

```
In [120.]: # Import Libraries
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt

In [126.]: # Read the Data and Showed 10 rows
companies = pd.read_csv(r'C:\Users\HP\Desktop\Advance Data Analyst\3. Go beyond Numbers. turning Data into insight\2. Module 2\1. Discovering\lab\Unicorn_Companies.csv')
companies.head(10)

Out[126.]:
   Company  Valuation  Date Joined  Industry  City  Country/Region  Continent  Year Founded  Funding  Select Investors
0  Bytedance  $100B  4/7/17  Artificial intelligence  Beijing  China  Asia  2012  $8B  Sequoia Capital China, SIG Asia Investments, S...
1  Spacex  $100B  12/2/12  Other  Hawthorne  United States  North America  2002  $7B  Founders Fund, Draper Fisher Jurvetson, Rother...
2  SHEIN  $100B  7/3/18  E-commerce & direct-to-consumer  Shenzhen  China  Asia  2008  $2B  Tiger Global Management, Sequoia Capital China...
3  Stripe  $95B  1/23/14  Fintech  San Francisco  United States  North America  2010  $2B  Khosla Ventures, LowercaseCapital, capitalG...
4  Klarna  $46B  12/12/11  Fintech  Stockholm  Sweden  Europe  2005  $4B  Institutional Venture Partners, Sequoia Capita...
5  Canva  $40B  1/6/18  Internet software & services  Surry Hills  Australia  Oceania  2012  $572M  Sequoia Capital China, Blackbird Ventures, Mat...
6  Check.com  $40B  5/2/19  Fintech  London  United Kingdom  Europe  2012  $2B  Tiger Global Management, Insight Partners, DST...
7  Instacart  $39B  12/30/14  Supply chain, logistics, & delivery  San Francisco  United States  North America  2012  $3B  Khosla Ventures, Kleiner Perkins Caufield & By...
8  JUUL Labs  $38B  12/20/17  Consumer & retail  San Francisco  United States  North America  2015  $14B  Tiger Global Management
9  Databricks  $38B  2/5/19  Data management & analytics  San Francisco  United States  North America  2013  $3B  Andreessen Horowitz, New Enterprise Associates...

In [128.]: # How large the dataset is
companies.size

Out[128.]: 10740

In [130.]: # Shape of the dataset
companies.shape

Out[130.]: (1074, 10)

In [132.]: # Get information
companies.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/Region  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 [134.]: # Get descriptive statistics
companies.describe()

Out[134.]:
   Year Founded
count  1074.000000
mean   2012.896717
std     5.686673
min    1919.000000
25%    2011.000000
50%    2014.000000
75%    2016.000000
max     2021.000000

In [138.]: # Step 1: Use pd.to_datetime() to convert Date Joined column to datetime
# Step 2: Update the column with the converted values
companies["Date Joined"] = pd.to_datetime(companies["Date Joined"])
companies.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     datetime64[ns]
3  Industry  1074 non-null     object
4  City  1058 non-null     object
5  Country/Region  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: datetime64[ns](1), int64(1), object(8)
memory usage: 84.0+ KB

In [140.]: # Step 3: Use .dt.year to extract year component from Date Joined column
# Step 2: Add the result as a new column named Year Joined to the DataFrame
companies["Year Joined"] = companies["Date Joined"].dt.year
companies.head(10)

Out[140.]:
   Company  Valuation  Date Joined  Industry  City  Country/Region  Continent  Year Founded  Funding  Select Investors  Year Joined
