PART:1

Steps taken during the project

STEPS TAKEN DURING THE PROJECT

- First, I converted the Chinese currency into Indian by multiplying the Chinese currency by 10.

- Then the variables like curr_age, age_phone and annual income(in Rs) were segmented.

- Then gender was changes from a categorical variable to a numerical variable.

- After all these steps I performed an exploratory data analysis using pivot tables just to get an idea about the dataset.

- Then the data was divided in ration 70:30 as training and test dataset respectively.

- Then a classification model was built using the training dataset and taking 0.5 as a random cut-off.

STEPS TAKEN DURING THE PROJECT

- Then using the obtained information a table was formed between various cut-offs and their metrices(accuracy, sensitivity and specificity).

- 0.5 was found as the optimum cut off.

- Then these coefficients and the cut-off of 0.5 was used on test dataset.

- Then the metrices of test and training data set were compared.

- After the successful evaluation of the model it was applied to the Indian data set to find out the number of potential customers.

- Then the data of potential customers was scaled(the scaled variables were age_phone, curr_age and annual salary).

STEPS TAKEN DURING THE PROJECT

- Then the potential customers were put into clusters of 3 and 4(k=3 and k=4) using k-means method in excel.

- Then the clusters were compared, and it was found that the clusters in K=4 were more discrete and hence more useful.

- Then a detailed EDA was done on all four clusters.

- Then the clusters were matched to various branches of the given flow chart in order to find the budget of each cluster.

- Then the formula (expected revenue=expected units sold*price per unit) the expected revenue was calculated

- Based on the given conditions and obtained results a business decision was made.

PART-2

Analysis

TASK 1: POTENTIAL CUSTOMERS IN INDIA

- A classification model is built after modifying the given dataset(like changing currency, segmentation of data etc.) with a cut-off of 0.5.

-

b0	-2.1896152
b1	-0.0111879
b2	-0.2256097
b3	0.96952081
b4	2.3245E-07

- These are the coefficients obtained. Here, x1=curr_age, x2=gender, x3=segmented_age_phone, x4=annual income(in Rs).

TASK 1: POTENTIAL CUSTOMERS IN INDIA

- So, by looking at the coefficients we can interpret that :

1. If the age of person increases(keeping other variables constant) probability of them buying the phone decreases.

2. Males have more probability of buying phone than females (keeping other variables constant).

3. As the age of phone increases the probability of buying new phone increases (keeping other variables constant).

4. As the annual income of a person increases probability of buying new phone increases (keeping other variables constant).

TASK 1: POTENTIAL CUSTOMERS IN INDIA

- The table on the left represents the accuracy, sensitivity and specificity of various cut offs.

- As we can see 0.5 and 0.6 cut-offs have close values in all three metrics.

- But the sensitivity in 0.5 is more.

- Since our business decision is based on the count of potential customers so I think we should prioritize sensitivity to have maximum true positives.

- So, *0.5 is our optimum cut-off* .

Probability cut off	Accuracy	Senstivity	Specificity
0.1	0.57	1.00	0.00
0.2	0.59	0.99	0.05
0.3	0.62	0.94	0.20
0.4	0.65	0.89	0.33
0.5	0.68	0.75	0.58
0.6	0.68	0.63	0.75
0.7	0.63	0.46	0.87
0.8	0.54	0.23	0.96
0.9	0.46	0.06	0.99
1	0.43	0.00	1.00

TASK 1: POTENTIAL CUSTOMERS IN INDIA

Test dataset	
Accuracy	0.68
Sensitivity	0.76
Specificity	0.58

| Training dataset	
accuracy	0.68
Sensitivity	0.75
Specificity	0.58

TASK 1: POTENTIAL CUSTOMERS IN INDIA

- As we can see in the previous slide our training dataset metrics matches very well with the test data set metrics so, hence the evaluation of the model is successful.

- Now after applying the model on Indian dataset we get a total of *31849* potential customers.
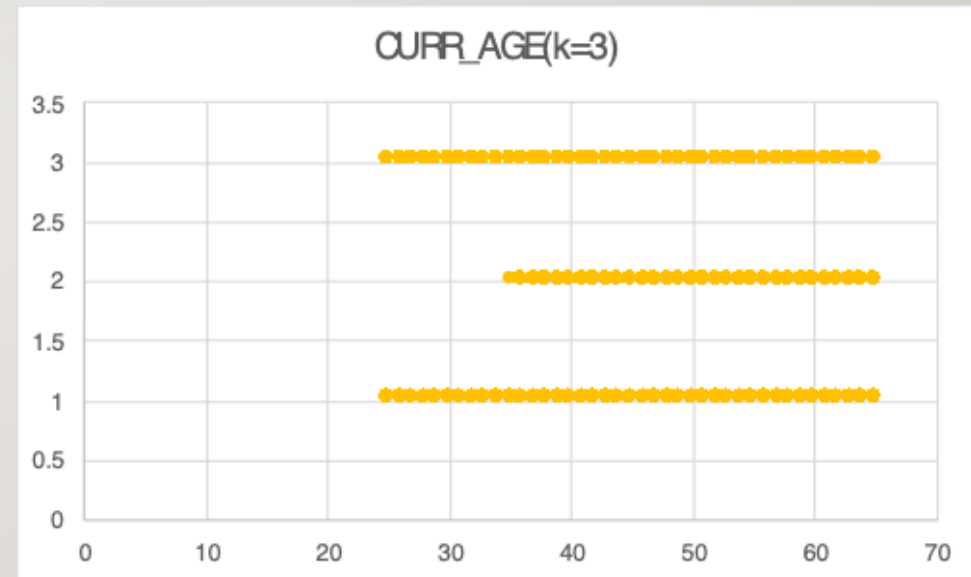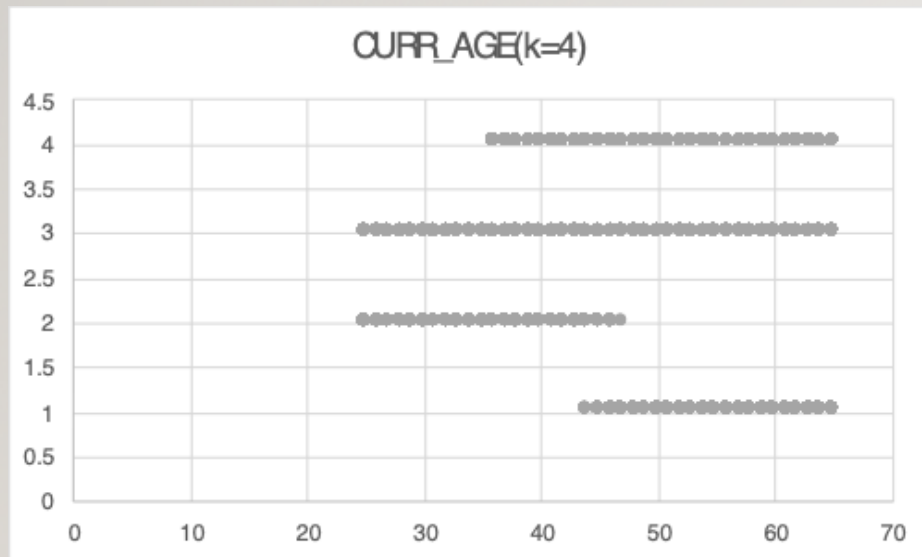
TASK 2: CUSTOMER SEGMENTATION

- Here are a few scatter plots of different variables at different values of k(3 and 4).
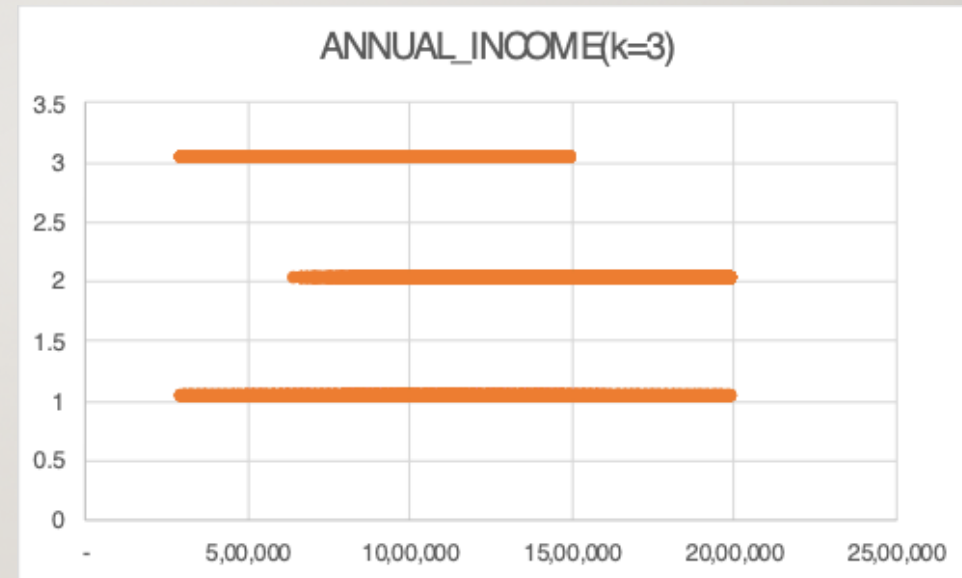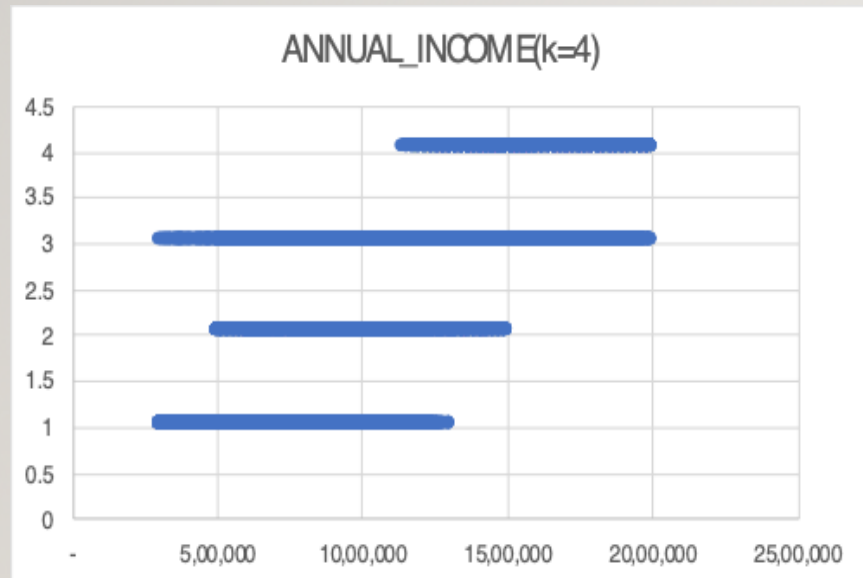
TASK 2: CUSTOMER SEGMENTATION
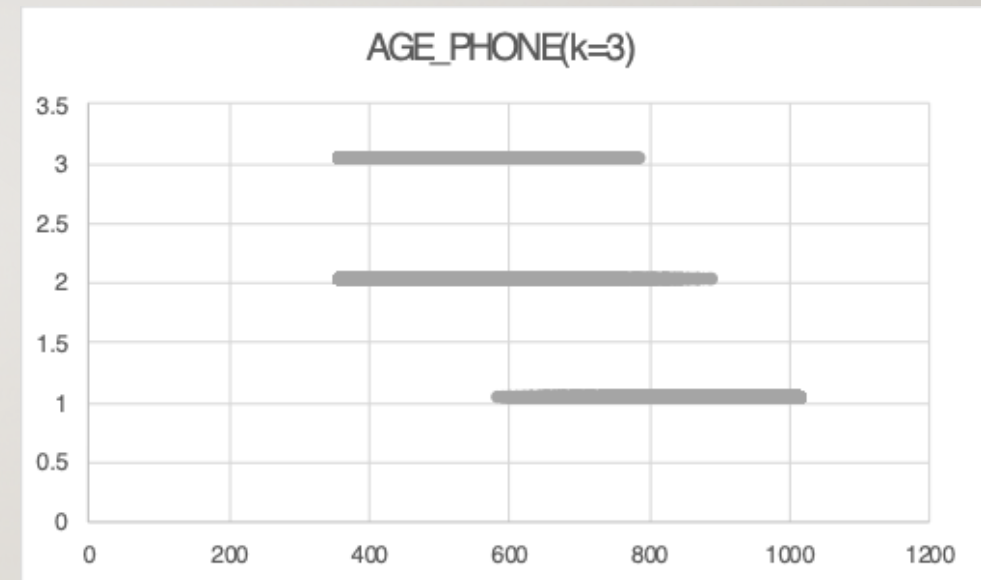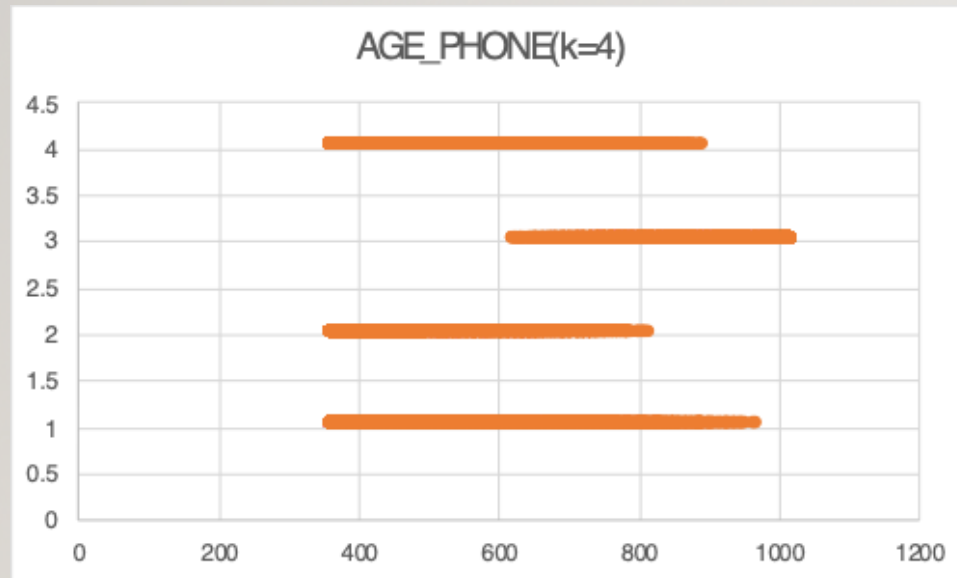
TASK 2: CUSTOMER SEGMENTATION

- SO AFTER SEEING THESE SCATTER PLOTS, WE CAN SAY THAT K=4 IS A BETTER CHOICE BECAUSE THE CLUSTERS ARE MORE DISCRETE WHEN WE TAKE K=4 THAN WHEN K=3.

TASK 2: CUSTOMER SEGMENTATION

- **TRAITS OF EACH CLUSTER:**

- *CLUSTER 1:*

1. Majority of its people are of age between 50-60.

2. Most of them have an annual income between 800000-1200000.

3. The age of phone is mostly 360-500 days.

4. It majorly includes old people with low annual income and with age of phone less than 9 months.

- *CLUSTER 2:*

1. Majority of its people are of age between 30-40 years.

2. Most of them have an annual income between 800000-1200000.

3. The age of phone is mostly 360-500 days.

4. It majorly comprises of young people with low/medium annual income.

TASK 2: CUSTOMER SEGMENTATION

- **_CLUSTER 3:_**

1. Majority of its people are of age between 30-50.

2. Most of them have an annual income between 800000-1500000.

3. The age of phone is mostly more than 500 days.

4. It majorly comprises of middle-aged people with medium annual income and age of phone more than 1.25 years.

- **_CLUSTER 4:_**

1. Majority of its people are of age between 40-60.

2. Most of them have an annual income between 1300000-1600000.

3. The age of phone is mostly 360-500 days.

4. It majorly comprises of old people with high income and with age of phone more than 1 year.

TASK 3: BUSINESS DECISION

- Here is a table depicting expected revenue under each cluster:

CLUSTERS	PRICE PER UNIT	EXPECTED UNITS SOLD	EXPECTED REVENUE
1	9000	5854	52686000
2	10000	10095	100950000
3	20500	7195	147497500
4	32500	8705	282912500
TOTAL		31849	₹58,40,46,000

TASK 3: BUSINESS DECISION

- As it can be seen the expected revenue is approximately *58 crores* which is more than the required expected revenue(20 crores).

- Also, the expected units to be sold are *31850* which is also more than the requirement(12000) .

- Therefore, *it is advisable for the company to enter the Indian market* as it has a lot of potential customers and high expected revenue.