# Top 50 US Tech Companies Analysis

# Collecting Dataset :

We have collected this Dataset from kaggle Dataset link :
https://www.kaggle.com/datasets/ogbuzurukelechi/2022-to-2023s-top-50-us-tech-companies

# Importing python Libraries

In [1]:
```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
import plotly.express as px
```

# Load the dataset

In [2]:
```
df = pd.read_csv("G:/Project/Ptyhon Project/Top 50 US Tech Company Analysis/Top 50 US Te
df
```

Out[2]:

| | Company Name | Industry | Sector | HQ State | Founding Year | Annual Revenue 2022-2023 (USD in Billions) | Market Cap (USD in Trillions) | Stock Name | Annual Income Tax in 2022-2023 (USD in Billions) | Er |
|---|---|---|---|---|---|---|---|---|---|---|
| **0** | Apple Inc. | Technology | Consumer Electronics | California | 1905 | 387.53 | 2.520 | AAPL | 18.314 | |
| **1** | Microsoft Corporation | Technology | Software Infrastructure | Washington | 1905 | 204.09 | 2.037 | MSFT | 15.139 | |
| **2** | Alphabet (Google) | Technology | Software Infrastructure | California | 1905 | 282.83 | 1.350 | GOOG | 11.356 | |
| **3** | Amazon | Technology | Software Application | Washington | 1905 | 513.98 | 1.030 | AMZN | -3.217 | |
| **4** | NVIDIA Corporation | Technology | Semiconductors | California | 1905 | 26.97 | 0.653 | NVDA | 0.189 | |
| **5** | Tesla | Technology | Software Infrastructure | Texas | 1905 | 81.46 | 0.625 | TSLA | 1.132 | |
| **6** | Meta Platforms | Technology | Software Infrastructure | California | 1905 | 116.60 | 0.524 | META | 5.619 | |
| **7** | Broadcom Inc. | Technology | Semiconductors | California | 1905 | 34.41 | 0.266 | AVGO | 0.939 | |
| **8** | Oracle Corporation | Technology | Software Infrastructure | Texas | 1905 | 46.07 | 0.236 | ORCL | 0.932 | |
| **9** | Cisco Systems Inc. | Technology | Communication Equipments | California | 1905 | 53.16 | 0.208 | CSCO | 2.665 | |

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| **10** | Salesforce Inc. | Technology | Software Application | California | 1905 | 31.35 | 0.189 | CRM | 0.452 |
| **11** | Adobe Inc. | Technology | Software Infrastructure | California | 1905 | 17.60 | 0.172 | ADBE | 1.252 |
| **12** | Texas Instruments Inc. | Technology | Semiconductors | Texas | 1905 | 20.02 | 0.162 | TXN | 1.283 |
| **13** | Advanced Micro Devices (AMD) Inc. | Technology | Semiconductors | California | 1905 | 23.60 | 0.155 | AMD | -0.122 |
| **14** | Qualcomm Inc. | Technology | Semiconductors | California | 1905 | 42.95 | 0.138 | QCOM | 2.012 |
| **15** | Netflix | Technology | Software Application | California | 1905 | 31.61 | 0.136 | NFLX | 0.772 |
| **16** | Intel Corporation | Technology | Semiconductors | California | 1905 | 63.05 | 0.118 | INTC | -0.249 |
| **17** | Intuit Inc. | Technology | Software Application | California | 1905 | 13.68 | 0.118 | INTU | 0.476 |
| **18** | IBM Corporation | Technology | IT Services | New York | 1905 | 60.52 | 0.113 | IBM | -0.626 |
| **19** | Applied Materials Inc. | Technology | Semiconductors | California | 1905 | 26.25 | 0.102 | AMAT | 1.074 |
| **20** | Booking Holdings | Technology | Software Application | Connecticut | 1905 | 17.09 | 0.097 | BKNG | 0.865 |
| **21** | Analog Devices Inc. | Technology | Semiconductors | Massachusetts | 1905 | 12.57 | 0.095 | ADI | 0.350 |
| **22** | ServiceNow Inc. | Technology | Software Application | California | 1905 | 7.24 | 0.090 | NOW | 0.074 |
| **23** | Automatic Data Processing | Technology | Software Application | New Jersey | 1905 | 16.67 | 0.090 | ADP | 0.855 |
| **24** | PayPal Holdings Inc. | Technology | Software Infrastructure | California | 1905 | 27.51 | 0.087 | PYPL | 0.947 |
| **25** | Airbnb | Technology | Software Application | California | 1905 | 8.39 | 0.078 | ABNB | 0.096 |
| **26** | Fiserv Inc. | Technology | IT Services | Wisconsin | 1905 | 17.73 | 0.071 | FISV | 0.551 |
| **27** | Lam Research Corporation | Technology | Semiconductors | California | 1905 | 19.04 | 0.069 | LRCX | 0.588 |
| **28** | Uber Technologies Inc. | Technology | Software Application | California | 1905 | 31.87 | 0.066 | UBER | -0.181 |
| **29** | Micron Technology | Technology | Semiconductors | Idaho | 1905 | 27.15 | 0.064 | MU | 0.888 |
| **30** | Equinix | Technology | IT Services | California | 1905 | 7.26 | 0.064 | EQIX | 0.125 |
| **31** | Activision Blizzard | Technology | Software Application | California | 1905 | 7.52 | 0.063 | ATVI | 0.231 |
| **32** | Palo Alto Networks Inc. | Technology | Software Infrastructure | California | 1905 | 6.15 | 0.059 | PANW | 0.060 |
| **33** | Synopsys Inc. | Technology | Software Infrastructure | California | 1905 | 5.17 | 0.057 | SNPS | 0.137 |

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| **34** | Cadence Design Systems Inc. | Technology | Software Application | California | 1905 | 3.56 | 0.057 | CDNS | 0.196 |
| **35** | KLA Corporation | Technology | Semiconductors | California | 1905 | 10.48 | 0.053 | KLAC | 0.167 |
| **36** | Arista Networks Inc. | Technology | Computer Hardware | California | 1905 | 4.38 | 0.052 | ANET | 0.229 |
| **37** | VMware Inc. | Technology | Software Infrastructure | California | 1905 | 13.34 | 0.051 | VMW | 0.265 |
| **38** | Workday Inc. | Technology | Software Application | California | 1905 | 6.21 | 0.049 | WDAY | 0.107 |
| **39** | Fortinet Inc. | Technology | Software Infrastructure | California | 1905 | 4.41 | 0.049 | FTNT | 0.031 |
| **40** | Block Inc. | Technology | Software Infrastructure | California | 1905 | 17.53 | 0.047 | SQ | -0.012 |
| **41** | Snowflake Inc. | Technology | Software Application | Montana | 1905 | 2.06 | 0.046 | SNOW | 0.003 |
| **42** | Roper Technologies | Technology | Electronic Components | Florida | 1905 | 5.61 | 0.046 | ROP | 0.296 |
| **43** | Microchip Technology Inc. | Technology | Semiconductors | Arizona | 1905 | 8.05 | 0.045 | MCHP | 0.197 |
| **44** | Autodesk Inc. | Technology | Software Application | California | 1905 | 5.00 | 0.045 | ADSK | 0.068 |
| **45** | GlobalFoundries | Technology | Semiconductors | New York | 1905 | 8.10 | 0.038 | GFS | 0.086 |
| **46** | IQVIA Holdings | Technology | Software Application | North Carolina | 1905 | 14.41 | 0.037 | IQV | 0.260 |
| **47** | Marvell Technology Inc. | Technology | Semiconductors | California | 1905 | 5.91 | 0.035 | MRVL | 0.249 |
| **48** | Dell Technologies Inc. | Technology | Computer Hardware | Texas | 1905 | 102.30 | 0.028 | DELL | 0.981 |
| **49** | HP Inc. | Technology | Computer Hardware | California | 1905 | 59.78 | 0.028 | HPQ | 1.238 |

In [3]:
```
# Upper 5 recoords
df.head()
```

Out[3]:

| | Company Name | Industry | Sector | HQ State | Founding Year | Annual Revenue 2022-2023 (USD in Billions) | Market Cap (USD in Trillions) | Stock Name | Annual Income Tax in 2022-2023 (USD in Billions) | Employee Size |
|---|---|---|---|---|---|---|---|---|---|---|
| **0** | Apple Inc. | Technology | Consumer Electronics | California | 1905 | 387.53 | 2.520 | AAPL | 18.314 | 164000 |
| **1** | Microsoft Corporation | Technology | Software Infrastructure | Washington | 1905 | 204.09 | 2.037 | MSFT | 15.139 | 221000 |
| **2** | Alphabet (Google) | Technology | Software Infrastructure | California | 1905 | 282.83 | 1.350 | GOOG | 11.356 | 190234 |
| **3** | Amazon | Technology | Software | Washington | 1905 | 513.98 | 1.030 | AMZN | -3.217 | 1541000 |

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | | Application | | | | | | | |
| **4** | NVIDIA Corporation | Technology | Semiconductors | California | 1905 | 26.97 | 0.653 | NVDA | 0.189 | 22473 |

In [4]:
```
# Lower  5 Records
df.tail()
```

Out[4]:

| | Company Name | Industry | Sector | HQ State | Founding Year | Annual Revenue 2022-2023 (USD in Billions) | Market Cap (USD in Trillions) | Stock Name | Annual Income Tax in 2022-2023 (USD in Billions) | Employ S |
|---|---|---|---|---|---|---|---|---|---|---|
| **45** | GlobalFoundries | Technology | Semiconductors | New York | 1905 | 8.10 | 0.038 | GFS | 0.086 | 146 |
| **46** | IQVIA Holdings | Technology | Software Application | North Carolina | 1905 | 14.41 | 0.037 | IQV | 0.260 | 850 |
| **47** | Marvell Technology Inc. | Technology | Semiconductors | California | 1905 | 5.91 | 0.035 | MRVL | 0.249 | 66 |
| **48** | Dell Technologies Inc. | Technology | Computer Hardware | Texas | 1905 | 102.30 | 0.028 | DELL | 0.981 | 133( |
| **49** | HP Inc. | Technology | Computer Hardware | California | 1905 | 59.78 | 0.028 | HPQ | 1.238 | 51( |

In [5]:
```
# number of row & column
df.shape
```

Out[5]:
```
(50, 10)
```

In [6]:
```
#Number of rows
df.shape[0]
```

Out[6]:
```
50
```

In [7]:
```
# Number of column
df.shape[1]
```

Out[7]:
```
10
```

In [8]:
```
print("The number of column is",df.shape[1])
print("The number of row is",df.shape[0])
```

```
The number of column is 10
The number of row is 50
```

# Data Summary

In [9]:
```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 50 entries, 0 to 49
Data columns (total 10 columns):
 #   Column                                                Non-Null Count  Dtype
---  ------                                                --------------  -----
```

```
 0   Company Name                                        50 non-null     object
 1   Industry                                            50 non-null     object
 2   Sector                                              50 non-null     object
 3   HQ State                                            50 non-null     object
 4   Founding Year                                       50 non-null     int64
 5   Annual Revenue 2022-2023 (USD in Billions)          50 non-null     float64
 6   Market Cap (USD in Trillions)                       50 non-null     float64
 7   Stock Name                                          50 non-null     object
 8   Annual Income Tax in 2022-2023 (USD in Billions)    50 non-null     float64
 9   Employee Size                                       50 non-null     int64
dtypes: float64(3), int64(2), object(5)
memory usage: 4.0+ KB
```

In [10]:
```
df.describe() # Numerically describe
```

Out[10]:

| | Founding Year | Annual Revenue 2022-2023 (USD in Billions) | Market Cap (USD in Trillions) | Annual Income Tax in 2022-2023 (USD in Billions) | Employee Size |
|---|---|---|---|---|---|
| **count** | 50.0 | 50.00000 | 50.000000 | 50.000000 | 5.000000e+01 |
| **mean** | 1905.0 | 51.20440 | 0.252160 | 1.386780 | 8.324962e+04 |
| **std** | 0.0 | 97.41288 | 0.490377 | 3.687916 | 2.205869e+05 |
| **min** | 1905.0 | 2.06000 | 0.028000 | -3.217000 | 2.993000e+03 |
| **25%** | 1905.0 | 7.65250 | 0.051250 | 0.098750 | 1.415000e+04 |
| **50%** | 1905.0 | 17.66500 | 0.082500 | 0.280500 | 2.472500e+04 |
| **75%** | 1905.0 | 40.81500 | 0.160250 | 0.945000 | 7.015575e+04 |
| **max** | 1905.0 | 513.98000 | 2.520000 | 18.314000 | 1.541000e+06 |

In [11]:
```
round(df.describe(),2) # upto 2 decimal
```

Out[11]:

| | Founding Year | Annual Revenue 2022-2023 (USD in Billions) | Market Cap (USD in Trillions) | Annual Income Tax in 2022-2023 (USD in Billions) | Employee Size |
|---|---|---|---|---|---|
| **count** | 50.0 | 50.00 | 50.00 | 50.00 | 50.00 |
| **mean** | 1905.0 | 51.20 | 0.25 | 1.39 | 83249.62 |
| **std** | 0.0 | 97.41 | 0.49 | 3.69 | 220586.93 |
| **min** | 1905.0 | 2.06 | 0.03 | -3.22 | 2993.00 |
| **25%** | 1905.0 | 7.65 | 0.05 | 0.10 | 14150.00 |
| **50%** | 1905.0 | 17.66 | 0.08 | 0.28 | 24725.00 |
| **75%** | 1905.0 | 40.82 | 0.16 | 0.94 | 70155.75 |
| **max** | 1905.0 | 513.98 | 2.52 | 18.31 | 1541000.00 |

In [12]:
```
round(df.describe(),2).T # it will Transpose the matrics. here 25% is Q1, 50% is Q2(medi
```

Out[12]:

| | count | mean | std | min | 25% | 50% | 75% | max |
|---|---|---|---|---|---|---|---|---|
| **Founding Year** | 50.0 | 1905.00 | 0.00 | 1905.00 | 1905.00 | 1905.00 | 1905.00 | 1905.00 |
| **Annual Revenue 2022-2023 (USD in Billions)** | 50.0 | 51.20 | 97.41 | 2.06 | 7.65 | 17.66 | 40.82 | 513.98 |
| **Market Cap (USD in Trillions)** | 50.0 | 0.25 | 0.49 | 0.03 | 0.05 | 0.08 | 0.16 | 2.52 |
| **Annual Income Tax in 2022-2023 (USD in Billions)** | 50.0 | 1.39 | 3.69 | -3.22 | 0.10 | 0.28 | 0.94 | 18.31 |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| **Employee Size** | 50.0 | 83249.62 | 220586.93 | 2993.00 | 14150.00 | 24725.00 | 70155.75 | 1541000.00 |

In [13]:
```
round(df.describe(include='all'),2).T   #it will describe the whole data includimg objec
```

Out[13]:

| | count | unique | top | freq | mean | std | min | 25% | 50% | 75% | max |
|---|---|---|---|---|---|---|---|---|---|---|---|
| **Company Name** | 50 | 50 | Apple Inc. | 1 | NaN | NaN | NaN | NaN | NaN | NaN | NaN |
| **Industry** | 50 | 1 | Technology | 50 | NaN | NaN | NaN | NaN | NaN | NaN | NaN |
| **Sector** | 50 | 8 | Software Application | 15 | NaN | NaN | NaN | NaN | NaN | NaN | NaN |
| **HQ State** | 50 | 13 | California | 33 | NaN | NaN | NaN | NaN | NaN | NaN | NaN |
| **Founding Year** | 50.0 | NaN | NaN | NaN | 1905.0 | 0.0 | 1905.0 | 1905.0 | 1905.0 | 1905.0 | 1905.0 |
| **Annual Revenue 2022-2023 (USD in Billions)** | 50.0 | NaN | NaN | NaN | 51.2 | 97.41 | 2.06 | 7.65 | 17.66 | 40.82 | 513.98 |
| **Market Cap (USD in Trillions)** | 50.0 | NaN | NaN | NaN | 0.25 | 0.49 | 0.03 | 0.05 | 0.08 | 0.16 | 2.52 |
| **Stock Name** | 50 | 50 | AAPL | 1 | NaN | NaN | NaN | NaN | NaN | NaN | NaN |
| **Annual Income Tax in 2022-2023 (USD in Billions)** | 50.0 | NaN | NaN | NaN | 1.39 | 3.69 | -3.22 | 0.1 | 0.28 | 0.94 | 18.31 |
| **Employee Size** | 50.0 | NaN | NaN | NaN | 83249.62 | 220586.93 | 2993.0 | 14150.0 | 24725.0 | 70155.75 | 1541000.0 |

# Data Cleaning

# Checkimg for duplicates

In [14]:
```
df.duplicated(keep =False).sum()
```

Out[14]:
```
0
```

# Checking for missing Values,wehere true there will be null

In [15]:
```
df.isnull()
```

Out[15]:

| | Company Name | Industry | Sector | HQ State | Founding Year | Annual Revenue 2022-2023 (USD in Billions) | Market Cap (USD in Trillions) | Stock Name | Annual Income Tax in 2022-2023 (USD in Billions) | Employee Size |
|---|---|---|---|---|---|---|---|---|---|---|
| **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 | False | 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 |
| **5** | False | False | False | False | False | False | False | False | False | False |
| **6** | False | False | False | False | False | False | False | False | False | False |
| **7** | False | False | False | False | False | False | False | False | False | False |
| **8** | False | False | False | False | False | False | False | False | False | False |
| **9** | False | False | False | False | False | False | False | False | False | False |
| **10** | False | False | False | False | False | False | False | False | False | False |
| **11** | False | False | False | False | False | False | False | False | False | False |
| **12** | False | False | False | False | False | False | False | False | False | False |
| **13** | False | False | False | False | False | False | False | False | False | False |
| **14** | False | False | False | False | False | False | False | False | False | False |
| **15** | False | False | False | False | False | False | False | False | False | False |
| **16** | False | False | False | False | False | False | False | False | False | False |
| **17** | False | False | False | False | False | False | False | False | False | False |
| **18** | False | False | False | False | False | False | False | False | False | False |
| **19** | False | False | False | False | False | False | False | False | False | False |
| **20** | False | False | False | False | False | False | False | False | False | False |
| **21** | False | False | False | False | False | False | False | False | False | False |
| **22** | False | False | False | False | False | False | False | False | False | False |
| **23** | False | False | False | False | False | False | False | False | False | False |
| **24** | False | False | False | False | False | False | False | False | False | False |
| **25** | False | False | False | False | False | False | False | False | False | False |
| **26** | False | False | False | False | False | False | False | False | False | False |
| **27** | False | False | False | False | False | False | False | False | False | False |
| **28** | False | False | False | False | False | False | False | False | False | False |
| **29** | False | False | False | False | False | False | False | False | False | False |
| **30** | False | False | False | False | False | False | False | False | False | False |
| **31** | False | False | False | False | False | False | False | False | False | False |
| **32** | False | False | False | False | False | False | False | False | False | False |
| **33** | False | False | False | False | False | False | False | False | False | False |

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| **34** | False | False | False | False | False | False | False | False | False | False |
| **35** | False | False | False | False | False | False | False | False | False | False |
| **36** | False | False | False | False | False | False | False | False | False | False |
| **37** | False | False | False | False | False | False | False | False | False | False |
| **38** | False | False | False | False | False | False | False | False | False | False |
| **39** | False | False | False | False | False | False | False | False | False | False |
| **40** | False | False | False | False | False | False | False | False | False | False |
| **41** | False | False | False | False | False | False | False | False | False | False |
| **42** | False | False | False | False | False | False | False | False | False | False |
| **43** | False | False | False | False | False | False | False | False | False | False |
| **44** | False | False | False | False | False | False | False | False | False | False |
| **45** | False | False | False | False | False | False | False | False | False | False |
| **46** | False | False | False | False | False | False | False | False | False | False |
| **47** | False | False | False | False | False | False | False | False | False | False |
| **48** | False | False | False | False | False | False | False | False | False | False |
| **49** | False | False | False | False | False | False | False | False | False | False |

# Total of the null value in each column

In [16]:
```
df.isnull().sum()
```

Out[16]:
```
Company Name                                        0
Industry                                            0
Sector                                              0
HQ State                                            0
Founding Year                                       0
Annual Revenue 2022-2023 (USD in Billions)          0
Market Cap (USD in Trillions)                       0
Stock Name                                          0
Annual Income Tax in 2022-2023 (USD in Billions)    0
Employee Size                                       0
dtype: int64
```

In [17]:
```
df.columns
```

Out[17]:
```
Index(['Company Name', 'Industry', 'Sector', 'HQ State', 'Founding Year',
       'Annual Revenue 2022-2023 (USD in Billions)',
       'Market Cap (USD in Trillions)', 'Stock Name',
       'Annual Income Tax in 2022-2023 (USD in Billions)', 'Employee Size'],
      dtype='object')
```

In [18]:
```
#Copy original dataframe
df2 =df.copy()
df2
```

Out[18]:

| | Company Name | Industry | Sector | HQ State | Founding Year | Annual Revenue 2022-2023 (USD in Billions) | Market Cap (USD in Trillions) | Stock Name | Annual Income Tax in 2022-2023 | Er |
|---|---|---|---|---|---|---|---|---|---|---|

| | | | | | | | | | (USD in Billions) |
|---|---|---|---|---|---|---|---|---|---|
| **0** | Apple Inc. | Technology | Consumer Electronics | California | 1905 | 387.53 | 2.520 | AAPL | 18.314 |
| **1** | Microsoft Corporation | Technology | Software Infrastructure | Washington | 1905 | 204.09 | 2.037 | MSFT | 15.139 |
| **2** | Alphabet (Google) | Technology | Software Infrastructure | California | 1905 | 282.83 | 1.350 | GOOG | 11.356 |
| **3** | Amazon | Technology | Software Application | Washington | 1905 | 513.98 | 1.030 | AMZN | -3.217 |
| **4** | NVIDIA Corporation | Technology | Semiconductors | California | 1905 | 26.97 | 0.653 | NVDA | 0.189 |
| **5** | Tesla | Technology | Software Infrastructure | Texas | 1905 | 81.46 | 0.625 | TSLA | 1.132 |
| **6** | Meta Platforms | Technology | Software Infrastructure | California | 1905 | 116.60 | 0.524 | META | 5.619 |
| **7** | Broadcom Inc. | Technology | Semiconductors | California | 1905 | 34.41 | 0.266 | AVGO | 0.939 |
| **8** | Oracle Corporation | Technology | Software Infrastructure | Texas | 1905 | 46.07 | 0.236 | ORCL | 0.932 |
| **9** | Cisco Systems Inc. | Technology | Communication Equipments | California | 1905 | 53.16 | 0.208 | CSCO | 2.665 |
| **10** | Salesforce Inc. | Technology | Software Application | California | 1905 | 31.35 | 0.189 | CRM | 0.452 |
| **11** | Adobe Inc. | Technology | Software Infrastructure | California | 1905 | 17.60 | 0.172 | ADBE | 1.252 |
| **12** | Texas Instruments Inc. | Technology | Semiconductors | Texas | 1905 | 20.02 | 0.162 | TXN | 1.283 |
| **13** | Advanced Micro Devices (AMD) Inc. | Technology | Semiconductors | California | 1905 | 23.60 | 0.155 | AMD | -0.122 |
| **14** | Qualcomm Inc. | Technology | Semiconductors | California | 1905 | 42.95 | 0.138 | QCOM | 2.012 |
| **15** | Netflix | Technology | Software Application | California | 1905 | 31.61 | 0.136 | NFLX | 0.772 |
| **16** | Intel Corporation | Technology | Semiconductors | California | 1905 | 63.05 | 0.118 | INTC | -0.249 |
| **17** | Intuit Inc. | Technology | Software Application | California | 1905 | 13.68 | 0.118 | INTU | 0.476 |
| **18** | IBM Corporation | Technology | IT Services | New York | 1905 | 60.52 | 0.113 | IBM | -0.626 |
| **19** | Applied Materials Inc. | Technology | Semiconductors | California | 1905 | 26.25 | 0.102 | AMAT | 1.074 |
| **20** | Booking Holdings | Technology | Software Application | Connecticut | 1905 | 17.09 | 0.097 | BKNG | 0.865 |
| **21** | Analog Devices Inc. | Technology | Semiconductors | Massachusetts | 1905 | 12.57 | 0.095 | ADI | 0.350 |
| **22** | ServiceNow Inc. | Technology | Software Application | California | 1905 | 7.24 | 0.090 | NOW | 0.074 |

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| **23** | Automatic Data Processing | Technology | Software Application | New Jersey | 1905 | 16.67 | 0.090 | ADP | 0.855 |
| **24** | PayPal Holdings Inc. | Technology | Software Infrastructure | California | 1905 | 27.51 | 0.087 | PYPL | 0.947 |
| **25** | Airbnb | Technology | Software Application | California | 1905 | 8.39 | 0.078 | ABNB | 0.096 |
| **26** | Fiserv Inc. | Technology | IT Services | Wisconsin | 1905 | 17.73 | 0.071 | FISV | 0.551 |
| **27** | Lam Research Corporation | Technology | Semiconductors | California | 1905 | 19.04 | 0.069 | LRCX | 0.588 |
| **28** | Uber Technologies Inc. | Technology | Software Application | California | 1905 | 31.87 | 0.066 | UBER | -0.181 |
| **29** | Micron Technology | Technology | Semiconductors | Idaho | 1905 | 27.15 | 0.064 | MU | 0.888 |
| **30** | Equinix | Technology | IT Services | California | 1905 | 7.26 | 0.064 | EQIX | 0.125 |
| **31** | Activision Blizzard | Technology | Software Application | California | 1905 | 7.52 | 0.063 | ATVI | 0.231 |
| **32** | Palo Alto Networks Inc. | Technology | Software Infrastructure | California | 1905 | 6.15 | 0.059 | PANW | 0.060 |
| **33** | Synopsys Inc. | Technology | Software Infrastructure | California | 1905 | 5.17 | 0.057 | SNPS | 0.137 |
| **34** | Cadence Design Systems Inc. | Technology | Software Application | California | 1905 | 3.56 | 0.057 | CDNS | 0.196 |
| **35** | KLA Corporation | Technology | Semiconductors | California | 1905 | 10.48 | 0.053 | KLAC | 0.167 |
| **36** | Arista Networks Inc. | Technology | Computer Hardware | California | 1905 | 4.38 | 0.052 | ANET | 0.229 |
| **37** | VMware Inc. | Technology | Software Infrastructure | California | 1905 | 13.34 | 0.051 | VMW | 0.265 |
| **38** | Workday Inc. | Technology | Software Application | California | 1905 | 6.21 | 0.049 | WDAY | 0.107 |
| **39** | Fortinet Inc. | Technology | Software Infrastructure | California | 1905 | 4.41 | 0.049 | FTNT | 0.031 |
| **40** | Block Inc. | Technology | Software Infrastructure | California | 1905 | 17.53 | 0.047 | SQ | -0.012 |
| **41** | Snowflake Inc. | Technology | Software Application | Montana | 1905 | 2.06 | 0.046 | SNOW | 0.003 |
| **42** | Roper Technologies | Technology | Electronic Components | Florida | 1905 | 5.61 | 0.046 | ROP | 0.296 |
| **43** | Microchip Technology Inc. | Technology | Semiconductors | Arizona | 1905 | 8.05 | 0.045 | MCHP | 0.197 |
| **44** | Autodesk Inc. | Technology | Software Application | California | 1905 | 5.00 | 0.045 | ADSK | 0.068 |
| **45** | GlobalFoundries | Technology | Semiconductors | New York | 1905 | 8.10 | 0.038 | GFS | 0.086 |
| **46** | IQVIA Holdings | Technology | Software Application | North Carolina | 1905 | 14.41 | 0.037 | IQV | 0.260 |
| **47** | Marvell | Technology | Semiconductors | California | 1905 | 5.91 | 0.035 | MRVL | 0.249 |

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | Technology Inc. | | | | | | | | |
| **48** | Dell Technologies Inc. | Technology | Computer Hardware | Texas | 1905 | 102.30 | 0.028 | DELL | 0.981 |
| **49** | HP Inc. | Technology | Computer Hardware | California | 1905 | 59.78 | 0.028 | HPQ | 1.238 |

In [19]:
```
df2.drop_duplicates(inplace= True )
```

In [20]:
```
df2
```

Out[20]:

| | Company Name | Industry | Sector | HQ State | Founding Year | Annual Revenue 2022-2023 (USD in Billions) | Market Cap (USD in Trillions) | Stock Name | Annual Income Tax in 2022-2023 (USD in Billions) | Er |
|---|---|---|---|---|---|---|---|---|---|---|
| **0** | Apple Inc. | Technology | Consumer Electronics | California | 1905 | 387.53 | 2.520 | AAPL | 18.314 | |
| **1** | Microsoft Corporation | Technology | Software Infrastructure | Washington | 1905 | 204.09 | 2.037 | MSFT | 15.139 | |
| **2** | Alphabet (Google) | Technology | Software Infrastructure | California | 1905 | 282.83 | 1.350 | GOOG | 11.356 | |
| **3** | Amazon | Technology | Software Application | Washington | 1905 | 513.98 | 1.030 | AMZN | -3.217 | |
| **4** | NVIDIA Corporation | Technology | Semiconductors | California | 1905 | 26.97 | 0.653 | NVDA | 0.189 | |
| **5** | Tesla | Technology | Software Infrastructure | Texas | 1905 | 81.46 | 0.625 | TSLA | 1.132 | |
| **6** | Meta Platforms | Technology | Software Infrastructure | California | 1905 | 116.60 | 0.524 | META | 5.619 | |
| **7** | Broadcom Inc. | Technology | Semiconductors | California | 1905 | 34.41 | 0.266 | AVGO | 0.939 | |
| **8** | Oracle Corporation | Technology | Software Infrastructure | Texas | 1905 | 46.07 | 0.236 | ORCL | 0.932 | |
| **9** | Cisco Systems Inc. | Technology | Communication Equipments | California | 1905 | 53.16 | 0.208 | CSCO | 2.665 | |
| **10** | Salesforce Inc. | Technology | Software Application | California | 1905 | 31.35 | 0.189 | CRM | 0.452 | |
| **11** | Adobe Inc. | Technology | Software Infrastructure | California | 1905 | 17.60 | 0.172 | ADBE | 1.252 | |
| **12** | Texas Instruments Inc. | Technology | Semiconductors | Texas | 1905 | 20.02 | 0.162 | TXN | 1.283 | |
| **13** | Advanced Micro Devices (AMD) Inc. | Technology | Semiconductors | California | 1905 | 23.60 | 0.155 | AMD | -0.122 | |
| **14** | Qualcomm Inc. | Technology | Semiconductors | California | 1905 | 42.95 | 0.138 | QCOM | 2.012 | |
| **15** | Netflix | Technology | Software Application | California | 1905 | 31.61 | 0.136 | NFLX | 0.772 | |

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| **16** | Intel Corporation | Technology | Semiconductors | California | 1905 | 63.05 | 0.118 | INTC | -0.249 |
| **17** | Intuit Inc. | Technology | Software Application | California | 1905 | 13.68 | 0.118 | INTU | 0.476 |
| **18** | IBM Corporation | Technology | IT Services | New York | 1905 | 60.52 | 0.113 | IBM | -0.626 |
| **19** | Applied Materials Inc. | Technology | Semiconductors | California | 1905 | 26.25 | 0.102 | AMAT | 1.074 |
| **20** | Booking Holdings | Technology | Software Application | Connecticut | 1905 | 17.09 | 0.097 | BKNG | 0.865 |
| **21** | Analog Devices Inc. | Technology | Semiconductors | Massachusetts | 1905 | 12.57 | 0.095 | ADI | 0.350 |
| **22** | ServiceNow Inc. | Technology | Software Application | California | 1905 | 7.24 | 0.090 | NOW | 0.074 |
| **23** | Automatic Data Processing | Technology | Software Application | New Jersey | 1905 | 16.67 | 0.090 | ADP | 0.855 |
| **24** | PayPal Holdings Inc. | Technology | Software Infrastructure | California | 1905 | 27.51 | 0.087 | PYPL | 0.947 |
| **25** | Airbnb | Technology | Software Application | California | 1905 | 8.39 | 0.078 | ABNB | 0.096 |
| **26** | Fiserv Inc. | Technology | IT Services | Wisconsin | 1905 | 17.73 | 0.071 | FISV | 0.551 |
| **27** | Lam Research Corporation | Technology | Semiconductors | California | 1905 | 19.04 | 0.069 | LRCX | 0.588 |
| **28** | Uber Technologies Inc. | Technology | Software Application | California | 1905 | 31.87 | 0.066 | UBER | -0.181 |
| **29** | Micron Technology | Technology | Semiconductors | Idaho | 1905 | 27.15 | 0.064 | MU | 0.888 |
| **30** | Equinix | Technology | IT Services | California | 1905 | 7.26 | 0.064 | EQIX | 0.125 |
| **31** | Activision Blizzard | Technology | Software Application | California | 1905 | 7.52 | 0.063 | ATVI | 0.231 |
| **32** | Palo Alto Networks Inc. | Technology | Software Infrastructure | California | 1905 | 6.15 | 0.059 | PANW | 0.060 |
| **33** | Synopsys Inc. | Technology | Software Infrastructure | California | 1905 | 5.17 | 0.057 | SNPS | 0.137 |
| **34** | Cadence Design Systems Inc. | Technology | Software Application | California | 1905 | 3.56 | 0.057 | CDNS | 0.196 |
| **35** | KLA Corporation | Technology | Semiconductors | California | 1905 | 10.48 | 0.053 | KLAC | 0.167 |
| **36** | Arista Networks Inc. | Technology | Computer Hardware | California | 1905 | 4.38 | 0.052 | ANET | 0.229 |
| **37** | VMware Inc. | Technology | Software Infrastructure | California | 1905 | 13.34 | 0.051 | VMW | 0.265 |
| **38** | Workday Inc. | Technology | Software Application | California | 1905 | 6.21 | 0.049 | WDAY | 0.107 |
| **39** | Fortinet Inc. | Technology | Software Infrastructure | California | 1905 | 4.41 | 0.049 | FTNT | 0.031 |

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| **40** | Block Inc. | Technology | Software Infrastructure | California | 1905 | 17.53 | 0.047 | SQ | -0.012 |
| **41** | Snowflake Inc. | Technology | Software Application | Montana | 1905 | 2.06 | 0.046 | SNOW | 0.003 |
| **42** | Roper Technologies | Technology | Electronic Components | Florida | 1905 | 5.61 | 0.046 | ROP | 0.296 |
| **43** | Microchip Technology Inc. | Technology | Semiconductors | Arizona | 1905 | 8.05 | 0.045 | MCHP | 0.197 |
| **44** | Autodesk Inc. | Technology | Software Application | California | 1905 | 5.00 | 0.045 | ADSK | 0.068 |
| **45** | GlobalFoundries | Technology | Semiconductors | New York | 1905 | 8.10 | 0.038 | GFS | 0.086 |
| **46** | IQVIA Holdings | Technology | Software Application | North Carolina | 1905 | 14.41 | 0.037 | IQV | 0.260 |
| **47** | Marvell Technology Inc. | Technology | Semiconductors | California | 1905 | 5.91 | 0.035 | MRVL | 0.249 |
| **48** | Dell Technologies Inc. | Technology | Computer Hardware | Texas | 1905 | 102.30 | 0.028 | DELL | 0.981 |
| **49** | HP Inc. | Technology | Computer Hardware | California | 1905 | 59.78 | 0.028 | HPQ | 1.238 |

# Modify Column names

In [21]:
```
df2= df2.rename(columns={'Annual Revenue 2022-2023 (USD in Billions)':'Annual Revenue(Bi
                         'Market Cap (USD in Trillions)': 'Market Cap(Trillions $)',
                         'Annual Income Tax in 2022-2023 (USD in Billions)': 'Annual Inco

df2
```

Out[21]:

