**Data Analysis Plan (Work in Progress)**

[Reintegration Economic Survey (RES) 2](#_Toc137464542)

[Reintegration Sustainability Survey (RSS) 5](#_Toc137464543)

[Appendix 1: Overall Frequencies (RES) 7](#_Toc137464544)

[Meta Variables 7](#_Toc137464545)

[Dependent Variables 8](#_Toc137464546)

[Independent Variables 11](#_Toc137464547)

[Appendix 2: Sample Binary Logistic Regression Model (Business Success) 19](#_Toc137464548)

[Appendix 3: Frequencies of Mission (RSS) 22](#_Toc137464549)

[Appendix 4: Distribution of RSS Scores 22](#_Toc137464550)

Sources to use for RSS:

MIMOSA VARIABLES

KOBO VARIABLES

# Reintegration Economic Survey (RES)

All dependent variables of interest are categorical, and I therefore plan to use logistic regression (more precisely, binary logistic regression to avoid factor imbalance and easier interpretation). A nice property of logistic regression is that we can compute odds ratio, which are easy to understand for non-specialists. Likewise, model performance can be translated to percentages, which is intuitive and straightforward for interpretation.

A sample logistic regression model, which predicts Business Success, is presented in Appendix II (this is for illustrative purposes only, since I did not check assumptions yet).

As other regression-based approaches, logistic regression comes with assumptions, some of which are mentioned below. The key message here is that because of these assumptions, some planned analyses might not be feasible (worst case scenario), or, more likely, with a smaller number of predictors (or comparisons) than initially planned. It is important to keep that in mind when reading this data analysis plan.

Prerequisites and potential caveats

* Data preparation. Dependent variables will likely be recoded to binary. Other variables might be dummy-coded. Since data are mostly categorical, it will difficult to replace NA, and these will most likely be discarded (or, if NA is too high, the entire variable will be dropped from the model); all answers options will need to be translated (to be discussed, since this is not data preparation per se)
* Number of observations. A minimum number of observations is needed for each comparison. Often, this means that we cannot use as many predictors as we would want
* Imbalance. Even when there are enough observations for a given variable or comparison, there might be imbalance, meaning that a category is overrepresented (for example, intention to migrate again in the RES dataset). We can deal with this most of the time, but not always
* Collinearity. Collinearity will be assessed with VIF and Tolerance statistics. Variables who are highly colinear (threshold will be fixed in due time) will be removed from the relevant model
* Model performance. Will be assessed with the Accuracy statistic (rather than McFadden’s R2). Baseline performance will also be reported. For comparison between groups, we will report the odds ratio (with confidence intervals)
* Representativity and inference. These data are not representative, and we cannot therefore make inferences on the population. That said, we can say something about the participants we interviewed. This is a limitation that needs to be acknowledged in all outputs

All that being said, here are the models that I plan to build:

Model 1: Determinants of Business Success

Binary logistic regression will be used to identify the significant predictors of whether a business is rather successful or rather unsuccessful.

Dependent variable:

* "Comment se porte votre entreprise ou business actuellement ?"

Predictor/control variables:

* "Pays",
* "Pays d’où le migrant est de retour :",
* "Sexe",
* "Age (l'enquête est destinée aux personnes agées de 14 ans et plus)",
* "Durée de l’absence du pays d’origine Mettre 0 si moins d'un an",
* "Situation de handicap",
* "Combien de temps entre votre retour et la réception de l’aide à la réintégration (ou sa première fourniture) ? En semaines",
* "Par quel moyen avez-vous reçu cette assistance économique ?",
* ?",
* "Type de business bis",
* "Qui sont les membres de cette entreprise ?",
* "Niveau microbusiness",
* "L'OIM ou un de ses partenaires vous a-t-elle formé sur la façon de gérer une entreprise ?",
* "L’entreprise emploie-t-elle du personnel ?",
* "Si oui, combien des personnes sont employées par votre entreprise ?",
* "Est-ce votre entreprise a été affectée par la maladie de Coronavirus ?" (I think that using just one control on Covid is a good idea)
* Est-ce que le type d'assistance économique que vous avez reçu correspondait à votre premier choix ?
* Délai entre la réception de l’assistance et je jour d’enquête

Model 2: Determinants of Business Profitability

Same as Model 1, except that the dependent variable will be:

* "L’entreprise vous permet -elle de gagner assez d’argent pour subvenir à vos besoins et à celle de votre famille ?"

Model 3: Impact of Training on Business Success

Same as Model 1, except that data will be subset to participants who received training only (indeed, 1,401 observations are missing).

Dependent variables:

* "Comment se porte votre entreprise ou business actuellement ?"
* If time permits: "L’entreprise vous permet -elle de gagner assez d’argent pour subvenir à vos besoins et à celle de votre famille ?"

Independent variables:

* "Type de formation"
* Training duration (to be computed)
* As many predictor variables from Model 1 as possible

Model 4 (if time permits): Determinants of Future Intentions to Migrate

Dependent variable:

* "Avez-vous déjà planifié de migrer de nouveau ?"

