

50 Startup EDA

```
In [3]: import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
import numpy as np
```

```
In [4]: data = pd.read_csv('C:/Users/SW20407278/Desktop/Final AI/Hands-On/Use_Cases_EDA/50_Startups.csv')
```

```
In [5]: data.info()

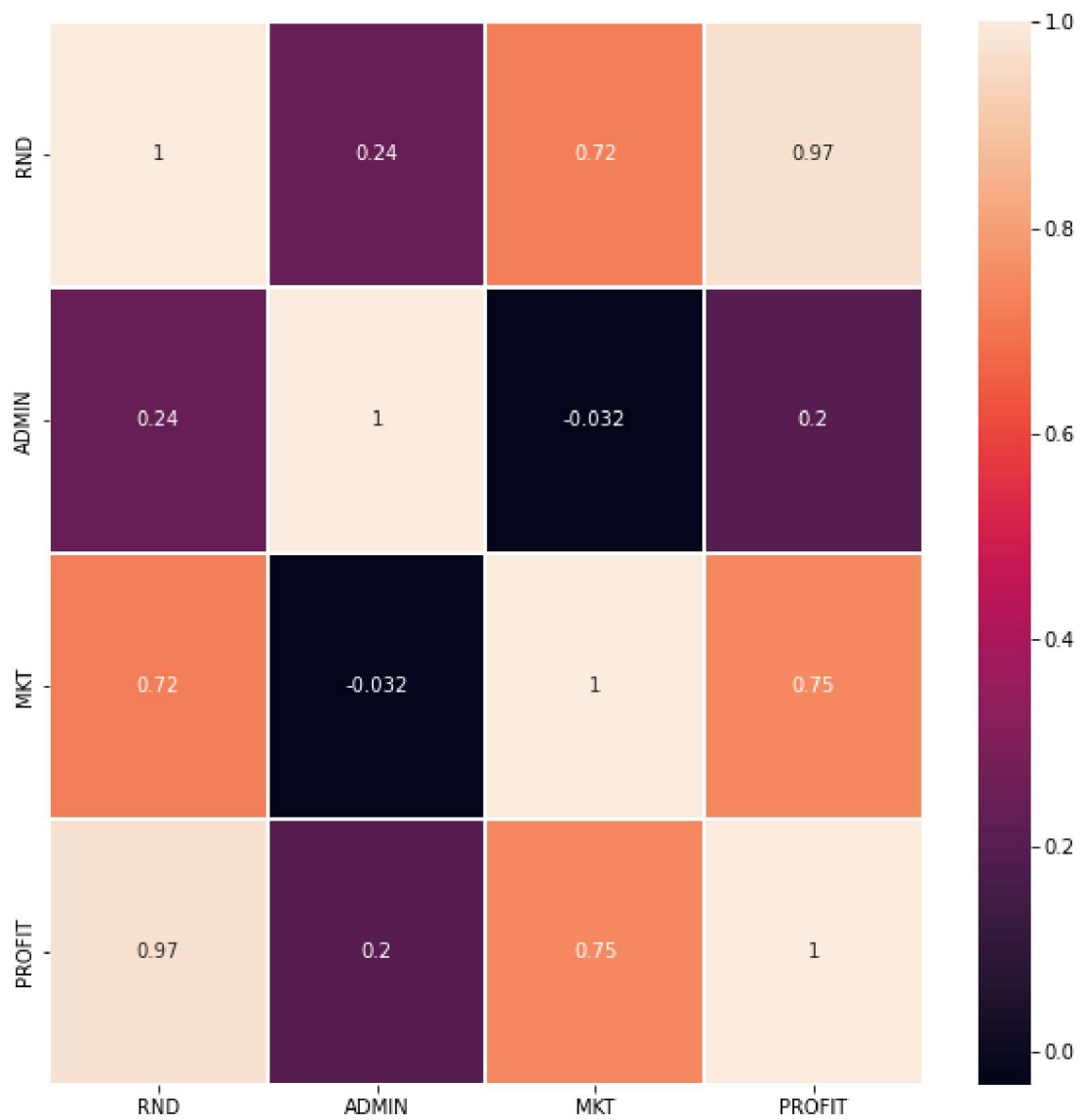
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 50 entries, 0 to 49
Data columns (total 5 columns):
 #   Column  Non-Null Count  Dtype  
---  --
 0   RND      50 non-null     float64
 1   ADMIN    50 non-null     float64
 2   MKT      50 non-null     float64
 3   STATE    50 non-null     object  
 4   PROFIT   50 non-null     float64
dtypes: float64(4), object(1)
memory usage: 2.1+ KB
```

```
In [7]: ### Shows statistical summary for the dataset
data.describe()
```

```
Out[7]:
```

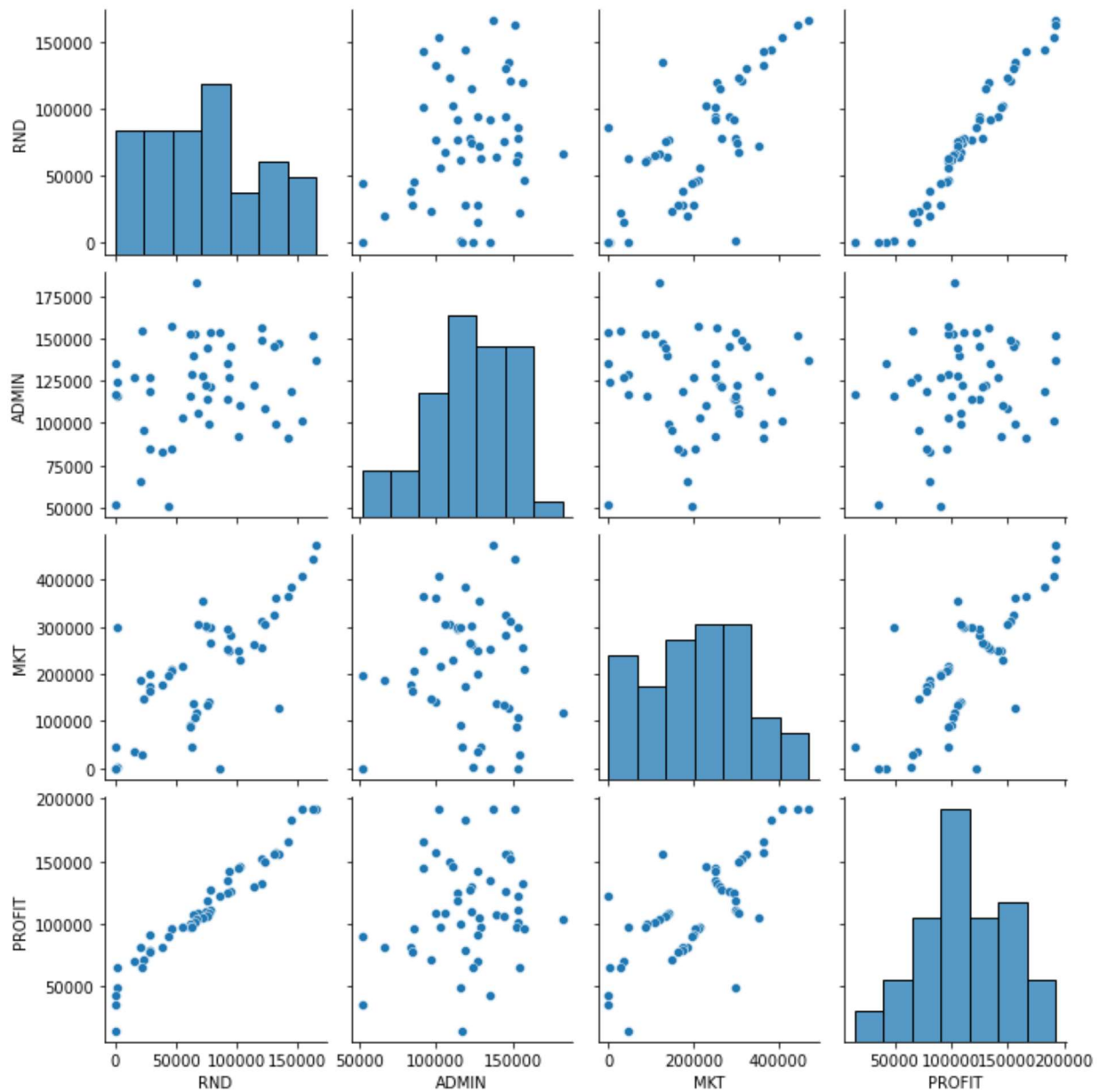
	RND	ADMIN	MKT	PROFIT
count	50.000000	50.000000	50.000000	50.000000
mean	73721.615600	121344.639600	211025.097800	112012.639200
std	45902.256482	28017.802755	122290.310726	40306.180338
min	0.000000	51283.140000	0.000000	14681.400000
25%	39936.370000	103730.875000	129300.132500	90138.902500
50%	73051.080000	122699.795000	212716.240000	107978.190000
75%	101602.800000	144842.180000	299469.085000	139765.977500
max	165349.200000	182645.560000	471784.100000	192261.830000

```
In [8]: plt.figure(figsize=(10,10))
sns.heatmap(data.corr(),annot=True, linewidth=2)
plt.show()
```



```
In [9]: sns.pairplot(data)
```

```
Out[9]: <seaborn.axisgrid.PairGrid at 0x1df838d37f0>
```

