

EXPT NO : 0
DATE : 01/02/23

STUDY ON NUMPY, SCIPY, STATSMODEL AND PANDAS PACKAGES

AIM:

To study about the Python packages required to work with data analytics

THEORY :

a. Pandas

Pandas is an open-source library that is made mainly for working with relational and labelled data. It provides various data structures and operations for manipulating data.

Functions in Pandas

S. No	Description	Example
1.	read_csv() : This helps to read a csv [comma-separated-values] file into a pandas dataframe. It can also read files separated by delimiter	data_e = pd.read_csv (r‘c:user_ds.csv’)
2.	head() : The head(n) is used to return the first n rows of a dataset. By default, the function returns 5 rows of the dataframe	data_e.head(10)
3.	describe() : It is used to generate descriptive statistics of data in a pandas data frame or services	data_e.describe()
4.	memory_usage() : It returns a pandas series having the memory usage of each column in a dataframe	data_e.memory_usage(deep = true)
5.	loc[] : It helps to access a group of rows and columns in a dataset	data_e.loc[0:4, [‘Name’, ‘Age’, ‘State’]]

b. Numpy

Numpy is a Python library used for working with arrays. It also has functions for working in domain of linear algebra, Fourier transformation and matrices.

Functions in Numpy

S. No	Description	Example
1.	numpy.reshape : This function allows us to change the dimension of the array without hampering the data value	res = arr.reshape(3, 2)
2.	numpy.concatenate() : This function is used to join two arrays of same size, either in a row-wise or column-wise way	res = numpy.concatenate((arr1, arr2), axis =1)
3.	numpy.char.add() : It concatenates the data value of two arrays, merge them and represent a new array as a result	res = numpy.char.add([‘python’], [‘simple’])
4.	numpy.median() : It calculates the median of an ordered array	med = numpy.median(X)
5.	numpy.average() : It returns the average of all data values of the passed array	avg = numpy.average(X)

c. Scipy

Scipy contains a variety of sub packages which help to solve the most common issue related to scientific computation. It is built on top of the numpy library.

Functions in Scipy

S. No	Description	Example
1.	special.logsumexp(x) : log sum exponential computes the log of sum exponential input element	np.log(np.sum (np.exp(a)))
2.	linalg.det : linear algebra of scipy is an implementation of BLAS and ALAS LAPACK libraries. It accepts 2D arrays and gives a @D array	ar = np.array([4, 5], [5, 2]) linalg.det(ar)
3.	linalg.eig() : The most common problem in linear algebra is Eigen value and Eigen value	eg_val, eg_vec = linalg.eig(ar)
4.	Scipy.sparse : It is used for creating sparse matrix using multiple data structures	from scipy.sparse import csr_matrix ar = np.array([0, 0], [1, 1]) print(csr_matrix(ar).data)

d. Stats Model

It provides classes and functions for the estimation of many different statistical models for conducting statistical test and statistical data exploration.

Functions in Statsmodel

S. No	Description	Example
1.	get_rdataset : It is used to download any dataset we want	data = sm.datasets. get_rdataset("Guerry", "Hist").data
2.	add_constant(X) : It is used to add a constant column to input dataset	X = sm.addconstant(X)
3.	OLS(y, x).fit() : It is a type of linear square method for estimating unknown parameters in linear regression	res = sm.OLS(y, x).fit()
4.	linear_rainbow() : The null hypothesis is the fit of the model using full sample. It is the same as using a central subset. Rainbow test has power against many different forms of non-linearity	print(sm.stats.linear_rainbow.__doc__)

Result:

Thus, a study on the Python packages used to work with data analysis has been made.

EXPT NO : 1	
DATE : 01/02/23	

READ DATA / DATASET

AIM :

To write python programs to read and display content from text file, csv file and web.

CODES AND OUTPUT :

a. Reading from Text File

Code:

```
lines = []
with open('input.txt', 'r') as f:
    lines = f.readlines()
count = 0
for line in lines:
    print(line)
    count += 1
print('Total lines: ', count)
```

Output :

```
Output exceeds the size limit. Open the full output data in a text editor
df.head() - returns the first n rows (default 5) of a DataFrame
df.tail() - returns the last n rows (default 5) of a DataFrame
df.shape - returns the shape of the DataFrame (rows, columns)
df.columns - returns the column names of the DataFrame
df.info() - returns a summary of the DataFrame, including the number of non-null values, data type, and memory usage
df.describe() - returns summary statistics of the numerical columns in the DataFrame
df.dtypes - returns the data types of the columns in the DataFrame
df.sort_values(by) - sorts the DataFrame by one or multiple columns
df.drop(columns) - drops one or multiple columns from the DataFrame
df.dropna() - drops the rows with missing values from the DataFrame
df.fillna() - fills the missing values in the DataFrame with a specified value or method
df.astype() - changes the data type of one or multiple columns
df.rename(columns) - renames one or multiple columns in the DataFrame
...
df.apply() - applies a function to each element or row/column of the DataFrame
df.to_csv() - saves the DataFrame to a CSV file
Total lines: 20
```

b. Reading from a CSV File :

Code :

```
import csv
count = 0
with open('../datasets/iris.csv', 'r') as f:
    reader = csv.reader(f)
    for row in reader:
        count += 1
        print(row)
        if(count == 5):
            break
```

Output:

```
['Id', 'SepalLengthCm', 'SepalWidthCm', 'PetalLengthCm', 'PetalWidthCm', 'Species']
['1', '5.1', '3.5', '1.4', '0.2', 'Iris-setosa']
['2', '4.9', '3.0', '1.4', '0.2', 'Iris-setosa']
['3', '4.7', '3.2', '1.3', '0.2', 'Iris-setosa']
['4', '4.6', '3.1', '1.5', '0.2', 'Iris-setosa']
```

c. Reading from Web :

Code :

```
import requests
from bs4 import BeautifulSoup
r = requests.get('https://www.python.org/')
soup = BeautifulSoup(r.text, 'html.parser')
print(soup.find_all('h1'))
```

Output :

```
[<h1 class="site-headline">
<a href="/"></a>
</h1>, <h1>Functions Defined</h1>, <h1>Compound Data Types</h1>, <h1>Intuitive Interpretation</h1>, <h1>All the Flow You'd Expect</h1>, <h1>Quick & Easy to Learn</h1>]
```

RESULT :

Thus, the Python programs to read and display content from text file, CSV file, and web have been written and verified.

EXPT NO : 2 a)	DESCRIPTIVE ANALYSIS
DATE : 01/02/23	

DESCRIPTIVE ANALYSIS

AIM :

To explore the various commands for performing descriptive data analysis on the health insurance Dataset.

CODES AND OUTPUTS :

Code:

```
import pandas as pd
import numpy as np
df = pd.read_csv('../datasets/insurance.csv')
df.head()
```

Output:

	age	sex	bmi	children	smoker	region	charges
0	19	female	27.900	0	yes	southwest	16884.92400
1	18	male	33.770	1	no	southeast	1725.55230
2	28	male	33.000	3	no	southeast	4449.46200
3	33	male	22.705	0	no	northwest	21984.47061
4	32	male	28.880	0	no	northwest	3866.85520

Code:

```
df.tail()
```

Output:

	age	sex	bmi	children	smoker	region	charges
1333	50	male	30.97	3	no	northwest	10600.5483
1334	18	female	31.92	0	no	northeast	2205.9808
1335	18	female	36.85	0	no	southeast	1629.8335
1336	21	female	25.80	0	no	southwest	2007.9450
1337	61	female	29.07	0	yes	northwest	29141.3603

Code:

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

Output:

	age	sex	bmi	children	smoker	region	charges
0	19	female	27.90	0	yes	southwest	16884.9240
1	18	male	33.77	1	no	southeast	1725.5523

Code:

```
df.shape
```

Output:

```
(1338, 7)
```

Code:

```
df.info()
```

Output:

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1338 entries, 0 to 1337
Data columns (total 7 columns):
 #   Column      Non-Null Count  Dtype  
 --- 
 0   age         1338 non-null   int64  
 1   sex         1338 non-null   object  
 2   bmi         1338 non-null   float64 
 3   children    1338 non-null   int64  
 4   smoker      1338 non-null   object  
 5   region      1338 non-null   object  
 6   charges     1338 non-null   float64 
dtypes: float64(2), int64(2), object(3)
memory usage: 73.3+ KB
```

Code:

```
df.describe()
```

Output:

	age	bmi	children	charges
count	1338.000000	1338.000000	1338.000000	1338.000000
mean	39.207025	30.663397	1.094918	13270.422265
std	14.049960	6.098187	1.205493	12110.011237
min	18.000000	15.960000	0.000000	1121.873900
25%	27.000000	26.296250	0.000000	4740.287150
50%	39.000000	30.400000	1.000000	9382.033000
75%	51.000000	34.693750	2.000000	16639.912515
max	64.000000	53.130000	5.000000	63770.428010

Code:

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

Output:

```
age      0
sex      0
bmi      0
children 0
smoker   0
region   0
charges  0
dtype: int64
```

Code:

```
df.value_counts('children')
```

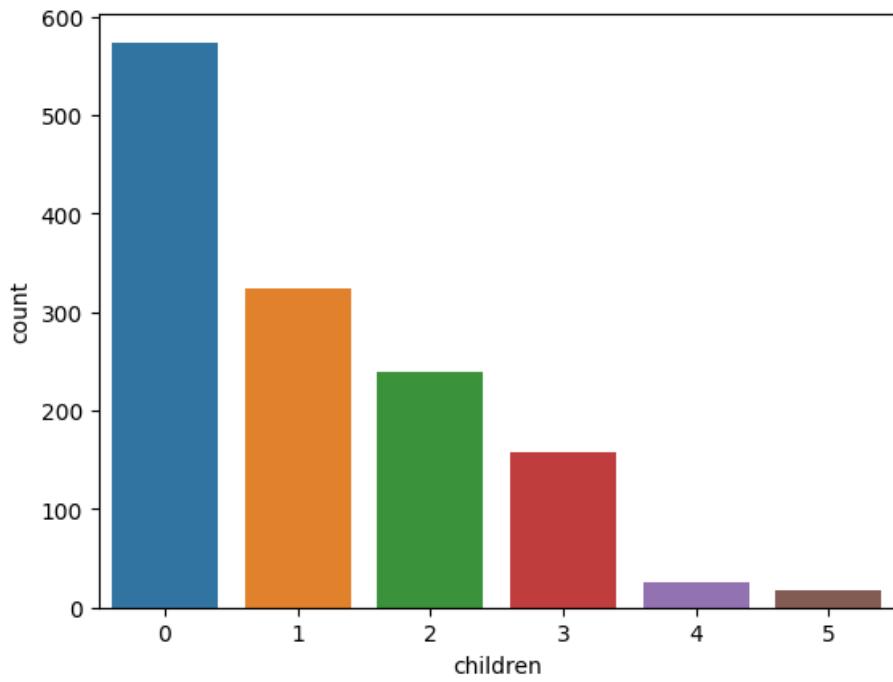
Output:

```
children
0    574
1    324
2    240
3    157
4     25
5     18
dtype: int64
```

Code:

```
import seaborn as sns
import matplotlib.pyplot as plt
sns.countplot(x='children', data=df)
plt.show()
```

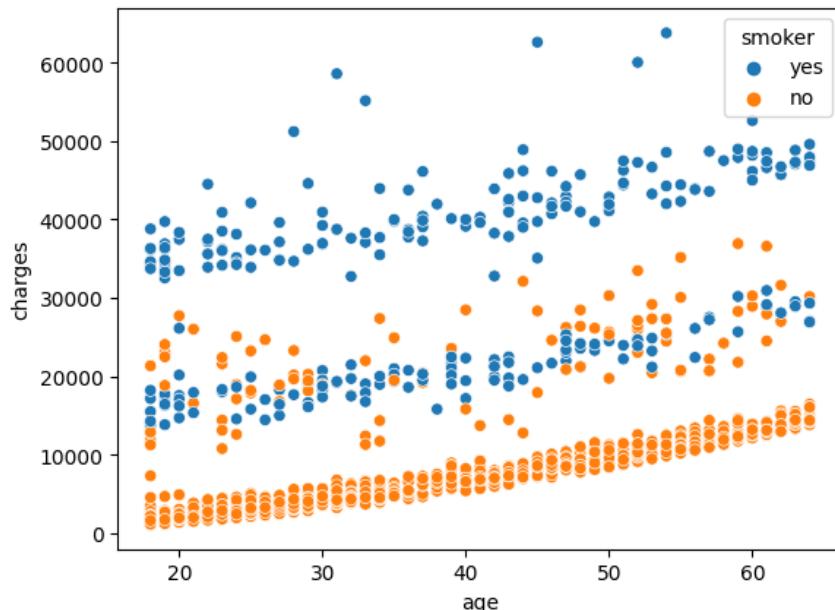
Output:



Code:

```
sns.scatterplot(x='age', y='charges', hue='smoker', data=df)  
plt.show()
```

Output:

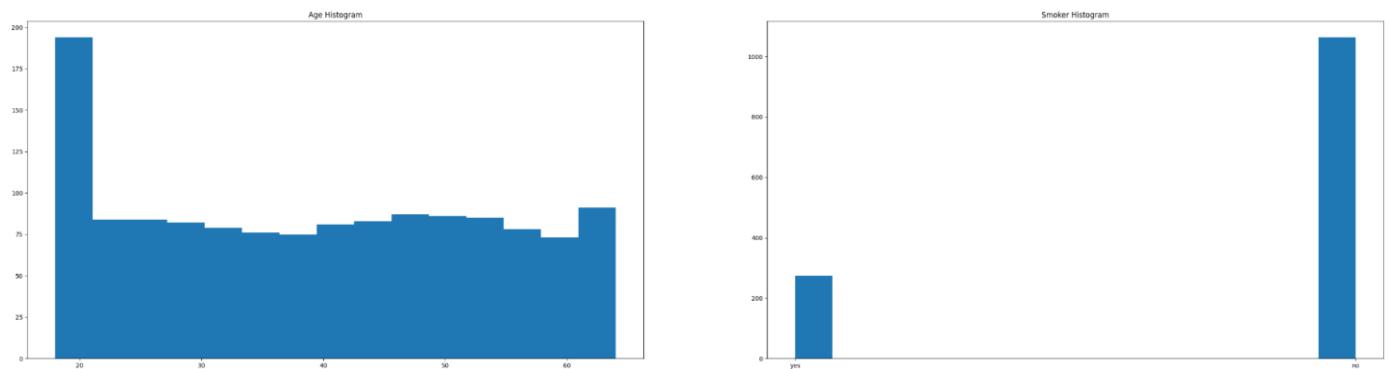


Code:

```
fig, axes = plt.subplots(1, 2, figsize=(40,10))  
axes[0].set_title("Age Histogram")  
axes[0].hist(df['age'], bins=15)  
axes[1].set_title("Smoker Histogram")  
axes[1].hist(df['smoker'], bins=15)
```

Output:

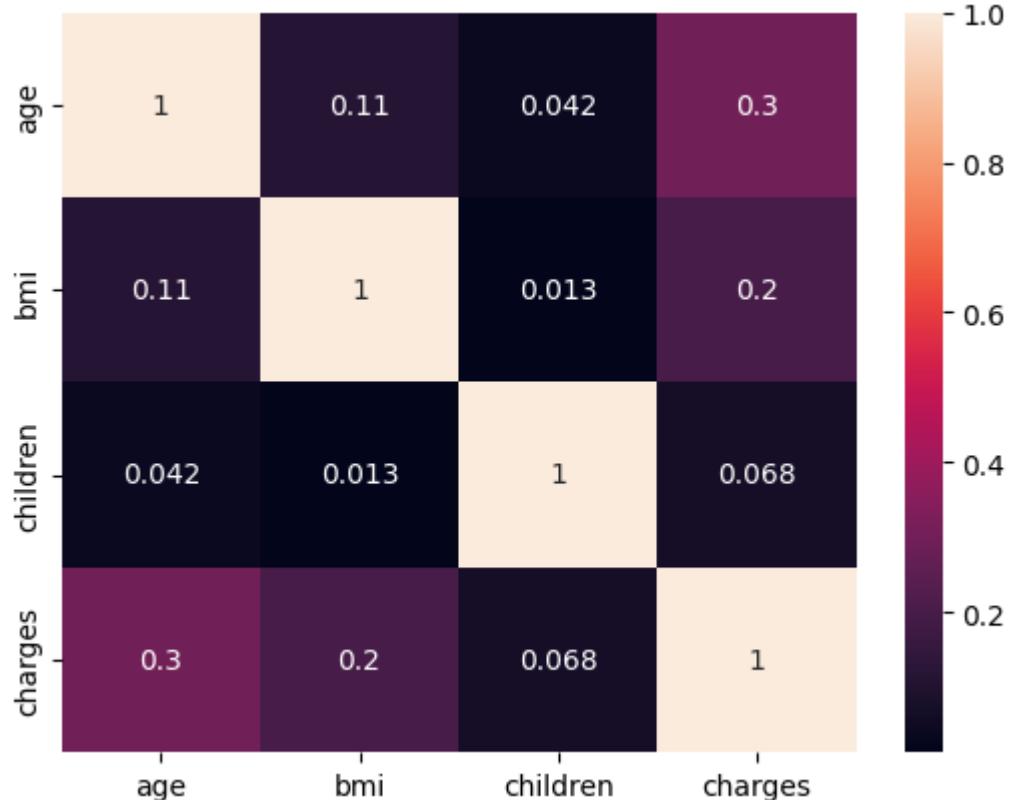
```
(array([ 274.,     0.,     0.,     0.,     0.,     0.,     0.,     0.,     0.,
         0.,     0.,     0.,     0.,     0., 1064.]),
 array([0.          , 0.06666667, 0.13333333, 0.2          , 0.26666667,
        0.33333333, 0.4          , 0.46666667, 0.53333333, 0.6          ,
        0.66666667, 0.73333333, 0.8          , 0.86666667, 0.93333333,
       1.          ]),
<BarContainer object of 15 artists>)
```



Code:

```
sns.heatmap(df.corr(method='pearson'), annot = True);
plt.show()
```

Output:



Code:

```
out_arr = np.partition(df.columns, 3)
```

```
out_arr
```

Output:

```
array(['age', 'bmi', 'charges', 'children', 'region', 'sex', 'smoker'],  
      dtype=object)
```

RESULT :

Thus, the various commands for performing descriptive data analysis on the given dataset have been executed and verified.

EXPT NO : 2 b)	
DATE : 08/02/23	

UNIVARIATE ANALYSIS

AIM :

To perform Univariate Analysis on the Credit Card Dataset.

CODES AND OUTPUTS:

Code:

```
import pandas as pd
import numpy as np
import seaborn as sns
import matplotlib.pyplot as plt
df = pd.read_csv('../datasets/insurance.csv')
df.head()
```

Output:

	age	sex	bmi	children	smoker	region	charges
0	19	female	27.900	0	yes	southwest	16884.92400
1	18	male	33.770	1	no	southeast	1725.55230
2	28	male	33.000	3	no	southeast	4449.46200
3	33	male	22.705	0	no	northwest	21984.47061
4	32	male	28.880	0	no	northwest	3866.85520

Code:

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

Output:

39.20702541106129

Code:

```
df['charges'].median()
```

Output:

9382.033

Code:

```
df['bmi'].mode()
```

Output:

```
0    32.3  
Name: bmi, dtype: float64
```

Code:

```
df['bmi'].var()
```

Output:

```
37.18788360977323
```

Code:

```
df['bmi'].std()
```

Output:

```
6.098186911679014
```

Code:

```
# univariate analysis of age column  
df['age'].describe()
```

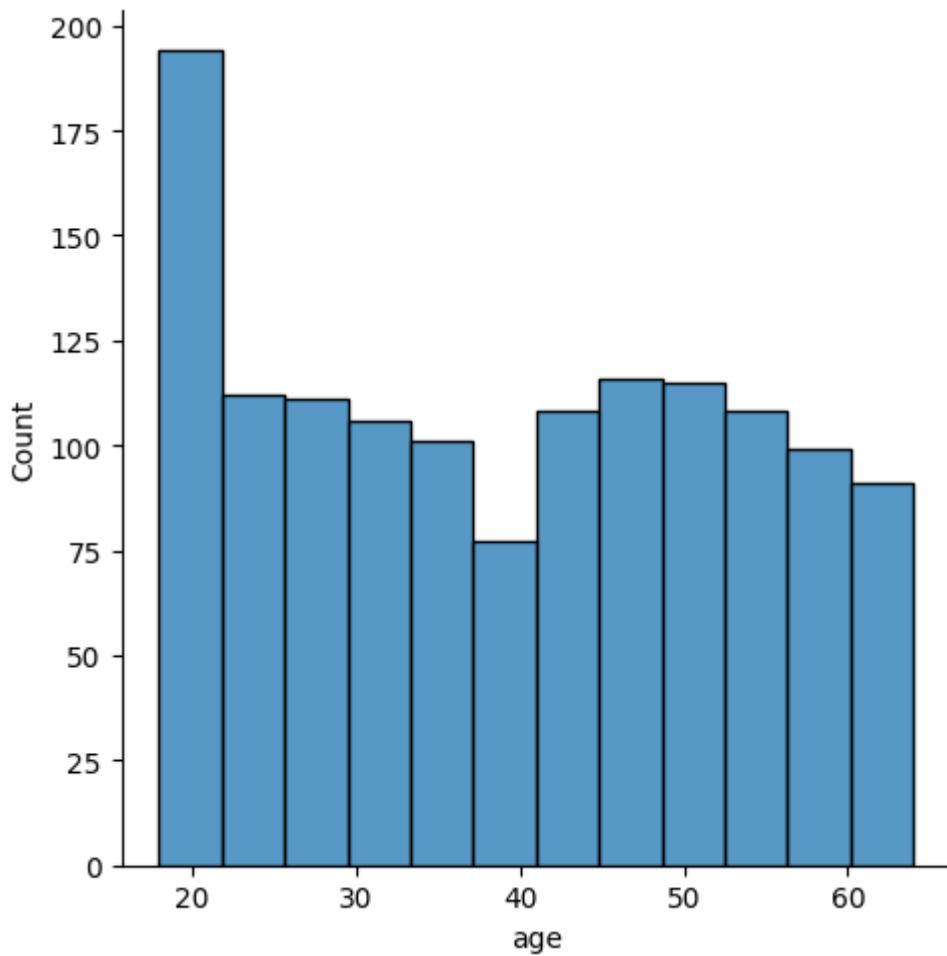
Output:

```
count    1338.000000  
mean     39.207025  
std      14.049960  
min      18.000000  
25%     27.000000  
50%     39.000000  
75%     51.000000  
max     64.000000  
Name: age, dtype: float64
```

Code:

```
sns.displot(df['age'])
```

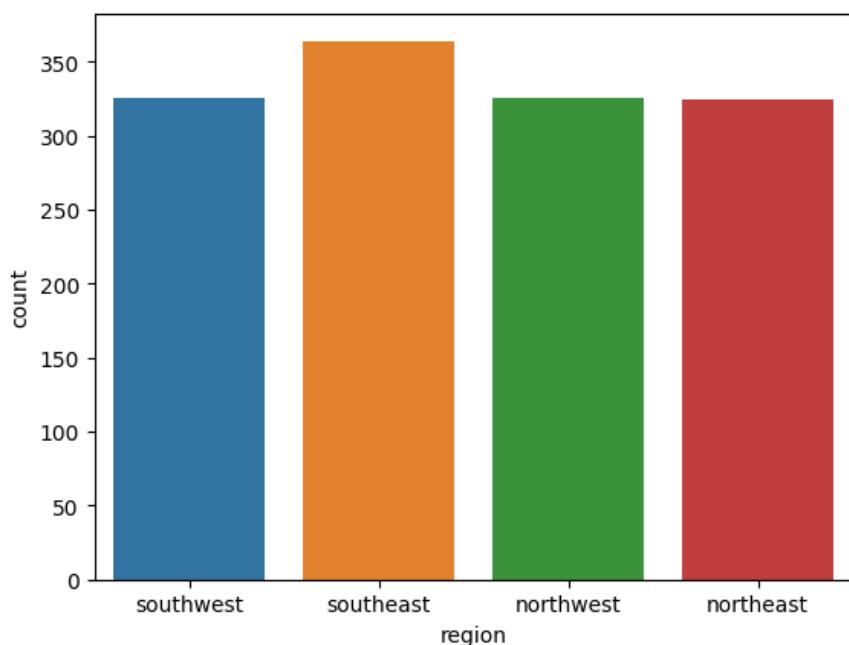
Output:



Code:

```
sns.countplot(data=df, x='region')
```

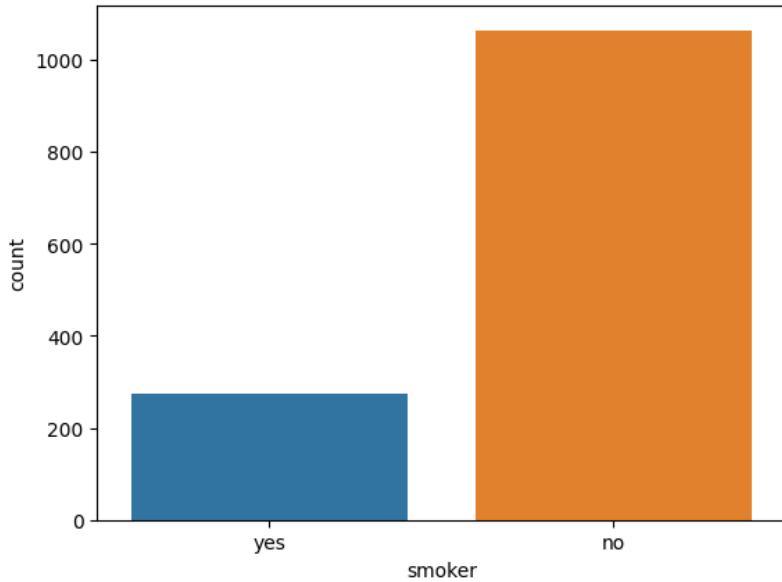
Output:



Code:

```
sns.countplot(data=df, x='smoker')
```

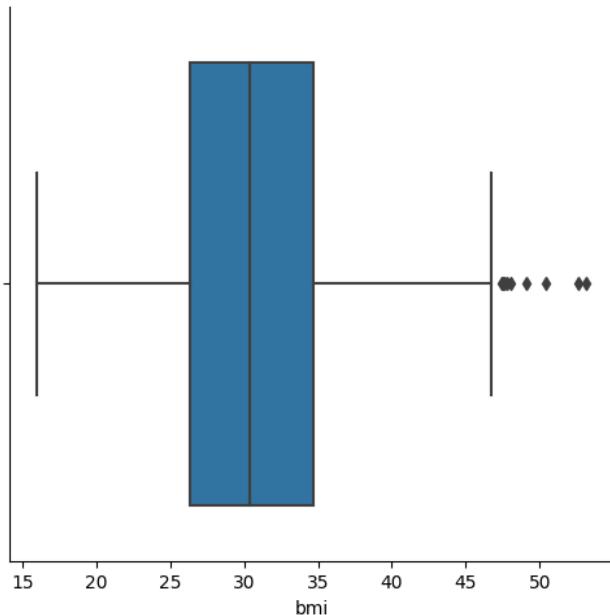
Output:



Code:

```
sns.catplot(x='bmi', kind='box', data=df)
```

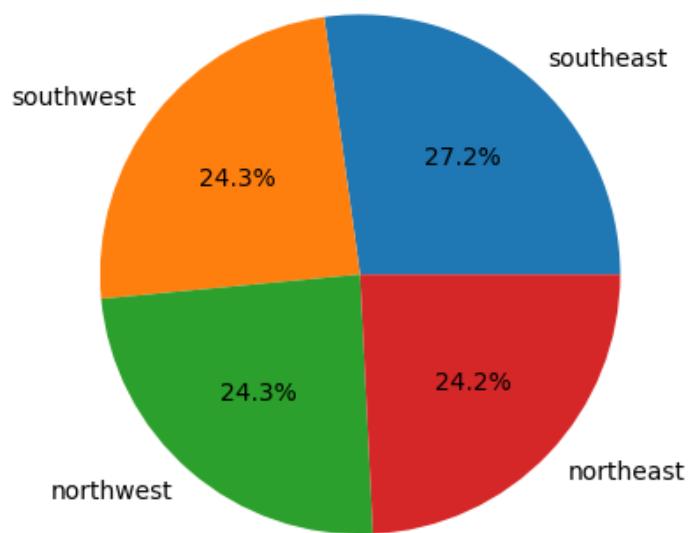
Output:



Code:

```
x = df['region'].value_counts()  
plt.pie(x.values, labels=x.index, autopct='%1.1f%%')  
plt.show()
```

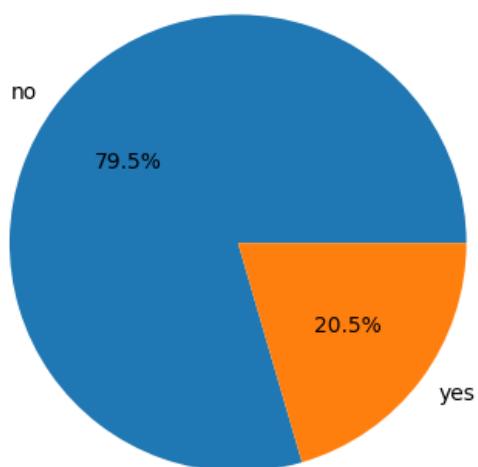
Output:



Code:

```
x2 = df['smoker'].value_counts()  
plt.pie(x2.values,labels=x2.index,autopct='%1.1f%%')  
plt.show()
```

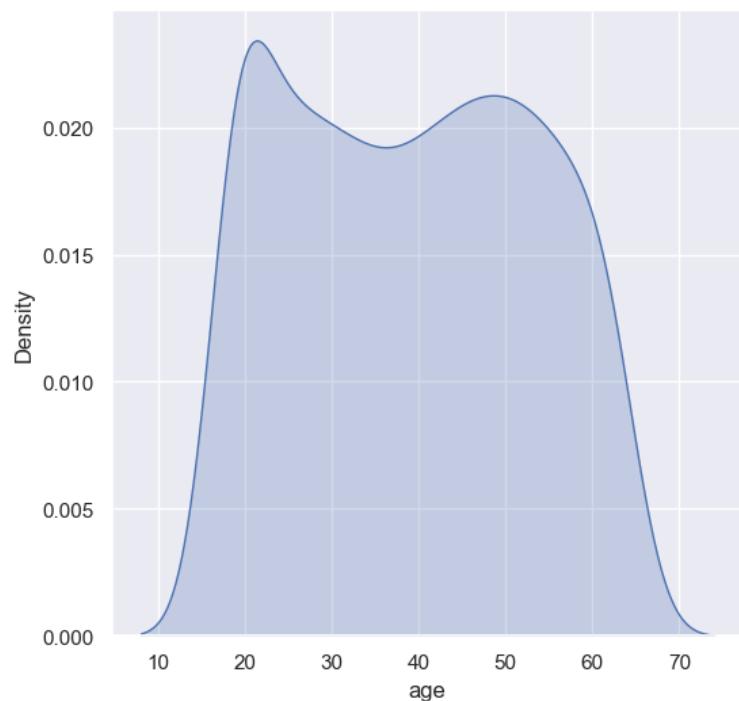
Output:



Code:

```
sns.set(rc={'figure.figsize':(6,6)})  
sns.kdeplot(df['age'],fill='True')
```

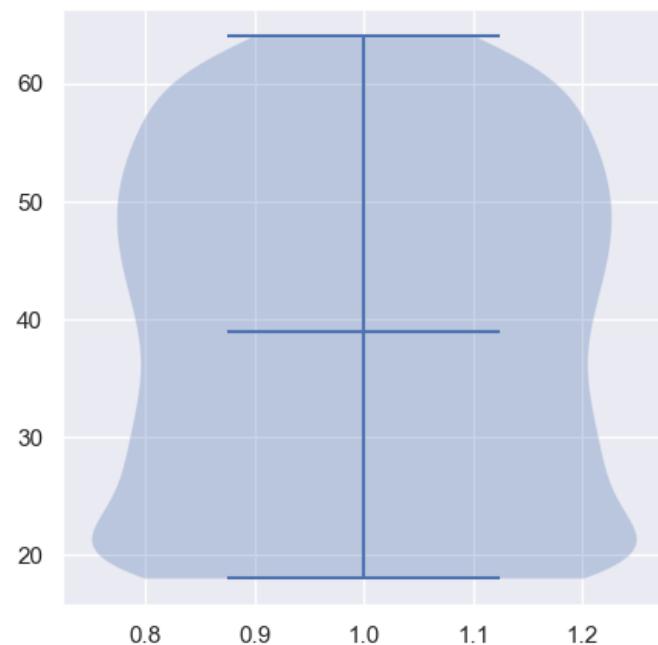
Output:



Code:

```
plt.figure(figsize=(5,5))
plt.violinplot(df['age'], showmedians=True)
```

Output:



RESULT :

Hence Univariate Analysis for the given Dataset was successfully performed .

EXPT NO : 2 c)	
DATE : 08/02/23	

BIVARIATE ANALYSIS

AIM :

To perform Bivariate Analysis on the health insurance Dataset.

CODES AND OUTPUTS:

Code:

```
import numpy as np
import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt

df2 = pd.read_csv('../datasets/insurance.csv')
df2.head()
```

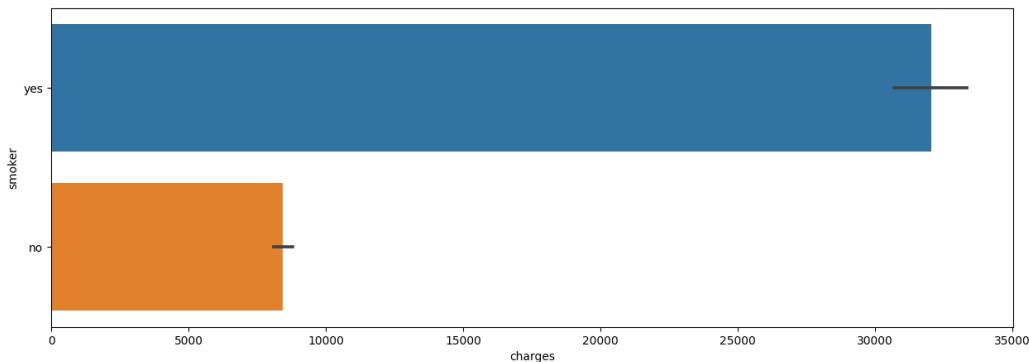
Output:

	age	sex	bmi	children	smoker	region	charges
0	19	female	27.900	0	yes	southwest	16884.92400
1	18	male	33.770	1	no	southeast	1725.55230
2	28	male	33.000	3	no	southeast	4449.46200
3	33	male	22.705	0	no	northwest	21984.47061
4	32	male	28.880	0	no	northwest	3866.85520

Code:

```
plt.figure(figsize=(15,5))
sns.barplot(x=df2['charges'],y=df2['smoker'])
plt.show()
```

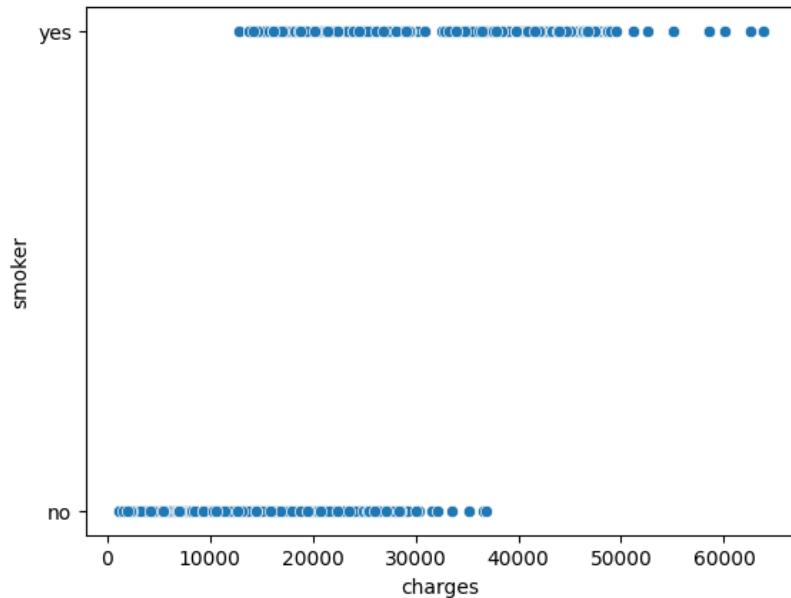
Output:



Code:

```
sns.scatterplot(x=df2['charges'],y=df2['smoker'])
```

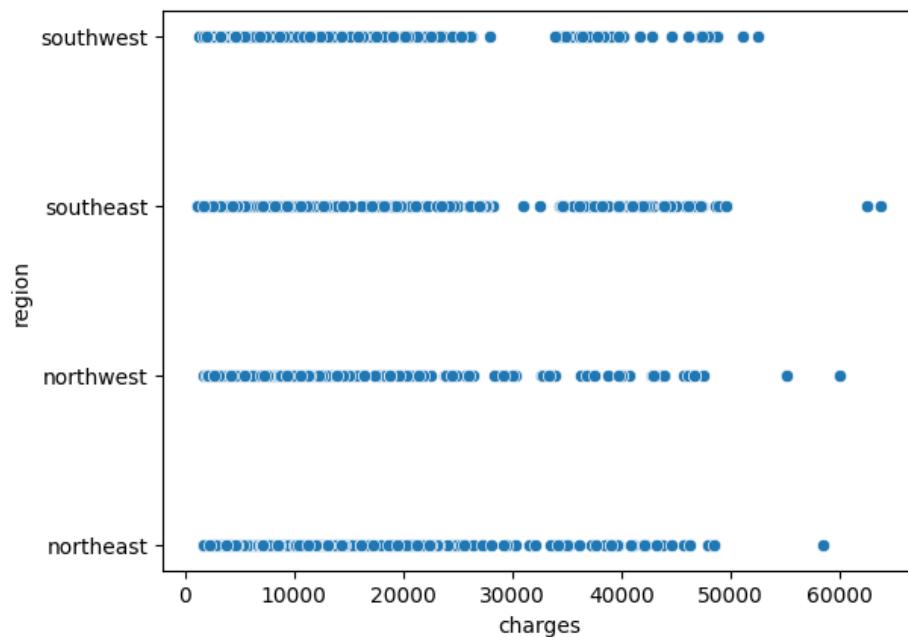
Output:



Code:

```
sns.scatterplot(x=df2['charges'],y=df2['region'])
```

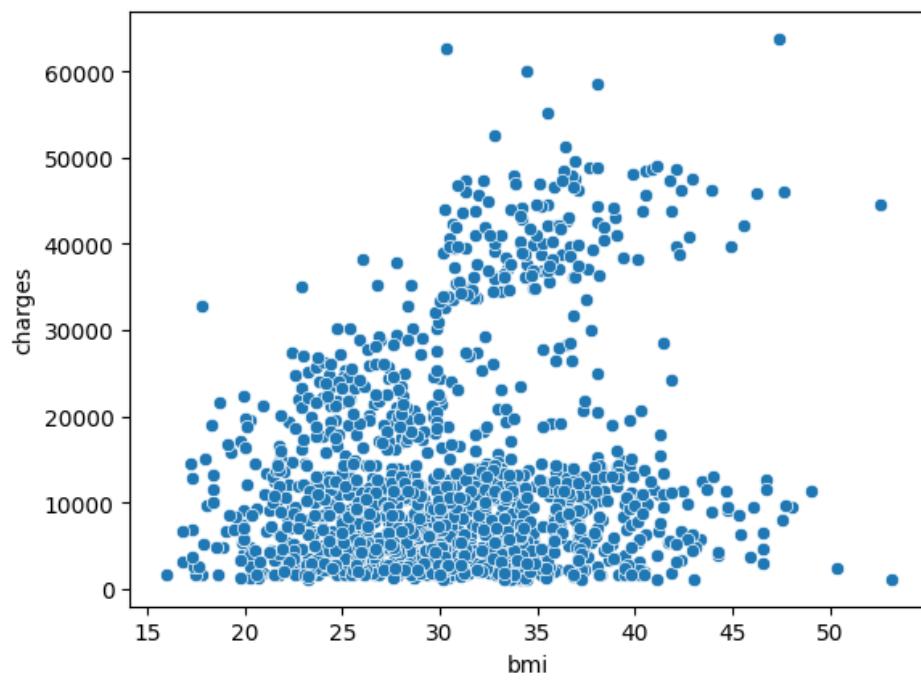
Output:



Code:

```
sns.scatterplot(x=df2['bmi'],y=df2['charges'])
```

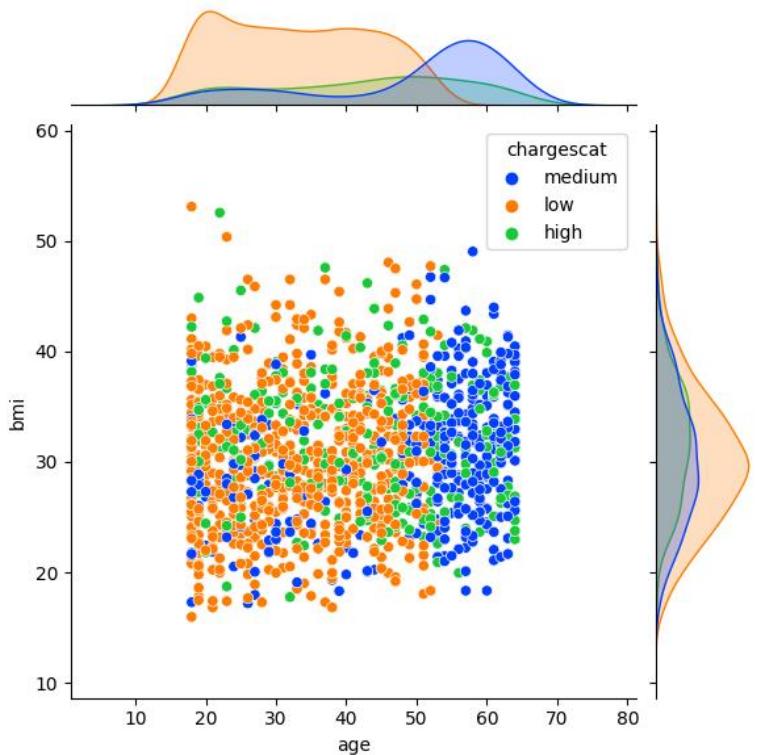
Output:



Code:

```
# create a new column which categorizes the charges
df2['chargescat']= df2['charges'].apply(lambda x: 'low' if x<10000 else 'medium' if
x<20000 else 'high')
sns.jointplot(x=df2['age'],y=df2['bmi'],hue='chargescat',palette='bright',data=df2)
plt.show()
```

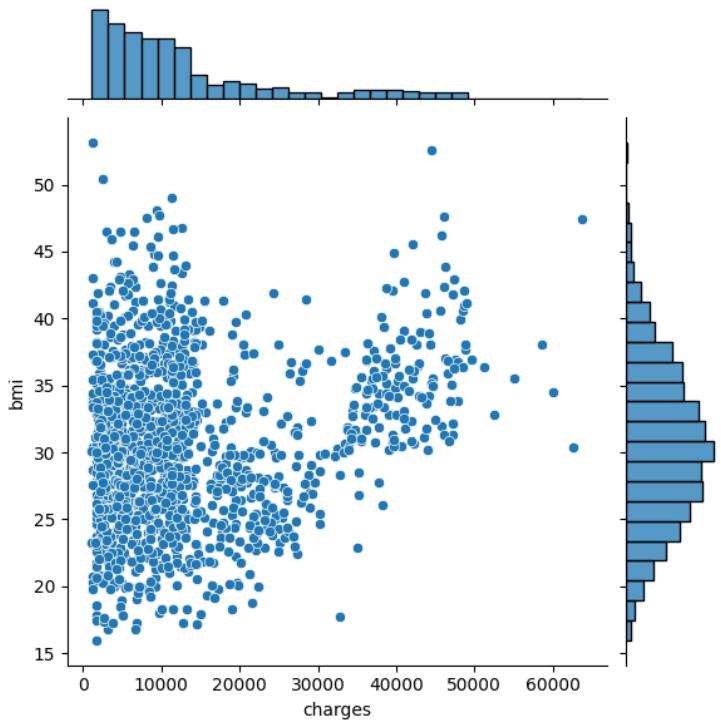
Output:



Code:

```
sns.jointplot(data = df2, x = 'charges', y ='bmi');  
plt.show()
```

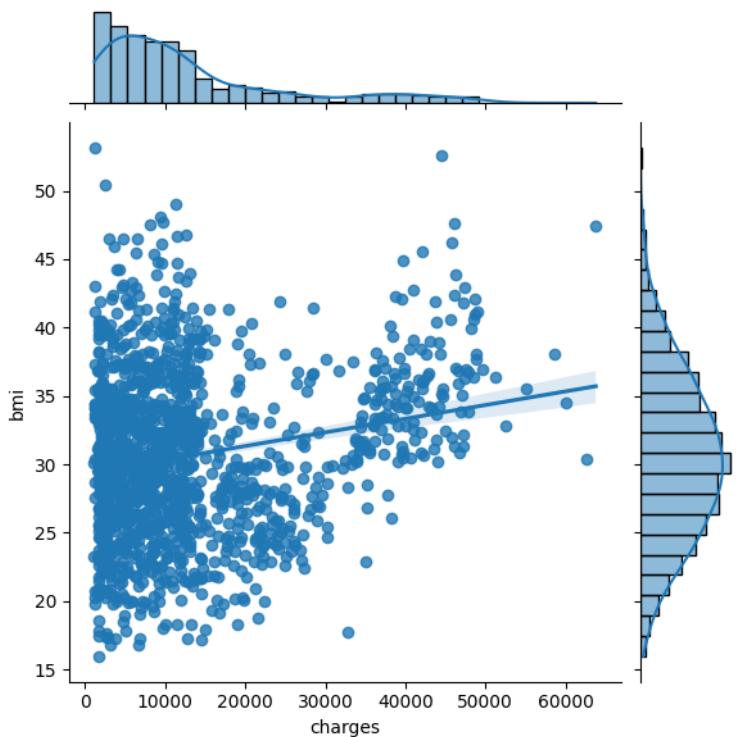
Output:



Code:

```
sns.jointplot(data = df2, x = 'charges', y ='bmi', kind='reg');  
plt.show()
```

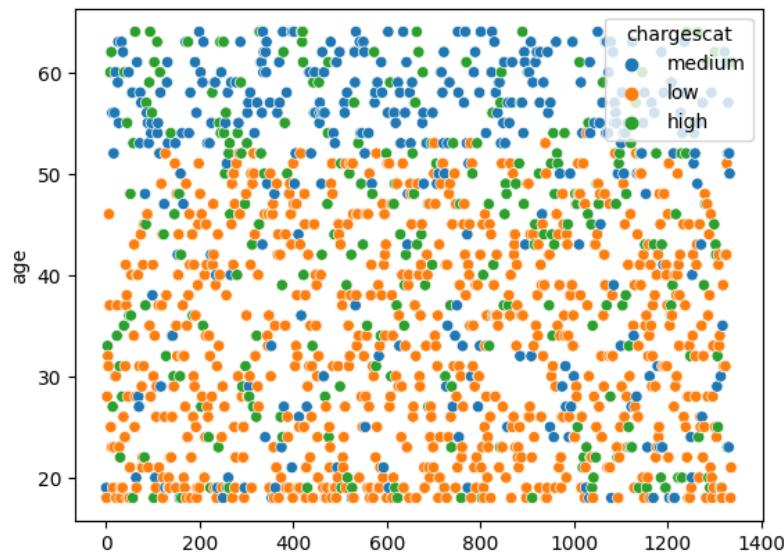
Output:



Code:

```
sns.scatterplot(x=df2.index,y=df2['age'],hue = df2['chargescat'])
```

Output:



RESULT :

Hence Bivariate Analysis for the given dataset was performed successfully.

EXPT NO : 2 d)	
DATE : 08/02/23	

MULTIVARIATE ANALYSIS

AIM :

To perform Multivariate Analysis on the health insurance Dataset.

CODES AND OUTPUTS:

Code:

```
#necessary libraries
import numpy as np
import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt
df = pd.read_csv("../datasets/insurance.csv")
df.head()
```

Output:

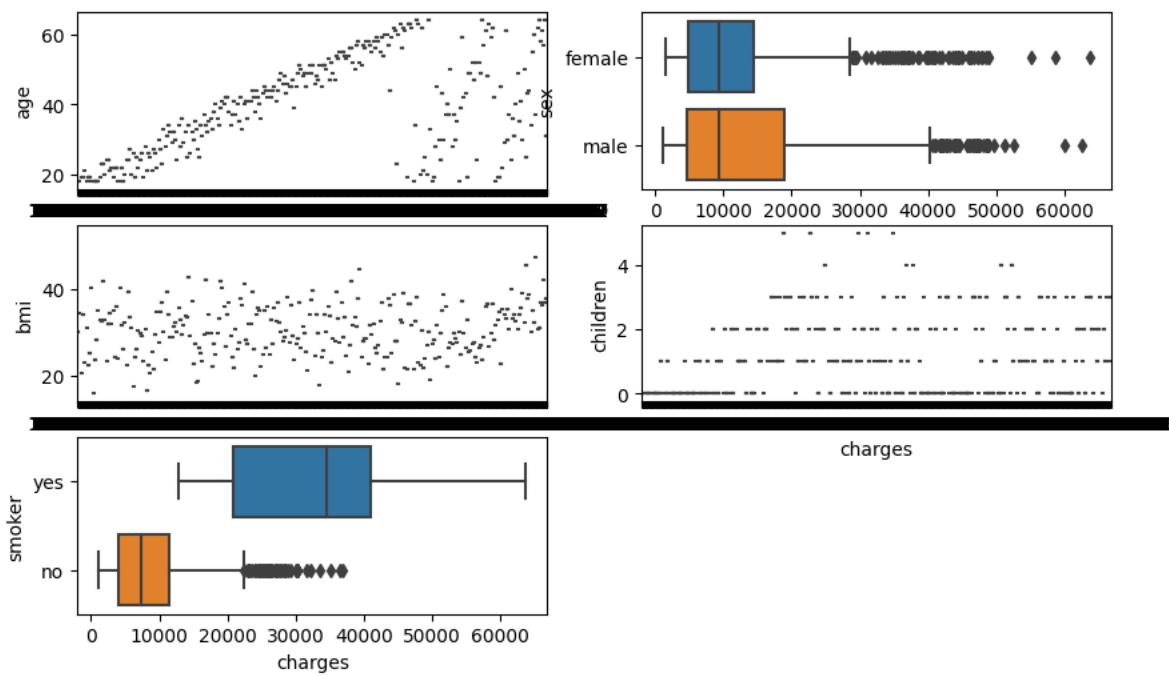
	age	sex	bmi	children	smoker	region	charges
0	19	female	27.900	0	yes	southwest	16884.92400
1	18	male	33.770	1	no	southeast	1725.55230
2	28	male	33.000	3	no	southeast	4449.46200
3	33	male	22.705	0	no	northwest	21984.47061
4	32	male	28.880	0	no	northwest	3866.85520
...
1333	50	male	30.970	3	no	northwest	10600.54830
1334	18	female	31.920	0	no	northeast	2205.98080
1335	18	female	36.850	0	no	southeast	1629.83350
1336	21	female	25.800	0	no	southwest	2007.94500
1337	61	female	29.070	0	yes	northwest	29141.36030

1338 rows × 7 columns

Code:

```
def graph(y):
    sns.boxplot(x="charges", y=y, data=df)
plt.figure(figsize=(10,10))
c=0;
for i in df.columns[0:5]:
    plt.subplot(521+c)
    c+=1
    graph(i)
plt.show()
```

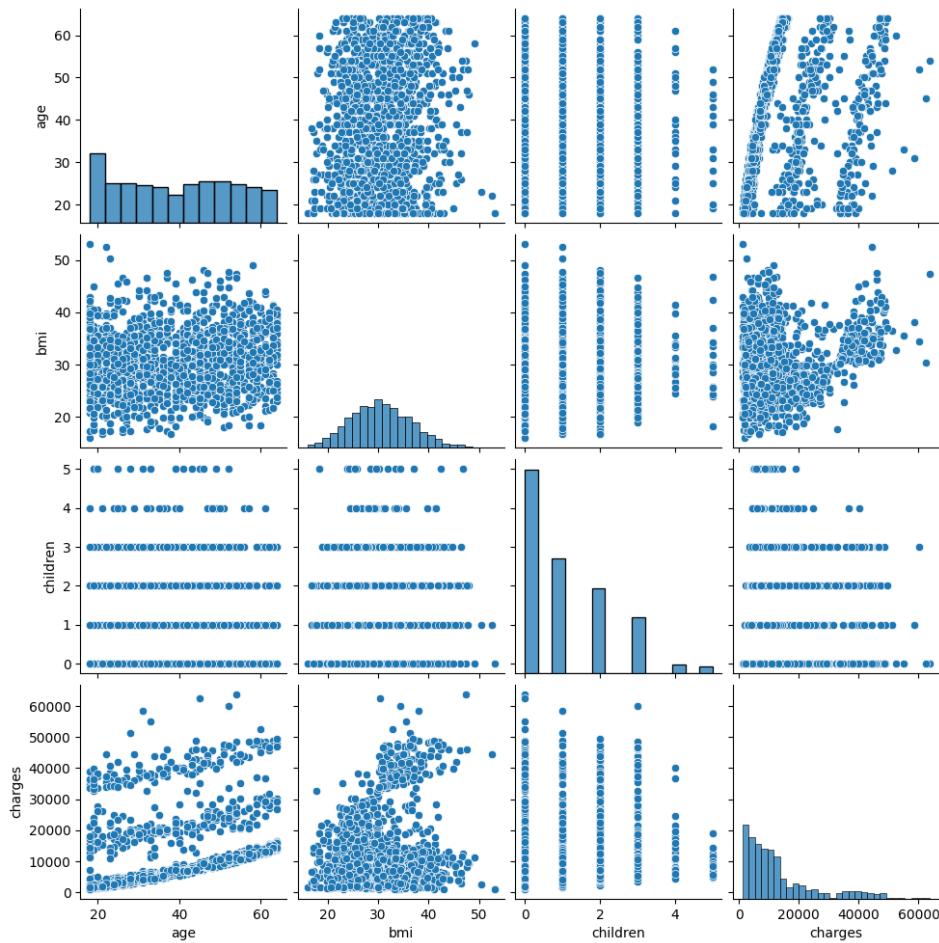
Output:



Code:

```
# pairplot
sns.pairplot(df)
plt.show()
```

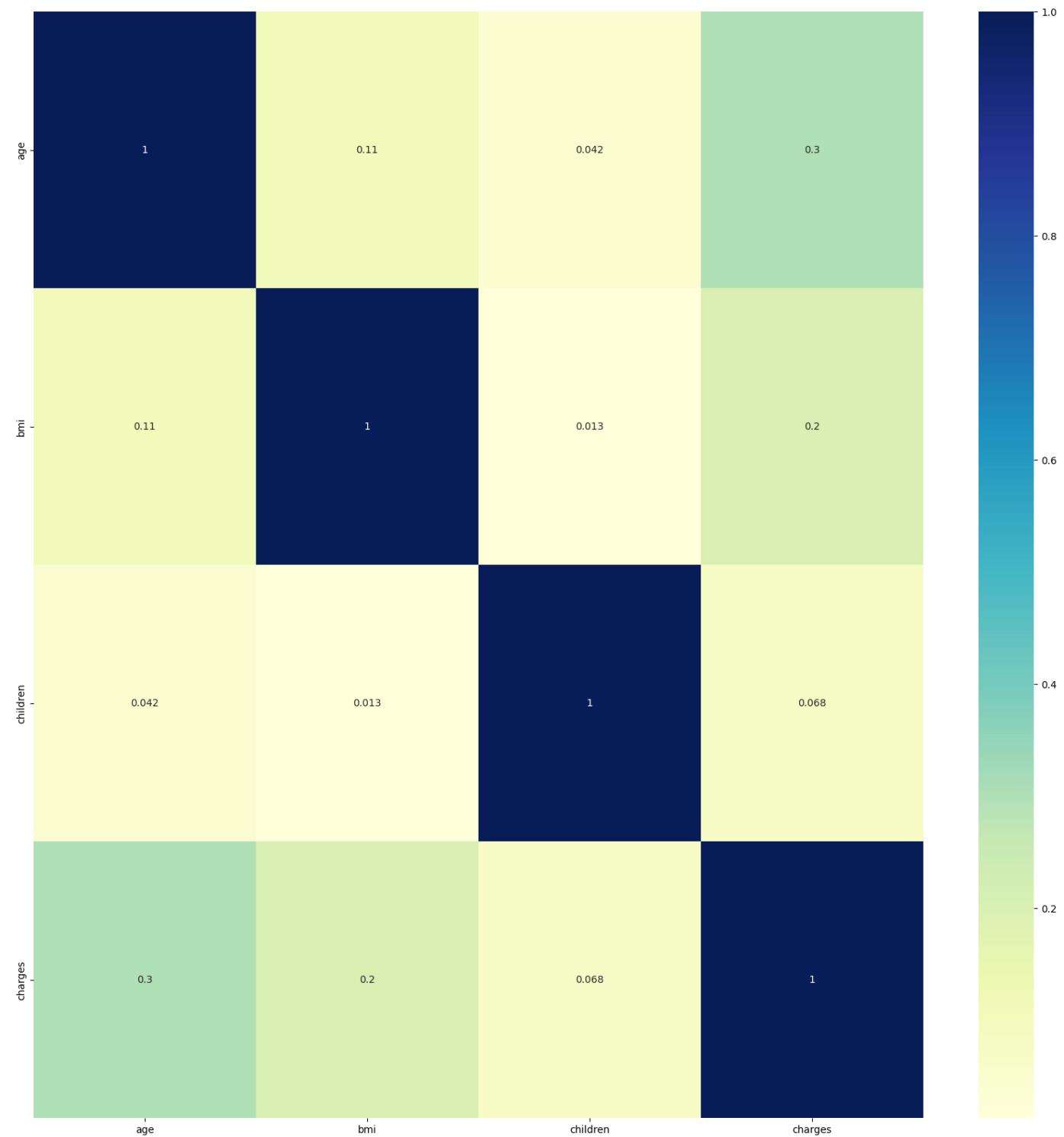
Output:



Code:

```
correlation = df.corr()
fig, ax = plt.subplots(figsize=(20,20))
sns.heatmap(correlation, xticklabels = correlation.columns, yticklabels =
correlation.columns,cmap="YlGnBu",annot= True)
```

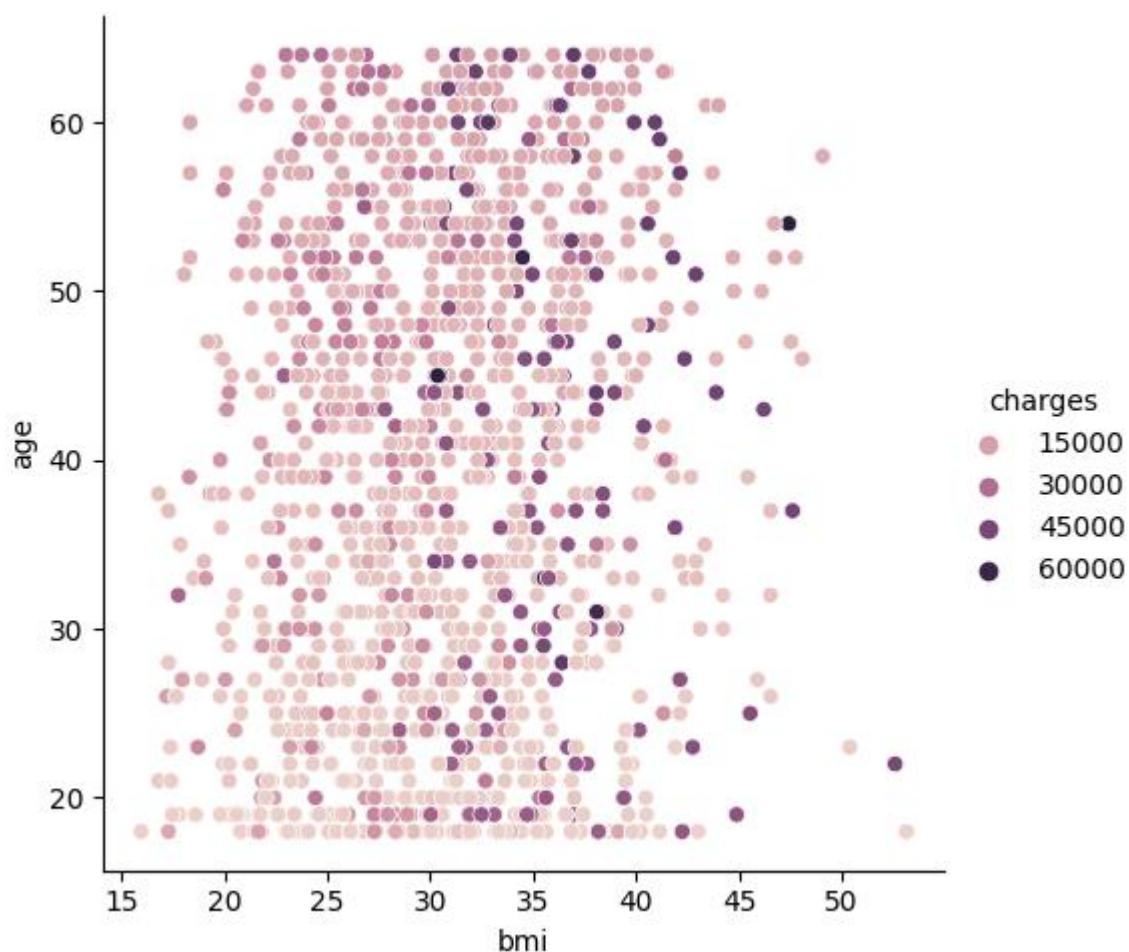
Output:



Code:

```
sns.relplot(x='bmi',y='age',hue='charges',data=df)
```

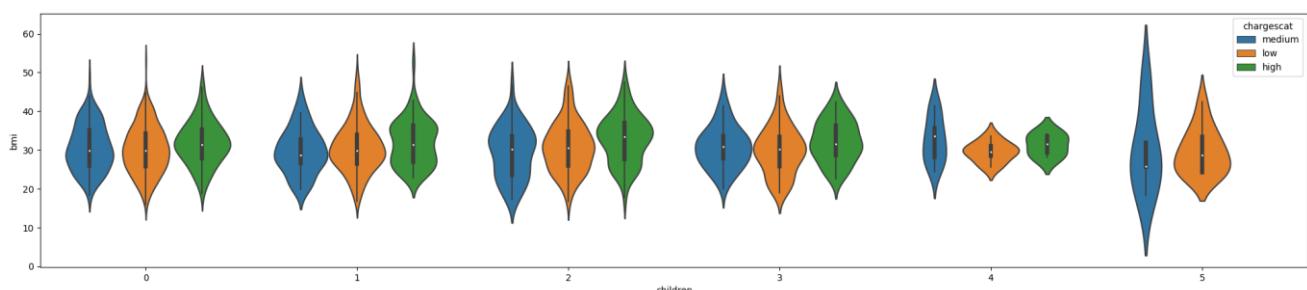
Output:



Code:

```
df['chargescat']= df['charges'].apply(lambda x: 'low' if x<10000 else 'medium' if x<20000  
else 'high')  
plt.figure(figsize=(25,5))  
sns.violinplot(x=df['children'],y=df['bmi'],hue='chargescat',data=df)
```

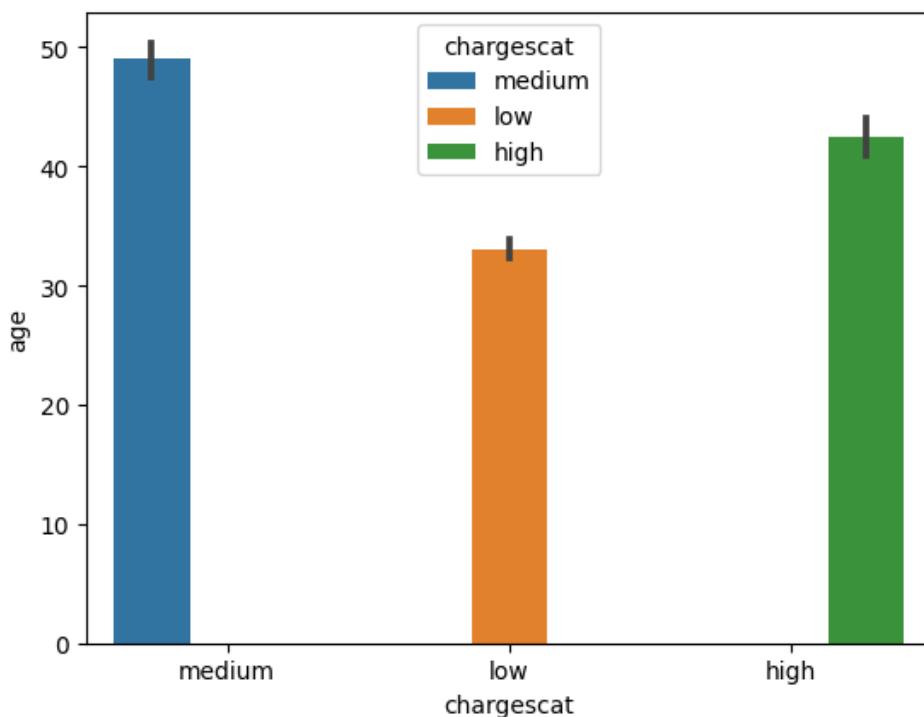
Output:



Code:

```
sns.barplot(x=df['chargescat'],y=df['age'],hue='chargescat',data=df)
```

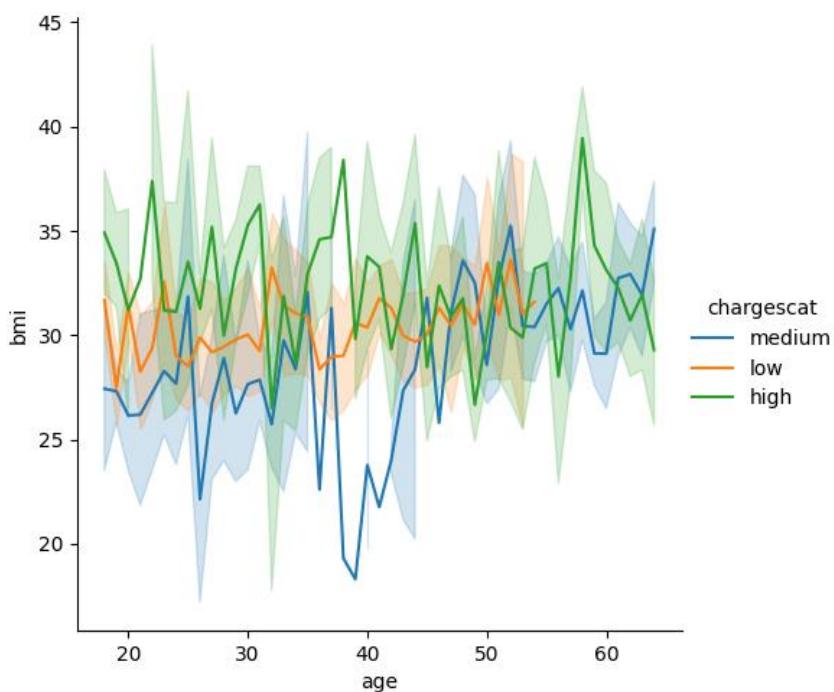
Output:



Code:

```
#high temeperature means higher respiratory rate->running or fever  
sns.relplot(data=df,x=df['age'],y=df['bmi'],hue=df['chargescat'],kind='line')
```

Output:



RESULT :

Hence Multivariate Analysis for the given dataset was performed successfully.

EXPT NO : 3 a)

LINEAR REGRESSION

DATE :
01/03/23

AIM:

To implement the linear regression using the dataset in python

CODES AND OUTPUTS :

1) Import libraries:

```
import pandas as pd  
import numpy as np
```

2) Read dataset

```
df = pd.read_csv('../datasets/insurance.csv')  
df.head()
```

	age	sex	bmi	children	smoker	region	charges
0	19	female	27.900	0	yes	southwest	16884.92400
1	18	male	33.770	1	no	southeast	1725.55230
2	28	male	33.000	3	no	southeast	4449.46200
3	33	male	22.705	0	no	northwest	21984.47061
4	32	male	28.880	0	no	northwest	3866.85520

```
df.shape
```

(1338, 7)

```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>  
RangeIndex: 1338 entries, 0 to 1337  
Data columns (total 7 columns):  
 #   Column      Non-Null Count  Dtype     
---  --    
 0   age         1338 non-null    int64    
 1   sex         1338 non-null    object   
 2   bmi         1338 non-null    float64  
 3   children    1338 non-null    int64    
 4   smoker      1338 non-null    object   
 5   region      1338 non-null    object   
 6   charges     1338 non-null    float64  
dtypes: float64(2), int64(2), object(3)  
memory usage: 73.3+ KB
```

```
df
```

	age	sex	bmi	children	smoker	region	charges
0	19	0	27.900	0	1	3	16884
1	18	1	33.770	1	0	2	1725
2	28	1	33.000	3	0	2	4449
3	33	1	22.705	0	0	1	21984
4	32	1	28.880	0	0	1	3866
...
1333	50	1	30.970	3	0	1	10600
1334	18	0	31.920	0	0	0	2205
1335	18	0	36.850	0	0	2	1629
1336	21	0	25.800	0	0	3	2007
1337	61	0	29.070	0	1	1	29141

1338 rows × 7 columns

Linear regression:

```
import math
import sklearn.metrics as sm
from sklearn.linear_model import LinearRegression
X=df.drop('charges',axis=1)
y=df['charges']
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.3, random_state=0)
model = LinearRegression()
model.fit(X_train, y_train)
y_pred=model.predict(X_test)
x=math.sqrt(sm.mean_squared_error(y_test, y_pred))
```

```
x
```

5771.616138591527

```
x/(df.charges.min()+df.charges.max())
```

0.8894324542065196

```

import numpy as np
import matplotlib.pyplot as plt

def estimate_coef(x, y):
    # number of observations/points
    n = np.size(x)

    # mean of x and y vector
    m_x = np.mean(x)
    m_y = np.mean(y)

    # calculating cross-deviation and deviation about x
    SS_xy = np.sum(y*x) - n*m_y*m_x
    SS_xx = np.sum(x*x) - n*m_x*m_x

    # calculating regression coefficients
    b_1 = SS_xy / SS_xx
    b_0 = m_y - b_1*m_x

    return (b_0, b_1)

def plot_regression_line(x, y, b):
    # plotting the actual points as scatter plot
    plt.scatter(x, y, color = "m",marker = "o", s = 30)

    # predicted response vector
    y_pred = b[0] + b[1]*x

    # plotting the regression line
    plt.plot(x, y_pred, color = "g")

    # putting labels
    plt.xlabel('x')
    plt.ylabel('y')

    # function to show plot
    plt.show()

def main():
    # observations / data
    x=df["age"]
    y=df['charges']

    # estimating coefficients
    b = estimate_coef(x, y)
    print("Estimated coefficients:\nb_0 = {} \nb_1 = {}".format(b[0], b[1]))

    # plotting regression line
    plot_regression_line(x, y, b)

if __name__ == "__main__":
    main()

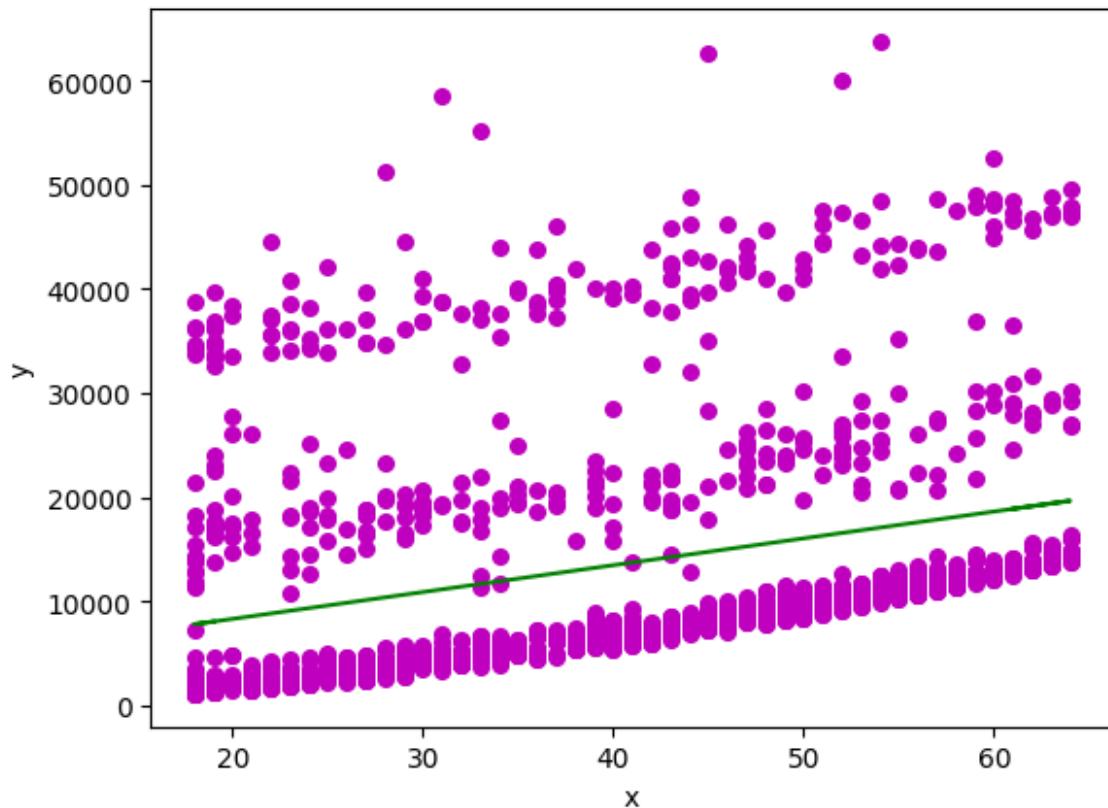
```

Output:

Estimated coefficients:

$b_0 = 3165.373765194823$

$b_1 = 257.72307711106436$



Result:

Thus linear regression has been implemented on given dataset and the results are verified

EXPT NO : 3 b)	LOGISTIC REGRESSION
DATE : 01/03/23	

LOGISTIC REGRESSION

AIM:

To implement the logistic regression using the dataset in python

CODES AND OUTPUTS :

1) Import libraries:

```
import pandas as pd  
import numpy as np
```

2) Read dataset

```
df = pd.read_csv('../datasets/insurance.csv')  
df.head()
```

	age	sex	bmi	children	smoker	region	charges
0	19	female	27.900	0	yes	southwest	16884.92400
1	18	male	33.770	1	no	southeast	1725.55230
2	28	male	33.000	3	no	southeast	4449.46200
3	33	male	22.705	0	no	northwest	21984.47061
4	32	male	28.880	0	no	northwest	3866.85520

```
df.shape
```

(1338, 7)

```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>  
RangeIndex: 1338 entries, 0 to 1337  
Data columns (total 7 columns):  
 #   Column      Non-Null Count  Dtype     
---    
 0   age         1338 non-null    int64    
 1   sex         1338 non-null    object    
 2   bmi         1338 non-null    float64   
 3   children    1338 non-null    int64    
 4   smoker      1338 non-null    object    
 5   region      1338 non-null    object    
 6   charges     1338 non-null    float64   
dtypes: float64(2), int64(2), object(3)  
memory usage: 73.3+ KB
```

```
df['charges'] = df['charges'].astype('int')
df.head()
```

	age	sex	bmi	children	smoker	region	charges
0	19	female	27.900	0	yes	southwest	16884
1	18	male	33.770	1	no	southeast	1725
2	28	male	33.000	3	no	southeast	4449
3	33	male	22.705	0	no	northwest	21984
4	32	male	28.880	0	no	northwest	3866

Regression :

Code:

```
# df2 = df.drop('chargescat', axis =1)
X=df.drop('charges',axis=1)
y=df['charges']
```

```
import numpy
from sklearn import preprocessing

label_encoder = preprocessing.LabelEncoder()
for i in df.columns:
    if type(df[i][2])==numpy.bool_ or type(df[i][0])==str:
        df[i]= label_encoder.fit_transform(df[i])
```

```
from sklearn.model_selection import train_test_split
from sklearn.metrics import balanced_accuracy_score
```

```
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.30, random_state=0)
from sklearn.linear_model import LogisticRegression
model = LogisticRegression(random_state=0,max_iter=1000)
model.fit(X_train, y_train)
y_pred=model.predict(X_test)
```

Output:

```
balanced_accuracy_score(y_test, y_pred)
0.7556675062972292
```

RESULT:

Thus logistic regression has been implemented on given dataset and the results are verified

EXPT NO : 4	
a)	
DATE :	
01/03/23	

NAÏVE BAYES CLASSIFIER

AIM:

To implement the naïve bayes using the given dataset in python

CODES AND OUTPUTS :

Code :

Import libraries:

```
import pandas as pd  
import numpy as np
```

Upload Dataset:

```
df = pd.read_csv('../datasets/insurance.csv')  
df.head()
```

	age	sex	bmi	children	smoker	region	charges
0	19	female	27.900	0	yes	southwest	16884.92400
1	18	male	33.770	1	no	southeast	1725.55230
2	28	male	33.000	3	no	southeast	4449.46200
3	33	male	22.705	0	no	northwest	21984.47061
4	32	male	28.880	0	no	northwest	3866.85520

```
df.shape
```

(1338, 7)

```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>  
RangeIndex: 1338 entries, 0 to 1337  
Data columns (total 7 columns):  
 #   Column      Non-Null Count  Dtype     
---    
 0   age         1338 non-null   int64    
 1   sex         1338 non-null   object   
 2   bmi         1338 non-null   float64  
 3   children    1338 non-null   int64    
 4   smoker      1338 non-null   object   
 5   region      1338 non-null   object   
 6   charges     1338 non-null   float64  
dtypes: float64(2), int64(2), object(3)  
memory usage: 73.3+ KB
```

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

Output

```
age      0  
sex      0  
bmi      0  
children 0  
smoker   0  
region   0  
charges  0  
dtype: int64
```

```
# convert categorical data to numerical data  
df['charges'] = df['charges'].astype('int')  
df.head()
```

	age	sex	bmi	children	smoker	region	charges
0	19	female	27.900	0	yes	southwest	16884
1	18	male	33.770	1	no	southeast	1725
2	28	male	33.000	3	no	southeast	4449
3	33	male	22.705	0	no	northwest	21984
4	32	male	28.880	0	no	northwest	3866

```
from sklearn.model_selection import train_test_split  
from sklearn.naive_bayes import GaussianNB  
from sklearn.metrics import balanced_accuracy_score
```

```
df.columns
```

```
Index(['age', 'sex', 'bmi', 'children', 'smoker', 'region', 'charges'], dt  
ype='object')
```

```
df['chargecat']= df['charges'].apply(lambda x: 'low' if x<10000 else 'medium' if x<2000
```

```
df.head()
```

	age	sex	bmi	children	smoker	region	charges	chargescat
0	19	female	27.900	0	yes	southwest	16884	medium
1	18	male	33.770	1	no	southeast	1725	low
2	28	male	33.000	3	no	southeast	4449	low
3	33	male	22.705	0	no	northwest	21984	high
4	32	male	28.880	0	no	northwest	3866	low

```
import numpy
from sklearn import preprocessing

label_encoder = preprocessing.LabelEncoder()
for i in df.columns:
    if type(df[i][2]) == numpy.bool_ or type(df[i][0]) == str:
        df[i] = label_encoder.fit_transform(df[i])
```

```
df2 = df.drop('charges', axis=1)
X=df2.drop('chargescat',axis=1)
y=df2['chargescat']
```

```
print(y)
print(X)
```

```
0      2
1      1
2      1
3      0
4      1
..
1333   2
1334   1
1335   1
1336   1
1337   0
Name: chargescat, Length: 1338, dtype: int32
   age  sex     bmi  children  smoker  region
0    19    0  27.900       0      1      3
1    18    1  33.770       1      0      2
2    28    1  33.000       3      0      2
3    33    1  22.705       0      0      1
4    32    1  28.880       0      0      1
...
1333   50    1  30.970       3      0      1
1334   18    0  31.920       0      0      0
1335   18    0  36.850       0      0      2
1336   21    0  25.800       0      0      3
1337   61    0  29.070       0      1      1
[1338 rows x 6 columns]
```

```
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.30, random_state=0

X=df.drop('chargescat',axis=1)
y=df['charges']

X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.3, random_state=0)
gnb = GaussianNB()
y_pred = gnb.fit(X_train, y_train).predict(X_test)
```

Output:

```
balanced_accuracy_score(y_test, y_pred)
```

0.9858942065491

Result:

Thus Classification using Naïve bayes classifier has been implemented on given dataset and the results are verified

EXPT NO : 4
a)
DATE : 01/03/23

SUPPORT VECTOR MACHINE

AIM:

To implement the naïve bayes using the given dataset in python

CODES AND OUTPUTS :

Code :

Import libraries:

```
import pandas as pd
import numpy as np
```

Upload Dataset:

```
df = pd.read_csv('../datasets/insurance.csv')
df.head()
```

	age	sex	bmi	children	smoker	region	charges
0	19	female	27.900	0	yes	southwest	16884.92400
1	18	male	33.770	1	no	southeast	1725.55230
2	28	male	33.000	3	no	southeast	4449.46200
3	33	male	22.705	0	no	northwest	21984.47061
4	32	male	28.880	0	no	northwest	3866.85520

df.shape

(1338, 7)

df.info()

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1338 entries, 0 to 1337
Data columns (total 7 columns):
 #   Column      Non-Null Count  Dtype  
--- 
 7   age         1338 non-null   int64  
 8   sex         1338 non-null   object  
 9   bmi         1338 non-null   float64 
 10  children    1338 non-null   int64  
 11  smoker      1338 non-null   object  
 12  region      1338 non-null   object  
 13  charges     1338 non-null   float64 
dtypes: float64(2), int64(2), object(3)
memory usage: 73.3+ KB
```

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

Out[24]:

```
age      0  
sex      0  
bmi      0  
children 0  
smoker   0  
region   0  
charges  0  
dtype: int64
```

```
# convert categorical data to numerical data  
df['charges'] = df['charges'].astype('int')  
df.head()
```

	age	sex	bmi	children	smoker	region	charges
0	19	female	27.900	0	yes	southwest	16884
1	18	male	33.770	1	no	southeast	1725
2	28	male	33.000	3	no	southeast	4449
3	33	male	22.705	0	no	northwest	21984
4	32	male	28.880	0	no	northwest	3866

```
from sklearn.model_selection import train_test_split  
from sklearn.naive_bayes import GaussianNB  
from sklearn.metrics import balanced_accuracy_score
```

```
df.columns
```

```
Index(['age', 'sex', 'bmi', 'children', 'smoker', 'region', 'charges'], dt  
ype='object')
```

```
df['chargecat']= df['charges'].apply(lambda x: 'low' if x<10000 else 'medium' if x<2000
```

```
df.head()
```

	age	sex	bmi	children	smoker	region	charges	chargescat
0	19	female	27.900	0	yes	southwest	16884	medium
1	18	male	33.770	1	no	southeast	1725	low
2	28	male	33.000	3	no	southeast	4449	low
3	33	male	22.705	0	no	northwest	21984	high
4	32	male	28.880	0	no	northwest	3866	low

```
import numpy
from sklearn import preprocessing

label_encoder = preprocessing.LabelEncoder()
for i in df.columns:
    if type(df[i][2]) == numpy.bool_ or type(df[i][0]) == str:
        df[i] = label_encoder.fit_transform(df[i])
```

```
df2 = df.drop('charges', axis=1)
X=df2.drop('chargescat',axis=1)
y=df2['chargescat']
```

```
print(y)
print(X)
0      2
1      1
2      1
3      0
4      1
..
1333    2
1334    1
1335    1
1336    1
1337    0
Name: chargescat, Length: 1338, dtype: int32
   age  sex     bmi  children  smoker  region
0    19    0  27.900       0      1      3
1    18    1  33.770       1      0      2
2    28    1  33.000       3      0      2
3    33    1  22.705       0      0      1
4    32    1  28.880       0      0      1
...
1333   50    1  30.970       3      0      1
1334   18    0  31.920       0      0      0
1335   18    0  36.850       0      0      2
1336   21    0  25.800       0      0      3
1337   61    0  29.070       0      1      1
```

```
[1338 rows x 6 columns]
```

```
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.30, random_state=0)
```

```
from sklearn import svm  
  
clf = svm.SVC()  
y_pred=clf.fit(X, y).predict(X_test)
```

Output:

```
balanced_accuracy_score(y_test, y_pred)
```

0.816384589421

Result:

Thus Classification using SVM has been implemented on given dataset and the results are verified

Aim:

To understand OpenStack deployment, its implementation and its applications.

Theory:

OpenStack is a cloud operating system that controls large pools of compute, storage and networking resources throughout a data-centre, all managed and provisioned through APIs with common authentication mechanisms. A dashboard is also available, giving administrators control while empowering their users to provide resources through a web interface. It began in 2010 as a joint project of Rackspace Hosting and NASA. It was managed by the OpenStack Foundation, a non-profit entity.

Working of OpenStack

It is essentially a series of commands known as scripts. These scripts are bundled into packages called projects that relay tasks that create cloud environments. To create these environments, OpenStack relies on

- Virtualization that creates a layer of virtual resources abstracted from hardware
- A base as that carries out commands given by OpenStack scripts

OpenStack uses virtualized resources to build clouds. It doesn't execute commands, rather delays them to base OS.

Components of OpenStack

OpenStack's architecture is made up of numerous open source projects. These are used to set up OpenStack's undercloud and overcloud. Overcloud is used by cloud users and undercloud is used by system admins. Underclouds contain the core components system admins need to set up and manage end user's environments called overclouds.

There are 6 stable, core services that handle computing, networking, storage, identity and images. These make up the infrastructure of OpenStack that allows the rest of the projects to handle dashboarding, orchestration, etc.

The 6 components are:

- Nova: It is a full management tool that helps compute resource-handling, scheduling, creation, deletion, etc.
- Neutron: It connects the networks across other services
- Swift: It is a highly fault-tolerant object storage service that stores and retrieves unstructured data objects
- Cinder: Provides persistent block storage

- Keystone: Authenticates and authorises all services
- Glance: Stores and retrieves VM disc images from various locations

Deployment of OpenStack

It is mostly deployed as infrastructure-as-a-Service (IaaS) in both public and private clouds where virtual servers and other resources are made available to users. The software platform consists of interrelated components control diverse, multivendor hardware. Lifecycle management tools and packaging are used to help instances maintain the lifecycle of deployments. Frameworks for the above are Tripleo, OpenStack-helm, kolla-ansible, kayole, etc.

Challenges in Implementation

- Installation challenges
- Documentation
- Upgrading OpenStack
- Long Term Support
- Deployment Models
- OpenStack-based public cloud
- On-premises distribution
- Hosted OpenStack Private Cloud
- Appliance based OpenStack
- Applications of OpenStack
- Private clouds and Public clouds
- Network function virtualization
- Containers

Result:

Thus, OpenStack deployment, its implementation and its applications have been studied.

Aim:

To install OpenStack and implement Infrastructure-as-a-service using it.

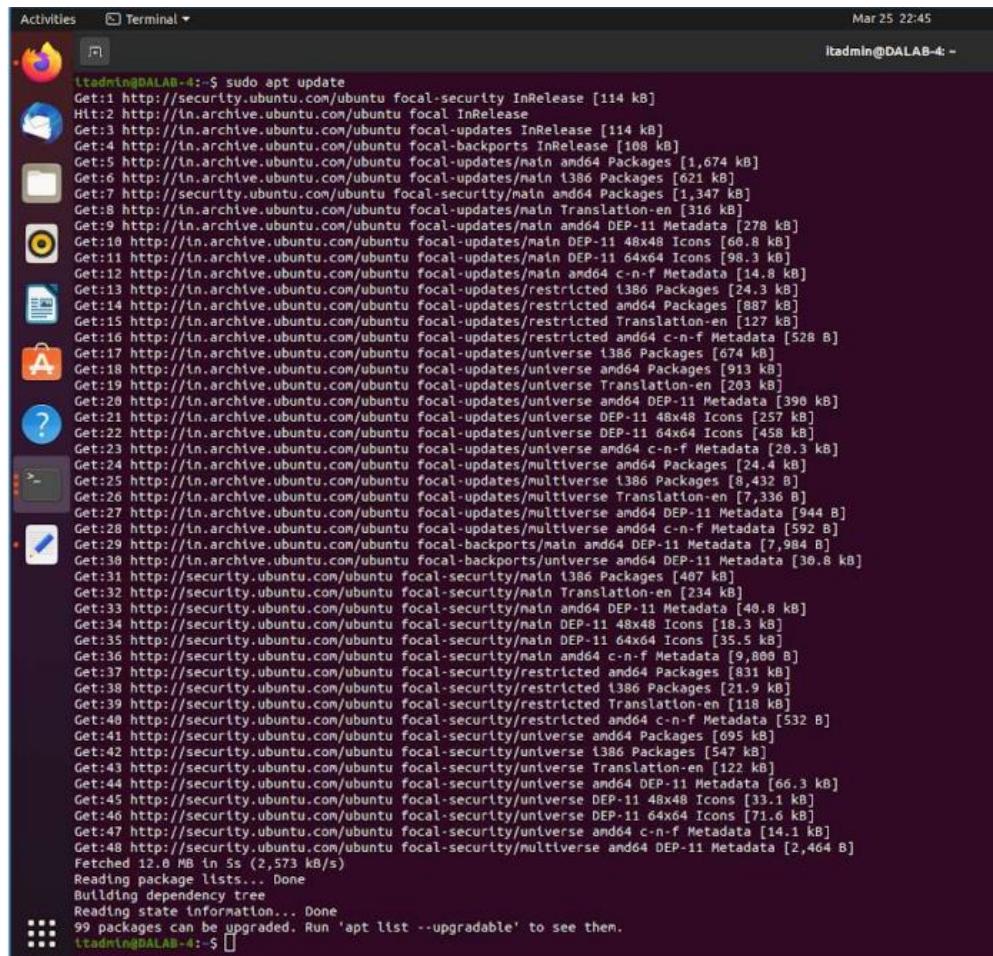
Procedure:

Installation of OpenStack in Ubuntu

1. Login to sudo ("superuser do") user
2. Open Terminal
3. In terminal, type:

```
sudo apt – update
```

This downloads package information from all configured sources.



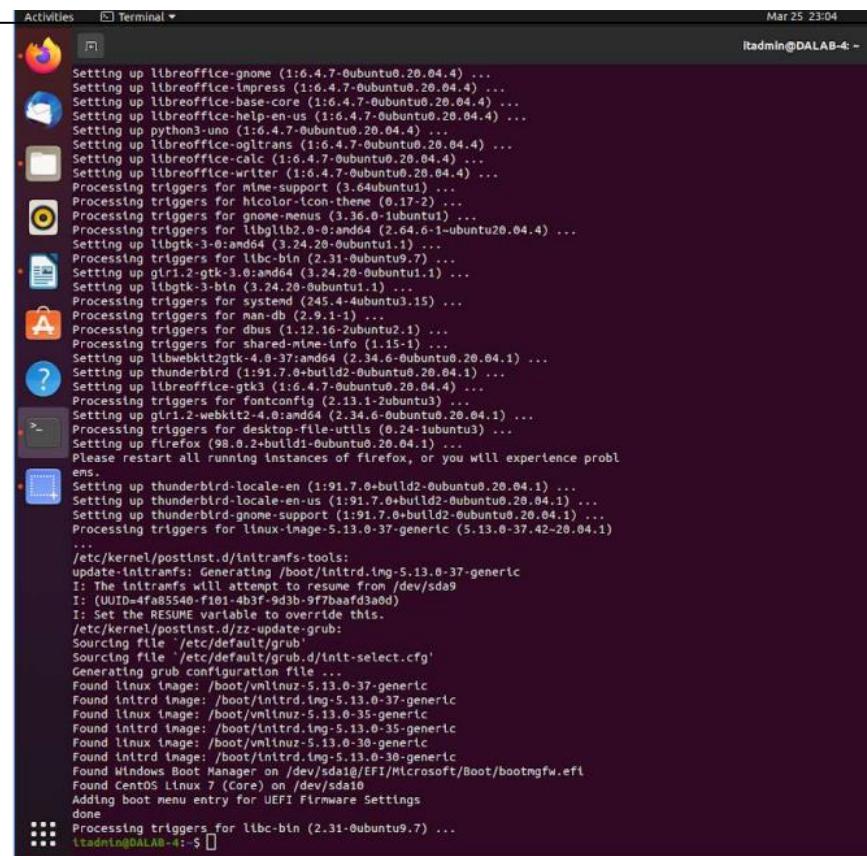
The screenshot shows a terminal window titled "Terminal" with the command "sudo apt update" entered. The output lists numerous package requests and their details, such as URLs, file sizes, and metadata. The process is shown to have fetched 12.0 MB in 5s (2,573 kb/s) and found 99 packages upgradable. The terminal prompt "ttadmin@DALAB-4:~\$ " is visible at the bottom.

```
ttadmin@DALAB-4:~$ sudo apt update
Get:1 http://security.ubuntu.com/ubuntu focal-security InRelease [114 kB]
Hit:2 http://in.archive.ubuntu.com/ubuntu focal InRelease
Get:3 http://in.archive.ubuntu.com/ubuntu focal-updates InRelease [114 kB]
Get:4 http://in.archive.ubuntu.com/ubuntu focal-backports InRelease [108 kB]
Get:5 http://in.archive.ubuntu.com/ubuntu focal-updates/nain amd64 Packages [1,674 kB]
Get:6 http://in.archive.ubuntu.com/ubuntu focal-updates/nain i386 Packages [621 kB]
Get:7 http://security.ubuntu.com/ubuntu focal-security/main amd64 Packages [1,347 kB]
Get:8 http://in.archive.ubuntu.com/ubuntu focal-updates/nain Translation-en [316 kB]
Get:9 http://in.archive.ubuntu.com/ubuntu focal-updates/main amd64 DEP-11 Metadata [278 kB]
Get:10 http://in.archive.ubuntu.com/ubuntu focal-updates/main DEP-11 48x48 Icons [60.8 kB]
Get:11 http://in.archive.ubuntu.com/ubuntu focal-updates/main DEP-11 64x64 Icons [98.3 kB]
Get:12 http://in.archive.ubuntu.com/ubuntu focal-updates/main amd64 c-n-f Metadata [14.8 kB]
Get:13 http://in.archive.ubuntu.com/ubuntu focal-updates/restricted i386 Packages [24.3 kB]
Get:14 http://in.archive.ubuntu.com/ubuntu focal-updates/restricted amd64 Packages [887 kB]
Get:15 http://in.archive.ubuntu.com/ubuntu focal-updates/restricted Translation-en [127 kB]
Get:16 http://in.archive.ubuntu.com/ubuntu focal-updates/restricted amd64 c-n-f Metadata [528 kB]
Get:17 http://in.archive.ubuntu.com/ubuntu focal-updates/universe i386 Packages [674 kB]
Get:18 http://in.archive.ubuntu.com/ubuntu focal-updates/universe amd64 Packages [913 kB]
Get:19 http://in.archive.ubuntu.com/ubuntu focal-updates/universe Translation-en [203 kB]
Get:20 http://in.archive.ubuntu.com/ubuntu focal-updates/universe amd64 DEP-11 Metadata [390 kB]
Get:21 http://in.archive.ubuntu.com/ubuntu focal-updates/universe DEP-11 48x48 Icons [257 kB]
Get:22 http://in.archive.ubuntu.com/ubuntu focal-updates/universe DEP-11 64x64 Icons [456 kB]
Get:23 http://in.archive.ubuntu.com/ubuntu focal-updates/universe amd64 c-n-f Metadata [28.3 kB]
Get:24 http://in.archive.ubuntu.com/ubuntu focal-updates/multiverse amd64 Packages [24.4 kB]
Get:25 http://in.archive.ubuntu.com/ubuntu focal-updates/multiverse i386 Packages [8,432 kB]
Get:26 http://in.archive.ubuntu.com/ubuntu focal-updates/multiverse Translation-en [7,336 kB]
Get:27 http://in.archive.ubuntu.com/ubuntu focal-updates/multiverse amd64 DEP-11 Metadata [944 kB]
Get:28 http://in.archive.ubuntu.com/ubuntu focal-updates/multiverse amd64 c-n-f Metadata [592 kB]
Get:29 http://in.archive.ubuntu.com/ubuntu focal-backports/main amd64 DEP-11 Metadata [7,984 kB]
Get:30 http://in.archive.ubuntu.com/ubuntu focal-backports/universe amd64 DEP-11 Metadata [30.8 kB]
Get:31 http://security.ubuntu.com/ubuntu focal-security/nain i386 Packages [407 kB]
Get:32 http://security.ubuntu.com/ubuntu focal-security/nain Translation-en [234 kB]
Get:33 http://security.ubuntu.com/ubuntu focal-security/nain amd64 DEP-11 Metadata [40.8 kB]
Get:34 http://security.ubuntu.com/ubuntu focal-security/nain DEP-11 48x48 Icons [18.3 kB]
Get:35 http://security.ubuntu.com/ubuntu focal-security/nain DEP-11 64x64 Icons [35.5 kB]
Get:36 http://security.ubuntu.com/ubuntu focal-security/nain amd64 c-n-f Metadata [9,800 B]
Get:37 http://security.ubuntu.com/ubuntu focal-security/restricted amd64 Packages [831 kB]
Get:38 http://security.ubuntu.com/ubuntu focal-security/restricted i386 Packages [21.9 kB]
Get:39 http://security.ubuntu.com/ubuntu focal-security/restricted Translation-en [118 kB]
Get:40 http://security.ubuntu.com/ubuntu focal-security/restricted amd64 c-n-f Metadata [532 B]
Get:41 http://security.ubuntu.com/ubuntu focal-security/universe amd64 Packages [695 kB]
Get:42 http://security.ubuntu.com/ubuntu focal-security/universe i386 Packages [547 kB]
Get:43 http://security.ubuntu.com/ubuntu focal-security/universe Translation-en [122 kB]
Get:44 http://security.ubuntu.com/ubuntu focal-security/universe amd64 DEP-11 Metadata [66.3 kB]
Get:45 http://security.ubuntu.com/ubuntu focal-security/universe DEP-11 48x48 Icons [33.1 kB]
Get:46 http://security.ubuntu.com/ubuntu focal-security/universe DEP-11 64x64 Icons [71.6 kB]
Get:47 http://security.ubuntu.com/ubuntu focal-security/universe amd64 c-n-f Metadata [14.1 kB]
Get:48 http://security.ubuntu.com/ubuntu focal-security/multiverse amd64 DEP-11 Metadata [2,464 B]
Fetched 12.0 MB in 5s (2,573 kB/s)
Reading package lists... done
Building dependency tree
Reading state information... done
99 packages can be upgraded. Run 'apt list --upgradable' to see them.
ttadmin@DALAB-4:~$
```

4. Upgrade everything in the system, all the packages, and the kernel to the latest versions as supported by the repositories:

5. Upgrade existing packages, installs new dependencies that are not in the system, and deletes those that are not needed:

```
sudo apt -y upgrade
```



```
Setting up libreoffice-gnome (1:6.4.7-0ubuntu0.20.04.4) ...
Setting up libreoffice-impress (1:6.4.7-0ubuntu0.20.04.4) ...
Setting up libreoffice-base-core (1:6.4.7-0ubuntu0.20.04.4) ...
Setting up libreoffice-help-en-us (1:6.4.7-0ubuntu0.20.04.4) ...
Setting up python3-uno (1:6.4.7-0ubuntu0.20.04.4) ...
Setting up libreoffice-ogtrans (1:6.4.7-0ubuntu0.20.04.4) ...
Setting up libreoffice-calc (1:6.4.7-0ubuntu0.20.04.4) ...
Setting up libreoffice-writer (1:6.4.7-0ubuntu0.20.04.4) ...
Processing triggers for mime-support (3.64ubuntu1) ...
Processing triggers for hicolor-icon-theme (0.17-2) ...
Processing triggers for gnome-menus (3.36.0-1ubuntu1) ...
Processing triggers for libglib2.0-0:amd64 (2.64.6-1ubuntu20.04.4) ...
Setting up libgtk-3-0:amd64 (3.24.20-0ubuntu1.1) ...
Processing triggers for liblc-bin (2.31-0ubuntu9.7) ...
Setting up libglib2.0-0:amd64 (3.24.20-0ubuntu1.1) ...
Setting up libgtk-3-bin (3.24.20-0ubuntu1.1) ...
Processing triggers for systemd (245.4-4ubuntu3.15) ...
Processing triggers for man-db (2.9.1-1) ...
Processing triggers for dbus (1.12.16-2ubuntu2.1) ...
Processing triggers for shared-mime-info (1.15-1) ...
Setting up libwebkit2gtk-4.0-37:amd64 (2.34.6-0ubuntu0.20.04.1) ...
Setting up thunderbird (1:91.7.0+build2-0ubuntu0.20.04.1) ...
Setting up libreoffice-gtk3 (1:6.4.7-0ubuntu0.20.04.4) ...
Processing triggers for fontconfig (2.13.1-2ubuntu3) ...
Setting up gir1.2-webkit2-4.0:amd64 (2.34.6-0ubuntu0.20.04.1) ...
Setting up firefox (98.0.2+build1-0ubuntu0.20.04.1) ...
Please restart all running instances of firefox, or you will experience problems.
Setting up thunderbird-locale-en (1:91.7.0+build2-0ubuntu0.20.04.1) ...
Setting up thunderbird-locale-en-us (1:91.7.0+build2-0ubuntu0.20.04.1) ...
Setting up thunderbird-gnome-support (1:91.7.0+build2-0ubuntu0.20.04.1) ...
Processing triggers for linux-image-5.13.0-37-generic (5.13.0-37.42-20.04.1) ...
...
/etc/kernel/postinst.d/initramfs-tools:
update-initramfs: Generating /boot/initrd.lmg-5.13.0-37-generic
I: The initramfs will attempt to resume from /dev/sda9
I: (UUID=4f8a5540-f101-4b3f-9d3b-977baaf3a0d)
I: Set the RESUME variable to override this.
/etc/kernel/postinst.d/zz-update-grub:
Sourcing file /etc/default/grub
Sourcing file /etc/default/grub.d/init-select.cfg
Generating grub configuration file ...
Found linux image: /boot/vmlinuz-5.13.0-37-generic
Found initrd image: /boot/initrd.lmg-5.13.0-37-generic
Found linux image: /boot/vmlinuz-5.13.0-35-generic
Found initrd image: /boot/initrd.lmg-5.13.0-35-generic
Found linux image: /boot/vmlinuz-5.13.0-30-generic
Found initrd image: /boot/initrd.lmg-5.13.0-30-generic
Found Windows Boot Manager on /dev/sda1@/EFI/Microsoft/Boot/bootmgfw.efi
Found CentOS Linux 7 (Core) on /dev/sda10
Adding boot menu entry for UEFI Firmware Settings
done
Processing triggers for libc-bin (2.31-0ubuntu9.7) ...
itadmin@DALAB-4:~$
```

5. Upgrade existing packages, installs new dependencies that are not in the system, and deletes

Those that are not needed:

```
sudo apt -y dist-upgrade
```



```
itadmin@DALAB-4:~$ sudo apt -y dist-upgrade
[sudo] password for itadmin:
Reading package lists... Done
Building dependency tree
Reading state information... Done
Calculating upgrade... Done
The following packages were automatically installed and are no longer required:
  libfwupdplugin1 linux-headers-5.13.0-30-generic
  linux-hwe-5.13-headers-5.13.0-30 linux-image-5.13.0-30-generic
  linux-modules-5.13.0-30-generic linux-modules-extra-5.13.0-30-generic
Use 'sudo apt autoremove' to remove them.
0 upgraded, 0 newly installed, 0 to remove and 0 not upgraded.
itadmin@DALAB-4:~$
```

6. Add user called stack to run DevStack. It should be run as non-root user with sudo enabled:

```
sudo useradd -s /bin/bash -d /opt/stack -m stack
```

```
base) student@DALAB-4:~/Desktop$ sudo useradd -s /bin/bash -d /opt/stack -m stack
sudo] password for student:
student is not in the sudoers file. This incident will be reported.
base) student@DALAB-4:~/Desktop$ su - itadmin
password:
itadmin@DALAB-4:~$ sudo useradd -s /bin/bash -d /opt/stack -m stack
sudo] password for itadmin:
itadmin@DALAB-4:~$ █
```

7. User should have sudo privileges to make changes to the system:

```
echo "stack ALL=(ALL) NOPASSWD: ALL" | sudo tee /etc/sudoers.d/stack
```

```
itadmin@DALAB-4:~$ echo "stack ALL=(ALL) NOPASSWD: ALL" | sudo tee /etc/sudoers.d/stack  
stack ALL=(ALL) NOPASSWD: ALL  
itadmin@DALAB-4:~$ █
```

8. Log in to the stack once user is created:

sudo su – stac

```
itadmin@DALAB-4:~$ sudo su - stack  
stack@DALAB-4:~$ sudo su -
```

1. Install git:

```
sudo apt -y install git
```

```
stack@DALAB-4: ~ $ sudo apt -y install git
Reading package lists... Done
Building dependency tree
Reading state information... Done
The following packages were automatically installed and are no longer required:
  libfwupdplugin1 linux-headers-5.13.0-30-generic
  linux-hwe-5.13-headers-5.13.0-30 linux-image-5.13.0-30-generic
  linux-modules-5.13.0-30-generic linux-modules-extra-5.13.0-30-generic
Use 'sudo apt autoremove' to remove them.
The following additional packages will be installed:
  git-man liberror-perl
Suggested packages:
  git-daemon-run | git-daemon-sysvinit git-doc git-el git-email git-gui gitk
  gitweb git-cvs git-mediawiki git-svn
The following NEW packages will be installed:
  git git-man liberror-perl
0 upgraded, 3 newly installed, 0 to remove and 0 not upgraded.
Need to get 5,465 kB of archives.
After this operation, 38.4 MB of additional disk space will be used.
Get:1 http://in.archive.ubuntu.com/ubuntu focal/main amd64 liberror-perl all 0.17029-1 [26.5 kB]
Get:2 http://in.archive.ubuntu.com/ubuntu focal-updates/main amd64 git-man all 1:2.25.1-1ubuntu3.2 [884 kB]
9% [git-man 114 kB/884 kB 13%]
pdates/main amd64 git amd64 1:2.25.1-1ubuntu3.2 [4,554 kB]
Fetched 5,465 kB in 4min 3s (22.5 kB/s)
Selecting previously unselected package liberror-perl.
(Reading database ... 213052 files and directories currently installed.)
Preparing to unpack .../liberror-perl_0.17029-1_all.deb ...
Unpacking liberror-perl (0.17029-1) ...
Selecting previously unselected package git-man.
Preparing to unpack .../git-man_1x3a2.25.1-1ubuntu3.2_all.deb ...
Unpacking git-man (1:2.25.1-1ubuntu3.2) ...
Selecting previously unselected package git.
Preparing to unpack .../git_1x3a2.25.1-1ubuntu3.2_amd64.deb ...
Unpacking git (1:2.25.1-1ubuntu3.2) ...
Setting up liberror-perl (0.17029-1) ...
Setting up git-man (1:2.25.1-1ubuntu3.2) ...
Setting up git (1:2.25.1-1ubuntu3.2) ...
Processing triggers for man-db (2.9.1-1) ...
```

2. Download devstack from its repository into system:

```
cd devstack
```

```
stack@DALAB-4:~$ git clone https://github.com/openstack-dev/devstack.git
cloning into 'devstack'...
remote: Enumerating objects: 48499, done.
remote: Counting objects: 100% (1725/1725), done.
remote: Compressing objects: 100% (686/686), done.
remote: Total 48499 (delta 1177), reused 1444 (delta 1033), pack-reused 46774
Receiving objects: 100% (48499/48499), 15.48 MiB | 4.31 MiB/s, done.
Resolving deltas: 100% (33808/33808), done.
```

3. Download devstack setup configurations files for it. Need to navigate devstack folder by running:

```
 nano local.conf  
vi local.conf  
git clone https://
```

```
stack@DALAB-4:~$ cd devstack  
stack@DALAB-4:~/devstack$ nano local.conf  
stack@DALAB-4:~/devstack$ vi local.conf
```

4. Add following inside local.conf:

```
[local|localrc] # Password for KeyStone, Database, RabbitMQ and Service  
ADMIN_PASSWORD=StrongAdminSecret  
DATABASE_PASSWORD=$ADMIN_PASSWORD  
RABBIT_PASSWORD=$ADMIN_PASSWORD  
SERVICE_PASSWORD=$ADMIN_PASSWORD
```

After setting above, in local.conf, press Esc key and type :wq to write/save and quit

13. Run following to setup Openstack on system:

./stack.sh

After installation, terminal:

```
=====
DevStack Component Timing
(times are in seconds)
=====
wait_for_service      18
pip_install          280
apt-get              942
run_process           68
dbsync                66
git_timed             1034
apt-get-update        1
test_with_retry       51
async_wait            1564
osc                   283
-----
Unaccounted time     816
=====
Total runtime         5123

=====
Async summary
=====
Time spent in the background minus waits: 2015 sec
Elapsed time: 5123 sec
Time if we did everything serially: 7138 sec
Speedup:  1.39332

This is your host IP address: 192.168.112.197
This is your host IPv6 address: ::1
Horizon is now available at http://192.168.112.197/dashboard
Keystone is serving at http://192.168.112.197/identity/
The default users are: admin and demo
The password: StrongAdminSecret

Services are running under systemd unit files.
For more information see:
https://docs.openstack.org/devstack/latest/systemd.html

DevStack Version: zed
Change: 0ed70e3f7687ffa62a8a4a38cdad14abdc8c7fa7 Merge "Update DEVSTACK_SERIES to zed" 2022-03-28 13:02:03 +0000
OS Version: Ubuntu 20.04 focal

2022-03-29 15:26:41.917 | stack.sh completed in 5123 seconds.
```

14. Browse URL on browser:

<http://localhost/dashboard>



15. Enter credentials. Log in as admin using username ‘admin’ and password as given in local.conf file:

The screenshot shows the OpenStack dashboard interface. The URL in the address bar is `localhost/dashboard/project/images`. The top navigation bar includes the OpenStack logo, user name 'admin', and a dropdown menu. The left sidebar has a tree structure with 'Project' expanded, showing 'Compute' (selected), 'Key Pairs', 'Server Groups', 'Volumes', 'Network', 'Admin', and 'Identity'. Under 'Compute', 'Instances' and 'Images' are listed. The main content area is titled 'Images' and displays a table with one item. The table columns are: Name, Type, Status, Visibility, Protected, Disk Format, and Size. The single row shows 'cirros-0.5.2-x86_64-disk' as the Name, 'Image' as the Type, 'Active' as the Status, 'Public' as the Visibility, 'No' as the Protected status, 'QCOW2' as the Disk Format, and '15.55 MB' as the Size. There are 'Create Image' and 'Delete Image' buttons at the bottom right of the table.

Result:

Thus, OpenStack has been successfully installed.

Exp no :5 c
Date : 8.3.23

Creation of VMs in OpenStack

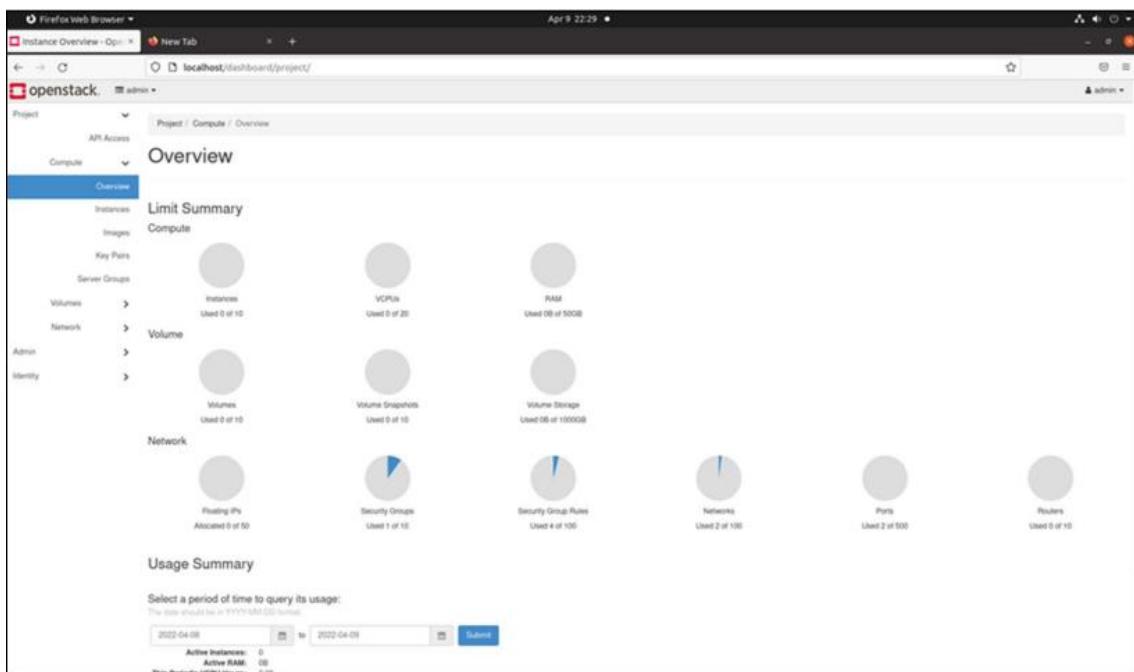
Aim:

To deploy a VM in OpenStack and execute a simple application on it.

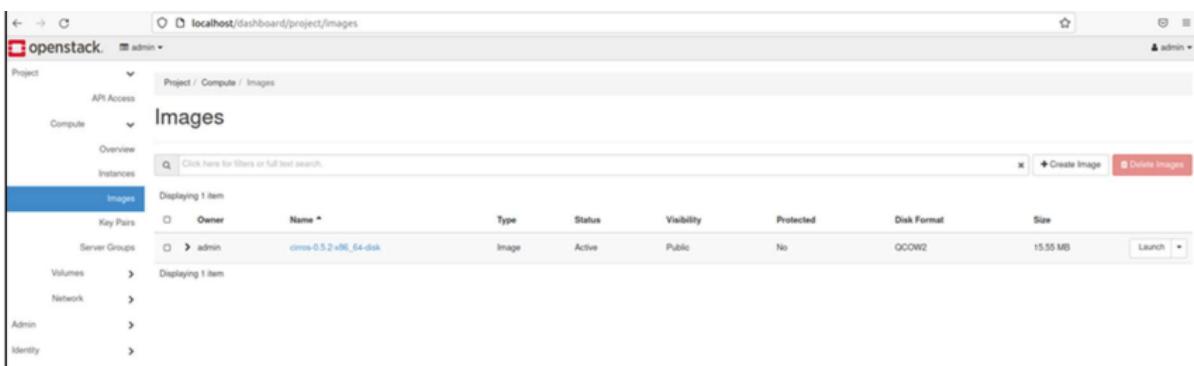
Procedure:

Deploying VM in OpenStack

1. Login to OpenStack using credential



2. Launch an Existing Image Instance in OpenStack



Once launch option is selected, pop-up window is displayed.

3. Give instance name of your choice. Leave the other fields as default values and click next

Launch Instance

Details

Please provide the initial hostname for the instance, the availability zone where it will be deployed, and the instance count. Increase the Count to create multiple instances with the same settings.

Source	Project Name	Total Instances (10 Max)
Flavor *	admin	 0 Current Usage 1 Added 9 Remaining
Networks *	Instance Name *	
Network Ports	Description	
Security Groups		
Key Pair	Availability Zone	
Configuration	nova	
Server Groups	Count *	
Scheduler Hints	1	
Metadata		

< Cancel **Next >** **Launch Instance**

4. Source tab will be next screen in the launch instance window where storage is created and click next

Launch Instance

Source

Instance source is the template used to create an instance. You can use an image, a snapshot of an instance (image snapshot), a volume or a volume snapshot (if enabled). You can also choose to use persistent storage by creating a new volume.

Flavor *	Select Boot Source	Create New Volume
Networks *	Image	Yes No
Network Ports	Volume Size (GB) *	Delete Volume on Instance Delete
Security Groups	1	Yes No
Key Pair	Allocated	
Configuration	Displaying 1 item	
Server Groups	Name Updated Size Format Visibility	
Scheduler Hints	cirros-0.5.2-x86_64-disk 4/5/22 10:37 PM 15.55 MB QCOW2 Public	
Metadata	Available	Select one
	<input type="text"/> Click here for filters or full text search.	
	Displaying 0 items	
	Name Updated Size Format Visibility	
	No items to display.	
	Displaying 0 items	

< Back **Next >** **Launch Instance**

5. Allocate VM resources by adding a flavour best suitable for your needs and click on Next

Launch Instance

Allocated						
Name	VCPUS	RAM	Total Disk	Root Disk	Ephemeral Disk	Public
cirros256	1	256 MB	1 GB	1 GB	0 GB	Yes

Available						
Name	VCPUS	RAM	Total Disk	Root Disk	Ephemeral Disk	Public
m1.nano	1	128 MB	1 GB	1 GB	0 GB	Yes
m1.micro	1	192 MB	1 GB	1 GB	0 GB	Yes
cirros256	1	256 MB	1 GB	1 GB	0 GB	Yes
m1.tiny	1	512 MB	1 GB	1 GB	0 GB	Yes
ds512M	1	512 MB	5 GB	5 GB	0 GB	Yes
ds1G	1	1 GB	10 GB	10 GB	0 GB	Yes
m1.small	1	2 GB	20 GB	20 GB	0 GB	Yes
ds2G	2	2 GB	10 GB	10 GB	0 GB	Yes
m1.medium	2	4 GB	40 GB	40 GB	0 GB	Yes
ds4G	4	4 GB	20 GB	20 GB	0 GB	Yes
m1.large	4	8 GB	80 GB	80 GB	0 GB	Yes
m1.xlarge	8	16 GB	160 GB	160 GB	0 GB	Yes

Cancel < Back Next > Launch Instance

6. From available instance, add instance which already launched so that the instance moves under allocated

Launch Instance

Allocated						
Name	VCPUS	RAM	Total Disk	Root Disk	Ephemeral Disk	Public
cirros256	1	256 MB	1 GB	1 GB	0 GB	Yes

Available						
Name	VCPUS	RAM	Total Disk	Root Disk	Ephemeral Disk	Public
m1.nano	1	128 MB	1 GB	1 GB	0 GB	Yes
m1.micro	1	192 MB	1 GB	1 GB	0 GB	Yes
cirros256	1	256 MB	1 GB	1 GB	0 GB	Yes
m1.tiny	1	512 MB	1 GB	1 GB	0 GB	Yes
ds512M	1	512 MB	5 GB	5 GB	0 GB	Yes
ds1G	1	1 GB	10 GB	10 GB	0 GB	Yes
m1.small	1	2 GB	20 GB	20 GB	0 GB	Yes
ds2G	2	2 GB	10 GB	10 GB	0 GB	Yes
m1.medium	2	4 GB	40 GB	40 GB	0 GB	Yes
ds4G	4	4 GB	20 GB	20 GB	0 GB	Yes
m1.large	4	8 GB	80 GB	80 GB	0 GB	Yes
m1.xlarge	8	16 GB	160 GB	160 GB	0 GB	Yes

Cancel < Back Next > Launch Instance

7. Finally, add shared network from available networks in OpenStack to your instance using upward arrow-mark button and hit on ‘Launch Instance’ to start the virtual machine

Launch Instance

Details Networks provide the communication channels for instances in the cloud.

Allocated Select networks from those listed below.

Network	Subnets Associated	Shared	Admin State	Status
1 shared	shared-subnet	Yes	Up	Active

Available Select at least one network

Network	Subnets Associated	Shared	Admin State	Status
public	public-subnet ipv6-public-subnet	No	Up	Active

< Back Next > Launch Instance

8. Once launch is clicked, select instance in right tab. The added instance is seen under it

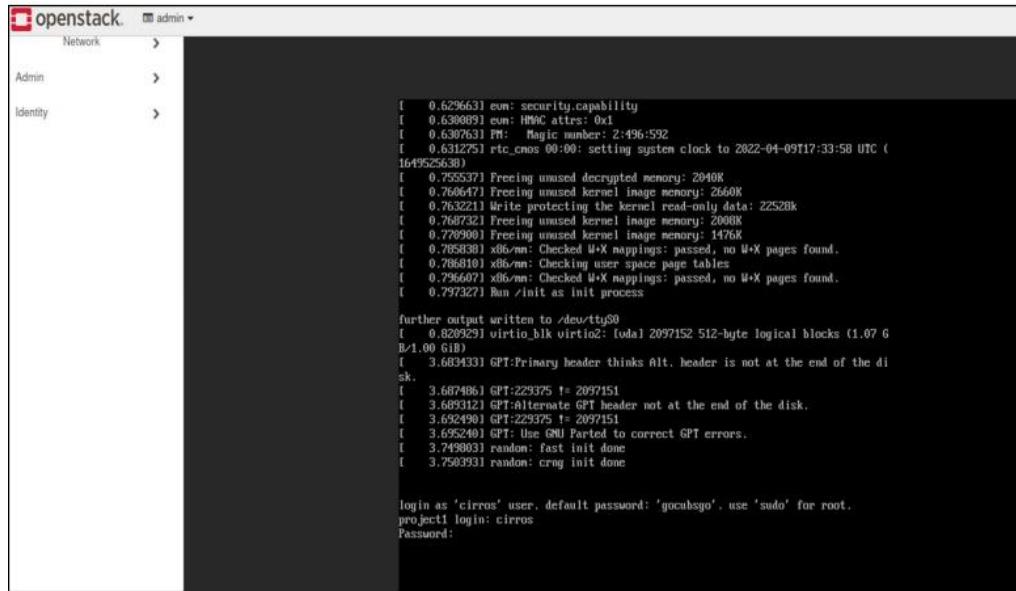
openstack. admin

Instances

Instance Name	Image Name	IP Address	Flavor	Key Pair	Status	Availability Zone	Task	Power State	Age	Actions
Project	-	192.168.233.104	unstable	-	Active	nova	None	Running	3 minutes	Create Snapshot

Associate Floating IP
Attach Interface
Detach Interface
Edit Instance
Attach Volume
Detach Volume
Update Metadata
Edit Security Groups
Edit Port Security Groups
Console
View Log
Rescue Instance
Pause Instance
Suspend Instance
Shelve Instance
Rescind Instance
Lock Instance
Soft Reboot Instance
Hard Reload Instance
Shut Off Instance
Relaunch Instance
Delete Instance

9. Click 'console'



The screenshot shows a terminal window titled 'Openstack' with the user 'admin'. The terminal displays a log of system bootup and initialization processes. Key logs include:

- [0.629663] evm: security.capability
- [0.630089] evm: HMC attrs: 0x1
- [0.630763] PM: Magic number: 2:496:592
- [0.631275] rtc_cmos 00:00: setting system clock to 2022-04-09T17:33:58 UTC (1649525630)
- [0.755537] Freeing unused decrypted memory: 2040K
- [0.760647] Freeing unused kernel image memory: 2660K
- [0.763221] Write protecting the kernel read-only data: 22520K
- [0.767321] Freeing unused kernel image memory: 2080K
- [0.779900] Freeing unused kernel image memory: 1476K
- [0.785381] x86-mm: Checked W+X mappings: passed, no W+X pages found.
- [0.786810] x86-mm: Checking user space page tables
- [0.790607] x86-mm: Checked W+X mappings: passed, no W+X pages found.
- [0.797327] Run /init as init process

further output written to /dev/ttys0
[0.820929] virtio_blk virtio2: [fda] 2097152 512-byte logical blocks (1.07 G B/1.00 GiB)
[3.680433] GPT:Primary header thinks Alt. header is not at the end of the disk.
[3.687486] GPT:229375 != 2097151
[3.689312] GPT:Alternate GPT header not at the end of the disk.
[3.692490] GPT:229375 != 2097151
[3.695240] GPT: Use GNU Parted to correct GPT errors.
[3.749803] random: fast init done
[3.750393] random: crng init done

login as 'cirros' user. default password: 'gocubsgo', use 'sudo' for root.
project1 login: cirros
Password:

10. Create a text file:

```
vi test.txt
```

11. In console type:

```
ls
```

12. Display text content using:

```
Cat test.txt
```

Result:

Thus, VM has been deployed in OpenStack and a simple application has been executed.

EXPT NO : 6
DATE :19/04/23

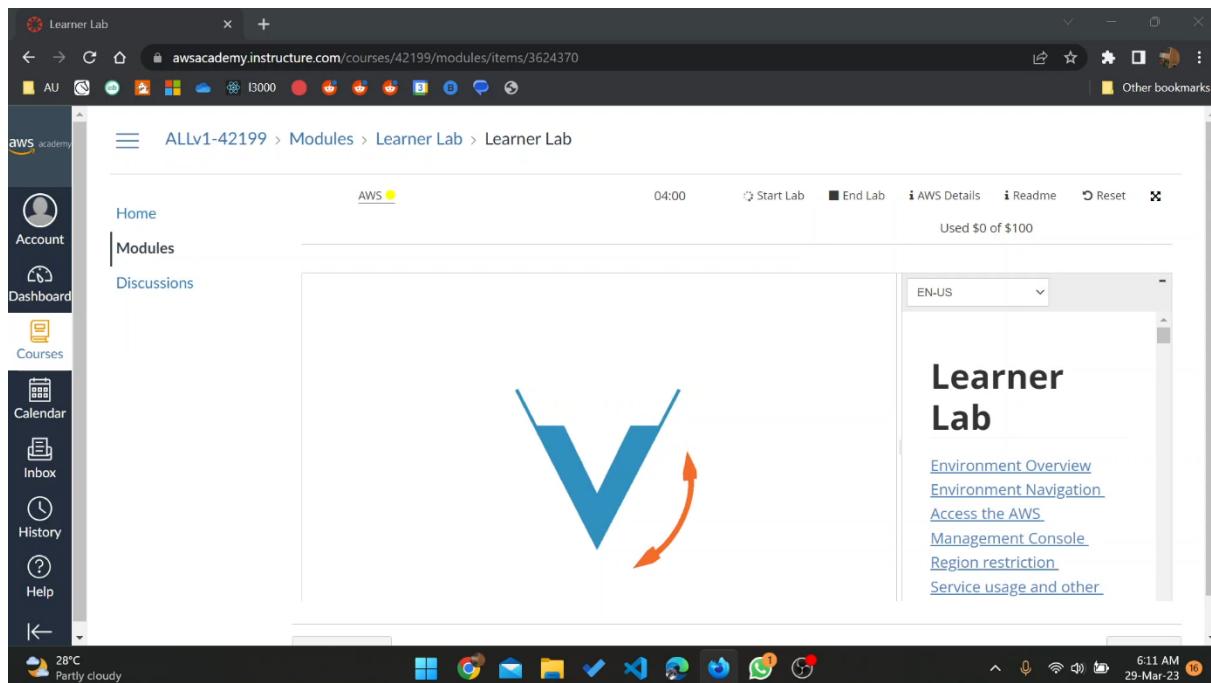
Creation of EC2 instance in AWS

AIM :

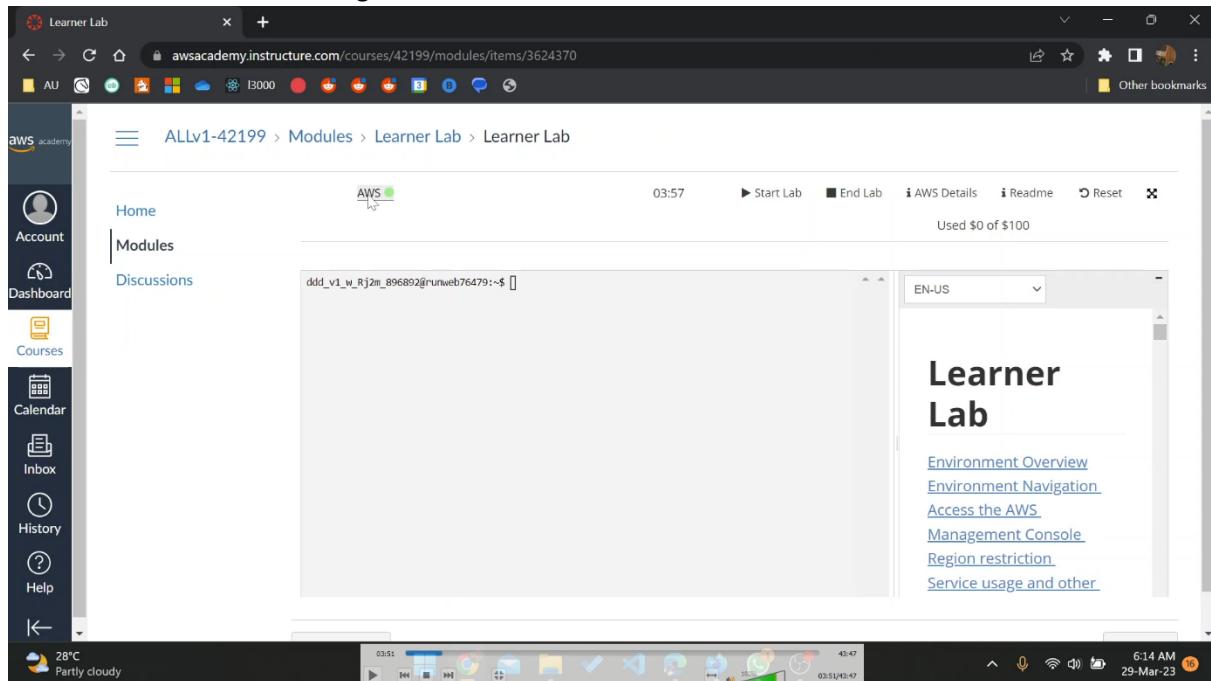
To create an EC2 instance in AWS.

PROCEDURE :

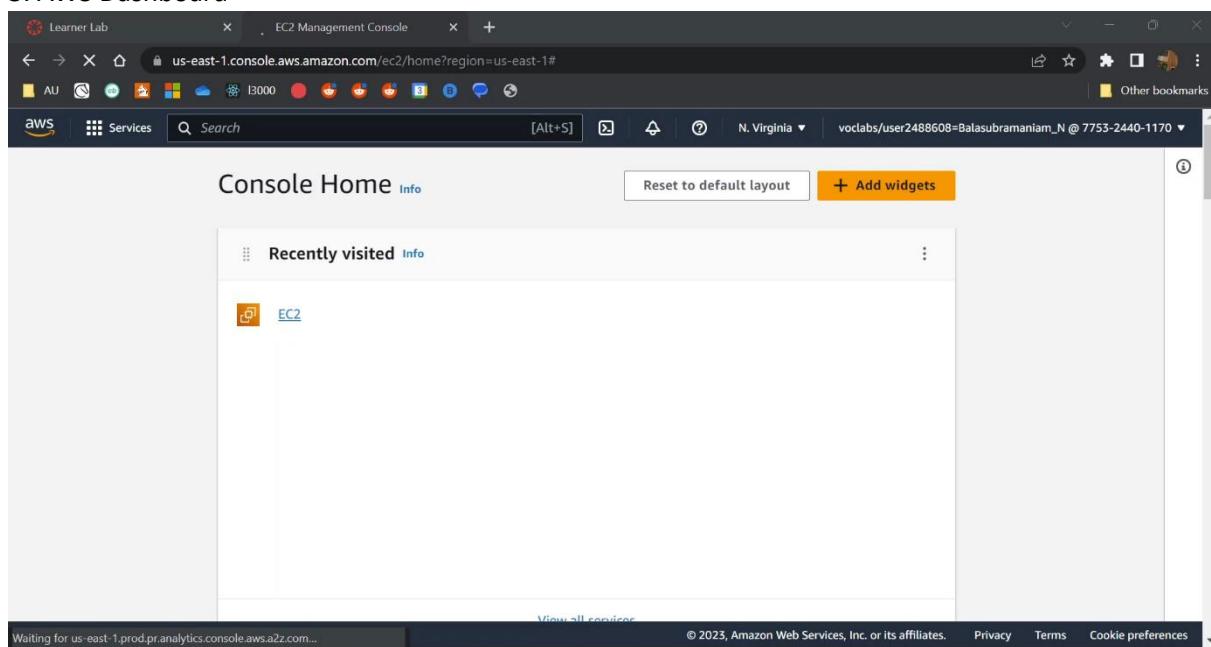
1. Login into AWS Canva Account



2. Now the terminal is running and is enabled for execution



3. AWS Dashboard



4. Instance Creation Page

The screenshot shows the 'Launch an instance' page in the AWS Management Console. The left sidebar lists 'Name and tags', 'Application and OS Images (Amazon Machine Image)', and 'Storage (volumes)'. The main area shows a 'Name and tags' section with a 'Name' field containing 'e.g. My Web Server'. Below it is a 'Virtual server type (instance type)' dropdown set to 't2.micro'. A 'Software Image (AMI)' dropdown is set to 'Amazon Linux 2023 AMI 2023.0.2...'. A 'Firewall (security group)' dropdown is set to 'New security group'. A 'Storage (volumes)' dropdown is set to 'Free tier eligible'. On the right, a 'Summary' panel shows 'Number of instances' as 1. At the bottom right are 'Cancel' and 'Launch instance' buttons.

5. Choosing the Operating System

The screenshot shows the 'Choose an Amazon Machine Image (AMI)' page. It features a search bar at the top and a 'Quick Start' section with icons for Amazon Linux, macOS, Ubuntu, Windows, and Red Hat. Below this is a grid of AMIs, with one highlighted: 'Microsoft Windows Server 2022 Base' (ami-0e38fa17744b2f6a5). This item is labeled as 'Free tier eligible'. To the right, there's a 'Browse more AMIs' link and a note about including AMIs from AWS Marketplace and the Community. The right side of the screen displays the same 'Summary' panel as the previous screenshot, showing 1 instance and the 'Launch instance' button.

6. Choosing the instance type

The screenshot shows the 'Launch an instance | EC2 Manager' wizard on the AWS console. The 'Instance type' section is selected, displaying the 't2.micro' option. This instance type is listed under the 'Free tier eligible' column. The 'AMI ID' field contains 'ami-0e38fa17744b2f6a5'. On the right side, the 'Summary' panel shows the number of instances set to 1, the software image as Microsoft Windows Server 2022, and the virtual server type as t2.micro. At the bottom right is a prominent orange 'Launch instance' button.

7. Creating a key pair

The screenshot shows the 'Create key pair' dialog box overlaid on the EC2 wizard. The dialog title is 'Create key pair'. It contains a descriptive text about key pairs and their purpose. Below it, there is a 'Key pair name - required' input field where 'keypair1' has been typed. Underneath, there is a note about character limits and spaces. A 'Private key file format' section contains two options: '.pem' (selected) and '.ppk'. The 'Create key pair' button is highlighted in orange at the bottom right of the dialog.

8. Choosing the network settings

The screenshot shows the 'Launch an instance | EC2 Manager' page. In the 'Security group' section, the 'Create security group' option is selected. A note says: 'We'll create a new security group called 'launch-wizard-1' with the following rules:'. Under 'Allow RDP traffic from Anywhere', the IP range is set to '0.0.0.0/0'. A warning message states: '⚠ Rules with source of 0.0.0.0/0 allow all IP addresses to access your instance. We recommend setting security group rules to allow access from known IP addresses only.' On the right, the 'Summary' panel shows 'Number of instances' as 1, 'Software Image (AMI)' as Microsoft Windows Server 2022, 'Virtual server type (instance type)' as t2.micro, and the 'Launch instance' button.

9. Selecting the File Storage type

The screenshot shows the 'Launch an instance | EC2 Manager' page. In the 'Storage' section, it shows '1x 30 GiB gp2 Root volume (Not encrypted)'. A note says: 'ⓘ Free tier eligible customers can get up to 30 GB of EBS General Purpose (SSD) or Magnetic storage'. Below this, there's a note: 'The selected AMI contains more instance store volumes than the instance allows. Only the first 0 instance store volumes from the AMI will be accessible from the instance'. The 'File systems' section shows '0 x File systems'. On the right, the 'Summary' panel shows 'Number of instances' as 1, 'Software Image (AMI)' as Microsoft Windows Server 2022, 'Virtual server type (instance type)' as t2.micro, and the 'Launch instance' button. A hand cursor is hovering over the 'Launch instance' button.

10. Instance Created

The screenshot shows the AWS EC2 Management Console with the URL us-east-1.console.aws.amazon.com/ec2/home?region=us-east-1#LaunchInstances. The main content area displays a green success message: "Successfully initiated launch of instance (i-005ee7f464f1b4830)". Below this, there is a link to "Launch log". A "Next Steps" section contains a search bar with the placeholder "What would you like to do next with this instance, for example "create alarm" or "create backup"" and a navigation bar with links for "Create billing and free tier usage alerts", "Connect to your instance", "Connect an RDS database", and "Create EBS snapshot policy". At the bottom, there are links for "Feedback", "Language", "keypair1.pem", and copyright information.

11. Instance Creation in Process

The screenshot shows the AWS EC2 Management Console with the URL us-east-1.console.aws.amazon.com/ec2/home?region=us-east-1#Instances. The left sidebar is expanded to show the "Instances" section, with "Instances" selected. The main content area displays a table titled "Instances (1) Info" with one row: "Instance 1" (Instance ID: i-005ee7f464f1b4830, Instance state: Pending, Instance type: t2.micro). Below the table, a modal window titled "Select an instance" is open, showing the same instance information. The bottom of the screen shows the standard AWS footer with links for "Feedback", "Language", "keypair1.pem", and copyright information.

12. Instance successfully created

The screenshot shows the AWS EC2 Management Console interface. The top navigation bar includes tabs for 'Learner Lab' and 'EC2 Management Console'. The URL is 'us-east-1.console.aws.amazon.com/ec2/home?region=us-east-1#Instances:'.

The main content area displays a table titled 'Instances (1) Info' with one row:

Name	Instance ID	Instance state	Instance type	Status check	Alarm status	Ava
Instance 1	i-005ee7f464f1b4830	Running	t2.micro	-	No alarms	us-e

Below the table, a modal window titled 'Select an instance' is open, showing the same instance information: 'Instance 1' (i-005ee7f464f1b4830), 'Running', 't2.micro', and 'No alarms'.

The left sidebar shows the navigation menu with 'Instances' selected under 'Services'.

RESULT :

Hence an EC2 instance has been successfully created using AWS Console, and the output was verified.

Aim:

To study about the basics of MongoDB and its operations.

Theory:

MongoDB is a general-purpose document database designed for modern application development and for the cloud. Its scale-out architecture allows you to meet the increasing demand for your system by adding more nodes to share the load.

MongoDB is a cross-platform, document oriented database that provides, high performance, high availability, and easy scalability. MongoDB works on concept of collection and document.

Document

MongoDB stores data as JSON documents. The document data model maps naturally to objects in application code, making it simple for developers to learn and use. A document is a set of key-value pairs. Documents have dynamic schema. Dynamic schema means that documents in the same collection do not need to have the same set of fields or structure, and common fields in a collection's documents may hold different types of data. Documents can be nested to express hierarchical relationships and to store structures such as arrays.

Collection

Collection is a group of MongoDB documents. It is the equivalent of an RDBMS table. A collection exists within a single database. Collections do not enforce a schema. Documents within a collection can have different fields. Typically, all documents in a collection are of similar or related purpose.

`_id` is a 12 bytes hexadecimal number which assures the uniqueness of every document. The first 4 bytes for the current timestamp, next 3 bytes for machine id, next 2 bytes for process id of MongoDB server and remaining 3 bytes are simple incremental VALUE.

Sample Document

```
{  
    '_id': ObjectId('7df78ad8902c'),  
    'Name': 'ABC',  
    'Age': 20,  
    'Reg_No': '201901a',  
    'Subjects': [ 'Math', 'English' ],  
    'Details': { 'Father': 'XY',  
                'Mother': 'PQ'  
              }  
}
```

Operations

1. Use

- To create and use a database. If the database exists, it returns existing database
- Syntax:
 - use Database_name

2. Create

- To create/insert a new document to a collection
- Syntax:
 - db.collection_name.insertOne()
 - db.collection_name.insertMany()

3. Read

- To retrieve documents from a collection
- Syntax:
 - db.collection_name.find()

4. Update

- To modify existing documents in a collection
- Syntax:
 - db.collection_name.updateOne()
 - db.collection_name.updateMany()
 - db.collection_name.replaceOne()

5. Delete

- To remove documents from a collection
- Syntax:
 - db.collection_name.deleteOne()
 - db.collection_name.deleteMany()

Result:

Thus, a study on the basics of MongoDB and its operations has been made.

2. Pick any system as an example, create database and its corresponding tables. Also populate the tables with data. Apply insert /search operations

Aim:

To design a student management system. This system contains: name, age and 12th marks.

Code:

- i. Create and Use Database ‘lab’

```
use lab
```

Output:

```
> use lab
switched to db lab
> show dbs
admin          0.000GB
blogDB         0.000GB
blogsDB        0.000GB
config          0.000GB
hirechain       0.000GB
library1        0.000GB
local           0.000GB
poolcarz        0.000GB
projectmanager  0.000GB
studentDB        0.000GB
tenderAuction    0.000GB
todolistdb       0.000GB
userDB           0.000GB
wikiDB          0.000GB
```

ii. Insert

- Insert one document into collection

```
db.demo.insertOne({"Name": "Bala Subramaniam", "Age":20, "Marks":482})
```

Output:

```
> db.demo.insertOne({"Name": "Bala Subramaniam", "Age":20, "Marks":482})  
{  
    "acknowledged" : true,  
    "insertedId" : ObjectId("6446c73d07631c44291c2885")  
}  
>
```

- Find all the documents

```
db.demo.find()
```

Output:

```
> db.demo.find()  
{ "_id" : ObjectId("6446c73d07631c44291c2885"), "Name" : "Bala Subramaniam", "Age" : 20, "Marks" : 482 }  
{ "_id" : ObjectId("6446c79a07631c44291c2886"), "Name" : "Jawahar A S", "Age" : 21, "Marks" : 483 }  
{ "_id" : ObjectId("6446c79b07631c44291c2887"), "Name" : "Sarukesh S", "Age" : 19, "Marks" : 484 }
```

iii. Search

- Find document where ‘Marks’ is 482

```
db.demo.find({"Marks":482})
```

Output:

```
> db.demo.find({"Marks":482})  
{ "_id" : ObjectId("6446c73d07631c44291c2885"), "Name" : "Bala Subramaniam", "Age" : 20, "Marks" : 482 }  
> -
```

iv. Update

Update the age as 23 if the present age is 21.

```
db.demo.updateMany({"Age": {$gt: 21}}, {$set: {"Age": 23}})
```

Output:

```

> db.demo.find()
{
  "_id" : ObjectId("6446c73d07631c44291c2885"),
  "Name" : "Bala Subramaniam",
  "Age" : 20,
  "Marks" : 482
}
{
  "_id" : ObjectId("6446c79a07631c44291c2886"),
  "Name" : "Jawahar A S",
  "Age" : 21,
  "Marks" : 483
}
{
  "_id" : ObjectId("6446c79b07631c44291c2887"),
  "Name" : "Sarukesh S",
  "Age" : 19,
  "Marks" : 484
}
> db.demo.updateMany({ "Age" : {$gt: 20} }, { $set : { 'Age' : 23 } })
{
  "acknowledged" : true,
  "matchedCount" : 1,
  "modifiedCount" : 1
}
> db.demo.find()
{
  "_id" : ObjectId("6446c73d07631c44291c2885"),
  "Name" : "Bala Subramaniam",
  "Age" : 20,
  "Marks" : 482
}
{
  "_id" : ObjectId("6446c79a07631c44291c2886"),
  "Name" : "Jawahar A S",
  "Age" : 23,
  "Marks" : 483
}
{
  "_id" : ObjectId("6446c79b07631c44291c2887"),
  "Name" : "Sarukesh S",
  "Age" : 19,
  "Marks" : 484
}
>

```

v. Delete

Delete all documents that have age greater than 19.

```
db.demo.deleteMany({ "Age" : {$gt : 19} })
```

Output:

```

> db.demo.find()
{
  "_id" : ObjectId("6446c73d07631c44291c2885"),
  "Name" : "Bala Subramaniam",
  "Age" : 20,
  "Marks" : 482
}
{
  "_id" : ObjectId("6446c79a07631c44291c2886"),
  "Name" : "Jawahar A S",
  "Age" : 23,
  "Marks" : 483
}
{
  "_id" : ObjectId("6446c79b07631c44291c2887"),
  "Name" : "Sarukesh S",
  "Age" : 19,
  "Marks" : 484
}
> db.demo.deleteMany({ "Age" : {$gt : 20} })
uncaught exception: SyntaxError: illegal character :
@(shell):1:20
> db.demo.deleteMany({ "Age" : {$gt : 20} })
{
  "acknowledged" : true,
  "deletedCount" : 1
}
> db.demo.find()
{
  "_id" : ObjectId("6446c73d07631c44291c2885"),
  "Name" : "Bala Subramaniam",
  "Age" : 20,
  "Marks" : 482
}
{
  "_id" : ObjectId("6446c79b07631c44291c2887"),
  "Name" : "Sarukesh S",
  "Age" : 19,
  "Marks" : 484
}

```

Result:

A student management system has been designed using MongoDB and CRUD operations have been performed on it successfully.

Hadoop Installation

Aim:

To study about the architecture of Hadoop and its components.

Procedure:

1. Update system before installation:

```
sudo apt – update
```

```
itadmin@PRLAB-22:~$ sudo apt update
[sudo] password for itadmin:
Hit:1 http://in.archive.ubuntu.com/ubuntu focal InRelease
Get:2 http://in.archive.ubuntu.com/ubuntu focal-updates InRelease [114 kB]
Ign:3 https://repo.mongodb.org/apt/ubuntu focal/mongodb-org/5.0 InRelease
Get:4 http://security.ubuntu.com/ubuntu focal-security InRelease [114 kB]
Get:5 http://in.archive.ubuntu.com/ubuntu focal-backports InRelease [108 kB]
Get:6 http://in.archive.ubuntu.com/ubuntu focal-updates/main amd64 Packages [1,790 kB]
Get:7 https://repo.mongodb.org/apt/ubuntu focal/mongodb-org/5.0 Release [4,406 B]
Get:8 http://in.archive.ubuntu.com/ubuntu focal-updates/main i386 Packages [650 kB]
Get:9 http://in.archive.ubuntu.com/ubuntu focal-updates/main Translation-en [330 kB]
Get:10 http://security.ubuntu.com/ubuntu focal-security/main i386 Packages [432 kB]
Get:11 https://repo.mongodb.org/apt/ubuntu focal/mongodb-org/5.0 Release.gpg [801 B]
Get:12 http://in.archive.ubuntu.com/ubuntu focal-updates/main amd64 DEP-11 Metadata [278 kB]
Get:13 http://in.archive.ubuntu.com/ubuntu focal-updates/main amd64 c-n-f Metadata [15.1 kB]
Get:14 http://in.archive.ubuntu.com/ubuntu focal-updates/restricted amd64 Packages [975 kB]
Get:15 http://in.archive.ubuntu.com/ubuntu focal-updates/restricted Translation-en [139 kB]
Get:16 http://in.archive.ubuntu.com/ubuntu focal-updates/universe amd64 Packages [920 kB]
Get:17 http://in.archive.ubuntu.com/ubuntu focal-updates/universe i386 Packages [679 kB]
Get:18 http://in.archive.ubuntu.com/ubuntu focal-updates/universe amd64 DEP-11 Metadata [390 kB]
Get:19 http://in.archive.ubuntu.com/ubuntu focal-updates/universe amd64 c-n-f Metadata [20.6 kB]
Get:20 http://in.archive.ubuntu.com/ubuntu focal-updates/multiverse amd64 DEP-11 Metadata [944 B]
Get:21 http://in.archive.ubuntu.com/ubuntu focal-backports/main amd64 DEP-11 Metadata [9,588 B]
Get:22 http://security.ubuntu.com/ubuntu focal-security/main amd64 Packages [1,451 kB]
Get:23 http://in.archive.ubuntu.com/ubuntu focal-backports/universe amd64 DEP-11 Metadata [30.8 kB]
Get:24 https://repo.mongodb.org/apt/ubuntu focal/mongodb-org/5.0/multiverse arm64 Packages [13.4 kB]
Get:25 https://repo.mongodb.org/apt/ubuntu focal/mongodb-org/5.0/multiverse amd64 Packages [15.5 kB]
Get:26 http://security.ubuntu.com/ubuntu focal-security/main Translation-en [249 kB]
Get:27 http://security.ubuntu.com/ubuntu focal-security/main amd64 DEP-11 Metadata [40.6 kB]
Get:28 http://security.ubuntu.com/ubuntu focal-security/main amd64 c-n-f Metadata [10.1 kB]
Get:29 http://security.ubuntu.com/ubuntu focal-security/restricted amd64 Packages [911 kB]
Get:30 http://security.ubuntu.com/ubuntu focal-security/restricted Translation-en [130 kB]
Get:31 http://security.ubuntu.com/ubuntu focal-security/universe amd64 Packages [702 kB]
Get:32 http://security.ubuntu.com/ubuntu focal-security/universe i386 Packages [550 kB]
Get:33 http://security.ubuntu.com/ubuntu focal-security/universe amd64 DEP-11 Metadata [66.3 kB]
Get:34 http://security.ubuntu.com/ubuntu focal-security/universe amd64 c-n-f Metadata [14.4 kB]
Get:35 http://security.ubuntu.com/ubuntu focal-security/multiverse amd64 DEP-11 Metadata [2,464 B]
Fetched 11.2 MB in 4s (2,700 kB/s)
Reading package lists... Done
Building dependency tree
Reading state information... Done
123 packages can be upgraded. Run 'apt list --upgradable' to see them.
```

2. Install OpenJDK

```
sudo apt install openjdk-11-jdk
```

```
itadmin@PRLAB-22:~$ sudo apt install openjdk-8-jdk -y
Reading package lists... Done
Building dependency tree
Reading state information... Done
The following package was automatically installed and is no longer required:
libfwupdplugin1
Use 'sudo apt autoremove' to remove it.
The following additional packages will be installed:
ca-certificates-java fonts-dejavu-extra java-common libatk-wrapper-java libatk-wrapper-java-jni libICE-dev libpthread-stubs0-dev libSM-dev libXi-dev libXau-dev libxcb1-dev libxdmcp-dev libxt-dev
openjdk-8-jdk-headless openjdk-8-jre openjdk-8-jre-headless x11proto-core-dev x11proto-dev xorg-sgml-doctools xtrans-dev
Suggested packages:
default-jre libICE-doc libSM-doc libXi-doc libxau-doc openjdk-8-source visualvmicedtea-8-plugin fonts-ipafont-gothic fonts-ipafont-mincho fonts-wqy-microhei
fonts-wqy-zenhei
The following NEW packages will be installed:
ca-certificates-java fonts-dejavu-extra java-common libatk-wrapper-java libatk-wrapper-java-jni libICE-dev libpthread-stubs0-dev libSM-dev libXi-dev libXau-dev libxcb1-dev libxdmcp-dev libxt-dev
openjdk-8-jdk openjdk-8-jdk-headless openjdk-8-jre openjdk-8-jre-headless x11proto-core-dev x11proto-dev xorg-sgml-doctools xtrans-dev
0 upgraded, 0 newly installed, 0 to remove and 123 not upgraded.
Need to get 41.9 MB of archives.
After this operation, 162 MB of additional disk space will be used.
Get:1 http://in.archive.ubuntu.com/ubuntu focal/main amd64 java-common all 0.72 [6,816 B]
Get:2 http://in.archive.ubuntu.com/ubuntu focal-updates/universe amd64 openjdk-8-jre-headless amd64 8u312-b07-0ubuntu1-20.04 [28.2 MB]
Get:3 http://in.archive.ubuntu.com/ubuntu focal/main amd64 ca-certificates-java all 20190405ubuntu1 [12.2 kB]
Get:4 http://in.archive.ubuntu.com/ubuntu focal/main amd64 fonts-dejavu-extra all 2.37-1.1,953 kB
Get:5 http://in.archive.ubuntu.com/ubuntu focal/main amd64 libatk-wrapper-java-jni amd64 0.37.1-1 [1.1 kB]
Get:6 http://in.archive.ubuntu.com/ubuntu focal/main amd64 libatk-wrapper-java-jni amd64 0.37.1-1+45.1 kB
Get:7 http://in.archive.ubuntu.com/ubuntu focal/main amd64 xorg-sgml-doctools all 1:1.11.1 [12.9 kB]
Get:8 http://in.archive.ubuntu.com/ubuntu focal/main amd64 x11proto-dev all 2019.2-0ubuntu1 [594 kB]
Get:9 http://in.archive.ubuntu.com/ubuntu focal/main amd64 x11proto-core-dev all 2019.2-0ubuntu1 [2,620 B]
Get:10 http://in.archive.ubuntu.com/ubuntu focal/main amd64 libICE-dev amd64 2:1.0.10-0ubuntu1 [47.8 kB]
Get:11 http://in.archive.ubuntu.com/ubuntu focal/main amd64 libpthread-stubs0-dev amd64 0.4-1 [5,384 B]
Get:12 http://in.archive.ubuntu.com/ubuntu focal/main amd64 libSM-dev amd64 2:1.2.3-1 [17.0 kB]
```

3. Verify intalled Java version

```
java -version; javac -version
```

```
itadmin@PRLAB-22:~$ java -version; javac -version
openjdk version "1.8.0_312"
OpenJDK Runtime Environment (build 1.8.0_312-8u312-b07-0ubuntu1~20.04-b07)
OpenJDK 64-Bit Server VM (build 25.312-b07, mixed mode)
javac 1.8.0_312
```

4. Install OpenSSH Server

```
sudo apt install openssh-server openssh-client -y
```

```
itadmin@PRLAB-22:~$ sudo apt install openssh-server openssh-client -y
Reading package lists... Done
Building dependency tree
Reading state information... Done
The following package was automatically installed and is no longer required:
libfwupdplugin1
Use 'sudo apt autoremove' to remove it.
The following additional packages will be installed:
openssh-sftp-server
Suggested packages:
Kerberos libpam-ssh monkeysphere ssh-askpass molly-guard
The following packages will be upgraded:
openssh-client openssh-server openssh-sftp-server
3 upgraded, 0 newly installed, 0 to remove and 120 not upgraded.
Need to get 1.099 kB of archives.
After this operation, 0 B of additional disk space will be used.
Get:1 http://in.archive.ubuntu.com/ubuntu focal-updates/main amd64 openssh-sftp-server amd64 1:8.2p1-4ubuntu0.5 [51.5 kB]
Get:2 http://in.archive.ubuntu.com/ubuntu focal-updates/main amd64 openssh-server amd64 1:8.2p1-4ubuntu0.5 [377 kB]
Get:3 http://in.archive.ubuntu.com/ubuntu focal-updates/main amd64 openssh-client amd64 1:8.2p1-4ubuntu0.5 [671 kB]
Fetched 1,099 kB in 0s (2,271 kB/s)
Preconfiguring packages...
(Reading database ... 213340 files and directories currently installed.)
Preparing to unpack .../openssh_sftp-server_1%3a8.2p1-4ubuntu0.5_amd64.deb ...
Unpacking openssh_sftp-server (1:8.2p1-4ubuntu0.5) over (1:8.2p1-4ubuntu0.4) ...
Preparing to unpack .../openssh_server_1%3a8.2p1-4ubuntu0.5_amd64.deb ...
Unpacking openssh-server (1:8.2p1-4ubuntu0.5) over (1:8.2p1-4ubuntu0.4) ...
Preparing to unpack .../openssh-client_1%3a8.2p1-4ubuntu0.5_amd64.deb ...
Unpacking openssh-client (1:8.2p1-4ubuntu0.5) over (1:8.2p1-4ubuntu0.4) ...
Setting up openssh-client (1:8.2p1-4ubuntu0.5) ...
Setting up openssh-sftp-server (1:8.2p1-4ubuntu0.5) ...
Setting up openssh-server (1:8.2p1-4ubuntu0.5) ...
rescue-ssh.target is a disabled or a static unit, not starting it.
Processing triggers for systemd (245.4-4ubuntu3.15) ...
Processing triggers for man-db (2.9.1-1) ...
Processing triggers for ufw (0.36-6ubuntu1) ...
```

5. Create a Hadoop user

```
sudo adduser hdoop
```

```
itadmin@PRLAB-22:~$ sudo adduser hdoop
Adding user `hdoop' ...
Adding new group `hdoop' (1002) ...
Adding new user `hdoop' (1002) with group `hdoop' ...
Creating home directory `/home/hdoop' ...
Copying files from `/etc/skel' ...
New password:
Retype new password:
passwd: password updated successfully
Changing the user information for hdoop
Enter the new value, or press ENTER for the default
    Full Name []:
    Room Number []:
    Work Phone []:
    Home Phone []:
    Other []:
Is the information correct? [Y/n] y
```

6. Switch to Hadoop user

```
su - hdoop
```

```
itadmin@PRLAB-22:~$ su - hdoop
Password:
hdoop@PRLAB-22:~$
```

7. Generate an SSH key pair and define the location is is to be stored in

```
ssh-keygen -t rsa -P "" -f ~/.ssh/id_rsa
```

```
hdoop@PRLAB-22:~$ ssh-keygen -t rsa -P '' -f ~/.ssh/id_rsa
Generating public/private rsa key pair.
Created directory '/home/hdoop/.ssh'.
Your identification has been saved in /home/hdoop/.ssh/id_rsa
Your public key has been saved in /home/hdoop/.ssh/id_rsa.pub
The key fingerprint is:
SHA256:ML0nTmQjrrLU8f7Y+ogsbBvRvyK8hur4xvz6CTYUA4 hdoop@PRLAB-22
The key's randomart image is:
+---[RSA 3072]---+
| . . .
| + o * .
| ooo o . S .
| o==. o o o
| =B+. .
| *=0oo. o
| O&O+.o+..
+---[SHA256]---+
```

8. Use the cat command to store the public key as authorized_keys in the ssh directory

```
cat ~/.ssh/id_rsa.pub >> ~/.ssh/authorized_keys
```

9. Set the permissions for your user with the chmod command

```
chmod 0600 ~/.ssh/authorized_keys
```

```
[SHELL] hdoop@PRLAB-22:~$ cat ~/.ssh/id_rsa.pub >> ~/.ssh/authorized_keys  
hdoop@PRLAB-22:~$ chmod 0600 ~/.ssh/authorized_keys
```

10. Verify everything is set up correctly by using the hdoop user to SSH to localhost

```
ssh localhost
```

```
hdoop@PRLAB-22:~$ ssh localhost  
The authenticity of host 'localhost (127.0.0.1)' can't be established.  
ECDSA key fingerprint is SHA256:SaTAp+60cPT8NCkg14kfRqJ8pzC8qlN5mV2T5rgLflE.  
Are you sure you want to continue connecting (yes/no/[fingerprint])? yes  
Warning: Permanently added 'localhost' (ECDSA) to the list of known hosts.  
Welcome to Ubuntu 20.04.4 LTS (GNU/Linux 5.13.0-39-generic x86_64)  
  
* Documentation: https://help.ubuntu.com  
* Management: https://landscape.canonical.com  
* Support: https://ubuntu.com/advantage  
  
118 updates can be applied immediately.  
67 of these updates are standard security updates.  
To see these additional updates run: apt list --upgradable  
  
Your Hardware Enablement Stack (HWE) is supported until April 2025.  
  
The programs included with the Ubuntu system are free software;  
the exact distribution terms for each program are described in the  
individual files in /usr/share/doc/*copyright.  
  
Ubuntu comes with ABSOLUTELY NO WARRANTY, to the extent permitted by  
applicable law.
```

11. Visit the [official Apache Hadoop project page](#), and select the version of Hadoop you want to implement. Click the link to download locally in a machine or install the Hadoop framework in the system using wget command

```
wget https://dlcdn.apache.org/hadoop/common/hadoop-3.2.3/hadoop-3.2.3.tar.gz
```

```
hdoop@PRLAB-22:~$ wget https://dlcdn.apache.org/hadoop/common/hadoop-3.2.3/hadoop-3.2.3.tar.gz  
--2022-05-17 23:23:55-- https://dlcdn.apache.org/hadoop/common/hadoop-3.2.3/hadoop-3.2.3.tar.gz  
Resolving dlcdn.apache.org (dlcdn.apache.org)... 151.101.2.132, 2a04:4e42::844  
Connecting to dlcdn.apache.org (dlcdn.apache.org)|151.101.2.132|:443... connected.  
HTTP request sent, awaiting response... 200 OK  
Length: 492241961 (460M) [application/x-gzip]  
Saving to: 'hadoop-3.2.3.tar.gz'  
  
hadoop-3.2.3.tar.gz          29%[=====] 140.20M  561KB/s   eta 6m 57s  
  
hadoop-3.2.3.tar.gz          100%[=====] 469.44M  1.02MB/s   in 8m 22s  
2022-05-17 23:32:17 (958 KB/s) - 'hadoop-3.2.3.tar.gz' saved [492241961/492241961]
```

Extract the files to initiate the Hadoop installation.

```
tar xvzf hadoop-3.2.3.tar.gz
```

```
hadoop-3.2.3/share/doc/hadoop/hadoop-distcp/css/maven-base.css  
hadoop-3.2.3/share/doc/hadoop/hadoop-distcp/css/maven-theme.css  
hadoop-3.2.3/share/doc/hadoop/hadoop-distcp/css/site.css  
hadoop-3.2.3/share/doc/hadoop/hadoop-distcp/css/print.css  
hadoop-3.2.3/share/doc/hadoop/hadoop-distcp/dependency-analysis.html  
hadoop-3.2.3/lib/  
hadoop-3.2.3/lib/native/  
hadoop-3.2.3/lib/native/libhdfspp.so.0.1.0  
hadoop-3.2.3/lib/native/libhadooppipes.a  
hadoop-3.2.3/lib/native/libhdfs.so.0.0.0  
hadoop-3.2.3/lib/native/libhadooputils.a  
hadoop-3.2.3/lib/native/libhadoop.so  
hadoop-3.2.3/lib/native/libhdfspp.so  
hadoop-3.2.3/lib/native/libhadoop.so.1.0.0  
hadoop-3.2.3/lib/native/libnativetask.a  
hadoop-3.2.3/lib/native/examples/  
hadoop-3.2.3/lib/native/examples/wordcount-part  
hadoop-3.2.3/lib/native/examples/wordcount-nopipe  
hadoop-3.2.3/lib/native/examples/pipes-sort  
hadoop-3.2.3/lib/native/examples/wordcount-simple  
hadoop-3.2.3/lib/native/libhdfs.a  
hadoop-3.2.3/lib/native/libhdfs.so  
hadoop-3.2.3/lib/native/libhadoop.a  
hadoop-3.2.3/lib/native/libnativetask.so.1.0.0  
hadoop-3.2.3/lib/native/libnativetask.so  
hadoop-3.2.3/lib/native/libhdfspp.a  
hadoop-3.2.3/LICENSE.txt
```

12. Rename the extracted directory by executing the command

```
mv hadoop-3.2.3 hadoop
```

```
hadoop@PRLAB-22:~$ mv hadoop-3.2.3 hadoop
```

13. Configure Java environment variables for setting up Hadoop

```
dirname $(dirname $(readlink -f $(which java)))
```

```
hadoop@PRLAB-22:~$ dirname $(dirname $(readlink -f $(which java)))  
/usr/lib/jvm/java-8-openjdk-amd64/jre
```

14. Open the “~/.bashrc” file in “vi” text editor. Define the Hadoop environment variables.

```
vi ~/.bashrc
```

```
#Hadoop Related Options  
export JAVA_HOME=/usr/lib/jvm/java-8-openjdk-amd64  
export HADOOP_HOME=/home/hdoop/hadoop  
export HADOOP_INSTALL=$HADOOP_HOME  
export HADOOP_MAPRED_HOME=$HADOOP_HOME  
export HADOOP_COMMON_HOME=$HADOOP_HOME  
export HADOOP_HDFS_HOME=$HADOOP_HOME  
export YARN_HOME=$HADOOP_HOME  
export HADOOP_COMMON_LIB_NATIVE_DIR=$HADOOP_HOME/lib/native  
export PATH=$PATH:$HADOOP_HOME/sbin:$HADOOP_HOME/bin  
export HADOOP_OPTS="-Djava.library.path=$HADOOP_HOME/lib/native"
```

15. Apply the changes to the current running environment

```
source ~/.bashrc
```

```
hdoop@PRLAB-22:~$ vi ~/.bashrc  
hdoop@PRLAB-22:~$ source ~/.bashrc  
hdoop@PRLAB-22:~$
```

16. Add the full path to the OpenJDK installation on your system

```
export JAVA_HOME=/usr/lib/jvm/java-8-openjdk-amd64
```

```

hadoop@PRLAB-22:~$ ls
hadoop  hadoop-3.2.3.tar.gz  test.txt
hadoop@PRLAB-22:~$ cd hadoop
hadoop@PRLAB-22:~/hadoop$ ls
bin  include  libexec      NOTICE.txt  sbin
etc  lib       LICENSE.txt  README.txt  share
hadoop@PRLAB-22:~/hadoop$ cd etc
hadoop@PRLAB-22:~/hadoop/etc$ ls
hadoop
hadoop@PRLAB-22:~/hadoop/etc$ cd hadoop
hadoop@PRLAB-22:~/hadoop/etc/hadoop$ ls
capacity-scheduler.xml          kms-log4j.properties
configuration.xsl                kms-site.xml
container-executor.cfg           log4j.properties
core-site.xml                   mapred-env.cmd
hadoop-env.cmd                  mapred-env.sh
hadoop-env.sh                   mapred-queues.xml.template
hadoop-metrics2.properties      mapred-site.xml
hadoop-policy.xml               shellprofile.d
hadoop-user-functions.sh.example ssl-client.xml.example
hdfs-site.xml                   ssl-server.xml.example
httpfs-env.sh                   workers
httpfs-log4j.properties         yarn-env.cmd
httpfs-signature.secret         yarn-env.sh
httpfs-site.xml                 yarnservice-log4j.properties
kms-acls.xml                   yarn-site.xml
kms-env.sh
hadoop@PRLAB-22:~/hadoop/etc/hadoop$ vi hadoop-env.sh
hadoop@PRLAB-22:~/hadoop/etc/hadoop$
```

17. Configure Apache Hadoop on the Ubuntu system. For this, the step is to create two directories: datanode and namenode, inside the home directory of Hadoop.

```

mkdir -p ~/dfsdata/namenode
mkdir -p ~/dfsdata/datanode
```

18. Specify the URL for your NameNode, and the temporary directory Hadoop uses for the map and reduce process. Go to specific directory and open core-site.xml

```

hadoop@PRLAB-22:~/hadoop/etc/hadoop$ vi core-site.xml
```

Add the following configuration to override the default values for the temporary directory and add your HDFS URL to replace the default local file system setting

```

<property>
  <name>hadoop.tmp.dir</name>
  <value>/home/hadoop/tmpdata</value>
</property>
<property>

  <name>fs.default.name</name>
  <value>hdfs://127.0.0.1:9000</value>
</property>
```

19. Configure the file by defining the NameNode and DataNode storage directories in hdfs-site.xml

```
hadoop@PRLAB-22:~/hadoop/etc/hadoop$ vi hdfs-site.xml
```

Add the following configuration to the file and, if needed, adjust the NameNode and DataNode directories to your custom locations.

```
<configuration>
<property>
  <name>dfs.data.dir</name>
  <value>/home/hadoop/dfsdata/namenode</value>
</property>
<property>
  <name>dfs.data.dir</name>
  <value>/home/hadoop/dfsdata/datanode</value>
</property>
<property>
  <name>dfs.replication</name>
  <value>1</value>
</property></configuration>
```

20. Use the following command to access the mapred-site.xml file and define MapReduce values:

```
hadoop@PRLAB-22:~/hadoop/etc/hadoop$ vi mapred-site.xml
```

Add the following configuration to change the default MapReduce framework name value to yarn

```
<configuration>
<property>
  <name>mapreduce.framework.name</name>
  <value>yarn</value>
</property> </configuration>
```

21. Edit yarn-site.xml File

Open the yarn-site.xml file in a text editor:
nano \$HADOOP_HOME/etc/hadoop/yarn-site.xml

Append the following configuration to the file

```
hadoop@PRLAB-22:~/hadoop/etc/hadoop$ vi yarn-site.xml
```

22. Format HDFS NameNode

hdfs namenode -format

23. Start Hadoop Cluster

/start-dfs.sh

```
hadoop@PRLAB-22:~$ cd hadoop  
hadoop@PRLAB-22:~/hadoop$ ls  
bin etc include lib libexec LICENSE.txt logs NOTICE.txt README.txt sbin share  
hadoop@PRLAB-22:~/hadoop$ cd sbin  
hadoop@PRLAB-22:~/hadoop/sbin$ ./start-dfs.sh
```

```
| Starting namenodes on [localhost]
| Starting datanodes
| Starting secondary namenodes [PRLAB-22]
|
```

```
./start-yarn.sh
```

```
hadoop@PRLAB-22:~/hadoop/sbin$ ./start-yarn.sh
Starting resourcemanager
Starting nodemanagers
[...]
```

25. Check if all the daemons are active and running as Java processes

```
jps
```

```
hadoop@PRLAB-22:~/hadoop/sbin$ jps
36208 NodeManager
36627 Jps
35547 DataNode
36059 ResourceManager
35372 NameNode
35789 SecondaryNameNode
```

26. Access Hadoop UI from Browser

```
http://localhost:9870
```

The screenshot shows a web browser window titled "Namenode information" with the URL "localhost:9870/dfshealth.html#tab-overview". The page has a green header bar with tabs for "Hadoop", "Overview", "Datanodes", "Datanode Volume Failures", "Snapshot", "Startup Progress", and "Utilities". The main content area is titled "Overview 'localhost:9000' (active)". It displays several configuration details in a table:

Started:	Fri May 20 03:13:28 +0530 2022
Version:	3.2.3, rabe5358143720085498613d399be3bbf01e0f131
Compiled:	Sun Mar 20 06:48:00 +0530 2022 by ubuntu from branch-3.2.3
Cluster ID:	CID-2d693310-a1b5-41df-aa5e-6b2e0fa5e167
Block Pool ID:	BP-1597927012-127.0.1.1-1652996439092

Below this is a "Summary" section with various system statistics:

Configured Capacity:	186.74 GB
Configured Remote Capacity:	0 B
DFS Used:	24 KB (0%)
Non DFS Used:	9.54 GB
DFS Remaining:	167.65 GB (89.78%)
Block Pool Used:	24 KB (0%)

27. The default port 9864 is used to access individual DataNodes directly from browser

http://localhost:9864

The screenshot shows a web browser window with the title "DataNode Information" and the URL "localhost:9864/datanode.html". The page has a green header bar with tabs for "Hadoop", "Overview", and "Utilities".

DataNode on PRLAB-22:9866

Cluster ID:	CID-2d693310-a1b5-41df-aa5e-6b2e0fa5e167
Version:	3.2.3, rabe5358143720085498613d399be3bbf01e0f131

Block Pools

Namenode Address	Block Pool ID	Actor State	Last Heartbeat	Last Block Report	Last Block Report Size (Max Size)
localhost:9000	BP-1597927012-127.0.1.1-1652996439092	RUNNING	1s	3 minutes	0 B (64 MB)

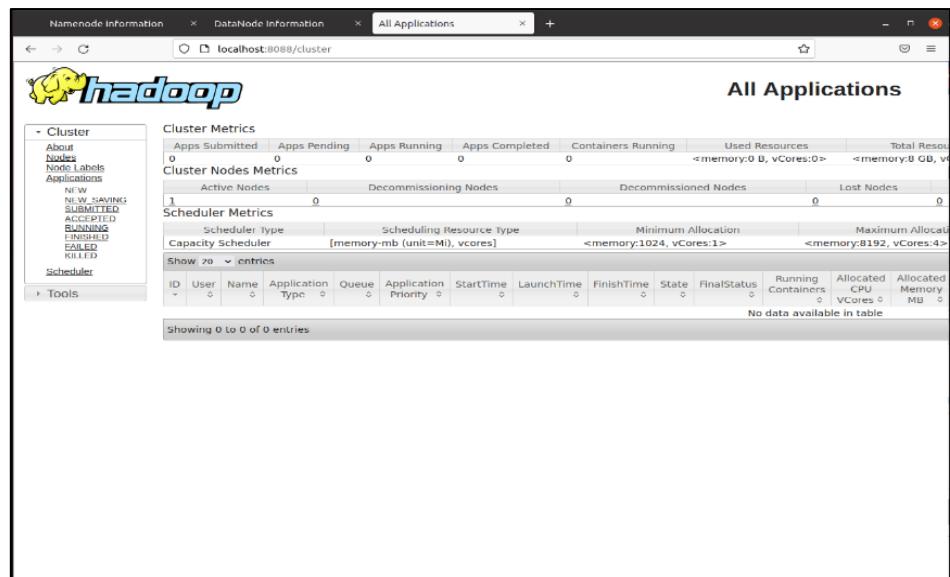
Volume Information

Directory	StorageType	Capacity Used	Capacity Left	Capacity Reserved	Reserved Space for Replicas	Blocks
/home/hadoop/dfsdata/datanode	DISK	24 KB	167.65 GB	0 B	0 B	0

Hadoop, 2022.

28. The YARN Resource Manager is accessible on port 8088. The Resource Manager is an invaluable tool that allows to monitor all running processes in Hadoop cluster

http://localhost:8088



29. Stop the NameNode, DataNode and YARN

```
./~stop-yarn.sh
```

```
hadoop@PRLAB-22:~/hadoop/sbin$ ./stop-yarn.sh
Stopping nodemanagers
Stopping resourcemanager
hadoop@PRLAB-22:~/hadoop/sbin$ ./stop-dfs.sh
Stopping namenodes on [localhost]
Stopping datanodes
Stopping secondary namenodes [PRLAB-22]
hadoop@PRLAB-22:~/hadoop/sbin$ jps
37827 Jps
```

Result: Thus, Hadoop has been successfully installed.

Aim: To implement file management in the Hadoop File System of Hadoop.

Codes and Output:

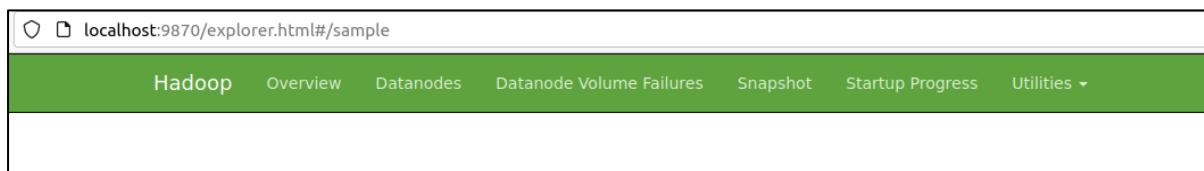
- a) Adding Directories

```
hadoop fs -mkdir /sample
```

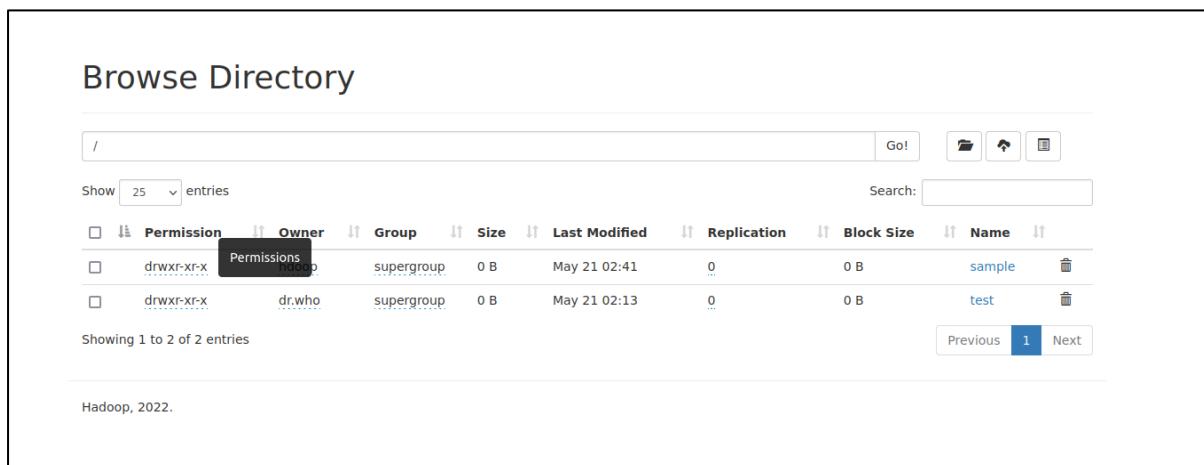
```
hadoop@PRLAB-22:~$ hadoop fs -mkdir /sample
```

Login to the web browser

```
http://localhost:9870
```



Under Utilities → click (Browse the file system option)



- b) Adding Files

Create a new file in the terminal

```
vi eg.txt
```

```
hadoop@PRLAB-22:~$ vi eg.txt
```

Type contents and save it and exit out of the file. Move the file to the newly created directory.

```
hadoop fs -put eg.txt /sample
```

The screenshot shows the Hadoop Web UI at localhost:9870/explorer.html#/sample. The page title is "Browse Directory". The URL in the address bar is /sample. The top navigation bar includes links for Hadoop, Overview, Datanodes, Datanode Volume Failures, Snapshot, Startup Progress, and Utilities. Below the navigation is a search bar with placeholder "Search:" and a "Go!" button. There are also icons for file operations like copy, move, and delete. A table lists one entry: "eg.txt" with details: Permission: -rw-r--r--, Owner: hdoop, Group: supergroup, Size: 18 B, Last Modified: May 21 02:46, Replication: 1, Block Size: 128 MB, Name: eg.txt. The table has columns for Permission, Owner, Group, Size, Last Modified, Replication, Block Size, Name, and a delete icon. Below the table, it says "Showing 1 to 1 of 1 entries". At the bottom left, it says "Hadoop, 2022.".

c) Retrieve File Contents

```
hadoop fs -cat /sample/eg.txt
```

```
hadoop@PRLAB-22:~$ hadoop fs -cat /sample/eg.txt
welcome to hadoop
```

d) Delete File

Create a new file and move to the directory 'sample'

```
vi hsampole.txt
hadoop fs -put hsampole.txt /sample
hadoop fs -cat /sample/hsampole.txt
```

```
hadoop@PRLAB-22:~$ vi hsampole.txt
hadoop@PRLAB-22:~$ hadoop fs -put hsampole.txt /sample
hadoop@PRLAB-22:~$ hadoop fs -cat /sample/hsampole.txt
Newly created file for hadoop
```

Browse Directory

/sample

File	Permission	Owner	Group	Size	Last Modified	Replication	Block Size	Name
eg.txt	-rw-r--r--	hdoop	supergroup	18 B	May 21 02:46	1	128 MB	eg.txt
hsample.txt	-rw-r--r--	hdoop	supergroup	30 B	May 21 02:54	1	128 MB	hsample.txt

Showing 1 to 2 of 2 entries

Hadoop, 2022.

```
hadoop fs -rm /sample/hsample.txt
```

```
hadoop@PRLAB-22:~$ hadoop fs -rm /sample/hsample.txt
Deleted /sample/hsample.txt
```

After deletion

Browse Directory

/sample

File	Permission	Owner	Group	Size	Last Modified	Replication	Block Size	Name
eg.txt	-rw-r--r--	hdoop	supergroup	18 B	May 21 02:46	1	128 MB	eg.txt

Showing 1 to 1 of 1 entries

Hadoop, 2022.

Result: Thus, file management in HDFS of Hadoop has been successfully carried out.

Aim:

To implement the concept of Hadoop multinode using multithreading

Theory:

Hadoop is a framework for distributed processing of large datasets across a cluster of computers. A Hadoop cluster consists of a single master node and multiple worker nodes. The master node manages the system and coordinates the distribution of work across the worker nodes, while the worker nodes process the data and run the tasks assigned to them by the master node.

In a multinode Hadoop cluster, data is stored across multiple worker nodes and processed in parallel to speed up the processing time. Data is divided into smaller chunks and replicated across several worker nodes for fault tolerance. Tasks are then distributed across the worker nodes and executed in parallel, allowing for faster processing of large datasets. To set up a Hadoop multinode cluster, you would need to install Hadoop on each worker node and configure the cluster settings to connect the nodes together. Once the cluster is set up, you can submit data processing jobs to the cluster and monitor their progress using the Hadoop web interface or command-line tools.

Code

```
import java.io.*;
import java.util.*;
public class WordCount {
    public static void main(String[] args) {
        List<WordCountThread> threads = new ArrayList<>();
        Map<String, Integer> wordCount = new HashMap<>();
        try (BufferedReader br = new BufferedReader(new FileReader("input.txt"))) {
            String line;
            StringBuilder paragraph = new StringBuilder();
            while ((line = br.readLine()) != null) {
                if (!line.isEmpty()) {
                    // Add the line to the current paragraph
                    paragraph.append(line).append(" ");
                } else if (paragraph.length() > 0) {
```

```

// Create a WordCountThread for the current paragraph and start the thread
final String text = paragraph.toString();
WordCountThread thread = new WordCountThread(text);
threads.add(thread);
thread.start();

// Reset the paragraph
paragraph.setLength(0);

}

}

if (paragraph.length() > 0) {
// Create a WordCountThread for the last paragraph and start the thread
final String text = paragraph.toString();
WordCountThread thread = new WordCountThread(text);
threads.add(thread);
thread.start();
}

} catch (IOException e) {
System.err.println("Error: " + e.getMessage());
System.exit(1);
}

// Wait for all the threads to finish and combine the word counts
for (WordCountThread thread : threads) {
try {
thread.join();

Map<String, Integer> partialWordCount = thread.getWordCount();
for (Map.Entry<String, Integer> entry : partialWordCount.entrySet()) {
String word = entry.getKey();
int count = entry.getValue();
wordCount.put(word, wordCount.getOrDefault(word, 0) + count);
}
}
}

```

```
        } catch (InterruptedException e) {
            System.err.println("Error: " + e.getMessage());
            System.exit(1);
        }
    }

    // Print the total number of words and word count
    System.out.println("Total words: " + wordCount.size());
    for (Map.Entry<String, Integer> entry : wordCount.entrySet()) {
        System.out.println(entry.getKey() + ":" + entry.getValue());
    }
}

private static class WordCountThread extends Thread {
    private String text;
    private Map<String, Integer> wordCount;
    public WordCountThread(String text) {
        this.text = text;
    }
    public void run() {
        wordCount = new HashMap<>();
        String[] words = text.split("\\s+");
        for (String word : words) {
            wordCount.put(word, wordCount.getOrDefault(word, 0) + 1);
        }
    }
    public Map<String, Integer> getWordCount() {
        return wordCount;
    }
}
```

Output:

```
bala9@XPS13 MSYS /c/Storage/College/SEM_6/Data Analytics and Cloud Lab/code/exp08 (main)
$ java WordCount.java
Total words: 16
which: 1
a: 1
be: 1
data: 1
going: 1
testing: 1
for: 1
is: 2
data.: 1
used: 1
multithreading.: 1
file: 1
cotains: 1
This: 2
text: 1
to: 1
```

Result:

Thus multi threading has been implemented using a text file to count the words.

EXPT NO : 9

DATE :
19/04/23

VISUALIZATION TOOLS PYTHON

AIM :

To visually see the data present data set usg many visualization tools.

CODES AND OUTPUTS:

Matplotlib:

```
import pandas as pd
df = pd.read_csv('../datasets/surance.csv')
print(df)
```

```
   age   sex   bmi  children  smoker    region  charges
0   19  female  27.900       0    yes  shwest  16884.92400
1   18     male  33.770       1     no  sheast  1725.55230
2   28     male  33.000       3     no  sheast  4449.46200
3   33     male  22.705       0     no northwest  21984.47061
4   32     male  28.880       0     no northwest  3866.85520
...
1333  50     male  30.970       3     no northwest  10600.54830
1334  18  female  31.920       0     no northeast  2205.98080
1335  18  female  36.850       0     no  sheast  1629.83350
1336  21  female  25.800       0     no  shwest  2007.94500
1337  61  female  29.070       0    yes northwest  29141.36030
```

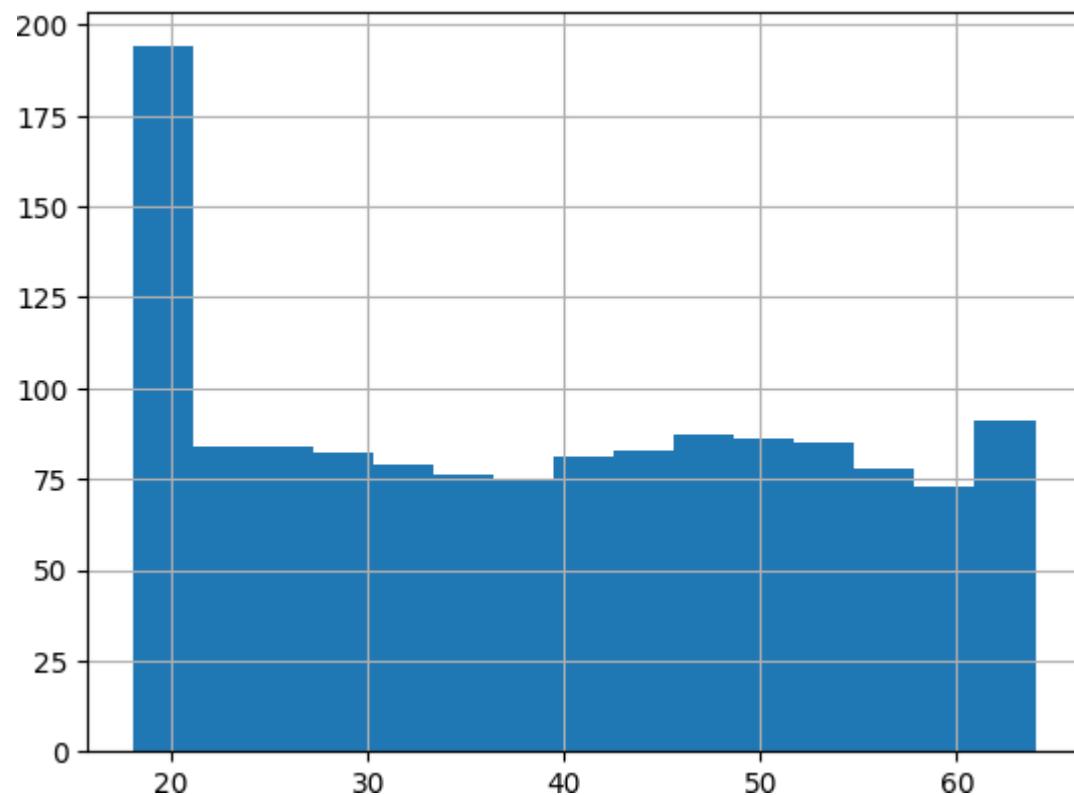
[1338 rows x 7 columns]

```
from matplotlib import pyplot as plt
import numpy as np
import pandas as pd
import matplotlib
import matplotlib.pyplot as plt
import matplotlib.image as mpimg
%matplotlib le
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Rangedex: 1338 entries, 0 to 1337
Data columns (total 7 columns):
 #   Column      Non-Null Count  Dtype  
--- 
 0   age         1338 non-null   int64  
 1   sex          1338 non-null   object  
 2   bmi          1338 non-null   float64 
 3   children     1338 non-null   int64  
 4   smoker        1338 non-null   object  
 5   region        1338 non-null   object  
 6   charges       1338 non-null   float64 
dtypes: float64(2), int64(2), object(3)
memory usage: 73.3+ KB
```

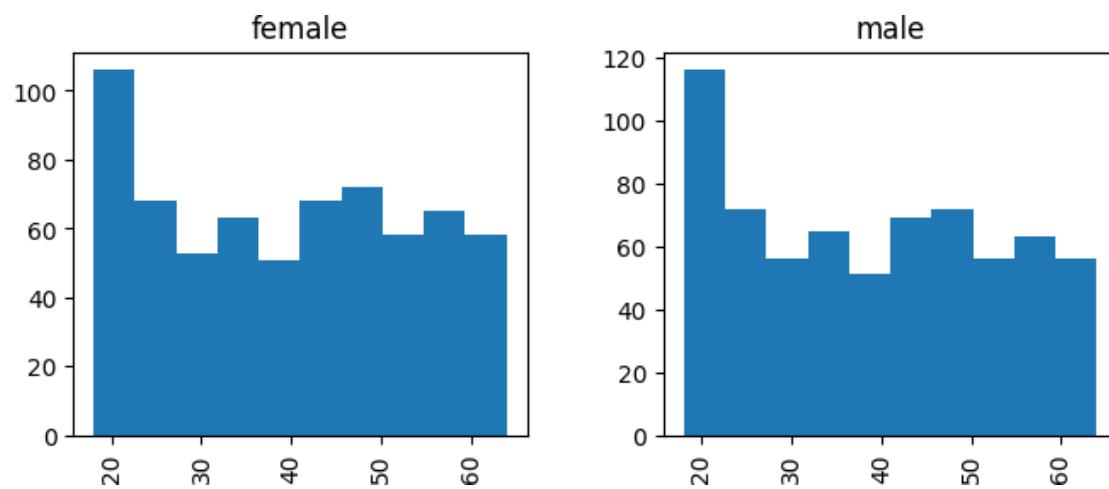
```
df[ 'age' ].hist(bs=15)
```

<AxesSubplot: >



```
df.hist(column='age', by='sex', lay=(1,3), figsize=(12,3), sharex=True)
```

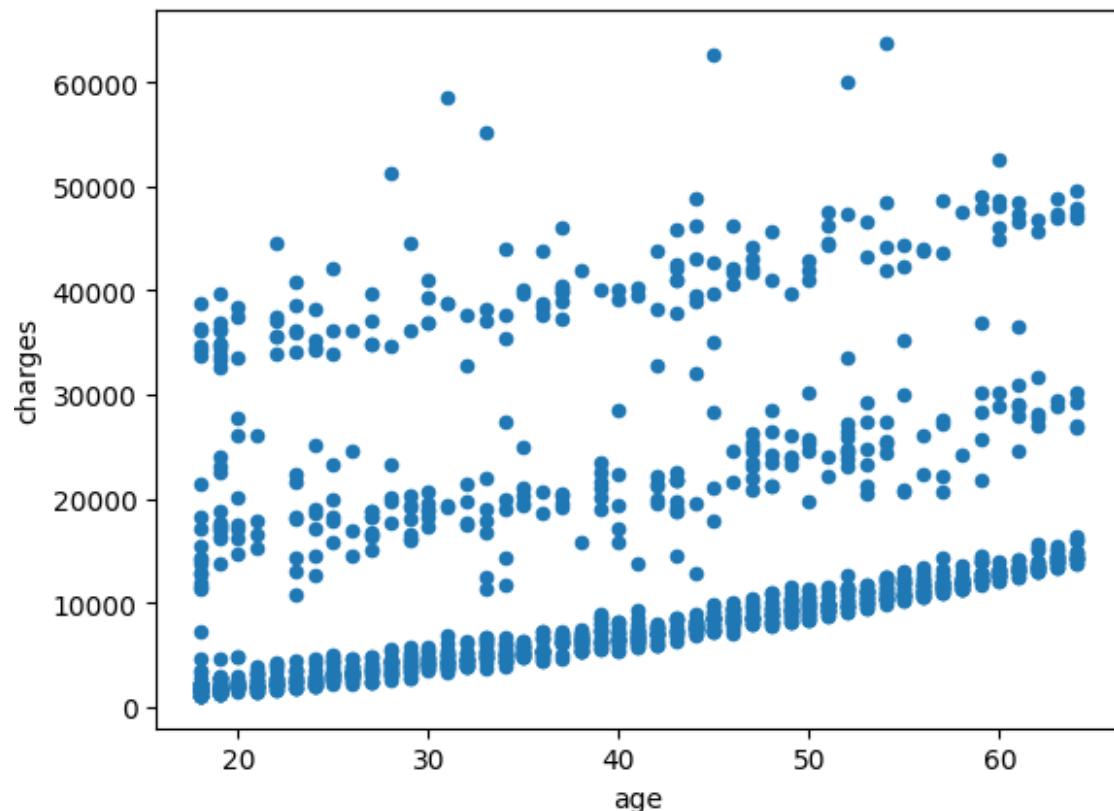
```
array([<AxesSubplot: title={'center': 'female'}>,
       <AxesSubplot: title={'center': 'male'}>, <AxesSubplot: >],
      dtype=object)
```



```
# Scatter plot
```

```
plt.scatter(df['age'],df['charges'])
```

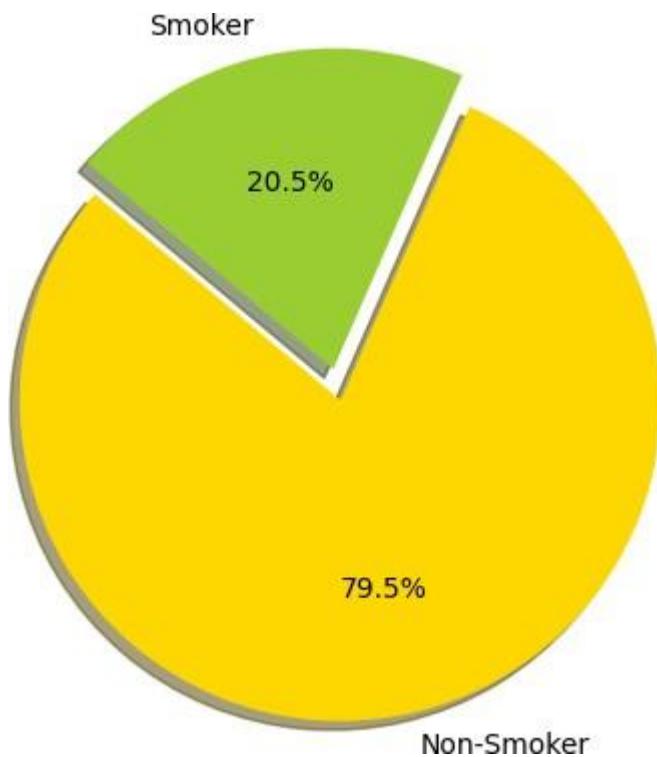
```
<AxesSubplot: xlabel='age', ylabel='charges'>
```



```
df1 = df['smoker'].value_counts()  
prt(df1)
```

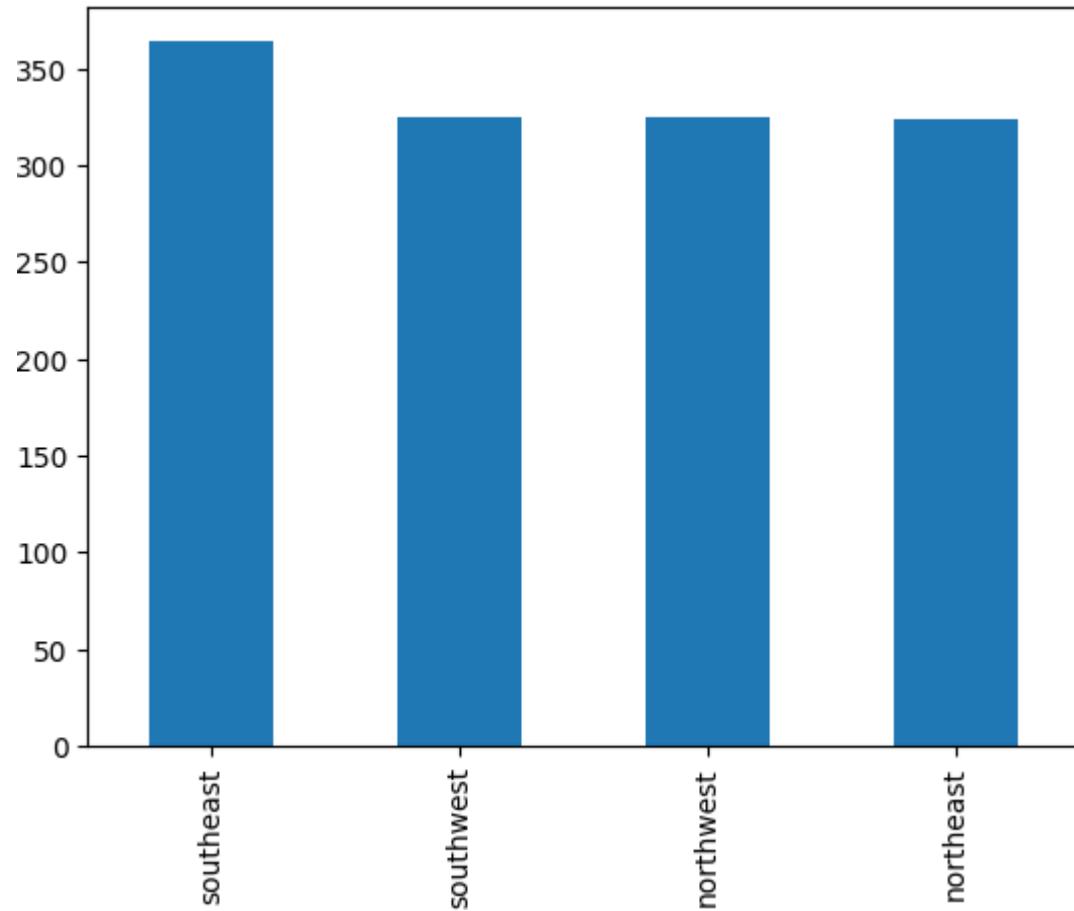
```
no      1064  
yes     274  
Name: smoker, dtype: int64
```

```
# pie chart
labels = 'Non-Smoker', 'Smoker'
sizes = [1064, 274]
colors = ['gold', 'yellowgreen']
explode = (0.1, 0) # explode 1st slice
plt.pie(sizes, explode=explode, labels=labels, colors=colors,
autopct='%1.1f%%', shadow=True, startangle=140)
plt.axis('equal')
plt.show()
```



```
# bar graph of region  
df1 = df['region'].value_counts()  
df1.plot(kd='bar')
```

<AxesSubplot: >



Seaborn

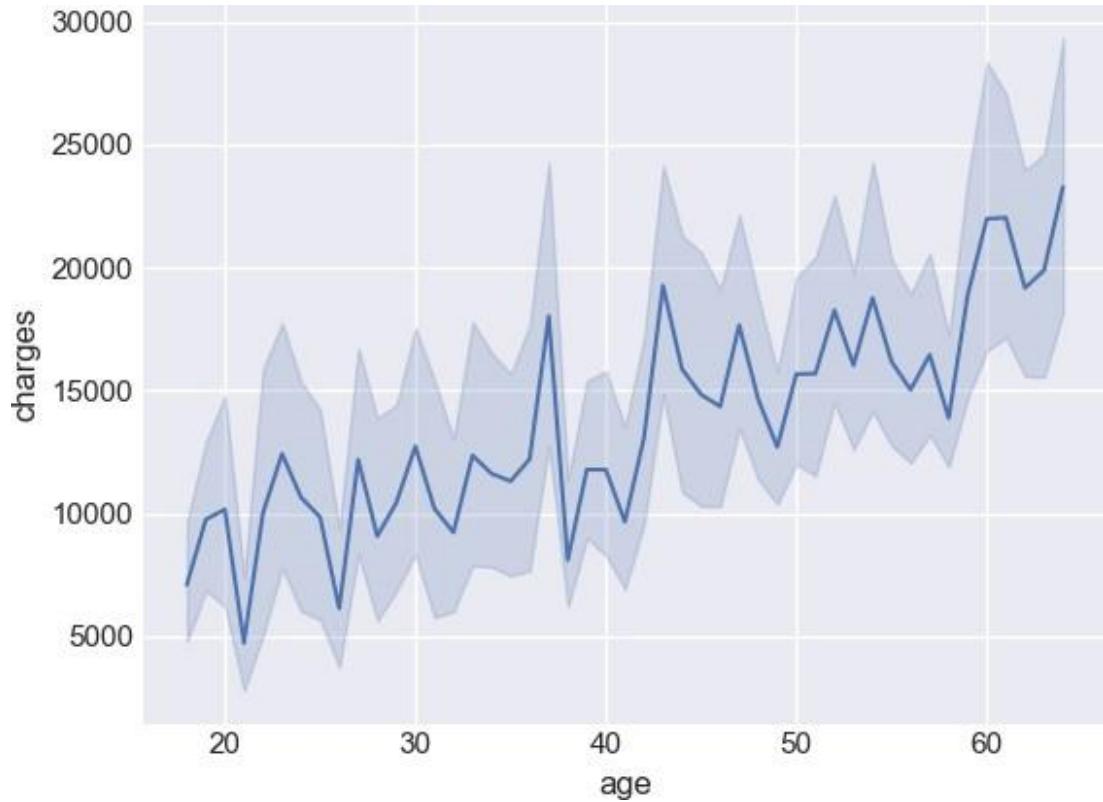
```
# import libraries
import numpy as np
import pandas as pd
import seaborn as sns
sns.set(color_codes=True)
import matplotlib.pyplot as plt
import matplotlib as mpl
import os
```

```
df = pd.read_csv('../datasets/surance.csv')
df.head()
```

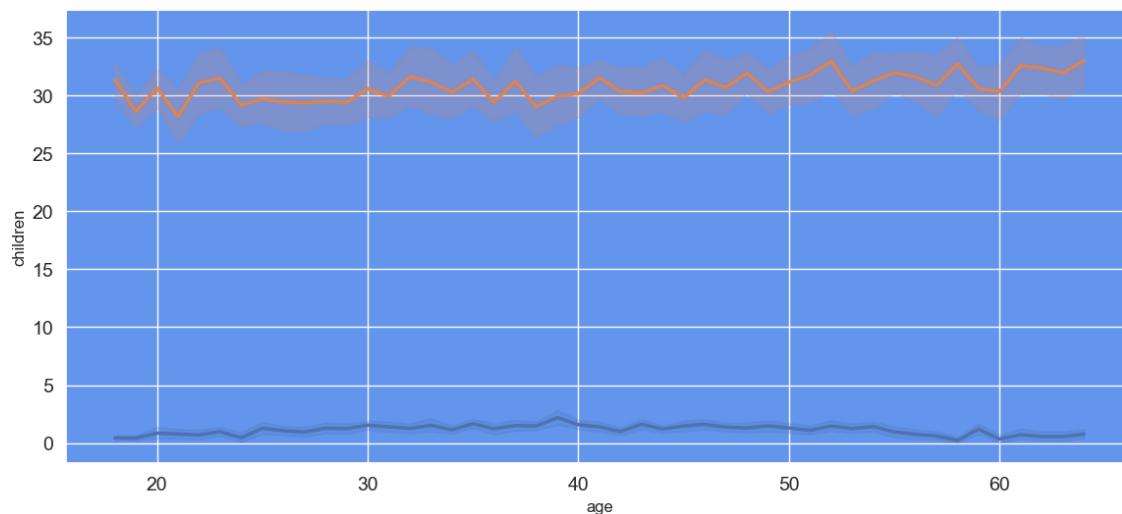
	age	sex	bmi	children	smoker	region	charges
0	19	female	27.900	0	yes	shwest	16884.92400
1	18	male	33.770	1	no	sheast	1725.55230
2	28	male	33.000	3	no	sheast	4449.46200
3	33	male	22.705	0	no	northwest	21984.47061
4	32	male	28.880	0	no	northwest	3866.85520

```
# chart
plt.style.use('seaborn-darkgrid')
%matplotlib le
sns.leplot(x='age',y='charges',data=df)
plt.show()
```

```
C:\Users\bala9\AppData\Local\Temp\ipykernel_12528\629716739.py: MatplotlibDeprecationWarning: The seaborn styles shipped by Matplotlib are deprecated since 3.6, as they no longer correspond to the styles shipped by seaborn. However, they will remain available as 'seaborn-v0_8-<style>'. Alternatively, directly use the seaborn API instead.
plt.style.use('seaborn-darkgrid')
```

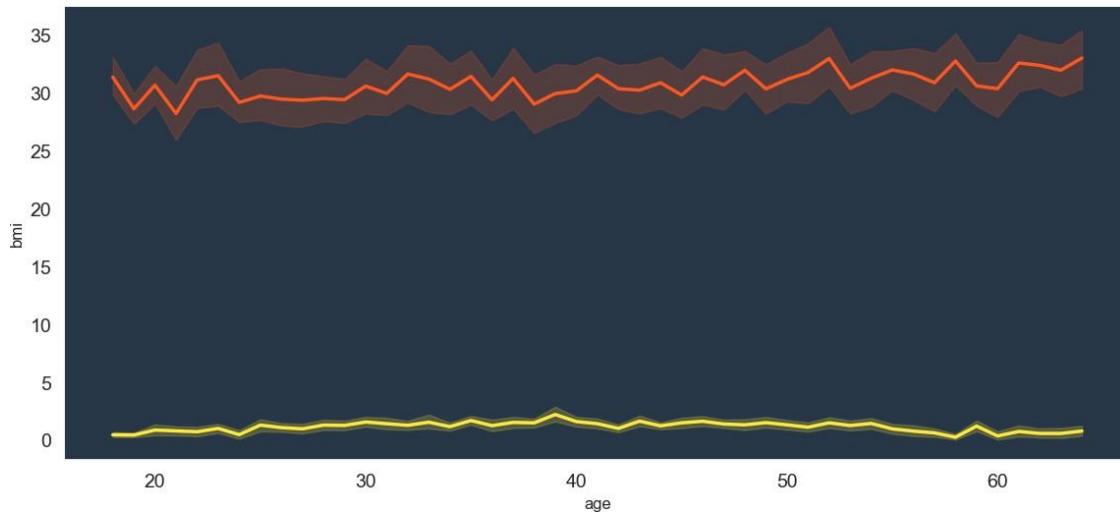


```
plt.figure(figsize=(14,6))
sns.set(rc={'axes.facecolor':'cornflowerblue','axes.grid":True,'xtick.labelsize':14,'ytic
sns.leplot(x=df.age,y=df.children,data=df , lewidth = 2.5)
sns.leplot(x=df.age,y=df.bmi,data=df, lewidth = 2.5)
plt.show()
```

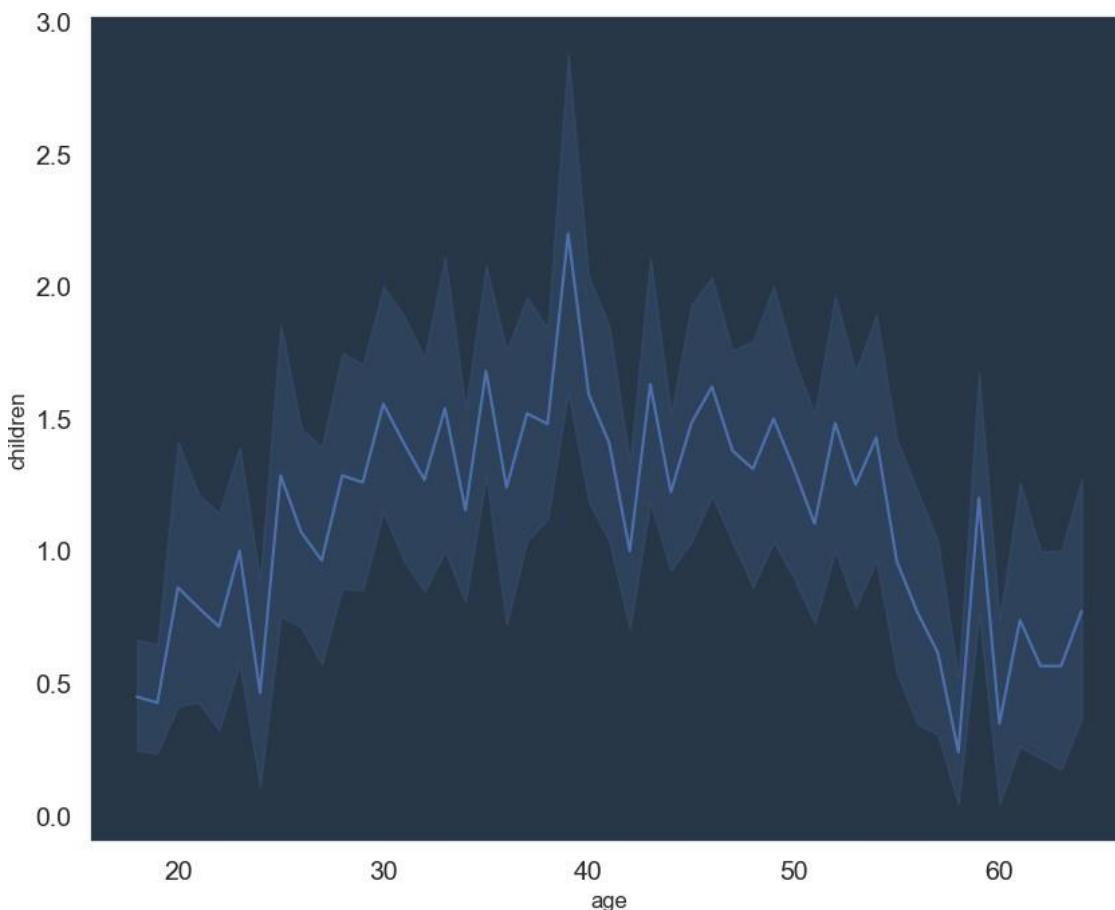


Line Plot

```
plt.figure(figsize=(14,6))
sns.set(rc={"axes.facecolor":"#283747", "axes.grid":False,'xtick.labelsize':14,'ytick.la
sns.leplot(x=df.age,y=df.bmi,data=df , color = "#FF5722",lewidth = 2.5)
sns.leplot(x=df.age,y=df.children,data=df,color = "#FFEB3B", lewidth = 2.5)
plt.show()
```

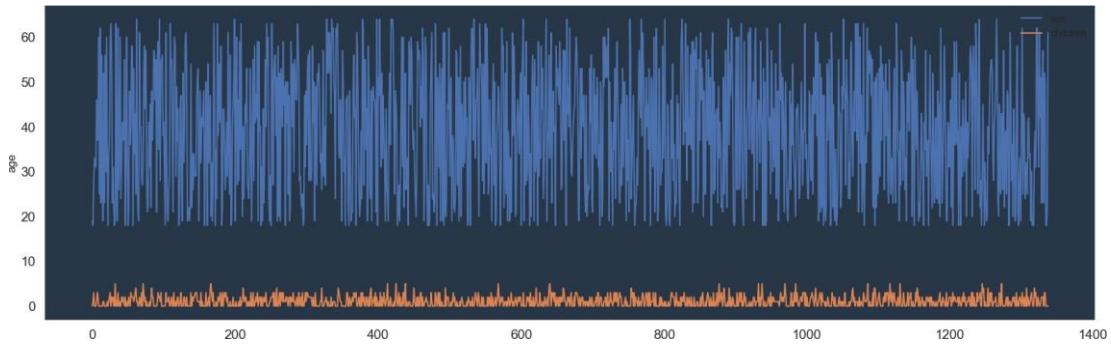


```
plt.figure(figsize=(10,8))
sns.leplot(x='age',y='children',data=df)
plt.show()
```

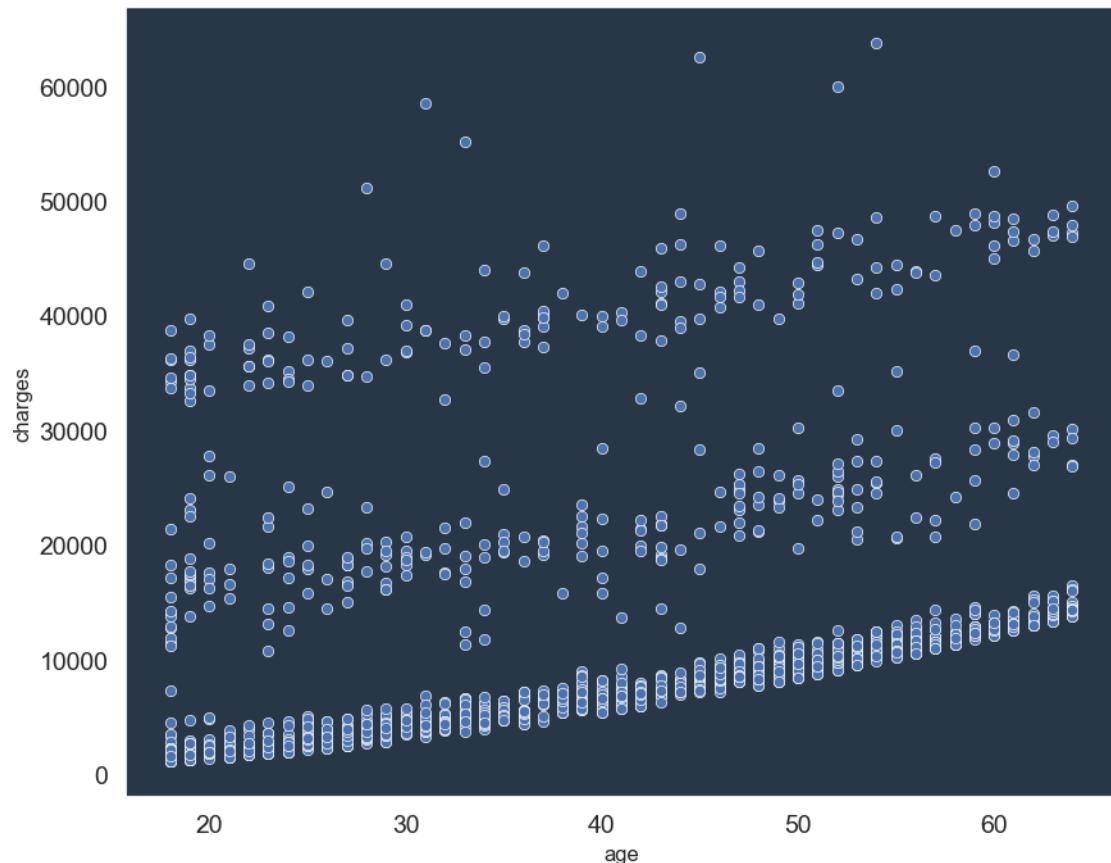


Code

```
plt.figure(figsize=(20,6))
sns.leplot(data=df['age'],lewidth = 1.5 , label = 'age')
sns.leplot(data=df['children'],lewidth = 1.5 , label = 'children')
plt.show()
```



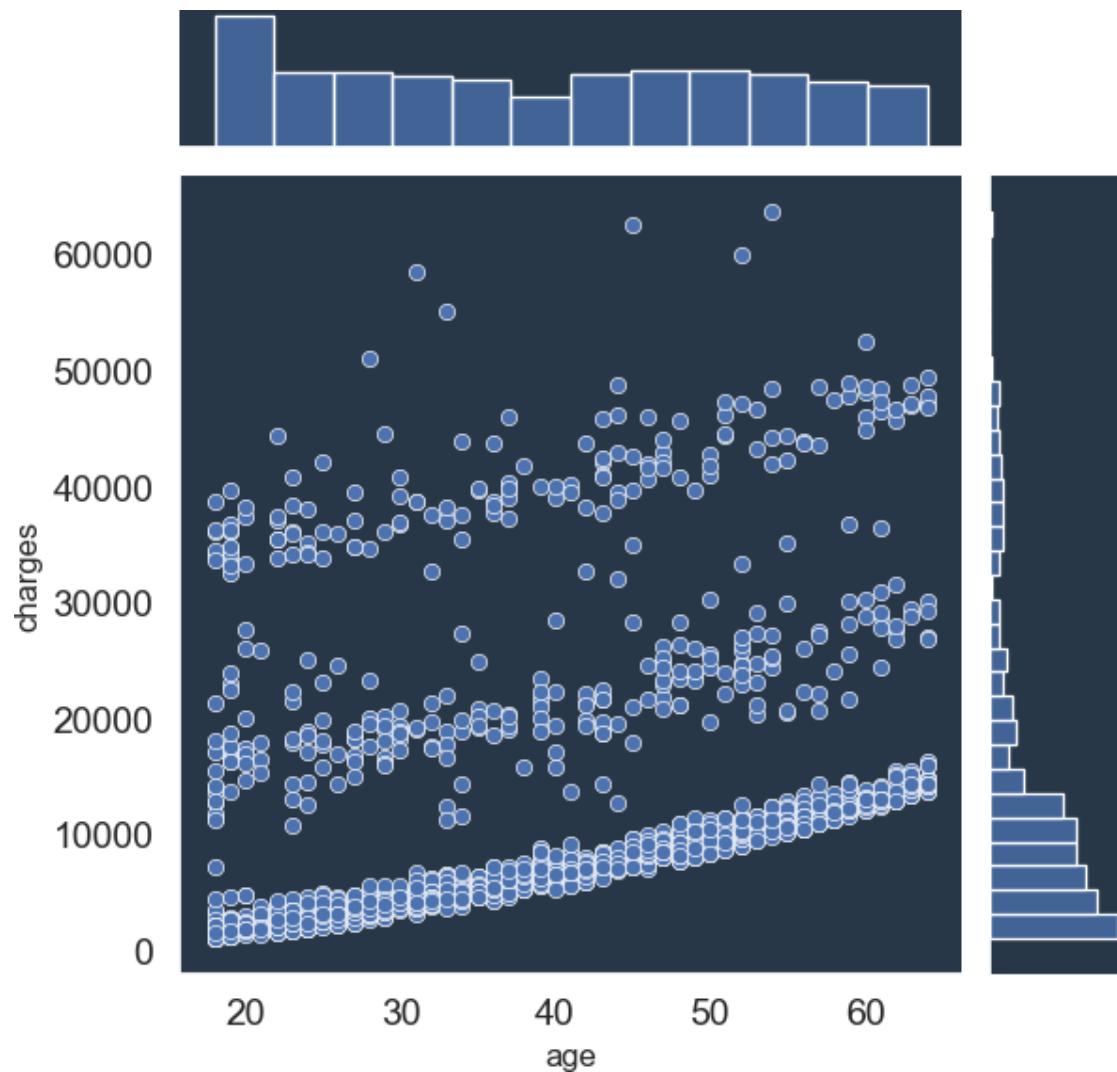
```
# scatter plot
plt.figure(figsize=(10,8))
sns.scatterplot(x='age',y='charges',data=df)
plt.show()
```



```
sns.jotplot(x='age',y='charges',data=df)
```

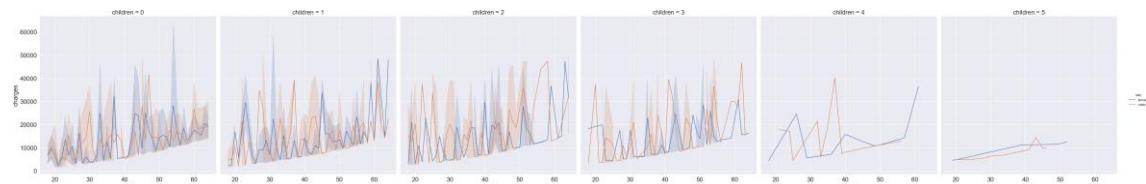
[31]:

```
<seaborn.axisgrid.JotGrid at 0x1796f69c1d0>
```

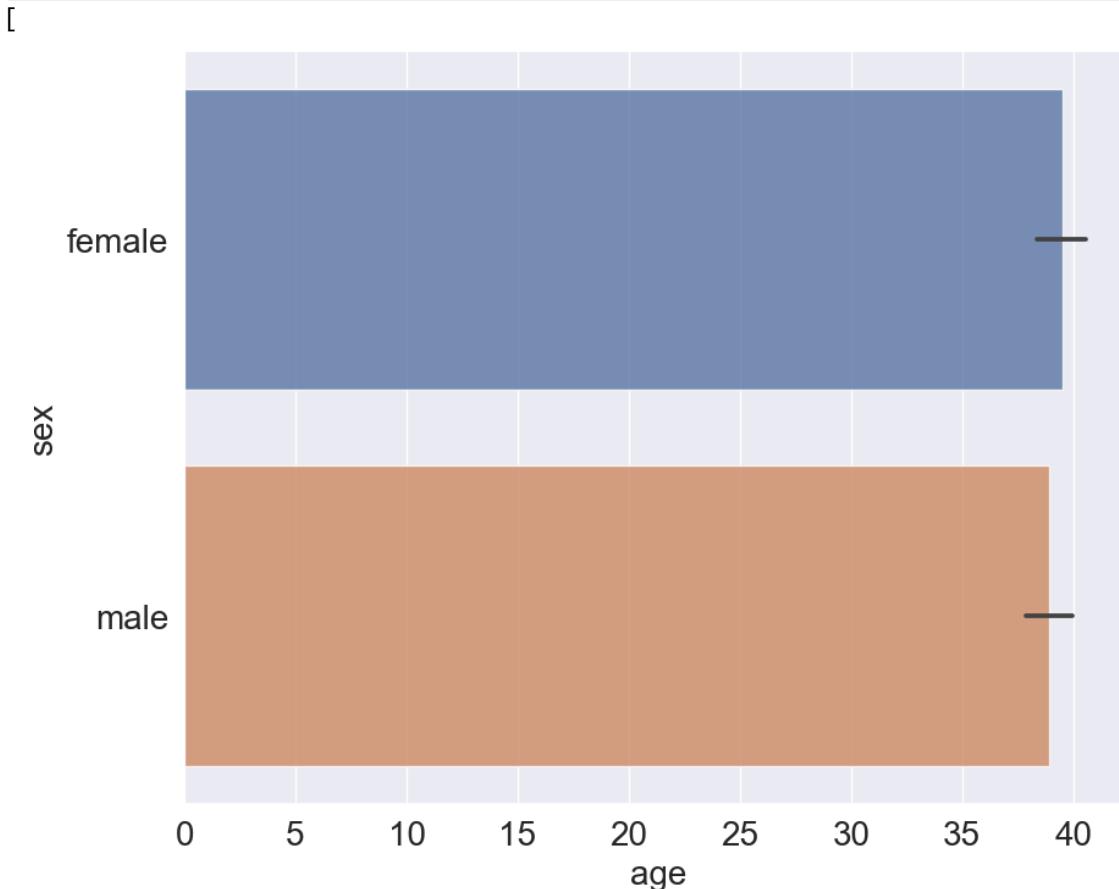


```
# rel plot
plt.figure(figsize=(14,20))
sns.set(rc={'xtick.labelsize':20,'ytick.labelsize':20,'axes.labelsize':20})
sns.relplot(x='age',y='charges',hue='sex',kd='le',col="children", height=8.5,aspect
plt.show()
```

```
<Figure size 1400x2000 with 0 Axes>
```

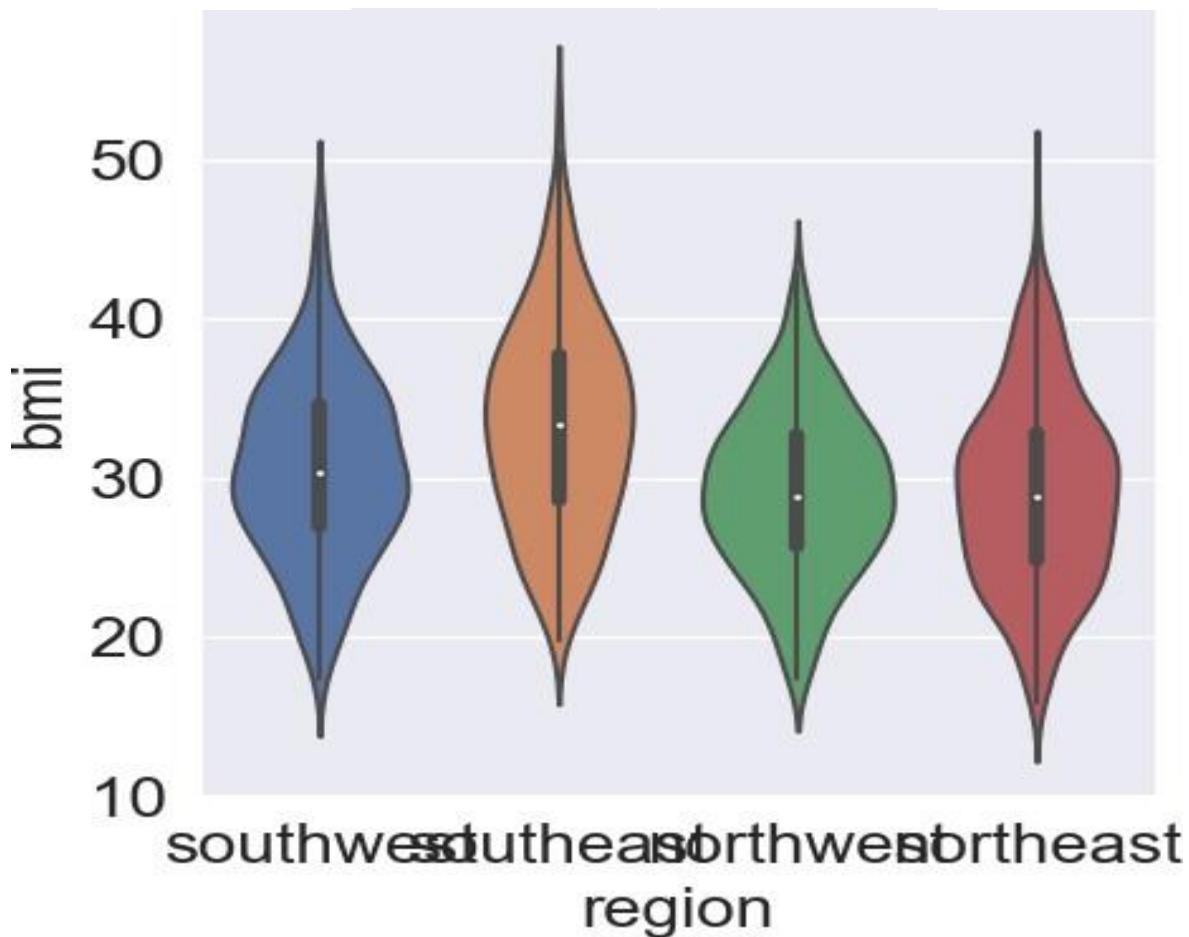


```
# bar plot
plt.figure(figsize=(10,8))
sns.barplot(x=df.age, y=df.sex, alpha=0.8)
plt.show()
```



```
plt.figure(figsize=(10,10))
sns.catplot(y ='region', kd = "count",data = df)
plt.show()
```

<Figure size 1000x1000 with 0 Axes>



```
# car plot

plt.figure(figsize=(12,10))
sns.catplot(x ='region' , y = "bmi" ,kind="violin" ,data = df)
plt.show()
```

<Figure size 1200x1000 with 0 Axes>

```
# histogram
plt.figure(figsize=(10,10))
sns.distplot(df['age'])
plt.show()
```

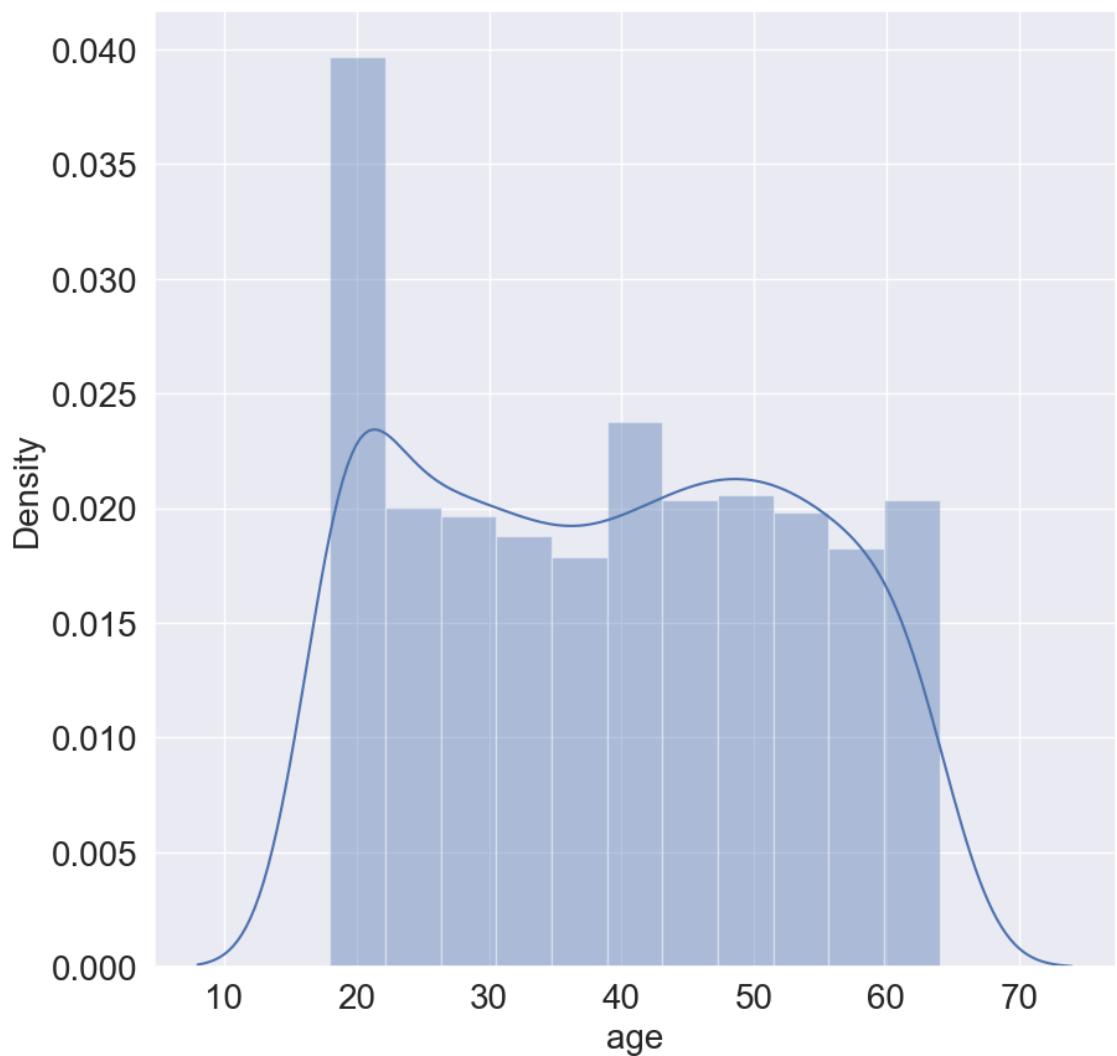
C:\Users\bala9\AppData\Local\Temp\ipykernel_12528\1806645068.py: UserWarning:

`distplot` is a deprecated function and will be removed in seaborn v0.14.0.

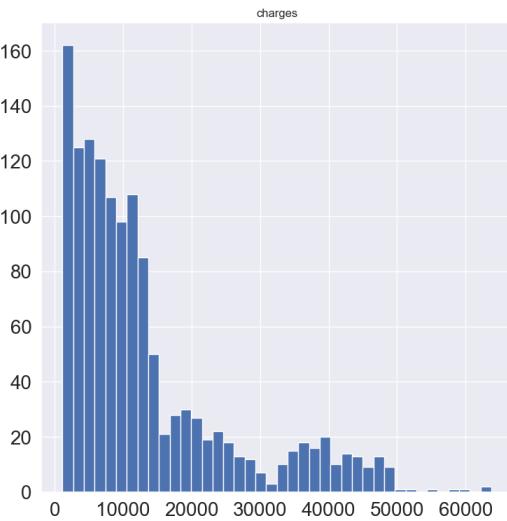
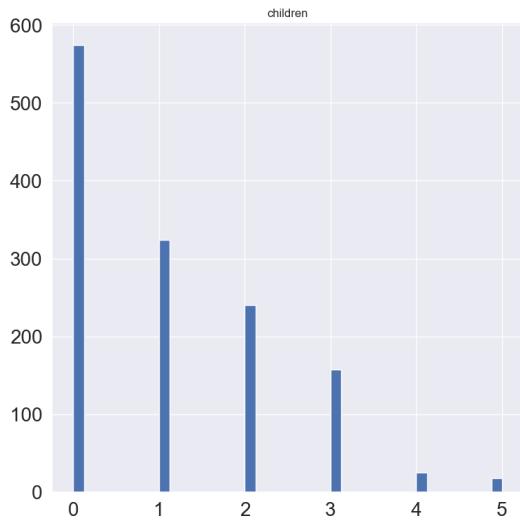
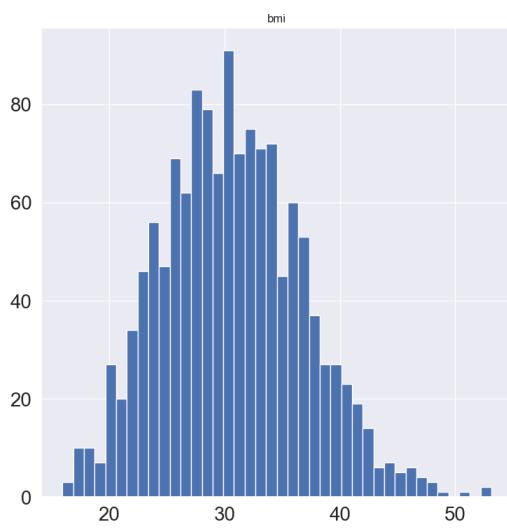
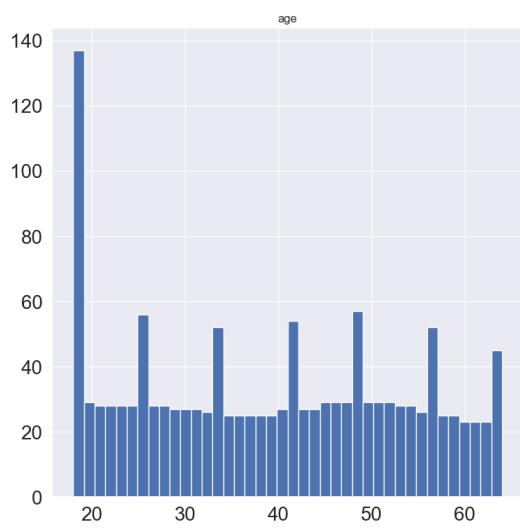
Please adapt your code to use either `displot` (a figure-level function with similar flexibility) or `histplot` (an axes-level function for histograms).

For a guide to updating your code to use the new functions, please see <https://gist.github.com/mwaskom/de44147ed2974457ad6372750bbe5751> (<https://gist.github.com/mwaskom/de44147ed2974457ad6372750bbe5751>)

```
sns.distplot(df['age'])
```



```
df.hist(bs=40 , figsize=(20,20))
plt.show()
```



```
# KDE plot

plt.figure(figsize=(10,8))
sns.kdeplot(x=df.age, y=df.charges, zorder=0, n_levels=6, shade=True,
             cbar=True, shade_lowest=False, cmap='viridis')
plt.show()
```

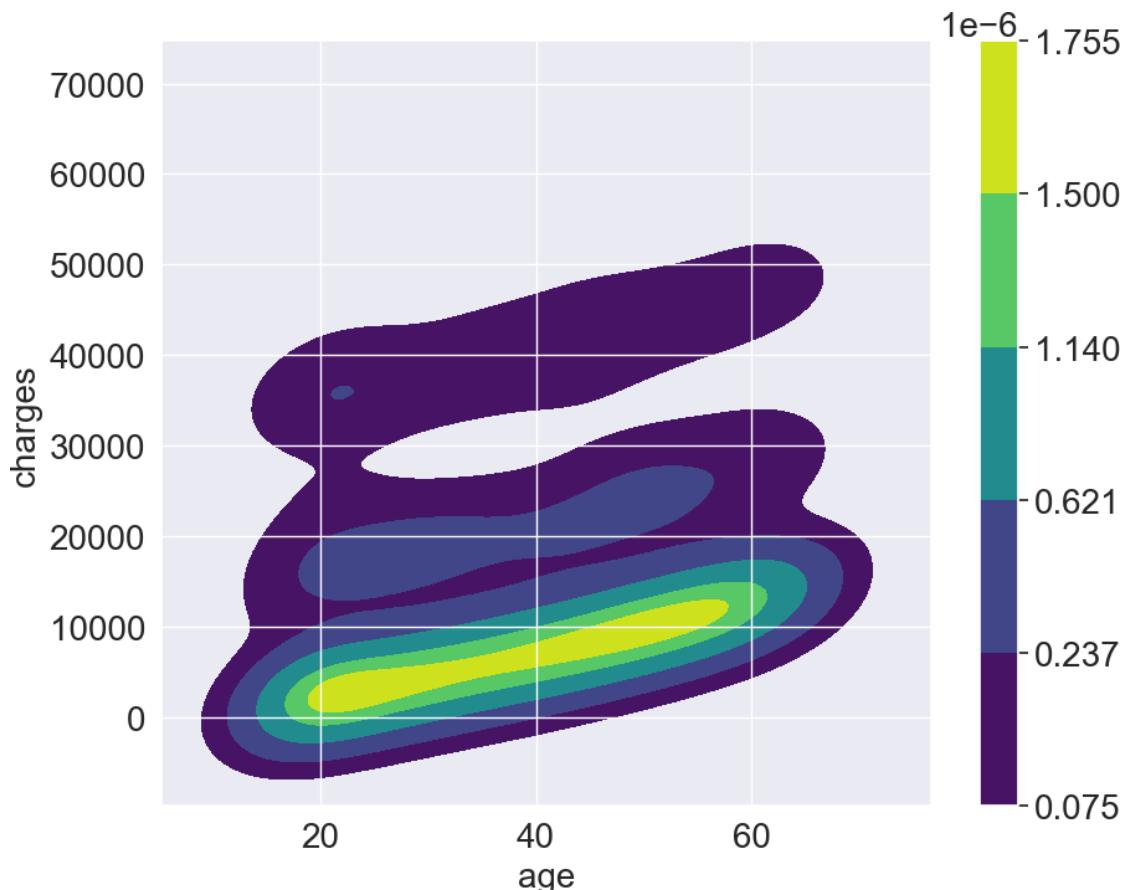
C:\Users\bala9\AppData\Local\Temp\ipykernel_12528\1412160463.py: UserWarning:

`shade_lowest` has been replaced by `thresh`; setting `thresh=0.05`. This will become an error seaborn v0.13.0; please update your code.

C:\Users\bala9\AppData\Local\Temp\ipykernel_12528\1412160463.py: FutureWarning:

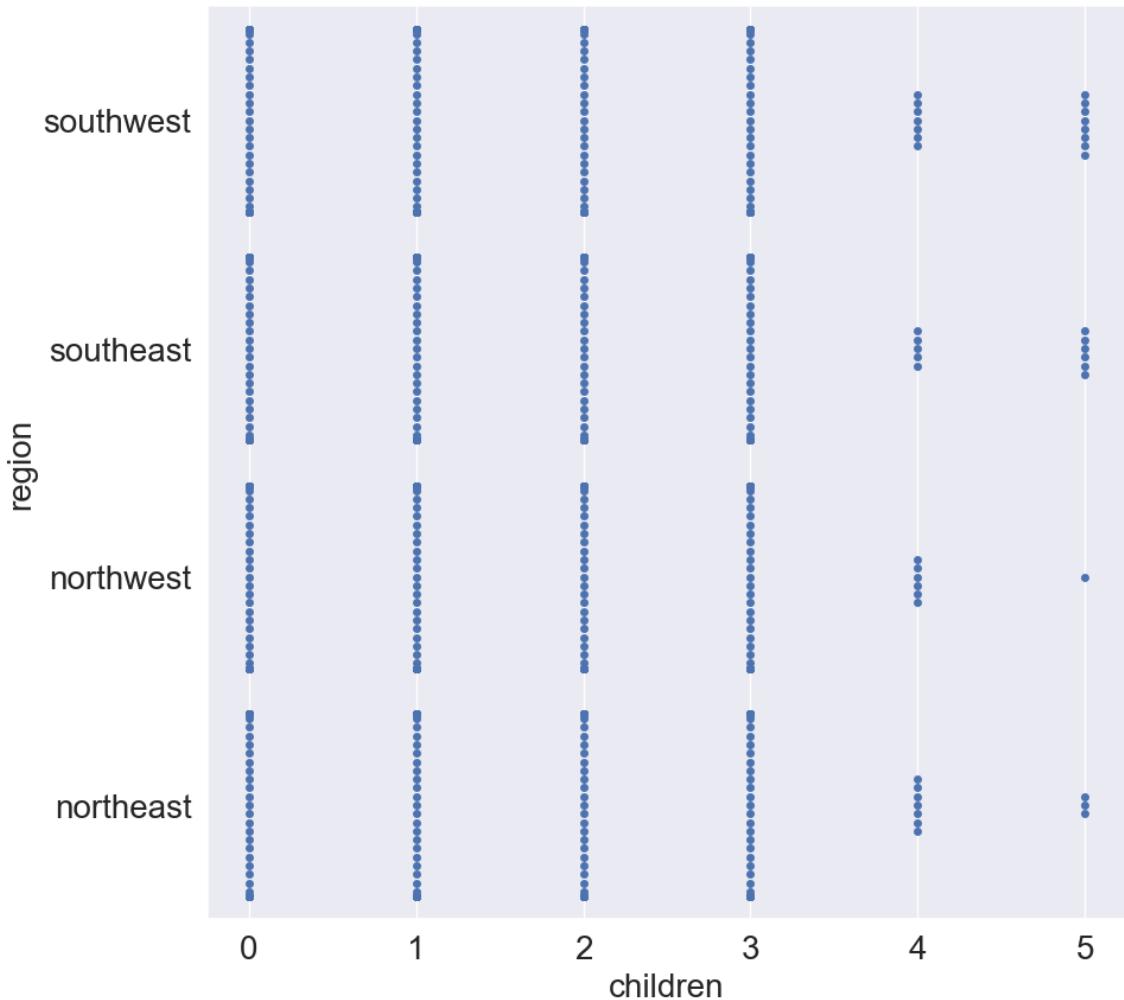
`shade` is now deprecated favor of `fill`; setting `fill=True`. This will become an error seaborn v0.14.0; please update your code.

sns.kdeplot(x=df.age, y=df.charges, zorder=0, n_levels=6, shade=True,



```
# swarm plot
plt.figure(figsize=(10,10))
sns.swarmplot(x=df['children'], y=df['region'])
plt.show()
```

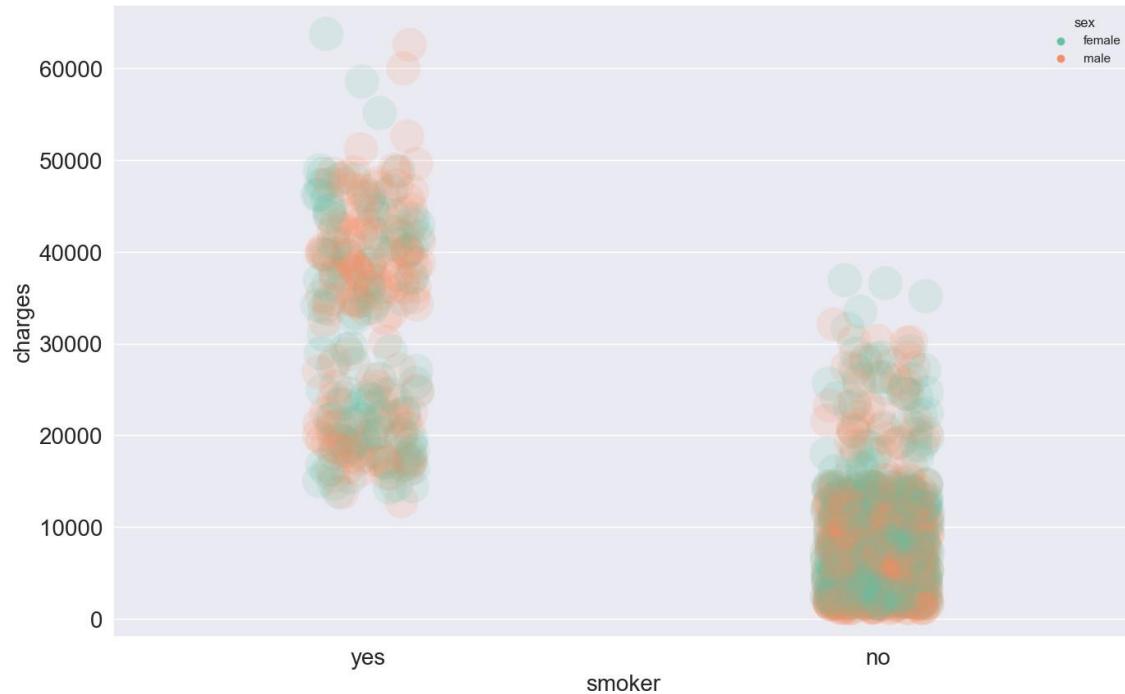
)



```
plt.figure(figsize=(7,7))
sns.swarmplot(x= "sex", y = "charges", size = 9 , linewidth= 2 , edgecolor="black" , data=df)
plt.show()
```

```
# strip plot
```

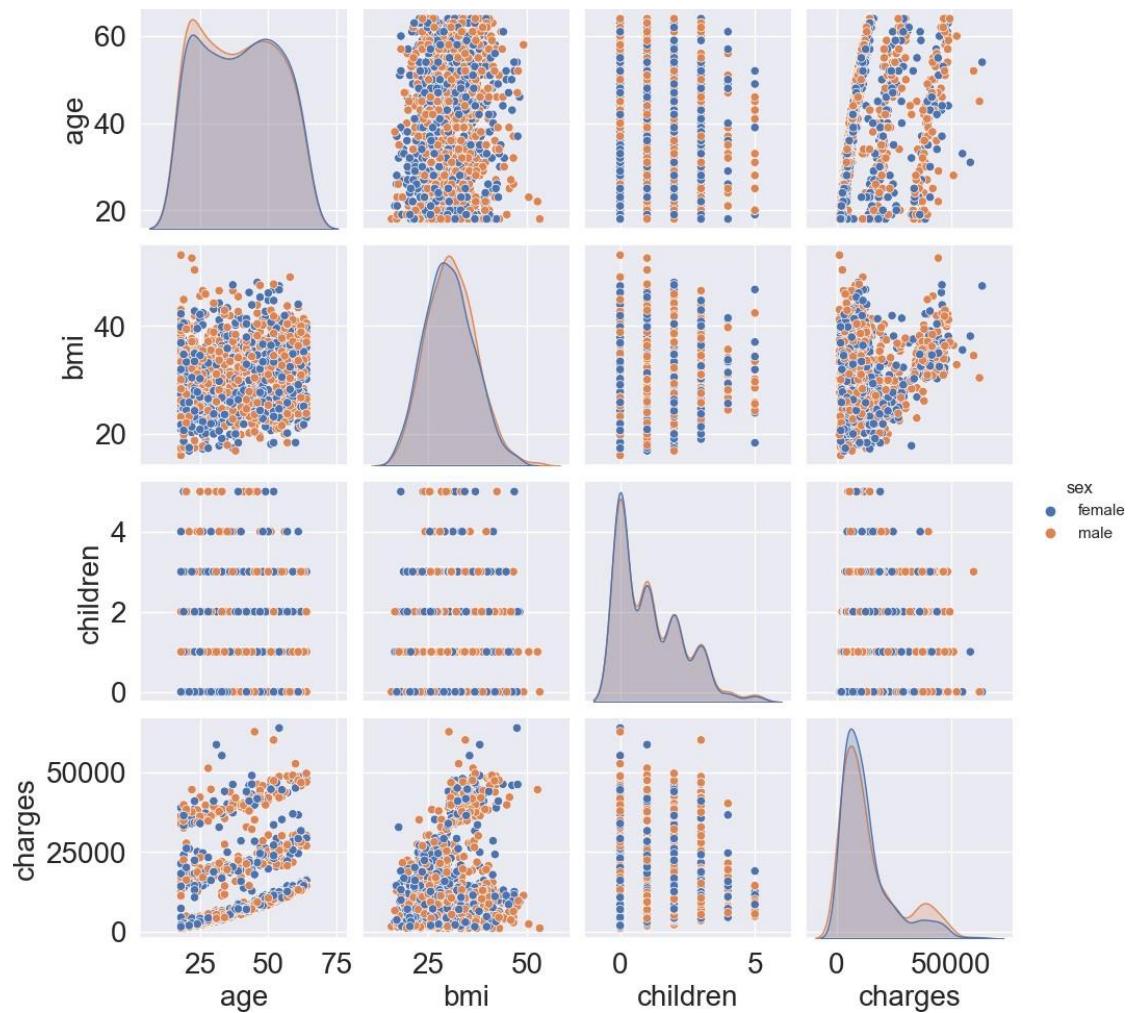
```
plt.figure(figsize=(16,10))
sns.stripplot(x=df[ "smoker" ] ,palette="Set2", y = df[ "charges" ],hue=df[ "sex" ],marker = ".")
plt.show()
```



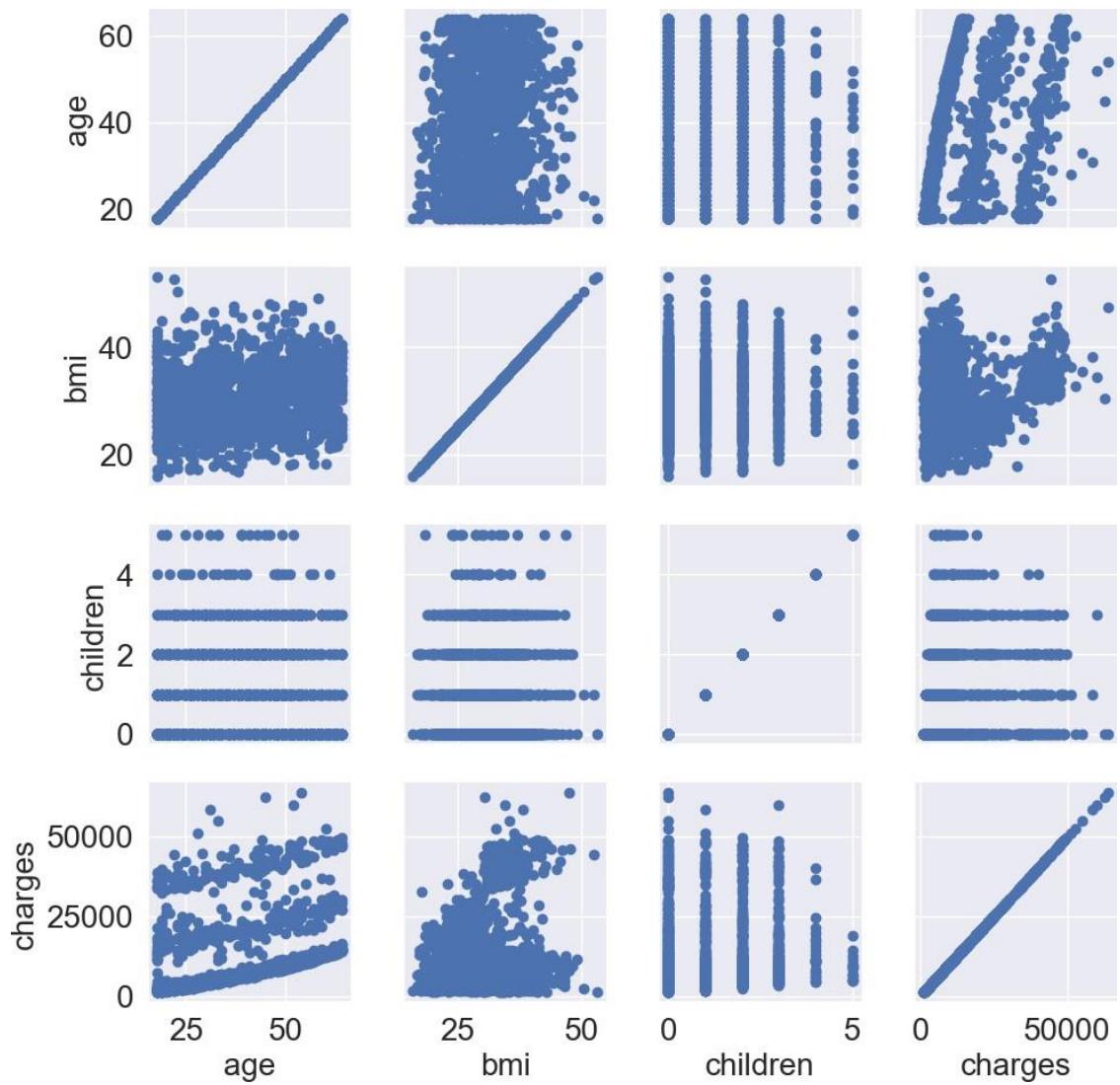
```
# pair plot  
sns.pairplot(df, hue='sex')
```

[55]:

<seaborn.axisgrid.PairGrid at 0x1796ed33f50>



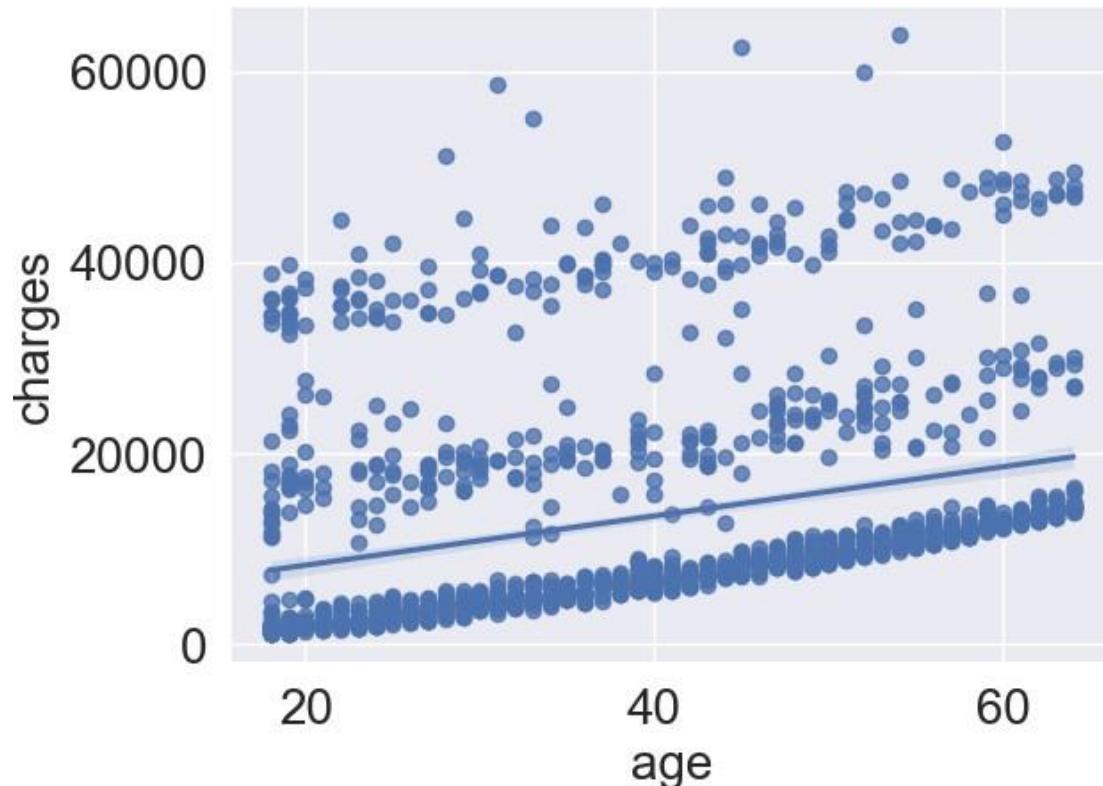
```
# grid plot
g = sns.PairGrid(df)
g = g.map(plt.scatter)
plt.show()
```



```
# regression plot  
sns.regplot(x=df['age'], y=df['charges'])
```

[57]:

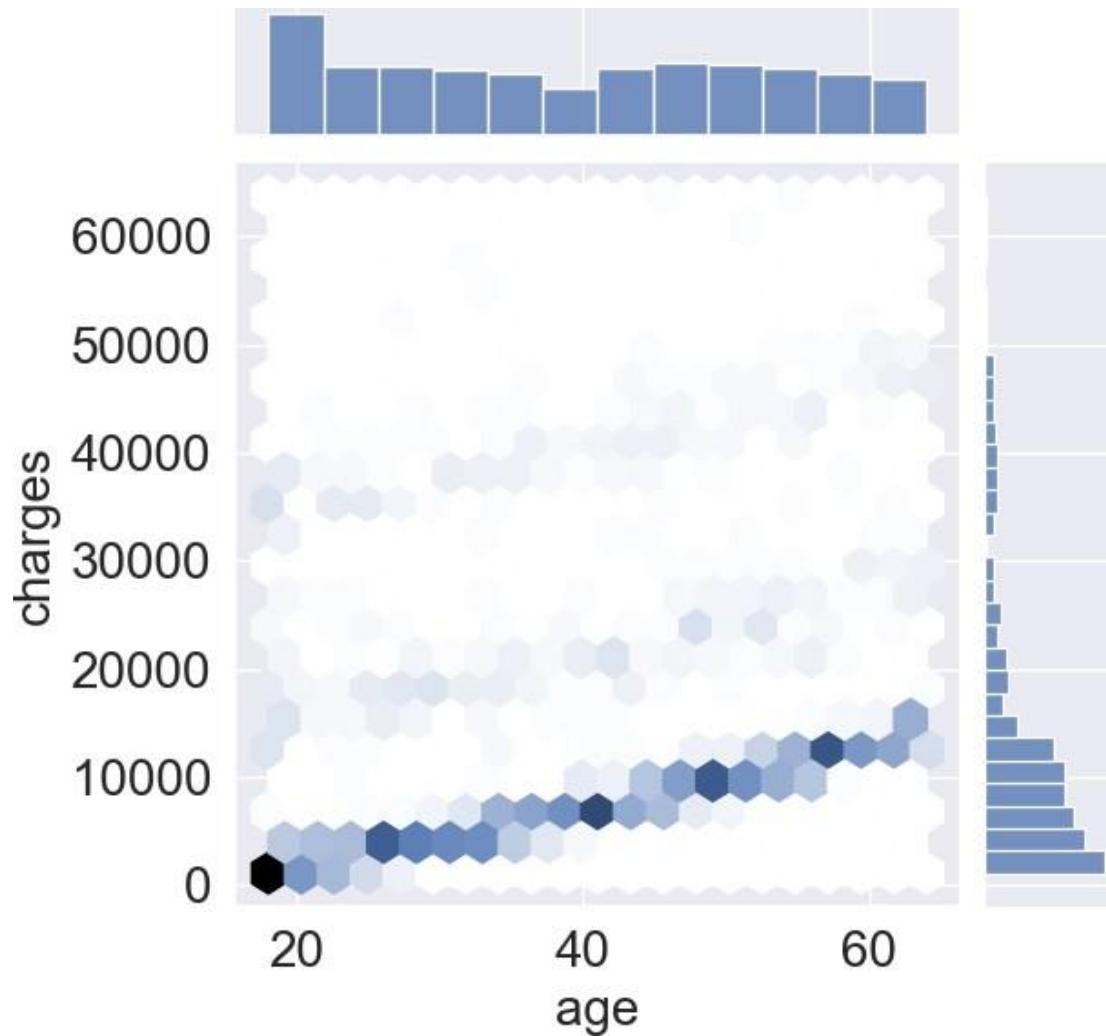
```
<AxesSubplot: xlabel='age', ylabel='charges'>
```



```
# jot plot  
sns.jotplot(x=df['age'], y=df['charges'], kd='hex')
```

[60]:

<seaborn.axisgrid.JotGrid at 0x1797890b010>

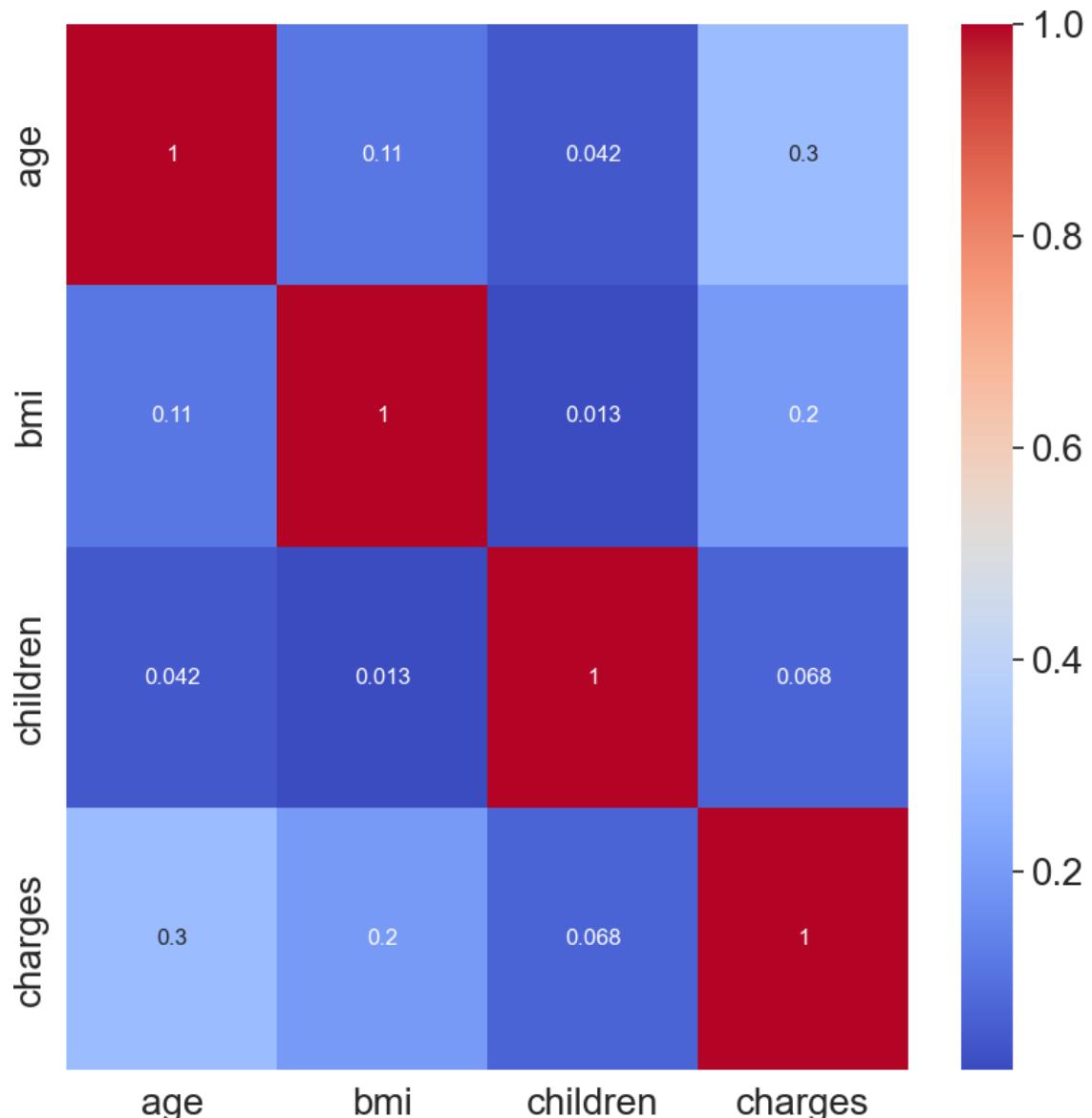


```
# heat map
```

```
plt.figure(figsize=(10,10))
sns.heatmap(df.corr(), annot=True, cmap='coolwarm')
plt.show()
```

C:\Users\bala9\AppData\Local\Temp\ipykernel_12528\62995110.py: FutureWarning: The default value of numeric_only DataFrame.corr is deprecated. In a future version, it will default to False. Select only valid columns or specify the value of numeric_only to silence this warning.

```
sns.heatmap(df.corr(), annot=True, cmap='coolwarm')
```



Result:

Thus visualization tools has been visualized and the results are verified.