Covid-19 case analysis

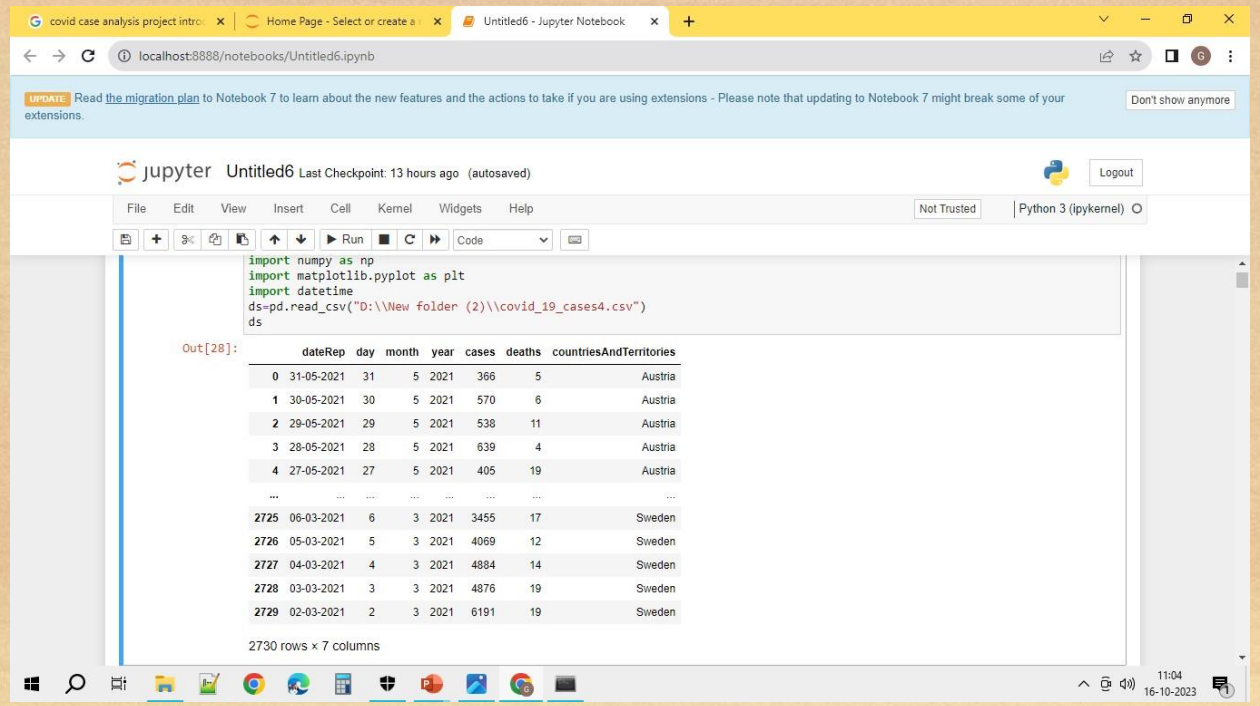
- NAME : C.PRIYA DHARSINI
- PHASE_3 : DEVLOPMENT PART 1

Covid - 19 case analysis

INTRODUCTION

Covid can be very contagious and spreads quickly. Over one million people have died from COVID-19 in the United States. COVID-19 most often causes respiratory symptoms that can feel much like a cold, the flu, or pneumonia. COVID-19 may attack more than your lungs and respiratory system.

Given Dataset



	dateRep	day	month	year	cases	deaths	countriesAndTerritories
0	31-05-2021	31	5	2021	366	5	Austria
1	30-05-2021	30	5	2021	570	6	Austria
2	29-05-2021	29	5	2021	538	11	Austria
3	28-05-2021	28	5	2021	639	4	Austria
4	27-05-2021	27	5	2021	405	19	Austria
...	...	...	...	...	...	...	...
2725	06-03-2021	6	3	2021	3455	17	Sweden
2726	05-03-2021	5	3	2021	4069	12	Sweden
2727	04-03-2021	4	3	2021	4884	14	Sweden
2728	03-03-2021	3	3	2021	4876	19	Sweden
2729	02-03-2021	2	3	2021	6191	19	Sweden

Necessary steps to follow

Import libraries:

start by importing the necessary libraries

```
import numpy as np
import matplotlib.pyplot as plt
import datetime
from sklearn.model_selection import train_test_split
from sklearn.preprocessing import standardscaler
```

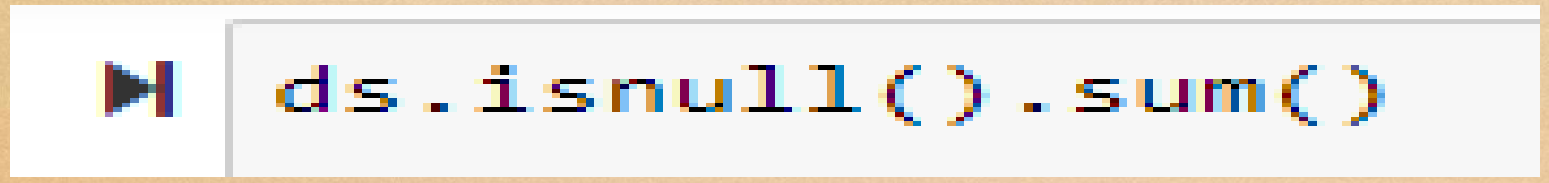
Load The Dataset

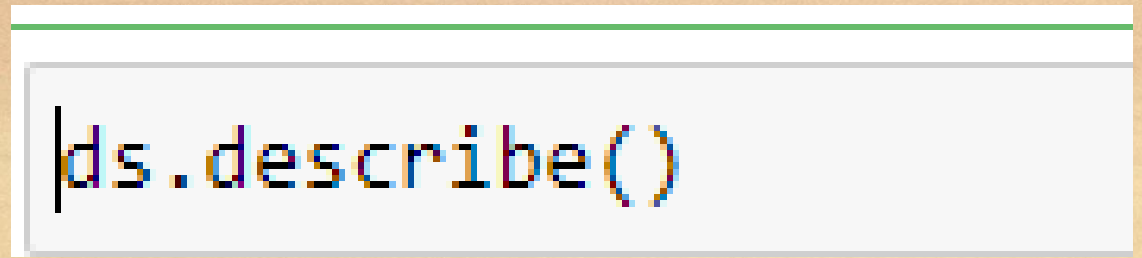
- Load your dataset into Pandas data frame.You can tupicaly find covid-19 case analysis datasets in CSV format,but you can adapt this code to other formats as needed.

```
ds=pd.read_csv("D:\\New folder (2)\\covid_19_cases4.csv")
ds
```

EXPLORATORY DATA ANALYSIS

Perform EDA to understand your data better. This include checking for missing value,exploring the data's statics,and visualizing it to identify pattens.





IMPORTANCE OF LOADING AND PROCESSING DATASET

- Loading and preprocessing the dataset is an important first step in building any machine learning model. However, it is especially important for house price prediction models, as house price datasets are often complex and noisy.

- By loading and preprocessing the dataset, we can ensure that the machine learning algorithm is able to learn from the data effectively and accurately.

LOADING THE DATASET

To Load a dataset in a jupyter notebook , you can use various python libraries,depending on the format of the dataset.

CSV files

Dataset is a csv format,you can use the ‘pandas’ library to read it.

```
ds=pd.read_csv("D:\\New folder (2)\\covid_19_cases4.csv")
ds
```

	dateRep	day	month	year	cases	deaths	countriesAndTerritories
0	31-05-2021	31	5	2021	366	5	Austria
1	30-05-2021	30	5	2021	570	6	Austria
2	29-05-2021	29	5	2021	538	11	Austria
3	28-05-2021	28	5	2021	639	4	Austria
4	27-05-2021	27	5	2021	405	19	Austria
...	...	...	...	...	...	...	...
2725	06-03-2021	6	3	2021	3455	17	Sweden
2726	05-03-2021	5	3	2021	4069	12	Sweden
2727	04-03-2021	4	3	2021	4884	14	Sweden
2728	03-03-2021	3	3	2021	4876	19	Sweden
2729	02-03-2021	2	3	2021	6191	19	Sweden

2730 rows × 7 columns

Handling missing values

- One way of handling missing values is the deletion of the rows or columns having null values. If any columns have more than half of the values as null then you can drop the entire column. In the same way, rows can also be dropped if having one or more columns values as null.

```
ds.isnull().sum()
```

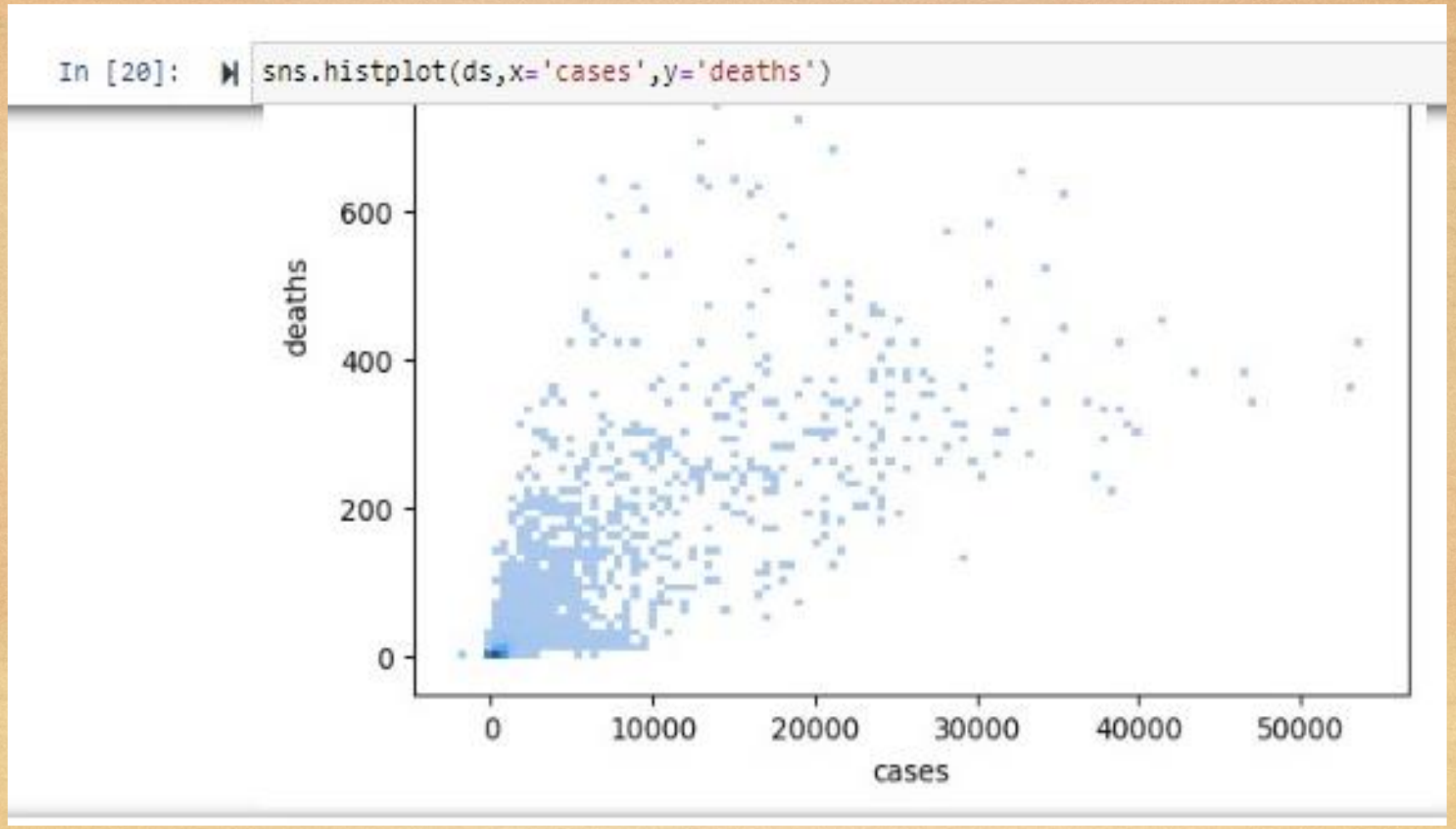
```
dateRep                    0
day                        0
month                      0
year                       0
cases                      0
deaths                     0
countriesAndTerritories    0
dtype: int64
```

Preprocessing the dataset

→Data preprocessing is the process of cleaning, Transforming and intragating data in order to make it Ready for analysis.

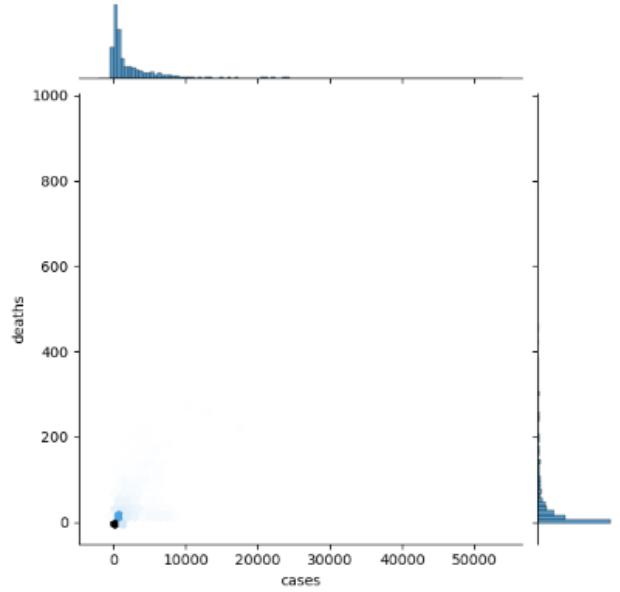
→this may involve removing errors and Inconsistencies, handling missing values, Transforming consistent format and scaling the data to a suitable range

Visualization and pre processing of data



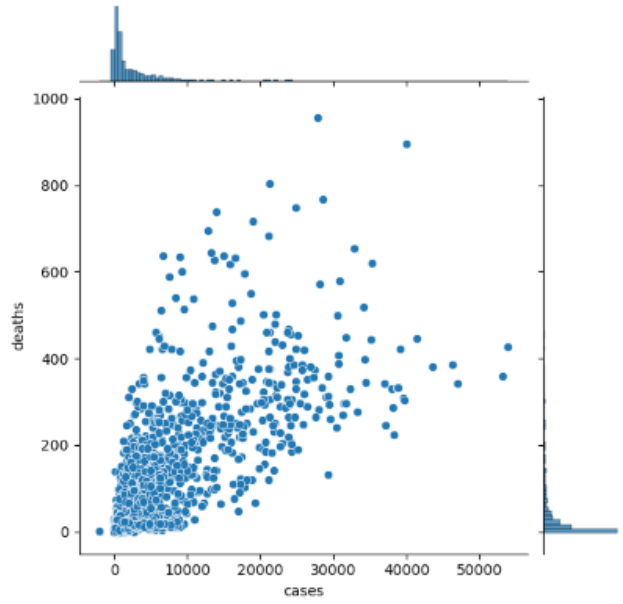
```
In [21]: sns.jointplot(ds,x='cases',y='deaths',kind='hex')
```

```
Out[21]: <seaborn.axisgrid.JointGrid at 0x25521ded690>
```



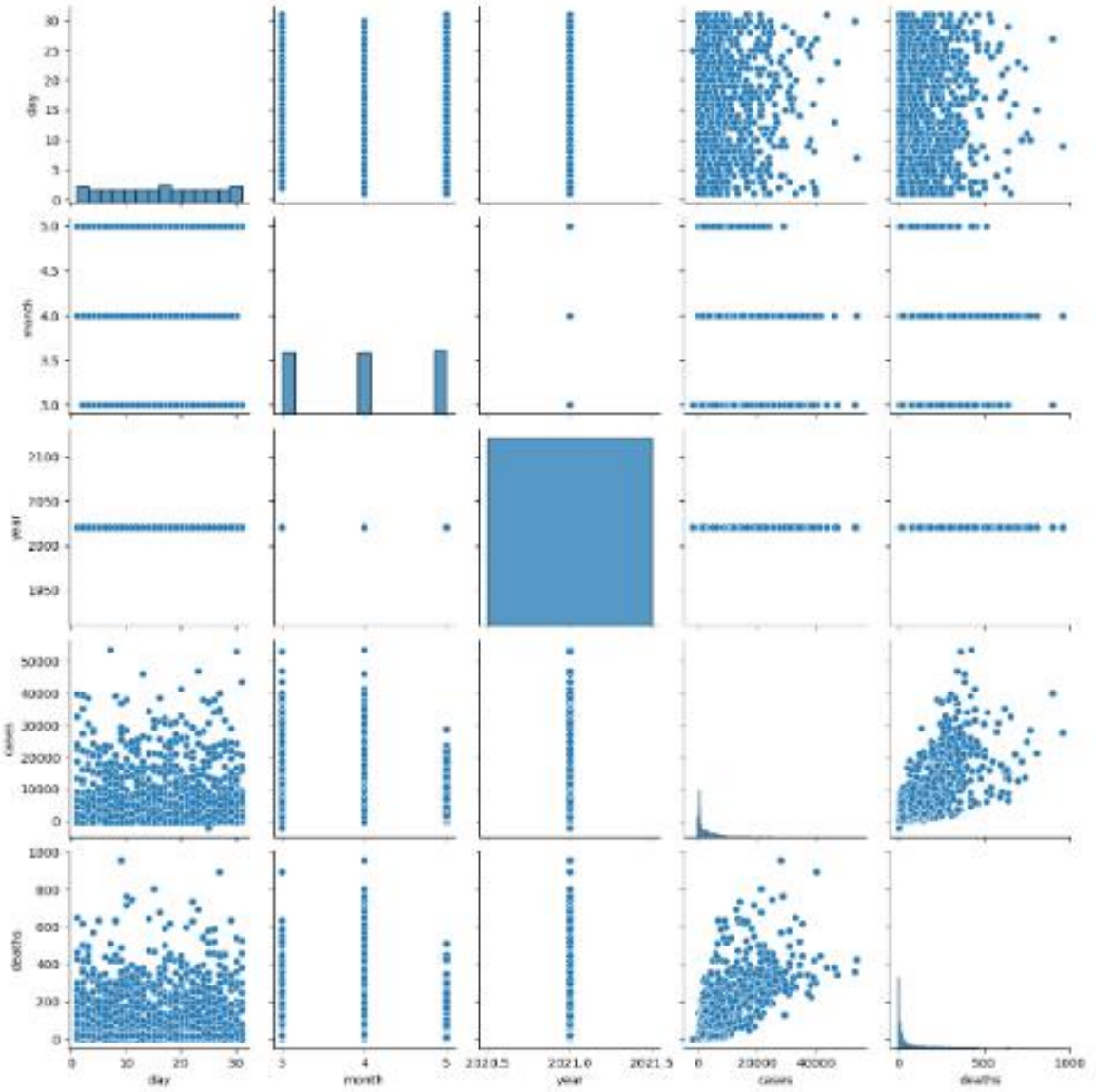

```
In [22]: sns.jointplot(ds,x='cases',y='deaths')
```

```
Out[22]: <seaborn.axisgrid.JointGrid at 0x2552218750>
```



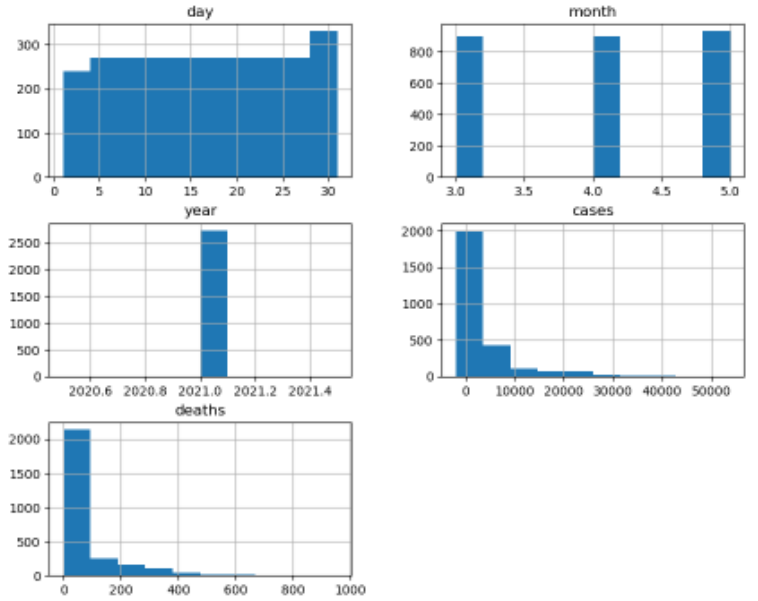
In [29]:
```
sns.pairplot(dataset)
```

Out[29]: <seaborn.axisgrid.PairGrid at 0x25522c03790>



```
In [30]: ds.hist(figsize=(10,8))
```

```
Out[30]: array([[<Axes: title={'center': 'day'}>,
                 <Axes: title={'center': 'month'}>],
                [<Axes: title={'center': 'year'}>,
                 <Axes: title={'center': 'cases'}>],
                [<Axes: title={'center': 'deaths'}>, <Axes: >]], dtype=object)
```



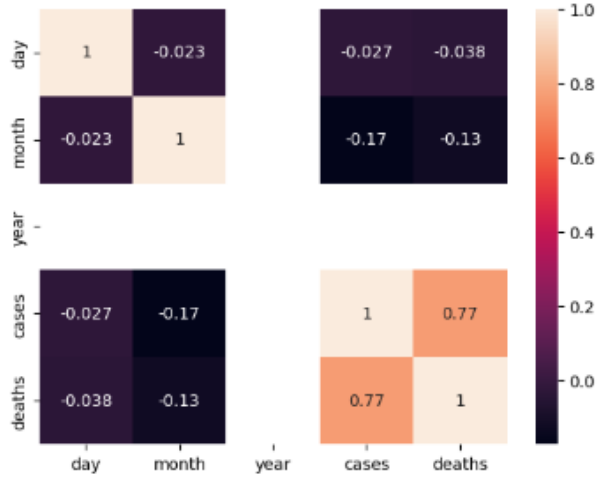

```
In [31]: ds.corr(numeric_only=True)
```

Out[31]:

	day	month	year	cases	deaths
day	1.000000	-0.022973	NaN	-0.026988	-0.038128
month	-0.022973	1.000000	NaN	-0.172412	-0.126515
year	NaN	NaN	NaN	NaN	NaN
cases	-0.026988	-0.172412	NaN	1.000000	0.766309
deaths	-0.038128	-0.126515	NaN	0.766309	1.000000

```
In [34]: sns.heatmap(ds.corr(numeric_only=True),annot=True)
```

```
Out[34]: <Axes: >
```



conclusion

Due to the time sensitivity of the pandemic, current scholarship predominantly put the discussion of covid 19 as a central subject in epidemiology.

Which this project , we hope to educate people on how easily viruses can be transmitted and infect an entire population and how social distancing and home isolation are key atrategies in halting the spread of the virus.