Predictor variables

* As many from Model 1 and 3 as statistically sound

Model 5 (if time permits): Determinants of Support Satisfaction

I believe it would be interesting to say something about the respondent’s satisfaction with the programme overall. It would also be interesting to compare this with Business Success.

Dependent variable:

* "Êtes-vous satisfait de l’aide à la réintégration de manière globale ?" OR
* "Pensez-vous que le retour a été une bonne décision ?"

Independent variables:

* As many as statistically sound
* Also include Business Succes and/or Business Profitability (this time, as predictor variables) in an alternative model

# Reintegration Sustainability Survey (RSS)

The main dependent variable of interest (Composite Reintegration Score) is continuous, and I therefore plan to use multiple linear regression. I might also use ANOVA, although regression is preferable to control for the influence of extraneous variables.

As for the RES questionnaire, I will first check assumptions, including assumptions of normality and linearity. The RSS data are more messy and will require more manipulations than the RES data. I will also recompute all the weighted scores to check for anomalies.

To measure model performance, I’ll use R2, which expresses the proportion of the variance that is explained by the predictors.

Model 6: Determinants of sustainable reintegration (Composite)

Dependent variable:

* Composite reintegration score

Predictor/control variables:

* ~~Assistance reçue, composée de:~~
  + ~~Counselling~~
  + ~~economic support~~
  + ~~financial services~~
  + ~~job placement~~
  + ~~microbusiness~~
  + ~~training~~
  + ~~training duration computed from training start date/end date social support~~
  + ~~childcare~~
  + ~~education~~
  + ~~housing~~
  + ~~legal services~~
  + ~~material assistance~~
  + ~~medical support~~
  + ~~social protection~~
  + ~~psychosocial support~~
* ~~"Sexe",~~
* ~~"Age",~~
* ~~"Durée de l’absence du pays d’origine Mettre 0 si moins d'un an",~~
* ~~"Pays de retour :",~~
* ~~pays d’origine =>"Mission OIM" dans la base kobo~~
* ~~vulnérabilités ( 3 variables distinctes: VoTs, UMINIR, Healthcondition)~~
* ~~microbusiness level~~
* ~~microbusiness delivered by IOM or referral~~
* ~~microbusiness assistance received in cash versus reception in-kind or mixte (MicrobusinessFormOfAssistance)~~
* ~~delai entre retour (arrival date) et reception assistance microbusiness (MicrobusinessEndDate)~~
* delai entre reception assistance microbusiness(MicrobusinessEndDate) et enquête kobo

Model 7: Determinants of sustainable reintegration (Economic Reintegration Score)

Dependent variable:

* Economic Reintegration Score

Predictor/control variables:

* Same as Model 6

# Appendix 1: Overall Frequencies (RES)

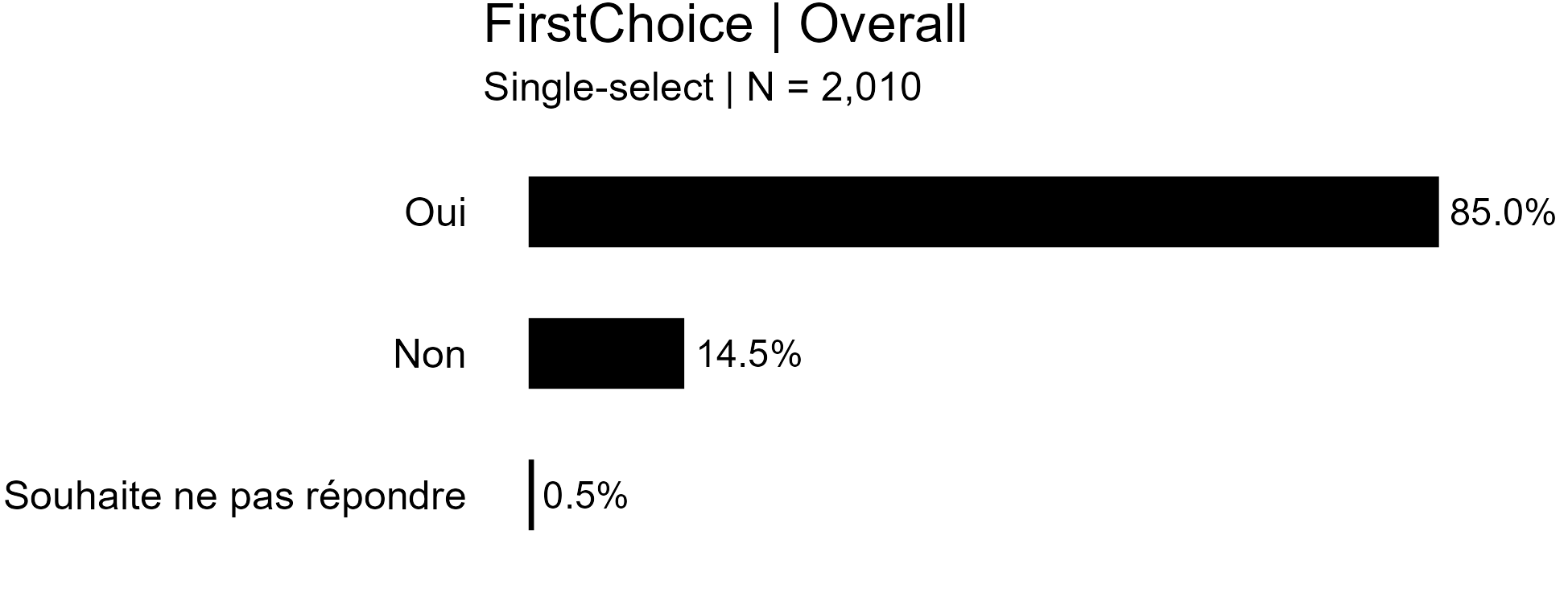
## Meta Variables

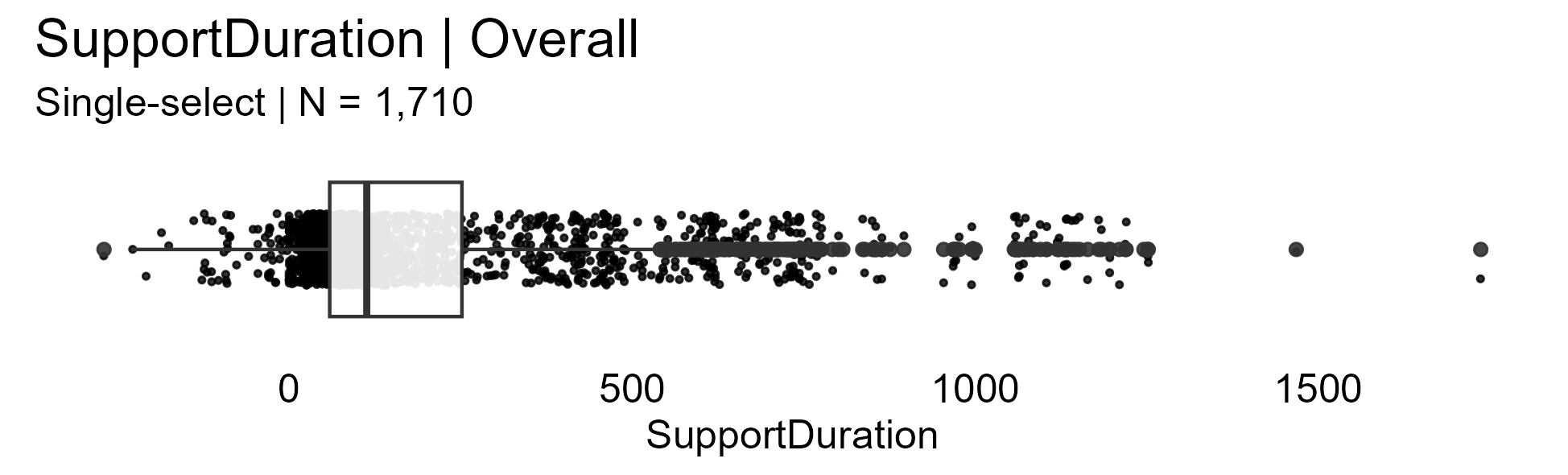
A picture containing black, darkness

Description automatically generated

A picture containing black, darkness

Description automatically generated





## Dependent Variables

A picture containing black, darkness

Description automatically generated

A picture containing black, darkness

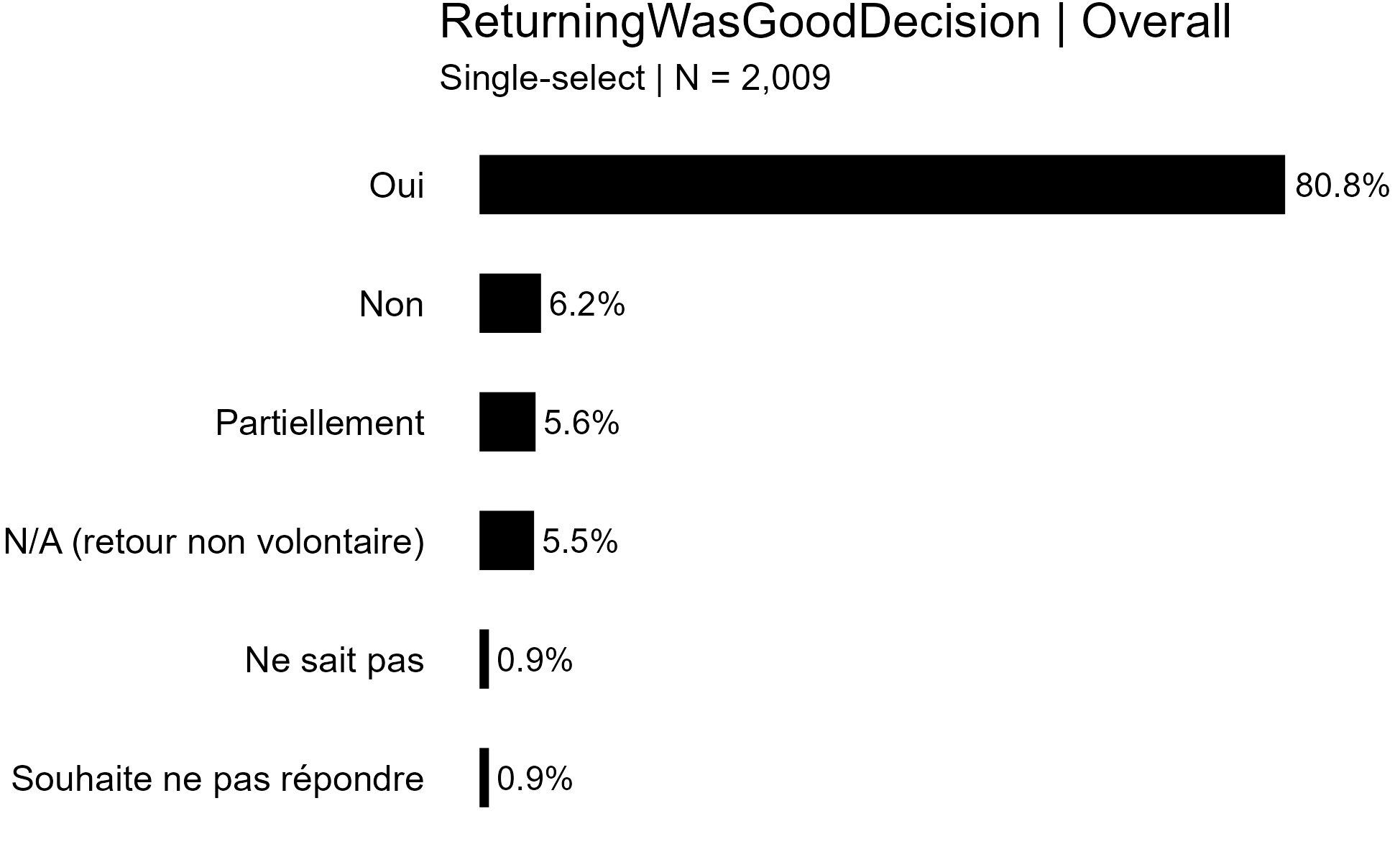
Description automatically generated

A picture containing black, darkness

Description automatically generated

A picture containing black, darkness

Description automatically generated



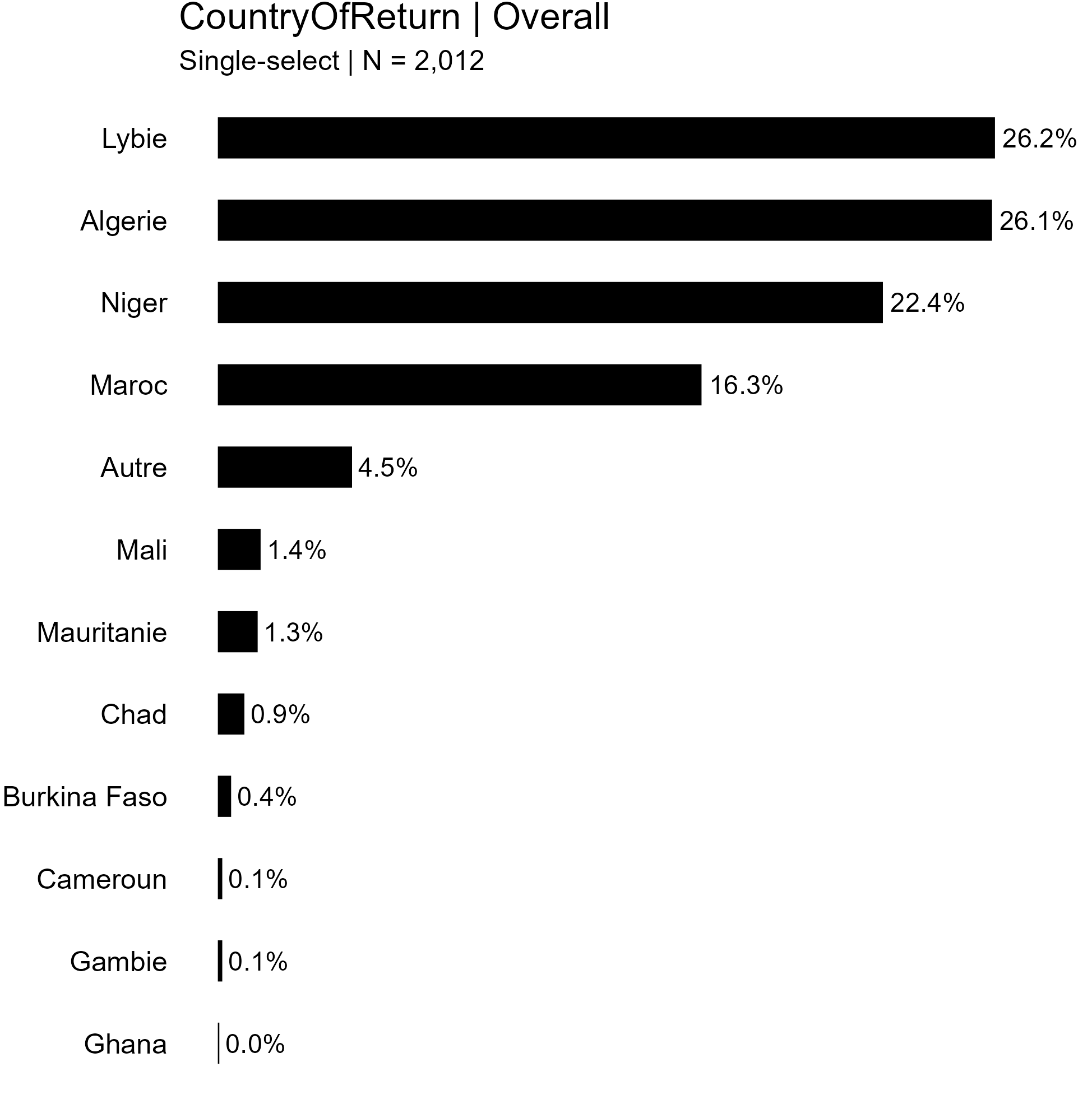
A picture containing black, darkness

Description automatically generated

## Independent Variables

A picture containing black, darkness

Description automatically generated



A picture containing black, darkness

Description automatically generated

A picture containing black, darkness

Description automatically generated

A picture containing black, darkness, screenshot, black and white

Description automatically generated

A picture containing black, darkness

Description automatically generated

A picture containing black, darkness

Description automatically generated

A picture containing screenshot, black

Description automatically generated

A picture containing black, darkness

Description automatically generated

A picture containing black, darkness

Description automatically generated

A picture containing black, darkness

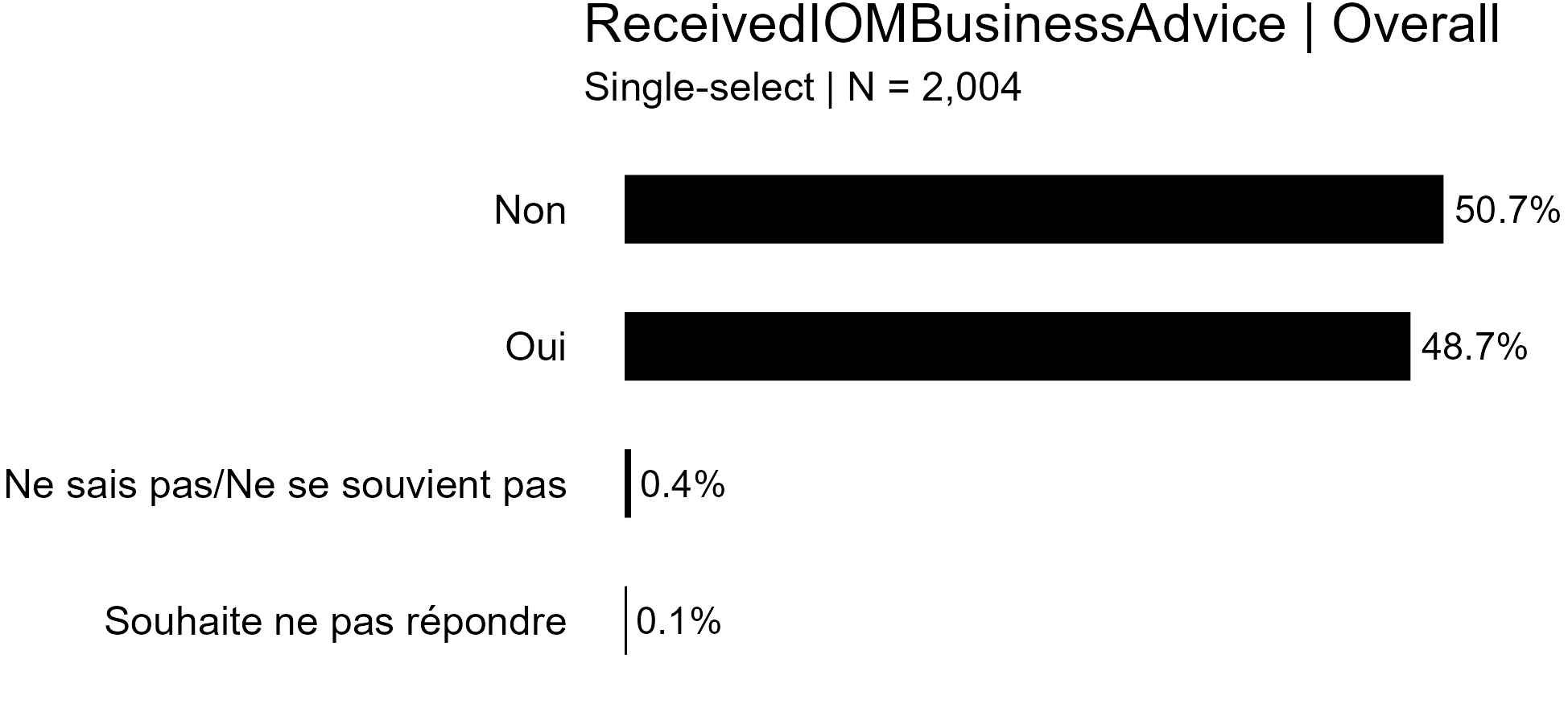
Description automatically generated

A picture containing black, darkness

Description automatically generated

A picture containing black, darkness

Description automatically generated



A picture containing black, darkness

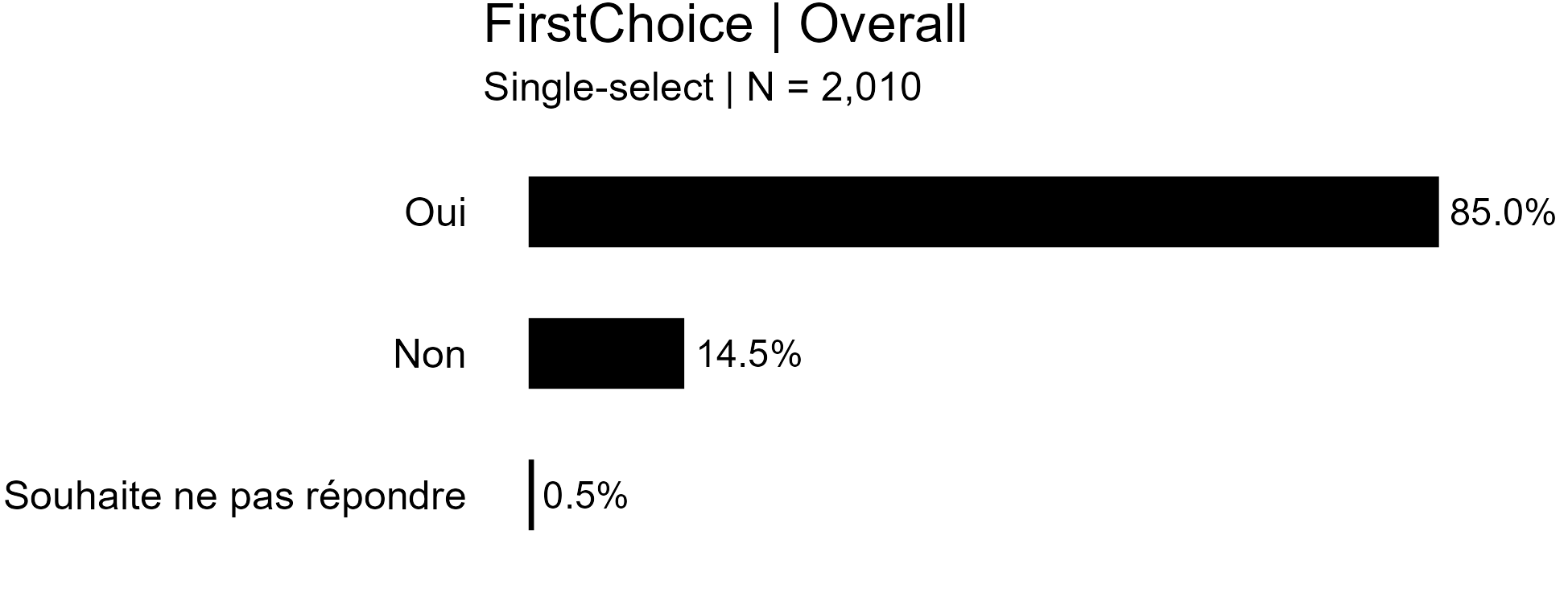
Description automatically generated

A picture containing black, darkness

Description automatically generated

A picture containing black, darkness

Description automatically generated



# Appendix 2: Sample Binary Logistic Regression Model (Business Success)

**Binomial Logistic Regression**

| Model Fit Measures | | | | | | | | | | | | | |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | | | | | | | | **Overall Model Test** | | | | | |
| **Model** | | **Deviance** | | **AIC** | | **R²McF** | | **χ²** | | **df** | | **p** | |
| 1 |  | 2375 |  | 2463 |  | 0.135 |  | 370 |  | 43 |  | < .001 |  |
|  | | | | | | | | | | | | | |

| Omnibus Likelihood Ratio Tests | | | | | | | |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  |  |  |  |  |  |  |  |
| **Predictor** | | **χ²** | | **df** | | **p** | |
| Country |  | 168.44 |  | 13 |  | < .001 |  |
| CountryOfReturn |  | 12.37 |  | 11 |  | 0.336 |  |
| Gender |  | 13.53 |  | 2 |  | 0.001 |  |
| AgeGroup |  | 19.24 |  | 3 |  | < .001 |  |
| BusinessType |  | 69.71 |  | 11 |  | < .001 |  |
| Project |  | 1.08 |  | 1 |  | 0.298 |  |
| InterviewType |  | 8.81 |  | 2 |  | 0.012 |  |
|  | | | | | | | |

| Model Coefficients - BusinessSucess | | | | | | | | | | | |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  |  |  |  |  |  |  |  |  |  |  |  |
| **Predictor** | | **Estimate** | | **SE** | | **Z** | | **p** | | **Odds ratio** | |
| Intercept |  | 0.1960 |  | 1.042 |  | 0.1880 |  | 0.851 |  | 1.217 |  |
| Country: |  |  |  |  |  |  |  |  |  |  |  |
| Cameroun – Burkina Faso |  | -14.1541 |  | 606.278 |  | -0.0233 |  | 0.981 |  | 7.13e-7 |  |
| Côte D'Ivoire – Burkina Faso |  | 0.0755 |  | 0.235 |  | 0.3216 |  | 0.748 |  | 1.078 |  |
| Gambie – Burkina Faso |  | -0.9322 |  | 0.277 |  | -3.3663 |  | < .001 |  | 0.394 |  |
| Ghana – Burkina Faso |  | 2.3788 |  | 0.297 |  | 8.0143 |  | < .001 |  | 10.792 |  |
| Guinée – Burkina Faso |  | 0.6500 |  | 0.273 |  | 2.3815 |  | 0.017 |  | 1.916 |  |
| Mali – Burkina Faso |  | 0.1783 |  | 0.348 |  | 0.5126 |  | 0.608 |  | 1.195 |  |
| Mauritanie – Burkina Faso |  | 14.9820 |  | 882.743 |  | 0.0170 |  | 0.986 |  | 3.21e+6 |  |
| Niger – Burkina Faso |  | -13.0686 |  | 882.743 |  | -0.0148 |  | 0.988 |  | 2.11e-6 |  |
| Nigeria – Burkina Faso |  | 0.3670 |  | 0.296 |  | 1.2404 |  | 0.215 |  | 1.443 |  |
| Sierra-Leone – Burkina Faso |  | -0.3447 |  | 0.419 |  | -0.8224 |  | 0.411 |  | 0.708 |  |
| Sénégal – Burkina Faso |  | 0.1562 |  | 0.214 |  | 0.7308 |  | 0.465 |  | 1.169 |  |
| Tchad – Burkina Faso |  | -0.6208 |  | 0.436 |  | -1.4242 |  | 0.154 |  | 0.537 |  |
| Togo – Burkina Faso |  | -0.1471 |  | 0.813 |  | -0.1808 |  | 0.856 |  | 0.863 |  |
| CountryOfReturn: |  |  |  |  |  |  |  |  |  |  |  |
| Autre – Algerie |  | -0.1093 |  | 0.263 |  | -0.4155 |  | 0.678 |  | 0.896 |  |
| Burkina Faso – Algerie |  | -0.0629 |  | 0.783 |  | -0.0803 |  | 0.936 |  | 0.939 |  |
| Cameroun – Algerie |  | 0.8820 |  | 1.243 |  | 0.7094 |  | 0.478 |  | 2.416 |  |
| Chad – Algerie |  | -0.5272 |  | 0.536 |  | -0.9836 |  | 0.325 |  | 0.590 |  |
| Gambie – Algerie |  | -0.3078 |  | 1.253 |  | -0.2457 |  | 0.806 |  | 0.735 |  |
| Ghana – Algerie |  | 14.7543 |  | 882.743 |  | 0.0167 |  | 0.987 |  | 2.56e+6 |  |
| Lybie – Algerie |  | 0.2633 |  | 0.155 |  | 1.6937 |  | 0.090 |  | 1.301 |  |
| Mali – Algerie |  | 0.3406 |  | 0.447 |  | 0.7627 |  | 0.446 |  | 1.406 |  |
| Maroc – Algerie |  | 0.1662 |  | 0.172 |  | 0.9684 |  | 0.333 |  | 1.181 |  |
| Mauritanie – Algerie |  | -0.6963 |  | 0.502 |  | -1.3873 |  | 0.165 |  | 0.498 |  |
| Niger – Algerie |  | 0.2499 |  | 0.154 |  | 1.6238 |  | 0.104 |  | 1.284 |  |
| Gender: |  |  |  |  |  |  |  |  |  |  |  |
| Masculin – Féminin |  | 0.6572 |  | 0.190 |  | 3.4595 |  | < .001 |  | 1.929 |  |
| Ne souhaite pas répondre – Féminin |  | -13.9794 |  | 882.743 |  | -0.0158 |  | 0.987 |  | 8.49e-7 |  |
| AgeGroup: |  |  |  |  |  |  |  |  |  |  |  |
| 14-17 ans – +65 ans |  | 1.1410 |  | 1.082 |  | 1.0541 |  | 0.292 |  | 3.130 |  |
| 18-35 ans – +65 ans |  | 0.0473 |  | 0.822 |  | 0.0575 |  | 0.954 |  | 1.048 |  |
| 36-65 ans – +65 ans |  | -0.5064 |  | 0.828 |  | -0.6115 |  | 0.541 |  | 0.603 |  |
| BusinessType: |  |  |  |  |  |  |  |  |  |  |  |
| Artisan-Ouvrier – Agriculture |  | 0.1077 |  | 0.330 |  | 0.3266 |  | 0.744 |  | 1.114 |  |
| Autre – Agriculture |  | -1.0074 |  | 0.359 |  | -2.8072 |  | 0.005 |  | 0.365 |  |
| Aviculture – Agriculture |  | -0.3561 |  | 0.323 |  | -1.1032 |  | 0.270 |  | 0.700 |  |
| Bâtiment/construction – Agriculture |  | -0.7142 |  | 0.454 |  | -1.5725 |  | 0.116 |  | 0.490 |  |
| Coiffure - Salon de beauté – Agriculture |  | 0.6617 |  | 0.480 |  | 1.3780 |  | 0.168 |  | 1.938 |  |
| Commerce – Agriculture |  | 0.5288 |  | 0.235 |  | 2.2511 |  | 0.024 |  | 1.697 |  |
| Couture – Agriculture |  | 0.8676 |  | 0.357 |  | 2.4311 |  | 0.015 |  | 2.381 |  |
| Elevage – Agriculture |  | 0.1888 |  | 0.258 |  | 0.7311 |  | 0.465 |  | 1.208 |  |
| Mécanique – Agriculture |  | 12.3955 |  | 882.743 |  | 0.0140 |  | 0.989 |  | 241715.733 |  |
| Restauration – Agriculture |  | -0.0829 |  | 0.446 |  | -0.1860 |  | 0.852 |  | 0.920 |  |
| Transport (Moto - Auto) – Agriculture |  | 0.8867 |  | 0.295 |  | 3.0085 |  | 0.003 |  | 2.427 |  |
| Project: |  |  |  |  |  |  |  |  |  |  |  |
| SMP – JI |  | 0.1577 |  | 0.151 |  | 1.0411 |  | 0.298 |  | 1.171 |  |
| InterviewType: |  |  |  |  |  |  |  |  |  |  |  |
| Par téléphone – Bureau OIM |  | -1.3541 |  | 0.523 |  | -2.5913 |  | 0.010 |  | 0.258 |  |
| Sur le terrain (lieu de travail ou d’habitation du migrant…) – Bureau OIM |  | -1.3587 |  | 0.534 |  | -2.5422 |  | 0.011 |  | 0.257 |  |
| Note. Estimates represent the log odds of "BusinessSucess = High" vs. "BusinessSucess = Low" | | | | | | | | | | | |
|  | | | | | | | | | | | |

**Prediction**

| Classification Table – BusinessSucess | | | | | | | |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  | | **Predicted** | | | |  | |
| **Observed** | | **Low** | | **High** | | **% Correct** | |
| Low |  | 535 |  | 340 |  | 61.1 |  |
| High |  | 310 |  | 819 |  | 72.5 |  |
| Note. The cut-off value is set to 0.5 | | | | | | | |
|  | | | | | | | |

| Predictive Measures | | | | | |
| --- | --- | --- | --- | --- | --- |
|  |  |  |  |  |  |
| **Accuracy** | | **Specificity** | | **Sensitivity** | |
| 0.676 |  | 0.611 |  | 0.725 |  |
| Note. The cut-off value is set to 0.5 | | | | | |
|  | | | | | |

# Appendix 3: Frequencies of Mission (RSS)

| Frequencies of Mission | | | | | | | |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  |  |  |  |  |  |  |  |
| **Levels** | | **Counts** | | **% of Total** | | **Cumulative %** | |
| Burkina Faso |  | 66 |  | 4.8 % |  | 4.8 % |  |
| Cameroun |  | 103 |  | 7.6 % |  | 12.4 % |  |
| Cote d'ivoire |  | 103 |  | 7.6 % |  | 20.0 % |  |
| Gambie |  | 100 |  | 7.3 % |  | 27.3 % |  |
| Ghana |  | 103 |  | 7.6 % |  | 34.9 % |  |
| Guinee Conakry |  | 149 |  | 10.9 % |  | 45.8 % |  |
| Mali |  | 148 |  | 10.9 % |  | 56.7 % |  |
| Niger |  | 196 |  | 14.4 % |  | 71.1 % |  |
| Nigeria |  | 107 |  | 7.9 % |  | 79.0 % |  |
| Senegal |  | 100 |  | 7.3 % |  | 86.3 % |  |
| Sierra Leone |  | 44 |  | 3.2 % |  | 89.6 % |  |
| Tchad |  | 142 |  | 10.4 % |  | 100.0 % |  |
|  | | | | | | | |

# Appendix 4: Distribution of RSS Scores

**Descriptives**

| Descriptives | | | | | |
| --- | --- | --- | --- | --- | --- |
|  |  |  |  |  |  |
|  | | **Composite reintegration score** | | **Economic reintegration Score** | |
| N |  | 1360 |  | 1361 |  |
| Missing |  | 1 |  | 0 |  |
| Mean |  | 0.669 |  | 0.621 |  |
| Median |  | 0.673 |  | 0.635 |  |
| Standard deviation |  | 0.113 |  | 0.167 |  |
| Minimum |  | 0.217 |  | 0.117 |  |
| Maximum |  | 0.942 |  | 0.970 |  |
|  | | | | | |

A picture containing screenshot, design

Description automatically generated

A picture containing screenshot, black

Description automatically generated

A picture containing screenshot, design

Description automatically generated

A picture containing screenshot, rectangle, square, design

Description automatically generated

A screenshot of a computer

Description automatically generated with low confidence

A picture containing screenshot, black

Description automatically generated