CHALLENGE 2

OGUNLEYE SUNDAY SOLOMON

In [250]:

```
#The data was wrangled using excel.
#I created four folders for the datasets(covid11pm, covidother, worldpm, worldother).
#I combined the dataset and loaded them using excel power query.
#I did the joining,sorting and cleaning using excel power query.
```

In [251]:

```
import pandas as pd
import xlrd
import matplotlib.pyplot as plt
import seaborn as sns
from datetime import date
import numpy as np
plt.style.use('ggplot')
```

In [252]:

```
# Export all data
covid11pm = pd.read_excel('C:/Users/sogunleye/Documents/PYTHON/CHALLENGE/covid11pm.xlsx')
covidother =  pd.read_excel('C:/Users/sogunleye/Documents/PYTHON/CHALLENGE/covidother.xlsx')
world11pm = pd.read_excel('C:/Users/sogunleye/Documents/PYTHON/CHALLENGE/world11pm.xlsx')
worldother = pd.read_excel('C:/Users/sogunleye/Documents/PYTHON/CHALLENGE/worldother.xlsx')
```

In [253]:

```
#get a day of the week and month columns
world11pm['month'] = pd.to_datetime(world11pm['Date']).dt.month_name()
world11pm['day'] = pd.to_datetime(world11pm['Date']).dt.day_name()

worldother['month'] = pd.to_datetime(worldother['Date']).dt.month_name()
worldother['day'] = pd.to_datetime(worldother['Date']).dt.day_name()

covid11pm['month'] = pd.to_datetime(covid11pm['Date']).dt.month_name()
covid11pm['day'] = pd.to_datetime(covid11pm['Date']).dt.day_name()

covidother['month'] = pd.to_datetime(covidother['Date']).dt.month_name()
covidother['day'] = pd.to_datetime(covidother['Date']).dt.day_name()
```

2. CASES RECORDED IN USA AND GERMANY IN APRIL (EACH)

In [254]:

```
world11pm.tail()
```

Out[254]:

	Date	Time	Country	Total confirmed	Total_confirmed_today	Total_deaths	Total_death
8431	2020-05-23	23:49:05	zimbabwe	56	5	4	
8432	2020-05-24	23:49:07	zimbabwe	56	5	4	
8433	2020-04-28	23:49:03	zimbabwe	32	1	4	
8434	2020-04-29	23:48:59	zimbabwe	32	1	4	
8435	2020-04-30	23:49:11	zimbabwe	40	8	4	

In [255]:

```
headers = [line.lower().replace(' ', '_') for line in world11pm.columns]
world11pm.columns = headers
world11pm.head()
```

Out[255]:

	date	time	country	total_confirmed	total_confirmed_today	total_deaths	total_dea
0	2020-05-01	23:49:16	afghanistan	2335	164	68	
1	2020-05-02	23:49:17	afghanistan	2469	134	72	
2	2020-05-03	23:49:02	afghanistan	2704	235	85	
3	2020-05-04	23:49:07	afghanistan	2894	190	90	
4	2020-05-05	23:49:08	afghanistan	3224	330	95	

In [256]:

```
#2a CASES CONFIRMED IN USA IN APRIL
cases_us=world11pm[(world11pm['country'] =='USA') & (world11pm['month'] =='April')]
```

In [257]:

```
cases_us.head()
```

Out[257]:

	date	time	country	total_confirmed	total_confirmed_today	total_deaths	total_dea
7559	2020-04-07	23:49:27	USA	396215	29609	12790	
7560	2020-04-07	23:49:27	USA	141942	5267	14045	
7561	2020-04-07	23:49:27	USA	135586	3039	17127	
7562	2020-04-07	23:49:27	USA	109069	11059	10328	
7563	2020-04-07	23:49:27	USA	107663	4289	2016	

In [258]:

```
usa_confirmed_cases = cases_us.total_confirmed.sum()
f"The total number of CONFIRMED CORONAVIRUS CASES for the month of APRIL in USA is {usa
_confirmed_cases}."
```

Out[258]:

```
'The total number of CONFIRMED CORONAVIRUS CASES for the month of APRIL in
USA is 4543913.'
```

In [259]:

```
#2b CASES CONFIRMED IN GERMANY IN APRIL
cases_germany=world11pm[(world11pm['country'] =='germany') & (world11pm['month'] =='Apr
il')]
```

In [260]:

```
cases_germany.head()
```

Out[260]:

	date	time	country	total_confirmed	total_confirmed_today	total_deaths	total_dea
2786	2020-04-13	23:49:52	germany	130072	2218	3194	
2788	2020-04-14	23:49:30	germany	132210	2138	3495	
2790	2020-04-15	23:49:27	germany	134753	2543	3804	
2792	2020-04-16	23:49:31	germany	137698	2945	4052	
2794	2020-04-17	23:49:23	germany	141397	3699	4352	

In [261]:

```
germany_confirmed_cases = cases_germany.total_confirmed.sum()
f"The total number of CONFIRMED CORONAVIRUS CASES for the month of APRIL in GERMANY is {germany_confirmed_cases}."
```

Out[261]:

```
'The total number of CONFIRMED CORONAVIRUS CASES for the month of APRIL in
GERMANY is 1894601.'
```

3. CASES RECORDED IN USA AND GERMANY IN MAY (EACH)

In [262]:

```
#3a CASES CONFIRMED IN USA IN MAY
maycases_us=world11pm[(world11pm['country'] =='USA') & (world11pm['month'] =='May')]
```

In [263]:

```
mayusa_confirmed_cases = maycases_us.total_confirmed.sum()
f"The total number of CONFIRMED CORONAVIRUS CASES for the month of MAY in USA is {mayusa_confirmed_cases}."
```

Out[263]:

```
'The total number of CONFIRMED CORONAVIRUS CASES for the month of MAY in U
SA is 0.'
```

In [264]:

```
#3b CASES CONFIRMED IN GERMANY IN MAY
maycases_germany=world11pm[(world11pm['country'] =='germany') & (world11pm['month'] == 'May')]
```

In [265]:

```
maygermany_confirmed_cases = maycases_germany.total_confirmed.sum()
f"The total number of CONFIRMED CORONAVIRUS CASES for the month of MAY in GERMANY is {maygermany_confirmed_cases}."
```

Out[265]:

```
'The total number of CONFIRMED CORONAVIRUS CASES for the month of MAY in G
ERMANY is 4155362.'
```

4. CASES RECORDED IN ITALY AND GERMANY IN APRIL (EACH)

In [266]:

```
covidother.head()
```

Out[266]:

	Date	Time	countries	infected	infected_today	deaths	deaths_today	recovered	re
0	2020-05-01	10:02:27	Afghanistan	2171	0	64	0	260	
1	2020-05-02	10:02:32	Afghanistan	2469	134	72	4	331	
2	2020-05-03	10:02:26	Afghanistan	2704	235	85	13	345	
3	2020-05-04	10:02:27	Afghanistan	2704	0	85	0	345	
4	2020-04-05	10:35:27	Afghanistan	337	NaN	7	NaN	15	

In [267]:

```
headers = [line.lower().replace(' ', '_').replace('ies','y') for line in covidother.col
umns]
covidother.columns = headers
covidother = covidother.replace(np.nan, 0)
covidother.head()
```

Out[267]:

	date	time	country	infected	infected_today	deaths	deaths_today	recovered	re
0	2020-05-01	10:02:27	Afghanistan	2171	0	64	0	260	
1	2020-05-02	10:02:32	Afghanistan	2469	134	72	4	331	
2	2020-05-03	10:02:26	Afghanistan	2704	235	85	13	345	
3	2020-05-04	10:02:27	Afghanistan	2704	0	85	0	345	
4	2020-04-05	10:35:27	Afghanistan	337	0	7	0	15	

In [268]:

```
#4a CASES CONFIRMED IN ITALY IN APRIL
italy_cases = covidother[(covidother['country']=='Italy') & (covidother['month']=='Apri
l')]
```

In [269]:

```
aprilitaly_cases = italy_cases.infected.sum()
f"The total number of CONFIRMED CORONAVIRUS CASES for the month of APRIL in ITALY is {a
prilitaly_cases}."
```

Out[269]:

```
'The total number of CONFIRMED CORONAVIRUS CASES for the month of APRIL in
ITALY is 3976163.'
```

In [270]:

```
#4b CASES CONFIRMED IN RUSSIA IN APRIL
russia_cases = covidother[(covidother['country']=='Russia') & (covidother['month']=='Ap
ril')]
```

In [271]:

```
aprilrussia_cases = russia_cases.infected.sum()
f"The total number of CONFIRMED CORONAVIRUS CASES for the month of APRIL in RUSSIA is {aprilrussia_cases}."
```

Out[271]:

```
'The total number of CONFIRMED CORONAVIRUS CASES for the month of APRIL in
RUSSIA is 753231.'
```

CASES RECORDED IN ITALY AND RUSSIA IN MAY (EACH)

In [272]:

```
#5a CASES CONFIRMED IN ITALY IN MAY
italy_cases = covidother[(covidother['country']=='Italy') & (covidother['month']=='May')]
mayitaly_cases = italy_cases.infected.sum()
f"The total number of CONFIRMED CORONAVIRUS CASES for the month of MAY in ITALY is {mayitaly_cases}."
```

Out[272]:

```
'The total number of CONFIRMED CORONAVIRUS CASES for the month of MAY in I
TALY is 5054679.'
```

In [273]:

```
#3b CASES CONFIRMED IN RUSSIA IN MAY
russia_cases = covidother[(covidother['country']=='Russia') & (covidother['month']=='May')]
mayrussia_cases = russia_cases.infected.sum()
f"The total number of CONFIRMED CORONAVIRUS CASES for the month of MAY in RUSSIA is {mayrussia_cases}."
```

Out[273]:

```
'The total number of CONFIRMED CORONAVIRUS CASES for the month of MAY in R
USSIA is 5434752.'
```

KICKER

In [274]:

```
headers = [line.lower().replace(' ', '_') for line in world11pm.columns]
world11pm.columns = headers
world11pm.head()
```

Out[274]:

	date	time	country	total_confirmed	total_confirmed_today	total_deaths	total_dea
0	2020-05-01	23:49:16	afghanistan	2335	164	68	
1	2020-05-02	23:49:17	afghanistan	2469	134	72	
2	2020-05-03	23:49:02	afghanistan	2704	235	85	
3	2020-05-04	23:49:07	afghanistan	2894	190	90	
4	2020-05-05	23:49:08	afghanistan	3224	330	95	

In [275]:

```
cases = world11pm.query('country in ["USA","Italy","china","russia","germany"] & month
 == ["April", "May"]')
```

In [276]:

```
cases.head()
```

Out[276]:

	date	time	country	total_confirmed	total_confirmed_today	total_deaths	total_dea
1577	2020-05-01	23:49:16	china	83743	0	4634	
1578	2020-05-02	23:49:17	china	83744	1	4634	
1579	2020-05-03	23:49:02	china	83744	1	4634	
1580	2020-05-04	23:49:07	china	83750	6	4634	
1581	2020-05-05	23:49:08	china	83751	1	4634	

In [277]:

```
confirmed_cases=cases.total_confirmed.sum()
f"The total number of CONFIRMED CORONAVIRUS CASES for the month of APRIL and MAY in US
A,ITALY,CHINA,RUSSIA and GERMANY is {confirmed_cases}."
```

Out[277]:

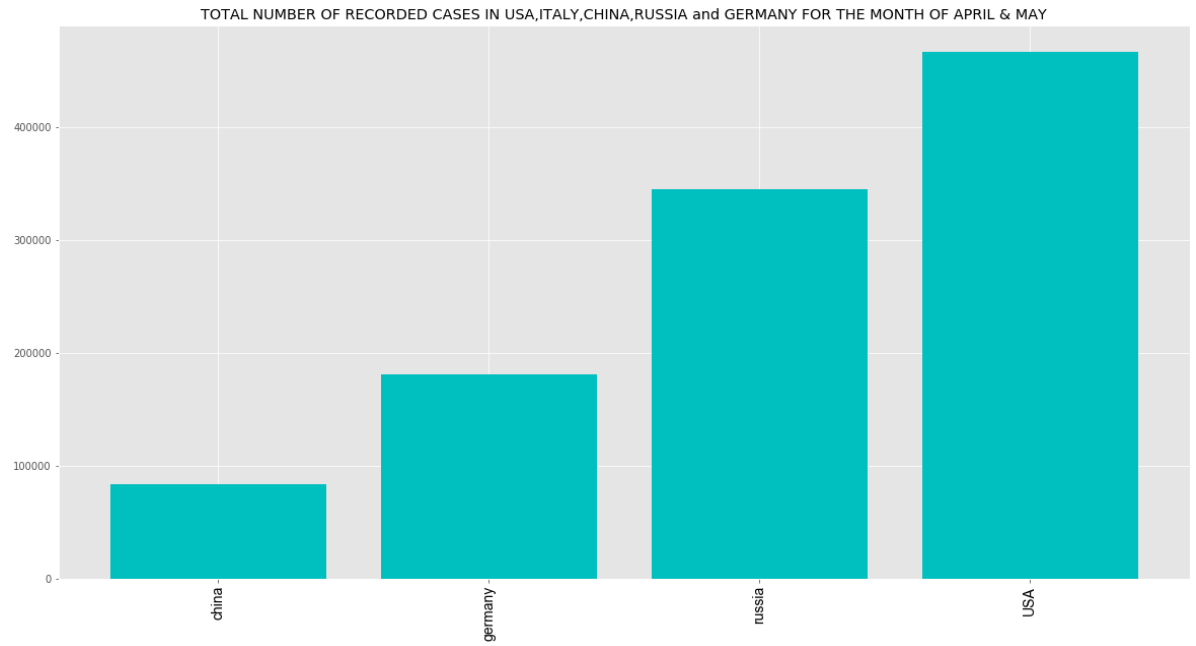
```
'The total number of CONFIRMED CORONAVIRUS CASES for the month of APRIL an
d MAY in USA,ITALY,CHINA,RUSSIA and GERMANY is 19944755.'
```

In [279]:

```
plt.figure(figsize=(20,10))
plt.bar(x=cases['country'], height = cases['total_confirmed'], color='c')
plt.xticks(rotation=90, family='Arial', color='black', size=14)
plt.title('TOTAL NUMBER OF RECORDED CASES IN USA,ITALY,CHINA,RUSSIA and GERMANY FOR THE
MONTH OF APRIL & MAY')
```

Out[279]:

```
Text(0.5, 1.0, 'TOTAL NUMBER OF RECORDED CASES IN USA,ITALY,CHINA,RUSSIA a
nd GERMANY FOR THE MONTH OF APRIL & MAY')
```



In [ ]: