Business Intelligence and Visualization

CA_ONE

Data Analysis & Visualization Report

NAME: MAUSAM SUDHIR JADHAV

STUDENT NO: 10605724

INTRODUCTION:

The following report entitles the data analysis and visualization of the Indian Start ups spread sheet obtained from the Open data source websites.

OBJECTIVES:

To visualize and analyze the Startups dataset to gain meaningful insights and explain them accordingly

SOURCE AND GENERAL INFORMATION ON DATASET:

- The **Startups** Excel file was obtained from the Kaggle data source website. It provides information about the Indian Startup companies and their general details like the amount invested.
- The original dataset is in Excel spreadsheet format and it consists of 8 columns and 665 rows.
- Following are the titles of the columns and their details:
    1. **Company/Brand**: It gives us the details of various startups and brands registered in India. Each row in this column consist of different Startup
    2. **Founded**: It signifies the year at which a particular startup was founded. It's basically ranging from year 1989 to 2021
    3. **Headquarter**: It gives us the location at which the headquarters of particular startup was founded. There are total of 41 locations across India at which Headquarters of different companies are situated.
    4. **Sector**: It gives us the sector or area of interest at which a particular startup belongs.
    5. **Founders:** It gives us the names of the founder or CEO's of Startups.
    6. **Investor**: It gives us the data on Investors names. It can be name of a Enterprise or a person's name.
    7. **Amount**: It is the amount or capital invested in this startup companies. It is denoted in Indian Rupees and has values ranging from one Lakh to around 1 billion.

8. **Stage:** It signifies the Investment stage at which the startup is currently positioned depending on the interest level of potential investors. For example "Seed" funding which is one of the term in stage column that signifies close friends or family, it can be also called as angel funding.

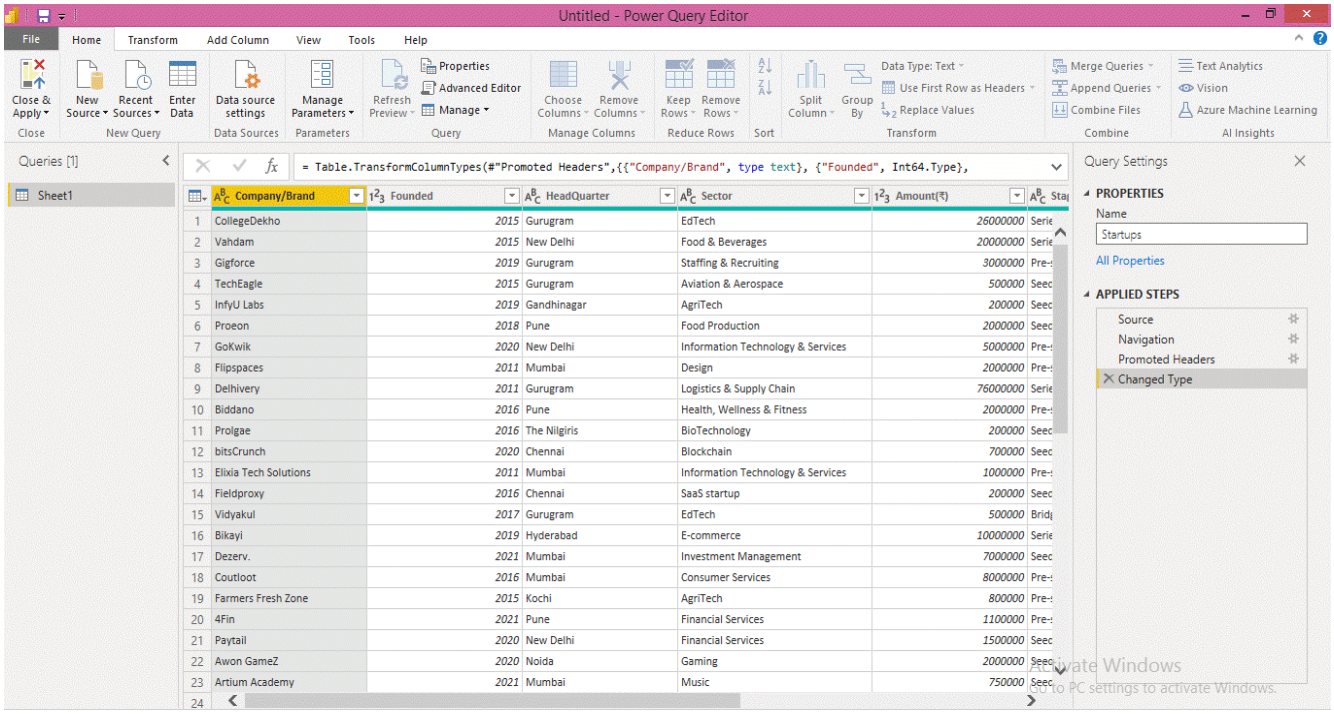
As the dataset obtained from kaggle is raw and unfiltered, we have to do the further cleaning process to get the proper filtered dataset. So we load the data in power BI tool for further cleaning.

Power BI is interactive Data visualization software developed by Microsoft for the sole purpose of Business Insights. It consists of various visualization techniques and Query Editor to perform various tasks on dataset.

DATA VISUALIZATION, CLEANING AND INTERPRETATION ON POWER BI:

Intial Steps on Power BI:

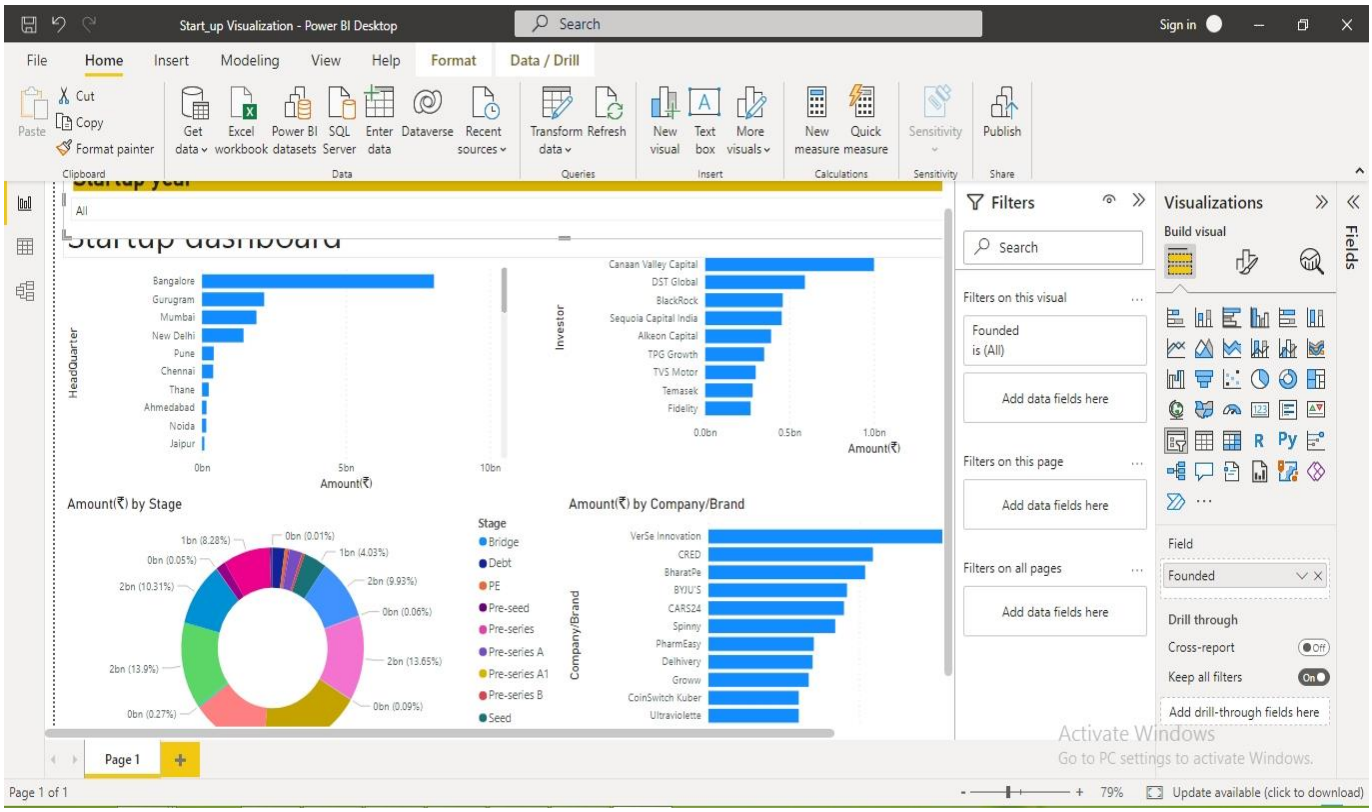
- Firstly we load the dataset by selecting **GetData** Tab on power BI.
- After that we select transform data and then load the data after which the table appears on Power BI query editor as shown in the image below.



- Now that we've connected to a data source, we can shape the data in query editor as per our needs.
- There are various options on the query editor through which we can filter the unwanted rows and remove columns or delete columns and all the steps we perform are recorded by the query editor in the applied steps section as shown in the above figure.
- We also change the data types of columns by right-clicking on particular column and then change type, so as to assign them appropriate data type.
- After that we check if there is any illogical or irrelevant data in the sheet and we remove them by filtering options if required.
- Because of the cleaning process the number of rows is considerably reduced to 658.
- After that we select **Close & Apply** from the Home ribbon and the data gets loaded into Power BI Desktop where we can perform visualizations.

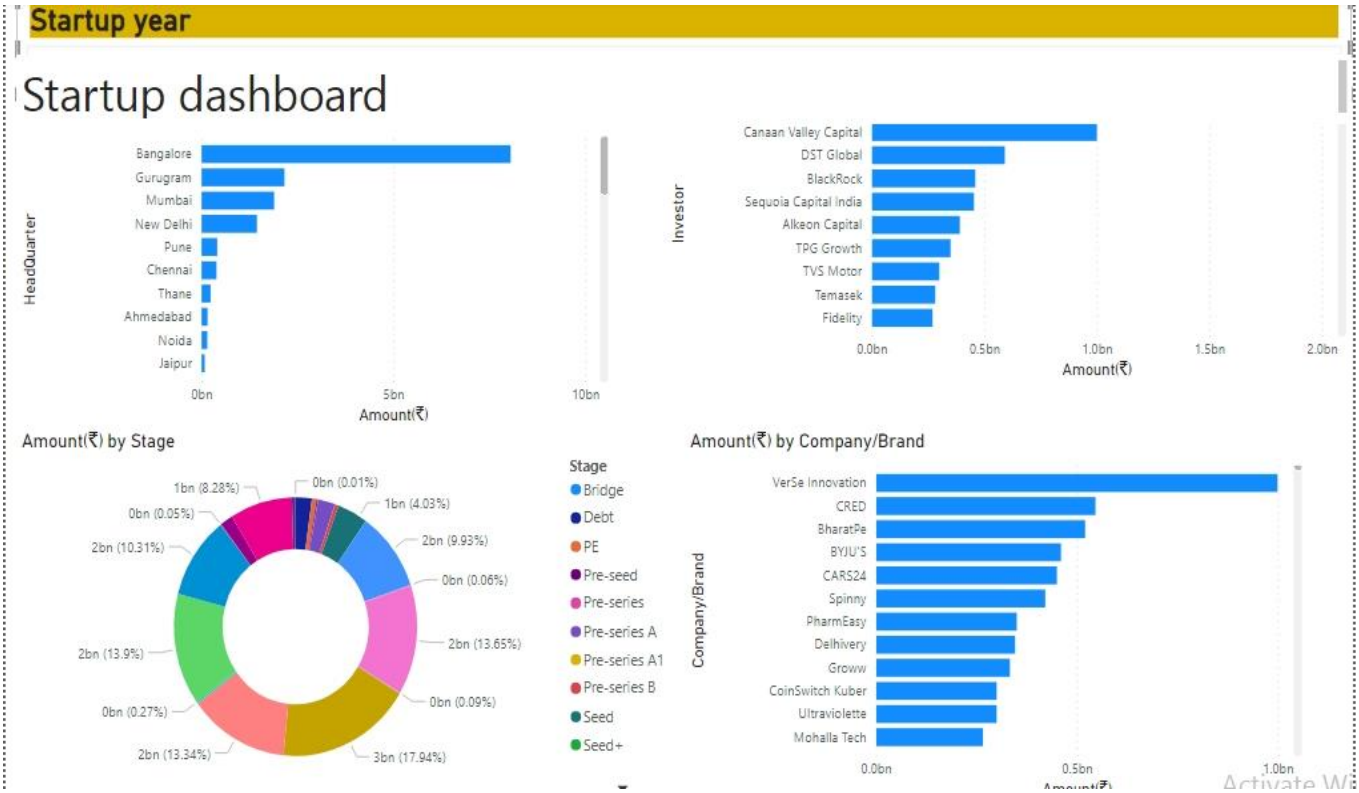
Visualizations on Power BI:

After performing the cleaning process on Power BI, we commence the visualization process. There are various pallets and tabs on the Power BI tool. There is visualizations palette which consists of various bar charts and fields to place particular columns on horizontal or vertical axis. The Dataset is available just beside visualization as see in the picture below.



So using the visualization tools we build 4 charts among which one is donut chart and 3 stacked bar charts as presented on the dashboard. We also install slicer function on the dashboard which is a sort of filter function for the years at which start ups were founded and with this we can highlight the statistics at particular year ranging from year 1989 to 2021.

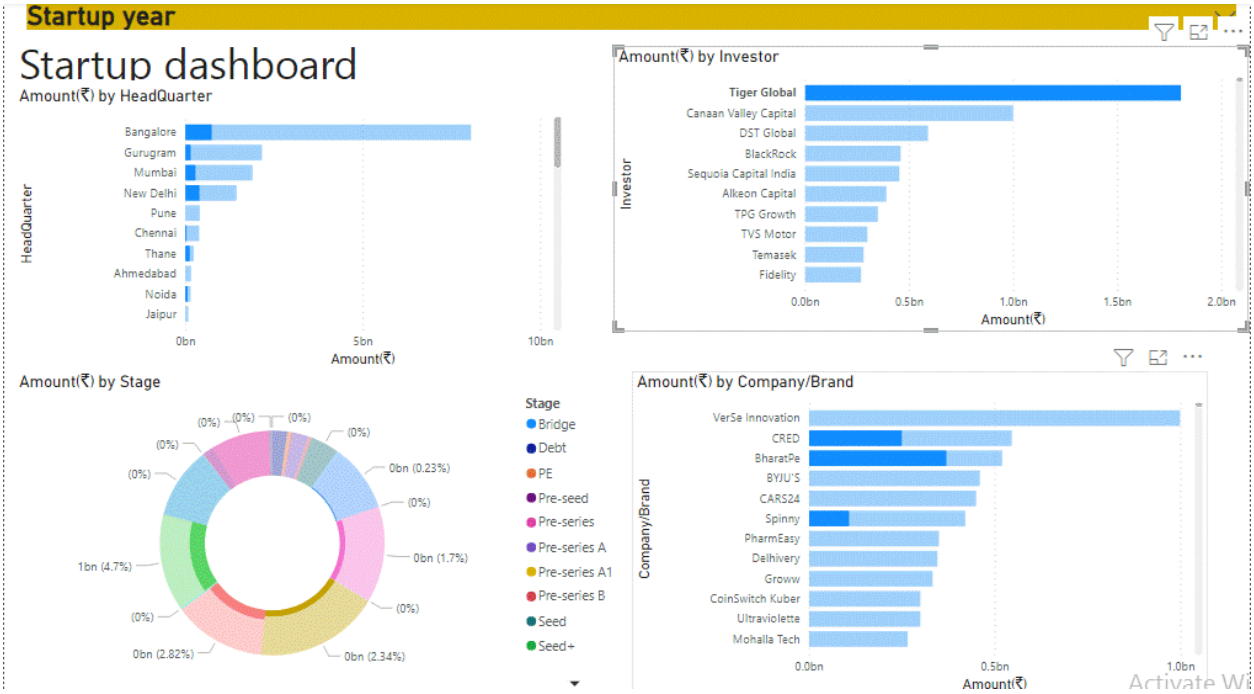
We can see the zoomed in version of the visualization dashboard below:



As seen in the above figure we have there are three stacked bar charts and each have amount variable in x-axis and we are plotting multiple variable against amount in those bar charts. So from Amount by Company/Brand chart we observe that Amount(₹) invested was highest for VerSe Innovation at 1000000000 rupees, followed by CRED and BharatPe. The VerSe Innovation belongs to Innovation Management sector and it was founded in year 2007. Also for the rest 586 Company/Brand, the Amount(₹) ranged from 100000 to 1000000000. Also we get to know that VerSe Innovation accounted for 6.52% of Amount(₹) and the company that invested in Verse was Canaan valley capital which is located at Bangalore and it was funded at pr-series stage.

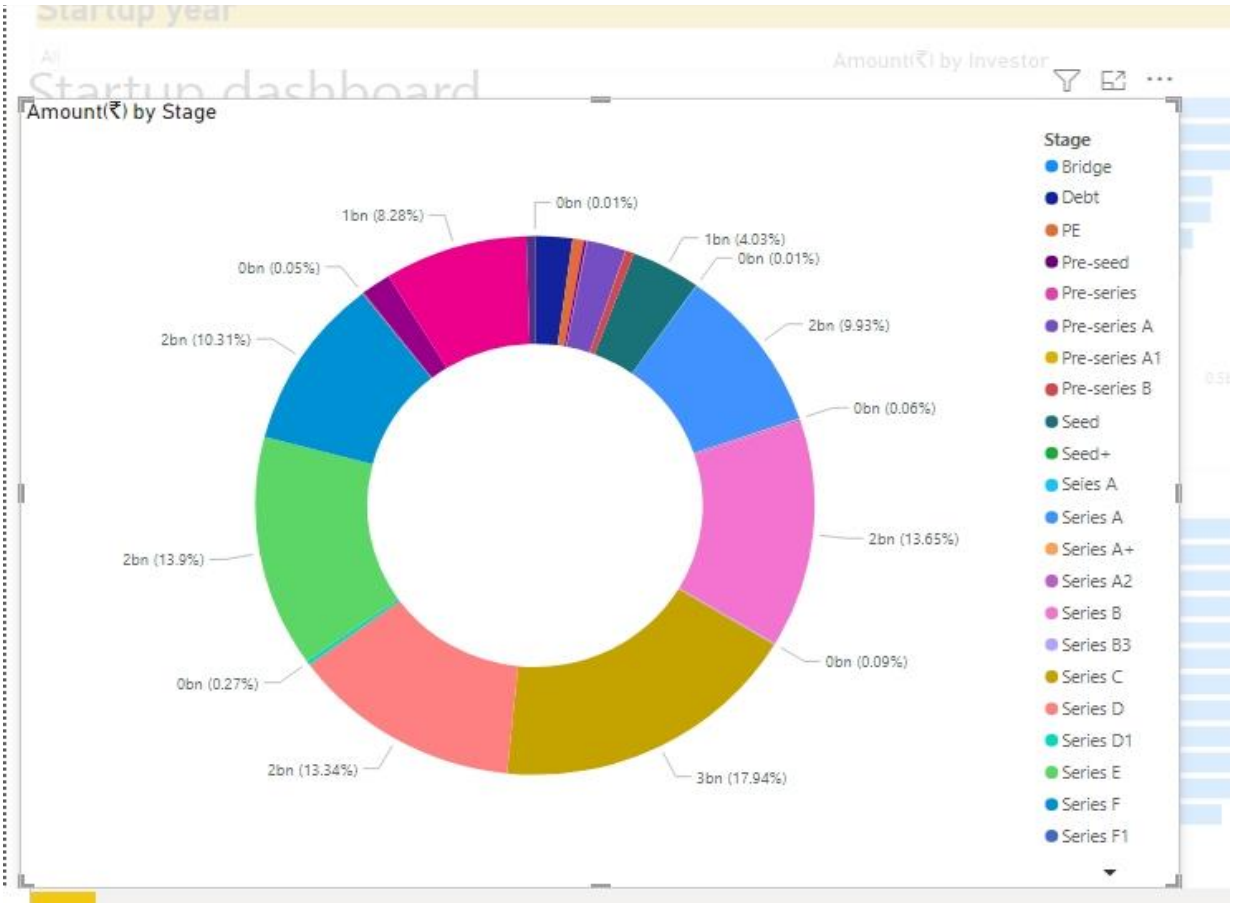
From the first chart in the top left which is amount by Headquarter, we observe that the total Amount(₹) invested was highest for Bangalore city which is 8,05,4015000 billion rupees, followed by Gurugram and Mumbai City. Also Bangalore city accounted for 52.53% of the Amount(₹).

Looking at the Amount by Investor Bar chart as given in figure below we observe that Tiger Global Investment Firm contributed the largest capital as investment for various Startups.



It accounted to the sum total of 1,807,000000 Amount(₹),followed by Canaan Valley Capital and DST Global as second and Third highest investment firms. The Companies/Brand That Tiger Global Invested in were M2P Fintech, Progcap, Classplus, GoMechanic, Spinny, Neobank Jupiter, Bizongo, infra.market, BharatPe, Apna, Porter, CRED, Slice, G.O.A.T, and Captain Fresh. Tiger Global accounted for 11.79% of Amount(₹) and there were variety of sectors in which the amount was invested.

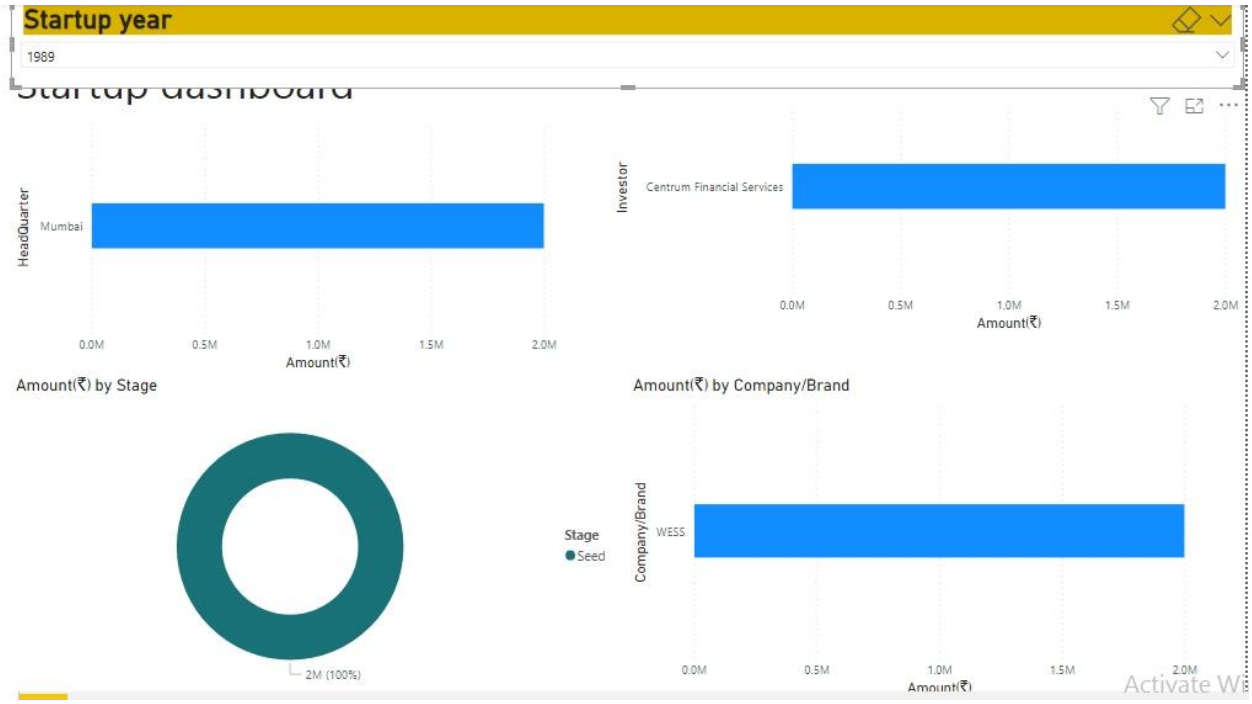
Next we analyze the Donut Chart which represents the stage at which the funding takes place while plotting it against the Amount(₹) as given in Figure below.



So as we can see from above figure that all the stages are denoted by certain colors and all of them represent a certain section of donut chart based on the amount percentage. Also we observe that Series C accounted for 17.94% of total Amount(₹) which is the highest. So this means that most of the investors injected capital in the successful businesses, in an effort to receive more returns.

We also add slicer function given in visualization palette in Power BI and we assign it to filter by year. So by using this we can get the corresponding visualizations or particular year.

For example, for the year 1989 we get the visualizations as given in the dashboard image below:



So, as we can see from above in 1989 only one company was founded i.e. WESS and its Investor was Centrum Financial Services. Its location is at Mumbai and the total amount invested was 2 million.

So, this is all that we can interpret about the Startup file from the visualization Dashboard.

CONCLUSION:

I faced lots of challenges while acquiring the dataset as there was lot on conundrum as to what dataset to choose and what is the best source. Some of the datasets were very complicated and making their visualization was bit of a task.

After carrying out lot of research I was able to get the appropriate dataset on Kaggle which is an online community platform or website which allows users to find and publish dataset and do variety of other tasks.

Next challenge was to find a appropriate tool or platform for visualization. Now, there are lot of visualizations tool some are very user-friendly, while some are too complicated. I had to choose the tool which can give even better visualization and is also easy to use.

So, I thought it is best to use power BI because it is very simplistic and flexible. Additionally we can extract data from multiple sources and perform powerful dashboard visualizations which also includes geo-mapping.

So Power BI enabled me to gain various meaningful insights on this data, and it can be further documented and can be used by the government agencies or investors to keep track of the Startups and which sector is booming with the large number of startup.

In this case we can observe that most of the companies belonged to Agritech Sector. We also get to know which company has the highest amount invested and this information can be used to attract further investors. It can be also usedby goverment agencies to invest in that startup if it has potential and use the returns for further social welfare.

We also get to know the year at which the startups were founded and from that we can derive whether we are progressing or not. However, in this case we can observe that the rate at which startups are being founded is increasing every year and it is a good sign of development for India.

REFERENCES:

https://www.investopedia.com/articles/personal-finance/102015/series-b-c-funding-what-it-all-means-and-how-it-works.asp

https://www.kaggle.com/datasets?tags=11102-Businesshttps://elearning.dbs.ie/pluginfile.php/1558110/mod_resource/content/1/Tutorial%201%20-%20Power%20BI%20Updated.pdf