Financial Independence Survey - Logistic Regression Model

Overview

The objective of this project is to develop a Logistic Regression Model to address the Research Question:

What factors contribute to an individual's perception of financial independence?

To accomplish this objective, we will analyze a dataset that is a subset of the official results from the 2020 Financial Independence Survey on Reddit (r/financialindependence). This subset exclusively comprises responses from individuals representing themselves (excluding contributions from other household members) and excludes retired individuals. The dataset encompasses 1,998 rows and 65 variables, covering information such as income contributors, the financial impact of the pandemic, political affiliation, demographics, details about financial independence, employment status, and various financial aspects. The data has been sourced from Reddit.

For additional details regarding the data dictionary and source information, please refer to the www.openintro.org website.

Data cleaning

Which variables required cleaning? Are there any missing values? Did you make any assumptions during the data cleaning process?

A series of transformations was carried out during the data cleaning process of their 65 original variables, among which the following stand out:

- Out of a total of 30 variables, which constituted numeric-type variables such as Children Expenses, Luxury Expenses, Transportation Expenses, Taxes, Medical Debt, etc., it was observed that they did not contain the number 0 but did have many NA (Not Available) values. The assumption was made that individuals who responded with 'NA' in fields such as Children Expenses did so because they had no associated expenses in that category. For this reason, in these 30 variables, NA values were replaced with 0.
-

Modeling

```
Call:
glm(formula = fin_indy ~ age + political_grouped + edu_grouped +
    housing + housing + total_debts + total_assets + total_expenses +
    `2020_gross_inc` + `2020_invst_save`, family = "binomial",
    data = reddit_finance_sub)

Coefficients:
                                          Estimate Std. Error z value
(Intercept)                             -5.745e+00  9.606e-01  -5.981
age                                      8.591e-02  1.453e-02   5.913
political_groupedLibertarian Party       2.836e-01  4.195e-01   0.676
political_groupedOther                   4.517e-01  2.220e-01   2.034
political_groupedRepublican              7.677e-01  3.091e-01   2.484
edu_groupedSome College or Trade School -1.200e+00  8.779e-01  -1.367
edu_groupedBachelor's or Associate's Degree -5.363e-01  7.098e-01  -0.756
edu_groupedDoctorate or Master's Degree -4.854e-01  7.147e-01  -0.679
housingOther                             1.076e+00  7.479e-01   1.438
housingOwn                               1.246e-01  5.490e-01   0.227
housingRent                              1.366e-01  5.376e-01   0.254
total_debts                             -1.627e-06  4.384e-07  -3.711
total_assets                             5.302e-07  7.363e-08   7.202
total_expenses                          -2.833e-06  1.046e-06  -2.709
`2020_gross_inc`                         1.226e-06  6.438e-07   1.905
`2020_invst_save`                       -2.416e-06  1.296e-06  -1.865
                                        Pr(>|z|)
(Intercept)                             2.22e-09 ***
age                                     3.36e-09 ***
political_groupedLibertarian Party      0.498987
political_groupedOther                  0.041910 *
political_groupedRepublican             0.013007 *
edu_groupedSome College or Trade School 0.171501
edu_groupedBachelor's or Associate's Degree 0.449912
edu_groupedDoctorate or Master's Degree 0.497045
housingOther                            0.150337
housingOwn                              0.820388
housingRent                             0.799376
total_debts                             0.000207 ***
total_assets                            5.95e-13 ***
total_expenses                          0.006740 **
```

```
`2020_gross_inc`                            0.056824 .
`2020_invst_save`                           0.062172 .
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 1037.51  on 1981  degrees of freedom
Residual deviance:  788.88  on 1966  degrees of freedom
AIC: 820.88

Number of Fisher Scoring iterations: 6
```

Results

Future work