Data Analysis for Unicorn Companies

**Problem Statement**

Unicorn Companies must make decisions that will focus on companies with high growth potential, diversify investment portfolio and prioritize companies with experienced leadership teams.

**Steps**

Data Cleaning and Manipulation

* *Achievements*

Managed to upload Data Frame on the Jupiter Notebook and deduced the information of the data set using df.info () method. Columns Valuation, Date Joined, Year Founded, Funding had their data represented as ‘objects’. Achieved on converting the datatype of column Valuation to a ‘float’ using the codes:

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

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

df

Converted datatype of column Date Joined to Date-Time data type.

* Achieved replacing the missing values on column City and Select Investors by replacing with ‘unknown’ and imputing mode respectively.
* *Challenges*

Had challenges in converting the data type of Year founded to Date-Time datatype because the data set was specified as in ‘Year’ .it is impossible to convert a string to a float.

Had a challenge in converting the Funding Column datatype from object to float because the dataset contained string (‘unknown’)

Exploratory Analysis

When performing univariate and bivariate analysis the following was observed

1. Descriptive statistics – using the method. descibe()

Most companies were founded with one billion with a frequency as high as 60 companies over a century. Though funding went up to as far as 946M giving a mean of one billion on observation the funding of many companies multiplied over a short period of time (2020 to 2022) as compared to 1919 t0 2020.

The founded companies were able to join the Unicorn companies over the years though most of them remained with a valuation of one billion giving a total of 471 companies.

1. Top 5 Investors -using the code value\_count().head

Sequoia Capital is the top Investor amongst all investing on a total of 3 companies.

1. Top 5 Countries – plotted a bar graph to determine the following;

United Staes of America has the most Investors having San Fransisco and New York being the leading cities where the investments are.

1. Summary Statistics on Year Founded

First company was founded in 1919 by 2014 50% of the companies were founded .in 2021they reached their maximum.

1. Top Industries with a high Valuation

Cyber security industry has the most Valuation followed by Mobile and telecommunication. Though Fintech rates third Industry has more companies invested on it.

Visualization

Achieved plotting line graph showing Companies versus the year they joined the Unicorn Companies.

Used a scatter plot to show different types of companies when they were Founded in comparison with their Valuation.

Illustrated the top 4 Industries in a pie chart

On determining the number of occurrences of top companies in major cities, used a bar graph and a heatmap. Considered the heatmap over the bar graph because it could present the information better.

On finding the top 10 industries used a bar graph. Aiming to visualise the possible investing industries besides the top 5.

Using a heatmap, the number of companies were shown across all Continents.

To be able to recommend the most experienced companies ,used a bar graph showing the top 5 companies years of experience and their count on that very year.