# Data Analysis of Unicorn Companies

**CIS 5270 Python Project**

**Navyasree Sriramoju**

**Suman Chauhan**

# Data Analysis of Unicorn Companies

## Introduction

Unicorn companies are a rare breed of startups that have achieved extraordinary success by surpassing the elusive $1 billion valuation mark. These privately held companies, like the mythical creature they are named after, capture the imagination and intrigue of the business world.

Unicorn companies share several key characteristics that set them apart from ordinary startups:

***Rapid Growth:*** Unicorns exhibit exceptional growth rates, often scaling their operations and customer base at an unprecedented pace. They disrupt traditional industries and carve out new markets with innovative products or services.

***Valuation Milestone:*** Crossing the $1 billion valuation threshold is a significant achievement for unicorns. This remarkable valuation reflects investor confidence in the company's potential to disrupt industries and generate substantial returns.

***Disruptive Innovation:*** Unicorns are known for their ability to introduce groundbreaking innovations that challenge established norms and business models. They leverage technology and new approaches to solve existing problems and create value for customers.

***Funding Success:*** Securing substantial funding is a crucial aspect of a unicorn's journey. They attract large investments from venture capitalists, private equity firms, and other institutional investors who recognize their potential for exponential growth.

***Scalable Business Models:*** Unicorns build scalable business models that allow them to expand rapidly while maintaining high-profit margins. They leverage technology, network effects, and economies of scale to drive their growth trajectory.

Unicorn companies represent the pinnacle of entrepreneurial achievement, defying the odds and transforming industries through disruptive innovation and rapid growth. Their ability to reach astronomical valuations reflects their impact on the business landscape and investor confidence in their potential. By inspiring entrepreneurs, creating jobs, disrupting industries, and driving investor confidence, unicorns play a vital role in shaping the entrepreneurial ecosystem and driving economic progress.

The article ***"How Unicorns Grow"*** published in Harvard Business Review in January 2016 explores the growth strategies and patterns of unicorn companies. The author examines the characteristics that enable these startups to achieve extraordinary valuations and sustain rapid growth.

The key findings of the article are as follows:

- ***Market Expansion:*** Unicorns prioritizes market expansion and aims to capture a significant share of their target markets. They often adopt a hyper-growth mindset and pursue aggressive customer acquisition strategies.

- ***Network Effects:*** Unicorns leverage network effects to their advantage. By building platforms that connect users and facilitate interactions, they create value for both customers and suppliers. The network effects help drive growth and increase barriers to entry for potential competitors.
- ***Multi-Sided Platforms:*** Many unicorns operate as multi-sided platforms, connecting different groups of users or stakeholders. They create an ecosystem where one side of the platform attracts the other, leading to exponential growth. Examples include ride-hailing services connecting drivers and passengers or e-commerce platforms connecting buyers and sellers.
- ***Scaling Operations:*** Unicorns prioritize scaling their operations quickly and efficiently. They invest in infrastructure, technology, and talent to support growth while optimizing costs and maintaining high-profit margins.
- ***Continuous Innovation:*** Unicorns are committed to continuous innovation and improving their offerings. They invest heavily in research and development, product enhancements, and user experience to stay ahead of the competition and meet evolving customer needs.
- ***Strategic Partnerships:*** Unicorns form strategic partnerships with other companies to expand their reach and enhance their value proposition. These partnerships can involve distribution agreements, integration with complementary services, or collaborations to access new markets or technologies.
- ***Funding Strategies:*** Unicorns employ various funding strategies to fuel their growth. They often secure substantial venture capital investments, raise multiple funding rounds, and sometimes rely on debt financing or strategic investments from established companies.

The article concludes by highlighting the challenges that unicorns face as they grow, including increased competition, regulatory hurdles, and the need to maintain profitability while pursuing rapid expansion. It emphasizes the importance of strategic decision-making, adaptability, and a focus on sustainable growth for long-term success.

The article titled **"What Is a Unicorn Company?"** published on Fortune's website provides an overview of unicorn companies and explores their significance in the business world. It offers insights into the characteristics, challenges, and impact of these highly valued startups.

The main points covered in the article are as follows:

- ***Definition of a Unicorn:*** The article begins by defining a unicorn company as a privately held startup that achieves a valuation of $1 billion or more. It explains that the term "unicorn" was coined to represent the rarity and exceptional nature of such companies.
- ***Factors Contributing to Unicorn Status:*** The article discusses the factors that contribute to a company attaining unicorn status. It highlights the importance of disruptive

innovation, rapid growth, scalability, and the ability to attract substantial funding as key elements in achieving such high valuations.

- ***Rise of Unicorns:*** The article explores the rise of unicorn companies in recent years, noting that the number of unicorns has increased significantly. It attributes this growth to factors such as technological advancements, supportive investment climate, and the pursuit of disruptive business models.
- ***Challenges Faced by Unicorns:*** The article acknowledges that while being labeled a unicorn is often associated with success, these companies face unique challenges. It highlights issues like sustaining growth, achieving profitability, managing investor expectations, and facing increased scrutiny as they navigate their growth trajectories.
- ***Impact on the Economy:*** The article emphasizes the broader impact of unicorn companies on the economy. It notes that these startups contribute to job creation, drive innovation in their respective industries, and attract talent and investment to the startup ecosystem.
- ***Risks and Concerns:*** The article also addresses concerns related to unicorn companies. It mentions potential risks associated with high valuations, market volatility, and the possibility of some unicorns failing to deliver on their promises.
- ***Future Outlook:*** The article concludes by discussing the future outlook for unicorn companies. It suggests that while the concept of unicorns will continue to captivate the business world, investors and entrepreneurs should approach them with caution, focusing on sustainable growth and long-term value creation.

Overall, the article provides a concise overview of unicorn companies, highlighting their defining characteristics, challenges, and impact. It aims to provide readers with a broader understanding of the unicorn phenomenon and its implications for the business and investment landscape.

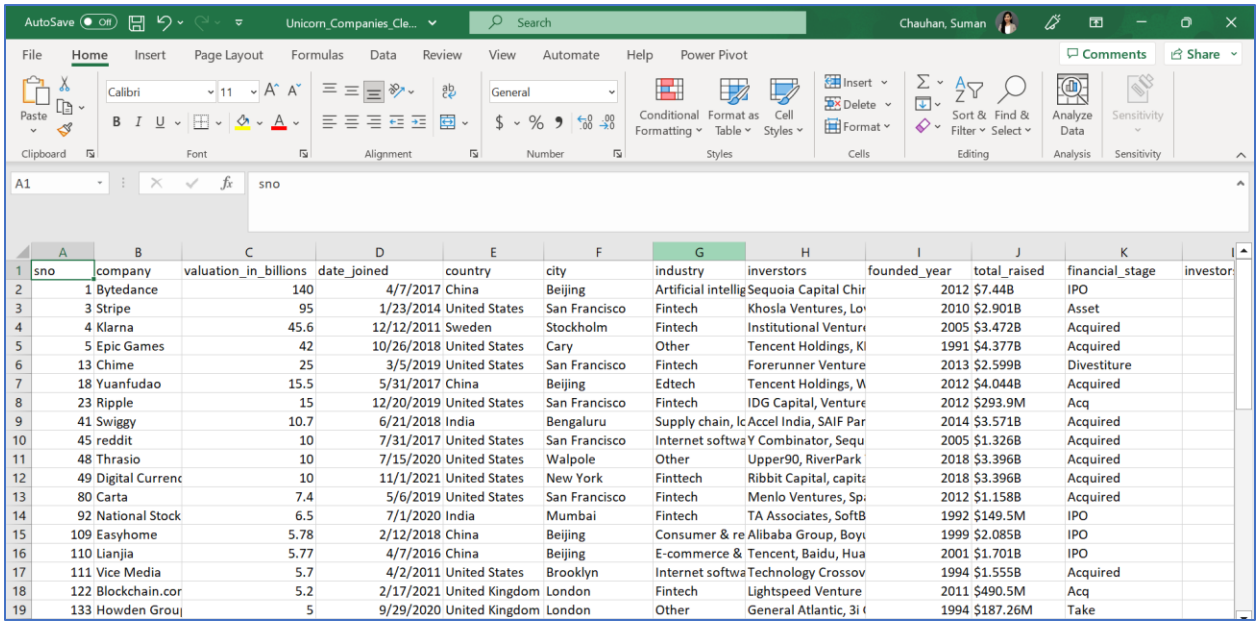
## Dataset Description

| Column No. | Field Name | Description | Example Value |
|---|---|---|---|
| 1 | company | This column has the Name of the Unicorn Company | Space X |
| 2 | valuation_in_billions | This column indicates Value of a company in billions of dollars | $140 |
| 3 | date_joined | This column shows date a company has | 4/7/17 |

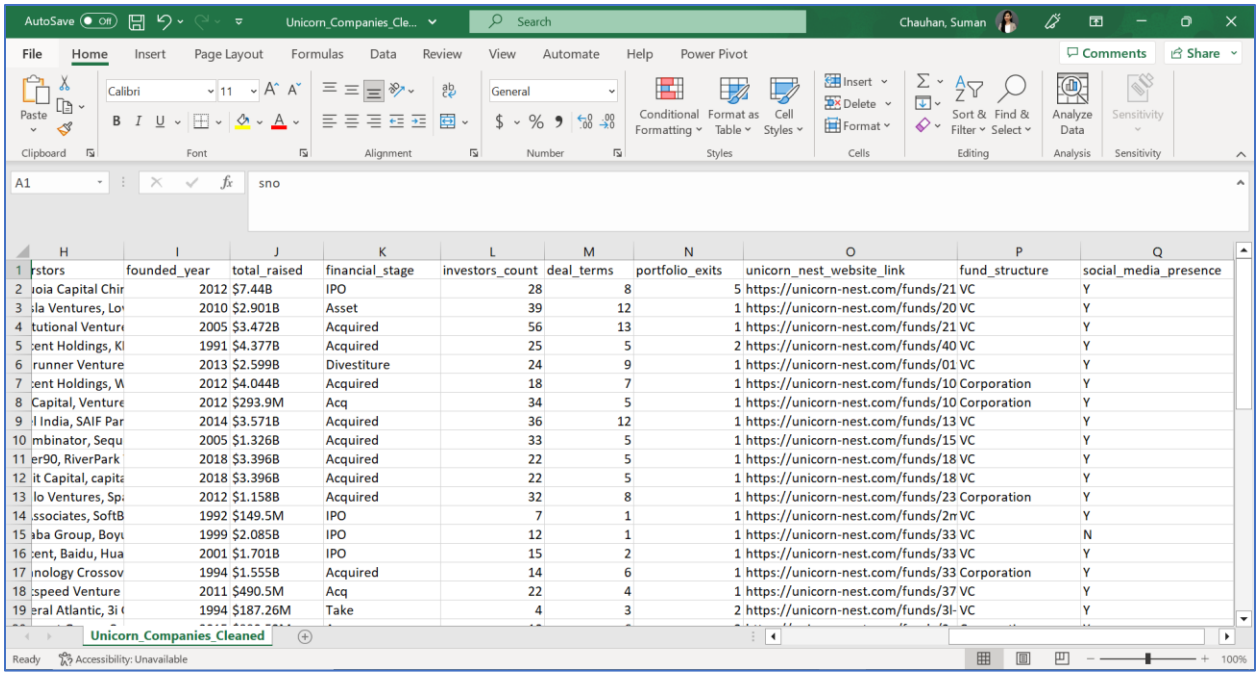
| | | | |
|---|---|---|---|
| | | been valued as a unicorn | |
| 4 | country | This column has Unicorn's country of origin | United States |
| 5 | city | This column is Place in a country where the Unicorn exists | San Francisco |
| 6 | industry | This column is Type of Unicorn industry | Supply Chain, Logistics, and delivery |
| 7 | select_investors | This column is Any person/entity who commits capital in Unicorn with the expectation of receiving financial returns | Sequoia Capital China, SIG Asia Investments, Softbank Group |
| 8 | founded_year | This column is In the year unicorn was founded | 2012 |
| 9 | total_raised | This column is to date, amount raised by investors for a unicorn | $7.44B |
| 10 | financial_stage | This column indicates in what stage of financial development is unicorn company now | IPO, Acquired |
| 11 | investors_count | This column is Number of investors who have invested in Unicorn | 28 |
| 12 | deal_terms | This column is Total number of deal terms agreed | 8 |

| | | | |
|---|---|---|---|
| | | between Unicorns and investors | |
| 13 | portfolio_exits | This column is Number of investors who have exited Unicorns | 1 |
| 14 | unicorn_nest_website_link | This column provides the website link to the company's profile on the Unicorn Nest website | https://unicorn-nest.com/funds/2020/ |
| 15 | fund_structure | This column specifies the fund structure of the company | Corporation |
| 16 | social_media_presence | This column indicates whether the company has presence on social media | Y |

**URL to the data set: [https://www.kaggle.com/datasets/soniya1809/unicorn-companies-data](https://www.kaggle.com/datasets/soniya1809/unicorn-companies-data)**
**Screenshot of dataset:**



| sno | company | valuation_in_billions | date_joined | country | city | industry | inverstors | founded_year | total_raised | financial_stage | investor |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | Bytedance | 140 | 4/7/2017 | China | Beijing | Artificial intellig | Sequoia Capital Chir | 2012 | $7.44B | IPO | |
| 3 | Stripe | 95 | 1/23/2014 | United States | San Francisco | Fintech | Khosla Ventures, Lo | 2010 | $2.901B | Asset | |
| 4 | Klarna | 45.6 | 12/12/2011 | Sweden | Stockholm | Fintech | Institutional Venture | 2005 | $3.472B | Acquired | |
| 5 | Epic Games | 42 | 10/26/2018 | United States | Cary | Other | Tencent Holdings, Kl | 1991 | $4.377B | Acquired | |
| 13 | Chime | 25 | 3/5/2019 | United States | San Francisco | Fintech | Forerunner Venture | 2013 | $2.599B | Divestiture | |
| 18 | Yuanfudao | 15.5 | 5/31/2017 | China | Beijing | Edtech | Tencent Holdings, W | 2012 | $4.044B | Acquired | |
| 23 | Ripple | 15 | 12/20/2019 | United States | San Francisco | Fintech | IDG Capital, Venture | 2012 | $293.9M | Acq | |
| 41 | Swiggy | 10.7 | 6/21/2018 | India | Bengaluru | Supply chain, lc | Accel India, SAIF Par | 2014 | $3.571B | Acquired | |
| 45 | reddit | 10 | 7/31/2017 | United States | San Francisco | Internet softwa | Y Combinator, Sequ | 2005 | $1.326B | Acquired | |
| 48 | Thrasio | 10 | 7/15/2020 | United States | Walpole | Other | Upper90, RiverPark | 2018 | $3.396B | Acquired | |
| 49 | Digital Currenc | 10 | 11/1/2021 | United States | New York | Finttech | Ribbit Capital, capita | 2018 | $3.396B | Acquired | |
| 80 | Carta | 7.4 | 5/6/2019 | United States | San Francisco | Fintech | Menlo Ventures, Spi | 2012 | $1.158B | Acquired | |
| 92 | National Stock | 6.5 | 7/1/2020 | India | Mumbai | Fintech | TA Associates, SoftB | 1992 | $149.5M | IPO | |
| 109 | Easyhome | 5.78 | 2/12/2018 | China | Beijing | Consumer & re | Alibaba Group, Boyu | 1999 | $2.085B | IPO | |
| 110 | Lianjia | 5.77 | 4/7/2016 | China | Beijing | E-commerce & | Tencent, Baidu, Hua | 2001 | $1.701B | IPO | |
| 111 | Vice Media | 5.7 | 4/2/2011 | United States | Brooklyn | Internet softwa | Technology Crossov | 1994 | $1.555B | Acquired | |
| 122 | Blockchain.cor | 5.2 | 2/17/2021 | United Kingdom | London | Fintech | Lightspeed Venture | 2011 | $490.5M | Acq | |
| 133 | Howden Grou | 5 | 9/29/2020 | United Kingdom | London | Other | General Atlantic, 3i | 1994 | $187.26M | Take | |



| | H | I | J | K | L | M | N | O | P | Q |
|---|---|---|---|---|---|---|---|---|---|---|
| 1 | rstors | founded_year | total_raised | financial_stage | investors_count | deal_terms | portfolio_exits | unicorn_nest_website_link | fund_structure | social_media_presence |
| 2 | oia Capital Chir | 2012 | $7.44B | IPO | 28 | 8 | 5 | https://unicorn-nest.com/funds/21 | VC | Y |
| 3 | la Ventures, Lo | 2010 | $2.901B | Asset | 39 | 12 | 1 | https://unicorn-nest.com/funds/20 | VC | Y |
| 4 | tutional Venture | 2005 | $3.472B | Acquired | 56 | 13 | 1 | https://unicorn-nest.com/funds/21 | VC | Y |
| 5 | ent Holdings, Kl | 1991 | $4.377B | Acquired | 25 | 5 | 2 | https://unicorn-nest.com/funds/40 | VC | Y |
| 6 | runner Venture | 2013 | $2.599B | Divestiture | 24 | 9 | 1 | https://unicorn-nest.com/funds/01 | VC | Y |
| 7 | ent Holdings, W | 2012 | $4.044B | Acquired | 18 | 7 | 1 | https://unicorn-nest.com/funds/10 | Corporation | Y |
| 8 | Capital, Venture | 2012 | $293.9M | Acq | 34 | 5 | 1 | https://unicorn-nest.com/funds/10 | Corporation | Y |
| 9 | l India, SAIF Par | 2014 | $3.571B | Acquired | 36 | 12 | 1 | https://unicorn-nest.com/funds/13 | VC | Y |
| 10 | mbinator, Sequ | 2005 | $1.326B | Acquired | 33 | 5 | 1 | https://unicorn-nest.com/funds/15 | VC | Y |
| 11 | er90, RiverPark | 2018 | $3.396B | Acquired | 22 | 5 | 1 | https://unicorn-nest.com/funds/18 | VC | Y |
| 12 | it Capital, capita | 2018 | $3.396B | Acquired | 22 | 5 | 1 | https://unicorn-nest.com/funds/18 | VC | Y |
| 13 | lo Ventures, Sp | 2012 | $1.158B | Acquired | 32 | 8 | 1 | https://unicorn-nest.com/funds/23 | Corporation | Y |
| 14 | ssociates, SoftB | 1992 | $149.5M | IPO | 7 | 1 | 1 | https://unicorn-nest.com/funds/2r | VC | Y |
| 15 | ba Group, Boyu | 1999 | $2.085B | IPO | 12 | 1 | 1 | https://unicorn-nest.com/funds/33 | VC | N |
| 16 | ent, Baidu, Hua | 2001 | $1.701B | IPO | 15 | 2 | 1 | https://unicorn-nest.com/funds/33 | VC | Y |
| 17 | nology Crossov | 1994 | $1.555B | Acquired | 14 | 6 | 1 | https://unicorn-nest.com/funds/33 | Corporation | Y |
| 18 | speed Venture | 2011 | $490.5M | Acq | 22 | 4 | 1 | https://unicorn-nest.com/funds/37 | VC | Y |
| 19 | eral Atlantic, 3i | 1994 | $187.26M | Take | 4 | 3 | 2 | https://unicorn-nest.com/funds/3l | VC | Y |

## Data Cleaning

### Problem Statement

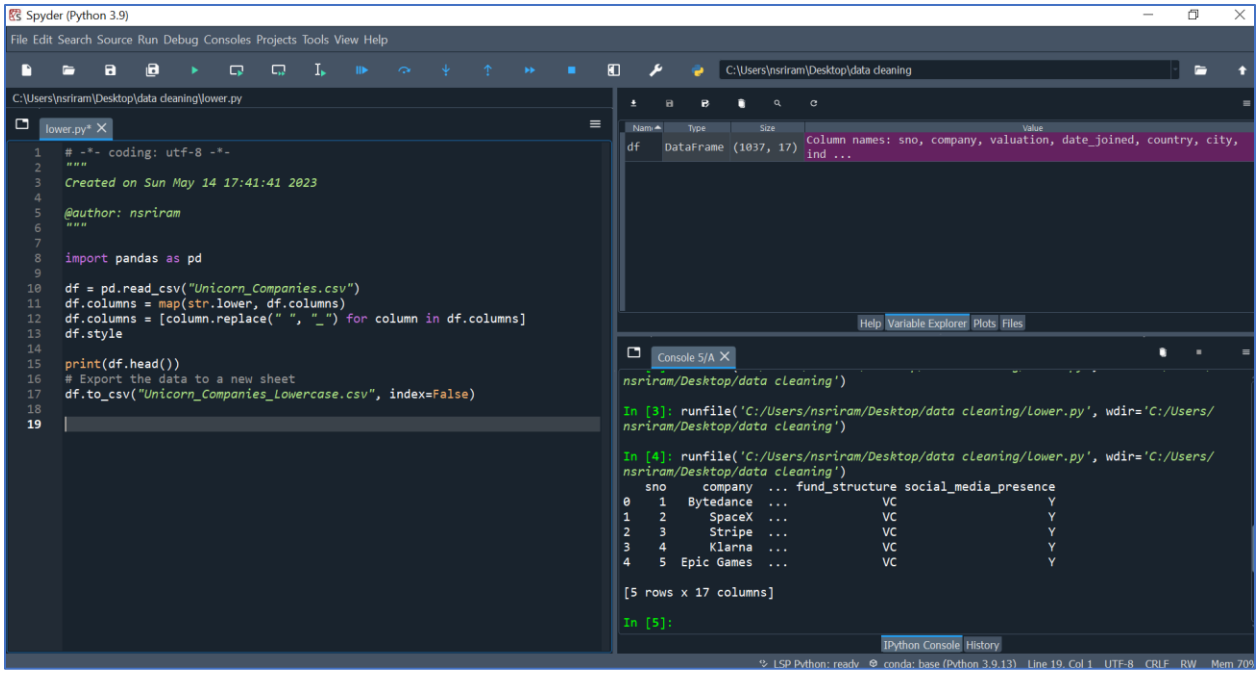
In the raw data, all column names are not in one case. They also have spaces in between which makes it difficult to analyze specific column data in Python.

### Steps Taken

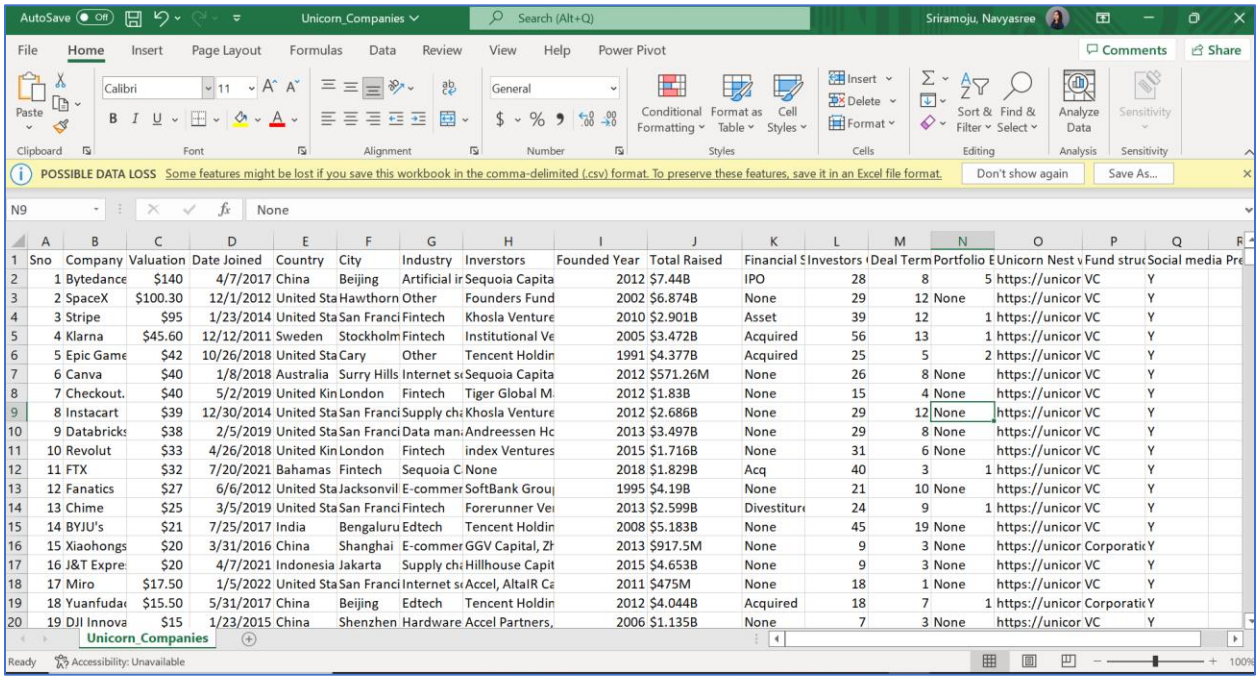
1. Changed all the column names to single case- lower case
2. Replaced spaces (‘ ‘) with underscore(_)

**Code:**

```
import pandas as pd
df = pd.read_csv("Unicorn_Companies.csv")
df.columns = map(str.lower, df.columns)
df.columns = [column.replace(" ", "_") for column in df.columns]
df.style
print(df.head())
# Export the data to a new sheet
df.to_csv("Unicorn_Companies_Lowercase.csv", index=False)
```

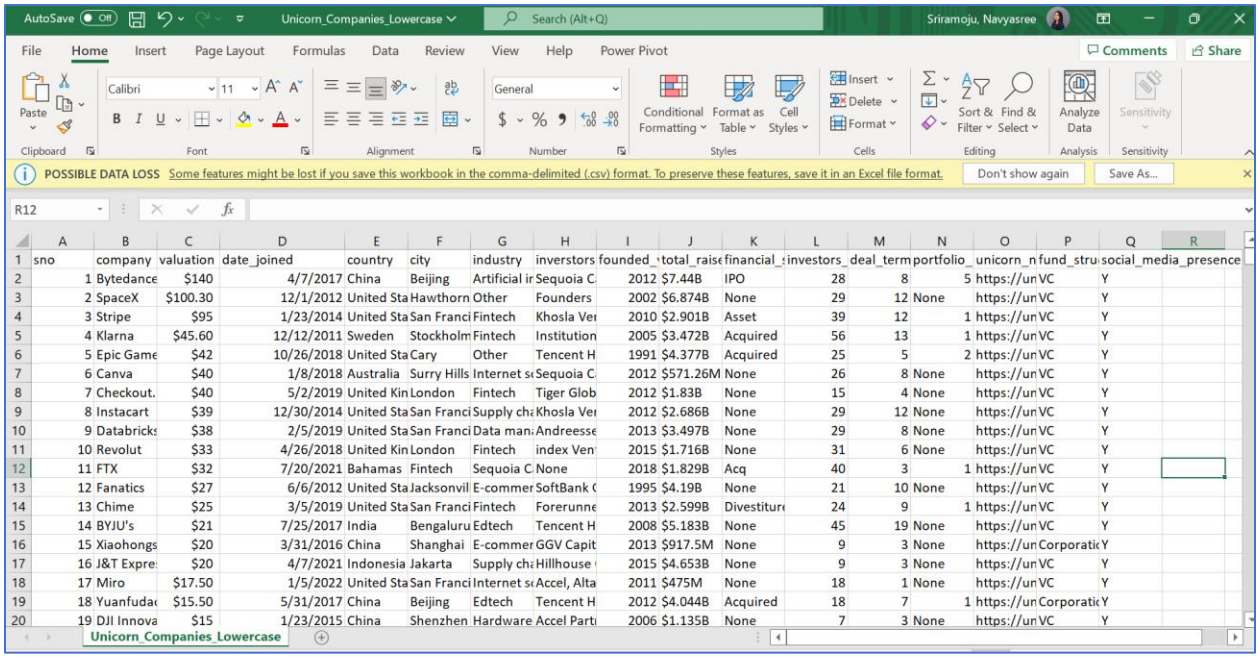



## Before Cleaning:




| | Sno | Company | Valuation | Date Joined | Country | City | Industry | Inverstors | Founded Year | Total Raised | Financial S | Investors ( | Deal Term | Portfolio E | Unicorn Nest v | Fund struc | Social media Pre |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 2 | 1 | Bytedance | $140 | 4/7/2017 | China | Beijing | Artificial ir | Sequoia Capita | 2012 | $7.44B | IPO | 28 | 8 | 5 | https://unicor | VC | Y |
| 3 | 2 | SpaceX | $100.30 | 12/1/2012 | United Sta | Hawthorn | Other | Founders Fund | 2002 | $6.874B | None | 29 | 12 | None | https://unicor | VC | Y |
| 4 | 3 | Stripe | $95 | 1/23/2014 | United Sta | San Franci | Fintech | Khosla Venture | 2010 | $2.901B | Asset | 39 | 12 | 1 | https://unicor | VC | Y |
| 5 | 4 | Klarna | $45.60 | 12/12/2011 | Sweden | Stockholm | Fintech | Institutional Ve | 2005 | $3.472B | Acquired | 56 | 13 | 1 | https://unicor | VC | Y |
| 6 | 5 | Epic Game | $42 | 10/26/2018 | United Sta | Cary | Other | Tencent Holdin | 1991 | $4.377B | Acquired | 25 | 5 | 2 | https://unicor | VC | Y |
| 7 | 6 | Canva | $40 | 1/8/2018 | Australia | Surry Hills | Internet s | Sequoia Capita | 2012 | $571.26M | None | 26 | 8 | None | https://unicor | VC | Y |
| 8 | 7 | Checkout. | $40 | 5/2/2019 | United Kin | London | Fintech | Tiger Global M | 2012 | $1.83B | None | 15 | 4 | None | https://unicor | VC | Y |
| 9 | 8 | Instacart | $39 | 12/30/2014 | United Sta | San Franci | Supply cha | Khosla Venture | 2012 | $2.686B | None | 29 | 12 | None | https://unicor | VC | Y |
| 10 | 9 | Databricks | $38 | 2/5/2019 | United Sta | San Franci | Data mani | Andreessen Ho | 2013 | $3.497B | None | 29 | 8 | None | https://unicor | VC | Y |
| 11 | 10 | Revolut | $33 | 4/26/2018 | United Kin | London | Fintech | index Ventures | 2015 | $1.716B | None | 31 | 6 | None | https://unicor | VC | Y |
| 12 | 11 | FTX | $32 | 7/20/2021 | Bahamas | Fintech | Sequoia C | None | 2018 | $1.829B | Acq | 40 | 3 | 1 | https://unicor | VC | Y |
| 13 | 12 | Fanatics | $27 | 6/6/2012 | United Sta | Jacksonvil | E-commer | SoftBank Grou | 1995 | $4.19B | None | 21 | 10 | None | https://unicor | VC | Y |
| 14 | 13 | Chime | $25 | 3/5/2019 | United Sta | San Franci | Fintech | Forerunner Ve | 2013 | $2.599B | Divestiture | 24 | 9 | 1 | https://unicor | VC | Y |
| 15 | 14 | BYJU's | $21 | 7/25/2017 | India | Bengaluru | Edtech | Tencent Holdin | 2008 | $5.183B | None | 45 | 19 | None | https://unicor | VC | Y |
| 16 | 15 | Xiaohongs | $20 | 3/31/2016 | China | Shanghai | E-commer | GGV Capital, ZH | 2013 | $917.5M | None | 9 | 3 | None | https://unicor | Corporatic | Y |
| 17 | 16 | J&T Expre | $20 | 4/7/2021 | Indonesia | Jakarta | Supply cha | Hillhouse Capit | 2015 | $4.653B | None | 9 | 3 | None | https://unicor | VC | Y |
| 18 | 17 | Miro | $17.50 | 1/5/2022 | United Sta | San Franci | Internet s | Accel, AltaIR Ca | 2011 | $475M | None | 18 | 1 | None | https://unicor | VC | Y |
| 19 | 18 | Yuanfudao | $15.50 | 5/31/2017 | China | Beijing | Edtech | Tencent Holdin | 2012 | $4.044B | Acquired | 18 | 7 | 1 | https://unicor | Corporatic | Y |
| 20 | 19 | DJI Innova | $15 | 1/23/2015 | China | Shenzhen | Hardware | Accel Partners, | 2006 | $1.135B | None | 7 | 3 | None | https://unicor | VC | Y |

**After Cleaning:**



| sno | company | valuation | date_joined | country | city | industry | inverstors | founded_ | total_raise | financial_s | investors_ | deal_term | portfolio_ | unicorn_n | fund_stru | social_media_presence |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | Bytedance | $140 | 4/7/2017 | China | Beijing | Artificial ir | Sequoia C | 2012 | $7.44B | IPO | 28 | 8 | 5 | https://un | VC | Y |
| 2 | SpaceX | $100.30 | 12/1/2012 | United Sta | Hawthorn | Other | Founders | 2002 | $6.874B | None | 29 | 12 | None | https://un | VC | Y |
| 3 | Stripe | $95 | 1/23/2014 | United Sta | San Franci | Fintech | Khosla Ver | 2010 | $2.901B | Asset | 39 | 12 | 1 | https://un | VC | Y |
| 4 | Klarna | $45.60 | 12/12/2011 | Sweden | Stockholm | Fintech | Institution | 2005 | $3.472B | Acquired | 56 | 13 | 1 | https://un | VC | Y |
| 5 | Epic Game | $42 | 10/26/2018 | United Sta | Cary | Other | Tencent H | 1991 | $4.377B | Acquired | 25 | 5 | 2 | https://un | VC | Y |
| 6 | Canva | $40 | 1/8/2018 | Australia | Surry Hills | Internet s | Sequoia C | 2012 | $571.26M | None | 26 | 8 | None | https://un | VC | Y |
| 7 | Checkout. | $40 | 5/2/2019 | United Kin | London | Fintech | Tiger Glob | 2012 | $1.83B | None | 15 | 4 | None | https://un | VC | Y |
| 8 | Instacart | $39 | 12/30/2014 | United Sta | San Franci | Supply cha | Khosla Ver | 2012 | $2.686B | None | 29 | 12 | None | https://un | VC | Y |
| 9 | Databricks | $38 | 2/5/2019 | United Sta | San Franci | Data man | Andreesse | 2013 | $3.497B | None | 29 | 8 | None | https://un | VC | Y |
| 10 | Revolut | $33 | 4/26/2018 | United Kin | London | Fintech | index Ven | 2015 | $1.716B | None | 31 | 6 | None | https://un | VC | Y |
| 11 | FTX | $32 | 7/20/2021 | Bahamas | Fintech | Sequoia C | None | 2018 | $1.829B | Acq | 40 | 3 | 1 | https://un | VC | Y |
| 12 | Fanatics | $27 | 6/6/2012 | United Sta | Jacksonvil | E-commer | SoftBank ( | 1995 | $4.19B | None | 21 | 10 | None | https://un | VC | Y |
| 13 | Chime | $25 | 3/5/2019 | United Sta | San Franci | Fintech | Forerunne | 2013 | $2.599B | Divestiture | 24 | 9 | 1 | https://un | VC | Y |
| 14 | BYJU's | $21 | 7/25/2017 | India | Bengaluru | Edtech | Tencent H | 2008 | $5.183B | None | 45 | 19 | None | https://un | VC | Y |
| 15 | Xiaohongs | $20 | 3/31/2016 | China | Shanghai | E-commer | GGV Capit | 2013 | $917.5M | None | 9 | 3 | None | https://un | Corporatic | Y |
| 16 | J&T Expre | $20 | 4/7/2021 | Indonesia | Jakarta | Supply cha | Hillhouse | 2015 | $4.653B | None | 9 | 3 | None | https://un | VC | Y |
| 17 | Miro | $17.50 | 1/5/2022 | United Sta | San Franci | Internet s | Accel, Alta | 2011 | $475M | None | 18 | 1 | None | https://un | VC | Y |
| 18 | Yuanfuda | $15.50 | 5/31/2017 | China | Beijing | Edtech | Tencent H | 2012 | $4.044B | Acquired | 18 | 7 | 1 | https://un | Corporatic | Y |
| 19 | DJI Innova | $15 | 1/23/2015 | China | Shenzhen | Hardware | Accel Parti | 2006 | $1.135B | None | 7 | 3 | None | https://un | VC | Y |

**2. Renaming column Name, replacing string value '$' with ' ', changing datatype to float**

**Problem Statement**

During analysis, the column valuation provides information regarding the company's valuation in billions of dollars. However, this column contains special characters that make it difficult to read in python. Additionally, the data type of this column is not numeric.

**Steps Taken**

1. Renamed the column name from Valuation to valuation_in_billions
2. Removed the special character '$'
3. Changed the data type to float

**Code:**

```
import pandas as pd

import numpy as np

df = pd.read_csv("/Users/nsriram/Unicorn_Companies.csv")

df.columns = map(str.lower, df.columns)

df.columns = [column.replace(" ", "_") for column in df.columns]
```

df.rename(columns={'valuation':'valuation_in_billions'}, inplace=True)

df['valuation_in_billions'] = df['valuation_in_billions'].str.replace('$','')

df['valuation_in_billions'] = df['valuation_in_billions'].astype('float')

df.style

df.to_csv("Unicorn_Companies_modified.csv", index=False)

```
In [42]: import pandas as pd
         import numpy as np
         df = pd.read_csv("/Users/nsriram/Unicorn_Companies.csv")
         df.columns = map(str.lower, df.columns)
         df.columns = [column.replace(" ", "_") for column in df.columns]
         df.rename(columns={'valuation':'valuation_in_billions'}, inplace=True)
         df['valuation_in_billions'] = df['valuation_in_billions'].str.replace('$','')
         df['valuation_in_billions'] = df['valuation_in_billions'].astype('float')
         df.style

         df.to_csv("Unicorn_Companies_modified.csv", index=False)
```

**Before Cleaning:**

| Sno | Company | Valuation | Date Joined | Country | City | Industry | Inverstors | Founded Year | Total Raised | Financial Stage | Investors Count | Deal Terms | Portfo Ex |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | Bytedance | $140 | 4/7/2017 | China | Beijing | Artificial intelligence | Sequoia Capital China, SIG Asia Investments, Sina Weibo, Softbank Group | 2012 | $7.44B | IPO | 28 | 8 | |
| 2 | SpaceX | $100.30 | 12/1/2012 | United States | Hawthorne | Other | Founders Fund, Draper Fisher Jurvetson, Rothenberg Ventures | 2002 | $6.874B | None | 29 | 12 | No |
| 3 | Stripe | $95 | 1/23/2014 | United States | San Francisco | Fintech | Khosla Ventures, LowercaseCapital, capitalG | 2010 | $2.901B | Asset | 39 | 12 | |
| 4 | Klarna | $45.60 | 12/12/2011 | Sweden | Stockholm | Fintech | Institutional Venture Partners, | 2005 | $3.472B | Acquired | 56 | 13 | |

**After Cleaning:**

```
In [16]: import pandas as pd
         import numpy as np
         df = pd.read_csv("/Users/nsriram/Unicorn_Companies.csv")
         df.columns = map(str.lower, df.columns)
         df.columns = [column.replace(" ", "_") for column in df.columns]
         df.rename(columns={'valuation_($b)':'valuation_in_billions'}, inplace=True)
         df['valuation_in_billions'] = df['valuation_in_billions'].str.replace('$','')
         df['valuation_in_billions'] = df['valuation_in_billions'].astype('float')
         df.style
```

```
ers\nsriram\AppData\Local\Temp\ipykernel_22672\2795418673.py:7: FutureWarning: The default value of regex will change fr
ue to False in a future version. In addition, single character regular expressions will *not* be treated as literal stri
hen regex=True.
'valuation_in_billions'] = df['valuation_in_billions'].str.replace('$','')
```

| sno | company | valuation_in_billions | date_joined | country | city | industry | inverstors | founded_year | total_raised | financial_stage | inv |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | Bytedance | 140.000000 | 4/7/2017 | China | Beijing | Artificial intelligence | Sequoia Capital China, SIG Asia Investments, Sina Weibo, Softbank Group | 2012 | $7.44B | IPO | |
| 2 | SpaceX | 100.300000 | 12/1/2012 | United States | Hawthorne | Other | Founders Fund, Draper Fisher Jurvetson, Rothenberg Ventures | 2002 | $6.874B | None | |
| 3 | Stripe | 95.000000 | 1/23/2014 | United States | San Francisco | Fintech | Khosla Ventures, LowercaseCapital, | 2010 | $2.901B | Asset | |

### 3. Handling None values

**Problem Statement: Analyzed the columns which have None values in them**

**Code**

```
import pandas as pd

# Read the file into a DataFrame

df = pd.read_csv('/Users/nsriram/Unicorn_Companies_modified.csv')

# Loop through each column and check for 'None' in the unique values

for column in df.columns:

 if 'None' in df[column].unique():

 print(column)
```

```
In [44]: import pandas as pd

         # Read the Unicorn_Companies.csv file into a DataFrame
         df = pd.read_csv('/Users/nsriram/Unicorn_Companies_modified.csv')

         # Loop through each column and check for 'None' in the unique values
         for column in df.columns:
             if 'None' in df[column].unique():
                 print(column)

         inverstors
         founded_year
         total_raised
         financial_stage
         investors_count
         deal_terms
         portfolio_exits

         C:\Users\nsriram\AppData\Local\Temp\ipykernel_22752\1275729036.py:8: FutureWarning: elementwise comparison failed; returning sc
         alar instead, but in the future will perform elementwise comparison
           if 'None' in df[column].unique():
```

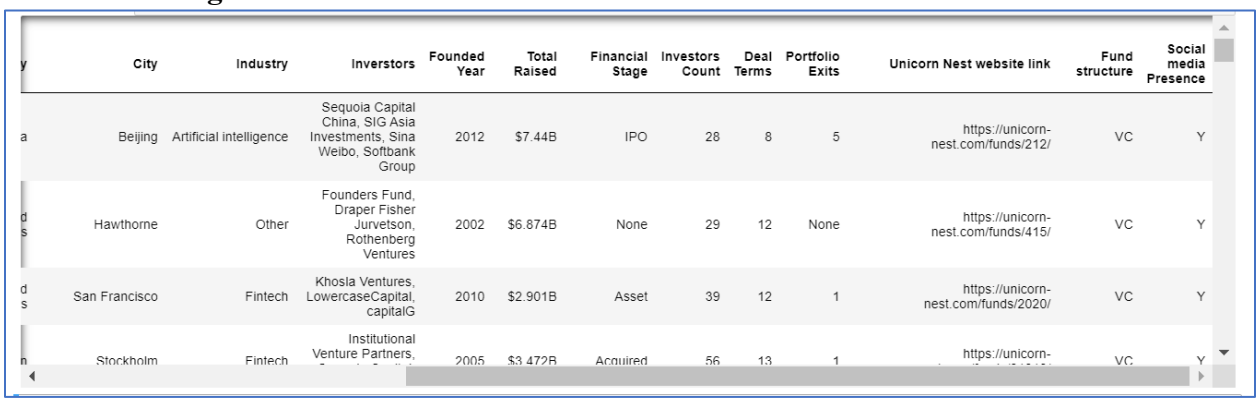
**Replaced None values with numpy.nan**

**Code**

```
import pandas as pd

df = pd.read_csv('/Users/nsriram/Unicorn_Companies.csv')

df['Inverstors'] = df['Inverstors'].replace('None', np.nan)
df['Total Raised'] = df['Total Raised'].replace('None', np.nan)
df['Deal Terms'] = df['Deal Terms'].replace('None', np.nan)
df['Portfolio Exits'] = df['Portfolio Exits'].replace('None', np.nan)
df['Founded Year'] = df['Founded Year'].replace('None', np.nan)
df['Investors Count'] = df['Investors Count'].replace('None', np.nan)
df['Financial Stage'] = df['Financial Stage'].replace('None', np.nan)

df.style
```

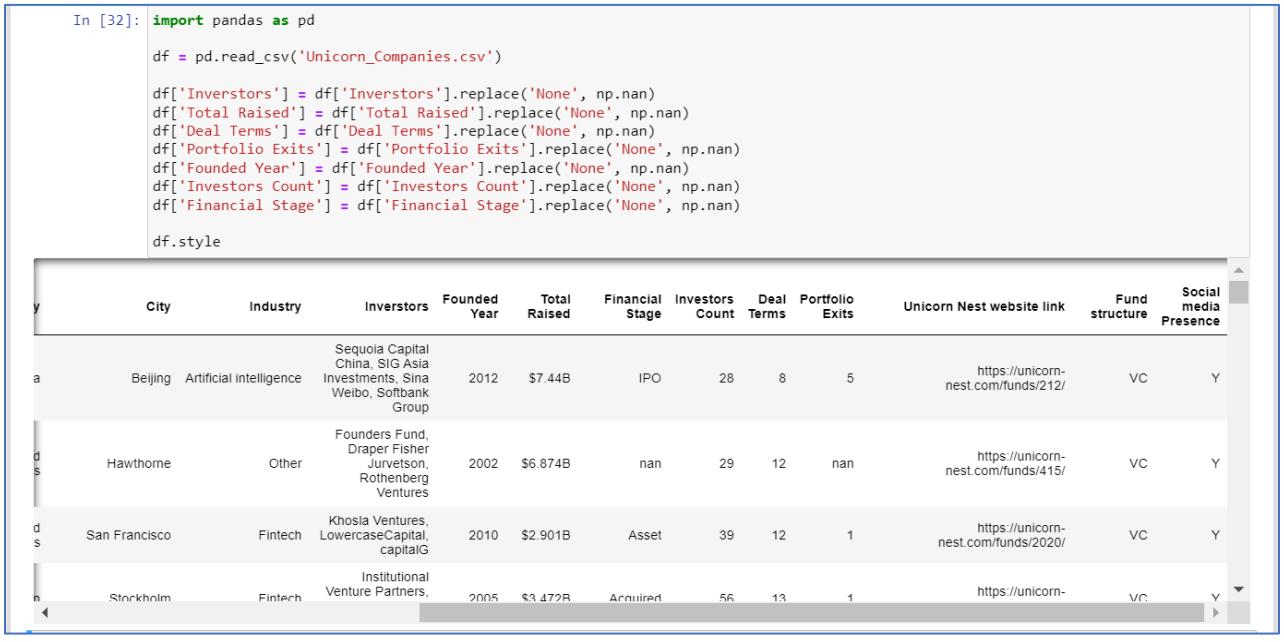
```
#Export to new csv
df.to_csv("Unicorn_Companies_None.csv", index=False)
```

## Before Cleaning:



| y | City | Industry | Inverstors | Founded Year | Total Raised | Financial Stage | Investors Count | Deal Terms | Portfolio Exits | Unicorn Nest website link | Fund structure | Social media Presence |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| a | Beijing | Artificial intelligence | Sequoia Capital China, SIG Asia Investments, Sina Weibo, Softbank Group | 2012 | $7.44B | IPO | 28 | 8 | 5 | https://unicorn-nest.com/funds/212/ | VC | Y |
| d s | Hawthorne | Other | Founders Fund, Draper Fisher Jurvetson, Rothenberg Ventures | 2002 | $6.874B | None | 29 | 12 | None | https://unicorn-nest.com/funds/415/ | VC | Y |
| d s | San Francisco | Fintech | Khosla Ventures, LowercaseCapital, capitalG | 2010 | $2.901B | Asset | 39 | 12 | 1 | https://unicorn-nest.com/funds/2020/ | VC | Y |
| h | Stockholm | Fintech | Institutional Venture Partners, | 2005 | $3.472B | Acquired | 56 | 13 | 1 | https://unicorn- | VC | Y |

## After Cleaning:



```
In [32]: import pandas as pd

df = pd.read_csv('Unicorn_Companies.csv')

df['Inverstors'] = df['Inverstors'].replace('None', np.nan)
df['Total Raised'] = df['Total Raised'].replace('None', np.nan)
df['Deal Terms'] = df['Deal Terms'].replace('None', np.nan)
df['Portfolio Exits'] = df['Portfolio Exits'].replace('None', np.nan)
df['Founded Year'] = df['Founded Year'].replace('None', np.nan)
df['Investors Count'] = df['Investors Count'].replace('None', np.nan)
df['Financial Stage'] = df['Financial Stage'].replace('None', np.nan)

df.style
```

| y | City | Industry | Inverstors | Founded Year | Total Raised | Financial Stage | Investors Count | Deal Terms | Portfolio Exits | Unicorn Nest website link | Fund structure | Social media Presence |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| a | Beijing | Artificial intelligence | Sequoia Capital China, SIG Asia Investments, Sina Weibo, Softbank Group | 2012 | $7.44B | IPO | 28 | 8 | 5 | https://unicorn-nest.com/funds/212/ | VC | Y |
| d s | Hawthorne | Other | Founders Fund, Draper Fisher Jurvetson, Rothenberg Ventures | 2002 | $6.874B | nan | 29 | 12 | nan | https://unicorn-nest.com/funds/415/ | VC | Y |
| d s | San Francisco | Fintech | Khosla Ventures, LowercaseCapital, capitalG | 2010 | $2.901B | Asset | 39 | 12 | 1 | https://unicorn-nest.com/funds/2020/ | VC | Y |
| h | Stockholm | Fintech | Institutional Venture Partners, | 2005 | $3.472B | Acquired | 56 | 13 | 1 | https://unicorn- | VC | Y |

## Statistical Functions

### 1. Summary Statistics for Valuation_in_billions

**Code:**

```
import pandas as pd
import matplotlib.pyplot as plt
df = pd.read_csv("/Users/nsriram/Unicorn_Companies_Cleaned.csv")
print("Minimum valuation in billions:" + str(df['valuation_in_billions'].min()))
print("Maximum valuation in billions:" + str(df['valuation_in_billions'].max()))
print("Median of valuation in billions :" + str(df['valuation_in_billions'].median()))
print("Mode of valuation in billions:\n" + str(df['valuation_in_billions'].mode()))
print("Mean of valuation in billions:" + str(df['valuation_in_billions'].mean()))
print("Standard Deviation of valuation in billions:" + str(df['valuation_in_billions'].std()))
print("Describe:\n" + str(df['valuation_in_billions'].describe()))
```

```
In [11]: import pandas as pd
         import matplotlib.pyplot as plt
         df = pd.read_csv("/Users/nsriram/Unicorn_Companies_Cleaned.csv")
         print("Minimum valuation in billions:" + str(df['valuation_in_billions'].min()))
         print("Maximum valuation in billions:" + str(df['valuation_in_billions'].max()))
         print("Median of valuation in billions :" + str(df['valuation_in_billions'].median()))
         print("Mode of valuation in billions:\n" + str(df['valuation_in_billions'].mode()))
         print("Mean of valuation in billions:" + str(df['valuation_in_billions'].mean()))
         print("Standard Deviation of valuation in billions:" + str(df['valuation_in_billions'].std()))
         print("Describe:\n" + str(df['valuation_in_billions'].describe()))

         Minimum valuation in billions:1.0
         Maximum valuation in billions:140.0
         Median of valuation in billions :4.0
         Mode of valuation in billions:
         0    1.0
         Name: valuation_in_billions, dtype: float64
         Mean of valuation in billions:11.91022727272727
         Standard Deviation of valuation in billions:25.60217309895796
         Describe:
         count     44.000000
         mean      11.910227
         std       25.602173
         min        1.000000
         25%        1.687500
         50%        4.000000
         75%        8.050000
         max      140.000000
         Name: valuation_in_billions, dtype: float64
```

