Dropping not available values in the columns “InvestorsName” and ”StartupName” and “InvestmentType” using dropna() method

jupyter 4th Last Checkpoint: 18 hours ago (autosaved)

Logout

File Edit View Insert Cell Kernel Widgets Help Trusted Python 3 (ipykernel)

Run Code

```
[2372 rows x 10 columns]
```

In [56]:
```
df.dropna(subset=["InvestmentType","InvestorsName","StartupName"],inplace=True)
df.reset_index(drop = True, inplace = True)
print(df)
```

```
       SNo        Date       StartupName   IndustryVertical  \
0        0  01/08/2017          TouchKin         Technology
1        1  02/08/2017           Ethinos         Technology
2        2  02/08/2017      Leverage Edu  Consumer Internet
3        3  02/08/2017              Zepo  Consumer Internet
4        4  02/08/2017      Click2Clinic  Consumer Internet
...    ...         ...               ...                ...
2358  2367  29/01/2015        Printvenue                NaN
2359  2368  29/01/2015          Graphene                NaN
2360  2369  30/01/2015    Mad Street Den                NaN
2361  2370  30/01/2015         Simplotel                NaN
2362  2371  31/01/2015  couponmachine.in                NaN

                                    SubVertical CityLocation  \
0                         Predictive Care Platform    Bangalore
1                         Digital Marketing Agency       Mumbai
2     Online platform for Higher Education Services    New Delhi
3                           DIY Ecommerce platform       Mumbai
4                    healthcare service aggregator    Hyderabad
...                                            ...          ...
2358                                         NaN          NaN
2359                                         NaN          NaN
2360                                         NaN          NaN
2361                                         NaN          NaN
2362                                         NaN          NaN

                                      InvestorsName  InvestmentType  \
0                                       Kae Capital  Private Equity
1                           Triton Investment Advisors  Private Equity
2     Kashyap Deorah, Anand Sankeshwar, Deepak Jain,...    Seed Funding
3     Kunal Shah, LetsVenture, Anupam Mittal, Hetal ...    Seed Funding
4                          Narottam Thudi, Shireesh Palle    Seed Funding
...                                                  ...             ...
2358                       Asia Pacific Internet Group  Private Equity
```

Correcting the spelling mistake of startup names and ignoring ”undisclosed investors” in column “InvestorsName”

In [57]:
```
df["StartupName"].replace("Flipkart.com","Flipkart",inplace=True)
df["StartupName"].replace("Ola Cabs","Ola",inplace=True)
df["StartupName"].replace("Olacabs","Ola",inplace=True)
df["StartupName"].replace("OlaCabs","Ola",inplace=True)
df["StartupName"].replace("Oyo Rooms","Oyo",inplace=True)
df["StartupName"].replace("OyoRooms","Oyo",inplace=True)
df["StartupName"].replace("Oyorooms","Oyo",inplace=True)
df["StartupName"].replace("OYO Rooms","Oyo",inplace=True)
df["StartupName"].replace("Paytm Marketplace","Paytm",inplace=True)
df = df[df.InvestorsName != 'Undisclosed Investors']
df = df[df.InvestorsName != 'Undisclosed investors']
df = df[df.InvestorsName != 'undisclosed investors']
df = df[df.InvestorsName != 'undisclosed investor']
print(df)
```

```
       SNo        Date       StartupName   IndustryVertical  \
0        0  01/08/2017          TouchKin         Technology
1        1  02/08/2017           Ethinos         Technology
2        2  02/08/2017      Leverage Edu  Consumer Internet
3        3  02/08/2017              Zepo  Consumer Internet
4        4  02/08/2017      Click2Clinic  Consumer Internet
...    ...         ...               ...                ...
2358  2367  29/01/2015        Printvenue                NaN
2359  2368  29/01/2015          Graphene                NaN
2360  2369  30/01/2015    Mad Street Den                NaN
2361  2370  30/01/2015         Simplotel                NaN
2362  2371  31/01/2015  couponmachine.in                NaN

                                    SubVertical CityLocation  \
0                         Predictive Care Platform    Bangalore
1                         Digital Marketing Agency       Mumbai
2     Online platform for Higher Education Services    New Delhi
3                           DIY Ecommerce platform       Mumbai
4                     healthcare service aggregator    Hyderabad
...                                             ...          ...
2358                                            NaN          NaN
2359                                            NaN          NaN
2360                                            NaN          NaN
2361                                            NaN          NaN
2362                                            NaN          NaN
```

Correcting the spelling mistakes in the column “InvestmentType” and selecting only “Crowd Funding” and “Seed Funding” investors

jupyter 4th Last Checkpoint: 18 hours ago (autosaved) Logout

File Edit View Insert Cell Kernel Widgets Help Trusted Python 3 (ipykernel)

Run Code

In [58]:
```
df["InvestmentType"].replace("Crowd funding","Crowd Funding",inplace=True)
df["InvestmentType"].replace("PrivateEquity","Private Equity",inplace=True)
df["InvestmentType"].replace("SeedFunding","Seed Funding",inplace=True)

df=df[(df["InvestmentType"]=="Crowd Funding") | (df["InvestmentType"]=="Seed Funding")]

print(df.head())
```

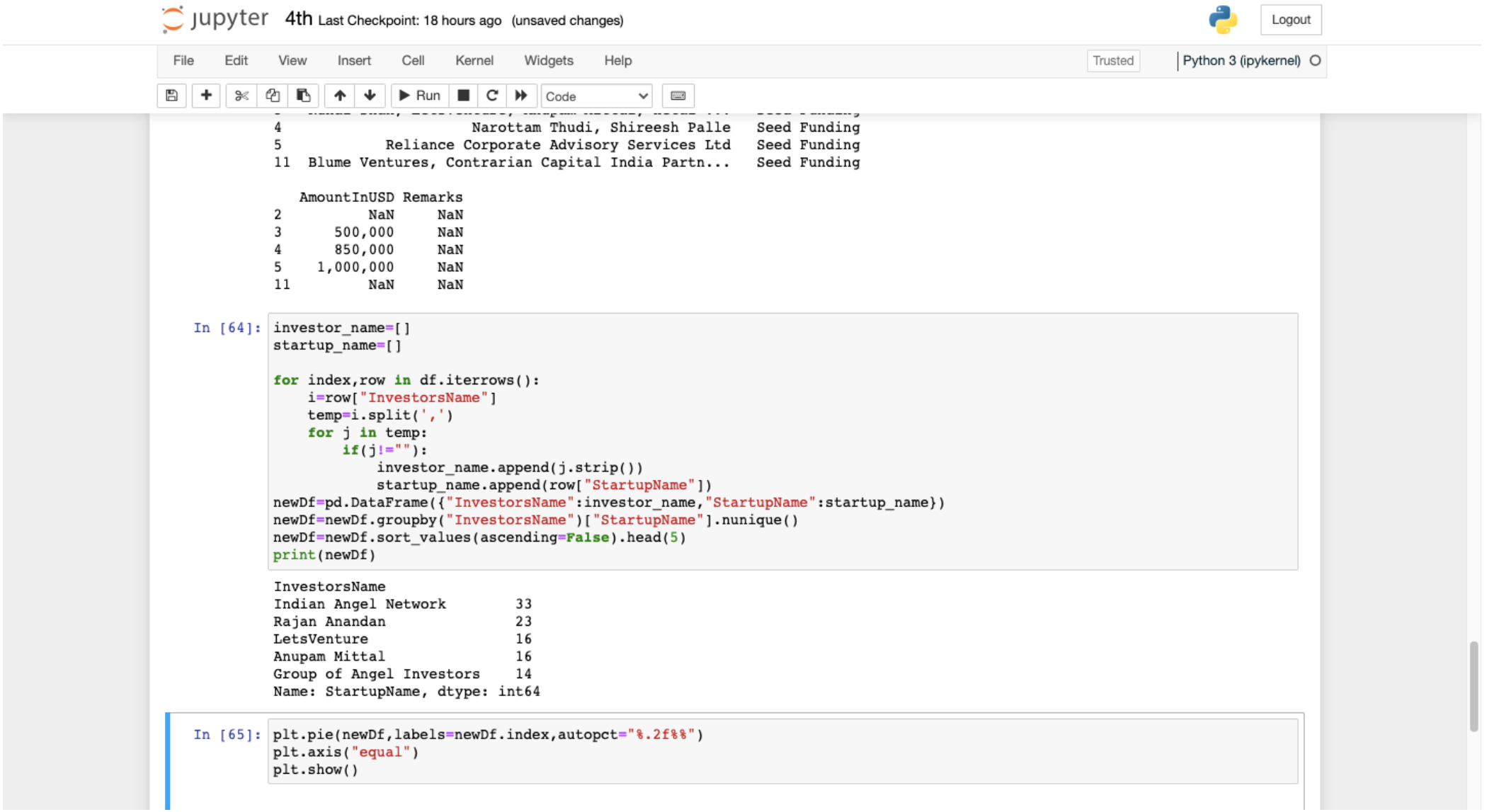
```
    SNo        Date    StartupName   IndustryVertical  \
2     2  02/08/2017   Leverage Edu  Consumer Internet
3     3  02/08/2017           Zepo  Consumer Internet
4     4  02/08/2017   Click2Clinic  Consumer Internet
5     5  01/07/2017  Billion Loans  Consumer Internet
11   11  06/07/2017         Minjar         Technology

                                SubVertical CityLocation  \
2   Online platform for Higher Education Services    New Delhi
3                         DIY Ecommerce platform       Mumbai
4                  healthcare service aggregator    Hyderabad
5                  Peer to Peer Lending platform    Bangalore
11                     Cloud Solutions provider    Bangalore

                                    InvestorsName InvestmentType  \
2   Kashyap Deorah, Anand Sankeshwar, Deepak Jain,...   Seed Funding
3   Kunal Shah, LetsVenture, Anupam Mittal, Hetal ...   Seed Funding
4                     Narottam Thudi, Shireesh Palle   Seed Funding
5           Reliance Corporate Advisory Services Ltd   Seed Funding
11  Blume Ventures, Contrarian Capital India Partn...   Seed Funding

   AmountInUSD Remarks
2          NaN     NaN
3      500,000     NaN
4      850,000     NaN
5    1,000,000     NaN
11         NaN     NaN
```

Used iterrow() method to iterate through the each rows of the dataframe and then splitting the investors names and append Them to investor_names list and also corresponding startup name to startup_name list and then created a new Data frame having ”InvestorsName” and “StartupName” columns and then using nunique() method to get the count of Unique values in column “StartupName” for each investorsName and then sorted the dataframe by values in descending order



Top 5 Investors of type Crowd Funding and Seed Funding

jupyter 4th Last Checkpoint: 18 hours ago (autosaved)

Logout

File Edit View Insert Cell Kernel Widgets Help

Trusted Python 3 (ipykernel)

Run Code

```
Rajan Anandan                23
LetsVenture                  16
Anupam Mittal                16
Group of Angel Investors     14
Name: StartupName, dtype: int64
```

In [62]:
```
plt.pie(newDf,labels=newDf.index,autopct="%.2f%%")
plt.axis("equal")
plt.show()
```