Employment determinants among Ukrainian refugees: a comparative analysis across European host countries

Logistic regression and random forest models for employment prediction

Table of contents

Abstract

Introduction and context

The inflow of more than 8 million Ukrainian refugees into Europe has introduced both challenges and opportunities for labour market outcomes. People fleeing the war in Ukraine are allowed to live and work in the European Union for up tp 3 years under the Temporary Protection Status. Although the education and qualification levels of Ukrainian refugees are generally high, difficulties such as lack of language skills, childcare services and credential recognition processes pose significant barriers to employment. Instances of ethnic discrimination in the labor market have been reported, highlighting additional challenges faced by Ukrainian refugees (Londar et al., 2024). According to Preut, expanding language courses, childcare services, and accelerating credential recognition can enhance employment likelihood (Preut, 2023).

Data and Methods

This study employs Multi-Sectoral Needs Assessments (MSNA) of 2023, which provide data on Ukrainian refugees' needs and priorities in Estonia, Slovakia, Poland, Romania, Moldova, and Czechia. The assessment employs both household-level and individual-level data collected through structured surveys in all the countries.

Data Wrangling and Preparation

To enable cross-regional comparisons, variables that were consistently present across all six national datasets were selected and merged into a single dataset. Questions allowing multiple responses (e.g., types of support received, reasons for unemployment, barriers to accessing services) were transformed using one-hot encoding into a series of binary indicator variables (1 = selected, 0 = not selected) to show each of the possible response categories in disaggregated form. Where categorical responses varied slightly across countries (e.g., differences in wording or coding), the values were recoded into a uniform format for comparison.

For the convenience of exploratory data analysis (EDA) and model building, the outcome variable (employment status) was defined as follows:

- Respondents were classified as employed (1) if they reported any form of employment.
- All others (unemployed, in education, retired, or not working due to other factors) were classified as unemployed (0).

This two-category outcome variable, Employed, was used as the dependent variable for all the following regression and classification models.

Based on the proportions of missing data and their impact on model performance, non-available observations were excluded or imputed. Individual aggregated dataframes were created so that the distribution of binary-coded multi-response variables could be visualized and country-level comparisons could be made.

Statistical Modeling

Logistic regression and random forest classification were applied in an attempt to identify influential factors concerning employment among Ukrainian refugees:

- Random Forest Classifier

A random forest model was trained to account for potential non-linearities and interactions between variables. The model was trained using predictors selected based on feature importance scores. Its performance was evaluated using classification measures such as accuracy, precision, recall, and area under the ROC curve (AUC). Random forest allowed for selection of the most significant predictors of job-finding based on mean decrease in impurity (Gini importance) and permutation importance.

- Logistic Regression

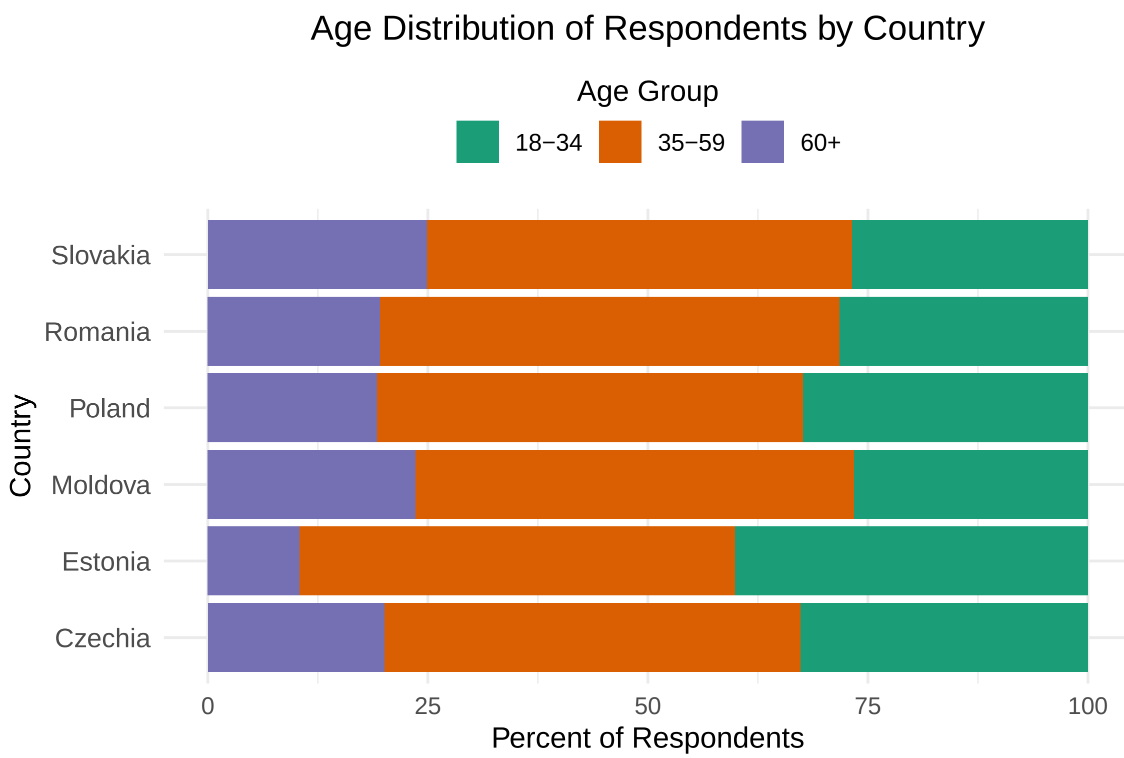
Binary logistic regression analysis was used to estimate the probability of employment (Employed = 1). Independent variables included demographic and socio-economic ones such as age, education level, country of residence, language skills, accommodation, and social services consumption of social services, among others. Odds ratios were estimated to define direction and strength of associations.

Pooled Analysis Across Six Countries

Exploratory Data Analysis

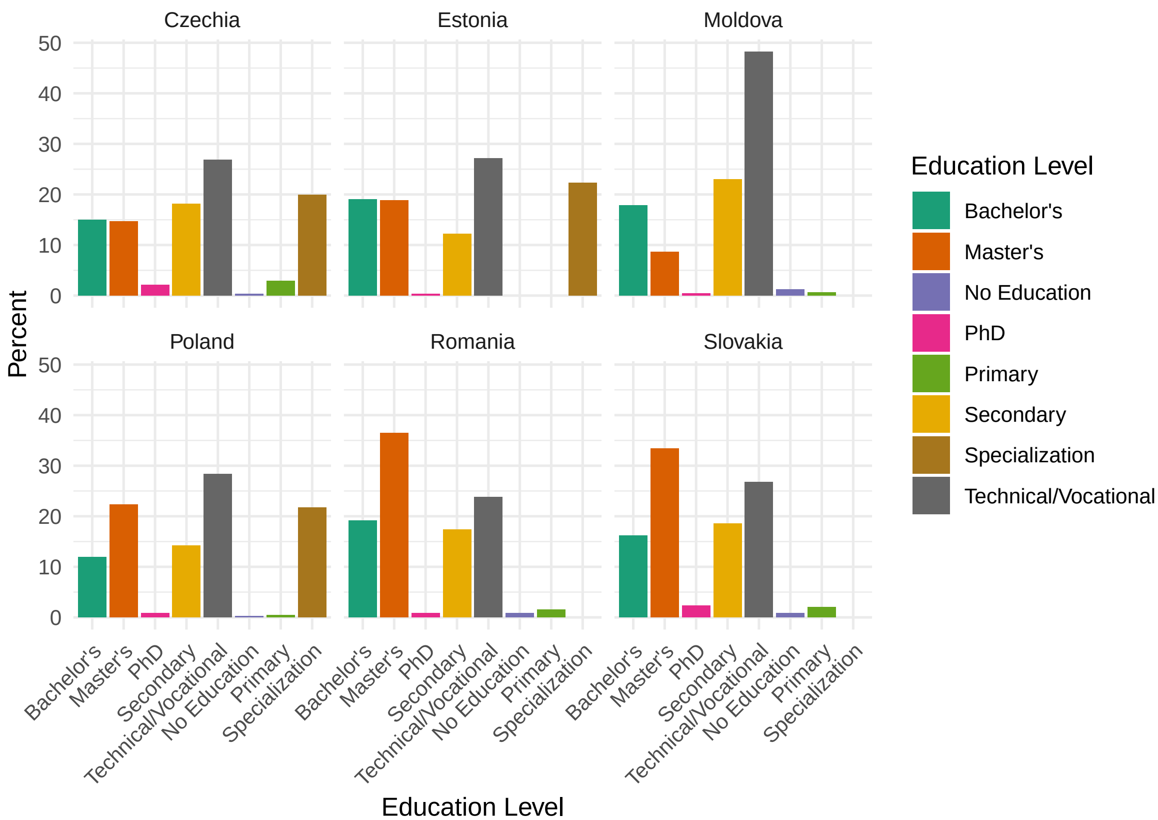
Demographics

The bar chart displays the distribution of survey respondents by age group (18–34, 35–59, 60+) across six Central and Eastern European countries. Across all countries, the 35–59 age group consistently has the highest number of respondents, followed by the 18–34 group, with the 60+ group being the least represented. Romania stands out with a particularly high concentration of respondents in the 35–59 group compared to the other age categories. Estonia shows a more balanced distribution between the younger and middle-aged groups, though the same pattern persists.



