ASSIGNMENT

Name: Ambuj Singh

Problem Statement

Based on the data of both the Chinese and the Indian markets, task is to help the company decide whether to enter the Indian market. Since you can't analyze the entire Indian and the Chinese market, the company has asked you to analyze sample data over two major cities, one from each country, to understand the sales pattern. Therefore, you are provided with the data for Shanghai and Mumbai.

Assignment Objective

- Identify Potential Market
- Potential Customer Base In India
- Business Decision and Estimated Sales and Revenue

PART I: Potential Customer Base in India

Identify Potential Market / Classification Model

Classification Model			
Actual/Pred.	Not purchase	Purchase	
Not purchase	9753	7216	16969
Purchase	5550	17481	23031
	15303	24697	

Accuracy	68%
Cutoff	0.5

probability Cutoff	TP	TN	FP	FN	Accuracy	Senstivity	Specificity
0.1	23031	0	16969	0	0.58	1.00	0.00
0.2	22834	893	16076	197	0.59	0.99	0.05
0.3	21631	3325	13644	1400	0.62	0.94	0.20
0.4	20557	5693	11276	2474	0.66	0.89	0.34
0.5	17481	9753	7216	5550	0.68	0.76	0.57
0.6	14531	12756	4213	8500	0.68	0.63	0.75
0.7	10693	14605	2364	12338	0.63	0.46	0.86
0.8	5523	17508	757	16212	0.54	0.24	0.96
0.9	1350	16850	119	21681	0.46	0.06	0.99
1	0	23031	0	16969	0.42	0.00	1.00

Variable	Coefficient	Interpretation
CURR_AGE	-2.450374419	This coefficient doesn't have any dependency on model because of negative coefficient
Gender_N	-0.011399128	This coefficient doesn't have any dependency on model because of negative coefficient
ANN_INCOME	0.237585958	This coefficie have impact on model because of possitive value
Age_phn_category	0.000002467	This coefficient doesn't have any dependency on model because of less possitive value
PURCHASE	0.970918248	This coefficient des have high dependency on model because of high possitive coefficient value

PART I : Potential Customer Base in India

Identify Potential Market

- According to the classification model we can see that accuracy of this model is up to 68 % with a cut-off value of 0.5.
- We have segmented data on behalf of the phone age category in 4 Segments as described in the figure.
- Gender-wise distribution, in that 55.71% of all population belongs to Male Category.
- Age Wise Distribution, most of the person belongs to the (25-39),(40-54) Age Group and majority in male persons those who have purchased a new phone.

- We have further analyzed the Classification model for Accurate Results. By Accuracy, Sensitivity of model, and Specificity of model

Count of PURCHASE	Column Labels		
Phone Age Catego	0	1	Grand Total
3	6.22%	23.02%	29.24%
2	23.77%	17.54%	41.31%
4	2.13%	11.12%	13.25%
1	10.31%	5.90%	16.21%
Grand Total	42.42%	57.58%	100.00%

Phone Age Category

Days	Segment
<200	1
200-360	2
360-500	3
>500	4

	Total Sellphone purchase		
	0	1	Grand Total
Count of PURCHASE	16969	23031	40000

Classification Model			
Actual/Pred.	Not purchase	Purchase	
Not purchase	9753	7216	16969
Purchase	5550	17481	23031
	15303	24697	

Count of PURCHASE	Gender Wise Distribution			
Row Labels	0	1	Grand Total	
25-39	16.88%	19.80%	36.67%	Youth Category
40-54	14.06%	22.52%	36.57%	Senior Category
55-69	13.36%	13.40%	26.76%	Old Category
Grand Total	44.29%	55.71%	100.00%	

0	Female
1	Male

PART I : Potential Customer Base in India

Identify Potential Market

- According to Salary wise distribution we can see that those who belong to the medium income category have a maximum purchase ratio of up to 25.07%
- Those who have higher income have purchased fewer smartphones according to EDA.
- On further investigation of the Classification model we have analyzed the sensitivity of the model for different y and of 0.6% model is having maximum Accuracy, Sensitivity, and Specificity.
- So on Cut-off 0.6% Accuracy, Sensitivity, Specificity are as follows-0.68%,0.76%,0.57%
- Confusion Matrix for different Probability Cut-off is given in table.

Salary wise purchase 0 Not Purchase/1 Purchase				
Row Labels	0	1	Grand Total	
70089-320089	23.10%	23.48%	46.58%	Low Income
320089-570089	15.03%	25.07%	40.10%	Medium Income
570089-820089	4.29%	9.03%	13.32%	High income
Grand Total	42.42%	57.58%	100.00%	

Accuracy	Senstivity	Specificity
Cutoff = 0.6	Coefficient Considered for probability cutoff	
0.68	0.76	0.57

probability Cutoff	TP	TN	FP	FN	Accuracy	Senstivity	Specificity
0.1	23031	0	16969	0	0.58	1.00	0.00
0.2	22834	893	16076	197	0.59	0.99	0.05
0.3	21631	3325	13644	1400	0.62	0.94	0.20
0.4	20557	5693	11276	2474	0.66	0.89	0.34
0.5	17481	9753	7216	5550	0.68	0.76	0.57
0.6	14531	12756	4213	8500	0.68	0.63	0.75
0.7	10693	14605	2364	12338	0.63	0.46	0.86
0.8	5523	17508	757	16212	0.54	0.24	0.96
0.9	1350	16850	119	21681	0.46	0.06	0.99
1	0	23031	0	16969	0.42	0.00	1.00

PART II : Customer Segmentation

Clusters for Effective Marketing

Cluster Summary	X_CURR_AGE	Y_Gender_N	Z_Inc (Lac)	M_Age_phn_category	N_PURCHASE	Category
1	33.29	0.57	276630.6	2.34	0.56	This Cluster belongs to Youth Category and having income more than Senior Category People
2	55.44	0.48	251287	2.42	0.51	This Cluster Belongs to Old Category and having Least Income.
3	49.88	0.61	570866.5	2.45	0.66	This Cluster Belongs to Senior Category and having Superior Income

- We have used K Means Clustering to optimize the solution for the different categories.
- We found three Different clusters having different properties.
- Purchase probability for these clusters more than Cut-off value 0.5%

Market Share by 40% in each Cluster			
Cluster 1	15771	6308.4	
Cluster 2	11942	4776.8	
Cluster 3	12275	4910	
Market Entry Condition		15995.2	>12000

	Optimized	Iteration 6				
	Cluster Centre	X	Y	Z	M	N
for Euc dist 7	1	33.43	0.57	27.73	2.34	0.56
	2	55.59	0.48	25.10	2.42	0.51
	3	49.91	0.61	57.17	2.45	0.66
		Current	Previous			
	Clusster 1	15772	15991			
	Clusster 2	11947	11799			
	Clusster 3	12281	12210			
	Total	40000				
	Optimized	Iteration 7				
	Cluster Centre	X	Y	Z	M	N
for Euc dist 8	1	33.29	0.57	27.66	2.34	0.56
	2	55.44	0.48	25.13	2.42	0.51
	3	49.88	0.61	57.09	2.45	0.66
		Current	Previous			
	Clusster 1	15771	15772			
	Clusster 2	11942	11947			
	Clusster 3	12275	12281			
	Total	39988				

PART II : Customer Segmentation

Clusters for Effective Marketing

We can see that from Iteration for Optimization of clusters we will get three clusters for different category

- Cluster 1 belongs to Youth Category and has an income more than Senior. Category People and Phone Purchasing probability 0.56% cut off is 0.50% then they buy New. Phone
- Cluster 2 This Cluster Belongs to Old Category and has Least Income. Category People and Phone Purchasing probability 0.51% cut off is 0.50% then they buy New. Phone
- Cluster 3 This Cluster Belongs to Senior Category and has Superior Income. Category People and Phone Purchasing probability 0.66% cut off is 0.50% then they buy New. Phone

Market Share by 40% in each Cluster			
Cluster 1	15771	6308.4	
Cluster 2	11942	4776.8	
Cluster 3	12275	4910	
Market Entry Condition		15995.2	>12000

Count of PURCHASE Gender Wise Distribution				
Row Labels	0	1	Grand Total	
25-39	16.88%	19.80%	36.67%	Youth Category
40-54	14.06%	22.52%	36.57%	Senior Category
55-69	13.36%	13.40%	26.76%	Old Category
Grand Total	44.29%	55.71%	100.00%	

0	Female
1	Male

Salary wise purchase 0 Not Purchase/1 Purchase				
Row Labels	0	1	Grand Total	
70089-320089	23.10%	23.48%	46.58%	Low Income
320089-570089	15.03%	25.07%	40.10%	Medium Income
570089-820089	4.29%	9.03%	13.32%	High income
Grand Total	42.42%	57.58%	100.00%	

PART III: Business Decision

Expected Sales and Revenue

- From Indian Market data Set we have interpreted some data points that are as follows.

- For the Indian data set most of the persons to use phones up to four years so as per the analysis we have performed on the Chinese Data Set, it is predicted that they have exceeded the limit to using the phone so they are willing to buy a new phone.
- As per the Indian Data Set Annual Income of People belonging to Different Categories having high ratio than in the Chinese data set.
- As per the Analysis Market Entry Condition is fulfilled for the given data set.
- Sum of 40 % Shares of All Clusters (15995>12000) [Market Entry Condition for Indian Market] [Total Unit Sold >12000]

Salary Segmentation	Average of phn_age_years	Count of ID	40 % of each grp
300033-1050032	4.128583248	30077	12030.8
1050033-1800032	4.129819383	34936	13974.4
1800033-2550032	4.126192307	4987	1994.8
Grand Total	**4.129029848**	**70000**	28000

Age Description	Average of phn_age_years	Average of ANN_INCOME
25-39	4.12602942	1045438
40-54	4.125059094	1256606
55-69	4.138538363	1143048
Grand Total	**4.129029848**	**1148678.519**

Market Share by 40% in each Cluster			
Cluster 1	15771	6308.4	
Cluster 2	11942	4776.8	
Cluster 3	12275	4910	
Market Entry Condition		15995.2	>12000

PART III: Business Decision

Expected Sales and Revenue

- The expected revenue that the company can generate under each cluster
- Cluster 1- [119859600] Rs Approx.
- Cluster 2- [90759200] Rs Approx.
- Cluster 3- [93290000] Rs Approx.
- Market Entry Condition for Company Establishment is Sales Should be Grater than 12000 units.
- Total Revenue generated for withstanding in the market is also fulfilling the condition.
- 2. The final decision is that company should enter the Indian market with business justification as per the Classification Model

Market Share by 40% in each Cluster		
Cluster 1	15771	6308.4
Cluster 2	11942	4776.8
Cluster 3	12275	4910
Market Entry Condition		15995.2

If Average price for cell-phone is lie bw 19000-21000 then Revenue Generation in Rs.					
Market Share by 40% in each Cluster		40% of each	Revenue	Total Revenue Condition (Revenue> 200000000	TRUE
Cluster 1	15771	6308.4	119859600		
Cluster 2	11942	4776.8	90759200		
Cluster 3	12275	4910	93290000	303908800	
Market Entry Condition		15995.2	>12000	Total revenue Approx ↑↑	