**Summary**:

Based on the summary statistics, the highest valuation observed is $140B, indicating that unicorn companies are in a favorable position. It is evident that the lowest valuation is $1B, which serves as the threshold for being categorized as a unicorn. The 25th percentile value stands at $1B, implying that a significant portion of companies have surpassed the minimum valuation. In fact,

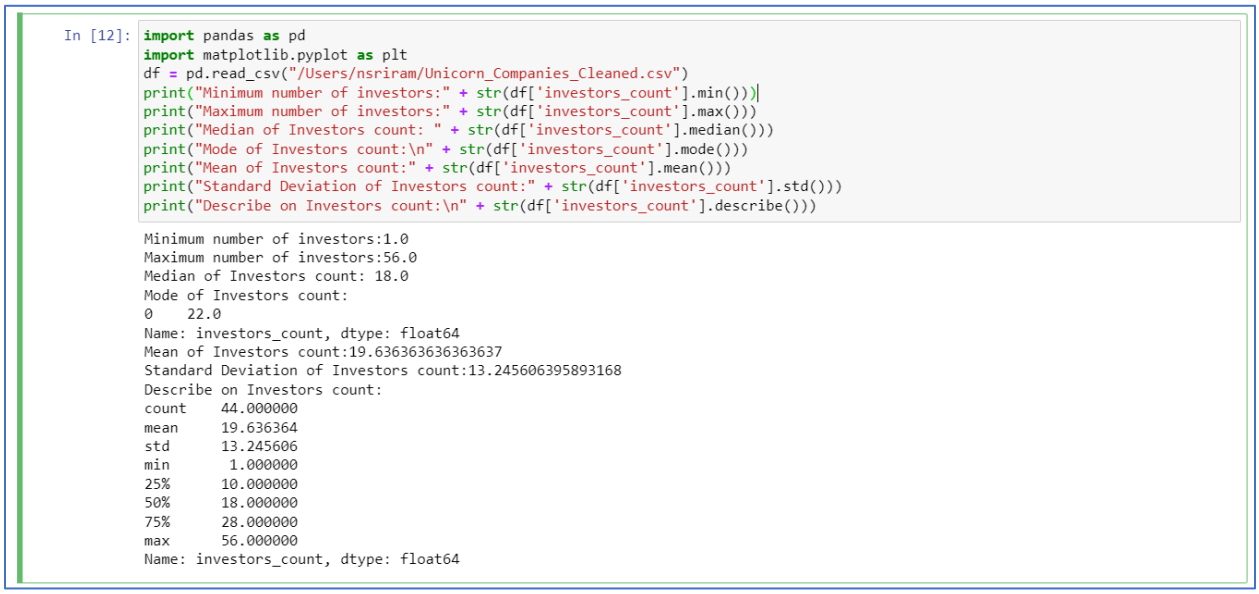
the mean valuation is 3.29B, indicating that the majority of companies have already exceeded the minimum requirement. Additionally, the standard deviation is $7.3B, twice the mean value, suggesting that valuations of unicorn companies span a wide range from $1B to $140B, rather than being concentrated solely at the $1B minimum valuation.

**2. Summary Statistics for investors_count**

**Code**

**#Statistical summary of investors_count**

```
import pandas as pd
import matplotlib.pyplot as plt
df = pd.read_csv("/Users/nsriram/Unicorn_Companies_Cleaned.csv")
print("Minimum number of investors:" + str(df['investors_count'].min()))
print("Maximum number of investors:" + str(df['investors_count'].max()))
print("Median of Investors count: " + str(df['investors_count'].median()))
print("Mode of Investors count:\n" + str(df['investors_count'].mode()))
print("Mean of Investors count:" + str(df['investors_count'].mean()))
print("Standard Deviation of Investors count:" + str(df['investors_count'].std()))
print("Describe on Investors count:\n" + str(df['investors_count'].describe()))
```



```
In [12]: import pandas as pd
         import matplotlib.pyplot as plt
         df = pd.read_csv("/Users/nsriram/Unicorn_Companies_Cleaned.csv")
         print("Minimum number of investors:" + str(df['investors_count'].min()))
         print("Maximum number of investors:" + str(df['investors_count'].max()))
         print("Median of Investors count: " + str(df['investors_count'].median()))
         print("Mode of Investors count:\n" + str(df['investors_count'].mode()))
         print("Mean of Investors count:" + str(df['investors_count'].mean()))
         print("Standard Deviation of Investors count:" + str(df['investors_count'].std()))
         print("Describe on Investors count:\n" + str(df['investors_count'].describe()))

         Minimum number of investors:1.0
         Maximum number of investors:56.0
         Median of Investors count: 18.0
         Mode of Investors count:
         0    22.0
         Name: investors_count, dtype: float64
         Mean of Investors count:19.636363636363637
         Standard Deviation of Investors count:13.245606395893168
         Describe on Investors count:
         count    44.000000
         mean     19.636364
         std      13.245606
         min       1.000000
         25%      10.000000
         50%      18.000000
         75%      28.000000
         max      56.000000
         Name: investors_count, dtype: float64
```

**Summary**:

The summary statistics reveal that the highest number of investors observed is 91, while the lowest is 1. Factors such as interest rates, technological advancements, and regulatory advantages have collectively led to an increase in investments in unicorn companies in recent times. The 75th percentile value of 19 suggests that the majority of unicorns have approximately more than 15 investors. This is further supported by the mean value, which indicates that most companies have received investments from approximately 15 individuals or entities. The standard deviation, which is 9.9, signifies the variability in the number of investors among unicorn companies.

**3. Summary Statistics for deal_terms**

**Code**

```
#Statistical summary of deal_terms

import pandas as pd
import matplotlib.pyplot as plt

df = pd.read_csv("/Users/nsriram/Unicorn_Companies_Cleaned.csv")

print("Minimum number of deal terms:" + str(df['deal_terms'].min()))
print("Maximum number of deal terms:" + str(df['deal_terms'].max()))
print("Median of deal terms: " + str(df['deal_terms'].median()))
print("Mode of deal terms:\n" + str(df['deal_terms'].mode()))
print("Mean of deal terms:" + str(df['deal_terms'].mean()))
print("Standard Deviation of deal terms:" + str(df['deal_terms'].std()))
print("Describe on deal terms:\n" + str(df['deal_terms'].describe()))
```

In [17]:
```
import pandas as pd
import matplotlib.pyplot as plt

df = pd.read_csv("/Users/nsriram/Unicorn_Companies_Cleaned.csv")

print("Minimum number of deal terms:" + str(df['deal_terms'].min()))
print("Maximum number of deal terms:" + str(df['deal_terms'].max()))
print("Median of deal terms: " + str(df['deal_terms'].median()))
print("Mode of deal terms:\n" + str(df['deal_terms'].mode()))
print("Mean of deal terms:" + str(df['deal_terms'].mean()))
print("Standard Deviation of deal terms:" + str(df['deal_terms'].std()))
print("Describe on deal terms:\n" + str(df['deal_terms'].describe()))
```

```
Minimum number of deal terms:1.0
Maximum number of deal terms:13.0
Median of deal terms: 5.0
Mode of deal terms:
0    1.0
1    5.0
Name: deal_terms, dtype: float64
Mean of deal terms:4.795454545454546
Standard Deviation of deal terms:3.1147029083346225
Describe on deal terms:
count    44.000000
mean      4.795455
std       3.114703
min       1.000000
25%       2.000000
50%       5.000000
75%       6.250000
max      13.000000
Name: deal_terms, dtype: float64
```

**Summary:**

According to the summary statistics, the highest number of deal conditions recorded is 19. Companies facing a larger number of deal conditions face a significant challenge in securing financing from investors. Based on the summary statistics, it is observed that 75% of the unicorns have deal conditions 4B, and the mean value is 3. This suggests that the majority of investors have invested in unicorns with less than 4 deal terms, indicating that many unicorns may share similar or comparable deal terms. The standard deviation, which is 2.15, is close to the mean of 3, further supporting the notion that the deal terms among unicorns are relatively consistent.

# Analysis &Visualizations

## 1. Top 5 Countries by Unicorn Company Valuation

### Code

```
import matplotlib.pyplot as plt

import seaborn as sns

import pandas as pd


# Load CSV file into pandas dataframe

df = pd.read_csv('Unicorn_Companies_Cleaned.csv')

# Calculate total unicorn valuation by country

country_valuation = df.groupby('country')["valuation_in_billions '].sum().reset_index()

# Sort dataframe by valuation in descending order

country_valuation = country_valuation.sort_values("valuation_in_billions ',
ascending=False)


# Take top 5 countries by valuation

top_5 = country_valuation[:5]


# Create bar plot using Seaborn

sns.barplot(x='country', y="valuation_in_billions ', data=top_5)


# Set plot title and axis labels

plt.title('Top 5 Countries by Unicorn Company Valuation')

plt.xlabel('country')

plt.ylabel("valuation_in_billions ')


# Display the plot

plt.show()
```

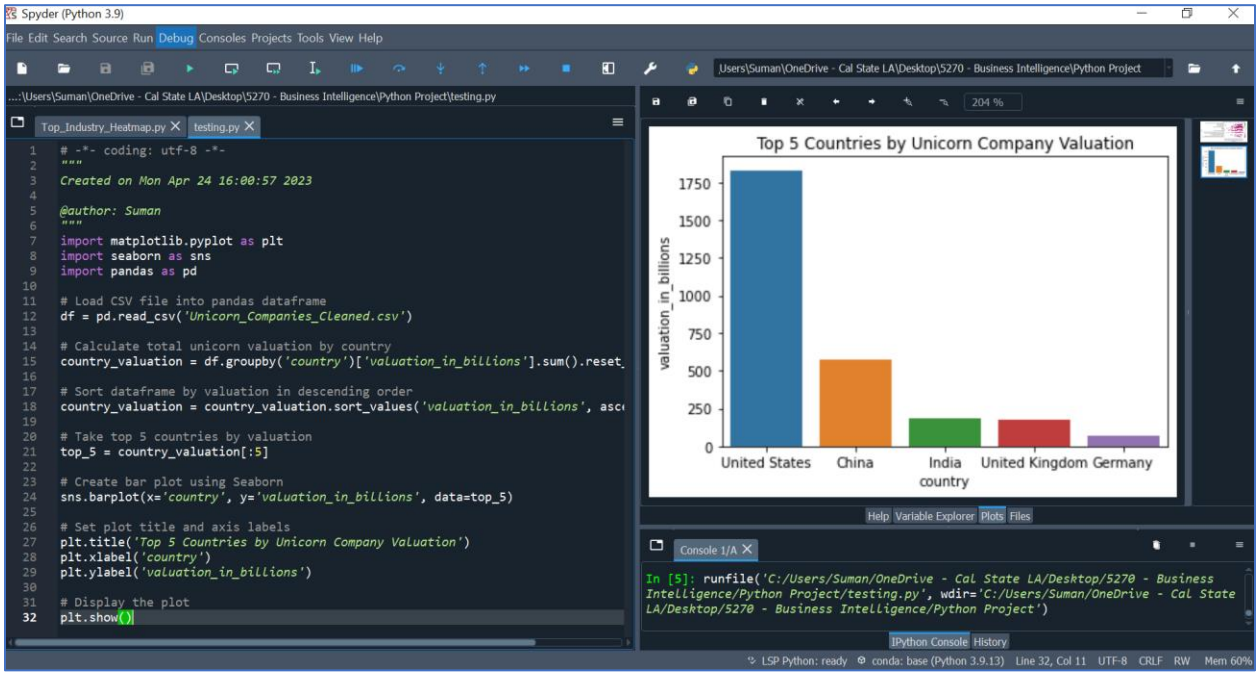
Spyder (Python 3.9)

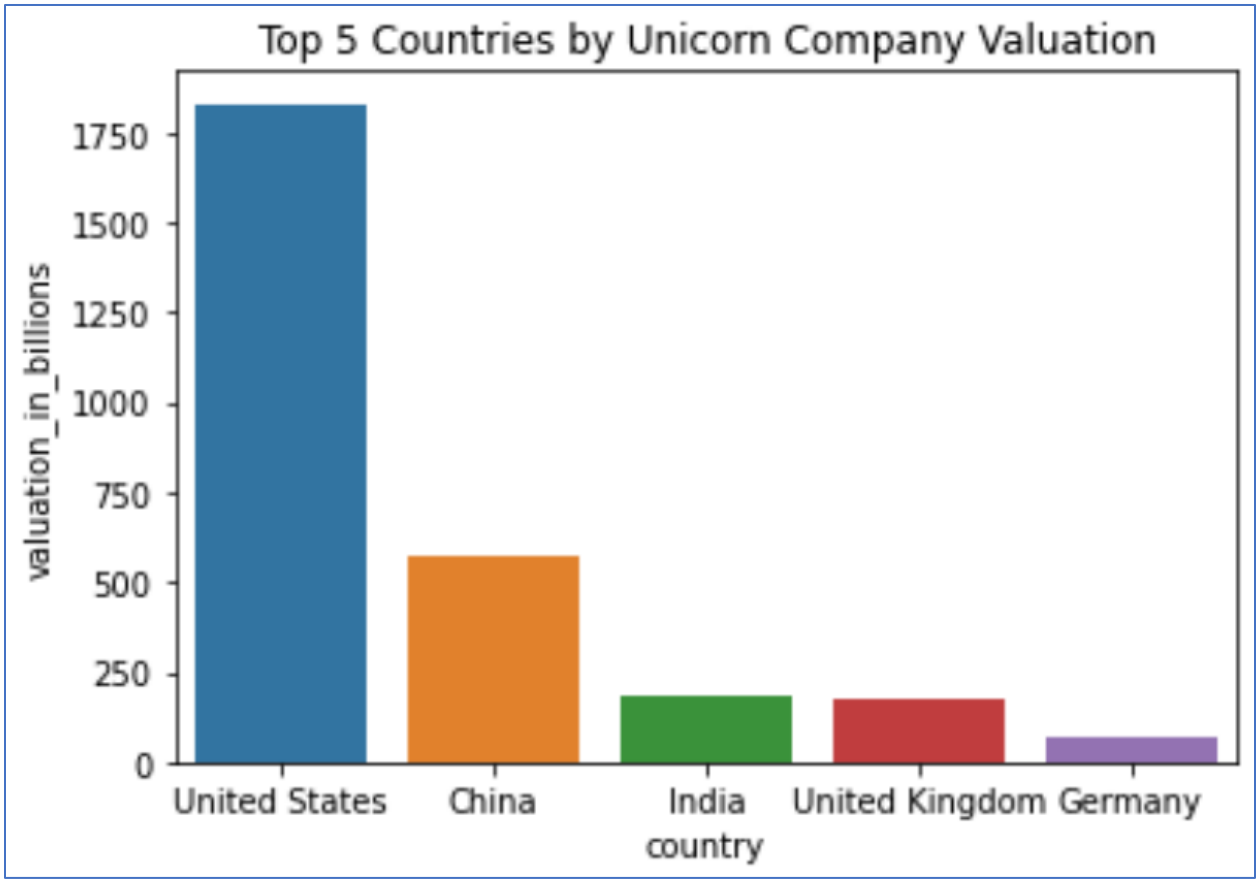
File Edit Search Source Run Debug Consoles Projects Tools View Help

...:\Users\Suman\OneDrive - Cal State LA\Desktop\5270 - Business Intelligence\Python Project\testing.py

Top_Industry_Heatmap.py | testing.py

```
# -*- coding: utf-8 -*-
"""
Created on Mon Apr 24 16:00:57 2023

@author: Suman
"""
import matplotlib.pyplot as plt
import seaborn as sns
import pandas as pd

# Load CSV file into pandas dataframe
df = pd.read_csv('Unicorn_Companies_Cleaned.csv')

# Calculate total unicorn valuation by country
country_valuation = df.groupby('country')['valuation_in_billions'].sum().reset_index()

# Sort dataframe by valuation in descending order
country_valuation = country_valuation.sort_values('valuation_in_billions', ascending=False)

# Take top 5 countries by valuation
top_5 = country_valuation[:5]

# Create bar plot using Seaborn
sns.barplot(x='country', y='valuation_in_billions', data=top_5)

# Set plot title and axis labels
plt.title('Top 5 Countries by Unicorn Company Valuation')
plt.xlabel('country')
plt.ylabel('valuation_in_billions')

# Display the plot
plt.show()
```






**Explanation**:

Through our analysis, we have identified a concentration of unicorn companies and their valuations in specific countries. The top five countries in this regard are the United States, China, India, the United Kingdom, and Germany. Notably, the United States stands out with a significantly higher valuation compared to the combined valuations of the other four countries. To perform this analysis, we utilized the groupby() function, which allows data to be grouped for more insightful computations and analysis. In 2021, the number of unicorn companies founded in China surpassed the combined count of 11 other countries, including Germany, India, and the UK. However, the country with the highest number of unicorns was the United States, featuring notable companies like SpaceX, Stripe, and Instacart, which boasted substantial market valuations.

2. **Top Industries by valuation and founding year.**

   **Code**

   import pandas as pd

   import matplotlib.pyplot as plt

   import seaborn as sns

```
# Load CSV file into pandas dataframe
df = pd.read_csv('Unicorn_Companies_Cleaned.csv')

# Calculate total unicorn valuation by industry and year
industry_year_valuation = df.groupby(['industry', 'founded_year'])["valuation_in_billions
'].sum().reset_index()

# Pivot dataframe to create heatmap data
heatmap_data = pd.pivot_table(industry_year_valuation, values='valuation_in_billions',
index=['industry'], columns=['founded_year'])

# Create heatmap using Seaborn
sns.heatmap(heatmap_data, cmap='PiYG')

# Set plot title and axis labels
plt.title('Total Unicorn Valuation by Industry and Founded Year')
plt.xlabel('founded_year')
plt.ylabel('industry')

# Display the plot
plt.show()
```

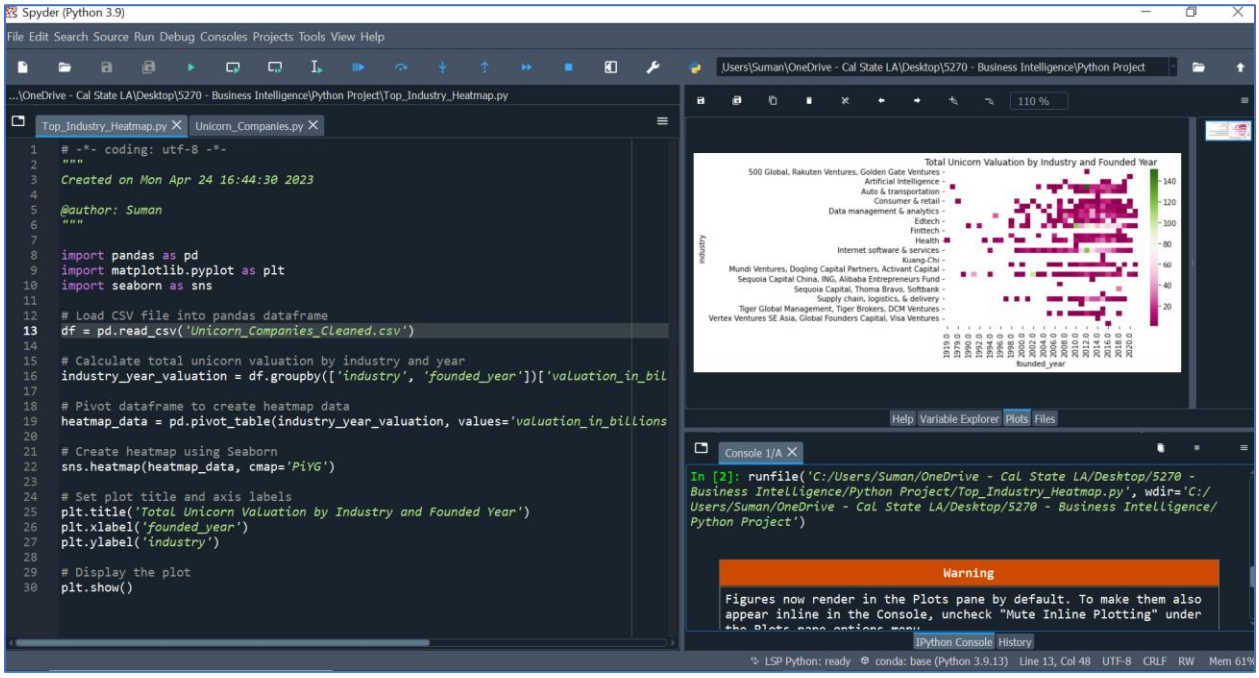
Spyder (Python 3.9)

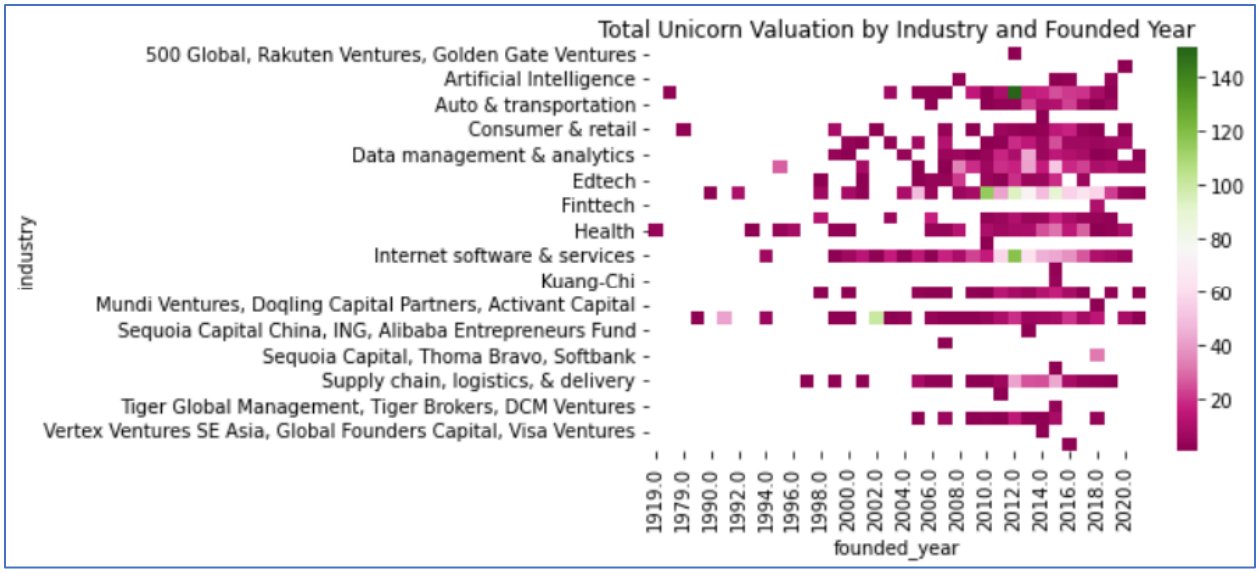
File Edit Search Source Run Debug Consoles Projects Tools View Help

...\OneDrive - Cal State LA\Desktop\5270 - Business Intelligence\Python Project\Top_Industry_Heatmap.py

Top_Industry_Heatmap.py | Unicorn_Companies.py

```
# -*- coding: utf-8 -*-
"""
Created on Mon Apr 24 16:44:30 2023

@author: Suman
"""

import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns

# Load CSV file into pandas dataframe
df = pd.read_csv('Unicorn_Companies_Cleaned.csv')

# Calculate total unicorn valuation by industry and year
industry_year_valuation = df.groupby(['industry', 'founded_year'])['valuation_in_billions'].sum().reset_index()

# Pivot dataframe to create heatmap data
heatmap_data = pd.pivot_table(industry_year_valuation, values='valuation_in_billions', index=['industry'], columns=['founded_year'])

# Create heatmap using Seaborn
sns.heatmap(heatmap_data, cmap='PiYG')

# Set plot title and axis labels
plt.title('Total Unicorn Valuation by Industry and Founded Year')
plt.xlabel('founded_year')
plt.ylabel('industry')

# Display the plot
plt.show()
```






**Explanation**

The provided code utilizes the groupby() function to analyze the relationship between industry and valuation. By grouping the data by industry and calculating valuations, a heatmap visualization is generated. The heatmap highlights the Fintech industry as having the highest total valuation, reflecting the significant growth of Fintech startups in recent years. Embracing fintech has become crucial for traditional banking institutions, as demonstrated by the integration of fintech into strategic models and the incorporation of emerging technologies in products and services. The heatmap provides a concise overview of valuation distribution across industries, offering insights into industry trends within the unicorn company landscape.

3. **Show the distribution of the top 10 industries among the Unicorn Companies?**

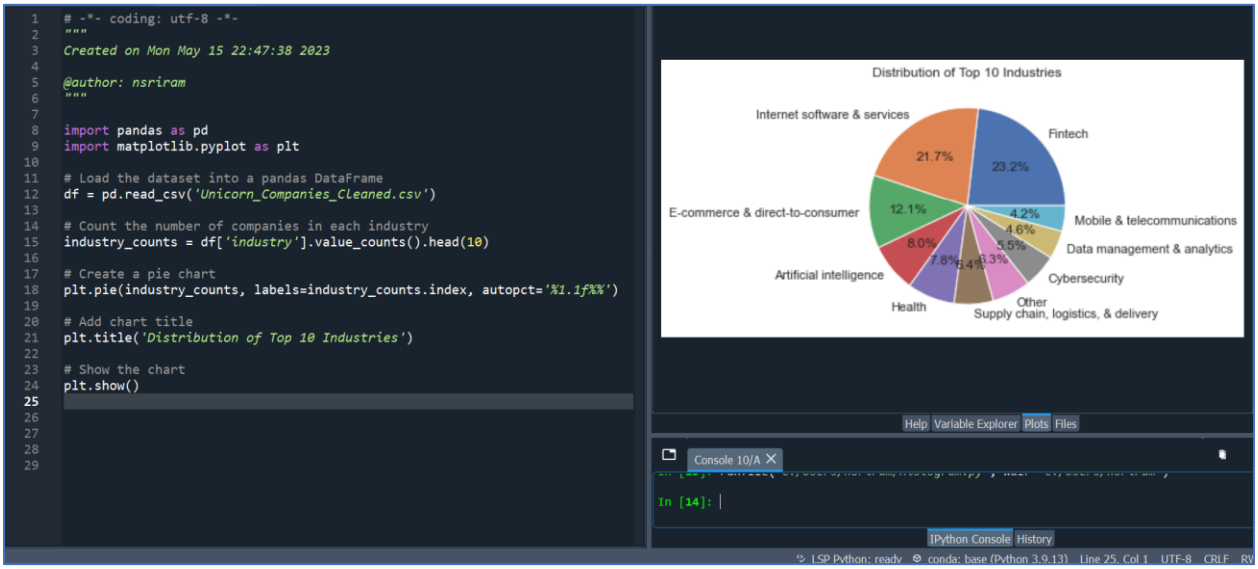
**Code**:

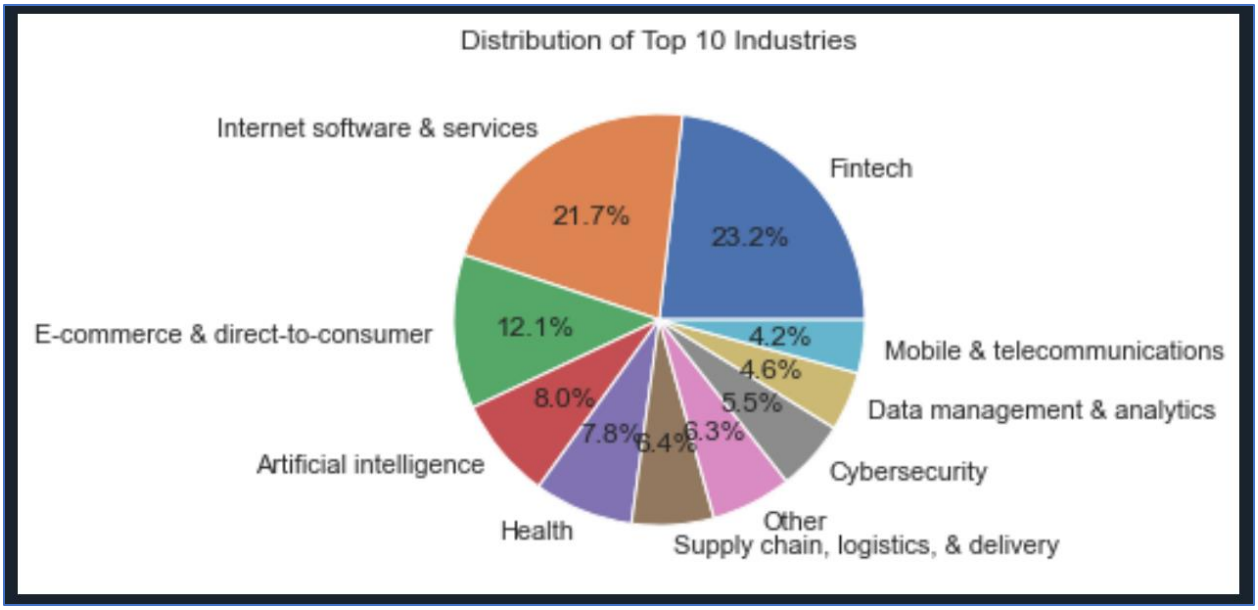
```
import pandas as pd
import matplotlib.pyplot as plt

# Load the dataset into a pandas DataFrame
df = pd.read_csv('Unicorn_Companies_Cleaned.csv')

# Count the number of companies in each industry
industry_counts = df['industry'].value_counts().head(10)

# Create a pie chart
plt.pie(industry_counts, labels=industry_counts.index, autopct='%1.1f%%')
```

# Add chart title
plt.title('Distribution of Top 10 Industries')

# Show the chart
plt.show()

```python
# -*- coding: utf-8 -*-
"""
Created on Mon May 15 22:47:38 2023

@author: nsriram
"""

import pandas as pd
import matplotlib.pyplot as plt

# Load the dataset into a pandas DataFrame
df = pd.read_csv('Unicorn_Companies_Cleaned.csv')

# Count the number of companies in each industry
industry_counts = df['industry'].value_counts().head(10)

# Create a pie chart
plt.pie(industry_counts, labels=industry_counts.index, autopct='%1.1f%%')

# Add chart title
plt.title('Distribution of Top 10 Industries')

# Show the chart
plt.show()
```






The above chart represents the distribution of the top 10 industries among unicorn companies in a dataset. The chart shows the percentage distribution of companies in each industry. Fintech stands out as the industry with the highest percentage, accounting for 23% of the total companies. Following closely is the Internet Software & Services industry, representing 21.7% of the companies. The Ecommerce industry also holds a notable position among the top 10, with 12.1%. The pie chart effectively illustrates the dominance of Fintech, the strong presence of Internet Software & Services, and the significant representation of the Ecommerce sector within the unicorn company landscape.

4. **What is the relationship between total raised and valuation for the top 10 unicorn companies?**

**Code**:

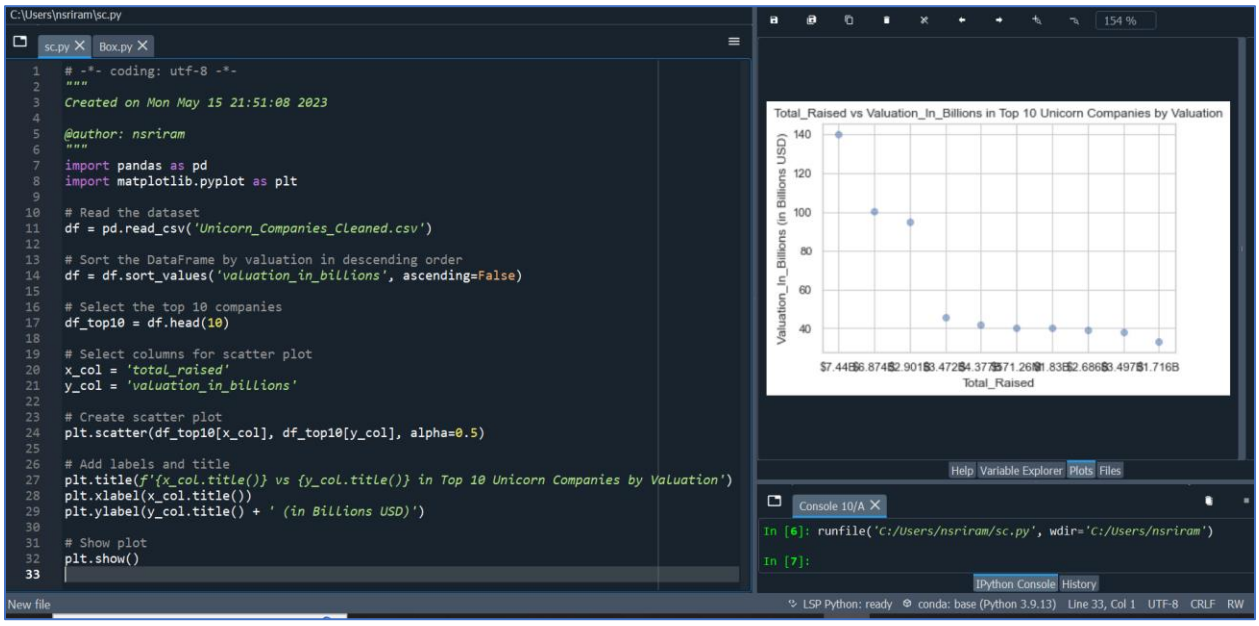
```
import pandas as pd

import matplotlib.pyplot as plt


# Read the dataset

df = pd.read_csv('Unicorn_Companies_Cleaned.csv')


# Sort the DataFrame by valuation in descending order

df = df.sort_values('valuation_in_billions', ascending=False)
```

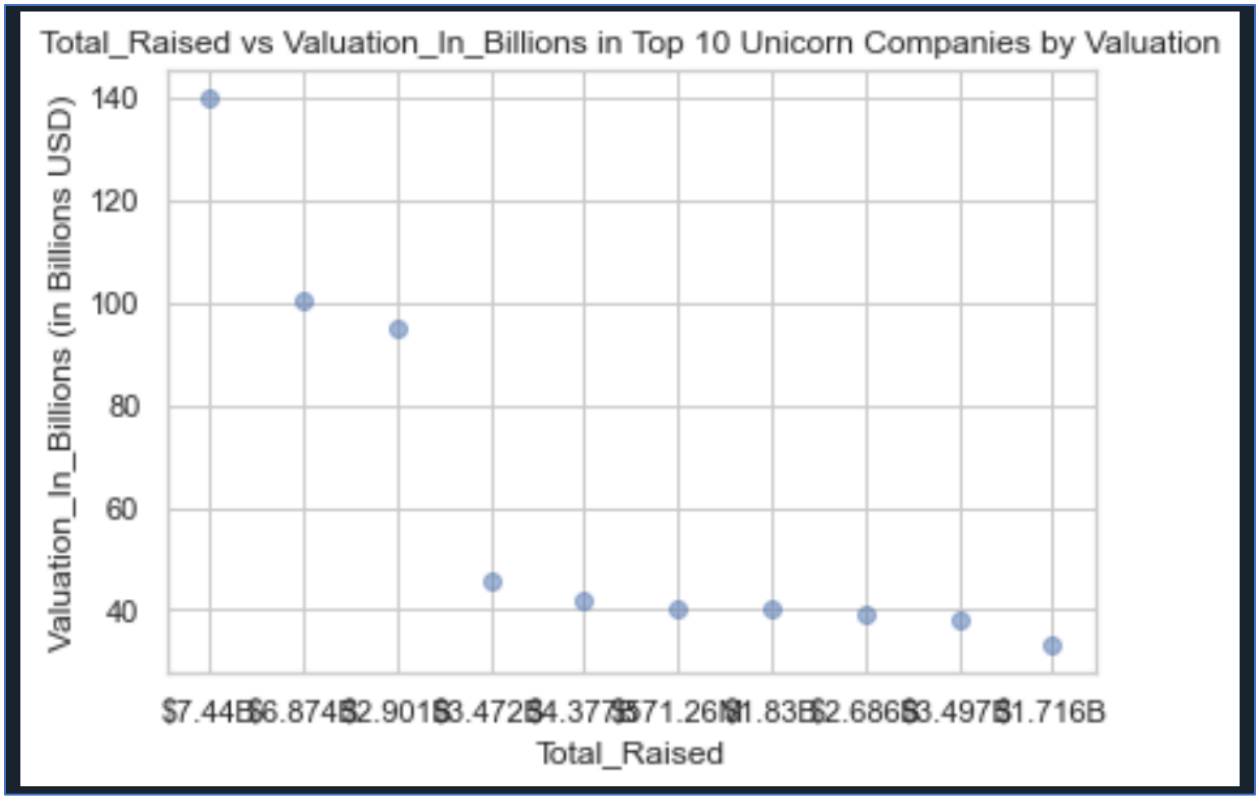
```
# Select the top 10 companies
df_top10 = df.head(10)

# Select columns for scatter plot
x_col = 'total_raised'
y_col = 'valuation_in_billions'

# Create scatter plot
plt.scatter(df_top10[x_col], df_top10[y_col], alpha=0.5)

# Add labels and title
plt.title(f'{x_col.title()} vs {y_col.title()} in Top 10 Unicorn Companies by Valuation')
plt.xlabel(x_col.title())
plt.ylabel(y_col.title() + ' (in Billions USD)')
# Show plot
plt.show()
```

C:\Users\nsriram\sc.py

sc.py | Box.py

```python
# -*- coding: utf-8 -*-
"""
Created on Mon May 15 21:51:08 2023

@author: nsriram
"""
import pandas as pd
import matplotlib.pyplot as plt

# Read the dataset
df = pd.read_csv('Unicorn_Companies_Cleaned.csv')

# Sort the DataFrame by valuation in descending order
df = df.sort_values('valuation_in_billions', ascending=False)

# Select the top 10 companies
df_top10 = df.head(10)

# Select columns for scatter plot
x_col = 'total_raised'
y_col = 'valuation_in_billions'

# Create scatter plot
plt.scatter(df_top10[x_col], df_top10[y_col], alpha=0.5)

# Add labels and title
plt.title(f'{x_col.title()} vs {y_col.title()} in Top 10 Unicorn Companies by Valuation')
plt.xlabel(x_col.title())
plt.ylabel(y_col.title() + ' (in Billions USD)')

# Show plot
plt.show()
```






The scatter plot illustrates the relationship between the total amount raised and valuation for the top 10 unicorn companies. For specific data points, when the valuation is 140 billion USD, the corresponding total amount raised is approximately $7.44 billion, and when the valuation is 100 billion USD, the total amount raised is approximately $6.874 billion. The scatter plot allows us to observe the overall trend and potential correlation between the total amount raised and valuation among the top 10 companies, providing insights into how the total amount raised may vary for different valuation levels.

5. **What is the distribution of company valuations by industry?**

**Code:**

```
import pandas as pd

import seaborn as sns

import matplotlib.pyplot as plt


# Load the dataset into a Pandas DataFrame

df = pd.read_csv("Unicorn_Companies_Cleaned.csv")
```

# Create a box and whisker chart to visualize the distribution of company valuations by industry

sns.set(style="whitegrid")

plt.figure(figsize=(12,8))

ax = sns.boxplot(x='industry', y='valuation_in_billions', data=df)

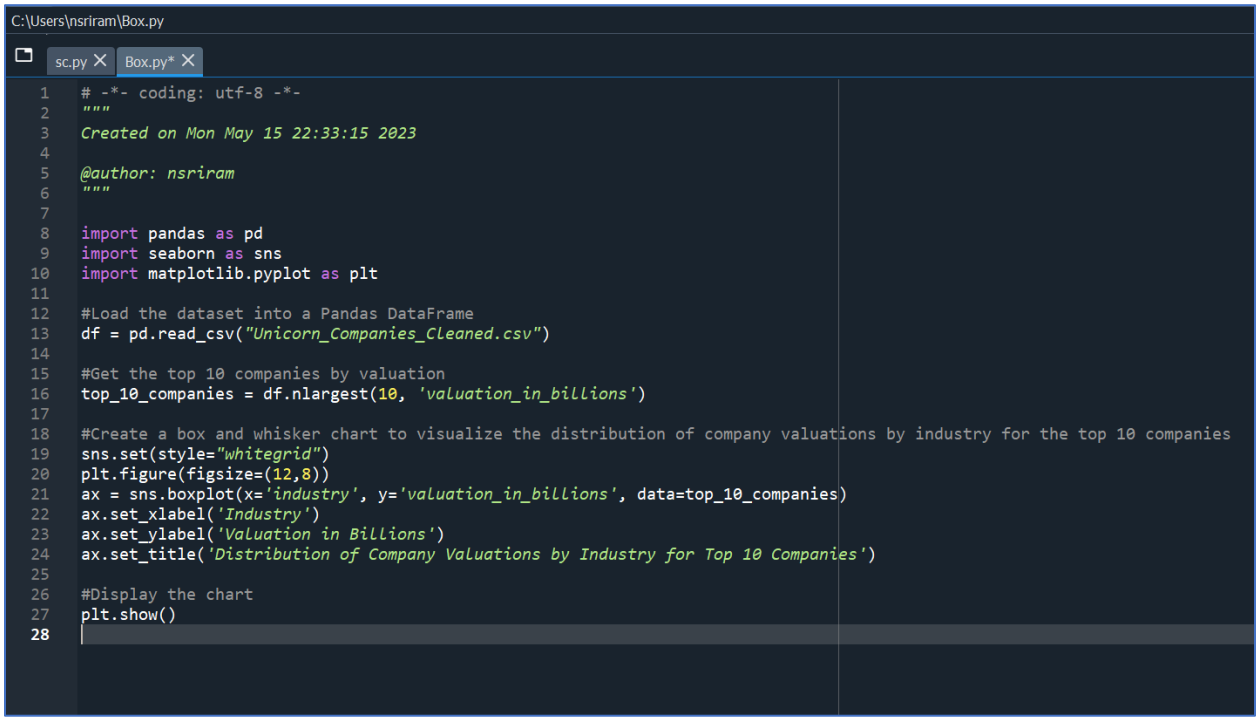
ax.set_xlabel('Industry')

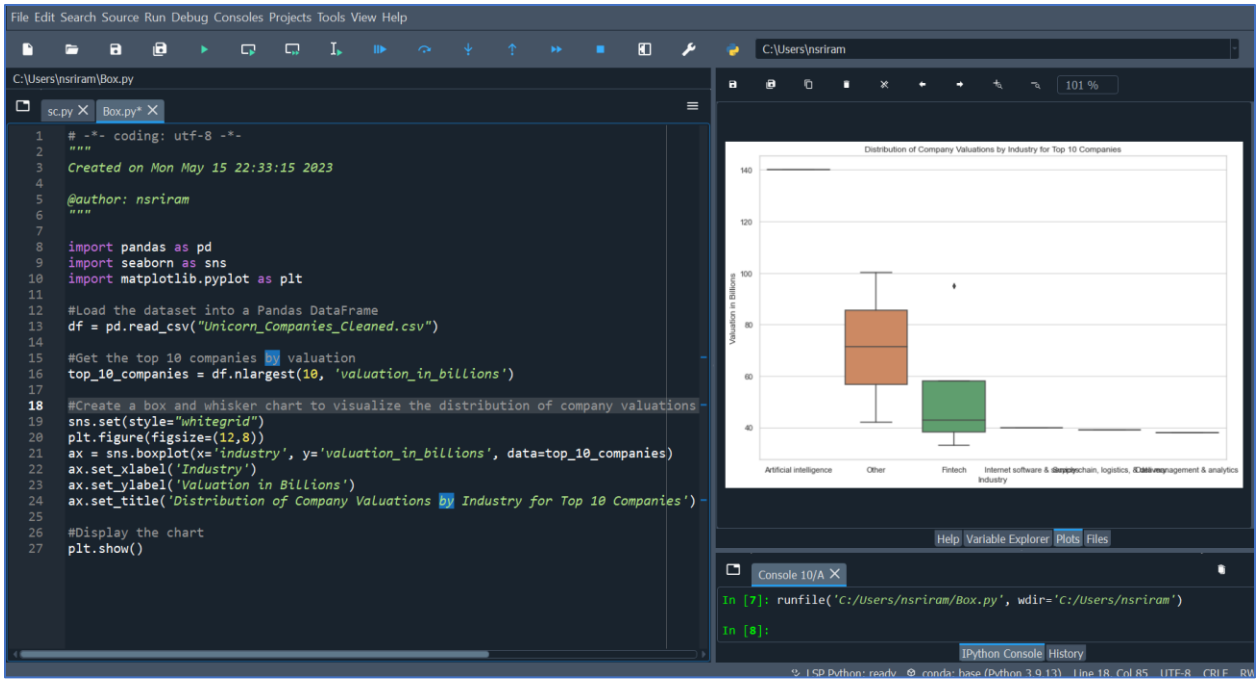
ax.set_ylabel('Valuation in Billions')

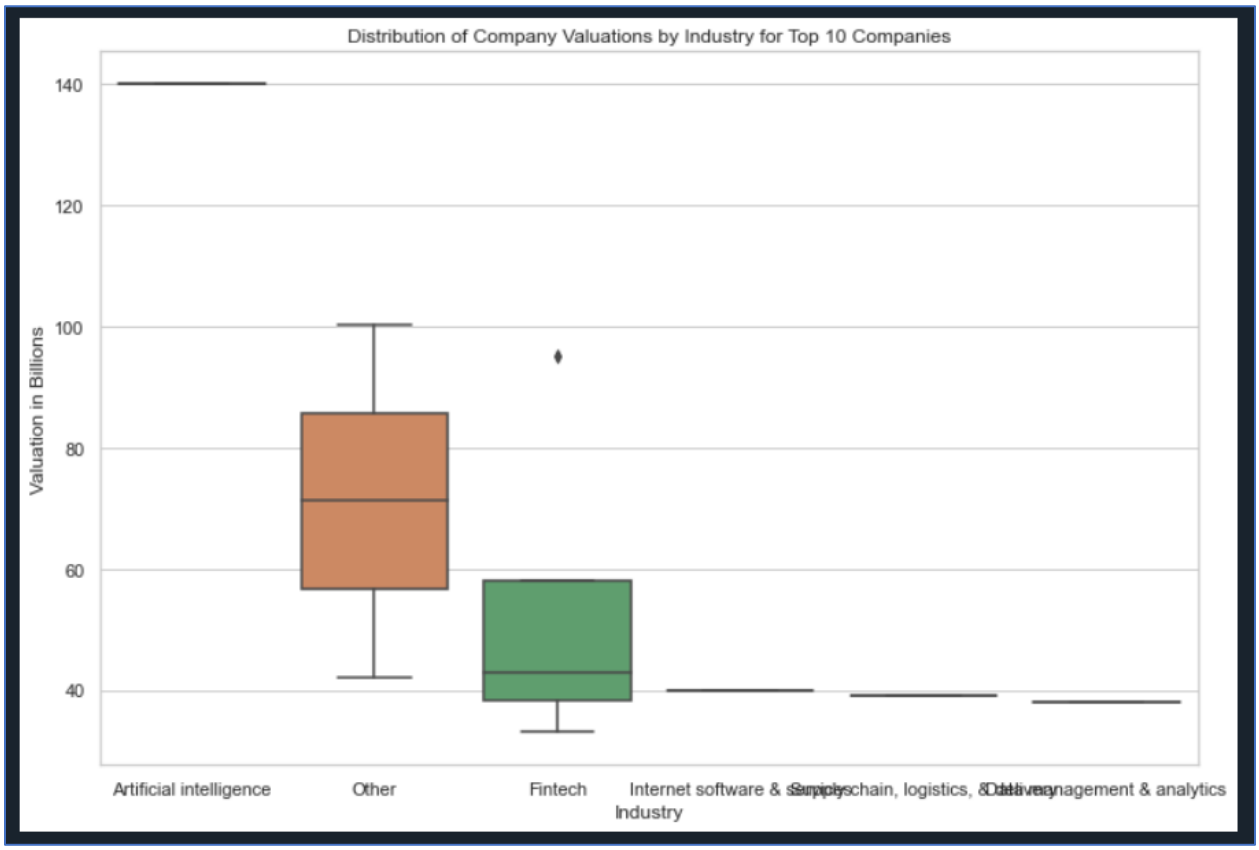
ax.set_title('Distribution of Company Valuations by Industry')

# Display the chart

plt.show()




File Edit Search Source Run Debug Consoles Projects Tools View Help
C:\Users\nsriram
C:\Users\nsriram\Box.py
sc.py
Box.py*
# -*- coding: utf-8 -*-
"""
Created on Mon May 15 22:33:15 2023
@author: nsriram
"""
import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt
#Load the dataset into a Pandas DataFrame
df = pd.read_csv("Unicorn_Companies_Cleaned.csv")
#Get the top 10 companies by valuation
top_10_companies = df.nlargest(10, 'valuation_in_billions')
#Create a box and whisker chart to visualize the distribution of company valuations
sns.set(style="whitegrid")
plt.figure(figsize=(12,8))
ax = sns.boxplot(x='industry', y='valuation_in_billions', data=top_10_companies)
ax.set_xlabel('Industry')
ax.set_ylabel('Valuation in Billions')
ax.set_title('Distribution of Company Valuations by Industry for Top 10 Companies')
#Display the chart
plt.show()
Distribution of Company Valuations by Industry for Top 10 Companies
Valuation in Billions
Artificial intelligence
Other
Fintech
Industry
Help
Variable Explorer
Plots
Files
Console 10/A
In [7]: runfile('C:/Users/nsriram/Box.py', wdir='C:/Users/nsriram')
In [8]:
IPython Console
History


Distribution of Company Valuations by Industry for Top 10 Companies
Valuation in Billions
140
120
100
80
60
40
Artificial intelligence
Other
Fintech
Industry

The chart visualizes the distribution of company valuations by industry in a dataset of unicorn companies. The chart reveals that the Fintech industry has a lower median valuation than the other industries, while the Other Industries category has the highest median valuation. The Internet Software and Services industry also exhibits a relatively high median valuation. The interquartile range (IQR) displayed in the chart indicates the spread of valuations within each industry, with the Other Industries category having the largest IQR and the Fintech industry showing a narrower range. Overall, the chart effectively showcases the variations in company valuations across different industries, emphasizing differences in median valuations and levels of variability.

## Citations


Hagiu, A. (2016, January). How Unicorns Grow. Harvard Business Review. Retrieved from https://hbr.org/2016/01/how-unicorns-grow

Ivana Pino. Fortune. (2022, December 20). What Is a Unicorn Company? Fortune. Retrieved from https://fortune.com/recommends/investing/what-is-a-unicorn-company/

Team, RingCentral. “Unicorn Startups: What They Are and How to Build One.” RingCentral, 30 Dec. 2022, www.ringcentral.com/us/en/blog/what-is-a-unicorn-startup/#:~:text=In%20the%20plainest%20terms%2C%20a%20unicorn%20startup%20is,with%20help%20from%20venture%20capitalists%20and%20other%20investors.