| | Company Name | Industry | Sector | HQ State | Founding Year | Annual Revenue(Billion $) | Market Cap(Trillions $) | Stock Name | t |
|---|---|---|---|---|---|---|---|---|---|
| **0** | Apple Inc. | Technology | Consumer Electronics | California | 1905 | 387.53 | 2.520 | AAPL | |
| **1** | Microsoft Corporation | Technology | Software Infrastructure | Washington | 1905 | 204.09 | 2.037 | MSFT | |
| **2** | Alphabet (Google) | Technology | Software Infrastructure | California | 1905 | 282.83 | 1.350 | GOOG | |
| **3** | Amazon | Technology | Software Application | Washington | 1905 | 513.98 | 1.030 | AMZN | |
| **4** | NVIDIA Corporation | Technology | Semiconductors | California | 1905 | 26.97 | 0.653 | NVDA | |
| **5** | Tesla | Technology | Software Infrastructure | Texas | 1905 | 81.46 | 0.625 | TSLA | |
| **6** | Meta Platforms | Technology | Software Infrastructure | California | 1905 | 116.60 | 0.524 | META | |
| **7** | Broadcom Inc. | Technology | Semiconductors | California | 1905 | 34.41 | 0.266 | AVGO | |
| **8** | Oracle | Technology | Software | Texas | 1905 | 46.07 | 0.236 | ORCL | |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | Corporation | | Infrastructure | | | | | |
| **9** | Cisco Systems Inc. | Technology | Communication Equipments | California | 1905 | 53.16 | 0.208 | CSCO |
| **10** | Salesforce Inc. | Technology | Software Application | California | 1905 | 31.35 | 0.189 | CRM |
| **11** | Adobe Inc. | Technology | Software Infrastructure | California | 1905 | 17.60 | 0.172 | ADBE |
| **12** | Texas Instruments Inc. | Technology | Semiconductors | Texas | 1905 | 20.02 | 0.162 | TXN |
| **13** | Advanced Micro Devices (AMD) Inc. | Technology | Semiconductors | California | 1905 | 23.60 | 0.155 | AMD |
| **14** | Qualcomm Inc. | Technology | Semiconductors | California | 1905 | 42.95 | 0.138 | QCOM |
| **15** | Netflix | Technology | Software Application | California | 1905 | 31.61 | 0.136 | NFLX |
| **16** | Intel Corporation | Technology | Semiconductors | California | 1905 | 63.05 | 0.118 | INTC |
| **17** | Intuit Inc. | Technology | Software Application | California | 1905 | 13.68 | 0.118 | INTU |
| **18** | IBM Corporation | Technology | IT Services | New York | 1905 | 60.52 | 0.113 | IBM |
| **19** | Applied Materials Inc. | Technology | Semiconductors | California | 1905 | 26.25 | 0.102 | AMAT |
| **20** | Booking Holdings | Technology | Software Application | Connecticut | 1905 | 17.09 | 0.097 | BKNG |
| **21** | Analog Devices Inc. | Technology | Semiconductors | Massachusetts | 1905 | 12.57 | 0.095 | ADI |
| **22** | ServiceNow Inc. | Technology | Software Application | California | 1905 | 7.24 | 0.090 | NOW |
| **23** | Automatic Data Processing | Technology | Software Application | New Jersey | 1905 | 16.67 | 0.090 | ADP |
| **24** | PayPal Holdings Inc. | Technology | Software Infrastructure | California | 1905 | 27.51 | 0.087 | PYPL |
| **25** | Airbnb | Technology | Software Application | California | 1905 | 8.39 | 0.078 | ABNB |
| **26** | Fiserv Inc. | Technology | IT Services | Wisconsin | 1905 | 17.73 | 0.071 | FISV |
| **27** | Lam Research Corporation | Technology | Semiconductors | California | 1905 | 19.04 | 0.069 | LRCX |
| **28** | Uber Technologies Inc. | Technology | Software Application | California | 1905 | 31.87 | 0.066 | UBER |
| **29** | Micron Technology | Technology | Semiconductors | Idaho | 1905 | 27.15 | 0.064 | MU |
| **30** | Equinix | Technology | IT Services | California | 1905 | 7.26 | 0.064 | EQIX |
| **31** | Activision Blizzard | Technology | Software Application | California | 1905 | 7.52 | 0.063 | ATVI |
| **32** | Palo Alto Networks Inc. | Technology | Software Infrastructure | California | 1905 | 6.15 | 0.059 | PANW |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| **33** | Synopsys Inc. | Technology | Software Infrastructure | California | 1905 | 5.17 | 0.057 | SNPS |
| **34** | Cadence Design Systems Inc. | Technology | Software Application | California | 1905 | 3.56 | 0.057 | CDNS |
| **35** | KLA Corporation | Technology | Semiconductors | California | 1905 | 10.48 | 0.053 | KLAC |
| **36** | Arista Networks Inc. | Technology | Computer Hardware | California | 1905 | 4.38 | 0.052 | ANET |
| **37** | VMware Inc. | Technology | Software Infrastructure | California | 1905 | 13.34 | 0.051 | VMW |
| **38** | Workday Inc. | Technology | Software Application | California | 1905 | 6.21 | 0.049 | WDAY |
| **39** | Fortinet Inc. | Technology | Software Infrastructure | California | 1905 | 4.41 | 0.049 | FTNT |
| **40** | Block Inc. | Technology | Software Infrastructure | California | 1905 | 17.53 | 0.047 | SQ |
| **41** | Snowflake Inc. | Technology | Software Application | Montana | 1905 | 2.06 | 0.046 | SNOW |
| **42** | Roper Technologies | Technology | Electronic Components | Florida | 1905 | 5.61 | 0.046 | ROP |
| **43** | Microchip Technology Inc. | Technology | Semiconductors | Arizona | 1905 | 8.05 | 0.045 | MCHP |
| **44** | Autodesk Inc. | Technology | Software Application | California | 1905 | 5.00 | 0.045 | ADSK |
| **45** | GlobalFoundries | Technology | Semiconductors | New York | 1905 | 8.10 | 0.038 | GFS |
| **46** | IQVIA Holdings | Technology | Software Application | North Carolina | 1905 | 14.41 | 0.037 | IQV |
| **47** | Marvell Technology Inc. | Technology | Semiconductors | California | 1905 | 5.91 | 0.035 | MRVL |
| **48** | Dell Technologies Inc. | Technology | Computer Hardware | Texas | 1905 | 102.30 | 0.028 | DELL |
| **49** | HP Inc. | Technology | Computer Hardware | California | 1905 | 59.78 | 0.028 | HPQ |

# Exploratory Data Analysis

# Summary Statistics

In [22]:
```
# Calculate the sum of Annual Revenue
annual_revenue_sum = df2['Annual Revenue(Billion $)'].sum()
print(f"Sum of Annual Revenue: {annual_revenue_sum} billion USD")
```

```
Sum of Annual Revenue: 2560.2200000000003 billion USD
```

In [23]:
```
# Calculate the Sum of market Cap
Market_cap_sum = df2['Market Cap(Trillions $)'].sum()
```

```
print(f"Sum of Market Cap: {Market_cap_sum} trillion USD")
```

```
Sum of Market Cap: 12.608000000000002 trillion USD
```

In [24]:
```
# Calculate the sum of employees
Employee_size_sum= df2['Employee Size'].sum()
print(f"sum of Employee Size:{Employee_size_sum}")
```

```
sum of Employee Size:4162481
```

In [25]:
```
# Calculate the sum of annual Income Tax
Income_tax_sum= df2['Annual Income tax(Billion $)'].sum()
print(f'sum of Annual Income Tax:{Income_tax_sum} billion USD')
```

```
sum of Annual Income Tax:69.339 billion USD
```

# Ranking Companies : annual Revenue , Market Cap, income tax, employee size

# Annual Revenue

In [26]:
```
# Grouping the data
annual_revenue = df2.groupby('Company Name')['Annual Revenue(Billion $)'].sum().reset_in


#sort the result in descending order to find the company highest annual revenue
annual_revenue = annual_revenue.sort_values(by='Annual Revenue(Billion $)',ascending=Fal

annual_revenue.head(10)
```

Out[26]:

| | Company Name | Annual Revenue(Billion $) |
|---|---|---|
| **0** | Amazon | 513.98 |
| **1** | Apple Inc. | 387.53 |
| **2** | Alphabet (Google) | 282.83 |
| **3** | Microsoft Corporation | 204.09 |
| **4** | Meta Platforms | 116.60 |
| **5** | Dell Technologies Inc. | 102.30 |
| **6** | Tesla | 81.46 |
| **7** | Intel Corporation | 63.05 |
| **8** | IBM Corporation | 60.52 |
| **9** | HP Inc. | 59.78 |

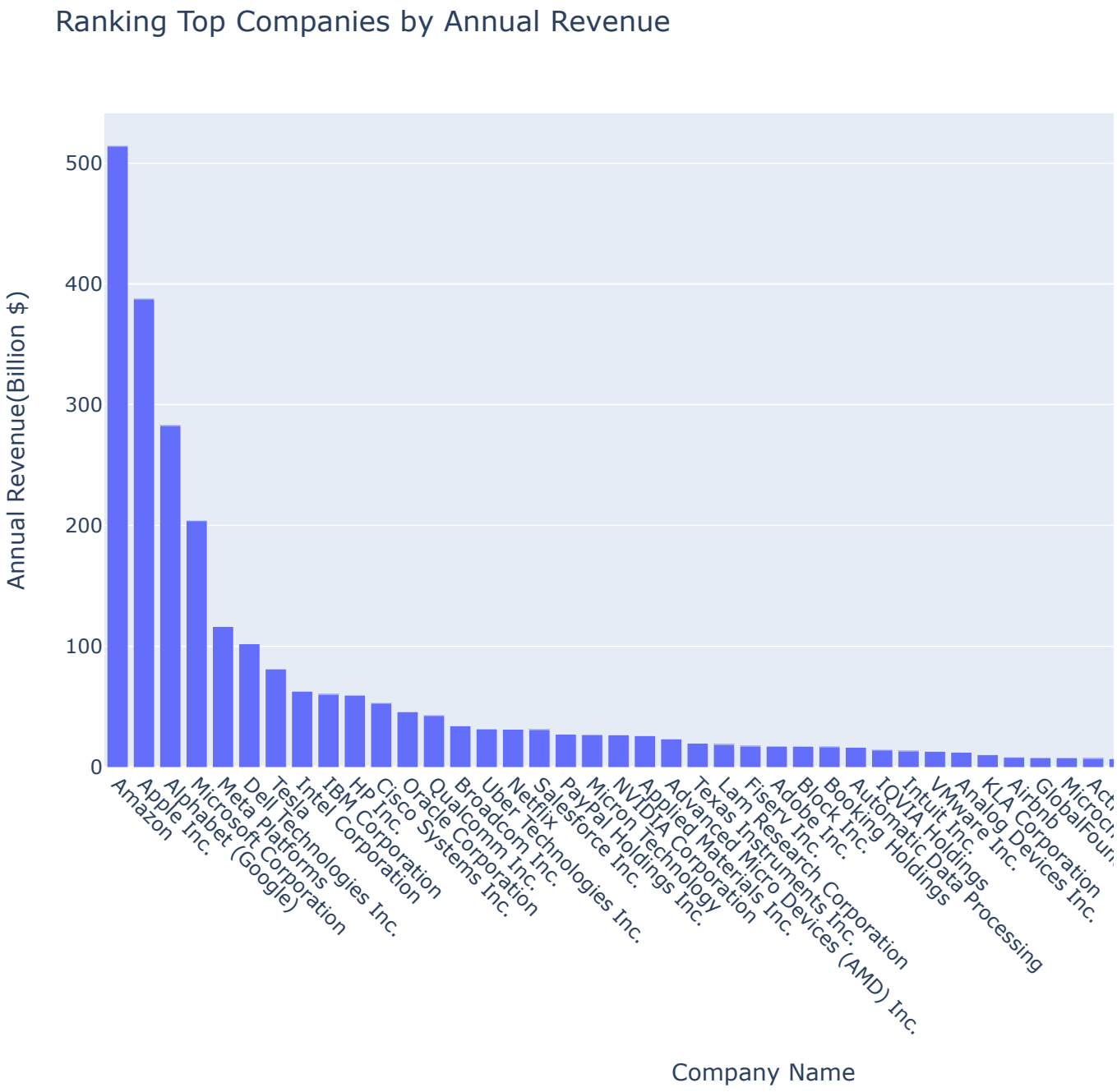
In [27]:
```
#Lowest 5 company by Revenue
annual_revenue.tail()
```

Out[27]:

| | Company Name | Annual Revenue(Billion $) |
|---|---|---|
| **45** | Autodesk Inc. | 5.00 |
| **46** | Fortinet Inc. | 4.41 |
| **47** | Arista Networks Inc. | 4.38 |

| | | |
|---|---|---|
| **48** | Cadence Design Systems Inc. | 3.56 |
| **49** | Snowflake Inc. | 2.06 |

In [28]:
```
# Creating a bar graph to visualize this data
fig= px.bar(annual_revenue, x ='Company Name',y='Annual Revenue(Billion $)', title='Rank

# Customize the layout (e.g., axis labels, rotation of  x-axis labels)
fig.update_xaxes(tickangle=45)


# set the figure size
fig.update_layout(width=1000, height = 700)

#show the plot
fig.show()
```



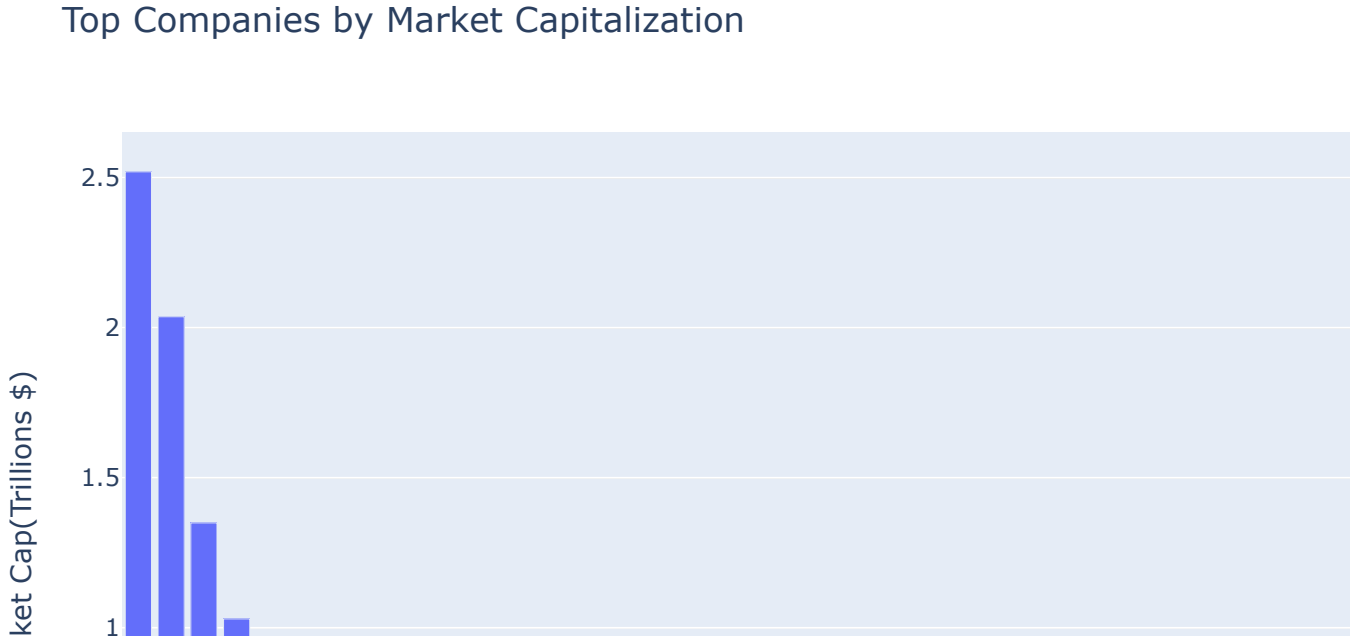
# Market Cap

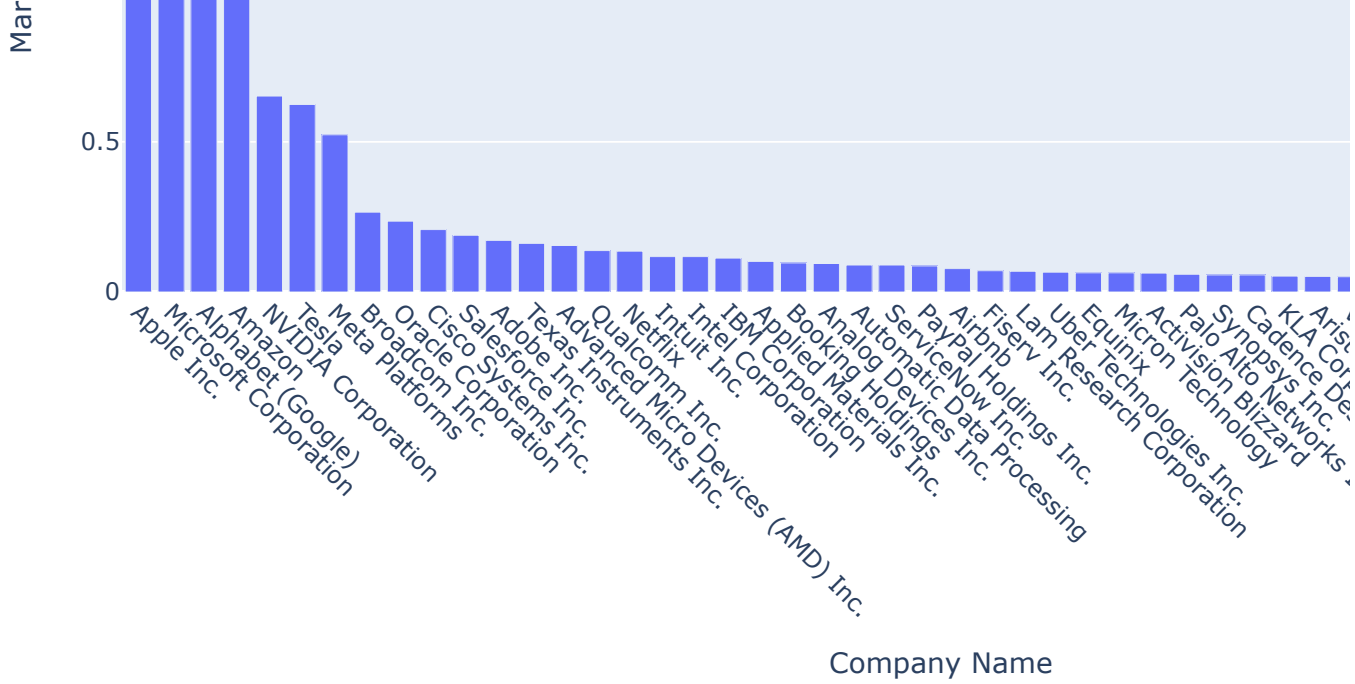
In [29]:
```
#Grouping the data
market_cap = df2.groupby('Company Name')['Market Cap(Trillions $)'].sum().reset_index()

#sort the resut in descending order to find the companies with the highest cap
market_cap =market_cap.sort_values(by ='Market Cap(Trillions $)', ascending = False).res
market_cap.head(10)
```

Out[29]:

| | Company Name | Market Cap(Trillions $) |
|---|---|---|
| **0** | Apple Inc. | 2.520 |
| **1** | Microsoft Corporation | 2.037 |
| **2** | Alphabet (Google) | 1.350 |
| **3** | Amazon | 1.030 |
| **4** | NVIDIA Corporation | 0.653 |
| **5** | Tesla | 0.625 |
| **6** | Meta Platforms | 0.524 |
| **7** | Broadcom Inc. | 0.266 |
| **8** | Oracle Corporation | 0.236 |
| **9** | Cisco Systems Inc. | 0.208 |

In [30]:
```
# Creating a bar graph to visualize this data
fig = px.bar(market_cap, x = "Company Name", y = "Market Cap(Trillions $)",
             title = "Top Companies by Market Capitalization")

# Customize the layout (e.g., axis labels, rotaion of x-axis labels)
fig.update_xaxes(tickangle= 45)

# set the figure size
fig.update_layout(width =1000, height= 700)

#show the plot
fig.show()
```






# Annual Income Tax

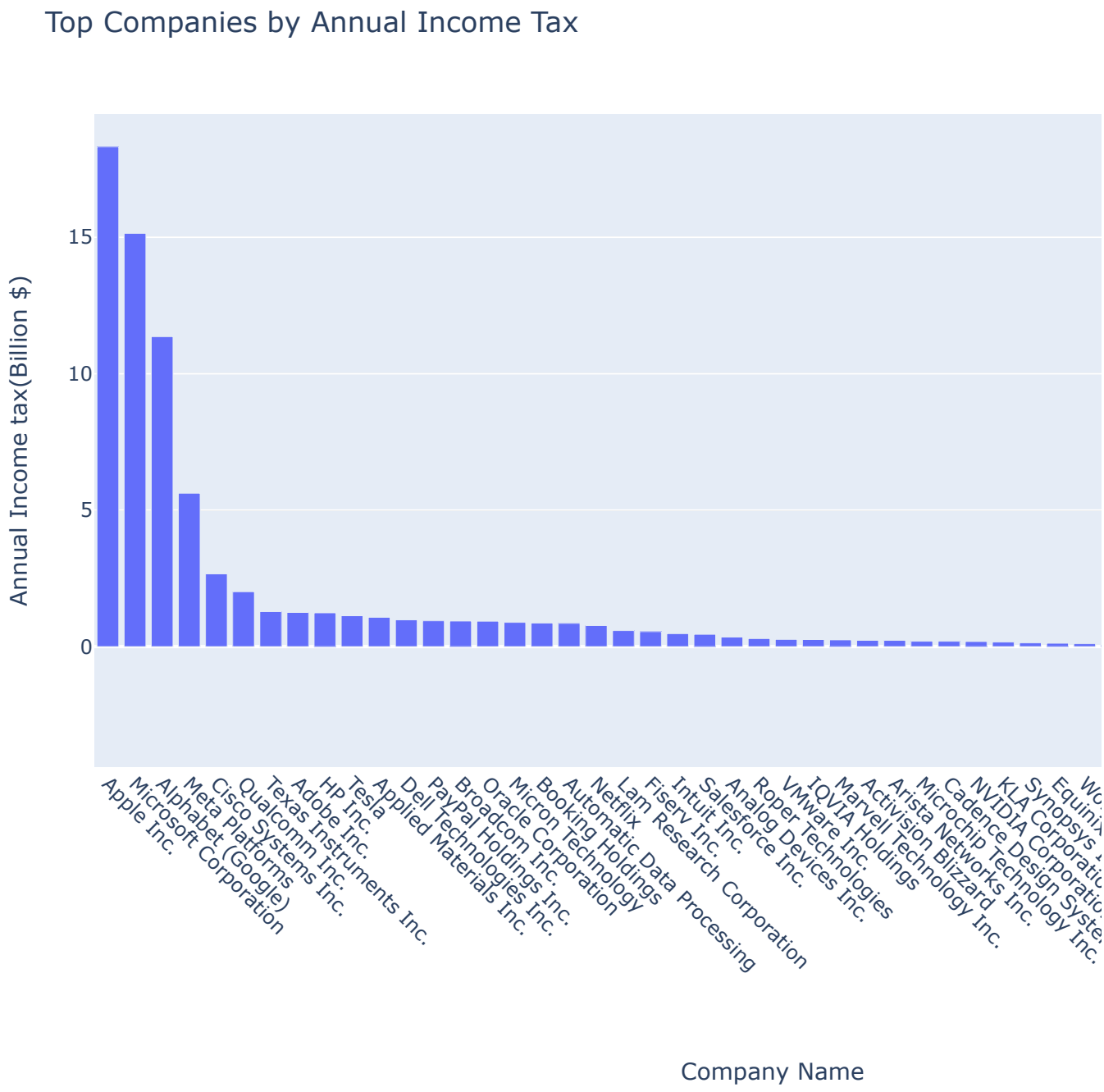
In [31]:
```
# Grouping the data
income_tax =df2.groupby('Company Name')['Annual Income tax(Billion $)'].sum().reset_inde
#income_tax.head(10)

#sort the data in descending order to find the company highest Income tax
income_tax= income_tax.sort_values(by= 'Annual Income tax(Billion $)', ascending= False)
income_tax.head(10)
```

Out[31]:

| | Company Name | Annual Income tax(Billion $) |
|---|---|---|
| **0** | Apple Inc. | 18.314 |
| **1** | Microsoft Corporation | 15.139 |
| **2** | Alphabet (Google) | 11.356 |
| **3** | Meta Platforms | 5.619 |
| **4** | Cisco Systems Inc. | 2.665 |
| **5** | Qualcomm Inc. | 2.012 |
| **6** | Texas Instruments Inc. | 1.283 |
| **7** | Adobe Inc. | 1.252 |
| **8** | HP Inc. | 1.238 |
| **9** | Tesla | 1.132 |

In [32]:
```
# Creating the bar graph to visualize this data
fig = px.bar(income_tax, x = "Company Name", y ="Annual Income tax(Billion $)",
             title= "Top Companies by Annual Income Tax")

# Customize the layout (e.g., axis labels, rotation of x-axis label)
fig.update_xaxes(tickangle=45)

#Set the figure size
fig.update_layout(width =1000, height =700)
```

```
#show the plot
fig.show()
```




# Employee Size

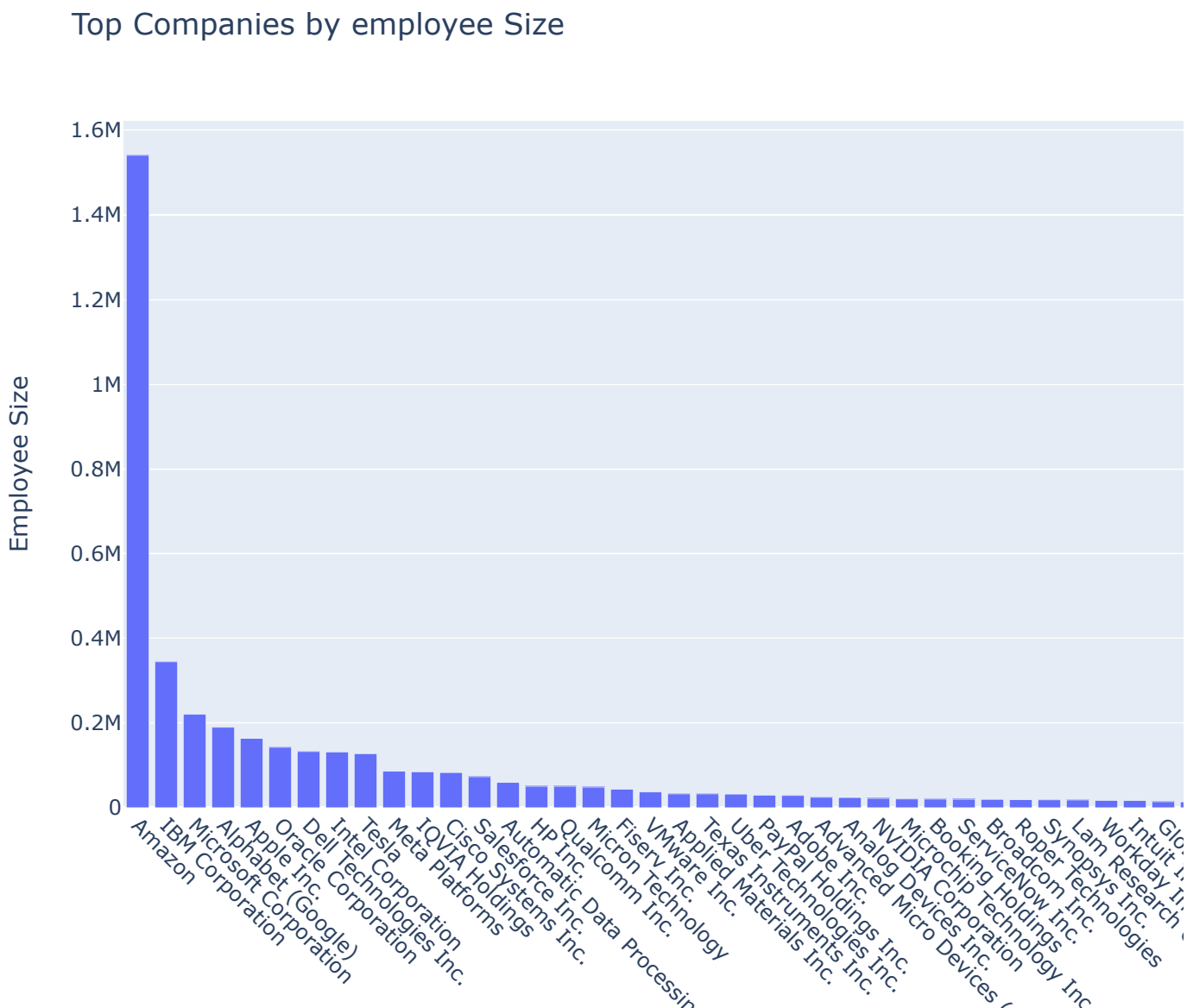
In [33]:
```
#Grouping the data
employee_size =df2.groupby("Company Name")["Employee Size"].sum().reset_index()
#employee_size

# Sorting the data to find Companies highest employee size
employee_size= employee_size.sort_values(by = 'Employee Size', ascending = False).reset_
employee_size.head(10)
```

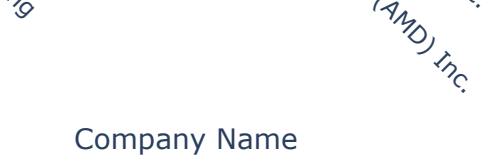
Out[33]:

| | Company Name | Employee Size |
|---|---|---|
| **0** | Amazon | 1541000 |
| **1** | IBM Corporation | 345000 |

| | | |
|---|---|---|
| **2** | Microsoft Corporation | 221000 |
| **3** | Alphabet (Google) | 190234 |
| **4** | Apple Inc. | 164000 |
| **5** | Oracle Corporation | 143000 |
| **6** | Dell Technologies Inc. | 133000 |
| **7** | Intel Corporation | 131900 |
| **8** | Tesla | 127855 |
| **9** | Meta Platforms | 86482 |

In [34]:
```
# Creating a bar graph visulize this data
fig = px.bar(employee_size, x = 'Company Name', y = 'Employee Size', title= 'Top Compani

# Customize the layout(e.g, axis labels , rotation of x-axis labels)
fig.update_xaxes(tickangle =45)

# Set the figure size
fig.update_layout(width =1000, height =700)

# show the plot
fig.show()
```






# Geographical Analysis

In [35]:
```
df2.head()
```

Out[35]:

| | Company Name | Industry | Sector | HQ State | Founding Year | Annual Revenue(Billion $) | Market Cap(Trillions $) | Stock Name | Annu Incom tax(Billic |
|---|---|---|---|---|---|---|---|---|---|
| **0** | Apple Inc. | Technology | Consumer Electronics | California | 1905 | 387.53 | 2.520 | AAPL | 18.3 |
| **1** | Microsoft Corporation | Technology | Software Infrastructure | Washington | 1905 | 204.09 | 2.037 | MSFT | 15.13 |
| **2** | Alphabet (Google) | Technology | Software Infrastructure | California | 1905 | 282.83 | 1.350 | GOOG | 11.35 |
| **3** | Amazon | Technology | Software Application | Washington | 1905 | 513.98 | 1.030 | AMZN | -3.2 |
| **4** | NVIDIA Corporation | Technology | Semiconductors | California | 1905 | 26.97 | 0.653 | NVDA | 0.18 |

In [36]:
```
# Number of unique state
df2['HQ State'].nunique()
```

Out[36]:
```
13
```

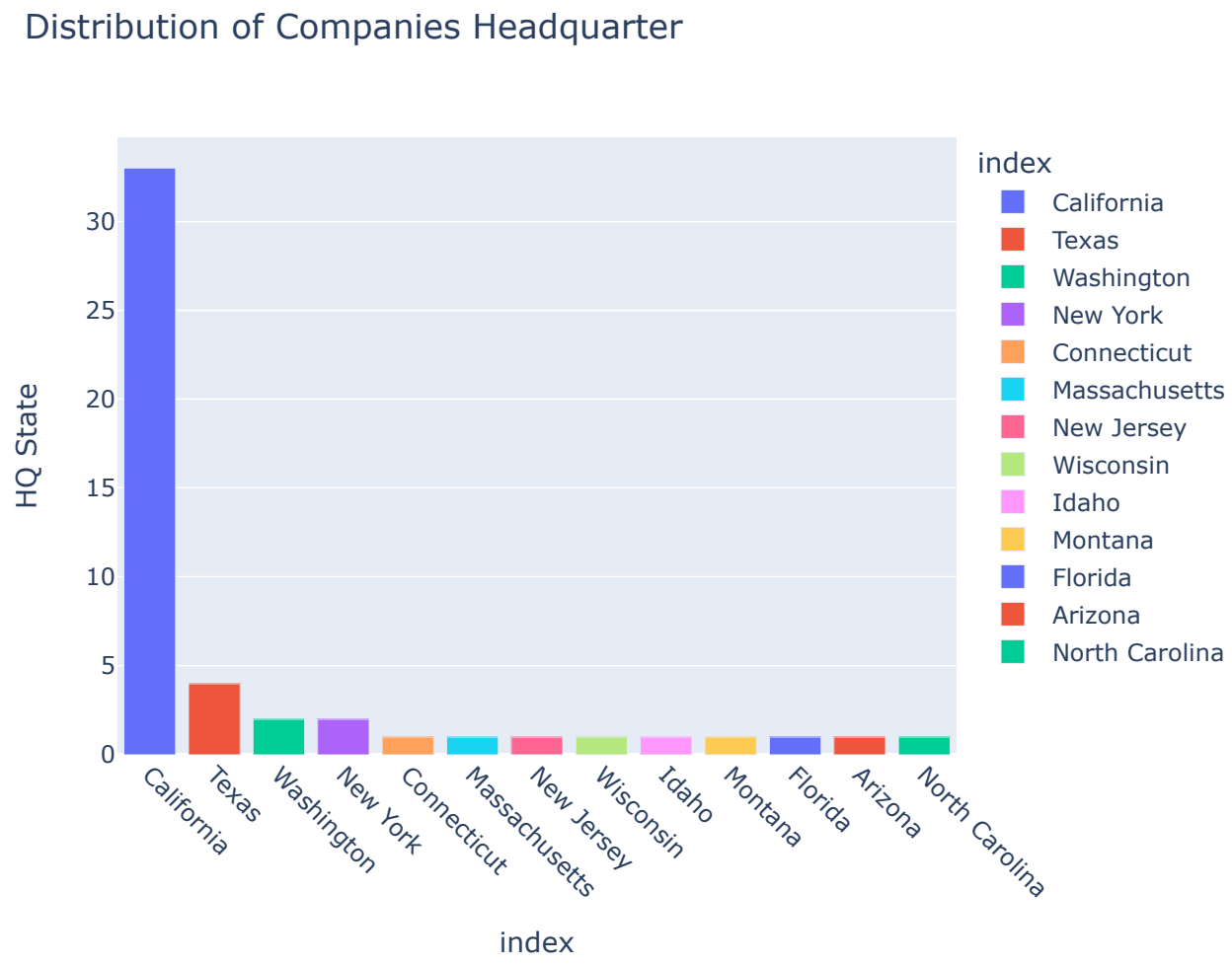
In [37]:
```
#Count the number of companies in each state
hq_state = df2['HQ State'].value_counts().reset_index()
hq_state
```

Out[37]:

| | index | HQ State |
|---|---|---|
| **0** | California | 33 |
| **1** | Texas | 4 |
| **2** | Washington | 2 |
| **3** | New York | 2 |
| **4** | Connecticut | 1 |
| **5** | Massachusetts | 1 |
| **6** | New Jersey | 1 |
| **7** | Wisconsin | 1 |
| **8** | Idaho | 1 |
| **9** | Montana | 1 |
| **10** | Florida | 1 |
| **11** | Arizona | 1 |

| | | |
|---|---|---|
| **12** | North Carolina | 1 |

In [38]:
```
# Create a bar graph using plotly
fig = px.bar(hq_state, x= 'index', y= 'HQ State', title = 'Distribution of Companies Hea

# Customize the layout (e.g axis labels, rotation of x-axis labels)
fig.update_xaxes(tickangle =45)

#show the plot
fig.show()
```




# Sector Analysis

In [39]:
```
df2
```

Out[39]:

| | Company Name | Industry | Sector | HQ State | Founding Year | Annual Revenue(Billion $) | Market Cap(Trillions $) | Stock Name | t |
|---|---|---|---|---|---|---|---|---|---|
| **0** | Apple Inc. | Technology | Consumer Electronics | California | 1905 | 387.53 | 2.520 | AAPL | |
| **1** | Microsoft Corporation | Technology | Software Infrastructure | Washington | 1905 | 204.09 | 2.037 | MSFT | |
| **2** | Alphabet (Google) | Technology | Software Infrastructure | California | 1905 | 282.83 | 1.350 | GOOG | |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| **3** | Amazon | Technology | Software Application | Washington | 1905 | 513.98 | 1.030 | AMZN |
| **4** | NVIDIA Corporation | Technology | Semiconductors | California | 1905 | 26.97 | 0.653 | NVDA |
| **5** | Tesla | Technology | Software Infrastructure | Texas | 1905 | 81.46 | 0.625 | TSLA |
| **6** | Meta Platforms | Technology | Software Infrastructure | California | 1905 | 116.60 | 0.524 | META |
| **7** | Broadcom Inc. | Technology | Semiconductors | California | 1905 | 34.41 | 0.266 | AVGO |
| **8** | Oracle Corporation | Technology | Software Infrastructure | Texas | 1905 | 46.07 | 0.236 | ORCL |
| **9** | Cisco Systems Inc. | Technology | Communication Equipments | California | 1905 | 53.16 | 0.208 | CSCO |
| **10** | Salesforce Inc. | Technology | Software Application | California | 1905 | 31.35 | 0.189 | CRM |
| **11** | Adobe Inc. | Technology | Software Infrastructure | California | 1905 | 17.60 | 0.172 | ADBE |
| **12** | Texas Instruments Inc. | Technology | Semiconductors | Texas | 1905 | 20.02 | 0.162 | TXN |
| **13** | Advanced Micro Devices (AMD) Inc. | Technology | Semiconductors | California | 1905 | 23.60 | 0.155 | AMD |
| **14** | Qualcomm Inc. | Technology | Semiconductors | California | 1905 | 42.95 | 0.138 | QCOM |
| **15** | Netflix | Technology | Software Application | California | 1905 | 31.61 | 0.136 | NFLX |
| **16** | Intel Corporation | Technology | Semiconductors | California | 1905 | 63.05 | 0.118 | INTC |
| **17** | Intuit Inc. | Technology | Software Application | California | 1905 | 13.68 | 0.118 | INTU |
| **18** | IBM Corporation | Technology | IT Services | New York | 1905 | 60.52 | 0.113 | IBM |
| **19** | Applied Materials Inc. | Technology | Semiconductors | California | 1905 | 26.25 | 0.102 | AMAT |
| **20** | Booking Holdings | Technology | Software Application | Connecticut | 1905 | 17.09 | 0.097 | BKNG |
| **21** | Analog Devices Inc. | Technology | Semiconductors | Massachusetts | 1905 | 12.57 | 0.095 | ADI |
| **22** | ServiceNow Inc. | Technology | Software Application | California | 1905 | 7.24 | 0.090 | NOW |
| **23** | Automatic Data Processing | Technology | Software Application | New Jersey | 1905 | 16.67 | 0.090 | ADP |
| **24** | PayPal Holdings Inc. | Technology | Software Infrastructure | California | 1905 | 27.51 | 0.087 | PYPL |
| **25** | Airbnb | Technology | Software Application | California | 1905 | 8.39 | 0.078 | ABNB |
| **26** | Fiserv Inc. | Technology | IT Services | Wisconsin | 1905 | 17.73 | 0.071 | FISV |
| **27** | Lam Research | Technology | Semiconductors | California | 1905 | 19.04 | 0.069 | LRCX |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | Corporation | | | | | | | |
| **28** | Uber Technologies Inc. | Technology | Software Application | California | 1905 | 31.87 | 0.066 | UBER |
| **29** | Micron Technology | Technology | Semiconductors | Idaho | 1905 | 27.15 | 0.064 | MU |
| **30** | Equinix | Technology | IT Services | California | 1905 | 7.26 | 0.064 | EQIX |
| **31** | Activision Blizzard | Technology | Software Application | California | 1905 | 7.52 | 0.063 | ATVI |
| **32** | Palo Alto Networks Inc. | Technology | Software Infrastructure | California | 1905 | 6.15 | 0.059 | PANW |
| **33** | Synopsys Inc. | Technology | Software Infrastructure | California | 1905 | 5.17 | 0.057 | SNPS |
| **34** | Cadence Design Systems Inc. | Technology | Software Application | California | 1905 | 3.56 | 0.057 | CDNS |
| **35** | KLA Corporation | Technology | Semiconductors | California | 1905 | 10.48 | 0.053 | KLAC |
| **36** | Arista Networks Inc. | Technology | Computer Hardware | California | 1905 | 4.38 | 0.052 | ANET |
| **37** | VMware Inc. | Technology | Software Infrastructure | California | 1905 | 13.34 | 0.051 | VMW |
| **38** | Workday Inc. | Technology | Software Application | California | 1905 | 6.21 | 0.049 | WDAY |
| **39** | Fortinet Inc. | Technology | Software Infrastructure | California | 1905 | 4.41 | 0.049 | FTNT |
| **40** | Block Inc. | Technology | Software Infrastructure | California | 1905 | 17.53 | 0.047 | SQ |
| **41** | Snowflake Inc. | Technology | Software Application | Montana | 1905 | 2.06 | 0.046 | SNOW |
| **42** | Roper Technologies | Technology | Electronic Components | Florida | 1905 | 5.61 | 0.046 | ROP |
| **43** | Microchip Technology Inc. | Technology | Semiconductors | Arizona | 1905 | 8.05 | 0.045 | MCHP |
| **44** | Autodesk Inc. | Technology | Software Application | California | 1905 | 5.00 | 0.045 | ADSK |
| **45** | GlobalFoundries | Technology | Semiconductors | New York | 1905 | 8.10 | 0.038 | GFS |
| **46** | IQVIA Holdings | Technology | Software Application | North Carolina | 1905 | 14.41 | 0.037 | IQV |
| **47** | Marvell Technology Inc. | Technology | Semiconductors | California | 1905 | 5.91 | 0.035 | MRVL |
| **48** | Dell Technologies Inc. | Technology | Computer Hardware | Texas | 1905 | 102.30 | 0.028 | DELL |
| **49** | HP Inc. | Technology | Computer Hardware | California | 1905 | 59.78 | 0.028 | HPQ |

In [40]:
```
# group the data
sector_distribution = df2['Sector'].value_counts().reset_index()
```

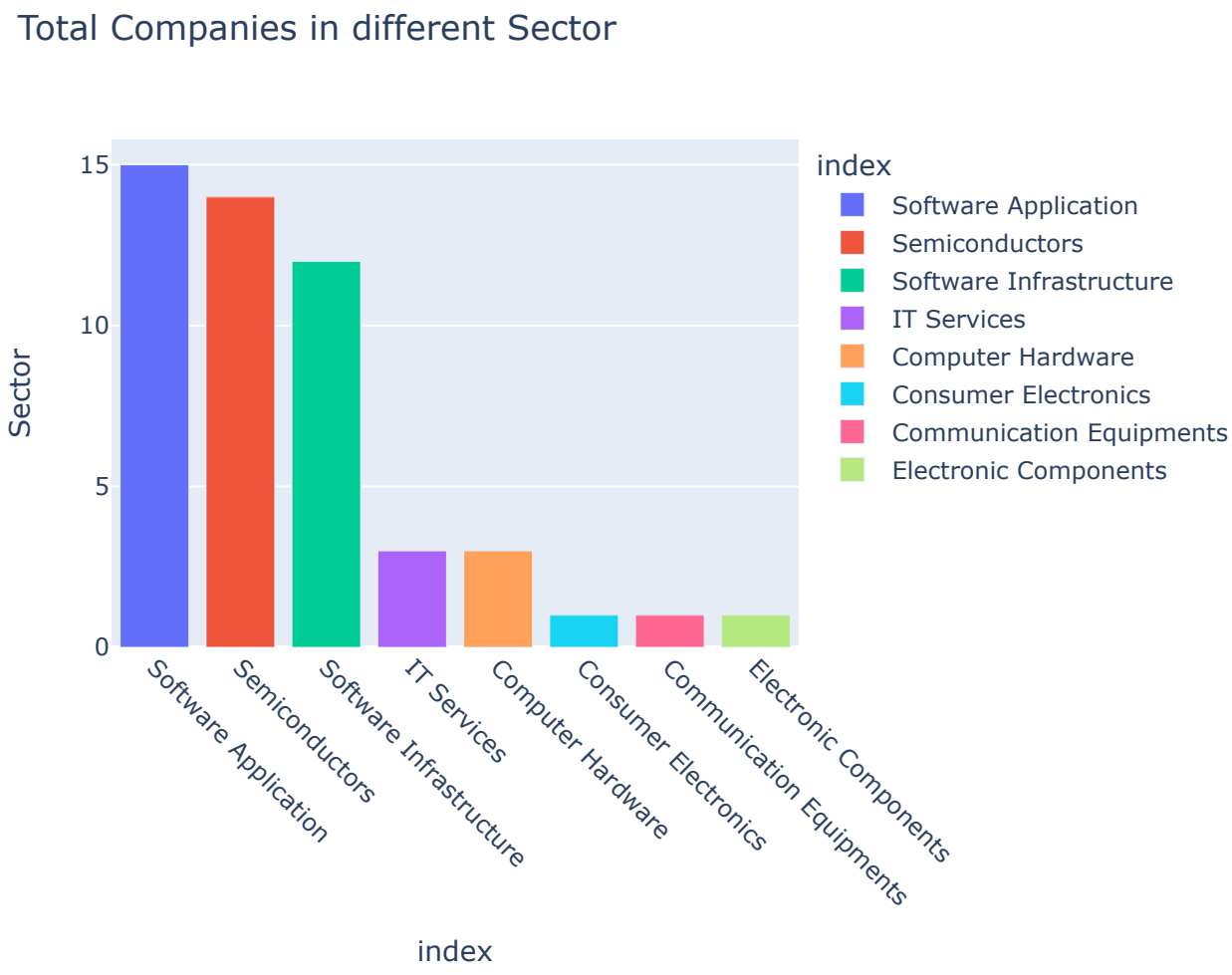
```
sector_distribution
```

Out[40]:

| | index | Sector |
|---|---|---|
| **0** | Software Application | 15 |
| **1** | Semiconductors | 14 |
| **2** | Software Infrastructure | 12 |
| **3** | IT Services | 3 |
| **4** | Computer Hardware | 3 |
| **5** | Consumer Electronics | 1 |
| **6** | Communication Equipments | 1 |
| **7** | Electronic Components | 1 |

In [41]:

```
# Create a bar graph using plotly
fig = px.bar(sector_distribution, x= 'index', y= 'Sector', title = 'Total Companies in d

# Customize the layout (e.g axis labels, rotation of x-axis labels)
fig.update_xaxes(tickangle =45)

#show the plot
fig.show()
```




# Number of Employees in each HQ State

In [42]:
```
#Grouping the data
employee_size_hq =df2.groupby("HQ State")["Employee Size"].sum().reset_index()
#employee_size

# Sorting the data to find Companies highest employee size
employee_size_hq= employee_size_hq.sort_values(by = 'Employee Size', ascending = False).
employee_size_hq.head(10)
```

Out[42]:

| | HQ State | Employee Size |
|---|---|---|
| **0** | Washington | 1762000 |
| **1** | California | 1275585 |
| **2** | Texas | 436855 |
| **3** | New York | 359600 |
| **4** | North Carolina | 85000 |
| **5** | New Jersey | 60000 |
| **6** | Idaho | 49000 |
| **7** | Wisconsin | 44000 |
| **8** | Massachusetts | 24450 |
| **9** | Arizona | 21000 |

In [43]:
```
# Create a line plot using plotly
fig = px.line(employee_size_hq, x ='HQ State', y = 'Employee Size', title = 'Number of e

# Customize a plot layout , labels, and stylig if needed
fig.update_xaxes(title_text= 'HQ State')
fig.update_yaxes(title_text= 'Number of employees')

# Show the plot
fig.show()
```


# Sectoral Analysis using different Metrics

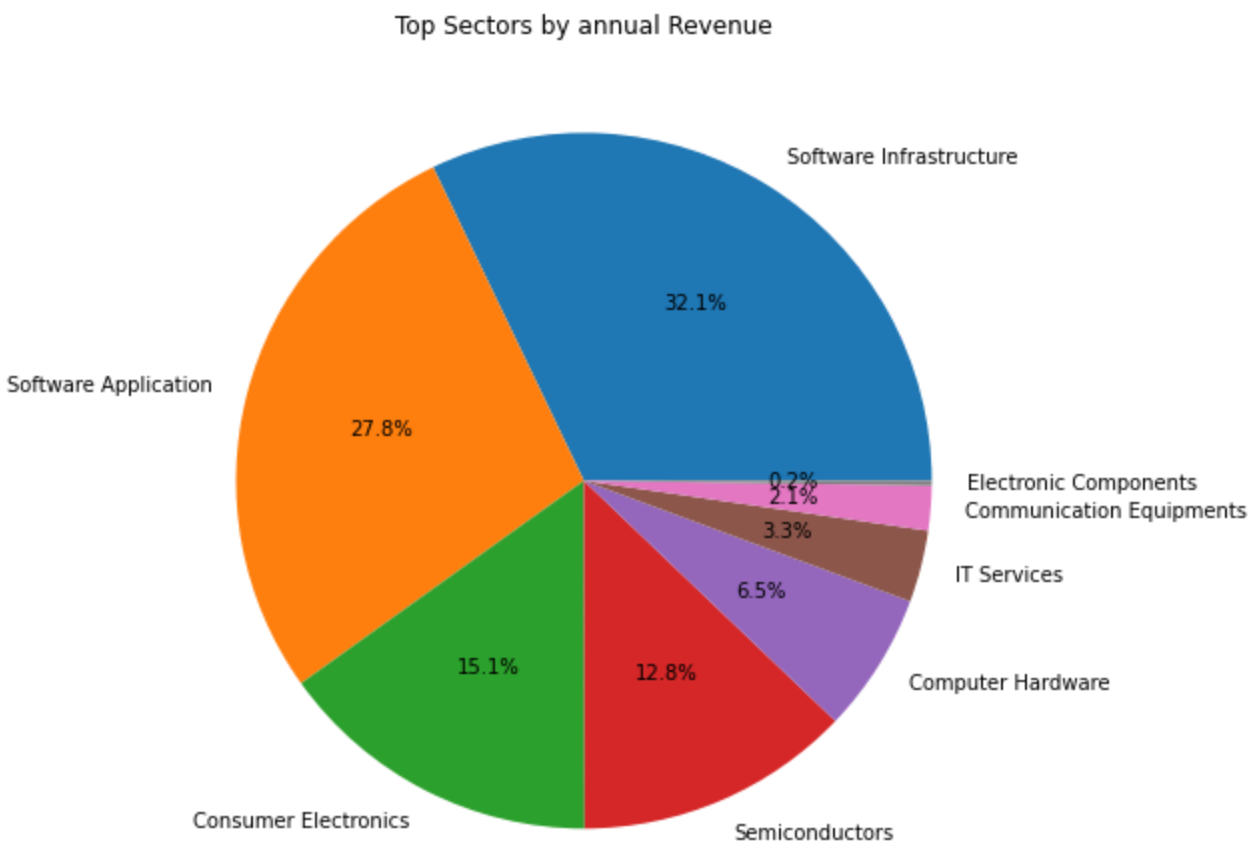
# Sector Vs Annual Revenue

In [44]:
```
#Group the data
sector_revenue =df2.groupby('Sector')['Annual Revenue(Billion $)'].sum().reset_index()

# Sort the result in descending order
sector_revenue= sector_revenue.sort_values(by= 'Annual Revenue(Billion $)', ascending =

#Print the Top sectors company with the hghest Annual revenue
sector_revenue
```

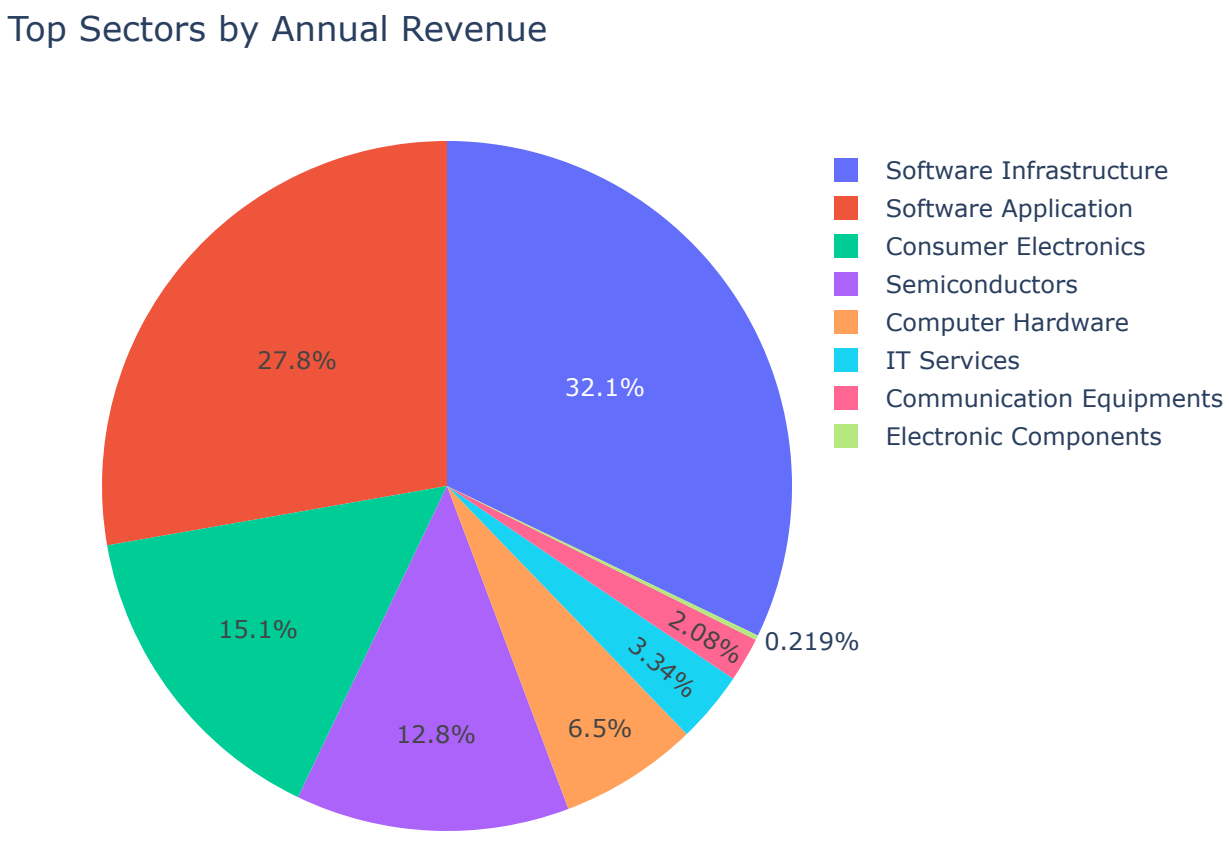
Out[44]:

| | Sector | Annual Revenue(Billion $) |
|---|---|---|
| **0** | Software Infrastructure | 822.76 |
| **1** | Software Application | 710.64 |
| **2** | Consumer Electronics | 387.53 |
| **3** | Semiconductors | 328.55 |
| **4** | Computer Hardware | 166.46 |
| **5** | IT Services | 85.51 |
| **6** | Communication Equipments | 53.16 |
| **7** | Electronic Components | 5.61 |