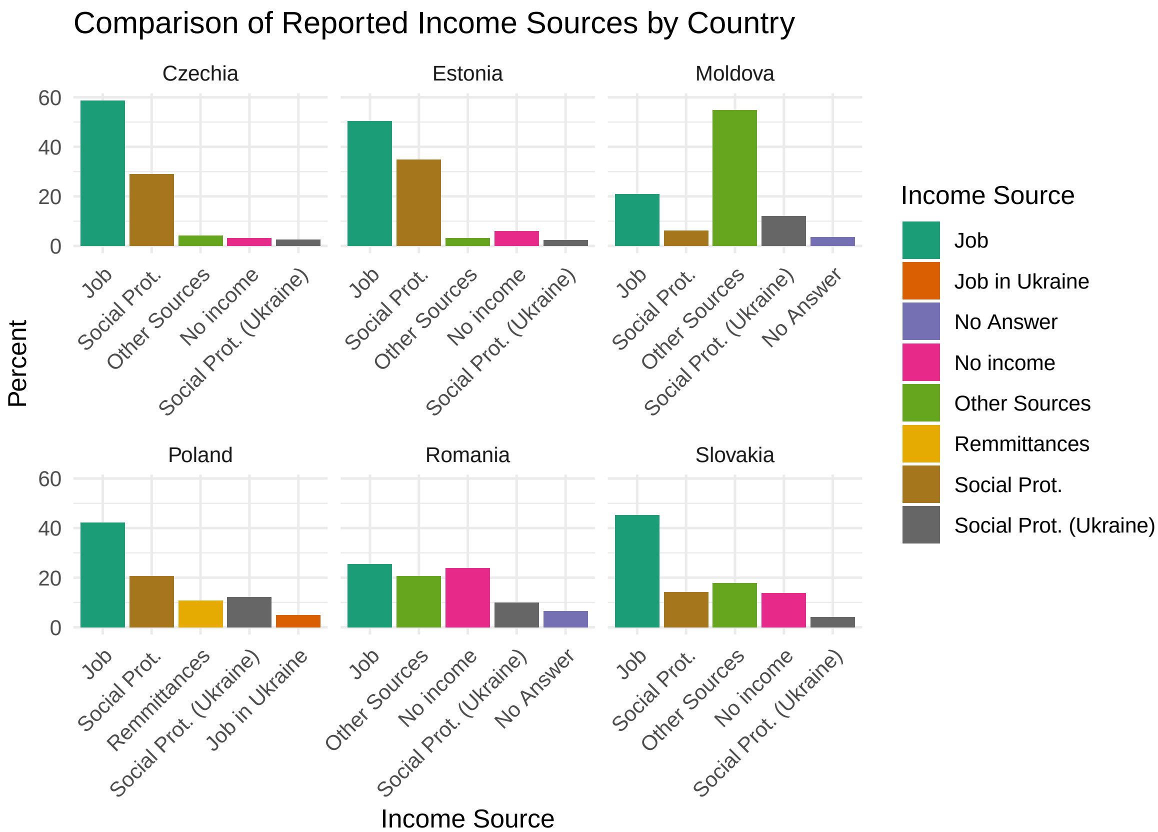
While the data on education of Ukrainian refugees is limited, the snapshot of age distribution reveals that the majority of arrivals are part of labor force. Notably, in Estonia, 49.46% of arrivals hold a vocational and specialization education, followed by around 38% of refugees with Master or Bachelor degrees. People with secondary education are moderately represented, but there are very few people with PhD education. The largest group is employed, indicating that many refugees were active in the workforce before displacement. Other categories include students, self-employed, and those in housekeeping.

The educational categories range from primary and secondary education to higher levels such as Bachelor's, Master's, and PhD degrees, as well as technical or vocational training and no formal education. "Technical/Vocational" category appears frequently across multiple sections, suggesting a significant portion of the workforce in these regions may have either technical training. This trend could indicate a reliance on skilled labor or informal employment in certain areas. The number of people with Bachelor's degrees is almost the same across all countries and ranges from 10 to 20%. Slovakia and Romania have the highest proportion of people with Master's degrees. There are very few people with no education in all the countries. Slovakia and Czechia appear the most prominent for respondents with PdD degrees.

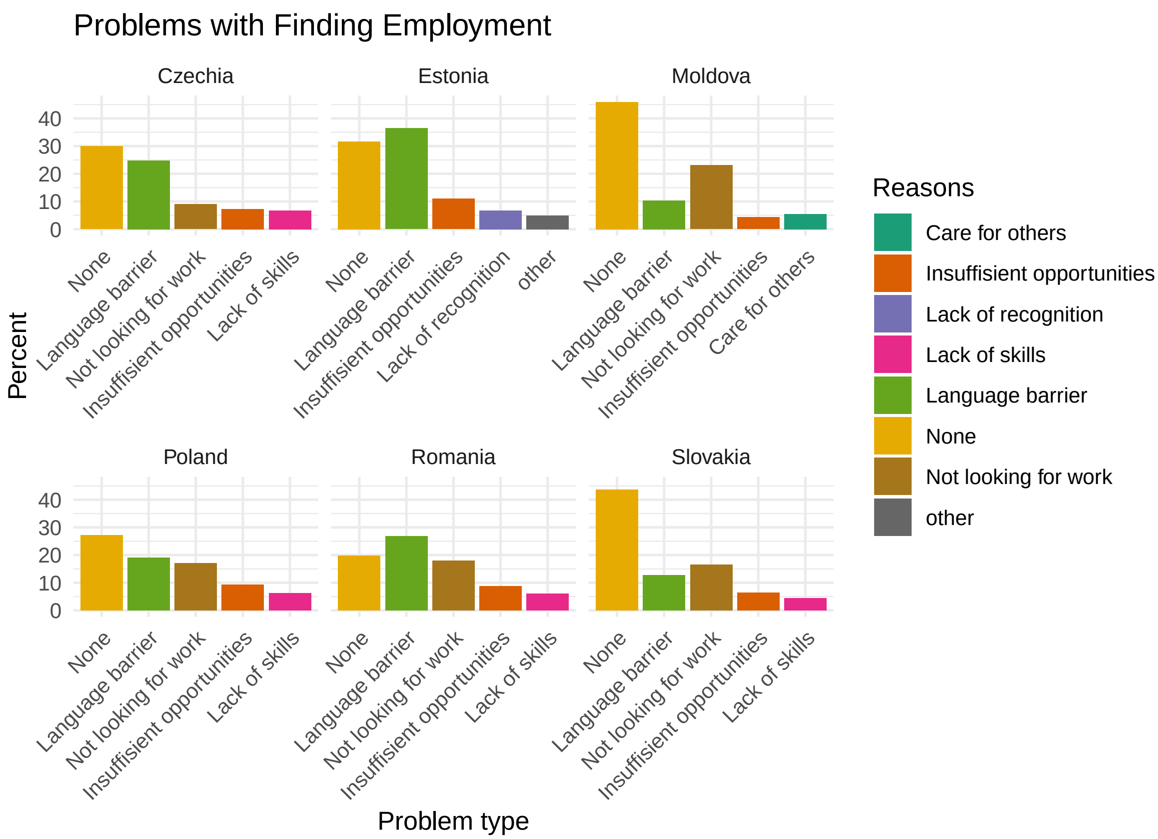


Economic capacity and Aid

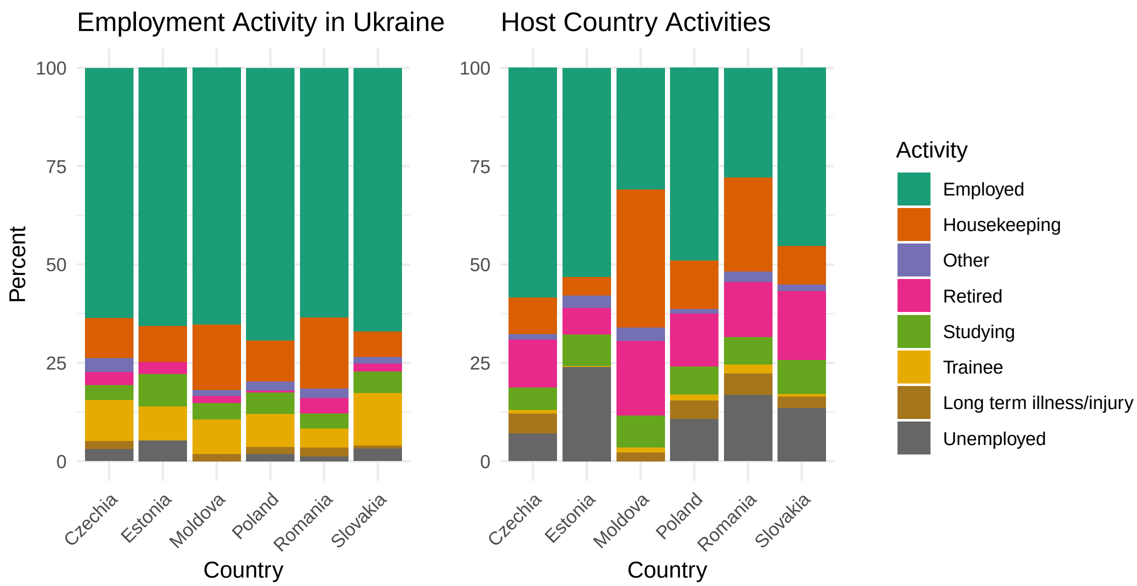
The faceted bar chart indicates the sources of declared income. The most apparent trend is the dominant role of "employment in the host country" as the principal source of income in most countries, especially in Czechia, Estonia, and Slovakia, where it accounts for over 40–50%. Romania has a more balanced split, with employment percentages, other sources, and maintained employment from Ukraine all being similar, however, it also has the highest number of people with no income. Remote work and remittances in Romania and Poland respectively also arise more than anywhere else. Across all countries, rates like "no answer", and “no income” are low.



