```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
df = pd.read_csv('startup_funding.csv')
df['InvestmentType'].replace('SeedFunding','Seed Funding',inplace = True)
df['InvestmentType'].replace('Crowd funding','Crowd Funding',inplace = True)
df['InvestmentType'].replace('PrivateEquity','Private Equity',inplace = True)
df['AmountInUSD'].fillna('0',inplace = True) # nan handling of funding column

x = dict.fromkeys(['OYO Rooms', 'Oyo Rooms','OyoRooms','Oyorooms'], 'Oyo')
y = dict.fromkeys(['Ola Cabs','Olacabs'],'Ola')
df['StartupName']=df['StartupName'].replace(x)
df['StartupName']=df['StartupName'].replace(y)
df['StartupName']=df['StartupName'].replace({"Flipkart.com":"Flipkart"})
df['StartupName']=df['StartupName'].replace({'Paytm Marketplace':'Paytm'})
df['InvestorsName'].fillna('0',inplace = True) # unknown investor names replaced by
'0'

d = {}
# d is a dictionary which uses investors as keys for which value is another
dictionary with
# startup names as keys and number of investments made in that startup as value
# for ex 'Sequoia Capital': {'Ola':5, 'Flipkart':2} could be a key value pair in d
(so Sequoia would have invested
# 5 times in Ola and twice in Flipkart)

for _, row in df.iterrows():
    x = row['InvestorsName'].strip().rstrip(',') # trailing comma caused ' ' to
appear as investor name so it had to be removed
    itype = row['InvestmentType']
    if itype=='Seed Funding' or itype=='Crowd Funding':
        startup = row['StartupName'].strip()
        names = x.split(',')
        for name in names:
            if 'undisclosed' in name.lower(): #ignoring undisclosed investors
                break
            else:
                name = name.strip()
                if name in d:
                    if startup in d[name]:
                        d[name][startup]+=1
                    else:
                        d[name][startup]=1
                else:
                    d[name]={}
                    d[name][startup]=1


d2 = {} #counts number of startups an investor has invested in, as is required by
question
# ex: 'Sequoia Capital': 2 could be a key value pair in d2 corresponding to example
key-value pair in d
# i.e. Sequoia Capital invested in 2 startups (ola and flipkart)

for investor in d:
    d2[investor] = len(d[investor])

d2 = sorted(d2.items(), key=lambda x: x[1], reverse=True)[:5] # d2 is now a list of
tuples (investor, count)
```

```
# sorted in descending order of count

investors = [t[0] for t in d2]
counts = [t[1] for t in d2]
for i in range(len(d2)):
    print(investors[i],counts[i])
plt.plot(investors,counts)
plt.scatter(investors,counts)
plt.xticks(rotation = 45, horizontalalignment = 'center')
plt.title('Seed Funding and Crowd Funding by Investors')
plt.xlabel('Investor')
plt.ylabel('Number of startups funded')
plt.show()
```