1. Data cleaning Process

The sentences covered by yellow are several “dangerous” steps. They can change the data.

1. Combine two data sets together to get a new data set named as “combined\_data”

The two data sets are the original excel file you give me, one has the details information of Organization such as “Cam\_Cause\_” , another has the details of campaign.

2. Only several columns are useful for us, so just extract the specific columns of combined\_data into extract\_data

Here are all the useful columns:

Text

Description automatically generated

3. By visualizing the data we found "Impact\_Message3" "Impact\_Message4" and "Impact\_Message5" etc are many missing data. So fill them with “0”. That is mainly because blank cells easily cause some unexpected problems during programming.

This step fills all the blank cell with “0”

Text

Description automatically generated

4. Calculate the Number of deductibility

5. Calculate the numbers of "organization causes" and "campaign causes”

6. Calculate the numbers of images

7. Change the format of date and calculate the duration of campaign

8. Delete the rows where “duration\_day” less than 1

9. Delete the rows where “Actual\_Donation\_Amount” is 0 but “Distinct\_Donors” bigger than 0

10. Classify the “Impact\_Message” into 3 categories

11. Calculate Polarity and subjectivity of every Impact\_Message

12. Calculate the number of words of Num\_desc\_cam and Num\_desc\_NPO(Already fill the blank cell with 0)

13. Convert “NPO\_Tax\_Deductibility” and “duration\_day” data type into numeric

14. Calculate the “correlations and Variance inflation factor”

15. Calculate “Variance, Average, Max, Min, Median”

16. Do regression

2. Data cleaning Process

1. Can you show me examples where, you have campaigns that Words of Campaign =1 and Words of NPO =1

Actually, all the campaigns whose Words of Campaign = 1 are the blank cell initially. Because they are filled into “0” just as the above mentioned.

2. Also who me the example where Avg Customer Amount = 900,000? I want to see that it cannot be through though.

**Table

Description automatically generated**Here it is, show we do more operation on it

0 – no impact message

1-           same thank you

2-            same thank you but with different amount

3-            thank you with different amount  and explanation of how the donations will be used

Table

Description automatically generated

I think it is ok. You can check it.

3. New Model

Table1 correlation Matrix and descriptive statistics

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  |  | Mean | Std.Dev | Min | Max | [1] | [2] | [3] | [4] | [5] | [6] | [7] | [8] | [9] | [10] | [11] | [12] |
| [1] | DV:Donation Amount | 9813 | 62978 | 0 | 3.4\*10^6 | 1 |  |  |  |  |  |  |  |  |  |  |  |
| [2] | Campaign Goal | 44797 | 154417 | 100 | 5\*10^6 | 0.48 | 1 |  |  |  |  |  |  |  |  |  |  |
| [3] | Tax Deductibility | 0.939 | 0.240 | 0 | 1 | 0.04 | -0.02 | 1 |  |  |  |  |  |  |  |  |  |
| [4] | Video use | 0.404 | 0.491 | 0 | 1 | -0.14 | -0.04 | 0.015 | 1 |  |  |  |  |  |  |  |  |
| [5] | Impact message | 1.183 | 1.205 | 0 | 3 | 0.15 | 0.49 | -0.03 | -0.041 | 1 |  |  |  |  |  |  |  |
| [6] | Number of images | 2.870 | 1.589 | 0 | 5 | 0.19 | 0.076 | -0.059 | -0.076 | 0.013 | 1 |  |  |  |  |  |  |
| [7] | Words of campaign | 138 | 67.5 | 1 | 309 | 0.18 | 0.2 | -0.026 | -0.0003 | 0.26 | 0.082 | 1 |  |  |  |  |  |
| [8] | Words of NPO | 101 | 56.1 | 1 | 179 | 0.41 | 0.064 | 0.069 | -0.1 | 0.002 | 0.18 | 0.054 | 1 |  |  |  |  |
| [9] | Organization Cause | 3.033 | 1.536 | 0 | 5 | 0.45 | 0.057 | 0.017 | -0.13 | -0.044 | 0.16 | 0.049 | 0.52 | 1 |  |  |  |
| [10] | Campaign Cause | 3.456 | 0.982 | 0 | 4 | -0.08 | -0.046 | 0.069 | 0.086 | 0.024 | -0.12 | 0.022 | -0.043 | 0.27 | 1 |  |  |
| [11] | Description\_Campaign\_polarity | 0.165 | 0.155 | -0.8 | 1 | -0.12 | -0.058 | 0.011 | 0.086 | 0.081 | -0.078 | 0.034 | -0.1 | -0.13 | 0.05 | 1 |  |
| [12] | Description\_Campaign\_subjectivity | 0.422 | 0.143 | 0 | 1 | 0.12 | -0.003 | -0.029 | -0.083 | -0.009 | 0.11 | 0.095 | 0.13 | 0.14 | -0.04 | 0.35 | 1 |
| [13] | Avg\_Custom\_Amount | 243.40 | 7312 | 0.0 | 900000 | 0.13 | 0.5 | -0.034 | -0.014 | 0.86 | -0.043 | 0.24 | -0.084 | -0.11 | 0.02 | 0.095 | -0.044 | 1 |

The new regression result.

A picture containing text, receipt

Description automatically generated

The new category of impact message seems performs better!