

```
1 import pandas as pd
2 import matplotlib.pyplot as plt
3 import seaborn as sns
```

```
1 df = pd.read_csv('/content/financial_services_companies.csv')
2 df.head()
```

	Rank	Company	Industry	Revenue in (USD Million)	Net Income in (USD Millions)	Total Assest in (USD Millions)	Headquarters
0	1	Transamerica Corporation	Conglomerate	245510	42521	873	United States
1	2	Ping An Insurance Group	Insurance	191509	20738	1460	China
2	3	ICBC	Banking	182794	45783	5110	China
3	4	China Construction Bank	Banking	172000	39282	4311	China
4	5	Agricultural Bank of	Banking	152884	21202	4160	China

Next steps:

[Generate code with df](#)[View recommended plots](#)[New interactive sheet](#)

```
1 import matplotlib.pyplot as plt
2 # Check for missing values
3 print(df.isnull().sum())
```

```
Rank      0
Company   0
Industry   0
Revenue in (USD Million)  0
Net Income in (USD Millions)  0
Total Assest in (USD Millions)  0
Headquarters  0
dtype: int64
```

```
1 # Summary statistics
2 print(df.describe())
```

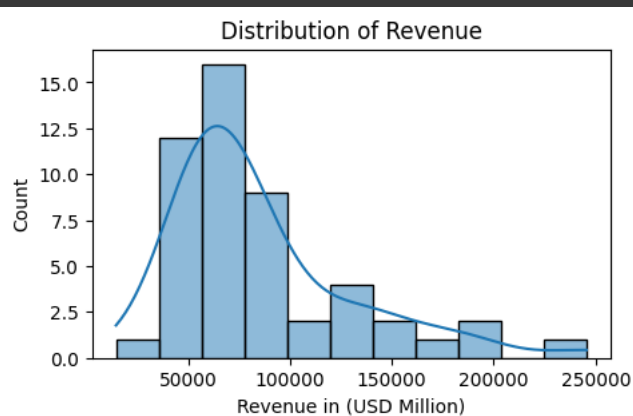
```
Rank Revenue in (USD Million) Net Income in (USD Millions) \
count  50.00000      50.000000      50.000000
mean   25.50000      85435.120000      9369.320000
std    14.57738      44689.154251     11101.187832
min     1.00000      14592.000000      169.000000
25%    13.25000      56189.250000      3030.250000
50%    25.50000      70736.000000      4963.000000
75%    37.75000      96284.250000     10660.750000
max    50.00000     245510.000000     45783.000000

Total Assest in (USD Millions)
count      50.00000
mean     1480.46000
std     1282.83769
min       13.00000
25%       588.00000
50%     1024.50000
75%     2183.75000
max     5110.00000
```

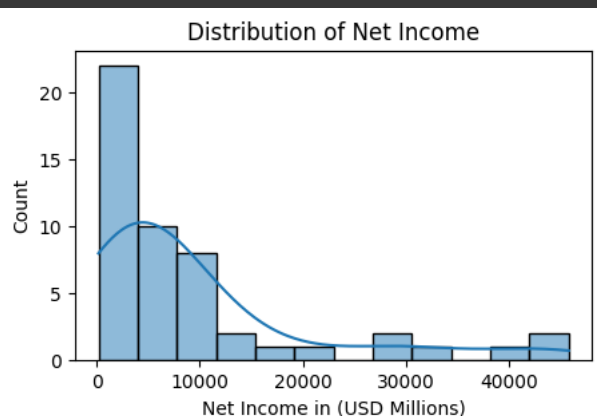
```
1 # Data types of each column
2 print(df.dtypes)
```

```
Rank      int64
Company   object
Industry   object
Revenue in (USD Million)  int64
Net Income in (USD Millions)  int64
Total Assest in (USD Millions)  int64
Headquarters  object
dtype: object
```

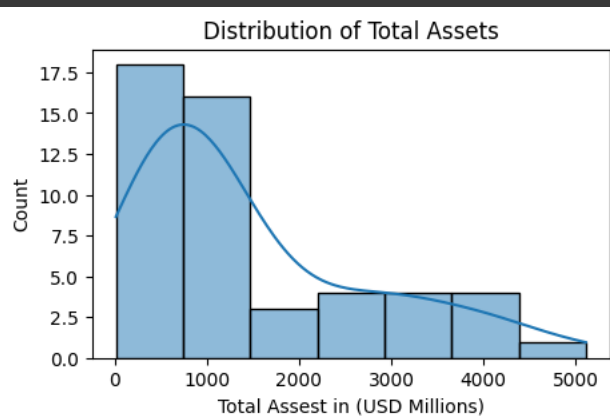
```
1 # Distribution of numerical features
2 plt.figure(figsize=(5, 3))
3 sns.histplot(df['Revenue in (USD Million)'], kde=True)
4 plt.title('Distribution of Revenue')
5 plt.show()
```



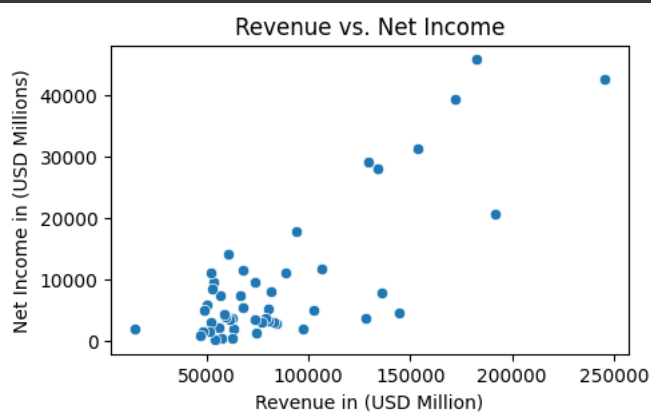
```
1 plt.figure(figsize=(5, 3))
2 sns.histplot(df['Net Income in (USD Millions)'], kde=True)
3 plt.title('Distribution of Net Income')
4 plt.show()
```



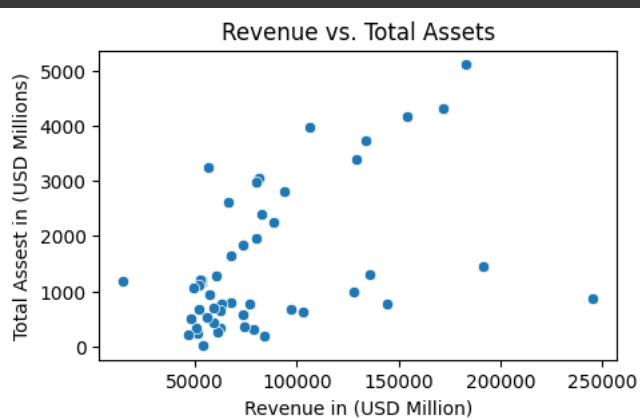
```
1 plt.figure(figsize=(5, 3))
2 sns.histplot(df['Total Assest in (USD Millions)'], kde=True)
3 plt.title('Distribution of Total Assets')
4 plt.show()
```



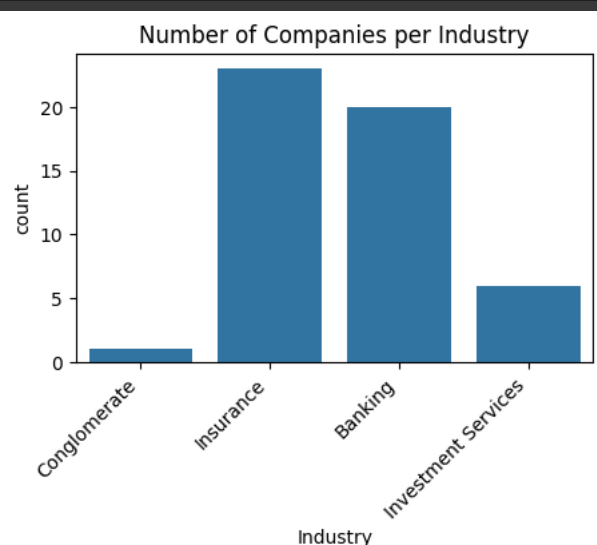
```
1 # Relationship between revenue and net income
2 plt.figure(figsize=(5, 3))
3 sns.scatterplot(x='Revenue in (USD Million)', y='Net Income in (USD Millions)', data=df)
4 plt.title('Revenue vs. Net Income')
5 plt.show()
```



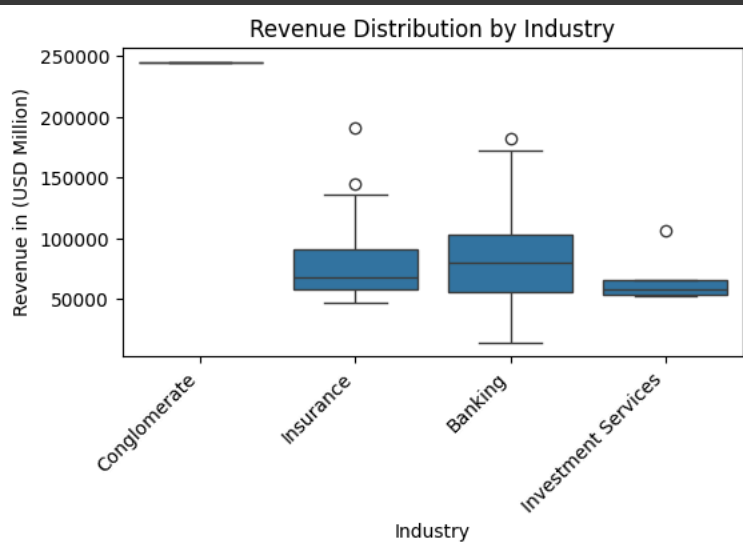
```
1 # Relationship between revenue and total assets
2 plt.figure(figsize=(5, 3))
3 sns.scatterplot(x='Revenue in (USD Million)', y='Total Assest in (USD Millions)', data=df)
4 plt.title('Revenue vs. Total Assets')
5 plt.show()
```



```
1 # Count of companies by industry
2 plt.figure(figsize=(5, 3))
3 sns.countplot(x='Industry', data=df)
4 plt.xticks(rotation=45, ha='right') # Rotate x-axis labels for better readability
5 plt.title('Number of Companies per Industry')
6 plt.show()
```



```
1 # Boxplot of revenue by industry
2 plt.figure(figsize=(6, 3))
3 sns.boxplot(x='Industry', y='Revenue in (USD Million)', data=df)
4 plt.xticks(rotation=45, ha='right')
5 plt.title('Revenue Distribution by Industry')
6 plt.show()
```



```
1 # Correlation matrix
2 correlation_matrix = df[['Revenue in (USD Million)', 'Net Income in (USD
3 Millions)', 'Total Assest in (USD Millions)']].corr()
4 plt.figure(figsize=(4, 3))
5 sns.heatmap(correlation_matrix, annot=True, cmap='coolwarm')
6 plt.title('Correlation Matrix')
7 plt.show()
```

