In [36]:

```
#Load the required libraries
import pandas as pd
import numpy as np
import seaborn as sns
import matplotlib.pyplot as plt
#Load the data
unicorn = pd.read_csv('C:/Users/dabir/Downloads/EDA FILE/data_set/unicorn/Unicorn_Compani
```

In [2]:

```
unicorn.head()
```

Out[2]:

	Company	Valuation	Date Joined	Industry	City	Country	Continent	Year Founded	Fundin
0	Bytedance	$180B	2017-04-07	Artificial intelligence	Beijing	China	Asia	2012	$8
1	SpaceX	$100B	2012-12-01	Other	Hawthorne	United States	North America	2002	$7
2	SHEIN	$100B	2018-07-03	E-commerce & direct-to-consumer	Shenzhen	China	Asia	2008	$2
3	Stripe	$95B	2014-01-23	Fintech	San Francisco	United States	North America	2010	$2
4	Klarna	$46B	2011-12-12	Fintech	Stockholm	Sweden	Europe	2005	$4

In [3]:

```
unicorn.shape
```

Out[3]:

```
(1074, 10)
```

In [4]:

```
#Basic information

unicorn.info()

#Describe the data

unicorn.describe()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1074 entries, 0 to 1073
Data columns (total 10 columns):
 #   Column            Non-Null Count  Dtype 
---  ------            --------------  ----- 
 0   Company           1074 non-null   object
 1   Valuation         1074 non-null   object
 2   Date Joined       1074 non-null   object
 3   Industry          1074 non-null   object
 4   City              1058 non-null   object
 5   Country           1074 non-null   object
 6   Continent         1074 non-null   object
 7   Year Founded      1074 non-null   int64 
 8   Funding           1074 non-null   object
 9   Select Investors  1073 non-null   object
dtypes: int64(1), object(9)
memory usage: 84.0+ KB
```

Out[4]:

	Year Founded
count	1074.000000
mean	2012.895717
std	5.698573
min	1919.000000
25%	2011.000000
50%	2014.000000
75%	2016.000000
max	2021.000000

In [25]:

```
#Datatypes

unicorn.dtypes
```

Out[25]:

```
Company             object
Valuation           object
Date Joined         object
Industry            object
City                object
Country             object
Continent           object
Year Founded         int64
Funding             object
Select Investors    object
dtype: object
```

In [5]:

```
#the duplicates

unicorn.duplicated().sum()
```

Out[5]:

```
0
```

In [19]:

```
#unique values

unicorn['Company'].unique()

unicorn['Continent'].unique()

unicorn['Country'].unique()
```

Out[19]:

```
array(['China', 'United States', 'Sweden', 'Australia', 'United Kingdom',
       'Bahamas', 'India', 'Indonesia', 'Turkey', 'Estonia', 'Germany',
       'Hong Kong', 'South Korea', 'Mexico', 'Canada', 'Netherlands',
       'France', 'Finland', 'Israel', 'Lithuania', 'Denmark', 'Belgium',
       'Colombia', 'Brazil', 'Singapore', 'Austria', 'Ireland',
       'United Arab Emirates', 'Switzerland', 'Vietnam', 'South Africa',
       'Thailand', 'Norway', 'Chile', 'Argentina', 'Bermuda', 'Japan',
       'Spain', 'Malaysia', 'Senegal', 'Philippines', 'Luxembourg',
       'Nigeria', 'Czech Republic', 'Croatia', 'Italy'], dtype=object)
```

In [23]:

```
# null values

unicorn.isnull().sum()
```

Out[23]:

```
Company              0
Valuation            0
Date Joined          0
Industry             0
City                16
Country              0
Continent            0
Year Founded         0
Funding              0
Select Investors     1
dtype: int64
```

In [24]:

```
#Replace null values

unicorn.replace(np.nan,'0',inplace = True)

#Check the changes now
unicorn.isnull().sum()
```

Out[24]:

```
Company             0
Valuation           0
Date Joined         0
Industry            0
City                0
Country             0
Continent           0
Year Founded        0
Funding             0
Select Investors    0
dtype: int64
```

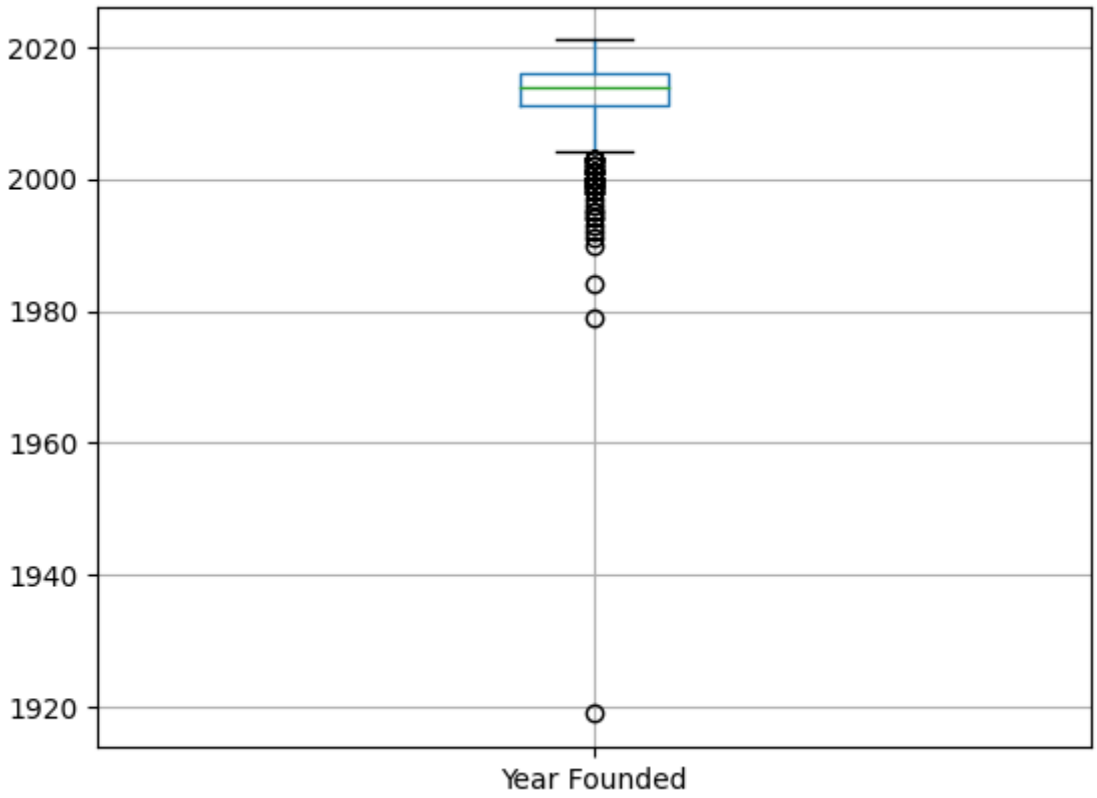
In [28]:

```
#Boxplot

unicorn[['Year Founded']].boxplot()
```

Out[28]:

<AxesSubplot: >



In [33]:

```
from datetime import date
unicorn.today().year
unicorn['Car_Age']=date.today().year-data['Year Founded']
unicorn.head()
```

```
---------------------------------------------------------------------------
-
AttributeError                            Traceback (most recent call las
t)
Cell In[33], line 2
      1 ##from datetime import date
----> 2 unicorn.today().year
      3 unicorn['Car_Age']=date.today().year-data['Year Founded']
      4 unicorn.head()

File ~\anaconda3\lib\site-packages\pandas\core\generic.py:5902, in NDFram
e.__getattr__(self, name)
   5895 if (
   5896     name not in self._internal_names_set
   5897     and name not in self._metadata
   5898     and name not in self._accessors
   5899     and self._info_axis._can_hold_identifiers_and_holds_name(name)
   5900 ):
   5901     return self[name]
-> 5902 return object.__getattribute__(self, name)

AttributeError: 'DataFrame' object has no attribute 'today'
```

In [39]:

```
unicorn.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1074 entries, 0 to 1073
Data columns (total 10 columns):
 #   Column            Non-Null Count  Dtype
---  ------            --------------  -----
 0   Company           1074 non-null   object
 1   Valuation         1074 non-null   object
 2   Date Joined       1074 non-null   object
 3   Industry          1074 non-null   object
 4   City              1058 non-null   object
 5   Country           1074 non-null   object
 6   Continent         1074 non-null   object
 7   Year Founded      1074 non-null   int64
 8   Funding           1074 non-null   object
 9   Select Investors  1073 non-null   object
dtypes: int64(1), object(9)
memory usage: 84.0+ KB
```

In [52]:

```
pip install pandas_profiling
```

```
Requirement already satisfied: pandas_profiling in c:\users\dabir\anaconda
3\lib\site-packages (3.6.6)Note: you may need to restart the kernel to use
updated packages.

Requirement already satisfied: ydata-profiling in c:\users\dabir\anaconda3
\lib\site-packages (from pandas_profiling) (4.1.2)
Requirement already satisfied: statsmodels<0.14,>=0.13.2 in c:\users\dabir
\anaconda3\lib\site-packages (from ydata-profiling->pandas_profiling) (0.1
3.5)
Requirement already satisfied: PyYAML<6.1,>=5.0.0 in c:\users\dabir\anacon
da3\lib\site-packages (from ydata-profiling->pandas_profiling) (6.0)
Requirement already satisfied: pandas!=1.4.0,<1.6,>1.1 in c:\users\dabir\a
naconda3\lib\site-packages (from ydata-profiling->pandas_profiling) (1.5.
3)
Requirement already satisfied: htmlmin==0.1.12 in c:\users\dabir\anaconda3
\lib\site-packages (from ydata-profiling->pandas_profiling) (0.1.12)
Requirement already satisfied: tqdm<4.65,>=4.48.2 in c:\users\dabir\anacon
da3\lib\site-packages (from ydata-profiling->pandas_profiling) (4.64.1)
Requirement already satisfied: typeguard<2.14,>=2.13.2 in c:\users\dabir\a
naconda3\lib\site-packages (from ydata-profiling->pandas_profiling) (2.13.
3)
Requirement already satisfied: requests<2.29,>=2.24.0 in c:\users\dabir\an
aconda3\lib\site-packages (from ydata-profiling->pandas_profiling) (2.28.
1)
Requirement already satisfied: multimethod<1.10,>=1.4 in c:\users\dabir\an
aconda3\lib\site-packages (from ydata-profiling->pandas_profiling) (1.9.1)
Requirement already satisfied: imagehash==4.3.1 in c:\users\dabir\anaconda
3\lib\site-packages (from ydata-profiling->pandas_profiling) (4.3.1)
Requirement already satisfied: numpy<1.24,>=1.16.0 in c:\users\dabir\anaco
nda3\lib\site-packages (from ydata-profiling->pandas_profiling) (1.23.5)
Requirement already satisfied: scipy<1.10,>=1.4.1 in c:\users\dabir\anacon
da3\lib\site-packages (from ydata-profiling->pandas_profiling) (1.9.3)
Requirement already satisfied: phik<0.13,>=0.11.1 in c:\users\dabir\anacon
da3\lib\site-packages (from ydata-profiling->pandas_profiling) (0.12.3)
Requirement already satisfied: pydantic<1.11,>=1.8.1 in c:\users\dabir\ana
conda3\lib\site-packages (from ydata-profiling->pandas_profiling) (1.10.7)
Requirement already satisfied: matplotlib<3.7,>=3.2 in c:\users\dabir\anac
onda3\lib\site-packages (from ydata-profiling->pandas_profiling) (3.6.3)
Requirement already satisfied: jinja2<3.2,>=2.11.1 in c:\users\dabir\anaco
nda3\lib\site-packages (from ydata-profiling->pandas_profiling) (3.1.2)
Requirement already satisfied: seaborn<0.13,>=0.10.1 in c:\users\dabir\ana
conda3\lib\site-packages (from ydata-profiling->pandas_profiling) (0.12.2)
Requirement already satisfied: visions[type_image_path]==0.7.5 in c:\users
\dabir\anaconda3\lib\site-packages (from ydata-profiling->pandas_profilin
g) (0.7.5)
Requirement already satisfied: PyWavelets in c:\users\dabir\anaconda3\lib
\site-packages (from imagehash==4.3.1->ydata-profiling->pandas_profiling)
(1.4.1)
Requirement already satisfied: pillow in c:\users\dabir\anaconda3\lib\site
-packages (from imagehash==4.3.1->ydata-profiling->pandas_profiling) (9.4.
0)
Requirement already satisfied: networkx>=2.4 in c:\users\dabir\anaconda3\l
ib\site-packages (from visions[type_image_path]==0.7.5->ydata-profiling->p
andas_profiling) (2.8.4)
Requirement already satisfied: tangled-up-in-unicode>=0.0.4 in c:\users\da
bir\anaconda3\lib\site-packages (from visions[type_image_path]==0.7.5->yda
ta-profiling->pandas_profiling) (0.2.0)
Requirement already satisfied: attrs>=19.3.0 in c:\users\dabir\anaconda3\l
ib\site-packages (from visions[type_image_path]==0.7.5->ydata-profiling->p
andas_profiling) (22.1.0)
Requirement already satisfied: MarkupSafe>=2.0 in c:\users\dabir\anaconda3
```

```
\lib\site-packages (from jinja2<3.2,>=2.11.1->ydata-profiling->pandas_prof
iling) (2.1.1)
Requirement already satisfied: fonttools>=4.22.0 in c:\users\dabir\anacond
a3\lib\site-packages (from matplotlib<3.7,>=3.2->ydata-profiling->pandas_p
rofiling) (4.25.0)
Requirement already satisfied: packaging>=20.0 in c:\users\dabir\anaconda3
\lib\site-packages (from matplotlib<3.7,>=3.2->ydata-profiling->pandas_pro
filing) (22.0)
Requirement already satisfied: cycler>=0.10 in c:\users\dabir\anaconda3\li
b\site-packages (from matplotlib<3.7,>=3.2->ydata-profiling->pandas_profil
ing) (0.11.0)
Requirement already satisfied: contourpy>=1.0.1 in c:\users\dabir\anaconda
3\lib\site-packages (from matplotlib<3.7,>=3.2->ydata-profiling->pandas_pr
ofiling) (1.0.5)
Requirement already satisfied: python-dateutil>=2.7 in c:\users\dabir\anac
onda3\lib\site-packages (from matplotlib<3.7,>=3.2->ydata-profiling->panda
s_profiling) (2.8.2)
Requirement already satisfied: pyparsing>=2.2.1 in c:\users\dabir\anaconda
3\lib\site-packages (from matplotlib<3.7,>=3.2->ydata-profiling->pandas_pr
ofiling) (3.0.9)
Requirement already satisfied: kiwisolver>=1.0.1 in c:\users\dabir\anacond
a3\lib\site-packages (from matplotlib<3.7,>=3.2->ydata-profiling->pandas_p
rofiling) (1.4.4)
Requirement already satisfied: pytz>=2020.1 in c:\users\dabir\anaconda3\li
b\site-packages (from pandas!=1.4.0,<1.6,>1.1->ydata-profiling->pandas_pro
filing) (2022.7)
Requirement already satisfied: joblib>=0.14.1 in c:\users\dabir\anaconda3
\lib\site-packages (from phik<0.13,>=0.11.1->ydata-profiling->pandas_profi
ling) (1.1.1)
Requirement already satisfied: typing-extensions>=4.2.0 in c:\users\dabir
\anaconda3\lib\site-packages (from pydantic<1.11,>=1.8.1->ydata-profiling-
>pandas_profiling) (4.4.0)
Requirement already satisfied: charset-normalizer<3,>=2 in c:\users\dabir
\anaconda3\lib\site-packages (from requests<2.29,>=2.24.0->ydata-profiling
->pandas_profiling) (2.0.4)
Requirement already satisfied: idna<4,>=2.5 in c:\users\dabir\anaconda3\li
b\site-packages (from requests<2.29,>=2.24.0->ydata-profiling->pandas_prof
iling) (3.4)
Requirement already satisfied: certifi>=2017.4.17 in c:\users\dabir\anacon
da3\lib\site-packages (from requests<2.29,>=2.24.0->ydata-profiling->panda
s_profiling) (2022.12.7)
Requirement already satisfied: urllib3<1.27,>=1.21.1 in c:\users\dabir\ana
conda3\lib\site-packages (from requests<2.29,>=2.24.0->ydata-profiling->pa
ndas_profiling) (1.26.14)
Requirement already satisfied: patsy>=0.5.2 in c:\users\dabir\anaconda3\li
b\site-packages (from statsmodels<0.14,>=0.13.2->ydata-profiling->pandas_p
rofiling) (0.5.3)
Requirement already satisfied: colorama in c:\users\dabir\anaconda3\lib\si
te-packages (from tqdm<4.65,>=4.48.2->ydata-profiling->pandas_profiling)
(0.4.6)
Requirement already satisfied: six in c:\users\dabir\anaconda3\lib\site-pa
ckages (from patsy>=0.5.2->statsmodels<0.14,>=0.13.2->ydata-profiling->pan
das_profiling) (1.16.0)
```

In [55]:

```
import pandas_profiling as pp
profile = pp.ProfileReport(unicorn)
profile.to_file(output_file="unicorn_before_preprocessing.html")
```

Summarize dataset: 21/21 [00:02<00:00, 9.04it/s, Completed]

100%

Generate report structure: 1/1 [00:01<00:00, 1.77s/it]

100%

Render HTML: 1/1 [00:00<00:00, 1.52it/s]

100%

Export report to file: 1/1 [00:00<00:00, 69.04it/s]

100%

In [ ]:

```

```

In [57]:

```
import os
```

In [58]:

```
os.getcwd
```

Out[58]:

```
<function nt.getcwd()>
```

In [ ]:

```

```