

# Logistic Regression Modeling (cleaned\_1-OCT-modelling.csv)

Ziyan Zhao

07/10/2020

## Attributes:

- 1: StageName (factor)
- 2: RecordType.Name 3: RICE\_Supported\_\_\_c
- 4: CreatedDate
- 5: Actual\_Close\_Date\_\_\_c (factor)
- 6: Lead\_Faculty\_\_\_c
- 7: Lead\_School\_\_\_c
- 8: Parent\_Opportunity\_\_\_c (factor)
- 9: RecordType.Name.1
- 10: Industry
- 11: Industry\_Sub\_Type\_\_\_c
- 12: Business\_Type\_\_\_c
- 13: Is\_External\_\_\_c (factor)
- 14: ParentId (factor)
- 15: CreatedYear
- 16: CreatedMonth

```
# Reading Data In
data <- read.csv("/Users/ChantelleChiu/Documents/GitHub/Project-90106-G28/Chantelle/cleaned_1-OCT-modelling.csv",
                 header = TRUE, stringsAsFactors = T)
names(data) <- c("X1", "X2", "X3", "X4", "X5", "X6", "X7", "X8", "X9", "X10",
                 "X11", "X12", "X13", "X14", "X15", "X16")

summary(data)
```

##	X1	X2
##	Min. :0.0000	Research Contract :2565
##	1st Qu.:0.0000	Custom Education (MSPACE Included):1105
##	Median :0.0000	Grants : 662
##	Mean :0.3737	Consultancy (Non-research) : 462

```

## 3rd Qu.:1.0000 Technology Transfer : 304
## Max. :1.0000 Competitive Bid : 89
## (Other) : 221
## X3 X4 X5
## Not supported : 413 2017-10-05T03:34:17.000Z: 6 Min. :0.0000
## RIC RE&D and BD&I:2649 2019-03-19T02:03:51.000Z: 6 1st Qu.:1.0000
## RIC-BD&I :2075 2019-05-29T02:49:22.000Z: 6 Median :1.0000
## RIC-RE&D : 271 2016-08-03T02:21:55.000Z: 5 Mean :0.9798
## 2017-07-20T06:46:12.000Z: 5 3rd Qu.:1.0000
## 2017-07-20T06:46:14.000Z: 5 Max. :1.0000
## (Other) :5375
## X6 X7 X8
## 0012e000002ZGfbAAG:1033 NotGiven :2397 Min. :0.00000
## 0012e000002ZLg7AAG: 874 0012e000002Zt0cAAC: 386 1st Qu.:0.00000
## NotGiven : 820 0012e000002Zt0mAAC: 368 Median :0.00000
## 0012e000002Zmo0AAC: 755 0012e000002Z3pPAAS: 190 Mean :0.09874
## 0012e000002ZmnxAAC: 598 0012e000002Zt10AAC: 190 3rd Qu.:0.00000
## 0012e000002ZLgJAAW: 272 0012e000002Zt0bAAC: 158 Max. :1.00000
## (Other) :1056 (Other) :1719
## X9 X10
## Administrative : 1 Services :2048
## Business Organization:5368 Health :1146
## University Department: 39 Public Administration : 693
## Defence : 386
## Agriculture & Horticulture : 212
## Technology: IT & Telephone Services: 205
## (Other) : 718
## X11
## Education :1210
## Public Administration : 693
## Health Care & Healthy Aging: 587
## Defence : 386
## Community : 281
## Pharmaceuticals : 278
## (Other) :1973
## X12 X13 X14
## Government: Australia: Local & State :1126 Min. :0.0000 Min. :0.0000
## Government: Australia: Federal :1007 1st Qu.:1.0000 1st Qu.:0.0000
## SME (small to medium enterprise) : 756 Median :1.0000 Median :0.0000
## Not for profit : 689 Mean :0.9926 Mean :0.1627
## Multinational / Other Large Corporate: 680 3rd Qu.:1.0000 3rd Qu.:0.0000
## University : 446 Max. :1.0000 Max. :1.0000
## (Other) : 704
## X15 X16
## Min. :2015 Min. : 1.000

```

```
## 1st Qu.:2017    1st Qu.: 4.000
## Median :2018    Median : 6.000
## Mean   :2018    Mean   : 6.323
## 3rd Qu.:2019    3rd Qu.: 9.000
## Max.   :2020    Max.   :12.000
##
```

Build regression model with 1 predictor (“StageName”) and 15 features (X2-X16)