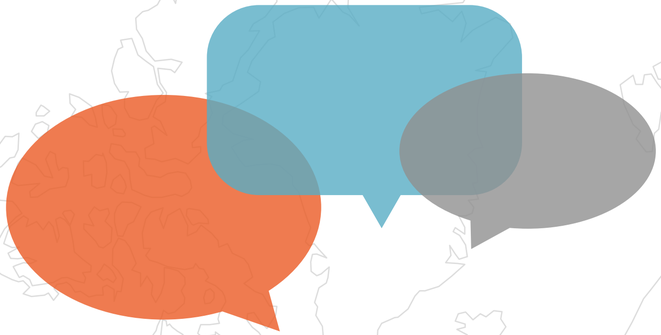
ReDI School of Digital Integration



Servus!

and

Welcome!

We believe in the **power of conversation**

DIGITAL CAREER PROGRAM

Tech Talents of Tomorrow

ReDI School of
Digital Integration

My name is Azuka Benjamin Nebonta.

The following is a presentation of my final project pursuant to ReDi Data Science fall 2020 program.

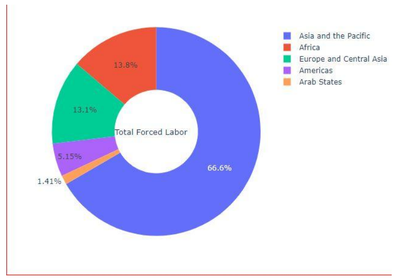
Project Title:

Concept of "Trends in Global Human Trafficking and exploitation of Persons"

29.11.2020

ReDI School of Digital Integration

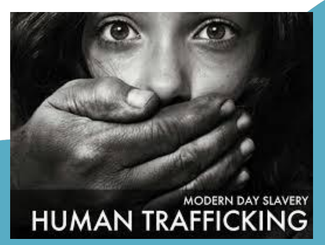
Hello, and welcome to my final project.



My name is Benjamin Nebonta, and by the time you read this, I would bewell on my way to be called a ReDi Digital school alumni.

The aim of this work is take a close look at a world social order crime - **Human Trafficking, and Victim exploitation.**

We are ReDi



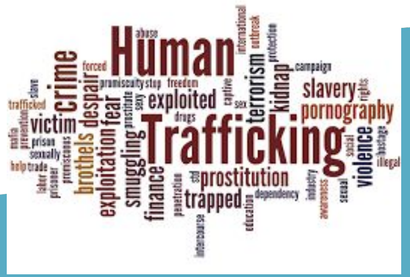
ReDI School of Digital Integration

What are my Project Goals?

What I aim to achieve with this work is to use exploratory data analysis and visualization to show the trends in the Human trafficking and exploitation eco system.

Among other variables, I want to explore the victims demographic information. To use statistical data to show among others variables, who is being trafficked, how, by who, and the relationship between the exploiter and the exploited.

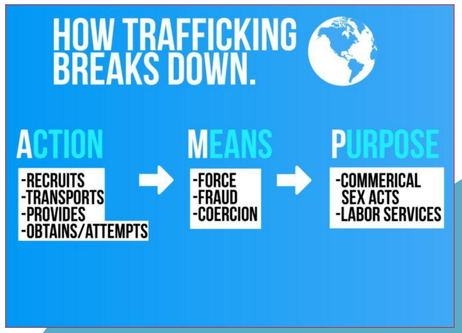
In addition to exploring the relationship between these 2 groups, i will analyse and expose the pertinent data to the following:



ReDI School of Digital Integration

Questions:

- Who is been trafficked?
- What gender is trafficked the most? (Gender disparity)
- What age group is trafficked the most?
- What are the victims Nationalities?
- Which are the destination countries? (Receiving Countries)
- How are the trafficked victims exploited?
- How are the Victims exploited (Mode of Exploitation)
- Relationship between the trafficked and Trafficker



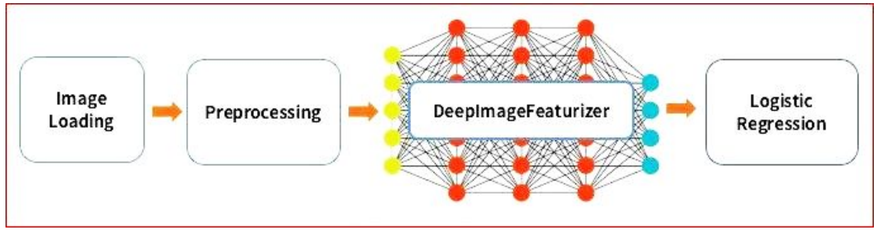
ReDI School of Digital Integration

Methodology:

I will employ Python programming language in jupyter notebook as the primary development and deployment platform. I will also work with it's dependent frameworks such as:

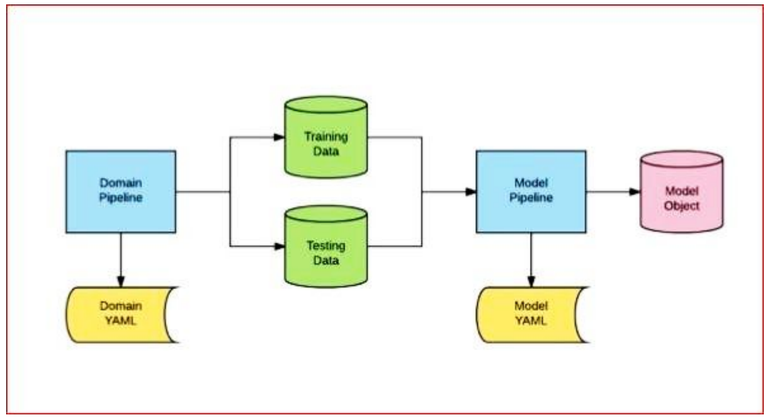
- Numpy
- Pandas
- Matplotlib
- Seaborn
- Scikit-learn

ReDI School of Digital Integration



What are my Pipelines:

- Data Collection
- Data Cleaning
- Exploratory Data Analysis
- Visualization
- Model Building



ReDI School of Digital Integration

What is Human Trafficking?

According to Wikipedia.org, human trafficking is the trade of humans for the purpose of forced labour, sexual slavery, or commercial sexual exploitation for the trafficker or others. This may also encompass providing a spouse in the context of forced marriage, or the extraction of organs or tissues, including for surrogacy and ova removal.

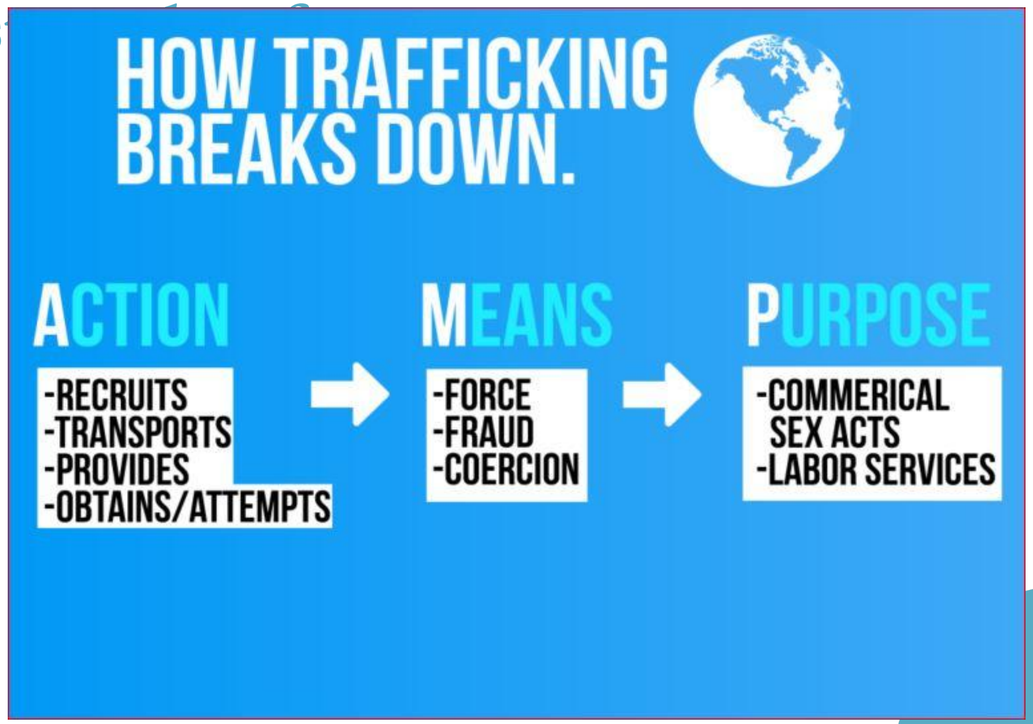
Human trafficking takes place across Continents, Countries, within a single country, or even within a single community.

