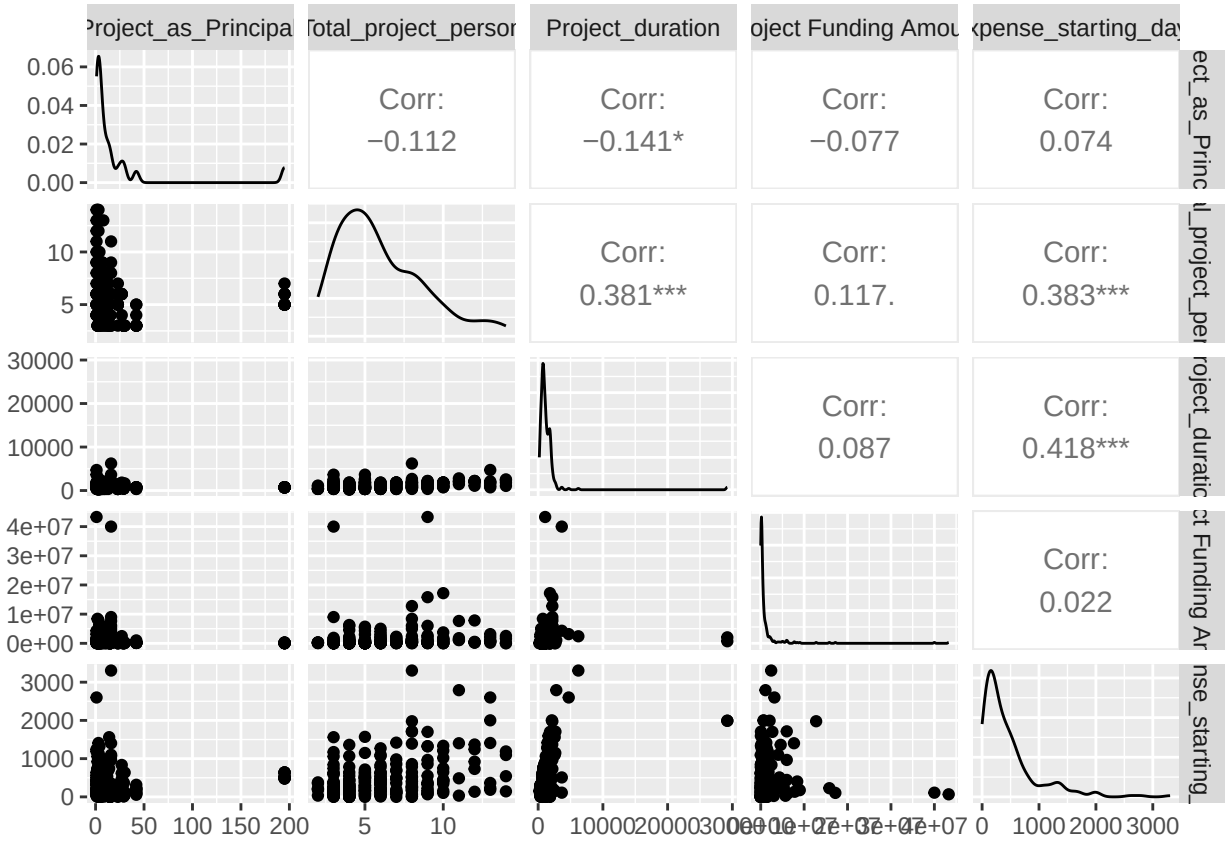
Spending start days prediction: Regression Approach

Md Ismail Hossain

2025-03-03

Data Structure:




```
##
## Call:
## lm(formula = Expense_starting_days ~ Number_of_Project_as_Principal_Investigator +
##     Total_project_person + Project_duration + `Project Funding Amount` +
##     `Project Funding Type` + `Project Type`, data = project_expenditure_indirect_cost_selected)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -817.75 -191.81  -10.36  135.42 1044.12
##
## Coefficients:
##                                              Estimate Std. Error t value
## (Intercept)                                -2.658e+02  6.446e+01  -4.124
## Number_of_Project_as_Principal_Investigator  1.891e+00  5.011e-01   3.773
## Total_project_person                         3.229e+01  8.727e+00   3.700
## Project_duration                             3.430e-01  3.084e-02  11.123
## `Project Funding Amount`                    -1.777e-05  5.021e-06  -3.540
## `Project Funding Type`Federal Passthrough    2.034e+02  6.036e+01   3.370
## `Project Funding Type`Internal               1.031e+02  2.973e+02   0.347
## `Project Funding Type`Non-Federal            1.503e+02  5.161e+01   2.912
## `Project Type`UW Grant Cost Share           -1.498e+02  1.094e+02  -1.370
##                                             Pr(>|t|)
## (Intercept)                                 5.51e-05 ***
```

```
## Number_of_Project_as_Principal_Investigator 0.000214 ***
## Total_project_person                        0.000281 ***
## Project_duration                             < 2e-16 ***
## `Project Funding Amount`                    0.000501 ***
## `Project Funding Type`Federal Passthrough   0.000905 ***
## `Project Funding Type`Internal              0.729013
## `Project Funding Type`Non-Federal           0.004006 **
## `Project Type`UW Grant Cost Share           0.172290
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 295.3 on 195 degrees of freedom
##   (91 observations deleted due to missingness)
## Multiple R-squared:  0.4898, Adjusted R-squared:  0.4689
## F-statistic:  23.4 on 8 and 195 DF,  p-value: < 2.2e-16


##
## Call:
## glm(formula = Expense_starting_days ~ Number_of_Project_as_Principal_Investigator +
##     Total_project_person + Project_duration + `Project Funding Amount` +
##     `Project Funding Type` + `Project Type`, family = poisson(link = "log"),
##     data = project_expenditure_indirect_cost_selected)
##
## Coefficients:
##                                             Estimate Std. Error z value
## (Intercept)                                4.646e+00  1.116e-02 416.322
## Number_of_Project_as_Principal_Investigator 4.817e-03  8.084e-05  59.591
## Total_project_person                        8.264e-02  1.350e-03  61.235
## Project_duration                            4.371e-04  3.045e-06 143.527
## `Project Funding Amount`                   -3.890e-08  1.327e-09 -29.323
## `Project Funding Type`Federal Passthrough   5.151e-01  1.007e-02  51.167
## `Project Funding Type`Internal              3.489e-01  5.414e-02   6.444
## `Project Funding Type`Non-Federal           3.293e-01  9.216e-03  35.734
## `Project Type`UW Grant Cost Share          -6.289e-01  2.296e-02 -27.391
##                                            Pr(>|z|)
## (Intercept)                                 < 2e-16 ***
## Number_of_Project_as_Principal_Investigator  < 2e-16 ***
## Total_project_person                        < 2e-16 ***
## Project_duration                            < 2e-16 ***
## `Project Funding Amount`                    < 2e-16 ***
## `Project Funding Type`Federal Passthrough   < 2e-16 ***
## `Project Funding Type`Internal             1.16e-10 ***
## `Project Funding Type`Non-Federal           < 2e-16 ***
## `Project Type`UW Grant Cost Share           < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for poisson family taken to be 1)
##
##     Null deviance: 66571  on 203  degrees of freedom
## Residual deviance: 37100  on 195  degrees of freedom
##   (91 observations deleted due to missingness)
## AIC: 38562
##
```

```
## Number of Fisher Scoring iterations: 5
```

```
## [1] 452.8169
```

```
## [1] 239580
```

```
##
## Call:
## glm.nb(formula = Expense_starting_days ~ Number_of_Project_as_Principal_Investigator +
##     Total_project_person + Project_duration + `Project Funding Amount` +
##     `Project Funding Type` + `Project Type`, data = project_expenditure_indirect_cost_selected,
##     init.theta = 1.140417393, link = log)
##
## Coefficients:
##                                             Estimate Std. Error z value
## (Intercept)                                4.497e+00  2.048e-01  21.956
## Number_of_Project_as_Principal_Investigator 5.419e-03  1.591e-03   3.405
## Total_project_person                        8.228e-02  2.772e-02   2.969
## Project_duration                            5.364e-04  9.794e-05   5.477
## `Project Funding Amount`                   -4.280e-08  1.602e-08  -2.671
## `Project Funding Type`Federal Passthrough   5.450e-01  1.917e-01   2.843
## `Project Funding Type`Internal              4.017e-01  9.441e-01   0.425
## `Project Funding Type`Non-Federal           3.973e-01  1.640e-01   2.423
## `Project Type`UW Grant Cost Share          -8.692e-01  3.480e-01  -2.498
##                                            Pr(>|z|)
## (Intercept)                                 < 2e-16 ***
## Number_of_Project_as_Principal_Investigator 0.000661 ***
## Total_project_person                        0.002991 **
## Project_duration                            4.32e-08 ***
## `Project Funding Amount`                    0.007570 **
## `Project Funding Type`Federal Passthrough   0.004470 **
## `Project Funding Type`Internal              0.670520
## `Project Funding Type`Non-Federal           0.015386 *
## `Project Type`UW Grant Cost Share           0.012499 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for Negative Binomial(1.1404) family taken to be 1)
##
##     Null deviance: 312.94  on 203  degrees of freedom
## Residual deviance: 237.04  on 195  degrees of freedom
##   (91 observations deleted due to missingness)
## AIC: 2783.8
##
## Number of Fisher Scoring iterations: 1
##
##
##               Theta:  1.140
##           Std. Err.:  0.106
##
##  2 x log-likelihood:  -2763.801
```

```
##               df       AIC
```

```
## ols_model     10  2910.455
## poisson_model  9 38561.620
## negbin_model  10  2783.801
```

Looks like Negative Binomial is best so far!