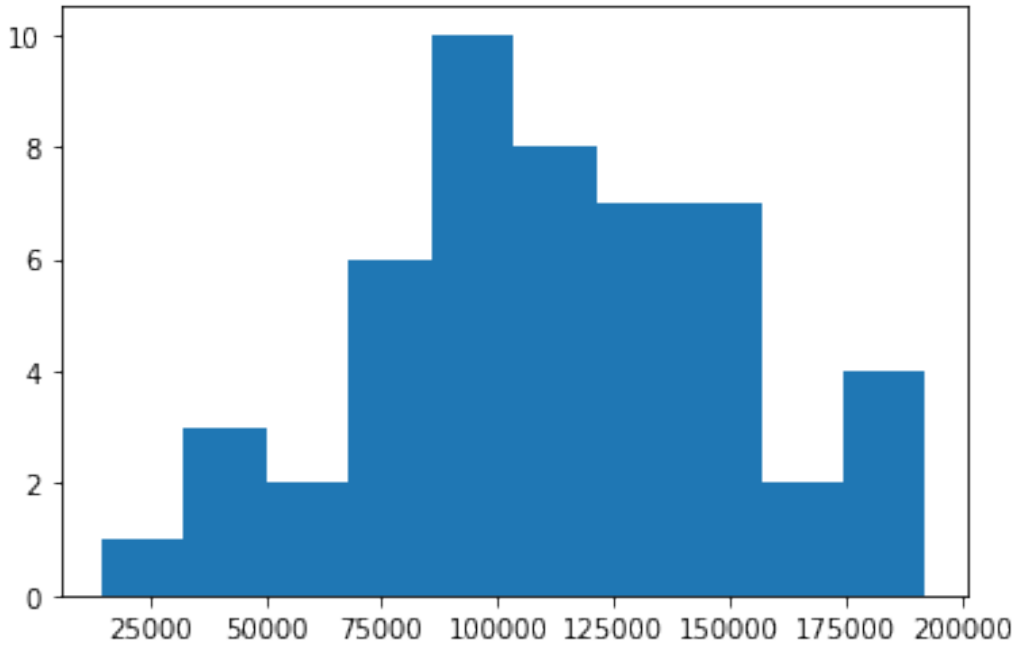
```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
import missingno as msno
%matplotlib inline

#Using same datasetset from Aditi Dhali ma'ams class
dataset =
pd.read_csv("https://raw.githubusercontent.com/arnob016/PyClass/master
/AI%20Lab%20Report/data.csv")
dataset.head()
```

	R&D Spend	Administration	Marketing Spend	State	Profit
0	165349.20	136897.80	471784.10	New York	192261.83
1	162597.70	151377.59	443898.53	California	191792.06
2	153441.51	101145.55	407934.54	Florida	191050.39
3	144372.41	118671.85	383199.62	New York	182901.99
4	142107.34	91391.77	366168.42	Florida	166187.94

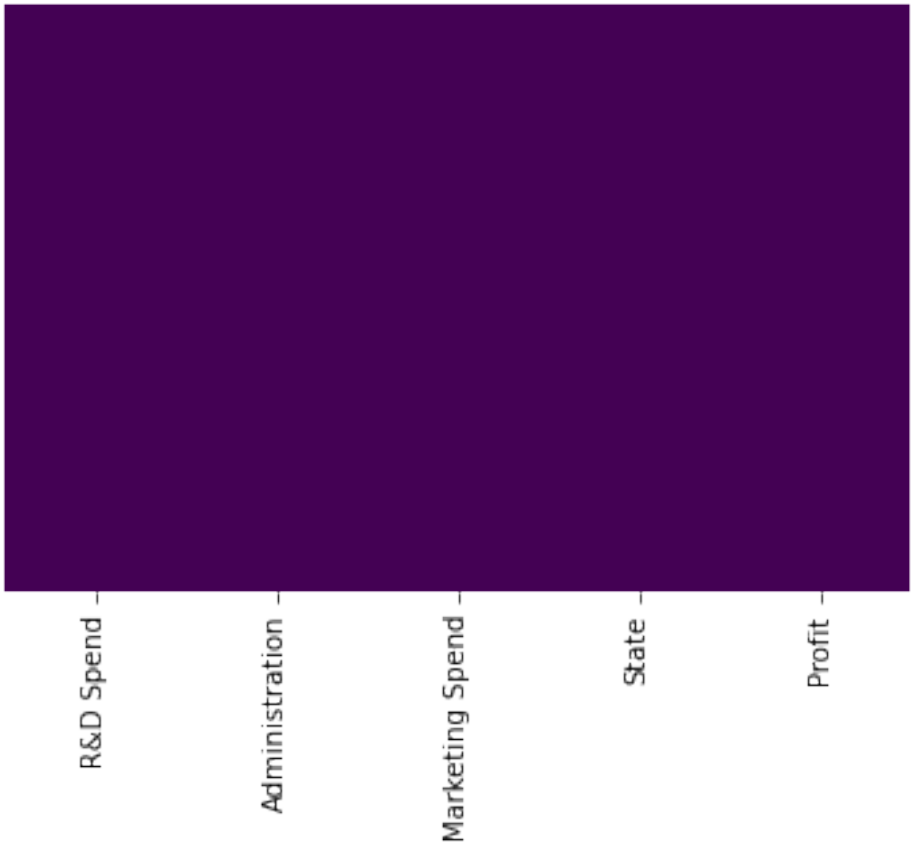
```
print(dataset['Profit'].head())
plt.hist(dataset['Profit'])
```

```
0    192261.83
1    191792.06
2    191050.39
3    182901.99
4    166187.94
Name: Profit, dtype: float64

(array([ 1.,  3.,  2.,  6., 10.,  8.,  7.,  7.,  2.,  4.]),
 array([ 14681.4  ,  32439.443,  50197.486,  67955.529,  85713.572,
        103471.615, 121229.658, 138987.701, 156745.744, 174503.787,
        192261.83 ]),
 <a list of 10 Patch objects>)
```




```
import seaborn as sns

sns.heatmap(dataset.isnull(),yticklabels = False, cbar = False, cmap =
'viridis')
```

```
<matplotlib.axes._subplots.AxesSubplot at 0x7ff1987c2430>
```




```
nullset =
pd.read_csv("https://raw.githubusercontent.com/arnob016/PyClass/master
/AI%20Lab%20Report/nullcopy.csv")
nullset.head(10)
```

	R&D Spend	Administration	Marketing Spend	State	Profit
0	165349.20	136897.80	471784.10	New York	192261.83
1	162597.70	151377.59	443898.53	California	191792.06
2	153441.51	101145.55	NaN	Florida	191050.39
3	144372.41	118671.85	NaN	New York	182901.99
4	142107.34	91391.77	NaN	Florida	166187.94
5	131876.90	99814.71	362861.36	New York	156991.12
6	NaN	NaN	NaN	California	156122.51
7	NaN	NaN	NaN	Florida	155752.60
8	NaN	NaN	NaN	New York	152211.77
9	NaN	NaN	NaN	California	149759.96

```
pd.isnull(nullset).head()
```

	R&D Spend	Administration	Marketing Spend	State	Profit
0	False	False	False	False	False
1	False	False	False	False	False
2	False	False	True	False	False

```
3      False           False             True  False   False
4      False           False             True  False   False
```

```
nulldata = nullset[pd.isnull(nullset).any(axis=1)]
nulldata.head()
```

```
   R&D Spend  Administration  Marketing Spend       State     Profit
2  153441.51       101145.55              NaN     Florida  191050.39
3  144372.41       118671.85              NaN    New York  182901.99
4  142107.34        91391.77              NaN     Florida  166187.94
6        NaN             NaN              NaN  California  156122.51
7        NaN             NaN              NaN     Florida  155752.60
```

```
nulldata.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 26 entries, 2 to 47
Data columns (total 5 columns):
 #   Column           Non-Null Count  Dtype  
---  ------           --------------  -----  
 0   R&D Spend        13 non-null     float64
 1   Administration   17 non-null     float64
 2   Marketing Spend  13 non-null     float64
 3   State            19 non-null     object 
 4   Profit           20 non-null     float64
dtypes: float64(4), object(1)
memory usage: 1.2+ KB
```

```
msno.bar(nulldata)
```

```
<matplotlib.axes._subplots.AxesSubplot at 0x7ff195dcc520>
```