In [45]:
```
# Create a pie chart using Matplotlib
plt.figure(figsize=(8,8))
plt.pie(sector_revenue['Annual Revenue(Billion $)'], labels = sector_revenue['Sector'],
plt.title("Top Sectors by annual Revenue")
plt.show()
```




In [46]:
```
# Create a pie chart using plotly
fig = px.pie(sector_revenue, values ='Annual Revenue(Billion $)', names= 'Sector',
             title = 'Top Sectors by Annual Revenue')

#show the pie chart
fig.show()
```



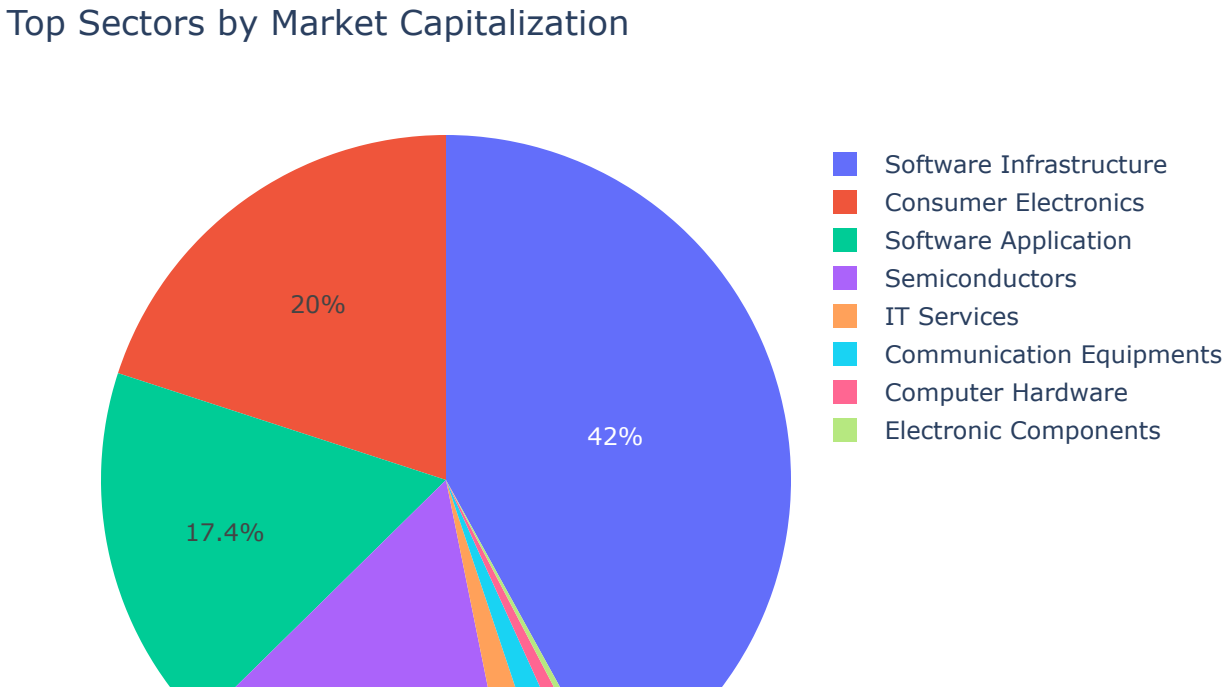
# Sector vs Market Cap

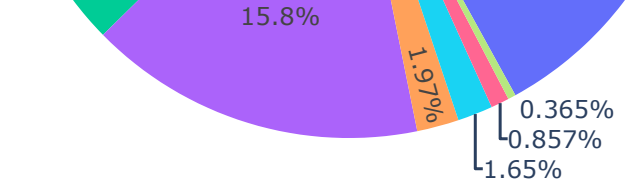
In [47]:
```
#Group the data
sector_marketcap =df2.groupby('Sector')['Market Cap(Trillions $)'].sum().reset_index()

# Sort the result in descending order
sector_marketcap= sector_marketcap.sort_values(by= 'Market Cap(Trillions $)', ascending
#Print the Top sectors company with the hghest Market Cap
sector_marketcap
```

Out[47]:

| | Sector | Market Cap(Trillions $) |
|---|---|---|
| **0** | Software Infrastructure | 5.294 |
| **1** | Consumer Electronics | 2.520 |
| **2** | Software Application | 2.191 |
| **3** | Semiconductors | 1.993 |
| **4** | IT Services | 0.248 |
| **5** | Communication Equipments | 0.208 |
| **6** | Computer Hardware | 0.108 |
| **7** | Electronic Components | 0.046 |

In [48]:
```
# Create a pie chart using plotly
fig = px.pie(sector_marketcap, values ='Market Cap(Trillions $)' , names= 'Sector',
             title = 'Top Sectors by Market Capitalization')

#show the pie chart
fig.show()
```





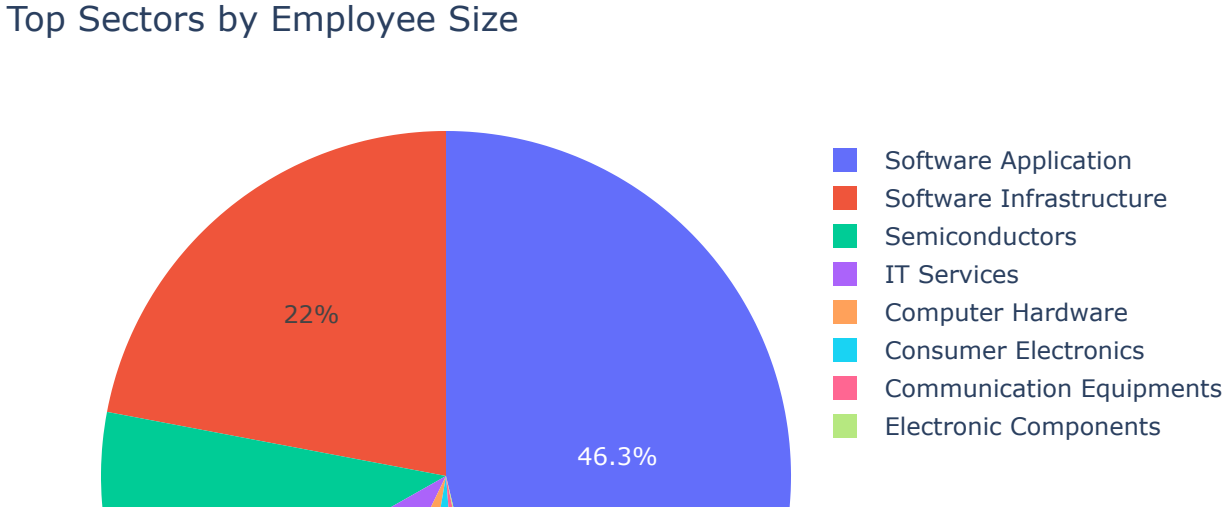

# Sector vs Employee Size

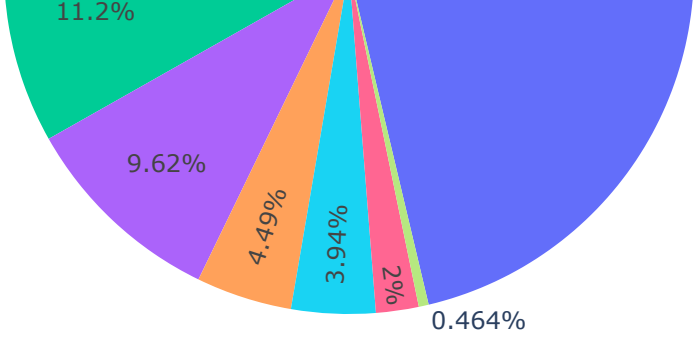
In [49]:
```
#Group the data
sector_employee =df2.groupby('Sector')['Employee Size'].sum().reset_index()

# Sort the result in descending order
sector_employee= sector_employee.sort_values(by= 'Employee Size', ascending = False).res

#Print the Top sectors company with the hghest Market Cap
sector_employee
```

Out[49]:

| | Sector | Employee Size |
|---|---|---|
| **0** | Software Application | 1926819 |
| **1** | Software Infrastructure | 916800 |
| **2** | Semiconductors | 464818 |
| **3** | IT Services | 400451 |
| **4** | Computer Hardware | 186993 |
| **5** | Consumer Electronics | 164000 |
| **6** | Communication Equipments | 83300 |
| **7** | Electronic Components | 19300 |

In [50]:
```
# Create a pie chart using plotly
fig = px.pie(sector_employee, values ='Employee Size' , names= 'Sector',
             title = 'Top Sectors by Employee Size')

#show the pie chart
fig.show()
```

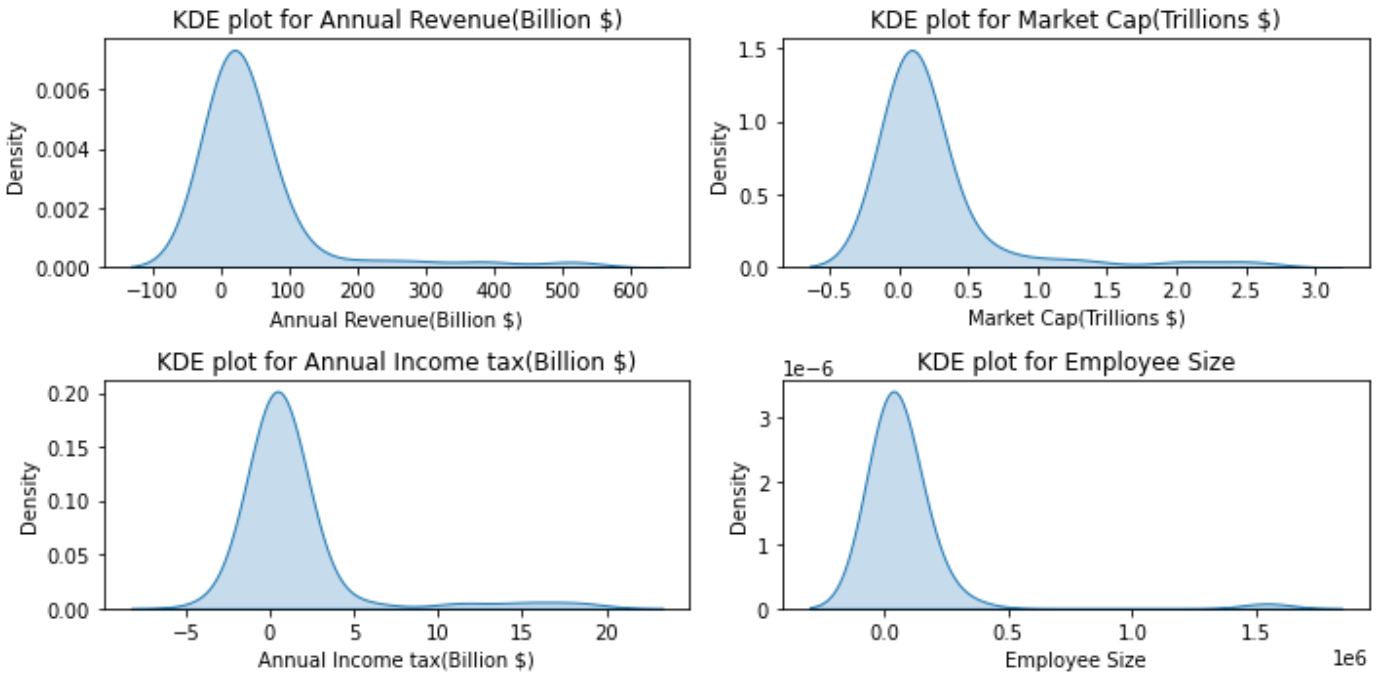





# Correlation Analysis

# Data distribution in numeric columns

In [51]:
```
#List of columns
columns =['Annual Revenue(Billion $)', 'Market Cap(Trillions $)', 'Annual Income tax(Bil

# Creating a subplot for each column
plt.figure(figsize=(10,5))

for i , column in enumerate(columns,1):
    plt.subplot(2,2,i)
    sns.kdeplot(data= df2, x= column, fill =True)
    plt.title(f"KDE plot for {column}")

plt.tight_layout()
plt.show()
```




In [ ]: