Module 6

- By Suprateek Halsana

In [143]:
```python
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
from sklearn.cluster import KMeans
import warnings
warnings.filterwarnings('ignore')
sns.set(style='darkgrid')
```

Problem [6.1]

In [144]:
```python
# Dataframe Created to store Close Price of 30 Stocks
data=pd.DataFrame()
```

In [145]:
```python
# stocks of 10 Large Cap Companies

data['infy']=pd.read_csv(r"C:\Users\Suprateek Halsana\Documents\Python Scripts\Aspiring Mind Internship\Prerequisites\Large_Cap\Large_Cap\INFY.csv")['Close Price']
data['ioc']=pd.read_csv(r"C:\Users\Suprateek Halsana\Documents\Python Scripts\Aspiring Mind Internship\Prerequisites\Large_Cap\Large_Cap\IOC.csv")['Close Price']
data['itc']=pd.read_csv(r"C:\Users\Suprateek Halsana\Documents\Python Scripts\Aspiring Mind Internship\Prerequisites\Large_Cap\Large_Cap\ITC.csv")['Close Price']
data['lt']=pd.read_csv(r"C:\Users\Suprateek Halsana\Documents\Python Scripts\Aspiring Mind Internship\Prerequisites\Large_Cap\Large_Cap\LT.csv")['Close Price']
data['mm']=pd.read_csv(r"C:\Users\Suprateek Halsana\Documents\Python Scripts\Aspiring Mind Internship\Prerequisites\Large_Cap\Large_Cap\M&M.csv")['Close Price']
data['maruti']=pd.read_csv(r"C:\Users\Suprateek Halsana\Documents\Python Scripts\Aspiring Mind Internship\Prerequisites\Large_Cap\Large_Cap\MARUTI.csv")['Close Price']
data['reliance']=pd.read_csv(r"C:\Users\Suprateek Halsana\Documents\Python Scripts\Aspiring Mind Internship\Prerequisites\Large_Cap\Large_Cap\RELIANCE.csv")['Close Price']
data['sunpharma']=pd.read_csv(r"C:\Users\Suprateek Halsana\Documents\Python Scripts\Aspiring Mind Internship\Prerequisites\Large_Cap\Large_Cap\SUNPHARMA.csv")['Close Price']
data['tcs']=pd.read_csv(r"C:\Users\Suprateek Halsana\Documents\Python Scripts\Aspiring Mind Internship\Prerequisites\Large_Cap\Large_Cap\TCS.csv")['Close Price']
data['titan']=pd.read_csv(r"C:\Users\Suprateek Halsana\Documents\Python Scripts\Aspiring Mind Internship\Prerequisites\Large_Cap\Large_Cap\TITAN.csv")['Close Price']
```

In [146]:
```python
# stocks of 10 Mid Cap Companies

data['idbi']=pd.read_csv(r"C:\Users\Suprateek Halsana\Documents\Python Scripts\Aspiring Mind Internship\Prerequisites\Mid_Cap\Mid_Cap\IDBI.csv")['Close Price']
data['igl']=pd.read_csv(r"C:\Users\Suprateek Halsana\Documents\Python Scripts\Aspiring Mind Internship\Prerequisites\Mid_Cap\Mid_Cap\IGL.csv")['Close Price']
data['jindals']=pd.read_csv(r"C:\Users\Suprateek Halsana\Documents\Python Scripts\Aspiring Mind Internship\Prerequisites\Mid_Cap\Mid_Cap\JINDALSTEL.csv")['Close Price']
data['jublfood']=pd.read_csv(r"C:\Users\Suprateek Halsana\Documents\Python Scripts\Aspiring Mind Internship\Prerequisites\Mid_Cap\Mid_Cap\JUBLFOOD.csv")['Close Price']
data['mindtree']=pd.read_csv(r"C:\Users\Suprateek Halsana\Documents\Python Scripts\Aspiring Mind Internship\Prerequisites\Mid_Cap\Mid_Cap\MINDTREE.csv")['Close Price']
data['mrpl']=pd.read_csv(r"C:\Users\Suprateek Halsana\Documents\Python Scripts\Aspiring Mind Internship\Prerequisites\Mid_Cap\Mid_Cap\MRPL.csv")['Close Price']
data['muthootfin']=pd.read_csv(r"C:\Users\Suprateek Halsana\Documents\Python Scripts\Aspiring Mind Internship\Prerequisites\Mid_Cap\Mid_Cap\MUTHOOTFIN.csv")['Close Price']
data['nbcc']=pd.read_csv(r"C:\Users\Suprateek Halsana\Documents\Python Scripts\Aspiring Mind Internship\Prerequisites\Mid_Cap\Mid_Cap\NBCC.csv")['Close Price']
data['pnb']=pd.read_csv(r"C:\Users\Suprateek Halsana\Documents\Python Scripts\Aspiring Mind Internship\Prerequisites\Mid_Cap\Mid_Cap\PNB.csv")['Close Price']
data['rblbank']=pd.read_csv(r"C:\Users\Suprateek Halsana\Documents\Python Scripts\Aspiring Mind Internship\Prerequisites\Mid_Cap\Mid_Cap\RBLBANK.csv")['Close Price']
```

In [147]:
```python
# stocks of 10 Small Cap Companies

data['luxind']=pd.read_csv(r"C:\Users\Suprateek Halsana\Documents\Python Scripts\Aspiring Mind Internship\Prerequisites\Small_Cap\Small_Cap\LUXIND.csv")['Close Price']
data['mindaind']=pd.read_csv(r"C:\Users\Suprateek Halsana\Documents\Python Scripts\Aspiring Mind Internship\Prerequisites\Small_Cap\Small_Cap\MINDAIND.csv")['Close Price']
data['ncc']=pd.read_csv(r"C:\Users\Suprateek Halsana\Documents\Python Scripts\Aspiring Mind Internship\Prerequisites\Small_Cap\Small_Cap\NCC.csv")['Close Price']
data['pvr']=pd.read_csv(r"C:\Users\Suprateek Halsana\Documents\Python Scripts\Aspiring Mind Internship\Prerequisites\Small_Cap\Small_Cap\PVR.csv")['Close Price']
data['raymond']=pd.read_csv(r"C:\Users\Suprateek Halsana\Documents\Python Scripts\Aspiring Mind Internship\Prerequisites\Small_Cap\Small_Cap\RAYMOND.csv")['Close Price']
data['rcom']=pd.read_csv(r"C:\Users\Suprateek Halsana\Documents\Python Scripts\Aspiring Mind Internship\Prerequisites\Small_Cap\Small_Cap\RCOM.csv")['Close Price']
data['sonatsoftw']=pd.read_csv(r"C:\Users\Suprateek Halsana\Documents\Python Scripts\Aspiring Mind Internship\Prerequisites\Small_Cap\Small_Cap\SONATSOFTW.csv")['Close Price']
data['suzlon']=pd.read_csv(r"C:\Users\Suprateek Halsana\Documents\Python Scripts\Aspiring Mind Internship\Prerequisites\Small_Cap\Small_Cap\SUZLON.csv")['Close Price']
data['venkeys']=pd.read_csv(r"C:\Users\Suprateek Halsana\Documents\Python Scripts\Aspiring Mind Internship\Prerequisites\Small_Cap\Small_Cap\VENKEYS.csv")['Close Price']
data['vipind']=pd.read_csv(r"C:\Users\Suprateek Halsana\Documents\Python Scripts\Aspiring Mind Internship\Prerequisites\Small_Cap\Small_Cap\VIPIND.csv")['Close Price']
```

In [148]:
```python
# As the data contains missing values
# So We Shall Remove the missing values by dropna function
print('### Before Null Removal ###\n')
print(data.isnull().sum())

data.dropna(inplace=True)


print('\n### After Null Removal ###')
data.isnull().sum()
```

```
### Before Null Removal ###

infy          0
ioc           1
itc           0
lt            2
mm            0
maruti        0
reliance      1
sunpharma     1
tcs           0
titan         0
idbi          2
igl           0
jindals       2
jublfood      1
mindtree      2
mrpl          2
muthootfin    0
nbcc          2
pnb           2
rblbank       2
luxind        2
mindaind      1
ncc           0
pvr           1
raymond       2
rcom          2
sonatsoftw    2
suzlon        2
venkeys       2
vipind        2
dtype: int64

### After Null Removal ###
```

Out[148]:
```
infy          0
ioc           0
itc           0
lt            0
mm            0
maruti        0
reliance      0
sunpharma     0
tcs           0
titan         0
idbi          0
igl           0
jindals       0
jublfood      0
mindtree      0
mrpl          0
muthootfin    0
nbcc          0
pnb           0
rblbank       0
luxind        0
mindaind      0
ncc           0
pvr           0
raymond       0
rcom          0
sonatsoftw    0
suzlon        0
venkeys       0
vipind        0
dtype: int64
```

In [149]:
```python
data
```

Out[149]:

	infy	ioc	itc	lt	mm	maruti	reliance	sunpharma	tcs	titan	...	luxind	mindaind	ncc	pvr
0	951.55	442.10	277.95	1742.40	1375.65	6823.90	1344.10	654.45	2365.10	483.65	...	819.60	517.90	97.35	1524.95
1	955.00	446.60	283.45	1740.70	1366.35	6953.95	1356.30	652.35	2429.15	488.30	...	817.50	545.60	100.40	1539.00
2	952.80	444.25	281.65	1755.45	1378.80	6958.20	1353.10	650.90	2455.35	481.75	...	819.80	560.10	101.45	1530.30
3	961.75	439.90	277.90	1722.60	1352.85	6831.05	1327.35	653.95	2536.20	471.65	...	820.05	545.25	97.00	1515.05
4	957.95	435.40	286.20	1724.55	1333.40	6790.55	1318.85	653.15	2507.15	471.15	...	817.80	560.80	95.45	1513.55
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
489	723.60	156.70	304.25	1366.60	645.90	6710.00	1384.90	453.15	2260.35	1145.05	...	1314.40	351.60	97.55	1802.30
490	718.40	153.45	307.00	1360.35	645.80	6709.65	1343.50	448.20	2215.40	1082.85	...	1268.50	347.05	96.75	1780.10
491	724.55	152.20	301.90	1356.45	638.50	6702.00	1299.45	439.15	2132.00	1086.50	...	1248.20	350.35	93.70	1748.50
492	719.35	150.60	300.65	1355.50	636.00	6650.15	1256.45	439.35	2157.85	1088.70	...	1239.95	342.00	92.65	1758.45
493	721.05	147.60	299.85	1317.65	621.65	6624.95	1251.15	437.75	2151.95	1104.45	...	1211.55	345.15	93.60	1715.40

494 rows × 30 columns

In [150]:
```python
# list of Stocks
stock_list=list(data.columns)
```

Problem [6.2]

In [151]:
```python
# Dataframe Containing the Daily Return of 30 stocks
daily_ret=pd.DataFrame()
for i in stock_list:
    daily_ret[i]=data[i].pct_change()
    daily_ret[i][0]=0
```

In [152]:
```python
daily_ret
```

Out[152]:

	infy	ioc	itc	lt	mm	maruti	reliance	sunpharma	tcs	titan	...	luxind	mindaind
0	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	...	0.000000	0.00000
1	0.003626	0.010179	0.019788	-0.000976	-0.006760	0.019058	0.009077	-0.003209	0.027081	0.009614	...	-0.002562	0.05348
2	-0.002304	-0.005262	-0.006350	0.008474	0.009112	0.000611	-0.002359	-0.002223	0.010786	-0.013414	...	0.002813	0.02657
3	0.009393	-0.009792	-0.013314	-0.018713	-0.018821	-0.018273	-0.019030	0.004686	0.032928	-0.020965	...	0.000305	-0.02651
4	-0.003951	-0.010230	0.029867	0.001132	-0.014377	-0.005929	-0.006404	-0.001223	-0.011454	-0.001060	...	-0.002744	0.02851
...	...	...	...	...	...	...	...	...	...	...	...	...	...
489	-0.009852	0.003522	-0.002459	0.010276	0.000930	0.004003	-0.017000	-0.000221	0.009738	0.006062	...	-0.014471	0.00457
490	-0.007186	-0.020740	0.009039	-0.004573	-0.000155	-0.000052	-0.029894	-0.010924	-0.019886	-0.054321	...	-0.034921	-0.01294
491	0.008561	-0.008146	-0.016612	-0.002867	-0.011304	-0.001140	-0.032787	-0.020192	-0.037646	0.003371	...	-0.016003	0.00950
492	-0.007177	-0.010512	-0.004140	-0.000700	-0.003915	-0.007736	-0.033091	0.000455	0.012125	0.002025	...	-0.006610	-0.02383
493	0.002363	-0.019920	-0.002661	-0.027923	-0.022563	-0.003789	-0.004218	-0.003642	-0.002734	0.014467	...	-0.022904	0.00921

494 rows × 30 columns

In [153]:
```python
print('### Annual Return ###')
round(daily_ret.mean()*252,2)
```

```
### Annual Return ###
```

Out[153]:
```
infy           -0.02
ioc            -0.40
itc             0.06
lt             -0.09
mm             -0.27
maruti          0.01
reliance        0.09
sunpharma      -0.15
tcs             0.08
titan           0.47
idbi           -0.30
igl            -0.13
jindals         0.27
jublfood        0.27
mindtree        0.39
mrpl           -0.32
muthootfin    104.39
nbcc           -0.51
pnb            -0.24
rblbank         0.11
luxind          0.25
mindaind        0.08
ncc             0.08
pvr             0.11
raymond         0.07
rcom           -0.88
sonatsoftw      0.46
suzlon         -0.47
venkeys         0.38
vipind          0.48
dtype: float64
```

In [154]:
```python
print('### Volatility ###')
round(((daily_ret.std())*(252**0.5)),2)
```

```
### Volatility ###
```

Out[154]:
```
infy           0.42
ioc            0.50
itc            0.22
lt             0.31
mm             0.44
maruti         0.23
reliance       0.44
sunpharma      0.34
tcs            0.43
titan          0.32
idbi           0.45
igl            0.65
jindals        0.49
jublfood       0.50
mindtree       0.34
mrpl           0.37
muthootfin    18.28
nbcc           0.51
pnb            0.54
rblbank        0.26
luxind         0.34
mindaind       0.61
ncc            0.44
pvr            0.31
raymond        0.36
rcom           0.98
sonatsoftw     0.40
suzlon         0.59
venkeys        0.63
vipind         0.43
dtype: float64
```

Problem [6.3]

In [155]:
```python
returns = daily_ret.mean() * 252
volatility = data.pct_change().std() * (252**0.5)
returns.columns = ["Returns"]
volatility.columns = ["Volatility"]
#Concatenating the returns and variances into a single data-frame
ret_var = pd.concat([returns, volatility], axis = 1).dropna()
ret_var.columns = ["Annual Returns","Volatility"]
```

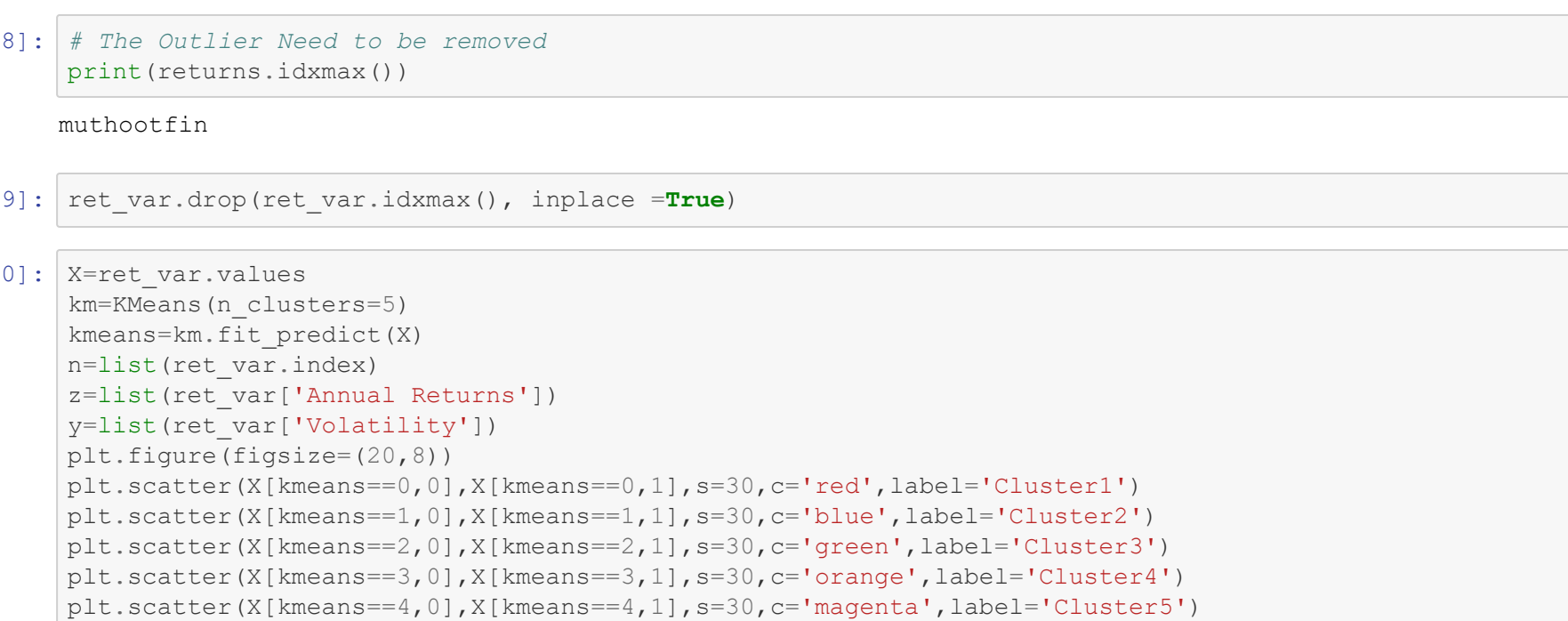
In [156]:
```python
ret_var
```

Out[156]:

	Annual Returns	Volatility
infy	-0.022626	0.418963
ioc	-0.395583	0.499514
itc	0.063885	0.223863
lt	-0.085730	0.314323
mm	-0.268613	0.440357
maruti	0.010174	0.225258
reliance	0.094928	0.439091
sunpharma	-0.146316	0.342426
tcs	0.080307	0.431387
titan	0.472152	0.324361
idbi	-0.295409	0.453787
igl	-0.131495	0.651128
jindals	0.274814	0.492980
jublfood	0.266973	0.501004
mindtree	0.391775	0.341485
mrpl	-0.322303	0.365898
muthootfin	104.385879	18.299167
nbcc	-0.514133	0.508672
pnb	-0.238955	0.545390
rblbank	0.113267	0.258732
luxind	0.254644	0.335409
mindaind	0.081559	0.607996
ncc	0.076986	0.441393
pvr	0.108310	0.310191
raymond	0.069879	0.363834
rcom	-0.879944	0.985366
sonatsoftw	0.455334	0.401937
suzlon	-0.467204	0.593556
venkeys	0.380938	0.627383
vipind	0.478693	0.432069

In [157]:
```python
X = ret_var.values
distorsions = []
for k in range(2, 20):
    k_means = KMeans(n_clusters=k)
    k_means.fit(X)
    distorsions.append(k_means.inertia_)
fig = plt.figure(figsize=(15, 5))
plt.plot(range(2, 20), distorsions)
plt.grid(True)
plt.title('Elbow curve')
```

Out[157]:
```
Text(0.5, 1.0, 'Elbow curve')
```



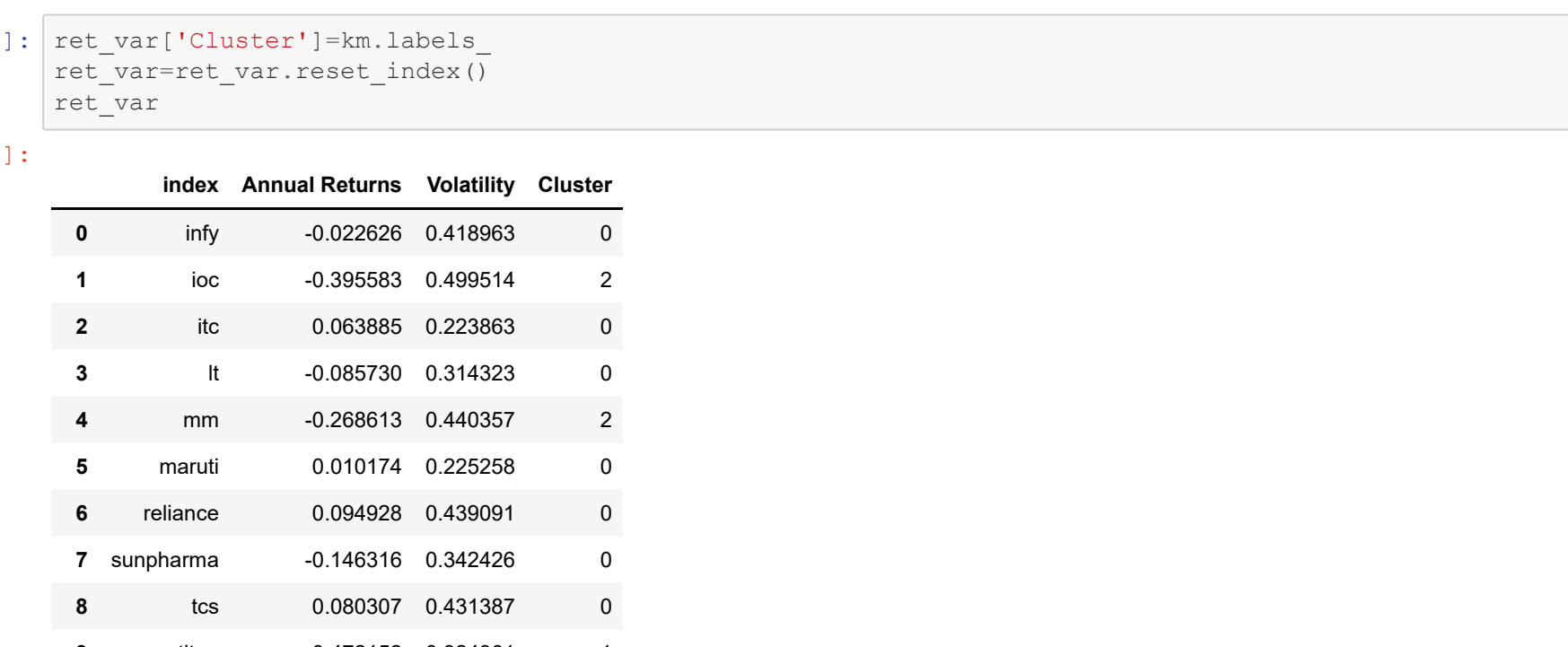
Problem [6.4]

In [158]:
```python
# The Outlier Need to be removed
print(returns.idxmax())
```

```
muthootfin
```

In [159]:
```python
ret_var.drop(ret_var.idxmax(), inplace =True)
```

In [160]:
```python
X=ret_var.values
km=KMeans(n_clusters=5)
kmeans=km.fit_predict(X)
n=list(ret_var.index)
z=list(ret_var['Annual Returns'])
y=list(ret_var['Volatility'])
plt.figure(figsize=(20,8))
plt.scatter(X[kmeans==0,0],X[kmeans==0,1],s=30,c='red',label='Cluster1')
plt.scatter(X[kmeans==1,0],X[kmeans==1,1],s=30,c='blue',label='Cluster2')
plt.scatter(X[kmeans==2,0],X[kmeans==2,1],s=30,c='green',label='Cluster3')
plt.scatter(X[kmeans==3,0],X[kmeans==3,1],s=30,c='orange',label='Cluster4')
plt.scatter(X[kmeans==4,0],X[kmeans==4,1],s=30,c='magenta',label='Cluster5')
plt.legend()
for i,txt in enumerate(n):
    plt.annotate(txt,(z[i],y[i]))
```



In [161]:
```python
ret_var['Cluster']=km.labels_
ret_var=ret_var.reset_index()
ret_var
```

Out[161]:

	index	Annual Returns	Volatility	Cluster
0	infy	-0.022626	0.418963	0
1	ioc	-0.395583	0.499514	2
2	itc	0.063885	0.223863	0
3	lt	-0.085730	0.314323	0
4	mm	-0.268613	0.440357	2
5	maruti	0.010174	0.225258	0
6	reliance	0.094928	0.439091	0
7	sunpharma	-0.146316	0.342426	0
8	tcs	0.080307	0.431387	0
9	titan	0.472152	0.324361	1
10	idbi	-0.295409	0.453787	2
11	igl	-0.131495	0.651128	2
12	jindals	0.274814	0.492980	4
13	jublfood	0.266973	0.501004	4
14	mindtree	0.391775	0.341485	1
15	mrpl	-0.322303	0.365898	2
16	nbcc	-0.514133	0.508672	2
17	pnb	-0.238955	0.545390	2
18	rblbank	0.113267	0.258732	0
19	luxind	0.254644	0.335409	1
20	mindaind	0.081559	0.607996	4
21	ncc	0.076986	0.441393	0
22	pvr	0.108310	0.310191	0
23	raymond	0.069879	0.363834	0
24	rcom	-0.879944	0.985366	3
25	sonatsoftw	0.455334	0.401937	1
26	suzlon	-0.467204	0.593556	2
27	venkeys	0.380938	0.627383	4
28	vipind	0.478693	0.432069	1

In [162]:

```
# Group by Cluster
g=ret_var.groupby(['Cluster'])

for name,group in g:
    print('\n Stocks of Cluster  :',name)
    print(list(group['index']),'\n')
```

```
 Stocks of Cluster  : 0
['infy', 'itc', 'lt', 'maruti', 'reliance', 'sunpharma', 'tcs', 'rblbank', 'ncc', 'pvr', 'raymond']


 Stocks of Cluster  : 1
['titan', 'mindtree', 'luxind', 'sonatsoftw', 'vipind']


 Stocks of Cluster  : 2
['ioc', 'mm', 'idbi', 'igl', 'mrpl', 'nbcc', 'pnb', 'suzlon']


 Stocks of Cluster  : 3
['rcom']


 Stocks of Cluster  : 4
['jindals', 'jublfood', 'mindaind', 'venkeys']
```