Article

Analysis of Access to Emergency Funds in Sub-Saharan Countries– A Human Rights-Based Approach

Rose Porta[1,‡,*], Alejandra Munoz Garcia[1,‡], Margaret Bassney[1,‡], Aushanae Haller[1,‡]

1 Department of Statistical and Data Sciences, Smith College, Northampton MA;
* Correspondence: rporta@smith.edu
† These authors contributed equally to this work.

Version October 20, 2022 submitted to Water

Simple Summary: A Simple summary goes here.

Abstract: Having access to emergency funds is a valuable resource that many people end up needing at least once in their lives. Those who have access to emergency funding and other financial services have the capacity to remain afloat when unexpected predicaments arise, while those who are without this privilege have no choice but to endure crises and simply hope for the best. The purpose of our project is to analyze the access adults have to emergency funds and financial services in Sub-Saharan countries using a 2017 dataset from the Global Findex Database. Additionally, an important goal of our project is to employ a variety of different approaches in an attempt to minimize bias and maximize fairness, particularly when examining the performance for males and females. We also aim to determine how adults in the Sub-Saharan African region access financial services as well as establish the amount of bias we have within our models using exploratory data analysis, a baseline model, and a variety of fairness metrics. We hope to implement our findings in a Jupyter notebook where this information can be made accessible to a broader undergraduate audience.

Keywords: keyword 1; keyword 2; keyword 3 (list three to ten pertinent keywords specific to the article, yet reasonably common within the subject discipline.).

```python
# !pip install seaborn
# !pip install numpy
# !pip install pandas
# !pip install matplotlib
```

```python
# test python chunk

# Imports
import pandas as pd
import numpy as np
import seaborn as sns
# from plotnine import *
import matplotlib.pyplot as plt
# from sklearn.model_selection import train_test_split
# from sklearn.tree import DecisionTreeRegressor
# import statsmodels.api as sm
```

```
# from sklearn.compose import make_column_transformer
# from sklearn.preprocessing import OneHotEncoder
# from sklearn.linear_model import LogisticRegression, LogisticRegressionCV
```

1. Version

This Rmd-skeleton uses the mdpi Latex template published 2019/02. However, the official template gets more frequently updated than the 'rticles' package. Therefore, please make sure prior to paper submission, that you're using the most recent .cls, .tex and .bst files (available here).

2. Introduction

Science is often viewed as a way to offer trustworthy research backed solutions and answers. A lot of that research involves statistical methods performed on data however, what happens when the data and statistical methods are not as objective and trustworthy as is so often assumed? The conclusions drawn from the data are biased and unfair, most often towards minorities and protected classes of people. To contribute to a human rights based approach to data analysis, we evaluate fairness metrics on a machine learning algorithm to measure bias. We use a Global Findex data set which contains financial information about 35 Sub Saharan countries. Specifically, we create models to predict access to emergency funds, then analyze the fairness of those models. We focus on group and individual fairness metrics for the protected attribute sex. In addition we investigate the data set itself to understand where potential biases might have been implanted.

Data sets and algorithms have real world impacts on real people. The inherent bias in data sets can carry over into machine learning algorithms that are used to profile and categorize people [1,2]. Since data set's are not collected in a vacuum and often represent the discriminatory environments in which they are collected [3], we must find ways to make data sets and statistical methods more equitable. In this study we explore fairness methods that can be used to evaluate machine learning models. The "impossibility theorem" is the idea that not all fairness metrics can be satisfied at the same time[4]. Although fairness is complex and there are multiple approaches to make a model fair [5,6], it's important to continue to question how data and algorithms can be biased and how to mitigate that bias.

While there have been previous studies implementing fairness techniques in different contexts[7,8], we implement them in an exploratory context meant to teach how and when to use these techniques thus giving us more freedom to branch beyond a specific question while supporting previous work about the importance of these fairness metrics[3,9]. We analyse the data, data collection methods, prediction models, and the fairness metrics to assess how biased our data is and understand how we can de-bias when possible.

