Personal Task1 Author- Sagar Bhoi

1. Python program to Create a excel file

In [1]:
```
# !pip install xlsxwriter

#importing xlsxwriter module
import xlsxwriter
```

In [2]:
```
#Creating new excel file
workbook= xlsxwriter.Workbook("YoshopsTask1.xlsx")
```

In [3]:
```
#writing in Excel file worksheet
worksheet = workbook.add_worksheet()
worksheet.write('A1','Yoshops idea is to create India’s most reliable and regeniabl
workbook.close()
```

In [ ]:
```

```

2. Python program for Import data from an excel file

In [4]:
```
import pandas as pd
df=pd.read_excel("Yoshops Survey_1021_16_Jan_2023_Updated.xlsx")
df
```

Out[4]:

	S.NO	Submitted Time	1. Name	3. Location , City Name	4. What type of Tution are you paying?	5. Study in Class	6. Which Price range for Tution Monthly Fees You like must	7. Laptop and Mobile which Price range you like most	8. Do you like biriyani and which biriyani you like the more ?
0	1	24-11-2022	Kavita Israni	No answer	Offline Class room	No answer	Rs.501 to 999	No answer	Veg Biryani | Paneer Tikka Biryani
1	2	24-11-2022	Kunal Anand	No answer	Offline Class room	No answer	Rs.1 to Rs.499	No answer	Chicken Biryani
2	3	25-11-2022	Deepak parmal	No answer	Online Zoom meeting	Graduation with Internship	Rs.1 to Rs.499	No answer	Chicken Biryani
3	4	25-11-2022	Nidhi Gupta	No answer	Online Zoom meeting | Offline Class room	LKG to STD 5 | STD 6 to STD 10	Rs.1001 to 1499	No answer	Chicken Biryani
4	5	25-11-2022	Rohan Pandey	No answer	Online Zoom meeting	STD 6 to STD 10	Rs.1001 to 1499	No answer	Chicken Biryani
...	...	...	...	...	...	...	...	...	...
1007	1008	12-01-2023	Ritesh Kumar	Sasaram	Online Zoom meeting	Graduation with Internship	Rs.1 to Rs.499	Rs15000 to Rs.30000	Paneer Tikka Biryani
1008	1009	12-01-2023	Swati S	Chennai	Online Zoom meeting	Post Graduation with Internship	Rs.1001 to 1499	Rs.31000 to Rs.50000 | Rs.51000 to Rs.100000	Chicken Biryani | Mutton Biryani
1009	1010	12-01-2023	K Kiran	Chennai	Online Zoom meeting	Post Graduation with Internship	Rs.501 to 999 | Rs.1001 to 1499	Rs.31000 to Rs.50000	Chicken Biryani | Mutton Biryani

	S.NO	Submitted Time	1. Name	3. Location , City Name	4. What type of Tution are you paying?	5. Study in Class	6. Which Price range for Tution Monthly Fees You like must	7. Laptop and Mobile which Price range you like most	8. Do you like biriyani and which biriyani you like the more ?
1010	1011	13-01-2023	Ave Maria	Angamaly, Kerala	Offline Class room	STD 6 to STD 10	Rs.1 to Rs.499	Rs15000 to Rs.30000	Chicken Biryani
1011	1012	13-01-2023	Sara Shaju	Angamaly, Kerala	Offline Class room	STD 6 to STD 10	Rs.1 to Rs.499	Rs15000 to Rs.30000	Chicken Biryani

1012 rows × 12 columns

In [ ]:

3. Python program for Format data in excel sheet

In [5]:
```
import openpyxl
import re as re
import os as os
from openpyxl.styles import numbers
from openpyxl.styles import Font, Color
```

In [6]:
```
wb=openpyxl.load_workbook("Yoshops Survey_1021_16_Jan_2023_Updated.xlsx")
ws=wb['Responses']
ws['B1']= 'Submission'
wb.save("Yoshops Survey_1021_16_Jan_2023_Updated_new.xlsx")

##BOLD HEADERS
Font_style=Font(name="Calibri",size=14,bold=True,color="661111")
a4=ws['B1']
a4.font=Font_style
wb.save("Yoshops Survey_1021_16_Jan_2023_Updated_new.xlsx")
for  i in range (1,8):
    ws.cell(row=1,column=i).font=Font_style

wb.save("Yoshops Survey_1021_16_Jan_2023_Updated_new.xlsx")
```

In [7]:
```
df
```

Out[7]:

	S.NO	Submitted Time	1. Name	3. Location , City Name	4. What type of Tution are you paying?	5. Study in Class	6. Which Price range for Tution Monthly Fees You like must	7. Laptop and Mobile which Price range you like most	8. Do you like biriyani and which biriyani you like the more ?
0	1	24-11-2022	Kavita Israni	No answer	Offline Class room	No answer	Rs.501 to 999	No answer	Veg Biryani | Paneer Tikka Biryani
1	2	24-11-2022	Kunal Anand	No answer	Offline Class room	No answer	Rs.1 to Rs.499	No answer	Chicken Biryani
2	3	25-11-2022	Deepak parmal	No answer	Online Zoom meeting	Graduation with Internship	Rs.1 to Rs.499	No answer	Chicken Biryani
3	4	25-11-2022	Nidhi Gupta	No answer	Online Zoom meeting | Offline Class room	LKG to STD 5 | STD 6 to STD 10	Rs.1001 to 1499	No answer	Chicken Biryani
4	5	25-11-2022	Rohan Pandey	No answer	Online Zoom meeting	STD 6 to STD 10	Rs.1001 to 1499	No answer	Chicken Biryani
...	...	...	...	...	...	...	...	...	...
1007	1008	12-01-2023	Ritesh Kumar	Sasaram	Online Zoom meeting	Graduation with Internship	Rs.1 to Rs.499	Rs15000 to Rs.30000	Paneer Tikka Biryani
1008	1009	12-01-2023	Swati S	Chennai	Online Zoom meeting	Post Graduation with Internship	Rs.1001 to 1499	Rs.31000 to Rs.50000 | Rs.51000 to Rs.100000	Chicken Biryani | Mutton Biryani
1009	1010	12-01-2023	K Kiran	Chennai	Online Zoom meeting	Post Graduation with Internship	Rs.501 to 999 | Rs.1001 to 1499	Rs.31000 to Rs.50000	Chicken Biryani | Mutton Biryani

	S.NO	Submitted Time	1. Name	3. Location , City Name	4. What type of Tution are you paying?	5. Study in Class	6. Which Price range for Tution Monthly Fees You like must	7. Laptop and Mobile which Price range you like most	8. Do you like biriyani and which biriyani you like the more ?
1010	1011	13-01-2023	Ave Maria	Angamaly, Kerala	Offline Class room	STD 6 to STD 10	Rs.1 to Rs.499	Rs15000 to Rs.30000	Chicken Biryani
1011	1012	13-01-2023	Sara Shaju	Angamaly, Kerala	Offline Class room	STD 6 to STD 10	Rs.1 to Rs.499	Rs15000 to Rs.30000	Chicken Biryani

1012 rows × 12 columns

In [ ]:

4. Python program for Prepare Yoshops Survey and Order excel charts Like = Pie Chart and Bar Chart Weekly, Monthly and Yearly Reports.

In [8]:
```
import pandas as pd
import numpy as np
import os
import re as re
import matplotlib.pyplot as plt
import seaborn as sns
%matplotlib inline
import warnings
warnings.filterwarnings("ignore", category=FutureWarning)
```

In [9]:
```
url=r"C:\Users\Sagar\Yoshops Data Science Intern\Yoshops_Order_List.xlsx"
```

In [10]:
```
df.shape
```

Out[10]: (1012, 12)

In [11]:
```
df1=pd.read_excel(url,sheet_name="Mobile")
```

In [12]:
```
df1
```

Out[12]:

	Name	Address 1	Adress 2	Adress 3	City	pincode	
0	HARMESH RANI GABA	316 JAGRATI ENCLAVE	0	0	NEWDELHI	110092.0	
1	B KRISHNA BAI	204 CHITRAPUR HSG SOCIETY	15TH CROSS RD MALLESWARAM	BANGALORE	BANGALORE	560055.0	
2	GANGA SHREEDHAR	6 NUNGAMBAKKAM HIGH RD	MADRAS	0	CHENNAI	600034.0	
3	JOHN PINTO	NaN	NaN	NaN	NaN	NaN	
4	ROBIN GHOSH	C-O MR JOY DEEP KAR ADVOCATE	7 OLD POST OFFICE STREET	0	KOLKATTA	700001.0	
...	...	...	...	...	...	...	
366	VIJAYA ANANTHA NARAYANAN	A-3 SHREYAS APARTMENTS	C O D ROAD MALAD E	0	MUMBAI	400097.0	M
367	SULAXANA PRATAPRAI VYAS	204 CHANDRALOKA	97 NEAPEAN SEA ROAD	0	MUMBAI	400006.0	M
368	S B MOHANTY	217 GANESHNAGAR ADARSH C H S	TITWALA P O MANDA	TAL KALYAN DIST THANE	THANE	421605.0	M
369	JAYSHREE MODI	C 4 YESHWANT CO OP HSG SOC	236 NATH PAI NAGAR	GHATKOPAR EAST	MUMBAI	400077.0	M
370	SHAILESH K PUJARA H U F	54-10 NEELKANTH PRAKASH	GARODIA NAGAR GHATKOPAR EAST	0	MUMBAI	400077.0	M

371 rows × 7 columns

In [13]:
```
df1.isnull().sum()
```

Out[13]:
```
Name         0
Address 1    1
Adress 2     1
Adress 3     1
City         1
pincode      1
state        1
dtype: int64
```

In [14]:
```
df1_Customers=df1.groupby(['state'])['Name'].size().abs()
```

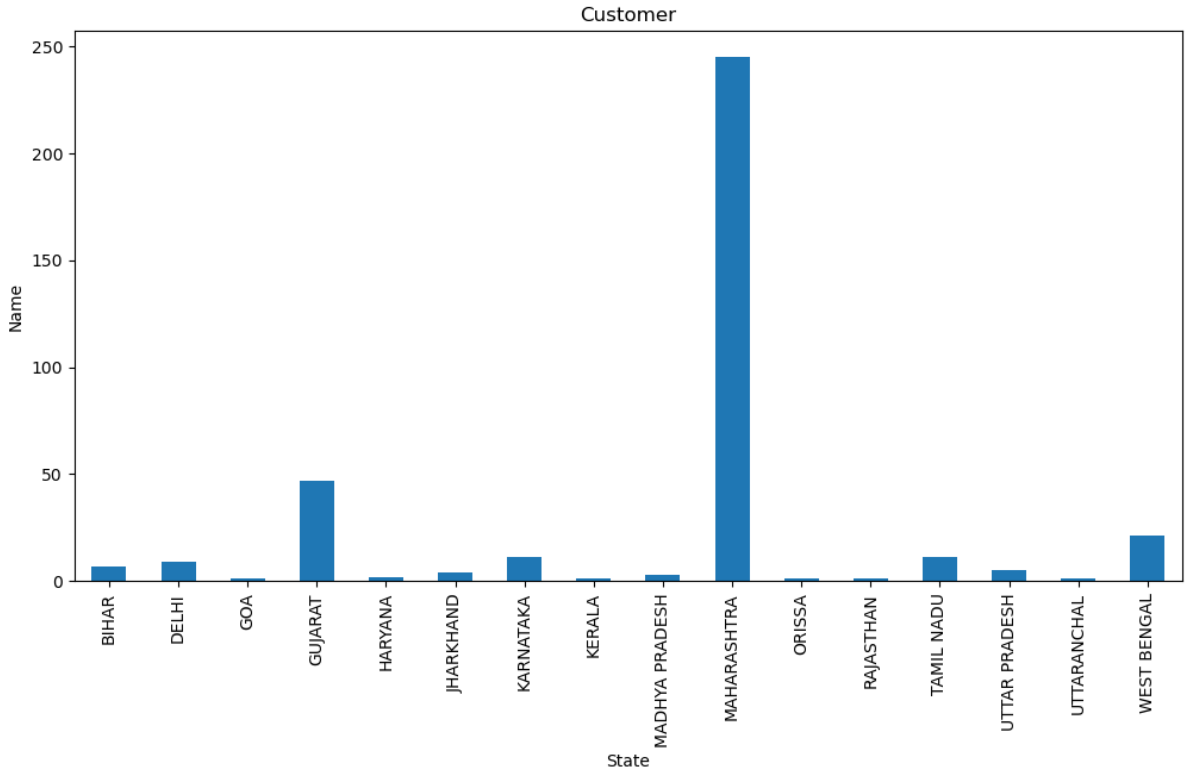
In [15]:
```
State=df1.state.unique()
```

In [16]:
```
State
```

Out[16]:
```
array(['DELHI', 'KARNATAKA', 'TAMIL NADU', nan, 'WEST BENGAL',
       'MAHARASHTRA', 'GUJARAT', 'ORISSA', 'UTTAR PRADESH', 'UTTARANCHAL',
       'KERALA', 'BIHAR', 'MADHYA PRADESH', 'GOA', 'HARYANA', 'JHARKHAND',
       'RAJASTHAN'], dtype=object)
```

In [17]:
```
df1_Customers.plot(kind='bar', title='Customer', ylabel='Name',
        xlabel='State', figsize=(12, 6))
```

Out[17]:
```
<AxesSubplot: title={'center': 'Customer'}, xlabel='State', ylabel='Name'>
```



In [18]:
```
Toys=pd.read_excel(url,sheet_name="Toys",header=1)
```

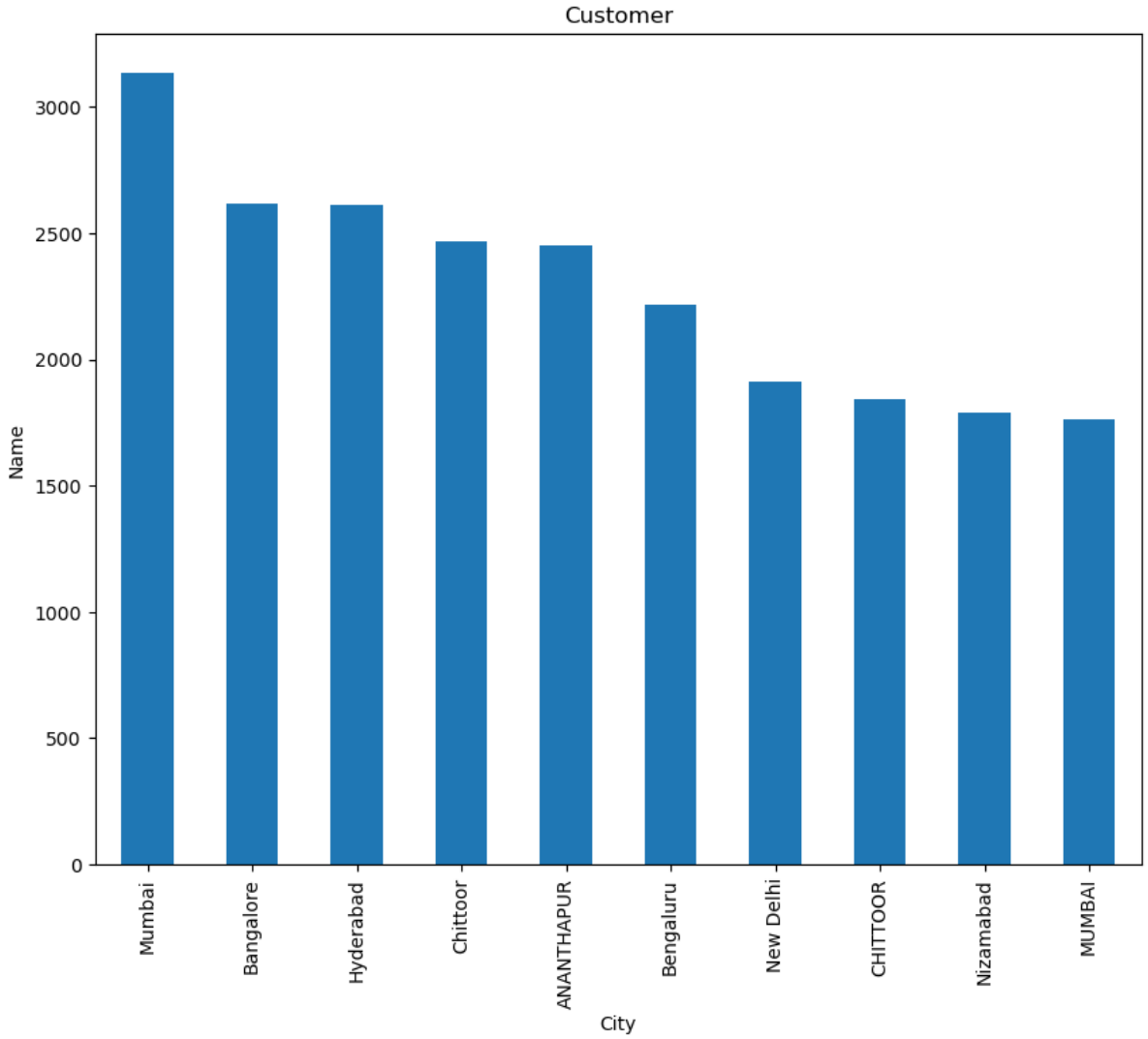
In [19]:
```
city_toys=Toys.groupby(['City'])["Name"].size().nlargest(10)
```

In [20]:
```
city_toys
```

Out[20]:
```
City
Mumbai        3134
Bangalore     2617
Hyderabad     2613
Chittoor      2470
ANANTHAPUR    2454
Bengaluru     2216
New Delhi     1913
CHITTOOR      1843
Nizamabad     1788
MUMBAI        1763
Name: Name, dtype: int64
```

In [21]:
```
city_toys.plot(kind='bar', title='Customer', ylabel='Name',
        xlabel='City', figsize=(10, 8))
```

Out[21]: <AxesSubplot: title={'center': 'Customer'}, xlabel='City', ylabel='Name'>



In [22]:
```
df1.state.value_counts().nlargest(10)
```

Out[22]:
```
MAHARASHTRA       245
GUJARAT            47
WEST BENGAL        21
KARNATAKA          11
TAMIL NADU         11
DELHI               9
BIHAR               7
UTTAR PRADESH       5
JHARKHAND           4
MADHYA PRADESH      3
Name: state, dtype: int64
```

In [23]:
```
Online_class=pd.read_excel(url,sheet_name="Online Class")
```

In [24]:
```
Online_class.head()
```

Out[24]:

	Account Name	Address 1	Adress 2	City	PIN Code	State	amount
0	REVENUE DIVISIONAL OFFICER	NaN	NaN	MAHATMA GANDHI ROAD, VIJAYAWADA (PADMAVATHI HO...	NaN	ANDHRA PRADESH	NaN
1	A.KARUNA SHREE	NaN	NaN	MAHATMA GANDHI ROAD, VIJAYAWADA (PADMAVATHI HO...	NaN	ANDHRA PRADESH	NaN
2	T RAMA RAO	NaN	NaN	MAHATMA GANDHI ROAD, VIJAYAWADA (PADMAVATHI HO...	NaN	ANDHRA PRADESH	NaN
3	MULLAPATI SATYAVANI	NaN	NaN	MAHATMA GANDHI ROAD, VIJAYAWADA (PADMAVATHI HO...	NaN	ANDHRA PRADESH	NaN
4	KESAMNENI ANNAPOORNA,	NaN	NaN	MAHATMA GANDHI ROAD, VIJAYAWADA (PADMAVATHI HO...	NaN	ANDHRA PRADESH	NaN

In [33]:
```
Online_class.head()
```

Out[33]:

	Account Name	Address 1	Adress 2	City	PIN Code	State	amount
0	REVENUE DIVISIONAL OFFICER	NaN	NaN	MAHATMA GANDHI ROAD, VIJAYAWADA (PADMAVATHI HO...	NaN	ANDHRA PRADESH	NaN
1	A.KARUNA SHREE	NaN	NaN	MAHATMA GANDHI ROAD, VIJAYAWADA (PADMAVATHI HO...	NaN	ANDHRA PRADESH	NaN
2	T RAMA RAO	NaN	NaN	MAHATMA GANDHI ROAD, VIJAYAWADA (PADMAVATHI HO...	NaN	ANDHRA PRADESH	NaN
3	MULLAPATI SATYAVANI	NaN	NaN	MAHATMA GANDHI ROAD, VIJAYAWADA (PADMAVATHI HO...	NaN	ANDHRA PRADESH	NaN
4	KESAMNENI ANNAPOORNA,	NaN	NaN	MAHATMA GANDHI ROAD, VIJAYAWADA (PADMAVATHI HO...	NaN	ANDHRA PRADESH	NaN

In [34]:
```
city=Online_class.City.unique()
```

In [35]:
```
State_value=Online_class.State.unique()
```

In [36]:
```
State_value
```

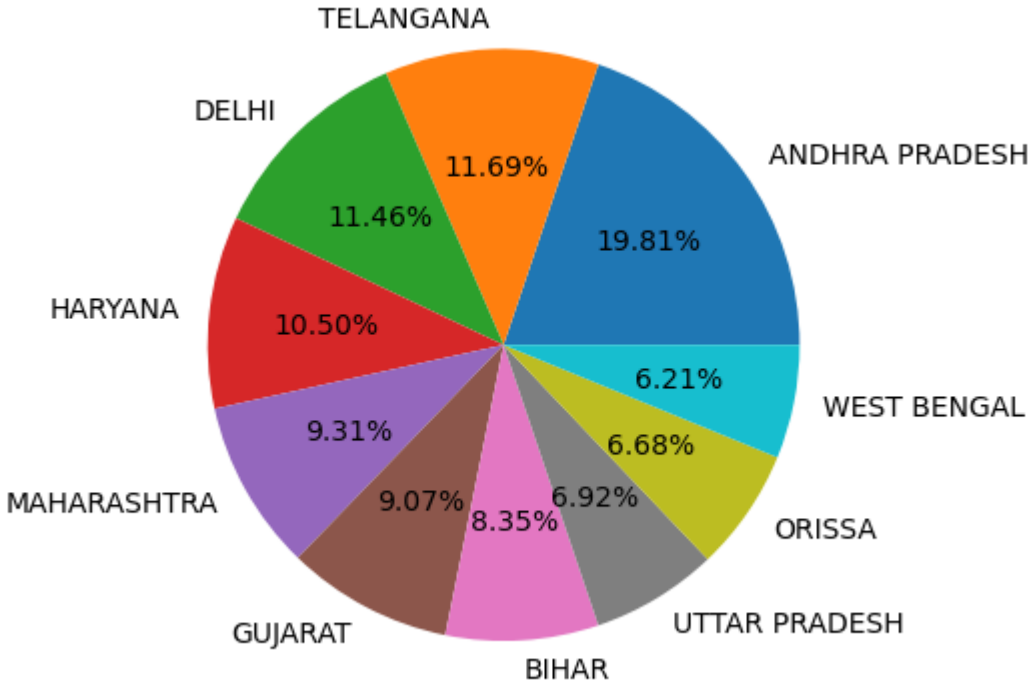
Out[36]:
```
array(['ANDHRA PRADESH', 'TELANGANA', 'DELHI', 'HARYANA', 'MAHARASHTRA',
       'GUJARAT', 'BIHAR', 'UTTAR PRADESH', 'ORISSA', 'WEST BENGAL',
       'CHANDIGARH', 'KERALA', 'PUNJAB', 'MADHYA PRADESH', 'GOA',
       'RAJASTHAN', 'PUDUCHERRY', 'TAMILNADU', 'KARNATAKA'], dtype=object)
```

In [37]:
```
Account_holder=Online_class["Account Name"].value_counts()
Account_holder
```

Out[37]:
```
ACTO HANAMKONDA                 83
ACTO I KARIMNAGAR               49
EXECUTIVE OFFICER               48
PRINCIPAL JUNIOR CIVIL JUDGE    44
ACTO II KARIMNAGAR              39
                                ..
B SASHIKALA                      1
B SASIKALA                       1
B Sathyanarayana                 1
B SATHYANARAYANA LF 80/16        1
ZYDUS PHARMA                     1
Name: Account Name, Length: 150492, dtype: int64
```

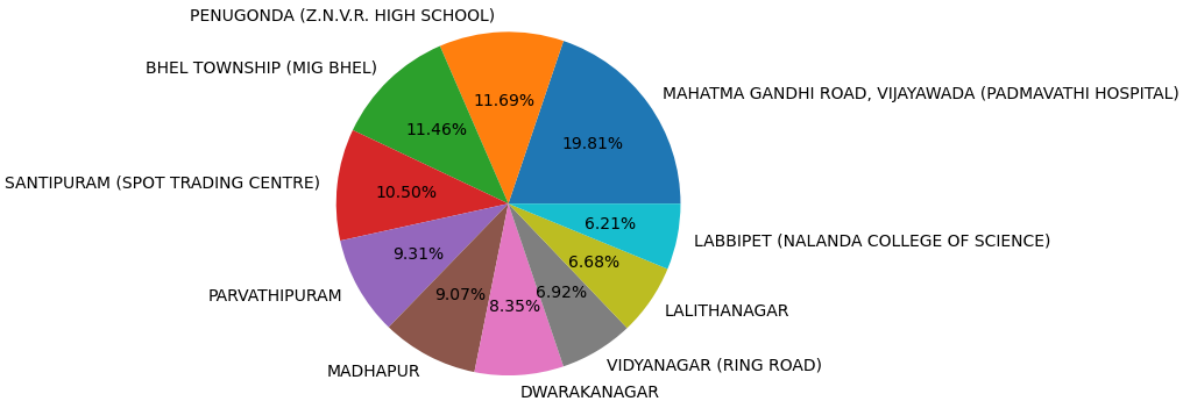
In [38]:
```
plt.pie(Account_holder[:10],labels=State_value[:10],autopct='%1.2f%%')
```

Out[38]:
```
([<matplotlib.patches.Wedge at 0x1f6d5067760>,
  <matplotlib.patches.Wedge at 0x1f6d5067dc0>,
  <matplotlib.patches.Wedge at 0x1f6d5070490>,
  <matplotlib.patches.Wedge at 0x1f6d5070b20>,
  <matplotlib.patches.Wedge at 0x1f6d50781f0>,
  <matplotlib.patches.Wedge at 0x1f6d50788b0>,
  <matplotlib.patches.Wedge at 0x1f6d5078f40>,
  <matplotlib.patches.Wedge at 0x1f6d507f610>,
  <matplotlib.patches.Wedge at 0x1f6d507fca0>,
  <matplotlib.patches.Wedge at 0x1f6d5087370>],
 [Text(0.8937809334005918, 0.6412141943918327, 'ANDHRA PRADESH'),
  Text(-0.045349020088905394, 1.0990648144568074, 'TELANGANA'),
  Text(-0.7645855130470348, 0.7908280427745356, 'DELHI'),
  Text(-1.0930504105191774, 0.12345363528004374, 'HARYANA'),
  Text(-0.9600980233054592, -0.5368535979621912, 'MAHARASHTRA'),
  Text(-0.5114763359796779, -0.9738541768318314, 'GUJARAT'),
  Text(0.07005733014298224, -1.0977668106174632, 'BIHAR'),
  Text(0.5689358731423637, -0.9414414332563319, 'UTTAR PRADESH'),
  Text(0.9079767710240082, -0.6209494208716326, 'ORISSA'),
  Text(1.0791644397471654, -0.2130824065594966, 'WEST BENGAL')],
 [Text(0.4875168727639591, 0.34975319694099966, '19.81%'),
  Text(-0.024735829139402938, 0.5994898987946221, '11.69%'),
  Text(-0.41704664348020076, 0.4313607506042921, '11.46%'),
  Text(-0.5962093148286421, 0.06733834651638748, '10.50%'),
  Text(-0.5236898308938868, -0.2928292352521043, '9.31%'),
  Text(-0.27898709235255154, -0.5311931873628171, '9.07%'),
  Text(0.03821308916889939, -0.5987818967004345, '8.35%'),
  Text(0.31032865807765286, -0.5135135090489082, '6.92%'),
  Text(0.4952600569221862, -0.33869968411179957, '6.68%'),
  Text(0.5886351489529993, -0.11622676721427086, '6.21%')])
```



In [39]:
```
plt.pie(Account_holder[:10],labels=city[:10],autopct='%1.2f%%')
```

Out[39]:
```
([<matplotlib.patches.Wedge at 0x1f6d50b40d0>,
  <matplotlib.patches.Wedge at 0x1f6d50b4790>,
  <matplotlib.patches.Wedge at 0x1f6d50b4d00>,
  <matplotlib.patches.Wedge at 0x1f6d50ba3d0>,
  <matplotlib.patches.Wedge at 0x1f6d50baa60>,
  <matplotlib.patches.Wedge at 0x1f6d50c3130>,
  <matplotlib.patches.Wedge at 0x1f6d50c37c0>,
  <matplotlib.patches.Wedge at 0x1f6d50c3e50>,
  <matplotlib.patches.Wedge at 0x1f6d50cb520>,
  <matplotlib.patches.Wedge at 0x1f6d50cbbb0>],
 [Text(0.8937809334005918, 0.6412141943918327, 'MAHATMA GANDHI ROAD, VIJAYAWADA (P
ADMAVATHI HOSPITAL)'),
  Text(-0.045349020088905394, 1.0990648144568074, 'PENUGONDA (Z.N.V.R. HIGH SCHOO
L) '),
  Text(-0.7645855130470348, 0.7908280427745356, 'BHEL TOWNSHIP (MIG BHEL)'),
  Text(-1.0930504105191774, 0.12345363528004374, 'SANTIPURAM (SPOT TRADING CENTR
E)'),
  Text(-0.9600980233054592, -0.5368535979621912, 'PARVATHIPURAM'),
  Text(-0.5114763359796779, -0.9738541768318314, 'MADHAPUR'),
  Text(0.07005733014298224, -1.0977668106174632, 'DWARAKANAGAR'),
  Text(0.5689358731423637, -0.9414414332563319, 'VIDYANAGAR (RING ROAD)'),
  Text(0.9079767710240082, -0.6209494208716326, 'LALITHANAGAR'),
  Text(1.0791644397471654, -0.2130824065594966, 'LABBIPET (NALANDA COLLEGE OF SCIE
NCE)')],
 [Text(0.4875168727639591, 0.34975319694099966, '19.81%'),
  Text(-0.024735829139402938, 0.5994898987946221, '11.69%'),
  Text(-0.41704664348020076, 0.4313607506042921, '11.46%'),
  Text(-0.5962093148286421, 0.06733834651638748, '10.50%'),
  Text(-0.5236898308938868, -0.2928292352521043, '9.31%'),
  Text(-0.27898709235255154, -0.5311931873628171, '9.07%'),
  Text(0.03821308916889939, -0.5987818967004345, '8.35%'),
  Text(0.31032865807765286, -0.5135135090489082, '6.92%'),
  Text(0.4952600569221862, -0.33869968411179957, '6.68%'),
  Text(0.5886351489529993, -0.11622676721427086, '6.21%')])
```



In [ ]:

5. Python program for Extract mobile no from PDF, Json and MS word file and save into MS excel

Extracting Mob No. from MS word File

In [40]:
```
pip install docx2txt
```

```
Collecting docx2txt
  Downloading docx2txt-0.8.tar.gz (2.8 kB)
  Preparing metadata (setup.py): started
  Preparing metadata (setup.py): finished with status 'done'
Building wheels for collected packages: docx2txt
  Building wheel for docx2txt (setup.py): started
  Building wheel for docx2txt (setup.py): finished with status 'done'
  Created wheel for docx2txt: filename=docx2txt-0.8-py3-none-any.whl size=3966 sha
256=fb708256b4896f303798340ee81867825a758e606356440dfb1f1e2ff4ecfdca
  Stored in directory: c:\users\sagar\appdata\local\pip\cache\wheels\27\87\87\6c7e
cf671f38e277e9b77e3a93e47e14bab847dd939d84cd25
Successfully built docx2txt
Installing collected packages: docx2txt
Successfully installed docx2txt-0.8
Note: you may need to restart the kernel to use updated packages.
```

import docx2txt import re my_doc=docx2txt.process(r"D:\autoCV.docx") my_doc pattern = re.compile(r'[789]\d{9}]*') matches=pattern.finditer(my_doc) for match in matches: print(match.group(0))

Extracting Mobile No. from JSon File

In [41]:
```
import json
import os as os
import re as re
import pandas as pd
```

In [42]:
```
path=r'C:\Users\Sagar\Yoshops Data Science Intern\contact data\1657173630381_504257
def get_file(path):
    files=[]
    file_list=[]
    #r=root, d=directory, f=files
    for r,d,f in os.walk(path):
        for file in f:
            if '.json' in file:
                files.append(os.path.join(r,file))
    for f in files:

        file_list.append(f)
    return file_list
```

In [43]:
```
list_of_paths=get_file(path)
```

In [44]:
```
mob1=[]
mob2=[]
for i in list_of_paths:
    with open(i,'r') as f:
        docs=json.loads(f.read())
    number1=''
    dic1 = docs['messages'][0]
    dic2 = docs['messages'][1]
```

```
    for i in dic1['msg']:
        if i.isdigit():
            number1+=i
    mob1.append(number1)
    number2=''
    for j in dic2['msg']:
        if j.isdigit():
            number2+=j
    mob2.append(number2)
df = pd.DataFrame.from_dict({'Yoshops':mob1,'Customers':mob2})
df.to_excel('test.xlsx', header=True, index=False)
```

In [45]:
```
mob1
```

Out[45]:
```
['919080749858',
 '919080749858',
 '919080749858',
 '919080749858',
 '919080749858',
 '919080749858',
 '919080749858',
 '919080749858',
 '919080749858',
 '919080749858',
 '919080749858',
 '919080749858',
 '919080749858',
 '919080749858',
 '919080749858',
 '919080749858',
 '919080749858',
 '919080749858',
 '919080749858',
 '919080749858']
```

In [46]:
```
mob2
```

Out[46]:
```
['8459599718',
 '9528692288',
 '9123557647',
 '8280021014',
 '9304522101',
 '6200028123',
 '03181982707',
 '9391193459',
 '9799818360',
 '9962278886',
 '9569402314',
 '8767852141',
 '8815007851',
 '7073128350',
 '6382105605',
 '08164881200',
 '8815007851',
 '9515654885',
 '9154940621',
 '7277862208']
```

Extracting Mobile No. from Pdf file

importing required modules

import PyPDF2

creating a pdf file object

f = open('survey report pdf.pdf', 'rb') pdfReader = PyPDF2.PdfFileReader(f)

printing number of pages in pdf file

print(pdfReader.numPages)

creating a page object

pageone = pdfReader.getPage(0)

extracting text from page

print(pageone.extractText())

f.close()

In [ ]:

6.Prepare python program for data cleaning process to removing unnecessary data

In [60]:
```
df1 = pd.read_excel('Yoshops_Feedback.xlsx')
```

In [62]:
```
df1.head()
```

Out[62]:

	11. Any IDEA or Suggestions for Yoshops Startup
0	No Answer
1	No Answer
2	Management should be better, I Think If you wo...
3	No
4	Marketing

In [63]:
```
df1.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1012 entries, 0 to 1011
Data columns (total 1 columns):
 #   Column                                          Non-Null Count  Dtype 
---  ------                                          --------------  ----- 
 0   11. Any IDEA or Suggestions for Yoshops Startup  950 non-null    object
dtypes: object(1)
memory usage: 8.0+ KB
```

In [64]:
```
df1.isnull().sum()
```

Out[64]:
```
11. Any IDEA or Suggestions for Yoshops Startup     62
dtype: int64
```

In [65]:
```
df1 = df1.dropna()
```

In [66]:
```
# remove duplicate rows
df1 = df1.drop_duplicates()
```

In [67]:
```
print(df1)
```

```
      11. Any IDEA or Suggestions for Yoshops Startup
0                                                No Answer
2     Management should be better, I Think If you wo...
3                                                      No
4                                               Marketing
5                                                      no
...                                                   ...
984                                             Work hard
986                                 doing good keep it up
989                                          No Nathing
1004                                         Good survey
1006                                  No I don't have

[367 rows x 1 columns]
```

In [68]:
```
# write cleaned data to a new excel file
df1.to_excel('cleaned_data.xlsx', index=False)
print('New cleaned excel file created')
```

```
New cleaned excel file created
```

In [ ]: