INSIGHTS PROJECT

Amazon Job Analysis Mini-Project

Data-Set Description:

It is a dataset including information on amazon job opening around the world from June 2011 to March 2018. This dataset is collected using Selenium and BeautifulSoup by scraping all of the jobs for Amazon job site.

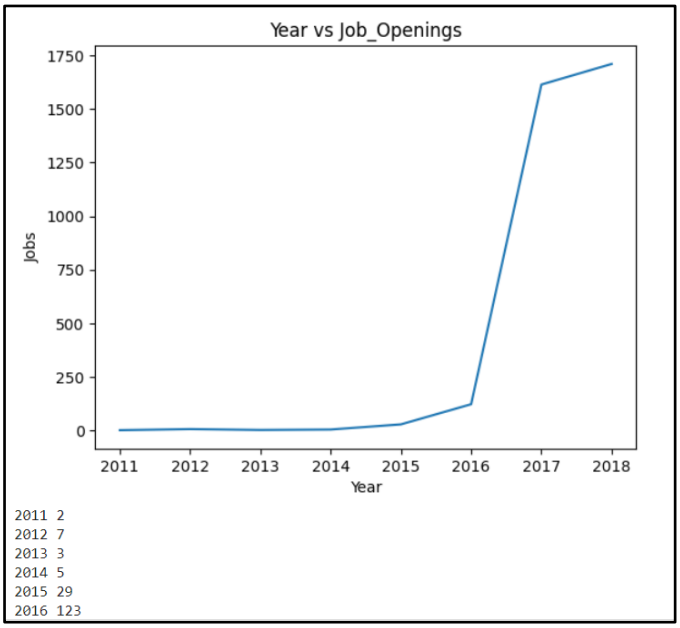
Question-1:

Plot the line graph between no. of Job postings with respect to year. Print the year and the number of job posting as integer value.

```
'''
FIRSTLY, I CONVERTED THE DATA FROM CSV FILE TO DICTIONARY FOR EASE IN DATA PROCESSING.
DID SUMMATION OF THE FREQUENCY OF INDIVIDUAL YEARS.
LASTLY, I CHOSE LINE GRAPH AS IT IS THE MOST SUITED ONE TO GAIN THE INSIGHTS.
'''
import csv
import matplotlib.pyplot as plt
import collections
with open('amazon_jobs_dataset.csv', encoding ='UTF-8') as file_obj:
    file_data = csv.DictReader(file_obj, skipinitialspace=True)

    dct = {}
    for row in file_data:
        date = row['Posting_date'].split()
        key = date[2]
        if key in dct:
            dct[key] += 1
        else:
            dct[key] = 1

    ord_dct = collections.OrderedDict(sorted(dct.items()))
    plt.plot(list(ord_dct.keys()),list(ord_dct.values()))
    plt.xlabel("Year")
    plt.ylabel("Jobs")
    plt.title('Year vs Job_Openings')
    plt.show()

    for i in ord_dct.keys():
        print(i,end=" ")
        print(ord_dct[i])
```

Insight-1 Code

Output-1:



Output of above Question

Insights:

Based on the analysis of the data, it is evident that there has been a significant increase in the number of job openings starting from the year 2011. The most substantial growth occurred in the years 2015 and 2016. Prior to 2015, the number of job openings remained relatively constant from 2011 to 2014. However, this trend shifted when Amazon began establishing offices in India in 2015, resulting in a significant increase in the number of job openings.

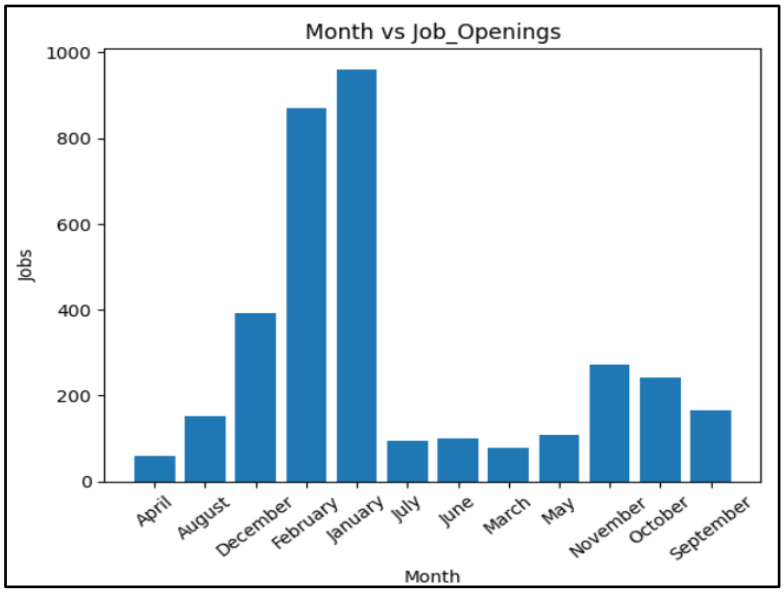
Question-2:

Plot the Bar graph between Month vs Job Openings. Print the month name and the number of job posting as integer value. Order of months doesn't matter.

```
'''
SAME AS ABOVE, FIRSTLY GENERATED THE DICTIONARY.
INSTEAD OF YEARS, NEED TO FIND OUT THE NUMBER OF JOB OPENINGS ON MONTHLY BASIS.
'''
with open('amazon_jobs_dataset.csv', encoding ='UTF-8') as file_obj:
    file_data = csv.DictReader(file_obj, skipinitialspace=True)

    dct = {}
    for row in file_data:
        date = row['Posting_date'].split()
        key = date[0]
        if key in dct:
            dct[key] += 1
        else:
            dct[key] = 1

    ord_dct = collections.OrderedDict(sorted(dct.items()))
    plt.bar(list(ord_dct.keys()),list(ord_dct.values()))
    plt.xlabel("Month")
    plt.ylabel("Jobs")
    plt.title('Month vs Job_Openings')
    plt.xticks(rotation = 40)
    plt.show()

    for i in ord_dct.keys():
        print(i,end=" ")
        print(ord_dct[i])
```

Code to find the answer of Question-2

Output-2:



Output of the Question-2

```
April 58
August 153
December 393
February 869
January 961
July 95
June 99
March 78
May 108
November 271
October 243
September 165
```

Output of the Question-2

INSIGHTS:

Based on our analysis, it is apparent that the number of job openings consistently increased from October to February, regardless of the year. This trend can be attributed to several factors. In the United States, the fiscal year begins in October, and in India, the fiscal year ends in March. Therefore, this period marks a time of significant shuffling and attrition in the job market, resulting in an increase in job openings. Additionally, many students in India begin their job search around October, which also contributes to the rise in job openings during this period. Therefore, it is crucial to consider these factors when analyzing the job market trends during this time of the year.

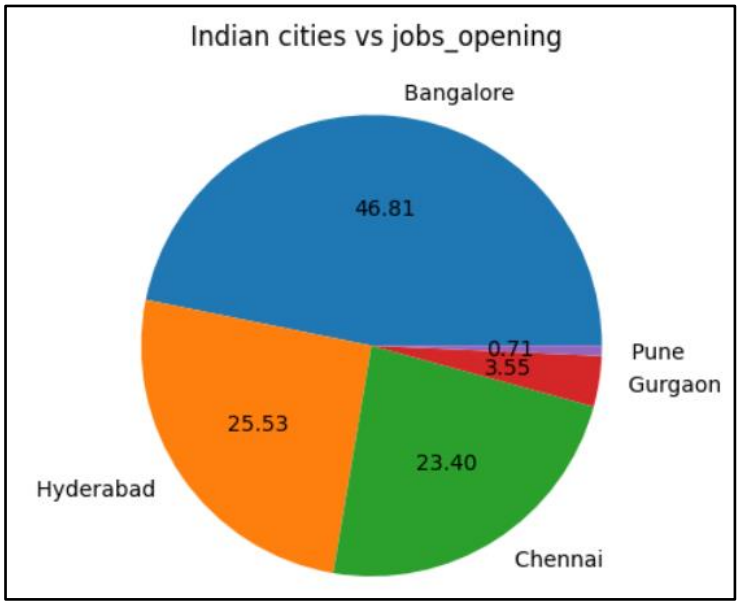
Question-3:

Plot the Pie chart between Indian cities vs No. of job openings.

```
1
2 with open('amazon_jobs_dataset.csv', encoding ='UTF-8') as file_obj:
3     file_data = csv.DictReader(file_obj, skipinitialspace=True)
4
5     dct = {}
6     for row in file_data:
7         country = row['location'].split(',')[0]
8         if country == 'IN' :
9             key = row['location'].split(',')[2]
10            if key in dct:
11                dct[key] += 1
12            else:
13                dct[key] = 1
14
15    ord_dct = collections.OrderedDict(sorted(dct.items()))
16    plt.pie(dct.values(),autopct='%0.2f',labels=dct.keys())
17
18    plt.title(' Indian cities vs jobs_opening')
19    plt.xticks(rotation = 40)
20    plt.show()
```

Code for Question-3

Output:



Output of Question-3

Insights:

Based on our analysis, it is apparent that the majority of job openings, specifically more than 45 percent, are from the Bangalore office. This comes as no surprise as Bangalore is not only the headquarters of the company in India but also known as the Silicon Valley of India, attracting a larger talent pool. The second highest number of job openings is from Hyderabad, which is also an emerging Silicon Valley in India. The remaining offices have a minor contribution to the overall job openings.

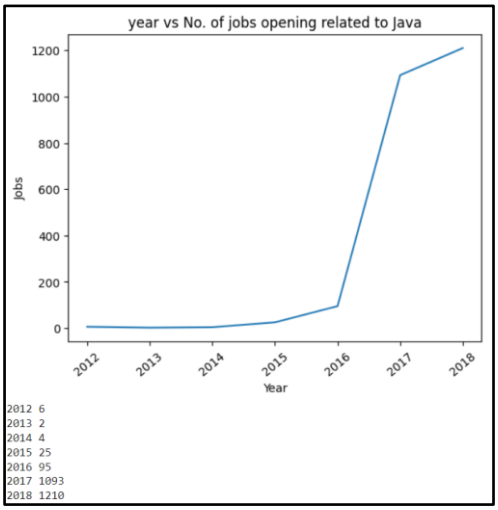
Question-4:

Plot the scatter graph between year vs No. of jobs opening related to Java. Print the year and number of Jobs opening in Java Profile.

```
with open('amazon_jobs_dataset.csv', encoding ='UTF-8') as file_obj:
    file_data = csv.DictReader(file_obj, skipinitialspace=True)

    dct = {}
    for row in file_data:
        qlfn = row['BASIC QUALIFICATIONS']
        if 'Java' in qlfn or 'java' in qlfn :
            key = row['Posting_date'].split()[2]
            if key in dct:
                dct[key] += 1
            else:
                dct[key] = 1

    ord_dct = collections.OrderedDict(sorted(dct.items()))
    plt.plot(list(ord_dct.keys()),list(ord_dct.values()))

    plt.xlabel("Year")
    plt.ylabel("Jobs")
    plt.title('year vs No. of jobs opening related to Java')
    plt.xticks(rotation = 40)
    plt.show()

    for i in ord_dct.keys():
        print(i,ord_dct[i])
```

Code for Question-4

Output-4:



Output for Question-5

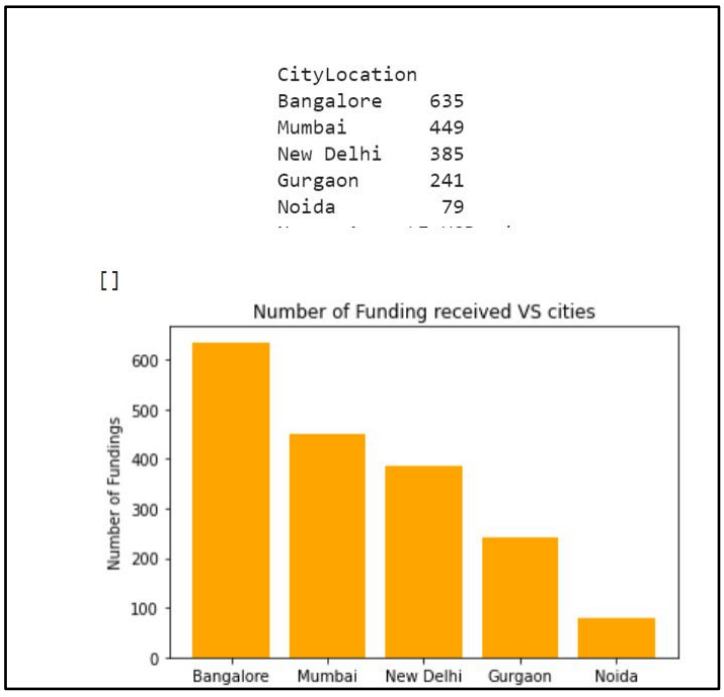
Insights:

Upon careful analysis, it is evident that as the company expands and grows, the number of job openings is likely to increase. Furthermore, it is observed that jobs requiring Java skills are in high demand in India, which is not surprising given the importance of Java in the tech industry.

Start-Up Funding Problem

Question-1:

Your Friend has developed the Product and he wants to establish the product startup and he is searching for a perfect location where getting the investment has a high chance. But due to its financial restriction, he can choose only between three locations - Bangalore, Mumbai, and NCR. As a friend, you want to help your friend deciding the location. NCR include Gurgaon, Noida and New Delhi. Find the location where the greatest number of funding is done. That means, find the location where startups have received funding maximum number of times. Plot the bar graph between location and number of funding. For few startups multiple locations are given, one Indian and one Foreign. Consider the startup if any one of the city lies in given locations.



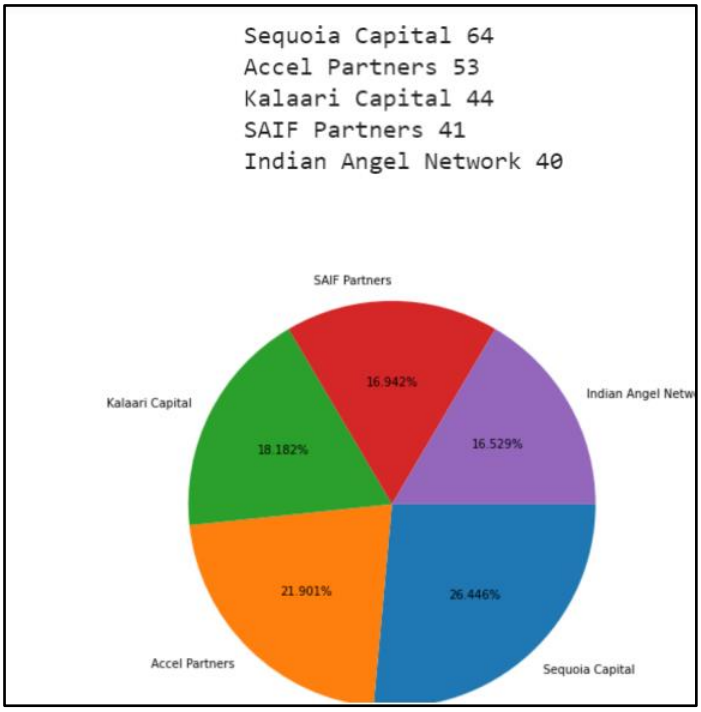
Output of Question-1

Insights:

Based on the analysis of funding data, it is evident that Bangalore has received the highest amount of funding compared to other cities. However, it is important to note that if we combine the total fundings received in Delhi, Gurgaon, and Noida (collectively known as NCR), the sum amounts to 705. Hence, when we compare the individual cities, Bangalore has received the highest funding, but when we compare Bangalore, Mumbai, and NCR as regions, NCR emerges as the clear winner. Therefore, I recommend my friend to consider NCR as it will provide more opportunities to explore as it comprises three cities. This can lead to better networking and collaboration opportunities, ultimately leading to more significant growth and development.

Question-2:

Even after trying for so many times, your friend's startup could not find the investment. So you decided to take this matter in your hand and try to find the list of investors who probably can invest in your friend's startup. Your list will increase the chance of your friend startup getting some initial investment by contacting these investors. Find the top 5 investors who have invested maximum number of times (consider repeat investments in one company also). In a startup, multiple investors might have invested. So, consider each investor for that startup. Ignore undisclosed investors.



Output of Question-2

Insights:

After analyzing the data, it has become evident that there are certain investors who have heavily invested in multiple startups. Therefore, I would highly recommend my friend to approach these top 5 investors in order to seek funding. It is important to note that this recommendation is based on the data analysis conducted on the available information regarding the investors.

I must highlight that due to the undisclosed nature of some investments, there may be other investors who have invested heavily but are not represented in the data. Nonetheless, based on the available data, the top 5 investors who have invested in multiple startups are listed below:

- Sequoia Capital
- Accel Partners
- Kalaari Capital
- SAIF Partners
- Nexus Venture Partners

These investors have a track record of investing in startups across various sectors and have a reputation of providing strategic support to their portfolio companies. Therefore, approaching them for funding may prove to be highly beneficial for my friend's startup.