3. Data

Our data is derived from The World Bank in The Global Findex Database, comprising the most comprehensive data sets on how adults save, borrow, make payments, and manage risk in more than 140 economies around the world. The data set was created to record various measures of financial equity and inclusion, with the intention that such information could reveal opportunities to expand access to financial services and to promote greater use of digital financial services for individuals who do not have a bank account. Conducted by Gallup, Inc for the annual Gallup World Poll, the participants responded to the questionnaire either on the phone or in-person. There were several variables of interest in this dataset when creating models to predict access to emergency funds, including demographic and financial information.

```
# Read in Data
file = '~/Desktop/Fall-2022/SDS410/women-at-table/findex_SubSahAfrica.csv'
```

```
df = pd.read_csv(file, index_col=0)
print(f'There are {df.shape[0]} entries and {df.shape[1]} features')
```

```
## There are 35000 entries and 105 features
```

```
df.head()
```

```
##        economy economycode  ... pay_cash  pay_cash_mobintbuy
## 12138   Benin          BEN  ...      0.0                 NaN
## 12139   Benin          BEN  ...      0.0                 NaN
## 12140   Benin          BEN  ...      0.0                 0.0
## 12141   Benin          BEN  ...      0.0                 NaN
## 12142   Benin          BEN  ...      0.0                 NaN
##
## [5 rows x 105 columns]
```

```
# Set Gender Palette
red = '#FF7377'
blue = '#00B2EE'
gender_palette = [blue, red]

purple = '#BF3EFF'
green = '#1B851B'
yes_no_pal = [purple, green]

pink = '#FFC0CB'
```

```
# select vars of interest
df2 = df[['female', 'age', 'emp_in', 'account_fin', 'fin24', 'fin25', 'fin32', 'fin48', 'educ', 'ec
df2 = df2.query('fin24 < 3')
# Recode fin24 values
df2.loc[df2['fin24'] == 1, "fin24"] = 'Yes'
df2.loc[df2['fin24'] == 2, "fin24"] = 'No'
df2.loc[df2['fin24'] == 3, "fin24"] = 'Don\'t Know'
df2.loc[df2['fin24'] == 4, "fin24"] = 'Refuse'
# rename fin24 to has_access
df2.rename(columns = {'fin24': 'has_access'}, inplace = True)
# recode gender
df2.loc[df2['female'] == 1, "female"] = 'male'
df2.loc[df2['female'] == 2, "female"] = 'female'
df2.rename(columns = {'female': 'gender'}, inplace = True)
# recode account_fin
df2.loc[df2['account_fin'] == 0, "account_fin"] = 'No'
df2.loc[df2['account_fin'] == 1, "account_fin"] = 'Yes'
# recode emp_in
df2.loc[df2['emp_in'] == 0, "emp_in"] = 'No'
df2.loc[df2['emp_in'] == 1, "emp_in"] = 'Yes'
# Recode fin32 values: Recieved Wage Payments
df2.loc[df2['fin32'] == 1, "fin32"] = 'Yes'
df2.loc[df2['fin32'] == 2, "fin32"] = 'No'
```

```
df2.loc[df2['fin32'] == 3, "fin32"] = 'Don\'t Know'
df2.loc[df2['fin32'] == 4, "fin32"] = 'Refuse'
# recode fin48 values: National ID
df2.loc[df2['fin48'] == 1, "fin48"] = 'Yes'
df2.loc[df2['fin48'] == 2, "fin48"] = 'No'
df2.loc[df2['fin48'] == 3, "fin48"] = 'Don\'t Know'
df2.loc[df2['fin48'] == 4, "fin48"] = 'Refuse'
#recode educ values: Highest Level of Education
df2.loc[df2['educ'] == 1, "educ"] = 'Primary'
df2.loc[df2['educ'] == 2, "educ"] = 'Secondary'
df2.loc[df2['educ'] == 3, "educ"] = 'Tertiary'
df2.loc[df2['educ'] == 4, "educ"] = 'Don\'t Know'
df2.loc[df2['educ'] == 5, "educ"] = 'Refuse'
#recode fin25 values: Main Source of Emergency Funds
df2.loc[df2['fin25'] == 1, "fin25"] = 'Savings'
df2.loc[df2['fin25'] == 2, "fin25"] = 'Family, relatives, or friends'
df2.loc[df2['fin25'] == 3, "fin25"] = 'Money from working'
df2.loc[df2['fin25'] == 4, "fin25"] = 'Borrowing from a bank/employer/private lender'
df2.loc[df2['fin25'] == 5, "fin25"] = 'Selling assets'
df2.loc[df2['fin25'] == 6, "fin25"] = '(Some other source)'
df2.loc[df2['fin25'] == 7, "fin25"] = 'Don\'t Know'
df2.loc[df2['fin25'] == 8, "fin25"] = 'Refuse'
#recode fin32 values: Recieved Wage Payments
df2.loc[df2['fin32'] == 1, "fin32"] = 'Yes'
df2.loc[df2['fin32'] == 2, "fin32"] = 'No'
df2.loc[df2['fin32'] == 3, "fin32"] = 'Don\'t Know'
df2.loc[df2['fin32'] == 4, "fin32"] = 'Refuse'

df2.rename(columns = {'fin32': 'Receive Wage Payments', 'fin48': 'National ID', 'educ': 'Education'
df2.head()
```

```
##        gender   age emp_in  ... National ID Education economy
## 12138    male  60.0     No  ...          No   Primary   Benin
## 12139    male  45.0    Yes  ...         Yes   Primary   Benin
## 12140    male  27.0    Yes  ...         Yes  Tertiary   Benin
## 12141  female  24.0    Yes  ...          No   Primary   Benin
## 12142    male  22.0    Yes  ...          No   Primary   Benin
##
## [5 rows x 10 columns]
```

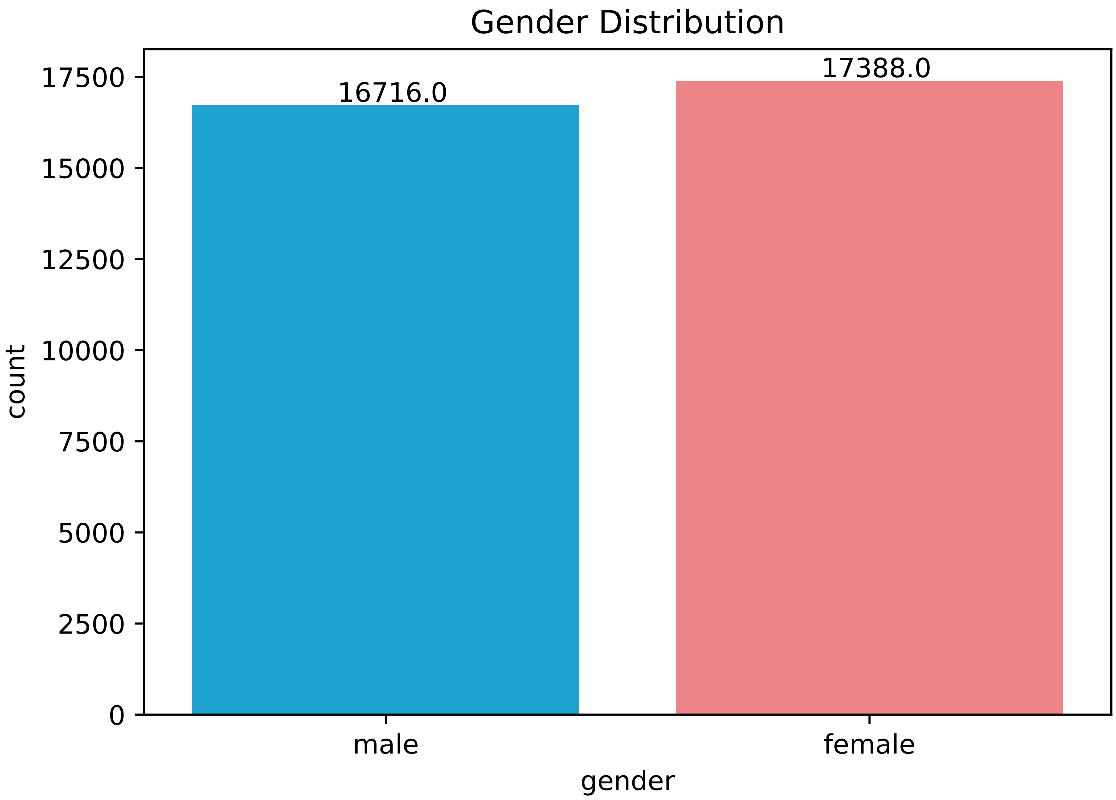
3.1. Demographics

3.1.1. *gender*

The variable *gender* distinguishes gender. There are 16,716 males in this dataset and 17,388 females. This is a fairly equal distribution that we can see in the graph below.

```
# barplot female versus not
g = sns.countplot(x = 'gender', data = df2, palette = gender_palette)
for p in g.patches:
```

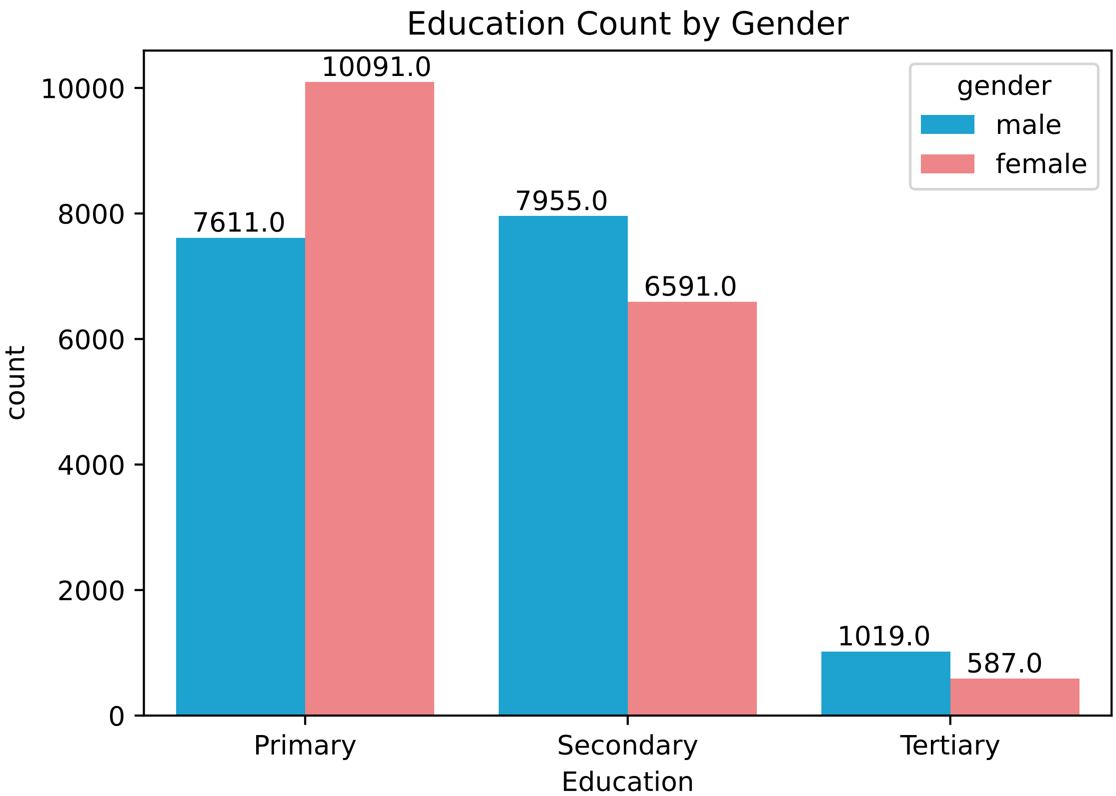
```
    g.annotate('{:.1f}'.format(p.get_height()), (p.get_x()+ .3, p.get_height()+100))
g.set(title = "Gender Distribution")
```



3.1.2. *Education*

The *Education* variable corresponds to the highest level of education attained with ‘Primary’, ‘Secondary’ and ‘Tertiary’ being the three options. Here is the distribution of education by gender:

```
ebg = sns.countplot(x="Education", hue= "gender", palette= gender_palette, order= ['Primary', 'Seco
for p in ebg.patches:
    ebg.annotate('{:.1f}'.format(p.get_height()), (p.get_x()+.05, p.get_height()+100))
ebg.set(title= "Education Count by Gender")
```

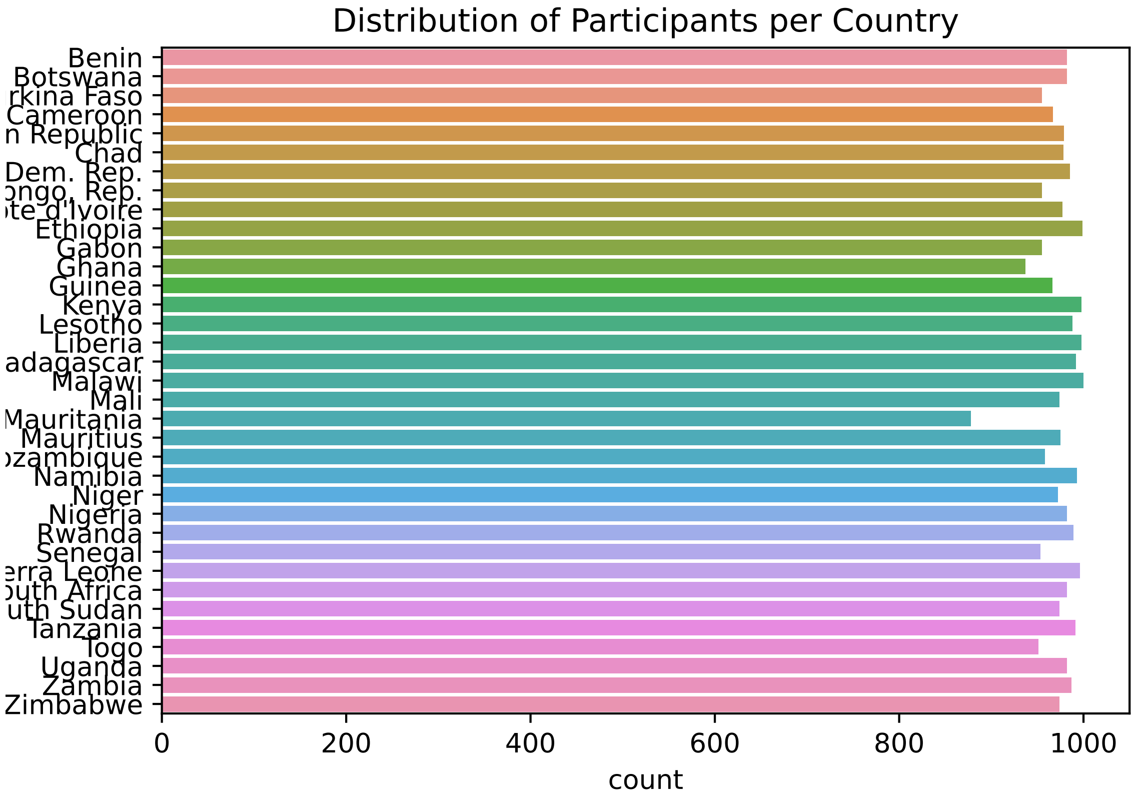


The bar plot above shows us that there are more women with primary education, but more men with secondary or tertiary education. Overall, we can see that there are more men with higher education than women. About 1,000 more men have received a secondary education and there is about double the amount of men with tertiary education compared to women showing a clear disparity.

3.1.3. *economy*

The final demographic variable of interest is the *economy* variable that separates respondents by which country they live in. There are 35 different countries with exactly 1000 respondents from each.

```
# Number of Observations per Country
plt.rcParams["figure.figsize"] = (10,7)
sns.countplot(y = 'economy', data = df2).set(title = "Distribution of Participants per Country")
```



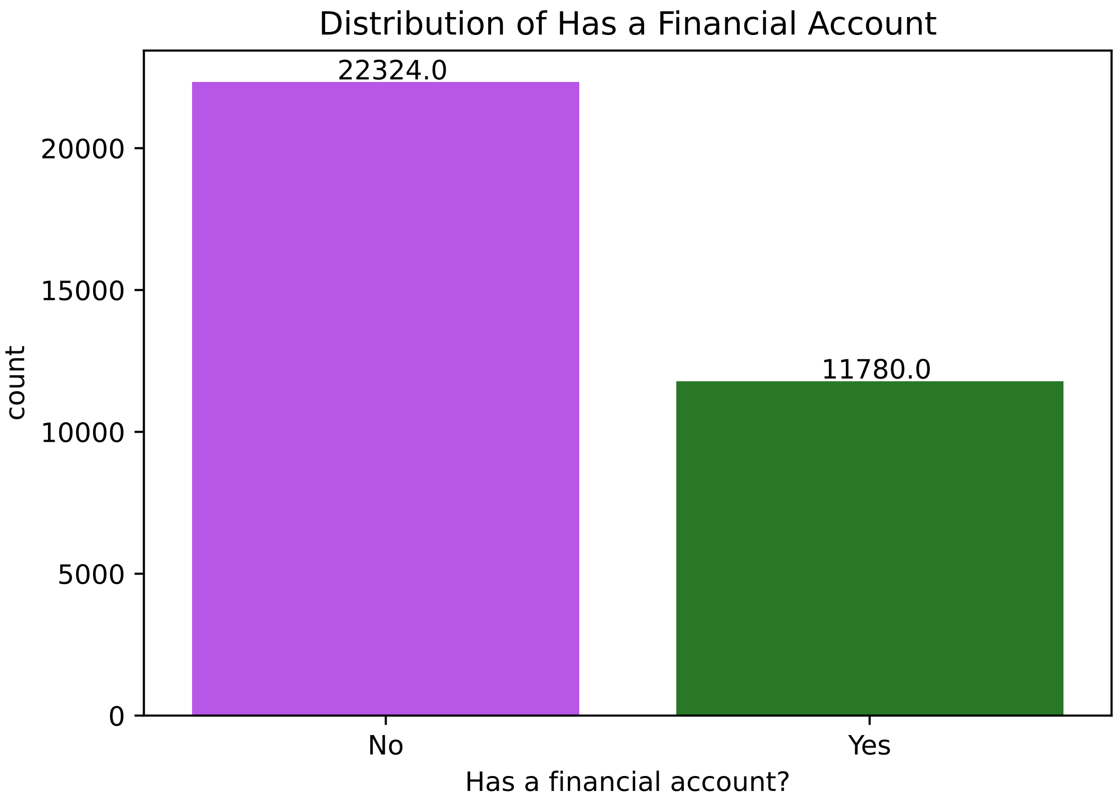
3.2. *Financial*

From the financial related variables, we were most interested in a few specific financial variables that we thought would have an impact on access to emergency funds.

3.2.1. *account_fin*

The first variable being *account_fin* which distinguishes those who have a financial account from those who don't:

```
# barplot of number of people who have a bank account
g = sns.countplot(x = 'account_fin', data = df2, palette = yes_no_pal)
for p in g.patches:
   g.annotate('{:.1f}'.format(p.get_height()), (p.get_x()+ .3, p.get_height()+100))
g.set(title = "Distribution of Has a Financial Account", xlabel = 'Has a financial account?')
```



We can see that about two thirds of individuals do not have an account. This is likely connected to the lack of access to emergency funds displayed above given that if an individual does not have a financial account, we would expect they are less likely to have a source of emergency funds, as emergency funds are generally stored in an account.

3.2.2. *reason*

Those who do not have a financial account were asked why in the *reason* variable, that provides a list of possible reasons for not having a financial account:

```
# reasons for no financial account
# pivot data to long format
df_long = df[['fin11a', "fin11b", "fin11c", "fin11d", "fin11e", "fin11f", "fin11g", "fin11h"]]\
.stack()\
.reset_index()
df_long.rename(columns = {'level_1':'reason', 0:'value'}, inplace = True)
df3 = df_long.query('value == 1.0') # filter to only "yes" answers for each reason
# recode reason values
df3.loc[df3['reason'] == 'fin11a', "reason"] = 'distance'
df3.loc[df3['reason'] == 'fin11b', "reason"] = 'expense'
df3.loc[df3['reason'] == 'fin11c', "reason"] = 'no_documentation'
df3.loc[df3['reason'] == 'fin11d', "reason"] = 'lack_trust'
df3.loc[df3['reason'] == 'fin11e', "reason"] = 'religious'
df3.loc[df3['reason'] == 'fin11f', "reason"] = 'lack_money'
df3.loc[df3['reason'] == 'fin11g', "reason"] = 'family_member_has'
df3.loc[df3['reason'] == 'fin11h', "reason"] = 'no_need'
df3.head()
```

```python
# aggregate counts by reason
```

```
##     level_0            reason  value
## 0     12138          distance    1.0
## 2     12138  no_documentation    1.0
## 11    12140        lack_trust    1.0
## 15    12140           no_need    1.0
## 18    12141  no_documentation    1.0
```