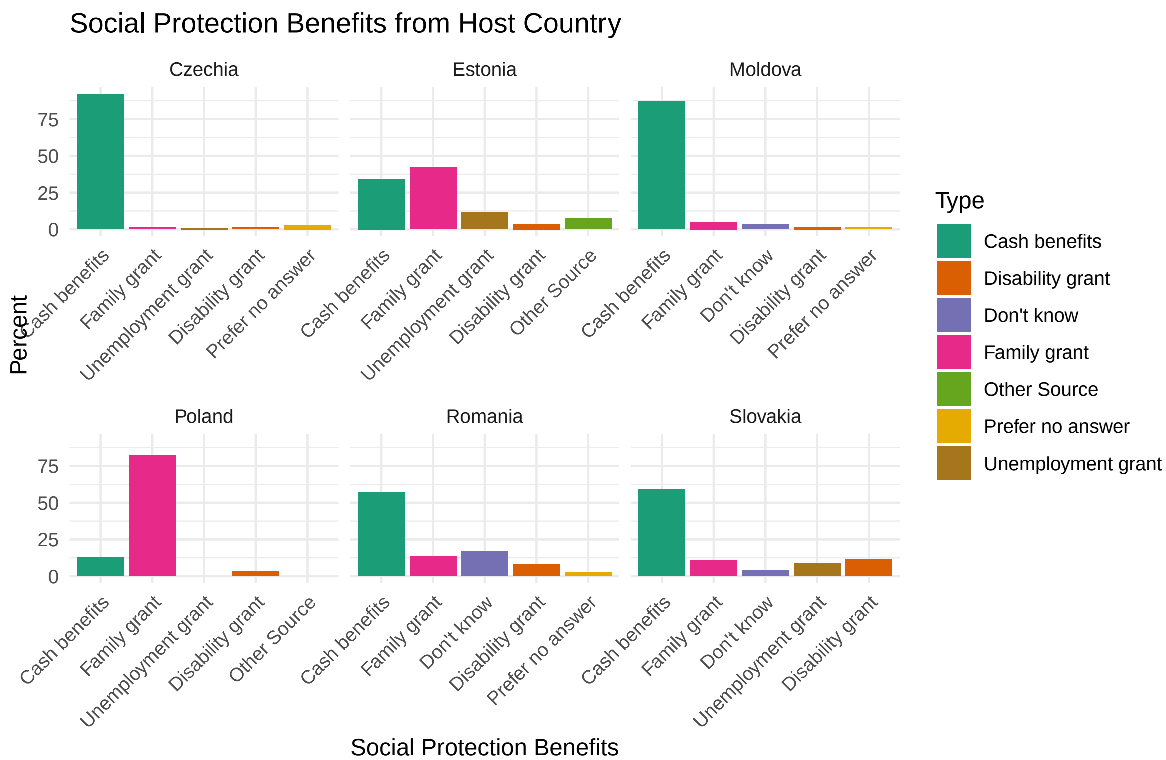
The chart shows that the most commonly reported reason for not experiencing employment problems is “None,” particularly in Moldova and Slovakia, where over 40% of respondents indicated no difficulties. In contrast, Estonia stands out with a high incidence of the “Language barrier” (around 37%), suggesting it is a major obstacle there. Across all countries, language barriers and the fact that many are “Not looking for work” are consistently among the top reasons cited, though to varying degrees. “Lack of skills” and “Lack of recognition” are less frequently mentioned, typically below 10%. Moldova, Romania, and Poland also report relatively high percentages for “Not looking for work” and “Insufficient opportunities,” indicating structural and motivational challenges.



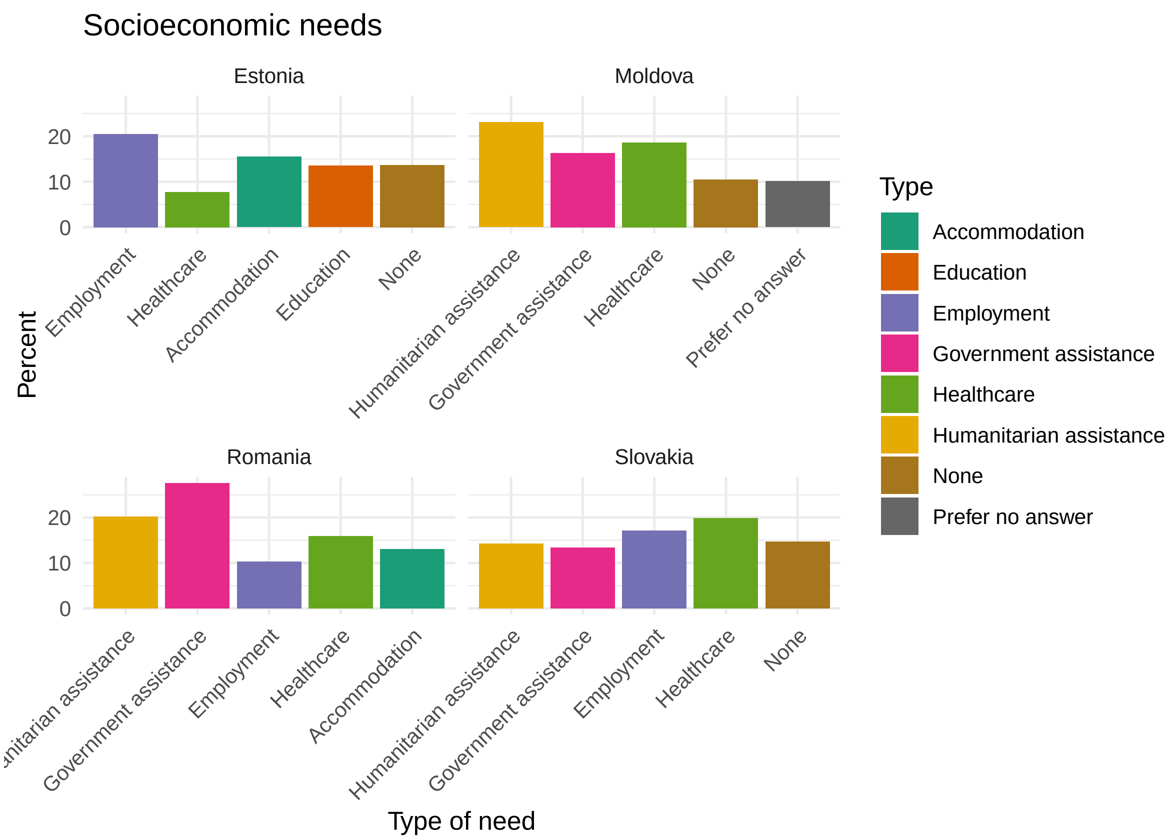
The graph compares employment activity of Ukrainian refugees prior to displacement with their current activities in host countries. In Ukraine, the vast majority of refugees across all countries were employed, with employment rates ranging from approximately 60% to over 70%. Housekeeping and retirement were the next most common categories, but far less prevalent. In contrast, in host countries, employment drops significantly, while unemployment, studying, and housekeeping rise sharply. For example, in Poland, Romania, and Slovakia, a substantial portion - around one-third - of refugees are now unemployed. Studying categories also grow considerably, particularly in Romania and Moldova. This shift indicates a major transition for many refugees, from stable employment in Ukraine to navigating unemployment or reskilling opportunities in their new environments.



The chart illustrates the types of social protection benefits received by Ukrainian refugees across six host countries. In most countries—especially Czechia, Moldova, Romania, and Slovakia—cash benefits are the dominant form of support, reaching around 80–90% in some cases. Poland stands out as the only country where the majority (over 80%) report receiving family grants instead of cash benefits. Estonia also presents a more balanced distribution, with both cash benefits and unemployment grants being common, and a notable share receiving unemployment grants or reporting "Other Source." Across all countries, disability and unemployment grants remain minor components. Overall, the data suggest that cash-based support remains the primary mechanism, though the specific type of aid can vary significantly by host country.

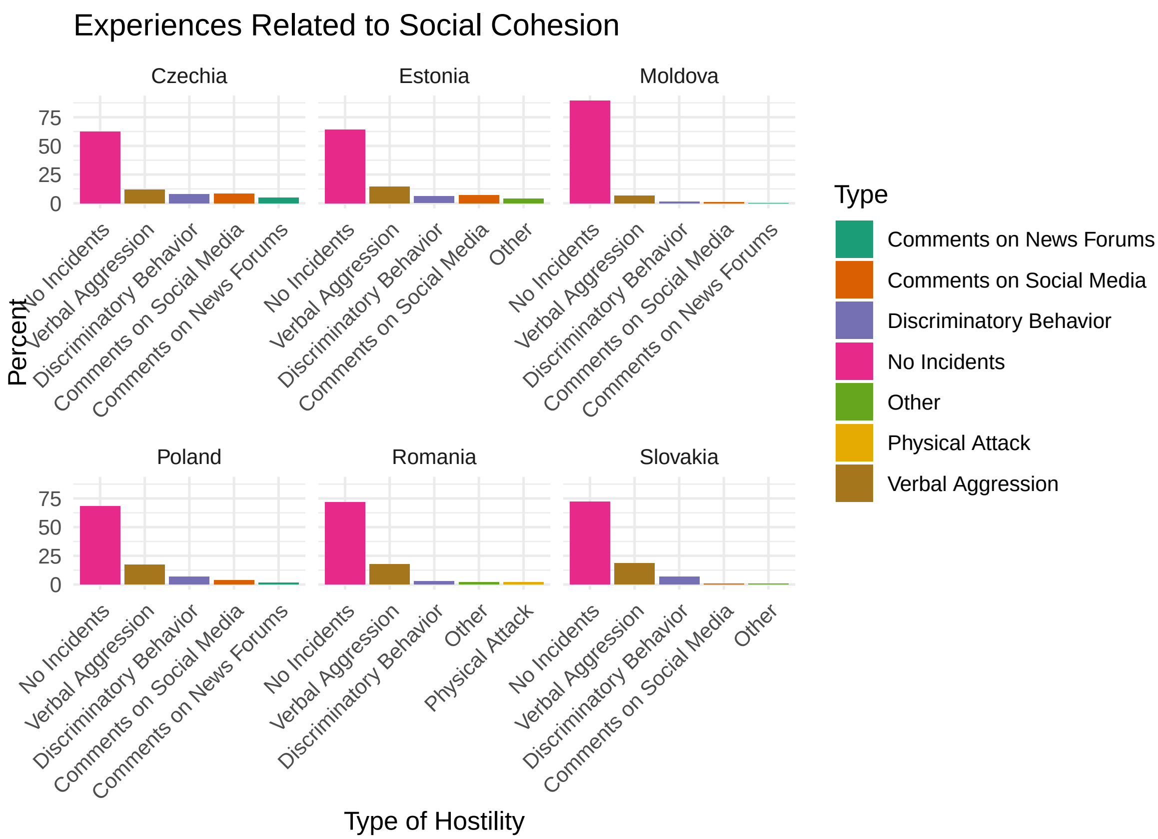


Social Protection Benefits from Host Country
Czechia
Estonia
Moldova
Poland
Romania
Slovakia
Percent
0
25
50
75
Cash benefits
Family grant
Unemployment grant
Disability grant
Prefer no answer
Other Source
Don't know
Social Protection Benefits
Type
Cash benefits
Disability grant
Don't know
Family grant
Other Source
Prefer no answer
Unemployment grant



Safety and Social Cohesion

According to data, a majority of respondents reported no incidents of aggression or discrimination across all countries. Moldova (~80%) and Slovakia (~75%) report the highest rates of "No Incidents". Verbal aggression and discriminatory behavior are the most common types of hostility experienced after "No Incidents". Romania, Poland, and Slovakia show moderate levels of verbal aggression and discriminatory behavior, likely reflecting higher public tension. Romania shows physical attack in its top five, and even there, the percentage is minimal, so physical safety, therefore, seems relatively high across the surveyed countries. Moldova stands out positively with the highest share of refugees reporting no hostility.



