Following is the code used to make the interpretation

Giriraj

2022-11-15

```
library(dplyr)
```

```
##
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':
##
##     filter, lag

## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```
# loading files for data
ppei_features <- read.csv('Datasets/poverty_and_ecomnomic_indicators.csv')
ppei_labels   <- read.csv('Datasets/train_labels.csv')

# merging loaded files as one
data <- merge(ppei_features, ppei_labels, by = c('row_id'), all.x=T)

# generating glimpse of data
glimpse(data)
```

```
## Rows: 12,600
## Columns: 60
## $ row_id                          <int> 0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10~
## $ country                         <chr> "C", "C", "A", "A", "D", "A", "C~
## $ is_urban                        <chr> "False", "True", "False", "False~
## $ age                             <dbl> 18, 30, 20, 61, 26, 36, 35, 33, ~
## $ female                          <chr> "True", "True", "True", "False",~
## $ married                         <chr> "True", "True", "True", "True", ~
## $ religion                        <chr> "P", "P", "Q", "Q", "X", "Q", "P~
## $ relationship_to_hh_head         <chr> "Other", "Other", "Spouse", "Hea~
## $ education_level                 <dbl> 1, 1, 1, 0, 1, 1, 1, 1, 2, 1, 0,~
## $ literacy                        <chr> "True", "True", "True", "False",~
## $ can_add                         <chr> "True", "True", "True", "True", ~
## $ can_divide                      <chr> "True", "True", "True", "True", ~
## $ can_calc_percents               <chr> "True", "False", "True", "False"~
## $ can_calc_compounding            <chr> "True", "False", "False", "True"~
## $ employed_last_year              <chr> "False", "False", "True", "True"~
```

```
## $ employment_category_last_year         <chr> "housewife_or_student", "housewi~
## $ employment_type_last_year             <chr> "not_working", "not_working", "i~
## $ share_hh_income_provided              <dbl> 1, NA, 1, NA, 2, 1, 5, 3, 5, 1, ~
## $ income_ag_livestock_last_year         <chr> "False", "False", "False", "Fals~
## $ income_friends_family_last_year       <chr> "False", "False", "False", "Fals~
## $ income_government_last_year           <chr> "False", "False", "False", "Fals~
## $ income_own_business_last_year         <chr> "False", "False", "False", "Fals~
## $ income_private_sector_last_year       <chr> "False", "False", "False", "Fals~
## $ income_public_sector_last_year        <chr> "False", "False", "False", "Fals~
## $ num_times_borrowed_last_year          <int> 0, 0, 1, 0, 0, 1, 0, 0, 3, 0, 0,~
## $ borrowing_recency                     <int> 0, 0, 2, 0, 0, 2, 0, 0, 2, 0, 0,~
## $ formal_savings                        <chr> "False", "False", "False", "Fals~
## $ informal_savings                      <chr> "False", "False", "False", "Fals~
## $ cash_property_savings                 <chr> "False", "False", "False", "Fals~
## $ has_insurance                         <chr> "False", "False", "False", "Fals~
## $ has_investment                        <chr> "False", "False", "False", "Fals~
## $ bank_interest_rate                    <dbl> NA, NA, NA, NA, NA, NA, NA, NA, ~
## $ mm_interest_rate                      <dbl> NA, NA, NA, NA, NA, NA, NA, NA, ~
## $ mfi_interest_rate                     <dbl> NA, NA, NA, NA, NA, NA, NA, NA, ~
## $ other_fsp_interest_rate               <dbl> NA, NA, NA, NA, NA, NA, NA, NA, ~
## $ num_shocks_last_year                  <int> 0, 0, 0, 0, 0, 1, 0, 1, 1, 0, 0,~
## $ avg_shock_strength_last_year          <dbl> 0, 0, 0, 0, 0, 4, 0, 1, 5, 0, 0,~
## $ borrowed_for_emergency_last_year      <chr> "False", "False", "False", "Fals~
## $ borrowed_for_daily_expenses_last_year <chr> "False", "False", "False", "Fals~
## $ borrowed_for_home_or_biz_last_year    <chr> "False", "False", "False", "Fals~
## $ phone_technology                      <int> 0, 1, 1, 0, 0, 0, 0, 2, 3, 1, 0,~
## $ can_call                              <chr> "True", "True", "True", "True", ~
## $ can_text                              <chr> "True", "False", "False", "False~
## $ can_use_internet                      <chr> "False", "False", "False", "Fals~
## $ can_make_transaction                  <chr> "False", "False", "False", "Fals~
## $ phone_ownership                       <int> 1, 2, 2, 0, 1, 1, 1, 2, 2, 2, 1,~
## $ advanced_phone_use                    <chr> "False", "False", "False", "Fals~
## $ reg_bank_acct                         <chr> "True", "True", "False", "False"~
## $ reg_mm_acct                           <chr> "False", "False", "False", "Fals~
## $ reg_formal_nbfi_account               <chr> "False", "False", "False", "Fals~
## $ financially_included                  <chr> "True", "True", "False", "False"~
## $ active_bank_user                      <chr> "True", "True", "False", "False"~
## $ active_mm_user                        <chr> "False", "False", "False", "Fals~
## $ active_formal_nbfi_user               <chr> "False", "False", "False", "Fals~
## $ active_informal_nbfi_user             <chr> "False", "False", "False", "Fals~
## $ nonreg_active_mm_user                 <chr> "False", "False", "False", "Fals~
## $ num_formal_institutions_last_year     <int> 1, 1, 0, 0, 1, 1, 1, 0, 3, 1, 1,~
## $ num_informal_institutions_last_year   <int> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,~
## $ num_financial_activities_last_year    <int> 1, 0, 0, 0, 3, 0, 1, 0, 7, 1, 1,~
## $ poverty_probability                   <dbl> 0.515, 0.981, 0.982, 0.879, 0.79~
```

```
# performing descriptive analysis on education level
summary(data$education_level)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.    NA's
##   0.000   1.000   1.000   1.316   2.000   3.000     236
```

```r
# replacing NA with 4 and removing all datapoints with value 4 from the dataset
data$education_level[is.na(data$education_level) == TRUE] <- 4
data <- data[data$education_level != 4,]
```

```r
# checking for levels of the variables education and literacy
table(data$education_level)
```

```
##
##    0    1    2    3
## 2545 4550 4083 1186
```

```r
table(data$literacy)
```

```
##
## False  True
##  4668  7696
```

```r
# factoring education_level to have in-depth analysis
data$education_level.f <- as.factor(data$education_level)

# performing descriptive analysis on education level after treating NA values
summary(data$education_level)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.000   1.000   1.000   1.316   2.000   3.000
```

```r
# factoring literacy to have in-depth analysis
data$literacy.f <- as.factor(data$literacy)

# re-leveling literacy to use Literacy = TRUE as level 1
data$literacy.f <- relevel(data$literacy.f, "True")
levels(data$literacy.f)
```

```
## [1] "True"  "False"
```

```r
# generating an Ordinary Least Square model
print("Following is the OLS model summary: ")
```

```
## [1] "Following is the OLS model summary: "
```

```r
ols_model <- lm(poverty_probability ~ data$education_level.f + data$literacy.f, data = data)
summary(ols_model)
```

```
##
## Call:
## lm(formula = poverty_probability ~ data$education_level.f + data$literacy.f,
##     data = data)
##
## Residuals:
```

```
##     Min      1Q  Median      3Q     Max
## -0.7128 -0.2021  0.0451  0.2295  0.6009
##
## Coefficients:
##                           Estimate Std. Error t value Pr(>|t|)
## (Intercept)               0.709734   0.007870  90.178  < 2e-16 ***
## data$education_level.f1 -0.026831   0.007408  -3.622 0.000294 ***
## data$education_level.f2 -0.188195   0.008488 -22.172  < 2e-16 ***
## data$education_level.f3 -0.316599   0.011042 -28.671  < 2e-16 ***
## data$literacy.fFalse     0.003057   0.006409   0.477 0.633383
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.272 on 12359 degrees of freedom
## Multiple R-squared:  0.1321, Adjusted R-squared:  0.1318
## F-statistic: 470.4 on 4 and 12359 DF,  p-value: < 2.2e-16
```

```
# generating an ANOVA model
print("Following is the ANOVA model summary: ")
```

```
## [1] "Following is the ANOVA model summary: "
```

```
anova_model <- aov(poverty_probability ~ data$education_level.f + data$literacy.f, data = data)
summary(anova_model)
```

```
##                            Df Sum Sq Mean Sq F value Pr(>F)
## data$education_level.f      3  139.1   46.38 627.091 <2e-16 ***
## data$literacy.f             1    0.0    0.02   0.228  0.633
## Residuals               12359  914.1    0.07
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```