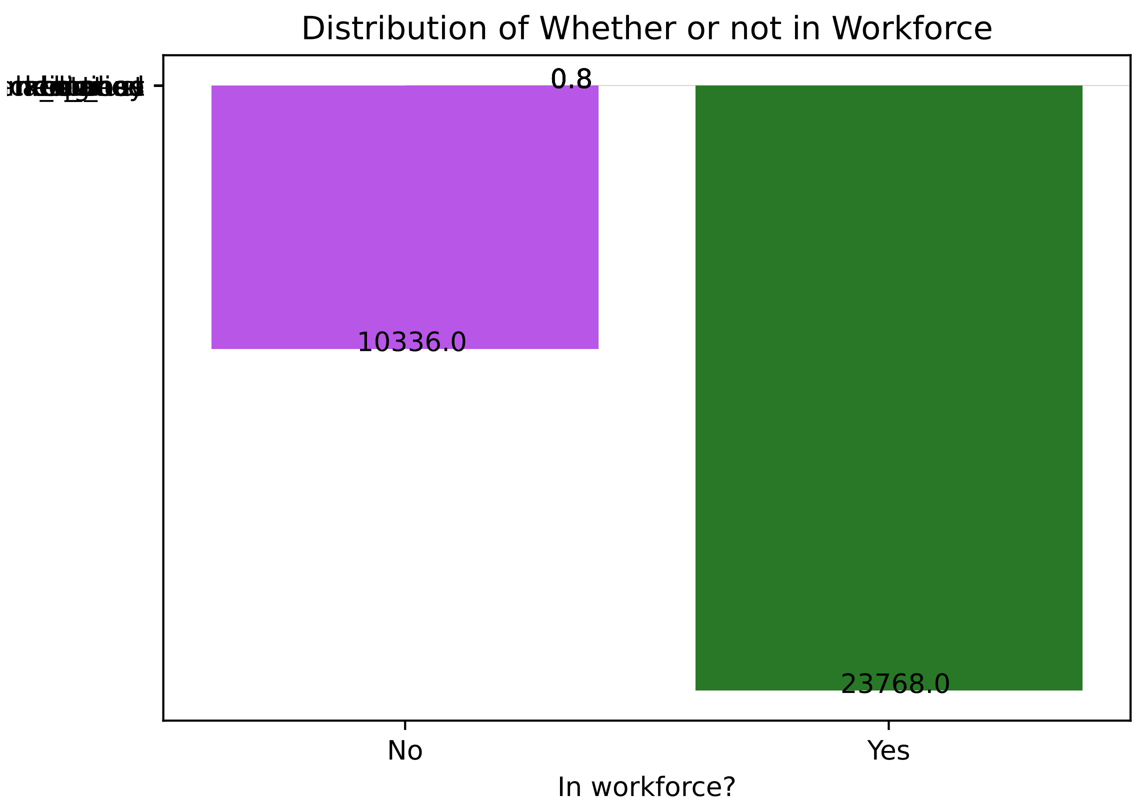
```python
reason_counts = df3.groupby('reason').agg('count').reset_index()
```

```python
# barplot of count by reason-- seaborn
reasons_plot = sns.barplot(y = 'reason', x = 'value', data = reason_counts, palette = [pink, pink, 
```

3.2.3. *emp_in*

Employment status was another financial variable of interest represented by *emp_in*, which asks whether or not the participant is in the workforce. It appears that about three-fourths of individuals are in the workforce:

```python
# Distribution by in workforce
g = sns.countplot(x = 'emp_in', data = df2, palette = yes_no_pal)
for p in g.patches:
   g.annotate('{:.1f}'.format(p.get_height()), (p.get_x()+ .3, p.get_height()+100))
g.set(title = "Distribution of Whether or not in Workforce", xlabel = "In workforce?")
```



3.2.4. *inc_q*

And lastly, we evaluated *inc_q*, which represents income quantile. Income quantile is separated into 5 quantiles with 1 being the poorest and 5 being the richest. The mean for all of the countries in the dataset is 3.241. This means that all the countries average out to be about middle class.

```
df_inq = df['inc_q']
df["inc_q"].mean()
# fig = plt.figure()
# ax1 = fig.add_subplot(2,2,1)
# sns.countplot(df_inq, ax = ax1)
# ax1.set_xlabel("Income Quantile")
# ax1.set_xticklabels(['Poorest 20%','Second 20%','Middle 20%','Fourth 20%','Richest 20%'])
# ax1.title.set_text('Income Quantile  Distribution')
```

```
## 3.241085714285714
```

The majority of the data set has individuals within the richest quantile, Quantile 5.

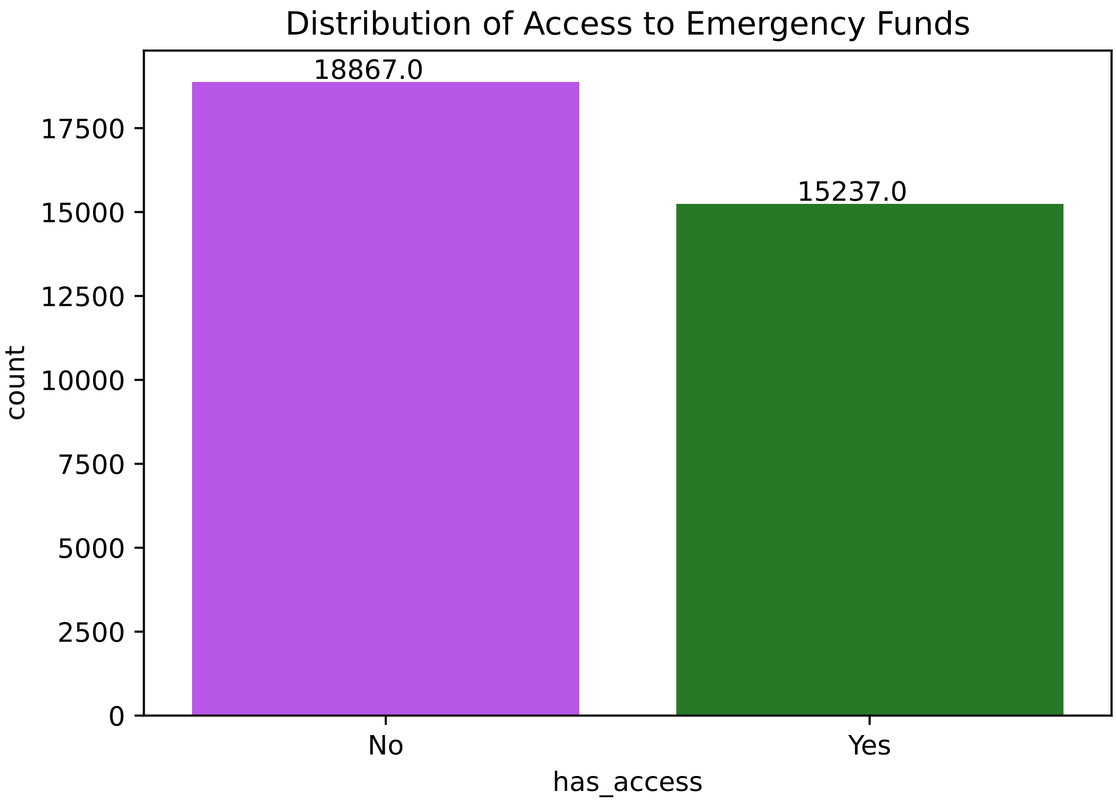
3.3. Emergency Funds

To explore access to emergency funds in our dataset, we were interested 3 variables we thought could be related:

3.3.1. *has_access*

The variable *has_access* directly asks participants if they have access to emergency funds:

```
# barplot of access to emergency funds
g = sns.countplot(x = 'has_access', data = df2, palette = yes_no_pal)
for p in g.patches:
   g.annotate('{:.1f}'.format(p.get_height()), (p.get_x()+ .25, p.get_height()+100))
g.set(title = "Distribution of Access to Emergency Funds")
```



The barchart above displays the overall distribution of access to emergency funds. We can see that over half of individuals represented in the data do not have access.

3.3.2. *main_source_funds*

We proceeded to explore the source of emergency funds using the *main_source_funds* variable, which provides a list of options for where participants receive their main source of emergency funds:

```
# barplot of main source of emergency funds
chart = sns.countplot(y = 'main_source_funds',
                      data = df2,
                      order = ["Money from working","Family, relatives, or friends",
                               "Savings", "Selling assets",
                               "Borrowing from a bank/employer/private lender", "(Some other source
                      title = "Distribution of the Main Source of Emergency Funds")
```

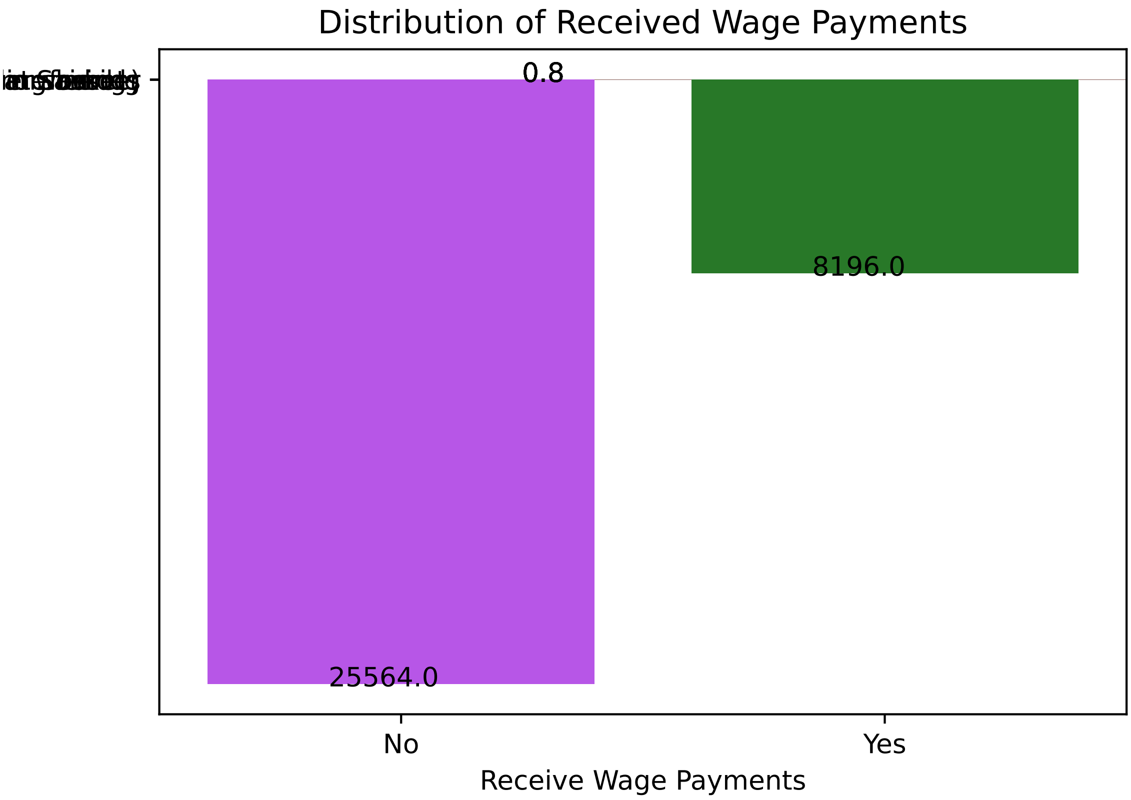
The barchart above displays the overall distribution of the main source of emergency funds. Most of the individuals with access to emergency funds receive their funding from work, their family and friends, or their savings.

3.3.3. *Recieve Wage Payments*

Diving further into the "Money from Working" category, we can see that only 8196 individuals receive wage payments from the *Receive Wage Payments* variable. This analysis suggests that receiving wage payments may be a key factor in determining access to emergency funds.

```
# barplot of recieved wage payments
chart = sns.countplot(x = 'Receive Wage Payments',data = df2, order = ["No", "Yes"], palette = yes_
```

```
for p in chart.patches:
   chart.annotate('{:.1f}'.format(p.get_height()), (p.get_x()+ .25, p.get_height()+100))
chart.set(title = "Distribution of Received Wage Payments")
```



3.3.4. *gender*, *economy* and *Education* in relation to Emergency Funds

Finally, we sought to find if there were disparities in access to emergency funds by *gender*, *economy*, and *Education*.

```
df_gender = df2[['gender', 'economy']].query("gender == 'female'").groupby('economy').count()
df_gender = df_gender.assign(male = lambda df_gender: 1000 - df_gender['gender']).rename(columns = 

#Percent of females
df_gender = df_gender.assign(Female = lambda df_gender: (df_gender['female'])/10)

#Percent of males
df_gender = df_gender.assign(Male = lambda df_gender: 100 - df_gender['Female'])

df_gender = df_gender[['Male', 'Female']]
df_gender['economy'] = df2.economy.unique()
df_gender = df_gender.sort_values('Male')

# Counting by Access to Emergency Funds by gender
f_access = df2[['has_access', 'economy', 'gender']].query("has_access == 'Yes' & gender == 'female'
m_access = df2[['has_access', 'economy', 'gender']].query("has_access == 'Yes'& gender != 'female'")
```

```
# Defining acccess vs no access by country columms
f_count = f_access['has_access']
m_count = m_access['has_access']

# Merging columns
df3 = pd.merge(f_count, m_count, how='inner', on = 'economy')
df3 = df3.assign(total = lambda df3: df3['has_access_x']+df3['has_access_y'])

#Percent of females with access
df3 = df3.assign(f_percent_access = lambda df3: (df3['has_access_x']/df3["total"])*100)

#Percent of males with access
df3 = df3.assign(m_percent_access = 100 - df3['f_percent_access'])

df3['economy'] = df2.economy.unique()
df3 = df3[['m_percent_access', 'f_percent_access',   'economy']]
df3 = df3.sort_values('m_percent_access')
df3 = df3.rename(columns = {'m_percent_access':'Males With Acesss', 'f_percent_access': 'Females Wi
```

```
# create new figure
fig = plt.figure()
#add sub plot
```

```
## <string>:1: MatplotlibDeprecationWarning: The resize_event function was deprecated in Matplotlib
```

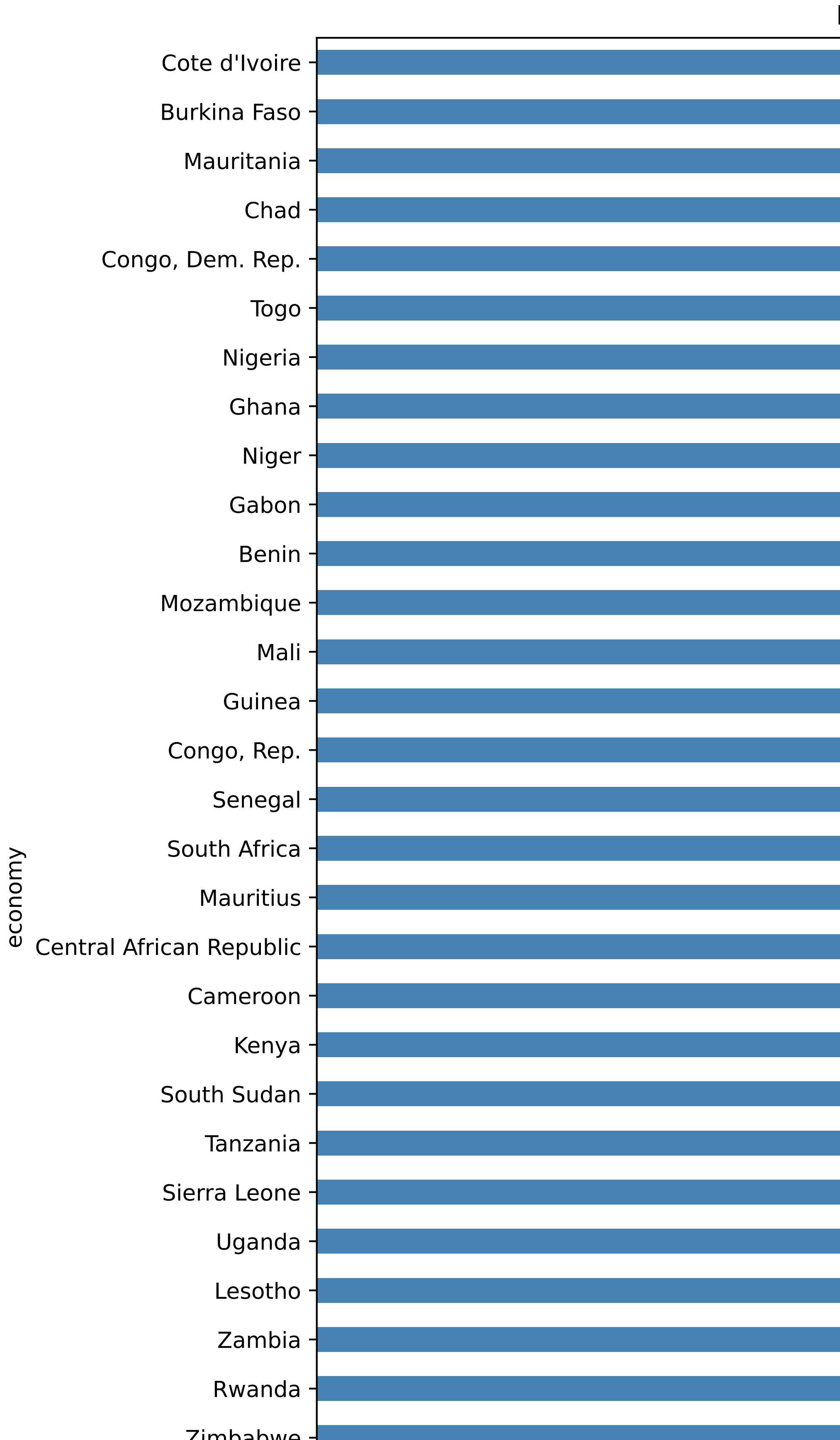
```
ax1 = fig.add_subplot(1,2,1)
# fig 1
df_gender.set_index('economy').plot(kind='barh', stacked=True, color=['steelblue', 'pink'], title =
#add sub plot
ax2 = fig.add_subplot(1,2,2)
# fig 2
fig2 = df3.set_index('economy').plot(kind='barh', stacked=True, color=['steelblue', 'pink'], title
# set fig size
fig.set_size_inches(29, 14)
```

In the side-by-side barplots above, we can see that although only about 50% of the countries have a higher percentage of men represented in the questionnaire (left bar plot), in 75% of the countries more men have access to emergency funds than women (right bar plot).

```
# Counting by Access to Emergency Funds by country
access = df2[['has_access', 'economy']].query("has_access == 'Yes'").groupby('economy').count()
no_access = df2[['has_access', 'economy']].query("has_access == 'No'").groupby('economy').count()

# Defining acccess vs no access by country columms
a_count = access['has_access']
no_a_count = no_access['has_access']

# Merging columns
df3 = pd.merge(a_count, no_a_count, how='inner', on = 'economy')#.assign(total = lambda df3: df3['h
#df3 = df3.assign(percent_access = lambda df3: (df3['has_access_x']/df3["total"])*100)
```

```
df3['economy'] = df2.economy.unique()
df3 = df3.sort_values('has_access_x')


educ_perc = (df2.groupby(['Education'])['has_access']
                     .value_counts(normalize=True)
                     .rename('percentage')
                     .mul(100)
                     .reset_index()
                     )
```

```
educ_access = sns.barplot(x = "Education", y = "percentage", order= ['Primary', 'Secondary', 'Terti
for p in educ_access.patches:
   educ_access.annotate('{:.1f}'.format(p.get_height()), (p.get_x()+.12, p.get_height()+.5))
educ_access.set(title = "Access to Emergency Funds by Education Level", xlabel = 'Education Level')
```


economy
Cote d'Ivoire
Burkina Faso
Mauritania
Chad
Congo, Dem. Rep.
Togo
Nigeria
Ghana
Niger
Gabon
Benin
Mozambique
Mali
Guinea
Congo, Rep.
Senegal
South Africa
Mauritius
Central African Republic
Cameroon
Kenya
South Sudan
Tanzania
Sierra Leone
Uganda
Lesotho
Zambia
Rwanda

Additionally, in the barplot above we can see the distribution of funds based on an individual's highest education level. 63% of people with only a primary education do not have access to emergency funds compared to 37% of people who do. These numbers are more evenly distributed for those with secondary education, with about 49% of people not having access to emergency funds, while 51% of people do have access. Finally, for those with a tertiary level of education we can see that about 72% of people have access to emergency funds while only 28% of that group does not have access. Overall, we can make the assumption that people with a higher level of education are more likely to have access to emergency funds.

4. Materials and Methods

Materials and Methods [10] should be described with sufficient details to allow others to replicate and build on published results. Please note that publication of your manuscript implicates that you must make all materials, data, computer code, and protocols associated with the publication available to readers. Please disclose at the submission stage any restrictions on the availability of materials or information. New methods and protocols should be described in detail while well-established methods can be briefly described and appropriately cited.[5]

Research manuscripts reporting large datasets that are deposited in a publicly available database should specify where the data have been deposited and provide the relevant accession numbers. If the accession numbers have not yet been obtained at the time of submission, please state that they will be provided during review. They must be provided prior to publication.[11]

Interventionary studies involving animals or humans, and other studies require ethical approval must list the authority that provided approval and the corresponding ethical approval code.

5. Results

This section may be divided by subheadings. It should provide a concise and precise description of the experimental results, their interpretation as well as the experimental conclusions that can be drawn.[3]

5.1. Subsection Heading Here

Subsection text here.

5.1.1. Subsubsection Heading Here

Bulleted lists look like this:

- First bullet
- Second bullet
- Third bullet

Numbered lists can be added as follows:

1. First item
2. Second item
3. Third item

The text continues here.

All figures and tables should be cited in the main text as Figure 1, Table 1, etc.

Figure 1. This is a figure, Schemes follow the same formatting. If there are multiple panels, they should be listed as: (**a**) Description of what is contained in the first panel. (**b**) Description of what is contained in the second panel. Figures should be placed in the main text near to the first time they are cited. A caption on a single line should be centered.

Table 1. This is a table caption. Tables should be placed in the main text near to the first time they are cited.

Title 1	**Title 2**	**Title 3**
entry 1	data	data
entry 2	data	data

This is an example of an equation:

$$\mathbb{S} \tag{1}$$

Example of a theorem:

Theorem 1. *Example text of a theorem.*

The text continues here. Proofs must be formatted as follows:

Example of a proof:

Proof of Theorem 1. Text of the proof. Note that the phrase 'of Theorem 1' is optional if it is clear which theorem is being referred to. □

The text continues here.

6. Discussion

Authors should discuss the results and how they can be interpreted in perspective of previous studies and of the working hypotheses. The findings and their implications should be discussed in the broadest context possible. Future research directions may also be highlighted.

7. Conclusion

This section is not mandatory, but can be added to the manuscript if the discussion is unusually long or complex.

8. Patents

This section is not mandatory, but may be added if there are patents resulting from the work reported in this manuscript.

Acknowledgments: All sources of funding of the study should be disclosed. Please clearly indicate grants that you have received in support of your research work. Clearly state if you received funds for covering the costs to publish in open access.

Author Contributions: For research articles with several authors, a short paragraph specifying their individual contributions must be provided. The following statements should be used "X.X. and Y.Y. conceive and designed the experiments; X.X. performed the experiments; X.X. and Y.Y. analyzed the data; W.W. contributed reagents/materials/analysis tools; Y.Y. wrote the paper.'' Authorship must be limited to those who have contributed substantially to the work reported.

Conflicts of Interest: Declare conflicts of interest or state 'The authors declare no conflict of interest.' Authors must identify and declare any personal circumstances or interest that may be perceived as inappropriately influencing the representation or interpretation of reported research results. Any role of the funding sponsors in the design of the study; in the collection, analyses or interpretation of data in the writing of the manuscript, or in the decision to publish the results must be declared in this section. If there is no role, please state 'The founding sponsors had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, an in the decision to publish the results'.

Abbreviations

The following abbreviations are used in this manuscript:

MDPI	Multidisciplinary Digital Publishing Institute
DOAJ	Directory of open access journals
TLA	Three letter acronym
LD	linear dichroism

Appendix A

Appendix A.1

The appendix is an optional section that can contain details and data supplemental to the main text. For example, explanations of experimental details that would disrupt the flow of the main text, but nonetheless remain crucial to understanding and reproducing the research shown; figures of replicates for experiments of which representative data is shown in the main text can be added here if brief, or as Supplementary data. Mathematical proofs of results not central to the paper can be added as an appendix.

Appendix B

All appendix sections must be cited in the main text. In the appendixes, Figures, Tables, etc. should be labeled starting with 'A', e.g., Figure A1, Figure A2, etc.

References

1. Navarro, C.L.A.; Damen, J.A.; Takada, T.; Nijman, S.W.; Dhiman, P.; Ma, J.; Collins, G.S.; Bajpai, R.; Riley, R.D.; Moons, K.G.; others. Risk of bias in studies on prediction models developed using supervised machine learning techniques: systematic review. *bmj* **2021**, *375*.
2. Hellström, T.; Dignum, V.; Bensch, S. Bias in Machine Learning–What is it Good for? *arXiv preprint arXiv:2004.00686* **2020**.
3. Barocas, S.; Hardt, M.; Narayanan, A. Fairness and machine learning.
4. Kleinberg, J.; Mullainathan, S.; Raghavan, M. Inherent trade-offs in the fair determination of risk scores. *arXiv preprint arXiv:1609.05807* **2016**.
5. Kypraiou, S. What is Fairness? **2021**. Publisher: PubPub.
6. Green, B.; Hu, L. The myth in the methodology: Towards a recontextualization of fairness in machine learning. Proceedings of the machine learning: the debates workshop, 2018.
7. Deho, O.B.; Zhan, C.; Li, J.; Liu, J.; Liu, L.; Duy Le, T. How do the existing fairness metrics and unfairness mitigation algorithms contribute to ethical learning analytics? *British Journal of Educational Technology* **2022**.
8. Kim, J.Y.; Cho, S.B. An Information Theoretic Approach to Reducing Algorithmic Bias for Machine Learning. *Neurocomputing* **2022**.
9. Anahideh, H.; Asudeh, A.; Thirumuruganathan, S. Fair active learning. *Expert Systems with Applications* **2022**, *199*, 116981.
10. Knight, C. Automated Decision-making and Judicial Review. *Judicial Review* **2020**. Publisher: Routledge.
11. Gebru, T.; Morgenstern, J.; Vecchione, B.; Vaughan, J.W.; Wallach, H.; Daumé III, H.; Crawford, K. Datasheets for Datasets, 2021. arXiv:1803.09010 [cs], doi:10.48550/arXiv.1803.09010.

Sample Availability: Samples of the compounds ...... are available from the authors.