0  Bytedance  $100B  2017-04-07  Artificial intelligence  Beijing  China  Asia  2012  $8B  Sequoia Capital China, SIG Asia Investments, S...  2017
1  Spacex  $100B  2012-12-01  Other  Hawthorne  United States  North America  2002  $7B  Founders Fund, Draper Fisher Jurvetson, Rother...  2012
2  SHEIN  $100B  2018-07-03  E-commerce & direct-to-consumer  Shenzhen  China  Asia  2008  $2B  Tiger Global Management, Sequoia Capital China...  2018
3  Stripe  $95B  2014-01-23  Fintech  San Francisco  United States  North America  2010  $2B  Khosla Ventures, LowercaseCapital, capitalG...  2014
4  Klarna  $46B  2011-12-12  Fintech  Stockholm  Sweden  Europe  2005  $4B  Institutional Venture Partners, Sequoia Capita...  2011
5  Canva  $40B  2018-01-08  Internet software & services  Surry Hills  Australia  Oceania  2012  $572M  Sequoia Capital China, Blackbird Ventures, Mat...  2018
6  Check.com  $40B  2019-05-02  Fintech  London  United Kingdom  Europe  2012  $2B  Tiger Global Management, Insight Partners, DST...  2019
7  Instacart  $39B  2014-12-30  Supply chain, logistics, & delivery  San Francisco  United States  North America  2012  $3B  Khosla Ventures, Kleiner Perkins Caufield & By...  2014
8  JUUL Labs  $38B  2017-12-20  Consumer & retail  San Francisco  United States  North America  2015  $14B  Tiger Global Management  2017
9  Databricks  $38B  2019-02-05  Data management & analytics  San Francisco  United States  North America  2013  $3B  Andreessen Horowitz, New Enterprise Associates...  2019

In [144.]: # Sample the data
companies_sample = companies.sample(n = 50, random_state = 42)
companies_sample

Out[144.]:
   Company  Valuation  Date Joined  Industry  City  Country/Region  Continent  Year Founded  Funding  Select Investors  Year Joined
542  Aiven  $2B  2021-10-18  Internet software & services  Helsinki  Finland  Europe  2016  $210M  Institutional Venture Partners, Atomico, Early...  2021
370  Jusfun Big Data  $2B  2018-07-09  Data management & analytics  Beijing  China  Asia  2010  $137M  Boxin Capital, DT Capital Partners, IDG Capital  2018
307  Innovacore  $3B  2021-02-19  Health  San Francisco  United States  North America  2014  $379M  M12, WestBridge Capital, Lightspeed Venture Pa...  2021
493  Algolia  $2B  2021-07-28  Internet software & services  San Francisco  United States  North America  2012  $334M  Accel, Aiven Capital, Storm Ventures  2021
350  SouChe Holdings  $3B  2017-11-01  E-commerce & direct-to-consumer  Hangzhou  China  Asia  2012  $1B  Morningside Ventures, Warburg Pincus, CreditEa...  2017
237  Wellflow  $4B  2021-01-13  Internet software & services  San Francisco  United States  North America  2010  $335M  Accel, SilverSmith Capital Partners, capitalG  2021
475  Gousto  $2B  2021-11-03  Supply chain, logistics, & delivery  London  United Kingdom  Europe  2012  $311M  MMC Ventures, BGF Ventures, Unilever Ventures  2020
578  Midfys Market  $2B  2021-04-21  E-commerce & direct-to-consumer  Pennsauken  United States  North America  2018  $527M  Accel, D1 Capital Partners, Greenoaks Capital M...  2021
462  Hibob  $2B  2021-10-12  Internet software & services  Tel Aviv  Israel  Asia  2015  $274M  Bessemer Venture Partners, Eight Roads Venture...  2021