Logistic Regression (Pooled Model)

Random forest

To identify the most important predictors of employment among Ukrainian refugees, a Random Forest (RF) classifier was applied using the randomForest package in R. The analysis included 23 theoretically relevant independent variables related to demographics, barriers to employment, social benefits, and hostile experiences (see variable list below).

```
set.seed(123)
important_vars <- c(
  "introduction_resp_age",
  "demographics_educ_level_grouped",
  "demographics_resp_activity",
  "income_social_protection_host_govt",
  "income_remmittances",
  "income_social_protection_ukr_govt",
```

```r
  "diff_lack_of_lang",
  "diff_lack_childcare",
  "diff_lack_of_skills",
  "diff_lack_of_education_skills",
  "diff_lack_of_info",
  "diff_none",
  "diff_discrimination",
  "diff_lack_of_documentation",
  "needs_medical",
  "needs_accommodation",
  "needs_host_govt_assistance",
  "benefits_cash_benefits",
  "benefits_unemployment_grant",
  "hostile_comments_social_media",
  "hostile_verbal_aggression",
  "hostile_discriminatory_behavior",
  "hostile_none"
)
```

Only respondents with non-missing values across all included variables were selected for the analysis. `employed_binary` was converted into a factor with two levels: "Yes" and "No". A Random Forest model was fit with the following parameters:

- Number of trees (ntree): 500
- Number of variables randomly sampled at each split (mtry): 2
- Importance metric: Mean Decrease in Accuracy

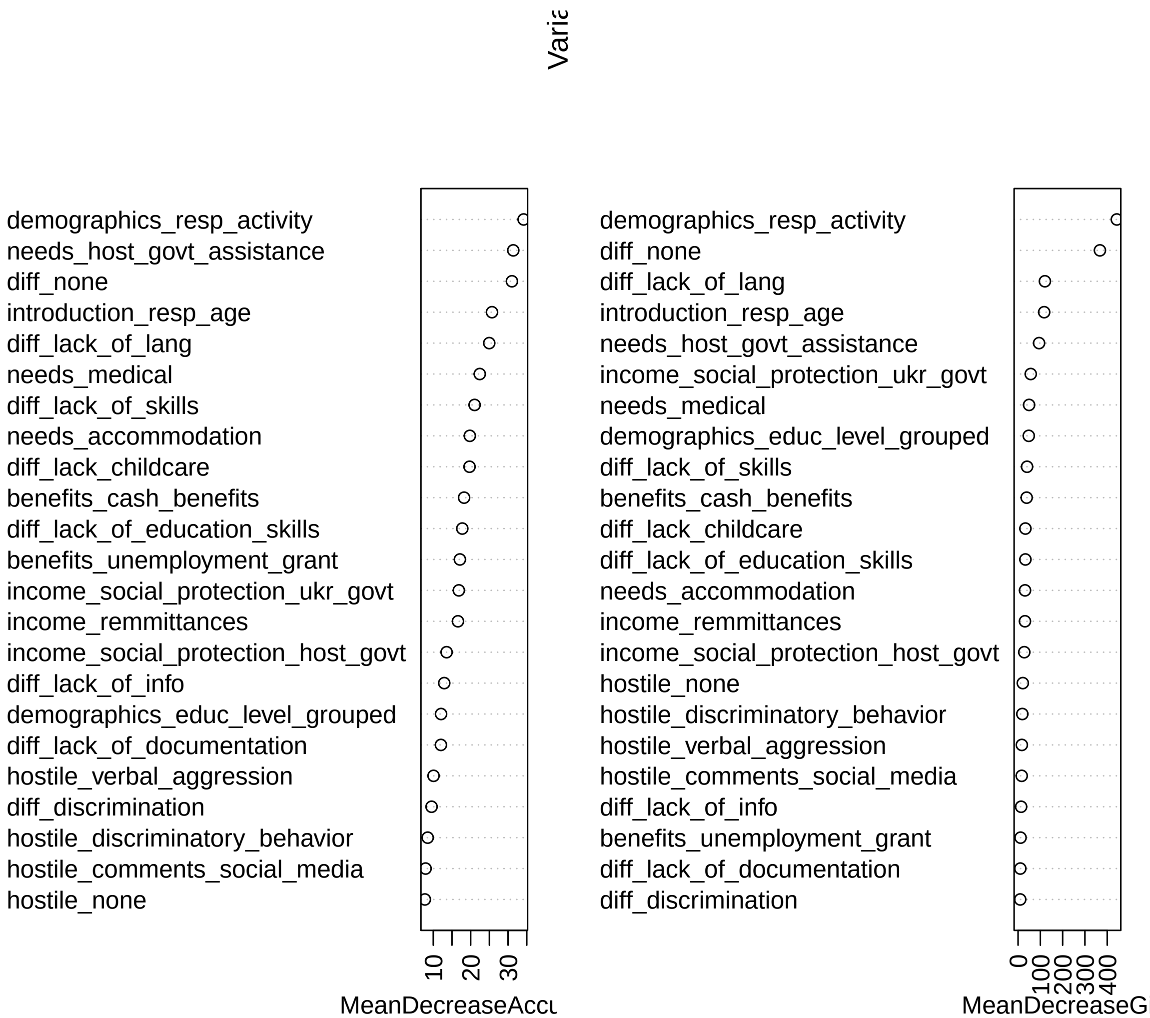
```r
model_data <- combined_data |>
  select(employed_binary, all_of(important_vars)) |>
  drop_na() |>
  mutate(employed_binary = factor(employed_binary, levels = c(0, 1),
                                  labels = c("No", "Yes")))

rf_model <- randomForest(employed_binary ~ ., data = model_data,
                         ntree = 500, mtry = 2, importance = TRUE)
```

The top 15 variables were selected based on their Mean Decrease in Accuracy, which reflects the contribution of each variable to prediction performance. The variable importance plot (varImpPlot) was used to visually assess the relative influence of each predictor.

The selected variables were then retained for subsequent logistic regression modeling, to estimate the direction and significance of their effects.

```r
imp <- importance(rf_model)
top_vars <- rownames(imp)[order(imp[, "MeanDecreaseAccuracy"],
                                decreasing = TRUE)][1:15]
par(mar = c(7, 6, 4, 2), las = 2)
varImpPlot(rf_model, main = "Variable Importance: Random Forest Model")
```




```r
model_data_selected <- model_data[, c("employed_binary", top_vars)]
model_data_selected$employed_binary <-
  as.integer(model_data_selected$employed_binary == "Yes")
```

Logistic Regression

A logistic regression model was fit to predict the binary employment outcome `employed_binary`, where 0 = Not Employed/1 = Employed, among Ukrainian refugees using selected explanatory variables identified as important in prior analysis.

The logistic regression results reveal several important predictors of employment status among Ukrainian refugees. For instance, respondents reporting no difficulties (diff_none) have substantially higher odds of employment, over 7 times greater, compared to those facing challenges. Conversely, individuals engaged in housekeeping activities show about 83% lower odds of employment relative to the baseline group, and those with long-term illness or injury exhibit even more pronounced reductions in employment odds. Older respondents (age 60+) also have significantly lower chances of being employed. Interestingly, reporting lack of language skills is associated with more than a threefold increase in the odds of employment (60.7%), suggesting complex underlying factors such as self-selection or varying integration experiences. Other barriers such as lack of childcare or medical needs, 35.8% and 41.9% lower respectively, correspond to decreased employment odds, as do receiving unemployment grants (76.9% lower) or social protection benefits from either host countries (11.5% lower) or Ukraine (34.5% lower).

```r
logit_model <- glm(employed_binary ~ .,
                   data = model_data_selected, family = binomial)
summary(logit_model)
```

```
Call:
glm(formula = employed_binary ~ ., family = binomial, data = model_data_selected)

Coefficients:
                                              Estimate Std. Error z value
(Intercept)                                    0.06838    0.05968   1.146
demographics_resp_activityHousekeeping        -1.77701    0.07683 -23.129
demographics_resp_activityLong term illness/injury -3.14444    0.27829 -11.299
demographics_resp_activityOther               -0.68796    0.15579  -4.416
demographics_resp_activityRetired             -1.83316    0.16432 -11.156
demographics_resp_activityStudying            -0.91515    0.08130 -11.256
demographics_resp_activityTrainee             -0.64276    0.15695  -4.095
demographics_resp_activityUnemployed          -0.90545    0.14530  -6.232
needs_host_govt_assistance                    -0.90319    0.06943 -13.008
diff_none                                      2.12265    0.05848  36.300
introduction_resp_age35-59                    -0.02881    0.05097  -0.565
introduction_resp_age60+                      -0.96521    0.11838  -8.153
diff_lack_of_lang                              1.15577    0.05094  22.690
```

```
needs_medical                                       -0.54324    0.07345  -7.396
diff_lack_of_skills                                  0.47595    0.06318   7.534
needs_accommodation                                 -0.22780    0.07952  -2.865
diff_lack_childcare                                 -0.44331    0.09888  -4.483
benefits_cash_benefits                              -0.61771    0.08576  -7.202
diff_lack_of_education_skills                        0.55654    0.09308   5.979
benefits_unemployment_grant                         -1.46624    0.27581  -5.316
income_social_protection_ukr_govt                   -0.42288    0.06681  -6.330
income_remmittances                                 -0.58374    0.07257  -8.043
income_social_protection_host_govt                  -0.12153    0.05645  -2.153
                                                    Pr(>|z|)
(Intercept)                                          0.25191
demographics_resp_activityHousekeeping               < 2e-16 ***
demographics_resp_activityLong term illness/injury   < 2e-16 ***
demographics_resp_activityOther                     1.01e-05 ***
demographics_resp_activityRetired                    < 2e-16 ***
demographics_resp_activityStudying                   < 2e-16 ***
demographics_resp_activityTrainee                   4.22e-05 ***
demographics_resp_activityUnemployed                4.61e-10 ***
needs_host_govt_assistance                           < 2e-16 ***
diff_none                                            < 2e-16 ***
introduction_resp_age35-59                           0.57199
introduction_resp_age60+                            3.54e-16 ***
diff_lack_of_lang                                    < 2e-16 ***
needs_medical                                       1.40e-13 ***
diff_lack_of_skills                                 4.93e-14 ***
needs_accommodation                                  0.00417 **
diff_lack_childcare                                 7.35e-06 ***
benefits_cash_benefits                              5.92e-13 ***
diff_lack_of_education_skills                       2.24e-09 ***
benefits_unemployment_grant                         1.06e-07 ***
income_social_protection_ukr_govt                   2.45e-10 ***
income_remmittances                                 8.73e-16 ***
income_social_protection_host_govt                   0.03132 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 17791  on 12889  degrees of freedom
Residual deviance: 13047  on 12867  degrees of freedom
AIC: 13093
```

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

Model Fit and Evaluation

The logistic regression model demonstrates a good overall fit to the data. McFadden's pseudo-R^2 is approximately 0.267, indicating the model explains around 27% of the variance in employment outcomes, which reflects a moderate level of explanatory power for this type of social data. In terms of predictive performance, the model achieves a classification accuracy of about 75.1%, correctly predicting employment status for three out of four individuals. The confusion matrix reveals that the model more accurately identifies those who are employed (5,677 correctly classified) compared to those not employed (4,002 correctly classified), though there are some misclassifications in both groups. Additionally, the Area Under the Receiver Operating Characteristic Curve (AUC) is 0.828, signifying strong discriminatory ability to distinguish between employed and unemployed individuals. Together, these metrics suggest the model is both statistically sound and practically useful for understanding employment determinants among Ukrainian refugees.

```r
pR2(logit_model)
```

```
fitting null model for pseudo-r2
```

```
          llh       llhNull            G2      McFadden          r2ML
-6523.6443129 -8895.6827665  4744.0769072     0.2666505     0.3079127
         r2CU
    0.4113823
```

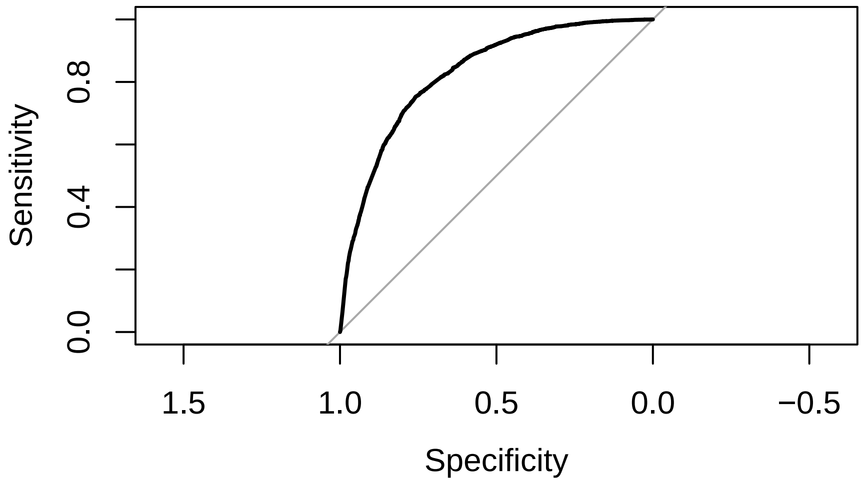
```r
pred_probs <- predict(logit_model, type = "response")
pred_class <- ifelse(pred_probs > 0.5, 1, 0)
table(Predicted = pred_class, Actual = model_data_selected$employed_binary)
```

```
         Actual
Predicted    0    1
        0 4002 1269
        1 1942 5677
```

```r
mean(pred_class == model_data_selected$employed_binary)
```

```
[1] 0.7508922
```

```r
roc_obj <- roc(model_data_selected$employed_binary, pred_probs)
plot(roc_obj)
```




```r
auc(roc_obj)
```

```
Area under the curve: 0.8279
```

Country-Specific Analysis

A consistent analytical procedure was applied for each country included in the study to identify key factors associated with employment among Ukrainian refugees. Initially, exploratory data analysis was conducted by plotting relevant predictor variables against the binary employment outcome to visually assess differences between employed and unemployed groups. Variables showing notable group differences or supported by sufficient sample sizes were selected for further modeling. Multiple logistic regression models were then fitted for each country, with statistically insignificant predictors being iteratively removed, in order to isolate the most influential variables affecting employment outcomes within each country. The significance alpha level for p-values was generally set at 0.05; however, in some cases, a more lenient threshold of 0.1 was used when exploratory data analysis (EDA) plots indicated meaningful differences. Additionally, the conditions for logistic regression were met, as the predictor variables included in the models were categorical. The following sections present regression results for each country, highlighting both common and unique factors influencing refugee employment.

Poland

Logistic regression

In Poland, multiple individual and structural factors were significantly associated with the likelihood of employment among Ukrainian refugees. Compared to respondents aged 18–34, those aged 60 and over had 69.3% lower odds of being employed, while the 35–59 group showed a marginal reduction in odds by 13.2%, though not statistically significant on the alpha 0.01 level ($p = 0.059$). Employment odds were also considerably lower for respondents engaged in housekeeping (80% lower), those with long-term illness or injury (96% lower), and retirees (80% lower), all highly significant. Studying and trainee statuses were associated with 53% and 42% lower odds of employment, respectively.

Income sources were also relevant: receiving social protection from the host government was associated with 16% lower odds of employment, remittances with 49% lower odds, and Ukrainian government benefits with 39% lower odds. Language barriers were a positive predictor: those citing lack of language skills as a difficulty had 134% higher odds of being employed, possibly indicating higher employment motivation or selection bias. Lack of childcare decreased employment odds by 57%, while difficulties with age-related opportunities reduced them by 26%. In contrast, reporting "no difficulties" was associated with a 373% increase in the odds of being employed. Not actively looking for work had a strong negative effect, reducing odds by 84%. Finally, challenges related to lack of skills and lack of educational qualifications slightly increased employment odds by 29% and 78%, respectively — possibly due to their association with active job-seeking behaviors.

```
Call:
glm(formula = employed_binary ~ ., family = binomial, data = poland_data)

Coefficients:
                                              Estimate Std. Error z value
(Intercept)                                    0.78438    0.09423   8.324
introduction_resp_age35-59                    -0.14252    0.07556  -1.886
introduction_resp_age60+                      -1.17855    0.17723  -6.650
demographics_resp_activityHousekeeping        -1.61056    0.11261 -14.302
demographics_resp_activityLong term illness/injury -3.22135    0.38360  -8.398
demographics_resp_activityOther               -1.19764    0.40752  -2.939
demographics_resp_activityRetired             -1.60703    0.22360  -7.187
demographics_resp_activityStudying            -0.76296    0.12028  -6.343
demographics_resp_activityTrainee             -0.53943    0.23431  -2.302
demographics_resp_activityUnemployed          -1.03238    0.21568  -4.787
income_social_protection_host_govt            -0.17779    0.06587  -2.699
income_remmittances                           -0.66915    0.07867  -8.506
income_social_protection_ukr_govt             -0.49970    0.08566  -5.834
diff_lack_of_lang                              0.84912    0.07634  11.122
diff_lack_childcare                           -0.85765    0.12434  -6.898
diff_lack_of_skills                            0.25544    0.08526   2.996
diff_lack_of_education_skills                  0.57518    0.11676   4.926
diff_none                                      1.56400    0.09414  16.614
diff_not_looking_for_work                     -1.84200    0.11139 -16.536
diff_lack_of_age_opportunities                -0.30076    0.11766  -2.556
                                              Pr(>|z|)
(Intercept)                                    < 2e-16 ***
introduction_resp_age35-59                     0.05928 .
introduction_resp_age60+                      2.94e-11 ***
demographics_resp_activityHousekeeping         < 2e-16 ***
demographics_resp_activityLong term illness/injury  < 2e-16 ***
demographics_resp_activityOther                0.00329 **
demographics_resp_activityRetired             6.63e-13 ***
demographics_resp_activityStudying            2.25e-10 ***
demographics_resp_activityTrainee              0.02132 *
demographics_resp_activityUnemployed          1.70e-06 ***
income_social_protection_host_govt             0.00696 **
income_remmittances                            < 2e-16 ***
income_social_protection_ukr_govt             5.42e-09 ***
diff_lack_of_lang                              < 2e-16 ***
diff_lack_childcare                           5.29e-12 ***
diff_lack_of_skills                            0.00274 **
```

```
diff_lack_of_education_skills                         8.38e-07 ***
diff_none                                              < 2e-16 ***
diff_not_looking_for_work                              < 2e-16 ***
diff_lack_of_age_opportunities                         0.01058 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 9386.3  on 6909  degrees of freedom
Residual deviance: 6369.8  on 6890  degrees of freedom
AIC: 6409.8

Number of Fisher Scoring iterations: 5
```

Model performance

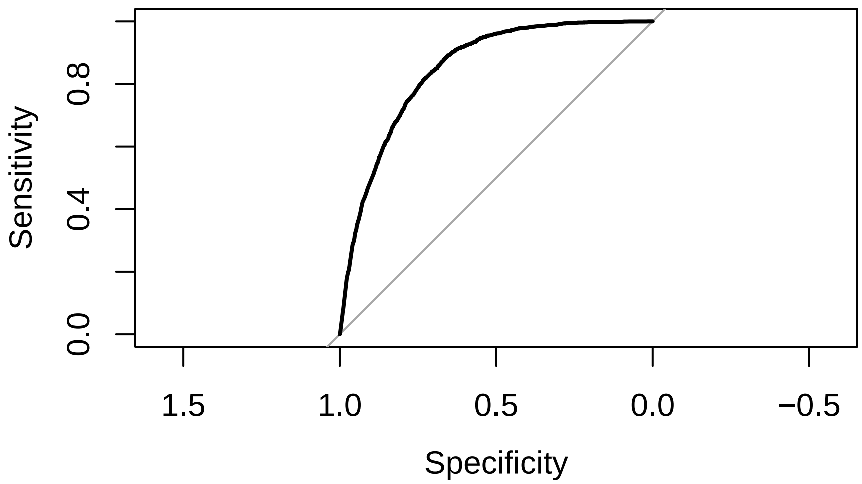
In terms of model performance, the McFadden pseudo R^2 was 0.321, indicating a substantial improvement in model fit compared to the null model, and suggesting that the selected predictors explain a meaningful portion of the variation in employment outcomes.In terms of predictive accuracy, the model correctly classified approximately 79.3% of cases. The confusion matrix showed 1886 true negatives and 3593 true positives, with relatively balanced false positives (993) and false negatives (438). The area under the ROC curve (AUC) was 0.848, reflecting strong discriminative ability and indicating that the model effectively distinguishes between employed and unemployed individuals based on the included variables.

```
fitting null model for pseudo-r2
```

```
          llh       llhNull            G2      McFadden          r2ML 
-3184.9182312 -4693.1693810  3016.5022996     0.3213716     0.3537324 
         r2CU 
    0.4761375 
```

```
         Actual
Predicted    0    1
        0 1886  438
        1  993 3593
```

```
[1] 0.7929088
```




```
Area under the curve: 0.848
```

Estonia

Logistic regression

A logistic regression model was fitted to assess the predictors of employment among Ukrainian refugees in Estonia. Compared to the reference group, respondents previously engaged in housekeeping, while living in Ukraine, had a 91.5% lower odds of employment, while those who were retired had a 96.8% lower odds. Individuals who were studying had a 92.0% lower odds, and those who were unemployed had a 66.0% lower odds of being employed. Respondents who identified their main activity as "Other" also had a 75.7% lower odds.

Some barriers and support factors also had strong negative associations with employment. Reporting lack of documentation was associated with a 94.6% decrease in odds of employment, while receiving cash benefits or an unemployment grant was linked to an 88.1% and an 85.3% decrease in employment odds. Those who reported a need for employment support had an 85.2% lower odds of being employed.

Conversely, some variables were positively associated with employment. Surprisingly, those who reported language barriers had a 131.4% higher odds of employment. Experiencing hostile comments on social media was associated with a 533.1% increase in the odds of employment, and perceiving a loss of benefits was linked to a 1,628% increase in odds, though the latter should be interpreted cautiously due to wide error margins.

```
Call:
glm(formula = employed_binary ~ ., family = binomial, data = estonia_data)

Coefficients:
                                      Estimate Std. Error z value Pr(>|z|)
(Intercept)                             1.2360     0.1988   6.217 5.06e-10 ***
demographics_resp_activityHousekeeping -2.3784     0.4171  -5.702 1.18e-08 ***
demographics_resp_activityOther        -1.4186     0.6311  -2.248 0.024585 *
demographics_resp_activityRetired      -3.4393     0.6679  -5.149 2.61e-07 ***
demographics_resp_activityStudying     -2.5225     0.4597  -5.487 4.09e-08 ***
demographics_resp_activityUnemployed   -1.0777     0.4744  -2.272 0.023103 *
diff_lack_of_lang                       0.8389     0.2518   3.332 0.000862 ***
diff_lack_of_documentation             -2.9101     1.2807  -2.272 0.023070 *
benefits_cash_benefits                 -2.1139     0.3006  -7.032 2.03e-12 ***
benefits_unemployment_grant            -1.9297     0.4701  -4.105 4.04e-05 ***
hostile_comments_social_media           1.8576     0.4340   4.281 1.86e-05 ***
diff_loss_benefits                      2.8724     1.1600   2.476 0.013277 *
needs_employment                       -1.8743     0.3113  -6.021 1.74e-09 ***
---
```

```
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 768.45  on 555  degrees of freedom
Residual deviance: 467.63  on 543  degrees of freedom
AIC: 493.63

Number of Fisher Scoring iterations: 6
```

Model performance

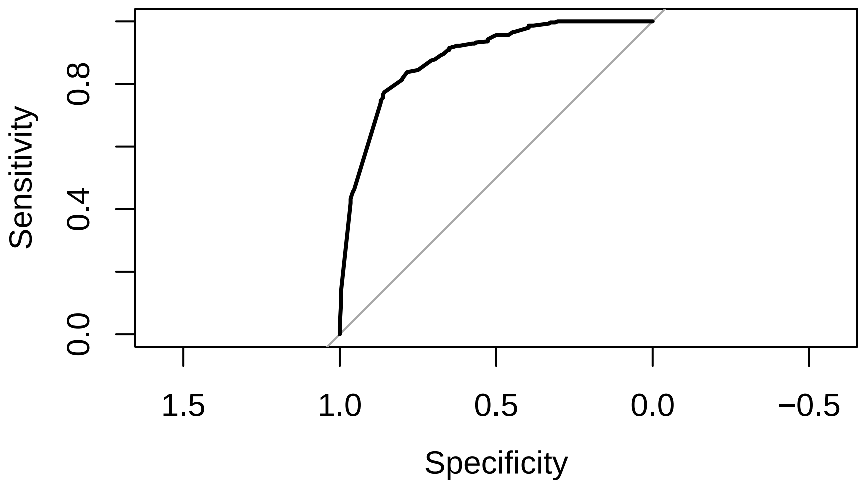
The Estonia model demonstrated strong overall performance. The McFadden pseudo-R^2 was 0.39, indicating a good model fit relative to the null model. The classification accuracy was 80.0%, correctly predicting 250 out of 296 employed individuals and 195 out of 260 unemployed individuals. The confusion matrix showed a balance between sensitivity and specificity. The Area Under the Receiver Operating Characteristic Curve (AUC) was 0.884, suggesting high discriminative ability between employed and unemployed groups. These results indicate that the model effectively distinguishes between employment statuses based on the selected predictors.

```
fitting null model for pseudo-r2

         llh      llhNull           G2     McFadden         r2ML         r2CU
-233.8139308 -384.2235491  300.8192365    0.3914638    0.4178585    0.5579266

         Actual
Predicted   0   1
        0 195  46
        1  65 250

[1] 0.8003597
```




```
Area under the curve: 0.8842
```

Czechia

Logistic regression

In Czechia, respondents who were engaged in housekeeping were 73.6% less likely to be employed than those working. Those reporting a long-term illness or injury were 98.2% less likely to be employed, while retired individuals had 86.9% lower odds of being employed. Respondents who were studying had 70.1% lower odds of employment, and those who were unemployed at the time of the survey had 68.1% lower odds of being employed.

Access to social protection from the host government was negatively associated with employment: recipients were 78.2% less likely to be employed. On the other hand, individuals who reported no major difficulties integrating or accessing work were nearly 6.5 times more likely to be employed. Experiences of verbal aggression were positively associated with employment, with those experiencing it having 85.7% higher odds of being employed.

Difficulties such as lack of decent employment and lack of childcare were also significant: those who faced a lack of decent employment had 53.4% higher odds of being employed, whereas those who reported childcare difficulties had 46.7% lower odds of being employed. Finally, individuals who were not actively looking for work had 87.0% lower odds of being employed, indicating a strong behavioral link to employment outcomes. These findings highlight both structural and personal factors influencing refugee employment in the Czech context.

```
Call:
glm(formula = employed_binary ~ ., family = binomial, data = czechia_data)

Coefficients:
                                                Estimate Std. Error z value
(Intercept)                                       1.3440     0.1974   6.808
demographics_resp_activityHousekeeping           -1.2947     0.2415  -5.361
demographics_resp_activityLong term illness/injury  -4.0272     0.7728  -5.212
demographics_resp_activityOther                  -0.0595     0.4072  -0.146
demographics_resp_activityRetired                -2.0217     0.4088  -4.945
demographics_resp_activityStudying               -1.2062     0.2364  -5.102
demographics_resp_activityTrainee                -0.7219     0.4587  -1.574
demographics_resp_activityUnemployed             -1.1449     0.4022  -2.846
income_social_protection_host_govt               -1.5296     0.1579  -9.689
diff_lack_of_lang                                 0.4990     0.1812   2.754
diff_none                                         1.8667     0.2523   7.398
hostile_verbal_aggression                         0.6194     0.1941   3.191
diff_lack_of_decent_employment                    0.4282     0.1925   2.224
diff_lack_childcare                              -0.6328     0.2586  -2.447
```

```
diff_not_looking_for_work                          -2.0256     0.2688  -7.536
                                                   Pr(>|z|)
(Intercept)                                        9.91e-12 ***
demographics_resp_activityHousekeeping             8.29e-08 ***
demographics_resp_activityLong term illness/injury 1.87e-07 ***
demographics_resp_activityOther                     0.88383
demographics_resp_activityRetired                  7.60e-07 ***
demographics_resp_activityStudying                 3.36e-07 ***
demographics_resp_activityTrainee                   0.11553
demographics_resp_activityUnemployed                0.00442 **
income_social_protection_host_govt                  < 2e-16 ***
diff_lack_of_lang                                   0.00589 **
diff_none                                          1.39e-13 ***
hostile_verbal_aggression                           0.00142 **
diff_lack_of_decent_employment                      0.02613 *
diff_lack_childcare                                 0.01439 *
diff_not_looking_for_work                          4.84e-14 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 1860.0  on 1486  degrees of freedom
Residual deviance: 1145.3  on 1472  degrees of freedom
AIC: 1175.3

Number of Fisher Scoring iterations: 6
```

Model performance

The logistic regression model for Czechia demonstrated strong performance in predicting employment status among Ukrainian refugees. The McFadden pseudo-R^2 was 0.384, indicating a substantial improvement over the null model and suggesting that the predictors explained a meaningful portion of the variation in employment outcomes. The area under the ROC curve (AUC) was 0.878, which reflects excellent discriminatory ability of the model in distinguishing between employed and unemployed individuals.

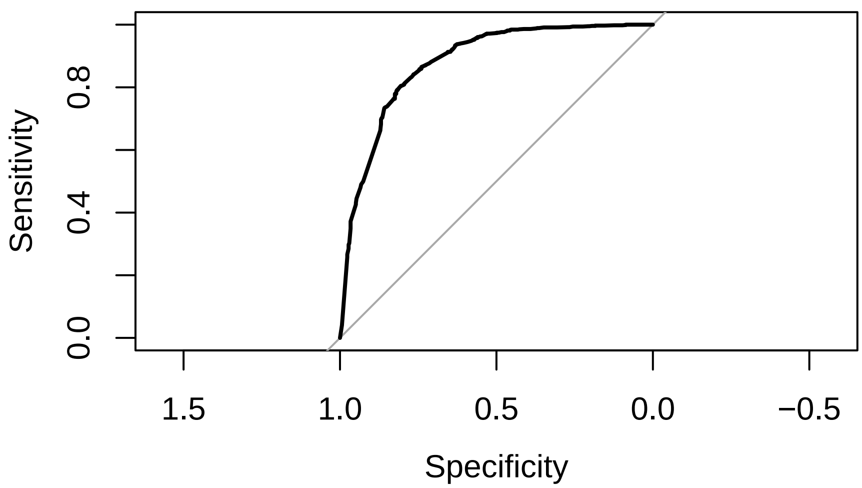
In terms of classification accuracy, the model correctly predicted 83.8% of all cases. Among the unemployed, it correctly classified 295 out of 358 individuals (82.4%), while for the employed group, it correctly predicted 951 out of 1,129 cases (84.2%). These results suggest that the model is both accurate and reliable in identifying employment patterns within this population.

```
fitting null model for pseudo-r2

          llh       llhNull            G2      McFadden          r2ML          r2CU
-572.6510292 -930.0016949  714.7013315     0.3842473     0.3816082     0.5346625

         Actual
Predicted   0   1
        0 295  63
        1 178 951

[1] 0.8379287
```




```
Area under the curve: 0.8782
```

Slovakia

Logistic regression

The logistic regression analysis for Slovakia identified that being aged 60 or older reduced the odds of employment by approximately 67.2%. Compared to other activity categories, engaging in housekeeping decreased the odds of employment by 79.5%, having a long-term illness or injury by 78.2%, being retired by 81.4%, studying by 70.9%, and being unemployed by 65.1%. Receipt of social protection benefits from the host government was associated with a 50.1% lower likelihood of employment. Not looking for work was linked to a 73.4% decrease in employment odds, while needing host government assistance or humanitarian assistance decreased the odds by 46.5% and 42.3%, respectively. Reporting no barriers to employment increased the odds by 400.5%, and surprisingly, reporting a lack of skills increased the odds by 78.0%. Finally, receiving cash benefits showed a marginal negative association, with odds of employment decreased by 52.8%.

These results highlight that older age, non-working activity statuses back in Ukraine, and reliance on benefits or assistance are strongly linked to lower employment probabilities, while self-reported absence of barriers greatly improves employment odds. The positive association with reported lack of skills can be linked to underlying dynamics in skill recognition.

```
Call:
glm(formula = employed_binary ~ ., family = binomial, data = slovakia_data)

Coefficients:
                                               Estimate Std. Error z value
(Intercept)                                    0.735414   0.210802   3.489
introduction_resp_age35-59                     0.121755   0.200053   0.609
introduction_resp_age60+                      -1.079908   0.390158  -2.768
demographics_resp_activityHousekeeping        -1.587135   0.337215  -4.707
demographics_resp_activityLong term illness/injury -1.520569   0.756290  -2.011
demographics_resp_activityOther               -0.005407   0.566287  -0.010
demographics_resp_activityRetired             -1.684344   0.514602  -3.273
demographics_resp_activityStudying            -1.203479   0.267835  -4.493
demographics_resp_activityTrainee             -1.491922   0.783914  -1.903
demographics_resp_activityUnemployed          -1.086592   0.406088  -2.676
income_social_protection_host_govt            -0.704119   0.291831  -2.413
diff_lack_of_skills                            0.576793   0.281010   2.053
diff_not_looking_for_work                     -1.323255   0.267911  -4.939
needs_host_govt_assistance                    -0.626295   0.188424  -3.324
needs_humanitarian_assistance                 -0.550940   0.189311  -2.910
diff_none                                      1.610766   0.182710   8.816
```

```
benefits_cash_benefits                             -0.743778   0.389486  -1.910
                                                   Pr(>|z|)    
(Intercept)                                        0.000485 ***
introduction_resp_age35-59                         0.542782    
introduction_resp_age60+                           0.005642 ** 
demographics_resp_activityHousekeeping             2.52e-06 ***
demographics_resp_activityLong term illness/injury 0.044372 *  
demographics_resp_activityOther                    0.992382    
demographics_resp_activityRetired                  0.001064 ** 
demographics_resp_activityStudying                 7.01e-06 ***
demographics_resp_activityTrainee                  0.057018 .  
demographics_resp_activityUnemployed               0.007456 ** 
income_social_protection_host_govt                 0.015832 *  
diff_lack_of_skills                                0.040114 *  
diff_not_looking_for_work                          7.85e-07 ***
needs_host_govt_assistance                         0.000888 ***
needs_humanitarian_assistance                      0.003612 ** 
diff_none                                           < 2e-16 ***
benefits_cash_benefits                             0.056180 .  
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 1425.3  on 1038  degrees of freedom
Residual deviance: 1000.8  on 1022  degrees of freedom
AIC: 1034.8

Number of Fisher Scoring iterations: 5
```

Model performance

The logistic regression model for Slovakia showed a moderate fit, with a McFadden's pseudo-R^2 of approximately 0.298, indicating that about 29.8% of the variability in employment status is explained by the predictors. Classification accuracy was 77.2%, with the model correctly predicting employment status for the majority of cases. The area under the ROC curve (AUC) was 0.843, indicating good discrimination ability between employed and unemployed individuals.

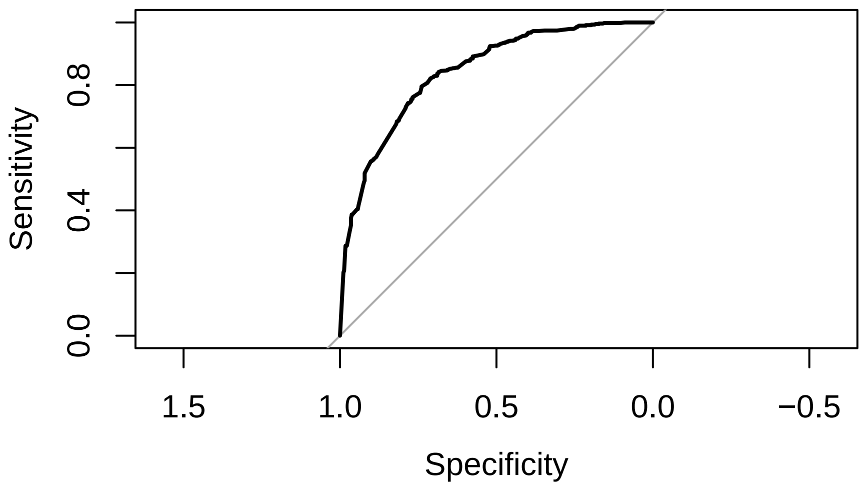
```
fitting null model for pseudo-r2

         llh      llhNull           G2     McFadden         r2ML          r2CU
```

```
-500.3917968 -712.6424265  424.5012595    0.2978361    0.3353982    0.4493868
```

```
         Actual
Predicted   0   1
        0 311  91
        1 146 491
```

```
[1] 0.7718961
```




```
Area under the curve: 0.8431
```

Moldova

Logistic regression

The logistic regression model for Moldova showed that being engaged in housekeeping in Ukraine was associated with a 86.5% decrease in the odds of employment, while having a long-term illness or injury reduced the odds by 93.1%. Respondents reporting “other” activities had a 65.3% lower chance of being employed, and those studying had their odds reduced by 63.2%. Those identified as trainees showed a 79.1% decrease in odds.

Conversely, lacking language skills increased the odds of employment by 336%, and lack of skills was also positively associated, increasing the odds by 297%. Lack of information had the strongest positive effect, increasing employment odds by 169% (OR = 17.18). Respondents reporting no particular difficulties had odds more than 10 times higher. Medical needs and caregiving responsibilities negatively impacted employment odds, reducing them by 50.3% and 87% respectively.

```
Call:
glm(formula = employed_binary ~ ., family = binomial, data = moldova_data)

Coefficients:
                                               Estimate Std. Error z value
(Intercept)                                     -0.9833     0.1848  -5.320
demographics_resp_activityHousekeeping          -2.0080     0.2682  -7.488
demographics_resp_activityLong term illness/injury  -2.6039     1.0590  -2.459
demographics_resp_activityOther                 -1.0485     0.5190  -2.020
demographics_resp_activityRetired              -17.0721   517.3108  -0.033
demographics_resp_activityStudying              -1.0026     0.2611  -3.840
demographics_resp_activityTrainee               -1.5737     0.7118  -2.211
diff_lack_of_lang                                1.4756     0.2359   6.256
diff_lack_of_skills                              1.3830     0.4008   3.450
diff_lack_of_info                                2.8429     0.8857   3.210
diff_none                                        2.3018     0.1974  11.662
needs_medical                                   -0.6993     0.1590  -4.398
diff_need_to_take_care_of_others                -2.0762     0.6061  -3.425
needs_humanitarian_assistance                   -0.2771     0.1521  -1.821
                                               Pr(>|z|)
(Intercept)                                    1.04e-07 ***
demographics_resp_activityHousekeeping         7.00e-14 ***
demographics_resp_activityLong term illness/injury 0.013944 *
demographics_resp_activityOther                0.043380 *
demographics_resp_activityRetired              0.973673
```

```
demographics_resp_activityStudying                0.000123 ***
demographics_resp_activityTrainee                 0.027040 *
diff_lack_of_lang                                 3.95e-10 ***
diff_lack_of_skills                               0.000560 ***
diff_lack_of_info                                 0.001329 **
diff_none                                          < 2e-16 ***
needs_medical                                     1.09e-05 ***
diff_need_to_take_care_of_others                  0.000614 ***
needs_humanitarian_assistance                     0.068569 .
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 1497.3  on 1119  degrees of freedom
Residual deviance: 1071.2  on 1106  degrees of freedom
AIC: 1099.2

Number of Fisher Scoring iterations: 16
```

Model performance

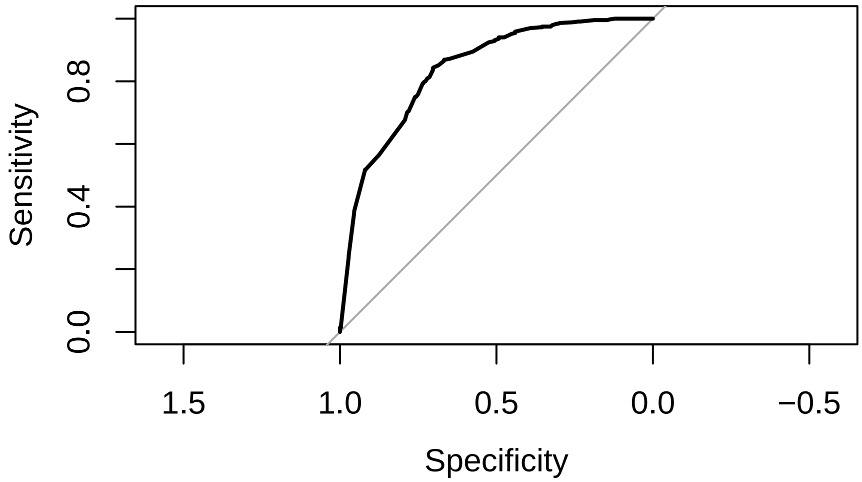
The logistic regression model for Moldova demonstrated a reasonable fit, with a McFadden's pseudo-R^2 of approximately 0.285, indicating that about 28.5% of the variation in employment status is explained by the model. The classification accuracy was 75.4%. Additionally, the model showed good discrimination ability with an area under the ROC curve (AUC) of 0.84, suggesting a strong capacity to distinguish between employed and unemployed respondents.

```
fitting null model for pseudo-r2

         llh      llhNull           G2     McFadden         r2ML         r2CU
-535.6231813 -748.6388062  426.0312499    0.2845372    0.3164019    0.4291187

         Actual
Predicted   0   1
        0 513 105
        1 171 331

[1] 0.7535714
```




```
Area under the curve: 0.84
```

Romania

Logistic regression

The logistic regression model for Romania reveals several significant predictors of employment status. Being aged 60 or older reduces the odds of being employed by about 60%. Respondents engaged in housekeeping activities have approximately 82% lower odds of employment. Those with long-term illness or injury show a strong negative effect, with odds reduced by about 97%. Other activity categories such as "Other," "Retired," and "Studying" also significantly reduce employment odds by roughly 45%, 66%, and 54% respectively.

Lack of childcare decreases employment odds by about 66%, while lacking skills reduces odds by about 38%. Respondents needing to care for others have about 57% lower odds, and those not looking for work show a large negative association with employment, with 73% lower odds. Conversely, respondents reporting no barriers to employment ("diff_none") have more than 4 times higher odds of being employed. Interestingly, needing host government assistance is associated with a 49% increase in employment odds, whereas income from other sources reduces employment odds by about 57%.

```
Call:
glm(formula = employed_binary ~ ., family = binomial, data = romania_data)

Coefficients:
                                             Estimate Std. Error z value
(Intercept)                                   -0.3581     0.1969  -1.819
introduction_resp_age35-59                    -0.1837     0.1384  -1.328
introduction_resp_age60+                      -0.9131     0.3420  -2.670
demographics_resp_activityHousekeeping        -1.6896     0.2206  -7.661
demographics_resp_activityLong term illness/injury -3.4553  1.0334  -3.344
demographics_resp_activityOther               -0.6077     0.2873  -2.115
demographics_resp_activityRetired             -1.0784     0.4908  -2.197
demographics_resp_activityStudying            -0.7808     0.2850  -2.740
demographics_resp_activityTrainee             -0.2453     0.3828  -0.641
demographics_resp_activityUnemployed          -0.7930     0.5770  -1.374
diff_lack_of_decent_employment                 0.2715     0.1541   1.762
diff_lack_childcare                           -1.0849     0.3599  -3.015
diff_none                                      1.5422     0.1599   9.647
needs_host_govt_assistance                     0.3998     0.1603   2.494
income_other_sources                          -0.8606     0.1487  -5.787
diff_lack_of_skills                           -0.4809     0.1729  -2.782
diff_need_to_take_care_of_others              -0.8419     0.2580  -3.264
diff_not_looking_for_work                     -1.3312     0.2136  -6.232
```

```
                                                   Pr(>|z|)
(Intercept)                                        0.068986 .
introduction_resp_age35-59                         0.184322
introduction_resp_age60+                           0.007586 **
demographics_resp_activityHousekeeping             1.85e-14 ***
demographics_resp_activityLong term illness/injury 0.000827 ***
demographics_resp_activityOther                    0.034414 *
demographics_resp_activityRetired                  0.028008 *
demographics_resp_activityStudying                 0.006143 **
demographics_resp_activityTrainee                  0.521733
demographics_resp_activityUnemployed               0.169320
diff_lack_of_decent_employment                     0.078151 .
diff_lack_childcare                                0.002573 **
diff_none                                           < 2e-16 ***
needs_host_govt_assistance                         0.012644 *
income_other_sources                               7.18e-09 ***
diff_lack_of_skills                                0.005410 **
diff_need_to_take_care_of_others                   0.001100 **
diff_not_looking_for_work                          4.62e-10 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 2255.5  on 1777  degrees of freedom
Residual deviance: 1726.4  on 1760  degrees of freedom
AIC: 1762.4

Number of Fisher Scoring iterations: 6
```

Model performance

The model shows a McFadden's pseudo-R^2 of approximately 0.23, indicating a moderate fit. The likelihood ratio test statistic (G^2) of 529.08 suggests the model fits significantly better than the null model. Classification accuracy is around 77.2%, with an area under the ROC curve (AUC) of 0.81, indicating good discrimination ability between employed and unemployed respondents.

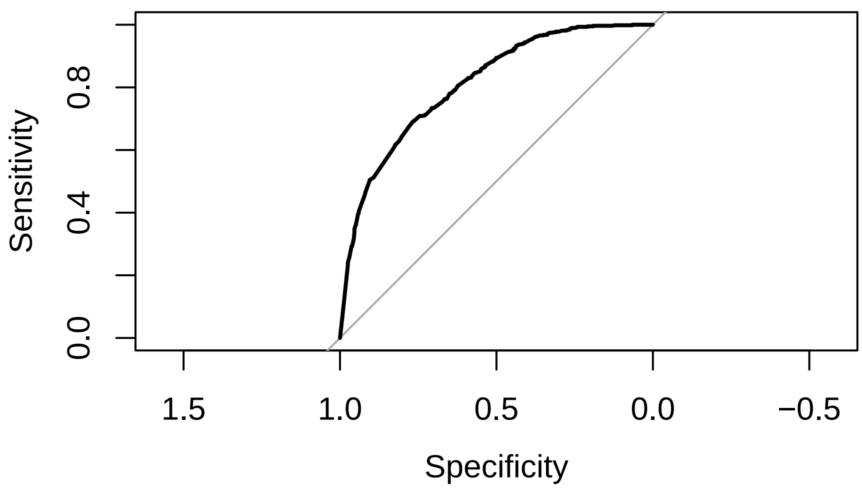
```
fitting null model for pseudo-r2

          llh       llhNull            G2      McFadden          r2ML
```

```
 -863.2127389 -1127.7536564   529.0818350     0.2345733     0.2573804
         r2CU
    0.3580870

         Actual
Predicted    0    1
        0 1078  292
        1  113  295

[1] 0.772216
```




```
Area under the curve: 0.8102
```

Discussion

Limitations

Conclusion

Citations

- SocioFactor, IOM, Impact-REACH, SHC, Sociofactor, TARKI Social Research Institute, Ipsos, UNHCR (2023). Poland, Slovak Republic, Hungary, Czech Republic, Moldova, Romania, Bulgaria: Multi-Sector Needs Assessment (MSNA) - 2023. Accessed from: https://microdata.unhcr.org
- UNHCR (2023). Estonia: Multi-Sector Needs Assessment (MSNA) - 2023. Accessed from: https://microdata.unhcr.org
- Preut, Holger. " : ." Zeitschrift Der Koreanisch-Deutschen Gesellschaft Für Sozialwissenschaften, vol. 33, no. 3, Koreanisch-Deutsche Gesellschaft fuer Sozlaiwssenschaften, Sept. 2023, pp. 3–33. Crossref, doi:10.19032/zkdgs.2023.09.33.3.3.
- Londar, S., et al. "Challenges for Ukrainian Refugees in the EU Labour Market: The Case of Poland and Germany." Educational Analytics of Ukraine, no. 5, State Scientific Institution - Institute of Educational Analytics, 2024, pp. 5–16. Crossref, doi:10.32987/2617-8532-2024-5-5-16.