Introduction

For the next step of the interview process, we would like you to collect some data and compile it in a suitable way. World Data Lab is embarking on an ambitious project together with the Mastercard Foundation to build a 'Youth employment clock' with data from across 7 countries: Kenya, Rwanda, Ghana, Senegal, Ethiopia, Nigeria and Uganda that will track employment trends in real time. With Rwanda being a key contributor to this, to assess your competencies, kindly follow the prompt below:

Prompt

We are interested in data for Rwanda on the following variables:

● Number of employed by region, age (17-35 and 36-64 years), gender, education, and sector,

● Number of unemployed by region, age (17-35 and 36-64 years), gender, education, and

● Number of inactive by region, age (17-35 and 36-64 years), gender, education, and reason for inactivity.

Please note that the data should be representative on the subnational level! Feel free to use any credible data sources (e.g., national statistical offices or central banks) you can find. Clean the data (either manually or using code; if you use code, please also share it) and put it into an Excel file (can also be multiple sheets). Please make sure to properly document the different steps (incl. sources) and share that information as well! The data should be compiled in such a way that it can be easily read using R or Python codes.

Understanding the experimental design

The experimental design for this project is as follows:

1. Loading the libraries and dataset
2. Data cleaning and preparation
3. Excel file generation

Data Source

This data set presents the results of the labour Force Survey for the fourth quarter (Q1) of the year 2022 whose data collection was conducted from 1st to 28th February 2022 in all the districts of Rwanda using Telephone based interviewing mode by the Rwanda National institute of statistics.

The survey and data can be accessed here: https://www.statistics.gov.rw/publication/1808

Data relevance

Below is a description of the dataset. The data set is an excel file with multiple sheets containing tables showing various labour indicators.

LIST OF TABLES

1. Table B.1: Summary labour force indicators, February-22 (Q1)
2. Table B.2: Population by sex, age group and urban/rural area, February-22 (Q1)
3. Table B.3: Households by household size, sex of head of household and urban/rural area, February-22 (Q1)
4. Table B.4: Population 16 years old and over by labour force status, sex, age group, and urban/rural area, February-22 (Q1)
5. Table B.5: Population 16 years old and over by labour force status and level of educational attainment , February-22 (Q1)
6. Table B.6: Population 16 years old and over by labour force status and marital status, February-22 (Q1)
7. Table B.7:Employed population by sex, age group, and urban/rural area, February-22 (Q1)
8. Table B.8: Employed population by sex, occupation group, and urban/rural area, February-22 (Q1)
9. Table B.9: Employed population by sex, educational attainment, and urban/rural area, February-22 (Q1)
10. Table B.10:Employed population by sex, branch of economic activity, and urban/rural area, February-22 (Q1)
11. Table B.11: Educational attainement and field of Education by Labour market status, February-22 (Q1)
12. Table B.12: Employed population by sex, status in employment, and urban/rural area, February-22 (Q1)
13. Table B.13: Employed population by sex, hours usually worked per week at all jobs, and urban/rural area, February-22 (Q1)
14. Table B.14: Youth Population by sex, and residential area, February-22 (Q1)

15. Table B.15: Youth Unemployed by sex, duration of seeking employment, and urban/rural area, February-22 (Q1)

16. Table B.16:Youth not in employment and not currently in education or training by sex, age group, and urban/rural area, February-22 (Q1)

17. Table B.17:Unemployed population by sex, broad age group and urban/rural area, February-22 (Q1)
18. Table B.18: Unemployed population by sex, level of educational, and urban/rural area, February-22 (Q1)
19. Table B.19A: Unemployed population(who looked for a job) by sex,method of seeking employment, and urban/rural area, February-22 (Q1)
20. Table B.20: Unemployed population(who looked for a job) by sex, duration of seeking employment, and urban/rural area, February-22 (Q1)
21. Table B.21: Time related under employment by age group sex and area of residence, February-22 (Q1)

From the prompt directions we need only tables 4,5,7,9,10,17 and 18. However, for the inactive labour I have failed to find a data source that specifies the reason for inactivity from official and unofficial sites. I assumed Rwanda doesn't collect that type of data.

The region data provided at a sub national level only has 2 variables, rural and urban.

The age range specified in the prompt, 17-35 yrs and 36-64 yrs , years are presented as 16-34 yrs and 35-64 yrs respectively as per the data source.

Load the dataset

Loading the data set into a pandas dataframe after importing our python libraries.

In [375…

```
#import libraries
import numpy as np
import pandas as pd
pd.options.mode.chained_assignment = None  # default='warn'

# Read all sheets in your File
df = pd.read_excel('RLFS Tables_ February_2022_Q1.xls', sheet_name=None)
```

Preview our dataset

In [375…

```
#preview our dataset

#uncomment the line below to preview the dataset(avoided on pdf due to messy pages)
#df
```

The CSV has multiple datasets on multiple sheets. We'll print the sheet names in an ordered dictionary so as to easily load only the dataframes we need from the 7 sheets (tables 4,5,7,9,10,17,18).

In [375…

```
# Prints all the sheets name in an ordered dictionary
print(df.keys())
```

```
dict_keys(['List Of Tables', 'Table 1', 'Table 2-3', 'Table 4', 'Table 5', 'Table 6', 'T
able 7-8 ', 'Table 9', 'Table10', 'Table 11', 'Table 12-13', 'Table 14', 'Table15', 'Tab
le 16 ', 'Table17-18', 'Table 19-20', 'Table 21'])
```

Since we need only tables (4,5,7,9,10,17,18) I'm creating the first dataframe that will be the first sheet on our excel file that will contain our list/table of contents. This is useful for knowing which sheet contains which labour force demographic indicator.

Dataframe 0

Created a dataframe for the table of contents which will be the first sheet of our excel output

In [375…

```
#First sheet of the csv includes our created list of tables for containing sheets infor
# initialize list elements
data = [['Table 1','Employed population by gender, age, and region'],['Table 2','Employ

# Create the pandas DataFrame with column name is provided explicitly
df0= pd.DataFrame(data, columns=['Tables', 'Labour Force Indicator'])

# print dataframe.
df0
```

Out[375…

	Tables	Labour Force Indicator
0	Table 1	Employed population by gender, age, and region
1	Table 2	Employed population by education
2	Table 3	Employed population by sector
3	Table 4	Unemployed population by gender,age and region
4	Table 5	Unemployed population by education
5	Table 6	Inactive by age,gender and region
6	Table 7	Inactive by education

We'll load the sheets into dataframes from the 7 sheets of our CSV source file (tables 4,5,7,9,10,17,18).

In [375…

```
#Load the sheets needed into seperate dataframes newly created
df1 = pd.read_excel('RLFS Tables_ February_2022_Q1.xls', sheet_name='Table 7-8 ')
df2 = pd.read_excel('RLFS Tables_ February_2022_Q1.xls', sheet_name='Table 9')
df3 = pd.read_excel('RLFS Tables_ February_2022_Q1.xls', sheet_name='Table10')
df4 = pd.read_excel('RLFS Tables_ February_2022_Q1.xls', sheet_name='Table17-18')
df6 = pd.read_excel('RLFS Tables_ February_2022_Q1.xls', sheet_name='Table 4')
df7 = pd.read_excel('RLFS Tables_ February_2022_Q1.xls', sheet_name='Table 5')

#seperate tables on sheet 4 into 2 dataframes
df_4 = df4.iloc[:8,:]
df5 = df4.iloc[11:,:]
```

Data Cleaning

As the dataset is small I checked for anomalies and replaced them manually with code as I worked on each dataframe/sheet.

Dataframe 1

In [376…

```
#preview our dataframe
df1
```

Out[376…

	Table B.7:Employed population by sex, age group, and urban/rural area, February-22 (Q1)	Unnamed: 1	Unnamed: 2	Unnamed: 3	Unnamed: 4	Unnamed: 5	Unnamed: 6	Unnamed: 7
0	NaN	Total	Sex	NaN	Residential area	NaN	Participated in	Not participated
1	NaN	NaN	Male	Female	Urban	Rural	subsistence	in subsistence
2	NaN	NaN	NaN	NaN	NaN	NaN	agriculture	agriculture
3	Employed population 16+	3585651	1949119	1636532	838756	2746895	1495054	2090597
4	16-19 yrs	255131	130850	124282	46705	208426	75506	179625
5	20-24 yrs	460470	248410	212060	94837	365632	147131	313338
6	25-29 yrs	521936	278458	243477	121022	400914	199191	322745
7	30-34 yrs	543455	315782	227673	147619	395836	210348	333107
8	35- 39 yrs	528368	285830	242538	126859	401509	240545	287823
9	40-44 yrs	441428	241770	199658	118781	322647	208816	232612
10	45-49 yrs	301756	155797	145959	72960	228796	143861	157895
11	50-54 yrs	204672	102941	101731	44361	160311	113540	91132
12	55-59 yrs	145359	83882	61478	34451	110908	56306	89053
13	60-64 yrs	105416	56065	49351	17057	88358	58741	46675
14	65-69 yrs	48365	30026	18339	7010	41355	32555	15811
15	70-74 yrs	23108	13734	9374	5511	17597	8514	14594
16	75+	6187	5575	612	1582	4605	0	6187
17	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN

	Table B.7:Employed population by sex, age group, and urban/rural area, February-22 (Q1)	Unnamed: 1	Unnamed: 2	Unnamed: 3	Unnamed: 4	Unnamed: 5	Unnamed: 6	Unnamed: 7
18	Table B.8: Employed population by sex, occupat...	NaN	NaN	NaN	NaN	NaN	NaN	NaN
19	NaN	Total	Sex	NaN	Residential area	NaN	Participated in	Not participated
20	NaN	NaN	Male	Female	Urban	Rural	subsistence	in subsistence
21	NaN	NaN	NaN	NaN	NaN	NaN	agriculture	agriculture
22	Occupation group (ISCO High level)	3585651	1949119	1636532	838756	2746895	1495054	2090597
23	Managers	44042	33982	10060	27335	16707	2883	41160
24	Professionals	211286	122742	88544	103929	107357	20375	190910
25	Technicians and associate professionals	24767	20064	4703	18511	6256	672	24096
26	Clerical support workers	24286	9982	14304	18608	5678	0	24286
27	Service and sales workers	507256	240308	266948	220528	286729	128610	378646
28	Skilled agricultural, forestry and fishery wor...	160010	86720	73291	20503	139508	19417	140593
29	Craft and related trades workers	253785	182049	71736	90595	163191	67034	186751
30	Plant and machine operators and assemblers	99043	94796	4247	56960	42083	9945	89099
31	Elementary occupations	2261175	1158475	1102699	281788	1979386	1246118	1015056
32	NaN	NaN		NaN		NaN		NaN

In [376...

```
#drop unnecessary columns
```

```
df1.drop(
    columns=["Unnamed: 6","Unnamed: 7"],inplace = True
)

#drop the lower table from our datframe since it is not needed
df1.drop(range(14,33), inplace = True)

#Sum rows using the correct age group intervals
df1.loc[4] += df1.loc[5]
df1.loc[4]+= df1.loc[6]
df1.loc[4]+= df1.loc[7]
df1.loc[8] += df1.loc[9]
df1.loc[8]+= df1.loc[10]
df1.loc[8]+= df1.loc[11]
df1.loc[8]+= df1.loc[12]
df1.loc[8]+= df1.loc[13]

#drop unnecessary rows
df1.drop(df1.index[[2,3,5,6,7,9,10,11,12,13]], inplace = True)

#rename age intervals after summation.
df1.at[4, 'Table B.7:Employed population by sex, age group, and urban/rural area, Febru
df1.at[8, 'Table B.7:Employed population by sex, age group, and urban/rural area, Febru

df1.at[0, 'Table B.7:Employed population by sex, age group, and urban/rural area, Febru
df1.at[0, 'Unnamed: 2']= "Gender"
df1.at[0, 'Unnamed: 4']= "Region"
#view our dataframe
df1
```

Out[376...

	Table B.7:Employed population by sex, age group, and urban/rural area, February-22 (Q1)	Unnamed: 1	Unnamed: 2	Unnamed: 3	Unnamed: 4	Unnamed: 5
0	Age	Total	Gender	NaN	Region	NaN
1	NaN	NaN	Male	Female	Urban	Rural
4	16-34 yrs	1780992	973500	807492	410183	1370808
8	35-64 yrs	1726999	926285	800715	414469	1312529

Dataframe 2

Preview our dataframe

In [376...

```
#view dataframe
df2
```

Out[376...

	Unnamed: 0	Unnamed: 1	Unnamed: 2	Unnamed: 3	Unnamed: 4	Unnamed: 5	Unnamed: 6	Unnamed: 7
0	Table B.9: Employed population by sex, educati...	NaN	NaN	NaN	NaN	NaN	NaN	NaN

	Unnamed: 0	Unnamed: 1	Unnamed: 2	Unnamed: 3	Unnamed: 4	Unnamed: 5	Unnamed: 6	Unnamed: 7
1	NaN	Total	Sex	NaN	Residential area	NaN	Participated in	Not participated
2	NaN	NaN	Male	Female	Urban	Rural	subsistence	in subsistence
3	NaN	NaN	NaN	NaN	NaN	NaN	agriculture	agriculture
4	Employed population	3585419	1949119	1636300	838525	2746895	1495054	2090366
5	None	1681699	903279	778420	200462	1481237	897742	783957
6	Primary	1151227	625677	525550	255436	895792	477539	673688
7	Lower secondary	218596	106134	112462	90775	127821	65309	153287
8	Upper secondary	306344	173173	133172	134492	171852	40588	265756
9	University	227554	140856	86697	157360	70194	13876	213678

In [376…

```
#drop unnecessary columns
df2.drop(
    columns=["Unnamed: 6","Unnamed: 7"],inplace = True
)

#drop unnecessary rows
df2.drop(df2.index[[0,3,4]], inplace = True)

#rename Education column header
df2.at[1, 'Unnamed: 0']= "Education"

#rename Gender column header
df2.at[1, 'Unnamed: 2']= "Gender"

#rename Region column header
df2.at[1, 'Unnamed: 4']= "Region"

#view our dataframe
df2
```

Out[376…

	Unnamed: 0	Unnamed: 1	Unnamed: 2	Unnamed: 3	Unnamed: 4	Unnamed: 5
1	Education	Total	Gender	NaN	Region	NaN
2	NaN	NaN	Male	Female	Urban	Rural
5	None	1681699	903279	778420	200462	1481237
6	Primary	1151227	625677	525550	255436	895792
7	Lower secondary	218596	106134	112462	90775	127821
8	Upper secondary	306344	173173	133172	134492	171852
9	University	227554	140856	86697	157360	70194

Dataframe 3

Preview our dataframe

In [376…

```
#view dataframe
df3
```

Out[376…

	Table B.10:Employed population by sex, branch of economic activity, and urban/rural area, February-22 (Q1)	Unnamed: 1	Unnamed: 2	Unnamed: 3	Unnamed: 4	Unnamed: 5	Unnamed: 6	Unnamed: 7
0	NaN	Total	Sex	NaN	Residential area	NaN	Participated in	Not participated
1	NaN	NaN	Male	Female	Urban	Rural	subsistence	in subsistence
2	NaN	NaN	NaN	NaN	NaN	NaN	agriculture	agriculture
3	Employed population	3585651	1949119	1636532	838756	2746895	1495054	2090597
4	Agriculture, forestry and fishing	1881040	887627	993414	109108	1771933	1133164	747876
5	Mining and quarrying	27353	26752	602	0	27353	8545	18808
6	Manufacturing	158469	72513	85956	53493	104976	37523	120946
7	Electricity, gas, steam and air conditioning s...	4318	3937	382	611	3708	337	3981
8	Water supply, sewerage and waste management	5191	4387	804	1496	3695	0	5191
9	Construction	328658	273912	54746	97456	231202	106059	222599
10	Wholesale, retail trade, repair of motor vehic...	360222	156448	203774	169004	191218	97665	262557
11	Transportation and storage	166739	162200	4539	62469	104271	27255	139484
12	Accommodation and food service activities	59181	35242	23939	27430	31750	10404	48777

	Table B.10:Employed population by sex, branch of economic activity, and urban/rural area, February-22 (Q1)	Unnamed: 1	Unnamed: 2	Unnamed: 3	Unnamed: 4	Unnamed: 5	Unnamed: 6	Unnamed: 7
13	Information and communication	9905	6399	3506	7754	2151	0	9905
14	Financial and insurance activities	29714	13742	15972	18620	11094	2080	27634
15	Real estate activities	1472	1238	234	1220	252	0	1472
16	Professional, scientific and technical activities	25109	15055	10054	15736	9373	950	24158
17	Administrative and support service activities	41692	32976	8716	24585	17107	4233	37459
18	Public administration and defence	59168	47602	11566	28800	30368	5816	53352
19	Education	134330	70660	63670	46263	88067	26097	108233
20	Human health and social work activities	46600	21496	25104	34963	11637	3175	43425
21	Arts, entertainment and recreation	8852	4949	3903	6116	2736	829	8022
22	Other service activities	103002	61739	41262	37167	65834	26378	76624
23	Activities of households as employers	130875	48710	82165	92706	38169	4543	126332
24	Activities of extraterritorial organizations a...	3760	1537	2223	3760	0	0	3760

In [376...

```
#drop unnecessary columns
df3.drop(
    columns=["Unnamed: 6","Unnamed: 7"],inplace = True
)
#drop unnecessary rows
df3.drop(df3.index[[2,3]], inplace = True)
```

```
#rename Sector column header
df3.at[0, 'Table B.10:Employed population by sex, branch of economic activity, and urba

#rename Gender column header
df3.at[0, 'Unnamed: 2']= "Gender"

#rename Region column header
df3.at[0, 'Unnamed: 4']= "Region"

#view our dataframe
df3
```

Out[376...

	Table B.10:Employed population by sex, branch of economic activity, and urban/rural area, February-22 (Q1)	Unnamed: 1	Unnamed: 2	Unnamed: 3	Unnamed: 4	Unnamed: 5
0	Sector	Total	Gender	NaN	Region	NaN
1	NaN	NaN	Male	Female	Urban	Rural
4	Agriculture, forestry and fishing	1881040	887627	993414	109108	1771933
5	Mining and quarrying	27353	26752	602	0	27353
6	Manufacturing	158469	72513	85956	53493	104976
7	Electricity, gas, steam and air conditioning s...	4318	3937	382	611	3708
8	Water supply, sewerage and waste management	5191	4387	804	1496	3695
9	Construction	328658	273912	54746	97456	231202
10	Wholesale, retail trade, repair of motor vehic...	360222	156448	203774	169004	191218
11	Transportation and storage	166739	162200	4539	62469	104271
12	Accommodation and food service activities	59181	35242	23939	27430	31750
13	Information and communication	9905	6399	3506	7754	2151
14	Financial and insurance activities	29714	13742	15972	18620	11094
15	Real estate activities	1472	1238	234	1220	252
16	Professional, scientific and technical activities	25109	15055	10054	15736	9373
17	Administrative and support service activities	41692	32976	8716	24585	17107
18	Public administration and defence	59168	47602	11566	28800	30368
19	Education	134330	70660	63670	46263	88067
20	Human health and social work activities	46600	21496	25104	34963	11637
21	Arts, entertainment and recreation	8852	4949	3903	6116	2736
22	Other service activities	103002	61739	41262	37167	65834
23	Activities of households as employers	130875	48710	82165	92706	38169

	Table B.10:Employed population by sex, branch of economic activity, and urban/rural area, February-22 (Q1)	Unnamed: 1	Unnamed: 2	Unnamed: 3	Unnamed: 4	Unnamed: 5
24	Activities of extraterritorial organizations a...	3760	1537	2223	3760	0

Dataframe 4

Preview our dataframe

In [376...
```
#view dataframe
df_4
```

Out[376...

	Table B.17:Unemployed population by sex, broad age group and urban/rural area, February-22 (Q1)	Unnamed: 1	Unnamed: 2	Unnamed: 3	Unnamed: 4	Unnamed: 5	Unnamed: 6	Unnamed: 7
0	NaN	Total	Sex	NaN	Residencial area	NaN	Participated in subsistence agriculture	Not participated in subsistence agriculture
1	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
2	NaN	NaN	Male	Female	Urban	Rural	NaN	NaN
3	Unemployed population 16+	707654	344270	363384	181462	526192	383418	324236
4	16-24 yrs	211757	104656	107101	38954	172803	106710	105048
5	25-34 yrs	248148	110426	137723	70042	178106	113835	134314
6	35-54 yrs	208982	105715	103267	65529	143453	131745	77237
7	55-64 yrs	32035	20022	12013	6711	25324	26769	5266

In [376...
```
#drop unnecessary columns and rows
df_4.drop(
    columns=["Unnamed: 6","Unnamed: 7"],inplace = True
)
df_4.drop(df_4.index[1], inplace = True)
df_4.drop(df_4.index[2], inplace = True)

#Sum rows using the correct age group intervals while dropping uneccessary rows
df_4.loc[4] += df_4.loc[5]
df_4.drop([5], inplace=True)
df_4.loc[6] += df_4.loc[7]
df_4.drop([7], inplace=True)

#rename age intervals after summation.
```

```
df_4.at[4, 'Table B.17:Unemployed population by sex, broad age group and urban/rural ar
df_4.at[6, 'Table B.17:Unemployed population by sex, broad age group and urban/rural ar

#rename Age column header
df_4.at[0, 'Table B.17:Unemployed population by sex, broad age group and urban/rural ar

#rename Gender column header
df_4.at[0, 'Unnamed: 2']= "Gender"

#rename Region column header
df_4.at[0, 'Unnamed: 4']= "Region"


#view our dataframe
df_4
```

Out[376...

	Table B.17:Unemployed population by sex, broad age group and urban/rural area, February-22 (Q1)	Unnamed: 1	Unnamed: 2	Unnamed: 3	Unnamed: 4	Unnamed: 5
0	Age	Total	Gender	NaN	Region	NaN
2	NaN	NaN	Male	Female	Urban	Rural
4	16-34 yrs	459905	215082	244824	108996	350909
6	35-64 yrs	241017	125737	115280	72240	168777

Dataframe 5

Preview our dataframe

In [376...

```
#view dataframe
df5
```

Out[376...

	Table B.17:Unemployed population by sex, broad age group and urban/rural area, February-22 (Q1)	Unnamed: 1	Unnamed: 2	Unnamed: 3	Unnamed: 4	Unnamed: 5	Unnamed: 6	Unnamed: 7
11	NaN	Total	Sex	NaN	Area of Residence	NaN	Participated in subsistence agriculture	Not participated in subsistence agriculture
12	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
13	NaN	NaN	Male	Female	Urban	Rural	NaN	NaN
14	Unemployed population 16+	707654	344270	363384	181462	526192	383418	324236
15	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN

	Table B.17:Unemployed population by sex, broad age group and urban/rural area, February-22 (Q1)	Unnamed: 1	Unnamed: 2	Unnamed: 3	Unnamed: 4	Unnamed: 5	Unnamed: 6	Unnamed: 7
16	None	260394	136316	124078	28418	231976	171240	89154
17	Primary	203533	106414	97119	40456	163076	138565	64967
18	Lower secondary	55290	20551	34739	18593	36697	26054	29236
19	Upper secondary	144066	61657	82409	61900	82166	42552	101514
20	University	44372	19332	25040	32094	12277	5007	39365

In [376...

```
#drop unnecessary rows
df5.drop(df5.index[[1,3,4,]], inplace = True)

#drop unnecessary columns
df5.drop(
    columns=["Unnamed: 6","Unnamed: 7"],inplace = True
)
#rename Education column header
df5.at[11, 'Table B.17:Unemployed population by sex, broad age group and urban/rural ar

#rename Gender column header
df5.at[11, 'Unnamed: 2']= "Gender"

#rename Region column header
df5.at[11, 'Unnamed: 4']= "Region"

#view our dataframe
df5
```

Out[376...

	Table B.17:Unemployed population by sex, broad age group and urban/rural area, February-22 (Q1)	Unnamed: 1	Unnamed: 2	Unnamed: 3	Unnamed: 4	Unnamed: 5
11	Education	Total	Gender	NaN	Region	NaN
13	NaN	NaN	Male	Female	Urban	Rural
16	None	260394	136316	124078	28418	231976
17	Primary	203533	106414	97119	40456	163076
18	Lower secondary	55290	20551	34739	18593	36697
19	Upper secondary	144066	61657	82409	61900	82166
20	University	44372	19332	25040	32094	12277

Dataframe 6

Preview our dataframe

In [377…

```
#view our dataframe
df6
```

Out[377…

	Unnamed: 0	Table B.4: Population 16 years old and over by labour force status, sex, age group, and urban/rural area, February-22 (Q1)	Unnamed: 2	Unnamed: 3	Unnamed: 4	Unnamed: 5	Unnamed: 6	Unnamed: 7	Un
0	NaN	15	Total	Labour force status	NaN	NaN	NaN	Labour force participation rate (%)	Em
1	NaN	NaN	NaN	Labour force	Employed	Unemployed	Outside labour force	NaN	
2	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	
3	NaN	Total Population 16 yrs and over	7873326	4293305	3585651	707654	3580022	54.52975	
4	1	16-24 yrs	2407580	927358	715601	211757	1480222	38.518263	
5	2	25-34 yrs	1759320	1313539	1065391	248148	445780	74.661744	
6	3	35-54 yrs	2403970	1685206	1476224	208982	718764	70.100958	
7	4	55-64 yrs	674731	282810	250775	32035	391921	41.914481	
8	5	65+ yrs	627725	84391	77660	6731	543335	13.443944	
9	1	NaN	NaN	0	NaN	NaN	NaN	NaN	
10	Male	Male Pop. 16+ yrs	3709665	2293389	1949119	344270	1416277	61.821997	
11	3	16-24 yrs	1181932	483915	379259	104656	698016	40.942711	
12	4	25-34 yrs	846080	704666	594240	110426	141414	83.285978	
13	5	35-54 yrs	1107221	892052	786337	105715	215168	80.566752	
14	6	55-64 yrs	302305	159969	139947	20022	142337	52.916425	
15	7	65+ yrs	272127	52786	49335	3451	219342	19.397561	
16	2	NaN	NaN	NaN	NaN	NaN	NaN	NaN	

	Unnamed: 0	Table B.4: Population 16 years old and over by labour force status, sex, age group, and urban/rural area, February-22 (Q1)	Unnamed: 2	Unnamed: 3	Unnamed: 4	Unnamed: 5	Unnamed: 6	Unnamed: 7	Ur
17	Female	Female Pop. 16+ yrs	4163661	1999916	1636532	363384	2163745	48.032633	
18	2	16-24 yrs	1225648	443443	336342	107101	782205	36.18029	
19	3	25-34 yrs	913240	608874	471151	137723	304367	66.67185	
20	4	35-54 yrs	1296750	793154	689887	103267	503596	61.164758	
21	5	55-64 yrs	372425	122841	110828	12013	249584	32.984091	
22	6	65+ yrs	355598	31605	28325	3280	323994	8.887845	
23	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	
24	Urban	Urban Pop. 16+ yrs	1611279	1020218	838756	181462	591061	63.317278	
25	1	16-24 yrs	495942	180496	141542	38954	315446	36.394578	
26	2	25-34 yrs	441555	338683	268641	70042	102872	76.702336	
27	3	35-54 yrs	506208	428490	362961	65529	77718	84.647023	
28	4	55-64 yrs	95446	58219	51508	6711	37227	60.996794	
29	5	65+ yrs	72127	14328	14103	225	57799	19.86496	
30	NaN	NaN	NaN	NaN	NaN	NaN	NaN	0	
31	Rural	Rural Pop. 16+ yrs	6262047	3273087	2746895	526192	2988960	52.268643	
32	1	16-24 yrs	1911638	746862	574059	172803	1164776	39.069217	
33	2	25-34 yrs	1317765	974856	796750	178106	342909	73.977985	
34	3	35-54 yrs	1897762	1256716	1113263	143453	641046	66.220949	
35	4	55-64 yrs	579284	224591	199267	25324	354694	38.770448	
36	5	65+ yrs	555599	70063	63557	6506	485536	12.610354	

Tidying the dataframe

In [377…

```
#replace anomaly "15" with nan
df6.at[0, 'Table B.4: Population 16 years old and over by labour force status, sex, age
```

```
#drop unnecessary columns
df6.drop(
    columns=["Unnamed: 2","Unnamed: 3","Unnamed: 4","Unnamed: 5","Unnamed: 7","Unnamed:
)

#drop more unnecessary rows before we sum the rows for easier and cleaner coding
df6.drop(df6.index[[2,3,8,9,10,15,16,17,22,23,24,29,30,31,36]], inplace = True)

#view our dataframe
df6
```

Out[377…

	Unnamed: 0	Table B.4: Population 16 years old and over by labour force status, sex, age group, and urban/rural area, February-22 (Q1)	Unnamed: 6
0	NaN	NaN	NaN
1	NaN	NaN	Outside labour force
4	1	16-24 yrs	1480222
5	2	25-34 yrs	445780
6	3	35-54 yrs	718764
7	4	55-64 yrs	391921
11	3	16-24 yrs	698016
12	4	25-34 yrs	141414
13	5	35-54 yrs	215168
14	6	55-64 yrs	142337
18	2	16-24 yrs	782205
19	3	25-34 yrs	304367
20	4	35-54 yrs	503596
21	5	55-64 yrs	249584
25	1	16-24 yrs	315446
26	2	25-34 yrs	102872
27	3	35-54 yrs	77718
28	4	55-64 yrs	37227
32	1	16-24 yrs	1164776
33	2	25-34 yrs	342909
34	3	35-54 yrs	641046
35	4	55-64 yrs	354694

Now we adjust the ages by summing rows to attain age in the ranges 16-34 and 35-64

In [377…

```
#Sum the rows in the age ranges 16-34 and 35-64
df6.loc[4] += df6.loc[5]
df6.loc[6] += df6.loc[7]
```

```
df6.loc[11] += df6.loc[12]
df6.loc[13] += df6.loc[14]
df6.loc[18] += df6.loc[19]
df6.loc[20] += df6.loc[21]
df6.loc[25] += df6.loc[26]
df6.loc[27] += df6.loc[28]
df6.loc[32] += df6.loc[33]
df6.loc[34] += df6.loc[35]

#drop unnecesary rows
df6.drop(df6.index[[1,3,5,7,9,11,13,15,17,19,21]], inplace=True)

#rename age intervals after summation.
df6.at[4,'Table B.4: Population 16 years old and over by labour force status, sex, age
df6.at[6, 'Table B.4: Population 16 years old and over by labour force status, sex, age
df6.at[11,'Table B.4: Population 16 years old and over by labour force status, sex, age
df6.at[13, 'Table B.4: Population 16 years old and over by labour force status, sex, ag
df6.at[18,'Table B.4: Population 16 years old and over by labour force status, sex, age
df6.at[20, 'Table B.4: Population 16 years old and over by labour force status, sex, ag
df6.at[25,'Table B.4: Population 16 years old and over by labour force status, sex, age
df6.at[27, 'Table B.4: Population 16 years old and over by labour force status, sex, ag
df6.at[32,'Table B.4: Population 16 years old and over by labour force status, sex, age
df6.at[34, 'Table B.4: Population 16 years old and over by labour force status, sex, ag

#rename Inactive column header
df6.at[0, 'Unnamed: 6']= "Inactive"

#rename Age column header
df6.at[0, 'Table B.4: Population 16 years old and over by labour force status, sex, age

#view our dataframe
df6
```

Out[377…

	Unnamed: 0	Table B.4: Population 16 years old and over by labour force status, sex, age group, and urban/rural area, February-22 (Q1)	Unnamed: 6
0	NaN	Age	Inactive
4	3	16-34 yrs	1926002
6	7	35-64 yrs	1110685
11	7	16-34 yrs	839430
13	11	35-64 yrs	357505
18	5	16-34 yrs	1086572
20	9	35-64 yrs	753180
25	3	16-34 yrs	418318
27	7	35-64 yrs	114945
32	3	16-34 yrs	1507685
34	7	35-64 yrs	995740

Finish cleaning the dataframe

In [377…

```
#replace first column values with np.nan and demographic indicators where necessary
```

```
df6.at[6, 'Unnamed: 0']= np.nan
df6.at[13, 'Unnamed: 0']= np.nan
df6.at[20, 'Unnamed: 0']= np.nan
df6.at[27, 'Unnamed: 0']= np.nan
df6.at[34, 'Unnamed: 0']= np.nan
df6.at[4, 'Unnamed: 0']= "Population"
df6.at[11, 'Unnamed: 0']= "Male"
df6.at[18, 'Unnamed: 0']= "Female"
df6.at[25, 'Unnamed: 0']= "Urban"
df6.at[32, 'Unnamed: 0']= "Rural"

#view data
df6
```

Out[377…

	Unnamed: 0	Table B.4: Population 16 years old and over by labour force status, sex, age group, and urban/rural area, February-22 (Q1)	Unnamed: 6
0	NaN	Age	Inactive
4	Population	16-34 yrs	1926002
6	NaN	35-64 yrs	1110685
11	Male	16-34 yrs	839430
13	NaN	35-64 yrs	357505
18	Female	16-34 yrs	1086572
20	NaN	35-64 yrs	753180
25	Urban	16-34 yrs	418318
27	NaN	35-64 yrs	114945
32	Rural	16-34 yrs	1507685
34	NaN	35-64 yrs	995740

Dataframe 7

Preview our dataframe

In [377…

```
df7
```

Out[377…

	Unnamed: 0	Unnamed: 1	Unnamed: 2	Unnamed: 3	Unnamed: 4	Unnamed: 5	Unnamed: 6	Unnamed: 7	U
0	Table B.5: Population 16 years old and over by...	NaN	NaN	NaN	NaN	NaN	NaN	NaN	
1	Marital status	Total	Labour force status	NaN	NaN	NaN	Labour force participation rate (%)	Employment-population ratio (%)	Un

	Unnamed: 0	Unnamed: 1	Unnamed: 2	Unnamed: 3	Unnamed: 4	Unnamed: 5	Unnamed: 6	Unnamed: 7	U
2	NaN	NaN	Labour force	Employed	Unemployed	Outside labour force	NaN	NaN	
3	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	
4	Population 16 yrs and over	7873095	4293073	3585419	707654	3580022	54.528403	45.540147	
5	None	3572196	1942093	1681699	260394	1630104	54.366922	47.077456	
6	Primary	2576349	1354760	1151227	203533	1221589	52.584491	44.684435	
7	Lower secondary	734020	273886	218596	55290	460135	37.313152	29.78066	
8	Upper secondary	676805	450410	306344	144066	226395	66.549449	45.263259	
9	University	313724	271926	227554	44372	41798	86.676824	72.533182	

In [377…

```
#drop unnecessary columns and rows
df7.drop(df7.index[[0,2,3,4]], inplace = True)
df7.drop(df7.tail(1).index)
df7.drop(
    columns=["Unnamed: 1","Unnamed: 2","Unnamed: 3","Unnamed: 4","Unnamed: 6","Unnamed:
)

#rename education level column header
df7.at[1, 'Unnamed: 0']= "Education"

#rename Inactive column header
df7.at[1, 'Unnamed: 5']= "Inactive"

#view data
df7
```

Out[377…

	Unnamed: 0	Unnamed: 5
1	Education	Inactive
5	None	1630104
6	Primary	1221589
7	Lower secondary	460135
8	Upper secondary	226395
9	University	41798

Generate Excel file

Write Our DataFrames to an excel file

In [377…

```
# Create a Pandas Excel writer using XlsxWriter as the engine.
writer = pd.ExcelWriter('rwanda.xlsx', engine='xlsxwriter')

# Write each dataframe to a different worksheet. dropping the headers and indices where
df0.to_excel(writer, sheet_name='List_Of_Tables',index=False,header=True)
df1.to_excel(writer, sheet_name='Table 1',index=False,header=False)
df2.to_excel(writer, sheet_name='Table 2',index=False,header=False)
df3.to_excel(writer, sheet_name='Table 3',index=False,header=False)
df_4.to_excel(writer, sheet_name='Table 4',index=False,header=False)
df5.to_excel(writer, sheet_name='Table 5',index=False,header=False)
df6.to_excel(writer, sheet_name='Table 6',index=False,header=False)
df7.to_excel(writer, sheet_name='Table 7',index=False,header=False)
# Close the Pandas Excel writer and output the Excel file.
writer.save()
```