ReDI School of Digital Integration

What are the 3 elements of Human Trafficking?

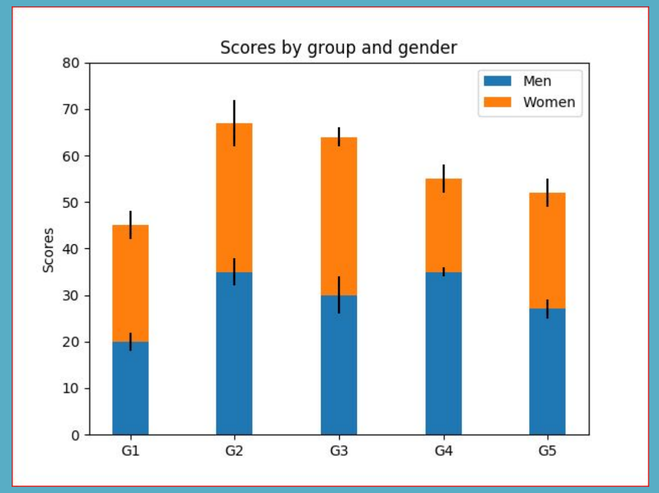
The three key elements that must be present for a situation of trafficking in persons to exis

1. Action (recruitment)
2. Means (threat)
3. purpose (exploitation)



ReDI School of Digital Integration

Now, Lets dig into the problem.



The Source data will be code collected from:
Http://www.ctdatacollaborative.org

Problem and solution statement:

The source dataset has 64 columns. One of the columns was useless for the purpose of the analysis I propose, so I will remove it.

I then proceeded to clean the data using Numpy and Pandas libraries.

The resulting dataframe have 63 columns which I will use for this work.

Loading project data

The data used for the project will be obtained in url from https://www.ctdatacollaborative.org

In [9]:
```
df = pd.read_csv('https://www.ctdatacollaborative.org/sites/default/files/The%20Global%20Dataset%2014%20Apr%202020.csv', dtyp
```

Data Cleaning

Remove an unrequired index column

In [10]:
```
df.drop(df.columns[0], axis=1, inplace=True)
```

Columns, Row Count and Type

We now have 63 different columns and some missing data in across the columns

Data Cleaning Contd...

The source dataset had number values which will diffuse the results of the analysis ('-99', '00', -99).

The solution is not to drop them as that will make the dataset unusable due to its composition. The solution was to convert the values to Numpy non values(NaN).

Lets's handle missing values

In [12]:
```
df.replace('-99', np.nan, inplace=True)
df.replace(-99, np.nan, inplace=True)
df.replace('00', np.nan, inplace=True)
```

ReDI School of Digital Integration

These are questions I asked, and the Answers obtained:

- In what year was most cases recorded

```
2016    16399
```

- What types of sexual Exploitation were forced on the victims

	0
1	Prostitution
2	Pornography
3	Private sexual services

- Which types of Victim Exploitation Practiced

	0
0	Sexual exploitation
2	Forced labour
3	Other
4	Forced labour;Other
5	Slavery and similar practices
6	Forced marriage
7	Forced labour;Sexual exploitation;Combined sex...

ReDI School of Digital Integration

These are questions I asked, and the Answers obtained: contd...

- In What forced labour occupations were victims engaged

1	Not specified
2	Other;Not specified
3	Hospitality
4	Domestic work
5	Domestic work;Not specified
6	Begging
7	Agriculture
8	Construction
9	Manufacturing
10	Construction;Not specified
11	Agriculture;Not specified
12	Other
13	Aquafarming
14	Peddling
15	Domestic work;Other

- How many Victims reported themselves as abducted

16371

- In Which countries were victims Exploited

58 Countries (too many to illustrate in the limited space)

ReDI School of Digital Integration

These are questions I asked, and the Answers obtained: contd...

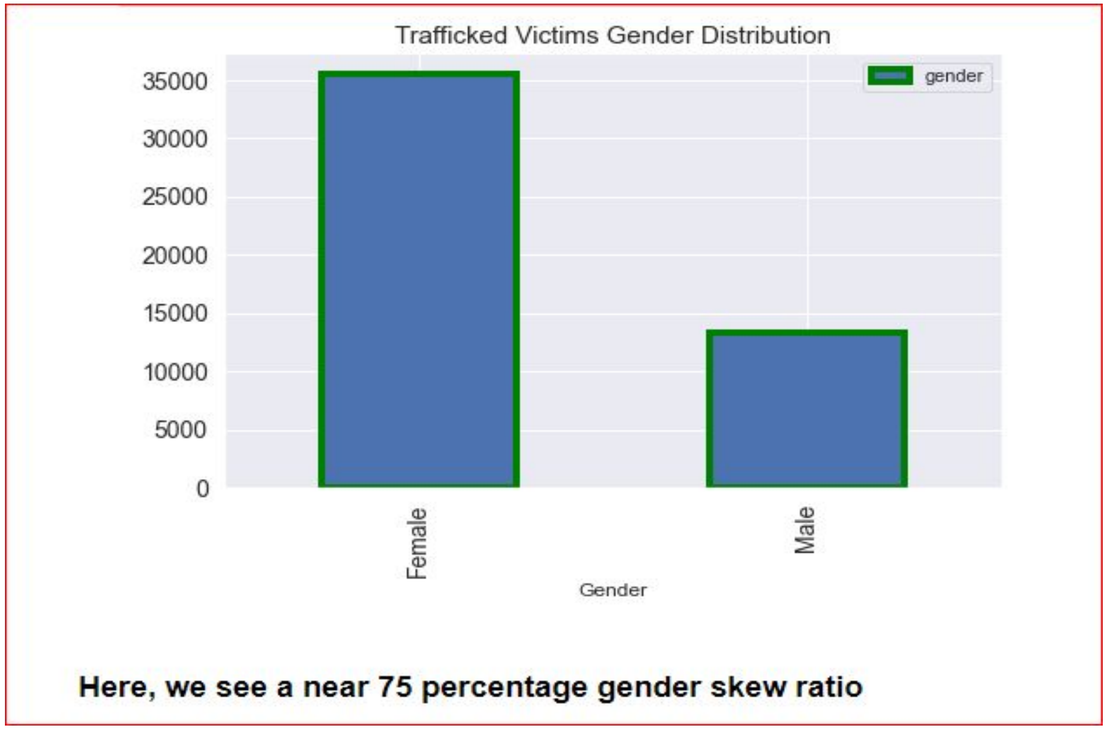
- How was the dataset information acquire

```
['Case Management', 'Hotline']
```

- What number of Trafficked Victims are of unknown Citizenship

```
Trafficked Victims of unknown Citizenship are:

9136 (18.72%)
```



Here, we see a near 75 percentage gender skew ratio

- What number of Victims were exploited in Unknown Countries

```
 11019 (22.58%)
```

- Victims Gender Distribution.

```
Female    35534
Male      13267
```

ReDI School of Digital Integration

These are questions I asked, and the Answers obtained: contd...

- ## Number of reported cases by Country

```
PH    11365
UA     7761
MD     5901
US     3636
KH     1979
```

(These are the first leading 5 countries. Too many for this space.)

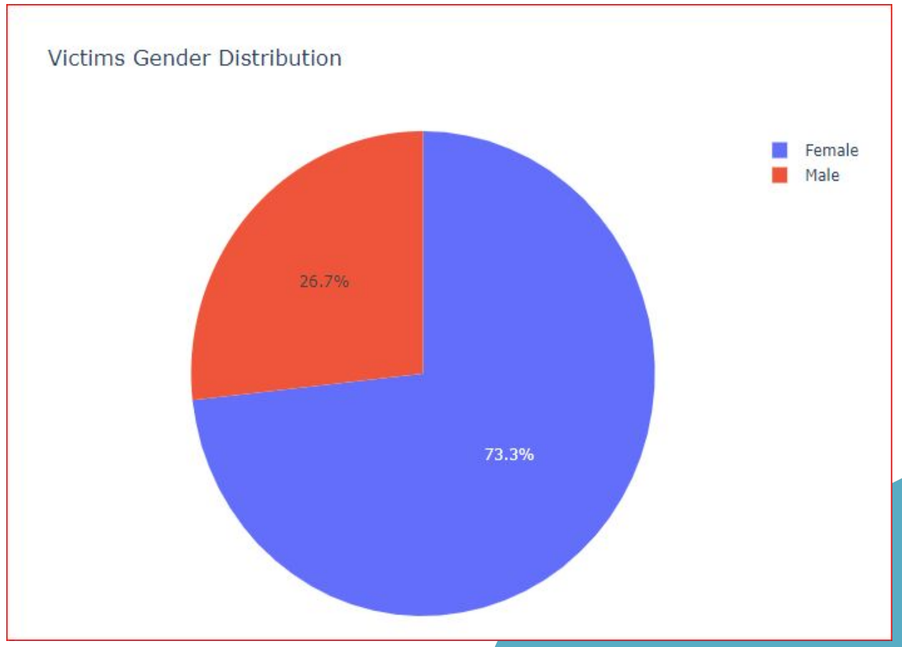
- ## Victims Age and Gender Distribution.

```
    gender ageBroad  Number of Trafficked Individuals
0   Female      0--8                               841
1   Female    18--20                              4227
2   Female    21--23                              3734
3   Female    24--26                              2816
4   Female    27--29                              1869
5   Female    30--38                              4222
6   Female    39--47                              1366
7   Female       48+                               865
8   Female     9--17                              6754
9     Male      0--8                               715
10    Male    18--20                               221
11    Male    21--23                               479
12    Male    24--26                               648
13    Male    27--29                               772
14    Male    30--38                              2869
15    Male    39--47                              1455
16    Male       48+                               695
17    Male     9--17                              1891

'Age and Gender distribution of Trafficked persons'
```

Cumulative Victims Age Distribution

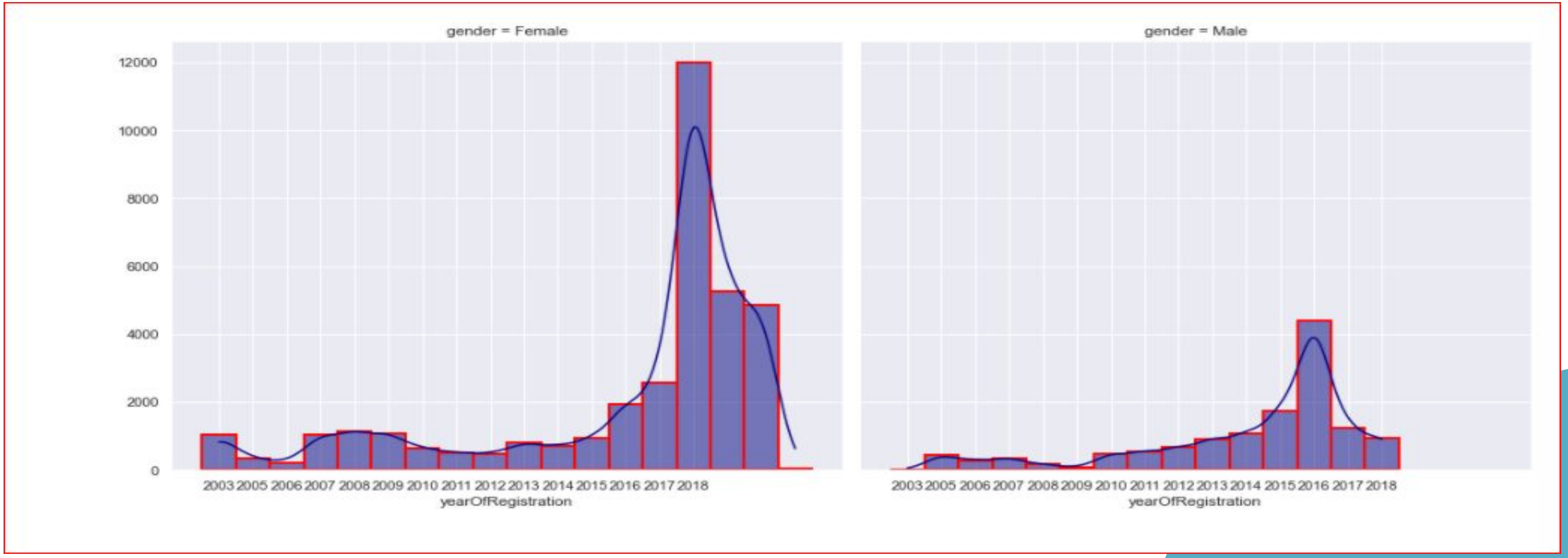
```
00--08    1556
09--17    8645
18--20    4448
21--23    4213
24--26    3464
27--29    2641
30--38    7091
39--47    2821
48+       1560
```



ReDI School of Digital Integration

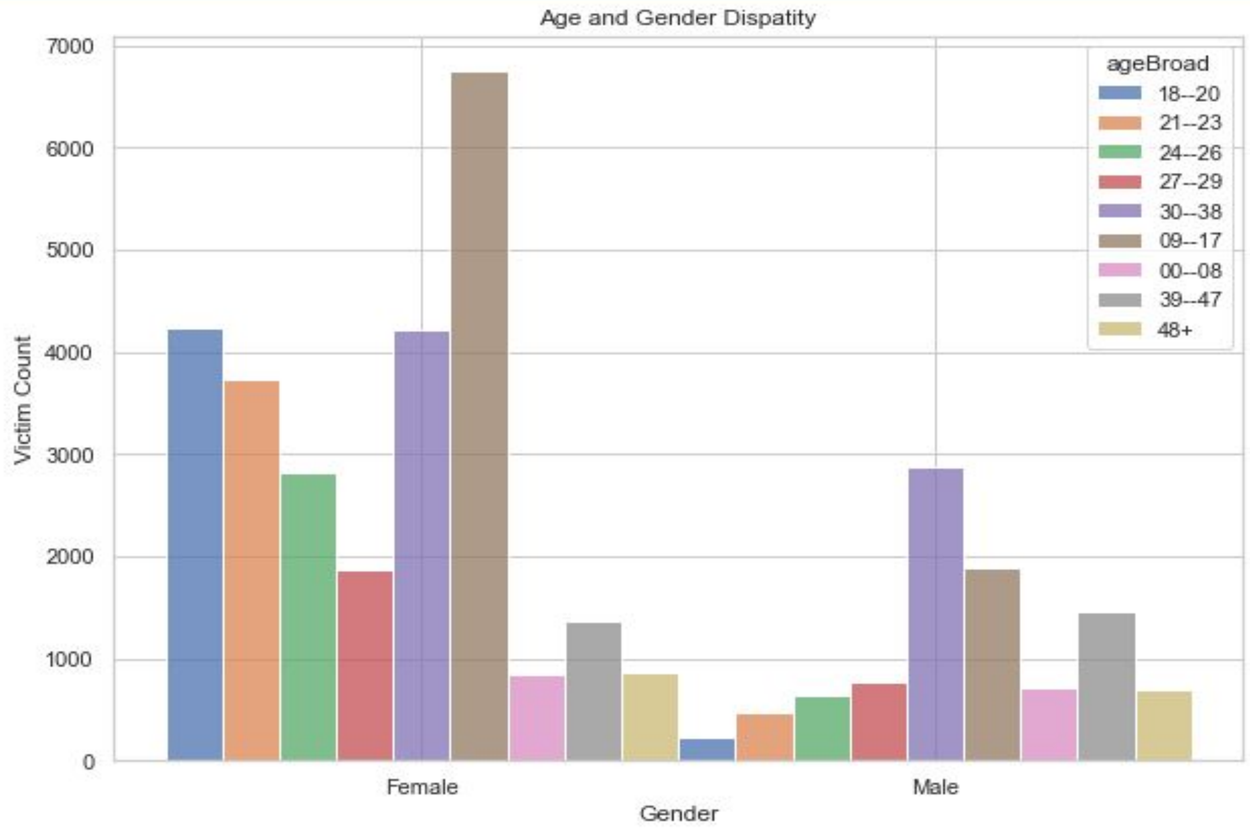
Human trafficking development over time

The following graph shows the time development of trafficking in Humans between 2003 and 2018. We continue to observe the incremental lopsidedness in victims population against females plateauing in 2018. Conversely, the male graph plateaued two years earlier in 2016.¶



ReDI School of Digital Integration

Victims age range and Gender Disparity



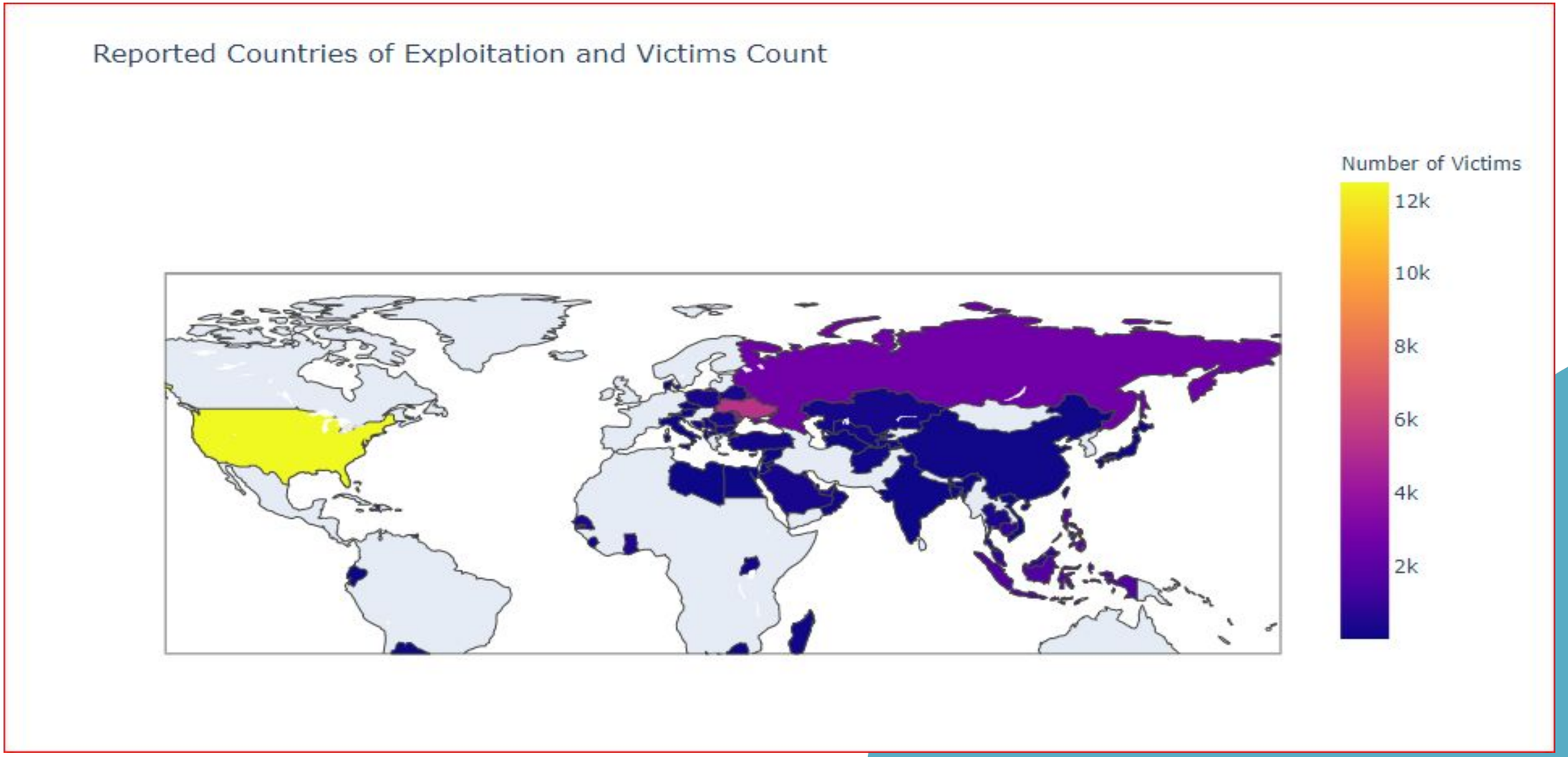
ReDI School of Digital Integration

I also added a Pivot table showing countries where Trafficked victims were exploited, and numbers of exploited victims.

CountryOfExploitation / yearOfRegistration	-99	00	AE	AF	AL	AR	AT	BA	BD	BG	BH	BY	CN	CY	CZ	DK	EC	EG	GH	HK	HT	ID	IN	IT	JO	JP	KH
2002	562	0	0	0	0	0	0	41	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
2003	252	0	0	0	0	0	0	86	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
2004	29	0	11	0	0	0	0	0	0	0	0	0	0	11	0	0	0	0	0	0	0	0	0	0	0	0	0
2005	0	0	54	0	22	0	0	0	0	130	0	0	0	0	0	0	0	0	514	0	0	0	0	12	0	0	0
2006	0	0	32	0	0	0	0	0	0	0	0	0	0	0	14	0	0	0	30	0	0	21	0	0	0	0	0
2007	0	0	68	0	0	0	24	0	0	226	0	0	0	0	24	0	0	0	0	0	0	0	0	0	0	0	0
2008	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	12	0	0	0
2009	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
2010	0	11	0	0	0	0	0	0	0	0	0	0	0	0	0	0	13	0	0	0	0	0	0	0	0	0	0
2011	0	672	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
2012	0	161	0	0	0	18	0	0	0	0	0	0	0	0	0	0	0	0	0	0	285	0	0	0	32	0	0
2013	0	0	0	12	0	0	0	0	0	0	0	301	0	0	0	0	0	0	0	0	54	29	0	0	53	0	0
2014	141	0	290	71	0	0	0	0	0	0	22	58	0	0	0	0	0	13	0	0	0	249	0	0	23	0	0
2015	0	0	0	0	0	0	0	0	0	0	0	22	0	0	0	17	0	0	0	0	0	1238	0	0	0	0	0
2016	8663	0	13	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	240	0	0	0	111	0
2017	528	0	36	0	0	0	0	0	0	0	0	0	65	0	0	0	0	0	0	31	0	0	0	0	0	0	86
2018	0	0	0	0	0	0	0	0	26	0	0	19	14	0	0	0	0	0	0	0	0	0	46	0	0	0	914
2019	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

ReDI School of Digital Integration

This is an interactive world map that details countries of exploitation and number of exploited victims.



Next Stop:

Machine Learning: Work in progress

Machine Learning

Clean up the data for analysis. The target is to determine the type of exploration. correlateted features were removed.

In [47]:

```
'''df2 = df.copy()                                                    # still requires work
df2 = df2.loc[df.yearOfRegistration > 2016
df2 = df2.drop(['meansOfControlConcatenated',
                        'isForcedLabour',
                        'isSexualExploit',
                        'isOtherExploit',
                        'isSexAndLabour',
                        'isForcedMarriage',
                        'isForcedMilitary',
                        'isOrganRemoval',
                        'typeOfLabourAgriculture',
                        'typeOfLabourAquafarming',
                        'typeOfLabourBegging',
                        'typeOfLabourConstruction',
                        'typeOfLabourDomesticWork',
                        'typeOfLabourHospitality',
                        'typeOfLabourIllicitActivities',
                        'typeOfLabourManufacturing',
                        'typeOfLabourMiningOrDrilling',
                        'typeOfLabourPeddling',
                        'typeOfLabourTransportation',
                        'typeOfLabourOther',
                        'typeOfLabourNotSpecified',
                        'typeOfLabourConcatenated',
                        'typeOfSexProstitution',
                        'typeOfSexPornography',
                        'typeOfSexRemoteInteractiveServices',
                        'typeOfSexPrivateSexualServices',
                        'typeOfSexConcatenated'], axis=1)

df2.shape
```

ReDI School of Digital Integration

Thats all she wrote for now folks. So, what did we do here?.

- I imported data.
- Cleaned data.
- Analysed Data.
- Visualized Data.
- And, I am currently working on the machine Learning Algorithm.

ReDI School of Digital Integration

This need to be said:

ReDi is a great Environment to Sturdy.

Great Volunteer teachers, and a fantastic initiative geared towards knowledge equality.

The world needs ReDi and to them I say super big,

THANK YOU!

ReDI School of
Digital Integration

And you?

Thank you for coming along on this journey.

I enjoyed every bit of my time working on this project.

THANK YOU!