no of fundings

Check the trend of investments over the years. To check the trend, find - Total number of fundings done in each year. Plot a line graph between year and number of fundings. Take year on x-axis and number of fundings on y-axis. Print year-wise total number of fundings also. Print years in ascending order. Note : There is some error in the 'Date' feature. Make sure to handle that.

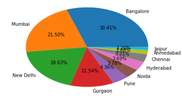
In [2]:
```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
df_start=pd.read_csv('startup_funding.csv',encoding='utf-8')
df_start['Date'].replace("12/05.2015","12/05/2015",inplace=True)
df_start['Date'].replace("13/04.2015","13/04/2015",inplace=True)
df_start['Date'].replace("15/01.2015","15/01/2015",inplace=True)
df_start['Date'].replace("22/01//2015","22/01/2015",inplace=True)
def convertDate(date):
    return date.split('/')[-1]
df_start['Year']=df_start['Date'].apply(convertDate)
year_count=df_start['Year'].value_counts()

year_fund=list(zip(year_count.index,year_count.values))

year_fund=np.array(year_fund,dtype=int)
year_fund=year_fund[year_fund[:,0].argsort()]

year=year_fund[:,0]

funding_round=year_fund[:,1]
plt.plot(year,funding_round,marker = 'o')
plt.xticks(year)
plt.title('Year vs No. of Funding Round')
plt.xlabel('Year')
plt.ylabel('No. of Funding Round')
plt.show()
for i in range(len(year)):
    print(year[i],funding_round[i])
```

```
<matplotlib.figure.Figure at 0x1eb0f8b77b8>

2015 936
2016 993
2017 443
```

top indian cities

Find out which cities are generally chosen for starting a startup. Find top 10 Indian cities which have most number of startups ? Plot a pie chart and visualise it. Print the city name and number of startups in that city also. Note : Take city name "Delhi" as "New Delhi". Check the case-sensitiveness of cities also. That means - at some place, instead of "Bangalore", "bangalore" is given. Take city name as "Bangalore". For few startups multiple locations are given, one Indian and one Foreign. Count those startups in Indian startup also. Indian city name is first. Print the city in descending order with respect to the number of startups.

In [3]:
```
import pandas as pd
import matplotlib.pyplot as plt
import numpy as np
df_start=pd.read_csv('startup_funding.csv',encoding='utf-8')
df_start['CityLocation'].dropna(inplace=True)
def separateCity(city):
    return city.split('/')[0].strip()
df_start['CityLocation']=df_start['CityLocation'].apply(separateCity)
df_start['CityLocation'].replace("Delhi","New Delhi",inplace=True)
df_start['CityLocation'].replace("bangalore","Bangalore",inplace=True)

city_count=df_start['CityLocation'].value_counts()[0:10]
city=city_count.index
numCity=city_count.values
plt.pie(numCity,labels=city,autopct="%.2f%%")
plt.show()
for i in range(len(city)):
    print(city[i],numCity[i])
```




```
Bangalore 635
Mumbai 449
New Delhi 389
Gurgaon 241
Pune 91
Noida 79
Hyderabad 77
Chennai 67
Ahmedabad 35
Jaipur 25
```

funding amount

Problem Statement : Find out if cities play any role in receiving funding. Find top 10 Indian cities with most amount of fundings received. Find out percentage of funding each city has got (among top 10 Indian cities only). Print the city and percentage with 2 decimal place after rounding off. Note: Take city name "Delhi" as "New Delhi". Check the case-sensitiveness of cities also. That means - at some place, instead of "Bangalore", "bangalore" is given. Take city name as "Bangalore". For few startups multiple locations are given, one Indian and one Foreign. Count those startups in Indian startup also. Indian city name is first. Print the city in descending order with respect to the percentage of funding.