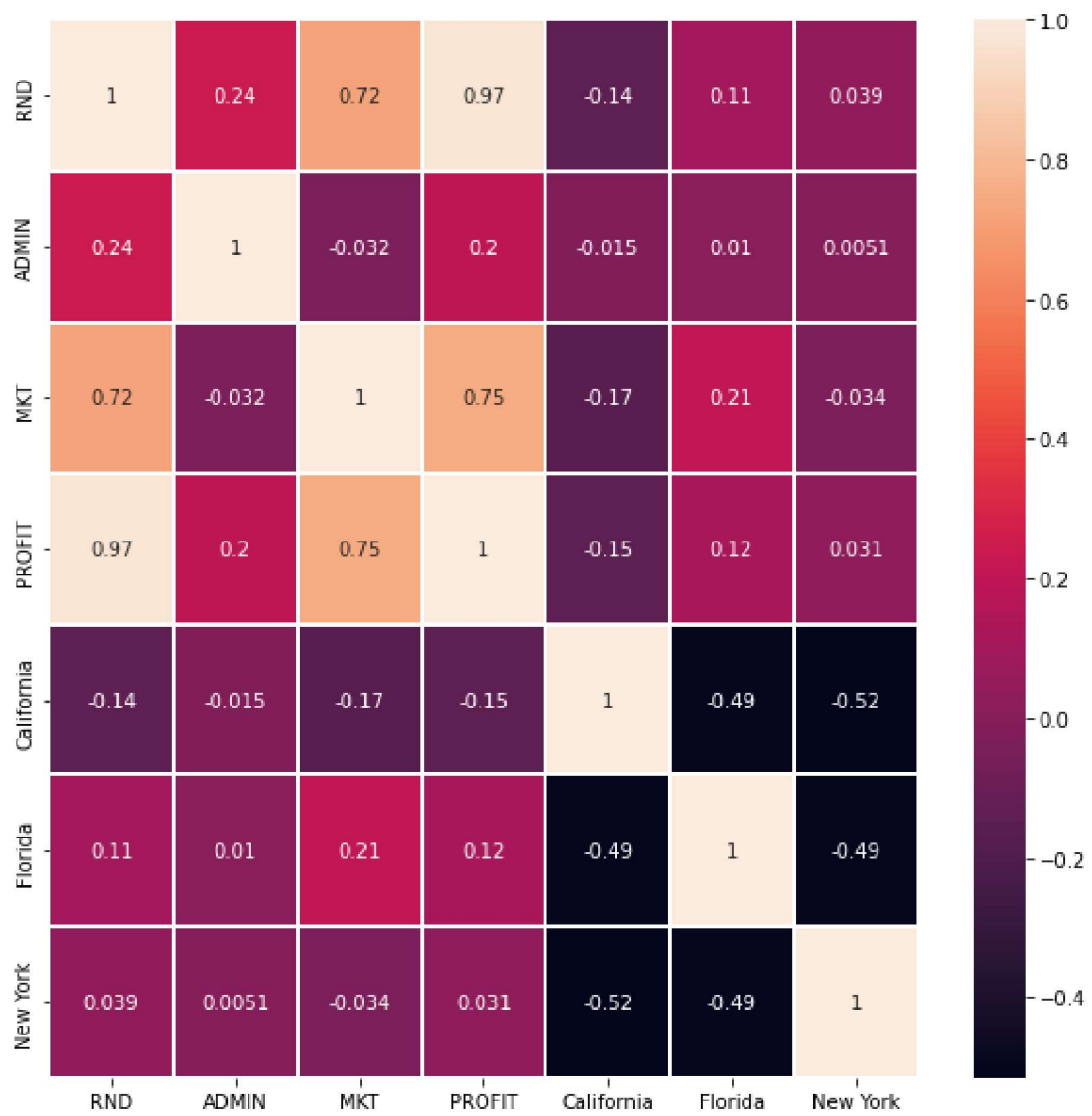


```
In [10]: dummy_state = pd.get_dummies(data.STATE)
df = data.join(dummy_state)
df.head()
```

```
Out[10]:
```

	RND	ADMIN	MKT	STATE	PROFIT	California	Florida	New York
0	165349.20	136897.80	471784.10	New York	192261.83	0	0	1
1	162597.70	151377.59	443898.53	California	191792.06	1	0	0
2	153441.51	101145.55	407934.54	Florida	191050.39	0	1	0
3	144372.41	118671.85	383199.62	New York	182901.99	0	0	1
4	142107.34	91391.77	366168.42	Florida	166187.94	0	1	0

```
In [11]: plt.figure(figsize=(10,10))
sns.heatmap(df.corr(),annot=True,linewidth=2)
plt.show()
```



In []: