	O	O	V										
B11	O	O	O	O											
B12	O	O	O												
B13	O	O													
B14	O														
B15															

Table2: Establishment of Adjacency Matrix

A	B1	B2	B3	B4	B5	B6	B7	B8	B9	B10	B11	B12	B13	B14	B15
B1	1	0	0	0	0	0	0	0	0	1	0	0	0	0	0
B2	0	1	0	0	0	0	0	0	1	0	0	0	0	0	0
B3	0	0	1	0	0	0	0	1	0	1	0	0	0	0	0
B4	0	0	0	1	1	1	0	0	1	0	0	1	1	1	1
B5	0	0	0	0	1	0	0	0	0	0	0	0	1	0	0
B6	0	0	0	0	0	1	1	1	1	0	0	0	0	0	0
B7	0	0	0	0	0	0	1	0	0	0	1	0	0	0	0
B8	0	0	0	0	0	0	0	1	0	0	1	0	0	0	0
B9	0	1	0	0	0	0	0	0	1	1	1	0	0	0	0
B10	0	0	0	0	0	1	0	0	1	1	1	0	0	0	0
B11	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0
B12	0	0	0	0	1	0	0	0	0	0	0	1	1	0	0
B13	0	0	0	0	0	1	0	0	0	0	0	0	1	0	0
B14	0	0	0	0	0	1	0	0	0	0	0	0	0	1	0
B15	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1

Table3: Establishment of Reachability Matrix

R	B1	B2	B3	B4	B5	B6	B7	B8	B9	B10	B11	B12	B13	B14	B15
B1	1	1	0	0	0	1	1	1	1	1	1	0	0	0	0
B2	0	1	0	0	0	1	1	1	1	1	1	0	0	0	0
B3	0	1	1	0	0	1	1	1	1	1	1	0	0	0	0
B4	0	1	0	1	1	1	1	1	1	1	1	1	1	1	1
B5	0	1	0	0	1	1	1	1	1	1	1	0	1	0	0
B6	0	1	0	0	0	1	1	1	1	1	1	0	0	0	0
B7	0	0	0	0	0	0	1	0	0	0	1	0	0	0	0
B8	0	0	0	0	0	0	0	1	0	0	1	0	0	0	0
B9	0	1	0	0	0	1	1	1	1	1	1	0	0	0	0
B10	0	1	0	0	0	1	1	1	1	1	1	0	0	0	0
B11	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0
B12	0	1	0	0	1	1	1	1	1	1	1	1	1	0	0
B13	0	1	0	0	0	1	1	1	1	1	1	0	1	0	0
B14	0	1	0	0	0	1	1	1	1	1	1	0	0	1	0
B15	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1

Table 4: Level 1 Partition

Barrier	Reachability Set	Antecedent Set	Intersection Set	Level
1	1,2,6,7,8,9,10,11	1	1	
2	2,6,7,8,9,10,11	1,2,3,4,5,6,9,10,12,13,14	2,6,9,10	
3	2,3,6,7,8,9,10,11	3	3	
4	2,4,5,6,7,8,9,10,11,12,13,14,15	4	4	
5	2,5,6,7,8,9,10,11,13	4,5,12	5	
6	2,6,7,8,9,10,11	1,2,3,4,5,6,9,10,12,13,14	2,6,9,10	
7	7,11	1,2,3,4,5,6,7,9,10,12,13,14	7	L1
8	8,11	1,2,3,4,5,6,8,9,10,12,13,14	8	
9	2,6,7,8,9,10,11	1,2,3,4,5,6,9,10,12,13,14	2,6,9,10	
10	2,6,7,8,9,10,11	1,2,3,4,5,6,9,10,12,13,14	2,6,9,10	
11	11	1,2,3,4,5,6,7,8,9,10,12,13,14	11	
12	2,5,6,7,8,9,10,11,12,13	4,12	12	
13	2,6,7,8,9,10,11,13	4,5,12,13	13	
14	2,6,7,8,9,10,11,14	4,14	14	
15	15	4,15	15	

Table 5: Summary of Level Partition results

Level	Barriers
L1	B11—Limited market for bamboo Products
L2	B7—Research and development (R&D) B8—Production Costs B15—Technology transfer mechanism
L3	B2—Knowledge about bamboo Products B6—Enterprise participation B9—Education B10—Traditional bias against bamboo Products
L4	B1—Durability B3—Industrialization for production B13—Code of practice in the bamboo industry B14—Incentive policy
L5	B5—Standards and specifications
L6	B12—Certification mechanism for bamboo materials
L7	B4—Ineffective action by government departments

Instrumentation of Data

- Data collection
- Pre-processing of Data
- Binary classification of parameters
- Statistical Analysis

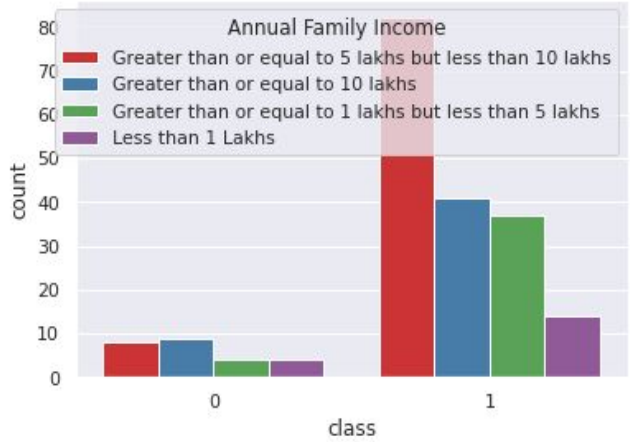


Fig 1: Annual Family Income

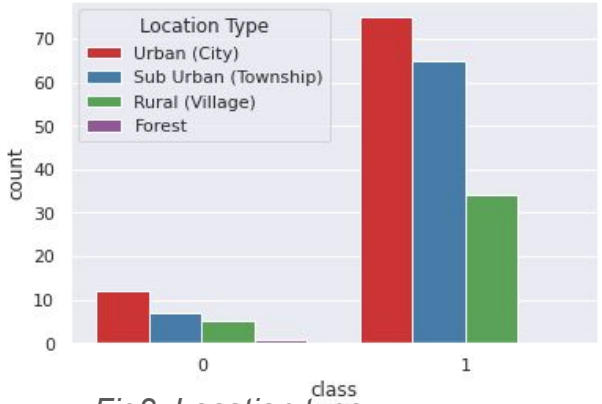


Fig2: Location type

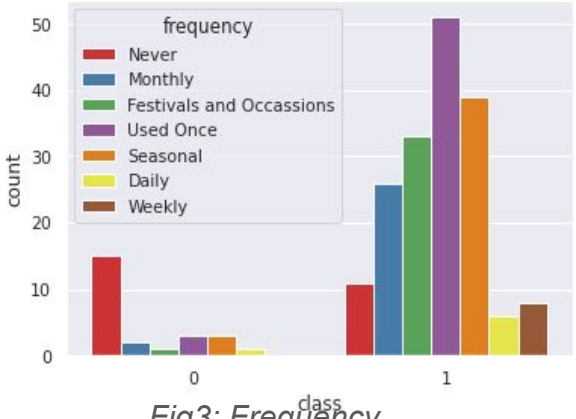


Fig3: Frequency

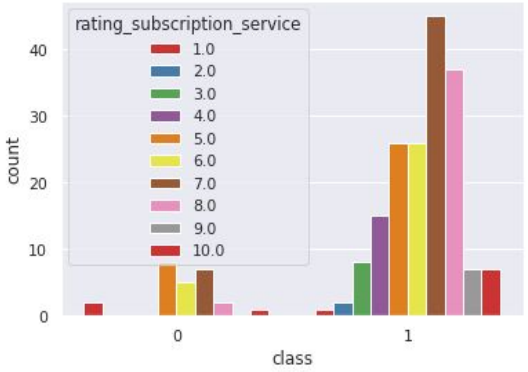


Fig4: Rating Subscription Service

Logistic Regression

The model predicts that about 86% of the people who filled the survey would be interested in bamboo based products subscription service

Mathematical Modelling

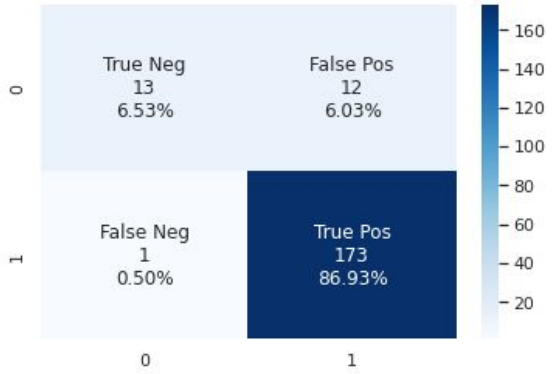


Fig 5: Confusion Matrix

1. Hypothetical function for logistic regression
   h(x) = sigmoid (wx + b)
   Where x is the feature vector obtained from pre-processing of the bamboo products
   w is the weight that is assigned to each feature, b is the bias we defined

2. Updating Weights
   sigmoid(z) = 1 / (1+e(z))
   ***z** or pi(X)*
   will be called the log(odd) or logit or log(p/1-p))

$$\pi(\mathbf{X}) = \frac{\exp(\beta_0 + \beta_1 X_1 + \ldots + \beta_k X_k)}{1 + \exp(\beta_0 + \beta_1 X_1 + \ldots + \beta_k X_k)} = \frac{\exp(\mathbf{X}\beta)}{1 + \exp(\mathbf{X}\beta)} = \frac{1}{1 + \exp(-\mathbf{X}\beta)},$$

3. Optimization algorithm used is gradient descent

$$
\begin{aligned}
&Repeat\ \{ \\
&\theta_j := \theta_j - \frac{\alpha}{m} \sum_{i=1}^{m} (h_\theta(x^{(i)}) - y^{(i)}) x_j^{(i)} \\
&\}
\end{aligned}
$$

4. probability is represented mathematically

$\mathbf{P(Y = y|X = x) = \sigma(\beta^T x)^y \cdot [1 - \sigma(\beta^T x)]^{(1-y)}}$; *where y is either 0 or 1*

J = -ylog(h(x))- (1- y)log(1- h(x)) here, y is the real target value i.e. the willingness of customer to purchase.

h(x)= sigmoid(wx + b)

For y = 0, J = -log(1- h(x)) and y = 1, J= - log(h(x))

Target- we need to maximize the probability by minimizing the loss function.

Support Vector Machine

The model predicts that about 90% of the people who filled the survey would be interested in bamboo based products subscription service

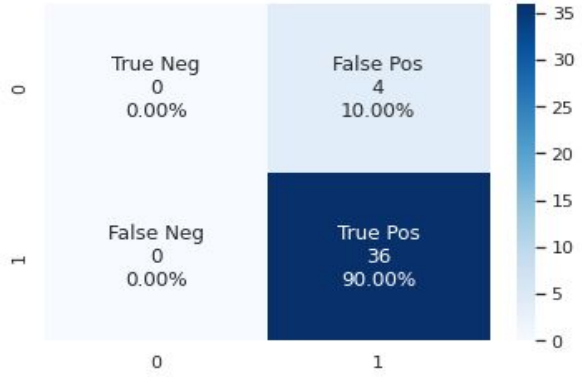


Fig 6: Confusion Matrix

Mathematical Modelling

When the marginal distance between these two classes is at their maximum, they become the optimal solution for maximizing margin and minimizing risk.

Hyperplane equation dividing the points (for classifying) can be written as:

H: $wT(x) + b = 0$

The goal of the algorithm involved behind SVM:

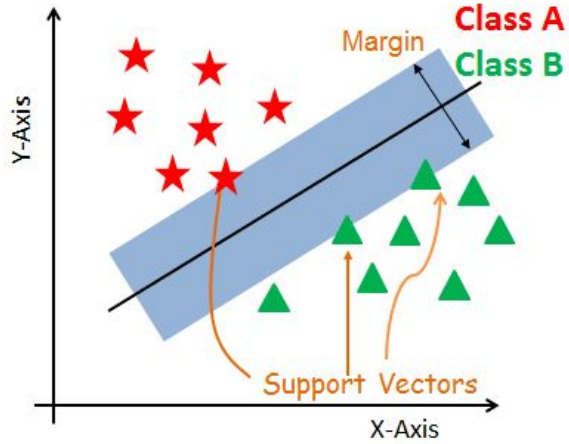


Fig 6: svm

The goal is to maximize the minimum distance.

$$d_H(\phi(x_0)) = \frac{|w^T(\phi(x_0)) + b|}{||w||_2} \qquad w^* = arg_w max\ [min_n\ d_H(\phi(x_n))]$$

$$y_n[w^T\phi(x) + b] = \begin{cases} \geq 0\ if\ correct \\ < 0\ if\ incorrect \end{cases}$$

For perfectly separable datasets, the optimal hyperplane classifies all the points correctly, further substituting the optimal values in the weight equation.

$$w^* = arg_w max\left[min_n \frac{|w^T(\phi(x_n)) + b|}{||w||_2}\right] = arg_w max\left[min_n \frac{y_n|w^T(\phi(x_n)) + b|}{||w||_2}\right] \quad \because\ perfect\ seperation$$

Python Modelling Of LR and SVM from scratch

Logistic Regression

```
# Function for model training
                # no_of_training_examples, no_of_features

                # weight initialization
                # gradient descent learning
                # Helper function to update weights in
gradient descent
                # calculate gradients
                # update weights
               # Hypothetical function h( x )
                predict
                Z = 1 / ( 1 + np.exp( - ( X.dot( self.W ) +
self.b ) ) )
                Y = 0,1
                Constraint
                Z > 0.5
```

Support Vector Machine

```
# Gradient Descent logic
                    # Calculating the Hinge Loss
            # Appending all losses
              losses.append(l)

            # Starting from 0 to the number of samples
with batch_size as interval
                                # Calculating the gradients


                          #w.r.t w
                           # w.r.t b
                          # Updating weights and bias
                w = w - learning_rate * w + learning_rate
* gradw
                b = b + learning_rate * gradb
```

Results and Discussion

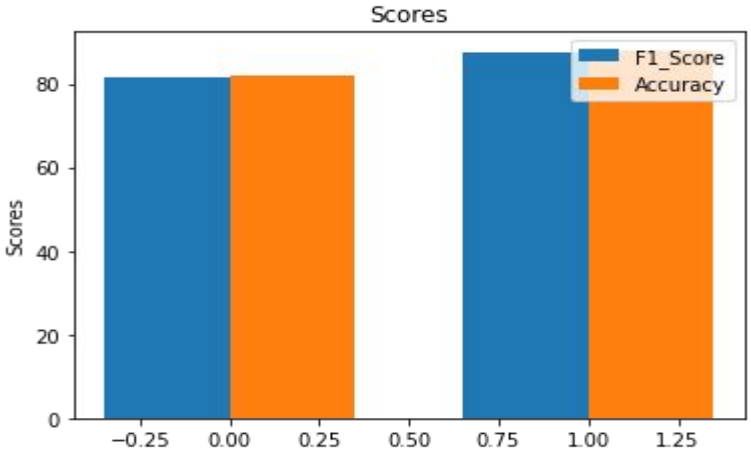


Fig 7: Scores

Logistic Regression SVM

1. Logistic Regression

Accuracy on test set by mathematical model : 90.0
Accuracy on test set by sklearn model : 92.5

2. SVM

Accuracy on test set by mathematical model : 90.0
Accuracy on test set by sklearn model : 90.0

According to both the models around 80 to 90 per cent of the people will be willing to buy subscription based bamboo products.

True positives
LR- 86
SVM- 90

The models show that SVM is more consistent
Logistic regression by sk method has the highest accuracy but this also includes true positives

F1 score of the models give us better understanding about willingness of people to pay

Data Analysis

Properties of Bamboo Products Rating, Factors Negatively affecting Bamboo Products

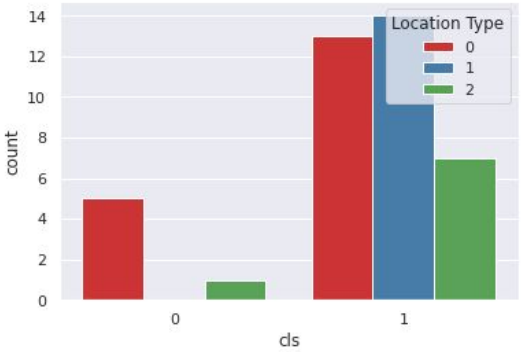


Fig 8: Location Type

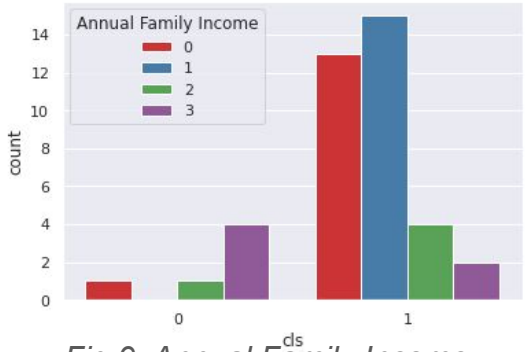


Fig 9: Annual Family Income

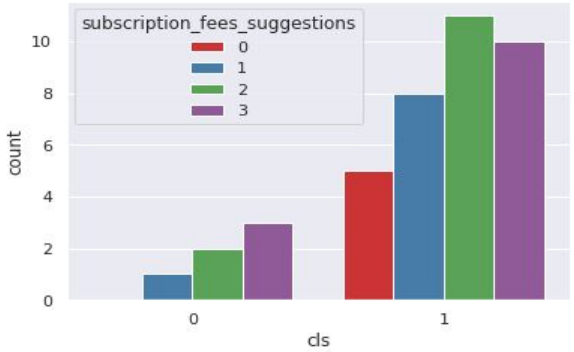


Fig 10: Subscription Fees Suggestion

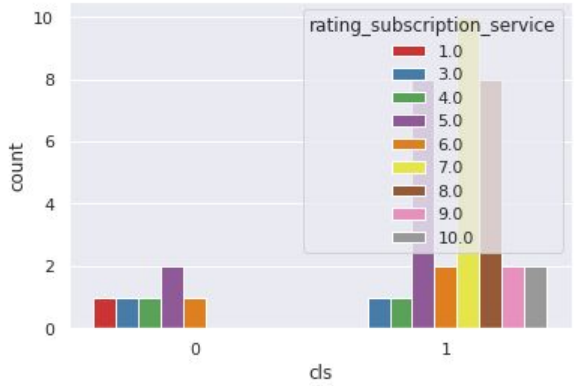


Fig11: Rating Subscription Service

Conclusion and future Prospective

- It was found that ineffective action by the government department and limited market was the most essential barrier affecting the application of bamboo Products
- Bamboo, an incomparably renewable grass, is been touted as the sustainable solution
- Switching to bamboo based products is quite impactful for the society and hence we plan to create strategy for subscription-based services for bamboo products.
- ISM helps top-level authorities make effective policies, standards, and regulations to guide the development of bamboo products
- Barriers with strong driving powers for the application of bamboo products included standards and specifications, durability, industrialization for production.

- Using this information we will have a target customer's characteristics like which state they belong to, their income level etc..
- This customer information is relevant to a model. We also have the states where bamboo production is taking place.
- With this data we will be able to model the logistic inventories of the bamboo subscription.
- Also we will be able to make a supply chain modeling using this data.
- As a prospective future project we can collect more data in the survey and then extensively create better models.

Thank you