```
# introduction in pandas---->
# pandas is an open source library of python which is used for handle data m
```

```
import pandas as pd
```

```
# data structure of pandas---->
#(1)series
#(2)dataframe
```

```
# series----> it is one dimensional array and it return only value not colum
```

```
a=pd.Series([1,2,3,4,5])
a
```

```
type(a)
```

```
# dataframe---->  it is multi dimensional array and it return column with va
```

```
# task:: we have to create a dataframe of employees in which column are
# emp_id , emp_name,department,salary,working_hours
```

```
a={"emp_id":['101','102','103','104','105','106'],
    "emp_name":['divya','chiya','sourabh','anil','vartika','keshav'],
    "department":['abc','dfg','tyu','yio','sdf','hjkl'],
    "salary":['1000','6789','34546','34567','34567','34567'],
    "working hours":['1','2','4','6','8','8']}
```

```
import pandas as pd
df=pd.DataFrame(a)
df
```

```
# how can we export our dataframe into csv data ?
```

```
df.to_csv("emp_info.csv")
df                                          # TO_CSV MEAN EXPORT YOUR DATA
```

```
df.to_csv('new_emp_info.csv',index=False)            # INDEX=FALSE MEAN IT
```

```
# HOW TO IMPORT CSV DATA USING PANDAS
```

```

```

```
import pandas as pd
df=pd.read_csv("/content/covid_toy - covid_toy.csv")
df.head()
```

```
# how can we check total column in this dataframe
```

```
df.columns
```

```
# how can we check top 5 data from the dataframe.
```

```
df.head(2)
```

```
# how can we check bottom data from the dataframe.
```

```
df.tail(2)
```

```
df.sample(2)
```

```
# how can we check datatype of all column
```

```
df.dtypes
```

```
# how can we check statically view of dataframe.
```

```
df.describe()
```

```
# how can we check ovarall information in the dataframe
```

```
df.info()
```

```
df.head(2)
```

```
# how can we check missing value in a dataframe
```

```
df.isnull().sum()
```

```
df['fever'].mean()
```

```
df['fever']=df['fever'].fillna(df['fever'].mean())
```

```
df.isnull().sum()
```

```
# how can we check total sub-categories in a column.
```

```
df['gender'].value_counts()
```

```
df['has_covid'].value_counts()
```

```
df['gender']=df['gender'].map({'male':0,'female':1})
```

```
df['cough']=df['cough'].map({'mild':0,'strong':1})
df['cough']=df['cough'].map({'mild':0,'strong':1})
df['has_covid']=df['has_covid'].map({'yes':1,'no':0})
```

```
# data
# categorical(descride)----> we cannot divide futher
# numerical(continuous)-----> we can divide futher
```

```
# loc() and iloc()
# df.loc[row_range , column_name]  ----> start value include and last value
# df.iloc[row_range , column_range]   ----> start value include but last val
```

```
import pandas as pd
df.loc[5:10,['age','fever','cough']]
```

```
df.iloc[5:10 ,[0,2]]=55
```

```
 df.head(11)
```

```
df['gender'].value_counts()
```

```
df.head(2)
```

```
df['gender']=df['gender'].map({'male':0,'female':1})
df['cough']=df['cough'].map({"mild":0,"strong":1})
```

```
df['has_covid']=df['has_covid'].map({"yes":1,"NO":0})
```

```
df.sample(3)
```

```
df.drop(columns=['city'])
```

```
df.sort_index()      # sort the data
```

```
import pandas as pd
df = df.rename(columns={"cough":"updated_cough"})
```

```
import pandas as pd
df.head(3)
```

```
df.rename(columns={"updated_city":"city"})
```

```
df.iloc[2:5 ,0:4:2]
```

```
df['Age']=df['age'].add(10)
```

```
df.head(2)
```

```
df['Age']=df['age'].add(10)
```

```
df[['age','fever']]      # using this method ,we can filter data
```

```
df.describe()
```

```
df.loc[5:10,['gender','age']]
```

```
s=df.axes
s
```

```
p=df.dtypes
p
```

```
b=df.empty
b                        ### it return true if there any missing data in dat
```

```
df.isnull().sum      ### for check missing values.
```

```
dropna
```

```
i = df.ndim
i                         # number of axes(it is 2)
```

```
t=df.shape
t
df.shape[0]
df.shape [1]                    #shape=n(rows),n(column)
                                # total row
                                # total column
```

```
d=df.size                      # row count* column count
d
```

```
a=df.values                  #get a numpy array for df
a
```

```
df=df.copy()
df
```

```
p=df.sort_values(by='resume_score')
p
```

```
r=df.sort_index()
r
```

```
df['salary']=df['salary'].astype(int)
df
```

```
df.head(3)
```

```
t=df.add(4)
t
df['cgpa']=df['cgpa'].add(4)
df
```

```
s=df.count()
s
```

```
p=df.max()
p
```

```
q=df.min()
q
```

```
g=df.filter(items=['cgpa','placed'])
g
df[['cgpa','placed']]                # second option of filtering
```

```
s=df.filter(items=[5,6],axis=0)     #axis=0 means row wise operation
s
```

```
t=df.filter(like='5',axis=0)
t
```

```
p=df.rename(columns={'cgpa':'updated_cgpa','resume_sc
p
```

```
import pandas as pd
df['half']=df['cgpa'].where(df['cgpa']>7,other=0)
df.head(10)
```

```
df.groupby('has_covid').size()
```