Market Entry Analysis Project

-SUBMITTED BY

ARNAB MONDAL

PROBLEM STATEMENT :

- This Project is based on a mobile company 'XYZ Mobiles', a fictional China-based mobile company .
- XYZ Mobiles believes that the Indian market is very similar to China, in which the company currently operates.
- Before entering the new market, the company wants to be sure that the whole process will be profitable for them.

Hence, we are given the task to check for the following conditions that must be fulfilled in the Indian market for the company to enter:

- **Sale of a minimum of 12,000 phones** over the sample data in one year .
- **Collection of at least Rs. 20 crores** over the sample data in one year .

T1: JUSTIFICATION DURING MODEL DEVELOPMENT :

- Gender was classified into binary data as male (1) and female(0). The annual income was converted into INR for matching the situation of Indian currency.
- From the Chinese customer data it is clear that purchase decision depends on 4 factors. Customer age, Gender, last phone life, and annual income. The following changes were made on the raw data
- The phone life was classified into 4 category as below:

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

- Then after dataset is converted into training and test set with 70:30 Rule and then K-Means Clustering Modelling Is Performed On both of them.
- And then ROC Curve, Beta Values and Conversion Matrix (Including Accuracy, Sensitivity, Precision etc.)Is Computed from train and test data through K-means Clustering.

T1: PIVOT ANALYSIS AFTER CLEANING OF DATASET:

- **Table: 1**

GENDER	SUM OF PURCHASE	COUNT OF LEAD	CONVERSION RATE
0	9836	17715	55.52
1	13195	22285	59.21

- **Table: 2**

PHONE AGE CATEGORY			
PHONE AGE CATEGORY	SUM OF PURCHASE	COUNT OF LEAD	CONVERSION RATE
1	2351	6459	36.40
2	7023	16545	42.45
3	9208	11697	78.72
4	4449	5299	83.96

- **Note :**

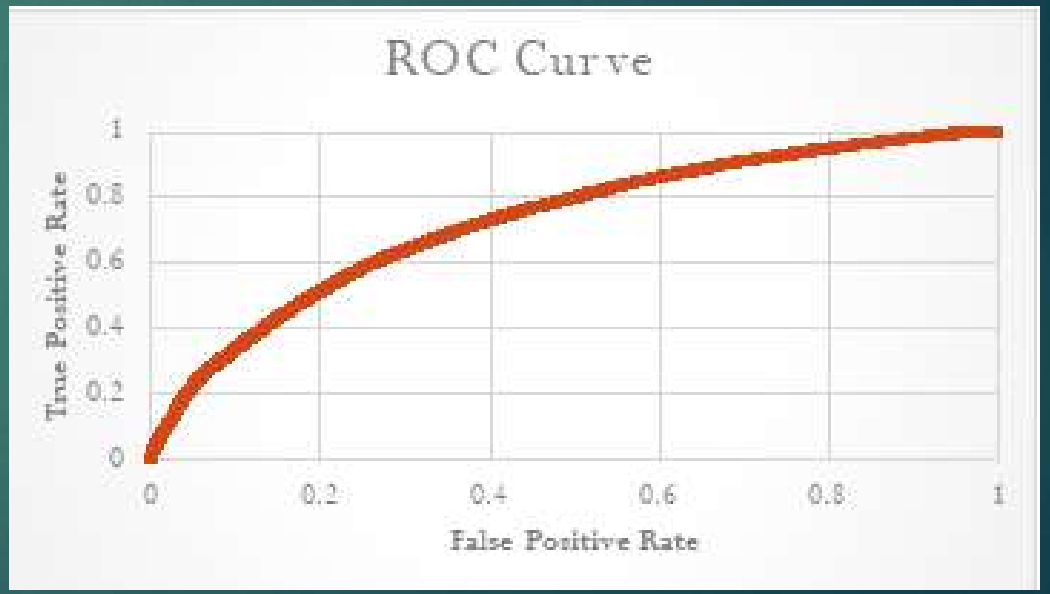
1. From Table 1 we came to know that conversion rate for Males is High As Compared to Females that is, **59.21 %.**
2. From Table 2 we can analyze that category 4 (i.e., Phone Age > 500) is having highest Conversion Rate that is, **83.96 %.**

T1: CLASSIFICATION MODEL BASED ON CHINESE DATASET:

- The logistic regression was made on the formatted data by considering the factors like age, gender, income, phone life and purchase.
- The coefficients B0 to B4 and ROC Curve of training Data are computed as follows:

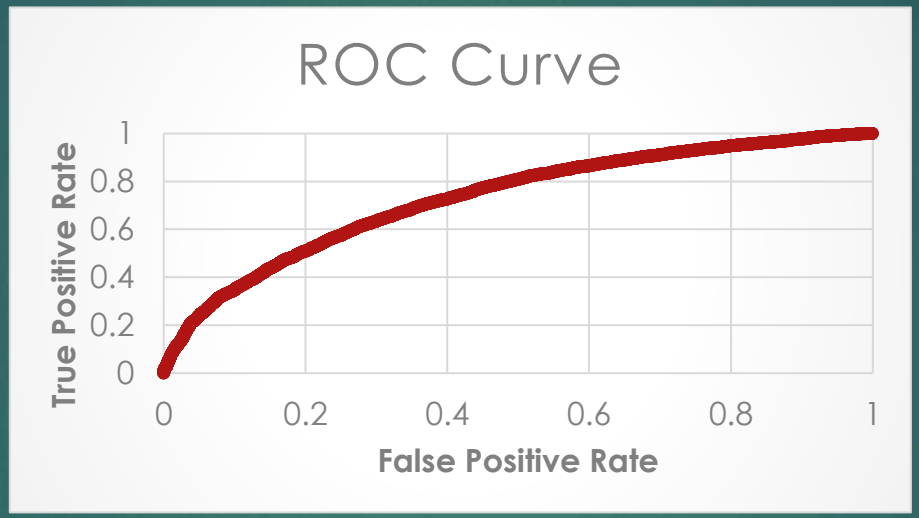
Coefficients	
B0	-1.518413636915270
B1	-0.01185574166408 6
B2	0.217181358351312
B3	0.000002250497046
B4	0.004185135548266

Conversion Matrix		
Actual	Predicted	
	Class 0	Class 1
Class 0	6381	5585
Class 1	3612	12422
Accuracy	0.672	
Precision	0.690	
Recall	0.775	
F1-Score	0.730	
Sensitivity	0.775	
Specificity	0.533	
TPR	0.775	
FPR	0.467	



T1: CLASSIFICATION MODEL BASED ON CHINESE DATASET AND METRICS ASSOCIATED:

- The coefficients B0 to B4 and ROC Curve of test Data are computed as follows:



- From the coefficient it is clear that age has a negative impact. Old people are less likely to buy a phone.
- Gender plays a significant role, which is evident in the pivot analysis.
- The annual income plays an insignificant role as the coefficient is very small.
- The phone life also plays an important role and the categorial variation has an impact of its own.

T1: COUNT OF POTENTIAL CUSTOMERS IN INDIA BASED ON MODEL:

- The data set is formatted such that gender is converted into a binomial model and the phone age is calculated by considering the purchase date as 1st July 2019.
- The phone life is converted into 4 categories as follow:

DAYS	SEGMENT
<200	1
200-360	2
360-500	3
>500	4

- The probability is computed based on coefficients (B0 to B4) obtained from the Chinese dataset and the no of potential customers in India based on a cut-off 0.5 is 31573 with a conversion ratio of 45.10%.

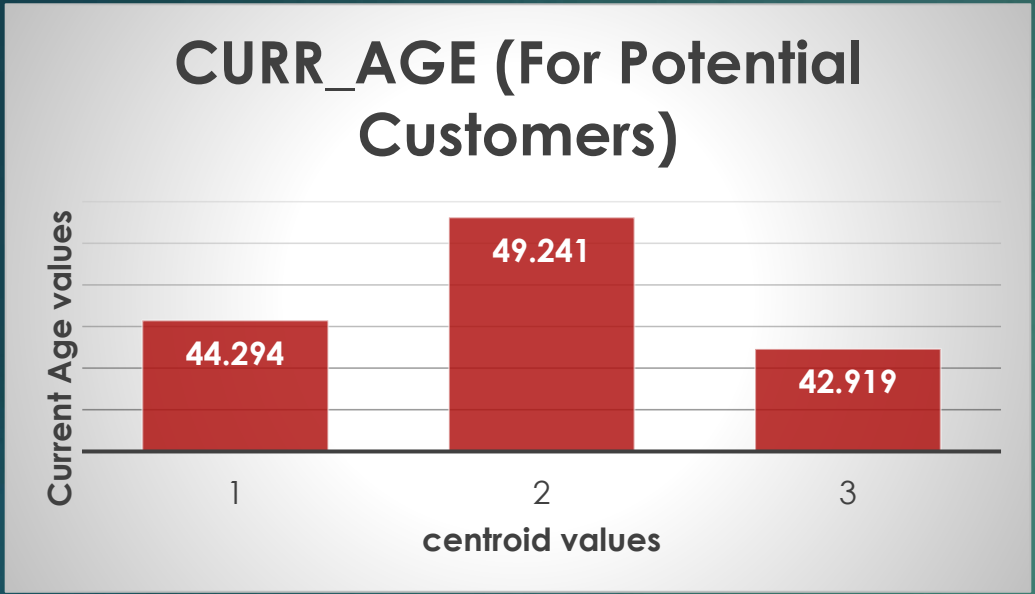
T2: JUSTIFICATION DURING CLUSTERING :

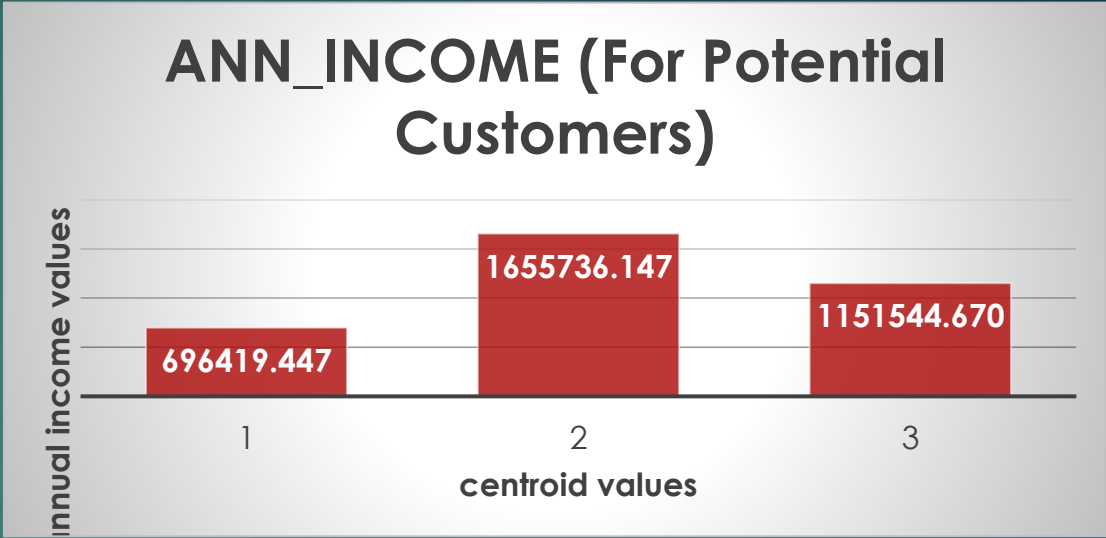
- Clustering is Performed on 3 & 4 Clusters and their error terms (For more scaled and standardized data) is also found and with that centroid values were generated.
- In our analysis 3 clusters results were taken for further analysis and predicting results and then EDA is performed on each cluster & Centroid values and following results were obtained:

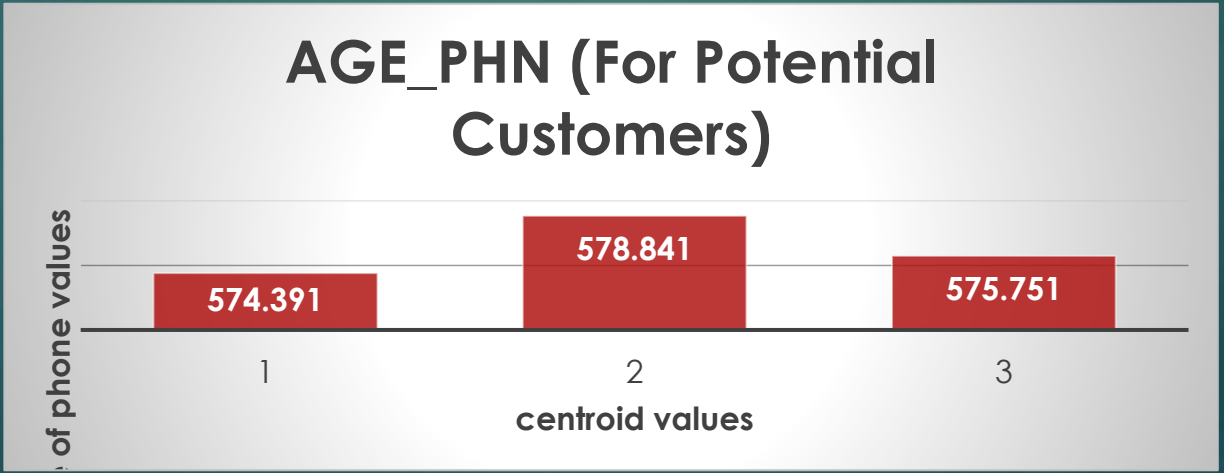
Centroid values (For Potential Customers)			
	1	2	3
CURR_AGE	44.294	49.241	42.919
GENDER	0.437	0.610	0.477
ANN_INCOME	696419.447	1655736.147	1151544.670
AGE_PHN	574.391	578.841	575.751
Centroid values (Error Terms)			
	1	2	3
CURR_AGE	0.460	-1.212	0.525
GENDER	0.998	-0.072	-1.002
ANN_INCOME	0.343	-0.539	0.086
AGE_PHN	-0.002	-0.004	0.005

- From the table above, **centroid value 2** is to be taken for business decision and It has error terms in negative as well which is good **so most peoples were clustered (i.e., their centroid) around with current age of 49, Annual income around 1655736 and age of phone around 579 days is to target For** through our Clustering analysis and EDA analysis of that is also done in further slides.

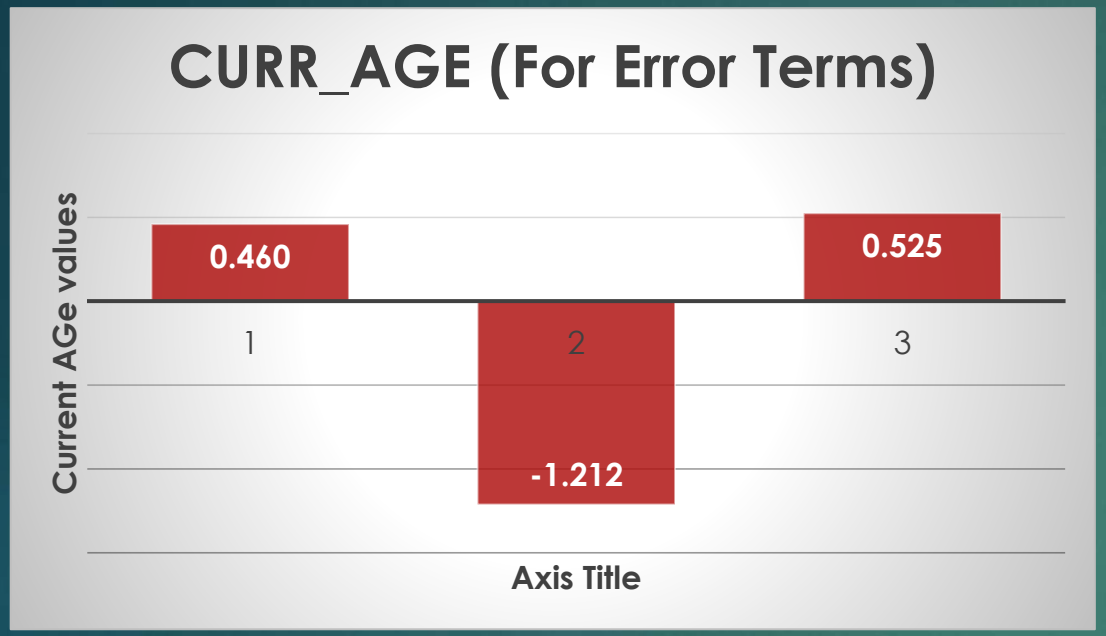
T2 : EDA ANALYSIS :

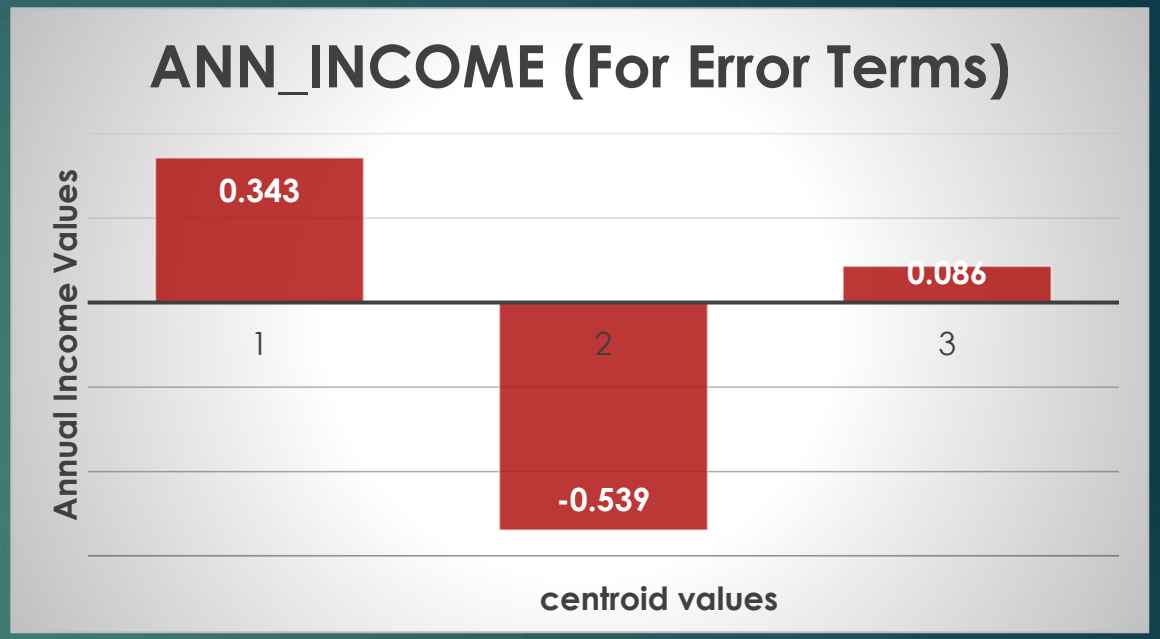


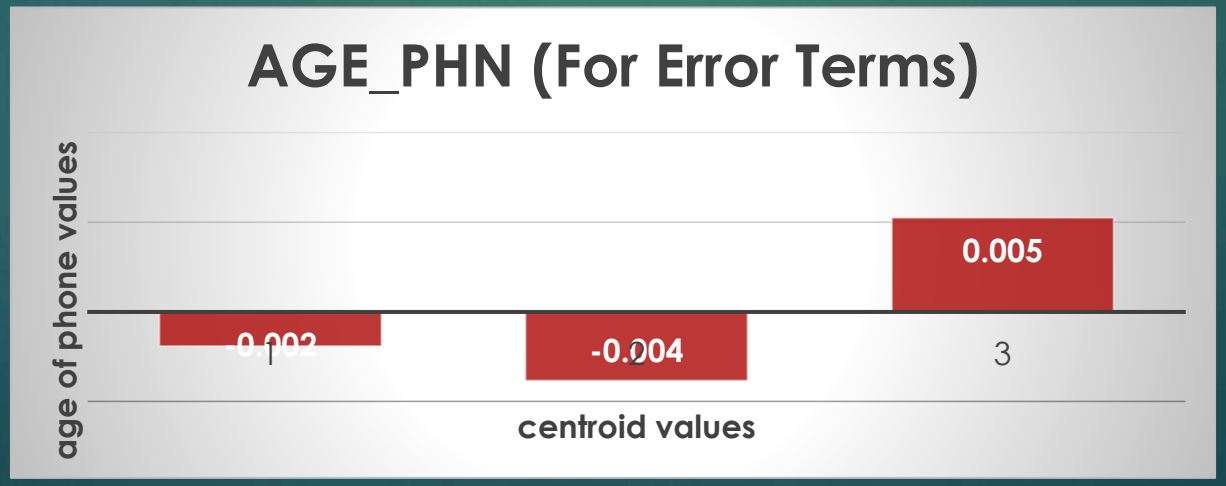




T2 : EDA ANALYSIS :







T3: JUSTIFICATION FOR FINAL RESULTS:

- Now we need to segment our customers through their ages to check for our final results i.e.,

	Segments	Age Criteria
Young Age	1	25-35
Mid Age	2	35-55
Old Age	3	55-65

- **Sale of a minimum of 12,000 phones** over the sample data in one year .
- **Collection of at least Rs. 20 crores** over the sample data in one year .

T3: JUSTIFICATION FOR FINAL RESULTS:

- From the table below we can Conclude that **revenue collection is way over Rs. 20 Crores for all 3 segment customers (that is Young, Mid and Old):**

Young Aged Customers			
	Average Revenue	Expected Revenue	IF >20 CRORE
Low Income	<8000	252384000	YES
Medium Income	10000	315730000	YES
High Income	12500	394662500	YES
	Average	320992167	

Mid Aged Customers			
	Average Revenue	Expected Revenue	IF >20 CRORE
Low Income	12500	394662500	YES
Medium Income	22375	706445875	YES
High Income	31250	986656250	YES
	Average	695921542	

Old Aged Customers			
	Average Revenue	Expected Revenue	IF >20 CRORE
Low Income	9500	299943500	YES
Medium Income	22375	706445875	YES
High Income	31250	986656250	YES
	Average	664348541.7	

T3: JUSTIFICATION FOR FINAL RESULTS:

- From the Table Below we can conclude that Customer Age Segment 2 ,**count of purchase is 15283 which is way over our required result that is minimum 12000 Phones, sum of their ages is very high as 30566 and their Conversion rate is also high of 47.67 % in INDIA** as compared to other segments so XYZ company must target Middle aged customers and so, **yes they can enter in INDIAN market as there is no loss for XYZ Mobile Company.**

Customer Age Segments	Count of Purchase	Sum of Customers Age Segmented	Conversion Rate
1	7660	7660	11.95
2	15283	30566	47.67
3	8630	25890	40.38

THANK YOU