In [1]:

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
```

In [2]:

```
df =pd.read_csv("D:\\Future-500.csv")
df
```

Out[2]:

	ID	Name	Industry	Inception	Employees	State	City	Revenue
0	1	Over-Hex	Software	2006.0	25.0	TN	Franklin	$9,684,527
1	2	Unimattax	IT Services	2009.0	36.0	PA	Newtown Square	$14,016,543
2	3	Greenfax	Retail	2012.0	NaN	SC	Greenville	$9,746,272
3	4	Blacklane	IT Services	2011.0	66.0	CA	Orange	$15,359,369
4	5	Yearflex	Software	2013.0	45.0	WI	Madison	$8,567,910
...	...	...	...	...	...	...	...	...
495	496	Buretteadmirable	IT Services	2009.0	93.0	ME	Portland	$15,407,450
496	497	Inventtremendous	Construction	2009.0	24.0	MN	Woodbury	$9,144,857
497	498	Overviewparrot	Retail	2011.0	7125.0	TX	Fort Worth	$11,134,728
498	499	Belaguerra	IT Services	2010.0	140.0	MI	Troy	$17,387,130
499	500	Allpossible	IT Services	2011.0	24.0	CA	Los Angeles	$11,949,706

500 rows × 11 columns

In [3]:

```
df.head()
```

Out[3]:

	ID	Name	Industry	Inception	Employees	State	City	Revenue	Expenses	
0	1	Over-Hex	Software	2006.0	25.0	TN	Franklin	$9,684,527	1,130,700 Dollars	8
1	2	Unimattax	IT Services	2009.0	36.0	PA	Newtown Square	$14,016,543	804,035 Dollars	13
2	3	Greenfax	Retail	2012.0	NaN	SC	Greenville	$9,746,272	1,044,375 Dollars	8
3	4	Blacklane	IT Services	2011.0	66.0	CA	Orange	$15,359,369	4,631,808 Dollars	10
4	5	Yearflex	Software	2013.0	45.0	WI	Madison	$8,567,910	4,374,841 Dollars	4

In [4]:

```
df.isnull()
```

Out[4]:

	ID	Name	Industry	Inception	Employees	State	City	Revenue	Expenses	Profit
0	False	False	False	False	False	False	False	False	False	False
1	False	False	False	False	False	False	False	False	False	False
2	False	False	False	False	True	False	False	False	False	False
3	False	False	False	False	False	False	False	False	False	False
4	False	False	False	False	False	False	False	False	False	False
...	...	...	...	...	...	...	...	...	...	...
495	False	False	False	False	False	False	False	False	False	False
496	False	False	False	False	False	False	False	False	False	False
497	False	False	False	False	False	False	False	False	False	False
498	False	False	False	False	False	False	False	False	False	False
499	False	False	False	False	False	False	False	False	False	False

500 rows × 11 columns

In [5]:

```
df.isnull().count().sum()
```

Out[5]:

```
5500
```

In [6]:

```
df.isnull().sum()
```

Out[6]:

```
ID           0
Name         0
Industry     2
Inception    1
Employees    2
State        4
City         0
Revenue      2
Expenses     3
Profit       2
Growth       1
dtype: int64
```

In [7]:

```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 500 entries, 0 to 499
Data columns (total 11 columns):
 #   Column     Non-Null Count  Dtype  
---  ------     --------------  -----  
 0   ID         500 non-null    int64  
 1   Name       500 non-null    object 
 2   Industry   498 non-null    object 
 3   Inception  499 non-null    float64
 4   Employees  498 non-null    float64
 5   State      496 non-null    object 
 6   City       500 non-null    object 
 7   Revenue    498 non-null    object 
 8   Expenses   497 non-null    object 
 9   Profit     498 non-null    float64
 10  Growth     499 non-null    object 
dtypes: float64(3), int64(1), object(7)
memory usage: 43.1+ KB
```

In [8]:

```
df['Industry'].isnull()
```

Out[8]:

```
0      False
1      False
2      False
3      False
4      False
       ...
495    False
496    False
497    False
498    False
499    False
Name: Industry, Length: 500, dtype: bool
```

In [9]:

```
df.describe()
```

Out[9]:

	ID	Inception	Employees	Profit
count	500.000000	499.000000	498.000000	4.980000e+02
mean	250.500000	2010.174349	148.610442	6.539474e+06
std	144.481833	3.228211	397.353657	3.869934e+06
min	1.000000	1999.000000	1.000000	1.243400e+04
25%	125.750000	2009.000000	27.250000	3.272074e+06
50%	250.500000	2011.000000	56.000000	6.513366e+06
75%	375.250000	2012.000000	126.000000	9.303951e+06
max	500.000000	2014.000000	7125.000000	1.962453e+07

In [10]:

```
df['Revenue']=df['Revenue'].str.lstrip("$")
```

In [11]:

```
df.head()
```

Out[11]:

	ID	Name	Industry	Inception	Employees	State	City	Revenue	Expenses	
0	1	Over-Hex	Software	2006.0	25.0	TN	Franklin	9,684,527	1,130,700 Dollars	85
1	2	Unimattax	IT Services	2009.0	36.0	PA	Newtown Square	14,016,543	804,035 Dollars	132
2	3	Greenfax	Retail	2012.0	NaN	SC	Greenville	9,746,272	1,044,375 Dollars	87
3	4	Blacklane	IT Services	2011.0	66.0	CA	Orange	15,359,369	4,631,808 Dollars	107
4	5	Yearflex	Software	2013.0	45.0	WI	Madison	8,567,910	4,374,841 Dollars	41

In [12]:

```
df['Revenue']=df['Revenue'].replace("," , "",regex=True)
```

In [13]:

```
df['Revenue']=df['Revenue'].replace(" " , "",regex=True)
```

In [14]:

```
df.head()
```

Out[14]:

	ID	Name	Industry	Inception	Employees	State	City	Revenue	Expenses	
0	1	Over-Hex	Software	2006.0	25.0	TN	Franklin	9684527	1,130,700 Dollars	855
1	2	Unimattax	IT Services	2009.0	36.0	PA	Newtown Square	14016543	804,035 Dollars	1321
2	3	Greenfax	Retail	2012.0	NaN	SC	Greenville	9746272	1,044,375 Dollars	870
3	4	Blacklane	IT Services	2011.0	66.0	CA	Orange	15359369	4,631,808 Dollars	1072
4	5	Yearflex	Software	2013.0	45.0	WI	Madison	8567910	4,374,841 Dollars	419

In [15]:

```
df['Expenses']=df['Expenses'].replace(",", "",regex=True)
```

In [16]:

```
df['Revenue']=df['Revenue'].replace(" " , "",regex=True)
```

In [17]:

```
df.head()
```

Out[17]:

	ID	Name	Industry	Inception	Employees	State	City	Revenue	Expenses	
0	1	Over-Hex	Software	2006.0	25.0	TN	Franklin	9684527	1130700 Dollars	855
1	2	Unimattax	IT Services	2009.0	36.0	PA	Newtown Square	14016543	804035 Dollars	1321
2	3	Greenfax	Retail	2012.0	NaN	SC	Greenville	9746272	1044375 Dollars	870
3	4	Blacklane	IT Services	2011.0	66.0	CA	Orange	15359369	4631808 Dollars	1072
4	5	Yearflex	Software	2013.0	45.0	WI	Madison	8567910	4374841 Dollars	419

In [18]:

```
df["Growth"]=df["Growth"].str.rstrip("%")
```

In [19]:

```
df.head()
```

Out[19]:

	ID	Name	Industry	Inception	Employees	State	City	Revenue	Expenses	
0	1	Over-Hex	Software	2006.0	25.0	TN	Franklin	9684527	1130700 Dollars	855
1	2	Unimattax	IT Services	2009.0	36.0	PA	Newtown Square	14016543	804035 Dollars	1321
2	3	Greenfax	Retail	2012.0	NaN	SC	Greenville	9746272	1044375 Dollars	870
3	4	Blacklane	IT Services	2011.0	66.0	CA	Orange	15359369	4631808 Dollars	1072
4	5	Yearflex	Software	2013.0	45.0	WI	Madison	8567910	4374841 Dollars	419

In [20]:

```
df.describe()
```

Out[20]:

	ID	Inception	Employees	Profit
count	500.000000	499.000000	498.000000	4.980000e+02
mean	250.500000	2010.174349	148.610442	6.539474e+06
std	144.481833	3.228211	397.353657	3.869934e+06
min	1.000000	1999.000000	1.000000	1.243400e+04
25%	125.750000	2009.000000	27.250000	3.272074e+06
50%	250.500000	2011.000000	56.000000	6.513366e+06
75%	375.250000	2012.000000	126.000000	9.303951e+06
max	500.000000	2014.000000	7125.000000	1.962453e+07

In [21]:

```
df.isnull()
```

Out[21]:

	ID	Name	Industry	Inception	Employees	State	City	Revenue	Expenses	Profit
0	False	False	False	False	False	False	False	False	False	False
1	False	False	False	False	False	False	False	False	False	False
2	False	False	False	False	True	False	False	False	False	False
3	False	False	False	False	False	False	False	False	False	False
4	False	False	False	False	False	False	False	False	False	False
...	...	...	...	...	...	...	...	...	...	...
495	False	False	False	False	False	False	False	False	False	False
496	False	False	False	False	False	False	False	False	False	False
497	False	False	False	False	False	False	False	False	False	False
498	False	False	False	False	False	False	False	False	False	False
499	False	False	False	False	False	False	False	False	False	False

500 rows × 11 columns

In [22]:

```
df[df.isnull().any(axis=1)]
```

Out[22]:

	ID	Name	Industry	Inception	Employees	State	City	Revenue	Exp
2	3	Greenfax	Retail	2012.0	NaN	SC	Greenville	9746272	1(
7	8	Rednimdox	Construction	2013.0	73.0	NY	Woodside	NaN	
10	11	Canecorporation	Health	2012.0	6.0	NaN	New York	10597009	7
13	14	Techline	NaN	2006.0	65.0	CA	San Ramon	13898119	5
14	15	Cityace	NaN	2010.0	25.0	CO	Louisville	9254614	6
16	17	Ganzlax	IT Services	2011.0	75.0	NJ	Iselin	14001180	
21	22	Lathotline	Health	NaN	103.0	VA	McLean	9418303	7
43	44	Ganzgreen	Construction	2010.0	224.0	TN	Franklin	NaN	
83	84	Drilldrill	Software	2010.0	30.0	NaN	San Francisco	7800620	2
266	267	Circlechop	Software	2010.0	14.0	NaN	San Francisco	9067070	5
331	332	Westminster	Financial Services	2010.0	NaN	MI	Troy	11861652	5
378	379	Stovepuck	Retail	2013.0	73.0	NaN	New York	13814975	5

In [23]:

```
count = df[["State","City"]].isna().sum()
```

In [24]:

```
print(count)
```

```
State    4
City     0
dtype: int64
```

In [25]:

```
df[df["State"].isna()]
```

Out[25]:

	ID	Name	Industry	Inception	Employees	State	City	Revenue	Expens
10	11	Canecorporation	Health	2012.0	6.0	NaN	New York	10597009	7591 Doll
83	84	Drilldrill	Software	2010.0	30.0	NaN	San Francisco	7800620	2785 Doll
266	267	Circlechop	Software	2010.0	14.0	NaN	San Francisco	9067070	5929 Doll
378	379	Stovepuck	Retail	2013.0	73.0	NaN	New York	13814975	5904 Doll

In [26]:

```
df[df["City"]=="New York"]
```

Out[26]:

	ID	Name	Industry	Inception	Employees	State	City	Revenue	Expenses
10	11	Canecorporation	Health	2012.0	6.0	NaN	New York	10597009	7591189 Dollars
65	66	Zathtone	Software	2009.0	121.0	NY	New York	8953759	4069681 Dollars
131	132	Nimlux	Financial Services	2013.0	285.0	NY	New York	7753213	1117206 Dollars
214	215	Nervousdispatcher	Financial Services	2011.0	80.0	NY	New York	6955439	409237 Dollars
324	325	Monotrichous	Software	2011.0	12.0	NY	New York	6203337	3660171 Dollars
330	331	Aspic	Retail	2001.0	130.0	NY	New York	10498414	4645617 Dollars
336	337	Asynchronous	Health	2013.0	22.0	NY	New York	8720504	2632756 Dollars
369	370	Haulbittern	IT Services	2010.0	138.0	NY	New York	15104143	4177864 Dollars
378	379	Stovepuck	Retail	2013.0	73.0	NaN	New York	13814975	5904502 Dollars
418	419	Twistedelephant	IT Services	2011.0	150.0	NY	New York	16917213	2814394 Dollars
473	474	Valuesbilled	IT Services	2013.0	95.0	NY	New York	13214550	7479813 Dollars

In [27]:

```
df["State"][10]='NY'
```

```
C:\Users\dell\AppData\Local\Temp\ipykernel_17024\1982128690.py:1: SettingW
ithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame

See the caveats in the documentation: https://pandas.pydata.org/pandas-doc
s/stable/user_guide/indexing.html#returning-a-view-versus-a-copy (https://
pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-
view-versus-a-copy)
  df["State"][10]='NY'
```

In [28]:

```
df[df["City"]=="New York"]
```

Out[28]:

	ID	Name	Industry	Inception	Employees	State	City	Revenue	Expenses
10	11	Canecorporation	Health	2012.0	6.0	NY	New York	10597009	7591189 Dollars
65	66	Zathtone	Software	2009.0	121.0	NY	New York	8953759	4069681 Dollars
131	132	Nimlux	Financial Services	2013.0	285.0	NY	New York	7753213	1117206 Dollars
214	215	Nervousdispatcher	Financial Services	2011.0	80.0	NY	New York	6955439	409237 Dollars
324	325	Monotrichous	Software	2011.0	12.0	NY	New York	6203337	3660171 Dollars
330	331	Aspic	Retail	2001.0	130.0	NY	New York	10498414	4645617 Dollars
336	337	Asynchronous	Health	2013.0	22.0	NY	New York	8720504	2632756 Dollars
369	370	Haulbittern	IT Services	2010.0	138.0	NY	New York	15104143	4177864 Dollars
378	379	Stovepuck	Retail	2013.0	73.0	NaN	New York	13814975	5904502 Dollars
418	419	Twistedelephant	IT Services	2011.0	150.0	NY	New York	16917213	2814394 Dollars
473	474	Valuesbilled	IT Services	2013.0	95.0	NY	New York	13214550	7479813 Dollars

In [29]:

```
df["State"][378]='NY'
```

```
C:\Users\dell\AppData\Local\Temp\ipykernel_17024\1440361811.py:1: SettingW
ithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame

See the caveats in the documentation: https://pandas.pydata.org/pandas-doc
s/stable/user_guide/indexing.html#returning-a-view-versus-a-copy (https://
pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-
view-versus-a-copy)
  df["State"][378]='NY'
```

In [30]:

```
df[df["City"]=="New York"]
```

Out[30]:

	ID	Name	Industry	Inception	Employees	State	City	Revenue	Expenses
10	11	Canecorporation	Health	2012.0	6.0	NY	New York	10597009	7591189 Dollars
65	66	Zathtone	Software	2009.0	121.0	NY	New York	8953759	4069681 Dollars
131	132	Nimlux	Financial Services	2013.0	285.0	NY	New York	7753213	1117206 Dollars
214	215	Nervousdispatcher	Financial Services	2011.0	80.0	NY	New York	6955439	409237 Dollars
324	325	Monotrichous	Software	2011.0	12.0	NY	New York	6203337	3660171 Dollars
330	331	Aspic	Retail	2001.0	130.0	NY	New York	10498414	4645617 Dollars
336	337	Asynchronous	Health	2013.0	22.0	NY	New York	8720504	2632756 Dollars
369	370	Haulbittern	IT Services	2010.0	138.0	NY	New York	15104143	4177864 Dollars
378	379	Stovepuck	Retail	2013.0	73.0	NY	New York	13814975	5904502 Dollars
418	419	Twistedelephant	IT Services	2011.0	150.0	NY	New York	16917213	2814394 Dollars
473	474	Valuesbilled	IT Services	2013.0	95.0	NY	New York	13214550	7479813 Dollars

In [31]:

```
df[df.isnull().any(axis=1)]
```

Out[31]:

	ID	Name	Industry	Inception	Employees	State	City	Revenue	Expens
2	3	Greenfax	Retail	2012.0	NaN	SC	Greenville	9746272	10443 Doll
7	8	Rednimdox	Construction	2013.0	73.0	NY	Woodside	NaN	N
13	14	Techline	NaN	2006.0	65.0	CA	San Ramon	13898119	54703 Doll
14	15	Cityace	NaN	2010.0	25.0	CO	Louisville	9254614	62494 Doll
16	17	Ganzlax	IT Services	2011.0	75.0	NJ	Iselin	14001180	N
21	22	Lathotline	Health	NaN	103.0	VA	McLean	9418303	75672 Doll
43	44	Ganzgreen	Construction	2010.0	224.0	TN	Franklin	NaN	N
83	84	Drilldrill	Software	2010.0	30.0	NaN	San Francisco	7800620	27857 Doll
266	267	Circlechop	Software	2010.0	14.0	NaN	San Francisco	9067070	59298 Doll
331	332	Westminster	Financial Services	2010.0	NaN	MI	Troy	11861652	52451 Doll

In [32]:

```
count=df["Industry"].isna().sum()
```

In [33]:

```
count
```

Out[33]:

2

In [34]:

```
df[df["Industry"].isna()]
```

Out[34]:

	ID	Name	Industry	Inception	Employees	State	City	Revenue	Expenses	P
13	14	Techline	NaN	2006.0	65.0	CA	San Ramon	13898119	5470303 Dollars	84278
14	15	Cityace	NaN	2010.0	25.0	CO	Louisville	9254614	6249498 Dollars	30051

In [35]:

```
df[df["Inception"]==2010.0]
```

Out[35]:

	ID	Name	Industry	Inception	Employees	State	City	Revenue	E
9	10	Stripfind	Financial Services	2010.0	25.0	FL	Boca Raton	12329371	
14	15	Cityace	NaN	2010.0	25.0	CO	Louisville	9254614	
26	27	Ranktech	Government Services	2010.0	607.0	FL	Tampa	10515557	
29	30	Tambase	Software	2010.0	29.0	AZ	Scottsdale	9141817	
36	37	D-Zoocare	Financial Services	2010.0	98.0	WA	Seattle	9153440	
...	...	...	...	...	...	...	...	...	
468	469	Hoeforeach	IT Services	2010.0	25.0	WI	Wauwatosa	13000248	
476	477	Soddenfurther	Financial Services	2010.0	20.0	CA	San Diego	9243036	
479	480	Excellencerecent	Health	2010.0	381.0	TX	Farmers Branch	10611567	
489	490	Suckedregister	Retail	2010.0	100.0	TX	Houston	11511235	
498	499	Belaguerra	IT Services	2010.0	140.0	MI	Troy	17387130	

83 rows × 11 columns

In [36]:

```
df[df["Name"]=='Cityace']
```

Out[36]:

	ID	Name	Industry	Inception	Employees	State	City	Revenue	Expenses	Pr
14	15	Cityace	NaN	2010.0	25.0	CO	Louisville	9254614	6249498 Dollars	300511

In [37]:

```
df[df["City"]=='Louisville']
```

Out[37]:

	ID	Name	Industry	Inception	Employees	State	City	Revenue	Expense
14	15	Cityace	NaN	2010.0	25.0	CO	Louisville	9254614	624949 Dollar
255	256	Colouredrecipe	IT Services	2010.0	34.0	KY	Louisville	13884717	566309 Dollar

In [38]:

```
df["Industry"][14]='IT Services'
```

```
C:\Users\dell\AppData\Local\Temp\ipykernel_17024\3449398422.py:1: SettingW
ithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame

See the caveats in the documentation: https://pandas.pydata.org/pandas-doc
s/stable/user_guide/indexing.html#returning-a-view-versus-a-copy (https://
pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-
view-versus-a-copy)
  df["Industry"][14]='IT Services'
```

In [39]:

```
df[df["City"]=='San Ramon']
```

Out[39]:

	ID	Name	Industry	Inception	Employees	State	City	Revenue	Expens
8	9	Lamtone	IT Services	2009.0	55.0	CA	San Ramon	11757018	6482 Dol
13	14	Techline	NaN	2006.0	65.0	CA	San Ramon	13898119	5470 Dol
464	465	Dashedcompounds	IT Services	2014.0	59.0	CA	San Ramon	12791846	7879 Dol

In [40]:

```
df["Industry"][13]='IT Services'
```

```
C:\Users\dell\AppData\Local\Temp\ipykernel_17024\2065895781.py:1: SettingW
ithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame

See the caveats in the documentation: https://pandas.pydata.org/pandas-doc
s/stable/user_guide/indexing.html#returning-a-view-versus-a-copy (https://
pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-
view-versus-a-copy)
  df["Industry"][13]='IT Services'
```

In [41]:

```
df[df["Industry"].isna()]
```

Out[41]:

	ID	Name	Industry	Inception	Employees	State	City	Revenue	Expenses	Profit	Growth

In [42]:

```
df[df["Employees"].isna()]
```

Out[42]:

	ID	Name	Industry	Inception	Employees	State	City	Revenue	Expenses
2	3	Greenfax	Retail	2012.0	NaN	SC	Greenville	9746272	1044375 Dollars
331	332	Westminster	Financial Services	2010.0	NaN	MI	Troy	11861652	5245126 Dollars

In [43]:

```
df[df["Industry"]=='Retail'].describe()
```

Out[43]:

	ID	Inception	Employees	Profit
count	48.000000	48.000000	47.000000	4.800000e+01
mean	265.750000	2010.416667	209.276596	7.482728e+06
std	142.938835	3.375968	1033.738965	2.897292e+06
min	3.000000	1999.000000	1.000000	8.153810e+05
25%	134.000000	2009.000000	15.500000	5.658476e+06
50%	286.000000	2011.000000	28.000000	7.326357e+06
75%	386.250000	2013.000000	70.000000	9.490338e+06
max	498.000000	2014.000000	7125.000000	1.336925e+07

In [44]:

```
df["Employees"][2]=28.0
```

```
C:\Users\dell\AppData\Local\Temp\ipykernel_17024\2945548489.py:1: SettingW
ithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame

See the caveats in the documentation: https://pandas.pydata.org/pandas-doc
s/stable/user_guide/indexing.html#returning-a-view-versus-a-copy (https://
pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-
view-versus-a-copy)
  df["Employees"][2]=28.0
```

In [45]:

```
df[df["Industry"]=='Financial Services'].describe()
```

Out[45]:

	ID	Inception	Employees	Profit
count	54.000000	54.000000	53.000000	5.400000e+01
mean	273.351852	2009.888889	211.132075	8.295883e+06
std	156.460877	2.717287	326.567170	2.167587e+06
min	7.000000	2001.000000	3.000000	3.259485e+06
25%	112.250000	2009.000000	33.000000	6.638672e+06
50%	318.000000	2010.000000	80.000000	8.301998e+06
75%	399.000000	2011.000000	250.000000	1.019811e+07
max	495.000000	2014.000000	1628.000000	1.220510e+07

In [46]:

```
df["Employees"][331]=80.0
```

```
C:\Users\dell\AppData\Local\Temp\ipykernel_17024\1842307310.py:1: SettingW
ithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame

See the caveats in the documentation: https://pandas.pydata.org/pandas-doc
s/stable/user_guide/indexing.html#returning-a-view-versus-a-copy (https://
pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-
view-versus-a-copy)
  df["Employees"][331]=80.0
```

In [47]:

```
df["Expenses"]=df["Expenses"].str.rstrip(" Dollars")
```

In [48]:

```
df.head()
```

Out[48]:

	ID	Name	Industry	Inception	Employees	State	City	Revenue	Expenses	
0	1	Over-Hex	Software	2006.0	25.0	TN	Franklin	9684527	1130700	855
1	2	Unimattax	IT Services	2009.0	36.0	PA	Newtown Square	14016543	804035	1321
2	3	Greenfax	Retail	2012.0	28.0	SC	Greenville	9746272	1044375	870
3	4	Blacklane	IT Services	2011.0	66.0	CA	Orange	15359369	4631808	1072
4	5	Yearflex	Software	2013.0	45.0	WI	Madison	8567910	4374841	419

In [49]:

```
df.head()
```

Out[49]:

	ID	Name	Industry	Inception	Employees	State	City	Revenue	Expenses	
0	1	Over-Hex	Software	2006.0	25.0	TN	Franklin	9684527	1130700	855
1	2	Unimattax	IT Services	2009.0	36.0	PA	Newtown Square	14016543	804035	1321
2	3	Greenfax	Retail	2012.0	28.0	SC	Greenville	9746272	1044375	870
3	4	Blacklane	IT Services	2011.0	66.0	CA	Orange	15359369	4631808	1072
4	5	Yearflex	Software	2013.0	45.0	WI	Madison	8567910	4374841	419

In [50]:

```
df[df.isnull().any(axis=1)]
```

Out[50]:

	ID	Name	Industry	Inception	Employees	State	City	Revenue	Expense
7	8	Rednimdox	Construction	2013.0	73.0	NY	Woodside	NaN	Na
16	17	Ganzlax	IT Services	2011.0	75.0	NJ	Iselin	14001180	Na
21	22	Lathotline	Health	NaN	103.0	VA	McLean	9418303	756723
43	44	Ganzgreen	Construction	2010.0	224.0	TN	Franklin	NaN	Na
83	84	Drilldrill	Software	2010.0	30.0	NaN	San Francisco	7800620	278579
266	267	Circlechop	Software	2010.0	14.0	NaN	San Francisco	9067070	592982

In [51]:

```
df.isnull().sum()
```

Out[51]:

```
ID           0
Name         0
Industry     0
Inception    1
Employees    0
State        2
City         0
Revenue      2
Expenses     3
Profit       2
Growth       1
dtype: int64
```

In [52]:

```
df[df["City"]=="San Francisco"]
```

Out[52]:

	ID	Name	Industry	Inception	Employees	State	City	Revenue	Expenses	
83	84	Drilldrill	Software	2010.0	30.0	NaN	San Francisco	7800620	2785799	5
266	267	Circlechop	Software	2010.0	14.0	NaN	San Francisco	9067070	5929828	3

In [53]:

```
df["State"][83]='SF'
```

```
C:\Users\dell\AppData\Local\Temp\ipykernel_17024\3788798315.py:1: SettingW
ithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame

See the caveats in the documentation: https://pandas.pydata.org/pandas-doc
s/stable/user_guide/indexing.html#returning-a-view-versus-a-copy (https://
pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-
view-versus-a-copy)
  df["State"][83]='SF'
```

In [54]:

```
df["State"][266]='SF'
```

```
C:\Users\dell\AppData\Local\Temp\ipykernel_17024\1755374528.py:1: SettingW
ithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame

See the caveats in the documentation: https://pandas.pydata.org/pandas-doc
s/stable/user_guide/indexing.html#returning-a-view-versus-a-copy (https://
pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-
view-versus-a-copy)
  df["State"][266]='SF'
```

In [55]:

```
df[df.isnull().any(axis=1)]
```

Out[55]:

	ID	Name	Industry	Inception	Employees	State	City	Revenue	Expenses
7	8	Rednimdox	Construction	2013.0	73.0	NY	Woodside	NaN	NaN
16	17	Ganzlax	IT Services	2011.0	75.0	NJ	Iselin	14001180	NaN
21	22	Lathotline	Health	NaN	103.0	VA	McLean	9418303	7567233
43	44	Ganzgreen	Construction	2010.0	224.0	TN	Franklin	NaN	NaN

In [56]:

```
df["Revenue"]=df.Revenue.astype(float)
```

In [57]:

```
df["Expenses"]=df.Expenses.astype(float)
```

In [58]:

```
revenue_constrction=df[df["Industry"]=='Construction']
```

In [59]:

```
revenue_constrction
```

Out[59]:

	ID	Name	Industry	Inception	Employees	State	City	Rever
7	8	Rednimdox	Construction	2013.0	73.0	NY	Woodside	N
24	25	Tampware	Construction	2011.0	13.0	TX	Houston	978598
28	29	Groovecane	Construction	2011.0	17.0	UT	South Jordan	681116
38	39	Biotain	Construction	2012.0	62.0	IA	Davenport	1164254
43	44	Ganzgreen	Construction	2010.0	224.0	TN	Franklin	N
44	45	Zamhouse	Construction	2009.0	39.0	ME	Lisbon Falls	754369
45	46	Openjocon	Construction	2013.0	75.0	IL	Midlothian	1137434
48	49	Villadox	Construction	2012.0	20.0	CA	Carlsbad	865164
62	63	Sumzoomit	Construction	2010.0	32.0	MO	Lee's Summit	1842957
66	67	Vivacode	Construction	2009.0	28.0	LA	Lake Charles	941064
73	74	Conin	Construction	2009.0	23.0	CA	Rancho Dominguez	990348
84	85	Tristone	Construction	2013.0	43.0	MD	Millersville	1175598
86	87	Tontam	Construction	2007.0	26.0	VA	Ashburn	763593
99	100	Labam	Construction	2010.0	150.0	FL	Winter Garden	812411
100	101	Black-Tax	Construction	2012.0	22.0	MD	Columbia	516632
106	107	Basetrax	Construction	2011.0	28.0	NY	Buffalo	939101
110	111	Vivaity	Construction	2009.0	23.0	WA	Lakewood	752914
132	133	Gogofan	Construction	2011.0	75.0	FL	Winter Garden	782643
144	145	Sanzatrax	Construction	2012.0	49.0	CA	San Diego	621999
158	159	Lamdombase	Construction	2012.0	7.0	TX	Houston	1142845
165	166	Lamcare	Construction	1999.0	108.0	NY	Staten Island	754771
166	167	Keytouch	Construction	2013.0	35.0	CA	Vista	859726
190	191	Silit	Construction	2007.0	26.0	VA	Chesapeake	1069241
197	198	Singlefind	Construction	2014.0	25.0	OK	Oklahoma City	1172567
207	208	Countslovenly	Construction	2010.0	45.0	FL	Spring Hill	838036
251	252	Estermourn	Construction	2010.0	20.0	ID	Worley	774013
252	253	Splashnuts	Construction	2012.0	272.0	FL	Apopka	1100454
271	272	Presslute	Construction	2011.0	37.0	CA	El Cajon	1061824
278	279	Hypergalaxypegasus	Construction	2009.0	103.0	UT	Centerville	912775
294	295	Optimismexterior	Construction	2009.0	28.0	IL	Bellwood	842702
295	296	Shovelerablaze	Construction	1999.0	150.0	IL	Downers Grove	601486
298	299	Damedash	Construction	2012.0	48.0	CA	San Diego	521583
305	306	Fatcoffle	Construction	2011.0	28.0	OR	Portland	927410

	ID	Name	Industry	Inception	Employees	State	City	Rever
306	307	Decoratorsdecipher	Construction	2011.0	5.0	VA	Ashburn	898235
316	317	Nomarchy	Construction	2010.0	24.0	NC	Charlotte	1410217
326	327	Cyaneous	Construction	2012.0	58.0	IL	Oakbrook Terrace	791472
335	336	Refertilization	Construction	2013.0	13.0	TX	San Antonio	1006429
359	360	Remembergabbro	Construction	2012.0	45.0	UT	Lindon	1087857
403	404	Profilecliche	Construction	2000.0	125.0	NY	Melville	939866
410	411	Miaowgrinder	Construction	2000.0	176.0	NY	Niagara Falls	769746
437	438	Tausnash	Construction	2009.0	95.0	FL	Sarasota	655461
440	441	Grippingstirring	Construction	2012.0	120.0	TX	Mansfield	441927
447	448	Testingear	Construction	2012.0	10.0	TN	Nashville	1065114
452	453	Flaskssunny	Construction	2012.0	175.0	IL	Rolling Meadows	858986
459	460	Rustyentertain	Construction	2012.0	23.0	TX	Carrollton	964681
462	463	Friendlego	Construction	2012.0	38.0	NJ	Westfield	917678
482	483	Abusivebong	Construction	2009.0	8.0	WI	Oshkosh	854332
486	487	Genusequ	Construction	2007.0	45.0	NC	Greensboro	849846
490	491	Comparejson	Construction	2013.0	125.0	IL	Glendale Heights	1235953
496	497	Inventtremendous	Construction	2009.0	24.0	MN	Woodbury	914485

In [60]:

```
revenue_constrction.describe().T
```

Out[60]:

	count	mean	std	min	25%	50%	75%
ID	50.0	236.440	1.576288e+02	8.0	90.25	230.0	354.00
Inception	50.0	2009.940	3.530682e+00	1999.0	2009.00	2011.0	2012.00
Employees	50.0	61.260	5.943091e+01	5.0	23.25	37.5	75.00
Revenue	48.0	9158737.125	2.404913e+06	4419277.0	7729470.75	9055058.5	10626473.75
Expenses	48.0	4453204.500	1.793322e+06	214470.0	3539327.50	4506975.5	5376596.25
Profit	48.0	4705532.625	2.805089e+06	96073.0	2442148.25	4573280.5	6801062.00

In [61]:

```
df[df["Revenue"].isna()]
```

Out[61]:

	ID	Name	Industry	Inception	Employees	State	City	Revenue	Expenses
7	8	Rednimdox	Construction	2013.0	73.0	NY	Woodside	NaN	NaN
43	44	Ganzgreen	Construction	2010.0	224.0	TN	Franklin	NaN	NaN

In [62]:

```
df["Revenue"][7]=9055058.5
```

```
C:\Users\dell\AppData\Local\Temp\ipykernel_17024\4011270014.py:1: SettingW
ithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame

See the caveats in the documentation: https://pandas.pydata.org/pandas-doc
s/stable/user_guide/indexing.html#returning-a-view-versus-a-copy (https://
pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-
view-versus-a-copy)
  df["Revenue"][7]=9055058.5
```

In [63]:

```
df["Revenue"][43]=9055058.5
```

```
C:\Users\dell\AppData\Local\Temp\ipykernel_17024\781329487.py:1: SettingWi
thCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame

See the caveats in the documentation: https://pandas.pydata.org/pandas-doc
s/stable/user_guide/indexing.html#returning-a-view-versus-a-copy (https://
pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-
view-versus-a-copy)
  df["Revenue"][43]=9055058.5
```

In [64]:

```
df[df.isnull().any(axis=1)]
```

Out[64]:

	ID	Name	Industry	Inception	Employees	State	City	Revenue	Expense
7	8	Rednimdox	Construction	2013.0	73.0	NY	Woodside	9055058.5	Nal
16	17	Ganzlax	IT Services	2011.0	75.0	NJ	Iselin	14001180.0	Nal
21	22	Lathotline	Health	NaN	103.0	VA	McLean	9418303.0	7567233.
43	44	Ganzgreen	Construction	2010.0	224.0	TN	Franklin	9055058.5	Nal

In [65]:

```
df["Expenses"][7]='4506975.5'
```

```
C:\Users\dell\AppData\Local\Temp\ipykernel_17024\2246218118.py:1: SettingW
ithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame

See the caveats in the documentation: https://pandas.pydata.org/pandas-doc
s/stable/user_guide/indexing.html#returning-a-view-versus-a-copy (https://
pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-
view-versus-a-copy)
  df["Expenses"][7]='4506975.5'
```

In [66]:

```
df["Expenses"][43]='4506975.5'
```

In [67]:

```
df[df.isna().any(axis=1)]
```

Out[67]:

	ID	Name	Industry	Inception	Employees	State	City	Revenue	Expense
7	8	Rednimdox	Construction	2013.0	73.0	NY	Woodside	9055058.5	4506975.
16	17	Ganzlax	IT Services	2011.0	75.0	NJ	Iselin	14001180.0	Nal
21	22	Lathotline	Health	NaN	103.0	VA	McLean	9418303.0	7567233.
43	44	Ganzgreen	Construction	2010.0	224.0	TN	Franklin	9055058.5	4506975.

In [68]:

```
df["Expenses"][16]=(df["Revenue"][16])-(df["Profit"][16])
```

```
C:\Users\dell\AppData\Local\Temp\ipykernel_17024\219174223.py:1: SettingWi
thCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame

See the caveats in the documentation: https://pandas.pydata.org/pandas-doc
s/stable/user_guide/indexing.html#returning-a-view-versus-a-copy (https://
pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-
view-versus-a-copy)
  df["Expenses"][16]=(df["Revenue"][16])-(df["Profit"][16])
```

In [69]:

```
df[df["Industry"]=='Health'].describe()
```

Out[69]:

	ID	Inception	Employees	Revenue	Profit
count	86.000000	85.000000	86.000000	8.600000e+01	8.600000e+01
mean	260.813953	2010.894118	205.872093	8.811122e+06	2.929281e+06
std	148.923234	3.008019	305.918590	1.978820e+06	2.075214e+06
min	11.000000	2000.000000	6.000000	1.614585e+06	1.243400e+04
25%	123.000000	2010.000000	31.000000	7.588070e+06	1.311362e+06
50%	282.000000	2012.000000	86.500000	8.855710e+06	2.514786e+06
75%	389.000000	2013.000000	228.750000	1.001363e+07	4.480563e+06
max	485.000000	2014.000000	1600.000000	1.531230e+07	9.174395e+06

In [70]:

```
df["Inception"][21]=2010.0
```

```
C:\Users\dell\AppData\Local\Temp\ipykernel_17024\3376397242.py:1: SettingW
ithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame

See the caveats in the documentation: https://pandas.pydata.org/pandas-doc
s/stable/user_guide/indexing.html#returning-a-view-versus-a-copy (https://
pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-
view-versus-a-copy)
  df["Inception"][21]=2010.0
```

In [71]:

```
df[df.isna().any(axis=1)]
```

Out[71]:

	ID	Name	Industry	Inception	Employees	State	City	Revenue	Expenses
7	8	Rednimdox	Construction	2013.0	73.0	NY	Woodside	9055058.5	4506975.5
43	44	Ganzgreen	Construction	2010.0	224.0	TN	Franklin	9055058.5	4506975.5

In [72]:

```
df[df["Industry"]=='Construction']
```

Out[72]:

	ID	Name	Industry	Inception	Employees	State	City	Rever
7	8	Rednimdox	Construction	2013.0	73.0	NY	Woodside	905505
24	25	Tampware	Construction	2011.0	13.0	TX	Houston	978598
28	29	Groovecane	Construction	2011.0	17.0	UT	South Jordan	681116
38	39	Biotain	Construction	2012.0	62.0	IA	Davenport	1164254
43	44	Ganzgreen	Construction	2010.0	224.0	TN	Franklin	905505
44	45	Zamhouse	Construction	2009.0	39.0	ME	Lisbon Falls	754369
45	46	Openjocon	Construction	2013.0	75.0	IL	Midlothian	1137434
48	49	Villadox	Construction	2012.0	20.0	CA	Carlsbad	865164
62	63	Sumzoomit	Construction	2010.0	32.0	MO	Lee's Summit	1842957
66	67	Vivacode	Construction	2009.0	28.0	LA	Lake Charles	941064
73	74	Conin	Construction	2009.0	23.0	CA	Rancho Dominguez	990348
84	85	Tristone	Construction	2013.0	43.0	MD	Millersville	1175598
86	87	Tontam	Construction	2007.0	26.0	VA	Ashburn	763593
99	100	Labam	Construction	2010.0	150.0	FL	Winter Garden	812411
100	101	Black-Tax	Construction	2012.0	22.0	MD	Columbia	516632
106	107	Basetrax	Construction	2011.0	28.0	NY	Buffalo	939101
110	111	Vivaity	Construction	2009.0	23.0	WA	Lakewood	752914
132	133	Gogofan	Construction	2011.0	75.0	FL	Winter Garden	782643
144	145	Sanzatrax	Construction	2012.0	49.0	CA	San Diego	621999
158	159	Lamdombase	Construction	2012.0	7.0	TX	Houston	1142845
165	166	Lamcare	Construction	1999.0	108.0	NY	Staten Island	754771
166	167	Keytouch	Construction	2013.0	35.0	CA	Vista	859726
190	191	Silit	Construction	2007.0	26.0	VA	Chesapeake	1069241
197	198	Singlefind	Construction	2014.0	25.0	OK	Oklahoma City	1172567
207	208	Countslovenly	Construction	2010.0	45.0	FL	Spring Hill	838036
251	252	Estermourn	Construction	2010.0	20.0	ID	Worley	774013
252	253	Splashnuts	Construction	2012.0	272.0	FL	Apopka	1100454
271	272	Presslute	Construction	2011.0	37.0	CA	El Cajon	1061824
278	279	Hypergalaxypegasus	Construction	2009.0	103.0	UT	Centerville	912775
294	295	Optimismexterior	Construction	2009.0	28.0	IL	Bellwood	842702
295	296	Shovelerablaze	Construction	1999.0	150.0	IL	Downers Grove	601486
298	299	Damedash	Construction	2012.0	48.0	CA	San Diego	521583
305	306	Fatcoffle	Construction	2011.0	28.0	OR	Portland	927410

	ID	Name	Industry	Inception	Employees	State	City	Rever
306	307	Decoratorsdecipher	Construction	2011.0	5.0	VA	Ashburn	898235
316	317	Nomarchy	Construction	2010.0	24.0	NC	Charlotte	1410217
326	327	Cyaneous	Construction	2012.0	58.0	IL	Oakbrook Terrace	791472
335	336	Refertilization	Construction	2013.0	13.0	TX	San Antonio	1006429
359	360	Remembergabbro	Construction	2012.0	45.0	UT	Lindon	1087857
403	404	Profilecliche	Construction	2000.0	125.0	NY	Melville	939866
410	411	Miaowgrinder	Construction	2000.0	176.0	NY	Niagara Falls	769746
437	438	Tausnash	Construction	2009.0	95.0	FL	Sarasota	655461
440	441	Grippingstirring	Construction	2012.0	120.0	TX	Mansfield	441927
447	448	Testingear	Construction	2012.0	10.0	TN	Nashville	1065114
452	453	Flaskssunny	Construction	2012.0	175.0	IL	Rolling Meadows	858986
459	460	Rustyentertain	Construction	2012.0	23.0	TX	Carrollton	964681
462	463	Friendlego	Construction	2012.0	38.0	NJ	Westfield	917678
482	483	Abusivebong	Construction	2009.0	8.0	WI	Oshkosh	854332
486	487	Genusequ	Construction	2007.0	45.0	NC	Greensboro	849846
490	491	Comparejson	Construction	2013.0	125.0	IL	Glendale Heights	1235953
496	497	Inventtremendous	Construction	2009.0	24.0	MN	Woodbury	914485

In [73]:

```
df["Growth"]=df.Growth.astype(float)
```

In [74]:

```
df[df["Industry"]=='Construction'].describe()
```

Out[74]:

	ID	Inception	Employees	Revenue	Profit	Growth
count	50.00000	50.000000	50.00000	5.000000e+01	4.800000e+01	49.000000
mean	236.44000	2009.940000	61.26000	9.154590e+06	4.705533e+06	10.061224
std	157.62875	3.530682	59.43091	2.355412e+06	2.805089e+06	3.071429
min	8.00000	1999.000000	5.00000	4.419277e+06	9.607300e+04	5.000000
25%	90.25000	2009.000000	23.25000	7.761712e+06	2.442148e+06	8.000000
50%	230.00000	2011.000000	37.50000	9.055058e+06	4.573280e+06	10.000000
75%	354.00000	2012.000000	75.00000	1.047976e+07	6.801062e+06	12.000000
max	497.00000	2014.000000	272.00000	1.842958e+07	1.261618e+07	19.000000

In [75]:

```
df["Growth"][7]=10.0
```

```
C:\Users\dell\AppData\Local\Temp\ipykernel_17024\3110848792.py:1: SettingW
ithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame

See the caveats in the documentation: https://pandas.pydata.org/pandas-doc
s/stable/user_guide/indexing.html#returning-a-view-versus-a-copy (https://
pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-
view-versus-a-copy)
  df["Growth"][7]=10.0
```

In [76]:

```
df[df.isnull().any(axis=1)]
```

Out[76]:

	ID	Name	Industry	Inception	Employees	State	City	Revenue	Expenses
7	8	Rednimdox	Construction	2013.0	73.0	NY	Woodside	9055058.5	4506975.5
43	44	Ganzgreen	Construction	2010.0	224.0	TN	Franklin	9055058.5	4506975.5

In [77]:

```
df['Expenses']=df.Expenses.astype(float)
```

In [78]:

```
type(df['Revenue'])
```

Out[78]:

```
pandas.core.series.Series
```

In [79]:

```
df['Profit'][7]=(df['Revenue'][7])-(df['Expenses'][7])
```

```
C:\Users\dell\AppData\Local\Temp\ipykernel_17024\128114387.py:1: SettingWi
thCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame

See the caveats in the documentation: https://pandas.pydata.org/pandas-doc
s/stable/user_guide/indexing.html#returning-a-view-versus-a-copy (https://
pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-
view-versus-a-copy)
  df['Profit'][7]=(df['Revenue'][7])-(df['Expenses'][7])
```

In [80]:

```
df['Profit'][43]=(df['Revenue'][43])-(df['Expenses'][43])
```

```
C:\Users\dell\AppData\Local\Temp\ipykernel_17024\1405208390.py:1: SettingW
ithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame

See the caveats in the documentation: https://pandas.pydata.org/pandas-doc
s/stable/user_guide/indexing.html#returning-a-view-versus-a-copy (https://
pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-
view-versus-a-copy)
  df['Profit'][43]=(df['Revenue'][43])-(df['Expenses'][43])
```

In [81]:

```
df[df.isna().any(axis=1)]
```

Out[81]:

	ID	Name	Industry	Inception	Employees	State	City	Revenue	Expenses	Profit	Growth

In [82]:

```
df.isnull().count()
```

Out[82]:

```
ID           500
Name         500
Industry     500
Inception    500
Employees    500
State        500
City         500
Revenue      500
Expenses     500
Profit       500
Growth       500
dtype: int64
```

In [83]:

```
df.describe()
```

Out[83]:

	ID	Inception	Employees	Revenue	Expenses	Profit	
count	500.000000	500.000000	500.000000	5.000000e+02	5.000000e+02	5.000000e+02	500.
mean	250.500000	2010.174000	148.232000	1.083801e+07	4.306501e+06	6.531508e+06	14.
std	144.481833	3.224985	396.605024	3.189223e+06	2.111938e+06	3.864219e+06	6.
min	1.000000	1999.000000	1.000000	1.614585e+06	7.121900e+04	1.243400e+04	-3.
25%	125.750000	2009.000000	27.750000	8.696767e+06	2.758423e+06	3.297251e+06	8.
50%	250.500000	2011.000000	56.000000	1.064063e+07	4.366960e+06	6.503795e+06	15.
75%	375.250000	2012.000000	126.000000	1.310111e+07	5.805174e+06	9.283554e+06	20.
max	500.000000	2014.000000	7125.000000	2.181005e+07	9.860686e+06	1.962453e+07	30.

In [84]:

```
df.head()
```

Out[84]:

	ID	Name	Industry	Inception	Employees	State	City	Revenue	Expenses	
0	1	Over-Hex	Software	2006.0	25.0	TN	Franklin	9684527.0	1130700.0	8
1	2	Unimattax	IT Services	2009.0	36.0	PA	Newtown Square	14016543.0	804035.0	13
2	3	Greenfax	Retail	2012.0	28.0	SC	Greenville	9746272.0	1044375.0	8
3	4	Blacklane	IT Services	2011.0	66.0	CA	Orange	15359369.0	4631808.0	10
4	5	Yearflex	Software	2013.0	45.0	WI	Madison	8567910.0	4374841.0	4

In [85]:

```
Industry_Software=df[df["Industry"]=='Software']
```

In [86]:

```
Industry_Software['Revenue']
```

Out[86]:

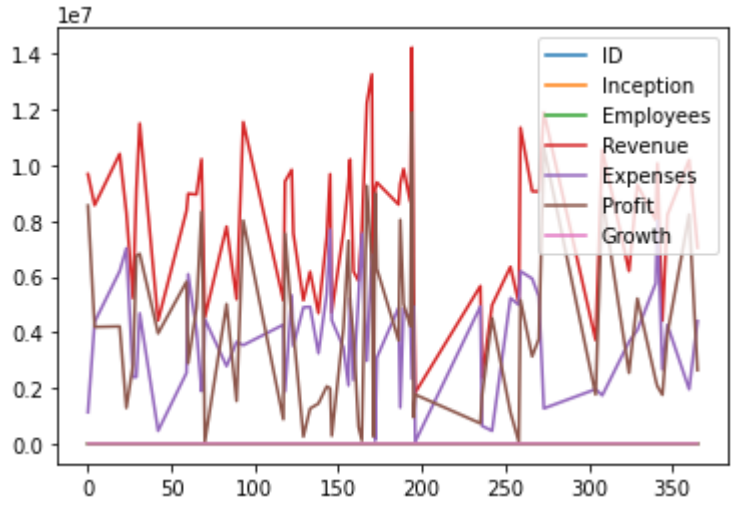
```
0       9684527.0
4       8567910.0
19     10410628.0
23      8304480.0
27      5231275.0
           ...
341    10067223.0
344     4422083.0
347     8215508.0
360    10188123.0
365     7047541.0
Name: Revenue, Length: 64, dtype: float64
```

In [87]:

```
Industry_Software.plot()
```

Out[87]:

```
<AxesSubplot:>
```



In [88]:

```
df['Industry'].unique()
```

Out[88]:

```
array(['Software', 'IT Services', 'Retail', 'Financial Services',
       'Construction', 'Health', 'Government Services'], dtype=object)
```

In [89]:

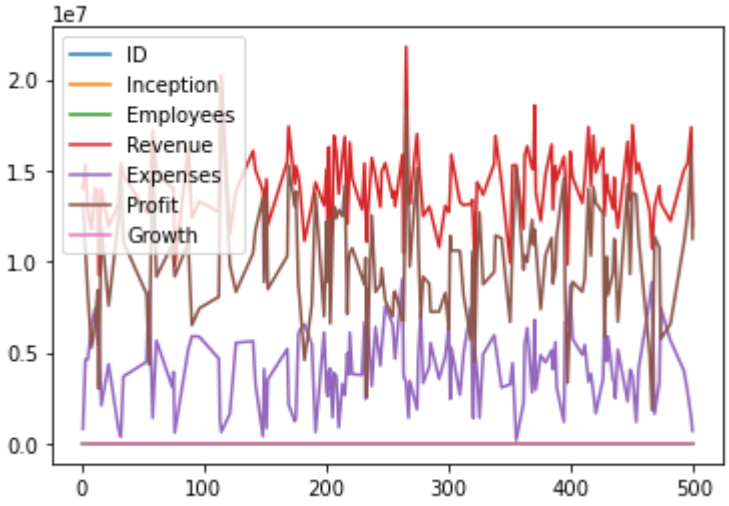
```
Industry_IT_Services=df[df["Industry"]=='IT Services']
```

In [90]:

```
Industry_IT_Services.plot()
```

Out[90]:

<AxesSubplot:>



In [91]:

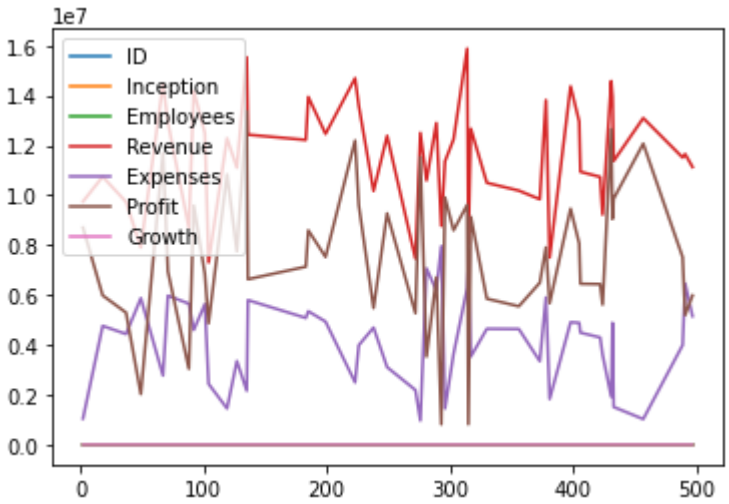
```
Industry_Retail=df[df["Industry"]=='Retail']
```

In [92]:

```
Industry_Retail.plot()
```

Out[92]:

<AxesSubplot:>



In [93]:

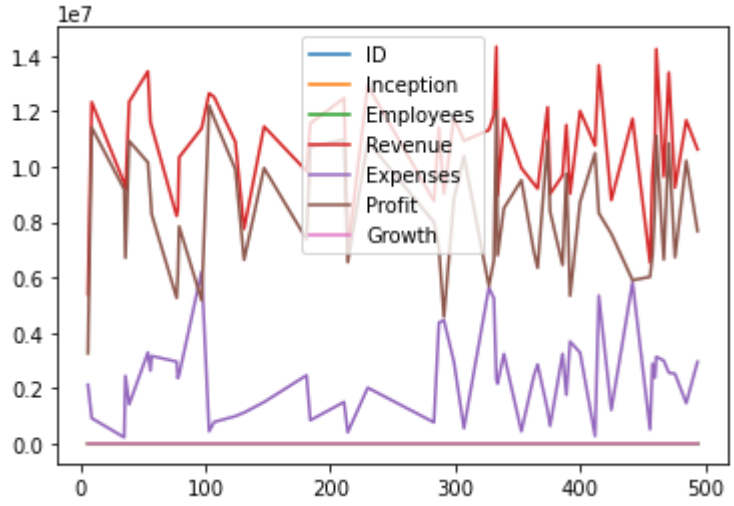
```
Industry_Financial_Services=df[df["Industry"]=='Financial Services']
```

In [94]:

```
Industry_Financial_Services.plot()
```

Out[94]:

<AxesSubplot:>



In [95]:

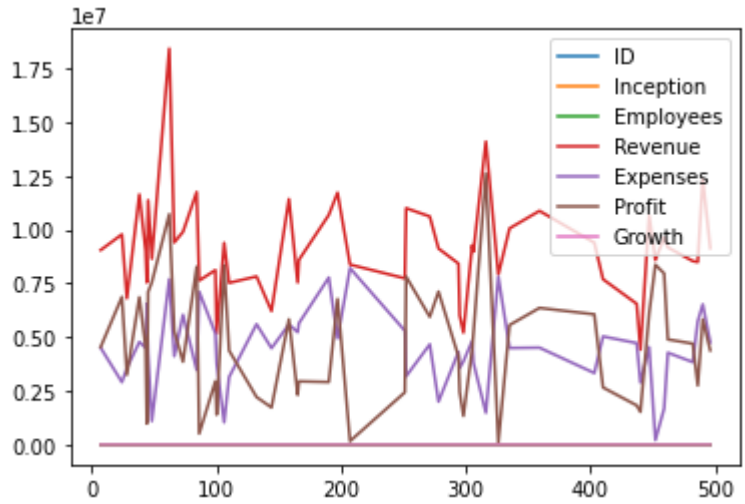
```
Industry_Construction=df[df["Industry"]=='Construction']
```

In [96]:

```
Industry_Construction.plot()
```

Out[96]:

<AxesSubplot:>



In [97]:

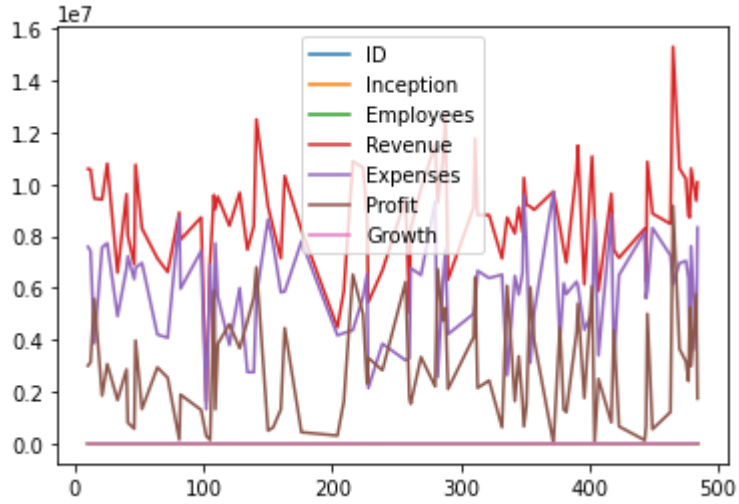
```
Industry_Health=df[df["Industry"]=='Health']
```

In [98]:

```
Industry_Health.plot()
```

Out[98]:

```
<AxesSubplot:>
```



In [99]:

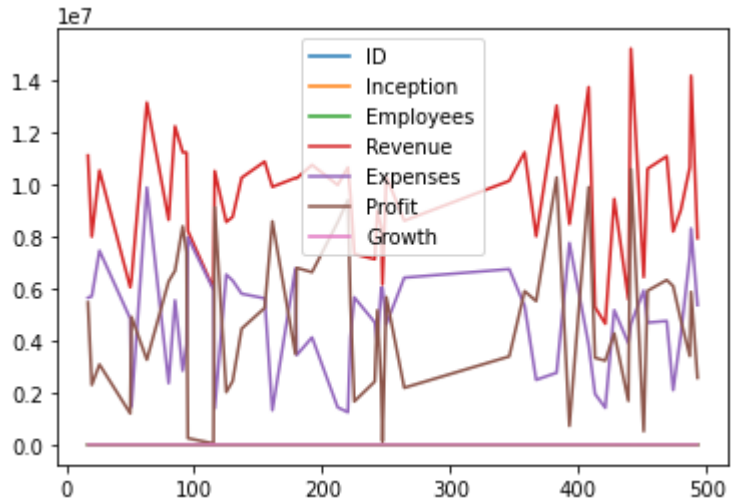
```
Industry_Government_Services=df[df["Industry"]=='Government Services']
```

In [100]:

```
Industry_Government_Services.plot()
```

Out[100]:

```
<AxesSubplot:>
```



In [101]:

```
insights=df[["Revenue","Expenses"]]
```

In [102]:

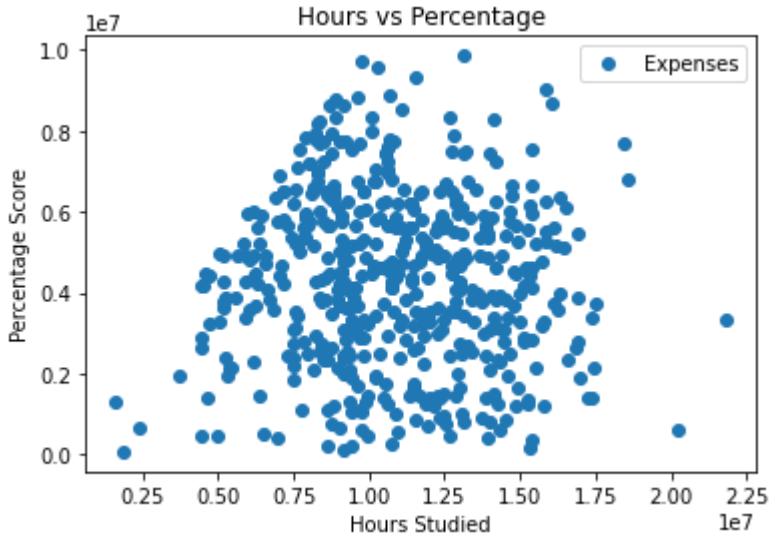
```
insights.describe()
```

Out[102]:

	Revenue	Expenses
count	5.000000e+02	5.000000e+02
mean	1.083801e+07	4.306501e+06
std	3.189223e+06	2.111938e+06
min	1.614585e+06	7.121900e+04
25%	8.696767e+06	2.758423e+06
50%	1.064063e+07	4.366960e+06
75%	1.310111e+07	5.805174e+06
max	2.181005e+07	9.860686e+06

In [103]:

```
df.plot(x = 'Revenue', y = 'Expenses', style = 'o')
plt.title('Hours vs Percentage')
plt.xlabel('Hours Studied')
plt.ylabel('Percentage Score')
plt.show()
```



In [ ]:

```

```