**Exploratory Data Analysis of Covid**

**Step\_1 🡪 I have imported all these libraries for data handling**

import pandas as pd

import numpy as np

import matplotlib.pyplot as plt

**Step\_2🡪 Following command use for data read**

df= pd.read\_csv("forecasts\_custom\_prophet1.csv")

**Step\_3 🡪 I used this following command for checking the dimension of data, data types, no. of columns and no of rows in dataset.**

df.head(10)

df.shape

df.columns

df.describe()

df.dtypes

**Step\_4 🡪 after used “ df.isnull().sum()” by this code I described all null values but in our data there is no null values. But in our dataset there are multiple zero values so by following code I have printed all the 0 values for every column.**

list1 = ['ObservationDate','Province/State','Country/Region','Confirmed','Deaths','Recovered']

for i in list1:

print("Column name :-",i,"\n")

print(df[i].value\_counts().head(1)/df.shape[0]\*100)

print("\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*")

**Step\_5 🡪 then I saw that first three columns out of six have 46% null values so as per the rule I drop all those rows by using following code.**

for i in list1:

df1.drop(df1.index[df1[i] == '0'], inplace=True)

**Step\_6 🡪 then I check the shape of dataframe by using following code to check If all values are dropped or not**

df.shape

**Step\_7🡪 by this code dropped timestamp from date column that is unusable**

df1['ObservationDate'] = df1['ObservationDate'].apply(lambda row : row.replace(" 00:00",""))

**Step\_8🡪 in this step extracted month column for visualization of the data according to the month**

df1['Month']=pd.to\_datetime(df1['ObservationDate']).dt.month

df1.**head(10)**

**Step\_9🡪By this code grouped all the data according to month**

con\_grouped\_month = df1.groupby('Month').Confirmed.sum()

death\_grouped\_month = df1.groupby('Month').Deaths.sum()

rec\_grouped\_month = df1.groupby('Month').Recovered.sum()

**Step\_10🡪plotting bar graph by month for confirmed cases, deaths and recovery.**

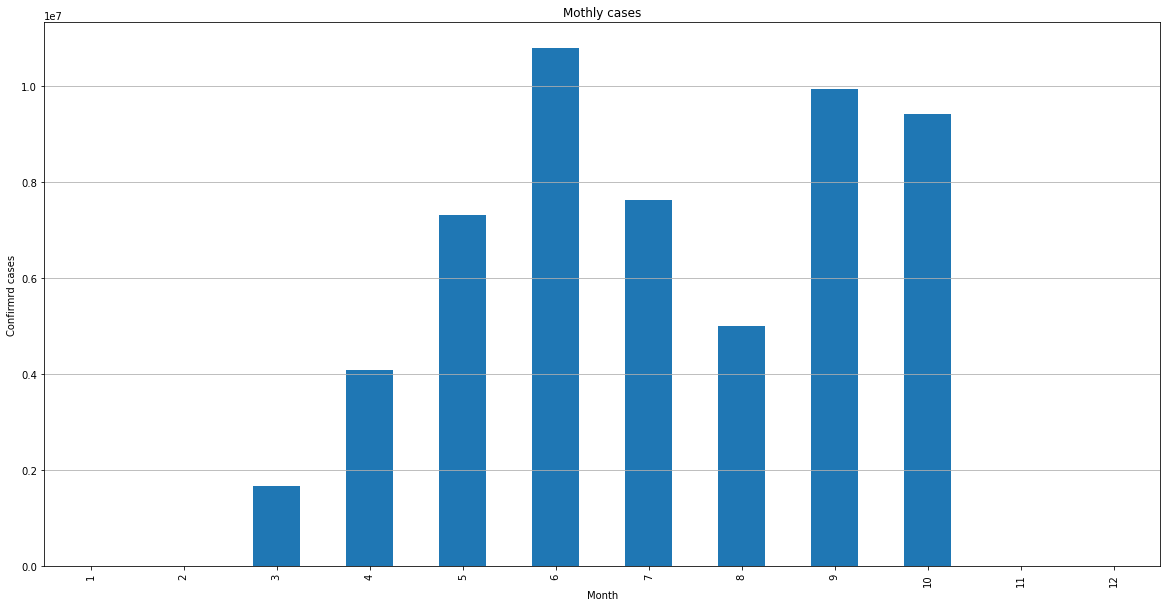
con\_grouped\_month.plot(kind='bar',figsize=(20,10))

plt.xlabel("Month")

plt.ylabel("Confirmrd cases")

plt.grid(axis='y')

plt.title("Mothly cases")



