**Just Do it**

**General Work flow**

1. Clean data

2. Fill data

2.1 Fill 0 into missing data

2.2 Calculate “duration” and add more columns of date

3. Create Sentiment Analysis variables

3.1 Calculate “polarity”\* and “subjectivity”\* of message1-message5

3.2 Calculate “similarity” \*of message1-message5

3.3 Calculate “distance”\* between message1-message5

4. Show statistics information of useful variables

5. Linear regression

**As for the detail, I will show you the codes and explanation.**

\*\* “Polarity”\* item is the positiveness of the text, which is a floating point number in the range of [-1.0, 1.0]

\*\* “Subjectivity” item is a subjective score, which is a floating point number in the range of [0.0, 1.0], where 0.0 is very objective and 1.0 is very subjective

\*\* “Similarity” item is a score, which in range of [0.0, 1.0]. 0 means this two sentences are totally different and 1 means there are the same.

\*\* “Distance” Between two strings, the minimum number of editing operations required to convert one into another, if the distance between them is greater, it means that they are more different

**Question & Problems**

1. In the email I notice that “NPC status  **( this one ignore first, something wrong with this entry. )”** So I add it in variable. I tried to add it in independent variables I found this variable is helpless to improve the R-Square. You can see the result in codes.
2. I used the “Leon Dataset 3 Nov 2022.xlsx” as data set. I think this data set have more information but there is no “Number of causes the campaign identifies with” in this data set. Could you tell me how to add these information into it or change data set?

1. Descriptive statistics

Table1 Descriptive statistics

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Descriptive statistics | | | | | | |
| statistics | Number | Median | Min | Max | Variance | Average |
| Actual\_Donation\_Amount | 15957 | 1300.0 | 0 | 3431670 | 4788325250 | 9813 |
| Donation\_per\_donor | 15957 | 84.3 | 0 | 163050 | 2492432 | 199 |
| Campaign\_Goal | 15957 | 5000 | 100 | 5000000 | 23847804493 | 44800 |
| Campaign\_Video | 15957 | 0 | 0 | 1 | 0.2409 | 0.4046 |
| Campaign\_Image\_num | 15957 | 3 | 0 | 5 | 2.524 | 2.868 |
| duration\_day | 15957 | 60 | 0 | 630 | 12089 | 107 |
| Msg1\_category | 15957 | 1 | 0 | 2 | 0.7603 | 0.8325 |
| Msg2\_category | 15957 | 1 | 0 | 2 | 0.7772 | 0.8326 |
| Msg3\_category | 15957 | 1 | 0 | 2 | 0.7762 | 0.8247 |
| Msg4\_category | 15957 | 1 | 0 | 2 | 0.7444 | 0.7708 |
| Msg5\_category | 15957 | 1 | 0 | 2 | 0.6741 | 0.7680 |

If need more info about this please tell directly.

2. Correlation matrix

2.1 Correlation between important variables and actual amount donation.

Table2 Correlations Matrix

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Correlations Matrix | | | | | | | | | | | | | |
|  | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 |
| 0.Actual Donation |  |  |  |  |  |  |  |  |  |  |  |  |  |
| 1.Campaign goal | 0.47 |  |  |  |  |  |  |  |  |  |  |  |  |
| 2.Duration | 0.31 | 0.43 |  |  |  |  |  |  |  |  |  |  |  |
| 3.Images Number | 0.19 | 0.079 | 0.11 |  |  |  |  |  |  |  |  |  |  |
| 4.Video | -0.14 | -0.04 | -0.05 | -0.08 |  |  |  |  |  |  |  |  |  |
| 5.impact message1 | 0.18 | 0.5 | 0.25 | -0.05 | -0.08 |  |  |  |  |  |  |  |  |
| 6.impact message2 | 0.19 | 0.51 | 0.26 | -0.055 | -0.08 | 0.97 |  |  |  |  |  |  |  |
| 7.impact message3 | 0.19 | 0.5 | 0.25 | -0.05 | -0.08 | 0.95 | 0.98 |  |  |  |  |  |  |
| 8.impact message4 | 0.17 | 0.48 | 0.23 | -0.034 | -0.06 | 0.89 | 0.92 | 0.93 |  |  |  |  |  |
| 9.impact message5 | 0.17 | 0.51 | 0.23 | -0.032 | -0.06 | 0.93 | 0.92 | 0.91 | 0.88 |  |  |  |  |
| 10.Total\_polarity | 0.12 | 0.22 | 0.21 | -0.065 | -0.06 | 0.44 | 0.45 | 0.44 | 0.38 | 0.42 |  |  |  |
| 11.Total\_subjectivity | 0.24 | 0.4 | -0.22 | 0.15 | -0.14 | 0.78 | 0.79 | 0.78 | 0.72 | 0.76 | 0.57 |  |  |
| 12.Total\_similarity | -0.14 | -0.45 | 0.23 | 0.024 | 0.034 | -0.85 | -0.83 | -0.82 | -0.73 | -0.8 | -0.45 | -0.66 |  |
| 13.Total\_distance | 0.16 | 0.46 | 0.23 | 0.0004 | -0.04 | 0.87 | 0.21 | 0.86 | 0.79 | 0.84 | 0.48 | 0.71 | -0.97 |

The video correlation turned out to be negative, which is unexpected totally. As for others, I think they are reasonable.

The same, pls tell me if need more.

2.2 Other correlations

In this picture I add some other control variables into it. You can take it as reference.

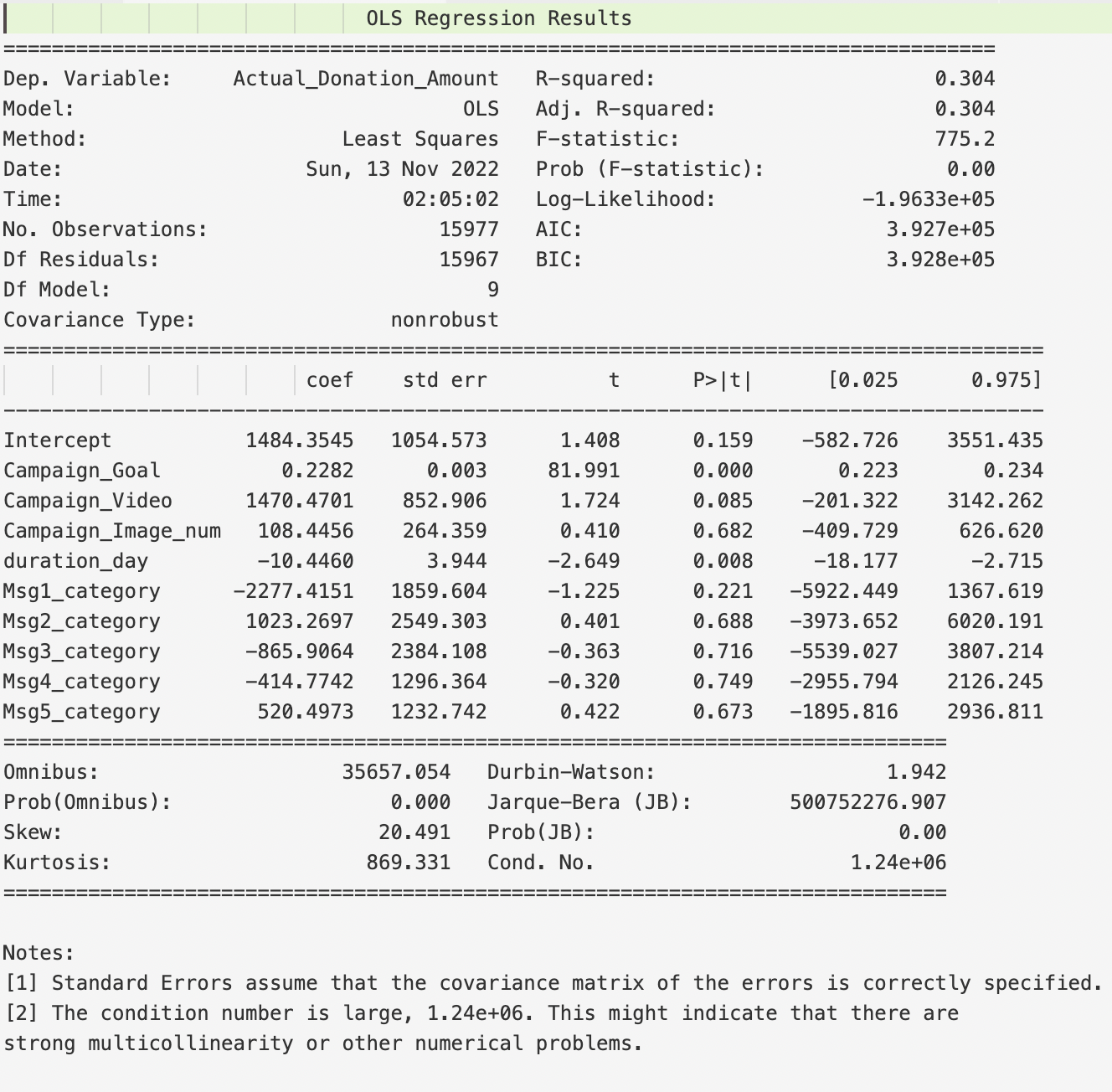
Chart

Description automatically generated with medium confidence

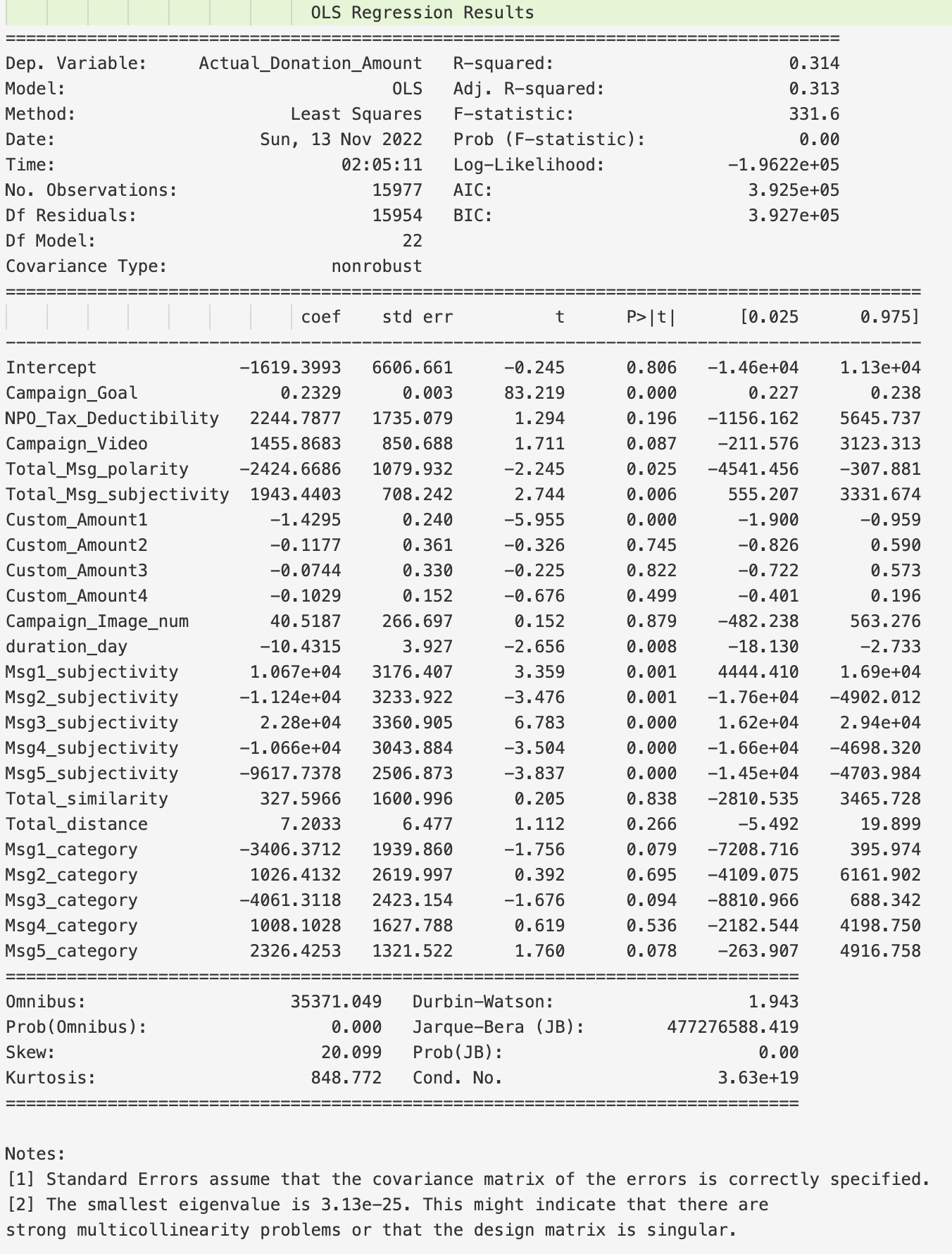
Figure1 Correlations Matrix

3. Linear Regression Result

3.1 The Result of selected independent variables



3.2 The Result of selected + other control variables



Actually, the additional variables improve the result a little.

3.3 The regression result of donation of per donor

The R-squared is very low, I want to make sure “donation of per donor = Amount donation/ Distinct Donors” right? If not please tell me.