978  News Break  $1B  2021-01-07  Internet software & services  Mountain View  United States  North America  2015  $151M  IDG Capital, Francisco Partners, ZhenFund  2021
545  AppFlyer  $2B  2020-01-21  Mobile & telecommunications  San Francisco  United States  North America  2011  $293M  Magna Venture Partners, Plango Venture Capita...  2020
713  VideoAck  $1B  2021-10-21  Internet software & services  Los Angeles  United States  North America  2014  $382M  Simon Equity Partners, Wavemaker Partners, Ant...  2021
141  Checkr  $5B  2019-09-19  Internet software & services  San Francisco  United States  North America  2014  $559M  Y Combinator, Accel, T. Rowe Price  2019
275  Axonius  $3B  2021-03-01  Cybersecurity  New York  United States  North America  2017  $395M  Vertex Ventures Israel, Bessemer Venture Part...  2021
428  Zenoti  $2B  2020-12-15  Internet software & services  Bellevue  United States  North America  2010  $331M  Norwest Venture Partners, Accel, Tiger Global ...  2020
453  Tonal  $2B  2021-03-31  E-commerce & direct-to-consumer  San Francisco  United States  North America  2015  $450M  Mayfield Fund, Shasta Ventures, L. Calleton  2021
808  Acko General Insurance  $1B  2021-10-28  Fintech  Bengaluru  India  Asia  2016  $458M  Intact Ventures, Munich Re Ventures, General A...  2021
361  CFGT  $2B  2021-09-15  Fintech  Boston  United States  North America  2000  $19M  The Carlyle Group, CVC Capital Partners  2021
1073  Zwill  $1B  2020-09-16  E-commerce & direct-to-consumer  Long Beach  United States  North America  2014  $620M  Novator Partners, True, Causeway Media Partners  2020
737  Island  $1B  2022-03-23  Internet software & services  Dallas  United States  North America  2020  $215M  Insight Partners, Sequoia Capital, Stripes Group  2022
327  BookFi  $3B  2021-03-11  Fintech  Jersey City  United States  North America  2017  $1B  ConsenSys Ventures, Valar Ventures, PUC  2021
842  Rubicon  $1B  2017-08-25  Other  Atlanta  United States  North America  2008  $223M  Goldman Sachs, Leonardo DiCaprio, Prometheus  2017
718  Alloy  $1B  2021-09-30  Fintech  New York  United States  North America  2015  $156M  Bessemer Venture Partners, Eniac Ventures, Can...  2021
192  Applied Inuition  $4B  2020-10-22  Data management & analytics  Sunnyvale  United States  North America  2017  $352M  Andreessen Horowitz, Lux Capital, General Cata...  2020
657  Hello TransTech  $1B  2018-06-01  Auto & transportation  Shanghai  China  Asia  2016  $2B  Ant Financial Services Group, GGV Capital  2018
549  Black Sesame Technologies  $2B  2021-09-22  Artificial intelligence  Shanghai  China  Asia  2016  $115M  Northern Light Venture Capital, Xiaomi, Future...  2021
626  Keenon Robotics  $1B  2021-09-15  Supply chain, logistics, & delivery  Shanghai  China  Asia  2010  $229M  Yung Partners, SoftBank Group, Ivision Ventures  2021
215  SSENSE  $4B  2021-06-08  E-commerce & direct-to-consumer  Montreal  Canada  North America  2003  Unknown  Sequoia Capital  2021
912  Fabric  $1B  2021-10-26  Supply chain, logistics, & delivery  New York  United States  North America  2015  $294M  Innovation Endeavors, Alaph, Temasek  2021
685  Ouyee  $1B  2019-06-27  Internet software & services  Shanghai  China  Asia  2015  $432M  Taigang Venture Capital  2019
199  Papaya Global  $4B  2021-03-04  Internet software & services  New York  United States  North America  2016  $440M  Bessemer Venture Partners, Insight Partners, N...  2021
290  Nexiva  $3B  2021-09-14  Internet software & services  Scottsdale  United States  North America  2008  $200M  Goldman Sachs Asset Management  2021
107  Deel  $6B  2021-04-21  Fintech  San Francisco  United States  North America  2018  $629M  Andreessen Horowitz, Spark Capital, Y Combinator  2021
670  DT Dream  $1B  2017-06-08  Data management & analytics  Hangzhou  China  Asia  2015  $267M  Alibaba Group, China Everbright Investment Man...  2017
1008  Razor  $1B  2021-11-08  Consumer & retail  Berlin  Germany  Europe  2020  $167M  Global Founders Capital, 468 Capital, RedaPhn...  2021
243  OpenAI  $3B  2019-07-22  Artificial intelligence  San Francisco  United States  North America  2015  $1B  Khosla Ventures  2019
570  Kaseya  $2B  2019-03-27  Cybersecurity  Miami  United States  North America  2000  $545M  Insight Partners, TPG Alternative & Renewable ...  2019
88  Attentive  $7B  2020-09-23  Mobile & telecommunications  Hoboken  United States  North America  2016  $863M  NextView Ventures, Eniac Ventures, Sequoia Cap...  2020
51  Alchemy  $10B  2021-10-28  Fintech  San Francisco  United States  North America  2017  $564M  DFJ Growth Fund, Coast Capital Management, Additi...  2021
736  InSightec  $1B  2020-03-06  Health  Tirat Carmel  Israel  Asia  1999  $633M  York Capital Management, GE Healthcare, Koch D...  2020
381  MindMaze  $2B  2016-01-28  Health  Lausanne  Switzerland  Europe  2012  $340M  Hinduja Group  2016
308  Intelize  $3B  2021-02-18  Internet software & services  Sacramento  United States  North America  2014  $334M  Accel, Sequoia Capital, Y Combinator  2022
746  Yugabyte  $1B  2021-10-28  Data management & analytics  Sunnyvale  United States  North America  2016  $291M  Lightspeed Venture Partners, Dell Technologies...  2021
318  HighRadius  $3B  2020-01-07  Fintech  Houston  United States  North America  2006  $475M  Susquehanna Growth Equity, Citi Ventures, ICON...  2020
31  Faire  $12B  2019-10-30  Artificial intelligence  San Francisco  United States  North America  2017  $1B  Khosla Ventures, Forrester Ventures, Sequoia ...  2019
424  Uplight  $2B  2021-03-03  Other  Boulder  United States  North America  2019  Unknown  Rubicon Technology Partners, Max Ventures, Inc...  2021
867  Assent  $1B  2022-01-06  Supply chain, logistics, & delivery  Ottawa  Canada  North America  2010  $534M  Vista Equity Partners, Warburg Pincus, First A...  2022
1009  Reid Ventures  $1B  2015-01-07  Other  Fort Mill  United States  North America  2000  $250M  Silver Lake Partners, General Atlantic  2015
359  Cambridge Mobile Telematics  $2B  2018-12-19  Mobile & telecommunications  Cambridge  United States  North America  2010  $503M  SoftBank Group  2018
342  Lattice  $3B  2021-03-23  Internet software & services  San Francisco  United States  North America  2015  $329M  Khosla Ventures, Thrive Capital, Y Combinator  2021

In [116.]: # Prepare data for plotting
# Create new 'years_till_unicorn' column
companies_sample["years_till_unicorn"] = companies_sample["Year Joined"] - companies_sample["Year Founded"]

# Group the data by 'Industry'. For each industry, get the max value in the 'years_till_unicorn' column.
grouped = (companies_sample[["Industry", "years_till_unicorn"]]
          .groupby("Industry")
          .max()
          .sort_values(by="years_till_unicorn")
          )
grouped

Out[116.]:
   Industry
Consumer & retail  1
Auto & transportation  2
Artificial intelligence  5
Data management & analytics  8
Mobile & telecommunications  9
Supply chain, logistics, & delivery  12
Internet software & services  13
Other  15
E-commerce & direct-to-consumer  18
Cybersecurity  19
Fintech  21
Health  21

In [118.]: # Create bar plot
# With Industry column as the categories of the bars
# and the difference in years between Year Joined column and Year Founded column as the heights of the bars
plt.bar(grouped.index, grouped["years_till_unicorn"])

# Set title
plt.title("Bar plot of maximum years taken by company to become unicorn per industry (from sample)")

# Set x-axis label
plt.xlabel("Industry")

# Set y-axis label
plt.ylabel("Maximum number of years")

# Rotate labels on the x-axis as a way to avoid overlap in the positions of the text
plt.xticks(rotation=45, horizontalalignment='right')

# Display the plot
plt.show()

Bar plot of maximum years taken by company to become unicorn per industry (from sample)

```

```
In [150.]: # Create a column representing company valuation as numeric data
# Create new column
companies_sample["valuation_billions"] = companies_sample["Valuation"]
# Remove the '$' from each value
companies_sample["valuation_billions"] = companies_sample["valuation_billions"].str.replace('$', '')
# Remove the 'B' from each value
companies_sample["valuation_billions"] = companies_sample["valuation_billions"].str.replace('B', '')
# Convert column to type int()
companies_sample["valuation_billions"] = companies_sample["valuation_billions"].astype('int')
companies_sample.head()

Out[150.]:
   Company  Valuation  Date Joined  Industry  City  Country/Region  Continent  Year Founded  Funding  Select Investors  Year Joined  years_till_unicorn  valuation_billions
542  Aiven  $2B  2021-10-18  Internet software & services  Helsinki  Finland  Europe  2016  $210M  Institutional Venture Partners, Atomico, Early...  2021  5  2
370  Jusfun Big Data  $2B  2018-07-09  Data management & analytics  Beijing  China  Asia  2010  $137M  Boxin Capital, DT Capital Partners, IDG Capital  2018  8  2
307  Innovacore  $3B  2021-02-19  Health  San Francisco  United States  North America  2014  $379M  M12, WestBridge Capital, Lightspeed Venture Pa...  2021  7  3
493  Algolia  $2B  2021-07-28  Internet software & services  San Francisco  United States  North America  2012  $334M  Accel, Aiven Capital, Storm Ventures  2021  9  2
350  SouChe Holdings  $3B  2017-11-01  E-commerce & direct-to-consumer  Hangzhou  China  Asia  2012  $1B  Morningside Ventures, Warburg Pincus, CreditEa...  2017  5  3

In [152.]: # Prepare data for modeling
grouped = (companies_sample[["Industry", "valuation_billions"]]
          .groupby("Industry")
          .max()
          .sort_values(by="valuation_billions")
          )
grouped

Out[152.]:
   Industry
Auto & transportation  1
Consumer & retail  1
Other  2
Supply chain, logistics, & delivery  2
Cybersecurity  3
Health  3
Data management & analytics  4
E-commerce & direct-to-consumer  4
Internet software & services  5
Mobile & telecommunications  7
Fintech  10
Artificial intelligence  12

In [154.]: # Create bar plot
# With Industry column as the categories of the bars
# and new valuation column as the heights of the bars
plt.bar(grouped.index, grouped["valuation_billions"])

# Set title
plt.title("Bar plot of maximum unicorn company valuation per industry (from sample)")

# Set x-axis label
plt.xlabel("Industry")

# Set y-axis label
plt.ylabel("Maximum valuation in billions of dollars")

# Rotate labels on the x-axis as a way to avoid overlap in the positions of the text
plt.xticks(rotation=45, horizontalalignment='right')

# Display the plot
plt.show()

Bar plot of maximum unicorn company valuation per industry (from sample)

```

```
In [ ]: # Conclusion
# Functions in the pandas library can be used to gather characterization about the data at hand.
# The info() and describe() functions were especially useful for gathering basic information about a dataset and finding descriptive statistics, respectively.
# Functions in the matplotlib.pyplot module can be used to create visualization to further understand specific aspects of the data.
# The bar() function allowed you to create bar plots that helped visualize categorical information about the data.
# You were able to visualize the maximum years to become a unicorn and maximum valuation for each industry represented in the sample taken from the data.
```