**Confirmed cases started to find from 3rd month of 2020, since then case kept rising.**

**In year 2020, highest number of confirmed cases found in 6th month: then cases started to decline but in 9th month there was sudden rise in cases.**

**After 10th month there is sudden fall started in case according to data**

**Step\_11🡪 here we plotted deaths by month.**

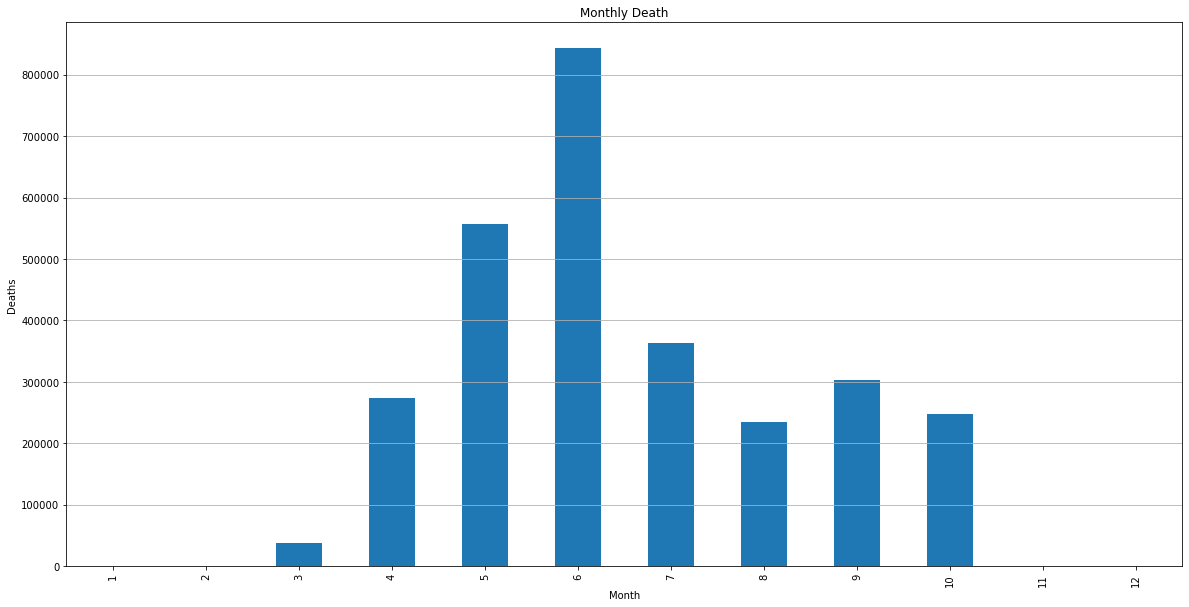
death\_grouped\_month.plot(kind='bar',figsize=(20,10))

plt.xlabel("Month")

plt.ylabel("Deaths")

plt.grid(axis='y')

plt.title("Monthly Deaths")



**For deaths from 3rd month started peoples dying. In 6th month more than 800000 peoples are died and this month is peak of this category. After 6th month graph felled down.**

**Step\_12🡪 here we plotted recovery by month.**

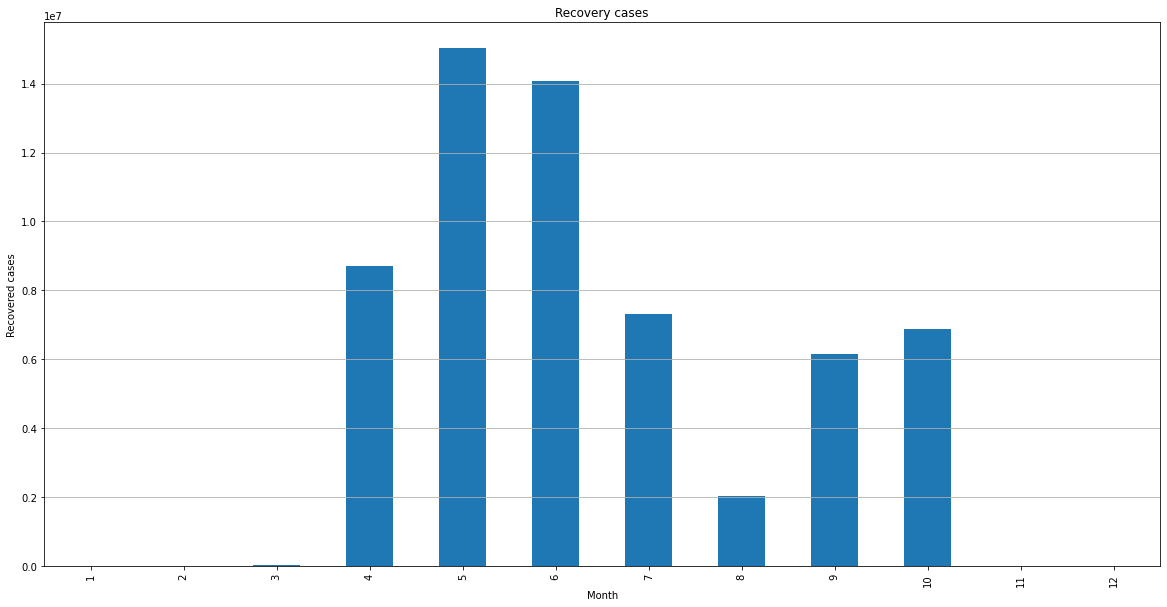
rec\_grouped\_month.plot(kind='bar',figsize=(20,10))

plt.xlabel("Month")

plt.ylabel("Recovered cases")

plt.grid(axis='y')

plt.title("Recovery cases")



**On this graph we can say that majority people recovered in 5th and 6th month of 2020**

**Step\_13🡪 then plotted line graph to see how the changes occurred in death, confirmed cases and recovery.**

fig, ax = plt.subplots()

ax.plot(con\_grouped\_month,label = 'Confirmed Cases')

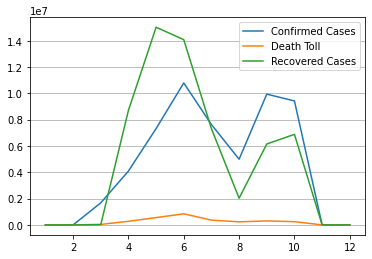
ax.plot(death\_grouped\_month,label = 'Death Toll' )

ax.plot(rec\_grouped\_month, label = 'Recovered Cases')

plt.grid(axis='y')

plt.legend()

plt.show()



Step\_14🡪 **then plotted graphs according to the country to see which country got more affected in confirmed cases.**

con\_grouped\_country = df1.groupby('Country/Region').Confirmed.sum()

death\_grouped\_country = df1.groupby('Country/Region').Deaths.sum()

rec\_grouped\_country = df1.groupby('Country/Region').Recovered.sum()

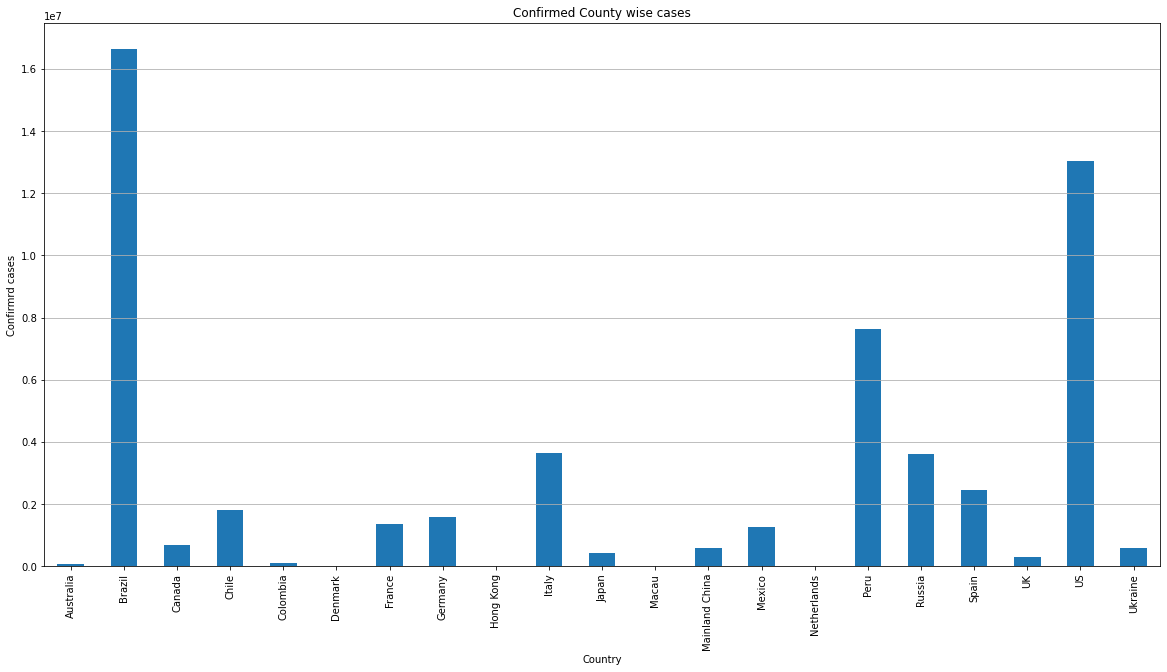
con\_grouped\_country.plot(kind='bar',figsize=(20,10))

plt.xlabel("Country")

plt.ylabel("Confirmrd cases")

plt.grid(axis='y')

plt.title("Country Wise Confirmed Cases")



**In brazil there are more cases recorded than remaining countries. Us is the 2nd highest country according to cases.**

**Step\_15🡪 here we plotted graphs according to the country to see which country had more deaths happened.**

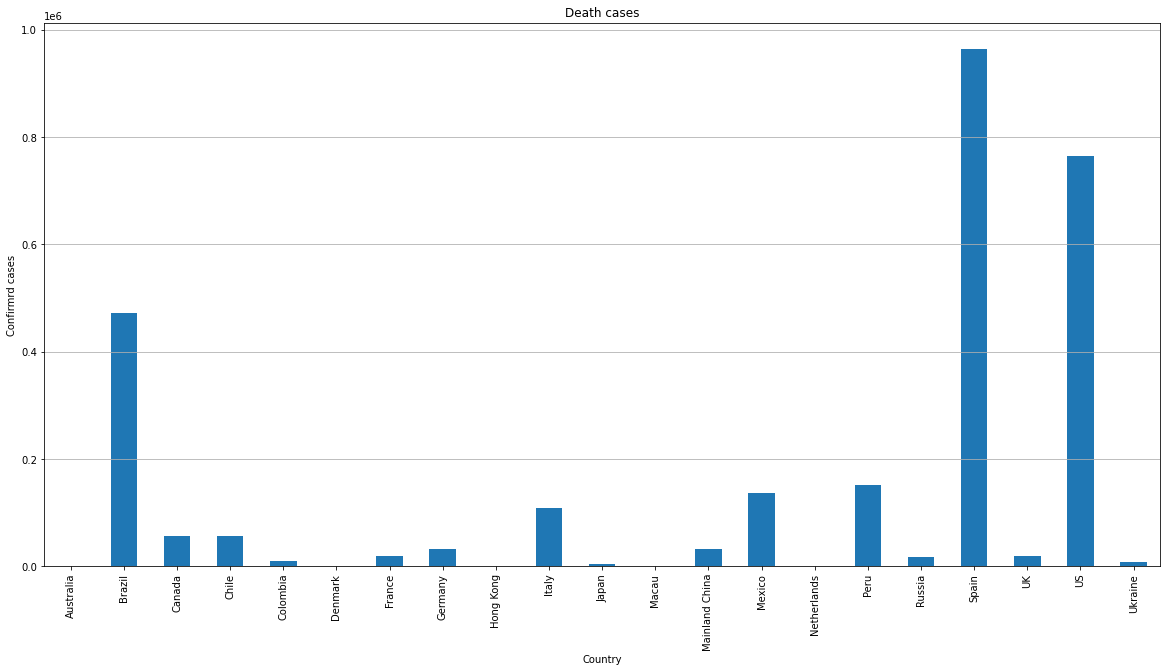
death\_grouped\_country.plot(kind='bar',figsize=(20,10))

plt.xlabel("Country")

plt.ylabel("Confirmrd Deaths")

plt.grid(axis='y')

plt.title("Death cases")



**According to this graph most people died in spain because of corona virous. In Us and brazil also more people die**

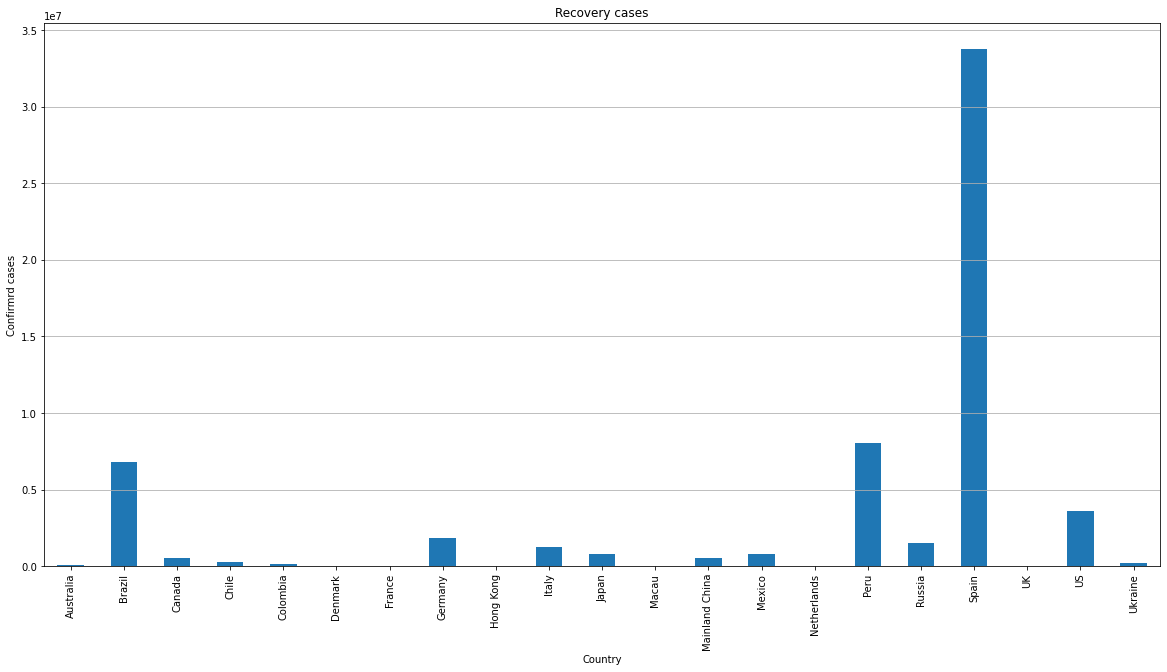
**Step\_16🡪 Recovery cases by country**

rec\_grouped\_country.plot(kind='bar',figsize=(20,10))

plt.xlabel("Country")

plt.ylabel("Confirmrd Recovery")

plt.grid(axis='y')

plt.title("Recovery cases")

**This graph shows us that most of the people are recovered in spain. Also we know that more people died.**

**Step\_17🡪 this step we performed for to see how corona virus is spread according to region wise of top 4 country.**

**USA**

fig, ax = plt.subplots()

ax.plot(con\_grouped\_US,label = 'Confirmed Cases')

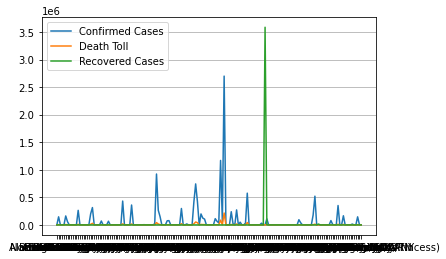
ax.plot(death\_grouped\_US,label = 'Death Toll' )

ax.plot(rec\_grouped\_US, label = 'Recovered Cases')

plt.grid(axis='y')

plt.legend()

plt.show()



**Step\_18🡪**

**Brazil**

con\_grouped\_Brazil = Brazil.groupby('Province/State').Confirmed.sum()

death\_grouped\_Brazil = Brazil.groupby('Province/State').Deaths.sum()

rec\_grouped\_Brazil = Brazil.groupby('Province/State').Recovered.sum()

fig, ax = plt.subplots()

ax.plot(con\_grouped\_Brazil,label = 'Confirmed Cases')

ax.plot(death\_grouped\_Brazil,label = 'Death Toll' )

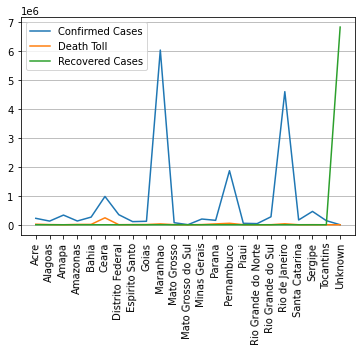
ax.plot(rec\_grouped\_Brazil, label = 'Recovered Cases')

plt.grid(axis='y')

plt.xticks(rotation=89)

plt.legend()

plt.show()



**Step\_19🡪**

**Peru**

con\_grouped\_Peru = Peru.groupby('Province/State').Confirmed.sum()

death\_grouped\_Peru = Peru.groupby('Province/State').Deaths.sum()

rec\_grouped\_Peru = Peru.groupby('Province/State').Recovered.sum()

fig, ax = plt.subplots()

ax.plot(con\_grouped\_Peru,label = 'Confirmed Cases')

ax.plot(death\_grouped\_Peru,label = 'Death Toll' )

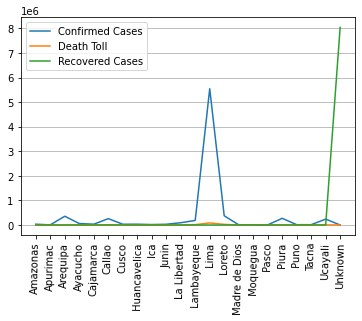
ax.plot(rec\_grouped\_Peru, label = 'Recovered Cases')

plt.grid(axis='y')

plt.xticks(rotation=89)

plt.legend()

plt.show()



**Step\_20🡪**

**Spain**

con\_grouped\_Spain = Spain.groupby('Province/State').Confirmed.sum()

death\_grouped\_Spain = Spain.groupby('Province/State').Deaths.sum()

rec\_grouped\_Spain = Spain.groupby('Province/State').Recovered.sum()

fig, ax = plt.subplots()

ax.plot(con\_grouped\_Spain,label = 'Confirmed Cases')

ax.plot(death\_grouped\_Spain,label = 'Death Toll' )

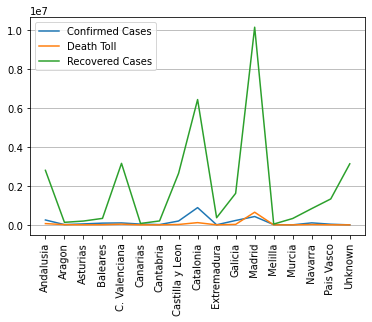
ax.plot(rec\_grouped\_Spain, label = 'Recovered Cases')

plt.grid(axis='y')

plt.xticks(rotation=89)

plt.legend()

plt.show()



This is spain region wise graph according to this we can say that in madrid there is highest recovered cases.

**Conclusion –**

1. Spain has highest number of recovered cases.
2. Brazil and Us has highest number of confirmed cases.
3. Spain and Us has highest number of deaths.
4. Confirmed cases started to find from 3rd month of 2020, since then case kept rising.
5. In year 2020, highest number of confirmed cases found in 6th month: then cases started to decline but in 9th month there was sudden rise in cases.
6. After 10th month there is sudden fall started in confirmed cases according to the data.
7. from 3rd month people started to dying. In 6th month more than 800000 peoples are died and this is peak. After 6th month graph felled down.
8. On the graph observation we saw that majority people recovered in 5th and 6th month of 2020