COVID-19 DATASET In [1]: import pandas as pd import numpy as np import seaborn as sns from matplotlib import pyplot as plt **%matplotlib** inline data=pd.read_csv(r'C:\Users\mujjj\Downloads\4. covid_19_data.csv') In [2]: In [3]: Region Confirmed Deaths **Date State** Recovered Out[3]: 0 4/29/2020 NaN Afghanistan 1939 252 1 4/29/2020 NaN Albania 766 30 455 **2** 4/29/2020 NaN 3848 444 1702 Algeria 3 4/29/2020 42 423 Andorra 743 NaN 4 4/29/2020 NaN Angola 27 2 7 316 4/29/2020 Wyoming US 545 7 0 **317** 4/29/2020 Xinjiang Mainland China 76 73 4/29/2020 11 0 Yukon Canada Mainland China 181 4/29/2020 185 Yunnan **320** 4/29/2020 Zhejiang Mainland China 1268 1263 321 rows × 6 columns data.shape In [4]: (321, 6) Out[4]: data.count() In [5]: 321 Date Out[5]: State 140 321 Region 321 Confirmed Deaths 321 Recovered 321 dtype: int64 In [6]: data.columns Index(['Date', 'State', 'Region', 'Confirmed', 'Deaths', 'Recovered'], dtype='object') Out[6]: data.isnull().sum() In [7]: Date 0 Out[7]: 181 State 0 Region Confirmed 0 Deaths 0 Recovered 0 dtype: int64 sns.heatmap(data.isnull()) In [8]: plt.show <function matplotlib.pyplot.show(close=None, block=None)> 16 32 48 64 80 96 112 128 144 160 176 - 0.8 - 0.6 192 - 0.4 208 224 240 256 272 288 304 0.2 Date Region Confirmed Deaths Recovered Q:1) Show the number of confirmed deaths and recoverd ceases in each regin data.columns In [9]: Index(['Date', 'State', 'Region', 'Confirmed', 'Deaths', 'Recovered'], dtype='object') Out[9]: data.groupby('Region').sum().head(20) In [10]: Out[10]: Confirmed Deaths Recovered Region **Afghanistan** 1939 60 252 30 **Albania** 766 455 Algeria 3848 444 1702 **Andorra** 743 42 423 2 7 Angola 27 3 Antigua and Barbuda 24 11 1192 **Argentina** 4285 214 Armenia 1932 30 900 Australia 6752 91 5715 15402 580 **Austria** 12779 Azerbaijan 1766 23 1267 **Bahamas** 80 11 23 1455 **Bahrain** 2921 8 163 Bangladesh 7103 150 **Barbados** 80 7 39 13181 84 2072 **Belarus Belgium** 47859 7501 11283 2 **Belize** 18 9 **Benin** 64 1 33 **Bhutan** 0 5 data.groupby('Region')['Confirmed'].sum() In [11]: Region Out[11]: Afghanistan 1939 Albania 766 Algeria 3848 Andorra 743 Angola 27 West Bank and Gaza Western Sahara 6 6 Yemen Zambia 97 Zimbabwe 32 Name: Confirmed, Length: 187, dtype: int64 In [12]: data.columns Index(['Date', 'State', 'Region', 'Confirmed', 'Deaths', 'Recovered'], dtype='object') Out[12]: data.groupby('Region')['Deaths'].sum().sort_values(ascending=False).head(10) In [13]: Out[13]: US 60967 Italy 27682 26166 UK Spain 24275 France 24121 7501 Belgium Germany 6467 5957 Iran Brazil 5513 Netherlands 4727 Name: Deaths, dtype: int64 data.groupby('Region')['Confirmed', 'Recovered'].sum() In [15]: C:\Users\mujjj\AppData\Local\Temp\ipykernel_11504\3760683653.py:1: FutureWarning: Indexing with multiple keys (implicitly converted to a tuple of keys) will be deprecated, use a list instead. data.groupby('Region')['Confirmed', 'Recovered'].sum() Confirmed Recovered Out[15]: Region **Afghanistan** 1939 252 **Albania** 766 455 Algeria 3848 1702 Andorra 743 423 7 **Angola** 27 West Bank and Gaza 344 71 Western Sahara 6 5 Yemen 6 1 Zambia 97 54 5 **Zimbabwe** 32 187 rows × 2 columns Remove all records where confirmed cases is less then 10 In [16]: data.head() Region Confirmed Deaths Recovered Out[16]: Date State 4/29/2020 1939 252 NaN Afghanistan 60 **1** 4/29/2020 NaN Albania 766 30 455 4/29/2020 Algeria 3848 1702 NaN 444 743 42 423 4/29/2020 NaN Andorra **4** 4/29/2020 NaN Angola 27 7 data[data.Confirmed<10] State Region Confirmed Deaths Recovered Out[17]: **Date 18** 4/29/2020 5 NaN Bhutan 7 0 2 98 4/29/2020 NaN MS Zaandam 9 0 105 4/29/2020 NaN Mauritania 8 1 6 4/29/2020 0 126 NaN Papua New Guinea 8 0 4/29/2020 0 140 Sao Tome and Principe 8 4 NaN 4/29/2020 6 0 177 NaN Western Sahara 5 4/29/2020 178 NaN Yemen 6 0 1 4/29/2020 3 0 184 Anguilla UK 3 192 4/29/2020 Bonaire, Sint Eustatius and Saba Netherlands 5 0 0 4/29/2020 6 1 3 194 British Virgin Islands UK 4/29/2020 203 Diamond Princess cruise ship Canada 0 1 0 272 4/29/2020 **Northwest Territories** Canada 5 0 0 284 4/29/2020 Canada 0 0 20327 Recovered 4/29/2020 0 0 120720 285 Recovered US 4/29/2020 288 Saint Barthelemy France 6 0 6 4/29/2020 0 289 Saint Pierre and Miquelon France 1 0 305 4/29/2020 **Tibet** Mainland China 1 0 1 data[data.Confirmed<10] In [18]: State Out[18]: Date Region Confirmed Deaths Recovered **18** 4/29/2020 7 0 5 NaN Bhutan 4/29/2020 MS Zaandam NaN 9 2 0 4/29/2020 6 105 NaN Mauritania 8 1 4/29/2020 NaN Papua New Guinea 8 0 126 4/29/2020 Sao Tome and Principe 8 0 140 NaN 4 4/29/2020 NaN Western Sahara 6 0 5 177 4/29/2020 178 NaN Yemen 6 0 1 4/29/2020 Anguilla 3 0 3 184 4/29/2020 Bonaire, Sint Eustatius and Saba 5 0 0 192 Netherlands 194 4/29/2020 British Virgin Islands UK 6 1 3 4/29/2020 0 1 0 203 Diamond Princess cruise ship Canada 4/29/2020 272 **Northwest Territories** Canada 0 0 284 4/29/2020 Recovered 0 0 20327 Canada 285 4/29/2020 US 0 0 120720 Recovered 0 288 4/29/2020 Saint Barthelemy 6 6 France 289 4/29/2020 Saint Pierre and Miquelon 0 France **305** 4/29/2020 Tibet Mainland China 1 0 1 data=data[~(data.Confirmed<10)]</pre> In [19]: data In [20]: **Region Confirmed** Out[20]: **Date** State Deaths Recovered 0 4/29/2020 NaN Afghanistan 1939 60 252 4/29/2020 Albania 766 30 455 NaN 4/29/2020 NaN 3848 444 1702 Algeria **3** 4/29/2020 NaN 743 423 Andorra 4 4/29/2020 7 NaN Angola 27 **316** 4/29/2020 Wyoming US 7 0 545 Xinjiang Mainland China **317** 4/29/2020 73 0 **318** 4/29/2020 0 Yukon Canada 11 4/29/2020 Yunnan Mainland China 181 185 4/29/2020 1268 1263 Zhejiang Mainland China 304 rows × 6 columns in which Region maximum number of confirmed cases were recorded In [22]: data.groupby('Region').Confirmed.sum().sort_values(ascending=False) Out[22]: US 1039909 Spain 236899 Italv 203591 France 166536 UK 166432 Seychelles 11 Burundi 11 Suriname 10 Gambia 10 Holy See 10 Name: Confirmed, Length: 180, dtype: int64 in which Region minimum number of confirmed death cases were recorded data.columns In [23]: Index(['Date', 'State', 'Region', 'Confirmed', 'Deaths', 'Recovered'], dtype='object') Out[23]: data.groupby('Region').Deaths.sum().sort_values(ascending=True).head(12) In [26]: Region Out[26]: Cambodia 0 Seychelles 0 Saint Lucia 0 Central African Republic 0 Saint Kitts and Nevis 0 South Sudan 0 Rwanda Grenada Macau Madagascar Nepal Namibia Name: Deaths, dtype: int64 how many confirmed, deaths and Recovered cases were reported from nepal till 29 April 2020 data.head() In [27]: Region Confirmed **Deaths** Out[27]: State Date **0** 4/29/2020 NaN Afghanistan 1939 60 252 4/29/2020 Albania 766 455 4/29/2020 444 1702 NaN Algeria 3848 4/29/2020 423 NaN Andorra 743 2 7 4/29/2020 NaN 27 Angola data[data.Region=='Nepal'] In [28]: Date State Region Confirmed Deaths Recovered Out[28]: **115** 4/29/2020 NaN Nepal 57 16 data[data.Region=='US'].head(5) In [33]: State Region Confirmed Deaths Recovered Out[33]: **Date 181** 4/29/2020 US 6912 256 0 Alabama **182** 4/29/2020 US 355 9 0 Alaska 4/29/2020 US 7209 308 186 Arizona 0 187 4/29/2020 Arkansas US 3193 57 4/29/2020 California US 48747 1946 Sort the entire data wrt no of confirmed cases in ascending order data.sort_values(by = ['Confirmed'], ascending=True) In [35]: Out[35]: Date State Region Confirmed Deaths Recovered **156** 4/29/2020 10 8 NaN Suriname 4/29/2020 NaN Holy See 4/29/2020 NaN Gambia 10 8 4/29/2020 Yukon Canada 0 **217** 4/29/2020 Greenland 11 0 11 Denmark **57** 4/29/2020 France 24087 NaN 165093 48228 168 4/29/2020 NaN 165221 26097 27682 **80** 4/29/2020 NaN Italy 203591 71252 **153** 4/29/2020 NaN Spain 236899 24275 132929 **265** 4/29/2020 US 299691 23477 0 New York 304 rows × 6 columns Sort the entire data wrt no of Recovered cases in descending order data.columns In [38]: Index(['Date', 'State', 'Region', 'Confirmed', 'Deaths', 'Recovered'], dtype='object') Out[38]: data.sort_values(by=['Recovered'], ascending=False) In [40]: Region Confirmed Deaths Recovered **Date State** Out[40]: **153** 4/29/2020 NaN 236899 24275 132929 Spain **61** 4/29/2020 NaN Germany 161539 6467 120400 **76** 4/29/2020 NaN 93657 5957 73791 Iran 4/29/2020 NaN 203591 27682 71252 229 4/29/2020 68128 4512 63616 Hubei Mainland China 4/29/2020 US 4934 230 0 258 Nevada 257 4/29/2020 Nebraska US 3851 0 255 4/29/2020 US 451 0 Montana 16 254 4/29/2020 Missouri US 7660 338 0 274 4/29/2020 US 17303 937 0 Ohio 304 rows × 6 columns