ASSIGNMENT – ACQUSITION ANALYTICS

SUBMITTED BY –Simran Jain

Dated – 27th August 2020

Business Objective:

- Objective is to predict the probability of a response from each prospect and target the ones most likely to respond to the next telemarketing campaign. The steps were as follows:

1. Identify relevant predictor variables for a response using EDA.
2. Build predictive models and choose the best one.
3. Sort the prospects in order of decreasing probability of response (predicted by the best model) and target the top X% (or top Y deciles), where X would be determined by your business objective (e.g., maximising the overall response rate/number of responders at a fixed marketing cost).

4.To solve these problems, you should resolve to build another model without including the variable 'duration'. This will help you understand the relationship of the other variables with the response.

Checkpoints:-

The checkpoints for the assignment are as follows:

1. Perform data preparation (no marks are awarded for this step)
    1. You can use the code provided in the lectures to complete all the data preparation steps
2. Build a logistic regression model without using the variable 'duration'
    1. Select variables using the usual methods
    2. Sort the data points in decreasing order of the probability of response
    3. Find the optimal probability cut-off and report the relevant evaluation metrics
3. Create a data frame with the variables prospect ID, actual response, predicted response, predicted probability of response, duration of the call in seconds and cost of the call
    1. While creating the data frame, list the cost of call for each prospect in a new column
4. Find the number of top X% prospects you should target to meet the business objective
    1. Report the average call duration for targeting the top X% prospects to the CMO
5. Create a lift chart
    1. The x-axis should show the number of prospects contacted; the y-axis should show the ratio of the response rate using the model and the response rate without using the model
6. Determine the cost of acquisition
    1. Consider cost = 1*number of contacts made in the current campaign; determine the cost incurred for acquiring 80% of customers using the predictive model
7. Create a small presentation for the CMO highlighting your findings and the methodology used.

Methodology used:

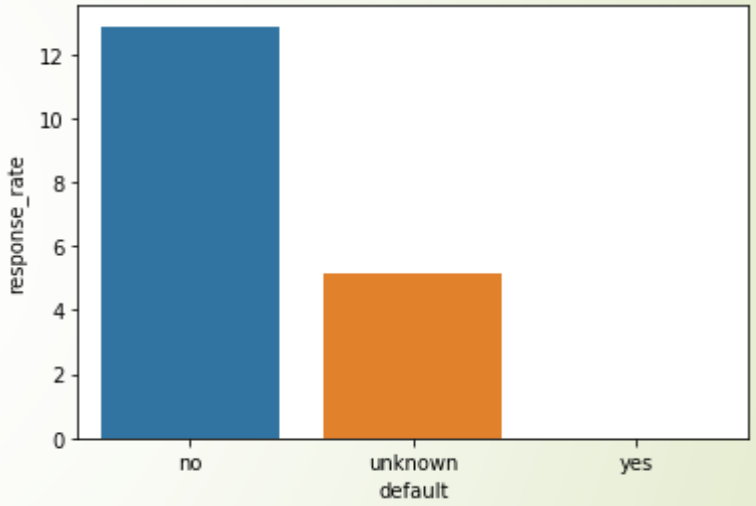
- 1. Read and understood the data
- 2. Cleaning of the data that is removing all the unnecessary variables from the data frame.
- 3.Performed Exploratory data analysis that is understanding about the outliers and how does the data behave .
- 4. Splitting the data to test and train dataset that is test data was 30% of the entire data while train data was 70% of the entire data.
- 5. Standardization of the test as well as train data .
- 6. Building models:
  - Using Logistic regression
  - Logistic regression using pca

  Last is we check the data with accuracy , specificity , sensitivity.

  We then created lift chart and did calculate cost of acquisition

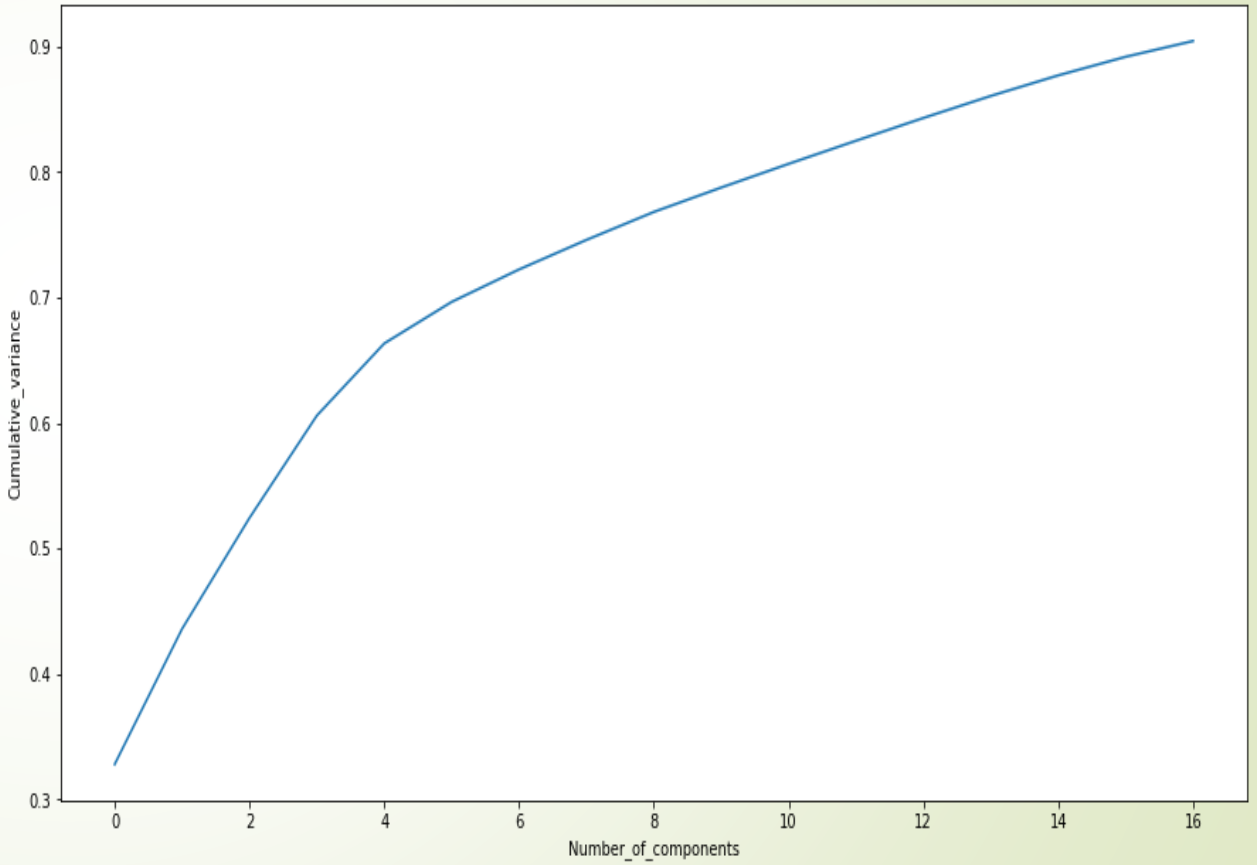
Model Building using logistic regression:-

- Using logistic regression we found out that our data has a lot of insignificant variables that needs to be removed.
- Thus tuning the model will help in dropping the variables.
- Class has imbalance thus to deal with imbalance we tried first removing outliers then dealt with it.
- We used logistic regression with pca

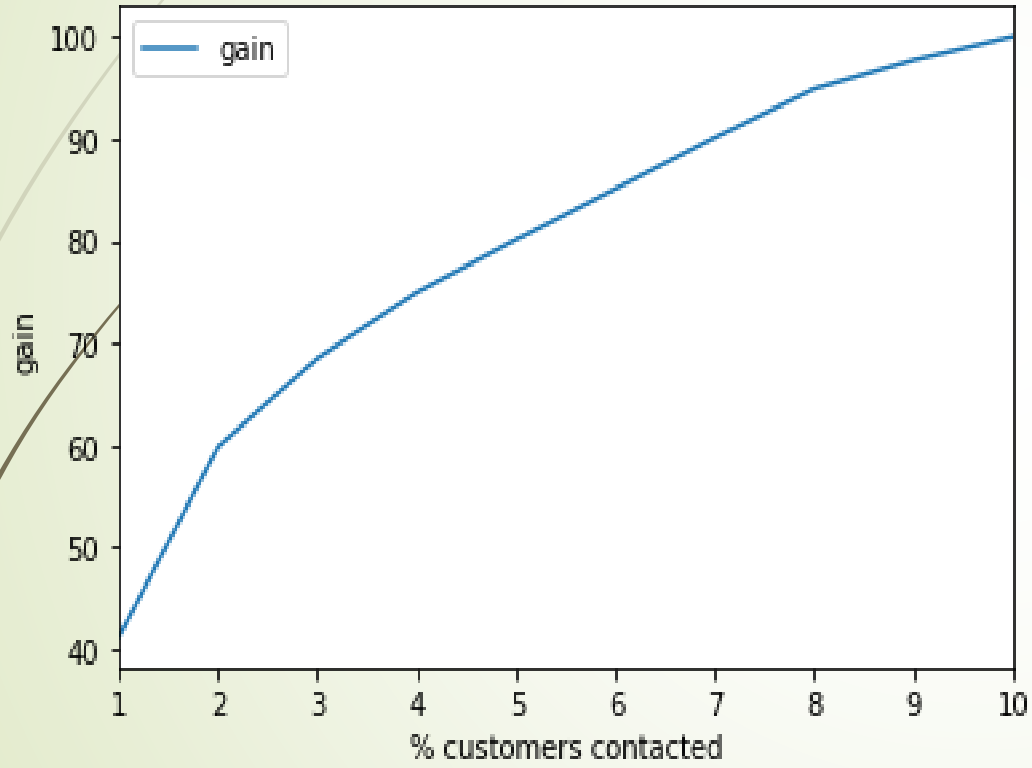


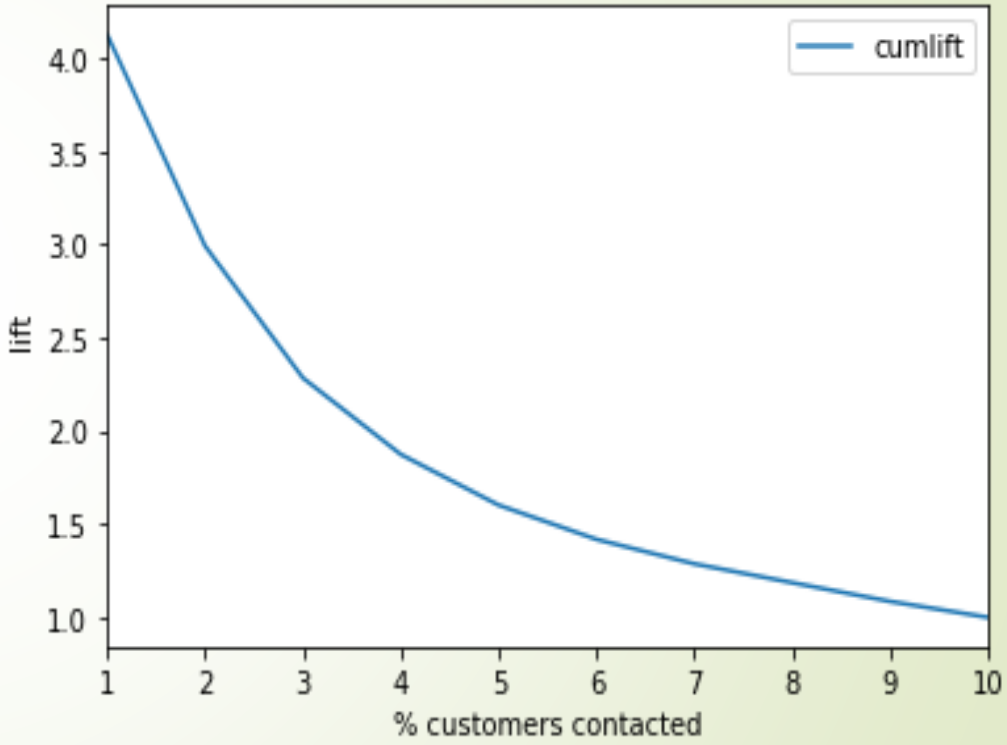
Scree Plot (No_of_components v/s cumulative_frequency):

1.Thus we see 17 variable explain 90% variance in the data.
2.ROC_AUC score: 0.7789
3.Threshold: 0.3930
4.Accuracy Score on test data: 0.808
5.Sensitivity: 0.71201
6.Specificity: 0.7077
7.false postive rate: 0.292
8. positive predictive value: 0.2347
9.Negative predictive value: 0.9512
10.Misclassification Rate: 0.291



Gain chart v/s Lift chart:-





Cost of acquisition:-

- Determine the cost of acquisition
- Consider
- cost = 1*number of contacts made in the current campaign;
- determine the cost incurred for acquiring 80% of customers using the predictive model
- As per the formula given above, the cost will be equal to
- cost = 1* (50 % of 41188) = 20594
- Consider average duration of call and cost of the call based on data available, then acquisition cost will be equal to
- cost = (duration of call) (cost of call) (number of contacts made)
- If we consider cost based on formula given to us, than we have reduced cost by 50% as we have to contact only 50 % of prospects to get 80 percent conversion.

Conclusion

- • To achieve our business objective of acquiring 80% of total responders at the minimum possible cost;
- we will need to target 50% of the total customer base for entire dataset. In case of test data, it is 60% of the test dataset.
- • Significant predictor variables identified by model:
- - job_retired
- - month_mar
- - poutcome_success
- - job_student
- - month_may
- - cons.price.idx
- - contact_telephone
- - previous_Nevercontacted
- - euribor3m
- • Through our model we have improved 50% efficiency; as instead of calling the entire customer base, we can now achieve our objective by targeting just 50% of the entire customer base.