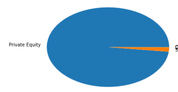
In [4]:
```
# Open and read data file as specified in the question
# Print the required output in given format
import pandas as pd
import numpy as np
df_start=pd.read_csv('startup_funding.csv',encoding='utf-8')
df_start['CityLocation'].dropna(inplace=True)
def separateCity(city):
    return city.split('/')[0].strip()
df_start['CityLocation']=df_start['CityLocation'].apply(separateCity)
df_start['CityLocation'].replace("Delhi","New Delhi",inplace=True)
df_start['CityLocation'].replace("bangalore","Bangalore",inplace=True)
## Converting "AmountInUSD" into numeric format
df_start["AmountInUSD"] = df_start["AmountInUSD"].apply(lambda x: float(str(x).replace(",",
""")))
df_start["AmountInUSD"] = pd.to_numeric(df_start["AmountInUSD"])
city_amount=df_start.groupby('CityLocation')['AmountInUSD'].sum().sort_values(ascending=Fals
e)[0:10]
city=city_amount.index
amountCity=city_amount.values
perAmount=np.true_divide(amountCity, amountCity.sum())*100
for i in range(len(city)):
    print(city[i],format(perAmount[i],'.2f'))
```

```
Bangalore 49.71
New Delhi 16.63
Mumbai 13.90
Gurgaon 12.21
Chennai 2.43
Pune 2.16
Hyderabad 1.15
Noida 1.01
Ahmedabad 0.58
Jaipur 0.21
```

investment type

There are 4 different type of investments. Find out percentage of amount funded for each investment type. Plot a pie chart to visualise. Print the investment type and percentage of amount funded with 2 decimal places after rounding off. Note : Correct spelling of investment types are - "Private Equity", "Seed Funding", "Debt Funding", and "Crowd Funding". Keep an eye for any spelling mistake. You can find this by printing unique values from this column. Print the investment type in descending order with respect to the percentage of the amount funded.

In [6]:
```
# Open and read data file as specified in the question
# Print the required output in given format
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import csv
df=pd.read_csv("startup_funding.csv", encoding='utf-8')
df['InvestmentType'].replace("PrivateEquity","Private Equity",inplace=True)
df['InvestmentType'].replace("SeedFunding","Seed Funding",inplace=True)
df['InvestmentType'].replace("Crowd funding","Crowd Funding",inplace=True)

df["AmountInUSD"] = df["AmountInUSD"].apply(lambda x: float(str(x).replace(",","","")))
df["AmountInUSD"] = pd.to_numeric(df["AmountInUSD"])
amount=df.groupby('InvestmentType')['AmountInUSD'].sum().sort_values(ascending=False)[::]
type=amount.index
invest=amount.values
perAmount=np.true_divide(invest, invest.sum())*100
plt.pie(perAmount, labels=type)
plt.show()
for i in range(len(type)):
    print(type[i],format(perAmount[i],'.2f'))
```




```
Private Equity 98.15
Seed Funding 1.81
Debt Funding 0.04
Crowd Funding 0.00
```

top industries

Problem Statement : Which type of companies get more easily funding. To answer this question, find - Top 5 industries and percentage of the total amount funded to that industry. (among top 5 only) Print the industry name and percentage of the amount funded with 2 decimal place after rounding off. Note : Ecommerce is the right word in industryVertical, so correct it. Print the industry in descending order with respect to the percentage of the amount funded.

In [8]:
```
import pandas as pd
import numpy as np
df=pd.read_csv('startup_funding.csv',encoding='utf-8')
df['IndustryVertical'].replace("eCommerce","ECommerce",inplace=True)
df['IndustryVertical'].replace('ECommerce','Ecommerce',inplace=True)
df['IndustryVertical'].replace('ecommerce','Ecommerce',inplace=True)


## Converting "AmountInUSD" into numeric format
df["AmountInUSD"] = df["AmountInUSD"].apply(lambda x: float(str(x).replace(",","","")))
df["AmountInUSD"] = pd.to_numeric(df["AmountInUSD"])
amount=df.groupby('IndustryVertical')['AmountInUSD'].sum().sort_values(ascending=False)[0:5]
industry=amount.index
amountCity=amount.values
perAmount=np.true_divide(amount, amount.sum())*100
for i in range(5):
    print(industry[i],format(perAmount[i],'.2f'))
```

```
ECommerce 40.53
Consumer Internet 35.95
Technology 10.45
Online Marketplace 6.63
E-Commerce & M-Commerce platform 6.44
```

top startups

Find top 5 startups with most amount of total funding. Print the startup name in descending order with respect to amount of funding. Note: Ola, Flipkart, Oyo, Paytm are important startups, so correct their names. There are many errors in startup names, ignore correcting all, just handle important ones.

In [9]:
```
import pandas as pd
import numpy as np
df_start=pd.read_csv('startup_funding.csv',encoding='utf-8')
df_start['StartupName'].replace('Olacabs','Ola',inplace=True)
df_start['StartupName'].replace('Ola Cabs','Ola',inplace=True)
df_start['StartupName'].replace('Flipkart.com','Flipkart',inplace=True)
df_start['StartupName'].replace('Paytm Marketplace','Paytm',inplace=True)
df_start['StartupName'].replace('Oyo Rooms','Oyo',inplace=True)
df_start['StartupName'].replace('Oyorooms','Oyo',inplace=True)
df_start['StartupName'].replace('OyoRooms','Oyo',inplace=True)
df_start['StartupName'].replace('OYO Rooms','Oyo',inplace=True)
## Converting "AmountInUSD" into numeric format
df_start["AmountInUSD"] = df_start["AmountInUSD"].apply(lambda x: float(str(x).replace(",",
""")))
df_start["AmountInUSD"] = pd.to_numeric(df_start["AmountInUSD"])
start_fund=df_start.groupby('StartupName')['AmountInUSD'].sum().sort_values(ascending=False)
[0:5]
startup=start_fund.index
for i in startup:
    print(i)
```

```
Paytm
Flipkart
Ola
Snapdeal
Oyo
```

funding rounds

Problem Statement : Find the top 5 startups who received the most number of funding rounds. That means, startups which got fundings maximum number of times. Print the startup name in descending order with respect to the number of funding round as integer value. Note: Ola, Flipkart, Oyo, Paytm are important startups, so correct their names. There are many errors in startup names, ignore correcting all, just handle important ones.

In [10]:
```
# Open and read data file as specified in the question
# Print the required output in given formatimport pandas as pd
import pandas as pd
import numpy as np
import csv
df_start=pd.read_csv("startup_funding.csv", encoding='utf-8')

df_start['StartupName'].replace('Olacabs','Ola',inplace=True)
df_start['StartupName'].replace('Ola Cabs','Ola',inplace=True)
df_start['StartupName'].replace('Flipkart.com','Flipkart',inplace=True)
df_start['StartupName'].replace('Paytm Marketplace','Paytm',inplace=True)
df_start['StartupName'].replace('Oyo Rooms','Oyo',inplace=True)
df_start['StartupName'].replace('Oyorooms','Oyo',inplace=True)
df_start['StartupName'].replace('OyoRooms','Oyo',inplace=True)
df_start['StartupName'].replace('OYO Rooms','Oyo',inplace=True)
amount=df_start['StartupName'].value_counts()
a=amount.index
b=amount.values
np_b=np.array(b)
np_b=np.sort(b)[::-1]

for i in range(5):
    print(a[i],np_b[i])
```

```
Ola 9
Swiggy 7
UrbanClap 6
Oyo 6
Paytm 6
```

top investors

Problem Statement : Find the Investors who have invested maximum number of times. Print the investor name and number of times invested as integer value. Note: In startup, multiple investors might have invested. So consider each investor for that startup. Ignore the undisclosed investors.

In [4]:
```
import numpy as np
import pandas as pd
import csv
import matplotlib.pyplot as plt
df=pd.read_csv('startup_funding.csv')
df.InvestorsName.fillna('',inplace=True)
```